

**QUANTUM COMPUTATIONAL DESIGN OF A PYROLIDINE-BASED ORGANOCATALYST
WITH ENHANCED SELECTIVITY AND ACTIVITY TOWARD ASYMMETRIC ALDOL
REACTION**

BY

Usman IBRAHIM

DEPARTMENT OF CHEMISTRY, FACULTY OF PHYSICAL SCIENCES

AHMADU BELLO UNIVERSITY, ZARIA

NIGERIA.

MARCH, 2019

**QUANTUM COMPUTATIONAL DESIGN OF A PYROLIDINE-BASED ORGANOCATALYST
WITH ENHANCED SELECTIVITY AND ACTIVITY TOWARD ASYMMETRIC ALDOL
REACTION**

BY

Usman IBRAHIM, BSc. (KUST) 2011

P13SCCH8001

**A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES,
AHMADU BELLO UNIVERSITY, ZARIA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF
MASTER DEGREE IN PHYSICAL CHEMISTRY.**

**DEPARTMENT OF CHEMISTRY, FACULTY OF PHYSICAL SCIENCES
AHMADU BELLO UNIVERSITY, ZARIA
NIGERIA.**

MARCH, 2019

DECLARATION

I declare that the work in this Dissertation entitled ‘Quantum Computational Design of aPyrolidine-Based OrganocatalystWith Enhanced Selectivity And Activity Toward Asymmetric Aldol Reaction’ has been carried out by me in the Department of Chemistry, Ahmadu Bello University, Zaria. The information derived from the literature has been duly acknowledged in the text and the reference list. No part of this Dissertation was previously presented for another degree or diploma at this or any other Institution.

Name of Student

Signature

Date

CERTIFICATION

This Dissertation entitled QUANTUM COMPUTATIONAL DESIGN OF A PYROLIDINE-BASED ORGANOCATALYST WITH ENHANCED SELECTIVITY AND ACTIVITY TOWARD ASYMMETRIC ALDOL REACTION by Usman IBRAHIM meets the regulations governing the award of the degree of M.Sc. Physical Chemistry of the Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

Prof. A Uzairu _____
Chairman, Supervisory Committee Signature Date

Dr. S.E Abechi _____
Member, Supervisory Committee Signature Date

Dr. A.A Siaka _____
Member, Supervisory Committee Signature Date

Prof. A. Uzairu _____
Head of Department Signature Date

Prof. S.Z Abubakar _____
Dean, School of Postgraduate Studies Signature Date

ACKNOWLEDGEMENTS

All praise is due to Allah for his intervention in all my affairs including the conduct of this research. Sincere thanks goes to the management of Abubakar Tafawa Balewa University, Bauchi for funding the research. I am grateful to our beloved leader, teacher and guide in the Physical Chemistry unit who doubles as the major supervisor of this work; the noble and humble Professor A. Uzairu. I thank him for everything he has done towards making the research a reality. Other members of the supervisory team, Dr. S.E. Abechi and Dr. A.A. Siaka have assisted greatly in improving the quality of this work. I acknowledge the role of the Head of Department and the entire staff of the Department of Chemistry, Ahmadu Bello University, Zaria in providing a conducive research environment. My senior colleague, David Arthur, deserves my gratitude for providing technical support in software operations throughout the experimental period of the research. I remain grateful to my course mates; Usman Abdulfatai, Abdullahi Bello Umar, Ibrahim Mansur Lawal and Stephen Ejeh for contributing literature materials and all the support. I received enormous support and encouragement from my family, which made it easier for me to overcome all the challenges faced during this exercise. My special gratitude goes to members of my family: my brothers and sisters; fabulous supporters and companions; my mother, Malama Maimuna, who knows well how to raise a chemist and my wife, Fatima, who knows well how to live with a chemist.

DEDICATION

This work is dedicated to my first teacher; my mother, MalamaMaimunaMuhammadu.

TABLE OF CONTENTS

	Page
Cover Page.....	i
Fly Leaf.....	ii
Title Page.....	iii
Declaration Page.....	iv
Certification Page.....	v
Acknowledgements.....	vi
Dedication.....	vii
Table of Contents.....	viii
List of Tables.....	xiii
List of Plates.....	xv
List of Figures.....	xvii
List of Schemes.....	xix
List of Abbreviations.....	xx
Abstract.....	xxii
CHAPTER ONE.....	1
1.0 INTRODUCTION.....	1
1.1 Background of the Study.....	1
1.1.1 Properties of a good catalyst.....	1
1.1.2 Classification of catalysts.....	4
1.1.3 Organocatalysis.....	6
1.1.4 Theoretical Framework.....	7
1.2 Research Problems.....	12
1.3 Justification of the Research.....	12
1.4 Research Questions.....	13
1.5 Research Hypothesis.....	13
1.6 Aim and Objectives.....	14
1.7 Scope and Limitations of the Study.....	15

CHAPTER TWO.....	16
2.0 LITERATURE REVIEW.....	16
2.1 Pyrrolidine-based Organocatalyst.....	16
2.2 Computational Chemistry.....	19
2.2.1 Methods adopted in computational chemistry.....	19
2.2.2 Density Functional Theory.....	20
2.2.3 Basis set: 6-31G*.....	22
2.2.4 Geometry optimization and descriptors calculation.....	23
2.2.5 Quantum modeling.....	24
2.3 Enantioselectivity.....	26
2.3.1 Quantitative Structure-Property Relationship on Enantioselectivity.....	27
2.4 Asymmetric Aldol reaction.....	31
2.5 Mechanism of Organocatalyzed Asymmetric Aldol reaction.....	32
2.6 Rate of Aldol reaction.....	35
CHAPTER THREE.....	36
3.0 MATERIALS AND METHODS.....	36
3.1 Materials.....	36
3.1.1 The Dataset of catalyst compounds.....	36
3.2 Methods.....	43
3.2.1 Sketching the molecular structures using Chemdraw ultra and optimization using ChemBio.....	43
3.2.2 Optimization using Spartan 14 V.1.1.0.....	46
3.2.3 Generating descriptors using PaDel-descriptor.....	63
3.2.4 Variable elimination and selection.....	63
3.2.5 Data arrangement and dataset division.....	66
3.2.6 Model development using Material studio 8.0.....	66
3.2.7 Model validation.....	74
3.2.8 Designing of new catalyst.....	75

3.2.9 Predicting the enantioselectivity of the catalyst.....	75
3.2.10 Determination of the mechanism of the reaction.....	76
3.2.11 Determination of the catalyst effect on the rate of reaction.....	82
CHAPTER FOUR.....	84
4.0 RESULTS.....	84
4.1 Model Development.....	84
4.1.1 Models.....	84
4.1.2 Comparison between actual and calculated values of ee for the training set.....	85
4.2 Model Validation.....	89
4.2.1 Internal validation.....	89
4.2.2 External validation.....	89
4.3 Catalyst Design.....	89
4.4 Optimized Structures.....	89
4.4.1 Reaction species for proline-catalyzed aldol reaction.....	89
4.4.2 Reaction species for aldol reaction catalyzed by the candidate catalyst.....	96
4.5 Graphical Models (Surfaces).....	96
4.5.1 Frontier orbitals, density maps and electrostatic potential maps for proline-catalyzed aldol reaction.....	96
4.5.2 Frontier orbitals, density maps and electrostatic potential maps for aldol reaction catalyzed by the candidate catalyst.....	105
4.6 Results of E-HOMO and E-LUMO Computations.....	105
4.6.1 Results of E-HOMO and E-LUMO computations for proline-catalyzed system.....	105
4.6.2 Results of E-HOMO and E-LUMO computations for candidate-catalyzed system.....	106
4.7 Mechanisms of the Reaction.....	116
4.7.1 Mechanism of the reaction for proline-catalyzed system.....	116
4.7.2 Mechanism of the reaction for candidate-catalyzed system.....	116

4.8 Computed Thermodynamic Parameters.....	116
4.8.1 Computed thermodynamic parameters for proline-catalyzed reaction system	116
4.8.2 Computed thermodynamic parameters for candidate-catalyzed reaction system	119
4.9 Calculated Potential Energy Barriers.....	119
4.9.1 Calculated potential energy barriers for proline-catalyzed system	119
4.9.2 Calculated potential energy barriers for candidate-catalyzed system	120
4.10 Calculated thermodynamic and kinetic parameter.....	120
4.10.1 Calculated thermodynamic and kinetic parameters for the reaction of acetone and 4-nitrobenzaldehyde catalyzed by proline.....	120
4.10.2 Calculated thermodynamic and kinetic parameters for the reaction of acetone and 4-nitrobenzaldehyde catalyzed by the candidate catalyst.....	121
CHAPTER FIVE.....	128
5.0 DISCUSSION.....	128
5.1 Model Development and Validation.....	128
5.1.1 Model development.....	128
5.1.2 Internal Validation.....	128
5.1.3 External validation.....	130
5.2 Catalyst Design	131
5.2.1 New Catalyst Candidates.....	132
5.3 Graphical Models (Surfaces).....	133
5.4 Density maps and Electrostatic potential maps	135
5.4.1 Density maps and Electrostatic potential maps for proline-catalyzed aldol reaction.....	135
5.4.2 Density maps and Electrostatic potential maps for candidate-catalyzed aldol reaction.....	136
5.5 Mechanism of the Reaction.....	137
5.5.1 Mechanism of the reaction for proline-catalyzed system.....	138
5.5.2 Mechanism of the reaction for candidate-catalyzed system.....	141

5.6 Thermodynamic Parameters	143
5.7 Potential Energy Barriers	144
5.7.1 Potential energy barriers for proline-catalyzed system.....	144
5.7.2 Potential Energy Barriers for the candidate-catalyzed system.....	145
5.8 Determination of Kinetic parameters (activation energy and rate constant)	147
5.8.1 Kinetic parameters of the proline-catalyzed reaction.....	147
5.8.2 Kinetic parameters of the candidate-catalyzed reaction.....	149
CHAPTER SIX	154
6.0 CONCLUSIONS AND RECOMMENDATIONS	154
6.1 Conclusions	154
6.2 Recommendations	155
REFERENCES	156
APPENDICES	165

List of Tables

Table 3.1	Data set of pyrrolidine-based catalyst for asymmetric aldol reaction	37
Table 4.1a	Comparison between actual and calculated values of ee for the training set using equations 1 and 2.	86
Table 4.1b	Comparison between actual and calculated values of ee for the training set using equations 3 to 5.	87
Table 4.2	Results of internal validation	90
Table 4.3	Results of external validation	91
Table 4.4	Calculated Enantiomeric Excess values for the candidate catalysts.	94
Table 4.5a	Bond Lengths of the Transition Bonds for the proline-catalyzed reaction between acetone and 4-nitrobenzaldehyde, and the Imaginary frequencies.	97
Table 4.5b	Bond Lengths of the Transition Bonds for the candidate-catalyzed reaction between acetone and 4-nitrobenzaldehyde, and the Imaginary frequencies.	99
Table 4.6	Results of E-HOMO and E-LUMO computations for reaction with proline	110
Table 4.7	Results of HOMO Center and LUMO Center computations for reaction with proline	111
Table 4.8	HOMO-LUMO energy Gaps for proline-catalyzed system.	112
Table 4.9	Results of E-HOMO and E-LUMO computations for reaction with candidate catalyst	113
Table 4.10	Results of HOMO Center and LUMO Center computations for reaction with the candidate catalyst	114
Table 4.11	HOMO-LUMO energy Gaps for the candidate-catalyzed system.	115

Table 4.12	Results of thermodynamic computations for proline-catalyzed system	122
Table 4.13	Results of thermodynamic computations for candidate-catalyzed system.	123
Table 4.14	Results of Enthalpy change of activation, Entropy change of activation, Activation Energies and Rate Constants calculations for proline-catalyzed system.	126
Table 4.15	Results of Enthalpy change of activation, Entropy change of activation, Activation Energies and Rate Constants calculations for candidate-catalyzed system.	127

List of Plates

Plate 3.1	Chemdraw Ultra 12.0 user interface showing the menu bar	44
Plate 3.2	Chemdraw Ultra 12.0 user interface showing the drawn proline structure	45
Plate 3.3	Chem3D Pro 12.0 user interface showing the imported proline structure	47
Plate 3.4	SPARTAN 14 user interface showing the menu and the tool bars	48
Plate 3.5	File menu interface showing the molecule and substituents tool kit	50
Plate 3.6	Interface showing the selected and pasted cyclopentane ring	51
Plate 3.7	Interface showing one C atom deleted from the cyclopentane ring	52
Plate 3.8	Interface showing a pyrrolidine ring formed by addition N atom	53
Plate 3.9	Interface showing selected and added Csp2 tool on the pyrrolidine ring	54
Plate 3.10	Interface showing selected and added Osp2 tool on the Csp2 carbon	55
Plate 3.11	Interface showing selected and added O fragment on the Csp2 fragment	56
Plate 3.12	Interface showing completely built proline molecule	57
Plate 3.13	Interface showing completely set semi empirical calculation	59
Plate 3.14	Display window showing the results of semi empirical calculation.	60
Plate 3.15	Calculations window showing completely set DFT calculation	61

Plate 3.16	Display window showing the DFT calculation	62
Plate 3.17	The Interface of PaDel-descriptor software	64
Plate 3.18	The Interface of PaDel-descriptor software showing imported file and calculation set up.	65
Plate 3.19	Opening window of the materials studio	68
Plate 3.20	Materials studio interface showing a blank project.	69
Plate 3.21	Materials studio interface showing a data set on the study table	70
Plate 3.22	Materials studio interface showing a correlation matrix	71
Plate 3.23	Genetic Functional Approximation window	72
Plate 3.24	Materials studio interface showing generated model equations	73
Plate 3.25	Spartan interface showing surfaces calculation	77
Plate 3.26	Spartan interface showing IR plot	81

List of Figures

Figure 1.1	Energy gap diagram between the frontier orbitals of molecules A and B	10
Figure 3.1	Energy gap diagram between the frontier orbitals of acetone and candidate catalyst	78
Figure 3.2	A typical transition state search for the reaction between acetone and proline	80
Figure 4.1	Plot of ee values of the training set calculated by the five equations as function of experimental ee values.	88
Figure 4.2	Candidate catalyst 1	92
Figure 4.3	Candidate catalyst 2	93
Figure 4.4	Optimized geometries for the reaction species of proline-catalyzed aldol reaction system	95
Figure 4.5	Optimized geometries for the reaction species of aldol reaction system catalyzed by the candidate catalyst	98
Figure 4.6a	Frontier orbitals of the reacting species for the proline-catalyzed aldol reaction	100
Figure 4.6b	Frontier orbitals of the intermediates for the proline-catalyzed aldol reaction.	101
Figure 4.7a	Density maps of LUMO and HOMO of the reacting species for the proline-catalyzed aldol reaction.	102
Figure 4.7b	Density maps of LUMO and HOMO of the intermediates for the proline-catalyzed aldol reaction	103
Figure 4.8	Electrostatic potential map of the reacting species and intermediates for the proline-catalyzed aldol reaction	104
Figure 4.9	Frontier orbitals of the intermediates involved in the candidate-catalyzed aldol reaction	107
Figure 4.10	Density maps of LUMO and HOMO of the intermediates for the candidate-catalyzed aldol reaction	108

Figure 4.11	Electrostatic potential map of the candidate catalyst, water and intermediates for the candidate-catalyzed aldol reaction	109
Figure 4.12	Energy Barrier Chart for the mechanistic steps of the proline-catalyzed reaction of acetone and 4-nitrobenzaldehyde	124
Figure 4.13	Energy Barrier Chart for the mechanistic steps of the candidate-catalyzed reaction of acetone and 4-nitrobenzaldehyde.	125

List of Schemes

Scheme 1.1	Hajos–Parrish–Eder–Sauer–Wiechert reaction	7
Scheme 1.2	Enantiomeric Reaction of methyl 2-oxocyclopentanecarboxylate	8
Scheme 2.1	A reaction between acetone and 4-nitrobenzaldehyde	30
Scheme 2.2	Reported mechanism of asymmetric aldol reaction of acetone and 4-nitrobenzaldehyde catalysed by proline	31
Scheme 4.1	Reaction mechanism of proline-catalyzed system showing (a) enamine formation step and (b) C-C bond formation step	117
Scheme 4.2	Reaction mechanism of candidate-catalyzed system showing (a) enamine formation step and (b) C-C bond formation step	118
Scheme 5.1	The elementary steps in the mechanism of proline-catalyzed aldol reaction of acetone and 4-nitrobenzaldehyde	147
Scheme 5.2	The elementary steps in the mechanism of the candidate-catalyzed aldol reaction of acetone and 4-nitrobenzaldehyde	149

List of Abbreviations

6-31G* 6-Gaussians Pople's split-valence double-zeta with d polarization.

Å Angström

Au Atomic unit

B3LYP Beck three parameter Lee Yang, Parr

CSV Comma-Separated Value

DFT Density Functional Theory

DMF Dimethylformamide

DMSO Dimethylsulfoxide

E_a Activation energy

E_e Electronic energy

ee Enantiomeric Excess

eV Electron volt

GFA Genetic Function Approximation

GTO Gaussian Type Orbital

h Planck's constant

HOMO Highest Occupied Molecular Orbital

I Intermediate

IR Infra-Red

k Rate constant

k_B Boltzmann constant

LUMO Lowest Unoccupied Molecular Orbital

P Product

QSPR Quantitative Structure-Property Relationship

SDF Structure-Data File

SSA	Steady State Approximation
SSE	Sum of squares of errors
STO	Slater Type Orbital
T	Temperature
TS	Transition state
ΔG^\ddagger	Gibbs free energy of activation
ΔH^\ddagger	Enthalpy of activation
ΔS^\ddagger	Entropy of activation

ABSTRACT

In this study, a data set of 34 pyrrolidine-based organocatalysts applicable on asymmetric aldol reaction were considered. The data set was subjected to Density Functional Theory(DFT) optimization using BLY3P/6-31G* model and Quantitative Structure-Property Relationship(QSPR) calculation. From the optimization and calculation, several electronic, topological and stereochemical-based descriptors were generated which led to the development of five regression models. The best model (Model one), with $R^2=0.84$ and $R_{\text{ext}}=0.88$, was used to adjust the proline structure for finding new catalyst candidates

A simulation reaction was carried out in which the candidate catalyst was applied over asymmetric aldol reaction between acetone and 4-nitrobenzaldehyde in an acetone medium. A plausible mechanism of the reaction was developed using HOMO-LUMO energies calculated using density functional theory at 6-31G*/B3LYP level of theory. From the itemized elementary steps, the fourth step was identified as the rate-determining step, with the highest activation energy of 54.20 kJmol^{-1} . The mechanism was used to derive the rate law from which the overall rate constant was calculated and found to be $0.787 \text{ M}^{-1} \text{ s}^{-1}$. The simulation reaction was also conducted on the same reaction catalyzed by proline and the mechanism was developed. New mechanistic steps that followed the previously reported iminium-enamine route of typical class 1 aldolase enzymes were proposed in greater detail. From the elementary steps, the first step which involves a bimolecular collision of acetone and proline was considered as the rate-determining step, having the highest activation energy of 59.07 kJmol^{-1} . The mechanism was used to develop the rate law from which the overall rate constant was calculated and found to be $4.04 \times 10^{-8} \text{ dm}^3 \text{ Mol}^{-1} \text{ s}^{-1}$. The new mechanistic insights and the explicit computation of the rate constant further improved the kinetic knowledge of the reaction.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background of the Study

It is conventional to define a catalyst as a substance which increases the rate of a chemical reaction without altering the thermodynamic equilibrium of the reaction and without its self being consumed in the process. It is therefore important to emphasise that a catalyst can never change for the better, the fate of a reaction that is thermodynamically unfavourable. Furthermore, a catalyst is a substance which accelerates the rate at which chemical bonds are formed or broken by lowering the activation energy in a reaction, thereby, improving the product yield and increasing the amount of desirable products while decreasing the amount of undesirable ones. Therefore, a catalyst is understood as a substance which alters the kinetics of the reaction for better yield without changing the thermodynamics.

Chemical catalysis affects our lives in many ways, in the sense that catalysts are involved in most of the reactions taking place in food, chemical, pharmaceutical, petroleum and petrochemical industries. Catalysts are the workhorses of chemical transformations in the industry. Approximately 85–90 % of the products of chemical industry are made in catalytic processes (Lloyd, 2011). Catalysts are indispensable for their role in prevention of pollution, by avoiding formation of waste products in catalytic reactions and suspension of pollutants in end-of-pipe solutions of automotive and industrial exhaust.

Summarily, a catalyst offers an alternative, energetically favourable mechanism to the non-catalytic reaction, thus enabling processes to be carried out under safe and feasible conditions of pressure and temperature in industry.

1.1.1 Properties of a good catalyst

The following properties describe the viability and affectivity of a catalyst:

1.1.1.1 Activity

The activity of a catalyst is regarded as the rate of a particular reaction over a given catalyst volume. This can be related to the turnover number, or frequency, which is the number of molecules of product produced by each active site per unit time under standard conditions.

Because it is not practically possible to calculate the active sites on a commercial catalyst, the space-time yield is measured instead. The space-time yield is the quantity of product produced per unit volume of catalyst per unit time, during industrial operations (Lloyd, 2011).

1.1.1.2 Selectivity

Selectivity of a catalyst is defined as the proportion of desirable product obtained from the total amount of reactants converted. Thus, it is possible to have a catalyst with almost 100% selectivity but economically less desirable if its activity is very low. However, selectivity in a catalyst is equally important because low-selectivity catalysts are uneconomic, not only because feed is wasted and by-products have to be separated from products, but also because side reactions are often more exothermic, making the reactor design difficult. Many processes operate at less than 100% conversion to limit heat evolution or to achieve higher selectivity (Lloyd, 2011). In these instances, the unconverted feed must be recycled and conversion per cycle can still be relatively low, but economically appreciable. Selectivity in enantiomeric reactions is termed 'Enantioselectivity'. Enantioselective reactions are the result of the competition between different possible diastereomeric reaction pathways, through diastereomeric transition states, when the prochiral substrate being complexed to the chiral catalyst reacts with the corresponding reagent. The efficiency of the chirality transfer depends on electronic and steric factors and is measured as 'enantiomeric excess' (ee), which is computed using equation 1.1 (Xuet *al.*, 2012).

$$(\% ee) = \frac{[R] - [S]}{[R] + [S]} \times 100 \quad (1.1)$$

Where $[R]$ is the concentration of R isomer in the product formed and $[S]$ is the concentration of S isomer in the product. A simple calculation shows that difference in energy of only 2 kcal/mol between the transition states is enough to obtain more than 90% ee, and small changes in any of the participants in the catalytic process can modify this difference in energy to a significant extent (Lloyd, 2011). Those modifications may occur in the near environment of the catalytic centre, at less than 1 nm scale, but also at longer distances in the catalyst, substrate, reagent, solvent, or support in the case of immobilized catalysts. This is the case because asymmetric catalysis can be considered an anometric phenomenon that requires a careful control of different variables.

1.1.1.3 Stability

Catalysts may lose their activity and selectivity over their operational lifetime before eventually needing replacement. However, certain aspects of malfunction can cause premature damage to a catalyst, leading to premature replacement. This can be caused by overheating the catalyst thereby decreasing the surface area or fusing the active sites, losing a volatile component at high operating temperature or presence of poisons in the feed that can deactivate the catalyst. Therefore, efforts must be made to obtain a more stable alternative or to prevent deactivation by modifying the operation. The performance of most catalysts can deteriorate during relatively short periods of malfunction, so the expected performance should be checked at regular intervals. By recording operating details such as feed and effluent composition and temperature profiles in the catalyst bed, it is possible to assess abnormal operating features or feed purity. Some catalysts can be regenerated in situ while activity can be restored in others by identifying temporary poisons in the feed and removing them.

1.1.1.4 Strength

The catalyst must be strong enough to resist various forms of regeneration or reactivation. Some of the situations that may want to destroy the catalyst during plant operation include: dust

forming from catalyst disintegration, cementing of the catalyst by entrained liquids, blockage of the catalyst bed by entrained solids and formation of carbon from organic feedstock. These problems generally affect the temperature profiles in the bed and, possibly, the overall reaction.

1.1.2 Classification of catalysts

1.1.2.1 Heterogeneous catalysts

In situations where the catalysts are in different phases from the reactants, the catalysts are referred to as heterogeneous catalysts. Most heterogeneous catalysts are solids acting upon liquid or gaseous reaction mixtures (Lloyd, 2011; Ye *et al*, 2017). A heterogeneous catalyst has active sites, which are the atoms or crystal faces where the reaction actually occurs. Depending on the mechanism, the active site may be either a planar exposed metal surface, a crystal edge with imperfect metal valence or a complicated combination of the two. Thus, not only most of the volume, but also most of the surface of a heterogeneous catalyst may be catalytically inactive. Finding out the nature of the active site requires technically challenging research. Thus, empirical research for finding out new metal combinations for catalysis continues. The total surface area of solid catalyst has an important effect on the reaction rate: The smaller the catalyst particle size, the larger the surface area for a given mass of particles.

Heterogeneous catalysts typically require to be dispersed on a second material that enhances the effectiveness or minimizes their cost. These second materials are termed 'supports'. Supports are porous materials with a high surface area. Alumina, zeolites or various kinds of activated carbon are most commonly used as supports. Specialized supports include silicon dioxide, titanium dioxide, calcium carbonate, and barium sulphate.

1.1.2.2 Homogeneous catalysts

Homogeneous catalysts function mechanistically like heterogeneous catalysts, only that they operate in the same phase as the reactants. Application of homogeneous catalysts calls for the use of reaction medium in which both the catalyst and the reactants are dissolved. In inorganic chemistry, homogeneous catalysis is often synonymous with organometallic catalysts (Elschenbroich, 2006).

1.1.2.3 Electrocatalysts

Electrocatalysts, mostly metals, are used in electrochemistry, to enhance the rates of the half reactions that form the fuel cell. One common type of fuel cell electrocatalyst is based upon nanoparticles of platinum that are supported on slightly larger carbon particles. When in contact with one of the electrodes in a fuel cell, this platinum increases the rate of oxygen reduction either to water, hydroxide or hydrogen peroxide (Liu *et al.*, 2016).

1.1.2.4 Enzymes and Biocatalysts

Enzymes are protein-based catalysts that function in metabolic and catabolic reactions. The term biocatalysts refer to all biomolecules, both enzymes and other non-protein based molecules that function in biological reactions (Nelson and Cox, 2000).

Biocatalysts can be regarded as intermediate between homogeneous and heterogeneous catalysts; most certainly, soluble enzymes are homogeneous catalysts while membrane-bound enzymes are heterogeneous.

1.1.2.5 Organocatalysts

Organocatalysts are small organic molecules without metals which can exhibit catalytic properties. This is the class of catalyst that is studied in this work.

1.1.3 Organocatalysis

The term organocatalysis is used to describe a form of catalysis in which the rate of chemical process is increased by an organic catalyst referred to as "organocatalyst" (*Berkessel and Groeger, 2005*) 'Organocatalyst' is the term invented to distinguish catalytically active small organic molecules from enzymes with which they have comparable effects on reaction rates (Ostwald, 1900). Organocatalysis is important to green chemistry for the fact that it is metal-free catalysis (*Dziedzic et al., 2006*).

Organocatalysts with secondary amine functional group can perform either Enamine catalysis; by forming catalytic quantities of an active enamine nucleophile, or Iminium catalysis; by forming catalytic quantities of an activated iminium electrophile. If the organocatalyst is chiral, it facilitates asymmetric catalysis; for example, the use of proline in aldol reactions is an example of chirality (*Hajos and Parrish, 1974*.)

Regular achiral organocatalysts are nitrogen-based molecules such as piperidine, used in the Knoevenagel condensation (*List, 2010*), 4-Dimethylaminopyridine (DMAP) used in esterification (*Dhimitruka and SantaLucia, 2006*) and 1,4-Diazabicyclo[2.2.2]octane (DABCO) used in the Baylis-Hillman reaction (*Park et al., 2007*).

These catalysts and reactions have a long history but current interest in organocatalysis is on asymmetric catalysis with chiral catalysts, called asymmetric organocatalysis or enantioselective organocatalysis. A pioneering reaction developed in the 1970s is named after the inventors as the Hajos–Parrish–Eder–Sauer–Wiechert reaction, which is an intramolecular Aldol reaction depicted in scheme 1.1.

Scheme 1.1: Hajos–Parrish–Eder–Sauer–Wiechert reaction

In this reaction, naturally occurring chiral proline is the chiral catalyst in an Aldol reaction. The starting material is an achiral triketone and it requires only 3% of proline to obtain the reaction product, a ketol in 93% enantiomeric excess. This is the first example of an amino acid-catalysed asymmetric aldol reaction (Hajos and Parrish, 1971 and 1974).

This catalyst works by forming an iminium ion with carbonyl groups of α,β -unsaturated aldehydes (enals) and enones in a rapid chemical equilibrium. This iminium activation is similar to the activation of carbonyl groups by a Lewis acid and both catalysts lower the substrate's LUMO (*Lelais and MacMillan, 2006*)

The transient iminium intermediate is chiral which is transferred to the reaction product via chiral induction. The catalysts have been used in Diels-Alder reactions, Michael additions, Friedel-Crafts alkylations, transfer hydrogenations and epoxidations (Trost *et al.*, 2016; El-Bakouriet *al.*, 2018)

1.1.4 Theoretical Framework

One of the most important properties that determine catalyst action is selectivity, which is termed ‘enantioselectivity’ in enantiomeric reactions (Furuta *et al.*, 2016). Enantioselectivity of a catalyst is therefore its ability to influence production of one enantiomer in greater yield than the other when there is competition between different possible enantiomeric reaction

pathways (Lloyd, 2011). Enantioselective capacity of a catalyst is a key factor in determining the product in asymmetric reaction. Scheme 1.2 shows how a given catalyst influences production of a desired isomer (*anti*) over the non-desired isomer (*syn*) in an intramolecular aldol reaction of methyl 2-oxocyclopentanecarboxylate. The reaction gives syn/anti ratio of 1:99, resulting in an enantioselectivity of 95% ee (Helmchen *et al.*, 1995).

Scheme 1.2: Enantiomeric Reaction of methyl 2-oxocyclopentanecarboxylate

a: methyl 2-oxocyclopentanecarboxylate

b: (1R,2R)-methyl 2-hydroxycyclopentanecarboxylate

c: (1S,2R)-methyl 2-hydroxycyclopentanecarboxylate

One of the techniques that can be used to design a new catalyst and determine its selectivity is Quantitative-Structure-Property-Relationship (QSPR). This is a technique that uses the information extracted from the structural properties of a set of compounds to develop a mathematical model that can predict some desired properties of any other compound that falls within the applicability domain of the model (Yoshino, 2010).

Another important determining factor of the catalyst action is the activity. Activity of a catalyst is regarded as the rate of a particular reaction over a given catalyst volume (Lloyd, 2011). Activity of the catalyst has direct effect on the turnover of product, or frequency, which is the number of molecules of product produced by each active site per unit time under standard conditions. Activity of the catalyst, which is equivalent to the reaction rate constant, can be

determined after deducing the reaction mechanism of the catalyst and deriving the rate law of the reaction.

To develop the reaction mechanism, the idea of using frontier orbitals to lead the reaction pathway proves to be very effective. Using quantum mechanical descriptors of E-HOMO and E-LUMO to propose a reaction mechanism is a technique that proves to be valuable. This minimizes trial and error in identification of mechanistic steps of the reaction. More so, this allows linking the quantum mechanical parameters to the reaction rate. This is in the sense that, the rate constant depends on the thermodynamic parameters (enthalpy of activation and entropy of activation) of the elementary steps which can only be itemized after identifying the reaction intermediates. The intermediate structures are built on the basis of the HOMO-LUMO gaps of the combining molecules. Figure 1.1 shows the energy gaps between the frontier orbitals of two interacting molecules; A and B (Siaka, 2017).

Fig. 1.1: Energy gap diagram between the frontier orbitals of molecules A and B.

Central to the computation of the mechanism is the determination of transition state structures. Unlike stable molecules, the transition state structures do not exist practically. Therefore, they cannot be measured experimentally (Ges'u *et al.*, 2016; Bhoorasingh and West, 2015). However, measured activation energies can help to determine the transition state energies relative to reactants

This is achieved by transition state theory which assumes that all molecules proceed over a single transition state towards product formation (Ges'u *et al.*, 2016). Like reactants, stable intermediates and products, the transition states correspond to well-defined structures which can be found and characterized by theoretical calculations almost as routinely as finding equilibrium geometry (Jacobson *et al.*, 2017). Having found the transition state geometry, it needs to be tested to verify that it is actually the transition route, among many others, with the minimum energy. The first test requires carrying out normal mode analysis on the proposed transition state geometry to verify that the Hessian gives one and only one imaginary frequency in the range of 400-2000 cm^{-1} (Maeda *et al.*, 2015). Secondly, the normal coordinate corresponding to imaginary frequency must smoothly connect reactants and products (Maeda *et al.*, 2015). That is, we should be able to walk along this coordinate without any additional optimization. The extent to which transition states incorporate partially or completely broken bonds may prompt anticipating that very simple theoretical models cannot satisfactorily describe the transition state structures.

The rate law can be derived after identifying all the intermediates involved in the reaction mechanism and determining the rate-limiting step. The rate-limiting step, also known as the rate-determining step, is the slowest step of a reaction mechanism. The derivation of the rate law is achieved using the approximation known as Steady State Approximation (SSA). This approximation states that when a short-lived reaction intermediate, C (equation 1.2), exists at

low and constant concentration throughout the course of reaction, the rate of formation of the intermediate is negligibly small.



$$\text{Therefore, } \frac{d[C]}{dt} = 0 \quad (1.3)$$

1.2 Research Problem

The discovery of truly efficient methods of obtaining chiral substances is a substantial challenge for synthetic chemists and it is apparently important in biological reactions to obtain single-enantiomer product than racemic mixture. Until recently, catalyst design relied on trial-and-error and the experience of individuals, but it is becoming more exact science nowadays.

Although intermediate-level computations can provide insight into how a catalyst works, the true computational design of practical catalysts for industrial and commercial applications will require accurate determination of thermodynamic and kinetic results, since a factor of 2 to 4 in catalyst efficiency can determine the economic feasibility of a process (Lloyd, 2011). This research targets at providing an efficient method for catalyst design via calculating thermodynamic and kinetic quantities with very high accuracy.

1.3 Justification of the Research

In QSPR approach, it is easy to extract what is important and easy to search a good catalyst. However, this is not based on reaction mechanism, rather, only on statistical estimations. On the other hand, *ab Initio* calculations of transition states is based on reaction mechanism but it is difficult to predict what is important and to search for a good catalyst without intuition. It is therefore wise to combine the two methods in order to overcome the deficiencies in both.

More so, most of the previous works on geometry of transition states relied on guessed transition state, in the sense that the starting structure would be arbitrarily chosen. This work

adopts a more specific method by using quantum mechanical descriptors of HOMO and LUMO to propose the transition states, thereby, minimizing trial and error in identification of mechanistic steps of the reaction.

1.4 Research Questions

- i. How can the enantioselectivity of pyrrolidine based catalysts be predicted using QSPR approach?
- ii. Can a new catalyst be designed using the QSPR model?
- iii. How can the newly designed catalyst be used to carry out *ab initio* transition state search for a selected asymmetric aldol reaction?
- iv. Is the new catalyst an improvement over the currently used catalysts like proline?

1.5 Research Hypothesis

The scheme of the study was based on the following null and alternative hypothesis:

- I. Null hypothesis:** The enantioselectivities of the catalysts are correlated with their molecular structures and QSPR technique can be used to design a viable catalyst with improved activity and selectivity over proline for asymmetric Aldol reaction
- II. Alternative hypothesis:** The enantioselectivities of catalysts are not correlated with their molecular structures and QSPR technique cannot be used to design a viable catalyst with improved activity and selectivity over proline for asymmetric Aldol reaction.

1.6 Aim and Objectives

This research aims at designing a new organocatalyst with improved enantioselectivity and determining its mechanism of action and effect on the reaction rate of a selected aldol reaction, via quantum computational method.

