

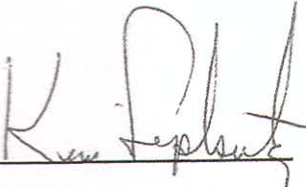
# **CoMFA STUDY OF CHIRAL CATALYSTS**

**Meeta Pradhan**

**Submitted to the faculty of the University Graduate School  
in partial fulfillment of the requirements for the  
degree of Master of Science in the  
School of Informatics Indiana University**

**May - 2003**

Accepted by the Graduate Faculty, Indiana University in partial fulfillment  
of the requirements for the degree of Master of Science.

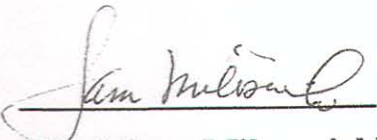
  

---

**(Prof. K. B. Lipkowitz)**

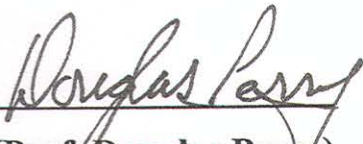
**Master's Thesis**

**Committee**

---

**(Prof. Sam Milosevich)**

---

**(Prof. Douglas Perry)**

## ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Professor K. B. Lipkowitz for giving me an opportunity to work under him. His patience in teaching me the subject and also creating my interest in computational chemistry is unaccountable.

I would like to thank my committee members for showing interest in the work.

A special thanks to Dr. Sabine Schefzick for helping me thought-out my thesis work. I would like to thank all my labmates including Dr. Ya Yin Fang, MeiMei Zang, Taka Sakamoto and Jon Stack for helping me and having made my working in the lab a pleasant memory.

I would like to thank Dr. Kelsey Forsythe for all his help with computers.

I would also like to thank Professor Rajiv Raje (Computer Science) for encouraging me to take up the Informatics course. I would like to thank my friends Mrs. Bhuvani Jelaji and Mrs. Mita Vagal for helping me in need.

Finally, a special thanks to my daughter Pranoti and husband Prashant who supported me through out my work.

## TABLE OF CONTENTS

Table of Contents.....	iv
Table of Figures.....	vi
Table of Tables.....	viii
ABSTRACT.....	ix
INTRODUCTION .....	1
The Diels-Alder cycloaddition reaction .....	7
Approaches in computer aided molecular design.....	9
Molecular modeling tools.....	12
Quantitative structural activity relationship.....	13
Objective of QSAR/CoMFA.....	15
Mathematics of QSAR/CoMFA.....	16
Creating a CoMFA.....	21
OBJECTIVE OF THIS THESIS.....	23
Introduction .....	23
Details of the compounds selected.....	24
Assumptions.....	27
COMPUTATIONAL METHODS.....	28
RESULTS AND DISCUSSION.....	30
Parameters studied .....	30
Part 1: CoMFA with internal validation.....	31
Analysis of Table 1.....	31
Part II: CoMFA with external validation.....	36
Analysis of Table II.....	36
Part III: Understanding how the catalysts work.....	42

CONCLUSIONS.....47  
REFERENCES.....48  
APPENDIX.....50

## TABLE OF FIGURES

Figure 1: Pictorial representation of enantiomeric amino acids.....2 depicting their handedness	2
Figure 2: The enantiomers of limonene; the (R) isomer has a fresh.....3 citrus orange-like odor while the (S) isomer has a harsh turpentine-like limonene odor	3
Figure 3: Structure of Thalidomide. Left and right-handed molecular .....4 forms exist. One form produced sedation and the other form was responsible for fetal abnormalities	4
Figure 4: Example of molecule with chiral carbon atom.....5	5
Figure 5: Example of a chiral molecule with a quadrivalent chiral .....5 sulfur atom	5
Figure 6: Example of a chiral structure with a lone pair of electrons.....5	5
Figure 7: Example of a molecule exhibiting atropisomerism.....6	6
Figure 8: Example of chirality due to helical shape- Hexahelicene .....6	6
Figure 9: Example of chirality due to a bridge structure.....6	6
Figure 10: Example of a molecule exhibiting chirality- Mobius strip.....7	7
Figure 11: The Diels-Alder cycloaddition reaction.....8	8
Figure 12: The real world situation involving taking proprietary ..... 11 information and selecting the correct pathway to produce the desired knowledge or application	11
Figure 13: Overall QSAR method..... 16	16
Figure 14: The regression equation is the best fit line through the data.....19 that minimizes the sum of the deviations	19

Figure 15: Application of chiral catalysts in Diels-Alder reaction.....	24
Figure 16: Structural details of the compounds used in this study.....	26
Figure 17: Template molecule.....	29
Figure 18: Alignment of all catalysts. Hydrogen atoms have been omitted for clarity. Only the lowest energy conformer is considered in the set of catalysts	29
Figure 19: Graph of predicted (ee's) vs actual (ee's).....	41
Figure 20: Electrostatic potential (grid) surrounding the van der Waals surface of catalyst 23	42
Figure 21: CoMFA steric STDEV*COEFF contour plot. Shown inside the field is the aligned set of 23 chiral catalysts with hydrogen atoms removed for clarity. Placement of bulky groups near the green region (contoured at contribution level = 93) and/or removal of steric bulk near the yellow region (contoured at contribution level = 7) should increase ee for those catalysts that are not very stereoselective	44
Figure 22: CoMFA steric STDEV*COEFF contour plot. Shown inside the field is the highly efficient catalyst 3 (ee = 96%). It is to be noted that significant steric bulk lies in the green region while the yellow region is devoid of steric bulk confirming the model	45
Figure 23: CoMFA electrostatic STDEV*COEFF contour plot. Shown is the aligned set of 23 chiral catalyst with hydrogen atom removed for clarity. Placement of groups near blue gives more electrostatic and / or removal near red gives more negative electrostatic	46

## TABLE OF TABLES

Table 1: Influence of Selected Variables on CoMFA Results for 23 Catalysts.....	33
Table 2: Influence of Selected Variables on CoMFA Results for Training.....	36
Set of 19 Catalysts	



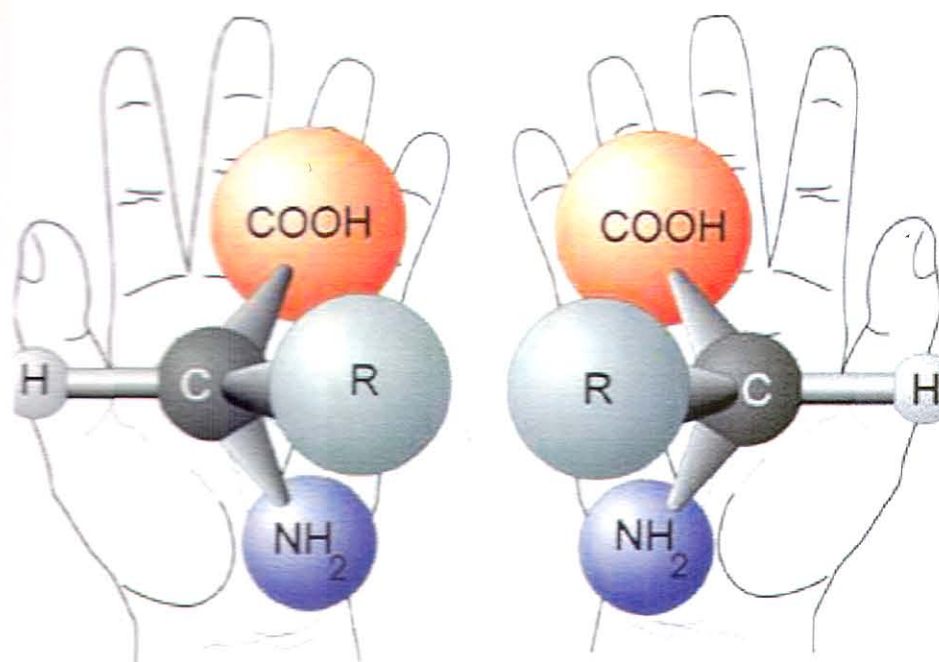
## ABSTRACT

A QSAR using Comparative Molecular Field Analysis (CoMFA) is developed for a set of 23 catalysts containing bis-oxazoline or phosphino-oxazoline ligands that are known to induce asymmetry during the Diels-Alder reaction of N-2-alkenoyl-1,3-oxazolidine-2-one with cyclopentadiene. It is shown that extremely high  $q^2$  statistics can be derived using standard modeling protocols when internal validation alone is done as well as when an external test set is used. From these models it is shown that approximately 70% of the variance in the observed enantiomeric excess can be attributed to the steric field and the remainder of the variance to the electrostatic field. Suggestions about how to improve the performance of inefficient catalysts are given the thesis.

## INTRODUCTION

Nature is asymmetric and molecular asymmetry in particular plays a crucial role in science and technology.<sup>1</sup> A variety of significant biological functions emerge through molecular recognition and this requires strict matching of chirality.<sup>1</sup> The property of the nonsuperimposability of an object on its mirror image is called chirality.<sup>1</sup>

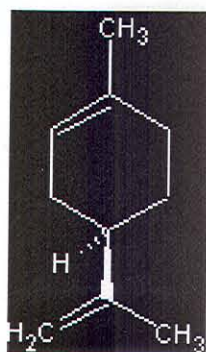
If a molecule is non superimposable on its mirror image, the mirror image is a different molecule. For a pure compound, which is optically active, there are two and only two isomers, called enantiomers, which differ in structure only in the left and right handedness of their orientations (shown in Figure 1). These two forms are non-superimposable mirror images of each other, i.e., they are related like our left and right hands. Hence this property is called chirality, from the Greek word for hand. Contrarily, molecules that are superimposable lack this chirality and are thus achiral. These two enantiomeric forms are also optical isomers because they rotate the plane of polarized light in opposite directions, though in equal amounts. The isomer that rotates the plane to the left is called the levo isomer and is designated (-), while the one that rotates the plane to the right is called the dextro isomer and is designated (+). Because they differ in this property they are often called optical antipodes.



