Energy Decomposition Analysis of Drug/Protein Interactions

Lorena Ruano de Domingo

Máster en Química Teórica y Modelización Computacional

MÁSTERES DE LA UAM

2021-2022

Facultad de Ciencias

UAM Universidad Autónoma de Madrid

excelencia Campus Internacional UAM CSIC+

UNIVERSIDAD AUTONOMA DE MADRID

Master Erasmus Mundus in

Theoretical Chemistry and Computational Modelling

Energy Decomposition Analysis of Drug/Protein Interactions

Lorena Ruano de Domingo

Master Thesis. Course 2021-2022

Director: Juan José Nogueira Pérez
Codirector: Marcos Mandado Alonso
Places where the project was carried out:
Autonomous University of Madrid / Department of Chemistry
University of Vigo / Department of Physical Chemistry

FACULTAD DE CIENCIAS

ACKNOWLEDGMENT

I would like to thank my director Juan José Nogueira Pérez for trusting, teaching and helping me since my Bachelor Thesis, and my codirector Marcos Mandado Alonso for teaching me and for his help during my stay in Vigo. Of course, thanks both for your patience.

Thanks all the members of MoBioChem group for their support and specially Gustavo, who has been helping me since my Bachelor Thesis too.

Thanks my friends and classmates of the EMTCCM for accompanying me in this stage, specially Javi, Nuria, Sergio, etc. and my non-theoretical chemist friends that although they do not understand anything, they listen to my troubles in programming and trust me more than me, Ana.

My family, who has worked hard to give me the opportunity to get here. Thanks for all your support and for cheering me up.

Finally, thanks CCC-UAM and Vigo's cluster for the computational time.

ABSTRACT

The presence of proteins in our organisms is essential for their correct functioning due to the large number of processes in which they are involved, in particular, as target to drugs. Therefore, the study of drug/protein interactions is very important to understand the mode of action of those drugs and develop improved ones. In this study, a Python code has been developed to automatically compute the interaction energy between 20 amino acids and 300 drugs and to perform an energy decomposition analysis based on deformation electron densities. The results of these calculations reveal that, among the attractive energy contributions of the systems, namely electrostatic, dispersion and induction, the electrostatic energy contribution dominates for most of the drug/amino acid pairs. Moreover, the distributions of these attractive energy terms are relatively broad, which might indicate that there exist a structure/EDA relation. Thus, an analysis of the EDA based on the presence of eight of the most frequent polar functional groups in bioactive molecules and some non-polar groups was carried out, leading to the conclusion that the classification of the systems based only in the presence of one functional group is not possible. In the future, a more complex analysis, taking into account different groups simultaneously, will be performed to find out the relation between chemical structure and EDA.

RESUMEN

La presencia de proteínas en nuestros organismos es esencial para su buen funcionamiento debido al gran número de procesos en los que están involucradas, en particular, actuando como moléculas diana de medicamentos. Por lo tanto, el estudio de las interacciones medicamento/proteína es muy importante para comprender el modo de acción de estos medicamentos y desarrollar otros con propiedades mejoradas. En este estudio, se ha desarrollado un código de Python que calcula de manera automática la energía de interacción entre 20 aminoácidos y 300 medicamentos, y realiza un análisis de la descomposición de la energía basado en la densidad de deformación electrónica. Los resultados de estos cálculos revelan que, de entre las diferentes contribuciones de la energía atractiva de los sistemas, electrostática, dispersión e inducción, la contribución de energía electrostática es la que domina para la mayoría de los complejos medicamento/aminoácido. Además, las distribuciones de estos términos de energía atractivos son anchas, indicando que podría existir alguna relación entre la estructura del medicamento y los resultados del análisis EDA. Por lo que se realiza un análisis de la presencia de ocho de los grupos funcionales polares más frecuentes en moléculas bioactivas y algunos grupos no polares, lo que lleva a la conclusión de que no es posible una clasificación de los sistemas basada únicamente en la presencia de un único grupo funcional. En el futuro, se realizará un análisis más complejo teniendo en cuenta varios grupos al mismo tiempo para encontrar la relación entre la estructura del fármaco y los resultados obtenidos del EDA.

CONTENT

1 INTRODUCTION

Proteins are crucial components of every organism because almost all processes that cells carry out need their presence. There are thousands of different proteins in all cellular systems and they are the major constituents of our organisms.[1-3] The importance of this group of molecules was pointed out by the German chemist Gerardus Mulder in 1838, who employed for the first time the word "protein" from the greek "*proteios*", that means "fundamental" or "essential".[4] Proteins are macromolecules mainly formed by small building blocks called amino acids joined together by peptide bonds, although many proteins need other components called cofactors or prosthetic groups to function correctly. There are hundreds of amino acids in nature, but just 20 amino acids make up the proteins. They are composed by a central carbon (α carbon) surrounded by an amino (NH_2) and a carboxyl (COOH) groups, a hydrogen atom and a side chain (R-chain) which depends on the amino acid. The first and last amino acids of the polypeptide chain have their amino and carboxyl groups free and are usually called amino- or N-terminus and carboxy- or C-terminus, respectively.[5] (see Figure 1)

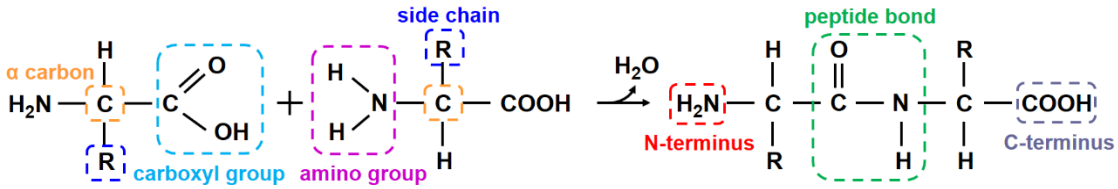


Figure 1. Schematic representation of the formation of peptide bond and labels of the principal moieties of an amino acid.

Every protein has a specific structure that is determined by the sequence of amino acids and, therefore, by the sequence of nucleotides in the DNA, whose proper folding in a unique three-dimensional structure allows the protein to be biologically active and functional.[4] Due to the huge number of combinations of the monomeric units (amino acids), proteins execute myriad functions.[1, 6] Likely, the most important functions are performed by enzymes, which act as catalysts speeding up many reactions, for example, the lactase enzyme, which catalyses the degradation of lactose into simpler sugar molecules. But there are many other types of proteins: (i) transport proteins embedded in cell membranes, *e.g.*, ion channels, which allow the flow of substances through the lipid

bilayer; (ii) transport proteins present in the blood stream, *e.g.*, the haemoglobin protein, in charge of the oxygen transport through the organism; (iii) defensive proteins acting as immune protection, *e.g.*, antibodies; (iv) signalling proteins such as hormones, *e.g.*, insulin, or other cell surface receptors, which transmit signals into or between cells or transmit nervous impulses; (v) structural proteins which are involved in the rigidity of the cell and in the contractile function in the muscle cells, *e.g.*, collagen, actin and myosin; and (vi) regulatory proteins which control the activity of other proteins. These are some of the most relevant protein classes, but there are many more in charge of a range of relevant biological functions.[2, 4-7]

As explained above, proteins are implied in many biological processes in our organisms by interacting with other molecules, including the interaction with drugs.[8] Pharmacology aims to investigate the metabolic route of the drugs once they have entered our bodies. Specifically, the drug/protein interactions are intensively investigated in many research studies since the drug/protein binding process determines, in a great extent, the distribution, toxicity and activity of the drugs and, therefore, their therapeutic efficiency.[9] One example is the binding of drugs with blood proteins as albumin, that controls osmotic pressure, among other functions.[10] Hence, the study of drug/protein interactions is crucial to understand the mode of action of those drugs once they are in our organism, and to develop novel therapeutic agents with enhanced efficacy.[8] The application of experimental techniques often remains the most trustworthy approach, but the experimental characterization of a huge number of drug/protein pairs is very time-consuming and costly due to sample volume and instrumentation.[11, 12] Thus, computational methods have gained popularity since they can be applied in a systematic and efficient way and can provide molecular information which is not attainable by experimental measurements.[11]

The complexity of biological systems, such as proteins, requires the use of approximations in the theoretical models to handle a large number of atoms and run longer simulation times. Most of the theoretical approximations simplify the calculation of the interatomic interactions, which is the most time-consuming step in computational modelling. This is the case of force fields, which are simple analytical functions that are parameterized based on quantum mechanical calculations or experimental measurements. Force fields are widely employed in classical molecular dynamics simulations of

biological systems.[13-15] The aim of a force field is to capture the nature of the interatomic interactions by describing the dependency of the potential energy on the coordinates of the system in a simplistic way and, thus, at a low computational cost.[16] There are many different force fields but a common expression employed is the one of Equation (2), which split the potential energy in bonding and non-bonding interactions:

$$V = \sum_{i=1}^{N_b} V_b + \sum_{i=1}^{N_a} V_a + \sum_{i=1}^{N_d} V_d + \sum_{i>j}^{N_{nb}} V_{nb} \tag{1}$$

The first three terms of this equation compose the bonding potential energy, where bond distances and angles are usually defined by harmonic potentials, while dihedral angles are defined by a Fourier transform. Non-bonding interactions have three components: Coulomb, van der Waals and repulsion interactions. The last two interactions are modelled by Lennard-Jones potentials, while the interactions between charges are defined by Coulomb potential. One of the most common potential shapes reads as:

$$V = \sum_{i=1}^{N_b} \frac{1}{2} k_b (r - r_0)^2 + \sum_{i=1}^{N_a} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{i=1}^{N_d} k_t [1 + cos(n\omega - \gamma)] + \sum_{i>j}^{N_{nb}} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} + 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2}$$

This equation has four different terms. In the first one, which corresponds to the potential energy of the bond distances (bond stretching), k_b is the force constant, r is the distance between the two atoms with the interaction of interest, and r_0 is the equilibrium distance between these two atoms. For the second term, which corresponds to the potential energy of the angles (angle bending), k_a is the force constant, θ is the angle between the atoms for which the interaction is calculated, and θ_0 is the equilibrium angle between those atoms. For the third term, which defines the potential energy of the dihedral angles or torsions, ω is the dihedral angle, n is the number of minima that are presented in the potential energy curve, k_t is the energy barrier that has to be overcome to go from one minimum to the other, and γ is the angle which determines the position of the minimum. The fourth term represents the non-bonding potential energy, in which the first fraction corresponds to electrostatic or coulombic interaction, where r_{ij} is the distance between

atom i and atom j, q_i and q_j are the charges of atoms i and atom j, and ε is the permittivity of the medium. The second fraction corresponds to the repulsion interaction between electrons, and the third fraction corresponds to the van der Waals attraction interactions, which are dipole-dipole interactions. In both cases, σ_{ij} is the distance between i and j atoms when the potential energy is zero and ε_{ij} is the maximum attraction energy between atoms.[16, 17]

The accuracy of the simulations depends on the quality of the force field parameters and how they are obtained, *i.e.*, the conditions used in the parametrization process. Moreover, many force fields are often developed for specific systems. Therefore, the accuracy and transferability of the force fields are limited.[16] The conventional force fields that describe the interactions with Equation (2) (or a similar one) are usually called first generation or class I force fields. This class of force fields have some disadvantages: chemical reactions cannot be modelled because bonds cannot be formed or broken, and atomic charges are fixed (non-reactive and non-polarizable force fields).[16, 18] This is the case, for example, of the AMBER[19] and CHARMM[20] force fields, among others.[21] There is a second generation or class II of force fields that include cross terms as coupling between stretching, bending and torsion terms.[16] For example, the CFF[22] (consistent force field) and UFF[23], among others, are class II force fields. More advance force fields also include the description of polarization interactions. Polarization is the redistribution of electron density between atoms chemically bonded due to an electric field created by another molecule.[21, 24] This redistribution of charge is usually modelled by three different methods: fluctuating charges, Drude oscillators and induced point dipoles. In the three methods, the charge distribution of the molecules depends on their chemical environment. The use of polarizable force fields to model large systems is computationally demanding and, therefore, they are barely applied.[21] In addition, this type of force fields are usually system specific.

These conventional force fields, including the polarizable ones, have the advantage of having physical meaning, although they present a clear limitation: the low flexibility of their analytical functions, which precludes an accurate description of the systems and processes which were not employed in their parameterization. Machine learning force fields (ML-FFs) overcome the limitations of the common force fields by using more flexible architectures, but at the price of lacking physical meaning. ML-FFs are very

efficient algorithms which are trained with high-level electronic structure methods and, therefore, combine the accuracy of *ab initio* methods and the low computational cost of conventional force fields.[25]

ML is based on the development and application of algorithms which improve their performance by learning from data, imitating, thus, the human learning process carried out by our brain. The algorithms receive training data and, then, create a model able to make predictions based on the behaviour of the test data. There are different models which can be trained, one of them being the artificial neural networks (NNs), which are composed by interconnected nodes or neurons that form the input layers with the input data, one or more hidden layers where the data is processed during the learning process, and the output layers, which provides the prediction made by the model.[26] Despite the convenient feature of being very flexible, force fields based on ML methods also present disadvantages: they need a large number of data to be trained and they do not provide physical insight into the nature of the interactions.[25] These drawbacks can be circumvent by introducing physically inspired data into the ML model. ML-FFs are usually trained with quantum mechanical energies and, something, with energy gradient. Here, it is suggested to train the ML-FF not only with interaction energies but also with their different contributions, namely electrostatic, repulsion, dispersion and induction interactions, which are obtained from energy decomposition analysis (EDA) calculations based on electron density.[27, 28] This EDA approach is used to provide a novel insight in the nature of the interactions between the amino acids and a set of drugs taken from ZINC database.[29] Finally, the employed drugs are classified taking into account their interaction energy constituents and structural features.

2 OBJECTIVES

The general goal of this thesis is to characterize by means of computational methods the interaction between drugs and amino acids. In order to achieve this general goal, the following specific goals are envisaged:

- Set up all possible complexes formed by 300 drugs from the ZINC database[29] and the 20 amino acids that compose our proteins.
- Find the most favourable relative orientations between monomers for each of the complexes.
- Compute the interaction energy curve for all the complexes.
- Perform an EDA analysis of the interaction energy.
- Find a relation between the EDA components and the chemical structure of the complexes.

All these steps will be achieved by developing a Python code able to perform each of the tasks in an automatic way. The long-term goal of this project is to develop a ML-FF based on the training of a NN with interaction energies and their EDA components to model drug/protein interactions.

3 METHODS

In this master thesis the interaction energies of a set of systems have been calculated using semiempirical and Density Functional Theory (DFT) methods and then, those interactions have been decomposed into their different contributions by means of EDA calculations. Therefore, these three methods are going to be explain in this section.

3.1 Interaction Energy and Basis Set Superposition Error (BSSE)

As has been said previously, the goal of this research is to calculate the interaction energy between the amino acids and a set of drugs. The interaction energy between two fragments A and B is defined as the difference between the energy of the AB complex and the energies of the isolated fragments.[30]

$$E_{int} = E_{AB}^{AB} - E_{A}^{A} - E_{B}^{B} \tag{3}$$

where the superscripts indicate the basis sets used to calculate the energies. If these basis sets were infinite, Equation (3) were correct but, this is never the case.

The energy of the isolated fragments is overestimated with respect to the energy of the complex because the calculation of the fragments uses a smaller basis set. To be correct and have comparable energies of complex and fragments, the basis sets have to be of the same size. Furthermore, the interaction of the basis between the fragments within the complex leads to a decrease of their energies, *i.e.*, the energies of the fragments are lower when they are forming the complex than when they are isolated. This is known as the Basis Set Superposition Error (BSSE).[31] To solve this problem, it is necessary to introduce the counterpoise (CP) technique, in which the energies of the complex and the fragments are calculated in the same basis set: the one of the complex using Equation (4). These energies will be computed by semiempirical and DFT methods.

$$E_{int}^{CP} = E_{AB}^{AB} - E_{A}^{AB} - E_{B}^{AB} \tag{4}$$

3.2 Semiempirical Methods

One of the principal issues in quantum chemistry is the resolution of the time-independent Schrödinger equation

$$H(r,R)\psi(r,R) = E(R)\psi(r,R) \tag{5}$$

where H is the Hamiltonian operator, represented in atomic units in Equation (6), ψ is the wavefunction which describes the state of the system, E is the energy of the system, and r are the coordinates of the electrons and R are the nuclear coordinates.

$$H(r,R) = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2}\nabla_A^2 + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} + \sum_{A=1}^{M}\sum_{i=1}^{N}\frac{Z_A}{r_{iA}} \quad (6)$$

where ∇_i^2 and ∇_A^2 are the Laplacian operators of the i electron and A nucleus, r_{ij} is the distance between i and j electrons, R_{AB} is the distance between A and B nuclei, r_{iA} is the distance between i electron and A nucleus, Z_A and Z_B are the atomic numbers of the nuclei, N is the number of electrons and M is the number of nuclei.

The Schrödinger equation cannot be solved analytically even for simple molecules. To simplify this problem, it is necessary to apply Born-Oppenheimer approximation, which considers that the positions of the nuclei are fixed while electrons move because nuclei are heavier than electrons. Then, Equation (6) can be simplified as follows

$$H_{el}(r,R) = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{i=1}^{N}\frac{Z_A}{r_{iA}} \quad (7)$$

Where the kinetic energy of the nuclei is considered to be zero and the interaction between nuclei is constant. One of the most important *ab initio* methods based on the Born-Oppenheimer approximation is the Hartree-Fock self-consistent field (SCF) method, which starts with a trial set of spin orbitals used to solve the Fock operator

$$f_1 = h_1 + \sum_u \left(J_u(1) - K_u(1)\right) \quad (8)$$

where h_1 is the core Hamiltonian of electron 1, $J_u(1)$ and $K_u(1)$ are the Coulomb and the exchange operators of electron 1, respectively, and $u = a, b, \ldots$ is the sum over all spin orbitals.

$$h_1 = -\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \quad (9)$$

$$J_u(1)\phi_a(1) = \left[\int \phi_u^*(2)\frac{1}{r_{12}}\phi_u(2)dx_2\right]\phi_a(1) \quad (10)$$

$$K_u(1)\phi_a(1) = \left[\int \phi_u^*(2)\frac{1}{r_{12}}\phi_a(2)dx_2\right]\phi_u(1) \quad (11)$$

Then, Hartree-Fock (HF) equations are solved to get a new set of spin orbitals and so on until convergence.

Semiempirical methods simplify the calculations of the electronic structure introducing approximations and parameters fitted to reproduce experimental measurements and from high level calculations. This limit its accuracy but instead, they can be applied to large systems like biomolecules.[32]

These methods are based on three approximations: (i) Neglection of the core electrons, Equation (9), from the Hamiltonian calculations; (ii) Use of minimum number of basis set; (iii) Reduction of the number of two-electron integrals, Coulomb and exchange, Equation (10) and (11), respectively.[33]

In this work, PM6 (parameterized model 6) has been applied because it is one of the most employed semiempirical method when describing interaction energies and provide accurate results, taking into account the intrinsic limitations of the semiempirical methods.[34]

3.3 Density Functional Theory

The Density Functional Theory is an alternative, based on electron density, to *ab initio* methods because it introduces electron correlation overcoming the poor treatment of this contribution in the HF method, but it is less time-consuming than post-HF methods. It is a compromise between accuracy and computational cost.[35] Formally, while wavefunction-based methods in a system of N electrons depend on $4N$ variables (3 spatial and 1 spin coordinates), the electron density depends only on the 3 spatial coordinates without taking into account the size of the system.[30, 33] However, in the formulation used to deal with chemical problems, the Kohn-Sham formulation that will be explained later, the electron density is computed from the molecular orbitals, which depends on 4N coordinates. Therefore, the real reason of the speed of DFT calculations for molecular

systems is not because it depends on a lower number of coordinates but because the exchange-correlation energy is simplified by parameterized functions, instead of computing the exchange integrals, as HF does. Thus, DFT is widely used to compute large molecules as biomolecules.

According to DFT, the energy of the electronic ground state is defined by a specific electron density. The problem is that the functional which relates the energy of the ground state and the electron density is unknown. The search of this functional is the goal of the DFT methods,[30] whose development started in 1964 with the Hohenberg and Kohn theorems.[36]

3.3.1 Hohenberg and Kohn Theorems

First Theorem

Any observable of a stationary non-degenerate ground state can be obtained, in principle exactly, from the ground state density, *i.e.*, every observable can be written as a functional of the electron density of the ground state. To prove this statement, it will be shown that the electron density determines the external potential via *reductio ad adsurdum*, supposing that this statement is not correct and leading to a contradiction.[33, 35]

Let's consider an exact density of a non-degenerate ground state and let's assume that, for the same density $\rho(r)$, there are two external potentials ($V_{ext,1}(r)$, $V_{ext,2}(r)$) that generate two Hamiltonians (H_1, H_2) with two different wave functions (ψ_1 , ψ_2) whose ground state energies are

$$\begin{aligned} E_1^0 &= \langle \psi_1 | H_1 | \psi_1 \rangle \\ E_2^0 &= \langle \psi_2 | H_2 | \psi_2 \rangle \end{aligned} \tag{12}$$

Now, the expectation value of the energy of ψ_2 is calculated with H_1 using the variational principle:

$$E_1^0 < \langle \psi_2 | H_1 | \psi_2 \rangle = \langle \psi_2 | H_2 | \psi_2 \rangle + \langle \psi_2 | H_1 - H_2 | \psi_2 \rangle \tag{13}$$

and we know that

$$\langle \psi_2 | H_1 - H_2 | \psi_2 \rangle = \int \rho(r) \big[V_{ext,1}(r) - V_{ext,2}(r) \big] dr \tag{14}$$

Therefore,

$$E_1^0 < E_2^0 + \int \rho(r)\big[V_{ext,1}(r) - V_{ext,2}(r)\big]dr \tag{15}$$

Similarly, the expectation value of the energy of ψ_1 with H_2 can be calculated as

$$E_2^0 < \langle\psi_1|H_2|\psi_1\rangle = \langle\psi_1|H_1|\psi_1\rangle + \langle\psi_1|H_2 - H_1|\psi_1\rangle \tag{16}$$

and knowing that

$$\langle\psi_1|H_2 - H_1|\psi_1\rangle = \int \rho(r)\big[V_{ext,2}(r) - V_{ext,1}(r)\big]dr \tag{17}$$

Hence,

$$E_2^0 < E_1^0 + \int \rho(r)\big[V_{ext,2}(r) - V_{ext,1}(r)\big]dr \tag{18}$$

From Equations (15) and (18), we obtain the following inequality

$$E_1^0 + E_2^0 < E_2^0 + E_1^0 \tag{19}$$

This leads to a contradiction proving that the initial assumption of having two different external potentials generated by the same electron density was not correct. Thus, each external potential is determined by a unique density.

The total ground state energy can be expressed as a functional of the electron density as

$$E(\rho) = V_{ne}(\rho) + T(\rho) + V_{ee}(\rho) = \int \rho(r)V_{ext}(r)dr + T(\rho) + V_{ee}(\rho) \tag{20}$$

then, the kinetic and electron-electron interaction energies can be grouped in a Hohenberg-Kohn functional, $F_{HF}(\rho)$ [33, 35]

$$E(\rho) = \int \rho(r)V_{ext}(r)dr + F_{HF}(\rho) \tag{21}$$

Second Theorem

The non-degenerate ground state electron density can be calculated, in principle exactly, by searching the electron density which minimizes the ground state energy using the

variational method.[35] Any trial density $\tilde{\rho}(r)$ defines a Hamiltonian $\tilde{H}$ from which we can obtain the wavefunction $\tilde{\psi}$ for the ground state, according to the variational method.[33]

$$\langle\tilde{\psi}|H|\tilde{\psi}\rangle = E(\tilde{\rho}) \geq E(\rho) = \langle\psi|H|\psi\rangle \tag{22}$$

This means that the total energy calculated from the trial density has to be larger or equal than the exact energy of the ground state.

If we apply the minimum condition to the energy, which is constrained by the N-representability, *i.e.*, $\int \rho(r)dr = N$ where N is the number of electrons, we obtain

$$\delta E(\rho) - \mu\delta\left[\int \rho(r)dr - N\right] = 0 \tag{23}$$

$$\mu = \frac{\delta E(\rho)}{\delta\rho(r)} = V_{ext}(r) + \frac{\delta F_{HF}(\rho)}{\delta\rho(r)} \tag{24}$$

where μ is the Lagrange multiplier at the minimum.

3.3.2 The Kohn and Sham Method

The previous Equations (12-*(24*) describe a method to minimize the energy by changing its density. The problem of Equation (24) is that the functional which relates the energy and the electron density is unknown, in particular, the relation between the kinetic energy and the electron density. Since the kinetic energy can be calculated from the wave function, Kohn and Sham proposed a method which consists of combining the electron density and wave function frameworks. It considers a reference system of N non-interacting electrons under a external potential V_r which provides a wave function with the same density as the real system.[33, 35]

The wave function of this ideal system can be exactly calculated by the Hartree-Fock method because there are no electron-electron interactions. The Hamiltonian of this system depends on single-electron terms: the kinetic energy of the electrons and the nuclear-electron interactions described by the external potential:

$$H_r = \sum_{i=1}^{N} h(i) = \sum_{i=1}^{N} -\frac{1}{2}\nabla^2(i) + \sum_{i=1}^{N} V_r(i) \tag{25}$$

and the exact wave function is expressed by a Slater determinant

$$\psi = \frac{1}{\sqrt{N}} |\phi_1(1)\phi_2(2) \dots \phi_N(N)| \tag{26}$$

where ϕ_i are the molecular orbitals expressed by a linear combination of basis functions θ_j, where c_{ji} are the coefficients of the linear combination:

$$\phi_i = \sum_{j=1}^{N} c_{ji}\, \theta_j \tag{27}$$

The orbital coefficients are obtained by solving the HF equations:

$$\left(-\frac{1}{2}\nabla^2 + V_r\right)\phi_i = E_i\phi_i \tag{28}$$

Once we know the wave function, the exact density and kinetic energy can be calculated by Equation (29) and (30), respectively:

$$\rho(r) = \sum_{i=1}^{N} |\phi_i|^2 \tag{29}$$

$$T_r(\rho) = \sum_{i=1}^{N} \left\langle \phi_i \middle| -\frac{1}{2}\nabla^2 \middle| \phi_i \right\rangle \tag{30}$$

Then, the total energy is written as

$$E_r(\rho) = T_r(\rho) + \int \rho(r) V_r(r) dr \tag{31}$$

The fundamental equation of DFT is the equation that minimizes the energy with respect to the electron density

$$\mu = \frac{\delta E_r(\rho)}{\delta \rho(r)} = V_r(r) + \frac{\delta T_r(\rho)}{\delta \rho(r)} \tag{32}$$

Now, for a real system with interacting electrons, Equation (31) is transformed in

$$E(\rho) = T(\rho) + V_{ee}(\rho) + \int \rho(r) V_{ext}(r) dr \tag{33}$$

If we add and subtract the kinetic energy $T_r(\rho)$ and the Coulomb repulsion $J(\rho)$ terms of the reference system of non-interacting electrons defined above, we obtain

$$E(\rho) = T_r(\rho) + [T(\rho) - T_r(\rho)] + J(\rho) + [V_{ee}(\rho) - J(\rho)] + \int \rho(r) V_{ext}(r) dr \tag{34}$$

where the Coulomb repulsion $J(\rho)$ term is defined as

$$J(\rho) = \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 \tag{35}$$

The difference between the kinetic energy of the real system $T(\rho)$ and that of the reference one $T_r(\rho)$ is the correlation kinetic energy $T_c(\rho)$, and the difference between electron-electron interaction $V_{ee}(\rho)$ and the classical Coulomb repulsion $J(\rho)$ is called exchange-correlation electronic energy $W_{XC}(\rho)$. These last two terms are usually grouped in a single term called exchange-correlation energy $E_{XC}(\rho)$

$$E_{XC}(\rho) = T_c(\rho) + W_{XC}(\rho) \tag{36}$$

and its derivative with respect to the density is the exchange-correlation potential

$$\frac{\partial E_{XC}(\rho)}{\partial \rho} = V_{XC}(r) \tag{37}$$

If we introduce Equation (35) and (36) into Equation (34)

$$E(\rho) = T_r(\rho) + E_{XC}(\rho) + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 + \int \rho(r) V_{ext}(r) dr \tag{38}$$

If we apply Equation (32)

$$\frac{\partial E(\rho)}{\partial \rho} = \frac{\partial T_r(\rho)}{\partial \rho} + \frac{\partial E_{XC}(\rho)}{\partial \rho} + V_{ext}(r) + \int \frac{\rho(r_2)}{r_{12}} dr_2 \tag{39}$$

where the Coulomb potential $V_c(r)$ is

$$V_c(r) = V_{ext}(r) + \int \frac{\rho(r_2)}{r_{12}} dr_2 \tag{40}$$

If Equation (37) and (40) are introduced in Equation (39)

$$\mu = \frac{\partial E(\rho)}{\partial \rho} = \frac{\partial T_r(\rho)}{\partial \rho} + V_c(r) + V_{XC}(r) \tag{41}$$

And the sum of the Coulomb $V_c(r)$ and exchange-correlation potentials $V_{XC}(r)$ is expressed as the effective potential $V_{eff}(r)$

$$\mu = \frac{\partial T_r(\rho)}{\partial \rho} + V_{eff}(r) \tag{42}$$

and compare with Equation (32), we realise that both equations are the same but changing $V_r(r)$ for $V_{eff}(r)$. Therefore, we can write the monoelectronic Schrödinger equation as

$$\left(-\frac{1}{2}\nabla^2 + V_c(r) + V_{XC}(r)\right)\phi_i = E_i\phi_i \tag{43}$$

where the orbitals ϕ_i are called Kohn-Sham orbitals, which allow the calculation of the electron density by Equation (29). Therefore, as in the HF method, the equations have to be solved iteratively. First, from a set of orbitals the electron density can be calculated using Equation (29). Then, the density is used to compute the coulomb and exchange-correlation potentials needed for solving the Kohn-Sham equations, Equation (43). This process is repeated until convergence. The problem is that the exchange-correlation energy and, therefore, the exchange-correlation potential are not known.

3.3.3 Exchange-Correlation Functionals

The difference in the results between different DFT methods is the choice of the exchange-correlational energy functional.[30] There is a classification in different types, the Jacob's ladder, depending on the dependency of the exchange correlation energy with the electron density: (i) in the local density approximation (LDA) the exchange correlation energy depends on the electron density $\rho(r)$ of a uniform electron gas; (ii) the generalized gradient approximation (GGA) considers a non-uniform electron gas and the exchange correlation energy depends on the electron density $\rho(r)$ and its gradient $\nabla\rho(r)$; (iii) the meta-GGA exchange-correlation functionals depend also on higher-order derivatives of the electron density, like the Laplacian $\nabla^2\rho(r)$; (iv) in the hybrid or hyper-GGA the exchange-correlation energy depends on $\rho(r)$, $\nabla\rho(r)$, $\nabla^2\rho(r)$ and part of the exact HF exchange energy is used to compute the exchange energy; (v) double hybrid functionals combines HF, DFT and Møller-Plesset perturbation theory (MP2).[30]

In this work, the M06-2X exchange-correlation energy functional has been used because it has been shown to be correct for small systems, like the ones studied here,[28] and it has the good feature of including implicitly most of the dispersion energy leaving the other energy terms untouched and, therefore, there is no need to use Grimme's empirical correction.

This functional is classified as hybrid meta-generalized gradient approximation (hybrid meta-GGAs) whose hybrid exchange correlation energy is define as

$$E_{XC}^{hyb} = \frac{X}{100} E_X^{HF} + \left(1 - \frac{X}{100}\right) E_X^{DFT} + E_C^{DFT} \tag{44}$$

where E_X^{HF} is the non-local HF exchange energy, E_X^{DFT} is the local DFT exchange energy, E_C^{DFT} is the local DFT correlation energy, and X is the percentage of HF exchange in the hybrid functional,[37] which corresponds to 54%.[38]

3.4 Energy Decomposition Analysis Based on the Electron Density

The energy of the AB complex, E_{AB}^{AB} in Equation (4), can be expressed in terms of the one- and two-electron densities, $\rho(r_1)$ and $\rho(r_1, r_2)$, respectively, and the exchange-correlation density, $\rho_{XC}(r_1, r_2)$,[27] as follows:

$$\begin{aligned} E_{AB} = & -\frac{1}{2}\int \nabla^2 \rho(r_1, r_1')_{r_1'=r_1}\, dr_1 + \int v_M\, \rho(r_1) dr_1 \\ & + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{|r_2 - r_1|} dr_1 dr_2 + \frac{1}{2}\iint \frac{\rho_{XC}(r_1, r_2)}{|r_2 - r_1|} dr_1 dr_2 \\ & + \sum_{I=1}^{M}\sum_{J>I}^{M} \frac{Z_I Z_J}{|R_I - R_J|} \end{aligned} \tag{45}$$

where ∇^2 is the Laplacian operator, $v_N = \frac{Z}{R}$ represents the electrostatic potential created by the nuclei, M is the number of nuclei and Z the nuclear charge. [27]

In Equation (45), the five terms correspond to the kinetic energy of electrons, electron-nucleus attraction, electron-electron classical repulsion, exchange-correlation and nuclear repulsion energies, respectively. Note that the electron kinetic energy term is calculated from the density matrix of first order, $\rho(r_1, r_1')$, instead of the one-electron density.

There are some terms that can be rewritten in terms of their contributions for the non-interacting systems. They are the nuclear electrostatic potential and energy, the one-electron density and the exchange-correlation density, Equation (46)-(49)

$$\sum_{I=1}^{M}\sum_{J>I}^{M}\frac{Z_I Z_J}{|R_I - R_J|} = \sum_{I=1}^{M_A}\sum_{J>I}^{M_A}\frac{Z_I Z_J}{|R_I - R_J|} + \sum_{I=1}^{M_B}\sum_{J>I}^{M_B}\frac{Z_I Z_J}{|R_I - R_J|} + \sum_{I=1}^{M_A}\sum_{J=1}^{M_B}\frac{Z_I Z_J}{|R_I - R_J|} \tag{46}$$

$$v_M = v_{M_A} + v_{M_B} \tag{47}$$

$$\rho(r_1) = \rho_A(r_1) + \rho_B(r_1) + \Delta\rho(r_1) \tag{48}$$

$$\rho_{XC}(r_1, r_2) = \rho_{XC,A}(r_1, r_2) + \rho_{XC,B}(r_1, r_2) + \rho_{XC,AB}(r_1, r_2) + \Delta\rho_{XC}(r_1, r_2) \tag{49}$$

where $\Delta\rho$ is the one-electron deformation density, that is formed by the contributions from Pauli repulsion, $\Delta\rho_{Pauli}$, and polarization energy, $\Delta\rho_{Pol}$, so Equation (*48*) is transformed in Equation (*50*)

$$\rho(r_1) = \rho_A(r_1) + \rho_B(r_1) + \Delta\rho_{Pauli}(r_1) + \Delta\rho_{Pol}(r_1) \tag{50}$$

and $\rho_{XC,AB}(r_1, r_2)$, $\Delta\rho_{XC}(r_1, r_2)$ are exchange and polarization terms.

Also, we can define the electrostatic potentials of fragments *A* and *B* by Equation (51) and (52)

$$v_A(r_1) = v_{M_A}(r_1) + \int \frac{\rho_A(r_2)}{|r_2 - r_1|} dr_2 \tag{51}$$

$$v_B(r_1) = v_{M_B}(r_1) + \int \frac{\rho_B(r_2)}{|r_2 - r_1|} dr_2 \tag{52}$$

By substituting Equations (46)-(52) into Equation (45) and removing the energies of the isolated (unperturbed) fragments, a new expression for the interaction energy, decomposed in electrostatic, E_{elec}, exchange, E_{exch}, repulsion, E_{rep}, and polarization terms, E_{pol}, is obtained, Equation (53)

$$E_{int} = E_{elec} + E_{exch} + E_{rep} + E_{pol} \tag{53}$$

where each of these terms are represented by Equations (54)-(57)

$$E_{elec} = \int \nu_{M_A}\, \rho_B(r_1) dr_1 + \int \nu_{M_B}\, \rho_A(r_1) dr_1 + \iint \frac{\rho_A(r_1)\rho_B(r_2)}{|r_2 - r_1|} dr_1 dr_2 + \sum_{I=1}^{M_A} \sum_{J=1}^{M_B} \frac{Z_I Z_J}{|R_I - R_J|} \tag{54}$$

$$E_{exch} = \frac{1}{2} \iint \frac{\rho_{XC,AB}(r_1, r_2)}{|r_2 - r_1|} dr_1 dr_2 \tag{55}$$

$$E_{rep} = \int \nu_A\, \Delta\rho_{Pauli}(r_1) dr_1 + \int \nu_B\, \Delta\rho_{Pauli}(r_1) dr_1 + \frac{1}{2} \iint \frac{\Delta\rho_{Pauli}(r_1)\Delta\rho_{Pauli}(r_2)}{|r_2 - r_1|} dr_1 dr_2 - \frac{1}{2} \int \nabla^2 \Delta\rho_{Pauli}(r_1, r_1')_{r_1' = r_1}\, dr_1 \tag{56}$$

$$E_{pol} = \int \nu_A\, \Delta\rho_{Pol}(r_1) dr_1 + \int \nu_B\, \Delta\rho_{Pol}(r_1) dr_1 + \frac{1}{2} \iint \frac{\Delta\rho_{Pol}(r_1)\Delta\rho_{Pol}(r_2)}{|r_2 - r_1|} dr_1 dr_2 + \iint \frac{\Delta\rho_{Pauli}(r_1)\Delta\rho_{Pol}(r_2)}{|r_2 - r_1|} dr_1 dr_2 - \frac{1}{2} \int \nabla^2 \Delta\rho_{Pol}(r_1, r_1')_{r_1' = r_1}\, dr_1 + \frac{1}{2} \iint \frac{\Delta\rho_{XC}(r_1, r_2)}{|r_2 - r_1|} dr_1 dr_2 \tag{57}$$

Exchange and repulsion terms arise from Pauli exclusion principle, so they are usually grouped in one term called Pauli energy, E_{Pauli}. On the other hand, the polarization term can be split exactly in two contributions, induction and dispersion, by means of 2nd-order perturbation theory (PT).[28, 39] The induction term is given by Equation (58)

$$E_{ind} = \sum_{m \neq 0} \frac{[\int \nu_A \rho_B^{m0}(r_1) dr_1]^2}{E_B^m - E_B^0} + \sum_{n \neq 0} \frac{[\int \nu_B \rho_A^{n0}(r_1) dr_1]^2}{E_A^n - E_A^0} \tag{58}$$

where ρ_A^{n0} and ρ_B^{m0} are the induced transition one-electron densities of fragments A and B from the ground state configuration 0 to the n and m excited states, respectively.

The 1st-order correction to the electron density required for the calculation of the 2nd-order energy is

$$\Delta\rho(r) = \Delta\rho_A(r) + \Delta\rho_B(r)$$

$$\Delta\rho(r) = 2\sum_{n\neq 0}\frac{\int v_B\rho_A^{n0}(r_1)dr_1}{E_A^n - E_A^0}\rho_A^{n0} + 2\sum_{m\neq 0}\frac{\int v_A\rho_B^{m0}(r_1)dr_1}{E_B^m - E_B^0}\rho_B^{m0} \tag{59}$$

The first and second terms of Equation (57) composed the charge-induction energy.

$$E_{charge-ind} = \int v_A\,\Delta\rho_{Pol}(r_1)dr_1 + \int v_B\,\Delta\rho_{Pol}(r_1)dr_1 \tag{60}$$

If we introduce Equation (59) into Equation (60),

$$\begin{aligned} E_{charge-ind} = {} & 2\sum_{m\neq 0}\frac{[\int v_A\rho_B^{m0}(r_1)dr_1]^2}{E_B^m - E_B^0} + 2\sum_{n\neq 0}\frac{[\int v_B\rho_A^{n0}(r_1)dr_1]^2}{E_A^n - E_A^0} \\ & + 2\sum_{m\neq 0}\frac{\int v_A\rho_B^{m0}(r_1)dr_1\int v_B\rho_B^{m0}(r_1)dr_1}{E_B^m - E_B^0} \\ & + 2\sum_{n\neq 0}\frac{\int v_B\rho_A^{n0}(r_1)dr_1\int v_A\rho_A^{n0}(r_1)dr_1}{E_A^n - E_A^0} \end{aligned} \tag{61}$$

Comparing Equation (58) with Equation (61) we realise that the induction energy is half of the first two terms

$$\begin{aligned} E_{ind} = \frac{1}{2}\Bigg[& E_{charge-ind} - \sum_{m\neq 0}\frac{\int v_A\rho_B^{m0}(r_1)dr_1\int v_B\rho_B^{m0}(r_1)dr_1}{E_B^m - E_B^0} \\ & - \sum_{n\neq 0}\frac{\int v_B\rho_A^{n0}(r_1)dr_1\int v_A\rho_A^{n0}(r_1)dr_1}{E_A^n - E_A^0}\Bigg] \end{aligned} \tag{62}$$

$$E_{ind} = \frac{1}{2}\left[E_{charge-ind} - \int v_A\,\Delta\rho_A(r_1)dr_1 - \int v_B\,\Delta\rho_B(r_1)dr_1\right]$$

This can be rewritten as

$$E_{ind} = \frac{1}{2}\left[\int v_A\,\Delta\rho_B(r_1)dr_1 + \int v_B\,\Delta\rho_A(r_1)dr_1\right] \tag{63}$$

which matches with the classical expression of the induction energy.

4 RESULTS AND DISCUSION

4.1 Computational Details

In this study, the setup for 6000 complexes formed by one of the amino acids and one of the 300 drugs from ZINC database[29] has been performed. The goal is to compute the interaction energy between the molecules of the complex and perform an EDA. To this aim, a Python script code, explained in detail in the next section, has been developed.

The script determines the orientation between the molecules of the complexes for which the interaction energy is the most favourable one. Then, for this orientation the potential energy curve is computed at PM6 level of theory. Five geometries around the PM6 minimum are selected and the interaction energy is recalculated at DFT level, using the M06-2X functional and the 6-31G* basis set as implemented in the Gaussian09[40] software. The DFT input files have been generated by MoBioTools,[41] a toolkit which was developed to automatically set up the input files of massive quantum mechanical calculations. Then, the DFT interaction energies are decomposed by an EDA in their different contributions.

4.2 Procedure and Code Development

The general workflow of this project, represented in Figure 2, consists of different steps. First, 6000 systems composed by one of the 20 amino acids and one of the 300 drugs randomly chosen from the ZINC database[29] are built. In the second step, for each of the systems, the individual molecules are located at a calculated distance and randomly rotated until having 15 different orientations. Then, a single point energy at PM6 level in all the orientations have been computed, and the lowest energy orientation is chosen. In the third step, the PM6 interaction energy between the two molecules of each complex is computed for different intermolecular distances. Then, 5 geometries around the PM6 minimum are chosen and, in the fourth step, the DFT interaction energy and EDA calculations are carried out. Each of these four steps will be explained in more detail in the following.

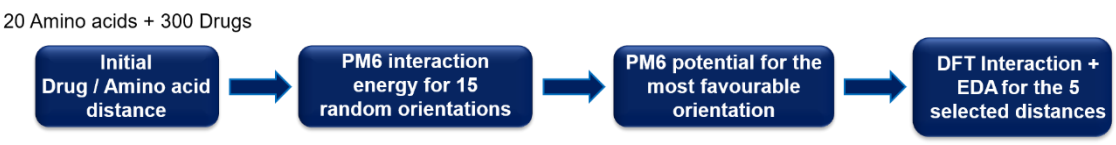


Figure 2. General workflow of the procedure performed.

For this study, the geometries of the amino acids have been taken from Ropo *et al.*[42] and the ones of the drugs from the ZINC database.[29] For the amino acids, there are neutral, acid and basic amino acids but, for simplicity, to avoid the use of large basis sets, the protonated or deprotonated species, respectively, are used to have all the amino acids in their neutral form. They are represented in Figure 3. Initially, 300 drugs were randomly selected. Thus, by combining the 20 amino acids with the 300 drugs, 6000 systems are obtained. However, at the time being, the calculation of 291 systems finished and are the ones presented in this thesis. In this section, a step-by-step explanation of the setup of the systems with a specific example will be given showing relevant fragments of the codes developed in Python.

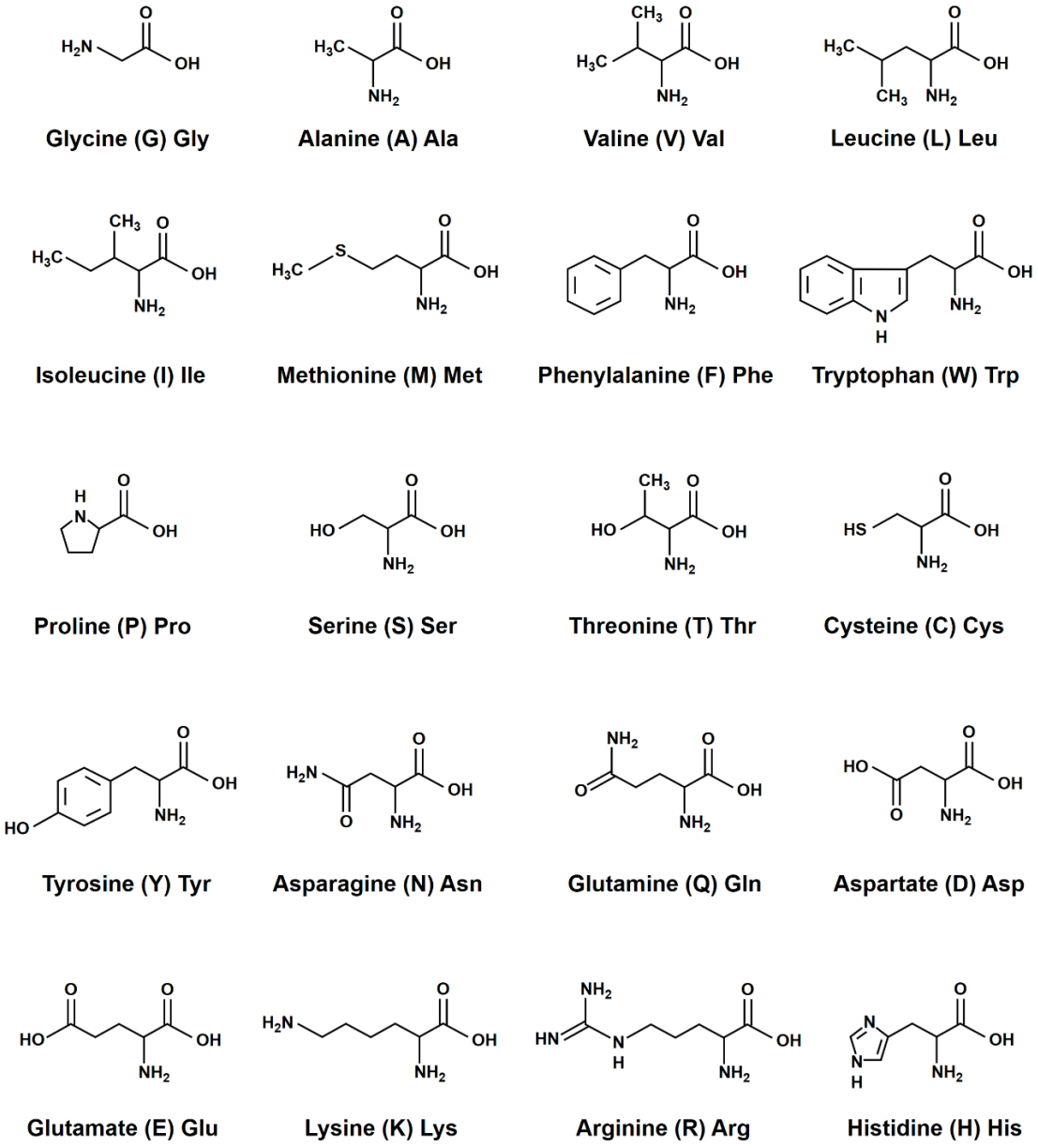


Figure 3. Molecular formulas of the amino acids.

The selected system is formed by the glycine amino acid and 3-(2-methoxyphenoxy)propane-1,2-diol molecule (Figure 4). From now on, they will also be called monomer1 and monomer2, respectively.

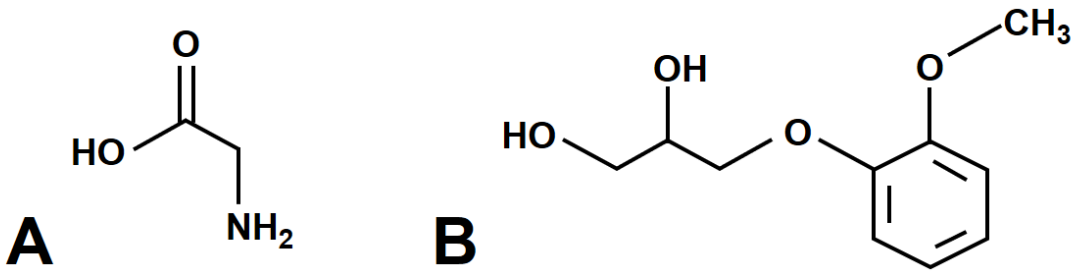


Figure 4. Chemical formula of glycine amino acid (A) and 3-(2-methoxyphenoxy)propane-1,2-diol molecule (B).

The goal is to obtain orientations at a distance where the interactions are likely to be attractive. The idea is to be close to the interaction energy minimum. To do so, the first step is to calculate the geometric centre of monomer1 and centre it in the origin of coordinates. In the code there is a list, called "atoms1", in which each element is a sub-list with the x-, y-, and z-coordinates of an atom. Figure 5 shows the fragment of the code that makes one list containing the x-coordinates of all atoms in the list "atoms1" and another two lists for y- and z- coordinates. Then, to centre monomer1, it is necessary to calculate its geometric centre. This is also represented in Figure 5, where "x_mean1" is the mean value of the x- coordinates of all the atoms in monomer1 and "round(…,6)" means that the result is approximated to a number with 6 decimal figures. The same for y- and z-coordinates.

```
x_coord1 = [item[0] for item in atoms1]
y_coord1 = [item[1] for item in atoms1]
z_coord1 = [item[2] for item in atoms1]
x_mean1 = round(sum(x_coord1)/len(x_coord1),6)
y_mean1 = round(sum(y_coord1)/len(y_coord1),6)
z_mean1 = round(sum(z_coord1)/len(z_coord1),6)
centroid1 = [x_mean1,y_mean1,z_mean1]
```

Figure 5. Fragment of the code to calculate the geometric centre of monomer1.

In the next step, the geometric centre of monomer1 is set to be the origin of coordinates. In Figure 6, it is observed that first, a new list for each coordinate is created, called "x_, y_ or z_coord1_centred" by subtracting to each coordinate the mean value of the corresponding coordinate, *i.e.* each value of coordinate x minus the mean value of

coordinate x, and then, a list with the new coordinates of each atom called "atoms1_centred" is created. These lists are created using a *for loop* from 0 to "n1" that is the number of atoms in "atoms1" list. Afterwards, this monomer is rotated with a rotational matrix that have random angles.

```
for i in range(0,n1):
    x_coord1_centred[i] = x_coord1[i] - x_mean1
    y_coord1_centred[i] = y_coord1[i] - y_mean1
    z_coord1_centred[i] = z_coord1[i] - z_mean1

for i in range(0,n1):
    atoms1_centred[i] = [round(x_coord1_centred[i],6),
                         round(y_coord1_centred[i],6),
                         round(z_coord1_centred[i],6)]
```

Figure 6. Fragment of the code to centre monomer1 in the centre of coordinates.

Following the same procedure as with monomer1, the geometric centre of monomer2 is calculated, centred in the origin of coordinates and rotated with another rotation matrix different from the one used to rotate monomer1. The next step is to translate monomer2 to a desired distance. This distance is defined between the geometric centres of the two molecules and it is calculated in the following manner. First, the interatomic distances between all the atoms within the monomers is calculated, as shown in Figure 7. First, the pairs between the atoms coordinates from which the distance is calculated "pairs1x, pairs1y, pairs1z" are formed. Then, the Euclidean distance between each pair of atoms is computed and stored in a list called "differences1". Later, the largest distances "max_value1" for each monomer are searched, and called diameters d_1 and d_2 (see Figure 8). Then, imaginary spheres are created around each monomer to avoid the overlap between the atoms of the molecules. The radius of these spheres will be half of the "max_value1", called radius r_1 and r_2 (see Figure 8), plus twice the van der Waals radius of the biggest atom of each monomer. In the case of the system of study, for monomer1, the largest distance (d_1) corresponds to the distance between the oxygen atom forming the C=O double bond of the carboxyl group and the hydrogen atom of the amino group, 4.33 Å. Half of this distance will be the radius r_1, to which twice the van der Waals radius of carbon atom, 1.70 Å, which is the biggest one in this molecule, will be added. The same for monomer2, where the maximum distance is 10.45 Å, and the atom with the largest van der Waals radius is carbon, as in the case of monomer1. This is represented in

Figure 8. This procedure is repeated until obtaining 15 different orientations at the same distance. Also, a geometry in which both monomers are at very large distance is prepared to be able to calculate the relative energy of each geometry with respect to the geometry at large distance, *i.e.*, the interaction energy. This interaction energy is computed at PM6 level of theory using Gaussian09[40] software.

```
pairs1x = [(x_coord1[i],x_coord1[j]) for i in range(len(x_coord1))
            for j in range(i+1,len(x_coord1))]
pairs1y = [(y_coord1[i],y_coord1[j]) for i in range(len(y_coord1))
            for j in range(i+1,len(y_coord1))]
pairs1z = [(z_coord1[i],z_coord1[j]) for i in range(len(z_coord1))
            for j in range(i+1,len(z_coord1))]

differences1 = []
for i in range(len(pairs1x)):
    difference1 = math.sqrt((pairs1x[i][0] - pairs1x[i][1])**2
                  + (pairs1y[i][0] - pairs1y[i][1])**2
                  + (pairs1z[i][0] - pairs1z[i][1])**2)
    differences1.append(abs(difference1))

max_value1 = max(differences1) + 2*max_radius1
r1 = max_value1/2
```

Figure 7. Fragment of the code to calculate the distances between atoms.

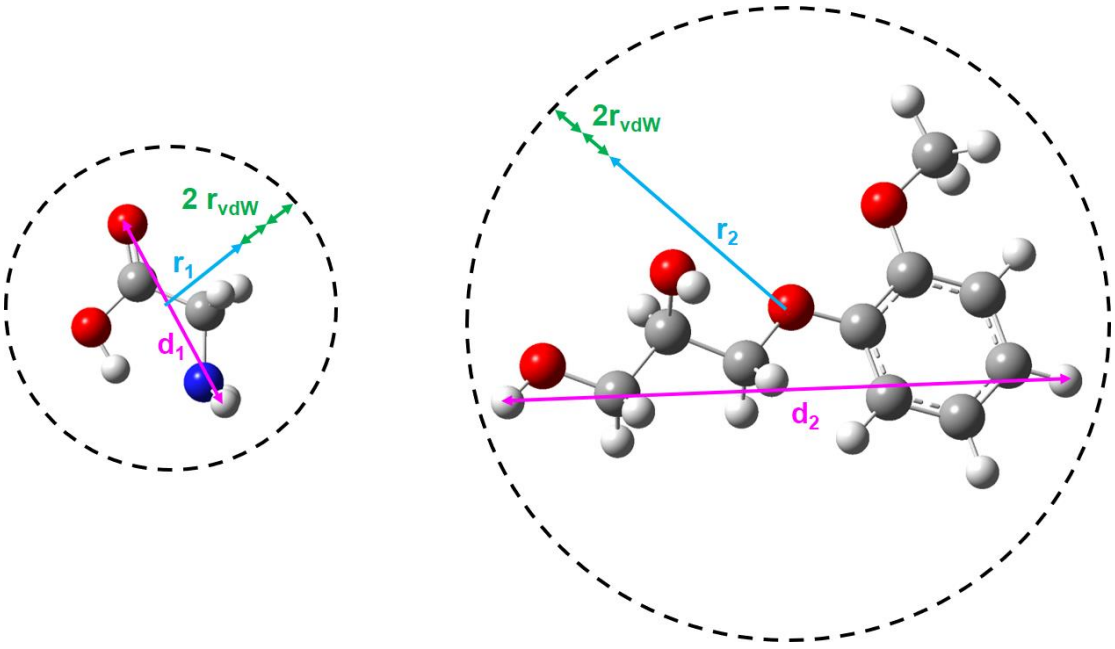


Figure 8. Schematic representation of the calculation of the distance between monomers.

The next goal is to find a few distances around the potential energy minimum for which to run the DFT calculations. For doing that, another code is used, whose first step searches the previously computed interaction energies for the 15 orientations and selects the geometry with the lowest PM6 interaction energy. At a second step, it changes the distance between the geometric centres of the monomers of the most favourable orientation in order to compute the potential energy curve at PM6 level for this orientation. For doing that, it calculates the geometric centres of the monomers and then, it changes the distance between these geometric centres. First decreasing the distance 0.1 Å until the interaction energy reaches a value of 20 kcal/mol and then, increasing 0.1 Å until the relative energy is 10 % of the interaction energy at the minimum (see Figure 9). The potential energy curve of the system is thus obtained, see Figure 10. It shows a minimum of -5.47 kcal/mol at 3.49 Å. The shorter distance is 3.06 Å with an energy of 22.82 kcal/mol and the largest distance is at 6.24 Å with an energy of -0.53 kcal/mol.

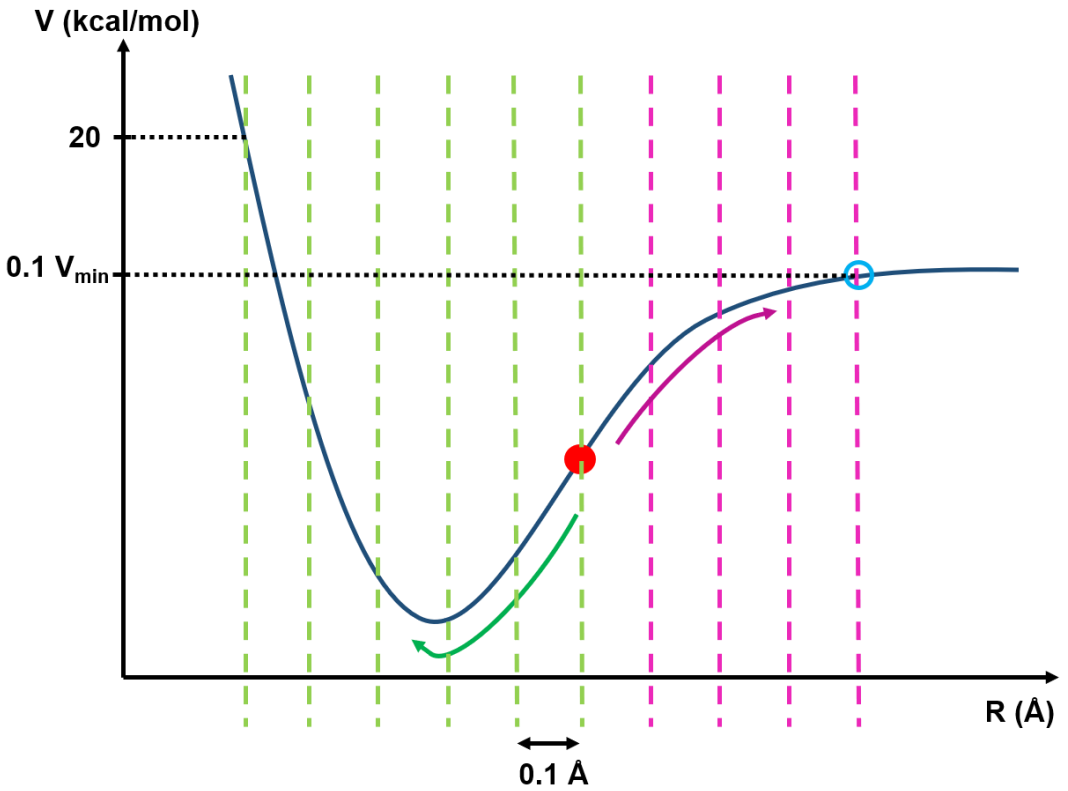


Figure 9. Schematic representation of the process followed to select the distances for which the interaction energy is computed.

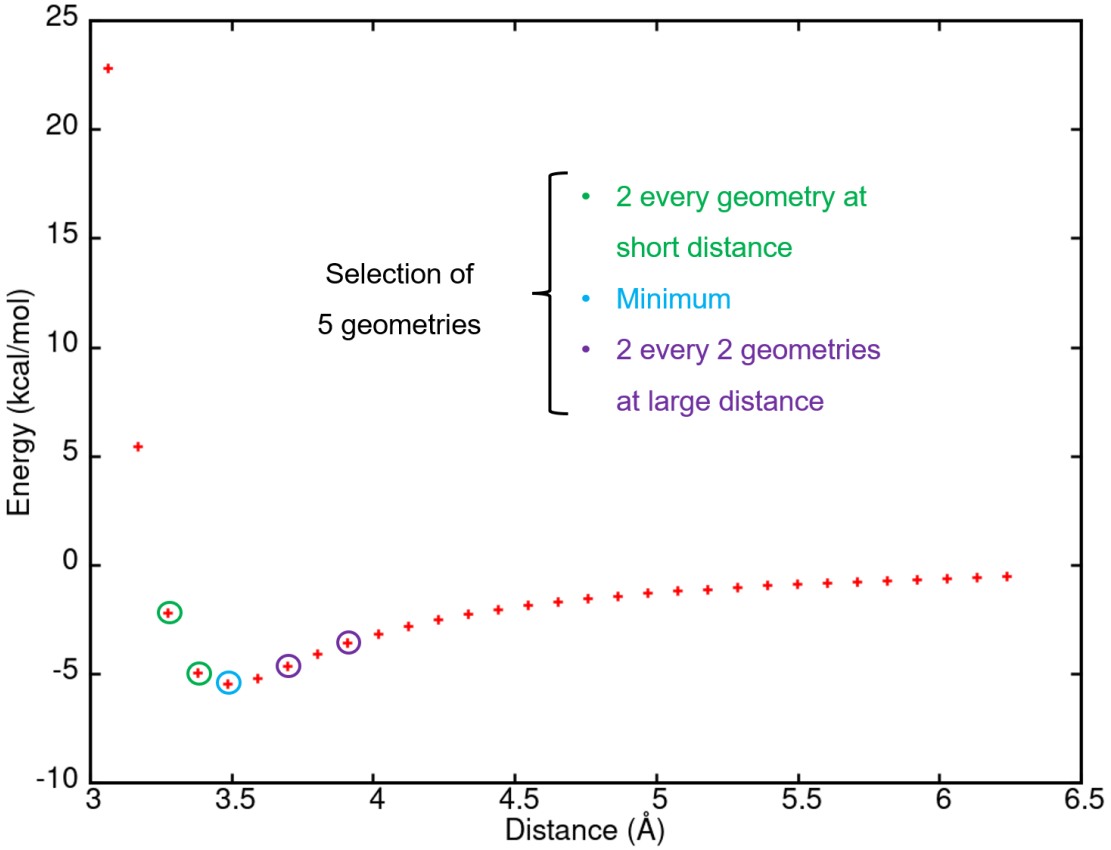


Figure 10. Potential energy surface for glycine - 3-(2-methoxyphenoxy)propane-1,2-diol system at PM6 level.

In the next step, a code was developed to select five geometries from the PM6 potential energy curve: the one corresponding to the energy minimum, two consecutive geometries at shorter distances and two geometries every two geometries at larger distances. Then, it changes the format of the coordinate files from *xyz* to *mol2* and join the five geometries in a single *mol2* file.

The *mol2* files are employed as input by the MoBioTools kit[41] to create the input files for the energy calculation with the M06-2X functional and 6-31G* basis function using the Gaussian09[40] software. In order to use MoBioTools kit,[41] we need to create two files ("main.inp" and "tpl.inp"). The first file, "main.inp" (Figure 11A), has one section (*&main*) in which one provides the QM software (*tpl*), the trajectory file where the geometries are located (*traj*), the QM region (*qmmask*), and the indices of the geometries for which the input files want to be generated. The second file, "tpl.inp" file (Figure 11B), is divided in five sections: (i) *&header* corresponds to the general instructions in Gaussian; (ii) *&route* is specific for Gaussian and defines the methodology, the basis set and the type of calculation; (iii) *&chgspin* includes the charge and spin multiplicity of the

monomers and the complex; and (iv) *&bsse* provides the atom indices to divide the system into two fragment, which are needed to calculate the interaction energy and to correct for the basis set superposition error (BSSE).[31] After running MoBioTools,[41] the input files for the energy calculation for Gaussian for each monomer and for the complex are generated.

A

```
&main
tpl = gaussian
traj = pes_full.mol2
qmmask = :GLY1,LIG1
geoms = 0 4
&end
```

B

```
&header
%chk=geom.chk
&end

&route
#p M062X/6-31g*
&end

&chgspin
0 ,1,0,1,0 ,1
&end

&bsse
mon1 = @1-10
mon2 = @11-38
&end
```

C

```
mon1_geom0
mon2_geom0
complex_geom0
HF
!prin
!vect
thre 9
```

Figure 11. (A) "main.inp" and (B) "tpl.inp" input files for MoBioTools[41] and (C) EDA input files.

Once the three files for the system (monomer1, monomer2 and complex) are created, the energy calculation with the M06-2X functional and 6-31G* basis set is performed using the Gaussian09[40] software, from which the EDA program will take the electron densities to carry out the interaction energy decomposition. However, before running the EDA calculation, it is necessary to change the *chk* files from Gaussian to *fchk* format (human readable), create an input file which contains the names of the monomers and complex of the Gaussian formatted checkpoint files (*fchk*), the density code, that is the method used to calculate the density, in this case DFT although HF is written, because this "HF" keyword is valid for both methods, and the threshold employed for the two-electron integrals (9 is an optimal value often used), see Figure 11C.

Then, the EDA program developed by Mandado *et al*.[27] is run to obtain the decomposition of the energy in its different contributions, namely, electrostatic, Pauli, dispersion and induction, represented in Figure 12, along the intermolecular distance. As can be seen, for this particular complex, the electrostatic, polarization, dispersion and induction contributions go to more negative energies when the monomers get closer. On the other hand, the Pauli repulsion term, becomes more positive when the monomers approach each other. Thus, the DFT total energy presents a minimum of -7.36 kcal/mol at 3.49 Å, which coincides with the minimum obtained from the PES at PM6 level of theory. Also from this EDA calculation, the change of polarization involved in the interaction of the monomers can be obtained such that the chemical groups which modify its electron density can be identified. Figure 13 displays the polarization deformation density of the system of study at the minimum of energy obtained from EDA method, where it is seen that there is electron density transfer from the amino acid to the alcohol group of the drug. All the steps described here were performed for the 291 created systems and the results are represented in the Annexes.

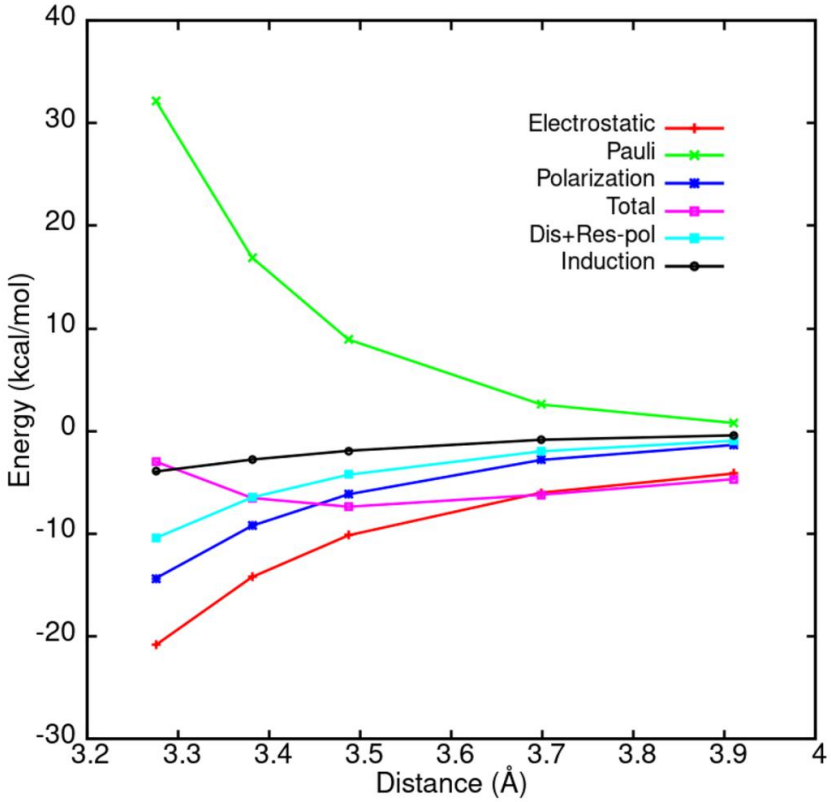


Figure 12. Energy decomposition analysis of glycine amino acid and 3-(2-methoxyphenoxy)propane-1,2-diol molecule.

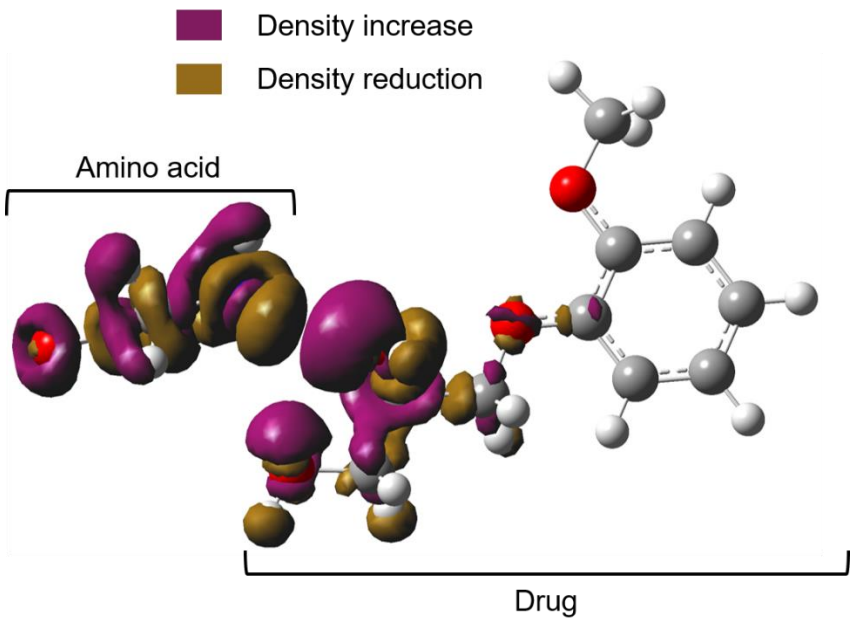


Figure 13. Polarization deformation density change in the geometry of minimum energy.

4.3 Interaction Energy

In this section the total interaction energy between the monomers for all the systems investigated will be discussed. Figure 14 and Figure 15 represent the distribution of the total interaction energy for all the systems, whose total average energy is -4.89 kcal/mol, and the distribution of the total interaction energy per atom, whose average is 0.10 kcal/mol. In addition, the system with highest (negative) total interaction energy and the system with the highest total interaction energy per atom are represented in Figure 14 and Figure 15. The first one provides the drug that interact in the strongest way with the amino acids, while the second one tells us which drug presents the strongest interaction types independently on the size of the molecule. It can be observed that the complex corresponding to the highest total interaction energy is very large size, a fact which is not surprising since a large number of atoms leads to a large number of attractive interactions. It is more interesting to see that the drug with the strongest interaction per atom is a very small molecule consisting of only 12 atoms. However, the molecule contains 6 fluorine atoms, which are very electronegative and induce large permanent charge separation in the molecule, leading to strong electrostatic interactions. As we will discuss later, electrostatic interactions are dominant for many of the drugs investigated here.

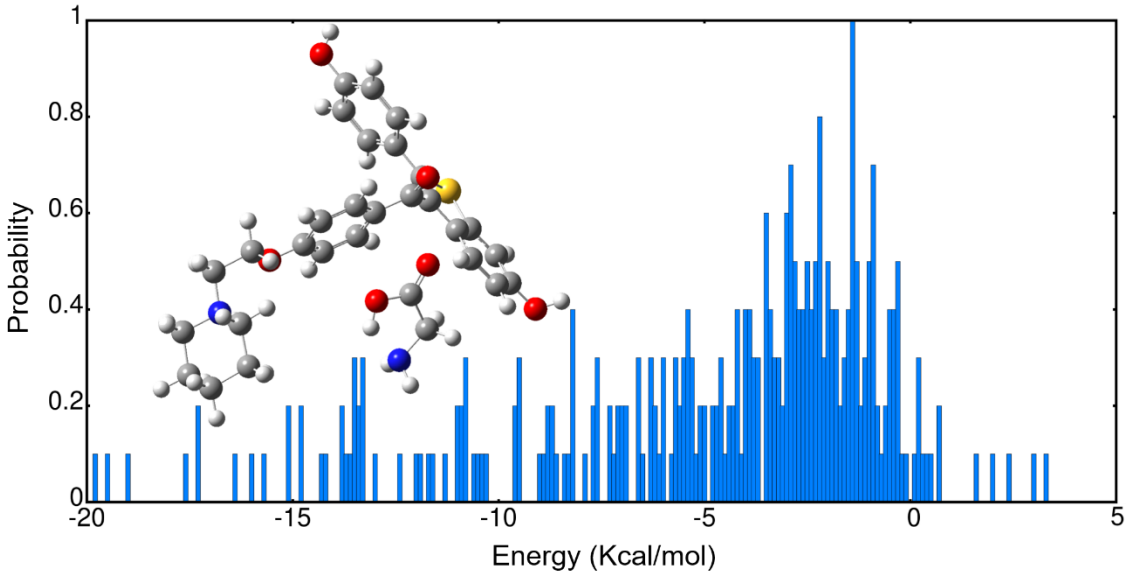


Figure 14. Total interaction energy distribution of all the systems and representation of the system with highest energy. Colour code: carbon atoms in grey, hydrogens in white, oxygens in red, nitrogens in blue and sulphurs in yellow.

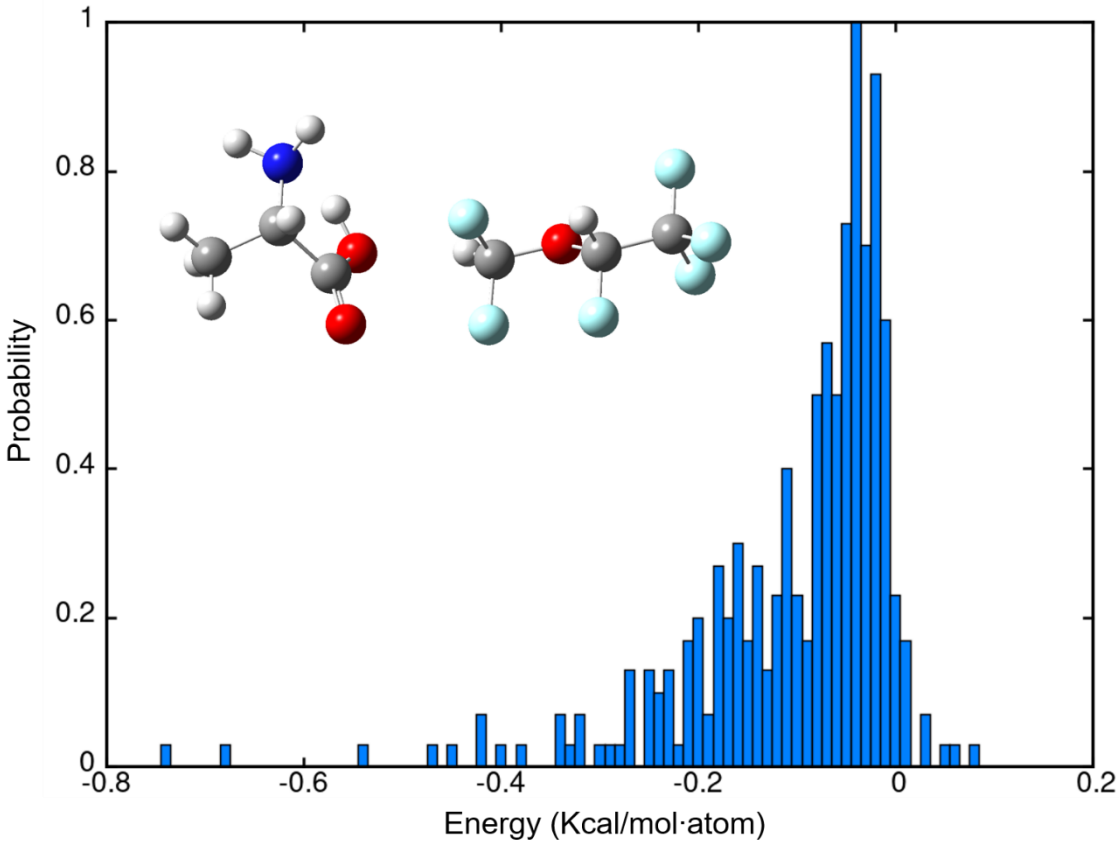


Figure 15. Total interaction energy per atom distribution of all the systems and representation of the system with highest energy. Colour code: carbon atoms in grey, hydrogens in white, oxygens in red, nitrogens in blue and fluorines in light blue.

4.4 Pauli Repulsion

The Pauli repulsion energy contribution is now analysed. Figure 16 and Figure 17 shows the Pauli repulsion distribution and the Pauli repulsion per atom distribution for all the systems, whose mean values are 9.15 and 0.19 kcal/mol, respectively. Moreover, the insets of both figures display the molecule with the largest Pauli repulsion and the molecule with the largest Pauli repulsion per atom. In this case, both molecules are the same. This can be understood by realizing that the Pauli repulsion energy is a short-range energy contribution, and, therefore depends only on the atoms which are in close contact and is independent on the size of the molecule.

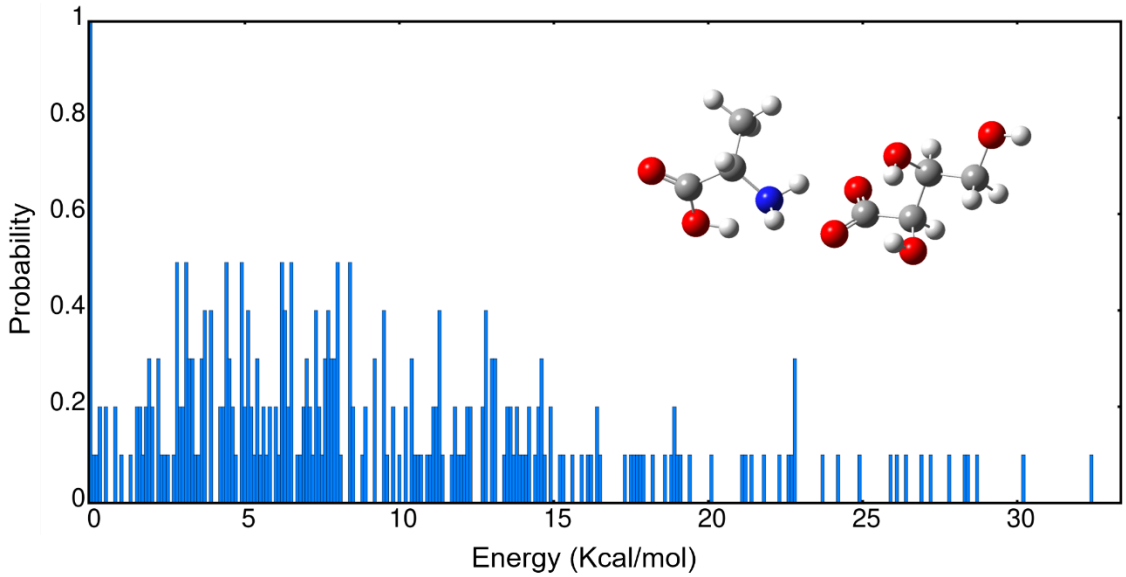


Figure 16. Pauli repulsion energy distribution of all the systems and representation of the system with highest repulsion energy. Carbon atoms in grey, hydrogens in white, oxygens in red and nitrogens in blue.

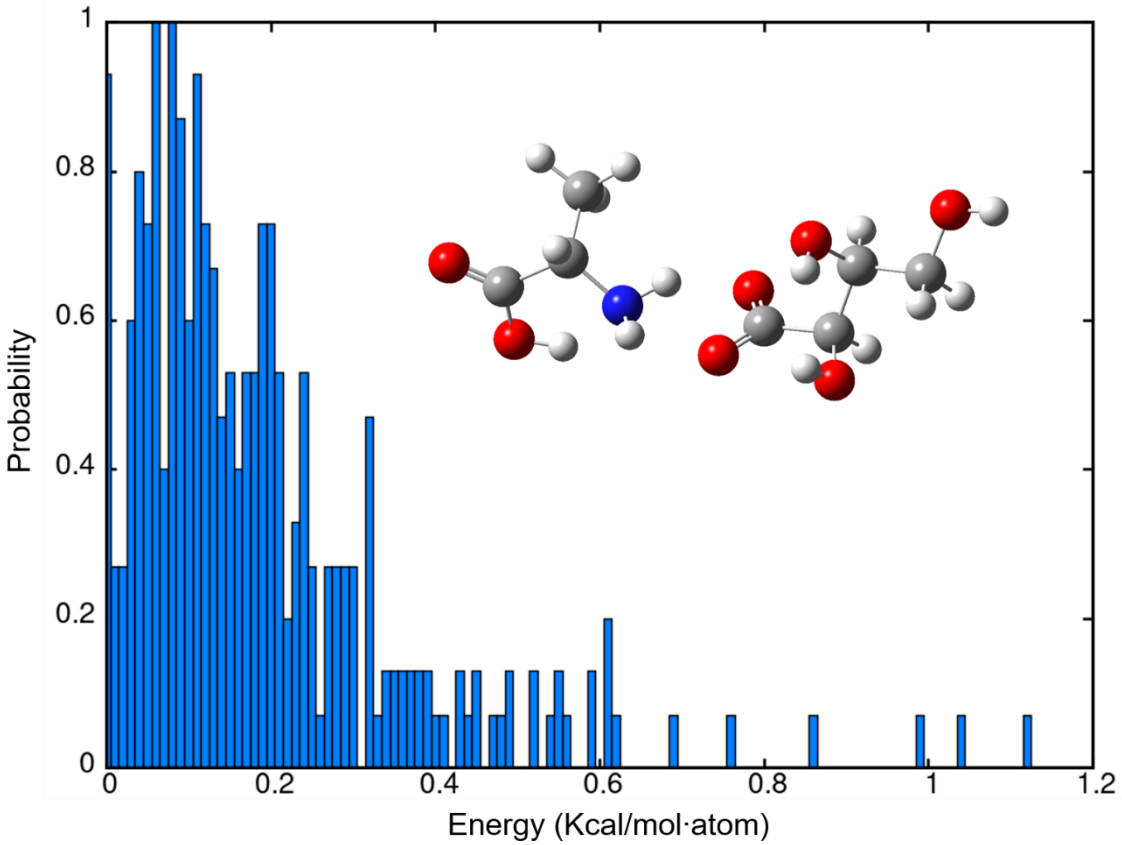


Figure 17. Pauli repulsion energy per atom distribution of all the systems and representation of the system with highest repulsion energy. Carbon atoms in grey, hydrogens in white, oxygens in red and nitrogens in blue.

4.5 Attraction Energy

In this section, only the attractive energy terms, namely electrostatic, dispersion and induction, will be taken into account. The first analysis performed consists of determining the percentage of each of the three attractive terms with respect to the total attractive energy for each system and calculate the distribution of percentages of each of these three terms. In Figure 18, it can be observed that most of the systems have a percentage of electrostatic contribution between 55-70% with a maximum value at 60% and an average value of 58%. In Figure 19, it is shown that dispersion represent between 20-40% of the attractive energy with a maximum and a mean value around 30%. Finally, Figure 20 shows that most of the systems present an induction energy contributing between 5-20% to the total attraction, with a maximum value of 8-9% and an average value of 12%.

