CHEMICAL MODELING OF THE ACTIVE SITE REACTION MECHANISMS OF SUPEROXIDE DISMUTASE.

by

Krishna Kishore Mandava.

Submitted in partial Fulfillment of the Requirements

for the Degree of

Master of Science

in the

Chemistry

Program

YOUNGSTOWN STATE UNIVERSITY

AUGUST, 2005.

CHEMICAL MODELING OF THE ACTIVE SITE REACTION MECHANISMS

OF SUPEROXIDE DISMUTASE.

Krishna Kishore Mandava.

I hereby release this thesis to the public. I understand this thesis will be housed at the Circulation Desk of the University library and will be available for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

M. Krishna Kishore 06/29/05

Krishna Kishore Mandava Date

Approvals:

Howard Mettee June 29 '05

Dr. Howard Mettee Date
Thesis Advisor

Steven M. Schildcrout 6/29/05

Dr. Steven Schildcrout Date
Committee member

Daryl W. Mincey 6/29/05

Dr. Daryl Mincey Date
Committee member

Peter J. Kasvinsky 2/5/05

Dr. Peter J. Kasvinsky Date
Dean of Graduate Studies

Abstract:

Computational work on thermodynamics of dismutation of superoxide anion radicals ($O_2^{\cdot-}$) to hydrogen peroxide and molecular oxygen was conducted with various theoretical models using a variety of basis sets. The models include semi-empirical, Hartree-Fock, density functional, and electron correlation contained in the Spartan Pro® 2004® program suite. Standard free energies for most levels of theory for the reaction of -395 ± 3 kcal/mol were compared favorably with calorimetric value of -403.56. The catalytic role of Cu^{2+} in the superoxide dismutase enzyme was reflected by observing its effect on reducing the activation energies. Activation energies were computed for various transition state complexes involved in concerted, two step and three step mechanisms catalyzed by Cu^{2+}. The corresponding values of 17.5, 28, and 36 kcal/mol for E_a show an increase with the complexity of the mechanism. A simplified mimic of the active site was optimized by the density functional method and the geometry of it was in accord with that found by X-ray diffraction experiments of the crystal and calculations by the PAW method.

Acknowledgements

I would like to thank Professor Howard Mettee for supervising this project and for his devoted guidance during the period of this project, especially his revisions and corrections of the thesis which made it successful. I would like to thank Professor Steven Schildcrout and Professor Daryl Mincey for being on my thesis committee. Special thanks are also due to the Chemistry Department and faculty for their guidance and financial support to pursue masters. In addition I would like to thank my family and friends Armend Axhemi, Bhargavi Vemulapalli, Severina Dimova, and Tarek Musslimani for their moral support.

Table of Contents

	Page
Title page	i
Signature page	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vi
List of Figures	viii
List of Symbols and Abbreviations	x
I) Chapter 1 Introduction	1
II) Chapter 2 Computational methods	10
A) Introduction	10
B) Semi-empirical methods	11
C) Molecular mechanics	19
D) Hartree-Fock (HF) or *ab initio* model	23
E) Hartree-Fock wavefunctions and basis sets	24
F) Density functional theory	28
G) Geometry optimization	33
H) Transition state optimization	35
I) Thermodynamics and kinetics	36
J) Modeling chemical reactions and electron correlation effects	39
K) Effect of solvation	40

Table of Contents (cont'd)

III) Chapter 3 Results and Discussion 41

A) Thermodynamics 41

B) Building and locating transition states 45

C) Building the transition state with copper (Cu^{2+}) 50

D) Product confirmation 55

E) Reaction energy profile calculation with and without copper 57

F) Modeling the two step and three step mechanisms 62

G) Active site simulation 63

H) Computation times of the dismutation reaction at different levels 66

IV) Chapter 4 Conclusion 70

V) Chapter 5 Future work 74

VI) Bibliography 77

List of Tables

Page

1) Statistical analysis of difference between experimental and calculated heats of formation. (kcal/mol) 14

2) Average errors in bond distances at different levels of semi-empirical methods. 15

3) Heats of formation of dismutation reaction components (kcal or au) . 42

4) Absolute entropies of the dismutation reaction components (cal /K mol). 43

5) ΔH_{Rxn}, ΔS_{Rxn}, and ΔG_{Rxn} of dismutation reaction at various levels of theory (kcal/ K mol). 44

6) ΔH values of reactants ($2O_2^-$ and $2H_3O^+$), and transition state at different levels of theory without copper. (kcal or au). 46

7) Activation energy without copper at various levels of theory (kcal/mol). 47

8) Number of optimization steps required for Dismutation reaction without copper to reach transition state. 48

9) Values of ΔH's of reactants and, transition state with copper computed at different levels. (au) 51

10) Activation energies of dismutation reaction with copper at different levels of theory (kcal). 52

11) Optimization steps required for the dismutation reaction with copper to reach transition state. 53

12) Activation energies with and without copper. 54

13) Computed bond-lengths of O-O in O_2, O-O in H_2O_2, and O-H in H_2O_2 at initial, transition and the final states. 56

14) Calculated bond lengths of O-O in O_2, O-O in H_2O_2, and O-H in H_2O_2 at final states (products) and comparison with experimental values. 57

15) Comparison between bond distances in imidazole molecule in the Optimized structure by PAW, Crystal structure, and Optimized structure by B3LYP/6-31G* in SPARTAN 2004. 66

16) Computation times of reactant molecules (ground state) to reach the transition state (excited state) with and without copper at different levels of theory 68

List of Figures

	Page
1) Two active sites in SOD.	5
2) Minimal dismutation reaction illustrating hexagonal planar initial state, twisted transition state and repulsive final state.	59
3) The effect of copper ion on the minimal dismutation reaction (geometries).	60
4) Energy profile of copper ion catalyzed and uncatalyzed minimal dismutation reaction ($2H^+ + 2O_2^-$).	61
5) Biological active site of SOD.	64
5) Optimized structure of the simplified active site.	65
6) Surface complementarity of the protein streptavidin (purple) and the ligand protein (white) which shows molecular surfaces of biotin bounded to streptavidin.	75

List of Symbols and abbreviations

AM1	Austin model 1, a semi-empirical method
B3LYP	A density functional method due to Lee, Yang, and Parr
DFT	Density functional theory
HF	Hartree-Fock method
LDA	Local density approximation
MNDO	Modified neglect of diatomic overlap
PM3	Parametrized model 3, a semi-empirical method
ΔG^0	Standard free energy change between reactants and products
ΔH^0	Standard enthalpy change between reactants and products
ΔS^0	Standard entropy change between reactants and products
$\Delta G^\dagger$	Free energy change between reactants and transition state
$\Delta H^\dagger$	Enthalpy change between reactants and transition state
$\Delta S^\dagger$	Entropy change between reactants and transition state
R	Universal gas constant
k	Boltzmann constant
h	Planck constant
STO-3G	Slater type orbitals in terms of three gaussians
3-21G	Split valence basis set using three gaussians for inner shell orbitals
6-21G	Split valence basis set using six gaussians for inner shell orbitals

List of Symbols and abbreviations (cont'd)

6-31G*	Split valence basis set with d-orbitals
6-31G**	Split valence basis set 6-31G** with additional p-orbitals
K_e	Equilibrium constant between reactants and products
$K^{\dagger}$	Equilibrium constant between reactants and transition state.
PAW	Projector augmented-wave method (density functional method).

Chapter 1 - Introduction

The use of modern molecular orbital theories to model actual chemical reactions has become increasingly possible due principally to advances in the sophistication and capacity of today's software and desktop computers.[1] The ability of theory to compute both thermodynamic ground state properties of reactants and products, as well as transition state energetics (though with less reliability), is in many cases approaching the level of experimental error of real bench experiments. While solid state and extended network systems are less easily treated, even enzymatic active sites can in many cases be portrayed with clarity and understanding. The typical open question is, what level of theory and with what basis set can the reaction be modeled successfully?

Superoxide dismutase (SOD) is one of the most important enzymes in the front line of defense against oxidative stress in mammalian tissues, blood, plants, algae and aerobic bacteria.[2,3] It catalyses the decomposition of the toxic superoxide ion radical, O_2^-, which is a product of numerous metabolic oxidative pathways.

$$2\,O_2^- + 2\,H^+ \xrightarrow{\text{SOD}} O_2 + H_2O_2 \qquad [1]$$

Catalase then continues to remove the still toxic hydrogen peroxide in another reaction.

$$2\,H_2O_2 \xrightarrow{\textit{catalase}} 2\,H_2O + O_2 \qquad [2]$$

It has been found that SOD inhibits the peroxidation of fatty acids in inflammation and hemostasis, is an anti-inflammatory in animal cells, and is involved in anti aging and anti cancer processes.[4] Although reaction [1] is the simplest representation of the conversion of superoxide, it is probable that the proton is hydrated as in most biological environments. Thus it is more realistic to consider the following in water solutions.

$$2\ O_2^- + 2\ H_3O^+ \xrightarrow{SOD} O_2 + H_2O_2 + 2\ H_2O \qquad [3]$$

There has been little published effort to model the thermodynamic and kinetic aspects of the dismutation of superoxide anion, both as an isolated chemical process, and then in an environment that mimics the active site of the dismutase enzyme. The present work is an attempt to investigate the adequacy of computational modeling theory in predicting known properties of this reaction, and to map out a reaction energy profile leading from reactants to products in an environment that resembles the SOD active site.

1 Modeling Introduction.

This project is presented as a series of molecular orbital calculations which include molecular mechanics, semiempirical, Hartree-Fock, density functional and electron correlation methods as incorporated in the Spartan Pro® 2004® suites of programs available from Wavefunction, Inc. The computers used were dell OPTIPLEX GX-270 Pentium-4 3.00Ghz. These methods have their origins in the well known Schrödinger equation, which predicts the energy in a system of attractive and repulsive interactions such as found in typical molecules, but the mathematical details of the Hamiltonian operators and representative eigenfunctions need not be proposed by the

program user, as they are contained in the modeling software. It is normally sufficient to graphically describe the molecules or radicals/ions using ordinary bonding theories and then specify a method and a type of orbital approximation in order to develop energy, enthalpy and entropy characteristics of a molecule (or transition state), not to mention electron polarization, dipole moments, orbital symmetries, and numerous other useful properties.

Molecular mechanics assembles atoms like springs(with characteristic force constants) acting as bonds between them in their familiar geometry, based on the previously known and accepted values for bond lengths, angles and strengths. This is perhaps the most primitive form of theory since there are no provisions to account for individual atomic nuclei and electrons, only their collective behavior as atoms. In contrast, the semiempirical and Hartree-Fock (HF) methods consider the individual atomic particles set out as a geometrically accurate, frozen nuclear (or "core") skeleton followed by the addition of electrons into molecular orbitals (MOs) made up of atomic orbitals on each nuclear center. The density functional method does not use MOs in the usual sense to describe the electron distribution, but it uses instead a density gradient formed around nuclear framework using the shapes of the HF orbitals as a guide to produce the "density functional." Finally the electron correlation methods, Moller-Plesset or coupled clusters, invoke the use of excited states (electrons promoted to higher energy orbitals) to combine with ground state configurations in order to represent the total molecule. This is an effort to reduce the inaccuracies of neglecting electron-electron repulsions, which are ignored in the simpler HF methods, making these energies too large.

Regardless of the method of investigation, the modeling procedure is the same. First the molecule is constructed atom by atom, the positions being prompted by the program and the modeler's knowledge of bonding. Then a selection of the method and the basis set is made, along with the properties desired, and the calculation is started. When convergence of the energy is reached in an energy minimization, iterative process by subtle adjustments of bond distances and angles, the other properties can then be calculated (vibrational frequencies for example). Finally the results must be critically analyzed, not only to understand the properties themselves but also to see that the calculation has been performed correctly.

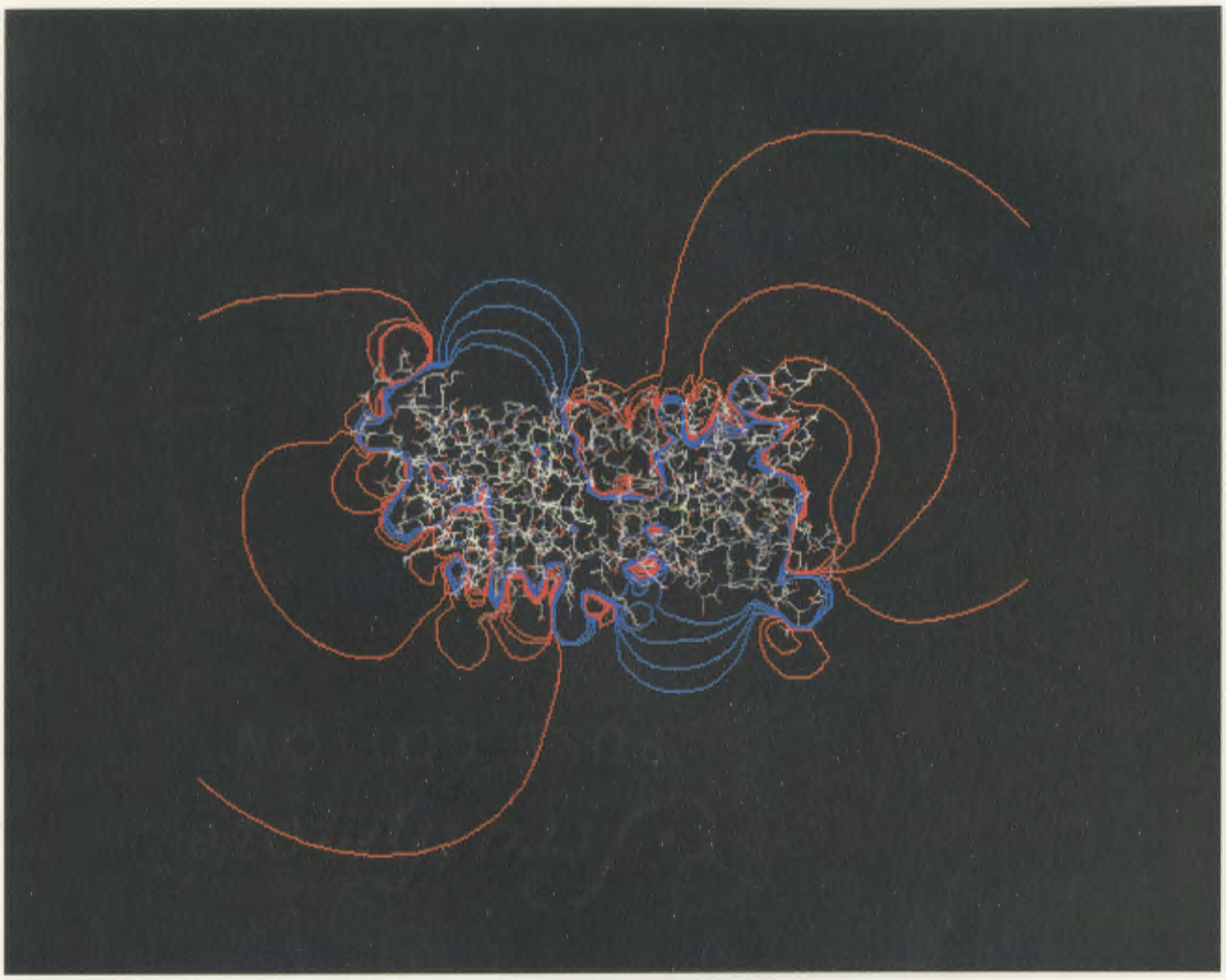
2 Enzyme Active Site.

