TOWARDS A MORE COMPLETE UNDERSTANDING OF STACKING INTERACTIONS INVOLVING DNA

by

DREW PAUL HARDING

(Under the Direction of Steven Wheeler)

ABSTRACT

Despite many years of experimental effort, the factors that impact stacking interactions involving DNA have proved difficult to unravel. This incomplete understanding of these fundamental interactions hinders efforts to design molecules that harness these interactions for sensing and therapeutic purposes. Modern computational tools are well-suited to model DNA-containing systems, and this dissertation describes the application of a variety of these tools to understand stacking interaction involving nucleobase through studies of nucleobase targeting sensors, DNA intercalators, and the stacking of natural and artificial nucleobases. First, I analyze the stacking and hydrogen bonding interactions operating in a recently developed sensor for guanine. The results show that the binding strength can be modulated by a number of factors, including both directly through changing the size and shape of a pendant aryl group that engages in stacking interactions with the bound guanine and indirectly via changes in the number and quality of hydrogen-bonding interactions between the bound guanine and the linker connecting the pendant aryl group and the rest of the sensor. Second, I establish guidelines for the quantum mechanical (QM) study of DNA intercalators, demonstrated

the performance of a range of density functional theory (DFT) methods and their limitations when applied to intercalated DNA. Furthermore, I show that the required size of the DNA binding site model in such computations depends on the binding mode of the intercalator and its protonation state. Finally, I develop a quantitative and qualitative model for the intrinsic stacking of natural and artificial nucleobases. Using this predictive model, stacking interaction energies between artificial and natural nucleobases can be rapidly predicted based solely on a descriptor derived from the electrostatic potentials (ESPs) of the nucleobases as well as their heavy-atom counts. This ESP-derived descriptor plainly illustrates how the unique electrostatic character of the different nucleobases modulates the strength of stacking interactions. Together, these studies provide key new insights into the nature of stacking interactions, with important implications for understanding their role in biological systems and exploiting these interactions in numerous biological and non-biological contexts.

INDEX WORDS: computational chemistry, density functional theory, DNA, noncovalent interactions, molecular descriptors, nucleobases, electrostatic potentials, multivariate linear regression

TOWARDS A MORE COMPLETE UNDERSTANDING OF STACKING INTERACTIONS INVOLVING DNA

by

DREW PAUL HARDING

B.S., Louisiana State University, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

TOWARDS A MORE COMPLETE UNDERSTANDING OF STACKING INTERACTIONS INVOLVING DNA

by

DREW PAUL HARDING

Major Professor:	Steven E. Wheeler
Committee:	Henry F. Schaefer III
	Robert Woods

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
December 2019

DEDICATION

To my wife Jennifer, you give me all the strength I need for success. You are the reason this has materialized.

ACKNOWLEDGEMENTS

I want to thank everyone who helped and supported me throughout this PhD program, I could not have done it without you. First, I want to thank my research advisor Dr. Steven E. Wheeler, who took me into his group, cultivated my research, and taught me so much. Thank you for all the opportunities you put before me. I also want to thank all the Wheeler group members who have listened to my struggles, sat through my presentations, and given me great advice over the years: SharathChandra, Stephen, Bryan, Andrea, Yanfei, Rajat, Trevor, Yi, Victoria, Laura, and Tony. I want thank the undergraduate group members as well, who have trusted me to mentor them and in return have taught me how to be a better teacher and leader: Juni, Lauren, Lou, Patrick, Bella, Deena, and Hannah. I want to thank my committee members for their support and continued interest in my success. Thank you for all your advice and wisdom.

I want to thank my family and friends for everything they have done to support me getting to this point. Mom and Dad, thank you for always ensuring I have everything I need to succeed. I also want to thank my mother and father-in-law, Lisette and Paul West, for their constant generosity and guidance. Reese and Sissy, thank you for your encouragement and always making time for me when I would sporadically and hurriedly visit. I also want to thank my uncle Damon and step-father Jim, who both never missed an opportunity to cheer me up and give me confidence. Thank you to all my friends, I can’t list you all but I want you to know that you are responsible for me making it here. I

also want to thank my church family at Resurrection Presbyterian Church, your prayers and encouragement have had a lasting impression on me.

All my grand-parents who couldn't be here, I would have loved for you to see me achieve this.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

As a fundamental component to almost all life, DNA holds information required to promote health and battle illness. Our understanding of its behavior has brought about techniques that range from repairing DNA that may have a mutated gene to destroying the DNA of cancer cells. Beyond its uses as biological blueprints and therapeutic targets, the utility of DNA has been expanded to include materials, sensors, and supramolecular architectures.[1-6] Yet, there are many behaviors of DNA we still do not completely understand.

Paradoxically, the DNA of life in its many forms is chemically simple, with a homogeneous backbone connecting only two naturally occurring nucleobase pairs stacked upon each other. It is the sequence of these two base pairs that results in the functional diversity of DNA. Key to the structure and stability of DNA are stacking interactions between adjacent base pairs. Stacking interactions, most broadly defined as roughly parallel non-covalent interactions between planar π-systems,[7-11] have long been known.[12] However, we still lack a complete understanding of these interactions in complex systems such as DNA. A deeper understanding of stacking interactions in these systems will give us more insight into intracellular DNA interactions and small molecule-DNA interactions while also expanding the utility of DNA in nanotechnology.

There have been tremendous advances in our understanding of stacking interactions involving small aromatic systems over the last two decades.[7-11,13-18] Many of

these advances have been based on computational studies of simple aromatic systems using either high-accuracy *ab initio* methods or dispersion-corrected density functional theory (DFT). Despite this progress in understanding the stacking of simple substituted benzenes and monocyclic heterocycles, the stacking interactions of DNA nucleobases are still not well understood. This arises partly because the main driving force for stacking interactions differs markedly between the gas-phase and aqueous solution. This has led to vastly different views of stacking interactions between experimentalists (who tend to focus on solution-phase stacking interactions) and quantum chemists (who are often stuck in the gas-phase).

Chapters 2-4 of this dissertation describes three projects in which modern computational tools have been used to gain insights into stacking interactions involving DNA nucleobases. The first of these is related to the development of improved molecular sensors for guanine. Guanine-rich DNA sequences can form non-canonical structurers such as G-quadruplexes or Z-DNA, which can impact a number of biological processes.[19-20] Beyond improving sensor design, a more detailed understanding of the these sensors can provide insights into non-covalent interactions with nucleobases for molecules ranging from RNA-targeting inhibitors[21-22] to DNA intercalators.[23-27] I describe a study of a guanine sensor developed by Fang *et al*.[28] in 2006. This sensor is based on a 1,8-naphthyridine-2,7-diamine unit that engages in hydrogen bonding interactions with guanine combined with a pendant pyrenyl group to further enhance binding through stacking interactions. I used *ab initio* methods and DFT to study both model guanine···1,8-naphthyridine dimers as well as guanine bound to Fang's sensor as well as analogs bearing different stacking arenes. The results reveal that guanine-binding

energies can be modulated both directly through changes in the stacking interactions with the pendant arene as well as through the secondary effects of this arene on hydrogen-bonding interactions with the linker connecting the arene to the rest of the sensor.

The second project involves DNA intercalators, which are small molecules that non-covalently bind DNA by inserting between adjacent stacked base pairs. Upon intercalator binding, the change in DNA structure inhibits various biological functions.[29] Some naturally occurring DNA intercalators, such as doxorubicin and proflavine, have had their pharmaceutical significance realized and are used to treat a range of illnesses.[30-31] As the therapeutic potential of DNA intercalators is further explored, computational tools able to accurately model these systems will be needed.[32] Unfortunately, computational drug-design tools have been mainly developed to target proteins and as a result are less well adapted for DNA.

Docking is a key example of a computational tool that, while widely used in protein based computational drug discovery, is difficult to apply in the context of DNA. Given that most docking protocols utilize a rigid model approach, a binding pocket must already be present. This requires either a crystal structure of intercalated DNA or the creation of an artificial binding pocket.[33-34] Moreover, the scoring functions used in docking still lack sufficient accuracy to provide reliable predictions for DNA intercalation.[33] At the same time, classical molecular mechanics (MM) force fields have only recently achieved sufficient accuracy to reliably reproduce canonical DNA structures over significant simulation times.[35] Since these force fields were parameterized for canonical DNA structures, it is unlikely that they will adequately account for the structural changes that accompany intercalation, let alone reliably predict

the strength and specificity of intercalator binding. One route to improved docking scoring functions and MM potentials for DNA intercalation is through parameterization based on reliable quantum mechanical (QM) computations. However, this requires that a sufficiently reliable yet computationally tractable QM method be identified.[36]

There have been numerous QM studies of intercalators.[37-55] For example, in 2011, Pulay *et al.*[42] reported an analysis of the intercalation of ethidium into a UA/AU step of RNA using second order Møller-Plesset perturbation theory (MP2) with the 6-31G(*d*,*p*) basis set in an effort to quantify the importance of pairwise interactions with individual RNA components (nucleobases, sugar-phosphate backbone, *etc*). To this end, they considered truncated models containing the adjacent base pairs with and without the sugar-phosphate backbone. While they were able to devise fragmentation schemes that reproduced the interaction energy of the intact system, they noted that the success of such schemes will vary on a case-by-case basis.[42] That same year, Sherrill *et al.*[43] investigated protonated proflavine intercalated into a CGA trinucleotide using functional-group based symmetry-adapted perturbation theory (F-SAPT). They found that including the backbone in such computations is essential to compute reliable interaction energies, because the interaction of the intercalator with the backbone contributed a third of the total interaction energy. In 2016, Kokoschka *et al*.[56] compared QM intercalator-nucleobase interaction energies to the change in electrochemical impedance spectroscopy measurements of DNA from intercalation. The QM computations included SCS(MI)-MP2/cc-pVTZ interaction energies evaluated at PBE-D3/cc-pVTZ optimized structures of proflavine, ellipticine, and 1-pyrenemethylamine interacting with a single base pair or a single nucleobase as models of double and single stranded DNA. They concluded that

these models were enough to capture qualitative trends observed in their experimental results. While these and other studies[37-55] have provided key insights into DNA intercalation, they employed a wide range of binding site model sizes (with little justification) since there exist no guidelines for the number of bases and backbones that must be included to compute reliable intercalator interaction energies.

To address this shortcoming, in Chapter 3 I quantify errors in QM-computed intercalator binding energies arising from the choice of QM method and DNA binding site model size. The results reveal several DFT methods that provide reliable binding energies for small models, compared to data from reliable *ab initio* methods. Additionally, I show that the size of the binding site model required to compute accurate interaction energies is dependent on the protonation state and binding mode of the intercalator. In particular, neutral intercalators that bind parallel the DNA-base pairs require only the inclusion of the nearest base pairs and backbones, while intercalators that are either charged or bind perpendicularly can require models comprising several flanking nucleotides. To address the need to consider large binding site models in QM studies of DNA intercalators, I show that the computationally inexpensive HF-3c method[57] provides accurate binding energies and can be routinely applied to large binding site models in order to estimate binding energies.

These two studies revealed an inadequate understanding of the *intrinsic* stacking interactions of nucleobases, which we define as the maximum possible stacking interaction for a given pair of nucleobases independent of geometric constraints imposed in biological macromolecules. There is a long history of experimental efforts aimed at informing general models to explain trends in nucleobase stacking interactions. For

example, in the 1990s Gellman *et al.*[58-59] proposed that complementary local charges (*i.e.* local dipole moments) drive nucleobase stacking based on ^{1}H NMR data for model stacked dimers. In Gellman's model, desolvation and dispersion effects are inconsequential. In 2000, Kool *et al.*[60] indirectly countered Gellman's assertion in a broad study of the factors that impact the stacking of both natural and artificial nucleobases. Kool attempted to correlate hydrophobicity, polarizability, excluded solvent surface area, and dipole moments of natural and artificial nucleobases with experimental stacking data derived from measures of the helical stability of DNA modified with the addition of terminal natural or artificial nucleobase. Kool found that solvation driven hydrophobic effects, solvent excluded surface-area, and polarizability correlated the strongest to stacking interactions for both artificial and natural nucleobases, suggesting that electrostatic effects play a negligible role. However, these correlations were much weaker when considering the natural nucleobases alone. One reason that these simple molecular descriptors appear capable of capturing stacking trends for non-natural nucleobases, yet fail for the natural nucleobases, is that global molecular quantities such as polarizability and molecular dipole can only reliably capture intermolecular interactions at long distances. Similarly, relating quantities such as solvent excluded surface area and overall aromatic size to stacking strength neglects the role of heterogeneities in the electrostatic environment around aromatic systems. Such variations will be small in homogenous planar molecules, such as the polycyclic aromatic hydrocarbons that constituted the artificial nucleobases considered by Kool *et al.*,[60] but this is not the case for the natural nucleobases.

Overall, Kool *et al.*[60] found that the stacking strength of the natural nucleobases followed the trend A > G > T = C. This is in conflict with other experimental results.[61-64] For example, G|C stacking was shown to increase nearest-neighbor stability over A|T from melting experiments of oligonucleotides.[61] In a separate study,[62-63] experimentally measured equilibrium stacking energies revealed stronger stacking interactions of G-over A|T. These results are also supported by single-molecule dissection studies of nucleobase stacking,[64] which also showed the stacking strength of G|C is greater than A|T. Overall, these experiments consistently show nucleobase stacking strength is ranked G > C $\geq$ A > T.[61-62,64] These results are further supported by experimentally derived thermodynamic parameters of natural nucleobases used to predict DNA thermal denaturation.[61,65-66]

Running in parallel with these experimental efforts to reveal trends in nucleobase stacking is an extensive body of quantum chemical studies of nucleobase stacking,[67-96] many of which aimed to elucidate the nature of these interactions. The most comprehensive of these is the review by Šponer *et al.*,[93] in which they conclude that a balance of short-range repulsion and dispersion interactions govern stacking interactions of nucleobases. They further note that while the overall dipolar character of the nucleobases leads to antiparallel alignment of the molecular dipoles, any resulting enhanced electrostatic interactions are largely counterbalanced by the desolvation cost of bringing the two nucleobases together. Overall, these quantum chemical studies[70-80] have contributed to our understanding of the role of key interactions (electrostatics, dispersion, etc.) in nucleobase stacking. However, a robust physical model of the underlying trends has remained elusive.

The net result is that we lack both a practical means of quantitatively predicting the stacking strength of aromatic molecules with nucleobases (short of expensive QM computations on stacked dimers) and a qualitative understanding of the origin of the established trends in stacking of the natural nucleobases. Computational quantum chemistry can be used to fill these gaps. One approach is the utilize more nuanced molecular descriptors combined with multivariate linear regression (MLVR) methods to develop predictive models of stacking interactions. For instance, Bootsma *et al*. introduced new heterocycle descriptors[97] based on DFT-computed electrostatic potentials (ESPs) and electric fields that capture the heterogeneities in the electrostatic environment of the heterocycle. These descriptors have been used to develop robust models of stacking interactions of amides,[98] Asp-Arg salt-bridges,[97] and aromatic amino acids[99] with drug-like heterocycles. Such predictive models can be used to rapidly predict stacking interactions based on readily computed molecular properties of the monomers, avoiding the need to tedious searches for local stacked energy minima. Analogous models for nucleobase stacking would be advantageous both for predicting the potential stacking interactions of natural and non-natural nucleobases and could provide physical insight into the factors that drive trends in nucleobase stacking.

Chapter 4 describes a systematic study of stacking interactions of natural and artificial nucleobases resulting in the development of a quantitative predictive model of these interactions as well as a deeper physical understanding of the factors that underlie trends in stacking interactions among nucleobases. A thorough search of local stacked energy minima reveals clear trends in intrinsic stacking abilities of the natural nucleobases: G > C $\geq$ A > T > U. Further analysis using symmetry-adapted perturbation

theory (SAPT) confirms previous findings[93] that dispersion interactions dominate these interactions but identifies electrostatic effects as the main modulator of stacking strength. Equipped with this insight, I used recently developed heterocycle descriptors from Bootsma *et al.*[97] in combination with MVLR methods to develop a predictive model of the intrinsic stacking of nucleobase-like heterocycles with the natural nucleobases. Further examination of the electrostatic potentials that underlie these descriptors provides a physical understanding of the above-mentioned trend in intrinsic stacking strengths.

CHAPTER 2

BETTER SENSING THROUGH STACKING: THE ROLE OF NON-COVALENT INTERACTIONS IN GUANINE-BINDING SENSORS[a]

[a] Harding, D. P; Bootsma, A. N; and S. E. Wheeler. *J. Phys. Chem. B* **123**,487 (2019).
Reprinted here with permission of the publisher

Abstract

A series of aryl-substituted naphthyridine-based sensors for 9-alkylguanine was analyzed using density functional theory (DFT) and correlated *ab initio* methods. First, the 2-acetamido-1,8-naphthyridine backbone of these sensors was examined with rigorous *ab initio* methods and was shown to exhibit a guanine-binding energy commensurate with that of cytosine. Second, computational analyses of a guanine-specific fluorescent sensor from Fang and co-workers (*Org. Lett.* **2016,** *18*, 1724) resulted in a revised binding model and showed that π-stacking interactions with a pendant pyrenyl group are vital for strong guanine binding. Finally, 24 related guanine sensors with varying aryl groups were studied. Overall, it was found that both the geometry and point of attachment of the pendant aryl groups significantly impact the guanine binding affinity. This occurs through both the direct modulation of the π-stacking interactions with guanine as well as secondary geometric effects that influence the strength and number of hydrogen bonds between guanine and the ethylenediamine linker connecting the arene to the naphthyridine backbone.

I. Introduction

There have been considerable advances in the last decade in the development of small molecular receptors for everything from inorganic ions and clusters[100-104] to arenes and biomolecules.[7-11] Much of this growth has been driven by an increased understanding of the underlying non-covalent interactions.[7-18,105] The development of sensors for specific DNA bases is a particularly important application of molecular receptors, with many potential biomedical uses. Guanine-rich DNA sequences, for

example, can form G-quadruplexes that impact a number of processes including the activity of telomerase.[19] Moreover, detailed computational studies of these sensors can provide insights into the nature of non-covalent interactions involving nucleobases, which play a role in everything from the structure of DNA and RNA to the behavior of DNA intercalators[23-27] and the development of RNA-targeting small molecule inhibitors.[21-22]

A selective guanine sensor should take advantage of non-covalent interactions that are unique to guanine. The most obvious of these is its ADD (A = acceptor, D = donor) hydrogen bonding motif, and guanine sensors typically exhibit the same complementary DAA hydrogen bonding pattern found in cytosine. However, the ability of guanine to engage in strong π-stacking interactions with arenes can also be exploited, which has the ancillary benefit of limiting the conformational flexibility of the bound base with the hydrogen bonding unit and leading to increased binding affinity.

The desired DAA hydrogen bonding pattern can be found in 2-acetamido-1,8-naphthyridine (**1**, Figure 1), derivatives of which were first utilized for guanine binding by Giese *et al.*[106] in the 1980s for hydrogen bonding chromatography. Subsequently, derivatives of **1** have been exploited in a number of guanine binding applications.[28,107-123] For instance, Fang *et al.*[121-122] developed a series of ethynyl bridged naphthyridine derivatives capable of high-yield fluorescence-based sensing, including sensors **2** and **3** (Figure 1). These sensors, which form up to four hydrogen-bonds with guanine, exhibit binding constants up to $(9.1 \pm 0.2) \times 10^4$ M^{-1}.

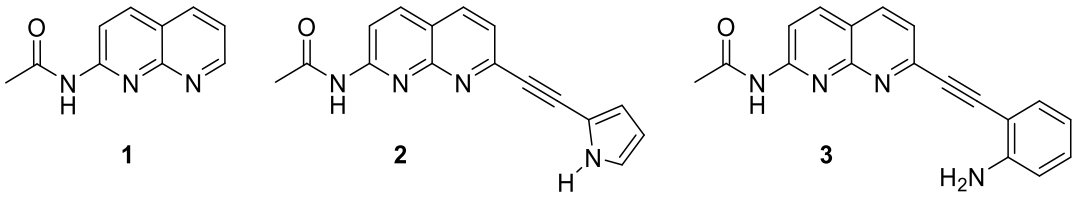

Figure 1. 2-acetamido-1,8-naphthyridine (**1**) and two of the first-generation guanine sensors from Fang *et al*. (**2** and **3**).[121]

In 2016, Fang *et al*.[28] developed a new generation of fluorescent sensors for 9-decylguanine in dichloromethane based on derivatives of **3.** Through the introduction of an ethylenediamine linker on the phenylamine nitrogen, the binding constant was improved to 2.3 x 10^5 M^{-1}. Incorporation of a pyrenyl substituent on the ethylenediamine chain (sensor **8b**, Figure 2) further enhanced the binding constant to 1.4 x 10^6 M^{-1}. Fang *et al*.[28] proposed the binding mode presented in Figure 3a, which features three hydrogen bonds with the acetamido-naphthyridine, a hydrogen bond with the amino linker, and π-stacking interactions of guanine with the pyrene.

Despite the demonstrated utility of naphthyridine derivatives in guanine-binding applications,[28,107-123] there have been only limited computational studies of hydrogen bonded guanine-naphthyridine complexes. For instance, in 2007, Fang *et al*.[121] reported semi-empirical PM3 optimized geometries of complexes of 9-alkylguanine with sensors **2** and **3**. Cywinski *et al*.[113] used B3LYP/6-31+G(*d*) computations in their development of a selective fluorescent guanosine sensor similar to **3**, while Liang *et al.*[116] used DFT computations to design a 1,8-naphthyridine-2,7-diamine to selectively bind Watson-Crick base pairs. More recently, Dyubankova *et al.*[123] used molecular dynamics simulations to study the binding of a derivative of **1** with a guanine bulge of an RNA hairpin. However,

there have been no rigorous *ab initio* computational studies of complexes between guanine and 2-acetamido-1,8-naphthyridine derivatives to date, nor has anyone computationally studied bound complexes of sensor **8b** with guanine. This has resulted in an incomplete understanding of the factors that lead to sensitive, specific guanine binding sensors, which has hampered the further development of improved sensors.

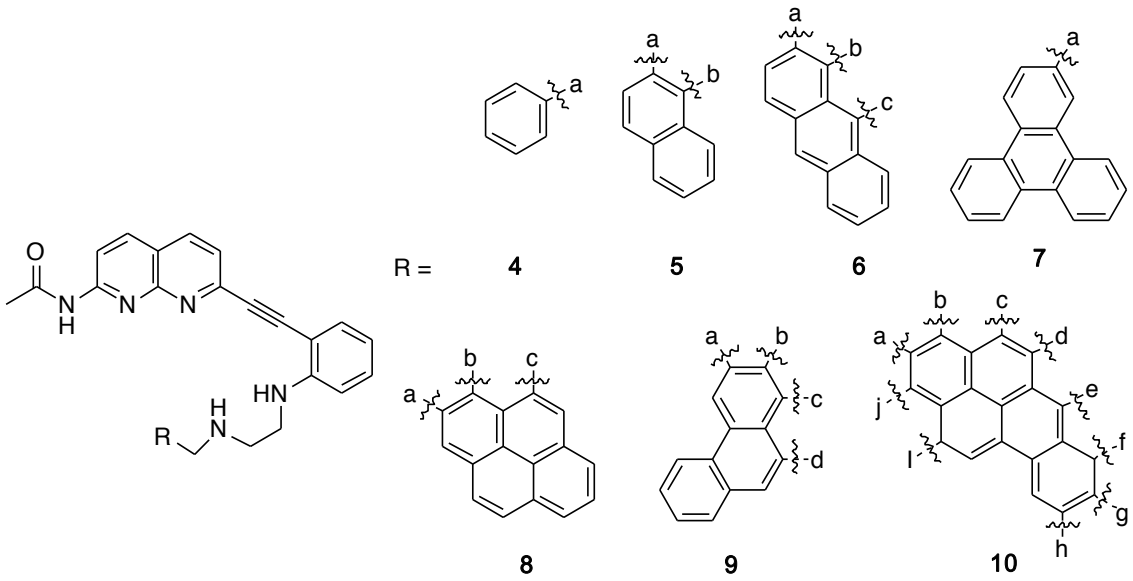

Figure 2. Second-generation guanine sensor from Fang *et al*.[28] (**8b**), along with analogs featuring other aryl substituents R.

Herein, we report a computational study of the hydrogen bonding complex formed by 9-methylguanine (G, as a model of 9-alkylguanines) and 2-acetamido-1,8-naphthyridine (**1**), followed by an exploration of complexes of guanine with Fang's fluorescent sensor (**8b)**. Finally, we consider how the size and shape of the pendant aryl group in 23 analogs of **8b** (**4**-**10**, Figure 2) impact the binding strength. Ultimately, we show that 2-acetamido-1,8-naphthyridine is a near perfect binding partner for guanine, the sensor devised by Fang *et al*.[28] (**8b**) binds guanine through more non-covalent

interactions than originally proposed, and that maximizing the guanine binding affinity of these sensors requires the inclusion of an aryl group that can engage in strong π-stacking interactions with both guanine and the arene core of the sensor while not disrupting the hydrogen-bonding network between the ethylenediamine linker and guanine.

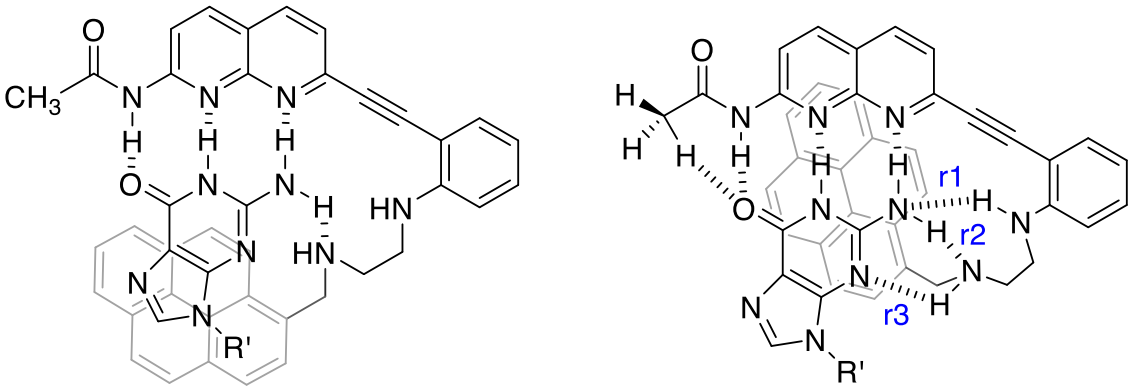

Figure 3. (a) Binding mode proposed by Fang *et al*. [28] for **8b** with 9-alkylguanine and (b) revised binding mode based on the present work. Hydrogen bonds that play a key role in determining the guanine-binding strength are labeled (r1-3).

II. Theoretical Methods

Given the conformational flexibility of sensors **4**-**10**, a comprehensive conformational search was required to identify the lowest-lying conformation of both the isolated sensors and their complexes with guanine. This was achieved by taking 10,000 steps using the enhanced systematic torsional sampling method in MacroModel[124] with the MMFF force-field and chloroform as the solvent (as an approximation to dichloromethane). The resulting structures that were within 3.0 kcal mol^{-1} of the global minimum were further optimized at the B97-D/3-21G level of theory.[125-126] The lowest lying 33% of these structures were then optimized at the M06-2X/6-31G(*d*) level of

theory.[127] Finally, single point energies were computed at the M06-2X/6-311+G(*d*) level of theory for the lowest-lying M06-2X/6-31G(d) optimized geometries. The solvent (dichloromethane) was accounted for using IEF-PCM in all DFT computations,[128-129] which were done using Gaussian09.[130]

For each sensor, we present several computed quantities to characterize the binding. First, the binding energy (E_{bind}) is defined as the difference in energy (excluding zero-point vibrational energies) between the bound complex and the global minimum optimized geometry of the isolated sensor and guanine. The interaction energy (E_{int}) corresponds to the difference in energy between the bound complex and the isolated sensor and guanine in the geometry of the complex. Finally, the distortion energy (E_{dist}) is the difference between the binding and interaction energy, which is dominated by changes in the sensor geometry. For each complex, E_{int} was approximately decomposed into individual pairwise interactions between guanine and different components of the sensor, as discussed in the text.

To assess the reliability of the chosen level of theory, we also considered model hydrogen bonding complexes of G with **1** and with 1-methyl-cytosine (C). Interaction energies were computed at M06-2X/6-31G(d) optimized geometries using multiple levels of theory, including M06-2X/6-311+G(d) and ωB97X-D/def2-TZVP,[131-134] and compared to data from rigorous *ab initio* methods. For the latter, we relied on DLPNO-CCSD(T)/cc-pVXZ[135-138] energies (computed using Orca 4.0[139]) extrapolated to the complete basis set limit within the focal point approach of Allen and co-workers.[140-142] These focal point extrapolated energies are well-converged to the *ab initio* limit (see Appendix A, Table A-1) and hence serve as highly-accurate benchmark data.

Additionally, symmetry-adapted perturbation theory (SAPT)[143-146] was applied at the SAPT0/jun-cc-pVDZ level of theory using Psi4[147] to decompose the interaction energies in these model complexes into contributions from electrostatic, dispersion, induction, and exchange-repulsion effects.[148] Recent benchmarks by Sponer and co-workers[149] have shown that SAPT0 provides reliable interaction energies for DNA bases.