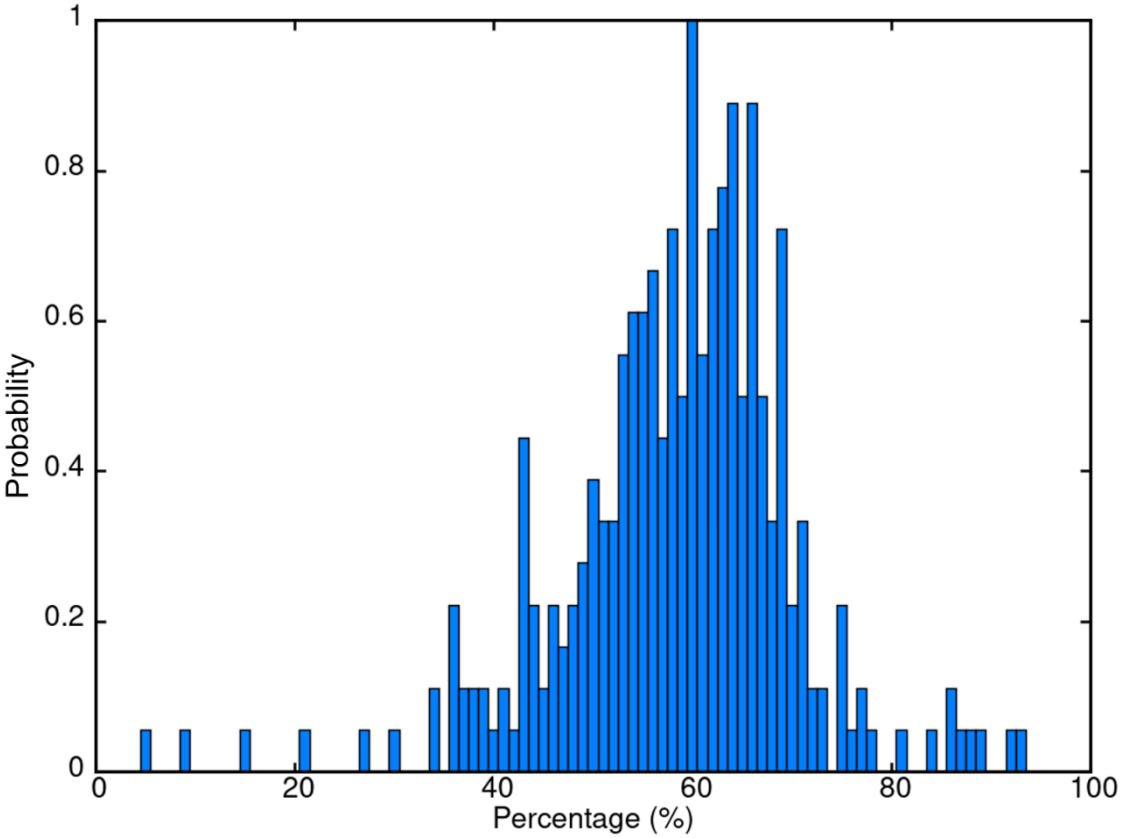


Figure 18. Distribution of electrostatic percentages to the total attraction energy.

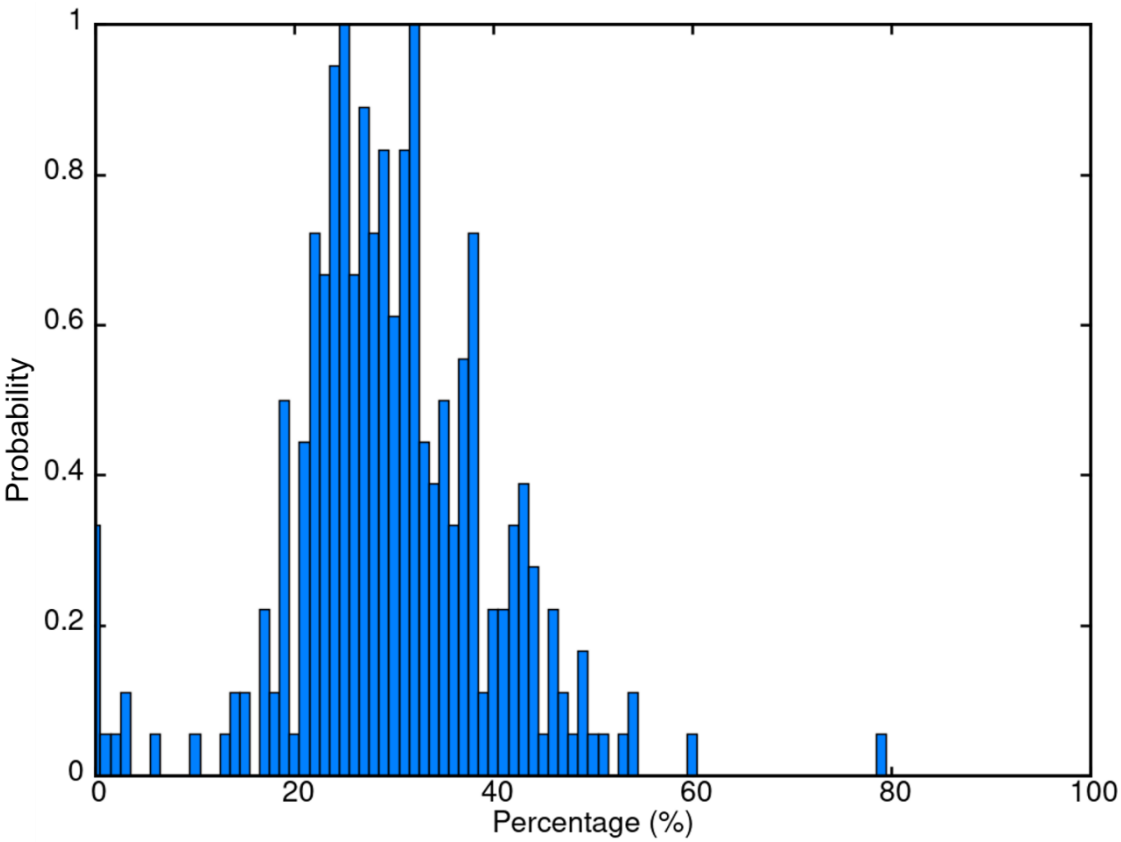


Figure 19. Distribution of dispersion percentages to the total attraction energy.

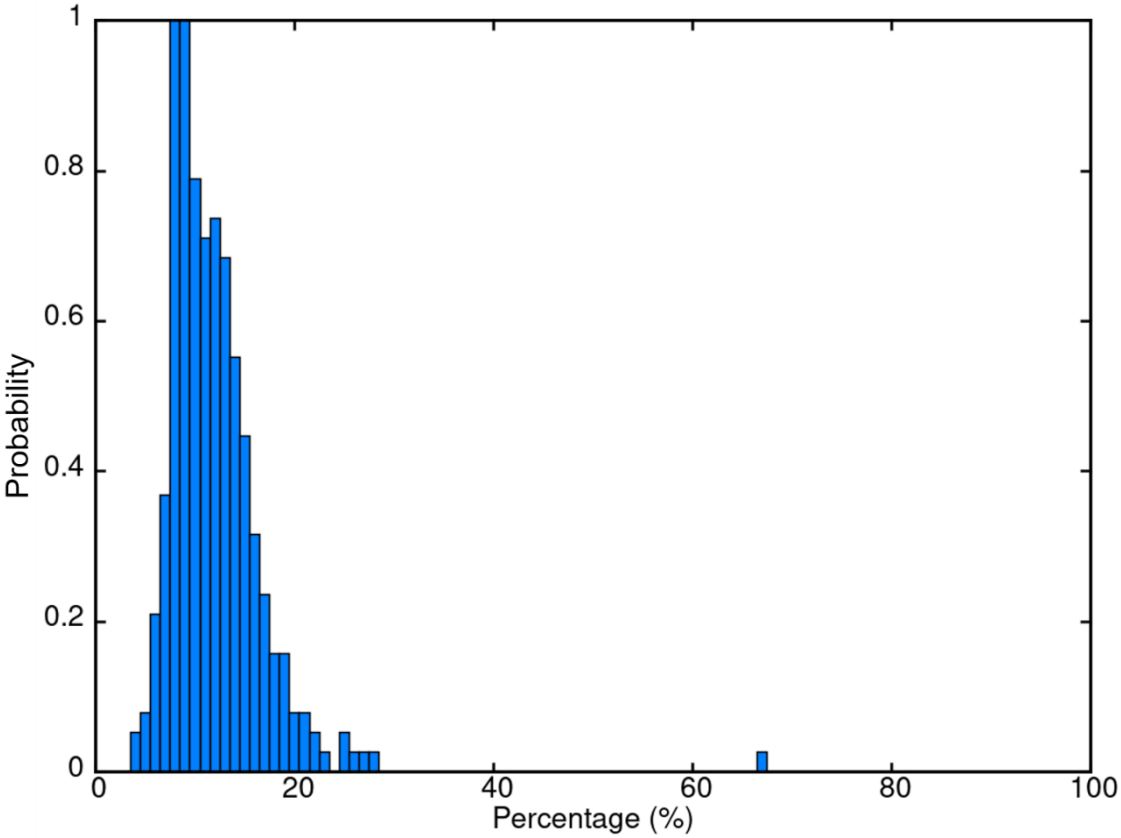


Figure 20. Distribution of induction percentages to the total attraction energy.

Then, in order to analyse whether there exists a relation between the chemical structure of the drugs and the EDA contributions, different analysis have been carried out. First, the complexes with percentage values of electrostatic, dispersion and induction contributions higher than a selected threshold have been selected and represented to see if there exist any evident feature common to most of them, which can be identified by simple visualization. It was found that 31 systems have an electrostatic contribution higher than 70%, 44 systems have the dispersion percentage higher than 40% and 65 systems have the induction percentage higher than 15%. When those systems were plotted, no similar structural characteristic was identified. Therefore, the percentage thresholds were increased until only 10 systems satisfy the new chosen tighter criterion, resulting in percentages of 78%, 49% and 22% for electrostatic, dispersion and induction, respectively. However, once again no common features among the molecules were evident by visualization.

Therefore, it was clear that another strategy was needed to find a relation between EDA and the chemical structure. A new analysis was performed where it is analyse the presence of eight of the most common functional polar groups presented in bioactive molecules[43]

and four aromatic rings common in the studied drugs. These functional groups are shown in Figure 21, where *R* means any aliphatic or aromatic carbon. For doing that, it is necessary to change the *mol2* format to *SMILES*, which is a file format with a linear text that describes the connectivity and chirality of the molecules. This is performed using OpenBabel[44] software. In addition, "rdkit"[45] Python module has been used to check if the functional group is presented and how many times it is present.

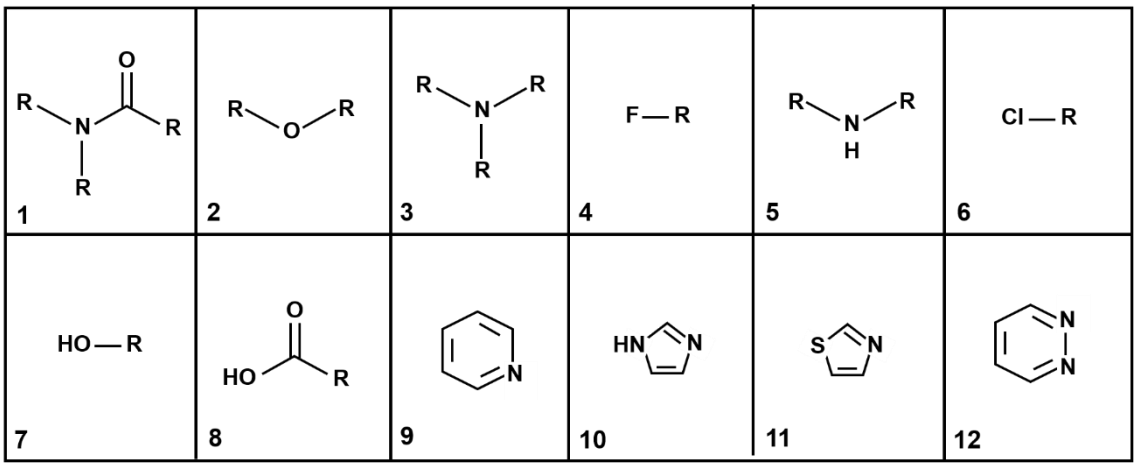


Figure 21. Most frequent functional groups in bioactive molecules.

From this analysis, the percentage of systems that present each of the functional groups is obtained. In Figure 22, it can be observed that the most frequent functional group in our systems is the functional group 7 (an alcohol), which is present in around 74% of the investigated drugs, followed by a secondary amine (group 5) with 72% of population. In the next level, the tertiary amine (group 3) and the ether (group 2) with 46% and 42% population, respectively, are found. The carboxyl group 8 is found in 29% of the drugs, then the groups 4 and 6, which contain halogen atoms, are found in 11% of the drugs. Finally, group 1 and the four aromatic rings, with 10% and 5% population, respectively, are the least common chemical groups.

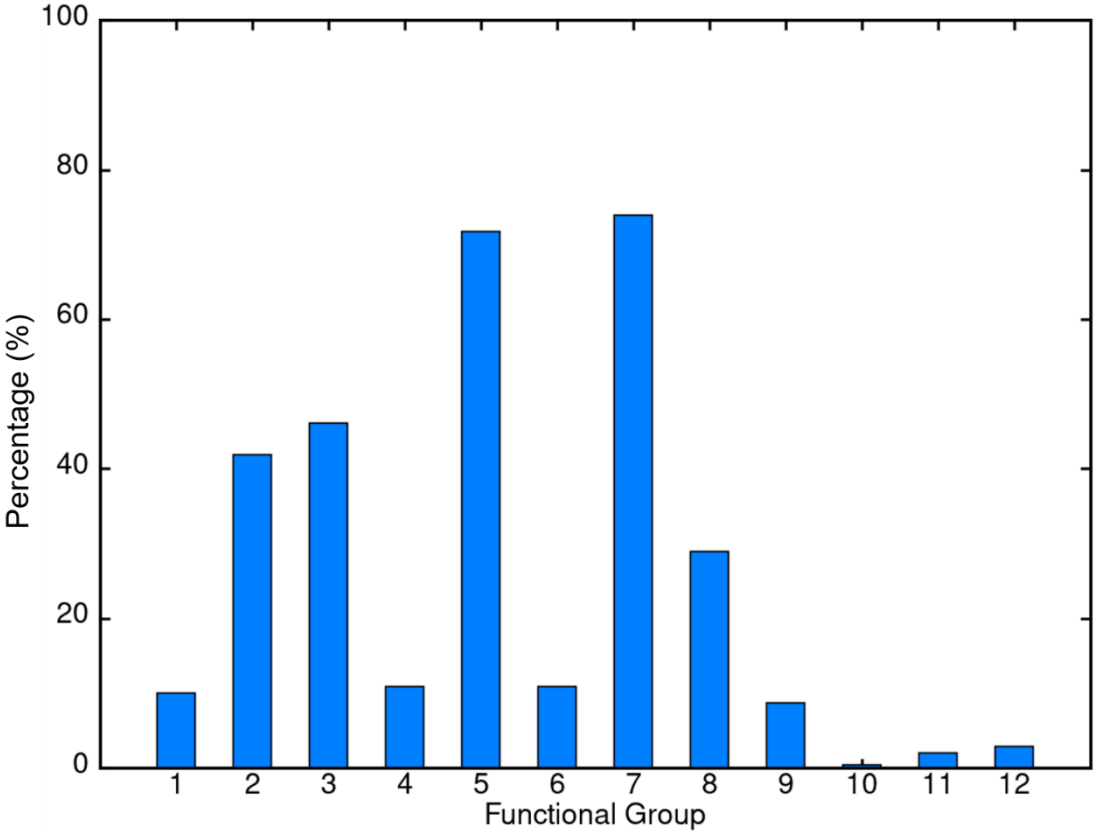


Figure 22. Percentage of systems that contain each of the 12 functional groups.

Afterwards, the average of each attractive energy contribution (electrostatic, dispersion and induction) is computed for the sets of systems that present each of the functional groups mentioned above. This is represented in Figure 23, where it is observed that the predominant term for all the groups is the electrostatic contribution. This can be explained by the fact that all the functional groups are polar with electronegative atoms able to form strong interactions between the permanent charge distributions of the molecules. Although the percentages are similar along all the sets of molecules, the group with highest electrostatic percentage is the one that contains the group 11. As can be seen in Figure 21, this group is a small ring with a high electron density due to the presence of a nitrogen and a sulphur atom which can likely participate in strong electrostatic interactions. In the case of dispersion and induction, the drugs with the largest contributions are the ones that possess the group 6, the one with the chlorine atom. Since this atom is easily polarizable due to its size, its electron cloud is more deformable and can participate in polarization interactions.

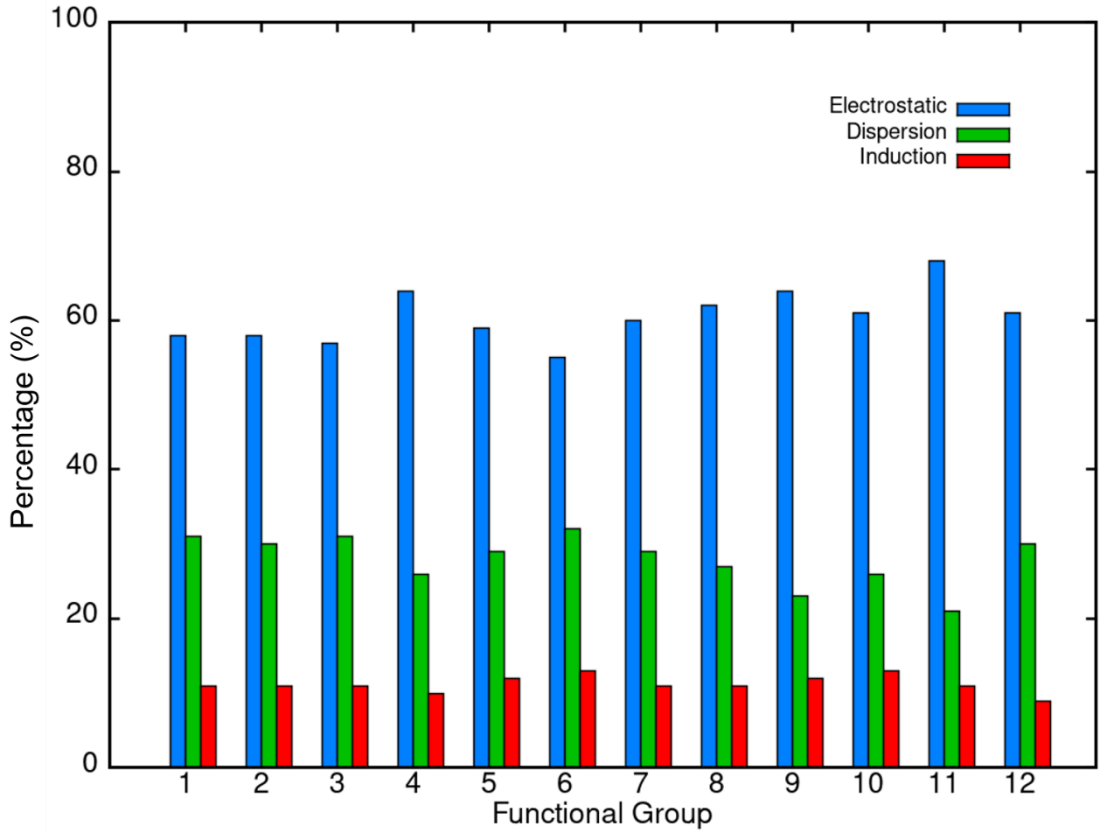


Figure 23. Electrostatic, dispersion and induction percentages of the systems which present each of the 12 functional groups.

Now, the two systems with the highest electrostatic, dispersion and induction percentages will be analysed in detail to check if there are some common features. In Figure 24, the two systems with the highest electrostatic percentages are represented. System 97 is composed by glycine amino acid and 4-(3-chloro-4-(3-cyclopropylureido)phenoxy)-7-methoxyquinoline-6-carboxamide molecule, and it presents 93% of electrostatic contribution. On the other hand, the system 152 is formed by glycine amino acid and 2-(((4-(3-methoxypropoxy)pyridin-2-yl)methyl)thio)-1H-benzo[d]imidazole molecule, presenting 92% of electrostatic contribution. In both systems, the presence of electronegative atoms in the organic molecule near the amino acid could lead to the formation of hydrogen bonds contributing to the high electrostatic percentages, although an analysis of hydrogen bond population performed with *cpptraj*[46] software did not reveal the presence of hydrogen bonding. This means that the large electrostatic energy is caused by long-range interactions.

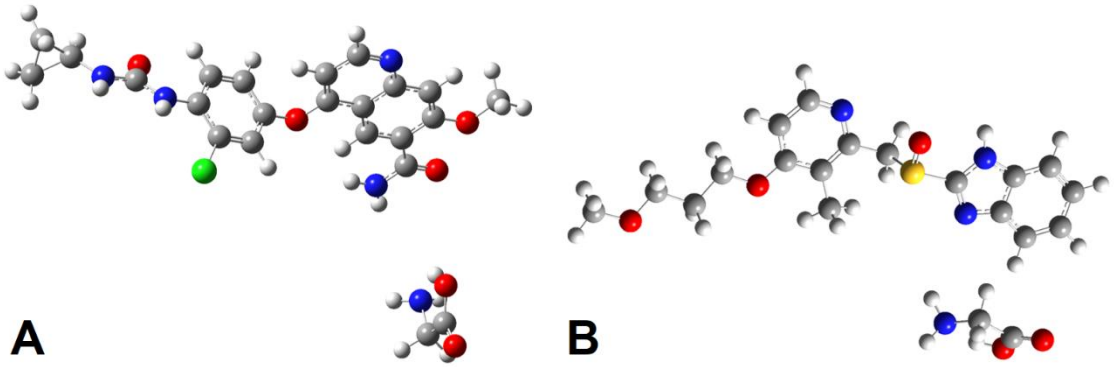


Figure 24. Systems with highest electrostatic percentage. (A) System 97 and (B) system 152. Colour code: carbon atoms in grey, hydrogens in white, oxygens in red, nitrogens in blue, chlorine in green and sulphur in yellow.

The two systems with the highest dispersion percentages to the total attractive energy are represented in Figure 25. System 84, with a 79% of dispersion, is composed by the glycine amino acid and the (Z)-(2-(4-(4-chloro-1,2-diphenylbut-1-en-1-yl)phenoxy)ethyl)dimethyl-l4-azane molecule, and system 5, which is formed by the glycine amino acid and the 1-(2-((4-chlorobenzyl)thio)-2-(2,4-dichlorophenyl)ethyl)-1H-imidazole drug, has a dispersion percentage of 60%. It is shown that system 84, which has only one chlorine atom, has higher dispersion than system 5, which present three chlorine atoms and a sulphur atom. The initial hypothesis that systems with a higher number of big atoms, whose electron cloud is more polarizable, could have higher dispersion percentage, is not satisfied or, at least, not exclusively. This could be due to the presence of other functional groups, for example, aromatic rings, which are also easily polarizable. However, Figure 23 shows that the drugs that contain aromatic rings have higher electrostatic contributions than dispersion and induction ones. This suggests that a more exhaustive analysis, including more functional groups at the same time during the analysis and classification of drugs, has to be performed.

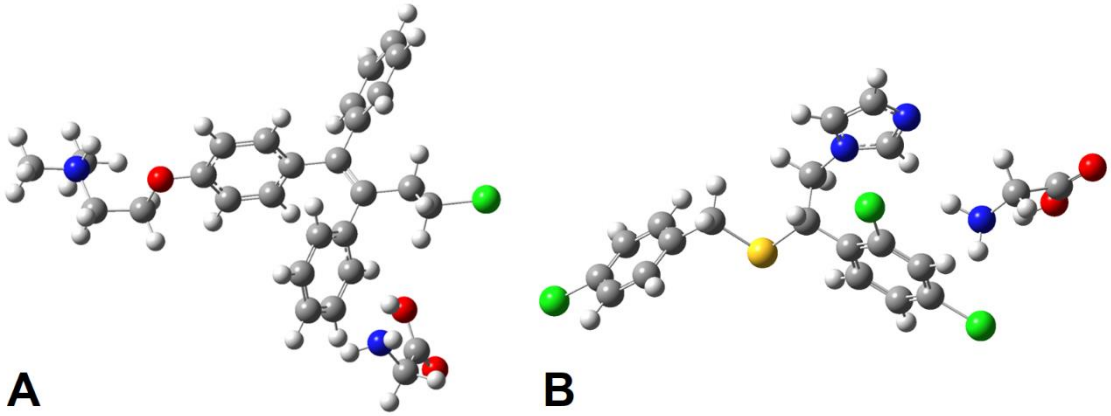


Figure 25. Systems with the highest dispersion percentage. (A) System 84 and (B) system 5. Colour code: carbon atoms in grey, hydrogens in white, oxygens in red, nitrogens in blue, chlorines in green and sulphurs in yellow.

The systems with largest induction percentages are represented in Figure 26. They are the complex 91, composed by glycine amino acid and 2,6-dichloro-N1-(imidazolidin-2-yl)benzene-1,4-diamine molecule, with a 67% of induction and the complex 277 with an induction percentage of 28%, which is formed by alanine amino acid and the same organic molecule of system 5, which also presents the largest dispersion. The analysis of these two systems is similar to the one of the systems with the highest dispersion percentages because dispersion and induction are encompassed in the polarization energy term of the EDA.

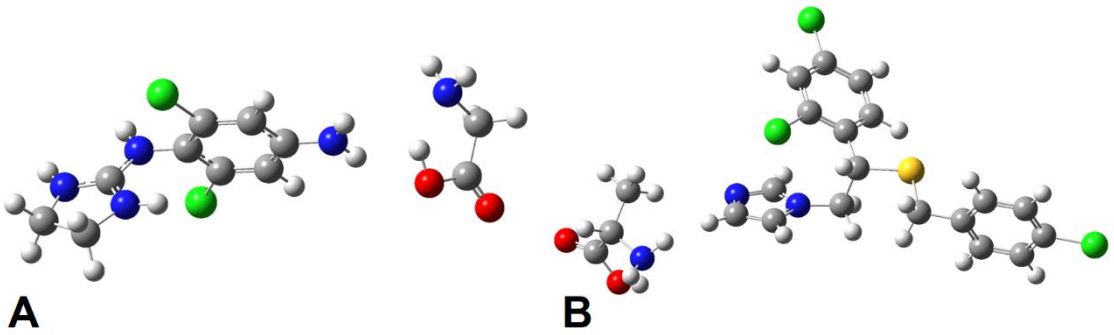


Figure 26. Systems with the highest induction percentage. (A) System 91 and (B) system 277. Colour code: carbons atom in grey, hydrogens in white, oxygens in red, nitrogens in blue, chlorines in green and sulphurs in yellow.

5 CONCLUSIONS

Proteins are involved in many processes in our organisms and their presence is essential for their appropriate functioning. Because one of most important functions carried out by proteins is the reception of drugs, the study of drug/protein interactions is crucial to understand the mode of action and side effects of those drugs and, thus, develop new ones with improved properties.

In this study, 20 amino acids and 300 drugs has been used to create 6000 systems, although 291 has finished at the time being and are the ones presented in this thesis. Then, different steps are performed in order to characterize the interaction energy: (i) find the most favourable orientation between the molecules of the complexes at PM6 level; (ii) compute the PM6 potential energy curve for this favourable orientation; (iii) compute the interaction energy at DFT level for 5 geometries around the minimum; (iv) perform an EDA to obtain the different energy components; (v) find a relation between the structure and the EDA. To accomplish all these steps, a Python code to carry out different tasks in an automatic manner has been developed.

From these calculations, the distribution of the total and Pauli repulsion energies and total and Pauli repulsion energies per atom have been obtained and the system with highest energy on each of these distributions has been compared. The system with the largest total interaction energy is much bigger than the one with the largest total interaction energy per atom because the long-range electrostatic contribution is the major component of the total interaction energy and, therefore, all atoms of the molecule are involved in the interaction. Contrary, the systems with the highest Pauli repulsion energy and highest Pauli repulsion energy per atom are the same because this energy contribution is a short-range term and, therefore, nearly independent on the system size.

In the case of attractive energy terms, namely electrostatic, dispersion and induction energies, all the distributions are relatively broad showing that the set of selected drugs is very diverse. The energy contribution that dominates the attractive energy for most of the systems is the electrostatic contribution.

A classification analysis of the drugs has been performed in order to try to classify the systems based on any structure/EDA relation. Specifically, the occurrence of eight of the most frequent polar functional groups in bioactive molecules and four non-polar groups was computed for the set of 291 drug/amino acid pairs. The most frequent functional group in the set of drugs employed was an alcohol and the least frequent ones were the

non-polar aromatic rings. However, the presence or absence of these functionals groups were not related with the EDA contributions. Therefore, more sophisticated analyses are needed in order to reveal any possible structure/energy relation.

6 REFERENCES

1. McKee, T.; McKee, J. R., *Bioquímica: La Base Molecular de la Vida*. 3 ed.; McGraw Hill Editorial: 2009.

2. Berg, J. M.; Tynoczko, J. L.; Stryer, L., *Bioquímica*. 6 ed.; Editorial Reverté: 2007.

3. Pollard, T. D.; Earnshaw, W. C., *Cell Biology*. 3 ed.; Elsevier: 2017.

4. Boyer, R., *Conceptos de Bioquímica*. 1 ed.; International Thomson Editores: 2000.

5. Taylor, M. R.; Dickey, J. L.; Simon, E. J.; Hogan, K.; Reece, J. B., *Campbell Biology: Concepts & Connections*. 9 ed.; Pearson Education: 2020.

6. Alberts, B.; Bray, D.; Hopkin, K.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P., *Essential Cell Biology*. 4 ed.; Garland Science: 2013.

7. Lodish, H.; Berk, A.; Matsudaira, P.; Kaiser, C. A.; Scott, M. P.; Zipursky, L.; Darnell, J., *Molecular Cell Biology*. 5 ed.; W. H. Freeman and Company: 2003.

8. Wang, Y. C.; Zhang, C. H.; Deng, N. Y.; Wang, Y., Kernel-based data fusion improves the drug-protein interaction prediction. *Computational Biology and Chemistry* **2011,** 35, (6), 353-362.

9. Lombardi, D.; Dittrich, P. S., Droplet microfluidics with magnetic beads: A new tool to investigate drug-protein interactions. *Analytical and Bioanalytical Chemistry* **2011,** 399, (1), 347-352.

10. Hage, D. S.; Jackson, A.; Sobansky, M. R.; Schiel, J. E.; Yoo, M. J.; Joseph, K. S., Characterization of drug-protein interactions in blood using high-performance affinity chromatography. *Journal of Separation Science* **2009,** 32, (5-6), 835-853.

11. Zheng, S.; Li, Y.; Chen, S.; Xu, J.; Yang, Y., Predicting drug-protein interaction using quasi-visual question answering system. *Nature Machine Intelligence* **2020,** 2, (2), 134-140.

12. Sharma, H.; Navalkar, A.; Maji, S. K.; Agrawal, A., Analysis of drug-protein interaction in bio-inspired microwells. *SN Applied Sciences* **2019,** 1, (8), 819.

13. Adcock, S. A.; McCammon, J. A., Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews* **2006,** 106, (5), 1589-1615.

14. Karplus, M.; McCammon, J. A., Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **2002,** 9, (9), 646-652.

15. Ruano, L.; Cárdenas, G.; Nogueira, J. J., The Permeation Mechanism of Cisplatin Through a Dioleoylphosphocholine Bilayer. *ChemPhysChem* **2021,** 22, (12), 1251-1261.

16. González, M. A., Force fields and molecular dynamics simulations. *Journées de la Neutronique* **2011,** 12, 169-200.

17. Nogueira, J. J. Molecular Dynamics - Chapter 2: Force Fields. https://www.youtube.com/watch?v=vpqo0q3zvJU&t=1526s (15/05/2022).

18. Harrison, J. A.; Schall, J. D.; Maskey, S.; Mikulski, P. T.; Knippenberg, M. T.; Morrow, B. H., Review of force fields and intermolecular potentials used in atomistic computational materials research. *Applied Physics Reviews* **2018,** 5, (3), 031104.

19. Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Advances in Protein Chemistry* **2003,** 66, 27-85.

20. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D., CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2016,** 14, (1), 71-73.

21. Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J., Polarization effects in molecular mechanical force fields. *Journal of Physics Condensed Matter* **2009,** 21, (33), 333102.

22. Niketic, S. R.; Rasmussen, K., *The Consistent Force Field: A Documentation*. 1 ed.; Springer: 1977.

23. Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *Journal of the American Chemical Society* **1992,** 114, (25), 10024-10035.

24. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biology* **2011,** 9, 71.

25. Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K. R., Machine Learning Force Fields. *Chemical Reviews* **2021,** 121, (16), 10142-10186.

26. Hurwitz, J.; Kirsch, D., *Machine Learning For Dummies*. IBM Limited Edition: 2018.

27. Mandado, M.; Hermida-Ramón, J. M., Electron density based partitioning scheme of interaction energies. *Journal of Chemical Theory and Computation* **2011,** 7, (3), 633-641.

28. Ramos-Berdullas, N.; Pérez-Juste, I.; Van Alsenoy, C.; Mandado, M., Theoretical study of the adsorption of aromatic units on carbon allotropes including explicit (empirical) DFT dispersion corrections and implicitly dispersion-corrected

functionals: The pyridine case. *Physical Chemistry Chemical Physics* **2015,** 17, (1), 575-587.

29. Sterling, T.; Irwin, J. J., ZINC 15 - Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling* **2015,** 55, (11), 2324-2337.

30. Jensen, F., *Introduction to Computational Chemistry*. 3 ed.; John Wiley & Sons: 2017.

31. Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **2002,** 100, (1), 65-73.

32. Thiel, W., Semiempirical quantum-chemical methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014,** 4, (2), 145-157.

33. Ramachandran, K. I.; Deepa, G.; Namboori, K., *Computational Chemistry and Molecular Modeling: Principles and Applications*. 1 ed.; Springer: 2008.

34. Rezác, J.; Fanfrlík, J.; Salahub, D.; Hobza, P., Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes. *Journal of Chemical Theory and Computation* **2009,** 5, (7), 1749-1760.

35. Andrés, J.; Bertran, J., *Theoretical and Computational Chemistry: Foundations, Methods and Techniques*. 1 ed.; Publicacions de la Universitat Jaume I: 2007.

36. Hohenberg, P.; Kohn, W., Inhomogeneous electron gas. *Physical Review* **1964,** 136, (3B), B864-B871.

37. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** 120, (1-3), 215-241.

38. Valero, R.; Costa, R.; De P. R. Moreira, I.; Truhlar, D. G.; Illas, F., Performance of the M06 family of exchange-correlation functionals for predicting magnetic coupling in organic and inorganic molecules. *Journal of Chemical Physics* **2008,** 128, (11), 114103.

39. Cárdenas, G.; Pérez-Barcia, A.; Mandado, M.; Nogueira, J. J., Characterization of cisplatin/membrane interactions by QM/MM energy decomposition analysis. *Physical Chemistry Chemical Physics* **2021,** 23, (36), 20533-20540.

40. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe,

D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 09*, Wallingford CT, 2016.

41. Cárdenas, G.; Lucia-Tamudo, J.; Mateo-delaFuente, H.; Palmisano, V. F.; Anguita-Ortiz, N.; Ruano, L.; Pérez-Barcia, A.; Díaz-Tendero, S.; Mandado, M.; Nogueira, J. J., MoBioTools: A Toolkit to Setup QM/MM Calculations. *ChemRxiv. Cambridge: Cambridge Open Engage* **2022**.

42. Ropo, M.; Schneider, M.; Baldauf, C.; Blum, V., First-principles data set of 45,892 isolated and cation-coordinated conformers of 20 proteinogenic amino acids. *Scientific Data* **2016,** 3, 160009.

43. Ertl, P.; Altmann, E.; McKenna, J. M., The Most Common Functional Groups in Bioactive Molecules and How Their Popularity Has Evolved over Time. *Journal of Medicinal Chemistry* **2020,** 63, (15), 8408-8418.

44. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An Open chemical toolbox. *Journal of Cheminformatics* **2011,** 3, (10), 33.

45. RDKit: Open-Source cheminformatics. https://www.rdkit.org

46. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* **2013,** 9, (7), 3084-3095.

7 ANNEXES

The potential energy curve of the 291 systems are represented on the left-hand side with the insets displaying the molecules. On the right-hand side, the decomposition of the energy in its different contributions along the intermolecular distance is represented.

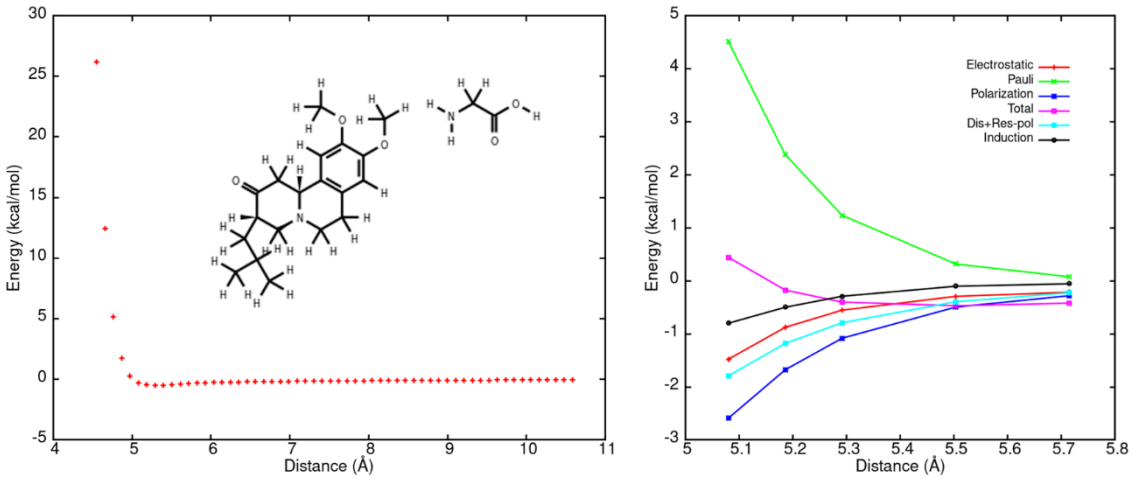


Figure 1: System 1

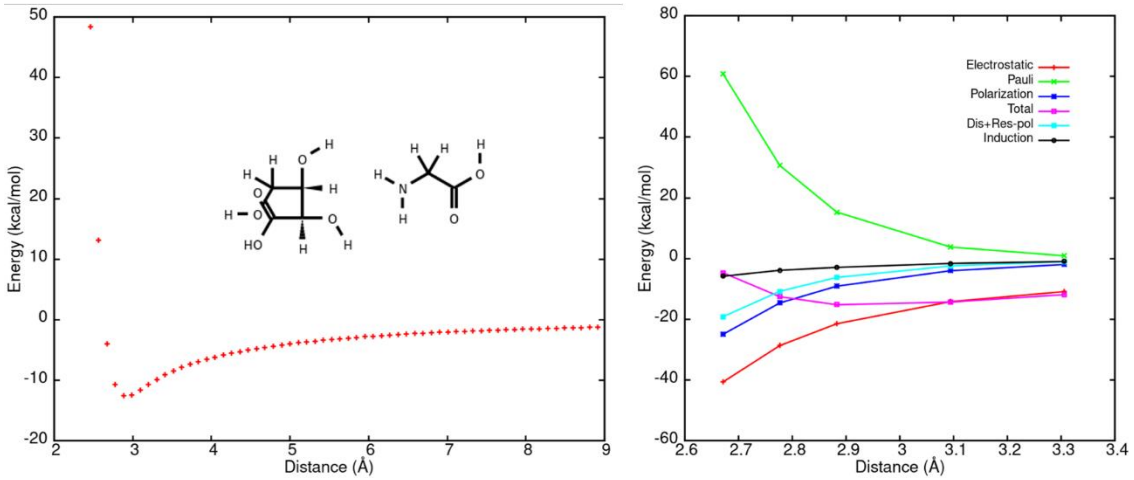


Figure 2: System 2

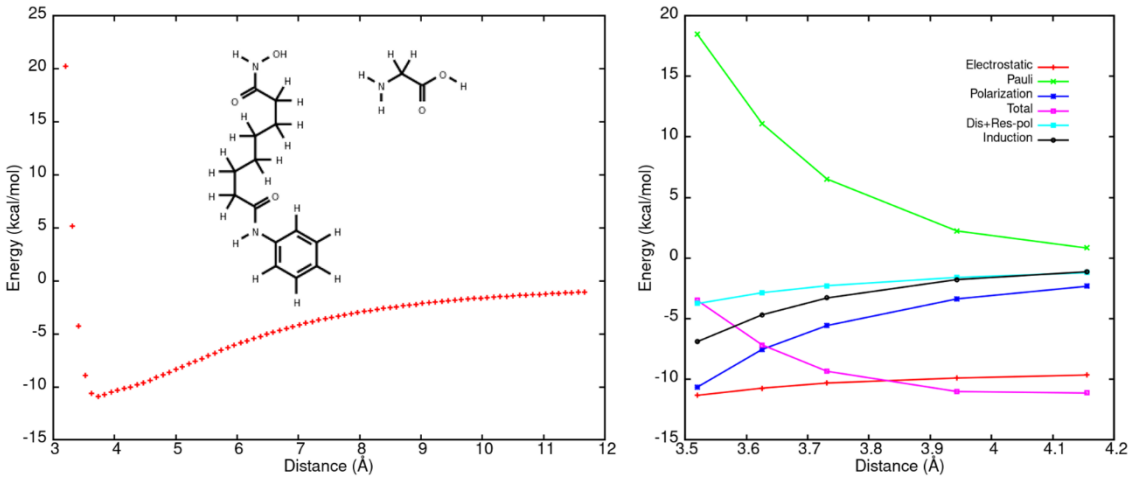


Figure 3: System 3

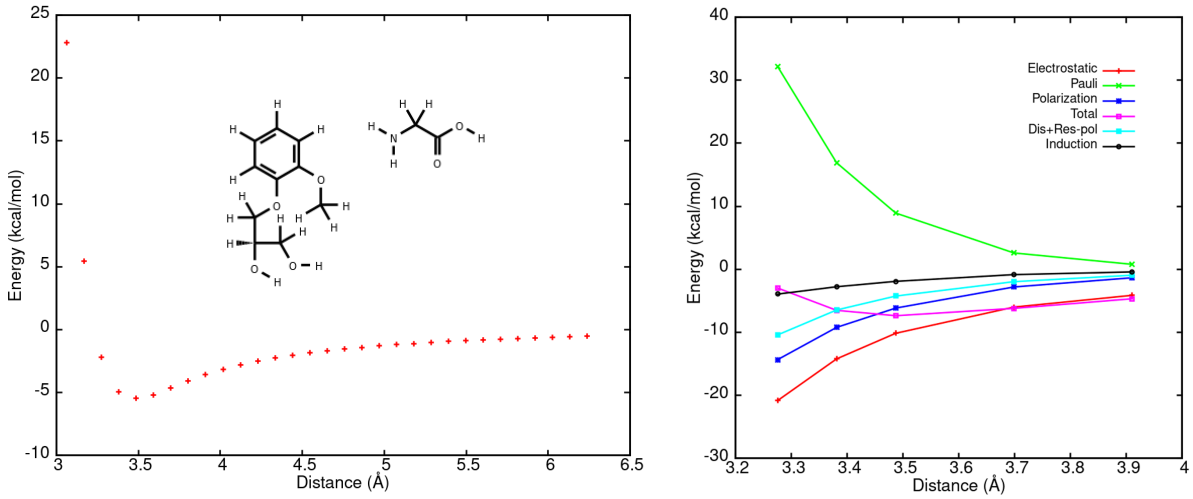


Figure 4: System 4

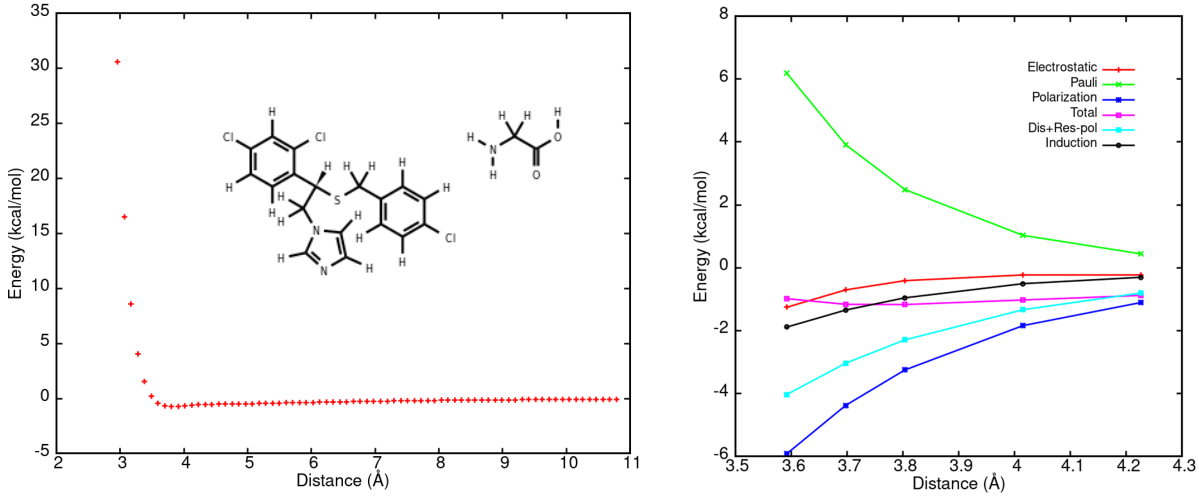


Figure 5: System 5

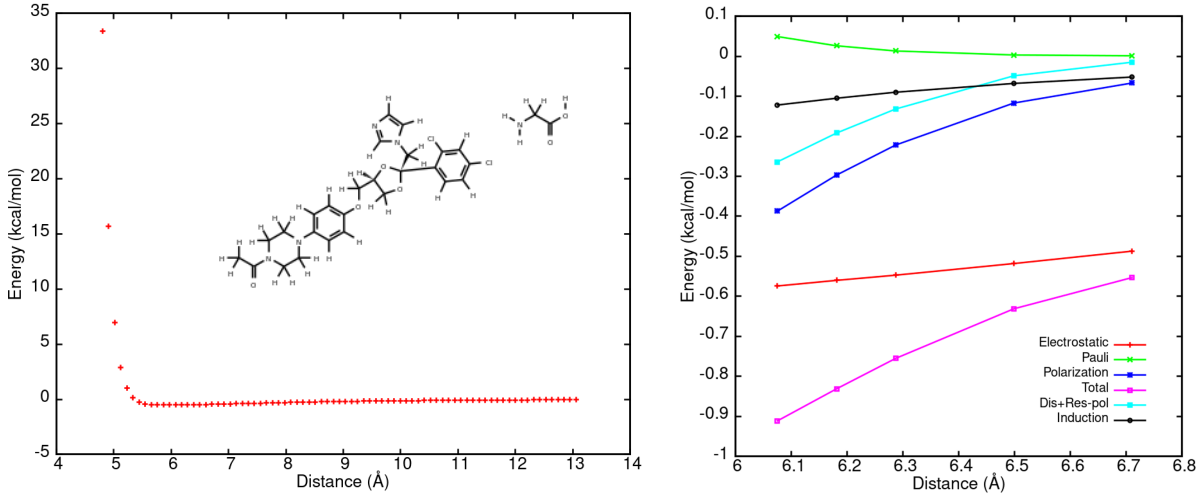
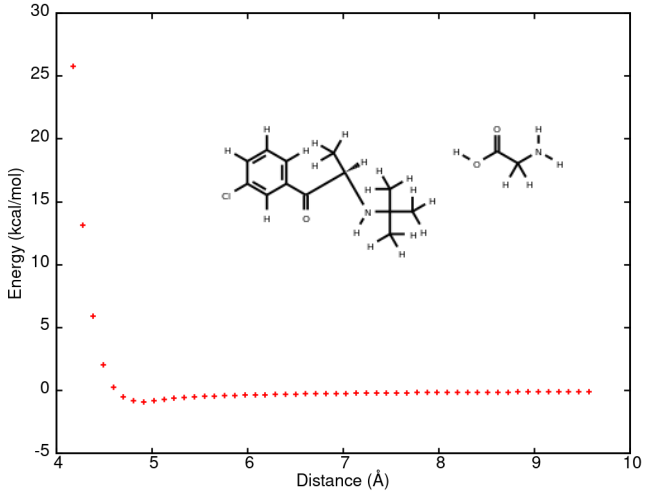


Figure 6: System 6



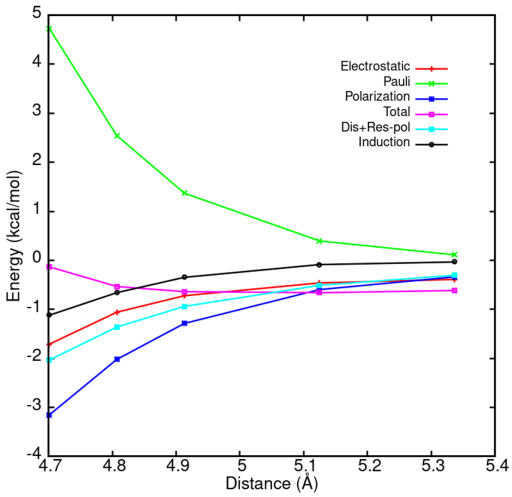
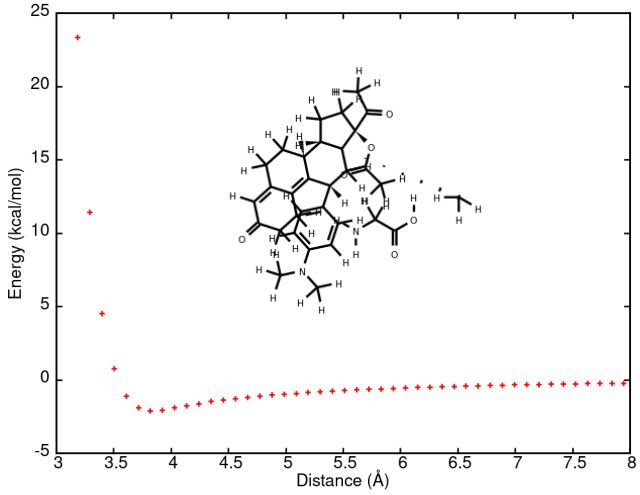


Figure 7: System 7



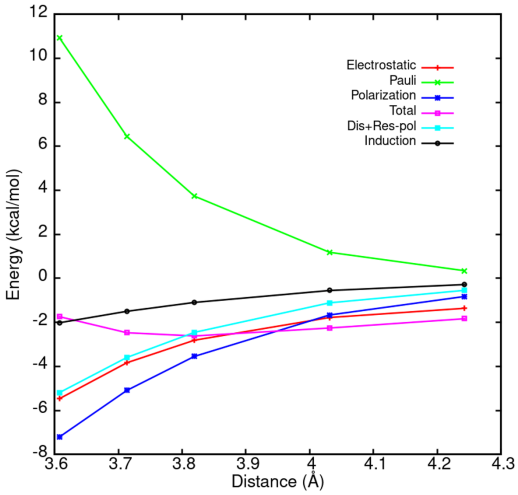
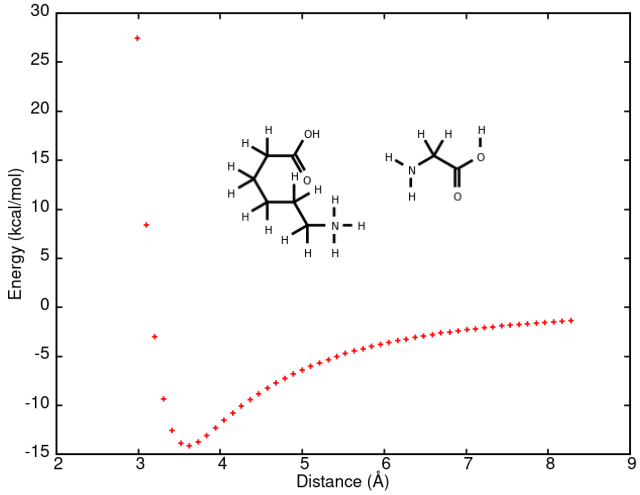


Figure 8: System 8



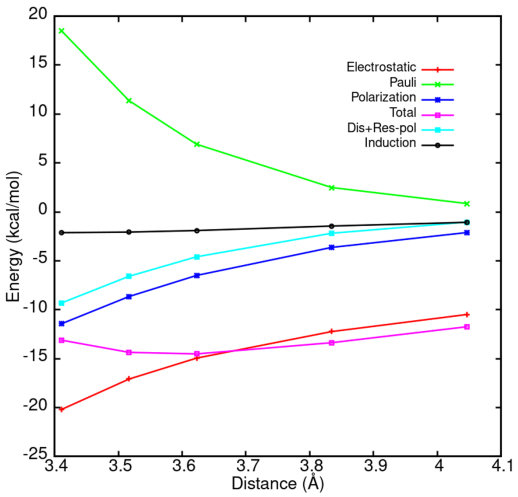


Figure 9: System 9

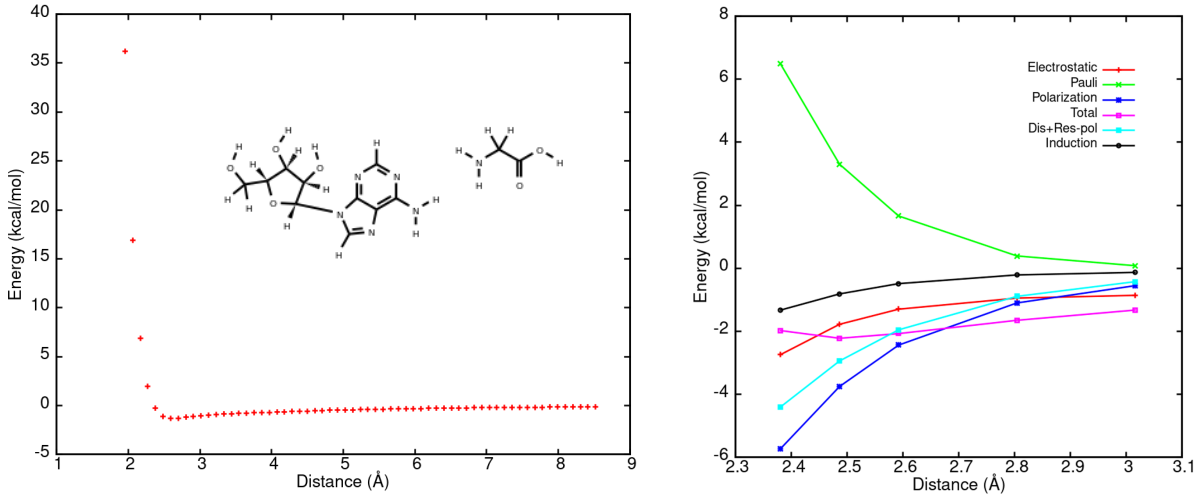


Figure 10: System 10

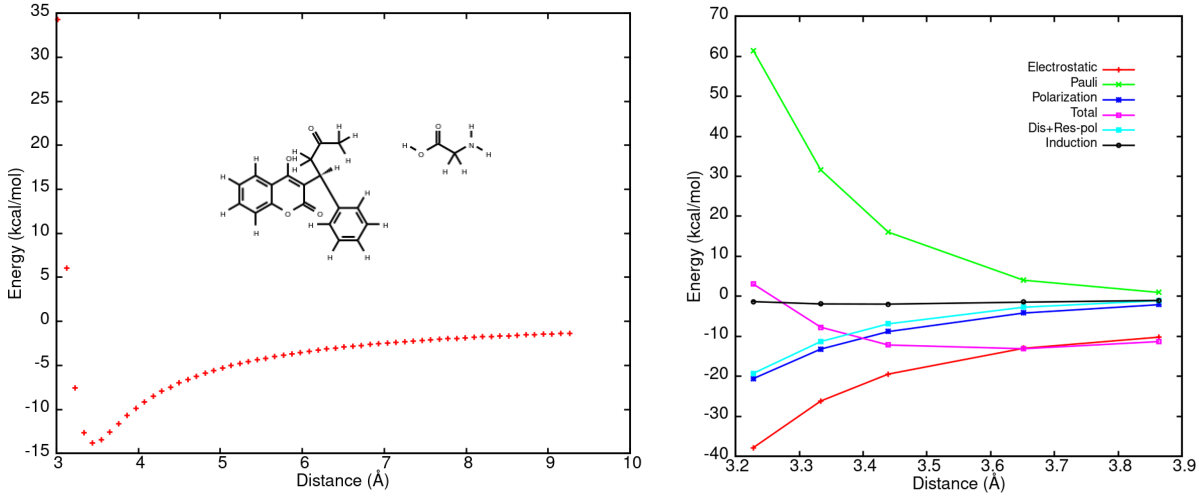


Figure 11: System 11

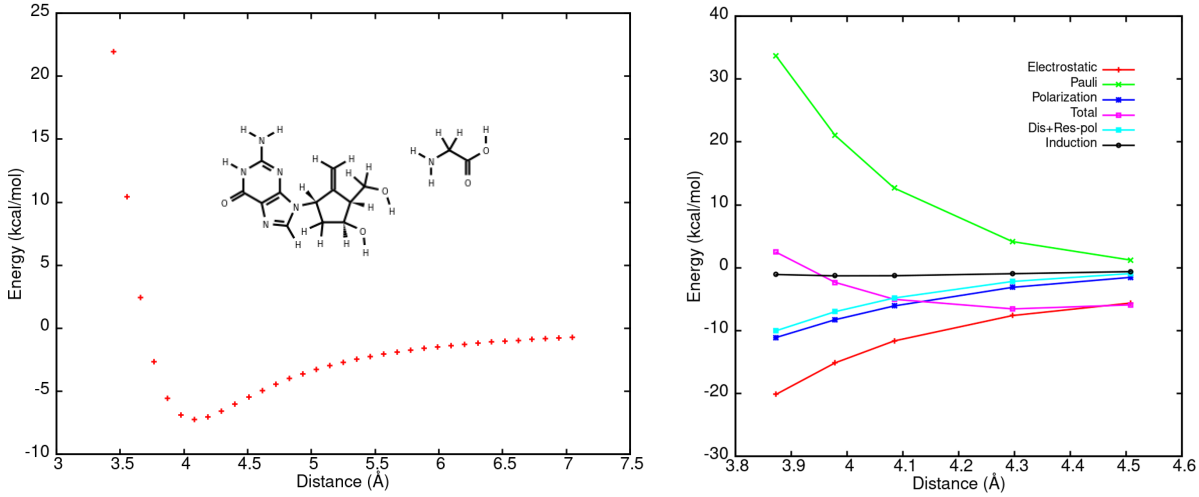
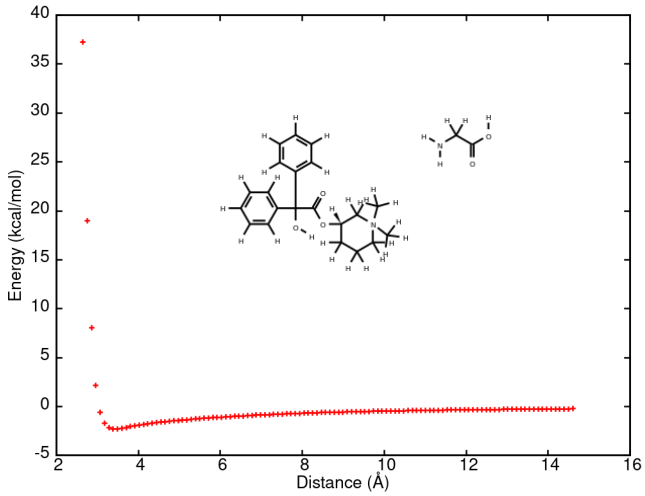


Figure 12: System 12



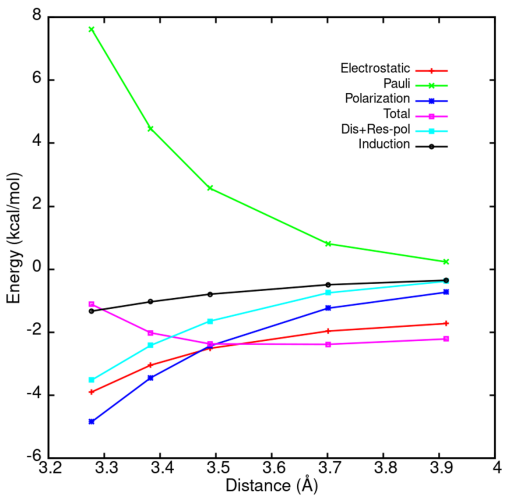
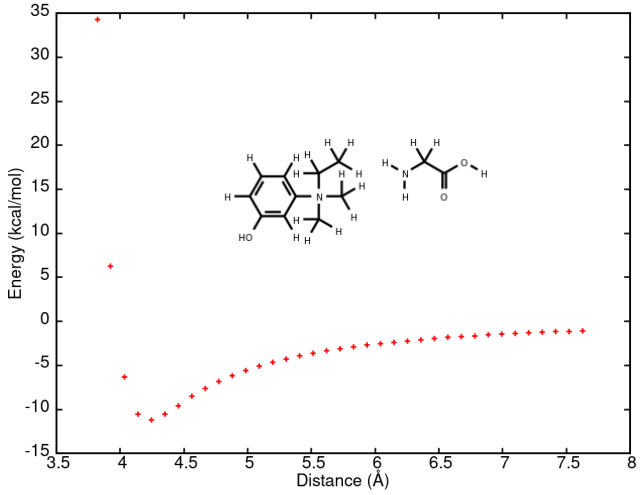


Figure 13: System 13



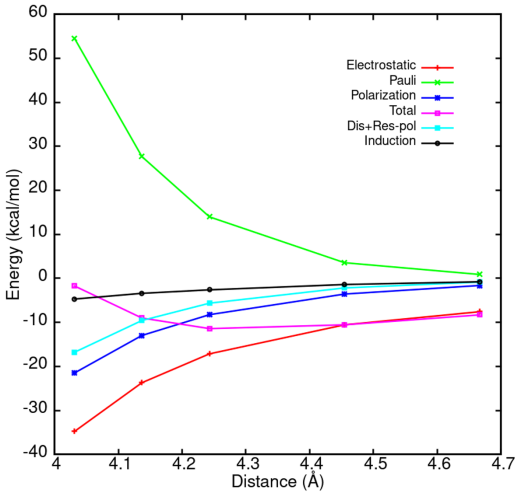
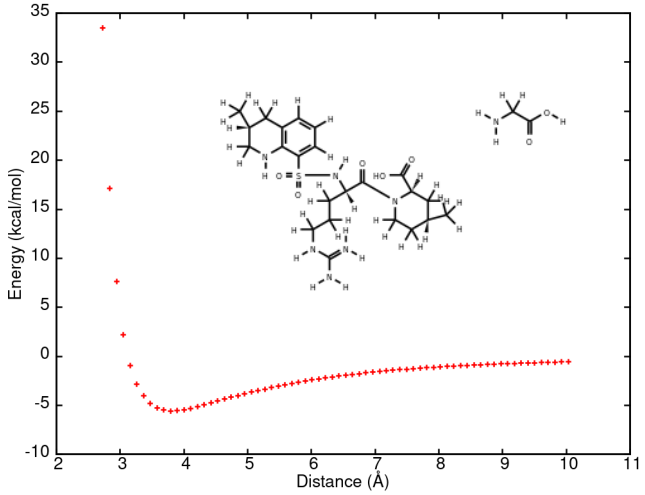


Figure 14: System 14



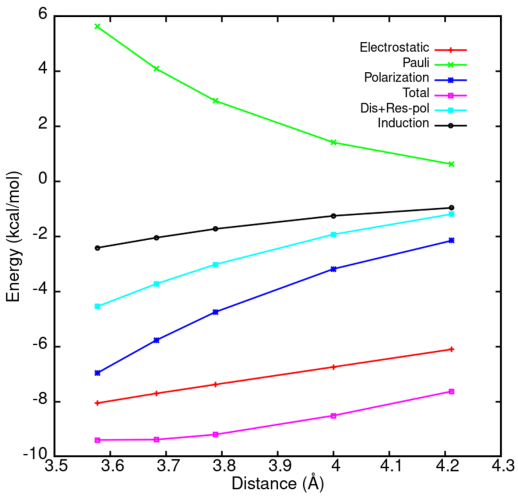


Figure 15: System 15

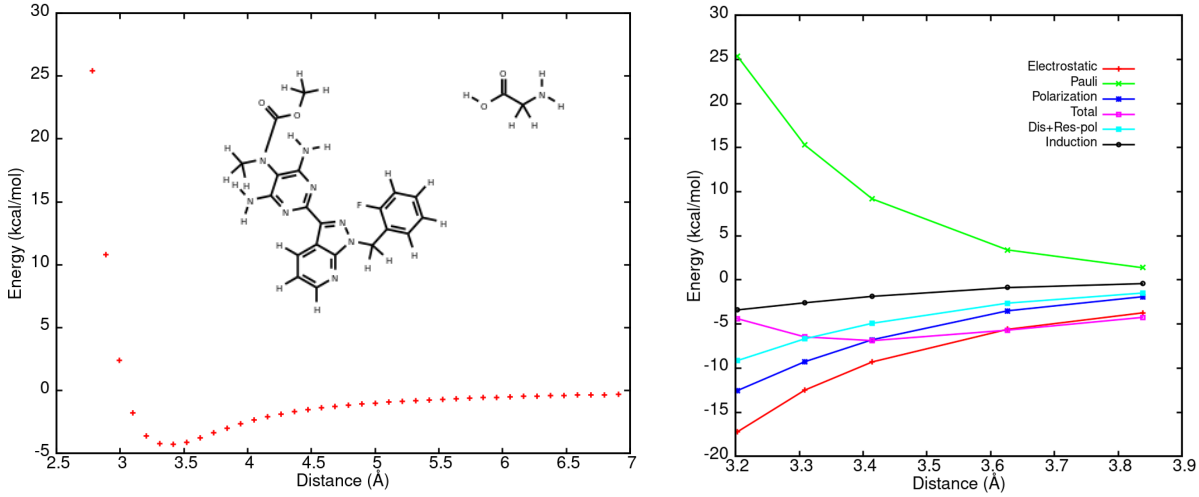


Figure 16: System 16

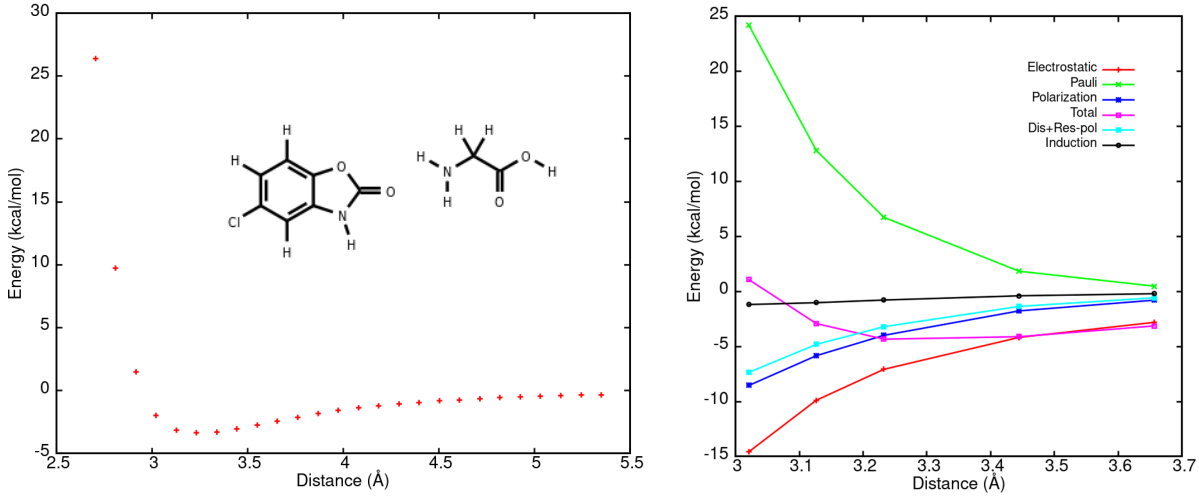


Figure 17: System 17

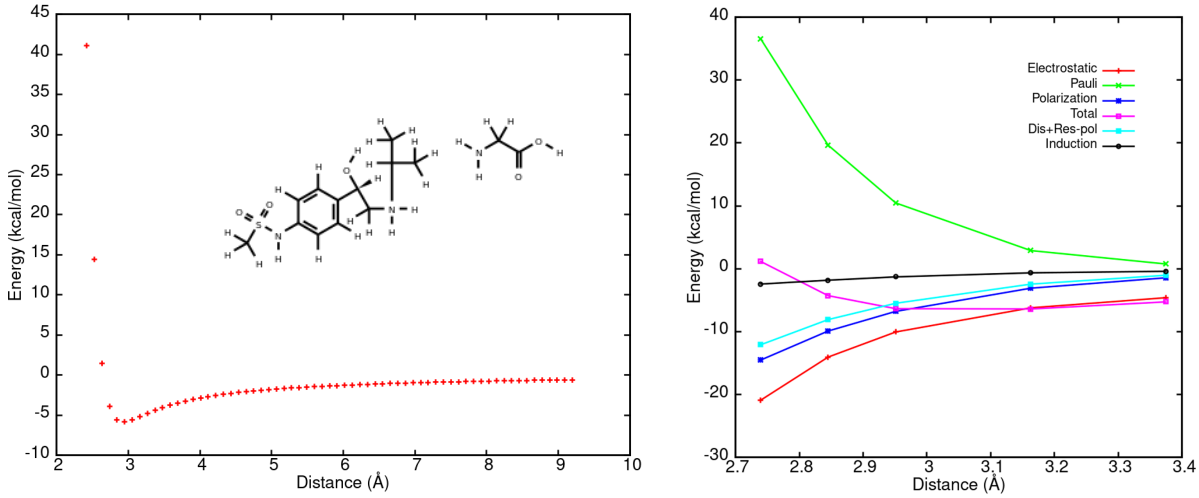


Figure 18: System 18

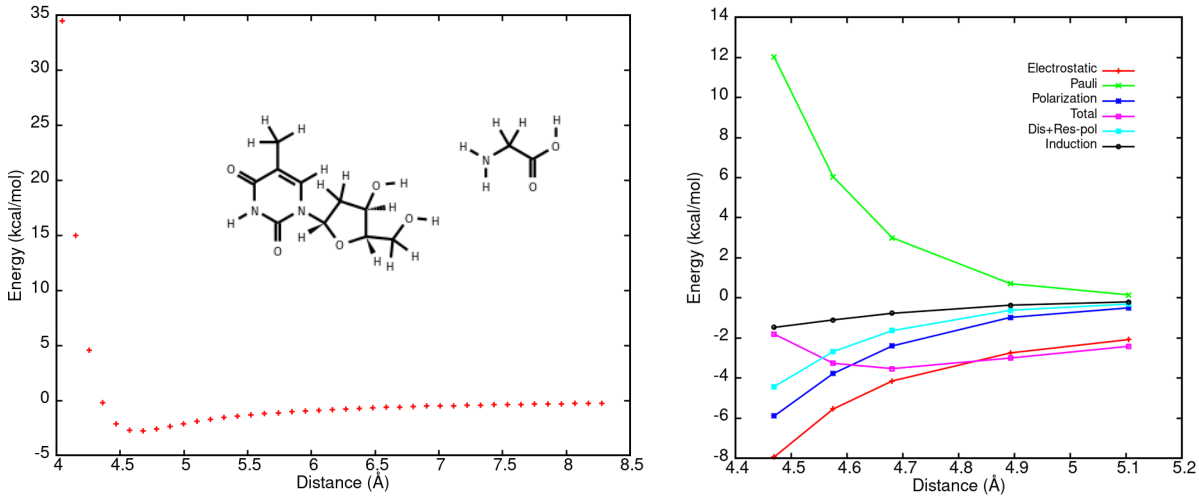


Figure 19: System 19

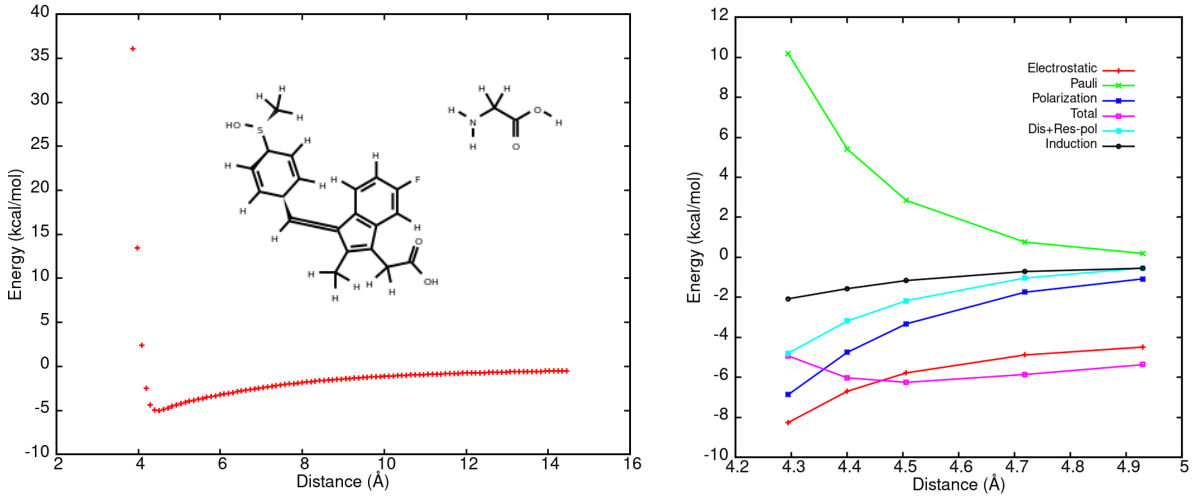


Figure 20: System 20

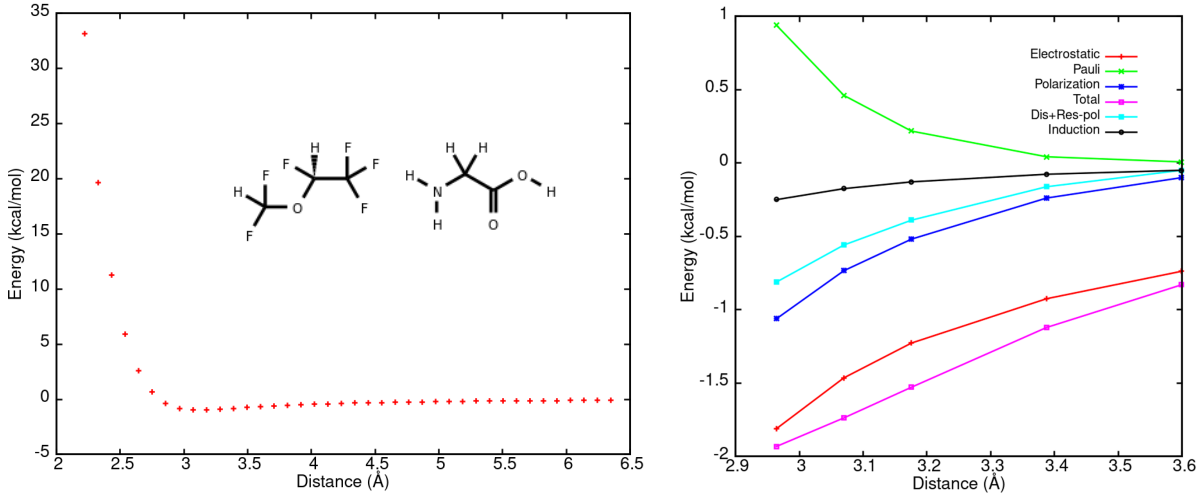
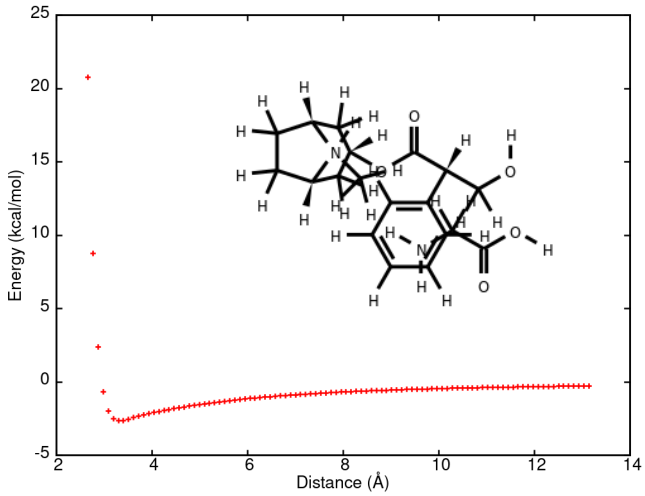


Figure 21: System 21



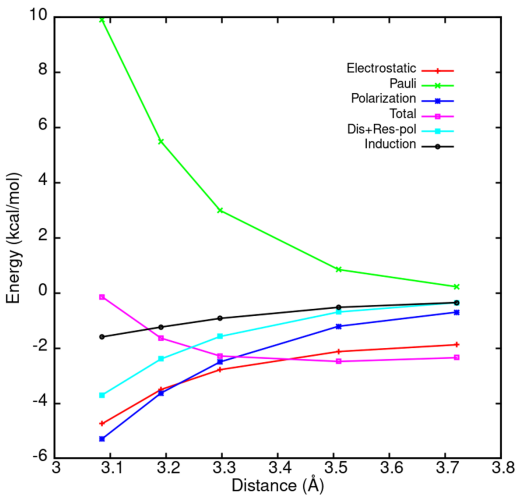
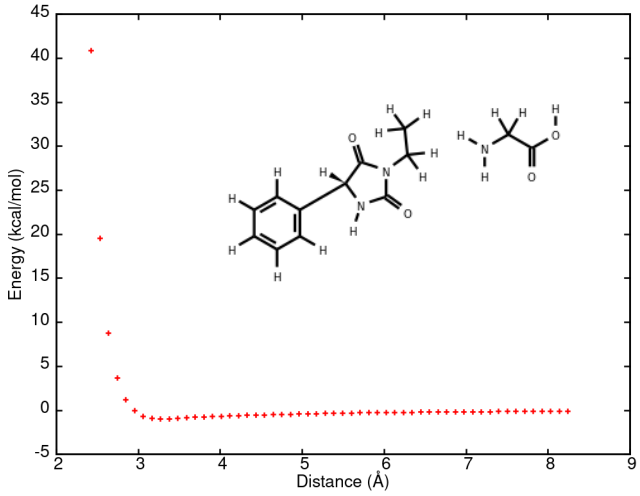


Figure 22: System 22



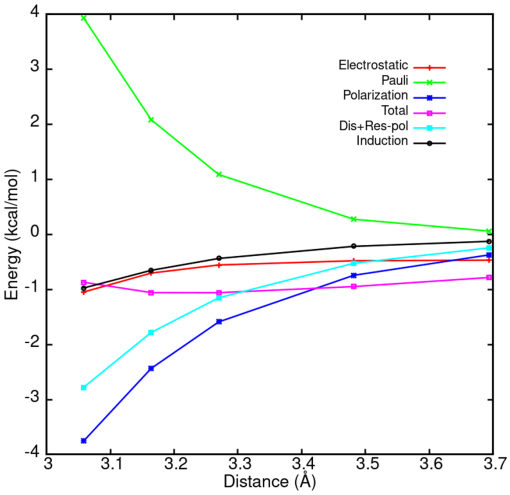
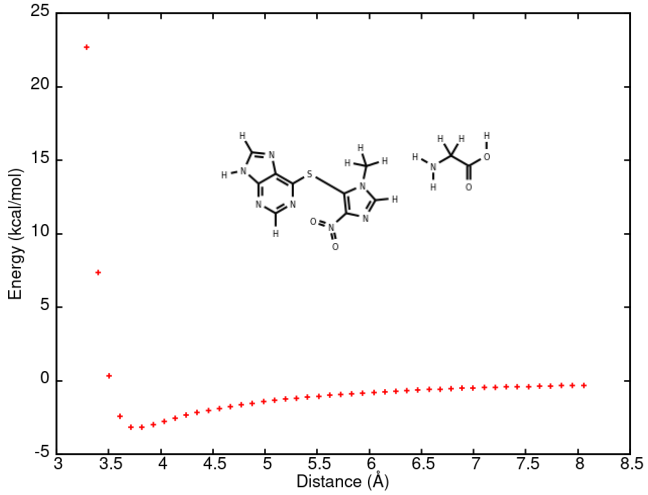


