# QUANTIFYING NON-COVALENT INTERACTIONS IN BIOLOGICAL SYSTEMS THROUGH COMPUTATIONAL CHEMISTRY

A Dissertation

by

YI AN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Steven E. Wheeler |
| Committee Members, | Robert R. Lucchese |
| | Christian Hilty |
| | Wonmuk Hwang |
| Head of Department, | Francois P. Gabbai |

December 2015

Major Subject: Chemistry

# ABSTRACT

Non-covalent interactions, especially those involving arenes, perform critical roles in biological processes, including ligand-protein and ligand-nucleic acid interactions, as well as protein and DNA/RNA stability and function. We investigated non-covalent interactions involving aromatic rings using computational chemistry in four contexts: sequence-selective DNA-binding antibodies, nitroarene-protein interactions, heterocycle-adenine interactions, and the tuning of amino acid $pK_a$'s. We first surveyed recent advances in our understanding of alkali metal cation-π interactions and their role in both experimental and theoretical fields. Then quantum mechanical computations and classical molecular mechanics based molecular dynamics simulations were utilized as the main tools to explore non-covalent interactions involving π-system in biological systems.

We present computational analyses of the binding of four dinucleotides to a sequence-selective single-stranded DNA (ssDNA) binding antibody (ED-10) and selected point-mutants using MMGBSA as well as DFT applied to a cluster model of the binding site. The results indicate that the sequence selectivity arises primarily from differences in the strength of π-stacking and XH/π interactions with the surrounding aromatic residues; hydrogen bonds play little role.

Stacking interactions in nitroarene binding sites of proteins were studied through analyses of structures in the protein data bank (PDB), as well as DFT and *ab initio* computations applied to model systems. The results show that the interactions between

aromatic amino acids and nitroarenes are very strong, and the regiochemistry of the nitro substituents plays a significant role in the relative monomer orientations and strength of the interaction.

Next, complexes of 9-methyladenine with 46 heterocycles commonly found in drugs were studied using dispersion-corrected density function theory, revealing 268 unique stacked dimers. The predicted binding energies for each heterocycle span a broad range, highlighting the strong dependence of heterocycle stacking interactions on the relative orientation of the two arenes. The data reveal several key local, direct interactions that strongly influence the preferred orientation and strength of the most favorable stacked complexes.

Finally, the impact of π-stacking interactions on the $pK_a$ of ionizable amino acid side chains was investigated by DFT computations. π-stacking interactions, edge-to-face interactions, CH/π interactions, and NH/π interactions abound in these complexes, and impact the acidity of these ionizable groups.

# DEDICATION

I dedicate this dissertation to my high school math teacher Liu Yin. I wish I had learned more about vector applications from him.

# ACKNOWLEDGEMENTS

First, I would like to thank my committee chair Dr. Wheeler for all his help and advice on my research and study. Dr. Wheeler taught me every aspect of research without any micro-management. I would like to thank my committee members Dr. Lucchese, Dr. Hilty, and Dr. Hwang for their support and suggestions throughout the course of this research.

Second, I would like to thank my former co-workers Dr. Tongxiang Lu, Dr. Jacob W.G. Bloom, and Dr. Diana Sepulveda for their valuable discussions which often trigger my research ideas. I also would like to thank the department of chemistry and Texas A&M University for providing me with excellent research and study experience.

Finally, I would like to thank my parents and girlfriend for their encouragement during my years at Texas A&M University. My cousin Song An, who got his doctoral degree at Texas A&M three years ahead of me, is a role model in my entire life. I'm very grateful that he taught me the best way to deal with pressure and how to balance work and life.

# NOMENCLATURE

| | |
|---|---|
| ADMET | Absorption, Distribution, Metabolism, Excretion and Toxicity |
| AMP | Adenosine monophosphate |
| ARM | Antibody Recruiting Molecules |
| CABP | Community-Acquired Bacterial Pneumonia |
| CCSD(T) | Coupled Cluster Singles and Doubles with Perturbed Triple Excitation |
| DART | D-amino Acid Antibody Recruitment Therapy |
| DFT | Density Functional Theory |
| DNP | Dinitrophenol |
| MD | Molecular Dynamics |
| MM | Molecular Mechanics |
| MM/GBSA | Molecular Mechanics/Generalized Born Surface Area |
| MP2 | 2nd Order Møller-Plesset Perturbation Theory |
| PDB | Protein Data Bank |
| PSMA | Prostate Specific Membrane Antigen |
| RMSD | Root Mean Square Deviation |
| SAPT | Symmetry Adapted Perturbation Theory |
| ssDNA | single-stranded DNA |

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF CHARTS

Page

# CHAPTER I

# INTRODUCTION AND BACKGROUND

## Nature of Non-covalent Interaction

Macromolecular structure is determined and maintained by a combination of covalent and non-covalent interactions. Non-covalent interactions occur through a variety of mechanisms, including electrostatic, van der Walls, hydrogen bonding, and hydrophobic interactions. Electrostatic interactions, in the narrow sense, are the Coulombic interactions between charged species. Van der Waals interactions are weak interactions that occur between all molecules, and are dominated by London dispersion forces. These result from the interaction between instantaneous dipoles formed via transient distortion of electron clouds within a molecule. Molecules also interact with each other through interactions of permanent dipoles and dipole induced-dipole interactions. Finally, hydrogen bonds are formed between a hydrogen atom attached to an electronegative donor atom and a neighboring acceptor atom. Hydrogen bonding is a relatively strong and directional non-covalent interaction, with its magnitude and directionality determined primarily by the electrostatic forces between the donor hydrogen atom and the acceptor atom.

In this research, we focus on the impact of π-stacking and XH/π interactions on various biological systems. At the supramolecular level, aromatic rings can interact in three prototypical configurations: stacked, edge-to-face, and T-shaped (see Figure I-1). The non-covalent interaction we are primarily interested in is the π–stacking interaction.

Although on average π–stacking interactions are weaker than hydrogen bonding, π–stacking interactions play a critical role in many self-assembly and molecular recognition processes. XH/π interactions, in which the hydrogen of a NH or CH group is directed toward the face of an aromatic ring, also make important contributions to biomolecular structure and function. These weakly polar interactions can also play essential roles in macromolecular function. XH/$\pi$ interactions are best classified somewhere between conventional hydrogen bonds and weaker interactions dominated by dispersion, and are dependent on the orientation of the donor and acceptor group.

London dispersion interactions, which occur between all molecules, are attractive non-covalent interactions arising from instantaneous polarization due to unequal distributions of electrons. In general, the strength of dispersion interaction increases with molecular size, because larger molecules tend to have larger polarizabilities. For nonpolar molecules, dispersion is the dominant non-covalent interaction, much stronger than other Van der Waals forces. In this research, the significance of dispersion interaction will be discussed thoroughly, since aromatic systems in biological environment are quite nonpolar.

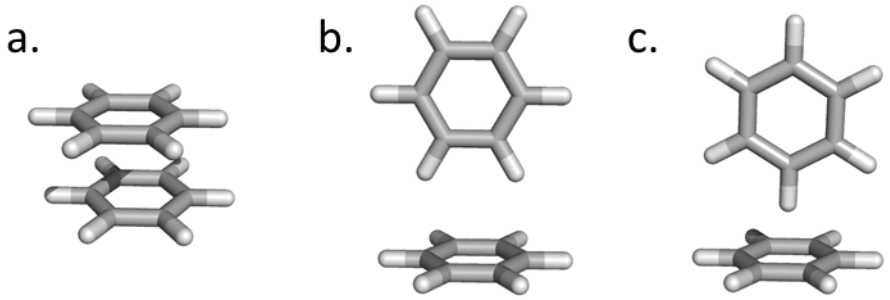



**Figure I-1.** Model aromatic interactions showing a. stacking, b. edge-to-face, c. T-shaped configurations**.**

**Non-covalent Interaction with Aromatic Rings in Biology**

Non-covalent interactions involving aromatic rings are critical for organization and assembly in supramolecular and biological chemistry.[1,2] They are indispensable for proteins and nucleic acids to adopt their correct three dimensional configurations and perform their functions.[3] These non-covalent forces also play a significant role in protein-DNA interactions, protein-ligand interactions, nucleic acid-ligand interactions, and interactions between amino acids both within a given protein and in protein-protein interactions.

Protein-DNA interactions participate in a variety of key biological processes, such as DNA replication, transcription, translation, and control of gene expression. For instance, proteins can bind both single-stranded DNA and double-stranded DNA,[4-7] and many DNA-binding antibodies show sequence-specific recognition dependent on various non-covalent interactions. Exploring non-covalent interactions between antibody and DNA and elucidating the general mechanism of base selectivity at the molecular level is crucial to further understanding mechanism of many auto-immune diseases.

Protein–ligand interactions are also vitally important in biological processes, including molecular recognition, conformational stability, signal transmission, regulation of enzyme activity, and regulation of gene replication.[8] Detailed structural analyses at the atomic level reveal that non-covalent interactions central for the binding of ligands by proteins.[9,10] Thus, understanding protein–ligand binding affinities, especially for drug-like small molecules, is central to pharmaceutical design.[11] Since computational approaches have become an essential part of the drug discovery process,[12,13] accurate computations of non-covalent interactions in biological contexts and detailed analyses of ligand binding sites are fundamentally important in medicinal chemistry.

A large number of therapeutic agents with antitumor, antibacterial, and antiviral activities perform their functions through noncovalent interaction between nucleic acids (both DNA and RNA) and small molecules.[14] For instance, DNA intercalating antitumor agents, such as distamycin, berenil, and actinomycin, bind the minor groove of double-stranded oligodeoxynucleotides non-covalently.[15] Sundry RNA molecules, including ribosomal RNA, tRNA, trans-activating response RNA, and rev response element RNA, are antibacterial and antiviral drug targets.[16-18] Thus, elucidating the specific mechanism of non-covalent interaction between small molecule and DNA/RNA is beneficial for rational drug design.

Finally, non-covalent interactions with aromatic amino acids are fundamentally crucial in protein stability and function.[19-23] For instance, π-stacking interactions may play a vital role in the molecular recognition and formation of amyloid fibrils self-assembly.[24] Aromatic residues not only engage in π-stacking interactions but also other

non-covalent interactions with aromatic amino acids.[25,26] For example, XH/π interactions are present in many dimers of aromatic amino acids. Although such interactions are ubiquitous in protein crystal structures,[27-30] their origin and significance in biomolecular process is still largely unsettled. A thorough understanding of how aromatic residues interact with nearby amino acids, and how XH/π interactions are involved, will shed light on our knowledge about protein-protein interaction.

**Computational Chemistry Methods in Biology**

Computational chemistry is a branch of chemistry that applies mathematical algorithms and statistical theories based on either classical- or quantum-mechanical treatment of molecular systems in order to determine the various properties of a chemical system, particularly the structure and energy of the species.[31,32] Modern computational chemistry finds increasing applications in areas spanning from chemical catalysis, novel material design, molecular spectroscopy, to solid-state chemistry, enzymatic reactions, and drug discovery, *etc*.[33-42]

One major field of computational chemistry is molecular mechanics (MM), which utilizes classical mechanics to model (typically large) molecules. In such simulations, the atoms are treated as masses and bonds treated as springs with empirically-determined force constants. The force-constants for all relevant bends, angles, dihedral angles, and non-covalent interactions between atoms (Lennard-Jones and Coulombic Interactions) are incorporated into a fixed functional form to find the potential energy of the system.[43-45] These force constants are optimized to fit

experimental results or highly accurate quantum mechanical computations. As the force field is built from such empirical parameters, calculations using force field methods are in general less accurate but faster compared to quantum mechanical calculations. Therefore, force field techniques are mainly applied in large systems where tens of thousands of atoms/molecules are involved, such as proteins, polymers, and other organic/inorganic materials.[46]

Based on molecular mechanics, molecular dynamics (MD) is the computational technique to simulate the classical motion of atoms as a function of time.[47,48] In MD simulations, trajectories are computed by numerically solving Newton equations of motion based on the forces given by the empirical force field. In biological systems, molecular dynamics simulation finds excellent applications in areas like protein and RNA folding pathways, binding modes and energies between ligand and protein/nucleic acid, and solvation free energies of biological macromolecules, etc.[13,49-53]

The concept of free energy is fundamental to chemistry, as free energy changes of molecular systems define their tendencies to associate and react. Therefore, being able to accurately predict the free energy change in various biological processes using MD simulation is advantageous. The main types of free energy computations applied to biological systems are solvation/transfer free energy, conformational free energy, and binding free energy. The two most conventional schemes to calculate free energy in biological context are thermodynamic integration and free energy perturbation theory.[54-56] Both methods rely on statistical mechanics and thermodynamic cycles to calculate the free energy difference between two end-states based on a series of (possibly fictitious)

intermediate states. In practice these two methods are very computationally expensive because the process of free energy change has to be divided into a sufficient number of simulation "windows" and the ensemble average needs to be evaluated for each of these windows. Alternative approaches that are less computational costly are "end-state methods". For a macromolecule-ligand complex, the binding free energy is defined as

$$\Delta G = G_{complex} - G_{macromolecule} - G_{ligand} \quad \text{(I-1)}$$

MM-PBSA is a popular end-state method focusing on implicit solvent models, where MM means this method is based on molecular mechanics, PB refers to the fact that in the original method the electrostatic energy was evaluated by Poisson–Boltzmann model, and SA means the surface area is used to calculate the non-polar energy in solution.[57] MM-GBSA is based on ensembles derived from molecular dynamics or Monte Carlo simulations. The MM-PBSA method integrated in the AMBER package[58] provides a variety of implicit solvent models, including the Poisson–Boltzmann, the Generalized Born, and the Reference Interaction Site models.[59] For more rigorous results, contributions to the entropy from solute vibrations may be computed using normal mode or quasi-harmonic analysis.

Accuracy is still a challenge for force field methods built on empirical parameters. For highly accurate results, quantum mechanical computations based on (approximately) solving the Schrödinger equation are necessary. The Schrödinger equation describes quantum mechanical behavior of physical systems, where the wavefunction defines the quantum state of the system. The wavefunction contains all information about the system but it does not have classical physical significance. The

square of the wave function has physical significance as a probability density. The time-dependent Schrödinger equation describes how the wavefunction changes with time,

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = \hat{H}\Psi(r,t) \tag{I-2}$$

where $\Psi(r,t)$ is the wavefunction and $\hat{H}$ is the Hamiltonian (total energy) operator. In this research we only focus on time-independent Schrödinger equation, which is the more commonly used form for computations in chemical and biological context. The general form of time-independent Schrödinger equation is:

$$\left[\frac{\hbar^2}{2\mu}\nabla^2 + \mathrm{V(r)}\right]\Psi(r) = E\Psi(r) \tag{I-3}$$

where $\mathrm{V(r)}$ is the potential energy and $\Psi(r)$ is the (time-independent) stationary state wavefunction. For systems with more than one electron, the Schrödinger equation cannot be solved analytically due to electron-electron interactions. Therefore, approximations have to be made to get desired wave functions and energies of the system.

In *ab initio* electronic structure methods, the electronic wavefunction is expressed as one or more anti-symmetrized linear combination of molecular orbitals (Slater determinants). These molecular orbitals are expressed as a linear combination of atom-centered Gaussian basis functions (the basis set).[60] In Hartree-Fock (HF) theory, the wavefunction is represented by a single Slater determinant, with the underlying molecular orbitals derived by minimizing the quantum mechanical energy (the variational method). In Hartree-Fock theory, each electron is treated as if it is subjected to the repulsion of the mean field of all of the other electrons. A variety of post-HF *ab initio* methods are designed to improve upon the Hartree-Fock energy and wavefunction

by incorporating the effects of other Slater determinants. Common post Hartree-Fock methods include configuration interaction, coupled cluster, and Møller–Plesset perturbation theory methods. As the number of determinants and size of the basis set size are increased, these approach the exact solution of the non-relativistic Schrödinger equation. Unfortunately, this comes with a sharp increase in computational cost.

Density functional theory (DFT), initially developed by physicists to solve solid-state and condensed phase problems, has recently become the major computational tool for electronic structure problems due to its excellent performances in balancing accuracy and computational cost.[33,61,62] A functional is a function that acts on another function to give a scalar. Unlike *ab initio* methods, in DFT the energy of a many-electron system are computed from electron density, $\rho(r)$, rather than the full many-dimensional wavefunction. As a result, DFT is computationally much less expensive compared to *ab initio* methods, but modern DFT methods can provide accuracy rivaling that of many *ab initio* methods. In particular, DFT computations scale as $N^3$ to $N^4$, where N is the size of the system, whereas *ab initio* methods scales as $N^4$ in the best case (Hartree-Fock), or worse for more accurate ab initio methods. The Hohenberg-Kohn (HK) theorems laid the theoretical foundation for modern DFT.[63] The first HK theorem states that the exact energy of a quantum mechanical system (under the Born-Oppenheimer approximation) can be determined by a universal functional of the exact density, $\rho(r)$. The second HK theorem states that when this universal functional is applied to an approximate electron density, the resulting energy is an upper bound to the exact energy. Unfortunately, the form of this universal functional is unknown.

Kohn and Sham introduced orbitals into density functional theory.[64] In the Kohn-Sham approach, a fictitious system of non-interacting electrons is constructed to generate the same density as the true system. Since the system only consists of non-interacting electrons, the wave function can be expressed as a single Slater determinant constructed from molecular orbitals. In the Kohn-Sham energy expression, the total energy is given by

$$E[\rho(\mathrm{r})] = T_s[\rho(\mathrm{r})] + V_{ne}[\rho(\mathrm{r})] + J[\rho(\mathrm{r})] + E_{xc}[\rho(\mathrm{r})] \qquad \text{(I-4)}$$

where the terms are, from left to right, the kinetic energy of the non-interacting electrons, the Coulombic attraction of the electrons to the nuclei, the Coulombic interaction of the electrons, and the exchange-correlation energy.

The exchange-correlation energy is where the KS energy expression differs from Hartree-Fock, and is the only component of this energy for which an exact expression is not known. The exchange effect was discovered by Heisenberg and Dirac separately,[65,66] and result from the fact that wave function of indistinguishable electrons must change sign when the coordinates of two electrons are exchanged. The correlation energy arises from the correlated motion of interacting electrons, which is conventionally defined as the difference between the exact energy and Hartree-Fock energy. Typical exchange-correlation energy expressions are split into separate components to describe these two effects. Together, the exchange functional and correlation functional define a given DFT functional. There are many popular DFT functionals that provide reliable molecular structures and energies. Unfortunately, many standard DFT functionals fail to describe dispersion interactions. However, there has been tremendous progress the last few years

developing DFT methods that capture dispersion interactions.[67] The simplest approach is the DFT-D method, in which an empirical dispersion term is simply added to the energy from a conventional DFT functional.[68,69] Examples of DFT-D functionals include B97-D and B3LYP-D3. We will use the former method, which provides reliable predictions of π-stacking, CH/π, and other non-covalent interactions, almost exclusively in this work.

# CHAPTER II

# ALKALI METAL CATION-Π INTERACTIONS*

## Introduction

Cation−π interactions, in which positively-charged atomic or polyatomic ions interact with the faces of aromatic rings (see Figure II-1), are prevalent in chemical systems, playing vital roles in the fields of chemistry, materials science, biology, and allied areas.[70-72] Although these interactions were only recognized a few decades ago,[73-75] their significance and importance has been firmly established in recent years. Cation−π interactions are primarily electrostatic in origin, arising from the interaction of a cation with the negatively charged π-electron cloud of an arene, and are often among the strongest non-covalent interactions. However, under some circumstances, cation−π interactions may also be relatively weak, depending on the nature of the cation and the arene. Indeed, the strength of cation–π interactions can be tuned over a surprisingly large range, and it is this variability that underlies the utility of cation–π interactions in many contexts. It also leads to challenges quantifying their role in many chemical systems. Overall, understanding the nature, strength, and relevance of cation−π interactions is of paramount importance in the design of functional molecules and materials.

* Reprinted with permission from "Cation-π Interactions" by Y. An and S. E. Wheeler, 2015. *The Lightest Metals: Science and Technology from Lithium to Calcium*, T. Hanusa (Ed) 173, Copyright © 2015 John Wiley & Sons Inc.

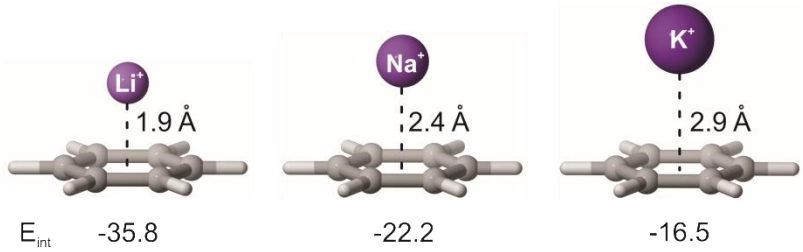



**Figure II-1.** Prototypical cation–π interactions of alkali metal cations with benzene. Optimized distances above the ring center and gas-phase interaction energies (Eint, in kcal $mol^{-1}$) are based on high-accuracy gas-phase computations from Sherrill and co-workers.[76]

There have been a number of reviews of cation–π interactions in recent years,[73-75,77,78] including the seminal work of Ma and Dougherty[73] and more recent reviews from Dougherty[74] and by Mahadevi and Sastry.[75] Although cation–π interactions often involve polyatomic cations, most notably quaternary ammoniums, in this chapter we focus on cation−π interactions involving the alkali metal cations $Li^+$, $Na^+$, and $K^+$. Complexes involving these simple cations have long served as models for more general cation–π interactions, and also play key roles in many areas of chemistry. Here, we survey their roles in chemistry, biology, and materials science, with particular attention paid to understanding the factors responsible for modulating the strength of cation−π interactions, along with their structural and functional significance.

## Nature of Cation–π Interactions

Cation−π interactions result from the interaction of a closed shell cation with a neutral π-system.[79,80] Although this moniker was initially used to refer to the interaction

of a cation with either aromatic or non-aromatic π-systems, popular use in recent years has centered on interactions with arenes. Consequently, such interactions will be the focus here. Cation–π interactions have a rich history spanning the last three decades, and our understanding of the nature of these interactions has continued to evolve over this time.[70-72]

The Genesis of Cation–π Interactions

The history of cation–π interactions begins in the early 1980s, when Kebarle and co-workers reported[81] the first measurements of $K^+$ interacting with benzene, based on mass spectrometric data. Remarkably, they found that $K^+$ preferentially binds benzene over a water molecule, suggesting that interactions of alkali metal cations with benzene could possibly be favorable in aqueous solutions. These mass spectrometric results were accompanied by *ab initio* computations, which established that cation is located along the $C_6$ axis in these complexes, directly above the face of benzene (see Figure II-1).[81] Subsequent mass spectrometric studies by Guo *et al.*[82] on non-bonded complexes of additional metal cations with benzene were in general agreement with the data from Kebarle and co-workers.[81] In particular, Guo *et al.*[82] reported stronger binding enthalpies for $Na^+$ with benzene (−28.0 kcal $mol^{-1}$), compared to $Na^+$ with $H_2O$ (−24.0 kcal $mol^{-1}$).

Cation–π interactions came to the forefront following pioneering work by Dougherty and co-workers[73,74] in the early 1990s. This was spurred in part by the demonstration that cation–π interactions play key roles in biological systems.[83-86] Moreover, it was Dougherty and co-workers who originally coined the term "cation–π interaction" to describe interactions of positive ions with either aromatic rings or non-

aromatic alkenes (*e.g.* ethene).[79,80] Initially, in the early 1980s, Dougherty *et al*. focused[87-89] on interactions of cations with aromatic host−guest systems through gas-phase measurements and studies of receptors in aqueous media. Accompanying computational studies of prototypical cation–π complexes were later reported, emphasizing structures of relevance to biological receptors and heterocycles of importance in medicinal chemistry.

In the 1990s, Lisy and co-workers[90] studied the interaction of $K^+$ with benzene and water through vibrational spectroscopy. In particular, they explored the competition between the interaction of $K^+$ with benzene and water, leading to important observations on the preferential binding of $K^+$ to benzene and the dehydration of $K^+$ by benzene. Lisy *et al*.[91] also showed that $K^+$ displays higher selectivity towards aromatic systems in aqueous environments, as compared to $Na^+$. This later led to a proposal of a size-specific mechanism for the interaction of alkali metal ions with aromatic side chains in some $K^+$ channel proteins.

In 1996, Dougherty *et al*.[88,89] showed that trends in cation–π interaction energies could be rationalized based on computed electrostatic potentials of aromatic rings. This provided a means of quantitatively predicting cation–π interactions based on straightforward computations applied to just the aromatic system, while also showing that molecular electrostatic potential (MEP) plots, such as those shown in Figure II-2, provide a powerful tool for qualitative predictions of both the geometry and strength of cation–π interactions. That is, cations will preferentially bind to areas of highly negative electrostatic potential, with the strength of binding proportional to the magnitude of the

negative electrostatic potential. Ultimately, the success in analyzing cation–π interactions in terms of electrostatic potentials solidified the now widespread view of cation–π interactions as primarily electrostatic in origin, as originally proposed by Kebarle.[81]

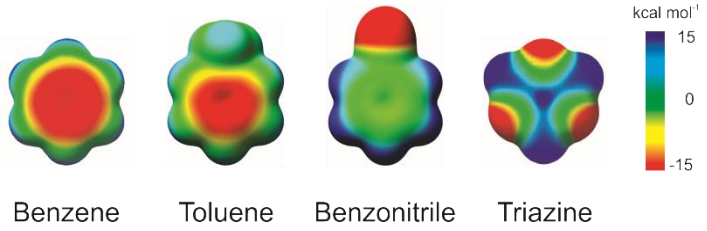



**Figure II-2.** Molecular electrostatic potential (MEP) plots of benzene, toluene, benzonitrile, and triazine. In these plots, the electrostatic potential is shown on an electron density isosurface ($\rho = 0.001\ e\ \text{bohr}^{-3}$).

Although electrostatic effects have long been understood to underlie cation–π interactions, the significance of polarization (*i.e.* induction) effects in cation−π interactions has also been emphasized,[76,92-95] most notably by Cubero *et al*.[92] Such effects, in which the electron density of the arene is polarized by the presence of the cation, are mainly dependent on the size of the π system, while dependence on the nature of substitution is negligible.

Physical Origin of Cation–π Interactions

Superficially, the origin of attractive interactions between cations and the faces of aromatic rings is straightforward—the electron rich faces of arenes bind cations. For

example, MEP plots (see Figure II-2) reveal regions of negative electrostatic potential above the faces of aromatic rings such as benzene, indicating that cations will engage in attractive electrostatic interactions with the faces of such rings. More quantitatively, cation–π interactions are often discussed in terms of molecular quadrupole moments. In particular, the $z^2$-component of the molecular quadrupole moment of benzene is –8.7 B.[96,97] This negative $Q_{zz}$ value arises from the abundance of negative charge along the *z*-axis (*i.e.* the π-electron cloud) and positive charge away from the *z*-axis (*i.e.* the partially positive charged hydrogens). The result is a sizeable quadrupole moment suited to interact favorably with a positive charge located along the *z*-axis. Thus, as a first approximation, cation–π interactions can be viewed as electrostatic interactions between a charge and an electric quadrupole.

However, as noted by Ma and Dougherty in their 1997 review,[73] the cation–π interaction cannot be modeled quantitatively as a charge-quadrupole interaction. This is because the size of the arene is commensurate with the distance between the arene and cation (2–3 Å); any multipolar expansion of electrostatic interactions at such close proximities will be divergent. This weakness of the charge-quadrupole picture in describing cation–π interactions will be even more severe when considering larger arenes.

Alternatively, one can view benzene as a collection of six radially-oriented dipoles, corresponding to the polarized C-H bonds. From this perspective, favorable cation–π interactions arise because a cation located above the center of the ring can interact with the positive end of each of these six local C-H dipoles. This local multipole

view of aromatic rings is more physically sound when discussing cation–π interactions with larger arenes, and also leads to a more satisfactory understanding of substituent effects in cation–π interactions.[78,98-100] Such local multipole descriptions of the electrostatic component of non-covalent interactions have long been championed by Stone and co-workers,[101-104] and are finding increasing use in modern molecular mechanics force fields.[105]

Electrostatic potentials, which are widely used in discussions of non-covalent interactions, can provide a more rigorous means of predicting the electrostatic interaction between a cation and an aromatic ring.[88,89] Formally, the electrostatic potential at a given point near a molecule is the electrostatic interaction that a positive test charge would experience at that point. Electrostatic potentials provide reliable predictions of relative interaction energies for cation–π complexes (*vide infra*), particularly in cases in which the arenes involved all have similar polarizabilities. However, despite their widespread use and utility in predicting cation–π interactions, MEP plots are widely misinterpreted,[98,106,107] adding an additional layer of confusion to many discussions of cation–π interactions.

Furthermore, as noted early on by Cubero and co-workers,[92] as well as others,[76,93-95] there are other, significant contributions to cation–π interactions beyond the simple electrostatic effects captured by electrostatic potentials and molecular quadrupole moments. Chief among these are due to induction effects, which refers to the enhanced interaction that arises from the relaxation of the electron density in each monomer in the presence of the other monomer (*i.e.* the mutual polarization of one monomer by the

other). Such effects are strongly dependent on the polarizability of the interacting species. Although molecular quadrupole moments provide a reasonably accurate description of relative cation–π interaction energies involving sets of similar arenes, this correlation can break down when the polarizability of the arenes differ significantly.[108] In these cases, variations in the contribution of induction lead to changes in the interaction energy that are not captured by the $Q_{zz}$ values.

Sherrill and co-workers[76] have performed density functional theory (DFT) based symmetry-adapted perturbation theory (SAPT) computations on model cation–π complexes. SAPT computations[109,110] provide a rigorous dissection of non-covalent interaction energies into contributions from electrostatics, induction, dispersion, and exchange-repulsion (Pauli repulsion). Sherrill *et al*.[76] confirmed previous reports[93-95] that the contribution of dispersion interactions in cation–π interactions are small, contributing only 4% to the total binding of $Na^+$ by benzene, for example. They also showed that, for this prototypical cation–π complex, induction effects are responsible for nearly half of the total binding energy, with the remainder being electrostatic in origin.

Of course, induction alone is typically not sufficient for strong cation–π interactions, although it can lead to overall favorable binding, and a strong electrostatic component can be considered to be a defining characteristic of cation–π interactions.[74] For example, despite the greater polarizability of cyclohexane, compared to benzene, the former does not bind alkali metal cations nearly as well as the latter.[111] Intriguingly, Sherrill and co-workers[76] also revealed that alkali metal cations can also interact favorably with benzene in a "side-on" configuration. In this geometry, the mildly

repulsive electrostatic component of the interaction is overcome by strongly attractive induction terms, with a small contribution from dispersion effects. However, in our view, such complexes should not be referred to as cation–π complexes, because they are fundamentally different in origin.

Finally, we turn briefly to the role of aromaticity in cation–π interactions. Although the original definition[79,80] of cation–π interactions included interactions of cations with both aromatic and non-aromatic π-systems, there has been greater focus in recent years on cation–π interactions involving aromatic systems. This presumably arises in part from the prevalence of cation–π interactions in biological systems, in which cations interact with aromatic amino acid side chains. Somewhat unexpectedly, in 2011, Bloom and Wheeler recently showed[112] that aromaticity actually enhances the strength of cation–π interactions. That is, an aromatic ring will bind a cation more strongly over its center than an otherwise identical non-aromatic cyclic conjugated polyene. This can be explained by the net influx of π-electron density toward the center of a cyclic polyene upon aromatic delocalization, and suggests that cation–π interactions are indeed "special", unlike π–stacking and other non-covalent interactions involving aromatic rings.[112]

Tuning the Strength of Cation–π Interactions

One of the key features of cation–π interactions is the possibility of tuning their strength over a broad range. First, the strength of cation–π interactions depends strongly on the identity of the cation, with $Li^+$ interacting with benzene far more strongly than either $Na^+$ or $K^+$ (see Figure II-1). However, for a given alkali metal cation, the strength

of cation–π interactions can also be widely tuned through substituent and heteroatom effects.[98] Although this has long been established, our understanding of the origin of substituent and heteroatom effects in the strength of cation–π interactions continues to evolve.[78,99,100]

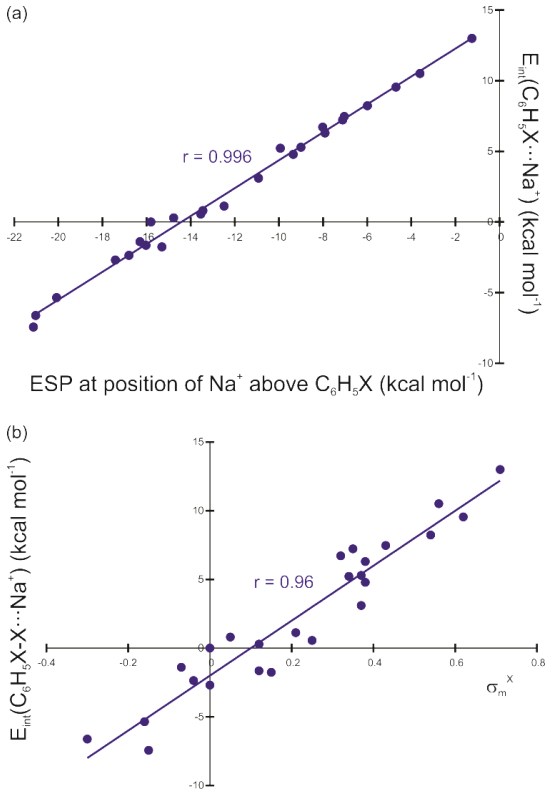



**Figure II-3.** DFT computed interaction energies, relative to the unsubstituted case (X = H), for model complexes of $Na^+$ above the center of monosubstituted benzenes, $C_6H_5X$, vs (a) the electrostatic potential computed at the position of the $Na^+$ above the ring and (b) the Hammett $\sigma_m$ constant for the substituent. Data from Ref. [98].

