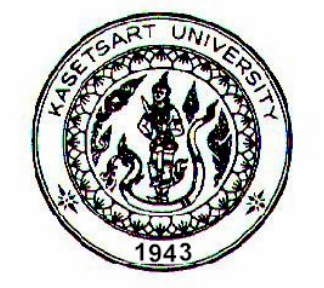


THESIS

HUMAN ANGIOTENSIN-CONVERTING ENZYME-2 (ACE2): THE EVALUATION OF THE PARTICULAR INTERACTIONS OF INHIBITOR, (S,S)-2-{1-CARBOXY-2-[3-(3,5-DICHLOROBENZYL)-3H-IMIDAZOL4-YL]-ETHYLAMINO}-4-METHYLPENTANOIC ACID (MLN-4760), IN AN ACTIVE SITE USING QUANTUM CHEMICAL APPROACH

PRAPASIRI PONGPRAYOON

GRADUATE SCHOOL, KASETSART UNIVERSITY

2007

THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Chemistry)
DEGREE

Chemistry	Chemistry
FIELD	**DEPARTMENT**

TITLE: Human Angiotensin-Converting Enzyme-2 (ACE2): the Evaluation of the Particular Interactions of Inhibitor, (*S,S*)-2-{1-carboxy-2-[3-(3,5-dichlorobenzyl)-3*H*-imidazol4-yl]-ethylamino}-4-methylpentanoic acid (MLN-4760), in an Active Site Using Quantum Chemical Approach

NAME: Miss Prapasiri Pongprayoon

THIS THESIS HAS BEEN ACCEPTED BY

Supa Hannongbua **THESIS ADVISOR**
(Associate Professor Supa Hannongbua, Dr.rer.nat.)

Chak Sangma **COMMITTEE MEMBER**
(Mr. Chak Sangma, Ph.D.)

Noojaree Prasitpun **COMMITTEE MEMBER**
(Assistant Professor Noojaree Prasitpun, Ph.D.)

Noojaree Prasitpun **DEPARTMENT HEAD**
(Assistant Professor Noojaree Prasitpun, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON 02/03/07

Vinai Artkongharn **DEAN**
(Associate Professor Vinai Artkongharn, M.A.)

THESIS

HUMAN ANGIOTENSIN-CONVERTING ENZYME-2 (ACE2): THE EVALUATION OF THE PARTICULAR INTERACTIONS OF INHIBITOR, (S,S)-2-{1-CARBOXY-2-[3-(3,5-DICHLOROBENZYL)-3H-IMIDAZOL4-YL]-ETHYLAMINO}-4-METHYLPENTANOIC ACID (MLN-4760), IN AN ACTIVE SITE USING QUANTUM CHEMICAL APPROACH

PRAPASIRI PONGPRAYOON

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Chemistry)
Graduate School, Kasetsart University
2007

Prapasiri Pongprayoon 2007: Human Angiotensin-Converting Enzyme-2 (ACE2): the Evaluation of the Particular Interactions of Inhibitor, (*S,S*)-2-{1-carboxy-2-[3-(3,5-dichlorobenzyl)-3*H*-imidazol4-yl]-ethylamino}-4-methylpentanoic acid (MLN-4760), in an Active Site Using Quantum Chemical Approach. Master of Science (Chemistry), Major Field: Chemistry, Department of Chemistry. Thesis Advisor: Associate Professor Supa Hannongbua, Dr. rer. nat. 59 pages.

Angiotensin-converting enzyme-2 (ACE2) is a membrane protein with its active site exposed to the extracellular surface of endothelial cells associated with hypertension, heart, and kidney disease. A counter-regulatory role to ACE has been proposed which makes it an interesting new cardio-renal disease target as well as it also acted as the SARS (the Severe Acute Respiratory Syndrome) receptor-binding domain. Recently, human ACE2 structure was firstly solved by X-ray crystallography in three forms: apo-bound, SARS- and inhibitor- bound (MLN-4760). However, some interactions cannot directly be evaluated by the structural distances from the crystal structure. So, the objective in this study is to determine the tangible interactions of the inhibitor (MLN-4760) in the pocket as well as reveal certain structural differences that exist between the native and the SARS-bound complex. The quantum-based density functional theory was performed to understand the particular interaction between the inhibitor and each nearby amino acids. The structural analysis shows that the conformation of SARS-bound ACE2 and the native is virtually identical and the lobe revealed as the SARS receptor domain in all of three structures (native, SARS-bound, and inhibitor-bound) does not change its conformation indicating that SARS virus are able to attach the receptor at any state. From quantum chemical calculations, we suggest that changing the functional groups of MLN-4760 around Glu145, Asn149, and Asp368 may help to improve the inhibiting affinity as well as Tyr510 may play a crucial role in changing the native conformation of a catalytic residue, His345, to the functional form.

PRAPASIRI PONGPRAYOON — Student's signature

Supa Hannongbua — Thesis Advisor's signature

2 / 03 / 07

ACKNOWLEDGEMENTS

This thesis could not have been completed in the time available to me without the dedication and advice of my kind and nice supervisor, Dr Supa Hannongbua, especially, her tolerance and understanding as well as my co-supervisors, Dr Chak Sangma and Dr Noojaree Prasitpun.

Moreover, I am also very grateful to Miss Panvika Pannopard and Miss Kornvika Tangpong for supplying me with their big hands.

My thanks also go to all members of LCAC fully supporting me in whatever I have needed to do in this research.

Finally, I am also thankful for having the opportunity to attend this course as well as the Postgraduate Education and Research Programs in Petroleum and Petrochemical Technology (ADB-MUA) in Physical Chemistry providing me funding for this study.

Prapasiri Pongprayoon

January, 2007

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (cont'd)

ABBREVIATIONS

ACE	=	Angiotensin-Converting Enzyme
ACE2	=	Angiotensin-Converting Enzyme2
Ala	=	Alanine
Ang I	=	Angiotensin I
Ang II	=	Angiotensin II
Arg	=	Arginine
Asn	=	Asparagine
Asp	=	Aspartic acid
AT1	=	Angiotensin type 1
AT2	=	Angiotensin type 2
B3LYP	=	Becke 3 Lee-Young-Parr
cDNA	=	Complementary Deoxyribonucleic Acid
CPDA	=	Carboxypeptidase A
COO^-	=	Carboxylate group
Cys	=	Cysteine
DFT	=	Density Functional Theory
GEA	=	Gradient-Expansion Approximation
GGA	=	Generalized-Gradient Approximation
Gln	=	Glutamine
Glu	=	Glutamic acid
Gly	=	Glycine
His	=	Histidine
Ile	=	Isoleucine
Leu	=	Leucine
LDA	=	Local Density Approximation
Lys	=	Lysine
NCPL	=	Lysyl carboxy-terminus of *N*-[(*S*)-1-carboxy-3-phenylpropyl]-L-lysine
Met	=	Methionine

MLN-4760	=	(S,S) 2-{1-carboxy-2-[3-(3,5-dichloro-benzyl)-3H-imidazol-4-yl]-ethylamino}-4-methyl-pentanoic
PDB	=	Protein Data Bank
Phe	=	Phenylalanine
Pro	=	Proline
RAS	=	Renin-Angiotensin System
RBD	=	Receptor-Binding Domain
RMSD	=	Root Mean Square Deviation
SARS	=	the Severe Acute Respiratory Syndrome
SARS-CoV	-	the Severe Acute Respiratory Syndrome-Corona Virus
tACE	=	testis Angiotensin-Converting Enzyme
Thr	=	Threonine
Trp	=	Tryptophan
Val	=	Valine
Tyr	=	Tyrosine

HUMAN ANGIOTENSIN-CONVERTING ENZYME-2 (ACE2): THE EVALUATION OF THE PARTICULAR INTERACTIONS OF INHIBITOR, (S,S)-2-{1-CARBOXY-2-[3-(3,5-DICHLOROBENZYL)-3H-IMIDAZOL4-YL]-ETHYLAMINO}-4-METHYLPENTANOIC ACID (MLN-4760), IN AN ACTIVE SITE USING QUANTUM CHEMICAL APPROACH

INTRODUCTION

Angiotensin-converting enzyme 2 (ACE2) is a newly discovered membrane-bound aminopeptidase (Huentelman *et al*., 2004). The extracellular 3D structure of ACE2 protease domain is divided into two subdomains, which are defined by the movement of the subdomains relative to each other upon inhibitor binding. Subdomain I (N-terminal) contains the zinc-binding site, which faces into the deep cleft formed by the two subdomains connected at the base of the cleft, which is ~42% identical to the corresponding domains of human somatic and testicular ACE (Towler *et al*., 2004). This enzyme has been proven to be critical in impacting cardiovascular and immune systems (Huentelman *et al*., 2004). ACE has long been known as a major regulator of the Renin-Angiotensin System (RAS) (Figure 1) and represents a well-established drug target for the treatment of hypertension and heart disease. The discovery of ACE2 increases the complexity of the RAS-indicating a counter-regulatory role to ACE that makes it a promising new cardiorenal disease target (Rella *et al.*, 2006). ACE2 catalyzes the cleavage of Angiotensin I to Angiotensin 1-9 and may have a unique role in the local rennin-angiotensin system in heart and kidney (Natesh *et al*., 2003).

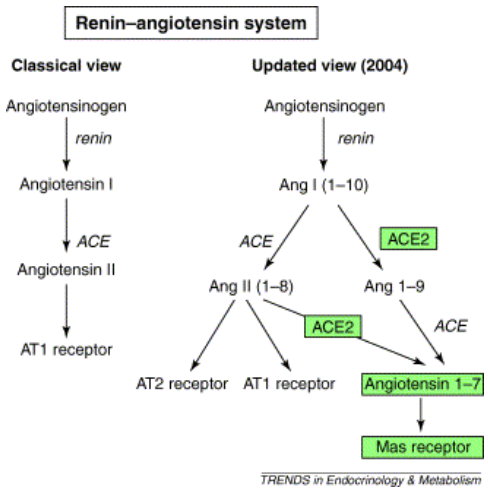


Figure 1 Renin–angiotensin system (RAS) pathway. Schematic diagram of the classical RAS (left), showing the main pathway for angiotensin II (Ang II) generation from Ang I via angiotensin-converting enzyme (ACE). In the classical view, Ang II mediates all known effects via the AT1 receptor. On the right is the updated view of the RAS, showing the role of ACE2 in degrading Ang I to Ang 1–9, and Ang II to the vasodilator Ang 1–7. In this version of the RAS, Ang II also mediates effects via the G-protein-coupled AT2 receptor, whereas Ang 1–7 acts through the Mas receptor.

Source: Burrell *et al.* (2004)

In a recent discovery, ACE2 was also identified as a functional receptor for the coronavirus that causes the severe acute respiratory syndrome (SARS) and its binding site on the SARS-CoV glycoprotein was localized between residues 303 and 537 (Li *et al.*, 2003; Wang *et al.*, 2004) (Figure 2). Therefore, ACE2 is proposed to be important for the development of novel pharmacotherapy against hypertension, related cardiovascular diseases, and SARS (Tikellis *et al*., 2003). The full-length, human ACE2 cDNA predicts a protein of 805 amino acids that has 42% homology with the N-terminal catalytic domain of ACE, 61% sequence similarity in a region surrounding the active site, and a hydrophobic region near the C-terminus, which serves as a membrane anchor. Unlike somatic ACE, ACE2 has only one active enzymatic site and functions as a carboxypeptidase rather than a dipeptidyl carboxypeptidase. Thus, ACE2 removes a single C-terminal Leu residue from Ang I

to generate Ang 1–9, a peptide with no known function. Although ACE2 was described originally for its ability to generate Ang 1–9 from Ang I, it also degrades Ang II to the biologically active peptide, Ang 1–7 (Figure 1). The high level of sequence similarity allowed many scientists to build a robust homology model of the ACE2 and used it to understand ACE2 nature. Despite the high sequence similarity, the substrate specificity and inhibitor sensitivity between ACE and ACE2 are distinct. Hence, the ACE inhibitors, captopril and lisinopril, are not inhibitors of ACE2 (Turner *et al*., 2002; Wang *et al*., 2004). In addition, the physiological and pathophysiological role of ACE2 is not yet clearly understood.

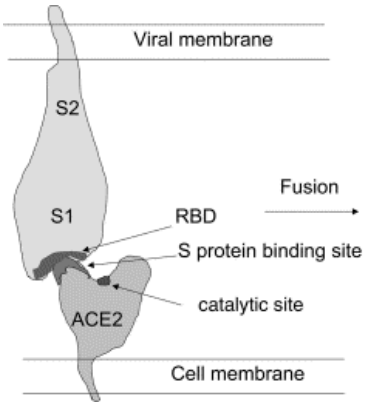


Figure 2 Schematic representation of the interaction between ACE2 and the SARS-CoV S-glycoprotein leading to binding and fusion. The RBD (receptor-binding domain) is depicted as a surface containing a cavity(s) that binds a ridge(s) close to the deep channel containing the catalytic site.

Source: Prabakaran *et al.* (2004)

Recently, the structure of human ACE2 was firstly solved by X-ray crystallography in two forms: apo-bound and inhibitor-bound (MLN-4760) (PDB entry: 1R42 and 1R4L) (Towler *et al.*, 2004). However, some interactions cannot directly be evaluated by the structural distances from the crystal structure. So, the objective in this study is to determine the tangible interactions of the inhibitor (MLN-4760) in the pocket. Based on these data, we test the hypothesis that understanding of ACE2 inhibitor binding interactions and implication for substrate specificity can be developed with the use of quantum chemical calculations approach.

LITERATURE REVIEW

Angiotensin-Converting Enzyme (ACE) is a popular drug target for hypertension and heart disease. Recently, a new ACE homologue has been identified, termed ACE2 whose physiological and pathophysiological roles are currently under investigation. Angiotensin-converting enzyme-2 (ACE2) is a membrane protein with its active site exposed to the extracellular surface of endothelial cells, the renal tubular epithelium and also the epithelia of the lung and the small intestine associated with hypertension, heart, and kidney disease and a counter-regulatory role to ACE has been proposed which makes it an interesting new cardio-renal disease target.