This aim will be achieved through the following objectives:

- i. Collecting a data set of the pyrrolidine based molecules with established catalytic effect and recorded enantiomeric excess (ee) from previous works and performing quantum mechanical calculations on the molecules of the data set to optimize their structures and compute several molecular descriptors using the softwares: Spartan “14”, PaDel and Chem Pro.
- ii. Establishing a correlation between the descriptors and the enantioselectivity of the compounds and developing a QSPR model of robust predictability using Genetic Function Approximation (GFA) method.
- iii. Designing a new catalyst by adjusting the structure of proline holistically, being guided by the model, to achieve an analogue and predicting the enantioselectivity of the proposed catalyst using the model.
- iv. Performing *in silico* aldol reaction involving the proposed catalyst to obtain the HOMO, LUMO and thermodynamic parameters of all the steps involved in the mechanism using Spartan 14 and repeating same with proline as the catalyst.
- v. Using the HOMO-LUMO gap to suggest the mechanisms of the reactions and using the thermodynamic parameters to calculate the rate constants of all the consecutive reactions by applying Eyring equation and subsequently computing the overall rate constants, by hand.

1.7 Scope and Limitations of the Study

This research is aimed at developing a QSPR model that is capable of predicting the enantioselectivity of only pyrrolidine-based organocatalysts that fall within the applicability domain of the model. It involves proposing a viable catalyst with known enantioselectivity and activity values in form of enantiomeric excess and rate constant, respectively, for an aldol reaction between acetone and 4-nitrobenzaldehyde.

The process of transition state search is excessively time-consuming; therefore, the study could not be extended to other forms of aldol reaction due to time constraint.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Pyrrolidine-based Organocatalyst

Pyrrolidine belongs to the class of heterocyclic amines having a saturated five-membered ring with the molecular formula, C_4H_9N . Many derivatives of pyrrolidine, like prolines and prolinamides, have been proven to have catalytic effect. Proline is a secondary, cyclic, pyrrolidine-base, α -amino acid. A unique advantage of these characteristics, different from other amino acids, is the improved pKa value of the amine group compared to primary amino acids. Another advantage is the proline's effective aminocatalysis- a Lewis-base-type catalyst that alleviates iminium-enamine-based transformations. This unique nucleophilic activity is due to the presence of the pyrrolidine part, which reacts with carbonyl compounds to form iminium ions and enamine more readily than most other amines. More so, the carboxylate part acts as a general Brønsted cocatalyst, and hence, promotes the catalytic activity. Being bifunctional, with an acid group and an amine group, proline can act both as an acid or base and also behave in a similar way as enzymes, in the sense that enzymes typically use several different functional groups in their catalytic machinery. This also makes it to act as a bidentate ligand that can form catalytically active metal complexes. Like other classes of amino acids, proline is very cheap and abundant chiral molecule, available in both enantiomeric forms.

Proline was shown to possess catalytic activity as well as enantioselectivity (List *et al.*, 2000; De *et al.*, 2017) on asymmetric aldol reaction (Kutzscher *et al.*, 2016) in an extensive number of researches that include the pioneering work of Hajos-Parrish-Eder-Sauer-Wiechert (1974). It is an effective catalyst for several other asymmetric transformations, such as Mannich and Michael additions (Trost *et al.*, 2016; Jin *et al.*, 2016; Klalwani and Sudaai, 2016).

The work of Enders and Grondal, published in 2005, reported catalytic action of proline upon aldol reaction with up to 97% yield and selectivity in dimethylformamide (DMF). Northrup

and MacMillan(2002) worked with L-proline in DMF and recorded an enantioselectivity above 99% and a yield of above 88%. Repeating the work in dimethylsulfoxide (DMSO) medium, the enantioselectivity fell to 97% enantiomeric excess (ee) and the yield rose to 97% (Northrup and MacMillan, 2004). D-proline also has catalytic ability as reported in the later work in which it was used to achieve an ee value of 98% in DMSO medium (Mangion and MacMillan, 2005).

Hayashi *et al.*(2004) found out that trans-4-tert-Butyldimethylsiloxy-L-proline, a derivative of proline, displayed a better catalytic activity than the proline its self. The reactivity of the catalyst was investigated in the α -aminooxylation of cyclohexanone in the presence of 30 mol% of the catalyst. Proline scarcely promoted the reaction in CH₂Cl₂ and tetrahydrofuran (THF) solvents, while trans-4-tert-Butyldimethylsiloxy-L-proline gave favourable results in the same solvents. This was attributed to the poor solubility of proline in a variety of organic solvents compared to the trans-4-tert-Butyldimethylsiloxy-L-proline. Moreover, while disappearance of nitrosobenzene over 30 mol % proline occurred in 60 min, it only took the same reaction over 30 mol % trans-4-tert-Butyldimethylsiloxy-L-proline, 1 min to be completed. The catalyst loading can also be reduced in the case of trans-4-tert-Butyldimethylsiloxy-L-proline. As low as 10 mol% of trans-4-tert-Butyldimethylsiloxy-L-proline was enough to finish the reaction in 15 min, while 5.5 h was necessary to complete the reaction with the same catalyst loading in the case of proline.

L-Prolinamide was initially observed to be ineffective in catalyzing the direct aldol reaction (Sakthivel *et al.*, 2001). However, it was found that in the presence of 20 mol% L-prolinamide, the model reaction of 4-nitrobenzaldehyde with neat acetone proceeded with 80% yield and 30% ee. This encouraged further investigation on the catalytic activity of a family of L-prolinamides. In particular, the enantioselectivity increases as the aryl substituent varies from electron-donating to electron-withdrawing, which makes the amide group more acidic and thus a better hydrogen bond donor.

In the work of Tang *et al.* (2005), L-Prolinamides were found to be active catalysts for the direct aldol reaction of 4-nitrobenzaldehyde with acetone at room temperature. Moderate enantioselectivities of up to 46% ee were achieved. The enantioselectivity increases as the amide group becomes a better hydrogen bond donor. L-Prolinamides, derived from the reaction of L-proline with α -hydroxyamines such that there was a terminal hydroxyl group, showed more efficient catalysis and higher enantioselectivities. In particular, catalysts, prepared from L-proline and (1*S*,2*S*)-diphenyl-2-aminoethanol, exhibits high enantioselectivities of up to 93% ee for aromatic aldehydes and up to 99% ee for aliphatic.

Furthermore, prolinamides were also found to possess catalytic activity in other works including the one in which large number of secondary *N*-arylprolinamides with increasing amine acidity and steric crowding were synthesized and tested for catalytic activity upon enantioselective aldol reactions. It was found that in DMF containing 10 mol-% of TFA, all catalysts with moderate to high amine acidity catalyzed the reaction between cyclohexanone and a variety of electrophilic aldehydes leading to aldols in excess of 90% yield and 95% enantioselectivity (Moorthy and Saha, 2009).

In 2011, Kokotos *et al.* published their work in which they proposed a novel pyrrolidine-based organocatalyst. The compound, combining the pyrrolidine with thioxotetrahydropyrimidinone was found to be a highly effective catalyst on asymmetric Michael addition. Low catalyst loading of 1-2.5% resulted in high yield and excellent stereoselectivities in reaction between cyclic ketone and nitroolefins (Kokotos *et al.*, 2011).

2.2 Computational Chemistry

Computational chemistry is a field in chemistry under which theoretical chemistry methods are applied in solving chemical problems through computer simulation. It incorporates various

levels of theory developed to calculate the structures and properties of compounds into expeditious computer programs, thus, making the quantum calculation on many-electron system an easy task(Lewers, 2016). Therefore, computational chemistry is a way of, more or less, mimicking what takes place in the laboratory (*in vitro* experiments) by the use of computer (*in silico* experiment). Although, computational results are normally complementary to the findings of laboratory experiments, it can, in many cases, predict as yet undiscovered chemical phenomena. This is widely seen in the design of new ligands (e.g: drugs) and catalysts.

Looking at the complexity and length of quantum mechanical calculations of many-electron system, it is expected that high speed and large memory computers are generally needed for this task. However, two factors affect the cost and computational demands of a calculation; the size of the molecules and the method chosen. In all cases, the computation time and memory increase rapidly with the size of the system being studied. More so, computational chemistry methods range from highly accurate to very approximate; highly accurate methods of course have more computational demands than the approximate ones. Some of these methods are briefly highlighted below.

2.2.1 Methods adopted in computational chemistry

Some of the methods adopted are *Ab initio*; they are based on the first principle, relying entirely on the fundament of quantum mechanics and primary physical constants. Other methods are called Empirical or Semi-Empirical, in the sense that they include additional experimental parameters upon the quantum mechanics calculations(Lewers, 2016).

Both *ab initio* and semi-empirical approaches make certain approximations. The *ab initio* calculations adopts the Born-Oppenheimer approximation, which greatly simplifies the underlying Schrödinger equation by assuming that the nuclei of atoms remain fixed in place

during the calculation, while the electrons revolve round the nuclei. On the other hand, semi-empirical calculations assume many of the energy integrals to be zero.

In some methods like Molecular Mechanics, the details of electronic structure are less important than the long-time phase space behaviour of molecules (Lewers, 2016). This is considered in conformational studies of proteins and protein-ligand binding thermodynamics. In this case, classical approximations to the potential energy surface are employed, as they are computationally less intensive than electronic calculations, to enable longer simulations of molecular dynamics. Furthermore, chemoinformatics uses even more empirical and computationally cheaper methods like machine learning based on physicochemical properties. The method used in this study is an *ab initio* method called Density Functional Theory (DFT)

Based on the concept of molecular theory, a variety of chemical phenomena depend mainly on the changes of electronic state in molecules. Chemical reactions are among such chemical phenomena, in which bond formations and cleavages are observed from changes in electronic states among reactants, reagents, and catalysts. A wide variety of actions of catalysts are largely dependent on the electronic states belonging to the atoms bearing the frontier orbitals. To explore high-performance catalysts, analytical investigation based on one electronic state is necessary, and theoretical calculations are indispensable for determining the electronic states. (Tanaka *et al.*, 2013)

2.2.2 Density Functional Theory (DFT)

Density functional theory (DFT) methods are considered to be *ab initio* methods for determining the molecular electronic structure, even though many of the most common functionals use parameters derived from empirical data, or from more complex calculations. In this method, the total energy is expressed in terms of the total one-electron density rather than the wave function. In this type of calculation, there is an approximate Hamiltonian and an

approximate expression for the total electron density. DFT methods can be very accurate but has high computational cost (Lewers, 2016).

In the last and current decade, quantum chemistry calculations have been employed for studying different chemical phenomena such as reaction pathway analysis and spectroscopic assignments due to the theoretical developments and high-speed parallel computers.

By the 1980s, most of the major *ab initio* molecular orbital theories had been established. Among these, the use of density functional theory (DFT) has overwhelmingly increased after 1990 (Tanaka *et al.*, 2013), and it has made an important contribution to promotion of theoretical calculations due to accuracy improvement of functionals (Hammer and Nørskov, 2000). With DFT, the electronic state of the chemical compounds is expressed by electron density functionals, which have been modified and proposed by many types of approaches for improving computation accuracy. It is note-worthy that DFT calculation is widely applied not only to the assignment of experimental spectra, but also to the study of reaction mechanisms (Honkala, *et al.*, 2005; Kandoi, *et al.*, 2006; Rankin, *et al.*, 2002). More so, theoretical methods were used to describe surface chemical reactions in detail and to explain variations in catalytic activity from one catalyst to another. This led to using computational methods to design new catalysts by screening for catalysts with enhanced activity and the ones with improved selectivity (Nørskov, *et al.*, 2009). DFT calculation is now readily adopted owing to its low-cost computing, convenient programs, and many high-performance functional developments (Zhao and Truhlar, 2008).

However, appropriate functionals in DFT should be employed to suit the problem in question, because the accuracy and reproducibility of the results depend on the functionals used (Tanaka *et al.*, 2013). As such, comparison among obtained results with available experimental data or high-level computations is necessary to avoid misleading results.

2.2.3 Basis set: 6-31G*

In addition to the choice of method, some of which we introduced thereof, it is necessary to choose a basis set. Generically, a basis set is a collection of vectors which defines a space in which a problem is solved (Balabin, 2010). In quantum chemistry, it is referred to a set of nonorthogonal functions, usually centred on different atoms in a molecule, which are used to expand the molecular orbitals with the linear combination of atomic orbitals (LCAO). All *ab initio* calculations need to define a level of theory (the method) chosen and a basis set selected. An “orbital” is a one-electron function. Atomic orbital (AO) is represented by atom-centred Slater type orbital (STO) which is its self a combination of Gaussian type orbitals (GTOs).

For STO, equation 2.1 is used to describe the orbital.

$$\phi_{abc}^{STO}(x, y, z) = Nx^a y^b z^c e^{-\zeta r} \quad (2.1)$$

Where: N is a normalization constant, ζ is a constant related to the effective [charge](#) of the nucleus, it controls the width of the orbital (large ζ gives tight function, small ζ gives diffuse function), r is the radial distance of the electron from the nucleus, a is the component of angular momentum along x axis, b is the component of angular momentum along y axis, and c is the component of angular momentum along z axis (Balabin, 2010).

These are H-atom-like orbitals, at least for 1s; however, they lack radial nodes and are not pure spherical harmonics.

For GTO, equation 2.2 is used in describing the orbital.

$$\phi_{abc}^{GTO}(x, y, z) = Nx^a y^b z^c e^{-\zeta r^2} \quad (2.2)$$

Again, N is a normalization constant, ζ controls the width of the orbital, r is the radial distance of the electron from the nucleus and a, b, c control angular momentum (Balabin, 2010).

STO's are more accurate, but it takes longer to compute integrals using them. Thus, quantum chemists choose to use a linear combination of many GTO's to mimic an STO.

A combination of Gaussian type orbitals (CGTO) containing n Gaussians to mimic an STO is generally represented as “STO- n G” basis.

A minimal basis set contains one basis function (STO, GTO, or CGTO) for each atomic orbital in the atom. Double-Zeta has two basis functions for each AO, while triple-zeta contains three basis functions for each AO. A split-valences basis uses only one basis function for each core AO, and a larger basis for the valence AO's. Polarization function arises when other atoms approach an atom's orbitals, which might want to shift to one side or the other (polarization). An s orbital can polarize in one direction if it's mixed with a p orbital. The p orbitals can polarize if mixed with d orbitals

In general, to polarize a basis function with angular momentum l , it is mixed with basis functions of angular momentum $l + 1$.

Pople basis set as developed by the late Nobel Laureate, John Pople, and popularized by the Gaussian set of programs, provides that if STO-3G is a minimal basis set, it means each AO is represented by 3 Gaussians (3G), chosen to mimic the behaviour of an STO. Pople's split-valence double-zeta basis set is called 6-31G; it means the core orbital is a CGTO made up of 6 Gaussians, and the valence is described by two orbitals; one CGTO made of 3 Gaussians, and one single Gaussian. While 6-31G* [or 6-31G(d)] is 6-31G with added d polarization functions on non-hydrogen atoms (Pople, 1971; Legler, 2015)

2.2.4 Geometry Optimization and Descriptors Calculations

The term, electron geometry of the molecule refers to the nuclear arrangement of the molecule that minimizes the molecular electronic energy [the potential energy (P.E) term in Schrodinger equation], including inter nuclear repulsion. The electronic energy changes with the change in bond angle and bond length. Any given molecular formula of a certain molecule can represent a number of molecular isomers with each isomer considered to be a local minimum on the energy surface (a plot of potential energy). A local minimum is a stationary point where all

such displacements lead to an increase in energy, while the stationary point is that geometry such that the derivative of the energy with respect to all displacements of the nuclei is zero. The lowest local minimum is called the global minimum and corresponds to the most stable isomer or conformation. If there is one particular coordinate change that leads to a decrease in the total energy in both directions, the stationary point is a transition structure and the coordinate is the reaction coordinate. Many mathematical techniques are used to find, by hand, the stationary point and, hence, local minimum of a function of several variables. These procedures will find a local minimum in electronic energy of the neighbourhood of the initially assumed geometry. The process of finding such minimum is called Geometry optimization.

With the help of computational chemistry, this can be achieved by applying any of the computational methods highlighted in 2.2.1 above.

Molecular descriptors are the essential information of a molecule in terms of its physicochemical properties such as constitutional, electronic, geometrical, hydrophobic, lipophilicity, solubility, steric, quantum mechanical and topological properties, obtainable in quantitative form.

2.2.5 Quantum modelling

Molecular modelling embraces all theoretical methods and computational techniques used to model or mimic the behaviour of molecules ranging from small diatomics to large biological molecules and material assemblies. The simplest calculations can be performed by hand, but inevitably computers are required to perform molecular modelling of any system of interest. The common feature of molecular modelling techniques is the atomistic level description of the molecular systems. This may include treating atoms as the smallest individual unit as in molecular mechanics approach, or explicitly modelling electrons of each atom as in quantum chemistry approach.

Herein, the descriptors obtained by computational chemistry calculation are used to set up an equation which can predict certain properties of interest of either the ligand or the receptor.

Quantitative Structure-Property Relationship (QSPR) can be regarded as the means of predicting and understanding physical and chemical properties of compounds in terms of their structures. It is a technique used in the chemical and biological sciences and engineering to reduce the laborious work of survey of library of compounds for search of the compounds of interest. It involves a series of quantum computational and statistical calculations that translate to developing a statistical model with appreciable predictive ability.

Quantitative Structure-Activity/Property/Toxicity Relationship (QSAR/QSPR/QSTR) models are regression models which, like other regression models, relate a set of "predictor" variables (X) to the potency of the response variable (Y).

In QSAR/QSPR/QSTR modelling, the predictors consist of physico-chemical properties or theoretical molecular descriptors of chemicals; the QSAR response-variable could be a biological activity of the chemicals. QSAR/QSPR/ QSTR models first summarize a supposed relationship between chemical structures and biological activity, physico-chemical property or toxicity in a data-set of chemicals. Subsequently, QSAR models predict the activities of new chemicals.

As an example, biological activity can be expressed quantitatively as the concentration of a substance required to give a certain biological response. Additionally, when physicochemical properties or structures are expressed by numbers, one can find a mathematical relationship, or quantitative structure-activity relationship, between the two. The mathematical expression, if carefully validated can then be used to predict the modelled response of other chemical structures.

If the activity/property/toxicity is y and the combination of physiochemical/structural properties is x , a QSAR/QSPR/QSTR model has the form of a mathematical equation as seen in equation 2.3.

$$y = f(x) + \text{error} \quad (2.3)$$

The work generally begins by collecting a data set of compounds of interest, then subjecting them to geometric optimization in order to obtain their minimum energy state and various descriptors which are used for the model development. Various quantum mechanical methods are adopted in that respect, some of which include Semi empirical method, Density Functional Theory, Hartree-fock, e.t.c.

2.3 Enantioselectivity

The enantioselectivity determining step of the mechanism was identified and shown to be influenced by proline (Agamiet *al.*, 1984; Benaglia *et al.*, 2001).

Calculating the enantioselectivity-determining transition state geometry is complicated by a number of factors, namely, (1) the size of the system (2) ability of the catalyst to exhibit conformational isomerism (3) several competing H-bondings at the transition state (Lambert *et al.*, 2014)

So many researches have demonstrated that the C-C bond formation step is the key step in determining the enantioselectivity, in aldol and related reactions. Perrin and Chang(2016) reported that the rate-limiting step in the base-catalyzed aldol condensation of benzaldehydes with acetophenones to produce chalcones, is the final loss of hydroxide and formation of the C=C bond.

Simulation reactions of acetone and benzaldehyde with three enamines derived from the condensation of prolinamides was investigated via quantum computational calculations.

The results showed that the amide and the terminal hydroxyl groups form hydrogen bonds with the benzaldehyde substrate. These hydrogen bonds caused high enantioselectivity. (Tang *et al.*, 2004)

It has been reported that the acidic proton of proline is critical for the reactivity and stereoselectivity of the proline-catalyzed direct aldol reaction (List *et al.*, 2000; Sakthivel *et al.*, 2001).

Experimental ^{13}C kinetic isotope effects (KIE) have been used by Lambert *et al.* (2014) to interrogate the rate-limiting step of the Michael addition of glycinate imines to benzyl acrylate catalyzed by a chiral 2,3bis(dicyclohexylamino) cyclopropenimine catalyst. The research discovered that the reaction proceeded via rate-limiting carbon-carbon bond formation. The origins of enantioselectivity and a key noncovalent interaction ($\text{CH}\cdots\text{O}$) responsible for transition state organization are identified theoretically at DFT level of theory and probed using experimental labelling studies.

2.3.1 Quantitative Structure-Property Relationship (QSPR) on Enantioselectivity

Application of QSPR methodology to predict enantioselectivity of a catalyst is an established practice in the field of computational chemistry. The pioneer work of Norrby *et al.* (1997), studied the steric influence on the selectivity in palladium-catalyzed allylation. The structure of π -allyl Pd was generated using molecular mechanics (MM2).

A set of descriptors including breaking Pd-C bond length, Dihedrals N1-N2-Pd-Ct, Dihedrals Du-Pd-Cc-Ct, Steric Interaction with Nu, were used to carry out regression using Levenberg-Marquardt (multi linear regression). Leave one out (LOO) $Q^2 = 0.86$, Leave several out (LSO) $Q^2 = 0.87$. However, multi linear regression method adopted in the work has the limitation of not being able to generate variety of models in order to be tested and choose the best one.

Kozlowski *et al.* (2003) worked on predicting the enantioselectivity of β -amino alcohol catalysts in aldehyde alkylation. A data set of 13 β -amino alcohol catalysts was utilized to

develop the QSPR using normal least square method. Probe Interaction Energies (PIE) descriptors were correlated with ΔG values which could be converted to enantiomeric excess (ee) values using $\Delta G = RT \ln K$, where K is the enantiomeric ratio. The best individual model was found to be; $\Delta G = a + c_1(\text{PIE}_1) + c_2(\text{PIE}_2)$, where, $a = 5.48$ kcal/mol, $c_1 = -0.27$, $c_2 = 0.36$, average standard deviation (SD) = 0.17 kcal/mol, $R^2 = 0.93$. In a work reported by Dixon *et al.* (2004) energy was calculated as PIE by quantum mechanical method and regression was performed using n-variables regressions by simulated-annealing.

Heck reaction, wherein unsaturated halide combines with alkene to form substituted alkene in the presence of a base and palladium catalyst, was investigated using QSPR. Selected descriptors which include, HOMO, LOMO, GAP, dipole, A, change on ligating atom, Hammett(+), Hammett(-), V, percentage of sphere occupation, R_{\max} , and O_{\max} were used for the model development. (Farruseing *et al.*, 2004)

QSPR approach was used for asymmetric lithiation of *N*-Boc-pyrrolidine by Kozlowski *et al.* (2004). Their work investigated a chiral amine, spartine for its essential moiety with regard to enantioselectivity. A training set of 16 diamine lithium complexes was used to develop the QSSR. HF/3-21G* was used for structure optimization and B3LYP/6-31G* (DFT) was used for PIE calculation. The model was used to predict (*R*) 22-25% ee whereas experimental result was (*R*) 35% ee. The LOO Cross Validation: $r_{cv}^2 = 0.68$ and the Correlation Coefficient (CC) = 0.82. Further prediction for new catalysts was carried out by Kozlowski *et al.* (2006) in which a training set of 13 compounds and a test set of 8 compounds were utilized. The predicted values of ee were plotted against the experimental values with $r_{pred}^2 = 0.87$.

More effective method, Genetic Function Algorithm, was adopted in predicting enantiomeric excess in a catalytic processing chirality codes (Xu *et al.*, 2012). In this study, a library of chiral catalysts was used for the asymmetric hydrogen transfer to acetophenone utilizing the QSPR methods. The molecular descriptors of the catalysts, including chirality codes, were used

to construct the QSPR models for the quantitative prediction of enantiomeric excesses. The results revealed that the molecular codes derived from the structure of catalysts can be applied to construct robust predictive models of the catalytic performance. However, the work did not apply the model for prediction of the ee of any compound.

Cross validation by Leave One Out (LOO) or Leave Several Out (LSO) is usually employed for validation of QSPR model. In LOO, correlation model is re-calculated with training set where one compound is excluded. Then, activity of excluded compound is predicted using the new model and compared with real value (Cheng *et al.*, 2017). In LSO, almost same as LOO, several compounds were excluded in this case. Prediction of activity of other compounds can be calculated using model equation (Cheng *et al.*, 2017).

2.4 Asymmetric Aldol Reaction

Generally, asymmetric aldol reaction is a powerful method for the construction of carbon-carbon bonds in an enantioselective fashion. Historically, the aldol reaction, was first discovered independently Alexander by Borodin in 1869 and Charles-Adolphe Wurtz in 1872 (Ritcher, 1969; Wurtz, 1872), and is one of the most powerful transformations in organic chemistry. The process unites two carbonyl partners to give β -hydroxyketones with up to two new stereocenters. The reaction requires two carbonyl compounds which may or may not be the same. One of the carbonyl compounds must contain a C-H group bonded to the carbonyl (C=O) group. The hydrogen is called α -hydrogen (Martínez *et al.*, 2014).

This reaction has significant application in biochemistry (Liu *et al.*, 2016). For example, an aldol condensation reaction occurs in the synthesis of glucose, and the reverse of this reaction occurs in the catabolism (breakdown) of glucose.

The breakdown of fructose-1,6-bisphosphate into dihydroxyacetone and glyceraldehyde-3-phosphate in the second stage of glycolysis, is an example of a reverse aldol reaction catalyzed by the enzyme aldolase A (also known as fructose-1,6-bisphosphate aldolase).

Also, in the glyoxylate cycle of plants and some prokaryotes, isocitrate lyase produces glyoxylate and succinate from isocitrate. Following deprotonation of the OH group, isocitrate lyase cleaves isocitrate into the four-carbon succinate and the two-carbon glyoxylate via an aldol cleavage reaction. This cleavage is very similar mechanistically to the aldolase A reaction of glycolysis (Hu *et al.*, 2016).

This study focuses on an intermolecular asymmetric aldol reaction, particularly, a reaction between acetone and 4-nitrobenzaldehyde shown in scheme 2.1.

Scheme 2.1: A reaction between acetone and 4-nitrobenzaldehyde.

The aldol reaction may proceed via two fundamentally different mechanisms. Carbonyl compounds, such as aldehydes and ketones, can be converted to enols or enol ethers. These species, being nucleophilic at the α -carbon, can attack especially reactive protonated carbonyls such as protonated aldehydes. This is the 'enol' mechanism. Carbonyl compounds, being carbon acids, can also be deprotonated to form enolates that are much more nucleophilic than

enols or enol ethers and can attack electrophiles directly. The usual electrophile is an aldehyde, since ketones are much less reactive. This is the 'enolate' mechanism. The mechanism of the reaction, as seen in Scheme 2.2, was reported by List *et al.* (List *et al.*, 2000).



Scheme 2.2: Reported mechanism of asymmetric aldol reaction of acetone and 4-nitrobenzaldehyde catalysed by proline.

The discovery of aldol reaction can be traced back to the work of Aleksandr Porfir'evich Borodin in 1869, where he first observed the formation of the aldol, 3-hydroxybutanal, from acetaldehyde under the influence of inorganic catalysts such as hydrochloric acid or zinc chloride (Richter, 1869).

The importance of producing enantiomerically pure compounds cannot be over emphasised so also the fact that asymmetric catalysis is the best method of achieving such is beyond doubt. In this method, a chiral catalyst is used to transmit chiral information to the target product through formation of a transition state (TS). Though chiral transition metal complexes are used as catalysts in this regard, small organic molecules containing chiral centres are more suitable for this task, most especially, biological reaction, as their mechanism of action resembles that of enzymes, aldolase and decarboxylase (Zlotin *et al.*, 2009). This is the area called organocatalysis.

A breakthrough research in organocatalysis was the discovery of the Robinson asymmetric annulation using (S)-proline as a catalyst by two independent groups of scientists: Hajos-Parrish and Eder-Sauer-Wiechert. The reaction is useful in the synthesis of steroids and so many other chiral and biologically active compounds.(Hajos and Parrish, 1971; Eder *et al.*, 1971)

The development of asymmetric organocatalysis as a field of study was triggered by the work of List *et al.*(2000) which exhibited the ability of (S)-proline to catalyze intermolecular aldol reactions. One of the carbonyl compounds, usually ketone acts as the nucleophile while the other, mostly aldehyde serves as the electrophile in the formation of the new C-C bond.

2.5 Mechanism of organocatalyzed asymmetric aldol reaction

Determination of the transition states is crucial in theoretical study of mechanism and because the transition structures only exist for a short period they cannot be measured experimentally (Engel and Reid, 2006). However, measurement of activation energies can help to ascertain the energies of transition states relative to reactants.

Different automated routines have been developed to efficiently locate transition states; however, for many methods, there is no guarantee that a TS would ever be found.

A highly effective but cumbersome approach is to explicitly calculate photoemission spectra (PES) by manually varying specific constrained coordinates. In this approach, unconstrained coordinates are allowed to freely relax to minimum energy positions in a standard geometry optimization, and the result is a single data point on a PES surface. After multiple points have been sampled, the PES can be interpolated with minimum curvature or radial basis function algorithms.

The most ubiquitous TS searches use the modified Newton–Raphson routine and start from anatomic configuration close to that of the TS. The optimization procedure successively moves along the normal mode of the Hessian with the lowest magnitude in an attempt to find the point of maximum negative curvature. These methods are rather efficient if the starting geometry is near the real TS and the utilized Hessian contains only one negative mode (Gasper, 1954). When neither of these criteria is applicable, this procedure may fail spectacularly.

Other QM routines incorporate more information either by requiring coordinates from the initial reactant and product configurations or sometimes by including an initial guess for the TS as well. Such routines are quite efficient for locating transition states involving one or two bonds breaking in molecules, but for very large systems such as solids where atom movement influences a larger environment, alternative methods are needed.

Hajos proposed a mechanism in which one of the enantiotropic acceptor carbonyl groups was activated as a carbinol amine (Hajos and Parrish, 1974). However, the stereochemistry of this mechanism was strongly questioned soon after its proposition by a work on review of annulation (Jung, 1976).

Intra- and inter-molecular proline-catalyzed aldol reaction of aldehydes and ketones of different kinds have been extensively investigated (Blackmond *et al.*, 2007; 2009; Houk and Bahmanyar, 2001; Sharma and Sunoj, 2010; Sanchez *et al.*, 2012; and List *et al.*, 2003). In particular, Houk and List provided the evidence of the reaction mechanism, which involves the

stereoselectivity model commonly known as the Houk-List Model (Houk and Bahmanyar, 2001; List *et al.*, 2003).

The mechanism of the reaction, as seen in scheme 2.2, was proposed to proceed via iminium-enamine transformation (List *et al.*, 2000) as was discussed earlier by Jung in 1976 and Eschenmoser *et al.* in 1978, and several other groups, way back in the 1970s and 1980s (Agami *et al.*, 1986). This mechanism resembles that of a typical class I aldolase enzymes that catalyse the aldol reaction in a biological system (Drauz and Waldmann, 1995; Wong and Whiteside, 1994). A side-chain enamine mechanism that involves two proline molecules in the C-C bond formation step was proposed by Agami *et al.* (1986). One proline molecule acted on enamine formation while the other one served as a proton transfer mediator. Although widely accepted, but this two-proline mechanism was lately challenged when a homogeneous one-proline enamine mechanism was proposed in which the various protons transfers were mediated by proline's carboxylic acid functional group (List *et al.*, 2000). In the work of Houk and Bahmanyar, (2001), a very similar mechanism for the intramolecular variant was developed. Working on both intra- and inter-molecular aldol reactions, more evidence for the involvement of only one proline in the transition state of C-C bond formation was further provided by List *et al.*, in 2003. This was achieved based on kinetic studies, by observing that the reaction was first-order in proline. Another strong evidence for one-proline mechanism came in the studies of intra- and inter-molecular aldolization catalysed by polymer-supported prolines (List, 2000). In addition, stable intermediates, formed in reversible reactions between ketone and proline, were detected by H-NMR (List *et al.*, 2004).

DFT was used to investigate proline-catalyzed aldol reaction between acetone and acetaldehyde and found to proceed via enamine mechanism with the amino function of proline performing the catalytic action. Despite the fact that many proline-catalyzed aldol reactions are

completely homogeneous, Swaminathan *et al.*(1999), proposed a heterogeneous aldolization mechanism on the surface of crystalline proline.

Houk and Bahmanyar(2001) and List *et al.*(2004) focused on finding new mechanistic insights into proline-catalysed aldol reaction using different combinations of aldehydes and ketones. However, despite the apparent interest in this area of research, very limited experimental data were gathered in support of the various mechanisms proposed.

2.6 Rate of aldol Reaction

Blackmond *et al.*(2009) used RPKA to investigate the dependence of reaction rate on the reactants' concentrations and proposed a rate law. Ceotto *et al.*(2016) improved upon the rate law, taking into consideration the reversibility of the elementary steps. Previous works indicated that enamine formation and/or C-C bond formation is the rate-determining step (List *et al.*, 1998). However the work of Rankin *et al.*(2002), utilizing density functional method in acetone medium, found that the initial complexation between proline and acetone required the highest activation energy, which qualified the step as the rate limiting. More so, upon carrying out the experiment in different medium (DMSO), the activation energy dropped significantly (Rankin *et al.*, 2002). It was observed that solvents exert great effect on the activation energies of the reactions. In the work of Rankin *et al.*(2002), the solvent effect was determined using Onsager model and found to increase the activation energy by up to 120 kJ/mol. Tang *et al.*(2004) found that the hydrogen bond formed between the amide group and the terminal hydroxyl with the benzaldehyde substrate in the simulation reactions of acetone and benzaldehyde was responsible for the reduced activation energy observed.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Materials

This work was carried out *in silico*, and the hardware and software that were used include:

- I. HP Pavilion (core I 5), 12.00GB RAM, 1TB storage disk, Windows 8.1 Home Premium
Operating system
- II. Spartan 14v.1.1.0 (Wavefunction)
- III. Chem3D Pro12.0.1v,
- IV. Padel descriptor tool kit,
- V. Acceryl Material studios 8.0,
- VI. Chemdraw ultra 12.0,
- VII. Microsoft office Excel 2013.

3.1.1 The Dataset of the catalyst compounds

A data set of 35 pyrrolidine based molecules (prolines, prolinamides, others) with catalytic effect on a chosen asymmetric aldol reaction was collected from previously published works with their respective enantiomeric excess (ee) expressed in percentage of the total amount of reactants converted in the catalytic reactions. The structures of the compounds were drawn using Chemdraw software. The detail structure of all the sample compounds are given in Table 3.1 with their respective enantiomeric excess and the reaction conditions (medium and temperature).

Table 3.1: Data set of pyrrolidine-based catalyst for asymmetric aldol reaction.

S/N	CATALYSTS	SOLVENT	TEMP(⁰ C)	ee(%)	Reference
-----	-----------	---------	-----------------------	-------	-----------

1		Acetone	25	86	Sakhivel <i>et al.</i> , 2001
2		Acetone	25	81	Tang <i>et al.</i> , 2006
3		DMSO	25	78	Sakhivel <i>et al.</i> , 2001 and List, 2000
4		DMSO	25	62	Sakhivel <i>et al.</i> , 2001; List, 2000
5		DMSO	25	73	Sakhivel <i>et al.</i> , 2001; List, 2000
6		DMSO	25	76	Sakhivel <i>et al.</i> , 2001

7		DMSO	25	62	Sakthivel <i>et al.</i> , 2001 and List, 2000
8		DMSO	25	74	Sakthivel <i>et al.</i> , 2001 and List, 2000
9		DMSO	25	74	Sakthivel <i>et al.</i> , 2000; List, 2000
10		Acetone	20	99	Tang <i>et al.</i> , 2005
11		Acetone	-15	64	Tang <i>et al.</i> , 2006
12	(S)-N-((R)-3-(benzylthio)-1-hydroxy-1,1-diphenylpropan-2-yl)pyrrolidine-2-carboxamide	Acetone	20	80	Almasi <i>et al.</i> , 2008

13		DMSO		82	DineÂr, and Amedjkouh, 2006
14		DMSO	20	76	Hartikka and Arvidsson, 2004
15		DMSO	20	88	Tong <i>et al.</i> , 2008
16		DMSO		86	Sakthivel <i>et al.</i> , 2001
17		Acetone		94	Gryko, Lipin' ski, 2005 and 2006
18		Acetone	0	80	Almasi <i>et al.</i> , 2008

19	Acetone	-5	82	Lacoste <i>et al.</i> , 2004
20	Acetone	25	84	Singhet <i>al.</i> , 2006
21	Acetone		92	Leyet <i>al.</i> , 2005
22	Acetone		93	Berkessel <i>et al.</i> , 2004
23	Acetone		90	Berkessel <i>et al.</i> , 2004

24		Acetone	20	71	Berkessel <i>et al.</i> , 2004
25		Acetone	20	92	Cheng <i>et al.</i> , 2006
26	(S)-N'-((S)-pyrrolidine-2-carbonyl)pyrrolidine-1-benzyl-2-carbohydrazide	DMF	20	89	Kokotoset <i>al.</i> , 2005
27		DMF	20	71	Moorthy and Saha, 2009
28		DMF		72	Moorthy and Saha, 2009

29	DMF	20	72	Moorthy and Saha, 2009
30	DMF	20	72	Moorthy and Saha, 2009
31	DMF		71	Moorthy and Saha, 2009
32	DMF		72	Moorthy and Saha, 2009
33	DMF		90	Moorthy and Saha, 2009

34		DMF		68	Moorthy and Saha, 2009
35		Acetone	-35	98	Zhao <i>et al.</i> , 2005

(S)-N-((1S,2S)-2-acetamido-1,2-diphenylethyl)pyrrolidine-2-carboxamide

3.2. Methods

The general procedure adopted in this study involved building and optimising all the compounds in the dataset by using Spartan '14' quantum mechanical software. Quantum chemical descriptors and thermodynamics properties were generated using Spartan "14" V.1.1.0 and Chem3D Pro 12.0.1V, while topological descriptors, constitutional descriptors and geometrical descriptors were generated using Padel descriptor software. Other important descriptors were calculated manually. The material studio 8.0 software was used to generate correlation matrix, perform outlier analysis and subsequently develop the QSPR models. The

models were thoroughly validated and the best performing model was used to search for a viable pyrrolidine-based catalyst. The catalytic surface reaction to determine the activity of the designed catalyst in comparison with that of proline was finally carried out using the Spartan '14' software.