Copper-zinc *superoxide dismutase* is a dimeric enzyme containing copper and zinc ions in each subunit, and there are two active sites in each subunit. The copper ion is coordinated by four histidine residues in a distorted square planar geometry. One residue acts as a bridge between the copper and the zinc ions. An aspartate and two other histidine residues complete the coordination of the zinc ion. The catalytic copper ion is located at the bottom of a shallow protein cavity, which has numerous charged residues on its surface which are believed to play a role in steering the likewise charged superoxide(s) to the copper ion. The zinc ion is completely buried in the protein.

Figure 1 illustrates the electrostatic potential around the Cu-Zn superoxide dismutase according to McRee et al. Red contours denote negative electrostatic potential and blue contours indicate positive electrostatic potentials. Two active sites are located in each dimer, one in each lobe. They are located at the top left and lower right, where there is a significant concentration of positive electrostatic potential, perhaps related to the

steering effects required to attract and navigate the superoxide anion to the vicinity of Cu^{2+}.

Figure 1. Two active sites in SOD[19].



3 Mechanisms.

Two mechanisms have been proposed in the literature for the dismutation reaction. In the most widely accepted mechanism[7], copper ion is repeatedly oxidized and reduced in a two step process. In the first step, the superoxide ion itself is oxidized by

donating its electron to copper, immediately forming molecular oxygen [4]. In the second step, a second superoxide is reduced by simultaneously accepting an electron, from the now reducing copper, and two protons to form hydrogen peroxide [5]. (The source of these protons could be hydronium ions.)

$$Cu^{2+} + O_2^{\cdot-} \rightarrow Cu^{+} + O_2 \quad [4]$$

$$Cu^{+} + O_2^{\cdot-} + 2\,H^{+} \rightarrow Cu^{2+} + H_2O_2 \quad [5]$$

Within this scheme, it has been suggested that the bond between copper and the His61 (the bridging histidine to zinc, HI^{+}) breaks down losing a H^{+} in the first step, forming a highly basic imidazolate group (**I**), which would in turn take a proton from either water or H_3O^{+}. In the second step, the H^{+} is transferred from the His61 to the second superoxide molecule, forming an intermediate peroxide radical, $HO_2^{\cdot}$ [7].

$$Zn\text{--}HI^{+}\text{--}Cu^{2+} + O_2^{\cdot-} \rightarrow Zn\text{--}I + Cu^{+} + O_2 + H^{+} \quad [6]$$

$$H^{+} + O_2^{\cdot-} \rightarrow HO_2^{\cdot} \quad [7]$$

$$Cu^{+} + HO_2 + H^{+} \rightarrow Cu^{2+} + H_2O_2 \quad [8]$$

$$Cu^{2+} + Zn\text{--}I + H_3O^{+} \rightarrow Zn\text{--}HI^{+}\text{--}Cu^{2+} + H_2O \quad [9]$$

In the second step as well, the His61 must pick up another H^{+} from either water or H_3O^{+} in order to re-establish the bridge between copper and zinc and to be prepared to lose this proton again in subsequent first steps[9].

On the basis of quantum mechanical calculations, Osman, Bash and Rosi [9,10] proposed another mechanism. They suggested the formation of a stable copper-superoxide intermediate complex that is able to oxidize a second superoxide to form O_2. In this case, no breaking of the copper-His61 bond occurs.

$$Cu^{2+} + O_2^{\cdot-} \rightarrow (CuO_2)^{+} \quad [10]$$

$$(CuO_2)^{+} + O_2^{\cdot-} \rightarrow (CuO_2) + O_2 \quad [11]$$

$$(CuO_2) + 2\,H^{+} \rightarrow Cu^{2+} + H_2O_2 \quad [12]$$

The source of the two H^+ in the last step of the second mechanism generating hydrogen peroxide is not specified.

We chose to model at first a primitive transition state using all the principal reactants including Cu^{2+}, 2 H^+ and 2 $O_2^{\cdot-}$, using Cu^{2+} as an electron transfer agent to oxidize one superoxide and reduce the other with the same electron [13]. Then, we attempted to define minimum energy pathways in all three mechanistic schemes to help determine the most reasonable mechanism.

$$Cu^{2+} + 2\,O_2^{\cdot-} + 2\,H^{+} \rightarrow Cu^{2+} + O_2 + H_2O_2 \quad [13]$$

4 Thermodynamics.

As a criterion of the ability of theory to predict the experimental values, we chose to evaluate the thermodynamic parameters of the dismutation reaction [14], where the

source of H^+ was H_3O^+, since the energy of the bare H^+ is difficult to calculate without an electron.

$$2\ H_3O^+ \ + \ 2\ O_2^{\cdot-} \ \rightarrow \ O_2 \ + \ H_2O_2 \ + \ 2\ H_2O \qquad [14]$$

Gas phase values for the enthalpies of formation and standard entropies at 298 K are available for all these from the JANAF-NIST Thermochemical Tables, [22] published in 1998. Whether or not a given theoretical model was successful could in the first instance be checked by comparing its predictions with these experimental quantities.

	H_3O^+	$O_2^{\cdot-}$	O_2	H_2O_2	H_2O
ΔH_f° (kJ/mol):	581.16	-48.593	0.00	-136.106	-241.826
S° (J/K mol):	192.257	209.591	205.147	232.991	188.834

Using the definition $\Delta G = \Delta H - T\Delta S$ for the reaction, and converting to cal/kcal, we have

ΔH_{rxn} = -402.70 kcal/mol ΔS_{rxn} = 0.00289 kcal/K mol ΔG_{rxn} = - 403.56 kcal/mol

It's clear that this reaction has a large thermodynamic driving force, mainly derived from the electrostatic attractions of the oppositely charged ions and the subsequent rebonding of the protons from H_3O^+ to form H_2O_2.

On a more technical note, the Hartree-Fock and higher level treatments gave their energies for each substance in atomic units (au), not the more familiar joules or calories.

Since one au is equivalent to 627.5 kcal, and the standard states for the higher level calculations were the isolated nuclei and electrons, no direct energy comparison could be made between theory and experiment, compound to compound. However, when the energy differences for the reaction were computed in au, the corresponding values in calories or joules could easily be made using the conversion factor above.

Chapter 2 Computational methods:

A) Introduction:

The chemical reaction that forms the basis for the present model calculations includes the superoxide and acid reactants and oxygen, peroxide and water products as thermodynamic ground states.

$$2O_2\cdot^- + 2H_3O^+ \rightarrow H_2O_2 + O_2 + 2H_2O$$

In so far as the thermodynamic driving force for the reaction is concerned, no reference to a transition state is required. The general approach is to compute the enthalpies (ΔH's) and absolute entropies (S's) for the individual molecules by specifying the type of calculation (SCF, semi-empirical, etc) and the basis set of orbitals if required and then to calculate ΔH, ΔS, and ΔG for the reaction. The program adjusts the molecular geometry so as to minimize the energy, at which point the thermodynamic parameters (ΔH, ΔS) are generated from the moments of inertia, normal vibrational modes and coulombic effects of electrons surrounding a frozen nuclear skeleton.

These thermodynamic parameters are independent of any mechanistic pathway connecting reactants and products. The project is done using several model versions to calculate thermodynamic and kinetics using semi-empirical, Hartree-Fock, density functional, etc models. With the limitations of protons in modeling (not comparable with experimental values) we included two hydroniums (H_3O^+) in place of the protons, so the representation of the reaction is altered slightly. Moreover protons in biological system don't freely exist and their presence is nearly always found in water or hydronium ions, though histidine can be a proton donor as well.

The kinetic approach used here was a concerted one where the reaction is visualized as occurring in one step. This is computationally convenient since the proposed two step and three step mechanisms involve additional composite groups including zinc ion and protein. Our active site however, does include copper ion which plays the most important redox catalytic role. We also compare the kinetics of both the proposed mechanisms.

The energy and configuration of the transition states, with and without copper were identified in each case by including the four (or five) principal species, $2O_2^-$, $2H_3O^+$ (with or without Cu^{2+}) arranged in such a geometry which minimizes the maximum energy pathway between the initial and final states. In this manner the electron transfer role of Cu^{2+} was incorporated without the added complexity of a particular active site transition state in the proposed mechanisms. These were examined later. We did not calculate rate constants but only activation energies based on comparing the energy differences between the transition and initial states.

B) <u>Semi-empirical methods:</u>

Austin Model One (AM1) [11] and Parameterized Model Three (PM3) [12] are the most commonly used semi-empirical methods. Semi-empirical models follow in a straightforward way from *ab initio* molecular orbital theories. These are based on the generalization that considers only the valence electrons in the calculations whereas the inner shell electrons are considered as a part of a fixed core. Only a minimal set of valence orbitals are considered on each atom [e.g. 1s (core) + 2s, $2p_x$, $2p_y$, $2p_z$ (valence) on carbon).These are a part of quantum chemical models which are derived from the Schrödinger equation, where all the molecules are treated as a collection of nuclei (or

cores) and electrons with no consideration for chemical bond. The key assumption is made in the semi-empirical methods that the atomic orbitals residing on different atoms do not overlap which is considered as the central approximation.

The semi-empirical method is based on the parameterization of atomic orbital constants. These are the adjustable parameters which are chosen to mimic equilibrium geometries, enthalpies of formation, ionization potentials and dipole moments of well known compounds. The choice of parameters, more than anything else, appears to be the key to formulating successful semi-empirical methods. These parameters (a, b) are multipliers and exponents of the Gaussian functions (ae^{-br^2}) chosen to represent the atomic orbitals. Semi-empirical methods lead directly to a reduction in computation effort from the fourth power of the total number of basis functions (in *ab initio* models) to the square of the number of basis functions. The key approximation, termed as **NDDO** (neglect of diatomic differential overlap) **approximation**, is made which is very severe in that it completely eliminates overlap of atomic basis functions on different, even adjacent atoms thus ignoring bonding effects. To speed up the computation the innermost atomic shells are kept frozen. The frozen Core Orbitals (CO), which are solutions of a large-basis, all-electron calculation on the isolated atom, are expressed in an auxiliary set of (Slater-type) basis functions. The core basis set and the expansion coefficients that give the COs expressed in them are stored in the database data files. AM1 and PM3 are extensions to the MNDO method, which signifies Modified Neglect of Diatomic Overlap. MNDO does not neglect the overlap of orbitals in adjacent atoms, and thus contains a measure of bonding interactions.

The earlier AM1 is parameterized for H, C, N and O, the so called organic

elements. PM3 has the advantage over AM1 in that it is parameterized for more atoms (i.e., 3rd row nonmetals plus aluminum and halogens) than is AM1. It thus allows a wider range of molecules in the computations. AM1 and PM3 are parameterized in large part to yield the energies of molecules in the form of standard enthalpies of formation (where the enthalpies of the elements are set to zero at 298K and 1 bar). This stands in contrast to the Hartree-Fock models, which yield total energy E_{tot} at 0K (where the zero-energy state is a frozen nuclear framework and the separated electrons).

There are several limitations of AM1 which are noted as follows. AM1 predicts hydrogen bond strengths approximately correctly but geometry is often wrong which is occasionally bifurcated. The results of activation energies are much improved over MNDO but there are still significant errors in case of hypervalent (e.g. pentabonded carbon) molecules. Nitro groups are too unstable and peroxide bonds are too short. Heats of formation may be unreliable for molecules with a large amount of charge localized on one atom. The limitations of PM3 are as follows. Barriers for rotation about peptide bonds are too low and proton affinities are not accurate. Hydrogen bonds are too short by $0.1A^0$ and non-bonded distances are often too short. Table 1 summarizes the differences between AM1, PM3 and MNDO with respect to the heats of formation of many reference compounds.

Table 1 Statistical analysis of difference between experimental and calculated heats of formation.[21](kcal/mol)

Statistical analysis of difference between experimental and calculated heats of formation.

Type of compounds	No. of compounds	Averages (unsigned)			Average (signed)			Root mean square		
		PM3	MNDO	AM1	PM3	MNDO	AM1	PM3	MNDO	AM1
Hydrogen	485	6.3	12.9	11.2	1.5	4.8	5.0	8.6	24.1	23.5
Carbon	463	6.2	12.8	11.1	1.1	3.4	3.9	8.8	23.1	22.1
Nitrogen	118	6.9	18.8	9.9	0.2	10.1	4.1	9.4	26.9	12.9
Oxygen	256	9.7	31.7	33.5	0.1	25.0	28.8	13.9	54.7	61.2
Fluorine	148	6.9	41.9	46.6	1.5	34.1	38.4	12.0	80.3	79.7
Aluminum	46	14.7	29.9	38.6	2.6	6.7	33.2	21.2	34.2	53.3
Silicon	78	10.1	22.4	14.5	1.9	13.0	3.6	14.2	32.9	20.8
Phosphorus	71	12.5	37.3	53.6	−0.5	23.8	44.8	16.1	55.1	83.0
Sulfur	101	12.0	50.3	53.5	−1.4	36.9	41.5	16.2	79.8	81.7
Chlorine	105	9.6	23.0	22.3	0.2	12.8	10.6	13.3	51.1	41.4
Bromine	70	11.5	26.8	27.1	1.1	17.1	14.9	15.7	47.1	44.7
Iodine	77	10.7	30.3	27.3	0.2	12.9	11.8	15.9	61.1	54.0
Set of Compounds used in Refs. 3 and 16	138	4.4	6.2	5.5	0.0	−1.4	0.7	6.3	9.1	7.3
Compounds of C, H, N, O, only	276	5.7	11.2	7.5	0.4	3.8	1.4	7.9	18.5	10.5
Nitro compounds	29	5.2	39.6	15.7	2.5	36.1	14.5	6.2	44.1	18.5
Organophosphorus-V compounds	15	10.9	53.9	75.6	3.6	50.2	75.6	14.3	58.7	80.1
Normal valent compounds	657	7.8	13.9	12.7	0.7	3.3	3.7	11.4	25.1	24.3
Hypervalent	106	13.6	75.8	63.1	−0.8	67.2	74.7	17.3	104.5	110.0
All compounds	763	8.6	22.5	22.4	0.5	12.1	13.6	12.4	45.5	46.8
All compounds except Al, P, and S	547	7.1	15.5	11.5	0.9	6.8	2.9	10.1	35.1	26.5

In general, PM3 shows smaller errors over the wide range of compounds and even AM1 is not as accurate for the organic atoms for which it is parameterized. Table 2 compares the same theories with respect to the average errors in bond distances in many of the same compounds.