Cation–π interactions show marked sensitivity to substituents. For example, the interaction of $Na^+$ with benzene in the gas-phase (–23.5 kcal $mol^{-1}$)[98] can be tuned over a range of at least 20 kcal $mol^{-1}$ by adding a single substituent to the benzene. This can be seen in Figure II-3, in which relative gas-phase interaction energies of $Na^+$ over the center of 25 monosubstituted benzenes are plotted. These interaction energies are strongly correlated with the computed electrostatic potential computed at the position of the anion above the ring (see Figure II-3a). They are also reasonably well-correlated with Hammett $\sigma_m$ constants (see Figure II-3b), which can provide a qualitative guide to the strength of these interactions in substituted aromatic systems. Moreover, as with other non-covalent interactions involving aromatic rings,[78,99,100,113] substituent effects in cation–π interactions are additive, providing access to an even greater range of interaction energies through the introduction of multiple substituents.

Traditionally, substituent effects in cation–π interactions were rationalized based on the π-electron density of the substituted ring (see Figure II-4a). That is, electron donating groups ($CH_3$, $NH_2$, *etc*.) enhance cation–π interactions by increasing the π-electron density of the ring. Similarly, electron-withdrawing groups ($NO_2$, CN, etc) lead to weaker cation–π interactions through the depletion of the π-electron density of the aromatic ring. These notions are seemingly supported by MEP plots (see Figure II-2), and are reminiscent of the seminal model of aromatic interactions from Hunter and Sanders.[114] For example, toluene binds $Na^+$ more strongly than does benzonitrile, in accordance with qualitative predictions from the corresponding MEP plots, as well as conventional notions of the π-electron densities of these two aromatic systems.

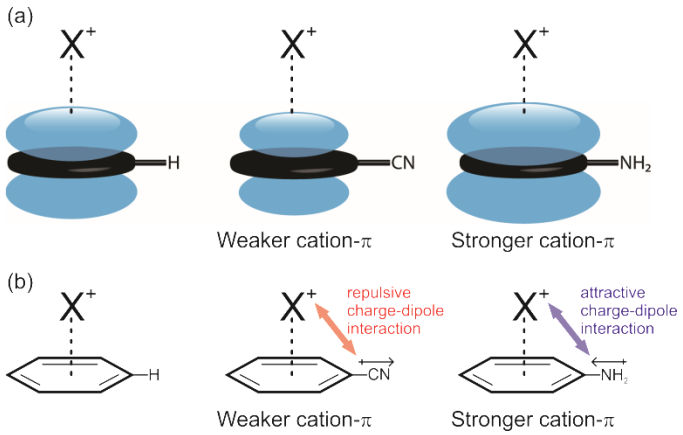



**Figure II-4.** Competing models of substituent effects in cation–π interactions. (a) In conventional views, substituent effects arise from the modulation of the aromatic π-electron density by the substituents. (b) In the local dipole model of Wheeler and Houk,[98] substituent effects in cation–π interactions are dominated by the through-space electrostatic interaction of the local dipole associated with the substituent and the cation. Both models tend to provide similar predictions with regard to substituent effect trends.

However, the origin of differences in electrostatic potentials of substituted arenes, as well as the interactions of these systems with cations, are widely misunderstood.[98,106] In particular, substituent effects in cation–π interactions are not necessarily due to substituent-induced changes in the π-electron density. In 2009, Wheeler and Houk[98] showed that both cation–π interactions and the electrostatic potentials above the centers of substituted benzenes can be accounted for by the through-space effects of the substituents. There is no need to consider changes in the π-electron density caused by the substituents. This mirrors similar revelations regarding substituent effects in other non-covalent interactions involving aromatic rings, including π–stacking and anion–π interactions.[99,100,115-120]

Wheeler and Houk[106] later extended these findings to MEP plots, which were similarly shown to have little to do with the distribution of π-electrons. In other words, the dramatic differences in the MEP plots of aniline, benzene, and benzonitrile (see Figure II-2) do not arise from underlying differences in π-electron density. Instead, Wheeler and Houk[106] showed that substituent effects in the electrostatic potentials of arenes are dominated by through-space effects of the substituents. This can be understood in terms of the effect of the local dipole moment associated with the substituent on the electrostatic potential. This can be seen in Figure II-5, in which the electrostatic potential in the plane bisecting benzonitrile (leftmost image) is qualitatively reproduced by simply adding the electrostatic potential of HCN to that of benzene. In other words, the change in the electrostatic potential above benzene from negative to positive by adding a single CN substituent simply reflects the positive end of the local CN dipole, not any substantive changes in the π-electron density of the ring.

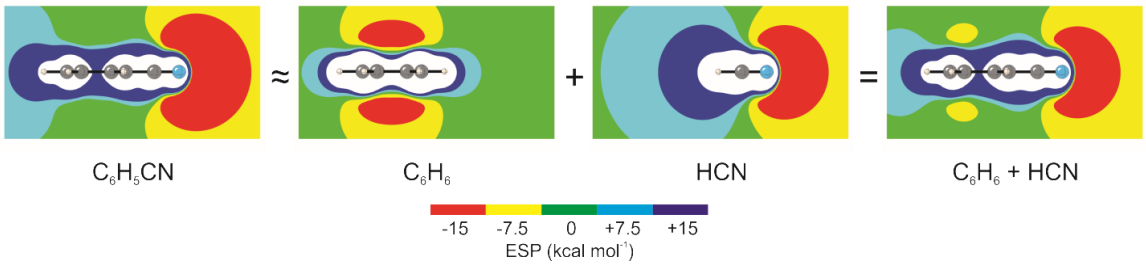



**Figure II-5.** The electrostatic potential in the plane bisecting benzonitrile (left), as well as an additive approximation of this electrostatic potential constructed by adding the electrostatic potential in the plane bisecting benzene to the electrostatic potential of HCN. Reprinted with permission from Wheeler.[391]

Thus, the dominant factor in substituent effects in cation–π interactions is the interaction of the local dipole moment associated with the substituent and the cation (see Figure II-4b). This can be seen explicitly in Figure II-6, in which an accurate interaction potential for $Na^+$ above the center of benzonitrile is reproduced quantitatively by adding a simple charge-dipole interaction to the interaction of $Na^+$ with benzene. That is, no changes in the interaction of the cation with the aromatic ring itself are needed; instead, one only needs to account for the interaction of the cation with the local dipole associated with the nitrile substituent. This strongly supports the local dipole picture of substituent effects in π–stacking interactions,[98] and precludes any substantial role of π-electron density changes caused by the substituents.

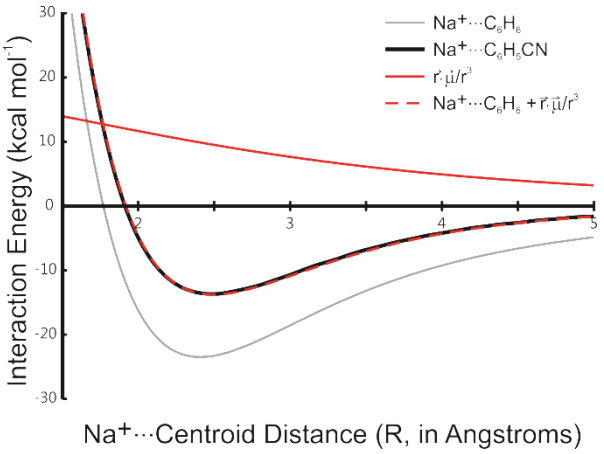



**Figure II-6.** Accurate gas-phase interaction energy of $Na^+$ with benzene (gray line) and benzonitrile (black line) as a function of distance above the ring. Also shown are a simple charge-dipole interaction (red line) and the result of adding this charge-dipole term to the $Na^{+}\cdots$benzene interaction potential (red dashed line). Modified with permission from Wheeler.[391]

Recent work from Wheeler and Bloom[107] suggests that there are similar misconceptions regarding the impact of heteroatoms on the ability of *N*-heterocycles to bind cations, which is often discussed in terms of π-electron densities. For example, Wheeler and Bloom showed that the increasingly positive electrostatic potentials of the symmetric azines pyridine, 1,4-pyrazine, 1,2,3-triazine, 1,2,4,5-tetrazine, and the hypothetical molecule hexazine have little to do with changes in the π-electron density. Indeed, the effect of introducing nitrogens into an aromatic ring is a net influx of π-electron density towards the ring center! For example, despite the strongly positive electrostatic potential above 1,3,5-triazine (see Figure II-2), there is actually more π-electron density near the center of triazine than benzene. Although the impact of these effects on cation–π interactions has not been considered explicitly, related work on anion–π interactions[107] suggests that the effect of *N*-heteroatoms on strength of cation–π interactions are not due to changes in the π-electron density.

Regardless of the origin of the effects, both heteroatoms and substituents provide a simple and predictable means of tuning the strength of cation–π interactions. This tuning can be exploited in a myriad of chemical contexts, and contributes to the power of cation–π interactions.

**Spectroscopic Characterizations of Cation−π Interactions**

Over the past decade, mounting spectroscopic data on cation−π interactions[121-123] has provided many key insights into their nature and abundance. Nuclear magnetic resonance (NMR), mass spectrometry (MS), infrared spectroscopy (IR) and other

techniques have been widely applied in the characterization of these interactions. In particular, NMR has been used to detect and characterize cation−π interactions in aqueous solution and in the solid state. For example, Wu and co-workers[124,125] established NMR signatures for complexes containing cation−π interactions through solid-state NMR studies of tetraphenylborate salts. Alkali metal ions exhibit highly negative chemical shifts when bound to tetraphenylborate compounds, and this highly-shielded environment at the metal center can be used to detect cation−π interactions. Gokel *et al.*[126] used solid-state NMR to study cation−π interactions between $Na^+$ and two lariat ether molecules, in which the sodium cation interacts with an indolyl group, while Cuc and co-workers[127] investigated an inclusion complex of calix[4]arenes and $Cs^+$ via diffusion NMR spectroscopy.

Mass spectrometry has also been widely used as a tool to investigate the properties of cation−π complexes, as well as the intrinsic binding mode of different cations. For instance, Dunbar and co-workers[128-131] utilized Fourier-transform ion cyclotron resonance (FT-ICR) ion trapping mass spectrometry to explore the binding mode of cations upon the formation of metal ion−π complexes with coronene,[128] complexes of $Na^+$ with Ala and Phe amino acid side chains,[130] and encapsulation of metal cations by Phe−Phe ligands.[131] Armentrout *et al.*[132] determined the binding energies of mono and bis-benzene complexes with alkali metal cations using Collision-induced dissociation (CID). The observed trend in binding energies depended on the magnitudes of electrostatic interactions and ligand−ligand repulsions. Rodgers and co-workers[133-141] performed number of experiments on metal ion−π complexes using

threshold CID along with complementary computational studies. They showed that the strength of cation−π interactions of alkali metal ions with a variety of π-systems, including azines and other *N*-heterocycles, halobenzenes, toluene, and the aromatic amino acids Phe, Tyr, and Trp, depend both on nature of the π-group and any substituents present. Finally, mass spectrometry has been use to show that cation−π interactions enhance the formation of sandwich benzo-crown ether-alkali metal cation complexes with $Na^+$, $K^+$, and $Cs^+$ ions.[142]

Cation−π interactions of alkali metal−crown ether complexes were studied in the gas phase by Lisy *et al.*[143] using infrared predissociation (IRPD) spectroscopy, accompanied by DFT computations and SAPT analyses. Recently, Dunbar *et al.*[131,144,145] investigated the encapsulation of metal cations by aromatic peptides using Infrared multiple photon dissociation (IRMPD) spectroscopy. The interplay between alkali ions and different di-, tripeptides, and Phe ligands were characterized. Alkali metal-ion complexes were shown to prefer charge-solvated binding modes since the zwitterion (salt bridge) complexes are less stable. Binding thermochemistry of cation−π complexes were also studied by Morel and co-worker.[146]

A number of other spectroscopic techniques have been brought to bear on cation–π interactions. Schlamadinger *et al.*[147] used UV resonance Raman (UVRR) spectroscopy to study the changes in vibrational structure upon formation of a cation−π complex. They observed systematic shifts in relative intensity in the 760−780 $cm^{-1}$ region when $Na^+$ or $K^+$ and the indole moieties of a crown ether coordinate through cation−π interactions. Ultraviolet second-derivative absorbance spectroscopy was

utilized by Lucas and co-workers[148] to understand cation−π interactions involving aromatic residues in protein structures. In particular, structural information on eight different proteins was obtained by probing cation−π interactions and evaluating the empirical UV spectral data.

Solid state structures have also been explored recently in order to shed additional light on cation−π interactions.[149,150] For example, Gokel and co-workers[22,48-58] employed their lariat ether model systems as a tool to detect cation−π interactions in the solid state for synthetic receptors bearing the aromatic amino acid side chains of Phe, Tyr, and Trp.[126,151-159] Cation complexation has been attributed in part to the presence of flexible side arms that can compete with the counteranion. A diaza-18-crown-6 scaffold in the synthetic receptor model systems was determined to be the primary cation binding site in these systems, and modifications of its side arms modulate its ability to bind cations. Gokel *et al.*[160] also studied cation transport through synthetic hydraphile channels comprising side arms with varying degrees of π-electron density. Cation−π interactions were demonstrated to play a vital role in obtaining a stable, active ion channel conformation.

Finally, interaction energies between alkali metal ions and model π-systems in different phases were compared by Fukin *et al.*[161] Cation–π interaction energies were found to follow the order $Cs^+ \approx Rb^+ > K^+ > Na^+$ in the solid-state for 1:1 complexes of metal cations with hexakis-(methoxymethyl) benzene. This trend is notably different from the order observed in the gas-phase. In these systems, the capture of alkali-metal cations is facilitated by a pair of ether "tentacles", which loosely position the cation over

the phenyl ring. X-ray crystallographic analyses[162] of N-[3,5-bis(α,α-dimethylbenzyl-2-hydroxybenzyl)] iminodiacetic acid, in which a potassium ion interacts with aromatic groups, reveal strong cation−π interactions of the metal ion with α,α-dimethylbenzyl groups. Finally, intramolecular interactions of tris(2,6-diisopropylphenoxy)-silanethiol and its $Na^+$ salts were investigated by Dolega and co-workers.[163] The sodium ion was shown to coordinate with the neutral phenyl rings by analyses of the crystal structure and through FT-IR spectroscopy.

**Cation−π Interactions in Host-Guest and Biological Systems**

Host guest complexes have been employed as powerful tools for characterizing molecular recognition between synthetic receptors and small guest molecules facilitated by cation−π interactions, shedding light on cation–π interactions with aromatic amino acid side chains. For example, in 1995, selective recognition of alkali metal cations by cyclophanes was reported by Inokuchi *et al.*[164] Cage-like molecules featuring phenyl walls and caps were synthesized by Kim and co-workers[165] in order to examine cation complexation, in which the selectivity for $Li^+$ and $NH_4^+$ was governed by the gate size of the cage. Lelias-Vanderperre *et al.*[166] studied cation recognition by macrotricycles comprising aromatic rings, reporting preferential binding of $NH_4^+$ over $K^+$. Rathore and co-workers[167,168] designed and synthesized a hexaaryl benzene (HAB) based receptor featuring a bipolar receptor site, which enhanced the binding ability of $K^+$ by the synergistic effect of a polar ethereal fence and the central benzene ring through a cation−π interaction. Cametti and co-workers[169] studied cation−π interactions between

alkali metal halide and uranyl-salophen receptors. Solid-state structures of complexes formed by the two armed uranyl−salophen complex receptors with CsF and with chlorides from KCl, RbCl, and CsCl reveal the existence of dimeric supramolecular assemblies in which two receptor units assemble into capsules fully enclosing $(MX)_2$ ion quartets. Nissinen *et al.*[170,171] used NMR and X-ray structure analyses to investigate cation−π interactions between the alkali metals $K^+$, $Rb^+$, and $Cs^+$ and pyrogallerene and resorcinarene receptors. Strong cation−π interactions were identified, and both electrostatic complementarity and preorganization were shown to be critical of forming molecular cages.

Host-guest complexes in which the cation is not located above the center of an aromatic ring have also explored computationally by Macias *et al.*[172] Their computational results indicate that multiple weak, non-optimal cation–π interactions can contribute significantly to the overall binding strength. This analysis underscores the importance of neighboring aromatic faces and provides new insight into the significance of cation–π binding, not only for calix[4]arenes, but also for other supramolecular and biological systems.

Cation−π interactions can play important roles in enzyme conformation and activity. For example, Wouters *et al.*[173] showed that a cation−π interaction between $Na^+$ and a nearby tryptophan in the crystal structure of tetragonal lysozyme adopts a conformation similar to that in model cation–π systems, while Pan and co-workers[174] found that a Phe residue is highly conserved in integrin $\alpha_4\beta_7$, and engages in cation–π interactions that are crucial to the regulation of integrin affinity and signaling. Dougherty

*et al*.[175] explored the interaction of $Na^+$ and $Ca^{2+}$ with aromatic residues in the external vestibule of the ion-conducting pores of a sodium channel. A cation–π interaction involving $Ca^{2+}$ was confirmed by the linear response of $Ca^{2+}$ blocking induced by increasingly fluorinating the aromatic ring.

Williams and co-workers[176,177] investigated the potential role for both divalent and monovalent cations in disrupting DNA and RNA structures and stabilizing the unstacked conformations through cation−π interactions. Cation−π interactions, as well as hydrogen bonds and π–stacking interactions between loop residues and the outer G-quartets, have been suggested to all contribute to the formation and stability of DNA G-quadruplexes.[178-181] Analysis of a data set of mesophilic and thermophilic protein single chain models showed that cation−π interactions stabilize protein secondary structures, leading to higher rigidity and enhanced thermal stability. Finally, Matsumura *et al*.[182,183] demonstrated the presence of a unique cation–π interaction between $Na^+$ and a Phe residue in the wild type T1 lipase of *Geobacillus zalihae* (see Figure II-7), the impact of which was quantified through molecular dynamics simulations.

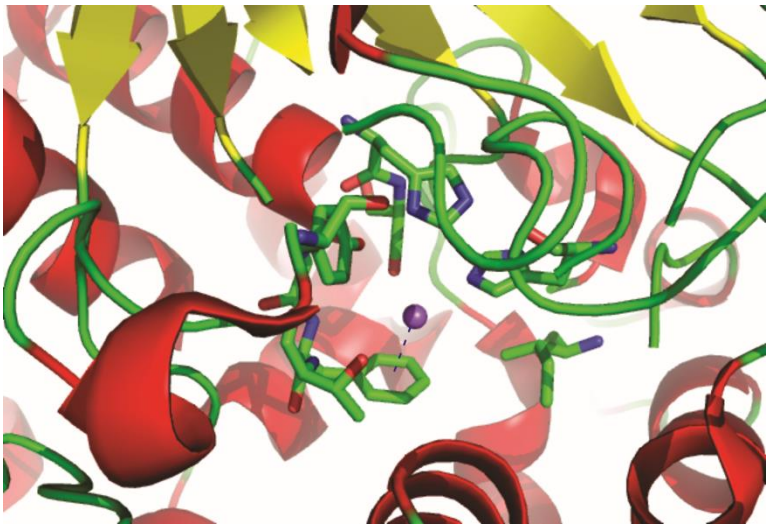


**Figure II-7.** Structure of the T1 lipase studied by Matsumura *et al.*,[182] which exhibits a well-defined cation–π interaction between $Na^+$ and a nearby Phe residue (PDB: 2DSN).

**Cation−π Interactions in Materials Science**

Cation–π interactions have also received some interest in the realm of materials. For example, Ramamurthy and co-workers[184] showed that aromatic molecules tend to aggregate within cation exchanged Y zeolites. Computations indicated a strong stabilizing effect arising from cation−π interactions, which is consistent with the formation of ground-state π-stacked aggregates of naphthalene molecules within Y zeolites. Ramamurthy *et al.*[185] also probed the impact of cations on photochemistry inside zeolites, showing that cation–π interactions control the photochemical and photophysical behaviors of supercages of zeolite guest molecules.

Although the interaction between neutral alkali metals and graphene or carbon nanotubes has been widely studied, there are surprisingly limited computational studies of cation−π interactions in nanosystems.[186-188] Tachikawa *et al.*[189] reported

computational studies on the interaction between $Li^+$ and the surface of amorphous carbon based on semiempirical MO theory and dynamics simulations. The results showed that $Li^+$ tends to bind at the edges of model amorphous carbon clusters through peripheral cation−π interactions. More recently, Tachikawa *et al.*[190] investigated the interaction of $Li^+$ with $C_{60}$, identifying distinct stable binding sites in which $Li^+$ interacts with either a five-membered ring or six-membered ring.

Umadevi *et al.*[191] used DFT to study the interaction of small molecules and ions with graphene, considering a series of linear and branched polycyclic aromatic hydrocarbon (PAH) systems as models. In a more comprehensive follow-up study,[192] the effect of chirality and curvature of carbon nanotubes (CNT) on their interaction with alkali and alkali earth metal ions was explored, revealing that monocationic ions preferentially bind zigzag CNTs, whereas the dications bind more strongly to armchair CNTs. Valencia *et al.*[193] studied the adsorption of alkali metal cations on graphite, using DFT to understand the structural, energetic, and electronic properties of these cation−π complexes. They suggested that one valence electron is entirely transferred from the alkali metal to the graphite surface, leading to a strong interaction between the lithium ion and the π-electrons of the graphite.

Mpourmpakis *et al.*[194] used *ab initio* computations and molecular mechanics simulations to show that the interactions of $Li^+$, $Na^+$, and $K^+$ with carbon nanotubes are independent of the nanotube curvature. Cation−π interactions between alkali metal cations and $C_{24}$-fullerene were investigated by Moradi and co-workers[195] using DFT. They found that the ideal adsorption site is above the center of one six-membered ring

on the exterior surface of $C_{24}$. Overall, the trend in binding of $Li^+$, $Na^+$, $K^+$ follows that for smaller arenes, with larger cations binding more weakly. Related boron-doped fullerenes were shown to have stronger interaction energies, while a nitrogen-doped analogue bound cations more weakly. Finally, Hilder *et al.*[196] used MD simulations to study the selectivity of cations permeating boron nitride nanotubes, showing that these nanotubes selectively allow monovalent cations to pass through while blocking anions.[197]

**Cation−π Interactions in Chemical Catalysis**

Recently, cation−π interactions involving alkali metal cations have also gained some attention from synthetic chemists, due to their potential utility in catalytic reactions. For example, Ramamurthy and co-workers[198] reported the selective photoisomerization of diarylcyclopropanes arising from strong cation−π interactions that lead to preferential binding of alkali metal cations to ideally placed phenyl rings. Wooten *et al.*[199] studied the impact cation−π interactions on the structure and binding of $M_3(sol)_n(BINOLate)_3Ln$ catalysts, while Monje and co-workers[200] explored the stabilizing effect of cation–π interactions in organolithium chemistry via Sn-Li exchange equilibria. The interaction energy between $Li^+$ and the π-system of α-oxy-organolithium molecules was demonstrated to be comparable to that observed in systems containing Li-N or Li-O interactions. Finally, a series of alkali metal cation functionalized titanosilicate molecular sieves were reported in which cyclohexene epoxidation with

$H_2O_2$ was demonstrated. Ultimately, a distinct relationship between catalytic activity and intermolecular cation–π interaction energies were reported.[201]

## Conclusions

Cation–π interactions involving alkali metal cations have emerged as key non-covalent interactions in modern chemistry, biology, and materials science. Although our understanding of the nature of these interactions continues to evolve, the confluence of spectroscopic and computational analyses has unveiled many key properties of these interactions, while also identifying cation–π interactions as important non-covalent interactions in a number of chemical contexts. Importantly, the strength of cation–π interactions can be systematically modulated through the introduction of substituents and heteroatoms into aromatic rings, providing a powerful means of tuning these interactions for uses in everything from materials to catalysis.

# CHAPTER III

# AROMATIC INTERACTIONS MODULATE THE 5'-BASE SELECTIVITY OF THE DNA-BINDING AUTOANTIBODY ED-10[*]

## Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease whose underlying cause is not fully understood.[202] Anti-DNA antibodies (Abs) often play a crucial role in the disease pathogenesis by triggering kidney inflammation via antibody-DNA complex deposition.[203] Several non-DNA ligands have been identified that bind to anti-DNA Abs, including both small molecules and macro biomolecules.[204,205] Molecules that bind pathogenic anti-DNA Abs may prevent tissue damage and are considered potential therapeutic agents for the treatment of SLE.[206] Thus, it is crucially important to elucidate the mechanisms by which anti-DNA Abs recognize diverse ligands in order to understand the roles that anti-DNA Abs play in autoimmune disease pathology, and for inhibitor design.[207]

DNA-binding antibodies can also serve as models of DNA and RNA recognition by proteins,[208] which underlie many key biological phenomena.[209-222] These recognition phenomena depend on the interplay of sundry non-covalent interactions, and a complete elucidation of the general mechanism of base recognition by proteins has remained elusive. Selective recognition of DNA and RNA bases has long been attributed to

[*] Reprinted with permission from "Aromatic Interactions Modulate the 5'-Base Selectivity of the DNA-binding Autoantibody ED-10" by Y. An and S. E. Wheeler, 2014. *J. Phys. Chem. B* **118**, 5653. Copyright © 2014 American Chemical Society.

specific hydrogen bonding interactions.[223-227] On the other hand, the role of aromatic amino acid residues [tyrosine (Tyr), tryptophan (Trp), phenylalanine (Phe), and histidine (His)] in these recognition processes has been a somewhat contentious issue, with previous attempts to quantify their role in DNA and RNA recognition based on statistical analyses of known protein-DNA structures leading to mixed conclusions.[228-232] However, because the strength of π-stacking interactions varies depending on the identity of the aromatic amino acid side chain and nucleobase,[233-237] these interactions will necessarily impact base selectivity. What remains to be quantified, however, is the extent to which π-stacking interactions can impact base selectivity in specific DNA and RNA binding proteins.

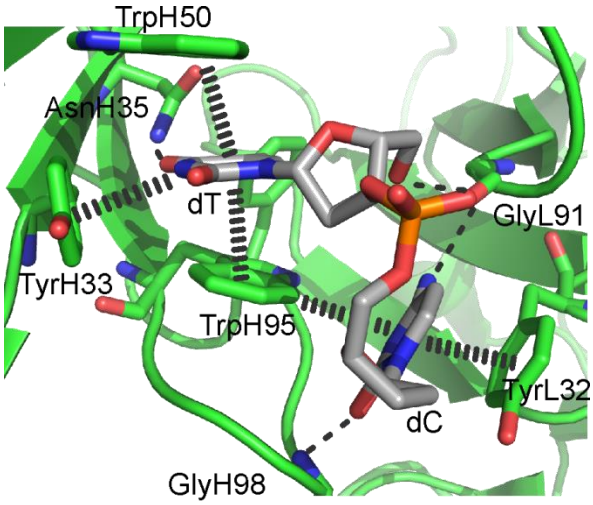



**Figure III-1.** 5'-dTdC-3' binding site in ED-10. The dC binding pocket is the D-ARM of Tanner and co-workers.[238]

Anti-DNA antibodies are classified as either single-stranded (ss) or double-stranded (ds) DNA-antibodies depending on what type of DNA they bind.[239-241] So far, only a few crystal structures of ss-DNA-antibody complexes are available.[242-248] Among these structures, ED-10, BV04-01, and DNA-1 are the best examples, with detailed information about affinity, structure, and sequence specificity.[242-248] Based on available structural data, Tanner and co-workers identified a conserved and fundamental structural element responsible for the recognition of ssDNA, termed the "ssDNA-antibody recognition module" (D-ARM).[238] The D-ARM consists of a tyrosine residue that stacks with the base and a glycine that forms a hydrogen bond with the base.

Sanguineti *et al.*[249] investigated the binding of a number of oligonucleotides to monoclonal Ab ED-10, which exhibits sub-nanomolar binding affinities. Some of their reported relative binding free energies are shown in Table III-1. They showed via fluorescence binding experiments and thermodynamic measurements that the 5' end of ssDNA has the highest binding affinity,[249] and that only the first two bases from the 5' end (dTdC in the case of the native antigen) interact strongly with the antibody. Moreover, it was found that the binding affinity is insensitive to the oligonucleotide length, as long as the sequence 5'-dTdC is present. On the other hand, if the 5'-base is anything other than thymine, the binding free energy is about five kcal $mol^{-1}$ less favorable.

Sanguineti *et al.*[249] also reported a 1.89 Å resolution crystal structure with a truncated 6-mer nucleotide bound. We note that in this crystal structure, the 6-mer has been cleaved, leaving 5'-dTdC-3' bound, with the C3' hydroxyl group missing from the

cytosine nucleotide. The binding site of ED-10 from this crystal structure is depicted in Figure III-1. Cytosine is located in the D-ARM, while thymine is bound in a similarly structured binding pocket. π-stacking interactions from the light chain complementarity-determining region (LCDR) 1 and heavy chain complementarity-determining region (HCDR) 3 as well as hydrogen bonding interactions from LCDR3 are observed in the D-ARM. The cytosine ring engages in π-stacking interactions with TyrL32 and is roughly perpendicular to the TrpH95 ring system in the tip of HCDR3. Cytosine also forms a hydrogen bond with the main-chain carbonyl from GlyL91. Missing from this binding pocket is the hydrogen bonding interaction with the L91 side chain, as seen in other examples of the D-ARM.[242-248] Finally, cytosine also accepts a hydrogen bond from the main-chain amide of GlyH98. Sanguineti *et al.*[249] showed that TyrH97 played an integral role in the binding of this dinucleotide, rearranging with large conformational changes to allow sub-nanomolar binding affinity.

**Table III-1.** Experimental/MMGBSA predicted binding free energies/enthalpies (kcal $mol^{-1}$), relative to the native antigen, 5'-dTdC, for four 18mer oligonucleotides/dinucleotides bound to the native antibody (ED-10) and three point mutants.

| | ED-10 | | Point Mutants[b] | | |
|---|---|---|---|---|---|
| | Exp[a] | MMGBSA | AlaH50 | PheH50 | TyrH50 |
| 5'-dTdC | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5'-dAdC | 5.2 | 8.7 | 2.8 | 6.5 | 5.4 |
| 5'-dGdC | 5.1 | 0.7 | - | - | - |
| 5'-dCdC | 4.9 | 14.7 | - | - | - |

[a] Experimental binding free energies are for 18-mer oligonucleotides starting with the listed sequence.

[b] For point mutants, enthalpies are calculated instead of free energies.

The binding site for thymine is more extensive, featuring π-stacking interactions with TrpH50 and TrpH95, an edge-to-face interaction with TyrH33, and a hydrogen bond donated from AsnH35 (see Figure III-1). The presence of three flanking aromatic amino acid residues in the dT binding pocket appears to be unique among ss-DNA binding antibodies for which detailed structural information is available.[242-248] As shown below, edge-to-face XH/π interactions with TyrH33 are vital for the 5'-dT selectivity exhibited by this antibody.[249]

To date, there have been no detailed computational studies of the origin of base selectivity in anti-ssDNA binding antibodies. Herein, we present computational analyses of the binding of four dinucleotides to ED-10, as well as selected point-mutants. We utilize classical molecular mechanics based dynamics simulations[50] and density functional theory (DFT) computations applied to a cluster model of the 5'-nucleotide binding sites.[250-254] The primary aim is to identify the non-covalent interactions that give rise to the pronounced selectivity for 5'-dT exhibited by ED-10, and hence shed light on the origin of sequence selectivity and mode of antigen recognition in this anti-ssDNA antibody.

**Theoretical Methods**

We have explored the ssDNA antibody complex (ED-10) with the native antigen 5'-dTdC bound as well as complexes with 5'-dAdC, 5'-dCdC, and 5'-dGdC. The crystal structure of ED-10 was extracted from PDB file 2OK0, in which the 5'-thymidine presents in an *anti* conformation. The 5' nucleobase was mutated from thymine to

adenine, cytosine, and guanine using Pymol,[255] with each mutated 5' nucleoside kept as an *anti* conformer. The starting coordinates of each complex were generated by the AMBER package with hydrogens added.[256-258] Each complex was solvated in a 100.0 Å $\times$ 74.2 Å $\times$ 82.3Å rectangular box of TIP3P water molecules.[259] $Na^+$ ions were added to neutralize the system.