Before the ACE2 tertiary structure was solved, Dales *et al*. (2002) determined the minimal structural features required for binding to the ACE2 catalytic domain, several experiments under the assay conditions offered clues to the structural requirements for ACE2 inhibitors. Hence, many candidates were designed and tested their ability to inhibit ACE2. For design strategy, they conceived nonhydrolyzable His-Leu-like molecules bearing a C-terminal carboxylate and including a centrally located carboxylate to serve as a zinc coordinating element. They chose the carboxylate as the zinc binding fragment on the basis of precedence with a known carboxypeptidase A (CPDA) inhibitor, benzylsuccinic acid. The amino dicarboxylate core selected for evaluation is exemplified by 1a, which inhibits ACE2 in the micromolar range, Figure 3. This molecule appears to be capable of occupying both S1 and S1' and was amenable to further elaboration.

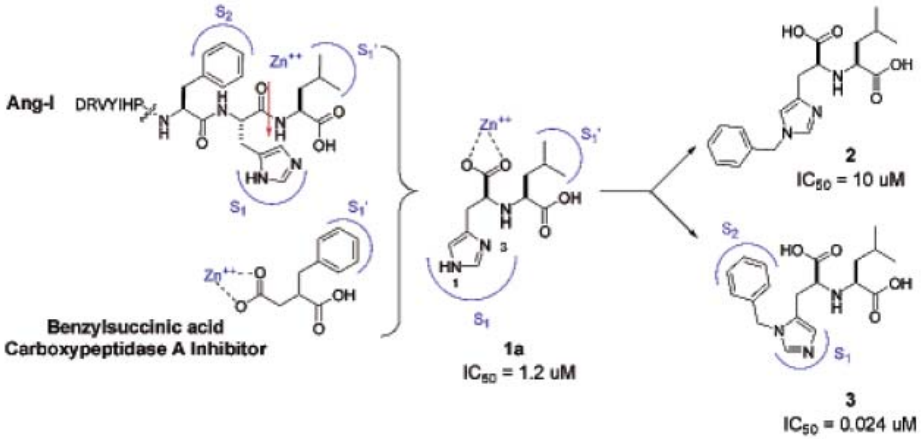


Figure 3 Design of ACE2 inhibitors.
Source: Dales *et al*. (2002)

In 2004, Towler *et al*. (2004) reported the first crystal structure of the extracellular metallopeptidase domain of human ACE2 in its native and inhibitor-bound. The protein structures were obtained by X-ray crystallographic study with resolution of 2.2 Å and 3.0 Å, respectively (PDB codes: 1R4L, 1R42). Comparison with ACE proteins indicated that human ACE2 is mostly similar to human testicular ACE. It is 33% identical overall to human testicular ACE. Based on the solved ACE2 crystal structure with its inhibitor, a structure-based drug design project is undertaken to identify novel potent and selective inhibitors.

Nonetheless, before the first ACE2 crystal structures were solved, Guy *et al.* (2003) and Prabakaran *et al*. (2004) proposed a model of the ACE2 using comparative modeling, based on the crystal structure of testicular ACE. Because a comprehensive analysis of the structure and function of the catalytic site was very recently reported by Guy *et al.* (2003), Prabakaran *et al*. (2004) used the enzymatic site location for reference purposes. They suggested a possible binding region for the SARS-CoV *S*-glycoprotein on ACE2 and could help in the design of experiments to further elucidate the structure and function of ACE2. Nevertheless, in case of a comprehensive analysis of the structure and function of the catalytic site, Guy *et al.* (2003) found that structural differences exist between the active site of ACE (dipeptidyl carboxypeptidase) and ACE2 (carboxypeptidase) that are responsible for

the differences in specificity. The main differences occur in the ligand-binding pockets, particularly at the S2' subsite and in the binding of the peptide carboxy-terminus. The model based on the crystal structure of tACE, indicated that the catalytic mechanism of ACE2 resembles that of ACE. The predicted structural differences between the active site of ACE (dipeptidyl carboxypeptidase) and ACE2 (carboxypeptidase) were also verified by the ACE2 structure. ACE2 functions as a carboxypeptidase and is not susceptible to inhibition by the classical ACE inhibitors. The ACE inhibitors, lisinopril and captopril, do not have an ability to inhibit the ACE2. The evidence from Guy *et al*. (2003) provided by the comparative modeling indicates that the main differences occur in the ligand-binding pockets, particularly at the S2' subsite and the binding of the peptide carboxy-terminus. The cavity in tACE is larger than that of ACE2, allowing an extra amino acid to bind in the specificity pocket. In particular, tACE Gln281, which points into the binding pocket, forms a hydrogen bond with the carboxy-terminus of lisinopril (*N*-[(*S*)-1-carboxy-3-phenylpropyl]-L-lysyl-L-proline) (Figure 4B). This residue is Arg273 in ACE2, representing the insertion of an ethylene group into the side chain at this position. This causes steric conflict with the proline carboxy-terminus when the tACE conformation of lisinopril is docked into an identical position in the ACE2 model. Removing the proline carboxy-terminus allows Arg273 from the model to make a salt bridge with the lysyl carboxy-terminus of *N*-[(*S*)-1-carboxy-3-phenylpropyl]-L-lysine (NCPL). His505 and His345 of ACE2, also present in tACE (as His513 and His353), can hydrogen bond with the new carboxy-terminus. Other changes that may facilitate the switch in substrate specificity are the conversion of Lys511 and Tyr520 in tACE (which form a salt bridge and hydrogen bond, respectively, to the proline carboxy-terminus) to the non-hydrogen bonding residues Leu503 and Phe512, respectively, and conversion of Thr282 in tACE to the more bulky Phe274 in ACE2. These changes in ACE2 serve to fill out the S2' binding pocket present in tACE. Also of note is the presence of Arg518 in ACE2, which is modeled to partially occupy the S2' subsite but can be remodeled in a fully extended conformation to make a hydrogen bond with the lysyl carboxy-terminus of NCPL. However, it is unlikely that the guanidinium groups of Arg273 and Arg518 can simultaneously bind the carboxy-terminus due to their close proximities. The model therefore provides a rational

explanation for the observed substrate specificity differences between ACE (acting as a peptidyl dipeptidase) and ACE2 (acting as a carboxypeptidase). Furthermore, with the S2' site in ACE2 being occupied, it explains why none of the commonly used ACE inhibitors, such as lisinopril, captopril, and enalaprilat (which are based on dipeptide structures and bind in the S1' and S2' pockets of ACE), inhibit ACE2 (Figure 4).

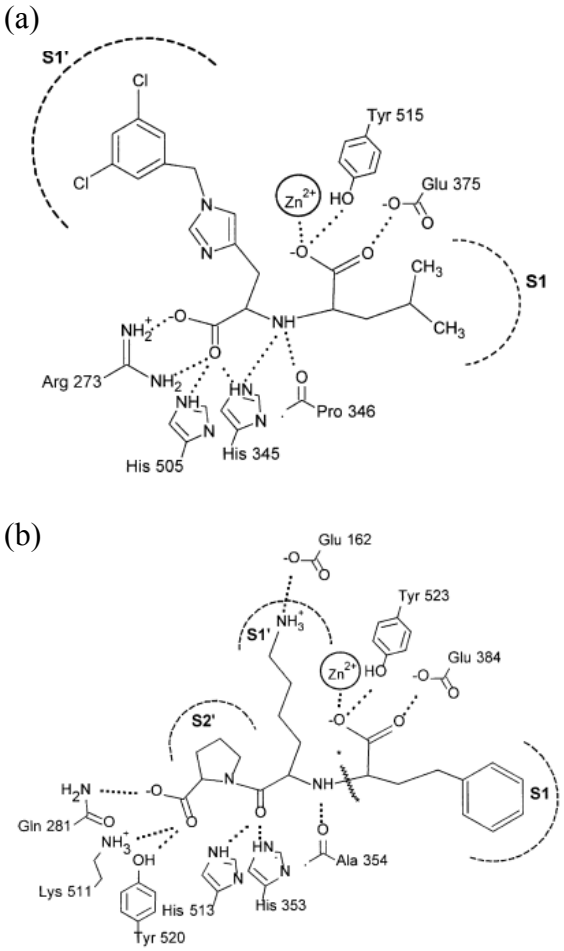


Figure 4 Schematic view of the active site of ACE2 and tACE. Binding interactions of the inhibitor (a) MLN-4760 at the active site of ACE2 and (b) lisinopril at the active site of tACE. Hydrogen bonds to the ligand are shown (dotted lines). The different binding subsites are labeled.

Source: Guy *et al*. (2005)

Admittedly, the inhibitor-bound crystal structure, MLN-4760, reveals important residues responsible for inhibitor binding and presumably substrate binding and catalysis. It was found that the important difference between ACE2 and ACE is the substitution of Arg273 for Gln. Guy *et al*. (2005) has highlighted the importance of Arg273 in binding of the C-terminus of peptide substrates by ACE2. The complete loss of activity observed as a result of replacing this single residue could not have been predicted simply by analysis of the active-site structure. Moerover, it was found that Arg273 plays a key role in stabilizing the MLN-4760 inhibitor and peptide substrate by making a salt-bridge with the C-terminus of the ACE2 inhibitor, MLN-4760, and is hence proposed to be involved in binding of the C-terminus of peptide substrates.

In 2006, Rella *et al*. have attempted a structure-based approach to identify novel potent and selective inhibitors. Computational approaches focus on pharmacophore-based virtual screening of large compound databases. The software packages LigandScout and Catalyst were used for identification and visualization of protein-ligand interaction sites, pharmacophore model generation, and database searching, respectively. Chemical feature-based pharmacophore models were generated based on the ACE2 X-ray crystal structure in complex with a potent inhibitor. Selectivity was ensured by initial screening for ACE inhibitors within an internal database and the Derwent World Drug Index. Seventeen compounds were selected based on high fit values as well as diverse structure and subjected to experimental validation in a bioassay. They also show that the six most promising candidates exhibiting IC_{50} values in the range of 62-179 μM.

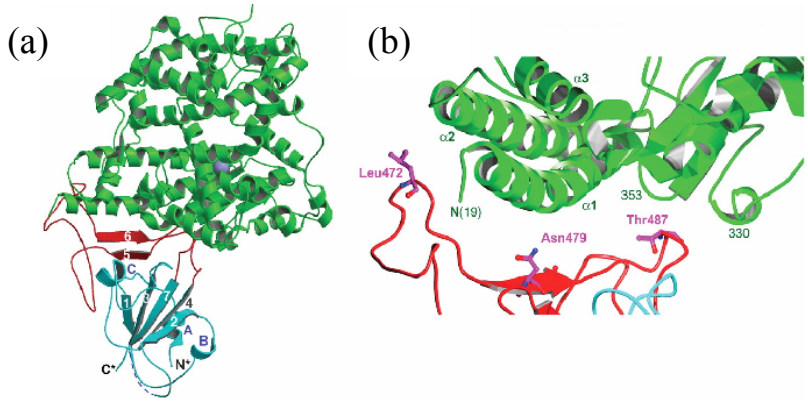


Figure 5 (a) Crystal structure of the receptor-binding domain (core structure in cyan and receptor-binding motif in red) in complex of the human receptor ACE2 (green). (b) Detail of the binding interface, with side chains of three residues (Leu472, Asn479, and Thr487 from left to right) critical for cross-species and human-to-human transmission of SARS-CoV.

Source : Li *et al*. 2005

Moreover, ACE2 was found as a receptor of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), which emerged as a serious epidemic in 2002 to 2003, with over 8,000 infected cases and a fatality rate of ~10%. Also, in 2005, the first crystal structure at 2.9 Å of SARS-CoV spike receptor-binding domain complexed with ACE2 was solved by Li *et al.* (2005) (Figure 5). The atomic details at the interface between the two proteins clarify the importance of residue changes that facilitate efficient cross-species infection and human-to-human transmission.

METHODS OF CALCULATIONS

1. Structure collection

All structures of human angiotensin converting enzymes-related carboxy-petidase investigated in this study (PDB code: 1R4L, 1R42, and 2AJF) were collected from Protein data bank (www.rcsb.org). Because the X-ray structure has no hydrogen atoms, they are added to the geometrical structure to generate the complete structure of the model by using Gview03W.

2. Structural analysis

In order to compare the conformations of ACE2 in term of the inhibitor-bound, SARS-bound, and apo-bound structures, the set of RMSDs were calculated by the Spdbv program.

3. Density Functional Theory

Density-functional theory is one of the most popular and successful quantum mechanical approaches to matter. It is nowadays routinely applied for calculating, e.g., the binding energy of molecules in chemistry and the band structure of solids in physics. First applications relevant for fields traditionally considered more distant from quantum mechanics, such as biology and mineralogy are beginning to appear. Superconductivity, atoms in the focus of strong laser pulses, relativistic effects in heavy elements and in atomic nuclei, classical liquids, and magnetic properties of alloys have all been studied with DFT.

DFT owes this versatility to the generality of its fundamental concepts and the flexibility one has in implementing them. In spite of this flexibility and generality, DFT is based on quite a rigid conceptual framework. This section introduces some aspects of this framework in general terms. The following sections then deal in detail

with two core elements of DFT, the Hohenberg-Kohn theorem, the Kohn-Sham equations, a (necessarily less detailed) description of approximations typically made in practical DFT calculations, and of some extensions and generalizations of DFT.