Table 2 Average errors in bond distances at different levels of semi-empirical methods.[21]

Table 3.2 Average errors in bond distances (Å)

Bonds to:	MNDO	AM1	PM3
H	0.015	0.006	0.005
C	0.002	0.002	0.002
N	0.015	0.014	0.012
O	0.017	0.011	0.006
F	0.023	0.017	0.011
Si	0.030	0.019	0.045

Clearly the smaller difference noted for PM3 show its superiority over this range of compounds.

AM1 was developed because MNDO method failed to reproduce steric interference and hydrogen bonding. A common cause of errors in MNDO is the tendency to overestimate repulsions between neighboring atoms. AM1 deals with this repulsion by adding additional gaussian functions, in the form of a core repulsion function (CRF), for any pair of atoms AB in the molecule.

$$CRF\ (AB) = z_A z_B \gamma_{ss} [1 + F\ (A) + F\ (B)]$$

$$\text{where } F\ (A) = \exp\ (-\alpha_A R_{AB}) + \sum_I K_{Ai} \exp\ [L_{Ai} (R_{AB} - M_{Ai})^2\ ]$$

$$\text{and } F(\ B\ ) = \exp(\ -\alpha_B R_{AB}\ ) + \sum_j K_{Bj} \exp[L_{Bj}\ (\ R_{AB} - M_{Bj}\ )^2]$$

where α, L, K and M are further adjustable parameters. Two ways were used to reduce the excessive interatomic repulsions at large separation. In the first, one or more attractive gaussians were added to compensate directly. In the second, repulsive gaussian functions were centered at smaller internuclear separations.

The MNDO method, which is the basis for AM1 and PM3, is itself based on Hartree-Fock theory. The treatment with MNDO is confined to the valence electrons of closed shell molecules. These electrons are assumed to move in the field of the fixed core potentials of the nuclei and of the inner shell electrons (core repulsion). The valence shell MO's, Ψ_i are represented by linear combination of a minimum basis set of valence shell AO's, Φ_ν and Φ_μ, on the same atom:

$$\Psi_i = \sum_\nu C_{\nu i}\, \Phi_\nu$$

The coefficients $C_{\nu i}$ are found from the Roothan-Hall equations (the variation method) which assume the form:

$$\sum_\nu (F_{\mu\nu} - E_i\, \delta_{\mu\nu)} = 0$$

where E_i is the eigenvalue of the MO Ψ_i and $\delta_{\mu\nu}$ is the Kronecker δ. The elements $F_{\mu\nu}$ of the Fock matrix are the sum of a one electron part $H_{\mu\nu}$ (core Hamiltonian), and a two-electron part $G_{\mu\nu}$;

$$F_{\mu\nu} = H_{\mu\nu} + G_{\mu\nu}$$

and then the electronic energy E_{el} is given by;

$$E_{el} = \tfrac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu})$$

where $P_{\mu\nu}$ is an element of the bond order matrix.

It is assumed that AO's Φ_μ and Φ_ν are centered at atom A and the AO's Φ_λ and Φ_σ are at atom B ($A \neq B$). The Fock matrix elements then are:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_B V_{\mu\mu, B} + \sum_{\nu, A} P_{\nu\nu} [(\mu\mu, \nu\nu) - \tfrac{1}{2} (\mu\nu, \mu\nu)] + \sum_B \sum_{\lambda, \sigma} P_{\lambda\sigma} (\mu\mu, \lambda\sigma)$$

$$F_{\mu\nu} = \sum_B V_{\mu\nu, B} + \tfrac{1}{2} P_{\mu\nu} [3(\mu\nu, \mu\nu) - (\mu\mu, \nu\nu)] + \sum_B \sum_{\lambda, \sigma} (\mu\nu, \lambda\sigma)$$

$$F_{\mu\lambda} = \beta_{\mu\lambda} - \tfrac{1}{2} \sum_{\nu, A} \sum_{\sigma, B} P_{\nu\sigma} (\mu\nu, \lambda\sigma)$$

The Fock matrix has the following six terms:

1) One-center one-electron energies $U_{\mu\nu}$ which represent the sum of the kinetic energy of an electron in AO Φ_μ at atom A and its potential energy due to the attraction of the core of atom A. $U_{\mu\nu}$ assumes the following form:

$$U_{\mu\nu} = (\partial^2 \Phi_\mu / \partial \tau_1) - (e^2 / r_1) \Phi_\mu \, d\tau$$

2) One-center two-electron repulsion integrals, i.e., Coulomb integrals.

$$(\mu\mu, \nu\nu) = g_{\mu\nu}$$

3) Exchange integrals on atom A.

$$(\mu\nu, \mu\nu) = h_{\mu\nu}$$

4) Two-center one-electron core resonance integrals $\beta_{\mu\lambda}$ which have the following form:

$$\beta_{\mu\lambda} = \int \Phi_\mu (e^2 / r_B) \Phi_\lambda \, d\tau_\lambda$$

5) Two center one electron attractions $V_{\mu\nu, B}$ between an electron in the distribution $\Psi_\mu \Psi_\nu$ at atom A and the core of atom B.

6) Two-center two-electron repulsion integrals $(\mu\nu, \lambda\sigma)$. These integrals represent the energy of interaction between the charge distribution at atom A and atom B, and they have the following form:

$$(\mu\nu, \lambda\sigma) = \iint \Phi_\mu^{*(1)} \Phi\nu^{(1)} [1/r_{12}] \Phi\lambda^{*(2)} \Phi\sigma^{(2)} \, d\tau_1 d\tau_2$$

The total energy of the molecule E_{tot} is the sum of the electronic energy E_{el} and the repulsions between the cores of atoms A and B (E_{AB}^{core}).

$$E_{tot}^{mol} = E_{el} + \sum_A \sum_B E_{AB}^{core}$$

The heat of formation of the molecule is obtained from its total energy by subtracting the electronic energies and adding the experimental heats of formation of the atoms in the molecule:

$$\Delta H_f^{mol} = E_{tot}^{mol} - \sum_A E_{el}^A + \sum_A \Delta H_f^A$$

The electronic energies of the atoms are calculated from restricted single-determinantal wave functions using the same approximations and parameters as in molecular NDDO calculations.

C) Molecular mechanics:

The section provides a brief description of molecular mechanics method, which are not quantum mechanical in nature. This method is based on the classical Hooke's law (harmonic oscillator), which treats the chemical bond between two adjacent atoms as a spring with a force constant k. The restoring force between these two atoms is proportional to the displacement from the rest position, equilibrium (x):

$$F = -kx$$

The potential energy of the spring is:

$$U = \frac{1}{2} kx^2$$

This method tries to reproduce experimental data using parameterization of mechanical and electronic interactions. The potential energy of the molecule is a summation of stretching, bending, torsional motions and van der Waals interactions over the molecule, in addition to charge-charge and dipole-dipole effects.

$$E = E_{str} + E_{bend} + E_{torsion} + E_{non\text{-}bond}$$

The potential energy curves for individual terms are approximately transferable (e.g. CH stretches in ethane almost the same as in octane). The terms consist of functional forms and parameters, where, the parameters are chosen to fit structures (in some cases vibrational spectra, steric energies). The force field is comprised of functional forms, parameters and atom types where each atom is assigned an atomic type based on its bonding and environment (e.g. carbon: sp^3, sp^2, sp, aromatic, carbonyl, etc). The parameters are assigned based on these atom types (e.g. different C-C bond length and force constant for sp^3-sp^3 vs. sp^2-sp^2).

Examples of molecular mechanical models are MM2, MM3, AMBER, SYBYL, DREIDING, UFF, MMFF, etc which choose slightly different values for these parameters. Molecular mechanics force fields are designed to be transferable, and can be used for broad classes of molecular systems (but stay within the scope of the original parameterization), whereas force fields used for vibrational analysis, and analytical potential energy surfaces used for dynamics, are custom fit for individual systems. For bond stretching many force fields use a quadratic term, but this energy is too large for very elongated bonds

$$E_{str} = \Sigma\, k_i\, (r - r_0)^2$$

The Morse potential is more accurate, but is usually not used because of the added inefficiency.

$$E_{str} = \Sigma\, D_e\, [1\text{-}\exp\, (-\beta(r - r_0)]^{\,2}$$

A cubic polynomial has the wrong asymptotic form, but a quartic polynomial is often a suitable fit.

$$E_{str} = \Sigma\, \{k_i\, (r - r_0)^{\,2} + k'_i\, (r - r_0)^{\,3} + k''_i\, (r - r_0)^{\,4}\}$$

For bend terms a quadratic polynomial is usually sufficient.

$$E_{bend} = \Sigma\, k_i\, (\theta - \theta_0)^2$$

For very strained systems (e.g. cyclopropane) a higher polynomial is necessary.

$$E_{bend} = \Sigma\, k_i\, (\theta - \theta_0)^{\,2} + k_i'\, (\theta - \theta_0)^{\,3} + k_i''\, (\theta - \theta_0)^{\,4} + \ldots$$

For most torsional motions force fields with a single cosine function by appropriate barrier multiplicity, n are used.

$$E_{tors} = \Sigma\, V_i \cos\, [n\, (\theta - \theta\, _0)]$$

The non-bonded terms include van der Waals, electrostatic and hydrogen bonded interactions.

$$E_{non\text{-}bond} = E_{vdW} + E_{es} + E_{Hbond}$$

These non-bonded terms are expressed as Lennard-Jones potential

$$E_{vdW} = \Sigma\, 4\, \varepsilon_{ij}\, ((\sigma_{ij} / r_{ij})^{12} - (\sigma_{ij} / r_{ij})^{6})$$

or Buckingham potentials which are harder to compute since A, B and C are difficult to assign.

$$E_{vdW} = \Sigma\, A \exp(-B\, r_{ij}) - C\, r_{ij}^{-6}$$

σ and ε for each atom(A or B) are tabulated, and for pairs of atoms are obtained by the arithmetic and geometric means.

$$\sigma_{AB} = (\sigma_{AA} + \sigma_{BB})/2; \quad \varepsilon_{AB} = (\varepsilon_{AA}\, \varepsilon_{BB})^{1/2}$$

Some force fields add an extra term for hydrogen bond interactions. This requires them to be identified before the calculation is carried out.

$$E_{Hbond} = \Sigma\, A\, r_{ij}^{-12} - C\, r_{ij}^{-10}$$

More accurate representation of the potential energy surface (e.g. for vibrational frequencies) requires interaction cross terms between stretch, bend and torsional motions themselves, which are coupling effects.

The most important terms are,

$$E_{str\text{-}str} = \Sigma\, k_{ij}\, (r_i - r_{i0})\, (r_j - r_{j0})$$

$$E_{str\text{-}bend} = \Sigma\, k_{ij}\, (r_i - r_{i0})\, (\theta_j - \theta_{j0})$$

$$E_{bend\text{-}bend} = \Sigma\, k_{ij}\, (\theta_i - \theta_{i0})\, (\theta_j - \theta_{j0})$$

$$E_{bend\text{-}bend\text{-}tors} = \Sigma\, V_{ij}\, (\theta_i - \theta_{i0})\, (\theta_j - \theta_{j0}) \cos [n\, (\theta_{ij} - \theta_{ij0})]$$

Nonbonding energetics can be explained as follows. Steric energy is the energy relative to an artificial structure with no interactions which can be used to compare different conformers of the same molecule. Strain energy is the energy relative to a strainless molecule also with no interactions e.g. all trans hydrocarbons. It is not always safe to decompose energy into individual components (stretch, bend torsion, non-bonded etc) because different force fields can give similar energies and structures but quite different arbitrary components. The heat of formation is computed by averaging the bond energies (added to the strain energy) to get approximate atomization energy.

Heat of formation of the molecule = Atomization energy of the molecule – Heat of formation of the atoms.

To compare with experimental quantities, the heats of formation of the atoms must be calculated from the elements in their standard states at 298 K and 1 bar.

With molecular mechanics, the effects of substituents on geometry and strain energy can be found. Molecular mechanics is well parameterized for organic molecules, less so for inorganic molecules. In molecular mechanics specialty force fields are available for proteins, DNA, and liquid simulation. Molecular mechanics cannot be used for reactions that break bonds. It is useful for simple organic problems like ring strain in cycloalkanes, conformational analysis etc. High end biochemistry problems are solved, such as docking of substrates into active sites, refining x-ray structures, determining structures from NMR data, free energy simulations etc.

D) Hartree-Fock (HF) or *ab initio* model:

Hartree-Fock (HF) or *ab initio* theory approximates the true many-electron wave function as a product of N single electron wave functions (Scheme 1), in a determinantal form which incorporates the Pauli principle that it changes sign when any two electrons are exchanged.

$$\Psi(1, 2 \ldots N) = 1/\sqrt{N!} \begin{vmatrix} U_1(1) & U_2(1) \ldots\ldots U_n(1) \\ U_1(2) & U_2(2) \ldots\ldots U_n(2) \\ \vdots & \\ U_1(n) & U_2(n) \ldots\ldots U_n(N) \end{vmatrix}$$

Scheme 1.

where the U's are orthonormal spin orbitals and $1/\sqrt{N!}$ is a normalizing factor. The energy of the system is obtained using the variation method, which minimizes the energy with respect to the coefficients of the orbitals. It is based on the Born-Oppenheimer approximation, which states that nuclear motion is independent of electronic motion. Furthermore, it states that the nuclei are frozen in their equilibrium positions on the time scale of electronic motion. In the Hartree-Fock model, a Slater type orbital is used, to represent atomic orbitals, which has the following exponential form.

$$S_{n, l, m} = N_{(n, l)}\, r^{n-1}\, e^{-\zeta r}\, Y_m(\theta, \varphi)$$

Here ζ is the adjustable parameter and $Y_m(\theta, \varphi)$ are spherical harmonics. This function is approximated using a linear combination of one to six gaussian type orbitals

according to the following equation:

$$\Phi(r) = (2\alpha/\pi)^{3/4} e^{-\alpha r^{**}2}$$

The HF approximation which states that the many-electron wave function can be approximated using a multiplication of many one-electron-functions follows from the assumption that the electron solutions for the many-electron molecule will closely resemble a combination of one-electronic energy solutions for the hydrogen atom. Various levels of HF theory are possible, and some selection is required depending on the system to be investigated, the accuracy desired, and the computer resources at hand.

E) Hartree-Fock wavefunctions and basis sets:

1) Closed-shell and Open-Shell determinantal wavefunctions: [13]

Closed shell single determinant wave functions characterize the most commonly used form of HF theory, and they are appropriate for the description of the ground states of most molecules with an even number of electrons. Open shell single determinant wavefunctions are appropriate for the description of systems with an odd number of electrons, as well as for systems with an even number of electrons characterized by electronic states other than closed-shell singlets.