III. Results and Discussion

A. Complex of Guanine with 2-Acetamido-1,8-naphthyridine

We first consider model hydrogen bonded complexes of G with **1** and with 1-methylcytosine (C) to both identify suitable computational methods for studying complexes of guanine with sensors **4-10** and to assess the guanine-binding abilities of **1**. Optimized geometries (at the M06-2X/6-31G(d) level of theory) of the **1**···G and C···G complexes are shown in Figures 4a and 4b. It is established that M06-2X accurately describes both hydrogen bonding and dispersion-driven non-covalent interactions.[127] However, to assess the sensitivity of interaction energies for these two complexes to computational method, we computed gas-phase single point energies at a number of levels of theory (see Table 1 and Appendix A, Table A-2). DLPNO-CCSD(T) interaction energies extrapolated to the complete basis set (CBS) limit for **1**···G and C···G are –29.2 and –31.5 kcal mol^{-1}, respectively. Satisfyingly, the three DFT methods considered provide similar predicted values. Moreover, the M06-2X/6-311+G(d)//M06-2X/6-31G(d) data (–29.7 and –32.1 kcal mol^{-1}) are in excellent agreement with these robust *ab initio* methods, with the relative interaction energy within 0.1 kcal mol^{-1}. While counterpoise corrections[150] provide improved interaction energies at the M06-2X/6-31G(d) and

ωB97X-D/def2-TZVP levels, they do not for M06-2X/6-311+G(d). Given the excellent agreement with DLPNO-CCSD(T)/CBS interaction energies (and binding energies, see Appendix A, Table A-2), data for the larger systems examined below were computed at the M06-2X/6-311+G(d)//M06-2X/6-31G(d) level of theory (without counterpoise corrections).

All the methods in Table 1 predict a ~2.0 kcal mol^{-1} difference in interaction energy between these two complexes, favoring C···G. We applied symmetry adapted perturbation theory at the SAPT0/jun-cc-pVDZ level[143-146,151] to understand the origin of this difference (see Table 1). The most significant contributor to the stronger interaction energy in C···G compared to **1**···G is the electrostatic component, which favors C···G by 3 kcal mol^{-1}. This is complemented by induction effects but tempered somewhat by dispersion, which contribute –0.7 and +2.0 kcal mol^{-1} to the difference in interaction energies, respectively. Exchange repulsion effects are essentially equal in these two complexes, contributing only 0.2 kcal mol^{-1} to the relative interaction energies.

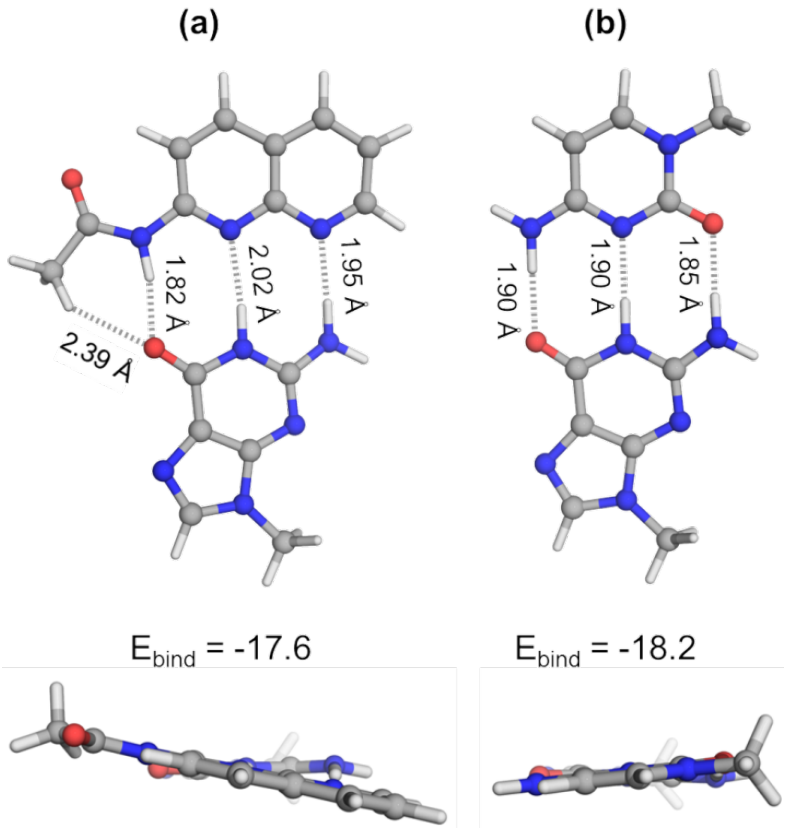

Figure 4. Two views of the M06-2X/6-31G(d) optimized complexes of guanine (G) with (a) **1** and (b) cytosine (C), along with M06-2X/6-311+G(*d*) predicted binding energies (E_{bind}) in dichloromethane (kcal mol^{-1}).

This difference in electrostatic components is reflected in the geometries of the hydrogen bonding interactions. That is, the two NH···N hydrogen bonding distances in C···G are more than 0.1 Å shorter than the corresponding distances in **1**···G. These effects are offset slightly by the NH···O interaction, which is shorter for **1**···G than for C···G and the presence of a non-classical hydrogen bond[152] between this carbonyl oxygen and a methyl hydrogen that is absent in C···G. Once monomer relaxation and solvation effects are accounted for, the solution-phase binding energy for **1**···G (–17.6 kcal mol^{-1}) is only slightly less favorable than that for C···G (–18.2 kcal mol^{-1}). Thus, 2-acetamido-1,8-

naphthyridine (**1**) provides an excellent biomimetic hydrogen bonding partner for guanine, forming the backbone of the sensors developed by Fang *et al*.[28] and others.[107-123]

Table 1. Gas-phase interaction energies (in kcal mol^{-1}) for complexes of 9-methyl-guanine (G) with **1** and 1-methyl-cytosine (C), as well as the difference between the two, at various levels of theory computed at M06-2X/6-31G(d) optimized geometries. For SAPT0, the individual components of the interaction energy are also provided.

	1···G	**C···G**	**Diff.**
M06-2X/6-31G(d)	–32.7	–35.2	–2.5
M06-2X/6-31G(d)[b]	–28.6	–31.4	–2.8
M06-2X/6-311+G(d)	–29.7	–32.1	–2.4
M06-2X/6-311+G(d)[b]	–28.4	–30.8	–2.4
ωB97X-D/def2-TZVP	–30.7	–32.5	–1.8
ωB97X-D/def2-TZVP[b]	–29.8	–31.9	–2.1
DLPNO-CCSD(T)/CBS[a]	–29.2	–31.5	–2.3
SAPT0/jun-cc-pVDZ	–32.7	–34.8	–2.1
Electrostatics	–41.7	–44.8	–3.1
Exchange Repulsion	37.5	37.3	–0.2
Induction	–17.1	–17.8	–0.7
Dispersion	–11.4	–9.4	2.0

[a] DLPNO-CCSD(T) energies extrapolated to the complete basis set limit.
[b] Counterpoise corrected[150]

B. Fang's Fluorescent Sensor

Next, we consider Fang's fluorescent sensor (**8b**) to assess the accuracy of the proposed binding mode in Figure 3a[28] and to understand the contributions of various non-covalent interactions to the overall binding energy. The optimized geometries of **8b** both with and without guanine bound are shown in Figure 5. Overall, Fang's proposed binding model is qualitatively correct, with only slight deviations from the DFT optimized structure. A revised version of Fang's proposed binding mode is depicted in Figure 3b. The most significant difference between Fang's proposed binding mode and the fully optimized bound complex is the position of the pyrenyl group. In particular, in the optimized structure this group is engaged in π-stacking interactions with both the guanine and the naphthyridine backbone of the sensor. This intramolecular π-stacking interaction between the aryl substituent and naphthyridine is instrumental in determining the overall binding geometry, and will prove important when considering analogs of **8b** (*vide infra*). The optimized complex of **8b** with guanine also shows two hydrogen bonds between the exocyclic amino group of guanine and the ethylenediamine linker of the sensor in addition to a non-classical CH···O interaction[152] between the amide group on the sensor and the carbonyl oxygen of guanine.

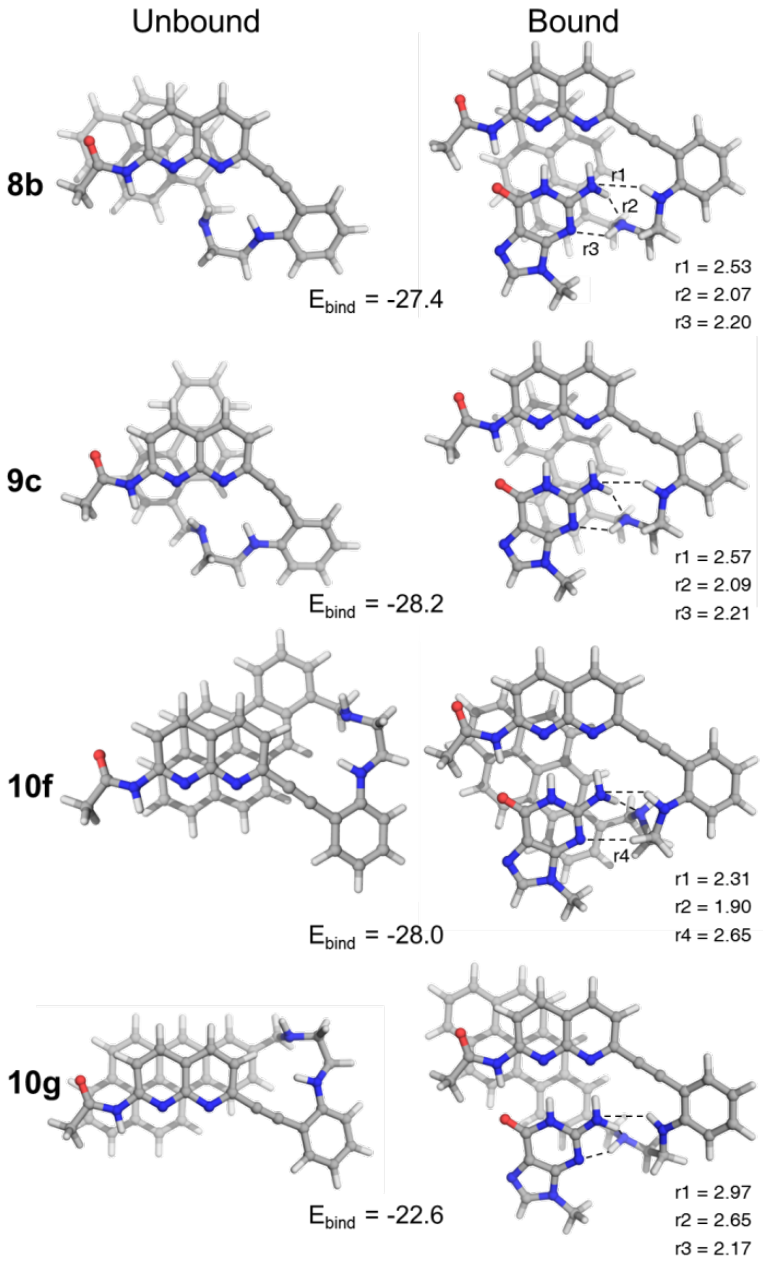

Figure 5. Optimized geometries of selected sensors with (right) and without (left) guanine bound, along with predicted solution-phase binding energies in kcal mol^{-1}.

The predicted binding energy for the complex of **8b** with guanine is –27.4 kcal mol^{-1} in dichloromethane. This is nearly 10 kcal mol^{-1} stronger than with 2-acetamido-1,8-naphthyridine (**1**) alone. Unlike **1**, **8b** is flexible, and the sensor conformation differs

significantly in the bound and unbound state. The result is a significant amount of distortion energy required for the sensor to adopt the conformation found in the bound sensor. In the gas-phase, the interaction energy (–51.7 kcal mol^{-1}), which excludes contributions from distortion, is 20 kcal mol^{-1} more favorable for **8b** compared to **1**. The contributions of the additional hydrogen bonding and π-stacking interactions exhibited by **8b** can be seen by decomposing the total interaction energy into individual contributions from different functional groups in the sensor (see Figure 6). This decomposition was achieved by first severing the indicated bonds. For each resulting fragment, the open valence was capped with a hydrogen atom (the position of which was optimized with the remaining atoms fixed in space) and the interaction energy computed. Based on this approximate decomposition, the ethylenediamine linker (model M3) and pyrenyl group (model M4) contribute approximately 11.1 and 13.1 kcal mol^{-1} to the total interaction energy, respectively (see Table 2), accounting for the bulk of the ~20 kcal mol^{-1} difference.

Finally, we note that Fang *et al*.[34] presented data for only one of three potential regioisomers of **8b**, which differ by the connectivity of the pyrenyl substituent to the ethylenediamine linker (*i.e*. **8b** *vs* **8a** or **8c** in Figure 2). Binding energies for these other potential pyrene-based sensors with guanine were computed to be 1.4 kcal mol^{-1} less favorable than Fang's sensor **8b** (see Appendix, Table A-3). The sensitivity of these binding energies to the regiochemistry of the pendant aryl group raises the question of how to maximally exploit π-stacking interactions in the design of a guanine sensor.

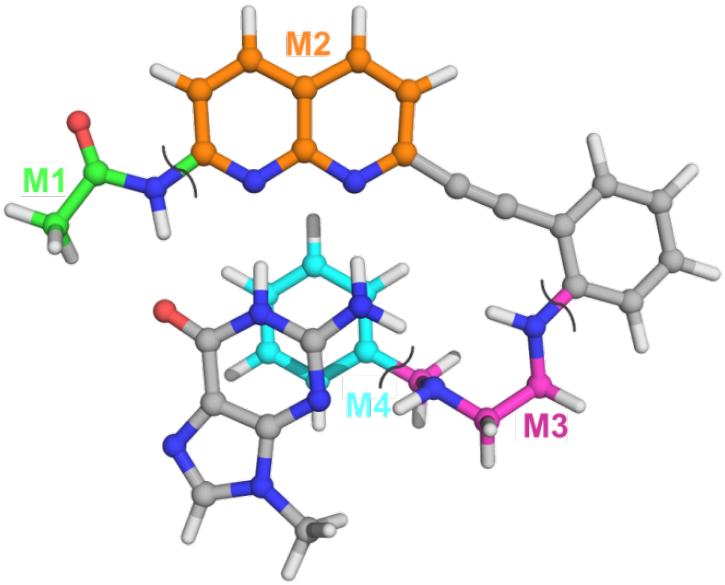

Figure 6. Truncated model regions (M1-M4) used to decompose the total interaction energies into individual pairwise interactions (shown here for sensor **4a**). For example, model M1 captures the interaction between guanine and the amide component of the sensor, while model M4 captures the π-stacking interaction between guanine and the aryl substituent. In the truncated model systems all severed bonds (denoted by black curves) were replaced with hydrogens, whose positions were then optimized. For more details, Appendix A-5.

C. Impact of Aryl Substituents on Guanine Binding

Observing the impact of the regiochemistry of the pyrenyl substituent on the predicted guanine binding energy of **8a-c** prompted the exploration of different aryl substituents. We considered the seven aryl groups depicted in Figure 2. For each arene, we included all unique regioisomers except for those requiring bonding to the bay regions of phenanthrene, triphenylene, and benzo[a]pyrene. In these excluded cases, significant steric congestion would render these systems synthetically impractical. Solution-phase binding energies for these 24 distinct sensors are plotted in Figure 7 (for data, see Appendix Table A-3). Optimized geometries of selected sensors both with and without guanine bound are shown in Figure 5. Overall, the predicted binding energies in Figure 7 span more than 5 kcal mol^{-1}, with sensor **9c** exhibiting the most favorable guanine binding energy of -28.2 kcal mol^{-1}. Even for a given arene there can be a considerable spread in guanine-binding energies. For instance, the computed binding energies for **10f** (-28.0 kcal mol^{-1}) and **10g** (-22.6 kcal mol^{-1}) differ by more than 5 kcal mol^{-1}. Clearly, optimizing guanine binding through π-stacking interactions is not as simple as incorporating a π-stacking group.

To gain a deeper understanding into the underlying trends in binding energies, we computed gas-phase binding energies as well as the corresponding distortion and interaction energies for the best sensors for each aryl substituent and all regioisomers of **10** (see Table 2). First, we note that there are slight differences in the ordering of gas-phase binding energies compared those in solution. For example, even though **7a** is predicted to bind guanine more weakly in dichloromethane than either **8b** or **9c**, the gas-phase binding energy of **7a** exceeds those of **8b** and **9c**. Similarly, although **9c** exhibits

the strongest guanine binding in solution, **10f** is predicted to have the strongest gas-phase binding energy and interaction energy. Presumably, these changes are a consequence of changes in the amount of solvent-accessible surface area in the bound and unbound geometries. For some isomers of **10**, for example, the geometry of the benzo[a]pyrene allows for close contact between the arene and the rest of the sensor in the unbound state and between the arene and the sensor plus the guanine in the bound state. For other regioisomers, on the other hand, the geometry of benzo[a]pyrene precludes close association of the arene and the naphthyridine in the unbound geometry. Regardless, the gas-phase data in Table 2 should provide a reasonable starting point for understanding the binding energies of these different potential sensors in solution.

Considering the sensors with different aryl groups, there is no correlation between the distortion energies and the gas-phase binding energies ($R^2 = 0.03$, see Table 2), while the interaction energies show a weak correlation with the total binding energies ($R^2 = 0.41$). Overall, neither of these individual components dominate the trend in binding energies. Looking more closely at the total interaction energies, we have approximately decomposed these into contributions from individual components of the sensor (see models M1-M4 in Figure 6). As seen in Table 2, the sum of these pairwise interaction energies is strongly correlated ($R^2 = 0.99$) with the total interaction energy, indicating little impact of cooperativity or errors resulting from severing covalent bonds.[153] Among these pair-wise interactions, those involving the ethylenediamine linker and the aryl group interacting with guanine are most strongly correlated with the total interaction energy ($R^2 = 0.98$ and 0.98 for M3 and M4, respectively). That is, both of these components strongly impact the interaction of guanine with the sensor. This is perhaps

unsurprising for the aryl substituent (M4), which engages in highly variable π-stacking interactions with the guanine based on the size and positioning of the aryl ring relative to guanine.

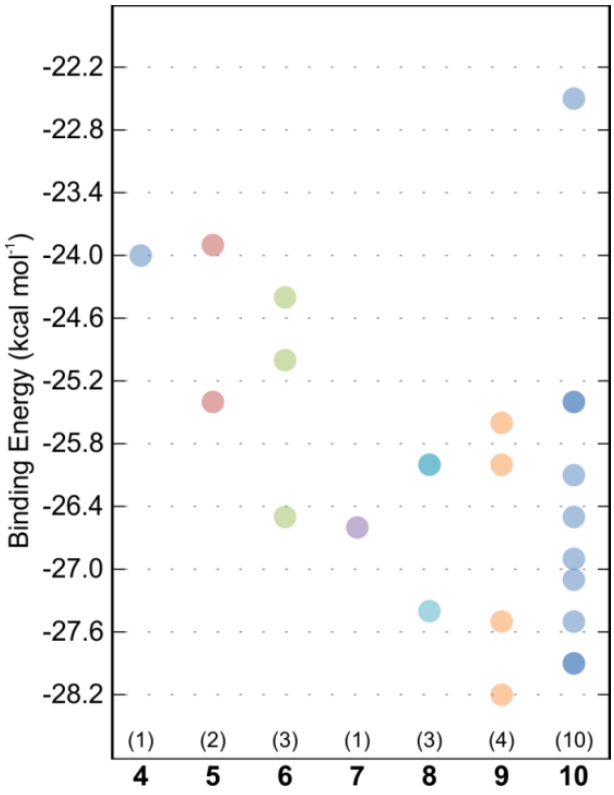

Figure 7. Distribution of predicted binding energies (kcal mol^{-1}) for different regioisomers of sensors **4**-**10** in dichloromethane. The number of distinct sensors in each column is listed in parentheses.

The varying hydrogen-bonding interactions with the ethylenediamine linker, however, appear to be secondary effects due to changes in the binding pose for the different aryl substituents. In order to identify which of these hydrogen-bonding interactions are most significant with regard to the total interaction energy, we tabulated the various N···H-N distances (see Figure 5 and Appendix A, Table A-4) as an indicator

of the strengths of these hydrogen bonds. In general, the shorter these hydrogen bond distances the more favorable the overall interaction energy between the sensor and guanine. Moreover, shorter H-bonding distances correlate with the strength of the stacking interaction between the pendant aryl group and guanine (model M4). Distance r1, which corresponds to the hydrogen bonding interaction between an amino hydrogen of the linker and the exocyclic amino nitrogen of guanine, is most strongly correlated with both the total interaction energy (R^2=0.97) and the M4 interaction energy (R^2 = 0.92). Notably, this hydrogen bonding interaction was not included in the binding model proposed by Fang *et al*. (see Fig. 3), but appears to be a vital determinant of the guanine-binding affinities of these sensors.

Both the π-stacking interactions of the different aryl substituents with guanine and the hydrogen bonding interactions between the ethylenediamine linker and guanine strongly correlate with the total interaction energy. The picture that emerges is one in which the aryl substituent is positioned to maximize π-stacking interactions with both guanine and the naphthyridine component of the sensor, which in turn impacts the geometry of the ethylenediamine linker and its hydrogen-bonding interactions with guanine. This requirement to both engage in strong π-stacking interactions while being compatible with strong hydrogen bonding interactions between guanine and the linker underlies the sensitivity of these sensors to the size, shape, and point of connectivity of the pendant aryl group. Simply put, integrating a large aryl group does not guarantee improved guanine binding. For instance, sensor **9c** features the medium-sized phenanthryl substituent, yet is predicted to bind guanine more strongly than either **8** (with pyrenyl substituent) or **10** (with a benzo[a]pyrenyl substituent).

The same analysis used to understand different aryl substituents was also performed on the different regioisomers of **10** (see Table 2). We observed qualitatively similar trends to those seen for the best regioisomers of sensors **4**-**10**; however, for **10a-j** there is a slightly stronger correlation between E_{int} and E_{bind}. The interaction energy from model M4 again correlates with the total interaction energy, but this correlation is weaker for **10a-j** compared to the best sensors of **4**-**10** ($R^2 = 0.70$ compared to 0.98). This decreased correlation can be attributed to the relatively constant π-stacking interactions between the arene and guanine; the large size of benzo[a]pyrene means that these π-stacking interactions are essentially maximized regardless of the connectivity. Nonetheless, variations in π-stacking between guanine and the aryl group in regioisomers of **10** still have some direct impact on the overall binding. For example, **10f** and **10g** have vastly different π-stacking geometries with G, as seen in Figure 5. This is reflected in the 6 kcal mol^{-1} difference in interaction energy between **10f** and **10g,** of which more than 4 kcal mol^{-1} arises from the interaction energy of the arene with guanine (model M4).

As for interactions with the ethylenediamine linker, the correlation of the M3 interaction energies with the total interaction energy is still strong (R^2=0.94). In these cases, however, the correlation between the various H-bonding distances and the interaction energy are more modest than observed for the best of sensors **4**-**10**. Regardless, there is still a tendency for the sensors exhibiting the shortest (and therefore, presumably, strongest) H-bonding interactions with the ethylenediamine linker to have the most favorable predicted total interaction energies. In this case, however, there is less importance of any individual H-bonding interaction. Instead, strong guanine binding

correlates with the linker exhibiting at least two short H-bonding contacts. For instance, sensor **10f**, which exhibits the strongest predicted binding energy among the regioisomers of **10**, actually features a slightly different network of hydrogen bonds than the other sensors; however, the corresponding distances are by far the shortest. Again, the picture that emerges is one in which π-stacking interactions between the arene and guanine (as well as the naphthyridine component of the sensor) dictate the overall geometry of the bound complex. Whether this is compatible with strong hydrogen bonds between guanine and the ethylenediamine linker then determines the overall guanine binding ability.

Table 2. Gas-phase binding energies (E_{bind}), distortion energies (E_{dist}), and interaction energies (E_{int}), in kcal/mol, for the regioisomers of **4**-**10** exhibiting the most favorable binding energy, as well as all regioisomers of **10**. Also included are pairwise interaction energies for individual sensor components (M1-M4) as well as the sum of these four pairwise interaction energies.

				E_{int} for pairwise models				
Sensor	E_{bind}	E_{dist}	E_{int}	M1	M2	M3	M4	Sum
4a	–33.2	12.3	–45.5	–6.6	–18.6	–8.1	–8.8	–42.2
5a	–34.6	12.3	–46.9	–6.5	–18.7	–8.3	–10.5	–44.0
6b	–35.9	17.3	–53.1	–6.7	–18.2	–11.3	–14.2	–50.3
7a	–39.2	8.8	–47.9	–6.8	–18.5	–9.0	–11.2	–45.5
8b	–37.4	14.3	–51.7[a]	–6.9	–18.4	–11.1	–13.1	–49.5
9c	–37.5	13.8	–51.4[a]	–6.6	–18.4	–10.8	–12.8	–48.9
10f	–40.6	13.5	–54.1[a]	–6.7	–18.2	–12.3	–14.8	–52.0
Correl. with E_{bind} (R^2)		0.03	0.41					
Correl. with E_{int} (R^2)				0.17	0.87	0.98	0.98	0.99
				E_{int} for pairwise models				
Sensor	E_{bind}	E_{dist}	E_{int}	M1	M2	M3	M4	Sum
10a	–38.0	12.0	–50.0	–7.0	–18.6	–9.1	–13.6	–48.2
10b	–35.7	13.7	–49.4	–6.8	–18.8	–8.8	–13.1	–47.4
10c	–38.6	9.9	–48.5	–6.9	–18.7	–8.4	–12.3	–46.3
10d	–35.8	10.7	–46.5	–7.0	–18.2	–7.0	–11.5	–43.7
10e	–35.6	12.4	–48.1	–6.8	–18.5	–8.5	–12.3	–46.1
10f	–40.6	13.5	–54.1	–6.7	–18.2	–12.3	–14.8	–52.0
10g	–35.3	12.9	–48.1	–7.2	–18.3	–8.8	–10.6	–44.9
10h	–38.6	10.7	–49.3	–6.6	–18.7	–9.0	–12.6	–46.9
10i	–35.6	12.4	–48.0	–6.8	–18.6	–8.2	–12.5	–46.1
10j	–40.4	10.8	–51.2	–6.7	–18.3	–11.0	–13.0	–49.0
Correl. with E_{bind} (R^2)		0.09	0.65					
Correl. with E_{int} (R^2)				0.22	0.04	0.94	0.70	0.97

[a] with counterpoise corrections, the interaction energies for **8b**, **9c**, and **10f** are –47.1, –47.2, and –48.6 kcal/mol, respectively.

IV. Summary and Concluding Remarks

Computational studies are constantly broadening our understanding of noncovalent interactions, with important implications for understanding biological systems. While chemical intuition can lead to the development of approximate binding models for complex systems, computations are often required to provide a deeper understanding of the underlying non-covalent interactions. We have presented DFT-computed structures and binding energies as well as rigorous *ab initio* benchmarks for model hydrogen bonding complexes of **1 (**2-acetamido-1,8-naphthyridine) with guanine. The data show that **1** is an excellent hydrogen bonding partner for guanine, with a predicted binding energy within 1 kcal mol^{-1} to that of its Watson-Crick counterpart cytosine. Although the three conventional hydrogen bonds in the complex of **1** with guanine are weaker in than those between cytosine and guanine, complexation of guanine with **1** benefits from an additional, non-classical hydrogen bond. These results support the use of **1** as the backbone for the development of guanine-specific sensors.[28,107-123]

Computations further show that sensor **8b**, designed by Fang *et al.*[28] and utilizing a naphthyridine backbone, exhibits more than a 10 kcal mol^{-1} enhanced binding relative to **1** due to the introduction of a π-stacking pyrenyl group and the formation of additional hydrogen bonds with the ethylenediamine linker. Moreover, among the three possible isomers of **8b**, which differ by the point of connectivity of the pyrenyl substituent to the ethylenediamine linker, **8b** provides the most favorable predicted binding energy. While the binding model presented by Fang *et al.*[28] is qualitatively correct, it neglects essential intramolecular π-stacking interactions as well as some hydrogen bonding interactions that are vital for understanding the binding affinity for guanine.

We also examined 24 sensors differing in the identity and point of connectivity of the pendant aryl group. The data revealed that both the size and regiochemistry of this aryl group impacts the guanine binding energy. For instance, of the 24 sensors considered, those with a pendant benzo[a]pyrene group are among the best and worst guanine binders. Ultimately, the data show that further improving guanine sensing is not as simple as introducing larger arene substituents; the structure and regiochemistry of this aryl substituent must be chosen such that π-stacking interactions can be maximized without disrupting vital hydrogen bonding contacts with the ethylenediamine linker. These data underscore the rich interplay of non-covalent interactions that can impact the effectiveness of selective small-molecule sensors. At the same time, they demonstrate that careful computational analyses can unravel these interactions, providing insight into the means by which more effective sensors can be designed

CHAPTER 3

IMPORTANCE OF MODEL SIZE IN QUANTUM MECHANICAL STUDIES OF DNA INTERCALATION[2]

[2] Harding, D. P; L. J. Kingsley; G. Spraggon; and S. E. Wheeler. Submitted to the Journal of Computational Chemistry.

Abstract

The convergence of DFT-computed interaction energies with increasing binding site model size was assessed. The data show that while accurate intercalator interaction energies can be derived from binding site models featuring only the flanking nucleotides for uncharged intercalators that bind parallel to the DNA base pairs, errors remain significant even when including distant nucleotides for intercalators that are charged, exhibit groove-binding tails that engage in non-covalent interactions with distant nucleotides, or that bind perpendicular to the DNA base pairs. Consequently, binding site models that include at least three adjacent nucleotides are required to consistently predict converged binding energies. The computationally inexpensive HF-3c method[57] is shown to provide reliable interaction energies and can be routinely applied to such large models.

