The Development of an Empirically Corrected Semi-Empirical Method

And its

Application to Macromolecular Complexes

A Thesis

Submitted to the Faculty

of

Drexel University

by

Michael E. Foster

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

May 2011

Acknowledgments

I would like to thank my advisor Dr. Karl Sohlberg for helping and inspiring me to pursue my Ph.D. in theoretical and computational chemistry. Without his guidance and theoretical intellect this achievement would not have been possible. My future success will in part be accredited, to my mentor and colleague, Dr. Karl Sohlberg.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AM1	Austin Method 1
AM1-FS1	AM1 Method with Semi-Empirical Corrections for Dispersion and Hydrogen-Bonding Interactions
CBPQT	Cyclobis(paraquat-p-phenylene)
CCSD(T)	Coupled-Cluster Theory with Single and Double and Perturbative Triple Excitations
CM-distance	Center-of-Mass Distance
CP	Counterpoise Correction
DFT	Density Functional Theory
DFT-D	DFT with a Semi-Empirical Correction for Dispersion Interaction (e.g. B97-D, BLYP-D, TPSS-D)
F66	Database Consists of 66 Molecular Complexes for which H-bonding and Dispersion Interactions are Important; Used for the Parameterization of AM1-FS1
GAMESS	General Atomic and Molecular Electronic Structure System
HF	Hartree-Fock
PAH	Poly-Cyclic Aromatic Hydrocarbons
PEC	Potential Energy Curve

PM3	Parameterized Model Number 3
PM6	Parameterized Model Number 6
PM6-DH	PM6 Method with Semi-Empirical Corrections for Dispersion and Hydrogen-Bonding Interactions
RM1	Recife Model 1
RMSE	Root Mean Square Error
S22 Database	Database Consisting of 22 Molecular Complexes for which H-bonding and Dispersion Interactions are Important
SAPT	Symmetry-Adapted Perturbation Theory
SCF	Self-Consistent Field
SE methods	Semi-empirical methods (e.g. AM1, RM1, PM3, & PM6)
SE-D methods	Semi-Empirical Methods with a Semi-empirical Correction for Dispersion Interaction (e.g. AM1-D & PM3-D)
XC-functional	Exchange-Correlation functional (e.g. B97, BLYP, B3LYP, PBE, M06-L, M06-2X, M08, TPSS)

Abstract

The Development of An Empirically Corrected Semi-Empirical Method and its Application to Macromolecular Complexes

Michael E. Foster

Karl W. Sohlberg, Ph.D.

Computational chemistry is a growing field crossing interdisciplinary scientific fields because of the ability to predict physical properties while reducing costs and waste materials; however, there are limiting factors. Computationally modeling systems governed by non-bonded interactions, especially van der Waals (dispersion) interactions, is currently a difficult task, since many conventional quantum mechanical techniques neglect such interactions. Methods that are capable of modeling such interactions are computationally extremely expensive, limiting system size to only a few dozen atoms. Therefore, such computations are intractable for exploring in the upper limits of the nanoscopic world. One avenue of nanotechnology involves engineering machines at the molecular level that are capable of producing useful work. Such devices promise to be applicable in a wide range of areas, such as molecular-scale electronics, nanometer-scale engineering, medicine, and space science to name a few. In order to model such large systems, semi-empirical methods appear to be an attractive option; however, the popular semi-empirical methods (e.g. AM1) do not model long-range dispersion but this is not their only shortcoming. For weakly interacting systems, hydrogen bonding also poses a concern. Therefore, an empirically-corrected AM1 method that uses two empirical correction terms, one for dispersion and one for hydrogen-bonding interactions, has been developed and termed AM1-FS1. The AM1-FS1 method has been tested and used to study several carbon nanostructure complexes and rotaxane systems and is found to

produce results in good agreement with experimental and other first-principles calculations.

Chapter 1: Introduction

Intermolecular interactions such as dispersion and hydrogen-bonding play a critical role in determining the structure and properties of many molecular systems of current interest. For example: Protein folding is currently of vast interest in the scientific community[1-4] because of its involvement in many biological processes, including but not limited to "the trafficking of molecules to specific cellular locations and the regulation of cellular growth and differentiation[4]." The secondary structure of proteins is, in part, dependent on hydrogen bonding and van der Waals forces[5, 6], thus, an accurate and efficient computational model for non-covalent interactions is needed to model such processes. Additionally, the two strands in the double helix of DNA are held together by hydrogen-bonding[7, 8], so accurate modeling of non-bonding interactions is critical to the theoretical description of many life processes. Also of importance is the structure of molecular crystals[9, 10] arising from non-bonding interactions among the molecular monomers, and co-conformational selectivity[1-4, 11-13] in interlocked molecules such as catenanes[14, 15] and rotaxanes[15-17]. All these types of systems are dictated by non-bonding interactions.

Significant breakthroughs in the previously mentioned areas could be expected to result from the application of accurate theoretical/computational modeling to these systems. Such studies will require accurate and computationally efficient modeling techniques. In the toolbox of the theoretical/computational chemist, generally the most efficient techniques are those that fall into the category of molecular mechanics (MM). The MM methods are efficient because they rely upon an empirically parameterized

function of the atomic coordinates to express the energy of the system. Computation of the energy therefore involves only the evaluation of an algebraic expression. A major disadvantage of the MM methods is that they do not yield explicit electronic structure information. Such methods represent a fundamentally different theoretical approach from what will be addressed; therefore, MM approaches will not be considered any further.

In contrast to MM methods, quantum mechanical (QM) techniques yield explicit electronic structure information, but at significant computational expense. Currently, the most popular QM method for modeling molecular systems is density functional theory (DFT). This is due in part to its ability to accurately describe chemical and physical properties for a diversity of systems, often at modest computational expense. A major shortcoming with DFT is the inability of most popular XC-functionals (exchange-correlation functionals) to accurately model long-range van der Waals (dispersion) interactions. Therefore, these methods predict systems like the benzene dimer to be unbound. Currently, an increasingly popular approach to overcome this hurdle is to add an empirical correction to the DFT total energy. Empirically corrected DFT methods for dispersion interactions, coined DFT-D, have become popular due to their success with essentially no added computational expense. These methods have shown dramatic improvements for dispersion bound complexes. Not only are such complexes now predicted to be bound, but excellent agreement with CCSD(T) results have been achieved, the current "gold standard" in computational chemistry. For relatively small systems, DFT-D methods are computationally feasible and should provide quite accurate results, but modeling macromolecular host/guest complexes can be extremely computationally expensive; therefore, alternative methods need to be explored. A detailed

review of DFT-D methods is presented in Chapter 2, where both performance/accuracy and theory are discussed.

Intermediate between MM and first-principles methods are semi-empirical (SE) electronic structure methods. The SE techniques are based on a quantum mechanical description of the electronic structure, but to achieve computational efficiency, rely upon empirical parameterization to estimate the values of certain difficult-to-evaluate integrals. Some of the most widely used semi-empirical techniques are AM1[18], PM3[19], RM1[20], and PM6[21]. These methods are sufficiently computationally efficient for modeling systems composed of hundreds or even thousands of atoms, but typically perform poorly for dispersion and hydrogen-bonding. Like most XC-functionals, these semi-empirical methods are essentially incapable of modeling dispersion bound complexes (e.g. the benzene dimer) because the form of the semi-empirical wavefunction causes electron-correlation to be neglected. Even qualitatively reliable modeling of dispersion-bound macromolecular systems, such as complexes of carbon-nanostructures, is therefore out of the question; SE methods predict such complexes to be unbound. That is, the wrong sign of the interaction is predicted.

Like DFT methods, the accuracy of SE methods in modeling dispersion-bound systems can be dramatically improved by adding an empirical correction term. McNamara and Hillier[22] reported adding an empirical correction term to the AM1 and PM3 methods to incorporate dispersion interactions, but found the overall results to be unsatisfactory. Therefore, to gain further improvements, in particular to improve the accuracy with which hydrogen-bonding is modeled, they re-optimized a large portion of the original AM1 parameters, using a small data set consisting of 22 complexes (see

Appendix A). The resulting method, with *both* re-optimized semi-empirical parameters *and* an empirical correction term, is referred to as AM1-D. (They have also produced an analogous PM3-D.) These empirically-corrected methods show a substantial improvement in accuracy (over the corresponding original SE methods) for predicting intermolecular interaction energies, but at a significant cost. As described in chapter 3, AM1-D is nearly 25-fold less accurate in the prediction of heats of formation than the original AM1 method.

More recently Řezáč and colleagues[23] published an empirically corrected PM6 method for modeling dispersion and hydrogen-bonding (HB) interactions, named PM6-DH. This method incorporates an empirical correction for dispersion interactions and also includes a second correction term for HB interactions. The group identified 8 types of H-bonds and used a different set of 3 parameters for each type, for a total of 24 H-bonding parameters. In their defense it should be noted that they did use a relatively large training set to determine these H-bonding parameters. The major shortcoming of this method, as they acknowledge, is that knowledge of atom connectivity is required. One of the major benefits of QM techniques is that atom connectivity is not required, allowing bond formation/deformation to be modeled. Although, Řezáč and colleagues have obtained good results, for a method to be widely used, atom connectivity information should not be required. In addition to the limitations introduced, input of atom connectivity information is sufficiently burdensome to deter routine use, especially for macromolecular complexes where there may be thousands of atom-atom interactions that must be distinguished. These different empirically corrected SE methods are reviewed in Chapter 2.

To address some of the mentioned shortcomings, an in-house empirically corrected SE method has been developed, termed AM1-FS1. This method has been parameterized to a diverse training set that includes non-equilibrium structures and yields sub-kilocalorie accuracy in the prediction of intermolecular interaction energies. This has been achieved with substantially less parameterization than existing empirically-corrected SE methods and *without modification of the original AM1 parameters*. AM1-FS1 therefore retains the predictive power for thermochemical quantities of the original AM1 Hamiltonian, and does not require atom connectivity information. A detailed analysis and description of the AM1-FS1 method is presented in Chapter 3.

Ultimately, we seek a method that leads to good accuracy in the prediction of structural properties and intermolecular interaction energies for macromolecular complexes, without the sacrifice of some of the basic QM benefits. The performance of AM1-FS1 has been tested on several carbon nanostructure complexes and pseudorotaxanes and is found to produce results in very good agreement with the best first-principles calculations available. These results have prompted further investigations of carbon nanostructure complexes and rotaxanes systems, which are discussed in Chapters 4 and 5 respectively.

Chapter 2: Empirically Corrected DFT and Semi-Empirical Methods for Non-Bonding Interactions

2.1 Introduction

In density functional theory (DFT), the total energy of the ground state is expressed as a functional of the electron probability density (ρ). No knowledge of the molecular wave function is required. This is a statement of the Hohenberg-Kohn theorem; however, the theorem does not tell us how to find ρ without first finding the molecular wave function. This hurdle was overcome by Kohn and Sham who developed a method for determining ρ and from it the ground state energy[24]. The Kohn-Sham approach is the most widely used DFT technique. Based on the Kohn-Sham approach, the ground state energy can be expressed as:

$$E_0 = -\sum_{\alpha} Z_\alpha \int \frac{\rho(\mathbf{r}_1)}{r_{1\alpha}} d\mathbf{r}_1 - \frac{1}{2}\sum_{i=1}^{n} \left\langle \theta_i^{KS}(1) \middle| \nabla_1^2 \middle| \theta_i^{KS}(1) \right\rangle + \frac{1}{2}\iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho] \quad , \quad (2.1)$$

where the terms describe the electron-nuclear attraction, electronic kinetic energy, electron-electron repulsion, and the exchange-correlation energy of the electrons respectively. The kinetic energy term depends on the Kohn-Sham (KS) orbitals (θ_i^{KS}). The Kohn-Sham orbitals can be found by iterative self-consistent solution of the following equation:

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_{\alpha} \frac{Z_\alpha}{r_{1\alpha}} + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + \upsilon_{xc}(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS} \theta_i^{KS}(1) \quad (2.2)$$

where υ_{xc} is the exchange-correlation potential:

$$\upsilon_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \quad . \quad (2.3)$$

This quantity can be found if the $E_{xc}[p]$ is known; however, the exact functional form of $E_{xc}[p]$ is unknown, except for the free electron gas. In principle, DFT would provide the exact ground state energy if the exact exchange-correlation functional was known. In lieu of the exact exchange-correlation functional, numerous approximate forms have been proposed and the best choice is often unclear. The exact functional is a "holy grail" of computational chemistry.

Dispersion interactions depend on electron correlation, but DFT typically neglects long-range dispersion because the exchange-correlation term, V_{xc}, is typically assumed to be a functional of the local electron density $V_{xc}(\rho)$, or of the gradient of the electron density $V_{xc}(\partial\rho/\partial\mathbf{r})$. A consequence of this assumption is that only local contributions to the electron correlation are included. The typical DFT functionals therefore do not model correlation outside the Fermi hole, and thereby neglect long-range dispersion. There have recently been some functionals developed that are capable of modeling dispersion interactions, such as M05[25] and M06[26]. These functionals do not contain an explicit dispersion term; but they have been parameterized to systems governed by dispersion interactions and have shown some success for modeling dispersion bound complexes. Implicit inclusion of dispersion is conceptually very different from correcting the DFT total energy with an empirical dispersion term. This review focuses on DFT (and SE) methods that have been corrected with an empirical dispersion term.

The popular AM1 and PM3 semi-empirical methods are based on Hartree-Fock (HF) theory. HF theory utilizes the approximation that the total electronic wavefunction (Ψ) may be written as a Hartree-product of one-electron wavefunctions:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N) \approx \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\phi_3(\mathbf{r}_3)\cdots\phi_N(\mathbf{r}_N) \quad . \tag{2.4}$$

This approximation simplifies the wavefunction from 3N dimensional space to N wavefunctions in three dimensions, which begins to make solution to the Schrödinger's equation feasible, but it leads to the neglect of electron exchange and correlation. The Hartree-product also does not satisfy the antisymmetry principle (Pauli Exclusion Principle); however, this condition is satisfied by expressing the product in a Slater-determinant. A Slater-determinant insures that the wave function vanishes if two electrons have the same spin and occupy the same space. Ultimately HF theory is based on a single Slater-determinant, thus taking electron-exchange into consideration, but not electron-correlation[27]. As mentioned above, electron correlation is responsible for long-range dispersion interactions, thus the HF Hamiltonian is incapable of accurately modeling van der Waals complexes. Since electron-correlation is neglected in the wavefunction of HF theory, it is neglected in the wavefunction of the AM1 and PM3 methods as well, and as a consequence AM1 and PM3 fail to model dispersion interactions.

While the popular DFT and SE methods neglect long-range dispersion, the functional form of dispersion at long range is known. The functional form can be derived from basic physical principles. Here we outline a derivation presented in Ref. 28. Start by considering two non-polar molecules, *a* and *b*. Suppose each molecule is composed of a negative charge (-Q) that oscillates about a fixed positive charge (+Q) with angular frequency (ω_o) in the z-direction, as shown in Figure 2.1. On average, the two molecules possess no permanent dipole moment, however, at any instant in time (t), they have an instantaneous dipole moment ($\mu_a = Qz_a(t)$). To determine the functional form, first consider the molecules at infinite separation. This allows the two molecules to be

modeled separately. The Schrödinger wave equation for molecule *a* is (an equivalent equation can be written for molecule *b*):

$$\frac{1}{M}\frac{\partial^2\Psi_a}{\partial z_a^2}+\frac{2}{\hbar^2}\left(E_a-\frac{1}{2}kz_a^2\right)\Psi_a=0 \qquad (2.5)$$

where the first term is the kinetic energy and the second is the potential energy of the system (½kz_a^2 is the potential energy of the oscillator). This is just the simple harmonic oscillator problem, thus the eigenvalue equations for molecules *a* and *b* are:

$$E_a=\left(n_a+\frac{1}{2}\right)\hbar\omega_o\ , \qquad E_b=\left(n_b+\frac{1}{2}\right)\hbar\omega_o\ . \qquad (2.6)$$

Therefore, the total energy of the system in the ground state is:

$$E(\infty)=E_a+E_b=\hbar\omega_o\ . \qquad (2.7)$$

Now consider the case when the molecules are separated by some distance *r*, where *r* is much greater than the displacement of the positive and negative charges ($r >> z_a$ and z_b). Now, there will be an interaction between the two molecules, which can be modeled as the interaction between two dipoles, since at any instant in time each molecule has a dipole moment. Thus, the Schrödinger wave equation for this system can be written as follows:

$$\frac{1}{M}\frac{\partial^2\Psi}{\partial z_a^2}+\frac{1}{M}\frac{\partial^2\Psi}{\partial z_b^2}+\frac{2}{\hbar^2}\left(E-\frac{1}{2}kz_a^2-\frac{1}{2}kz_b^2-\frac{2z_az_bQ^2}{4\pi\varepsilon_0 r^3}\right)\Psi=0\ . \qquad (2.8)$$

The last term in the parentheses is the potential energy between the dipoles. If the following transformation is made:

$$Z_1=\frac{z_a+z_b}{\sqrt{2}}\ ,\quad Z_2=\frac{z_a-z_b}{\sqrt{2}}\ , \qquad (2.9)$$

the Schrödinger wave equation can be rewritten as:

$$\frac{1}{M}\frac{\partial^2\Psi}{\partial Z_1^2}+\frac{1}{M}\frac{\partial^2\Psi}{\partial Z_2^2}+\frac{2}{\hbar^2}\left(E-\frac{1}{2}k_1Z_1^2-\frac{1}{2}k_2Z_2^2\right)\Psi=0 \quad . \tag{2.10}$$

This transformation reduces the problem to two independent harmonic oscillators, thus the eigenvalue equation is the sum of the two independent oscillators:

$$E(r)=\left(n_1+\frac{1}{2}\right)\hbar\omega_1+\left(n_2+\frac{1}{2}\right)\hbar\omega_2 \quad . \tag{2.11}$$

The ground state energy is:

$$E(r)=\frac{1}{2}\hbar(\omega_1+\omega_2) \quad , \tag{2.12}$$

where:

$$\omega_1=\omega_0\left\{1-\frac{2Q^2}{4\pi\varepsilon_0 r^3 k}\right\}^{\frac{1}{2}} \quad , \quad \omega_2=\omega_0\left\{1+\frac{2Q^2}{4\pi\varepsilon_0 r^3 k}\right\}^{\frac{1}{2}} \quad . \tag{2.13}$$

If we substitute these frequencies into Eq. 2.12 and expand the solution by the binomial theorem, the energy of the ground can be expressed as follows:

$$E(r)=\hbar\omega_0-\frac{Q^4\hbar\omega_0}{2(4\pi\varepsilon_0)^2 r^6 k^2}+.... \quad . \tag{2.14}$$

Thus, for this simple model the interaction energy can be expressed as follows:

$$U_{disp}=E(r)-E(\infty)=-\frac{Q^4\hbar\omega_0}{2(4\pi\varepsilon_0)^2 r^6 k^2}+.... \quad , \tag{2.15}$$

where the leading term in the binomial expansion, which is the dipole-dipole dispersion energy, has $1/r^6$ dependency. This proves that long-range dispersion interactions are dominated by the $1/r^6$ term, which is the justification behind the popular Lennard-Jones potential, which will be briefly discussed in the next section.

Given that the popular DFT and SE methods neglect dispersion, but given also that the correct functional form of the dispersion interaction at long range is known, there has been much interest in empirically correcting DFT and SE methods to include dispersion. Herein we review recent advances in this area.

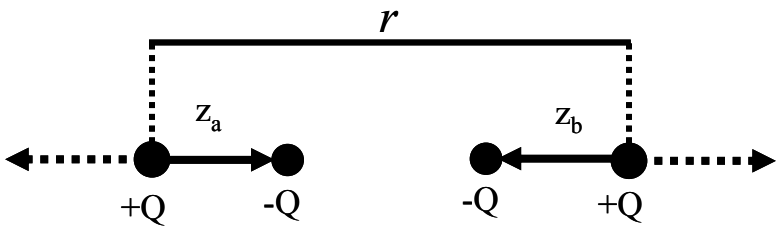


Figure 2.1: A schematic of two non-polar molecules, separated by some distance *r*. Each molecule is composed of a single positive (+Q) and negative (-Q) charge displaced by some distance z_a and z_b. This simple model may be used to derive the functional form of long-range dispersion interactions.

2.2 Empirical Dispersion Potentials

As noted above, dispersion interactions are solely of quantum mechanical origin[28], but they are neglected for the most part, if not completely, in commonly employed quantum chemical techniques (SE, HF, and DFT). One approach to incorporating dispersion interactions is to add an empirical potential to the quantum mechanical total energy. Empirical potentials are the basis of molecular mechanics and are computationally extremely efficient in comparison to QM methods. Therefore, an empirical potential can be added to a QM method without incurring any appreciable additional computational expense. Because of the promise of increased accuracy without

additional computational expense, this approach has recently been receiving considerable attention for modeling dispersion-bonded complexes.

A popular empirical potential for describing interactions between neutral atoms or molecules is the 6-12 Lennard-Jones (LJ) potential. The LJ potential attempts to describe both attractive and repulsive interactions. The attractive portion of the potential models the instantaneous dipole-dipole interactions, in other words, the dispersion interactions. The functional form of the LJ potential is:

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{2.16}$$

where ε is the depth of the well, σ is the intermolecular separation at which the potential equals zero, and r is the separation of the two particles. The two parameters ε and σ are specific to the two interacting particles. For two interacting atoms these quantities are usually combined:

$$C_{12} = \varepsilon\sigma^{12} \quad , \qquad C_6 = \varepsilon\sigma^{6} \quad , \tag{2.17}$$

and the LJ potential is then expressed in the general form:

$$V(r) = \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6}, \tag{2.18}$$

where C_{12} and C_6 are constants that depend on the atoms being considered; the C_6 constant is commonly referred to as the dispersion coefficient. In molecular mechanics these values are determined by parameterization; however, C_6 values can be determined with the knowledge of the polarizability of the atoms and their ionization potentials. To determine the dispersion interaction of a molecular system, the term is summed over the unique atom pairs. (Typically bonded pairs, i.e. “1-2 interactions” and atoms bonded to a

common atom, i.e."1-3 interactions" are removed. In some cases atoms sharing a common dihedral angle, i.e. "1-4 interactions" are also excluded from the sum.) The LJ function is composed of two terms: a repulsion term and an attractive term. The negative r^{-6} term describes the long range dispersion interactions; the origin of this functional form was derived above. The positive r^{-12} term attempts to describe the short range Pauli repulsive interactions; however, this term has no fundamental theoretical basis and is chosen for computational convenience. (r^{-12} is easily obtained by squaring r^{-6}.)[29] Figure 2.2 graphically shows the functional form of the LJ potential (solid line) for the interaction between two carbon atoms (the C_6 and C_{12} values for two interacting carbon atoms were obtained from AutoDock Version 1[30]). This figure also shows the behavior of the attraction dispersion term (dotted line) and repulsion term (dashed line). At large values of r, the attractive term dominates since the repulsion term approaches zero much more quickly with increasing r. The reverse is true at close distances; therefore, at close distances the LJ potential has a steep potential wall due to the (r^{-12}) term. The dispersion interaction between two molecules can be approximated by summing over all unique inter-component atom-atom pairs.

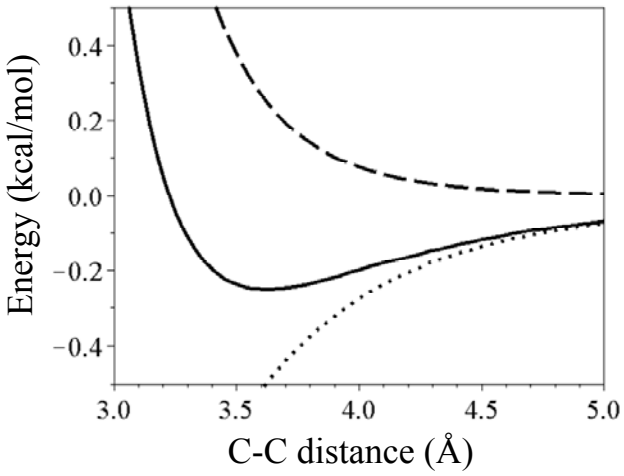


Figure 2.2: Graphical representation of the LJ potential (solid line) and the corresponding attractive (dotted line) and repulsive (dashed line) components. The curves correspond to two interacting carbon atoms (the C6 and C12 values for two interacting carbon atoms were obtained from the code, AutoDock Version 1[30]).

The LJ potential is a relatively reliable model for dispersion interactions, provided good dispersion coefficients are available, however, other variants are also in use. A variation proposed by Wu and Yang[31], has been receiving considerable attention for use in empirically correcting DFT methods. This function is of the form:

$$E_{dis} = -\sum_{i<j} \frac{C_6^{ij}}{r_{ij}^6} f_{damp}\left(r_{ij}\right) \tag{2.19}$$

where C_6^{ij} is the dispersion coefficient, f_{damp} is a damping function and the sum is over all unique atom pairs. The damping function attenuates the r^{-6} long-range dispersion interaction at short range and is expressed as follows:

$$f_{damp}\left(r_{ij}\right) = \frac{1}{1+e^{-d\left(r_{ij}/R_{vdw}-1\right)}} \tag{2.20}$$

where d is the damping coefficient and R_{vdw} is the equilibrium van der Waals separation for the atom-pair being considered. A graphical representation of this function for the interaction between two carbon atoms is shown in Figure 2.3; the figure also graphically shows the attractive portion of the function and the damping function.

While the LJ potential "blows-up" at short-range (small values of r) due to the r^{-12} term, Eq. 2.19 goes to zero at short range due to the damping function. The reason for this difference is that DFT and SE methods already model short-range repulsive (Pauli repulsive) interactions. The empirical correction function is therefore made to go to zero to avoid double counting the repulsive interactions. Energy decomposition analysis (EDA) and symmetry-adapted perturbation theory (SAPT) can be used with HF, DFT, or higher levels of theory to separate out the individual energy contributions such as: electrostatic, exchange repulsion, polarization and dispersion energy. This type of analysis can help identify the functional form of the various components that contribute to the total energy. This can help identify potential improvements that can/need to be made and is also very useful for the development of MM methods[32, 33].

Employing a damping function also allows intra-component dispersion interactions to be easily considered in addition to inter-component interaction even without knowledge of atom connectivity. In protein modeling for example, the inclusion of intra-component dispersion interactions is very important[34]. Due to the large size of proteins, intra-molecular dispersion interactions play a major role in determining their structure. Including both inter-molecular and intra-molecular terms in the dispersion correction is also beneficial from a computational standpoint. Including both inter-molecular and intra-molecular terms allows the total dispersion energy to be calculated

by summing over all the unique atom-atom pairs so that the number of atoms in each monomer does not need to be specified. While keeping track of the number of atoms in each monomer is not much of a problem for a dimer system, as the number of monomers increases, such bookkeeping becomes a cumbersome programming task.

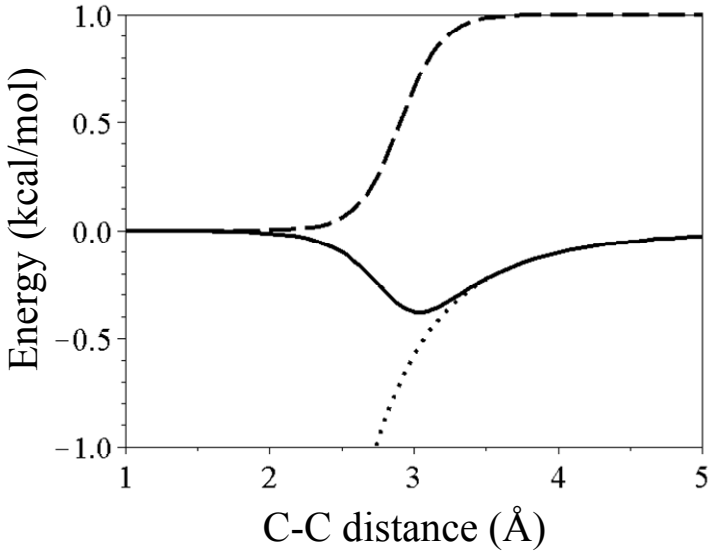


Figure 2.3: A graphical representation of the potential (solid line) proposed by Wu and Yang[31]. The attractive (dotted line) R^{-6} term and damping function (dashed line) are included. The curves correspond to two interacting carbon atoms. The dispersion coefficient (C_6^{ij}) was obtained from Ref. 35 and the damping coefficient (d) was set to 20.

Subsequent to the introduction of the functional form (Eq. 2.19) by Wu and Yang[31], Grimme[34] proposed to use of a global scaling factor (S_6), an adjustable parameter multiplying Eq. 2.19. The value of this parameter is obtained by a fitting procedure. Both the S_6 and d parameters can be adjusted depending on the method or DFT functional used. These parameters are unitless and typically have values close to 1.0 and 20

respectively. Grimme proposed and used a value of 20 for the damping constant and adjusted the global scaling factor depending on the DFT functional used (B97, S_6 = 1.25; BLYP, S_6 = 1.2; PBE, S_6 = 0.75)[35]. The effect of adjusting these two parameters is shown in Figure 2.4. Figure 2.4A shows the effect of changing the damping coefficient (d): d = 15 (dotted line), 20 (solid line), and 25 (dashed line); S_6 = 1.0 in all cases. As the damping coefficient increases, the function is more abruptly shut off at low r. The depth of the potential well also increases as d increases, however, the depth is more strongly influenced by the global scaling factor (Figure 2.4B: S_6 = 0.8 (dotted line), 1.0 (solid line), and 1.2 (dashed line); d = 20 in all cases). These parameters can be tailored for different methods. As in MM methods, fitting the parameters requires a training set; therefore, the values obtained may or may not be the best choice for a system outside the training set, especially if it is chemically significantly different from the species in the training set.

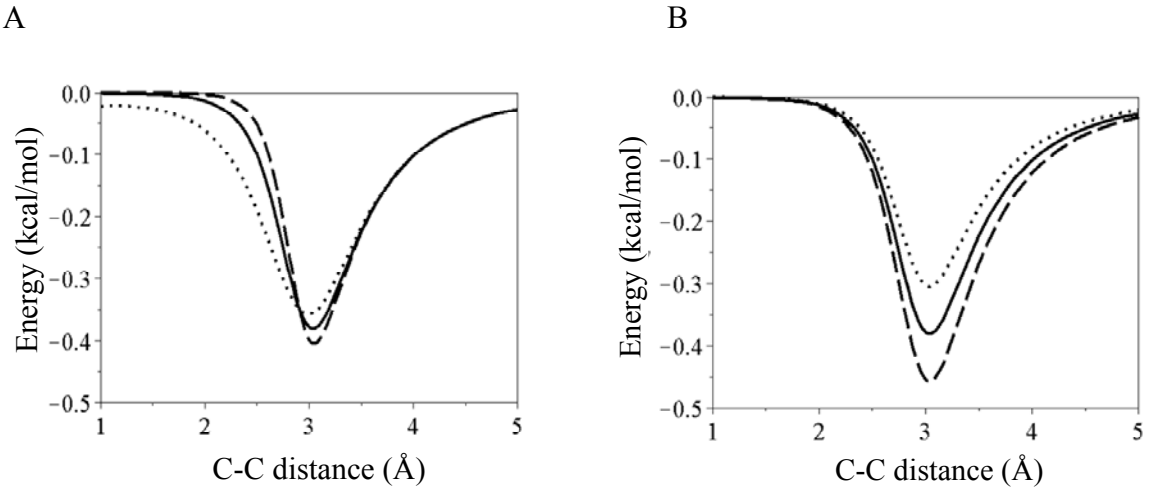


Figure 2.4: Figure 2.4A shows the effect of changing the damping coefficient (*d*) in Eq. 2.18: d = 15 (dotted line), 20 (solid line), and 25 (dashed line); S_6 = 1.0 in all cases. Figure 2.4B shows the effect of change the global scaling factor: S_6 = 0.8 (dotted line), 1.0 (solid line), and 1.2 (dashed line); *d* = 20 in all cases. The curves correspond to two interacting carbon atoms; the dispersion coefficient (C_6^{ij}) was obtained from Ref. 35.

Other functions have been proposed for modeling dispersion interactions, but, Eq. 2.19 appears to be the most widely used and has shown considerable success when used as a correction term for DFT techniques. Jurečka and colleagues[36] suggest a slight variation to the function proposed by Grimme. They recommend that the scaling factor (S_6) be moved inside of the damping function, yielding,

$$f_{damp}\left(r_{ij}\right) = \frac{1}{1 + e^{-d\left(\frac{r_{ij}}{S_R R_{vdw}} - 1\right)}} \quad . \tag{2.21}$$

The parameter, now termed S_R, scales the equilibrium van der Waals separation. A graphical representation of the dispersion function (Eq. 2.19) utilizing this damping function is shown in Figure 2.5, where it can be seen that scaling the sum of the van der Waals radii allows the short-to-medium range interactions to be adjusted while leaving

the long-range r^{-6} behavior unchanged. This differs from the effect of the global scaling (S_6) used by Grimme, which affects the function over all r, as shown in Figure 2.4B. Moving the scaling factor inside the damping function is therefore physically motivated, since only the short-range dispersion interactions should differ among different XC-functionals[36]. The S_R term attempts to correct for the deviation from r^{-6} behavior at short range and the parameter therefore needs to be optimized for the functional of choice. Jurečka and colleagues state that this term "allows for correction of the inaccuracy (or, better, fitness) of the absolute values of the vdW radii, and thus only relative values of the radii need to be correct[36]." The same could be said about the S_6 factor relating to the dispersion coefficient (C_6^{ij}). Moreover, the best position of the parameter might depend of the computational method being used.

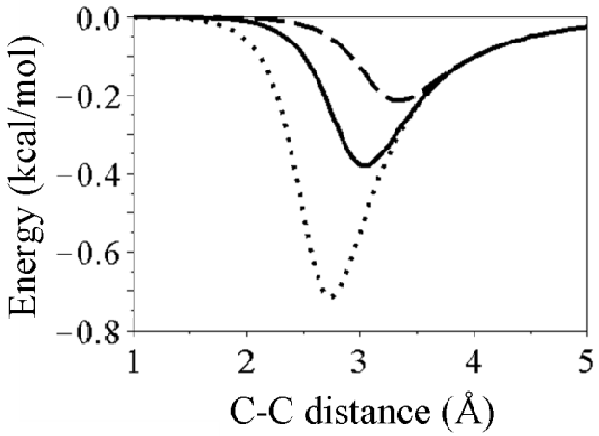


Figure 2.5: Graphical representation of the modified potential (solid line) suggested by Jurečka and colleagues[36]. This figure shows the effect of changing S_R in the damping function (Eq. 2.20): S_R = 0.9 (dotted line), 1.0 (solid line), and 1.1 (dashed line). The figure serves as a graphical representation of the effect of changing the value of the S_R term. The curves correspond to two interacting carbon atoms; the dispersion coefficient (C_6^{ij}) was obtained from Ref. 35.

It is worth briefly discussing the different combination rules being used to determine the pair-wise dispersion coefficient. Grimme originally[34] used a harmonic average of the form:

$$C_6^{ij} = 2\frac{C_6^i C_6^j}{C_6^i + C_6^j} \quad , \tag{2.22}$$

but later[35] switched to the geometric mean combination rule:

$$C_6^{ij} = \sqrt{C_6^i C_6^j} \quad . \tag{2.23}$$

Grimme clams that the geometric mean "yields much better results[35]." Jurečka and colleagues[36] are using the combination rule suggested by Wu and Yang[31], which is of the form:

$$C_6^{ij} = \frac{2\left((C_6^i C_6^j)^2 N_{eff\_i} N_{eff\_j}\right)^{1/3}}{\left(C_6^i N_{eff\_j}^2\right)^{1/3} + \left(C_6^j N_{eff\_i}^2\right)^{1/3}} \quad , \tag{2.24}$$

where N_{eff} is the effective number of electrons. This combination rule is based on the Slater-Kirkwood formula[37]. Different combination rules have also been used for the equilibrium van der Waals separation (R_{vdw}) term. Grimme uses the arithmetic mean, thus refers to this term as the sum of the atomic van der Waals radii. Jurečka and colleagues[36] used the cubic mean:

$$R_{vdw} = \frac{R_{ii}^3 + R_{jj}^3}{R_{ii}^2 + R_{jj}^2} \tag{2.25}$$

where R_{ii} and R_{jj} are the equilibrium van der Waals radii of two identically atoms (van der Waals diameter). Jurečka and colleagues stated that, "the cubic mean yielded lower errors," thus they have employed this combination rule. This combination rule was suggested by Halgren[38], and can be viewed as a weighted average. The cubic mean

increases the equilibrium van der Waals separation, relative to the arithmetic mean, as the difference between R_{ii} and R_{jj} become larger. In the next section, the accuracy of empirically corrected DFT and SE methods will be considered. These methods are referred to as DFT-D and SE-D, where -D signifies dispersion corrected.

2.3. Empirically Corrected DFT and SE Methods

2.3.1 DFT-D Methods

Empirically corrected DFT methods for modeling dispersion interactions are far more popular than empirically-corrected semi-empirical methods. This is mainly due to the general greater accuracy of DFT methods. Although most DFT methods include some electron correlation, as discussed in the previous section, long-range dispersion interactions are almost universally neglected by the popular DFT functionals[34]. An empirical correction for long-range dispersion interactions is therefore a desirable enhancement to DFT methods.

Grimme and colleagues[34, 35, 39-41] have published numerous articles devoted to developing and studying DFT-D methods, with the first in 2004[34]. In 2006, Grimme[35] published a revised method and provided dispersion coefficients (C_6) and van der Waals radii (R_{vdw}) for the elements H-Xe. These parameters were derived from high level computations. (The computational details will not be discussed here. For details the reader should consult Ref. 35.) Grimme's research has increased the applicability of DFT-D methods, allowing for molecules containing a wide variety of elements to be modeled. In addition to Grimme, numerous other groups are now pursuing this line of research[16, 22, 42-44].

Grimme and colleagues have developed empirical corrections for a variety of DFT functionals[35, 39]. The accuracy of the different empirically corrected DFT methods was tested on a database containing 22 complexes with accurately known structures and interaction energies (commonly referred to as the "S22" database, see Appendix A). The S22 database, provided by Jurecka et al.[45], is composed of 7 hydrogen bonded, 8 dispersion-bonded, and 7 mixed-bonded complexes. The binding energies reported are CCSD(T) quality and are extrapolated to the complete basis set (CBS) limit (see Table B-1 for names and binding energies of the different complexes in the S22 database). The geometries are mainly MP2 quality, with a few smaller complexes optimized at the CCSD(T) level. This database has been used by others as a standard; therefore, it will be used herein when comparing methods whenever possible.

Of the different XC-functionals considered by Grimme and colleagues in 2006[39], the dispersion corrected B97 functional produced the best results and the BLYP functional was a close second. The reported binding energies for the species in the S22 database as computed using these two different methods are reported in Table B-1. The root mean square errors (RMSE) and mean unsigned error (MUE) are reported in Table 2.1. The total RMSE values are 0.46 and 0.58 kcal/mol for B97-D and BLYP-D respectively. When the RMSEs are partitioned, it is found that the hydrogen bonded complexes have the greatest error in both cases (0.60 and 0.83 kcal/mol respectively). The larger error associated with the hydrogen bonded complexes may be due to the fact that their calculations are not counterpoise (CP) corrected. This is discussed further below.

More recently, Schwabe and Grimme[46] empirically corrected the B2PLYP XC-functional (a semiempirical hybrid functional with corrections from perturbation theory) and achieved even greater accuracy, as evaluated based on the S22 database. The statistical results are shown in Table 2.1 (detailed results for the individual complexes are shown in Table B-1). It should be noted that in their study the interaction energy was defined in a non-standard way as,

$$\Delta E = \tfrac{1}{2}\Delta E_{CP} + \tfrac{1}{2}\Delta E_{noCP}, \tag{2.26}$$

where CP indicates the energy has been counterpoise corrected for basis set superposition error (BSSE). In the group's earlier studies, the interaction energies were not CP corrected, because it was found that uncorrected results yielded lower errors, assuming a basis set of a least triple-zeta quality is used[39]. Using the B2PLYP-D method, the total RMSE for the S22 database reduced to 0.39 kcal/mol, with most of the improvement coming from the mixed complexes. The overall improvement is likely due to the B2PLYP functional, since this functional "seems to outperform all current hybrid GGAs and meta-GGAs[47]" functionals. It should be noted that a smaller global scaling factor was used (S_6 = 0.55) in the dispersion correction term (typical values are close to 1). This was needed due to the perturbation term in the B2PLYP functional recovering part of the dispersion effects. The B2PLYP functional underestimates long-range effects principally because the MP2 perturbation component recovers only *part* of the correlation that is otherwise neglected by the semi-local GGA component[46]. This is why a dispersion correction term is needed and why the optimal global scaling factor is small. Since the B2PLYP functional incorporates some long-range dispersion interactions, owing to the

perturbation correction, it would not be desirable to scale R_{vdw} (use S_R). Scaling R_{vdw} would result in double counting dispersion interactions in the long-range regime.

Table 2.1: Single-point interaction energy statistics (kcal/mol) for the S22 database[45]. The B97/TVZ(2df,2pd) and BLYP/TVZ(2df,2pd) interactions energies used to construct this table were obtained from Ref. 39 and the B2PLYP/TVZP values from Ref. 46.

