Use of MP2, DFT and semi-emperical calculations of protein-ligand interaction energies and *ab initio* QSAR in the development of novel statin drugs

Written at Rhodes College

Allison Marie Price

Department of Chemistry

Rhodes College

Memphis, Tennessee

2011

Submitted in partial fulfillment of the requirements for the

Bachelor of Science degree with Honors in Chemistry

This Honors paper by Allison Price has been read

and approved for Honors in Chemistry.

Dr. Mauricio Cafiero

Project Advisor

_____

Dr. Carolyn Jaslow

Second Reader

_____

Dr. Shubho Banerjee

Extra-Departmental Reader

_____

Dr. Darlene Loprete

Department Chair

_____

CONTENTS

Signature page	ii
Contents	iii
List of Tables and Figures	iv
Abstract	vi
Authorization page	vii
Introduction	1
Methods	12
Results and Discussion	17
Conclusions	27
Works Cited	28

Figure 1. Novel tail fragment in complex with Tyr479 residue from the HMG-CoA reductase active site. 11

Figure 2. Rosuvastatin candidates 1-4 as docked in the HMG-CoA reductase active site (above). Simvastatin candidates 1-3 as docked in the HMG-CoA reductase active site (below). Figures illustrate interaction of tail moiety with Tyr479 and the interaction of the statin moiety with other amino acid residues in the active site. 14

Figure 3. Secondary generations candidates, G2-G5, as docked in the HMG-CoA reductase active site. Figures illustrate interaction of tail moiety with Tyr479 and the interaction of the statin moiety with other amino acid residues in the active site. 15

Table 1. Molecular properties (calculated with HCTH/ 6-31g) for on-the-market and theoretical statins to be correlated with experimental potencies, pIC50 values. 18

Table 2. Interaction energies of substrate, drugs and drug candidates with the HMG-CoA reductase active site. Energies are in kcal/mol and were calculated with a basis set of 6-311+g*. *This value is not calculated in this work, see reference. 20

Table 3. Interaction energies for Generation 2 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*. 22

Table 4. Interaction energies for Generation 3 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*. 22

Table 5. Interaction energies for Generation 4 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*. 23

Table 6. Interaction energies for Generation 5 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*. 24

Table 7. Molecular properties (calculated with HCTH/6-31g) for theoretical statins which will be used to predict experimental potencies, pIC50 values. 24

Figure 4. NADP(H) bound in the allosteric site of HMG-CoA reductase, taken from the crystal structure. 26

Table 8. Interaction energies for NADPH in the HMG-CoA reductase allosteric site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*. 26

ABSTRACT

Use of MP2, DFT and semi-emperical calculations of protein-ligand interaction energies and *ab initio* QSAR in the development of novel statin drugs

by

Allison Marie Price

Previous work (*J. Phys. Chem. B*, 113, 14810, 2009) has shown that the residue Tyr479 in the active site of 3-hydroxy-3-methyglutaryl-coenzyme A (HMG-CoA) reductase exerts a strong attraction on ligands. Statin drugs moderate blood cholesterol levels by acting as competitive inhibitors of HMG-CoA reductase, blocking the biosynthesis of cholesterol early in the synthesis pathway. In this work, novel molecular fragments that bind strongly to all areas of the active site have been developed using *ab initio* correlated methods and attached to known statin drugs to create novel drug candidates that interact more strongly with the enzyme than the original drugs. Interaction energies between small molecule ligands and the target enzyme active site are calculated with all-electron correlated methods such as MP2 and DFT, as well as semi-empirical methods. Various molecular properties of known statin drugs were also correlated to the values of the *in vivo* potencies (the pIC50) of each drug in order to derive quantitative structure/activity trends. Based on the above work a second generation of drug candidates has been designed to be more specific for the total HMG-CoA reductase active site. Further, Tyr479 has been found to be present in an allosteric binding site, which upon analysis, appears to be a promising site for uncompetitive inhibition of HMG-CoA. The calculations show that the novel drugs candidates developed in this work will be effective inhibitors of both the active and the allosteric site.

I give permission for public access to my Honors paper and for any copying or digitization to be done at the discretion of the College Archivist and/or the College Librarian.

Signed_____

Allison Price

Date _____

1 Introduction

The purpose of this work is to develop novel inhibitors of cholesterol synthesis. Development of cholesterol inhibitors is important because pathology due to excess levels of cholesterol is becoming more and more prevalent in today's society. Computational chemistry is used to design novel statin drug candidates, and model how effective these candidates may be *in vivo* through quantum chemical calculations.

1.1 Cholesterol and Statins

Cholesterol ($C_{27}H_{46}O$) is a prevalent compound in the human body which can be ingested as well as synthesized *in vivo*[1]. Free cholesterol in the blood is necessary for the synthesis of cell membranes and the production of hormones, and comes from the hydrolysis of lipoproteins[1].

Nelson and Cox discuss that the synthesis of cholesterol begins with the basic two-carbon structure, acetyl-CoA. First, two acetyl-CoA units condense to form acetoacetyl-CoA, which then condenses with a third acetyl-CoA forming HMG-CoA (β-hydroxy-β-methylglutaryl-CoA). HMG-CoA is converted to mevalonate via the enzyme HMG-CoA reductase and the donation of four electrons by two NADPH. The change to mevalonate is the irreversible ("committed") rate-limiting step of the reaction series that ultimately leads to the formation of cholesterol[2].

Nelson and Cox also discuss that because cholesterol synthesis is a complex and energy expensive process, it is supposed to be tightly regulated. In healthy cells intracellular cholesterol levels regulate cholesterol synthesis. If levels of cholesterol in the cells are high, cholesterol synthesis is stemmed. Hormonal regulation of cholesterol synthesis occurs through the phosphorylation of HMG-CoA reductase. When glucagon is

secreted by the pancreas it phosphorylates the enzyme and inactivates it. Insulin has the opposite effect when secreted and promotes the dephosphorylation of the enzyme reactivating it[2].

While manufactured in the liver, most cholesterol does not stay there. As Nelson and Cox summarize, some cholesterol is used in the liver to form the membranes of hepatocytes, but most of it is exported as biliary cholesterol, bile acids, or cholesteryl esters. Cholesteryl esters are formed when a fatty acid is removed from coenzyme A and added to the cholesterol hydroxyl group. This addition makes the cholesterol more hydrophobic. The cholesteryl esters are then transported from the liver as part of lipoprotein particles. Cholesteryl esters are then taken to other tissues that need cholesterol for membrane synthesis or hormone synthesis[2].