I. Introduction

Some of the various biological roles of DNA[29] can be modulated by the non-covalent binding of small molecules. A key group of such small molecules are DNA intercalators, which insert between adjacent base pairs leading to partial unwinding and lengthening of the structure of DNA (see Figure 8). These deformed structures can hinder key biological functions, which can be exploited for therapeutic purposes. For instance, intercalators such as doxorubicin and proflavine have been used to treat acute leukemia and the parasite trypanosomiasis, respectively.[30-31] As the biological importance of intercalators has become more widely appreciated and their use as clinical drugs realized, there has been increased interest in the design of intercalator-based therapeutics.[32]

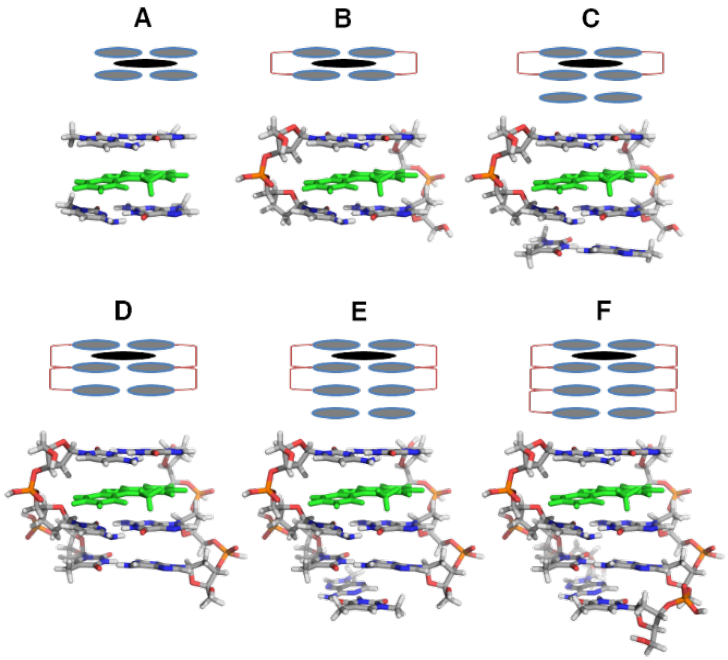

Figure 8. DNA intercalator binding site Models A-F (the intercalator is shown in green).

Despite advances in computational tools for designing small molecules that bind proteins, analogous tools for DNA-targeting small molecules are less well developed. Reasons for this are numerous both from a historical and physiological perspective. Historically, drug design has focused on proteins as drug targets, leading to the development of protein-centric computational tools and structures. From a purely physical standpoint, the flexible structure, helical framework, and high charge density of DNA coupled with numerous binding sites that are often exposed to solvent make specific DNA intercalation a difficult computational problem.[154] The unique features of DNA in comparison to proteins or peptides make it difficult to apply existing computational drug discovery tools to DNA. This is due in part to the inability of standard docking protocols and MM potentials to provide reliable intercalation binding energies.[33-35] One route to improving such tools is through parameterization based on

reliable quantum mechanical (QM) computations. However, this requires that a sufficiently reliable yet computationally tractable QM method be identified.[36]

There have been numerous QM studies of intercalators[37-55] employing a wide range of both methods and model sizes. While these studies have provided insights into DNA intercalation, there exist no guidelines for the number of bases and backbones that must be included to compute reliable intercalator interaction energies.

Herein, we show that errors due to the model size show unexpectedly slow convergence for certain classes of intercalators, requiring large binding site models in order to reliably predict converged intercalator binding energies. We also show that the minimal basis method HF-3c provides reliable binding energies yet is sufficiently computationally inexpensive to be routinely applied to large binding site models.

II. Methods

Structures of ten structurally distinct intercalators bound to 12 binding sites were selected from the PDB (see Figure 9). In nine of these structures, intercalation occurs between GC bases, which is reflective of the PDB in that the binding sites are heavily biased to intercalation at GC sites.[155] Further details about the intercalation sites and the binding stoichiometry of the selected systems can be found in Appendix B, Table B-1. Atoms not covalently bound to either DNA or the intercalator (*e.g.* solvent, counterions, *etc*.) were removed and all phosphate groups protonated.[83,156-157] For each intercalated DNA structure, six models were constructed (Model A-F in Figure 8). The largest of these consists of intercalated DNA truncated to the four base pairs surrounding the intercalators (see Model F in Figure 8). It should be noted that all intercalators

considered were bound between the two terminal bases of the helical strand, which is reflective of the binding position in the crystal structures. Hydrogen atoms were added to satiate all open valences and the positions of all hydrogen atoms were optimized at the B97-D/6-31G(d) level of theory[125-126,158] with the coordinates of the heavy atoms frozen. Some intercalators included ionizable nitrogen atoms. We considered protonated versions of these intercalators, which are denoted with a + (*e.g.* $\mathbf{3}^{+}$, Figure 10).

Interaction energies, defined as the difference in energy between the intercalated DNA and the isolated DNA and intercalator at the hydrogen-optimized crystal-structure geometry, were evaluated using rigorous *ab initio* methods, DFT methods, and the parameterized methods HF-3c and PBEh-3c. First, reliable interaction energies were determined within the focal point approach of Allen and co-workers[140-141,159] for four intercalators using the two smallest binding site models. In the focal point approach, dual extrapolations with respect to one-particle basis set and electron correlation were carried out using RI-MP2,[160] DLPNO-CCSD,[161-164] and DLPNO-CCSD(T)[161-164] with the Dunning correlation-consistent basis sets up to cc-pVQZ.[165] In particular, Hartree-Fock energies were extrapolated based on cc-pVXZ (X = D, T, Q) energies using an exponential functional form[166] while correlation energies were extrapolated separately from cc-pVXZ (X = T, Q) data following Helgaker *et al.*[167] Sponer and co-workers[168] recently demonstrated the reliability of extrapolated DLPNO-CC interaction energies on model stacked nucleotides. Incremental focal point tables are included in Appendix B, Table B2 and suggest that these interaction energies are converged to within 2.5 kcal mol^{-1} of the *ab initio* limit. DFT interaction energies were computed using the B3LYP-D3,[169-170] B97-D,[125-126] M06-2X,[171] and ωB97X-D[172] functionals in combination with the

6-31G(d),[158] 6-31+G(d),[158] def2-TZVP,[173] and cc-pVTZ[165] basis sets. The B3LYP-D3 data include three-body dispersion corrections and used the Becke-Johnson damping function.[174] Interaction energies were also computed using HF-3c and PBEh-3c.[57,175] The electrostatic component of these interaction energies was estimated by first computing the ESP due to the DNA at the positions of the intercalator atoms. The total electrostatic interaction energy was estimated as the sum of products of NPA atomic charges[176-177] for the isolated intercalator and the ESP values at the corresponding positions. All DFT computations carried out with Gaussian09[130] while Orca 4.0[178] was used to compute RI-MP2, DLPNO-CCSD, DLPNO-CCSD(T), HF-3c, and PBEh-3c energies. The DLPNO computations utilized default PNO cutoffs.

III. Results and Discussion

Ten intercalators (see Figure 9) were selected to represent typical intercalators found on the PDB. They range from planar anthracene-based structures (*e.g.* proflavine, **4**) to tetracyclic structures appended with flexible groove-binding tails (*e.g.* doxorubicin, **7**). Representative binding modes are depicted in Figure 11. The intercalated DNA crystal structures were truncated to six distinct models ranging from two to four base pairs both with and without the sugar-phosphate backbone (Models A-F, Figure 8). The impact of both level of theory and model size on computed interaction energies were quantified.

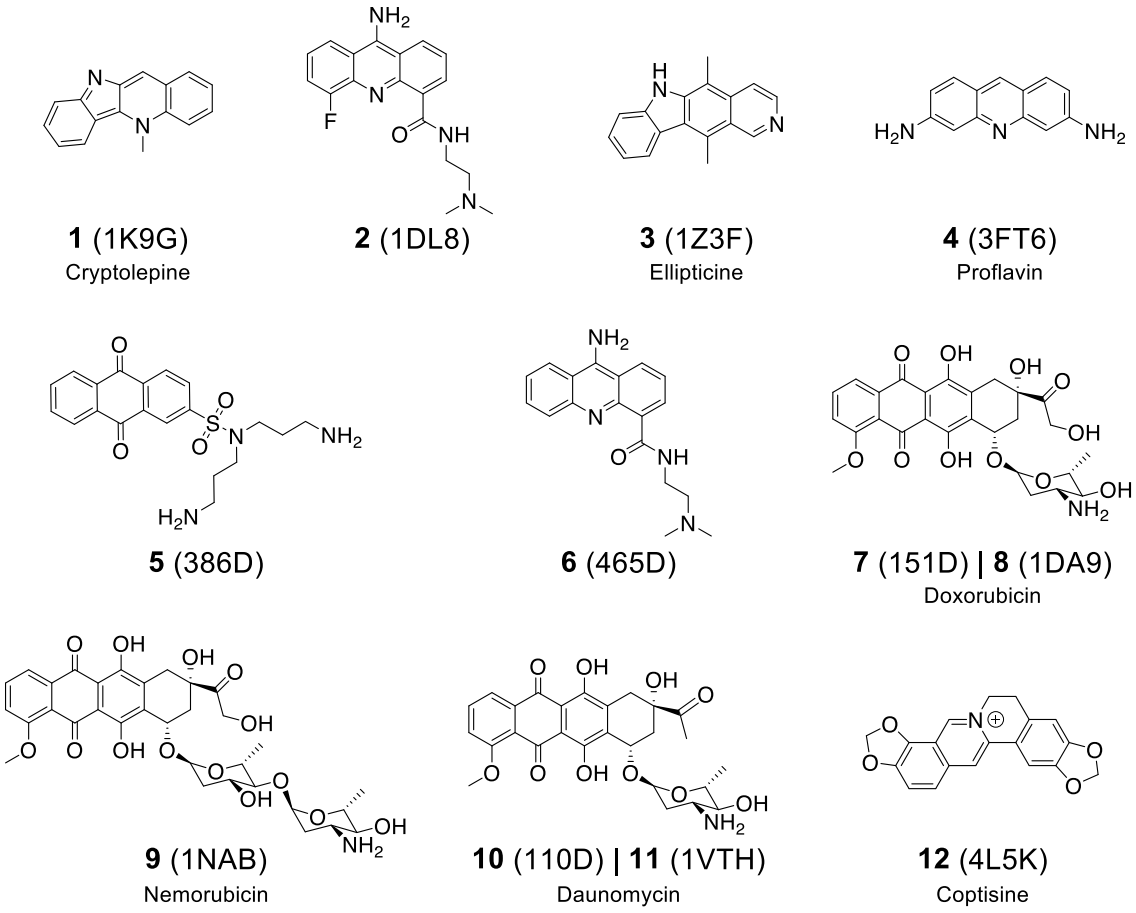

Figure 9. Intercalators studied along with corresponding PDB codes. For intercalators (**7**, **8**) and (**10**, **11**), the same intercalator is bound to two different binding pockets. Intercalators **1**-**3** were used to assess DFT methods, whereas all 12 were considered for assessing the convergence of binding energies with respect to model size.

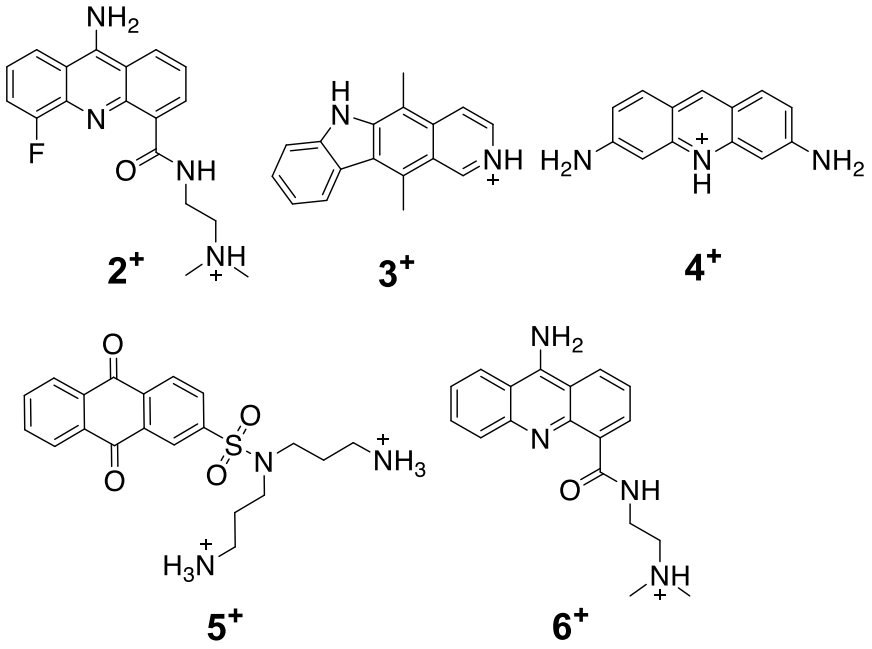

Figure 10. Protonated forms of intercalators **2**-**6**.

A. Performance of DFT Methods

Given the diversity of methods used in previous studies of DNA intercalators,[37-55] we first assess the performance of popular DFT methods in predicting intercalator interaction energies. Given the absence of experimental binding energies for the intercalated systems for which reliable crystal structures are available, we used DLPNO-CCSD(T) interaction energies extrapolated to the complete basis set (CBS) limit for intercalators **1**-**3**, as well as **3**$^+$, using the two smallest binding site models (Models A and B) to assess the performance of different DFT methods. As shown below, for these four systems Model B captures 95% of the interaction energy of the most complete model (Model F). The computed reference interaction energies are listed in Table 3.

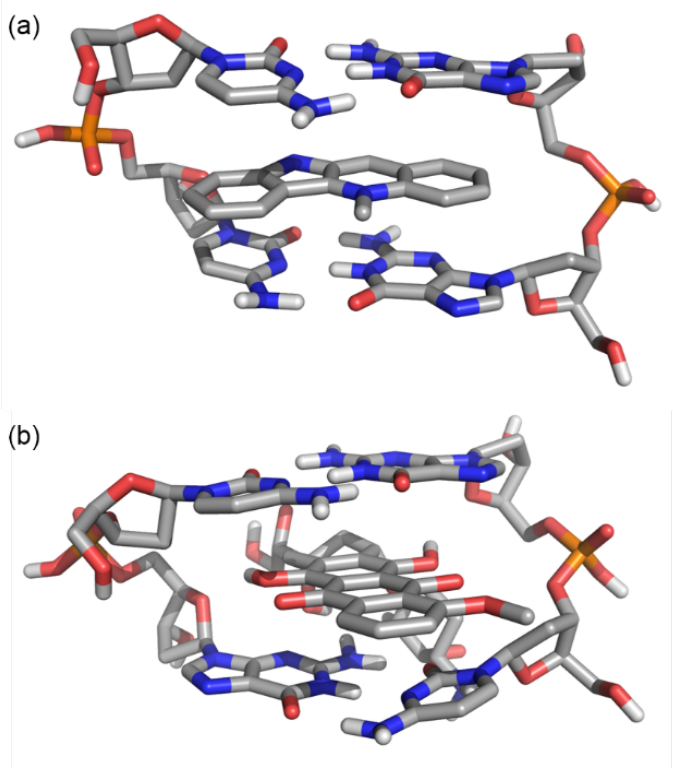

Figure 11. Representative parallel and perpendicular binding modes for (a) **1** and (b) **7**, respectively.

Sixteen combinations of four DFT functions paired with four popular basis sets were explored. Figure 12 shows the percent error of the DFT-predicted interaction energies compared to the ab initio data for 1-3 and 3+ using Model B; a table of the values can be found in Appendix B, Table B-3. Results for Model A are similar (see Appendix B Table B-4). There is considerable spread in errors in the DFT interaction energies, which can exceed 30%. This should serve as a clear warning that the poor choice of either functional or basis set for such systems can lead to significant errors in computed interaction energies. However, it should be noted that even these errors are considerably smaller than those using MP2 (see Appendix B, Table B-5), which is widely used in the literature.[38,42,48-54] Fortunately, for many functional/basis set combinations the errors are considerably smaller, sometimes below 1%.

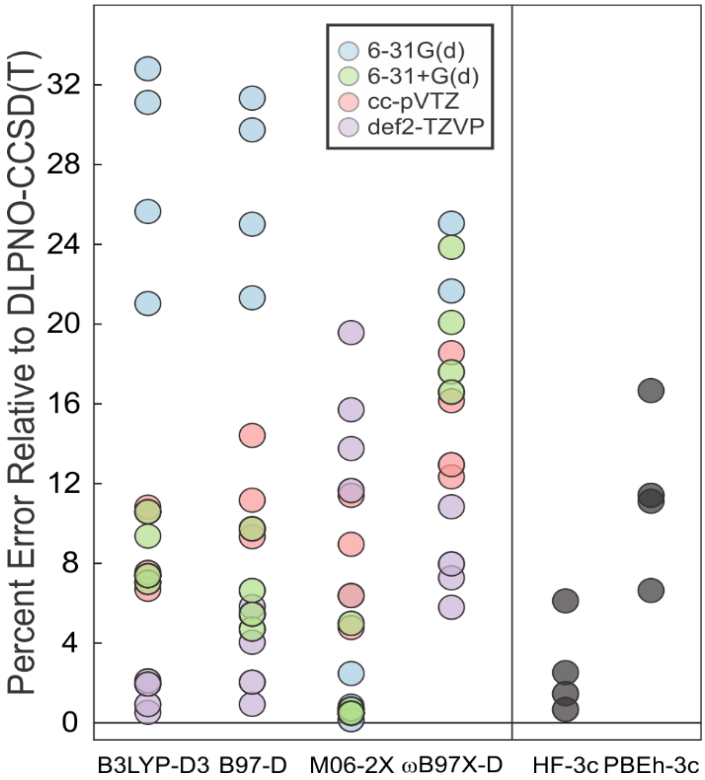

Figure 12. Percent errors in interaction energies computed with four DFT functionals paired with four popular basis sets as well as HF-3c and PBEh-3c compared to extrapolated DLPNO-CCSD(T) values for intercalators **1**-**3** and $\mathbf{3}^{+}$ using Model B. Data for Model A are available in Appendix B.

Table 3. B3LYP-D3(BJ)/def2-TZVP interaction energies (kcal mol^{-1}) for intercalators **1**-**12** using models A-F. DLPNO-CCSD(T)/CBS interaction energies are provided for selected systems in parentheses. Whether the intercalator binds parallel (par) or perpendicular (perp) to the base pair axes is also indicated.

	Binding Mode	A	B	C	D	E	F
1	par	-41.1 (-42.0)	-45.2 (-45.4)	-46.0	-45.9	-45.6	-45.9
2	par	-39.9 (-39.5)	-45.7 (-45.3)	-46.4	-46.2	-46.8	-46.6
$\mathbf{2}^{+}$	par	-72.8	-86.5	-90.4	-92.4	-93.8	-95.4
3	par	-34.4 (-34.6)	-40.2 (-39.4)	-41.2	-41.4	-41.4	-41.2
$\mathbf{3}^{+}$	par	-47.9 (-47.7)	-63.8 (-62.5)	-64.2	-65.6	-66.8	-67.9
4	par	-36.1	-45.5	-46.2	-46.7	-46.6	-46.7
$\mathbf{4}^{+}$	par	-51.8	-69.7	-71.5	-73.0	-73.6	-75.1
5	par	-41.4	-48.5	-49.0	-49.3	-49.6	-49.6
$\mathbf{5}^{+}$	par	-55.1	-71.9	-75.3	-79.1	-80.5	-83.3
6	par	-40.6	-46.5	-46.6	-47.1	-47.6	-47.8
$\mathbf{6}^{+}$	par	-77.3	-87.4	-90.9	-93.1	-94.4	-95.1
7	perp	-42.2	-61.0	-67.6	-73.9	-74.2	-76.0
8	perp	-35.8	-46.1	-50.9	-54.6	-55.6	-56.2
9	perp	-52.5	-63.8	-65.0	-69.2	-73.1	-73.3
10	perp	-51.4	-58.1	-60.2	-62.6	-63.0	-63.1
11	perp	-50.8	-40.6	-45.6	-48.7	-49.4	-53.3
12	perp	-38.5	-45.6	-48.3	-50.6	-51.4	-53.6

Overall, several trends emerge from these data, if M06-2X is excluded. First, errors for a given functional generally decrease going from 6-31G(d) to 6-31+G(d), then cc-pVTZ, and finally to def2-TZVP. The basis set dependence of the M06-2X predictions are almost exactly the opposite—errors are maximized with def2-TZVP and minimized with the smaller Pople-style basis sets. While not shown in Figure 12, we also considered B3LYP without an empirical dispersion correction. The resulting interaction energies for these dispersion-dominated systems exhibited errors as large as 140%. This provides yet another warning that B3LYP (and other functionals that do not capture dispersion interactions) should not be used to study DNA intercalation. In contrast to this, B3LYP-D3/def2-TZVP provides the most reliable interaction energies for these systems, with errors consistently below 5% (see Table 3). For instance, the largest error (1.3 kcal

mol^{-1}) occurs for **3**$^{+}$ with Model B; errors are all considerably less than 1 kcal mol^{-1} for the other systems considered. Other levels of theory that perform nearly as well include B97-D/def2-TZVP and M06-2X/6-31+G(d).

Table 4. HF-3c interaction energies (kcal mol^{-1}) using Models A-F.

	A	B	C	D	E	F
1	-40.3	-44.3	-45.2	-45.3	-45.3	-45.7
2	-35.7	-42.5	-43.8	-43.4	-44.0	-43.6
2^{+}	-70.8	-85.3	-90.3	-92.9	-94.2	-96.0
3	-32.4	-39.1	-40.2	-40.3	-40.4	-40.3
3^{+}	-44.5	-61.6	-62.0	-63.5	-64.9	-65.9
4	-34.8	-44.6	-45.4	-45.7	-45.7	-45.6
4^{+}	-49.0	-67.3	-69.4	-71.0	-71.6	-73.0
5	-38.8	-46.2	-47.4	-47.7	-48.0	-48.3
5^{+}	-49.3	-65.2	-69.5	-73.1	-74.5	-77.5
6	-37.0	-44.0	-44.4	-44.9	-45.4	-45.4
6^{+}	-76.4	-87.7	-92.5	-94.8	-96.1	-96.7
7	-37.5	-56.1	-62.8	-68.8	-69.1	-70.8
8	-32.5	-42.3	-46.6	-50.6	-51.6	-52.7
9	-47.7	-57.1	-58.2	-62.4	-66.4	-66.5
10	-45.6	-50.7	-53.6	-56.4	-56.8	-57.2
11	-44.9	-30.4	-34.6	-37.2	-38.1	-42.0
12	-34.6	-42.0	-45.0	-47.9	-48.7	-50.6

We also considered two parameterized minimal-basis set methods designed to provide reliable interaction energies at a much lower computational cost, HF-3c and PBEh-3c.[57,175] Data for HF-3c are provided in Table 4 (see Appendix B, Table B-6 for PBEh-3c results). Percent errors in interaction energies, relative to the DLPNO-CCSD(T) data, are plotted in Figure 12 for these four systems using Model B. While errors for PBEh-3c range from 7 to 17%, HF-3c provides interaction energy errors for **1**-**3** and **3**$^{+}$ consistently below 6%. This is well below the errors observed for many of the DFT/basis set combinations.

B. Convergence with Model Size

Computational expediency necessitates the use of a truncated models (*i.e.* using a small portion of the DNA) in applications of QM methods to DNA intercalation. Unfortunately, there has been no study that has identified the minimal model required to reliably capture interactions between an intercalator and DNA. To address this, we computed the interaction energy of each intercalator using six models of increasing size (Models A-F; see Figure 8). The models, which use coordinates from available crystal structures,[179-189] range from the inclusion of only two base pairs (Model A) to duplexes of tetra nucleotides (Model F). Interaction energies using these six models were computed for the 12 non-protonated intercalators for all 16 levels of DFT. Given the small errors observed for B3LYP-D3/def2-TZVP above, we focus on these data below; errors in predicted interaction energies for the 15 other levels of DFT can be found in Appendix B, Table B-6 (the trends discussed below apply across all DFT methods considered). B3LYP-D3/def2-TZVP computed interaction energies for all intercalators considered based on Models A-F are listed in Table 3. Overall, the gas-phase interaction energies of the non-protonated intercalators with Model F range from –41.2 kcal mol^{-1} to –76.0 kcal mol^{-1}.

Percent absolute errors in predicted interaction energies, compared to Model F, are plotted in Figure 13a as a function of model size. Excluding the obvious outlier of **11** with Model A, which can be attributed to a fortuitous cancelation of sizeable errors, these interaction energies show monotonic convergence with increasing model size. Notably, errors for the smallest model size (Model A, which features only the adjacent base pairs) are large, ranging from 10 to nearly 50% of the total interaction energy. Inclusion of the

sugar-phosphate backbone (*i.e.* Model B) leads to drastically reduced errors, which are below 25% for all 12 systems. For some systems (**1**-**6**), errors using Model B drop by nearly an order of magnitude compared to Model A, to well below 5% (see Figure 13b). This is consistent with previous work[42-43] showing that the backbone contributes significantly to the total interaction energy. However, for five of the systems (**7**-**12**), errors in predicted interaction energies remain above 10% with Model B (corresponding to errors in interaction energies of 5-8 kcal mol^{-1}) and show a very slow convergence with model size (see Figure 13c). Even with Model D, which includes three base pairs and the associated sugar-phosphate backbones, errors exceed 5% for three of these systems (**9**, **11**, and **12**).

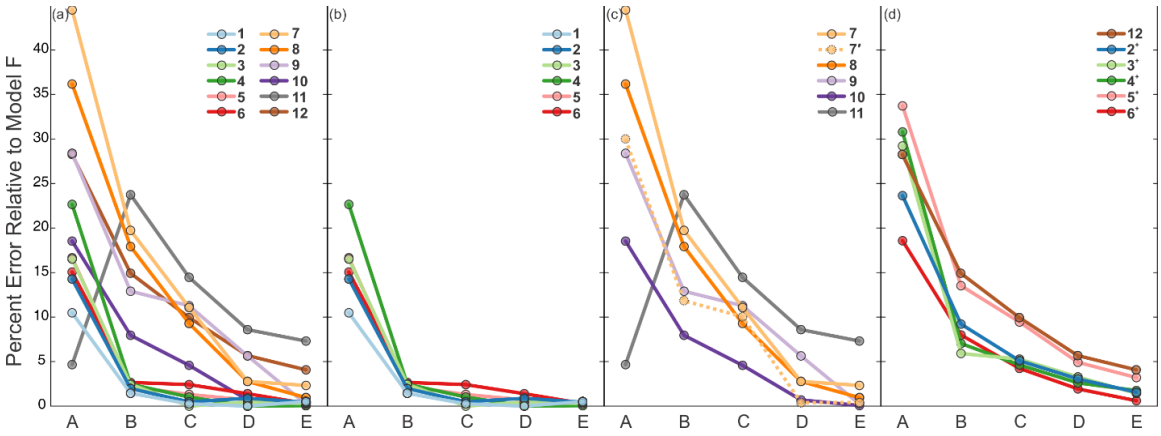

Figure 13. Percent absolute error of Model A-E, relative to Model F, at the B3LYP-D3/def2-TZVP level of theory for (a) all unprotonated intercalators (**1**-**12**), (b) intercalators that exhibit rapid convergence with respect to model size (**1**-**6**), (c) neutral intercalators that show slower convergence with respect to model size (**7**-**11**), and (d) charged intercalators (**12**, $\mathbf{2}^{+}$-$\mathbf{6}^{+}$).

Errors for **10** and **11** are instructive because they involve the same intercalator bound to two different sites. While errors for **10** drop below 5% after Model C, **11** shows much slower convergence with respect to the inclusion of distant nucleotides. Indeed, even for Model E the errors for **11** are in excess of 5%. This difference in behavior can be attributed to the position of the pendant sugar group on this intercalator in the two binding sites. In **11**, this group is engaged in several H-bonding interactions with a distant sugar-phosphate backbone, whereas it interacts with more proximal groups in the case of **10**. The result is that inclusion of four nucleotide pairs plus all sugar-phosphate backbones (*i.e.* Model F) is mandatory to capture the interaction energy in **11**; for **10**, smaller models suffice.

From a broader perspective, there are no obvious structural features delineating intercalators that exhibit slow vs fast convergence with model size. For instance, errors for **1**-**6** are all below 5% for Model B, but these intercalators have drastically different molecular structures. In particular, the intercalators in **2**, **5**, and **6** include groove-binding tails, yet the interaction energy is well-converged even with Model B. The main common feature of these six systems is that they all bind nearly parallel to the base pairs (*e.g.* Figure 4a). Moreover, for the intercalators with groove-binding tails, these tails mainly interact with the adjacent nucleobases, rendering Model B sufficient for capturing all direct non-covalent interactions.