3.2.1 Sketching the molecular structures using Chemdraw Ultra 12.0 and optimization using ChemBio3D Pro 12.0.1V.

The Chemdraw ultra 12.0 software was launched to display the interface as shown in Plate 3.1. Cyclopentane structure was selected from the toolkit at the left hand side of the software interface to begin the drawing for proline structure. Since the structure involves elements other than carbon and hydrogen, the periodic table was explored to select a nitrogen atom and the nitrogen atom was inserted in the cyclopentane ring to make a pyrrolidine ring. Oxygen atoms were attached to complete the structure of proline (Plate 3.2). The work was saved and same was repeated for all the compounds in the dataset. To optimize the structures, the Chemdraw ultra 12.0 was exited and the Chem3D Pro 12.0.1V was launched. The file button was clicked and 'import file' was selected from the pop up menu. This gave a dialog box to select the saved Chemdraw ultra 12.0 project.

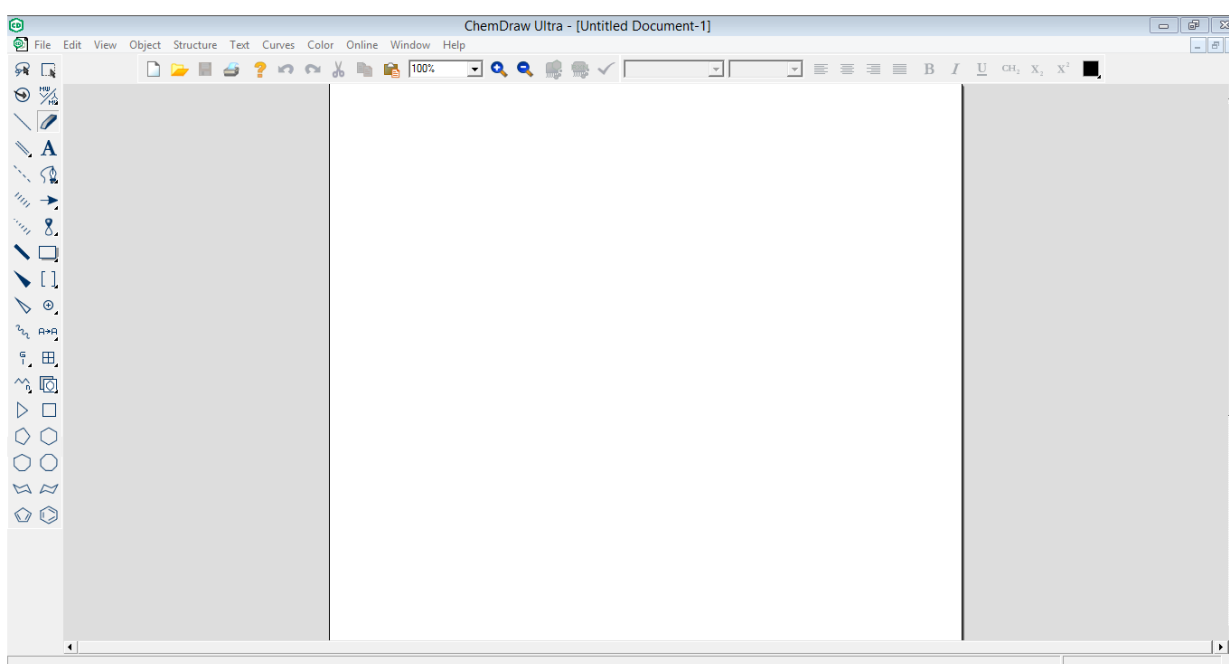


Plate 3.1: Chemdraw Ultra 12.0 user interface showing the menu bar

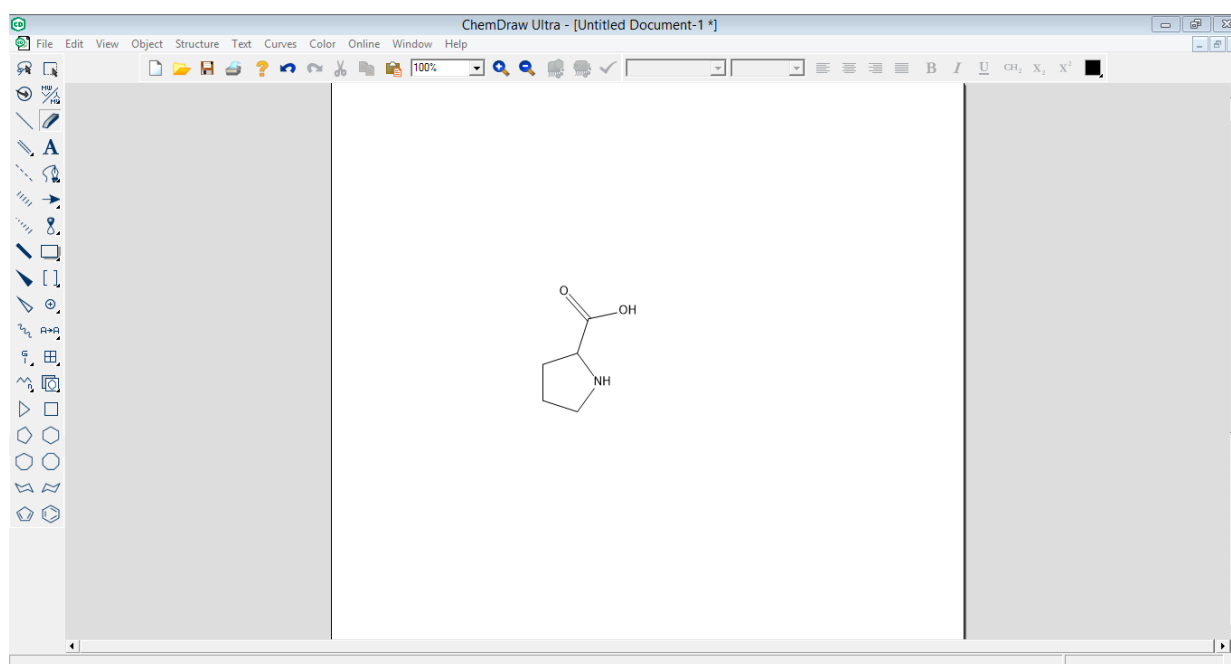


Plate 3.2: Chemdraw Ultra 12.0 user interface showing the drawn proline structure

Proline was selected and imported into the model area window as seen in Plate 3.3. From the menu bar, the calculation option was selected to display a window. On the window, GAMESS interface was located and then minimize (energy/geometry) was clicked. DFT was selected as the method of choice and then the run button was clicked to initiate the calculation. The display option on the menu bar was clicked to get the result of the calculation. When a window popped up, “properties” option was clicked. A small window containing the following options: chempro, chemprostd, CLogP, Deriver, molecular Topology was displayed. Each of the options was marked and “Ok” button was clicked to display the results. The same procedure was repeated for all the compounds in the dataset.

3.2.2 Optimization using Spartan “14” V.1.1.0

In this work, cascading computational method was adopted for optimization and calculation of all the structural geometries involved. The “cascade computational method” involves the use of molecular mechanics to remove strain energies of the molecules, followed by application of semi-empirical method to optimize the structures and finally applying more exact DFT methods to obtain improved results (Hehre, 1998). This method is significant because it has the ability to make computation less costly by relegating initial geometry calculation to less intensive methods.

3.2.2.1 Molecular mechanics and Stationary points geometry optimizations using semi-empirical (PM3) method

On the Microsoft windows 8.1 professional version 2014 SPI computer system, SPARTAN 14 program suit was lunched to open the user interface as shown in Plate 3.4. The “file” option was selected from the tool bar to launch the molecules (peptides, nucleotides, organic, inorganic) or substituents model kit depending on the molecule of interest. Atoms (with the correct geometry) that make up the molecule were selected and pasted on the SPARTAN interface. Illustration on building L-proline molecule is as given here in.

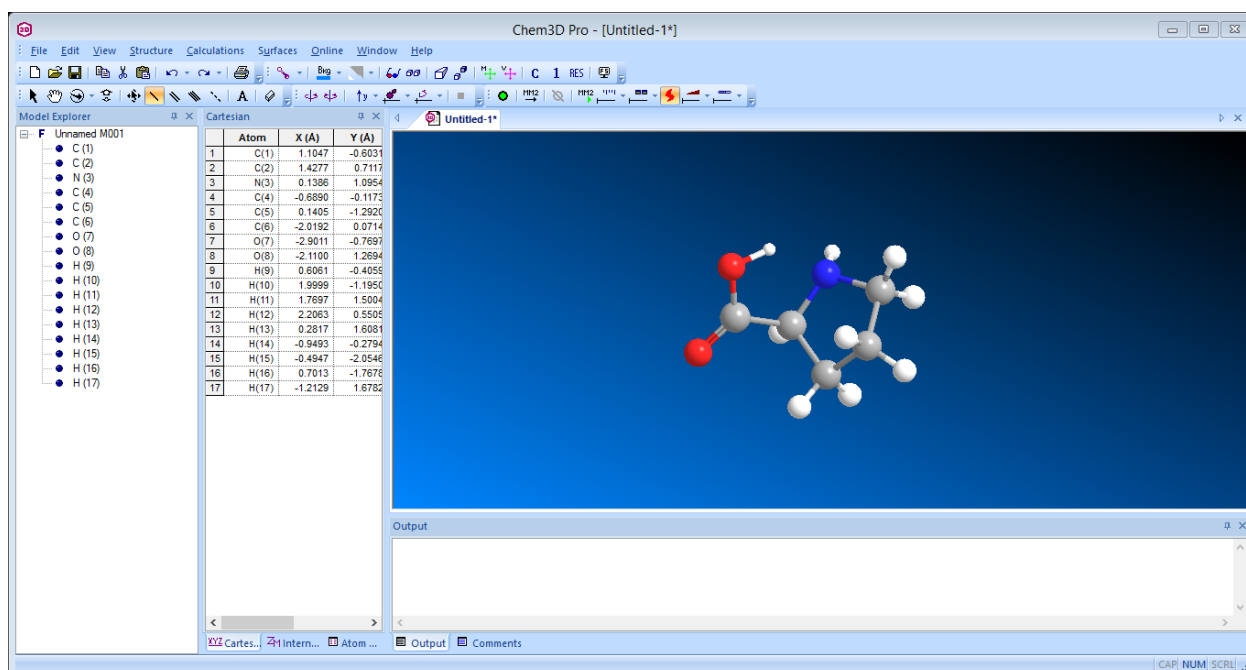


Plate 3.3: Chem3D Pro 12.0 user interface showing the imported proline structure.

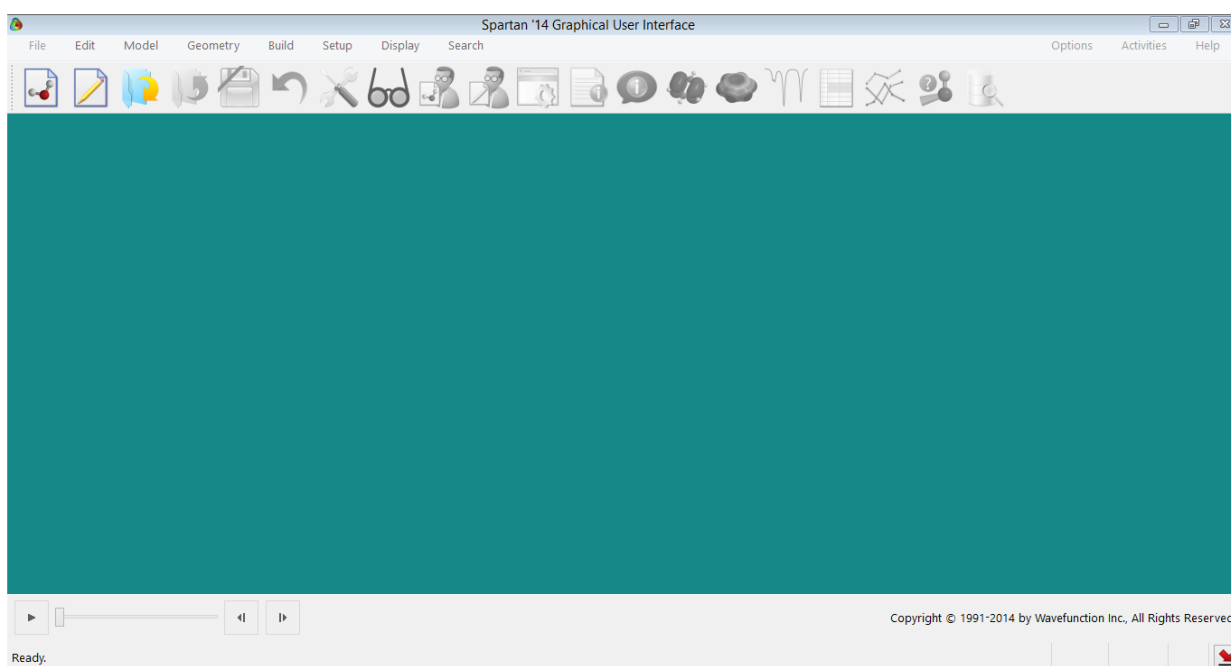


Plate 3.4: SPARTAN 14 user interface showing the menu and the tool bars

Firstly, “new build” was selected from the tool bar to launch the molecules (peptides, nucleotides, organic, inorganic) and substituents model kit as shown in Plate 3.5. This was followed by clicking on rings from the model kit from which the cyclopentane ring was selected and pasted on the user interface as shown in Plate 3.6. From the bottom right corner of the interface, ‘delete bond’ function was selected and one of the carbon atoms in the cyclopentane ring was clicked to be deleted (Plate 3.7). From the popped model kit hitherto, an N pyramidal atomic fragment was selected and added to the position of the previously deleted carbon atom. The ‘make bond’ function was selected from the bottom right corner of the interface and a click was made between the N atom and the neighboring carbon atom to establish a single bond between them, thereby completing the structure of a pyrrolidine ring as presented in Plate 3.8.

This was followed by selecting sp^2 -hybridized carbon (C_{sp^2}) atomic fragment and adding it to the α -carbon atom of the pyrrolidine ring as shown in Plate 3.9. sp^2 -hybridized oxygen (O_{sp^2}) atomic fragment was selected and added to the double bond on the C_{sp^2} fragment (Plate 3.10). O atomic fragment was then selected and added to the single bond on the C_{sp^2} fragment to give the structure seen in Plate 3.11. Finally, Molecular mechanics calculation was performed by clicking the ‘minimize’ button on the bottom right corner of the interface to remove strain energy and automatically add H atom to all free bonds on the structure. The final proline structure is given in Plate 3.12.

The proline molecule built was then subjected to semi empirical calculation to optimize the structure. The “set up” menu on the open window was selected and “calculation” on the drop down menu was clicked followed by “Equilibrium Geometry” at “ground state”. “Semi-empirical” at the desired level (PM3) of computation was selected on the same interface

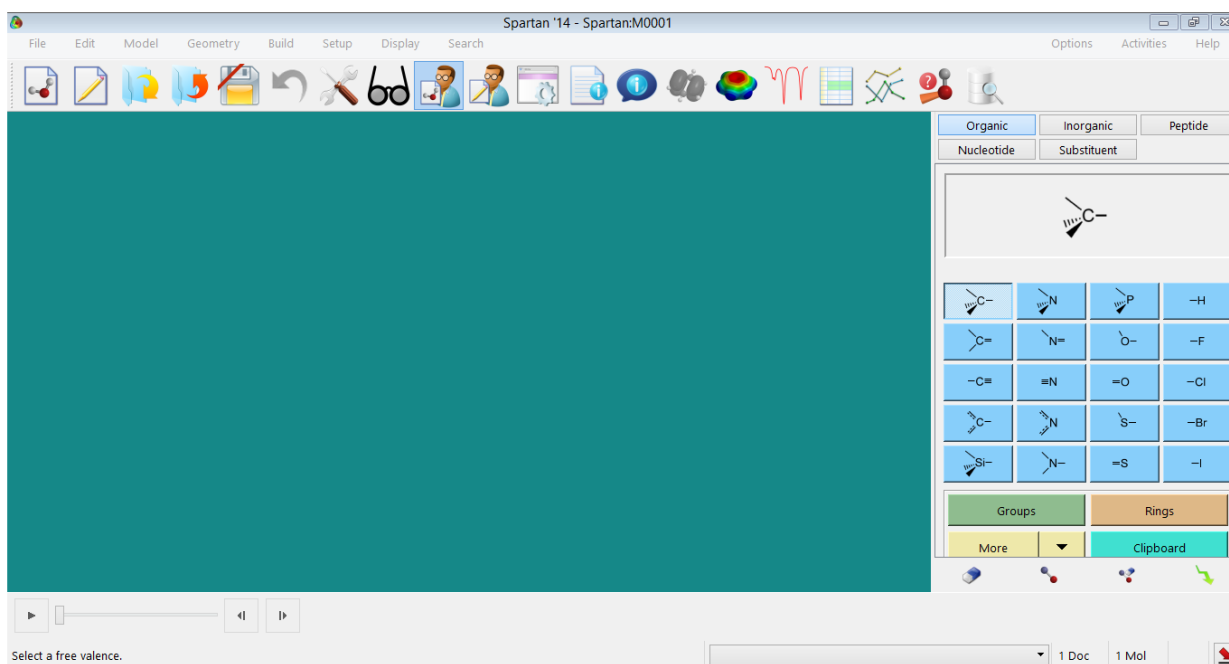


Plate 3.5: File menu interface showing the molecule and substituents tool kit

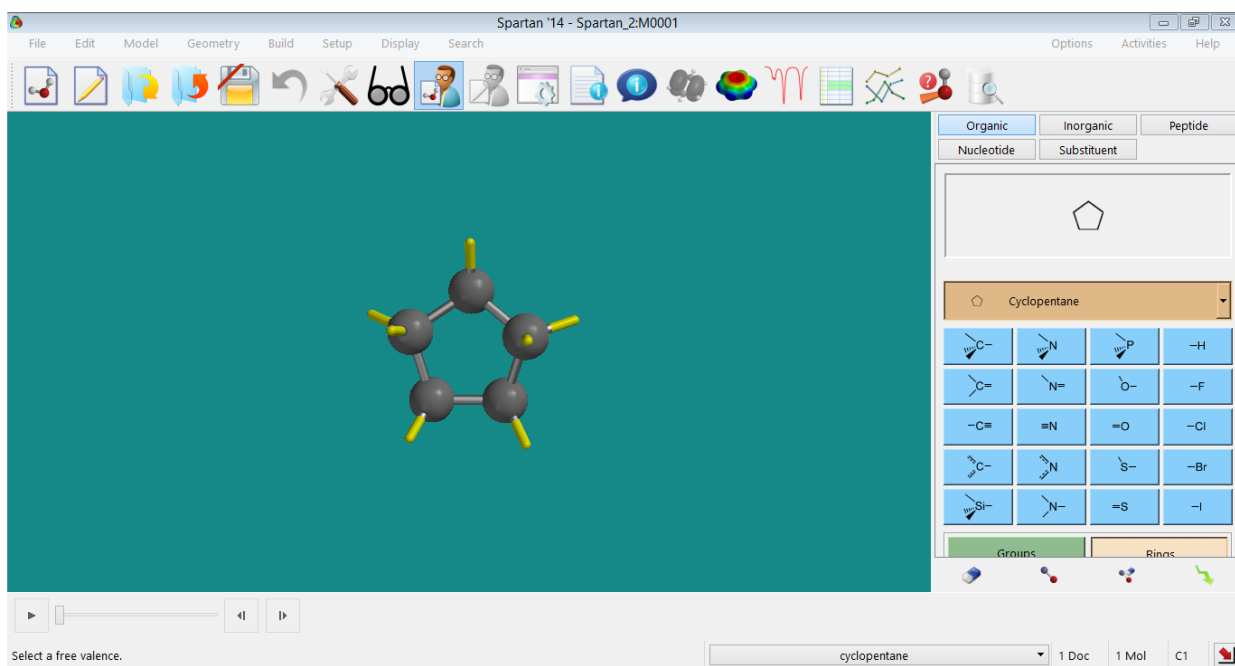


Plate 3.6: Interface showing the selected and pasted cyclopentane ring.

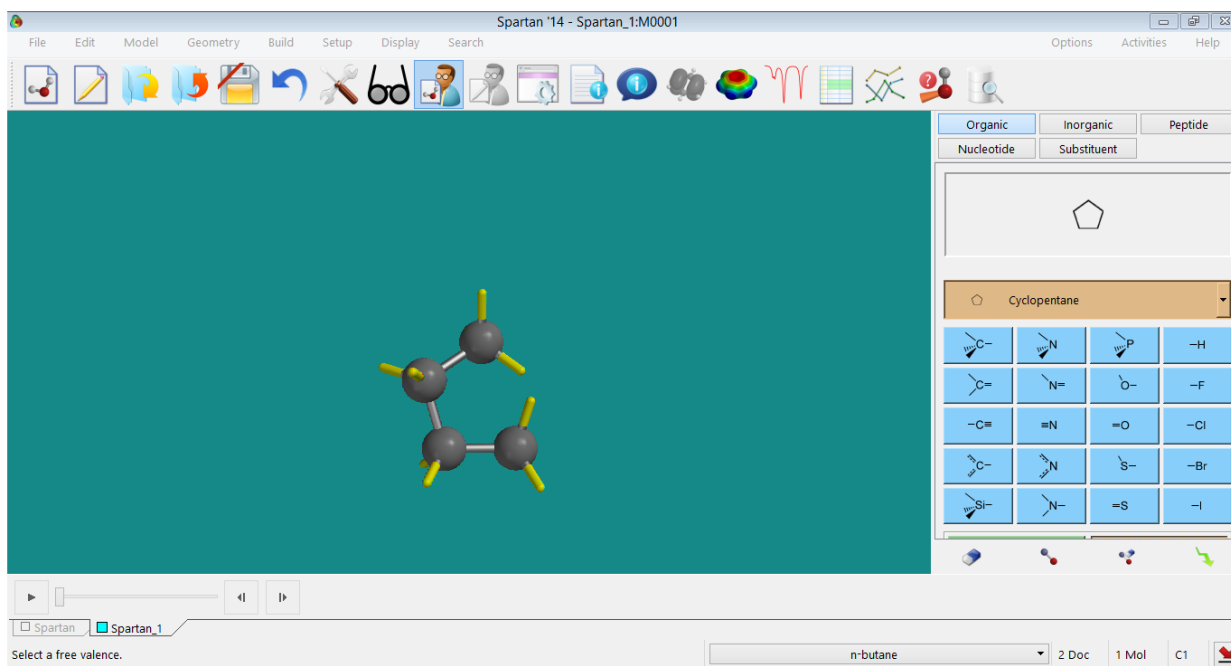


Plate 3.7: Interface showing one C atom deleted from the cyclopentane ring

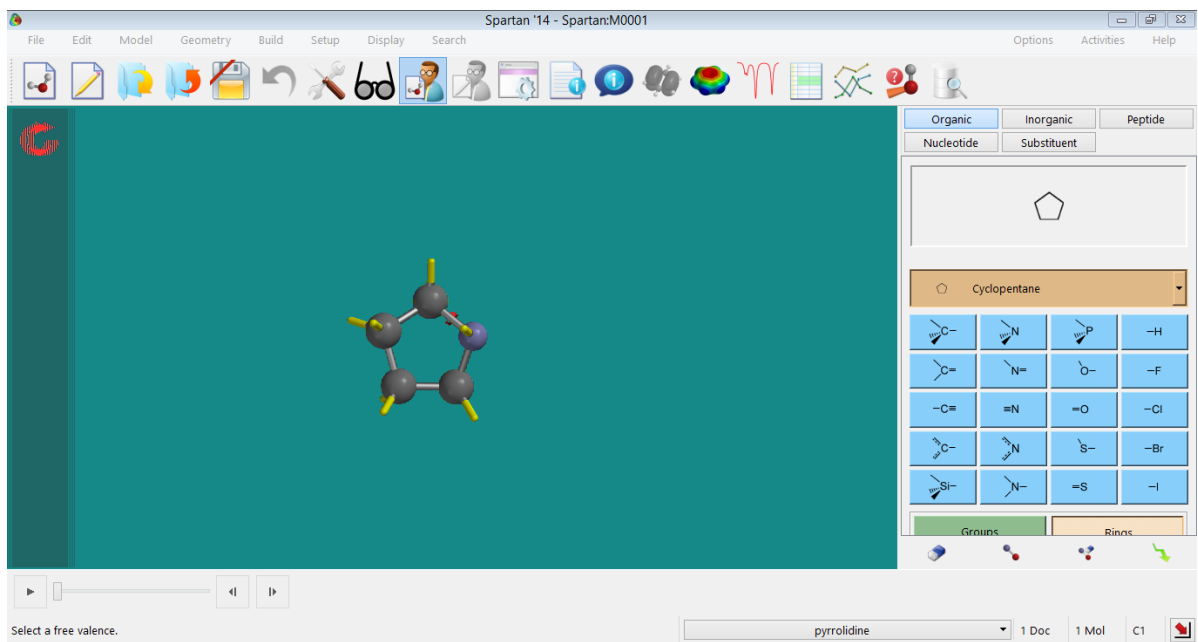


Plate 3.8: Interface showing a pyrrolidine ring formed by addition N atom

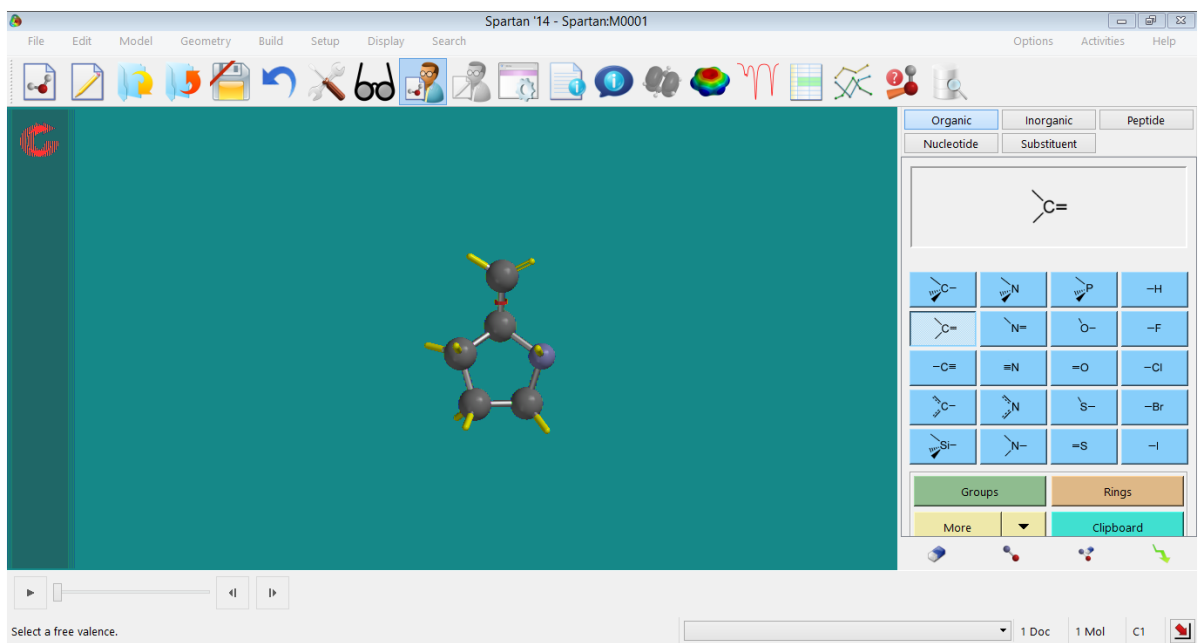


Plate 3.9: Interface showing selected and added Csp2 tool on the pyrrolidine ring

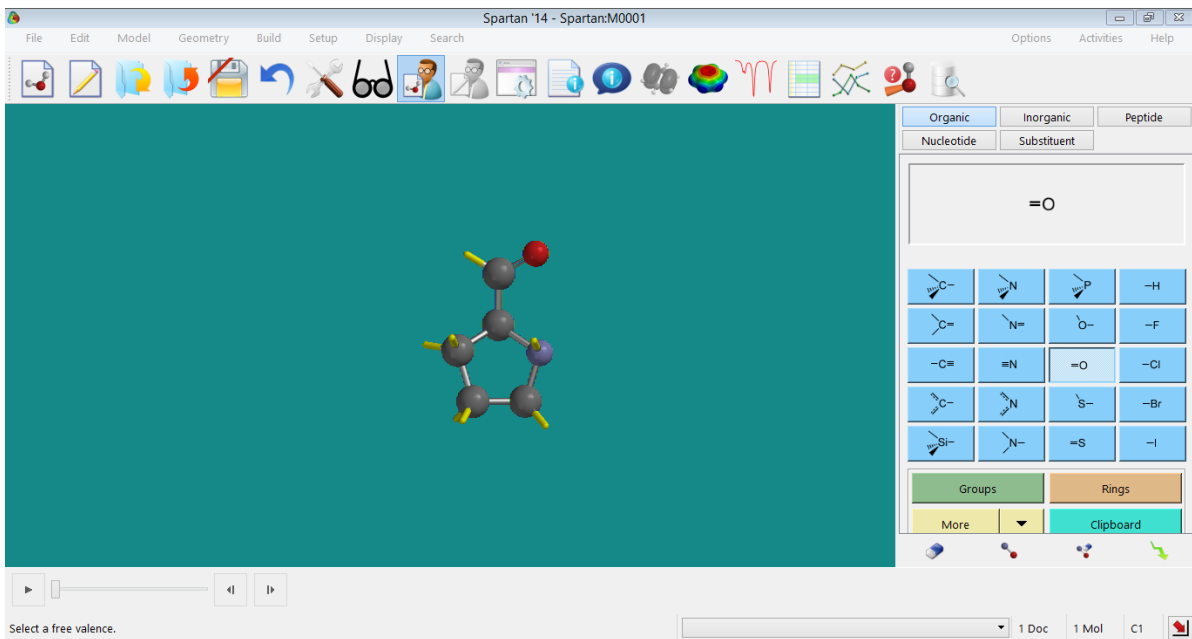


Plate 3.10: Interface showing selected and added Osp2 tool on the Csp2 carbon

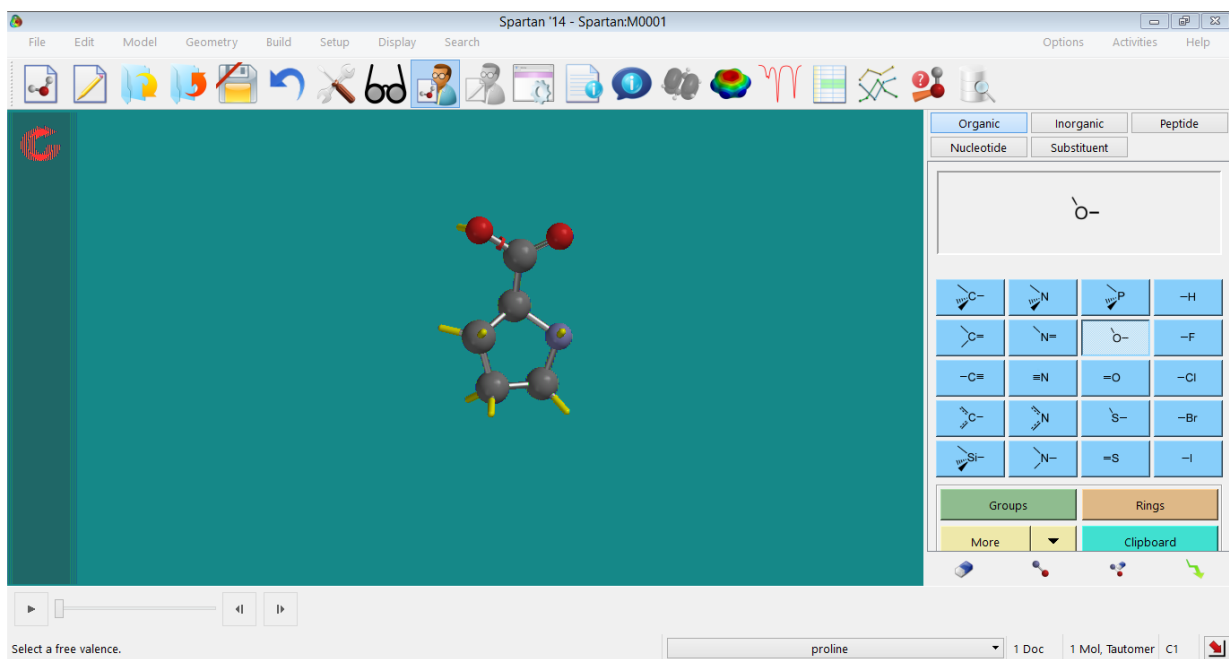


Plate 3.11: Interface showing selected and added O fragment on the Csp² fragment

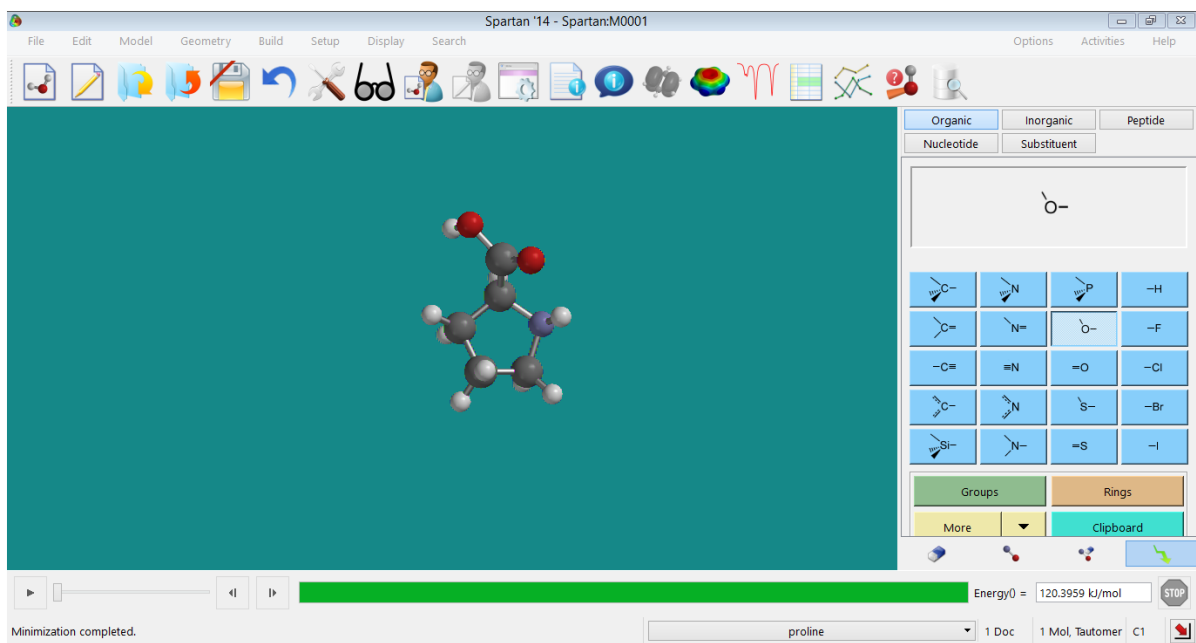


Plate 3.12: Interface showing completely built proline molecule.