To get a first idea of what density-functional theory is about, it is useful to take a step back and recall some elementary quantum mechanics. In quantum mechanics we learn that all information we can possibly have about a given system is contained in the system's wave function. Here we will exclusively be concerned with the electronic structure of atoms, molecules and solids. The nuclear degrees of freedom (e.g., the crystal lattice in a solid) appear only in the form of a potential *v(r)* acting on the electrons, so that the wavefunction depends only on the electronic coordinates. Nonrelativistically, this wave function is calculated from Schrödinger's equation, which for a single electron moving in a potential *v(r)* reads

$$\left[-\frac{\hbar^2\nabla^2}{2m}+v(r)\right]\Psi(r)=\varepsilon\Psi(r) \qquad (1)$$

If there is more than one electron (i.e., one has a many-body problem) Schrödinger's equation becomes

$$\left[\sum_i^N\left(-\frac{\hbar^2\nabla_i^{\,2}}{2m}+v(r)\right)+\sum_{i<j}U(r_i,r_j)\right]\Psi(r_1,r_2...,r_N)=E\Psi(r_1,r_2...,r_N), \qquad (2)$$

where N is the number of electrons and $U(r_i,r_j)$ is the electron-electron interaction. For a Coulomb system (the only type of system we consider here), one has

$$\hat{U}=\sum_{i<j}U(r_i,r_j)=\sum_{i<j}\frac{q^2}{|r_i-r_j|} \qquad (3)$$

Note that this is the same operator for any system of particles interacting via the Coulomb interaction, just as the kinetic energy operator

$$T = -\frac{\hbar^2}{2m}\sum_i \nabla_i^2 \tag{4}$$

is the same for any nonrelativistic system. Whether our system is an atom, a molecule, or a solid thus depends only on the potential $v(r_i)$. For an atom, e.g,

$$\hat{V} = \sum_i v(r_i) = \sum_i \frac{Qq}{|r_i - R|} \tag{5}$$

where Q is the nuclear charge and R the nuclear position. When dealing with a single atom, R is usually taken to be the zero of the coordinate system. For a molecule or a solid one has

$$\hat{V} = \sum_i v(r_i) = \sum_{ik} \frac{Qq}{|r_i - R_k|} \tag{6}$$

where the sum on k extends over all nuclei in the system, each with charge $Q_k = Z_k e$ and position R_k. It is only the spatial arrangement of the R_k (together with the corresponding boundary conditions) that distinguishes, fundamentally, a molecule from a solid. Similarly, it is only through the term $\hat{U}$ that the (essentially simple) single-body quantum mechanics of equation (1) differs from the extremely complex many-body problem posed by equation (2). These properties are built into DFT in a very fundamental way. The usual quantum-mechanical approach to Schrödinger's equation (SE) can be summarized by the following sequence

$$v(r) \overset{SE}{==>} \Psi(r_1, r_2 ..., r_N) \overset{<\Psi|...|\Psi>}{=====>} observables \tag{7}$$

i.e., one specifies the system by choosing $v(r)$, plugs it into Schrödinger's equation, solves that equation for the wave function, and then calculates observables by taking

expectation values of operators with this wave function. One among the observables that are calculated in this way is the particle density.

$$n(r) = N\int d^3r_2 \int d^3r_3 ... \int d^3r_N \Psi^*(r, r_2 ..., r_N)\Psi(r, r_2 ..., r_N) \qquad (8)$$

Many powerful methods for solving Schrödinger's equation have been developed during decades of struggling with the many-body problem. In physics, for example, one has diagrammatic perturbation theory (based on Feynman diagrams and Green's functions), while in chemistry one often uses configuration interaction (CI) methods, which are based on systematic expansion in Slater determinants. A host of more special techniques also exists. The problem with these methods is the great demand they place on one's computational resources: it is simply impossible to apply them efficiently to large and complex systems. Nobody has ever calculated the chemical properties of a 100-atom molecule with full CI, or the electronic structure of a real semiconductor using nothing but Green's functions.

It is here where DFT provides a viable alternative, less accurate perhaps, but much more versatile. DFT explicitly recognizes that nonrelativistic Coulomb systems differ only by their potential v(r), and supplies a prescription for dealing with the universal operators $\hat{T}$ and $\hat{U}$ once and for all. Furthermore, DFT provides a way to systematically map the many-body problem, with $\hat{U}$, onto a single-body problem, without $\hat{U}$. All this is done by promoting the particle density *n(r)* from just one among many observables to the status of key variable, on which the calculation of all other observables can be based. This approach forms the basis of the large majority of electronic-structure calculations in physics and chemistry. Much of what we know about the electrical, magnetic, and structural properties of materials has been calculated using DFT, and the extent to which DFT has contributed to the science of molecules is reflected by the 1998 Nobel Prize in Chemistry, which was awarded to Walter Kohn, the founding father of DFT, and John Pople, who was instrumental in

implementing DFT in computational chemistry. The density-functional approach can be summarized by the sequence.

$$n(r) \Longrightarrow \Psi(r_1,...,r_N) \Longrightarrow v(r) \qquad (8)$$

i.e., knowledge of *n(r)* implies knowledge of the wavefunction and the potential, and hence of all other observables. Although this sequence describes the conceptual structure of DFT, it does not really represent what is done in actual applications of it, which typically proceed along rather different lines, and do not make explicit use of many-body wavefunctions. The following sections attempt to explain both the conceptual structure and some of the many possible shapes and disguises under which this structure appears in applications.

The Hohenberg-Kohn theorem

At the heart of DFT is the Hohenberg-Kohn (HK) theorem. This theorem states that for ground states equation (8) can be inverted: given a ground state density $n_0(r)$ it is possible, in principle, to calculate the corresponding ground-state wavefunction $\Psi_0(r_1, r_2, ..., r_N)$. This means that Ψ_0 is a functional of n_0 . Consequently, all ground-state observables are functionals of n_0, too. If Ψ_0 can be calculated from n_0 and vice versa, both functions are equivalent and contain exactly the same information. At first sight this seems impossible: how can a function of one (vectorial) variable r be equivalent to a function of *N* (vectorial) variables $r_1 \ldots r_N$? How can one arbitrary variable contain the same information as *N* arbitrary variables?

The crucial fact which makes this possible is that knowledge of $n_0(r)$ implies implicit knowledge of much more than that of an arbitrary function *f(r)*. The ground-state wavefunction ψ_0 must not only reproduce the ground-state density, but also minimize the energy. For a given ground-state density $n_0(r)$, we can write this requirement as

$$E_{v,0} = \min_{\Psi->n_0} < \Psi \mid \hat{T} + \hat{U} + \hat{V} \mid \Psi > \qquad (9)$$

where $E_{v,0}$ denotes the ground-state energy in potential $v(r)$. The preceding equation tells us that for a given density $n_0(r)$ the ground-state wavefunction ψ_0 is that which reproduces this $n_0(r)$ and minimizes the energy.

For an arbitrary density $n(r)$, we define the functional

$$E_v[n] = \min_{\Psi->n} < \Psi \mid \hat{T} + \hat{U} + \hat{V} \mid \Psi > \qquad (10)$$

If n is a density different from the ground-state density n_0 in potential $v(r)$, then the ψ that produce this n are different from the ground-state wavefunction ψ_0, and according to the variational principle the minimum obtained from $E_v[n]$ is higher than (or equal to) the ground-state energy $E_{v,0} = E_v[n_0]$. Thus, the functional $E_v[n]$ is minimized by the ground-state density n_0, and its value at the minimum is $E_{v,0}$.

The total-energy functional can be written as

$$E_v[n] = \min_{\Psi->n} < \Psi \mid \hat{T} + \hat{U} \mid \Psi > + \int d^3 r n(r) v(r) =: F[n] + V[n] \qquad (11)$$

where the internal-energy functional $F[n] = \min_{\Psi->n} < \Psi \mid \hat{T} + \hat{U} \mid \Psi >$ (universal functionals) is independent of the potential $v(r)$, and thus determined only by the structure of the operators $\hat{U}$ and $\hat{T}$. This universality of the internal-energy functional allows us to define the ground-state wavefunction ψ_0 as that antisymmetric N-particle function that delivers the minimum of $F[n]$ and reproduces n_0. If the ground state is nondegenerate, this double requirement uniquely determines ψ_0 in terms of $n_0(r)$, without having to specify $v(r)$ explicitly.

equations (9) to (11) constitute the constrained-search proof of the Hohenberg-Kohn theorem. The original proof by Hohenberg and Kohn proceeded by assuming that ψ_0 was not determined uniquely by n_0 and showed that this produced a contradiction to the variational principle. Both proofs, by constrained search and by contradiction, are elegant and simple. In fact, it is a bit surprising that it took 38 years from Schrödinger's first papers on quantum mechanics to Hohenberg and Kohn's 1964 paper containing their famous theorem. Since 1964, the HK theorem has been thoroughly scrutinized, and several alternative proofs have been found. One of these is the so-called 'strong form of the Hohenberg-Kohn theorem', based on the inequality.

$$\int d^3 r \Delta n(r) \Delta v(r) < 0 \qquad (12)$$

Here $\Delta v(r)$ is a change in the potential, and $\Delta n(r)$ is the resulting change in the density. We see immediately that if $\Delta v \neq 0$ we cannot have $\Delta n(r) \equiv 0$, i.e., a change in the potential must also change the density. This observation implies again the HK theorem for a single density variable: there cannot be two local potentials with the same ground-state charge density. A given-particle ground-state density thus determines uniquely the corresponding potential, and hence also the wavefunction. Moreover, equation (12) establishes a relation between the signs of $\Delta n(r)$ and $\Delta v(r)$: if Δv is mostly positive, $\Delta n(r)$ must be mostly negative, so that their integral over all space is negative. This additional information is not immediately available from the two classic proofs, and is the reason why this is called the 'strong' form of the HK theorem. Equation (12) can be obtained along the lines of the standard HK proof, but it can be turned into an independent proof of the HK theorem because it can also be derived perturbatively.

Another alternative argument is valid only for Coulomb potentials. It is based on Kato's theorem, which states that for such potentials the electron density has a cusp at the position of the nuclei, where it satisfies

$$Z_k = -\frac{a_0}{2n(r)}\frac{dn(r)}{dr}\bigg|_{r->R_k} \qquad (13)$$

Here R_k denotes the positions of the nuclei, Z_k their atomic number, and $a_0 = \frac{\hbar^2}{2me}$ is the Bohr radius. For a Coulomb system one can thus, in principle, read off all information necessary for completely specifying the Hamiltonian directly from examining the density distribution: the integral over *n(r)* yields *N*, the total particle number; the position of the cusps of *n(r)* are the positions of the nuclei, R_k; and the derivative of *n(r)* at these positions yields Z_k by means of equation (13). This is all one needs to specify the complete Hamiltonian of equation (2) (and thus implicitly all its eigenstates). In practice one almost never knows the density distribution sufficiently well to implement the search for the cusps and calculate the local derivatives. Still, Kato's theorem provides a vivid illustration of how the density can indeed contain sufficient information to completely specify a nontrivial Hamiltonian.

The Kohn-Sham Equations

Density-functional theory can be implemented in many ways. The minimization of an explicit energy functional, discussed up to this point, is not normally the most efficient among them. Much more widely used is the Kohn- Sham approach. Interestingly, this approach owes its success and popularity partly to the fact that it does not exclusively work in terms of the particle (or charge) density, but brings a special kind of wavefunctions (single-particle orbitals) back into the game. As a consequence DFT then looks formally like a single-particle theory, although many-body effects are still included via the so-called exchange-correlation functional. We will now see in some detail how this is done.

Excange-correlation energy

The Thomas-Fermi approximation for *T[n]* is not very good. A more accurate scheme for treating the kinetic-energy functional of interacting electrons, *T[n]*, is

based on decomposing it into one part that represents the kinetic energy of noninteracting particles of density *n*, i.e., the quantity called above $T_s[n]$, and one that represents the remainder, denoted $T_c[n]$ (the sub-scripts *s* and *c* stand for 'single-particle' and 'correlation', respectively).

$$T[n] = T_s[n] + T_c[n] \tag{14}$$

$T_s[n]$ is not known exactly as a functional of *n*, but it is easily expressed in terms of the single-particle orbitals $\varphi i(r)$ of a noninteracting system with density *n*, as

$$T_s[n] = -\frac{\hbar^2}{2m}\sum_i^N \int d^3 r \phi_i^*(r)\nabla^2\phi_i(r) \tag{15}$$

because for noninteracting particles the total kinetic energy is just the sum of the individual kinetic energies. Since all $\varphi i(r)$ are functionals of *n*, this expression for T_s is an explicit orbital functional but an implicit density functional, $T_s[n] = T_s[\{\varphi_i[n]\}]$, where the notation indicates that T_s depends on the full set of occupied orbitals φ_i, each of which is a functional of *n*. Other such orbital functionals will be discussed later.

We now rewrite the exact energy functional as

$$E[n] = T[n] + U[n] + V[n] = T_s\left[\{\phi_i[n]\}\right] + U_H[n] + E_{xc}[n] + V[n] \tag{16}$$

where by definition E_{xc} contains the differences $T - T_s$ (i.e. T_c) and $U - U_H$. This definition shows that a significant part of the correlation energy E_c is due to the difference T_c between the noninteracting and the interacting kinetic energies. Equation (16) is formally exact, but of course E_{xc} is unknown — although the HK theorem guarantees that it is a density functional. This functional, $E_{xc}[n]$, is called the exchange-correlation (*xc*) energy. It is often decomposed as $E_{xc} = E_x + E_c$, where E_x is due to the Pauli principle (exchange energy) and E_c is due to correlations. (T_c is then a

part of E_c.) The exchange energy can be written explicitly in terms of the single-particle orbitals as

$$E_x\left[\{\phi_i[n]\}\right] = -\frac{q^2}{2}\sum_{jk}\int d^3r \int d^3r' \int \frac{\phi_i^*(r)\phi_k^*(r)\phi_j(r')\phi_k(r)}{|r-r'|} \qquad (17)$$

which is known as the Fock term, but no general exact expression in terms of the density is known.

There are basically three distinct types of approximations involved in a DFT calculation. One is conceptual, and concerns the interpretation of KS eigenvalues and orbitals as physical energies and wavefunctions. The second type of approximation is numerical, and concerns methods for actually solving the differential equation. A main aspect here is the selection of suitable basis functions. The third type of approximation involves constructing an expression for the unknown *xc* functional $E_{xc}[n]$, which contains all many-body aspects of the problem. It is with this type of approximation that we are concerned in the present section. This part is intended to give the reader an idea of what types of functionals exist, and to describe what their main features are, separately for local functionals (TF, LDA and Xα), semilocal, or gradient dependent, functionals (GEA and GGA), and nonlocal functionals (hybrids, orbital functionals such as meta-GGAs, EXX and SIC, and integral dependent functionals such as ADA). This chapter does deal only most superficially with the actual construction of these functionals.