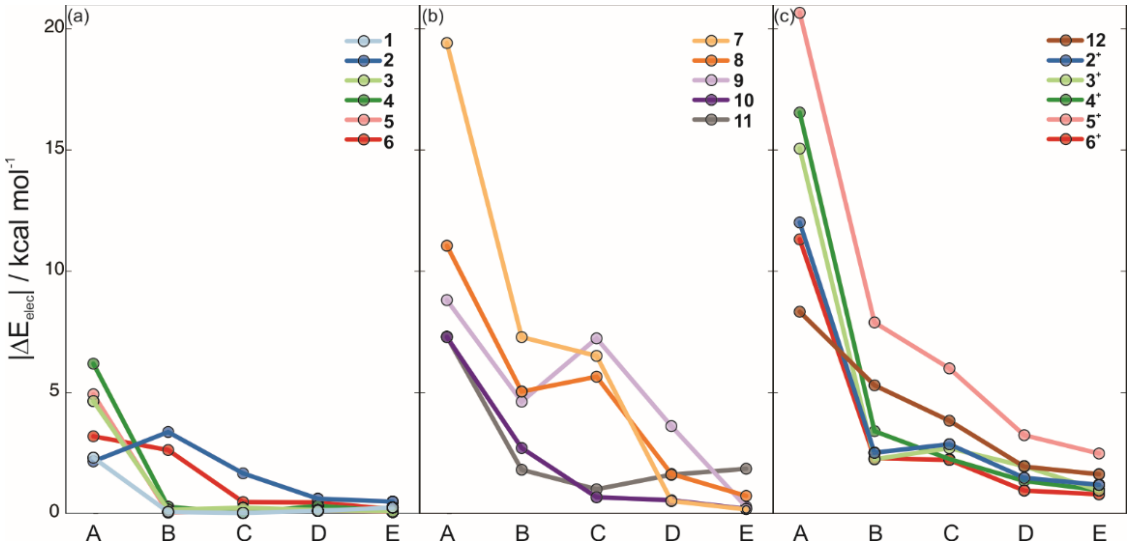

Figure 14. Absolute errors in the electrostatic component of the total interaction energy, relative to Model F, for a) **1**-**6**, b) **7**-**11**, and c) **12** and **2**$^{+}$-**6**$^{+}$

Systems **7**-**11**, on the other hand, all include groove-binding tails and all bind perpendicular to the base pairs (*e.g.* Figure 11b). To differentiate between these two effects, we considered **7′**, in which the sugar group of **7** was removed and replaced with a hydrogen. While removal of this pendant group results in a decrease in percent error (see Figure 13c), compared to **7**, the convergence with model size for **7′** still resembles the other perpendicular intercalators. This, combined with the comparison of **10** and **11**, suggests that the slow convergence with respect to model size for **7**-**11** is due to both the groove-binding tails and the perpendicular binding modes. The former effect is most severe when the groove-binding tail interacts with distant nucleotides.

Electrostatic Contributions to Errors

To further characterize the source of this slow convergence for selected intercalators, we approximated the electrostatic contribution to the total interaction

energies (see Computational Methods for details). The absolute differences in the electrostatic contribution ($|\Delta E_{elec}|$), compared to Model F, are plotted in Figure 14. Errors in the electrostatic component mirror those observed for the total interaction energy—they converge rapidly for **1**-**6** (see Figure 14a) but much more slowly for **7**-**11** (see Figure 14b). The large $|\Delta E_{elec}|$ values for Model A are unsurprising given Sherrill's demonstration[43] of the large electrostatic interaction between intercalators and the backbone, which is absent in Model A. For **1**-**6**, these errors are mostly eliminated by Model B. This rapid convergence of $|\Delta E_{elec}|$ for the parallel-binding intercalators can be attributed to the fact that they are buried between flanking nucleobases, so through-space electrostatic interactions with distant nucleotides are be screened by the intervening electron density.

For **7**-**11**, on the other hand, significant $|\Delta E_{elec}|$ values remain even for the larger models. This can be explained by the presence of unimpeded through-space electrostatic interactions with distant backbone and nucleobase atoms for the systems that bind perpendicular to the base pairs. Thus, the slow convergence of QM interaction energies for the perpendicular intercalators is partly an artifact of the use of gas-phase computations. Consequently, solution-phase computations that include counterions for systems such as **7**-**11** could converge more quickly with respect to model size, because the intervening solvent and ions will provide some dielectric screening.

Ranking Intercalator Binding

A key potential use of QM for studying intercalators is ranking intercalators by their interaction strength. To this end, we assessed the reliability of rankings from each

model size compared to Model F (see Figure 15). Rankings from Model A are qualitatively incorrect. For instance, using Model F, **1** is the 11th weakest binder, whereas Model A predicts it is ranked 6th. Indeed, only **10** and **12** are ranked correctly using Model A. While the rank order from Model B is considerably better, mirroring the drastic drop in percent error in interaction energies discussed above, still only **3** and **12** are ranked correctly. By Model C, the ranking is qualitatively correct, except for one outlier (**11**) and the ranking is correct apart from small swaps for Models D and E. As such, Model C should be considered a minimum model for reliably ranking intercalator binding, with Model D providing more robust results.

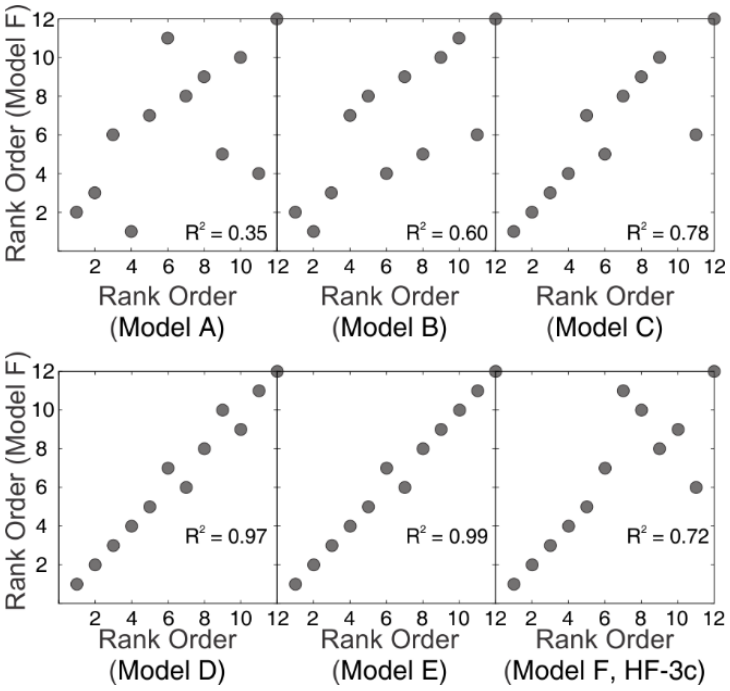

Figure 15. Rank-order correlation for intercalators 1-12 based on Models A-E using B3LYP-D3/def2-TZVP and Model F using HF-3c compared to Model F using B3LYP-D3/def2-TZVP.

Charged vs Neutral Intercalators

Of the intercalators considered, only **12** carries a permanent positive charge. This system, which lacks a groove-binding tail yet binds perpendicularly, exhibits similarly slow convergence with respect to model size as seen for intercalators **7**-**11**. To probe the impact of intercalator charge, we considered protonated versions of **2**-**6** (see Figure 10). Interaction energies for $\mathbf{2}^{+}$-$\mathbf{6}^{+}$ with Models A-F were computed with all 16 DFT methods. B3LYP-D3/def2TZVP results are listed in Table 1 and plotted in Figure 13d; the remaining results can be found in Appendix B, Table B-6. As expected, there is a sharp increase in interaction energies upon protonation of **2**-**6** across all models (see Table 3). Clearly, the protonation state will have a notable effect and intercalation energies computed using the incorrect protonation state will be qualitatively incorrect. This increase in interaction energy is accompanied by an increased percent error across all models and a much slower convergence, as seen for **12** (see Figure 13d).

As done for the neutral intercalators, we also estimated the electrostatic component of the total interaction energies for **12** and $\mathbf{2}^{+}$-$\mathbf{6}^{+}$. $|\Delta E_{elec}|$ is considerably larger for the charged vs uncharged intercalators. This is most severe for Model A, which lacks the highly polar sugar-phosphate backbone. As seen for intercalators **7**-**11**, $|\Delta E_{elec}|$ declines for Model B, but remains non-negligible for the larger models. Thus, the large errors for the charged intercalators can also be attributed to long-range electrostatic interactions with distant nucleotides.

Performance of HF-3c

Given that large binding site models are required to provide converged intercalator binding energies, combined with the accuracy of HF-3c demonstrated for

selected intercalators using Models A and B, we assessed the performance of this computationally inexpensive method for all models sizes compared to B3LYP-D3/def2-TZVP data. First, HF-3c interaction energies (Table 4) show the same convergence with respect to model size as B3LYP-D3 and the other DFT methods.

Considering all 12 systems with all six models, there is a strong correlation between the HF-3c results and those from B3LYP-D3 ($R^2 = 0.97$, see Figure 16), with a mean absolute deviation (MAD) and mean sign deviation (MSD) of 3.4 and –3.2 kcal mol^{-1}, respectively. That is, HF-3c provides interaction energies for these systems that are systematically slightly weaker than those from B3LYP-D3/def2-TZVP, but at a significantly reduced computational cost. While differences between HF-3c and B3LYP-D3 are as large as 11 kcal mol^{-1} for some systems, for other intercalators the differences are consistently below 1 kcal mol^{-1}. Overall, the largest deviations between the HF-3c and B3LYP-D3 data occur for the perpendicular binding intercalators (**7**-**12**) as well as the protonated intercalator $\mathbf{5}^+$ (see Figure 16); for the other systems the deviations are almost all below 3 kcal mol^{-1}.

With regard to ranking intercalator binding, HF-3c provides correct rankings for the five strongest binding intercalators using Model F (compared to B3LYP-D3) and also correctly predicts the most strongly bound intercalator (see Figure 15). Of the six intercalators that are ranked incorrectly, the spread of energies (4 kcal mol^{-1}) is commensurate with the average error of HF-3c, so its failure to rank these correctly is unsurprising. Overall, these data suggest that HF-3c constitutes a practical approach for studying intercalators that will allow for routine computations on large binding site models.

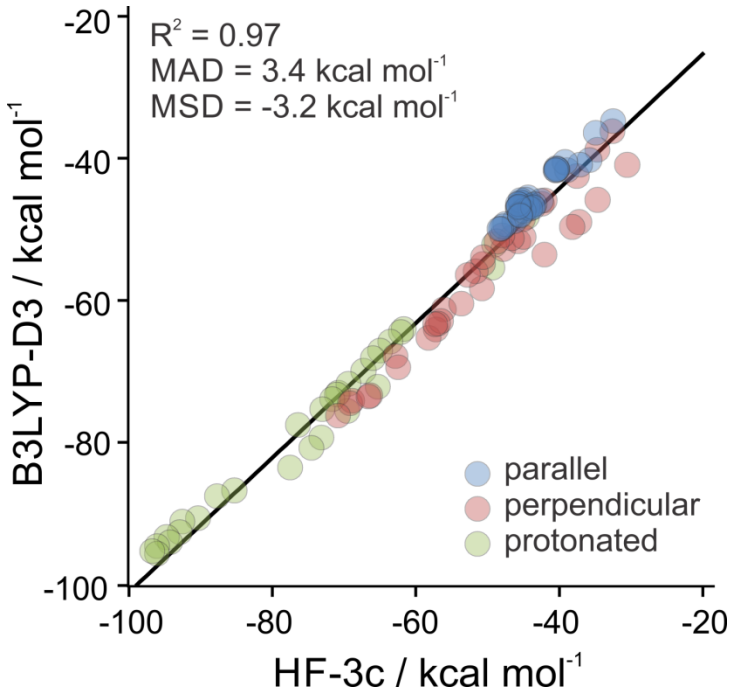

Figure 16. Correlation of B3LYP-D3/def2-TZVP interaction energies with data from HF-3c for all intercalators across all model sizes.

IV. Conclusions

Although quantum mechanical methods can provide reliable interactions for DNA intercalators, this requires the judicious choice of both level of theory and model size. Based on comparisons to reliable *ab initio* interaction energies for selected systems and small model sizes, we found that B3LYP-D3(BJ)/def2-TZVP, B97-D/def2-TZVP, and M06-2X/6-31+G(d) provide the most reliable interaction energies for DNA intercalation. B3LYP and other functionals that fail to capture dispersion interactions, as well as MP2, provide qualitatively incorrect interaction energies and should be not be applied to intercalated systems.

As for model size, recommendations depend on several factors. For neutral intercalators that bind parallel to the base-pairs and lack groove-binding tails that interact with distant nucleotides, errors in predicted interaction energies are already converged upon inclusion of the flanking base-pairs and sugar-phosphate backbone (Model B in Figure 8). However, for intercalators that carry a formal charge, bind perpendicular to the base-pairs, or include groove-binding tails that engage in non-covalent interactions with distant nucleotides, significantly larger models are required. For instance, for some intercalators errors approaching 10% can persist even for Model E.

Given the slow convergence of intercalator binding energies in some cases, it is particularly important to identify affordable computational approaches for such systems. To this end, we showed that HF-3c provides interaction energies that are in close agreement with the best tested DFT method across all model sizes, providing a pragmatic method that can be routinely applied to larger binding site models. Our hope is that these recommendations can help guide future applications of QM methods to intercalation, which in turn can be used to refine computational tools for designing nucleotide targeting small molecules.

CHAPTER 4

INTRINSIC STACKING INTERACTIONS OF NATURAL AND ARTIFICIAL NUCLEOBASES[3]

[3] Harding, D. P; L. J. Kingsley; G. Spraggon; and S. E. Wheeler. Submitted to the J. Am. Chem. Soc.

Abstract

The intrinsic gas-phase stacking energies of natural and artificial nucleobases were explored using density functional theory (DFT) and correlated *ab initio* methods. Ranking the stacking strength of natural nucleobase dimers revealed a preference in binding partner similar to that seen from experiments, namely G > C ≥ A > T > U. Decomposition of the energy components using symmetry-adapted perturbation theory (SAPT) revealed these dispersion dominated interactions are modulated by electrostatics. Artificial nucleobases showed a similar stacking preference for natural nucleobases and were also modulated by electrostatic interactions. Using molecular descriptors based on computed electrostatic potentials (ESPs), a robust predictive multivariate model was developed that quantitively predicts the maximum stacking interaction between natural and a wide range of artificial nucleobases. Further analysis of the descriptors in this model show that the superior stacking ability of guanine and cytosine stem from the clustering of amide and imino functional groups, which lead to widely varying values of the ESP over the face of the nucleobase.

I. Introduction

Both the natural biological functions of nucleobases and their utility in artificial materials, sensors, and supramolecular architectures are modulated by the ability of these heterocycles to engage in stacking interactions[1-6] (the roughly parallel face-to-face interaction of planar molecules). For example, given the chemical uniformity of the backbone and the existence of only two naturally occurring base-pairs, nucleobase stacking interactions can explain sequence dependent shapes, stability, and dynamics[4] of

both RNA and DNA.[93] In nanotechnology uses of DNA, stacking interactions can provide enhanced thermal stability and increased structural diversity, such as that seen in DNA origami motifs.[190-191] At the same time, stacking interactions underpin efforts to expand the genetic alphabet, especially in cases of non-hydrogen bonded artificial base pairs.[192] Given this central role of stacking interactions of nucleobases, it is unsurprising that considerable effort has been expended over the last half-century to understand these stacking interactions. Despite these efforts, our understanding of the factors that impact nucleobase stacking remains incomplete.

Several conceptual models of nucleobase stacking in aqueous environments have been developed based on experimental data. In the 1990s, Gellman and co-workers[58-59] argued, based on NMR data for model dimers, that nucleobase stacking is driven by the alignment of complementary local charges (*i.e.* local dipole moments) rather than solvent or dispersion interactions. In 2000, Kool and co-workers[60] presented a comprehensive study of the factors that impact the stacking of both natural and artificial nucleobases in water, such as hydrophobicity, polarizability, excluded solvent surface area, and dipole moment. These properties were correlated to the helical stability of DNA modified with a single dangling nucleobase. They found that solvation driven hydrophobic effects and solvent excluded surface-area best correlate with their data. However, this correlation is much weaker when considering only the natural nucleobases. Furthermore, the trend in natural nucleobase stacking strength identified in these experiments, A > G > T = C, is contrary to observations seen in other studies of the strength of nucleobase stacking,[61-64] which consistently show that G > C $\geq$ A > T.[61-62,64] This latter trend is also consistent

with experimentally derived thermodynamic parameters for stacked nucleobases, which are used to predict DNA thermal denaturation for a wide variety of techniques.[61,65-66]

In more recent years, there have also been a multitude of computational studies of nucleobase stacking in the gas phase,[67-96] including many aimed at uncovering the nature of these interactions. Chief among these is the review by Šponer *et al.*,[93] in which they discuss the intrinsic nature of nucleobase stacking as being governed by a balance of dispersion and short-range repulsion interactions. These effects boil down primarily to the size of the system, because the electrostatic effects arising from dipole-dipole interactions are expected to be counterbalanced by solvation effects. While these studies[70-80] have provided many insights into the relative contributions of electrostatic, dispersion, *etc*. interactions to nucleobase stacking, they did not provide a physical explanation for these trends. At the same time, there have been recent advances in our understanding of the impact of heteroatoms and substituents on stacking interactions based on computational work.[79,193-213] For instance, Bootsma *et al*.[99] recently developed a predictive model of stacking interactions between heterocycles the aromatic amino acids Phe, Tyr, and Trp based on heterocycle descriptors[97] derived from the ESP of the heterocycle.

Our lack of a comprehensive understanding of the factors that control nucleobase stacking has widespread implications. For example, the design of biomimetic and fluorescent nucleobases hinges on our ability to design arenes that engage in stacking interactions that are commensurate with those of the natural nucleobases. Below, we develop both a quantitative predictive model and qualitative conceptual model of intrinsic gas-phase stacking interactions of natural and artificial nucleobases. These models can

form a foundation upon which an understanding of solution-phase stacking of nucleobases in realistic biological contexts can be developed.

II. Results and Discussion

Below, we consider idealized stacked dimers of natural and artificial nucleobases[214-216] (see Figure 17). To provide a broad understanding of the stacking ability of different nucleobases, we systematically searched for all low-lying stacked energy minima constrained such that the heavy-atoms of the nucleobases were in parallel places. Few of these geometries are compatible with, for example, the structure of B-DNA; instead, they provide insight into the *intrinsic* stacking abilities of different nucleobases (under the constraint of parallel stacking). This enables an assessment of the stacking of nucleobases in different contexts and allows us to develop quantitative and qualitative models of the factors that control the maximum stacking interaction between a pair of nucleobases irrespective of constraints placed on these interactions within particular macromolecules.

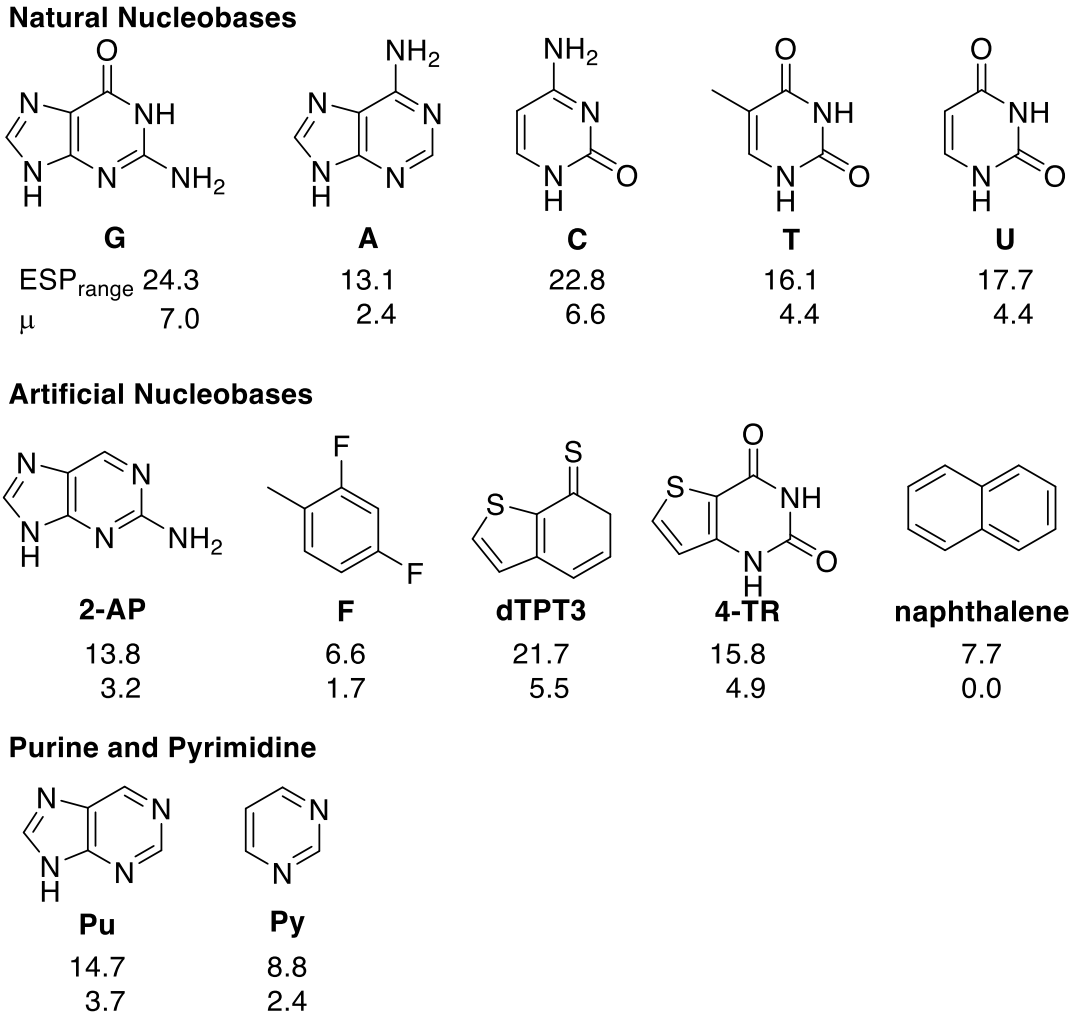

Figure 17. Structures of natural and artificial nucleobases as well as pyrimidine and purine along with computed ESP_{range} values (in kcal mol^{-1}) and dipole moments (in Debye).

A. Trends in Nucleobase Stacking

First, we considered all unique stacked dimers of the natural nucleobases (NNs). Accurate gas-phase interaction energies for the most strongly interacting stacked dimer for each pair of nucleobases are listed in Table 5. These nucleobase stacking interactions span 10 kcal mol^{-1}, from –19.7 kcal mol^{-1} for the G|G dimer to –8.8 kcal mol^{-1} for U|U.

Table 5. DLPNO-CCSD(T) interaction energies $\left(E_{int}^{calc}\right)$ as well as predicted interaction energies $\left(E_{int}^{pred}\right)$ from equation 1 and equation 2, in kcal mol^{-1} for the global minimum energy dimers.

		E_{int}^{calc}	E_{int}^{pred}				E_{int}^{calc}	E_{int}^{pred}	
			Eq 1	Eq 2				Eq 1	Eq 2
G	**G**	-19.7	-20.4	-19.9	**2-AP**	**G**	-16.1	-15.7	-15.3
G	**C**	-15.9	-15.6	-15.9	**2-AP**	**C**	-13.7	-12.2	-12.6
G	**T**	-15.6	-15.2	-15.3	**2-AP**	**T**	-12.1	-11.6	-11.7
G	**U**	-15.4	-14.4	-14.3	**2-AP**	**A**	-11.7	-11.6	-10.7
G	**A**	-14.7	-15.5	-14.5	**2-AP**	**U**	-11.5	-11.1	-11.1
C	**A**	-13.6	-12.1	-12.0	**F**	**G**	-11.4	-12.6	-13.1
C	**T**	-12.7	-11.9	-12.6	**F**	**A**	-9.6	-9.1	-9.1
C	**C**	-12.3	-12.2	-13.0	**F**	**C**	-9.0	-10.0	-11.0
C	**U**	-12.1	-11.3	-11.9	**F**	**T**	-8.5	-9.3	-10.1
A	**T**	-12.0	-11.4	-11.1	**F**	**U**	-8.0	-9.0	-9.7
A	**A**	-11.6	-11.4	-10.0	**dTPT3**	**G**	-17.7	-18.1	-17.4
A	**U**	-11.4	-11.0	-10.6	**dTPT3**	**A**	-14.4	-13.7	-12.6
T	**T**	-9.8	-11.4	-11.8	**dTPT3**	**C**	-13.8	-14.0	-14.1
T	**U**	-9.4	-10.9	-11.3	**dTPT3**	**T**	-13.3	-13.5	-13.4
U	**U**	-8.8	-10.4	-10.7	**dTPT3**	**U**	-13.1	-12.8	-12.6
Pu	**G**	-14.5	-14.8	-14.7	**4-TR**	**G**	-16.5	-17.5	-17.9
Pu	**C**	-12.3	-11.6	-12.2	**4-TR**	**C**	-13.2	-13.5	-14.5
Pu	**A**	-11.0	-11.1	-10.6	**4-TR**	**A**	-13.2	-12.9	-12.7
Pu	**T**	-10.4	-11.1	-11.4	**4-TR**	**T**	-11.5	-12.9	-13.6
Pu	**U**	-9.7	-10.6	-10.9	**4-TR**	**U**	-10.9	-12.3	-12.9
Py	**G**	-9.6	-10.2	-11.1	**Naph.**	**G**	-11.6	-13.9	-12.3
Py	**A**	-8.6	-7.8	-8.4	**Naph.**	**A**	-10.6	-9.9	-8.0
Py	**C**	-7.5	-8.4	-9.6	**Naph.**	**T**	-10.3	-10.1	-9.3
Py	**T**	-6.9	-7.9	-9.0	**Naph.**	**U**	-10.0	-9.8	-9.0
Py	**U**	-6.6	-7.7	-8.7	**Naph.**	**C**	-9.3	-10.9	-10.4

There is a clear trend in stacking interactions among the natural nucleobases, with stacking affinities following the trend G > C ≥ A > T > U. These are in general agreement with many experiments.[61-64] In order to identify the factors that underpin these trends, we performed symmetry-adapted perturbation theory (SAPT) computations on the most favorable stacked dimers.[217-220] SAPT enables the decomposition of interaction

energies into electrostatic, induction, exchange-repulsion, and dispersion contributions, which can then be used to unravel the effects that drive trends in non-covalent interactions.[221] First, the total SAPT interaction energies are strongly correlated with the DLPNO-CCSD(T) data, supporting the use of SAPT to unravel the dominant source of stacking interaction trends. These SAPT data indicate that while the average dispersion contribution to these interactions is comparable to the average electrostatic contribution, differences in the electrostatic component are the primary determinant of the strength of a given nucleobase stacking interaction (see Appendix C, Table C-1 and Figure C-1). In other words, without the electrostatic component, the stacking interactions of the natural bases are nearly constant at -1.6 ± 1.1 kcal mol^{-1}.

It is instructive to compare the stacking of the natural nucleobases with the those of the parent unsubstituted heterocycles, purine (Pu) and pyrimidine (Py). First, we consider the purine-derived nucleobases guanine and adenine. The intrinsic stacking interactions of guanine are systematically 3-5 kcal mol^{-1} stronger than those of purine. That is, the G|X dimers stack ~30% more strongly than the corresponding Pu|X dimers (X = G, C, T, U, and A). This can be contrasted with the A|X dimers, which stack on average only ~10% more strongly than the corresponding Pu|X dimers. Apparently, the substitution pattern in guanine leads to systematically much stronger stacking interactions, whereas the substituents on adenine have a much more modest effect on stacking. Turning to the pyrimidine-derived bases (cytosine, thymine, and uracil), the stacking interactions of the C|X dimers are ~40% stronger than those of the corresponding Py|X dimers. For thymine and uracil, stacking interactions are on average 33% and 30% stronger than for pyrimidine, respectively. Thus, the substitution pattern of

all three of these pyrimidine-derived bases drastically enhance stacking interactions compared to the parent heterocycle, with the effect strongest for cytosine. The SAPT data confirm that these enhanced interactions are electrostatic in nature.

Popular views of stacking interactions indicate that their strength is maximized for arenes with complementary electrostatic properties. In other words, 'electron-poor' arenes stack best with 'electron-rich' systems. This implies that a heterocycle that stacks best with one system might not stack best with others, seemingly in conflict with the above analysis showing that G stacks more strongly than the other natural nucleobases regardless of the nature of its stacking partner. To further probe this relationship, we also considered stacked dimers of the natural nucleobases with naphthalene (Naph.) and four artificial nucleobases (ANs) (see Figure 17). ANs are increasingly employed in a number of contexts, including as fluorescent indicators[214] to monitor nucleic acid interactions or as the genetic basis of semi-synthetic organisms.[216] The particular ANs we considered are 2-AP (a fluorescent isomer of adenine),[214] F (a non-hydrogen-bonding isostere of thymine),[215] dTPT2 (a semi-synthetic nucleobase),[216] and 4-TR (another fluorescent nucleobase).[214] Interaction energies for the most strongly interacting dimer of these systems with the five NAs are listed in Table 5.

Overall, naphthalene and the non-natural nucleobases exhibit the same trend as observed for the stacked dimers of natural nucleobases. Small deviations in the ordering occur when stacking interactions are very close in energy and are most prevalent in the case of naphthalene. Regardless, these data indicate that the ability of a given nucleobase to engage in strong stacking interactions is largely independent of the nature of the stacking partner, in accord with models of stacking interactions from Wheeler and co-

workers.[199,222] Furthermore, SAPT analysis again indicates that while dispersion interactions are the largest contribution to binding, the trend in stacking interactions is determined by the electrostatic part (see Appendix C, Table C-1 and Figure C-2).

B. Stacking Interactions in Biological Contexts

Armed with values for the maximum possible parallel stacking interactions between natural nucleobases, we can now assess the strength of stacking interactions between nucleobases in particular biological contexts. For instance, Spôner and co-workers[223] provided reliable optimized geometries for stacked nucleobase dimers in B-DNA. These geometries differ markedly from the optimal stacked dimer for each base pair, resulting in stacking interactions in B-DNA that are a fraction of the corresponding intrinsic values. For instance, Figure 18 shows the geometry of the global minimum stacked G|G dimer compared to that found in B-DNA. The geometry in B-DNA features one of the guanines flipped 180° relative to the global minimum structure, resulting in a ~75% reduction in the computed interaction energy from –19.7 kcal mol^{-1} to only –4.6 kcal mol^{-1}. The relatively weak G|G stacking observed in B-DNA can be contrasted with that in G-quadruplexes, which are geometrically much closer to the global minimum energy stacked structure (see Figure 18). This is reflected in the computed interaction energies in the G-quadruplex geometry, which is 65% of the maximum possible stacking interaction at –12.8 kcal mol^{-1}.