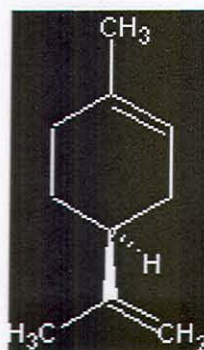
**Figure 1: Pictorial representation of enantiomeric amino acids depicting their handedness**

When chiral compounds occur in living things, only one of the two forms is usually present. Examples of such compounds are amino acids, carbohydrates and nucleic acids. Enzymes, nature's catalysts, are also chiral. They are extremely selective and produce or bind to only one of the enantiomers that fits the active site on an enzyme like a key fits a lock.

Enantiomers have identical properties in a symmetrical environment, but their properties may differ in an unsymmetrical environment. Enantiomers may react at different rates with achiral molecules if an optically active catalyst is present; they may have different solubility in an optically active solvent. For this reason, if a chiral compound interacts with a protein to induce a specific response in a biological organism, it is likely that its enantiomer will either not interact or produce a completely different response. Some of these differences can be quite startling, as for example *limonene* (Figure 2). One enantiomer produces the smell of oranges whereas the other gives rise to the smell of lemons.



(R)- (+)-limonene



(S) – (-) – limonene

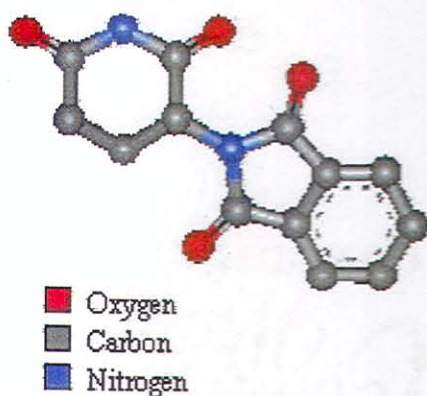
**Figure 2: The enantiomers of limonene; the (R) isomer has a fresh citrus orange-like odor while the (S) isomer has a harsh turpentine -like limonene odor**

Mixtures of equal amounts of enantiomers are optically inactive since the equal and opposite optical rotations cancel. Such mixtures are called racemic mixtures or racemates. Their properties are not always the same as those of the individual enantiomers. The properties in the gaseous or liquid state or in solution usually are the same, since a mixture is nearly ideal, but properties involving the solid state, such as melting points, solubilities and heats of fusion, are often different.<sup>1</sup> Thus, racemic tartaric acid has a melting point of 204-206°C and solubility in water at 20°C of 206-g/ liter, while for the (+) or the (-) enantiomer, the corresponding figures are 170°C and 139-g/liter. The separation of a racemic mixture into its two optically active components is called resolution. Clearly, unequal mixtures of two enantiomers will have a lower optical rotation than a pure enantiomer and the strength of this rotation will depend upon the enantiomeric excess (ee) of the mixtures. Mixtures of unequal amounts of enantiomers are referred to as scalemic.

Understanding chirality is extremely important in the preparation of therapeutic drugs. For example, one enantiomer of *penicillamine* is a potent anti-arthritis agent whereas the other enantiomer is highly toxic.<sup>2</sup> Perhaps the most startling example of the difference in activity between enantiomers is *Thalidomide*.

*Thalidomide* (Figure 3) is a sedative drug that was prescribed to pregnant women, from 1957 into the early 60's.<sup>2</sup> It was present in at least 46 countries under different brand names. When taken

during the first trimester of pregnancy, *Thalidomide* prevented the proper growth of the fetus, resulting in horrific birth defects in thousands of children around the world. The *Thalidomide* molecule is chiral. The drug that was marketed was a racemic mixture. One of the molecules was a sedative, whereas the other one was found later to cause fetal abnormalities. The tragedy is claimed to have been entirely avoidable had the physiological properties of the individual thalidomide [molecules] been tested prior to commercialization.



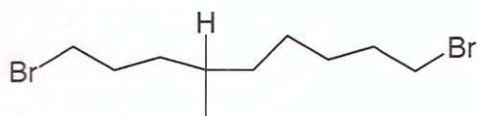
**Figure 3: Structure of Thalidomide. Left and right-handed molecular forms exist. One form produced sedation and the other form was responsible for fetal abnormalities**

Another example is that of *Aspartame* a sweetening agent, it is more than a hundred times sweeter than sucrose.<sup>2</sup> Yet, the mirror image molecule is bitter. Another example is carvone. (S)-carvone possesses the odor perception of caraway while [the mirror image molecule] (R)-carvone has a spearmint odor.

The only requirement for a molecule being chiral is that it not be superimposable on its mirror image. There exist a wide variety of chiral molecules and these molecules can be classified into several categories:

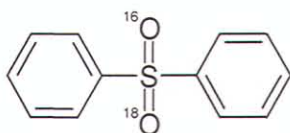
1. *Compounds with a chiral carbon atom:* If there is one such atom, the molecule must be chiral.

An example of such a system is depicted in Figure 4.



**Figure 4: Example of molecule with chiral carbon atom**

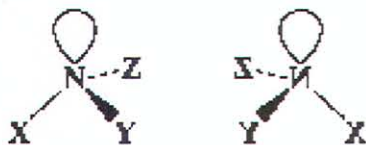
2. *Compounds with other quadrivalent chiral atoms:* Any molecule containing an atom that has four bonds pointing to the corners of the tetrahedron is chiral if the four groups are different. An example of this is shown in Figure 5, where it is noted that even isotopic substitutions result in chirality.



**Figure 5: Example of a chiral molecule with a quadrivalent chiral sulfur atom**

3. *Compounds with trivalent chiral atoms:* Atoms with pyramidal bonding give rise to optical activity if the atom is connected to three different groups since the unshared pair of electrons is analogous to a fourth group, which by necessity is different from others. An example of this is shown in Figure 6.

#### Pyramidal Inversion



**Figure 6: Example of a chiral structure with a lone pair of electrons**

4. *Restricted rotation giving rise to perpendicular dissymmetric planes:* Restricted rotation around single bonds can also give rise to chirality. An example of this, called atropisomerism, is depicted in Figure 7.

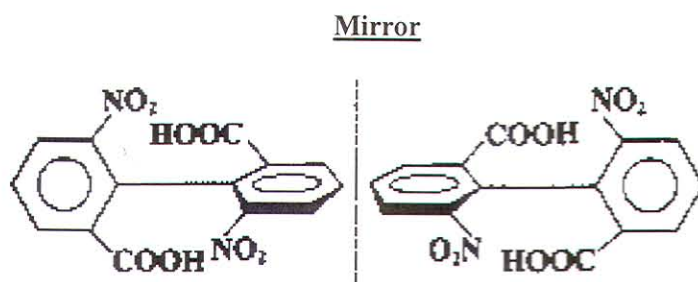


Figure 7: Example of a molecule exhibiting atropisomerism

5. *Chirality due to helical shapes:* Several compounds have been prepared that are chiral because they have a shape that is actually helical and can therefore be left or right handed in orientation as presented in Figure 8.

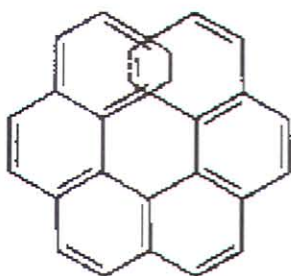


Figure 8: Example of chirality due to helical shape- Hexahelicene

6. *Chirality caused by restricted rotation of other types:* Substituted paracyclophanes and related systems may be optically active. In this case chirality results because the benzene ring cannot rotate in such a way that the carboxyl group goes through the alicyclic ring (Figure 9). Many chiral-layered cyclophanes have been prepared.

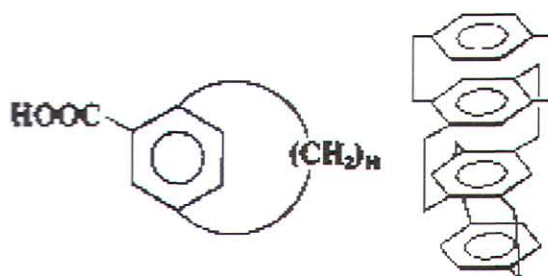
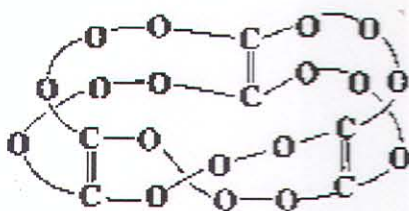


Figure 9: Example of chirality due to a bridge structure

The main molecular chain in the molecule shown below in Figure 10 has the form of a Mobius strip. This molecule has no chiral carbons, nor does it have a rigid shape, but it too has neither a plane nor an alternating axis of symmetry thus making it chiral.



**Figure 10: Example of a molecule exhibiting chirality- Mobius strip**

Chiral compounds have a widespread use in medicine and agriculture. Enantioselective synthesis of chiral organic compounds is an important task for synthetic chemists and the design of catalytic, asymmetric reactions that proceed with high enantioselectivity is an important goal in chemical synthesis. If a single enantiomer can be made more selectively, this could lead to a greater demand for that molecule on the chiral drug world market.

Since the early 1970s, a large number of research groups have become interested in discovering new and practical techniques for the control of absolute stereochemistry and there has been remarkable progress in the field of catalytic asymmetric synthesis employing chiral Lewis acids. One reaction that is especially important in synthetic organic chemistry is the Diels-Alder reaction. In a recent review, the Noble laureat E.J. Corey said, "If one chemical reaction had to be selected from all those in the repertoire of synthetic organic construction, it was clear by 1970 that the Diels-Alder reaction would be the logical choice." Because of its significance and because it forms the basis for the research described in this thesis, we present a modest background discussion of this reaction below.