Based on the molecule built on the interface, actual charge was input on the “Total charge” menu with the corresponding multiplicity also selected. Furthermore, to select “vibrational

modes”, “Thermodynamic”, “Orbitals and Energies”, “Charges and bond order” tabs, “IR” menu was selected. Finally, Global Calculation menu was selected before (Plate 3.13) submitting the calculation by clicking the submit menu, and the file saved with appropriate name. After the calculation was completed the output was accessed through the “Display” menu as seen in Plate 3.14.

3.2.2.2 DFT (B3LYP) geometry optimization for stationary points

The saved file of the optimized proline molecule from the semi-empirical optimization (section 3.2.2.1) was opened on a new interface for the density functional theory calculations. “Calculation” was selected from the “Set up” menu, and this was followed by selecting “Equilibrium Geometry” at “Ground” state, selecting “Density Functional” method, “B3LYP” model and “6-31G*” basis set. The medium was set to be “acetone” and “Current” geometry was selected too.

Furthermore, to select “vibrational modes”, “Thermodynamic”, “Orbitals and Energies”, “Charges and bond order” tabs, “IR” menu was selected. Finally, Global Calculation menu was selected. Plate 3.15 shows the interface for completely set DFT calculation. The calculation was submitted by clicking the submission menu, and the file was saved with appropriate name. After the calculation was completed, the output was accessed through the “Display” menu (Plate 3.16).

The same procedure as in sections 3.2.2.1 and 3.2.2.2 was repeated for the entire compounds in the dataset and the optimized structures were exported in structure-data file (SDF) format to be recognizable in the padel descriptors software.

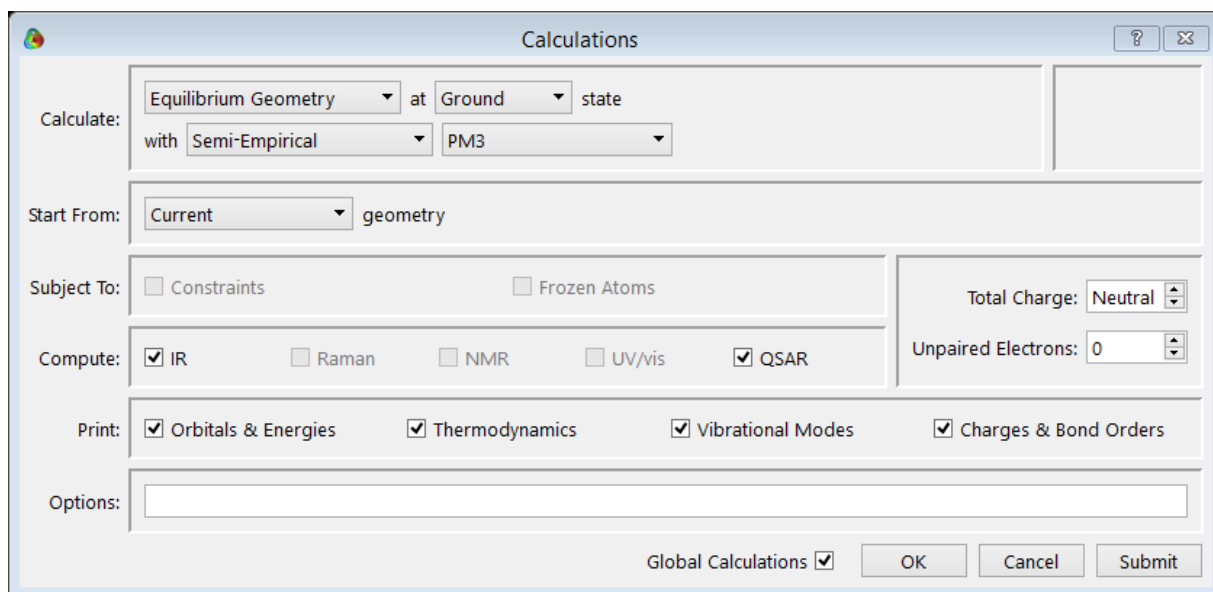


Plate 3.13: Interface showing completely set semi empirical calculation.

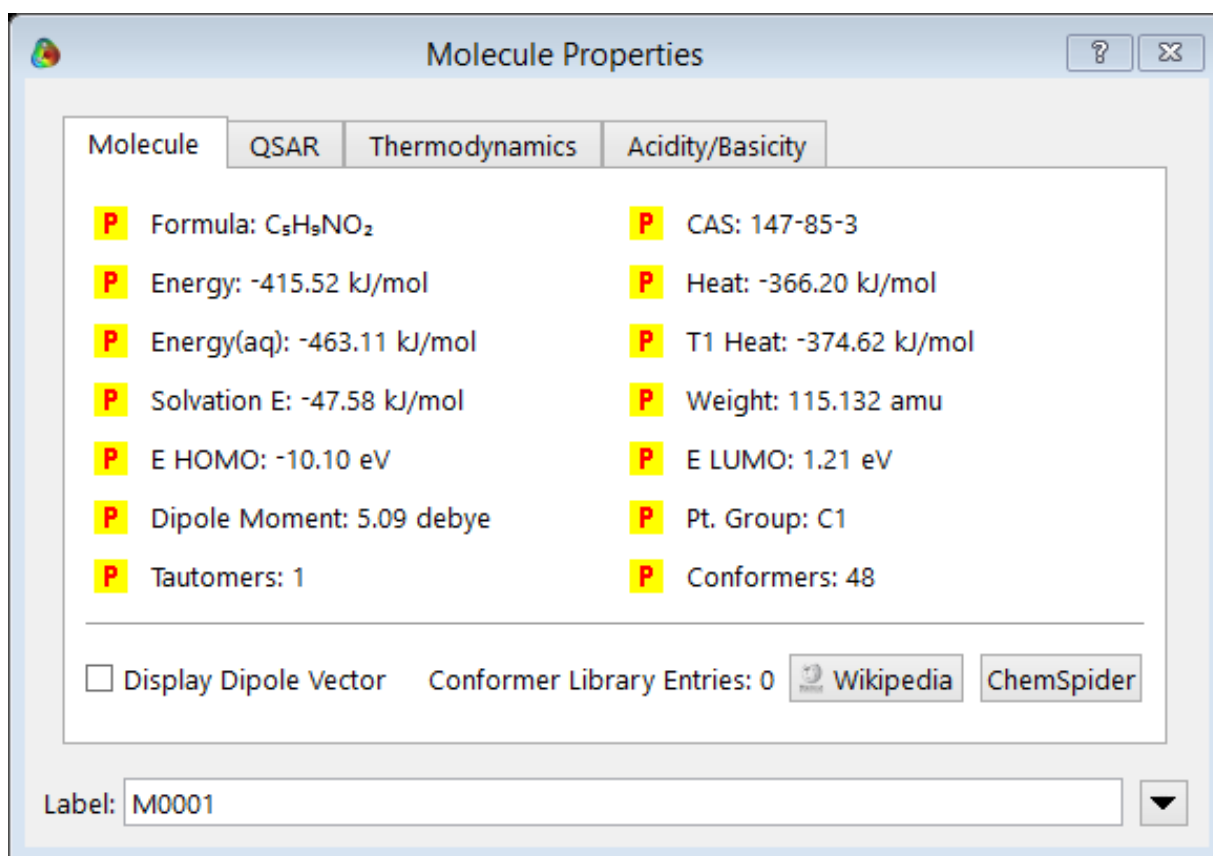


Plate 3.14: Display window showing the results of semi empirical calculation.

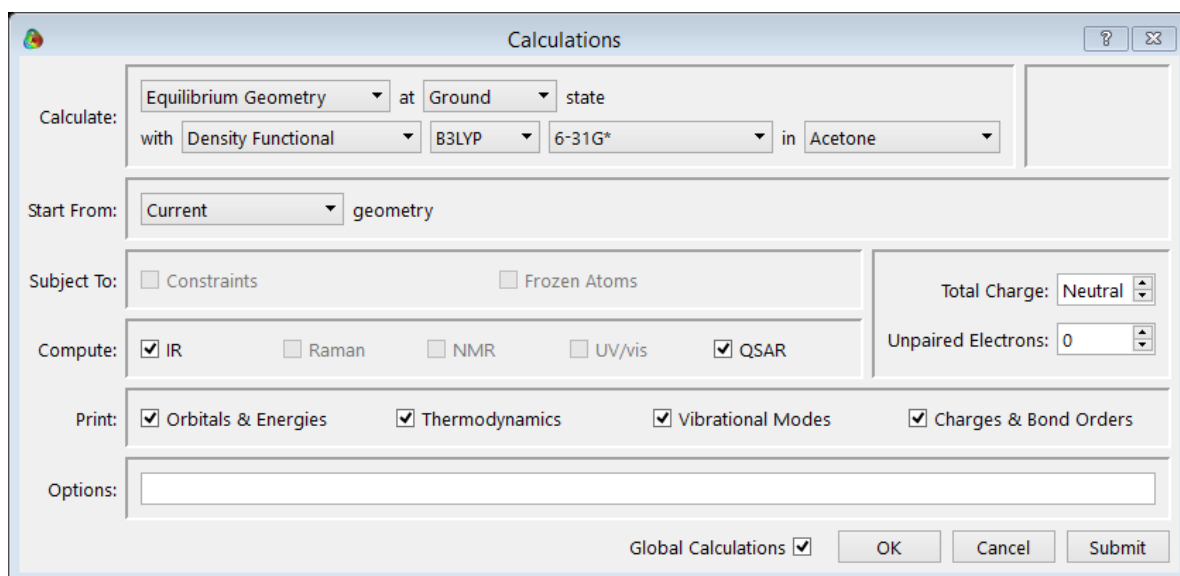


Plate 3.15: Calculations window showing completely set DFT calculation

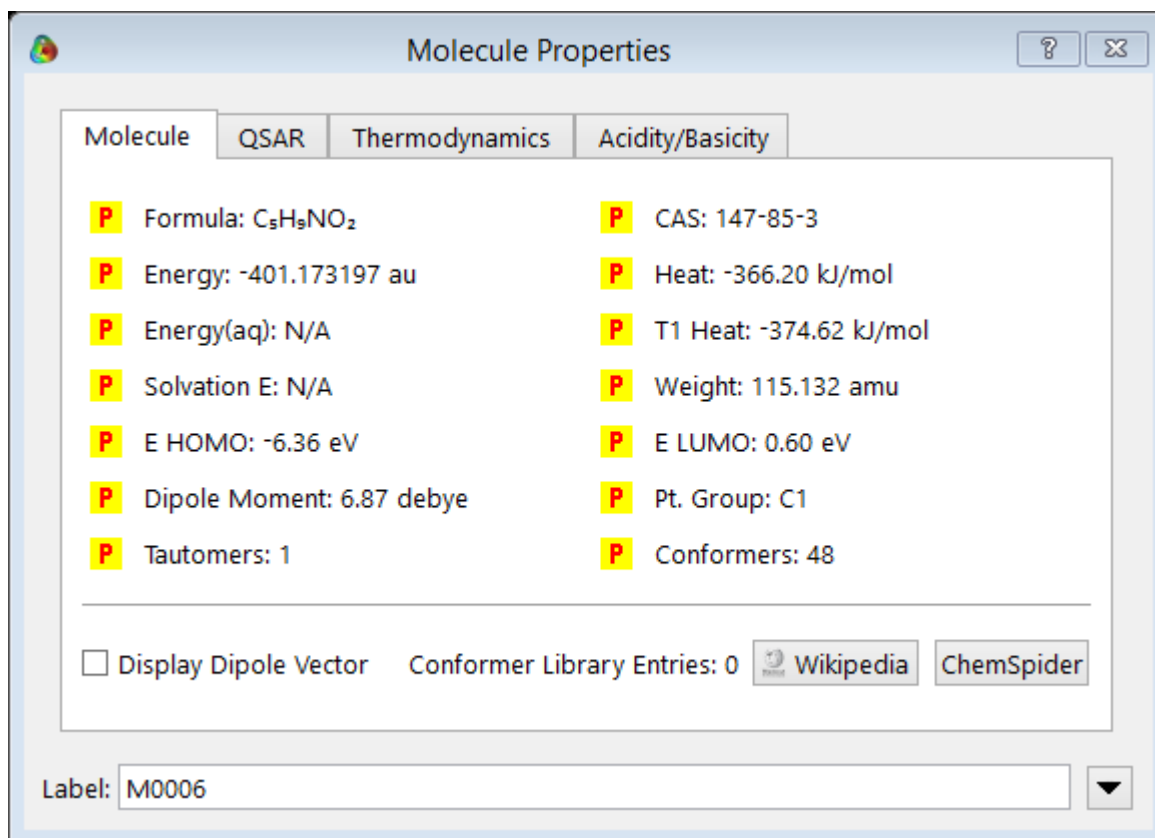


Plate 3.16: Display window showing the DFT calculation.

This procedure consumes a lot of time. It, for instance, took between 72 hours to 346 hours to compute a single transition state geometry for this work, using a core i5/12.00GB RAM personal computer.

3.2.3 Generating descriptors using PaDel-descriptor

The software PaDel-descriptor was launched to display the interface as seen in Plate 3.17. A directory containing the molecular structure file was selected by clicking on the molecular directory/file on the window. The most common file format that is supported by the software is SDF, which was used in this work. The file to save the calculated descriptors was selected by clicking on the descriptor output file on the interface. The options 0D, 1D, 2D, 3D were clicked successively to get the descriptors of choice. Plate 3.18 shows the interface completely set for calculation. The ‘start’ button at the bottom right of the interface was clicked to perform the calculation. The descriptors were automatically saved in comma-separated value (CSV) file format. The calculated descriptors were obtained in the output file and explored with Microsoft Excel.

3.2.4 Variable elimination and selection

After the descriptors had been obtained, it was assumed that many of these descriptors were more or less co-linear, that could cause disturbances for many statistical analysis methods. In order to obtain well-behaved model, variable elimination and selection techniques had to be carried out. Variable elimination is the process where variables are physically removed from the dataset. Variables are eliminated for two reasons: firstly, if they have a small variance that is below some threshold values, secondly, the existence of redundancy (exact linear dependencies between subsets of the variables) and multicollinearity (high multiple correlations between subsets of the variables). On the other hand, variable selection is the process where one searches for the best descriptor combinations (Sofie, 2009).

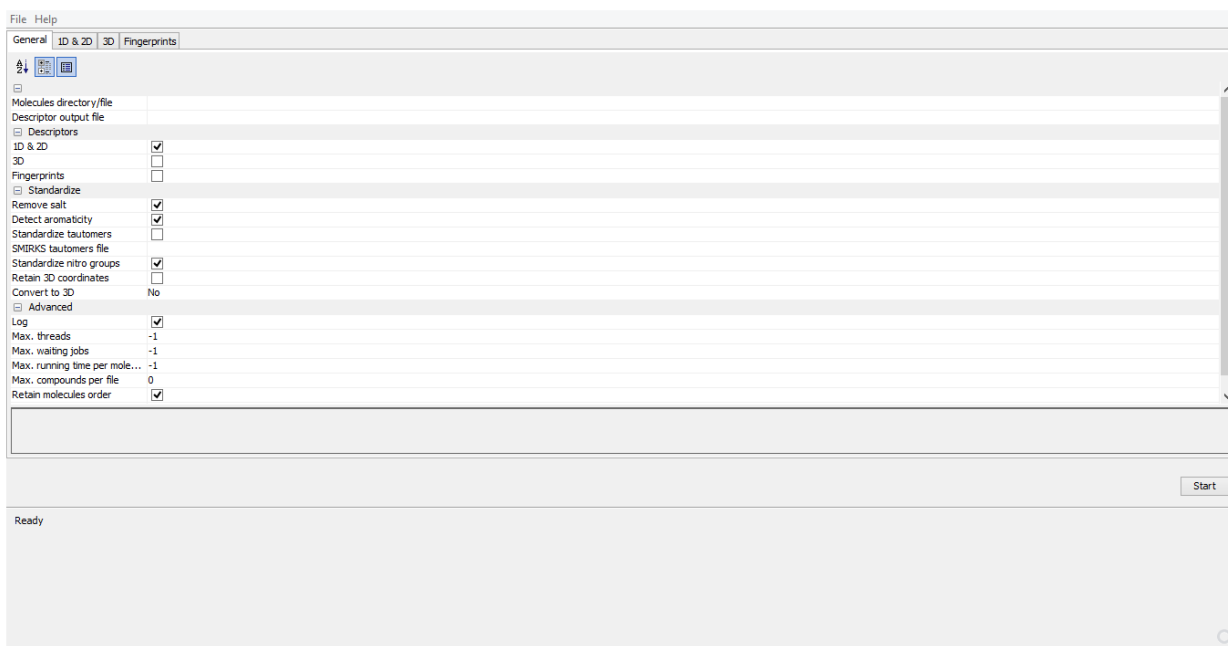


Plate 3.17: The Interface of PaDel-descriptor software

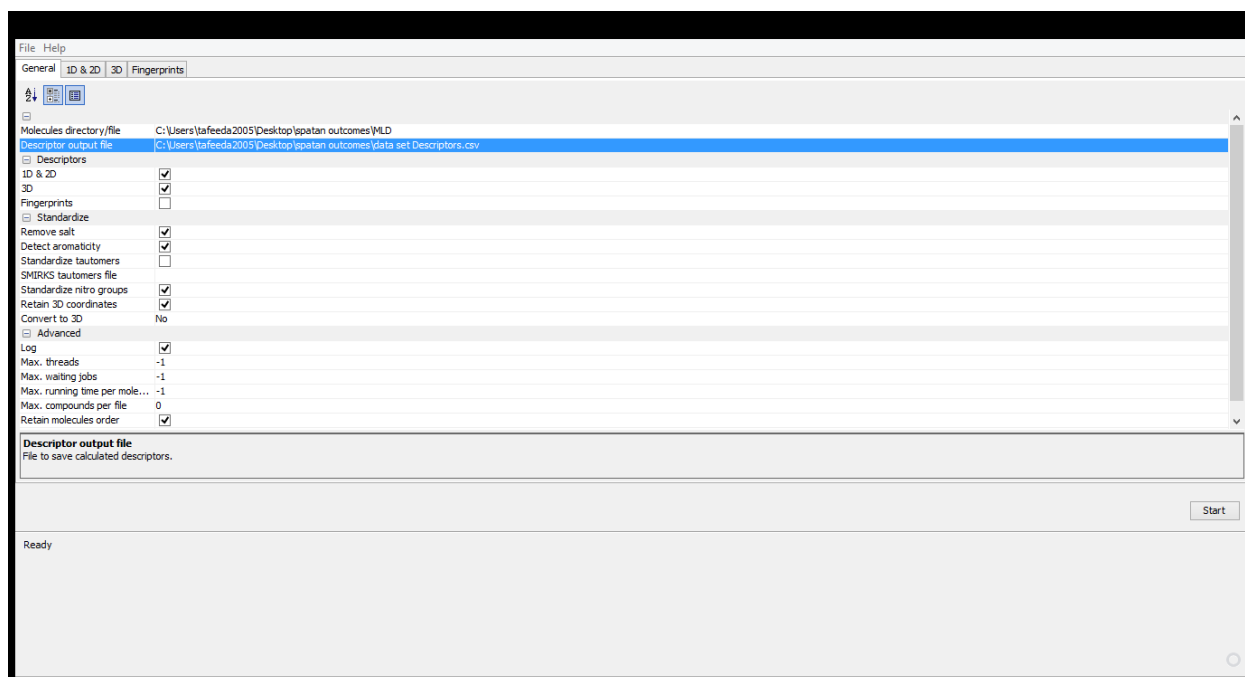


Plate 3.18: The Interface of PaDel-descriptorsoftware showing imported file and calculation set up.

The best descriptor combination was subjected to regression analysis in order to carry out correlation analysis between descriptors, and between descriptors and enantioselectivity values.

3.2.5 Data arrangement and data set division.

Microsoft excel was launched to display the spreadsheet. The names of all the descriptors in the topmost raw were arranged, thereby making the headings of the columns and the serial numbers of the compounds were entered in the first column. The values of the respective descriptors were entered in appropriate cells for compound 1. This was repeated for the entire compounds in the dataset. The molecules for the test set were selected and saved in a new page by means of copy and paste.

3.2.6 Model development using Materials studio 8.0

Genetic Function Approximation (GFA) statistical tool was used to generate the QSPR equations (models). These are mathematical relationship between sets of descriptors (independent variables), x_{ij} and the enantioselectivity values (dependent variables), y_i of the compound i. The linear relationship relating descriptors, $x_{i1}, x_{i2}, x_{i3} \dots$ to enantioselectivity values, y_i , is given as:

$$y_i = x_{i1}m_1 + x_{i2}m_2 + \dots + x_{in}b_n + e_i \quad (3.1)$$

Where m is the linear slope expressing the correlations between descriptors x_{ij} with the enantioselectivity values, y_i of the compound i, and e_i is a constant..

The Materials Studio software was launched by double-clicking on the icon. The software opened with a dialog box titled 'Welcome to Materials Studio' (Plate 3.19) that allows the user to either create a new project or load an existing project from this dialog. 'Create a new project' was selected and the OK button was clicked. This opened the New Project dialog box. The file location where the project would be stored was selected and the project name was entered as 'enantioselectivity model' in the File name field, then the OK button was clicked.

This got the user interface ready for the project as seen in Plate 3.20. The 'new' button on the tool bar was clicked to display a dialog from which 'study table' was selected and the 'OK' button was clicked. A spreadsheet was displayed and the training dataset table was added to the spreadsheet from Microsoft Excel by means of 'copy and paste' (Plate 3.21). Having transferred the data, the entire data was highlighted and the 'statistics' button on the menu bar was clicked. The 'initial analysis' from the pop up menu was clicked to display another pop up menu from which 'correlation matrix' was selected. From the correlation matrix generated (Plate 3.22), the descriptors that display high correlation (0.7 and above) with the ee value were identified.

To build the model, only the ee column was highlighted on the data table after clicking the 'statistics' button on the menu bar. The 'model building' from the pop up menu was clicked to display another pop up menu from which 'Genetic Function Approximation' was selected. This opened up a dialog with the title 'Genetic Function Approximation' as shown in Plate 3.23 with a selection menu right below the title from which 'variable selection' was already selected. The list of the descriptors displayed was followed and all those descriptors already considered to have high correlation with the ee were checked. The 'equation data' option from the selection menu was then selected and the field for 'additional items' was set as 'none' whereas, the fields of 'initial equation length' and the 'maximum equation length' were set to be the minimum and maximum numbers of descriptors respectively, desired to appear in the model. The rest of the fields were left unchecked and the next item 'parameters' on the selection menu was checked. The 'number of top equations returned' field was set to be '5' as the number of models desired to be generated. The 'OK' button was hit to generate the QSPR models (Plate 3.24).

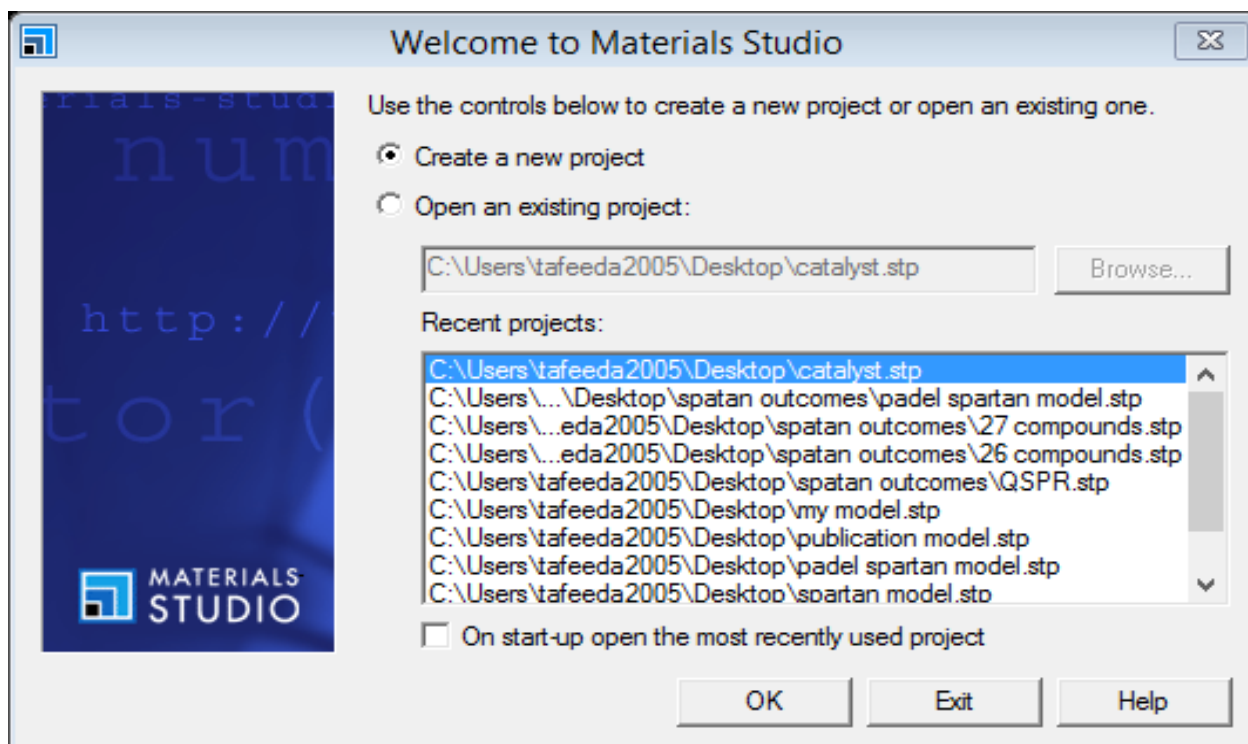


Plate 3.19: Opening window of the materials studio

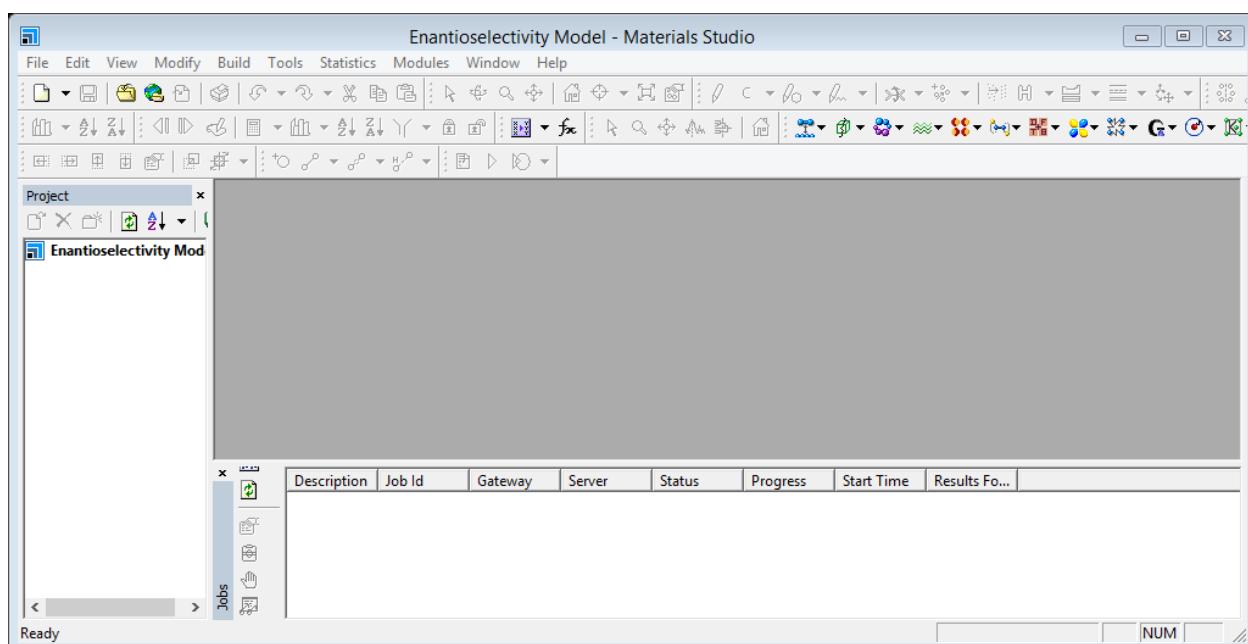


Plate 3.20: Materials studio interface showing a blank project.

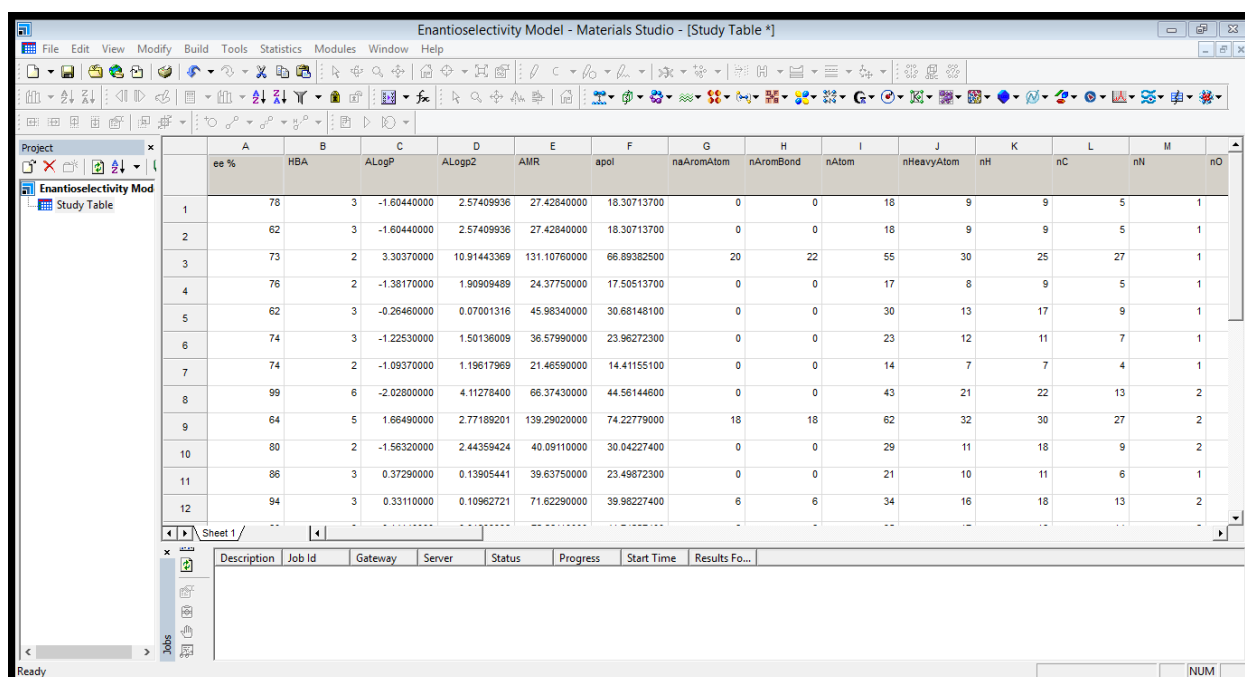


Plate 3.21: Materials studio interface showing a data set on the study table.

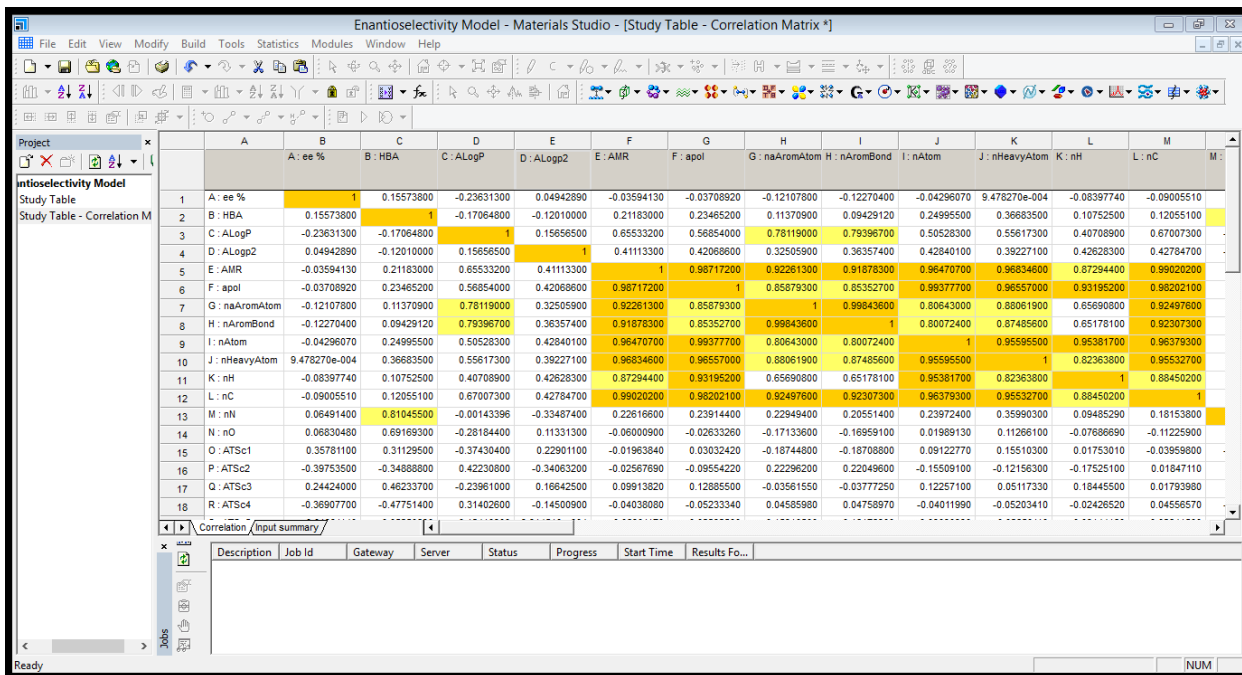


Plate 3.22: Materials studio interface showing a correlation matrix.

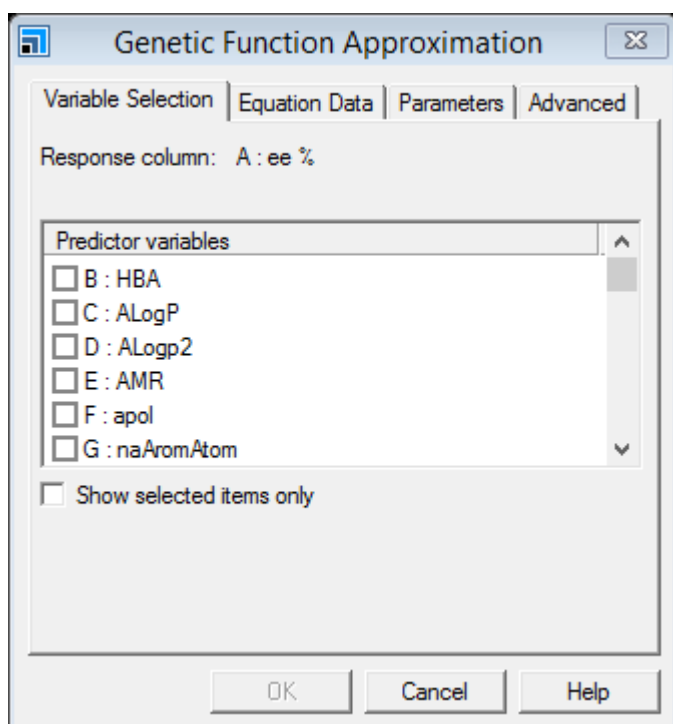


Plate 3.23: Genetic Functional Approximation window.