2) STO-3G minimal basis set: [14]

The basis set Slater Type Atomic Orbital termed STO-3G consists of a minimal basis uses Slater type orbitals and expands each one in terms of three gaussian functions. This is important because it approximates the Slater type functions in terms of three mathematically manageable gaussian functions. Slater type orbitals are mathematically difficult and their calculations are time consuming to perform because of the evaluation

of two electron-integrals. Gaussian functions can be evaluated analytically due to the exp ($-r^2$) term, which is not available in Slater type orbitals (which depend on exp (-r)). Two features of the gaussian expansions which comprise the STO-3G basis sets should be noted. First, the Slater 2s, 3s, 4s, and 5s functions are expanded in terms of the simplest s-type gaussians, that is, zero-order s gaussians, rather than as combinations of higher-order gaussian forms. This is also the case for the Slater 3p, 4p, and 5p atomic orbitals, all of which are written in terms of first-order p gaussian functions, and for the 4d orbitals, which are expressed as linear combination of second-order d gaussian primitives. A second feature to note is that the expansions for atomic functions of given principal quantum number n often share a common set of gaussian exponents, α.

The absolute energies of atoms and molecules calculated using the STO-3G minimal basis sets exhibit strong dependence on the value of properties such as energy differences, optimum geometries, charge distributions, and electric dipole moments. With the exception of the total energy, molecular properties calculated using the STO-3G basis set yields properties that are reasonably close to limiting values, the method of which is selected as an optimum compromise for widespread application with the important exception of equilibrium geometries.

3) 6-21G and 3-21G split valence basis sets: [14, 15]

The 6-21G and 3-21G basis sets are defined through the second and fourth row of the periodic table characterize representations in which two basis functions, instead of one have been allocated to describe each valence atomic orbital. These basis sets are representations in which each outer shell orbital (except for H) is represented by two functions, one contracted and the other diffuse. In each case, the inner functions are

composed of two gaussians while the outer (diffuse) functions are made up of only one gaussian. In the 6-21G basis set, each inner-shell atomic orbital ((1s) for second row atoms) is written in terms of six gaussian primitives. Except for hydrogen, the 6-21G and 3-21G basis sets are employed without rescaling of the valence functions to account for changes that might occur due to molecule formation. There are two main reasons behind this choice. First, optimum scale factors greater and less than unity have been found in molecules for all first-row atoms with the exception of boron. Second, if the basis sets are to be used for exploration of reaction potential surfaces, where molecules may be partly dissociated, a good description of the free atom is equally important.

4) 6-31G*and 6-31Gbasis set:**

These are the simplest of polarization basis sets originally proposed by Hariharan and Pople [16] for first-row atoms and later extended to second-row elements. 6-31G* basis set is the simplest of the two polarization basis sets, 6-31G* and 6-31G**, uses six gaussian primitives for the core orbitals, and a three/one split pair for each s and p valence orbital and a single set of six d-functions (*), second order gaussians which are equivalent to one s and five d orbitals, for all second row atoms. This basis set contains no provision for polarization of the s orbitals on hydrogen. 6-31G** basis set has in addition to 6-31G*, a set of p orbitals which has been added to each hydrogen to allow for polarization in hydrogen bonding. The p orbitals perform the same function for the s-valence orbital of hydrogen as the d-orbitals perform for the p valence orbitals for the second row elements.

5) 6-311G basis set:**

These are the larger polarization basis sets which are more flexible than the simple 6-31G* and 6-31G** but are generally limited to only small molecular systems. A higher level of valence shell orbital splitting is contained in the 6-311G basis set. As before, the inner-shell atomic orbitals are represented in terms of six gaussian primitives, but it splits each valence function into three parts instead of two, these being written in terms of three, one and one gaussian primitives, respectively. The basis is supplemented by a single set of five d-type gaussian functions for first-row atoms, and a single set of uncontracted p-type gaussians for hydrogen. One special feature of this basis set is that the gaussian exponents and expansion coefficients have been chosen so as to minimize the energy of the atomic ground state at the higher order second-order Moller-Plesset (MP2) perturbation level, rather than at the corresponding Hartree-Fock level. This allows for the possibility of excited state basis functions to incorporate electron correlation within the atomic valence region.

6) 6-31+G* and 6-311+G basis set:**

The 6-31+G* basis set is constructed for first and second-row elements from the underlying 6-31G* representations by the addition of single set of diffuse gaussian s and p-type functions. Several sets of data clearly show the effect of added diffuse functions to be far more significant for the negatively charged species than for the corresponding neutral species. 6-311+G** basis set represents six gaussians for 1s orbitals of all non H atoms, and a three way split valence for all H and valence orbitals of non H atoms, the inner part using three gaussians, a middle part and outer part each by one gaussian.

F) Density functional theory:

Although density functional theory (DFT) has been known for many years, it is only recently (since 1990) that its application to chemical problems has become popular. In the first stage of DFT, the energy is expressed as a "functional"of the density of a uniform electron gas (E (ρ)). This is then modified to express the electron density around the nuclear frame work. The choice of the functional is the only unspecified variable of the DFT method. At the present time, there is no systematic way of choosing the functional, and the most popular ones in the literature have been derived by careful comparison with experiment. Some of the most common methods are: BP86 - developed by Becke and Perdew in 1986,[17] BLYP - developed by Becke, Lee, Yang and Parr, B3LYP - a modification of BLYP in which a three-parameter functional developed by Axel Becke is used.

The Thomas-Fermi-Dirac model explained energy as a function of the one electron density, ρ, nuclear-electron attraction, V_{NE} and electron-electron repulsion J[ρ].

$$V_{NE}[\rho] = \sum_A Z_A \int \frac{\rho}{r_A} d\tau$$

$$J[\rho] = \int \frac{\rho(1)\rho(2)}{r_{12}} d\tau_1 d\tau_2$$

The Thomas-Fermi approximation for the kinetic energy is given as

$$T[\rho] = c\int \rho^{5/3} d\tau \quad c = \frac{3}{10}(3\pi^2)^{2/3}$$

The Slater approximation for the exchange energy appears as

$$K[\rho] = c\int \rho^{4/3} d\tau \quad c = -\frac{9\alpha}{8}\left(\frac{3}{\pi}\right)^{1/3} \quad \alpha = 1$$

Hohenberg and Kohn (1964) described energy as a functional of the density $E[\rho]$ where the functional is universal, independent of the system and the exact density minimizes $E[\rho]$, which applies only to the ground state. Kohn and Sham (1965) proposed variational equations for a local functional where the density can be written as a single determinant of orbitals (but orbitals are not exactly the same as Hartree-Fock).

Density Functional Theory (DFT) is a technique for solving the Schrödinger equation:

$$\mathbf{H\Psi = E\Psi}$$

where the Hamiltonian operator H is a combination of kinetic and potential energy operators, which account for the interacting electrons, the nuclear-electron attraction, coulombic repulsion of the nuclei and a final term for the exchange and correlation energy of the electrons. DFT is an accurate method for predicting the ground state energy of compounds. The idea behind DFT is to describe the interacting system of electrons by means of its charge density and not by as in other methods a many electron wave function. In DFT Ψ is a functional. The electron exchange correlation functionals (function of a function) are usually described as functions, which result from operating on the electronic density distribution. The following different techniques are used for approximation of exchange correlation function.

1) Local density approximation (LDA).

2) Non local extension to LDA.

3) Hybrid Functional Methods.

The LDA functional can be regarded as an exchange correlational functional for a homogeneous electron gas. It is based on assumption that there is existence of an energy functional and derived an expression for the kinetic energy based on the on the density of electrons, $\rho(r)$, in an infinite potential well. This kinetic energy functional is given by:

$$T_{TF}(\rho) = (3/10)\,(3\pi^2)^{2/3} \int dr\, \rho^{5/3}(r)$$

$$E_{TF}(\rho) = T_{TF}(\rho) - Z\int dr\, \rho(r)/(R-r) + \tfrac{1}{2} \int dr_1\, dr_2\, \rho(r_1)\, \rho(r_2)/(r_1 - r_2)$$

According to the above equation proposed by Thomas and Fermi, energy is given by a sum of the electronic kinetic energy (T), electron nucleus attraction energy, and the Hartree correlation energy where Z is the nuclear charge, and R is the position vector of the nucleus.

The first Hohenberg-Kohn theorem [18] states: "The external potential v (r) is determined, within a trivial additive constant, by the electron density ρ (r)". Since the electron density is determined by the number of electrons, it follows that the electron density determines the wave function and thereby all of the ground state properties of the system. The proof of the theorem assumes the existence of two external potentials which differ by more than a constant, and which lead to the same ground state density. This implies the existence of two different Hamiltonians, with differing wave functions, corresponding to the same ground state electron density. The electron density determines all properties of the ground state including the total ground state energy E_{tot}, the ground state kinetic energy T, the energy of the electrons in the external potential U, and the electron-electron interaction energies W, all of these are functionals of the electron density. The total energy can therefore be written as:

$$E^0(\text{tot}) = E_{tot}(\rho^0) = T(\rho^0) + U(\rho^0) + W(\rho^0)$$

where ρ^0 denotes the true ground state electron density of the system. The first Hohenberg-Kohn theorem thus establishes the existence of the total energy functional, but it does not provide the solution to the many body electron problems.

The second Hohenberg-Kohn theorem provides a variational principle:" For any given non-negative trial density, ρ (r), that integrates to the correct number of electrons, N, the true ground state energy E^0 (tot) satisfies the relation":

$$E^0(\text{tot}) \leq E_{tot}[\rho(r)].$$

This provides a way (by minimization of E_{tot}) to obtain solutions for the total energy if T (ρ), U (ρ) and W (ρ) are known functionals. They considered that an electronic system with a given Hamiltonian has ground state energy as well as a ground state wave function, which is completely determined by the minimization of the total energy as a functional of the wave function. They also assumed that the external potential, together with the number of electrons, completely determines the Hamiltonian, and that these two quantities determine all the properties of the ground state.

To treat the kinetic energy better, Kohn and L. J. Sham introduced non-interacting orbitals of neutral density. They did this to find a better approximation to the unknown T (ρ) and W (ρ) functionals. Using a non-interacting reference system, for which the ground state density is exactly equal to the ground state density of the fully interacting system, they succeeded in showing that any N-electron density can be uniquely decomposed into orbitals. These are called the Kohn-Sham orbitals (KS), and the expectation value of the kinetic energy operator using these orbitals is the non-interacting kinetic energy T_s (ρ). These orbitals may be based on normal HF orbitals, s, p, d, and f.

When the Hohenberg-Kohn variational principle is applied to the Kohn-Sham orbitals, the canonical Kohn-Sham equations are obtained:

$$[- \tfrac{1}{2} (\partial^2 / \partial \tau^2)_i + V_{eff}] \psi_i = \varepsilon_i \psi_i$$

where τ includes coordinate x, y, and z and i runs over all electrons, ε_i is the Kohn-Sham eigenvalue of electron i, and the effective potential is:

$$V_{eff}(r) = V_{ext}(r) + \int dr' \rho(r')/(r-r') + V_{xc}(r)$$

where $r^2 = x^2 + y^2 + z^2$. Here V_{ext} is the external potential, and the exchange-correlation potential is defined as the functional derivative of the exchange and correlation energy with respect to the density:

$$V_{xc}(r) = \partial E_{xc}[\rho(r)] / \partial \rho(r)$$

These equations are nonlinear like the Hartree-Fock equations and are thus solved by an equivalent self consistent (iterative) procedure. The resulting density is given when the energy converges to a preset limiting value.

$$\rho(r) = \Sigma_I \Psi_i^*(r) \Psi_i(r)$$

and the Kohn-Sham eigenvalues then give the total ground state energy using either of two equivalent solutions:

$$E_{tot}(\rho) = T_s(\rho) + U(\rho) + H(\rho) + E_{xc}(\rho) \quad \text{(or)}$$

$$E_{tot}(\rho) = \sum_I(\varepsilon_i) - \tfrac{1}{2}\int drdr'\rho(r)\rho(r')/(r-r') + E_{xc}(\rho) - \int drV_{xc}(r)\rho(r)$$

The non-local extension to LDA is necessary because molecules are logically in a heterogeneous electron gas. A hybrid functional employs Hartree-Fock treatment for the exchange term as repulsion integrals.

G) Geometry optimization: [19]

In molecular modeling we are interested in minimum points on energy surfaces. This is because the minimum energy arrangement of the atoms corresponds to the most stable equilibrium states of the system, and any movement away from this minimum gives a configuration with a higher energy (less stability). To identify those geometries of the system that correspond to minimum points on the energy surface a minimization algorithm is used. The different types of algorithms used are:

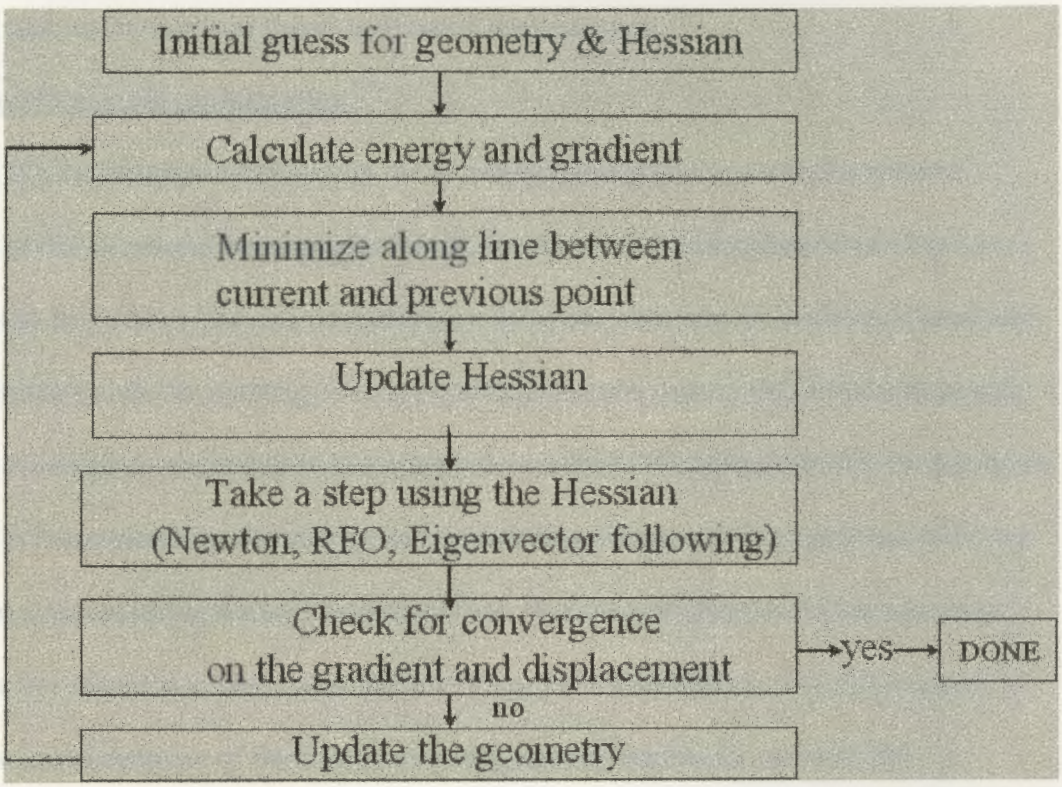
1) Univariate search, axial iteration which has slow convergence especially if there is strong coupling between two or more of the coordinates.