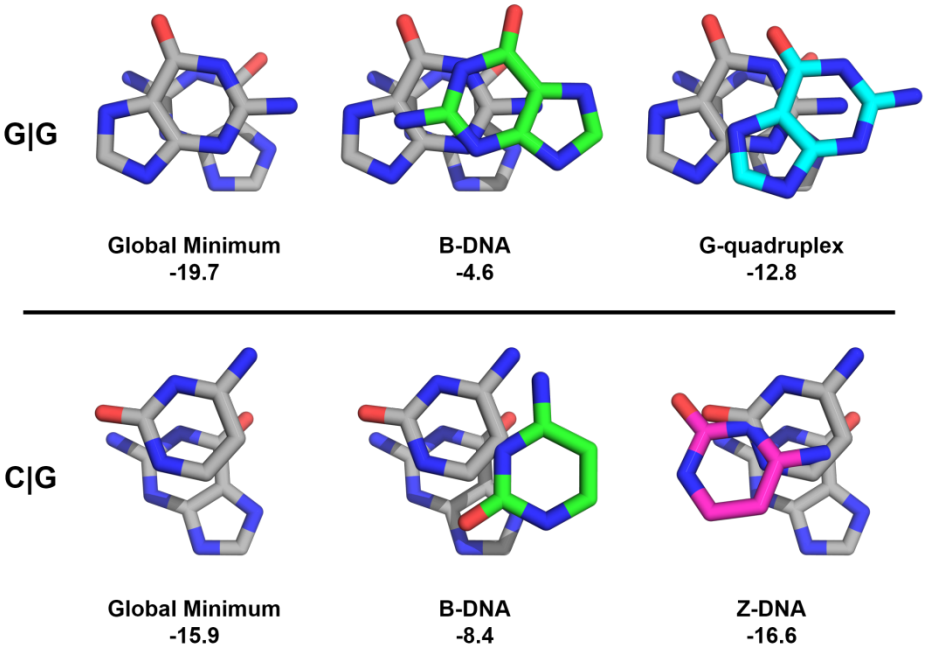

Figure 18. Comparison of global minimum stacked dimers (in gray) and those in B-DNA (green), a G-quadruplexe (cyan), and Z-DNA (magenta) for the G|G and C|G dimers. Hydrogens have been removed for clarity. Interaction energies in parenthesis are in kcal mol^{-1}.

We can compare other global minimum energy stacked dimers with their counterparts in specific macromolecules. For instance, Figure 18 also shows the lowest-lying stacked C|G dimer compared to those in B-DNA and Z-DNA. In both macromolecules cytosine is displaced laterally compared to the global minimum structure. In the case of B-DNA this results in a loss of nearly 50% of interaction strength. In Z-DNA, on the other hand, the stacking interaction exceeds that in the lowest-energy stacked dimer by 0.8 kcal mol^{-1}. This is primarily a result of the tilting of cytosine in the Z-DNA structure, which results in enhanced interactions not feasible in the constrained stacked optimized structures. Regardless, the point remains that in non-

canonical structures such as G-quadruplexes and Z-DNA, the nucleobase stacking can be considerably more favorable than it is in canonical B-DNA.

C. Predictive Models of Nucleobase Stacking

To further elucidate the factors that control the strength of nucleobase stacking, we explored a number of multivariate models based on conventional molecular properties as well as the recently-introduced heterocycle descriptors from Bootsma *et al*.[99] Bootsma *et al*. developed several predictive models of stacking interactions of drug-like heterocycles with aromatic amino acid side chains (Phe, Tyr, and Trp) based on ESP-derived descriptors combined with the number of heavy atoms (N_{HA}) in the heterocycle and the amino acid side chain. Inspired by these models,[99] we fit the following functional form to the gas-phase stacking interaction energies for the 15 unique dimers of the natural nucleobases:

$$E_{int}^{pred} = -0.00275 N_{HA}^{A} N_{HA}^{B} \left(ESP_{range}^{A} + ESP_{range}^{B} \right) - 4.18 \qquad (1)$$

In equation 1, N_{HA}^{A} and N_{HA}^{B} are the number of heavy atoms in nucleobases A and B, respectively, while ESP_{range}^{A} and ESP_{range}^{B} are the corresponding ESP_{range} (see Figure 17). Interaction energies predicted using equation 1 are listed in Table 5; these data are plotted against the DLPNO-CCSD(T)/cc-pVQZ values in Figure 19a. There are some outliers. For example, equation 1 overestimates the stacking of T|T, T|U, and U|U by about 1.5 kcal mol^{-1} and underestimates the stacking of C|A by about the same amount. Regardless, this two-parameter model provides surprisingly robust predictions (R^2 = 0.89), with a root mean square error (RMSE) of 0.9 kcal mol^{-1} and is robust to cross-

validation ($Q^2 = 0.87$). Moreover, the two parameters only impact the overall slope and intercept. In other words, the trend in nucleobase stacking can be captured reliably using a parameter-free model based on the products of N_{HA} and sum of ESP_{range} for the two nucleobases.

To probe the generality of equation 1, we utilized the 35 dimers of pyrimidine, purine, and the artificial nucleobases from Figure 17 with the natural nucleobases as a test set. Despite being trained on only 15 stacked dimers of natural nucleobases, equation 1 provides very accurate predictions for the dimers in this test set (see Figure 19a), with an R^2 of 0.90 and RMSE of 0.9 kcal mol^{-1}. That equation 1, which is built around a simple ESP-based descriptor, can reliably capture stacking interactions of natural nucleobases with this diverse set of artificial nucleobases, confirms the importance of electrostatic effects in determining the intrinsic stacking abilities of nucleobases.

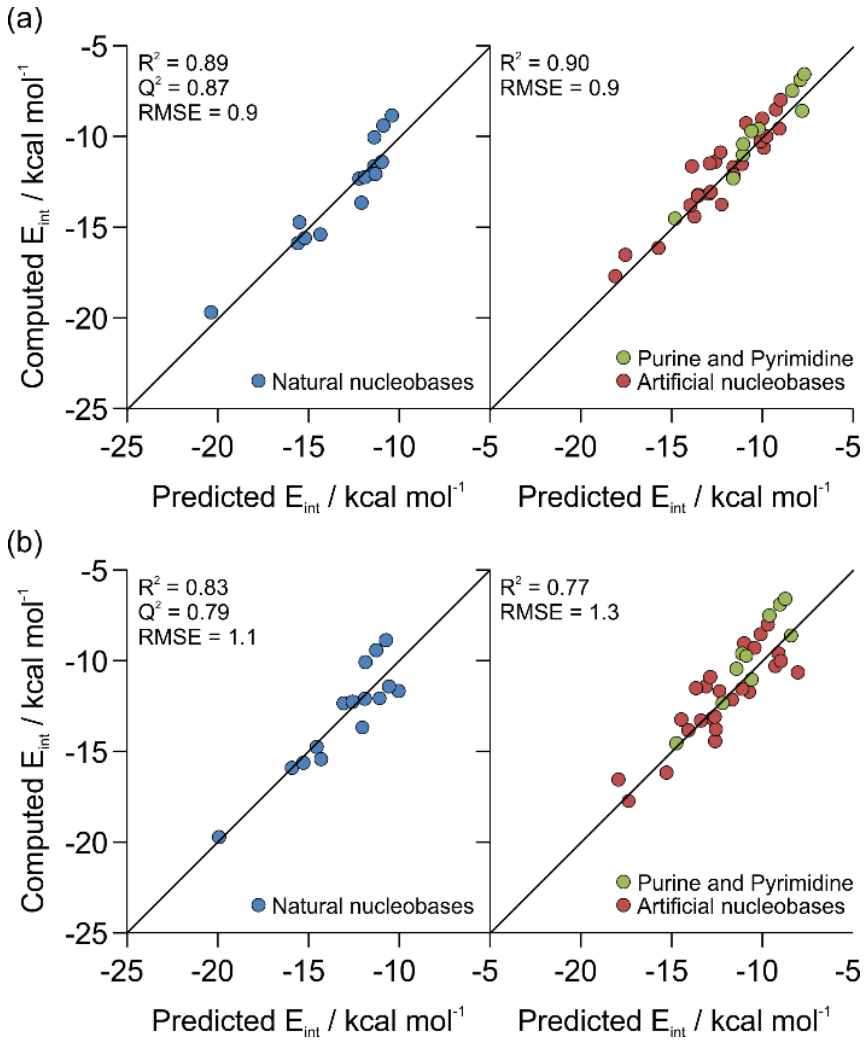

Figure 19. Correlation of predicted interaction energies from Equations 1 (a) and 2 (b) versus DLPNO-CCSD(T) data for the global minimum energy stacked dimers of the training set data (blue) and the test set data (red and green).

The ability to rapidly predict the ability of a heterocycle to stack with one of the natural nucleobases should also prove useful in several contexts. For example, there has been tremendous effort put into the development of non-natural nucleobase analogs, with uses in everything from expanding the genetic alphabet[224] to the development of fluorescent nucleobase analogs.[214] However, previously it was unclear whether these

non-natural nucleobase analogs would exhibit stacking interactions commensurate with the natural counterparts, and concerns have been raised[225] about the disruption of DNA helical stability that can accompany the incorporation of some non-natural nucleobases. Equation 1 can be used to rapidly predict the intrinsic stacking of potential new artificial nucleobases.

Bootsma *et al*.[67] noted that ESP_{range} provides similar information regarding the distribution of charge within a molecule as the molecular dipole moment (μ), which is a more familiar and intuitive descriptor for many chemists. This similarity can be seen by the correlation of the values of these descriptors for the nucleobases in Figure 17 (R^2 = 0.9). To see how well μ performs when paired with N_{HA}, we fit equation 2 to the same training data as used for equation 1:

$$E_{int}^{pred} = -0.00824 N_{HA}^{A} N_{HA}^{B} \left(\mu^{A} + \mu^{B}\right) - 6.04 \qquad (2)$$

Although nucleobase stacking can be predicted qualitatively using only the number of heavy atoms and the molecular dipole moments (see Figure 19b), the use of ESP_{range} instead of μ (*i.e.* equation 1) provides a more robust fit. The predicted energies from equation 2 are listed in Table 5.

D. Conceptual Understanding of Nucleobase Stacking

The simple form of equation 1 lends itself to the development of a conceptual understanding of the factors that determine the strength of nucleobase stacking. Overall, stacking is enhanced for nucleobases with more heavy-atoms (N_{HA}) and a larger range of ESP values within the van der Waals (vdW) projection of the nucleobase onto the plane 3.25 Å away (*i.e.* larger ESP_{range} value). The ESPs of the systems studied are shown in Figure 20; the corresponding ESP_{range} values are listed in Figure 17. The overall trend in

ESP_{range} values is G > C > T > U > A. In other words, G stacks strongly because of both its size (N_{HA}) and the large ESP_{range} value. The stacking of adenine, by contrast, is weaker, despite its similar size, due to its small ESP_{range} value. Similarly, the strong stacking exhibited by cytosine can be attributed to an ESP_{range} value that approaches that of G.

Bootsma and Wheeler[226] recently showed that ESP_{range} values for heterocycles follow predictable trends based on the number and relative positions of different heteroatom types (imino nitrogens, amino nitrogens, carbonyls, *etc*.). In particular, the impact of these heteroatoms on the ESP can be modeled as the sum of local dipoles located at the position of each heteroatom. Similar behavior can be seen for the nucleobases. The magnitude of these dipoles varies with functional group, following the general order amido > imino N (-N=) > endocyclic amino (-NH-) > exocyclic amino ($-NH_2$), with the direction of the dipole for amino groups opposite that of the others. The impact of multiple functional groups on ESP_{range} is then due to the sum of these local dipoles.

The ESPs of the nucleobases are shown in Figure 20. These ESPs can be decomposed into contributions of individual functional groups/heteroatoms. For instance, Figure 21a shows the ESP of guanine (far right) as the ESP of indole corrected for the two imino Ns, exocyclic NH_2, and amido group. Similarly, Figure 21b shows the ESP of adenine constructed from the ESP of indole corrected for three two imino Ns and an exocyclic NH_2. These additive ESPs are nearly identical to the actual ESPs of these nucleobases (compare the rightmost images in Figure 21 with the corresponding ESPs in Figure 20). Furthermore, this additivity applies to the other nucleobases, enabling the

visualization of contributions of each functional group to ESP_{range} and the intrinsic stacking interaction energy of these nucleobases.

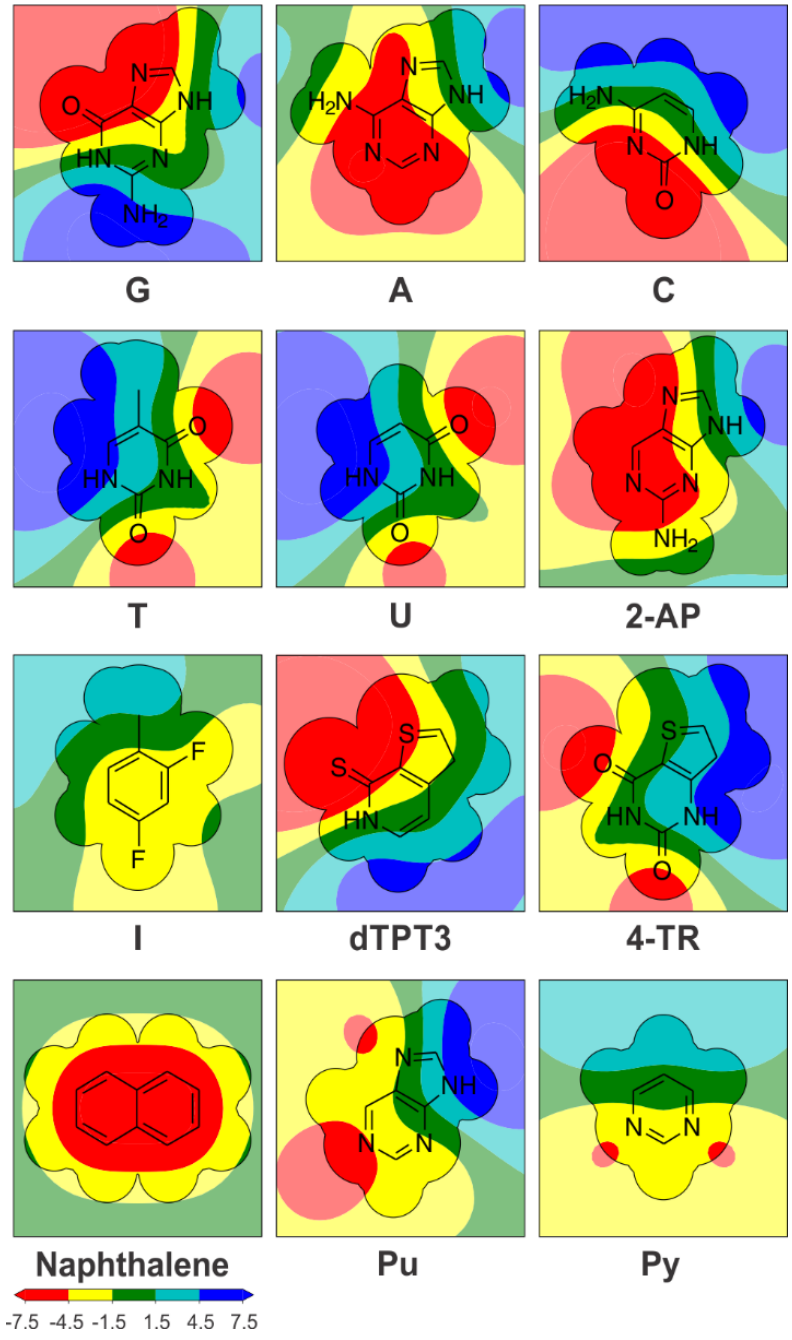

Figure 20. Electrostatic potentials (in kcal mol^{-1}, according to the scale shown) in the plane 3.25 Å from the nucleobases, along with the vdW projection onto this plane. ESP_{range} is the range of ESP values within the outlined area.

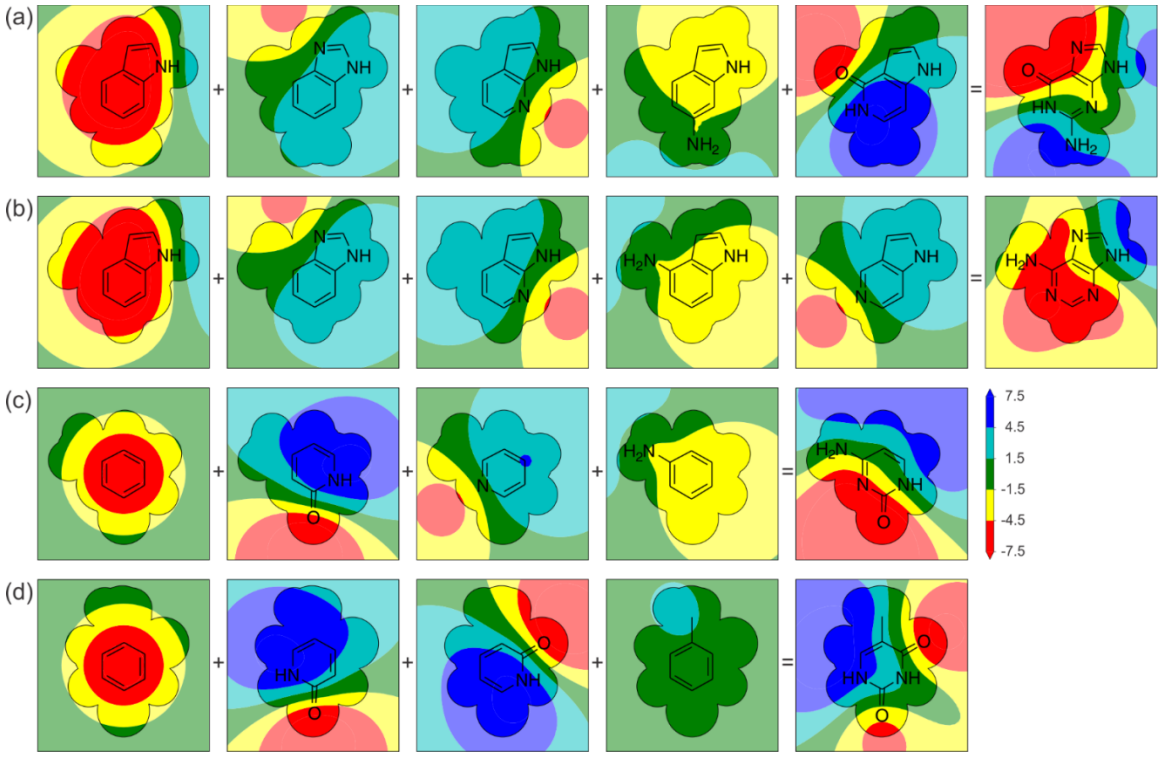

Figure 21. Construction of approximate ESPs (in kcal mol^{-1}, according to the scale shown) of (a) guanine, (b) adenine, (c) cytosine, and (d) thymine (far right) by appending the contributions of each heteroatom/functional group to the ESP of indole (for a and b) or benzene (for c and) (far left). The vdW projection of the complete nucleobase onto this plane is outlined in each panel.

For example, the drastically different ESP_{range} values for guanine and adenine can be understood by considering how the local dipoles corresponding to the different heteroatom/functional groups combine. In adenine, the three imino nitrogens have the biggest impact on the ESP. However, given their relative orientations, these effects largely cancel, leading to an ESP_{range} value similar to that of indole. This can be contrasted with guanine. While there is a cancelation of the impact of the two imino nitrogens, this is far outweighed by the strong additivity of the amido and adjacent imino

groups. The result is an ESP_{range} value (24.3 kcal mol^{-1}) significantly larger not only than that of the parent indole but of any of the nucleobases. Similarly, while the two amino groups in thymine partially cancel, the amino and imino groups exhibit strong additivity that leads to the large ESP_{range} value for cytosine.

Of course, this additivity of local dipoles associated with each heteroatom/functional group is partially captured by the molecular dipole moments. However, the consideration of ESP_{range} (and the underlying *local* dipole moments) provides a more nuanced view of the electrostatic contributions of nucleobase stacking (see Figures 4a and 4b). This echoes the observations of Gellman and co-workers from 25 years ago.[58-59]

III. Conclusions

We have studied the maximal stacking interactions possible between stacked pairs of natural and artificial nucleobases using robust quantum chemical methods. The data reveal that the strength of these intrinsic stacking interactions are primarily modulated by electrostatic effects. Comparisons of these intrinsic stacking interactions for the natural nucleobases with the stacking in particular biological macromolecules (B-DNA, Z-DNA, and G-quadruplexes) reveals a wide range of deviations from ideal stacking. That is, for some systems (e.g. B-DNA) geometric constraints force the nucleobases to adopt stacked geometries far from ideal, whereas in others (Z-DNA and G-quadruplexes) the stacked dimers can be much closer to their ideal geometries. Overall, we find that the superior intrinsic stacking interactions of guanine and cytosine, compared to adenine and thymine, respectively, can be traced to the clustering of amide and imino functional groups on in

the former cases. This picture is consistent with the local, direct interaction model of substituent effects in π-stacking interactions[196] as well as Gellman's model[58-59] of nucleobase stacking in which local charges, not hydrophobicity or dispersion interactions, govern nucleobase stacking. Thus, while differential desolvation effects undoubtedly impact trends in nucleobase stacking interactions, there are marked differences in the stacking affinities of both natural and non-natural nucleobases that are independent of the hydrophobic effect. We can now not only understand these differences, but readily predict the stacking of new nucleobase analogs.

IV. Computational Section

Structures of stacked dimers were optimized at the ωB97X-D/def2-TZVP level of theory[172-173] with the heavy atoms of each monomer constrained to lie in parallel planes. Initial geometries for these optimizations were constructed by placing a nucleobase at nine evenly spaced positions in a 2.8 x 2.8 Å square grid above the other nucleobase. At each position, we considered six orientations of the nucleobase generated by 60° rotations. For each of these dimer geometries the top base was flipped by 180° for a total of 108 unique initial geometries. Each of these dimers was optimized the nearest local minimum energy structure; unique stacked dimers were identified based on a RMSD of 0.4 Å. DLPNO-CCSD(T)/cc-pVQZ single point energies were computed for all dimers and all monomers in their corresponding dimer geometries (using default PNO cutoffs). Interaction energies, defined as the difference in energy between the optimized dimer and the monomers in the optimized dimer geometry, were computed for all unique stacked

dimers. Herein, we focus on dimers with the most favorable interaction energy for each nucleobase combination.

All DFT optimizations were performed using Gaussian09,[130] while Orca 4.0[178] was used for DLPNO-CCSD(T). SAPT computations were performed at the SAPT0/jun-cc-pVDZ level[217-220,227-228] using Psi4.[229] ESP_{range} values were determined by first computing the ESP on a 0.1 Å spaced grid of points 3.25 Å from the nucleobase at the ωB97X-D/def2-TZVP level of theory[172-173] and then identifying the minimum and maximum values within the projection of the van der Waals volume,[230] as done previously.[97]

Dimers of G|G and C|G were taken from PDB crystal structures of a G-quadruplex[231] and Z-DNA[232], respectively. Hydrogens were added to the crystal structure dimers and optimized using ωB97X-D/def2-TZVP level of theory. The selected dimers were the strongest interacting dimers computed at the DLPNO-CCSD(T)/cc-pVQZ level of theory.

CHAPTER 5

CONCLUSIONS

Stacking interactions are central to the behavior of nucleobases in biological and non-biological contexts. Unfortunately, we currently lack both computational and conceptual tools to leverage the power of these interactions in the design of both molecular sensors and DNA-targeting drugs. This work aimed to help fill this gap by quantifying the role of stacking and other non-covalent interactions in a guanine sensor, by developing best-practices for quantum mechanical studies of DNA intercalation, and by developing quantitative predictive models of intrinsic stacking interactions of natural and non-natural nucleobases combined with a conceptual understanding of these interactions.

In the case of guanine targeting sensors, I showed that the backbone, 2-acetemido-1,8-naphthyridine, is comparable to cytosine as a hydrogen bonding partner for guanine and demonstrated that the guanine binding strengths of a series of potential guanine sensors are not only influenced by varying the π-stacking area and point of connectivity of the aryl group but also the local orientation of the aryl-substituent upon binding. In the case of modeling DNA intercalators, I established guidelines for the choice of both DFT method and binding site model size (*i.e.* how many base pairs and corresponding backbones). Finally, I studied the intrinsic stacking of natural and artificial nucleobases, showing that nucleobase stacking is dominated by dispersion interactions but modulated by electrostatic effects. Furthermore, I developed a robust model to predict the intrinsic

stacking of natural and artificial nucleobases that provides key insights into the origin of trends in nucleobase stacking strengths.

REFERENCES

1. Sivakova, S.; Rowan, S. J., Nucleobases as supramolecular motifs. *Chem. Soc. Rev.* **2005,** *34*, 9-21.

2. Davis, J. T.; Spada, G. P., Supramolecular architectures generated by self-assembly of guanosine derivatives. *Chem. Soc. Rev.* **2007,** *36*, 296-313.

3. Jissy, A. K.; Datta, A., Design and Applications of Noncanonical DNA Base Pairs. *J. Phys. Chem. Lett.* **2014,** *5*, 154-166.

4. Šponer, J.; Shukla, M. K.; Wang, J.; Leszczynski, J., Computational modeling of DNA and RNA fragments. *Handbook of Computational Chemistry* **2017**, 1803-1826.

5. Zhou, L.; Retailleau, P.; Morel, M.; Rudiuk, S.; Baigl, D., Photoswitchable Fluorescent Crystals Obtained by the Photoreversible Coassembly of a Nucleobase and an Azobenzene Intercalator. *J. Am. Chem. Soc.* **2019,** *141*, 9321-9329.

6. Cai, H.; Huang, Y.-L.; Li, D., Biological metal–organic frameworks: Structures, host–guest chemistry and bio-applications. *Coord. Chem. Rev.* **2019,** *378*, 207-221.

7. Persch, E.; Dumele, O.; Diederich, F., Molecular Recognition in Chemical and Biological Systems. *Angewandte Chemie International Edition* **2015,** *54*, 3290-3327.

8. Bissantz, C.; Kuhn, B.; Stahl, M., A Medicinal Chemist's Guide to Molecular Interactions. *Journal of Medicinal Chemistry* **2010,** *53*, 5061-5084.

9. Meyer, E. A.; Castellano, R. K.; Diederich, F., Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem. Int. Ed.* **2003,** *42*, 1210-1250.

10. Salonen, L. M.; Ellermann, M.; Diederich, F., Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angewandte Chemie International Edition* **2011,** *50*, 4808-4842.

11. Schneider, H.-J., Binding Mechanisms in Supramolecular Complexes. *Angewandte Chemie International Edition* **2009,** *48*, 3924-3977.

12. Hunter, C. A.; Sanders, J. K. M., The Nature of π-π Interactions. *J. Am. Chem. Soc.* **1990,** *112*, 5525-5534.

13. Sinnokrot, M. O.; Sherrill, C. D., High-Accuracy Quantum Mechanical Studies of π-π Interactions in Benzene Dimers. *Journal of Physical Chemistry A* **2006,** *110*, 10656-10668.

14. Grimme, S., Do Special Noncovalent π-π Stacking Interactions Really Exist? *Angew. Chem. Int. Ed.* **2008,** *47*, 3430-3434.

15. Bloom, J. W.; Wheeler, S. E., Taking the aromaticity out of aromatic interactions. *Angew Chem Int Ed Engl* **2011,** *50*, 7847-7849.

16. Martinez, C. R.; Iverson, B. L., Rethinking the Term "π-Stacking". *Chem. Sci.* **2012,** *3*, 2191-2201.

17. Wheeler, S. E., Understanding Substituent Effects in Noncovalent Interactions Involving Aromatic Rings. *Accounts of Chemical Research* **2013,** *46*, 1029-1038.

18. Wheeler, S. E.; Bloom, J. W. G., Toward a More Complete Understanding of Noncovalent Interactions Involving Aromatic Rings. *The Journal of Physical Chemistry A* **2014,** *118*, 6133-6147.

19. Huppert, J. L.; Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33*, 2908-2916.

20. Suram, A.; Rao, J. K. S.; S, L. K.; A, V. M., First Evidence to Show the Topological Change of DNA from B-DNA to Z-DNA Conformation in the Hippocampus of Alzheimer's Brain. *NeuroMolecular Medicine* **2002,** *2*, 289-298.

21. Morgan, B. S.; Forte, J. E.; Culver, R. N.; Zhang, Y.; Hargrove, A. E., Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angew. Chem. Int. Ed.* **2017,** *56*, 13498-13502.

22. Patwardhan, N. N.; Ganser, L. R.; Kapral, G. J.; Eubanks, C. S.; Lee, J.; Sathyamoorthy, B.; Al-Hashimi, H. M.; Hargrove, A. E., Amiloride as a new RNA-binding scaffold with activity against HIV-1 TAR. *Med. Chem. Commun.* **2017,** *8*, 1022-1036.

23. Řeha, D.; Kabeláč, M.; Ryjáček, F.; Šponer, J.; Šponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P., Intercalators. 1. Nature of Stacking Interactions between Intercalators (Ethidium, Daunomycin, Ellipticine, and 4'6-Diaminide-2-phenylindole) and DNA Base Pairs. *Ab Initio* Quantum Chemical, Density Functional Theory, and Empirical Potential Study. *J. Am. Chem. Soc.* **2002,** *124*, 3366-3376.

