THE EFFECTS OF AROMATICITY GAIN IN MULTIPOINT HYDROGEN-BONDED ARRAYS

A Thesis Presented to

the Faculty of the Department of Chemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

By

Yu Zhang

December 2018

THE EFFECTS OF AROMATICITY GAIN IN MULTIPOINT HYDROGEN-BONDED ARRAYS

Yu Zhang

APPROVED:

Dr. Judy I. Wu, Chairman

Dr. Jakoah R. Brgoch

Dr. Arnold M. Guloy

Dr. Jeremy A. May

Dr. James M. Briggs

Dr. Dan E. Wells, Dean

College of Natural Sciences and Mathematics

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Judy Wu for your support during my two and a half years' study at the University of Houston. I am very grateful for your advice. Thank you for your trust to let me help Jowa complete his project and give me so many potential projects to try. By working with you, I have enough freedom to explore my own ideas, which is really the happiest thing for me. I deeply appreciate that you allow me to chase dream and give enough support to me. I can't imagine what myself and my research would be without your guidance, encouragement, and support.

I am very grateful to have Dr. May, Dr. Guloy, Dr. Brgoch, and Dr. Briggs in my committee. I have benefited a lot from your classes. You also helped me a lot during my thesis preparation. I also want to thank Dr. Teets and Tang (UW-Madison). I got lots of help from you both in classes and during my preparation for application.

I am lucky to meet so many friends in Houston. It is you guys to help me overcome all my depressed days. I am honored to have worked with: Jowa, Ranjita, Zhili, Mahsa, Lucas, David, Hari, and Sia. It's also my pleasure to get so many useful suggestions from: Chen, Ziyan, Ya, Yanxin, Hao, Mengfan, Xihong, Xiaojing, Meng, Yuwei, Yujia, and Qinxuan.

Lastly, my family is always the most important part in my life. My deepest appreciation goes to my parents, who have supported me unconditionally all the time through my life. I am lucky in the past couple years to know some of my best friends: Yunpu, Wenze, Huarui, Zhefu, Yulian, and Kejing. We have shared much happiness and pain. I wish all of you a bright future.

THE EFFECTS OF AROMATICITY GAIN IN MULTIPOINT HYDROGEN-BONDED ARRAYS

A Thesis Presented to

the Faculty of the Department of Chemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

By

Yu Zhang

December 2018

ABSTRACT

The aromaticity-modulated hydrogen bonding (AMHB) model was applied to understand how "aromaticity gain" influences the association strengths of multipoint hydrogen-bonded arrays and as a driving force for electronic complementarity in base pairing. The block-localized wavefuntion (BLW) method was used to quantify the degree of "aromaticity gain" (*i.e.*, the amount of increased cyclic π-electron delocalization) in arrays upon hydrogen bonding. An excellent linear relationship was found between the computed gas-phase association free energies and the amount of increased cyclic π-electron delocalization energies of 26 triply (r = 0.940) and 20 quadruply (r = 0.959). Computational analyses for 57 hydrogen-bonded base pairs also document excellent linear correlation between the gas-phase association energies and the degree of aromaticity gain of paired bases (r = 0.949). Hydrogen bonding interactions can polarize the ring π-electrons to increase (or decrease) cyclic $4n + 2$ π-electron delocalization, resulting in aromaticity gain (or loss) in complexes, and become strengthened (or weakened). Our findings point to important limitations of the secondary electrostatic interaction (SEI) model, suggesting the importance of considering aromaticity gain in arrays as a relevant factor for determining the stability of multipoint hydrogen-bonded complexes. This work shows that aromaticity gain increases the inherent association strengths of hydrogen-bonded complexes. Potential implications of the AMHB model for improving nucleic acid force-fields and for designing robust unnatural base pairs are also discussed.

CONTENTS

LIST OF ABBREVIATIONS

AMHB = Aromaticity-modulated hydrogen bonding

BLW = Block-localized wavefunction

SEI = Secondary electrostatic interaction

$-\Delta G_{assoc}$ = Association free energy

NICS = Nucleus independent chemical shifts

DE_{π} = π-electron delocalization energy

ΔDE_{π} = Increased π-electron delocalization energy

Ψ_{deloc} = Fully electron delocalized wavefunction

Ψ_{loc} = π-electron localized wavefunction

UPy = Ureidopyrimidone

NBO deletion = Natural bond orbital deletion

K_{assoc} = Association constant

$-\Delta E$ = Base pairing interaction energy

$-\Delta\Delta E$ = Interaction energy difference

ΔESP = Electrostatic potential difference map

PCM = Polarizable continuum model

IEF-PCM = Polarizable continuum model using the integral equation formalism variant

ΔNBO-DEL_{π} = Increased π-electron delocalization energy calculated by NBO deletion

A = Adenine

T = Thymine

U = Uracil

G = Guanine

C = Cytosine

In the AAA-DDD, AAD-DDA, ADA-DAD, AADD-DDAA, ADDA-DAAD, ADAA-DADD and ADAD-DADA series of hydrogen-bonded arrays, A indicates proton acceptor and D indicates proton donor.

CHAPTER 1

INTRODUCTION

Hydrogen bonding, with its many unique roles in chemistry, has attracted the attention of chemists for more than 100 years. Since the hydrogen bond participates in most intermolecular recognitions, *e.g.*, the recognition properties of nucleobase pairs,[1-4] enzyme-substrate interactions,[5] and drug-receptor interactions,[6] it has been recognized as a highly important bonding interaction in biochemical systems. From the structure of water clusters[7,8] to the secondary and tertiary structures of proteins and nucleic acids, the robustness of all these structures are dominated by the strengths of hydrogen bonding interactions. Fully understanding molecular features that might significantly influence the strengths of hydrogen bonding interactions will contribute to the design of new materials, to explain solvent effects, and to understand the structures of proteins.

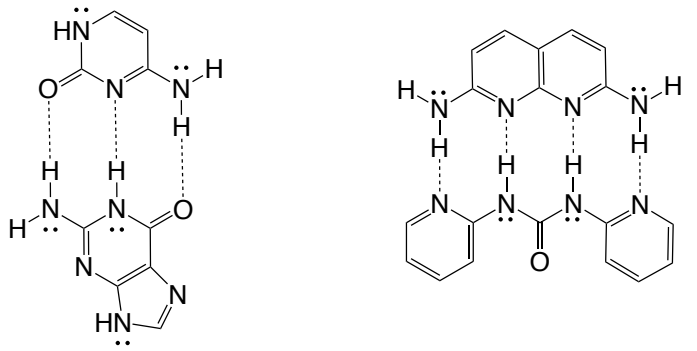
With the potential to be used in different fields of chemistry, multipoint hydrogen-bonded arrays stand out due to their self-assembling characters and strong binding strengths.[9-12] Such attractive features open new research topics, for example, using hydrogen-bond-mediated non-covalent polymers to design stimuli-responsive materials.[10] Kass and coworkers[5] demonstrated the importance of hydrogen-bonded arrays from another aspect, showing that a large energetic stabilization caused by hydrogen-bonded network may lead to catalytic rate enhancements. Wilson and coworkers[13] also reported using hydrogen-bonded arrays to tune self-assembled elastomers. Therefore, it is very important to understand key factors that influence the strengths of hydrogen bonding interactions.

1.1 Numbers and strengths of hydrogen bonds

It is well known that the stability of the hydrogen-bonded array depends on the numbers of hydrogen bonding interactions. For example, arrays with three hydrogen bonds are stronger than those arrays with two hydrogen bonds. Two hydrogen-bonded arrays are stronger than those arrays with only one hydrogen bond. However, although a good linear correlation was found between the association free energy and the number of hydrogen bonds,[14] it is still improper to evaluate the association strengths only by the hydrogen bond numbers since there are many other characters that matter a lot, for example, the pre-organization, secondary electrostatic interaction (SEI), substituent effects, and tautomerization.[10]

G•C base pairing and the Lüning[15] complex model (**Fig. 1**) are given to illustrate the limitations of evaluating the association strengths only by hydrogen bond numbers. Three and four hydrogen bonds exist in G•C pair and the Lüning complex model, respectively. Since the Lüning complex model contains one more hydrogen bond than G•C pair, the $-\Delta G_{assoc}$ values of the Lüning complex model is supposed to be higher than that of G•C pair. However, our computational results show the reverse trend, pointing to important limitations of common sense that more hydrogen bonds, more stable the structure is. There must be some other characters also influence the hydrogen bonding strengths other than the numbers of hydrogen bonds.

Therefore, comparing the association strengths of hydrogen-bonded complexes only by the numbers of hydrogen bonds is not always reliable. More factors should be considered when evaluating the association strengths of hydrogen bonding interactions.



G•C
$-\Delta G_{assoc}$ = 16.51 kcal/mol

Lüning complex model
$-\Delta G_{assoc}$ = 5.9 kcal/mol

Fig. 1 $-\Delta G_{assoc}$ values for G•C base pair and the Lüning complex model in the gas-phase. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level.

1.2 Secondary electrostatic interaction (SEI) model

In 1990, Jorgensen and Pranata proposed the secondary electrostatic interaction model,[16] demonstrating that adjacent proton donor and acceptor that are significantly close to each other have an effect on the association strength of hydrogen-bonded complex. It has long guided the understanding of multipoint hydrogen-bonded arrays and their association. According to the SEI model, those hydrogen-bonded arrays with all hydrogen bond donors on one fragment and all acceptors on the other fragment are the most robust, since this arrangement maximizes attractive electrostatic interactions. Thus, the strengths of triply hydrogen-bonded arrays should follow the order: AAA-DDD > AAD-DDA> ADA-DAD (see **Fig. 2**). (In the AAA-DDD, AAD-DDA, ADA-DAD hydrogen-bonded arrays, A indicates proton acceptor and D indicates proton donor.)

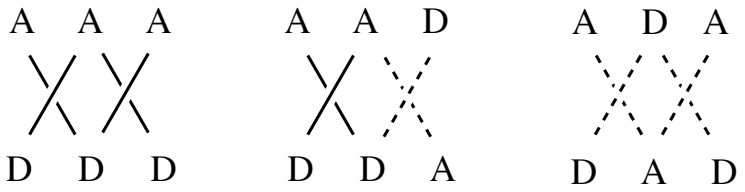
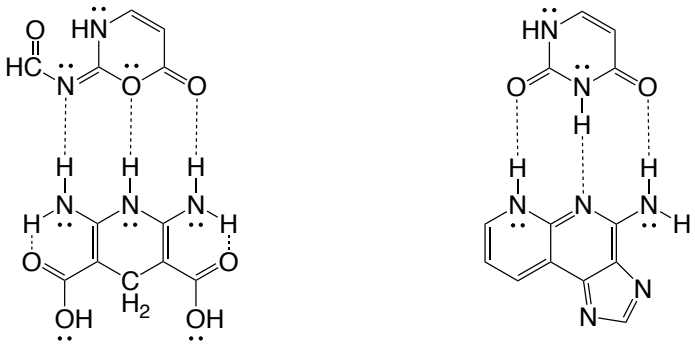


Fig. 2 Schematic illustration of the secondary electrostatic interaction model for triply hydrogen-bonded arrays. (– lines indicate attractive interactions, --- lines indicate repulsive interactions)

However, according to our computational results, this model limits the understanding of hydrogen-bonded arrays. For example, the selected AAA-DDD and ADA-DAD complexes (**Fig. 3**), which exhibit the same two NH...O bonds and one NH...N hydrogen bond but different patterns, violate the strength order suggested by the SEI model. Even with all hydrogen bond donors on one side and all acceptors on the other side and additional intramolecular hydrogen bonding interactions, the $-\Delta G_{assoc}$ value of the selected AAA-DDD complex is still much lower than that of the selected ADA-DAD complex.



AAA-DDD
$-\Delta G_{assoc}$ = 1.43 kcal/mol

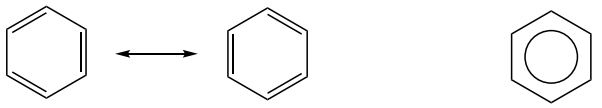
ADA-DAD
$-\Delta G_{assoc}$ = 9.29 kcal/mol

Fig. 3 $-\Delta G_{assoc}$ values for selected AAA-DDD and ADA-DAD complexes in the gas-phase. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level.

Our computed results show that the SEI model cannot be used to fully understand the association strengths of multipoint hydrogen-bonded arrays. Except the numbers, types and patterns of hydrogen bonds, here should be other factors that have significant effects on their complexation.

1.3 Aromaticity-modulated hydrogen bonding (AMHB) relationship

As one of the most important concepts in organic chemistry, aromaticity has been widely studied since 1855.[17] Although the term "aromaticity" has been proposed for more than 150 years, a precise definition is still missing so far. The most widely accepted way to determine whether the compound is aromatic or not, is by the Hückel's rule. According to the Hückel's rule, aromatic compounds should always adopt cyclic, planar structures with $4n + 2$ π-electron delocalization effect, which could highly stabilize the molecular structures. The best example to illustrate aromaticity is benzene (**Fig. 4**).[29] Due to the six π-electron delocalization effect in the six-membered planar ring, all C-C bond lengths in benzene are 1.39 Å between single bond length (1.47 Å) and double bond length (1.34 Å), and each bond angle is 120°. The direct evidence of aromaticity existing in benzene ring is to compare the heats of hydrogenation between similar compounds. If we compare 1,3-cyclohexadiene and benzene, the heats of hydrogenation for benzene should be higher than 1,3-cyclohexadiene since the number of double bonds is more in benzene. However, experimental results suggest that benzene is more stable than 1,3-cyclohexadiene. This stabilization can be attributed to the aromaticity of benzene, which is missing in 1,3-cyclohexadiene.



Kekulé resonance forms of benzene Delocalized D_{6h} benzene

Fig. 4 Model structures of benzene. Kekulé resonance forms of benzene (left) and delocalized D_{6h} benzene (right).

Aromaticity has been widely used to explain the behavior of organic molecules. It is attractive for chemists to solve significant chemical problems by applying quantum chemical approaches to measure aromaticity quantitatively. In 2005, Boldyrev and Wang[18] summarized the criteria for π-aromaticity and π-antiaromaticity based on different properties including electronic nature, energy, geometry, magnetic properties, reactivity and spectroscopy. Some popular methods of quantifying aromaticity include nucleus independent chemical shifts (NICS)[19] and block-localized wavefunction (BLW)[20-22] method. NICS is a computational method that can effectively measure the magnetic properties caused by the ring π-electrons delocalization, reflecting its aromatic character. Evaluation of π-electron delocalization energy is another way to measure aromaticity. BLW methods can artificially eliminate the undesired contaminating energetic effects to measure the energy difference between enabling and disabling the π-electrons delocalization. Both of them can give a direct comparison of different aromatic compounds, which would help to understand, interpret and quantify the abstract aromatic character.

In 1945, Dewar[23] first realized that the intramolecular hydrogen bonding interaction in tropolone (**Fig. 5**) could increase the π-electrons delocalization to enhance the aromatic character of nonbenzenoid rings. Frontera[24] and Storer[25] proposed that the superior

hydrogen bond donating ability of squaramide over urea-based organocatalysts could be explained by aromatic gain. Kieltyka and coworkers[26] also used aromatic gain to design self-assembling squaramide-based polymers. Although chemists gradually realized that aromatic character in hydrogen bonding systems has significant effects on their properties, however, a well-established relationship between aromaticity and hydrogen bonding interactions is still missing.

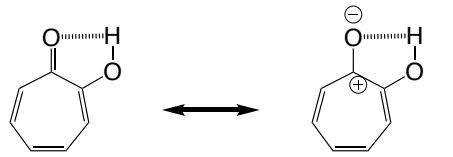


Fig. 5 Tropolone and aromatic sextet structures.

Until recently, Wu and her coworkers[27,28] proposed a clear and comprehensive reciprocal aromaticity-modulated hydrogen bonding (AMHB) relationship. Computed dissected nucleus-independent chemical shifts reveal a uniform pattern and document changes in the magnetic aromatic character of the heterocycles considered, suggesting that the hydrogen bonding interactions that increase (or decrease) cyclic $4n + 2$ π-electron delocalization in heterocycles are strengthened (or weakened) due to enhanced (or reduced) aromatic character in the hydrogen-bonded complex. This relationship can be used to tune the association strengths of hydrogen-bonded arrays through π-electron polarization effects, which is documented by results based on computations[27,28,30] and high-field NMR spectroscopy[31]. The AMHB relationship has valuable insights into modifying hydrogen bond strengths, and also has significant implications for organocatalysts and self-assembling materials design.

In this thesis, I apply computational quantum chemical tools to address two anomalies regarding hydrogen bonding in multipoint arrays. Nature achieves a wide range of weak to strong hydrogen bonds to accomplish complex chemical tasks, however, the aromaticity effects on the inherent association trends of multipoint hydrogen-bonded arrays are still not fully understood. Another question is related to the carrier of genetic information, DNA, which is formed by double helix bound with hydrogen bonds. The fundamental aspects of the recognition properties of nucleobase pairs remain puzzling. This thesis will focus on applying the AMHB relationship to understand: 1) The inherent association strengths of multipoint hydrogen-bonded arrays, and 2) how Hückel aromaticity influences the association strengths of nucleobases. Our computed results suggest that the AMHB relationship can clearly explain the discrepancy raised by the limitations of hydrogen bond numbers and the SEI model, providing a promising conclusion that the association strengths of multipoint hydrogen-bonded arrays correlate well with the increased aromatic character upon hydrogen bonding.

1.4 References

1. Szatyłowicz, H.; Sadlej-Sosnowska, N. *J. Chem. Inf. Model.* **2010**, *50*, 2151-2161.
2. Kyogoku, Y.; Lord, R. C.; Rich, A. *Biochim. Biophys. Acta* **1969**, *179*, 10-17.
3. Kyogoku, Y.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *57*, 250-257.
4. Zhang, Y.; Wu, C.-H.; Wu, J. I. *Org. Biomol. Chem.* DOI: 10.1039/c8ob01669k.
5. Shokri, A.; Schmidt, J.; Wang, X.-B.; Kass, S. R. *J. Am. Chem. Soc.* **2012**, *134*, 2094-2099.
6. Hadži, D.; Kidrič, J.; Koller, J.; Mavri, J. *J. Mol. Struct.* **1990**, *237*, 139-150.
7. Jorgenson, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381-1392.
8. Zielkiewicz, J. *J. Chem. Phys.* **2005**, *123*, 104501.
9. Zimmerman, S. C.; Corbin, P. S. *Struct. Bonding (Berlin, Ger.)* **2000**, *96*, 63.
10. Wilson, A. J. *Soft Matter* **2007**, *3*, 409-425.
11. de Greef, T. F. A.; Meijer, E. W. *Nature* **2008**, *453*, 171-173.
12. de Greef, T. F. A.; Smulders, M. M.; Wolffs, M.; Schenning, A. P.; Sijbesma, R. P.; Meijer, E. W. *Chem. Rev.* **2009**, *109*, 5687-5754.
13. Gooch, A.; Nedolisa, C.; Houton, K. A.; Lindsay, C. I.; Saiani, A.; Wilson, A. J. *Macromolecules* **2012**, *45*, 4723-4729.
14. Schneider, H.-J.; Juneva, R. K.; Simova, S. *Chem. Ber.* **1989**, *122*, 1211-1213.
15. Lüning, U.; Kühl, C. *Tetrahedron* **1998**, *39*, 5735-5738.
16. Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008-2010.
17. Hofmann, A. W. *Proc. Royal Soc. Lond.* **1855**, *8*, 1-3.
18. Boldyrev, A. I.; Wang, L.-S. *Chem. Rev.* **2005**, *105*, 3716-3757.