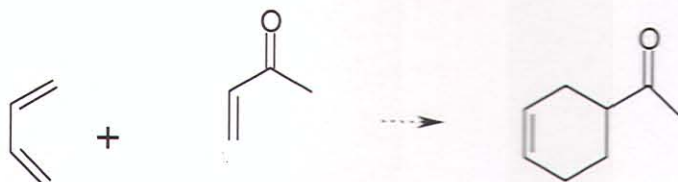
## **The Diels-Alder Cycloaddition Reaction**

This reaction is named after its discoverers.<sup>3</sup> It is extremely useful in organic synthesis because it forms two carbon-carbon bonds in a single step and is one of the few methods available for making



cyclic molecules. An example of this reaction is shown in Figure 11. The reaction occurs in a single step through a cyclic transition state in which the two new carbon-carbon bonds form simultaneously.

The 1950 Nobel Prize in chemistry was awarded to Diels and Alder in recognition of the importance of their discovery.



**Figure 11: The Diels-Alder cycloaddition reaction**

The mechanism of the Diels-Alder cycloaddition is different from that of other reactions because it is neither polar nor radical. Rather, the Diels-Alder reaction is a pericyclic process.<sup>3</sup> The two reactants simply join together through a cyclic transition state in which two new carbon-carbon bonds form at the same time.

The Diels-Alder cycloaddition reaction occurs most rapidly and in highest yield if the alkene component, or dienophile, has an electron-withdrawing substituent group. One of the most useful features of the Diels-Alder reaction is that it is stereospecific. The stereochemistry of the starting dienophile is maintained during the reaction and a single product stereoisomer results. In the Diels-Alder reaction the diene and dienophile partners line up so that the endo product, rather than the alternative exo product, is formed. Endo products result from Diels-Alder reactions because the amount of orbital overlap between diene and dienophile is higher when the reactants lie directly on top of one another so that the electron-withdrawing substituent on the dienophile is underneath the diene.

The asymmetric Diels-Alder reaction has been widely used as a useful method for the enantio and diastereoselective construction of six membered ring compounds.<sup>4</sup> A number of asymmetric Diels-Alder reactions have been developed so far mostly with the use of Lewis acidic catalysts such as aluminum, silicon, titanium, boron, zinc, magnesium derivatives and so on. Currently much attention is being devoted to more efficient catalytic asymmetric Diels-Alder reactions with transition metal

catalysts such as chromium, iron, cobalt, copper, ruthenium and rhodium and lanthanide catalysts have appeared.

Chiral metal catalysts are important tools for the synthesis of enantio pure organic compounds. The asymmetric Diels-Alder reaction is one of the most powerful and versatile reactions in organic synthesis. Catalytic enantioselective processes with chiral Lewis acid derived catalysts significantly extend the scope and utility of this reaction. In the design of a chiral Lewis acid the choice of the chiral ligand and the metal and its counterion is critical to superior endo/exo selectivity as well as endo enantioselectivity.

## **Approaches in Computer Aided Molecular Design**

Mathematical theories and the corresponding computational evaluations of chemical phenomena are not new. As early as 1929, P.A.M Dirac wrote that in view of the contemporary development of quantum mechanics- "*the underlying physical laws necessary for the mathematical theory of physics and the whole chemistry are thus completely known.*" For many years, chemists have developed simplifications, approximations and numerical methods that make it possible to evaluate mathematical theories for realistic molecules. These developments continue to the present day. Almost 50 years have passed since the first computation of the steric energy of a molecule in terms of nonbonded interactions was done. It is over 30 years ago that the first use of an electronic computer to calculate and optimize the molecular energy was made by Hendrickson.

Molecular modeling techniques are widely used in the chemical, pharmaceutical and agrochemical industries. Molecular modeling studies involve three stages. In the first stage a model is selected to describe the intra- and inter-molecular interactions in the system. The two most common models that are used in the molecular modeling are quantum mechanics and molecular mechanics. These models enable the energy of any arrangement of the atoms and molecules in the system to be

calculated. The second stage of a molecular modeling study is the calculation itself, such as an energy minimization, molecular dynamics or Monte Carlo simulation or a conformational search. Finally, the calculation must be analyzed to calculate the properties. The above-developed model is then used to predict the property of a new molecule. The flow sheet given in Figure12 shows the various stages involved in development of a new molecule.<sup>5</sup>

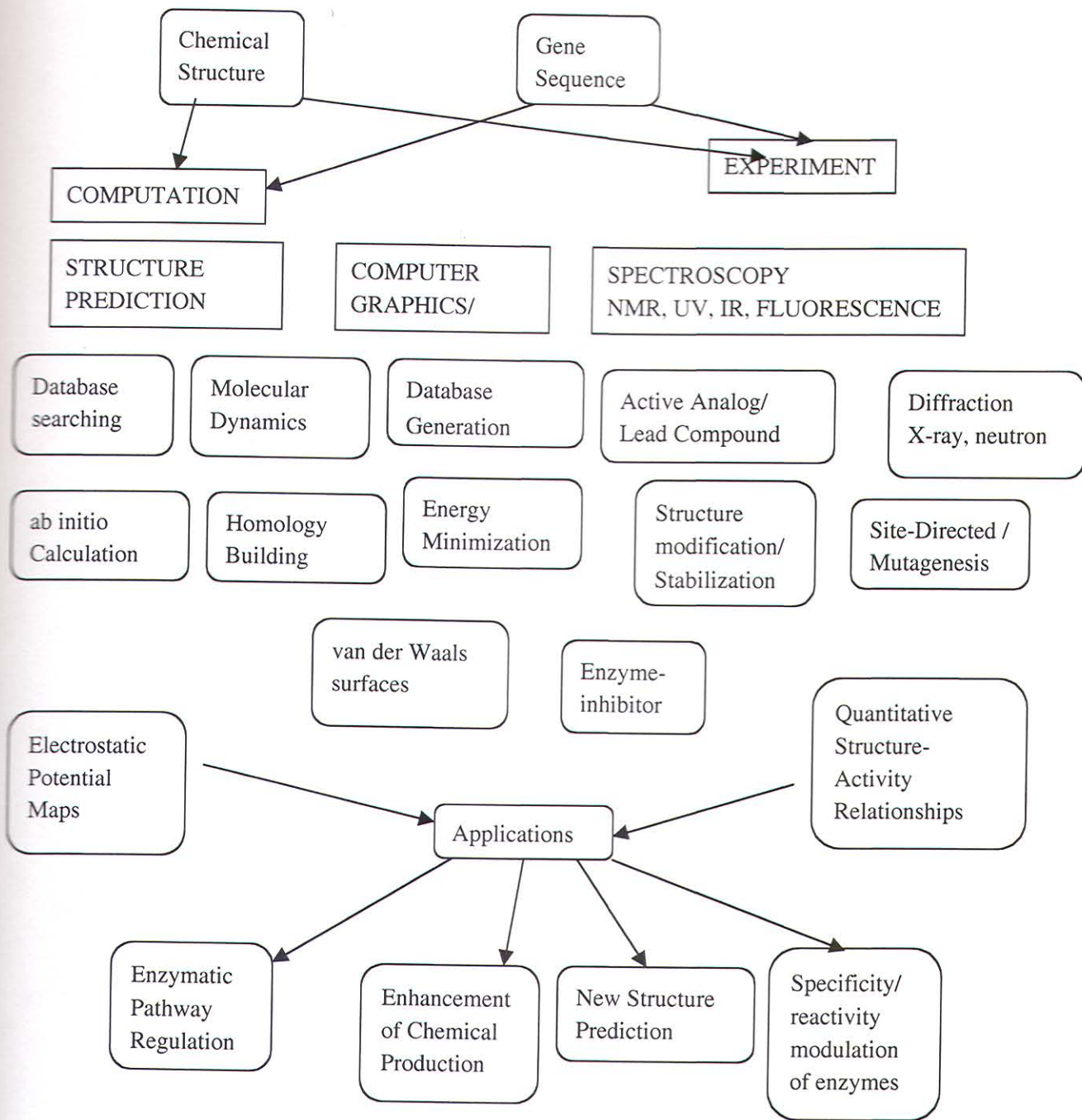


Figure 12: The real world situation involving taking proprietary information and selecting the correct pathway to produce the desired knowledge or application <sup>5</sup>

## Molecular Modeling Tools

There are many approaches in computational chemistry that are popular in molecular modeling. Many tools are available to determine the molecule's Potential Energy Surfaces (PES). From a Potential Energy Surfaces one can interpret directly the molecule's shape and reactivity. Popular molecular modeling tools used to determine a PES include:

*Molecular Mechanics (MM), Molecular Dynamics (MD) and Quantum Methods.* These methods allow one to calculate energies and structures of molecules. Solvent effects can be included explicitly or via empirical models.

*Molecular Mechanics (MM):* is a nonquantum mechanical (QM) way of computing the structures, their energies and some properties of the given molecule. MM relies on an empirical force field (FF), which is a numerical method for representing the molecule's PES. Because MM treats electrons in an implicit way, it is much faster than QM that treats electrons explicitly. A limitation of MM is that bond making and bond breaking processes cannot be modeled directly as with QM.

*Molecular Dynamics (MD):* Energy minimized structures are motionless and accordingly, are incomplete models of reality. In molecular dynamics, atomic motion is described with Newtonian mechanics. Dynamical properties of molecules can thus be modeled. Because simulation periods are typically in the nanosecond range, only very fast processes can be explored

*Quantum methods:* The objective is to describe the spatial positions of electrons and nuclei. These methods are substantially more computationally demanding than those that rely upon potential functions, for molecules of the same size. Quantum chemical methods can be roughly divided into:

*Semiempirical methods* - approximate methods in which some small quantum effects are neglected and some parameters are estimated by fitting to experimental data. They may only be used for chemical species for which they were parameterized. For distorted, uncommon bonding situations produce unreliable results.

*Nonempirical methods* - do not require empirical parameters and can be used for any molecular system. Two common non-empirical methods include traditional *ab initio* methods based on

Hartree-Fock theory and Density Functional methods where electron density is the primary way of describing the energy of the system.