2) Conjugate gradient and quasi-Newton methods have better convergence with numerical or analytical gradients.

3) Newton methods have rapid convergence which use not only first derivatives (i.e. the gradients) but also the second derivatives to locate a minimum. Second derivatives provide information about the curvature of the function.

When optimization fails there are certain steps that we need to take as follows: if the result (output) shows that a preset number of cycles is exceeded, then we check for very flexible coordinates and/or strongly coupled coordinates. We need to increase the number of cycles in the options menu when the calculation is set up. **OPT= (Restart Maxcyc=N)** is the command. The highest point on the pathway between two minima is

of especial interest and is known as the saddle point with the arrangement of atoms being in the transition state structure. The following chart presents the geometry optimization steps for energy minima.



The choice of which minimization algorithm is used depends on a number of factors, including the storage and computational requirements, the relative speeds with which the various parts of the calculation can be performed, the availability of analytical derivatives etc. Thus, any method that requires the Hessian matrix to be stored may present memory problems when applied to systems containing thousands of atoms. The Hessian gives the second derivative of the total energy with respect to each (i) normal coordinate ($\delta^2E/\delta q_i^2$). Calculations on systems of this size are invariably performed using

molecular mechanics, and clusters of atoms treated as one body. Quantum mechanical calculations are restricted to systems with relatively small number of atoms, so storing the Hessian matrix is not a problem. As the energy calculation is often the most time-consuming part of the complete problem, it is desirable that the minimization method chosen take as few steps as possible to reach the minimum.

H) Transition state optimization [19]

We are interested not only in the thermodynamics of a process (the relative energy of the initial and final species that define the total thermodynamic driving force) but also in its kinetics (the rate of conversion from one structure to another). Knowledge of minimum points on an energy surface enables thermodynamic data to be interpreted, but for kinetics it is necessary to investigate the nature of the energy surface at maximum points. We want to know the energy variation during changes in geometry and when we arrive at a successful optimized transition state. For the optimization of the transition state we use the term in Options as **OPT = TS** or **OPT (Saddle=*n*)**. For this we have to input an initial estimate of the transition state geometry and make sure that the coordinates dominating the decomposition vector are not strongly coupled to the remaining coordinates. We also have to make sure that the initial Hessian has a negative eigenvalue which corresponds to the decomposition notion. In case of optimization failure we use **CALCFC** or **READFC** to continue the calculation to reach optimization. These increase parameter flexibility.

We used PC Spartan Pro® and 2004® for building and locating transition states. The software assists in providing both an extensive and extendable library of calculated transition states and a facility for matching as closely as possible entries in the library

with the reaction in hand. If the reaction is unknown to the library, a fallback technique similar to the linear synchronous transit method is automatically invoked, which guesses the average of reactant and product geometries. We can use **OPT=QST2** and **OPT=QST3** in the Options menu where synchronous transit guided transition state search include optimization in redundant internal coordinates which is divided into:

1) QST2: Input a reactant-like structure and a product-like structure (initial estimate of transition state by linear interpolation in redundant internal coordinates). Atoms need to be specified in the same order in each structure (Input structures do not correspond to optimized structures).

2) QST3: Input reactant, product, and estimate of transition state.

The first few steps search for a maximum along path and the remaining steps use regular transition state optimization method (quasi-Newton with eigenvector following/RFO). If transition vector deviates too much from path it automatically chooses a better vector to follow. The maximum number of optimization steps required for the transition state geometries for the dismutation reaction without copper usually required 60 to 120 steps and with copper is150 to 180 steps, depending on the level of theory.

I) Thermodynamics and kinetics: [19]

A wide variety of thermodynamic properties can be calculated from computer simulations, and comparison of experimental and calculated values for such properties is an important way in which the accuracy of the simulation and the underlying energy model can be quantified. Simulation methods also enable predictions to be made of the thermodynamic properties of systems for which there is no experimental data, or for which experimental data is difficult or impossible to obtain. Simulations can also provide

structural information about the conformational changes in molecules and the distributions of molecules in a system. Unfortunately, the free energy is a difficult quantity to obtain in systems such as liquids or flexible macromolecules that have many minimum energy configurations separated by low-energy barriers and intermolecular interactions. Associated quantities such as the entropy and chemical potential are also difficult to calculate in such condensed phases.

The equilibrium constant for the reaction is a thermodynamic property of the reactants and products and is a function of the standard Gibbs free energy, ΔG^0, according to the following equation:

$$\Delta G^0 = -RT\ln K_e$$

where R is the gas constant and T is the absolute temperature and K_e is the equilibrium constant.

In general, $\Delta G > 0$ refers to a nonspontaneous reaction;

$\Delta G = 0$ refers to an equilibrium state of the reaction;

$\Delta G < 0$ refers to a spontaneous reaction;

ΔG^0 is related to the standard entropy and enthalpy of the reaction by the following equation. The superscript (^{0}) refers to the standard conditions at 298K and 1 bar pressure.

$$\Delta G^0 = \Delta H^0 - T\Delta S^0$$

Our calculations give ΔH^0 and S^0 for individual compounds, from which we calculate ΔG^0.

The rate constant of a reaction is a kinetic property and is related to the activation energy by the Arrhenius equation:

$$\ln k_1 = \ln A - Ea/RT$$

where k_1 is the rate constant and A is a frequency factor. The alternative Eyring equation gives a relation which links the rate constant and the equilibrium constant between the reactants and transition state:

$$k_1 = (kT/h)\, K^{\dagger}$$

where k is Boltzmann constant, h is Planck constant and $K^{\dagger}$ is the equilibrium constant between reactants and transition state. The superscript (†) refers to this equilibrium. The free energy of activation is related to the enthalpy of activation and entropy of activation by the equation below:

$$\Delta G^{\dagger} = \Delta H^{\dagger} - T\Delta S^{\dagger}$$

The activation energy E_a is related to the enthalpy of activation $\Delta H^{\dagger}$ by the following equation:

$$E_a = \Delta H^{\dagger} + 2RT$$

J) Modeling chemical reactions and electron correlation effects. [19]

The preferred technique for modeling chemical reactions is usually considered to be quantum mechanics, but if one wishes to represent the whole system explicitly, the large number of atoms that must be considered is rarely practical in case of *ab initio* quantum mechanics. Furthermore environmental effects such as solvent interactions are not readily taken into account. We consider three methods used to study chemical reactions that involve large molecules. One strategy is to use a purely empirical approach. An alternative is to divide the system into two parts and treat the reaction region using quantum mechanics, and the rest of the system being modeled using molecular mechanics. Third, we shall consider techniques such as density functional theory or Carloni-Parrinello method.[7, 19]

The empirical approaches employ force field models for studying reactions, which are used to estimate the activation energies of possible transition states to explain the reaction. The force field model is usually derived by extending an existing force field to enable the structures and relative energies of transition structures to be determined. This second approach to simulate chemical reactions in solution is to use a combination of quantum mechanics and molecular mechanics. The reacting parts of the system are treated quantum mechanically, with the remainder being modeled using the force field. The total energy E_{TOT} for the system can be written as

$$E_{TOT} = E_{QM} + E_{MM} + E_{QM/MM}$$

where E_{QM} is the energy of those parts of the system treated exclusively with quantum mechanics, and E_{MM} is the energy of the purely molecular mechanical parts of the

system. $E_{QM/MM}$ is entirely due to non-bonded interactions between the quantum mechanical and molecular mechanical parts of the system.

It is well known that ignoring electron correlation and bond harmonicities contribute an error of 10% of the energy of a typical calculation. Moller-Plesset theory (MP2, MP3 etc.) provides a way to correct for the electron correlation error by including electronically excited state configurations in the ground state wave function.

K) Effect of solvation:

Since nearly all computer models are based on gas phase molecules or ions, their results may be influenced by solvent effects. They are impractical to include in a specific way, but may be crudely reflected in molecular mechanics calculations. Generally the effects are small when considering entropy and geometry changes, but may become significant in process involving changes in polarity.

In our reactions involving charge dispersed initial and transition states, polar solvent stabilization will affect both states similarly. Since the products O_2 and H_2O_2 are not as highly polarized, they would not be stabilized as much as the reactants, so one would expect the solvated exothermocities to be slightly less.

Chapter 3 Results and Discussion.

Introduction:

This project is done at several theoretical levels to calculate thermodynamic and kinetic properties using semi-empirical, Hartree-Fock, density functional and electron correlation methods. Superoxide dismutase catalyses the dismutation reaction of the toxic superoxide radical anion to molecular oxygen and hydrogen peroxide, thus forming a crucial part of the cellular antioxidant defense mechanism.

$$2\ O_2^- + 2\ H_3O^+ \xrightarrow{SOD} O_2 + H_2O_2 + 2\ H_2O$$

Our model reaction replaces H^+ by H_3O^+ since the latter is much more amenable to theoretical treatment and accurate comparison with the experiment.

A) Thermodynamics:

The heat of formation and entropies (translational, rotational, and vibrational) were calculated for each reactant and product by the standard statistical thermodynamic methods in the Spartan programs. From these we calculated ΔH_{Rxn}, ΔS_{Rxn}, and ΔG_{Rxn} for the reaction at semi-empirical (AM1 and PM3), Hartree-Fock (STO-3G, 3-21G*, 6-31G**, 6-311+G**), density functional (B3LYP at 6-31G*, 6-31G**, 6-311+G**), and finally MP2 levels of treatment.

A1) Results:

Table 3 shows computed values of **heats of formation** of the reactants and products. Given below are the tables with calculated heat of formation, entropy (sum of translational, rotational, and vibrational), and ΔH_{Rxn}, ΔS_{Rxn}, ΔG_{Rxn} with the different theories and basis sets. For the heats of formation of the model reaction compounds (Table 3), we note the reasonable values relative to the experiment even with AM1 except for $O_2^{\cdot -}$, and the unusual non-zero ΔH_f of O_2 with PM3.

Table 3. Heats of formation of dismutation reaction components (kcal or au).

LEVEL OF THEORY	$O_2^{\cdot -}$	H_3O^+	O_2	H_2O_2	H_2O
SE (AM1)-kcal/mol	-22.181	143.478	0.725	-35.227	-59.240
SE (PM3) - kcal/mol	-13.275	159.077	18.384	-40.778	-53.427
HF (STO3G)-au	-147.384	-75.330	-147.551	-148.765	-74.965
HF (3-21[G*])-au	-148.723	-75.891	-148.687	-149.945	-75.585
HF (6-31G*)-au	-149.563	-79.310	-149.532	-150.776	-76.023
HF (6-31G**)-au	-149.743	-79.363	-149.591	-150.881	-76.096
DF-B3LYP (6-31G*)-au	-150.299	-76.689	-150.257	-151.532	-76.408
DF-B3LYP (6-31G**)-au	-150.299	-76.705	-150.257	-151.542	-76.419
DF-B3LYP (6-311+G**)-au	-150.285	-76.881	-150.265	-151.551	-76.423
MP2/6-31G** (au)	-149.910	-76.506	-149.90	-151.153	-76.213
EXPERIMENTAL-kcal/mol	-11.610	138.900	0.00	-32.53	-57.801

Table 4 shows computed values of **entropies (translational, rotational and vibrational)** of reactants and products. In the case of entropies, the computed values are close to the experimental quantities, again except for $O_2^{\cdot -}$ and O_2. $O_2^{\cdot -}$ is an open shell anion radical and O_2 is a ground state diradical. Electronic degeneracies are included in the modeling programs but open shell systems are still difficult to predict accurately.

Table 4. Absolute entropies of the dismutation reaction components (cal /K mol).

LEVEL OF THEORY	$O_2^{\cdot -}$	H_3O^+	O_2	H_2O_2	H_2O
SE (AM1)	46.688	46.199	46.379	56.362	45.09
SE (PM3)	46.96	46.039	46.672	54.782	44.991
HF (STO3G)	47.119	46.243	46.845	55.366	45.283
HF (3-21[G*])	47.333	46.43	46.918	56.877	45.087
HF (6-31G*)	47.04	46.117	46.60	54.001	44.958
HF (6-31G**)	47.06	46.114	46.659	54.001	44.956
DF-B3LYP(6-31G*)	47.257	46.186	46.80	53.353	45.10
DF-B3LYP (6-31G**)	47.25	46.21	46.80	53.28	45.08
DF-B3LYP (6-311+G**)	47.25	46.22	46.80	53.21	45.06
MP2/6-31G**	47.39	46.12	47.01	54.10	45.06
EXPERIMENTAL	50.09	45.95	49.03	55.69	45.13

Table 5 shows computed values of ΔH_{Rxn}, ΔS_{Rxn}, **and** ΔG_{Rxn} at different levels. Note that only the higher levels of theory predict ΔS_{Rxn} well, while AM1 (and not PM3) and most HF and DF levels are almost within the experimental error (+/- 0.1 kcal) of the calorimetric values.

Table 5. ΔH_{Rxn}, ΔS_{Rxn}, and ΔG_{Rxn} of dismutation reaction at various levels of theory (kcal/ K mol)

LEVEL OF THEORY	ΔH_{Rxn}	ΔS_{Rxn}	ΔG_{Rxn}
AM1	-395.025	0.007	-397.26
PM3	-457.62	0.005	-459.24
HF-STO3G	-514.15	0.006	-515.90
HF-[3-21G*]	-360.93	0.006	-362.82
HF-6-31G*	-380.15	0.004	-382.35
HF-6-31G**	-381.88	0.004	-383.14
DF-B3LYP-6-31G*	-395.19	0.003	-396.22
DF-B3LYP-6-31G**	-394.24	0.003	-395.23
DF-B3LYP 6311+G**	-394.13	0.003	-395.02
MP2	-395.75	0.003	-396.65
EXPERIMENTAL VALUES	-402.70	0.003	-403.56

A2) Discussion:

When we take up this reaction we want to see whether we can explain the kinetics and active site in a reasonable manner. For this we first started with the basic thermodynamic properties to see which levels of theory are adequate. A wide variety of

thermodynamic properties can be calculated from computer simulations, and comparison of experimental and calculated values for such properties is an important way in which the accuracy of the simulation and the underlying energy can be understood. Simulation methods also enable predictions to be made on the thermodynamic properties of systems for which there is no experimental data, or for which experimental data is difficult or impossible to obtain. As the molecules in the present reaction are not so complex it is not too surprising that many levels of theory are approximately correct, but only the highest levels of HF and DF gave the reasonably close (except AM1) values. The computed values in general gave good agreement when compared to the experimental values.

B) Building and locating transition states:

We used both Spartan Pro® and 2004® for building and locating transition states. The software assists in providing both an extensive and extendable library of calculated transition states and a facility for matching as closely as possible entries in the library with the reaction in hand. As the reaction is unknown to library, a fallback technique similar to the linear synchronous transit method is automatically invoked which guesses the average of reactant and product geometries.