19. Chen, Z.; Wannere, C. S.; Corminboeuf, C.; Puchta, R.; Schleyer, P. v. R. *Chem. Rev.* **2005**, *105*, 3842-3888.

20. Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530-5538.

21. Mo, Y.; Song, L.; Lin, Y. *J. Phys. Chem. A* **2007**, *111*, 8291-9301.

22. Mo, Y. in *The Chemical Bond: Fundamental Aspects of Chemical Bonding*, ed. Frenking, G.; Shaik, S. Wiley, Weinheim, Germany, **2014**, p,199.

23. Dewar, M. J. S. *Nature* **1945**, *155*, 50-51.

24. Quiñonero, D.; Prohens, R.; Garau, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, R. M. *Chem. Phys. Lett.* **2002**, *351*, 115-120.

25. Storer, R. I.; Aciro, C.; Jones, L. H. *Chem. Soc. Rev.* **2011**, *40*, 2330-2346.

26. Saez Talens, V.; Englebienne, P.; Trinh, T. T.; Noteborn, W. E. M.; Voets, I. K.; Kieltyka, R. E. *Angew. Chem., Int. Ed.* **2015**, *54*, 10502-10506.

27. Wu, J. I.; Jackson, J. E.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **2014**, *136*, 13526-13529.

28. Kakeshpour, T.; Wu, J. I.; Jackson, J. E. *J. Am. Chem. Soc.* **2016**, *138*, 3427-3432.

29. Krygowski, T. M.; Cyrański, M. K. *Chem. Rev.* **2001**, *101*, 1385-1419.

30. Anana, M.; Fernandez, I.; Schaefer, H. F.; Wu, J. I. *J. Comput. Chem.* **2016**, *37*, 59-63.

31. Kakeshpour, T.; Bailey, J. P.; Jenner, M. R.; Howell, D. E.; Staples, R. J.; Holmes, D.; Wu, J. I.; Jackson, J. E. *Angew. Chem., Int. Ed.* **2017**, *56*, 9842-9846.

CHAPTER 2

AROMATICITY GAIN INCREASES THE INHERENT ASSOCIATION STRENGTHS OF MULTIPOINT HYDROGEN-BONDED ARRAYS

Reproduced from Wu, C.-H.; Zhang, Y.; van Rickley, K.; Wu, J. I. *Chem. Commun.* **2018**, *54*, 3512-3515 with permission from the Royal Society of Chemistry.

2.1 Introduction

Multipoint hydrogen-bonded arrays are increasingly featured in the designs of supramolecular polymers, due to their rigid structures and high recognition specificity.[1-4] It is well known that the stability of the hydrogen-bonded array complexes depend on the numbers, types and patterns[5] of the hydrogen bond donor/acceptor pairs present. What is less clear, however, is whether or not other molecular features might significantly influence the hydrogen bonding interactions of arrays so that their association trends might be predicted more reliably *a priori*. In this paper, we report computational results documenting excellent linear correlation between the inherent association strengths of arrays and the amount of "aromaticity gain" in arrays upon complexation.

Although aromaticity and hydrogen bonding have long been considered as separate concepts in organic chemistry, we recently showed that changes in the aromatic character of heterocycles can significantly influence their hydrogen bonding capabilities through a reciprocal aromaticity-modulated hydrogen bonding (AMHB) relationship.[6-9] Results based on computations[6-8] and high-field NMR spectroscopy[9] revealed that hydrogen bonding interactions that increase cyclic $4n + 2$ π-electron delocalizations in heterocycles are strengthened as a result of enhanced aromatic character in the resulting hydrogen-bonded complex. Conversely, hydrogen bonding interactions that decrease cyclic $4n + 2$ π-electron delocalizations in heterocycles are weakened due to reduced aromatic character in the hydrogen-bonded complex. According to the AMHB relationship, we showed that heterocycles with the same numbers, types, and patterns of hydrogen bond

donors/acceptors moieties can exhibit surprisingly different hydrogen bond strengths depending on their π-conjugation patterns.

Here, we report the implication of AMHB as a model to understand and predict the inherent association trends of multipoint hydrogen-bonded arrays. Two examples, the guanine-cytosine (G•C) nucleobase pair **1•2** and the ureidopyrimidone (UPy) dimer **3•3**, are illustrated in **Fig. 1**.

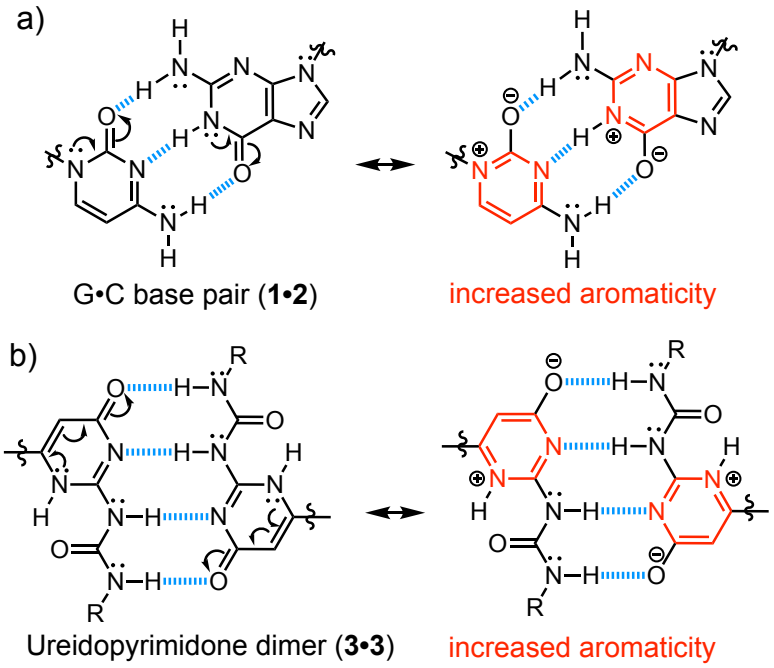


Fig. 1 Schematic illustration of aromaticity-modulated hydrogen bonding (AMHB) in (a) the guanine-cytosine (G•C) base pair, **1•2**, and (b) ureidopyrimidone (UPy) dimer, **3•3**.

Based on the Hückel definition of π-aromaticity for closed-shell planar rings, none of the six membered rings in G, C, and Upy are formally "aromatic" due to lack of a cyclic delocalization of $4n + 2$ π-electrons. However, in their hydrogen-bonded forms, the π-electrons of G, C, and Upy are polarized, resulting in increased cyclic $4n + 2$ π-electron delocalization in the six membered rings (see **Fig. 1a** and **b**, resonance structures in red),

which in return strengthens the corresponding hydrogen bonding interactions. This "extra" aromaticity gain stabilizes the G•C pair and in the Upy dimer, providing a possible explanation for their stronger than expected association strengths compared to analogous arrays with the same numbers, types, and patterns of hydrogen bonding interactions.[10-13]

2.2 Computational methods

Since aromaticity is associated with the cyclic delocalizatioin of π-electrons, aromaticity gain in arrays can be evaluated by the amount of increased π-electron delocalization energy (ΔDE_π) as two array monomers come together to form a hydrogen-bonded complex; $\Delta DE_\pi = DE_{\pi(A\bullet B)} - [DE_{\pi(A)} + DE_{\pi(B)}]$. Here, the block-localized wavefunction (BLW) method,[14-16] an *ab initio* valence bond approach, is applied to measure the π-electron delocalization energies (DE_π) of the monomers and complexes. DE_π is evaluated by the energy difference between that of the fully electron delocalized wavefuntion (Ψ_{deloc}) of the monomer or complex considered and that of the π-electron localized wavefuntion (Ψ_{loc}), in which all π-electron delocalization effects are disabled; $DE_\pi = \Psi_{loc} - \Psi_{deloc}$. Because of its computational efficiency and documented reliability in reproducing experimental trends, the BLW method has been widely applied to quantify and interpret the effects of π-electron delocalization in many chemical systems.[16] All BLW computations were performed at the B3LYP/6-31G(d) level using the GAMESS-2013-R1 program.[17] Geometries for all monomers and complexes were optimized at the ωB97X-D/6-311+G(d,p) level with an ultrafine grid employing the Gaussian09

program.[18]

Following the BLW procedure described above, large positive ΔDE_π values indicate substantial aromaticity gain in arrays upon hydrogen bonding. For example, in the 2-pyridone dimer (See **Fig. 2a**), two hydrogen bonding interactions polarize the N π-lone pairs and C=O π-bonds to increase cyclic six π-electron delocalization (see resonance form on right), giving rise to considerable aromaticity gain in the six membered rings and a large ΔDE_π = 26.1 kcal/mol value. Small positive ΔDE_π values indicate little to no aromaticity gain (or a decreased aromatic character) in arrays upon hydrogen bonding. For example, in the 2-hydroxypyridine dimer (see **Fig. 2b**), two hydrogen bonding interactions polarize the N π-lone pairs and C=N π-bonds to decrease cyclic six π-electron delocalization (see resonance form on right), resulting in reduced aromatic character in the six membered rings and a small ΔDE_π = 5.7 kcal/mol value. The effects of aromaticity gain (or loss) upon array complexation also may be considered as a manifestation of non-additivity in resonance-assisted hydrogen bonding.[19]

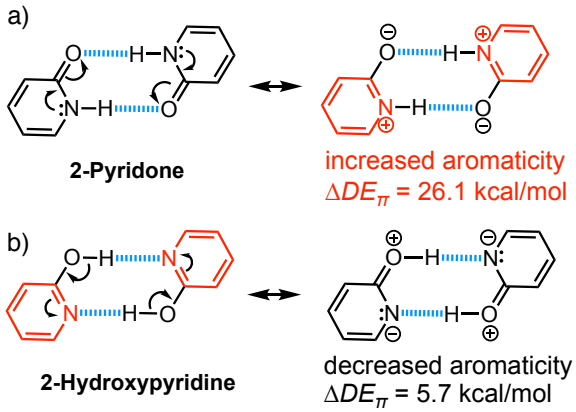


Fig. 2 AMHB in (a) the 2-pyridone dimer (note large ΔDE_π value due to aromaticity gain in the six-membered rings) and (b) the 2-hydroxypyridine dimer (note small ΔDE_π value, due to reduced aromatic character in the six-membered rings).

2.3 Results and discussion

Based on a survey of 46 hydrogen-bonded arrays, an excellent linear relationship was found between the computed gas-phase association free energies ($-\Delta G_{assoc}$, at 298 K) and ΔDE_{π} values of 26 triply (r = 0.940) and 20 quadruply (r = 0.959) hydrogen-bonded arrays (see **Fig. 3**), suggesting that the inherent association strengths of multipoint hydrogen-bonded arrays correlate well with the amount of aromaticity gain in arrays upon complexation. Depending on the π-conjugation pattern of the array monomers considered, hydrogen bonding interactions that increase cyclic $4n + 2$ π-electron delocalizations in arrays (as indicated by a large ΔDE_{π} value) are strengthened, while hydrogen bonding interactions that decrease cyclic $4n + 2$ π-electron delocalizations (as indicated by a small ΔDE_{π} value) are weakened. Computations in implicit chloroform solvation and analyses based on the natural bond orbital (NBO) deletion method[20] showing the same excellent correlation are presented in the 2.5 Computational data.

This finding points to important limitations of the secondary electrostatic interaction (SEI) model of Jorgensen and Pranata,[5] which has long guided the understanding of multipoint hydrogen-bonded arrays and their associations in supramolecular chemistry. According to the SEI model, it was suggested that for a given number of hydrogen bonds in an array, those with all hydrogen bond donors (D) on one fragment and all acceptors (A) on the other are the most robust, since this arrangement maximizes attractive electrostatic interactions. Thus, the association strengths of triply hydrogen-bonded array are expected to follow the order: AAA-DDD > AAD-DDA > ADA-DAD (**Fig. 3a**), while

those of quadruply hydrogen-bonded arrays are expected to follow the order: AADD-DDAA > ADDA-DAAD ≈ ADAA-DADD > ADAD-DADA (**Fig. 3b**).

Past studies both supporting and refuting the SEI model have appeared in the literature. Schneider *et al.*,[21] and later Zimmerman and coworkers,[22] have shown that empirical increments taking into account primary and secondary electrostatic interaction (as well as secondary CH...O interactions)[13] can be used to predict the experimental associations of hydrogen-bonded arrays satisfactorily. Based on a survey of more than 60 arrays, Vanka *et al.*[23] found excellent correlation between the computed array association energies and calculated electrostatic forces between the arrays. Popelier and Joubert showed, based on a study of 28 base pairs, that electrostatic interactions between many remote atom pairs also contribute importantly to array binding.[24] However, Lukin and Leszynski argued that the incremental approaches of Scheider and Zimmerman can be deceptive;[25] based on extensive quantum chemical calculations, these authors demonstrated that some ADD-DAA arrays appear to have weaker experimentally observed associations than their analogous AAA-DDD arrays only because of a more solvated ADD and DAA monomer in wet polar solvent.

Guerra *et al.* noted that effects other than electrostatic interactions play important roles in the hydrogen bonds of DNA base pairs.[26] Mo commented that changes in the electrostatic components of computed array association energies could arise from changes in the π-electron delocalization energies of monomers upon hydrogen bonding.[27,28] Although the SEI model has been criticized on the basis of both quantum

chemical calculations and experimental evidence, it remains the most widely applied concept for the design and synthesis of hydrogen-bonded molecular recognition units.

In sharp contrast to the SEI model, our computations show that arrays with the "best" electrostatic interaction patterns do not necessarily exhibit the strongest inherent association strengths. Surprisingly, the AAA-DDD complexes (in blue), despite having all hydrogen bond donors (D) on one fragment and all acceptors (A) on the other, exhibit lower $-\Delta G_{assoc}$ values compared to those of the AAD-DDA (in black) and ADA-DAD (in red) complexes (**Fig. 3a**). Even arrays with the same SEI patterns can exhibit a wide range of $-\Delta G_{assoc}$ values. Notably, the computed $-\Delta G_{assoc}$ values for the AAA-DDD, AAD-DDA, AADD-DDAA, and ADDA-DAAD sets vary over a range of *ca.* 10 kcal/mol, corresponding to a $K_{assoc} \approx 10^7$ difference! These trends violate the SEI model and illustrate the importance of considering aromaticity gain in arrays as a relevant factor for determining the stability of multipoint hydrogen-bonded complexes.

Clear exceptions to the SEI model may be explained when the effects of aromaticity gain in arrays are considered. For example, the quadruply hydrogen-bonded modules of Corbin-Zimmerman[22] ($K_{assoc} \geq 3 \times 10^7$ M^{-1} in chloroform, **Fig. 4a**) and Lüning[29] ($K_{assoc} \approx 2000$ M^{-1} in chloroform, **Fig. 4b**), exhibit the same ADDA-DAAD pattern, but display drastically different experimental K_{assoc} values. This disparity (a near 10^4 times difference) has been attributed to variances in the preorganization energies of the monomers,[1] but can arise in part due to the different π-conjugation patterns of the monomers (note orange highlight in **Fig. 4**).

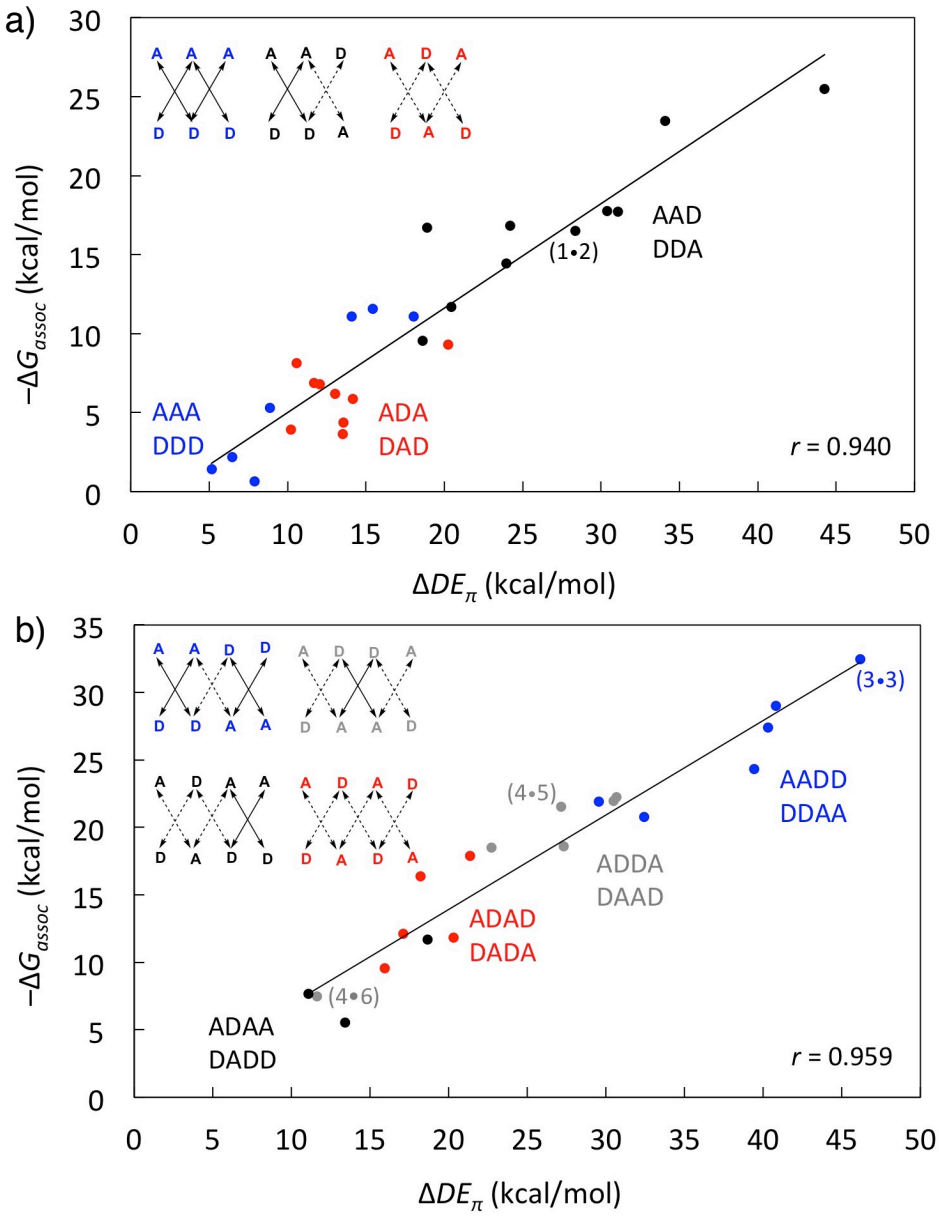


Fig. 3 Plot of $-\Delta G_{\text{assoc}}$ vs. ΔDE_{π} for (a) triply and (b) quadruply hydrogen-bonded arrays. The secondary electrostatic interaction (SEI) patterns for each array are color coded; see top left corner of each plot (– lines indicate attractive interactions, --- lines indicate repulsive interactions).