	B97-D / TZV(2df,2pd)	BLYP-D / TZV(2df,2pd)	B2PLYP-D/TZVPP
RMSE (Hydrogen bonded)	0.60	0.83	0.58
RMSE (Dispersion bonded)	0.22	0.40	0.24
RMSE (Mixed bonded)	0.48	0.43	0.29
RMSE	0.46	0.58	0.39
MUE	0.35	0.47	0.31

As discussed earlier, Jurečka and colleagues[36] proposed that the scaling factor be moved inside the damping function (Eq. 2.21). It is worth noting that they used the dispersion coefficients reported by Grimme but alternative van der Waals radii. They also used different combination rules for determining C_6^{ij} (Eq. 2.24) and R_{vdw} (Eq. 2.25). Although no direct comparisons were made, Jurečka and colleagues[36] claim to have achieved more accurate results, thereby justifying these changes. The group studied the accuracy of numerous DFT functionals with and without an empirical correction for dispersion interactions. They also considered a variety of basis sets and the effect of BSSE. In each case considered, optimal parameters were used (d and S_R). The best results were obtained with the TPSS functional using the 6-311++G(3df,3pd) basis set; therefore, we will limit our discussion to this functional and basis set. The article by

Jurečka and colleagues[36] is also useful for comparing the performance of various uncorrected DFT methods.

Jurečka et al.[36] used the S22 database to test the performance of the various different methods, allowing comparisons to the works of Grimme and others. The uncorrected DFT results reported clearly demonstrate why it is desirable to include an empirical correction for dispersion interactions. The statistical results for the empirically corrected and uncorrected TPSS functional are shown in Table 2.2 (detailed results for the individual complexes are shown in Table B-2). The empirical dispersion term lowers the RMSE from 4.17 kcal/mol to 0.40 kcal/mol, an improvement by more than a factor of 10 (a graphical representation showing the individual deviation from the reference values is shown in Figure 2.6). Comparable improvement is observed for all the XC-functionals considered in the study; clearly showing the benefit of an empirical dispersion term. Improvements are not only seen in the dispersion bound cases, but also for hydrogen bonded cases. The RMSE for the hydrogen bonded cases decreased from 1.75 to 0.59 kcal/mol. This presumably occurs because dispersion interactions still play an important role for systems dominated by hydrogen bonding. Table 2.2 also shows the effect of the CP-correction. In agreement with the findings of Grimme, the CP-corrected results increase the error. This is likely due to the fact that the BSSE tracks the dispersion interaction energy[16], so that in the absence of the CP correction, the BSSE artificially "recovers" a portion of the dispersion interaction. In addition, the dispersion parameters are optimized with non CP-corrected results, and since the parameters are basis set depended, it is not surprising the better results are obtained without CP-correcting.

Table 2.2: Single-point interaction energy statistics (kcal/mol) for the S22 database[45]. LP = 6-311++G(3df,3pd); CP indicates that the results have been counterpoise corrected. The interactions energies used to construct this table were obtained from Ref. 36.

	TPSS / LP	TPSS / LP CP	TPSS-D / LP	TPSS-D / LP CP
RMSE (Hydrogen bonded)	1.7	2.2	0.6	0.4
RMSE (Dispersion bonded)	6.3	7.2	0.3	0.6
RMSE (Mixed bonded)	2.5	3.0	0.2	0.4
RMSE	4.2	4.8	0.4	0.5
MUE	3.0	3.7	0.3	0.4

Unfortunately, a direct comparison between the methods of Jurečka and Grimme could not be made because the same XC-functional and basis set were not used in the different works reported in the literature. Currently the best results of Jurečka and Grimme are virtually identical, both having a mean unsigned error (MUE) of 0.3 kcal/mol and a RMSE of 0.4 kcal/mol. From this information alone it is hard to ascertain which method is better; however, avoiding the CP-correction, which was used for Grimme's B2PLYP-D method, significantly limits the computational expense. The two methods (TPSS-D and B2PLYP-D) are graphically compared in Figure 2.6, where the deviation from the reference interaction energy is plotted against the S22 complex number. It can be seen that both these methods show good agreement with the CCSD(T) reference values; however, it should be noted that both these methods have used the S22 database for parameterizing the dispersion correction term. Therefore, it is not too surprising to find very good correlation. It can be stated that both these methods perform very well and represent a drastic improvement over the uncorrected DFT methods, which is also shown in Figure 2.6 for the TPSS functional.

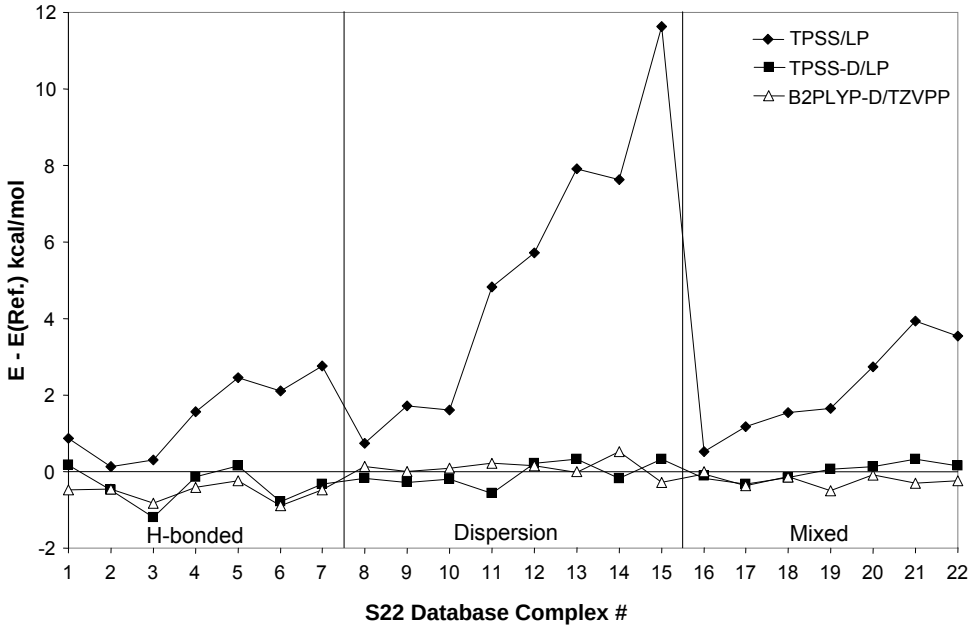


Figure 2.6: A graphical representation showing the interaction energy (kcal/mol) deviation from the S22 CCSD(T) reference values[45]. All DFT calculations are preformed at the reference geometries. Note a positive deviation indicates that the complex is under bound and if negative it is over bound. LP = 6-311++G(3df,3pd). Here and in subsequent analogous figures the line segments connecting adjacent data points are intended as a visual aid. The TPSS and TPSS-D interactions energies used to construct this figure were obtained from Ref. 36 and the B2PLYP values from Ref. 46.

Comparing the energies from DFT-D calculations to CCSD(T) energies at fixed reference geometries is not a completely realistic estimator of accuracy. First of all, the DFT-D energy is of little practical value if the CCSD(T) energy is already known. Secondly, it is possible for a method to predict a very accurate energy at a specific molecular geometry, yet yield a very inaccurate picture of the remainder of the potential energy surface, as shown schematically in Figure 2.7. Therefore, it is very important to consider the effect of structural optimization with DFT-D methods. Often, the geometry of the system(s) is the property of interest. We will now look at the effect on the energies and *geometries* when the complexes in the S22 database are optimized with different

DFT-D methods. DFT-D optimized energies and/or structures have not always been made available, (or even computed) therefore, limited comparisons are possible. Grimme and colleagues carried out DFT-D optimization work, but their optimized energies and structures for the species in the S22 database have apparently not been reported in the literature. Fortunately, another group that has adopted Grimme's method, (Morgado and colleagues[48]) has preformed geometry optimizations at the BLYP-D/TZV(2d,2p) level of theory and provided optimized structures as supplementary material. The method used was identical to the one used by Grimme and colleagues for the data reported in Table 2.1, except for the size of the basis set (TZV(2df,2pd) vs. TZV(2d,2p)). Morgado and colleagues[48] reported only optimized energies and not single-point energies for the S22 complexes, again preventing a direct comparison. The change in the RMSE upon optimization cannot be directly evaluated; however, it can be estimated based on Grimme's single-point BLYP-D/TZV(2df,2pd) results (Table 2.1).

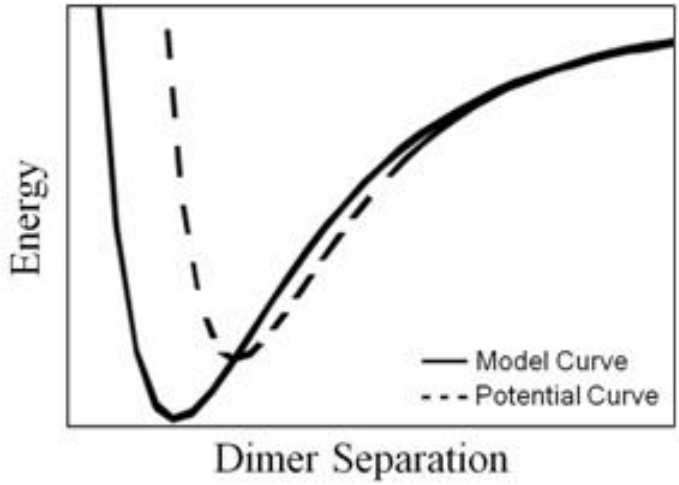


Figure 2.7: A schematic showing how it is possible for a method to predict a very accurate energy at a specific molecular geometry, yet yield a very inaccurate picture of the entire potential energy surface. Note that the model curve (solid-line) yields exactly the correct energy at the minimum of the potential (dashed-line), yet upon structural optimization based on the model would yield a wildly incorrect value for the equilibrium separation.

Statistical data evaluating the performance of the BLYP-D/TZV(2d,2p) method upon optimization for the species in the S22 database are reported in Table 2.3. The RMSE and MUE for the optimized energies and interaction distances are shown, along with the RMSE for the different categories of complexes (detailed results for the individual complexes are shown in Table B-3). As should be expected, the RMSE value for the energy is larger for the optimized geometries (0.90 kcal/mol), than the single-point energies (0.46 kcal/mol). (Note that this result is not a direct comparison due to the difference in basis sets as discussed above.) The structural distortion upon optimization is compared in two different ways: i) the interaction distance (see Figure S1 of Ref. 22) and ii) the center-of-mass distance (CM-distance) between the two monomers. The two different comparisons are used, because limited data is available in the literature. (Some

of the interaction distances are defined as the CM-distances, see Table 2.3.) The RMSE of the CM-distances was found to be 0.095 Å, whereas for the interaction distances it was found to be 0.183 Å for the BLYP-D/TZV(2d,2p) method. This indicates that the interaction distance is a more sensitive measurement of structural distortion upon optimization. The BLYP-D method produces very accurate interaction distances for the hydrogen bonded complexes (RMSE = 0.035 Å), but the RMSE for the interaction energies exceeds 1.0 kcal/mol. This is the highest RMSE for any category of complexes, but the *relative* error is the smallest of any category. The error associated with the interaction energy for the hydrogen bonded complexes would likely be even lower if the energies were CP-corrected, as this was the case for the TPSS functional (see Table 2.2). The relative error associated with the dispersion bonded complexes is the largest; however, the most drastic improvement is seen for these complexes; they are now predicted to be bound.

Table 2.3: A statistical comparison of the geometry optimized energies (kcal/mol), interaction distances (Angstroms), and CM-distance (Angstroms) for the complexes in the S22 database[45]. The interaction distance is defined as the CM-distance for complexes 11-15 and the 2nd distance reported for complex 22. All data used to construct this table was obtained from Ref. 48.

	BLYP-D / TZV(2d,2p)		
	Interaction Energies (kcal/mol)	Interaction Distances (Å)	CM-Distances (Å)
RMSE (Hydrogen bonded)	1.32	0.035	0.047
RMSE (Dispersion bonded)	0.76	0.144	0.104
RMSE (Mixed bonded)	0.40	0.304	0.117
RMSE	0.90	0.192	0.095
MUE	0.72	0.097	0.071

Above we have compared binding energies and interaction distances, determined with the BLYP-D method, to the benchmark S22 structures, and have seen that this method is relatively successful. It is also useful to consider the benefit that is gained by using a DFT-D method over a traditional (uncorrected) DFT method. Jurečka and colleagues[36] have made these data available; reporting optimized binding energies and CM-distances between the two monomers, using the TPSS functional both with and without dispersion correction. In addition, the TPSS-D optimized structures were graciously provided by the authors upon our request, allowing the interaction distances to be calculated for comparison. The statistical data for the optimized interaction energies, center-of-mass (CM) and interaction distances are reported in Table 2.4 (detailed results for the individual complexes are shown in Table B-4). The RMSE for the binding energies of the S22 complexes decreases from 2.7 to 1.1 kcal/mol upon inclusion of the empirical dispersion correction to the TSPP results. A more drastic improvement is observed for the CM-distances, the RMSE decreased from 0.524 to 0.062 Å. This clearly shows the benefit of using an empirical dispersion term with traditional DFT methods.

In Table 2.2 (single-point TPSS and TPSS-D analysis), it is shown that by applying the dispersion correction to the TPSS functional, the overall interaction energy error decreases for the hydrogen bonded complexes; however, this is not true upon geometry optimization. For TPSS-D optimized hydrogen bonded structures, the RMSE increases for the interaction energies from 1.3 to 1.8 kcal/mol. The error associated with the CM-distances also increases from 0.023 to 0.051 Å. These results indicate that it is better to use the uncorrected TSPP functional, if the dominant interaction is hydrogen bonding; (a conclusion that does not necessarily extend to other functionals). This is a

non-intuitive result; including theory otherwise neglected gives a worse result? This is just a consequence of BSSE, which is more prevalent in hydrogen bonded complexes due to their short inter-molecular distances. As observed in Table 2.2, CP-correcting the TPSS result causes the RMSE to decrease from 0.6 to 0.4 kcal/mol. Thus, if the calculations were performed at the complete basis set limit, it would likely be found that the TPSS-D method outperforms TPSS method for hydrogen bonded complexes upon geometry optimizations. It should be noted that a new set of dispersion parameters would be needed since these parameters are basis set depended[36].

Table 2.4: A statistical comparison of the geometry optimized interaction energies (kcal/mol) and CM-distances (Angstroms) for the complexes in the S22 database. LP = 6-311++G(3df,3pd). The interaction distance is defined as the CM-distance in complexes 11-15 and 2nd value reported for complex 22. The TPSS/LP interactions energies and CM-distances used to construct this table were obtained from Ref. 36. The interaction distances were computed from the optimized structures obtained from the authors[36].

	TPSS / LP		TPSS-D / LP		
	ΔE (kcal/mol)	CM-Distance (Å)	ΔE (kcal/mol)	CM-Distance (Å)	Interaction Distances (Å)
RMSE (Hydrogen bonded)	1.3	0.023	1.8	0.051	0.070
RMSE (Dispersion bonded)	3.9	0.812	0.3	0.039	0.040
RMSE (Mixed bonded)	2.0	0.331	0.2	0.089	0.084
RMSE	2.7	0.524	1.1	0.062	0.067
MUE	2.1	0.342	0.6	0.036	0.045

When the interaction energies for the structures in the S22 database as computed with the BLYP-D and TSPP-D methods are compared, it is found that the BLYP-D method performs slightly better. The total RMSEs for the S22 database are 0.90 and 1.1 kcal/mol respectively. On the other hand, if the interaction distances or CM-distances are

compared, the TSPP-D method outperforms BLYP-D. The interaction distance RMSE for the BLYP-D method is more than twice that of the TSPP-D method (see Table 2.3 and Table 2.4). Graphical comparisons are presented in Figure 2.8, where the deviation from the reference interaction energies (Figure 2.8A) and distances (Figure 2.8B) are plotted against the S22 complex number. Here it can be clearly seen that the TPSS-D method is superior for producing interaction energies upon optimization in almost all cases of dispersion and mixed complexes. What is really remarkable about the TSPP-D method, however, is that the RMSE in the energy does not change upon optimization for the dispersion and mixed bound complexes. This is not achieved by any other method considered. The TSPP-D method also performs considerably better then BLYP-D for interaction distances, which can be seen in Figure 2.8B and is backed by the statistical data in Table 2.3 and Table 2.4. The TSPP-D method, however does not outperform BLYP-D upon optimization for the hydrogen bonded complexes based on both the interaction energies and distances. In fact, the uncorrected TSPP method outperforms both corrected methods. (See Figure 2.8 and/or Table 2.3 & Table 2.4.) Again, CP-correcting the TSPP-D method should improve the error for the hydrogen bonded complexes. This is not the case for the BLYP-D method due to the nature of the CP-correction; which decreases the binding. As shown in Figure 2.8A for the BLYP-D method, most of the deviations are in the positive direction indicating the complexes are under bound. Thus, very careful consideration should be taken when picking a DFT-D method depending on the dominant interaction(s) in the system.

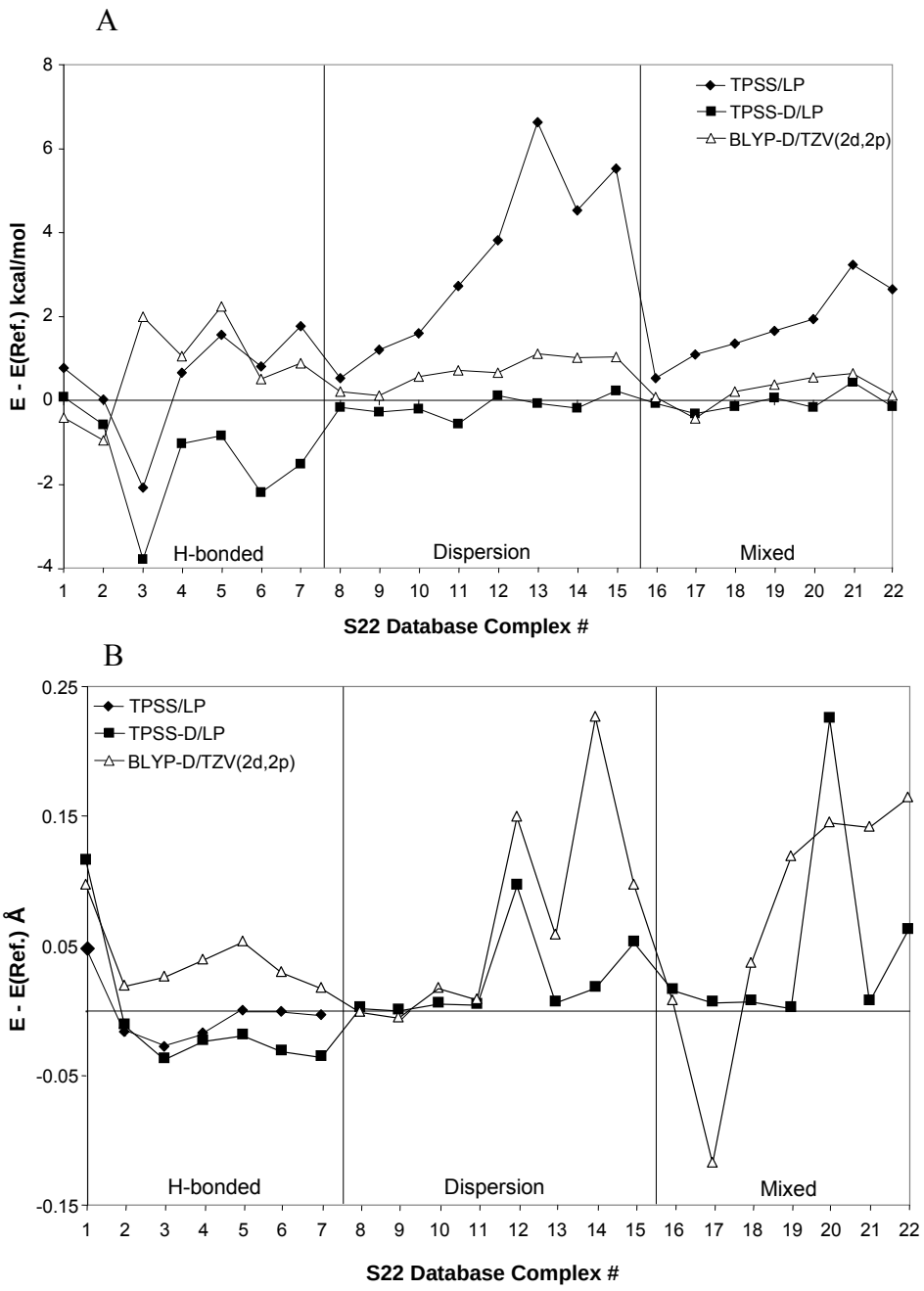


Figure 2.8: Graphical representation showing deviations of the optimized interaction energies (kcal/mol) (Figure A) and CM-distances (Å) (Figure B) from the S22 CCSD(T) reference values[45]. The TPSS values for the interaction distances are only shown for the hydrogen bonded complexes. This was done for clarity and axis scaling proposes, many of the other points are far off scale, see Table B-4. Note that a positive deviation indicates that the complex is under bound and if it is negative the complex is over bound. LP = 6-311++G(3df,3pd). The TPSS and TPSS-D data used to construct these figures were obtained from Ref. 36 and the BLYP-D data from Ref. 48.

Antony and Grimme[39] also subjected the B97-D and BLYP-D methods to testing with a larger database (*JSCH-2005*) containing DNA base pairs and amino acid pairs. The database is MP2 and CCSD(T) quality. (For further detail see Ref. 45). The *JSCH-2005* database plus the S22 database, consists of 161 complexes. The RMSE for the combined database are 0.92 and 0.85 kcal/mol for B97-D and BLYP-D respectively, showing BLYP-D to be more accurate. This is opposite to what is found when looking at only the S22 database. Antony and Grimme[39] still recommend using the B97 functional, however, "because it provides a more consistent description of non-covalent complexes compared to normal thermochemistry[39]." This statement is supported by the RMSE for the dispersion-bonded complexes in Table 2.1. Antony and Grimme conclude that, "the DFT-D ΔE values are essentially of coupled cluster quality[39]." DFT-D calculations (TZV(2df,2pd) basis) may be computational efficient in comparison to CCSD(T) (extrapolated to the CBS limit); however, even DFT-D calculations are not cheap and are typically only practical for systems consisting of less than a few hundred atoms.

2.3.2 SE-D Methods

Empirically corrected semi-empirical (SE) methods for modeling dispersion interactions have not received a lot of attention in comparison to DFT-D methods, but SE-D methods potentially have a huge computational efficiency advantage over DFT-D. Modeling very large systems (hundreds or even thousands of atoms) with accuracy is of significant interest in the scientific community. Systems such as, molecular devices, polypeptides, DNA and RNA are of great importance, but are currently essentially inaccessible to DFT-D calculations. SE-D methods have the potential of making accurate inter-component interaction energies for such systems computationally accessible. In this

section, the results of empirically correcting the popular AM1 and PM3 semiempirical methods for improved description of dispersion interactions will be considered.

As discussed in Section 2.1, the AM1 and PM3 methods neglect dispersion interactions because they are HF-based and HF theory assumes an uncorrelated electron wavefunction. Incorporating an empirical correction provides one possible approach to modeling van der Waals bonded complexes with SE methodology. McNamara and Hillier[22] have applied Grimme's empirical dispersion correction to the AM1 and PM3 methods. To achieve more accurate results, McNamara and Hillier re-optimized 18 of the AM1 and PM3 parameters (for H, C, N, and O) using the S22 database as the training set. They also considered different global scaling factors when optimizing the AM1 and PM3 parameters. They concluded that the final result is independent of the global scaling factor because the AM1 and PM3 parameters adjust to accommodate the different global scaling factors considered. This finding suggests that some of the AM1 and PM3 parameters did not need to be optimized, just the global scaling factor for each method and possibly also the damping constant. McNamara and Hillier used the same global scaling factor and damping constant for both semiempirical methods and optimized the AM1 and PM3 parameters. More accurate results with less parameterization might have been achieved if optimal parameters for the dispersion term were found for both AM1 and PM3. They choose to use the same parameters that Grimme[34] used with the BLYP functional because of possible QM/MM implications, since using the same dispersion parameters avoids a discontinuity at the QM/MM boundary.

McNamara and Hillier did attempt to adjust only the dispersion parameters, and not change the SE parameters, but "found the final results to be quite poor[22]." They did

not report these values; however, so we performed these calculations and found that the poor results arise almost exclusively from the hydrogen bonded complexes (see AM1-D[a] results in Table 2.5). The re-parameterization of the AM1 and PM3 methods significantly decreases the RMSE of hydrogen bonded complexes from 9.25 to 1.56 kcal/mol for the AM1 methods. By contrast, it has actually increased the RMSE for the dispersion and mixed complexes. This suggests that the carbon, and perhaps hydrogen, parameters should not have been changed. The empirically corrected methods of McNamara and Hillier are significant improvements over the traditional AM1 and PM3 methods. The RMSE for the uncorrected AM1 and PM3 methods are 8.47 and 7.73 kcal/mol[22], based on the S22 database. As can be seen, the empirically corrected AM1-D and PM3-D methods are far superior to their uncorrected counterparts, with RMSEs of 1.23 and 1.18 kcal/mol respectively. The AM1-D method outperforms the PM3-D method for dispersion-bonded complexes; on the other hand, PM3-D outperforms AM1-D for hydrogen bonded complexes. The PM3-D method surprisingly even outperforms the BLYP-D method for hydrogen bonded complexes, based on the RMSE for the S22 database. All of this error analysis data is presented in Table 2.5. Detailed information for the individual complexes is reported in Table B-5.

Table 2.5: Single-point interaction energies (kcal/mol) at the S22 geometries. [a] AM1-D results without re-parameterization of AM1 method (S_6=1.1 and d=23.0). The AM1, PM3, AM1-D, and PM3-D interactions energies used to construct this table were obtained from Ref. 22 and the PM3-D* and PM6-DH values from Ref. 49 and 23 respectively.

	AM1	PM3	AM1-D[a]	AM1-D	PM3-D	PM3-D*	PM6-DH
RMSE (Hydrogen bonded)	11.64	7.77	9.25	1.56	0.76	2.85	1.07
RMSE (Dispersion bonded)	8.21	10.13	0.61	0.82	1.68	0.81	0.54
RMSE (Mixed bonded)	3.57	3.22	0.77	1.25	0.72	0.92	0.57
RMSE	8.47	7.73	5.25	1.23	1.18	1.76	0.76
MUE	6.54	5.94	2.77	0.85	0.90	1.23	0.59

Thus far we have considered only the energies corresponding to the structures in the S22 database, but it is also important to consider the effect of optimizing with the AM1-D and PM3-D methods. McNamara and Hillier[22] optimized all systems in the S22 database with both methods. Upon optimization, the RMSEs for the binding energy increased to 2.47 and 1.60 kcal/mol for the AM1-D and PM3-D methods respectively. These results are reported in Table 2.6, along with the RMSE associated with each subgroup of complexes in the S22 database. (Detailed results for the individual complexes are shown in Table B-6.) The partitioning of errors shows that the dispersion-bonded complexes are drastically affected by optimization with the AM1-D method; the RMSE increased from 0.82 to 2.36 kcal/mol. The RMSE error in the interaction energy for the PM3-D method was hardly affected, suggesting that this method should be preferred over AM1-D. The uncorrected AM1 and PM3 methods actually show improvement for the energies of dispersion-bonded complexes upon optimization. This is just because the reference geometries are repulsive at the uncorrected AM1 and PM3

levels since dispersion interactions are neglected. Thus, upon optimization, the complexes dissociate. This is why it is important to consider structural changes upon optimization. Based solely on the interaction energy analysis, the PM3-D method outperforms the AM1-D method when optimizations are performed. This could just be a result of the parameterization, that is, better parameterization of the AM1 method might be achievable.

Table 2.6: Geometry optimized interaction energies (kcal/mol) for the S22 complexes. The AM1, PM3, AM1-D, and PM3-D interactions energies used to construct this table were obtained from Ref. 22 and the PM3-D* and PM6-DH values from Ref. 49 and 23 respectively.

	AM1	PM3	AM1-D	PM3-D	PM3-D*	PM6-DH
RMSE (Hydrogen bonded)	9.90	8.04	3.38	1.82	2.66	1.40
RMSE (Dispersion bonded)	3.73	3.65	2.36	1.71	2.79	0.80
RMSE (Mixed bonded)	2.96	2.40	2.09	1.40	0.89	0.82
RMSE	6.25	5.22	2.65	1.65	2.31	1.04
MUE	4.82	4.09	2.16	1.51	1.60	0.82

McNamara and Hillier[22] also considered the structural distortion resulting from optimization with the AM1-D and PM3-D methods. The group reported the interaction distances (see Ref. 22) for the AM1, PM3, AM1-D, and PM3-D methods. The resulting error analysis is shown in Table 2.7. (It should be noted that we disagree with the reference interaction distance reported by McNamara and Hillier[22], for the indole-benzene (S14) complex. They reported a value of 3.444 Å; we calculated the value to be 3.498 Å. We are in agreement for all other interaction distances reported.) The RMSE in

interaction distances for the AM1-D and PM3-D methods are 0.419 and 0.249 Å respectively. The PM3-D method more accurately reproduces the energies and structures, overall and in all categories, than the AM1-D method. Not surprisingly, both methods significantly outperform the uncorrected methods. The PM3-D interaction distance error is actually comparable to that of the BLYP-D (Table 2.3), which is 0.183 Å. Most of the difference can be related back to the hydrogen bonded complexes (PM3-D RMSE (H-bonded) = 0.134 Å and BLYP-D RMSE (H-bonded) = 0.035 Å. This suggests that the PM3 method, as well as the AM1 method, needs to be better parameterized for hydrogen bonding. One of the most surprising findings is that AM1-D performs quite poorly for the dispersion bound complexes upon optimization as measured by either the interaction energy or distance. It can be seen in Figure 2.9 that all the complexes are over bound based on interaction energy and most exhibit an over-bound (i.e. too short) interaction distance as well. The same can generally be said for the PM3-D method, however, the errors are not quite as drastic. It is conceivable that a more accurate semi-empirical method could be developed if the training set used to optimize the parameter contained complexes on both sides of the minimum of the potential well. McNamara and Hillier used the S22 database to obtain the new AM1 and PM3 parameters; but this database contains only optimized complexes, i.e. structures at the minimum of the potential energy surface. If a larger and more diverse training set were used, potentially more accurate geometries and energies could be obtained upon optimization. Such an approach could also benefit DFT-D methods.

Table 2.7: Geometry optimized interaction distances (Angstroms) for the S22 complexes. The interaction distance is defined as the CM-distance in complexes 11-15 and 2nd value reported for complex 22. All interactions energies used to construct this table were obtained from Ref. 22.

	AM1	PM3	AM1-D	PM3-D
RMSE (Hydrogen bonded)	0.387	0.257	0.137	0.134
RMSE (Dispersion bonded)	2.015	1.962	0.644	0.272
RMSE (Mixed bonded)	0.929	0.598	0.336	0.315
RMSE	1.277	1.171	0.419	0.249
MUE	0.853	0.691	0.301	0.199

Hillier and colleagues[49] more recently reported a re-parameterized version of the PM3-D method, named PM3-D*. (The results are summarized in Table 2.5 & Table 2.6, details information about the individual systems can be found in Table B-5 & Table B-6.) The training set for optimization of the parameters consisted of the S22 database along with 9 carbohydrate-benzene complexes calculated at the BLYP-D/TZV(2d,2p) level. The group also optimized more parameters associated with the PM3 method and added a tailored version of the core-core repulsion function developed by Voityuk and Rösch[50], in attempt to achieve a more accurate method. The group claims that the new PM3-D* method reduces the geometry optimized mean unsigned error (MUE) from 2.16 to 1.40 kcal/mol for the S22 database complexes; however, we are unable to reproduce these numbers. From the interaction energies reported by Hillier and colleagues, we find the MUE to be 1.51 and 1.60 kcal/mol for the PM3-D and PM3-D* method respectively. It therefore appears that not only are the RMSE numbers reported by Hillier and colleagues incorrect, but the PM3-D* method is worse for optimizing the S22 database complexes than PM3-D. The MUE value reported for the PM3-D optimized energies in Ref. 49

(2.16 kcal/mol), disagrees with what was previously reported in Ref. 22 (1.51 kcal/mol), which is reproducible. The reported MUE value of 2.16 kcal/mol is the same value previously reported for the AM1-D method. The origin of error related to the MUE value associated with the PM3-D* is unknown. To the credit of the PM3-D* method, it does show significant improvement for a variety of interactions of carbohydrates and amino acids with aromatic systems, which was the stated main goal in developing the new method.

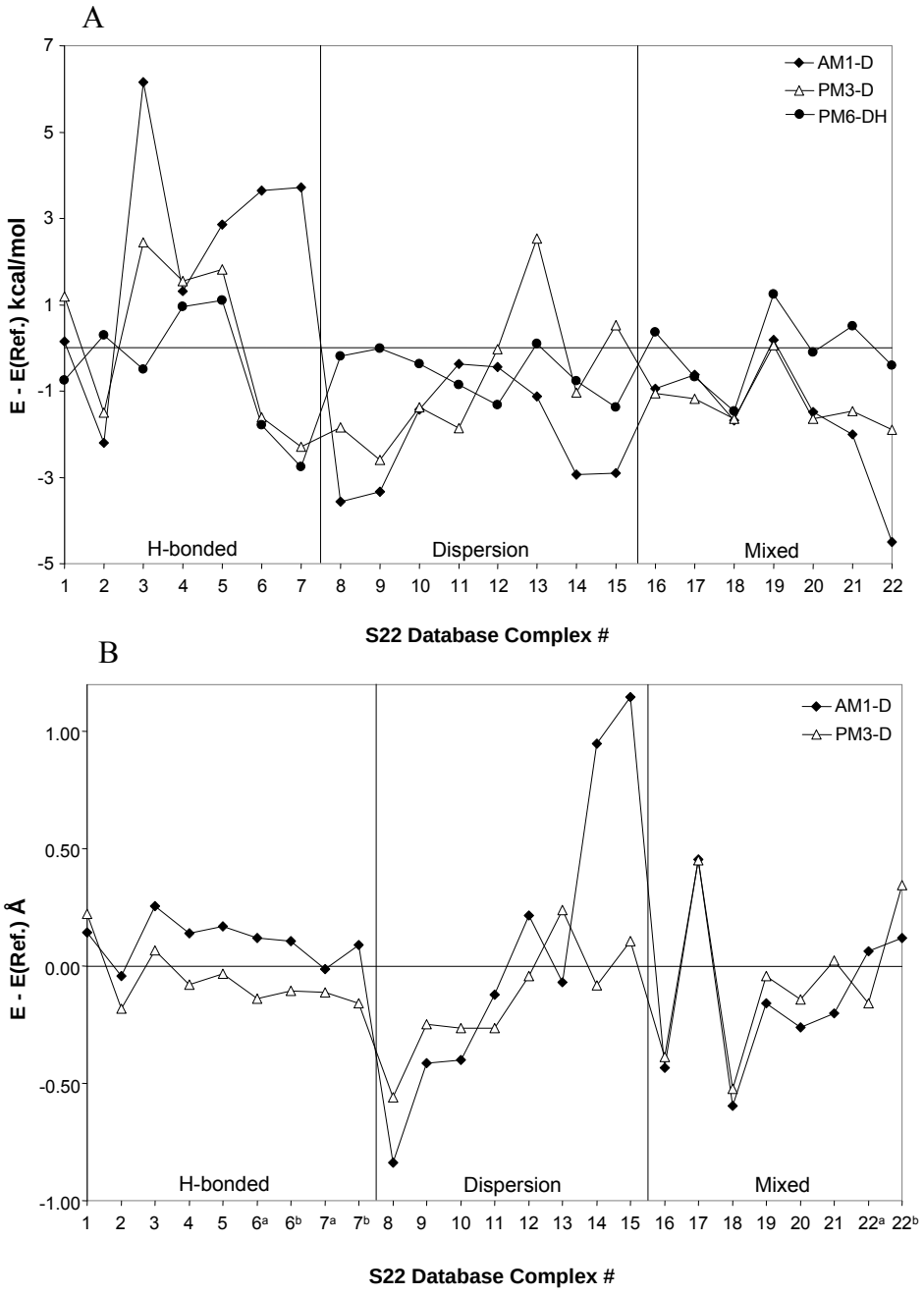


Figure 2.9: Graphical representation showing deviations in the optimized interaction energies (kcal/mol) (Figure A) and distances (Å) (Figure B) from the S22 CCSD(T) reference values[45]. Note a positive deviation indicates that the complex is under bound and if it is negative the complex is over bound. Complexes 6, 7, and 22 are defined by two interaction distances and are represented by superscript *a* and *b*. The AM1-D, and PM3-D interactions energies and distances used to construct this table were obtained from Ref. 22 and the PM6-DH interactions energies from Ref. 23.

Very recently Řezáč and colleagues[23] published an empirically corrected PM6[21] method for dispersion and hydrogen-bonding interactions, named PM6-DH. To incorporate dispersion interaction, the group used the empirical correction described by Jurecka et al.[36] that is they used Eq. 2.19 and Eq. 2.21. The group optimized the two dispersion parameters to complexes 8-22 of the S22 database, a very small training set. Thus, it is not surprising that they have achieved very accurate signal-point interaction energies for the dispersion bonded complexes. The RMSE for the dispersion bonded complexes is 0.54 kcal/mol (see Table 2.5). They also achieved good results for the optimized interaction energy, a RMSE of 0.80 kcal/mol (Table 2.6). Unfortunately, the group did not report any interaction distances or CM-distances, nor did they provide optimized structures. Therefore, we are unable to provide a structural distortion comparison. To gain a little insight into the structural distortion we have constructed a potential energy curve for the parallel benzene dimer. The graph is shown in Figure 2.10. This figure shows that the PM6-DH method severely over binds this dispersion bound system. Based on this one system, it seems that the PM6-DH method could benefit from more parameterization for the dispersion parameters.

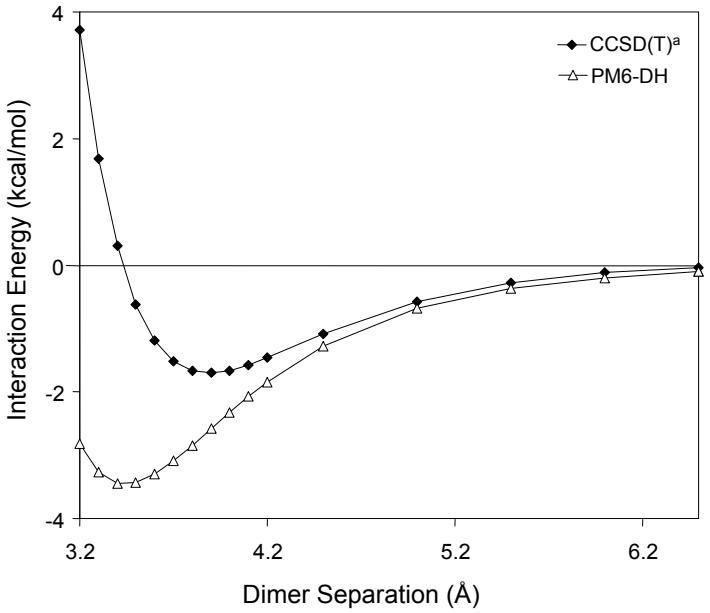


Figure 2.10: Potential energy curves for the parallel benzene dimer as determined with the PM6-DH method. [a] Ref. 51.

To improve the PM6 method for H-bonding Řezáč and colleagues included a second correction term involving 3 parameters. The correction term is applied to H-bonding situations, but not all types of H-bonds are modeled with the same parameters. The group identified 8 types of H-bonds and used a different set of 3 parameters for each type, for a total of 24 H-bonding parameters. The group did use a relatively large training set to determine these parameters. They used 14 point potential energy curves for 104 hydrogen bonded complexes; this is the kind of parameterization training set that could benefit the dispersion parameters. The major shortcoming with their hydrogen bonding correction term is that atom connectivity information is required. Therefore, PM6-DH is unable to model bond formation and bond breaking A major benefit of a quantum mechanical techniques is thereby lost. Even after all this hydrogen bonding

parameterization, the PM6-DH method is outperformed by McNamara and Hillier[22] PM3-D method for the single-point interaction energies (see Table 2.5), albeit not for the optimized interaction energies for the hydrogen complexes. The PM6-DH method results in a RSME for the optimized hydrogen bonded complexes of 1.4 kcal/mol (PM3-D RMSE = 1.82 kcal/mol). As noted before, the ability for a method to perform well upon optimization is arguably much more important.

The PM6-DH method currently produces the lowest RMSE for the full S22 database for both the signal-point (0.76 kcal/mol) and optimized energies (1.04 kcal/mol); when compared to other published empirically corrected SE methods. The optimized interaction energy and distance results for the AM1-D, PM3-D and PM6-DH (optimized interaction distances are not available) are graphically summarized in Figure 2.9. Figure 2.9A shows the optimized interaction energy deviations from the reference values as a function of S22 database complex number. Looking at the hydrogen bonding curves it can be seen that the points are distributed evenly (i.e. similar positive and negative deviations) for the PM3-D and PM6-DH methods. This likely indicates that optimal parameterization has been achieved. Thus, to achieve appreciable improvements an additional or different correction term would be needed. On the other hand, the dispersion complexes are virtually all over bound, indicting further benefits likely could be achieved with better parameterization. This over binding is not solely due to the parameters in the dispersion term; it is caused by the weak repulsive wall associated with the SE methods. This weak repulsive wall is an artifact of the minimal basis set used in the NDDO approximation[52] that is commonly employed in SE methods. The effect of using a minimal basis set is shown in Figure 2.11; where the HF method is used to model

the parallel benzene dimer with different basis sets. As the number of valence basis functions increase from 1 to 3 (STO-3G → 3-21G → 6-311G), the repulsive wall also increases. Thus, the use of a minimum basis severely limits the accuracy. Another issue, as mentioned earlier, is that SE methods are parameterized to experimental data. Since experimental data includes electron correlation, SE methods implicitly incorporate some correlation, although, not explicitly in the wavefunction.