Figure 23: System 23



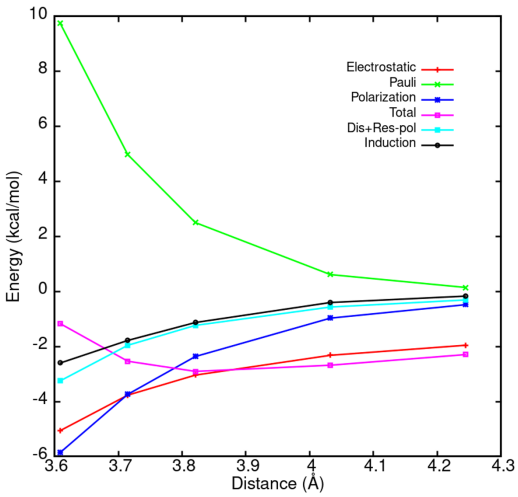


Figure 24: System 24

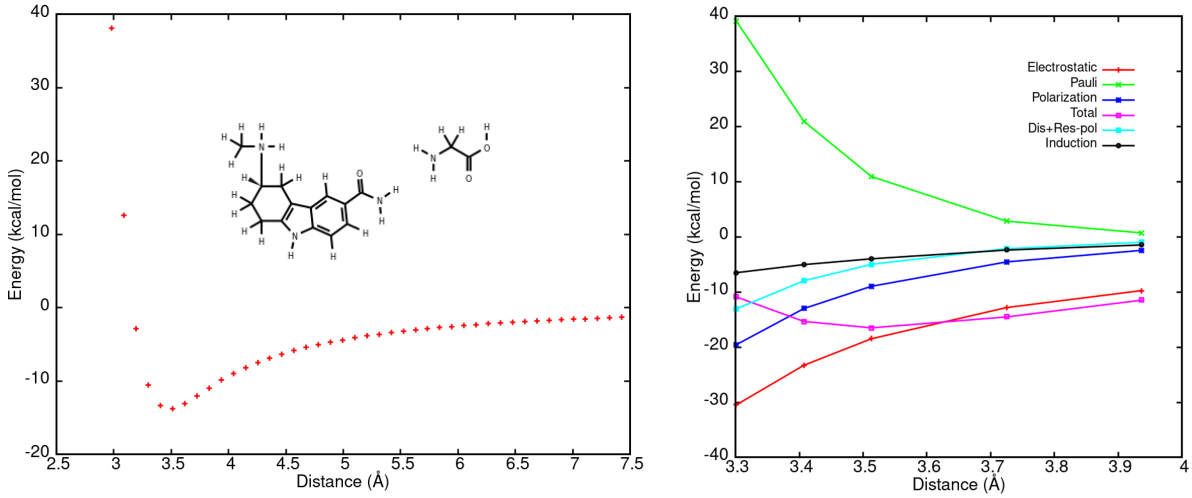


Figure 25: System 25

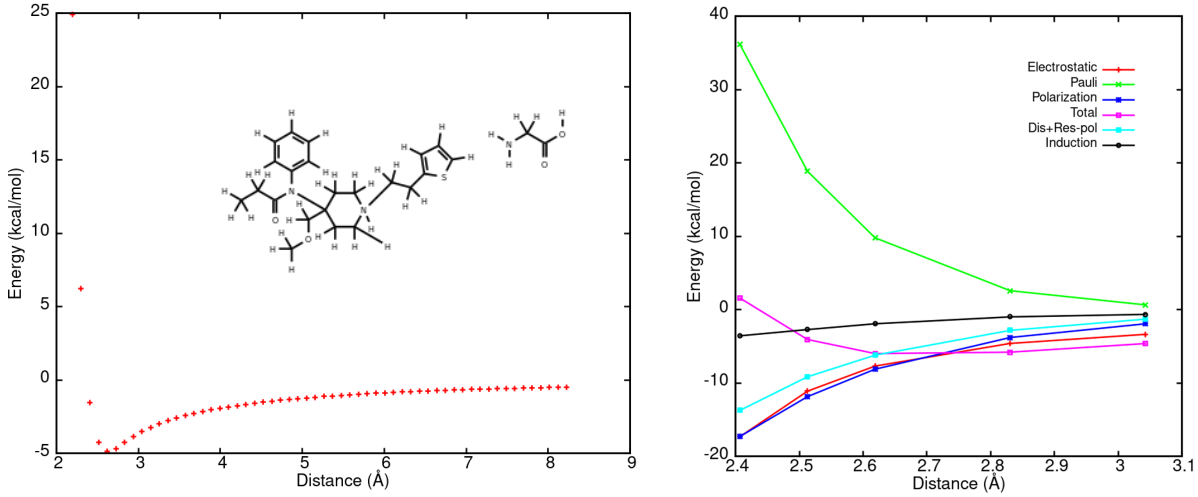


Figure 26: System 26

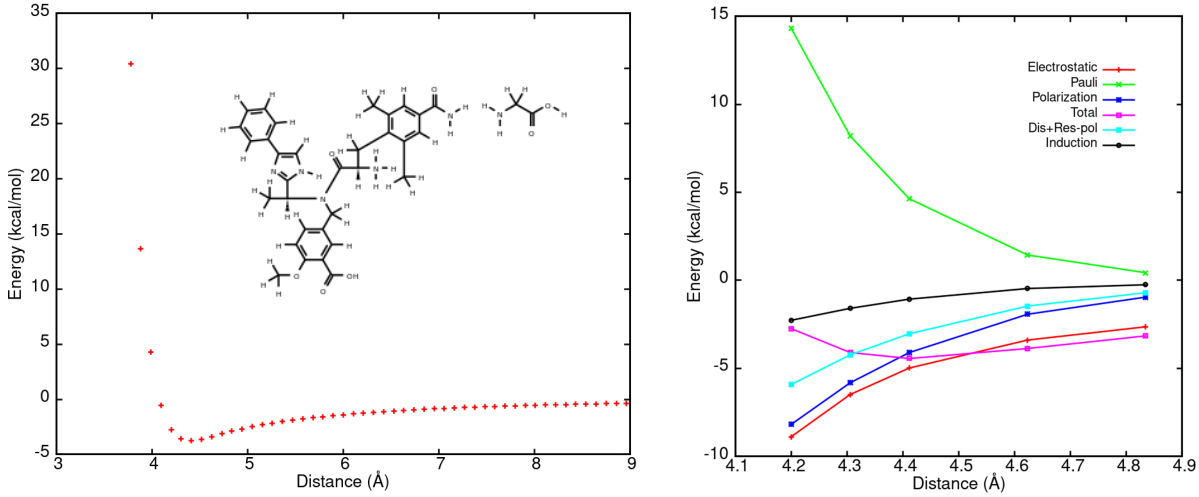
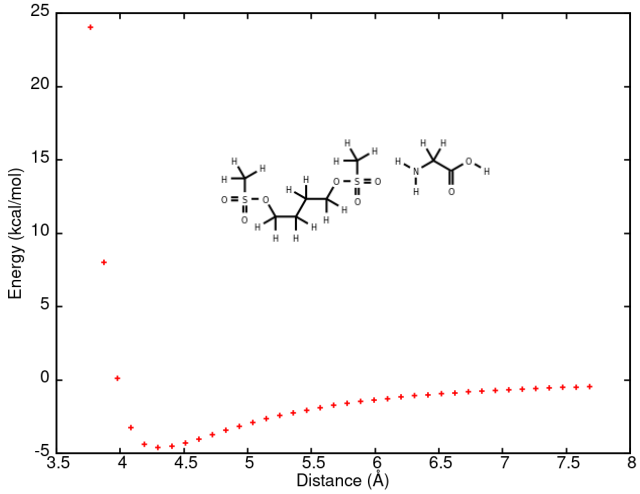


Figure 27: System 27



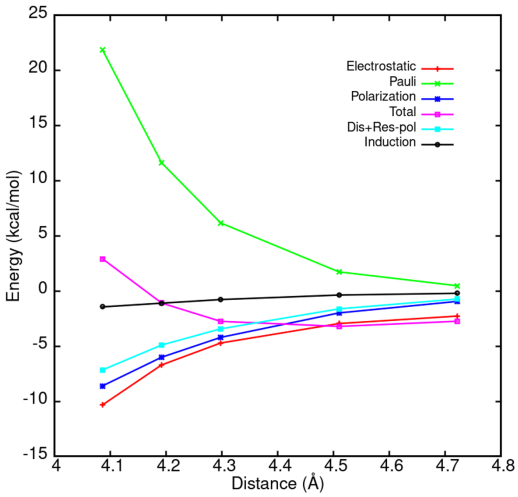
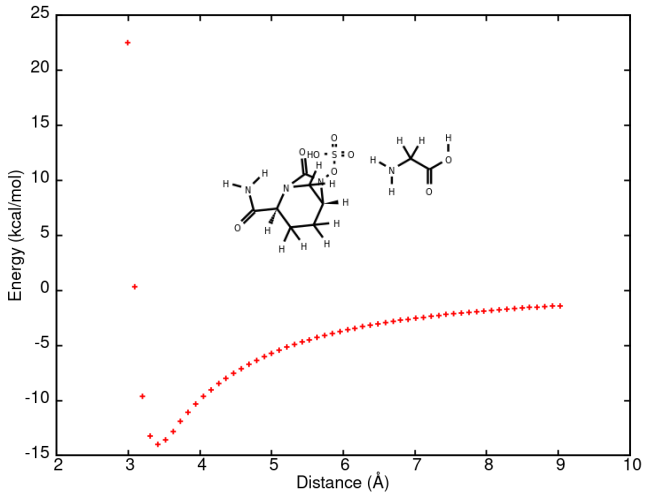


Figure 28: System 28



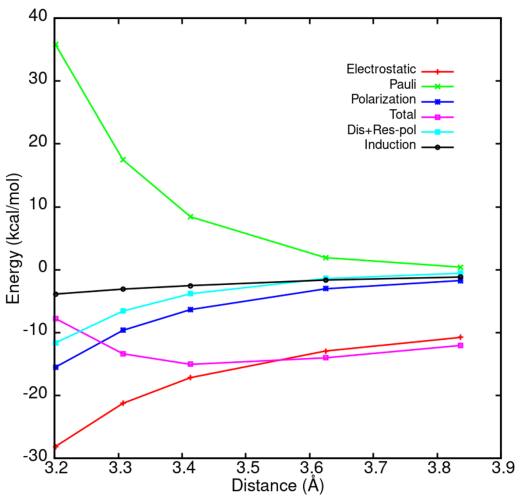
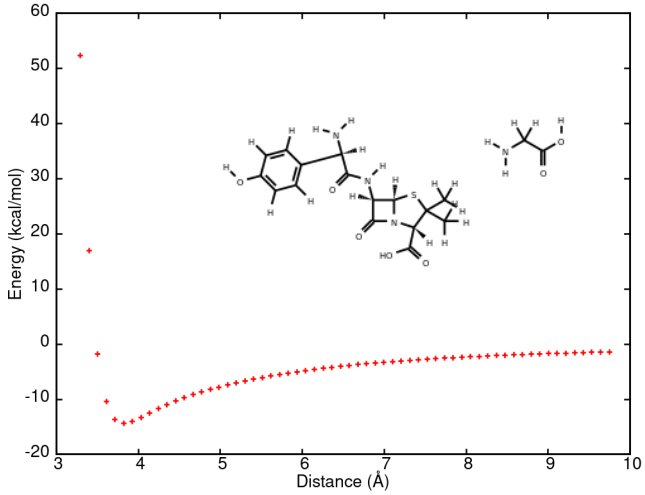


Figure 29: System 29



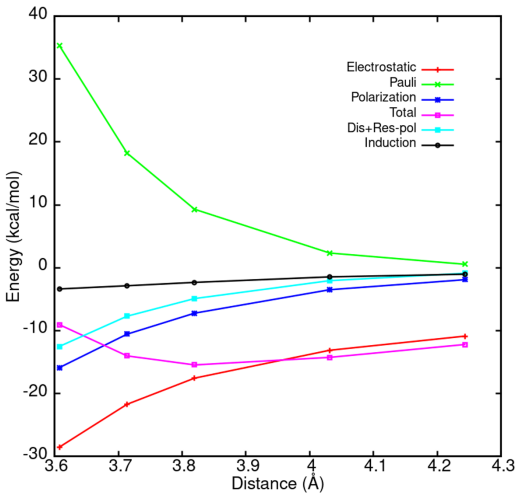
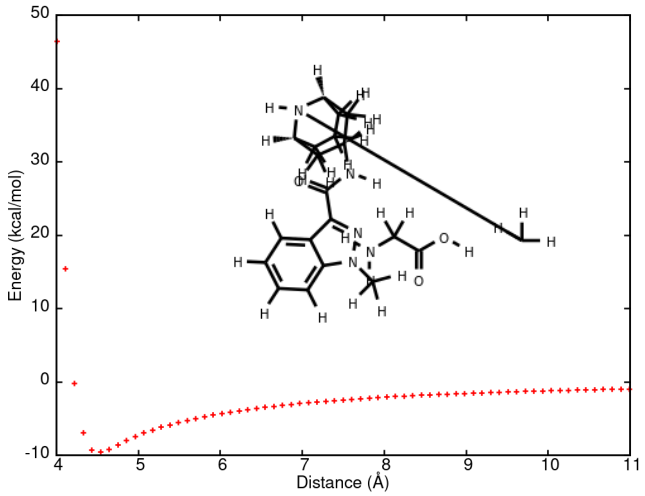


Figure 30: System 30



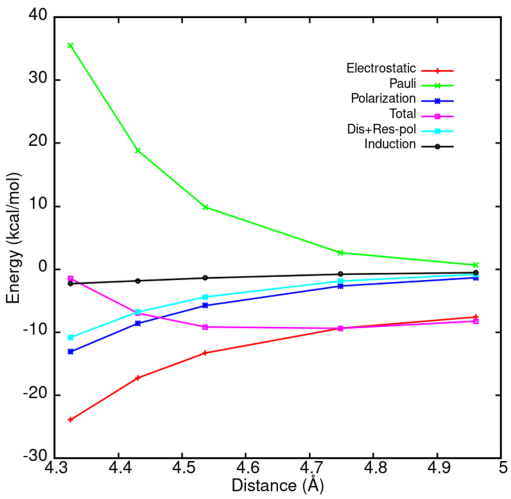
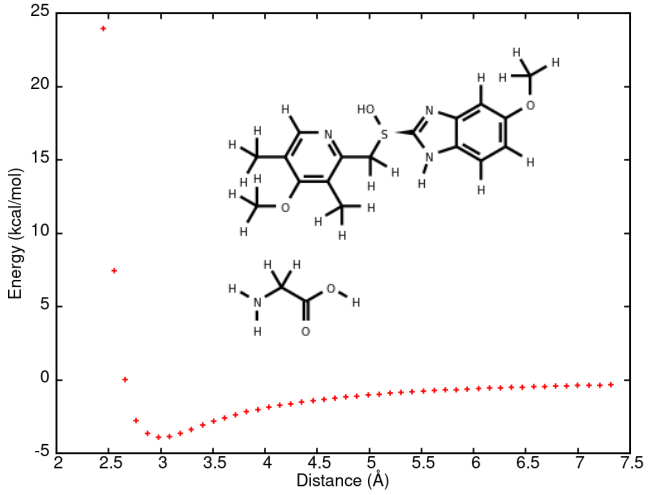


Figure 31: System 31



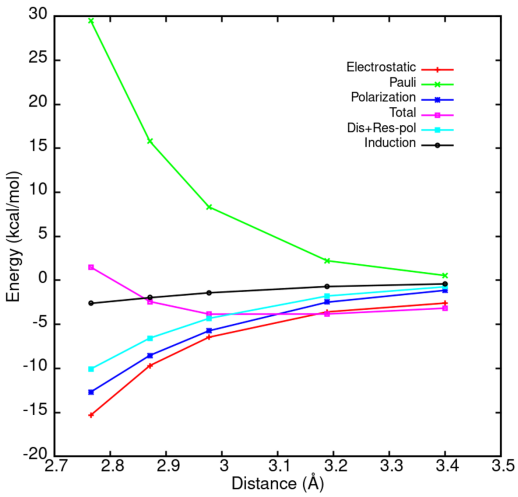
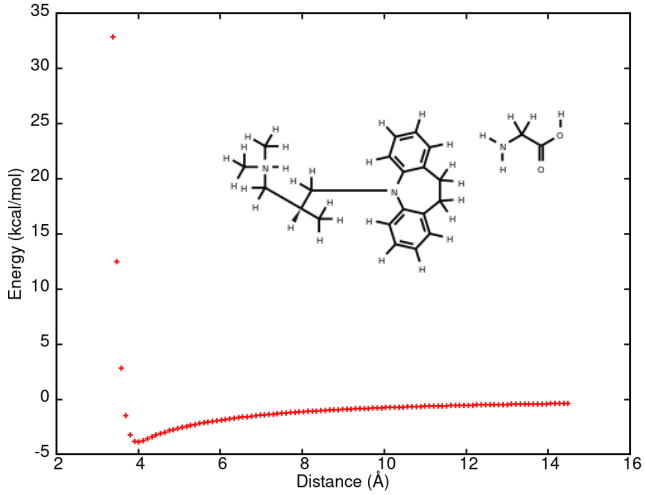


Figure 32: System 32



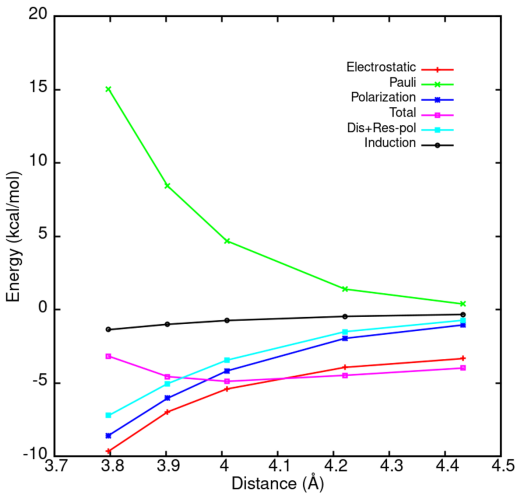


Figure 33: System 33

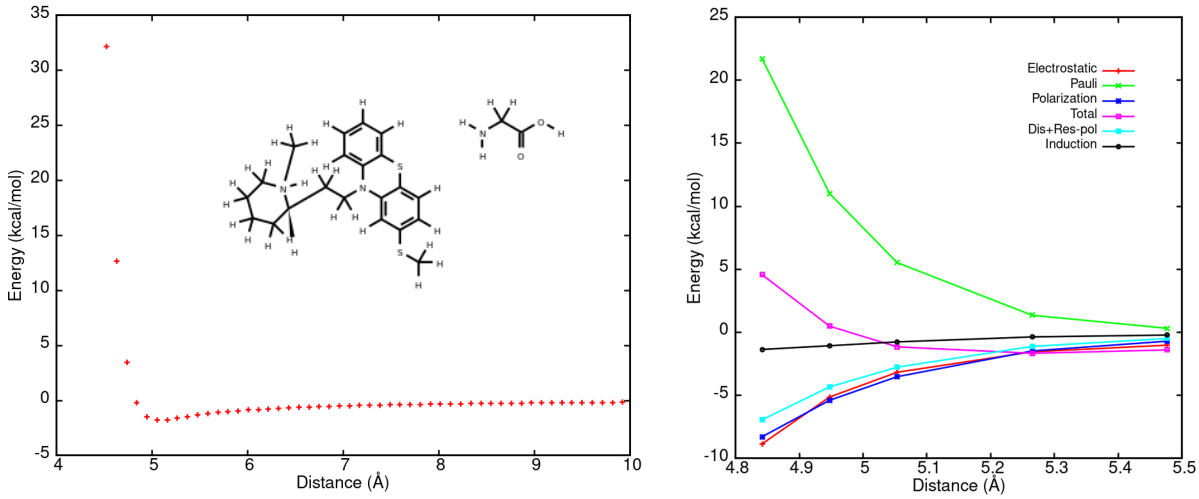


Figure 34: System 34

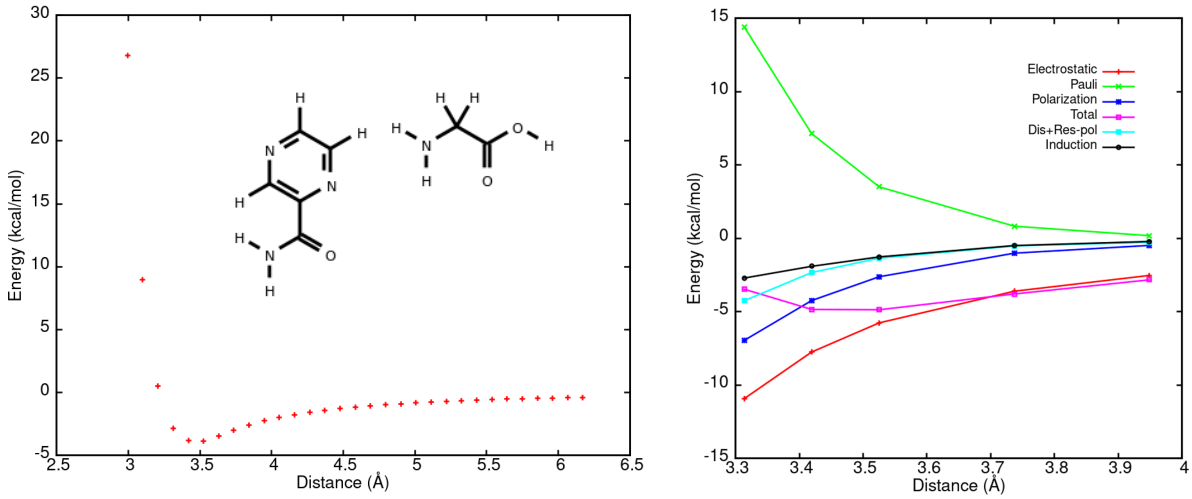


Figure 35: System 35

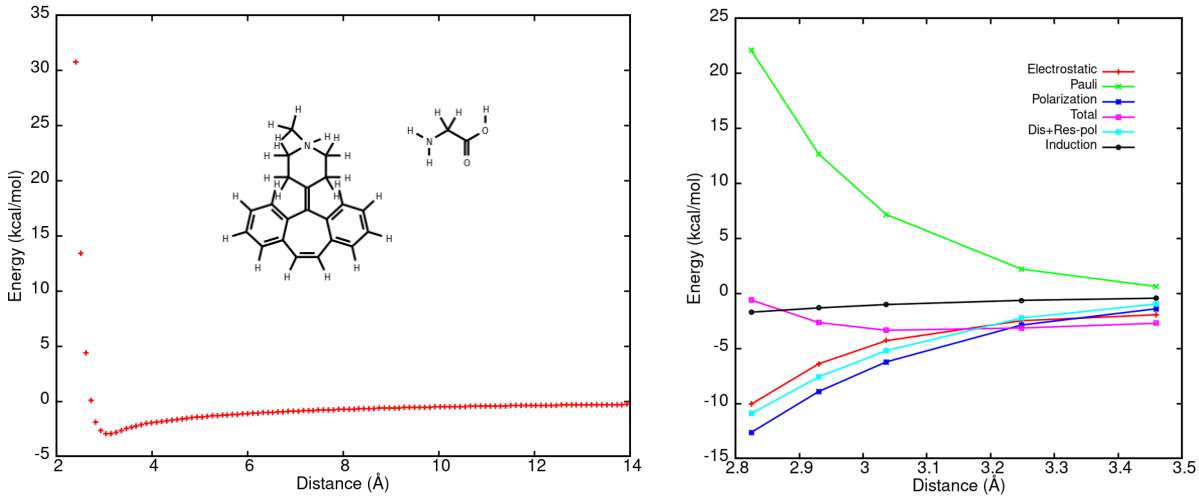


Figure 36: System 36

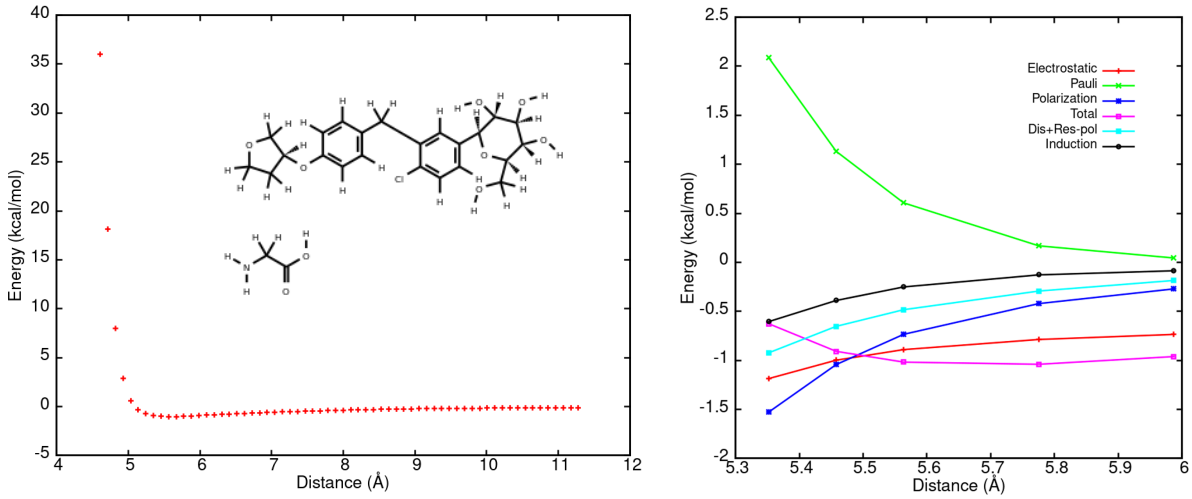


Figure 37: System 37

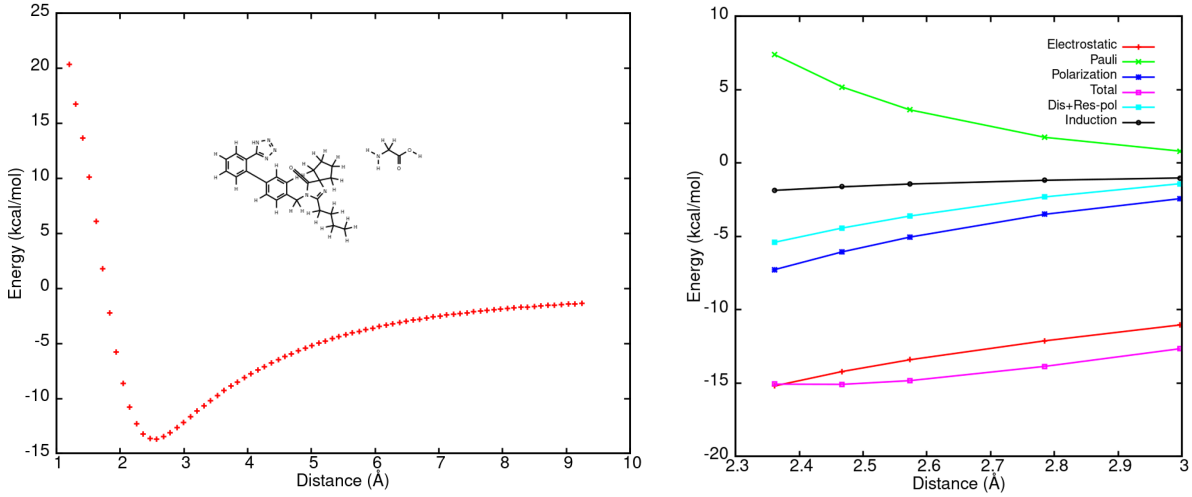


Figure 38: System 38

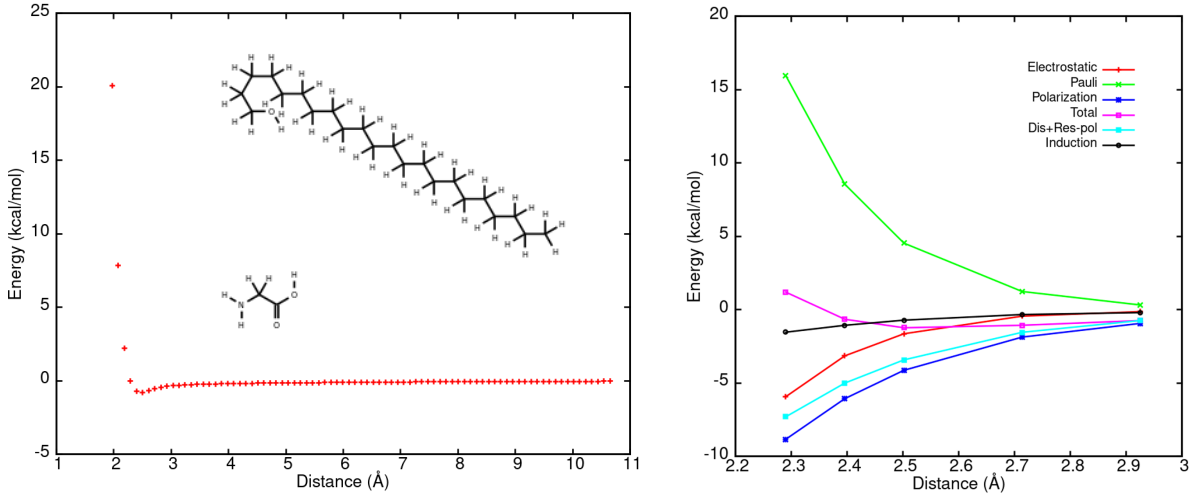
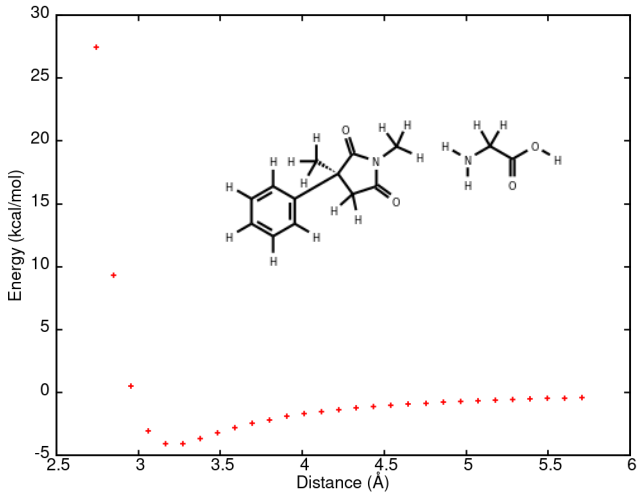


Figure 39: System 39



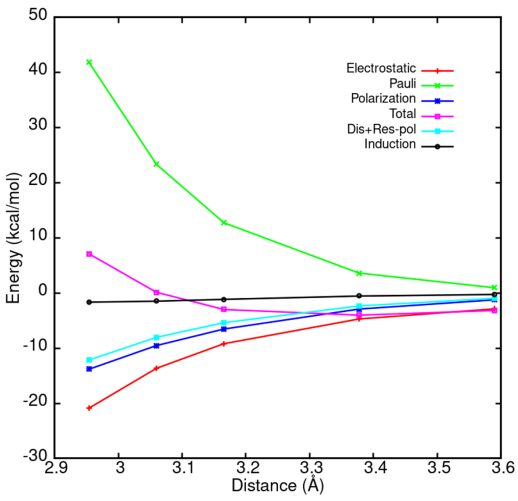
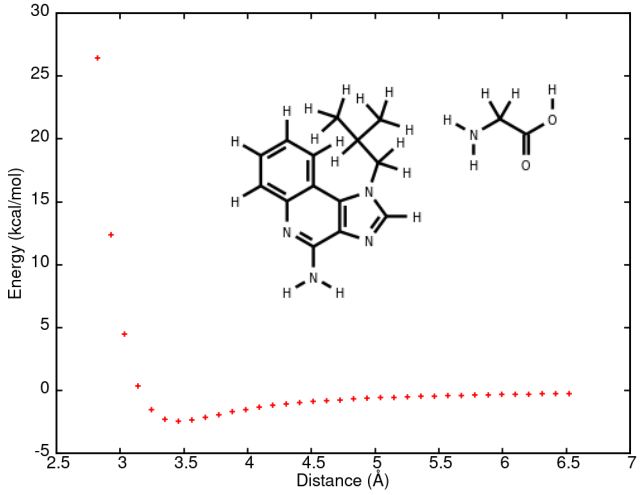


Figure 40: System 40



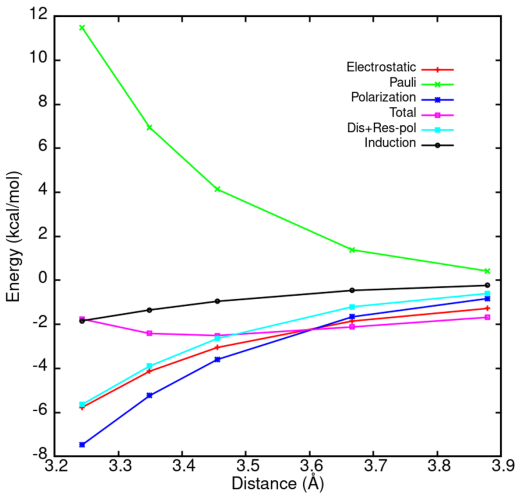
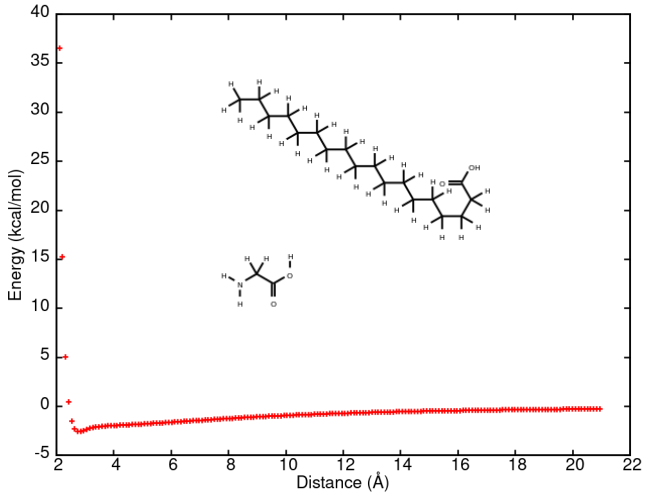


Figure 41: System 41



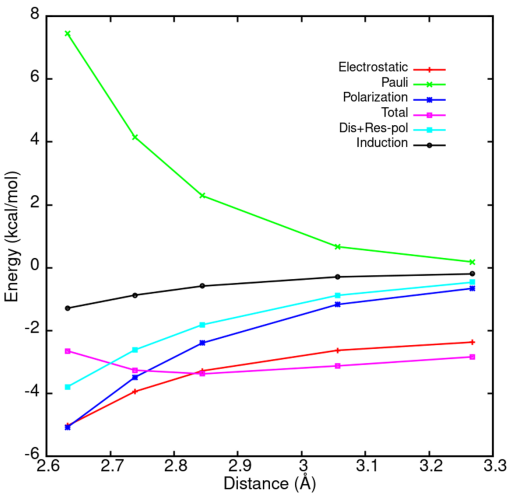


Figure 42: System 42

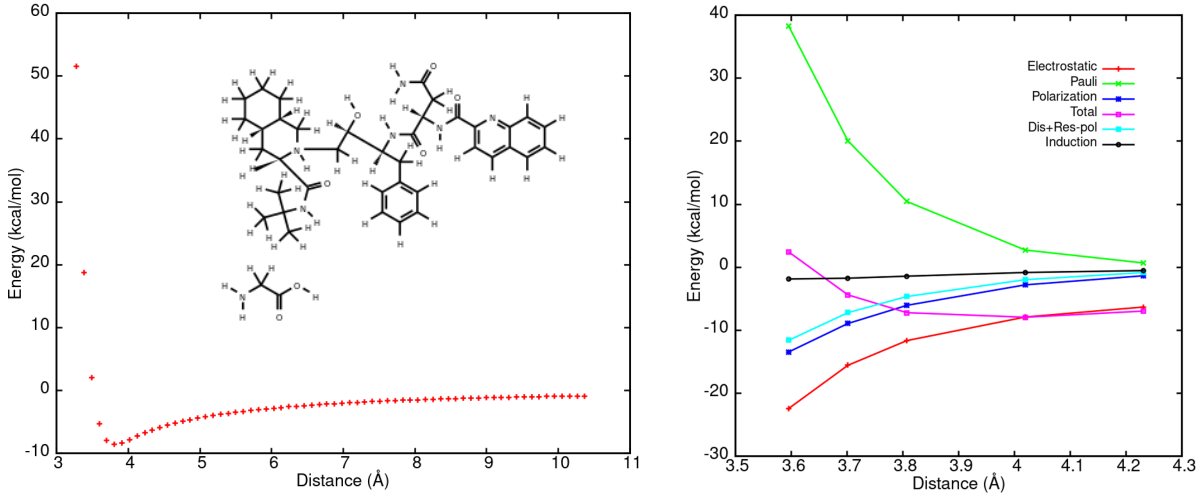


Figure 43: System 43

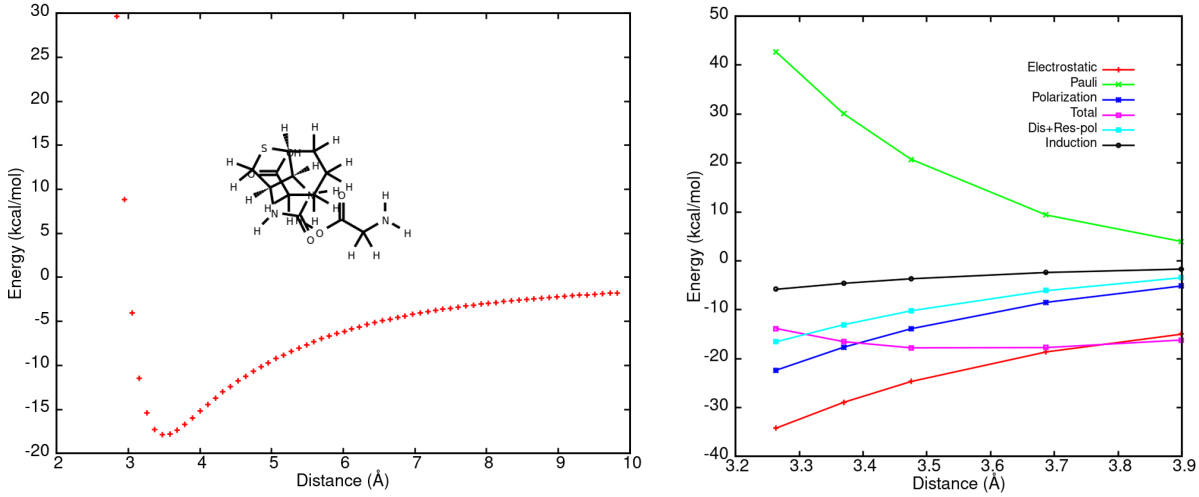


Figure 44: System 44

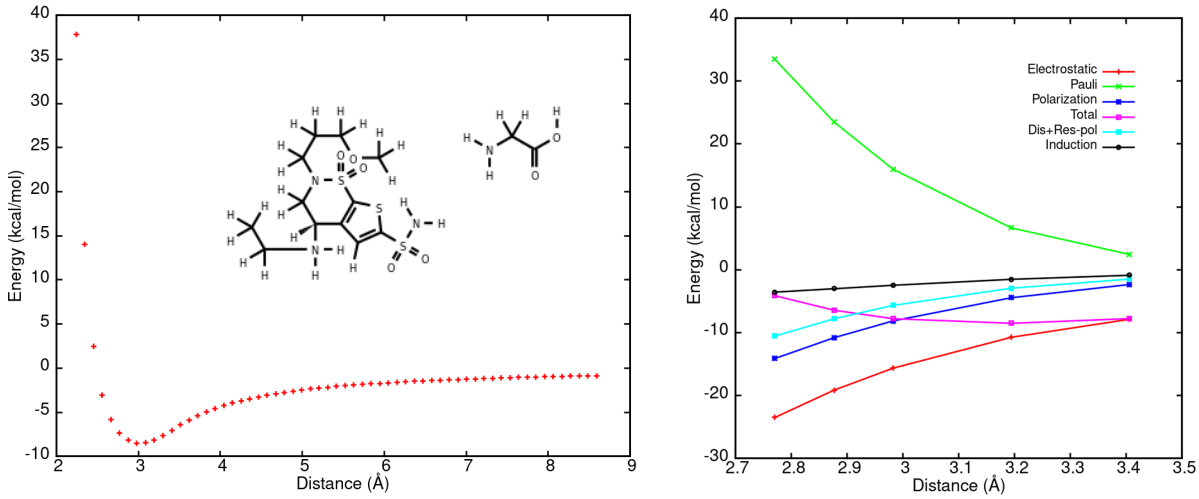


Figure 45: System 45

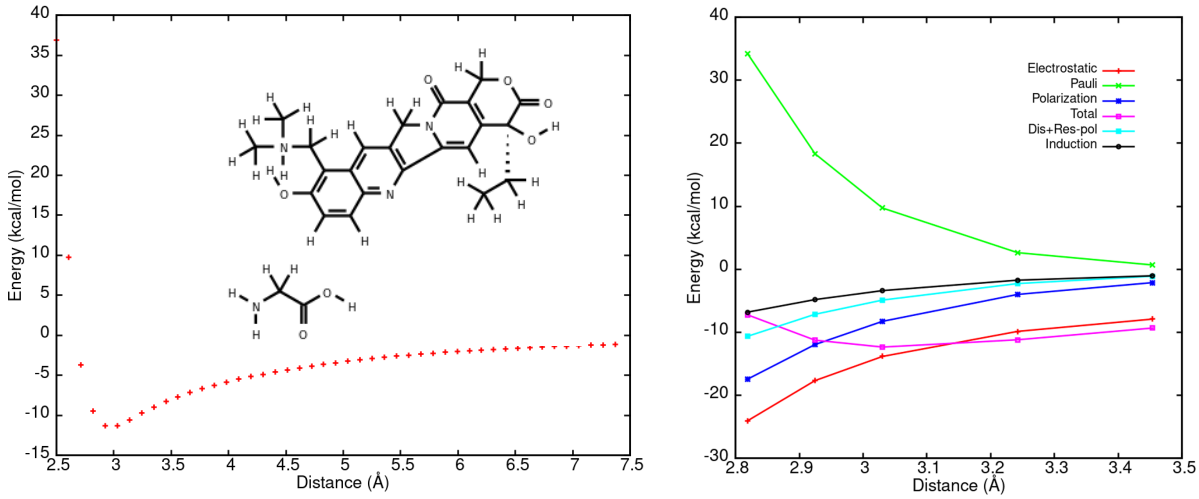


Figure 46: System 46

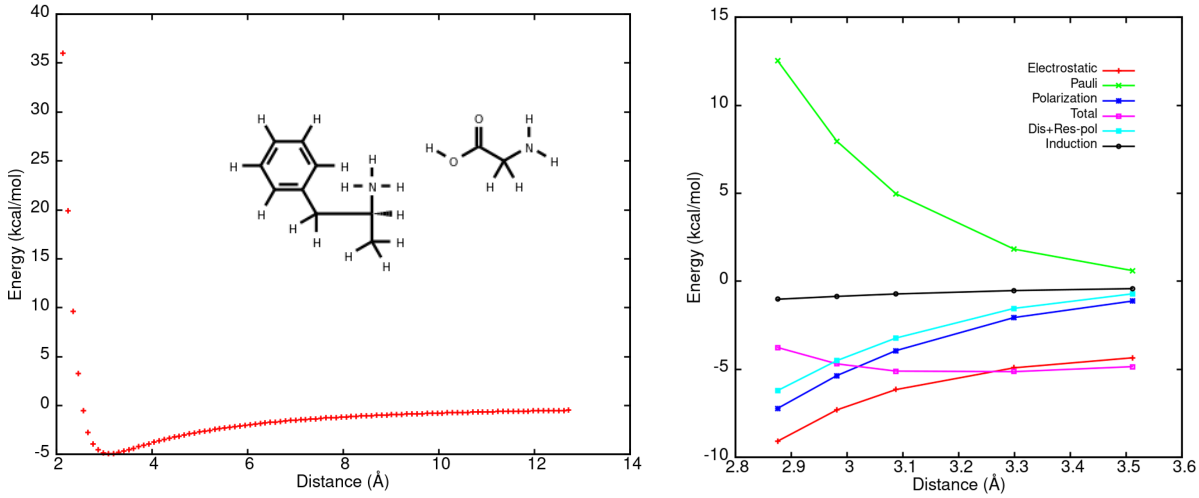


Figure 47: System 47

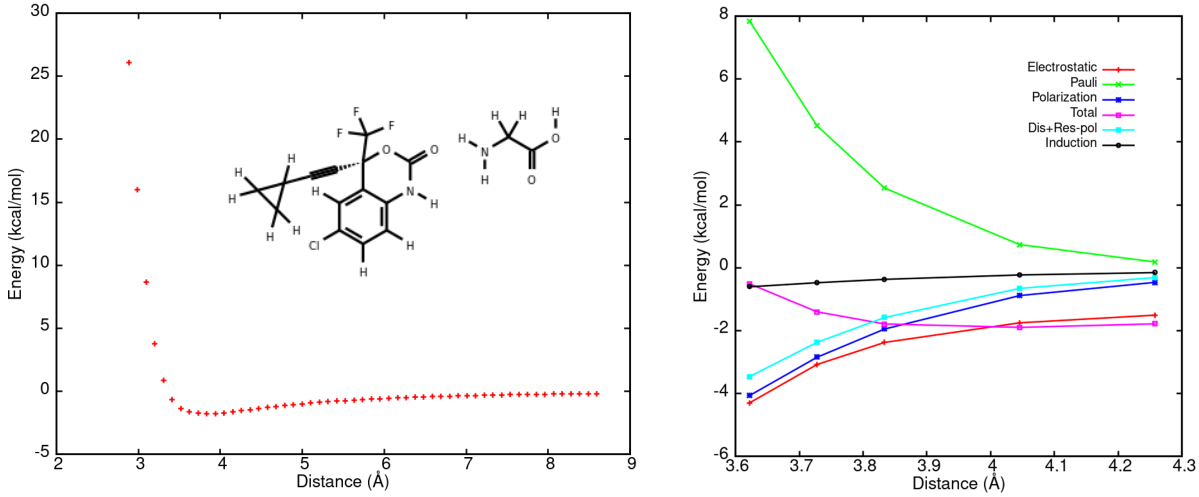
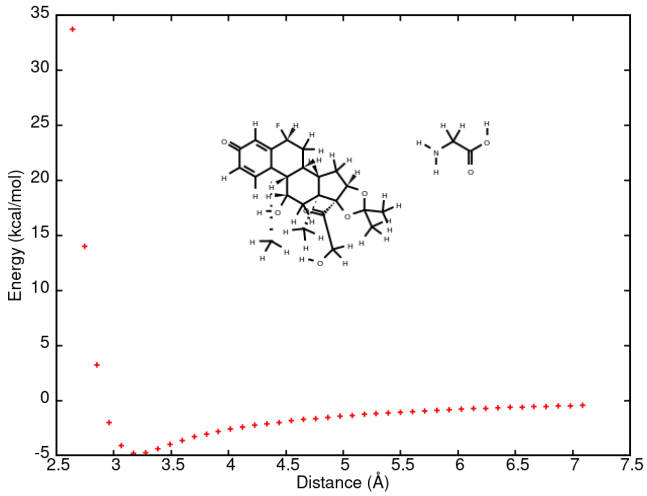


Figure 48: System 48



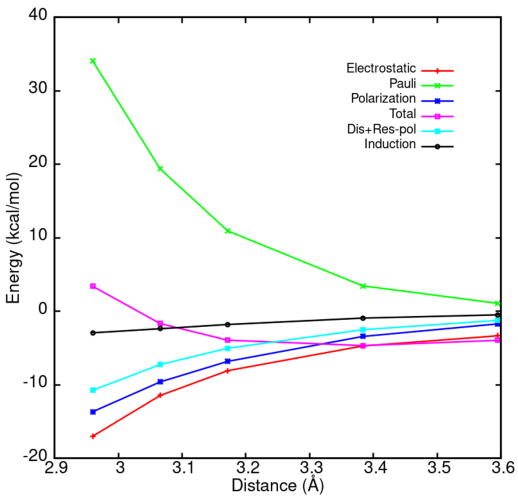
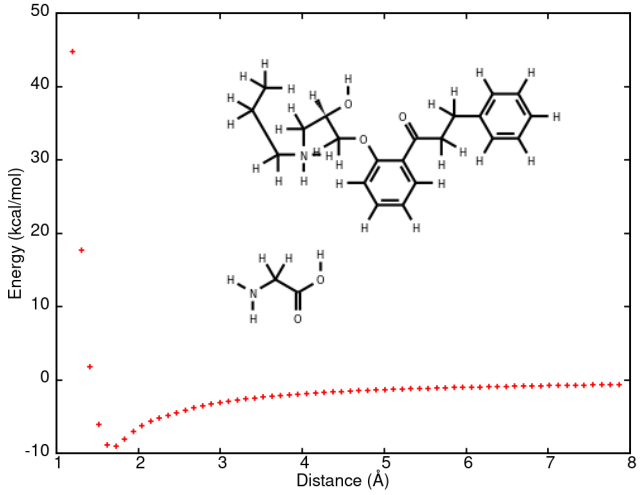


Figure 49: System 49



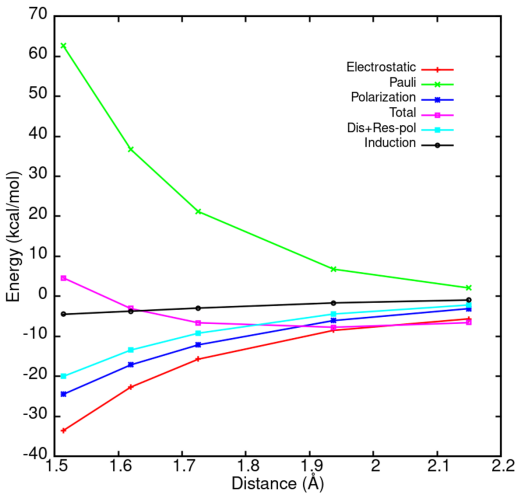
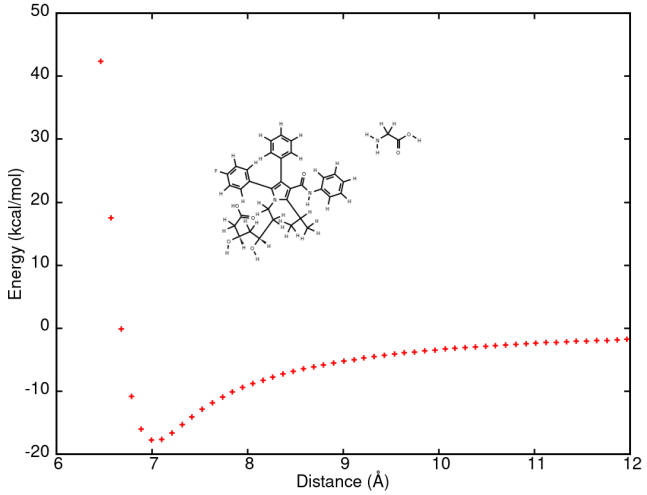


Figure 50: System 50



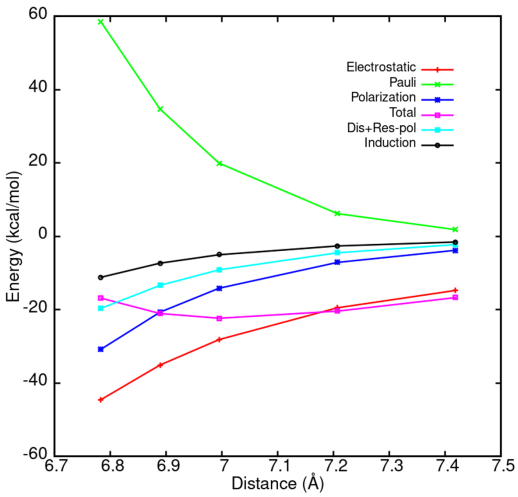


Figure 51: System 51

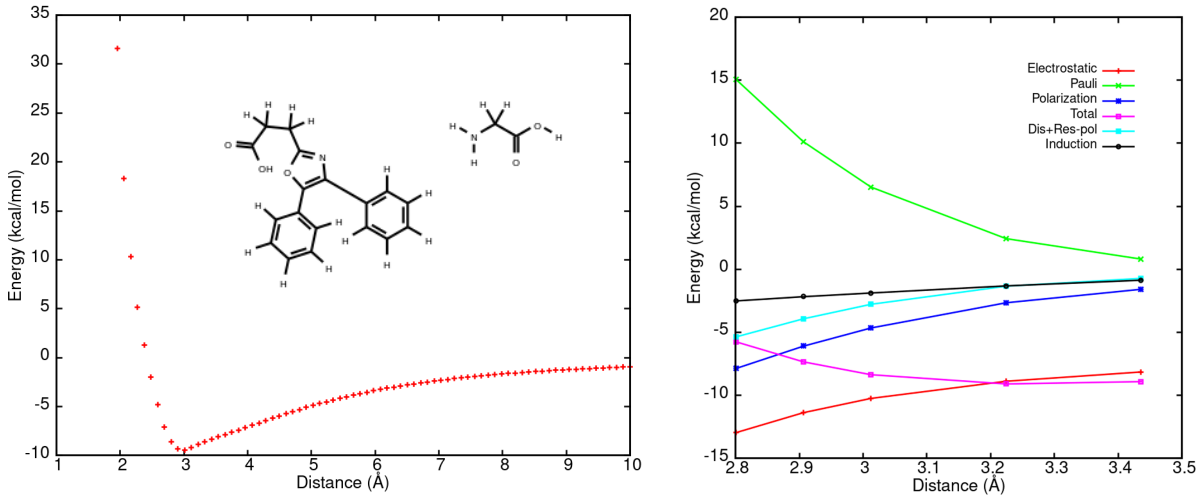


Figure 52: System 52

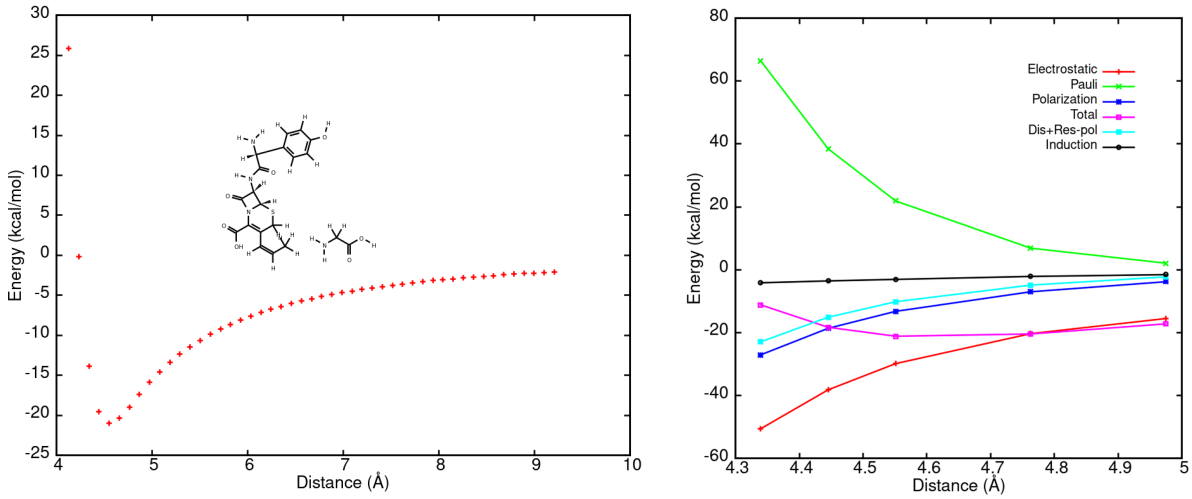


Figure 53: System 53

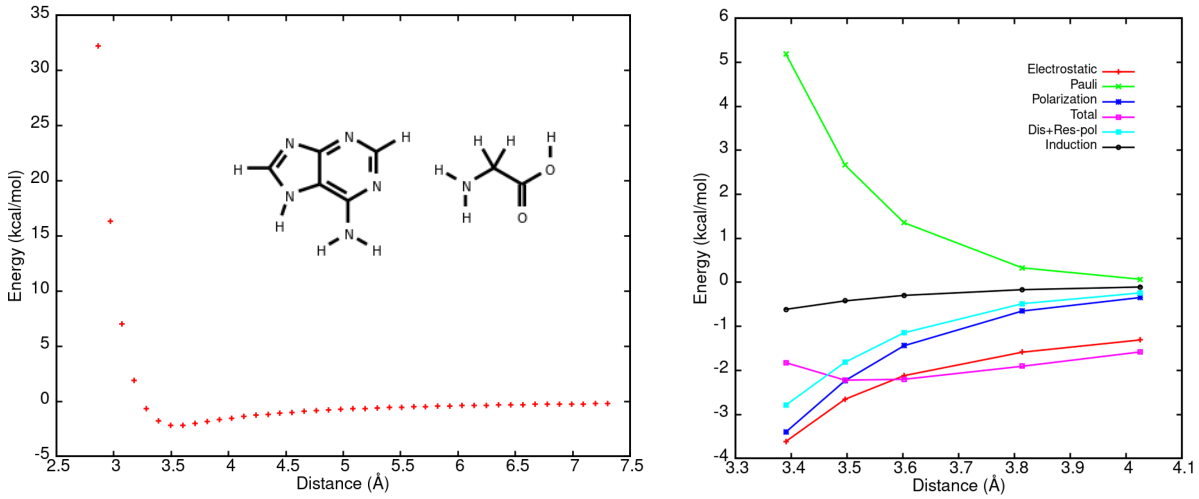


Figure 54: System 54

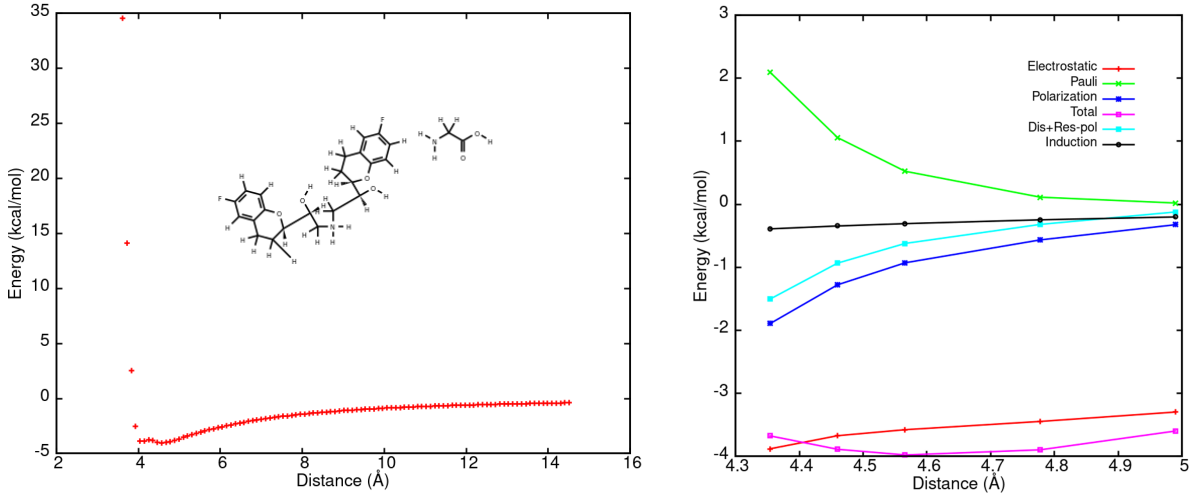


Figure 55: System 55

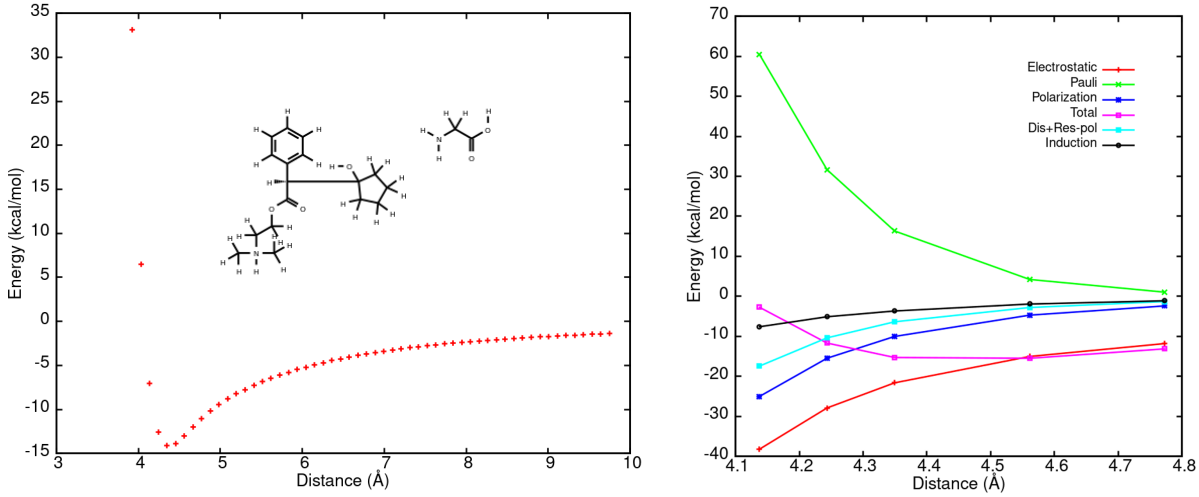


Figure 56: System 56

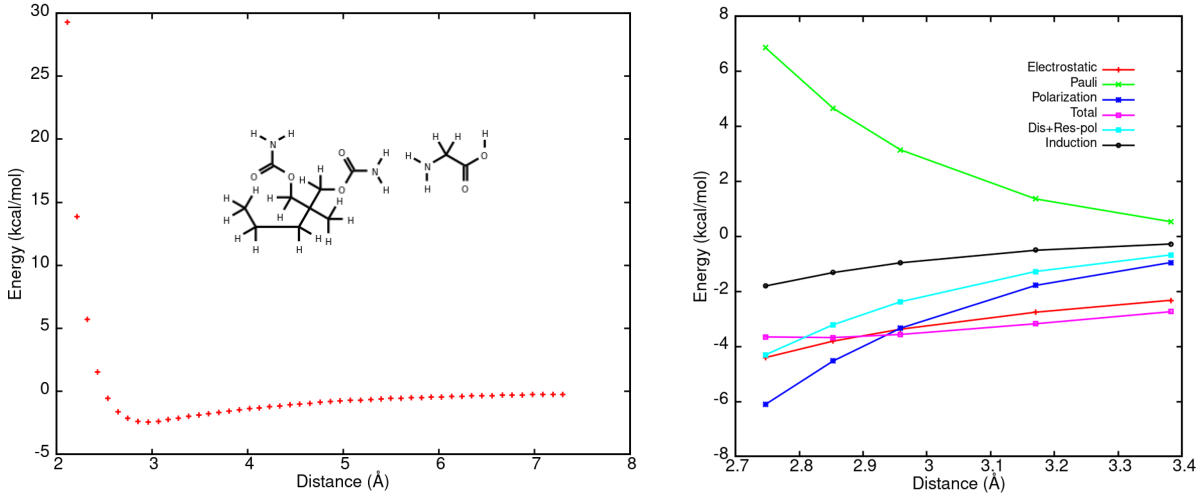


Figure 57: System 57

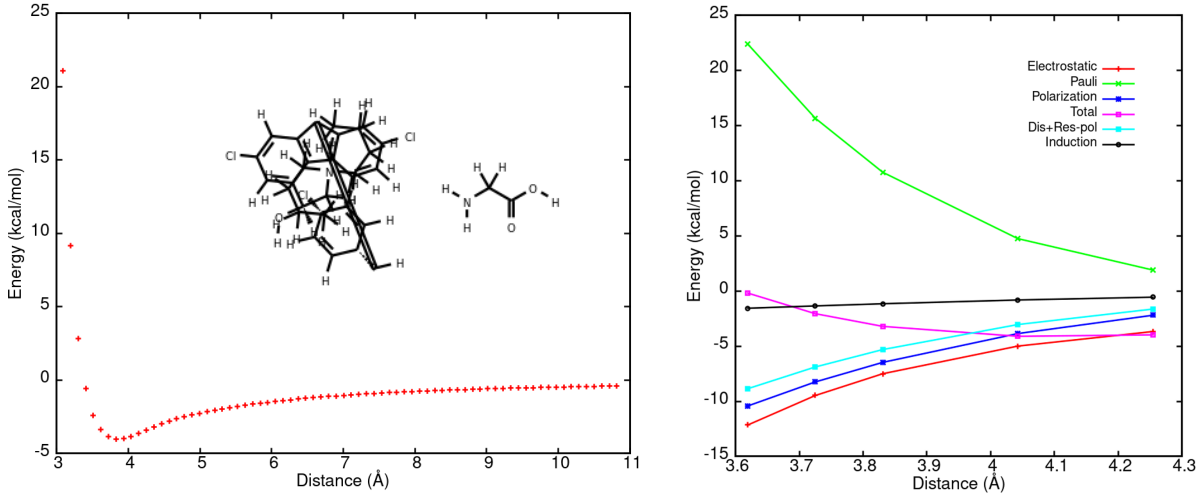


Figure 58: System 58

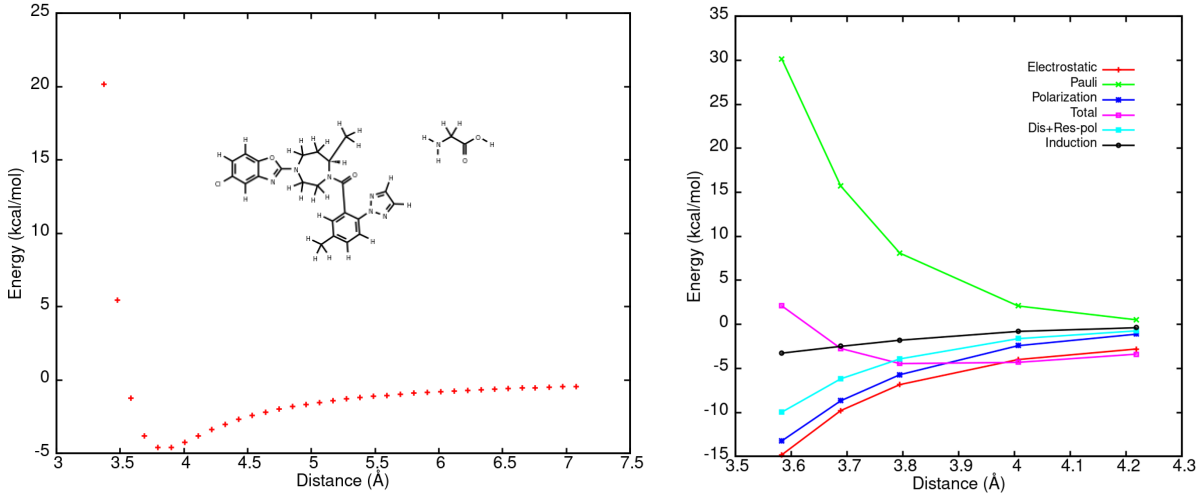


Figure 59: System 59

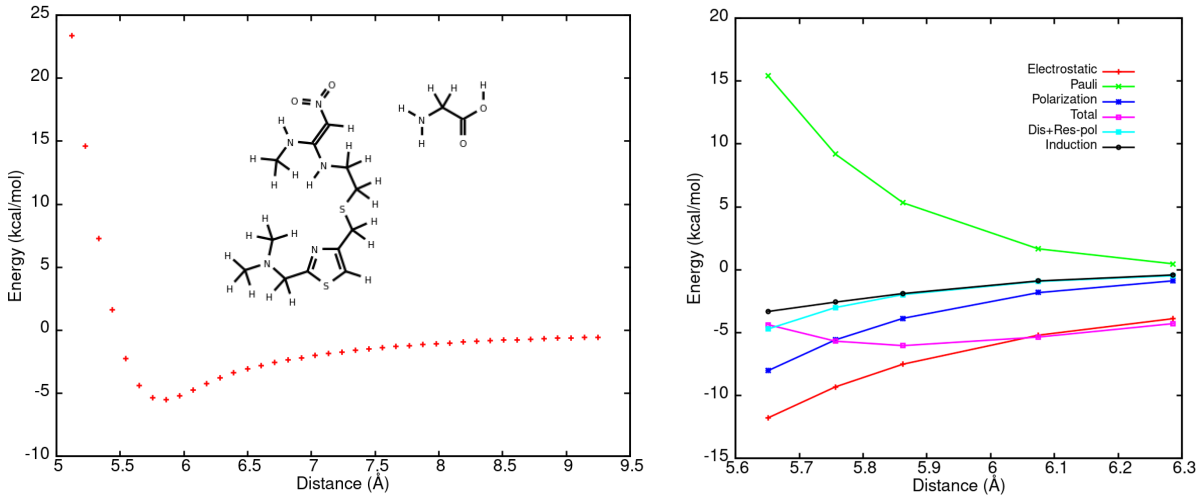
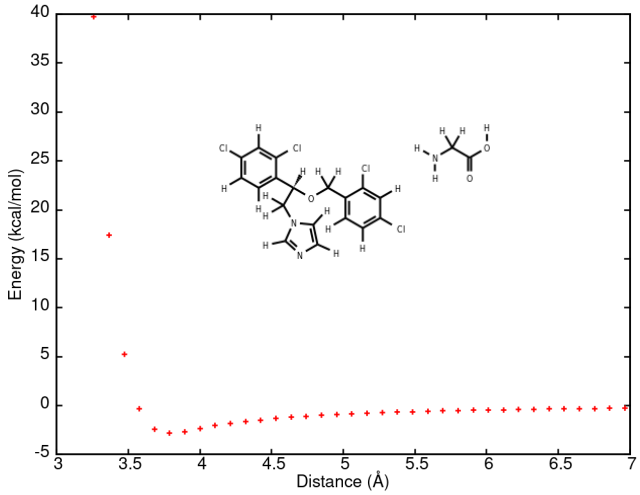


Figure 60: System 60



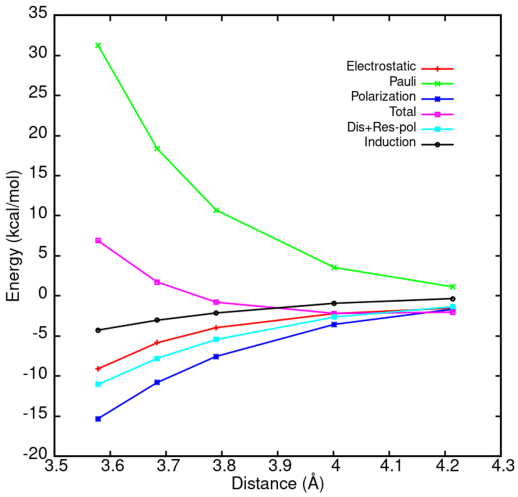
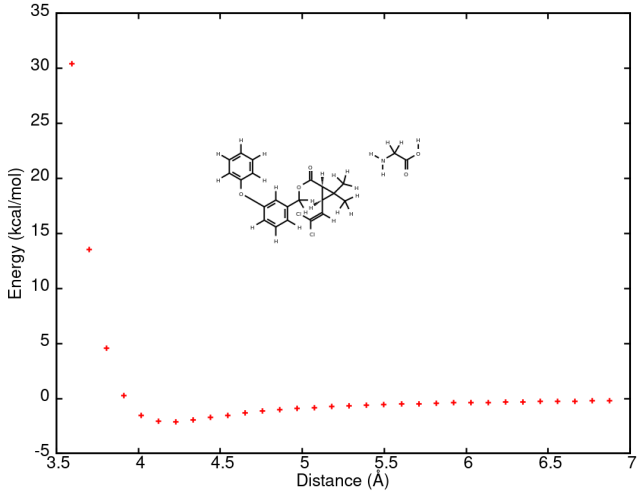


Figure 61: System 61



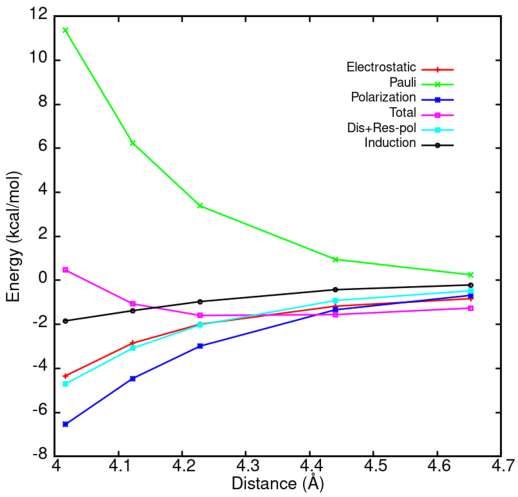
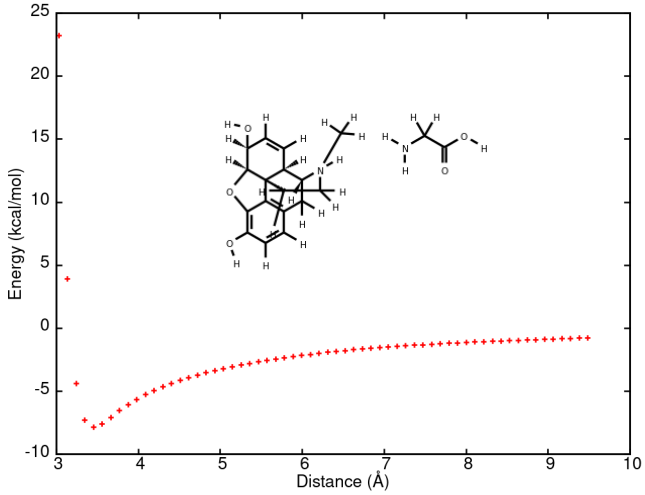


Figure 62: System 62



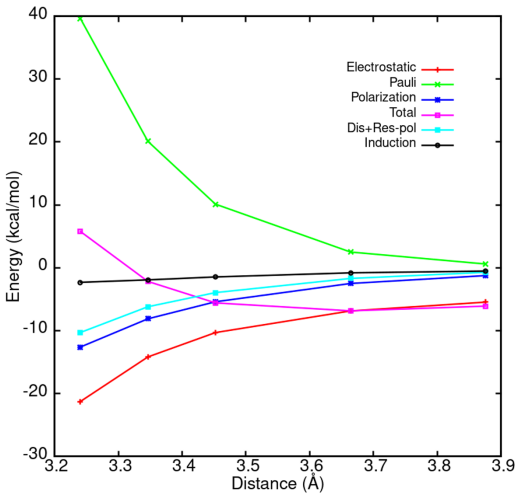


Figure 63: System 63

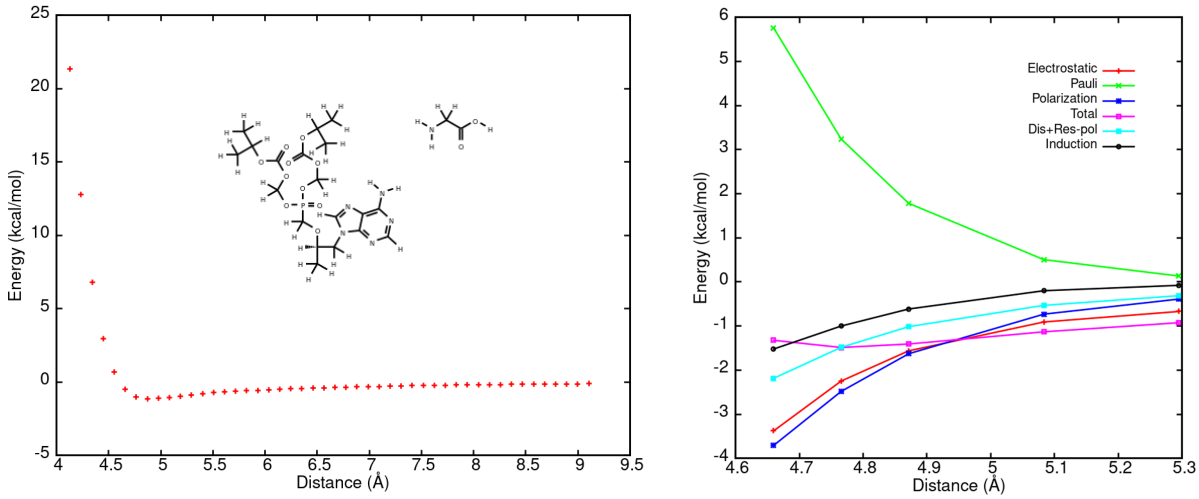


Figure 64: System 64

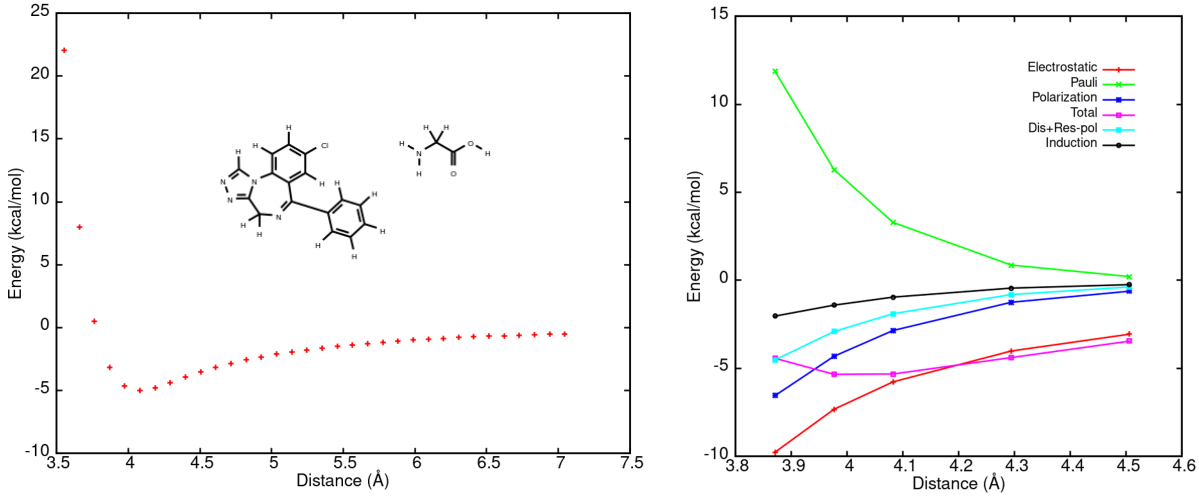


Figure 65: System 65

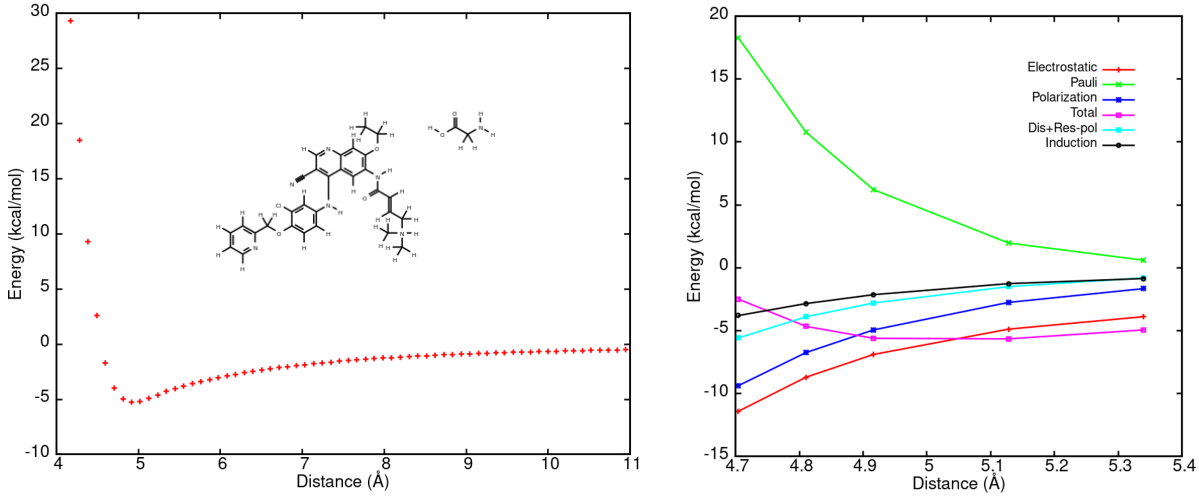


Figure 66: System 66

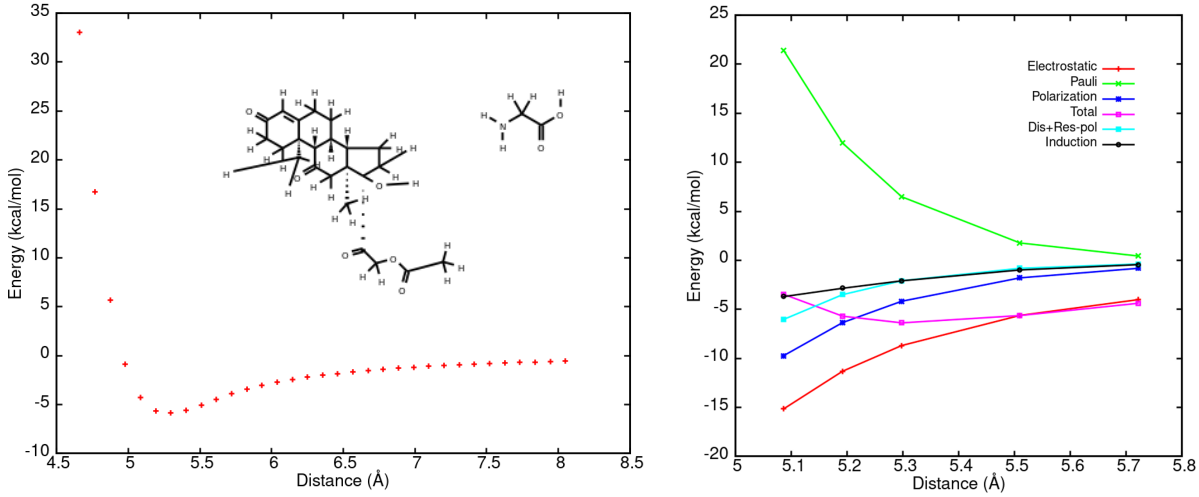


Figure 67: System 67

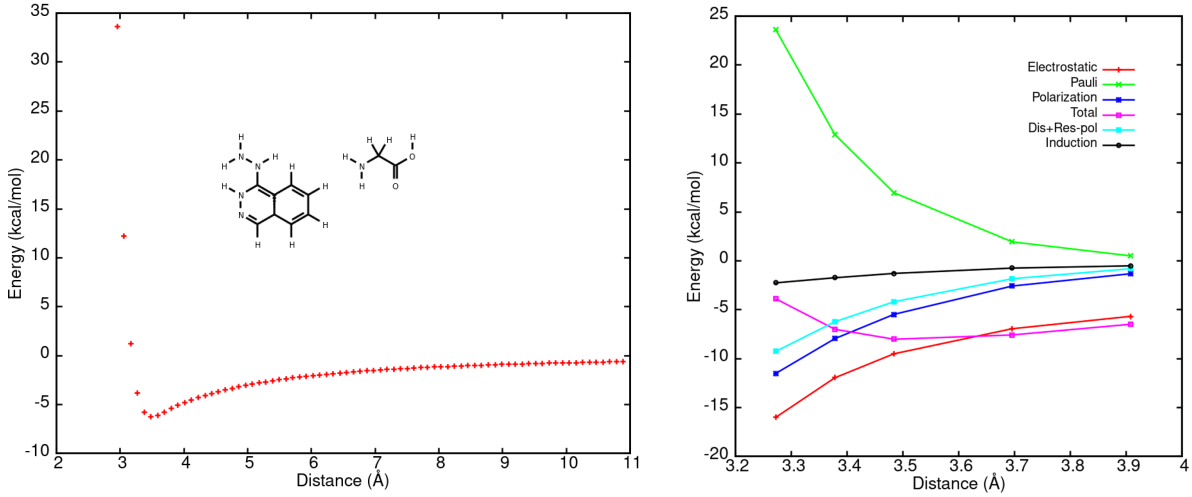


Figure 68: System 68

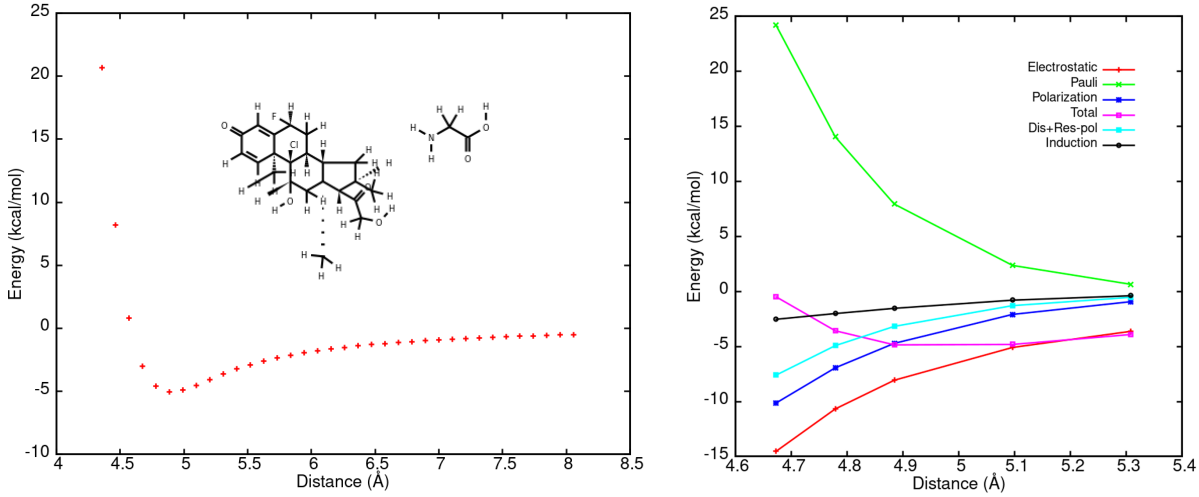


Figure 69: System 69

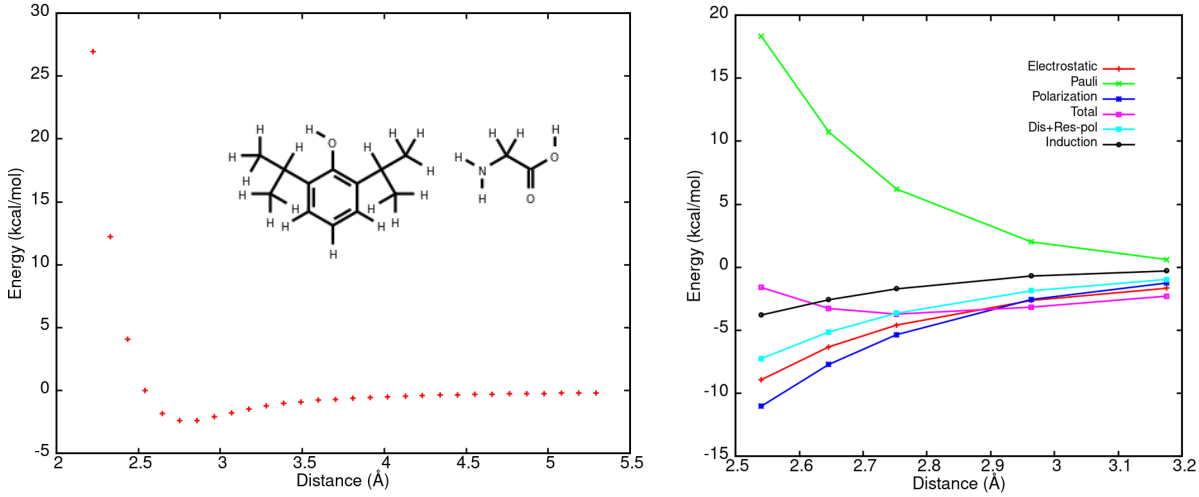


Figure 70: System 70

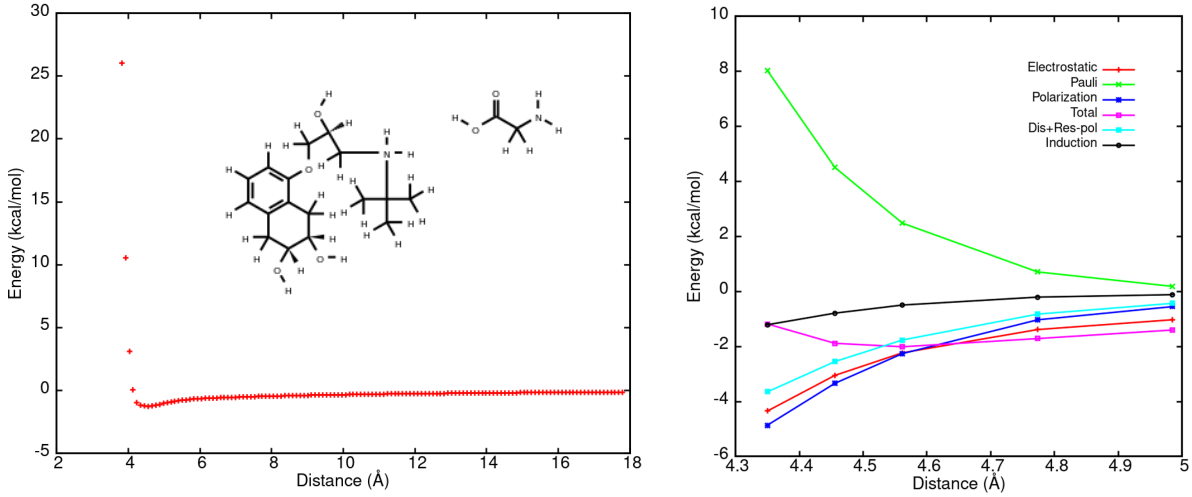


Figure 71: System 71

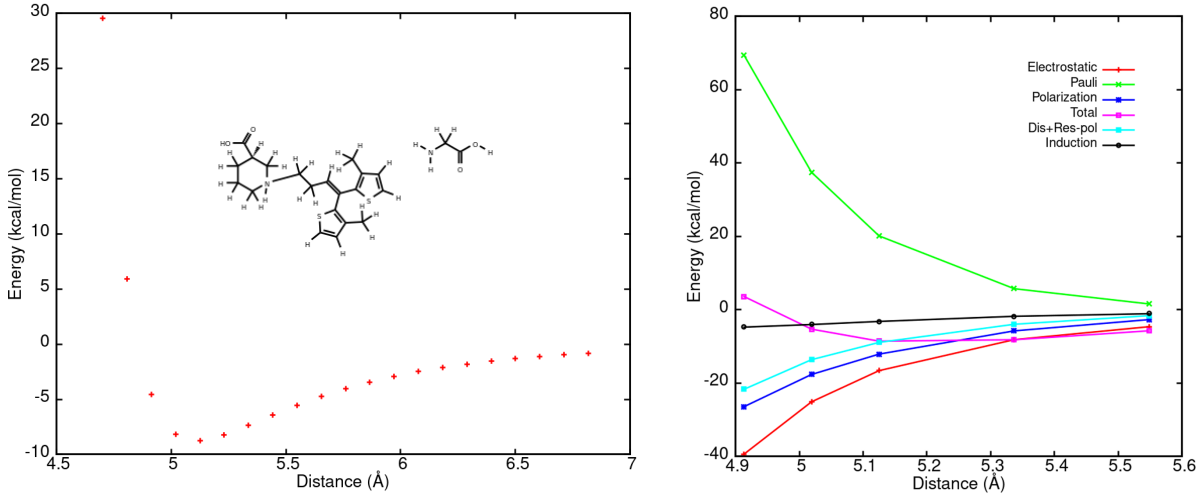
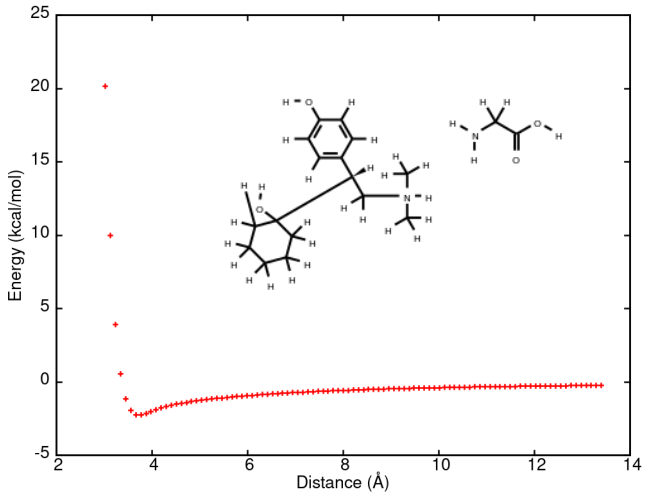