## Quantitative Structural Activity Relationship

Methods have been developed in the last decades to describe Quantitative Structure Activity Relationships (QSAR) and Quantitative Structure Property Relationships (QSPR) which deal with the modeling between structural, chemical, and biological properties.<sup>6</sup> Quantitative structure-activity relationships (QSAR) relate numerical properties of the molecular structure (using physicochemical descriptors) to its activity by a mathematical model. These physicochemical descriptors, which include parameters to account for hydrophobicity, topology, electronic properties, and steric effects, are determined empirically or, more recently, by computational methods. Activities used in QSAR include chemical measurements and biological assays. QSAR is currently being applied in many disciplines, with many pertaining to drug design and environmental risk assessment.<sup>6, 7, 8</sup>

One important aim of drug design is to correlate the three-dimension (3D) structure of the drug molecules with their biological activities, i.e., to derive a 3D-QSAR. The goal is to be able to design and predict the biological efficacy of new molecules prior to synthesis.

A structure-activity study can help to decide which features of a molecule give rise to its overall activity and help to make modified compounds with enhanced properties. The relationship between these numerical properties and the activity is often described by an equation of the general form:

$$v = f(p) \quad [1]$$

where

$v$  is the activity in question

$p$  is the structure-derived properties of the molecule

$f$  is the function

The history of QSAR dates back to the 19<sup>th</sup> century. In 1893, Charles Richter observed that the toxicity of alcohols decreases as the water solubility increases.<sup>9</sup> Around 1900, Hans Meyer and Charles

Ernest Overton noted that the toxicity of organic compounds depend on their lipophilicity.<sup>10,11</sup> In the 1930's, Louis Hammett established a relationship between electronic properties of substituents of organic acids and bases with their reactivity and their equilibrium constants. 1964, is considered to be the birth year of modern QSAR, when two important papers were published. Hansch and Fujita published a paper entitled “ $\rho - \sigma - \pi$  Analysis.<sup>10,12</sup> A Method for the Correlation of Biological Activity and Chemical Structure”. Free and Wilson published a paper with the title: “A Mathematical Contribution to Structure Activity Studies”. Both contributions started the development of two new methods of Quantitative Structure-Activity Relationships, later called Hansch analysis and Free-Wilson analysis, respectively. For steric effects, initially bulk parameters were chosen, such as the molar refractivity. Hansch developed an equation for correlation of biological activity to a molecules electronic characteristics and hydrophobicity.<sup>8</sup> An example of such a relationship is given in equation [2].

$$\log 1/C = -k_1 (\log P)^2 + k_2 \log P + k_3 \sigma + k_4 \quad [2]$$

In this equation, C is the molar concentration that produces a certain biological effect,  $\log P$  describes the distribution of a molecule partitioning between n-octanol and water,  $\sigma$  is the Hammett constant and  $k_1, k_2, k_3$  are coefficients determined from a least square regression. Steric parameters were later added to this general model.

The first Free-Wilson model developed was confusing and was complicated to calculate. Fujita and Ban described an improved version of the Free-Wilson model.<sup>9</sup> The Free-Wilson approach assumes that an unsubstituted compound contribute with a value  $\mu$  to the biological activity. Each additional substituent has, depending on the kind and position of the substituent  $x_i$ , an additional contribution  $a_i$  in equation [3]:

$$\log 1 / C = \sum a_i x_i + \mu \quad [3]$$

This approach is easy to apply, though it also has its own shortcomings. The derivation of a QSAR equation involves a number of distinct stages. First, it is obviously necessary to synthesize the compounds and determine their biological activities. QSAR's can be derived for very diverse sets of compounds, but it is more common to consider a related series of compounds that differ in just one part of the molecule. When planning which compounds to synthesize, it is important to cover the range of properties that may affect the activity. The properties chosen for inclusion in the QSAR equation should ideally be uncorrelated with each other.

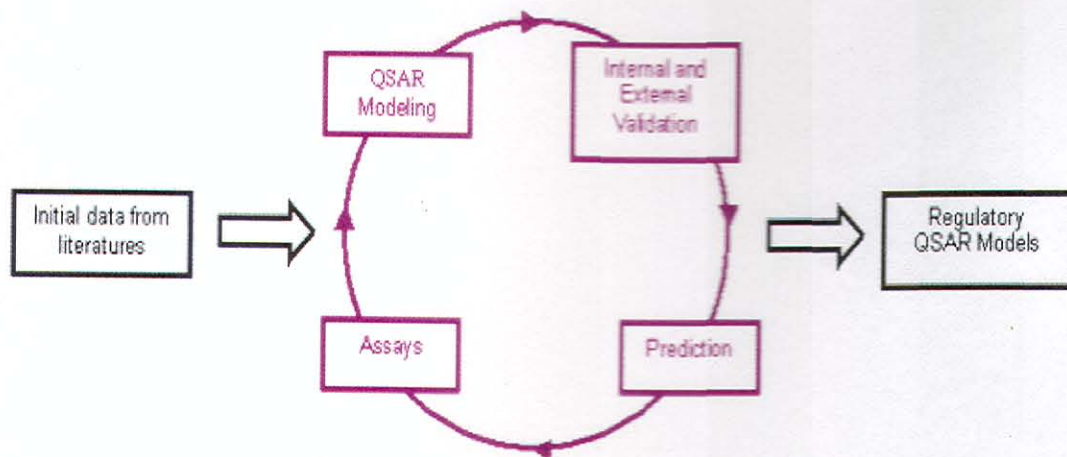
Since its introduction several years ago CoMFA (Comparative Molecular Field Analysis), first described by Cramer and co-workers, has become one of the most powerful tools for QSAR and drug design.<sup>8,13</sup> In fact, CoMFA has pioneered a new paradigm of three-dimensional QSAR where the shapes, properties, etc. of molecules are related to specific molecular features (substituents, etc.) and their spatial relationship. Thus molecular modification to improve biological performance based on QSAR can be more rooted in the actual chemistry of the involved molecules.

### **Objective of QSAR/CoMFA**

QSAR modeling employs statistical approaches to correlate and rationalize variations in the biological activity of a series of chemicals with variations in their molecular structures. The molecular structure is often represented by a set of quantities commonly known as molecular descriptors. QSAR's have been applied extensively in a wide range of scientific disciplines including chemistry, biology and toxicology. In both drug discovery and environmental toxicology, QSAR models are now regarded as a scientifically credible tool for predicting and classifying the biological activities of untested chemicals. As we enter the informatics era, QSAR has become essential in the drug discovery process as a screening and enrichment tool to eliminate, from further development, those chemicals lacking drug-like properties or those chemicals predicted to elicit a toxic response. This developing scenario spreads QSAR, beyond the pharmaceutical industry, to human and environmental regulatory authorities for use in toxicology. QSAR with CoMFA is fully integrated with SYBYL to enable visualization and analysis



of a structure-activity relationship. The idea underlying a **Comparative Molecular Field Analysis** (CoMFA) is that differences in a target property are often related to differences in the shapes of the non-covalent fields surrounding the tested molecules. Figure 13, illustrates the overall method of QSAR.



**Figure 13: Overall QSAR method**

### **Mathematics of QSAR/CoMFA**

CoMFA describes 3D structure-activity relationships in a quantitative manner. For this purpose, a set of molecules is selected. All the molecules in this set are assumed to interact with the same kind of receptor (or enzyme, ion channel) in the same manner. A certain subgroup of molecules is next selected to constitute the training set to derive the CoMFA model. Next, descriptors representing molecular structure of individual chemicals (i.e., hydrophobicity, structural fragments, charged surface area, the number of hydrogen bonds, solubility, etc.) are calculated. Then, a correlation between descriptors and activity for the training set is evaluated by employing various statistical approaches to determine the most statistically significant relationship (the QSAR model). A proper validation is required to ensure the model's predictive value for the chemicals not used in the training set. With adequately validated performance, such models can be used to predict activities of untested molecules.

While there are many possible adjustable parameters in CoMFA, certainly the most important is the relative alignment of the individual molecules when their fields are computed. Properly aligned molecules have a comparable conformation and a similar orientation in Cartesian space. Hence, alignment is the most crucial step of CoMFA analysis. All molecules in the data set must be overlapped using parts of the molecules (the template), which are assumed to have the same contribution to the biological effect. This is done by placing molecules, oriented the same way, at the center of a three dimensional grid. The CoMFA is done by evaluating at each grid point the interaction energy between a probe atom with each aligned molecule. Different atomic probes, e.g., a carbon atom, a positively or negatively charged atom, a hydrogen bond donor or acceptor or a lipophilic probe, are used to calculate field values in each grid point, i.e., the energy values which the probe would experience in the corresponding position of the regular 3D lattice. These fields correspond to tables, most often including several thousands of columns, which must be correlated with the binding affinities or with other biological activity value.

To put the shape of a molecular field into a QSAR table, the magnitude of its steric (Lennard-Jones) and electrostatic fields (Coulombic) are sampled at regular intervals throughout the region.

#### *Standard CoMFA Fields*

The standard potential energy fields produced by the out-of-the-box CoMFA program are steric (van der Waals) and electrostatic (Coulombic). The standard CoMFA probe is an  $sp^3$  hybridized carbon atom with an effective radius of 1.53 Å and a +1.0 charge. The probe atom to ligand atom distance-dependence of the potential functions (i.e., the standard 6-12 of the Lennard-Jones potential and r-square term of the Coulombic potential) result in steep changes as the probe nears the surface of the molecule. It is the convention to truncate the steric value at some arbitrary level (on the order of 4.0 to 30.0 kcal/mol) to eliminate points both within the van der Waals shells of molecules and at the periphery of the region so that effectively a shell of points is used. Electrostatic values are also truncated at similar levels and most commonly ignored at points inside the molecules. In 3D-QSAR

studies, the steric and electronic effects are approximated using molecular shape (Lennard-Jones) and charge distribution (Coulombic) potential energy fields.