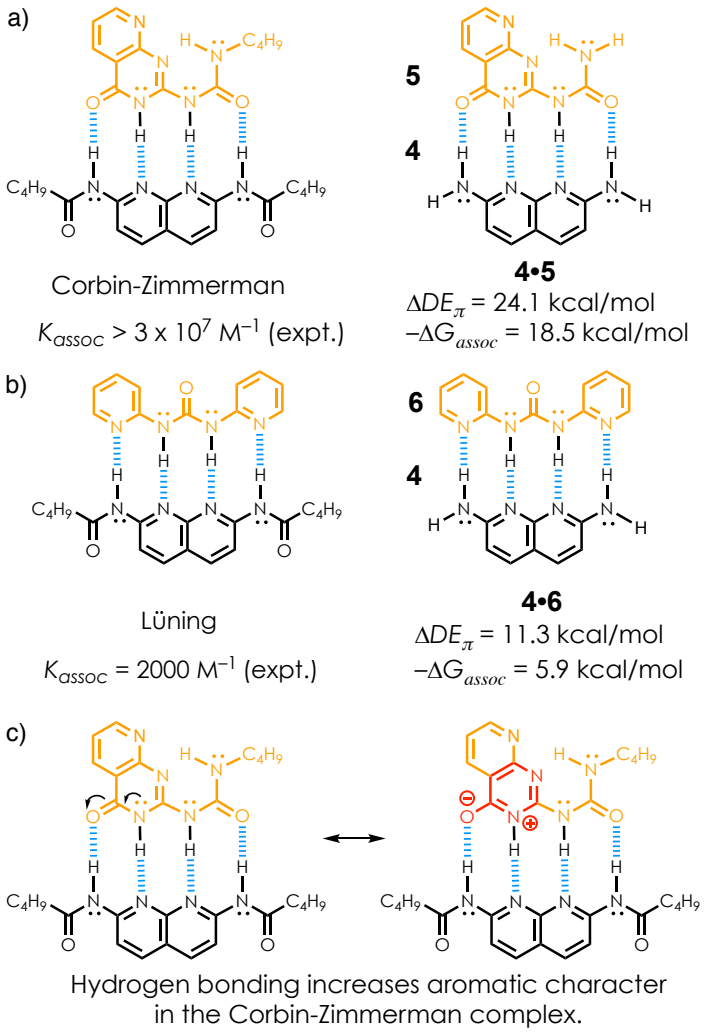


Fig. 4 Experimental Kassoc values (in chloroform) for the ADDA-DAAD modules of (a) Corbin-Zimmerman and (b) Lüning; see also model arrays, **4•5** and **4•6**, on right. Note π-conjugation pattern difference highlighted in orange. (c) Resonance form showing increased aromatic character in the Corbin-Zimmerman module upon hydrogen bonding.

In the Corbin-Zimmerman module, hydrogen bonding interactions can polarize the π-electrons to increase cyclic six π-aromatic character in the 4-pyridone moiety (see **Fig. 4c**, note resonance form in red), but such aromatization effects are absent in the Lüning complex. Indeed, BLW computations for models of the two ADDA-DAAD arrays, **4•5** and **4•6** (–COC_4H_9 groups replaced by H atom), show much greater π-conjugation gain

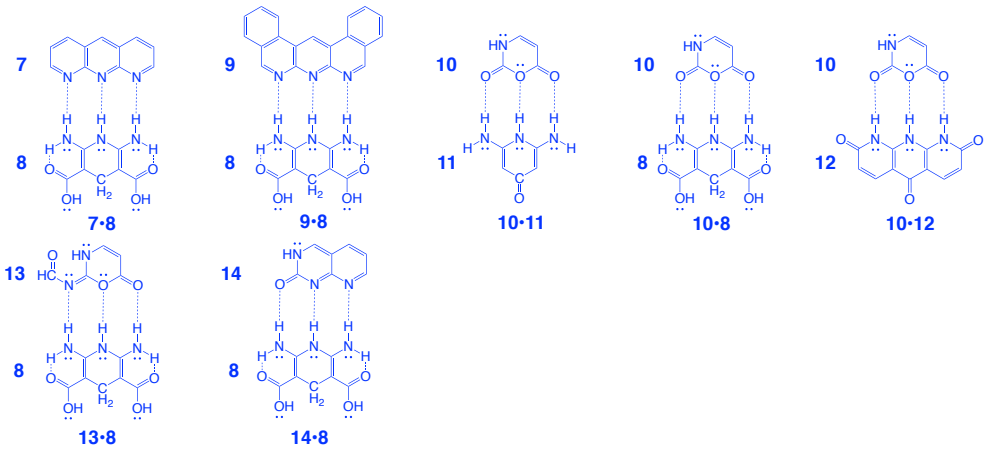
for **4•5** (ΔDE_π = 24.1 kcal/mol, **Fig. 4a**) than for **4•6** (ΔDE_π = 11.3 kcal/mol, **Fig. 4b**) (*cf.* **Fig. 2**, BLW analysis for 2-pyridone *vs.* 2-hydroxypyridine; fully aromatic rings exhibit less aromaticity gain upon hydrogen bond complexation).

2.4 Conclusions

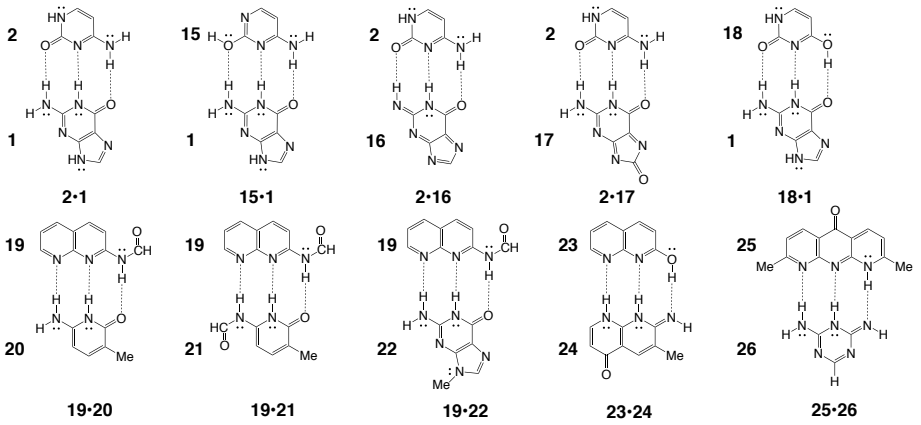
With its near 150 year old history, the term "aromatic rings" has evolved to adopt various shades of meanings in the chemical literature. Very often, rigid unsaturated rings are generally called aromatic rings, even if they do not follow the more stringent Hückel definition – a closed-shell π-conjugated ring having a cyclic delocalization of $4n + 2$ π-electrons. We show here that the traditional Hückel definition of aromaticity has chemical value for interpreting the inherent association trends of triply and quadruply hydrogen-bonded arrays. Of course blends of factors (*e.g.*, entropy, solvation, conformational and protomeric equilibria of the array monomers) can all influence the experimental associations of arrays. Nevertheless, our findings highlight the surprising impact of aromaticity gain on the association strengths of multipoint hydrogen-bonded arrays, suggesting that the potential for aromaticity gain in arrays should be considered in addition to the often used check-list (*i.e.*, numbers, types, and SEI patterns) for designing hydrogen-bonded molecular-recognition units.

2.5 Computational data

AAA–DDD



ADD–DAA



ADA–DAD

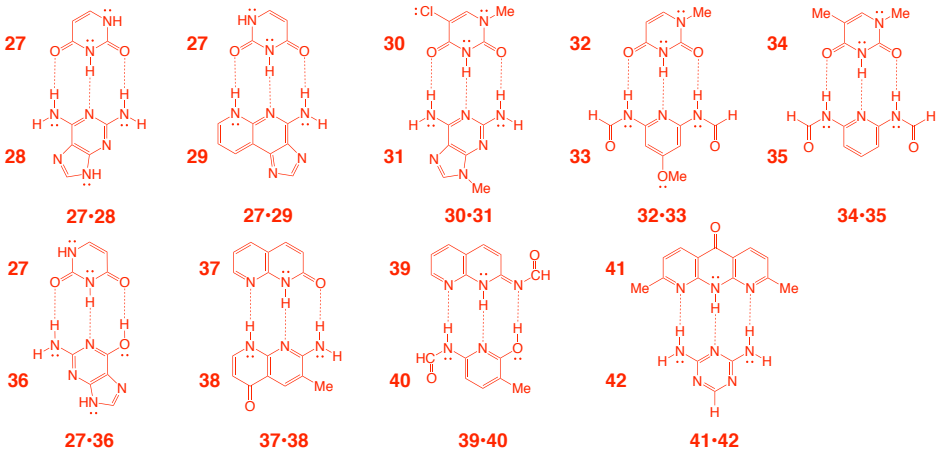
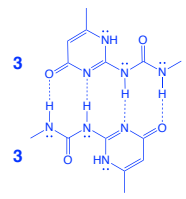


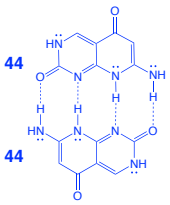
Figure S1. Structures for all triply hydrogen-bonded arrays considered.

AADD–DDAA

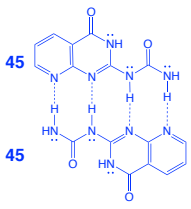


3·3 (UPy)

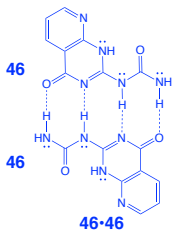
43·43



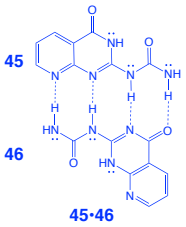
44·44



45·45

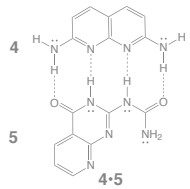


46·46

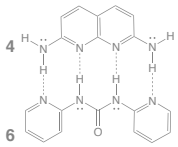


45·46

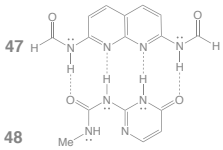
ADDA–DAAD



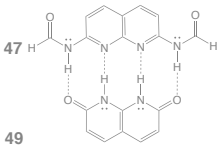
4·5



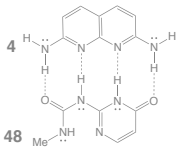
4·6



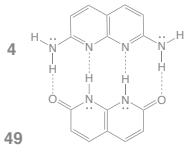
47·48



47·49

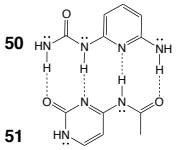


4·48

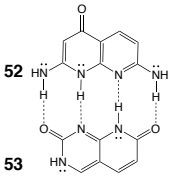


4·49

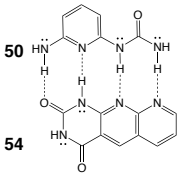
ADAA–DADD



50·51

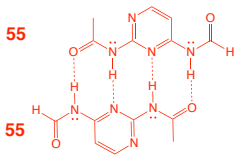


52·53

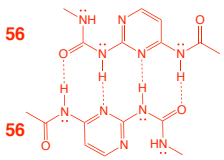


50·54

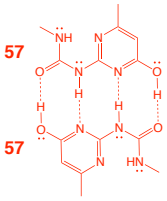
ADAD–DADA



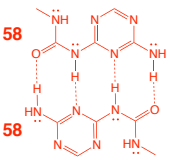
55·55



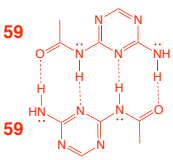
56·56



57·57



58·58



59·59

Figure S2. Structures for all quadruply hydrogen-bonded arrays considered.

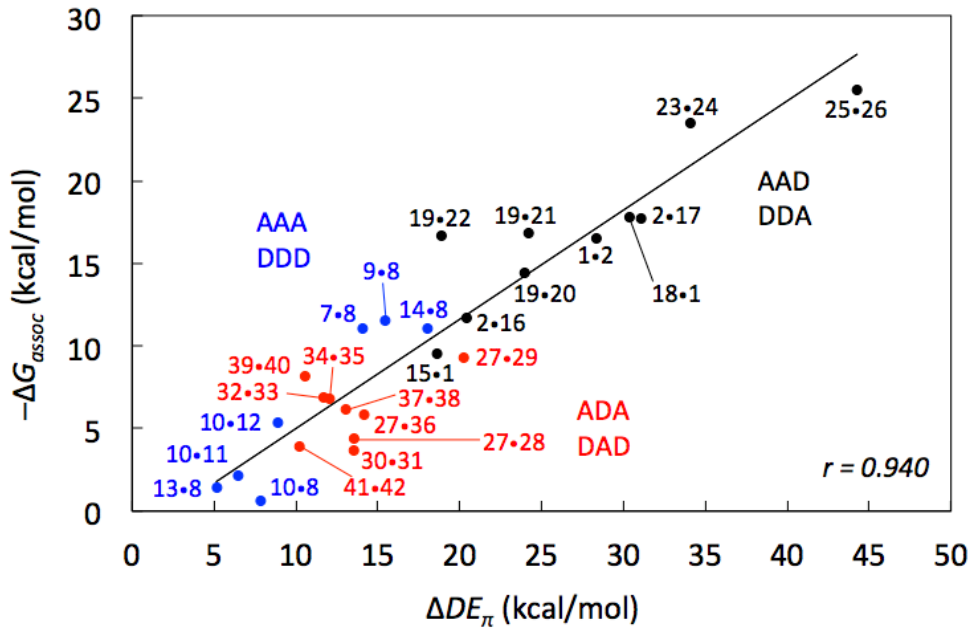


Figure S3. Plot of $-\Delta G_{assoc}$ *vs.* ΔDE_{π} for triply hydrogen-bonded arrays in the gas-phase. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level. BLW computations were performed at B3LYP/6-31G(d).

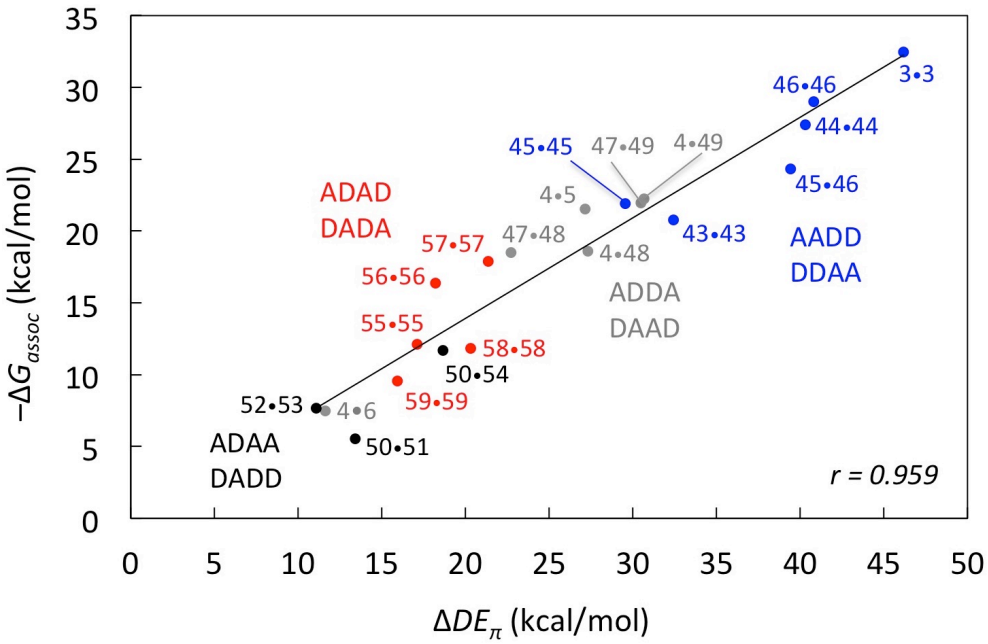


Figure S4. Plot of $-\Delta G_{assoc}$ *vs.* ΔDE_{π} for quadruply hydrogen-bonded arrays in the gas-phase. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level. BLW computations were performed at B3LYP/6-31G(d).