Nelson and Cox distinguish between "good" cholesterols and "bad" cholesterols. Cholesterol molecules are classified based on their densities. High and low density designations depend on the combination of lipids and proteins within the particle, and the particles range anywhere from chylomicrons (very loosely associated) to high-density lipoproteins. Chylomicrons are the largest and least dense of cholesterol containing particle and assist in the transport of dietary triglycerides[2]. Katzung elaborates that chylomicrons are formed in the intestine and carry ingested triglycerides as well as unesterified cholesterol and cholesteryl esters through the bloodstream. When the chylomicrons are broken down by lipases in the liver, the apoproteins and cholesterol are transferred to lipoproteins[3]. These complexes transport plasma lipids through the body. The basic structure of lipoproteins contains a hydrophobic core made up of cholesteryl esters and triglycerides. The core is surrounded by a layer of unesterified cholesterol,

phospholipids and apoproteins[3]. VLDL (very-low-density lipoprotein) is formed when the diet of a person contains more fatty acids than are immediately needed as fuel. VLDL can also be formed when a person's diet contains excess carbohydrates. LDL (low-density lipoprotein) forms when triglycerides are removed from VLDL, which leaves the lipoproteins with higher concentrations of cholesterol and cholesteryl esters. The purpose of LDL is to move cholesterol from the liver to extrahepatic tissues, especially growing tissues that require cholesterol for the formation of cell membranes[2]. Low-density lipoproteins contain apolipoprotein which inserts lipids into the walls of arteries and ultimately results in hardening and weakening of the vessels[3]. The last major lipoprotein is HDL (high-density lipoprotein). HDL is rich in protein and contains only minor amounts of cholesterol. Cholesteryl esters are completely absent in HDL[2]. Katzung discusses that to be synthesized, HDLs not only require cholesterol from the liver but they also require cholesterol from peripheral tissues which gives HDLs the ability to maintain cholesterol homeostasis of cells. They are able to maintain cellular levels of cholesterol by their ability to retrieve cholesterol from the artery walls that was inserted by the apolipoproteins. High-density lipoproteins (HDLs) have antiatherogenic effects. HDL can also inhibit the production of atherogenic lipoproteins[3].

Nelson and Cox explain that when excess lipoproteins are found in the body, one of the major clinical effects is atherosclerosis. Atherosclerosis is the leading cause of death in the United States and many other westernized countries[3].When the combined total amount of cholesterol synthesized and ingested exceeds the amount necessary for the synthesis of membranes, bile salts and steroids, pathological amounts of cholesterol can accumulate in the blood vessels. These pathological accumulations are often called

atherosclerotic plaques. Such plaques obstruct blood vessels causing atherosclerosis. High levels of cholesterol have been linked to atherosclerosis, especially high levels of LDL-bound cholesterol. As previously noted, there is an inverse relationship between high levels of HDLs and arterial disease, meaning that with higher levels of HDLs a person is less likely to suffer from disease. Low levels of HDL are an independent risk factor of atherosclerosis. Familial hypercholesterolemia, a condition where a person has defective LDL receptors and cannot uptake cholesterol carried by LDLs, leads to an increase of the concentration of cholesterol in the blood. The accumulation of cholesterol contributes to atherosclerotic plaques. This disorder is a vicious cycle because cholesterol cannot enter the cells and regulate cholesterol synthesis, so the body continues to synthesize cholesterol despite the high levels of cholesterol already found in the blood[2].

Cholesterol tends to be deposited with triglycerides, protein and dead cells in arteries causing the arteries to narrow—a major contributing factor of atherosclerosis. The deposits are known as an atheroma or atherscleortic plaque. At the point of disease, the arteries have narrowed so much that blood can barely pass through and increases risk for a complete blockage in circulation brought on by the formation of a blood clot[1]. As mentioned, healthy arteries can dilate in response to restricted blood supply, but atherogenic lipoproteins—apolipoproteins—impair this function, thus exacerbating the lack of blood supply when a plaque is obstructing the artery. Atherosclerosis is a multifaceted disease that requires modification and maintenance on many levels, but one way to stem the flow of potentially fatal outcomes is to reduce the levels of atherogenic lipoproteins. If the levels of atherogenic lipoproteins are reduced, the health of the arteries can be restored[3]. Depending on the location of the blockage, there can be a

number of disastrous outcomes. If a plaque occurs in the arteries leading to the brain, the result can be a stroke. If a plaque occurs in the arteries leading to the heart, it can result in a heart attack (heart failure) and/or coronary heart disease[1].

The risk of heart disease increases with rising concentrations of atherogenic lipoproteins. When simple diet and exercise is not enough to control levels of cholesterol, there is likely some sort of metabolic defect. HMG-CoA reductase inhibitors are added to the regimen to fight excess levels of endogenous cholesterol[2].

HMG-CoA reductase inhibitors, often referred to as statins, are widely prescribed to lower cholesterol. All statins contain an HMG-like moiety[4] and mimic the structure of HMG-CoA (3-hydroxy-3-methylglutaryl-coenzyme A). HMG-CoA reductase is part of the first mediated step in sterol biosynthesis[3]. Istvan and Deisenhofer show that the committed step of cholesterol biosynthesis, the step that statins inhibit, is when HMG-CoA is deacylated and splits into CoA and mevalonate. The reaction is as follows:

$$\text{HMG-CoA} + 2\ \text{NADPH} + 2\text{H}^+ \rightarrow \text{mevalonate} + 2\ \text{NADP}^+ + \text{CoASH}$$

In the body the drugs are hydrolyzed into their active hydroxy-acid form[4]. Statin drugs are 40-70% absorbed (see Katzung), with the exception of fluvastatin which is almost completely absorbed. Atorvastatin (commonly known as Lipitor) and rosuvastatin (commonly known as Crestor) are present in the circulatory system without being fully metabolized for 14 and 19 hours respectively, which correlates with the fact that these two drugs are potent and commonly prescribed. Statin drugs are most effective in reducing levels of LDLs. The inhibitors cause an increase in high-affinity LDL receptors which both increases the catabolic rate of LDLs as well as increases the amount of LDLs the liver extracts from the blood resulting in an overall decrease of LDL[3].