24. Langner, K. M.; Kedzierski, P.; Sokalski, W. A.; Leszczynski, J., Physical nature of ethidium and proflavine interactions with nucleic acid bases in the intercalation plane. *J. Phys. Chem. B* **2006,** *110*, 9720-9727.

25. Li, S.; Cooper, V. R.; Thonhauser, T.; Lundqvist, B. I.; Langreth, D. C., Stacking interactions and DNA intercalation. *J. Phys. Chem. B* **2009,** *113*, 11166-11172.

26. Rescifina, A.; Zagni, C.; Varrica, M. G.; Pistarà, V.; Corsaro, A., Recent advances in small organic molecules as DNA intercalating agents: Synthesis, activity, and modeling. *European Journal of Medicinal Chemistry* **2013,** *74*, 95-115.

27. Hargis, J. C.; Schaefer, H. F.; Houk, K. N.; Wheeler, S. E., Non-Covalent Interactions of a Benzo[a]pyrene Diol Epoxide with DNA Base Pairs: Insight into the Formation of Adducts of (+)-BaP DE-2 with DNA. *J. Phys. Chem. A* **2010,** *114*, 2038-2044.

28. Lu, S. H.; Phang, R.; Fang, J. M., Fluorescent Sensing of Guanine and Guanosine Monophosphate with Conjugated Receptors Incorporating Aniline and Naphthyridine Moieties. *Org. Lett.* **2016,** *18*, 1724-1727.

29. Leung, C. H.; Chan, D. S.; Ma, V. P.; Ma, D. L., DNA-binding small molecules as inhibitors of transcription factors. *Med. Res. Rev.* **2013,** *33*, 823-846.

30. Mross, K.; Massing, U.; Kratz, F., DNA-intercalators — the anthracyclines. In *Drugs Affecting Growth of Tumours. Milestones in Drug Therapy*, Pinedo, H. M.; Smorenburg, C. H., Eds. Birkhäuser Basel: 2006; pp 19-81.

31. Figgitt, D.; Denny, W.; Chavalitshewinkoon, P.; Wilairat, P.; Ralph, R., In vitro study of anticancer acridines as potential antitrypanosomal and antimalarial agents. *Antimicrobial Agents and Chemotherapy* **1992,** *36*, 1644-1647.

32. Mukherjee, A.; Sasikala, W. D., Drug-DNA intercalation: From discovery to the molecular mechanism. In *Advances in Protein Chemistry and Structural Biology*, 2013; Vol. 92, pp 1-62.

33. Soni, A.; Khurana, P.; Singh, T.; Jayaram, B., A DNA intercalation methodology for an efficient prediction of ligand binding pose and energetics. *Bioinformatics* **2017,** *33*, 1488-1496.

34. Ricci, C. G.; Netz, P. A., Docking studies on DNA-ligand interactions: building and application of a protocol to identify the binding mode. *J. Chem. Inf. Model.* **2009,** *49*, 1925-1935.

35. Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P., Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *Journal of Chemical Theory and Computation* **2015,** *11*, 5723-5736.

36. Mackerell, A. D., Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* **2004,** *25*, 1584-1604.

37. Jambrec, D.; Haddad, R.; Lauks, A.; Gebala, M.; Schuhmann, W.; Kokoschka, M., DNA Intercalators for Detection of DNA Hybridisation: SCS(MI)-MP2 Calculations and Electrochemical Impedance Spectroscopy. *ChemPlusChem* **2016,** *81*, 604-612.

38. Xu, P.; Wang, J.; Xu, Y.; Chu, H.; Shen, H., Advance in Structural Bioinformatics. 2015; Vol. 827, pp 187-203.

39. Terryn, R. J.; German, H. W.; Kummerer, T. M.; Sinden, R. R.; Baum, J. C.; Novak, M. J., Novel computational study on π -stacking to understand mechanistic interactions of Tryptanthrin analogues with DNA. *Toxicology Mechanisms and Methods* **2014,** *24*, 73-79.

40. Hardebeck, L. K. E.; Johnson, C. A.; Hudson, G. A.; Ren, Y.; Watt, M.; Kirkpatrick, C. C.; Znosko, B. M.; Lewis, M., Predicting DNA-intercalator binding: The development of an arene-arene stacking parameter from SAPT analysis of benzene-substituted benzene complexes. *Journal of Physical Organic Chemistry* **2013,** *26*, 879-884.

41. Hill, G. M.; Moriarity, D. M.; Setzer, W. N., Attenuation of cytotoxic natural product DNA intercalating agents by caffeine. *Scientia Pharmaceutica* **2011,** *79*, 729-747.

42. Langner, K. M.; Janowski, T.; Góra, R. W.; Dziekoński, P.; Sokalski, W. A.; Pulay, P., The Ethidium-UA/AU Intercalation Site: Effect of Model Fragmentation and

Backbone Charge State. *Journal of Chemical Theory and Computation* **2011,** *7*, 2600-2609.

43. Hohenstein, E. G.; Parrish, R. M.; Sherrill, C. D.; Turney, J. M.; Schaefer, H. F., Large-scale symmetry-adapted perturbation theory computations via density fitting and Laplace transformation techniques: Investigating the fundamental forces of DNA-intercalator interactions. *Journal of Chemical Physics* **2011,** *135*.

44. Hargis, J. C.; Schaefer, H. F.; Houk, K. N.; Wheeler, S. E., Noncovalent interactions of a benzo[a]pyrene diol epoxide with DNA base pairs: Insight into the formation of adducts of (+)-BaP DE-2 with DNA. *Journal of Physical Chemistry A* **2010**.

45. Grant Hill, J.; Platts, J. A., Local electron correlation descriptions of the intermolecular stacking interactions between aromatic intercalators and nucleic acids. *Chemical Physics Letters* **2009,** *479*, 279-283.

46. Li, S.; Cooper, V. R.; Thonhauser, T.; Lundqvist, B. I.; Langreth, D. C., Stacking interactions and DNA intercalation. *Journal of Physical Chemistry B* **2009,** *113*, 11166-11172.

47. Barone, G.; Guerra, C. F.; Gambino, N.; Silvestri, A.; Lauria, A.; Almerico, A. M.; Bickelhaupt, F. M., Intercalation of daunomycin into stacked dna base pairs. Dft study of an anticancer drug. *Journal of Biomolecular Structure and Dynamics* **2008,** *26*, 115-129.

48. Černý, J.; Hobza, P., Non-covalent interactions in biomacromolecules. *Physical Chemistry Chemical Physics* **2007,** *9*, 5291.

49. Langner, K. M.; Kedzierski, P.; Sokalski, W. A.; Leszczynski, J., Physical nature of ethidium and proflavine interactions with nucleic acid bases in the intercalation plane. *Journal of Physical Chemistry B* **2006,** *110*, 9720-9727.

50. Kubař, T.; Hanus, M.; Ryjáček, F.; Hobza, P., Binding of cationic and neutral phenanthridine intercalators to a DNA oligomer is controlled by dispersion energy: Quantum chemical calculations and molecular mechanics simulations. *Chemistry - A European Journal* **2005,** *12*, 280-290.

51. Dračínský, M.; Castaño, O., Calculations of interaction energies of ellipticine derivatives with DNA base pairs. *Phys. Chem. Chem. Phys.* **2004,** *6*, 1799-1805.

52. Kumar, A.; Elstner, M.; Suhai, S., SCC-DFTB-D study of intercalating carcinogens: Benzo(a)pyrene and its metabolites complexed with the G-C base pair. *International Journal of Quantum Chemistry* **2003,** *95*, 44-59.

53. Řeha, D.; Kabeláč, M.; Ryjáček, F.; Sponer, J.; Šponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P., Intercalators. 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, and 4′,6-diaminide-2-phenylindole) and DNA base pairs. Ab initio quantum chemical, density functional theory, and empirical potential study. *Journal of the American Chemical Society* **2002,** *124*, 3366-3376.

54. Bondarev, D. A.; Skawinski, W. J.; Venanzi, C. A., Nature of Intercalator Amiloride - Nucelobase Stacking . An Empirical Potential and ab Initio Electron Correlation Study. *J. Phys. Chem. B* **2000,** *104*, 815-822.

55. Medhi, C.; Mitchell, J. B. O.; Price, S. L.; Tabor, A. B., Electrostatic factors in DNA intercalation. *Biopolymers* **1999,** *52*, 84-93.

56. Jambrec, D.; Haddad, R.; Lauks, A.; Gebala, M.; Schuhmann, W.; Kokoschka, M., DNA Intercalators for Detection of DNA Hybridisation: SCS(MI)-MP2 Calculations and Electrochemical Impedance Spectroscopy. *Chempluschem* **2016,** *81*, 604-612.

57. Sure, R.; Grimme, S., Corrected small basis set Hartree-Fock method for large systems. *Journal of Computational Chemistry* **2013,** *34*, 1672-1685.

58. Newcomb, L. F.; Gellman, S. H., Aromatic Stacking Interactions in Aqueous-Solution - Evidence That Neither Classical Hydrophobic Effects nor Dispersion Forces Are Important. *J. Am. Chem. Soc.* **1994,** *116*, 4993-4994.

59. McKay, S. L.; Haptonstall, B.; Gellman, S. H., Beyond the hydrophobic effect: attractions involving heteroaromatic rings in aqueous solution. *J. Am. Chem. Soc.* **2001,** *123*, 1244-1245.

60. Guckian, K. M.; Schweitzer, B. A.; Ren, R. X. F.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T., Factors contributing to aromatic stacking in water: Evaluation in the context of DNA. *J. Am. Chem. Soc.* **2000,** *122*, 2213-2222.

61. SantaLucia, J., Jr.; Allawi, H. T.; Seneviratne, P. A., Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **1996,** *35*, 3555-3562.

62. Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M. D., Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.* **2004,** *342*, 775-785.

63. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D., Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006,** *34*, 564-574.

64. Kilchherr, F.; Wachauf, C.; Pelz, B.; Rief, M.; Zacharias, M.; Dietz, H., Single-molecule dissection of stacking forces in DNA. *Science* **2016,** *353*, aaf5508.

65. SantaLucia, J. In *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.*, Proc. Natl. Acad. Sci. U. S. A., 1998; pp 1460-1465.

66. Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H., Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **1998,** *37*, 14719-14735.

67. Šponer, J.; Leszczyński, J.; Hobza, P., Nature of nucleic acid-base stacking: Nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs. *J. Phys. Chem.* **1996,** *100*, 5590-5596.

68. Jurecka, P.; Hobza, P., True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine...cytosine, adenine...thymine, and their 9- and 1-methyl derivatives: complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. *J. Am. Chem. Soc.* **2003,** *125*, 15608-15613.

69. Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P., Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006,** *8*, 1985-1993.

70. Hill, G.; Forde, G.; Hill, N.; Lester, W. A.; Andrzej Sokalski, W.; Leszczynski, J., Interaction energies in stacked DNA bases? How important are electrostatics? *Chem. Phys. Lett.* **2003,** *381*, 729-732.

71. Hesselmann, A.; Jansen, G.; Schutz, M., Interaction energy contributions of H-bonded and stacked structures of the AT and GC DNA base pairs from the combined density functional theory and intermolecular perturbation theory approach. *J. Am. Chem. Soc.* **2006,** *128*, 11730-11731.

72. Langner, K. M.; Sokalski, W. A.; Leszczynski, J., Intriguing relations of interaction energy components in stacked nucleic acids. *J. Chem. Phys.* **2007,** *127*, 111102.

73. Sedlak, R.; Jurecka, P.; Hobza, P., Density functional theory-symmetry adapted perturbation treatment energy decomposition of nucleic acid base pairs taken from DNA crystal geometry. *J. Chem. Phys.* **2007,** *127*, 075104.

74. Sponer, J.; Riley, K. E.; Hobza, P., Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem. Phys.* **2008,** *10*, 2595-2610.

75. Cybulski, H.; Sadlej, J., Symmetry-Adapted Perturbation-Theory Interaction-Energy Decomposition for Hydrogen-Bonded and Stacking Structures. *J. Chem. Theory Comput.* **2008,** *4*, 892-897.

76. Czyżnikowska, Ż., On the importance of electrostatics in stabilization of stacked guanine–adenine complexes appearing in B-DNA crystals. *Journal of Molecular Structure: THEOCHEM* **2009,** *895*, 161-167.

77. Czyznikowska, Z.; Gora, R. W.; Zalesny, R.; Bartkowiak, W.; Baranowska-Laczkowska, A.; Leszczynski, J., The effect of intermolecular interactions on the electric dipole polarizabilities of nucleic acid base complexes. *Chem. Phys. Lett.* **2013,** *555*, 230-234.

78. Kasavajhala, K.; Bikkina, S.; Patil, I.; MacKerell, A. D., Jr.; Priyakumar, U. D., Dispersion interactions between urea and nucleobases contribute to the destabilization of RNA by urea in aqueous solution. *J. Phys. Chem. B* **2015,** *119*, 3755-3761.

79. Rezac, J.; Riley, K. E.; Hobza, P., S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011,** *7*, 2427-2438.

80. Poater, J.; Swart, M.; Bickelhaupt, F. M.; Fonseca Guerra, C., B-DNA structure and stability: the role of hydrogen bonding, pi-pi stacking interactions, twist-angle, and solvation. *Org. Biomol. Chem.* **2014,** *12*, 4691-4700.

81. Ladik, J.; Bende, A.; Bogar, F., The electronic structure of the four nucleotide bases in DNA, of their stacks, and of their homopolynucleotides in the absence and presence of water. *J. Chem. Phys.* **2008,** *128*, 105101.

82. Rutledge, L. R.; Durst, H. F.; Wetmore, S. D., Computational comparison of the stacking interactions between the aromatic amino acids and the natural or (cationic) methylated nucleobases. *Phys. Chem. Chem. Phys.* **2008,** *10*, 2801-2812.

83. Churchill, C. D.; Navarro-Whyte, L.; Rutledge, L. R.; Wetmore, S. D., Effects of the biological backbone on DNA-protein stacking interactions. *Phys. Chem. Chem. Phys.* **2009,** *11*, 10657-10670.

84. Rutledge, L. R.; Durst, H. F.; Wetmore, S. D., Evidence for Stabilization of DNA/RNA-Protein Complexes Arising from Nucleobase-Amino Acid Stacking and T-Shaped Interactions. *J. Chem. Theory Comput.* **2009,** *5*, 1400-1410.

85. Kim, S. K.; Ha, T.; Schermann, J. P., Biomolecular structures: from isolated molecules to the cell crowded medium. *Phys. Chem. Chem. Phys.* **2010,** *12*, 3334-3335.

86. Podeszwa, R.; Patkowski, K.; Szalewicz, K., Improved interaction energy benchmarks for dimers of biological relevance. *Phys. Chem. Chem. Phys.* **2010,** *12*, 5974-5979.

87. Sponer, J.; Sponer, J. E.; Petrov, A. I.; Leontis, N. B., Quantum chemical studies of nucleic acids: can we construct a bridge to the RNA structural biology and bioinformatics communities? *J. Phys. Chem. B* **2010,** *114*, 15723-15741.

88. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010,** *132*, 154104.

89. Hunter, R. S.; van Mourik, T., DNA base stacking: the stacked uracil/uracil and thymine/thymine minima. *J. Comput. Chem.* **2012,** *33*, 2161-2172.

90. Barone, G.; Fonseca Guerra, C.; Bickelhaupt, F. M., B-DNA Structure and Stability as Function of Nucleic Acid Composition: Dispersion-Corrected DFT Study of Dinucleoside Monophosphate Single and Double Strands. *ChemistryOpen* **2013,** *2*, 186-193.

91. McDonald, A. R.; Denning, E. J.; MacKerell, A. D., Jr., Impact of geometry optimization on base-base stacking interaction energies in the canonical A- and B-forms of DNA. *J. Phys. Chem. A* **2013,** *117*, 1560-1568.

92. Sedlak, R.; Janowski, T.; Pitonak, M.; Rezac, J.; Pulay, P.; Hobza, P., The accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **2013,** *9*, 3364-3374.

93. Sponer, J.; Sponer, J. E.; Mladek, A.; Jurecka, P.; Banas, P.; Otyepka, M., Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers* **2013,** *99*, 978-988.

94. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015,** *11*, 2315-2328.

95. Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M., The S66x8 benchmark for noncovalent interactions revisited: explicitly correlated ab initio methods and density functional theory. *Phys. Chem. Chem. Phys.* **2016,** *18*, 20905-20925.

96. Rezac, J.; Hobza, P., Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016,** *116*, 5038-5071.

97. Bootsma, A. N.; Wheeler, S. E., Tuning Stacking Interactions between Asp-Arg Salt Bridges and Heterocyclic Drug Fragments. *J. Chem. Inf. Model.* **2019,** *59*, 149-158.

98. Bootsma, A. N.; Wheeler, S. E., Stacking Interactions of Heterocyclic Drug Fragments with Protein Amide Backbones. *ChemMedChem* **2018,** *13*, 835-841.

99. Bootsma, A. N.; Doney, A. C.; Wheeler, S. E., Predicting the Strength of Stacking Interactions between Heterocycles and Aromatic Amino Acid Side Chains. *J. Am. Chem. Soc.* **2019**.

100. Wade, C. R.; Broomsgrove, A. E. J.; Aldridge, S.; Gabbaï, F. P., Fluoride ion complexation and sensing using organoboron compounds. *Chem. Rev.* **2010,** *110*, 3958-3984.

101. Jackson, M. N., Jr.; Kamunde-Devonish, M. K.; Hammann, B. A.; Wills, L. A.; Fullmer, L. B.; Hayes, S. E.; Cheong, P. H.-Y.; Casey, W. H.; Nyman, M.; Johnson, D. W., An Overview of Selected Current Approaches to the Characterization of Aqueous Inorganic Clusters. *Dalton Trans.* **2015,** *44*, 16982-17006.

102. Carroll, C. N.; Naleway, J. J.; Haley, M. M.; Johnson, D. W., Arylethynyl Receptors for Anions and Emerging Applications in Cellular Imaging. *Chem. Soc. Rev.* **2010,** *39*, 3875-3888.

103. Lee, S.; Chen, S.-H.; Flood, A. H., A pentagonal cyanostar macrocycle with cyanostilbene CH donors binds anions and forms dialkylphosphate [3]rotaxanes. *Nat. Chem.* **2013,** *5*, 704-710.

104. Li, Y.; Flood, A. H., Pure CH hydrogen bonding to chloride ions: A pre-organized and rigid macrocyclic receptor. *Angew. Chem. Int. Ed.* **2008,** *47*, 2649-2652.

105. Bloom, J. W. G.; Wheeler, S. E., Taking Aromaticity out of Aromatic Interactions. *Angew. Chem. Int. Ed.* **2011,** *50*, 7847-7849.

106. Feibush, B.; Saha, M.; Onan, K.; Karger, B.; Giese, R., Hplc Separation of DNA Adducts Based on Hydrogen-Bonding. *J. Am. Chem. Soc.* **1987,** *109*, 7531-7533.

107. Murray, T. J.; Zimmerman, S. C., New triply hydrogen bonded complexes with highly variable stabilities. *Journal of the American Chemical Society* **1992,** *114*, 4010-4011.

108. Nakatani, K.; Sando, S.; Saito, I., Recognition of a single guanine bulge by 2-acylamino-1,8-naphthyridine. *J. Am. Chem. Soc.* **2000,** *122*, 2172-2177.

109. Nakatani, K.; Sando, S.; Kumasawa, H.; Kikuchi, J.; Saito, I., Recognition of guanine-guanine mismatches by the dimeric form of 2-amino-1,8-naphthyridine. *J. Am. Chem. Soc.* **2001,** *123*, 12650-12657.

110. Nakatani, K.; Sando, S.; Yoshida, K.; Saito, I., Specific binding of 2-amino-1,8-naphthyridine into a single guanine bulge as evidenced by photooxidation of GG doublet. *Bioorg. Med. Chem. Lett.* **2001,** *11*, 335-337.

111. Nakatani, K.; Hagihara, S.; Sando, S.; Sakamoto, S.; Yamaguchi, K.; Maesawa, C.; Saito, I., Induction of a remarkable conformational change in a human telomeric sequence by the binding of naphthyridine dimer: Inhibition of the elongation of a

telomeric repeat by telomerase. *Journal of the American Chemical Society* **2003,** *125*, 662-666.

112. Nakatani, K.; Kobori, A.; Kumasawa, H.; Goto, Y.; Saito, I., The binding of guanine-guanine mismatched DNA to naphthyridine dimer immobilized sensor surfaces: kinetic aspects. *Bioorg. Med. Chem.* **2004,** *12*, 3117-3123.

113. Cywinski, P. J.; Moro, A. J.; Ritschel, T.; Hildebrandt, N.; Lohmannsroben, H. G., Sensitive and selective fluorescence detection of guanosine nucleotides by nanoparticles conjugated with a naphthyridine receptor. *Anal. Bioanal. Chem.* **2011,** *399*, 1215-1222.

114. Atsumi, H.; Maekawa, K.; Nakazawa, S.; Shiomi, D.; Sato, K.; Kitagawa, M.; Takui, T.; Nakatani, K., Tandem arrays of TEMPO and nitronyl nitroxide radicals with designed arrangements on DNA. *Chemistry* **2012,** *18*, 178-183.

115. Hagihara, M.; He, H.; Kimura, M.; Nakatani, K.; Ncd, C. C., A small molecule regulates hairpin structures in d(CGG) trinucleotide repeats. *Bioorganic & Medicinal Chemistry Letters* **2012,** *22*, 2000-2003.

116. Liang, F.; Lindsay, S.; Zhang, P., 1,8-Naphthyridine-2,7-diamine: a potential universal reader of Watson-Crick base pairs for DNA sequencing by electron tunneling. *Org. Biomol. Chem.* **2012,** *10*, 8654-8659.

117. Tanaka, K.; Murakami, M.; Jeon, J.-H.; Chujo, Y., Enhancement of affinity in molecular recognition viahydrogen bonds by POSS-core dendrimer and its application for selective complex formation between guanosine triphosphate and 1,8-naphthyridine derivatives. *Org. Biomol. Chem.* **2012,** *10*, 90-95.

118. Atsumi, H.; Nakazawa, S.; Dohno, C.; Sato, K.; Takui, T.; Nakatani, K., Ligand-induced electron spin-assembly on a DNA tile. *Chem Commun (Camb)* **2013,** *49*, 6370-6372.

119. Wang, Y. X.; Wang, C. P.; Bo, H. Y.; Gao, Q.; Qi, H. L.; Zhang, C. X., Specific recognition of a single guanine bulge in dsDNA using a surface plasmon resonance sensor with immobilized 2-(2-aminoacetyl)amino-5,6,7-trimethyl-1,8-naphthyridine. *Sensors and Actuators B-Chemical* **2013,** *177*, 800-806.

120. Hong, C.; Otabe, T.; Matsumoto, S.; Dohno, C.; Murata, A.; Hagihara, M.; Nakatani, K., Formation of a ligand-assisted complex of two RNA hairpin loops. *Chemistry* **2014,** *20*, 5282-5287.

121. Lu, S. H.; Selvi, S.; Fang, J. M., Ethynyl-linked (pyreno)pyrrole-naphthyridine and aniline-naphthyridine molecules as fluorescent sensors of guanine via multiple hydrogen bondings. *J. Org. Chem.* **2007,** *72*, 117-122.

122. Sun, Y. Y.; Liao, J. H.; Fang, J. M.; Chou, P. T.; Shen, C. H.; Hsu, C. W.; Chen, L. C., Fluorescent organic nanoparticles of benzofuran-naphthyridine linked molecules: formation and fluorescence enhancement in aqueous media. *Org. Lett.* **2006,** *8*, 3713-3716.

123. Dyubankova, N.; Froeyen, M.; Abramov, M.; Mattelaer, H. P.; Herdewijn, P.; Lescrinier, E., NMR study on the interaction of the conserved CREX 'stem-loop' in the Hepatitis E virus genome with a naphthyridine-based ligand. *Organic & biomolecular chemistry* **2015,** *13*, 9665-9672.

124. Mardirossian, N.; Head-Gordon, M., ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation,

designed by a survival-of-the-fittest strategy. *Physical Chemistry Chemical Physics* **2014,** *16*, 9904-9924.

125. Grimme, S., Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comp. Chem.* **2006,** *27*, 1787-1799.

126. Becke, A., Density-Functional Thermochemistry. V. Systematic Optimization of Exchange-Correlation Functionals. *J. Chem. Phys.* **1997,** *107*, 8554-8560.

127. Zhao, Y.; Truhlar, D., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Account* **2008,** *120*, 215-241.

128. Miertus, S.; Scrocco, E.; Tomasi, J., Electrostatic Interaction of a Solute with a Continuum - a Direct Utilization of Abinitio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.* **1981,** *55*, 117-129.

129. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum mechanical continuum solvation models. *Chem. Rev.* **2005,** *105*, 2999-3093.

130. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., JA; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; ramillo, J.; Gomperts,

R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas, O.; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. *Gaussian 09, Revision D.01*, Gaussian, Inc.: 2009.

131. Chai, J.-D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Physical Chemistry Chemical Physics* **2008,** *10*, 6615-6620.

132. Feller, D., The role of databases in support of computational chemistry calculations. *Journal of Computational Chemistry* **1996,** *17*, 1571-1586.

133. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005,** *7*, 3297-3305.

134. Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L., Basis Set Exchange: A Community Database for Computational Sciences. *Journal of Chemical Information and Modeling* **2007,** *47*, 1045-1052.

135. Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F., Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. *The Journal of Chemical Physics* **2015,** *143*, 034108.

136. Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F., Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *The Journal of Chemical Physics* **2013,** *139*, 134101.

137. Riplinger, C.; Neese, F., An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of Chemical Physics* **2013,** *138*, 034106.

138. Neese, F.; Hansen, A.; Liakos, D. G., Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *The Journal of Chemical Physics* **2009,** *131*, 064103.

139. Neese, F., Software update: the ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018,** *8*.

140. Allen, W. D.; East, A. L. L.; Császár, A. G., In *Structures and Conformations of Non-Rigid Molecules*, Laane, J.; Dakkouri, M.; van der Veken, B.; Oberhammer, H., Eds. Kluwer: Dordrecht, 1993; pp 343-373.

141. Császár, A. G.; Allen, W. D.; Schaefer, H. F., In Pursuit of the ab Initio Limit for Conformational Energy Prototypes *J. Chem. Phys.* **1998,** *108*, 9751-9764.

142. East, A. L. L.; Allen, W. D., The Heat of Formation of NCO. *Journal of Chemical Physics* **1993,** *99*, 4638-4650.

143. Jeziorski, B.; Moszyński, R.; Szalewicz, K., Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994,** *94*, 1887-1930.

144. Szalewicz, K., Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *WIREs Comp. Mol. Sci.* **2012,** *2*, 254-272.

145. Hohenstein, E. G.; Sherrill, C. D., Density Fitting of Intramonomer Correlation Effects in Symmetry-Adapted Perturbation Theory *Journal of Chemical Physics* **2010,** *133*, 014101.

146. Hohenstein, E. G.; Sherrill, C. D., Density Fitting and Cholesky Decomposition Approximations in Symmetry-Adapted Perturbation Theory: Implementation and Application to Probe the Nature of π-π Interactions in Linear Acenes. *Journal of Chemical Physics* **2010,** *132*, 184111.

147. Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D., Psi4: An Open-Source *ab Initio* Electronic Structure Program. *WIREs Comp. Mol. Sci.* **2012,** *2*, 556-565.

148. Sherrill, C. D., Energy Component Analysis of π Interactions. *Accounts of Chemical Research* **2013,** *46*, 1020-1028.

149. Kruse, H.; Banas, P.; Sponer, J., Investigations of stacked DNA base-pair steps: highly-accurate stacking interaction energies, energy decomposition and many-body stacking effects. *J Chem Theory Comput* **2018**.

150. Boys, S. F.; Bernardi, F., The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors *Molecular Physics* **1970,** *19*, 553-566.

151. Papajak, E.; Truhlar, D. G., Convergent Partially Augmented Basis Sets for Post-Hartree−Fock Calculations of Molecular Properties and Reaction Barrier Heights. *Journal of Chemical Theory and Computation* **2011,** *7*, 10-18.

152. Johnston, R. C.; Cheong, P. H.-Y., C-H···O non-classical hydrogen bonding in the stereomechanics of organic transformations: theory and recognition. *Organic & biomolecular chemistry* **2013,** *11*, 5057-5064.

153. This correlation also indicates that there is little variation in the interaction of guanine with the ethynylphenyl component of the sensor, which was not included in any of the four truncated models.

154. Rescifina, A.; Zagni, C.; Varrica, M. G.; Pistarà, V.; Corsaro, A., Recent advances in small organic molecules as DNA intercalating agents: Synthesis, activity, and modeling. *Eur. J. Med. Chem.* **2014,** *74*, 95-115.

155. Gilad, Y.; Senderowitz, H., Docking studies on DNA intercalators. *J. Chem. Inf. Model.* **2014,** *54*, 96-107.

156. Churchill, C. D.; Rutledge, L. R.; Wetmore, S. D., Effects of the biological backbone on stacking interactions at DNA-protein interfaces: the interplay between the backbone...pi and pi...pi components. *Phys. Chem. Chem. Phys.* **2010,** *12*, 14515-14526.

157. Lenz, S. A.; Kellie, J. L.; Wetmore, S. D., Glycosidic bond cleavage in deoxynucleotides: effects of solvent and the DNA phosphate backbone in the computational model. *J. Phys. Chem. B* **2012,** *116*, 14275-14284.

158. Hehre, W. J.; Stewart, R. F.; Pople, J. A., Self-consistent molecular-orbital methods. I. Use of gaussian expansions of slater-type atomic orbitals. *The Journal of Chemical Physics* **1969,** *51*, 2657-2664.