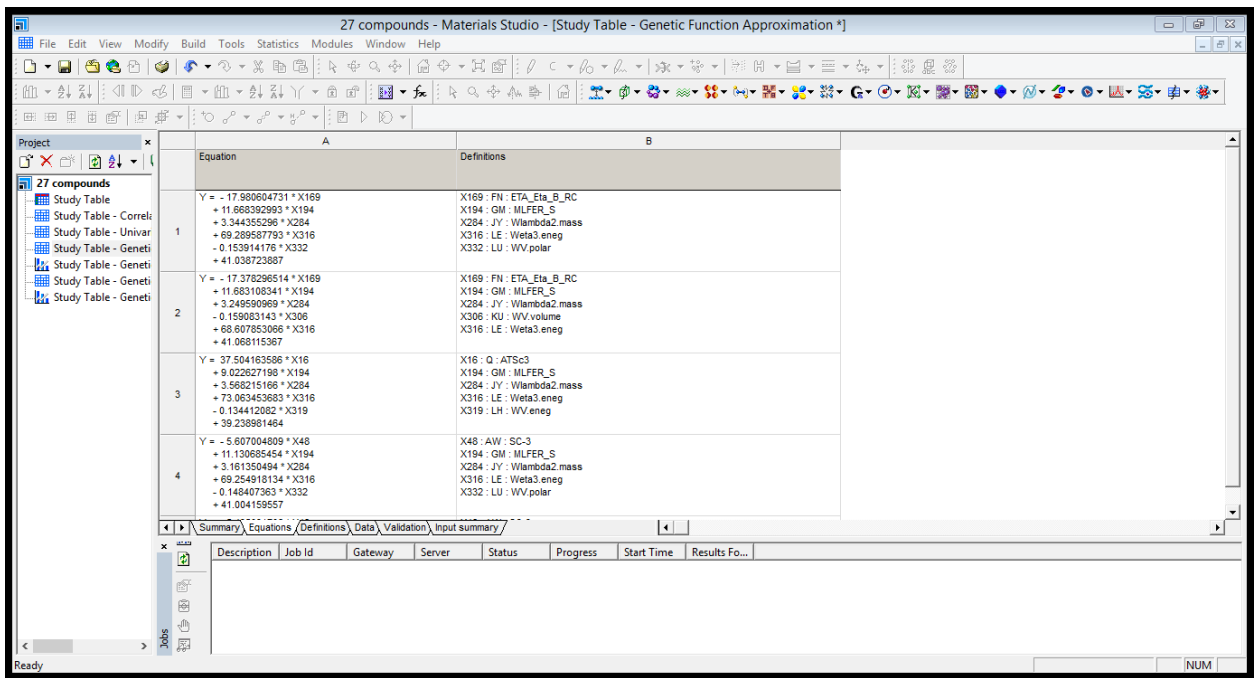


Plate 3.24: Materials studio interface showing generated model equations.

3.2.7 Model validation

The models developed were subjected to series of statistical tests to determine the robustness of the models and to prove that the correlation established was not as a result of happenstance.

3.2.7.1 *Internal validation*

Internal validation was carried out on the model using the compounds in the training set to ascertain the quality of the model, including how well it fitted the data, and to an extent, its predictive power (Konovalov *et al.*, 2004). In this regard, several parameters were calculated using Genetic Functional Algorithm method as built in in Material Studio 8.0, some of which include: R-squared, Adjusted R-squared and Cross validated R-squared..

3.2.7.2. *Cross validation techniques (external and internal)*

Since the real utility of a QSAR is its ability to accurately predict the modelled activity for new chemicals and interpret the model from the point of view of influential descriptors, a realistic assessment of the power of the model is necessary for a confident application. To accurately predict our model, we will have to apply cross validation technique which is defined as the techniques where a number of reduced data sets are created by deleting in each case, one or a small group of compounds from the data. The reduced data set is used as training set to develop a model while the remaining data are used to evaluate the model predictability. These excluded compounds however are still being used in the construction of the final model, so that the inclusion of their information in the analysis is ensured (Benigni and Bossa, 2008).

Leave-Many-Out (LMO) cross validation techniques in which number of the reduced data (M) was created by deleting in each case a small group of compounds (I') of at least 5 to 10 compounds out of the total 35 compounds (I). At each step, all the compounds belonging to this cancellation block (I'), were left out from the derivation of the model. The numbers of models (G) were constructed based on the number of the compounds in the reduced set (M). Because the number of the possible models increases exponentially with the number of compounds in the cancellation block (I'), it becomes impossible to construct systematically all

possible training sets. Therefore, at each run, the compounds were assigned randomly to the cancellation block (I') taking the number of randomly constituted cancellation block big enough, it was ensured that each compound had been left out at least once. The model predictive power is given by equation 3.2 (Benigni and Bossa, 2008).

$$q_{LMO}^2 = \sum_{i=1}^G \frac{\sum_{i=1}^M (y_i - \hat{y}_i)^2}{\sum_{i=1}^M (y_i - \bar{y})^2} \quad (3.2)$$

Where M is the number of compounds left out, y_i is the observed dependent variable, \bar{y} is the mean value of the dependent variable and \hat{y}_i is the calculated dependent variable

3.2.8 Designing the new catalyst.

The catalyst design was achieved by adjusting the values of the descriptors that appeared in the best model from the results of internal and external validation (Wold *et al.*, 1995). Increasing the values of the positive coefficients of the descriptors and decreasing the values of the negative coefficients of the descriptors would increase the ee value. The structural moieties described by each descriptor were identified. Proline, which is the commonest catalyst used for the selected reaction was chosen as the starting point (List, 2002; Ranklin *et al.*, 2002; Ceotto *et al.*, 2016). Adjustments were made on such moieties without totally distorting the chemistry of the compound to achieved improved values of the descriptors, and the new structure that emerged from such adjustments was presented as the new catalyst candidate

3.2.9 Predicting the enantioselectivity of the catalyst.

The structure of the candidate catalyst was built in Spartan 14 and set for the optimization as in 3.3.2. The medium of reaction was selected to be acetone in which the calculation was run. Such calculation was repeated for DMSO and DMF as the media of the reaction. The values of the descriptors that appeared in the selected model for the three optimization results were recorded. The values were then substituted in the model equation and the enantioselectivity values of the candidate catalyst in the three different media were as such deduced.

3.2.10 Determination of the mechanism of the reaction.

The structure of acetone was built and optimized as in section 3.2.2.2. Then 'setup' option was located from the menu bar and 'surfaces' was selected from the pop up menu. The 'add' button was located on the bottom left of the dialog box that appeared and was clicked. A pop up menu appeared from which the fields 'HOMO', 'LUMO', 'Electrostatic potential map' and '|LUMO| map' were checked. The fields appeared in the dialog box and the graphic calculations they represent ran automatically. Next to the 'add' button was 'more surfaces' which was clicked to display 'add surfaces' window. From the window, 'surfaces' was set as 'density' and 'property' was set as '|HOMO|' and 'apply' button at the bottom of the window was clicked. This gives additional field in the surfaces dialog box as density, |HOMO| whose graphic calculation, like the other fields, equally ran automatically. Plate 3.25 shows the surfaces window displaying completed graphic calculations. The 'Ok' button at the bottom of the 'add surface' window was clicked and the window disappeared. The |LUMO| map field was checked and the map appeared on the acetone molecule. The region of highest concentration of bluish colour was observed on the map and the atom located in the region was recorded as the centre of highest concentration of LUMO. The field was unchecked and the density, |HOMO| field was equally checked to display the map on the acetone molecule. The region of highest concentration of bluish colour on the compound was observed and the atom located in the region was recorded as the centre of highest concentration of HOMO. The same procedure was repeated for the 4-nitrobenzaldehyde and the proline catalyst and the shortest HOMO –LUMO gap was determined to be between HOMO of acetone and LUMO of proline as demonstrated in Figure 3.1.

Plate 3.25: Spartan interface showing surfaces calculation.

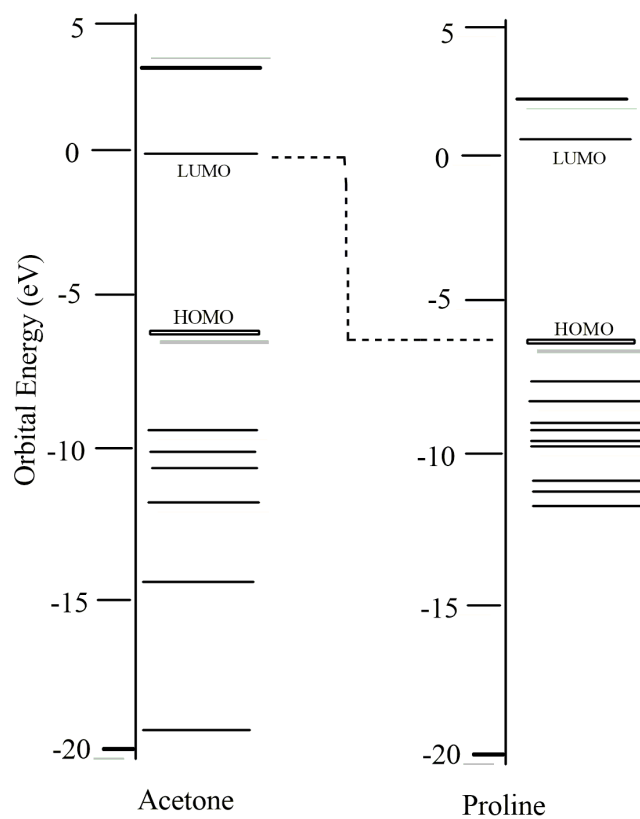


Fig. 3.1: Energy gap diagram between the frontier orbitals of acetone and proline catalyst.

To start the reaction, the reacting species that would collide to kick off the reaction were copied and pasted on a new SPARTAN interface. In this case, the reaction began by binding the proline to the acetone. From the menu bar, 'search' was clicked and 'Transition state' was selected from the pop up menu. The atom with the highest concentration of HOMO on proline was clicked and the atom with the highest concentration of LUMO on the acetone was clicked, the first atom was then clicked back. A dotted line joining the two atoms appeared with an arrow curved from the donor atom to the dotted line. This operation was repeated to connect the carbonyl oxygen of the acetone and the hydrogen atom attached to the nitrogen atom of the proline molecule.

To cleave a bond, the bond was clicked followed by clicking the atom that would accept the moved electrons. This resulted in a curved arrow being drawn from the centre of the bond to the atom accepting the moved electrons. This operation was performed to break the carbonyl π -bond on the acetone and the single bond between hydrogen and nitrogen on the proline. After all the arrows have been properly designated, a cyclic button on the right of the screen was selected to replace the conflated species with a guessed transition state structure. This transition state search step is illustrated in Figure 3.2. The activated complex was submitted for optimization, setting the calculation same as in the normal molecules but only changing the option of 'equilibrium geometry' with 'transition state geometry'. After the calculation was completed, 'spectra' was selected from the display menu to add 'IR calculated'. This displays the IR plot of the transition state structure and a list of frequencies at the left hand side of plot as seen in Plate 3.26. From the list of the frequencies, one and only imaginary frequency was identified to confirm that the structure searched was really a transition state. The same procedure was followed to search all the transition states involved in the reaction.

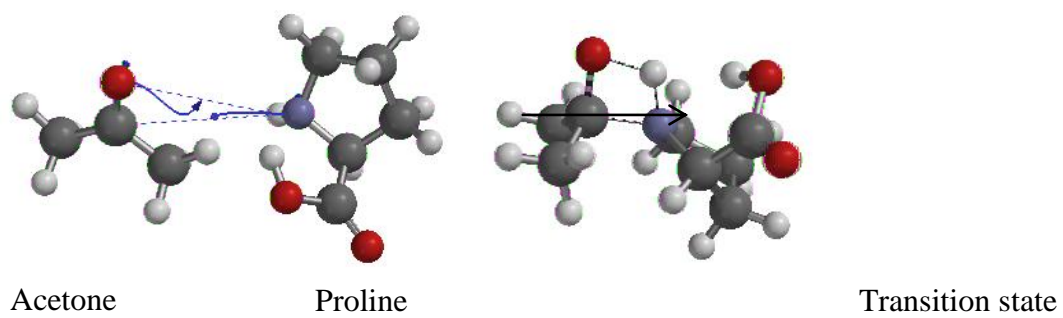


Fig. 3.2: A typical transition state search for the reaction between acetone and proline

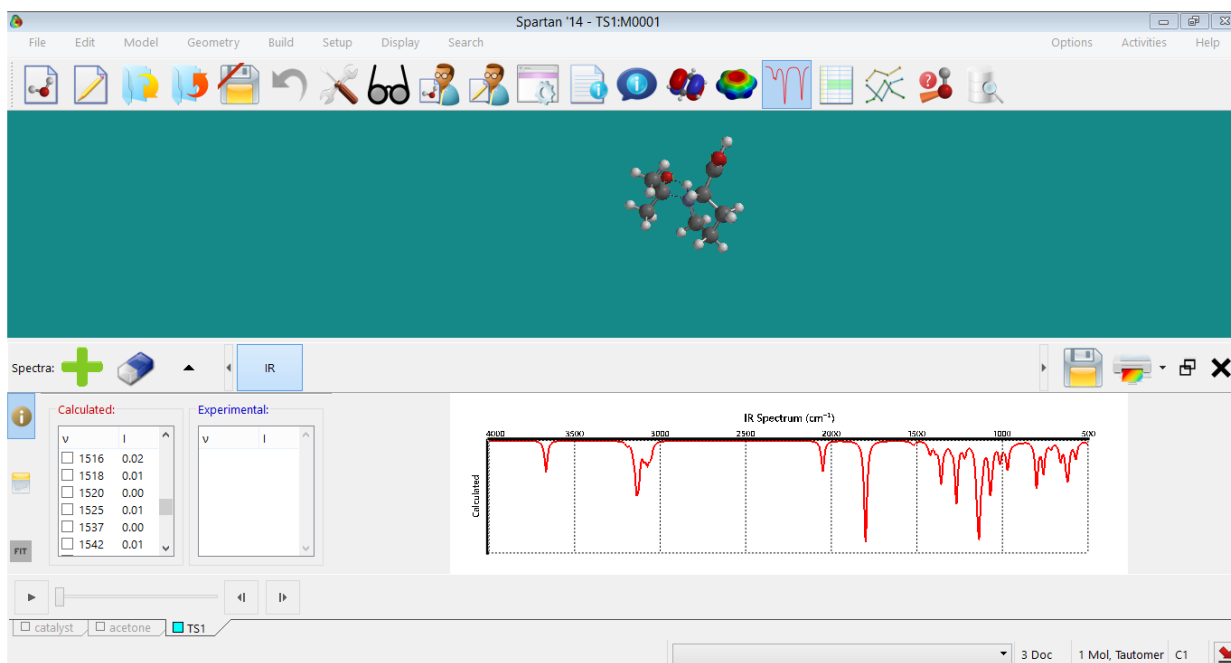


Plate 3.26: Spartan interface showing IR plot

This procedure was repeated for the same system investigated thereof using the designed catalyst in place of proline.

3.2.11 Determination of the catalyst effect on the rate of reaction.

After successful optimization of the entire species involved in the reaction, making sure that while setting the calculation, the ‘thermodynamics’ field was checked, the values of enthalpy, entropy and Gibbs free energy of all the species were determined. “Molecular properties” on the Display” menu was clicked and ‘Thermodynamics’ was selected from the window that appeared to view the thermodynamic properties; H^0 , S^0 and G^0 . The Eyring equation was applied to obtain the rate constant of each step

The heat of formation values obtained from the computation were calculated at 1 atmospheric pressure and temperature of 298.15K. They were used in calculating the enthalpies of activation, entropies of activation and Gibb’s energy of activation as formulated in the equations 3.3, 3.4 and 3.5, respectively (McQuaid *et al.*, 2004).

According to Engel and Reid, 2006, for transformation occurring from reactants to intermediates or products through transition state (activated complex), they have to attain the enthalpy of activation, entropy of activation and Gibb’s energy of activation. Equation 3.3 shows the expression for enthalpy of activation.

$$\Delta_r H^\#(298.15K) = \sum_{trans} n_{trans} \Delta_f H_{trans}^o(298.15K) - \sum_{react} n_{react} \Delta_f H_{react}^o(298.15K) \quad (3.3)$$

Where n_{react} and n_{trans} are the stoichiometric coefficients of the reactants and transition state respectively, while, at specified conditions, $\Delta_f H_{react}^o$, $\Delta_f H_{trans}^o$ and $\Delta_r H^\#$ are the standard heat of formation of reactants, standard heat of formation of transition state and enthalpy of activation respectively.

The entropy of activation was evaluated using the expression in equation (3.4).

$$\Delta_r S^\#(298.15K) = \sum_{trans} n_{trans} S_{trans}^o(298.15K) - \sum_{react} n_{react} S_{react}^o(298.15K) \quad (3.4)$$

where n_{react} and n_{trans} are the stoichiometric coefficients of the reactants and transition state respectively, while, at specified conditions, $\Delta_f S_{react}^o$, $\Delta_f S_{trans}^o$ and $\Delta_r S^\#$ are the standard entropy of formation of reactants, standard entropy of formation of transition state and activation entropy.

The Gibbs energy of activation was evaluated using the expression as presented in equation (3.5).

$$\Delta_r G^\# = \left\{ \sum_{trans} n_{trans} \Delta_f H_{trans}^o(298.15K) - \sum_{react} n_{react} \Delta_f H_{react}^o(298.15K) \right\} - T \left\{ \sum_{trans} n_{trans} S_{trans}^o(298.15K) - \sum_{react} n_{react} S_{react}^o(298.15K) \right\} \quad (3.5)$$

Activation energies for all the elementary steps were calculated using equation 3.6, which is applicable to both unimolecular and bimolecular solution-phase reactions (Puri *et al.*, 2014). Reaction rate evaluations were accomplished using transition state and Arrhenius theories as combined in Eyring equation given in equation 3.7 (Eyring, 1935).

$$E_a = \Delta H^\ddagger + RT \quad (3.6)$$

$$k(T) = e^{\frac{K_B T}{hc^0}} e^{\frac{\Delta S^\ddagger}{R}} e^{-E_a/RT} \quad (3.7)$$

Where E_a is the activation energy, h is the Planck's constant (6.626176×10^{-34} Js), R is molar gas constant ($8.31441 \text{ J mol}^{-1} \text{ K}^{-1}$), $C^o =$ concentration (taken to be unity), T is temperature (at 298.15K), k_B is Boltzmann constant ($1.38066 \times 10^{-23} \text{ JK}^{-1}$) and $k(T)$ is the rate constant at specified temperature. Equation 3.7 is also applicable to both bimolecular and unimolecular reactions steps provided they take place in a solution medium.

CHAPTER FOUR

4.0 RESULTS

4.1 Model Development

Using the values of the descriptors obtained from the exercises discussed in subsections 3.2.3 and 3.2.4, five regression models developed in form of equations are presented in subsection 4.4.1.

4.1.1 Models

The five QSPR models generated from the calculated descriptors are as follows:

$$\text{Model 1. } Y = (1.673122128 \times nC) - (34.176575810 \times VP-6) + (70.425593130 \times \text{maxssNH}) - (1.801773791 \times \text{PNSA-3}) + (0.063795412 \times \text{DPSA-1}) - 197.484885086$$

$$\text{Model 2. } Y = (1.457329708 \times nBondsS) - (34.951025741 \times VP-6) + (60.950716170 \times \text{maxssNH}) - (34.401517276 \times \text{ETA_Eta_B}) - (1.687263947 \times \text{PNSA-3}) - 167.808178222$$

$$\text{Model 3. } Y = -(30.377930179 \times VP-6) + (68.905751565 \times \text{maxssNH}) - (0.131476510 \times \text{TopoPSA}) - (1.413558796 \times \text{PNSA-3}) + (0.053016616 \times \text{WPSA-2}) - 159.764707999$$

$$\text{Model 4. } Y = -(60.413692411 \times VP-7) + (89.495366411 \times \text{maxssNH}) + (0.027050312 \times \text{WPATH}) - (0.145207436 \times \text{PNSA-1}) - (1.948394805 \times \text{PNSA-3}) - 223.289256845$$

$$\text{Model 5. } Y = (0.008207003 \times \text{ATSp4}) - (32.856482331 \times VP-6) + (69.613408769 \times \text{maxssNH}) - (1.705998619 \times \text{PNSA-3}) + (0.061866665 \times \text{DPSA-1}) - 184.459618277$$

In each equation, the dependent variable, Y, stands for the enantiomeric excess (ee) value, while the independent variables are a set of descriptors each accompanied by a coefficient constant. The full names of the descriptors in the models are given below.

nC: Number of carbon atoms

VP-6: Valence path, order 6

DPSA-1:	Difference in charged partial surface area
MaxssNH:	Maximum value of electrotopological state –NH-
PNSA-3:	Path weighted partial negative surface area
ETA_Eta B:	Branching index, Eta B
nBondsS:	Number of single bonds (including bonds with hydrogen)
TopoPSA:	Topological polar surface area
WPSA-2:	Partial positive surface area * total positive charge on the molecule
VP-7:	Valence path, order 7
WPATH:	Weiner path number
PNSA-1:	Partial negative surface area -- sum of surface area on negative parts of molecule for total molecular surface area /1000
ATSp4:	Autocorrelation descriptor, weighted by polarizability.

4.1.2: Comparison between actual and calculated values of ee for the training set.

The enantiomeric excess (ee) values calculated by the models were compared with the actual values, and the residual values were obtained as presented in Tables 4.1a and 4.1b. Table 4.1a compares the actual ee values of all the compounds in the training set with their respective calculated ee values from equations 1 and 2, while Table 4.1b compares with the calculated values from equations 3 to 5. The graphical presentations of these exercises are given in Figure 4.1.

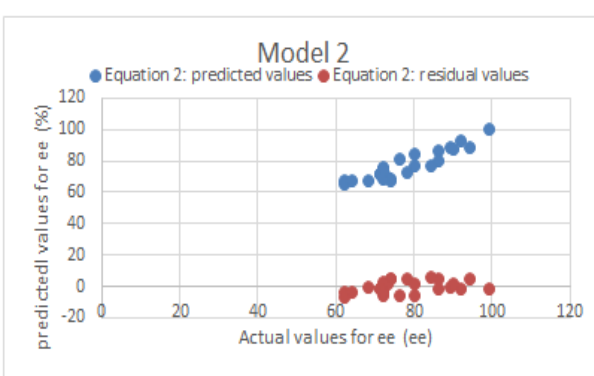
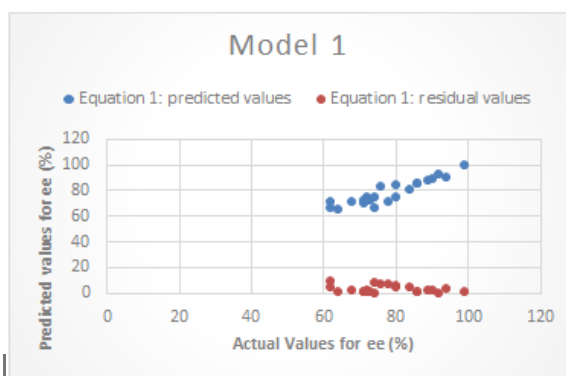
Table 4.1a: Comparison between actual and calculated values of ee for the training set using equations 1 and 2.

Catalysts	Actual values for ee (%)	Equation 1: predicted values	Equation 1: residual values	Equation 2: predicted values	Equation 2: residual values
<i>1</i>	86.00000000	84.77142500	1.22857500	87.18338100	1.18338100
3	78.00000000	71.06904300	6.93095700	72.81329500	5.18670500
4	62.00000000	66.04280700	4.04280700	68.07426100	6.07426100
5	73.00000000	72.36274700	0.63725300	70.38391400	2.61608600
6	76.00000000	83.25390900	7.25390900	81.31845300	5.31845300
7	62.00000000	71.18654100	9.18654100	65.70952600	3.70952600
8	74.00000000	74.07070300	0.07070300	69.22863200	4.77136800
9	74.00000000	65.74219300	8.25780700	68.41781500	5.58218500
10	99.00000000	99.80747000	0.80747000	100.3172710	1.31727100
11	64.00000000	65.11237700	1.11237700	67.58558500	3.58558500
12	80.00000000	74.37885400	5.62114600	77.90631100	2.09368900
16	86.00000000	85.55908300	0.44091700	81.08736300	4.91263700
17	94.00000000	90.44437800	3.55562200	89.08467300	4.91532700
18	80.00000000	84.23373900	4.23373900	84.96300200	4.96300200
20	84.00000000	80.13773100	3.86226900	77.19156400	6.80843600
21	92.00000000	91.93447300	0.06552700	92.96868400	0.96868400
23	71.00000000	70.28842700	0.71157300	71.86578600	0.86578600
24	89.00000000	87.59347200	1.40652800	88.79630500	0.20369500
26	71.00000000	70.75326400	0.24673600	72.31015200	1.31015200
27	72.00000000	74.38411000	2.38411000	76.82099800	4.82099800
28	72.00000000	72.32396100	0.32396100	68.93460400	3.06539600
29	72.00000000	73.16351800	1.16351800	75.53268000	3.53268000
30	71.00000000	71.71151500	0.71151500	72.10873500	1.10873500
31	72.00000000	72.80154200	0.80154200	75.51918100	3.51918100
32	90.00000000	88.50621800	1.49378200	87.63690800	2.36309200
33	68.00000000	70.36649700	2.36649700	68.24092200	0.24092200

Catalysts	Equation 3: predicted values	Equation 3: residual values	Equation 4: predicted values	Equation 4: residual values	Equation 5: predicted values	Equation 5: residual values
<i>1</i>	82.28177600	3.71822400	83.56642800	2.43357200	84.18377800	1.81622200
3	71.17233300	6.82766700	73.83339100	4.16660900	72.00247600	5.99752400
4	66.72676000	4.72676000	69.66242400	7.66242400	67.24420300	5.24420300
5	70.62283600	2.37716400	70.61476000	2.38524000	69.16973800	3.83026200
6	83.97362300	7.97362300	83.04003200	7.04003200	83.60185000	7.60185000
7	69.03547800	7.03547800	68.22257000	6.22257000	69.54555200	7.54555200
8	71.73547700	2.26452300	72.17821400	1.82178600	73.68535700	0.31464300
9	71.42968100	2.57031900	67.80308100	6.19691900	67.67658100	6.32341900
10	99.98902200	0.98902200	101.1373990	2.13739900	96.98974800	2.01025200
11	70.15463900	6.15463900	68.25369500	4.25369500	67.38580500	3.38580500
12	75.29125000	4.70875000	74.78731900	5.21268100	73.23170500	6.76829500
16	86.40364900	0.40364900	86.71780100	0.71780100	88.52394500	2.52394500
17	90.52249000	3.47751000	93.41992400	0.58007600	90.22546200	3.77453800
18	84.55119400	4.55119400	84.60055600	4.60055600	85.72068000	5.72068000
20	78.12488000	5.87512000	79.07934300	4.92065700	82.66278300	1.33721700
21	91.04671500	0.95328500	85.68415600	6.31584400	91.13151100	0.86848900

23	69.23386300	1.76613700	71.08723800	0.08723800	70.85708800	0.14291200
24	86.53921100	2.46078900	88.45111900	0.54888100	90.32581500	1.32581500
26	72.02479200	1.02479200	67.66126000	3.33874000	68.26993300	2.73006700
27	74.99876900	2.99876900	72.86834300	0.86834300	73.31954900	1.31954900
28	69.71951300	2.28048700	73.15357500	1.15357500	72.84226600	0.84226600
29	73.46164100	1.46164100	70.95262000	1.04738000	72.28460200	0.28460200
30	70.65307700	0.34692300	71.76602500	0.76602500	71.40439100	0.40439100
31	71.20552100	0.79447900	74.51353200	2.51353200	74.03073800	2.03073800
32	92.36577000	2.36577000	89.58637700	0.41362300	86.97496800	3.02503200
33	68.73604100	0.73604100	69.35881800	1.35881800	68.70947700	0.70947700

Table 4.1b: Comparison between actual and calculated values of ee for the training set using equations 3 to 5.



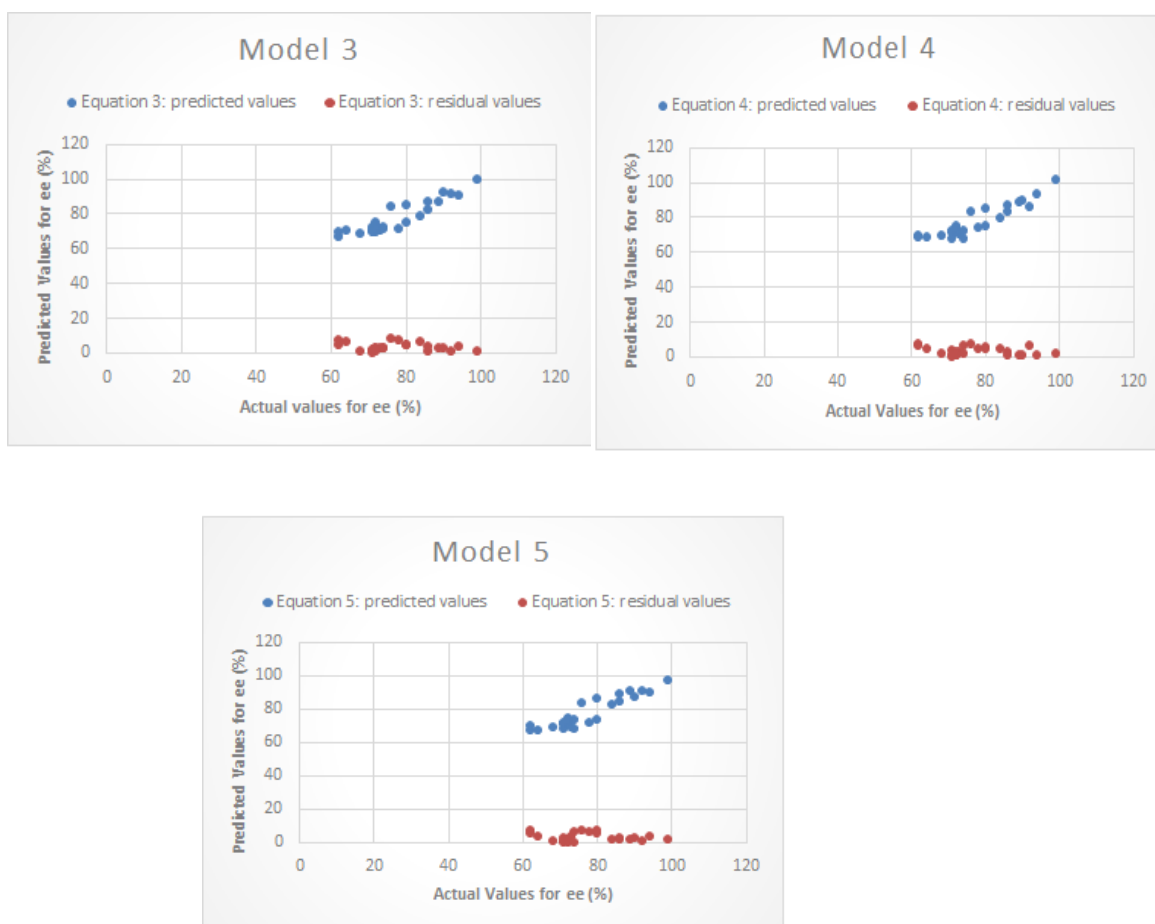


Fig. 4.1: Plot of ee values of the training set calculated by the five equations as function of experimental ee values.

4.2 Model Validation

4.2.1 Internal validation

The results of internal validation as computed by the material studio are presented in Table 4.2. The values of Correlation coefficient of multiple determination (R^2), Adjusted R^2 , Cross validated R^2 , Significance-of-regression (SOR) F- value and minimum EE for non-significant LOF are given for all the five model equations.

4.2.2 External validation

The results of external validation are presented in Table 4.3. The values of Q^2 and R_{ext} are given as calculated manually, for all the five model equations.

4.3 Catalyst Design

Two candidate catalysts were proposed using the selected model equation (equation 1) with the structures presented in Figures 4.1 and 4.2, respectively. The candidate catalysts were selected based on their improved ee values. The results of the ee values of the catalysts calculated using the selected model equation are given in Table 4.4, showing the values of the descriptors involved.

4.4 Optimized Structures

The optimized structures of all the species involved in the reactions investigated are presented in subsections 4.4.1 and 4.4.2.

4.4.1 Reaction species for proline-catalyzed aldol reaction

The optimized structures of the reaction species involved in proline-catalyzed aldol reaction system are presented in Figure 4.4. In this Figure, acetone is designated as A, 4-nitrobenzaldehyde is denoted as B and the catalyst (proline) is captioned C, while D designates water. Intermediates 1 to 6 are denoted by I₁, I₂, I₃, I₄, I₅ and I₆, respectively, while the transition structures 1 to 7 are designated as TS1, TS2, TS3, TS4, TS5, TS6 and TS7, respectively.

Table 4.2: Results of internal validation

Parameters	Equation 1	Equation 2	Equation 3	Equation 4	Equation 5
R-squared	0.85108000	0.84916600	0.84834900	0.84791200	0.84712400
Adjusted R-squared	0.81385000	0.81145700	0.81043700	0.80989000	0.80890600
Cross validated R-squared (R^2_{cv})	0.75853300	0.74419400	0.73659000	0.72975800	0.71562400
Significant Regression	Yes	Yes	Yes	Yes	Yes
SOR F-value	22.86008100	22.51914900	22.37643700	22.30052900	22.16506500

Critical SOR F-value (95%)	2.73293900	2.73293900	2.73293900	2.73293900	2.73293900
Min. EE for non-significant LOF (95%)	3.42572700	3.44767800	3.45699200	3.46197700	3.47092700

EE= experimental error

SOR=Significance of regression

LOF= Lack of fit

Table 4.3: Results of external validation

Models	Q^2_{ext}	R_{ext}
1	0.70	0.88
2	0.77	0.83
3	0.73	0.87
4	0.71	0.84
5	0.66	0.80

Fig. 4.2: Candidate catalyst 1:

Fig. 4.3: Candidate catalyst 2

Table 4.4: Calculated Enantiomeric Excess values for the candidate catalysts.

Candidate	nC	VP-6	MaxssNH	PNSA-3	DPSA-1	ee value (%)
1	9	0.412407	3.092136	-28.0615	221.8694	85

2 9 0.508986 3.057593 -29.4649 185.2646 80



Figure 4.4: Optimized geometries for the reaction species of proline-catalyzed aldol reaction system.

From the optimized transition states structures, the bond lengths of the partially formed and partially broken bonds were identified and the values are presented in Table 4.5a. Moreover, the Infra-Red spectra of the optimized transition state structures were plotted and the imaginary frequencies (ν) identified from the plots are given in brackets under the respective transition states in Table 4.5a.

4.4.2 Reaction species for aldol reaction catalyzed by the candidate catalyst.

The optimized structures (geometries) of the species involved in the aldol reaction catalyzed by the candidate catalyst are presented in Figure 4.5. In the Figure, acetone is designated as A, 4-nitrobenzaldehyde is denoted as B, and the candidate catalyst is denoted as C, while D designates water. The intermediates 1 to 6 are denoted by I₁, I₂, I₃, I₄, I₅ and I₆, respectively while the transition structures 1 to 7 are designated as TS1, TS2, TS3, TS4, TS5, TS6 and TS7, respectively.

From the optimized transition states structures, the bond lengths of the partially formed and partially broken bonds were identified and the values are presented in Table 4.5b. Moreover, the Infra-Red spectra of the optimized transition state structures were plotted and the imaginary frequencies (ν) identified from the plots are given in brackets under the respective transition states in Table 4.5b.

4.5 Graphical Models (Surfaces)

4.5.1 Frontier orbitals, density maps and Electrostatic potential maps for proline-catalyzed aldol reaction.

The frontier orbitals of the reacting species for the proline-catalyzed aldol reaction are presented in Figure 4.6a while the frontier orbitals for the intermediates are given in Figure 4.6b. Figure 4.7a shows the density maps of LUMO and HOMO for the reacting species involved in the reaction while Figure 4.7b presents the LUMO and HOMO density maps for the intermediates involved. Electrostatic potential maps for the reacting species and intermediates are given in Figure 4.8.

Table 4.5a: Bond Lengths of the Transition Bonds for the proline-catalyzed reaction between acetone and 4-nitrobenzaldehyde, and the Imaginary frequencies.

TS1	Bond Length (Å)	TS2	Bond Length (Å)	TS3	Bond Length(Å)	TS4	Bond Length (Å)
C==O	1.377	C==N	1.437	N==C	1.364	C==C	1.416
O---H	1.352	C---O	1.495	C---O	3.850	C---O	2.982
C---N	1.682	OH--H	1.643	(v =	1999 Hz)	C---H	1.386
N---H	1.217	O---H	1.003			O---H	1.240
(v =	1633 Hz)	(v =	1586 Hz)			(v =	1421 Hz)
TS5		TS6		TS7			
C==N	1.324	C==N	1.467	C---N	1.679		
C==C	1.496	N---H	1.517	C==O	1.332		
C---C	1.585	O---H	1.261	C=O--H	1.235		
C==O	1.396	C---O	1.537	C-O--H	1.096		
C=O--H	2.44	(v =	1983 Hz)	(v =	1275 Hz)		
C-O--H	0.975						
(v =	474 Hz)						

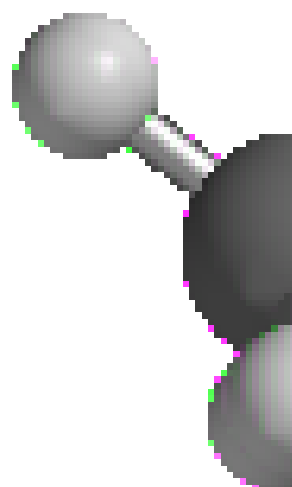


Figure 4.5: Optimized geometries for the reaction species of aldol reaction system catalyzed by the candidate catalyst

Table 4.5b: Bond Lengths of the Transition Bonds for the candidate-catalyzed reaction between acetone and 4-nitrobenzaldehyde, and the Imaginary frequencies.