Figure 72: System 72



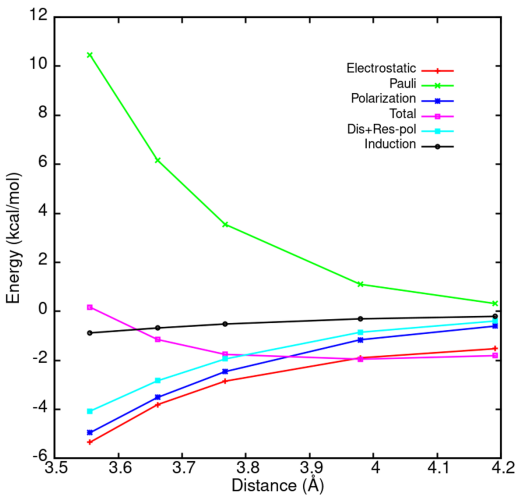
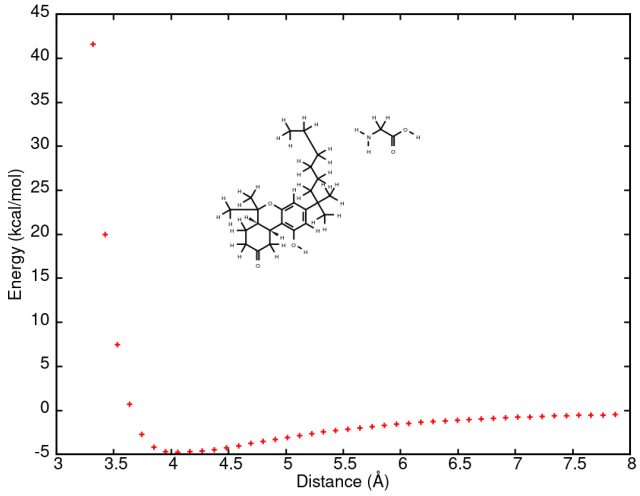


Figure 73: System 73



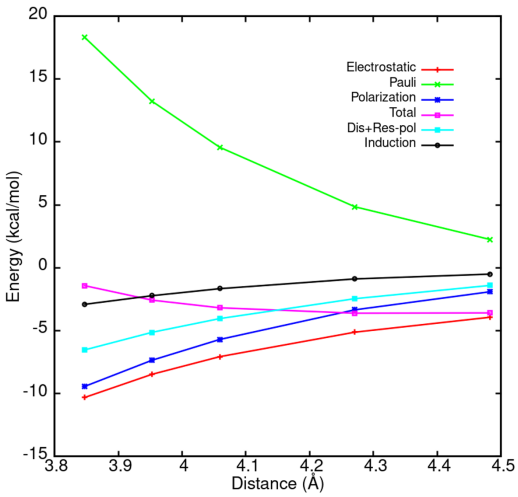
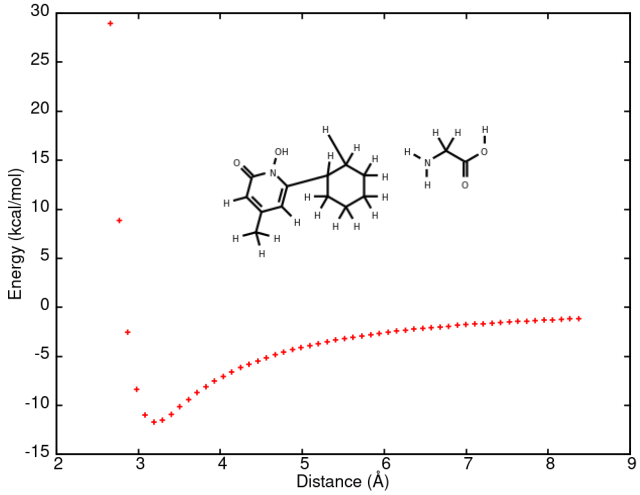


Figure 74: System 74



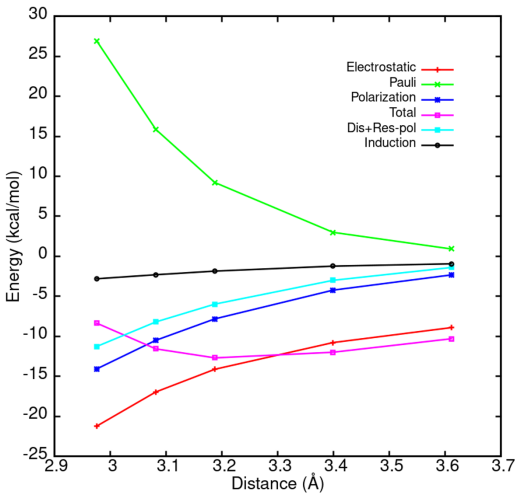


Figure 75: System 75

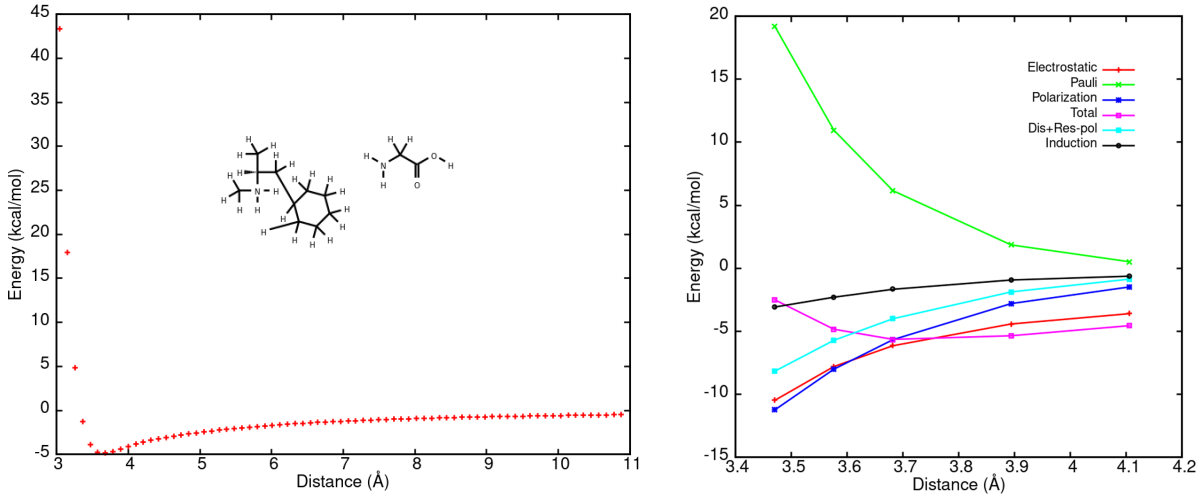


Figure 76: System 76

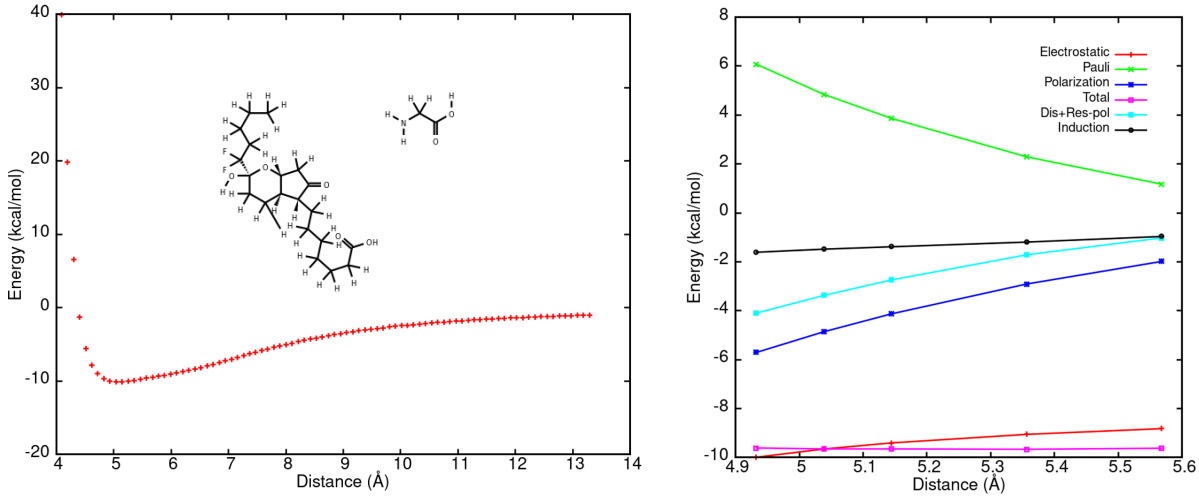


Figure 77: System 77

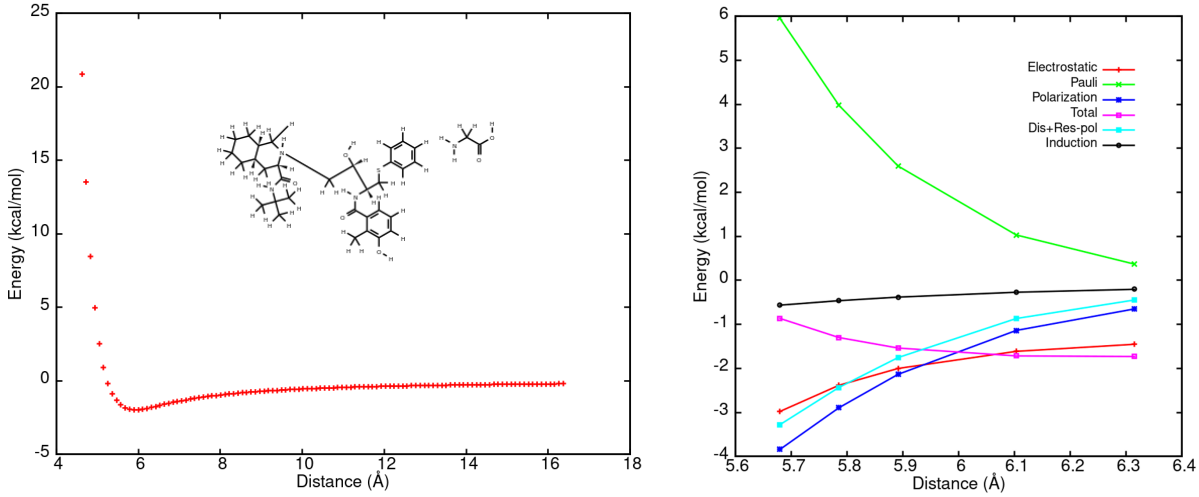
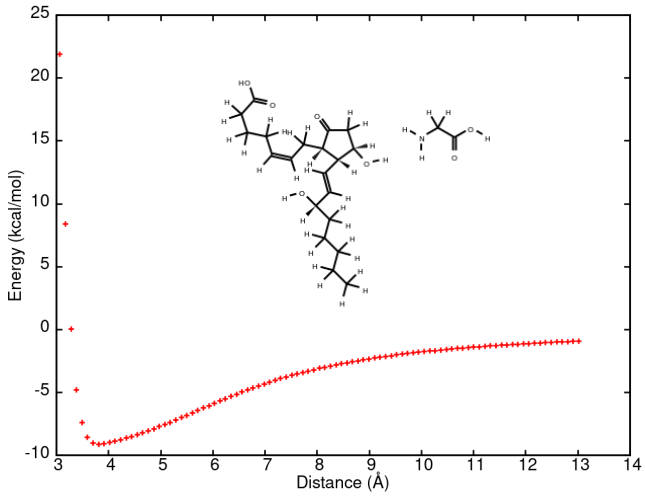


Figure 78: System 78



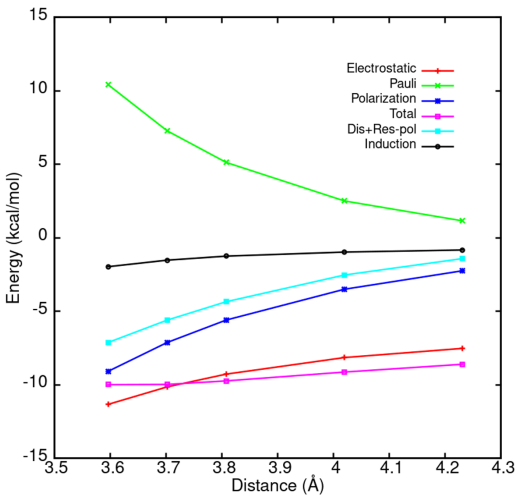
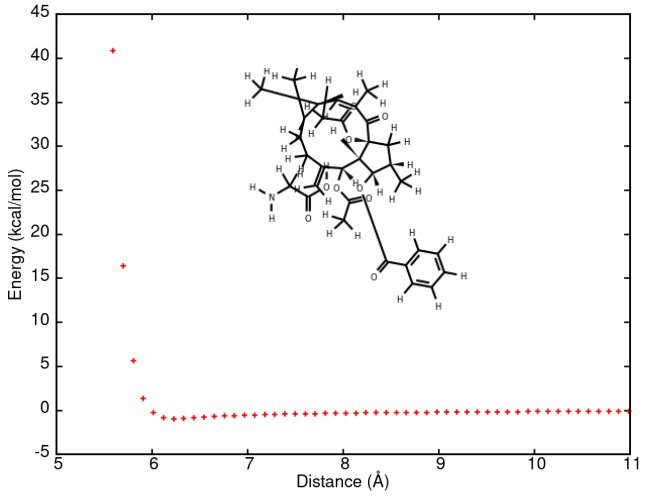


Figure 79: System 79



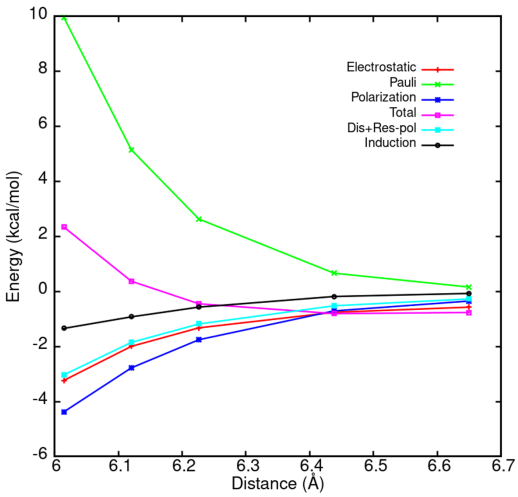
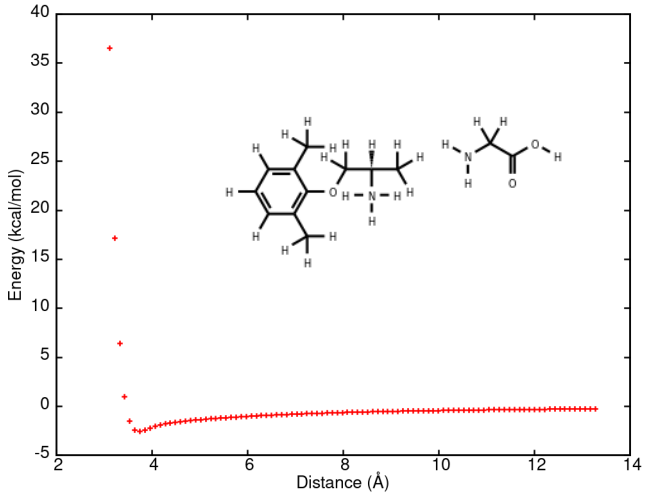


Figure 80: System 80



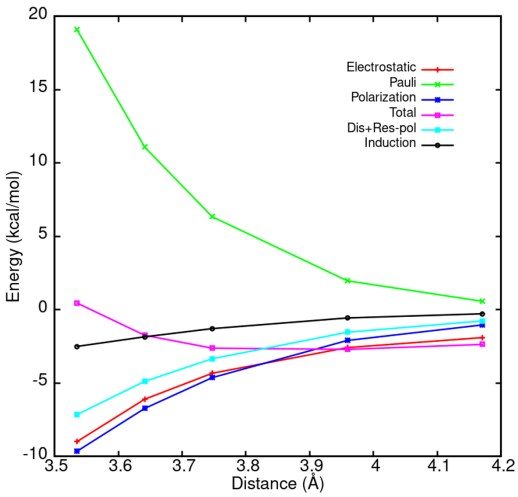


Figure 81: System 81

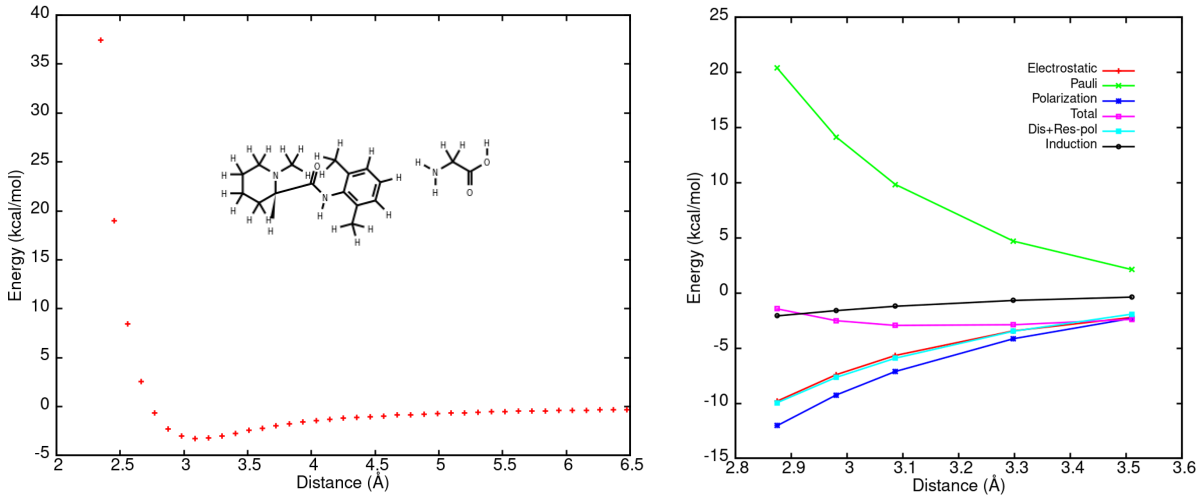


Figure 82: System 82

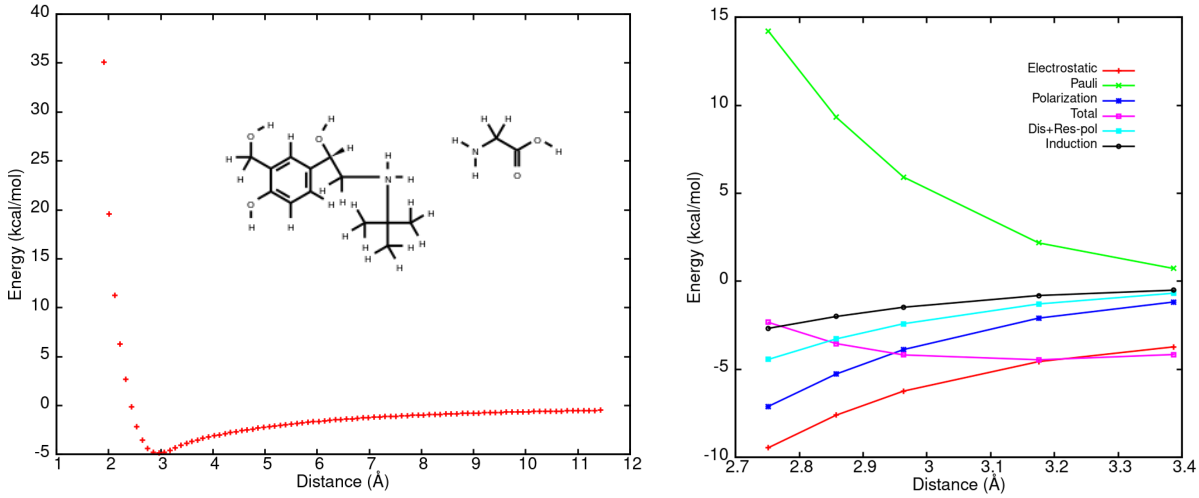


Figure 83: System 83

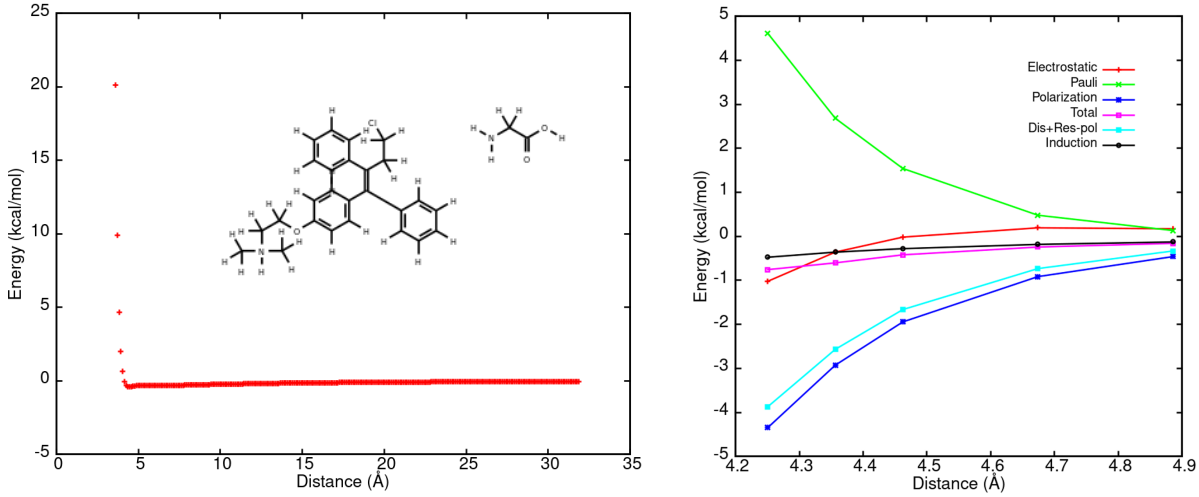


Figure 84: System 84

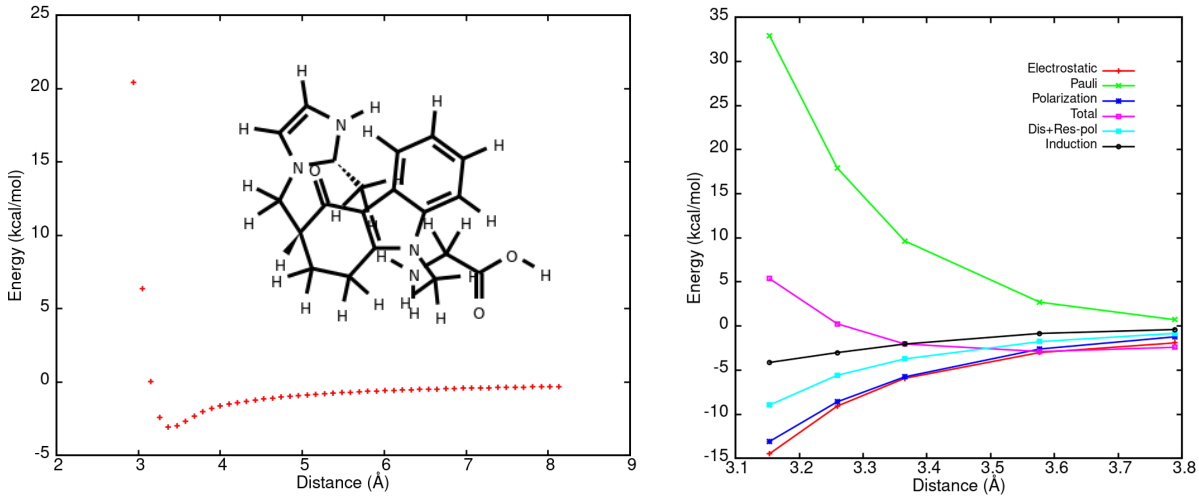


Figure 85: System 85

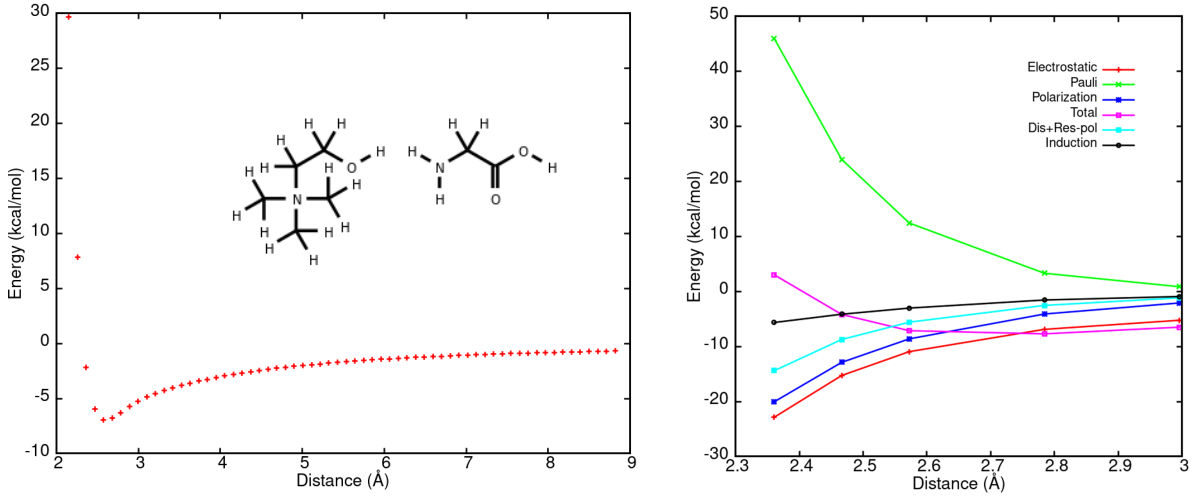


Figure 86: System 86

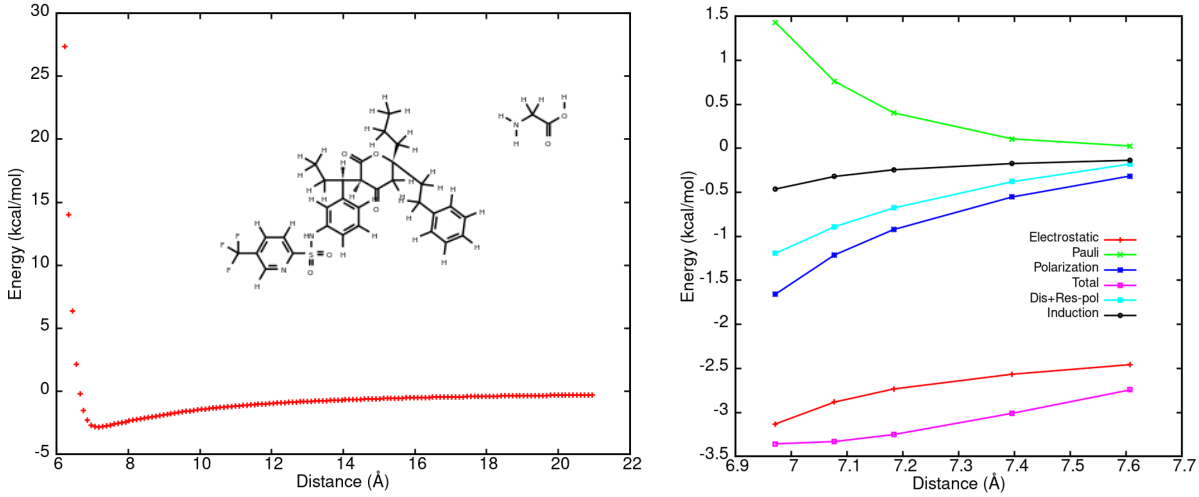


Figure 87: System 87

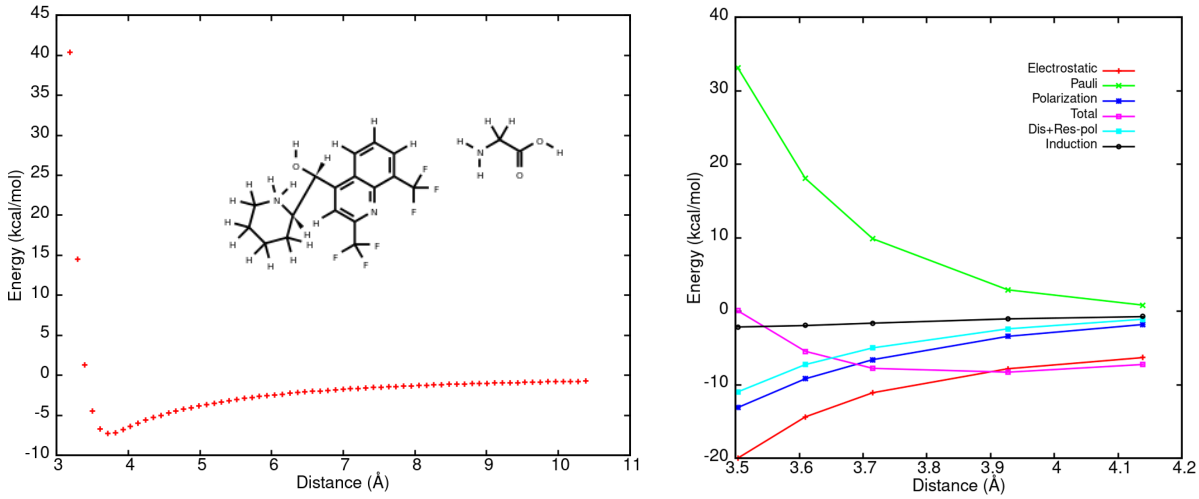


Figure 88: System 88

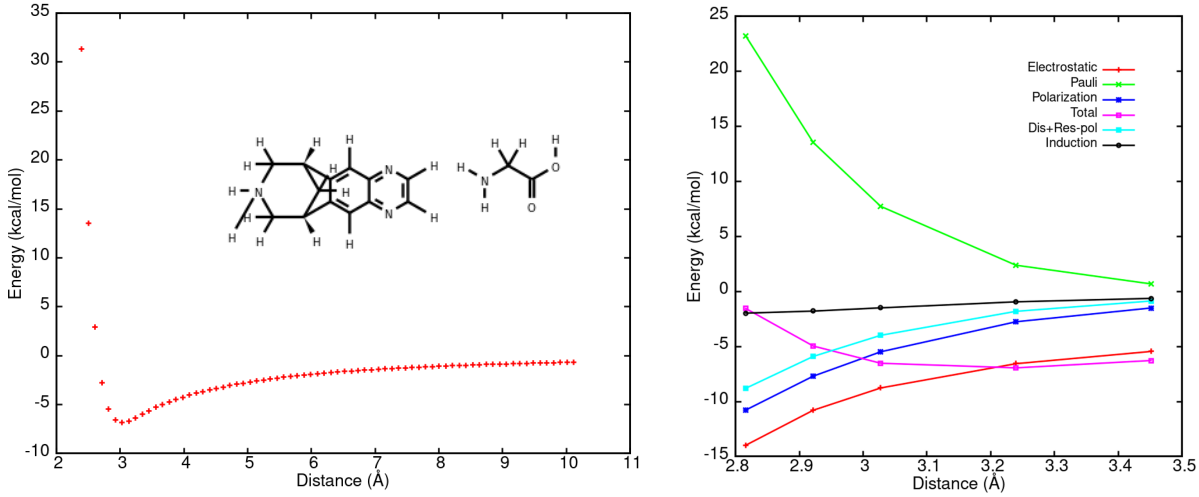


Figure 89: System 89

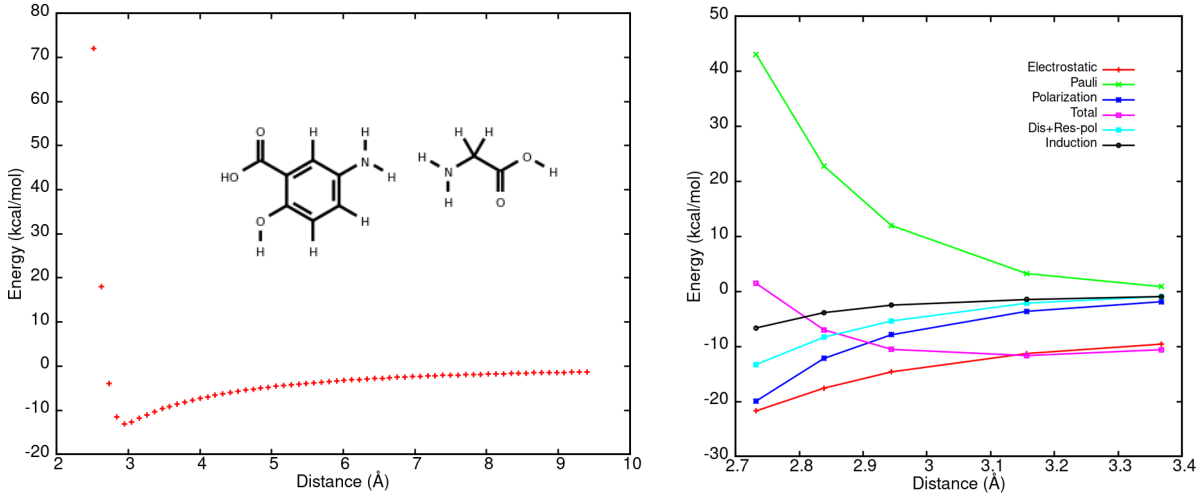


Figure 90: System 90

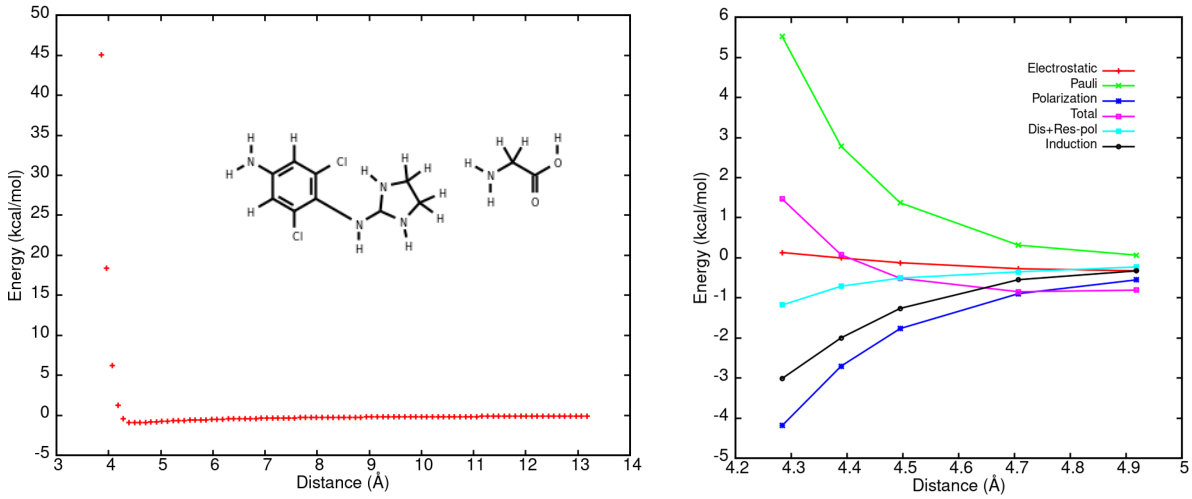


Figure 91: System 91

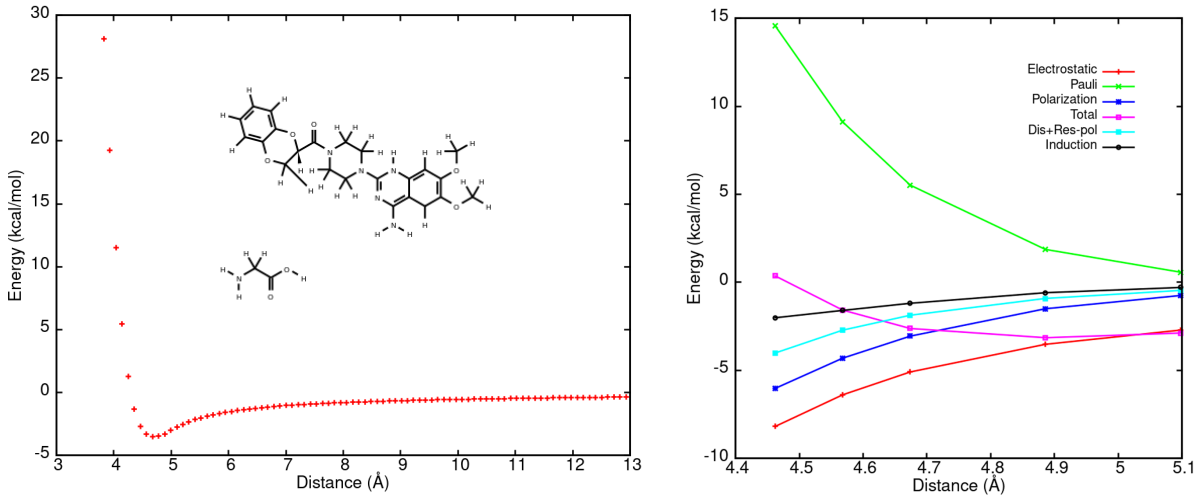


Figure 92: System 92

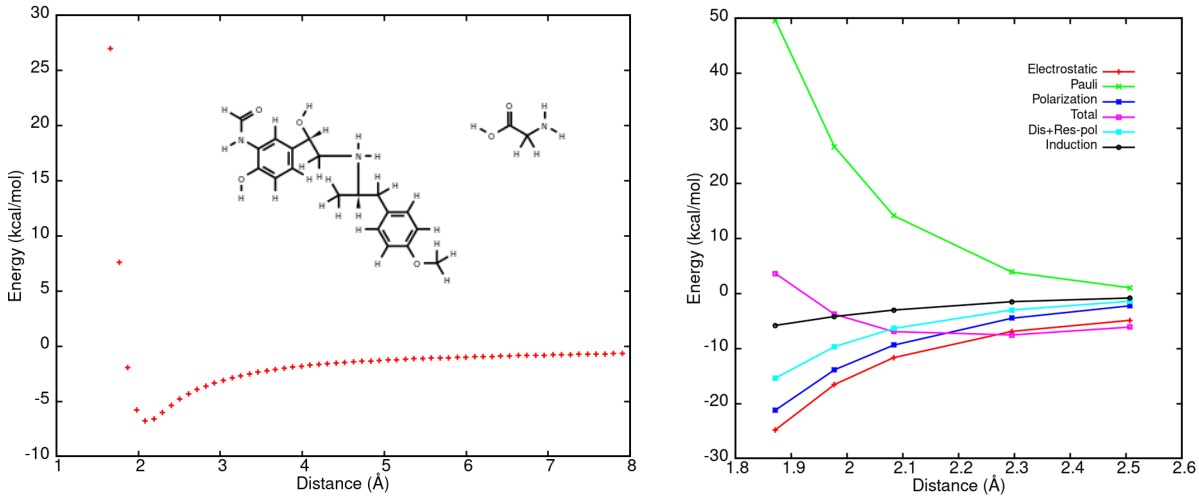


Figure 93: System 93

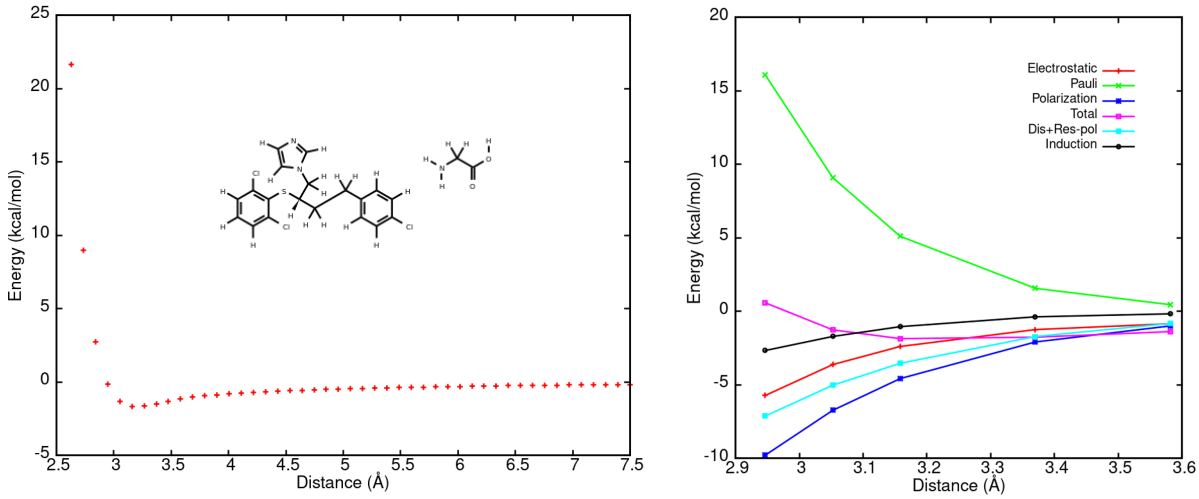


Figure 94: System 94

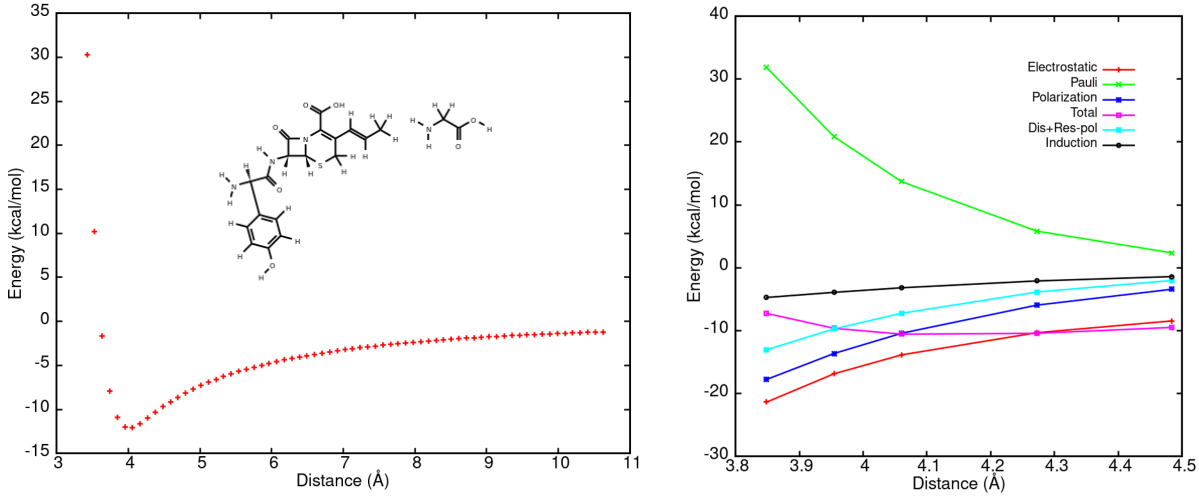


Figure 95: System 95

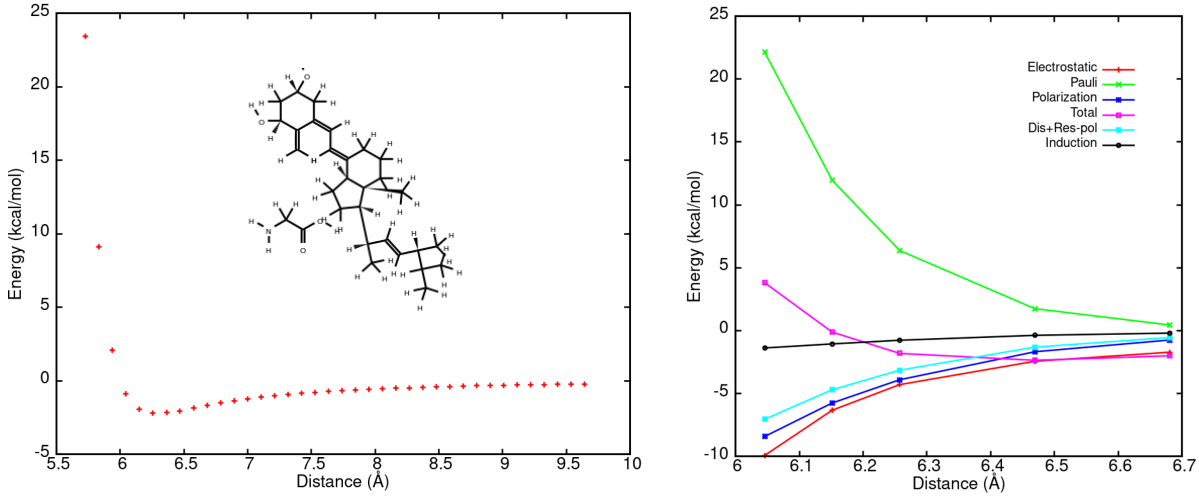


Figure 96: System 96

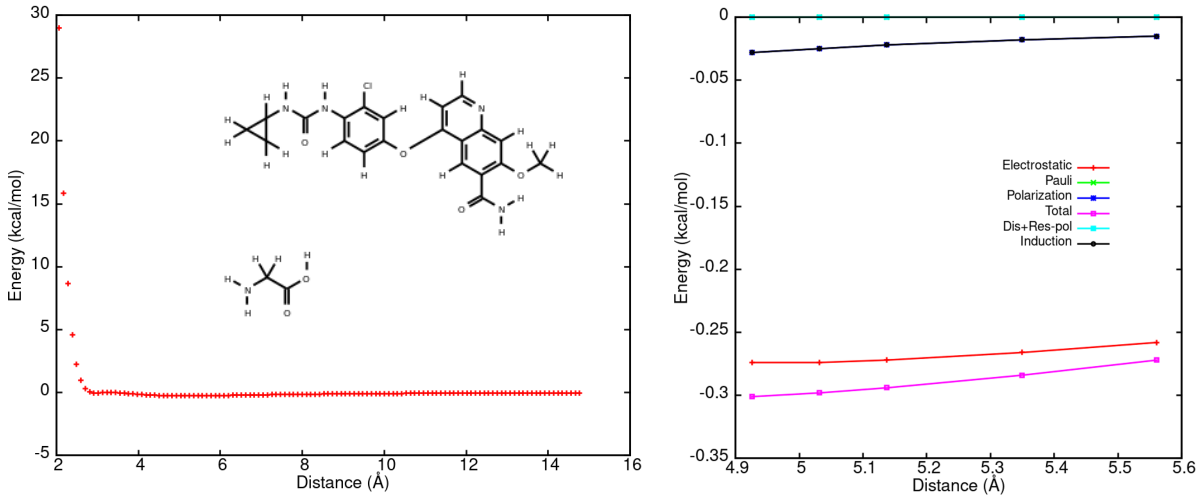


Figure 97: System 97

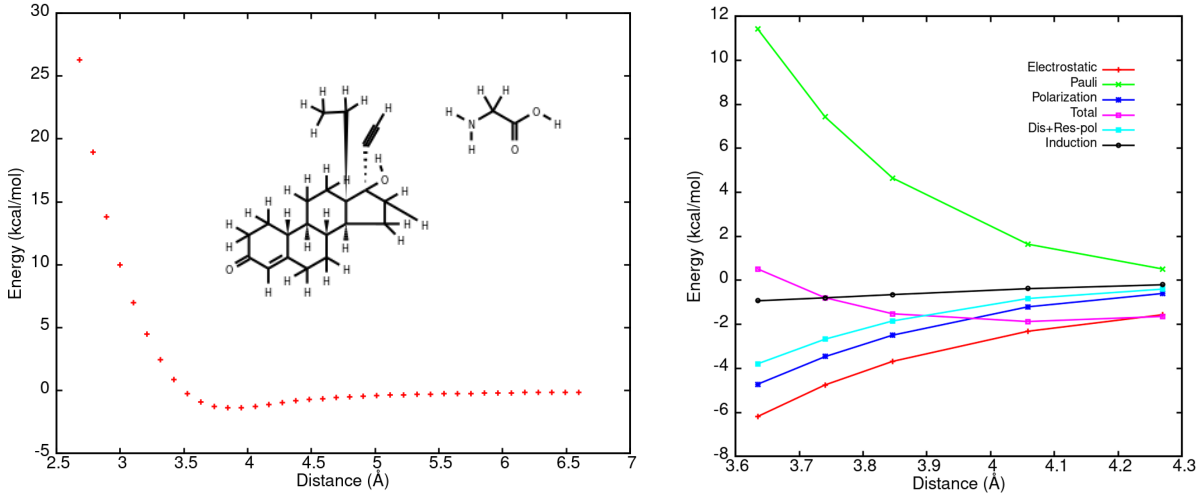


Figure 98: System 98

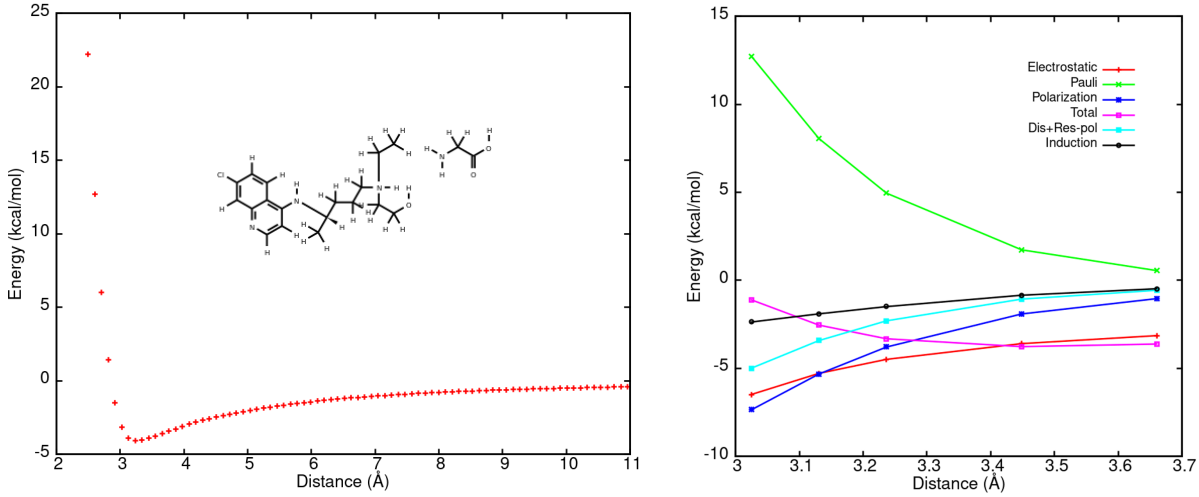


Figure 99: System 99

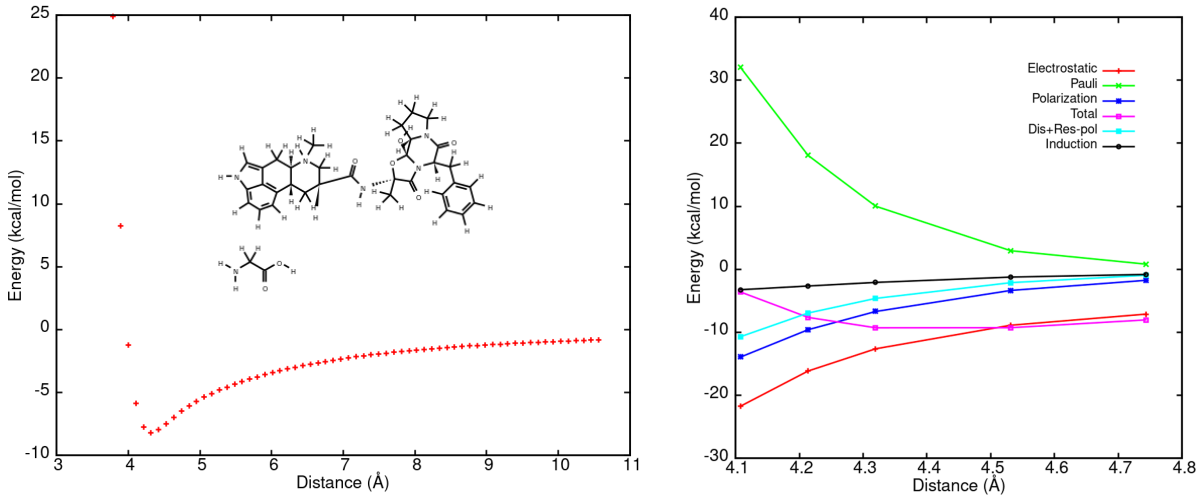


Figure 100: System 100

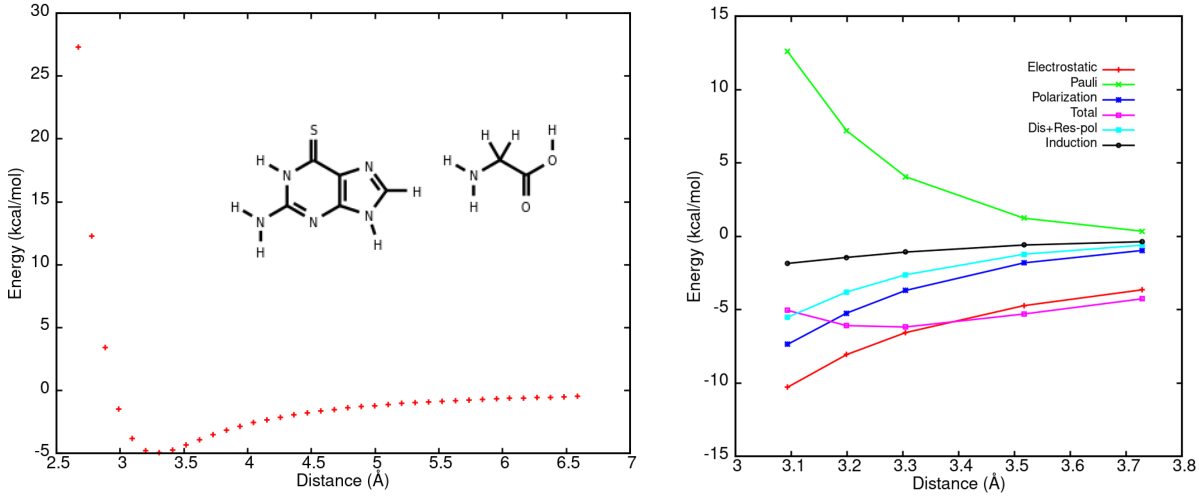


Figure 101: System 101

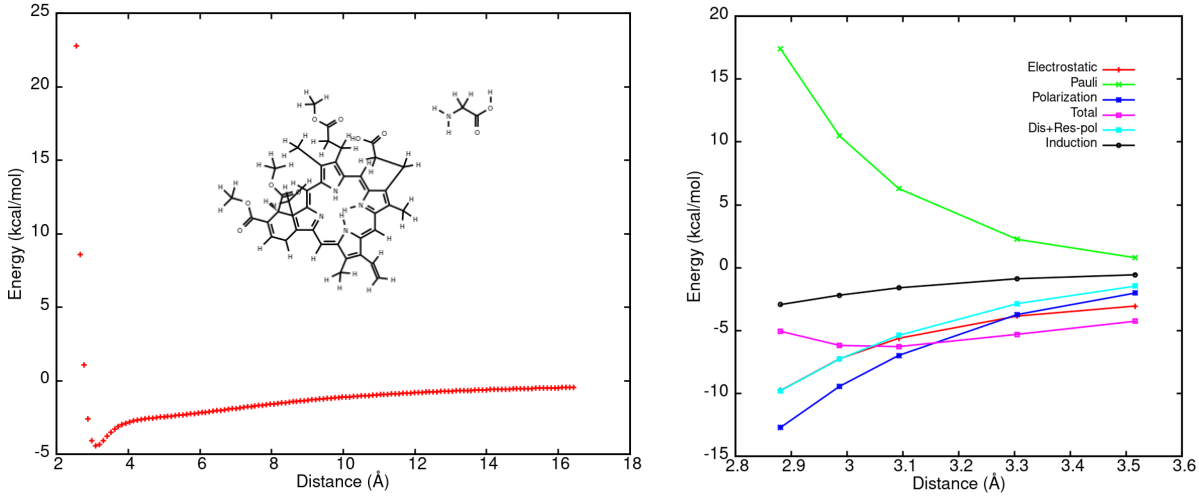
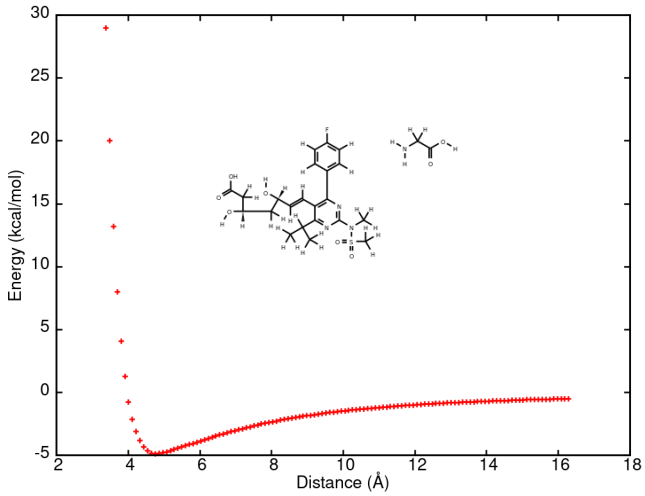


Figure 102: System 102



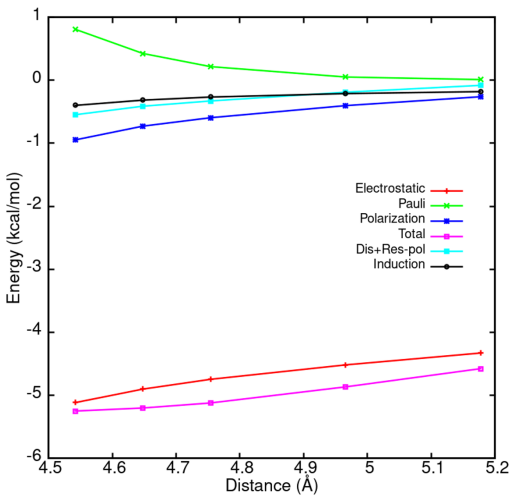
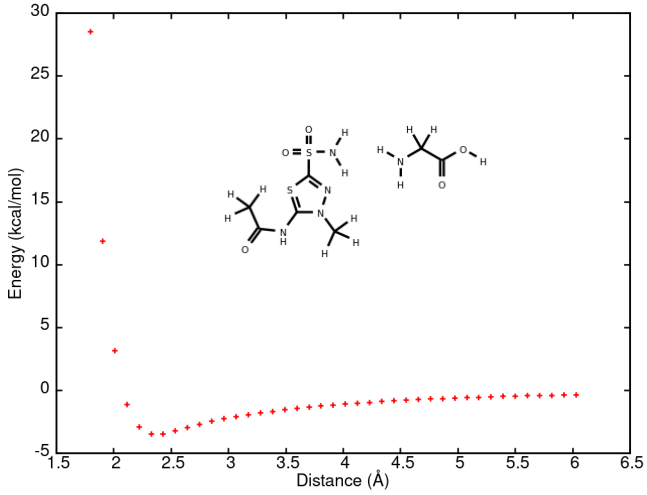


Figure 103: System 103



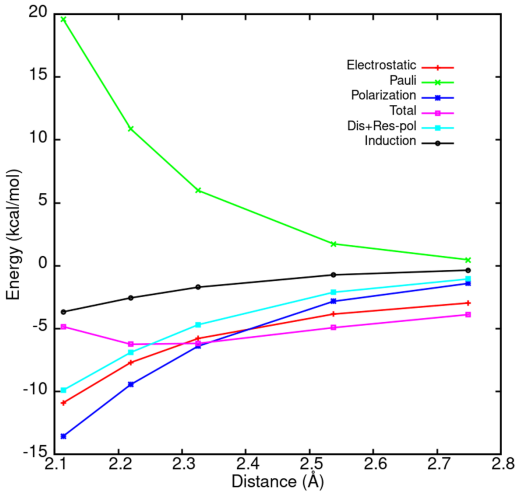
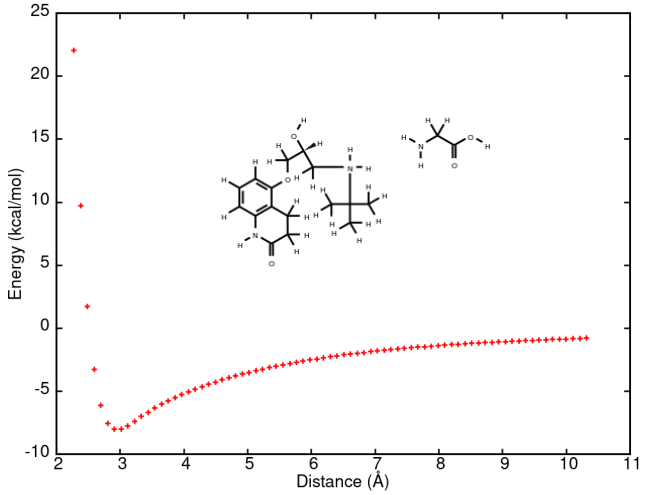


Figure 104: System 104



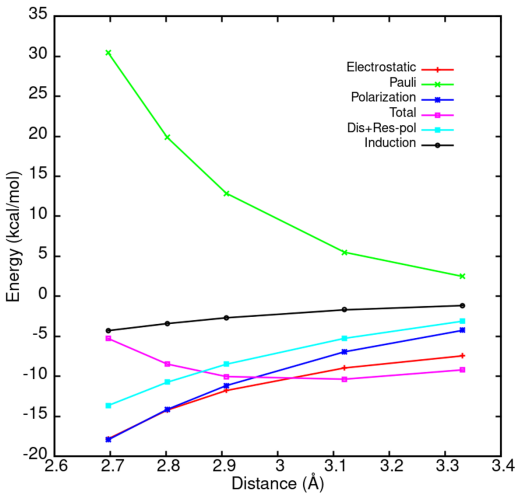
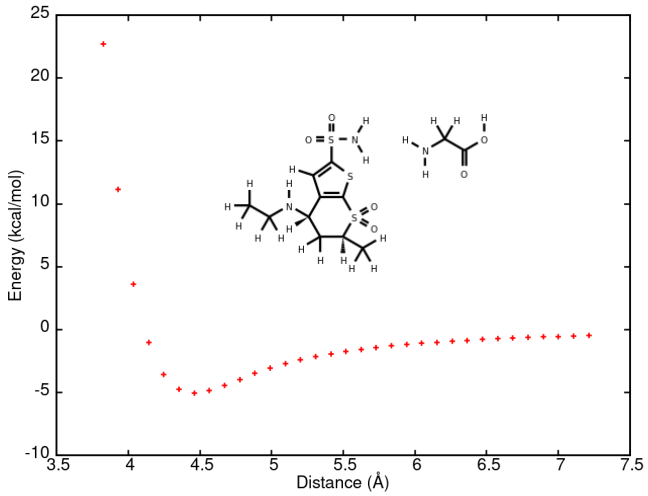


Figure 105: System 105



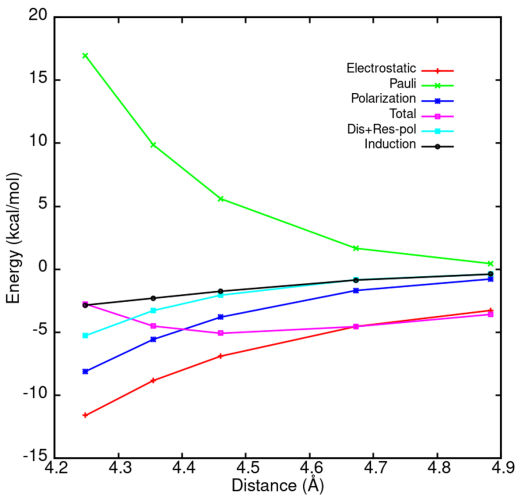
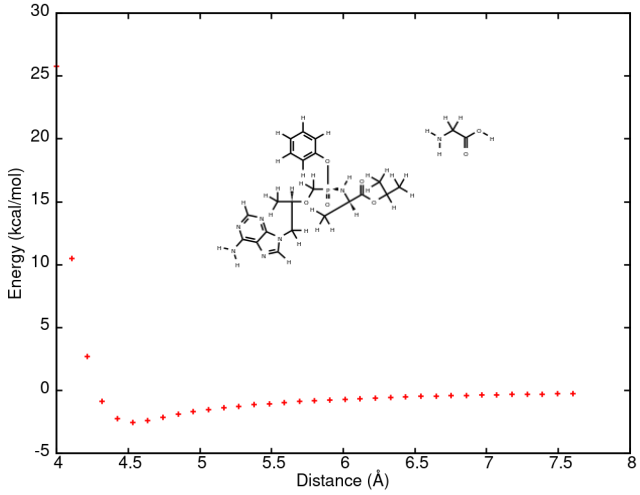


Figure 106: System 106



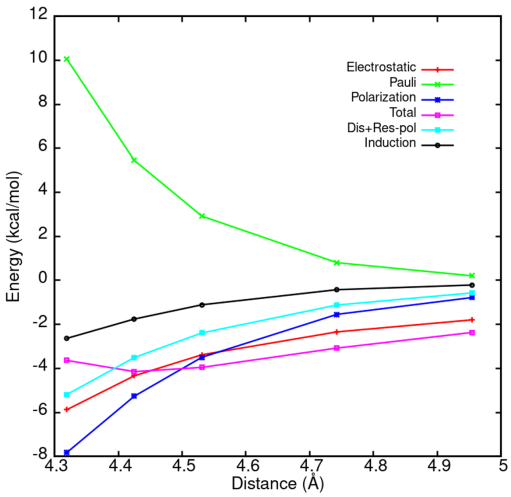
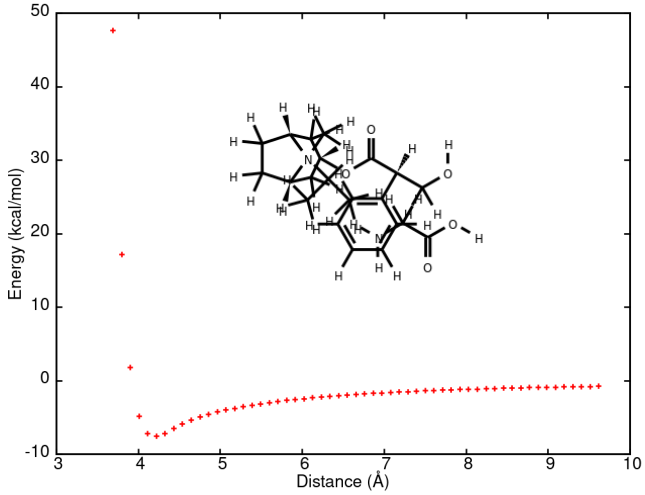


Figure 107: System 107



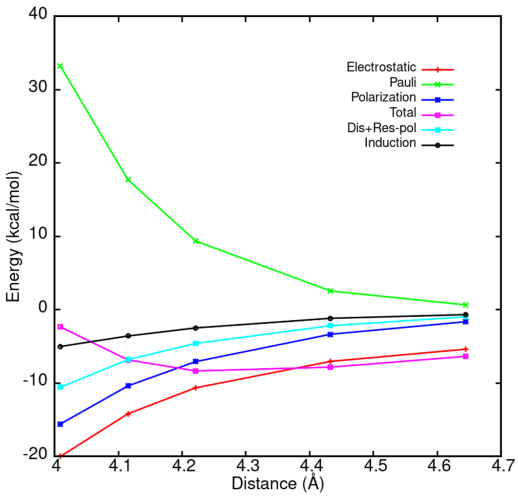
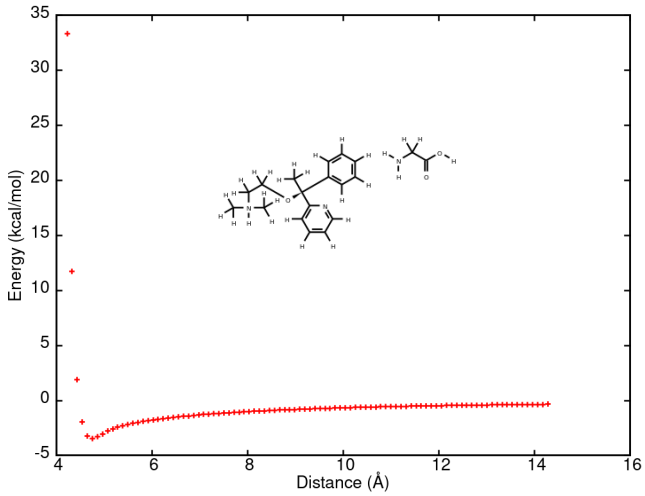


Figure 108: System 108



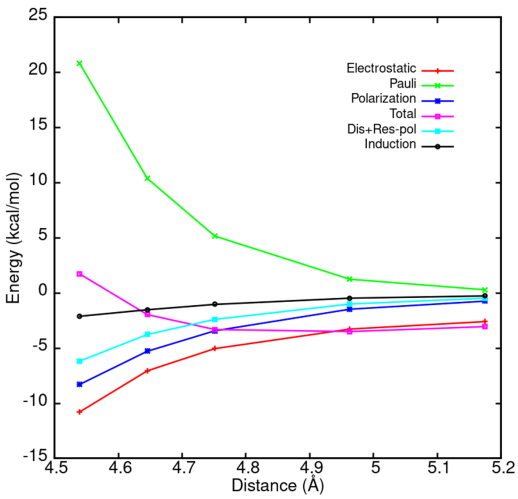
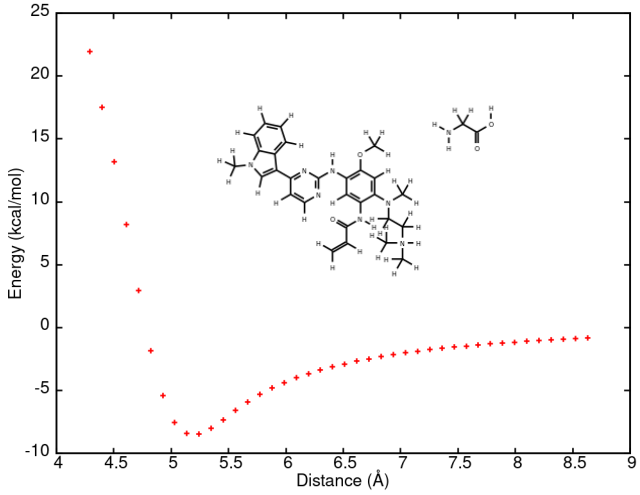


Figure 109: System 109



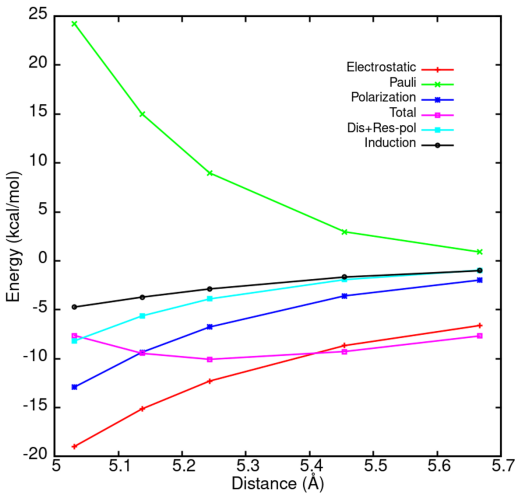
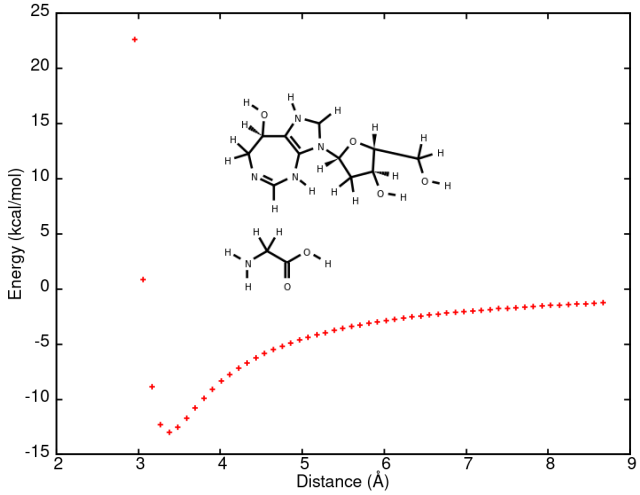