159. East, A. L. L.; Allen, W. D., The Heat of Formation of Nco. *J. Chem. Phys.* **1993,** *99*, 4638-4650.

160. Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R., RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chemical Physics Letters* **1998,** *294*, 143-152.

161. Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F., Sparse maps-A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple

linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals. *J. Chem. Phys.* **2015,** *143*, 034108.

162. Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F., Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013,** *139*, 134101.

163. Riplinger, C.; Neese, F., An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013,** *138*, 034106.

164. Neese, F.; Hansen, A.; Liakos, D. G., Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys.* **2009,** *131*, 064103.

165. Dunning, T. H., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989,** *90*, 1007-1023.

166. Feller, D., The Use of Systematic Sequences of Wave Functions for Estimating the Complete Basis Set, Full Configuration Interaction Limit in Water. *J. Chem. Phys.* **1993,** *98*, 7059-7071.

167. Helgaker, T.; Klopper, W.; Koch, H.; Noga, J., Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997,** *106*, 9639-9646.

168. Kruse, H.; Banáš, P.; Šponer, J., Investigations of Stacked DNA Base-Pair Steps: Highly Accurate Stacking Interaction Energies, Energy Decomposition, and Many-Body Stacking Effects. *Journal of Chemical Theory and Computation* **2019,** *15*, 95-115.

169. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.

170. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010,** *132*, 154104.

171. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008,** *120*, 215-241.

172. Chai, J. D.; Head-Gordon, M., Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008,** *10*, 6615-6620.

173. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7*, 3297-3305.

174. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32*, 1456-1465.

175. Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A., Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *Journal of Chemical Physics* **2015,** *143*, 054107.

176. E. D. Glendening, A. E. R., J. E. Carpenter, F. Weinhold, NBO Version 3.1.

177. Reed, A. E.; Weinstock, R. B.; Weinhold, F., Natural-Population Analysis. *J. Chem. Phys.* **1985,** *83*, 735-746.

178. Neese, F., Software update: the ORCA program system, version 4.0. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2018,** *8*.

179. Maehigashi, T.; Persil, O.; Hud, N. V.; Williams, L. D., Crystal Structure of Proflavine in Complex with a DNA hexamer duplex. *To be publsihed*.

180. Lisgarten, J. N.; Coll, M.; Portugal, J.; Wright, C. W.; Aymami, J., The antimalarial and cytotoxic drug cryptolepine intercalates into DNA at cytosine-cytosine sites. *Nat Struct Biol* **2002,** *9*, 57-60.

181. Canals, A.; Purciolas, M.; Aymami, J.; Coll, M., The anticancer agent ellipticine unwinds DNA by intercalative binding in an orientation parallel to base pairs. *Acta Crystallogr D Biol Crystallogr* **2005,** *61*, 1009-1012.

182. Ferraroni, M.; Bazzicalupi, C.; Gratteri, P., Crystal structure of the complex of DNA hexamer d(CGATCG) with Coptisine. *To be published*.

183. Adams, A.; Guss, J. M.; Collyer, C. A.; Denny, W. A.; Wakelin, L. P., Crystal structure of the topoisomerase II poison 9-amino-[N-(2-dimethylamino)ethyl]acridine-4-carboxamide bound to the DNA hexanucleotide d(CGTACG)2. *Biochemistry* **1999,** *38*, 9221-9233.

184. Lipscomb, L. A.; Peek, M. E.; Zhou, F. X.; Bertrand, J. A.; D., V.; L.D., W., Water ring structure at DNA interfaces: hydration and dynamics of DNA-anthracycline complexes. *Biochemistry* **1994,** *33*, 3649-3659.

185. Leonard, G. A.; Hambley, T. W.; McAuley-Hecht, K.; Brown, T.; Hunter, W. N., Anthracycline-DNA interactions at unfavourable base-pair triplet-binding sites: structures of d(CGGCCG)/daunomycin and d(TGGCCA)/adriamycin complexes. *Acta Crystallogr D Biol Crystallogr* **1993,** *49*, 458-467.

186. Gasper, S. M.; Armitage, B.; Shui, X.; Hu, G. G.; Yu, C.; Schuster, G. B.; Williams, L. D., Three-Dimensional Structure and Reactivity of a Photochemical Cleavage Agent Bound to DNA. *Journal of the American Chemical Society* **1998,** *120*, 12402-12409.

187. Nunn, C. M.; Van Meervelt, L.; Zhang, S.; Moore, M. H.; Kennard, O., DNA-drug interactions. *Journal of molecular biology* **1991,** *222*, 167-177.

188. Adams, A.; Guss, J. M.; Collyer, C. A.; Denny, W. A.; Prakash, A. S.; Wakelin, L. P. G., Acridinecarboxamide Topoisomerase Poisons: Structural and Kinetic Studies of the DNA Complexes of 5-Substituted 9-Amino-(N-(2-dimethylamino)ethyl)acridine-4-Carboxamides. *Molecular Pharmacology* **2000,** *58*, 649-658.

189. Temperini, C., The crystal structure of the complex between a disaccharide anthracycline and the DNA hexamer d(CGATCG) reveals two different binding sites involving two DNA duplexes. *Nucleic Acids Research* **2003,** *31*, 1464-1469.

190. Madsen, M.; Gothelf, K. V., Chemistries for DNA Nanotechnology. *Chemical Reviews* **2019,** *119*, 6384-6458.

191. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nature Reviews Materials* **2018,** *3*, 17068.

192. Malyshev, D. A.; Romesberg, F. E., The Expanded Genetic Alphabet. *Angewandte Chemie - International Edition* **2015,** *54*, 11930-11944.

193. Sinnokrot, M. O.; Sherrill, C. D., Substituent Effects in π-π Interactions: Sandwich and T-Shaped Configurations. *J. Am. Chem. Soc.* **2004,** *126*, 7690-7697.

194. Cockroft, S. L.; Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J., Electrostatic control of aromatic stacking interactions. *J. Am. Chem. Soc.* **2005,** *127*, 8594-8595.

195. Cockroft, S. L.; Perkins, J.; Zonta, C.; Adams, H.; Spey, S. E.; Low, C. M.; Vinter, J. G.; Lawson, K. R.; Urch, C. J.; Hunter, C. A., Substituent effects on aromatic stacking interactions. *Org. Biomol. Chem.* **2007,** *5*, 1062-1080.

196. Wheeler, S. E.; Houk, K. N., Substituent effects in the benzene dimer are due to direct interactions of the substituents with the unsubstituted benzene. *J. Am. Chem. Soc.* **2008,** *130*, 10854-10855.

197. Ringer, A. L.; Sherrill, C. D., Substituent Effects in Sandwich Configurations of Multiply Substituted Benzene Dimers Are Not Solely Governed By Electrostatic Control. *J. Am. Chem. Soc.* **2009,** *131*, 4574-4575.

198. Cockroft, S. L.; Hunter, C. A., Desolvation and substituent effects in edge-to-face aromatic interactions. *Chem Commun (Camb)* **2009**, 3961-3963.

199. Wheeler, S. E., Local nature of substituent effects in stacking interactions. *J. Am. Chem. Soc.* **2011,** *133*, 10262-10274.

200. Watt, M.; Hardebeck, L. K. E.; Kirkpatrick, C. C.; Lewis, M., Face-to-Face Arene-Arene Bidning Energies: Dominated by Dispersion but Predicted by Electrostatic and Dispersion/Polarizability Substituent Constants. *J. Am. Chem. Soc.* **2011,** *133*, 3854-3862.

201. Muchowska, K. B.; Adam, C.; Mati, I. K.; Cockroft, S. L., Electrostatic modulation of aromatic rings via explicit solvation of substituents. *J. Am. Chem. Soc.* **2013,** *135*, 9976-9979.

202. Wheeler, S. E.; Bloom, J. W., Toward a more complete understanding of noncovalent interactions involving aromatic rings. *J. Phys. Chem. A* **2014,** *118*, 6133-6147.

203. Hwang, J.; Li, P.; Carroll, W. R.; Smith, M. D.; Pellechia, P. J.; Shimizu, K. D., Additivity of substituent effects in aromatic stacking interactions. *J. Am. Chem. Soc.* **2014,** *136*, 14060-14067.

204. Mignon, P.; Loverix, S.; De Proft, F.; Geerlings, P., Influence of stacking on hydrogen bonding: Quantum chemical study on pyridine-benzene model complexes. *J. Phys. Chem. A* **2004,** *108*, 6038-6044.

205. Piacenza, M.; Grimme, S., Van der Waals interactions in aromatic systems: structure and energetics of dimers and trimers of pyridine. *ChemPhysChem* **2005,** *6*, 1554-1558.

206. Swart, M.; van der Wijst, T.; Fonseca Guerra, C.; Bickelhaupt, F. M., Pi-pi stacking tackled with density functional theory. *J. Mol. Model.* **2007,** *13*, 1245-1257.

207. Itahara, T.; Imaizumi, K., Role of nitrogen atom in aromatic stacking. *J. Phys. Chem. B* **2007,** *111*, 2025-2032.

208. Hohenstein, E. G.; Sherrill, C. D., Effects of heteroatoms on aromatic pi-pi interactions: benzene-pyridine and pyridine dimer. *J. Phys. Chem. A* **2009,** *113*, 878-886.

209. Geng, Y.; Takatani, T.; Hohenstein, E. G.; Sherrill, C. D., Accurately characterizing the pi-pi interaction energies of indole-benzene complexes. *J. Phys. Chem. A* **2010,** *114*, 3576-3582.

210. Karthikeyan, S.; Nagase, S., Origins of the stability of imidazole-imidazole, benzene-imidazole, and benzene-indole dimers: CCSD(T)/CBS and SAPT calculations. *J. Phys. Chem. A* **2012,** *116*, 1694-1700.

211. Li, P.; Zhao, C.; Smith, M. D.; Shimizu, K. D., Comprehensive experimental study of N-heterocyclic pi-stacking interactions of neutral and cationic pyridines. *J. Org. Chem.* **2013,** *78*, 5303-5313.

212. Jacobs, M.; Greff Da Silveira, L.; Prampolini, G.; Livotto, P. R.; Cacelli, I., Interaction Energy Landscapes of Aromatic Heterocycles through a Reliable yet Affordable Computational Approach. *J. Chem. Theory Comput.* **2018,** *14*, 543-556.

213. Gung, B. W.; Wekesa, F.; Barnes, C. L., Stacking interactions between nitrogen-containing six-membered heterocyclic aromatic rings and substituted benzene: studies in solution and in the solid state. *J. Org. Chem.* **2008,** *73*, 1803-1808.

214. Xu, W.; Chan, K. M.; Kool, E. T., Fluorescent nucleobases as tools for studying DNA and RNA. *Nat Chem* **2017,** *9*, 1043-1055.

215. Khakshoor, O.; Wheeler, S. E.; Houk, K. N.; Kool, E. T., Measurement and theory of hydrogen bonding contribution to isosteric DNA base pairs. *J. Am. Chem. Soc.* **2012,** *134*, 3154-3163.

216. Zhang, Y.; Ptacin, J. L.; Fischer, E. C.; Aerni, H. R.; Caffaro, C. E.; San Jose, K.; Feldman, A. W.; Turner, C. R.; Romesberg, F. E., A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **2017,** *551*, 644-647.

217. Jeziorski, B.; Moszynski, R.; Szalewicz, K., Perturbation-Theory Approach to Intermolecular Potential-Energy Surfaces of Van-Der-Waals Complexes. *Chem. Rev.* **1994,** *94*, 1887-1930.

218. Szalewicz, K., Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2012,** *2*, 254-272.

219. Hohenstein, E. G.; Sherrill, C. D., Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2010,** *133*, 014101.

220. Hohenstein, E. G.; Sherrill, C. D., Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and

application to probe the nature of pi-pi interactions in linear acenes. *J. Chem. Phys.* **2010,** *132*, 184111.

221. Bloom, J. W.; Raju, R. K.; Wheeler, S. E., Physical Nature of Substituent Effects in XH/pi Interactions. *J. Chem. Theory Comput.* **2012,** *8*, 3167-3174.

222. Raju, R. K.; Bloom, J. W.; Wheeler, S. E., Broad Transferability of Substituent Effects in pi-Stacking Interactions Provides New Insights into Their Origin. *J Chem Theory Comput* **2013,** *9*, 3479-3490.

223. Kruse, H.; Banas, P.; Sponer, J., Investigations of Stacked DNA Base-Pair Steps: Highly Accurate Stacking Interaction Energies, Energy Decomposition, and Many-Body Stacking Effects. *J Chem Theory Comput* **2019,** *15*, 95-115.

224. Malyshev, D. A.; Romesberg, F. E., The expanded genetic alphabet. *Angew Chem Int Ed Engl* **2015,** *54*, 11930-11944.

225. Hoshika, S.; Leal, N. A.; Kim, M. J.; Kim, M. S.; Karalkar, N. B.; Kim, H. J.; Bates, A. M.; Watkins, N. E., Jr.; SantaLucia, H. A.; Meyer, A. J.; DasGupta, S.; Piccirilli, J. A.; Ellington, A. D.; SantaLucia, J., Jr.; Georgiadis, M. M.; Benner, S. A., Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* **2019,** *363*, 884-887.

226. Bootsma, A. N.; Wheeler, S. E., Converting SMILES to Stacking Interaction Energies. *J. Chem. Inf. Model.* **2019**.

227. Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D., Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *The Journal of Chemical Physics* **2014,** *140*, 094106.

228. Papajak, E.; Truhlar, D. G., Convergent Partially Augmented Basis Sets for Post-Hartree−Fock Calculations of Molecular Properties and Reaction Barrier Heights. *J. Chem. Theory Comput.* **2011,** *7*, 10-18.

229. Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D., PSI4: an open-source ab initio electronic structure program. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2012,** *2*, 556-565.

230. Bondi, A., Van Der Waals Volumes + Radii. *J. Phys. Chem.* **1964,** *68*, 441-+.

231. Plavec, J.; Lenarcic Zivkovic, M.; Rozman, J., Adenine-driven structural switch from two- to three-quartet DNA G-quadruplex. *Angewandte Chemie (International ed. in English)* **2018**.

232. Drew, H. R.; Dickerson, R. E., Conformation and dynamics in a Z-DNA tetramer. *Journal of molecular biology* **1981,** *152*, 723-736.

APPENDIX A

SUPPLEMENTARY INFORMATION RELATED TO BETTER SENSING THROUGH STACKING: THE ROLE OF NON-COVALENT INTERACTIONS IN GUANINE-BINDING SENSORS

Table A-1. Incremental focal point tables for interaction energies and binding energies (in kcal mol^{-1}) for complexes of 9-methyl-guanine (G) with **1** and 1-methyl-cytosine (C).[a]

Interaction Energies

C···G	**df-HF**	**+δ(SL-MP2)**	**+δ(DLPNO-CCSD)**	**+δ[DLPNO-CCSD(T)]**	**=DLPNO-CCSD(T)**
cc-pVDZ	-29.67	-5.85	2.59	-0.35	[-33.28]
cc-pVTZ	-26.43	-7.28	2.21	-0.45	[-31.94]
cc-pVQZ	-25.64	-7.48	2.06	-0.43	[-31.48]
CBS limit	[-25.39]	[-7.63]	[+1.96]	[-0.41]	**[-31.47]**
1···G					
cc-pVDZ	-25.63	-6.60	1.67	-0.37	[-30.92]
cc-pVTZ	-22.34	-7.91	1.26	-0.40	[-29.39]
cc-pVQZ	-21.57	-8.18	1.11	-0.43	[-29.07]
CBS limit	[-21.34]	[-8.37]	[+1.01]	[-0.45]	**[-29.16]**

Binding Energies

C···G	**df-HF**	**+δ(SL-MP2)**	**+δ(DLPNO-CCSD)**	**+δ[DLPNO-CCSD(T)]**	**=DLPNO-CCSD(T)**
cc-pVDZ	-24.83	-7.19	3.20	-0.42	[-29.25]
cc-pVTZ	-21.65	-8.63	2.82	-0.55	[-28.01]
cc-pVQZ	-20.92	-8.88	2.67	-0.61	[-27.73]
CBS limit	[-20.71]	[-9.06]	[+2.56]	[-0.65]	**[-27.85]**
1···G					
cc-pVDZ	-21.51	-8.03	2.15	-0.51	[-27.91]
cc-pVTZ	-18.20	-9.57	1.83	-0.61	[-26.55]
cc-pVQZ	-17.46	-9.86	1.69	-0.68	[-26.31]
CBS limit	[-17.25]	[-10.07]	[+1.58]	[-0.73]	**[-26.48]**

[a] Numbers in brackets are the results of basis set extrapolations.

Table A-2. Gas-phase binding energies (in kcal mol^{-1}) for complexes of 9-methyl-guanine (G) with **1** and 1-methyl-cytosine (C), as well as the difference between the two, computed at M06-2X/6-31G(d) optimized geometries.

	1···G	**C···G**	**Diff.**
M06-2X/6-311+G(d)	–26.9	–28.4	–1.5
DLPNO-CCSD(T)/CBS[a]	–26.5	–27.9	–1.5

[a] DLPNO-CCSD(T) energies extrapolated to the complete basis set limit.

Table A-3. IEFPCM M06-2X/6-311+G(*d*)//6-31G(*d*) solution-phase (dichloromethane) binding energies of 1-methyl-cytosine and sensors **1**, **4**-**10** with 9-methyl-guanine in kcal mol^{-1}.

	E_{bind}		E_{bind}
C	-18.2	**9b**	-25.6
1	-17.6	**9c**	-28.2
4a	-24.0	**9d**	-26.0
5a	-25.4	**10a**	-25.4
5b	-23.9	**10b**	-27.5
6a	-24.4	**10c**	-25.5
6b	-26.5	**10d**	-26.6
6c	-25.0	**10e**	-27.2
7a	-26.6	**10f**	-28.0
8a	-26.0	**10g**	-22.6
8b	-27.4	**10h**	-26.2
8c	-26.0	**10i**	-26.9
9a	-27.4	**10j**	-27.9

Table A-4. Bond distances N···H-N hydrogen bonds between 1-methyl-guanine and sensors **4**-**9** and **10a**-**10j** and their respective correlation to gas phase interaction energies and M4 interaction energies.

Sensor	r1	r2	r3
4a	3.338	2.201	2.118
5a	3.286	2.172	2.144
6b	2.446	2.051	2.154
7a	3.232	2.125	2.204
8b	2.529	2.067	2.195
9c	2.567	2.085	2.208
10f	2.310	1.901	-
Correl. with E_{int} (r^2)	0.97	0.81	0.28
Correl. with E_{M4} (r^2)	0.92	0.82	0.34
Sensor	r1	r2	r3
10a	3.219	2.109	2.314
10b	3.216	2.172	2.300
10c	3.138	2.065	2.208
10d	2.954	2.681	1.895
10e	3.107	2.122	2.140
10f	2.310	1.901	-
10g	2.966	2.648	2.174
10h	3.275	2.152	2.230
10i	3.092	1.930	1.988
10j	2.552	2.069	2.198
Correl. with E_{int} (r^2)	0.49	0.36	0.57
Correl. with E_{M4} (r^2)	0.18	0.61	0.25

APPENDIX B

SUPPLEMENTARY INFORMATION RELATED TO IMPORTANCE OF MODEL SIZE IN QUANTUM MECHANICAL STUDIES OF DNA INTERCALATION

Table B-1. Details on DNA intercalator systems **1-12**.

PDB ID	Deposition Month-Year	Resolution (Å)	Sequence	Number of Base Pairs	Ligand Name	Ligand Smiles	Ligand: Duplex Stoicheometry
110D	Jan-93	1.9	C▪GGCC▪G	6	Daunomycin	COc1cccc2C(=O)c3c(O)c4C[C@](O)(C[C@H](O[C@H]5C[C@H](N)[C@H](O)[C@H](C)O5)c4c(O)c3C(=O)c12)C(C)=O	2:1
151D	Dec-93	1.4	C▪GATC▪G	6	Doxorubicin	COc1cccc2C(=O)c3c(O)c4C[C@](O)(C[C@H](O[C@H]5C[C@H](N)[C@H](O)[C@H](C)O5)c4c(O)c3C(=O)c12)C(=O)CO	2:1
1DA9	Jan-93	1.7	T▪GGCC▪A	6	Doxorubicin	COc1cccc2C(=O)c3c(O)c4C[C@](O)(C[C@H](O[C@H]5C[C@H](N)[C@H](O)[C@H](C)O5)c4c(O)c3C(=O)c12)C(=O)CO	2:1
1DL8	Dec-99	1.55	▪C▪GTAC▪G▪	6	5-Fluoro-9-Amino-(N-(2-Dimethylamino)Ethyl)Acridine-4-Carboxamide	CN(C)CCNC(=O)c1cccc2c(N)c3ccc(F)c3nc12	4:1
1K9G	Oct-01	1.4	▪C▪CTAG▪G▪	6	Cryptolepine	Cn1c2c(cc3ccccc13)nc1ccccc21	4:1
1NAB	Nov-02	2.15	C▪GATC▪G	6	7-[5-(4-Amino-5-Hydroxy-6-Methyl-Tetrahydro-Pyran-2-Yloxy)-4-Hydroxy-6-Methyl-Tetrahydro-Pyran-2-Yloxy]-6,9,11-Trihydroxy-9-(2-Hydroxy-Acetyl)-7,8,9,10-Tetrahydro-Naphthacene-5,12-Dione	C[C@@H]1O[C@H](C[C@H](N)[C@@H]1O)O[C@@H]1[C@H](C)O[C@H](C[C@@H]1O)O[C@H]1C[C@@](O)(Cc2c(O)c3C(=O)c4ccccc4C(=O)c3c(O)c12)C(=O)CO	2:1
1VTH	Mar-92	1.6	T▪GTAC▪A	6	Daunomycin	COc1cccc2C(=O)c3c(O)c4C[C@](O)(C[C@H](O[C@H]5C[C@H](N)[C@H](O)[C@H](C)O5)c4c(O)c3C(=O)c12)C(C)=O	2:1
1Z3F	Mar-05	1.5	C▪GATC▪G	6	Ellipticine	Cc1c2[nH]c3ccccc3c2c(C)c2c[nH+]ccc12	2:1
386D	Mar-98	1.8	C▪GTAC▪G	6	N,N-Bis(3-Aminopropyl)-2-Anthraquinonesulfonamide	[NH3+]CCCN(CCC[NH3+])S(=O)(=O)c1ccc2C(=O)c3ccccc3C(=O)c2c1	2:1
3FT	Jan-09	1.12	C▪GATC▪	6	Proflavine	Nc1ccc2cc3ccc(N)cc3nc2c1	2:1

6			G				
465D	Apr-99	1.6	C▪GTAC▪G	6	9-Amino-(N-(2-Dimethylamino)Ethyl)Acridine-4-Carboxamide	CN(C)CCNC(=O)c1cccc2c(N)c3ccccc3nc12	2:1
4L5K	Jun-13	2.71	C▪GATC▪G	6	Coptisine	C1Oc2cc3CC[n+]4cc5c6OCOc6ccc5cc4-c3cc2O1	2:1

▪ indicates ligand binding site

Table B-2. Incremental focal point tables for **1**-**3** and $\mathbf{3}^{+}$ with Models A and B in kcal mol^{-1}.

	HF	+δMP2	+δDLPNO-CCSD	+δDLPNO-CCSD(T)	NET
			1A		
cc-pVDZ	9.82	-69.77	16.05	-2.54	[-46.44]
cc-pVTZ	17.18	-80.98	18.68	[-2.54]	[-47.67]
cc-pVQZ	20.54	-80.71	[+18.68]	[-2.54]	[-44.04]
CBS LIMIT	[+22.39]	[-80.52]	[+18.68]	[-2.54]	[-41.99]
			1B		
cc-pVDZ	10.9	-79.06	18.75	-2.76	[-52.17]
cc-pVTZ	20.28	-90.72	20.81	[-2.76]	[-52.39]
cc-pVQZ	24.15	-90.06	[+20.81]	[-2.76]	[-47.85]
CBS LIMIT	[+26.08]	[-89.57]	[+20.81]	[-2.76]	[-45.44]
			2A		
cc-pVDZ	10.43	-70.48	15.04	-2.46	[-47.47]
cc-pVTZ	19.94	-82.18	17.38	[-2.46]	[-47.32]
cc-pVQZ	24.37	-81.65	[+17.38]	[-2.46]	[-42.36]
CBS LIMIT	[+26.82]	[-81.26]	[+17.38]	[-2.46]	[-39.52]
			2B		
cc-pVDZ	9.32	-78.59	17.34	-2.66	[-54.59]
cc-pVTZ	20.61	-91.16	19.43	[-2.66]	[-53.77]
cc-pVQZ	25.42	-90.42	[+19.43]	[-2.66]	[-48.22]
CBS LIMIT	[+27.85]	[-89.87]	[+19.43]	[-2.66]	[-45.25]
			3A		
cc-pVDZ	-0.08	-62.92	14.24	-2.13	[-50.89]
cc-pVTZ	6.23	-73.83	16.31	[-2.13]	[-53.42]
cc-pVQZ	9.48	-73.51	[+16.31]	[-2.13]	[-49.85]
CBS LIMIT	[+11.36]	[-73.28]	[+16.31]	[-2.13]	[-47.74]
			3B		
cc-pVDZ	-11.67	-72.21	17.11	-2.35	[-69.12]
cc-pVTZ	-2.86	-83.18	18.55	[-2.35]	[-69.84]
cc-pVQZ	1.05	-82.39	[+18.55]	[-2.35]	[-65.13]
CBS LIMIT	[+3.06]	[-81.81]	[+18.55]	[-2.35]	[-62.55]

			3^{+}A		
cc-pVDZ	14.2	-63.45	14.26	-2.14	[-37.14]
cc-pVTZ	20.21	-74.42	16.52	[-2.14]	[-39.84]
cc-pVQZ	23.33	-74.24	[+16.52]	[-2.14]	[-36.53]
CBS LIMIT	[+25.15]	[-74.10]	[+16.52]	[-2.14]	[-34.57]
			3^{+}B		
cc-pVDZ	13.78	-73.76	17.28	-2.27	[-44.96]
cc-pVTZ	22.1	-84.89	19.09	[-2.27]	[-45.97]
cc-pVQZ	25.74	-84.27	[+19.09]	[-2.27]	[-41.71]
CBS	[+27.61]	[-83.82]	[+19.09]	[-2.27]	[-39.39]

Table B-3. Percent error of the DFT-predicted interaction energies compared to the ab initio benchmarks using Model B

Density Functional	Basis Set	1B	2B	3B	3^{+} B
B3LYP-D3	6-31G(d)	26	33	31	21
B3LYP-D3	6-31+G(d)	7	9	11	7
B3LYP-D3	def2TZVP	1	1	2	2
B3LYP-D3	cc-pVTZ	7	11	11	8
B97D	6-31G(d)	25	30	31	21
B97D	6-31+G(d)	5	5	10	7
B97D	def2TZVP	2	1	6	4
B97D	cc-pVTZ	9	11	14	10
M06-2X	6-31G(d)	1	6	2	0
M06-2X	6-31+G(d)	1	5	1	0
M06-2X	def2TZVP	16	14	20	12
M06-2X	cc-pVTZ	9	5	11	6
wB97X-D	6-31G(d)	25	30	31	22
wB97X-D	6-31+G(d)	18	20	24	17
wB97X-D	def2TZVP	6	7	11	8
wB97X-D	cc-pVTZ	12	16	19	13
B3LYP	6-31G(d)	102	101	111	68
B3LYP	6-31+G(d)	120	124	132	82
B3LYP	def2TZVP	128	132	140	87
B3LYP	cc-pVTZ	120	123	132	82

Table B-4. Percent error of the RI-MP2 interaction energies compared to the ab initio benchmarks using Model B

		1B	2B	3B	3^{+} B
RI-MP2	cc-pVDZ	50	53	52	34
RI-MP2	cc-pVTZ	55	56	59	38

RI-MP2	cc-pVQZ	45	44	49	30

Table B-5. Percent error of the DFT-predicted interaction energies compared to the ab initio benchmarks using Model A

Density Functional	Basis Set	1A	2A	3A	3^{+}A
B3LYP-D3	6-31G(d)	19	32	23	18
B3LYP-D3	6-31+G(d)	3	7	5	5
B3LYP-D3	def2TZVP	2	1	0	0
B3LYP-D3	cc-pVTZ	4	11	7	6
B97D	6-31G(d)	14	24	19	16
B97D	6-31+G(d)	3	3	0	1
B97D	def2TZVP	4	4	1	0
B97D	cc-pVTZ	2	7	7	6
M06-2X	6-31G(d)	3	7	8	5
M06-2X	6-31+G(d)	1	5	4	2
M06-2X	def2TZVP	16	13	22	16
M06-2X	cc-pVTZ	10	3	15	10
wB97X-D	6-31G(d)	18	27	23	18
wB97X-D	6-31+G(d)	13	16	18	15
wB97X-D	def2TZVP	2	5	7	6
wB97X-D	cc-pVTZ	8	14	14	11

Table B-6. Interaction energies (kcal/mol) of all intercalator using all models at 15 different DFT levels of theory and the HF-3c and PBEh-3c methods

	A	B	C	D	E	F
B3LYP-D3/6-31G(d)						
1	-50.0	-57.1	-58.1	-58.0	-57.8	-58.3
2	-52.2	-60.1	-61.2	-61.0	-61.6	-61.3
3	-42.6	-51.7	-52.7	-53.0	-53.0	-52.8
4	-44.0	-57.5	-58.2	-58.7	-58.7	-58.6
5	-51.1	-60.5	-61.0	-61.3	-61.6	-61.6
6	-51.8	-59.9	-60.0	-60.6	-61.2	-61.3
7	-58.5	-82.7	-91.7	-99.6	-100.0	-102.1
8	-51.0	-64.6	-70.8	-75.8	-76.7	-77.3
9	-67.4	-83.0	-84.5	-89.6	-93.6	-93.7
10	-66.6	-77.3	-81.1	-84.1	-84.6	-84.5
11	-67.1	-60.7	-67.5	-71.4	-72.2	-76.2
12	-45.8	-54.4	-57.2	-59.4	-60.2	-62.3