TS1	Bond Length (Å)	TS2	Bond Length (Å)	TS3	Bond Length (Å)	TS4	Bond Length (Å)
C==O	1.365	C==N	1.471	N==C	1.348	C==C	1.417
O---H	1.451	C---O	1.428	C---O	5.827	C---O	2.762
C---N	1.583	OH--H	1.806	(v =	1996 Hz)	C---H	1.287
N---H	1.344	O---H	1.958			O---H	1.618
(v =	1951 Hz)	(v =	485 Hz)			(v =	830 Hz)
TS5		TS6		TS7			
C==N	1.325	C==N	1.513	C---N	1.928		
C==C	1.489	N---H	1.446	C==O	1.261		
C---C	1.581	O---H	1.242	C=O--H	1.797		
C==O	1.351	C---O	1.483	C-O--H	0.964		
C=O--H	1.178	(v =	1985 Hz)	(v =	1279 Hz)		
C-O--H	1.101						
(v =	863 Hz)						

Figure 4.6a: Frontier orbitals of the reacting species for the proline-catalyzed aldol reaction.



Figure 4.6b: Frontier orbitals of the intermediates for the proline-catalyzed aldol reaction.

Figure 4.7a: Density maps of LUMO and HOMO of the reacting species for the proline-catalyzed aldol reaction.

Figure 4.7b: Density maps of LUMO and HOMO of the intermediates for the proline-catalyzed aldol reaction.



Figure 4.8: Electrostatic potential map of the reacting species and intermediates for the proline-catalyzed aldol reaction.

In all the Figures, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, proline is denoted as C and water is denoted with D. Whereas the intermediates 1 to 6 are denoted with I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

4.5.2 Frontier orbitals, density maps and electrostatic potential maps for aldol reaction catalyzed by the candidate catalyst.

The frontier orbitals of the candidate catalyst and the intermediates for the candidate-catalyzed aldol reaction are presented in Figure 4.9. Figure 4.10 shows the density maps of LUMO and HOMO for the candidate and the intermediates involved in the reaction. Electrostatic potential maps for the candidate, water and the intermediates are given in Figure 4.11. In all the Figures, the candidate catalyst is denoted with C, water is denoted with D and the intermediates 1 to 6 are denoted with I₁, I₂, I₃, I₄, I₅ and I₆, respectively

4.6: Results of E-HOMO and E-LUMO Computations.

4.6.1: Results of E-HOMO and E-LUMO computations for proline-catalyzed system

The results of E-HOMO and E-LUMO computations for the proline-catalyzed aldol reaction of acetone and 4-nitrobenzaldehyde are given in Table 4.6. In the Table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, proline is denoted with C and water is denoted as D. The transitions states structures 1 to 7 are gives as TS1, TS2, TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

The HOMO and LUMO centres for the reaction are given in Table 4.7. In the Table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, proline is denoted with C and water is denoted as D. The intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

The energy gaps of the various combinations are given in Table 4.8. In the table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, proline is denoted with C, water is denoted with D and the product is represented by P. The intermediates 1 to 6 are denoted with

I₁, I₂, I₃, I₄, I₅ and I₆, respectively. The symbol of the specie bearing HOMO is written first, followed by the symbol of the species bearing LUMO in the combinations. The numerical figures in front of each combination are the values of the energy gap in (eV).

4.6.2: Results of E-HOMO and E-LUMO computations for candidate-catalyzed system

The results of E-HOMO and E-LUMO computations for the candidate-catalyzed aldol reaction of acetone and 4-nitrobenzaldehyde are given in Table 4.9. In the Table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, candidate catalyst is denoted with C, water is denoted as D and the product is represented by P. The transitions states structures 1 to 7 are gives as TS1, TS2, TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

The HOMO and LUMO centres for the reaction are given in Table 4.10. In the Table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, candidate catalyst is denoted as C and water is denoted as D. The intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

The energy gaps of the various combinations are presented in Table 4.11. In the Table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, candidate catalyst is denoted as C, water is denoted with D and the product is represented by P. The intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively. The symbol of the specie bearing HOMO is written first, followed by the symbol of the species bearing LUMO in the combinations. The numerical figures in front of each combination is the value of the energy gap in (eV).

C

Figure 4.9: Frontier orbitals of the intermediates involved in the candidate-catalyzed aldol reaction.

Catε

Figure 4.10: Density maps of LUMO and HOMO of the intermediates for the candidate-catalyzed aldol reaction.

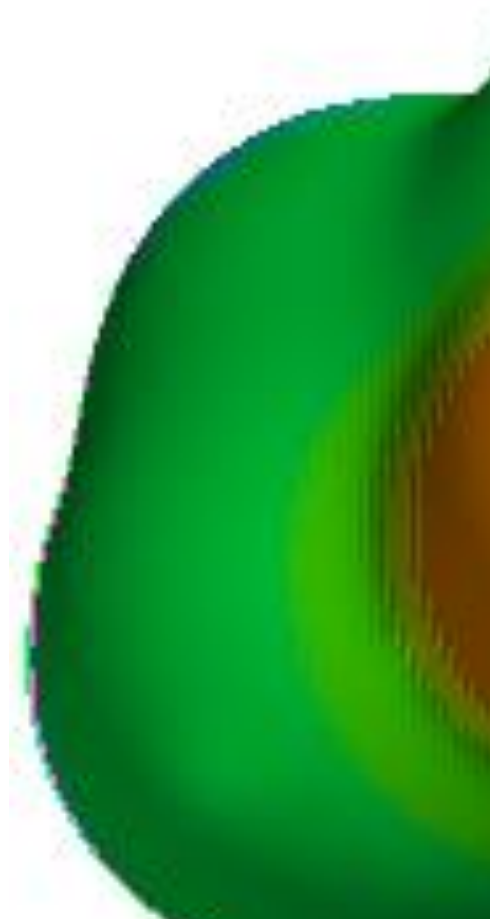


Figure 4.11: Electrostatic potential map of the candidate catalyst, water and intermediates for the candidate-catalyzed aldol reaction.

Table 4.6: Results of E-HOMO and E-LUMO computations for reaction with proline

Species	E-HOMO (eV)	E-LUMO (eV)
A	-6.56	-0.11
B	-7.18	-2.86
C	-6.36	0.60
D	-7.86	1.89
TS1	-5.44	0.04
I ₁	-5.65	-0.12
TS2	-6.05	0.37
I ₂	-5.44	-1.07
TS3	-6.28	-0.76
I ₃	-6.16	0.33
TS4	-5.63	-0.49
I ₄	-5.07	0.17
TS5	-5.23	-2.39
I ₅	-5.89	-2.41
TS6	-9.00	-4.04
I ₆	-5.53	-1.25
TS7	-6.21	-1.26
P	-6.79	-2.40

Table 4.7: Results of HOMO Centre and LUMO Centre computations for reaction with proline

Species	HOMO Center (nucleophile site)	LUMO Center (electrophile site)
A	=O	=C
B	=O-----C	=C-----O
C	-N-	=C
D	O	H
I ₁	-N-	=C, -C-
I ₂	O ⁻	=C
I ₃	O	H
I ₄	=C	H-----O
I ₅	=O-----C	=C-----N
I ₆	O ⁻	H-----O

Table 4.8: HOMO-LUMO energy Gaps for proline-catalyzed system.

Steps	HOMO-LUMO Gaps
1	A-C:7.16, C-A: 6.25, A-A: 6.45, C-C: 6.96
2	I ₁ -I ₁ : 5.53, I ₁ -A:5.54, A- I ₁ : 6.44, I ₁ -C: 6.25, C- I ₁ : 6.24
3	I ₂ -I ₂ : 4.37, I ₂ -A:5.33, A- I ₂ : 5.49, I ₂ -C: 6.04, C- I ₂ : 5.29
4	I ₃ - I ₃ : 5.83, I ₃ -A: 6.05, A- I ₃ : 6.23, I ₃ -C: 6.76, C-I ₃ : 6.03
5	I ₄ -I ₄ : 5.24, I ₄ -A: 4.96, A- I ₄ : 6.73, I ₄ -C: 5.67, C- I ₄ : 6.53, I ₄ -B: 2.21, B- I ₄ : 7.35
6	I ₅ - I ₅ : 3.48, I ₅ -A: 5.78, A- I ₅ : 4.15, I ₅ -C: 6.58, C-I ₅ :3.95, I ₅ -B: 3.03, B-I ₅ : 5.37, I ₅ -D: 7.78, D- I ₅ : 5.45
7	I ₆ - I ₆ : 4.28, I ₆ -A:5.42, A- I ₆ : 5.31, I ₆ -B: 3.03, B- I ₆ :5.93, I ₆ -C 5.59, C- I ₆ : 5.11, I ₆ -D 7.42, D-I ₆ : 6.61

Table 4.9: Results of E-HOMO and E-LUMO computations for reaction with candidate catalyst

Species	E-HOMO (eV)	E-LUMO (eV)
A	-6.56	-0.11
B	-7.18	-2.86
C	-6.52	2.07
D	-7.86	1.89
TS1	-9.74	1.29
I ₁	-6.08	-1.67
TS2	-8.64	1.48
I ₂	-5.31	-0.32
TS3	-7.19	0.75
I ₃	-5.11	-1.35
TS4	-8.43	-1.29
I ₄	-5.98	0.88
TS5	-8.76	-1.93
I ₅	-8.68	-1.34
TS6	-9.03	-0.97
I ₆	-6.44	-1.59
TS7	-10.12	-1.32

P	-6.79	-2.40
---	-------	-------

Table 4.10: Results of HOMO Centre and LUMO Centre computations for reaction with the candidate catalyst.

Species	HOMO Center (nucleophile site)	LUMO Center (electrophile site)
A	=O	=C
B	=O-----C	=C-----O
C	-N-	=C
D	O	H
I ₁	-N-	=C
I ₂	O [•]	=C
I ₃	O	H

I ₄	=C	H-----O
I ₅	=O-----C	=C-----N
I ₆	O ⁻	H-----O

Table 4.11: HOMO-LUMO energy Gaps for the candidate-catalyzed system.

Steps	HOMO-LUMO Gaps
1	A-C:8.63, C-A: 6.41, A-A: 6.45, C-C: 8.59
2	I ₁ - I ₁ : 4.41, I ₁ -A: 5.97, A- I ₁ : 4.89, I ₁ -C: 8.15, C- I ₁ : 4.85
3	I ₂ - I ₂ : 4.99, I ₂ -A: 5.2, A- I ₂ : 6.24, I ₂ -C: 7.38,C- I ₂ : 6.2
4	I ₃ - I ₃ : 3.76 , I ₃ -A: 5, A- I ₃ : 5.2, I ₃ -C: 7.18, C- I ₃ : 5.17
5	I ₄ - I ₄ : 6.86, I ₄ -A: 5.87, A- I ₄ : 7.44, I ₄ -C: 8.05, C- I ₄ :7.4, I ₄ -B: 3.12, B- I ₄ : 8.06

6	I ₅ - I ₅ :7.34, I ₅ -A: 8.57, A- I ₅ : 5.22, I ₅ -C: 10.75, C- I ₅ : 5.18, I ₅ -B: 5.82, B- I ₅ : 5.84, I ₅ -D: 10.57, D- I ₅ : 6.52
7	I ₆ - I ₆ : 4.85, I ₆ -A:6.33, A- I ₆ : 4.97, I ₆ -B: 3.58, B- I ₆ : 5.59, I ₆ -C 8.51, C- I ₆ : 4.93, I ₆ -D 8.33, D- I ₆ : 6.27

4.7: Mechanisms of the Reactions

The mechanism of the reaction investigated as catalyzed by both proline and the candidate catalyst are presented in subsections 4.7.1 and 4.7.2 as follows.

4.7.1: Mechanism of the reaction for proline-catalyzed system

The mechanism of aldol reaction between acetone and 4-nitrobenzaldehyde catalyzed by proline is given in Scheme 4.1. In the scheme, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, proline is denoted with C, water is denoted with D and the product is represented by P. The transitions states structures 1 to 7 are gives as TS1, TS2,

TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

4.7.2: Mechanism of the reaction for candidate-catalyzed system

The mechanism of aldol reaction between acetone and 4-nitrobenzaldehyde catalyzed by the candidate catalyst is given in Scheme 4.2. In the scheme, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, candidate catalyst is denoted with C, water is denoted with D and the product is represented by P. The transitions states structures 1 to 7 are gives as TS1, TS2, TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

4.8 Computed Thermodynamic Parameters

4.8.1 Computed thermodynamic parameters for proline-catalyzed reaction system

Reaction enthalpy, entropy, Gibbs free energy, and free energy of solvation for the reactants, intermediates and transition states of the proline-catalyzed aldol reaction between acetone and 4-nitrobenzaldehyde were computed and the results are presented in Table 4.12. In the Table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, proline is denoted with C, water is denoted with D and the product is represented by P. The transitions states structures



Scheme 4.1: Reaction mechanism of proline-catalyzed system showing (a) enamine formation step and (b) C-C bond formation step.

Scheme 4.2: Reaction mechanism of candidate-catalyzed system showing (a) enamine formation step and (b) C-C bond formation step.

1 to 7 are given as TS1, TS2, TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

4.8.2 Computed thermodynamic parameters for candidate-catalyzed reaction system

Reaction enthalpy, entropy, Gibbs free energy, and free energy of solvation for the reactants, intermediates and transition states of the candidate-catalyzed aldol reaction between acetone and 4-nitrobenzaldehyde were computed and the results are presented in Table 4.13. In the Table, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, candidate catalyst is denoted with C, water is denoted with D and the product is represented with P. The transitions states structures 1 to 7 are given as TS1, TS2, TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

4.9: Calculated Potential Energy Barriers

From the computed results of enthalpies of formation (H_f°), the energy barriers of the elementary steps were calculated and the results are presented in the subsections 4.9.1 and 4.9.2.

4.9.1: Calculated potential energy barriers for proline-catalyzed system

Figure 4.12 presents the two-dimensional potential energy profile of the aldol reaction between acetone and 4-nitrobenzaldehyde catalyzed by proline. This chart is only useful in displaying the energy barriers of all the reaction steps. It was deduced from the actual enthalpies and solvation energies presented in subsection 4.8.1. The y-axis is the relative energy of the system, and the x-axis is the reaction coordinate which corresponds to the geometry of the system at various points. The sum of enthalpies of formation of acetone and proline catalyst forms the energy minimum at 0 kJmol⁻¹. In the Figure, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, proline is denoted with C, water is denoted with D and the product is represented with P. The transitions states structures 1 to 7 are given as TS1, TS2,

TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

4.9.2: Calculated potential energy barriers for candidate-catalyzed system

Figure 4.13 presents the two-dimensional potential energy profile of the aldol reaction between acetone and 4-nitrobenzaldehyde catalyzed by the candidate catalyst. This chart is only useful in displaying the energy barriers of all the reaction steps. It was deduced from the actual enthalpies and solvation energies presented in subsection 4.8.2. The y-axis is the relative energy of the system, and the x-axis is the reaction coordinate which corresponds to the geometry of the system at various points.

The sum of enthalpies of formation of acetone and the candidate catalyst forms the energy minimum at 0 kJ mol⁻¹. In the Figure, the acetone is designated as A, 4-nitrobenzaldehyde is designated as B, candidate catalyst is denoted with C, water is denoted with D and the product is represented with P. The transitions states structures 1 to 7 are given as TS1, TS2, TS3, TS4, TS5, TS6, and TS7, respectively, whereas, the intermediates 1 to 6 are denoted as I₁, I₂, I₃, I₄, I₅ and I₆, respectively.

4.10. Calculated Thermodynamic and Kinetic Parameters

From the computed results of enthalpies of formation (H_f^o), entropies (S^o) and Gibb's Free energies (G^o), some thermodynamic and kinetic parameters were calculated and the results are presented in subsections 4.10.1 and 4.10.2.

4.10.1 Calculated thermodynamic and kinetic parameters for the reaction of acetone and 4-nitrobenzaldehyde catalyzed by proline.

The results of enthalpy change of activation (ΔH^\ddagger), entropy change of activation (ΔS^\ddagger), activation energy (E_a) and rate constant (k) for the elementary steps are presented in Table 4.14.

4.10.2 Calculated thermodynamic and kinetic parameters for the reaction of acetone and 4-nitrobenzaldehyde catalyzed by the candidate catalyst.

The results of enthalpy change of activation (ΔH^\ddagger), entropy change of activation (ΔS^\ddagger), activation energy (E_a) and rate constant (k) for the elementary steps are presented in Table 4.15.

Table 4.12: Results of thermodynamic computations for proline-catalyzed system

Species	H_f⁰(au)	H_f⁰(kJ/mol)	S⁰(J/mol⁰)	G⁰(au)	Solv. E (kJ/mol)
A	-193.075090	-506918.649	283.75	-193.107313	+25.33
B	-549.967067	-1443938.53	370.49	-550.009140	-45.78
C	-401.020650	-1052879.72	337.27	-401.058951	+40.1
D	-76.3930902	-200570.058	188.95	-76.4145469	-23.59
TS1	-594.023922	-1559609.81	424.80	-594.072162	-65.47
I₁	-594.082377	-1559763.28	419.70	-594.130037	+9.7
TS2	-594.067664	-1559724.65	416.89	-594.115006	-17.3
I₂	-517.676794	-1359160.42	402.47	-517.722498	+28.89
TS3	-517.632646	-1359044.51	402.49	-517.924462	-52.25
I₃	-517.686224	-1359185.18	388.87	-517.730384	+18.26
TS4	-517.654103	-1359100.85	397.64	-517.699258	-40.87
I₄	-517.664868	-1359129.11	399.75	-517.710264	-59.01
TS5	-1067.61355	-2803019.38	534.41	-1067.75048	-138.49
I₅	-1067.65016	-2803115.50	530.54	-1067.71041	-125.56
TS6	-1143.97796	-3003514.13	561.96	-1144.61354	-221.97
I₆	-1144.04502	-3003690.20	550.65	-1144.31160	-132.63
TS7	-1144.04168	-3003681.43	554.16	-1144.30268	-132.62
P	-743.045462	-1950865.86	443.03	-743.095772	-0.0038

Table 4.13: Results of thermodynamic computations for candidate-catalyzed system.

Species	H_f^o (au)	H_f^o (kJ/mol)	S^o(J/mol^o)	G^o (au)	Solv. E (kJ/mol)
A	-193.075090	-506918.649	283.75	-193.107313	+25.33
B	-549.967067	-1443938.53	370.49	-550.009140	-45.78
C	-732.563678	-1923345.94	337.27	-401.058951	+30.50
D	-76.3930902	-200570.058	188.95	-76.4145469	-23.59
TS1	-925.586543	-2430127.47	502.35	-925.685143	-40.10
I₁	-925.633009	-2430249.47	506.78	-925.703061	-51.05
TS2	-925.626063	-2430231.23	509.28	-925.726853	-63.82
I₂	-849.234131	-2229664.21	509.68	-849.301904	-92.57
TS3	-849.194288	-2229559.60	510.39	-849.371058	-166.5
I₃	-849.218756	-2229623.84	475.17	-849.294567	+20.70
TS4	-849.153286	-2229451.15	477.47	-849.210967	-100.26
I₄	-849.212391	-2229607.13	482.93	-849.328911	-15.79
TS5	-1399.15273	-3673475.49	609.29	-1399.23544	-110.68
I₅	-1399.17049	-3673522.12	616.10	-1399.26790	+60.76
TS6	-1475.50606	-3873941.16	626.69	-1475.61037	-68.55

I₆	-1475.58689	-3874153.38	631.84	-1475.67890	-76.55
TS7	1475.57415	-3874119.93	648.01	1475.668912	-99.97
P	-743.045462	-1950865.86	443.03	-743.1446 73	-0.0038

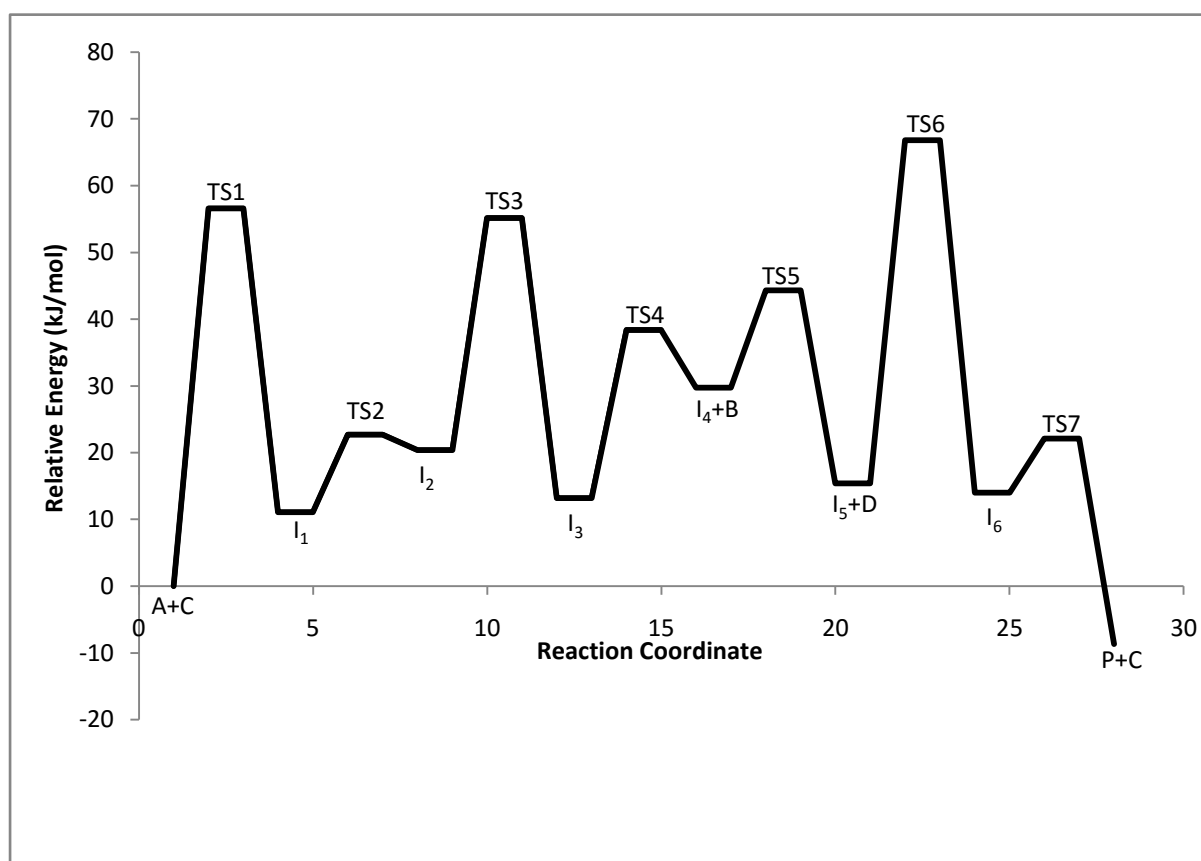


Figure 4.12: Energy Barrier Chart for the mechanistic steps of the proline-catalyzed reaction of acetone and 4-nitrobenzaldehyde.

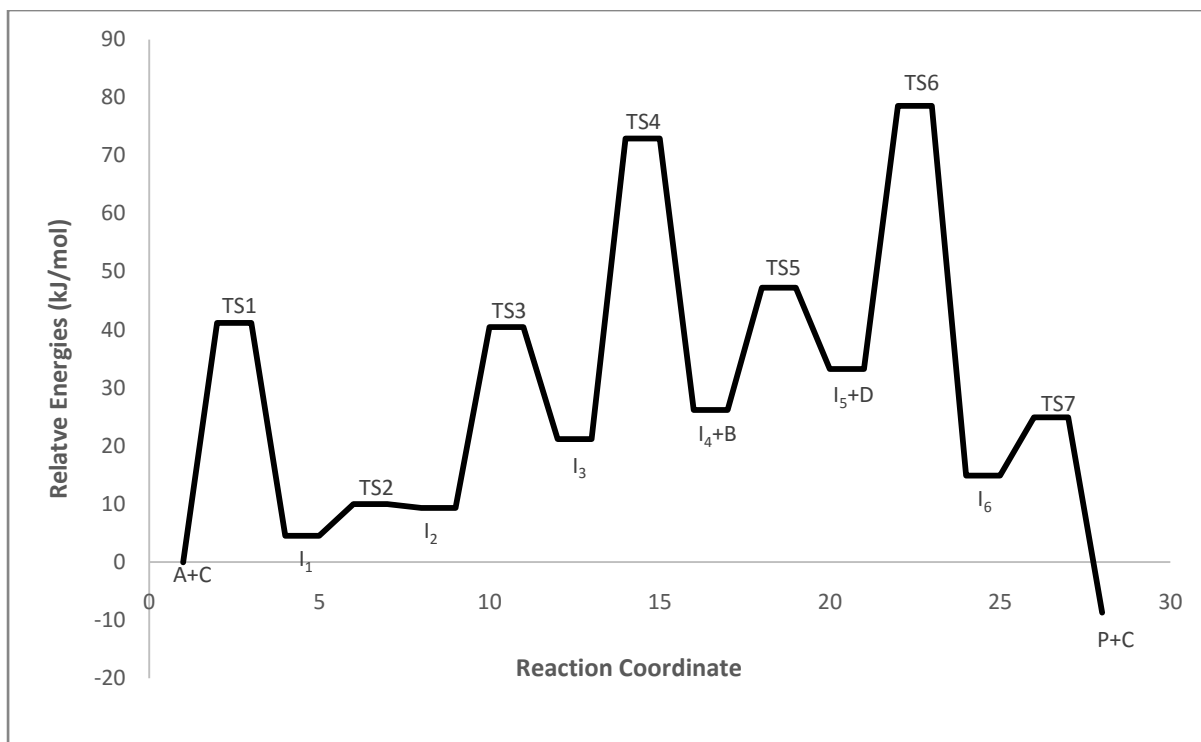


Figure 4.13: Energy Barrier Chart for the mechanistic steps of the candidate-catalyzed reaction of acetone and 4-nitrobenzaldehyde.

Table 4.14: Results of Enthalpy change of activation, Entropy change of activation, Activation Energies and Rate Constants calculations for proline-catalyzed system.

Elementary steps	$\Delta H^\ddagger(kJmol^{-1})$	$\Delta S^\ddagger(Jmol^{-1}K^{-1})$	$E_a(kJmol^{-1})$	k (a or b)
$A+C \xrightarrow{k_1} I_1$	56.6	-196.1827	59.07	$4.04 \times 10^{-8}a$
$I_1 \xrightarrow{k_2} I_2+D$	11.63	-2.81	14.10	4.08×10^{10}
$I_2 \xrightarrow{k_3} I_3$	34.77	0.02	37.24	4.82×10^6b
$I_3 \xrightarrow{k_4} I_4$	25.20	8.77	27.67	6.62×10^8b
$I_4+B \xrightarrow{k_5} I_5$	14.56	-235.83	17.03	$8.22 \times 10^{-3}a$
$I_5+D \xrightarrow{k_6} I_6$	51.43	-157.53	53.90	$3.24 \times 10^{-5}a$
$I_6 \xrightarrow{k_7} P+C$	8.77	3.51	11.24	$2.75 \times 10^{11}b$

$$a = dm^3mol^{-1}s^{-1}, \quad b = s^{-1}$$

Table 4.15: Results of Enthalpy change of activation, Entropy change of activation, Activation Energies and Rate Constants calculations for candidate-catalyzed system.

Elementary steps	$\Delta H^\ddagger(kJmol^{-1})$	$\Delta S^\ddagger(Jmol^{-1}K^{-1})$	$E_a(kJmol^{-1})$	k (a or b)
$A+C \xrightarrow{k_1} I_1$	41.9	-0.3980	44.45	2.473×10^5 a
$I_1 \xrightarrow{k_{-1}} A+C$	36.65	-0.0148	39.12	2.224×10^6 b
$I_1 \xrightarrow{k_2} I_2$	5.47	0.0084	7.94	6.805×10^{11} b
$I_2 \xrightarrow{k_{-2}} I_1$	0.69	-0.0013	3.16	4.688×10^{12} b
$I_2 \xrightarrow{k_3} I_3$	30.68	0.0024	33.15	2.504×10^7 b
$I_3 \xrightarrow{k_{-3}} I_2$	19.27	0.1181	21.74	2.566×10^9 b
$I_3 \xrightarrow{k_4} I_4$	51.73	0.0077	54.20	5.001×10^3 b

$$a = dm^3mol^{-1}s^{-1}, \quad b = s^{-1}$$

CHAPTER FIVE

5.0 DISCUSSION

5.1 Model Development and Validation.

5.1.1 Model development

The models developed involved several types of descriptors such as atom count (nC), bond count (nBondsS), charged partial surface area (PNSA-3, DPSA-1, WPSA-2), Chi path (VP-6, VP-7), autocorrelation (ATSp4), electrotopological state atom type (MaxssNH), extended topochemical atom (ETA_Eta B), wiener numbers (WPATH), and topological polar surface area (TopoPSA), descriptors. Among all the types of descriptors stated, only the Charged Partial Surface Area descriptors belong to 3D class, while the rest are of 2D class. Multiple models (up to five) were developed for systematic selection to get the best performing model. The comparison of the predicted ee values of all the compounds in the training set, calculated from each of the five models, with the actual values given in Table (4.1a and 4.1b) and Figure 4.1 show small residual values in all cases, falling within the range: 0.06-9.18%. This is an indication that the models do very well in predicting the ee values, since models with up to 20% residual values are accepted (Stanton and Jurs, 1990).

5.1.2 Internal Validation

The results of internal validation presented in subsection 4.2.1 show some statistical parameters that assess the quality of the models. The parameters include the following.

Correlation coefficient of multiple determinations (R^2):- This is the fraction of the total variance of the dependant variable (y) that is calculated from the sum of squares of regression using equation 1 of Appendix I (Box *et al.*, 1978).

It indicates the accuracy of the model. The closer the value of R^2 is to 1.0 the more accurate the model explains y variable. The general standard required for an accepted model must have the

coefficient of determination to be greater than or equal to 0.6 ($R^2 \geq 0.6$) (Konovalovet *al.*, 2004). However, R^2 has limitation in the sense that it always increases with increase in number of descriptors used in the model without being sensitive to over fitting. Model 1 shows the highest R^2 value (0.85), which means, it has the highest degree of accuracy.

Adjusted R^2 :- The R^2 is adjusted by reducing the proportion of variance to the size of the estimated model. It is calculated from equation 2 of Appendix I(Box *et al.*, 1978).

The adjusted R^2 is meaningful only when there are some degrees of freedom for lack-of-fit. For a model with an additional parameter to be an improvement, the addition of the parameter is required to decrease SSE at least proportionately to the reduction in the degrees of freedom. Compared to R^2 , this measure penalizes large equations.

Cross validated R-squared (R_{cv}^2):- This is the R^2 value derived from cross validation. It largely depends on the predictive sum of squares (PRESS) of a model. It is calculated from equation 3 of appendix I(Box *et al.*, 1978).

The R_{cv}^2 indicates the predictive power of the model. The predictive ability of the model increases as the R_{cv}^2 value approaches 1.0. For a model to be acceptable, the cross validation coefficient should be greater than 0.5 ($R_{cv}^2 > 0.5$). Model 1 has the highest predictive ability for having the highest value of R_{cv}^2 (0.75). Hitherto, for a good model, R_{cv}^2 should be very close to R^2 , so much so that the difference between R_{cv}^2 and R^2 does not exceed 0.3 (Konovalovet *al.*, 2004; Tropsha *et al.*, 2003). A very wide difference suggests that the model is over fitting the data. The differences of 0.09, 0.11, 0.11, 0.12 and 0.13 were found for the models 1 to 5, respectively. This demonstrates absence of over fitting in all the five models.

F Test:- The F test is a standard statistical test for the equality of the variances of two populations having normal distributions. Here, it is used to test whether the variance in the data

which is explained by the regression is much larger than the variance remaining due to errors. If this is the case, then the model is said to be significant, rather than one which simply fits the noise. To achieve meaningful interpretation, the result of F test is compared with that of the Critical SOR F-value. The F test is conducted by determining the significance-of-regression (SOR) F-value. This parameter is calculated using equation 4 of Appendix I (Box *et al.*, 1978).

Critical SOR F-value (95%) (F_{cr}): is the critical point of the F distribution evaluated for probability of 0.05 (at 95% confidence level). In conjunction with the calculated significance-of-regression (F value), it is used to determine whether or not the regression is significant. F to F_{cr} ratio provides an indication of statistical significance of the model. If the ratio is greater than 1, the model is significant, while if it is less than 1 the model is not significant. The ratios for all the model equations are greater than 1, having the values; 8.35, 8.24, 8.19, 8.16 and 8.12 for models 1 to 5, respectively. This shows that all the 5 models are significant.

5.1.3: External validation.

The purpose of this validation exercise is to further test the predictive power of the models. It is carried out upon the compounds in the test set to compute, mainly, correlation coefficient of external validation (Q_{ext}^2) calculated from equation 5 of Appendix I and Pearson correlation coefficient of prediction (R_{ext}) calculated from equation 6 of the appendix (Benigni and Bossa, 2008).

The general standard for an accepted model requires that the model must have $R_{ext} \geq 0.6$ and $Q_{ext}^2 > 0.5$ (Konovalov *et al.*, 2004; Berkessel and Koch, 2004). More so, difference between R_{ext} and Q_{ext}^2 should fall within the range: 0.2 to 0.3 for the model to acquire high quality (Eriksson *et al.*, 2003). All the five models are acceptable. For having the highest value of R_{ext} (0.88), which is the measure of predictive potential, model 1 is selected as the best

model. Moreover, the differences between R_{ext} and Q_{ext}^2 are 0.20, 0.10, 0.14, 0.13 and 0.14 for models 1 to 5, respectively. This shows that only model 1 falls within the standard range.

5.2 Catalyst Design

Since model 1 is selected as the best among the five models, we concentrate on using it to search for the new catalyst. Due to the qualities it possesses as discussed in Chapter 1, proline was selected as the central structure and model 1 was used to adjust the structure for higher ee value. The descriptors that appear in model 1 are highlighted below:

nC: number of carbons

This is the carbon count descriptor which stands for the total number of carbons in the compound. Having a positive coefficient in the model signifies that increasing the value of the descriptor favours the enantioselectivity. Increasing the value of the descriptor can be achieved by attaching any carbon rich substituent to the proline structure.

PNSA-3: path weighted partial negative surface area

This is a Charged Partial Surface Area (CPSA) descriptor. It is calculated using equation 1 of Appendix II (Stanton and Jurs, 1990).

Since it has negative coefficient, decreasing the value of the descriptor favours the enantioselectivity. Thus, we need to decrease the number of negative atoms in the molecule at the surface of contact with the solvent or replace the atoms with smaller ones of comparable charge. The large factor of the PNSA-3 (10^1) suggests that the enantioselectivity is most influenced by this descriptor.

DPSA-1: difference in charged partial surface area

This is another CPSA descriptor which is computed from the difference of the partial positive surface area-1 and partial negative surface area-1 as described by equations 2 to 4 of appendix II (Stanton and Jurs, 1990).

Increasing the value of the descriptor favours the enantioselectivity.

VP-6: valence path, order 6

This belongs to the chi connectivity indices. For small atoms like hydrogen, the descriptor is calculated using equation 5, while for atoms of higher principal quantum number, it is calculated from equation 6 of appendix II (Kier and Hall, 1986 and 1991).

Since the coefficient of the descriptor in the model equation is negative, decreasing the value of the descriptor favours the enantioselectivity.

MaxssNH: maximum value of electrotopological state –NH-

This is an electrotopological state atom type descriptor which is calculated from equation 7 of appendix II (Kier and Hall, 1986 and 1991).

Having a positive coefficient in the model equation, increasing the descriptor value leads to increase in enantioselectivity value. MaxssNH is definitely one of the most important descriptors to touch in order to bring a significant change to the value of the Y variable. This is because the descriptor is the one having the highest coefficient value (70.43) in the model equation.

5.2.1 New Catalyst Candidates

Two catalyst candidates are proposed.