In a typical QSAR or CoMFA analysis one must not only derive the mathematical model, one must then validate it. That is to say, providing only a statistical measure of agreement between experiment and theory, like a correlation coefficient from a plot of the computed activity versus experimental activities for the molecules in the data set, is not a sufficient validation. Accordingly a variety of validation schemes have been developed, the most common of which is the leave one out (LOO) cross validation method. Here one omits a single compound from the data set and a PLS model is developed using the remaining compounds. The model is used to predict the activity of the omitted molecule that was not included in the model. This procedure is repeated until all molecules in the data set have been eliminated once. The most salient statistic from such analysis is the cross-validated r squared ( $r^2_{cv}$ ) commonly published as  $q^2$ . A cross-validated r squared statistic,  $q^2$  is always smaller than a simple  $r^2$  from plots of predicted activity versus experimental activity.

#### *Deriving the QSAR Equation*<sup>8,10</sup>

The most widely used technique for deriving a QSAR equation is linear regression:

$$y = mx + c \quad [4]$$

where

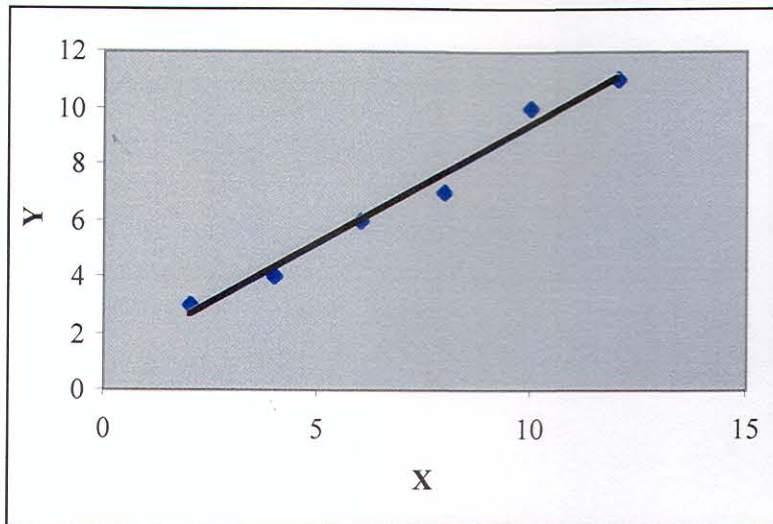
y is the dependent variable (the observations)

x is the independent variable.

The least-squares coefficients m and c in the linear regression equation are given by equation [5]:

$$m = \left\{ \sum_{(i=1..N)} [ (x_i - \bar{x}) (y_i - \bar{y}) ] \right\} / \left\{ \sum_{(i=1..N)} (x_i - \bar{x})^2 \right\} \quad [5a]$$

$$c = (\bar{y}) - m (\bar{x}) \quad [5b]$$



**Figure 14: The regression equation is the best fit line through the data that minimizes the sum of the deviations**

The term  $r^2$  is reported as the squared correlation coefficient, calculated as in equation [7]:

$$\text{TSS} = \sum_{(i=1..N)} (y_i - \bar{y})^2 \quad [6a]$$

$$\text{ESS} = \sum_{(i=1..N)} (y_{\text{calc},i} - \bar{y})^2 \quad [6b]$$

$$\text{RSS} = \sum_{(i=1..N)} (y_i - y_{\text{calc},i})^2 \quad [6c]$$

$$r^2 = \text{ESS} / \text{TSS} = (\text{TSS} - \text{RSS}) / (\text{TSS}) = (1 - (\text{RSS} / \text{TSS})) \quad [7]$$

where,

TSS = the total sum of squares

ESS = explained sum of squares

RSS = residual sum of squares

The value of  $r^2$  can be between 0 and 1. A value of 0.0 indicates that none of the variation in the observations is explained by variation in the independent variables, where as value of 1.0 indicates that all of the variation in the observation can be explained. This analysis can be extended to more than one variable and is known as multiple regression analysis.

### *Principal Component Regression and Partial Least Squares*<sup>8,14</sup>

Multiple linear regressions cannot deal with data sets where the variables are highly correlated and / or where the number of variables exceeds the number of data values. Two methods are used to deal with this: principal component regression(PCR) and partial least squares(PLS).

In principal component regression, the variables are subjected to a principal component analysis and then regression analysis performed using the first few principal components. PLS is a predictive analysis approach that derives coefficients for all the steric and electrostatic parameters (descriptors) from the CoMFA calculations:

$$\text{prediction} = \text{intercept} + (\text{coeff1} * \text{component1}) + (\text{coeff2} * \text{component2}) + \dots \dots \dots [8]$$

As described above CoMFA calculations produce a huge number of data points (x variables, independent variables). CoMFA generates a greater number of descriptors than the number of molecules. PLS combines descriptors to so-called components. Components are the linear combination of all descriptors, which are used to fit the shape of the parameters to the shape of the dependent variable. PLS is a predictive statistical method. In PLS, the components are constructed to explain not only the variation, but also to maximize the degree to which the variation in the observations can be explained. A PLS model is often evaluated according to its ability to predict the activity of the compounds not used to derive the model. For this cross-validation is used, where the predictive ability of the model is determined by dividing the data into a number of groups. This method is described below:

#### *Cross-validation*<sup>8,14</sup>

Cross-validation is a widely used and strongly recommended technique for checking the quality of a regression model. The most common form of cross validation is leave – one – out (LOO). Here each data value is left out in turn and a model derived using the remainder of the data. A value can then be predicted for the data left out and compared with the true observed value. This is repeated for every data point in the set and permits the calculation of a “cross-validated  $r^2$ ” value (also written  $r_{cv}^2$  or  $q^2$ ).

$r^2$  is the measure of goodness of fit,  $q^2$  is a measure of goodness of prediction. The cross-validated  $r^2$  is calculated as in equation [9]:

$$q^2 = 1 - \text{PRESS} / \left( \sum_{(i=1..N)} (y - y_i)^2 \right) \quad [9a]$$

$$\text{PRESS} = \left( \sum_{(i=1..N)} (y_{\text{pred}, i} - y)^2 \right) \quad [9b]$$

where

$y_{\text{pred}}$  is a predicted value

$y$  is an actual or experimental value

$y_i$  is the best estimate of the mean of all values that might be predicted.

The value of  $q^2$  can take a value between  $-\infty$  to 1. A value of 1 means an absolutely perfect predictive model. Values below zero imply that having no model is better than using the derived model. A good predictive model is said to exist when  $q^2 > 0.5$  are achieved. Furthermore, a cross validated calculation will specify the best prediction for the data set depending upon how many principal components have been used.

## Creating a CoMFA

The development of a CoMFA model depends upon the following factors - the selection of the active compounds, the different types of probe atoms that can be employed, the force-field models used to describe the interactions between the probe and each compound, the size and the spacing of points in the grid and indeed the way in which the PLS analysis is performed. One of the main requirements of the CoMFA technique is that it requires the structures of the molecules to be correctly overlaid. The vast number of grid field variables in a typical CoMFA analysis are closely coupled; even the smallest structural change in the compounds will cause changes in not just one variable but in a group of variables that are connected in space.

For better understanding of where the electrostatic and steric fields are positioned in space around the molecules, contour maps are used to display the 3D fields. These maps are 3 dimensional

iso-contour surfaces and are displayed with various colors. The contours of the steric map are shown traditionally in yellow and green and those of electrostatic are shown in red and blue. Greater value of the activity is correlated with more bulk near green, less bulk near yellow, more positive charge near blue and more negative charge near red.

Traditionally in a QSAR study one might evaluate the binding affinities of a series of drugs to a common receptor. The assumption being made is that all drugs bind to the same receptor and evoke biological responses via the same mechanism. In those studies the nature of the receptor is unknown. In our study we do the reverse, we know the shape of the catalysts (i.e., the receptor) and we want to correlate the steric and the electrostatic fields for those chiral catalysts with the experimental enantiomeric excesses (ee) for the reagents (the substrates) in a common reaction. We have built a QSAR using CoMFA for a set of 23 catalysts containing bis-oxazoline or phosphino-oxazoline ligands that are known to induce asymmetry during the Diels-Alder reaction of N-2-alkenoyl-1,3-oxazoline-2-one with cyclopentadiene.

# OBJECTIVE OF THIS THESIS

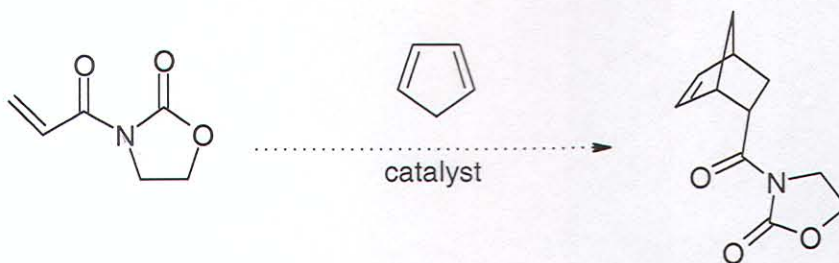
## Introduction

In a traditional QSAR study one might evaluate, say, the binding affinities of a series of drugs to a common receptor. The assumption being made is that all drugs bind to the same receptor and evoke biological responses via the same mechanism. In those studies the nature of the receptor is unknown. In our study, we know the shape of the catalysts (the receptor in our case). We want to correlate the steric and the electrostatic fields of those chiral catalysts with the experimental enantiomeric excesses (ee) for a common reaction.

Our objectives are two folds. First we want to see if commercially available software can be used to develop a CoMFA model for catalytic systems. To our knowledge this is the first study of chiral catalysts with CoMFA. This is important because the series of compounds we will study are comparable in number to typical, small library of compounds derived from combinatorial chemistry. Thus it is desirable for us to show whether or not CoMFA on small to medium sized libraries of chiral catalysts can be modeled quickly and efficiently. The second objective is to generate a high quality CoMFA. From this we will better understand how chiral catalysts work. Moreover, we will then be in a position to predict what kind of ligands and substituents will make the catalysts even more efficient as well as being able to tell the bench chemicals what to avoid.