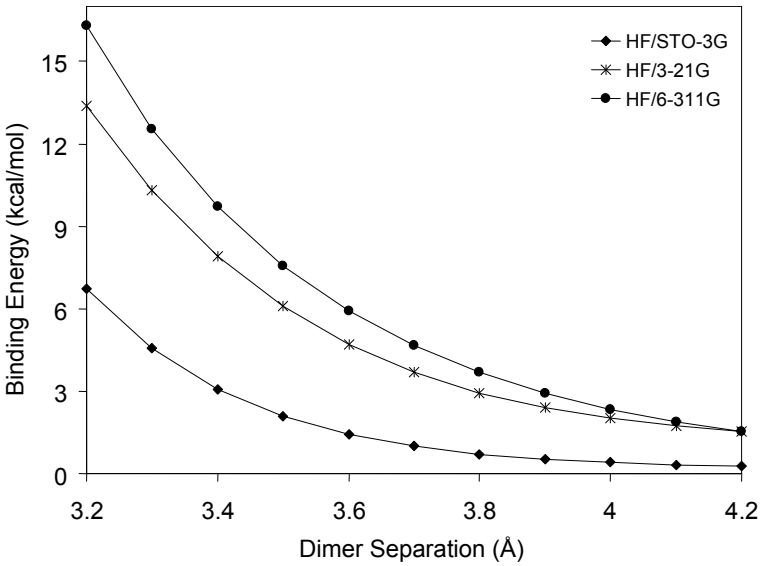


Figure 2.11: Potential energy curves for the parallel benzene dimer as determined with the HF method using if basis sets. This figure serves as a visual aid showing the affect of basis set on binding energy at short-range.

McNamara and Hillier[22] also considered the *JSCH-2005* database, as did Antony and Grimme[39]. They neglected complexes containing sulfur, however, because no systems containing sulfur are present in the S22 database, which was their training set for

parameterization. This reduced the size of the database to 156 complexes instead of 161 as used by Antony and Grimme. The RMSE for the total database at the reported geometries are 1.54 and 1.68 kcal/mol for AM1-D and PM3-D respectively. These errors represent a significant improvement over the standard AM1 and PM3 methods. RMSE for the SE-D methods is about twice that of the B97-D (0.92 kcal/mol) and BLYP-D (0.85 kcal/mol) methods; however, significantly less computational time is required. Hillier's group[43] later re-parameterized AM1-D and PM3-D for sulfur and claimed to achieve significant improvements. These empirically corrected SE methods considered are not of DFT-D quality; however, these methods are capable of model system far beyond the scope of DFT-D methods.

2.4. Conclusion

Empirically corrected DFT and SE methods for dispersion interactions have been used with considerable success. It is found, not unexpectedly, that DFT-D methods are more accurate than SE-D methods. This accuracy is a result of DFT methods more precisely describing the electrostatic and exchange-repulsion interactions. Consequently, DFT methods describe hydrogen bonding with considerably higher accuracy. In both cases, adding an empirical dispersion correction *significantly* increases the accuracy for modeling dispersion-bonded complexes, since in both cases (depending on the DFT functional) inter-component dispersion interactions are neglected. This seems to suggest that, in theory, the same accuracy could be achieved with a SE-D method when dispersion interactions are the only inter-component binding forces, if optimal empirical parameters are used. In reality this is not possible without an addition correction term for the repulsive wall, or at least not without significant re-parameterization of the SE

method itself. The empirically corrected and re-parameterized AM1-D and PM3-D methods are significant improvements over their uncorrected predecessors. Even greater accuracy has been achieved with the recently published PM6-DH method.

Partitioning the RMSE for the different subset of complexes showed that the dispersion-bonded complexes are described well with DFT-D and the SE-D methods, based on the S22 database geometries. Upon geometry optimization, the PM6-DH method outperforms both the AM1-D and PM3-D methods. For the DFT-D methods, the TPSS-D method performs the best. This method reproduces the same RMSE for dispersion complexes and mixed complexes upon optimization based on the interaction energy. The TPSS-D method also distorts the structure the least upon optimization. Both DFT-D and SE-D methods perform less-well for hydrogen bonded complexes.

In order for these methods to be widely applicable, they must produce reasonable optimized structures. Therefore, it is very important to consider the geometries produced upon optimization. Improvements could perhaps be achieved if the training set for optimization of the empirical parameters contained complexes not just at the potential minimum, as in the S22 database. Considering complexes on both sides of the potential minimum should help produce better results upon optimization. Such improvement would be at the expense of less accurate single-point energies though. Although DFT-D methods perform relatively well in this area, increasing the training set would very likely benefit the methods.

Another important issue is the computational efficiency of the two methods. Though DFT-D methods provide more accurate results, the computational expense can be 100-1000 times greater. Therefore, SE-D methods have a huge computational cost

advantage. If it is desired to accurately model systems composed of less than 100 atoms or so, DFT-D is a better choice. When modeling systems composed of several 100 or even 1000's of atoms, however, SE-D methods are far superior from the standpoint of computational efficiency.

Chapter 3: A New Empirical Correction to the AM1 Method for Macro-Molecular Complexes

3.1 Introduction

Herein, we present an empirical correction for the AM1 method that is suitable for modeling macromolecular complexes and avoids the noted shortcomings of existing techniques mentioned in Chapter 1. We have chosen to apply separate empirical correction terms for dispersion and hydrogen-bonding. Our method requires significantly less parameterization than the AM1-D and PM3-D methods of McNamara and Hillier[22] and also the PM6-DH method of Řezáč et al.[23]. Additionally, it is important to note that we have not altered any of the original AM1 parameters. Such changes can have deleterious effects on predictions of properties not based strictly on the total energy or its derivatives; such as heats of formation, ionization potentials, and dipole moments, if these quantities are not taken into consideration during re-parameterization. Our method also does not require knowledge of atom connectivity. We will henceforth refer to our new method as "AM1-FS1". AM1-FS1 achieves results that are comparable to, (and in many cases better than) those of other empirically corrected SE methods, with *significantly* less parameterization and with no reparameterization of the AM1 method. The main objective of AM1-FS1 is to accurately model macromolecular host/guest systems that are currently out of reach of DFT-D techniques. AM1-FS1 aims to not only accurately predict energies but also reasonable structures upon geometry optimization, since structural optimization is one of the main uses for such a technique. Herein, the accuracy of AM1-FS1 is tested by comparing interaction energies and distances to

CCSD(T) and SAPT results; comparisons are also made with other empirically corrected semi-empirical techniques.

3.2 Theory

3.2.1 Dispersion Correction

To correct the AM1 method for dispersion interactions, we have employed a method used by Grimme[35] with a slight modification suggested by Jurečka et al.[36]. The resulting dispersion correction is of the form:

$$E_{dis} = -\frac{C_6^{ij}}{r_{ij}^6} f_{damp}\left(r_{ij}\right) \quad , \tag{3.1}$$

where r_{ij} is the atom-atom separation, C_6^{ij} is the dispersion coefficient, and f_{damp} is a damping function of the form,

$$f_{damp}\left(r_{ij}\right) = \frac{1}{1 + e^{-d\left(\frac{r_{ij}}{S_R R_{vdw}} - 1\right)}} \quad . \tag{3.2}$$

This damping function depends on the equilibrium van der Waals separation (R_{vdw}) and the pair-wise atom separation (r_{ij}). The damping function also depends on two unitless parameters S_R and d, which have been optimized to a training-set as discussed at length in Section 3.2.3. The damping function operates as a switching function, turning off the dispersion term at short range. This is required because the SE wavefunction already models short-range repulsive interactions. Thus the popular 6-12 Lennard-Jones (LJ) potential is not suitable for use as a dispersion correction since a repulsive term is involved. (See Ref. 53 for a more detailed discussion and graphical representations).

It should be noted that we tried employing the global scaling factor, used by Grimme[34, 35], instead of scaling the equilibrium van der Waals separation (S_R); however, a smaller root mean square error (RMSE) was obtained on our training-set using S_R (discussed in section 3.2.3). Scaling R_{vdw} seems theoretically well motivated, since this allows only the short-range interactions to be tailored and leaves untouched the long-range interactions for which the correct functional form of the interaction is known to follow r^{-6}. (See Ref. 53 for a more detailed discussion and graphical representations).

Another decision concerns the choice of combination rules used for obtaining C_6^{ij} and R_{vdw}. We have chosen to employ the geometric mean and simple average combination rules for determining C_6^{ij} and R_{vdw}, respectively:

$$C_6^{ij} = \sqrt{C_6^i C_6^j} \quad , \quad R_{vdw} = \frac{R_i + R_j}{2} \quad . \tag{3.3}$$

The dispersion coefficients (C_6^i and C_6^j) and van der Waals radii (R_i and R_j) for the different atoms were obtained from Grimme's 2006 publication[35]. The decision of using these particular combination rules was not made without considering other options. For C_6^{ij}, both the harmonic mean[34] and the combination rule suggested by Wu and Yang[31], which uses the Slater-Kirkwood effective number of electrons, were considered. For R_{vdw}, the cubic mean suggested by Halgren[38] was also considered. We have also considered all possible combinations and found that the parameters (S_R and d) seemed to adjust to accommodate the different combination rules. The combination rules employed yielded the lowest RMSE for our training-set. It should be noted that only Grimme's 2006 published dispersion coefficients and van der Waals radii values were considered.

This dispersion correction scheme and its gradient (see Appendix C) has been added to the AM1 method and implemented into a locally modified version of GAMESS[54].

3.2.2 Hydrogen-Bonding Correction

Correcting the AM1 method for hydrogen-bonding is a more difficult task than correcting for its neglect of dispersion since hydrogen-bonding interactions are already in-part considered, given their partial electrostatic nature. It can be seen by looking at the H-bonded systems (1-7) in the S22 database (Table 3.1) that the AM1 method severely under binds such complexes. The AM1 method does, however, produce more reasonable interaction energies for hydrogen-bonded systems upon geometry optimizations (Table 3.2); this is because AM1 generally predicts dispersion-bound complexes to be unbound, while for H-bonded complexes it predicts some bonding, but generally with an unphysically large equilibrium separation (see Table 3.3). Thus, to improve the AM1 method for predicting H-bonding systems, the strength of these interactions needs to be increased at medium-to-short range. We have achieved this by adding a post-SCF (this term has since been added in a SCF manner, see Appendix D) pseudo-electrostatic term of the form:

$$E_{HB} = \alpha_1 \frac{Q_i Q_j}{r_{ij}} \cos^2(\theta) f_{damp2}(r_{ij}) \tag{3.4}$$

where α_1 is a global scaling factor, Q_i and Q_j are the AM1 Coulson charges[55] (which are referred to as MOPAC charges in GAMESS[54]), r_{ij} is the H---Y separation, θ is the XH---Y angle, and f_{damp2} is a damping function of the form:

$$f_{damp2}(r_{ij}) = e^{\frac{-(r_{ij}-\alpha_2 R_{vdw})^2}{\alpha_3^2 (1+\alpha_4 (r_{ij}-\alpha_2 R_{vdw}))^2}}, \tag{3.5}$$

where α_2, α_3, and α_4 are parameters and all other terms have the same meanings as in the dispersion correction. In this case, however, R_{vdw} is defined as the cubic mean,

$$R_{vdw} = \frac{R_i^3 + R_j^3}{R_i^2 + R_j^2} \quad . \tag{3.6}$$

The cubic mean is used in this case because it yields a slightly smaller RMSE for the F66 training set than using the simple average combination rule.

The damping function is an asymmetric distribution function (see Figure 3.1A) that turns the hydrogen-bonding function on/off over an appropriate range for correcting the AM1 method. To achieve an asymmetric distribution, three parameters (α_2, α_3, α_4) have been introduced, giving a total of 4 parameters in the H-bonding correction. We have optimized these four parameters to improve upon H-bonding for the AM1 method. A detailed discussion is presented in the next section.

The H-bonding correction function also depends on the square of the cosine of the XH---Y angle. This is motivated by the observation that H-bonding interactions are directionally dependent[56]. The cosine squared function was used instead of the cosine function because it approaches zero smoothly. We have also chosen to make the function zero for all angles less then 90 degrees. This helps exclude cases that are not H-bonding; such as an alpha-H atom in a carboxylic acid interacting with the adjacent carboxylate O atom. By using an appropriate summing scheme, we are able to identify highly likely H-bonding scenarios without knowledge of atom connectivity. This is done by first identifying H atoms for which the nearest neighbor is an N, O, or F atom. These H atoms are then allowed to interact with other N, O, or F atoms.

The overall function (Eq. 3.4) is shown graphically in Figure 3.1B, where it can be seen that the function is only turned on over a short range, peaking at approximately 2.8 bohr (1.5 Å), (for the specific bonding scenario depicted). This is the behavior that is needed to improve the AM1 method for H-bonding, since these interactions only need to be increased over a short range and only at short distances. The nature of the charges (MOPAC) of the atoms involved in H-bonding insures that Eq. 3.4 is negative, resulting in an attractive contribution. The charges are updated every optimization step. The optimization procedure also requires the gradient, which is determined by numerical differentiation. This correction scheme and its gradient (see Appendix C) has been implemented into a locally modified version of GAMESS[54].

The hydrogen bonding correction scheme as described above is continuous for proton transfer under most conditions. In most cases, the correction term effectively turns off, (i.e. is essentially equal to zero) before the proton reaches the half-way point in a proton transfer. For example, when a proton transfers between two formic acid molecules, the intercomponent oxygen-oxygen separation is about 2.7 Å; therefore, when the proton reaches the half-way point the H---Y distance is about 1.35 Å (2.55 bohr) and the H-bonding correction term is approximately zero (see Figure 3.1B). The function is also continuous when the molecules are separated by a greater distance even though there is a nonzero correction at the half-way point because at this point the function is identical in both directions (H---Y equals X---H). When the proton passes the half-way point, the H-bonding correction term corrects in the opposite direction. If a proton is transferring across an asymmetric system, however, a discontinuity can occur since the charges on the X and Y atoms may not be the same. This discontinuity can be eliminated by evaluating

Eq. 3.4 in both directions at all times and taking the correction to be a weighted sum of the two. This correction has been implemented into AM1-FS1 and has no effect on any of the binding energies reported in this manuscript, since evaluating the function in the opposite direction (considering the X---H bond to be H-bonding) leads to no correction because the X---H distance is essentially always less than 1 Å (1.9 bohr). In summary, the switching transition from one H-bonding situation to another is effectively continuous during a proton transfer. We currently cannot recommend AM1-FS1 for modeling proton transfers, however, since it has not been tested and more importantly because our training set does not contain data to parameterize for such situations. Nevertheless this correction scheme does not produce discontinuities. High quality (CCSD(T) and DFT-SAPT) proton transfer potential energy curve are scarce, rendering such a parameterization difficult at this time. We plan to explore this avenue in the future.

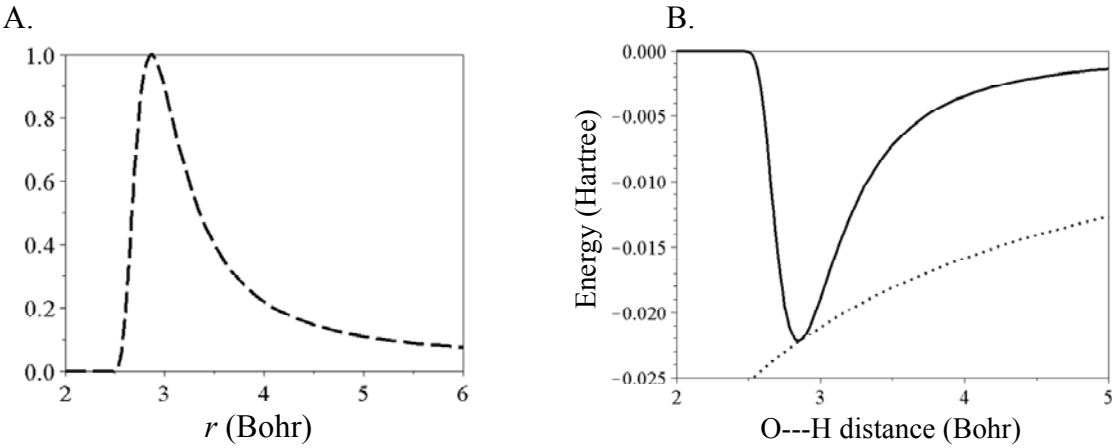


Figure 3.1: Graphical representation of the H-bonding damping function (A), the entire correction term (solid-line) and the electrostatic attractive portion (dotted-line) (B) used in the AM1-FS1 method. This model is for the case of the alpha-hydrogen atom (connected to the nitrogen atom) interacting with the parallel oxygen atom on the second monomer of the uracil dimer in the hydrogen bonding conformation. The MOPAC charges used correspond to the minimum energy structure (O---H R=1.77); this simplification has little effect on the functional form. This simplification has been used for graphical convenience.

3.2.3 Parameter Optimization

To improve the AM1 method for dispersion and hydrogen-bonding interactions two empirical correction terms have been added as discussed above. These two correction terms involve a total of 6 parameters: two for the dispersion term (Eq. 3.1) and four for the H-bonding term (Eq. 3.4). These 6 parameters have been mathematically optimized to the RMSE of the interaction energies of 66 complexes (the F66 training set, see Table E-1). All of the interaction energies in the training set are CCSD(T) or SAPT quality. The training set consists of complexes not only at their minimum energy structures but also at greater and lesser separation than the potential minimum. Inclusion of these non-equilibrium structures is intended to increase the reliability of geometry optimization with AM1-FS1.

Our F66 training set includes the complexes in the S22 database[45], which has been used by others for similar parameterization purposes[22, 36, 46]. We have also included the four additional H-bonded complexes[57] that were later introduced to the S22 database, now termed the S26 database. The additional interaction energies are also CCSD(T) quality. In our F66 training set the water dimer, T-shaped benzene dimer, and both uracil dimer structures from the S22 database have been replaced by 5 points on their respective interaction potential energy curves. In addition, 5 point potential energy curves have been added for the nitromethane dimer[58], parallel[51] and M1[59] benzene dimer, and three different benzene-acetylene[60] dimer conformations. For a detailed list of complexes in the training set refer to Table E-1 in the supplementary material. It would be desirable to have more potential energy curves in the training set, but there is limited high quality data

available. For training set proposes we have restricted ourselves to using only CCSD(T) or SAPT results, and only at or near the complete basis set limit.

Upon optimization of the parameters, the damping coefficient (*d*) in Eq. 3.2 optimized to infinity. This is because the AM1 method, as well as other semi-empirical methods, inaccurately models repulsive interactions at close range for dispersion bound complexes. This can be observed by comparing DFT and semi-empirical (AM1, PM3, RM1, and PM6) potential energy curves for the parallel benzene dimer, as shown in Figure 3.2. The figure clearly shows that at close separation the semi-empirical methods (AM1, PM3, RM1, and PM6) differ significantly from the DFT (BLYP/6-311G(d,p)) results, severely underestimating the repulsion at close separations. The inaccurate repulsive inner wall of the potential is a consequence of the minimal basis set and parameterization of the SE methods[53]. This inaccuracy is the origin of *d* optimizing to infinity. As *d* becomes larger, the dispersion correction is turned off more rapidly, however, the function cannot become positive as needed to correct for underestimation of the repulsion at short range by the SE method. This problem could potentially be improved if a 6-12 LJ potential was used and only inter-component atom pairs were considered; however, this introduces the requirement of atom connectivity information. It would also introduce a discontinuity in the potential and/or its derivative during bond breaking and formation processes, (not to mention that intra-component dispersion interactions would be neglected completely, thereby rendering the method ineffective for modeling conformational preference in macromolecules). We have therefore chosen to set *d* equal to 1000 and fully optimize the other 5 parameters. The damping coefficient was chosen to be 1000, because this is at the computational limit for evaluating the

derivative of Eq. 3.1 within double precision. (Derivative information is needed for structural optimizations.) The other 5 parameters optimized to the following values: S_6 = 1.1059, α_1 =0.4882, α_2 =0.6211, α_3 =0.3344, and α_4 =1.5451.

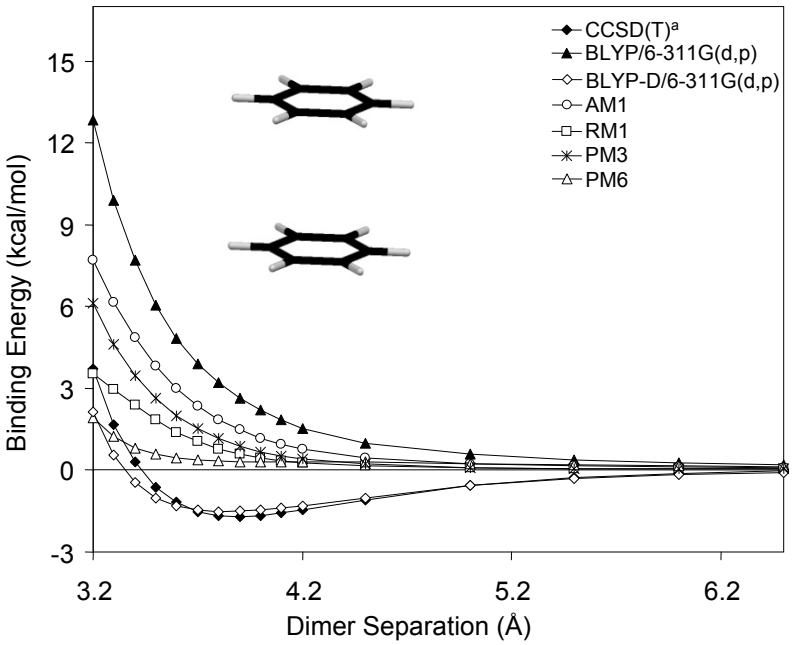


Figure 3.2: Potential energy curves for the parallel benzene dimer determined with various quantum mechanical methods. [a] Values from Ref. 51.

3.2.4 Why Begin with AM1?

AM1 has long been accepted as one of the most robust semi-empirical methods. This method has been used many times with success for modeling large systems, but this is not the only reason for choosing AM1. We applied the same correction scheme described above to the RM1 method, which is a reparameterized version of AM1. The

"corrected" RM1 method was actually less successful, based on the RMSE for the F66 training set. Upon further investigation, we found that the RM1 method, (as well as the PM3 and PM6 methods) performs worse than the AM1 method for the benzene dimer when compared to the DFT results that neglect dispersion interactions, as discussed above. Thus, if the same dispersion correction is applied to these mentioned SE methods, the AM1 method will produce the best result; even though other uncorrected SE methods produce potential energy curves closer to the CCSD(T) results. This is because the AM1 method has the strongest repulsive wall, therefore, producing a potential energy curve closest to the DFT result (see Figure 3.2). The functional form of the dispersion correction (Eq. 3.1) does not allow the term to become positive as is needed in some cases. This can be easily seen for the PM6 results in Figure 3.2. At close range the CCSD(T) results are more repulsive then the PM6 results, thus to make the PM6 curve identical to the CCSD(T) curve a repulsive correction would be needed. This problem is less severe for the AM1 method, rendering it more suitable for modeling dispersion interactions at close range. Note that these findings again might lead one to believe that using a function like the LJ potential would be beneficial, since a repulsive term is included. In fact, the LJ potential was among our many attempts to improve the AM1 method, but without success. This was due to the fact that we were/are unwilling to add the burden of requiring of atom connectivity information. We believe such a burden outweighs the potential added benefit.

3.3 Validation Studies

3.3.1 Single-Point Energies

In Table 3.1, the single-point interaction energies for the structures in the S22 database[45] are compared for various corrected and uncorrected SE methods. First, note that the interaction energies for the uncorrected AM1 and PM3 methods deviate significantly from the CCSD(T) reference values. The root mean square error (RMSE) for the AM1 and PM3 methods is 8.47 and 7.73 kcal/mol, respectively. Not only are these errors very large, but in many cases the sign of the interaction is predicted incorrectly. That is, the interactions are predicted to be repulsive not attractive. The addition of an empirical correction term(s) can drastically improve these methods. Our AM1-FS1 method reduces the RMSE to 1.18 kcal/mol, with the correct sign being predicted in all cases.

The results from McNamara and Hillier's[22] AM1-D and PM3-D methods and the PM6-DH method of Řezáč et al.[23] are also reported in Table 3.1. AM1-FS1 shows a slight improvement over the AM1-D method in two of the three subcategories and over all has a lower RMSE. AM1-FS1 achieves comparable accuracy to the PM3-D method for intermolecular interaction energies; the RMSE are both 1.18 kcal/mol with AM1-FS1 achieving a slightly lower MUE. While the overall improvement achieved by AM1-FS1 in the accuracy with which intermolecular binding energies are predicted is minor, we note that this has been achieved with significantly less parameterization and *no* modification of the original AM1 parameters.

The recently published PM6-DH method[23] slightly outperforms AM1-FS1 based on the single-point energies for the S22 database. Looking at the hydrogen bonded

complexes, the RMSE are 1.07 and 1.37 kcal/mol for the PM6-DH and AM1-FS1 methods respectively. Given that PM6-DH requires *different* parameters for each type of hydrogen bond, the 0.3 kcal/mol improvement in RMSE shown by PM6-DH is not especially significant. The group has identified 8 H-bonding scenarios resulting in a total of 24 parameters for their H-bond correction term (three parameters for each H-bonding type). AM1-FS1 only uses 4 parameters; AM1-FS1 also does not introduce the requirement of knowing atom connectivity. The PM6-DH method does show significant improvement for many dispersion bonded cases, but it performs poorly for modeling the potential energy surface of the benzene dimer. This is discussed below and shown graphically in Section 3.3.4. It should be noted that Řezáč et al.[23] only used complexes 8-22 of the S22 database for determining the dispersion parameters for PM6-DH. Thus, it is not unexpected that good agreement was achieved for the 8 dispersion bound complexes.

Table 3.1: Single-point interaction energies (kcal/mol) at the S22 geometries. The AM1-D and PM3-D results have been taken from Ref. 22, and the PM6-DH results from Ref. 23.

No.	Molecule (symmetry)	Ref. Values	AM1	PM3	AM1-D	PM3-D	PM6-DH	AM1-FS1
Hydrogen bonded complexes								
1	(NH3)2 (C2h)	-3.17	-0.78	0.77	-3.43	-1.77	-3.74	-1.60
2	(H2O)2 (Cs)	-5.02	-2.89	-2.79	-7.29	-5.14	-4.67	-5.53
3	Formic acid dimer (C2h)	-18.61	1.54	-9.91	-15.45	-18.57	-17.39	-16.06
4	Formamide dimer (C2h)	-15.96	-12.02	-8.08	-17.16	-15.37	-15.39	-15.75
5	Uracil dimer (C2h)	-20.65	-5.79	-11.32	-20.15	-20.30	-18.84	-20.80
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	-4.45	-7.46	-16.50	-17.52	-17.35	-14.73
7	Adenine thymine WC (C1)	-16.37	-4.28	-6.79	-16.58	-17.33	-17.83	-16.29
Complexes with predominant dispersion contribution								
8	(CH4)2 (D3d)	-0.53	0.21	-0.25	-0.94	-1.24	-0.73	-0.61
9	(C2H4)2 (D2d)	-1.51	-0.13	-1.11	-3.31	-3.60	-1.52	-2.27
10	Benzene CH4 (C3)	-1.50	0.40	-0.19	-2.12	-2.42	-1.75	-1.79
11	Benzene dimer (C2h)	-2.73	3.52	2.38	-2.90	-4.30	-3.62	-2.23
12	Pyrazine dimer (Cs)	-4.42	2.49	3.90	-4.57	-4.20	-5.41	-3.81
13	Uracil dimer (C2)	-10.12	0.12	5.80	-10.56	-6.78	-9.70	-8.47
14	Indole benzene (C1)	-5.22	5.39	4.04	-4.04	-6.09	-5.20	-3.23
15	Adenine thymine stack (C1)	-12.23	2.91	7.37	-12.20	-10.63	-12.78	-9.87
Mixed complexes								
16	Ethene ethyne (C2v)	-1.53	-0.35	-0.82	-1.61	-1.85	-1.11	-1.36
17	Benzene H2O (Cs)	-3.28	-0.69	-1.47	-3.43	-3.65	-3.41	-2.78
18	Benzene NH3 (Cs)	-2.35	-0.33	-0.59	-3.00	-2.96	-2.77	-2.65
19	Benzene HCN (Cs)	-4.46	-0.81	-1.63	-4.44	-4.43	-3.20	-3.17
20	Benzene dimer (C2v)	-2.74	0.37	-0.43	-3.85	-4.15	-2.84	-3.36
21	Indole benzene T-shape (C1)	-5.73	-1.05	-1.25	-7.10	-6.65	-5.30	-4.63
22	Phenol dimer (C1)	-7.05	-1.36	-1.37	-9.76	-7.52	-6.73	-6.91
	RMSE (Hydrogen bonded)		11.64	7.77	1.56	0.76	1.07	1.37
	RMSE (Dispersion bonded)		8.21	10.13	0.82	1.68	0.54	1.30
	RMSE (Mixed bonded)		3.57	3.22	1.25	0.72	0.57	0.72
	RMSE		8.47	7.73	1.23	1.18	0.76	1.18
	MUE		6.54	5.94	0.85	0.90	0.59	0.88

3.3.2 F66 Results

Many of the other empirically corrected SE methods discussed have been parameterized to the S22 database, thus should achieve accurate results for those complexes. AM1-FS1 has been parameterized to a larger training set consisting of 66 complexes. Parameterizing to this larger training set has led to an increase in RMSE for the S22 database. This is not unexpected and, in our opinion, is a worthwhile sacrifice that should make AM1-FS1 more versatile. (In fact, we tried optimizing AM1-FS1 solely to the S22 database and achieved near DFT-D level accuracy, but when the method was subsequently tested on the F66 training set, a larger RMSE resulted.) AM1-FS1 is

parameterized to the F66 training set and achieves a sub kilocalorie RMSE (0.99 kcal/mol) and MUE (0.69 kcal/mol). The individual results are reported in Table E-1 in the supplementary materials. Both AM1-D and PM3-D were parameterized solely to the S22 database, so high accuracy is not surprising when the S22 database is used as the "test set". We have performed calculations on the 66 complexes of the F66 set using McNamara and Hillier's AM1-D method for comparison. This provides for a much more comprehensive test of the method than the S22 database because it contains a wider variety of structures *and* non-equilibrium structures. AM1-D produces a RMSE and MUE of 1.49 and 1.02 kcal/mol respectively, approximately 50% less accurate than AM1-FS1. Upon close inspection of Table E-1, it can be seen that AM1-FS1 significantly outperforms AM1-D on the repulsive wall; an issue we will look at more closely in Section 3.3.4.

3.3.3 Optimized Energies and Structures

This section considers the effect of geometry optimization on interaction energy and structural distortion for systems in the S22 database. The ability of an empirically corrected SE method to perform accurately in this role is crucial because one of the principal uses of SE methodology is structural optimization of systems that are too large for optimization with first-principles methods. The ability of a method to reproduce interaction energies at *reference* geometries is not very useful, because if we know the CCSD(T) geometry, and therefore its energy, there is little value in knowing the SE energy for that structure. In Table 3.2, the interaction energies for the geometry optimized complexes are reported for a variety of corrected and uncorrected SE methods. Again, the uncorrected AM1 and PM3 methods perform poorly. Upon applying our

correction scheme to the AM1 method, the RMSE is lowered from 6.25 to 1.80 kcal/mol. AM1-FS1 is also an improvement over AM1-D; AM1-FS1 outperforms AM1-D in all subcategories by about 1 kcal/mol. This increase in performance for optimization, compared to the AM1-D method, presumably results from our use of a substantially larger training set that includes non-equilibrium structures.

The performance of McNamara and Hillier's PM3-D method is comparable to our AM1-FS1 method. Depending of the statistical metric selected, either may be said to outperform the other for predicting interaction energies upon geometry optimization of the structure in the S22 database. AM1-FS1 does perform better in two of the three categories, the dispersion and mixed bounded complexes. The PM6-DH method outperforms all the other methods; however, structural distortion should also be considered but, unfortunately, data for such a comparison is not available. Again, this aspect of a SE method is especially important since such a method will likely be used for optimization purposes.

To gauge the degree of structural distortion upon geometry optimization, select interaction distances are compared and are shown in Table 3.3. The interaction distances are defined as the center-of-mass separation and/or atom-atom distance(s) between the two monomers depending on the system (see Figure S1 of Ref. 22 for the specific interaction distances). Comparing the different empirically corrected SE methods we find that AM1-FS1 outperforms AM1-D in every category. Our method is generally comparable to PM3-D based on interaction distance. AM1-FS1 performs better in two of the three categories. This time AM1-FS1 outperforms PM3-D for the H-bonded complexes based on interaction distances, but not for the dispersion bound complexes.

Based solely on the total RMSE for the S22 database would be difficult to choose which method, AM1-FS1 or PM3-D, is better; however, AM1-FS1 does not require reoptimization of the AM1 parameters thereby preserving the predictive power of AM1 for calculation of heats of formation, discussed below. As noted above, interaction distances and/or structural geometries were not made available for the PM6-DH method preventing structural comparisons upon optimization of the S22 complexes.

Table 3.2: Geometry optimized energies (kcal/mol) for the complexes in the S22 database. The AM1-D and PM3-D results have been taken from Ref. 22, and the PM6-DH results from Ref. 23.

No.	Molecule (symmetry)	Ref. Values	AM1	PM3	AM1-D	PM3-D	PM6-DH	AM1-FS1
Hydrogen bonded complexes								
1	(NH3)2 (C2h)	-3.17	-1.39	-0.71	-3.03	-1.99	-3.92	-2.82
2	(H2O)2 (Cs)	-5.02	-3.30	-3.55	-7.22	-6.53	-4.73	-5.59
3	Formic acid dimer (C2h)	-18.61	-6.62	-9.58	-12.45	-16.16	-19.11	-17.76
4	Formamide dimer (C2h)	-15.96	-2.06	-6.99	-14.64	-14.42	-15.01	-15.83
5	Uracil dimer (C2h)	-20.65	-10.48	-10.70	-17.80	-18.83	-19.55	-25.06
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	-6.15	-7.06	-13.06	-18.32	-18.50	-15.16
7	Adenine thymine WC (C1)	-16.37	-5.06	-6.90	-12.66	-18.66	-19.12	-21.10
Complexes with predominant dispersion contribution								
8	(CH4)2 (D3d)	-0.53	-0.21	-0.32	-4.10	-2.38	-0.73	-2.46
9	(C2H4)2 (D2d)	-1.51	-0.13	-1.08	-4.85	-4.11	-1.53	-4.09
10	Benzene CH4 (C3)	-1.50	0.35	-0.20	-2.93	-2.88	-1.88	-2.84
11	Benzene dimer (C2h)	-2.73	0.01	-0.02	-3.10	-4.59	-3.59	-2.21
12	Pyrazine dimer (Cs)	-4.42	-0.34	-0.26	-4.87	-4.45	-5.74	-4.73
13	Uracil dimer (C2)	-10.12	-6.05	-4.26	-11.25	-7.59	-10.03	-9.99
14	Indole benzene (C1)	-5.22	-1.33	-1.65	-8.16	-6.26	-5.99	-6.51
15	Adenine thymine stack (C1)	-12.23	-5.15	-6.50	-15.13	-11.70	-13.61	-12.59
Mixed complexes								
16	Ethene ethyne (C2v)	-1.53	-0.57	-1.23	-2.47	-2.58	-1.17	-1.50
17	Benzene H2O (Cs)	-3.28	-1.03	-1.63	-3.90	-4.46	-3.95	-3.38
18	Benzene NH3 (Cs)	-2.35	-0.80	-0.93	-4.04	-3.99	-3.82	-4.70
19	Benzene HCN (Cs)	-4.46	-0.92	-1.85	-4.28	-4.40	-3.21	-2.46
20	Benzene dimer (C2v)	-2.74	-0.09	-0.52	-4.22	-4.39	-2.85	-2.15
21	Indole benzene T-shape (C1)	-5.73	-1.24	-1.67	-7.74	-7.20	-5.22	-5.88
22	Phenol dimer (C1)	-7.05	-3.39	-4.33	-11.55	-8.95	-7.46	-8.87
	RMSE (Hydrogen bonded)		9.90	8.04	3.38	1.82	1.40	2.55
	RMSE (Dispersion bonded)		3.73	3.65	2.36	1.71	0.80	1.34
	RMSE (Mixed bonded)		2.96	2.40	2.09	1.40	0.82	1.37
	RMSE		6.25	5.22	2.65	1.65	1.04	1.82
	MUE		4.82	4.09	2.16	1.51	0.82	1.28

Table 3.3: Interaction distances (Angstroms) for the complexes in the S22 database. The AM1-D and PM3-D results have been taken from Ref. 22.

No.	Molecule (symmetry)	Ref. Values	AM1	PM3	AM1-D	PM3-D	AM1-FS1
Hydrogen bonded complexes							
1	(NH3)2 (C2h)	2.504	2.784	3.241	2.646	2.726	2.668
2	(H2O)2 (Cs)	1.952	2.094	1.809	1.911	1.769	1.932
3	Formic acid dimer (C2h)	1.670	2.101	1.776	1.925	1.737	1.567
4	Formamide dimer (C2h)	1.841	2.072	1.807	1.981	1.763	1.916
5	Uracil dimer (C2h)	1.775	2.044	1.787	1.946	1.744	1.563
6	2-Pyridoxine2-aminopyridine (C1)	1.859, 1.874	2.511, 2.107	1.798, 1.815	1.980, 1.981	1.722,1.768	1.760, 1.878
7	Adenine thymine WC (C1)	1.819, 1.929	2.476, 2.101	1.780, 1.821	1.807, 2.018	1.708,1.769	1.597, 1.893
Complexes with predominant dispersion contribution							
8	(CH4)2 (D3d)	3.718	3.721	3.447	2.881	3.160	2.899
9	(C2H4)2 (D2d)	3.718	3.714	3.706	3.305	3.469	3.266
10	Benzene CH4 (C3)	3.716	3.746	3.718	3.315	3.450	3.457
11	Benzene dimer (C2h)	3.765	6.952	6.096	3.643	3.499	3.753
12	Pyrazine dimer (Cs)	3.479	4.848	4.760	3.695	3.437	3.681
13	Uracil dimer (C2)	3.166	5.805	6.732	3.097	3.406	3.007
14	Indole benzene (C1)	3.498	5.572	5.520	4.448	3.415	4.378
15	Adenine thymine stack (C1)	3.172	6.202	5.788	4.320	3.280	3.099
Mixed complexes							
16	Ethene ethyne (C2v)	2.752	2.468	2.429	2.319	2.366	2.374
17	Benzene H2O (Cs)	2.531	4.020	3.746	2.986	2.982	2.988
18	Benzene NH3 (Cs)	3.592	4.092	4.025	2.995	3.069	3.014
19	Benzene HCN (Cs)	3.387	3.472	3.694	3.228	3.343	3.303
20	Benzene dimer (C2v)	3.513	5.225	3.606	3.253	3.370	3.351
21	Indole benzene T-shape (C1)	3.210	3.811	3.807	3.010	3.233	3.208
22	Phenol dimer (C1)	1.937, 4.921	2.174, 5.925	1.829, 5.712	2.001, 5.040	1.778, 5.265	2.016, 4.937
	RMSE (Hydrogen bonded)		0.387	0.257	0.137	0.134	0.129
	RMSE (Dispersion bonded)		2.015	1.962	0.644	0.272	0.473
	RMSE (Mixed bonded)		0.929	0.598	0.336	0.315	0.301
	RMSE		1.277	1.171	0.419	0.249	0.326
	MUE		0.853	0.691	0.301	0.199	0.222

To further test the ability of AM1-FS1 to model H-bonding complexes, 16 additional hydrogen bonded DNA base pairs have been considered. The 16 additional complexes were chosen from Ref. 45 since these are the only complexes from the H-bonding subsection that have CCSD(T) quality binding energies. The geometries of these complexes, however, are from MP2 optimizations or experimental data. Therefore, these structures do not correspond to the CCSD(T) potential minimum, this is also the case for most of the S22 database structures. We have computed the binding energies for these complexes based on the reference geometries and also AM1-FS1 optimized geometries. The RMSE for the binding energies are 1.78 and 2.18 kcal/mol respectively. This error is consistent with the error associated with the hydrogen bonding complexes in

the S22 database, which were used for parameterization. The 16 complexes as well as the reference CCSD(T), single point and optimized AM1-FS1 binding energies are reported in Table E-2 of the supplementary material.

3.3.4 Potential Energy Curves

The value of a computational method is significantly enhanced if it is able to accurately describe the potential energy surface apart from the minimum. A given method could accurately predict the interaction energy at a specific molecular geometry, yet yield a very inaccurate picture of the remainder of the potential energy surface. (See Ref. 53 for a detailed discussion.) In this section, potential energy curves will be compared for various empirically corrected SE methods.

In Figure 3.3, potential energy curves for two different benzene dimer conformations are shown. Figure 3.3A, shows the interaction energy for the parallel dimer as a function of monomer separation. The parallel dimer is not the lowest energy conformation, but it is important to be able to model a variety of geometries correctly for the correct description of π-π interactions involved in large systems, and the parallel dimer represents a widely used test case, probably owing to the simplicity of its construction. The M1 benzene dimer, according to Ref. 59, is the lowest energy structure known. Comparing to the CCSD(T) and DFT(SAPT) reference values, among the empirically-corrected SE methods PM6-DH performs the worst for these systems. The PM6-DH method seems to over bind π-π interactions. This is due to the fact that the PM6 method performs poorly for dispersion bound complexes compared to DFT-BLYP results as discussed earlier and shown in Figure 3.2 (further discussion is presented in Ref. 53). McNamara and Hillier's AM1-D method performs very well for the M1 dimer,

however, not so well for the parallel dimer. On average AM1-FS1 performs the best for these two systems. AM1-FS1 has a very steep potential wall at close separation (Figure 3.3A), this is an artifact of using a large damping parameter (*d*) in the dispersion correction term (Eq. 3.2). Again, the large term is required because of the inability of the AM1 method to properly capture short-range repulsive interactions.