B1) Results:

Table 6 shows computed values of ΔH of formation of reactants ($2O_2^-$ and $2H_3O^+$), and transition state at different levels (no Cu^{2+}). Note that even the primitive AM1 is within 6% of the experimental values for the initial state.

Table 6. ΔH values of reactants ($2O_2^-$ and $2H_3O^+$), and transition state at different levels of theory without copper. (kcal or au)

LEVEL OF THEORY	ΔH of reactants	ΔH of transition state
AM1	241.318	291.335
PM3	185.998	235.157
HF-STO3G	-445.429	-445.344
HF-[3-21G*]	-449.229	-449.172
HF-6-31G*	-451.739	-451.687
HF-6-31G**	-451.745	-451.699
DF-B3LYP -6-31G*	-453.977	-453.937
DF-B3LYP-6-31G**	-454.010	-453.971
DF-B3LYP-6311+G**	-452.187	-452.138
MP2	-453.013	-452.997
EXPERIMENTAL	254.6	—

Table 7 shows **activation energies** at different levels of theory, with no copper present. Both semi empirical methods gave nearly the same high value for $\Delta H^{\neq}$, and the activation energy steadily lowers as more functions are added to the HF basis set.

Table 7. Activation energy without copper at various levels of theory (kcal/mol)

LEVEL OF THEORY	ACTIVATION ENERGY
AM1	50.017
PM3	49.159
HF-STO3G	53.337
HF-[3-21G*]	35.767
HF-6-31G*	32.630
HF-6-31G**	28.865
DF-B3LYP-6-31G*	25.100
DF-B3LYP-6-31G**	24.472
DF-B3LYP 6311+G**	30.747
MP2	22.59

Table 8 shows **number of optimization steps** required for the calculation of dismutation reaction without copper to reach transition state. It is clear from table 8 that the higher the level of theory the more cycles are required for optimization to be completed.

Table 8. Number of optimization steps required for Dismutation reaction without copper to reach transition state.

LEVEL OF THEORY	NUMBER OF OPTIMIZATION STEPS WITHOUT COPPER.
AM1	60
PM3	60
HF-STO3G	65
HF-[3-21G*]	65
HF-6-31G*	75
HF-6-31G**	80
DF-B3LYP-6-31G*	120
DF-B3LYP-6-31G**	120
DF-B3LYP 6311+G**	120
MP2	90

B2) Discussion:

The transition state (TS) is the highest point along the lowest energy pathway between reactants and products. It is very difficult to model transition states because unlike reactants and products, which are well defined entities, transition states are more distorted and their orbitals are more diffused. Transition states may exhibit elongated bonds, partial bonding and some aspects of excited states. Transition states cannot easily be observed experimentally, so it is difficult to devise parameters to model them. Existence of a unique transition state is not clear since it is very difficult to predict its structure. Moreover TS may involve partial bonding (very diffuse unconventional orbitals) and so lower levels of theory are not likely to help in modeling. We know relatively little about the geometries of TS, and most of what we know is based on calculations.

Based on the difficulties we followed an approach for predicting TS geometry. We first performed low level (AM1, PM3) semi-empirical MO calculation as a transition structure. We used this result as a starting point for the higher level calculation- (HF/3-21[G*], HF/6-31G*, HF/6-31G** in Spartan Pro® and density functional B3LYP/6-31G*, B3LYP/6-31G**, B3LYP/6-311+G** and MP2 in Spartan 2004®). For the best energy optimization values we did the single point energy calculations and verified them with the frequency calculation at different levels.

We calculate activation energy by taking the difference between **ΔH (transition state)** and **ΔH (initial state)**. The lower level (AM1&PM3) and HF/STO3G calculation yielded high activation energies of approximately 50 kcal/mol, and the remaining are in

the range of 24-36 kcal/mol. MP2 also gave slightly lower activation energy when compared to other levels of theory.

The final step in the process is verification and confirmation of the transition structure. This involves the calculation of an energy profile. In energy profile calculation a bond length or bond angle in the activated complex is varied incrementally along a reaction coordinate and the resulting structure is optimized at each step. The bond length that is the reaction coordinate is usually the one that is changing in the normal mode of vibration associated with the product formation. By observing the models of the structures at each step along the plots, one can confirm that the activated complex leads to products. We can confirm products formed from the activated complex by comparing the bond lengths of the products with the experimental values. The maximum number of steps for optimization depends on the level of theory for the dismutation reaction and is listed in the table 8 as 60 to 120 steps.

C) Building the transition state with copper (Cu^{2+}): Introduction:

When the Cu^{2+} is positioned at the center of the two $O_2^{\cdot-}$ and two H_3O^+, the possibility of electron transfer to produce Cu^+ and O_2 in the first instance, and from there $Cu^{2+} + HO_2^{\cdot} + H_2O$ in a second step, leading finally to H_2O_2 and H_2O, may occur with ease. Normally electron transfer reactions (to and from Cu^{2+}) occur with little or no activation energy when the donor and acceptor have opposite charges. In fact their cross section exceeds the hard sphere reaction values, leading to their characterization as "electron harpooning" mechanisms [23]. In this case, the activation energy comes from breaking and forming chemical bonds.

C1) Results:

Table 9 shows computed values of **ΔH of reactants and transition state with copper at different levels.** Note in general the greater stability of the more sophisticated theories, with the unexpected exception of the highest B3LYP level. However, as the table shows, when $\Delta H^{\neq}$ is calculated, this difference disappears.

Table 9. Values of ΔH's of reactants and, transition state with copper computed at different levels. (au)

LEVEL OF THEORY	ΔH of reactants with Cu^{2+}	ΔH of TS with Cu^{2+}
HF-[3-21G*]	-528.447	-528.428
HF-6-31G*	-537.124	-537.109
HF-6-31G**	-532.275	-532.261
DF-B3LYP-6-31G*	-533.975	-533.962
DF-B3LYP-6-31G**	-544.072	-544.062
DF-B3LYP 6311+G**	-526.311	-526.301
MP2	-561.245	-561.236

Table 10 shows computed **activation energies with copper** at different levels of theory. Comparison with Table 7 shows the dramatic lowering of the activation energy with Cu^{2+}. Table 12 contrasts both reactions.

Table 10. Activation energies of dismutation reaction with copper at different levels of theory (kcal).

LEVEL OF THEORY	ACTIVATION ENERGY WITH COPPER
HF-[3-21G*]	11.9
HF-6-31G*	9.4
HF-6-31G**	8.7
DF-B3LYP-6-31G*	8.1
DF-B3LYP-6-31G**	6.2
DF-B3LYP 6311+G**	6.2
MP2	5.6

Table 11 shows **number of optimization steps** required for the dismutation reaction with copper to reach transition state, a much greater number than 60-120 without copper

Table 11. Optimization steps required for the dismutation reaction with copper to reach transition state

LEVEL OF THEORY	NUMBER OF OPTIMIZATION STEPS WITH COPPER.
HF-[3-21G*]	150
HF-6-31G*	160
HF-6-31G**	160
DF-B3LYP-6-31G*	180
DF-B3LYP-6-31G**	180
DF-B3LYP 6311+G**	180
MP2	165

When copper is included with reactants and transition state, the higher levels of theory gave consistent results with the activation energies ranging from 5-12 kcal/mol approximately. Copper is clearly involved in the active site and participates in reducing the activation energy. The computational effort of transition state with copper is high, and the maximum number of steps to reach optimization is 150 to 180. This serves as an excellent source of information for the future research on this reaction.

Table 12 shows computed **activation energies with and without copper and their difference.** In general HF methods gave higher values of activation energies both with and without copper and greater reductions due to added copper. The consistent energy lowering effects of the density functional basis sets is no doubt due to their greater diffusiveness at distorted geometries.

Table 12. Activation energies with and without copper

LEVEL OF THEORY	ACTIVATION ENERGY(kcal) WITHOUT COPPER	ACTIVATION ENERGY (kcal) WITH COPPER.	REDUCTION OF E_{act} DUE TO CATALYSIS
HF-[3-21G*]	35.7	11.9	23.8
HF-6-31G*	32.6	9.4	23.2
HF-6-31G**	28.8	8.7	20.1
DF-B3LYP-6-31G*	25.1	8.1	17.0
DF-B3LYP-6-31G**	24.4	6.2	18.2
DF-B3LYP 6311+G**	30.7	6.2	24.5
MP2	22.5	5.6	16.9

C2) Metal ion effect:

It is known that Zn^{2+} has a structural, stabilizing role, in the active site, while Cu^{2+} is directly involved in the catalytic activity.[20] One of our goals is to establish its catalytic activity by modeling that it reduces the activation energy. We believe that copper also may play a role in steering protons (H_3O^+) to the superoxide at the active site as well as helping in the formation of H_2O_2 and O_2. Modeling transition state with copper had certain limitations. However the lower levels of theory AM1, PM3 and HF/STO3G failed to converge, and optimization exceeded the number of cycles. On the other hand higher levels of theory like HF/3-21[G*], HF/6-31G*, HF/631G**, DF/6-31G*, DF/6-31G**, DF/6311+G** and MP2 produced credible results for the illustration of catalytic activity.

D) Product confirmation:

After the calculations of activation energies we confirmed that the actual products formed from the activated complex are oxygen and hydrogen peroxide by comparing the bond lengths of these products with their experimental values. The values were also used to give a more detailed picture of the reaction.

D1) Results:

Table 13 shows computed bond-lengths of O-O in O_2, O-O in H_2O_2, and O-H in H_2O_2 at initial, transition and the final states. These values were used to construct a reaction profile as a series of single point energy calculations, where these distances were proportionally changed in 10-20 increments over the complete reaction.

Table 13. Computed bond-lengths of O-O in O_2, O-O in H_2O_2, and O-H in H_2O_2 at initial, transition and the final states.

LEVEL OF THEORY	BOND-LENGTH O-O in O_2 (A^o)			BOND-LENGTH O-O in H_2O_2 (A^o)			BOND-LENGTH O-H in H_2O_2 (A^o)		
	IS	TS	FS	IS	TS	FS	IS	TS	FS
HF-[3-21G*]	1.234	1.362	1.236	1.448	1.544	1.450	0.974	1.200	0.976
HF-6-31G*	1.234	1.340	1.242	1.448	1.520	1.452	0.974	1.196	0.976
HF-6-31G**	1.234	1.340	1.242	1.448	1.520	1.452	0.974	1.196	0.976
DF-B3LYP-6-31G*	1.234	1.342	1.240	1.448	1.522	1.454	0.974	1.198	0.976
DF-B3LYP-6-31G**	1.234	1.342	1.240	1.448	1.522	1.454	0.974	1.198	0.976
DF-B3LYP 6311+G**	1..234	1.340	1.240	1.448	1.522	1.454	0.974	1.198	0.976

Table 14 shows **calculated bond lengths of O-O in O_2, O-O in H_2O_2, and O-H in H_2O_2 at final states (products) and comparison with experimental values at different levels.** The close correspondence of the predicted and experimental values shows that theory is producing credible results.

Table 14. Calculated bond lengths of O-O in O_2, O-O in H_2O_2, and O-H in H_2O_2 at final states (products) and comparison with experimental values.

LEVEL OF THEORY	BOND LENGTHS IN A^0 OF FINAL STATES(PRODUCTS)					
	CALCULATED VALUE IN A^0			EXPERIMENTAL VALUE IN A^0		
	O-O O_2	O-O H_2O_2	O-H H_2O_2	O-O O_2	O-O H_2O_2	O-H H_2O_2
HF-[3-21G*]	1.236	1.450	0.976	1.240	1.454	0.976
HF-6-31G*	1.242	1.452	0.976	1.240	1.454	0.976
HF-6-31G**	1.242	1.452	0.976	1.240	1.454	0.976
DF-B3LYP-6-31G*	1.240	1.454	0.976	1.240	1.454	0.976
DF-B3LYP-6-31G**	1.240	1.454	0.976	1.240	1.454	0.976
DF-B3LYP 6311+G**	1.240	1.454	0.976	1.240	1.454	0.976

D2) Discussion:

Table 14 shows the level of accuracy with respectable values of results which can be expected when comparing theoretical with experimental values of bond lengths. There were problems in convergence when we tried to model the system for bond lengths with lower levels of theory like AM1, PM3 and HF/STO3G. These disappeared with the higher levels of theory. With all the results we can say that B3LYP is the most convenient method for transition state calculations for this model system (as bond lengths

computed by B3LYP are almost the same as experimental values) and with the higher level Hartree-Fock methods we observed a small deviation in calculated values.

E) Reaction energy profile calculation with and without copper:

This calculation is performed to observe how the energy of the system could depend upon smoothly connecting the minimum reactants ($2H^+$ and $2O_2^-$) necessary to form the observed products. Since the release of O_2 from the complex is the critical coordinate we chose to vary the H----O_2 distance to mimic this effect. Energy profile calculation without copper is performed at B3LYP/6-31G** for the dismutation reaction to verify that the activated complex in fact yields the products (O_2 and H_2O_2). For this the bond between the O-H which connects oxygen and hydrogen peroxide is constrained. We varied the bond-length from 1.37 A^0 to 3.97 A^0 and performed the optimization in 10 steps. In the calculation we selected the energy profiles with B3LYP at 6-31G** basis set. The calculation begins with an optimization of the activated complex geometry, and the geometry of the new structure is optimized at new values. It took approximately 40 hours for the calculation and gave a series of electronic energies (atomic units). Then a plot of electronic energies at the constrained distances is drawn through the 10 data points. Likewise we can calculate energy profiles at different levels. The software has an animation scheme where we can observe the animated pictures of initial state reaching transition state and finally separating into products.

E1) Results:

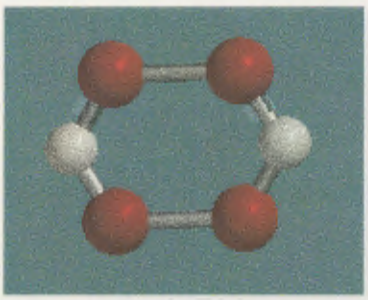
The pictures (Figure 2) below depict the structural geometries at initial transition and final states without copper. Note that the structure twists from a planar original alignment.

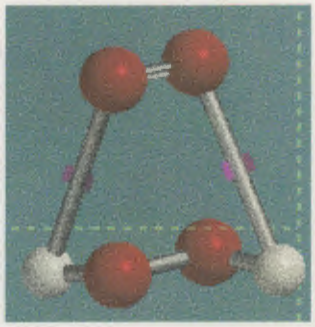
a) Optimized structure at B3LYP/6-31G**, at the initial state (reactants).