The reaction we evaluate is depicted in Figure 15. This reaction has become a de facto benchmark for synthetic chemists who want to demonstrate the utility of their chiral catalysts. We select the Diels-Alder reaction because the transition state for this concerted pericyclic reaction is not expected to change much from catalyst to catalyst.





**Figure 15: Application of chiral catalysts in Diels-Alder reaction**

The set of catalysts selected for our study is shown in Figure 16. All compounds selected contain at least one oxazole ring. These compounds are further divided into two groups- first containing a bis-oxazoline and the second with a phosphino-oxazoline substructure. The shapes, size and electrical properties of these molecules cover a suitable range of values for our analysis. The data for ee spans the range from 10 to 99, and is also given in Figure 16 in parentheses.

### Details of the compounds selected

For compounds (1) and (2) Ghosh and coworkers reported highly enantioselective cycloadditions between cyclopentadiene and various bidentate dienophiles in the presence of copper (II) –bis- (oxazoline) catalyst derived from *cis*-1-amino-2-indanol.<sup>4, 15, 16</sup>

For compounds (3) to (6) the Merck group reported the influence of ligand bite angle on spirobis-(oxazolines) in the enantioselectivity of copper(II)-catalyzed Diels-Alder reactions.<sup>4, 17</sup>

For Compounds (7), (8), (9), (10) and (12), as reported by Davies, the HIV protease inhibitor Crixivan is approved for the treatment of AIDS<sup>17</sup>. In the synthesis of Crixivan, an expedient way of producing large quantities of 1*S*,2*R*-*cis* aminoindanol was established. As a result of these studies commercial quantities of aminoindanol have become available. The oxazolidinones derived from aminoindanol are very efficient Evans-type auxiliaries for Diels-Alder reactions.<sup>18</sup> The ligands were prepared in 56-72% yield by alkylation of the parent ligand.<sup>19</sup>

For compound (13), Evans reported that bis(oxazoline) copper(II) complexes are highly enantioselective catalysts in Diels-Alder reactions involving bidentate dienophiles.<sup>19</sup> A comparative study was undertaken to elucidate the differences between the bis(oxazoline)- Cu(II) catalyst and the bis(oxazoline) catalysts derived from Fe(III), Mg(II) and Zn(II). Catalyst performance was found to be a function of a subtle relationship between bis(oxazoline) structure and transition metal.

For compounds (14) to (20), Helmchen reported that a general problem associated with complexes of Mg(II) and Fe(II) is that, they were fairly weak Lewis acid and therefore, had low catalytic activity.<sup>20</sup> Helmchen reported the first application of chiral (phosphino-oxazoline) Cu (II) complexes as catalysts in the Diels-Alder reaction of substituted N-acylamide dienophiles.<sup>20</sup>

For compounds (21) to (23), Evans and co-workers in 1993 reported the utility of chiral Cu(II)-bis-(oxazoline) complexes as Lewis acids in the catalysis of the Diels-Alder reactions of unsubstituted and  $\beta$ -substituted acrylimides with cyclopentadiene.<sup>4, 21, 22</sup>

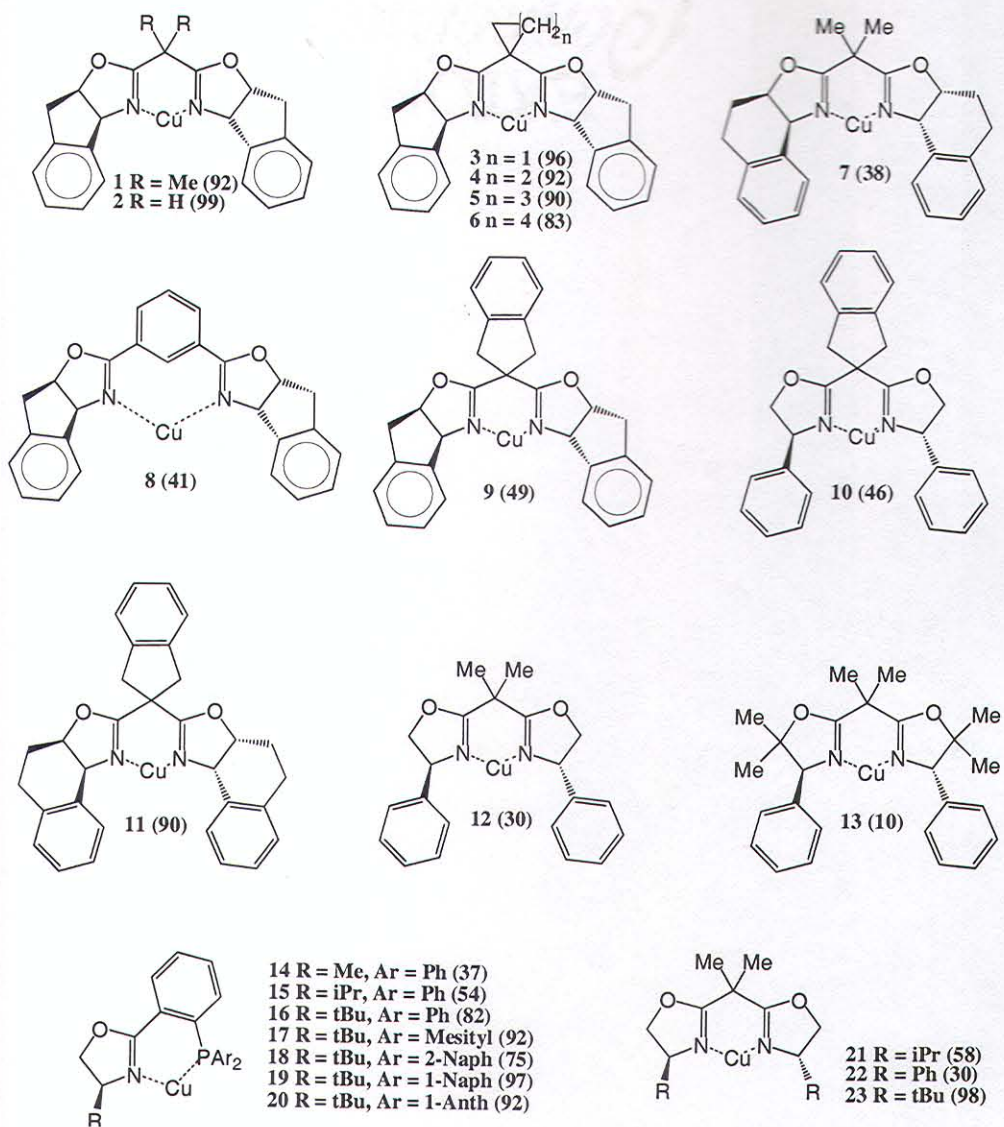


Figure 16: Structural details of the compounds used in this study

## Assumptions

The assumption made here is that a concerted pericyclic reaction of more or less similar synchronicity prevails around the catalyst's metal center for all of the systems studied in this thesis. The above systems that were selected also fulfilled the following requirements: -

- 1) The publication from which information was extracted should have a complete assessment of reaction conditions such that the reported ee is deemed reliable; the efficacy of many catalyzed reactions is often found to be dependent upon more than just the ligand used. In particular the metal (including its spin and oxidation state), the solvent, the temperature and the nature of the counter-ions associated with the catalyst before substrate binding all impact the observed stereochemistry. The systems selected were well studied in this regard by the author who created each catalyst. The ee's used in our analyses are thus considered to be the highest possible for a given ligand.
- 2) The reaction taking place should be as simple as possible. In this study it is presumed that the catalyzed cycloadditions are all single step, concerted, pericyclic reactions with no intermediates along the reaction pathway.
- 3) All reactions being catalyzed by the different catalysts have similar mechanisms.

## COMPUTATIONAL METHODS

All computations were done using commercially available software. Crystal structures when available were retrieved from the Cambridge Structural Database (CSD) and the CoMFA was done using SYBYL 6.8. Details of the molecular alignments, probes, grid spacing and the statistical analyses shall be given later in the thesis.

a) *Generation of Catalyst Structures*: Initial atomic coordinates were generated with the builder facility of Spartan or they were imported from the CSD. In all the cases the corresponding anions (usually triflates or antimony hexafluorides) were included in the calculations. The geometry optimization was stopped when the Spartan's default convergence criterion was met.

Geometry optimization of all initial conformations was done using no constraints at all. In some instances more than one conformation of catalyst is possible. Following a systematic (grid) conformer search, the lowest energy conformer was used for the CoMFA. The quantum mechanical calculations were done using the PM3tm Hamiltonian implemented in Spartan 5.1.3. It's known that PM3 is unreliable for partial charges on nitrogen. Thus, natural atomic charges as well as Mulliken charges were calculated. Finally, Gasteiger charges were used for the CoMFA calculations.

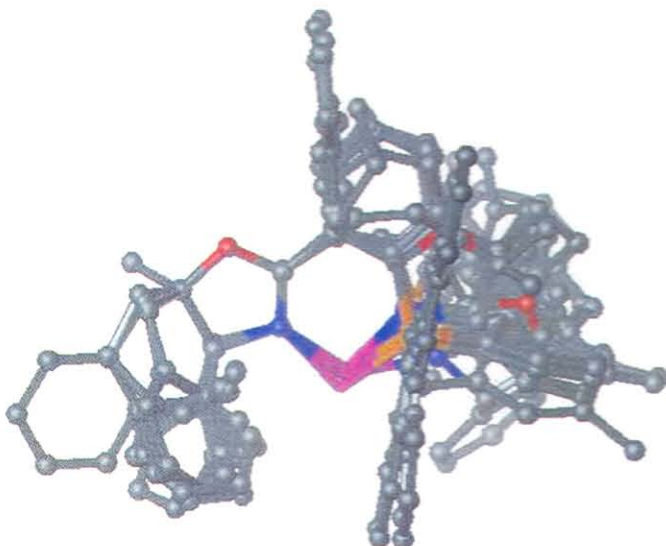
b) *Catalyst Alignment*: The lowest energy structure of each catalyst was imported in SYBYL (minus the counterion). There exist several possible alignment schemes. Alignments were done by least squares fitting the structures to a common template. The template is shown in Figure 17. It is a four atom template in the oxazoline ring which is common to all catalysts. This alignment was then used for the CoMFA. Figure 18 shows all 23 molecules that have been superimposed this way.