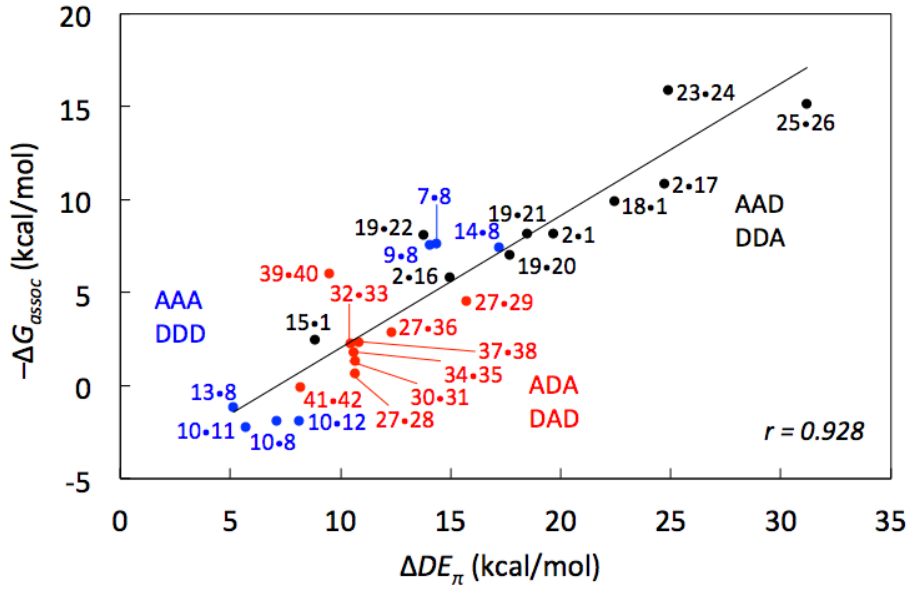


Figure S5. Plot of $-\Delta G_{\text{assoc}}$ *vs.* ΔDE_{π} for triply hydrogen-bonded arrays in implicit chloroform solvation. All geometries were optimized with C_s symmetry at the IEF-PCM-ωB97X-D/6-311+G(d,p) level. BLW computations were performed in implicit chloroform solvation at PCM-B3LYP/6-31G(d).

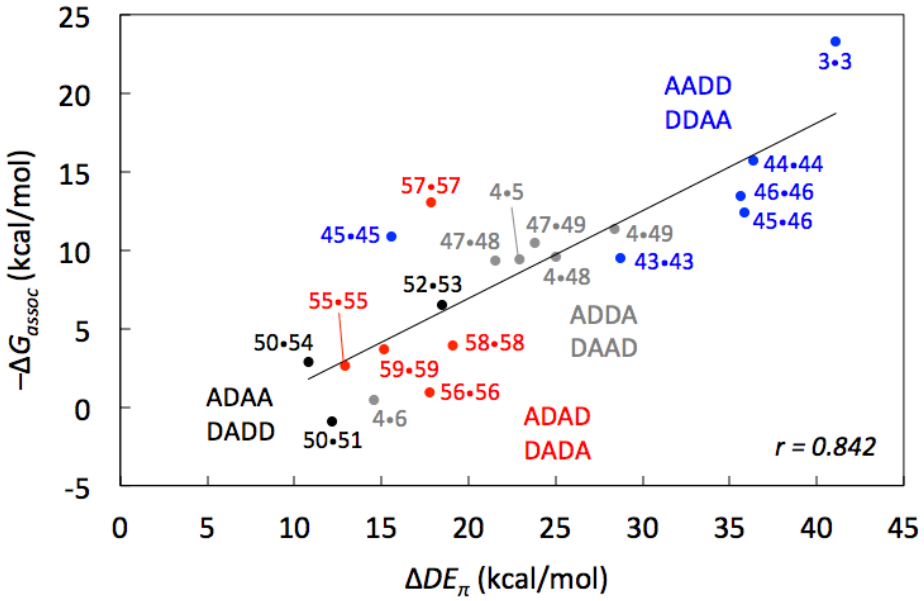


Figure S6. Plot of $-\Delta G_{\text{assoc}}$ *vs.* ΔDE_{π} for quadruply hydrogen-bonded arrays in implicit chloroform solvation. All geometries were optimized with C_s symmetry at the IEF-PCM-ωB97X-D/6-311+G(d,p) level. BLW computations were performed in implicit chloroform solvation at PCM-B3LYP/6-31G(d).

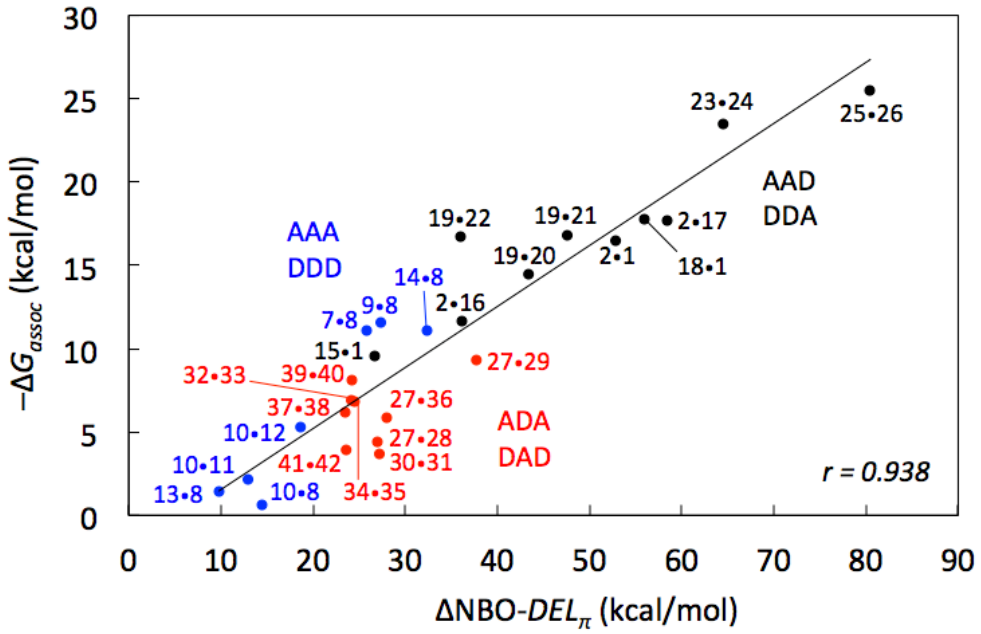


Figure S7. Plot of $-\Delta G_{assoc}$ *vs.* ΔNBO-DEL_π for triply hydrogen-bonded arrays in the gas-phase. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level. NBO deletion computations were performed at ωB97X-D/def2-TZVPP.

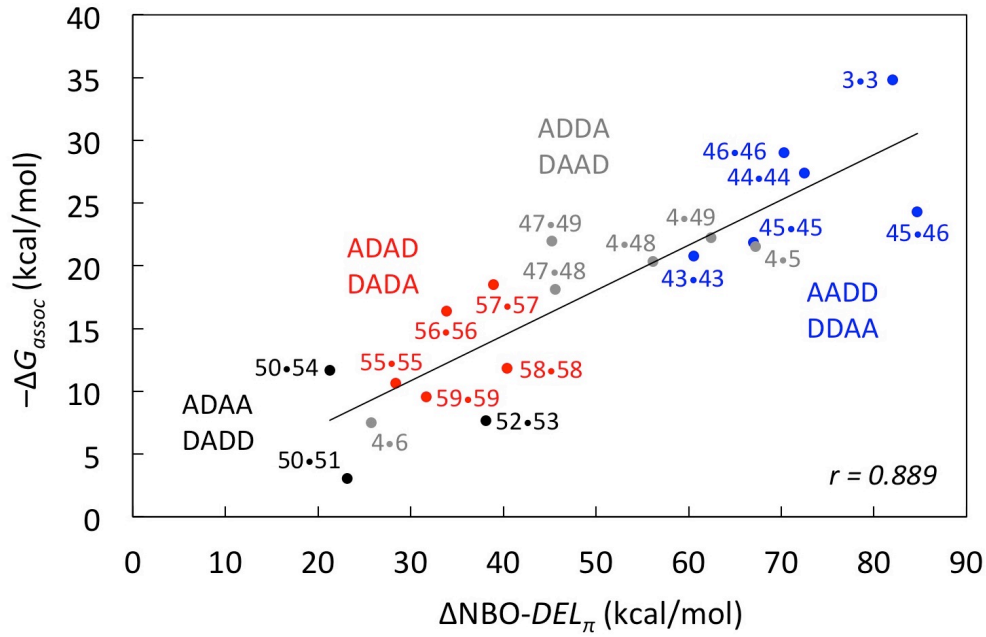


Figure S8. Plot of $-\Delta G_{assoc}$ *vs.* ΔNBO-DEL_π for quadruply hydrogen-bonded arrays in the gas-phase. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level. NBO deletion computations were performed at ωB97X-D/def2-TZVPP.

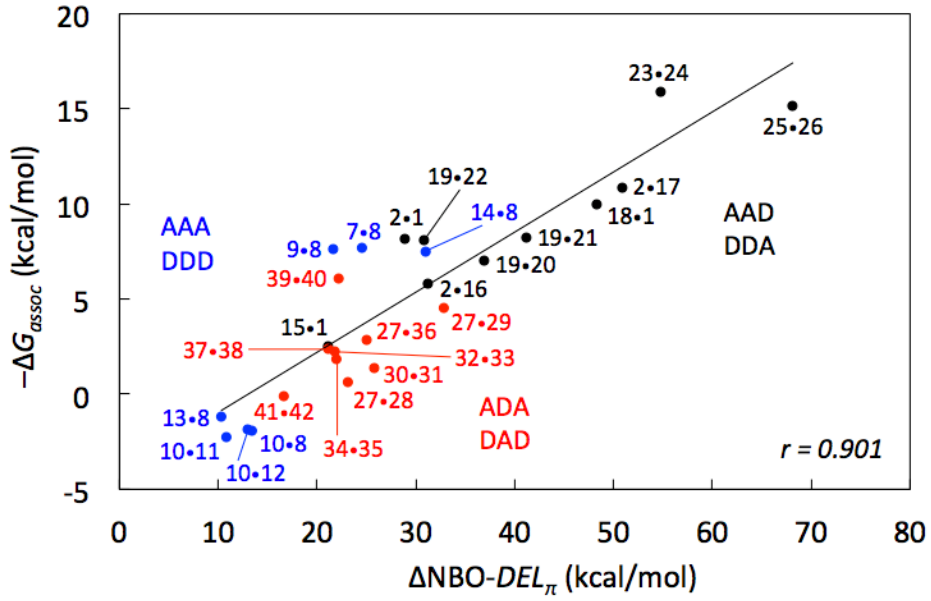


Figure S9. Plot of $-\Delta G_{assoc}$ *vs.* ΔNBO-DEL_π for triply hydrogen-bonded arrays in implicit chloroform solvation. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level. NBO deletion computations were performed at ωB97X-D/def2-TZVPP.

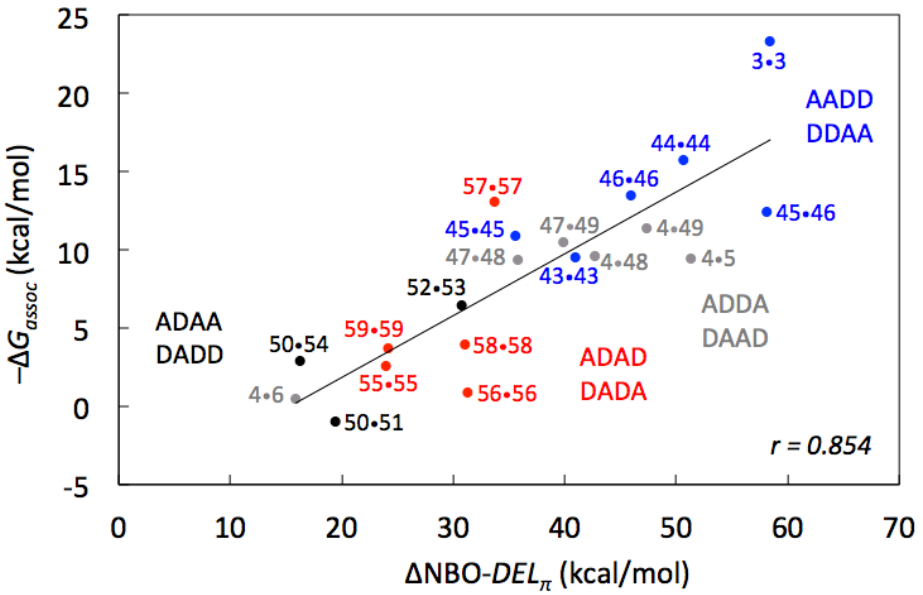


Figure S10. Plot of $-\Delta G_{assoc}$ *vs.* ΔNBO-DEL_π for triply hydrogen-bonded arrays in implicit chloroform solvation. All geometries were optimized with C_s symmetry at the ωB97X-D/6-311+G(d,p) level. NBO deletion computations were performed at ωB97X-D/def2-TZVPP.

2.6 References

1. Zimmerman, S. C.; Corbin, P. S. *Struct. Bonding (Berlin, Ger.)* **2000**, *96*, 63.
2. Wilson, A. J. *Soft Matter* **2007**, *3*, 409-425.
3. de Greef, T. F. A.; Meijer, E. W. *Nature* **2008**, *453*, 171-173.
4. de Greef, T. F. A.; Smulders, M. M.; Wolffs, M.; Schenning, A. P.; Sijbesma, R. P.; Meijer, E. W. *Chem. Rev.* **2009**, *109*, 5687-5754.
5. Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008-2010.
6. Wu, J. I.; Jackson, J. E.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **2014**, *136*, 13526-13529.
7. Kakeshpour, T.; Wu, J. I.; Jackson, J. E. *J. Am. Chem. Soc.* **2016**, *138*, 3427-3432.
8. Anand, M.; Fernandez, I.; Schaefer, H. F.; Wu, J. I. *J. Comput. Chem.* **2016**, *37*, 59-63.
9. Kakeshpour, T.; Bailey, J. P.; Jenner, M. R.; Howell, D. E.; Staples, R. J.; Holmes, D.; Wu, J. I.; Jackson, J. E. *Angew. Chem. Int. Ed.* **2017**, *56*, 9842-9846.
10. Beijer, F. H.; Kooijman, H.; Spek, A. L.; Sijbesma, R. P.; Meijer, E. W. *Angew. Chem. Int. Ed.* **1998**, *37*, 75-78.
11. Beijer, F. H.; Sijbesma, R. P.; Kooijman, H.; Spek, A. L.; Meijer, E. W. *J. Am. Chem. Soc.* **1998**, *120*, 6761-6769.
12. Ligthart, G. B. W. L.; Ohkawa, H.; Sijbesma, R. P.; Meijer, E. W. *J. Am. Chem. Soc.* **2005**, *127*, 810-811.
13. Quinn, J. R.; Zimmerman, S. C.; Del Bene, J. E.; Shavitt, I. *J. Am. Chem. Soc.* **2007**, *129*, 934-941.
14. Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530-5538.

15. Mo, Y.; Song, L.; Lin, Y. *J. Phys. Chem. A* **2007**, *111*, 8291-9301.

16. Mo, Y. In *The Chemical Bond: Fundamental Aspects of Chemical Bonding*; Frenking, G., and Shaik, S., Eds.; Wiley: Weinheim, Germany, 2014; pp 199.

17. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.;Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347-1363.

18. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2013.

19. Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. *J. Am. Chem. Soc.* **1989**, *111*, 1023-1028.

20. Glendening, E. D.; Landis, C. R.; Weinhold, F. *J. Comput. Chem.* **2013**, *34*, 1429-1437.

21. Sartorius, J.; Schneider, H.-J. *Chem. Eur. J.* **1996**, *2*, 1446-1452.

22. Corbin, P. S.; Zimmerman, S. C. *J. Am. Chem. Soc.* **1998**, *120*, 9710-9711.

23. Tiwari, M. K.; Vanka, K. *Chem. Sci.* **2017**, *8*, 1378-1390.

24. Popelier, P. L. A.; Joubert, L. *J. Am. Chem. Soc.* **2002**, *124*, 8725-8729.

25. Lukin, O.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 6775-6782.

26. Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J.* **1999**, *5*, 3581-3594.

27. Mo, Y. *J. Mol. Model.* **2006**, *12*, 665-672.

28. Beck, J. F.; Mo, Y. *J. Comput. Chem.* **2007**, *28*, 455-466.

29. Lüning, U.; Kühl, C. *Tetrahedron* **1998**, *39*, 5735-5738.

CHAPTER 3

WHY DO A•T AND G•C SELF-SORT? HÜCKEL AROMATICITY AS A DRIVING FORCE FOR ELECTRONIC COMPLEMENTARITY IN BASE PAIRING

Reproduced from Zhang, Y.; Wu, C.-H.; Wu, J. I. *Org. Biomol. Chem.* DOI: 10.1039/c8ob01669k with permission from the Royal Society of Chemistry.

3.1 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 nuclobase 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 aasociation 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 (~3 M^{-1}) and U•U (~6 M^{-1}), and the G•C (10^4–10^5 M^{-1}) pair, compared to G•G (10^3–10^4 M^{-1}) and C•C (~28 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 $4n + 2$ π-electron delocalization) of the interacting bases exhibit stronger than expected hydrogen bonds, while those that decrease aromaticity (*i.e.*, disrupt cyclic $4n + 2$ π-electron delocalization) of the interacting bases display weaker associations. In a related work, Cyrański *et al*. showed indeed that hydrogen bonding at the C=O position 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 arrays[18] 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 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.