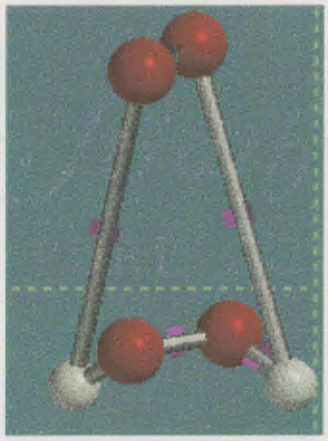
b) Optimized structure at B3LYP/6-31G**, at the Transition state.

c) Optimized structure at B3LYP/6-31G**, at the final state (products).

Figure 2. Minimal dismutation reaction illustrating hexagonal planar initial state, twisted transition state and repulsive final state.

a) 

b) 

c) 

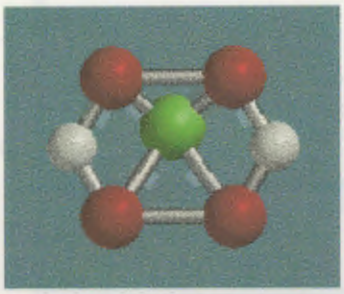
The pictures (Figure 3) below depict the structural geometries at initial, transition, and final states with copper.

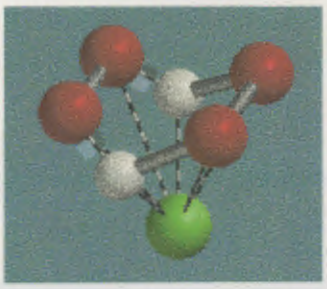
a) Optimized structure at B3LYP/6-31G**, with copper at the initial state (reactants).

b) Optimized structure at B3LYP/6-31G**, with copper at the Transition state.

c) Optimized structure at B3LYP/6-31G**, with copper at the final state (products).

Figure 3. The effect of copper ion on the minimal dismutation reaction. Both the initial hexagonal planar state and the transition state are contracted, facilitating both redox and bonding processes to form H_2O_2.

a) 

b) 

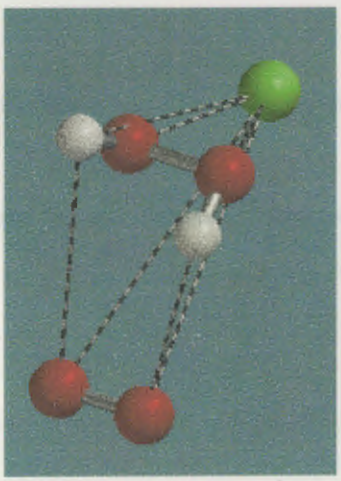
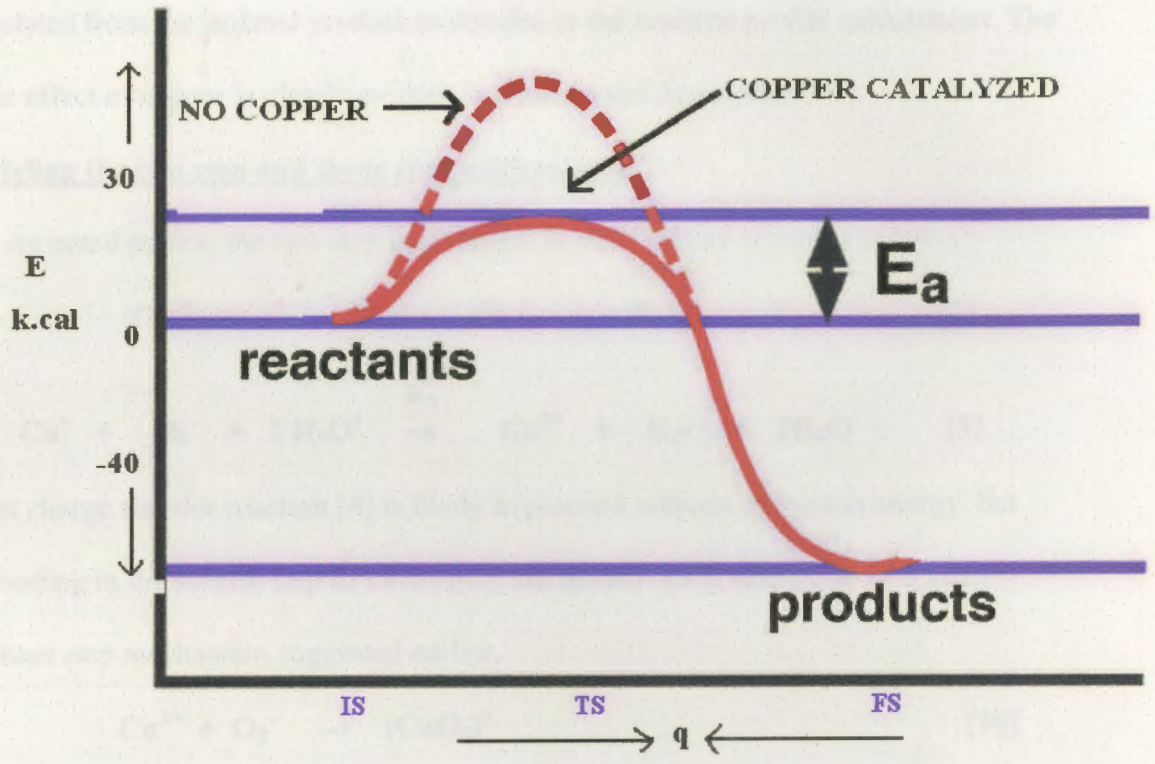
c) 

Figure 4. Energy profile of copper ion catalyzed and uncatalyzed minimal dismutation reaction ($2H^+ + 2O_2^-$). Energies for initial state (IS), transition state (TS) and final state (FS) are calculated based on isolated molecules. The reaction coordinate (q) contains all critical coordinates stepped proportionally.



E2) Discussion:

The initial structure is constructed and the bond-lengths are noted in the Tables 13 and 14. We have to calculate equilibrium geometries of this initial state, corresponding to isolated but close reactants, with success in convergence. The next step is to predict the transition state structure and its optimization by calculating transition state geometry. Another set of difficulties related to finding transition states is related to their incorporation of partial bonds. When the numbers of bonding and non-bonding electron pairs are changing, very large basis sets and significant inclusion of electron correlation

are required for accurate description. If one is attempting to use a semi-empirical MO method, one should be particularly concerned that these methods were empirically parameterized for stable molecules, and thus are much less likely to perform well in describing transition structures. After the successful convergence and optimization the bond-lengths of the transition state are noted in the table. The product bond lengths can be calculated from the isolated product molecules or the reaction profile calculations. The catalytic effect of copper is clearly evident in these model descriptions.

F) Modeling the two step and three step mechanisms:

As noted earlier, the two step mechanism is more widely accepted today.

$$Cu^{2+} + O_2^{\cdot-} \rightarrow Cu^{+} + O_2 \qquad [4]$$

$$Cu^{+} + O_2^{\cdot-} + 2\,H_3O^{+} \xrightarrow{E_a} Cu^{2+} + H_2O_2 + 2H_2O \qquad [5]$$

The first charge transfer reaction [4] is likely to proceed without activation energy, but the rebonding in the second step to form H_2O_2 has an activation energy of 17.5 kcal. In the three step mechanism suggested earlier,

$$Cu^{2+} + O_2^{\cdot-} \rightarrow (CuO_2)^{+} \qquad [10]$$

$$(CuO_2)^{+} + O_2^{\cdot-} + H_3O^{+} \xrightarrow{E_a} Cu^{2+} + O_2 + HO_2^{\cdot-} + H_2O \qquad [15]$$

the complex is likely to form without activation energy, but its fate to give peroxide has to proceed further and so does the formation of O_2, as both are simple electron transfers.

$$HO_2^{\cdot-} + 2\,H_3O^{+} \xrightarrow{E_a} H_2O_2 + 2H_2O \qquad [16]$$

We chose to examine the activation energies of the peroxide forming final steps, and the HO_2 forming second step, to see which had the lower activation energy and might therefore be preferred. Our results show the Cu^{2+} catalyzed E_a for the two step

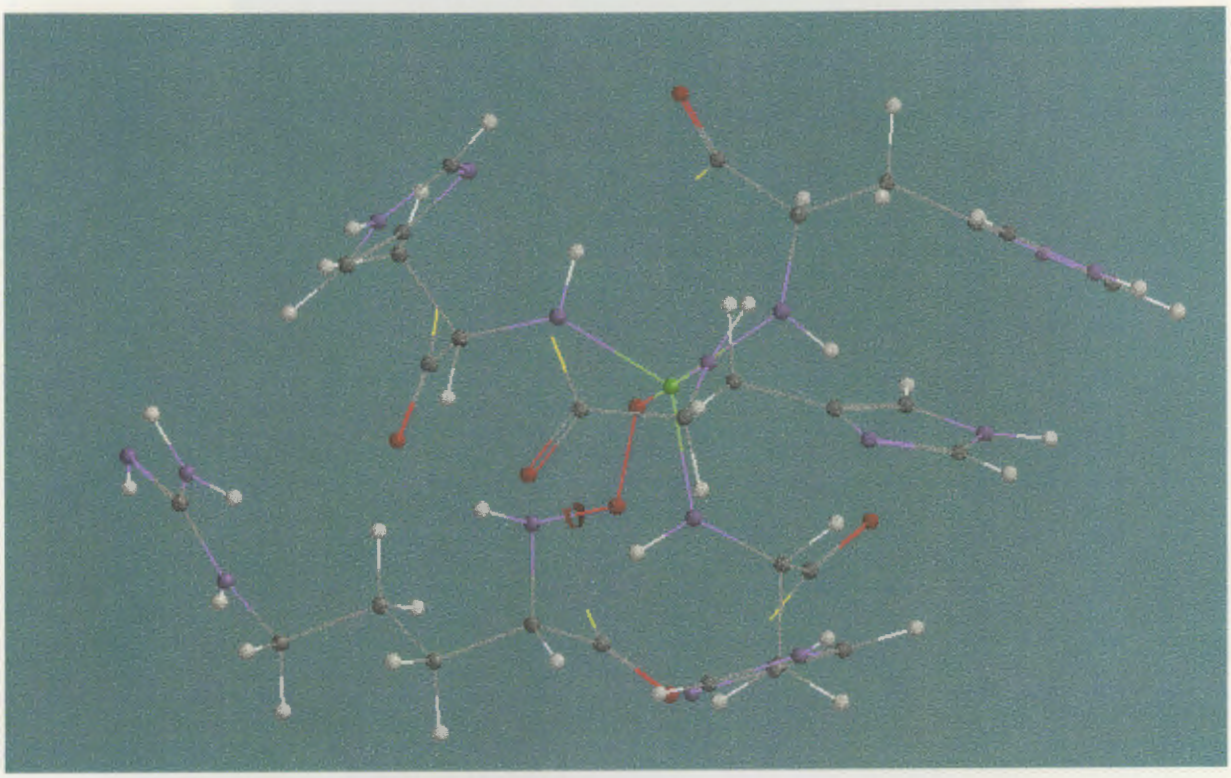
mechanism to be 17.5 kcal, while those of the bond forming steps of the three step mechanisms are 28 and 36 kcal, respectively at B3LYP/6-311+G**. It is therefore evident that the already preferred two step mechanism is more likely to occur than the three step alternative. Furthermore , if the concerted one step reaction could occur, it would be even more preferred at 6 kcal E_a.

G) Active site simulation:

At present it is not possible to perform simulations in the entire SOD (2 active sites) with density functional methods or other *ab initio* methods. Therefore we limit our considerations to one of the two independent active sites. Several simplified models are described in the past and mentioned in the literature[17]. This will reduce the effort of calculation without considerably altering the picture as the model contains the essential copper site, its four ligands, and one ammonium representing Arg. The earlier description of the active site given by Carloni et al.[7] used a special plane wave version of the density functional theory to explain the experimental bond distances called PAW. We attempted to compute these same bond distances using the optimized model structure by B3LYP/6-31G*. The particular bond distances in the imidazole structures (4) are compared.

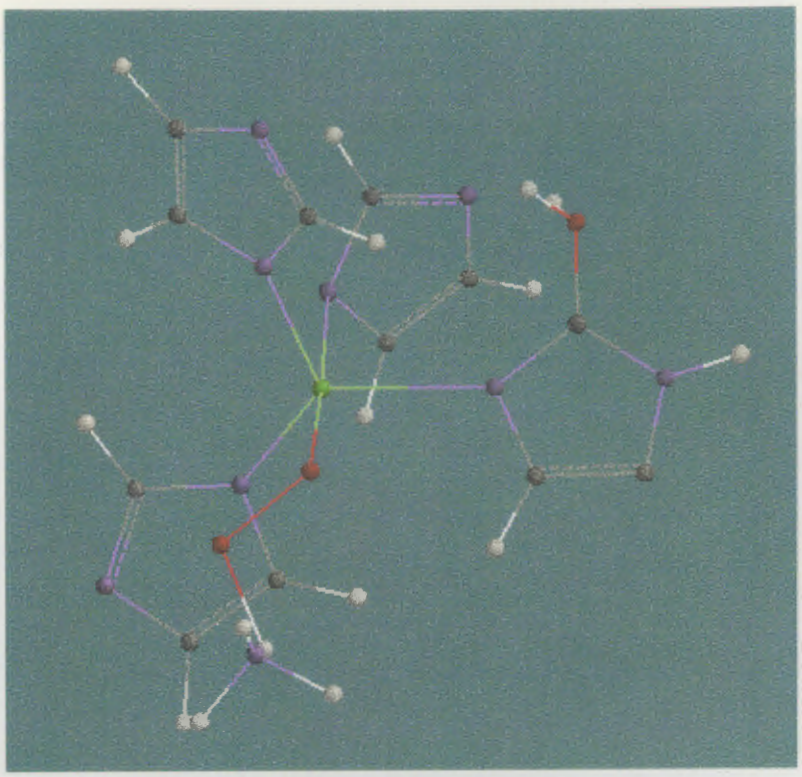
Picture below (Figure 5) is the structure of one of the active sites of SOD with four histidine groups and arginine.

Figure 5. Biological active site of SOD.



Picture below (Figure 6) is the optimized structure of simplified active site with the essential copper site, its four ligands, imidazole rings representing histidines, and one ammonium representing arginine.

Figure 6. Optimized structure of the simplified active site.



The bond distances in the constructed imidazole molecule in the optimized structure are studied. PAW is based on pseudopotential approach, which allows the use of simple plane-wave basis set which is computationally demanding for the transition-metal atoms. Carloni *et al.* calculated bond distances of imidazole molecule in a simplified active site of optimized structure in 1995 in which the calculations are performed by using the PAW program running on IBM/RS 6000 530,540 and 550 work stations[7]. The results produce decent comparison between bond distances of SPARTAN, PAW, and the crystal structure, showing that the DF theory with HF initial basis function are adequate to give realistic geometries.

Table 15. The table shows comparison between bond distances in imidazole molecule in the Optimized structure by PAW, Crystal structure, and Optimized structure by B3LYP/6-31G* in SPARTAN 2004.