A. B.

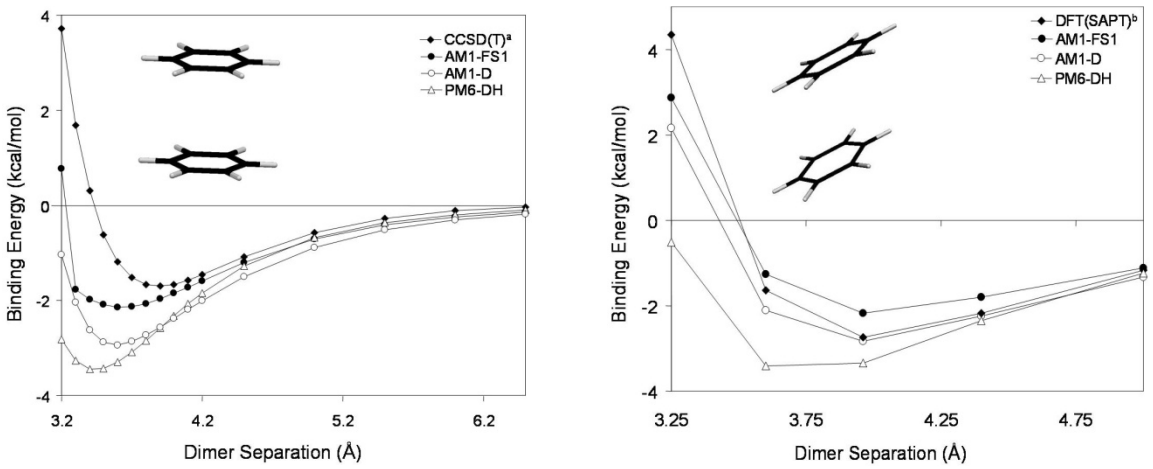


Figure 3.3: Parallel (Fig. A) and M1 (Fig. B) benzene dimer potential energy curves determined with various computational methods. [a] Ref. 51. [b] Ref. 59.

In Figure 3.4 and Figure 3.5, potential energy curves for the water dimer and the nitromethane dimer are shown respectively. The water dimer is a classic hydrogen bonding system. The potential energy curves in Figure 3.4 are shown as a function of O---O separation. The figure shows that AM1-FS1 dramatically improves upon the AM1 method, and outperforms McNamara and Hillier's AM1-D method. The correlation to

SCF-SAPT results[61] again shows that the hydrogen bonding correction term (Eq. 3.4) is a worthwhile addition to the AM1 method. The AM1-D method also performs relatively well for the water dimer. This means that the changes they have made to the AM1 parameters improve the results for this particular system; however, the same is not observed if we consider the nitromethane dimer. In Figure 3.5, we see that AM1-D performs poorly for the nitromethane dimer. The SCF-SAPT curve[58] is for the so-called "double hydrogen bond" configuration; however, nitromethane is not a classical H-bonding system. It lacks a hydrogen atom attached to a highly electronegative atom (N, O, or F), nevertheless, this system is said to form weak H-bonds[58]. As shown (Figure 3.5) the AM1 method performs relatively well for this system; whereas McNamara and Hiller's modification of the AM1 parameters has caused the AM1-D method to inadequately model this system. AM1-FS1, on the other hand, does not consider this a H-bonding case. Therefore, the H-bonding correction term is not turned on for this system. Consequently, AM1-FS1 performs well for this system by applying only the dispersion correction. This potential energy curve demonstrates that the AM1 parameters should not be changed in all cases. It should be noted that the AM1-D training set does not contain this system whereas the training set for AM1-FS1 does. (We have not compared the PM6-DH method of Řezáč et al.[23] in H-bonding cases because we do not have code for their elaborate H-bonding correction scheme.)

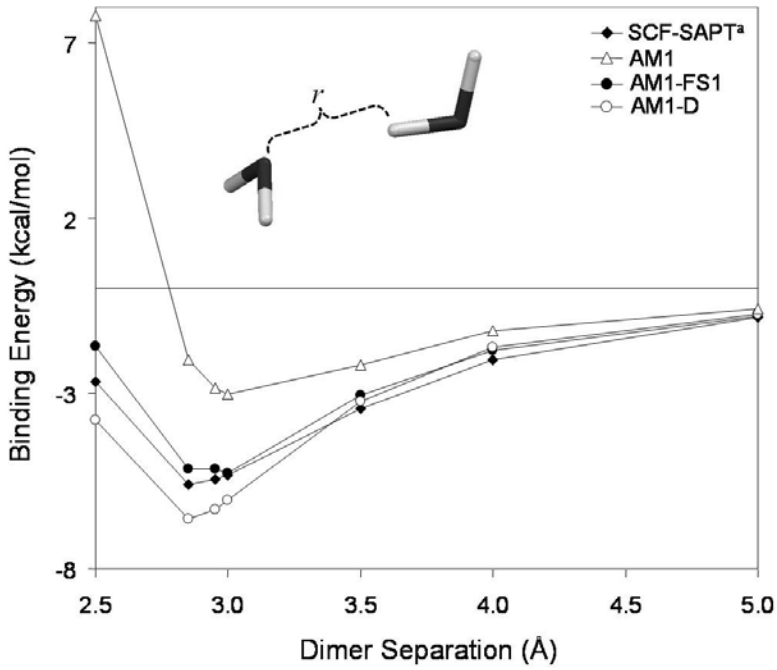


Figure 3.4: Water Dimer potential energy curves as a function of O---O separation, as determined with various computational methods. [a] Ref. 61.

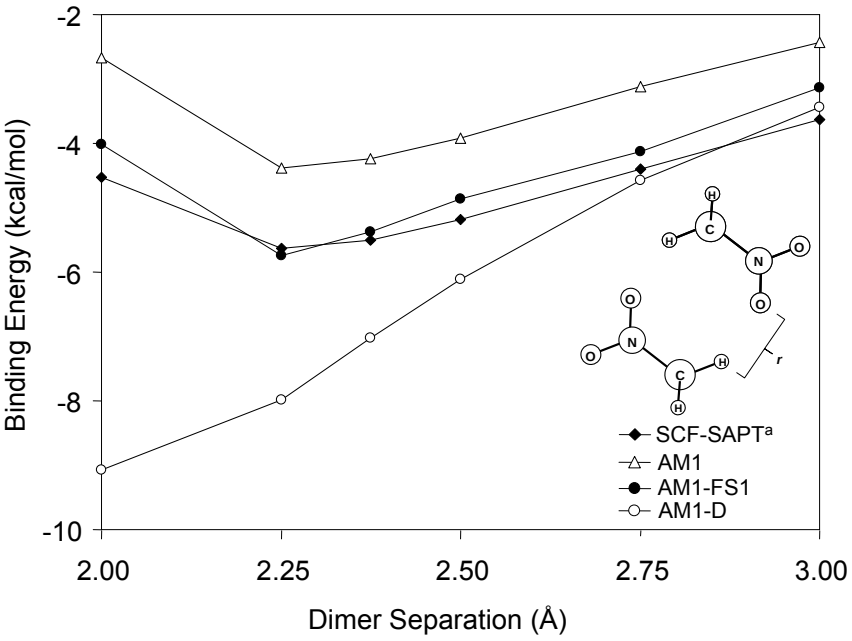


Figure 3.5: Nitromethane dimer potential energy curve in the "double hydrogen bond" configuration, as determined with various computational methods. [a] Ref. 58.

3.3.5 Heat of Formation

As mentioned earlier, modifying the original semi-empirical parameters can have deleterious effects, especially for thermodynamic properties. For example, the experimental heat of formation of benzene is 19.8 kcal/mol[62] and is predicted to be 22.0 kcal/mol by the AM1 method[18]. The AM1-D method, however, predicts a value of -12.9 kcal/mol. (PM3-D performs even more poorly, yielding -21.8 kcal/mol.) Re-parameterization has rendered AM1-D (and PM3-D) unreliable for predicting thermodynamic properties. On the other hand, AM1-FS1 does not change any of the original AM1 parameters and predicts the heat of formation of benzene to be 20.0 kcal/mol; in good agreement with experiment and, serendipitously, even a slight improvement over AM1. The AM1-FS1 empirical correction is designed to have little effect on quantities that are already predicted relatively well by the AM1 method. Table 3.4 collects results for calculations of heat of formation on 53 test molecules. Note that the RMSE in predictions of heat of formation with AM1-FS1 is comparable to that of the original AM1 method, but AM1-D is 24 times (2400%) less accurate. Reparameterization of the original PM3 method in the development of PM3-D, has also seriously degraded its predictive power for heats of formation (see Table 3.4). This clearly shows the negative consequences of changing the original semi-empirical parameters without the consideration of such quantities during optimization of the parameters.

While the original AM1 parameters have not been altered in AM1-FS1, the FS1 correction terms do influence the predicted heat of formation. This occurs because the heat of formation is in part determined from the total energy of the complex, which now contains the empirical correction energy, but also in part from the energies of the isolated

atoms. The isolated atom energies do not include any empirical correction energy since, by design, the correction terms are not implemented for a single atom. Therefore, the difference between the heat of formation as computed with AM1 and AM1-FS1 will generally become larger as the correction term(s) contribution increases. This will also be the case for DFT-D methods when the total energy is used in the determination of the heat of formation. The influence of the FS1 correction on the predicted heat of formation can have both undesirable and desirable consequences. For example, in Table 3.4, it can be seen that as the number of methylene units in the aliphatic hydrocarbons increases (methane → ethane → propane → etc.) the error in the predicted heat of formation increases. Fortunately, since the original AM1 parameters have not been altered, the AM1 heat of formation can be easily obtained by subtracting out the empirical correction energy from the AM1-FS1 heat of formation. This approach of subtracting the correction energy will also be effective for DFT-D methods when the total energy is used in the determination of the heat of formation. On the other hand, the correction to the total energy sometimes has a beneficial impact on the predicted heat of formation. For example, the experimental heat of formation of the benzene dimer is 30.4 kcal/mol[63] and is predicted to be 37.9 kcal/mol and 44.1 kcal/mol based upon structures optimized with the AM1-FS1 and AM1 methods, respectively. Note that these structures are significantly different upon optimization because the AM1 method does not consider dispersion interactions. If we determine the heat of formation of the AM1-FS1 optimized geometry with the AM1 method the error is even larger; the heat of formation is predicted to be 46.7 kcal/mol. Since the original AM1 parameters were optimized to give reliable predictions of the heat of formation at AM1-optimized geometries, it seems reasonable to

conclude that, in general, if the heat of formation of some large multicomponent carbon structure is desired, the AM1-FS1 method will likely produce a more accurate result at the AM1-FS1 geometry. Certainly, the AM1-FS1 structure will be more accurate since the dominate intercomponent interaction will be incorporated. This situation is much less complicated when a heat of reaction is of interest since the correction term(s) are applied to both the reactants and products.

Table 3.4: Heat of formation (kcal/mol). The experimental and AM1 results were obtained from Ref. 18. Note: the AM1-D and PM3-D results were obtained by coding the method outlined in Ref. 22; however, slight disagreements in the binding energies were observed with PM3-D for compounds containing oxygen, suggesting a misprint in the published PM3-D oxygen parameters.

	Heat of Formation (kcal/mol)				
Molecule	expt	AM1	AM1-D	PM3-D	AM1-FS1
methane	-17.8	-8.8	-78.4	-6.6	-8.8
ethane	-20.0	-17.4	-114.1	-12.2	-18.3
ethylene	12.5	16.5	-39.4	-11.0	16.1
acetylene	54.5	54.8	38.6	-9.8	54.8
propane	-25.0	-24.3	-148.4	-17.7	-27.1
propene	4.8	6.6	-75.6	-16.5	5.3
propyne	44.2	43.4	3.4	-15.3	43.0
allene	45.5	46.1	6.6	-15.3	45.7
n-butane	-30.0	-31.1	-182.7	-23.2	-36.1
isobutane	-32.0	-29.4	-181.1	-23.2	-35.3
but-1-ene	-0.1	0.4	-109.2	-22.0	-2.6
trans-2-butene	-2.8	-3.3	-111.8	-22.0	-5.6
cis-2-butene	-1.7	-2.2	-110.9	-22.0	-4.8
isobutene	-4.0	-1.2	-109.8	-22.0	-4.3
1,2-butadiene	38.8	37.1	-28.5	-20.8	35.9
trans-1,3-butadiene	26.3	29.9	-38.2	-20.8	28.3
1-butyne	39.5	37.5	-29.8	-20.8	35.6
2-butyne	34.8	32.0	-31.8	-20.8	31.1
vinylacetylene	72.8	67.9	42.1	-19.6	66.9
diacetylene	113.0	106.1	122.4	-18.4	105.8
n-pentane	-35.1	-37.9	-216.9	-28.7	-45.1
neopentane	-40.2	-32.8	-212.3	-28.7	-42.8
benzene	19.8	22.0	-12.9	-29.6	20.0
toluene	12.0	14.5	-46.6	-35.1	10.7
ammonia	-11.0	-7.3	-154.0	-7.7	-7.3
methylamine	-5.5	-7.4	-165.2	-13.3	-8.2
dimethylamine	-4.4	-5.6	-175.1	-18.8	-7.5
trimethylamine	-5.7	-1.7	-183.3	-24.3	-5.1
ethylamine	-11.3	-15.1	-200.4	-18.8	-17.5
n-propylamine	-16.8	-22.1	-234.7	-24.3	-26.5
isopropylamine	-20.0	-19.2	-231.9	-24.3	-23.9
tert-butylamine	-28.9	-21.2	-261.6	-29.8	-29.3
pyrrole	25.9	39.9	-56.0	-26.3	38.5
pyridine	34.6	32.1	-29.5	-30.7	30.6
pyridazine	66.5	55.3	-33.6	-31.8	54.2
water	-57.8	-59.2	-200.8	-12.0	-59.2
methanol	-48.2	-57.0	-193.7	-17.5	-57.5
ethanol	-56.2	-62.7	-225.8	-23.0	-64.4
1-propanol	-61.0	-70.6	-261.9	-28.5	-74.7
2-propanol	-65.2	-67.7	-258.2	-28.5	-71.8
t-butyl_alcohol	-74.7	-71.6	-288.7	-34.0	-78.6
dimethyl_ether	-44.0	-53.2	-185.7	-23.0	-53.8
diethyl_ether	-60.3	-64.4	-249.6	-34.0	-67.8
oxirane	-12.6	-8.9	-95.3	-21.8	-9.2
furan	-8.3	3.0	-52.7	-30.5	2.1
phenol	-23.0	-22.2	-120.0	-40.4	-24.9
anisole	-16.2	-15.8	-110.1	-45.9	-19.8
benzaldehyde	-8.8	-8.9	-58.5	-44.8	-12.1
formic_acid	-90.5	-97.4	-222.7	-27.2	-97.4
acetic_acid	-103.4	-103.0	-252.9	-32.7	-103.7
propionic_acid	-108.4	-108.0	-285.5	-38.2	-111.1
oxalic_acid	-173.0	-172.4	-370.9	-53.2	-172.8
benzoic_acid	-70.3	-68.0	-181.4	-55.6	-71.0
RMSE		4.9	132.5	46.7	5.6
MUE		3.6	118.4	36.1	4.3

3.4 Application to Macromolecular Complexes

The ultimate goal of AM1-FS1 is to be able to efficiently and accurately model large weakly-bound systems, such as complexes of carbon nanostructures and molecular devices. Such large systems are currently out of reach for CCSD(T), and extremely computationally demanding for DFT methods. Furthermore, most DFT functionals are incapable of modeling carbon nanostructure complexes due to the fact that these systems are governed by van der Waals interactions. Presumably, the current most accurate methods capable of modeling such systems are DFT-D methods and the M05[25], M06[26], and M08[64] family of functionals developed by Zhao and Truhlar. Performing geometry optimizations with DFT based methods on systems larger than 100 atoms is currently extremely computationally expensive, but such optimizations can be routinely performed with semiempirical based techniques even on "PC"-class computers. This great computational efficiency of semiempirical methods provides the central motivation for the present work.

3.4.1 Carbon Nanostructures

To test the performance of AM1-FS1 on complexes of carbon nanostructures, we have performed geometry optimizations and determined the binding energy of several inclusion complexes. The hosts considered are corannulene ($C_{20}H_{10}$), a double-concave hydrocarbon buckycatcher[65] ($C_{60}H_{28}$), and cyclic[6]paraphenylacetylene (6CPPA). The AM1-FS1 optimized structures are shown in Figure 3.6. (It is important to note that these complexes would be predicted to be unbound if the standard AM1 and most DFT methods were used.) To-date, the best binding energy values for these complexes are from DFT calculations reported by Zhao and Truhlar[66, 67], using the M06-2X functional.

The binding energies along with the AM1-FS1 and AM1-D results are reported in Table 3.5. The AM1-FS1 results are very comparable to the M06 values, however, AM1-FS1 overestimates the binding energy for 3,3@[6]CPPA and 4,4@[6]CPPA in comparison to DFT-M06-2X. This may be a result of the M06 functional underestimating dispersion interactions at long-range. This hypothesis is supported by the potential energy curve for the parallel benzene dimer shown in Figure 3.7, which clearly shows that the M06 functional fails to accurately model dispersion interactions in the long-range regime, where the predicted interaction energy even becomes slightly positive. This behavior might be easily overlooked since upon optimization of the parallel benzene dimer, a reasonable energy and structure will be produced. To show that this is the case for 3,3@[6]CPPA, the binding energy was determined using the BLYP-D functional. The resulting binding energy of 18.9 kcal/mol is in very good agreement with the AM1-FS1 result. The M06-2X functional underestimates binding for 3,3@[6]CPPA because the nearest intermolecular interaction is 4.5 Å, a distance at which M06 underestimates the interaction energy as exhibited by the benzene dimer potential energy curve.

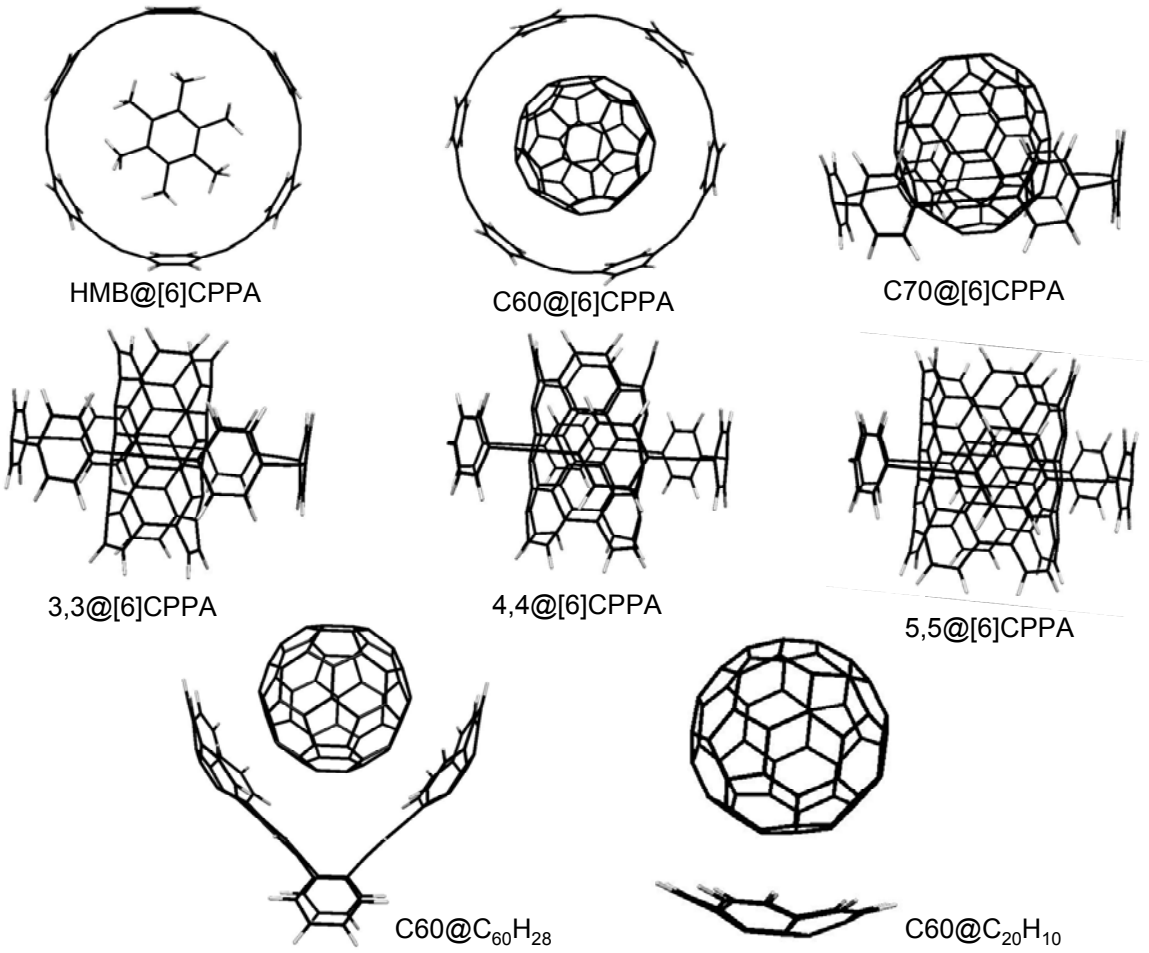


Figure 3.6: AM1-FS1 geometry optimized carbon nanostructure complexes.

The AM1-FS1 method significantly outperforms AM1-D in every case based on the current benchmark M06 values. AM1-FS1 achieves this correlation with only two added parameters (due to the nature of the systems only the dispersion correction term is "turned on" during the AM1-FS1 calculations), whereas, AM1-D utilizes 10 parameters. We credit the success of AM1-FS1 to parameterizing to a larger training set containing nonequilibrium complexes.

Table 3.5: Binding energies (kcal/mol) of carbon nanostructure complexes. The M06 results were obtained from Ref. 66. [a] Results were obtained from Ref. 67.

	AM1-FS1	AM1-D	M06-2X/6-31+G(d,p)//M06-L/MIDI!
HMB@6CPPA	-16.6	-17.6	-14.7
C60@6CPPA	-26.9	-30.1	-28.0
C70@6CPPA	-36.3	-41.0	-31.1
3,3@6CPPA	-17.7	-22.1	-5.4
4,4@6CPPA	-32.7	-42.0	-24.0
5,5@6CPPA	-43.2	-46.4	-43.3
C60@BuckyCatcher[a]	-29.3	-36.8	-26.4
C60@Coroanulene[a]	-13.4	-16.7	-12.4

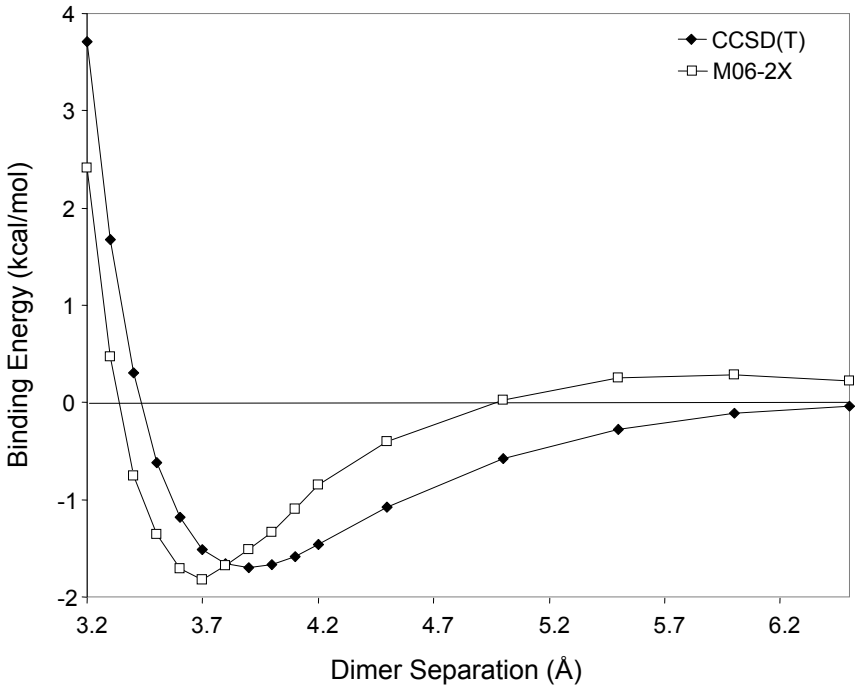


Figure 3.7: Parallel benzene dimer potential energy curve calculated with the M06-2X functional, using the 6-311G(d,p) basis. CCSD(T) results were obtained from Ref. 51.

3.4.2 Pseudorotaxanes

We also tested the performance of AM1-FS1 on six different pseudorotaxanes, since these types of complexes are of central interest to our research group. All of the systems considered incorporate cyclobis(paraquat-p-phenylene) ($CBPQT^{+4}$), a tetracationic ring structure. Six inclusion complexes with this ring have been formed with dimethoxybenzene and benzenedimethanamine in the *ortho*, *meta*, and *para* conformations. (AM1-FS1 optimized structures are shown in Figure 3.8.) We have performed geometry optimizations and determined the binding energies of these complexes and compared them to previously reported LMP2/6-311+G(d,p)//BHandHLYP/6-31G(d) results[68]. We also computed these binding energies at the M06-2X/6-311G(d,p)//M06-L/MIDI level of theory for additional comparisons. (All results are reported in Table 3.6.) Based on the results, the LMP2/6-311+G(d,p)//BHandHLYP/6-31G(d) results appear to underestimate the binding energy. This is likely a result of the geometry produced by the BHandHLYP functional and not the LMP2 method. This conclusion is based on the binding energies determined at the M06-2X/6-311G(d,p)//M06-L/MIDI level. The differences in binding energies between the isomers are sufficiently small that they may be taken to be insignificant given the level of theory and the large conformational space associated with these complexes. Based on these results, we believe AM1-FS1 is a valuable tool for modeling this class of macromolecular complexes.

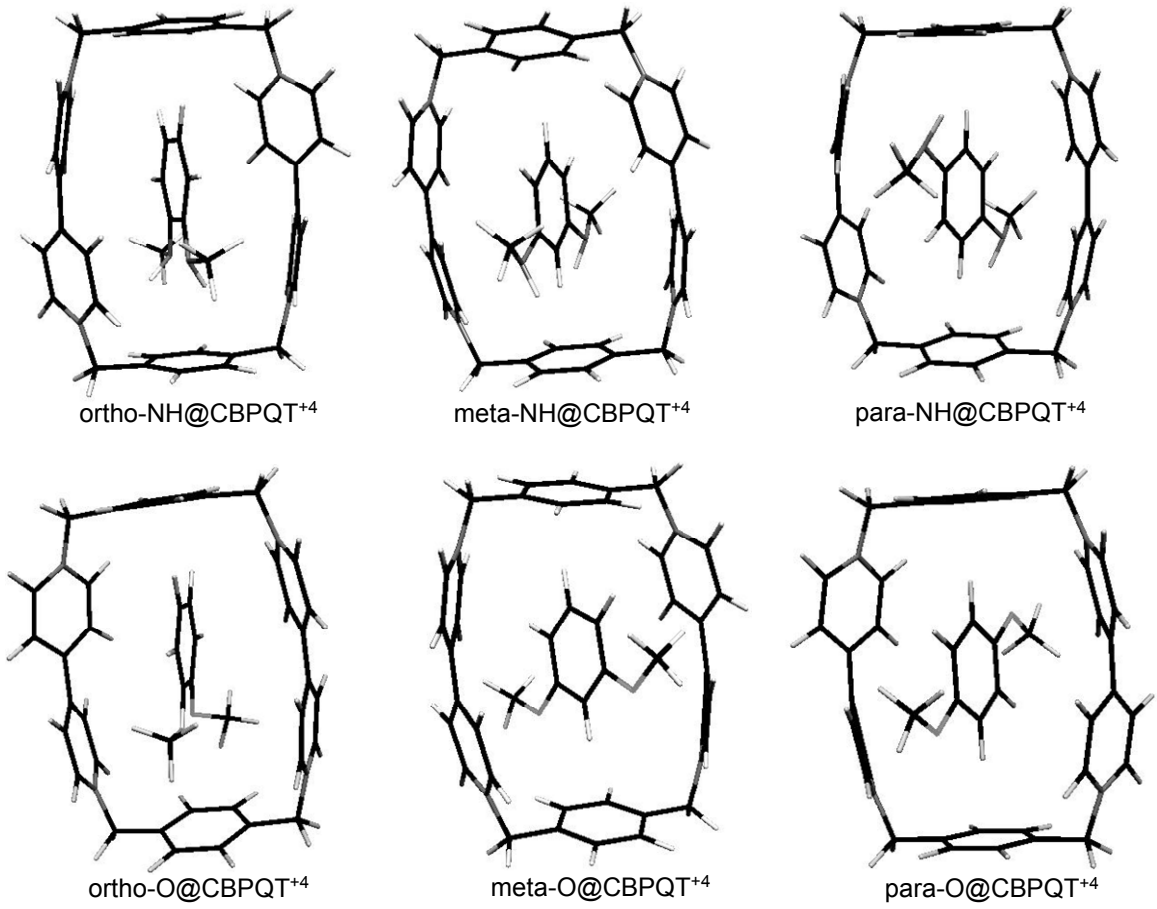


Figure 3.8: AM1-FS1 geometry optimized pseudorotaxane complexes.

Table 3.6: Binding energies (kcal/mol) of pseudorotaxane complexes. The LMP2 results were obtained from Ref. 68.

	Binding Energy (kcal/mol)		
	AM1-FS1	M06-2X/6-311G(d,p)// M06-L/MIDI	LMP2/6-311+G(d,p)// BHandHLYP/6-31G(d)
ortho-O@CBPQT^{+4}	-32.1	-34.7	-21.0
meta-O@CBPQT^{+4}	-31.7	-35.0	-16.1
para-O@CBPQT^{+4}	-33.4	-34.7	-21.3
ortho-NH@CBPQT^{+4}	-37.7	-41.7	-22.3
meta-NH@CBPQT^{+4}	-38.2	-36.9	-22.5
para-NH@CBPQT^{+4}	-38.5	-40.1	-23.9

3.5 Conclusions

AM1-FS1 is a new empirically-corrected semi-empirical method suitable for performing geometry optimizations on macromolecular complexes. AM1-FS1 displays considerable improvement over the traditional AM1 method for non-bonding interactions, yet it retains the computational efficiency and predictive power for thermo-chemical quantities of the original AM1 Hamiltonian. Validation testing shows that the method reduces the RMSE for the popular S22 database from 8.47 to 1.18 kcal/mol. More impressively, this new method has achieved kilocalorie accuracy on a training set of 66 complexes. This was accomplished with just 6 empirical parameters (2 for dispersion and 4 for hydrogen-bonding) and *no* reparameterization of AM1, (which we show here has led to serious consequences in existing empirically-corrected SE methods). This is a dramatic reduction in the total number of adjustable parameters compared to other previously published empirically corrected SE methods. Validation testing showed that the existing PM6-DH method outperforms AM1-FS1 based on the S22 database; however, the PM6-DH method has shown to be inaccurate for reproducing potential energy curves for the benzene dimer which is a classic test case for demonstrating the ability of a method to model dispersion interactions at various distances. This inability would result in inaccurate modeling of large dispersion bound systems like carbon nanostructures, due to the large number of interacting atoms at various distances. Moreover, unlike PM6-DH, AM1-FS1 does *not* require knowledge of atom connectivity. Based on the examples reported, on average AM1-FS1 is also the most reliable empirically corrected SE method for reproducing the potential energy curve away from

the global minimum. We credit this success to using a training set that contains non-equilibrium complexes.

This new AM1-FS1 method has shown to yield results comparable in accuracy to the best available calculations on complexes of carbon nanostructures and pseudorotaxanes. We believe AM1-FS1 is a useful computational tool for obtaining reliable results for such systems at limited computational expense. It should prove to be a valuable asset for routine modeling of macromolecular complexes that are currently at (or beyond) the limit of DFT based techniques, or out of reach of higher levels of theory.

Chapter 4: Designing Fullerene Separation Materials: A Theoretical Study

4.1 Introduction

The availability of reliable and efficient theoretical and computational schemes that could serve as design tools for nano-devices could revolutionize the process by which nanoscale devices are developed. Currently, the development process is closer akin to fundamental research than to design engineering. Synthetic chemists and materials scientists identify target materials based on chemical insights and experience concerning the effects of various structures and functional groups on molecular properties. Syntheses are then carried out and the targeted products tested. If the properties of the product don't match expectations, the findings become part of the knowledge base upon which new target materials are selected. By contrast, macro-scale device design makes extensive use of computational testing of hypothetical designs before prototypes are fabricated. The ultimate commercialization of nanotechnologies would benefit tremendously from similar computational design tools for nanoscale devices. Obviously the starting point for the design of nano-devices is quantum chemistry, as embodied in the time independent Schrödinger equation. Many robust codes for application of computational quantum chemistry are available, but the large size of nano-device systems makes identifying a technique that yields adequate reliability at a reasonable computational expense a great challenge.

Polycyclic aromatic hydrocarbons (PAHs) have been receiving a great deal of attention for possible use as the building blocks of nano-devices. PAHs are relatively

easy to synthesize from inexpensive starting materials and can form nano-particles in a stunning variety of shapes and sizes[69]. Among the PAH nanostructures, fullerenes are very promising as potential components for the development of nano-devices and nano-machines[70, 71]. One of the current hurdles to developing such devices is obtaining components/molecules of the same size in sufficient quantity. Typically, when fullerenes are synthesized, a distribution of sizes is generated. Therefore, an efficient separation technique is needed. A computational model that could identify suitable separation materials would be of considerable importance. Herein, we identify such a computational model and apply it to investigate fullerene complexation.

Since PAH nanostructures are composed solely of carbon and hydrogen atoms, the governing inter-molecular interactions are van der Waals-London (dispersion) forces. Studying systems in which the inter-molecular interactions are dominated by dispersion presents significant computational challenges. CCSD(T) and SAPT(DFT) methods accurately consider inter-component dispersion interactions, but are computationally extremely expensive, which renders calculations on even one fullerene intractable. In many of the more commonly used quantum mechanical methods, such as Hartree-Fock (HF) and most of the popular Density Function Theory (DFT) based techniques, dispersion interactions are neglected. Therefore, if optimizations are carried out on fullerene/PAH complexes with any of these common methods, the components will likely dissociate because the main attractive interaction, dispersion, is neglected in the calculations. Dispersion corrected DFT (DFT-D) methods and DFT functionals such as the M06[26] and M08[64] are capable of modeling inter-component dispersion interactions; however, modeling systems of the desired size can be extremely computationally

demanding even at the DFT level of theory. There have been several studies[72, 73] using the MPWB1K functional[74] to model the encapsulation of metallic and nonmetallic species into fullerenes; however, these systems are significantly smaller and, due to the nature of the systems, cannot spontaneously dissociate upon optimization even if dispersion interactions are not considered. More importantly, in order to get a good description of the potential energy surface a large degree of conformational space needs to be explored. This is where DFT based techniques become computationally inadequate from an efficiency standpoint. By contrast, semiempirical molecular orbital techniques (such as AM1 or PM6) are sufficiently efficient to model even large PAH nanostructures, but these techniques are based on HF theory and therefore cannot reliably model dispersion-bound complexes. In order to model the desired PAH nanostructures complexes, an empirically corrected semi-empirical method offers both the requisite efficiency and accuracy. Here we apply an empirically-corrected form of the AM1 Hamiltonian, termed AM1-FS1[75], to study the complexation of fullerenes by PAH macrocycles. In the present study many hundreds of thousands of SCF-calculations were required; a prohibitively expensive task for DFT based techniques. The AM1-FS1 results yield insight into the challenges involved in designing a selective fullerene separation system.

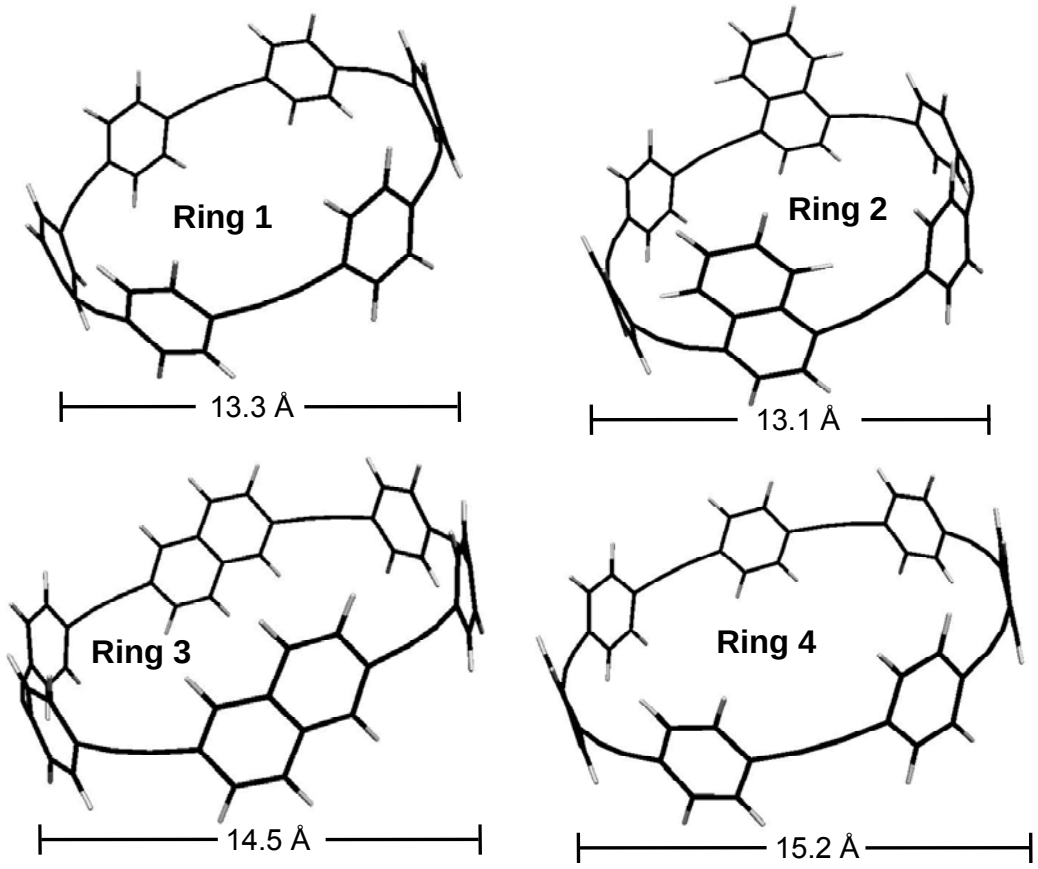


Figure 4.1: PAH nano-rings used in this investigation.

4.2 Theory

4.2.1 Structures

One avenue to fullerene separation that has seen considerable attention is complexation with PAH nano-rings[76-80] such as the [n]paraphenyleneacetylenes. In this study, four different carbon nano-ring structures, shown in Figure 4.1, were considered for their potential to aid in fullerene separation. The interactions of these four ring systems with eight different fullerenes ranging from C_{20} to C_{180} are considered here; the fullerenes are shown in Figure 4.2. These types of inclusion complexes have proven to

present a significant challenge to theoretical modeling[81]. Starting structures for the PAH nano-rings were constructed using a graphical-user-interface molecular editor[82]. Starting structures for the fullerenes were obtained from the Fullerene Structure library[83], except for the C_{180} capped nanotube structures ($C_{180\text{-}CNT}$) which was constructed using Nanotube Modeler[84]. The different conformations of the fullerene/capped nanotubes were chosen for their spherical or prolate geometry; therefore, the conformations may not be the lowest energy structures for the given stoichiometries. The ring/fullerene complexes were assembled manually with a graphical-user-interface molecular editor[82]. The host/guest systems considered contained between 92 and 264 atoms.

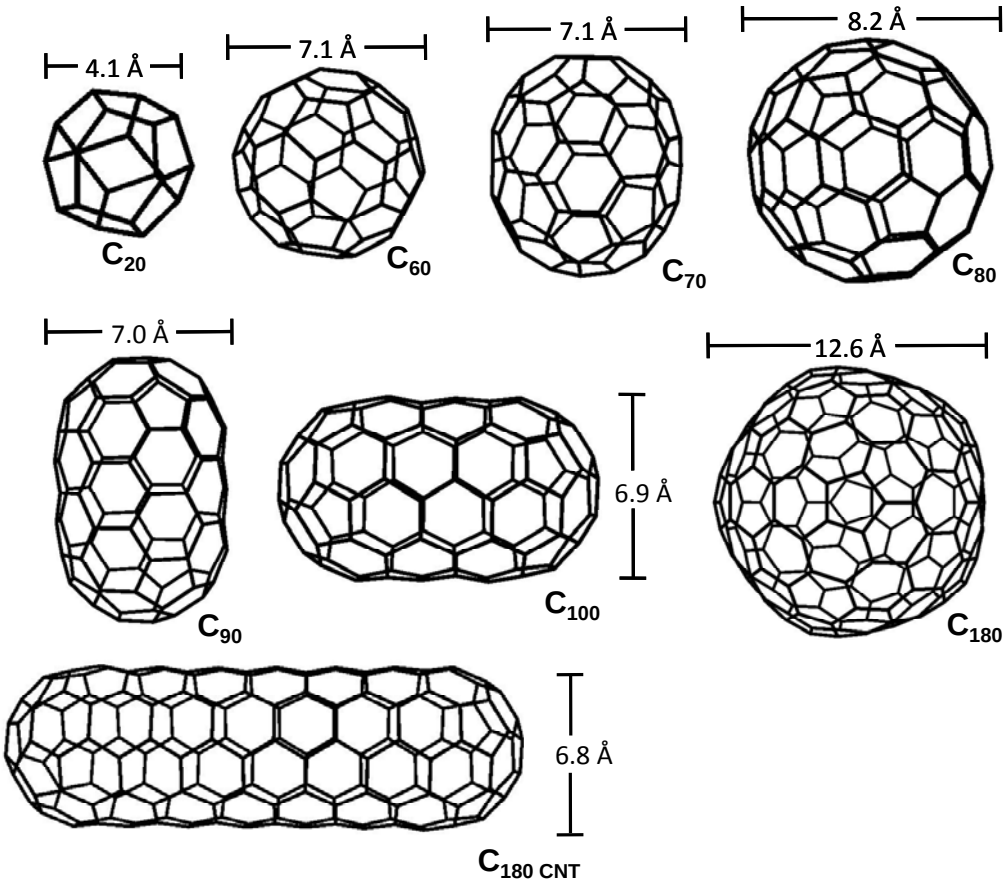


Figure 4.2: Fullerenes used in this investigation.