With the atoms of the ssDNA antibody complex fixed, the energy of each system was first minimized for 10,000 steps using steepest descent and conjugate gradient algorithms, followed by another 10,000 energy minimization steps without constraints. After energy minimization, each system was heated from 0K to 300K under constant pressure for a total of 300ps. A three staged density equilibration was applied, involving two initial steps with weak restraining forces and a third step without restraints for a total of 2ns. A 10ns simulation was then carried out for each protein-DNA complex using the Amber03 force field in an isothermal-isobaric (constant NPT) ensemble after the above 2ns equilibration.[46] The coordinates of each complex were saved every 2ps and the temperature maintained at 300K via Langevin dynamics.[260] The constant pressure of 1 atm was kept using a Nose-Hoover barostat.[261] The SHAKE algorithm was used to constrain bonds to hydrogen.[262] The particle mesh Ewald method was employed to treat the long-range electrostatic interactions.[263]

Binding enthalpies were estimated using MMGBSA (molecular mechanics/generalized Born model and surface area model) for each ssDNA antibody complex based on 1,000 frames recorded during the above-described 10 ns trajectories.[42] The single trajectory approach was used to calculate the binding enthalpy of dTdC,

dAdC, dGdC, and dCdC in ED-10. Due to the computational cost of normal mode analyses, the entropic contribution to the relative binding free energy was estimated based on normal mode analyses applied to 40 frames. The multiple trajectory approach to MMGBSA was also employed, and confirmed the trend of binding energy obtained using the single-trajectory approach.[59,264] Finally, MMGBSA was also used to compute pairwise interactions between the 5'-base and key binding site residues by only considering interactions between pairs of atoms contained within these interacting species.

These MD simulations were complemented by density functional theory (DFT) computations on cluster models of the 5' binding site at the B97-D/TZV(2*d*,2*p*) level of theory.[265-267] This was done primarily to quantify the non-covalent interactions between ssDNA and the binding site amino acid residues. The B97-D functional provides accurate predictions of dispersion-dominated non-covalent interactions, including the π-stacking, CH/π and NH/π interactions that are germane to this study.[68,268-270] Binding site amino acids were defined as all the residues within 5Å of the 5' nucleobase. The sugar ring and phosphate of the nucleotides were removed and open valences were capped with hydrogen atoms. To model an amino acid or groups of linked amino acids in the binding site, the terminal peptide bonds were cut. The open valence on NH was capped with hydrogen and a methyl group was added to the carbonyl. The positions of all of the hydrogen atoms were optimized with the remainder of the atoms fixed in space. For the native antigen, dTdC, we performed DFT computations on both the crystal structure binding site geometry as well as the average structure from the MD simulations. The

DFT computations for the other dinucleotides were based on the average MD structure, followed by 10,000 steps of energy minimization. Both the interaction energy of the overall binding pocket and the interaction energies between the 5' based and key residues were evaluated. The interaction energy was computed as the difference in energy between the bound complex and the separated interacting species, all at the complex geometry. All DFT predicted interaction energies correspond to gas-phase computations, which should provide a reasonable model for the interactions within the hydrophobic core of the protein. However, we also evaluated solution-phase interaction energies for the 5' bases with the entire 5'-binding site, to gauge the effect of solvation. The resulting interaction energies are in qualitative agreement with the gas-phase computations (see Table A-3). All DFT computations were performed using Gaussian09 and utilized density fitting techniques.[271]

**Results and Discussion**

Binding of the Native Antigen (dTdC)

First, we discuss the binding of the native antigen, dTdC, by ED-10. In the crystal structure geometry (Figure III-1), B97-D applied to a model of the binding site indicates that the strongest interactions involving thymine are the π-stacking interactions with TrpH95 and TrpH50, followed by the hydrogen-bonding interaction with AsnH35 and the NH/π interaction with TyrH33 (see Table III-2). It is notable the hydrogen bonding interaction with AsnH35 provides less stabilization of this complex than the π-stacking interactions with Trp residues. As seen below, hydrogen bonding also plays

only a minor role in the selectivity for the 5'-base. For cytosine, the strongest interaction stems from the loop containing TrpH95, ThrH96, TyrH97, GlyH98 and SerH99, which together contribute about –13.8 kcal $mol^{-1}$ to the overall binding energy. This is complemented by strong hydrogen bonding interactions with GlyL91 and the π-stacking interaction between cytosine and TyrL32 (Table A-1 in Appendix). Overall, the DFT-predicted interaction energy between dT and the 5' binding pocket residues in ED-10 is –35.0 kcal $mol^{-1}$.

**Table III-2.** B97-D total and pairwise interaction energies (kcal $mol^{-1}$) between the 5' base and key amino acid residues for four dinucleotides with ED-10.[a]

| | B97-D | | | |
|---|---|---|---|---|
| Res. | dTdC | dAdC | dGdC | dCdC |
| TrpH50 | –8.6 (–8.4) | –6.0 | –9.6 | –5.6 |
| TrpH95 | –7.7 (–8.9) | –8.0 | –9.9 | –3.5 |
| AsnH35 | –7.8 (–6.7) | –9.2 | – | 0.4 |
| TyrH33 | –8.9 (–6.3) | –2.9 | –17.8[c] | –3.1 |
| GlyL91 | –2.5 (–2.6) | –3.6 | –2.5 | –3.2 |
| TyrH97 | –0.3 (–2.6) | –0.3 | –3.8 | –7.9 |
| Total[b] | –38.8 (–35.0) | –32.7 | –45.2 | –22.8 |

[a] For dTdC, B97-D interaction energies for the crystal structure geometry are provided in parentheses.
[b] Total B97-D interaction energy of the 5' base with the binding site of ED-10.
[c] Interaction of G with TyrH33-MetH34-AsnH35.

The binding site of ED-10 from the averaged structures over the 10ns trajectory showed some qualitative deviations from the crystal structure. These differences could arise in part to the presence of the C3' hydroxyl group on the cytosine nucleotide, which was absent in the reported crystal structure. During the dynamics simulation, the

conformation of the dinucleotide does not undergo significant changes, but is accompanied by movement of Phe96, Trp268, Ser276, and Ile101 towards dT. The result is a slight compression of the binding site. However, the primary intermolecular interactions depicted in Figure III-1 are maintained throughout the trajectory. B97-D interaction energies based on the average MD structure indicate some re-ordering of the important non-covalent interactions relative to those in the crystal structure geometry, and an overall slight enhancement of the predicted binding of thymine to –38.8 kcal $mol^{-1}$. In particular, based on the average structure, DFT predicts that the NH/π interaction between thymine and TyrH33 is now stronger than the stacking interactions with tryptophan.

**Table III-3.** MMGBSA pairwise interaction enthalpies between the 5' base and key amino acid residues, as well as total binding ethalpies (ΔH) and binding free energies (ΔG) for four dinucleotides with ED-10. All values in kcal $mol^{-1}$.

| | MMGBSA | | | |
|---|---|---|---|---|
| Res. | dTdC | dAdC | dGdC | dCdC |
| TrpH50 | –4.8 | –2.5 | –5.5 | –2.9 |
| TrpH95 | –4.6 | –4.7 | –4.7 | –1.9 |
| AsnH35 | –3.5 | –3.7 | –3.8 | 0 |
| TyrH33 | –3.2 | –1.4 | –4.3 | –2.2 |
| GlyL91 | –2.8 | –1.8 | –2.6 | –1.1 |
| TyrH97 | –1.1 | –1.9 | –1.2 | –2.8 |
| ΔH | –42.7 | –33.4 | –41.8 | –25.7 |
| ΔG | –20.3 | –11.6 | –19.6 | –5.6 |

Origin of 5'-Base Selectivity

The total MMGBSA predicted binding enthalpy for dTdC in ED-10 is –42.7 kcal $mol^{-1}$. The main favorable non-covalent interactions involving thymine are π-stacking interactions with TrpH50 and TrpH95, hydrogen bonding interactions with AsnH35 and GlyL91, and an NH/π interaction with TyrH33 (see Table III-3), which is generally consistent with the DFT results for the binding site cluster model. For cytosine, the most favorable interaction is the hydrogen bonding interaction between NH on HisL27 and the phosphate group on cytosine. The interaction between the hydroxyl group on TyrH97 and the deoxyribose-phosphate backbone of 3' cytosine is also highly favorable. Other key interactions are the stacking interaction with TyrL32 and hydrogen bonding interactions with GlyH98 (Table A-1 in Appendix). Accounting for entropic effects, the predicted binding free energy for dTdC is –20.3 kcal $mol^{-1}$.

Experimentally, Sanguineti *et al*. showed that altering the 5' nucleobase drastically reduces the binding affinity (see Table III-1).[249] Below, we consider the binding of 5'-dTdC-3' and three other dinucleotides (dAdC, dGdC, and dCdC) to ED-10, in order to quantify the interactions that underlie the strong selectivity for 5'-dT. Table III-3 shows the total MMGBSA binding free energies and enthalpies for these four dinucleotides, as well as the pairwise interactions between the 5' base and its surrounding residues. Predicted binding enthalpies relative to dTdC are listed in Table III-1. MMGBSA results also show that for all four dinucleotides, the interactions involving the 3' base (cytosine) are relatively unchanged (Table A-1). That is, the base

selectivity at the 5' end stems from changes in interactions with the 5' base itself, not secondary effects impacting the 3' binding pocket.

The MMGBSA data indicate that dAdC, dGdC, and dGdC are all bound less strongly than the native antigen (dTdC), as observed experimentally.[249] However, the magnitude of the change in binding observed experimentally (about 5 kcal $mol^{-1}$, regardless of the base)[49] is overestimated in the case of dCdC and underestimated for dGdC. In the former case, this is potentially due to rearrangements of TyrH97 and TrpH95 in the binding site upon binding of dCdC[249] that were not captured by the MD simulations. Regardless, these results should still provide qualitative insight into the origin of 5'-dT selectivity in ED-10.

**Table III-4.** Mean distances (in Å) for key non-covalent interactions.[a]

| Res. | Interaction | dTdC | dAdC | dGdC | dCdC |
|---|---|---|---|---|---|
| TrpH50 | π-stacking | 3.6 (98%) | 4.5 (0%) | 3.6 (99%) | 3.8 (79%) |
| TrpH95 | π-stacking | 3.6 (98%) | 3.6 (97%) | 3.7 (94%) | 5.4 (0%) |
| TyrH33 | XH/π | 2.8 | 3.5 | 2.4 | 2.6 |
| AsnH35 | H-bond | 1.9 (79%) | 2.2 (13%) | 2.3 (53%) | >4 (0%) |

[a] Values in parentheses are the percent of the MD snapshots in which the centroid-centroid distance was less than 4 Å for π-stacking interactions and the H···X distance was less than 2 Å for hydrogen bonds.

For dAdC, MMGBSA predicts a binding enthalpy that is 9.3 kcal $mol^{-1}$ less favorable than dTdC. This arises primarily from the reduced strength of the π-stacking and edge-to-face interactions with TrpH50 and TyrH33. The decreased strength of these interactions is reflected in the average distances between relevant groups (see Table III-

4), which show a considerably larger separation between adenine and both TrpH50 and TyrH33 compared with dTdC. The hydrogen bond to GlyL91, which primarily involves the phosphate group, is also reduced slightly with dAdC bound compared to dTdC, which is again reflected in both the MMGBSA pairwise interaction energy and the average distance. Notably, the other major interactions in the binding pocket show little change. In particular, the hydrogen bonding interaction with AsnH35 is at least as strong in dAdC as in the native antigen, dTdC. Overall, MMGBSA predicts that 2.3 kcal $mol^{-1}$ of the preferential binding of dT over dA arises from differences in π-stacking interactions with TrpH50, while differences between the edge-to-face interactions of the nucleobase with TyrH33 account for 1.8 kcal $mol^{-1}$ of the different in binding enthalpy. After inclusion of the entropic contribution, the predicted binding free energy for dAdC is 8.7 kcal $mol^{-1}$ less favorable than for dTdC.

These findings are further supported by B97-D interaction energies, which show that the interaction of adenine with the binding site is 6.1 kcal $mol^{-1}$ less favorable than that of thymine. As seen in the MMGBSA results, this drop is due primarily to reductions in the strength of the π-stacking interactions with TrpH50 and the edge-to-face interaction with TyrH33. In the latter case, the NH/π interaction between thymine and TyrH33 (–8.9 kcal $mol^{-1}$) is replaced with a much weaker CH/π interaction between adenine and TyrH33 (–2.9 kcal $mol^{-1}$), contributing significantly to the preferential binding of dTdC over dAdC. Similarly, B97-D predicts that the π-stacking interaction between adenine and TrpH50 is 2.6 kcal $mol^{-1}$ less favorable than the corresponding interaction involving thymine. Thus, even though some of the other interactions

operative in the binding site are stronger for adenine (most notably the H-bond to AsnH35), these enhanced interactions are overshadowed by the reduced π-stacking and XH/π interactions with aromatic amino acid side chains.

For dCdC, both B97-D and MMGBSA predict a precipitous drop in binding strength compared to dTdC, due to a reduction in the strength of nearly all interactions with the binding site. Indeed, the strongest interaction with the 5'-dC is the π-stacking interaction with TrpH50, which is predicted to be 2 kcal $mol^{-1}$ weaker than for dTdC. Given the lack of a suitable H-bond acceptor on cytosine, the H-bonding interaction with AsnH35 observed for dTdC and dAdC is not present with dCdC bound. Indeed, B97-D predicts that the interaction between AsnH35 and the 5' cytosine is slightly repulsive. These significantly reduced interaction energies are offset slightly by the presence of TyrH97, which provides a favorable interaction with 5'-dC that is not present with any of the other bound dinucleotides. Overall, the present results indicate that selectivity for 5'-dT over dC arises from a combination of hydrogen bonding, edge-to-face, and π-stacking interactions.

For dGdC, MMGBSA predicts only a 0.9 kcal $mol^{-1}$ reduction in binding enthalpy, and a 0.7 kcal $mol^{-1}$ reduction in binding free energy, relative to dTdC. Moreover, B97-D predicts a more favorable overall interaction between guanine and the binding pocket compared to thymine. In particular, the π-stacking interactions of guanine with TrpH50 and TrpH95 and the NH/π interaction with TyrH33 are both predicted to be more favorable compared to the analogous interactions with thymine, and the π-stacking interactions present in the case of guanine are predicted by B97-D to be the most

favorable of any of the dinucleotides. These favorable effects for dGdC are somewhat compensated for by other effects. However, overall neither MMGBSA nor B97-D reproduce the experimentally observed 5 kcal $mol^{-1}$ drop in binding affinity for dGdC compared to dTdC.

We have investigated two potential sources of this discrepancy. First, we note that in all the above complexes, the 5' nucleosides (dT, dA, dC, and dG) adopt *anti* conformation throughout the simulations. We performed analogous simulations with the 5' nucleosides with a *syn* conformation, to test whether the non-native antigens, and dGdC in particular, binds in a different confirmation than dTdC. However, MMGBSA predicted binding enthalpies were less favorable for the *syn* conformer, compared to the anti conformer, for each of the dinucleotides considered. Thus, it does not appear that the 5'-base of the non-native antigens adopt a different conformation than in dTdC. Second, we carried out multiple trajectory MMGBSA computations to investigate whether differences in confirmations of the bound and free dinucleotides were contributing to the experimental observation of reduced binding free energies for the other dinucleotides. As seen with the single trajectory results, these multiple trajectory simulations (see Table A-2) show that as 5'-dT is replaced with either dA or dC, the binding enthalpy drops by a significant amount. However, dGdC is predicted to bind slightly more strongly than the native antigen dTdC. That is, both multiple and single trajectory approaches predict similar binding trend for ssDNA.

Impact of TrpH50 Mutations

The above results demonstrated that π-stacking interactions between the 5' nucleobase and TrpH50 contribute strongly to the preferential binding of dTdC over dAdC. To probe the importance of this interaction further, we considered mutation of TrpH50 to Ala, Phe, or Tyr (see Table III-5).

For all three point mutants, the binding enthalpy of dTdC becomes less favorable, while the binding of dAdC is predicted to be more favorable. In other words, although these point mutants are still predicted to be selective for 5'-dT, these mutations reduce the selectivity compared to the native antibody ED-10. The largest change in selectivity is predicted for the AlaH50 mutant, for which dTdC is predicted to be bound by only 2.8 kcal $mol^{-1}$ more than dAdC, compared to the 9.3 kcal $mol^{-1}$ difference for native ED-10.

The origin of this reduced selectivity can be seen by again examining the individual interactions within the binding pockets. For all three point mutants, the interactions in the cytosine pocket are unperturbed relative to those in the native antibody. Moreover, most of the interactions present in the binding site change very little, and the overall changes in binding enthalpy are due primarily to differences in the interaction of the 5'-base with the H50 residue as well as secondary effects on the interaction with TyrH33. In particular, for the AlaH50 mutant, the reduced selectivity for dTdC over dAdC arises from changes in the interactions involving GlyL91 and TyrH33.

**Table III-5.** MMGBSA predicted total binding enthalpy and pairwise interactions (kcal $mol^{-1}$) between the 5' base and key amino acid residues for TrpH50 mutants and the dTdC and dAdC dinucleotides.

| dTdC | TrpH50 (native) | AlaH50 | PheH50 | TyrH50 |
|---|---|---|---|---|
| ResH50 | –4.8 | –0.9 | –3.7 | –3.8 |
| TrpH95 | –4.6 | –4.2 | –4.5 | –4.5 |
| AsnH35 | –3.5 | –3.5 | –3.5 | –3.5 |
| TyrH33 | –3.2 | –3.5 | –3.5 | –3.2 |
| GlyL91 | –2.8 | –2.9 | –2.7 | –2.8 |
| TyrH97 | –1.1 | –1.0 | –1.4 | –1.3 |
| Total | –42.7 | –40.4 | –41.5 | –41.3 |
| dAdC | | | | |
| ResH50 | –2.5 | –0.6 | –1.6 | –1.3 |
| TrpH95 | –4.7 | –4.5 | –4.8 | –4.6 |
| AsnH35 | –3.7 | –3.9 | –3.8 | –3.1 |
| TyrH33 | –1.4 | –1.2 | –1.4 | –0.9 |
| GlyL91 | –1.8 | –2.3 | –2.4 | –2.9 |
| TyrH97 | –1.9 | –0.9 | –1.6 | –2.1 |
| Total | –33.4 | -37.6 | –35 | –35.9 |

**Summary and Concluding Remarks**

Binding free energies and enthalpies for four dinucleotides (dAdC, dGdC, dCdC, and the native antigen, dTdC) by ED-10 were computed using MMGBSA, and further analyzed using DFT applied to cluster models of the binding sites. Overall, MMGBSA provides relative binding enthalpies that are in qualitative agreement with experimental data for two of the three non–native dinucleotides (dAdC and dCdC), while for dGdC, MMGBSA appears to overestimate the binding affinity relative to dTdC. The MMGBSA data show that the affinity of ED-10 for the 3' cytosine is relatively constant across the four dinucleotides, and the strong selectivity for dTdC arises primarily through differences in interactions within the 5' binding site.

Analyses of the non–covalent interactions operative in the 5’ binding site reveal a number of key interactions. In particular, MMGBSA results show that for dTdC and dGdC, stacking interactions of the 5’ base with TrpH50 and TrpH95 contribute a total of about –10 kcal $mol^{-1}$ to the binding enthalpy. However, the other two dinucleotides (dAdC and dCdC) show weaker overall binding, and π-stacking interactions of the 5’ base with TrpH50 are reduced significantly. Hence, π-stacking interactions with TrpH50 not only contribute to dinucleotide binding, but are also central to the 5’-base selectivity of ED-10. Overall, the strong selectivity of ED-10 for dTdC over dAdC arises primarily from differential π-stacking interactions with TrpH95 and differences in XH/ interactions between the nucleobase and TyrH33. We note that this Tyr residue is not present in other ssDNA-binding autoantibodies, suggesting that these other antibodies will not exhibit the same strong 5’-base selectivity as ED-10. Similarly, MMGBSA computations predict that mutations of TrpH50 will provide measurable reductions in the selectivity for dTdC over dAdC, providing an experimental means of quantifying the role of π-stacking interactions in the 5’-base selectivity of this DNA-binding autoantibody.

Overall, both MMGBSA and B97-D data show that π-stacking, CH/π, NH/π, and hydrogen bonding interactions all contribute to the binding of the 5’-base in complexes of dinucleotides with ED-10. However, these results demonstrate that the strong selectivity for 5’-dT sequences is primarily a result of differences in aromatic interactions (XH/ π and π-stacking interactions), not hydrogen-bonding.

# CHAPTER IV

# QUANTIFYING THE Π-STACKING INTERACTIONS IN NITROARENE BINDING SITES OF PROTEINS*

## Introduction

Nitroarenes, which are widely used components of explosives, dyes, pesticides and pharmaceutical feedstocks, are also common environmental pollutants and potent carcinogens.[272,273] In nature, nitroarenes are metabolized primarily through bacterial degradation, for which they serve as a source of carbon, nitrogen, and energy.[273-276] In humans, nitroarenes can interact with proteins and DNA, leading to mutagenicity and cytotoxicity.[277] For instance, 1-10% of myeloma proteins bind nitrophenyl ligands.[278] Despite the prevalence of nitroarene binding sites in proteins, little is known about the nature of these binding sites, or how nitroarenes interact with aromatic amino acid side chains.

Some nitroaromatic compounds are bioactive, and are present in a limited number of pharmaceuticals.[279] For example, nitroarene-containing natural products exhibit antibiotic[272,280] and antifungal[281,282] activity. Similarly, synthetic nitroarenes have showed potential as therapeutic agents.[283] For example, Miller and co-workers recently devised anti-tuberculosis agents based on nitroaromatic compounds,[284] while Mânási-Csizmadia *et al.* synthesized a non-cytotoxic and photostable analog of the well-known myosin II inhibitor blebbistatin by nitrating C15.[285] Even though nitroaromatic

* Submitted to *Journal of Physical Chemistry B*. Copyright © 2015 American Chemical Society.

compounds are not ordinarily produced biosynthetically, nitroarene binding sites abound in human cells. For instance, in the human bloodstream, 1% of all immunoglobulins react with nitroarene epitopes,[286,287] and anti-dinitrophenyl (anti-DNP) antibodies have high prevalence in humans.[288,289] Ligands exploiting these nitroarene binding sites have been designed to mediate immunogenicity,[290-292] and dinitrophenyl containing ligands have been designed to target a number of diseases, including HIV,[293] lung cancer,[294-297] prostate cancer,[298] glioblastoma,[299] and B cell lymphoma.[300,301]

The Spiegel group has demonstrated the power of using nitroarene ligands as antibody recruiting molecules (ARM).[302-305] In 2009, Spiegel *et al.* reported a new antibody-recruiting nitroarene ligand targeting HIV gp120.[306] A ternary complex formed between anti-DNP antibodies, an ARM featuring a dinitrophenyl ring, and target cells was able to inactivate the virus. In the same year, Spiegel and co-workers designed a similar bifunctional small molecule drug targeting prostate cancer cells.[298] This ligand binds prostate specific membrane antigen (PSMA) on the membrane of prostate cancer cells, recruiting an anti-DNP antibody via its nitroarene terminus. In 2010 a previously unreported remote nitroarene binding site on PSMA was discovered by Spiegel and co-workers.[307] This binding site features a nitroarene-tryptophan stacking interaction, which was postulated to enhance the ligand binding affinity with PSMA. More recently, Pires *et al.* developed a D-amino acid antibody recruitment therapy (DART) approach utilizing anti-DNP antibodies.[308] This novel strategy involves the incorporation of non-natural D-amino acids onto the cell surfaces of bacteria, which then redirect antibodies to the bacterial cell surface. This provides a potential means of eliminating Gram-

positive bacteria by leveraging the power of the host immune system. Both of these novel biological applications could benefit from the more refined understanding of the binding of nitroarenes by proteins that can be provided by computational studies.

The majority of previous computational studies of substituent effects in π-stacking interactions have focused on the benzene dimer, providing insights into the factors that impact the strength of these interactions. Notably, Sherrill and co-workers[113,309] demonstrated strict additivity of substituent effects in stacked benzene dimers. That is, the impact of several substituents on a given stacked dimer is simply the sum of effects of individual substituents. We have shown that these, and other trends in substituent effects in π-stacking interactions, are readily explained by the local, direct interaction model of substituent effects in π-stacking interactions.[78,99,100,115,119,120] In contrast to venerable models of π-stacking interactions that center on the π-electron density of the interacting rings,[114,310-312] substituent effects in diverse π-stacking interactions can be explained in terms of direct, local interactions between the substituents and the other ring.[78,99,100,115,119,120]

These and other advances in our understanding of π-stacking interactions can be brought to bear on π-stacking interactions in biological contexts.[313,314] Wetmore and co-workers[233-236,254,315-324] have used computations to study π-stacking interactions of aromatic amino acid side chains with increasingly sophisticated models of DNA, providing key insights into DNA-repair mechanisms and DNA intercalation phenomena, among others. Tschumper *et al.*[237,325] recently presented computational studies of π-stacking interactions of phenylalanine and tyrosine with adenine in protein structures.

However, there has been no analogous study of the interactions of nitroarenes with aromatic amino acid side chains.

Many protein structures featuring bound nitroarene ligands have been deposited in the protein data bank (PDB), representing numerous families of enzymes (monooxygenases, dioxygenases, and reductases, among others). Two representative examples are depicted in Figure IV-1. In Figure IV-1a, a substituted *m*-dinitrophenyl ring is stacked with a Trp residue; in Figure IV-1b a similar nitroarene is stacked with a Trp and engages in a loose stacking interaction with a nearby Tyr residue. Despite the prevalence of nitroarene binding sites in proteins, and the demonstrated utility of these binding sites in the development of novel therapies,[294-308] there have been no detailed analyses of the non-covalent interactions responsible for the binding of nitroarenes by proteins. As such, a thorough understanding of non-covalent interaction between nitroarenes and proteins is needed.

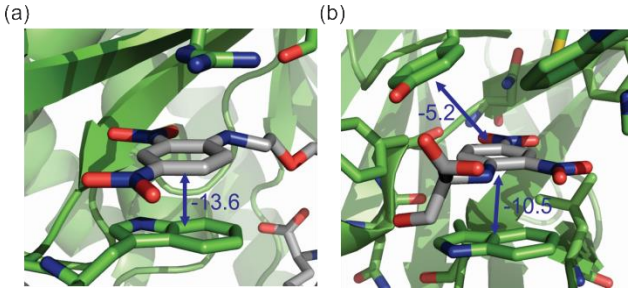



**Figure IV-1.** Representative nitroarene binding sites in proteins: (a) the remote DNP binding site of ARM in PMSA from Spiegel and co-workers (PDB: 2XEF)[307] and (b) a dinitrophenol binding site in the antibody SPE7 from Tawfik and co-workers (PDB: 1OAU)[326]. In this case, the nitroarene is sandwiched between Trp and Tyr side chains. In both cases, the strength of π-stacking interactions are provided in kcal $mol^{-1}$.

In 2009, Spiegel and co-workers[298] published a limited analysis of nitroarene binding sites in the PDB, reporting that 16 of 32 available nitroarene-protein complexes exhibit π-stacking interactions between the nitroarene and aromatic amino acid side chains. Herein, we present a systematic study of model π-stacking interactions between nitrobenzenes and aromatic amino acid side chains in protein, accompanied by a much broader analysis of nitroarene binding sites present in proteins in the PDB. The results reveal that nitroarene binding sites are prevalent in protein structures, with interactions between aromatic amino acid side chains contributing as much as 14 kcal $mol^{-1}$ to the binding of nitroarenes. Among the aromatic amino acids, Trp is particularly well-represented in these binding sites, which is consistent with our finding that Trp side chains provide stronger stacking interactions with nitroarenes than His, Phe, or Tyr.

**Theoretical Methods**

Stacking interactions between nitrobenzene (1-$NO_2$), *m*-dinitrobenzene (2-$NO_2$m), *p*-dinitrobenzene (2-$NO_2$p), and 1,3,5-trinitrobenzene (3-$NO_2$) and aromatic amino acids side chains (histidine, phenylalanine, tryptophan and tyrosine) were computed at several different levels of theory. The amino acid side chains were modeled as shown in Scheme IV-1. In order to locate all stacked complexes involving these rings, initial geometries were generated by separating the amino acid and nitroarene by 4.0 Å, with the two systems parallel. We then explored all possible relative orientations at 30 ° increments at each of five different starting locations.[327] Geometries were then fully optimized at the B97-D/TZV(2*d*,2*p*) level of theory. Binding energies were computed as the difference in

energy between the optimized dimer and the optimized monomers for all unique complexes, which were defined as the set of structures for which the root mean squared deviation (RMSD) with all other structures exceeds 0.5 Å. The binding energies for the global minimum structure for each nitroarene-amino acid pair were further quantified by higher-level *ab initio* methods. In particular, counterpoise-corrected CCSD(T)/aug-cc-pVTZ interaction energies were esimated by

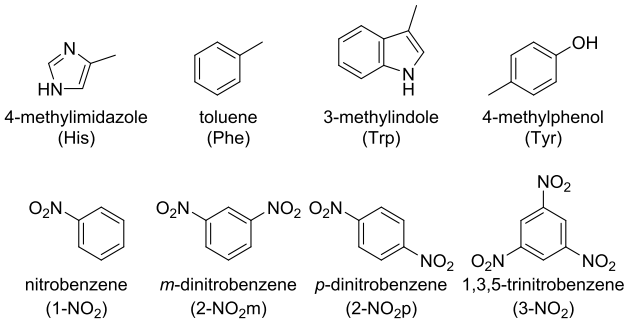
$$\text{CCSD(T)/AVTZ} \approx \text{MP2/AVTZ} + [\text{CCSD(T)/jun-cc-pVDZ} - \text{MP2/jun-cc-pVDZ}]$$

where AVDZ denotes the augmented correlation-consistent basis set (aug-cc-pVDZ) of Dunning[328] and jun-cc-pVDZ is the partially-augmented subset of the aug-cc-pVDZ basis devised by Papajak and Truhlar.[329] Symmetry adapted purtubation theory computations[109,110,330,331] (SAPT0) were also performed using the jun-cc-pVDZ basis set at these lowest-lying B97-D optimized structures. This level of SAPT-theory has been shown to provide excellent predictions across a broad range of non-covalent interactions at a very modest computational cost.[332]

Stacking interactions between nitroarenes and aromatic amino acids were also culled from crystal structures in the PDB. We identified protein structures with ligands containing a benzene ring with at least one nitro group, X-ray resolution less than 2.5Å, and protein sequence identity less than 70%. Aromatic residues (His, Phe, Trp, and Tyr) were identified that were engaged in π-stacking interactions with each nitroarene ligand. The position of aromatic amino acid residues relative to the bound nitroarenes were

characterized based on the intermolecular coordinates shown in Figure IV-2, and π-stacking interactions were defined as those for which R < 5Å, α < 45°, and β < 45 °. All other structures were excluded from the subsequent analyses.

**Scheme IV-1**




The structures extracted from the PDB files were further refined based on the structures of protein and ligand. First, in nearly all of the identified structures, the nitroarene is part of a larger ligand. To isolate the interaction of the protein with just the nitroarene component, we removed the non-nitroarene part at the closest carbon to the nitrophneyl ring (usually the first or the second carbon depending on the structure of the ligand). Small substituents (1-2 heavy atoms) on the nitroarene rings were retained. Ligands in which the nitroarene was fused with another aromatic ring were not considered in our analyses. Second, in cases in which the protein was an oligmer and the same nitroarene-amino acid pair occurs multiple times, we considered only one instance of this interaction in our analyses. Finally, in proteins in which the nitroarene engaged in

stacking interactions with multiple aromatic amino acids, we considered each interaction separately. The final number of stacking interactions meeting these criteria was 65.

To evaluate interaction energies for these 65 stacked complexes, hydrogens were added and their positions optimized at the B97-D/TZV(2*d*,2*p*) level of theory. Interaction energies were computed for the resulting "crystal structure" geometries as the energy difference between the complex and the separated monomers at the dimer geometry. These complexes were also subjected to full geometry optimizations to the nearest local minimum, and the binding energies evaluated at the B97-D/TZV(2*d*,2*p*) level of theory.

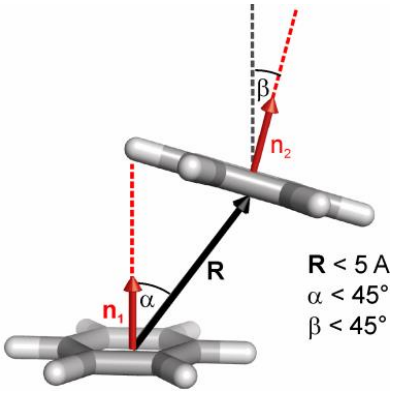



**Figure IV-2.** Intermolecular coordinates used to define stacking interactions between aromatic amino acid side chains and nitroarenes.

DFT computations were performed using Gaussian 09,[333] while the CCSD(T) and MP2 energies were computed using Molpro.[334] SAPT0 computations were carried out

using Psi4.[335] All *ab initio* energies were counterpoise corrected,[336] while the B97-D energies were not.