	SPARTAN 2004(A^0)	PAW(A^0)	CRYSTAL STRUCTURE(A^0)
N(1)-C(2)	1.378	1.374	1.358
N(1)-C(5)	1.396	1.384	1.381
C(2)-N(3)	1.338	1.329	1.333
C(5)-N(3)	1.396	1.384	1.389
C(4)-C(5)	1.382	1.378	1.378
N(1)-H(1)	1.042	1.036	1.053
C(2)-H(2)	1.108	1.103	1.087
C(4)-H(4)	1.106	1.102	1.086
C(5)-H(5)	1.104	1.101	1.087

H) Computation times of the dismutation reaction at different levels:

Our goal here is to provide information on computation times of dismutation reaction at different levels. We first started calculating computation times for individual reactants and products at different levels of theory. Then we calculated computational times for the reactants (ground state) to reach transition state. Searching for transition states and optimum geometries for the system in question is computationally intensive for a number of reasons. First, the species to be optimized are small clusters of molecules and this often results in long distance displacements for loosely bound entities. This

greatly increases the number of optimization steps per molecule. Also, to achieve converged geometries, exact Hessian matrices are often required, which often results in much longer computation times for each step in the geometry optimization. Finally, the description of transition state in the system demands inclusion of larger molecular species which also results in longer computation times due to poor scaling of the computational methods used. The computation times that we calculated help as information for any of the future research on this reaction. It will help in comparison of computation times on this reaction with other methods which use a different program. The calculations for individual species took less than one minute for each species at any level of theory.

Table 16 shows computation **times of reactant molecules (ground state) to reach the transition state (excited state) with and without copper** at different levels of theory.

Table 16. Computation times of reactant molecules (ground state) to reach the transition state (excited state) with and without copper at different levels of theory

LEVEL OF THEORY	Computation time taken for reactants from ground state to transition state without copper.(minutes)	Computation time taken for reactants from ground state to transition state with copper.(hours)
AM1	35	—
PM3	42	—
HF-STO3G	88	—
HF-[3-21G*]	115	4:06
HF-6-31G*	120	4:25
HF-6-31G**	245	8:50
DF-B3LYP -6-31G*	380	22:25
DF-B3LYP-6-31G**	426	35:20
DF-B3LYP-6311+G**	430	40:12
MP2	445	44:15

From the tables we can say that the computation time increased with the level of theory. Simpler modes of theory took very less time where as higher levels took more time particularly with B3LYP in case of transition state calculations. The result we produced might be useful tool in exploring speed and accuracy of the SPARTAN software for this reaction in the future. The result can be used in comparison of algorithms used in different levels of theory. The future investigators can conduct a

simulation study to investigate the computation time, the accuracy (probability to compare with the experimental values), and the expected similarity of true and computed values for several widely used programs and SPARTAN.

Chapter 4 Conclusion.

In an attempt to model the active site dismutation of superoxide to oxygen and hydrogen peroxide, we investigated the relevant thermodynamics and kinetics of three mechanisms. The thermodynamic values were checked against reported calorimetric data, and except for PM3 and lower level Hartree-Fock treatments, we found agreement within experimental error. The electron correlation method MP2 and even AM1 gave the most accurate results. Using ΔG for the reaction, all the methods could be compared.

LEVEL OF THEORY	ΔG_{Rxn} (kcal/mol)
AM1	-397.26
PM3	-459.24
HF-STO3G	-515.90
HF-3-21G*	-362.82
HF-6-31G*	-382.35
HF-6-31G**	-383.14
B3LYP-6-31G*	-396.22
B3LYP-6-31G**	-395.23
B3LYP-6311+G**	-395.02
MP2	-396.65
CALORIMETRIC	-403.56

After gaining confidence that these methods were suitable for ground states, we computed transition state energies and quasi-equilibrium geometries, examining in particular the role of copper ion in reducing the activation energy.

LEVEL OF THEORY	E_a (no copper) (kcal/mol)	E_a (with copper) (kcal/mol)
AM1	50.017	No convergence
PM3	49.159	No convergence
HF-STO3G	53.337	No convergence
HF-3-21G*	35.767	11.922
HF-6-31G*	32.630	9.412
HF-6-31G**	28.865	8.785
B3LYP-6-31G*	25.100	8.157
B3LYP-6-31G**	24.472	6.275
B3LYP-6311+G**	30.747	6.275
MP2	22.59	5.647

These activation energy values are lower limits since the initial state configuration, though the same for all methods brought the oppositely charged reactants to a near hexagonal geometry. The clear catalytic role of copper is manifest.

For the concerted one step dismutation reaction, a reaction energy profile was computed as a function of the normalized bond length changes to reflect all these changes in a single reaction coordinate, from reactants to products. This was constructed using

B3LYP/6-31G** (density functional) since it gave nearly the lowest value for the activation energy. MP2 was lower due to the inclusion of excited states to account for electron correlation. The density functional adds in electron repulsions without excited states.

We also evaluated the activation energies of all three mechanisms: (i) one step, concerted; (ii) two step, no complex; (iii) three step, one complex (CuO_2^+). For the two and three step mechanisms in the literature, we modeled the bond breaking/bond formation step since the others involved only electron transfer. As a general finding we see that if all the reacting species can be assembled at the same time and place, maximum use of bond formation exothermicities can be applied to bond breaking endothermicities, thus leading to the lowest E_a and most preferable mechanism. This circumstance is a high entropy requirement however, tending to raise the activation free energy. However the entropy term is not decisive.

MECHANISM	E_a(kcal)
ONE-STEP	5.647
TWO-STEP	17.5
THREE-STEP	28 and 36

Finally we attempted to simulate the active site of SOD itself. The biological active site is simplified by using imidazole rings in place of histidine and ammonium instead of arginine to make the computation practical. We used B3LYP/6-31G* to model

this and the other methods failed to converge. We optimized the structure and compared our results with the X-ray data and PAW calculations of Carloni et al.[7] Our results gave slightly larger (+0.002 to +0.012 A^0) bond lengths than PAW for the imidazole bond distance, and they in turn were slightly larger than the crystal structure values. Perhaps this reflects the absence of compressive crystalline forces in either model.

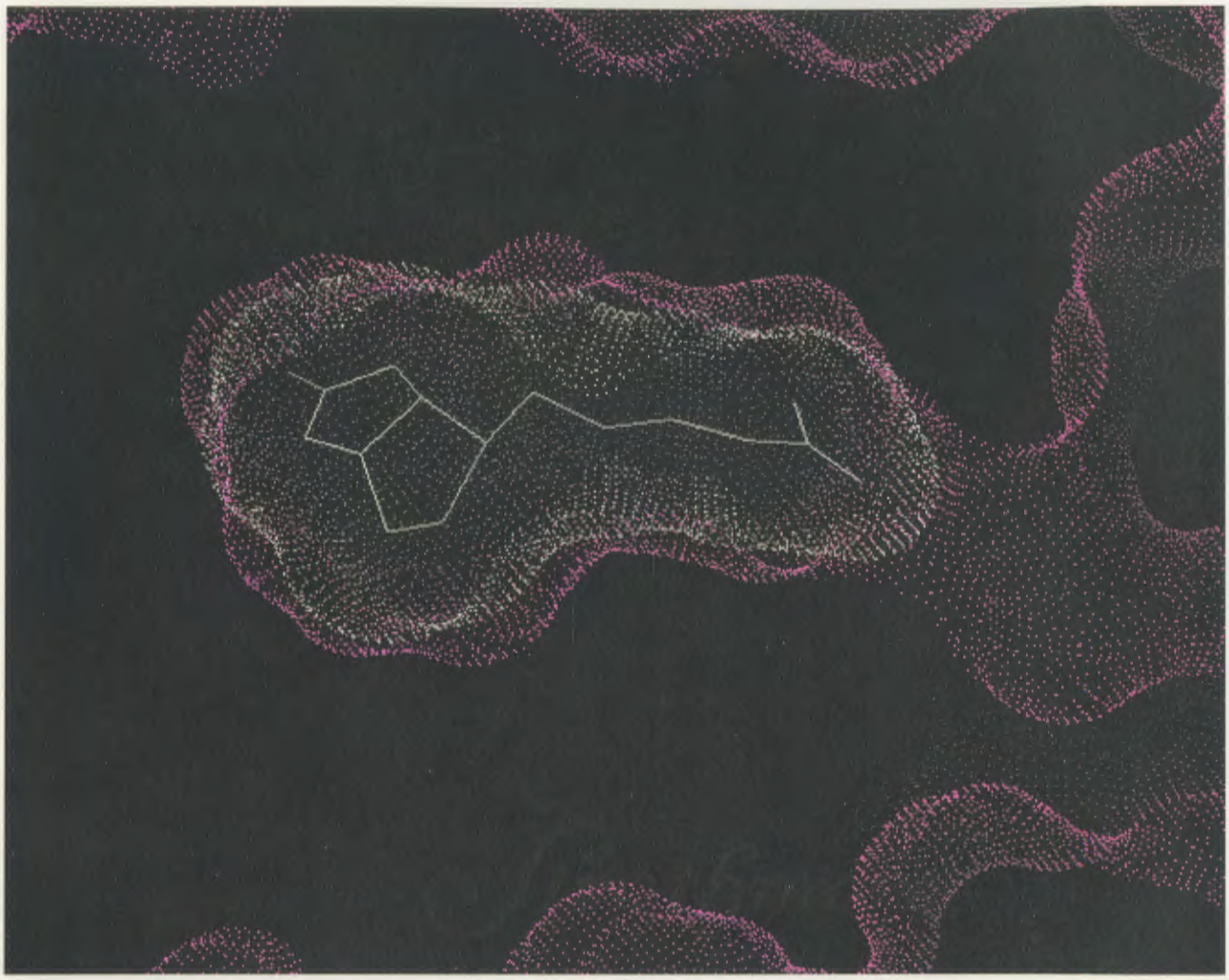
It could be added that the computation times required for various levels of theory and basis sets ranged from a few seconds to two days approximately. Accurate thermodynamic results could be obtained quickly, but those involving transition states took much longer and required a higher level of theory to converge in some cases.

Chapter 5 Future work:

Besides modeling enzyme-substrate interactions in biochemical conversions like SOD, many other applications of modeling to biomolecular systems are possible. For SOD, we should simulate the second and more complex active site in the natural system. For example the interaction between histidine, Cu^{2+} and O_2^- may differ there. Secondly, it would be valuable to base the activation parameter on an expanded initial state instead of contracted hexagonal configuration.

In cases of pharmaceutical modeling it is of interest to consider the interaction between drugs (or, ligands) and their targets (often, enzymes). Such interaction may be entirely due to non-bonded forces, but in some cases a covalent interaction may be involved. Many ligands show significant shape complementarity with the region of the macromolecule where they bind. This can be observed by constructing the molecular surfaces, which shows the molecular surfaces of ligand bound to target. For example Figure 7 [19] shows such a surface complementarity of the protein streptavidin (purple) and the ligand protein (white). The figure shows molecular surfaces of biotin bounded to streptavidin using Insight II, a sophisticated molecular modeling technique that provides powerful graphical interface algorithms for molecular dynamics, homology modeling, *de novo* design, and electrostatics—making it the suitable solution for protein modelers, computational chemists, and structural biologists.

Figure 7. Surface complementarity of the protein streptavidin (purple) and the ligand protein (white) which shows molecular surfaces of biotin bounded to streptavidin.



SPARTAN has shown itself capable of modeling various biological molecules and mechanisms, for example Vitamin E [6] which is very like SOD. Vitamin E also plays an active role in defending cells from attack by toxic oxidizing agents to give stable products that can be safely excreted. While the details of this mechanism are still uncertain, it seems likely that they might undergo hydrogen atom and electron transfer

reactions to give products, the process of which can be profiled by graphical models supported by SPARTAN. It would be useful to extend our methods to such important systems.

Bibliography:

1) Leach, A.R., "Molecular Modeling: Principles and Applications," 2nd ed., Pearson Prentice Hall, Harlow, England; **2001.**

2) Tinoco, I.; Sauer, K.; Wang, J. C. "Physical Chemistry, "2nd. ed., Prentice Hall, Englewood Cliffs, NJ; **1985.**

3) Niviere, V. and Fontecave, M. *J Biol Inorg Chem.* **2004;** *9(2),* 119-23.

4) Ciriolo, M.R.; Battistoni, A.; Falconi, M.; Filomeni, G; Rotilio, G. *Eur. J. Biochem.* **2001;** *268,* 737-742.

5) Hiroyuki, U. Assay of Enzyme Superoxide Dismutase, *Dojindo Newsletter Vol 3.* *http://www.dojindo.com/newsletter/review_vol3-3.html,* (accessed 03-20-2005).

6) Hehre, W. J; Deppmeier, B. J; Klunzinger, P. E. *A PC Spartan Pro Tutorial,* Wave function, Inc.; **1999.**

7) Carloni, P.T; Blichl, P.E; Parrinello, M. *J. Phys. Chem.* **1995;** *99,* 1338-1348.

8) Vladimir, P; Siegbahn, E. M. *Inorg. Chem.* **2005;** *44,* 3311-3320.

9) Rosi, M.; Sgamellotti, A; Tarantelli, F; Bertini, I; Luchinat, C. *Inorg. Chim. Acta,* **1985,** *107,* 21.

10) Rosi, M.; Sgamellotti, A; Tarantelli, F; Bertini, I; Luchinat, C. *Inorg. Chem.* **1986,** *25,* 1005.

11) Stewart, J. P. *J. Computational Chem.* **1989,** *10,* 209-220.

12) Hehre, W. J.; Stewart, R. F. and Pople, J. A. *J. Chem. Phys.* **1996**, *51*, 2657.

13) Hehre, W. J.; Radom, L.; Schleyer, P. V. R.; Pople, J. A. *Ab Initio Molecular Orbital theory,* John Wiley & Sons, Inc.; **1986.**

14) Hehre, W. J.; Ditchfield, R. and Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

15) Binkley, J. S.; Pople, J. A. and Hehre, W. J. *J. Chem. Soc.* **1980**, *102*, 939.

16) Hariharan, P. C. and Pople, J. A. *Chem. Phys. Lett.* **1972**, *66*, 217.

17) Parr, R. G. and Yang, W. *Density Functional Theory of Atoms and Molecules*, Oxford Univ. Press, **1989.**

18) Pederson, T. B. *www.fysik.dtu.dk/Bligaard/ms/node8.html*, (accessed 01-29-2005).

19) Leach, A. R. *Molecular modeling -Principles and Applications,* Pearson Education Limited Second edition; **2001.**

20) Worthington Biochemical Corporation. http://www.worthington-biochem.com/SODBE/default.html, (accessed 04-18-2005).

21) Wayne State University. http://chem.wayne.edu/~hbs/chm6440. (accessed 04-20-2005).

22) Chase, M.W. NIST-JANAF Thermochemical tables, *Amer.Inst.Phys. and Amer.Chem.Soc.* **1998.**

23) Laidler, K. J. and Meiser, J. H. "Physical Chemistry, "2nd. ed., Houghton Mifflin Company, Boston; **1995.**