Candidate 1: Adjustments were made upon proline to produce the catalyst structure in Figure 4.2. A hydrofuran ring was added to increase the carbon count without causing steric hindrance as might be with a smaller or much larger ring, due to ring strain. More so, the ring has been deliberately situated to remove the double bond on the carbonyl oxygen, and additional oxygen atom was added in order to increase the number of negatively charged atoms in the compound, thereby, increasing the negative value of PNSA-3. This consequently increases the value of DPSA-1. MaxSSNH also increases due to increase in number of atoms which increases the perturbation factor.

Candidate 2: Like in the case of candidate 1, proline was taken as the starting material to arrive at the candidate structure in Figure 4.3. A hydrofuran ring, containing carbons and oxygen was added to increase the carbon count, without causing steric hindrance. The oxygen serves to increase the number of negatively charged atoms in the compound, thereby, increasing the negative value of PNSA-3 and the value of DPSA-1. MaxSSNH also increases due to perturbation factor.

Candidate 1 is found to have higher ee value (85%) than the candidate 2 (80%). Candidate 1 is therefore considered for the simulation reaction, operating the candidate catalyst over asymmetric aldol reaction between acetone and 4-nitrobenzaldehyde in an acetone medium.

5.3 Graphical Models (Surfaces)

The LUMO and HOMO diagrams of the entire species involved in the mechanistic steps of asymmetric aldol reaction between acetone and 4-nitrobenzaldehyde as catalyzed both by proline and the candidate catalyst, are given in Figures 4.6a, 4.6b and 4.9. By convention, the blue and red colours show the positive and negative values of the orbital, respectively. Moreover, the shapes of the frontier orbitals can be used as a guide in determining reactivity. The frontier orbitals have to be oriented to undergo a symmetry-allowed interaction in such a way that the positive region of the HOMO interacts with the positive region of the LUMO, also the negative region of the HOMO interact with negative region of the LUMO, thereby resulting in positive overlap throughout. In this way, the two interactions reinforce, and the total frontier orbital interaction is non-zero. Therefore, according to Fukui-Woodward-Hoffmann rules (Bhoorasingh and West, 2015; Fukui, 1965; Hoffmann and Woodward, 1965), electron movement giving rise to chemical reaction can occur. On every species, the LUMO is delocalized onto several atoms and it is difficult to tell where exactly a pair of electrons (a nucleophile) will attack the molecule. A clearer picture is provided by the LUMO-map, which paints the absolute value of LUMO on the electron density surface. By convention, the blue

region represents maximum value of the LUMO and the red region represents the minimum value. Likewise, the HOMO delocalizes over several sites, but the largest contribution comes from one atom. This finding is supported by the HOMO-density which maps the absolute value of HOMO on to the electron density surface. By convention, the blue region represents maximum concentration of the HOMO and the red region represents the minimum concentration.

Electrostatic potential map (EPM) describes the charge distribution, which gives good insight into the physical and chemical properties of the molecule. This is because chemical reactions are associated with charged sites, in the sense that, the most highly-charged site in a molecule is often the most reactive. The sign of the charge describe the function which the molecule performs in the reaction. Positively-charged sites invite attack from bases and nucleophiles, while the negatively-charged sites invite attack from acids and electrophiles. Hence, the former functions as electrophile while the latter functions as nucleophile. Basically, electrostatic potential is the energy of interaction of a point positive charge with the nuclei and electrons of a molecule. If the point charge is located in a region of excess positive charge or electron poor region, the point charge-molecule interaction becomes repulsive and the electrostatic potential will be positive. However, if the point charge is located in a region of excess negative charge or electron rich region, the interaction is attractive and the electrostatic potential becomes negative. The simplest way to demonstrate molecular charge distribution is to construct an electrostatic potential map which is achieved by mapping electrostatic potential onto electron density surface corresponding to a space-filling model. On such map, red colour indicates the region of most negative electrostatic potential (nucleophiles), blue colour depicts the region of most positive electrostatic potential (electrophiles) and the colours between indicate intermediate values of the EPM. Therefore, the potential increases in the order: red<orange<yellow<green<blue.

5.4. Density maps and Electrostatic Potential Maps

5.4.1 Density maps and electrostatic potential maps for proline-catalyzed aldol reaction.

Section 4.5.1 gives the results of the graphical surfaces computations for the proline-catalyzed system. Centres of HOMO and LUMO concentration were deduced from the HOMO and LUMO density maps presented in Figures 4.7a and 4.7b, corroborated by the electrostatic potential maps presented in Figure 4.8. In Figure 4.6a, the LUMO of acetone is seen to be resided more on the carbonyl carbon (C=) and the HOMO is most resided on the carbonyl oxygen (O=). More so, from Figure 4.7a, the LUMO map of acetone shows the blue colour on the carbonyl carbon (C=) and the density of HOMO map shows green colour on the carbonyl oxygen (O=). Moreover, from the Figure 4.8, electrostatic potential map of acetone shows the red colour on the carbonyl oxygen, blue colour on terminal hydrogens and green colour on the carbonyl carbon. Conclusively, C= is the LUMO centre and a potential electrophile site, while O= is the HOMO center and a potential nucleophile site.

Similarly, Figure 4.6a shows the LUMO of proline residing mostly on the carbonyl carbon (C=) while the HOMO resides mostly on the nitrogen atom of the pyrrolidine group (-N-). The LUMO map as seen in Figure 4.7a shows the green colour on the carbonyl carbon (C=) of the proline molecule while density of HOMO displays the green colour on the the nitrogen atom (-N-). Furthermore, the electrostatic potential map of proline as seen in Figure 4.8 shows a deep blue colour on the carbonyl carbon and the redcolour on the nitrogen.

Complete analyses of frontier orbitals (Figures 4.6a and 4.6b), LUMO and HOMO map (Figures 4.7a and 4.7b) and the electrostatic potential maps (Figure 4.8) of the entire reacting species and the intermediates, yields the conclusions presented in Table 4.7.

5.4.2 Density maps and electrostatic potential maps for candidate-catalyzed aldol reaction.

Section 4.5.2 gives the results of the graphical surfaces computations for the candidate-catalyzed system. Centres of HOMO and LUMO concentration were deduced from the HOMO and LUMO density maps presented in Figures 4.10, corroborated by the electrostatic potential maps presented in Figure 4.11. In Figure 4.9, the LUMO is seen to be resided more on the carbonyl carbon (C=) and the HOMO is most resided on the carbonyl oxygen (O=). More so, from Figure 4.10, the LUMO map of acetone shows the blue colour on the carbonyl carbon (C=) and the density of HOMO map shows green colour on the carbonyl oxygen (O=). Moreover, from the Figure 4.11, electrostatic potential map of acetone shows the red colour on the carbonyl oxygen, blue colour on terminal hydrogens and green colour on the carbonyl carbon. Conclusively, C= is the LUMO centre and a potential electrophile site, while O= is the HOMO centre and a potential nucleophile site.

Similarly, Figure 4.9 shows the LUMO of candidate catalyst residing mostly on the carbonyl carbon (C=) while the HOMO resides mostly on the nitrogen atom of the pyrrolidine group (-N-). The LUMO map as seen in Figure 4.10 shows the green colour on the carbonyl carbon (C=) of the catalyst molecule while density of HOMO displays the green colour on the nitrogen atom (-N-). However, the electrostatic potential map of the candidate as seen in Figure 4.11 shows a faint blue colour on the central carbon atom of the pyrrolidine moiety (-C-), and the red colour is shown on the nitrogen atom.

Complete analyses of frontier orbitals (Figure 4.9), LUMO and HOMO map (Figure 4.10) and the electrostatic potential maps (Figure 4.11) of the entire reacting species and the intermediates, yields the conclusions presented in Table 4.10.

5.5 Mechanism of the Reaction.

The basis of the mechanisms of the reactions presented in subsections 4.7.1 and 4.7.2 is the idea of HOMO-LUMO energy gap. According to Molecular Orbital (MO) theory, HOMO and LUMO are regarded as frontier orbitals and considered to be the most involved MOs in

chemical reactions. It is imperative to note that most chemical reactions involve electron transfer between orbitals with HOMO being the donor orbital and LUMO being the acceptor orbital. For a successful interaction, the energy input required for electron movement is kept at the minimum. Where more than one combination of reagents can react, there exists a competition as to which combination would react first. Conventionally, we can suggest the most favourable combination by examining the energies of the frontier orbitals. It is reasonable to assume that the donor reagent with the highest energy HOMO will give up its electrons most easily and be the most reactive nucleophile. While the electron-acceptor reagent with the lowest energy LUMO will accept electrons most readily, and be the most reactive electrophile. The extent to which MOs combine depends on the degree of overlap (S) and the energy difference between MOs ($\Delta\epsilon$) in a relation, $S^2/\Delta\epsilon$. Hence, for a mixture of several combinations of competing nucleophiles and electrophiles, the fastest chemical reaction would be the one that involves the reagent combination of smallest HOMO-LUMO gap ($\Delta\epsilon$). This means, the HOMO and LUMO with narrower energy gap combine more readily than the ones with wider gap

5.5.1 Mechanism of the reaction for proline-catalyzed system.

The HOMO-LUMO energy gaps of all the possible combinations of reaction species involved in the mechanistic steps of the proline-catalyzed system were analyzed. The energy gaps of the various combinations are summarized in Table 4.8

From the results of E- HOMO and E-LUMO of all the reaction species given in Table 4.6, acetone (A) is seen to have E-HOMO value of -6.56 eV and E-LUMO value of -0.11eV. Proline (C) has the E-HOMO and E-LUMO values of -6.36 eV and 0.60 eV, respectively. The

results allowed the comparison of all possible HOMO-LUMO energy gaps between the reacting species in the system and the smallest gap is found to be between the HOMO of C and the LUMO of A (6.25 eV) as seen in Table 4.8. More so, from Table 4.7, the HOMO centre of C is at the nitrogen atom while the LUMO centre of A is at the carbonyl carbon atom.

The nitrogen atom of proline therefore attacks the carbonyl carbon of acetone, forming a C.....N bond of 1.628 Å, while O of acetone mechanically attacks H atom attached to N of proline, forming an O.....HN bond of 1.352 Å. Following such attacks, TS1 structure (Scheme 4.1a) is formed. TS1 was identified by an imaginary frequency of 1633 Hz. As C-N bond completely forms at 1.456 Å and N-H bond breaks, an intermediate I₁ is formed. Considering all the HOMO-LUMO gaps involved among the species in the reaction setting, as seen in Table 4.8, the shortest gap for possible interactions with the I₁ is found to be 5.53 eV occurring between the HOMO of I₁ (centred on N) and its LUMO (centred on C=). However, from the result of electrostatic potential map computation, the electrophile site is on -C- atom. This gives rise to competition between the C= and -C- atoms on which would accept the electrons from the N atom. Due to steric hindrance (Hofmann and Woodward, 1965; Fukui, 1965), the N atom could not attack C=, rather, -C-. This leads to the formation of TS2 as HO.....HO bond of 1.643 Å forms and C...OH bond stretches to 1.495 Å in preparation for the removal of a water molecule. TS2 was confirmed to be true transition state by an imaginary frequency from IR plot at 1586 Hz. When TS2 finally transforms to iminium ion I₂, water molecule (D) is completely removed but remains fused with I₂.

The next most favourable attack is suggested from the smallest HOMO-LUMO energy gap (4.37 eV) for possible interactions with I₂. This occurs between the HOMO of I₂ concentrating on O⁻, and its' LUMO centring on =C. The I₂ rearranges via formation of C-O bond, 3.85 Å long, to form the transition complex, TS3, fused with D. The IR plot of TS3 gave an imaginary frequency of 1999 Hz. As C=N π bond breaks and C-O bond contracts to 1.501 Å, TS3

stabilizes to I₃. The I₃, having the most favourable energy gap (5.83 eV) within its self, would rather rearrange than interact with any other chemical specie in the setting. The O atom, on which the HOMO is centred, therefore attacks either of the hydrogen atoms of H-CH₂ moieties, forming an O...H bond of 1.24 Å as seen in Scheme 4.1a. Mechanically driven breakage of C-H and C-O bonds begins as they stretch to 1.386 Å and 2.982 Å respectively. On the other hand, C-C π-bond occurs to complete the transition structure TS4. Analysis of IR plot confirmed TS4 to be true transition state giving an imaginary frequency of 1421 Hz. Upon completion of this transition step, enamine (I₄) is formed with newly established C=C bond.

At this stage, 4-nitrobenzaldehyde (B) is introduced into the system and involved in the analyses of HOMO-LUMO energy gaps. Among all the species present in the system, viz; A, C, I₄ and B, the most favourable combination is the HOMO of I₄ and LUMO of B, having the smallest gap of 2.21 eV. Enamine interacts with nitrobenzaldehyde via a C.....C bond of 1.585 Å that occurs between C= of enamine and =C...O of nitrobenzaldehyde. According to Agami *et al.* (1984), this particular step involves a second proline molecule, but a careful scrutiny of HOMO-LUMO gaps for the entire species in the reaction setting shows that the interaction with the second proline is energetically unfavourable. This is because the HOMO of I₄ and the LUMO of proline molecule has the energy gap of 5.67 eV, while the HOMO of proline and the LUMO of I₄ has the energy gap of 6.53 eV. In both cases the energy gap is the greater than the one in the direct combination of I₄ with B (2.21 eV). Hence, according to this analysis, there is no need for the participation of the second proline as a catalyst. Therefore, the interaction kicks up as aforementioned and proceeds by proton transfer from OH of proline to the carbonyl oxygen of B, via formation of O.....H bond of 0.975 Å and breakage of H....O bond by elongating to 2.44 Å. The net structure is the transition complex TS5 (Scheme 4.1b). An imaginary frequency of 474 Hz shows TS5 to be the true transition state. Upon complete proton transfer and contraction of C-C bond to 1.580 Å, a stable intermediate, I₅, forms. The

next interaction is between the HOMO of D, centred on O atom, and the LUMO of I₅, centred on C=...N, with energy gap of 5.45 eV. This interaction is not the most favourable in the system as regard the HOMO-LUMO energy gap. This is because the I₅-I₅, A-I₅, C-I₅, I₅-B and B-I₅ interactions have the energy gaps of 3.48 eV, 4.15eV, 3.95 eV, 3.03 eV and 5.37 eV, respectively, all of which are lower than that of the D- I₅ interaction (Table 4.8). However, the internal rearrangement (I₅-I₅ interaction) is discouraged by steric hindrance while the abundance of water allows the reaction with water to occur most readily than with either A, C or B. An O...C bond of 1.537 Å, therefore, forms between water molecule and intermediate I₅, and N⁺ entraps one H atom of D and forms a partial N...H bond of 1.450 Å. The resultant transition structure TS6, with imaginary frequency of 1983 Hz, forms before transforming to I₆. The next proposed step is the internal rearrangement of I₆ which has energy gap of 4.28eV, higher than I₆-B interaction with energy gap of 3.03 eV. The HOMO centre of I₆ is at O⁻ while the LUMO centre is at H atom of OH group at α position to N⁺. An O...H bond with bond length of 1.096 Å forms. This causes H to break from its previous bond with O and allows the O to condense to carbonyl group. As the C-O double bond forms, the C-N⁺ bond stretches to 1.679 Å. The entire process leads to the formation of transition state structure TS7, identified by an imaginary frequency of 1275 Hz, which splits to give back proline catalyst C and the aldol product, P.

5.5.2: Mechanism of the reaction for candidate-catalyzed system.

The HOMO-LUMO energy gaps of all the possible combinations of reaction species involved in the mechanistic steps of the candidate-catalyzed system were analyzed. The energy gaps of the various combinations are summarized in Table 4.11.

From the results of E- HOMO and E-LUMO of all the reaction species given in Table 4.9, acetone (A) is seen to have E-HOMO value of -6.56 eV and E-LUMO value of -0.11eV. Candidate catalyst (C) has the E-HOMO and E-LUMO values of -6.52 eV and 2.07 eV,

respectively. The results allowed the comparison of all possible HOMO-LUMO energy gaps between the reacting species in the system and the smallest gap is found to be between the HOMO of C and the LUMO of A (6.41eV) as seen in Table 4.11. More so, from Table 4.10, the HOMO centre of C is at the nitrogen atom while the LUMO centre of A is at the carbonyl carbon atom.

The reaction therefore begins by conflation of the acetone and the catalyst in which the N atom of the catalyst attacks the carbonyl carbon of A, forming a C.....N bond of 1.628 Å, while O of acetone mechanically attacks H atom attached to N of the catalyst, forming an O.....HN bond of 1.352 Å.

Following the attack of the catalyst upon acetone, TS1 structure (Scheme 4.2a) is formed. TS1 was identified as a true transition state by the presence of an imaginary frequency of 1951 Hz in its IR plot. As C-N bond completely forms and N-H bond on the catalyst moiety breaks, an intermediate, I₁, is formed as seen in Scheme 4.2a. After careful scrutiny of all the HOMO-LUMO gaps involved among the species in the reaction setting, viz, A, C and I₁, the shortest gap is found to be 4.41eV within the I₁, between its HOMO (centred on N) and its LUMO (distributed fairly among C= and C). This leads to the formation of TS2 (Scheme 4.2a) as HO...HO bond of 1.806 Å forms and C...OH bond stretches to 1.428 Å in preparation of removing a water molecule. TS2 was confirmed to be a true transition state by an imaginary frequency from IR plot at 485 Hz. When TS2 finally transforms to iminium ion (I₂), water molecule (D) is completely removed.

The next segment of the reaction involves iminium-enamine tautomerism. The shortest energy gap among all the possible combinations is 4.99 eV within the I₂. Therefore the I₂, having a HOMO concentrating on O⁻, and LUMO centring on =C, would rather rearrange than interact with any other chemical specie in the setting, forming C-O bond 5.827 Å. This bond formation

gives rise to a transition complex, TS3. The IR plot of TS3 gave an imaginary frequency of 1996 Hz. As C=N π bond breaks and C-O bond contracts, TS3 stabilizes to I_3 as seen in Scheme 4.2a. Bearing the E-HUMO of -5.11 eV and E-LUMO of -1.35 eV, I_3 has the most favourable combination within itself (I_3 - I_3 : 3.76 eV). I_3 , therefore, undergoes rearrangement. The O atom attacks either of the hydrogen atoms of H-CH₂ moieties, forming an O...H bond of 1.24 Å. Mechanically driven breakage of C-H and C-O bonds begins as they stretch to 1.287 Å and 2.762 Å respectively. On the other hand, C-C π bond occurs to complete the transition structure TS4 (Scheme 4.2a). Analysis of IR plot confirmed TS4 to be a true transition state, giving an imaginary frequency of 830 Hz. Upon completion of this transition step, enamine (I_4) is formed with C=C bond of 1.354 Å.

The third segment of the reaction in which the new C-C bond is established, is demonstrated in Scheme 4.2b. At this stage, 4-nitrobenzaldehyde (B) is introduced into the system and is involved in the analyses of HOMO-LUMO energy gaps. Among all the species present in the system, viz; A, C, I_4 and B, the most favourable combination is the HOMO of I_4 and LUMO of B, having the smallest gap of 3.12 eV. I_4 interacts with nitrobenzaldehyde (B) via a C.....C bond of 1.581 Å that occurs between C= of I_4 and carbonyl C of nitrobenzaldehyde. The interaction proceeds by proton transfer from OH of enamine to the carbonyl oxygen of nitrobenzaldehyde, via formation of an O.....H bond of 1.101 Å and breakage of another H....O bond by first elongating to 1.178 Å. The net result is the transition structure TS5 (Scheme 4.2b). An imaginary frequency of 863 Hz in the IR plot proves TS5 to be true transition state. Upon complete proton transfer and contraction of C-C bond, a stable intermediate, I_5 , forms. Considering all possible HOMO-LUMO gaps among the species present, the next most favourable interaction is between I_5 and water (D) with the HOMO-LUMO energy gap of 6.52 eV. The HOMO of D centring on O atom, attacks the LUMO of I_5 concentrating on C=....N. An O....C bond of 1.483 Å, therefore, forms between water molecule and the intermediate I_5 ;

and N^+ on the I_5 entraps one H of water and forms a partial $N\dots H$ bond of 1.446 Å. The resultant transition structure TS6 (Scheme 4.2b) forms with an IR imaginary frequency of 1985 Hz before eventually transforming to a complex, I_6 .

The final segment of the reaction is breaking of the complex I_6 to give the product with the recovery of the catalyst. HOMO centre of I_6 is at O^- with energy value of -6.44 eV while the LUMO centre is at H of OH group at α position to N^+ with energy value of -1.59 eV. An $O\dots H$ bond is formed from the HOMO-LUMO interaction with bond length of 0.964 Å. This causes H to break from its previous bond with O and allows the O to condense to carbonyl group. As the C-O double bond forms, the C- N^+ bond stretches to 1.928 Å. The entire process leads to the formation of transition state structure TS7 (Scheme 4.2b), identified by an imaginary frequency of 1279 Hz in the IR plot. TS7 finally splits to give back the catalyst C and the aldol product, P.

5.6 Thermodynamic Parameters.

The results of thermodynamic parameters presented in Tables 4.12 and 4.13 are the absolute values of the parameters. The results have no strong physical meaning since they depend on the chosen functional and basis set. However, they help to determine the energy barriers of the various transformations involved in the reactions, which is a very important parameter with strong physical application on both the kinetics and the thermodynamics of the reaction.

5.7 Potential Energy Barriers.

5.7.1 Potential energy barriers for proline-catalyzed system.

Figure 4.12 presents the two-dimensional potential energy profile of the reaction. This Figure is only useful in displaying the energy barriers of all the reaction steps. The y-axis is the relative energy of the system, and the x-axis is the reaction coordinate which corresponds to the geometry of the system at various points. The sum of enthalpies of formation of acetone (A) and proline catalyst (C) forms the energy minimum at 0 kJ mol

The first step which involves conflation of A and C forms TS1 at 56.6 kJ mol^{-1} above the ground level. The TS1 stabilizes to an intermediate, I_1 at $11.09 \text{ kJ mol}^{-1}$. After undergoing rearrangement, I_1 becomes TS2 at $22.70 \text{ kJ mol}^{-1}$. When TS2 finally transforms to iminium ion, I_2 ; lying slightly below TS2 at $20.39 \text{ kJ mol}^{-1}$, a water molecule (D) is completely removed but remains fused with I_2 . I_2 undergoes rearrangement to form a transition complex (TS3), fused with D, at $55.16 \text{ kJ mol}^{-1}$ above reactants' level. TS3 stabilizes to I_3 which, conflatd with D, occupies an energy position of 13.2 kJ mol^{-1} above ground level. Internal rearrangement of I_3 leads to the formation of the transition structure, TS4, with the relative energy of 38.4 kJ mol^{-1} , in coalescence with D. Upon completion of this transition step, I_4 is formed at $29.76 \text{ kJ mol}^{-1}$, conflatd with D. Some previous works assumed a single transition state in iminium-enamine conversion (Haindl *et al.*, 2015; Perrin and Chang, 2016). However, the supposed transition state, that would have been TS3, was found to have overwhelmingly higher energy barrier (91 kJ mol^{-1}) than the two-step pathway with TS3 ($34.77 \text{ kJ mol}^{-1}$) and TS4 (25.2 kJ mol^{-1}) as demonstrated in appendix III.

I_4 interacts with B to form the transition complex TS5 of relative energy $44.32 \text{ kJ mol}^{-1}$. Upon stabilization of TS5, a stable intermediate, I_5 , was formed at $15.41 \text{ kJ mol}^{-1}$ above the ground level. The I_5 rearranges to reach a transition state. The resultant transition structure, TS6, lies at the energy level of $66.84 \text{ kJ mol}^{-1}$, before transforming to I_6 at the energy level of $14.01 \text{ kJ mol}^{-1}$ above the ground state. The next elementary step is the formation of transition state structure TS7 at $22.12 \text{ kJ mol}^{-1}$, which splits to give back proline catalyst C and the aldol product, P at 8.68 kJ mol^{-1} lower than the ground level.

Analysing all the elementary steps, it is seen that steps 1, 2 and 4 are endergonic changes, while steps 3, 5, 6 and 7 are all exergonic changes (Figure 4.12). However, the entire process is an exothermic process with energy gain of $-8.68 \text{ kJ mol}^{-1}$. The most kinetically favourable step is step 7, with the least energy barrier (8.77 kJ mol^{-1}). Moreover, the most thermodynamically

favourable step is the same step (step 7), with the highest enthalpy change of formation ($-22.69 \text{ kJmol}^{-1}$). Even though, step 6 occupies the highest energy position at the activation complex (66.84 kJmol^{-1}), step 1 has the highest energy barrier (56.6 kJmol^{-1}), greater than that of step 6 (51.43 kJmol^{-1}). Therefore, step 1 is regarded as the rate-limiting step.

5.7.2 Potential energy barriers for the candidate-catalyzed system.

Figure 4.13 shows the two-dimensional potential energy diagram of the reaction progression for the elementary steps involved in the entire reaction. This is only useful in showing the energy barriers of all the reaction steps. The y axis corresponds to the relative energy of the system, and the x axis is the reaction coordinate which corresponds to the geometry of the system at various points during the reaction. The sum of enthalpies of formation of the acetone (A) and the candidate catalyst (C) is taken as the energy minimum at 0 kJmol^{-1} .

Following the attack of the catalyst on acetone, TS1 structure is formed with a barrier of $41.19 \text{ kJ mol}^{-1}$. Intermediate I_1 , results from TS1 at 4.54 kJ mol^{-1} above reactants' level. The next elementary step goes through formation of TS2 with a barrier of 5.47 kJmol^{-1} . When TS2 finally transforms to I_2 , lying slightly below TS2 at 9.32 kJ mol^{-1} , water molecule (D) is completely removed.

The I_2 rearrangement reaction gives rise to a transition complex (TS3) at 30.68 kJmol^{-1} above I_2 . TS3 then stabilizes to I_3 , which occupies energy position of 21.21 kJmol^{-1} above reactants' level. I_3 undergoes rearrangement to form the transition structure TS4, with a high energy barrier of 51.73 kJmol^{-1} . Upon completion of this transition step, I_4 is formed at 26.22 kJmol^{-1} , above ground level.

I_4 interacts with B to form the transition structure TS5 with a barrier of 21.06 kJmol^{-1} . Upon complete transformation, a stable intermediate, I_5 , forms at 7.01 kJmol^{-1} above I_4+B .

The next reaction occurs between water molecule and the intermediate I_5 ; the resultant transition structure, TS_6 , forms at energy level of 45.3 kJmol^{-1} above I_5 before eventually transforming to a complex, I_6 , lying lower at 14.93 kJmol^{-1} .

The final segment of the reaction is breaking of the complex I_6 to give the product with the recovery of the catalyst. The process leads to the formation of transition state structure TS_7 with a barrier of 10.03 kJmol^{-1} , which splits to give back the catalyst C and the aldol product, P , at 8.68 kJmol^{-1} lower than the ground level.

Analyzing all the elementary steps, it is seen that steps 1 to 5 are all endergonic changes, while steps 6 and 7 are exergonic changes (Figure 4.13). However, the entire process is an exothermic process with gain energy of -8.68 kJmol^{-1} . This suggests that the reaction is thermodynamically favourable. The most kinetically favourable step is step 2, with the least energy barrier (5.47 kJmol^{-1}) while the most thermodynamically favourable step is step 7 with the highest enthalpy change of formation ($-23.61 \text{ kJmol}^{-1}$). Even though, step 6 occupies the highest energy position at the activation complex (78.59 kJmol^{-1}), step 4 has the highest energy barrier (51.73 kJmol^{-1}), greater than that of step 6 (45.30 kJmol^{-1}). Therefore, step 4 is regarded as the rate-limiting step.

5.8 Determination of Kinetic Parameters (activation energy and rate constant).

5.8.1 Kinetic parameters of the proline-catalyzed reaction

Scheme 5.1 presents the elementary steps of the reaction mechanism. The activation energies were calculated using equation 5.1, which is applicable to both unimolecular and bimolecular solution-phase reactions (Lewers, 2016; Puri *et al.*, 2014).

Scheme 5.1: The elementary steps in the mechanism of proline-catalyzed aldol reaction of acetone and 4-nitrobenzaldehyde

$$E_a = \Delta H^\ddagger + RT \quad (5.1)$$

ΔH^\ddagger is the enthalpy of activation for a given elementary step, while T is the Kelvin temperature and R is the molar gas constant. The equation was applied using the temperature of the reactions (298.15 K) and molar gas constant of $8.314JK^{-1}mol^{-1}$.

For step 1; $\Delta H^\ddagger = H_f(TS1) - [H_f(A) + H_f(C)]$

$$\Delta H^\ddagger = 56.6 - 0 = 56.6 \text{ kJmol}^{-1}$$

And

$$E_a = 56.6 + 2.47 = 59.07 \text{ kJmol}^{-1}$$

From Table 4.14, the first step which involves the addition of proline to acetone has the highest activation energy (59.07 kJmol^{-1}), followed by the step involving the addition of water to C-C complex (53.90 kJmol^{-1}). We observed a reasonable trend is in the relationship between activation energies and the HOMO-LUMO gaps of the combining MOs, in the sense that the

two steps with the highest activation energies happen to have wider HOMO-LUMO gaps (6.25 eV and 5.45eV, respectively). This trend can be related to the fact that the HOMO with lower energy is comparably more stable and disinclined to release electron for bonding than the HOMO with higher energy. The separation of the HOMO from LUMO corresponds to the amount of energy needed to excite an electron. Therefore, higher activation energy is needed to achieve a reaction in the case of HOMOs at lower energy level further away from LUMOs than in the case of HOMOs at higher energy levels closer to LUMOs.

Having the highest reaction barrier (59.07 kJmol⁻¹) the rate-determining step (RDS) is, therefore, step 1, and the rate law is as seen in equation 5.2.

$$\frac{d[P]}{dt} = k_1[A][C] \quad (5.2)$$

Where $k_1=k_{\text{obs}}$.

The reaction is first order in A and C, zeroth order in B and D and second order overall. Table 4.10 presents the results of the rate constants of the various elementary steps calculated at 298.15 K from the Eyring equation (equation 5.3), in which ΔS^\ddagger is the entropy of activation (Engel and Reid, 2006; Puri *et al.*, 2014). The equation is applicable to only solution-phase bimolecular reactions (Engel and Reid, 2006).

$$k = e \frac{K_B T}{h c^0} e^{\Delta S^\ddagger / R} e^{-E_a / RT} \quad (5.3)$$

Since, $k_1=k_{\text{obs}}$, the overall rate constant (k_{obs}) is taken to be $4.04 \times 10^{-8} \text{ dm}^3 \text{ Mol}^{-1} \text{ s}^{-1}$.

5.8.2 Kinetic parameters of the candidate-catalyzed reaction.

The mechanism steps of the reaction are given in scheme 5.2, which involves seven elementary steps. The activation energies were calculated using equation 5.1, which is applicable to both unimolecular and bimolecular solution-phase reactions.

Scheme 5.2: The elementary steps in the mechanism of the candidate-catalyzed aldol reaction of acetone and 4-nitrobenzaldehyde.

The equation was applied using the temperature of the reactions (298.15 K) and molar gas constant of $8.314JK^{-1}mol^{-1}$. The values of the activation energies are presented in Table 4.15, from which step 4 appeared to have the highest E_a value ($54.20 kJmol^{-1}$). Step 4 is therefore considered as the rate-limiting step and eventually, we derived the rate law of the reaction. The detailed derivation is given below.

The rate equation is given as
$$\frac{d[p]}{dt} = \frac{d[I_4]}{dt} = k_4[I_3] \quad (5.4)$$

$$\frac{d[I_1]}{dt} = k_1[A][C] - k_{-1}[I_1] - k_2[I_1] + k_{-2}[I_2][D]$$

Applying steady state approximation (S.S.A),

$$k_1[A][C] - k_{-1}[I_1] - k_2[I_1] + k_{-2}[I_2][D] = 0 \quad (5.5)$$

$$\frac{d[I_2]}{dt} = k_2[I_1] - k_{-2}[I_2][D] - k_3[I_2] + k_{-3}[I_3]$$

Applying S.S.A, $k_2[I_1] - k_{-2}[I_2][D] - k_3[I_2] + k_{-3}[I_3] = 0$ (5.6)

Summing 5.5 and 5.6,

$$k_1[A][C] - k_{-1}[I_1] - k_2[I_1] + k_{-2}[I_2][D] + k_2[I_1] - k_{-2}[I_2][D] - k_3[I_2] + k_{-3}[I_3] = 0$$

=

$$k_1[A][C] - k_{-1}[I_1] - k_3[I_2] + k_{-3}[I_3] = 0 \quad (5.7)$$

$$\frac{d[I_3]}{dt} = k_3[I_2] - k_{-3}[I_3] - k_4[I_3]$$

Applying S.S.A, $k_3[I_2] - k_{-3}[I_3] - k_4[I_3] = 0 \quad (5.8)$

Summing 5.7 and 5.8,

$$k_1[A][C] - k_{-1}[I_1] - k_3[I_2] + k_{-3}[I_3] + k_3[I_2] - k_{-3}[I_3] - k_4[I_3] = 0$$

$$k_1[A][C] - k_{-1}[I_1] - k_4[I_3] = 0$$

Simplifying for $[I_1]$,

$$[I_1] = \frac{k_1[A][C] - k_4[I_3]}{k_{-1}} \quad (5.9)$$

Equation 5.6 can be solved in terms of $[I_2]$ as shown in equation 5.10, taking $[D] = 1$.

$$[I_2] = \frac{k_2[I_1] + k_{-3}[I_3]}{k_{-2} + k_3} \quad (5.10)$$

Similarly, equation 5.7 can be solved in terms of $[I_2]$ as:

$$[I_2] = \frac{k_1[A][C] - k_{-1}[I_1] + k_{-3}[I_3]}{k_3} \quad (5.11)$$

Substituting 5.9 for $[I_1]$ in 5.10 and 5.11:

Equation 5.10 becomes,

$$[I_2] = \frac{k_2 \left(\frac{k_1[A][C] - k_4[I_3]}{k_{-1}} \right) + k_{-3}[I_3]}{k_{-2} + k_3} \quad (5.12)$$

Equation 5.11 becomes,

$$[I_2] = \frac{k_1[A][C] + k_{-3}[I_3] - k_{-1} \left(\frac{k_1[A][C] - k_4[I_3]}{k_{-1}} \right)}{k_3}$$

Simplifying,

$$[I_2] = \frac{k_1[A][C] + k_{-3}[I_3] - k_1[A][C] + k_4[I_3]}{k_3}$$

Simplifying,

$$[I_2] = \frac{k_{-3}[I_3] + k_4[I_3]}{k_3} \quad (5.13)$$

Equating 5.12 and 5.13,

$$\frac{k_2 \left(\frac{k_1[A][C] - k_4[I_3]}{k_{-1}} \right) + k_{-3}[I_3]}{k_{-2} + k_3} = \frac{k_{-3}[I_3] + k_4[I_3]}{k_3}$$

Simplifying,

$$\frac{k_3 k_2 k_1 [A][C] - k_4 k_3 k_2 [I_3]}{k_{-1}} + k_3 k_{-3} [I_3] = (k_{-3} [I_3] + k_4 [I_3]) (k_{-2} + k_3)$$

Further simplifying,

$$\frac{k_3 k_2 k_1 [A][C] - k_4 k_3 k_2 [I_3]}{k_{-1}} = k_{-3} k_{-2} [I_3] + k_3 k_{-3} [I_3] + k_4 k_{-2} [I_3] + k_4 k_3 [I_3] - k_3 k_{-3} [I_3]$$

Reduced to,

$$\frac{k_3 k_2 k_1 [A][C] - k_4 k_3 k_2 [I_3]}{k_{-1}} = [I_3] (k_{-3} k_{-2} + k_4 k_{-2} + k_4 k_3)$$

Simplifying,

$$[I_3](k_{-3}k_{-2}k_{-1} + k_4k_{-2}k_{-1} + k_4k_3k_{-1}) = k_3k_2k_1[A][C] - k_4k_3k_2[I_3]$$

Further simplifying,

$$[I_3](k_{-3}k_{-2}k_{-1} + k_4k_{-2}k_{-1} + k_4k_3k_{-1} + k_4k_3k_2) = k_3k_2k_1[A][C]$$

Solving for $[I_3]$,

$$[I_3] = \frac{k_3k_2k_1[A][C]}{k_{-3}k_{-2}k_{-1} + k_4k_{-2}k_{-1} + k_4k_3k_{-1} + k_4k_3k_2}$$

Substituting for $[I_3]$ in the rate equation

$$\begin{aligned} \frac{d[p]}{dt} &= k_4 \left(\frac{k_3k_2k_1[A][C]}{k_{-3}k_{-2}k_{-1} + k_4k_{-2}k_{-1} + k_4k_3k_{-1} + k_4k_3k_2} \right) \\ \frac{d[p]}{dt} &= \left(\frac{k_4k_3k_2k_1}{k_{-3}k_{-2}k_{-1} + k_4k_{-2}k_{-1} + k_4k_3k_{-1} + k_4k_3k_2} \right) [A][C] \end{aligned} \quad (5.14)$$

Let

$$k_{obs} = \frac{k_4k_3k_2k_1}{k_{-3}k_{-2}k_{-1} + k_4k_{-2}k_{-1} + k_4k_3k_{-1} + k_4k_3k_2} \quad (5.15)$$

Therefore, the rate equation can be represented as,

$$\frac{d[p]}{dt} = k_{obs} [A][C] \quad (5.16)$$

The reaction is second order overall, being first order with respect to A and C; and zeroth order with respect to B and D.