4.2.2 Modeling

The ring/fullerene complexes studied here are non-bonded complexes held together by weak intermolecular interactions. The dominant non-bonding interactions involved are dispersion interactions. To study the complexation, the AM1-FS1 method[75] has been employed, which is a augmented version of the AM1 method incorporating dispersion and hydrogen-bonding correction terms. This method has been chosen for its previous success in modeling carbon nanostructure complexes[75]. While the AM1 method is unreliable for PAH complexes for structures, binding energies, and heat of formation[85], the FS1 correction greatly improves the accuracy in the prediction of structures and binding energies, and show slight improvements for heat of formations.

The AM1-FS1 method employs the empirical dispersion correction of Grimme[35, 86] with a slight modification suggested by Jurečka et al.[36] The resulting form is:

$$E_{dis} = -\frac{C_6^{ij}}{r^6} f_{damp}\left(r_{ij}\right), \tag{4.1}$$

where r_{ij} is the atom-atom separation, C_6^{ij} is the dispersion coefficient, and f_{damp} is a damping function of the form,

$$f_{damp}\left(r_{ij}\right) = \frac{1}{1 + e^{-d\left(\frac{r_{ij}}{S_R R_{vdw}} - 1\right)}}. \tag{4.2}$$

This damping function depends on the equilibrium van der Waals separation (R_{vdw}) and the pair-wise atom separation (r_{ij}). The damping function also depends on the empirical parameters S_R and d. To determine C_6^{ij} and R_{vdw} we employ the geometric mean and simple average combination rules as follows:

$$C_6^{ij} = \sqrt{C_6^i C_6^j} \quad , \quad R_{vdw} = \frac{R_{ii} + R_{jj}}{2} \quad . \tag{4.3}$$

The dispersion coefficients (C_6^i and C_6^j) and equilibrium van der Waals radii for two identical atoms (R_{ii} and R_{jj}) for the different atoms were obtained from Grimme[35, 86].

The two empirical parameters S_R and d in the dispersion correction term were determined by fitting to a database of intermolecular interaction energies for 66 weakly-bound complexes that contains both equilibrium (potential minima structures) and non-equilibrium structures (see Chapter 3 for parameterization details). It is believed that this parameterization scheme makes AM1-FS1 more suitable for modeling these PAH nanostructures than other current empirically corrected SE methods.

The systems involved in the study only contain H and C atoms; therefore, only the dispersion correction is utilized in the AM1-FS1 method. For details involving the hydrogen-bonding correction term refer to Foster and Sohlberg 2010.

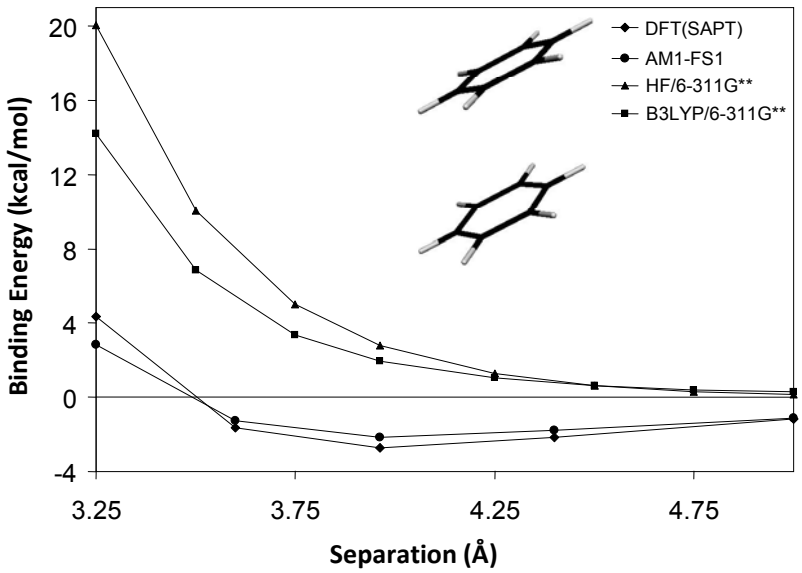


Figure 4.3: Potential energy curves for the *M1* benzene dimer computed at different levels of theory. The *M1* structure and DFT(SAPT) energies are reported in Ref. 59.

Table 4.1: AM1-FS1 and reported CCSD(T) binding energies (kcal/mol) for different conformations of the naphthalene dimer. The different dimer configurations and CCSD(T) energies are reported in Ref. 87.

Dimer	R1	R or R2	CCSD(T)	AM1-FS1
A		3.6	-3.15	-4.46
		3.8	-3.77	-4.24
		4.0	-3.60	-3.78
B		3.4	-4.77	-4.85
		3.6	-5.28	-4.74
		3.8	-4.91	-4.35
C	1.4	3.5	-5.32	-4.52
D	1.6	3.5	-5.32	-4.55
E	1.4	1.0	-5.73	-4.53
F		5.0	-4.34	-6.01
		5.2	-4.33	-4.55
G		5.2	-3.09	-4.40
		5.4	-3.07	-3.35
			RMSE	0.87

4.2.3 Validation

As a demonstration of the reliability of FS1-corrected AM1 for modeling PAH nanostructures, calculations were performed on the benzene and naphthalene dimers, and a coronene stack. Figure 4.3 shows interaction potential energy curves (PECs) for the

benzene dimer, in the *M1* conformation, as computed with various theoretical methods. Note that both HF-SCF and DFT-B3LYP predict the system to be unbound. The AM1-FS1 method yields an interaction potential curve in very good agreement with high-level calculations based on symmetry-adapted-perturbation-theory (DFT(SAPT))[59]. Table 4.1 compares the AM1-FS1 binding energy of the naphthalene dimer in several different conformations to published CCSD(T) results[87]. FS1-corrected-AM1 produces a sub kilocalorie root mean square error (RMSE) for the 13 different naphthalene structures considered. Again the AM1-FS1 method gives results in good agreement with the best first-principles calculations. A PEC for the coronene dimer in the *shifted graphite* conformation is shown in Figure 4.4. Again, it can be seen that the AM1-FS1 method is in good agreement with recent DFT(SAPT) results[88]. Note that both HF-SCF and DFT-B3LYP will (incorrectly) predict the structure to be un-bound, i.e. full geometry optimization will result in dissociation of the complex. This method has also been tested on a cluster composed of four perylenediimide (PTCDI) molecules. This system involves both dispersion and hydrogen bonding interactions; therefore, it utilizes both correction terms in the AM1-FS1 method. The optimized cluster, shown in Figure 4.5, is in good agreement with the experimentally determined structure[69]. Experimentally, the structure has a π-π stacking spacing (*d*-spacing in the longitudinal direction) of 3.6 Å and a 14 Å *d*-spacing in the perpendicular direction[69]. The FS1-corrected-AM1 method predicts a spacing of 3.5 Å and 14.5 Å respectively. This cluster was also optimized with the uncorrected and dispersion corrected (dispersion only) AM1 methods and in both cases the incorrect structure was predicted. In the AM1 case, the four monomers essentially merged into the same plane due to the inability of the AM1 method to model π-π stacking

(dispersion) interactions. (The HF as well as most DFT methods would likely result in similar unphysical structures.) When only the dispersion correction is added to the AM1 method, the structure maintains the π-π stacking configuration; however, the stacked pairs are shifted so that the four monomers are staggered. This shows the significance of including a hydrogen bonding correction term that depends on the XH---Y angle.

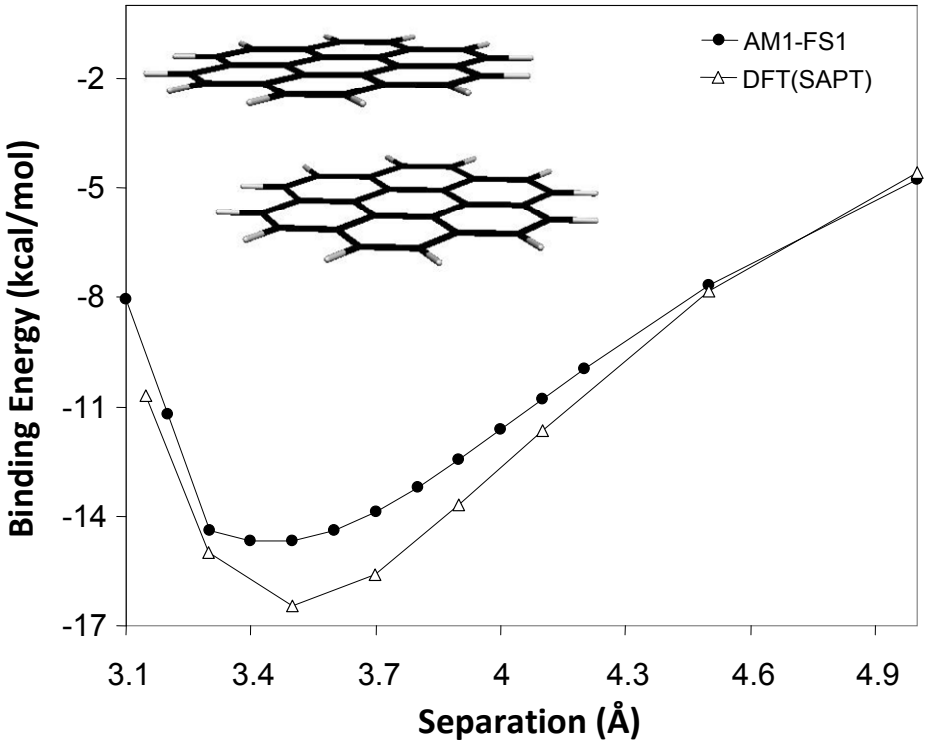


Figure 4.4: Potential energy curves for the *shifted graphite* coronene dimer. The structure and DFT(SAPT) energies are reported in Ref. 88.

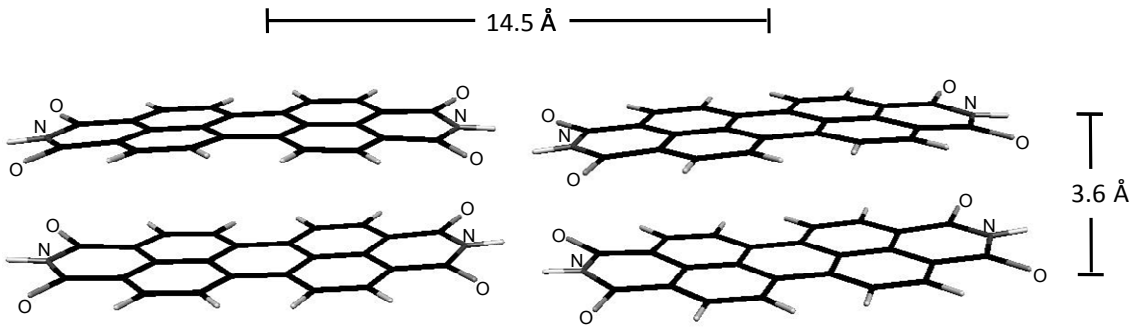


Figure 4.5: AM1-FS1 optimized perylenediimide (PTCDI) cluster. (Note, the hydrogen-bonding correction term is utilized).

It is important to note that in order to properly model large PAH nanostructures complexes, a method must be capable of accurately modeling dispersion interactions over a wide range of separations. These systems often contain a large number of mid-to-long range interactions; therefore, the mid-to-long range contributions have an appreciable effect on the interaction energy. Consequently, it is important that the computational method being used is cable of modeling the potential energy curve apart from the minimum energy geometry. We believe that the AM1-FS1 method is reasonably capable of modeling these types of system based on the potential energy curves provided as well as the other tests provided here and in Foster and Sohlberg 2010.

4.3 Results and Discussion

The computed AM1-FS1 binding energies for interactions of the eight fullerenes with the four rings are collected in Table 4.2. Literature values[66] computed with the DFT-M06 functional are available for comparison for ring1 interacting with C_{60} and C_{70}. These values are -28.0 and -31.1 kcal/mol respectively, in good agreement with the FS1-

corrected AM1 values. Note that in the prior first-principles investigation[66] of ring1 with C_{60}, certain symmetry elements were implicitly (or explicitly) assumed. We have therefore reported a calculation for the high symmetry case as well as for the fully optimized structure. The agreement of our empirically-corrected AM1 results with those of the best available calculations is generally excellent.

Table 4.2: Binding energies for fully optimized fullerene complexes in kcal/mole. The binding energy for the symmetry-constrained (C_{2h}) complex of ring1 with C_{60} is given in parenthesis.

	ring1	ring2	ring3	ring4
C_{20}	-11.3	-14.7	-11.0	-8.7
C_{60}	-33.5 (-26.9)	-37.3	-30.8	-24.3
C_{70}	-37.1	-40.7	-37.3	-32.1
C_{80}	-34.6	-39.9	-41.6	-36.9
C_{90}	-35.9	-44.4	-42.4	-32.8
C_{100}	-43.7	-50.0	-45.9	-42.3
C_{180}	-36.3	-42.1	-40.9	-42.0
$C_{180\text{-CNT}}$	-45.2	-52.2	-45.7	-34.9

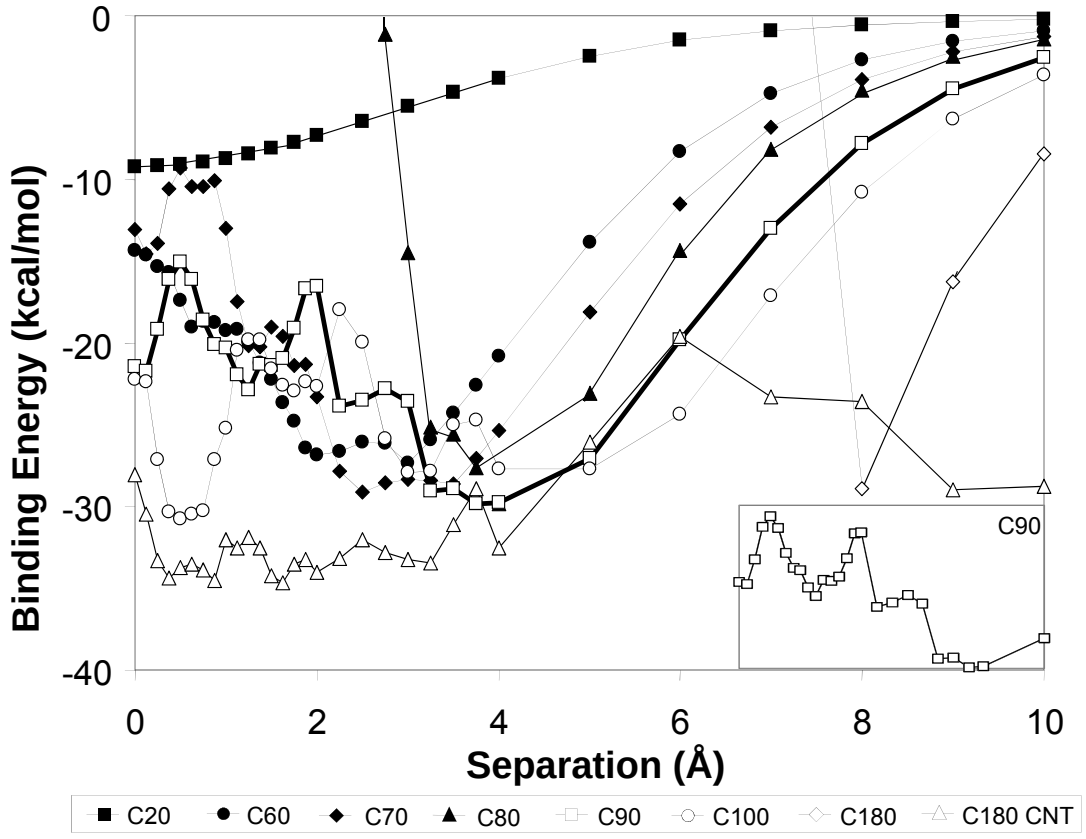


Figure 4.6: Single-point potential energy curves for interactions with ring1. The insert in lower right corner corresponds to the C90 curve (Bold line) in the range of 0 to 5 Å. This insert is attended for clarity of the observed oscillations.

One common theme in the search for fullerene complexation agents has been the idea that for best binding one should optimize the "fit" of the ring to the fullerene of interest[76, 78]. Consultation of Table 4.2 reveals no obvious pattern of selectivity of the rings based on the molecular weight of the fullerenes. To investigate the complexation more closely, potential energy curves were computed for interactions of each fullerene with each ring. Discrete representations of the PECs were generated by placing the fullerene at large separation from the ring on the axis that pierces the center of the ring

perpendicular to the nominal ring plane. (This axis is approximately collinear with the axis corresponding to the largest eigenvalue of the moment of inertia tensor for the isolated ring and is taken to be the z-axis here.) The single-point energies for the different fullerenes interacting with ring1 are shown in Figure 4.6. Corresponding PECs were generated for the other ring systems; however, more accurate PECs were desired. Therefore, constrained optimizations were performed by optimizing two of the spatial coordinates for every atom (x and y) while freezing the third (z). The two-dimensional optimizations preserved the fixed distance between the two components while allowing the complex to remove any close contacts. The results are shown in Figure 4.7 - 4.10. (These constrained optimization are well beyond what is currently computationally reasonable with DFT methods and are a testament to the computational efficiency of AM1-FS1.) Note that some complexes optimize with the fullerene at the center of the ring cavity, whereas others optimize with the ring offset from the center. This is also the case when full optimizations are performed; the displacements are reported in Figure 4.11. In general, this is unsurprising. The low molecular weight fullerenes "fit" into the cavity whereas the higher molecular weight ones are too large to insert into the cavity and instead find optimum interaction when displaced from the center, much as a large beach ball might rest on a basketball hoop whereas a basketball fits through the hoop.

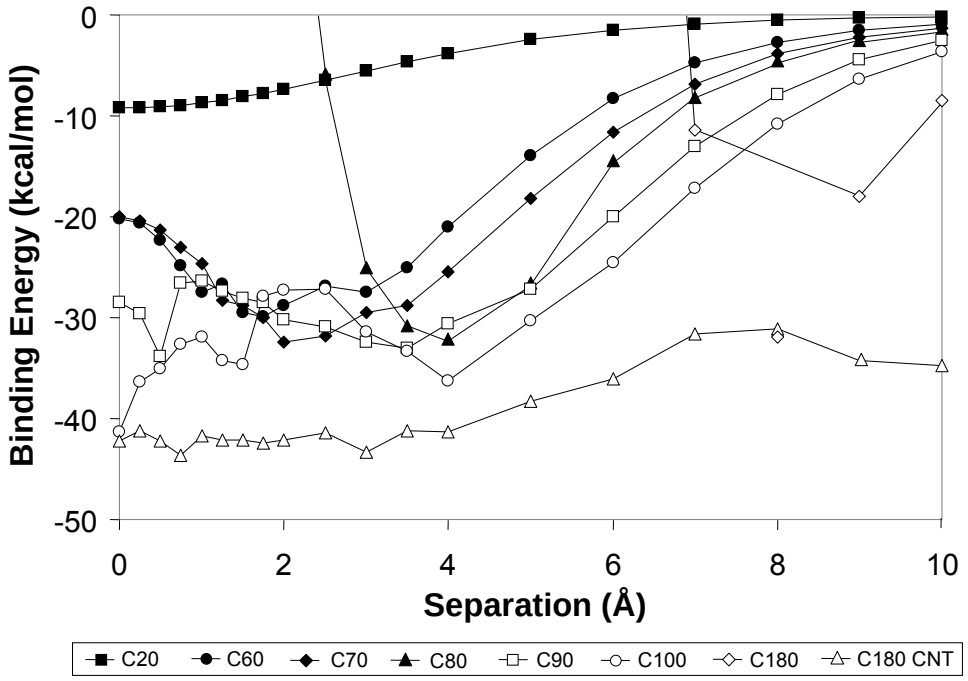


Figure 4.7: Constrained optimization potential energy curves for interactions with ring1.

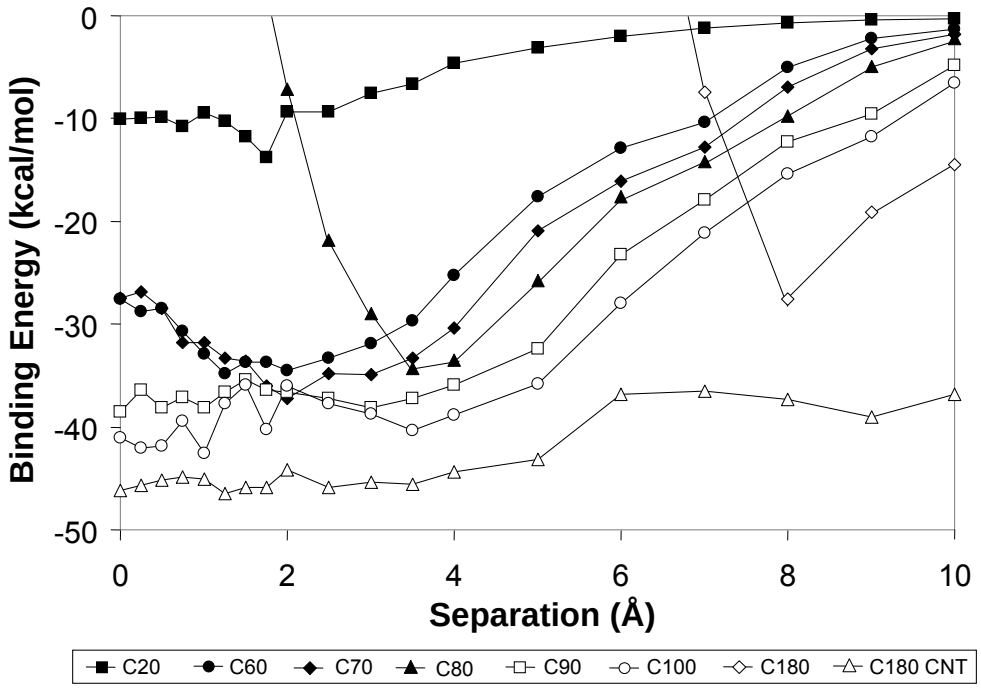


Figure 4.8: Constrained optimization potential energy curves for interactions with ring2.

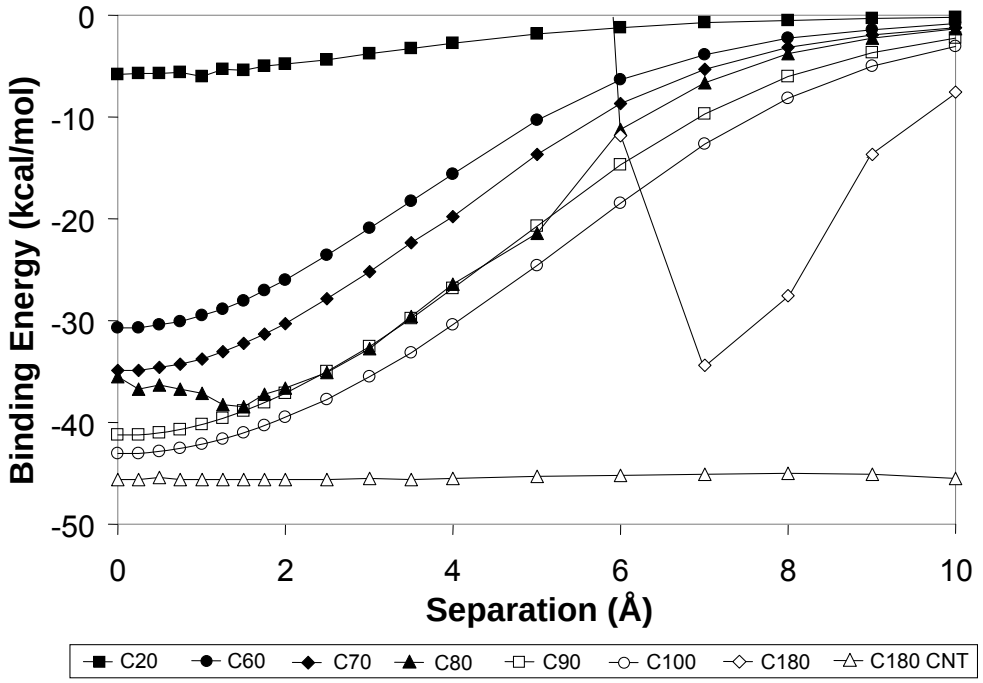


Figure 4.9: Constrained optimization potential energy curves for interactions with ring3.

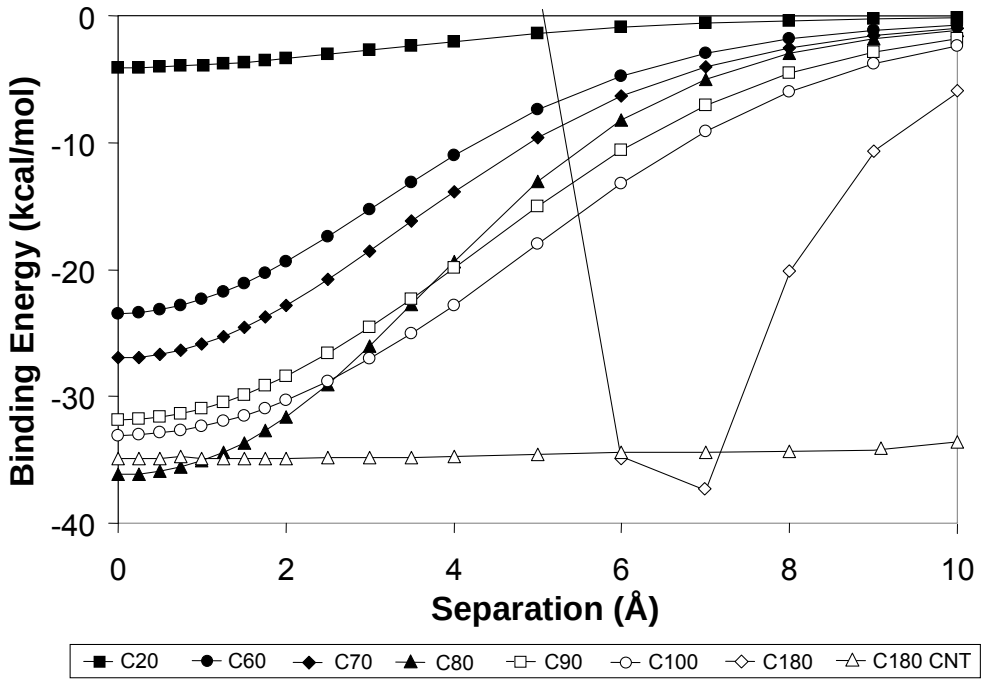


Figure 4.10: Constrained optimization potential energy curves for interactions with ring4.

Consider, for example, the fullerene interactions with ring3. C_{20} binds inside of the ring. C_{60} also binds inside the ring but the binding energy is greater because more inter-component atom-atom interactions are near their optimum van der Waals separation. The match is even better for C_{70} as evinced by the slightly greater binding energy. C_{80} is too large to fit within the cavity of the ring so the optimized structure exhibits the fullerene displaced from the center, (δ_{80}^{r2} = 0.7 Å, where δ denotes displacement, the subscript indicates the fullerene and the superscript indexes the ring). One might expect that the larger fullerenes optimize to positions farther from the ring center, and indeed C_{180} is much more displaced from the center (δ_{180} = 6.6 Å), but C_{90} and C_{100} appear to break the pattern. ($\delta_{90}^{r2} \approx \delta_{100}^{r2} \approx 0$). The origin of this discrepancy is revealed by closer examination of the structures of the fullerenes. C_{20}, C_{60}, C_{80}, and C_{180} are nearly spherical, but C_{70}, C_{90}, C_{100}, and $C_{180\text{-CNT}}$ are prolate so despite their molecular weights they share nearly identical cross-sectional diameters with C_{60} and therefore fit within the cavity of ring3.

If we consider only the spherical fullerenes, the trend in binding is clear: Starting from the smallest fullerene, C_{20}, the binding energy increases with molecular weight as long as the fullerene fits within the cavity of the ring. ($\Delta E_{20}^{r3} < \Delta E_{60}^{r3}$). For larger fullerenes that don't fit inside the ring, the larger the fullerene, the greater is its optimized displacement from the ring ($\delta_{80}^{r2} < \delta_{180}^{r2}$).

If we consider only the five fullerenes that have approximately 7 Å cross sectional diameter, (C_{60}, C_{70}, C_{90}, C_{100}, $C_{180\text{-CNT}}$) all fit within ring3 given proper orientation. The binding energy ΔE increases with molecular weight ($\Delta E_{60}^{r3} < \Delta E_{70}^{r3} < \Delta E_{90}^{r3} < \Delta E_{100}^{r3} \approx \Delta E_{180\,NT}^{r3}$) because more inter-component atom-atom dispersion interactions are possible.

However, the binding energy converges to a maximum at some molecular weight since making the fullerene longer eventually adds only very long-range inter-component interactions that make no appreciable contribution to the binding energy. This is evident by comparing the binding energies for C_{100} and $C_{180\text{-}CNT}$ interacting with ring3, both prolate-shaped fullerenes have approximately 7 Å cross-sectional diameter, but the computed binding energies differ by less than a half a percent.

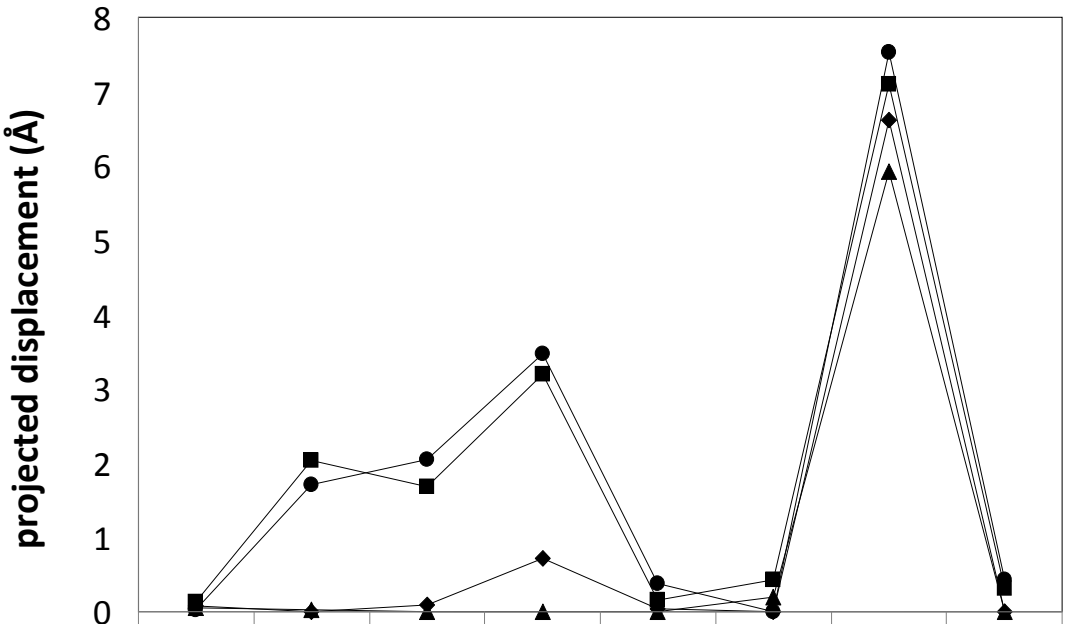
Both ring1 and ring2 have sufficiently smaller diameters than ring3 so that only C_{20} binds within the ring cavity, but otherwise the trends are the same. For the spherical fullerenes, the displacement increases with increasing molecular weight ($\delta_{20}^{r1} < \delta_{60}^{r2} < \delta_{80}^{r1} < \delta_{180}^{r1}$) and ($\delta_{20}^{r2} < \delta_{60}^{r2} < \delta_{80}^{r2} < \delta_{180}^{r2}$). The corresponding binding energies are generally larger for ring2 than ring1 because more intercomponent interactions are involved due to the increase in the number of atoms but not in the diameter in ring2 relative to ring1. The potential energy curves for interaction of the prolate fullerenes (C_{70}, C_{90}, C_{100}, $C_{180\text{-}CNT}$) with ring1 show interesting oscillations so that there are multiple local minima at close separation. (This is most evident for the single-point energy curves (Figure 4.6).) For clarity, the interactions of ring1 with C90 are highlighted in the insert of Figure 4.6. The oscillations are the strongest for C_{100}, which exhibits local minima near 0.5, 1.75, 3.0 and 4.25 Å. Note that these minima are separated by ~1.25Å and are a manifestation of translational periodicity in the interaction between the nanotube and the ring. Similar oscillations appear for ring2, but no such oscillations are present for interactions with ring3 or ring4 for interactions with these fullerenes. The oscillations will likely be present for any tube-like fullerene interacting with a ring in which the fit is "tight", i.e. the tube diameter is only slightly smaller than the diameter of the ring cavity.

Table 4.3: Fullerene-nanoring potential interaction distances (Angstroms). Values are defined as one-half of the difference of the diameter of the two components.

		C_{20}	C_{60}	C_{70}	C_{80}	C_{90}	C_{100}	C_{180}	$C_{180\text{-CNT}}$
		4.1	7.1	7.1	8.2	7.0	6.9	12.6	6.8
ring1	13.3	4.6	3.1	3.1	2.6	3.2	3.2	0.4	3.3
ring2	13.1	4.5	3.0	3.0	2.5	3.1	3.1	0.3	3.2
ring3	14.5	5.2	3.7	3.7	3.1	3.7	3.8	0.9	3.8
ring4	15.2	5.5	4.0	4.0	3.5	4.1	4.1	1.3	4.2

For the largest of the rings, all of the fullerenes except for C_{180} fit within the cavity. Again, to dissect the trends, we separate the spherical and prolate fullerenes. Among the spherical fullerenes that do fit within the cavity ($\Delta E_{20}^{r4} < \Delta E_{60}^{r4} < \Delta E_{80}^{r4}$), increasing in binding strength as the fullerene becomes a closer match to the cavity size. Among the prolate fullerenes, the binding energy generally increases with molecular weight because more inter-component atom-atom dispersion interactions are possible. Overall the deepest potential well for interaction with ring4 is exhibited by C_{80} (Figure 4.10). In this case the C_{80} fullerene, which has a diameter of 8.2 Å, interacting with ring4 of diameter 15.2 Å, yields an interaction distance of 3.5 Å, (See Table 4.3) which is the closest to the optimum separation of graphene sheets of 3.4 Å[89] that is possible with the set of fullerenes considered here. Upon full optimization, however, ring1, ring2 and ring3 all exhibit the strongest interaction with $C_{180\text{-CNT}}$. Here the interaction distance is never more than 0.4 Å from optimum, and $C_{180\text{-CNT}}$ presents sufficiently many more

opportunities for atom-atom dispersion interactions than any of the other fullerenes that its binding is strongest. Ring4 breaks the pattern, showing optimum binding with C_{80}. This is likely due to the fact that the diameter of $C_{180\text{-}CNT}$ is slight smaller than any of the other prolate fullerenes considered. As a test, a slightly different conformation of the C_{180} capped carbon nanotube was considered, which had a diameter of 7.9 Å (comparable to C80). The resulting binding energy was -50.9 kcal/mol, showing again that binding energy increases with molecular weight for the prolate fullerenes.



	C20	C60	C70	C80	C90	C100	C180	C180 CNT
ring1	0.0	1.7	2.0	3.5	0.4	0.0	7.5	0.4
ring2	0.1	2.0	1.7	3.2	0.2	0.4	7.1	0.3
ring3	0.1	0.0	0.1	0.7	0.0	0.0	6.6	0.0
ring4	0.1	0.0	0.0	0.0	0.0	0.2	5.9	0.0

Figure 4.11: Projected displacements of the fullerene from the ring centroid in the fully-optimized fullerene-nanoring complexes.

4.4 Conclusions

The interactions of eight fullerenes, (having molecular weights ranging from 240 to 2162) with four PAH nano-rings have been studied with AM1-FS1, an empirically-corrected semiempirical electronic structure theory. It is found that trends in structure and binding energy for the complexes only become apparent when the fullerenes are categorized by geometry. Among fullerenes of spheroidal geometry, those that fit within a ring cavity show strongest bonding for the largest structure. Among fullerenes too large to fit within the ring cavity, the structural displacement of the fullerene from the ring increases with fullerene molecular weight. For fullerenes having a prolate spheroidal geometry that fit within a ring cavity, binding strength increases with fullerene molecular weight, converging to a maximum for fullerenes of very high aspect ratio. Fullerene complexation is therefore highly conformationally dependent. Given the increasing prevalence of conformational isomerism with increasing fullerene molecular weight the challenge of separating fullerenes by complexation increases with molecular weight.

Chapter 5: A Computational Investigation of the Role of Counterions and Reorganization Energy in a Switchable Bistable [2]Rotaxane

5.1 Introduction

The development of molecular electronic devices is at the forefront of nanotechnology. Such devices not only have vast potential impact on the ever growing electronic world, but also are intrinsically intriguing to the scientific community. The fabrication of devices from molecules, termed the "bottom-up" approach, holds the potential of removing the size reduction limitations associated with the current "top-down" methodology typically used in device manufacture. A leading candidate in the development of molecular electronic devices is the "Stoddart-Heath-type" [2]rotaxane based switch[90-99]. Workable memory devices[93, 98, 99] have been fabricated using these switchable bistable [2]rotaxanes as memory elements and "promise ultimate scalability, minimal power consumption and low fabrication costs[97]."

A switchable bistable [2]rotaxane is composed of two components; a long chain molecule referred to as the "shaft" that passes through a "ring" structured molecule. The shaft contains two *stations* where the ring may bind and is terminated with bulky end groups, *stoppers*, preventing dissociation of the ring from the shaft. To achieve switchable functionality, the stations have different compositions and therefore different affinities for the ring. The complex therefore has two co-conformations with different stabilities, hence it is bistable. Some of the currently most promising systems are composed of a tetra-cationic ring, cyclobis(paraquat-p-phenylene) ($CBPQT^{4+}$), which *shuttles* between the electron-donating groups tetrathiafulvalene (TTF) and 1,5-dioxynaphthalene (DNP) on the shaft. Among these types of systems, considerable

synthetic and experimental effort has been expended with the hope of achieving more desirable properties[97]. The system studied here, (shown in Figure 5.1) has been successfully incorporated into a two-dimensional crossbar circuit[94]. This system, as well as similar ones, has the ability to function as a molecular switch because the two co-conformations exhibit different conductivities. The ground state co-conformation (GSCC), which is associated with the ring at the TTF station, exhibits low-conductance (off-state); whereas, the metastable state co-conformation (MSCC), wherein the ring is at the DNP station, exhibits high-conductance (on-state) (see Figure 5.2). This behavior has been demonstrated experimentally[93, 94, 96] in numerous systems of this type. While the exact theoretical reason for the conductivity change is unknown, the basic mechanism for functionality is agreed upon[92, 93, 100]. The inhabitants of the GSCC and MSCC states are controlled by oxidation of the system, which occurs predominantly in the TTF moiety. Oxidation causes an electrostatic repulsion of the tetracationic ring from TTF, rendering the DNP station the only stable binding site for the ring. Upon reduction, the ring remains at the DNP station. Now in the MSCC state the system is highly-conductive. After a period of time the system returns to the GSCC. The relaxation time depends on the reaction barrier (shuttling barrier) that must be surpassed[97]. Therefore, the shuttling barrier and site preference are key properties for designing functional and nonvolatile memory[95].

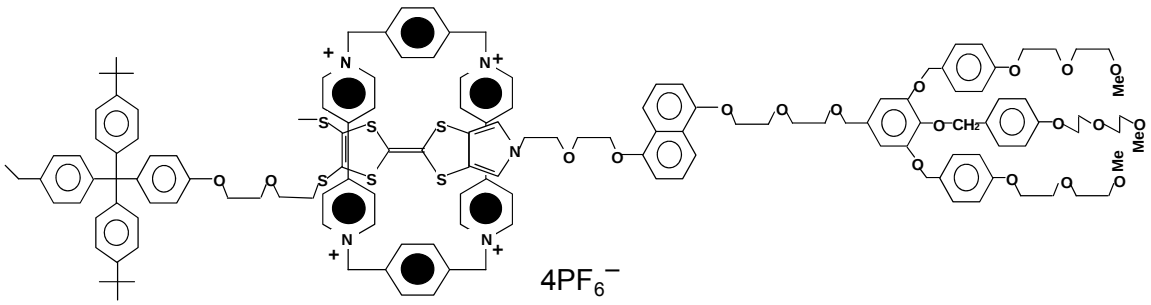


Figure 5.1: Schematic of the switchable bistable [2]rotaxane.