**Figure 17: Template molecule**

c) *Selection of Variables*: Once the molecules in the data set are aligned they are centered in a three-dimensional grid with uniformly placed grid points. At each grid point a test probe is selected and then intermolecular energy between the probe and each molecule is calculated. These interaction energies are the descriptors used by the SYBYL for the PLS regression. Many options are available for selection of grid dimensions, atomic charge assignments, treatment of the dielectric between probe and molecule, the probe to be used, etc.

Finally, the CoMFA was developed. A detailed analysis of the CoMFA developed is given in further chapters of this thesis.



**Figure 18: Alignment of all catalysts. Hydrogen atoms have been omitted for clarity; only the lowest energy conformer is considered in the set of catalysts**

## RESULTS AND DISCUSSION

### Parameters Studied

As already discussed in "OBJECTIVE" we developed and analyzed two sets of CoMFA:

- a) CoMFA with internal validation
- b) CoMFA with external validation

Using the aligned data set depicted in Figure 18 a standard CoMFA (INTERNAL VALIDATION) was performed. Several kinds of partial atomic charges were used initially for this evaluation including Mulliken, Gasteiger and Gasteiger -Marsilli. But the best results were obtained using the Gasteiger charges. Hence, the further analyzes was carried out using the Gasteiger charges only. The various parameters studied are as follows:

- 1) *Interaction Fields*: - The type of field used for the analyses i.e., steric or electrostatic. There are other kinds of fields that could be included, e.g., hydrogen bonding, hydrophobic, etc. The steric and electrostatics fields represents the best interactions that take place between this set of chiral catalysts and the reagents undergoing the Diels-Alder reaction so that they are used exclusively. When both fields are included in the analysis there are twice as many as descriptors as when electrical fields or steric fields are used alone. We studied the effect of fields, together and also individually.
- 2) *Magnitude of the energy cut off*: - At each grid point the interaction between a probe atom and each of the atoms in each of the aligned molecules is computed with a potential energy function. In some instances the grid points lie close to one or more atoms and consequently the computed energies at those points are unacceptably large. Thence, we invoke a cutoff that says, for example, remove all descriptors that exceed a repulsive energy of 30 Kcal/mol for sterics, and likewise remove data points containing Columbic attraction or repulsion of say, 20 Kcal/mol. We studied various combinations of the energy cutoffs and the analyses are given further in the thesis.
- 3) *Dielectric Treatment*: - This controls the form of the Columbic electrostatic energy calculations. We studied two variations in this- first we used *DISTANCE* where energy falls as  $1/r^2$ . Secondly, we used *CONSTANT* where energy falls as  $1/r$ , the pure Columbic expression.

4) *Column filtering*- We studied both the cases where it is ON. Here it omits from the analysis those columns (lattice points) whose energy variance is less than 2.0 Kcal/mol. When it is OFF it considers all the energy variance.

5) *Probes*- In the CoMFA field there are many probes used for the study. We varied probes to compute the Lennard-Jones steric field and the Coulomb electric fields respectively. The different types of probes used are given in Table 1.

6) *Latent Variables*- This is the optimum number of variables determined from the PLS analysis. We also carried out the analysis for the minimum number of latent variables. In traditional multiple regression methodology one uses ~ 5 observations per term in the QSAR equation. The reason for this is that the molecular descriptors used for QSAR's are not orthogonal to each other. Hence, using those non-orthogonal descriptors one would require about 4 terms in the QSAR model for the 23-molecule data set. Contrarily, because the latent variables of a PLS are orthogonal, one can use many LVs (Latent Variables) as needed to provide the best model. Nonetheless, to be conservative, we provide information about models using the minimum number of LVs in addition to the optimum number.

Finally we calculate  $r_{cv}^2$  from LOO (Leave one out) cross-validation and the squared correlation coefficient  $r^2$  (explained in the INTRODUCTION). Table 1 gives us the CoMFA studied results for the 23-compounds i.e. INTERNAL VALIDATION. We have 8 columns – Fields, Energy cutoff, Dielectric Function, Grid spacing, Probe type, Latent variables,  $r_{cv}^2$  and  $r^2$ .

## Part I: CoMFA with internal validation

### ANALYSIS OF TABLE 1

1) *SUB-SECTION (A)*: - In this part we studied the variation in Fields (steric / electrostatic) with energy cutoff varying from 30-0 kcal/mol for both steric and electrostatic. The dielectric function is kept as  $1/r^2$ ,  $C_{sp3}^+$  probe is used and the column filter is *OFF*. In this sub-section, we found the best



model to be one where the steric cutoff is 30 Kcal/mol and the electric cutoff is at 10 Kcal/mol. A  $q^2$  value of 0.836 for 6 Latent Variables and 0.734 for 3 latent variables and  $r^2$  value of 0.988 is obtained.

2) *SUB-SECTION (B)*: - In this part we considered all the above parameters except the column filter is kept *ON*. The best model is the one where the steric cutoff is 30 Kcal/mol and the electric cutoff is 20 Kcal/ mol. A  $q^2$  value of 0.833 for 6 Latent Variables, 0.746 for 3 Latent Variables and  $r^2$  value of 0.994 is obtained.

For rest of the study we considered column filter *ON*.

3) *SUB-SECTION (C)*: - In this part we considered the dielectric as  $1/r$ . We studied the effect of various force fields and their cutoff.  $C^+ sp^3$  probe is used for this study. The best results are obtained for the case where we had both the fields and their energy cutoff at 5/30 respectively. A  $q^2$  value of 0.840 for 5 variables, 0.772 for 3 latent variables and  $r^2$  value of 0.990 is obtained.

4) *SUB-SECTION (D)*:- In this case we kept the original conditions of *SUB-SECTION (B)* and studied the effect of grid spacing, for both the cases when dielectric function is  $1/r^2$  and  $1/r$ .

(i) For dielectric  $1/r^2$ - we get the best model when both the steric and electrostatic fields are considered, with energy cutoff 30/20 Kcal/mol, respectively and for the grid spacing 2.00. A  $q^2$  value of 0.833 for 6 latent variables, 0.732 for 3 latent variables and  $r^2$  value of 0.977 is obtained.

(ii) For dielectric  $1/r$ - we get the best model for the grid spacing 1.5, when both the fields are considered, the energy cutoff's for steric and electrostatics are 30/20 respectively. A  $q^2$  value of 0.734 for 6 latent variables, 0.606 for 3 latent variables and  $r^2$  value of 0.993 is obtained.

5) *SUB-SECTION (E)*:- We considered the variation with respect to probe type. The various combinations of fields and their cutoffs are given in Table 1 (sub-section E). Dielectric function is studied for both the cases  $1/r^2$  and  $1/r$ . We get the best model for the condition when both fields are considered and their cutoffs are 30/20 respectively, dielectric constant is  $1/r^2$ ,  $O^- sp^3$  probe is used. A  $q^2$  value of 0.756 for 6 latent variable, 0.653 for 3 latent variables and  $r^2$  value of 0.993 is obtained.

In this PART I of the thesis, we see that in all the models generated, we got a very high value of  $q^2$  for 6/3 LV's (Latent Variables) and also the value of  $r^2$  is high.

**Table 1: Influence of Selected Variables on CoMFA Results for 23**

**Catalysts**

Fields (a)	Energy Cutoff (b)	Dielectric Function	Grid Spacing (c)	Probe Type	Latent Variables (d)	$r_{cv}^2$	$r^2$ (e)
<b>SUB-SECTION (A)</b>							
<b>FIELD</b>							
Both	30/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.804//0.707	0.991
Both	30/20	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.833//0.732	0.977
Both	30/10	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.836//0.734	0.992
Both	30/5	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.832//0.729	0.993
E	30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.688//0.626	0.994
S	30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.784//0.659	0.995
Both	20/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.779//0.680	0.995
Both	10/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.774//0.656	0.994
Both	5/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.776//0.626	0.988
<b>SUB-SECTION (B)</b>							
<b>Column Filtering<sup>(h)</sup></b>							
Both	30/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.813//0.721	0.991
Both	30/20	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.833//0.746	0.994
Both	30/10	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.810//0.720	0.993
Both	30/5	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.832//0.729	0.993
Both	20/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.793//0.697	0.995

Both	10/30	$1/r^2$	2.00	$C_{sp3}^+$	6//3	0.789//0.683	0.994
Both	5/30	$1/r^2$	2.00	$C_{sp3}^+$	5//3	0.829//0.775	0.988

**SUB-SECTION (C)**

**Dielectric**

Both	30/30	$1/r$	2.00	$C_{sp3}^+$	6//3	0.825//0.717	0.994
Both	30/20	$1/r$	2.00	$C_{sp3}^+$	6//3	0.826//0.715	0.994
Both	30/10	$1/r$	2.00	$C_{sp3}^+$	6//3	0.822//0.721	0.993
Both	30/5	$1/r$	2.00	$C_{sp3}^+$	6//3	0.808//0.717	0.993
Both	20/30	$1/r$	2.00	$C_{sp3}^+$	6//3	0.808//0.717	0.993
Both	10/30	$1/r$	2.00	$C_{sp3}^+$	6//3	0.786//0.682	0.992
Both	5/30	$1/r$	2.00	$C_{sp3}^+$	6//3	0.786//0.659	0.990
Both	30/30	$1/r$	2.00	$C_{sp3}^+$	6//3	0.838//0.720	0.994
Both	30/20	$1/r$	2.00	$C_{sp3}^+$	6//3	0.801//0.702	0.994
Both	30/10	$1/r$	2.00	$C_{sp3}^+$	6//3	0.755//0.661	0.993
Both	30/5	$1/r$	2.00	$C_{sp3}^+$	6//3	0.810//0.686	0.993
Both	20/30	$1/r$	2.00	$C_{sp3}^+$	6//3	0.820//0.717	0.993
Both	10/30	$1/r$	2.00	$C_{sp3}^+$	6//3	0.814//0.696	0.992
Both	5/30	$1/r$	2.00	$C_{sp3}^+$	5//3	0.840//0.772	0.990