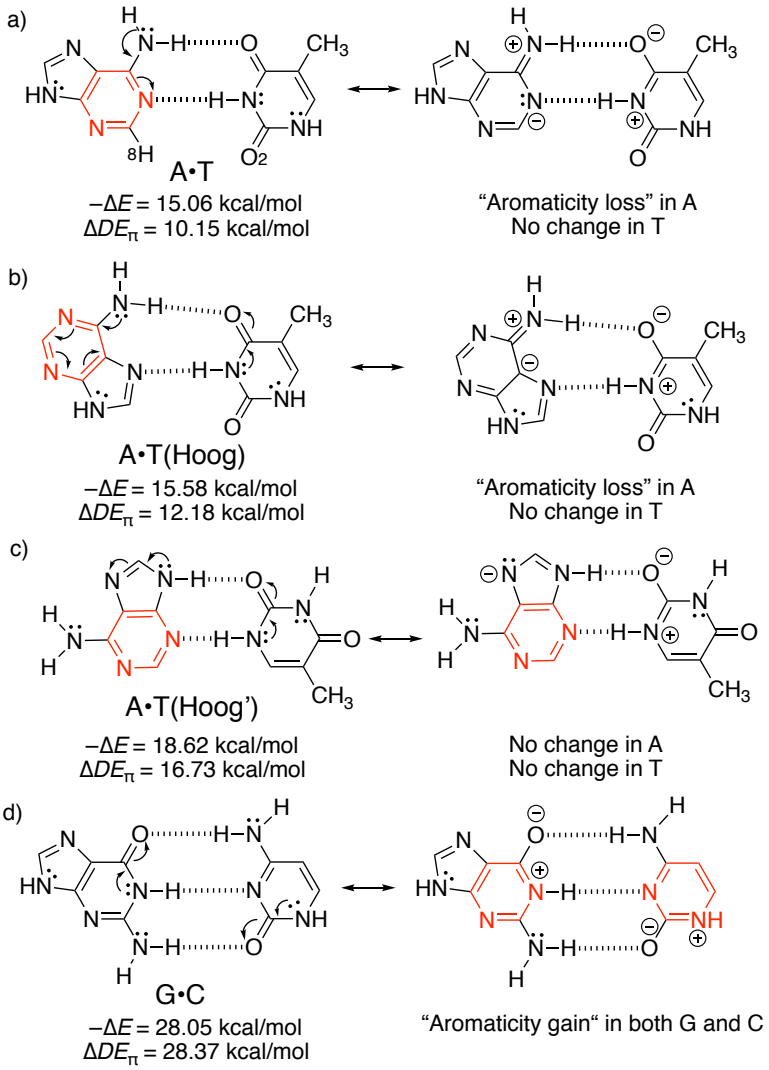


Fig. 1 Aromaticity-modulated hydrogen bonding (AMHB) in the (a) Watson–Crick A•T, (b) natural Hoogsteen A•T, (c) most stable Hoogsteen A•T, and (d) Watson–Crick G•C pairs. Resonance structures with formal cyclic $4n + 2\pi$ electron delocalizations are in red. Computed interaction energies ($-\Delta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

3.2 Computational methods

All geometries were optimized with a constrained C_s symmetry at ωB97X-D/6-311+G(d,p) level using ultrafine grid employing the Gaussian09 program[19]. Vibrational frequency calculations verified the nature of the stationary points, and the reported gas-

phase interaction energies for all 57 base pairs (see **Figure S1**) and the reference dimers (see **Figures 3** and **4** in the main text), 1•1, 2•2, 3•3, 4•4, 5•5, 3•1, 6•4, include zero-point vibrational energy (ZPE) corrections. The computed planarization energies for all nonplanar minima were less than 3 kcal/mol, except for the following: 1 (4.67 kcal/mol), 5 (3.22 kcal/mol), 6 (3.47 kcal/mol), reversed G•C (4.70 kcal/mol), 1•1 (3.47 kcal/mol), 5•5 (4.89 kcal/mol), and 6•4 (3.41 kcal/mol). Electrostatic potential (ESP) derived charges were computed for the planar geometries of A•T, G•C, A•A, I•I, and their isolated bases at ωB97X-D/6-311+G(d,p), using the pop=chelpg keyword in Gaussian09. Block-localized wavefunction (BLW) computations[20-22] quantified the π-electron delocalization energies (DE_π) of all isolated bases and hydrogen-bonded bases and were performed at the B3LYP/6-31G(d) level using the GAMESS-2013-R1 program[23].

The DE_π values were computed by the energy difference between the fully delocalized wavefunction (ψ_{deloc}) and the energy of a π-electron-localized wavefunction (ψ_{loc}) of the system considered; $DE_\pi = E(\psi_{loc}) - E(\psi_{deloc})$. Ψ_{loc} was computed by assigning all of the π-bonds and π-type lone-pairs to separate subspaces ("blocks") to disable π-electron delocalization. Based on this localization scheme, ψ_{loc} was computed by restricting the expansion of molecular orbitals over basis functions within a selected molecular subspace. Each block included two π-electrons and the p_z, d_{xz}, d_{yz} basis functions belonging to the heavy atoms assigned to the specific subspace. Direct comparisons of the computed DE_π values for the base pairs to the interacting bases provided a measure of the π-conjugation gain in the paired bases; $\Delta DE_\pi = DE_{a•b} - (DE_a + DE_b)$. The computed ΔDE_π values for all 57 base pairs were positive, indicating increased

π-conjugation for all paired bases upon hydrogen bonding. However, the degree of π-conjugation gain differs depending on the degree of gain or loss 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 change in aromatic character.

Natural Bond Orbital (NBO) computations[35] were performed at ωB97X-D/Def2-TZVPP//ωB97X-D/6-311+G(d,p) in the gas-phase using the planar geometries of selected isolated bases and hydrogen-bonded bases, to provide complementary insight. According to the NBO deletion procedure, all π* orbitals were deleted to quantify the effects of π-electron delocalization (DEL_π) for each of the selected hydrogen-bonded base (A•B) and the isolated bases (A and B). The difference, ΔNBO-DEL_π, calculated by the computed DEL_π of the complex minus that of its isolated nucleobases, provided a measure of the degree of increased π-electron delocalization in the monomers upon hydrogen bonding; $\Delta\text{NBO-}DEL_\pi = DEL_{\pi(\text{A•B})} - [DEL_{\pi(\text{A})} + DEL_{\pi(\text{B})}]$.

Electrostatic potential (ESP) values for the isolated and the paired A, T, G, C bases were computed using their optimized geometries at C_s, and differences of the computed ESP values were taken to generate the ΔESP plots. The computed ΔESP plots provide a measure of the change in electrostatic potential of A, T, G, C upon hydrogen bonding to form A•T and G•C (see **Fig. 3** in the main text and **Figure S4**). Positive ΔESP values (blue color) indicate a more repulsive surface, and negative ΔESP values (red color) indicate a more attractive surface upon base pairing.

For a given base, ESP values for the isolated and paired states were computed using a common isosurface, constructed based on the geometry of the isolated base, using an

electron density isosurface of $\rho = 0.001$ e au^{-3} and a 0.05 Bohr grid. For example, the change in electrostatic potential of guanine upon base pairing to form G•C was evaluated by: $\Delta ESP_G[G] = ESP_G[G•C\ pair] - ESP_G[G]$. Subscript "G" indicates that the isosurface used for computing the ESP values of guanine, in both the isolated and hydrogen-bonded states, were constructed based on the optimized geometry of isolated guanine. Isosurface coordinates were generated using Multiwfn program, followed by input to the Gaussian 09 program to compute the ESP values. All computations were performed at the ωB97X-D/6-311+G(d,p) level.

3.3 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 (ΔDE_π, see below). Geometries for all structures were optimized with a constrained C_s symmetry at ωB97X-D/6-311+G(d,p) employing Gaussian09[19]. 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 mathematically constrained to resemble a strict π-electron-localized Lewis structure; $DE_\pi = \psi_{loc} - \psi_{deloc}$. 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); $\Delta DE_\pi = DE_{a•b} - (DE_a + DE_b)$. 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, aromaticity loss in A, no change in T, **Figures 1a** and **b**) are lower compared to that of the most stable A•T configuration, A•T(Hoog'), (16.7 kcal/mol, no change in aromaticity for A or T, **Fig. 1c**). The computed ΔDE_π for G•C (28.4 kcal/mol) is even higher since base pairing increases aromaticity in both G and C (**Fig. 1a**).

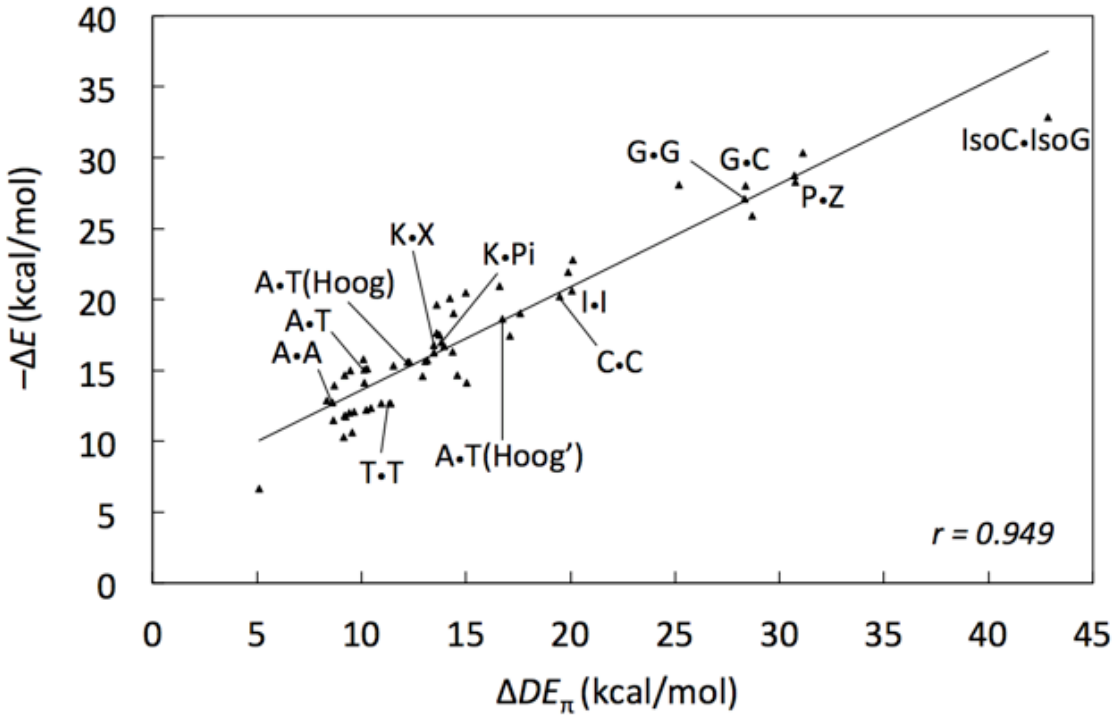


Fig. 2 Plot of base pairing interaction energy (–Δ*E*, in kcal/mol) *vs.* π-conjugation gain (ΔDE_π) in the gas-phase for all 57 base pairs. Plot of –Δ*E* vs. ΔDE_π for selected base pairs in chloroform is provided in **Fig. S8** of 3.5 Computational data.

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]. We note that previous benchmarking studies of the performance of various force-fields[26] 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 interaction (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 3.5 Computational data.

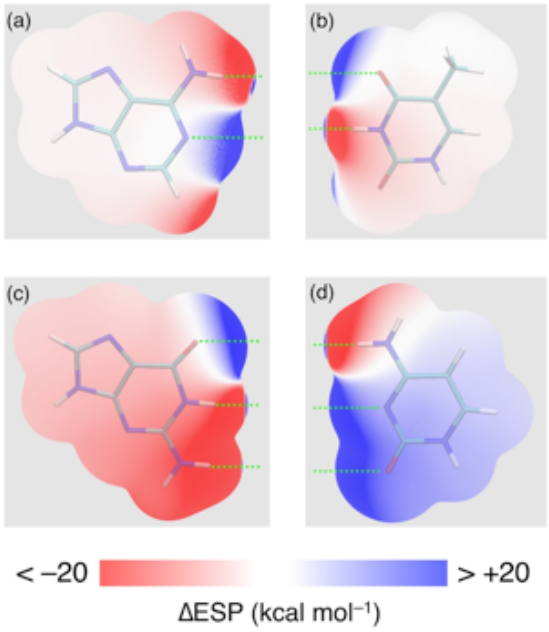
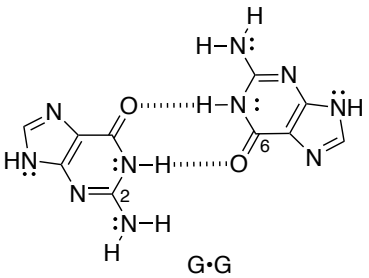


Fig. 3 Computed electrostatic potential difference maps, ΔESP, for (a) adenine, (b) thymine (c) guanine, and (d) cytosine, upon base pairing to A•T and G•C.

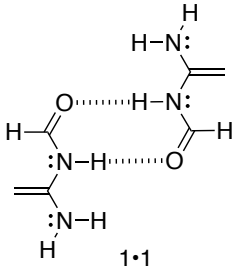
Direct comparisons of the computed $-\Delta E$ values for G•G, I•I, C•C, T•T, A•A, to those of their hydrogen-bonded acyclic dimer references (1•1, 2•2, 3•3, 4•4, 5•5) document the energetic effects of AMHB (**Fig. 4**). Notably, the computed $-\Delta E$ values for G•G (27.1 kcal/mol) and I•I (20.6 kcal/mol) are 6 to 8 kcal mol^{-1} higher compared to those of their acyclic references, 1•1 (19.6 kcal/mol) and 2•2 (14.3 kcal/mol), which display the same primary and secondary electrostatic interaction but are preclude of aromaticity gain. In contrast, the computed −ΔE values for C•C (20.2 kcal/mol) and T•T (12.7 kcal/mol) closely follow those of their acyclic references, 3•3 (21.1 kcal/mol) and 4•4 (10.8 kcal/mol), 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) is modestly lower than 5•5 (14.4 kcal/mol), 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 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.

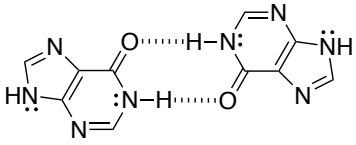
As shown in **Fig. 5**, the computed $-\Delta E$ values for both isoC•isoG (32.9 kcal/mol) and P•Z (28.3 kcal/mol) are 5 to 10 kcal/mol higher than their acyclic reference 3•1 (22.9 kcal/mol), due to increased aromaticity in the isoC, isoG, P, Z moieties upon base pairing. In sharp contrast, the computed $-\Delta E$ values for both K•Pi (17.0 kcal/mol) and K•X (16.8 kcal/mol) are close to that of their acyclic reference 6•4 (15.8 kcal/mol), 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 3.5 Computational data. A plot 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 3.5 Computational data.



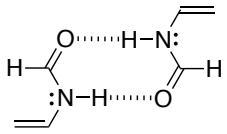
G•G
$-\Delta E$ = 27.10 kcal/mol
ΔDE_π = 28.33 kcal/mol



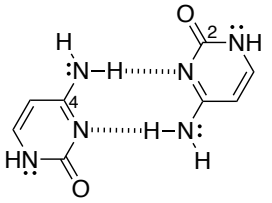
1•1
$-\Delta E$ = 19.57 kcal/mol



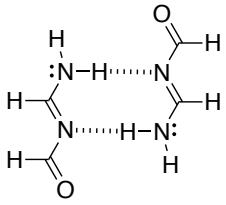
I•I
$-\Delta E$ = 20.63 kcal/mol
ΔDE_π = 20.06 kcal/mol



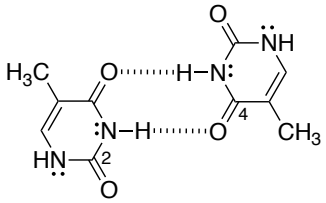
2•2
$-\Delta E$ = 14.32 kcal/mol



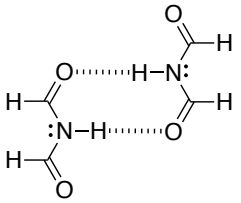
C•C
$-\Delta E$ = 20.18 kcal/mol
ΔDE_π = 19.48 kcal/mol



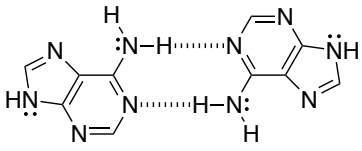
3•3
$-\Delta E$ = 21.07 kcal/mol



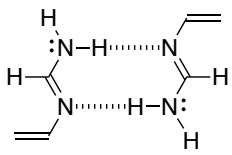
T•T
$-\Delta E$ = 12.69 kcal/mol
ΔDE_π = 11.36 kcal/mol



4•4
$-\Delta E$ = 10.80 kcal/mol



A•A
$-\Delta E$ = 12.76 kcal/mol
ΔDE_π = 8.60 kcal/mol



5•5
$-\Delta E$ = 14.39 kcal/mol

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 3.5 Computational data.

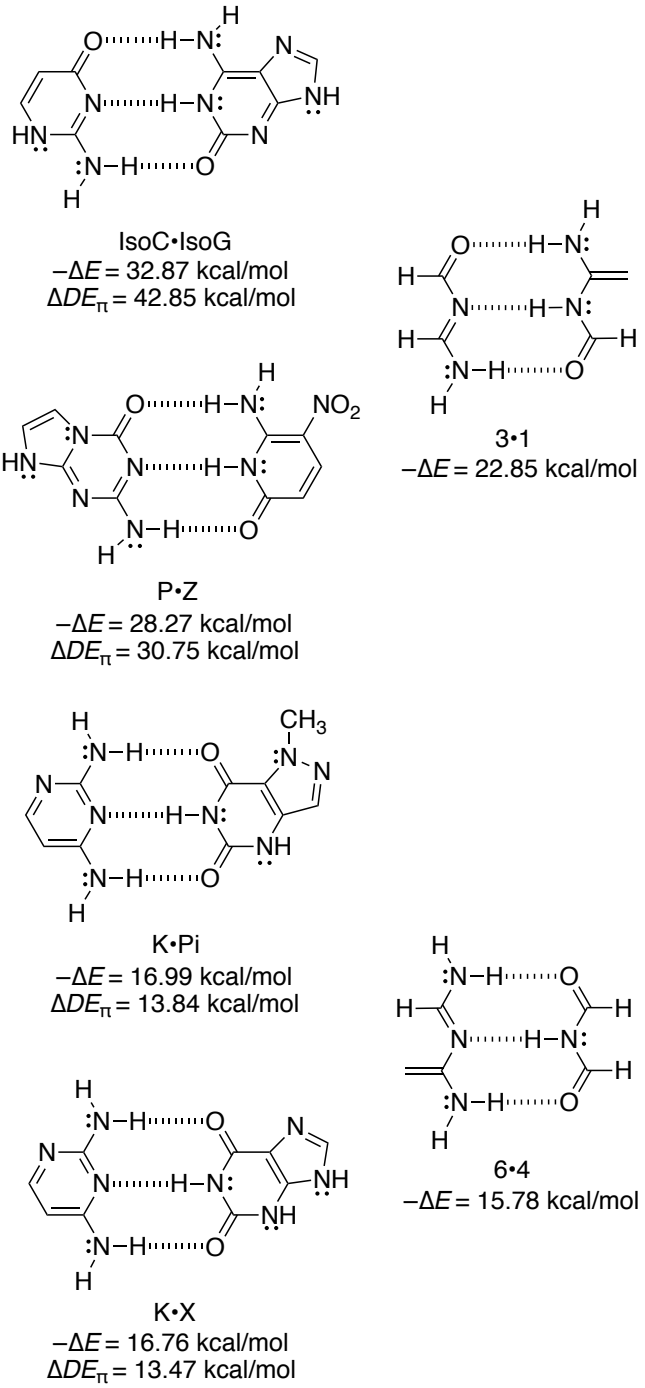


Fig. 5 Computed $-\Delta E$ and ΔDE_{π} values for isoC•isoG, P•Z, K•Pi, K•X, and $-\Delta E$ values of their acyclic references. See also **Fig. S3** in 3.5 Computational data.