**Results and Discussion**

Model Stacking Interactions

In total, 94 unique stacked dimers were located involving the four nitroarenes (1-$NO_2$, 2-$NO_2$m, 2-$NO_2$p, and 3-$NO_2$) and the four aromatic amino acid side chains (His, Phe, Trp, and Tyr). The distribution of binding energies of these unique complexes are displayed in Figure IV-3. The number of unique complexes for each amino acid/nitroarene pair was rather heterogeneous, ranging from only two for His···3-$NO_2$ to nine each for Tyr with 1-$NO_2$ and 2-$NO_2$m. Overall, the numbers of energy minima reflect the symmetry of the nitroarene, as expected, with fewer unique complexes for 2-$NO_2$p and 3-$NO_2$ compared to the less-symmetric 1-$NO_2$ and 2-$NO_2$m. For most of the dimers, there is a significant spread in binding energies depending on the relative orienation of the two stacked sytems.

The B97-D optimized structures of the lowest-lying complex for each amino acid/nitroarene pair are shown in Figure IV-4, along with CCSD(T) predicted binding energies. These CCSD(T) data are listed in Table III-1, along with the corresponding B97-D/TZVP(2*d*,2*p*) SAPT0 interaction energies. Overall, the B97-D energies are very strongly correlated with the CCSD(T) data (r = 0.98), lending credence to the use of this dispersion-corrected DFT method to optimize geometries. SAPT0 interaction energies are listed in Table III-1, along with the contribution of electrostatic, exchange-repulsion,

induction, and dispersion effects to the total interaction. The SAPT0 interaction energies are also very strongly correlated with the CCSD(T) binding energy data (r = 0.995), again supporting the use of this method to further understand these interactions. Overall, the SAPT data indicate that these interactions are driven by a combination of electrostatic and dispersion interactions. That is, removal of either of these two components results in very poor correlations with the total interaction energies, so both are vital for describing the magnitude of these non-covalent interactions.

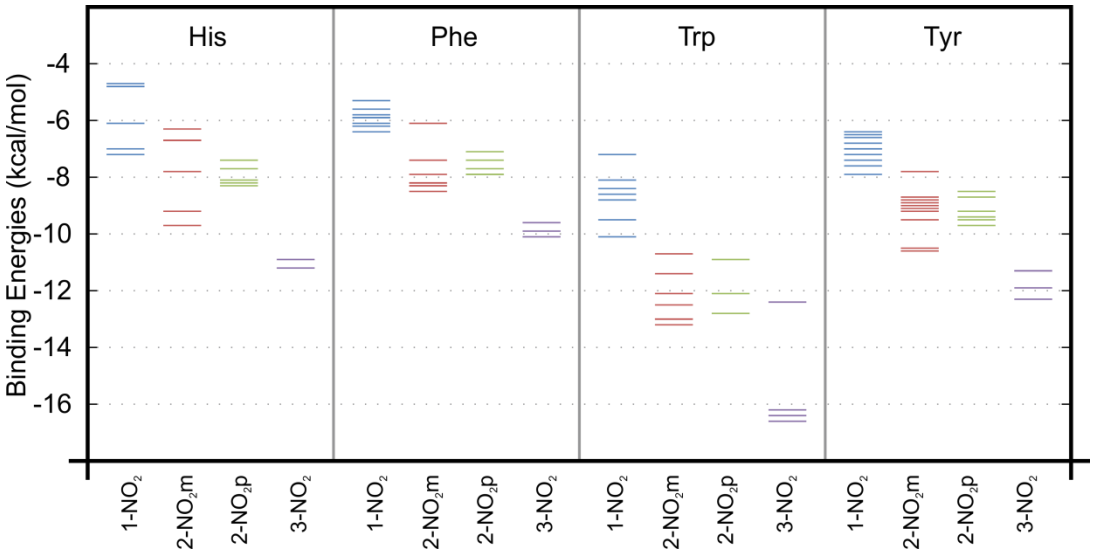



**Figure IV-3.** Distribution of B97-D/TZV(2*d*,2*p*) predicted binding energies for model stacked complexes between 1-NO2, 2-NO2m, 2-NO2p, and 3-NO2 with His, Phe, Trp, and Tyr side chains.

More insight can be obtained by looking at trends in the CCSD(T) interaction energies for a given amino acid side chain across all four nitroarenes. For each amino

acid, the interaction energy gets more favorable as the number of nitro groups increases, accompanied by a contraction of distance between the nitroarene and amino acid side chain by about 0.2 Å going from 1-$NO_2$ to 3-$NO_2$. These data are consistent with the well-established tendency of nitro-substituents to enhance π-stacking interactions.[309-312,337,338] However, the increase in interaction energy with the number of $NO_2$-groups is not uniform, in contrast to results for model stacked benzene dimers.[113,339] For example, the stacking interaction in 2-$NO_2$m···His is 2 kcal $mol^{-1}$ more favorable than that of 1-$NO_2$···His, yet the stacking interaction is only enhanced by 1 kcal $mol^{-1}$ upon addition of the third nitro group in 3-$NO_2$···His. This lack of additivity highlights the local nature of substituent effects in these π-stacking interactions, and the accompanying dependence of substituent effects on the local orientaiton of stacked rings.[119] More precisely, the introduction of additional nitro groups does not always lead to the same degree of enhancement due to differences in available local, direct interactions between the nitro group and functionality on the amino acid side chain.

The importance of local, direct interactions is also reflected in the geometries of these stacked dimers. In all cases, the local, direct interaction present in the 1-$NO_2$ dimer is preserved in the more heavily substituted dimers. For example, the placement of the $NO_2$ group directly above the NH of His in the 1$NO_2$···His dimer is also present in the global minimum complexes of His with 2-$NO_2$m, 2-$NO_2$p, and 3-$NO_2$. Similarly, the $NO_2$···$CH_3$ interaction present in the 1-$NO_2$···Phe dimer is also present in the other Phe complexes. The complexes involving Trp exhibit both of these interactions, although orienting the $NO_2$ group above the NH of the indole ring appears to take precedent. As a

result, the lowest-lying complex between 1-$NO_2$ and Trp features this $NO_2$···NH, whereas both this interaction and the Phe-like interaciton of the nitro group with the methyl group of Trp are present in the other Trp complexes. Similarly, for the Tyr complexes, the $NO_2$···OH interaction seems to dominate the direct interaction between the nitro and methyl groups that appears in the more substituted cases.

In the case of dinitrobenzene, 2-$NO_2$m exhibits stronger interactions with all four amino acid side chains than does 2-$NO_2$p. One could attribute this 1 kcal $mol^{-1}$ energy difference to the lack of a molecular dipole moment in *p*-dinitrobenzene. However, such an explanation falls flat when applied to 3-$NO_2$, which has no net dipole yet engages in significantly stronger interactions with the amino acid side chains than either 2-$NO_2$m or 2-$NO_2$p. Similarly, 2-$NO_2$p has no molecular dipole, yet exhibits much stronger interactions with the amino acid side chains than the decidedly polar 1-$NO_2$. Instead, the enhanced interaction of 2-$NO_2$m compared to 2-$NO_2$p results from geometric effects—the two nitro groups in *m*-dinitrobenzene are more ideally positioned to engage in favorable local, direct interactions with the key features of these four amino acids.

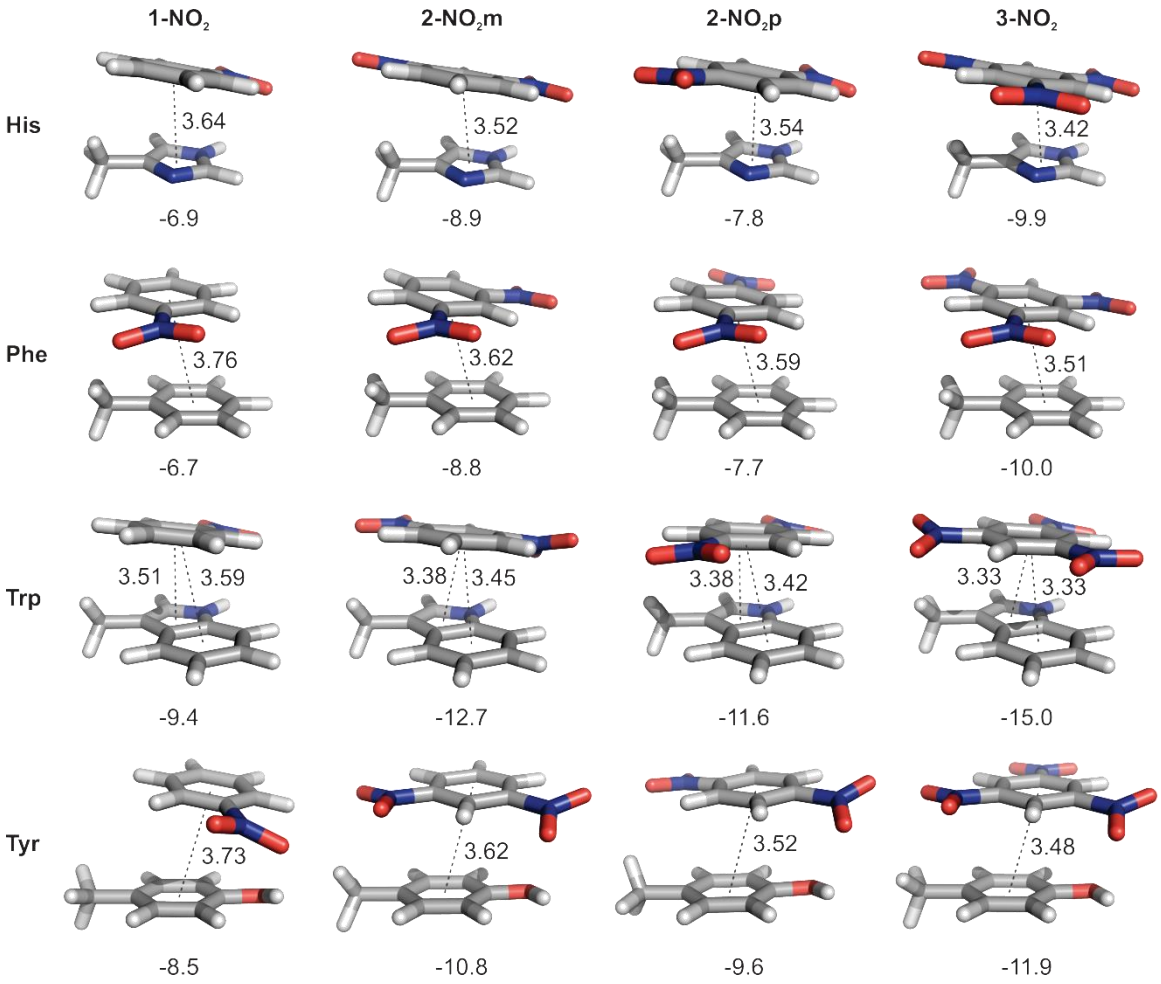



**Figure IV-4.** Global minimum energy structures for model stacked dimers of His, Phe, Tyr, and Trp amino acid side chains with 1-NO2, 2-NO2m, 2-NO2p, and 3-NO2 optimized at the B97-D/TZV(2d,2p) level of theory. Estimated CCSD(T)/aug-cc-pVTZ binding energies are provided in kcal $mol^{-1}$, while distances between ring centroids are given in Angstroms.

Considering instead the interactions of a single nitroarene with each of the aromatic amino acids, we find that the interaction energy follow the trend Trp > Tyr > Phe ≈ His. The reason can be attributed to a combination of dispersion interaction and direct interactions between the nitro-substituents and the local dipoles associated with

the heteroatoms in His and Trp. For example, even though complexes with phenylalanine have more favorable net dispersion interactions due to the larger size of the ring, this favorable interaction is compensated by direct interactions between the nitro-substituents of the nitroarenes and the NH group of the His. Similarly, even though both tryptophan and tyrosine exhibit favorable direct interactions with the nitro group (involving the NH group for Trp, and the OH for Tyr), the total interaction energy is greater for tryptophan complexes due to the additional dispersion interactions enabled by the larger surface area of Trp compared to Tyr.

These data enable other comparisons among different amino acids, yielding additional insight into the impact of substituents on π-stacking interactions in realistic stacked complexes. For example, the presence of the hydroxyl group in Tyr provides a nearly constant 2 kcal $mol^{-1}$ enhancement in the stacking interaction compared to Phe. This arises in part through weak OH···ON interactions, evidenced by the distortion of one of the nitro groups from planarity in all four of the complexes with Tyr. Moreover, the presence of the OH group in Tyr alters the prefered orientation of the nitroarene. This is most obvious for *p*-dinitrobenzene, which is rotated almost 90 ° between the Phe and Tyr global minimum structures; this stark difference in preferred stacking orientations will presumably play a role in the binding poses of nitroarenes in nitroarene binding sites featuring Phe vs Tyr residues.

**Table IV-1.** SAPT0/jun-cc-pVDZ, B97-D/TZV(2*d*,2*p*), and estimated CCSD(T)/aug-cc-pVTZ interaction energies, as well as SAPT0 components for stacked dimers of His, Phe, Trp, and Tyr amino acid side chains and 1-$NO_2$, 2-$NO_2$m, 2-$NO_2$p, and 3-$NO_2$.

| | | $E_{elec}$ | $E_{exch}$ | $E_{ind}$ | $E_{disp}$ | $E_{SAPT0}$ | $E_{B97D}$ | $E_{CCSD(T)}$ |
|---|---|---|---|---|---|---|---|---|
| His | 1-$NO_2$ | –7.2 | 11.3 | –1.7 | –10.2 | –7.7 | –7.2 | –6.9 |
| | 2-$NO_2$m | –10.0 | 15.5 | –2.8 | –12.9 | –10.3 | –9.6 | –8.9 |
| | 2-$NO_2$p | –8.4 | 13.2 | –2.2 | –11.5 | –8.8 | –8.3 | –7.8 |
| | 3-$NO_2$ | –11.7 | 17.7 | –3.9 | –14.2 | –12.2 | –11.2 | –9.9 |
| Phe | 1-$NO_2$ | –5.7 | 11.0 | –1.3 | –11.3 | –7.3 | –6.4 | –6.7 |
| | 2-$NO_2$m | –8.6 | 14.3 | –1.9 | –13.9 | –10.2 | –8.5 | –8.8 |
| | 2-$NO_2$p | –7.7 | 13.4 | –1.8 | –13.0 | –9.1 | –7.6 | –7.7 |
| | 3-$NO_2$ | –10.3 | 16.1 | –2.7 | –15.2 | –12.1 | –10.1 | –10.0 |
| Trp | 1-$NO_2$ | –10.0 | 17.6 | –2.5 | –16.6 | –11.6 | –9.6 | –9.4 |
| | 2-$NO_2$m | –14.4 | 22.8 | –3.9 | –20.0 | –15.6 | –13.2 | –12.7 |
| | 2-$NO_2$p | –13.6 | 22.6 | –3.8 | –19.9 | –14.7 | –12.9 | –11.6 |
| | 3-$NO_2$ | –18.0 | 28.2 | –6.1 | –23.6 | –19.6 | –16.6 | –15.0 |
| Tyr | 1-$NO_2$ | –9.0 | 12.9 | –2.0 | –11.6 | –9.7 | –7.9 | –8.5 |
| | 2-$NO_2$m | –11.6 | 17.0 | –2.8 | –15.1 | –12.5 | –10.6 | –10.8 |
| | 2-$NO_2$p | –11.1 | 16.6 | –2.6 | –14.3 | –11.3 | –9.7 | –9.6 |
| | 3-$NO_2$ | –13.6 | 19.5 | –3.8 | –16.6 | –14.5 | –12.2 | –11.9 |

Nitroarene Binding Sites in Proteins

The PDB contains a large number of proteins with nitroarene-containing ligands bound, and an analysis of these structures reveals several trends.[340] An examination of the distribution of amino acid residues in the 216 nitroarene binding sites revealed that the following amino acids represent more than 5% of binding site residues: Phe 8%; Tyr, Leu 7%; and Trp, Ser, Val, Gly 6%. The importance of aromatic amino acids in nitroarene binding is made more apparent by considering the proportion of each of the aromatic amino acid side chains that are located in the binding pocket, compared to all

instances of these amino acids in the proteins: Trp 17% > His 11% > Tyr 10% ≈ Phe 10%.

Of the 216 crystal structures containing nitroarene ligands, 191 have nearby aromatic amino acids. From these, we extracted 65 unique stacking interactions between an aromatic amino acid and nitroarenes that was not part of fused rings (see Table IV-2). These stacking complexes contain predominantly nitrophenyl rings, although seven structures contain *m*-dinitrophenyl groups. Trp residues are the most well-represented of the aromatic amino acid side chains in these stacking interactions, being involved in 23 of the 65 (35%). This is perhaps unsurprising in light of the model complexes studied above, which indicate that Trp engages in far stronger stacking interactions with nitroarenes than any of the other aromatic amino acid side chains. Trp is followed by Phe and Tyr residues, which are involved in 20 (31%) and 17 (26%) of the stacking interactions, respectively. His is far less commonly found among these stacked complexes in the PDB, appearing only five times (8%). This was initially somewhat surprising, since in the model complexes, stacking interactions between nitroarenes and the His sidechain is of similar strength to Phe, and not much weaker than Tyr. However, the small size of the imidazole ring of His, as well as the possibility of strongly favorable NH···π interactions, result in His having a greater tendency to engage in T-shaped interactions over π-stacking interactions when complexed with an arene. Such geometries were excluded from our analysis of the PDB structures, and likely account for the dearth of π-stacking interactions of His with nitroarenes in the analyzed nitroarene binding sites.

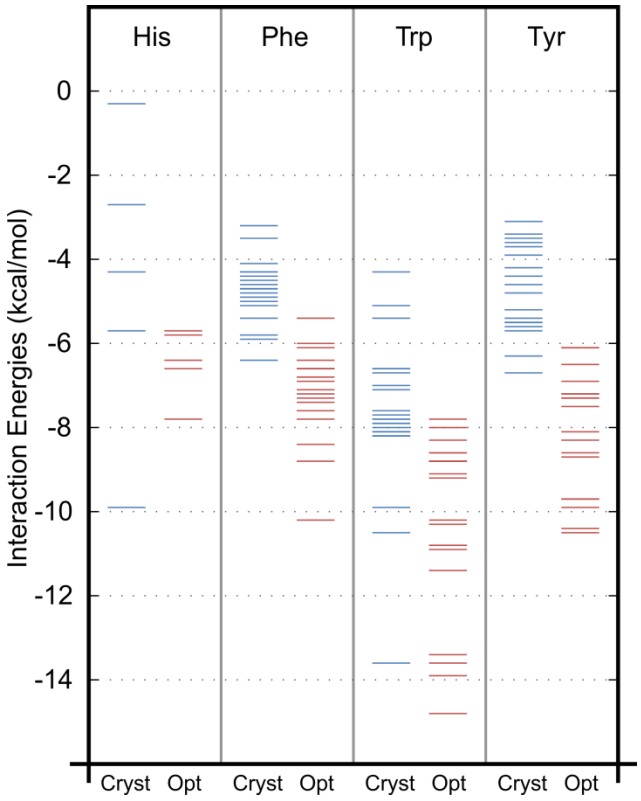



**Figure IV-5.** B97-D/TZV(2*d*,2*p*) predicted interaction energies for stacked dimers of nitroarenes with aromatic amino acid side chains (His, Phe, Trp, and Tyr) in the crystal structure geometry ('Cryst', blue lines) and the binding energies for the corresponding relaxed gas-phase energy minima ('Opt', red lines).

We have quantified the stacking interactions present in these binding sites at the B97-D/TZV(2*d*,2*p*) level of theory. Table IV-2 lists computed interaction energies for the crystal structure geometries, as well as the binding energies of the corresponding local minima resulting from full optimizations of these nitroarene-amino acid dimers. These data are plotted in Figure IV-5. The trend in energies both in the crystal structure geometries and the corresponding relaxed gas-phase dimers follow the results for the

model stacked dimers. That is, stacking interactions in nitroarene binding sites follow the trend Trp > Tyr > Phe ≈ His.

There is often significant deviation from the interaction energies in the crystal structure geometries and the binding energies of the corresponding gas-phase minima. For example, the stacking interaction between His and the nitroarene in 1NO3 is barely favorable (–0.3 kcal $mol^{-1}$), whereas the corresponding relaxed structure has an interaction energy of –5.7 kcal $mol^{-1}$. Other cases are less extreme, although similarly notable. For example, 3MK0 exhibits a stacking interaction between a Tyr side chain and a nitroarene, which is predicted to contribute –3.4 kcal $mol^{-1}$ to the binding. However, upon relaxation, this resulting binding energy of –8.7 kcal $mol^{-1}$ is more than 60% stronger. These differences mirror those found by Tschumper *et al*.[237,325] in their studies of stacking interactions of Tyr and Phe side with adenine in protein complexes, and could be due to a number of factors. The most likely contributor is the impact of other binding site residues as well as the additional components of the nitroarene ligand that were excised in these model complexes. The binding in the protein will obviously reflect the most favorable net interaction between the entire ligand and protein (not to mention entropic effects), whereas we are only considering the stacking interactions with the nitroarene component. Finally, in the case of His residues, we have assumed the most common protonation state, whereas a number of protonation states could be prevalent in the protein.

**Table IV-2.** B97-D/TZV(2*d*,2*p*) interaction energies (kcal mol$^{-1}$) for stacked dimers of 1-$NO_2$ and 2-$NO_2$m with His, Phe, Trp, and Tyr amino acid side chains found in the PDB. The interaction energies in the crystal structure geometry (Cryst) are provided, along with the binding energies of the corresponding gas-phase optimized energy minima (Opt).[a]

| Structure | Amino Acid | Cryst | Opt | Structure | Amino Acid | Cryst | Opt |
|---|---|---|---|---|---|---|---|
| 1NO3 | His | –0.3 | –5.7 | 2PDJ | Trp | –8.2 | –8.6 |
| 2Q95 | His | –2.7 | –5.8 | 2PDP | Trp | –7.7 | –8.8 |
| 3CHT | His | –9.9 | –7.8 | 2PDU | Trp | –8.2 | –8.8 |
| 3OJK | His | –4.3 | –6.6 | 2PZN | Trp | –8.2 | –8.8 |
| 43CA | His | –5.7 | –6.4 | 2V96 | Trp | –7.6 | –8.3 |
| 1AI5 | Phe | –4.3 | –7.1 | 2VCZ | Trp | –8.0 | –8.6 |
| 1GLQ | Phe | –4.3 | –7.6 | 2WGZ | Trp | –6.6 | –9.2 |
| 1JMZ | Phe | –3.2 | –6.1 | 2WGZ | Trp | –7.0 | –10.8 |
| 1LS6 | Phe | –4.7 | –6.4 | 2XEF[a] | Trp | –13.6 | –13.9 |
| 1LS6 | Phe | –4.4 | –6.0 | 3ABX | Trp | –5.4 | –10.3 |
| 1PX0 | Phe | –6.4 | –7.4 | 3AIR | Trp | –9.9 | –10.9 |
| 1RMH | Phe | –4.7 | –6.8 | 3M64 | Trp | –8.1 | –8.8 |
| 1T47 | Phe | –5.4 | –8.8 | 3MC5 | Trp | –8.1 | –8.8 |
| 1T47 | Phe | –5.0 | –8.4 | 3PQT[a] | Trp | –6.6 | –14.8 |
| 1T47 | Phe | –5.8 | –10.2 | 4A3H[a] | Trp | –7.9 | –13.6 |
| 1T47 | Phe | –4.6 | –5.4 | 1YEK | Tyr | –3.5 | –8.6 |
| 1V9T | Phe | –4.8 | –6.9 | 18GS[a] | Tyr | –6.7 | –10.4 |
| 1VBT | Phe | –4.9 | –6.8 | 1OAU[a] | Tyr | –5.2 | –10.5 |
| 1ZMT | Phe | –5.9 | –7.2 | 1RD4 | Tyr | –5.5 | –8.1 |
| 2GGX | Phe | –4.7 | –6.6 | 1XWK[a] | Tyr | –5.5 | –9.9 |
| 2OAC | Phe | –4.3 | –6.6 | 2D20 | Tyr | –4.4 | –7.2 |
| 2OAD | Phe | –4.5 | –6.6 | 2D20 | Tyr | –3.6 | –6.1 |
| 2V96 | Phe | –3.5 | –7.3 | 2NRU | Tyr | –3.9 | –7.5 |
| 3G4G | Phe | –5.1 | –7.8 | 2OAD | Tyr | –4.2 | –7.3 |
| 3O76 | Phe | –4.1 | –6.6 | 2WNB | Tyr | –4.8 | –9.7 |
| 1H2J[a] | Trp | –7.8 | –13.4 | 2WNF | Tyr | –4.6 | –9.7 |
| 1OAU[a] | Trp | –10.5 | –11.4 | 2Y59 | Tyr | –5.6 | –6.5 |
| 1RMH | Trp | –7.1 | –10.2 | 2XAP | Tyr | –6.3 | –8.3 |
| 1VAH | Trp | –5.1 | –10.8 | 3LM1 | Tyr | –3.7 | –6.1 |
| 1VBT | Trp | –6.7 | –8.0 | 3MK0 | Tyr | –3.4 | –8.7 |
| 1YEK | Trp | –7.8 | –9.1 | 3MK1 | Tyr | –3.1 | –7.2 |
| 2IKJ | Trp | –7.9 | –8.8 | 3O76 | Tyr | –5.4 | –6.9 |
| 2PDC | Trp | –8.2 | –8.8 | | | | |

[a] Complex involves 2-$NO_2$m

On the other hand, the computed interaction energies for some of the complexes extracted from crystal structures are strikingly similar to those of gas-phase optimized dimers, suggesting that in these cases the binding is dominated by these π-stacking interactions. For example, the interaction between Trp and the *m*-dinitrophnyl ring in 2XEF (–13.6 kcal $mol^{-1}$, see Figure IV-1) is nearly identical to that in the corresponding relaxed gas-phase complex (–13.9 kcal $mol^{-1}$). Similarly, the stacking interaction between the *m*-dinitrophenyl ring and the Trp in 1OAU (–10.5 kcal $mol^{-1}$) is about 90% of the strength in the corresponding gas-phase energy minimum (–11.4 kcal $mol^{-1}$). On the other hand, the interaction of this same nitroarene with the Tyr residue in 1OAU (–5.4 kcal $mol^{-1}$, see Figure IV-1) is only about half the strength of the corresponding relaxed complex (–10.5 kcal $mol^{-1}$).

**Summary and Concluding Remarks**

Nitroarene binding sites abound in proteins, and have been widely exploited in a number of therapeutic contexts.[290-308] However, despite the prevalence of these binding sites and their utility, little is known about the non-covalent interactions that drive the binding of nitroarenes by proteins. We have presented high-accuracy computational binding energies for model stacked complexes of nitroarenes with the amino acid side chains His, Phe, Trp, and Tyr. The data reveal that nitroarenes engage in strong π-stacking interactions with these aromatic amino acid side chains, with gas-phase interaction energies ranging from –6.7 kcal $mol^{-1}$ (Phe with nitrobenzene) to –15.0 kcal

$mol^{-1}$ (Trp with 1,3,5-trinitrobenzene). Overall, Trp engages in the strongest π-stacking interaction, followed by Tyr.

Unlike model stacked benzene dimers, we find that the impact of multiple nitro groups are not additive in complexes with His, Phe, Trp, and Tyr. This is attributed to the impact of local, direct interactions between the nitro groups and substituents and heteroatoms in the amino acid side chains. Further evidence of the importance of local, direct interactions in these stacking interactions comes from examination of the structures of the lowest-lying complexes for each amino acid/nitroarene pair.

In order to gain further insight into the π-stacking interactions in nitroarene binding sites of proteins, we have mined the PDB. Of 216 crystal structures that contain nitroarenes, 191 contain aromatic amino acids within 5 Å of the nitroarene. Of the binding sites containing aromatic amino acids, many exhibit clearly-defined π-stacking interactions between aromatic amino acids and the nitroarene component of the ligands. DFT computations applied to 65 such π-stacking complexes indicate that such interactions contribute significantly to the binding of the nitroarene. The resulting data should provide insight into the factors that lead to strong nitroarene binding, which should aid in both the identification of unrecognized nitroarene binding sites and the more effective exploitation of these interactions in biological applications.

# CHAPTER V

# STACKING INTERACTIONS BETWEEN 9-METHYLADENINE AND HETEROCYCLES COMMONLY FOUND IN PHARMACEUTICALS

## Introduction

π-stacking interactions are central to many areas of chemistry, including the binding of ligands to proteins and nucleic acids.[72,341,342] Harnessing such interactions lies at the heart of drug design.[343,344] In this context, understanding π-stacking interactions involving heterocycles is of paramount importance, because heterocyclic fragments abound in drugs and bioactive natural products.[345-352] Model π-stacking interactions involving heterocycles have been studied both experimentally[353-356] and theoretically,[357,358] with the primary aim of unraveling the many factors that control the strength and preferred orientation of these interactions. There have been numerous computational studies[359-367] of non-covalent interactions involving pyridine as a model *N*-heterocycle, building on the seminal work on the benzene-pyridine and pyridine-pyridine dimers by Sherrill and coworkers.[368] Stacking interactions of other six-membered *N*-heterocycles (pyrimidine, triazine, tetrazine, etc.),[369-372] as well as thiophene and benzothiophene, pyrrole, imidazole, indole, and larger bioactive molecules, have also been investigated computationally.[373-379]

Recently, Huber and co-workers[379] presented a comprehensive computational study of stacked dimers of benzene with 5- and 6-membered aromatic heterocycles found in small drug-like molecules. Ultimately, they concluded that even though

dispersion interactions are a significant driver of these stacking interactions, electrostatic effects play a key role in determining the preferred orientation of a given stacked dimer. Moreover, in accord with general guidelines for stacking interactions in drug design,[343] Huber and co-workers showed[379] that there was a correlation between the molecular dipole moments of the heterocycles and the predicted stacking interaction with benzene. On the other hand, Gellman *et al*.[380,381] previously emphasized the importance of local dipole moments[103,104,382-384] in understanding heterocycle stacking, as well as the unreliability of analyzing stacking interactions in terms of molecular dipoles.[385] This echoes our work on substituent effects in π-stacking interactions,[78,386-392] which can be understood in terms of the interaction of the local dipole moment associated with the substituent and the electric field of the other ring,[390] as well as the seminar model of heterocycle stacking interactions from Hunter and Sanders.[393] More generally, these effects can be understood in terms of direct electrostatic interactions between functional groups on the opposing rings.

Despite progress in understanding π-stacking interactions involving heterocyclic rings,[368-379] the rational design of heterocycle-containing molecules that maximize π-stacking interactions with a given aromatic group remains a challenge. A number of studies have been performed to identify common heterocyclic fragments within drug-like and bioactive molecules.[394,395352] For example, Broughton and Watson[394] mined heterocyclic motifs found in drugs that have reached Phase II clinic trial or later stages, reporting the 30 most frequently-identified heterocycles with good absorption, distribution, metabolism, excretion and toxicity (ADMET) characteristics for drug

design. Ertl and coworkers[395] analyzed an expansive virtual library of heteroaromatic scaffolds and investigated structural, bioactivity, and quantum mechanical properties of these molecules. Six heterocyclic fragments with fused rings were found to be most well-represented in bioactive molecules. Finally, Vitaku and coworkers[352] analyzed nitrogen heterocycles by exploring their structural diversity and substitution patterns, identifying 25 heterocycles that were found to be most commonly utilized among U.S. FDA approved pharmaceuticals.

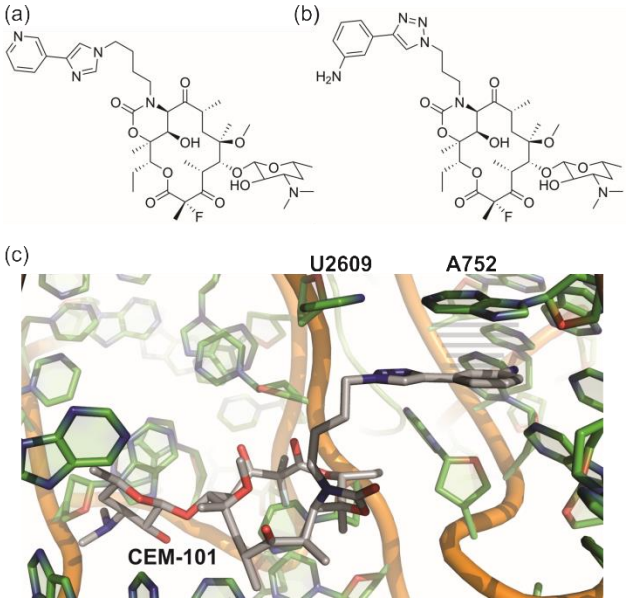
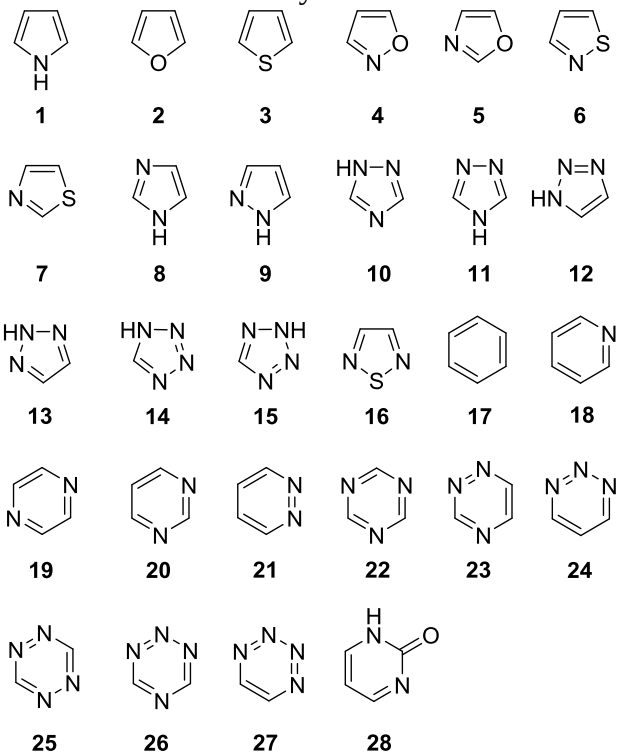



**Figure V-1.** Structures of telithromycin (a) and CEM-101 (b), as well as CEM-101 in the binding site of E. Coli ribosome (PDB ID: 4WWW).