The HMG-binding pocket, where HMG-CoA and statins bind, is characterized by a cis-loop, consisting of amino acid residues 682-694[4]. Statin drugs are structurally specific drugs, meaning they bind to a specific active site, and to avoid side effects, hopefully bind only at that site. The effectiveness of the drug is strongly dependent on the active site and thus is strongly affected even by small changes in the chemical structure of the active site[5]. Protein-ligand complexes like HMG-CoA reductase with HMG-CoA are sensitive to even the smallest of geometric changes within the active site[6]. When the binding of HMG-CoA is compared to the binding of statins, there is a clear rearrangement of the substrate-binding pocket (see Istvan and Deisenhofer). Current statin drugs are considerably smaller than the natural substrate HMG-CoA and bind to a small section of the HMG-CoA reductase active site, primarily to 4 residues: Asp690, Lys691, Lys692, and Ser684. These amino acid residues are located on the cis-loop and form polar interactions with the HMG-moiety of statins. There is charge and shape complementarity between the HMG-moiety and the protein due to the large number of hydrogen bonds and ion pairs (the interactions between the HMG-moiety and the protein are mostly ionic or polar). Rosuvastatin has the greatest number of bonding interactions with the active site, and best inhibits the binding of HMG-CoA in the active site of HMG-CoA reductase[4].

1.2 Drug Design

The need for novel HMG-CoA reductase inhibitors is increasing due to stricter guidelines for patients that are at risk for high cholesterol. Furthermore, Silverman discusses the profitability of statin drugs, prompting even more research into the next 'blockbuster drug.' Drug design can be a complicated process, especially if shooting in the dark. One

method of drug design is called random screening. Random screening is a method wherein natural products and synthetic chemicals are tested against the targets that the drug candidates are meant to act upon to see if the desired outcome is accomplished[5]. As drug design has developed, nonrandom screening has developed as an alternative to random screening. Nonrandom screening takes active compounds and modifies them to be more effective—the primary focus of this work. A third method for drug design is rational drug design. In this method information about the structure of the drug target is known. The drug candidates are tailored specifically to match the drug target. The philosophy of drug design used in the development of the novel fragment and the novel drug candidates is (loosely) the fragment-based lead discovery (FBLD) method[7]. This less costly and more efficient alternative to high-throughput screening optimizes the binding of small fragments of the overall drug molecule to the target active site and then puts the fragments together in the hope of increasing the overall molecule's affinity for its target (see Silverman). FBLD is a type of rational drug design that requires knowledge of the structure of the target active site so the fragments can be grown to fill the space. In the case of statin drug design, scientists are able to use HMG-CoA as a lead compound, because HMG-CoA that is normally synthesized by the body binds in the specific active site located within HMG-CoA reductase. Lead compounds can be trimmed to determine which part of the drug is responsible for the activity because it is important to know which part of the drug partakes in the desired interactions. Statin drugs were initially discovered through random screening, however, after the crystal structure of HMG-CoA reductase was discovered, rational drug design has been used to design improved statin

drugs. There is also an opposite approach to trimming drugs; increasing the complexity and/or the rigidity of a drug can potentially make the drug more potent[5].

1.3 Quantitative Structure-Activity Relationships

Silverman shows that once multiple drug candidates have been synthesized from the lead compound, a linear multiple regression analysis can be conducted to determine *C*, the dose that elicits the desired biological response, in this case the IC_{50} (the concentration of the drug that inhibits 50% of an enzyme's action). The pIC50 in the equation below is the negative logarithm of the C (concentration).

$$pIC50 = -k\pi^2 + k'\pi + \rho\sigma + k''$$

$$pIC50 = -k(\log P)^2 + k'(\log P) + \rho\sigma + k''$$

In these equations the pIC50 is correlated with π and σ which are the lipophilicity and electronic substituent constants of the drug candidates, respectively. The negative logarithm of the concentration reflects the fact that greater potency is associated with a lower dose, and the negative sign for the π^2 [or $(\log P)^2$] term reflects the expectation of an optimum lipophilicity, that is π_o or $\log P_o$. The terms k, k', ρ and k'' are the regression coefficients derived from the linear fitting[5].

1.4 Computational Methods

Modern computational chemistry is being applied to larger biological systems daily and with such advancement comes a concern regarding the amount of time and resources used to calculate information for such large systems. Examples of such systems are protein-ligand interactions, where the interactions of several hundred atoms and potentially thousands of electrons must be calculated at one time. If a protein-ligand system features interactions between the ligand and the active site that depend largely on

dispersion and induction forces, care must be taken to use a modeling method for each portion of the system that is accurate for dispersion and induction. This work examines an interesting and timely study of a system—statin drugs bound to their target enzyme—that requires careful treatment of dispersion and induction.

Dispersion and induction forces both arise from electron movement within molecular systems. When two separate molecules are brought together, the electron density within each molecule adjusts itself to accommodate the electron density in the other molecule. This polarization of the electron density gives rise to the dispersion and induction forces. The ability of a quantum chemical method to model this behavior is referred to as electron correlation.

Coupled-cluster singles, doubles, and perturbative triples (CCSD(T)) is an *ab initio* method that takes into account electron correlation for single electrons, pairs of electrons and sets of three electrons and can accurately calculate interaction energies for extensive complexes (see Riley and Hozba). CCSD(T) is a very accurate method, but is also very expensive and requires large amounts of time and computational resources. Second order Moller-Plesset perturbation theory (MP2) is also appropriate for calculating interaction energies for systems that are dominated by dispersion forces. MP2 is less accurate than CCSD(T) because it only takes into account the correlation energy for pairs of electrons. While it is a little less accurate, MP2 does a good job mimicking the results of CCSD(T) and does not require nearly as much computational time and resources. When MP2 is run with a large basis set, it accurately calculates the interaction energies due to hydrogen bonds, but overestimates the binding energies of stacked and dispersion-bound complexes. This flaw in MP2 can be corrected with a small basis set and

ultimately makes MP2 a viable option for accurate calculation of interaction energies[6].