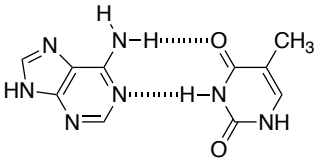
Overall, our findings suggest that while primary and secondary electrostatic interaction[29] have clear energetic consequences for base pairing (e.g., $-\Delta\Delta E$ = 8.8 kcal/mol for 1•1 *vs.* 4•4, and 7.1 kcal/mol for 3•1 *vs.* 6•4), the effects of AMHB are comparable in magnitude (e.g., $-\Delta\Delta E$ = 7.5 kcal/mol for 1•1 *vs*. G•G, and 10.0 kcal/mol for 3•1 *vs.* isoC•isoG), and therefore should be considered when evaluating base pairing strengths.

3.4 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.

3.5 Computational data

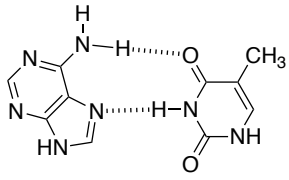
Doubly hydrogen-bonded pairs:



A•T

$-\Delta E$ = 15.06 kcal/mol

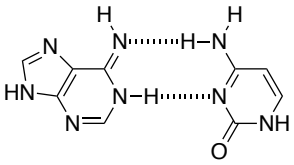
ΔDE_{π} = 10.15 kcal/mol



A•T(Hoog)

$-\Delta E$ = 15.58 kcal/mol

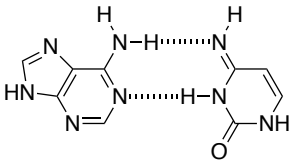
ΔDE_{π} = 12.18 kcal/mol



A_{imino}•C

$-\Delta E$ = 22.79 kcal/mol

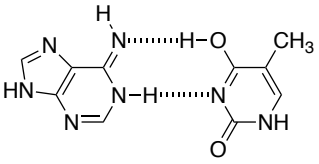
ΔDE_{π} = 20.12 kcal/mol



A•C_{imino}

$-\Delta E$ = 15.79 kcal/mol

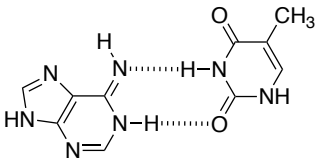
ΔDE_{π} = 10.09 kcal/mol



A_{imino}•T_{enol}

$-\Delta E$ = 28.05 kcal/mol

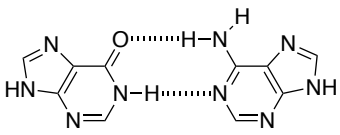
ΔDE_{π} = 25.19 kcal/mol



A_{imino}•T

$-\Delta E$ = 17.66 kcal/mol

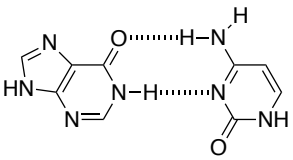
ΔDE_{π} = 13.61 kcal/mol



I•A

$-\Delta E$ = 16.25 kcal/mol

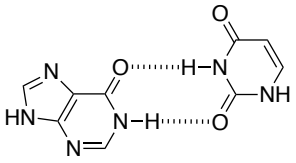
ΔDE_{π} = 13.47 kcal/mol



I•C

$-\Delta E$ = 21.96 kcal/mol

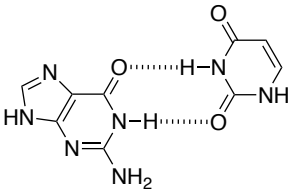
ΔDE_{π} = 19.89 kcal/mol



I•U

$-\Delta E$ = 14.60 kcal/mol

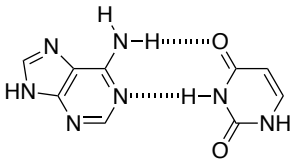
ΔDE_{π} = 12.92 kcal/mol



G•U

$-\Delta E$ = 16.32 kcal/mol

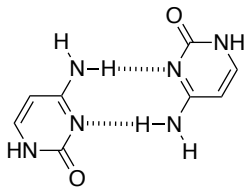
ΔDE_{π} = 14.36 kcal/mol



A•U

$-\Delta E$ = 15.14 kcal/mol

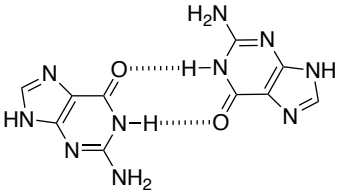
ΔDE_{π} = 10.26 kcal/mol



C•C

$-\Delta E$ = 20.18 kcal/mol

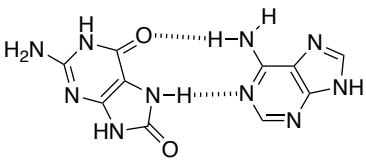
ΔDE_{π} = 19.48 kcal/mol



G•G

$-\Delta E$ = 27.10 kcal/mol

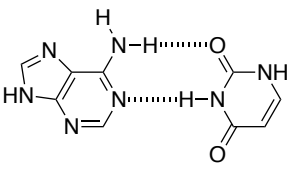
ΔDE_{π} = 28.33 kcal/mol



8-oxoG•A

$-\Delta E$ = 15.58 kcal/mol

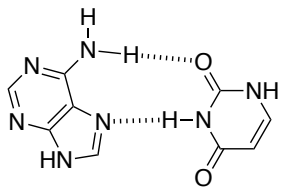
ΔDE_{π} = 12.28 kcal/mol



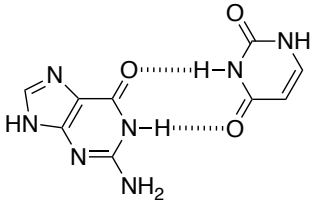
Reverse A•U

$-\Delta E$ = 14.67 kcal/mol

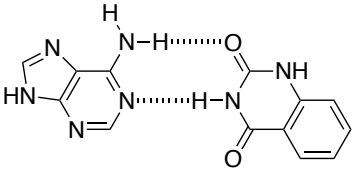
ΔDE_{π} = 9.19 kcal/mol



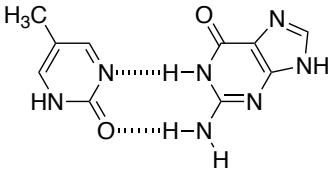
Reverse A•U(Hoog)
$-\Delta E$ = 15.31 kcal/mol
ΔDE_{π} = 11.52 kcal/mol



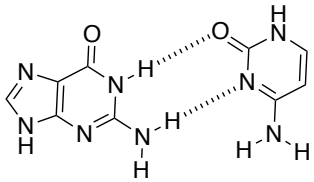
Reverse G•U
$-\Delta E$ = 17.43 kcal/mol
ΔDE_{π} = 17.12 kcal/mol



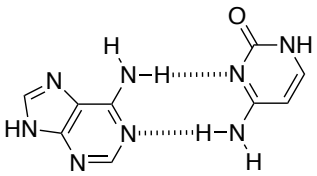
A•yT
$-\Delta E$ = 14.96 kcal/mol
ΔDE_{π} = 9.44 kcal/mol



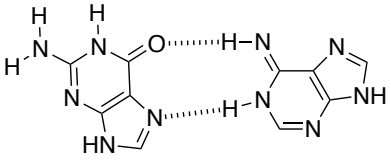
T(h)•G
$-\Delta E$ = 19.64 kcal/mol
ΔDE_{π} = 13.58 kcal/mol



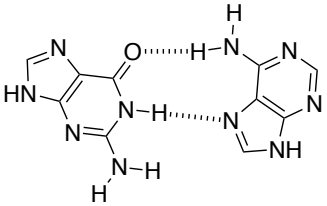
Reverse G•C
$-\Delta E$ = 14.11 kcal/mol
ΔDE_{π} = 10.13 kcal/mol



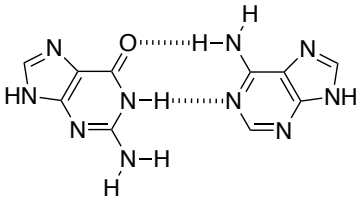
Reverse A•C
$-\Delta E$ = 15.62 kcal/mol
ΔDE_{π} = 13.06 kcal/mol



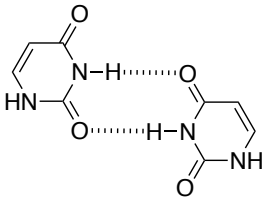
G•A_{imino}
$-\Delta E$ = 6.62 kcal/mol
ΔDE_{π} = 5.11 kcal/mol



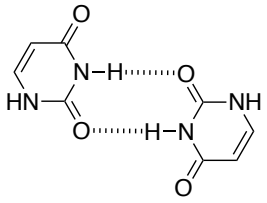
Syn-G•A
$-\Delta E$ = 14.65 kcal/mol
ΔDE_{π} = 14.59 kcal/mol



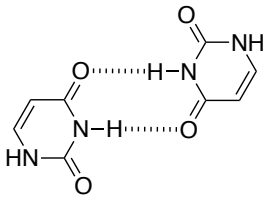
G•A
$-\Delta E$ = 16.73 kcal/mol
ΔDE_{π} = 13.95 kcal/mol



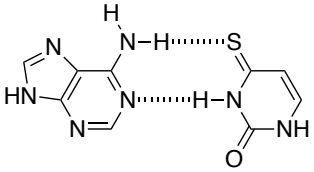
U•U 1
$-\Delta E$ = 12.18 kcal/mol
ΔDE_{π} = 10.22 kcal/mol



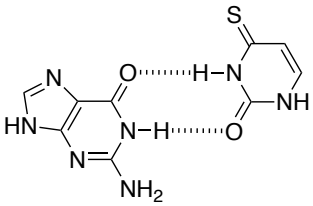
U•U 2
$-\Delta E$ = 11.73 kcal/mol
ΔDE_{π} = 9.25 kcal/mol



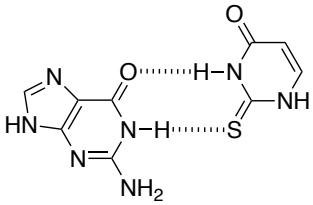
U•U 3
$-\Delta E$ = 12.68 kcal/mol
ΔDE_{π} = 11.41 kcal/mol



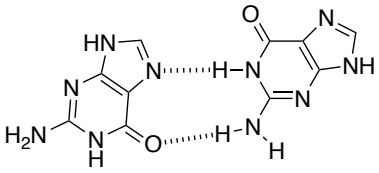
A•4-thioU
$-\Delta E$ = 13.92 kcal/mol
ΔDE_{π} = 8.69 kcal/mol



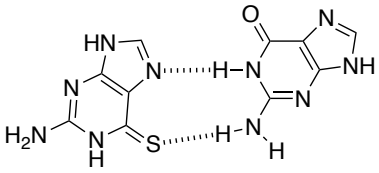
G•4-thioU
$-\Delta E$ = 15.69 kcal/mol
ΔDE_{π} = 13.14 kcal/mol



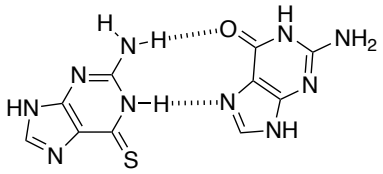
G•2-thioU
$-\Delta E$ = 14.14 kcal/mol
ΔDE_{π} = 15.04 kcal/mol



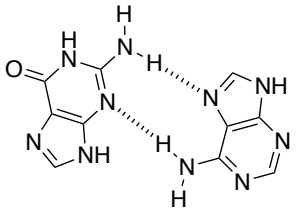
G•G 3
$-\Delta E$ = 20.08 kcal/mol
ΔDE_{π} = 14.20 kcal/mol



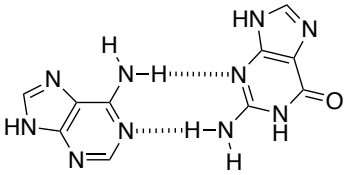
G•6-thioG 3
$-\Delta E$ = 20.49 kcal/mol
ΔDE_{π} = 15.01 kcal/mol



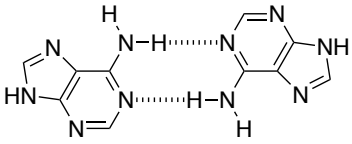
6-thioG•G 3
$-\Delta E$ = 20.97 kcal/mol
ΔDE_{π} = 16.63 kcal/mol



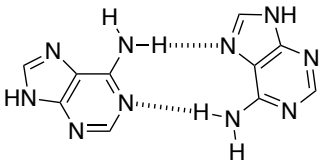
G•A 2
$-\Delta E$ = 11.46 kcal/mol
ΔDE_{π} = 8.63 kcal/mol



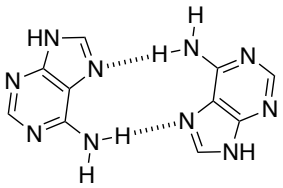
G•A 4
$-\Delta E$ = 12.84 kcal/mol
ΔDE_{π} = 8.34 kcal/mol



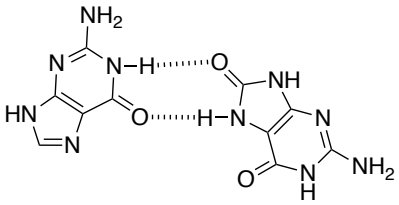
A•A 1
$-\Delta E$ = 12.76 kcal/mol
ΔDE_{π} = 8.60 kcal/mol



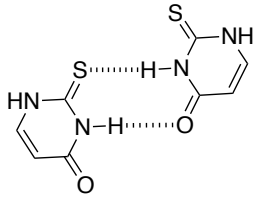
A•A 2
$-\Delta E$ = 11.82 kcal/mol
ΔDE_{π} = 9.20 kcal/mol



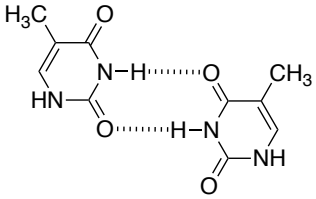
A•A 3
$-\Delta E$ = 10.26 kcal/mol
ΔDE_{π} = 9.16 kcal/mol



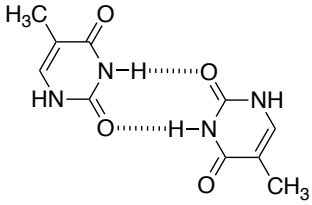
8-oxoG•G
$-\Delta E$ = 19.01 kcal/mol
ΔDE_{π} = 17.59 kcal/mol



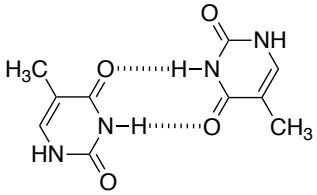
2-thioU•2-thioU
$-\Delta E$ = 10.61 kcal/mol
ΔDE_{π} = 9.53 kcal/mol



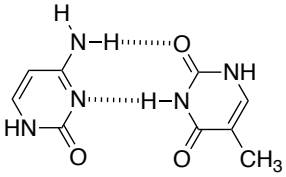
T•T 1
$-\Delta E$ = 12.37 kcal/mol
ΔDE_{π} = 10.47 kcal/mol



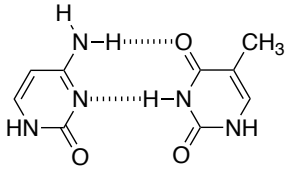
T•T 2
$-\Delta E$ = 12.07 kcal/mol
ΔDE_{π} = 9.65 kcal/mol



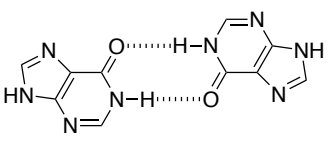
T•T 3
$-\Delta E$ = 12.69 kcal/mol
ΔDE_{π} = 11.36 kcal/mol



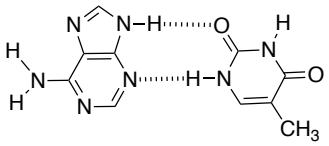
T•C 1
$-\Delta E$ = 12.03 kcal/mol
ΔDE_{π} = 9.41 kcal/mol



T•C 2
$-\Delta E$ = 12.64 kcal/mol
ΔDE_{π} = 10.97 kcal/mol

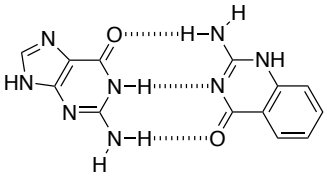


I•I
$-\Delta E$ = 20.63 kcal/mol
ΔDE_{π} = 20.06 kcal/mol

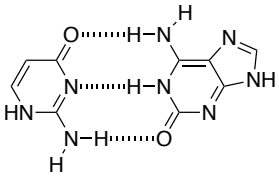


A•T(Hoog')
$-\Delta E$ = 18.62 kcal/mol
ΔDE_{π} = 16.73 kcal/mol

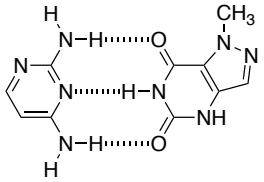
Triply hydrogen-bonded pairs:



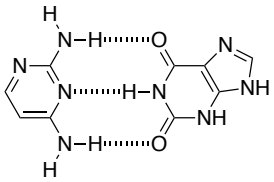
G•yC
$-\Delta E$ = 30.36 kcal/mol
ΔDE_{π} = 31.12 kcal/mol



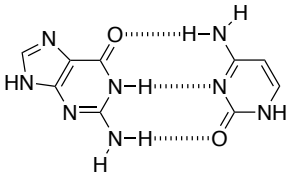
IsoC•IsoG
$-\Delta E$ = 32.87 kcal/mol
ΔDE_{π} = 42.85 kcal/mol



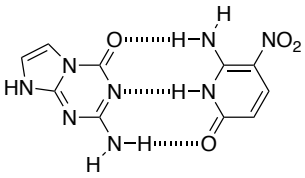
K•Pi
$-\Delta E$ = 16.99 kcal/mol
ΔDE_{π} = 13.84 kcal/mol



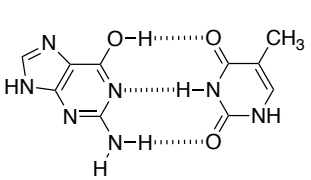
K•X
$-\Delta E$ = 16.76 kcal/mol
ΔDE_{π} = 13.47 kcal/mol



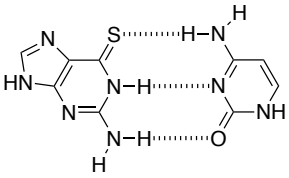
G•C
$-\Delta E$ = 28.05 kcal/mol
ΔDE_{π} = 28.37 kcal/mol



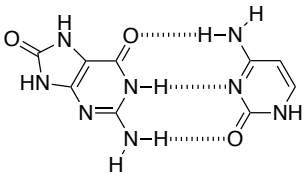
P•Z
$-\Delta E$ = 28.27 kcal/mol
ΔDE_{π} = 30.75 kcal/mol



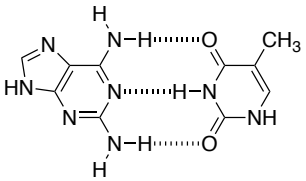
G*•T
$-\Delta E$ = 18.99 kcal/mol
ΔDE_{π} = 14.42 kcal/mol



6-thioG•C
$-\Delta E$ = 25.88 kcal/mol
ΔDE_{π} = 28.68 kcal/mol



8-oxoG•C
$-\Delta E$ = 28.75 kcal/mol
ΔDE_{π} = 30.70 kcal/mol



2-aminoA•T
$-\Delta E$ = 17.51 kcal/mol
ΔDE_{π} = 13.75 kcal/mol

Figure S1. Structures for all nucleobase pairs considered.