Local functionals: LDA

Historically (and in many applications also practically) the most important type of approximation is the local-density approximation (LDA). To understand the concept of an LDA recall first how the noninteracting kinetic energy $T_s[n]$ is treated in the Thomas-Fermi approximation: In a homogeneous system one knows that,

$$t_s^{\text{hom}}(n) = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3} n^{5/3} \tag{18}$$

where $n = const$. In an inhomogeneous system, with $n = n(r)$, one approximates locally

$$t_s(r) \approx t_s^{\text{hom}}(n(r)) = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3} n(r)^{5/3} \tag{19}$$

and obtains the full kinetic energy by integration over all space

$$T_s^{LDA}[n] = \int d^3 r t_s^{\text{hom}}(n(r)) = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3} \int d^3 r n(r)^{5/3} \tag{20}$$

For the kinetic energy the approximation $T_s[n] \approx T_s^{LDA}[n]$ is much inferior to the exact treatment of T_s in terms of orbitals, offered by the Kohn-Sham equations, but the LDA concept turned out to be highly useful for another component of the total energy, the exchange-correlation energy $E_{xc}[n]$. For the exchange energy $E_x[n]$ the procedure is very simple, since the per volume exchange energy of the homogeneous electron liquid is known exactly

$$e_x^{\text{hom}}(n) = -\frac{3q^2}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{4/3} \tag{21}$$

so that

$$E_x^{LDA}[n] = -\frac{3q^2}{4}\left(\frac{3}{\pi}\right)^{1/3} \int d^3 r n(r)^{4/3} \quad . \tag{22}$$

This is the LDA for E_x

For the correlation energy $E_c[n]$ the situation is more complicated since $e_c^{\text{hom}}(n)$ is not known exactly: the determination of the correlation energy of a homogeneous

interacting electron system (an electron liquid) is already a difficult many-body problem on its own! Early approximate expressions for $e_c^{\mathrm{hom}}(n)$ were based on applying perturbation theory (e.g. the random phase approximation) to this problem. These approximations became outdated with the advent of highly precise Quantum Monte Carlo (QMC) calculations for the electron liquid, by Ceperley and Alder. Modern expressions for $e_c^{\mathrm{hom}}(n)$ are parameterizations of these data. These expressions are implemented in most standard DFT program packages and in typical applications give almost identical results. On the other hand, the earlier parameterizations of the LDA, based on perturbation theory, can occasionally deviate substantially from the QMC ones, and are better avoided. Independently of the parameterizations, the LDA for $E_{xc}[n]$ formally consists in

$$E_{xc}[n] \approx E_{xc}^{LDA}[n] = \int d^3 r e_{xc}^{\mathrm{hom}}(n)\,|_{n->n(r)} = \int d^3 r e_{xc}^{\mathrm{hom}}(n(r)) \qquad (23)$$

where $e_{xc}^{\mathrm{hom}} = e_x^{\mathrm{hom}} + e_c^{\mathrm{hom}}$. The corresponding *xc* potential is simply

$$v_{xc}^{LDA}[n](r) = \left.\frac{\partial e_{xc}^{\mathrm{hom}}(n)}{\partial n}\right|_{n->n(r)} \qquad (24)$$

For many decades the LDA has been applied in, e.g., calculations of band structures and total energies in solid-state physics. In quantum chemistry it is much less popular, because it fails to provide results that are accurate enough to permit a quantitative discussion of the chemical bond in molecules (so-called 'chemical accuracy' requires calculations with an error of not more than about 1 kcal/mol = 0.04336 eV/particle). At this stage it may be worthwhile to recapitulate what practical DFT does, and where the LDA enters its conceptual structure: What real systems, such as atoms, molecules, clusters and solids, have in common, is that they are simultaneously inhomogeneous (the electrons are exposed to spatially varying electric fields produced by the nuclei) and interacting (the electrons interact via the Coulomb interaction). The way density-functional theory, in the local-density approximation, deals with this inhomogeneous

many-body problem is by decomposing it into two simpler (but still highly nontrivial) problems: the solution of a spatially homogeneous interacting problem (the homogeneous electron liquid) yields the uniform *xc* energy $e_c^{\mathrm{hom}}(n)$ and the solution of a spatially inhomogeneous noninteracting problem (the inhomogeneous electron gas described by the KS equations) yields the particle density. Both steps are connected by the local-density potential (24), which shows how the *xc* energy of the uniform interacting system enters the equations for the inhomogeneous noninteracting system. The particular way in which the inhomogeneous many-body problem is decomposed, and the various possible improvements on the LDA, is behind the success of DFT in practical applications of quantum mechanics to real materials. Some such improvements on the LDA are discussed in the next two sections.

Semilocal functionals: GEA, GGA and beyond

In the LDA one exploits knowledge of the density at point r. Any real system is spatially inhomogeneous, i.e., it has a spatially varying density *n(r),* and it would clearly be useful to also include information on the rate of this variation in the functional. A first attempt at doing this was the so-called 'gradient-expansion approximations' (GEA). In this class of approximation one tries to systematically calculate gradient-corrections of the form $|\Delta n(r)|$, $|\Delta n(r)|^2$, $\Delta^2 n(r)$, etc., to the LDA. A famous example is the lowest-order gradient correction to the Thomas-Fermi approximation for $T_s[n]$,

$$T_s[n] \approx T_s^W[n] = T_s^{LDA}[n] + \frac{\hbar^2}{8m}\int d^3r \frac{|\nabla n(r)|^2}{n(r)} \tag{25}$$

This second term on the right-hand side is called the Weizsacker term. Similarly, in

$$E_x[n] \approx E_x^{GEA(2)}[n] = E_x^{LDA}[n] - \frac{10q^2}{432\pi(3\pi^2)^{1/3}}\int d^3r \frac{|\nabla n(r)|^2}{n(r)^{4/3}} \tag{26}$$

the second term on the right-hand side is the lowest-order gradient correction to $E_x^{LDA}[n]$. In practice, the inclusion of low-order gradient corrections almost never improves on the LDA, and often even worsens it. Higher-order corrections (e.g., $\alpha|\nabla n(r)|^{\alpha}$ or $\beta\nabla^{\beta} n(r)$ with $\alpha, \beta > 2$), on the other hand, are exceedingly difficult to calculate, and little is known about them. In this situation it was a major breakthrough when it was realized, in the early eighties, that instead of power-series-like systematic gradient expansions one could experiment with more general functions of $n(r)$ and $\nabla n(r)$, which need not proceed order by order. Such functionals, of the general form

$$E_{xc}^{GGA}[n] = \int d^3 r f(n(r), \nabla n(r)) \qquad (27)$$

have become known as generalized-gradient approximations (GGAs)

Different GGAs differ in the choice of the function $f(n, \nabla n)$. Note that this makes different GGAs much more different from each other than the different parameterizations of the LDA: essentially there is only one correct expression for $e_{xc}^{\text{hom}}[n]$, and the various parameterizations of the are merely different ways of writing it. On the other hand, depending on the method of construction employed for obtaining $f(n, \nabla n)$ one can obtain very different GGAs. In particular, GGAs used in quantum chemistry typically proceed by fitting parameters to test sets of selected molecules. On the other hand, GGAs used in physics tend to emphasize exact constraints. Nowadays the most popular (and most reliable) GGAs are PBE (denoting the functional proposed in 1996 by Perdew, Burke and Ernzerhof (Perdew *et al.*, 1997)) in physics, and BLYP (denoting the combination of Becke's 1988 exchange functional (Becke *et al*., 1988) with the 1988 correlation functional of Lee, Yang and Parr (Lee *et al*., 1988)) in chemistry. Many other GGA-type functionals are also available, and new ones continue to appear. Quite generally, current GGAs seem to give reliable results for all main types of chemical bonds (covalent, ionic, metallic and hydrogen bridge). For van der waals interactions, however, common GGAs and LDA fail. To describe these very weak interactions several more specialized approaches

have been developed within DFT. Both in physics and in chemistry the widespread use of GGAs has lead to major improvements as compared to LDA. 'Chemical accuracy', as defined above, has not yet been attained, but is not too far away either.

The hybrid method

In spite of these advances, the quest for more accurate functionals goes ever on, and both in chemistry and physics various beyond-GGA functionals have appeared. Perhaps the most popular functional in quantum chemistry is B3LYP. This is a combination of the LYP GGA for correlation (Lee *et al.*, 1988) with Becke's three-parameter hybrid functional B3 for exchange (Becke *et al*., 1988). Common hybrid functionals, such as B3, mix a fraction of Hartree-Fock exchange into the DFT exchange functional (other mixtures are also possible). The construction of hybrid functional involves a certain amount of empiricism in the choice of functionals that are mixed and in the optimization of the weight factors given to the HF and DFT terms. Formally, this might be considered a drawback, but in practice B3 has proven to be the most successful exchange functional for chemical applications, in particular when combined with the LYP GGA functional for E_c. More extreme examples of this semiempirical mode of construction of functionals are Becke's 1997 hybrid functional (Becke *et al.,* 1997)

To get more reliable and reasonable results because this model comprise the metal ion and certain charged amino acid, we used the high level of calculation of the hybrid method (B3LYP) to investigate the system properties.

4. Investigation of the interaction energies

In order to study the individual interactions between the inhibitor, MLN-4760, and each amino acid in the binding site, the structure of inhibitor-bound ACE2 is cut into the small system. The studied binding pocket included residues surrounding the ACE2 binding pocket and MLN-4760 within the interatomic distance of 7 Å. Therefore, the system includes 22 residues (Glu145, Asn149, Arg273, Phe274,

His345, Pro346, Thr347, Met360, Thr362, Asp368, Thr371, His374, Glu375, His378, His401, Glu402, Phe504, His505, Tyr510, Ser511, Arg514, and Typ515), its inhibitor, MLN-4760, and also Zn^{2+} ion in order to understand the zinc-binding effect. The model is treated at a high level of calculations using the B3LYP/6-31G(d,p) methods in order to cover the electronic properties of Zn^{2+} ion in an active site. The backbones of all amino acids are fixed in the same position as those in X-ray structure to fully understand the particular effect of side chains on the inhibiting binding model as well as reduce the computational cost due to the massive size of this model. Furthermore, the particular interaction between inhibitor and the each nearby amino acids were calculated using the following equation,

$$E_{int} = E_{cplx} - (E_{MLN} + E_{AA}) \quad (19)$$

where E_{int} is the particular interaction energy, E_{cplx} is the energy of the complex between each amino acid and an inhibitor, E_{MLN} is the energy of MLN-4760, and E_{AA} is the energy of each amino acid. In case of the residue located in the zinc-binding area, we also calculated both the system containing zinc ion and another getting rid of zinc ion so as to elucidate the zinc-binding effect. So, we used equation 19 for the non-zinc model and equation 20 for the zinc-containing system.

$$E_{int} = E_{cplx+Zn^{2+}} - (E_{MLN} + E_{AA} + E_{Zn^{2+}}) \quad (20)$$

where $E_{Zn^{2+is}}$ the energy of Zn^{2+} and $E_{cplx+Zn^{2+}}$ is the energy of the complex including zinc ion.

RESULTS AND DISCUSSIONS

1. Structural analysis of SARS-bound, inhibitor-bound, and native ACE2

The differences of the extracellular region of human ACE2 enzyme in forms of SARS-bound, inhibitor-bound, and native three-dimensional structures (Figure 6) were investigated by backbone-fixed RMSD calculations where the native acted as a reference. It was found that the RMSDs of SARS-bound and MLN-4760-bound complexes were 0.48 Å (Figure 7) and 3.57 Å (Figure 8a), respectively.

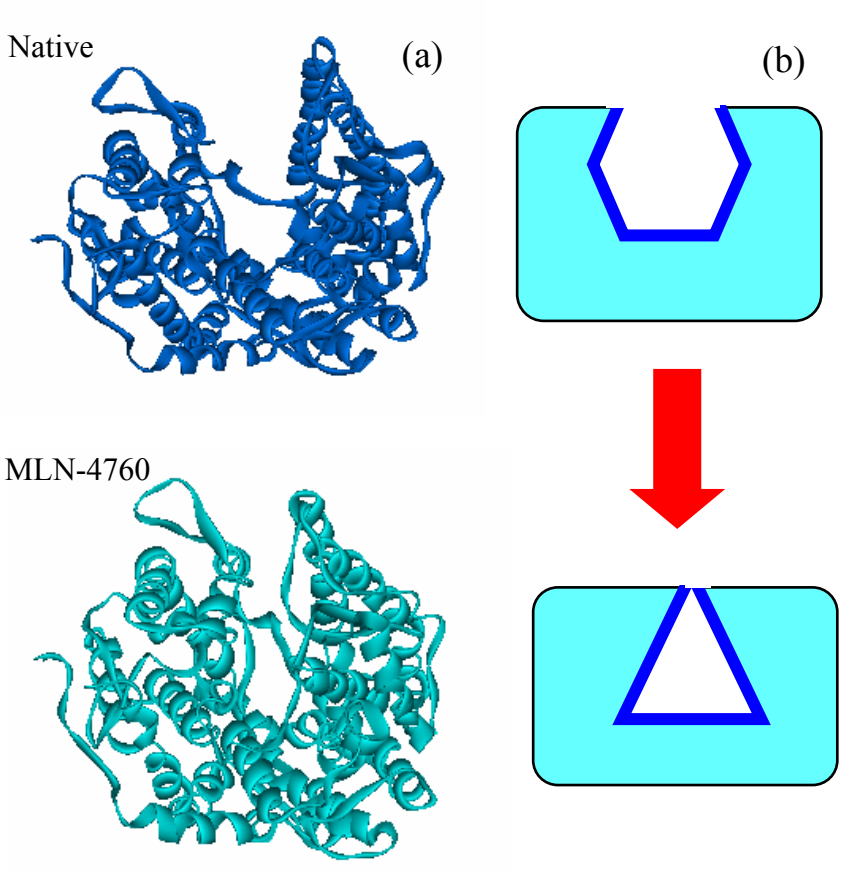


Figure 6 The ACE2 structures of the apo-bound and inhibitor-bound forms (a) the three-dimensional structures and (b) the cartoon pictures represent the open- and close- gating states.

The results indicate that the conformations of both native and SARS-bound complexes are virtually identical, while completely different in case of the native and MLN-4760-bound frameworks, especially the active site area. As shown in Figure 6, the different point obviously seeing is the changes of the positions of helices, loops and turns nearby the active site. These transitions generally indicate the close-gating state when the inhibitor is bound in the active site (Figure 8a).

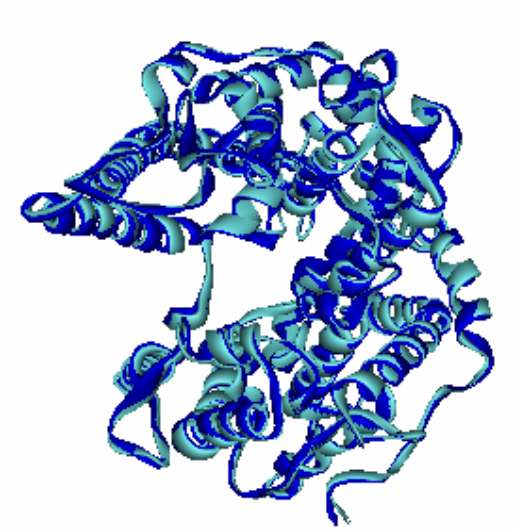


Figure 7 Superimposition of native ACE2 (cyan) and the SARS-bound structure (blue).