Interactions of heterocycles with aromatic amino acid side chains and nucleic acids abound in drug binding sites.[396-401] For example, the Andrade group has a long-

standing interest in ketolide antibiotics,[402-406] which feature an extended heterocyclic alkyl-aryl side chain attached to the macrocycle (see Figure V-1a)[407,408] and exhibit tighter binding to the ribosome compared to previous classes of macrolides.[409] Despite the excellent activity of telithromycin against many macrolide-resistant bacteria,[410] the adverse effects and safety concerns associated with telithromycin led to the U.S. FDA narrowing its approved clinical uses.[411] The search for newer ketolides with fewer side effects continues, and Andrade and coworkers[402-406] previously synthesized 4-desmethyl telithromycin, 4,8-didesmethyl telithromycin, 4,10-didesmethyl telithromycin, and 4,8,10-tridesmethyl telithromycin in order to explore structure-activity relationships.

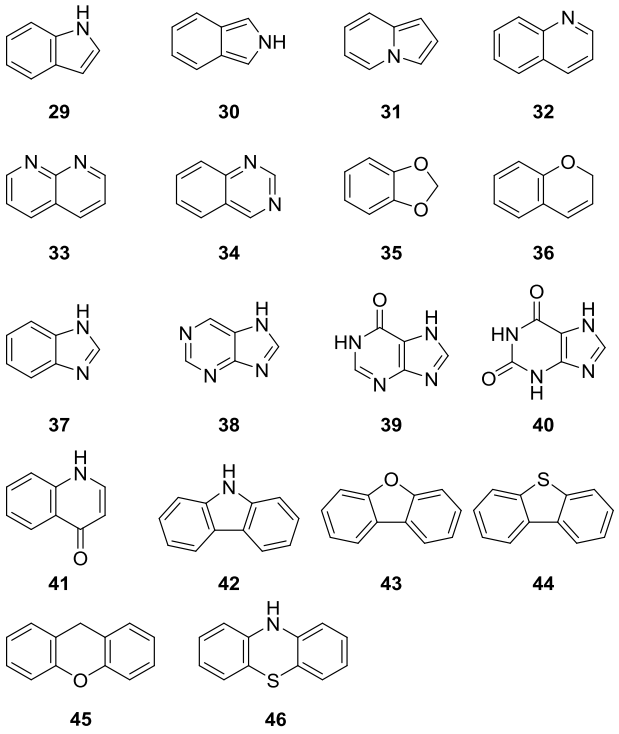
**Chart V-1.** Five and six-membered heterocycles.



A potential alternative means of tuning the activity of telithromycin is to change the alkyl-aryl side chain. For example, the structure of solithromycin (formerly known as CEM-101), which is now undergoing Phase 3 clinical development for the treatment of community-acquired bacterial pneumonia (CABP), urethritis and other infections,[412] differs from that of telithromycin in that the alkyl-aryl side chain contains triazole and alinine groups (see Figure V-1b). This side chain engages in well-defined π-stacking interactions with the uracil-adenine pair (A752 and U2609) in the binding site of the *E. Coli* ribosome (see Figure V-1c).[412] More precisely, the $NH_2$ of the aniline hydrogen bonds with the ribose backbone of A752, the triazole stacks with U2609, and the aniline fragment stacks with A752. However, due to the lack of a comprehensive understanding of π-stacking interactions involving heterocycles, it is not apparent what heterocycles can replace this aniline to provide enhanced stacking with the adenine.

Motivated by a desire to explore derivatives of CEM-101 that maximize π-stacking interactions with A752, herein we provide a computational study of model stacking interactions between 9-methyladenine and a set of heterocycles commonly found in pharmaceuticals (see Charts V-1 and V-2). The primary aim was to delineate the structural features that lead to the most favorable stacking interactions with adenosine, in order to guide the rational design of pharmaceuticals expected to engage in such interactions in the binding site.

**Chart V-2.** Fused heterocycles.




## Theoretical Methods

A representative collection of 46 heterocycles (see Charts V-1 and V-2) were selected to assess their propensity to form stacked dimers with 9-methyladenine as a mimic of adenosine. This set of heterocycles includes the conjugated systems identified by Vitaku and coworkers[352], as well as Broughton and Watson,[394] from their analyses of the structures of popular pharmaceuticals. Also included are various analogs of these, including structural isomers, isoelectronic congeners, *etc*. Although some of these other rings are not realistic as drug fragments due to lack of bioavailability or stability issues, whereas others are not the tautomer expected to be present under biological conditions,

they all serve to provide a more compehensive view of the heterocyclic features that lead to highly favorable π-stacking interactions with 9-methyladenine. This set of heterocycles has been divided into two main categories: five- and six-membered monocyclic species (Chart V-1) and heterocycles with two and three fused rings (Chart V-2).

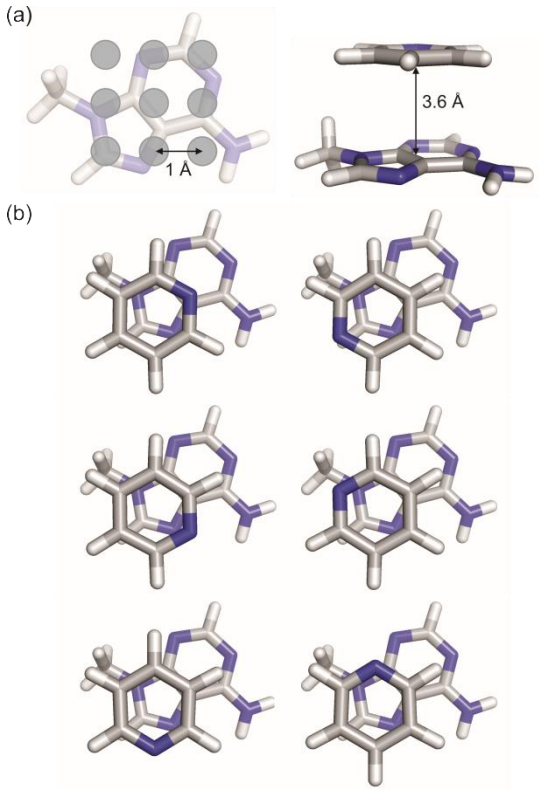



**Figure V-2.** (a) Schematic representation of the nine starting positions of the centroid of each heterocycle above the face of 9-methyladenine; (b) for each of these nine positions we considered six relative orienations of each heterocycle (pyridine used here as an example), for 54 total initial dimer geometries for each heterocycle.

The binding energy between each of these 46 heterocycles and 9-methyladenine was computed in the gas phase at the ωB97X-D/def2-TZVPP level of theory.[413,414] ωB97X-D, when paired with a triple-ζ basis set, was shown by Huber and co-workers[379] to provide robust predictions of stacked dimers of heterocycles with benzene. This level of theory has also been shown to provide accurate non-covalent binding energies for a wide range of model non-covalent complexes.[413,415-417] Although these gas-phase computations neglect differences in the desolvation costs of the different heterocycles, the present results should lay the groundwork for more detailed studies of the binding of these fragments in more realistic environments. Geometry optimizations were first carried out at the B97-D/def2-TZVPP level of theory.[265,266] To locate all stacked dimer configurations, we systematically searched over different relative positions and orientations of each heterocycle with 9-methyladenine (see Figure V-2). In particular, 54 initial stacked geometries were constructed for each heterocycle by considering nine initial positions of the heterocycle above the face of 9-methyladenine (see Figure V-2a), with the heterocycle located 3.6 Å above and parallel to the molecular plane of 9-methyladenine. At each of these nine positions, we considered six equally-spaced relative orientations of the heterocycle (*e.g.* see Figure V-2b). All of these initial structures were optimized to the nearest energy minimum. Unique optimized complexes were identified as those for which the root mean square deviation (RMSD) exceeded 0.5 Å with all other structures. Geometry optimizations of many of these structures resulted in edge-to-edge hydrogen-bonding arrangements, which were eliminated. In the case of heterocycle **12**, we were unable to locate any stationary point corresponding to a stacked

configuration with 9-methyladenine; instead, these two molecules had an overwhelming tendency to engage in hydrogen-bonding interactions. In total, 268 unique stacked complexes were identified. Gas-phase ωB97X-D/def2-TZVPP[413,414] single point energies were then carried out at all unique optimized stacked dimer geometries. Binding energies were evaluated at this level of theory as the difference in energy between the optimized dimers and the optimized isolated monomers. All computations were carried out using Gaussian09.[333]

**Results and Discussion**

For each heterocycle, numerous stacked energy minima were located, corresponding to different relative orientations and positions of the heterocycle above the face of 9-methyladenine. The overall distributions of binding energies for each heterocycle are plotted in Figures V-3 and V-4. All of the complexes considered are in a stacking arrangement (*i.e.* roughly parallel orientations with a significant degree of overlap); however, many of the heterocycles engage in additional interactions with 9-methyladenine. This includes CH/π and NH/π interactions, in which the edge of one ring is directed toward the face of the other. Moreover, the strengths of the π-stacking interactions are modulated by direct interactions[78,386-392] between functional groups on the heterocycle and the 9-methyladenine. These direct interactions are pivotal in determining the preferred orientation of the stacked rings, and can be understood as arising from the heterogeneous distribution of charge within these heterocycles.[380,381]

Below, we will focus primarily on the global energy minimum complex for each heterocycle, which are depicted in Figures V-5 and V-6.

In general, the larger rings (**29 - 46**) exhibited a far greater number of unique stacked energy minima than the smaller rings (**1 - 28**). In particular, whereas the small rings had, on average, 3.4 unique complexes, the larger rings had an average of 9.6 unique configurations. Similarly, many of the smaller rings exhibited only one or two (or even zero, in the case of **12**) unique stacked orientations, while there are at least four unique dimers for all of the larger heterocycles, many of which exhibit at least ten unique structures. Finally, even though the binding energies of the larger rings were typically more favorable, many of the smaller rings show substantial binding energies.

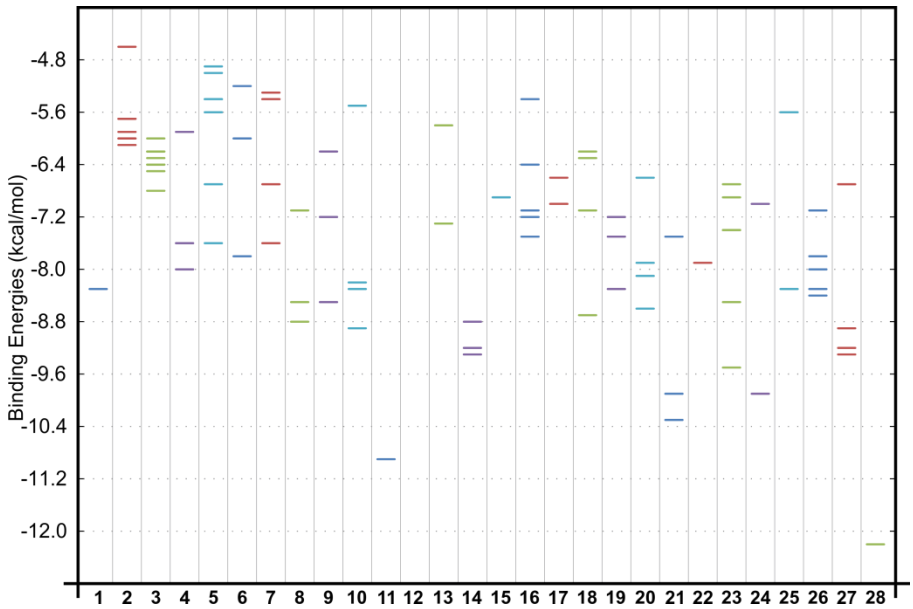



**Figure V-3.** Distribution of predicted binding energies (kcal $mol^{-1}$) for stacked dimers of 9-methyladenine with heterocycles **1-28**. For **12**, no stacked energy minima were identified at the B97-D/def2-TZVPP level of theory.

Starting with the 5-membered rings (**1-16**), we see a considerable spread in binding energies among the most favorable complexes for each heterocycle, ranging from –6.1 kcal $mol^{-1}$ for **2** to –10.9 kcal $mol^{-1}$ for **11**. Results for the triazoles (**10-13**) are particularly notable, revealing several trends that are seen again for the larger heterocycles. For example, **11** (4*H*-1,2,4-triazole), with a predicted binding energy of –10.9 kcal $mol^{-1}$, shows the most favorable binding energy of all 5-membered rings. For 1H-1,2,3-triazole (**12**), on the other hand, no stacked energy minima were located, whereas its tautomer 2H-1,2,3-triazole (**13**) exhibits a modestly favorable π-stacking interaction of –7.3 kcal $mol^{-1}$. It is important to note that this binding energy is actually less favorable than that predicted for pyrrole (**1**), highlighting the general trend that the incorporation of additional nitrogen atoms does not always lead to stronger stacking interactions. Finally, the predicted binding energy for **10** is 2 kcal $mol^{-1}$ less favorable than for its tautomer **11**.

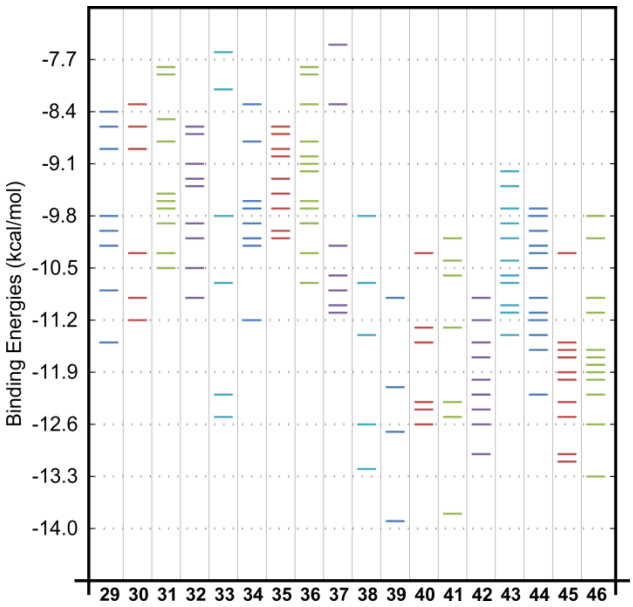



**Figure V-4.** Distribution of predicted binding energies (kcal $mol^{-1}$) for stacked dimers of 9-methyladenine with heterocycles **29**-**46**.

All of these data highlight the sensitivity of heterocycle stacking interactions to not only the regioisomer but also the tautomer that is present. This implies that the tautomer present in a given binding site could differ from that favored in solution, as long as the non-covalent interactions with 9-methyladenine can compensate for the energetic cost of forming a less favorable tautomer. The sensitivity to the tautomer present also extends to tautomeric tetrazoles **14** and **15**, for which the predicted binding energies differ by 2.4 kcal $mol^{-1}$.

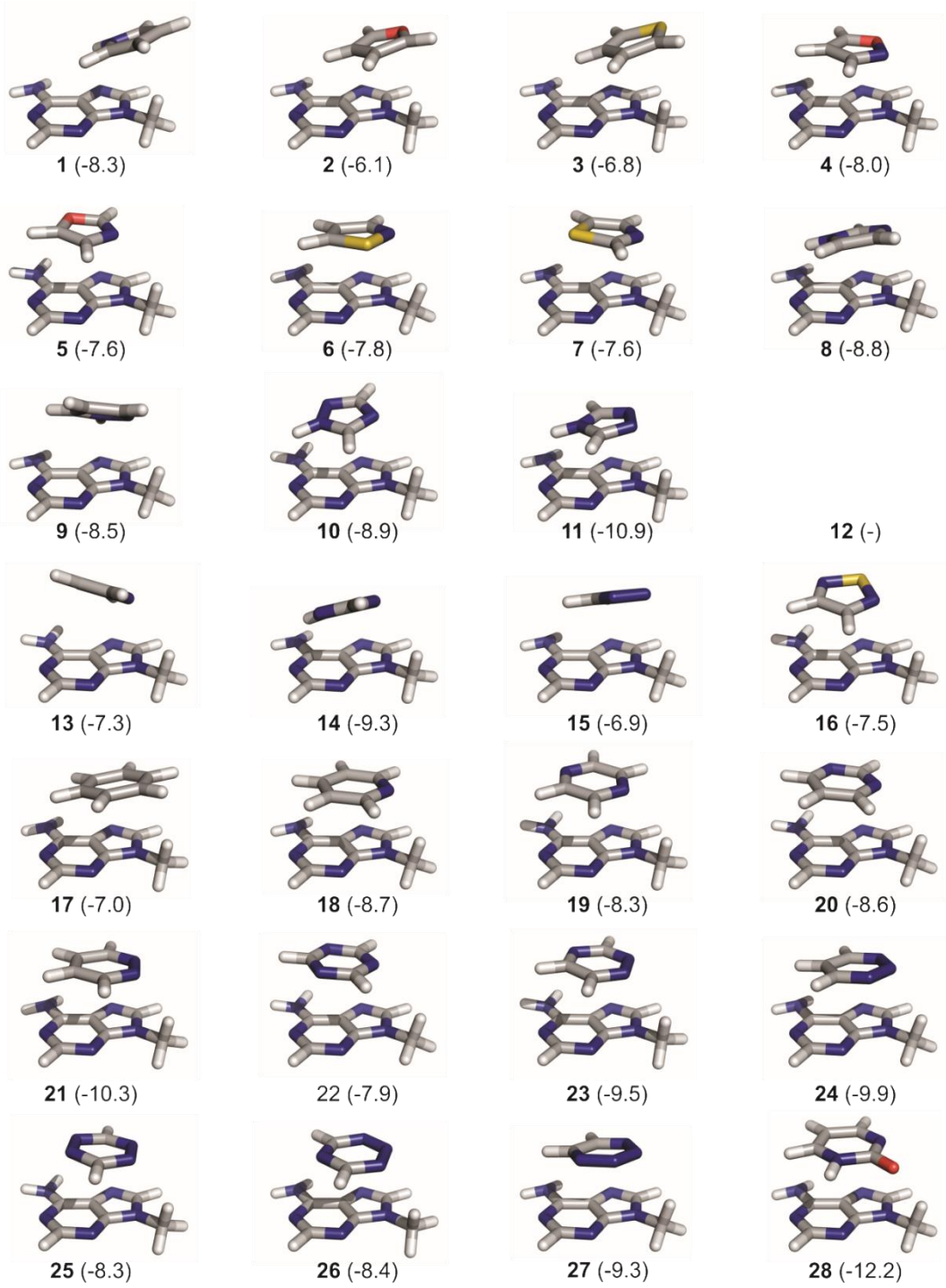



**Figure V-5.** Global minimum energy stacked dimers of 9-methyladenine and the heterocycles **1**-**28**. ωB97X-D/def2-TZVPP predicted binding energies are provided in kcal $mol^{-1}$. For **12**, no stacked energy minima were identified at this level of theory.

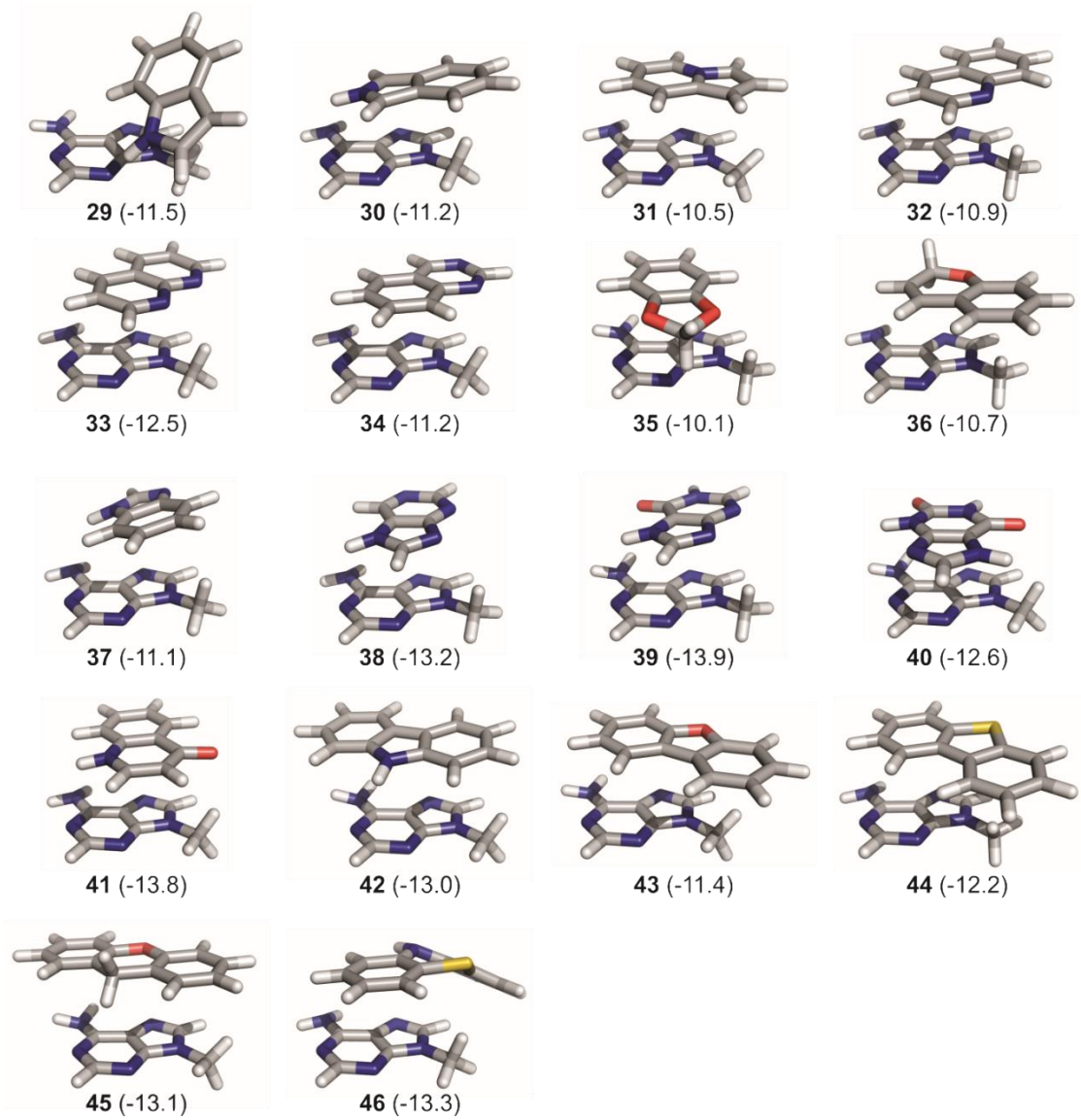



**Figure V-6.** Global minimum energy stacked dimers of 9-methyladenine and the heterocycles **29**-**46**. ωB97X-D/def2-TZVPP predicted binding energies are provided in kcal $mol^{-1}$.

In the case of **11**, the stacking interaction is enhanced by favorable direct interactions between the NH group of the triazole and a ring-nitrogen of 9-methyladenine, as well as between a ring nitrogen of **11** and the methyl group of 9-

methyladenine. Notably, the former interaction is not a conventional NH-donated hydrogen bond (the N···H-N angle is less than 112 °). Instead, this interaction is best described as the interaction of the local dipole moment associated with the NH group and the electric field of 9-methyladenine. The electric field in the region above the ring nitrogen will be dominated by the local dipole created by that nitrogen and the associated lone pair, similar to that observed above the nitrogens of triazine.[390] The latter interaction, occurring between a nitrogen and a decidedly non-polar CH group, is presumably driven largely by dispersion interactions. Notably, both of these interactions are present in many of the minimum energy complexes shown in Figures V-5 and V-6, and appear to be the primarily factors impacting the preferred orientation of stacked complexes of **1 - 46** with 9-methyladenine.

Comparing pyrrole (**1**), furan (**2**), and thiophene (**3**), binding energies follow the trend **1** > **3** > **2**. In the case of 1, the enhanced π-stacking appears to stem from the ability of pyrrole to engage in a strained hydrogen-bonding interaction between the NH of the pyrrole and the exocyclic $NH_2$ group of 9-methyladenine. This strained NH···N interaction also appears in the lowest-lying stacked dimers for imidazole (**8**) and pyrazole (**9**), which engage in stronger stacking interactions than their sulfur-analogs thiazole (**7**) and isothiazole (**6**). However, the tendency of sulfur-based heterocycles to provide more favorable π-stacking interactions than oxygen-based rings is not transferable to 5-membered rings with multiple heteroatoms. For example, comparing S-containing rings **6** and **7** with their oxygen analogs (**4** and **5**), we see that O-containing

rings engage in stacking interactions that are equal to or more favorable than the S-containing systems.

The six-membered rings (**17-28**) show a slightly narrower range of predicted binding energies. The rings examined primarily comprise benzene and the isoelectronic azines that result from the incorproation of between one and four nitrogens into the ring. In general, the nitrogen-containing heterocycles have more favorable π-stacking interactions with 9-methyladenine compared to benzene (**17**, –7.0 kcal $mol^{-1}$). For example, the single nitrogen atom in pyridine results in a 1.7 kcal $mol^{-1}$ enhancement of the binding energy, compared to benzene. However, the incorporation of a second nitrogen does not always lead to further enhancement of the interaction. For example, 1,4-pyrazine (**19**) and 1,3-pyrazine (**20**) engage in π-stacking interactions with 9-methyladenine that are slightly weaker than that of pyridine. Indeed, only 1,2-pyrazine (**21**) is predicted to provide greater stacking compared to pyridine. The triazines (**22-24**) show similar trends, with 1,3,5-triazine (**22**) leading to reduced stacking compared to pyridine, but the less-symmetric triazines (**23** and **24**) leading to more favorable stacking interactions than predicted for pyridine. The most favorable interaction for the triazines is predicted for the 1,2,3-isomer, **24**. For the tetrazines (**25 - 27**), only 1,2,3,4-tetrazine (**27**) provides a more favorable stacking interactions than pyridine. As seen with **11**, many of the six-member *N*-heterocycles display interactions between a ring-nitrogen and the methyl group of 9-methyladenine, again suggesting that this is an important interaction. Finally, we note that pyrimidine (**28**) engages in much more favorable stacking interactions with 9-methyladenine than any of the azines (–12.2 kcal $mol^{-1}$).

Turning next to the fused heterocycles (**29 - 46**), we see that these engage in much stronger π-stacking interactions with 9-methyladenine, on average, with predicted binding energies ranging from –10.1 kcal $mol^{-1}$ to –13.9 kcal $mol^{-1}$. Presumably, this arises from the greater surface area contact with 9-methyladenine, which will lead to more stabilizing dispersion interactions. However, even though the fused heterocycles exhibit stronger stacking interactions than the monocyclic heterocycles, there is no clear pattern with regard to the three-ring fused heterocycles and those with only two-rings. Any gain due to increased surface area quickly dissipates after the size of the heterocycle is comparable to that of adenine.

As with the small rings, the predicted stacking energies are sensitive to the distribution of heteroatoms within each ring. For instance, indole (**29**) and isoindole (**30**) exhibit stronger stacking interactions than indolizine (**31**). In the former case, a strained $NH^{\cdots}N$ hydrogen bonding interaction leads to indole deviating significantly from a parallel stacked arrangement. Isoindole (**30**), on the other hand, exhibits a nearly parallel configuration that is further enhanced by the unconventional $NH^{\cdots}N$ interaction seen for many of the smaller *N*-heterocycles. Similarly, 1,8-naphthyridine (**33**) exhibits stronger stacking interactions than quinazoline (**34**) due to the presence of two $N^{\cdots}H_3C$ interaction in the former case but only one in the latter case. The minimum energy complexes of **38** and **39** also exhibit this interaction. Overall, many of these larger heterocycles exhibit the same local, direct interactions as their smaller analogs. For example, the five-membered ring in **37** is in nearly the same orientation as seen for **8**. Similarly, the preferred stacking pose of **32** is similar to that of **18**. However, the

preferred orienation of the fused heterocycles can not always be predicted based on the preferred orientation of the component rings. For instance, the pyrimidine ring in quinazoline (**34**) is shifted significantly compared to pyrimidine itself (**20**).

With regard the broader trends, all of the heterocycles showed a marked sensitivity to orientation relative to 9-methyladenine, as indicated by the wide spread in predicted binding energies seen in Figures V-3 and V-4. This is perhaps unsurprising, given the established importance of electrostatic effects in heterocycle stacking and the heterogeneous distribution of charge around these rings.[373-379] However, this sensitivity has important implications for understanding stacking interactions in drug binding sites. In particular, even though a given heterocycle might engage in strong π-stacking interactions with 9-methyladenine, this might require a relative orientation that is not compatible with a given binding site. For instance, 1,2,3-triazine (**24**) is capable of engaging in strong π-stacking interactions with 9-methyladenine, with a predicted binding energy of –9.9 kcal $mol^{-1}$. However, this is only true for the orientation pictured in Figure V-5, in which the nitrogens of the triazine can interaction with the methyl group of the 9-methyladenine. In the other minimum energy orientation, the predicted binding energy is reduced to –7.0 kcal $mol^{-1}$, which is far weaker than many of the other, similarly-sized heterocycles. Similar observations hold for larger rings like **41**, for which the global minimum energy structure binds far more strongly than the next lowest-lying energy minimum.

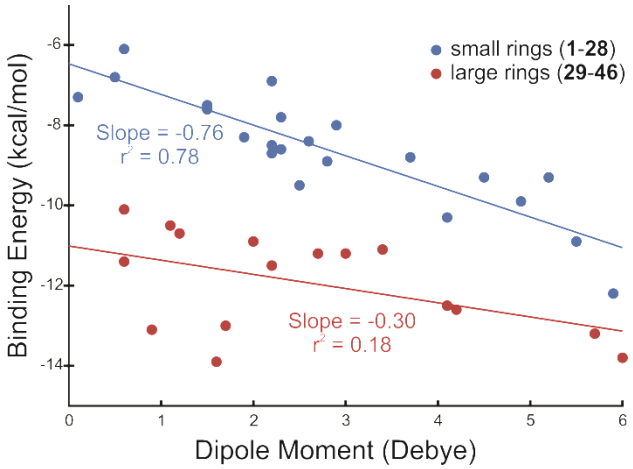



**Figure V-7.** Predicted binding energies (kcal $mol^{-1}$) for the global minimum energy stacked dimer of 9-methyladenine with heterocycles **1**-**46** versus the molecular dipole moments (in Debye) of the heterocycles. The slope and $r^2$ values apply to the best-fit lines for the small rings (**1**-**28**) and large rings (**29**-**46**) separately.

The above discussion has been cast primarily in terms of local, direct interactions, which we have shown to provide a more robust description of substituent effects in π-stacking intreactions.[78,386-392] However, given the common practice of discussing heterocycle stacking in terms of molecular dipole moments,[343,379] the ability of molecular dipole moments to predict the binding energies of these stacked complexes with 9-methyladenine was also considered. Overall, the molecular dipole moments of the heterocycles are poorly correlated with the total binding energies for the lowest-energy stacked complex for each heterocycle (see Figure V-7). However, considering only the small heterocycles (**1**-**28**), there is a good correlation ($r^2$ = 0.78) and in these cases dipole moments provide at least a qualitative predictor of π-stacking strengths with 9-methyladenine in these cases. This echos the recent work of Huber et al.[379] regarding

stacking interactions of small heterocycles with benzene. Moreover, for nearly all of the small heterocycles, the predicted global minimum energy stacked dimer corresponds to an orientation that features oppositely-oriented molecular dipoles (for examples, see Figure A-1 in Appendix). However, in the case of the fused heterocycles (**29-46**), there is essentially no correlation with the molecule dipole moment ($r^2$ = 0.18, see Figure V-7). This inability of molecular dipole-dipole interactions to predict the strength of π-stacking interactions for larger systems is unsurprising, because the multipole expansion, of which the dipole-dipole interaction is the leading term, will become less reliable as the size of the interacting systems grows with respect to the intermolecular distance. Thus, although the leading dipole-dipole term provides a reliable predictor of stacking interactions for small heterocycles, these interactions are more readily understood in terms of local, direct interactions of the functional groups on the two stacked systems (or, similarly, in terms of local dipole moments as championed by Gellman and co-workers).[380,381]

**Summary and Concluding Remarks**

Stacking interactions of 9-methyladenine with 46 heterocycles representative of common fragments of drug-like molecules were predicted at the ωB97X-D/def2-TZVPP//B97-D/def2-TZVPP level of theory. Overall, 268 unique stacked complexes were identified, exhibiting a broad range of predicted binding energies. Many of these interactions were highly favorable, and the introduction of heterocycles into the alkyl-aryl side chain of CEM-101 analogs should provide a powerful means of tuning its

binding affinity by modulating the stacking interaction of this group with A752 (see Figure V-1c). The broad range of predicted binding energies for each of the heterocycles highlights the impact of relative orientations on the stacking of heterocycles, with important implications for drug design.