Computational studies of molecular electronic devices have the potential of providing a fundamental understanding of the inner-workings of these types of systems. Ultimately, such studies will aid the future design process by providing more likely structural compositions that will provide the desired properties. The first step towards utilizing computational methods for design purposes is to identify a method that is capable of accurately predicting the key properties of known systems. When choosing a particular method, one typically is faced with the dilemma of balancing accuracy and computational cost. The accuracy needed for a particular problem depends on the property or properties in question. In practice, accuracy is limited by computational resources; therefore, as the size of the system increases, the ability to perform high level *ab initio* calculations diminishes. In an attempt to remove this obstacle, it is common to remove part of the system to focus in on individual sections. This allows higher level calculations to be performed with the goal of obtaining more accurate results than otherwise possible. The Goddard group[101] has demonstrated that qualitatively correct results can be achieved on these types of systems by such an analysis. An alternative

approach is to employ computationally efficient semi-empirical electronic structure methodology.

In this study we employ the AM1-FS1[75] method to investigate the binding site preference and energy profiles for the shuttling of the ring along the shaft in the neutral and dicationic states. In addition, the reorganization energy associated with the oxidation/reduction process, starting from both the GSCC and MSCC states, is determined. The AM1-FS1 method is a version of the AM1 Hamiltonian empirically-corrected for van der Waals (dispersion) and hydrogen-bonding interactions. The incorporation of dispersion interactions for systems of the type considered here is important because of the π-π stacking that occurs when the ring resides at either of the binding sites. These interactions are neglected in Hartree-Fock (HF) theory and by many exchange-correlation (XC) functionals used in density functional theory (DFT)[102]. Note that the system considered here does not have an alpha-hydrogen atom attached to a highly electronegative atom (N, O, or F); therefore, the H-bonding correction in AM1-FS1 method is not utilized. One of the main goals of this article is to further demonstrate that the AM1-FS1 method is capable of accurately modeling intermolecular bound complexes[75, 103]. As additional evidence, comparisons are made to previously published experimental and computational results as well by performing in-house DFT calculations. Choosing the AM1 Hamiltonian, over DFT methodologies, is desirable because semi-empirical methods can be 100-1000 times more computationally efficient. In addition to the mentioned studies, the presence of the hexafluorophosphate (PF_6^-) counterions on the computational results is considered.

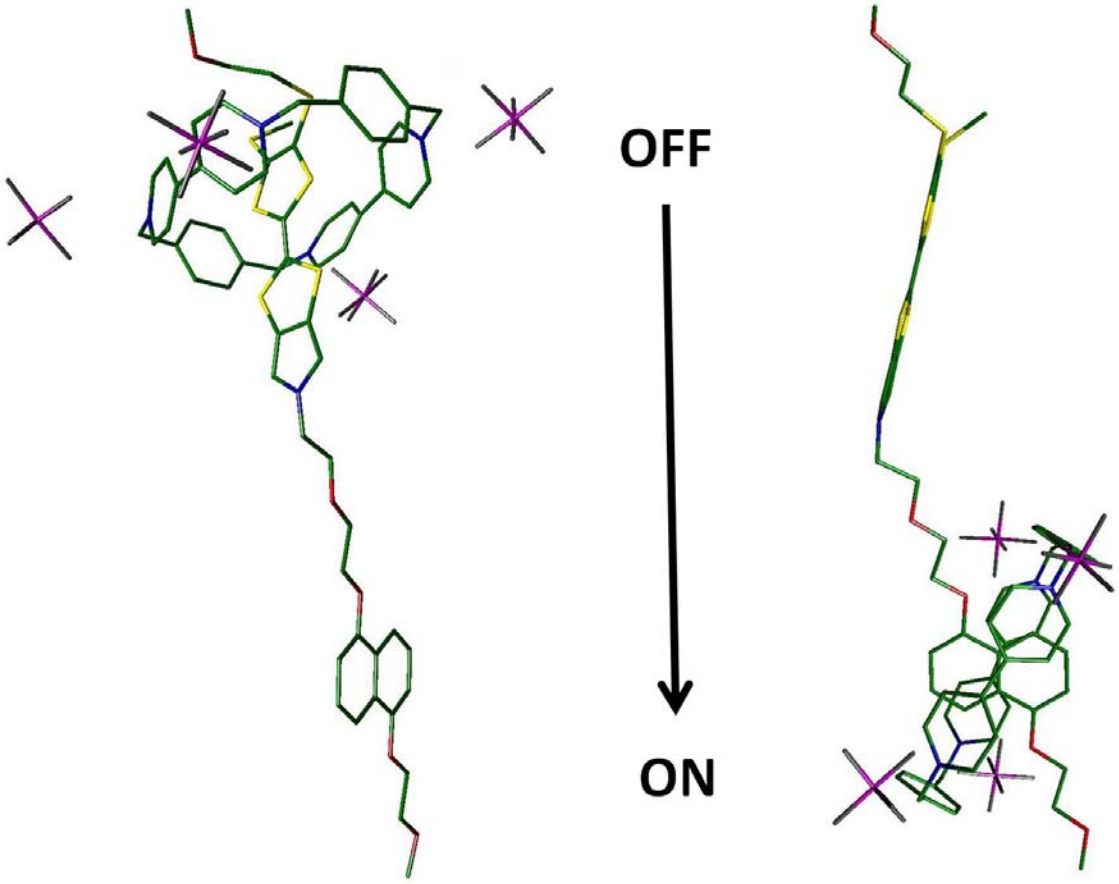


Figure 5.2: Full Geometry Optimized AM1-FS1 structures. On the left: the GSCC which is associated with the "off" state. On the right: the MSCC which is associated with the "on" state. The hydrogen atoms have been removed for clarity.

5.2 Computational Methods

Starting structures for the components of the pseudorotaxane considered here were generated using the HyperChem[104] gui. These structures were then fully optimized at the AM1-FS1[75] level of theory, using a locally modified version of GAMESS[105]. Hexafluorophosphate (PF_6^-) counterions were then symmetrically placed in the vicinity of each electron deficient nitrogen atom in the $CBPQT^{4+}$ ring and the $CBPQT(PF_6)_4$ complex was fully optimized.

To map the variation in energy with ring shuttling along the shaft, constrained optimizations were carried out for the complex with the center-of-mass (CM) of the optimized ring structure placed at 18 different positions along the shaft of the pseudorotaxane. Each of the positions corresponds to an atom, or midpoint between two atoms, along the shaft (see Figure 5.3). At each position, 6 different trial structures were generated by rotating the CBPQT$(PF_6)_4$ complex in equal increments between 0 and 90 degrees (this range covers full axial rotation due to the symmetry of the ring) about the axis that passes through the selected atom and its nearest neighbor on the backbone of the shaft, generating a total of 108 structures. All 108 structures were then optimized in the neutral (ground state) and oxidized (2+ state) states with AM1-FS1. To help maintain the desired conformations, three constraints were applied. First, a torsional constraint was applied between three "corner" carbon atoms on the ring and the selected atom position on the shaft. This constraint maintains the ring position relative to the shaft but allows for rotation of the ring during optimization. Second, the distance between the N atoms on the ring and the corresponding P atoms of the counterions were also fixed. This constraint prevented a large change in the position of the counterions, especially during the beginning steps of the optimization procedure. Without this constraint it was observed in some cases that the initial atomic forces were very large, (presumably due to unphysically close contacts created during structure generation) causing rapid counterion migration resulting in large differences in final energy between similar starting structures. The use of counterion constraints is intended to help more clearly map the shuttling barrier while limiting the computational space that needs to be explored. Third, the position of one atom on each end of the shaft was fixed, preventing the system from coiling. This

constraint is justified since in typical applications the ends of these systems are bound to surfaces preventing coiling. In practical application, the neighboring structures on the surface should also help prevent coiling of the complex. Note, approximately 10 structures in each state failed to converge; mainly because rotation of the ring near the DNP site causes unphysically close contacts. These structures were simply discarded.

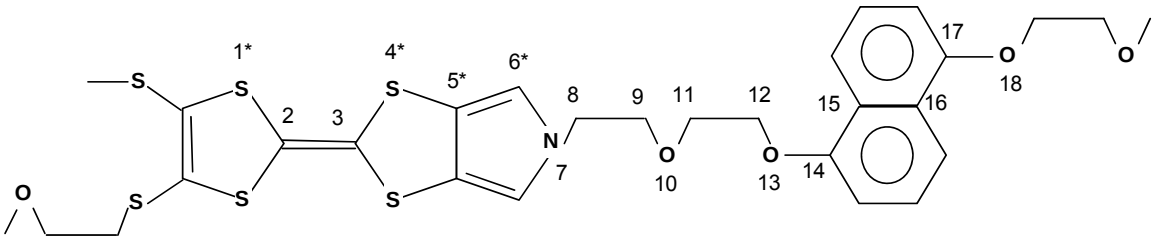


Figure 5.3: Schematic of the shaft of the pseudorotaxane. The different ring positions along the shaft are identified by integers; The " * " indicates that the midpoint between the two atoms was used for ring placement.

To further investigate this system, the lowest energy structure, (based on the AM1-FS1 optimizations) for each ring position along the shaft was identified and single-point DFT and AM1 calculations were preformed. This comparison is intended to remove methodological bias and show that AM1-FS1 is appropriate for the task at hand. All DFT calculations were performed using NWChem[106], with the PBE functional[107] and the 6-31G basis set. In addition, single-point calculations were carried out on these optimized structures with the counterions removed, since it has been suggested that the actual presence of these ions is uncertain[108]. Removing the counterions inherently changes the charge of the ground state of the system to 4+. To compare with results previously

reported by Subramanian et al.[108], single-point energies were also determined for the neutral state, again without counterions. This is an "unnatural" electronic configuration, however, we believe this was the electronic state previously considered. Evidence and further discussion of this matter will be addressed in the next section.

The lowest energy GSCC and MSCC structures, including counterions, were further investigated. These two structures were fully optimized, at the AM1-FS1 level, with all geometry constraints removed and are shown in Figure 5.2. A full geometry optimization was also performed on the lowest energy structure, including counterions, in the oxidized 2+ state with the ring positioned at the DNP site. (Only the DNP site was considered in the oxidized state; since the ring is energetically disfavored that the TTF site in the oxidized state.)

5.3 Results and Discussion

The variation in energy for the ring shuttling along the shaft of the pseudorotaxane in the neutral and oxidized (2+) states is shown in Figure 5.4. These curves are based on the AM1-FS1 constrained optimizations in the presence of counterions. The neutral state curve (more clearly depicted in Figure 5.7) correctly predicts the lowest energy co-conformation (GSCC) to be when the ring is positioned about the TTF site; a result that is in agreement with experimental[94] and other computational results[108]. Based on this curve (constrained optimizations), the GSCC state is favored by 1.6 kcal/mol over the MSCC state. This number is in excellent agreement with the value (1.5 kcal/mol) obtained when full geometry optimizations were carried out with the ring positioned at the TTF and DNP site. This provides evidence that the constraints applied during geometry optimization did not qualitatively alter the results.

The predicted reaction barrier for the ring moving from the GSCC state to the MSCC state is 14.4 kcal/mol, (and 12.8 kcal/mol in the reverse direction). These values agree with previously published[108] DFT (PBE functional) calculations using periodic-boundary-condition; where a 14 kcal/mol (600 meV) barrier was reported from the GSCC state to the MSCC state. Figure 5.4 also displays the reaction path in the oxidized state. This curve clearly indicates that upon oxidation DNP is the preferred site for ring binding. This site is predicted to be preferred by about 90 kcal/mol. This large energy preference is a result of the oxidation of the TTF group, leading to the repulsion of the tetracationic ring. The DNP site is therefore the only stable ring-binding site in the oxidized state. These two states are predicted to be separated by about 400 kcal/mol (~17 eV), roughly corresponding to the energy required to remove two electrons from the system.

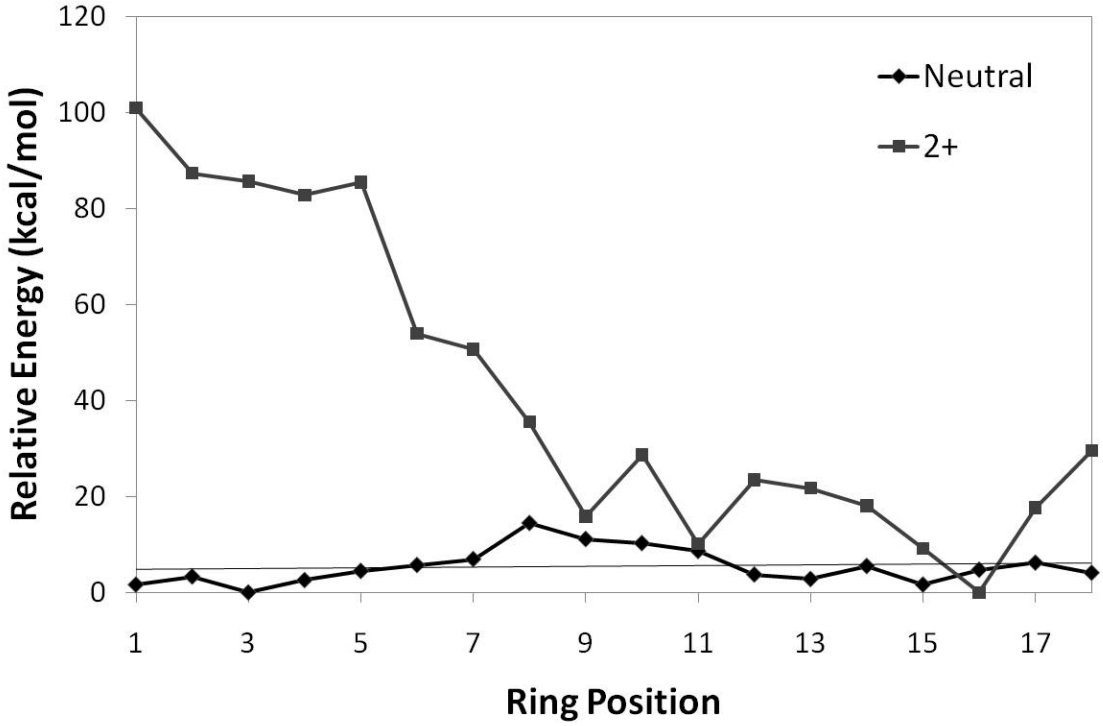


Figure 5.4: AM1-FS1 potential energy curves for the shuttling process of the pseudorotaxane with counter ions in the neutral and 2+ states obtained by restrained optimizations.

The binding site preference and the reaction barrier influence the functionality of the system. The binding site preference controls whether the system is in the "on" or "off" state. If the DNP site were preferred over the TTF site in the reduced state, the ring would never leave that site, rendering the system useless as a molecular switch. In this particular system, the TTF site is predicted to be more stable by about 1.5 kcal/mol. It might be desirable if these two states were separated further in energy to limit the initial population of the MSCC state, thereby decreasing the number of defects upon assembly. The reaction barrier is also important because it controls the rate at which the system returns to the GSCC from the MSCC. Increasing the reaction barrier should increase the

memory-retention characteristics. Tailoring these different properties by structural modifications is one area where accurate computational methods can potentially be extremely useful.

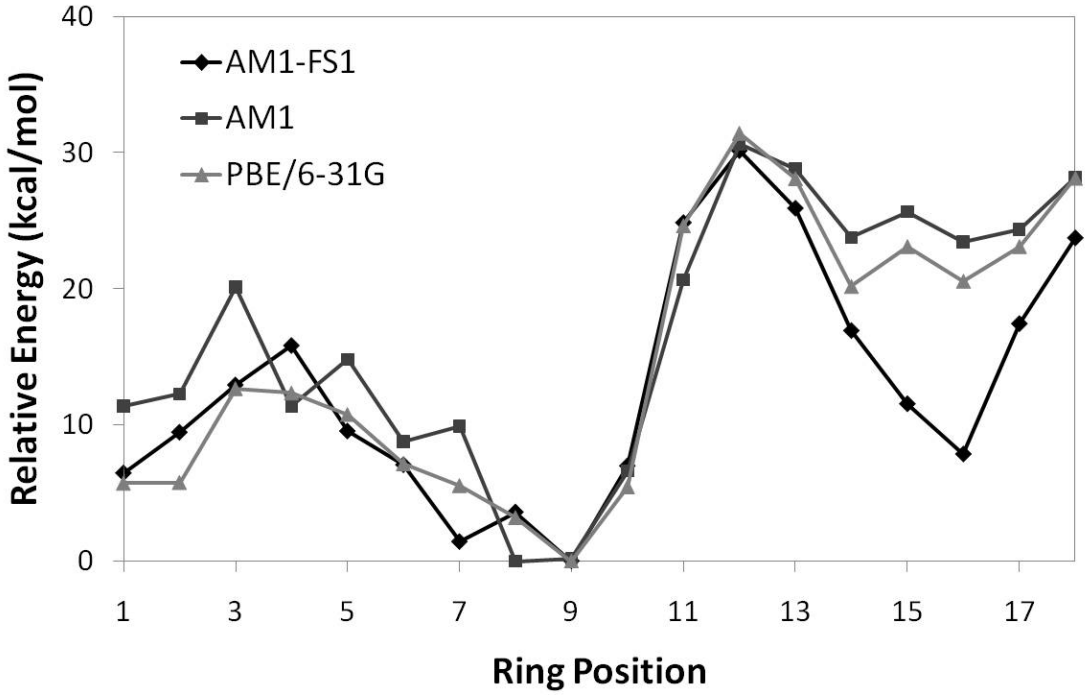


Figure 5.5: Potential energy curve for the shuttling process of the pseudorotaxane without counter ions in the neutral state (incorrect state). The pseudorotaxane structures were obtained by removing the counter ions from the AM1-FS1 optimized structures.

As mentioned above, this system was further studied by removing the counterions, since it has been suggested that "the presence and arrangement of these counterions has not been definitively established[108]." Subramanian et al[108], reported that in the absence of counterions the ring is energetically favored to bind between the TTF and DNP sites. We disagree with these findings and believe this is a result of a

computational error. Specifically, we believe that when the counterions where removed from the system the charge state was not changed to 4+, the natural charge state of the system in the absence of the four PF_6^- counterions, which balance the tetracationic ring ($CBPQT^{4+}$). To support this claim, we have carried out single-point calculations in the absence of counterions in the neutral (incorrect electronic state) and 4+ (correct electronic state) states. For comparison, these calculations were carried out at the AM1, AM1-FS1, and PBE/6-31G levels. The PBE functional was chosen since it was used in the previous study that found the ring to be favored between the two binding sites in absence of counterions. Figure 5.5 shows the energy profile for the ring moving along the shaft in the neutral state, clearly the ring is preferred at the center of the shaft. This is the same result reported by Subramanian et al.[108]; however, this curve represents the incorrect charge state. On the other hand, if we model the same system in the correct charge state (4+), a completely different energy profile is observed (see Figure 5.6). This curve more closely describes the correct behavior of the system, but does suggest that the ring is preferred at the DNP site in the absence of the counterions. The conclusion of Subramanian et al.[108], that the presence of counterions is needed to correctly describe this system, may be true, but the reasoning is entirely different.

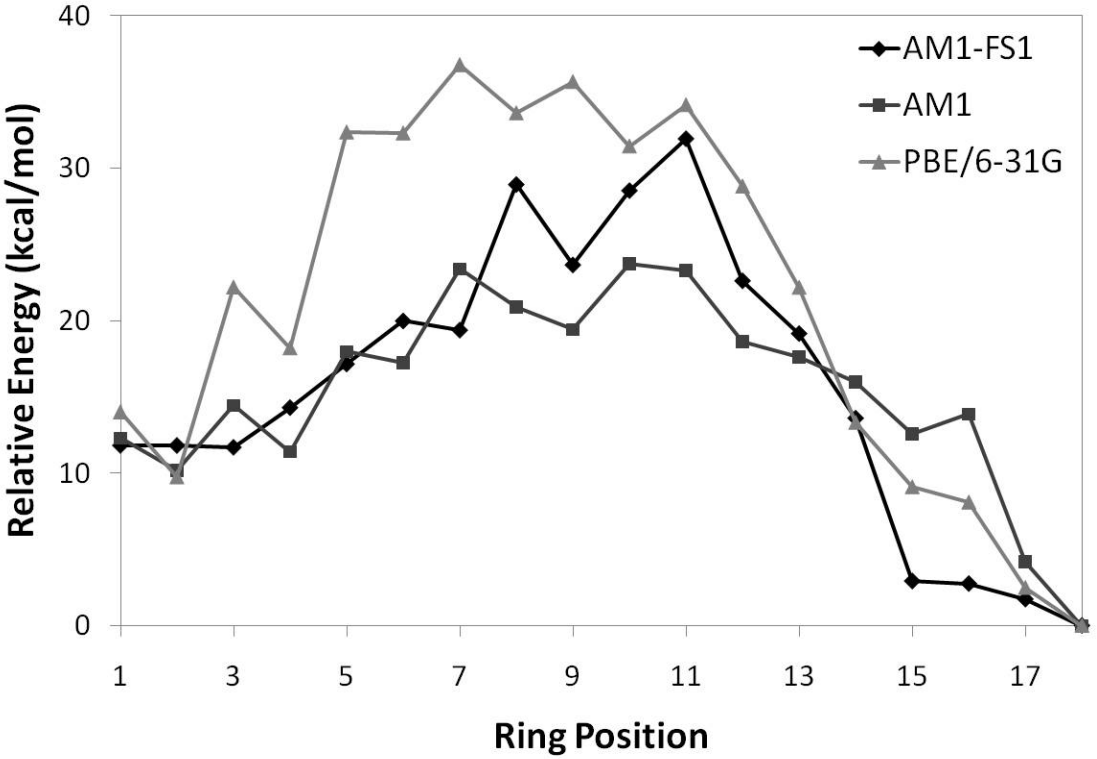


Figure 5.6: Potential energy curve for the shuttling process in the pseudorotaxane without counter ions in the 4+ state (correct state). The pseudorotaxane structures were obtained by removing the counter ions from the AM1-FS1 optimized structures.

To further support the accuracy of our AM1-FS1 calculations, single-point PBE/6-31G calculations in the presence of counterions were preformed. The results are shown in Figure 5.7. This figure shows that the PBE single-point calculations, based on the AM1-FS1 optimized structures, qualitatively predict the correct behavior of the system. These results are in agreement with those by Subramanian et al.[108], where the same XC-functional was used to study this system.

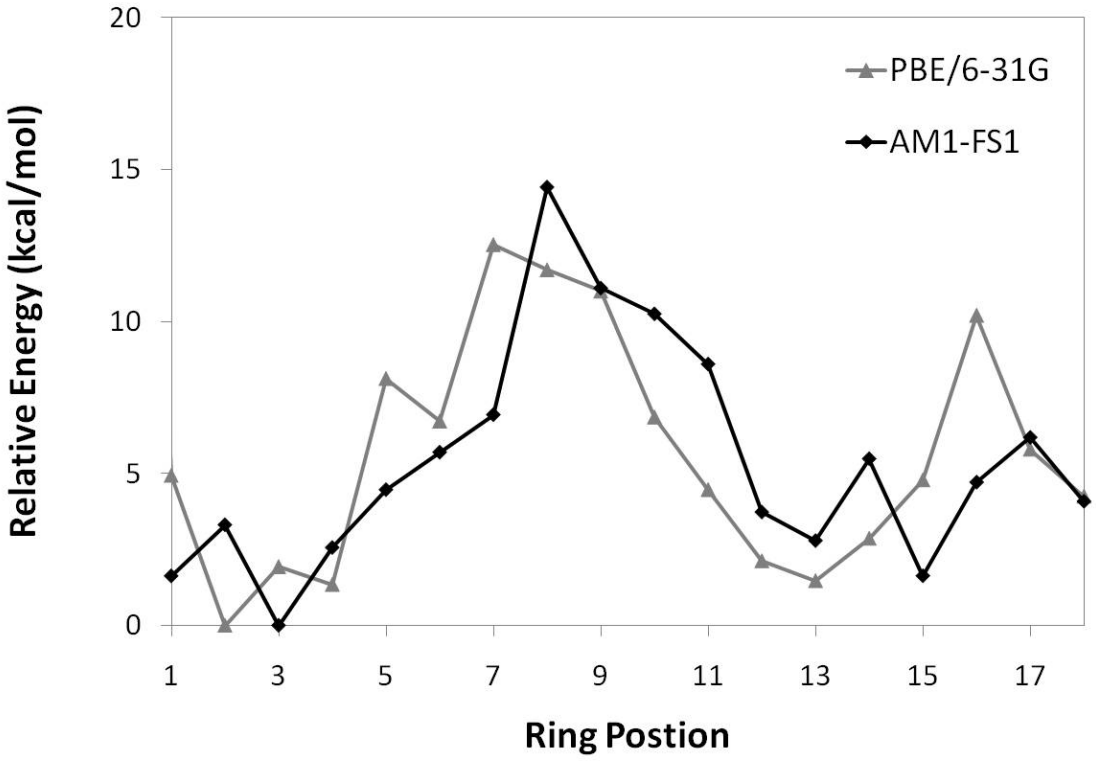


Figure 5.7: Potential energy curve for the shuttling process of the pseudorotaxane with counter ions. Single-point PBE/6-31G calculations were preformed on the AM1-FS1 optimized structures.

It has been recently suggested[109] that the reorganization energy associated with the oxidation/reduction process is the leading cause of the experimentally observed conductivity difference between the GSCC and MSCC states. This claim was theorized based on Deng and Goddard's implementation of Marcus hop rate theory[110]. More specifically, it is based on the fact that the reorganization energy enters exponentially into the expression for hop rate, which is directly proportional to the conductivity. Moreover, a decrease in reorganization energy corresponds to an increase in the hop rate and thus an increase in conductance. In a switchable bistable [2]rotaxane, as considered here, two minima exist in the reduced state (neutral or 4+ w/o counterions) namely the GCSS and

MSCC states. Therefore, the system can be in either of the two co-conformations at the moment of oxidation resulting in two different reorganization energies. These values can be determined from the energy associated with the diabatic charge transfer (a detailed prescription for determining these values is given in Ref. 109). The reorganization energies based on the AM1-FS1 full optimizations with counterions are 172 kcal/mol and 108 kcal/mol starting from the GSCC and MSCC respectively. This 64 kcal/mol difference has a huge affect (many orders-of-magnitude) on the difference in the predicted hoprate. The large difference in reorganization energy is not especially surprising since there is a large change in the nuclear coordinates between the minimum withe the ring at the TTF site in the reduced state, and the minimum with the ring at the DNP site in the oxidized state. On the other hand, when the system is oxidized in the MSCC, the ring remains at the DNP site; therefore, significantly less structural reorganization of the system is required upon oxidation. This energy difference is of the magnitude of the relaxation energy for the ring moving from the TTF site to the DNP site in the oxidized state. This result supports the claim[109] that the reorganization energy is the leading cause for the observed high-conductance in the MSCC state in comparison to the GSCC state.

5.4 Conclusions

Switchable bistable [2]rotaxanes have been previously studied both experimentally and computationally with the aim of aiding in the future development of molecular electronic devices. Although, many promising advances have been made, further investigations and advances are still needed before such devices will become integrated into practical electronic devices. To this end, we have studied one such

plausible system in hopes of indentifying a relatively efficient computational model for predicting key properties of the complex. We have demonstrated that the AM1-FS1 method is capable of determining binding site preference as well as quantitatively (relatively speaking to the more computationally expense DFT method) predicting the relative stabilities and the reaction barrier between the different co-conformational states. Since the functionality of the Stoddart-Heath type [2]rotaxanes depends on the difference in conductance resulting from the co-conformational state occupied, we have explored the idea that the reorganization energy is the possible origin of differences in conductivity between the two co-conformations. Our findings indeed support this prior claim. All of these properties, (binding site preference, barrier to co-conformational isomerism, reorganization energy) play important roles in the functionality of these systems; therefore, there is significant benefit in identifying an efficient computational technique for identifying them.

List of References

1. Xiao, Y., C. Chen, and Y. He, *Folding Mechanism of Beta-Hairpin Trpzip2: Heterogeneity, Transition State and Folding Pathways.* Int. J. Mol. Sci., 2009. **10**: p. 2838-2848.

2. Halperin, I., B. Ma, H. Wolfson, and R. Nussinov, *Principles of Docking: An Overviewof Search Algorithms and a Guide to Scoring Functions.* Proteins: Str. Func. and Genet., 2002. **47**: p. 409-443.

3. Jonikas, M.C., S.R. Collins, V. Denic, E. Oh, E.M. Quan, V. Schmid, J. Weibezahn, B. Schwappach, P. Walter, J.S. Weissman, and M. Schuldiner, *Comprehensive Characterization of Genes Required for Protein Folding in the Endoplasmic Reticulum.* Science, 2009. **323**: p. 1693-1697.

4. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**: p. 884-890.

5. Querol, E., J.A. Perez-Pons, and A. Mozo-Villarias, *Analysis of protein conformational characteristics related to thermostability.* Protein Eng., 1996. **9**(3): p. 265-271.

6. Aurora, R. and G.D. Rose, *Helix capping.* Protein Sci., 1998. **7**: p. 21-38.

7. Seeman, N.C., *An Overview of Structural DNA Nanotechnology.* Mol Biotechnol, 2007. **37**: p. 246-257.

8. Yang, W., *Poor base stacking at DNA lesions may initiate recognition by many repair proteins.* DNA Repair, 2006. **5**: p. 654-666.

9. Vallee, R., P. Damman, M. Dosiere, E. Toussaere, and J. Zyss, *Nonlinear Optical Properties and Crystalline Orientation of 2-Methyl-4-nitroaniline Layers Grown on Nanostructured Poly(tetrafluoroethylene) Substrates.* J. Am. Chem. Soc., 2000. **122**: p. 6701-6709.

10. Thalladi, V.R., S. Brasselet, H.-C. Weiss, D. Blaser, A.K. Katz, H.L. Carrell, R. Boese, J. Zyss, A. Nangia, and G.R. Desiraju, *Crystal Engineering of Some 2,4,6-Triaryloxy-1,3,5-triazines: Octupolar Nonlinear Materials.* J. Am. Chem. Soc., 1998. **120**: p. 2563-2577.

11. Hunter, C.A. and J.K.M. Sanders, *The nature of pi-pi interactions.* J. Am. Chem. Soc., 1990. **112**: p. 5525-5534.

12. Griffiths-Jones, S.R. and M.S. Searle, *Structure, Folding, and Energetics of Cooperative Interactions between the beta-Strands of a de Novo Designed Three-Stranded Antiparallel beta-Sheet Peptide.* J. Am. Chem. Soc. **122**: p. 8350-8356.

13. Burley, S.K. and G.A. Petsko, *Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization.* Science, 1985. **229**: p. 23-28.

14. Zheng, X. and K. Sohlberg, *Modeling bistability and switching in a [2]catenane.* Phys. Chem. Chem. Phys., 2004. **6**: p. 809 - 815.

15. Nepogodiev, S.A. and J.F. Stoddart, *Cyclodextrin-Based Catenanes and Rotaxanes.* Chem. Rev., 1998. **98**: p. 1959-1976.

16. Foster, M.E. and K. Sohlberg, *Theoretical Study of Binding Site Preference in [2]Rotaxanes.* J. Chem. Theory Comput., 2007. **3**: p. 2221-2233.

17. Zheng, X. and K. Sohlberg, *Modeling of a Rotaxane-based Molecular Device.* J. Phys. Chem., 2003. **107**: p. 1207.

18. Dewar, M.J.S., E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, *AM1: A new general purpose quantum mechanical molecular model.* J. Am. Chem. Soc., 1985. **107**: p. 3902.

19. Stewart, J.J.P., *Optimization of parameters for semiempirical methods I. Method.* J. Comp. Chem., 1989. **10**: p. 209-220.

20. Rocha, G.B., R.O. Freire, A.M. Simas, and J.J.P. Stewart, *RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I.* J. Comp. Chem., 2006. **27**: p. 1101-1111.

21. Stewart, J.J.P., *Optimization fo parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements.* J. Mol. Model, 2007. **13**: p. 1173-1213.

22. McNamara, J.P. and I.H. Hillier, *Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules.* Phys. Chem. Chem. Phys., 2007. **9**: p. 2362-2370.

23. Řezáč, J., J. Fanfrlik, D. Salahub, and P. Hobza, *Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes.* J. Chem. Theory Comput., 2009. **5**(7): p. 1749-1760.

24. Levine, I.N., *Quantum Chemistry*. 5 ed1999: Prentice Hall. 739.

25. Zhao, Y., N.E. Schultz, and D.G. Truhlar, *Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions.* J. Chem. Theory Comput., 2006. **2**: p. 364.

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

27. Szabo, A. and N.S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*1996: Dover Publications, Inc.

28. Maitland, G.C., M. Rigby, E.B. Smith, and W.A. Wakeham, *Intermolecular Forces: Their Origin and Determination*1981, New York: Oxford University Press.

29. Cramer, C.J., *Essentials of Computational Chemistry: Theories and Models*2002, Chichester: John Wiley & Sons, Ltd.

30. Goodsell, D.S. and A.J. Olson, *Automated Docking of Substrates to Proteins by Simulated Annealing.* Proteins: Str. Func. and Genet., 1990. **8**: p. 195-202.

31. Wu, Q. and W. Yang, *Empirical correction to density functional theory for van der Waals interactions.* J. Chem. Phys., 2002. **116**: p. 515-524.

32. Su, P. and H. Li, *Energy decomposition analysis of covalent bonds and intermolecular interactions.* J. Chem. Phys., 2009. **131**(014102): p. 1-15.

33. Jeziorski, B., R. Moszynski, and K. Szalewicz, *Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes.* Chem. Rev., 1994. **94**: p. 1887-1930.

34. Grimme, S., *Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections.* J. Comput. Chem., 2004. **25**: p. 1463-1473.

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

36. Jurečka, P., J. Cerny, P. Hobza, and D. Salahub, *Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with Ab Initio Quantum Mechanics Calculations.* J. Comp. Chem., 2007. **28**: p. 555-569.

37. Slater, J.C. and J.G. Kirkwood, *The van der Waals Forces in Gases.* Phys. Rev., 1931. **37**: p. 682-697.

38. Halgren, T.A., *Representation of van der Waals (vdW) Interaction in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters* J. Am. Chem. Soc., 1992. **114**: p. 7827-7843.

39. Antony, J. and S. Grimme, *Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules.* Phys. Chem. Chem. Phys., 2006. **8**: p. 5287-5293.

40. Antony, J. and S. Grimme, *Structures and interaction energies of stacked graphene–nucleobase complexes.* Phys. Chem. Chem. Phys., 2008. **10**: p. 2722-2729.

41. Parac, M., M. Etinski, M. Peric, and S. Grimme, *A Theoretical Investigation of the Geometries and Binding Energies of Molecular Tweezer and Clip Host-Guest Systems.* J. Chem. Theory Comput., 2005. **1**: p. 1110-1118.

42. van Mourik, T., *Assessment of Density Functionals for Intramolecular Dispersion-Rich Interactions.* J. Chem. Theory Comput., 2008. **4**: p. 1610-1619.

43. Morgado, C.A., J.P. McNamara, I.H. Hillier, N.A. Burton, and M.A. Vincent, *Density Functional and Semiempirical Molecular Orbital Methods Including Dispersion Corrections for the Accurate Description of Noncovalent Interactions Involving Sulfur-Containing Molecules.* J. Chem. Theory Comput., 2007. **3**: p. 1656-1664.

44. Kerber, T., M. Sierka, and J. Sauer, *Application of Semiempirical Long-Range Dispersion Corrections to Periodic Systems in Density Functional Theory.* J. Comp. Chem., 2008. **29**: p. 2088-2097.

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

46. Schwabe, T. and S. Grimme, *Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability.* Phys. Chem. Chem. Phys., 2007. **9**: p. 3397-3406.

47. Grimme, S., *Semiempirical hybrid density functional with perturbative second-order correlation.* J. Chem. Phys., 2006. **124**: p. 034108-1-16.

48. Morgado, C., M.A. Vincent, I.H. Hillier, and X. Shan, *Can the DFT-D method describe the full range of noncovalent interactions found in large biomolecules?* Phys. Chem. Chem. Phys., 2007. **9**: p. 448-451.

49. Sharma, R., J.P. McNamara, R.K. Raju, M.A. Vincent, I.H. Hillier, and C.A. Morgado, *The interaction of carbohydrates and amino acids with aromatic systems studied by density functional and semi-empirical molecular orbital calculations with dispersion correctionsw.* Phys. Chem. Chem. Phys., 2008. **10**: p. 2767-2774.

50. Voityuk, A.A. and N. Rosch, *AM1/d Parameters for Molybdenum.* J. Phys. Chem. A, 2000. **104**: p. 4089-4094.

51. Sinnokrot, M.O. and C.D. Sherrill, *Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations.* J. Phys. Chem. A, 2004. **108**: p. 10200-10207.

52. Dewar, M.J.S. and W. Thiel, *Ground states of molecules. 38. The MNDO method. Approximations and parameters.* J. Am. Chem. Soc., 1977. **99**: p. 4899-4907.

53. Foster, M.E. and K. Sohlberg, *Empirically corrected DFT and semi-empirical methods for non-bonding interactions.* Phys. Chem. Chem. Phys., 2010. **12**: p. 307-322.

54. Schmidt, M.W., K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, and J.A. Montgomery, *General Atomic and Molecular Electronic Structure System.* Journal of Computational Chemistry, 1993. **14**: p. 1347-1363.

55. Chirgwin, H. and C.A. Coulson, *The Electronic Structure of Conjugated Systems. VI.* Proc. R. Soc. Lond. A, 1950. **201**: p. 196-209.

56. Buckingham, A.D., J.E.D. Bene, and S.A.C. McDowell, *The hydrogen bond.* Chem. Phys. Lett., 2008. **463**: p. 1-10.

57. Riley, K.E. and P. Hobza, *Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes.* J. Phys. Chem. A, 2007. **111**: p. 8257-8263.

58. Cole, S.J., K. Szalewicz, G.D.P. III, and R.J. Bartlett, *Correlated calculation of the interaction inthe nitromethane dimer.* J. Chem. Phys., 1986. **84**(12): p. 6833-6836.

59. Podeszwa, R., R. Bukowski, and K. Szalewicz, *Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of pi-pi Interactions.* J. Phys. Chem. A, 2006. **110**: p. 10345-10354.

60. Tekin, A. and G. Jansen, *How accurate is the density functional theory combined with symmetry-adapted perturbation theory approach for CH-π and π-π interactions? A comparison to supermolecular calculations for the acetylene-benzene dimer.* Phys. Chem. Chem. Phys., 2007. **9**: p. 1680-1687.

61. Rybak, S., B. Jeziorski, and K. Szalewicz, *Many-body symmetry-adapted perturbation theory of intermolecular interactions. H20 and HF dimers.* J. Chem. Phys., 1991. **95**(9): p. 6576-6601.

62. Pedley, J.B. and J. Rylance, *Sussex-N.P.L. Computer Analysed Thermochemical Data: Organic and Organometallic Compounds*1977, University of Sussex.

63. Grover, J.R., *Dissociation Energies of the Benzene Dimer and Dimer Cation.* J. Phys. Chem., 1987. **91**: p. 3233-3237.

64. Zhao, Y. and D.G. Truhlar, *Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions.* J. Chem. Theory Comput., 2008. **4**: p. 1849-1868.

65. Sygula, A., F.R. Fronczek, R. Sygula, P.W. Rabideau, and M.M. Olmstead, *A Double Concave Hydrocarbon Buckycatcher.* J. Am. Chem. Soc., 2007. **129**: p. 3842-3843.

66. Zhao, Y. and D.G. Truhlar, *Size-selective supramolecular chemistry in a hydorcarbon nanoring.* J. Am. Chem. Soc., 2007. **129**: p. 8440-8442.

67. Zhao, Y. and D.G. Truhlar, *Computational Characterization and modeling of buckyball tweezers: density functional study of concave-convex pi---pi interactions.* Phys. Chem. Chem. Phys., 2008. **10**: p. 2813-2818.

68. Romero, C., L. Fomina, and S. Fomone, *How important is the dispersion interaction for cyclobis(paraquat-p-phenylene)-based molecular "shuttles"? A theoretical study.* Int. J. Quant. Chem., 2005. **102**: p. 200-208.

69. Ji, H.F., R. Majithia, X. Yang, X. Xu, and K. More, *Self-Assembly of Perylenediimide and Naphthalenediimide Nanostructures on Glass Substrates through Deposition from the Gas Phase.* J. Am. Chem. Soc. , 2008. **130**: p. 10056-10057.

70. Yeung, C.S., W.Q. Tian, L.V. Liu, and Y.A. Wang, *Chemistry of Single-Walled Carbon Nanotubes* Journal of Computational and Theoretical Nanoscience, 2009. **6**(6): p. 1213-1235.

71. Kang, J.W., H.J. Hwang, and Q. Jiang, *A Molecular Dynamics Study on Oscillation of a Carbon Nanotube Inside an Encapsulating Boron-Nitride Nanotube.* Journal of Computational and Theoretical Nanoscience, 2006. **3**(6): p. 880-884.

72. Slanina, Z., F. Uhlik, S.-L. Lee, L. Adamowicz, and S. Nagase, *MPWB1K calculations of stepwise encapsulations: Lix@C60.* Chem. Phys. Lett., 2008. **463**: p. 121-123.

73. Slanina, Z., P. Pulay, and S. Nagase, *H2, Ne, and N2 Energies of Encapsulation into C60 Evaluated with the MPWB1K Functional.* J. Chem. Theory Comput., 2006. **2**: p. 782-785.

74. Zhao, Y. and D.G. Truhlar, *Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van der Waals Interactions.* J. Phys. Chem. A, 2004. **108**: p. 6908-6918.

75. Foster, M.E. and K. Sohlberg, *A New Empirical Correction to the AM1 Method for Macromolecular Complexes.* J. Chem. Theory Comput., 2010. **6**(7): p. 2153-2166.

76. Kawase, T. and M. Oda, *Complexation of carbon nanorings with fullerenes.* Pure and Applied Chemistry, 2006. **78**(4): p. 831–839.