Interestingly, both SARS-bound and native conformations are nearly identical (according to very small RMSD). This result completely makes a good agreement with the previous experiment showing that the attachment of SARS spike protein on ACE2 has no effect on the three-dimensional conformation (Li *et al*., 2005). Also, it was found that the SARS spike domain contacted only the tip of one lobe. It does not contact the other lobe, nor does it occlude the peptidase active site (Li *et al.*, 2005). Moreover, in this study, it is shown that the conformation of SARS-contacting region of ACE2 in a MLN-4760 complex is very similar to that of the SARS-bound structure (Figure 8b) although the conformations of the rest are not. This is a good evidence to prove that the viral attachment do not influence the open- and closed- states.

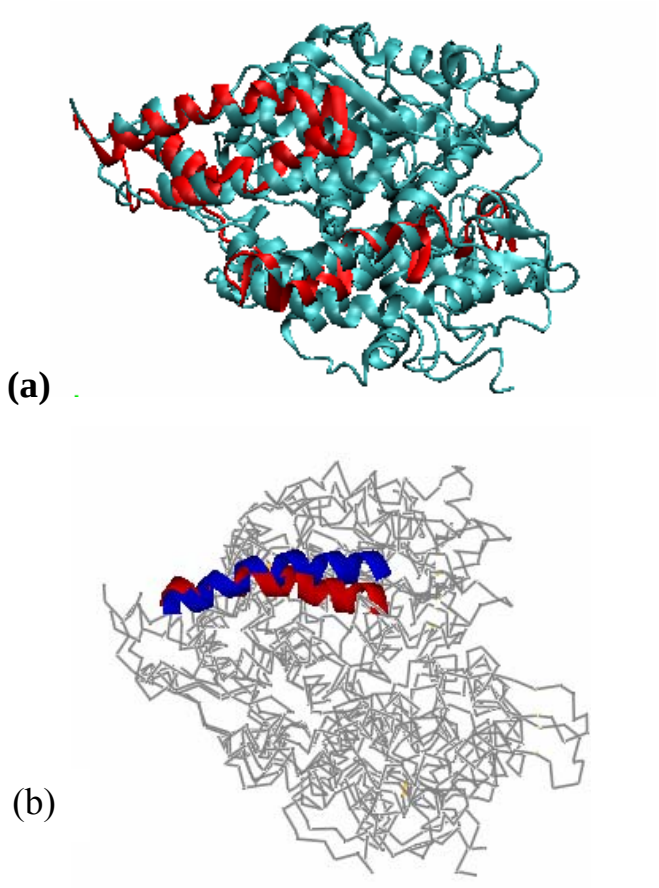


Figure 8 Superimposition of the native ACE2 and the inhibitor-bound (MLN-4760) complex (a) the native is in red, while the MLN-4760 complex is in cyan and (b) the SARS spike-binding domain of ACE2 from the native (blue) and inhibitor-bound (red) structure.

Admittedly, the comparison of the conformations of the amino acids in the active site between the native and MLN-4760-bound complexes is shown in Figure 9. Overall, the orientations of amino acids located in an active site of both apo- and MLN-4760-bound form are quite similar. As a rule, the positions of all residues in the active site are correspondingly moved down (Figure 9; blue) whose positions are a little bit lower than the native one (Figure 9; orange), but there are large conformation changes at position Phe504 and His345. As shown in Figure 10, in the MLN-4760 complex, the aromatic rings of Phe504 and His345 are flipped up in order to align in a parallel manner to each other. It seems that Phe504 and His345 act as the goalkeeper to remain the closed state of ACE2. Both distances of 3.43 Å and 3.49 Å and the

orientation between the aromatic rings of these two residues are possible to form the interaction like $\pi - \pi$ interaction, which may help to bind the inhibitor more efficient (Figure 10).

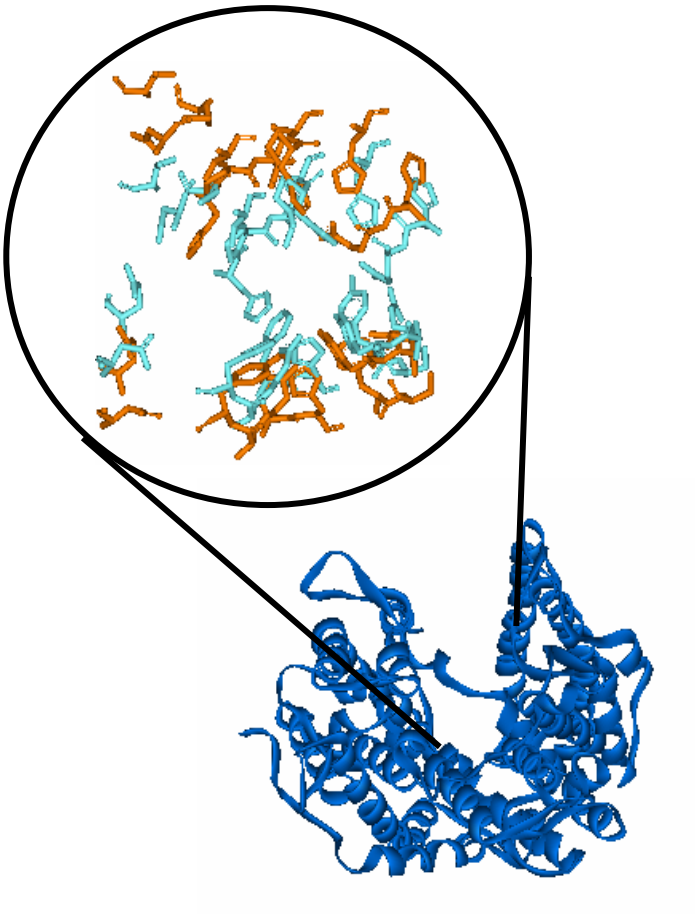


Figure 9 Amino acids aligning in the active site of the MLN-4760-bound complex (the blue color represents the MLN-4760 complex and the red one represent the native).

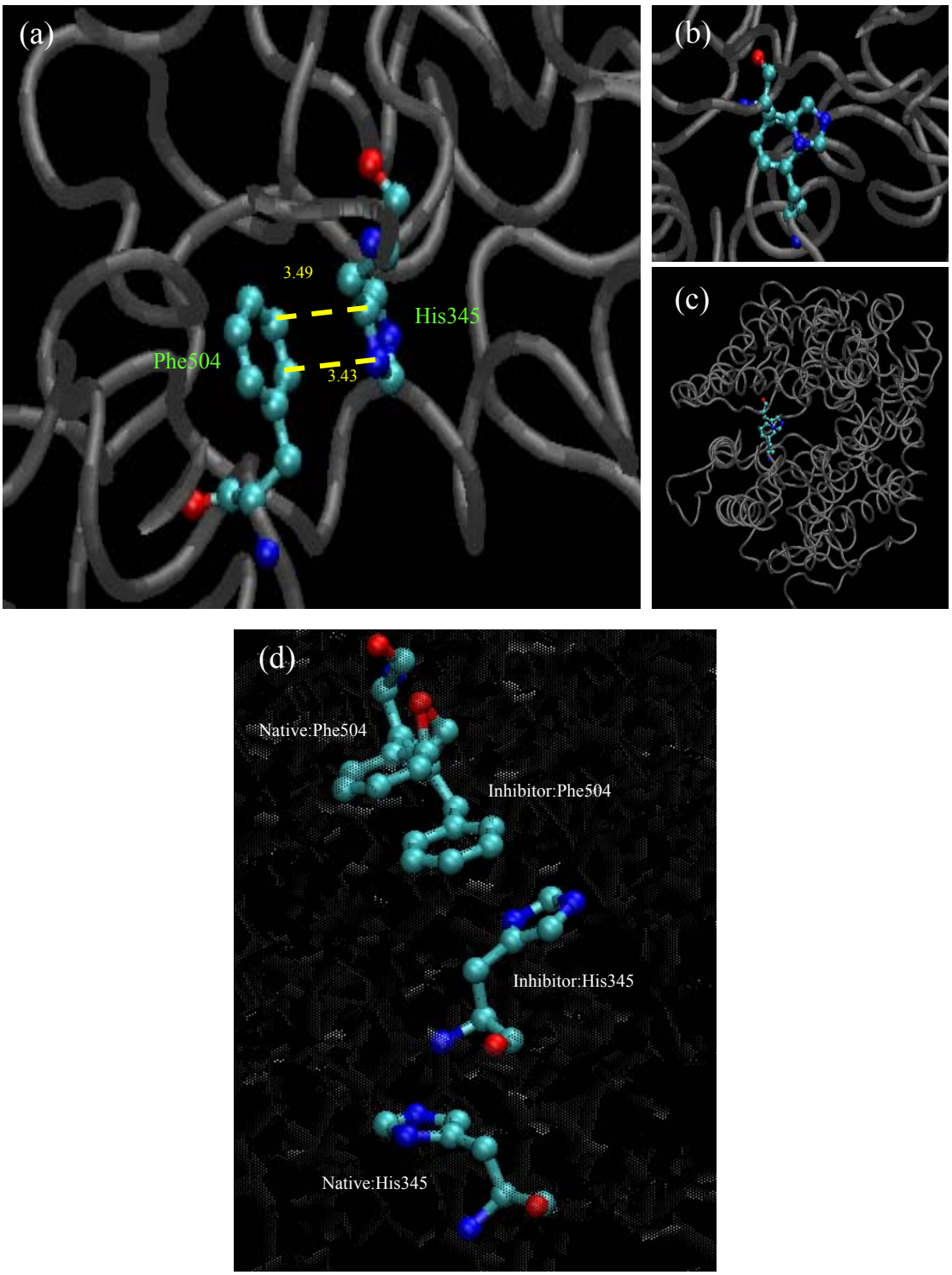


Figure 10 (a, b, and c) the orientation of Phe504 and His345 in the active site and (d) the comparison of positions of Phe504 and His345 in the native and inhibitor-bound complex.

2. Interaction energies of amino acids in the active site

2.1 Zinc-binding effect

All the interaction energies (E_{int}) between the inhibitor, MLN-4760, and its surrounding amino acids are shown in Table 1. The negative energy indicates the attractive interaction, while the positive represents the repulsion. From Table 1, it displays that almost all residues attractively interact with MLN-4760 and only certain residues give repulsive forces (Glu145, Asn149, Asp368, and Tyr510). However, even though almost all residues give the attractive interaction, the overall picture seems that the attractions in this system from each individual provide such a low level, except in the Zinc complex area (His374, Glu375, His378, and Glu402 ~ -0.44, -0.94, -1.47, amd -1.08 kcal/mol, respectively), especially the strongest attraction is at His378. To better understand the effect of Zn^{2+} ion, we also shed the light on the interaction energies of both the system including Zn^{2+} ion (Table 1) and without Zn^{2+} ion (Table 2). In the zinc area, the carboxylate group (COO^-) of MLN-4760 formed the complex with the Zn^{2+} and the charged residues, His374, Glu375, His378, and Glu402 (Figure 12). In the model including zinc ion, it is obvious that the negatively charged Glu372 and Glu402 can certainly form the zinc chelate giving high negative interaction energies, but in case of His374 and His378, which are positively charged residues, they also generate the high attraction forces although they are located nearby the Zn^{2+} ion. It is feasible that both His residues can form the positive complex with zinc ion making themselves so positive that help to increase the attraction in this area. The results show that the zinc ion completely participates and helps for enhancing the level of the attractions between residues and MLN-4760.

Table 1 List of the interaction energies of MLN-4760 with its surrounding amino acids in kcal/mol, calculated by B3LYP/6-31G(d, p) level of calculation.

Amino acid	E_{int}(kcal/mol)
Glu145	0.054
Asn149	0.004
Arg273	-0.269
Phe274	-0.027
His345	-0.244
Pro346	-0.011
Thr347	-0.015
Met360	-0.022
Thr362	-0.025
Asp368	0.067
Thr371	-0.018
His374	-0.435
Glu375	-0.940
His378	-1.468
His401	-0.158
Glu402	-1.078
Phe504	-0.025
His505	-0.258
Tyr510	0.122
Ser511	-0.016
Arg514	-0.184
Tyr515	-0.031

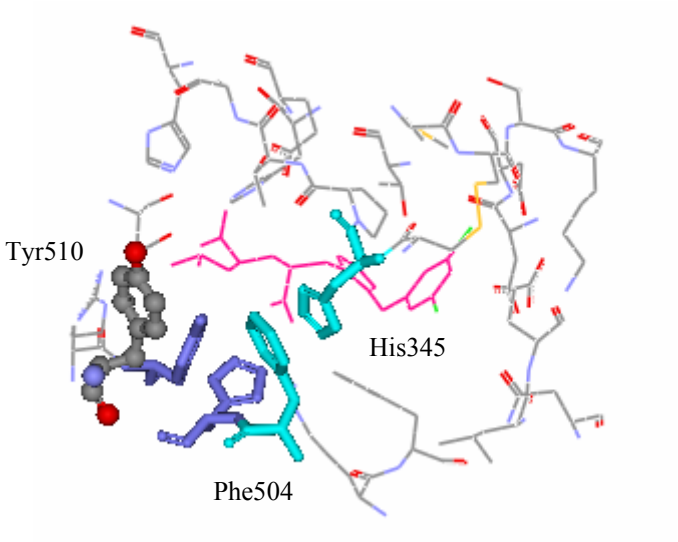


Figure 11 Orientation of Tyr510 in the active site surrounded by aromatic residues, His505 (blue), Tyr515 (blue), Phe504 (cyan), and His345 (cyan) and the MLN-4760 is shown in pink.

Moreover, as shown in Table 2, the interaction energies are rapidly decreased when getting rid of the zinc ion from the system. This phenomenon also supports that the metal ion plays a crucial role in the efficiency of inhibitor binding as well as the metal-bound area seems to be a future main focus of drug design to improve the drug binding affinity.