Overall, larger heterocycles engage in stronger π-stacking interactions than their smaller analogs. However, many of the features that impact the preferred orientation of the smaller rings are mirrored in the larger systems, leading to the identification of several structural features that lead to strong stacking interactions of heterocycles with 9-methyladenine. Most notably, heterocycles with NH groups exhibit strong π-stacking interactions with 9-methyladenine, presumably because of the favorable interaction between this NH group and the ring-nitrogens of adenine. Unlike a conventional NH···N hydrogen bond, in this direct NH···N interaction the NH···N angle approaches 90°. Similarly, unconventional N···$H_3$C interactions occur in many of the lowest-lying complex for each heterocycle, suggesting this a second key factor that enhances π-stacking interactions of *N*-heterocycles with 9-methyladenine.

For all of the heterocycles considered, the distribution of heteroatoms within each ring as well as the particular tautomer considered has a profound impact on predicted stacking interactions. One implication is that the tautomer present in the bound state could actually differ from that favored in solution. Although molecular dipole moments can provide a qualitative predictor of the strengths of stacking interactions of small heterocycles (5- and 6-membered rings) with 9-methyladenine, they are uncorrelated with the predicted binding energies of the larger, fused heterocycles.

Instead, both the preferred orientations and stacking interactions can be more readily explained in terms of local, direct interactions between functional groups on the two interacting systems.

# CHAPTER VI

# EFFECTS OF Π-STACKING INTERACTIONS ON THE PKA'S OF IONIZABLE AMINO ACID SIDE CHAINS

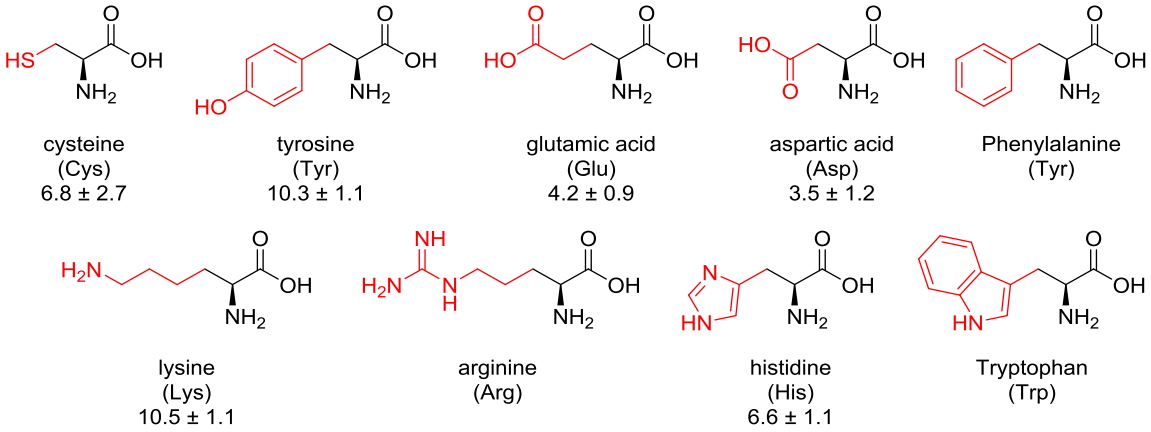
## Introduction

The protonation state of individual amino acids contributes to the structure, function, solubility, and stability of a protein,[418-427] and can impact everything from the catalytic activity of enzymes to the tendency of drugs to bind. Therefore, determining the $pK_a$ values of ionizable amino acids is indispensable in protein science. Aspartic acid (Asp), glutamic acid (Glu), cysteine (Cys), tyrosine (Tyr), histidine (His), lysine (Lys), and arginine (Arg) are the seven ionizable amino acid side chains, and their protonation state has great impact on the net charge and overall $pK_a$ of proteins (Chart VI-1). Among these seven ionizable side chains, Asp, Glu, Tyr and Cys are neutral at pH values below their $pK_a$ and negatively charged above the $pK_a$, while Arg, His, and Lys side chains are positively charged for pH values under their $pK_a$ and neutral at pH values above their $pK_a$.

The $pK_a$ values of folded proteins can be measured via nuclear magnetic resonance (NMR), potentiometric titration based on small peptide models, and other indirect methods,[422,428-432] which has led to a plethora of experimental $pK_a$ values for different amino acid side chains. Ionizable groups buried deeply in folded proteins often show large deviations from the intrinsic $pK_a$ values of the isolated side chains,[433] and changes in $pK_a$ values can be both positive and negative.[429,432,434-440] Average $pK_a$ values

for the ionizable amino acid side chains are presented in Chart VI-1, along with the standard deviation in values over a set of proteins.[441] As can been seen from these data, $pK_a$ values typically vary by 1-2 $pK_a$ units, although much larger changes are possible.[441] These $pK_a$ changes are typically attributed to three primary effects: dehydration (Born effect), charge-charge interaction (Coulombic interactions), and charge-dipole interactions (hydrogen bonds).[430,441-447]

**Chart VI-1.** Molecular structures of amino acids considered here. The side chains were modeled as the red portion of each structure, up to and including $C_\beta$. For the ionizable amino acid side chains, the average $pK_a$ values are provided along with typical ranges.[441]




Despite progress toward the accurate prediction of absolute and relative amino acid $pK_a$ values,[448-450] the influence of individual non-covalent interactions on $pK_a$ values, particularly those involving aromatic amino acid residues, are not well understood. We are particularly interested in cases where a neighboring aromatic residue forms close contact with an ionizable residue through π-stacking and XH/π

interactions.[268-270,451-453] These types of interactions have been shown to be strong stabilization forces in cases of protein-DNA complexes and protein-ligand complexes.[454] Upon ionization, the strength of these non-covalent interactions will change significantly. The resulting difference in the ability of the aromatic amino acid side chains to bind the protonated and deprotonated forms of the ionizable side chain could lead to potentially large changes in $pK_a$.

Motivated by the tabulated experimental $pK_a$ ranges of each ionizable side chain by Grimsley *et al.*,[441] we systematically explored the impact of π-stacking interactions between ionizable residues and the aromatic residues Phe, Tyr, Trp, and His using simplified model side chains. The results shed light on the impact of individual side chain interactions on $pK_a$-values, and the trends unveiled here could be used as a guide to understand environmental impacts on the $pK_a$'s of ionizable groups in proteins.

**Theoretical Methods**

Changes in $pK_a$ values ($\Delta pK_a$) for the seven ionizable amino acid side chains upon complexation with aromatic amino acid side chains were predicted. The amino acid side chains were modeled by replacing the backbone and $C_\beta$ with a methyl group (see Chart VI-1). For each ionizable group, we first determined all possible stacked complexes with with the aromatic residues Phe, Tyr, Trp, and His. Geometry optimizations were carried out at the B97-D/def2-TZVPP level of theory[68,455] in implicit water using the PCM model.[456,457] To locate all stacked complexes, we systematically searched over different relative positions and orientations of each ionizable group

stacked with aromatic residue (See Figure VI-1a). In particular, for each amino acid pair we constructed 54 initial stacked geometries by considering nine initial positions of the ionizable side chain above the face of aromatic residue, with the ionizable residue located 3.6 Å above and parallel to the molecular plane of aromatic side chain. At each of these nine positions, we considered six equally-spaced relative orientations of the ionizable group (see Figure VI-1b). All of these initial protonated structures were optimized to the nearest energy minimum. Many of these complexes optimized to edge-to-face or edge-to-edge hydrogen bonding configurations. Such structures were eliminated. Unique structures were identified as those that exhibit clearly-defined π-stacking interactions and for which the RMSD with all other structures exceeded 0.5 Å.

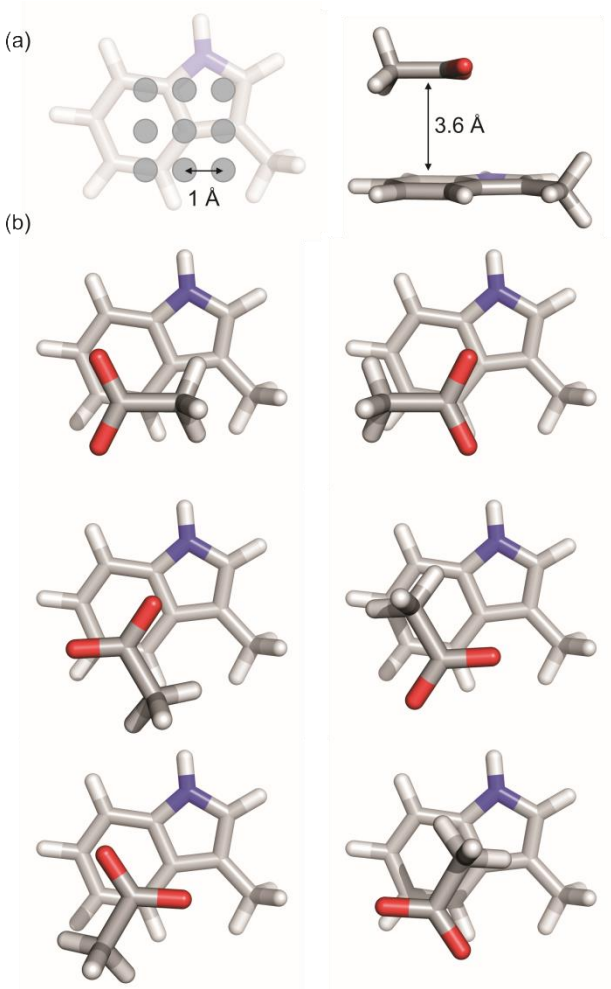



**Figure VI-1.** (a) Schematic representation of the nine starting positions of each ionizable side chain above the face of the aromatic side chain; (b) for each of these nine positions we considered six relative orienations of each ionizable side chain, for 54 total initial dimer geometries. (Asp interacting with Trp used as an example)

For each unique dimer structure, the deprotonable hydrogen from the ionizable residue was then removed. The resulting deprotonated dimers were then optimized while $C_\beta$ for each monomer was held fixed, to mimic the restraining effect of the protein backbone. In some cases, the deprotonated complexes optimized to structures with very

small imaginary vibrational frequencies indicating that the structures are not true minima on the potential energy surface at this level of theory. However, these small imaginary frequencies resulted from small numerical errors in the DFT energies, coupled with the very flat nature of the associated potential energy surfaces, and do not impact the predicted $pK_a$'s.

For each unique complex, the deprotonation energy ($\Delta E$) was defined as the difference in energy between the deprotonated dimer and the protonated dimer. From this, the absolute $pK_a$ can be computed via

$$pK_a = -log\left(e^{\frac{-\Delta E}{RT}}\right) \quad \text{(VI-1)}$$

The change in $pK_a$ upon complexation with an aromatic amino acid side chain, $\Delta pK_a$,was then computed from

$$\Delta pK_a = -log\left(e^{\frac{-\Delta\Delta E}{RT}}\right) \quad \text{(VI-2)}$$

where $\Delta\Delta E$ is the difference in deprotonation energies between the ionizable amino acid side chain complexed with an aromatic amino acid side chain and the corresponding isoliated ionizable amino acid side chain. That is, positive $\Delta pK_a$ values signify that the complexation of the ionizable amino acid side chain with the aromatic amino caid hinders deprotonation of the ionizable side chain.

For Lys and Arg, there are several possible protons that can be removed. For instance, upon complexation with another amino acid, the three protons of Lys are unique. We optimized the structure of the complexes resulting from removal of each of these protons, retaining the one that lead to the lowest-energy structure. Similarly, in the

case of Tyr···Tyr dimers, either one of the Tyr side chains can be deprotonated. In such dimers, we considered deprotonation of each Tyr side chain, and the reported $\Delta pK_a$ value results from the deprotonation that lead to the lower-energy structure.

For each ionizable residue we analyzed i) the conformational changes after deprotonation ii) the distribution and trend of $pK_a$ changes iii) the relationship between conformational changes and changes in $pK_a$. Due to the nature of ionizable amino acids, hydrogen bonded complexes and charge-dipole complexes are inevitably formed before and after deprotonation. Thus the conformational changes and the resulting $\Delta pK_a$ values for hydrogen bonding complexes are still worth investigating. Although we mainly focus on the effects of π-stacking and XH/π interactions between ionizable groups and aromatic groups, the effects of hydrogen bonding are discussed briefly.

Finally, we note that even though DFT computations paired with continuum solvent models are not expected to yield quantitatively correct absolute $pK_a$ values, our interest here is on the *change* in $pK_a$ values upon complexation with different aromatic amino acid side chains. In this case, the expectation is that errors arising from the approximate treatment of solvation effects will largely cancel, leading to reliable relative $pK_a$ values. Similarly, entropic contributions are neglected (*i.e.* we consider energy differences, rather than free energies), based on the assumption that entropic contributions will largely cancel leading to reliable relative $pK_a$ values.

**Results and Discussion**

The predicted $\Delta pK_a$ values for each ionizable amino acid side chain stacked with the aromatic residues Phe, Tyr, Trp, and His are displayed in Figure VI-2. In this plot, systems that maintain π-stacking configurations upon deprotonation are depicted with solid lines, whereas the dashed lines denote systems for which deprotonation leads to optimized structures with edge-to-face configurations. Overall, we predict significant changes in the $pK_a$ value of all seven ionizable amino acid side chains upon complexation. Notably, these predicted changes are at least as large as the standard deviation of measured $pK_a$ values for these groups in proteins (see Chart VI-1). The predicted $\Delta pK_a$ values are mostly positive for all but Arg, for which the $\Delta pK_a$ values are always negative. That is, for all but Arg, complexation with one of the aromatic amino acid side chains leads to a reduction in the acidity of these ionizable groups. More specific trends for individual ionizable amino acid side chains are discussed below.

$pK_a$ Changes of Asp and Glu

Aspartic and glutamic acids share many structural properties, and both bear a terminal carboxyl group that is deprotonated under physiological pH. Our computational results show that when Asp and Glu stack with aromatic amino acids, their $pK_a$ values only change in the positive direction (*i.e.* the protons become less acidic). Apparently, the aromatic amino acid side chains provide greater stabilization of the protonated Asp and Glu side chains than the deprotonated systems. This is consistent with the generally unfavorable interactions of anions with phenyl, phenol, indole, and imidazole rings. Overall, Asp exhibits fewer unique dimers across all four aromatic residues compared to

glutamic acid, presumably due to its smaller degree of conformational flexibility. Interestingly, for both Asp and Glu, the $\Delta pK_a$ values show a bimodal distribution, with $\Delta pK_a$ values either under 1 or above 2, suggesting that there are two families of interactions that lead to either large or relatively small changes in $pK_a$.

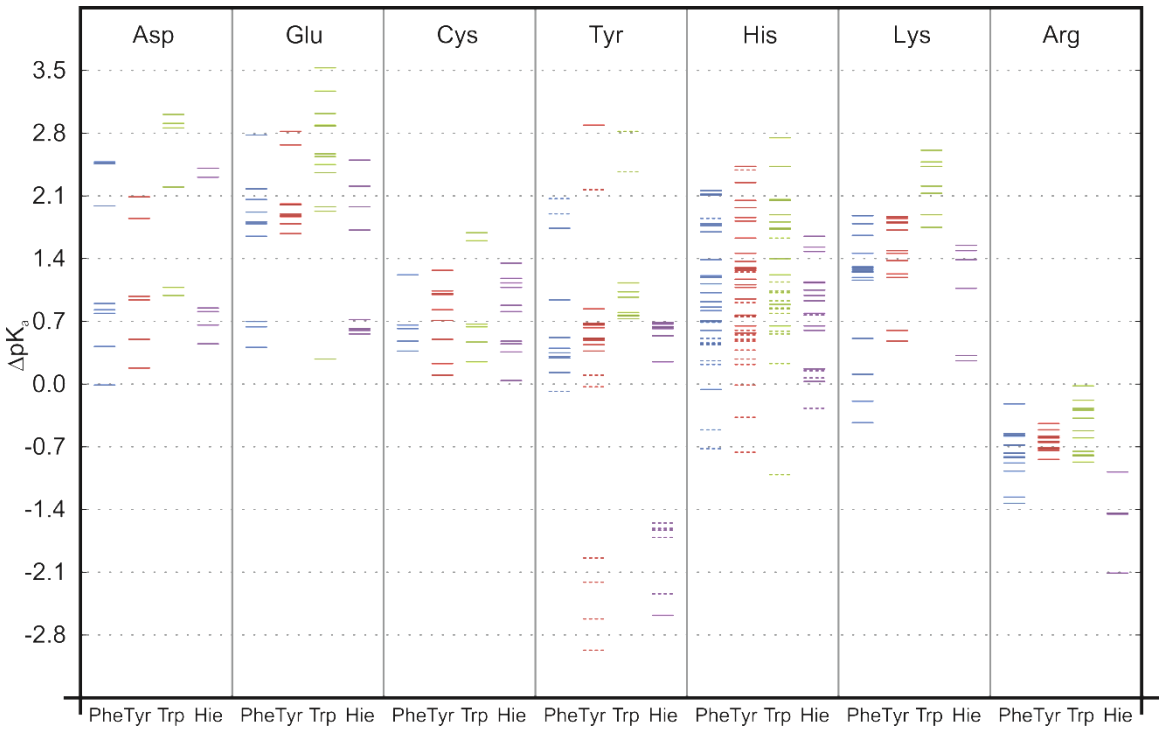



**Figure VI-2.** Distribution of $\Delta pK_a$ values for each ionizable residue-aromatic residue dimer. Dimers involving hydrogen bonding are not included in the distribution. The dotted lines correspond to complexes that change to edge-to-face configurations upon deprotonation, while the solid lines are in π-stacking orientations.

Although some of the complexes of Asp and Glu with the aromatic amino acid side chains showed significant conformational changes after deprotonation, there was no

correlation between the magnitude of $\Delta pK_a$ and the extent of conformational change. Moreover, most of these complexes do not show any notable conformational changes after deprotonation.

$pK_a$ Changes of Cys

Cysteine is a small amino acid with an ionizable SH group that is protonated at physiological pH. Our results indicate the number of unique dimers of model Cys is much smaller compared to other ionizable residues; again, this is presumably due to the smaller size of Cys, which leads to less conformational flexibility. The range of $\Delta pK_a$ values for Cys is also the smallest among all the seven ionizable groups, ranging from 0 to 1.7 with a mean $\Delta pK_a$ value close to 1 (Figure VI-2). All of the Cys complexes exhibit very small conformational changes upon deprotonation.

$pK_a$ Changes of Tyr

Tyrosine has an ionizable phenol group, which is neutral at physiological pH. Many of the complexes involving Tyr showed large conformational changes upon deprotonation, particularly for Tyr-Phe and Tyr-Tyr dimers. More precisely, after deprotonation, four of the eleven unique Tyr-Phe complexes adopt edge-to-face configurations. Similar conformational changes also occur to Tyr-Tyr dimers (see Figure VI-3), for which five out of 19 change from stacked configurations to edge-to-face complexes after removal of the acidic proton. For the Tyr-His dimers, on the other hand, nearly half of the complexes switch to edge-to-face configurations upon deprotonation. Moreover, there is a trend with regard to the tendency to engage in edge-to-face interactions and the $\Delta pK_a$ values. First, edge-to-face interactions tend to lead to negative

$\Delta pK_a$ values for Tyr stacked with each of the aromatic residues. Second, the complexes that switch to edge-to-face interactions typically lead to the largest $pK_a$ changes for Tyr-Phe, Tyr-Trp and Tyr-Tyr complexes. The complexes that maintain π-stacking interactions upon deprotonation tended to exhibit much smaller $\Delta pK_a$ values.

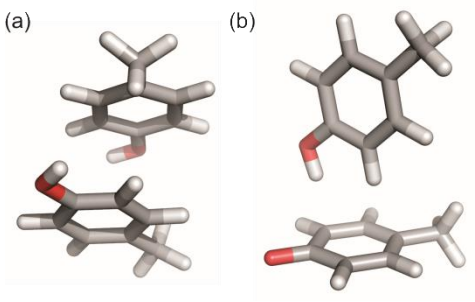



**Figure VI-3.** Example of Tyr dimer that changes from a parallel-displaced configuration (a) to an edge-to-face configuration upon deprotonation (b).

<u>$pK_a$ Changes of His</u>

Histidine features an imidazole ring that has two ionizable NH groups, leading to two potential deprotonated states (deprotonation of the δ hydrogen or the ε hydrogen). Histidine is positively charged at physiological pH with ε protonation dominant. For the purpose of a more complete understanding of the effects of neighboring aromatic rings on different hydrogens, we considered the deprotonation of both δ and ε hydrogens when considering stacked dimers of histidine with aromatic amino acid residues. Removal of these different hydrogens has a great impact on both the dimer geometry and the predicted $\Delta pK_a$ values. Overall, δ deprotonation leads to positive values of $\Delta pK_a$, while ε

deprotonation results in both positive and negative $\Delta pK_a$ values. Deprotonated complexes featuring edge-to-face interactions occur much more frequently in His dimers compared to the other systems. Indeed, for His-Phe, His-Tyr and His-Trp dimers, edge-to-face configurations predominate upon removal of the ε proton, nine out of 14 His-Phe dimers and six out of twelve His-Trp dimers switching from stacked configuration to edge-to-face configuration after ε deprotonation. Moreover, all of the negative $\Delta pK_a$ values are due to complexes in which deprotonation leads to edge-to-face interactions. On the other hand, π-stacking configurations predominate after removal of the δ proton across all four aromatic residues; out of 52 δ-deprotonated dimers, 45 retain π-stacking interactions upon deprotonation. These systems that maintain π-stacking interactions always lead to positive $\Delta pK_a$ values, with most stacked dimers leading to values from 0.7 to 2.1 $pK_a$ units.

$pK_a$ Changes of Lys and Arg

Both Lys and Arg carry a positive charge at physiological pH, and both have long side chains with terminal ammonium and guanidinium groups, respectively. The most distinctive difference between Lys and Arg dimers is that arginine only exhibits negative $\Delta pK_a$ values while nearly all of the Lys dimers show positive $\Delta pK_a$ values. In both cases, deprotonation leads to relatively small conformational changes. Lys exhibits both CH/π and NH/π interactions in these complexes, which lead to differences in predicted $\Delta pK_a$ values. For instance, complexes exhibiting NH/π interactions lead to the largest $\Delta pK_a$ values for Lys complexed with each of the aromatic amino acid residues. Such NH/π interaction occurs frequently for the Lys-Phe, Lys-Tyr, and Lys-Trp dimers.

This is particularly true of the complexes with Trp, for which four out of seven unique dimers exhibit an NH/π interaction. Indeed, on average the $\Delta pK_a$ value of Lys-Trp is higher than the other three mainly due to NH/π interactions in the former case. CH/π interactions, on the other hand, lead to more modest $\Delta pK_a$ values of 0 to +2.

Among all ionizable side chains, Arg is the only one that leads to negative $\Delta pK_a$ values when complexed with aromatic residues. This mainly arises because Arg engages in stabilizing NH/π interactions upon deprotonation for Arg-Phe, Arg-Tyr and Arg-Trp dimers.

The Effects of Hydrogen Bonding and Ion Dipole Interactions

In Figure VI-3 and the associated discussion above, we considered only complexes in which the ionizable side chain engaged only in π-stacking interactions with the aromatic amino acid side chain, eliminating complexes with hydrogen-bonding interactions. In such cases, the $\Delta pK_a$ values for all but Arg were almost all positive. Below, we consider the hydrogen-bonding complexes that were located. In these cases, Asp and Glu exhibit both positive and negative $\Delta pK_a$ values. Overall, when the dimer changes from hydrogen bonded configuration to stacking configuration after deprotonation, the $\Delta pK_a$ value is large and positive. If the dimer changes from stacking configuration to charge dipole configuration after deprotonation, $\Delta pK_a$ was large and negative. Tyr dimers show relative large negative $pK_a$ changes when hydrogen bonding effects are involved. Some Lys-His dimers exhibit large positive $\Delta pK_a$ values because they change from hydrogen-bonded conformation to stacked conformation after

deprotonation. For Arg-Tyr and Arg-His dimers, the few positive $\Delta pK_a$ values and large negative $\Delta pK_a$ values are also due to hydrogen bonding effects.

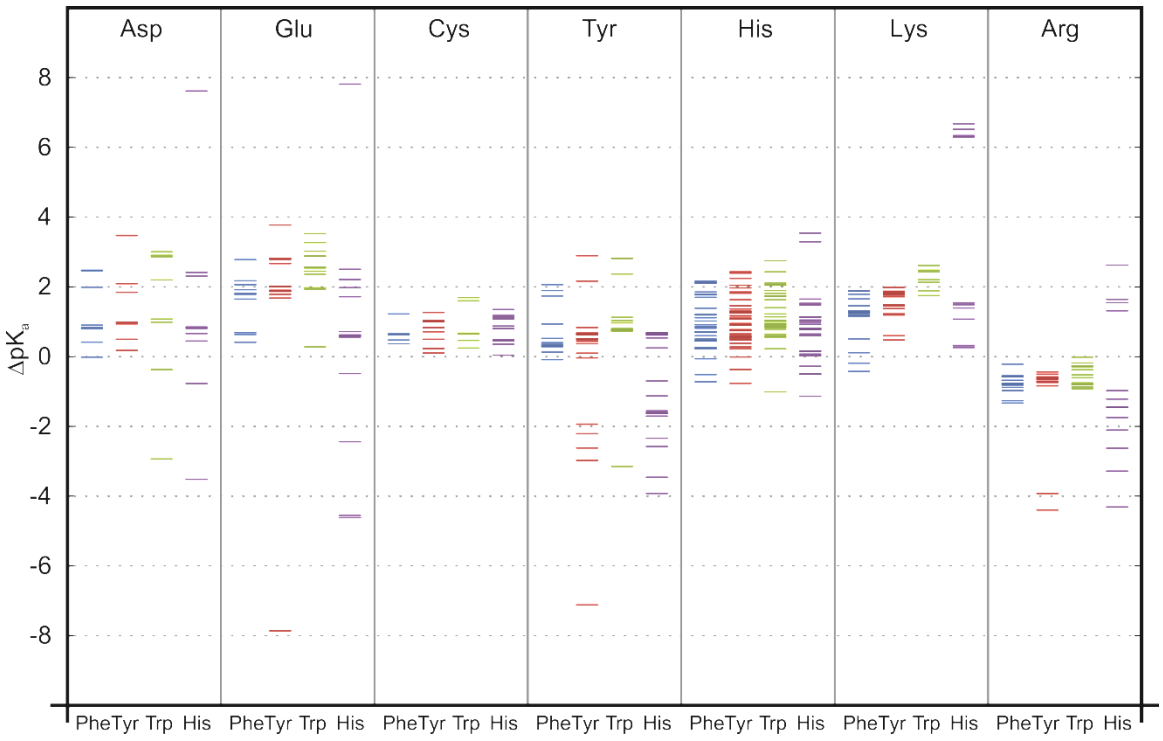



**Figure VI-4**. Distribution of $\Delta pK_a$ values for each ionizable residue-aromatic residue dimer with hydrogen bonded configurations. All Lys and arginine deprotonation states are included in this figure.

**Summary and Concluding Remarks**

The $pK_a$ values of ionizable amino acid side chains can be strongly influenced by their environment in proteins. Although changes in $pK_a$ for buried residues, relative to their intrinsic $pK_a$ of the isolated amino acid side chain, have long been understood to

arise from a combination of dehydration, Coulombic interactions, and hydrogen bonding interactions,[430,441-447] little attention has been given to the role of other non-covalent interactions operative in these systems. In particular, π-stacking interactions have the potential to significantly alter the $pK_a$ of these ionizable groups.

We predicted the change in $pK_a$ values for the seven ionizable amino acid side chains upon stacking with the aromatic amino acid side chains Phe, Tyr, Trp, and His. These π-stacking interactions can lead to significant changes in predicted $pK_a$ values compared to the isolated ionizable groups. Overall, the predicted changes in $pK_a$ values are mostly positive for all of the ionizable side chains except Arg. That is, the π-stacking of Asp, Glu, Cys, Tyr, His, or Lys with Phe, Tyr, Trp, or His typically leads to a decrease in the acidity of these ionizable side chains. With Arg, on the other hand, complexation with these aromatic side chains leads to an increase in the acidity of the guanidinium protons. The predicted change in $pK_a$ value is sensitive to the relative orientation of the complexed rings both before and after deprotonation. In particular, for Tyr and His dimers, negative $pK_a$ changes upon complexation with aromatic amino acid side chains were always accompanied by a change from a π-stacking to an edge-to-face configuration upon deprotonation. For His, deprotonation of tautomeric protons impacted both the dimer geometry and $pK_a$ change, with δ deprotonation leading to positive $\Delta pK_a$ values and primarily stacked deprotonated dimers. When hydrogen-bonding complexes were considered, there was a much broader range of predicted $pK_a$ changes, with both highly negative and positive values predicted for Asp, Glu, Tyr, Lys and Arg.

# CHAPTER VII

# CONCLUSIONS

In this research we first surveyed cation-π interactions involving alkali metal as a complement to our work on other non-covalent interactions in biological systems. Cation-π interactions were shown to be key non-covalent interactions in host-guest system, biology, materials science and chemical catalysis through spectroscopic and computational analyses. Furthermore, incorporating substituents and heteroatoms into aromatic rings systematically provides an effective means of altering the strength of cation-π interactions, and studying these effects will provide more thorough understanding of the nature of these interactions.

We next explored non-covalent interactions involving π systems in four contexts: protein-DNA interactions, protein-ligand interactions, nucleic acid-ligand interactions, and interaction between amino acids. First, the role of non–covalent interactions in the 5' binding site of the antibody ED-10 were examined. The marked selectivity of ED-10 for dTdC over dAdC was shown by MMGBSA and B97-D computations to result from differences in π-stacking and XH/π interactions between the nucleobase and a Trp residue in the binding site. Our computational results also predict that mutations of two key aromatic residues will severely impair in the selectivity for dTdC over dAdC, suggesting other antibodies without these key residues will not exhibit the same 5'-base selectivity as ED-10. Overall, both MMGBSA and B97-D data show that even though π-stacking, CH/π, NH/π, and hydrogen bonding interactions all participate in the binding

of dinucleotides with ED-10, the strong selectivity for 5'-dT sequences is primarily a result of differences in aromatic interactions (XH/π and π-stacking interactions), not hydrogen-bonding.

Next, π-stacking interactions between nitroarenes and model amino acid side chains, His, Phe, Trp, and Tyr, were shown to be strong, ranging from –6.7 kcal $mol^{-1}$ (Phe with nitrobenzene) to –15.0 kcal $mol^{-1}$ (Trp with 1,3,5-trinitrobenzene). Due to local, direct interactions between the nitro groups and substituents and heteroatoms in the amino acid side chains, the impact of multiple nitro groups were found to be non-additive in complexes with His, Phe, Trp, and Tyr. Out of 216 crystal structures that contain nitroarenes in PDB, 191 contain aromatic amino acids within 5 Å of the nitroarene, and 65 well-defined π-stacking complexes were considered for computational analyses. π-stacking interactions between the nitroarene and aromatic amino acid side chains exhibit a broad range of strengths, with many contributing significantly to binding. Our results on the nitroarene binding mode and strength will help the identification of unrecognized nitroarene binding sites in proteins, and lead to more effective means of exploiting these interactions in biological applications.

Next, stacking interactions of 9-methyladenine with 46 heterocycles representative of common fragments of drug-like molecules were examined, which show a broad range of interaction energies. Although larger heterocycles engage in generally stronger π-stacking interactions than their smaller analogs, the impact of relative orientations made some smaller rings also show highly favorable interactions. Heterocycles with NH groups exhibit strong π-stacking interactions with 9-

methyladenine, presumably because of unconventional interactions between this NH group and the ring-nitrogens of adenine. Similarly, $N^{\cdots}H_3C$ interactions could be a second key factor that enhances π-stacking interactions of *N*-heterocycles with 9-methyladenine. Since molecular dipoles only correlate with the binding strength of small heterocycles but not the larger, fused heterocycles, the preferred orientations and stacking interactions can be better explained by local, direct interactions between functional groups on the two interacting systems.

Finally, our computational results indicate π-stacking interactions can lead to significant changes in predicted $pK_a$ for ionizable amino acid side chains. Stacking interactions of Asp, Glu, Cys, Tyr, His, or Lys with Phe, Tyr, Trp or His mostly lead to an increase in the predicted $pK_a$ compared to the isolated amino acids, while the $pK_a$ of arginine is always suppressed by such interactions. That is to say, except for arginine, ionizable residue-aromatic residue interactions tend to exhibit decreased acidity when complexed with aromatic amino acids. For Tyr and His dimers, the conformational changes before and after deprotonation has great impact on the predicted change in $pK_a$ value. Negative $pK_a$ changes were always associated with a change from π-stacking to edge-to-face configurations upon deprotonation. For His, the removal of different tautomeric protons had diverse effects, with δ deprotonation leading to positive changes in the $pK_a$.

REFERENCES

(1) Cerny, J.; Hobza, P. *Physical Chemistry Chemical Physics* **2007**, *9*, 5291.