Finally, we applied the Eyring equation (5.3), which is applicable to both unimolecular and bimolecular solution-phase reactions, to calculate all the elementary rate constants that appear in the rate law.

Table 4.15 shows the values of the rate constants of the elementary steps at 298.15 K. From the Table, the overall rate constant (k_{obs}) is computed as follows:

$$k_4 k_3 k_2 k_1 = (5.001 \times 10^3)(2.504 \times 10^7)(6.805 \times 10^{11})(2.473 \times 10^5) = 2.107 \times 10^{28}$$

$$k_{-3} k_{-2} k_{-1} = (2.566 \times 10^9)(4.688 \times 10^{12})(2.224 \times 10^6) = 2.675 \times 10^{28}$$

$$k_4 k_{-2} k_{-1} = (5.001 \times 10^3)(4.688 \times 10^{12})(2.224 \times 10^6) = 5.214 \times 10^{22}$$

$$k_4 k_3 k_{-1} = (5.001 \times 10^3)(2.504 \times 10^7)(2.224 \times 10^6) = 2.785 \times 10^{17}$$

$$k_4 k_3 k_2 = (5.001 \times 10^3)(2.504 \times 10^7)(6.805 \times 10^{11}) = 8.522 \times 10^{22}$$

$$k_{obs} = \frac{2.107 \times 10^{28}}{2.675 \times 10^{28} + 5.214 \times 10^{22} + 2.785 \times 10^{17} + 8.522 \times 10^{22}}$$

$$k_{obs} = 0.787 \text{ M}^{-1} \text{ s}^{-1}$$

CHAPTER SIX

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Considering the large coefficient on the descriptor maxssNH in the selected model (70.42) and the large factor of the PNSA-3 (10^1), it is concluded that the enantioselectivity is most influenced by these two descriptors. This is not surprising as the maxssNH quantifies the effect of nitrogen, which is the key atom on the catalysts, while PNSA-3 is the only independent 3D descriptor in the model, and therefore as expected, it influences the stereochemical properties of the catalysts.

The proposed pyrrolidine-based organocatalysts (Candidates 1 and 2) were found to have higher ee values (85% and 80% respectively) than the proline (76%) in acetone medium.

The mechanism of direct aldol reaction between acetone and 4-nitrobenzaldehyde acted upon by the proposed catalyst has been investigated in acetone medium via density functional theory computation. Quantum mechanical descriptors of HOMO and LUMO have been used to explore various mechanistic steps. The results of this study agree with the already proposed iminium-enamine mechanism (Jung, 1976; Klalwani and Sudaai, 2016). However, an intermediate (I_3) and many transition states (TS3, TS4, TS6 and TS7) that were not seen in the previous works on this subject have been identified. The analysis of IR plots showing one and only imaginary frequency in each case, confirms them to be true transition states.

The enantiomeric excess value of proline is found to be 76% and 83 % from empirical and computational studies (using the model adopted in the current study), respectively. Whereas the enantiomeric excess value of the candidate catalyst is 85%. The rate constant of the asymmetric aldol reaction catalyzed by proline is found to be $4.04 \times 10^{-8} \text{ dm}^3 \text{ mol}^{-1} \text{ s}^{-1}$ while the rate constant of the reaction as catalyzed by the candidate catalyst is $0.787 \text{ dm}^3 \text{ mol}^{-1} \text{ s}^{-1}$. It can therefore be concluded that the designed catalyst is theoretically better than proline in both selectivity and activity. The appreciably high rate constant of the candidate-catalyzed reaction indicates high activity of the catalyst, making it a viable catalyst.

The energy position of the product in the candidate-catalyzed reaction ($-8.68 \text{ kJ mol}^{-1}$) is same as in the case of proline-catalyzed reaction. This proves that, irrespective of the catalyst used, the thermodynamic fate of the reaction is never altered.

6.2 Recommendations

It is recommended that the simulation reaction conducted using the candidate catalyst be repeated in different media (e.g: DMF, DMSO) to analyse the effect of solvent on the catalyst. More so, the catalyst should be applied on some other asymmetric aldol reactions of different carbonyl compounds combinations to test the diversity of the catalyst.

If it passes adequate computational assessment, it is suggested that the candidate should be synthesized for further testing by wet laboratory (*in vitro*) experiments.

REFERENCES

- Agami, C., Meynier, F., Puchot, C., Guilthem, J., and Pascard, C. (1984). Stereochemistry-59: New insights into the mechanism of the proline-catalyzed asymmetric robinson cyclization; structure of two intermediates. asymmetric dehydration. *Tetrahedron*. 40(6), 1031-1038.
- Almasi, D., Alonso, D.A., and NaÂjera, C. (2008). prolinamide versus prolinethionamides as recyclable catalysts in the enantioselective solvent-free inter- and intramolecular aldol reactions. *Advanced Synthesis and Catalysis*. 350, 2467-2480.
- Balabin, R.M. (2010). Intramolecular basis set superposition error as a measure of basis set incompleteness: Can one reach the basis set limit without extrapolation?. *Journal of Chemical Physics*. 132, 211103: 1032-1039. [doi: org/10.1063/1.3430647](https://doi.org/10.1063/1.3430647)
- Benaglia, M., Cinquini, M., Cozzi, F., and Puglisi, A. (2001). Enantioselective aldol condensations catalysed by poly(ethylene glycol)-supported proline. *Advanced Synthesis and Catalysis*. 344, 533-540.
- Benigni, R., and Bossa, C. (2008). Predictivity of QSAR. *Journal Chemical Information and Modelling*. 48, 971-980.
- Berkessel, A., Koch, B., and Lex, J. (2004). Proline-Derived *N*-Sulfonylcarboxamides: Readily Available, Highly Enantioselective and Versatile Catalysts for Direct Aldol Reactions.

- Advanced Synthesis and Catalysis*. 346(9-10), 1141-1146.
doi:10.1002/adsc.200404126
- Berkessel, A., and Groeger, H. (2005). *Asymmetric Organocatalysis*. Wiley-VCH, Weinheim, Germany, pp.57-58. doi:3.527/305173.
- Bhoorasingh, P. L., and West, R. H. (2015). Transition state geometry prediction using molecular group contributions. *Physical Chemistry Chemical Physics*.17, 32173.
- Bhoorasingh, P.L., and West, R.H. (2015). Transition state geometry prediction using molecular group contributions. *Physical Chemistry Chemical Physics*.17(48), 3217-3223. doi: 10.1039/C5CP04706D
- Blackmond, D.G., Zotova, N., Broadbelt, L.J., and Armstrong, A. (2009). Clarification of the Role of Water in Proline-Mediated Aldol Reactions. *Bioorganic and Medicinal Chemistry Letters*.19(1), 3934-3938.
- Blackmond, D.G. Zotova, N., Franzke, A., and Armstrong, A. (2007). Role of Water in Proline-Mediated Aldol Reactions. *Journal of American Chemical Society*.129(49), 15100-15101. doi:10.1021/ja0738881
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics For Experimenters: An introduction to Design, Data Analysis, and Model Building*. John Wiley & Son, New York, U.S.A, pp. 890-891.
- Boyd, D., and Lipkowitz, K. (Eds.). (1991). *The Molecular Connectivity Chi Indices and Kappa Shape Indices in StructureProperty Relations. In Reviews of Computational Chemistry*. VCH Publishers, Inc., New York, U.S.A, pp. 7422-7423.
- Ceotto, M., Orlandi, M., and Benaglia, M. (2016). Kinetics versus thermodynamics in the proline-catalyzed aldol reaction. *Chemical Science*. 7, 5421-5427. doi:10.1039/c6sc01328g
- Cheng, C., Zhang, M., Xue, C., Bai, F., and Zhao, X. (2017). Development of stress tolerant *Saccharomyces cerevisiae* strains by metabolic engineering: New aspects from cell flocculation and zinc supplementation. *Journal of Bioscience and Bioengineering*. 123(2), 141-146.
- Cheng, J.P., Lu, S., Mi, X., Zhang, L., Liu, S., and Xu, H. (2006). Functionalized Chiral Ionic Liquids as Highly Efficient Asymmetric Organocatalysts for Michael Addition to Nitroolefins. *Angewandte Chemie*. 118(19), 3164-3169. doi:10.1002/ange.200600048
- De, D., Bhattacharyya, A., and Bharadwaj, P. K. (2017). Enantioselective aldol reactions in water by a proline-derived cryptand and fixation of CO₂ by its exocyclic Co(II) complex. *Inorganic Chemistry*. 56 (18), 11443–11449.
- Dhimitruka, I., and SantaLucia, J. (2006). Synthesis of a Lux-S Enzyme Inhibitor Using an Improved Esterification Method, *Organic Letters*.8, 47-50.
- DineÂr, P., and Amedjkouh, M. (2006). Conscience synthesis of salinoprolineamide. *Organic and Biomolecular Chemistry*.42(10), 91-95.

- Dixon, S. L., Merz, K. M. Jr., Lauri, G., and Ianni, J.C. (2004). QM-QSAR: Utilization of a Semiempirical Probe Potential in a Field-Based QSAR Method. *Journal of Computational Chemistry*.26(1), 23-25.
- Drauz, K., and Waldmann, H. (1995). *Enzyme catalysis in organic synthesis: A comprehensive handbook*. VCH, Weinheim, Germany, pp851-890.
- Dziedzic, P., Zou, W., Hafren, J., and Cordova, A. (2006). The small peptide-catalyzed direct asymmetric aldol reaction in water. *Organo bimolecular chemistry*.4(1), 38-40. doi:[10.1039/b515880j](https://doi.org/10.1039/b515880j)
- Eder, U., Sauer, G., and Wiechert, R. (1971). New type of asymmetric cyclization to optically active steroid C.D partial structures. *Angewandte Chemie International Edition*.10(1) 496-470.
- El-Bakouri, O., Garcia-Borràs, M., Girón, R. M., Filippone, S., Martín, N., and Solà, M. (2018). On the regioselectivity of the Diels–Alder cycloaddition to C₆₀ in high spin states . *Physical Chemistry Chemical Physics*.20(17), 11577–11585. doi: 10.1039/C7CP07965F
- Elschenbroich, C. (2006). *Organometallics*. Wiley-VCH, Weinheim, Germany, pp. 220-221. doi: 978.3/527293902.
- Enders, D., and Grondal, C. Direct Organocatalytic De Novo Synthesis of Carbohydrates *Angewandte. Chemie. International Edition*.2005, 44(8), 1210-1212. [doi: 10.1002/anie.200462428](https://doi.org/10.1002/anie.200462428)
- Engel, T., and Reid, P. (2006). *Physical Chemistry*. Pearson Prentice Hall, New York, U.S.A, pp. 631-925.
- Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., and McDowell, R. M. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspective*. 111(10), 1361-1366
- Eyring, H. (1935). The Activated Complex and the Absolute Rate of Chemical Reactions. *Chemical Review*.17(1) 65-69.
- Farrusseing, D., Rothenberg G., and Burello, B. (2004). Combinatorial Explosion in Homogeneous Catalysis: Screening 60,000 Cross-Coupling Reactions. *Advanced Synthesis and Catalysis*.346(13), 1844.
- Fukui K. (1965). Stereospecificity with reference to some cyclic reaction. *Tetrahedron Letters*. 24, 2009.
- Furuta, T., Baba, K., Yamamoto, J., Hayashi, K., Sato, M., Yamanaka, M., and Kawabata, T. (2016). Catalytic discrimination between formyl groups in regio- and stereoselective intramolecular crossaldol reactions. *Chemical Science*.7, 3791-3791. doi: 10.1039/c5sc04594k
- Gaspar, R. (1954). Simulation of Heterogeneous Catalysis. *Acta Physica. Academiae Scientiarum Hungaricae*. 3, (1), 263-267.

- Ges`u, G.D., Leli`evre, T., Peutrecab, D.L., and Nectouxa, N. (2016). Jump Markov models and transition state theory: the quasi-stationary distribution approach. *Faraday Discuss.* **195**(1), 469-495. doi: 10.1039/C6FD00120C
- Gryko, D., and Lipin' ski, R. (2005). L-prolinamide-Efficient organocatalyst for Asymmetric Aldol Reaction. *Advanced Synthesis and Catalysis.* **347**(15), 1948-1953.
- Gryko, D., & Lipin'ski, R. (2006). Asymmetric Direct Aldol Reaction Catalysed by L-Prolinethioamides. *European Journal of Organic Chemistry.* **2006**(17), 3864-3876. doi: 10.1002/ejoc.200600219
- Haindl, M. H., Hioe, J., and Gschwind, R. M. (2015). The proline-enamine formation pathway revisited in DMSO—rate constants determined via NMR. *Journal of American Chemical Society.* **137**(40), 12835–12842.
- Hajos, Z. G. and Parrish, D. R. (1971). Asymmetrische Synthese polycyclischer organischer Verbindungen. *Ger. Pat.* **21**(10), 2623-2627.
- Hajos, Z.G, and Parrish, D.R (1974). Asymmetric synthesis of bicyclic intermediates of natural product chemistry. *Journal of Organic Chemistry.* **39** (1), 1615-1621.
- Hammer, B. Nørskov, J. K. (200). Theoretical surface science and catalysis — calculations and concepts. *Advanced Catalysis.* **45**, 71–129.
- Hartikka, A., and Arvidsson, P.I. (2004). Rational design of asymmetric organocatalysts—increased reactivity and solvent scope with a tetrazolic acid. *Tetrahedron: Asymmetry.* **15**(12), 1831-1834.
- Hayashi, Y., Yamaguchi, J., Hibino, K., Sumiya, T., Urushima, T. Shoji, M., Hashizume, D., and Koshino, H. (2004). A Highly Active 4-Siloxyproline Catalyst for Asymmetric Synthesis. *Advanced Synthesis and Catalysis.* **346**, 1435–1439. doi: 10.1002/adsc.200404166
- Hehre, W.J. (1998). *A brief guide to molecular mechanics and quantum chemical calculations / Warren J Hehre*. Wavefunction Inc., Irvine, U.S.A, pp. 298-299.
- Helmchen, G., Hoffmann, R.W., Mulzer, J., and Schaumann, E. (1995). *Hoube-Weyl Method of Organic Chemistry (4th Ed.)*. Goerge Thiem Verlag Stuttgart, New York, U.S.A, pp. 3961-3962.
- Hoffmann, R., and Woodward, R. B. (1965). Orbital symmetries and endo-exo relationships in concerted cycloaddition reactions. *Journal of American Chemical Society.* **87**, 4388-4389.
- Honkala, J.K., Hellman, A., Remediakis, I., Logadottir, A., Carlsson, A., Dahl, S., Christensen, C.H., and Norskov, J. K. (2005). Ammonia synthesis from first-principles calculations. *Science.* **307**(5709), 555-558. Doi: 10.1126/science.1106435
- Houk, K.N, & Bahmayar, S. (2001). The Origin of Stereoselectivity in Proline-Catalyzed Intramolecular Aldol Reactions. *Journal of American Chemical Society.* **123**, 12911-12912.

- Hu, H., Juvekar, A., Lyssiotis, C.A., Lien, E.C., Albeck, J.G., Oh, D., Verma, G., Hung, Y.P., Ullas, S., Lauring, J., Seth, P., Lundquist, M.R., Tolan, D.R., Grant, A.K., Needleman, D.J., Asara, J.M., Cantley, L.C., and Wulf, G.M. (2016). Phosphoinositide 3-kinase regulates glycolysis through mobilization of aldolase from the actin cytoskeleton. *Cell*. 164(3) 433–446. doi: [10.1016/j.cell.2015.12.042](https://doi.org/10.1016/j.cell.2015.12.042)
- Jacobson, L.D., Bochevarov, A.D., Watson, M.A., Hughes, T.F., Rinaldo, D., Ehrlich, S., and Friesner, R.A. (2017). Automated transition state search and its application to diverse types of organic reactions. *Journal of Chemical Theory and Computation*.13(11) 5780–5797. doi :[10.1021/acs.jctc.7b00764](https://doi.org/10.1021/acs.jctc.7b00764).
- Jin,H., Kim,S.T., Hwang, G.S., and Ryu, D.H. (2016).L-Proline derived bifunctional organocatalysts: enantioselective michael addition of dithiomalonates to trans- β -nitroolefins. *Journal of Organic Chemistry*. 81 (8): 3263–3274.
- Jung, M.E. (1976). A review of annulation.*Tetrahedron*. 32(1), 3-31.
- Kandoi, S., Greeley, J., Sanchez-Castillo, M. A., Evans, S. T., Gokhale, A.A, Dumesic, J. A., and Mavrikakis, M. (2006).Prediction of experimental methanol decomposition rates on platinum from first principles. *Top Catalysis*. 37(1), 17–28.
- Kier, L. B., and Hall, L. H. (1986).*Molecular Connectivity in Structure-Activity Analysis*.John Wiley and Sons, New York, U.S.A, pp88-89.
- Klalwani, K.G., and Sudaai, A. (2016). Proline-catalyzed enantioselective synthesis of 2,3-disubstituted piperidines. *Synlett*.27(1), 1339–1343.
- Kokotos, G., Bellis, E., and Vasilatou, K. (2005).4-Substituted Prolyl Sulfonamides as Enantioselective Organocatalysts for Aldol Reactions.*Synthesis*.2005(14).2407-2413. doi: 10.1055/s-2005-870026
- Kokotos, G., Limnios, D., Triggidou, D., Trifonidou, M. and Kokotos, C.G. (2011).Novel pyrrolidine-thiohydantoin/thioxotetrahydropyrimidinones as highly effective catalysts for the asymmetric Michael addition.*Organic and Biomolecular Chemistry*.9, 3386-3390. doi: 10.1039/c0ob01083a
- Kononov, D. A., Llewellyn, L.E., Heyden, Y. V., and Coomans, D. J. (2004).Robust cross-validation of linear regression QSAR models.*Journal of Chemical Information and Computer Science*. 44, 105-109.
- Kozlowski, M.C, Ianni, J.C., Annamalai, V., Phuan, P.W., and Panda, M. (2006). A Priori Theoretical Prediction of Selectivity in Asymmetric Catalysis: Design of Chiral Catalysts by Using Quantum Molecular Interaction Field. *Angewandte Chemie International Edition*.45, (1), 5502-5509.
- [Kozlowski](#), M. C., [Phuan](#), P. W., and [Ianni](#), J. C. (2004).Is the A-Ring of Sparteine Essential for High Enantioselectivity in the Asymmetric Lithiation–Substitution of *N*-Boc-pyrrolidine?.*Journal of American Chemical Society*.126(47), 15473–15479. doi: 10.1021/ja046321i

- [Kozłowski](#), M. C., Dixon, S. L., [Panda](#), M., and [Lauri](#), G. (2003). Quantum Mechanical Models Correlating Structure with Selectivity: Predicting the Enantioselectivity of β -Amino Alcohol Catalysts in Aldehyde Alkylation. *Journal American Chemical Society*. 125(22), 6614-6615. doi:10.1021/ja0293195
- Kutzscher, C., Nickerl, G., Senkovska, I., Bon, V., and Kaskel, S. (2016). Proline functionalized UiO-67 and UiO-68 type metal-organic frameworks showing reversed diastereoselectivity in aldol addition reactions. *Chemistry of Materials*. 28 (8), 2573–2580
- Lacoste, E., Landais, Y., Schenk, K., Verlhac, J.B., and Vincent, J.M. (2004). Benzoimidazole–pyrrolidine (BIP), a highly reactive chiral organocatalyst for aldol process. *Tetrahedron Letters*. 45(13), 8035-8038.
- Lambert, T. H., Bandar, J. S., Sauer, G. S., Wulff, W. D., and Veticatt, M. J. (2014). Transition State Analysis of Enantioselective Brønsted Base Catalysis by Chiral Cyclopropenimines. *Journal American Chemical Society*. 136 (30), 10700–10707. doi: 10.1021/ja504532d
- Legler, C.R., Brown, N.R., Dunbar, R.A., Harness, M.D., Nguyen, K., Oyewole, O., and Collier, W.B. (2015). Scaled quantum mechanical scale factors for vibrational calculations using alternate polarized and augmented basis sets with the B3LYP density functional calculation model. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 145(1), 15-24. doi: [10.1016/j.saa.2015.02.103](https://doi.org/10.1016/j.saa.2015.02.103)
- Lelais, G., and MacMillan, D.W.C. (2006). *Modern Strategies in Organic Catalysis: The Advent and Development of Iminium Activation (PDF)*. *Aldrich mica Acta* 39(3), 79.
- Lewers E. G. (2016). *Computational Chemistry*. Springer Nature, Peterborough, England, pp. 253-258.
- Ley, S.V., Cobb, A. J. A., Shaw, D.M., Longbottom, D.A., and Gold, J.B. (2005). Organocatalysis with proline derivatives: improved catalysts for the asymmetric Mannich, nitro-Michael and aldol reactions. *Organic and Biomolecular Chemistry*. 3, 84-96.
- List B. (2000). The Direct Catalytic Asymmetric Three-Component Mannich Reaction. *Journal of American Chemical Society*. 122(38), 9336-9337. doi: 10.1021/ja001923x
- List, B. (2010). *Emil Knoevenagel and the Roots of Aminocatalysis*. *Angewandte Chemie*. 49 (10): 1730–1734.
- List, B., Hoang, L., and Martin, H.J. (2004). New mechanistic studies on the proline-catalyzed aldol reaction. *National Academy of Science, USA*. 101(16), 5839-5842
- List, B., Hoang, L., Bahmanyar, S., and Houk, K.N. (2003). Kinetic and stereochemical evidence for the involvement of only one proline molecule in the Transition States of proline-catalyzed intra- and intermolecular aldol reactions. *Journal of American Chemical Society*. 125 (1), 16–17.
- List, B., Hoffmann, T., Zhong, G., Shabat, D., Anderson, J., and Barbas, C. F. III. (1998). Aldolase Antibodies of Remarkable Scope. *Journal of American Chemical Society*. 120(12), 2768-2779. doi: 10.1021/ja973676b

- List, B., Lerner, R.A., and Barbas, C.F. III. (2000). Proline-catalyzed direct Asymmetric aldol reactions. *Journal of American Chemical Society*. 122(10), 2395-2396. doi: [abs/10.1021/ja994280y](https://doi.org/10.1021/ja994280y)
- Liu, J.B., Liu, L.J., Dong, Z.Z., Yang, G.J, Leung, C.H. and Ma, D.L. (2016). An aldol reaction-based iridium (III) chemosensor for the visualization of proline in living cells. *Scientific Reports*. 6 (36) 509-515.
- Lloyd, L. (2011) *Handbook of industrial catalysts*. Springer Science Business Media, New York, U.S.A, pp. 6-8.
- Maeda, S., Harabuchi, Y., Ono, Y., Taketsugu, T., and Keiji, M. (2015). Intrinsic reaction coordinate: calculation, bifurcation, and automated search. *International Journal of Quantum Chemistry*. 115 (5), 258-269.
- Mangion, I. K., and MacMillan, D. W. C. (2005). Total Synthesis of Brasoside and Littoralisone. *Journal of American Chemical Society*. 127(11), 3696-3697.
- Martínez, A., Zumbansen, K., Döhring, A., Gemmeren, M., and List, B. (2014). Improved conditions for the proline-catalyzed aldol reaction of acetone with aliphatic aldehydes. *Synlett*. 25(7) 932-934.
- McQuaid, M. J., Sun, H. and Rigby, D. (2004). Development and Validation of COMPASS Force Field Parameters for Molecules with Aliphatic Azide Chains. *Journal of Computational Chemistry*, 25, 61-71.
- Moorthy, J. N., and Saha, S. (2009). Highly Diastereo- and Enantioselective Aldol Reactions in Common Organic Solvents Using *N*-Arylprolinamides as Organocatalysts with Enhanced Acidity. *European Journal of Organic Chemistry*. 2009(5), 739–748 doi: 10.1002/ejoc.200800960
- Nelson, D. L., and Cox, M. M. (2000) *Lehninger, Principles of Biochemistry* (3rd Ed). Worth Publishing, New York, U.S.A, pp.55-56. doi:1.572591536.
- [Norrby](#), P. O., [Oslob](#), J. D., [Åkermark](#), B., [and Helquist](#), P., (1997). Steric Influences on the Selectivity in Palladium-Catalyzed Allylation. *Organometallics*, 16(13), 3015-3021. doi:10.1021/om9700371
- Nørskov, J.K., Bligaard, T., Rossmeisl, J., and Christensen, C.H. (2009). Towards the computational design of solid catalysts. *Nature Chemistry*. 1, 37–46. doi:10.1038/nchem.121
- Northrup, A. B., and MacMillan, D. W. C. (2002). The First Direct and Enantioselective Cross-Aldol Reaction of Aldehydes. *Journal of American Chemical Society*. 124(24) 6798-6799. doi: 10.1021/ja0262378
- Northrup, A. B., and MacMillan, D. W. C. (2004). Enantioselective Organocatalysis. *Science*. 2004. 305, 1752-1755.
- Ochterski, J. W. (2000). *Thermochemistry in Gaussian, Technical Support Information*. Gaussian Inc, Pittsburgh, PA, pp. 1-19.

- Ostwald, W. (1900). Catalysis. *Zeitschrift für Physikalische Chemie*. 34(1), 495-503
- Park, K. S., Kim, J., Choo, H., and Chong, Y. (2007). Octanol-Accelerated Baylis-Hillman Reaction, *Synlett*. 3, 395-398. doi: 10.1055/s-2007-968032
- Perrin, C.L., and Chang, K. L. (2016). The complete mechanism of an aldol condensation. *Journal of Organic Chemistry*. 81 (13) 5631–5635. doi: 10.1021/acs.joc.6b00959
- Pople, J. A., Ditchfield, R., and Hehre, W.J. (1971). Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *Journal of Chemical Physics*. 54(2), 724–728.
- Puri B.R., Sharma L.R., and Pathania, M.S. (2014). *Engineering Chemistry*. Vishal publishing co., Chandigarh, India. pp.120-121.
- Rankin, K.N., Gauld, J.W., and Boyd, R. J. (2002). Density Functional Studies of the Proline-XCatalyzed Direct Aldol Reaction. *Journal of Physical Chemistry*. 106(20), 5155-5159. doi: 10.1021/jp020079p
- Richter, V. (1869). Condensation of heptanal and pentanal. *The Chemical News*. 20(9), 286-288.
- Sakthivel, K., Notz, W., Bui, T., and Barbas C.F. III (2001). Amino Acid Catalyzed Direct Asymmetric Aldol Reactions: A Bioorganic Approach to Catalytic Asymmetric Carbon–Carbon Bond-Forming Reactions. *Journal of American Chemical Society*. 123(22), 5260-5267. doi: 10.1021/ja010037z
- Sanchez, D., Bastida, D., Bures, J., Isart, C., Pineda, O., and Vilarrasa, O. (2012). Efficient Preparation of *N*-Phenylsulfenyl Ketimines from Oximes or Nitro Compounds without Racemization of α -Stereocenters. *Organic Letters*. 14(4), 536-540.
- Sharma, A.K., and Sunoj, R. B. (2010). Enamine versus oxazolidinone: what controls stereoselectivity in proline-catalyzed asymmetric aldol reactions? *Angewandte Chemie International Edition*. 49(36), 6373-6379.
- Siaka, A. A. (2017). *Computational Study of the Formation Mechanisms of Some Semicarbazones and Spirro Heterocycles*. Zaria: A.B.U Zaria. pp 43-45.
- Singh, V. K., Raj, M., and Ginotra, S.K. (2006). Highly Enantioselective Direct Aldol Reaction Catalyzed by Organic Molecules. *Organic Letters*. 8 (18), 4097–4099. doi:10.1021/ol0616081
- Sofie, V. (2009). Quantum Chemistry in QSAR: Quantum chemical descriptors, uses, benefits and drawbacks. *Ghent University*. 18, 110-113.
- Spartan' 14 Wavefunction, Inc. Irvine, CA. 2014.
- Stanton, D. T., and Jurs, P. S. (1990). Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. *Analytical Chemistry*. 62, 2323-2329.

- Swaminathan, S., Rajapopal, D., Moni, M. S., and Subramanian, S. (1999). *Proline mediated asymmetric ketol cyclization: a template reaction*. *Tetrahedron Asymmetry*. 10(9), 1631-1634.
- Tanaka, M., Takahashi, Y., and Yamaguchi, N. (2013). A study on adsorption mechanism of organoarsenic compounds on ferrihydrite by XAFS. *Journal of Physics Conference Series*.430(1), 12100-12110.
- Tang, X, Liegault, B, Renaud, J.L, Bruneau, C. (2006). *Tetrahedron: Asymmetry*. 17(2),187-191
- Tang, Z., Jiang, F., Cui, X., Gong, L. Z., Mi, A. Q., Jiang, Y. Z., and Wu, Y.D. (2004). Enantioselective direct aldol reactions catalyzed by L-prolinamide derivatives. *Proceedings of the National Academy of Sciences of the United States of America*. 101(16), 5755–5760 . doi_10.1073_pnas.0307176101
- Tang, Z., Yang, Z. H., Chen, X. H., Cun, L. F., Mi, A. Q., Jiang, Y. Z., and Gong, L. Z.. (2005). A highly efficient organocatalyst for direct aldol reactions of ketones with aldehydes. *Journal of American Chemical Society*.127, 9285-9289.
- Tong, S.T., Harris, P.W.R., Barker, D., and Brimble, M.A. (2008). Use of (S)-5-(2-Methylpyrrolidin-2-yl)-1H-tetrazole as a Novel and Enantioselective Organocatalyst for the Aldol Reaction. *European Journal of Organic Chemistry*.2008(1), 164-170. doi:10.1002/ejoc.200700834
- Tropsha, A., Gramatica, P., and Gombar. V. K. (2003). *Molecular informatics*.ACS *Combinatorial Science*. 22, 69-77.
- Trost, B. M., Saget, T., & Hung C. I. (2016). Direct catalytic asymmetric mannich reactions for the construction of quaternary carbon stereocenters. *Journal of American Chemical Society*.138 (11) 3659–3662.doi: 10.1021/jacs.6b01187
- Wold, S., Eriksson, L., and Vande, H. (1995). *Chemometrics Methods in Molecular design*. VCH, Weinheim, Germany, pp. 309-318.
- Wong, C.H., and Whiteside, G.M. (1994). *Enzymes in Synthetic Organic Chemistry (1st Ed.)*. Elsevier, New York, U.S.A, pp306-307.
- [Wurtz, C A.](#)(1872).*On an aldehyde alcohol*.[Bulletin de la Société Chimique de Paris](#).17(2) 436-442.
- Xu, L., Zhang, Q. Y., Zhang, D. D., Li, J. Y., and Long H.L. (2012). Prediction of Enantiomeric Excess in a Catalytic Process: A Chemoinformatics Approach Using Chirality Codes. *MATCH Communications in Mathematical and in Computational Chemistry*. 67, 773-786
- Ye, R., Zhao, J., Yuan, B., Liu, W.C., Faucher, F.F., Chang, M., Dereadt, C.V., Toste, F.D., and Somorjai, G.A. (2017).New insights into aldol reactions of methyl isocyanoacetate catalyzed by heterogenized homogeneous catalysts. *Nano Letters*.17 (1), 584–589. doi: 10.1021/acs.nanolett.6b04827

- Zhao, C. G., Samanta, S., Liu, J., and Dodda, R. (2005). *C*₂-Symmetric Bisprolinamide as a Highly Efficient Catalyst for Direct Aldol Reaction. *Organic Letters*. 7(23), 5321-5328.
- Zhao, Y., and Truhlar, D. G. (2008). Density functionals with broad applicability in chemistry. *Accounts of Chemical Research*. 2, 157–167.
- Zlotin, S.G., Kucherenko, A.S., and Beletskaya, I.P. (2009). Organocatalysis of asymmetric aldol reaction. *Catalysis and reagents. Russian Chemical Reviews*. 78 (8), 231-235.

APPENDICES

APPENDIX I

Equations for calculation of Internal Validation Parameters.

1. Correlation coefficient of multiple determinations (R^2)

$$R^2 = \frac{SSR}{SST} \quad (1)$$

Where, SSR= sum of squares of regression, SST = total sum of squares

2. Adjusted R^2

$$1 - \frac{SSE/(n-p)}{SST/(n-1)} \quad (2)$$

Where, p is the number of parameters in the regression model and n is the number of data points from which the model was built.

3. Cross validated R-squared (R_{cv}^2)

$$R_{cv}^2 = 1 - \frac{PRESS}{SST} \quad (3)$$

Where PRESS = predictive sum of squares, SST = total sum of squares

4. F Test

$$F = \frac{SSR/(p-1)}{SSE/(n-p)} \quad (4)$$

Where, SST; the total sum of squares is the mean-corrected sum of squares of the responses over the entire data set. SSE; the sum of squares of errors is the sum of squared differences between measurements and model predictions over the entire data set. SSR; the sum of squares due to regression is given by $SSR = SST - SSE$. n is the number of data points from which the model is built. p is the number of parameters in a regression model (including the intercept, when present).

5. Correlation Coefficient of External Validation (Q_{ext}^2)

$$Q_{ext}^2 = 1 - \frac{\sum_i (y_p - y_e)^2}{\sum_i (y_e - w)^2} \quad (6)$$

Where, y_p is the predicted value of y, w is the mean of the experimental y values of the training set, y_e is the experimental value of y,

6. Pearson correlation coefficient of prediction (R_{ext})

$$R_{ext} = \frac{\sum_i (y_e - \langle y_e \rangle)(y_p - \langle y_p \rangle)}{\sqrt{\sum_i (y_e - \langle y_e \rangle)^2} \sqrt{\sum_i (y_p - \langle y_p \rangle)^2}} \quad (7)$$

Where, y_p is the predicted value of y, y_e is the experimental value of y, $\langle y_e \rangle$ is the mean of y_e values and $\langle y_p \rangle$ is the mean of y_p values.

APPENDIX II

Equations for Calculating Model One Descriptors.

1. PNSA-3: path weighted partial negative surface area

$$PNSA_3 = \sum_a q_a^- SA_a^- \quad (1)$$

Where, q_a^- = partial charge over negatively charge atom a. SA_a^- = solvent accessible surface area of the negatively charged atom a.

2.DPSA-1: difference in charged partial surface area

$$DPSA_1 = PPSA_1 - PNSA_1 \quad (2)$$

$$PPSA_1 = \sum_{a^+} SA_a^+ \quad (3)$$

$$PNSA_1 = \sum_{a^-} SA_a^- \quad (4)$$

3. VP-6: valence path, order 6

$$6\chi_{vp} = \prod_{k=1}^{m+1} (\delta_k^v)^{1/2} = \prod_{k=1}^{m+1} (\sigma_k + \pi_k + n_k - h_k)^{1/2} \quad (5)$$

$$6\chi_{vp} = \prod_{k=1}^{m+1} \left[\frac{Z_k^v - h_k}{Z_k - Z_k^v - 1} \right]^{\frac{1}{2}} \quad (6)$$

Where, σ = number of electrons in sigma orbitals, π = number of electrons in pi orbitals, n = number of electrons unbounded, h = number of hydrogen atoms bonded to atom k , δ_k^v = valence vertex degree, Z_k = atomic number of atom k , Z_k^v = number of valence electrons in atom k .

4. MaxssNH: maximum value of electrotopological state –NH-

$$S_k = I_k + \sum_{i=1}^A \Delta I_{ki} = \frac{\left(\frac{2}{N}\right)^2 \delta_k^v + 1}{\delta_k} + \sum_{i=1}^A \left[\frac{[(I)_i - I_k]}{(d_{ki} + 1)^2} \right] \quad (7)$$

Where, N = principal quantum number, $\delta_k^v = Z_k^v - h_k$, d_{ki} = number of bonds that separate atom k and I .

APPENDIX III

Comparison between single step and two step mechanism of iminium-enamine transformation.