Figure 110: System 110



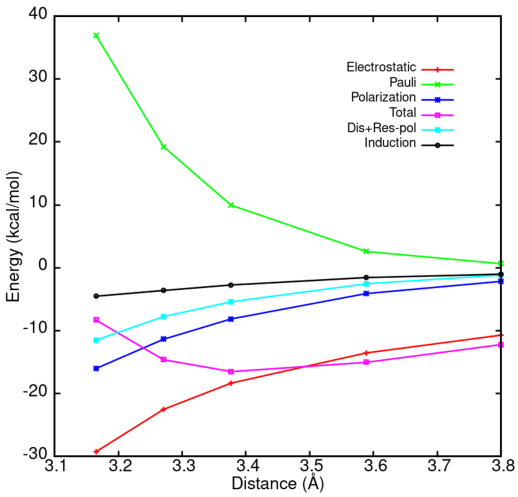


Figure 111: System 111

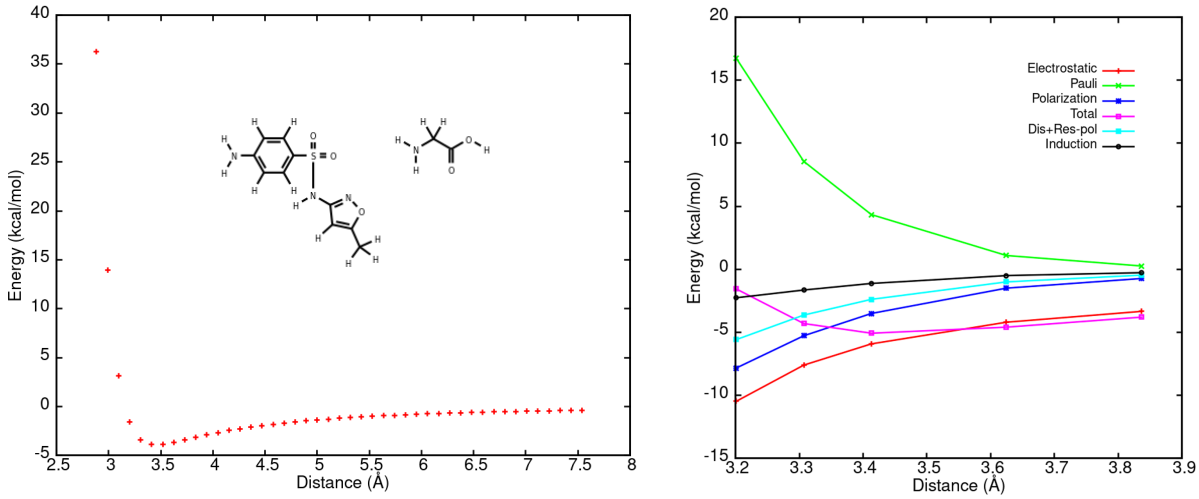


Figure 112: System 112

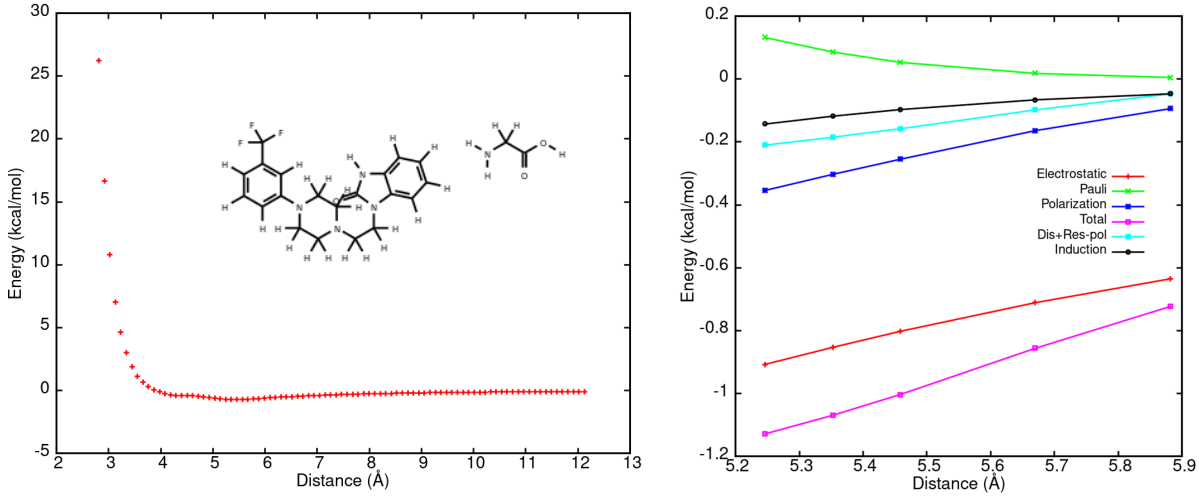


Figure 113: System 113

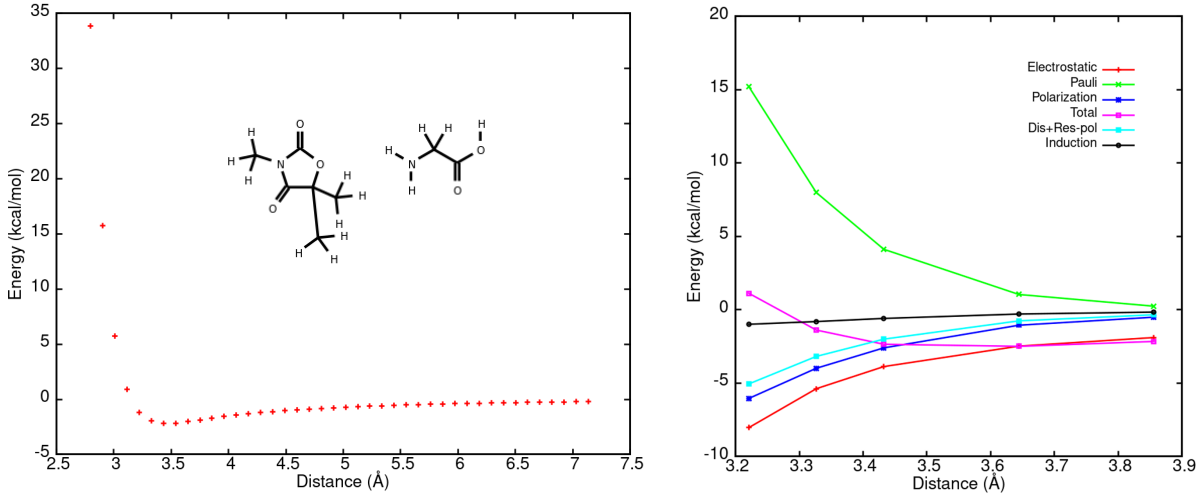


Figure 114: System 114

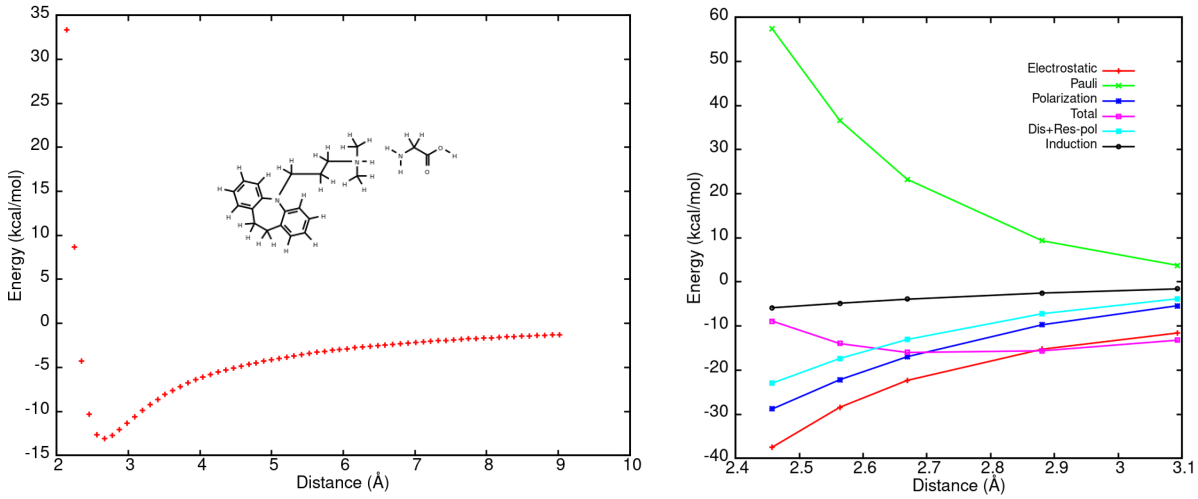


Figure 115: System 115

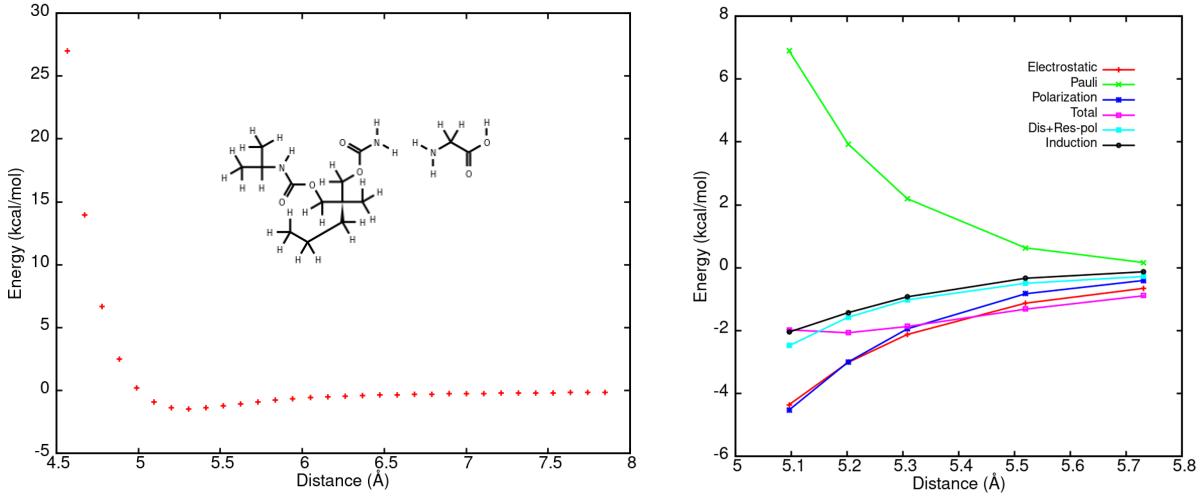


Figure 116: System 116

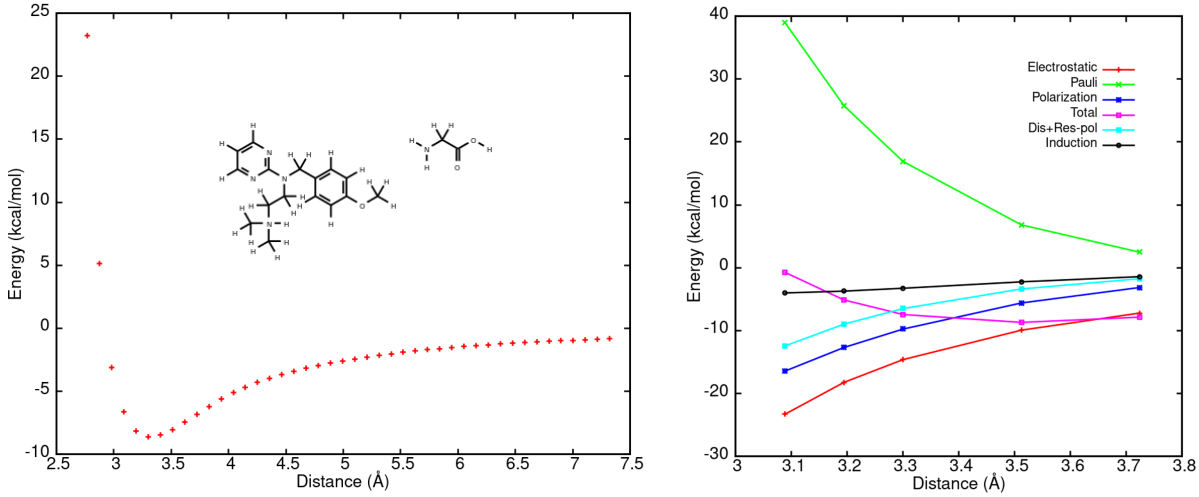


Figure 117: System 117

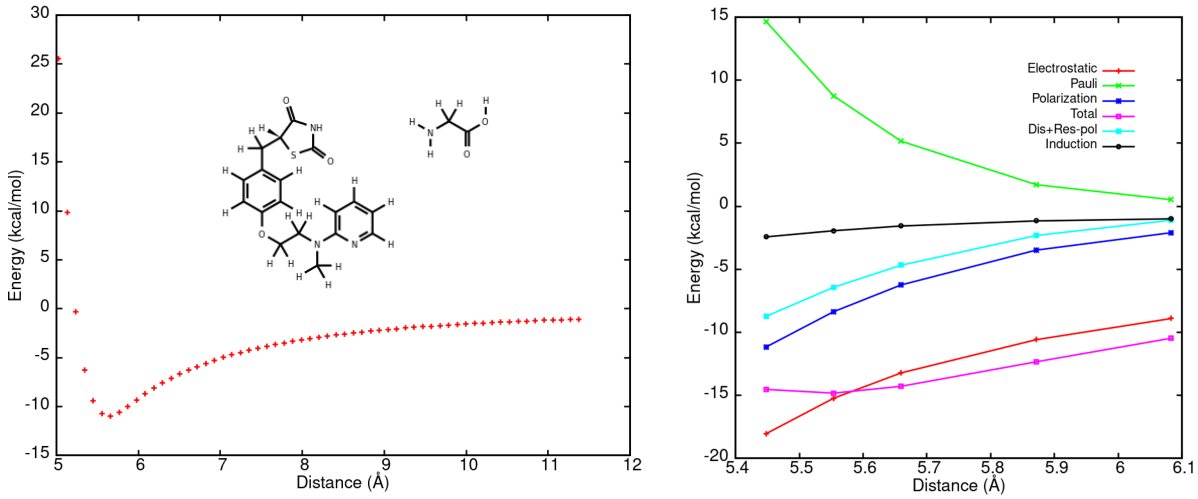


Figure 118: System 118

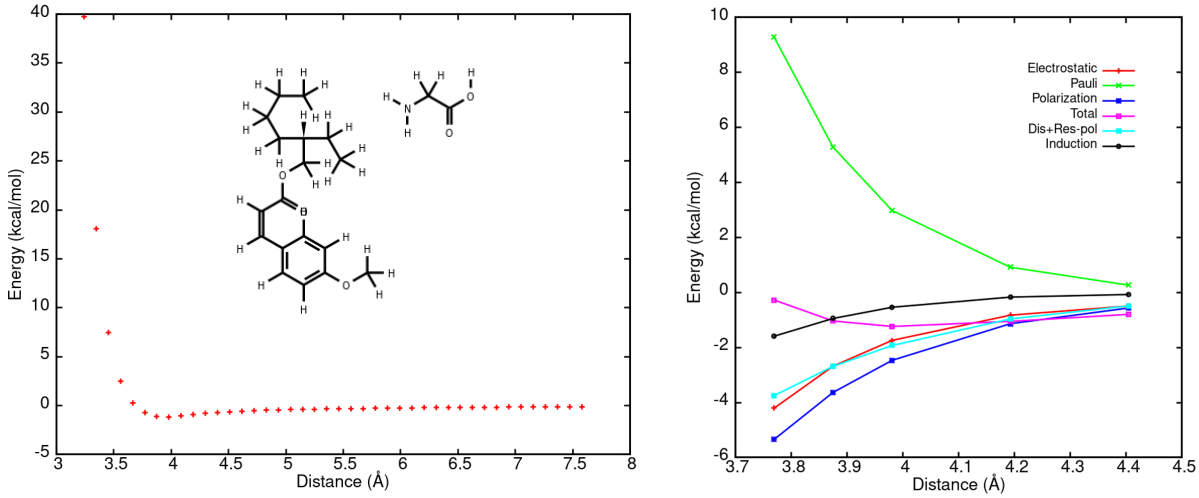


Figure 119: System 119

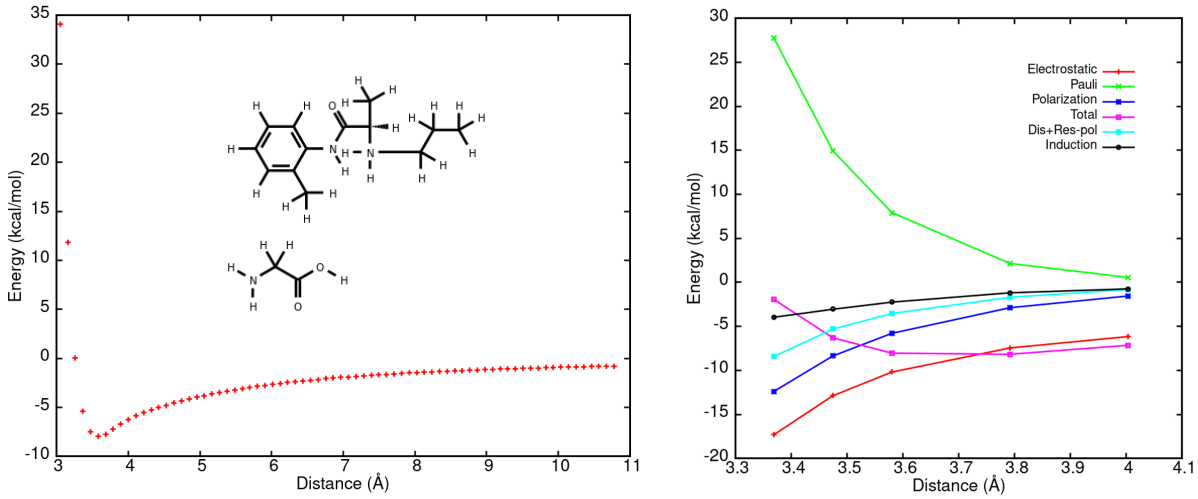
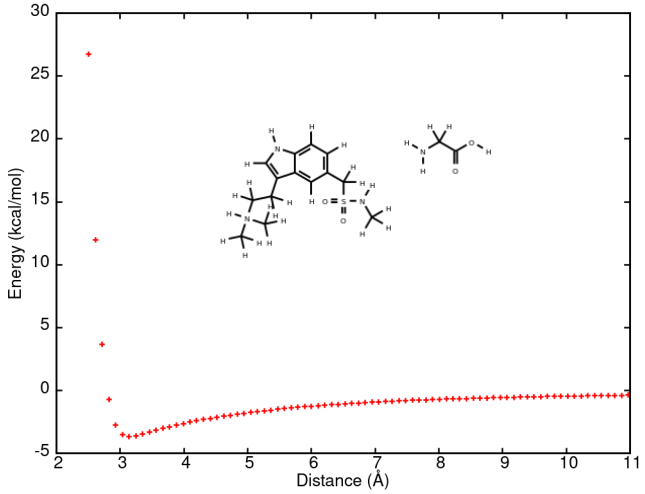


Figure 120: System 120



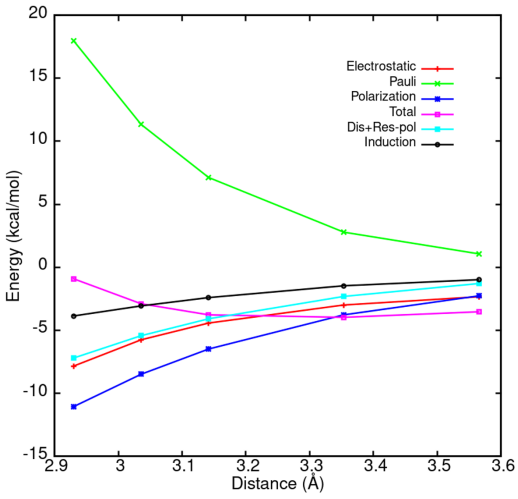
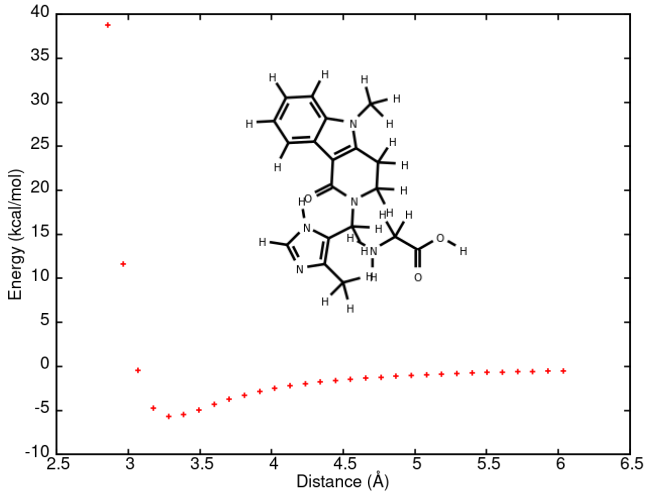


Figure 121: System 121



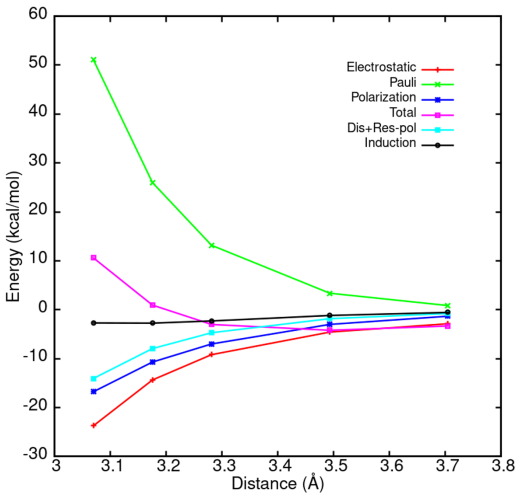
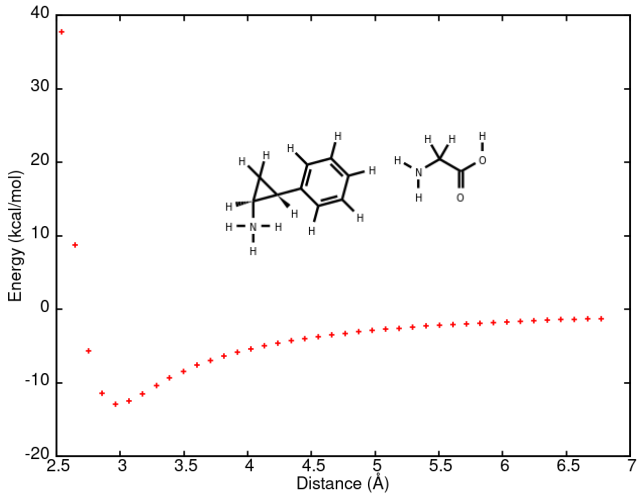


Figure 122: System 122



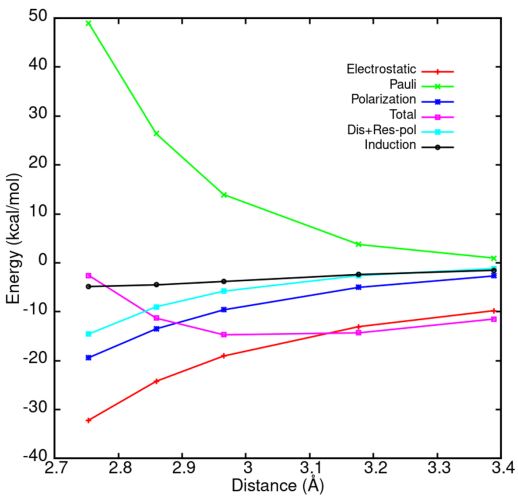
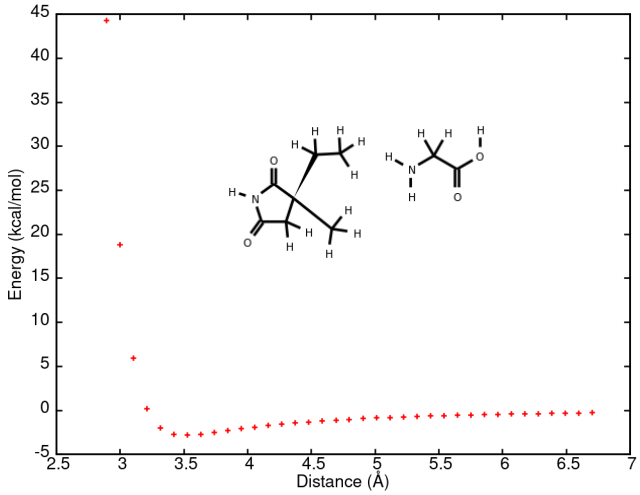


Figure 123: System 123



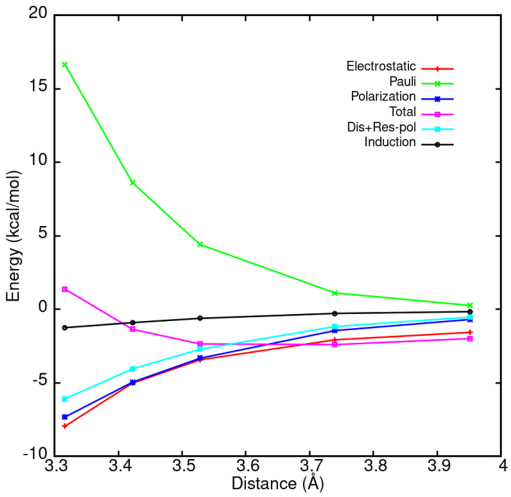
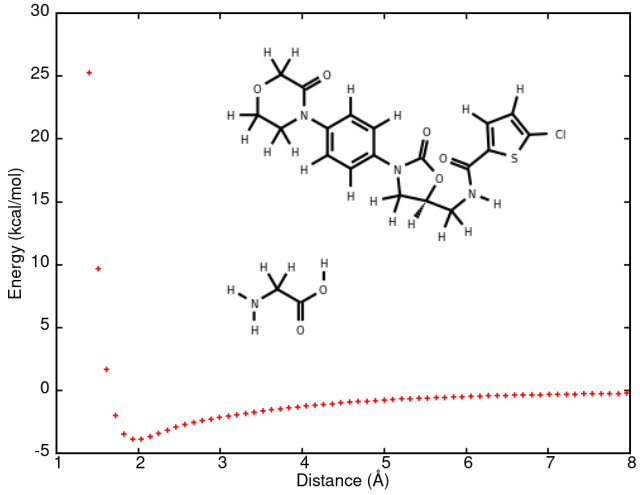


Figure 124: System 124



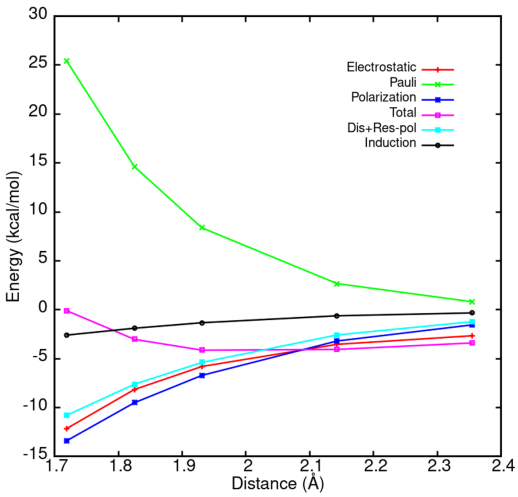
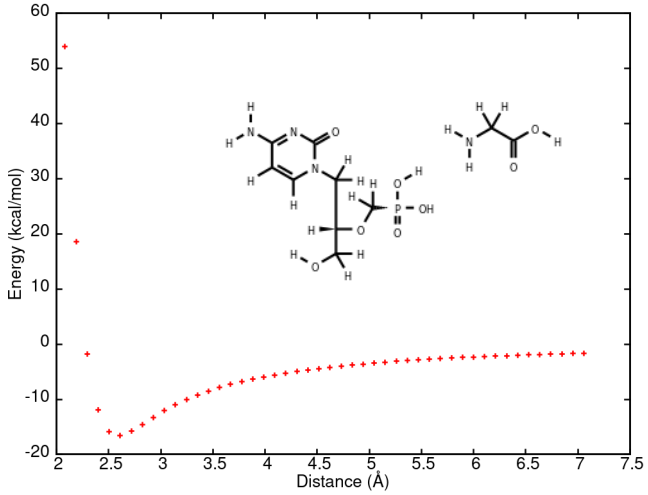


Figure 125: System 125



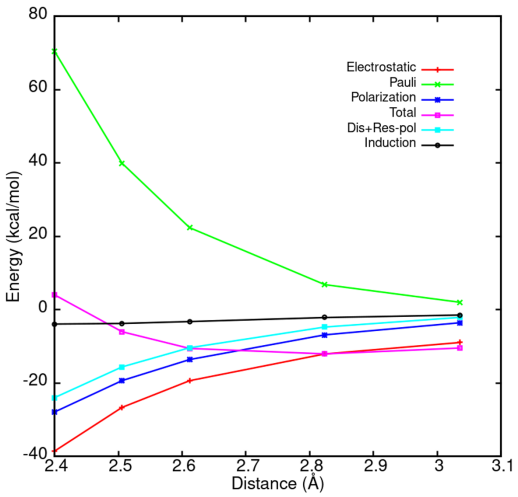
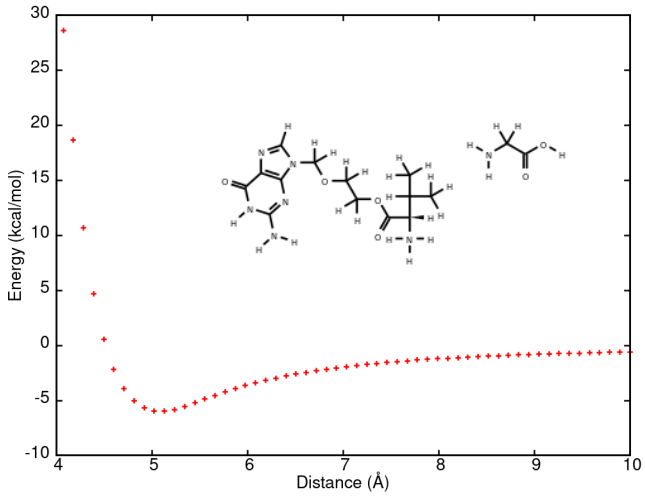


Figure 126: System 126



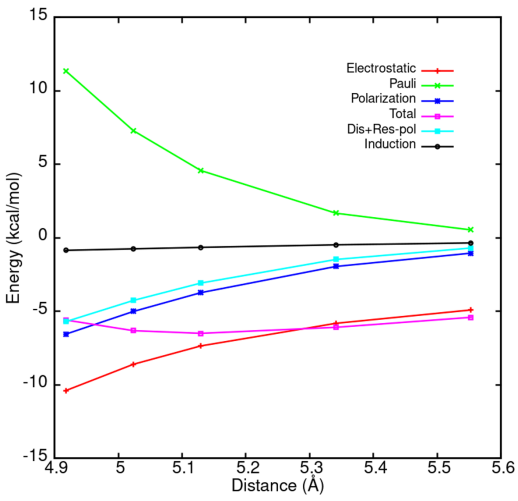
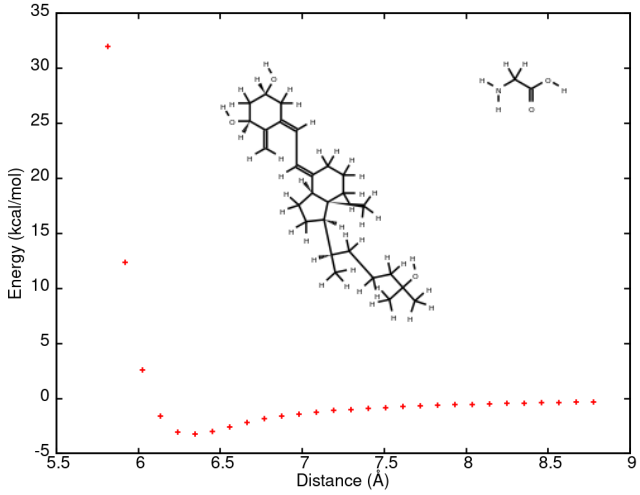


Figure 127: System 127



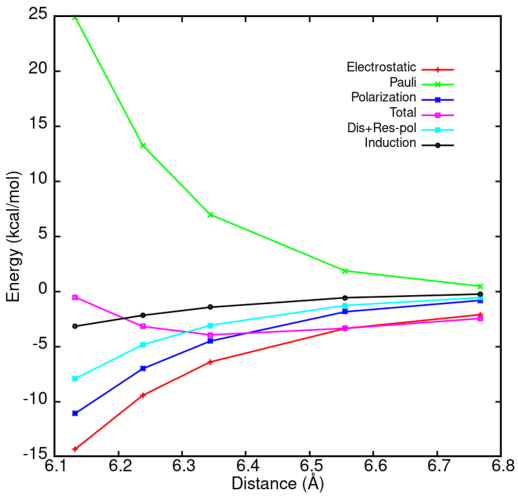
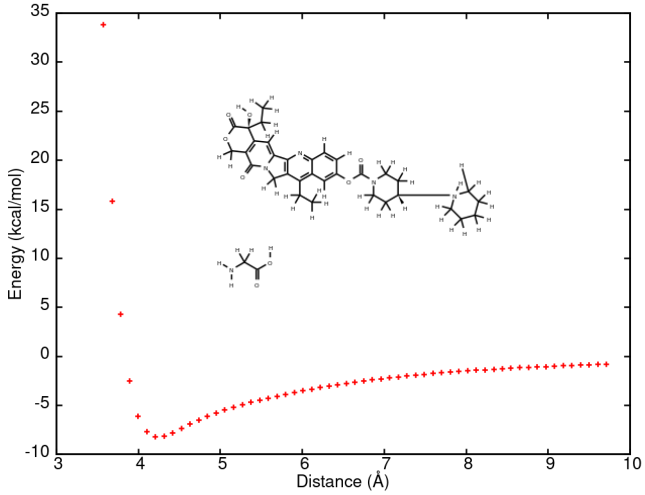


Figure 128: System 128



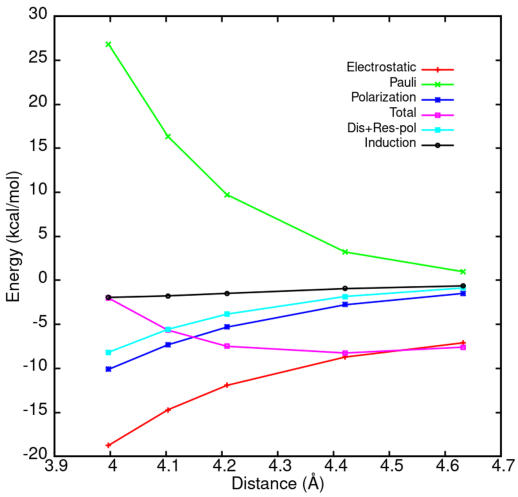
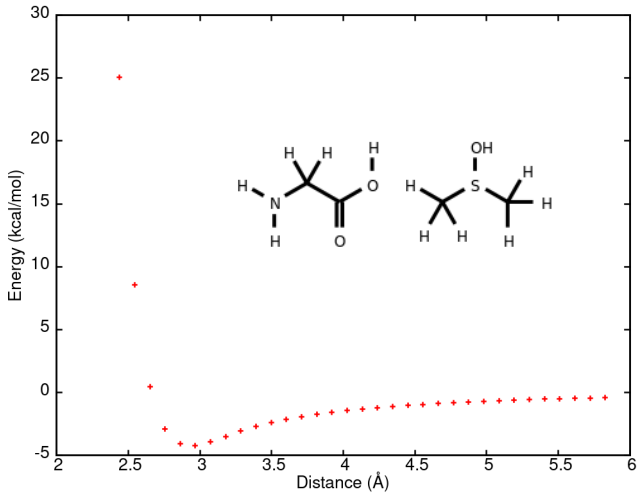


Figure 129: System 129



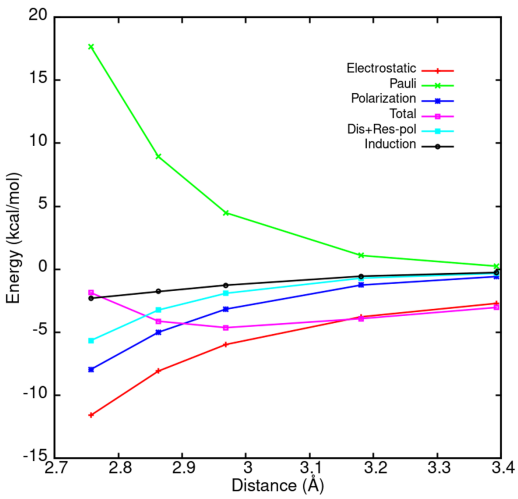
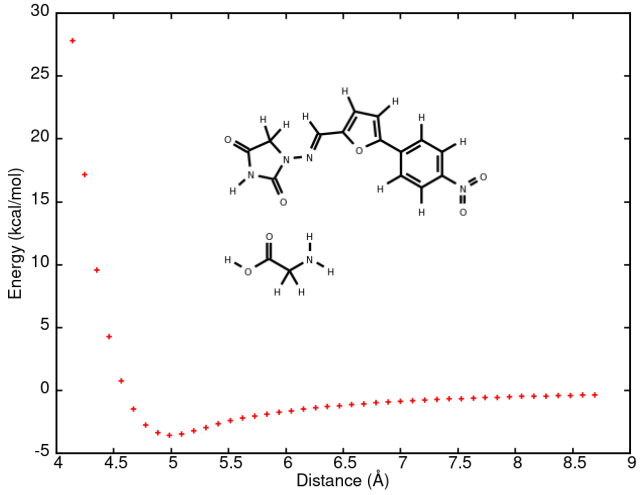


Figure 130: System 130



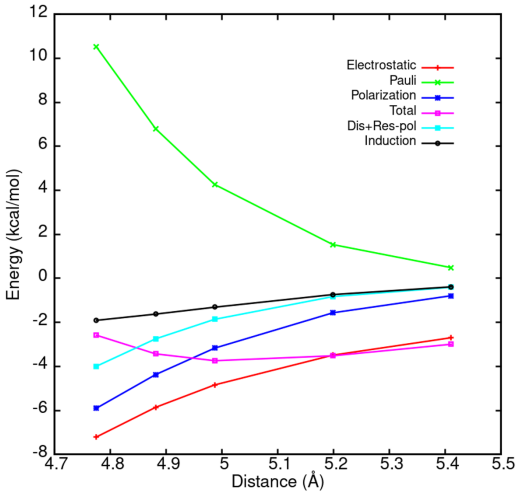
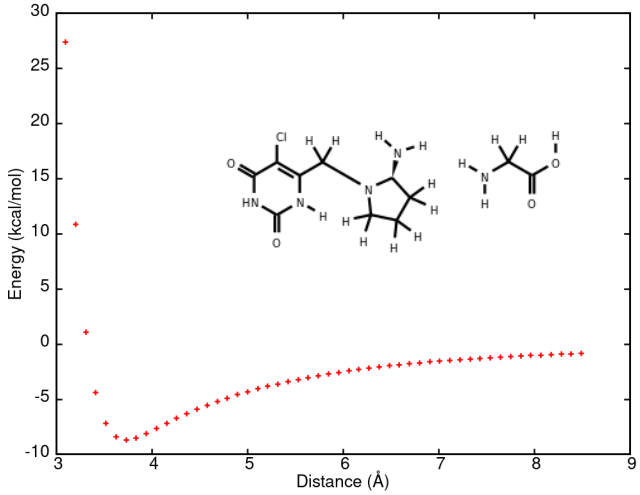


Figure 131: System 131



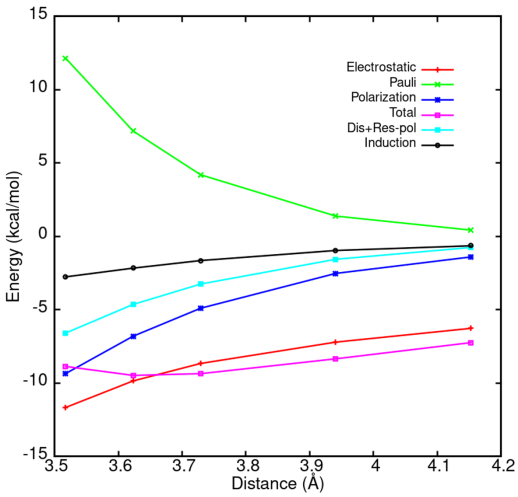
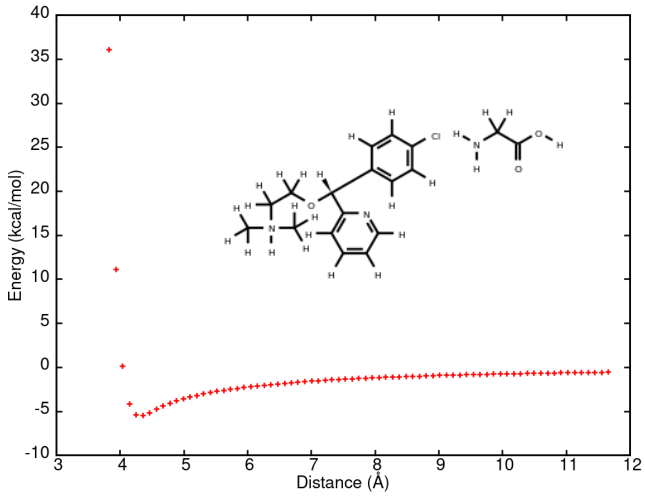


Figure 132: System 132



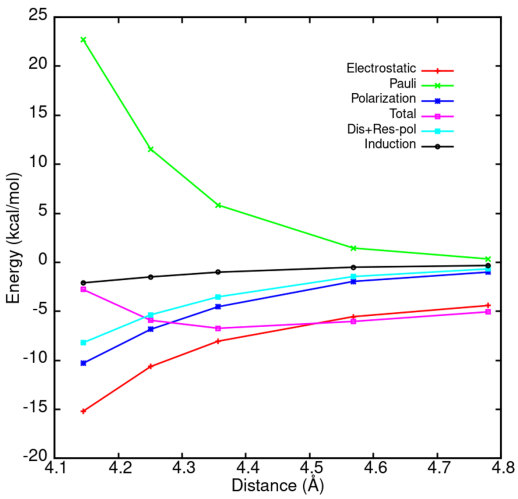
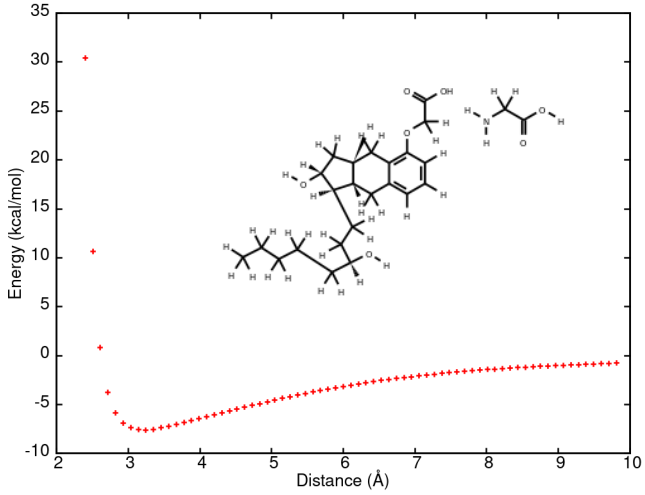


Figure 133: System 133



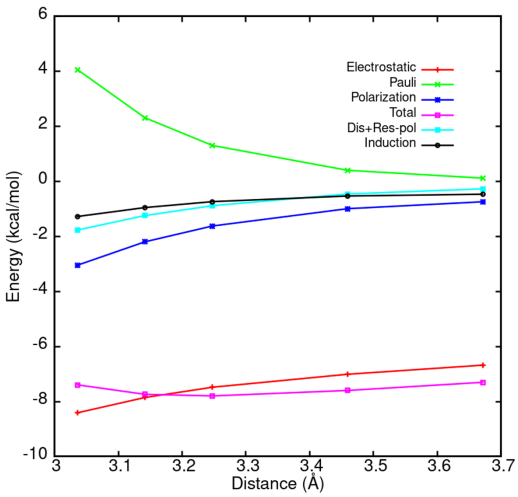
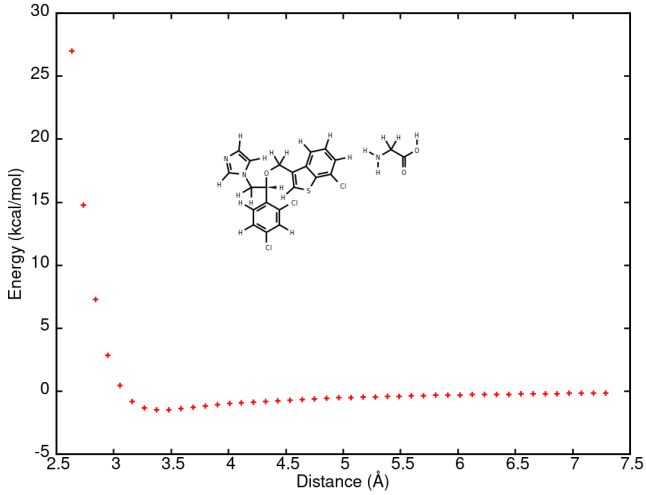


Figure 134: System 134



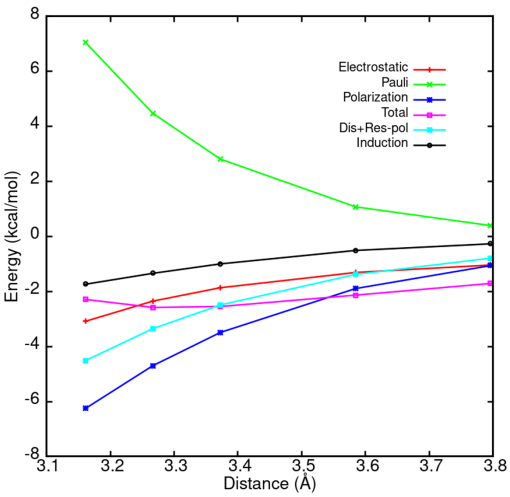
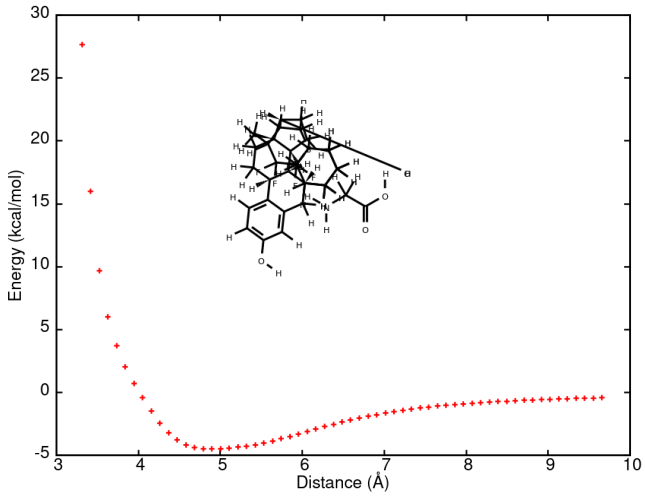


Figure 135: System 135



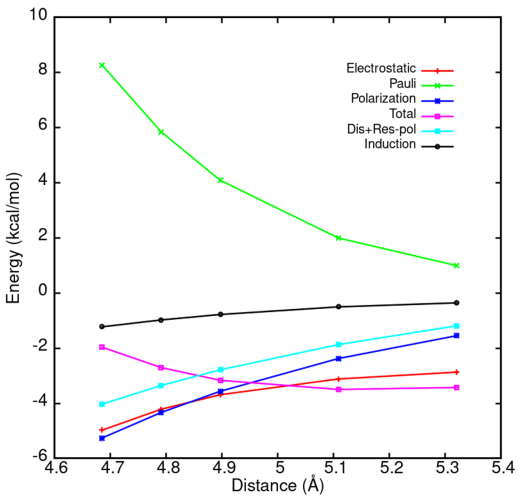
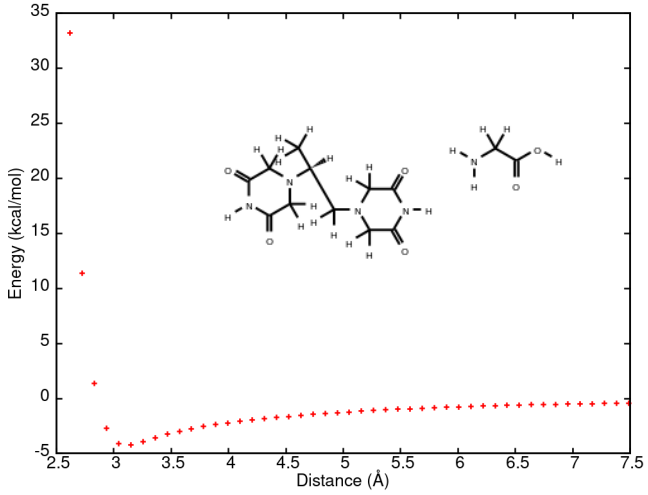


Figure 136: System 136



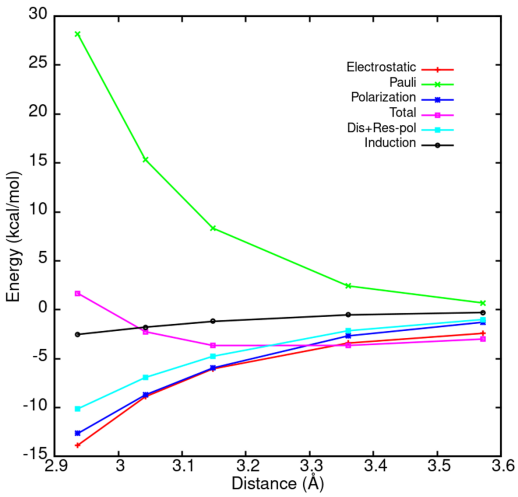
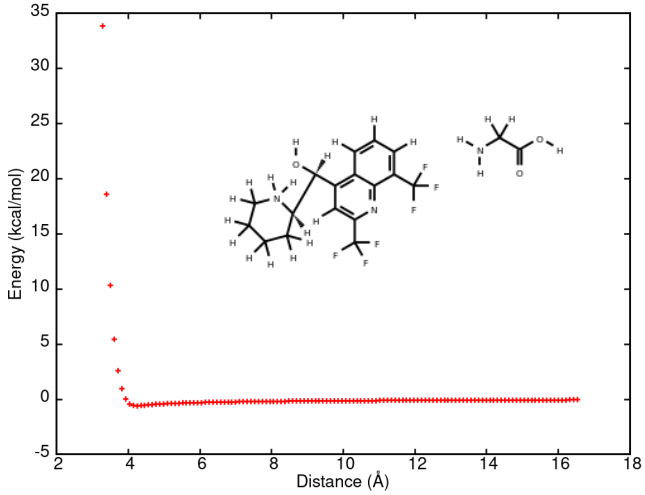


Figure 137: System 137



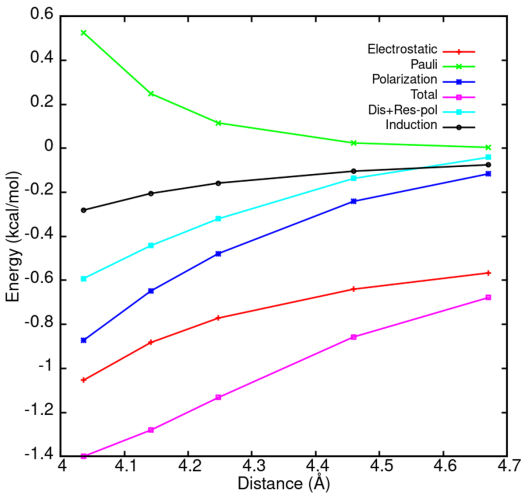


Figure 138: System 138

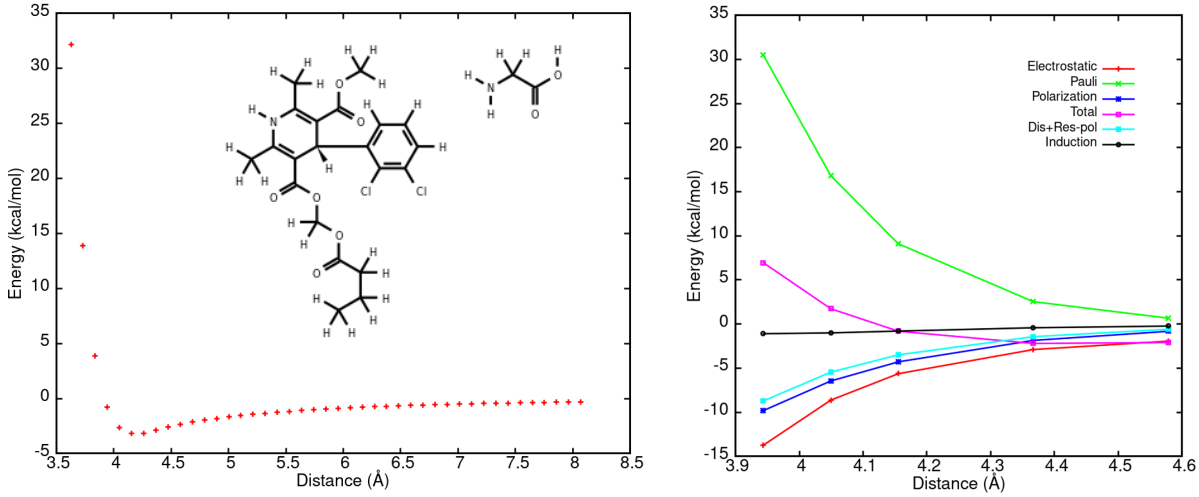


Figure 139: System 139

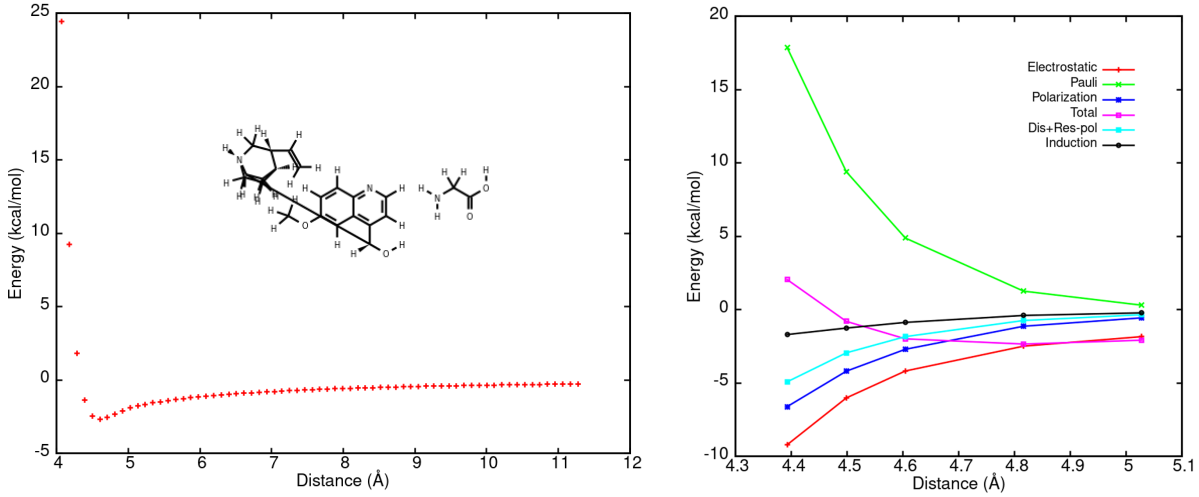


Figure 140: System 140

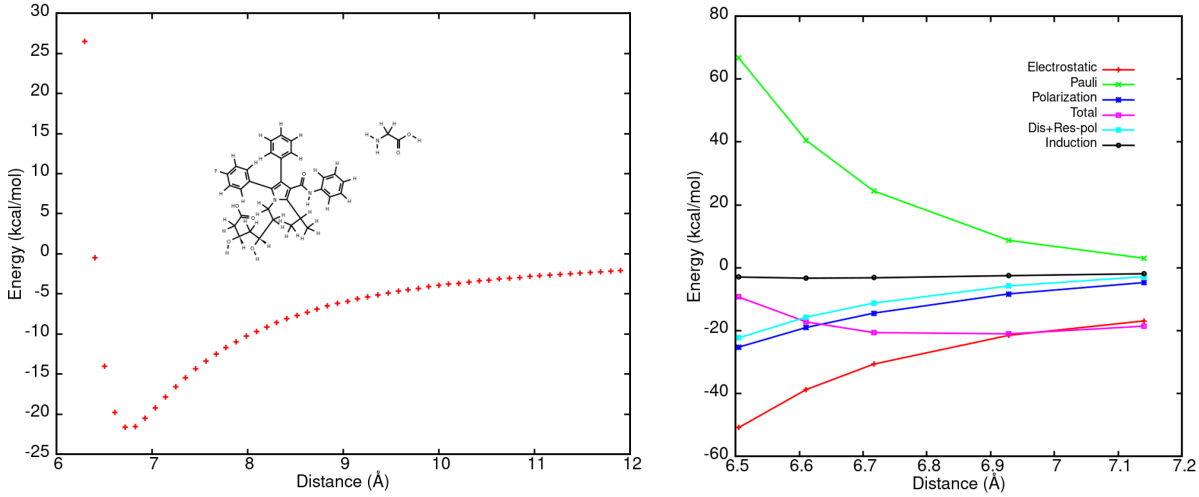


Figure 141: System 141

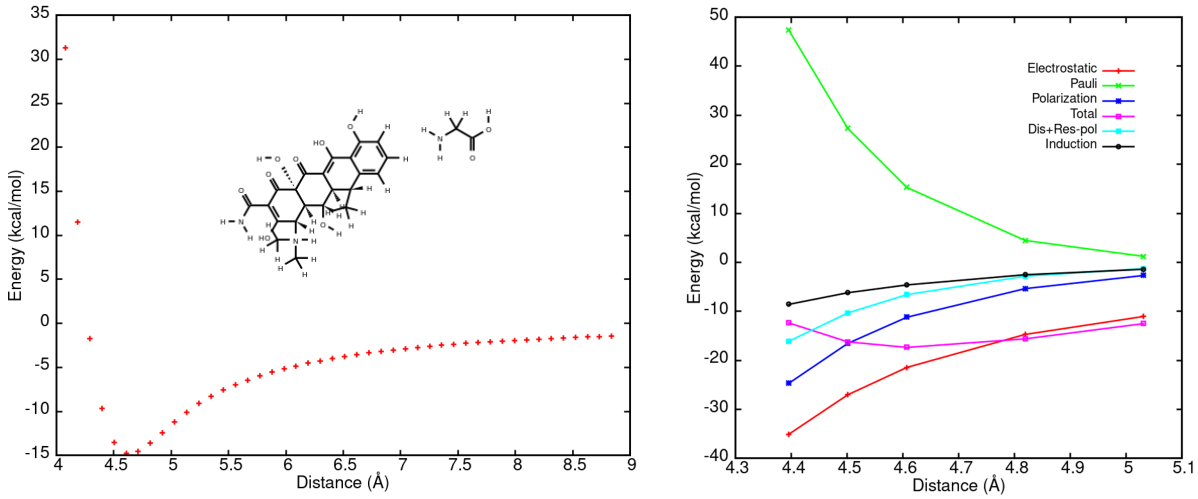


Figure 142: System 142

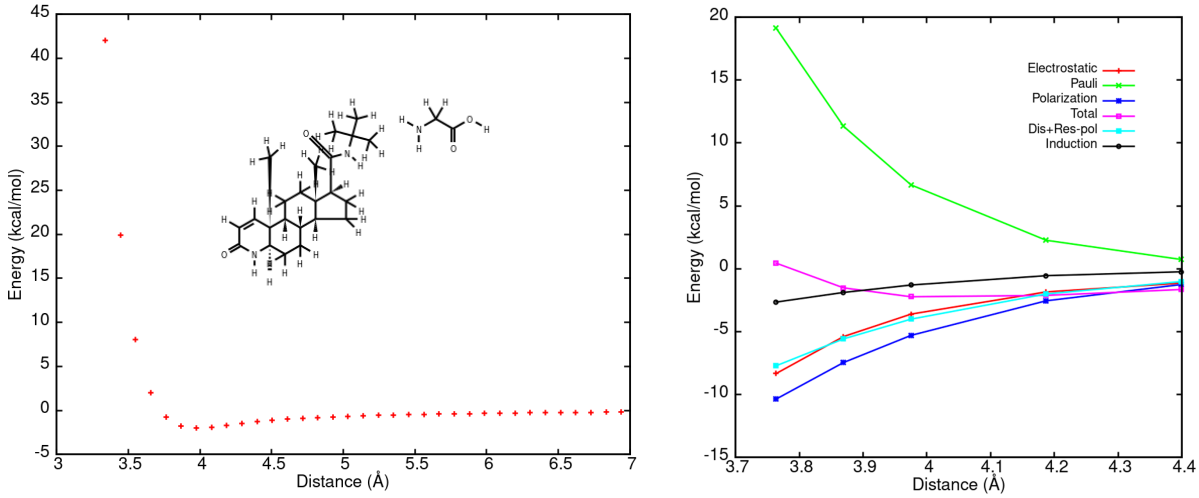


Figure 143: System 143

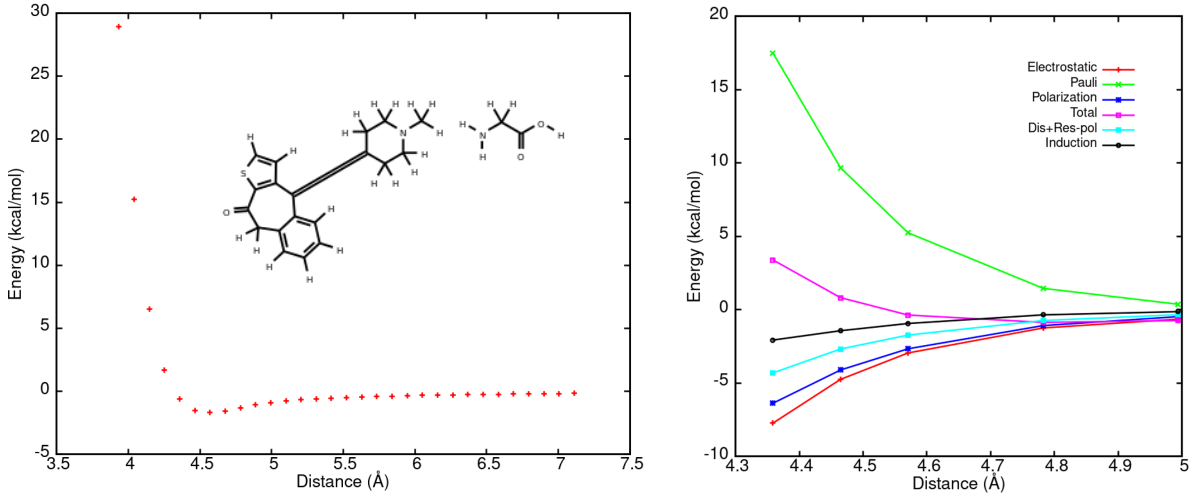
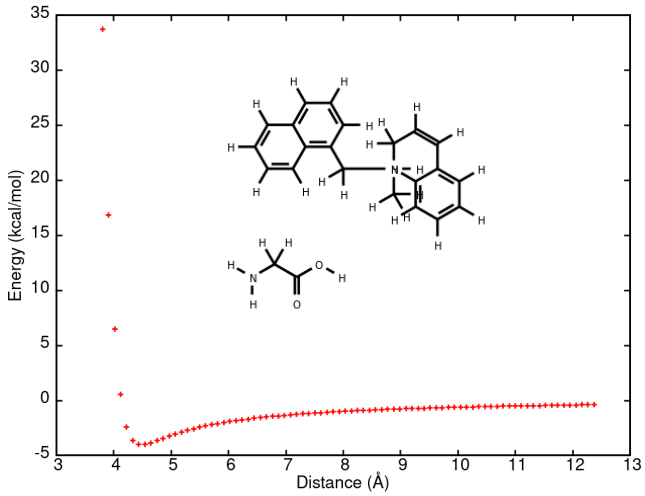


Figure 144: System 144



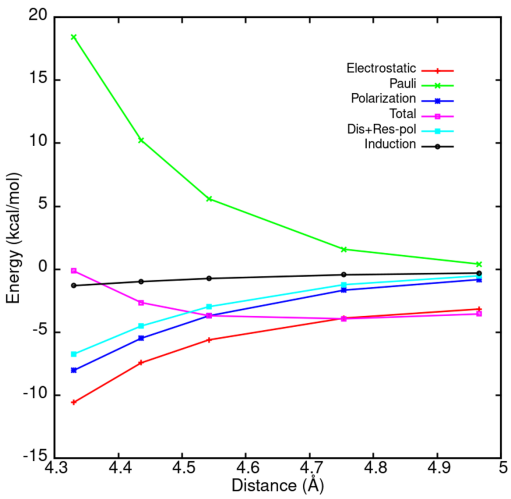
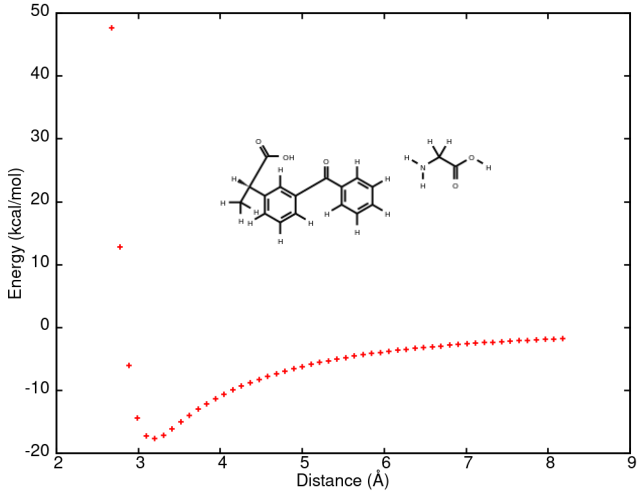


Figure 145: System 145



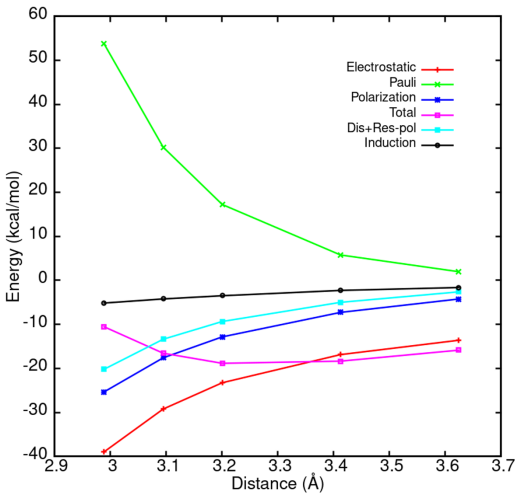
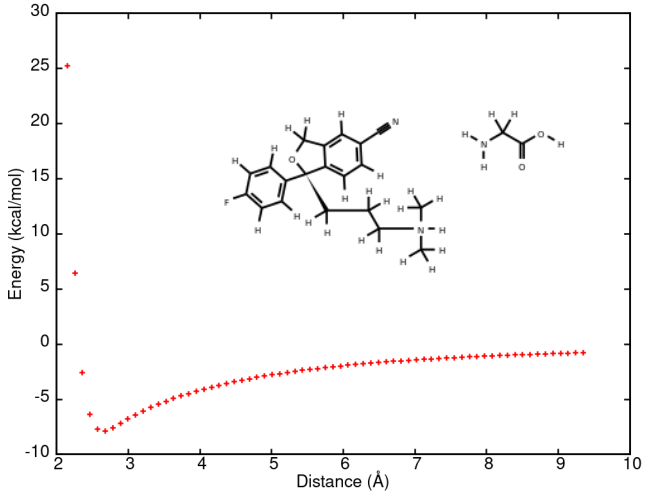


Figure 146: System 146



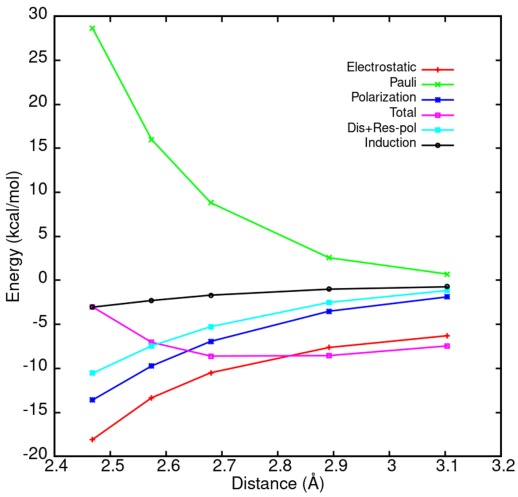
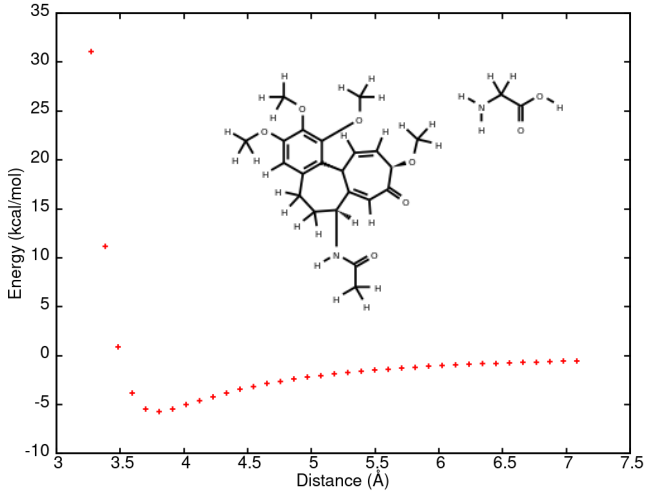


Figure 147: System 147



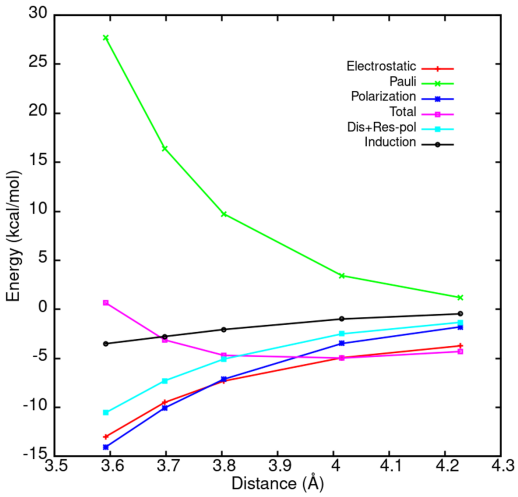
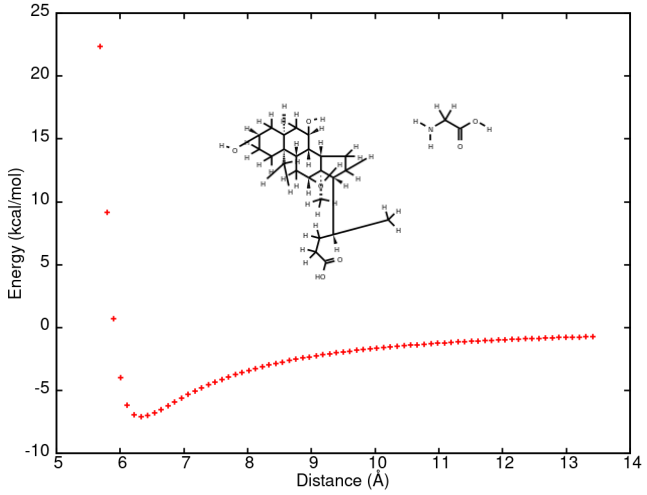


Figure 148: System 148



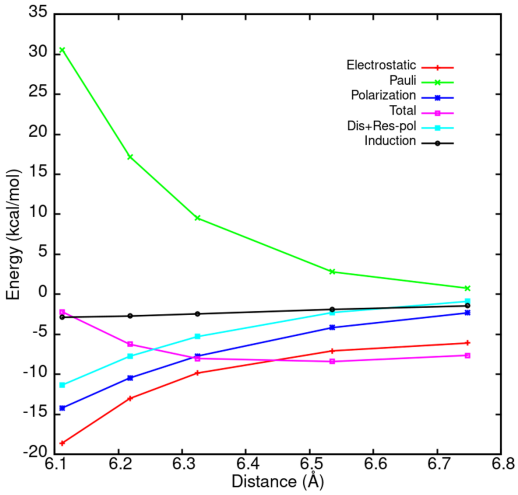
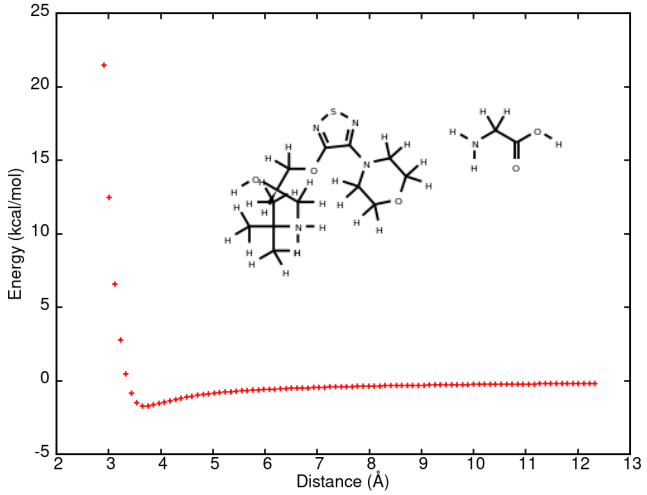


Figure 149: System 149



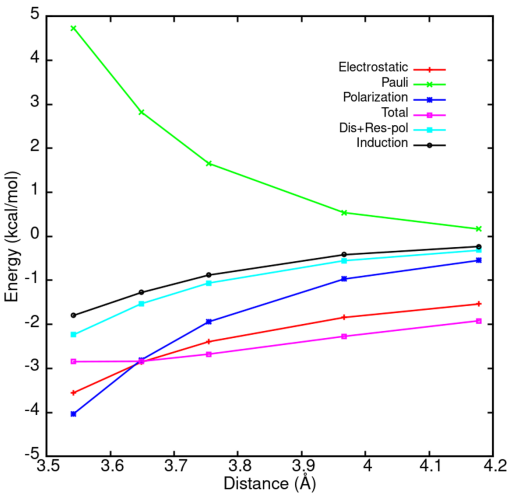
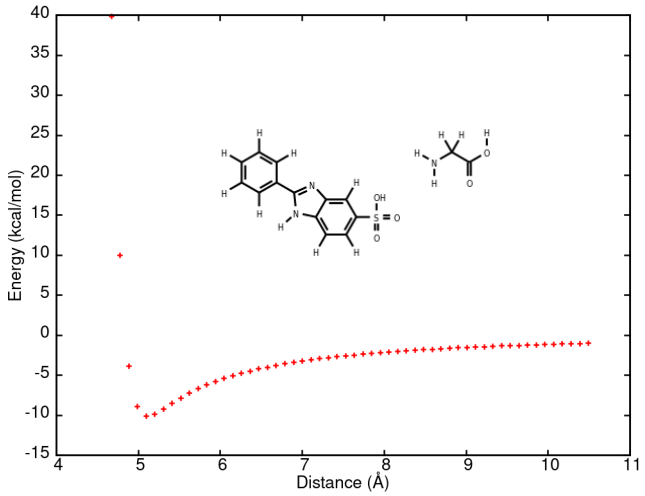


Figure 150: System 150



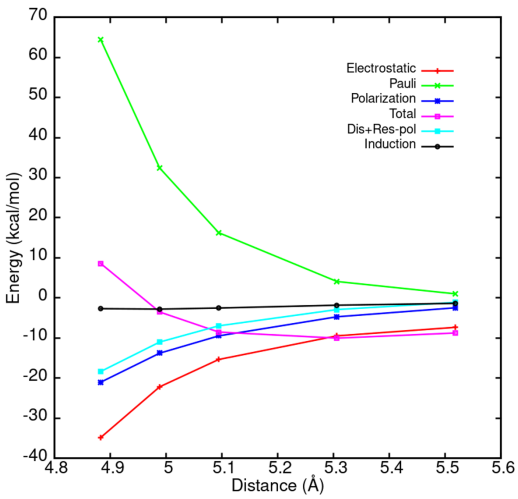
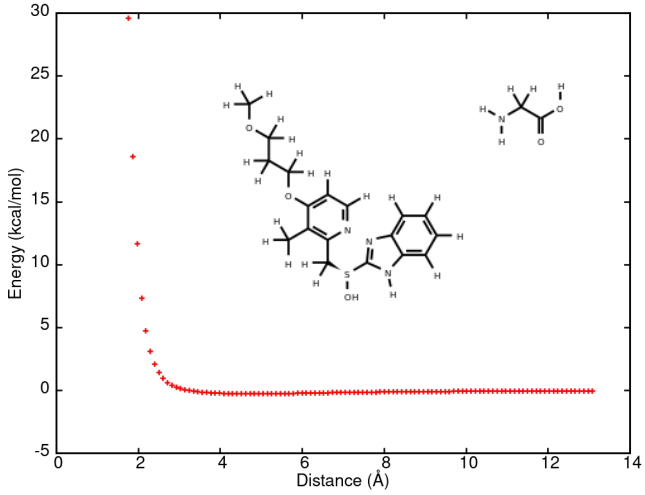


Figure 151: System 151



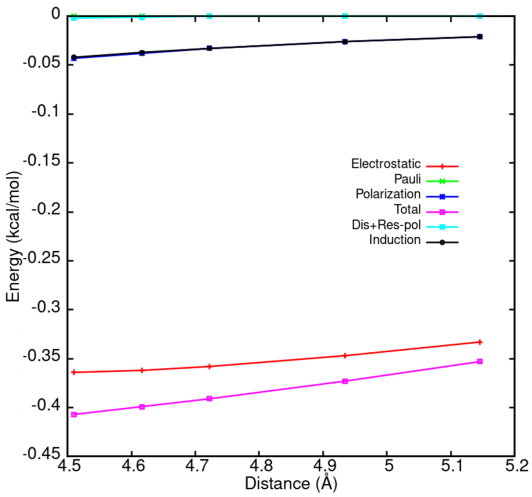
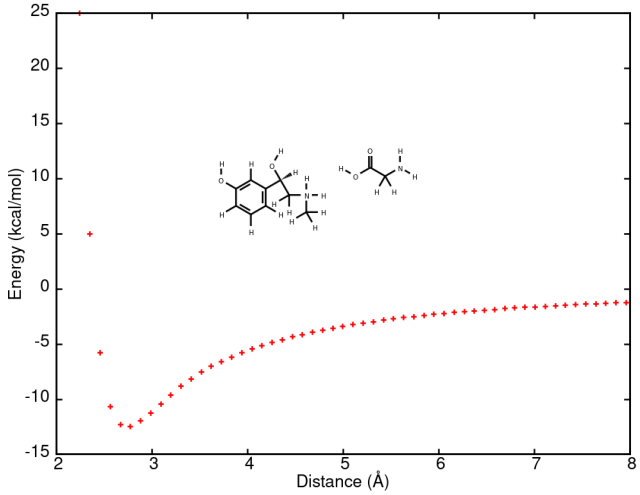


Figure 152: System 152



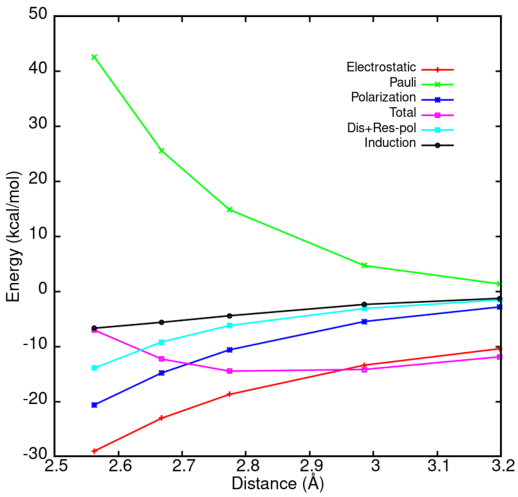
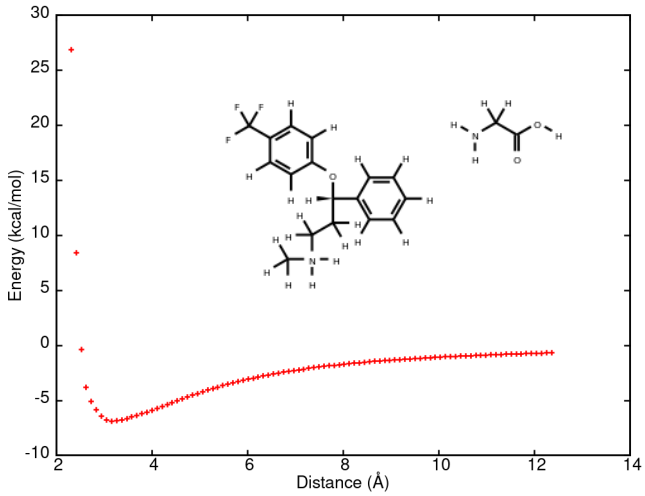


Figure 153: System 153



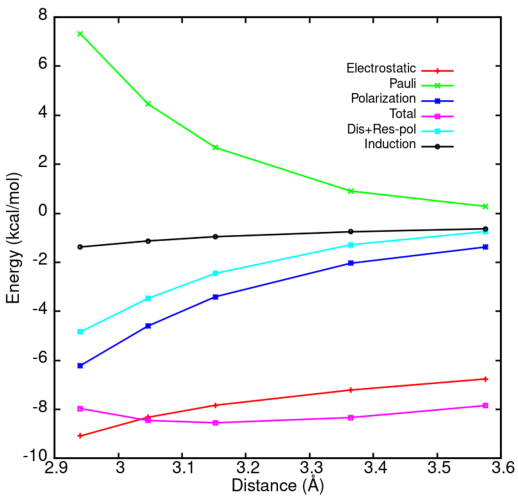
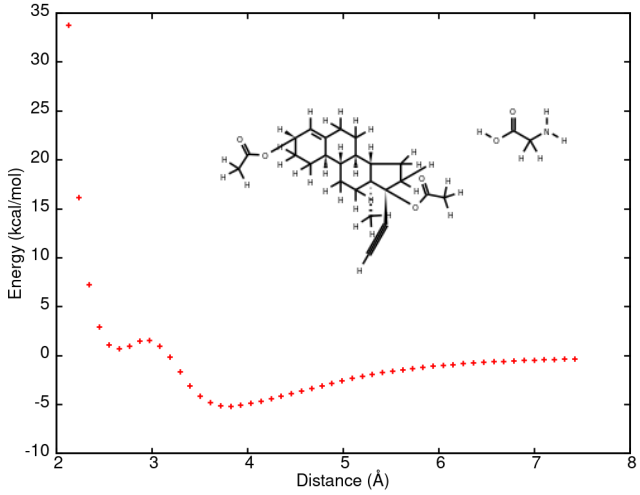


Figure 154: System 154



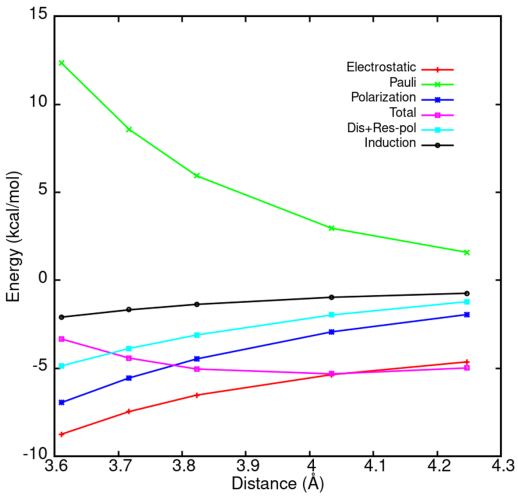
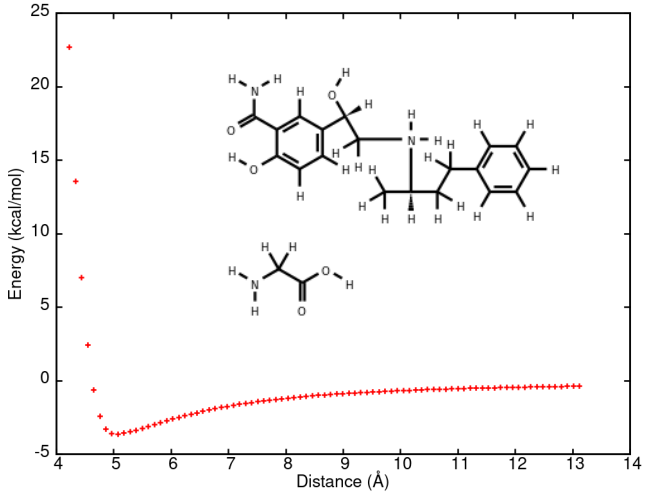


Figure 155: System 155



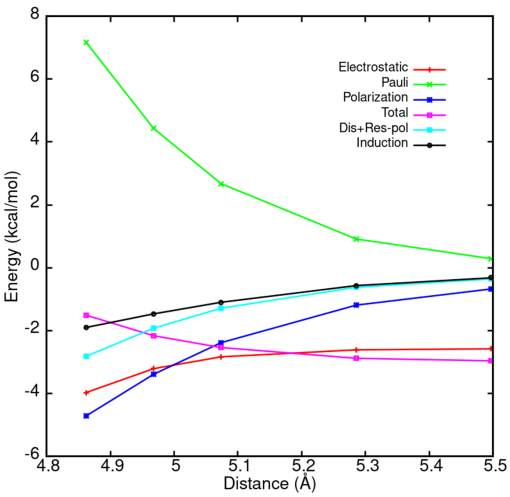
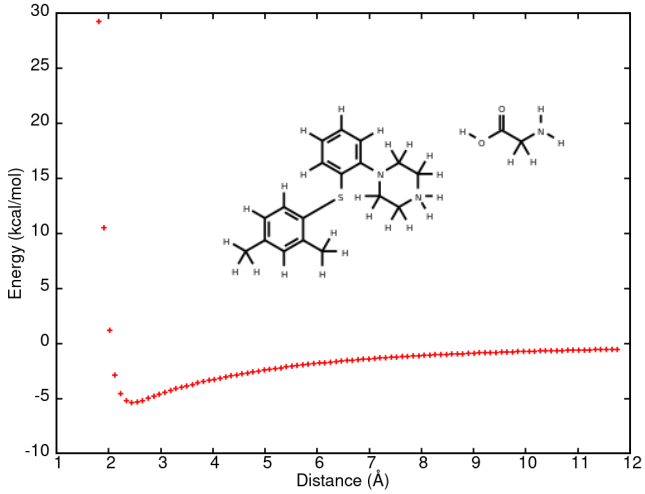


Figure 156: System 156



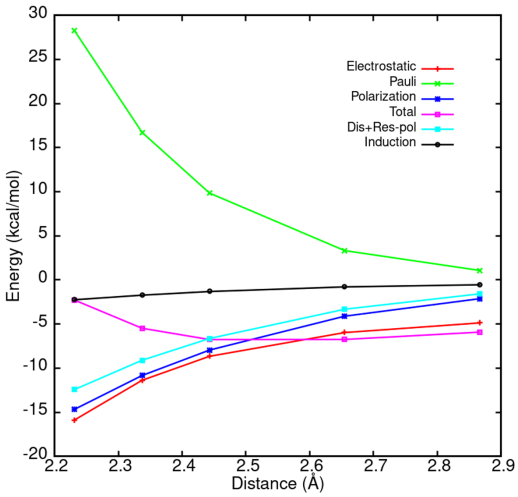
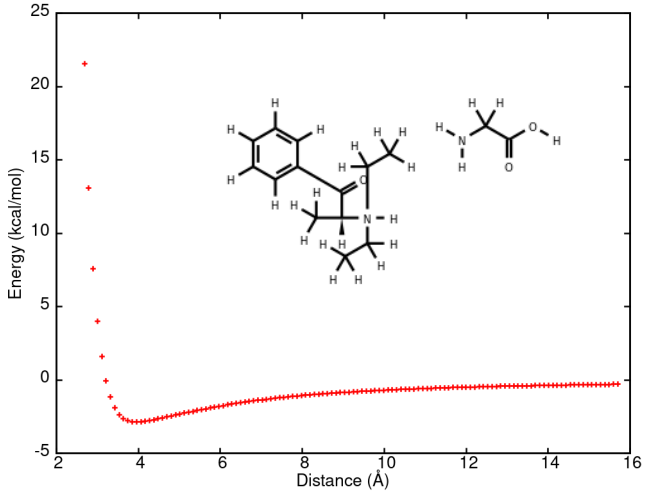


Figure 157: System 157



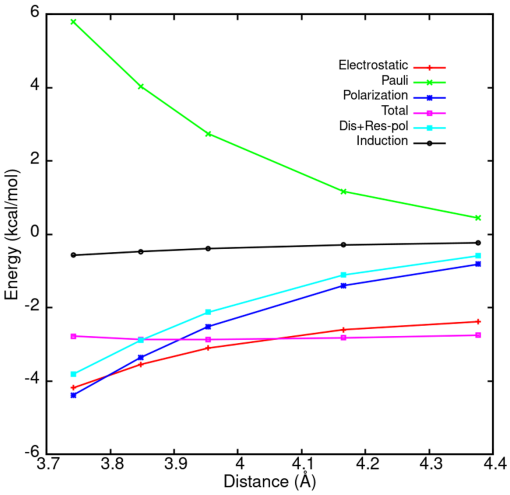
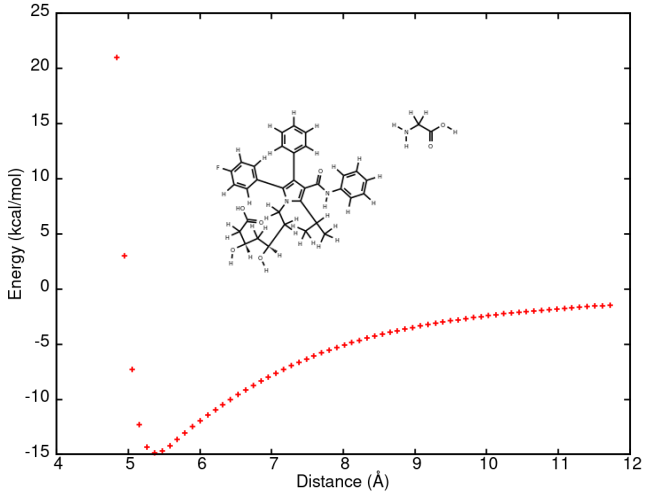


Figure 158: System 158



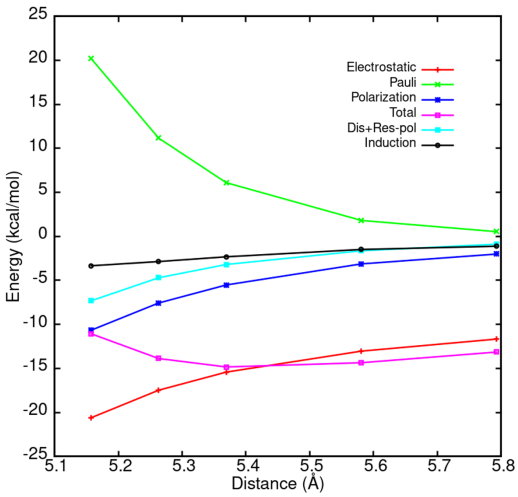
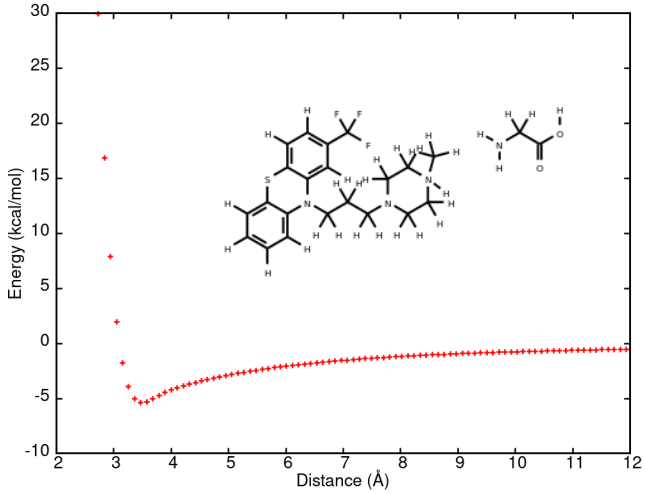