The current work uses density functional theory (DFT) calculations on the novel drug candidate molecules. Certain DFT methods have been shown through previous work in this group to adequately describe the weak forces in this system[5]. According to Riley and Hozba, DFT in theory can provide exact electronic energies for molecular systems which include electron correlation. The exact form of DFT is unknown so approximate forms have been developed over time which are parameterized to give accurate energies for different types of molecular systems. DFT is favorable over other methods because it is cheaper in terms of the computing time that it takes[6]. Most DFT functionals model electron correlation—and through that, dispersion—through their correlation potential. Certain DFT methods such as HCTH and B3LYP completely underestimate stacked and dispersion-bound complexes, and thus are not good methods to rely on when calculating these energies. HCTH and B3LYP do a decent job of describing the energies due to hydrogen bound complexes, although the exact numbers are still slightly underestimated. Other DFT methods use an empirical dispersion correction to model dispersion which improves the accuracy for interaction results not only of hydrogen bonded systems, but also stacked and dispersion-bound systems[6].

In this work, B3LYP is used as a point of comparison because it is a commonly used DFT functional in biological chemistry applications. B3LYP is shown here to be inadequate for this type of system as it is primarily successful at calculating the energies due to hydrogen bonding and this system is governed by dispersion forces as well. B3LYP is not a good method because it was designed without regards to noncovalent interactions[6]. Another DFT method shown in this work is HCTH. HCTH is a method that

is supposedly accurate at calculating all types of molecular interactions, but like B3LYP, has shown to be inadequate with this system of molecules. The local functional SVWN is also used as it has been shown to produce results remarkably similar to MP2 results for this type of system[8].

Truhlar has developed several DFT methods specifically designed to analyze dispersion and induction interactions[9, 10]. Vydrov and Voorhis have also designed a novel functional which models dispersion well[11], these new functionals are not included in the current work as they are not yet available in widely used software packages.

1.5 Previous Work

The crystal structure work of Istvan and Deisenhofer[4] (showing the crystal structure of statin drug molecules bound to HMG-CoA reductase) and previous work in this group[12] suggest that the interaction between Tyr479 in the active site and a ligand may be important in drug-binding and should be used in designing new ligands for this active site. In this work a novel molecular fragment that interacts strongly with Tyr479 developed in our group has been attached to two known statin drugs (rosuvastatin and simvastatin) to create new drug candidates that interact with the target enzyme active site more strongly than the original drugs, see Figure 1.

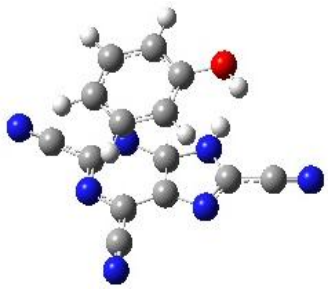

Figure 1. Novel tail fragment in complex with Tyr479 residue from the HMG-CoA reductase active site.

The novel drug fragment used here interacts with Tyr479 via dispersion and induction forces, thus an appropriate theoretical model that can describe these weak forces is needed to calculate the interaction energies. Second order Moller-Plessett perturbation theory (MP2) was the primary method used in the design of the novel drug fragment in this work[13].

Pfefferkorn *et al* prepared a set of prototype drugs by modifying an existing statin, atorvastatin, and subsequently adding groups through chemical synthesis[14]. The prototypes were tested via in vitro and in vivo methods. Pfefferkorn *et al* showed that smaller, less lipophilic molecules work more effectively as substrates for transport. The transport properties of the novel drug molecules designed in this work have been examined through calculation of solvation energies in water and chloroform for both the original drug and the new candidates using a conductor-like screening model (COSMO)[15]. The novel drug candidates are shown to have similar solvation properties to the original drug molecules.

2 Methods

2.1 QSAR

Molecular properties of a suite of statin drugs were examined to determine a correlation between molecular properties of statins and their pIC50 value. The pIC50 value is the negative logarithm of the amount (concentration) of drug needed in order to block fifty percent of cholesterol synthesis.

The drug molecules were each first structurally optimized in gas phase to get a foundation structure that works well in other solvents. The drugs were then optimized in water to mimic the aqueous environment of blood. Lastly, the drug molecules were

optimized in chloroform (a non-polar solvent) as a mimic of the cell membrane through which the drugs must pass. The dipole moments were also calculated because the dipoles are pertinent to molecular recognition of the drug molecules by the cells in the body. Successful drug candidates should generally have similar dipoles to statins that are known to be good cholesterol inhibitors. The HOMO/LUMO (highest occupied/lowest unoccupied molecular orbitals) gap was also examined as these are the orbitals from which and to which electrons will transfer between the binding site and the molecule (note: a smaller gap is more reactive than a larger gap). The LUMO energy was calculated because it seemed to best correlate with the effectiveness of the drug.

2.2 Interaction energies

In previous work, this research group developed seven statin drug candidates by adding a novel tail moiety to an existing statin in order to take full advantage of non-covalent interactions in the binding site[13]. This novel tail moiety, which was optimized to have maximum interaction with Tyr479, was added to current statins by a saturated hydrocarbon linker of differing lengths, see Figure 2. Rosuvastatin candidates 1-4 (RC1-RC4) had linkers from five to eight carbons respectively. Simvastain candidates 1-3 (SC1-SC3) had linkers from eight to ten carbons respectively. The number of carbons used in the linkers was tested because when this project was started, it was not known how many carbons would be needed to reach from the HMG-moiety to the Try479 residue on the active site.