(2) Zahradnik, R.; Muller Dethlefs, K.; Hobza, P.; Zahradnık, R.; Müller Dethlefs, K. *Collection of Czechoslovak Chemical Communications* **2006**, *71*, 443.

(3) Riley, K. E.; Hobza, P. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 3.

(4) Hunter, C. A. *Chemical Society Reviews* **1994**, *23*, 101.

(5) Hofstadler, S. A.; Griffey, R. H. *Chemical Reviews* **2001**, *101*, 377.

(6) Redinbo, M. R.; Stewart, L.; Kuhn, P.; Champoux, J. J.; Hol, W. G. J. *Science* **1998**, *279*, 1504.

(7) Cheng, X.; Harms, A. C.; Goudreau, P. N.; Terwilliger, T. C.; Smith, R. D. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93*, 7022.

(8) Dunn, M. F. In *eLS*; John Wiley & Sons, Ltd: 2001.

(9) Pramanik, B. N.; Bartner, P. L.; Mirza, U. A.; Liu, Y.-H.; Ganguly, A. K. *Journal of Mass Spectrometry* **1998**, *33*, 911.

(10) Conte, L. L.; Chothia, C.; Janin, J. *Journal of Molecular Biology* **1999**, *285*, 2177.

(11) Imming, P.; Sinning, C.; Meyer, A. *Nat Rev Drug Discov* **2006**, *5*, 821.

(12) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat Rev Drug Discov* **2004**, *3*, 935.

(13) Gilson, M. K.; Zhou, H.-X. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 21.

(14) Baguley, B. C. *Anti-cancer drug design* **1991**, *6*, 1.

(15) Wan, K. X.; Shibue, T.; Gross, M. L. *Journal of the American Chemical Society* **2000**, *122*, 300.

(16) Gallego, J.; Varani, G. *Accounts of Chemical Research* **2001**, *34*, 836.

(17) Thomas, J. R.; Hergenrother, P. J. *Chemical Reviews* **2008**, *108*, 1171.

(18) Tor, Y. *ChemBioChem* **2003**, *4*, 998.

(19) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.

(20) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *Journal of Molecular Biology* **1994**, *239*, 315.

(21) Dougherty, D. A. *Accounts of Chemical Research* **2013**, *46*, 885.

(22) Hunter, C. A.; Singh, J.; Thornton, J. M. *Journal of Molecular Biology* **1991**, *218*, 837.

(23) McGaughey, G. B.; Gagné, M.; Rappé, A. K. *Journal of Biological Chemistry* **1998**, *273*, 15458.

(24) Gazit, E. *The FASEB Journal* **2002**, *16*, 77.

(25) Burley, S. K.; Petsko, G. A. *FEBS Letters* **1986**, *203*, 139.

(26) Stites, W. E. *Chemical Reviews* **1997**, *97*, 1233.

(27) Umezawa, Y.; Nishio, M. *Bioorganic & Medicinal Chemistry* **1998**, *6*, 493.

(28) Plevin, M. J.; Bryce, D. L.; Boisbouvier, J. *Nat Chem* **2010**, *2*, 466.

(29) Nishio, M.; Umezawa, Y.; Hirota, M.; Takeuchi, Y. *Tetrahedron* **1995**, *51*, 8665.

(30) Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. *Journal of Molecular Biology* **2001**, *307*, 357.

(31) Boys, S. F.; Cook, G. B.; Reeves, C. M.; Shavitt, I. *Nature* **1956**, *178*, 1207.

(32) Allinger, N. L. *Journal of the American Chemical Society* **1977**, *99*, 8127.

(33) Zhao, Y.; Truhlar, D. G. *Accounts of Chemical Research* **2008**, *41*, 157.

(34) Andrés, J.-M. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 1011.

(35) Clementi, E. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 89.

(36) Garrett, B. C.; Truhlar, D. G. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 67.

(37) Hratchian, H. P.; Schlegel, H. B. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 195.

(38) Kapral, R.; Ciccotti, G. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 425.

(39) McCammon, J. A. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 41.

(40) Thiel, W. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 559.

(41) Wormer, P. E. S.; Avoird, A. v. d. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 1047.

(42) Zhao, L.; Zou, S.; Hao, E.; Schatz, G. C. In *Theory and Applications of Computational Chemistry*; Scuseria, G. E., Dykstra, C. E., Frenking, G., Kim, K. S., Eds.; Elsevier: Amsterdam, 2005, p 47.

(43) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *Journal of Computational Chemistry* **1990**, *11*, 440.

(44) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *Journal of the American Chemical Society* **1992**, *114*, 10024.

(45) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *Journal of the American Chemical Society* **1989**, *111*, 8551.

(46) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *Journal of Computational Chemistry* **2003**, *24*, 1999.

(47) Alder, B. J.; Wainwright, T. E. *The Journal of Chemical Physics* **1959**, *31*, 459.

(48) Rahman, A. *Physical Review* **1964**, *136*, A405.

(49) Sitkoff, D.; Sharp, K. A.; Honig, B. *The Journal of Physical Chemistry* **1994**, *98*, 1978.

(50) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.

(51) Ding, F.; Borreguero, J. M.; Buldyrey, S. V.; Stanley, H. E.; Dokholyan, N. V. *Proteins: Structure, Function, and Bioinformatics* **2003**, *53*, 220.

(52) Raval, A.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *Proteins: Structure, Function, and Bioinformatics* **2012**, *80*, 2071.

(53) Koehl, P.; Levitt, M. *Nat Struct Mol Biol* **1999**, *6*, 108.

(54) Kollman, P. *Chemical Reviews* **1993**, *93*, 2395.

(55) Zwanzig, R. W. *The Journal of Chemical Physics* **1954**, *22*, 1420.

(56) Straatsma, T. P.; McCammon, J. A. *Annual Review of Physical Chemistry* **1992**, *43*, 407.

(57) Massova, I.; Kollman, P. *Perspectives in Drug Discovery and Design* **2000**, *18*, 113.

(58) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham Iii, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Computer Physics Communications* **1995**, *91*, 1.

(59) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2012**, *8*, 3314.

(60) Davidson, E. R.; Feller, D. *Chemical Reviews* **1986**, *86*, 681.

(61) Kresse, G.; Furthmüller, J. *Computational Materials Science* **1996**, *6*, 15.

(62) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Physical Review B* **1992**, *46*, 6671.

(63) Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864.

(64) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133.

(65) Dirac, P. A. M. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1926**, *112*, 661.

(66) Heisenberg, W. *Z. Physik* **1926**, *38*, 411.

(67) Grimme, S. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 211.

(68) Grimme, S. *Journal of Computational Chemistry* **2006**, *27*, 1787.

(69) Grimme, S. *Journal of Computational Chemistry* **2004**, *25*, 1463.

(70) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew Chem Int Edit* **2003**, *42*, 1210.

(71) Salonen, L. M.; Ellermann, M.; Diederich, F. *Angew Chem Int Edit* **2011**, *50*, 4808.

(72) Schneider, H.-J. *Angew. Chem. Int. Ed.* **2009**, *48*, 3924.

(73) Ma, J. C.; Dougherty, D. A. *Chem Rev* **1997**, *97*, 1303.

(74) Dougherty, D. A. *Accounts Chem Res* **2013**, *46*, 885.

(75) Mahadevi, A. S.; Sastry, G. N. *Chem Rev* **2013**, *113*, 2100.

(76) Marshall, M. S.; Steele, R. P.; Thanthiriwatte, K. S.; Sherrill, C. D. *Journal of Physical Chemistry A* **2009**, *113*, 13628.

(77) Frontera, A.; Quiñonero, D.; Deyà, P. M. *WIREs Comp. Mol. Sci.* **2011**, *1*, 440.

(78) Raju, R. K.; Bloom, J. W. G.; An, Y.; Wheeler, S. E. *ChemPhysChem* **2011**, *12*, 3116.

(79) Stauffer, D. A.; Barrans, R. E., Jr.; Dougherty, D. A. *Angew. Chem. Int. Ed.* **1990**, *29*, 915.

(80) Stauffer, D. A.; Barrans, R. E., Jr.; Dougherty, D. A. *Journal of Organic Chemistry* **1990**, *55*, 2762.

(81) Sunner, J.; Nishizawa, K.; Kebarle, P. *J Phys Chem-Us* **1981**, *85*, 1814.

(82) Guo, B. C.; Purnell, J. W.; Castleman, A. W. *Chem Phys Lett* **1990**, *168*, 155.

(83) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.-C.; Lester, H. A.; Dougherty, D. A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12088.

(84) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9459.

(85) Dougherty, D. A. *Journal of Organic Chemistry* **2008**, *73*, 3667.

(86) Dougherty, D. A. *Science* **1996**, *271*, 163.

(87) Kumpf, R. A.; Dougherty, D. A. *Science* **1993**, *261*, 1708.

(88) Mecozzi, S.; West, A. P.; Dougherty, D. A. *P Natl Acad Sci USA* **1996**, *93*, 10566.

(89) Mecozzi, S.; West, A. P.; Dougherty, D. A. *J Am Chem Soc* **1996**, *118*, 2307.

(90) Cabarcos, O. M.; Weinheimer, C. J.; Lisy, J. M. *J Chem Phys* **1998**, *108*, 5151.

(91) Cabarcos, O. M.; Weinheimer, C. J.; Lisy, J. M. *J Chem Phys* **1999**, *110*, 8429.

(92) Cubero, E.; Luque, F. J.; Orozco, M. *P Natl Acad Sci USA* **1998**, *95*, 5976.

(93) Tsuzuki, S.; Yoshida, M.; Uchimaru, T.; Mikami, M. *Journal of Physical Chemistry A* **2001**, *105*, 769.

(94) Soteras, I.; Orozco, M.; Luque, F. J. *Phys Chem Chem Phys* **2008**, *10*, 2616.

(95) Tsuzuki, S.; Mikama, M.; Yamada, S. *J Am Chem Soc* **2007**, *129*, 8656.

(96) Battaglia, M. R.; Buckingham, A. D.; Williams, J. H. *Chem Phys Lett* **1981**, *78*, 421.

(97) Williams, J. H. *Accounts Chem Res* **1993**, *26*, 593.

(98) Wheeler, S. E.; Houk, K. N. *J Am Chem Soc* **2009**, *131*, 3126.

(99) Wheeler, S. E. *Accounts Chem Res* **2013**, *46*, 1029.

(100) Wheeler, S. E.; Bloom, J. W. G. *Journal of Physical Chemistry A* **2014**, *118*, 6133.

(101) Stone, A. J. *Chem Phys Lett* **1981**, *83*, 233.

(102) Stone, A. J.; Alderton, M. *Molecular Physics* **1984**, *56*, 1047.

(103) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 1996.

(104) Stone, A. J. *Science* **2008**, *321*, 787.

(105) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *Journal of Chemical Theory and Computation* **2013**, *9*, 4046.

(106) Wheeler, S. E.; Houk, K. N. *Journal of Chemical Theory and Computation* **2009**, *5*, 2301.

(107) Wheeler, S. E.; Bloom, J. W. G. *Chem. Commun.* **2014**, *50*, 11118.

(108) Bauzá, A.; Deyà, P. M.; Frontera, A.; Quiñonero, D. *Phys Chem Chem Phys* **2014**, *16*, 1322.

(109) Jeziorski, B.; Moszyński, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.

(110) Szalewicz, K. *WIREs Comp. Mol. Sci.* **2012**, *2*, 254.

(111) Premkumar, J. R.; Sastry, G. N. *Journal of Physical Chemistry A* **2014**, *118*, 11388.

(112) Bloom, J. W. G.; Wheeler, S. E. *Angew. Chem. Int. Ed.* **2011**, *50*, 7847.

(113) Ringer, A. L.; Sinnokrot, M. O.; Lively, R. P.; Sherrill, C. D. *Chem. Eur. J.* **2006**, *12*, 3821.

(114) Hunter, C. A.; Sanders, J. K. M. *J Am Chem Soc* **1990**, *112*, 5525.

(115) Wheeler, S. E.; Houk, K. N. *J Am Chem Soc* **2008**, *130*, 10854.

(116) Wheeler, S. E.; Houk, K. N. *Molecular Physics* **2009**, *107*, 749.

(117) Wheeler, S. E.; Houk, K. N. *J. Phys.Chem. A* **2010**, *114*, 8658.

(118) Wheeler, S. E.; McNeil, A. J.; Müller, P.; Swager, T. M.; Houk, K. N. *J Am Chem Soc* **2010**, *132*, 3304.

(119) Wheeler, S. E. *J Am Chem Soc* **2011**, *133*, 10262.

(120) Raju, R. K.; Bloom, J. W. G.; Wheeler, S. E. *Journal of Chemical Theory and Computation* **2013**, *9*, 3479.

(121) Zhu, D. Q.; Herbert, B. E.; Schlautman, M. A.; Carraway, E. R.; Hur, J. *J Environ Qual* **2004**, *33*, 1322.

(122) Zhu, D. Q.; Herbert, B. E.; Schlautman, M. A.; Carraway, E. R. *J Environ Qual* **2004**, *33*, 276.

(123) Gokel, George W.; De Wall, Stephen L.; Meadows, Eric S. *European Journal of Organic Chemistry* **2000**, *2000*, 2967.

(124) Wong, A.; Whitehead, R. D.; Gan, Z. H.; Wu, G. *Journal of Physical Chemistry A* **2004**, *108*, 10551.

(125) Wu, G.; Terskikh, V. *Journal of Physical Chemistry A* **2008**, *112*, 10359.

(126) Bryce, D. L.; Adiga, S.; Elliott, E. K.; Gokel, G. W. *Journal of Physical Chemistry A* **2006**, *110*, 13568.

(127) Cuc, D.; Canet, D.; Morel, J. P.; Morel-Desrosiers, N.; Mutzenhardt, P. *ChemPhysChem* **2007**, *8*, 643.

(128) Pozniak, B. P.; Dunbar, R. C. *J Am Chem Soc* **1997**, *119*, 10439.

(129) Ryzhov, V.; Dunbar, R. C.; Cerda, B.; Wesdemiotis, C. *J Am Soc Mass Spectr* **2000**, *11*, 1037.

(130) Gapeev, A.; Dunbar, R. C. *J Am Chem Soc* **2001**, *123*, 8360.

(131) Dunbar, R. C.; Steill, J. D.; Oomens, J. *J Am Chem Soc* **2011**, *133*, 9376.

(132) Amicangelo, J. C.; Armentrout, P. B. *Journal of Physical Chemistry A* **2000**, *104*, 11420.

(133) Amunugama, R.; Rodgers, M. T. *Int J Mass Spectrom* **2000**, *195*, 439.

(134) Amunugama, R.; Rodgers, M. T. *Journal of Physical Chemistry A* **2002**, *106*, 9718.

(135) Amunugama, R.; Rodgers, M. T. *Journal of Physical Chemistry A* **2002**, *106*, 9092.

(136) Yang, Z. B.; Rodgers, M. T. *J Am Chem Soc* **2004**, *126*, 16217.

(137) Ruan, C. H.; Rodgers, M. T. *J Am Chem Soc* **2004**, *126*, 14600.

(138) Ruan, C. H.; Yang, Z. B.; Hallowita, N.; Rodgers, M. T. *Journal of Physical Chemistry A* **2005**, *109*, 11539.

(139) Ruan, C. H.; Huang, H.; Rodgers, M. T. *Journal of Physical Chemistry A* **2007**, *111*, 13521.

(140) Hallowita, N.; Carl, D. R.; Armentrout, P. B.; Rodgerst, M. T. *Journal of Physical Chemistry A* **2008**, *112*, 7996.

(141) Hallowita, N.; Udonkang, E.; Ruan, C. H.; Frieler, C. E.; Rodgers, M. T. *Int J Mass Spectrom* **2009**, *283*, 35.

(142) Franski, R. *Rapid Commun Mass Sp* **2011**, *25*, 672.

(143) Rodriguez, J. D.; Kim, D.; Tarakeshwar, P.; Lisy, J. M. *Journal of Physical Chemistry A* **2010**, *114*, 1514.

(144) Dunbar, R. C.; Steill, J. D.; Polfer, N. C.; Oomens, J. *J Phys Chem B* **2009**, *113*, 10552.

(145) Dunbar, R. C.; Steill, J. D.; Oomens, J. *Phys Chem Chem Phys* **2010**, *12*, 13383.

(146) Morel, J. P.; Morel-Desrosiers, N. *Org Biomol Chem* **2006**, *4*, 462.

(147) Schlamadinger, D. E.; Daschbach, M. M.; Gokel, G. W.; Kim, J. E. *J Raman Spectrosc* **2011**, *42*, 633.

(148) Lucas, L. H.; Ersoy, B. A.; Kueltzo, L. A.; Joshi, S. B.; Brandau, D. T.; Thyagarajapuram, N.; Peek, L. J.; Middaugh, C. R. *Protein Sci* **2006**, *15*, 2228.

(149) Arnal-Herault, C.; Barboiu, M.; Petit, E.; Michau, M.; van der Lee, A. *New J Chem* **2005**, *29*, 1535.

(150) Spisak, S. N.; Zabula, A. V.; Filatov, A. S.; Rogachev, A. Y.; Petrukhina, M. A. *Angewandte Chemie-International Edition* **2011**, *50*, 8090.

(151) De Wall, S. L.; Meadows, E. S.; Barbour, L. J.; Gokel, G. W. *P Natl Acad Sci USA* **2000**, *97*, 6271.

(152) Gokel, G. W.; Barbour, L. J.; De Wall, S. L.; Meadows, E. S. *Coordin Chem Rev* **2001**, *222*, 127.

(153) Hu, J. X.; Barbour, L. J.; Gokel, G. W. *J Am Chem Soc* **2001**, *123*, 9486.

(154) Meadows, E. S.; De Wall, S. L.; Barbour, L. J.; Gokel, G. W. *J Am Chem Soc* **2001**, *123*, 3092.

(155) Gokel, G. W.; Barbour, L. J.; Ferdani, R.; Hu, J. X. *Accounts Chem Res* **2002**, *35*, 878.

(156) Hu, J. X.; Barbour, L. J.; Gokel, G. W. *P Natl Acad Sci USA* **2002**, *99*, 5121.

(157) Hu, J. X.; Barbour, L. J.; Gokel, G. W. *Chem Commun* **2002**, 1808.

(158) Hu, J. X.; Barbour, L. J.; Ferdani, R.; Gokel, G. W. *Chem Commun* **2002**, 1806.

(159) Gokel, G. W. *Chem Commun* **2003**, 2847.

(160) Weber, M. E.; Elliott, E. K.; Gokel, G. W. *Org Biomol Chem* **2006**, *4*, 83.

(161) Fukin, G. K.; Linderman, S. V.; Kochi, J. K. *J Am Chem Soc* **2002**, *124*, 8329.

(162) Schmitt, W.; Anson, C. E.; Hill, J. P.; Powell, A. K. *J Am Chem Soc* **2003**, *125*, 11142.

(163) Dolega, A.; Marynowski, W.; Baranowska, K.; Smiechowski, M.; Stangret, J. *Inorg Chem* **2012**, *51*, 836.

(164) Inokuchi, F.; Miyahara, Y.; Inazu, T.; Shinkai, S. *Angewandte Chemie-International Edition in English* **1995**, *34*, 1364.

(165) Kim, J.; Kim, Y. K.; Park, N.; Hahn, J. H.; Ahn, K. H. *Journal of Organic Chemistry* **2005**, *70*, 7087.

(166) Lelias-Vanderperre, A.; Chambron, J. C.; Espinosa, E.; Terrier, P.; Leize-Wagner, E. *Org Lett* **2007**, *9*, 2961.

(167) Shukla, R.; Lindeman, S. V.; Rathore, R. *J Am Chem Soc* **2006**, *128*, 5328.

(168) Shukla, R.; Lindeman, S. V.; Rathore, R. *Chem Commun* **2009**, 5600.

(169) Cametti, M.; Nissinen, M.; Cort, A. D.; Mandolini, L.; Rissanen, K. *J Am Chem Soc* **2005**, *127*, 3831.

(170) Ahman, A.; Nissinen, M. *Chem Commun* **2006**, 1209.

(171) Salorinne, K.; Nissinen, M. *Tetrahedron* **2008**, *64*, 1798.

(172) Macias, A. T.; Norton, J. E.; Evanseck, J. D. *J Am Chem Soc* **2003**, *125*, 2351.

(173) Wouters, J. *Protein Sci* **1998**, *7*, 2472.

(174) Pan, Y. P.; Zhang, K.; Qi, J. P.; Yue, J. A.; Springer, T. A.; Chen, J. F. *P Natl Acad Sci USA* **2010**, *107*, 21388.

(175) Santarelli, V. P.; Eastwood, A. L.; Dougherty, D. A.; Ahern, C. A.; Horn, R. *Biophysical Journal* **2007**, *93*, 2341.

(176) McFail-Isom, L.; Shui, X. Q.; Williams, L. D. *Biochemistry-Us* **1998**, *37*, 17105.

(177) Howerton, S. B.; Sines, C. C.; VanDerveer, D.; Williams, L. D. *Biochemistry-Us* **2001**, *40*, 10023.

(178) Cevec, M.; Plavec, J. *Biochemistry-Us* **2005**, *44*, 15238.

(179) Wu, G.; Wong, A.; Gan, Z. H.; Davis, J. T. *J Am Chem Soc* **2003**, *125*, 7182.

(180) Wu, G.; Wong, A. *Biochem Bioph Res Co* **2004**, *323*, 1139.

(181) Wong, A.; Ida, R.; Wu, G. *Biochem Bioph Res Co* **2005**, *337*, 363.

(182) Matsumura, H.; Yamamoto, T.; Leow, T. C.; Mori, T.; Salleh, A. B.; Basri, M.; Inoue, T.; Kai, Y.; Rahman, R. N. Z. R. A. *Proteins: Structure, Function, and Bioinformatics* **2008**, *70*, 592.

(183) Hagiwara, Y.; Matsumura, H.; Tateno, M. *J Am Chem Soc* **2009**, *131*, 16697.

(184) Thomas, K. J.; Sunoj, R. B.; Chandrasekhar, J.; Ramamurthy, V. *Langmuir* **2000**, *16*, 4912.

(185) Ramamurthy, V.; Shailaja, J.; Kaanumalle, L. S.; Sunoj, R. B.; Chandrasekhar, J. *Chem Commun* **2003**, 1987.

(186) Wang, X. L.; Zeng, Z.; Ahn, H.; Wang, G. X. *Appl Phys Lett* **2009**, *95*.

(187) Song, B.; Yang, J. W.; Zhao, J. J.; Fang, H. P. *Energ Environ Sci* **2011**, *4*, 1379.

(188) Umadevi, D.; Sastry, G. N. *J Phys Chem Lett* **2011**, *2*, 1572.

(189) Tachikawa, H.; Shimizu, A. *J Phys Chem B* **2005**, *109*, 13255.

(190) Tachikawa, H. *J Phys Chem C* **2011**, *115*, 20406.

(191) Umadevi, D.; Sastry, G. N. *J Phys Chem C* **2011**, *115*, 9656.

(192) Umadevi, D.; Sastry, G. N. *Chem Phys Lett* **2012**, *549*, 39.

(193) Valencia, F.; Romero, A. H.; Ancilotto, F.; Silvestrelli, P. L. *J Phys Chem B* **2006**, *110*, 14832.

(194) Mpourmpakis, G.; Tylianakis, E.; Papanikolaou, D.; Froudakis, G. *Rev Adv Mater Sci* **2006**, *11*, 92.

(195) Moradi, M.; Peyghan, A. A.; Bagheri, Z.; Kamfiroozi, M. *J Mol Model* **2012**, *18*, 3535.

(196) Hilder, T. A.; Gordon, D.; Chung, S. H. *Small* **2009**, *5*, 2870.

(197) Hilder, T. A.; Gordon, D.; Chung, S. H. *J Chem Phys* **2011**, *134*.

(198) Lakshminarasimhan, P.; Sunoj, R. B.; Chandrasekhar, J.; Ramamurthy, V. *J Am Chem Soc* **2000**, *122*, 4815.

(199) Wooten, A. J.; Carroll, P. J.; Walsh, P. J. *Org Lett* **2007**, *9*, 3359.

(200) Monje, P.; Paleo, M. R.; Garcia-Rio, L.; Sardina, F. J. *Journal of Organic Chemistry* **2008**, *73*, 7394.

(201) Kuwahara, Y.; Nishizawa, K.; Nakajima, T.; Kamegawa, T.; Mori, K.; Yamashita, H. *J Am Chem Soc* **2011**, *133*, 12462.

(202) Tan, E. M.; Cohen, A. S.; Fries, J. F.; Masi, A. T.; McShane, D. J.; Rothfield, N. F.; Schaller, J. G.; Talal, N.; Winchester, R. J. *Arthritis Rheum.* **1982**, *25*, 1271.

(203) Jang, Y. J.; Stollar, B. D. *Cell. Mol. Life Sci.* **2003**, *60*, 309.

(204) Blatt, N. B.; Glick, G. D. *Pharmacol. Ther.* **1999**, *83*, 125.

(205) Sibille, P.; Ternynck, T.; Nato, F.; Buttin, G.; Strosberg, D.; Avrameas, A. *Eur. J. Immunol.* **1997**, *27*, 1221.

(206) Amoura, Z.; Koutouzov, S.; Piette, J. G. *Curr. Opin. Rheumatol.* **2000**, *12*, 369.

(207) Schuermann, J. P.; Prewitt, S. P.; Davies, C.; Deutscher, S. L.; Tanner, J. J. *J. Mol. Biol.* **2005**, *347*, 965.

(208) Stevens, S. Y.; Glick, G. D. *Biochemistry* **1998**, *38*, 560.

(209) Braddock, D. T.; Baber, J. L.; Levens, D.; Clore, G. M. *EMBO J.* **2002**, *21*, 3476.

(210) Price, C. M.; Cech, T. R. *Genes Dev.* **1987**, *1*, 783.

(211) Gottschling, D. E.; Zakian, V. A. *Cell* **1986**, *47*, 195.

(212) Lin, J.-J.; Zakian, V. A. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13760.

(213) Baumann, P.; Cech, T. R. *Science* **2001**, *292*, 1171.

(214) Anderson, E. M.; Halsey, W. A.; Wuttke, D. S. *Biochem.* **2003**, *42*, 3751.

(215) Bochkarev, A.; Bochkareva, E. *Curr. Opin. Struct. Bio.* **2004**, *14*, 36.

(216) Nugent, C. I.; Hughes, T. R.; Lue, N. F.; Lundblad, V. *Science* **1996**, *274*, 249.

(217) Lohman, T. M.; Ferrari, M. E. *Annu. Rev. Biochem.* **1994**, *63*, 527.

(218) Wold, M. S. *Annu. Rev. Biochem.* **1997**, *66*, 61.

(219) Swamynathan, S. K.; Nambiar, A.; Guntaka, R. V. *FASEB J.* **1998**, *12*, 515.

(220) Raveh, S.; Vinh, J.; Rossier, J.; Agou, F.; Veron, M. *Biochem.* **2001**, *40*, 5882.

(221) Doe, R. C.; Bonanno, J. B.; Sonenberg, N.; Burley, S. K. *Cell* **1999**, *98*, 835.

(222) Varani, G.; Nagai, K. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 407.

(223) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 804.

(224) Cheng, A. C.; Chen, W. W.; Fuhrmann, C. N.; Frankel, A. D. *J. Mol. Biol.* **2003**, *327*, 781.

(225) Allers, J.; Shamoo, Y. *J. Mol. Biol.* **2001**, *311*, 75.

(226) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. *Nuc. Acid Res.* **2001**, *29*, 2860.

(227) Cheng, A. C.; Frankel, A. D. *J. Am. Chem. Soc.* **2003**, *126*, 434.

(228) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. *Nuc. Acid Res.* **2001**, *29*, 4294.

(229) Ahmad, S.; Gromiha, M. M.; Sarai, A. *Bioinformatics* **2004**, *20*, 477.

(230) Baker, C. M.; Grant, G. H. *Biopolymers* **2007**, *85*, 456.

(231) Lejeune, D.; Delsaux, N.; Charloteaux, B.; Thomas, A.; Brasseur, R. *Proteins: Struct. Funct. Bioinform.* **2005**, *61*, 258.

(232) Mandel-Gutfreund, Y.; Margalit, H. *Nuc. Acid Res.* **1998**, *26*, 2306.

(233) Rutledge, L. R.; Wetmore, S. D. *Can. J. Chem.* **2010**, *88*, 815.

(234) Rutledge, L. R.; Durst, H. F.; Wetmore, S. D. *J. Chem. Theory Comput.* **2009**, *5*, 1400.

(235) Rutledge, L. R.; Campbell-Verduyn, L. S.; Wetmore, S. D. *Chem. Phys. Lett.* **2007**, *444*, 167.

(236) Rutledge, L. R.; Campbell-Verduyn, L. S.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2006**, *110*, 19652.

(237) Copeland, K. L.; Anderson, J. A.; Farley, A. R.; Cox, J. R.; Tschumper, G. S. *J. Phys. Chem. B* **2008**, *112*, 14291.

(238) Tanner, J. J.; Komissarov, A. A.; Deutscher, S. L. *J. Mol. Biol.* **2001**, *314*, 807.

(239) Cygler, M.; Boodhoo, A.; Lee, J. S.; Anderson, W. F. *J. Biol. Chem.* **1987**, *262*, 643.

(240) Herron, J. N.; He, X. M.; Ballard, D. W.; Blier, P. R.; Pace, P. E.; Bothwell, A. L. M.; Voss, E. W.; Edmundson, A. B. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 159.

(241) Mol, C. D.; Muir, A. K. S.; Cygler, M.; Lee, J. S.; Anderson, W. F. *J. Biol. Chem.* **1994**, *269*, 3615.

(242) Ou, Z.; Bottoms, C. A.; Henzi, M. T.; Tanner, J. J. *J. Mol. Biol.* **2007**, *374*, 1029.

(243) Swanson, P. C.; Ackroyd, C.; Glick, G. D. *Biochemistry* **1996**, *35*, 1624.

(244) Lee, J. S.; Dombroski, D. F.; Mosmann, T. R. *Biochemistry* **1982**, *21*, 4940.

(245) Tan, E. M. *Adv. Immunol.* **1989**, *44*, 93.

(246) Rumbley, C. A.; Denzin, L. K.; Yantz, L.; Tetin, S. Y.; Voss, E. W. *J. Biol. Chem.* **1993**, *268*, 13667.

(247) Komissarov, A. A.; Calcutt, M. J.; Marchbank, M. T.; Peletskaya, E. N.; Deutscher, S. L. *J. Biol. Chem.* **1996**, *271*, 12241.

(248) Yokoyama, H.; Mizutani, R.; Satow, Y.; Komatsu, Y.; Ohtsuka, E.; Nikaido, O. *J. Mol. Biol.* **2000**, *299*, 711.

(249) Sanguineti, S.; Centeno Crowley, J. M.; Lodeiro Merlo, M. F.; Cerutti, M. L.; Wilson, I. A.; Goldbaum, F. A.; Stanfield, R. L.; de Prat-Gay, G. *J. Mol. Biol.* **2007**, *370*, 183.

(250) Churchill, C. D. M.; Wetmore, S. D. *J. Phys. Chem. B* **2009**, *113*, 16046.

(251) Rutledge, L. R.; Durst, H. F.; Wetmore, S. D. *J. Chem. Theory Comput.* **2009**, *5*, 1400.

(252) Rutledge, L. R.; Wetmore, S. D. *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **2010**, *88*, 815.

(253) Rutledge, L. R.; Churchill, C. D. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 3355.

(254) Rutledge, L. R.; Navarro-Whyte, L.; Peterson, T. L.; Wetmore, S. D. *J. Phys. Chem. A* **2011**, *115*, 12646.

(255) Schrodinger,, LLC: Portland, OR, 2010.

(256) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.

(257) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

(258) Case, D.; Darden, T. A.; Cheatham, T. E.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu,

X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. University of California, San Francisco, 2011.

(259) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(260) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523.

(261) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *The Journal of Chemical Physics* **1992**, *97*, 2635.

(262) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327.

(263) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of Chemical Physics* **1995**, *103*, 8577.

(264) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889.

(265) Grimme, S. *J. Comp. Chem.* **2006**, *27*, 1787.

(266) Becke, A. *J. Chem. Phys.* **1997**, *107*, 8554.

(267) Schafer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

(268) Raju, R. K.; Bloom, J. W. G.; An, Y.; Wheeler, S. E. *ChemPhysChem* **2011**, *12*, 3116.

(269) Wheeler, S. E. *J. Am. Chem. Soc.* **2011**, *133*, 10262.

(270) Bloom, J. W. G.; Raju, R. K.; Wheeler, S. E. *J. Chem. Theory Comput.* **2012**, *8*, 3167.

(271) Petersson, G. A.; Nyden, M. R. *J. Chem. Phys.* **1981**, *75*, 3223.

(272) Ju, K. S.; Parales, R. E. *Microbiol. Mol. Biol. R* **2010**, *74*, 250.

(273) Spain, J. C. *Annu. Rev. Microbiol.* **1995**, *49*, 523.

(274) Esteve-Núñez, A.; Caballero, A.; Ramos, J. L. *Microbiol. Mol. Biol. R* **2001**, *65*, 335.