Figure 159: System 159



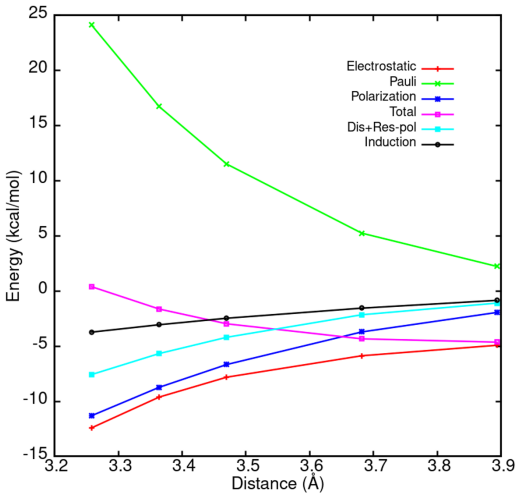
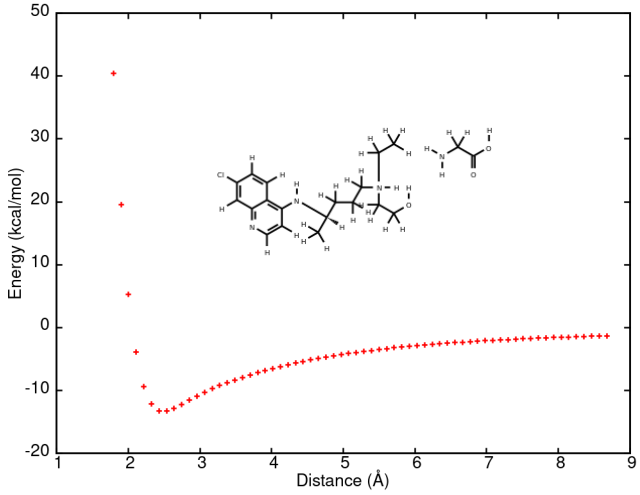


Figure 160: System 160



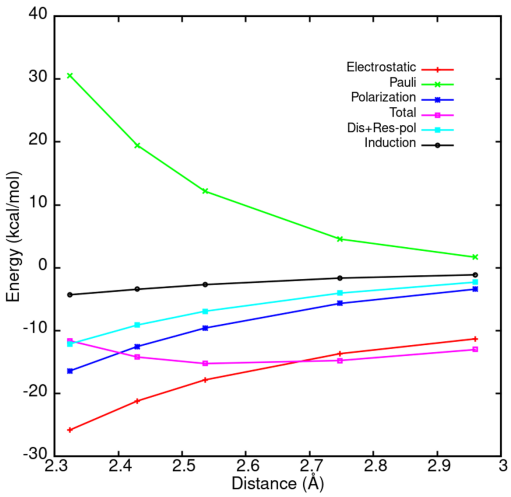
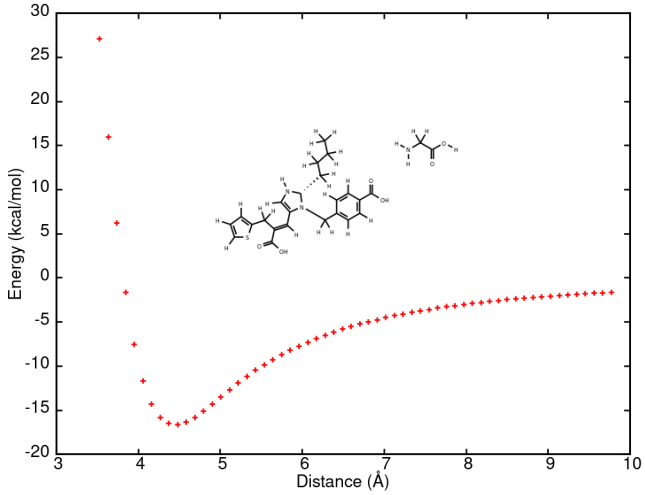


Figure 161: System 161



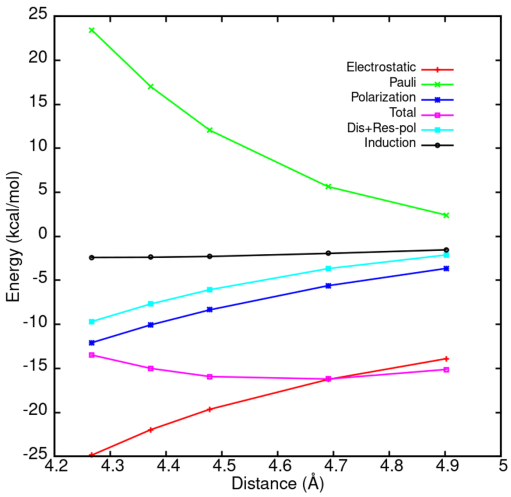


Figure 162: System 162

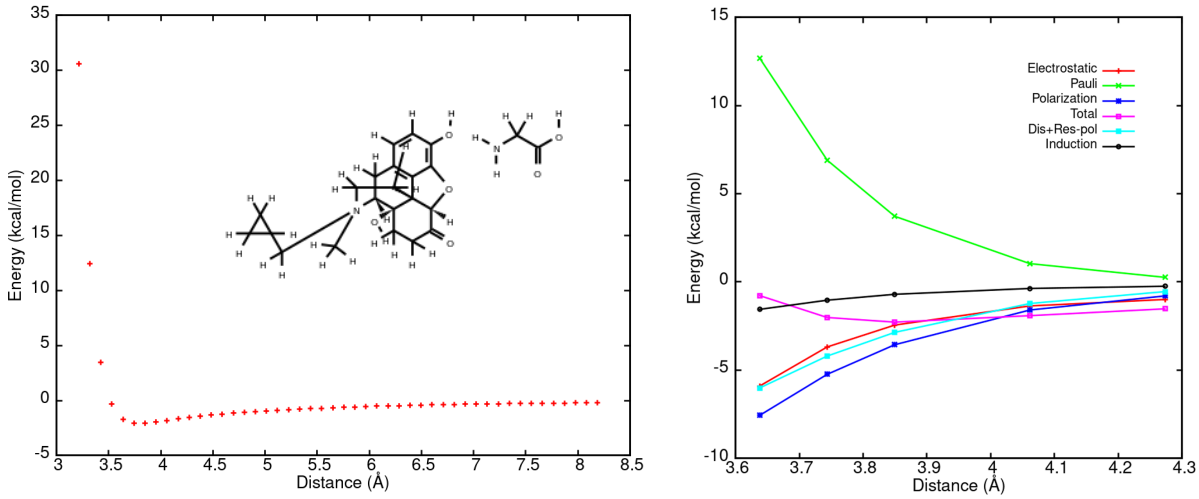


Figure 163: System 163

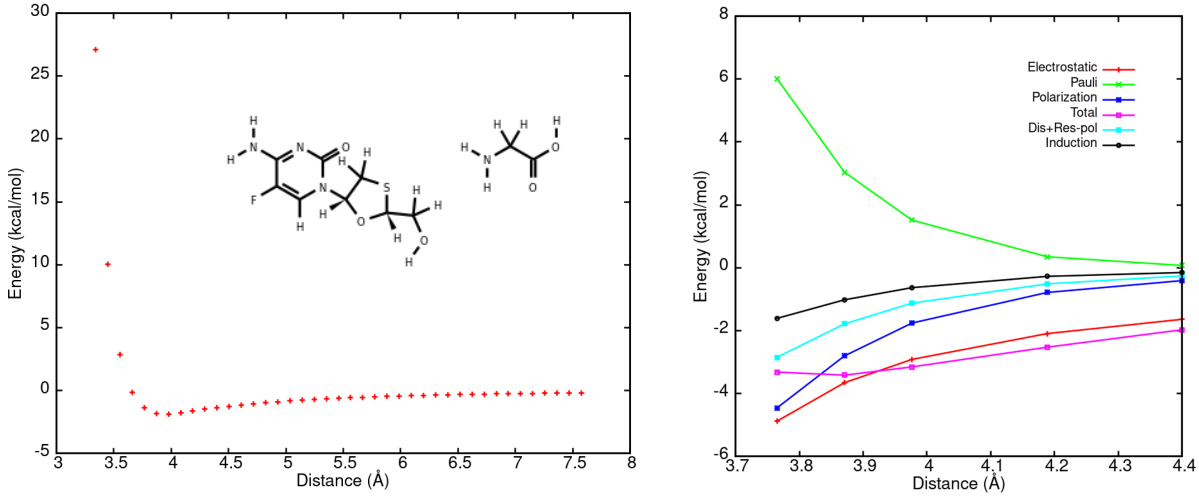


Figure 164: System 164

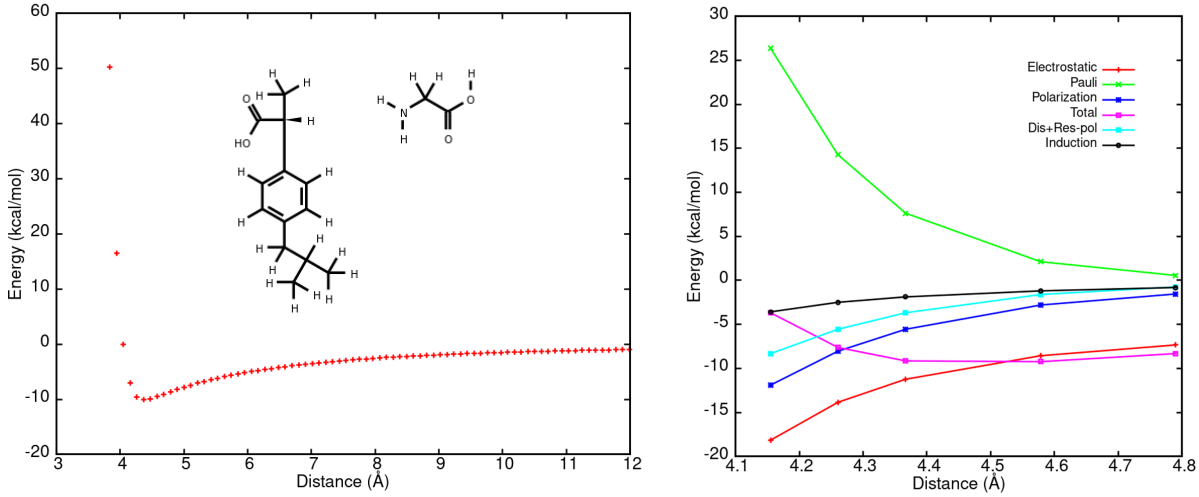
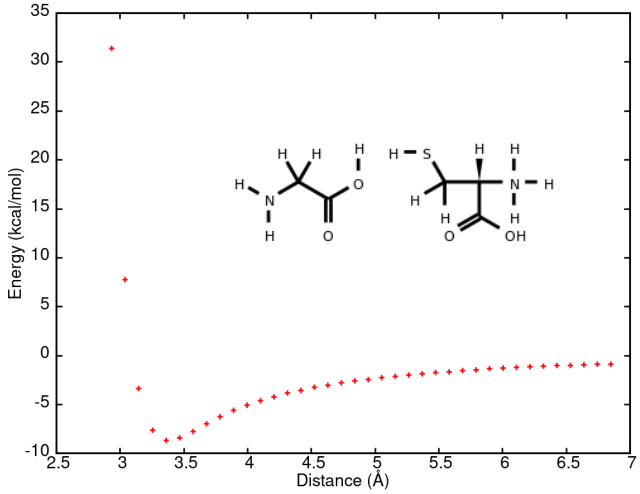


Figure 165: System 165



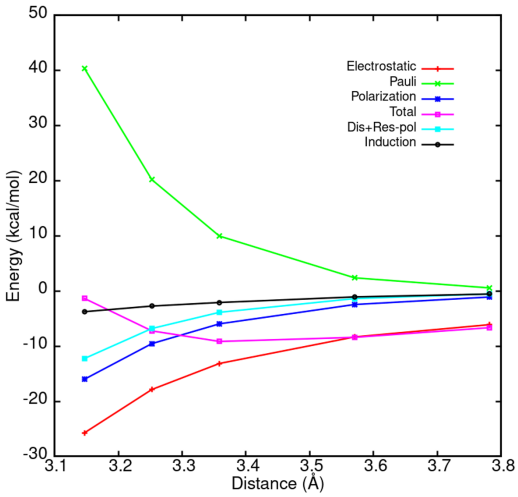
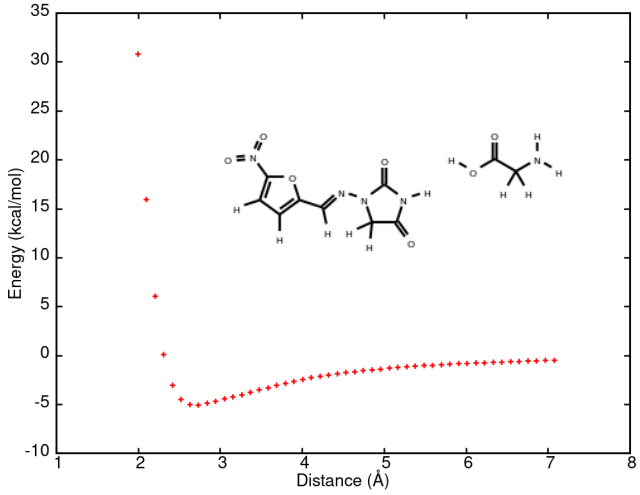


Figure 166: System 166



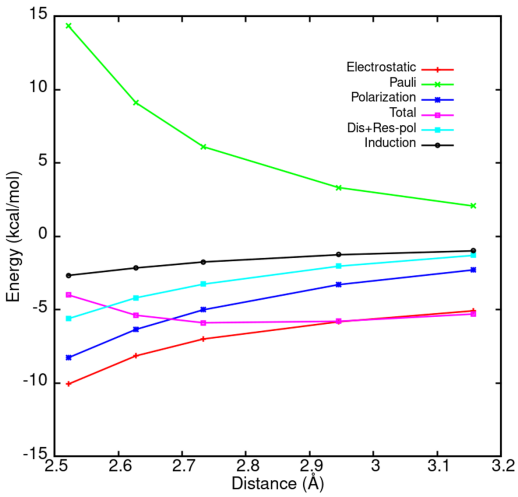
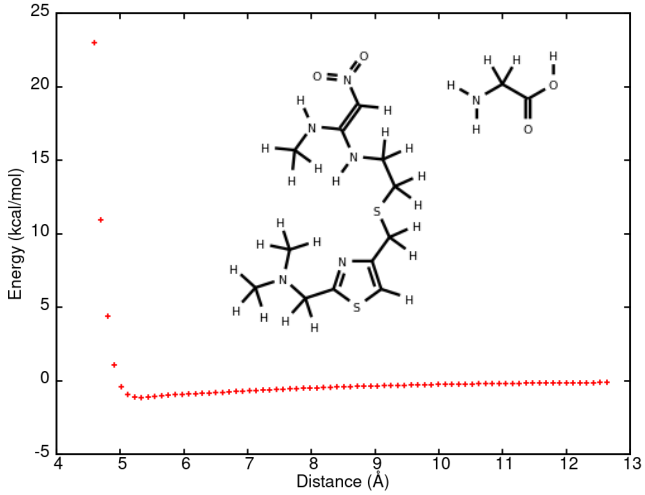


Figure 167: System 167



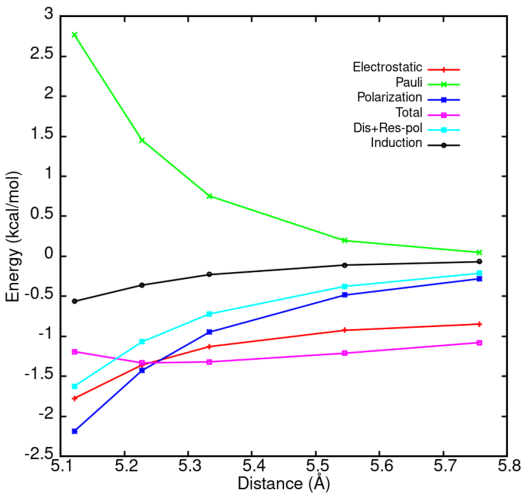


Figure 168: System 168

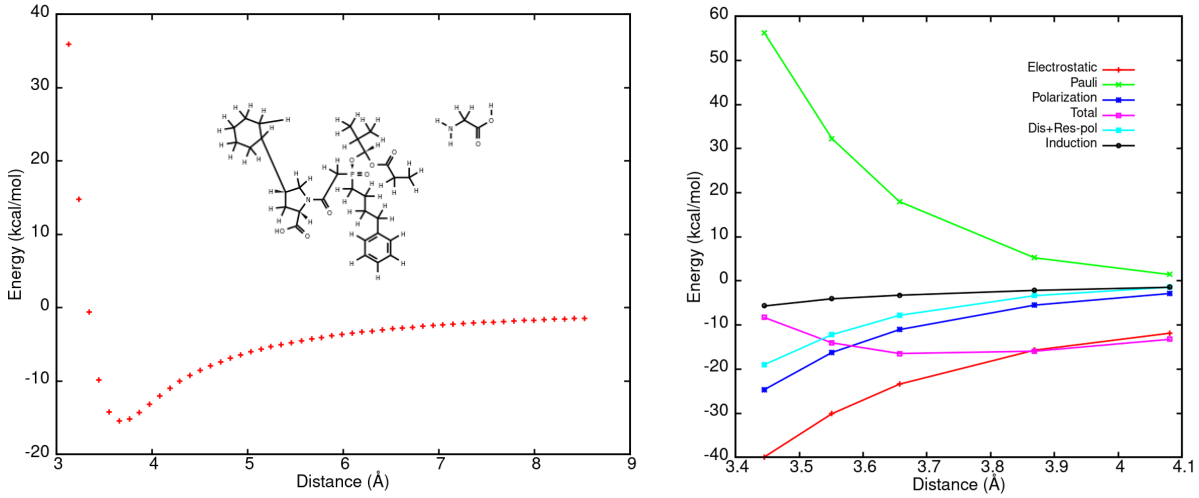


Figure 169: System 169

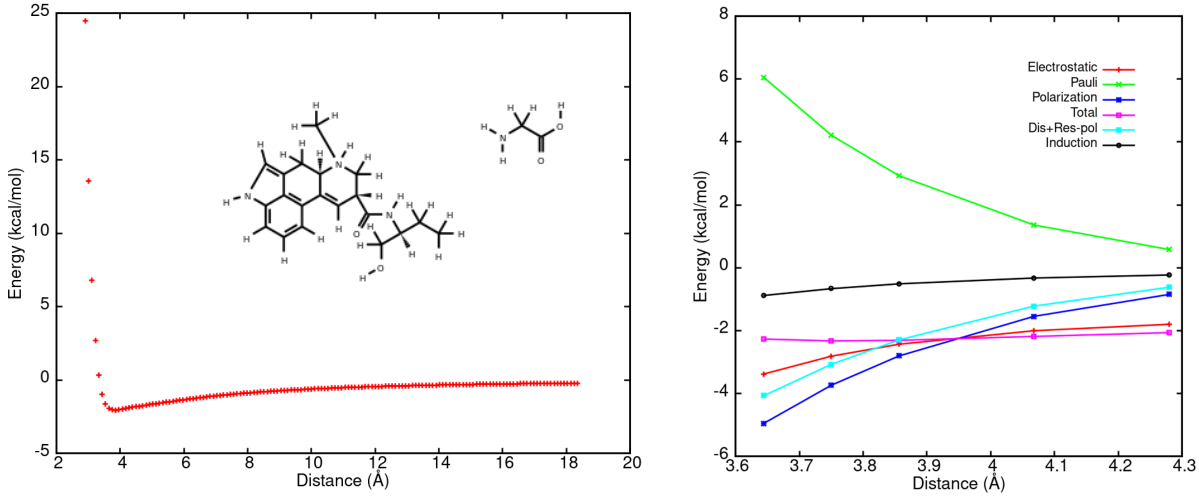


Figure 170: System 170

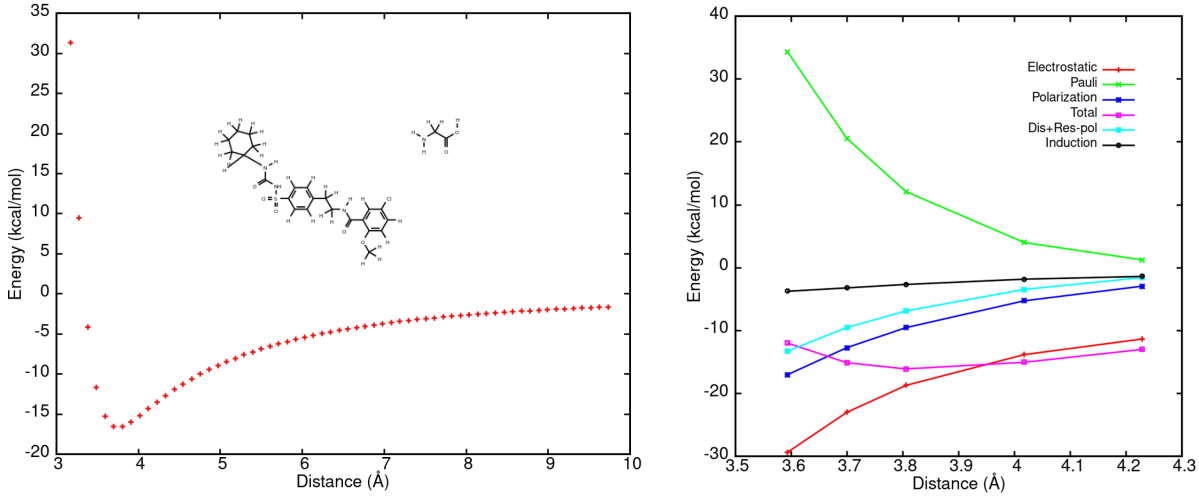
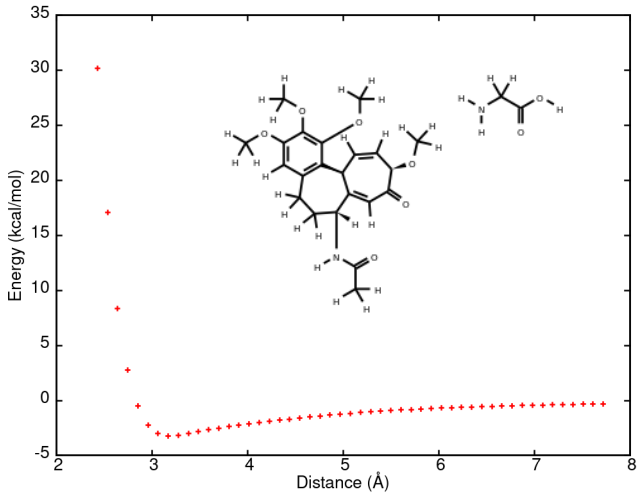


Figure 171: System 171



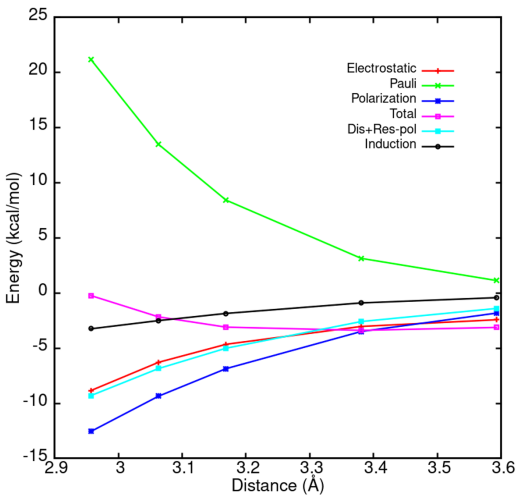
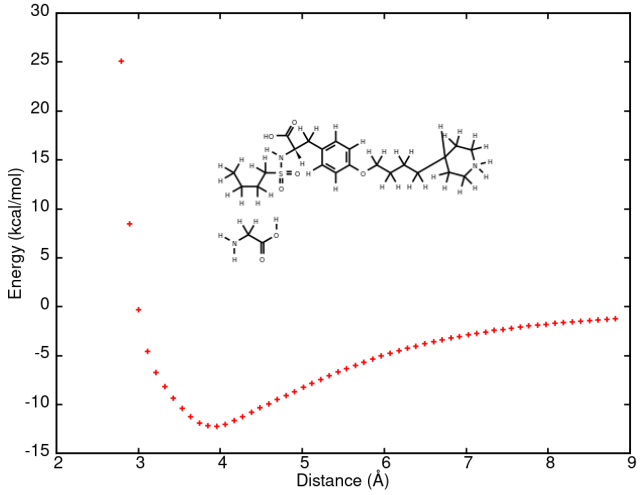


Figure 172: System 172



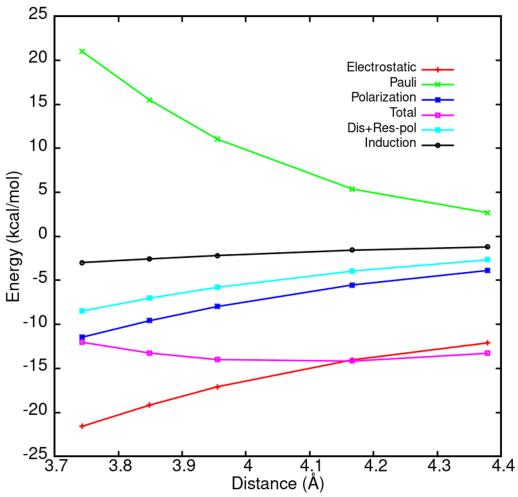
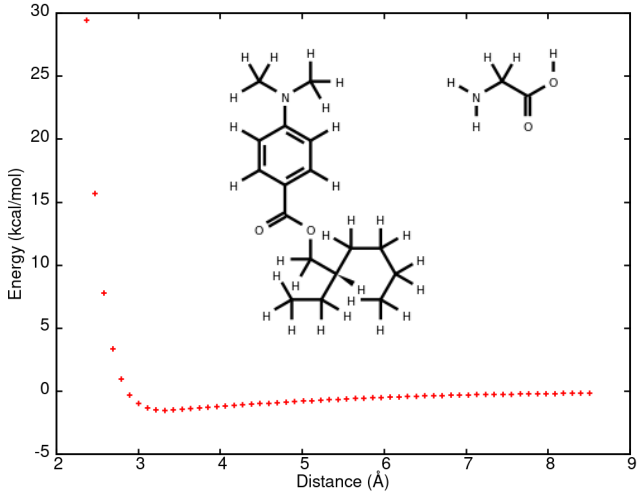


Figure 173: System 173



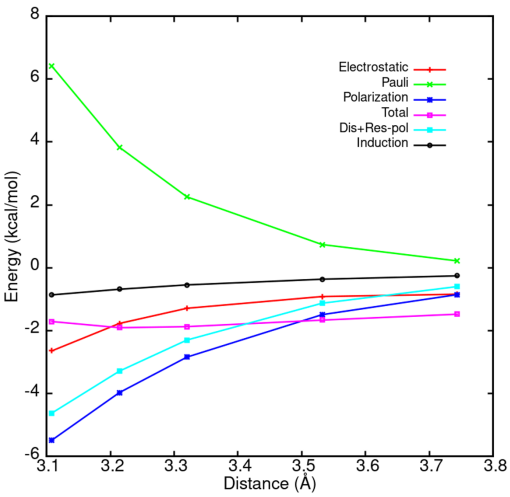
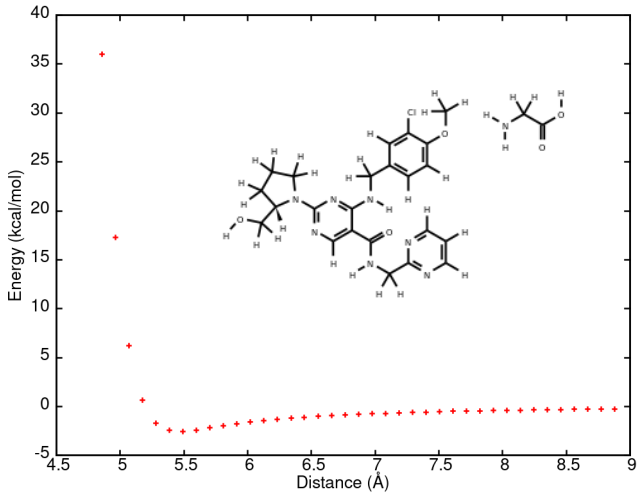


Figure 174: System 174



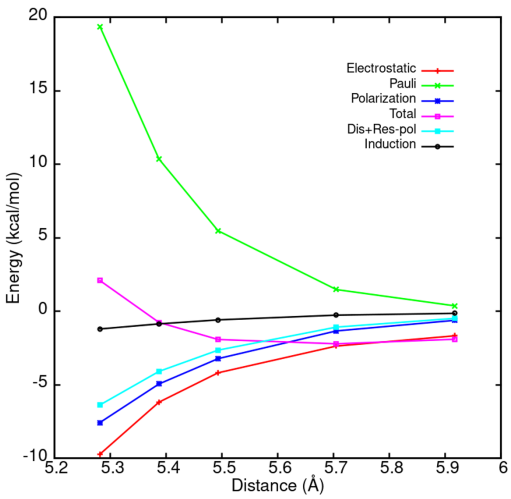
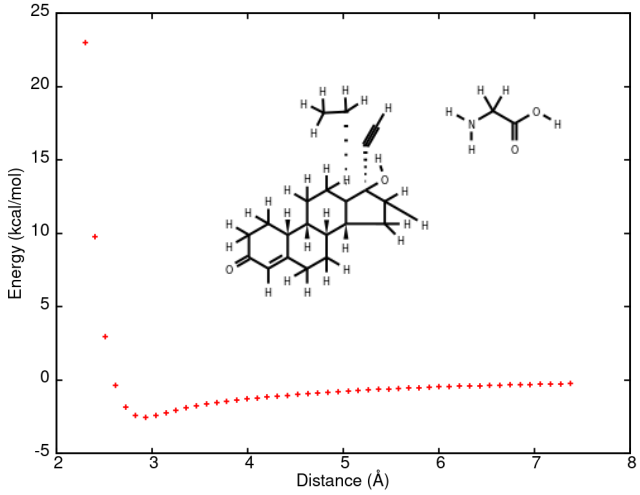


Figure 175: System 175



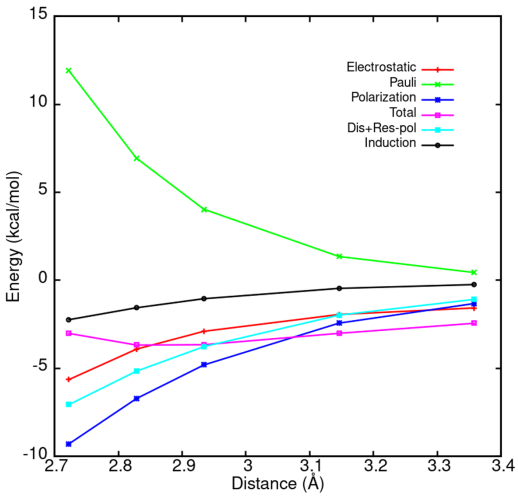
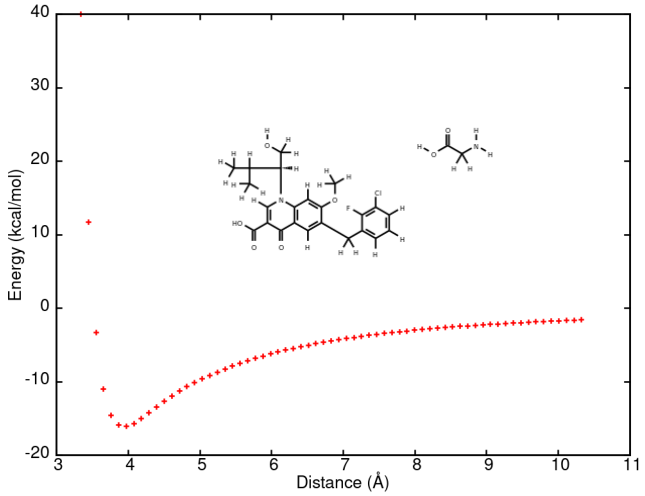


Figure 176: System 176



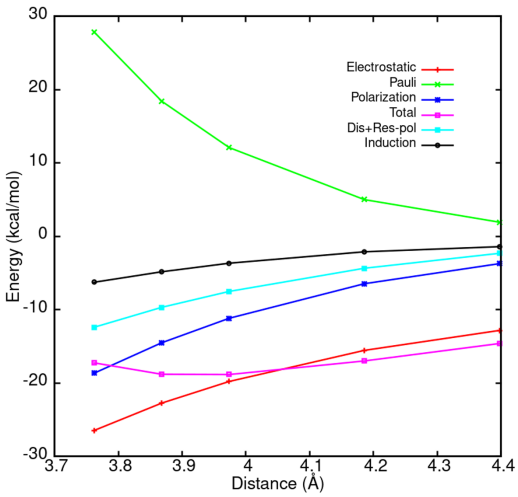
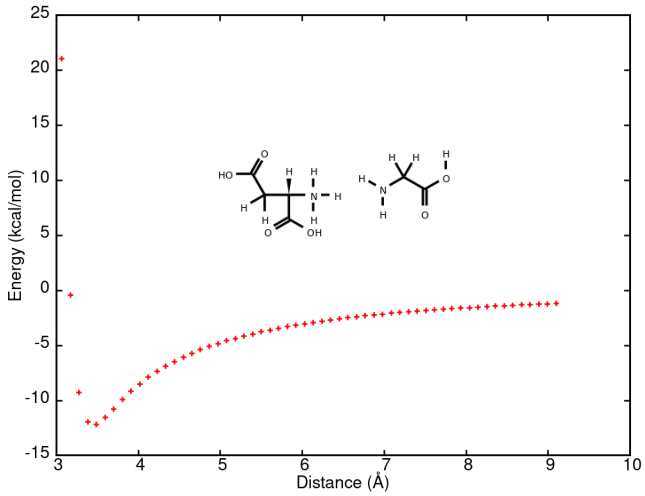


Figure 177: System 177



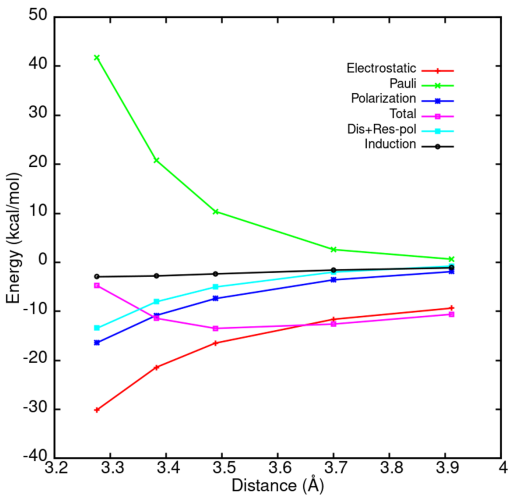
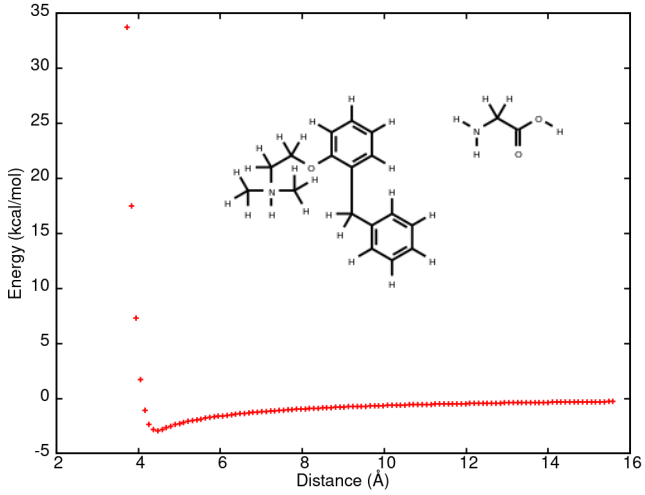


Figure 178: System 178



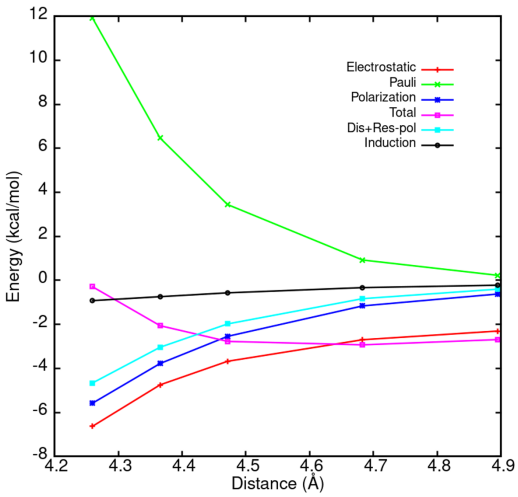
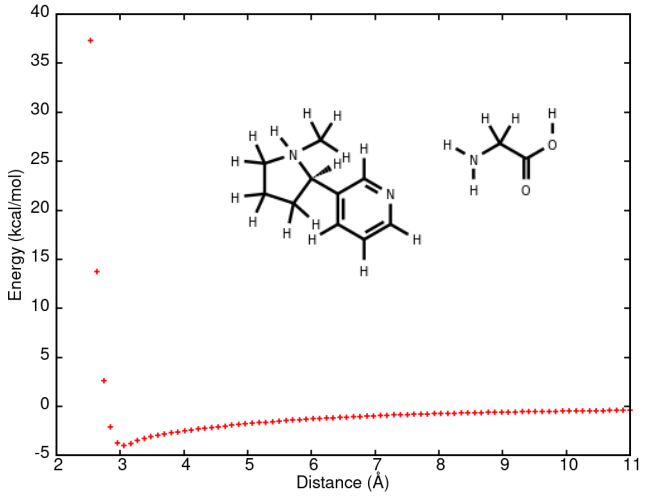


Figure 179: System 179



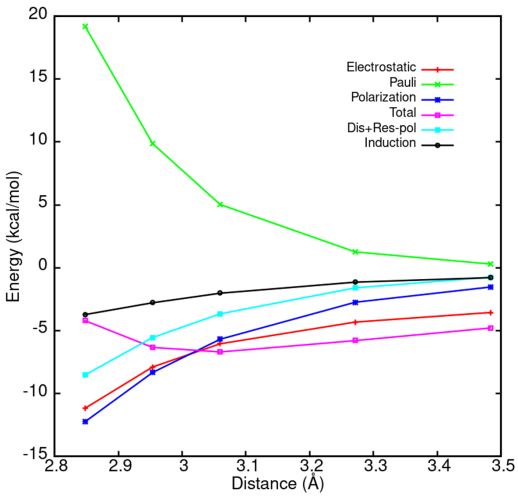
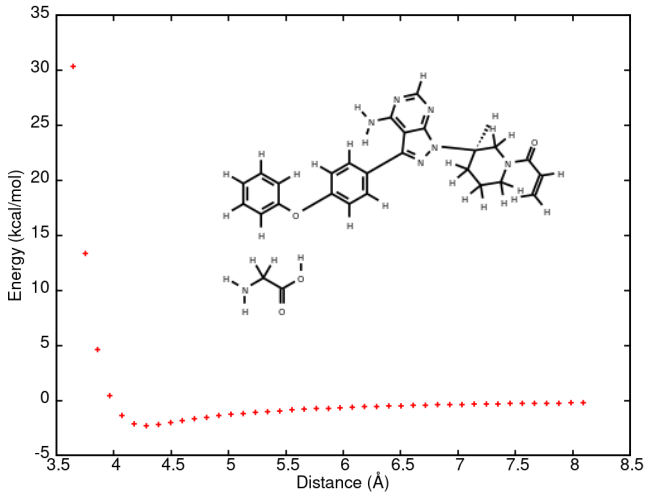


Figure 180: System 180



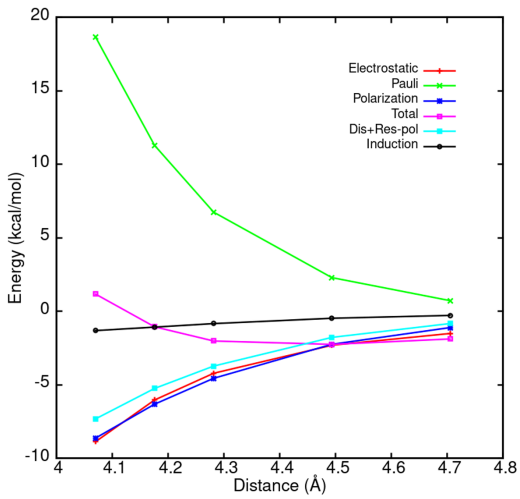
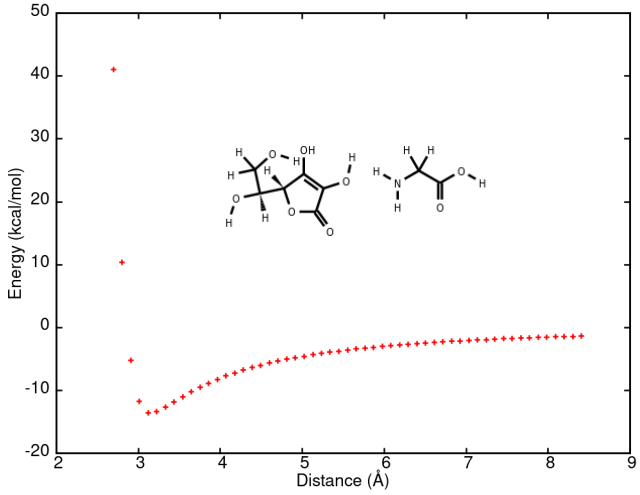


Figure 181: System 181



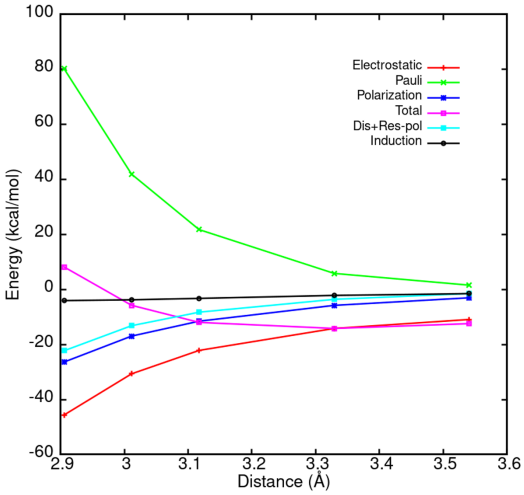
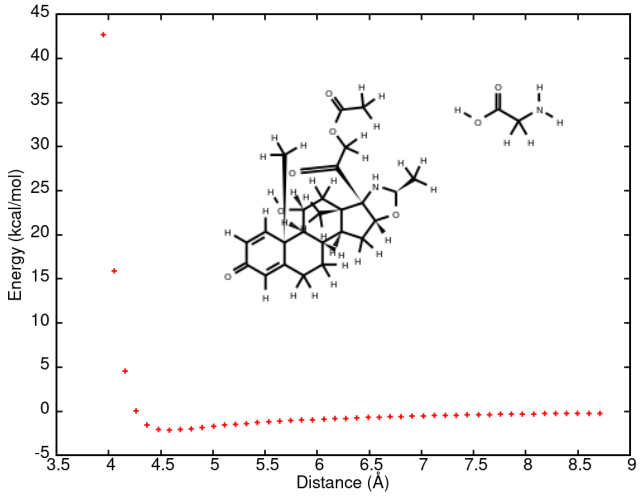


Figure 182: System 182



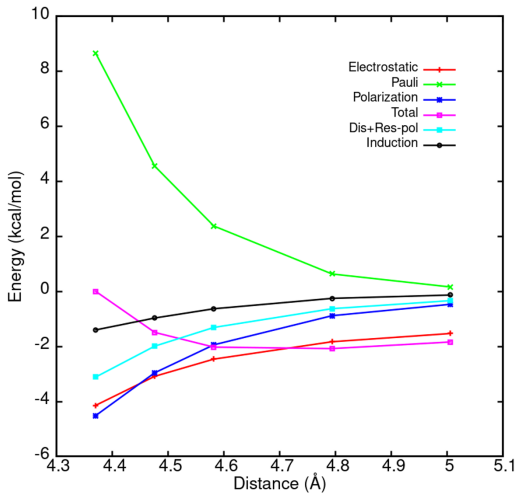


Figure 183: System 183

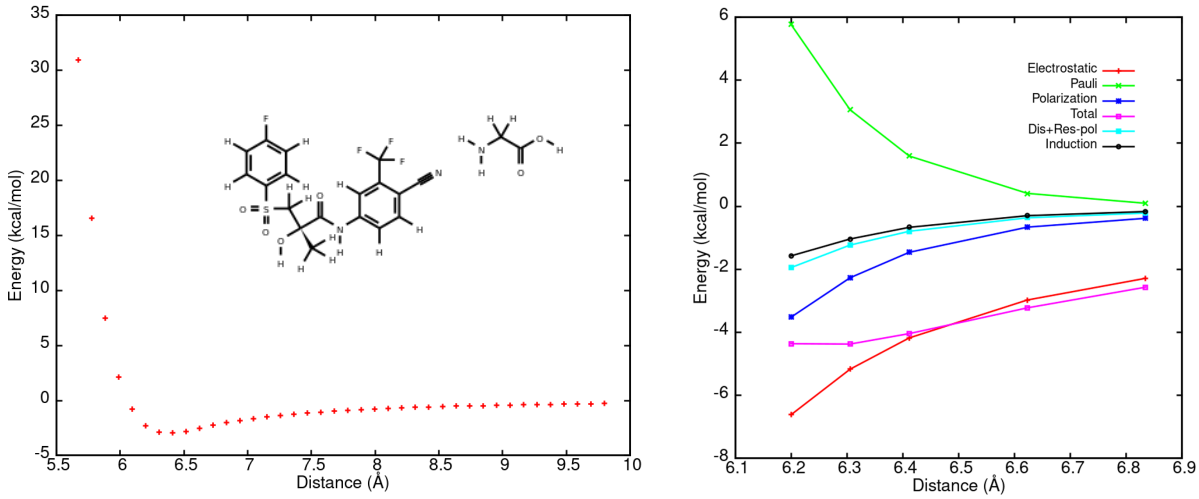


Figure 184: System 184

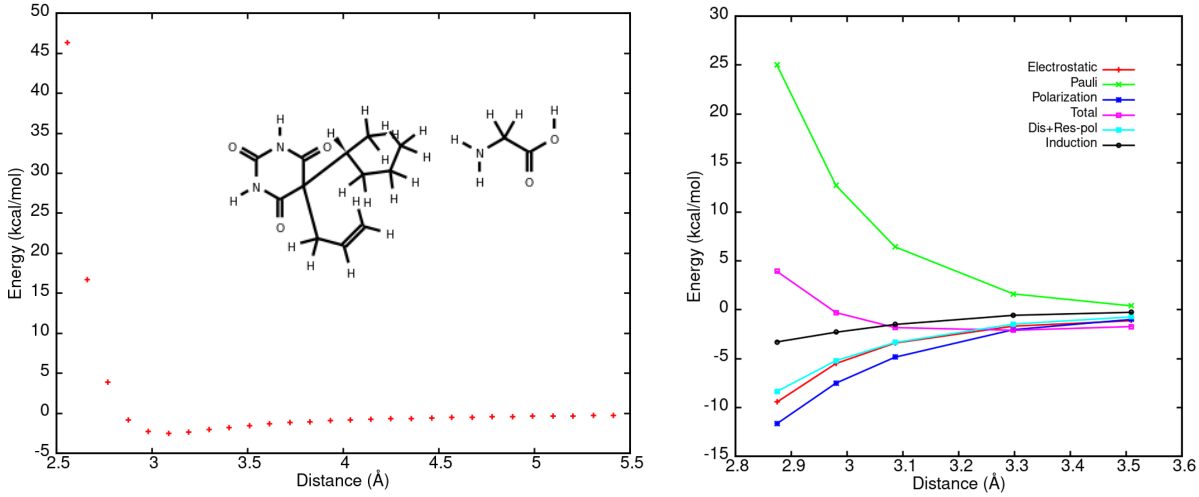


Figure 185: System 185

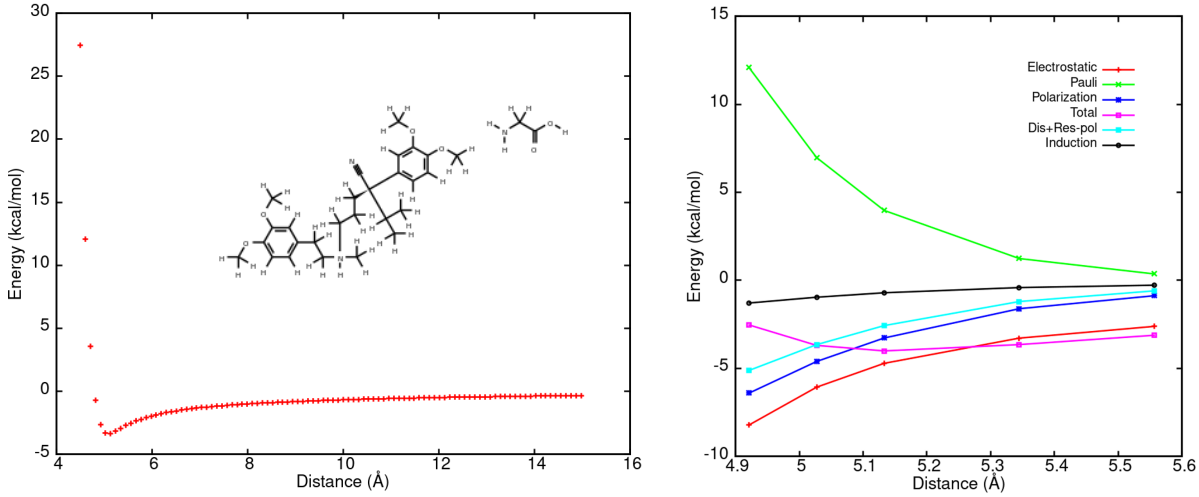


Figure 186: System 186

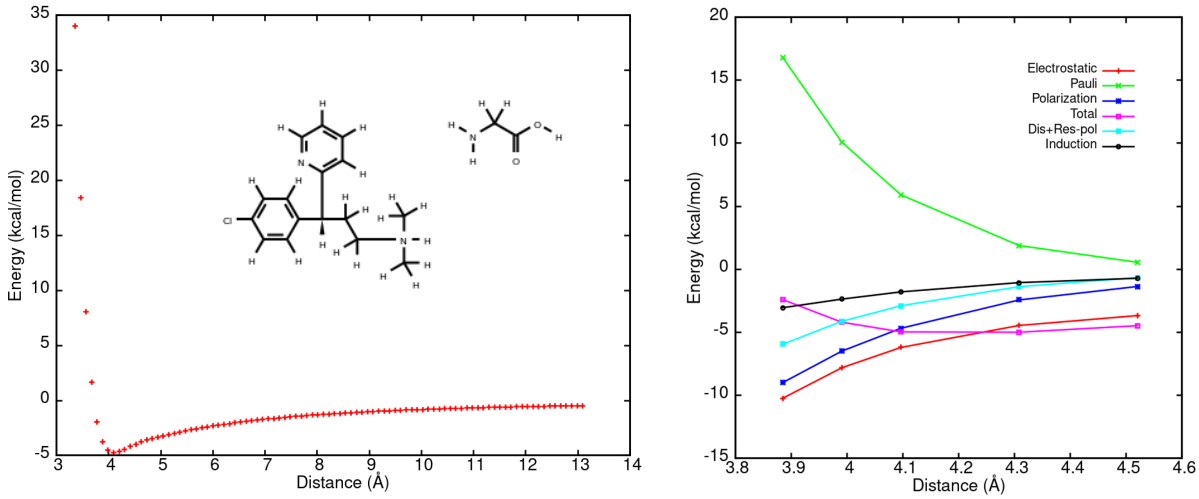


Figure 187: System 187

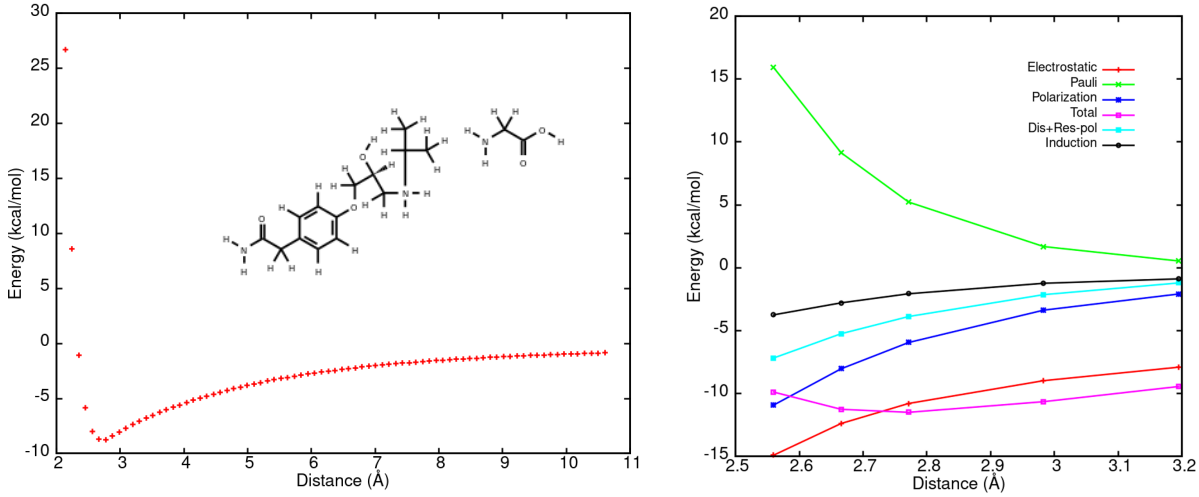


Figure 188: System 188

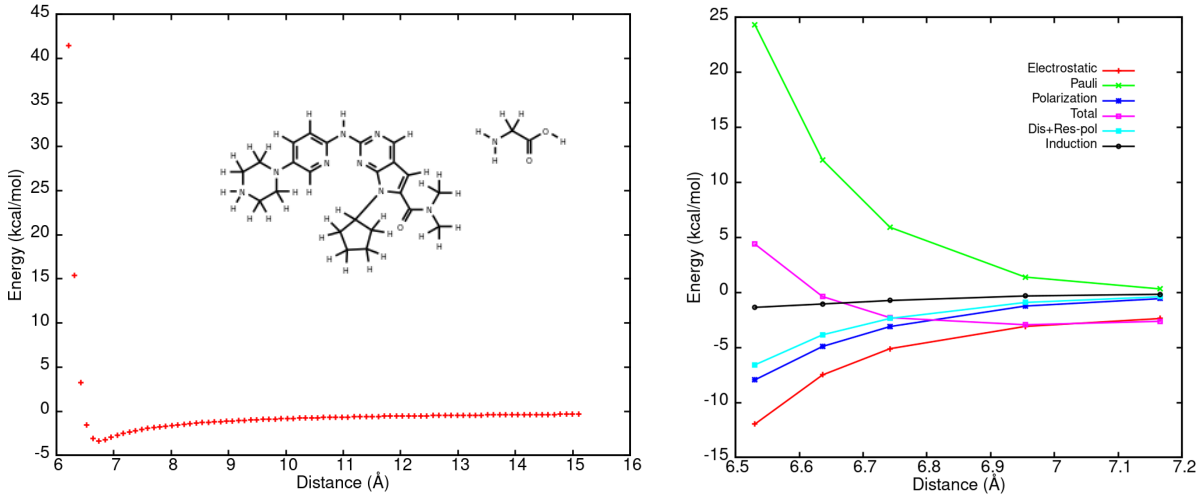
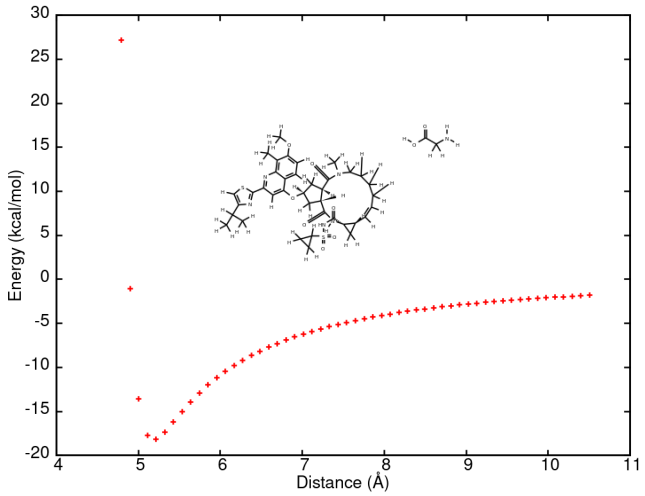


Figure 189: System 189



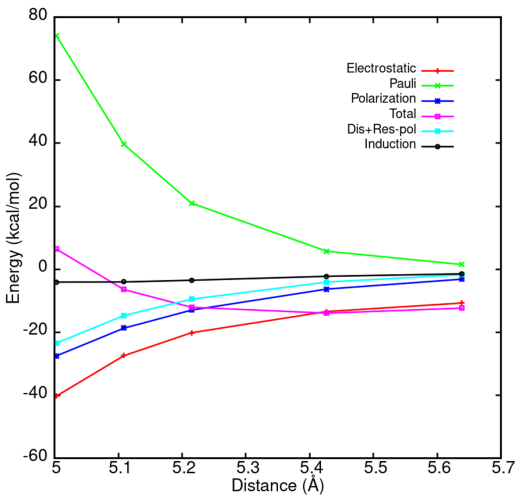
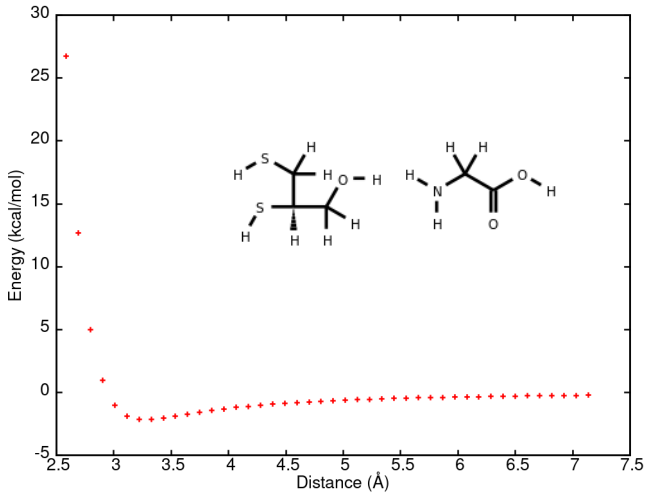


Figure 190: System 190



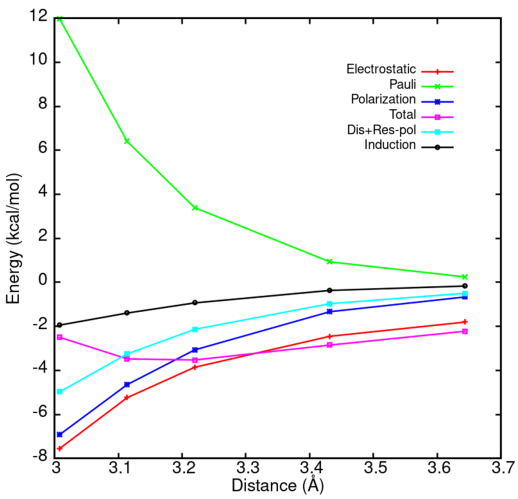
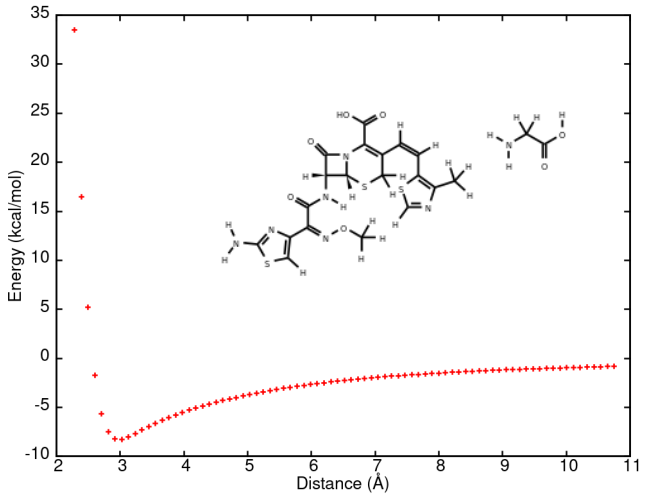


Figure 191: System 191



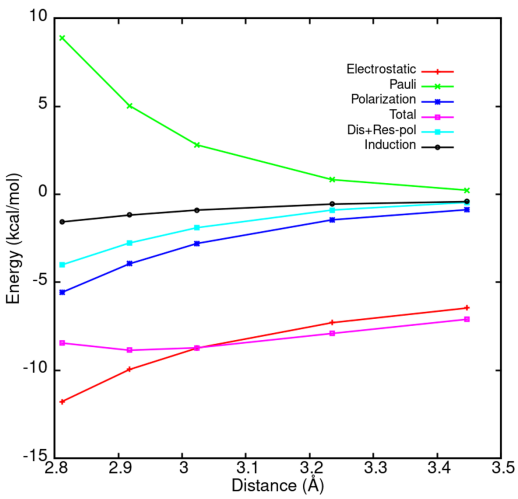


Figure 192: System 192

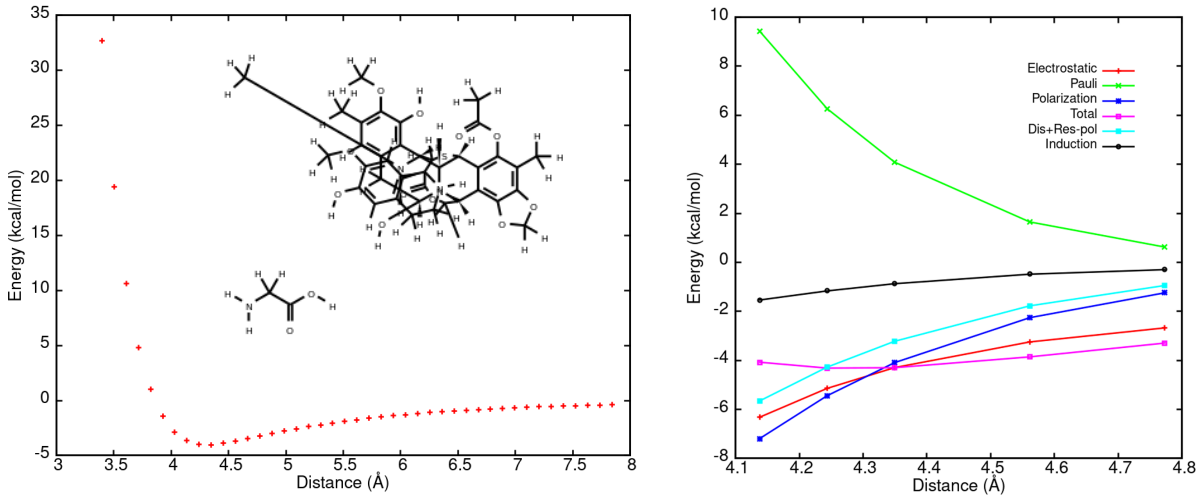


Figure 193: System 193

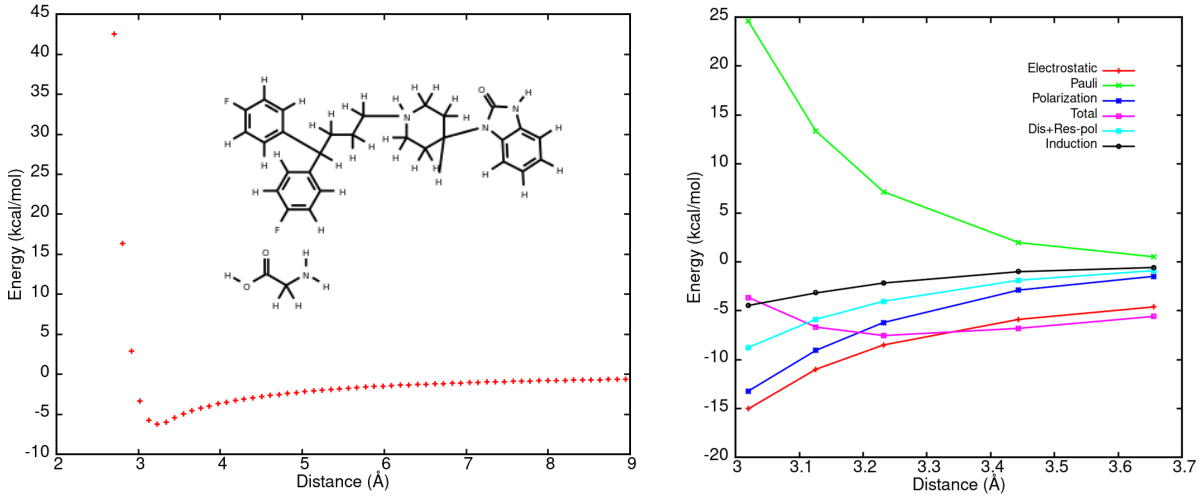


Figure 194: System 194

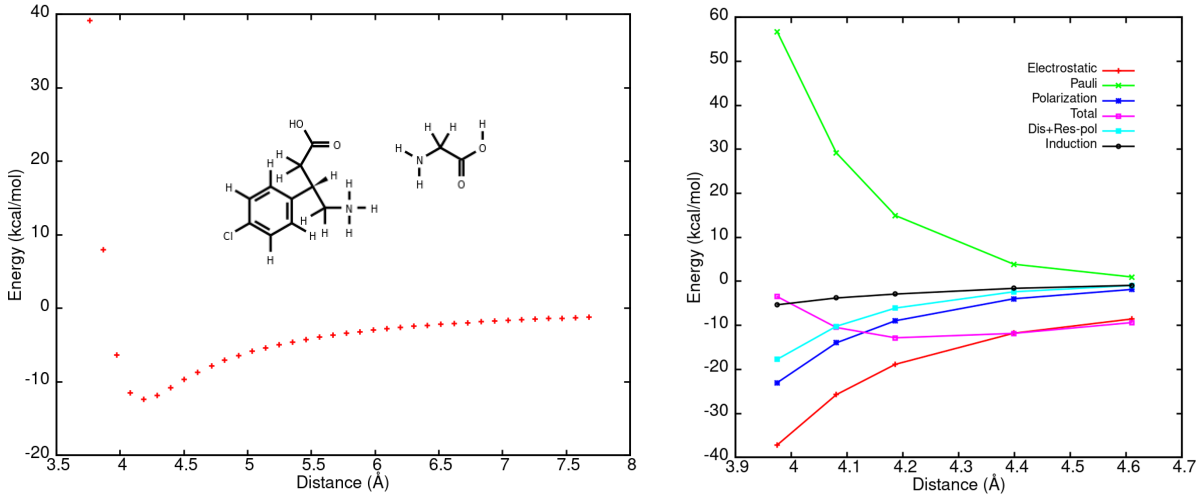
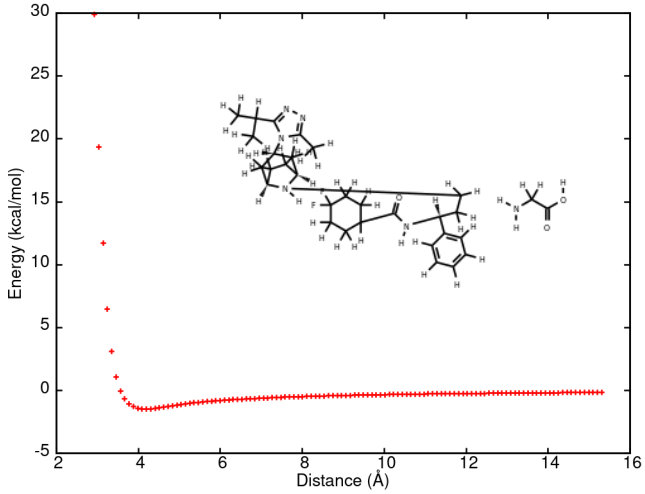


Figure 195: System 195



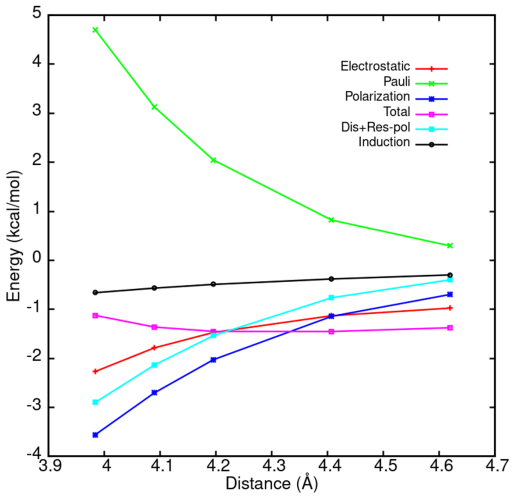
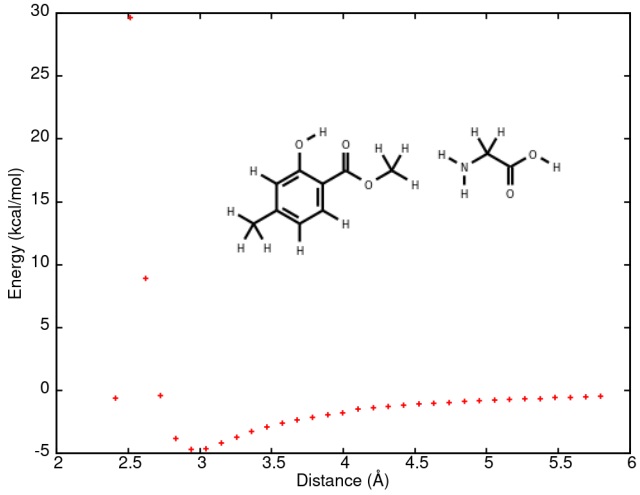


Figure 196: System 196



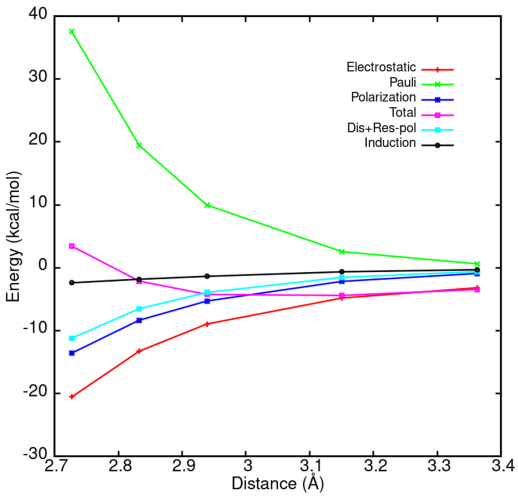
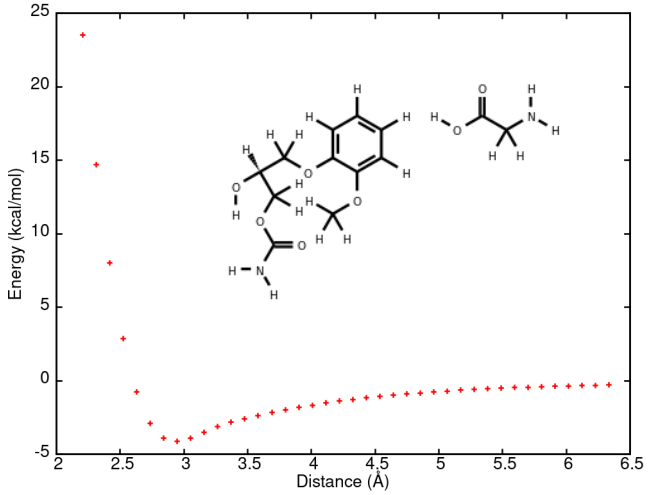


Figure 197: System 197



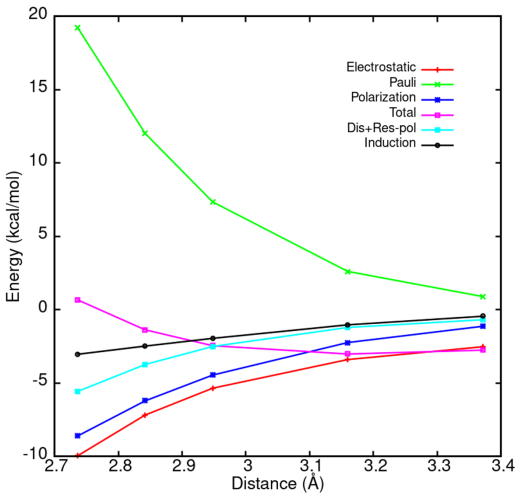


Figure 198: System 198

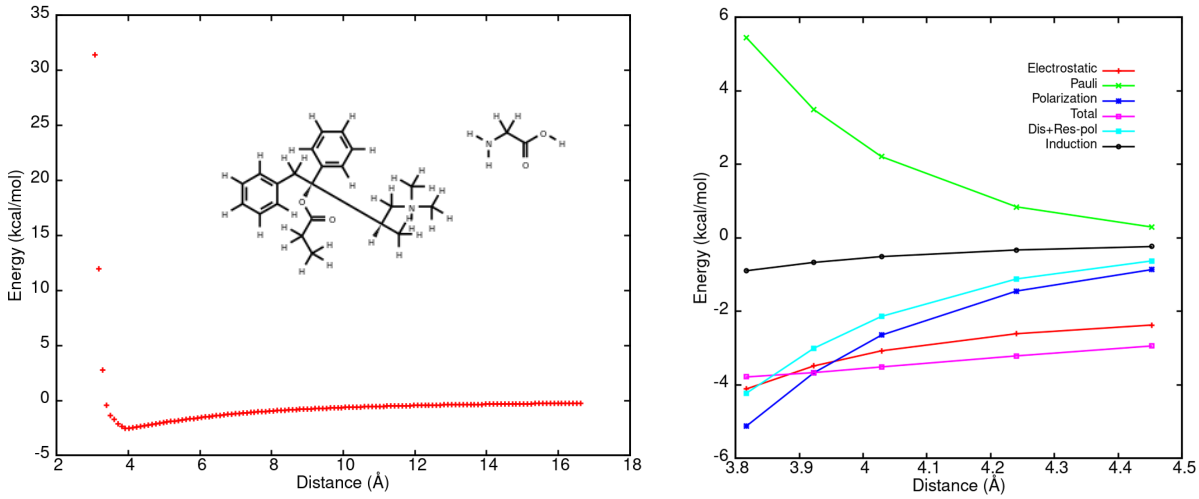


Figure 199: System 199

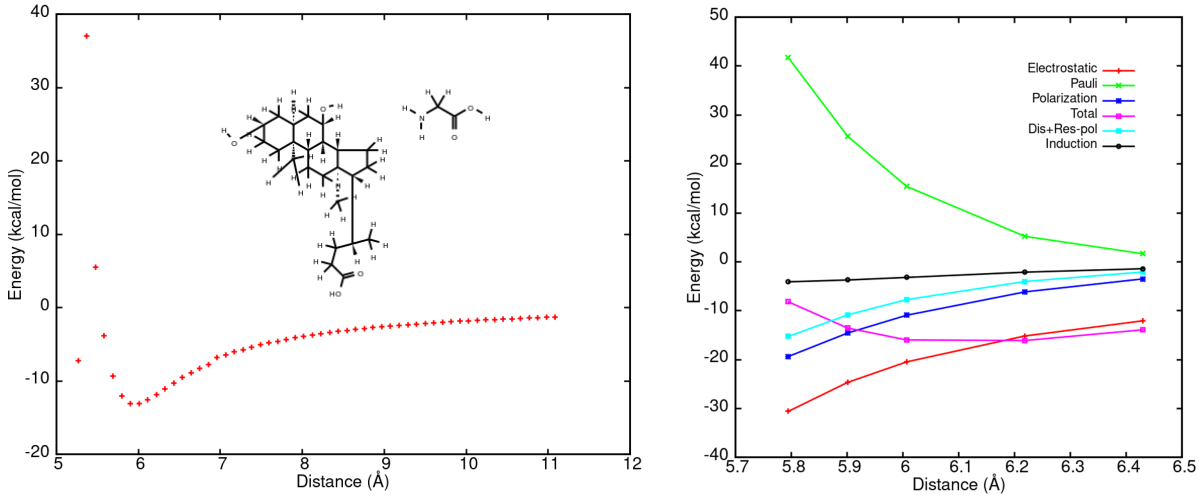


Figure 200: System 200

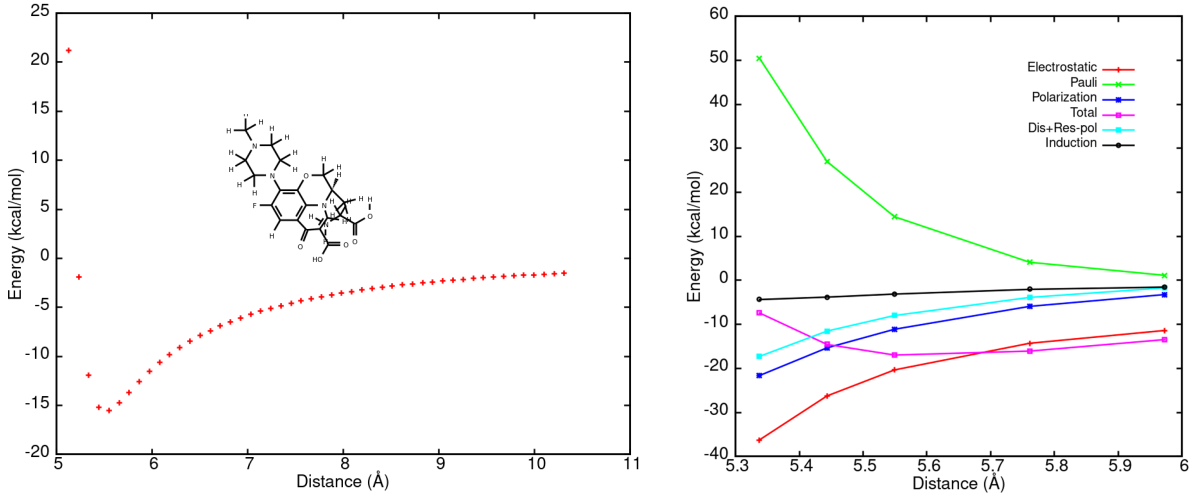


Figure 201: System 201

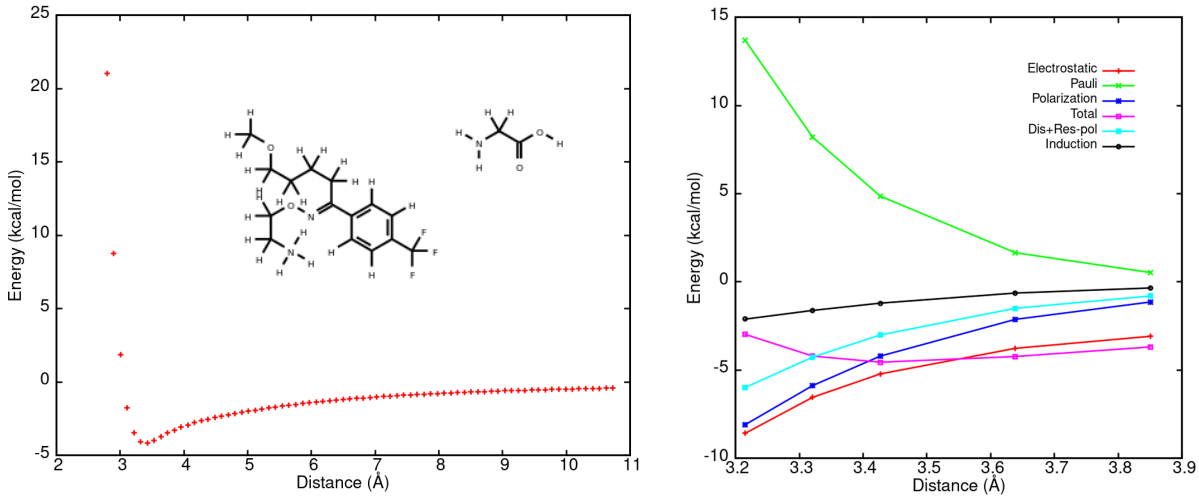


Figure 202: System 202

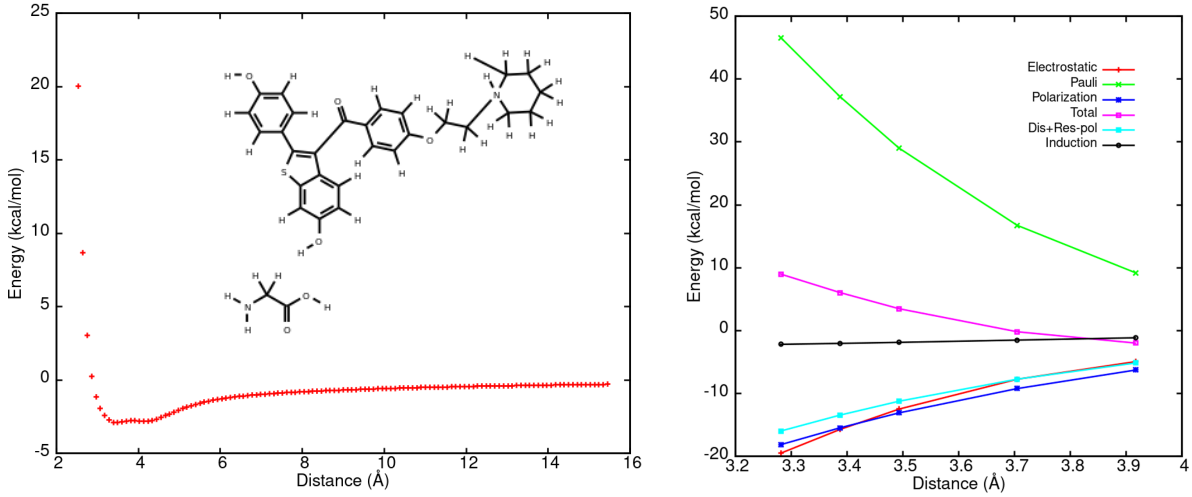


Figure 203: System 203

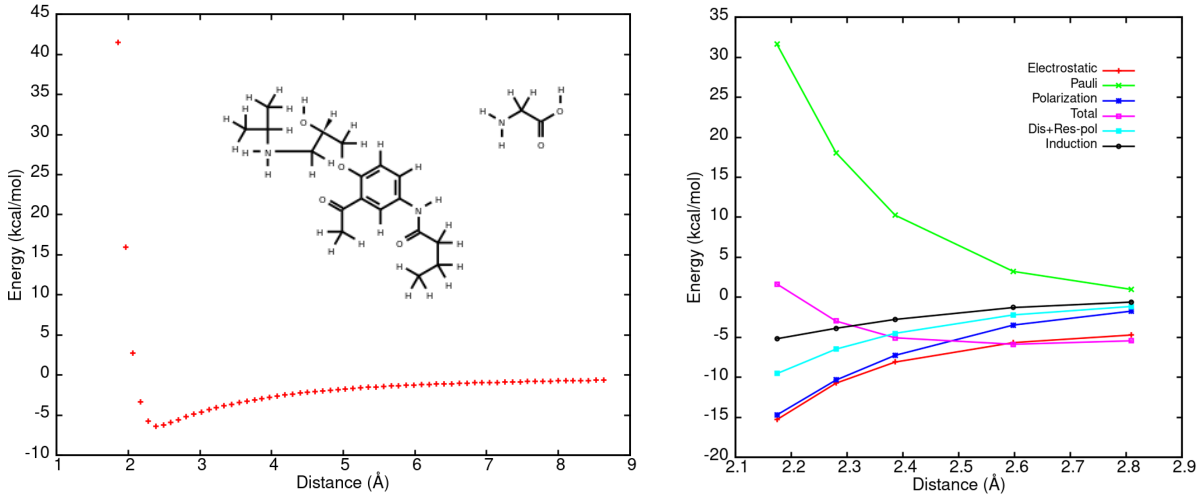
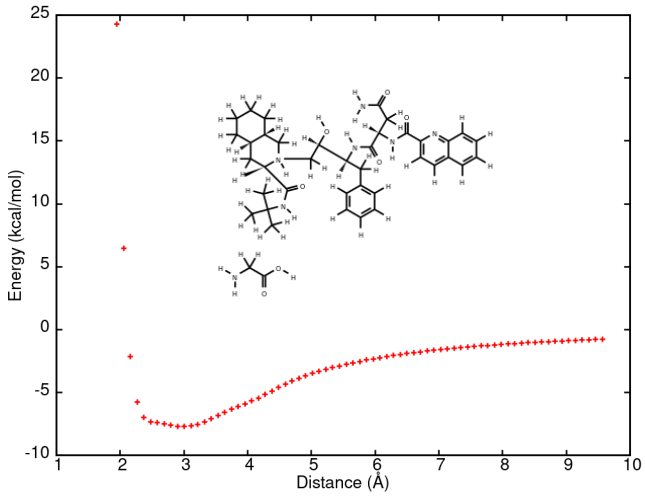