77. Kawase, T., K. Tanaka, N. Fujiwara, H.R. Darabi, and M. Oda, *Complexation of a Carbon Nanoring with Fullerenes.* Angew. Chem. - Int. Ed., 2003. **42**: p. 1624 – 1628.

78. Kawase, T., K. Tanaka, Y. Seirai, N. Shiono, and M. Oda, *Complexation of Carbon Nanorings with Fullerenes: Supramolecular Dynamics and Structural Tuning for a Fullerene Sensor.* 2003. **42**: p. 5597 –5600.

79. Kawase, T., N. Fujiwara, M. Tsutumi, M. Oda, Y. Maeda, T. Wakahara, and T. Akasaka, *Supramolecular Dynamics of Cyclic [6]Paraphenyleneacetylene Complexes with [60]- and [70]Fullerene Derivatives: Electronic and Structural Effects on Complexation.* Angew. Chem. - Int. Ed., 2004. **43**: p. 5060 –5062.

80. Perez, E.M. and N. Martin, *Curves ahead: molecular receptors for fullerenes based on concave–convex complementarity.* Chem. Soc. Rev., 2008. **37**: p. 1512–1519.

81. Cuesta, G.I., T.B. Pedersen, H. Koch, and A. Sanchez de Meras, *Carbon Nanorings: A Challenge to Theoretical Chemistry.* Chemphyschem, 2006. **7**: p. 2503 – 2507.

82. Hyperchem, 2000, Hypercube Inc.: Waterloo, Ontario, Canada.

83. Yoshida, M. *Fullerene Structure Library*. 1998; Available from: http://www.cochem2.tutkie.tut.ac.jp/Fuller/fsl/fsl.html.

84. *Nanotube Modeler*, 2005-2010, JCrystalSoft.

85. Slanina, Z., X. Zhao, N. Kurita, H. Gotoh, F. Uhlık, J.M. Rudzinski, K.H. Lee, and L. Adamowicz, *Computing the relative gas-phase populations of C60 and C70: Beyond the traditional $H_{f,298}$ scale.* J. Mol. Graphics Modell., 2001. **19**(2): p. 216-221.

86. Grimme, S., *Accurate description of van der Waals complexes by density functional theory including empirical corrections.* J. Comput. Chem., 2004. **25**: p. 1463-1476.

87. Tsuzuki, S., K. Honda, T. Uchimaru, and M. Mikami, *High-level ab initio computations of structures and interaction energies of naphthalene dimers: Origin of attraction and its directionality.* J. Chem. Phys., 2004. **120**: p. 647-659.

88. Podeszwa, R., *Interactions of graphene sheets deduced from properties of polycyclic aromatic hydrocarbons.* J. Chem. Phys., 2010. **132**(044704): p. 1-8.

89. Delhaes, P., *Graphite and Precursors*2001, Amsterdam: Gordon and Breach Science.

90. Collier, C.P., E.W. Wong, M. Belohradsky, F.M. Raymo, J.F. Stoddart, P.J. Kuekes, R.S. Williams, and J.R. Heath, *Electronically configurable molecular-based logic gates.* Science, 1999. **285**(5426): p. 391-394.

91. Heath, J.R. and M.A. Ratner, *Molecular electronics.* Physics Today, 2003. **56**(5): p. 43-49.

92. Flood, A.H., J.F. Stoddart, D.W. Steuerman, and J.R. Heath, *Whence Molecular Electronics?* Science, 2004. **306**: p. 2055-2056.

93. Green, J.E., J.W. Choi, A. Boukai, Y. Bunimovich, E. Johnston-Halperin, E. DeIonno, Y. Luo, B.A. Sheriff, K. Xu, Y.S. Shin, H.-R. Tseng, J.F. Stoddart, and J.R. Heath, *A 160-kilobit molecular electronic memory patterned at 10^{11} bits per square centimetre.* Nature, 2007. **445**: p. 414-417.

94. Luo, Y., C.P. Collier, J.O. Jeppesen, K.A. Nielsen, E. DeIonno, G. Ho, J. Perkins, H.-R. Tseng, T. Yamamoto, J.F. Stoddart, and J.R. Heath, *Two-dimensional molecular electronics circuits.* Chem. Phys. Chem., 2002. **3**: p. 519-525.

95. Choi, J.W., A.H. Flood, D. Steuerman, S. Nygaard, A.B. Braunschweig, N.N.P. Moonen, B.W. Laursen, Y. Luo, E. DeIonno, A.J. Peters, J.O. Jeppesen, K. Xu, J.F. Stoddart, and J.R. Heath, *Ground-State Equilibrium Thermodynamics and Switching Kinetics of Bistable [2]Rotaxanes Switched in Solution, Polymer Gels, and Molecular Electronic Devices.* Chem. Eur. J., 2006. **12**: p. 261-279.

96. Collier, C.P., J.O. Jeppesen, J.P. Yi Luo, E.W. Wong, J.R. Heath, and J.F. Stoddart, *Molecular-Based Electronically Switchable Tunnel Junction Devices.* J. Am. Chem. Soc., 2001. **123**: p. 12632-12641.

97. Dichtel, W.R., J.R. Heath, and J.F. Stoddart, *Designing bistable [2]rotaxanes for molecular electronic devices.* Phil. Trans. R. Soc. A, 2007. **365**: p. 1607-1625.

98. Beckman, R., K. Beverly, A. Boukai, Y. Bunimovich, J.W. Choi, E. DeIonno, J. Green, E. Johnston-Halperin, Y. Luo, B. Sheriff, J.F. Stoddart, and J.R. Heath, *Spiers Memorial Lecture - Molecular mechanics and molecular electronics.* Faraday Discussions, 2006. **131**: p. 9-22.

99. Liu, Y., A.H. Flood, P.A. Bonvallet, S.A. Vignon, B.H. Northrop, H.-R. Tseng, J.O. Jeppesen, T.J. Huang, B. Brough, M. Baller, S. Magonov, S.D. Solares, W.A. Goddard, C.-M. Ho, and J.F. Stoddart, *Linear Artificial Molecular Muscles.* J. Am. Chem. Soc., 2005. **127**: p. 9745-9759.

100. Deng, W.-Q., R.P. Muller, and W.A. Goddard, *Mechanism of the Stoddart-Heath Bistable Rotaxane Molecular Switch.* J. Am. Chem. Soc., 2004. **126**: p. 13562-13563.

101. Jang, Y.H., S. Hwang, Y.-H. Kim, S.S. Jang, and W.A.G. III, *Density Functional Theory Studies of the [2]Rotaxane Component of the Stoddart-Heath Molecular Switch.* J. Am. Chem. Soc., 2004. **126**: p. 12636-12645.

102. Foster, M.E. and K. Sohlberg, *Empirically corrected DFT and semi-empirical methods for non-bonding interactions.* Phys. Chem. Chem. Phys., 2009.

103. Foster, M.E. and K. Sohlberg, *Designing Fullerene Separation Materials: A Theoretical Study.* Fullerenes, Nanotubes and Carbon Nanostructures, 2011.

104. *HyperChem(TM) Professional 8.0.4, Hypercube, Inc.*: 1115 NW 4th Street, Gainesville, Florida 32601, USA.

105. Schmidt, M.W., K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S.J. Su, T.L. Windus, M. Dupuis, and J.A. Montgomery, *General Atomic and Molecular Electronic Structure System.* J. Comput. Chem., 1993 **14**: p. 1347-1363.

106. Valiev, M., E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J.v. Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, and W.A.d. Jong, *NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations.* Comput. Phys. Commun., 2010. **181**(9): p. 1477-1489.

107. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple.* Phys. Rev. Lett., 1996. **77**(18): p. 3865-3868.

108. Phoa, K., J.B. Neaton, and V. Subramanian, *First-Principles Studies of the Dynamics of [2]Rotaxane Molecular Switches.* Nano Lett., 2009. **9**: p. 3225-3229.

109. Rossi, M. and K. Sohlberg, *Origin of the redox-induced conductance transition in a thin film of switchable rotaxanes.* J. Comput. Theor. Nanosci., 2011.

110. Deng, W.-Q. and W.A. Goddard, *Predictions of Hole Mobilities in Oligoacene Organic Semiconductors from Quantum Mechanical Calculations*. J. Phys. Chem. B, 2004. **108**(25): p. 8614-8621.

111. Pitonak, M., K. E. Riley, P. Neograd y, and P. Hobza, *Highly Accurate CCSD(T) and DFT-SAPT Stabilization Energies of H-Bonded and Stacked Structures of the Uracil Dimer.* ChemPhysChem, 2008, **9**: p. 1636-1644.

Appendix A: S22 Database Complexes

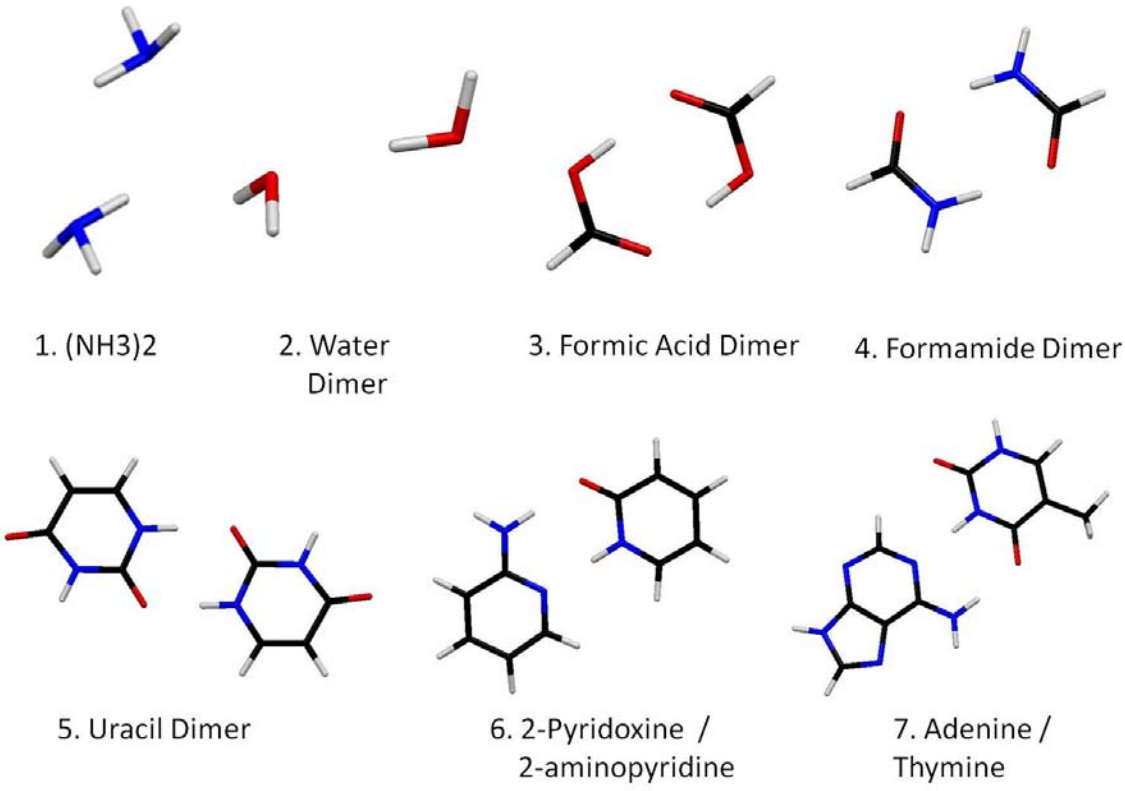


Figure A-1: Hydrogen-bonded complexes in the S22 database

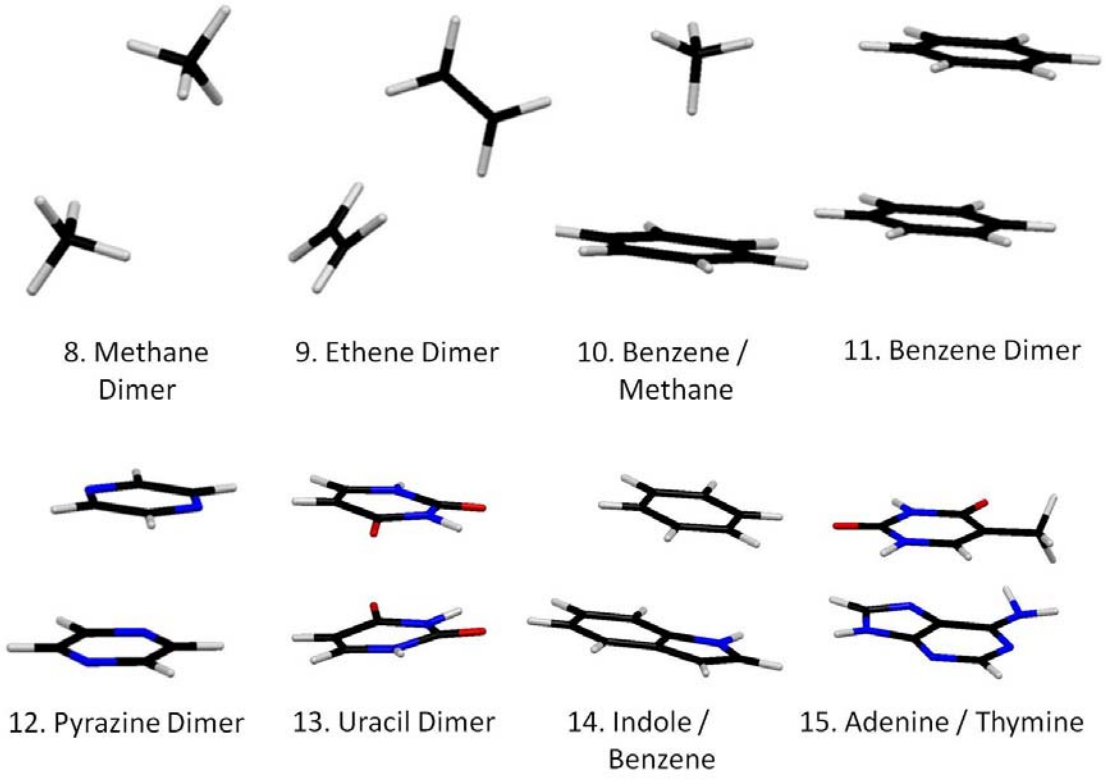


Figure A-2: Dispersion bonded complexes in the S22 database

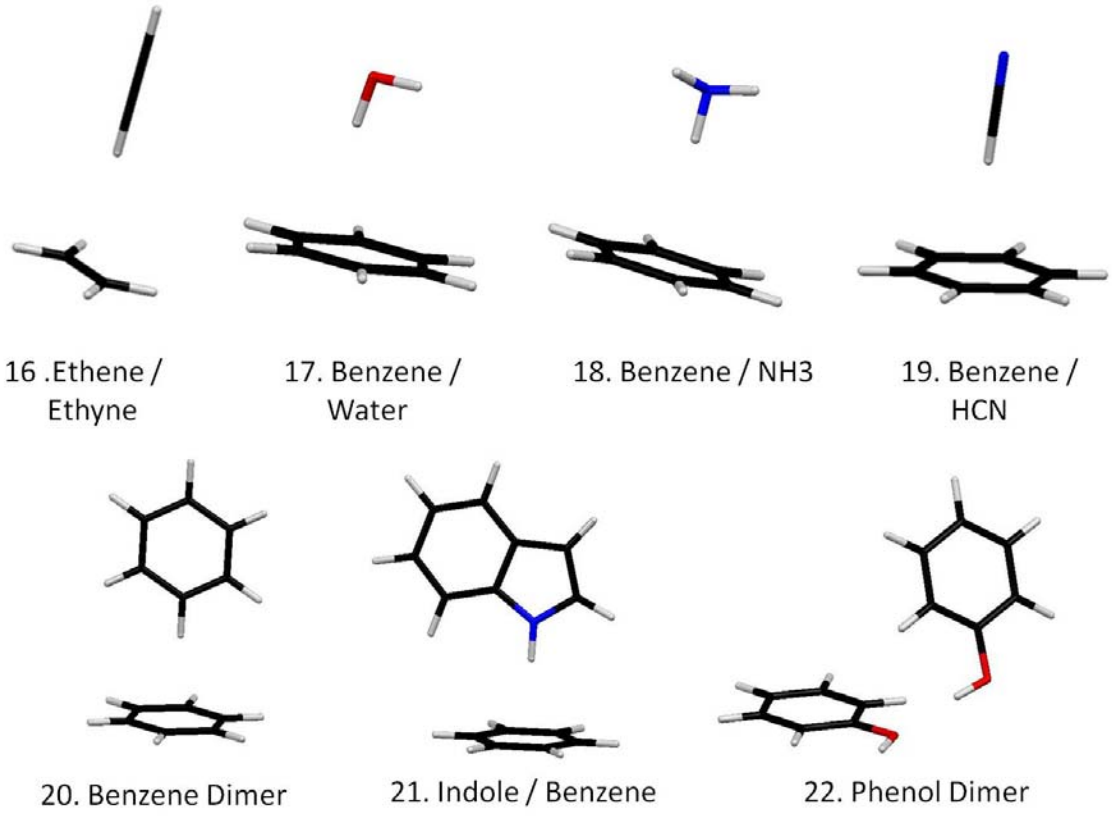


Figure A-3: Mixed bonded complexes in the S22 database

Appendix B: Supporting Information for Chapter 2

Table B-1: Single-point interaction energies (kcal/mol) at the S22 geometries.

No.	Molecule (symmetry)	Ref. Values	B97-D / TZV(2df,2pd)	BLYP-D / TZV(2df,2pd)	B2PLYP-D/TZVPP
Hydrogen bonded complexes					
1	(NH3)2 (C2h)	-3.17	-3.72	-4.16	-3.65
2	(H2O)2 (Cs)	-5.02	-5.07	-5.80	-5.47
3	Formic acid dimer (C2h)	-18.61	-18.25	-19.34	-19.43
4	Formamide dimer (C2h)	-15.96	-15.28	-16.39	-16.37
5	Uracil dimer (C2h)	-20.65	-19.45	-20.73	-20.88
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	-17.13	-18.05	-17.60
7	Adenine thymine WC (C1)	-16.37	-16.20	-17.19	-16.85
Complexes with predominant dispersion contribution					
8	(CH4)2 (D3d)	-0.53	-0.57	-0.36	-0.40
9	(C2H4)2 (D2d)	-1.51	-1.55	-1.55	-1.50
10	Benzene CH4 (C3)	-1.50	-1.51	-1.37	-1.42
11	Benzene dimer (C2h)	-2.73	-2.67	-2.35	-2.52
12	Pyrazine dimer (Cs)	-4.42	-4.07	-4.05	-4.26
13	Uracil dimer (C2)	-10.12	-10.02	-10.50	-10.14
14	Indole benzene (C1)	-5.22	-4.72	-4.55	-4.69
15	Adenine thymine stack (C1)	-12.23	-12.11	-12.85	-12.52
Mixed complexes					
16	Ethene ethyne (C2v)	-1.53	-1.73	-1.62	-1.60
17	Benzene H2O (Cs)	-3.28	-4.14	-4.16	-3.66
18	Benzene NH3 (Cs)	-2.35	-2.75	-2.66	-2.47
19	Benzene HCN (Cs)	-4.46	-4.88	-4.87	-4.96
20	Benzene dimer (C2v)	-2.74	-2.93	-2.76	-2.82
21	Indole benzene T-shape (C1)	-5.73	-6.26	-6.16	-6.04
22	Phenol dimer (C1)	-7.05	-6.60	-7.35	-7.28
	RMSE (Hydrogen bonded)		0.60	0.83	0.58
	RMSE (Dispersion bonded)		0.22	0.40	0.24
	RMSE (Mixed bonded)		0.48	0.43	0.29
	RMSE		0.46	0.58	0.39
	MUE		0.35	0.47	0.31

Table B-2: Single-point interaction energies (kcal/mol) at the S22 geometries. LP = 6-311++G(3df,3pd).

No.	Molecule (symmetry)	Ref. Values	TPSS / LP	TPSS / LP CP	TPSS-D / LP	TPSS-D / LP CP
Hydrogen bonded complexes						
1	(NH3)2 (C2h)	-3.17	-2.3	-2.2	-3.0	-3.1
2	(H2O)2 (Cs)	-5.02	-4.9	-4.4	-5.5	-5.2
3	Formic acid dimer (C2h)	-18.61	-18.3	-17.5	-19.8	-19.2
4	Formamide dimer (C2h)	-15.96	-14.4	-13.9	-16.1	-15.9
5	Uracil dimer (C2h)	-20.65	-18.2	-17.6	-20.5	-20.2
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	-14.6	-14.2	-17.5	-17.5
7	Adenine thymine WC (C1)	-16.37	-13.6	-13.1	-16.7	-16.7
Complexes with predominant dispersion contribution						
8	(CH4)2 (D3d)	-0.53	0.2	0.2	-0.7	-0.7
9	(C2H4)2 (D2d)	-1.51	0.2	0.3	-1.8	-2.5
10	Benzene CH4 (C3)	-1.50	0.1	0.5	-1.7	-1.5
11	Benzene dimer (C2h)	-2.73	2.1	3.0	-3.3	-2.6
12	Pyrazine dimer (Cs)	-4.42	1.3	1.9	-4.2	-3.9
13	Uracil dimer (C2)	-10.12	-2.2	-1.0	-9.8	-9.2
14	Indole benzene (C1)	-5.22	2.4	3.7	-5.4	-4.5
15	Adenine thymine stack (C1)	-12.23	-0.6	0.8	-11.9	-11.4
Mixed complexes						
16	Ethene ethyne (C2v)	-1.53	-1.0	-0.8	-1.7	-1.3
17	Benzene H2O (Cs)	-3.28	-2.1	-1.4	-3.6	-3.2
18	Benzene NH3 (Cs)	-2.35	-0.8	-0.3	-2.5	-2.3
19	Benzene HCN (Cs)	-4.46	-2.8	-2.2	-4.4	-4.1
20	Benzene dimer (C2v)	-2.74	0.0	0.6	-2.6	-2.3
21	Indole benzene T-shape (C1)	-5.73	-1.8	-1.0	-5.4	-4.8
22	Phenol dimer (C1)	-7.05	-3.5	-2.9	-6.9	-6.7
	RMSE (Hydrogen bonded)		1.7	2.2	0.6	0.4
	RMSE (Dispersion bonded)		6.3	7.2	0.3	0.6
	RMSE (Mixed bonded)		2.5	3.0	0.2	0.4
	RMSE		4.2	4.8	0.4	0.5
	MUE		3.0	3.7	0.3	0.4

Table B-3: Geometry optimized energies (kcal/mol), interaction distances (Angstroms), and CM-distance (Angstroms) for the complexes in the S22 database. The interaction distance is defined as the CM-distance for complexes 11-15 and the 2nd distance reported for complex 22.

		Interaction Energies (kcal/mol)		Interaction Distances (Å)		CM-Distances (Å)	
No.	Molecule (symmetry)	Ref. Values	BLYP-D / TZV(2d,2p)	Ref. Values	BLYP-D / TZV(2d,2p)	Ref. Values	BLYP-D / TZV(2d,2p)
Hydrogen bonded complexes							
1	(NH3)2 (C2h)	-3.17	-3.58	2.504	2.595	3.209	3.305
2	(H2O)2 (Cs)	-5.02	-5.98	1.952	1.961	2.909	2.927
3	Formic acid dimer (C2h)	-18.61	-16.62	1.670	1.674	2.993	3.017
4	Formamide dimer (C2h)	-15.96	-14.91	1.841	1.864	3.229	3.267
5	Uracil dimer (C2h)	-20.65	-18.41	1.775	1.789	6.075	6.127
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	-16.20	1.859, 1.874	1.846, 1.851	5.136	5.165
7	Adenine thymine WC (C1)	-16.37	-15.48	1.819, 1.929	1.786, 1.908	5.974	5.99
Complexes with predominant dispersion contribution							
8	(CH4)2 (D3d)	-0.53	-0.32	3.718	3.717	3.718	3.716
9	(C2H4)2 (D2d)	-1.51	-1.39	3.718	3.712	3.718	3.712
10	Benzene CH4 (C3)	-1.50	-0.94	3.716	3.732	3.716	3.732
11	Benzene dimer (C2h)	-2.73	-2.01	3.765	3.775	3.765	3.772
12	Pyrazine dimer (Cs)	-4.42	-3.76	3.479	3.631	3.479	3.628
13	Uracil dimer (C2)	-10.12	-9.01	3.166	2.807	3.166	3.223
14	Indole benzene (C1)	-5.22	-4.20	3.498	3.574	3.498	3.724
15	Adenine thymine stack (C1)	-12.23	-11.20	3.172	3.259	3.172	3.268
Mixed complexes							
16	Ethene ethyne (C2v)	-1.53	-1.46	2.752	2.755	4.422	4.428
17	Benzene H2O (Cs)	-3.28	-3.72	2.531	3.330	3.380	3.262
18	Benzene NH3 (Cs)	-2.35	-2.14	3.592	3.628	3.560	3.596
19	Benzene HCN (Cs)	-4.46	-4.08	3.387	3.510	3.950	4.068
20	Benzene dimer (C2v)	-2.74	-2.19	3.513	3.651	4.909	5.053
21	Indole benzene T-shape (C1)	-5.73	-5.09	3.210	3.409	4.884	5.024
22	Phenol dimer (C1)	-7.05	-6.94	1.937, 4.921	1.921, 5.084	4.921	5.084
	RMSE (Hydrogen bonded)		1.32		0.035		0.047
	RMSE (Dispersion bonded)		0.76		0.144		0.104
	RMSE (Mixed bonded)		0.40		0.304		0.117
	RMSE		0.90		0.192		0.095
	MUE		0.72		0.097		0.071

Table B-4: Geometry optimized interaction energies (kcal/mol) and CM-distances (Angstroms) for the complexes in the S22 database. LP = 6-311++G(3df,3pd). The interaction distance is defined as the CM-distance in complexes 11-15 and 2nd value reported for complex 22.

		Ref. Values		TPSS / LP		TPSS-D / LP		
		ΔE	CM Distance	ΔE	CM-Distance	ΔE	CM-Distance	Interaction Distances
Hydrogen bonded complexes								
1	(NH3)2 (C2h)	-3.17	3.209	-2.4	3.257	-3.1	3.324	2.635
2	(H2O)2 (Cs)	-5.02	2.909	-5.0	2.891	-5.6	2.897	1.928
3	Formic acid dimer (C2h)	-18.61	2.993	-20.7	2.964	-22.4	2.955	1.608
4	Formamide dimer (C2h)	-15.96	3.229	-15.3	3.210	-17.0	3.205	1.800
5	Uracil dimer (C2h)	-20.65	6.075	-19.1	6.074	-21.5	6.055	1.731
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	5.136	-15.9	5.134	-18.9	5.104	1.804,1.800
7	Adenine thymine WC (C1)	-16.37	5.974	-14.6	5.969	-17.9	5.938	1.752, 1.851
Complexes with predominant dispersion contribution								
8	(CH4)2 (D3d)	-0.53	3.718	0.0	3.923	-0.7	3.719	3.719
9	(C2H4)2 (D2d)	-1.51	3.718	-0.3	4.202	-1.8	3.718	3.735
10	Benzene CH4 (C3)	-1.50	3.716	0.1	3.723	-1.7	3.721	3.721
11	Benzene dimer (C2h)	-2.73	3.765	0.0	4.638	-3.3	3.769	3.769
12	Pyrazine dimer (Cs)	-4.42	3.479	-0.6	4.322	-4.3	3.575	3.575
13	Uracil dimer (C2)	-10.12	3.166	-3.5	3.800	-10.2	3.172	3.172
14	Indole benzene (C1)	-5.22	3.498	-0.7	4.637	-5.4	3.515	3.515
15	Adenine thymine stack (C1)	-12.23	3.172	-6.7	4.525	-12.0	3.224	3.224
Mixed complexes								
16	Ethene ethyne (C2v)	-1.53	4.422	-1.0	4.445	-1.6	4.437	2.768
17	Benzene H2O (Cs)	-3.28	3.380	-2.2	3.544	-3.6	3.386	2.531
18	Benzene NH3 (Cs)	-2.35	3.560	-1.0	3.845	-2.5	3.567	3.599
19	Benzene HCN (Cs)	-4.46	3.950	-2.8	3.956	-4.4	3.951	3.395
20	Benzene dimer (C2v)	-2.74	4.909	-0.8	5.411	-2.9	5.134	3.738
21	Indole benzene T-shape (C1)	-5.73	4.884	-2.5	5.279	-5.3	4.891	3.241
22	Phenol dimer (C1)	-7.05	4.921	-4.4	5.420	-7.2	4.983	1.924,4.983
	RMSE (Hydrogen bonded)			1.3	0.023	1.8	0.051	0.070
	RMSE (Dispersion bonded)			3.9	0.812	0.3	0.039	0.040
	RMSE (Mixed bonded)			2.0	0.331	0.2	0.089	0.084
	RMSE			2.7	0.524	1.1	0.062	0.067
	MUE			2.1	0.342	0.6	0.036	0.045

Table B-5: Single-point interaction energies (kcal/mol) at the S22 geometries. [a] AM1-D results without re-parameterization of AM1 method (S_6=1.1 and d=23.0).

No.	Molecule (symmetry)	Ref. Values	AM1	PM3	AM1-D[a]	AM1-D	PM3-D	PM3-D*
Hydrogen bonded complexes								
1	(NH3)2 (C2h)	-3.17	-0.78	0.77	-2.35	-3.43	-1.77	-0.67
2	(H2O)2 (Cs)	-5.02	-2.89	-2.79	-3.74	-7.29	-5.14	-4.53
3	Formic acid dimer (C2h)	-18.61	1.54	-9.91	-1.24	-15.45	-18.57	-17.46
4	Formamide dimer (C2h)	-15.96	-12.02	-8.08	-8.67	-17.16	-15.37	-10.32
5	Uracil dimer (C2h)	-20.65	-5.79	-11.32	-9.48	-20.15	-20.30	-19.88
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	-4.45	-7.46	-8.84	-16.50	-17.52	-13.50
7	Adenine thymine WC (C1)	-16.37	-4.28	-6.79	-8.95	-16.58	-17.33	-13.89
Complexes with predominant dispersion contribution								
8	(CH4)2 (D3d)	-0.53	0.21	-0.25	-0.70	-0.94	-1.24	-1.00
9	(C2H4)2 (D2d)	-1.51	-0.13	-1.11	-2.48	-3.31	-3.60	-2.44
10	Benzene CH4 (C3)	-1.50	0.40	-0.19	-2.00	-2.12	-2.42	-1.83
11	Benzene dimer (C2h)	-2.73	3.52	2.38	-2.79	-2.90	-4.30	-4.41
12	Pyrazine dimer (Cs)	-4.42	2.49	3.90	-4.38	-4.57	-4.20	-4.83
13	Uracil dimer (C2)	-10.12	0.12	5.80	-9.87	-10.56	-6.78	-10.78
14	Indole benzene (C1)	-5.22	5.39	4.04	-4.04	-4.04	-6.09	-5.68
15	Adenine thymine stack (C1)	-12.23	2.91	7.37	-11.74	-12.20	-10.63	-11.58
Mixed complexes								
16	Ethene ethyne (C2v)	-1.53	-0.35	-0.82	-1.45	-1.61	-1.85	-1.41
17	Benzene H2O (Cs)	-3.28	-0.69	-1.47	-3.29	-3.43	-3.65	-2.71
18	Benzene NH3 (Cs)	-2.35	-0.33	-0.59	-2.85	-3.00	-2.96	-1.97
19	Benzene HCN (Cs)	-4.46	-0.81	-1.63	-4.27	-4.44	-4.43	-3.05
20	Benzene dimer (C2v)	-2.74	0.37	-0.43	-3.68	-3.85	-4.15	-2.90
21	Indole benzene T-shape (C1)	-5.73	-1.05	-1.25	-6.96	-7.10	-6.65	-4.60
22	Phenol dimer (C1)	-7.05	-1.36	-1.37	-5.86	-9.76	-7.52	-5.57
	RMSE (Hydrogen bonded)		11.64	7.77	9.25	1.56	0.76	2.85
	RMSE (Dispersion bonded)		8.21	10.13	0.61	0.82	1.68	0.81
	RMSE (Mixed bonded)		3.57	3.22	0.77	1.25	0.72	0.92
	RMSE		8.47	7.73	5.25	1.23	1.18	1.76
	MUE		6.54	5.94	2.77	0.85	0.90	1.23

Table B-6: Geometry optimized interaction energies (kcal/mol) for the S22 complexes.

No.	Molecule (symmetry)	Ref. Values	AM1	PM3	AM1-D	PM3-D	PM3-D*
Hydrogen bonded complexes							
1	(NH3)2 (C2h)	-3.17	-1.39	-0.71	-3.03	-1.99	-2.12
2	(H2O)2 (Cs)	-5.02	-3.30	-3.55	-7.22	-6.53	-4.04
3	Formic acid dimer (C2h)	-18.61	-6.62	-9.58	-12.45	-16.16	-19.71
4	Formamide dimer (C2h)	-15.96	-2.06	-6.99	-14.64	-14.42	-10.05
5	Uracil dimer (C2h)	-20.65	-10.48	-10.70	-17.80	-18.83	-22.19
6	2-Pyridoxine2-aminopyridine (C1)	-16.71	-6.15	-7.06	-13.06	-18.32	-13.70
7	Adenine thymine WC (C1)	-16.37	-5.06	-6.90	-12.66	-18.66	-16.29
Complexes with predominant dispersion contribution							
8	(CH4)2 (D3d)	-0.53	-0.21	-0.32	-4.10	-2.38	-1.14
9	(C2H4)2 (D2d)	-1.51	-0.13	-1.08	-4.85	-4.11	-2.37
10	Benzene CH4 (C3)	-1.50	0.35	-0.20	-2.93	-2.88	-1.84
11	Benzene dimer (C2h)	-2.73	0.01	-0.02	-3.10	-4.59	-4.82
12	Pyrazine dimer (Cs)	-4.42	-0.34	-0.26	-4.87	-4.45	-6.36
13	Uracil dimer (C2)	-10.12	-6.05	-4.26	-11.25	-7.59	-15.47
14	Indole benzene (C1)	-5.22	-1.33	-1.65	-8.16	-6.26	-5.89
15	Adenine thymine stack (C1)	-12.23	-5.15	-6.50	-15.13	-11.70	-17.13
Mixed complexes							
16	Ethene ethyne (C2v)	-1.53	-0.57	-1.23	-2.47	-2.58	-1.36
17	Benzene H2O (Cs)	-3.28	-1.03	-1.63	-3.90	-4.46	-2.94
18	Benzene NH3 (Cs)	-2.35	-0.80	-0.93	-4.04	-3.99	-1.99
19	Benzene HCN (Cs)	-4.46	-0.92	-1.85	-4.28	-4.40	-2.88
20	Benzene dimer (C2v)	-2.74	-0.09	-0.52	-4.22	-4.39	-2.87
21	Indole benzene T-shape (C1)	-5.73	-1.24	-1.67	-7.74	-7.20	-4.87
22	Phenol dimer (C1)	-7.05	-3.39	-4.33	-11.55	-8.95	-5.63
	RMSE (Hydrogen bonded)		9.90	8.04	3.38	1.82	2.66
	RMSE (Dispersion bonded)		3.73	3.65	2.36	1.71	2.79
	RMSE (Mixed bonded)		2.96	2.40	2.09	1.40	0.89
	RMSE		6.25	5.22	2.65	1.65	2.31
	MUE		4.82	4.09	2.16	1.51	1.60

Table B-7: Geometry optimized interaction distances (Angstroms) for the S22 complexes. The interaction distance is defined as the CM-distance in complexes 11-15 and 2nd value reported for complex 22.

No.	Molecule (symmetry)	Ref. Values	AM1	PM3	AM1-D	PM3-D
Hydrogen bonded complexes						
1	(NH3)2 (C2h)	2.504	2.784	3.241	2.646	2.726
2	(H2O)2 (Cs)	1.952	2.094	1.809	1.911	1.769
3	Formic acid dimer (C2h)	1.670	2.101	1.776	1.925	1.737
4	Formamide dimer (C2h)	1.841	2.072	1.807	1.981	1.763
5	Uracil dimer (C2h)	1.775	2.044	1.787	1.946	1.744
6	2-Pyridoxine2-aminopyridine (C1)	1.859, 1.874	2.511, 2.107	1.798, 1.815	1.980, 1.981	1.722,1.768
7	Adenine thymine WC (C1)	1.819, 1.929	2.476, 2.101	1.780, 1.821	1.807, 2.018	1.708,1.769
Complexes with predominant dispersion contribution						
8	(CH4)2 (D3d)	3.718	3.721	3.447	2.881	3.160
9	(C2H4)2 (D2d)	3.718	3.714	3.706	3.305	3.469
10	Benzene CH4 (C3)	3.716	3.746	3.718	3.315	3.450
11	Benzene dimer (C2h)	3.765	6.952	6.096	3.643	3.499
12	Pyrazine dimer (Cs)	3.479	4.848	4.760	3.695	3.437
13	Uracil dimer (C2)	3.166	5.805	6.732	3.097	3.406
14	Indole benzene (C1)	3.498	5.572	5.520	4.448	3.415
15	Adenine thymine stack (C1)	3.172	6.202	5.788	4.320	3.280
Mixed complexes						
16	Ethene ethyne (C2v)	2.752	2.468	2.429	2.319	2.366
17	Benzene H2O (Cs)	2.531	4.020	3.746	2.986	2.982
18	Benzene NH3 (Cs)	3.592	4.092	4.025	2.995	3.069
19	Benzene HCN (Cs)	3.387	3.472	3.694	3.228	3.343
20	Benzene dimer (C2v)	3.513	5.225	3.606	3.253	3.370
21	Indole benzene T-shape (C1)	3.210	3.811	3.807	3.010	3.233
22	Phenol dimer (C1)	1.937, 4.921	2.174, 5.925	1.829, 5.712	2.001, 5.040	1.778, 5.265
	RMSE (Hydrogen bonded)		0.387	0.257	0.137	0.134
	RMSE (Dispersion bonded)		2.015	1.962	0.644	0.272
	RMSE (Mixed bonded)		0.929	0.598	0.336	0.315
	RMSE		1.277	1.171	0.419	0.249
	MUE		0.853	0.691	0.301	0.199

Appendix C:Determining the Gradient of the FS1 Correction

C.1 Dispersion Gradient

Determining the gradient of the dispersion correction term (Eq. 3.1) is relatively simple, since the only variable is r_{ij} (the distance between two atoms in the system). Taking the derivative of Eq. 3.1 yields:

$$\frac{dE_{dis}}{dr_{ij}} = \frac{6C_6^{ij}}{r_{ij}^7}\frac{1}{1+e^{-d\left(\frac{r_{ij}}{S_R R_{vdw}}-1\right)}} - \frac{C_6^{ij}}{r_{ij}^6}\frac{d\cdot e^{-d\left(\frac{r_{ij}}{S_R R_{vdw}}-1\right)}}{\left(1+e^{-d\left(\frac{r_{ij}}{S_R R_{vdw}}-1\right)}\right)^2 S_R R_{vdw}} \quad . \qquad \text{(C. 1)}$$

This derivative can be use to determine the gradient, with respect to the x, y, and z coordinates by applying the following transformations:

$$G_x = \frac{dE_{DISP}}{dr_{ij}}\left(x_i - x_j\right)r_{ij}\,, \qquad \text{(C.2)}$$

$$G_y = \frac{dE_{DISP}}{dr_{ij}}\left(y_i - y_j\right)r_{ij}\,, \qquad \text{(C.3)}$$

$$G_z = \frac{dE_{DISP}}{dr_{ij}}\left(z_i - z_j\right)r_{ij}\,. \qquad \text{(C.4)}$$

These values can now be simply added to the AM1 gradient of the *i*th atom and the negative of these values to the *j*th atom. The gradient needs to be computed for every unique atom pair. This allows geometry optimizations to be performed considering the dispersion correction.