2+	-86.3	-102.7	-106.8	-109.1	-110.7	-112.3
3+	-56.4	-75.7	-76.2	-77.5	-78.7	-79.8
4+	-60.0	-82.0	-83.9	-85.3	-85.9	-87.4
5+	-64.3	-83.1	-86.5	-90.0	-91.5	-94.2
6+	-90.5	-102.6	-106.5	-108.6	-110.0	-110.4
B3LYP-D3/6-31+G(d)						
1	-43.3	-48.6	-50.1	-50.1	-49.8	-50.4
2	-42.2	-49.5	-51.0	-51.1	-51.5	-51.5
3	-36.4	-43.6	-45.3	-45.6	-45.5	-45.4
4	-38.4	-49.3	-50.5	-51.3	-51.3	-51.0
5	-44.2	-52.9	-54.0	-54.6	-54.9	-55.1
6	-42.5	-50.2	-51.1	-51.8	10.0	-52.5
7	-47.0	-69.3	-77.0	-84.5	-85.2	-87.5
8	-40.3	-53.3	-59.0	-63.5	-64.8	-65.8
9	-56.1	-70.1	-72.2	-76.8	-81.3	-81.6
10	-55.7	-65.2	-68.6	-71.3	0.7	-72.2
11	-54.6	-47.0	-53.0	-56.7	-57.5	-62.1
12	-40.5	-48.5	-51.9	-54.3	-55.3	-57.5
2+	-74.9	-90.5	-95.0	-97.6	-98.8	-100.7
3+	-50.0	-67.2	-68.3	-69.8	-71.0	-72.2
4+	-54.5	-73.8	-76.3	-78.0	-78.7	-80.1
5+	-57.6	-75.8	-80.1	-84.1	-85.5	-88.5
6+	-79.9	-91.5	-95.9	-98.3	-99.6	-63.0
B3LYP-D3/cc-pVTZ						
1	-43.8	-48.5	-49.2	-49.0	-48.8	-49.3
2	-44.0	-50.1	-50.7	-50.6	-51.1	-50.9
3	-37.2	-43.6	-44.4	-44.6	-44.5	-44.4
4	-38.6	-48.9	-49.5	-49.9	-49.9	-49.7
5	-44.4	-52.1	-52.4	-52.7	-53.0	-53.0
6	-44.3	-50.6	-50.6	-51.1	-51.6	-51.7
7	-47.2	-67.2	-74.1	-80.8	-81.2	-83.0
8	-40.1	-51.4	-56.6	-60.6	-61.5	-62.1
9	-57.1	-69.7	-71.0	-75.6	-79.5	-79.7
10	-55.9	-63.5	-65.9	-68.5	-68.9	-68.9
11	-55.5	-46.4	-51.7	-54.9	-55.6	-59.6
12	-41.1	-48.8	-51.3	-53.6	-54.4	-56.6
2+	-76.9	-91.1	-94.6	-97.0	-98.3	-100.0
3+	-50.8	-67.3	-67.5	-68.9	-70.1	-71.2

4+	-54.2	-73.0	-74.7	-76.2	-76.8	-78.2
5+	-58.3	-75.6	-78.9	-82.6	-84.1	-86.9
6+	-81.0	-91.5	-95.0	-97.2	-98.4	-99.1
B97-D/6-31G(d)						
1	-49.4	-56.8	-58.3	-58.3	-58.2	-58.6
2	-50.4	-58.7	-60.6	-60.5	-61.4	-61.2
3	-42.1	-51.7	-53.2	-53.6	-53.6	-53.6
4	-42.1	-55.5	-56.5	-57.1	-57.2	-57.1
5	-49.3	-59.0	-59.9	-60.2	-60.6	-60.7
6	-50.0	-58.7	-59.6	-60.2	-61.1	-61.3
7	-57.9	-83.0	-93.4	-101.3	-102.3	-104.7
8	-50.6	-65.3	-73.4	-78.6	-79.4	-80.1
9	-66.8	-84.6	-87.5	-92.3	-96.4	-96.7
10	-66.5	-76.7	-81.9	-85.2	-86.0	-86.0
11	-66.1	-59.6	-67.3	-72.2	-73.4	-77.6
12	-47.1	-56.4	-59.7	-62.0	-62.9	-65.0
2+	-84.0	-100.5	-105.3	-107.6	-109.5	-111.0
3+	-56.3	-75.9	-76.9	-78.2	-79.4	-80.6
4+	-58.3	-80.1	-82.3	-83.8	-84.4	-85.9
5+	-62.8	-81.7	-85.4	-88.9	-90.4	-93.0
	-88.0	-100.5	-105.0	-107.1	-108.9	-109.3
B97-D/6-31+G(d)						
1	-42.4	-47.9	-49.9	-49.9	-49.7	-50.3
2	-39.8	-47.4	-49.6	-49.7	-50.4	-50.5
3	-35.5	-43.2	-45.4	-45.7	-45.7	-45.6
4	-36.3	-47.0	-48.4	-49.2	-49.3	-49.1
5	-41.5	-50.3	-51.8	-52.4	-52.9	-53.0
6	-40.1	-48.1	-49.8	-50.5	-51.3	-51.7
7	-45.7	-68.4	-77.4	-84.9	-85.9	-88.6
8	-39.0	-52.8	-60.0	-64.7	-66.0	-67.0
9	-54.9	-70.4	-73.9	-78.1	-82.6	-83.0
10	-54.7	-63.4	-67.8	-71.0	-72.0	-72.2
11	-52.6	-44.6	-51.1	-55.8	-57.1	-61.8
12	-41.0	-49.5	-53.2	-55.8	-56.8	-59.1
2+	-71.7	-87.4	-92.3	-94.9	-96.4	-98.2
3+	-49.2	-66.7	-68.3	-69.7	-70.9	-72.1
4+	-52.4	-71.4	-74.1	-75.8	-76.5	-77.9
5+	-55.3	-73.6	-78.1	-81.9	-83.5	-86.3
6+	-76.5	-88.3	-93.3	-95.6	-97.2	-98.0

B97-D/def2-TZVP						
1	-41.8	-46.4	-47.6	-47.6	-47.4	-48.0
2	-39.5	-45.7	-47.2	-47.1	-48.0	-47.9
3	-35.2	-41.7	-43.1	-43.4	-43.4	-43.3
4	-35.4	-44.9	-45.8	-46.4	-46.5	-46.4
5	-40.5	-48.0	-48.9	-49.2	-49.7	-49.7
6	-40.2	-46.7	-47.6	-48.2	-49.0	-49.3
7	-42.9	-62.7	-70.8	-77.1	-78.0	-80.1
8	-36.3	-47.9	-54.5	-58.4	-59.3	-59.9
9	-53.0	-66.6	-69.2	-73.1	-77.0	-77.3
10	-52.5	-58.8	-62.3	-65.2	-65.8	-66.0
11	-51.1	-41.0	-46.9	-51.1	-52.2	-56.2
12	-40.1	-47.9	-51.1	-53.4	-54.4	-56.6
2+	-71.8	-85.8	-89.9	-92.2	-94.0	-95.6
3+	-48.8	-65.1	-66.0	-67.4	-68.6	-69.7
4+	-51.0	-68.8	-70.9	-72.5	-73.2	-74.5
5+	-54.4	-71.5	-75.2	-78.8	-80.4	-83.0
6+	-76.2	-86.7	-90.9	-93.1	-94.6	-95.4
B97-D/cc-pVTZ						
1	-44.6	-49.7	-50.9	-50.8	-50.7	-51.2
2	-43.7	-50.3	-51.7	-51.7	-52.6	-52.4
3	-38.0	-45.1	-46.3	-46.6	-46.6	-46.6
4	-38.1	-48.3	-49.2	-49.8	-49.8	-49.8
5	-43.6	-51.6	-52.4	-52.7	-53.1	-53.2
6	-44.1	-51.0	-51.8	-52.4	-53.2	-53.4
7	-48.0	-69.0	-77.5	-84.2	-85.1	-87.3
8	-40.7	-53.3	-60.3	-64.5	-65.3	-66.0
9	-57.6	-72.4	-75.1	-79.5	-83.4	-83.7
10	-57.1	-64.4	-68.2	-71.2	-71.8	-72.0
11	-55.9	-46.9	-53.1	-57.5	-58.6	-62.7
12	-42.7	-51.2	-54.2	-56.5	-57.4	-59.6
2+	-76.1	-90.5	-94.7	-97.0	-98.7	-100.3
3+	-51.8	-68.6	-69.4	-70.8	-72.0	-73.1
4+	-53.6	-72.2	-74.2	-75.8	-76.4	-77.8
5+	-57.8	-75.3	-78.9	-82.5	-84.0	-86.7
6+	-80.1	-91.0	-95.1	-97.3	-98.9	-99.6
M06-2X/6-31G(d)						
1	-40.9	-45.8	-46.3	-46.0	-45.8	-46.1

2	-42.4	-48.1	-48.0	-49.7	-47.9	-47.7
3	-31.8	-38.4	-39.0	-39.1	-39.0	-38.9
4	-35.7	-47.1	-47.3	-47.5	-47.6	-47.2
5	-38.4	-44.9	-44.6	-44.8	-45.0	-45.0
6	-41.9	-48.3	-47.2	-47.6	-47.8	-47.9
7	-46.5	-66.7	-72.7	-79.6	-79.7	-80.6
8	-36.8	-46.1	-49.2	-53.5	-54.3	-54.4
9	-51.7	-61.9	-61.4	-65.4	-69.2	-69.0
10	-50.7	-58.8	-60.6	-62.5	-62.9	-62.4
11	-51.7	-42.8	-46.3	-49.1	-49.7	-52.9
12	-34.4	-40.7	-43.0	-45.2	-46.0	-47.9
2+	-75.8	-90.2	-93.2	-95.5	-96.7	-98.4
3+	-45.4	-62.4	-62.6	-63.7	-64.8	-66.1
4+	-51.9	-71.6	-73.2	-74.4	-75.1	-76.4
5+	-51.4	-67.3	-70.2	-73.7	-75.0	-77.8
6+	-80.2	-90.7	-93.3	-95.4	-96.4	-96.9
M06-2X/6-31+G(d)						
1	-41.4	-45.7	-46.7	-46.5	-46.1	-46.6
2	-41.3	-47.5	-48.2	-48.2	-48.3	-48.2
3	-33.3	-39.1	-40.4	-40.5	-40.3	-40.2
4	-37.2	-47.6	-48.4	-48.8	-48.9	-48.4
5	-39.3	-46.1	-46.5	-46.9	-47.0	-47.2
6	-41.3	-48.1	-47.9	-48.5	-48.6	-48.9
7	-45.7	-65.7	-72.2	-79.1	-79.5	-80.9
8	-36.1	-46.1	-49.7	-53.9	-55.0	-55.5
9	-50.8	-61.2	-61.9	-65.7	0.0	0.0
10	-50.1	-58.6	-60.9	-62.8	-63.4	-63.2
11	-50.2	-41.6	-45.5	-48.5	-49.2	0.0
12	-34.6	-41.1	-43.9	-46.3	-47.2	-49.4
2+	-73.8	-88.5	-92.2	-94.7	-95.7	-97.6
3+	-46.7	-62.8	-63.6	-64.8	-65.9	-67.2
4+	-53.6	-72.2	-74.2	-75.6	-76.4	-77.6
5+	-52.4	-68.9	-72.6	-76.5	-77.8	-80.9
6+	-78.4	-89.4	-92.9	-95.2	-96.1	-96.9
M06-2X/def2-TZVP						
1	-35.2	-38.3	-38.6	-38.3	-38.0	-38.4
2	-34.5	-39.0	-38.9	-38.6	-38.9	-38.7
3	-26.9	-31.7	-32.2	-32.3	-32.1	-31.9

4	-30.9	-39.8	-40.1	-40.3	-40.3	-39.9
5	-32.9	-38.4	-38.1	-38.3	-38.4	-38.4
6	-34.8	-39.8	-38.8	-39.3	-39.5	-39.6
7	-36.3	-52.9	-57.8	-63.5	-63.5	-64.5
8	-27.4	-34.7	-37.5	-40.7	-41.5	-41.7
9	-42.9	-50.3	-50.0	-53.4	-57.2	-57.1
10	-41.5	-47.0	-47.8	-49.5	-49.6	-49.3
11	-42.0	-30.3	-33.0	-35.7	-36.2	-39.4
12	-29.9	-35.4	-37.5	-39.7	-40.5	-42.7
2+	-66.9	-79.7	-82.5	-84.8	-85.9	-87.7
3+	-40.1	-55.3	-55.2	-56.5	-57.6	-58.7
4+	-46.7	-64.0	-65.4	-66.6	-67.3	-68.5
5+	-46.5	-61.8	-64.7	-68.4	-69.7	-72.5
6+	-71.3	-80.6	-83.2	-85.3	-86.3	-87.1
M06-2X/cc-pVTZ						
1	-37.7	-41.4	-41.5	-41.3	-41.0	-41.4
2	-38.2	-43.1	-42.9	-42.6	-42.9	-42.7
3	-29.5	-34.9	-35.4	-35.4	-35.2	-34.9
4	-33.2	-43.0	-43.1	-43.4	-43.4	-43.0
5	-35.8	-41.8	-41.4	-41.6	-41.8	-41.8
6	-38.2	-43.5	-42.6	-43.0	-43.2	-43.3
7	-41.0	-58.9	-64.2	-70.2	-70.4	-71.4
8	-31.5	-39.8	-43.0	-46.5	-47.2	-47.3
9	-47.1	-55.8	-55.6	-59.4	-63.1	-63.0
10	-45.9	-52.3	-53.4	-55.2	-55.3	-55.0
11	-46.5	-36.1	-39.3	-42.0	-42.6	-45.8
12	-32.6	-38.6	-40.7	-42.9	-43.7	-45.8
2+	-70.7	-83.9	-86.6	-88.9	-90.0	-91.8
3+	-42.8	-58.6	-58.4	-59.7	-60.8	-61.8
4+	-48.9	-67.0	-68.3	-69.6	-70.3	-71.5
5+	-49.6	-65.4	-68.1	-71.8	-73.2	-76.0
6+	-74.7	-84.3	-86.9	-89.0	-90.0	-90.8
ωB97X-D/6-31G(d)						
1	-49.7	-56.8	-58.0	-58.1	-58.0	-58.4
2	-50.2	-58.8	-60.3	-59.9	-60.8	-60.5
3	-42.4	-51.6	-52.9	-53.2	-53.1	-53.2
4	-42.9	-56.5	-57.5	-58.0	-57.9	-57.7
5	-48.5	-58.0	-58.7	-58.8	-59.0	-59.4

6	-50.1	-59.0	-59.3	-59.9	-60.6	-60.7
7	-56.2	-81.6	-91.0	-99.0	-99.5	-101.4
8	-48.3	-62.4	-68.7	-74.4	-74.7	-75.8
9	-64.6	-80.4	-82.4	-87.4	-91.7	-92.0
10	-63.6	-74.2	-78.3	-81.5	-82.1	-82.1
11	-63.7	-57.7	-64.7	-68.7	-69.9	-74.0
12	-44.6	-53.4	-56.6	-59.1	-60.1	-62.1
2+	-84.2	-101.4	-106.0	-108.4	-110.2	-111.8
3+	-56.5	-76.1	-76.8	-78.3	-79.4	-80.8
4+	-59.8	-81.9	-84.2	-85.6	-86.3	-87.7
5+	-61.5	-80.5	-84.2	-87.6	-89.0	-92.1
6+	-88.9	-102.0	-106.2	-108.3	-109.9	-110.3
ωB97X-D/6-31+G(d)						
1	-47.4	-53.4	-55.1	-55.2	-55.0	-55.5
2	-45.8	-54.3	-56.4	-56.3	-57.0	-56.9
3	-40.8	-48.8	-50.7	-51.1	-50.9	-51.0
4	-41.6	-53.6	-55.0	-55.6	-55.6	-55.3
5	-46.3	-55.6	-56.9	0.0	-57.5	-57.9
6	-46.2	-55.1	-56.2	-56.9	-57.5	-57.7
7	-51.3	-75.8	-85.0	-92.9	0.0	-95.9
8	-43.7	-58.0	0.0	0.0	0.0	-72.0
9	-59.6	-74.9	-77.7	-82.5	-87.2	0.0
10	-59.1	-69.4	-73.7	-76.8	-77.6	-77.8
11	-58.1	-51.9	-58.7	-62.7	-64.1	0.0
12	-42.7	-51.4	-55.2	-57.8	-58.9	-61.1
2+	-78.7	-95.7	0.0	0.0	0.0	0.0
3+	-54.7	-72.9	-74.2	-75.8	0.0	-78.3
4+	-58.6	-78.9	-81.7	-83.3	-84.0	-85.3
5+	-59.5	-78.5	-83.0	-86.8	-88.2	-91.5
6+	-83.8	-96.9	-101.7	-104.1	-105.5	-106.2
ωB97X-D/def2-TZVP						
1	-42.9	-48.1	-49.1	-49.2	-49.1	-49.5
2	-41.4	-48.5	-49.9	-49.6	-50.5	-50.3
3	-36.8	-43.7	-44.9	-45.2	-45.0	-45.2
4	-37.4	-48.0	-49.0	-49.4	-49.3	-49.1
5	-41.6	-49.7	-50.3	-50.4	-50.5	-50.9
6	-42.2	-49.5	-50.0	-50.5	-51.2	-51.3
7	-44.6	-65.7	-73.8	-80.3	-80.7	-82.6

8	-37.5	-49.2	-54.8	-59.5	-59.7	-60.9
9	-54.3	-67.0	-68.9	-73.2	-77.4	-77.7
10	-53.2	-60.7	-63.6	-66.5	-67.0	-67.1
11	-52.5	-43.5	-49.6	-53.0	-54.2	-58.4
12	-39.3	-47.2	-50.3	-52.8	-53.8	-55.9
2+	-74.3	-89.6	-93.8	-96.2	-97.8	-99.6
3+	-50.5	-67.5	-68.2	-69.8	-70.9	-72.3
4+	-53.6	-72.7	-74.8	-76.3	-76.9	-78.2
5+	-55.1	-72.9	-76.6	-80.1	-81.4	-84.6
6+	-79.0	-90.7	-94.7	-96.9	-98.3	-99.1
ωB97X-D/cc-pVTZ						
1	-45.4	-51.0	-52.0	-52.1	-51.9	-52.4
2	-45.1	-52.6	-53.9	-53.6	-54.5	-54.2
3	-39.3	-46.7	-47.8	-48.1	-47.9	-48.1
4	-39.7	-51.0	-51.9	-52.3	-52.2	-52.0
5	-44.4	-52.9	-53.4	-53.4	-53.7	-54.0
6	-45.6	-53.2	-53.6	-54.1	-54.8	-54.9
7	-49.3	-71.5	-79.9	-86.7	-87.2	-89.1
8	-41.5	-54.2	-60.1	-65.1	-65.2	-66.3
9	-58.5	-72.3	-74.3	-79.0	-83.1	-83.4
10	-57.5	-65.8	-68.9	-72.0	-72.4	-72.5
11	-56.9	-48.9	-55.3	-58.9	-60.2	-64.3
12	-41.8	-50.1	-53.1	-55.6	-56.6	-58.6
2+	-78.1	-93.7	-97.9	-100.2	-101.9	-103.6
3+	-53.1	-70.6	-71.2	-72.8	-73.8	-75.2
4+	-55.8	-75.6	0.0	-79.1	-79.7	-81.1
5+	-58.1	-76.3	-79.8	-83.3	-84.7	-87.8
6+	-82.4	-94.3	-98.3	-100.5	-101.9	-102.7
HF-3c						
1	-40.3	-44.3	-45.2	-45.3	-45.3	-45.7
2	-35.7	-42.5	-43.8	-43.4	-44.0	-43.6
3	-32.4	-39.1	-40.2	-40.3	-40.4	-40.3
4	-34.8	-44.6	-45.4	-45.7	-45.7	-45.6
5	-38.8	-46.2	-47.4	-47.7	-48.0	-48.3
6	-37.0	-44.0	-44.4	-44.9	-45.4	-45.4
7	-37.5	-56.1	-62.8	-68.8	-69.1	-70.8
8	-32.5	-42.3	-46.6	-50.6	-51.6	-52.7
9	-47.7	-57.1	-58.2	-62.4	-66.4	-66.5
10	-45.6	-50.7	-53.6	-56.4	-56.8	-57.2

11	-44.9	-30.4	-34.6	-37.2	-38.1	-42.0
12	-34.6	-42.0	-45.0	-47.9	-48.7	-50.6
2+	-70.8	-85.3	-90.3	-92.9	-94.2	-96.0
3+	-44.5	-61.6	-62.0	-63.5	-64.9	-65.9
4+	-49.0	-67.3	-69.4	-71.0	-71.6	-73.0
5+	-49.3	-65.2	-69.5	-73.1	-74.5	-77.5
6+	-76.4	-87.7	-92.5	-94.8	-96.1	-96.7
PBEh-3c						
1	-35.9	-40.3	-40.9	-41.1	-40.8	-40.9
2	-34.3	-40.2	-40.2	-39.6	-40.3	-39.5
3	-26.9	-32.8	-34.1	-34.3	-34.2	-34.0
4	-30.6	-40.1	-41.0	-41.2	-41.1	-41.1
5	-35.5	-42.3	-42.7	-43.2	-43.2	-43.4
6	-34.5	-40.6	-40.0	-40.4	-41.1	-40.2
7	-34.8	-51.8	-58.0	-62.8	-62.5	-63.8
8	-28.1	-37.7	-41.7	-45.8	-46.0	-47.1
9	-46.4	-56.2	-56.9	-60.8	-64.9	-64.6
10	-43.2	-46.1	-47.6	-49.9	-50.4	-50.5
11	-42.4	-28.6	-33.0	-34.6	-35.2	-38.8
12	-34.8	-41.8	-44.7	-47.0	-47.9	-50.1
2+	-66.5	-81.2	-84.3	-86.6	-88.2	-89.4
3+	-42.0	-58.4	-59.1	-60.7	-62.0	-63.0
4+	-48.1	-66.4	-68.6	-69.9	-70.7	-72.2
5+	-48.7	-65.2	-68.5	-72.5	-73.7	-76.7
6+	-71.8	-82.2	-85.2	-87.5	-89.1	-88.7

APPENDIX C

SUPPLEMENTARY INFORMATION RELATED INTRINSIC STACKING INTERACTIONS OF NATURAL AND ARTIFICIAL NUCLEOBASE

Table C-1. SAPT/jun-cc-pVDZ results for strongest interacting dimers: Total interaction Energy (E_{SAPT}), Electrostatic Energy (E_{Elec}), Exchange Energy (E_{Exch}), Induction Energy (E_{Ind}), Dispersion Energy (E_{Disp})

Dimer	E_{SAPT}	E_{Elec}	E_{Exch}	E_{Ind}	E_{Disp}
A|A	-10.6	-7.9	12.3	-1.4	-13.6
A|C	-13.4	-11.7	14.1	-2.4	-13.3
A|G	-14.4	-11.7	13.9	-2.4	-14.1
A|T	-11.0	-8.5	12.3	-1.6	-13.2
A|U	-10.6	-8.8	11.8	-1.8	-11.8
A|2-AP	-10.8	-8.0	12.0	-1.4	-13.4
A|F	-8.6	-6.3	12.0	-1.1	-13.2
A|dTPT3	-14.3	-10.5	14.4	-2.2	-16.0
A|4-TR	-12.3	-8.8	13.0	-1.5	-14.9
A|N	-10.2	-6.9	13.7	-1.6	-15.5
A|Pu	-10.0	-7.5	11.8	-1.4	-12.9
A|Py	-7.9	-5.9	9.8	-1.0	-10.8
C|C	-12.2	-11.8	11.5	-2.8	-9.1
C|G	-16.2	-15.3	14.8	-3.5	-12.1
C|T	-12.2	-10.8	11.7	-2.3	-10.8
C|U	-11.9	-11.1	11.4	-2.3	-9.9
C|2-AP	-13.3	-12.0	13.9	-2.3	-13.0
C|F	-8.2	-6.9	10.6	-1.8	-10.1
C|dTPT3	-13.7	-11.5	14.4	-2.7	-13.9
C|4-TR	-12.6	-10.8	11.8	-2.1	-11.5
C|N	-8.9	-6.9	12.6	-1.8	-12.7
C|Pu	-11.9	-10.3	12.6	-2.0	-12.3
C|Py	-7.3	-5.9	8.4	-1.4	-8.5
G|G	-19.6	-16.9	16.6	-3.0	-16.3
G|T	-15.2	-13.0	13.8	-2.4	-13.6
G|U	-15.2	-13.9	14.2	-2.9	-12.5
G|2-AP	-15.6	-12.9	15.0	-2.5	-15.2
G|F	-10.4	-8.1	12.4	-2.1	-12.5
G|dTPT3	-17.9	-14.3	16.9	-3.4	-17.1
G|4-TR	-16.1	-12.9	15.3	-2.7	-15.8
G|N	-11.3	-8.8	13.7	-2.6	-13.7

G|Pu	-14.0	-10.8	13.9	-2.5	-14.6
G|Py	-9.8	-7.8	10.3	-1.4	-10.9
TT	-8.3	-6.8	10.2	-1.2	-10.4
TU	-8.4	-7.4	10.4	-1.5	-9.8
T|2-AP	-11.0	-8.9	12.6	-1.7	-12.9
T|F	-7.4	-6.0	11.5	-1.5	-11.3
T|dTPT3	-13.1	-10.6	14.0	-2.1	-14.4
T|4-TR	-10.2	-8.0	11.9	-1.5	-12.7
T|Pu	-9.8	-7.4	11.2	-1.6	-12.0
T|Py	-6.8	-5.9	9.2	-1.1	-9.0
T|N	-9.8	-7.5	13.2	-1.9	-13.5
U|U	-8.0	-7.4	9.5	-1.5	-8.7
U|2-AP	-10.7	-9.3	12.1	-1.8	-11.7
U|F	-6.7	-5.6	10.1	-1.2	-10.0
U|dTPT3	-12.9	-10.8	13.2	-2.1	-13.1
U|4-TR	-9.7	-7.7	11.4	-1.7	-11.8
U|Pu	-9.0	-7.1	10.5	-1.6	-10.9
U|Py	-6.1	-5.6	8.3	-1.0	-7.8
U|N	-9.6	-7.0	11.0	-1.8	-11.8

SAPT Analysis

SAPT enables the decomposition of interaction energies into electrostatic (elec.), dispersion (disp.), induction (ind.), and exchange-repulsion (exch.) components. To determine which of these components underlie the variations in stacking interactions, we compared the total interaction energies for all local stacked minima with the total interaction energies without each component. For example, the non-dispersion part of ESAPT is

$$E_{SAPT} - E_{disp} = E_{elec.} + E_{ind.} + E_{exch.}$$

The results of this analysis for the natural nucleobases are displayed in Figure C-1 and show that electrostatic interactions are the primary driver of variations in these stacking interactions. In particular, the plot shows that if electrostatic interactions are eliminated, the sum of the other three components is essentially constant. This is not true for the other components.

A similar analysis of the unique stacked dimers of the non-natural nucleobases with the natural nucleobases (Figure C-2) reveals that these interactions are also primarily driven by electrostatic effects.

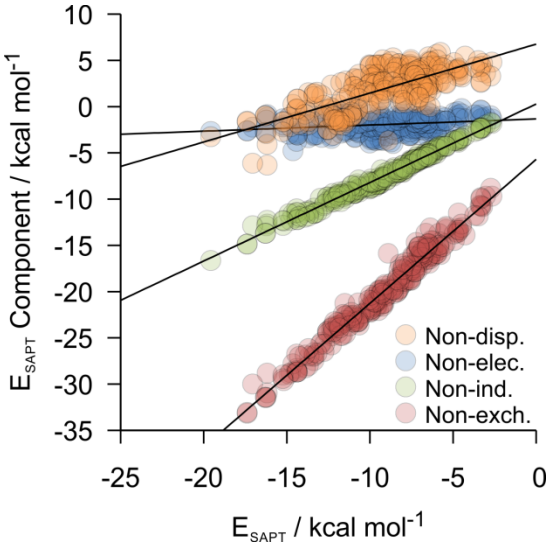

Figure C1. E_{SAPT} lacking individual components vs the total E_{SAPT} interaction energy for all unique stacked energy minima of the natural nucleobases.

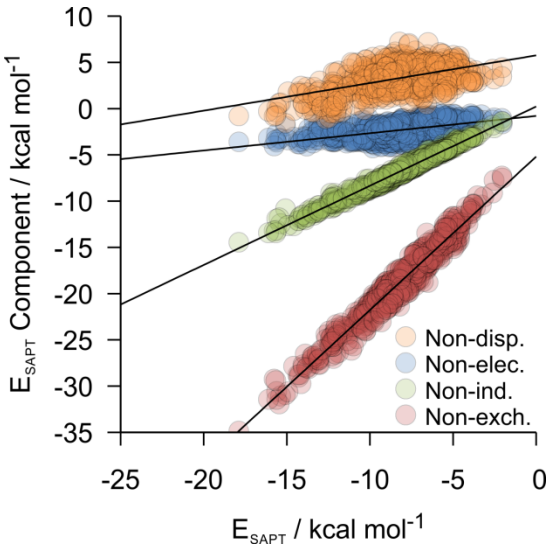

Figure C2. E_{SAPT} lacking individual components vs the total E_{SAPT} interaction energy for all unique stacked energy minima of the natural nucleobases.