Figure 204: System 204



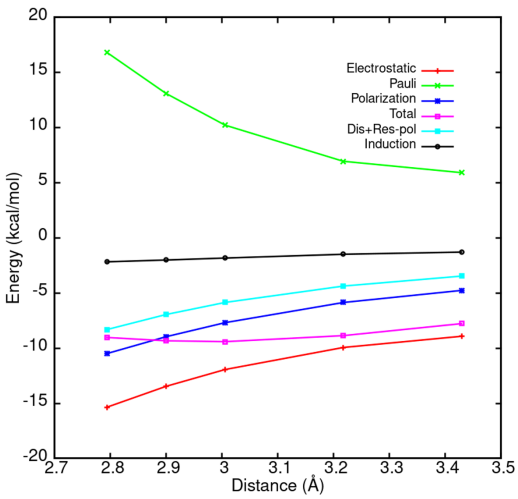
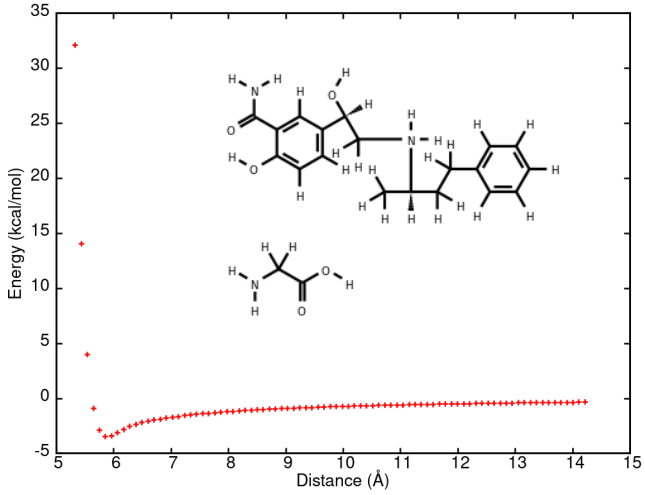


Figure 205: System 205



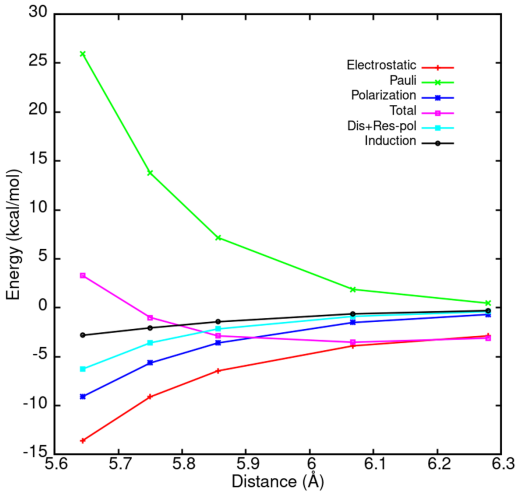
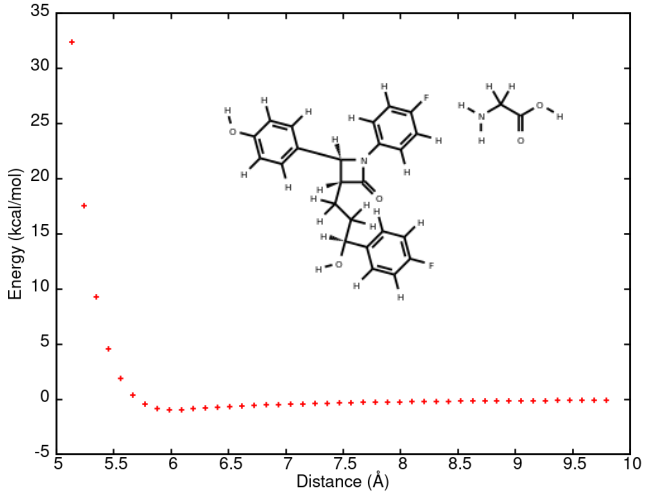


Figure 206: System 206



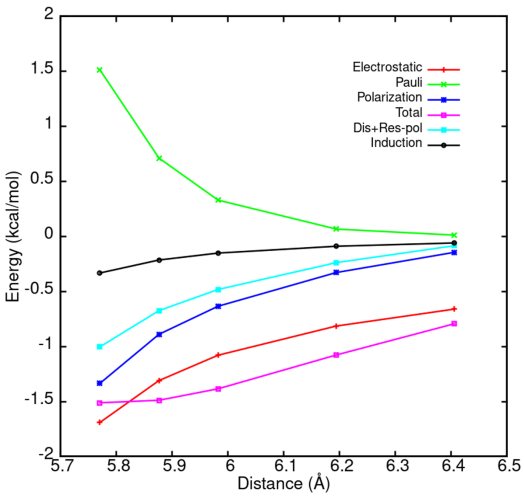


Figure 207: System 207

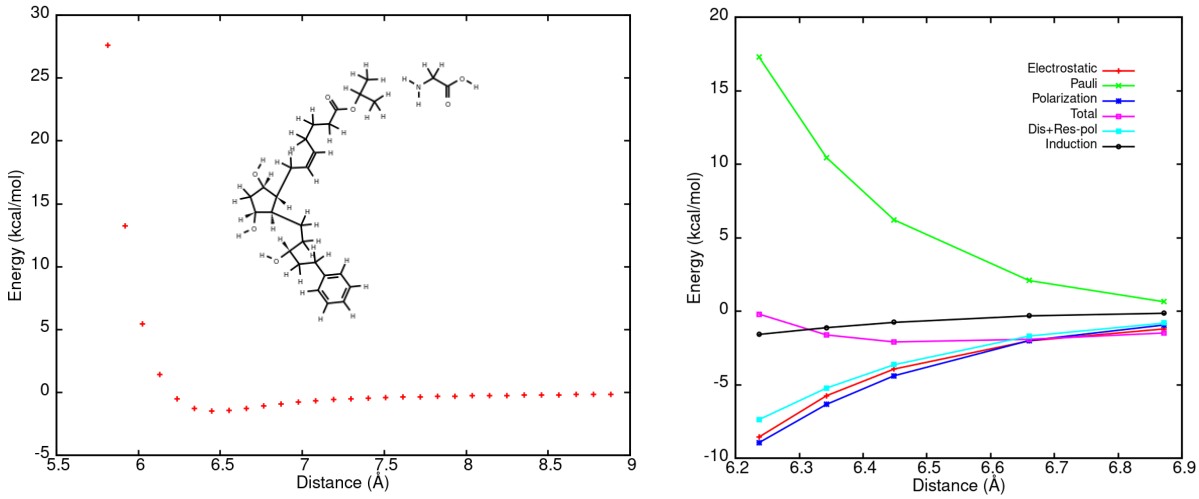


Figure 208: System 208

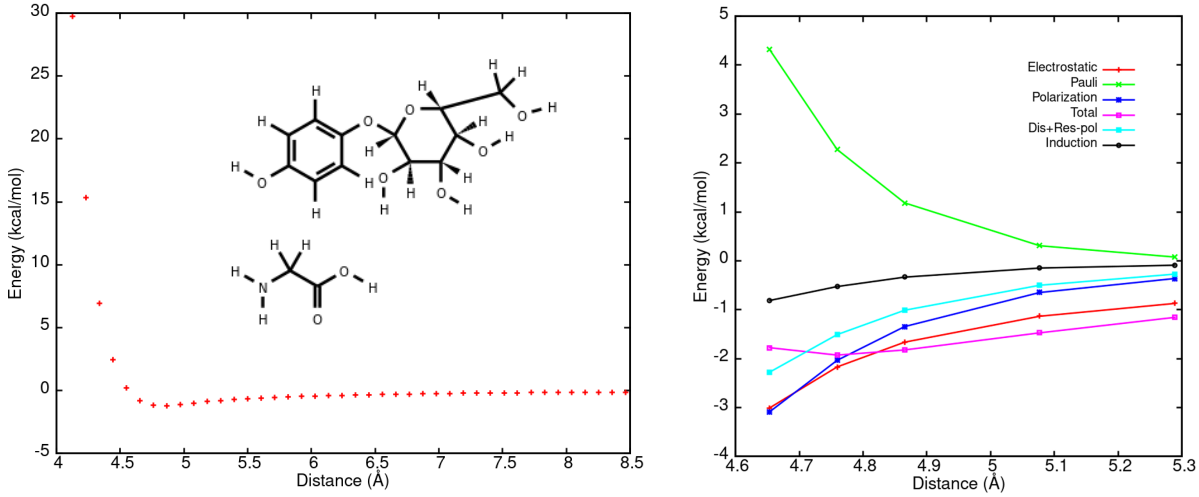


Figure 209: System 209

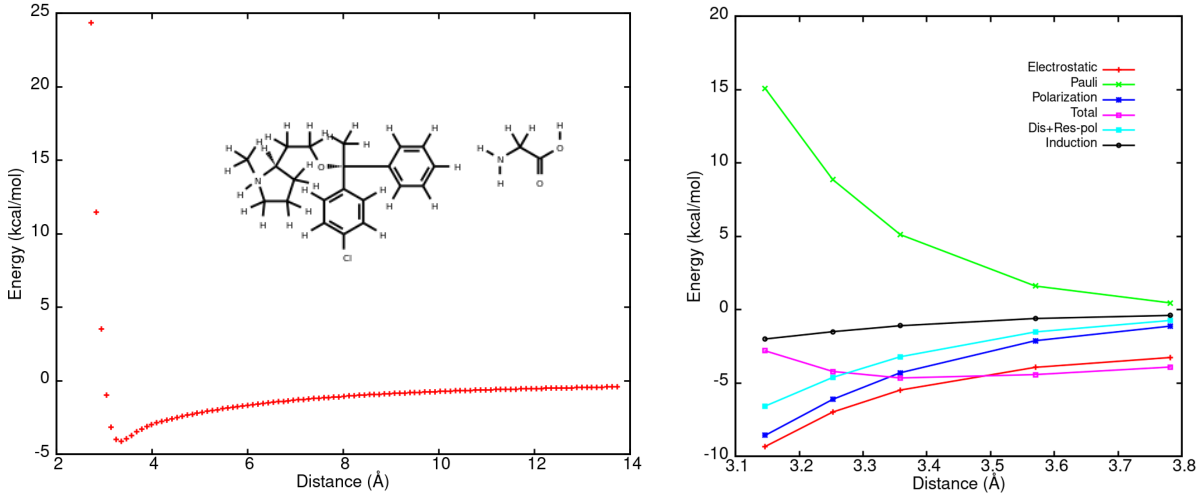


Figure 210: System 210

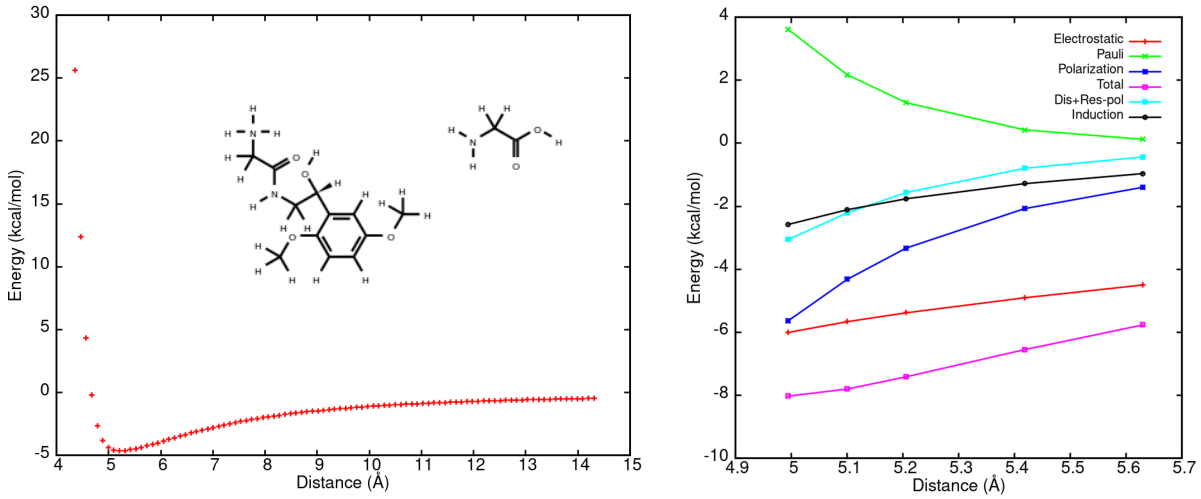


Figure 211: System 211

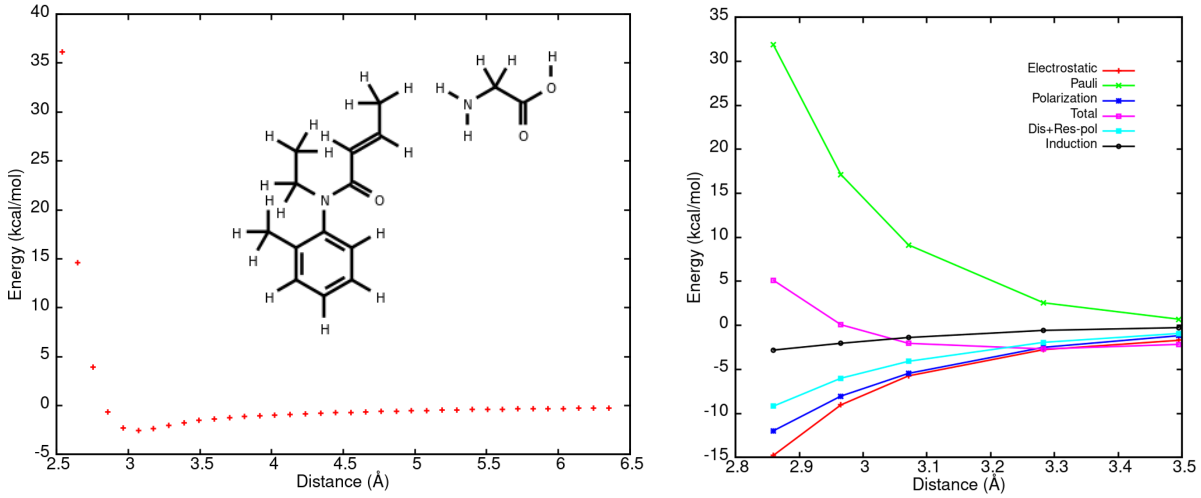


Figure 212: System 212

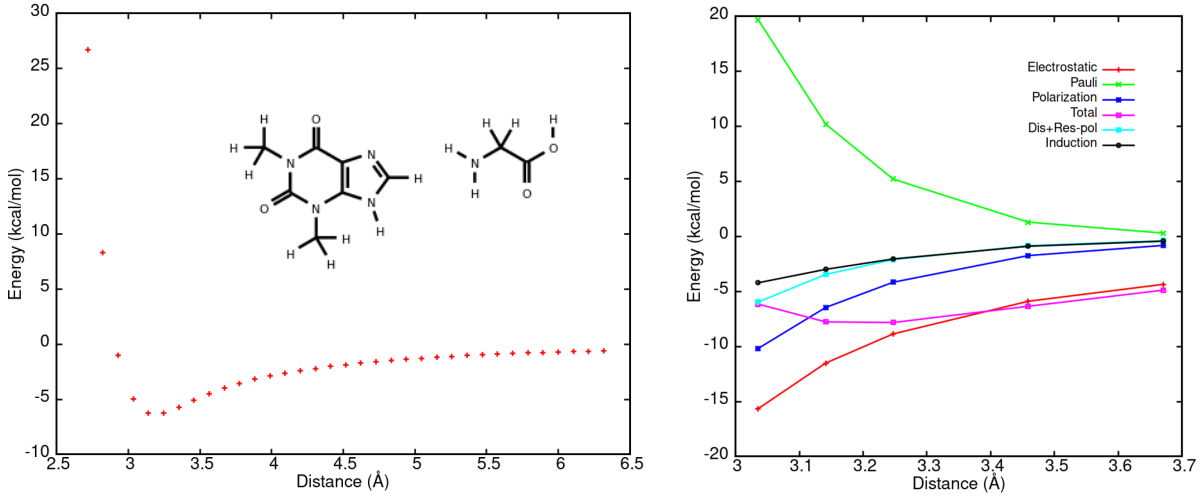
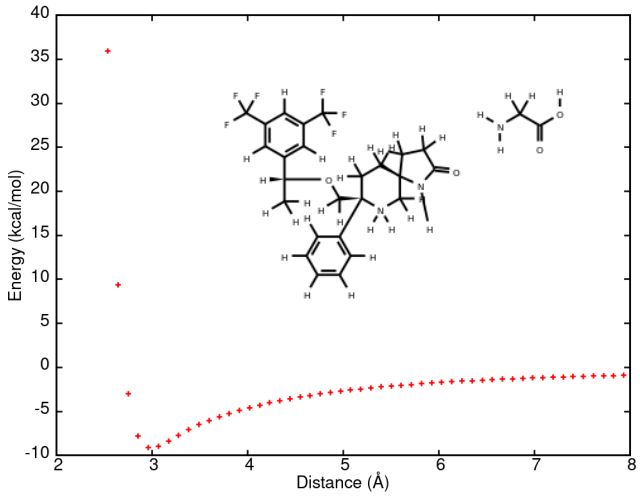


Figure 213: System 213



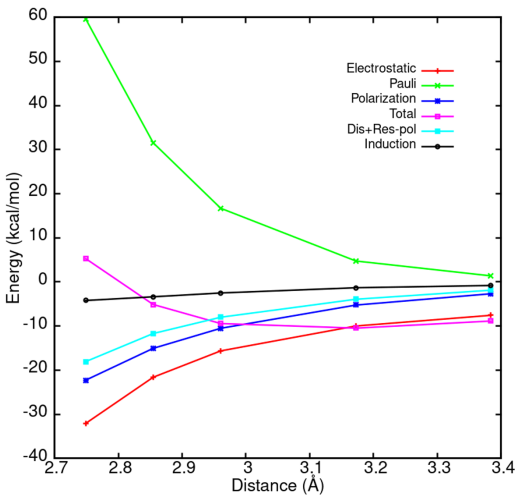
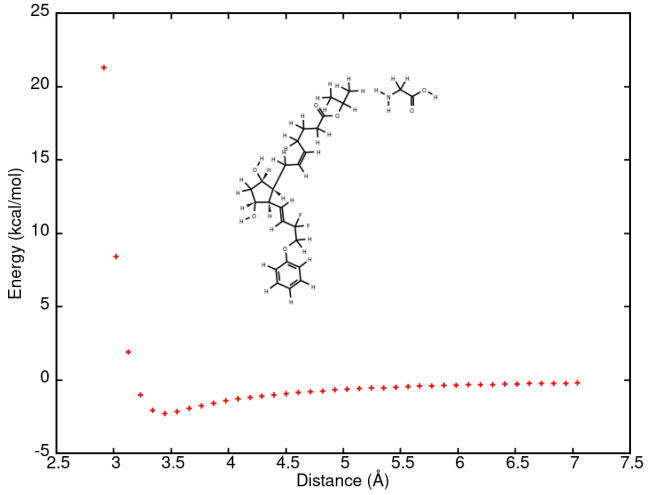


Figure 214: System 214



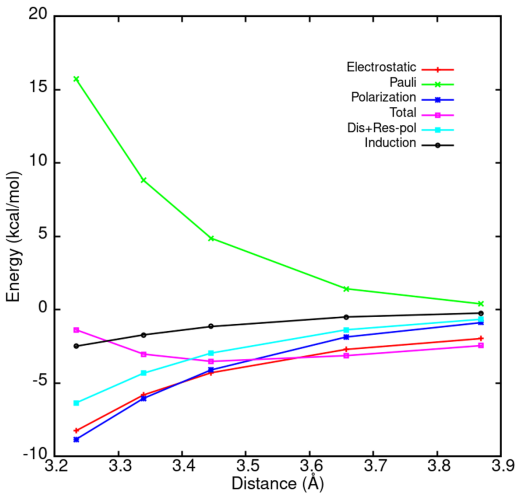
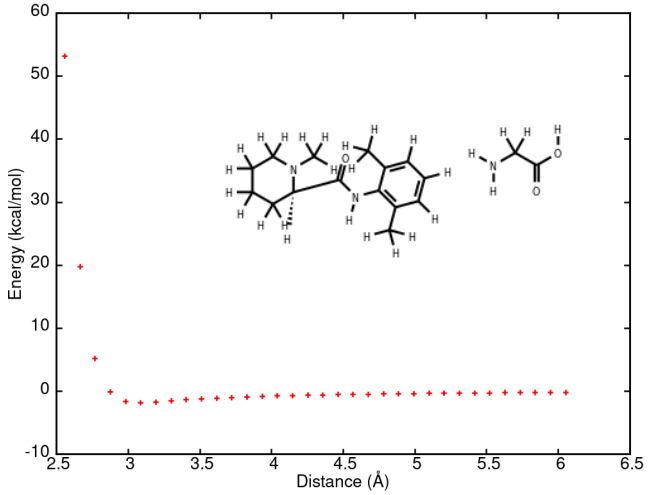


Figure 215: System 215



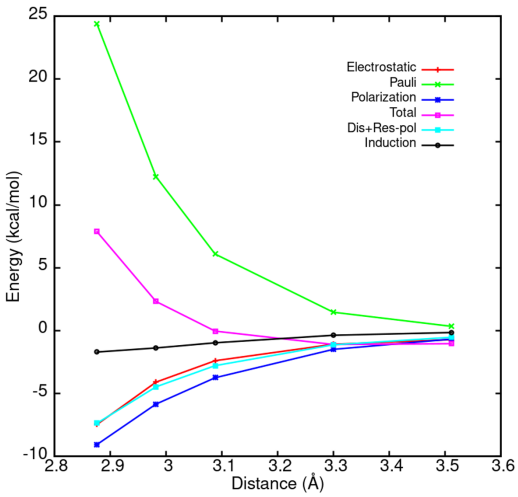


Figure 216: System 216

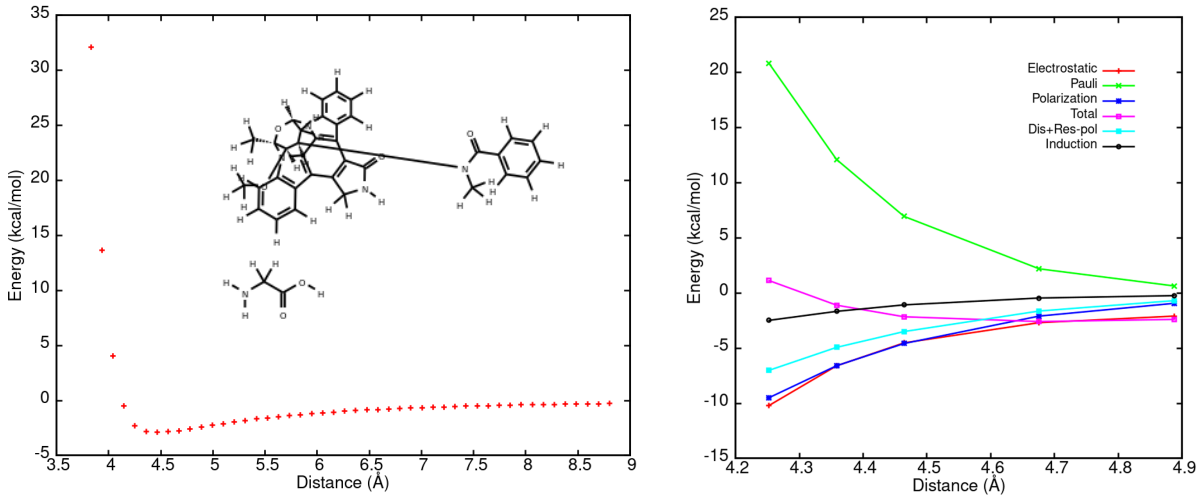


Figure 217: System 217

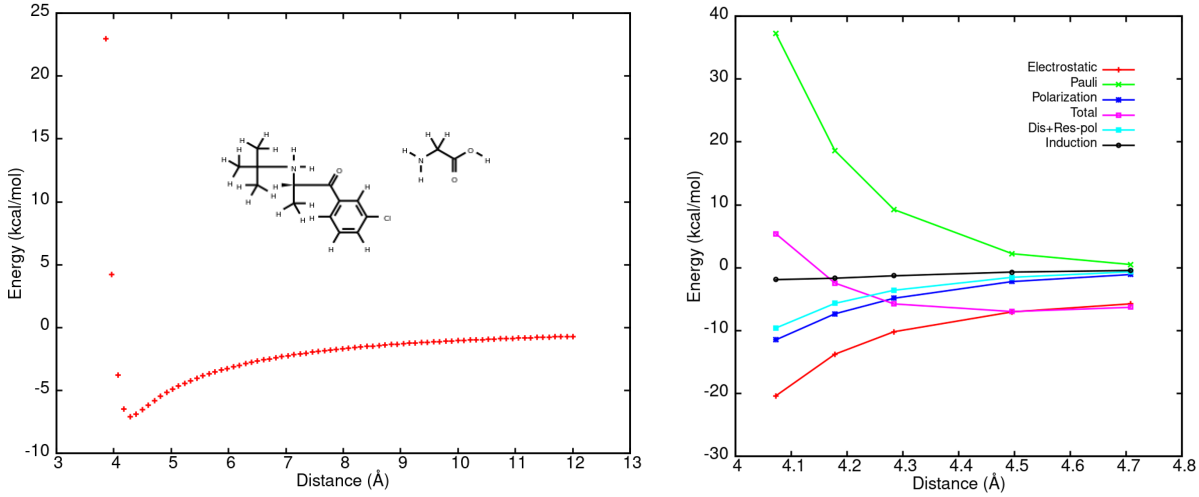


Figure 218: System 218

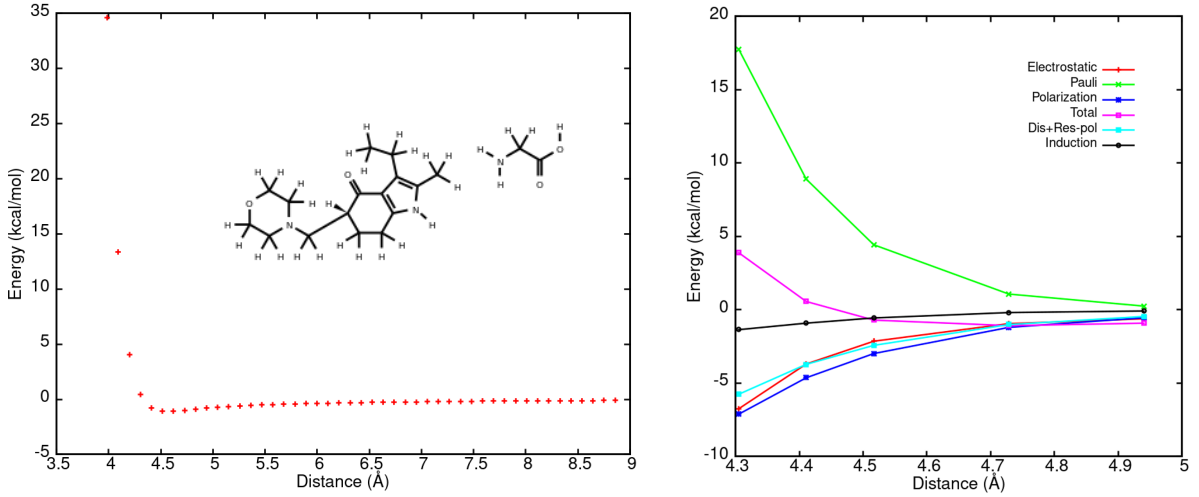


Figure 219: System 219

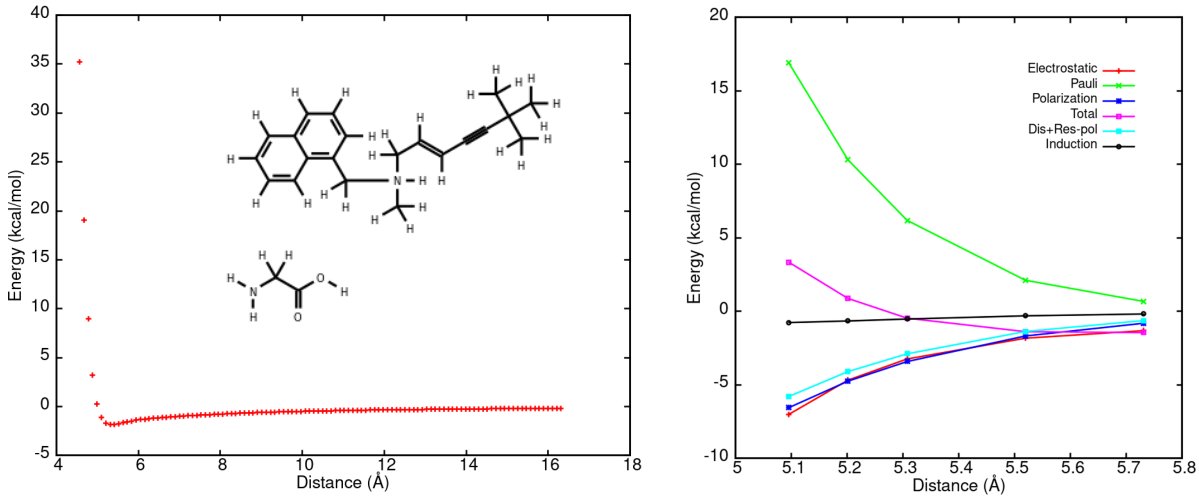


Figure 220: System 220

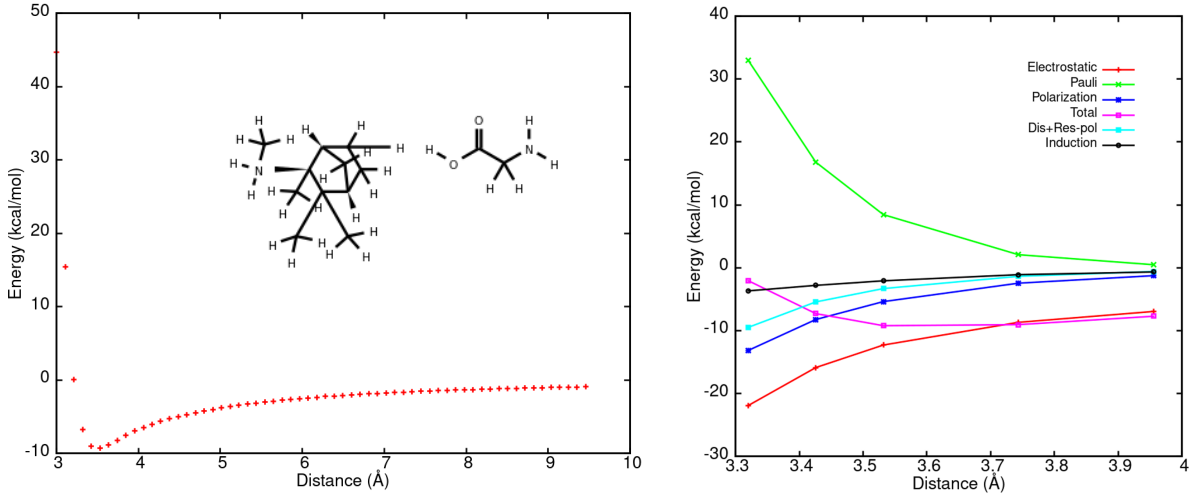


Figure 221: System 221

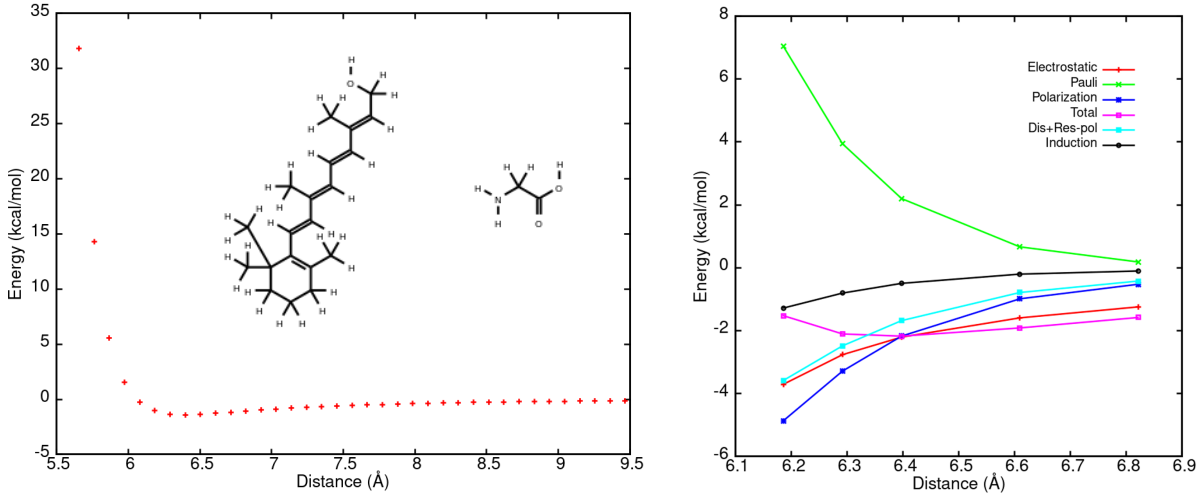
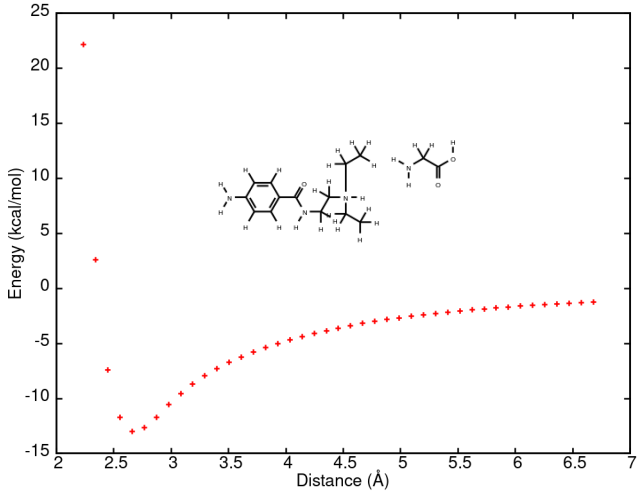


Figure 222: System 222



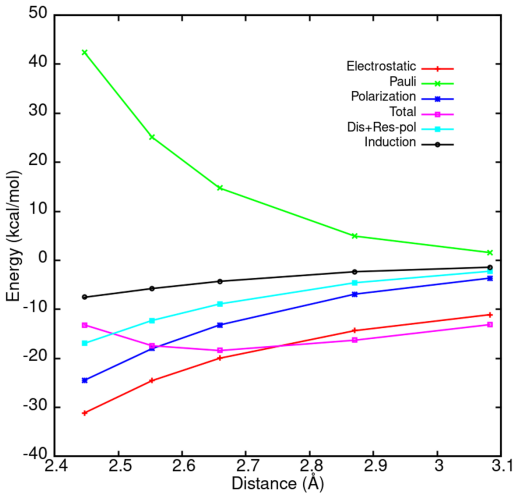
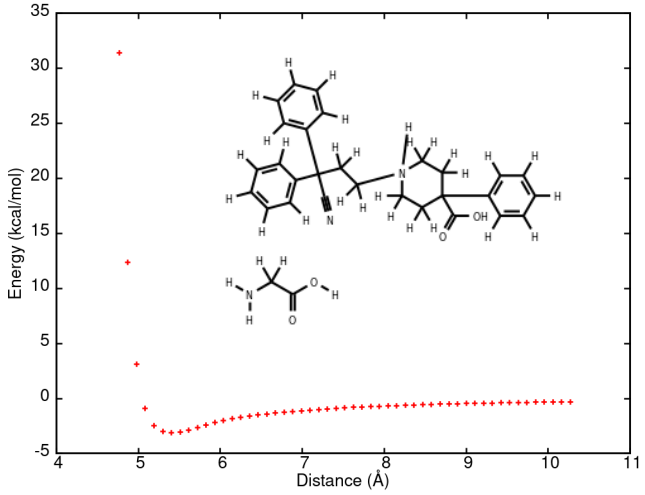


Figure 223: System 223



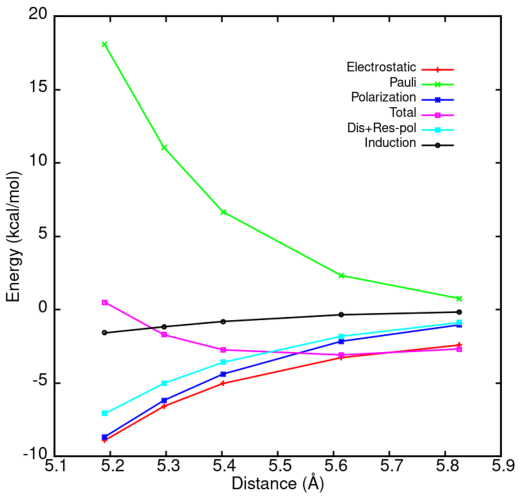
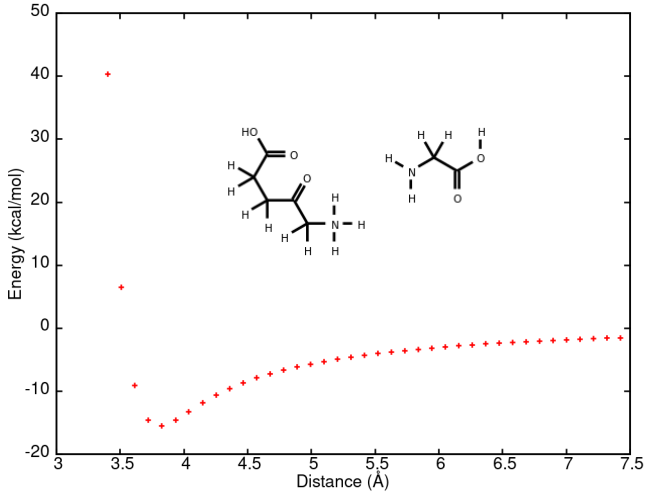


Figure 224: System 224



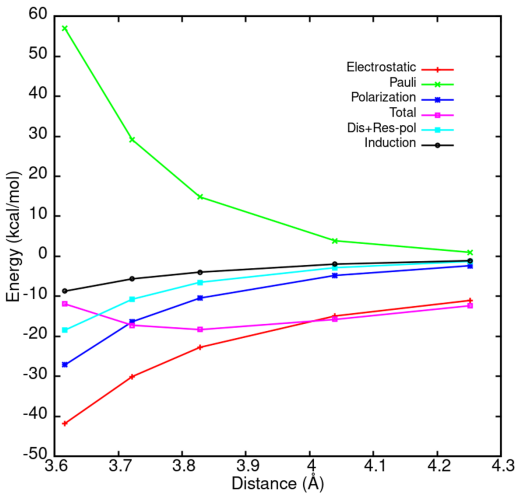


Figure 225: System 225

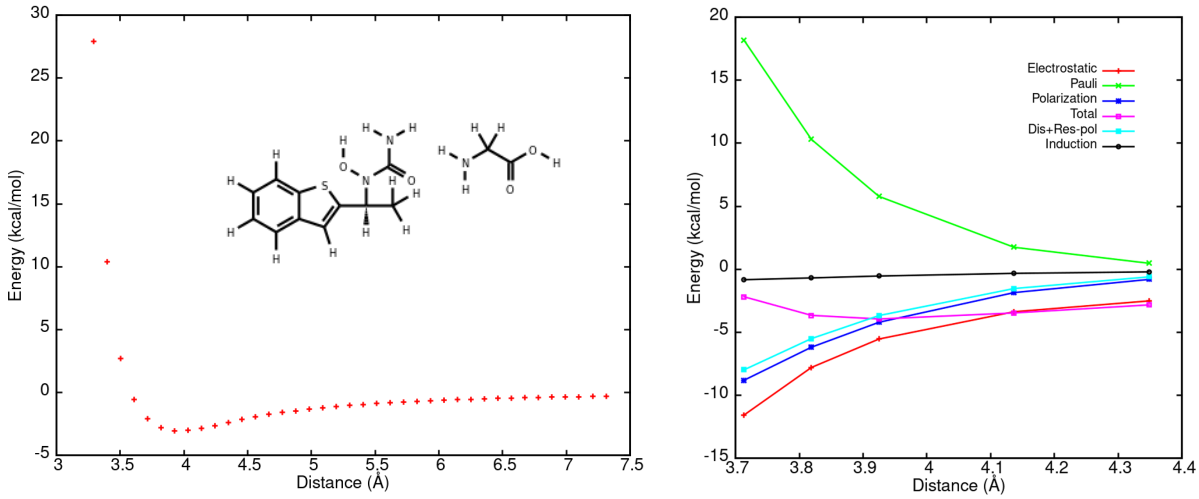


Figure 226: System 226

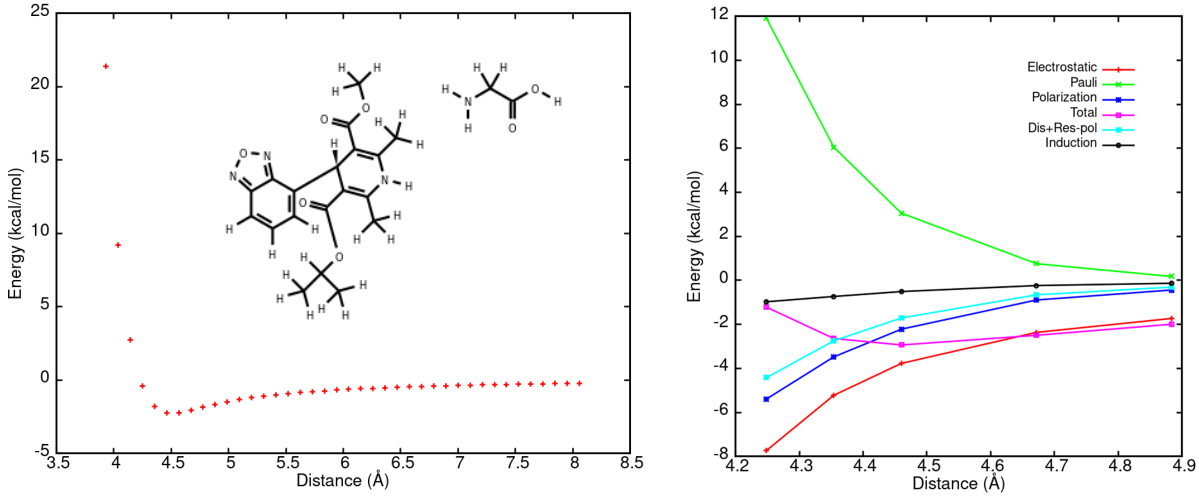


Figure 227: System 227

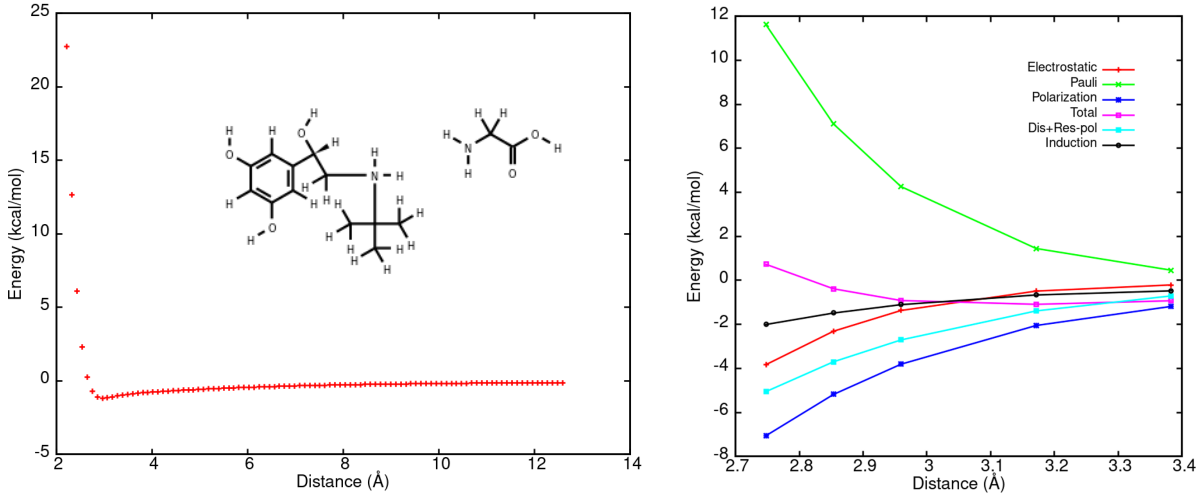
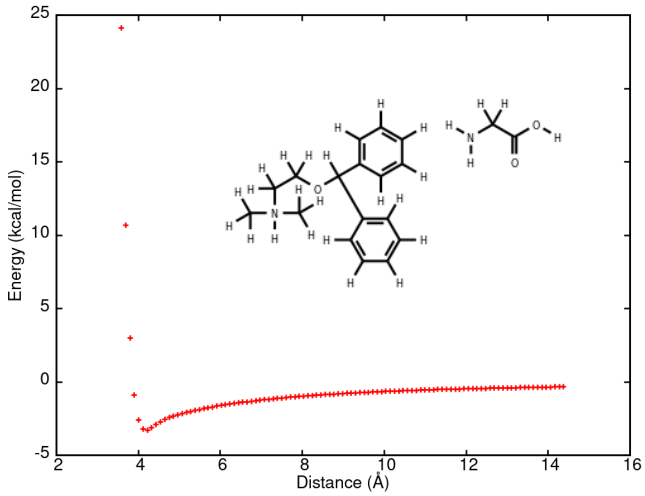


Figure 228: System 228



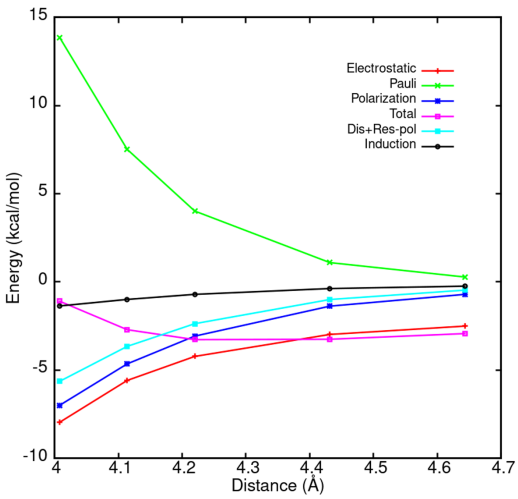
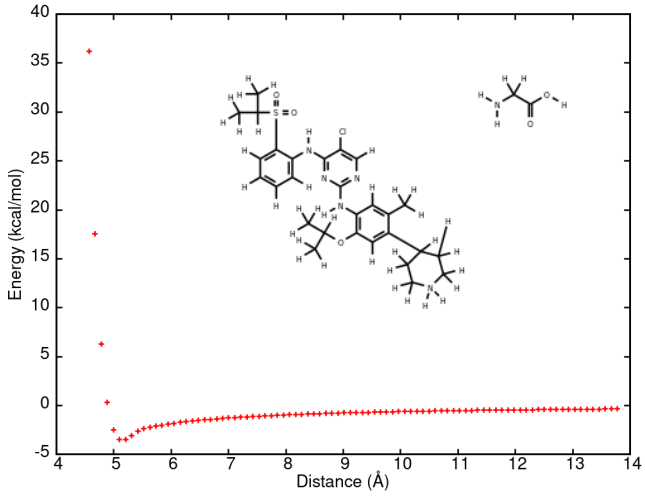


Figure 229: System 229



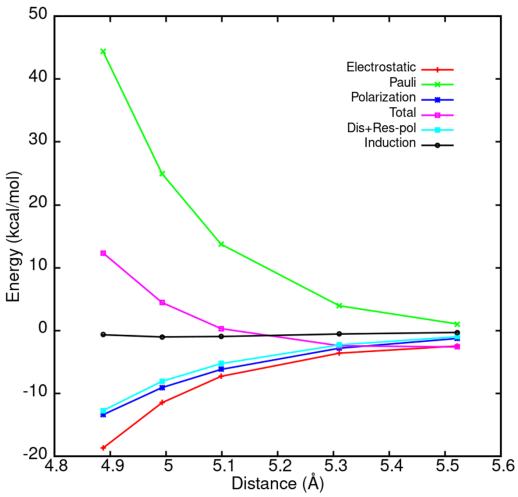
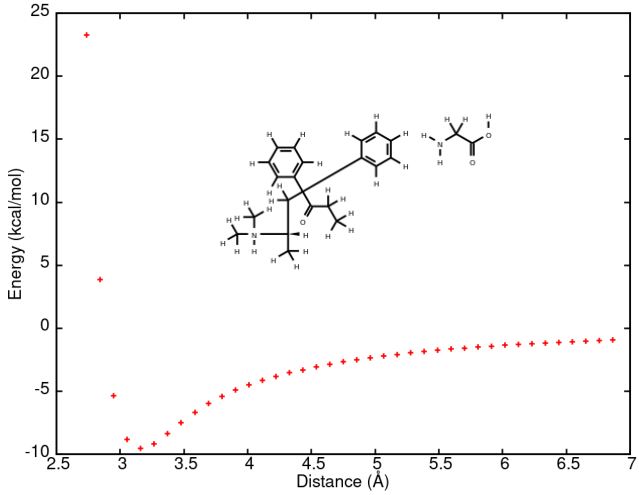


Figure 230: System 230



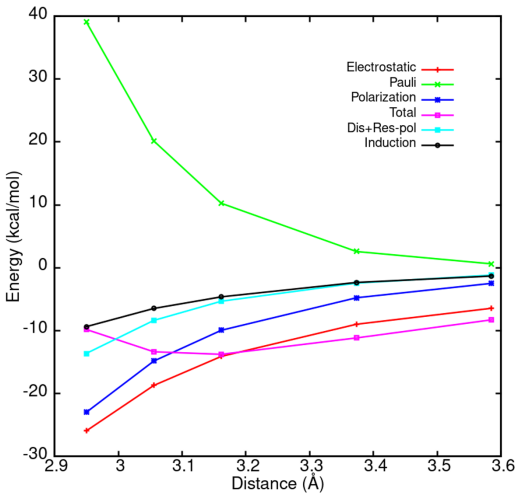
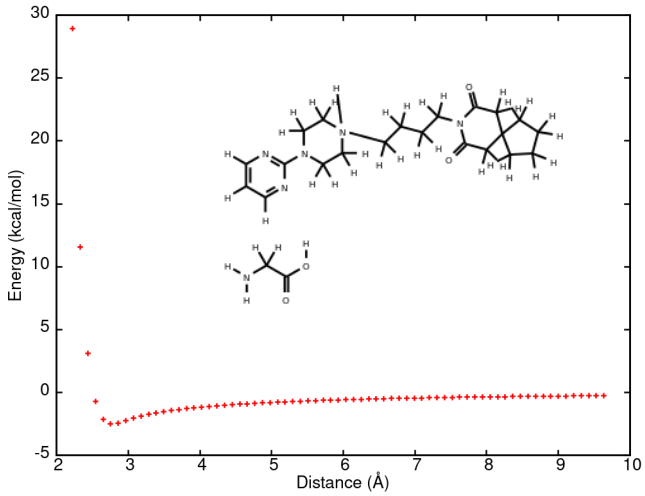


Figure 231: System 231



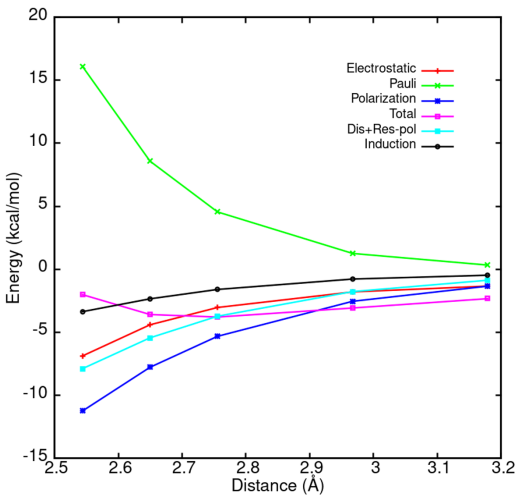
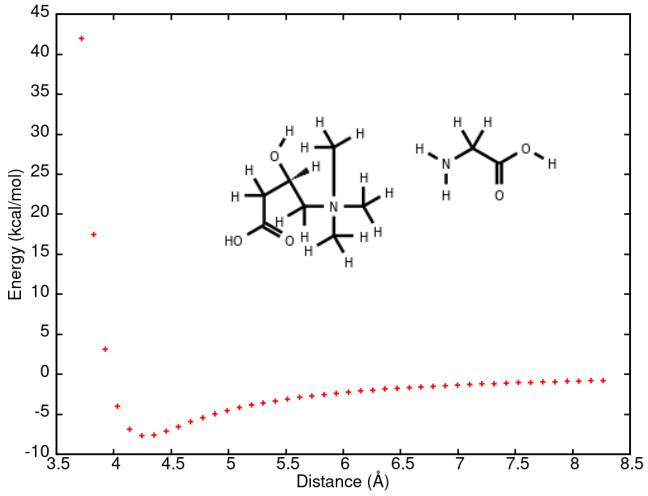


Figure 232: System 232



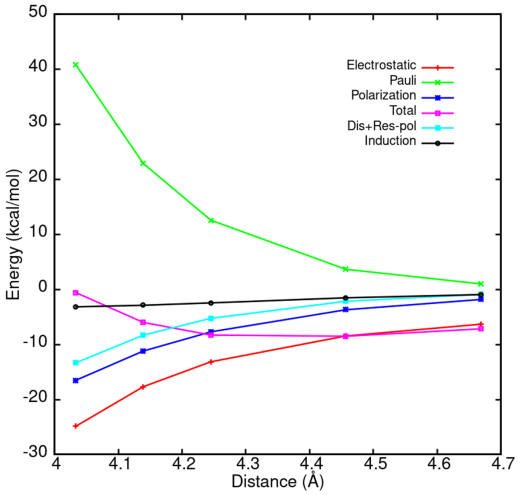
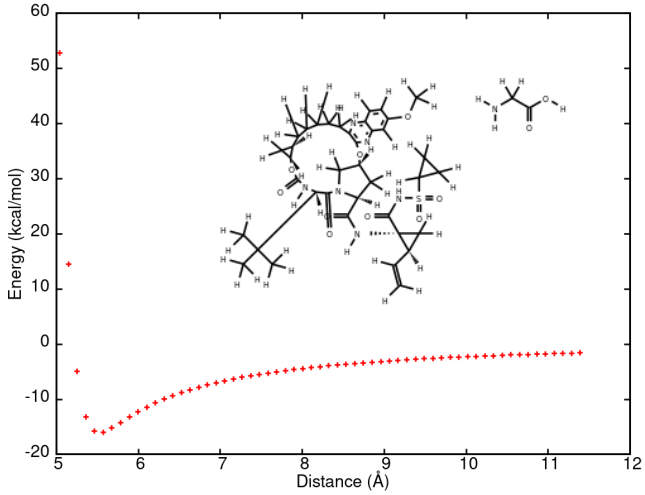