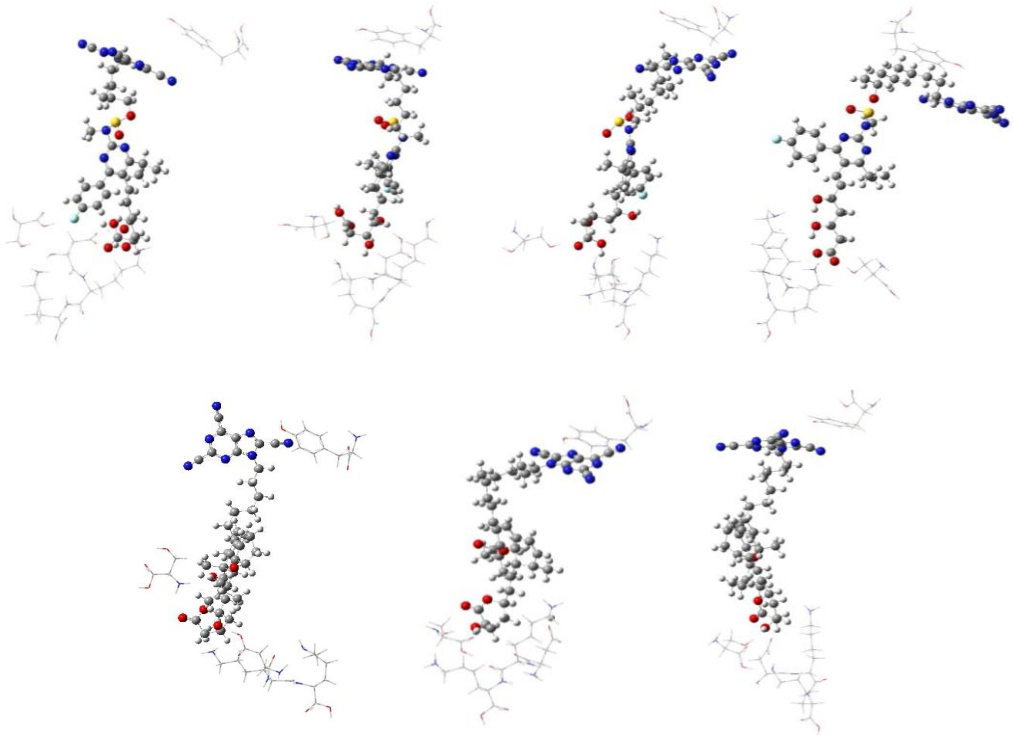

Figure 2. Rosuvastatin candidates 1-4 as docked in the HMG-CoA reductase active site (above). Simvastatin candidates 1-3 as docked in the HMG-CoA reductase active site (below). Figures illustrate interaction of tail moiety with Tyr479 and the interaction of the statin moiety with other amino acid residues in the active site.

The current work was performed on a previously obtained crystal structure of HMG-CoA reductase with HMG-CoA bound to the active site. The amino acid residues in the binding site for the statin drug candidate molecules were determined by proximity to the HMG-CoA substrate, and it was found that twenty-nine amino acids comprise the active site. Of the twenty nine amino acids in the active site, five form strong non-covalent bonds to the ligands: Tyr479, Ser684, Asp609, Lys691, Lys692[4]. Once the active site was isolated, the ArgusLab program was used to dock our drug candidates into the site and determine the most energetically favorable pose of the candidate in relation to the amino acid residues in the active site. The interaction energies were then calculated pair-wise, meaning the interaction energy between the drug and each of five amino acids

were calculated in turn, and then those energies were added together to find the total interaction energy. The DFT methods, SVWN, HCTH and B3LYP, with the 6-311+g* basis set, were used to calculate the counterpoise corrected interaction energies.

Information gathered from the initial seven drug candidates was used to design secondary generations of novel statin drugs called Generations 2-5 (G2-G5), see Figure 3. The secondary generations of drugs were comprised of modified forms of rosuvastatin candidate four because RC4 was the most successful candidate from the first generation of drug candidates. Modifications were made to RC4 that were believed to further enhance the success of this statin drug candidate by increasing the strength of non-covalent interactions in the active site.

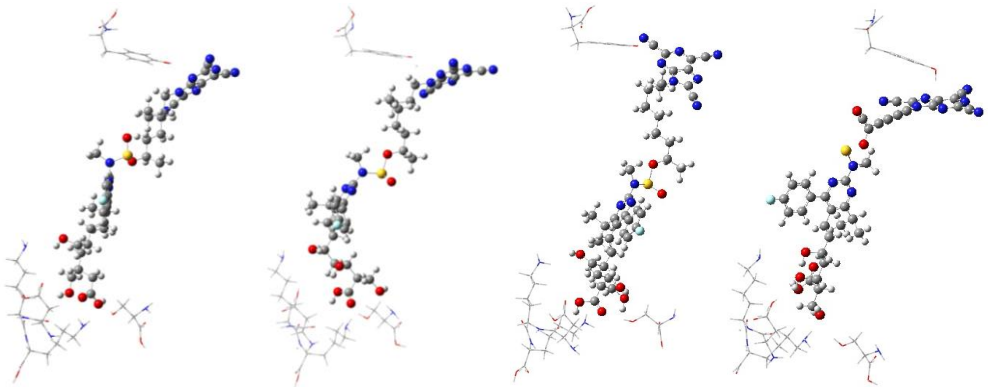

Figure 3. Secondary generations candidates, G2-G5, as docked in the HMG-CoA reductase active site. Figures illustrate interaction of tail moiety with Tyr479 and the interaction of the statin moiety with other amino acid residues in the active site.

The changes made to RC4 to generate G2 were the addition of a methyl group between each of the two hydroxyl groups on the front of the HMG-moiety. The purpose of this modification was to make the front end of the drug candidate broader in order to better span the entire width of the active site. The next modification made to RC4 was the replacement of a hydrogen attached to a carbon with a hydroxyl group on the very front end of the HMG-moiety to create G3. This modification was intended to form a hydrogen

bond with the Ser684 residue on the active site and thus bind the drug candidate more strongly to the active site. G4 was a hybrid between G2 and G3 with the modifications that made each of those drug candidates bind more strongly. One methyl group was left in between the hydroxyl group and the HMG-moiety and the hydroxyl group addition was also left in order to create G3. An entirely different approach was taken for the last second generation drug candidate, G5. The HMG-moiety was left as it is found in rosuvastatin and the linker was changed that attaches the modified ring structure to the rest of the drug. All of the previous drug candidates possessed a saturated hydrocarbon tail making the linker between the HMG-moiety and the engineered ring structure flexible allowing rotation around the single bonds. The sp^3 carbons of the linker were changed to sp carbons (carbons each with one triple bond and one single bond), making the linker rigid. The purpose of the stiff linker was to force the ring-ring interaction between the modified tail and the Tyr479 residue on the active site.

Binding of a ligand in an allosteric site in HMG-CoA reductase was also examined. This site binds the NADP(H) molecule, which is a cofactor for the reduction of HMG-CoA to mevalonate. The amino acid residues of the allosteric site were determined from the crystal structure of HMG-CoA reductase. The amino acid residues present were: Try479, Glu528, Asn529, Ala962, Asn965, Arg966 and Lys1120. The complex of these amino acids with the NADP(H) molecule was isolated; protons were added to this complex and the position of the protons were optimized with HF/6-31g. The DFT methods, SVWN, HCTH and B3LYP, with the 6-311+g* basis set, were used to calculate the counterpoise corrected interaction energies of the NADP(H) with each amino acid residue in turn.