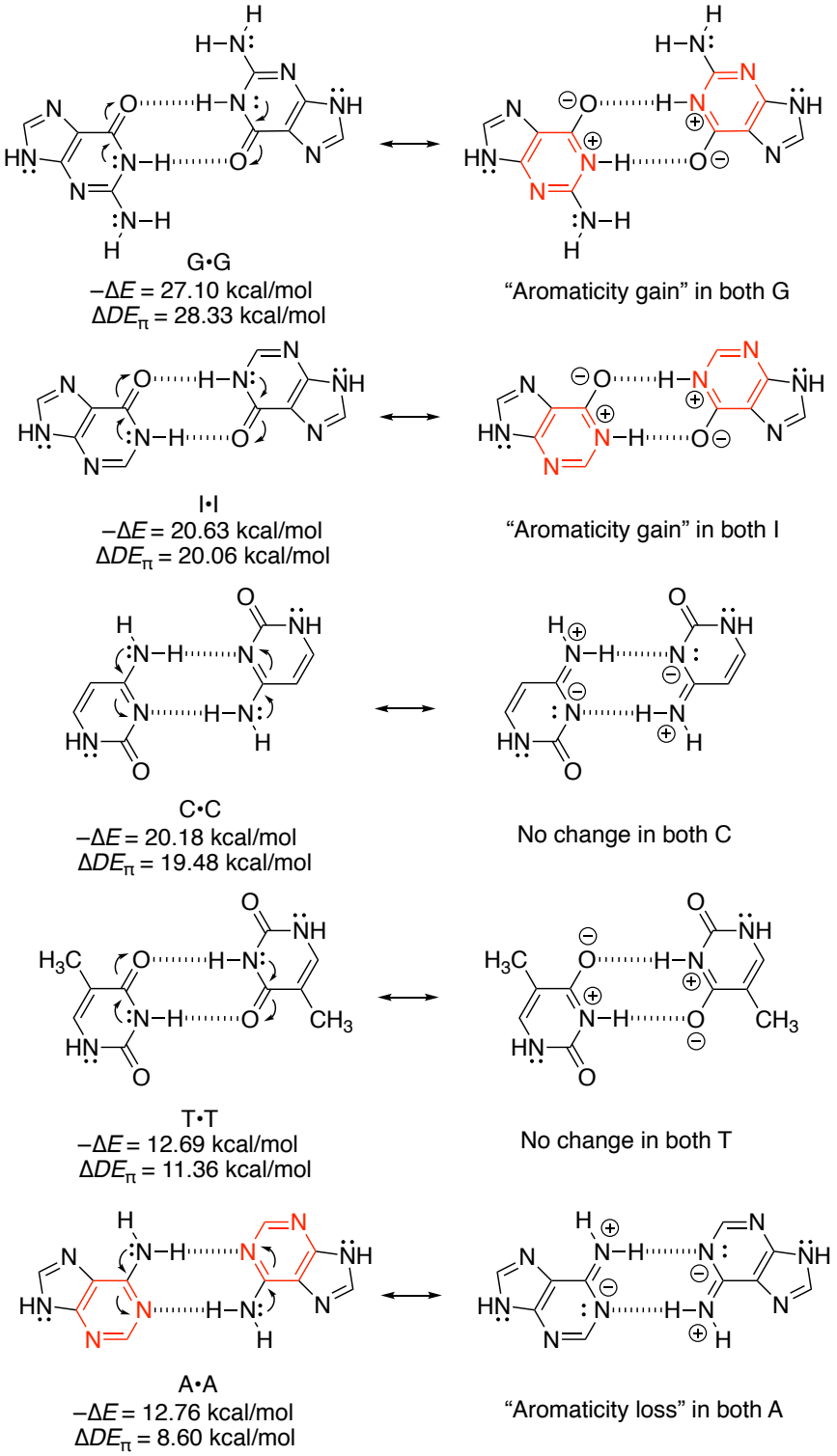


Figure S2. Resonance forms for base pairs shown in Figure 4 of the main text.

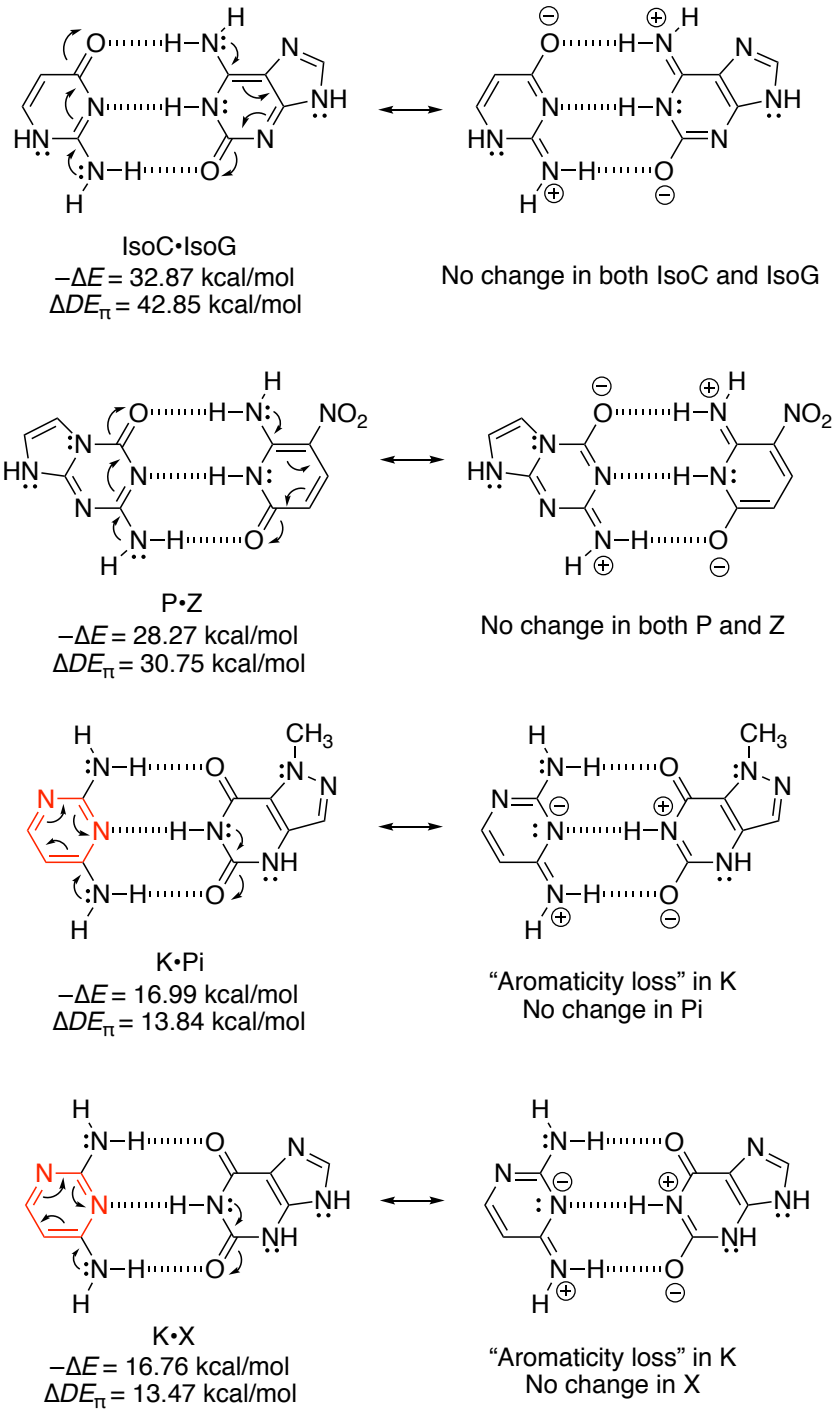
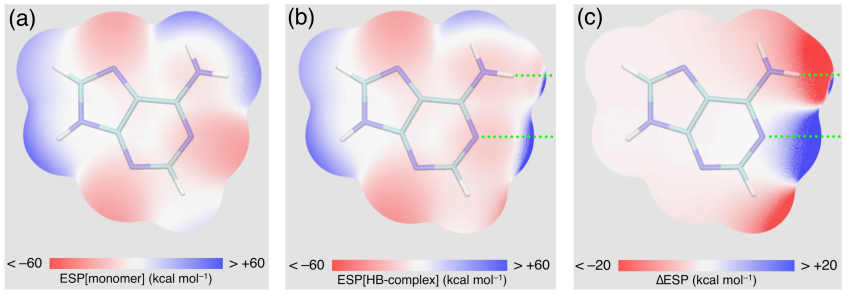
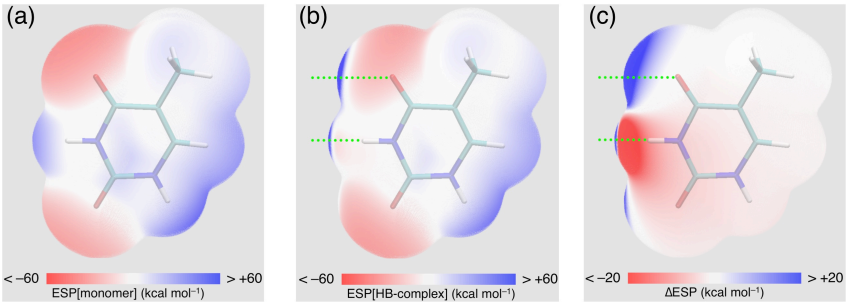


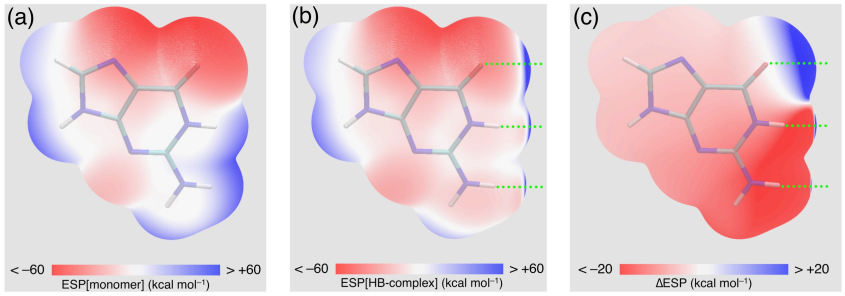
Figure S3. Resonance forms for base pairs shown in Figure 5 of the main text.



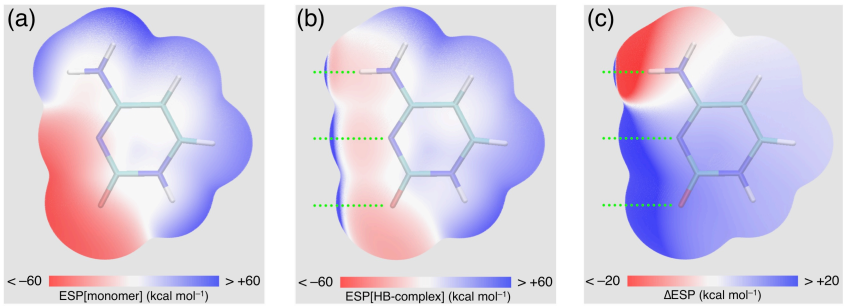
ESP plot of a) the adenine (A) monomer, b) adenine paired in A•T, and the c) ΔESP plot.



ESP plot of a) the thymine (T) monomer, b) thymine paired in A•T, and the c) ΔESP plot.



ESP plot of a) the guanine (G) monomer, b) guanine paired in G•C, and the c) ΔESP plot.



ESP plot of a) the cytosine (C) monomer, b) cytosine paired in G•C, and the c) ΔESP plot.

Figure S4. ESP and ΔESP plots for A, T, G, C upon base pairing to form A•T and G•C.

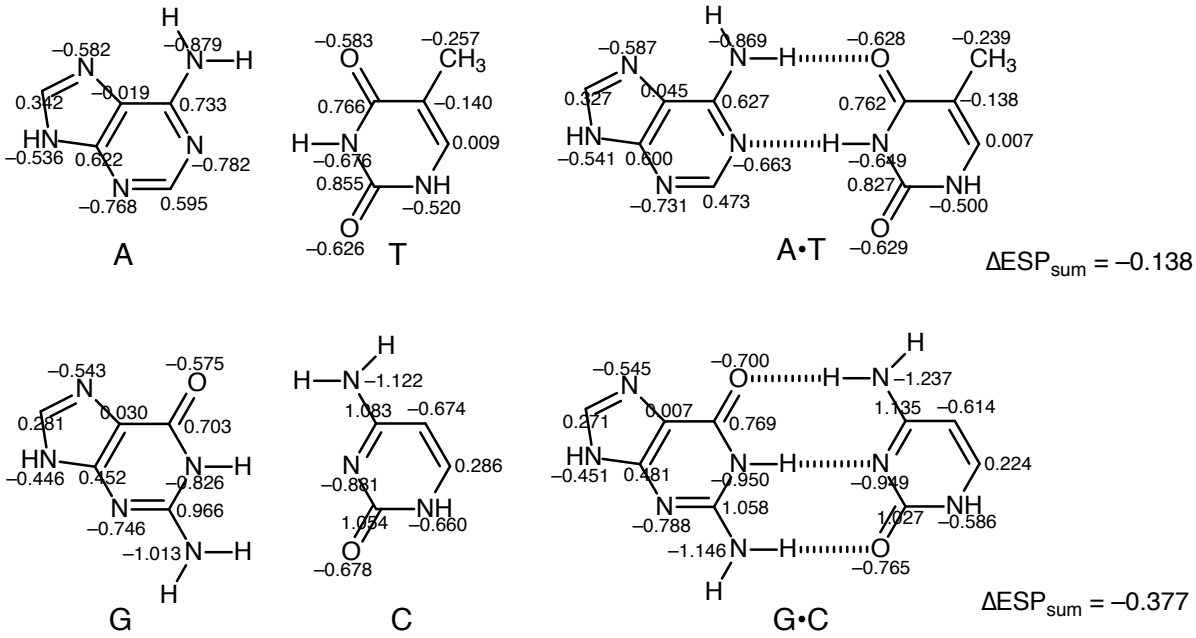
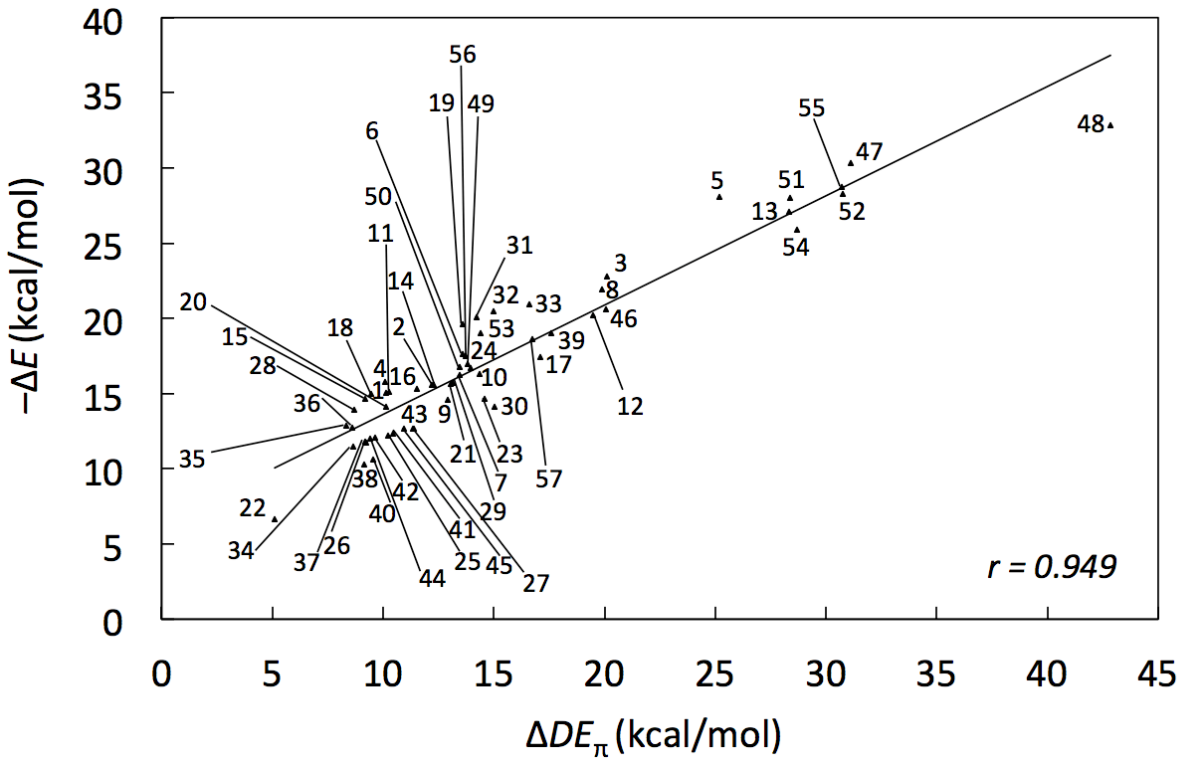


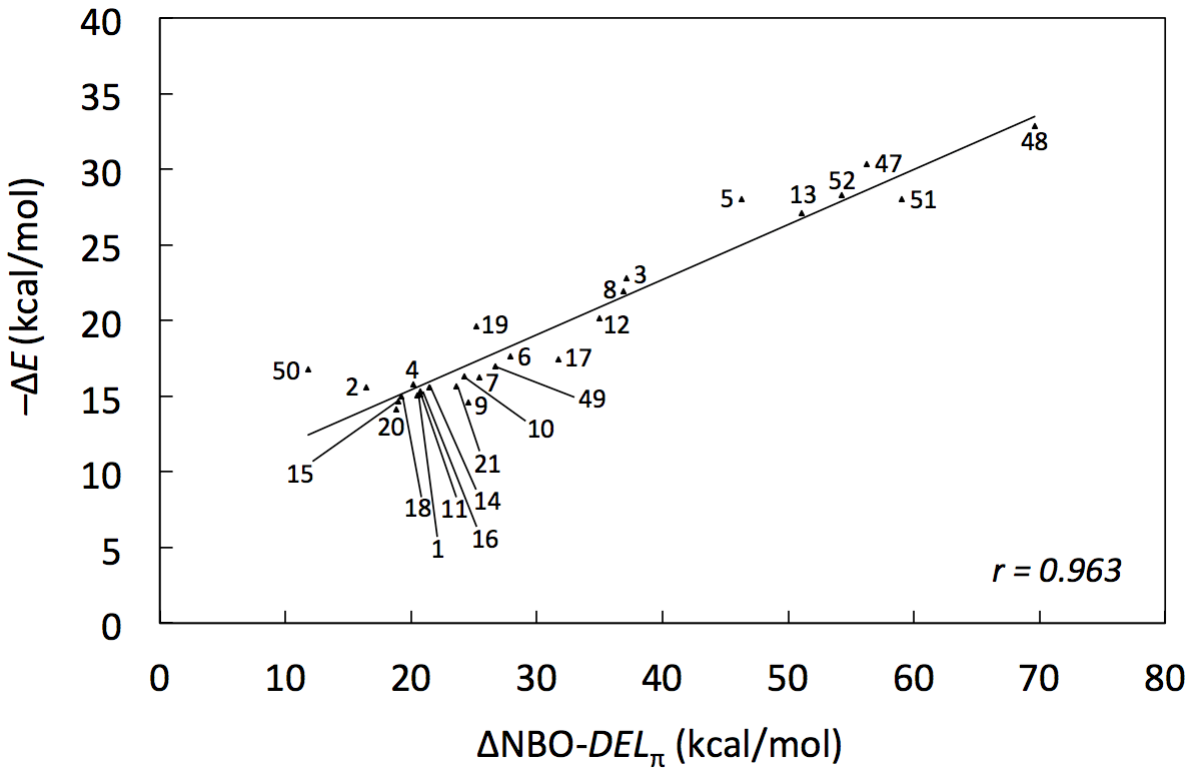
Figure S5. ESP-derived charges for A, T, G, C and the A•T, G•C pairs at a constrained C_s symmetry using the pop=chelpg keyword. Note the much larger change in charge distribution of G and C (upon pairing to form G•C) indicating the greater polarizability of these bases.