Figure 233: System 233



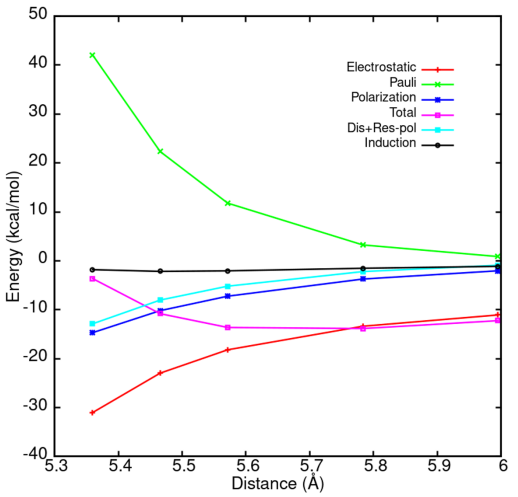


Figure 234: System 234

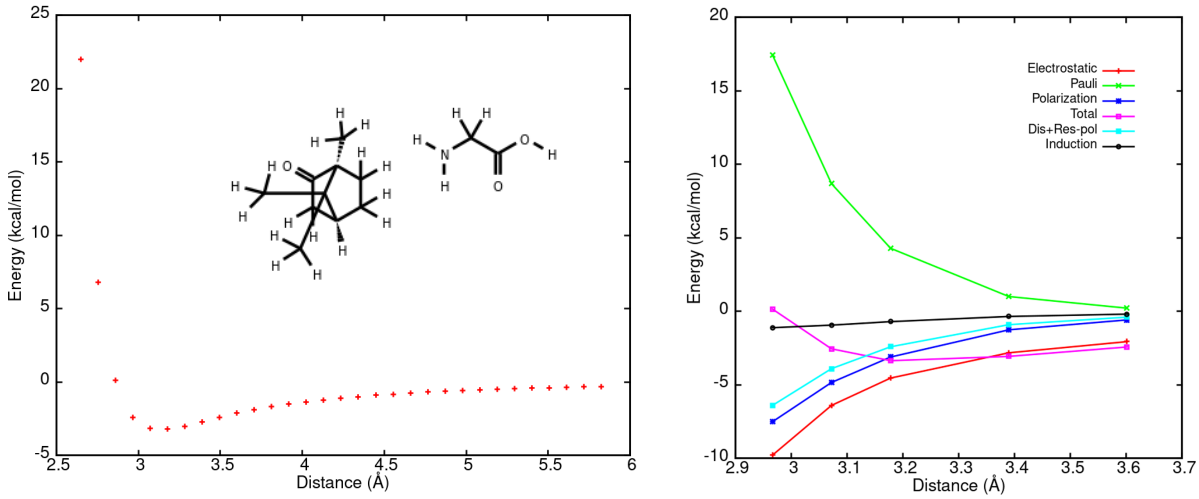


Figure 235: System 235

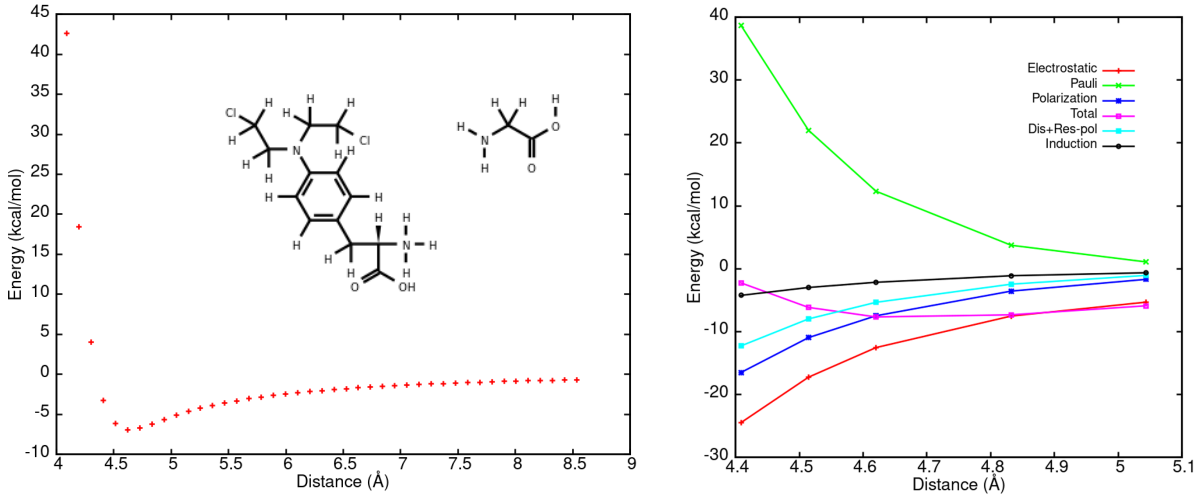


Figure 236: System 236

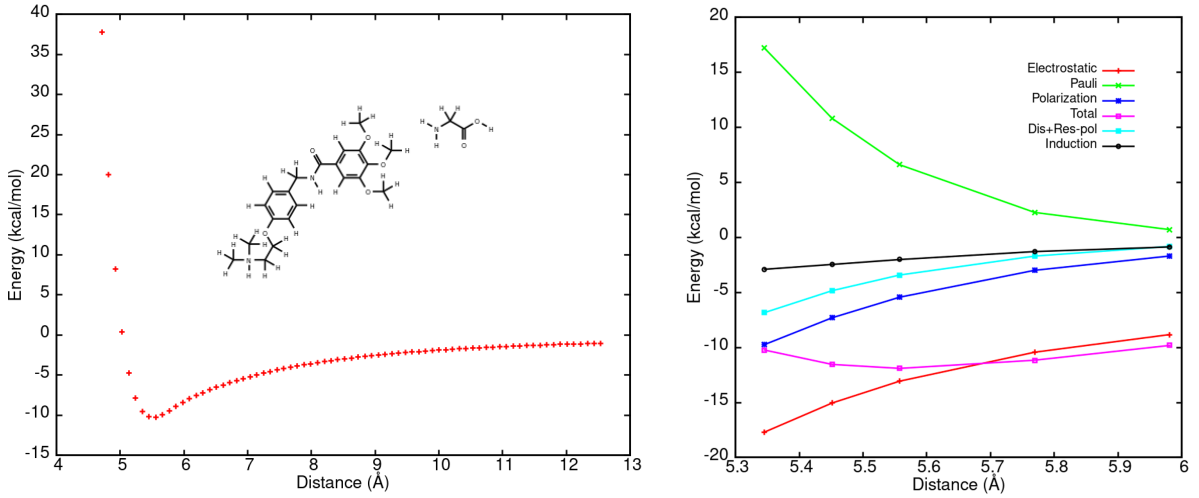


Figure 237: System 237

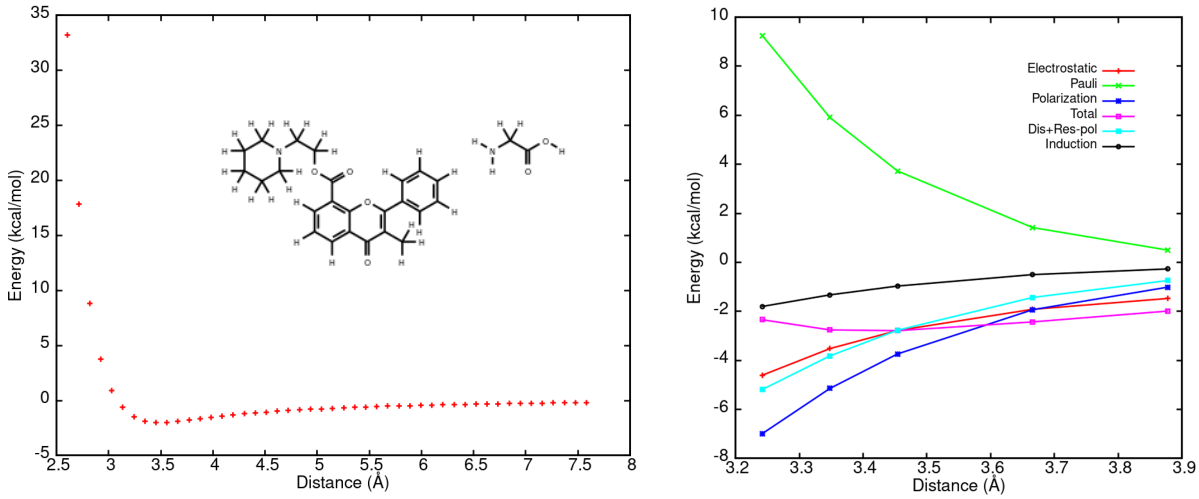


Figure 238: System 238

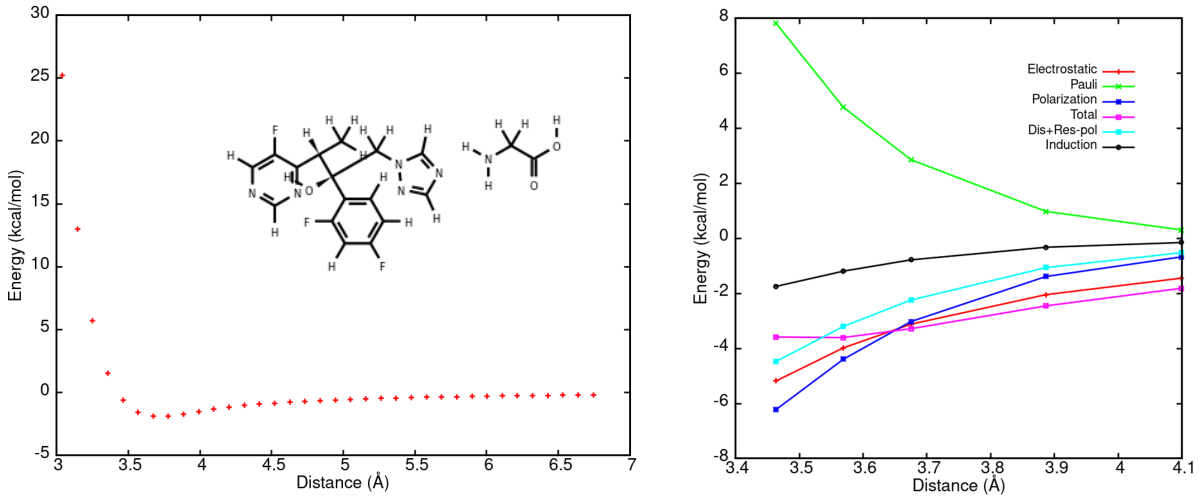


Figure 239: System 239

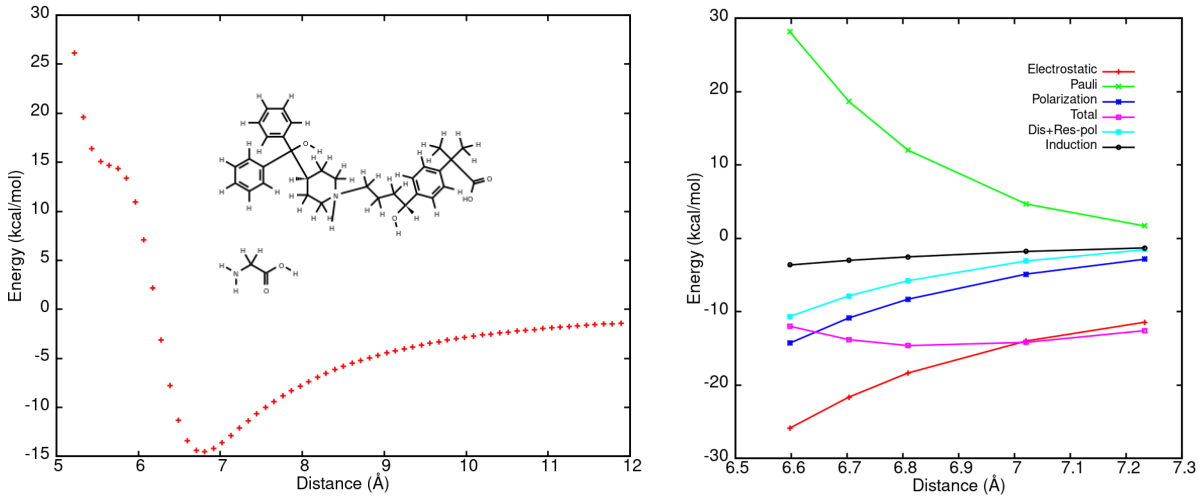


Figure 240: System 240

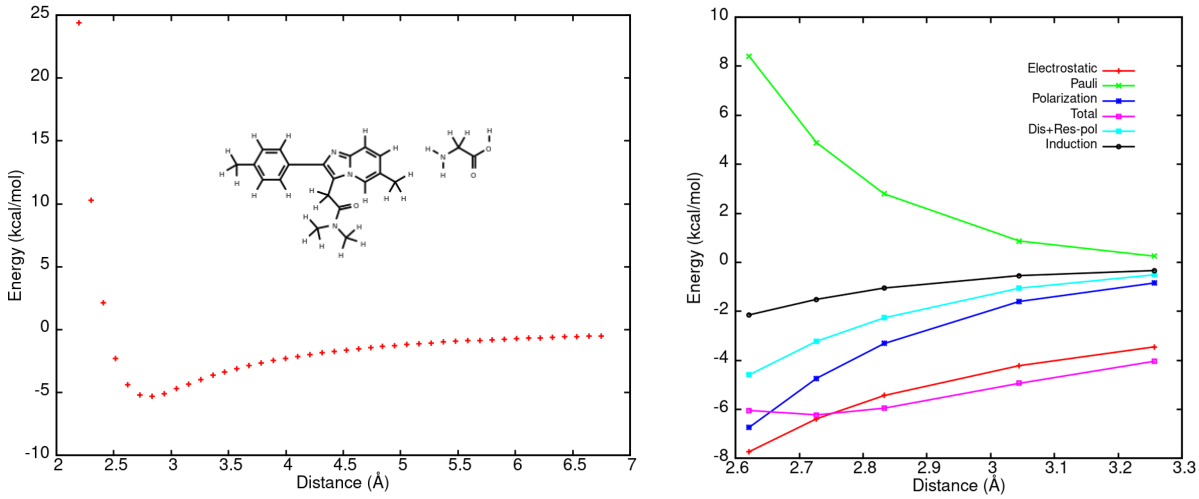


Figure 241: System 241

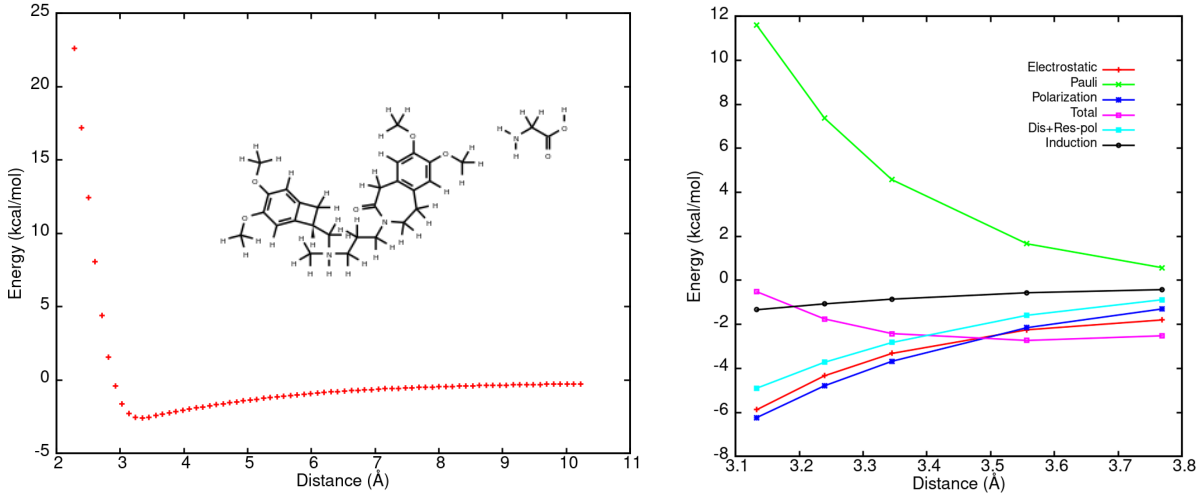


Figure 242: System 242

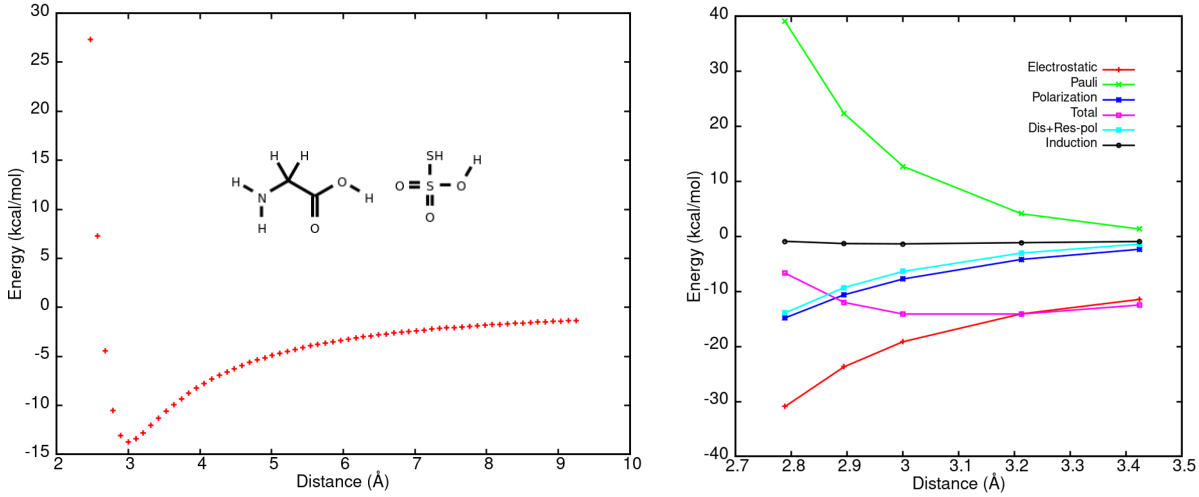


Figure 243: System 243

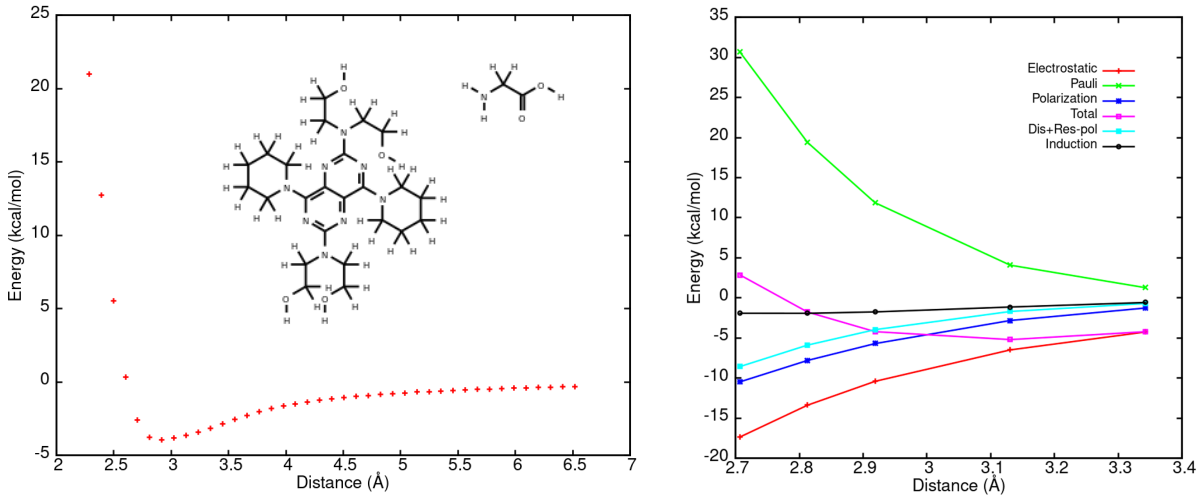


Figure 244: System 244

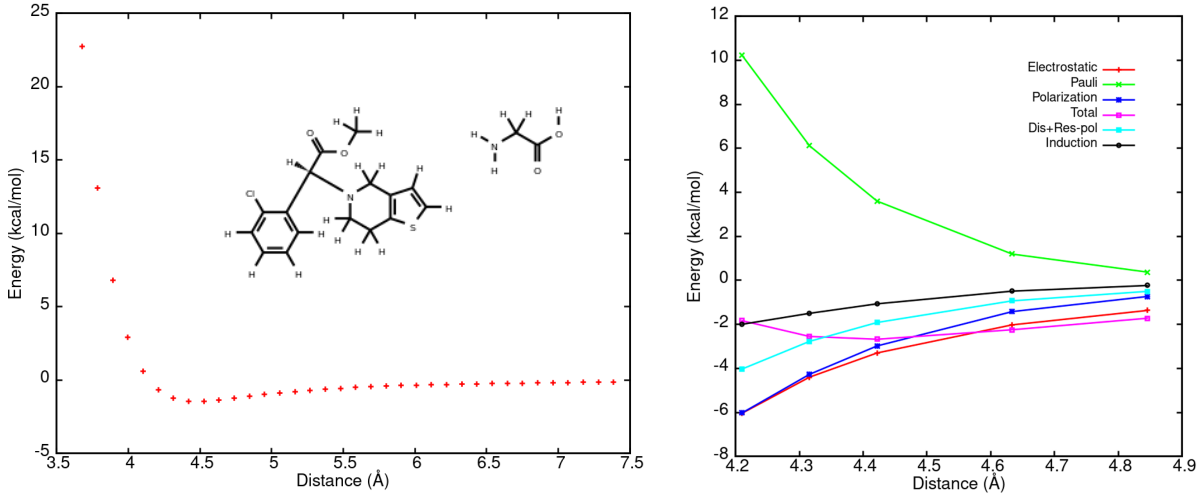


Figure 245: System 245

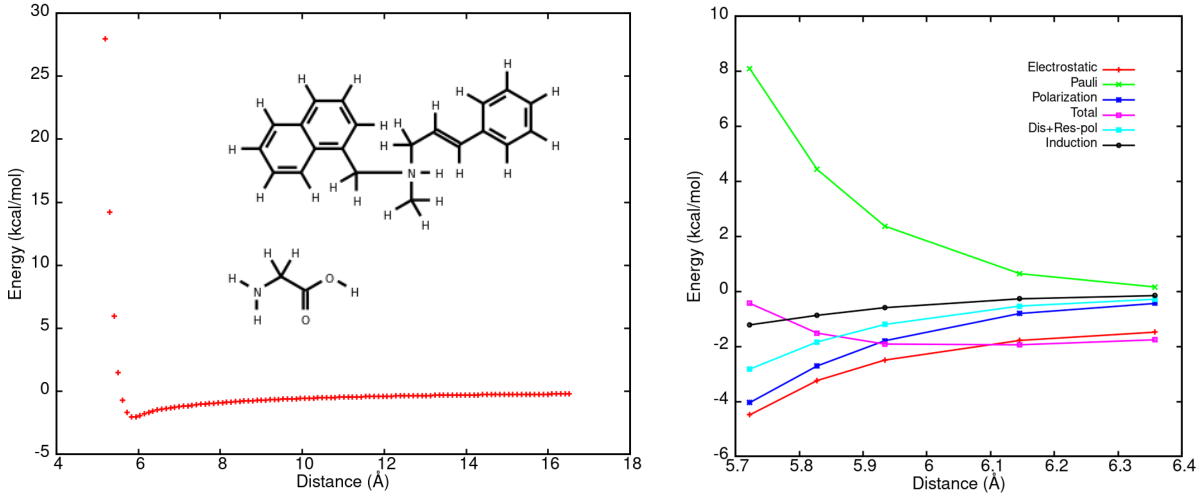


Figure 246: System 246

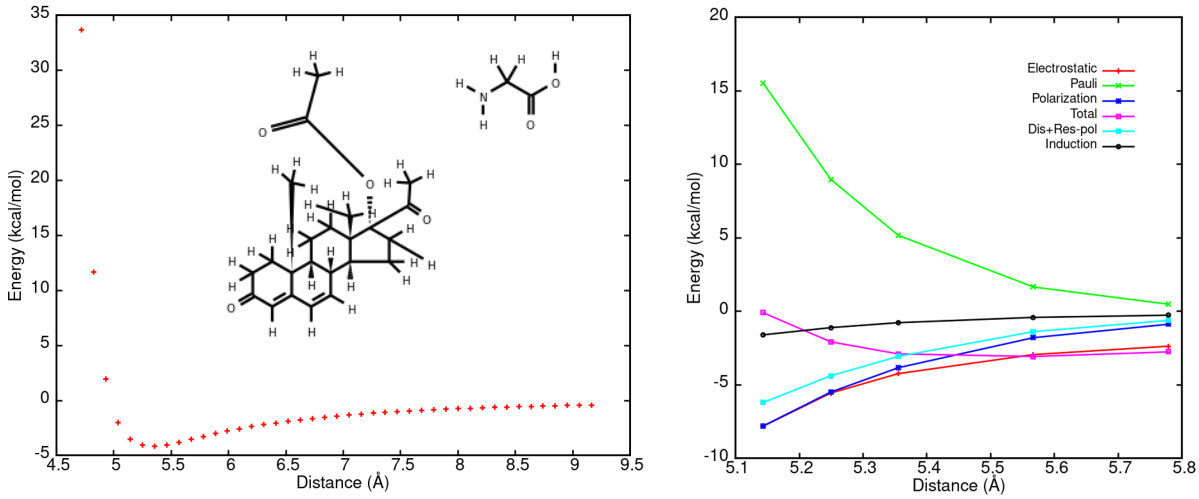


Figure 247: System 247

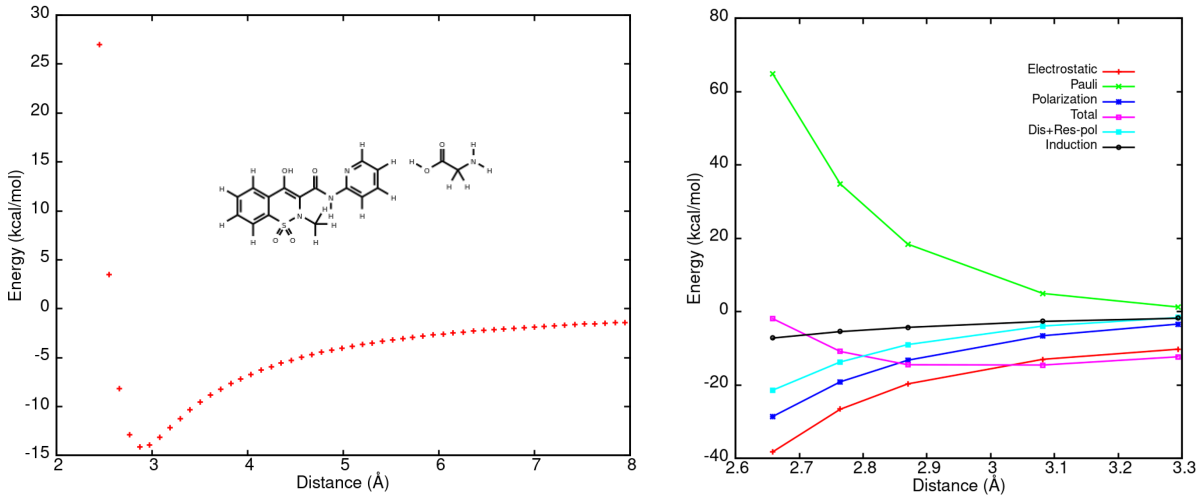


Figure 248: System 248

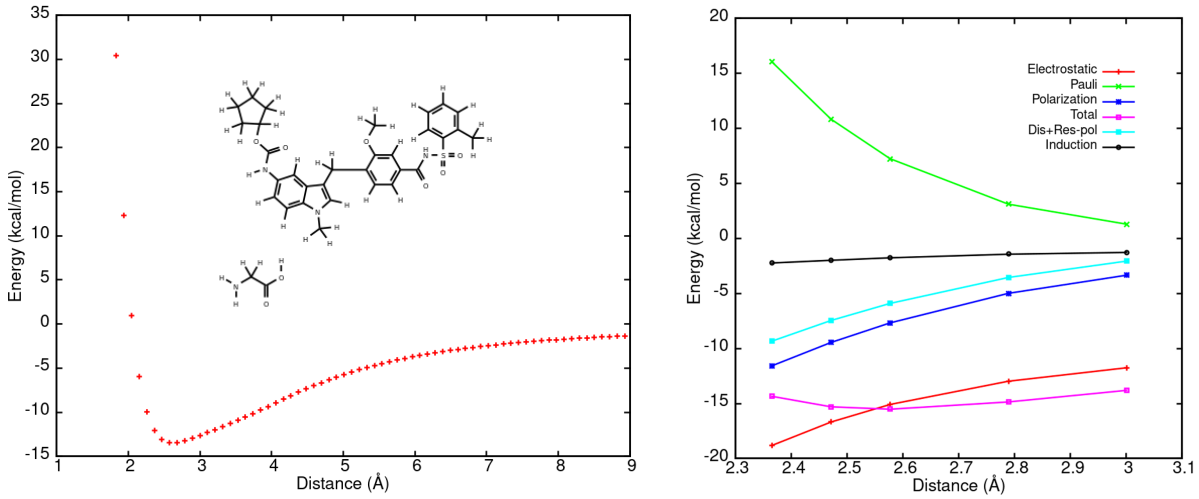


Figure 249: System 249

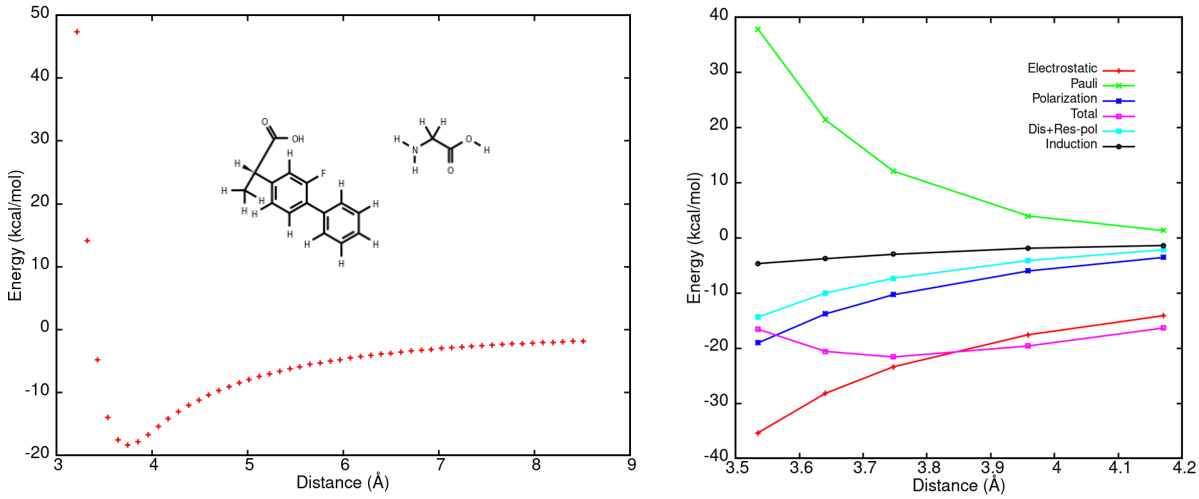


Figure 250: System 250

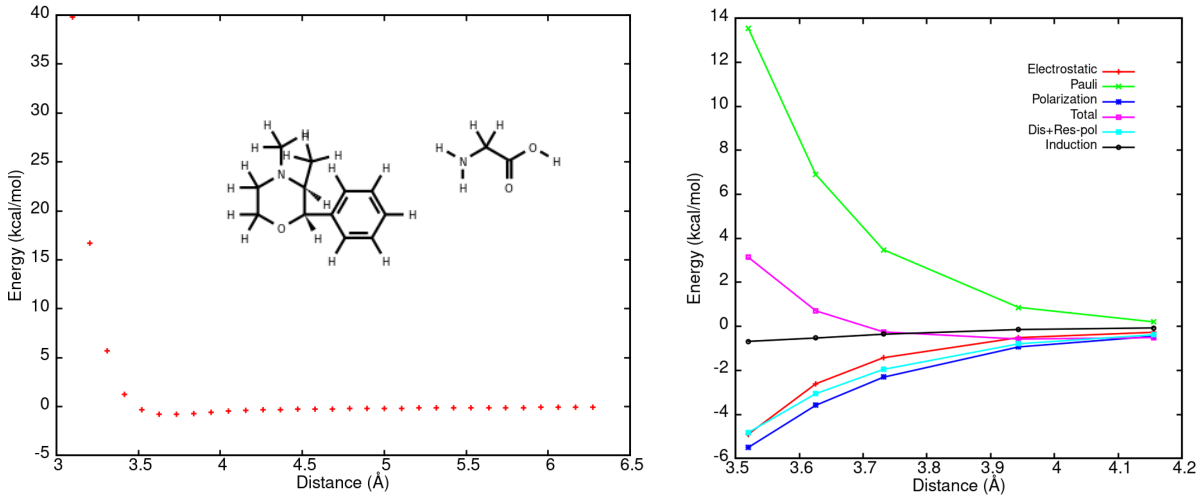


Figure 251: System 251

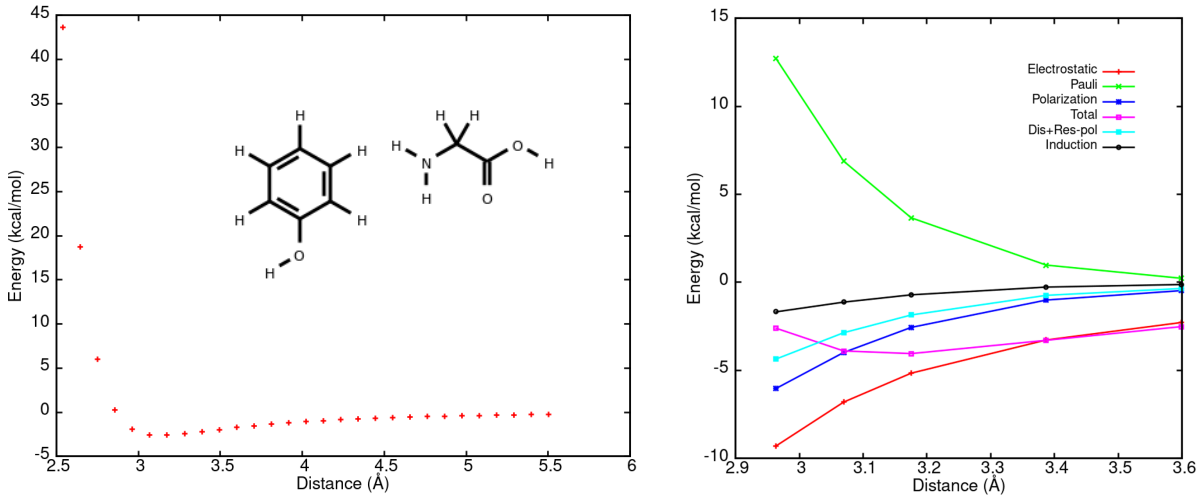


Figure 252: System 252

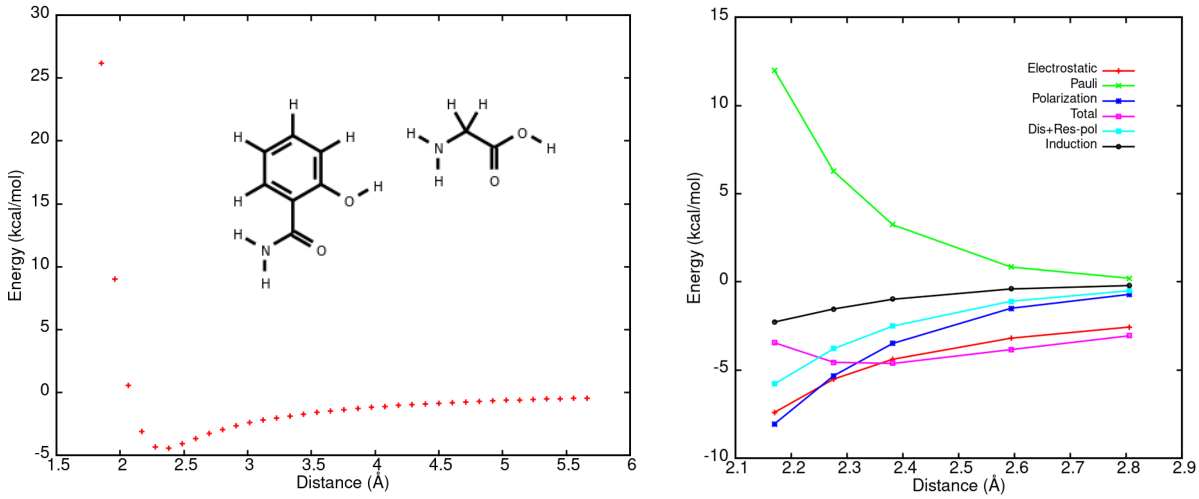


Figure 253: System 253

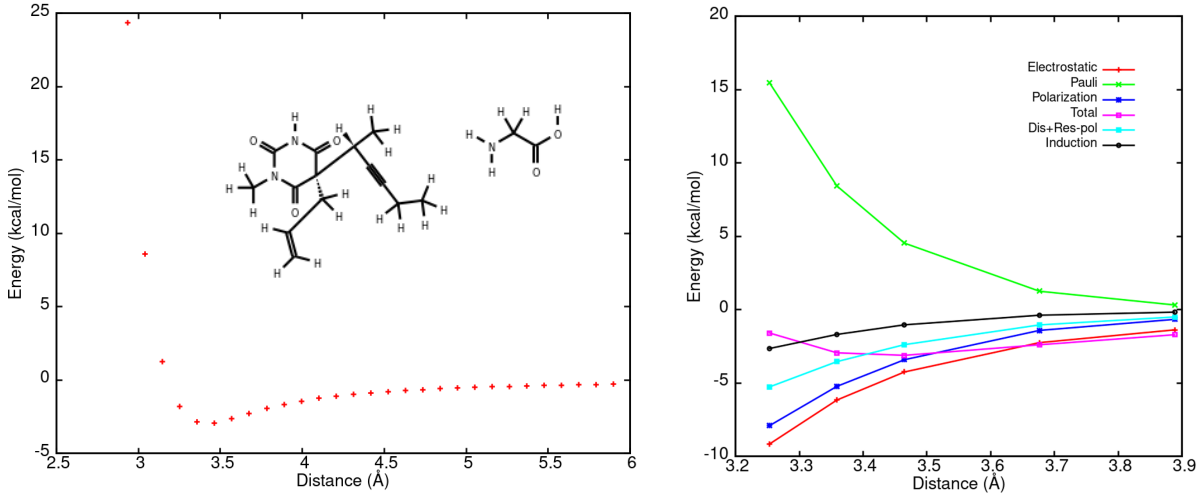


Figure 254: System 254

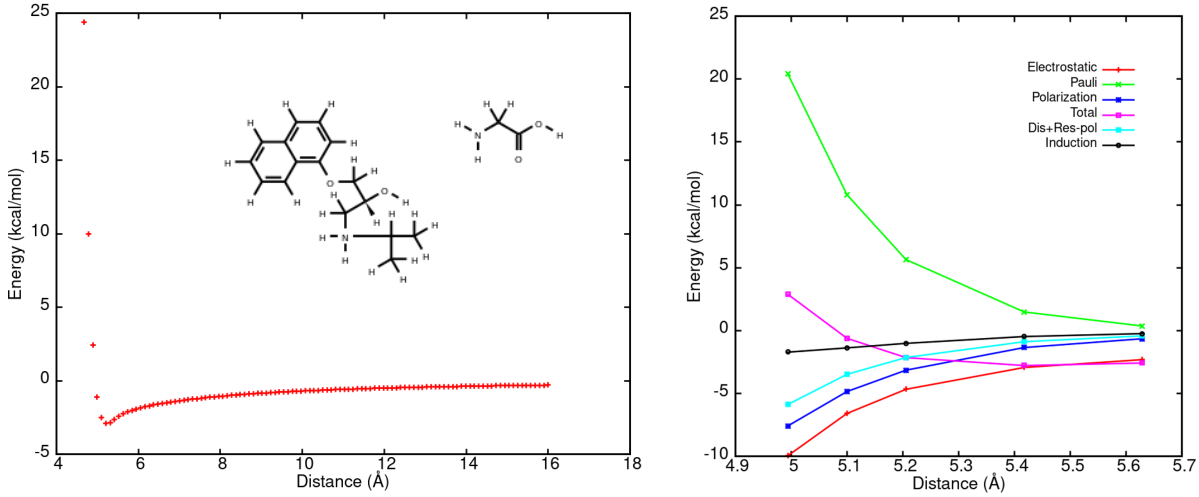
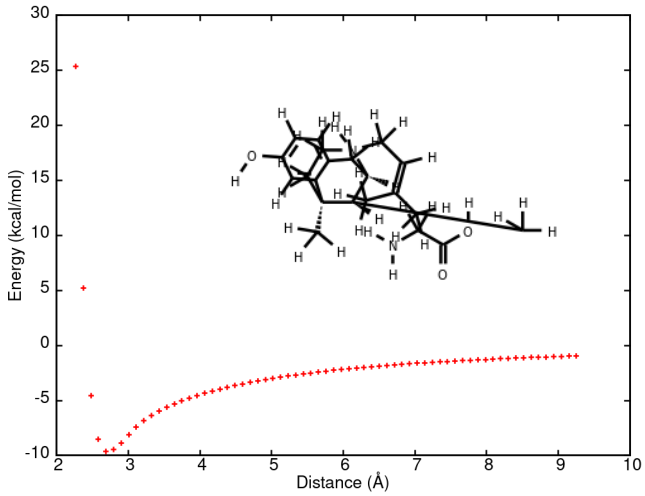


Figure 255: System 255



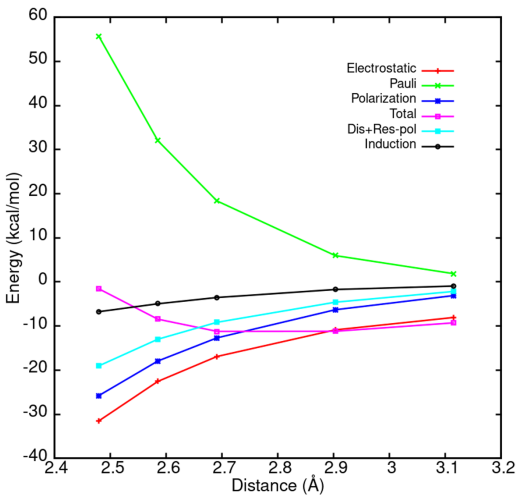
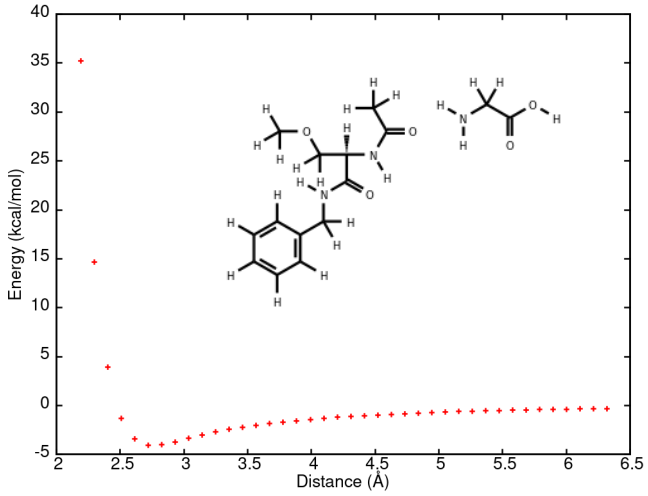


Figure 256: System 256



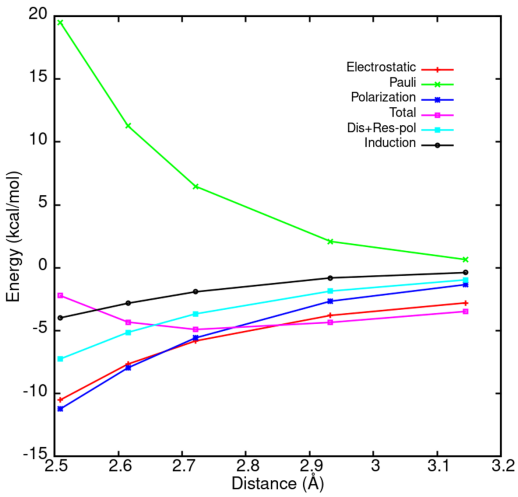
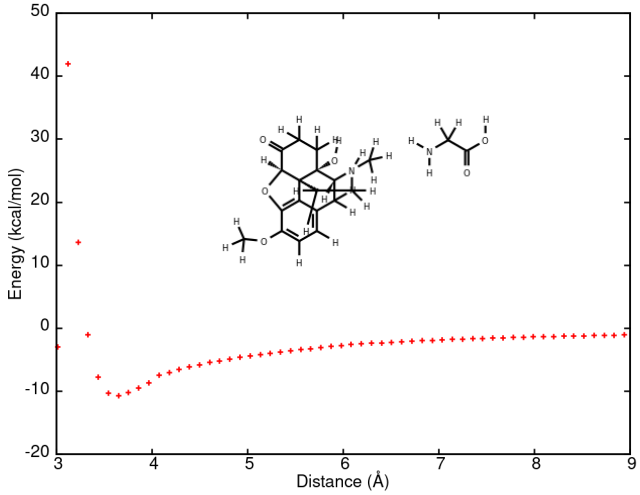


Figure 257: System 257



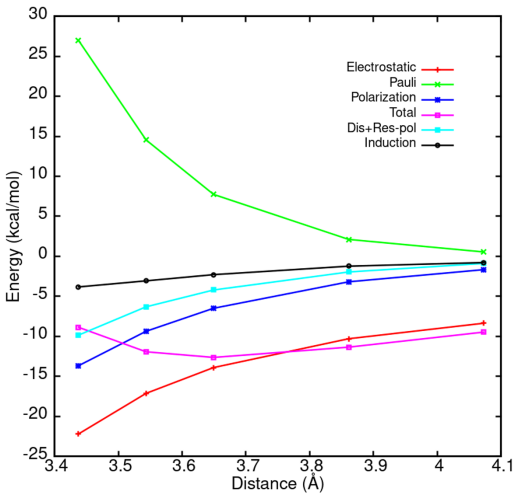
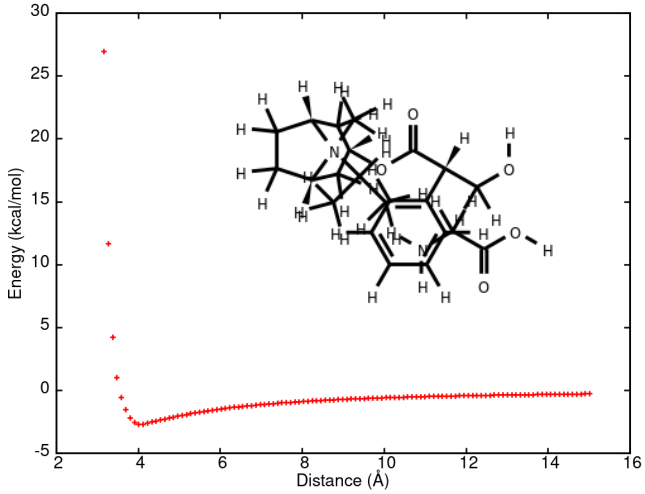


Figure 258: System 258



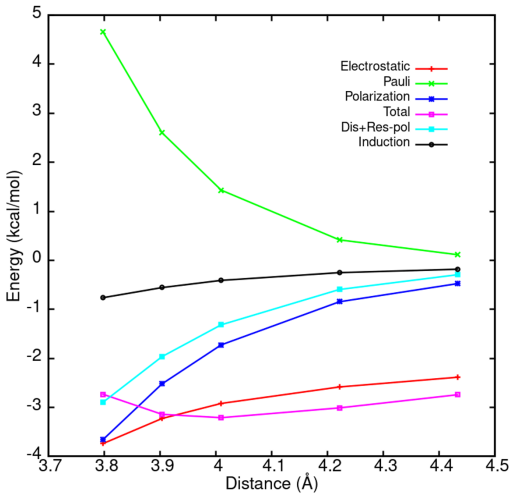
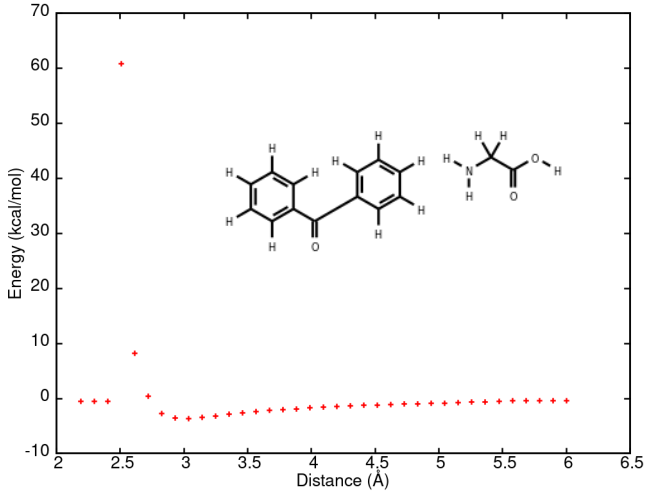


Figure 259: System 259



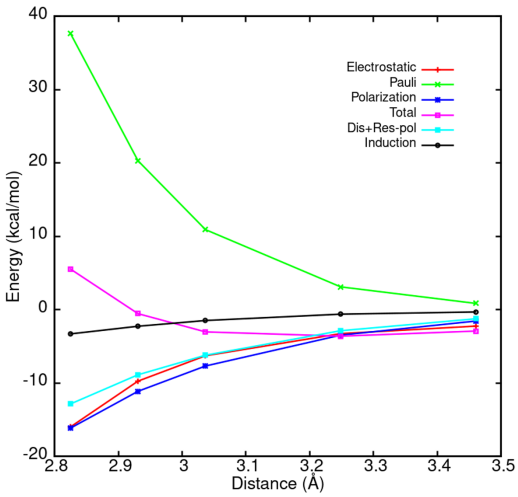
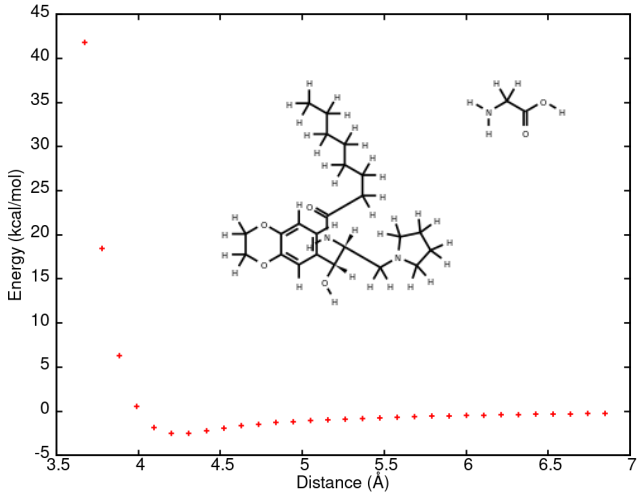


Figure 260: System 260



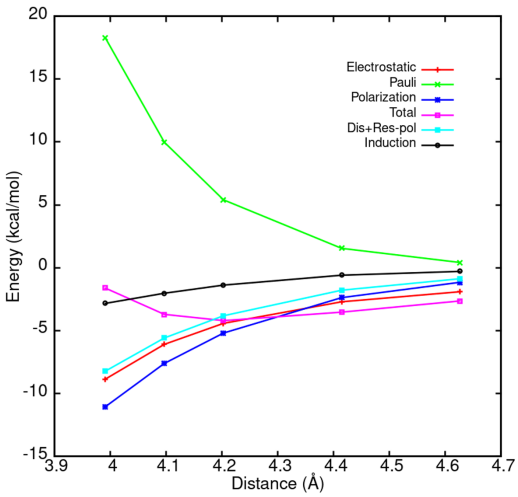
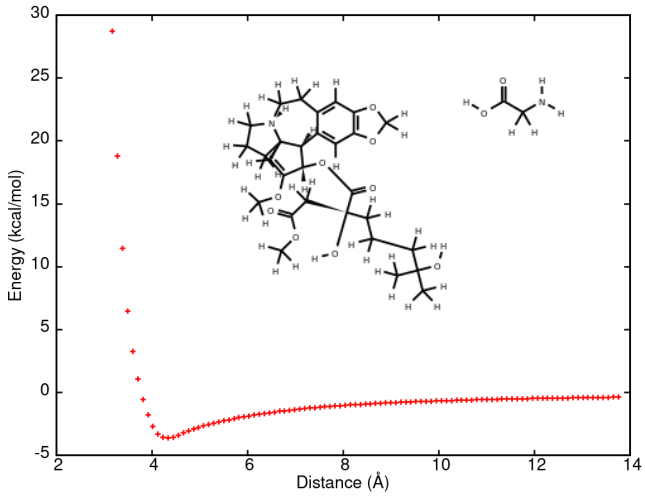


Figure 261: System 261



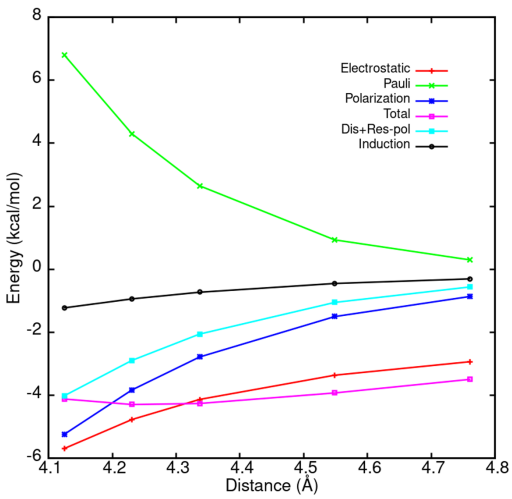
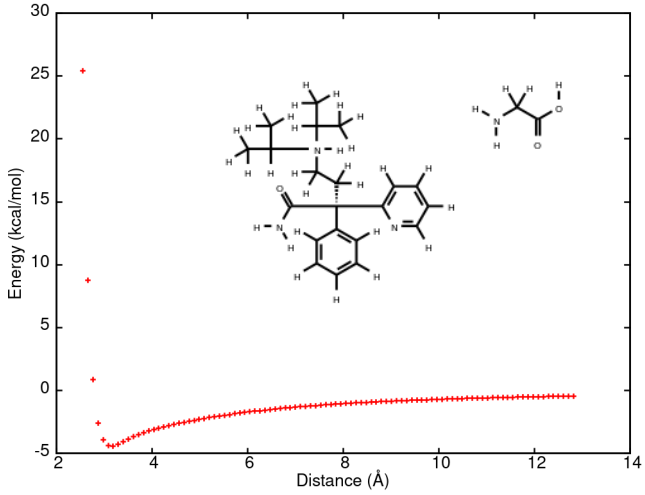


Figure 262: System 262



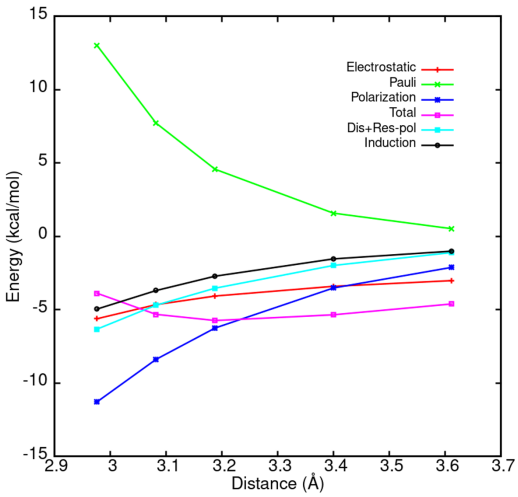
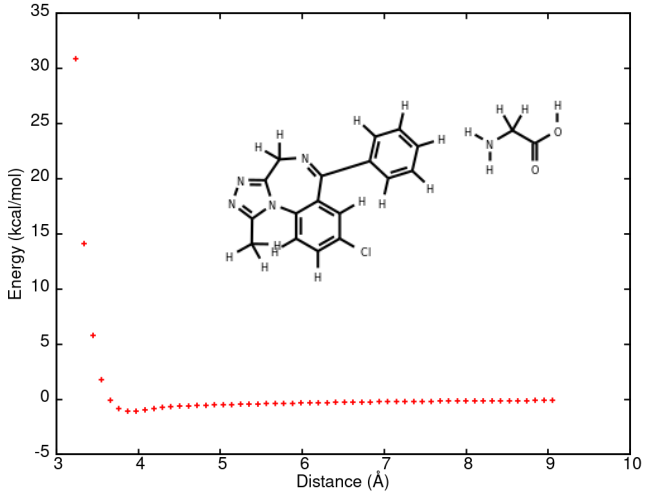


Figure 263: System 263



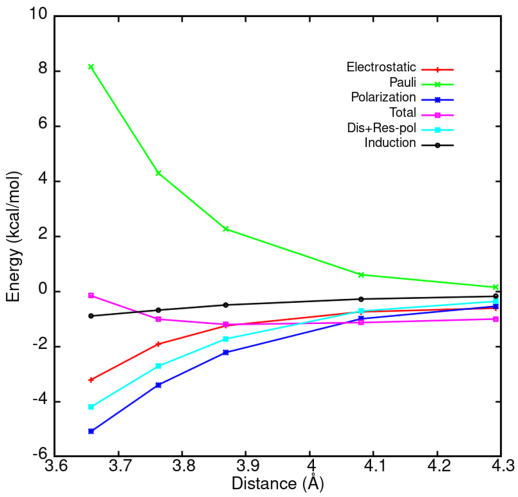


Figure 264: System 264

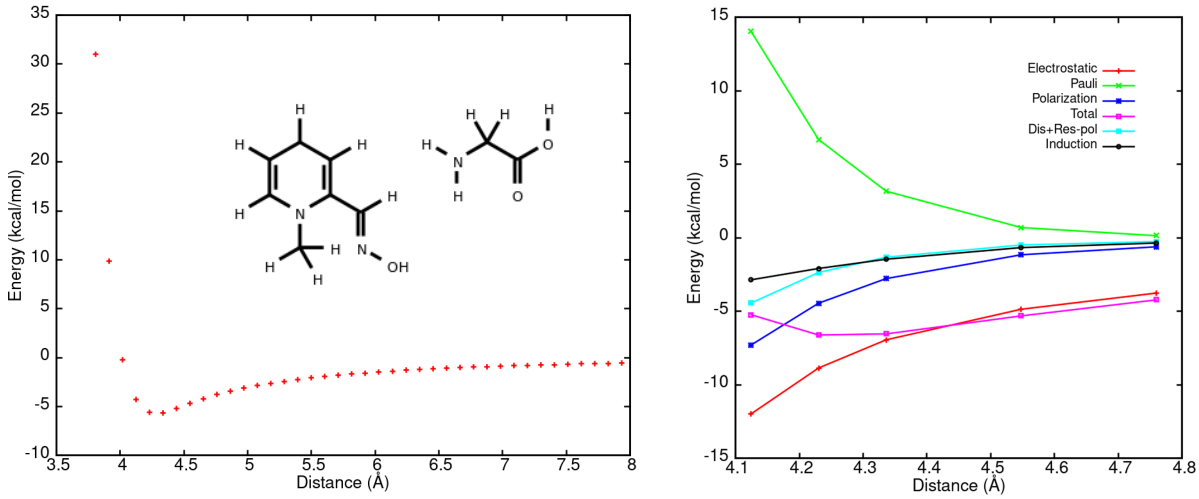


Figure 265: System 265

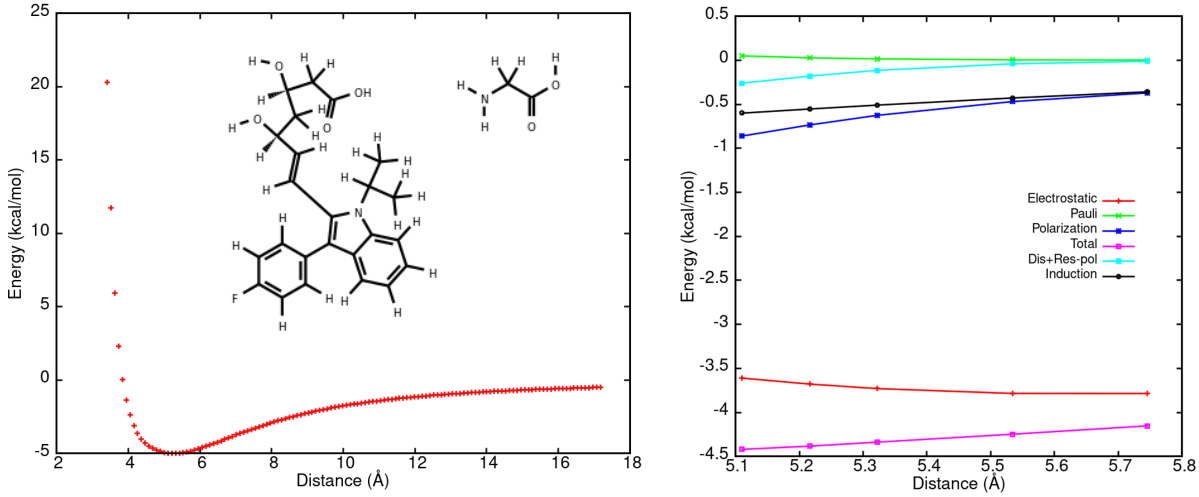


Figure 266: System 266

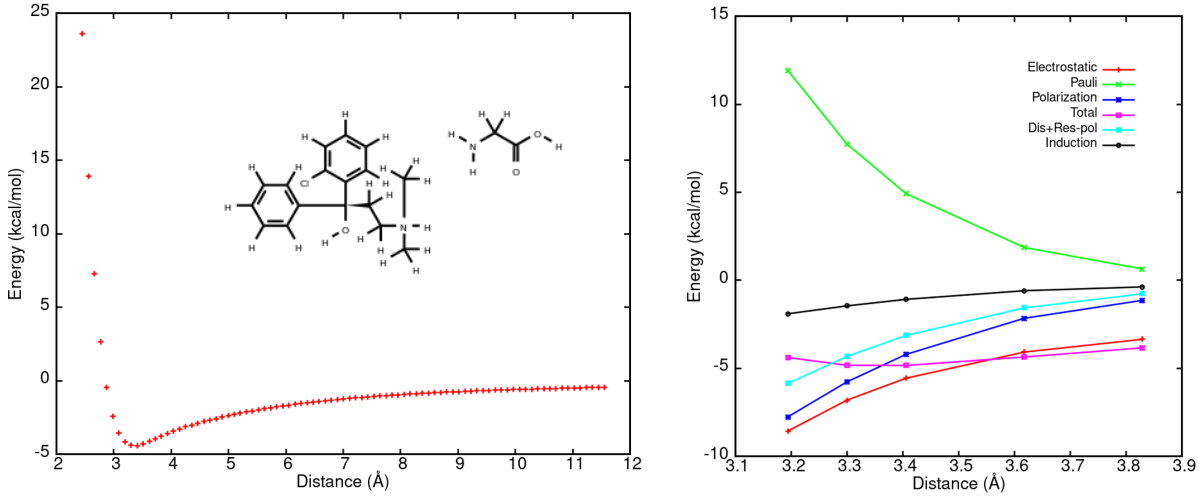
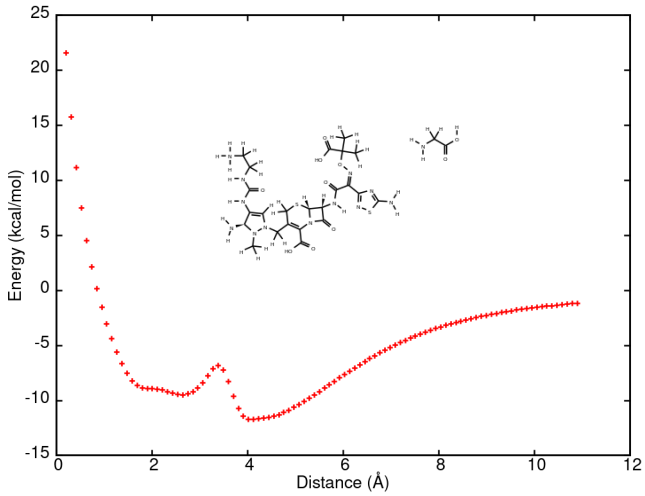


Figure 267: System 267



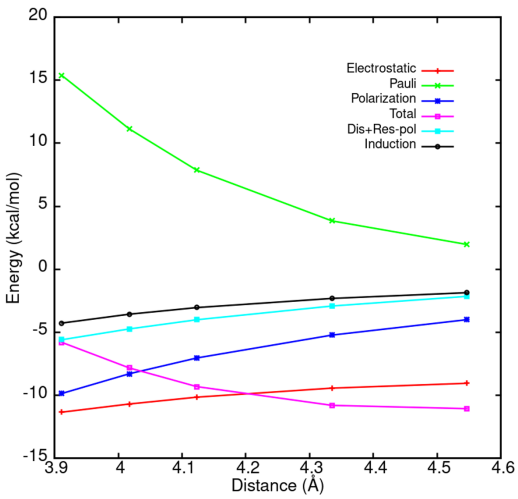
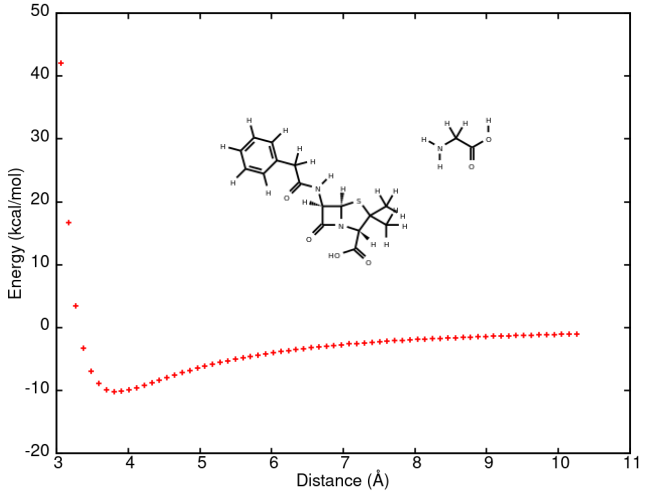


Figure 268: System 268



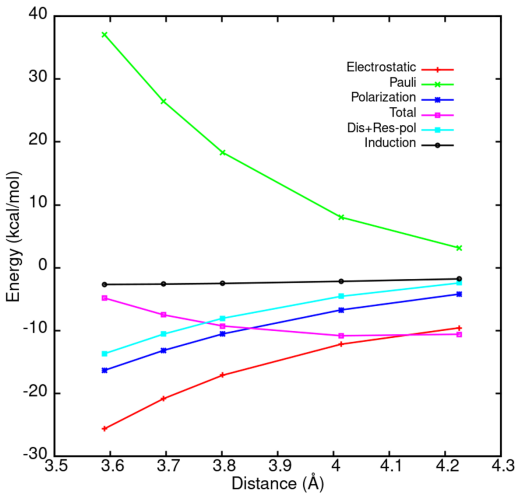
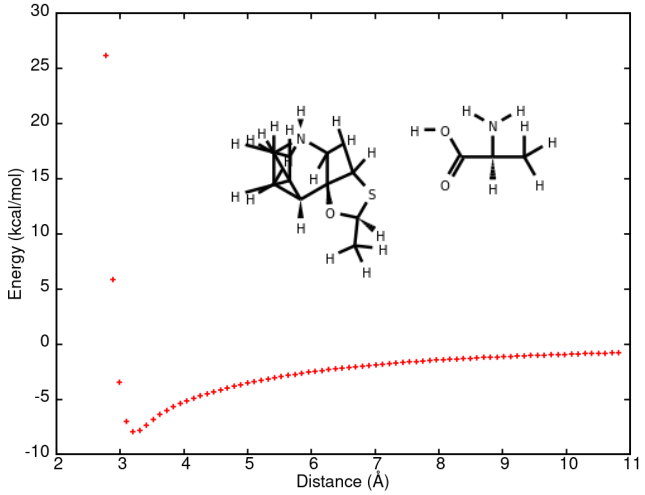


Figure 269: System 269



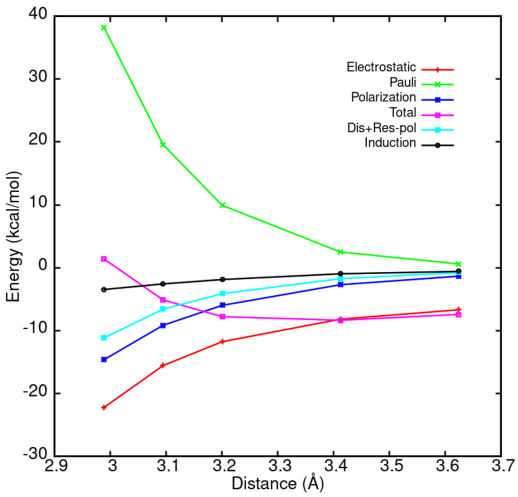
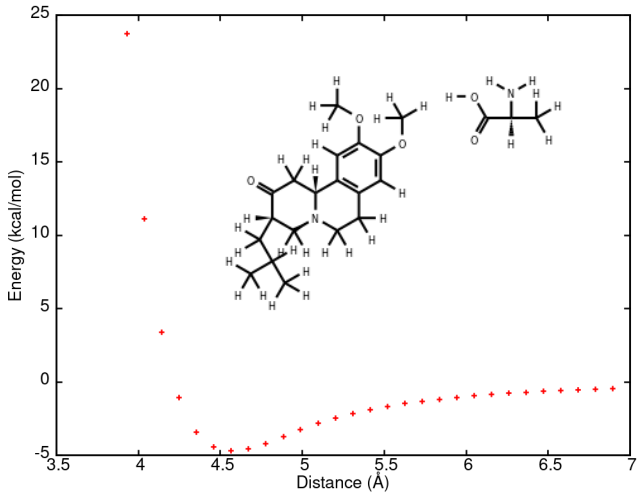


Figure 270: System 270



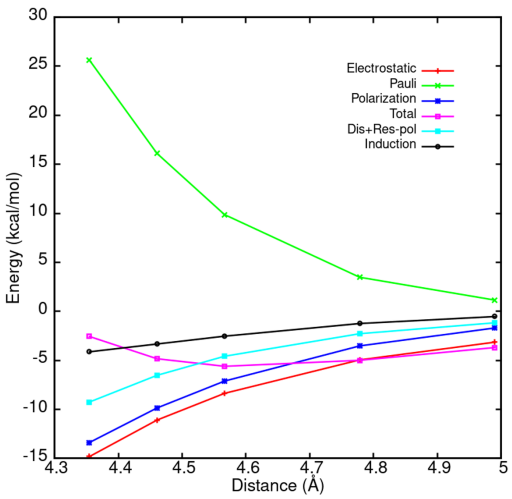
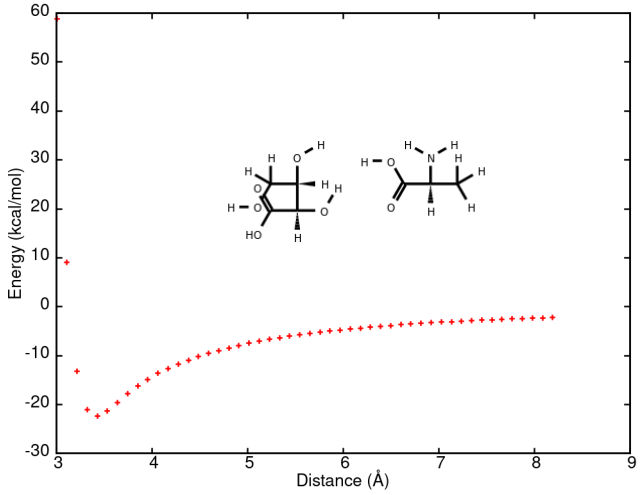


Figure 271: System 271



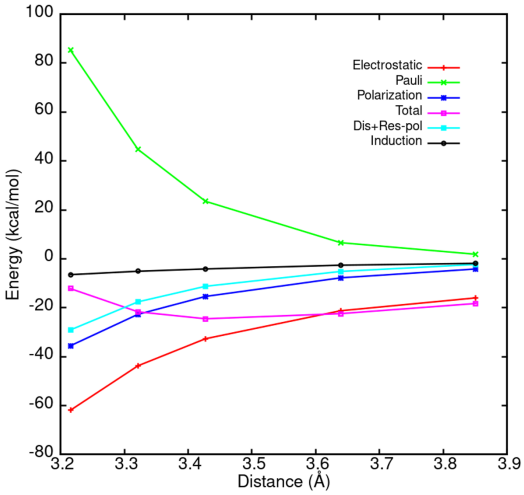
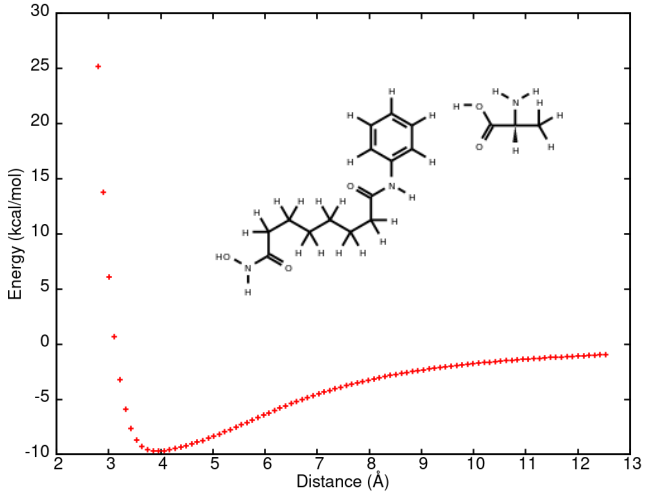


Figure 272: System 272



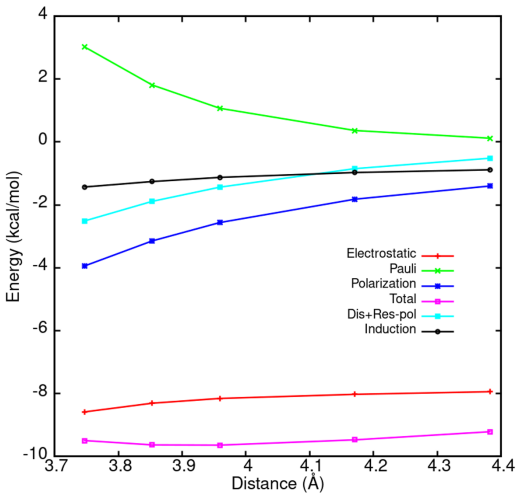
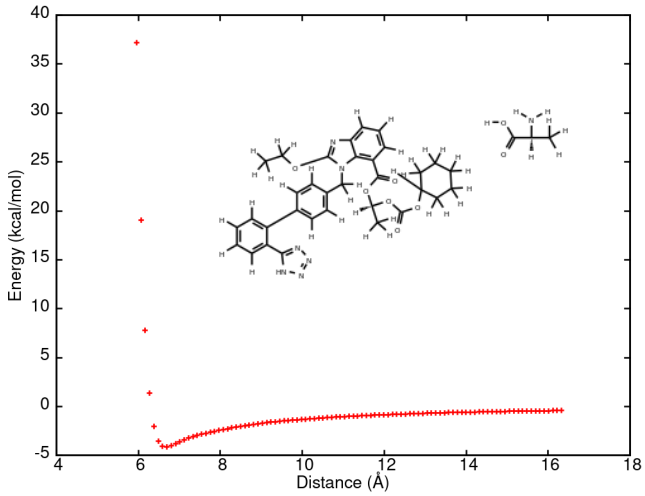


Figure 273: System 273



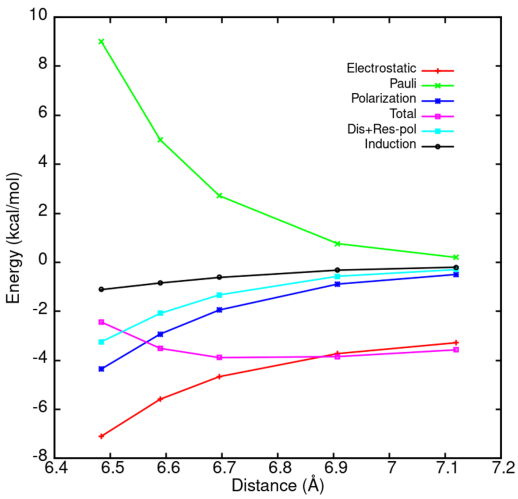
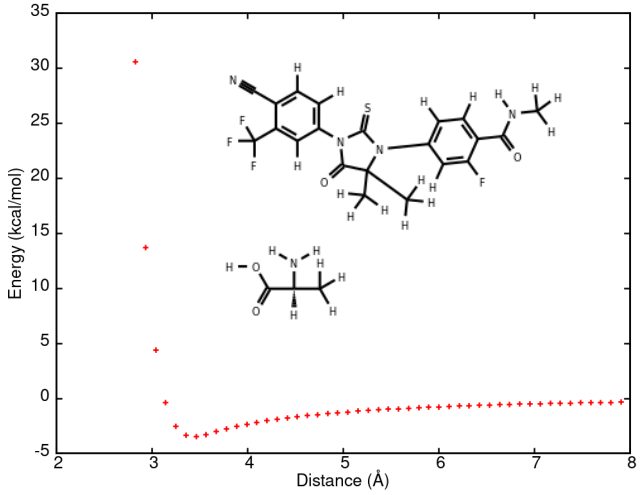


Figure 274: System 274



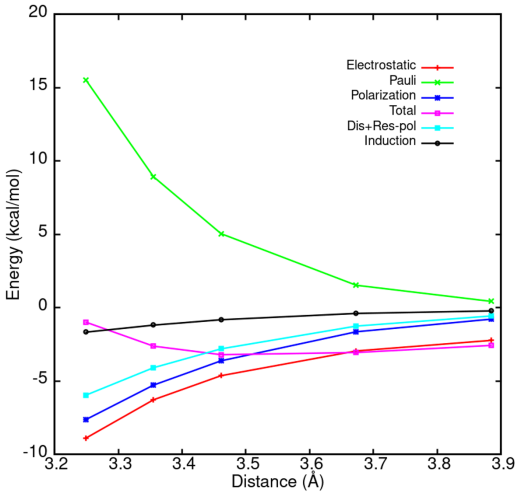
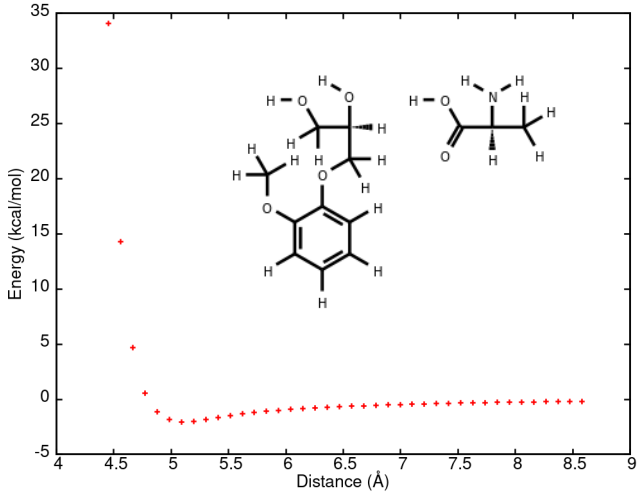


Figure 275: System 275



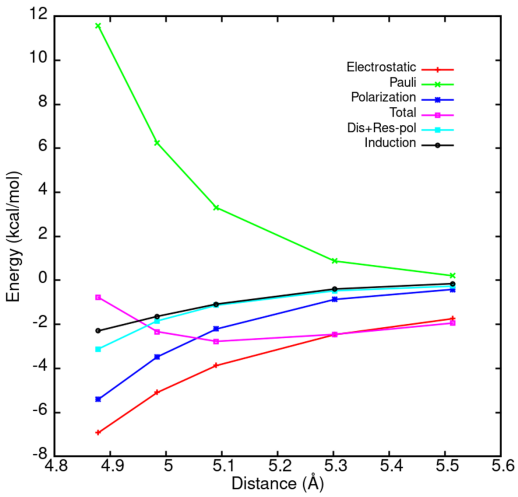


Figure 276: System 276

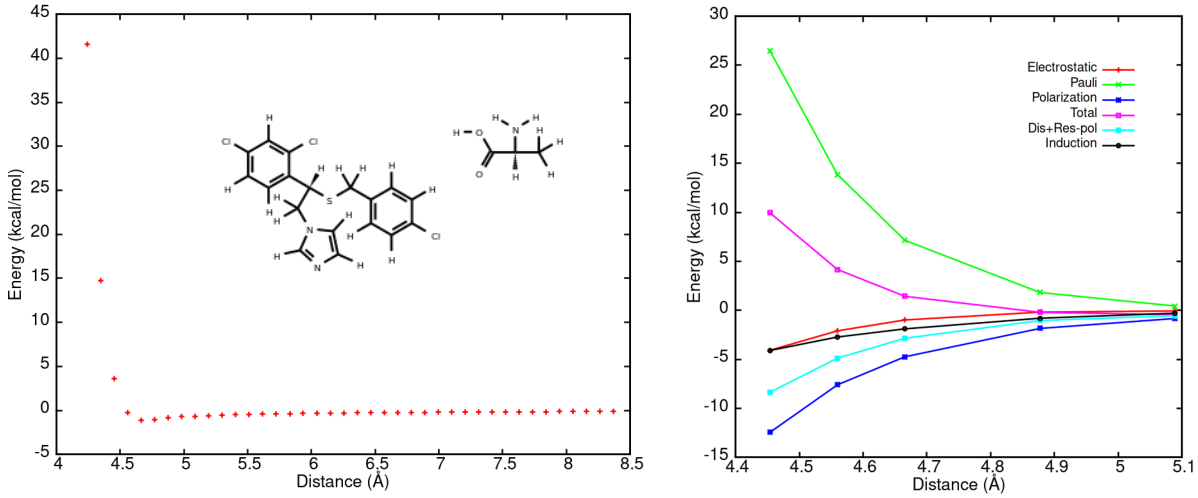


Figure 277: System 277

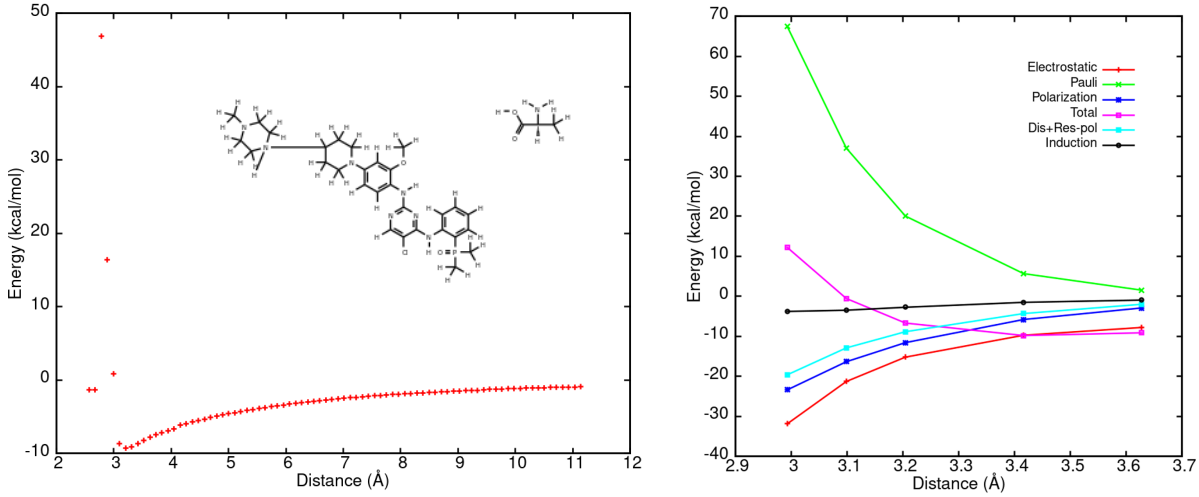


Figure 278: System 278

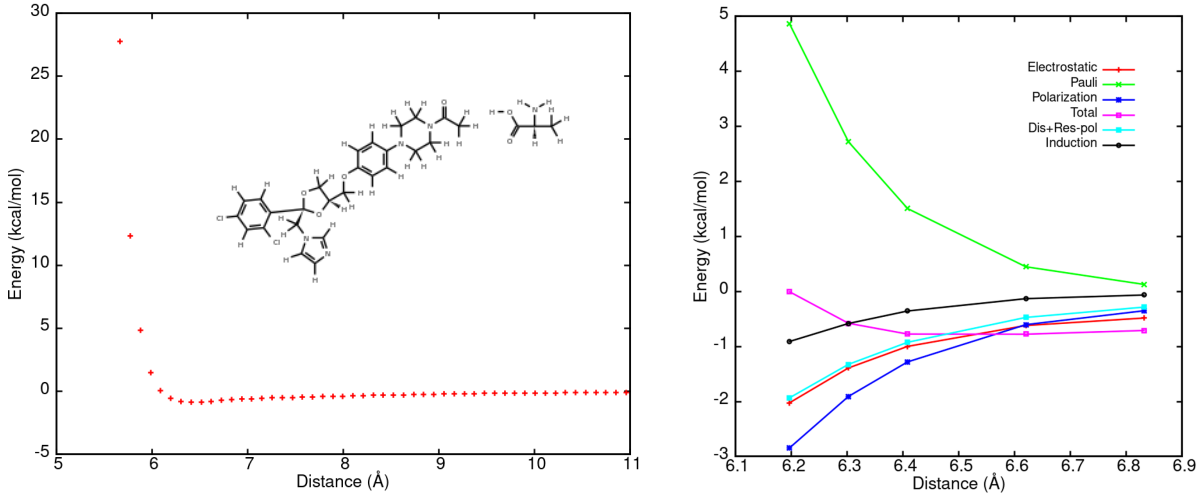
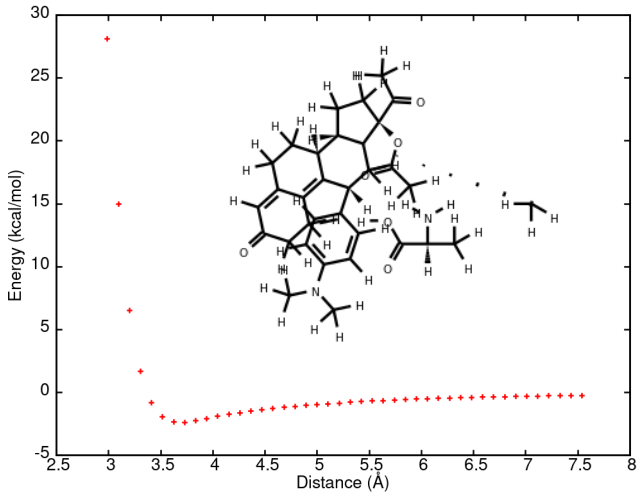


Figure 279: System 279



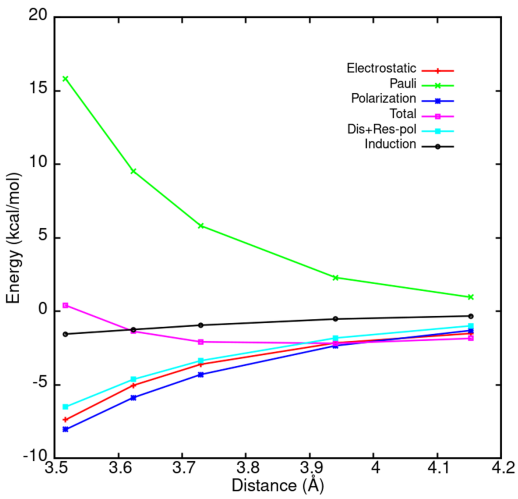
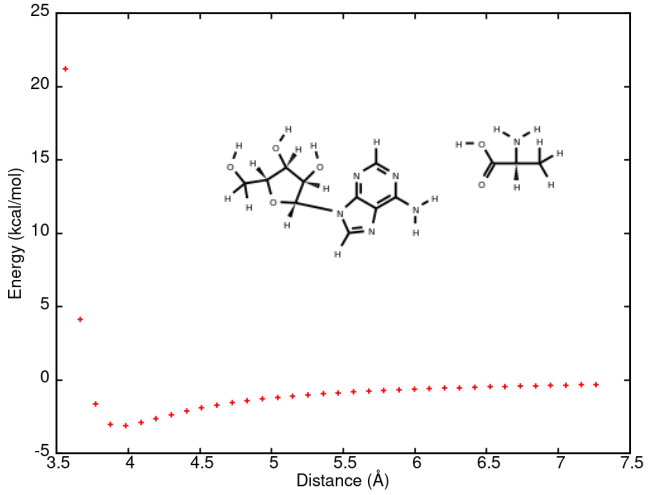


Figure 280: System 280



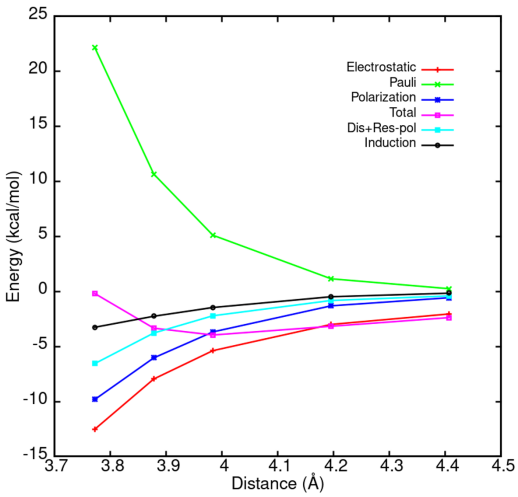
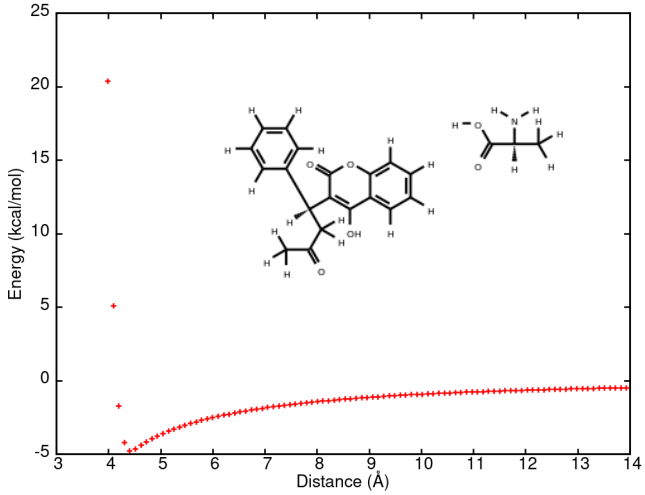


Figure 281: System 281



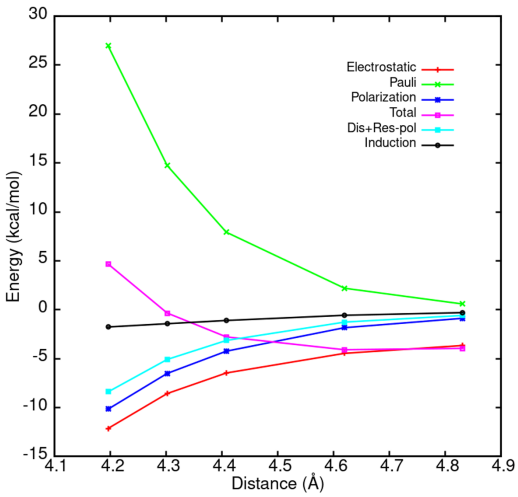


Figure 282: System 282

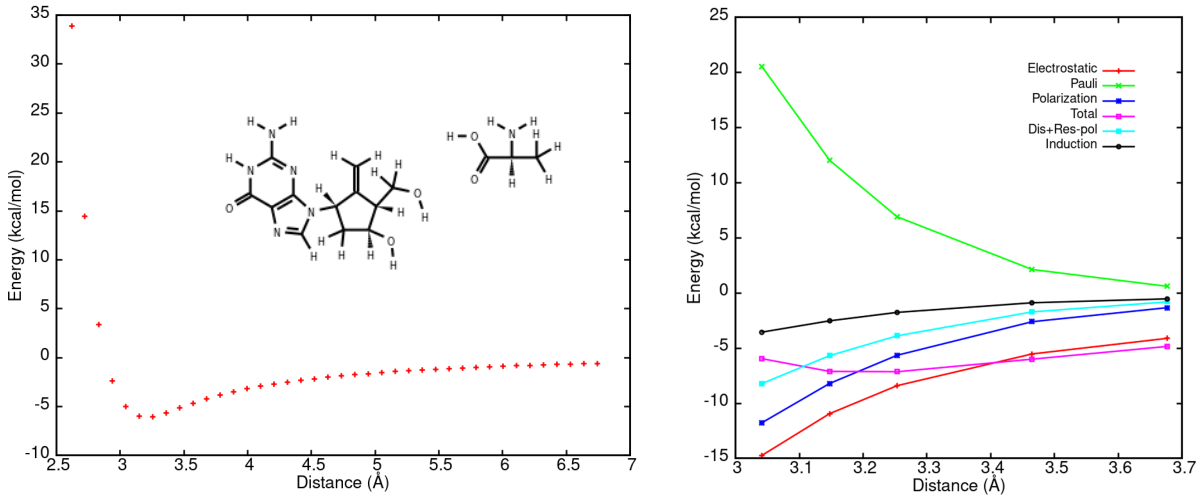


Figure 283: System 283

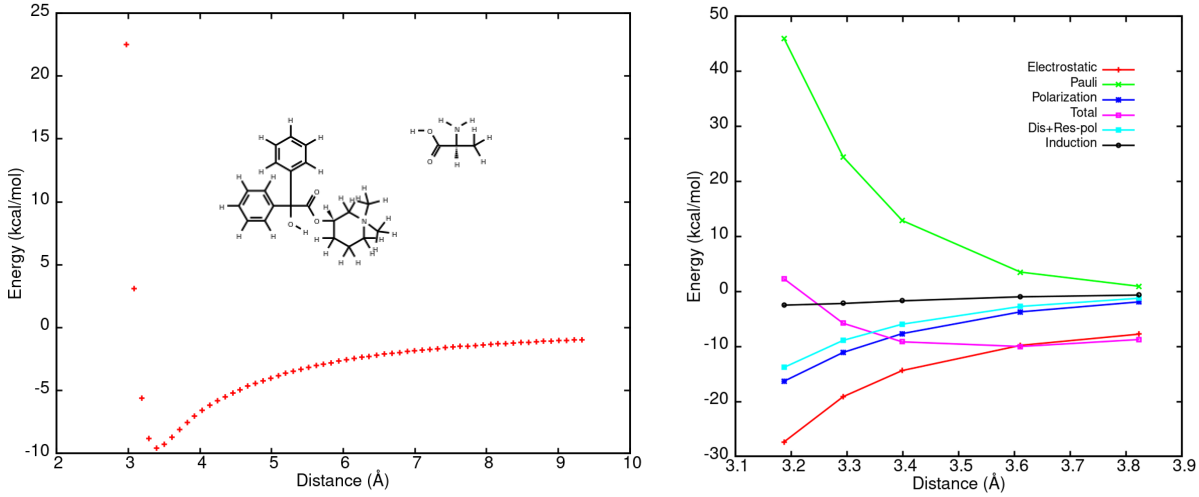


Figure 284: System 284

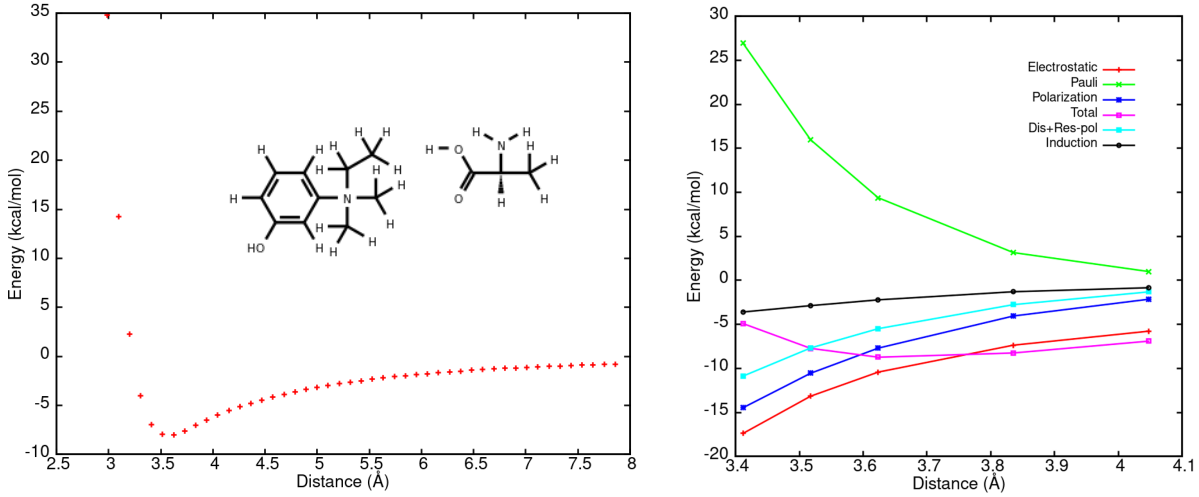
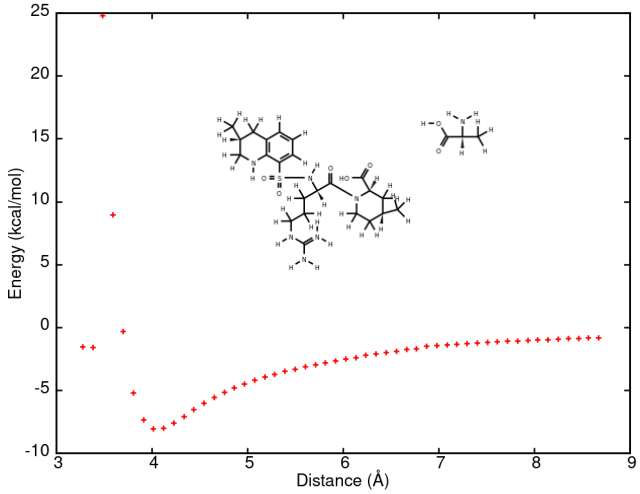


Figure 285: System 285



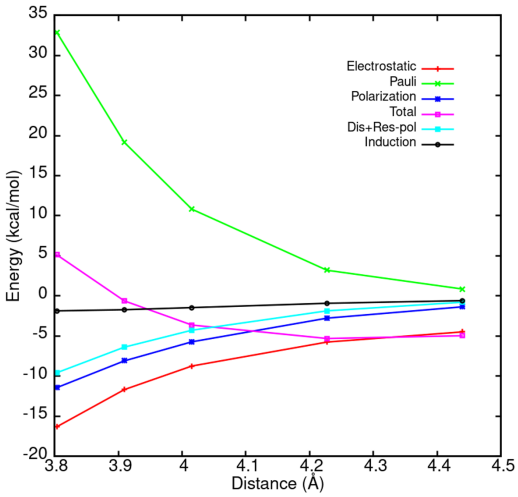
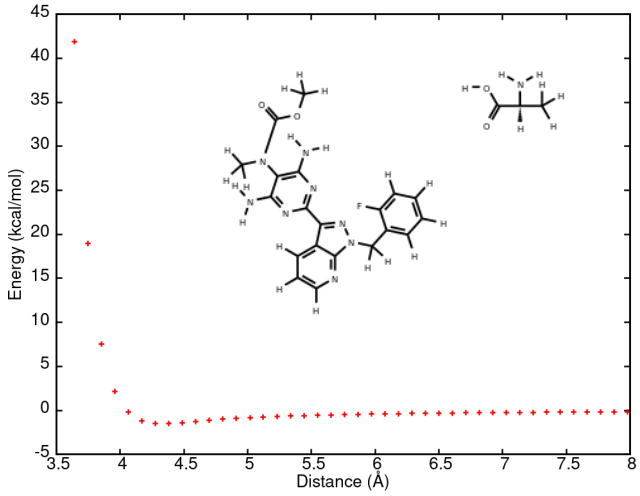


Figure 286: System 286



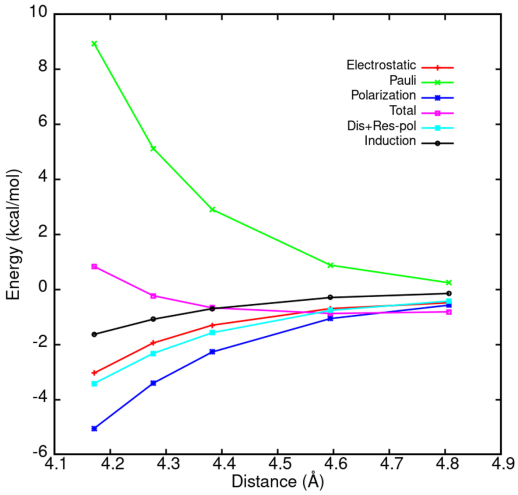
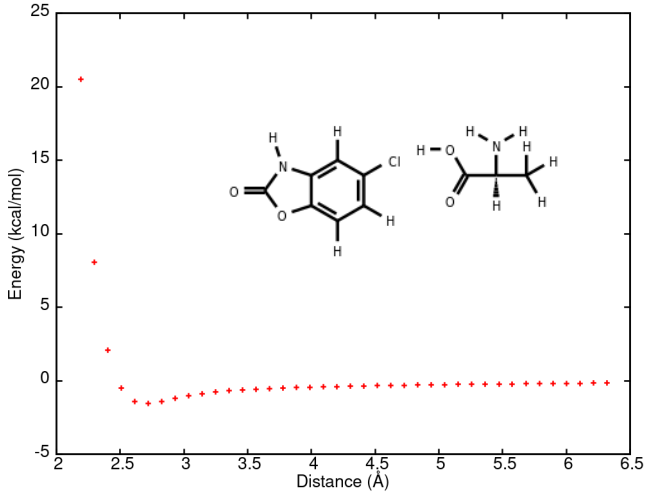


Figure 287: System 287



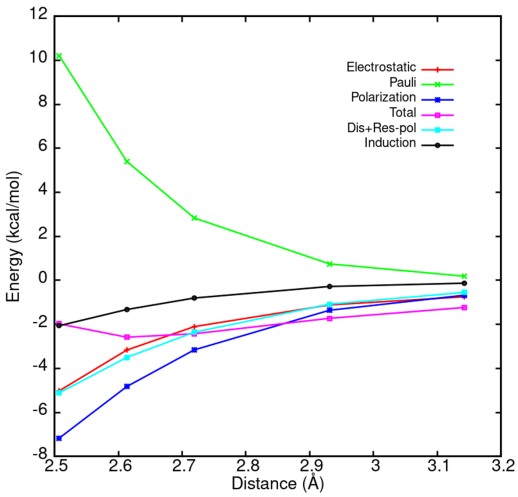
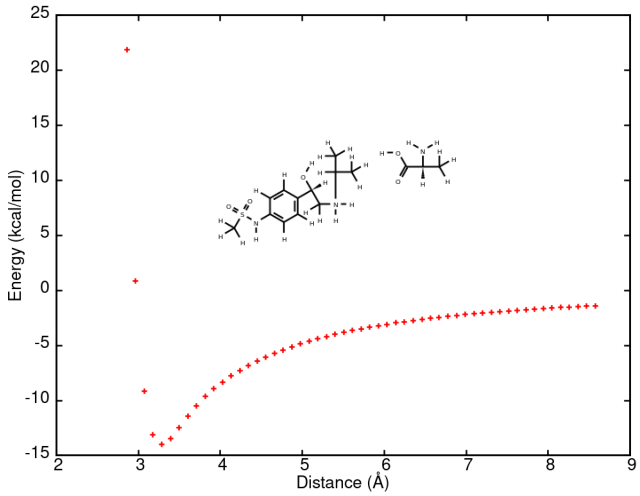


Figure 288: System 288



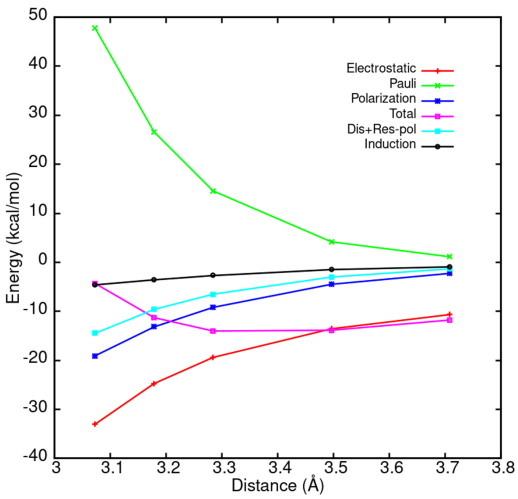
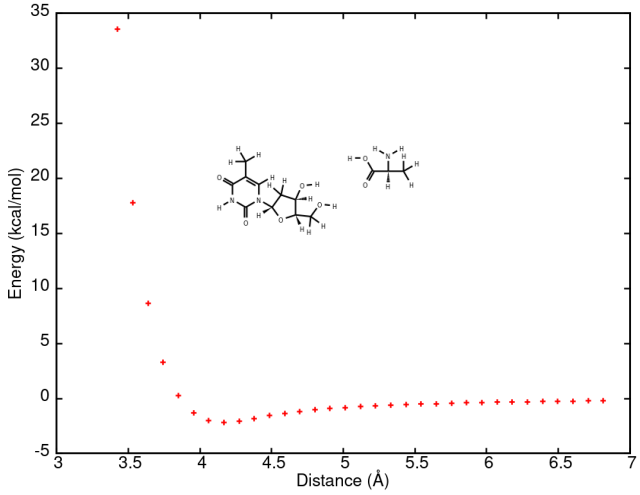


Figure 289: System 289



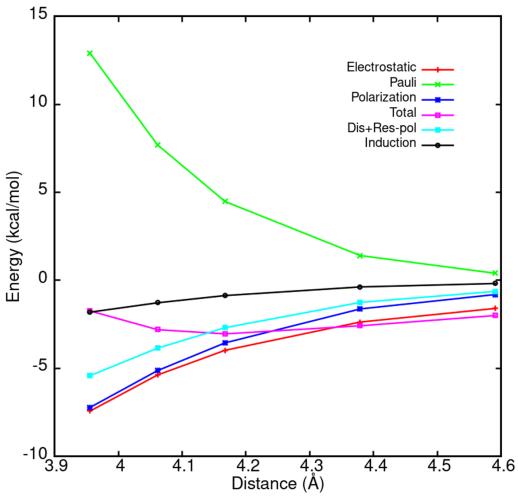
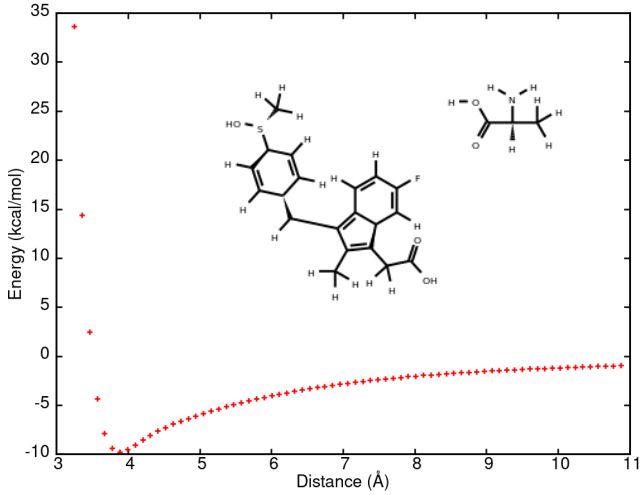


Figure 290: System 290



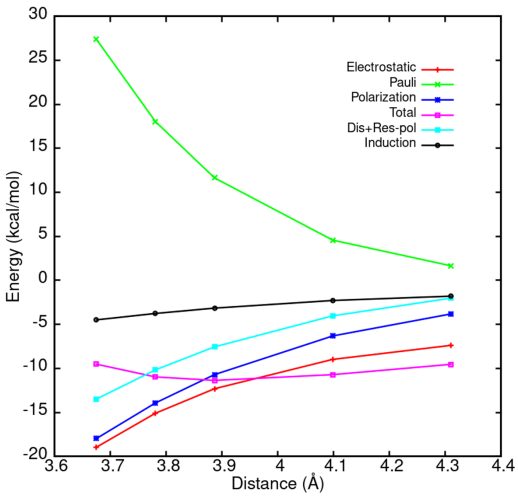


Figure 291: System 291

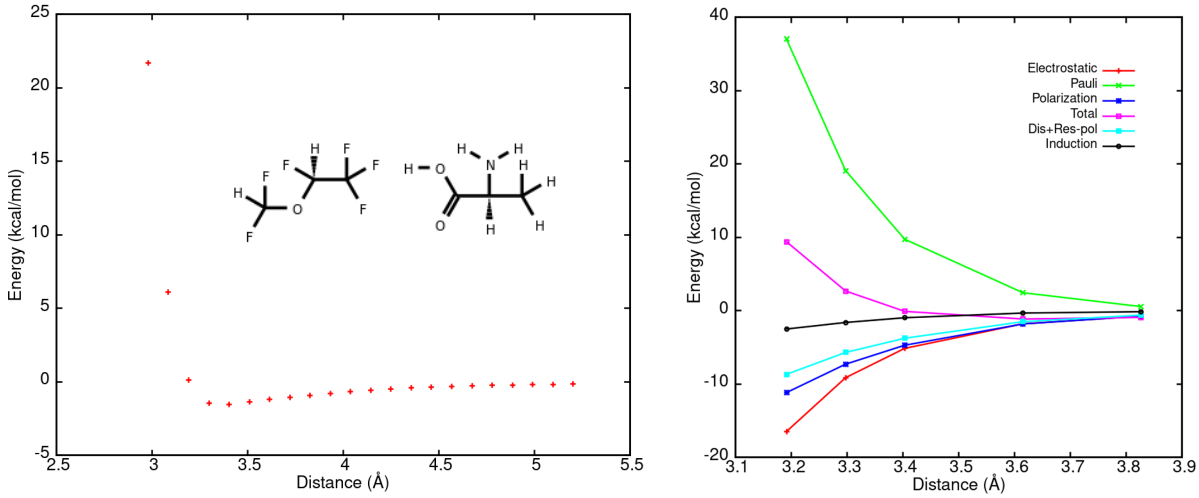


Figure 292: System 292

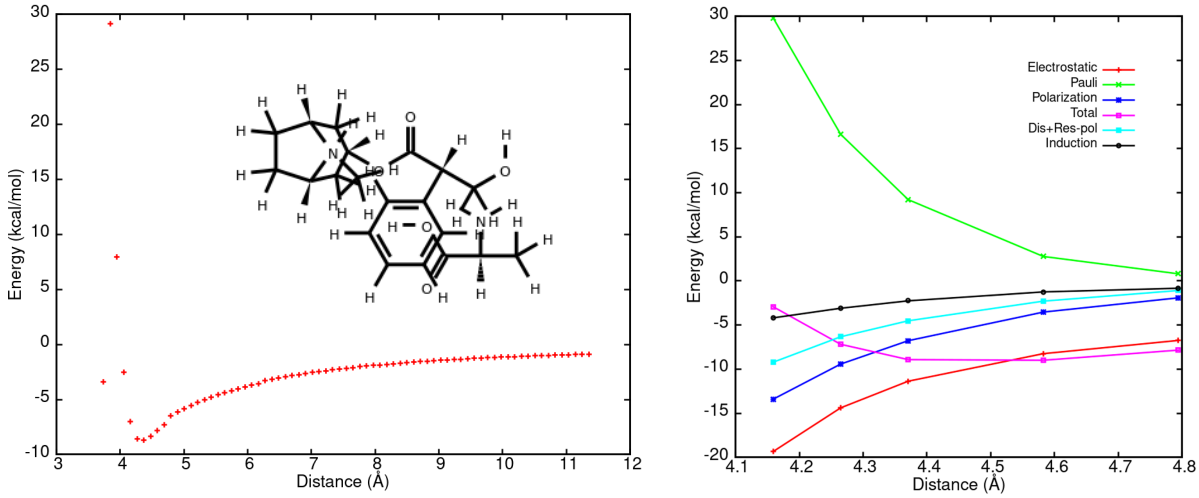


Figure 293: System 293