Table 2 List of the interaction energies of the amino acids nearby the Zn^{2+} area with the inhibitor, MLN-4760 (not including the zinc ion)

Amino acid	E_{int}(kcal/mol)
His374	-0.236
Glu375	0.150
His378	-0.249
Glu402	0.160

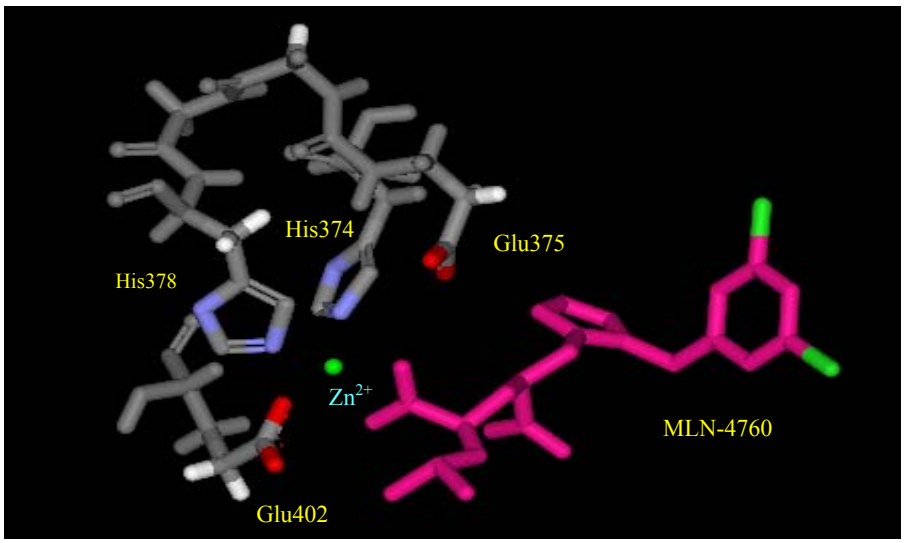


Figure 12 Cluster of amino acids and inhibitor with the zinc ion in the active site.

2.2 Repulsive interactions in the pocket

There are certain repulsive interactions originating from the amino acids surrounding MLN-4760 (Glu145, Asn149, Asp368, and Tyr510). From the structure (Figure 13), they should have been the attractive van der waals interaction between the methyl groups from MLN-4760 and the aromatic ring of Tyr510, but it is the fact that the interaction energy between MLN-4760 and Tyr510 is positive (repulsion) which can be divided into two opposite point of views. Firstly, the hydrophobic group of MLN-4760 at this position is not suitable for this area. It is likely that to get the better binding affinity, we should change the functional group here from methyl groups to others that can give the attraction instead. Secondly, it can be interpreted that even though these methyl groups do not fit and help to enhance the attraction between this inhibitor and Tyr510, in the overall picture, these methyl groups may fit with the whole residues located in this area or probably may help to improve either physical or chemical properties that we cannot recently know. Tyr510 is located nearby Phe504 and His345 as we mentioned before that the location of these two residues in the native and inhibitor-bound are completely different (Figure 10 and 11). Also, from the structural analysis, we suggest that these two residues may help to hold the inhibitor firmly in the active site. Our next suggestion is that Tyr510 may

participate or affect the conformation change of His345 and Phe504 in the transition of open-to-closed state.

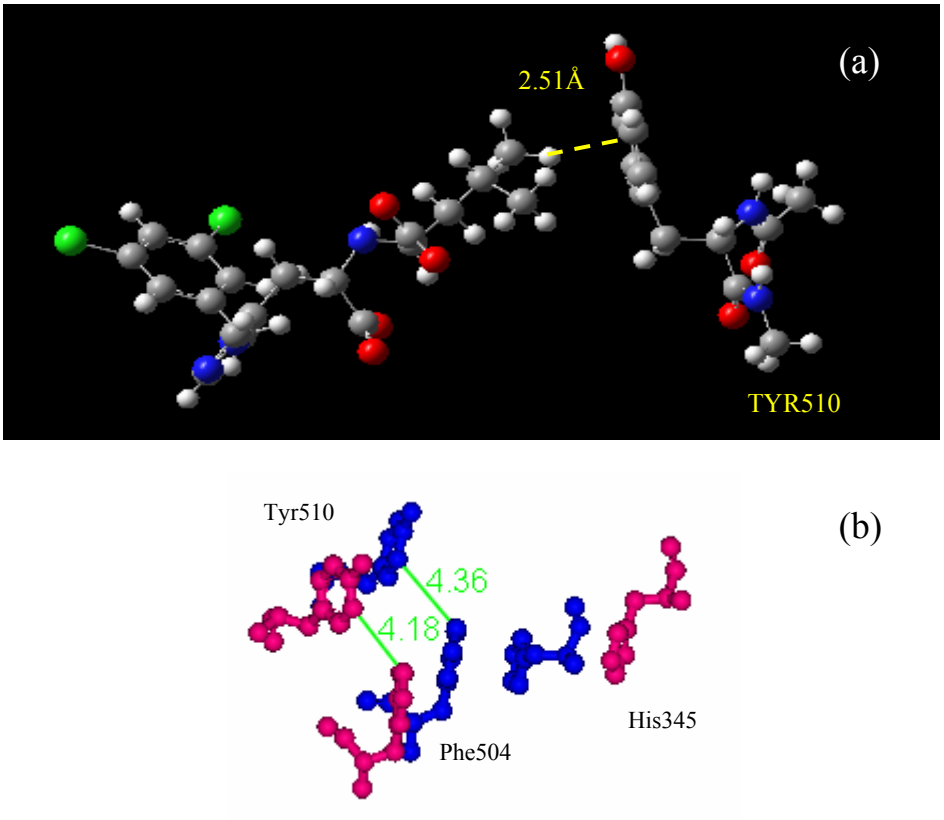


Figure 13 (a) Complex of MLN-4760 and Tyr510 and (b) Cluster of His345, Phe504, and Tyr510 in the native state (pink) and the MLN-4760 complex (blue)

As shown in Figure 13b, the orientation of the aromatic ring of Tyr510 in the native (pink) differs from that of the inhibitor-bound (blue). It seems that the ring of Tyr510 in a MLN-4760 complex does align parallelly with the phenyl ring of Phe504 although the distances between these two aromatic rings are quite far (4.18 Å in native and 4.36 Å in MLN-4760 complex). In summary, we suggest that to improve the inhibitor-binding activity, we should modify the functional groups nearby residue Glu145, Asn149, Asp368, and Tyr510 due to the repulsive interactions. However, this is only the preliminary evidences. We still need other approaches to understand and get the exact answer. The molecular dynamics simulations may help to understand the relationships and behavior between Tyr510 and its neighbors by the dynamics study of this system.

CONCLUSION

According to the previous work, it was found that the SARS receptor-binding domain of ACE2 is only located in the helix nearby the active site. Our structural analysis suggests that this helix is slightly changed its conformation in the inhibitor-bound structure. Moreover, these domains in the native and the SARS-bound are also virtually identical. Therefore, it indicates that the SARS spike domain may be able to attach ACE2 receptor even in the inhibitor- or substrate-bound conformation. However, we still need to understand more about the mechanism of SARS attachment as well as the effect of SARS spike binding on the transition sate of the open-to-closed state.

For the particular interactions of inhibitor, MLN-4760, in the active site, to develop the more potential drugs or inhibitor, we suggest that changing the functional groups of MLN-4760 around residue Glu145, Asn149, and Asp368 may help to improve the inhibitor-binding affinity. In case of Tyr510, even though the inhibitor does not gain such a strong effect from Tyr510, in the inhibitor-bound complex, we found that Tyr510 was unexpectedly aligned in a parallel manner with Phe504 and His345 found in our study that they had changed their conformations in order to bind to MLN-4760 firmly. Hence, we expect that Tyr510 may deal with the conformation changes of Phe504 and His345. So, it is necessary to get more understanding of its real responsibility in the inhibitor or substrate binding and the relationships between this residue with its nearby Phe504 and His345 which may help to understand more about the ACE2 activity.

The study we present here provides the insight into the conformational differences that exist between the apo-, inhibitor-, and SARS-bound structures in the active site, especially the particular interaction energies of the inhibitor, MLN-4760, with its surrounding amino acids that should aid the future design of novel and specific inhibitors of ACE2 activity. This may be useful for exploring the further physiological roles of ACE2 and suggesting the novel therapeutic agents.

LITERATURE CITED

Becke, A. D. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. **Physics Review A**. 38: 3098.

Becke, A. D. 1993. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. **Journal of Chemical Physics**. 98: 5648.

Boehm, M., and E. G. Nabel. 2002. Angiotensin-converting enzyme 2, a new cardiac regulator. **Journal of Medicine**. 347: 1795-1796.

Burrell, L. M., C. Tikellis, and M. E. Cooper. 2004. ACE2, a new regulator of the renin–angiotensin system. **TRENDS in Endocrinology and Metabolism**. 15(4): 166-169.

Capella, K. 2006. A Bird's-eye view of Density-Functional Theory. **Condensed Matter.** 5: 1-69.

Crackwer, M. A., S. Renu, G. Y. Oudit, C. Yagil, I. Kozieradzki, S. E. Scanga, A. J. Oliverira-dos-Santos, J. D. Costa, Y. Pei, J. Scholey, C. M. Ferrario, and A. S. Manoukain. 2002. Angiotensin-converting enzyme 2 is an essential regulator of heart function. **Nature Medicine**. 417: 823-828.

Dales, N. A., A. E. Gould, J. A. Brown, E. F. Calderwood, B. Guan, C. A. Minor, J. M. Gavin, P. Hales, V. K. Kaushik, J. Stewart, T. J. Peter, C. S. Vickers, T. D. Ocain, and M. A. Patane. 2002. Substract-Based Design of the first class of angiotensin-converting enzyme-related carboxypetidase. **Journal of American Chemical Society**. 124: 11852-11853.

Danilczyk, U., and J. Penninger. 2006. Angiotensin-converting enzyme II in the heart and the kidney. **Circulation Research**. 98: 463-471.

Dimitrov, D. S. 2003. The Secret Life of ACE2 as a Receptor for the SARS Virus. **Journal of Virology**. 77: 8801-8811.

Donoghue, M., F. Hsieh, E. Baronas, K. Godbout, M. Gosselin, N. Stagliano, M. Donovan, B. Woolf, K. Robinson, R. Jeyaseelan, R. E. Breitbart, and S. Acton. 2000. ACE2 gene expression is up-regulated in the human failing heart. **Circulation Research**. 87: E1-E9.

Douglas, G. C., M. K. O'bryan, M. P. Hedger, D. K. L. Lee, M. A. Yarski, A. I. Smith, and R. A. Lew. 2004. The Novel Angiotensin-Converting Enzyme (ACE) homolog, ACE2, is selectively expressed by adult leydig cells of the testis. **Endocrinology**. 145(10): 4703-4711.

Filippi, C., C. J. Umrigar, and M. Taut. 2005. Comparison of exact and approximate density functionals for an exactly soluble model. **Journal of Chemical Physics**. 100: 1290.

Guy, J. L., R. M. Jackson, K. R. Acharya, E. D. Sturrock, N. M. Hooper, and A. J. Turner. 2003. Angiotensin-converting enzyme-2 (ACE2): Comparative modelling of the active site, substrate specificity and chloride sensitivity. **Biochemistry** 42. 13185–13192.

Guy, J. L., R. M. Jackson, H. A. Jensen, N. M. Hooper, and A. J. Turner. 2005. Identification of critical active-site residues in angiotensin-converting enzyme-2 (ACE2) by site-directed mutagenesis. **The FEBS journal**. 272: 3512-3520.

Guy, J. L., D. W. Lambert, F. J. Warner, N. M. Hooper, and A. J. Turner. 2005. Membrane-associated zinc peptidase families: comparing ACE and ACE2. **Biochimica et Biophysica Acta**. 1751: 2-8.

Han, D., A. Penn-Nicholson, and M. Cho. 2006. Identification of critical determinats on ACE2 for SARS-CoV entry and developement of a potent entry inhibitor. **Virology**. 350: 15-25.

Huang, L., D. J. Sexton, K. Skogerson, M. Devlin, R. Smith, I. Sanyal, T. Parry, R. Kent, J. Enright, Q. Wu, G. Conley, D. Deolivira, L. Morganelli, M. Ducar, C. Wescott, and R. C. Ladner. 2003. Novel peptide inhibitors of angiotensin-converting enzyme 2. **Journal of Biological Chemistry**. 278(18): 15532-15540.

Huentelman, M. J., J. Zubcevic, J. A. H. Prada, X. Xiao, D. S. Dimitrov, M. K. Raizada, and D. A. Ostrov. 2004. Structure-based discovery of a novel angiotensin-converting enzyme 2 inhibitor. **Hypertension**. 44(6): 903-906.

Kuba, K., Y. Imai, S. Rao, H. Gao, F. Guo, B. Guan, Y. Huan, P. Yang, Y. Zhang, W. Deng, L. Bao, B. Zhang, G. Liu, Z. Wang, M. Chappell, and Y. Liu. 2005. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus induced lung injury. **Nature Medicine**. 11(8): 875-879.

Lee, C., W. Yang, and R. G. Parr. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. **Physics Review B**. 37: 785.

Li, F., W. Li, M. Farzan, and S. C. Harrison. 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. **Science**. 309: 1864 - 1868.

Li, W., M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. **Nature Medicine**. 426: 450-454.

Li, W., C. Zhang, J. Sui, J. H. Kuhn, S. K. Wong, I. C. Huang, K. Xu, N. Vasilieva, A. Murakami, Y. He, W. A. Marasco, Y. Guan, C. Hyeryun, and M. Farzan. 2005. Receptor and viral determinants of SARS-coronavirus adaption to human ACE2. **The EMBO Journal**. 24: 1634-1643.

Natesh, R., S. L. U. Schwager, E. D. Sturrock, and K. R. Acharya. 2003. Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. **Nature Medicine**. 421: 551-554.

Perdew, J. P., K. Burke, and M. Ernzerhof. 1997. Generalized gradient approximation made simple. **Physics Review letters**. 77:3865.

Prabakaran, P., X. Xiao, and D. S. Dimitrov. 2004. A model of the ACE2 structure and function as a SARS-CoV receptor. **Biochemical and Biophysical Research Communications**. 314: 235-241.

Rella, M., C. A. Rushworth, J. L. Guy, A. J. Turner, T. Langer, and R. M. Jackson. 2006. Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. **Journal of Chemical Information and modeling**. 46: 708-716.