C.2 Hydrogen-Bonding Gradient

Determining the Gradient of the hydrogen-bonding correction term (Eq. 3.4) is more difficult than for the dispersion term. In this case the there are two variables r_{ik} (the distance between the H---Y atoms in the system) and θ (the XH---Y angle). This is more difficult than just taking the partial derivative with respect to the two variables, since these variables are dependent. The gradient was achieved by writing the HB correction terms of a function of the distance between the x, y, and z components of the X-H and H-Y atom pairs:

$$X_{ij} = \left(x_i - x_j\right) \ , \tag{C.5}$$

$$Y_{ij} = \left(y_i - y_j\right) \ , \tag{C.6}$$

$$Z_{ij} = \left(z_i - z_j\right) \ , \tag{C.7}$$

$$X_{ik} = \left(x_i - x_k\right) \ , \tag{C.8}$$

$$Y_{ik} = \left(y_i - y_k\right) \ , \tag{C.9}$$

$$Z_{ik} = \left(z_i - z_k\right) \ , \tag{C.10}$$

where i is the H atom, j is the X atom, and k is the Y atom. The following substitutions can be made:

$$r_{ij} = \sqrt{X_{ij}^2 + Y_{ij}^2 + Z_{ij}^2} \ , \tag{C.11}$$

$$r_{ik} = \sqrt{X_{ik}^2 + Y_{ik}^2 + Z_{ik}^2} \ , \tag{C.12}$$

$$\theta = \arccos\left(\frac{X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}}{r_{ij}r_{ik}}\right) \ . \tag{C.13}$$

Now, the partial derivatives with respect to these 6 variables can be taken, since the variables are now independent. This yields the following 6 expressions:

$$\frac{\partial E_{HB}}{\partial X_{ij}} = 2\beta\left(X_{ik} - \frac{\left(X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}\right)}{r_{ij}^2} X_{ij} \right) , \tag{C.14}$$

$$\frac{\partial E_{HB}}{\partial Y_{ij}} = 2\beta\left(Y_{ik} - \frac{\left(X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}\right)}{r_{ij}^2} Y_{ij} \right) , \tag{C.15}$$

$$\frac{\partial E_{HB}}{\partial Z_{ij}} = 2\beta\left(Z_{ik} - \frac{\left(X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}\right)}{r_{ij}^2} Z_{ij} \right) , \tag{C.16}$$

$$\frac{\partial E_{HB}}{\partial X_{ik}} = 2\beta X_{ij} - 3\beta \frac{X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}}{r_{ik}^2} X_{ik} + \beta\left(X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}\right) \left(\frac{-2\left(r_{ik} - \alpha_2 R_{vdw}\right)}{\alpha_3^2\left(1 + \alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)\right)^2 r_{ik}} X_{ik} + \frac{2\alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)^2}{\alpha_3^2\left(1 + \alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)\right)^3 r_{ik}} X_{ik} \right) , \tag{C.17}$$

$$\frac{\partial E_{HB}}{\partial Y_{ik}} = 2\beta Y_{ij} - 3\beta \frac{X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}}{r_{ik}^2} Y_{ik} + \beta\left(X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}\right) \left(\frac{-2\left(r_{ik} - \alpha_2 R_{vdw}\right)}{\alpha_3^2\left(1 + \alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)\right)^2 r_{ik}} Y_{ik} + \frac{2\alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)^2}{\alpha_3^2\left(1 + \alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)\right)^3 r_{ik}} Y_{ik} \right) , \tag{C.18}$$

$$\frac{\partial E_{HB}}{\partial Z_{ik}} = 2\beta Z_{ij} - 3\beta \frac{X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}}{r_{ik}^2} Z_{ik} + \beta\left(X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}\right) \left(\frac{-2\left(r_{ik} - \alpha_2 R_{vdw}\right)}{\alpha_3^2\left(1 + \alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)\right)^2 r_{ik}} Z_{ik} + \frac{2\alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)^2}{\alpha_3^2\left(1 + \alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)\right)^3 r_{ik}} Z_{ik} \right) , \tag{C.19}$$

where:

$$\beta = \alpha_1 Q_i Q_j \frac{\left(X_{ij}X_{ik} + Y_{ij}Y_{ik} + Z_{ij}Z_{ik}\right)}{r_{ij}^2 r_{ik}^3} e^{\frac{-\left(r_{ik} - \alpha_2 R_{vdw}\right)^2}{\alpha_3^2\left(1 + \alpha_4\left(r_{ik} - \alpha_2 R_{vdw}\right)\right)^2}} . \tag{C.20}$$

Since, these partial derivatives are with respect to the difference of the x, y, and z coordinates, the values can be directly added to the AM1 gradient of the *i*th, *j*th, and *k*th atoms as follows; for the *i*th atom:

$$G_x = \frac{\partial E_{HB}}{\partial X_{ij}} + \frac{\partial E_{HB}}{\partial X_{ik}} \quad , \tag{C.21}$$

$$G_y = \frac{\partial E_{HB}}{\partial Y_{ij}} + \frac{\partial E_{HB}}{\partial Y_{ik}} \quad , \tag{C.22}$$

$$G_z = \frac{\partial E_{HB}}{\partial Z_{ij}} + \frac{\partial E_{HB}}{\partial Z_{ik}} \quad , \tag{C.23}$$

for the *j*th atom:

$$G_x = -\frac{\partial E_{HB}}{\partial X_{ij}} \quad , \tag{C.24}$$

$$G_y = -\frac{\partial E_{HB}}{\partial Y_{ij}} \quad , \tag{C.25}$$

$$G_z = -\frac{\partial E_{HB}}{\partial Z_{ij}} \quad , \tag{C.26}$$

for the *k*th atom:

$$G_x = -\frac{\partial E_{HB}}{\partial X_{ik}} \quad , \tag{C.27}$$

$$G_y = -\frac{\partial E_{HB}}{\partial Y_{ik}} \quad , \tag{C.28}$$

$$G_z = -\frac{\partial E_{HB}}{\partial Z_{ik}} \quad . \tag{C.29}$$

The gradient needs to be computed for every HB situation identified. This allows geometry optimizations to be performed considering the HB correction. The gradient considered here is the nuclear contribution; however, the charges on the atoms are also changing. This effect was not considered in the 1st version of AM1-FS1, but the method was later adjusted, as described is Appendix D.

Appendix D: SCF Addition of the HB Correction Term

In the first version of the empirically corrected AM1 method, termed AM1-FS1, the hydrogen-bonding (HB) correction term was added to the post self-consistent-field (SCF) energy. This causes the electronic gradient of HB correction to be neglected. This led to a slight error in the energy upon optimization. The error arises since the HB correction term depends on the Coulson charges for each atom involved, which depend on the electron density matrix (*P*). The Coulson charge of the *i*th atom is determined as follows:

$$Q_i = Z_i - R_i \quad ; \quad R_i = \sum_{\mu \in i} P_{\mu\mu} = trace(P) \ . \tag{D.1}$$

Solving the AM1 Hamiltonian for the electronic energy is an iterative process (SCF procedure); therefore, the density matrix changes throughout the process. Since, the FS1 HB correction term depends on the electron density matrix, the term needs to be added iteratively. To achieve this goal, let's first write the HB correction term involving R_A:

$$E_{HB} = \alpha_1 \frac{(Z_i - R_i)(Z_j - R_j)}{r_{ij}} \cos^2(\theta) f_{damp2}(r_{ij}) \ , \tag{D.2}$$

upon expanding $(Z_i - R_i)(Z_j - R_j)$, the expression can be broken up into the nuclear (E_{HB0}), 1-electron (E_{HB1}), and 2-electron terms (E_{HB2}):

$$E_{HB0} = \alpha_1 \frac{Z_i Z_j}{r_{ij}} \cos^2(\theta) f_{damp2}(r_{ij}) \ , \tag{D.3}$$

$$E_{HB1} = \alpha_1 \frac{(-Z_i R_j - Z_j R_i)}{r_{ij}} \cos^2(\theta) f_{damp2}(r_{ij}) \ , \tag{D.4}$$

$$E_{HB2} = \alpha_1 \frac{R_i R_j}{r_{ij}} \cos^2(\theta) f_{damp2}(r_{ij}) \quad . \tag{D.5}$$

Therefore, the total HB correction energy is:

$$E_{HB} = E_{HB0} + E_{HB1} + E_{HB2} \quad . \tag{D.6}$$

The E_{HB0} term is a constant because it does not depend on the electron density; therefore, it can simply be added at the end of the SCF procedure. The other two terms do need to be considered in the SCF procedure; however, this process is more complicated than simply adding E_{HB1} and E_{HB2} to the total energy at each iteration. The derivative of E_{HB1}, with respect to the electron density, needs to be added to the core-Hamiltonian matrix which is the one-electron part of the Fock operator. In addition, the derivative of E_{HB2}, with respect to the electron density, needs to be added to the Fock matrix; the two-electron part of the Fock operator. The derivatives of Eq. A.4 and A.5 with respect to R_i and R_j are:

$$\frac{\partial E_{HB1}}{\partial P_{\mu\nu}} = \alpha_1 \frac{(-Z_i - Z_j)}{r_{ij}} \cos^2(\theta) f_{damp2}(r_{ij}) \quad , \tag{D.7}$$

$$\frac{\partial E_{HB2}}{\partial P_{\mu\nu}} = \alpha_1 \frac{(R_i + R_j)}{r_{ij}} \cos^2(\theta) f_{damp2}(r_{ij}) \quad . \tag{D.8}$$

These derivatives were obtained by using the following relation for the trace of a matrix,

$$\frac{\partial R_i}{\partial P_{\mu\nu}} = \frac{\partial traceP}{\partial P_{\mu\nu}} = 1 \quad . \tag{D.9}$$

These derivatives with respect to R_i and R_j are added to the diagonal matrix elements of the core-Hamiltonian and Fock matrices centered on the *i*th and *j*th atoms respectively.

This process incorporates the total energy of the HB correction term, as well as, the derivative with respect to the atom charges.

Making these changes to the AM1-FS1 method required re-optimization of the HB parameters, leading to the following values: α_1 =0.3400, α_2 =0.6238, α_3 =0.4165, and α_4 =1.2409. Although, these values differ slightly from the original AM1-FS1, there is no appreciable effect on the tested interaction energies. The RMSE for the binding energies (see Chapter 3) of the F66 training set converged to the same value of 0.99 kcal/mol. The consequence of not considering the charge derivative causes the true minimum of the potential energy surface to differ from what is otherwise predicted.

Appendix E: Supporting Information for Chapter 3

Table E-1: F66 training set. All values are interaction energy (kcal/mol) at the reported geometry. The AM1-D results are from applying McNamara and Hillier's method. Complexes 1-18 Ref. 45 ; Complexes 19-22 Ref. 57; Complexes 23-27 Ref. 61; Complexes 28-32 Ref. 58; Complexes 33-37 Ref. 51; Complexes 38-42 Ref. 59; Complexes 43-47 Ref. 51; Complexes 48-57 Ref. 111; Complexes 58-66 Ref. 60.

	Ref. Values	AM1-D	AM1-FS1
1 adenine_thymine_wc (C1)	-16.37	-16.58	-16.29
2 adenine_thymine_stack (C1)	-12.23	-12.20	-9.87
3 ammonia_dimer (C2h)	-3.17	-3.43	-1.60
4 methane_dimer (D3d)	-0.53	-0.94	-0.61
5 ethene_dimer (D2d)	-1.51	-3.31	-2.27
6 ethene_ethine (C2v)	-1.53	-1.61	-1.36
7 formic_acid_dimer (C2h)	-18.61	-14.56	-16.06
8 formamide_dimer (C2h)	-15.96	-16.77	-15.75
9 benzene_water (Cs)	-3.28	-3.43	-2.78
10 benzene_ammonia (Cs)	-2.35	-3.00	-2.65
11 benzene_methane (C3)	-1.50	-2.12	-1.79
12 benzene_dimer_(C2h)	-2.73	-2.90	-2.23
13 indole_benzene_T-shape (C1)	-5.73	-7.10	-4.63
14 indole_benzene (C1)	-5.22	-4.04	-3.23
15 pyrazine_dimer (Cs)	-4.42	-4.57	-3.81
16 2-pyridoxine_2-aminopyridine (C1)	-16.71	-16.01	-14.73
17 phenol_dimer (C1)	-7.05	-9.07	-6.91
18 benzene_HCN (Cs)	-4.46	-4.44	-3.17
19 methanol_dimer	-5.70	-6.83	-4.82
20 methanol-formaldehyde	-5.31	-6.59	-4.24
21 methyl_amide_dimer_alpha	-6.69	-7.95	-7.39
22 methyl_amide_dimer_beta	-7.65	-8.62	-7.52
23 H2O_R=2.5	-2.68	-3.77	-1.66
24 H2O_R=2.85	-5.59	-6.59	-5.14
25 H2O_R=2.95	-5.45	-6.31	-5.16
26 H2O_R=3.0	-5.32	-6.06	-5.29
27 H2O_R=3.5	-3.43	-3.23	-3.05
28 nitromethane_R=2.0	-4.53	-9.08	-4.01
29 nitromethane_R=2.25	-5.63	-7.98	-5.75
30 nitromethane_R=2.375	-5.50	-7.02	-5.38
31 nitromethane_R=2.5	-5.19	-6.12	-4.86
32 nitromethane_R=2.75	-4.40	-4.58	-4.13
33 benzene_dimer_PAR_R=3.2	3.71	-1.04	0.77
34 benzene_dimer_PAR_R=3.5	-0.62	-2.88	-2.09
35 benzene_dimer_PAR_R=3.9	-1.70	-2.56	-1.97
36 benzene_dimer_PAR_R=4.5	-1.08	-1.50	-1.21
37 benzene_dimer_PAR_R=6.5	-0.04	-0.18	-0.14
38 benzene_dimer_M1_R=3.25	4.35	2.16	2.86
39 benzene_dimer_M1_R=3.6	-1.64	-2.11	-1.26
40 benzene_dimer_M1_R=3.96163	-2.74	-2.83	-2.17
41 benzene_dimer_M1_R=4.4	-2.18	-2.24	-1.80
42 benzene_dimer_M1_R=5.0	-1.17	-1.34	-1.12
43 benzene_dimer_T_R=4.4	1.10	-2.65	0.16
44 benzene_dimer_T_R=4.7	-2.03	-4.34	-2.07
45 benzene_dimer_T_R=5.0	-2.61	-3.51	-2.98
46 benzene_dimer_T_R=5.5	-1.98	-1.78	-1.50
47 benzene_dimer_T_R=7.9	-0.22	-0.20	-0.18
48 uracil_dimer_H-bonded_R=1.47	-12.19	-10.87	-12.29
49 uracil_dimer_H-bonded_R=1.67	-18.68	-17.89	-21.34
50 uracil_dimer_H-bonded_R=1.77	-19.22	-19.57	-20.80
51 uracil_dimer_H-bonded_R=2.77	-9.27	-7.68	-8.54
52 uracil_dimer_H-bonded_R=3.77	-3.82	-0.65	-3.76
53 uracil_dimer_Stacked_R=3.06	-6.91	-8.29	-6.13
54 uracil_dimer_Stacked_R=3.26	-9.38	-9.43	-9.43
55 uracil_dimer_Stacked_R=3.36	-9.58	-8.98	-8.47
56 uracil_dimer_Stacked_R=4.36	-4.60	-3.65	-3.54
57 uracil_dimer_Stacked_R=8.36	-0.30	-0.30	-0.31
58 benz-acetylene_S1_R=3.0	1.48	0.68	1.34
59 benz-acetylene_S1_R=3.75	-0.82	-1.07	-0.80
60 benz-acetylene_S1_R=4.5	-0.32	-0.56	-0.45
61 benz-acetylene_T_R=3.5	1.69	-0.07	2.98
62 benz-acetylene_T_R=4.25	-2.59	-2.71	-2.30
63 benz-acetylene_T_R=5.00	-1.52	-1.00	-0.88
64 benz-acetylene_SS1_R=4.5	-0.41	-2.13	-2.04
65 benz-acetylene_SS1_R=5.0	-1.32	-1.30	-1.05
66 benz-acetylene_SS1_R=6.0	-0.60	-0.43	-0.38
RMSE		1.492	0.988
MUE		1.017	0.685
MAXE-MINE		8.800	5.489

Table E-2: The reference CCSD(T), single point and optimized AM1-FS1 binding energies (kcal/mol) for 16 hydrogen bonded DNA base pair complexes. All reference values can be found in Ref. 45.

	Ref. Geometry	CCSD(T)/CBS noCP	AM1-FS1//Ref.	AM1-FS1
G…C WC	MP2	-32.06	-29.8	-32.7
mG…mC WC	MP2	-31.59	-30.4	-32.1
A…T WC	MP2	-16.86	-16.3	-18.5
mA…mT H	MP2	-18.16	-13.8	-12.8
8oG…C WC pl	MP2	-33.30	-33.4	-36.2
I…C WC pl	MP2	-24.90	-24.8	-25.5
G…U wobble	MP2	-19.10	-18.3	-17.6
CCH+	MP2	-51.40	-55.7	-54.2
U…U Calcutta pl	MP2	-10.30	-9.3	-10.1
U…U pl	MP2	-13.70	-13.5	-14.4
A…T WC	Exp.	-16.40	-17.0	-17.7
G…C WC*	Exp.	-35.80	-34.8	-32.4
A…T WC	Exp.	-18.40	-18.6	-17.5
G…A HB	Exp.	-11.30	-10.1	-9.9
C…G WC	Exp.	-30.70	-32.1	-32.8
G…C WC	Exp.	-31.40	-32.2	-33.2
RMSE			1.78	2.18
MUE			1.25	1.75

Appendix F: FS1 FORTRAN Code

This FORTRAN code is designed to be implemented with GAMESS and requires modifications of several GAMESS source files in order to pass information to and from the different subroutines. The below FORTRAN code contains the following subroutines:

FS1SETUP – Contains required parameters and setup information

EDISFS1 – Calculates the dispersion energy

EHBFS1 – Calculates the hydrogen-bonding energy (EHB0, EHB1, and EHB2) and electronic HB gradient.

GRADCORDIS – Calculates the dispersion gradient

GRADCORHB – Calculates the nuclear HB gradient

```
C*MODULE AM1FS1  *DECK FS1SETUP
      SUBROUTINE FS1SETUP(MODE)
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
C
      LOGICAL RHO
      PARAMETER (MXATM=2000, TOANGS=0.52917724924D0)
C GAMESS COMMON BLOCKS
      COMMON /INFOA / NAT,ICH,MUL,NUM,NQMT,NE,NA,NB,
     *              ZAN(MXATM),C(3,MXATM),IAN(MXATM)
      COMMON /IOFILE/ IR,IW,IP,IS,IPK,IDAF,NAV,IODA(950)
C FS1 PARAMETERS
      COMMON /FS1PAR/ SRVDW,D6,DTOL,ALFAP(4),C6(100),VDWD(100)
C
C     MODE=0: JUST PRINT
C     MODE=1: ABOUT TO DO DISPERION ENERGY TERM
C     MODE=2: ABOUT TO DO H-BOND ENERGY TERM
C     MODE=3: ABOUT TO DO DISPERION GRADIENT TERM
C     MODE=4: ABOUT TO DO H-BOND GRADIENT TERM
C
C
      C6(1)= 0.14D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(5)= 3.13D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
```

```
      C6(6)= 1.75D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(7)= 1.23D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(8)= 0.70D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(9)= 0.75D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(14)= 9.23D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(15)= 7.84D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(16)= 5.57D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
      C6(17)= 5.07D0*(1.0D0/(TOANGS/10.0D0))**6*3.809D-07
C
      VDWD(1)= 1.001D0/TOANGS
      VDWD(5)= 1.485D0/TOANGS
      VDWD(6)= 1.452D0/TOANGS
      VDWD(7)= 1.397D0/TOANGS
      VDWD(8)= 1.342D0/TOANGS
      VDWD(9)= 1.287D0/TOANGS
      VDWD(14)= 1.716D0/TOANGS
      VDWD(15)= 1.705D0/TOANGS
      VDWD(16)= 1.683D0/TOANGS
      VDWD(17)= 1.639D0/TOANGS
C
      SRVDW = 1.1058892D0
      D6 = 1000.0D0
      ALFAP(1)=0.3400377D0
      ALFAP(2)=0.6237877D0
      ALFAP(3)=0.4164925D0
      ALFAP(4)=1.2409020D0
C
C EXPONENTIAL CUTOFF IN DAMPING FUNCTION
      CTOL = 1.D-10
      DTOL = -LOG(CTOL/100.D0)
C
      IF(MODE.EQ.0) THEN
        WRITE(IW,*)'FS1 CORRECTION WILL BE APLIED TO THE AM1 METHOD'
        WRITE(IW,*)'SR (VDW SCALE FACTOR)=',SRVDW
        WRITE(IW,*)'D (DAMPING)=        ',D6
        WRITE(IW,*)'ALPHA1=             ',ALFAP(1)
        WRITE(IW,*)'ALPHA2=             ',ALFAP(2)
        WRITE(IW,*)'ALPHA3=             ',ALFAP(3)
        WRITE(IW,*)'ALPHA4=             ',ALFAP(4)
      END IF
C
      IF(MODE.GT.0) THEN
         IF(MODE.EQ.1) WRITE(IW,9000)
C          IF(MODE.EQ.2) WRITE(IW,9001)
         IF(MODE.EQ.3) WRITE(IW,9002)
         IF(MODE.EQ.4) WRITE(IW,9003)
```

```
      END IF
C
      RETURN
C
 9000 FORMAT(/1X,'SCF HAS CONVERGED, FS1 DISPERSION CORRECTION',
     *          ' ADDED TO AM1-HB ENERGY')
C 9001 FORMAT(/1X,'FS1 H-BOND CORRECTION TO ENERGY CONSIDERED')
 9002 FORMAT( 1X,'FS1 DISPERSION CORRECTION ADDED TO GRADIENT')
 9003 FORMAT( 1X,'FS1 H-BOND CORRECTION TO GRADIENT CONSIDERED')
C
      END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCCC
C*MODULE AM1FS1  *DECK EDISFS1
      SUBROUTINE EDISFS1(E_DISP)
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      PARAMETER (MXATM=2000)
      COMMON /INFOA / NAT,ICH,MUL,NUM,NQMT,NE,NA,NB,
     *              ZAN(MXATM),C(3,MXATM),IAN(MXATM)
      COMMON /FS1PAR/ SRVDW,D6,DTOL,ALFAP(4),C6(100),VDWD(100)
C
C      PROGRAM CALCULATES DISPERSION ENERGY
C      BASED ON GRIMME"S 2006 DISPERSION CORRECTION
C      NOTE THAT THIS IS DONE AFTER THE SCF PROCESS HAS CONVERGED
C
      CALL FS1SETUP(1)
C
      E_SUM = 0.0D0
      DO II=1,NAT-1
        C6I=C6(IAN(II))
        DO JJ=II+1,NAT
          C6J= C6(IAN(JJ))
C
          DX= C(1,II)-C(1,JJ)
          DY= C(2,II)-C(2,JJ)
          DZ= C(3,II)-C(3,JJ)
          RIJ= SQRT(DX*DX+DY*DY+DZ*DZ)
C
          RVDW= (VDWD(IAN(II))+VDWD(IAN(JJ)))
          C66= SQRT(C6I*C6J)
          SRRVDW= RVDW*SRVDW
          EXPARG=D6*(RIJ/SRRVDW-1.0D0)
          IF( EXPARG.GT.DTOL ) THEN
            DAMP = 1.D0
          ELSE IF( EXPARG.LT.-DTOL ) THEN
            DAMP = 0.D0
```

```
          ELSE
            DAMP = 1.D0/(1.D0+EXP(-EXPARG))
          ENDIF
          E_SUM = E_SUM-DAMP*C66/RIJ**6
        ENDDO
      ENDDO
      E_DISP=E_SUM
      RETURN
      END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCC
C*MODULE AM1FS1  *DECK EHBFS1
      SUBROUTINE EHBFS1(DMATRX,EHB0,EHB1,EHB2,MODE)
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      PARAMETER (MXATM=2000,PIVAL=3.141592653589793D0)
      DIMENSION DMATRX(*)
      DIMENSION HMHB(MXATM),FMHB(MXATM)
      COMMON /INFOA / NAT,ICH,MUL,NUM,NQMT,NE,NA,NB,
     *            ZAN(MXATM),C(3,MXATM),IAN(MXATM)
      COMMON /IOFILE/ IR,IW,IP,IS,IPK,IDAF,NAV,IODA(950)
C
      COMMON /MOLKST/
NUMAT,MNAT(MXATM),NFIRST(MXATM),NMIDLE(MXATM),
     *            NLAST(MXATM),NORBS,NELECS,NALPHA,NBETA,NCLOSE,
     *            NOPEN,NDUMY,FRACT
      COMMON /FS1PAR/ SRVDW,D6,DTOL,ALFAP(4),C6(100),VDWD(100)
      COMMON /HBSCF/ DPHB(MXATM)
      COMMON /CORE  / CORE(107)
C
C       CALCULATE H-BONDING CORRECTION TO THE ENERGY
C       NOTE THAT THIS IS DONE EVERY SCF STEP!
C
      CALL FS1SETUP(2)
C
      DO II=1,NAT
        IF (MODE.EQ.1) HMHB(II)=0.D0
        IF (MODE.EQ.2) FMHB(II)=0.D0
      ENDDO
C
      DO II=1,NAT
C FINDS A HYDROGEN ATOM
        NI=IAN(II)
        IF(NI.EQ.1) THEN
          RIJ=1.0D+99
          DO JTEMP=1,NAT
C FINDS NEAREST ATOM TO THE HYDROGEN ATOM
```

```
          DX=C(1,II)-C(1,JTEMP)
          DY=C(2,II)-C(2,JTEMP)
          DZ=C(3,II)-C(3,JTEMP)
          RIJTEMP=SQRT(DX*DX+DY*DY+DZ*DZ)
C FINDS NEAREST ATOM TO THE HYDROGEN ATOM
           IF((RIJTEMP.LT.RIJ).AND.(II.NE.JTEMP)) THEN
             RIJ=RIJTEMP
             JJ=JTEMP
           END IF
          ENDDO
C
       NJ=IAN(JJ)
       IF((NJ.EQ.7).OR.(NJ.EQ.8).OR.(NJ.EQ.9)) THEN
C
C       SEARCH FOR HYDROGEN BOND PARTNER
C
         DO KK=1,NAT
           NK=IAN(KK)
          IF( ((NK.EQ.7).OR.(NK.EQ.8).OR.(NK.EQ.9)) .AND.
     *          (JJ.NE.KK) ) THEN
C
           DX1=C(1,JJ)-C(1,II)
           DY1=C(2,JJ)-C(2,II)
           DZ1=C(3,JJ)-C(3,II)
C
           DX2=C(1,KK)-C(1,II)
           DY2=C(2,KK)-C(2,II)
           DZ2=C(3,KK)-C(3,II)
C
           D1D2=DX1*DX2 + DY1*DY2 + DZ1*DZ2
           RIJ2=DX1*DX1 + DY1*DY1 + DZ1*DZ1
           RIK2=DX2*DX2 + DY2*DY2 + DZ2*DZ2
           RIJ=SQRT(RIJ2)
           RIK=SQRT(RIK2)
           HBANG=ACOS(D1D2/(RIJ*RIK))
C
           IF((HBANG.LE.-PIVAL/2.0D0).OR.
     *          (HBANG.GE.PIVAL/2.0D0)) THEN
C
            WF=(D1D2/(RIJ*RIK))**2
C
            RVDW = ( (2.0D0*VDWD(IAN(II)))**3 +
     *                (2.0D0*VDWD(IAN(KK)))**3 ) /
     *              ( (2.0D0*VDWD(IAN(II)))**2 +
     *                (2.0D0*VDWD(IAN(KK)))**2 )
C
```

```
            DAMP2= EXP(-1.0D0*(RIK-RVDW*ALFAP(2))**2/
     *                 (ALFAP(3)**2*(1.0D0+ALFAP(4)*
     *                 (RIK-RVDW*ALFAP(2)))**2))
C
C PRE-SCF STEPS
            IF (MODE.EQ.1) THEN
              EHB0= EHB0+CORE(IAN(II))*CORE(IAN(KK))*
     *              ALFAP(1)*WF*DAMP2/RIK
C
              DHBE1I= -CORE(IAN(KK))*ALFAP(1)*WF*DAMP2/RIK
              DHBE1K= -CORE(IAN(II))*ALFAP(1)*WF*DAMP2/RIK
              HMHB(II)=HMHB(II)+DHBE1I
              HMHB(KK)=HMHB(KK)+DHBE1K
            END IF
C EVERY-SCF STEP
            IF (MODE.EQ.2) THEN
              EHB1= EHB1-(CORE(IAN(II))*DPHB(KK)+
     *              CORE(IAN(KK))*DPHB(II))*
     *              ALFAP(1)*WF*DAMP2/RIK
              EHB2= EHB2+DPHB(II)*DPHB(KK)*
     *               ALFAP(1)*WF*DAMP2/RIK
C CALCULATE THE 2-E DERIVATIVE
              DHBE2I= DPHB(KK)*ALFAP(1)*WF*DAMP2/RIK
              DHBE2K= DPHB(II)*ALFAP(1)*WF*DAMP2/RIK
C
              FMHB(II)=FMHB(II)+DHBE2I
              FMHB(KK)=FMHB(KK)+DHBE2K
            END IF
C
          END IF
        END IF
      ENDDO
    END IF
    END IF
  ENDDO
C ADD CHARGE DERIVATIVE TO FOCK MATRIX
    MM=0
    DO II=1,NAT
      IA=NFIRST(II)
      IB=NLAST(II)
        DO LL=IA,IB
          MM=MM+LL
           IF (MODE.EQ.1) DMATRX(MM)=DMATRX(MM)+HMHB(II)
           IF (MODE.EQ.2) DMATRX(MM)=DMATRX(MM)+FMHB(II)
        ENDDO
    ENDDO
```

```
      RETURN
      END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCCCC
C*MODULE AM1FS1  *DECK GRADCORDIS
      SUBROUTINE GRADCORDIS(GEDIS)
C
C     CALCULATE DISPERSION CORRECTION CONTRIBUTION TO NUCLEAR
GRAD
C
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      PARAMETER (MXATM=2000)
      DIMENSION GEDIS(3,MXATM),G(3,MXATM)
      COMMON /INFOA / NAT,ICH,MUL,NUM,NQMT,NE,NA,NB,
     *                ZAN(MXATM),C(3,MXATM),IAN(MXATM)
      COMMON /IOFILE/ IR,IW,IP,IS,IPK,IDAF,NAV,IODA(950)
      COMMON /FS1PAR/ SRVDW,D6,DTOL,ALFAP(4),C6(100),VDWD(100)
C
      CALL FS1SETUP(3)
C
      DO JJ=1,NAT
        DO II=1,3
          G(II,JJ) = 0.0D0
          GEDIS(II,JJ) = 0.0D0
        ENDDO
      ENDDO
C
      DO II=1,NAT-1
         C6I=C6(IAN(II))
         DO JJ=II+1,NAT
           C6J=C6(IAN(JJ))
C
            DX=C(1,II)-C(1,JJ)
            DY=C(2,II)-C(2,JJ)
            DZ=C(3,II)-C(3,JJ)
            RIJ=SQRT(DX*DX+DY*DY+DZ*DZ)
C
            RVDW =(VDWD(IAN(II))+VDWD(IAN(JJ)))
            C66  =SQRT(C6I*C6J)
            SRRVDW= RVDW*SRVDW
C
            EXPARG=D6*(RIJ/SRRVDW-1.0D0)
            IF( EXPARG.GT.DTOL ) THEN
              DAMP = 1.D0
              DAMP1 = 0.D0
            ELSE IF( EXPARG.LT.-DTOL ) THEN
```

```
          DAMP = 0.D0
          DAMP1 = 0.D0
        ELSE
          DAMP = 1.D0/(1.D0+EXP(-EXPARG))
          DAMP1 = EXP(-EXPARG)
        ENDIF
C
        DEDF= 6.0D0*DAMP/(RIJ**7)-D6/(SRRVDW*RIJ**6)*
     *          DAMP1*DAMP**2
C
        G(1,II)=G(1,II) + C66*DEDF*DX/RIJ
        G(2,II)=G(2,II) + C66*DEDF*DY/RIJ
        G(3,II)=G(3,II) + C66*DEDF*DZ/RIJ
        G(1,JJ)=G(1,JJ) - C66*DEDF*DX/RIJ
        G(2,JJ)=G(2,JJ) - C66*DEDF*DY/RIJ
        G(3,JJ)=G(3,JJ) - C66*DEDF*DZ/RIJ
      ENDDO
    ENDDO
C
    DO JJ=1,NAT
      DO II=1,3
        IF(ABS(G(II,JJ)).LT.1.0D-14) G(II,JJ)=0.0D+00
      ENDDO
    ENDDO
C
    DO JJ=1,NAT
      DO II=1,3
       GEDIS(II,JJ) = G(II,JJ)
      ENDDO
    ENDDO
C
    RETURN
    END
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCCC
C*MODULE AM1FS1  *DECK GRADCORHB
    SUBROUTINE GRADCORHB(GEHB)
C
C    CALCULATE H-BOND CORRECTION CONTRIBUTION TO GRADIENT
C    CHARGE CONTRIBUTION IS NEGLECTED HERE, THIS IS DONE IN EHBFS1
C
    IMPLICIT DOUBLE PRECISION(A-H,O-Z)
    PARAMETER (MXATM=2000,PIVAL=3.141592653589793D0)
    DIMENSION G2(3,MXATM),GEHB(3,MXATM)
    COMMON /INFOA / NAT,ICH,MUL,NUM,NQMT,NE,NA,NB,
    *             ZAN(MXATM),C(3,MXATM),IAN(MXATM)
```

```
      COMMON /IOFILE/ IR,IW,IP,IS,IPK,IDAF,NAV,IODA(950)
      COMMON /FS1PAR/ SRVDW,D6,DTOL,ALFAP(4),C6(100),VDWD(100)
      COMMON /HBSCF/ DPHB(MXATM)
      COMMON /CORE  / CORE(107)
C
      CALL FS1SETUP(4)
C
      DO JJ=1,NAT
       DO II=1,3
        G2(II,JJ) = 0.0D0
        GEHB(II,JJ) = 0.0D0
       ENDDO
      ENDDO
C
      DO II=1,NAT
C FINDS A HYDROGEN ATOM
        NI=IAN(II)
        IF(NI.EQ.1) THEN
          RIJ=1.0D+99
          DO JTEMP=1,NAT
            DX=C(1,II)-C(1,JTEMP)
            DY=C(2,II)-C(2,JTEMP)
            DZ=C(3,II)-C(3,JTEMP)
            RIJTEMP=SQRT(DX*DX+DY*DY+DZ*DZ)
C FINDS NEAREST ATOM TO THE HYDROGEN ATOM
            IF((RIJTEMP.LT.RIJ).AND.(II.NE.JTEMP)) THEN
              RIJ=RIJTEMP
              JJ=JTEMP
            END IF
          ENDDO
C
        NJ=IAN(JJ)
        IF((NJ.EQ.7).OR.(NJ.EQ.8).OR.(NJ.EQ.9)) THEN
C
C       SEARCH FOR HYDROGEN BOND PARTNER
C
        DO KK=1,NAT
          NK=IAN(KK)
          IF( ((NK.EQ.7).OR.(NK.EQ.8).OR.(NK.EQ.9)) .AND.
     *         (JJ.NE.KK) ) THEN
C
            DX1=C(1,JJ)-C(1,II)
            DY1=C(2,JJ)-C(2,II)
            DZ1=C(3,JJ)-C(3,II)
C
            DX2=C(1,KK)-C(1,II)
```

```
      DY2=C(2,KK)-C(2,II)
      DZ2=C(3,KK)-C(3,II)
C
      D1D2=DX1*DX2 + DY1*DY2 + DZ1*DZ2
      RIJ2=DX1*DX1 + DY1*DY1 + DZ1*DZ1
      RIK2=DX2*DX2 + DY2*DY2 + DZ2*DZ2
      RIJ=SQRT(RIJ2)
      RIK=SQRT(RIK2)
      HBANG=ACOS(D1D2/(RIJ*RIK))
C
      IF((HBANG.LE.-PIVAL/2.0D0).OR.
     *   (HBANG.GE.PIVAL/2.0D0)) THEN
      RVDW = ( (2.0D0*VDWD(NI))**3 +
     *         (2.0D0*VDWD(NK))**3 ) /
     *       ( (2.0D0*VDWD(NI))**2 +
     *         (2.0D0*VDWD(NK))**2 )
C
      TEMP1= D1D2/RIJ2
      TEMP2= 1.0D0/(RIK*RIK2)
      TEMP3= RIK - ALFAP(2)*RVDW
      TEMP4= 1.0D0/(ALFAP(3)**2)
      TEMP5= TEMP3**2*TEMP4
      TEMP6= 1.0D0 + ALFAP(4)*TEMP3
      TEMP7= TEMP6**2
      TEMP8= EXP(-TEMP5/TEMP7)
      TEMP9= D1D2**2/RIJ2**2
      TEMP10= TEMP2*TEMP8*DX1
      TEMP11= TEMP2*TEMP8*DY1
      TEMP12= TEMP2*TEMP8*DZ1
      TEMP13= D1D2**2/RIJ2
      TEMP14= TEMP8/(RIK*RIK2**2)
      TEMP15= TEMP3*TEMP4
      TEMP16= 1.0D0/(TEMP7*RIK)
      TEMP17= TEMP5/(TEMP7*TEMP6)
      TEMP18= ALFAP(4)/RIK
C
      QI=CORE(IAN(II))-DPHB(II)
      QK=CORE(IAN(KK))-DPHB(KK)
      Q2IK= QI*QK
C
      DDX1= Q2IK*(2.0D0*TEMP1*TEMP2*TEMP8*DX2
     *     - 2.0D0*TEMP9*TEMP10)
      DDY1= Q2IK*(2.0D0*TEMP1*TEMP2*TEMP8*DY2
     *     - 2.0D0*TEMP9*TEMP11)
      DDZ1= Q2IK*(2.0D0*TEMP1*TEMP2*TEMP8*DZ2
     *     - 2.0D0*TEMP9*TEMP12)
```

```
            DDX2= Q2IK*(2.0D0*TEMP1*TEMP10 - 3.0D0*TEMP13*TEMP14
     *           *DX2 + TEMP13*TEMP2*(-2.0D0*TEMP15*TEMP16
     *           *DX2 + 2.0D0*TEMP17*TEMP18*DX2)*TEMP8)
            DDY2= Q2IK*(2.0D0*TEMP1*TEMP11 - 3.0D0*TEMP13*TEMP14
     *           *DY2 + TEMP13*TEMP2*(-2.0D0*TEMP15*TEMP16
     *           *DY2 + 2.0D0*TEMP17*TEMP18*DY2)*TEMP8)
            DDZ2= Q2IK*(2.0D0*TEMP1*TEMP12 - 3.0D0*TEMP13*TEMP14
     *           *DZ2 + TEMP13*TEMP2*(-2.0D0*TEMP15*TEMP16
     *           *DZ2 + 2.0D0*TEMP17*TEMP18*DZ2)*TEMP8)
C
            G2(1,II)= G2(1,II)- ALFAP(1)*(DDX1 + DDX2)
            G2(2,II)= G2(2,II)- ALFAP(1)*(DDY1 + DDY2)
            G2(3,II)= G2(3,II)- ALFAP(1)*(DDZ1 + DDZ2)
C
            G2(1,JJ)= G2(1,JJ)+ ALFAP(1)*DDX1
            G2(2,JJ)= G2(2,JJ)+ ALFAP(1)*DDY1
            G2(3,JJ)= G2(3,JJ)+ ALFAP(1)*DDZ1
C
            G2(1,KK)= G2(1,KK)+ ALFAP(1)*DDX2
            G2(2,KK)= G2(2,KK)+ ALFAP(1)*DDY2
            G2(3,KK)= G2(3,KK)+ ALFAP(1)*DDZ2
C
            END IF
          END IF
        ENDDO
        END IF
      END IF
     ENDDO
C
    DO JJ=1,NAT
      DO II=1,3
        IF(ABS(G2(II,JJ)).LT.1.0D-14) G2(II,JJ)=0.0D+00
      ENDDO
    ENDDO
C
    DO JJ=1,NAT
      DO II=1,3
        GEHB(II,JJ)=G2(II,JJ)
      ENDDO
    ENDDO
    RETURN
    END
```

Vita

Michael E. Foster

EDUCATION:

Ph.D. in Theoretical and Computational Chemistry, Department of Chemistry Drexel University, Philadelphia, PA. Dissertation: *The Development of an Empirically Corrected Semi-Empirical Method and its Application to Macromolecular Complexes.*

Bachelor of Science, Chemistry (Minor in Mathematics), Drexel University, Philadelphia, PA June 2007.

PUBLICATIONS:

Michael E. Foster and Karl Sohlberg. "Designing Fullerene Separation Materials: A Theoretical Study", *Fullerenes, Nanotubes and Carbon Nanostructures*, 2010 (accepted)

Michael E. Foster and Karl Sohlberg. "A New Empirical Correction to the AM1 Method for Macromolecular Complexes", *J. Chem. Theory Comput*., 2010. 6(7): p. 2153-2166.

Michael E. Foster and Karl Sohlberg. "Empirically corrected DFT and semi-empirical methods for non-bonding interactions", *Phys. Chem. Chem. Phys.*, 2010. 12 (2): p. 307-322.

Dragancea, D., Addison A.W., Zeller M., Foster M.E., Prushan M.J., Thompson L.K., Revenco M.D., Hunter A.D. "A tetranuclear copper(II) complex with bis(o-aminobenzaldehyde)thiocarbohydrazone", *Inorganica Chimica Acta*., 2010. 363(10): p. 2065-2070.

Michael E. Foster and Karl Sohlberg, "Theoretical study of binding site preference in [2]rotaxanes", *J. Chem. Theory Comput.*, 2008. 3(6): p. 2221-2233.

Michael E. Foster and Karl Sohlberg, "Theoretical and Computational Studies of Mechanical Molecular Devices", *Encyclopedia of Nanoscience and Nanotechnology*, (accepted).

SCHOLARSHIPS and AWARDS:

September, 2008 Awarded the Graduate Assistance in Areas of National Need (GAANN) Fellowship
September, 2007 Awarded the Graduate Assistance in Areas of National Need (GAANN) Fellowship
June, 2007 Graduated – Magna Cum Laude at Drexel University
June, 2007 Robert O. Hutchins BIOMOL Prize for Research
April, 2007 2nd Place at Drexel CoAS Research Day
June, 2006 American Chemist Award
February, 2004 Dean's Scholar Award from Drexel University
May, 2004 Academic Excellence in Organic Chemistry Award from Bucks County Community College

CONFERENCE PRESENTATIONS:

April, 2010; ACS Mid-Atlantic Regional Meeting
April, 2010; Drexel Research Day
April, 2010; Drexel CoAS Research Day
April, 2009; Drexel Research Day
April, 2009; Drexel CoAS Research Day
August, 2008; ACS 236th National Meeting & Exposition
April, 2008; Drexel Research Day
April, 2008; Drexel CoAS Research Day
January, 2008; The Philadelphia Section, ACS Meeting
August, 2007; ACS 234th National Meeting & Exposition
May, 2007; ACS Mid-Atlantic Regional Meeting
April, 2007; 15th Sigma Xi Student Research Symposium

TEACHING EXPERIENCE:

Drexel University, Department of Chemistry
General Chemistry 101, 102, 103 (Recitations and Labs)
Physical Chemistry 557, 558 (Labs)