Label:

1. A•T	16. Reverse A•U(Hoog)	31. G•G 3	46. I•I
2. A•T(Hoog)	17. Reverse G•U	32. G•6-thioG 3	47. G•yC
3. A_{imino}•C	18. A•yT	33. 6-thioG•G 3	48. IsoC•IsoG
4. A•C_{imino}	19. T(h)•G	34. G•A 2	49. K•Pi
5. A_{imino}•T_{enol}	20. Reverse G•C	35. G•A 4	50. K•X
6. A_{imino}•T	21. Reverse A•C	36. A•A 1	51. G•C
7. I•A	22. G•A_{imino}	37. A•A 2	52. P•Z
8. I•C	23. Syn-G•A	38. A•A 3	53. G*•T
9. I•U	24. G•A	39. 8-oxoG•G	54. 6-thioG•C
10. G•U	25. U•U 1	40. 2-thioU•2-thioU	55. 8-oxoG•C
11. A•U	26. U•U 2	41. T•T 1	56. 2-aminoA•T
12. C•C	27. U•U 3	42. T•T 2	57. A•T(Hoog')
13. G•G	28. A•4-thioU	43. T•T 3	
14. 8-oxoG•A	29. G•4-thioU	44. T•C 1	
15. Reverse A•U	30. G•2-thioU	45. T•C 2	

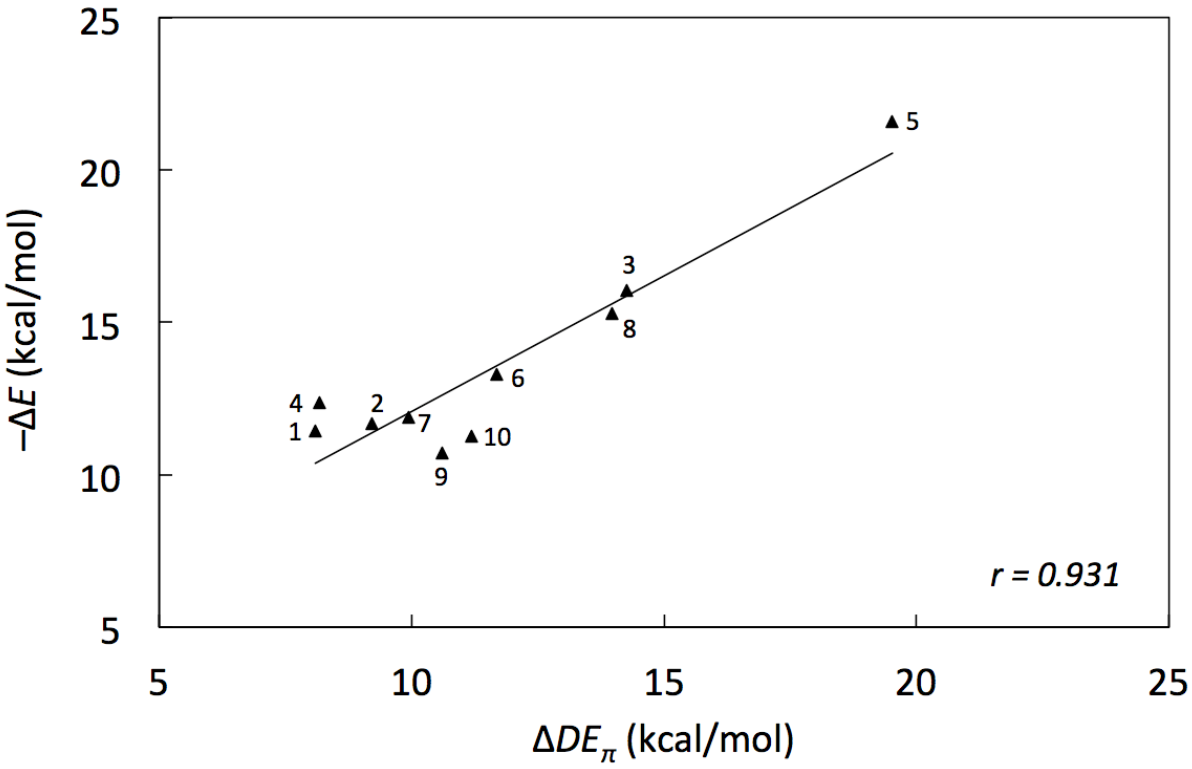
Figure S6. Plot of $-\Delta E$ vs. ΔDE_π for all nucleobase pairs. All geometries were optimized in the gas-phase with a constrained C_s symmetry at the ωB97X-D/6-311+G(d,p) level. BLW computations were performed at B3LYP/6-31G(d).



Label:

1. A•T	8. I•C	15. Reverse A•U	47. G•yC
2. A•T(Hoog)	9. I•U	16. Reverse A•U(Hoog)	48. IsoC•IsoG
3. A_{imino}•C	10. G•U	17. Reverse G•U	49. K•Pi
4. A•C_{imino}	11. A•U	18. A•yT	50. K•X
5. A_{imino}•T_{enol}	12. C•C	19. T(h)•G	51. G•C
6. A_{imino}•T	13. G•G	20. Reverse G•C	52. P•Z
7. I•A	14. 8-oxoG•A	21. Reverse A•C	

Figure S7. Plot of $-\Delta E$ vs. ΔNBO-DEL_π for selected nucleobase pairs in the gas-phase. All geometries were optimized with a constrained C_s symmetry at the ωB97X-D/6-311+G(d,p) level. NBO deletion computations were performed at ωB97X-D/def2-TZVPP.



Label:

1. A•T	4. A•C_{imino}	7. I•A	10. G•U
2. A•T(Hoog)	5. A_{imino}•T_{enol}	8. I•C	
3. A_{imino}•C	6. A_{imino}•T	9. I•U	

Figure S8. Plot of $-\Delta E$ vs. ΔDE_π for selected nucleobase pairs in chloroform. All geometries were optimized with a constrained C_s symmetry at the IEF-PCM-ωB97X-D/6-311+G(d,p) level. BLW computations were performed at B3LYP/6-31G(d).

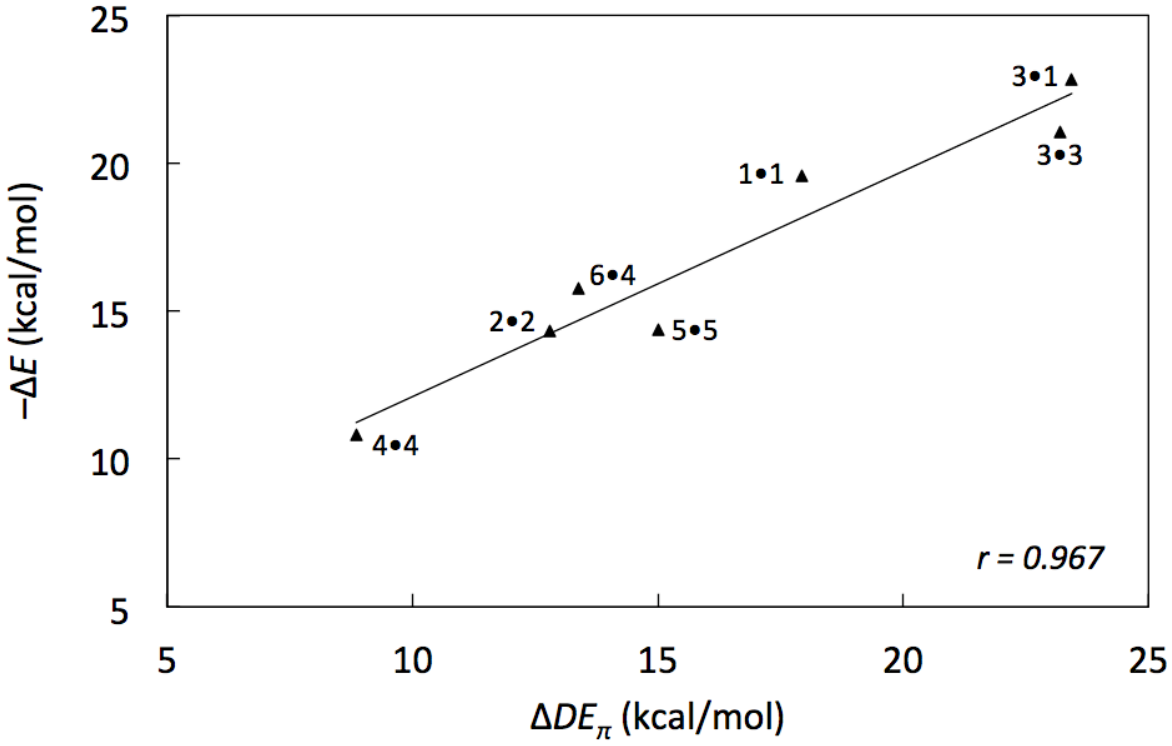


Figure S9. Plot of $-\Delta E$ vs. ΔDE_π for acyclic references considered in the gas phase. All geometries were optimized with a constrained C_s symmetry at the ωB97X-D/6-311+G(d,p) level. BLW computations were performed at B3LYP/6-31G(d).

3.6 References

1. Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.
2. Kyogoku, Y.; Lord, R. C.; Rich, A. *Biochim. Biophys. Acta* **1969**, *179*, 10-17.
3. Kyogoku, Y.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 250-257.
4. Jeong, K. S.; Tjivikua, T.; Muehldorf, A.; Deslongchamps, G.; Famulok, M.; Rebek Jr. J. *J. Am. Chem. Soc.* **1991**, *113*, 201-209.
5. Leonard, G. A.; McAuley-Hecht, K.; Brown, T.; Hunter, W. N. *Acta Crystallogr. Sect. D* **1995**, *51*, 136-139.
6. Quinn, J. R.; Zimmerman, S. C.; Del Bene, J. E.; Shavitt, I. *J. Am. Chem. Soc.* **2007**, *129*, 934-941.
7. Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J.* **1999**, *5*, 3581-3594.
8. Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441-6443.
9. Hobza, P.; Sponer, J.; Cubero, E.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2000**, *104*, 6286-6292.
10. Wu, J. I.; Jackson, J. E.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **2014**, *136*, 13526-13529.
11. Kakeshpour, T.; Wu, J. I.; Jackson, J. E. *J. Am. Chem. Soc.* **2016**, *138*, 3427-3432.
12. Cyrański, M. K.; Gilski, M.; Jaskólski, M.; Krygowski, T. M. *J. Org. Chem.* **2003**, *68*, 8607-8613.
13. Beck, J. F.; Mo, Y. *J. Comput. Chem.* **2006**, *28*, 455-466.
14. Fliegl, H.; Lehtonen, O.; Sundholm, D.; Kaila, V. R. I. *Phys. Chem. Chem. Phys.* **2011**, *13*, 434-437.

15. Quiñonero, D.; Frontera, A.; Ballester, P.; Deyà, P. M. *Tetrahedron Lett.* **2000**, *41*, 2001-2005.

16. Quiñonero, D.; Prohens, R.; Garau, C.; Frontera, A.; Ballester, P.; Costa, A.; Deyà, P. M. *Chem. Phys. Lett.* **2002**, *351*, 115-120.

17. Talens, V. S.; Englebienne, P.; Trinh, T. T.; Noteborn, W. E. M.; Voets, I. K.; Kieltyka, R. E. *Angew. Chem.* **2015**, *127*, 10648-10652.

18. Wu, C.-H.; Zhang, Y.; Rickley, K. v.; Wu, J. I. *Chem. Commun.* **2018**, *54*, 3512-3515.

19. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2013.

20. Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530-5538.

21. Mo, Y.; Song, L.; Lin, Y. *J. Phys. Chem. A* **2007**, *111*, 8291-9301.

22. Mo, Y. In *The Chemical Bond: Fundamental Aspects of Chemical Bonding*; Frenking, G., and Shaik, S., Eds.; Wiley: Weinheim, Germany, 2014; pp 199.

23. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.;Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347-1363.

24. Lu, T.; Chen, F. *J. Comput. Chem.* **2012**, *33*, 580-592.

25. Lu, T.; Chen, F. *J. Mol. Graph. Model.* **2012**, *38*, 314-323.

26. Hobze, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1996**, *18*, 1136-1150.

27. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

28. Case, D. A.; Cheatham, T. E. III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.

29. Jorgenson, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008-2010.

30. van der Lubbe, S. C. C.; Fonseca Guerra, C. *Chem. Eur. J.* **2017**, *23*, 10249-10253.

31. Kool, E. T. *Acc. Chem. Res.* **2002**, *35*, 936-943.

32. Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature*, **1990**, *343*, 33-37.

33. Ishikawa, M.; Hirao, I.; Yokoyama, S. *Tetrahedron Lett.* **2000**, *41*, 3931-3934.

34. Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4922-4925.

35. Glendening, E. D.; Landis, C. R.; Weinhold, F. *J. Comput. Chem.* **2013**, *34*, 1429-1437.

CHAPTER 4

CONCLUSION

Since the concept of aromaticity was proposed in 1855,[1] it has been considered as one of the most important molecular features in physical organic chemistry for a long time. Fully understanding the effects of aromaticity will significantly influence the development of physical organic chemistry. According to Hückel's definition of aromaticity, aromatic molecules all exhibit planar cyclic $4n + 2$ π-electron delocalization effects, which can be quantified by BLW and NBO deletion methods. Hence, understanding the effects of aromaticity by computational methods is one of main purposes in this thesis.

Another important research target in this thesis is the hydrogen bond. Hydrogen bonding interactions also play important roles in many different fields. For example, in biosystems, two nucleobases are linked through a hydrogen bond to give a nucleobase pair, which is the basis of nucleic acid double helix. The genetic information is also encoded in the order of base pairs. Another good example of the application of hydrogen bonding interactions is the enzyme catalysis reaction. Many chemical reactions (*e.g.*, enzymatic hydrolysis of acetylcholine with acetylcholinesterase), are catalyzed by enzymes, which are formed through hydrogen bonds linking with each other. In material science, hydrogen bonding interactions also stand out due to their rigid structures and high recognition specificity. For example, self-assembled elastomers can be tuned by hydrogen-bonded arrays. Hydrogen-bond-mediated stimuli-responsive materials also

[1] Hofmann, A. W. *Proc. Royal Soc. Lond.* **1855**, *8*, 1-3.

attract chemists a lot recently. Multipoint hydrogen-bonded arrays show a promising direction for designing materials and modulating their properties.

Therefore, we applied quantum computational tools to understand the effects of aromaticity gain on the association strengths of multipoint hydrogen-bonded arrays.

Our computational results based on BLW show that the inherent association free energies of multipoint hydrogen-bonded arrays correlate well with the degree of aromaticity gain upon complexation. Excellent correlations were found between the computed gas-phase association free energies and ΔDE_{π} values of 26 triply (r = 0.940) and 20 quadruply (r = 0.959) hydrogen-bonded arrays. Our findings point to the important limitations of the SEI model of Jorgensen and Pranata, which has long guided the understanding of multipoint hydrogen-bonded arrays and their associations in supramolecular chemistry. Clear exceptions to the SEI model were illustrated to emphasize the application of aromaticity gain to understand the hydrogen bonding association strength. Our results also suggested the potential for aromaticity gain in arrays to be listed as a key factor for designing hydrogen-bonded molecular-recognition units in addition to the often used checklist (*i.e.*, numbers, types, and SEI patterns).

Aromaticity gain can also be used to explain the electronic complementarity in base pairing. When given a mixture of A, T/U, G, and C in the primordial soup, A will pair with T/U and G with C instead of to themselves. Attractive idea was proposed in 1967 that the electronic complementarity directs their specific associations. Our computational results provide a possible explanation that the aromatic characters of nucleobases influence their association strengths to complementary bases through a reciprocal

aromaticity-modulated hydrogen bonding (AMHB) relationship. The computed electrostatic potential difference maps show the connection between the differences in the polarizability of nucleobases and their π-conjugation patterns. It demonstrates why fixed-charged approaches adopted by popular force fields, might understabilize certain interactions but overstabilize others. 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.

Overall, our findings emphasize the considerable effects of aromaticity gain on the association strengths of multipoint hydrogen-bonded arrays and nucleobase pairs, suggesting the chemical significance to consider aromaticity gain as a key factor when designing, evaluating and modifying hydrogen-bonded complexes.