Spyoulias, G. A., and P. Cardopatis. 2005. Current inhibition concepts of zinc metallopeptidases involved in blood Pressure regulation. **Current Enzyme Inhibition**. 1: 29-42.

Tikellis, C., C. I. Johnsto, J. M. Forbes, W. C. Burns, L. M. Burrell, and J. Risvanis. 2003. Characterization of renal angiotensin-converting enzyme 2 in diabetic nephropathy. **Hypertension**. 41(3): 392-397.

Towler, P., B. Staker, S. G. Prasad, J. Menon, J. Tang, T. Parsons, D. Ryan, M. Fisher, D. Williams, N. A. Dales, M. A. Patane, and M. W. Pantoliano. 2004. ACE2 X-ray structures reveal a large hinge-bending motion important for

inhibitor binding and catalysis. **Journal of Biological Chemistry**. 279(17): 17996-18007.

Turner, A. J., and N. M. Hooper. 2002. The angiotensin-converting enzyme gene family: genomics and pharmacology. **TRENDS in Pharmacological Science.** 23: 177-183.

Tzakos, A. G., and I. P. Gerothanasis. 2005. Domain-selective ligand-binding modes and atomic level pharmacophore refinement in angiotensin I converting enzyme (ACE) inhibitors. **ChemBioChem**. 6: 1089-1103.

Wang, P., J. Chen, A. Zheng, Y. Nie, W. Shi, W. Wang, G. Wang, H. Luo, H. Liu, L. Tan, X. Song, Z. Wang, X. Yin, X. Qu, X. Wang, T. Qing, M. Ding, and H. Deng. 2004. Longitudinally profiling neutralizing antibody response to SARS. **Biochemical Biophysical Research Communications.** 315: 235-241.

APPENDIX

Appendix: Theoretical Background

Theory of Quantum Chemical Calculations

Quantum mechanical methods for the study of molecules can be divided into two categories: *ab initio* and semiempirical models. *Ab initio* methods refer to quantum chemical methods in which all the integrals are exactly evaluated in the course of a calculation. *Ab initio* methods include Hartree-Fock (HF) or molecular orbital (MO) theory, configuration interaction (CI) theory, perturbation theory (PT), and density functional theory (DFT). *Ab initio* methods that include correlation can have accuracy comparable with experiment in structure and energy predictions. However, a drawback is that *ab initio* calculations are extremely demanding in computer resources, especially for large molecular systems. Semiempirical quantum chemical methods lie between *ab initio* and molecular mechanics (MM). Like MM, they use experimentally derived parameters to strive for accuracy; like *ab initio* methods, they are quantum-mechanical in nature. Semiempirical methods are computationally fast because many of the difficult integrals are neglected. The error introduced is compensated through the use of parameters. Thus, semiempirical procedures can often produce greater accuracy than *ab initio* calculations at a similar level. Table 6 attempts to show advantages and disadvantages of MM, semiempirical and *ab initio* methods.

1. Schrödinger Equation

The quantum chemical methods are based on finding solutions to the Schrödinger equation on molecular orbital theory. According to quantum mechanics postulates, the state of a system is fully described by a wave function that depends on the position of the electrons and nuclei in the system.

$$H\Psi = E\Psi \qquad (1)$$

where H is the Hamiltonian operator which gives the kinetic, T, and potential, V, energies of the system that is;

$$H = T + V \tag{2}$$

and

$$T = -\frac{\hbar^2}{2m}\nabla^2 \tag{3}$$

then

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V \tag{4}$$

Then, rewrite equation (1) is;

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + v\right\}\Psi = E\Psi \tag{5}$$

where the Laplacian operator, ∇^2, is ;

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{6}$$

h is Plank's constant divided by 2π. Ψ is the wave function which characterizes the particle's properties. E is the eigen energy of the particle corresponding to wave function.

2. Born-Oppenheimer approximation

Now consider N particle system H, the Hamiltonian operator, is composed of two parts, the kinetic (T) and potential (V) energy part which describes the electron (in atomic units) as shown in equation (12) so:

$$\mathrm{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \qquad (7)$$

The Hamiltonian can thus be separated into two main parts, which is one operator to describe the electron and the other to describe the nucleus.

$$\mathrm{H} = -\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + \mathrm{H}_{el} \qquad (8)$$

This focus on the electronic Hamiltonian, H_{el}, and try to solve the electronic Schrödinger equation in the field of the fixed nuclei. The nuclear-nuclear repulsion term (the final in equation (7)) appears as a constant in H_{el}. Further assume the wave function $\psi(\vec{r},\vec{R})$ to be a product of an electronic and a nuclear part:

$$\psi(\vec{r},\vec{R}) = \psi_{elec}(\vec{r},\vec{R})\psi_{nucl}(\vec{R}) \qquad (9)$$

The justification for this is that the electrons are much lighter than the nuclei. This is called the Born-Oppenheimer approximation (BO). The parametric $\vec{R}$ dependence of ψ_{elec} arises since the electron distribution depends implicitly on the particular nuclear arrangement for the system under study. The nuclear wave function, ψ_{nucl}, describes the vibrational, rotational and translational motion of the nuclei. From equation (1), (8) and (9) can obtain;

$$\mathrm{H}\psi(\vec{r},\vec{R}) = \mathrm{H}\psi_{elec}(\vec{r},\vec{R})\psi_{nucl}(\vec{R}) = \left(-\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + \mathrm{H}_{el}\right)\psi_{elec}(\vec{r},\vec{R})\psi_{nucl}(\vec{R})$$

$$= \left(-\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 + \mathrm{E}_{el}\right)\psi_{elec}(\vec{r},\vec{R})\psi_{nucl}(\vec{R}) = \mathrm{E}\psi_{elec}(\vec{r},\vec{R})\psi_{nucl}(\vec{R}) = \mathrm{E}\psi \qquad (10)$$

The electronic wavefunction $\psi_{elec}(\vec{r},\vec{R})$ can be divided out from both sides of equation (10), provided that terms in $\nabla^2\psi_{elec}(\vec{r},\vec{R})$ are small, i.e. the electronic wavefunction changes slowly upon small displacements of the nuclear positions. Thus, if we neglect the influence of the nuclear derivative on the electron wave function

(ψ_{el}) (i.e. the nuclei move slowly compared with the electrons) which can separate equation (10) into two equations, an electronic part:

$$\mathrm{H}_{el}\psi_{el}\left(\vec{r},\vec{R}\right)=\mathrm{E}_{el}\left(\vec{R}\right)\psi_{el}\left(\vec{r},\vec{R}\right) \tag{11}$$

where

$$\mathrm{H}_{el}=-\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2-\sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}}+\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}+\sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{12}$$

and a nuclear part:

$$\mathrm{H}_{nucl}\psi_{nucl}\left(\vec{R}\right)=\mathrm{E}\psi_{nucl}\left(\vec{R}\right) \tag{13}$$

where

$$\mathrm{H}_{nucl}=-\sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2+\mathrm{E}_{el}\left(\vec{R}\right) \tag{14}$$

Appendix Table 1 Synopsis of molecular modeling techniques.

Method	Advantages	Disadvantages	Best for
Ab initio .Uses quantum physics .Mathematically rigorous: no empirical parametes	.Useful for a broad range of systems .Does not depend on experimental data .Calculates transition states and excited states	.Computationally expensive	.Small systems .Electronic transitions .Systems without experimental data .Systems requiring high accuracy.
Semi-empirical .Uses quantum physics .Uses experimental parameters .Uses extensive approximations	.Less demanding computationally than *ab initio* methods. .Calculates transition states and excited states.	.Requires *ab initio* or experimental data for parameters. .Less rigorous than *ab initio* methods.	.Medium-sized systems (hundreds of atoms). .Electronic transitions.
Molecular Mechanics .Uses classical physics .Relies on force field with embedded empirical parameters.	.Computationally 'cheap', fast and useful with limited computer resources. .Can be used for large molecules like enzymes.	.No electronic properties .Requires *ab initio* or experimental data for parameters. .Commercial solfware applicable to limited range of molecules.	.Large systems (thousands of atoms) .System or processes that do not involve bond breaking.

3. Hartree Fock Theory

At this point it is important to mention the Born-Oppenheimer approximation, which states that since the mass of the nuclei is so much greater than the mass of the electron we can consider the nuclei to be stationary with respect to the electron and the kinetic energy term for the nuclei can be neglected. This greatly simplifies the expression. For a complete derivation of this please refer to standard textbooks on this subject. Also, we are neglecting relativistic effects and other operators such as spin-spin coupling effects by using the BO approximation.

Schrödinger equation:

$$\mathrm{H}\,\psi(\vec{r},\vec{R}) = \mathrm{E}\,\psi(\vec{r},\vec{R}) \tag{15}$$

An exact solution to the Schrödinger equation is not possible for any but the most trivial molecular systems. However, a number of simplifying assumptions and procedures do make an approximate solution possible for a large range of molecules. To simplify the treatment further, the next step is to assume that the electrons are non-interacting. This implies that (apart from the constant nuclear-nuclear repulsion term) which can rewrite the total n-electron Hamiltonian as a sum of n one-electron Hamiltonians,

$$\mathrm{H}_{el} = \sum_{i=1}^{N} h(i) \tag{16}$$

$$h(i) = \left(-\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} \right) \tag{17}$$

Through the wave functions, the effective potential is generated. This potential allows to refine wave functions, from which a new potential is obtained. The procedure is repeated until a stable, self-consistent solution is reached. Due to the iterative procedure, the initial guess of the wavefunction, can of course be chosen ad hoc. However, the better the initial guess is, the easier it is to reach a stable solution to the eigenvalue problems in a relatively short computational time, is provided by the

variation principle. This can be stated in the following way: Given any approximate wave function, satisfying the correct boundary conditions, the expectation value of the energy obtained by this wave function never lies below the exact energy of the ground state. Expressed in mathematical terms:

$$\mathrm{E}_e = \frac{\langle \psi | \mathrm{H}_e | \psi \rangle}{\langle \psi | \psi \rangle} \geq \mathrm{E}_{exact} \tag{18}$$

A conceptually appealing model for the (trial) wave function of our molecular system is to regard it as being constructed from molecular orbitals (MO). This description in analogous to the model used for the atomic orbitals (AO). The MO's, the elements of the wave function determinant, are in turn thought of as being constructed by a Linear Combination of Atomic Orbitals (LCAO),

$$\psi_i^{MO} = \sum_{\mu} c_{\mu i} \phi_{\mu}^{AO} \tag{19}$$

The variational principle leads to following equations describing the molecular orbital expansion coefficients, $c_{\nu i}$, derived by Roothaan and by Hall:

$$\sum_{\nu=1}^{N} \left(F_{\mu\nu} - \varepsilon_i S_{\mu\nu} \right) c_{\nu i} = 0 \quad \mu = 1,2,..., \mathrm{N} \tag{20}$$

Equation (16) can be rewritten in matrix form:

$$FC = SC\varepsilon \tag{21}$$

with

$$F_{\mu\nu} = H_{\mu\nu}^{core} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2} (\mu\lambda|\nu\sigma) \right] \tag{22}$$

$$F_{\mu\nu} = \mathrm{H}_{\mu\nu}^{core} + G_{\mu\nu} \tag{23}$$

where $\mathrm{H}_{\mu\nu}^{core}$, core-Hamiltonian matrix, defined as

$$\mathrm{H}_{\mu\nu}^{core} = \int dr_1 \phi_\mu^*(1) h(1) \phi_\nu(1) \tag{24}$$

The matrix P is the density matrix or charge- and bond-order matrix,

$$P_{\mu\nu} = 2\sum_a^{N/2} C_{\mu a} C_{\nu a}^* \tag{25}$$

The matrix S is the overlap matrix, indicating the overlap between orbitals.

$$S_{\mu\nu} = \int dr_1 \phi_\mu^*(1) \phi_\nu(1) \tag{26}$$

The term $(\mu\nu|\lambda\sigma)$ in equation 22 signified the two-electron repulsion integrals, defined as

$$(\mu\nu|\lambda\sigma) = \int dr_1 dr_2 \phi_\mu^*(1) \phi_\nu(2) r_{12}^{-1} \phi_\lambda^*(2) \phi_\sigma(2) \tag{27}$$

The (initial) wave function is used to generate an effective potential, which apply this potential in order to refine the coefficient matrix. The modified MO's form the new input in the Roothaan or Pople-Nesbet (1986) equations, and a new potential is generated. The iterative procedure is repeated until convergence is reached, i.e. when the changes in energy and/or charge density in two subsequent iterations are below a pre-set threshold value.