All calculations were performed using Gaussian03[16] or PQS software[17].

3 Results and Discussion

3.1 QSAR

The molecular properties for a suite of statin drugs were calculated and correlated with pIC50 values, see Table 1. The drugs used were either currently or previously offered on the market (simvastatin through lovastatin, see below) or were drug candidates with reported pIC50 values. The pIC50 values in the table below were experimentally determined[4, 14]. Initially, each drug was optimized in the gas phase to make the optimizations in solution easier and less costly. The drug candidates were then optimized in the presence of water and chloroform. Water was used as a solvent because it is akin to blood, and after being ingested the drug would have to be able to be carried through the blood. Chloroform was used as a solvent as it was the most non-polar solvent through which the drugs would run. The non-polar environment of chloroform mimics the non-polar environment of the cell membrane which the drugs must be able to diffuse across *in vivo*. The energies of the drug candidates in each of these phases were used to calculate the solvation energy of the drug candidates in each solvent. The solvation energies of the drugs in water and chloroform are important because as mentioned the drugs must be able to dissolve into both of these types of environments. The dipole moments (denoted by m) of the drugs in solution are important because the dipole moment is pertinent to how molecules bind to one another. The HOMO/LUMO gap (highest occupied molecular orbital/lowest unoccupied molecular orbital) is important because these are the orbitals from which and to which electrons will jump when molecules bind to one another.

	$DE_{solvation}(H_20$, kcal/mol)	$DE_{solvation}$ ($CHCl_3$, kcal/mol)	m (gas, debye)	m (H_20, debye)	m ($CHCl_3$, debye)	E_{LUMO} (ha)	HOMO/LUMO gap	pIC50
Simvastatin	-18.63	-11.22	5.44	7.47	6.70	-0.05	0.14	7.96
Atorvastatin	-33.46	-20.62	3.78	4.73	3.80	-0.06	0.13	8.10
Fluvastatin	-19.24	-12.00	4.55	7.25	5.67	-0.07	0.11	7.55
Cerivastatin	-21.66	-13.55	1.00	1.71	1.32	-0.06	0.13	8.00
Rosuvastatin	-28.47	-17.81	6.23	10.82	9.17	-0.12	0.08	8.30
Compactin	-21.90	-13.89	6.37	10.13	8.69	-0.05	0.14	7.64
Lovastatin	-21.25	-13.16	6.98	8.38	7.70	-0.05	0.14	7.70
Pfeff33	-24.88	-15.82	4.54	5.46	5.13	-0.07	0.12	8.02
Pfeff42	-28.63	-18.46	6.24	7.76	7.23	-0.07	0.12	8.72
Pfeff43	-47.88	-37.12	6.24	8.10	3.96	-0.07	0.12	8.54
Pfeff28	-25.31		2.98	4.14		-0.06	0.12	7.78

Table 1. Molecular properties (calculated with HCTH/6-31g) for on-the-market and theoretical statins to be correlated with experimental potencies, pIC50 values.

A multivariate linear regression was used to correlate the molecular properties with each of the drugs' pIC50 values and it was found that behind the solvation energy of the drug in water and the solvation energy of the drug in chloroform (R^2=0.482 and 0.4299 respectively) that the LUMO energy correlated most strongly with the potency of the drug (R^2=0.178). Further investigation of the LUMO orbitals on each drug showed that the successful drugs, when oriented in the same direction, all have LUMO orbitals that run horizontally; the strong correlation suggests that the orientation of the LUMO orbitals are important for the drug to bind properly in the active site.

The solvation energy of the drug in water, the solvation energy of the drug in chloroform and the LUMO energy were used to create a multiple regression predictive equation for our novel drugs. The equation is:

$$pIC50=0.625W+0.0417L-0.415C+6.833$$

The variable W represents the solvation energy of the drug in water, the variable L represents the LUMO energy and the variable C represents the solvation energy of the drug in chloroform. Results for this work are reported in section 3.3.

3.2 Interaction Energies

SVWN was used in this work as the standard method for measuring the interaction energies of drugs with HMG-CoA reductase. In the paper by Kee *et al*, it was shown that SVWN was the DFT method in closest agreement with MP2 which is taken to be accurate for these interactions[12]. Negative interaction energies mean the drug and the substrate attract one another and positive interaction energies means the drug and the substrate repels one another.

	B3LYP	SVWN	HCTH407	AM1
HMGCoA	-1.11	-39.93[12]*	1.06	--
Rosuva	-5.01	-31.51	-5.14	-106.14
Rosuva + T		-38.85		
RC1	-23.12	-57.72	-18.27	9.89
RC2	10.71	-26.84	12.78	5.40
RC3	19.18	-6.27	18.56	9.67
RC4	-74.22	-104.46	-73.45	-41.95
Simva	-5.81	-27.64	-7.01	1.20
Simva + T		-34.98		
SC1	-14.43	-40.28	-14.00	-6.30
SC2		-32.46	1.89	-1531
SC3	-6.56	-29.82	-5.53	-4.70

Table 2. Interaction energies of substrate, drugs and drug candidates with the HMG-CoA reductase active site. Energies are in kcal/mol and were calculated with a basis set of 6-311+g*. *This value is not calculated in this work, see reference.

HMG-CoA binds with an energy of -39.93 kilocalories[12], see Table 2. The interaction energies of rosuvastatin and simvastatin with the enzyme are -31.51 kcal per mole and -27.64 kcal per mole, respectively. These energies are comparable to the binding energy of HMG-CoA, although they bind a little less strongly, making them good competitive inhibitors. The way these candidates are referenced is "R" candidates one through four. All candidates possess the modified tail this research group previously designed and only vary by the length of saturated hydrocarbon linker attaching the tail to rosuvastatin. RC1 has a linker of five carbons all the way to RC4 which has a linker of eight carbons. "S" candidates have the modified tail attached to simvastatin and have linker lengths of eight to ten carbons, respectively. As made evident by the interaction energies, the most successful drug candidate was rosuvastatin candidate 4 with a binding energy of -104.46 kcal per mole. This not only shows that the drug is a promising competitive inhibitor, but

it also shows that RC4 binds to the active site more strongly and would be able to out compete HMG-CoA thus even further inhibiting the production of cholesterol. Rosuvastatin candidate 1 also bound more strongly than rosuvastatin itself with an interaction energy of -57.72 kcal per mole. Also notable, all of the modified simvastatin candidates had stronger interaction energies with the active site that simvastatin. Simvastatin has an interaction energy of -27.64 kcal per mole as compared to simvastatin candidates 1-3 that had interaction energies of -40.28 kcal per mole, -32.46 kcal per mole, and -29.82 kcal per mole, respectively.