(275) Gorontzy, T.; Drzyzga, O.; Kahl, M. W.; Bruns-nagel, D.; Breitung, J.; Loew, E. v.; Blotevogel, K. H. *Critical Reviews in Microbiology* **1994**, *20*, 265.

(276) Talmage, S.; Opresko, D.; Maxwell, C.; Welsh, C. E.; Cretella, F. M.; Reno, P.; Daniel, F. B. In *Reviews of Environmental Contamination and Toxicology*; Ware, G., Ed.; Springer New York: 1999; Vol. 161, p 1.

(277) Purohit, V.; Basu, A. K. *Chem. Res. Toxicol.* **2000**, *13*, 673.

(278) Michaeli, M. C.; Eisen, H. N. *J. Exp. Med.* **1974**, *140*, 687.

(279) Boelsterli, U. A.; Ho, H. K.; Zhou, S.; Leow, K. Y. *Curr. Drug Metab.* **2006**, *7*, 715.

(280) Ehrlich, J.; Bartz, Q. R.; Smith, R. M.; Joslyn, D. A.; Burkholder, P. R. *Science* **1947**, *106*, 417.

(281) Arima, K.; Fukuta, A.; Imanaka, H.; Kousaka, M.; Tamura, G. *Agr Biol Chem Tokyo* **1964**, *28*, 575.

(282) Leisinger, T.; Margraff, R. *Microbiol Rev* **1979**, *43*, 422.

(283) Patterson, S.; Wyllie, S. *Trends in Parasitology* **2014**, *30*, 289.

(284) Tiwari, R.; Mollmann, U.; Cho, S.; Franzblau, S. G.; Miller, P. A.; Miller, M. J. *Acs Med Chem Lett* **2014**, *5*, 587.

(285) Képiró, M.; Várkuti, B. H.; Végner, L.; Vörös, G.; Hegyi, G.; Varga, M.; Málnási-Csizmadia, A. *Angew. Chem. Int. Ed.* **2014**, *53*, 8211.

(286) Karjalainen, K.; Makela, O. *Eur. J. Immunol.* **1976**, *6*, 88.

(287) Farah, F. S. *Immunology* **1973**, *25*, 217.

(288) Ortega, E.; Kostovetzky, M.; Larralde, C. *Mol. Immunol.* **1984**, *21*, 883.

(289) Jormalai.S; Makela, O. *Eur. J. Immunol.* **1971**, *1*, 471.

(290) Shearer, G. M. *Eur. J. Immunol.* **1974**, *4*, 527.

(291) Nelson, D. L.; Poplack, D. G.; Holiman, B. J.; Henkart, P. A. *Clin. Exp. Immunol.* **1979**, *35*, 447.

(292) Six, H. R.; Uemura, K.; Kinsky, S. C. *Biochemistry* **1973**, *12*, 4003.

(293) Shokat, K. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1991**, *113*, 1861.

(294) Lu, Y. J.; Low, P. S. *Cancer Immunol Immun* **2002**, *51*, 153.

(295) Lu, Y. J.; Sega, E.; Low, P. S. *Int. J. Cancer* **2005**, *116*, 710.

(296) Lu, Y. J.; Sega, E.; Leamon, C. P.; Low, P. S. *Adv. Drug Delivery Rev.* **2004**, *56*, 1161.

(297) Lu, Y.; You, F.; Vlahov, I.; Westrick, E.; Fan, M.; Low, P. S.; Leamon, C. P. *Mol. Pharm.* **2007**, *4*, 695.

(298) Murelli, R. P.; Zhang, A. X.; Michel, J.; Jorgensen, W. L.; Spiegel, D. A. *J. Am. Chem. Soc.* **2009**, *131*, 17090.

(299) Jakobsche, C. E.; McEnaney, P. J.; Zhang, A. X.; Spiegel, D. A. *ACS Chem. Biol.* **2012**, *7*, 315.

(300) Cui, L.; Kitov, P. I.; Completo, G. C.; Paulson, J. C.; Bundle, D. R. *Bioconjugate Chem.* **2011**, *22*, 546.

(301) O'Reilly, M. K.; Collins, B. E.; Han, S.; Liao, L.; Rillahan, C.; Kitov, P. I.; Bundle, D. R.; Paulson, J. C. *J. Am. Chem. Soc.* **2008**, *130*, 7736.

(302) Jakobsche, C. E.; Parker, C. G.; Tao, R. N.; Kolesnikova, M. D.; Douglass, E. F.; Spiegel, D. A. *ACS Chem. Biol.* **2013**, *8*, 2404.

(303) McEnaney, P. J.; Parker, C. G.; Zhang, A. X.; Spiegel, D. A. *ACS Chem. Biol.* **2012**, *7*, 1139.

(304) Parker, C. G.; Dahlgren, M. K.; Tao, R. N.; Li, D. T.; Douglass, E. F.; Shoda, T.; Jawanda, N.; Spasov, K. A.; Lee, S.; Zhou, N.; Domaoal, R. A.; Sutton, R. E.; Anderson, K. S.; Krystal, M.; Jorgensen, W. L.; Spiegel, D. A. *Chem. Sci.* **2014**, *5*, 2311.

(305) Dubrovska, A.; Kim, C.; Elliott, J.; Shen, W. J.; Kuo, T. H.; Koo, D. I.; Li, C.; Tuntland, T.; Chang, J.; Groessl, T.; Wu, X.; Gorney, V.; Ramirez-Montagut, T.; Spiegel, D. A.; Cho, C. Y.; Schultz, P. G. *ACS Chem. Biol.* **2011**, *6*, 1223.

(306) Parker, C. G.; Domaoal, R. A.; Anderson, K. S.; Spiegel, D. A. *J. Am. Chem. Soc.* **2009**, *131*, 16392.

(307) Zhang, A. X.; Murelli, R. P.; Barinka, C.; Michel, J.; Cocleaza, A.; Jorgensen, W. L.; Lubkowski, J.; Spiegel, D. A. *J. Am. Chem. Soc.* **2010**, *132*, 12711.

(308) Fura, J. M.; Sabulski, M. J.; Pires, M. M. *ACS Chem. Biol.* **2014**, *9*, 1480.

(309) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.

(310) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651.

(311) Cockroft, S. L.; Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C., J. *J. Am. Chem. Soc.* **2005**, *127*, 8594.

(312) Cockroft, S. L.; Perkins, J.; Zonta, C.; Adams, H.; Spey, S. E.; Low, C. M. R.; Vinter, J. G.; Lawson, K. R.; Urch, C. J.; Hunter, C. A. *Org. Biomol. Chem.* **2007**, *5*, 1062.

(313) Hargis, J. C.; Schaefer, H. F.; Houk, K. N.; Wheeler, S. E. *J. Phys. Chem. A* **2010**, *114*, 2038.

(314) An, Y.; Raju, R. K.; Lu, T.; Wheeler, S. E. *J. Phys. Chem. B* **2014**, *118*, 5653.

(315) Rutledge, L. R.; Durst, H. F.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2801.

(316) Rutledge, L. R.; Wetmore, S. D. *J. Chem. Theory and Comput.* **2008**, *4*, 1768.

(317) Churchill, C. D. M.; Navarro-Whyte, L.; Rutledge, L. R.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10657.

(318) Churchill, C. D. M.; Wetmore, S. D. *J. Phys. Chem. B* **2009**, *113*, 16046.

(319) Churchill, C. D. M.; Rutledge, L. R.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14515.

(320) Rutledge, L. R.; Churchill, C. D. M.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 3355.

(321) Churchill, C. D. M.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16373.

(322) Kellie, J. L.; Mavarro-Whyte, L.; Carvey, M. T.; Wetmore, S. D. *J. Phys. Chem. B* **2012**, *116*, 2622.

(323) Wells, R. A.; Kellie, J. L.; Wetmore, S. D. *J. Phys. Chem. B* **2013**, *117*, 10462.

(324) Wilson, K. A.; Kellie, J. L.; Wetmore, S. D. *Nucleic Acids Res.* **2014**, *42*, 6726.

(325) Copeland, K. L.; Pellock, S. J.; Cox, J. R.; Cafiero, M. L.; Tschumper, G. S. *J. Phys. Chem. B* **2013**, *117*, 14001.

(326) James, L. C.; Roversi, P.; Tawfik, D. S. *Science* **2003**, *299*, 1362.

(327) The nitroarene was placed above the amino acid side chain so that the centers of mass were coincident, and also displaced +/-2.5 A along the x- and y-axes.

(328) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(329) Papajak, E.; Truhlar, D. G. *J. Chem. Theory and Comput.* **2011**, *7*, 10.

(330) Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 184111.

(331) Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *133*, 014101.

(332) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. *J. Chem. Phys.* **2014**, *140*, 094106.

(333)

(334) *MOLPRO, version 2010.1, is a package of ab initio programs written by H.-J. Werner, et al.*

(335) 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. *WIREs Comp. Mol. Sci.* **2012**, *2*, 556.

(336) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(337) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.

(338) Arnstein, S. A.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2646.

(339) Ringer, A. L.; Sherrill, C. D. *J. Am. Chem. Soc.* **2009**, *131*, 4574.

(340) The protein structures containing nitroarene ligands that met our criteria were as follows: 18GS, 1AI5, 1CLA, 1D0C, 1D0O, 1D0Y, 1D1Q, 1DOH, 1E2S, 1E36, 1E37, 1E38, 1E4I, 1EOC, 1F2B, 1F42, 1GLQ, 1GSQ, 1GVO, 1GVS, 1H2J, 1HNA, 1I9H, 1IDT, 1IJ8, 1IVB, 1IVD, 1J8V, 1JMZ, 1JR0, 1JYV, 1JYW, 1K0C, 1K4Q, 1KBQ, 1LOP, 1LS6, 1M2Q, 1M9T, 1NO3, 1OAU, 1OM5, 1PA9, 1PIP, 1PX0, 1Q3W, 1QIP, 1QW9, 1RD4, 1RMH, 1RTH, 1RXE, 1RXJ, 1RXK, 1SDA, 1T47, 1USQ, 1UYQ, 1V9T, 1VAH, 1VBS, 1VBT, 1VF3, 1VID, 1VYP, 1VYR, 1VYS, 1XGI, 1XI2, 1XWK, 1Y2K, 1YEK, 1YPM, 1YTJ, 1Z44, 1ZMT, 1ZWP, 1ZX1, 2ADP, 2AX6, 2B14, 2B15, 2B16, 2BMQ, 2BMR, 2BUU, 2BUZ, 2BZS, 2C70, 2CE2, 2CL0, 2CL6, 2CL7, 2CLC, 2D20, 2DQT, 2DQU, 2EVW, 2FHL, 2FHN, 2G70, 2GGX, 2GQN, 2GZ7, 2H5U, 2HJ4, 2HMO, 2I10, 2I6P, 2IKJ, 2IVD, 2JKJ, 2JKL, 2NRU, 2O5D, 2OAC, 2OAD, 2OF8, 2OFB, 2PB1, 2PDC, 2PDJ, 2PDP, 2PDU, 2PER, 2PZN, 2Q95, 2QMC, 2QN8, 2QP6, 2RAZ, 2SAM, 2UWX, 2V96, 2VCZ, 2VO4, 2W7X, 2WD1, 2WGZ, 2WHH, 2WNB, 2WNF, 2WVU, 2X4N, 2X4P, 2X4Q, 2X70, 2XAP, 2XEF, 2XOF, 2Y2I, 2Y59, 2ZHM,

2ZHN, 2ZOX, 2ZVP, 2ZYV, 2ZYW, 3A70, 3ABX, 3AI0, 3AI8, 3AIR, 3AIW, 3B0R, 3B7D, 3BGU, 3BWM, 3BWY, 3CHT, 3CZR, 3DH8, 3DIV, 3DS9, 3ESC, 3ESD, 3ETT, 3F64, 3G4G, 3G4I, 3GUS, 3H82, 3HV5, 3IE3, 3INQ, 3IOD, 3IX4, 3JVS, 3LE6, 3LM1, 3LXV, 3M3C, 3M3E, 3M3O, 3M64, 3M8T, 3MC5, 3MK0, 3MK1, 3MWI, 3NKS, 3NYD, 3NZ1, 3O5N, 3O76, 3OJK, 3PTQ, 3Q2D, 3QL6, 3QVO, 3T9U, 3T9V, 43CA, 4A3H, 5GST, and 5TLN.

(341) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210.

(342) Salonen, L. M.; Ellermann, M.; Diederich, F. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808.

(343) Bissantz, C.; Kuhn, B.; Stahl, F. *J. Med. Chem.* **2010**, *53*, 5061.

(344) Gomtsyan, A. *Chem Heterocycl Comp* **2012**, *48*, 7.

(345) Feher, M.; Schmidt, J. M. *J. Chem. Inf. Model.* **2003**, *43*, 218.

(346) Gakh, A. A.; Kirk, K. L. *Acs Sym Ser* **2009**, *1003*, 3.

(347) Pozharskii, A. F.; Soldatenkov, A. T.; Katritzky, A. R. In *Heterocycles in Life and Society*; John Wiley & Sons, Ltd: 2011, p 139.

(348) Wu, Y.-J. In *Progress in Heterocyclic Chemistry*; Gordon, W. G., John, A. J., Eds.; Elsevier: 2012; Vol. Volume 24, p 1.

(349) Kaushik, N. K.; Kaushik, N.; Attri, P.; Kumar, N.; Kim, C. H.; Verma, A. K.; Choi, E. H. *Molecules* **2013**, *18*, 6620.

(350) Salat, K.; Moniczewski, A.; Librowski, T. *Mini-Rev Med Chem* **2013**, *13*, 335.

(351) Baumann, M.; Baxendale, I. R. *Beilstein Journal of Organic Chemistry* **2013**, *9*, 2265.

(352) Vitaku, E.; Smith, D.; Njardarson, J. *Journal of medicinal chemistry* **2014**, *57*, 10257.

(353) Salonen, L. M.; Ellermann, M.; Diederich, F. *Angew Chem Int Edit* **2011**, *50*, 4808.

(354) Janjic, G. V.; Ninkovic, D. B.; Zaric, S. D. *Acta Crystallogr B* **2013**, *69*, 389.

(355) Kumar, S.; Das, A. *Journal of Chemical Physics* **2013**, *139*, 104311.

(356) Li, P.; Zhao, C.; Smith, M. D.; Shimizu, K. D. *The Journal of Organic Chemistry* **2013**, *78*, 5303.

(357) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. *Chemical Reviews* **2010**, *110*, 5023.

(358) Sherrill, C. D. *Accounts Chem Res* **2013**, *46*, 1020.

(359) Zhang, C. *Journal of Computational Chemistry* **2011**, *32*, 152.

(360) Smith, Q. A.; Gordon, M. S.; Slipchenko, L. V. *J Phys Chem A* **2011**, *115*, 4598.

(361) Majumder, M.; Sathyamurthy, N. *Theor Chem Acc* **2012**, *131*.

(362) Ninkovic, D. B.; Janjic, G. V.; Zaric, S. D. *Cryst Growth Des* **2012**, *12*, 1060.

(363) Gholipour, A. R.; Saydi, H.; Neiband, M. S.; Neyband, R. S. *Struct Chem* **2012**, *23*, 367.

(364) Zhikol, O. A.; Shishkin, O. V. *Int J Quantum Chem* **2012**, *112*, 3008.

(365) Ninkovic, D. B.; Andric, J. M.; Zaric, S. D. *Chemphyschem* **2013**, *14*, 237.

(366) Mottishaw, J. D.; Sun, H. R. *J Phys Chem A* **2013**, *117*, 7970.

(367) Lutz, P. B.; Bayse, C. A. *Phys Chem Chem Phys* **2013**, *15*, 9397.

(368) Hohenstein, E. G.; Sherrill, C. D. *J Phys Chem A* **2009**, *113*, 878.

(369) Wang, W. Z.; Hobza, P. *Chemphyschem* **2008**, *9*, 1003.

(370) Guin, M.; Patwari, G. N.; Karthikeyan, S.; Kim, K. S. *Phys Chem Chem Phys* **2009**, *11*, 11207.

(371) Mishra, B. K.; Arey, J. S.; Sathyamurthy, N. *J Phys Chem A* **2010**, *114*, 9606.

(372) Guin, M.; Patwari, G. N.; Karthikeyan, S.; Kim, K. S. *Phys Chem Chem Phys* **2011**, *13*, 5514.

(373) Mackie, I. D.; McClure, S. A.; DiLabio, G. A. *J Phys Chem A* **2009**, *113*, 5476.

(374) Geng, Y.; Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J Phys Chem A* **2010**, *114*, 3576.

(375) Karthikeyan, S.; Nagase, S. *J Phys Chem A* **2012**, *116*, 1694.

(376) Kumar, S.; Das, A. *Journal of Chemical Physics* **2012**, *137*.

(377) Koziol, L.; Kumar, N.; Wong, S. E.; Lightstone, F. C. *J Phys Chem A* **2013**, *117*, 12946.

(378) Gao, W.; Jiao, J. Q.; Feng, H. J.; Xuan, X. P.; Chen, L. P. *Theor Chem Acc* **2013**, *132*.

(379) Huber, R. G.; Margreiter, M. A.; Fuchs, J. E.; von Grafenstein, S.; Tautermann, C. S.; Liedl, K. R.; Fox, T. *Journal of Chemical Information and Modeling* **2014**, *54*, 1371.

(380) Newcomb, L. F.; Gellman, S. H. *J. Am. Chem. Soc.* **1994**, *116*, 4993.

(381) McKay, S. L.; Haptonstall, B.; Gellman, S. H. *J. Am. Chem. Soc.* **2001**, *123*, 1244.

(382) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.

(383) Stone, A. J.; Alderton, M. *Mol. Phys.* **1984**, *56*, 1047.

(384) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128.

(385) Briefly, multipole expansions of electrostatic interactions between molecules are typically divergent for large systems in stacking configurations. Thus, although the leading dipole-dipole term of this expansion can provide qualitatively correct predictions of the overall electrostatic interaction, it is the first term in an ultimately divergent expansion.

(386) Wheeler, S. E.; Houk, K. N. *J. Am. Chem. Soc.* **2008**, *130*, 10854.

(387) Wheeler, S. E.; Houk, K. N. *J. Chem. Theory Comput.* **2009**, *5*, 2301.

(388) Wheeler, S. E.; McNeil, A. J.; Müller, P.; Swager, T. M.; Houk, K. N. *J. Am. Chem. Soc.* **2010**, *132*, 3304.

(389) Wheeler, S. E. *J. Am. Chem. Soc.* **2011**, *133*, 10262.

(390) Raju, R. K.; Bloom, J. W. G.; Wheeler, S. E. *J. Chem. Theory Comput.* **2013**, *9*, 3479.

(391) Wheeler, S. E. *Acc. Chem. Res.* **2013**, *46*, 1029.

(392) Wheeler, S. E.; Bloom, J. W. G. *J. Phys. Chem. A* **2014**, *118*, 6133.

(393) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.

(394) Broughton, H.; Watson, I. *Journal of molecular graphics & modelling* **2004**, *23*, 51.

(395) Ertl, P.; Jelfs, S.; Muhlbacher, J.; Schuffenhauer, A.; Selzer, P. *Journal of medicinal chemistry* **2006**, *49*, 4568.

(396) Gao, Q.; Williams, L. D.; Egli, M.; Rabinovich, D.; Chen, S. L.; Quigley, G. J.; Rich, A. *Proceedings of the National Academy of Sciences* **1991**, *88*, 2422.

(397) Zimmermann, G. R.; Jenison, R. D.; Wick, C. L.; Simorre, J. P.; Pardi, A. *Nat Struct Biol* **1997**, *4*, 644.

(398) Weisberg, E.; Manley, P. W.; Breitenstein, W.; Brüggen, J.; Cowan-Jacob, S. W.; Ray, A.; Huntly, B.; Fabbro, D.; Fendrich, G.; Hall-Meyers, E.; Kung, A. L.; Mestan, J.; Daley, G. Q.; Callahan, L.; Catley, L.; Cavazza, C.; Mohammed, A.; Neuberg, D.; Wright, R. D.; Gilliland, D. G.; Griffin, J. D. *Cancer Cell* **2005**, *7*, 129.

(399) Seward, H. E.; Roujeinikova, A.; McLean, K. J.; Munro, A. W.; Leys, D. *Journal of Biological Chemistry* **2006**, *281*, 39437.

(400) Andersen, O. A.; Schonfeld, D. L.; Toogood-Johnson, I.; Felicetti, B.; Albrecht, C.; Fryatt, T.; Whittaker, M.; Hallett, D.; Barker, J. *Acta Crystallogr D* **2009**, *65*, 872.

(401) Rettig, M.; Langel, W.; Kamal, A.; Weisz, K. *Org Biomol Chem* **2010**, *8*, 3179.

(402) Velvadapu, V.; Paul, T.; Wagh, B.; Glassford, I.; DeBrosse, C.; Andrade, R. B. *The Journal of Organic Chemistry* **2011**, *76*, 7516.

(403) Velvadapu, V.; Paul, T.; Wagh, B.; Klepacki, D.; Guvench, O.; MacKerell, A.; Andrade, R. B. *ACS Medicinal Chemistry Letters* **2011**, *2*, 68.

(404) Velvadapu, V.; Glassford, I.; Lee, M.; Paul, T.; DeBrosse, C.; Klepacki, D.; Small, M. C.; MacKerell, A. D.; Andrade, R. B. *ACS Medicinal Chemistry Letters* **2012**, *3*, 211.

(405) Wagh, B.; Paul, T.; Glassford, I.; DeBrosse, C.; Klepacki, D.; Small, M. C.; MacKerell, A. D.; Andrade, R. B. *ACS Medicinal Chemistry Letters* **2012**, *3*, 1013.

(406) Glassford, I.; Lee, M.; Wagh, B.; Velvadapu, V.; Paul, T.; Sandelin, G.; DeBrosse, C.; Klepacki, D.; Small, M. C.; MacKerell, A. D.; Andrade, R. B. *ACS Medicinal Chemistry Letters* **2014**, *5*, 1021.

(407) Dunkle, J. A.; Xiong, L.; Mankin, A. S.; Cate, J. H. D. *Proceedings of the National Academy of Sciences* **2010**, *107*, 17152.

(408) Bulkley, D.; Innis, C. A.; Blaha, G.; Steitz, T. A. *Proceedings of the National Academy of Sciences* **2010**, *107*, 17158.

(409) Poehlsgaard, J.; Pfister, P.; Bottger, E. C.; Douthwaite, S. *Antimicrobial agents and chemotherapy* **2005**, *49*, 1553.

(410) Zhanel, G.; Walters, M.; Noreddin, A.; Vercaigne, L.; Wierzbowski, A.; Embil, J.; Gin, A.; Douthwaite, S.; Hoban, D. *Drugs* **2002**, *62*, 1771.

(411) Rafie, S.; MacDougall, C.; James, C. L. *Pharmacotherapy* **2010**, *30*, 290.

(412) Llano Sotelo, B.; Dunkle, J.; Klepacki, D.; Zhang, W.; Fernandes, P.; Zhang, W.; Mankin, A. *Antimicrobial agents and chemotherapy* **2010**, *54*, 4961.

(413) Chai, J.-D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.

(414) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

(415) Peverati, R.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2011**, *2*, 2810.

(416) Grimme, S. *WIREs Comp. Mol. Sci.* **2011**, *1*, 211.

(417) Vázquez-Mayagoitia, Á.; Sherrill, C. D.; Apra, E.; Sumpter, B. G. *J. Chem. Theory Comput.* **2010**, *6*, 727.

(418) Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. M. *J Mol Biol* **2002**, *324*, 105.

(419) Shaw, K. L.; Grimsley, G. R.; Yakovlev, G. I.; Makarov, A. A.; Pace, C. N. *Protein Science* **2001**, *10*, 1206.

(420) Riès-kautt, M.; Ducruix, A. In *Methods in Enzymology*; Academic Press: 1997; Vol. Volume 276, p 23.

(421) Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 2403.

(422) Matthew, J. B.; Gurd, F. R. N.; Garciamoreno, E. B.; Flanagan, M. A.; March, K. L.; Shire, S. J. *CRC Critical Reviews in Biochemistry* **1985**, *18*, 91.

(423) Pace, C. N.; Hermans, J. *CRC Critical Reviews in Biochemistry* **1975**, *3*, 1.

(424) Tanford, C. In *Advances in Protein Chemistry*; C.B. Anfinsen, M. L. A. J. T. E., Frederic, M. R., Eds.; Academic Press: 1968; Vol. Volume 23, p 121.

(425) Tanford, C. *Physical chemistry of macromolecules*; Wiley: New York,, 1961.

(426) Schulz, G. E.; Schirmer, R. H. *Principles of protein structure*; Springer-Verlag: New York, 1979.

(427) Fersht, A. *Enzyme structure and mechanism*; 2nd ed.; W.H. Freeman: New York, 1985.

(428) Markley, J. L. *Accounts of Chemical Research* **1975**, *8*, 70.

(429) Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D. *Proteins: Structure, Function, and Bioinformatics* **2002**, *48*, 388.

(430) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; Garc á-Moreno, E. B. *Biophysical Journal* **2002**, *82*, 3289.

(431) Nozaki, Y.; Tanford, C. *J Am Chem Soc* **1967**, *89*, 736.

(432) Giletto, A.; Pace, C. N. *Biochemistry* **1999**, *38*, 13379.

(433) Tanford, C. In *Advances in Protein Chemistry*; C.B. Anfinsen, K. B. M. L. A., John, T. E., Eds.; Academic Press: 1963; Vol. Volume 17, p 69.

(434) Garc á-Moreno, B. E.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. *Biophysical Chemistry* **1997**, *64*, 211.

(435) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; Garc á-Moreno E, B. *Biophysical Journal* **2000**, *79*, 1610.

(436) Edgcomb, S. P.; Murphy, K. P. *Proteins: Structure, Function, and Bioinformatics* **2002**, *49*, 1.

(437) Harris, T. K.; Turner, G. J. *IUBMB Life* **2002**, *53*, 85.

(438) Laurents, D. V.; Huyghues-Despointes, B. M. P.; Bruix, M.; Thurlkill, R. L.; Schell, D.; Newsom, S.; Grimsley, G. R.; Shaw, K. L.; Treviño, S.; Rico, M.; Briggs, J. M.; Antosiewicz, J. M.; Scholtz, J. M.; Pace, C. N. *J Mol Biol* **2003**, *325*, 1077.

(439) Horng, J.-C.; Cho, J.-H.; Raleigh, D. P. *J Mol Biol* **2005**, *345*, 163.

(440) Trevino, S. R.; Gokulan, K.; Newsom, S.; Thurlkill, R. L.; Shaw, K. L.; Mitkevich, V. A.; Makarov, A. A.; Sacchettini, J. C.; Scholtz, J. M.; Pace, C. N. *J Mol Biol* **2005**, *354*, 967.

(441) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Science : A Publication of the Protein Society* **2009**, *18*, 247.

(442) Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. *The Journal of Biological Chemistry* **2009**, *284*, 13285.

(443) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Science* **2006**, *15*, 1214.

(444) Gilson, M. K. *Current Opinion in Structural Biology* **1995**, *5*, 216.

(445) Schutz, C. N.; Warshel, A. *Proteins: Structure, Function, and Bioinformatics* **2001**, *44*, 400.

(446) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophysical Journal* **2002**, *83*, 1731.

(447) Braun-Sand, S.; Warshel, A. In *Protein Science Encyclopedia*; Wiley-VCH Verlag GmbH & Co. KGaA: 2008.

(448) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2005**, *61*, 704.

(449) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73*, 765.

(450) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 525.

(451) Wheeler, S. E.; Houk, K. N. *J Am Chem Soc* **2008**, *130*, 10854.

(452) Wheeler, S. E. *Accounts of Chemical Research* **2013**, *46*, 1029.

(453) Wheeler, S. E.; Bloom, J. W. G. *The Journal of Physical Chemistry A* **2014**, *118*, 6133.

(454) An, Y.; Raju, R. K.; Lu, T.; Wheeler, S. E. *The Journal of Physical Chemistry B* **2014**, *118*, 5653.

(455) Becke, A. D. *Journal of Chemical Physics* **1997**, *107*, 8554.

(456) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.

(457) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

APPENDIX

**Table A-1.** MMGBSA predicted pairwise binding enthalpies (kcal $mol^{-1}$) between the 3' cytosine and key amino acid residues for four dinucleotides with ED-10.

| Res. | dTdC | dAdC | dCdC | dGdC |
|---|---|---|---|---|
| HisL27D[a] | -5.8 | -5.4 | -5.5 | -5.8 |
| TyrH97[a] | -4.0 | -3.7 | -4.2 | -3.3 |
| TyrL32 | -3.2 | -3.6 | -3.2 | -3.4 |
| GlyH98 | -2.5 | -1.9 | -2.5 | -1.8 |
| SerH99 | -2.1 | -0.7 | -2.2 | -0.3 |
| AsnL28 | -1.6 | -1.3 | -1.4 | -1.2 |
| TrpH95 | -1.1 | -1.1 | -1.1 | -0.9 |
| GlyL91 | -0.8 | -1.9 | -0.6 | -2.2 |

[a]This residue interacts with the back bone of 5' cytosine.

**Table A-2.** MMGBSA predicted binding enthalpies (kcal $mol^{-1}$), relative to the native antigen 5'-dTdC, using single trjactory approach and multiple trajectory approach.

| | Single Traj | Multiple Traj |
|---|---|---|
| 5'-dTdC | 0.0 | 0.0 |
| 5'-dAdC | 8.7 | 15.5 |
| 5'-dGdC | 0.7 | -1.5 |
| 5'-dCdC | 14.7 | 21.7 |

**Table A-3.** Total interaction energy (kcal $mol^{-1}$) of the 5'-base with the 5'-binding site in the gas phase and in water using polarizable continuum model (PCM).

| | Gas Phase | PCM |
|---|---|---|
| 5'-dTdC | -38.8 | -25.7 |
| 5'-dAdC | -32.7 | -24.9 |
| 5'-dGdC | -22.8 | -13.2 |
| 5'-dCdC | -45.2 | -28.9 |

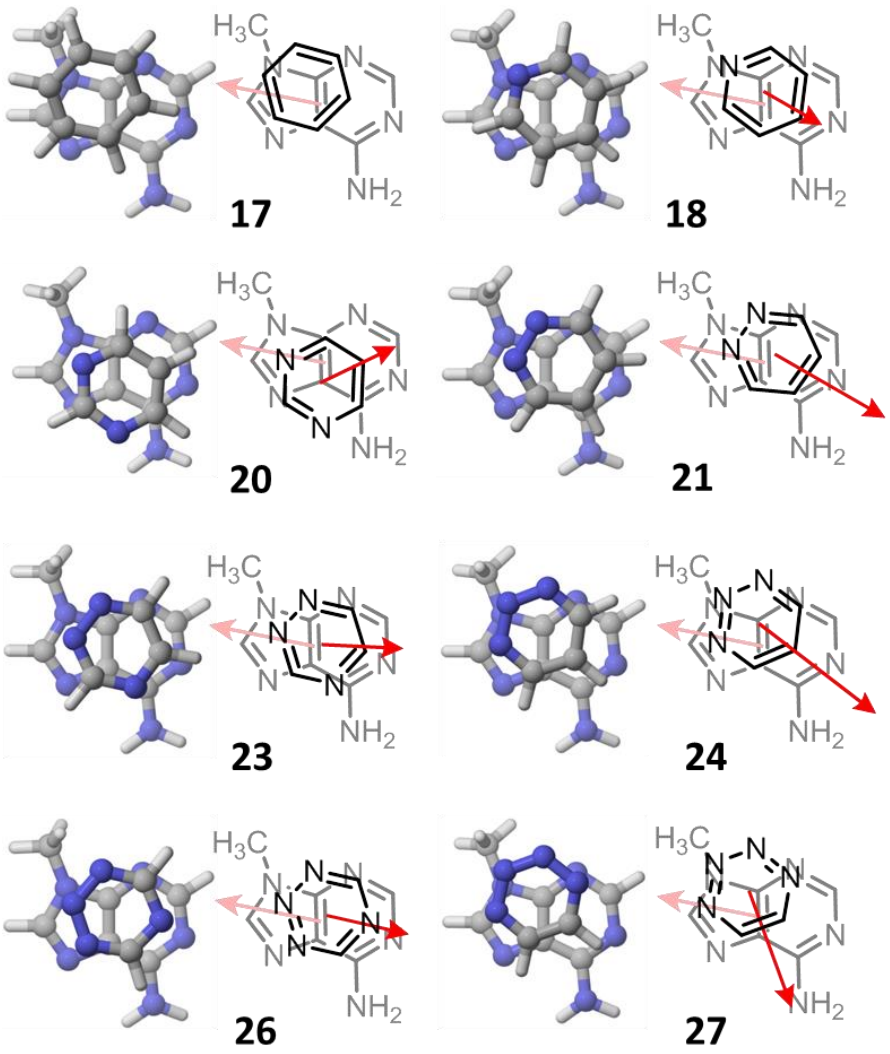



**Figure A-1.** Alignment of molecular dipole in stacked dimers of polar azines **18**, **20**, **21**, **23**, **24**, **26**, and **27** with 9-methyladenine. Even though molecular dipole moments provide qualitatively correct predictions of preferred orientations in these cases, this does not hold for many of the other small heterocycles.