Before a more technical description of the SCF-procedure is presented, first need to define a new transformation matrix X, used for orthogonalisation of the basis

set. This orthogonalisation can be either symmetric or canonical. A symmetric orthogonalisation implies that X is formed through the relation

$$X = S^{-1/2} = Us^{-1/2}U^{\tau} \tag{28}$$

where S is the overlap matrix, U is an unitary matrix which diagonalizes S, and the diagonal matrix of the eigenvalues of S is given by the relations. In the canonical orthogonalisation procedure, X is instead given by

$$X = Us^{-1/2} \tag{29}$$

Consider a new coefficient matrix C' related to the old coefficient matrix C by

$$C' = X^{-1}C, \qquad C = XC' \tag{30}$$

where assumed that X possesses an inverse. Substituting $C = XC'$ into the Roothaan equations gives

$$FXC' = SXC'\varepsilon \tag{31}$$

Multiplying on the left by X^{τ} gives

$$\left(X^{\tau}FX\right)C' = \left(X^{\tau}SX\right)C'\varepsilon \tag{32}$$

if define a new matrix F^{τ} by

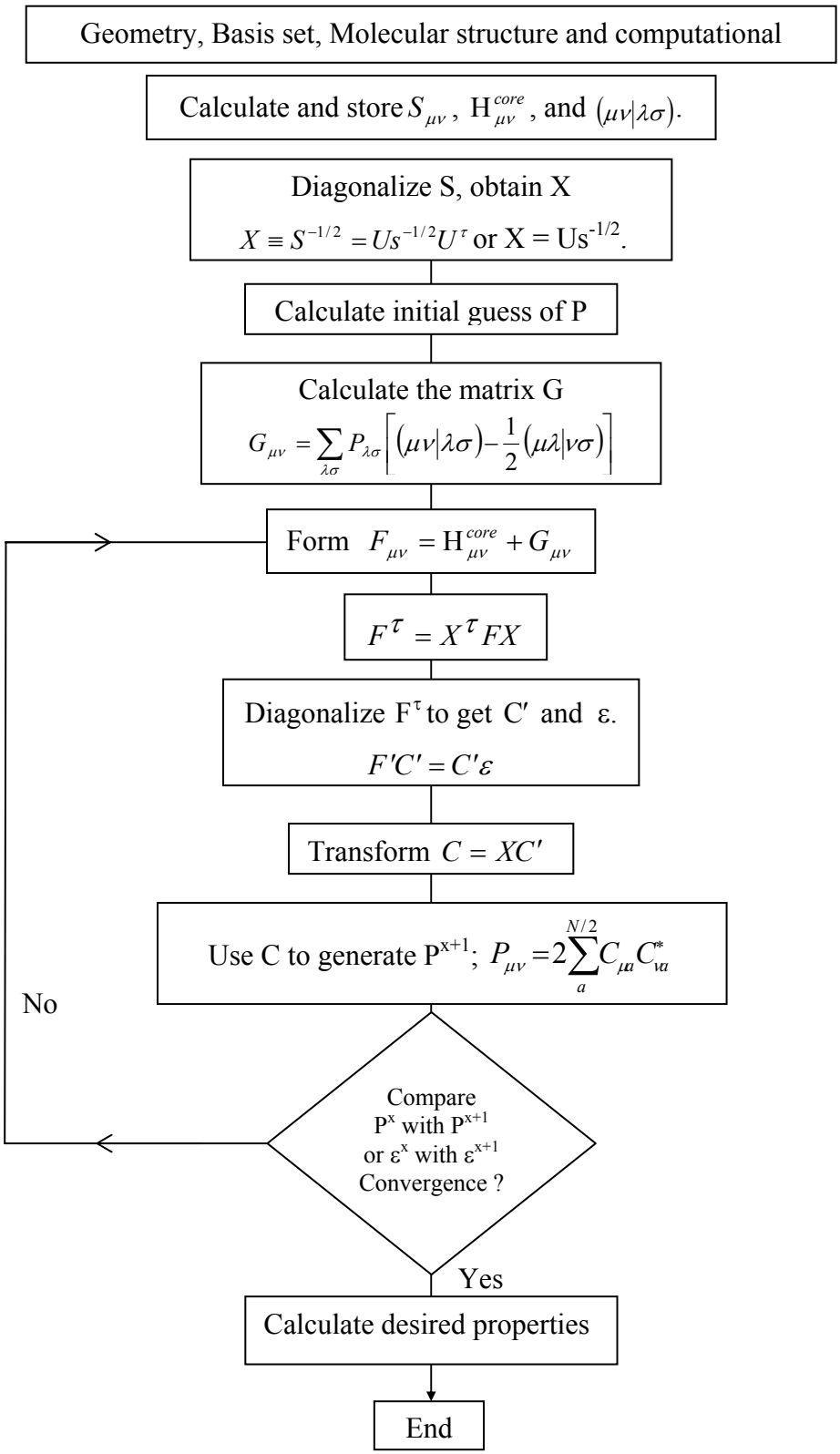
$$F^{\tau} = X^{\tau}FX \tag{33}$$

and use equation (26),then

$$F'C' = C'\varepsilon \tag{34}$$

The SCF procedure, outlined in Appendix Figure 1, is as follows

1. Specify a molecule (a set of nuclear coordinates $\{R_A\}$, atomic numbers $\{Z_A\}$, and number of electron N) and a basis set $\{\phi_\mu\}$.
2. Calculate all required molecular integrals, $S_{\mu\nu}$, $H_{\mu\nu}^{core}$ and $(\mu\nu|\lambda\sigma)$.
3. Diagonalize the overlap matrix S and obtain a transformation matrix X from either equation $X \equiv S^{-1/2} = Us^{-1/2}U^\tau$ or $X = Us^{-1/2}$.
4. Obtain a guess at the density matrix P.
5. Calculate the matrix G of equation $F_{\mu\nu} = H_{\mu\nu}^{core} + G_{\mu\nu}$ from the density matrix P and the two-electron integral $(\mu\nu|\lambda\sigma)$.
6. Add G to the core-Hamiltonian to obtain the Fock matrix $F = H^{core} + G$.
7. Calculate the transformed Fork matrix $F^\tau = X^\tau FX$
8. Diagonalize F^τ to obtain C' and ε.
9. Calculate $C = XC'$.
10. Form a new density matrix P from C using $P_{\mu\nu} = 2\sum_a^{N/2} C_{\mu a} C_{\nu a}^*$.
11. Determine whether the procedure has converged, i.e. determine whether the new density matrix of step (10) is the same as the previous density matrix within a specified criterion. If the procedure has not converged, return to step (5) with the new density matrix.
12. If the procedure has converged, then use the resultant solution, represented by C, P, F, etc., to calculate expectation values and other quantities of interest.



Appendix Figure 1 Schematic view of a Hartree-Fock self consistent field calculation

4. Semi-empirical Calculation

Semi-empirical methods increase the speed of computation by using approximations of *ab initio* techniques (*e.g.,* by limiting choices of molecular orbitals or considering only valence electrons) which have been fitted to experimental data. Until recently, the size of many energetic molecules placed them beyond the scope of *ab initio* calculations, so preliminary theoretical studies were performed using semi-empirical techniques. However, semi-empirical methods have been calibrated to typical organic or biological systems and tend to be inaccurate for problems involving hydrogen-bonding, chemical transitions or nitrated compounds.

Because both time and storage requirements of an *ab initio* Hartree-Fock calculation increase as the fourth power of the number of basis functions, calculations on large molecules even with the smallest basis sets are apt to be prohibitive. In such situations, the NDDO (neglect of diatomic differential overlap) formalism affords practical methods for calculating the electronic structure of large systems. Here only one- and two-centre, two-electron integrals are considered, and the Hartree-Fock matrix, consists only of elements for which basis functions μ and ν are on the same atom, and basis functions λ and σ are on another atom. The individual terms are below (the sum α is over all other atoms).

$$F_{\mu\mu} = \mathrm{H}_{\mu\mu}^{core} + \sum_{\nu} P_{\nu\nu}\left[\langle\mu\mu \mid \nu\nu\rangle - \langle\mu\nu \mid \mu\nu\rangle\right] + \sum_{\delta}\sum_{\lambda}\sum_{\sigma} P_{\lambda\sigma}\langle\mu\mu \mid \nu\nu\rangle$$

$$F_{\mu\nu} = \mathrm{H}_{\mu\nu}^{core} + P_{\mu\nu}\left[3\langle\mu\nu \mid \mu\nu\rangle - \langle\mu\mu \mid \nu\nu\rangle\right] + \sum_{\delta}\sum_{\lambda}\sum_{\sigma} P_{\lambda\sigma}\langle\mu\nu \mid \lambda\sigma\rangle$$

$$F_{\rho\lambda} = \beta_{\rho\lambda} - \frac{1}{2}\sum_{\nu}\sum_{\sigma} P_{\lambda\sigma}\langle\mu\nu \mid \lambda\sigma\rangle \qquad (35)$$

The elimination of three- and four-centre integrals greatly reduces the time and storage requirements for an NDDO calculation (which now increase as the square of the number of atoms) relative to that for a full Hartree-Fock treatment.

Three levels of NDDO theory are included in *SPARTAN'S SEMI EMPIRICAL* module:

- *Modified NDO (MNDO)* a method introduced to correct some of the problems associated with MINDO/3. MNDO does not work well for sterically crowded molecules, four-membered rings, hydrogen bonding, hypervalent compounds, nitro groups and peroxides. In general, MNDO overestimates activation barriers to chemical reactions.

- *Austin Method, version 1 (AM1)* a reparameterised version of MNDO which includes changes in nuclear repulsion terms. Although more accurate than MNDO, AM1 does not handle phosphorus-oxygen bonds, nitro compounds and peroxide bonds.

- *Parameterisation Model, version 3 (PM3)* a second reparameterisation of MNDO, functionally similar to AM1, but with some significant improvements. PM3 is a recently developed semi-empirical method that may contain as yet undiscovered defects.

In all of these formalisms, only the valence electrons are considered. The one electron terms are given by,

$$H_{\mu\nu}^{core} = U_{\mu\nu} - Z_A \sum_{B \neq A} Z_B \langle \mu\nu | \delta\delta \rangle \qquad (36)$$

Here, μ and ν are located on atom A and the summation is over all other atoms. $U_{\nu\nu}$ is related to the binding energy of an electron in atomic orbital ν, and is determined from spectroscopic data. $U_{\nu\mu}$ is set to zero for $\nu \neq \mu$. The second term in equation (36) represents the attraction on an electron on atom A from the nuclear framework. The two center integral involves only the s function on atom B. Z_A is the charge of atom A without its valence electrons.

All one-centre, two-electron integrals $(\nu\nu|\mu\mu)$ and $(\nu\mu|\nu\mu)$ are fitted to spectroscopic data. The two-centre, two-electron repulsion integrals $(\nu\mu|\lambda\sigma)$ are approximated by classical multipole-multipole charge interactions between atoms A

and B. The multipole charge separations within an atom are treated as adjustable parameters, i.e. optimized to fit the experimentally derived one-centre integrals.

The $\beta_{\rho\lambda}$ terms appearing in the Fock matrix (equation (35)) are the one-electron, two-centre core resonance integrals and are approximated as,

$$\beta_{\rho\lambda} = \frac{\beta_\rho + \beta_\lambda}{2} S_{\rho\lambda} \tag{37}$$

where $S_{\rho\lambda}$ is the overlap integral between Slater orbitals ρ and λ, and ρλ and λρ are adjustable parameters optimized using experimental thermo chemical data for simple molecules. Because all of the adjustable parameters are rooted in experimental data, these methods are known as semi-empirical. As in *ab initio* Hartree-Fock calculations, an SCF procedure is used to converge on a density matrix, and the electronic energy.

The three methods differ only in the core-repulsion terms (they also differ in the detailed parameterization). Core repulsion includes nuclear repulsion and non-valence electron-electron repulsion, which are not explicitly considered in the calculation of the electronic energy. In the MNDO model, the core repulsion energy is given by,

$$\mathrm{E}^{CR} = \sum_{A \neq B} \sum_{B \neq A} Z_A Z_B \left\langle \delta(A) \middle| \delta(B) \right\rangle \left(e^{-\sigma_a \mathrm{E}_{AB}} + e^{-\sigma_B \mathrm{E}_{AB}} \right) \tag{38}$$

In AM1 a sum of Gaussians is employed to better represent the core repulsion behaviour at van der Waals distances. PM3 uses a similar core repulsion function, but differs in the parameterisation procedure.

One advantage of methods parameterised using experimental data is their implicit inclusion of electron correlation effects. However, dependence on experimental data means that semi-empirical methods would not be expected to

perform well on unusual types of molecules for which no data are available from which to construct parameters.

4. Density Functional Theory

In case of the DFT, the basic concepts are mentioned in "METHODS OF CALCULATIONS" part. Therefore, to better understand, the advantages and disadvantage of this method are provided in this section.

Instead of focusing on wave functions and orbitals, DFT focuses on the electron density (although it usually employs orbitals to get the density). It includes an approximate treatment of electron correlation and therefore should be more accurate than Hartree-Fock theory. There are actually very many different DFT methods, depending on the particular treatment of correlation or "exchange".

The drawback of DFT is that nobody knows how to take a given DFT computation and improve it. This contrasts with all other *ab initio* methods, where an expert can always tell how to keep improving the results until the electronic Schrödinger equation is solved exactly. The reason anyone uses DFT is that it tends to give very accurate results much more cheaply than some competing methods.

Limitations

Despite the remarkable success of the LDA, its limitations mean that care must be taken in its application. For systems where the density varies slowly, the LDA tends to perform well, and chemical trends are well reproduced. In strongly correlated systems where an independent particle picture breaks down, the LDA is very inaccurate. The transition metal oxides XO (X=Fe, Mn, Ni) are all Mott insulators, but the LDA predicts that they are either semiconductors or metals. The LDA has been applied to high T_C superconductors, but finds several to be metallic, when in reality they are insulating at 0 K.

The LDA finds the wrong ground state for in many simpler cases. For example, the LDA finds the wrong ground state for the titanium atom. The LDA does not account for van der Waals bonding, and gives a very poor description of hydrogen bonding. These phenomena are essential for most of biochemistry: the structure of DNA of depends critically on hydrogen bonding, as do the changes in the structure of most molecules on solvation.

An obvious approach to improving the LDA is to include gradient corrections, by making E_{XC} a functional of the density and its gradient:

$$E_{XC}^{GGA}[n(r)] = \int \varepsilon_{XC}(n(r))n(r)dr + \int F_{XC}[n(r), |\nabla n(r)|]dr \qquad (68)$$

where F_{XC} is a correction chosen to satisfy one or several known limits for E_{XC}.

Clearly, there is no unique recipe for F_{XC}, and several dozen functionals have been proposed in the literature. They do not always represent a systematic improvement over the LDA and results must be carefully compared against experiment. The development of improved functionals is currently a very active area of research and although incremental improvements are likely, it is far from clear whether the research will be successful in providing the substantial increase in accuracy desired.

Summary

The main problem of the Hartree-Fock and DFT methods is the underlying treatment of electron correlation. In the Hartree-Fock method, electron correlations beyond a mean field picture are entirely neglected, whereas in DFT they are included approximately via a functional $E_{XC}[n]$. DFT methods consequently require careful calibration to establish their accuracy (or inaccuracy) on a case by case basis. Both methods provide a relatively inexpensive route to performing computational physics, chemistry and materials science, provided only trends and not highly accurate quantitative predictions are required.

CURRICULUM VITAE

PERSONAL DATA

Name:	Miss Prapasiri Pongprayoon
Date of birth:	October 8, 1982
Place of birth:	Udornthani, Thailand
Nationallity:	Thai

EDUCATION

: YEAR	INSTITUTION	DEGREE
2000-2004	Kasetsart University	B.Sc. (Chemistry)

SCHOLARSHIP: Scholarships from Charoen Pokphand Group of companies (2001-2004)

Scholarships from the Postgraduate on Education and research in Petroleum and Petrochemical Technology (ADB-MUA) (2004-2005)