3.3 Second Generation

Secondary generations of drug candidates were developed based on the above candidates. These drug candidates, called Generations 2-5 (G2-G5), were modified from RC4, the most successful drug candidate from the first round of modification based on its binding energy. These second generation drugs were optimized, docked, and the molecular properties and binding energies were calculated.

As calculated with SVWN, G2 had a total binding energy of -24.17 kcal per mole, see Table 3. When compared to RC4, G2 has a lower interaction energy. Also when compared to simvastatin and rosuvastatin, its binding energy with the active site is not as strong. Although it has all attractive interactions, the overall interactions with each amino acid residue are all relatively weak with the exception of Asp690.

	SVWN	B3LYP	HCTH
tyr479	-5.8443	2.1715	3.0606
ser684	-3.4767	-0.0201	-0.2055
asp690	-9.2728	-4.3111	-5.2962
lys691	-4.1799	-1.0817	-3.2502
lys692	-1.3950	2.9883	2.8057
Total Interaction	-24.1687	-0.2531	-2.8856

Table 3. Interaction energies for Generation 2 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*.

As calculated with SVWN, G3 had a total interaction energy of -35.36 kcal per mole, see Table 4. While still not as strong as the original RC4, this modification appears promising. The interaction energy of G3 was stronger than both simvastatin and rosuvastatin, which theoretically would make it a better competitive inhibitor than both of those drugs.

	SVWN	B3LYP	HCTH
tyr479	-6.3023	2.1259	2.9774
ser684	-18.5499	-6.3796	24.8870
asp690	-3.4442	-1.2563	-2.3958
lys691	-3.9056	-2.1099	-3.2711
lys692	-3.1562	-2.3881	-3.2234
Total Interaction	-35.3582	-10.0079	18.9741

Table 4. Interaction energies for Generation 3 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*.

As calculated with SVWN, G4 had a total binding energy of -38.18 kcal per mole, Table 5. Of all of the second generation candidates, G4 bound most strongly to the active site. Again, this interaction energy is not as strong as the original rosuvastatin candidate G4 was made from, but aligns well with the predicted pIC50 value (see Table 7). G4 had the highest predicted pIC50 value of all of the second generation candidates which

corresponds with the fact that it also binds most strongly. G4 binds almost as strongly as HMG-CoA, and binds stronger than rosuvastatin and simvastatin, theoretically making it the best inhibitor of HMG-CoA.

	SVWN	B3LYP	HCTH
tyr479	-2.9789	-1.5966	-2.9005
ser684	-14.3688	-4.4620	-1.9324
asp690	-11.1397	-3.1083	-2.0393
lys691	-3.0034	-1.2466	-3.1961
lys692	-6.6929	-3.3967	-3.9927
Total Interaction	-38.1837	-13.8103	-14.0609

Table 5. Interaction energies for Generation 4 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*.

As calculated with SVWN, G5 had a total binding energy of -25.32 kcal per mole, see Table 6. Like G2, this drug candidate is not as successful at binding to the active site as RC4, but it also binds more weakly than rosuvastatin and simvastatin. The length of the stiff linker was five carbons long. Five carbons were used as an educated guess based on the previous drug candidates at the length from where the HMG-moiety ended to the back end of the active site. We believe the weak interaction energy stems from the incorrect length of the linker. Because the linker was too short, the program we used to optimally dock the drug in the active site was forced to take an average between the strong interaction energy of the front end of the active site and the strong ring-ring interaction energy on the back end of the active site, thus resulting in an overall weaker interaction energy.

	SVWN	B3LYP	HCTH
tyr479	-7.9304	-1.1653	-1.2716
ser684	-6.0016	-3.7471	-4.1228
asp690	-6.4968	-1.5682	-1.5989
lys691	-3.7445	-1.8135	-3.3013
lys692	-1.1542	-0.6379	-1.3707
Total Interaction	-25.3276	-8.9320	-11.6654

Table 6. Interaction energies for Generation 5 candidate with the HMG-CoA reductase active site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*.

3.4 Quantitative Structure-Activity Relationships of the Second Generation

Molecular properties of these second generation drugs were used in the predictive equation discussed earlier to calculate the projected pIC50 values for these newest statin drug candidates, see Table 7.

	$DE_{solvation}$ (H_2O, kcal/mol)	$DE_{solvation}$ ($CHCl_3$, kcal/mol)	m (gas, debye)	m (H2O, debye)	m (CHCl3, debye)	E_{LUMO} (ha)	HOMO/LUMO gap
G2	-48.1968	-30.9849	10.0729	12.0986	11.4454	0.0077	0.3251
G3	-41.3342	-26.7274	11.5782	13.7630	13.0148	-0.1299	0.0896
G4	-44.3486	-28.3944	10.2355	12.9416	11.9781	-0.1299	0.0931
G5	-36.4093	-23.6753	13.9370	16.1217	15.4374	-0.1645	0.0328

Table 7. Molecular properties (calculated with HCTH/6-31g) for theoretical statins which will be used to predict experimental potencies, pIC50 values.

The pIC50 for G2 was 8.53. The pIC50 for G3 was 8.85. The pIC50 for G4 was 8.97. The pIC50 for G5 was 8.81. These projected pIC50 values were encouraging because the most potent statin offered has a pIC50 value of only 8.3. This means that by the current computational models the statin drugs designed here should be more effective than drugs currently offered.

So far, none of the second generation drugs have proven to be more successful at binding to the active site than the first generation. While Silverman proposed the idea that increasing the complexity of drugs could potentially make them bind more strongly, that

was not the case with the second generation of drugs[5]. There are two logical explanations as to why the second generation was not as successful as rosuvastatin candidate four even though that was the model drug from which each of the second generation candidates was made. The first potential explanation is that the addition of substituents to the drug made the candidate too complicated. Both the front end and the back end possessed the ability to form strong binding interactions and when docking the candidates, the program was not able to dock them in a way favorable to both strong interactions and ended up with an average between the two. Instead of increasing the interaction energies, this average actually weakened the interaction of the statin candidate with HMG-CoA reductase. The second potential explanation is that the addition of substituents made the front end of the drug candidates too bulky. In theory, changing the front end of the drug to be broader to better span the active site seemed reasonable, but in practice the drug was too bulky to fit comfortably into the active site.

3.5 Noncompetitive Inhibitors

At this point, the only statins that have been examined are competitive inhibitors of HMG-CoA binding to HMG-CoA reductase. These drugs do not affect the binding of NADPH in the allosteric site because no portions of the statins affect the allosteric site[4]. An idea currently being pursued by this research group is the creation of statins that are noncompetitive inhibitors and would exploit the allosteric site of HMG-CoA reductase where NADPH binds, see Figure 4 and Table 8.

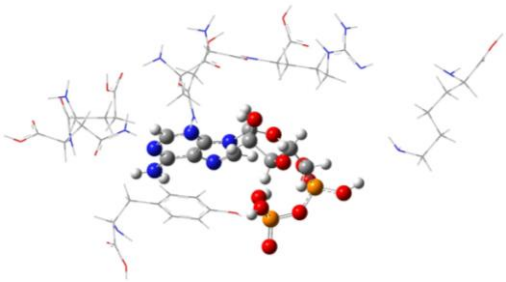

Figure 4. NADP(H) bound in the allosteric site of HMG-CoA reductase, taken from the crystal structure.

	SVWN	B3LYP	HCTH
tyr479	-21.4790	-5.9513	-4.3681
glu528	-2.3316	0.7913	-0.0298
asn529	-1.0381	3.0943	2.9604
ala962	-5.8178	-0.0356	-1.0861
asn965	-12.6650	-4.1606	-4.4186
arg966	-4.7023	0.7875	-0.1749
lys1120	-8.1595	-6.6753	-7.3977
Total Interaction	-56.1934	-12.1498	-14.5147

Table 8. Interaction energies for NADPH in the HMG-CoA reductase allosteric site. Energies are in kcal/mol. Calculated with a basis set of 6-311+g*.

As calculated with SVWN, NADPH binds in the allosteric site with and energy of -56.19 kcal per mole. This interaction energy is stronger than HMG-CoA binding in the active site, showing that the allosteric site would be a prime location to create a strongly binding drug. It was recently discovered that Tyr479 residue in the active site is also present in the allosteric site, providing the opportunity to use the novel tail again to interact strongly with Tyr479 on novel drugs created to bind in this site. If Try479 was pulled toward the allosteric site and simultaneously away from the active site, the SCoA tail of HMG-CoA would not be able to help HMG-CoA bind in the active site making the interaction of HMG-CoA with the active site weaker. If the binding of HMG-CoA with the active site is weak, the likelihood that it binds there and is able to be converted to mevalonate is much lower and thus the committed step of cholesterol synthesis is inhibited.

4 Conclusions

The study of molecular properties revealed that there is a strong correlation between the LUMO energy, the solvation energy of water and the solvation energy of chloroform and the potency of statin drugs (pIC50 values). The derived regression equation predicted high pIC50 values for the novel drug candidates. The results of the regression equation were further confirmed by the strong interaction energies of the drug candidates and the active site. Because these two completely different methods agreed and yielded the same result, it is believed that synthetic forms of these drug candidates would be successful inhibitors of cholesterol.

Due to the strong binding of NADPH in the allosteric site, it can be concluded that there is high possibility to create an uncompetitive inhibitor that will deform the active site and be more potent than statin drugs that are currently offered.

Works Cited

1. C. Pasternak, *The Molecules Within Us: Our Body in Health and Disease*; Plenum Press: New York, NY, 71 (1998).
2. D.L. Nelson and M.M. Cox, *Lehninger Principles of Biochemsitry*, 5th edition; W.H. Freeman and Company: New York, NY, 832 (2008).
3. B.G. Katzung, *Basic and Clinical Pharmacology*, 10th edition; The McGraw-Hill Companies: New York, 560, (2007).
4. E. S. Istvan and J. Deisenhofer, *Science* **292**, 1160 (2001).
5. R.B. Silverman, *The Organic Chemistry of Drug Design and Drug Action*; Academic Press Inc: San Diego, CA, 8 (1992).
6. K.E Riley and P. Hozba, *WIREs Comput. Mol. Sci*. **1**, 3 (2011).
7. S. Everts, *Chem. and Eng. News*. 15 (2008).
8. Hofto, M. E.; Cross, J. N.; Cafiero, M. *J. Phys. Chem.* 2007, **111**, 9651; Hofto, L. R.; Van Sickle, K.; Cafiero, M. *Int. J. Quant. Chem.* 2007, **108**, 112.; Hofto, M. E.; Godfrey-Kittle, A.; Cafiero, M. *THEOCHEM* 2007, **809**, 125.; Van Sickle, K.; Culberson, L. M.; Holzmacher, J. L.; Cafiero, M. *Int. J. Quant. Chem.* 2007, **107**, 1523.; Godfrey-Kittle, A.; Cafiero, M. *Int. J. Quant. Chem.* 2006, **106**, 2035.; Hofto, L. R., Lee, C., Cafiero, M. *J. Comp. Chem. A.* **30**, 1111 (2007).
9. Y. Zhao and D. G. Truhlar, *J. Phys. Chem.* **109**, 5656 (2005).
10. Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.* **3**, 289 (2007).
11. O.A. Vydrov and T. van Voorhis, *J. Chem. Phys.* **132**, 164113 (2010).
12. E. A. Kee, M. C. Livengood, E. E. Carter, M. L. McKenna, and M. Cafiero, *J. Phys. Chem. B.* **113**, 14810 (2009).
13. H. E. Utkov, A. M. Price and M. Cafiero, *Computational and Theoretical Chemistry*, in press.
14. J. A. Pfefferkorn, et al, *Bioorganic & Medicinal Chemistry Letters.* **17** (2007).
15. K. Baldridge and A. Klamt, *J. Chem. Phys.* **105**, 6622 (1997).
16. Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, (2004).
17. PQS version 3.1, Parallel Quantum Solutions, 2013 Green Acres Road, Fayetteville, Arkansas 72703.