**SUB-SECTION (D)**

**Grid Spacing**

Both	30/20	$1/r^2$	2.00	$C_{sp3}^+$	6//3	0.833//0.732	0.977
Both	30/20	$1/r^2$	1.75	$C_{sp3}^+$	6//3	0.717//0.598	0.990
Both	30/20	$1/r^2$	1.50	$C_{sp3}^+$	6//3	0.729//0.632	0.992
Both	30/20	$1/r^2$	1.0	$C_{sp3}^+$	6//3	0.696//0.585	0.994

Both	30/20	1/r	2.00	$C_{sp^3}^+$	6//3	0.630//0.510	0.992
Both	30/20	1/r	1.75	$C_{sp^3}^+$	6//3	0.697//0.563	0.994
Both	30/20	1/r	1.50	$C_{sp^3}^+$	6//3	0.734//0.606	0.993
Both	30/20	1/r	1.00	$C_{sp^3}^+$	6//3	0.692//0.565	0.994

SUB-SECTION (E)							
Probe Type							
Both	30/30	$1/r^2$	2.00	$H^+$	6//3	0.702//0.604	0.987
Both	30/20	$1/r^2$	1.75	$H^+$	6//3	0.717//0.598	0.990
Both	30/20	$1/r^2$	1.50	$H^+$	6//3	0.739//0.644	0.993
Both	30/20	$1/r^2$	1.00	$H^+$	6//3	0.726//0.629	0.993
Both	30/20	1/r	2.00	$H^+$	6//3	0.671//0.557	0.987
Both	30/20	1/r	1.75	$H^+$	6//3	0.686//0.567	0.991
Both	30/20	1/r	1.50	$H^+$	6//3	0.714//0.620	0.994
Both	30/20	1/r	1.00	$H^+$	6//3	0.700//0.599	0.993
Both	30/20	$1/r^2$	2.00	$O_{sp^3}^-$	6//3	0.684//0.610	0.989
Both	30/20	$1/r^2$	1.75	$O_{sp^3}^-$	6//3	0.697//0.602	0.992
Both	30/20	$1/r^2$	1.50	$O_{sp^3}^-$	6//3	0.756//0.653	0.993
Both	30/20	$1/r^2$	1.00	$O_{sp^3}^-$	6//3	0.739//0.638	0.995
Both	30/20	1/r	2.00	$O_{sp^3}^-$	6//3	0.619//0.533	0.989
Both	30/20	1/r	1.75	$O_{sp^3}^-$	6//3	0.684//0.563	0.992
Both	30/20	1/r	1.50	$O_{sp^3}^-$	6//3	0.731//0.625	0.993
Both	30/20	1/r	1.00	$O_{sp^3}^-$	6//3	0.704//0.595	0.990

(a) Fields are Steric (S) and Electrostatic (E) or Both

(b) Steric cutoff listed first/Coulomb cutoffs listed second, values in Kcal/mol

- (c) Angstrom
- (d) Optimum/Minimum number of components
- (e) These values are for three latent variables only
- (f) Column filtering set to 2.

## PART II: CoMFA with External Validation

### Analysis of Table II

To perform this assessment we selected, randomly, four of the 23 compounds to serve as an external test set (catalysts 10,15,18 and 22) and used the remaining 19 molecules to perform the CoMFA analysis. The same strategy described above was used here. The results are compiled in Table 2 where the column headings and the column sub-sections have the same meaning as described for Table 1.

**Table 2: Influence of Selected Variables on CoMFA Results for Training Set of 19 Catalysts**

Fields (a)	Energy Cutoff (b)	Dielectric Function	Grid Spacing ©	Probe Type	Latent Variables (d)	$r_{cv}^2$	$r^2$ (e)
<b>SUB-SECTION (A)</b>							
<b>FIELD</b>							
Both	30/30	$1/r^2$	2.00	$C_{sp^3}^+$	6//3	0.763//0.635	0.998
Both	30/20	$1/r^2$	2.00	$C_{sp^3}^+$	6//3	0.785//0.653	0.998
Both	30/10	$1/r^2$	2.00	$C_{sp^3}^+$	5//3	0.784//0.672	0.998
Both	30/5	$1/r^2$	2.00	$C_{sp^3}^+$	5//3	0.779//0.673	0.999
E	30	$1/r^2$	2.00	$C_{sp^3}^+$	6//3	0.666//0.573	0.997

S	30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.745//0.663	0.998
Both	20/30	$1/r^2$	2.00	$C^+_{sp3}$	5//3	0.739//0.632	0.999
Both	10/30	$1/r^2$	2.00	$C^+_{sp3}$	5//3	0.731//0.621	0.999
Both	5/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.723//0.566	0.999
<b>SUB-SECTION (B)</b>							
<b>COLUMN FILTERING <sup>(6)</sup></b>							
Both	30/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.764//0.650	0.998
Both	30/20	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.764//0.662	0.998
Both	30/10	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.734//0.648	0.998
Both	30/5	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.713//0.615	0.999
Both	20/30	$1/r^2$	2.00	$C^+_{sp3}$	5//3	0.758//0.663	0.999
Both	10/30	$1/r^2$	2.00	$C^+_{sp3}$	6//3	0.731//0.621	0.999
Both	5/30	$1/r^2$	2.00	$C^+_{sp3}$	3	0.615	0.999
<b>SUB-SECTION (C)</b>							
<b>Dielectric</b>							
Both	30/30	$1/r$	2.00	$C^+_{sp3}$	5//3	0.773//0.671	0.994
Both	30/20	$1/r$	2.00	$C^+_{sp3}$	5//3	0.789//0.682	0.999
Both	30/10	$1/r$	2.00	$C^+_{sp3}$	5//3	0.792//0.692	0.997
Both	30/5	$1/r$	2.00	$C^+_{sp3}$	5//3	0.781//0.689	0.999
Both	20/30	$1/r$	2.00	$C^+_{sp3}$	5//3	0.784//0.681	0.999
Both	10/30	$1/r$	2.00	$C^+_{sp3}$	5//3	0.756//0.653	0.999
Both	5/30	$1/r$	2.00	$C^+_{sp3}$	5//3	0.742//0.621	0.999
Both	30/30	$1/r$	2.00	$C^+_{sp3}$	5//3	0.783//0.678	0.999
Both	30/20	$1/r$	2.00	$C^+_{sp3}$	5//3	0.695//0.647	0.999
Both	30/10	$1/r$	2.00	$C^+_{sp3}$	4//3	0.608//0.555	0.999

Both	30/5	1/r	2.00	C <sup>+</sup> <sub>sp3</sub>	6//3	0.668//0.544	0.999
Both	20/30	1/r	2.00	C <sup>+</sup> <sub>sp3</sub>	5//3	0.811//0.702	0.999
Both	10/30	1/r	2.00	C <sup>+</sup> <sub>sp3</sub>	5//3	0.779//0.690	0.999
Both	5/30	1/r	2.00	C <sup>+</sup> <sub>sp3</sub>	4//3	0.699//0.671	0.999
<b>SUB-SECTION (D)</b>							
<b>GRID SPACING</b>							
Both	30/20	1/r <sup>2</sup>	2.00	C <sup>+</sup> <sub>sp3</sub>	6//3	0.785//0.653	0.998
Both	30/20	1/r <sup>2</sup>	1.75	C <sup>+</sup> <sub>sp3</sub>	6//3	0.618//0.439	0.998
Both	30/20	1/r <sup>2</sup>	1.50	C <sup>+</sup> <sub>sp3</sub>	6//3	0.671//0.521	0.998
Both	30/20	1/r <sup>2</sup>	1.0	C <sup>+</sup> <sub>sp3</sub>	6//3	0.605//0.465	0.998
Both	30/20	1/r	2.00	C <sup>+</sup> <sub>sp3</sub>	5//3	0.789//0.682	0.999
Both	30/20	1/r	1.75	C <sup>+</sup> <sub>sp3</sub>	5//3	0.561//0.423	0.999
Both	30/20	1/r	1.5	C <sup>+</sup> <sub>sp3</sub>	5//3	0.660//0.530	0.999
Both	30/20	1/r	1.00	C <sup>+</sup> <sub>sp3</sub>	5//3	0.632//0.424	0.999
<b>SUB-SECTION (E)</b>							
<b>Probe Type</b>							
Both	30/20	1/r <sup>2</sup>	2.00	H <sup>+</sup>	5//3	0.620//0.505	0.997
Both	30/20	1/r <sup>2</sup>	1.75	H <sup>+</sup>	6//3	0.663//0.500	0.999
Both	30/20	1/r <sup>2</sup>	1.50	H <sup>+</sup>	5//3	0.682//0.569	0.999
Both	30/20	1/r <sup>2</sup>	1.00	H <sup>+</sup>	5//3	0.658//0.531	0.999
Both	30/20	1/r	2.00	H <sup>+</sup>	5//3	0.615//0.517	0.998
Both	30/20	1/r	1.75	H <sup>+</sup>	6//3	0.655//0.522	0.999
Both	30/20	1/r	1.50	H <sup>+</sup>	5//3	0.691//0.584	0.999
Both	30/20	1/r	1.00	H <sup>+</sup>	5//3	0.646//0.534	0.996
Both	30/20	1/r <sup>2</sup>	2.00	H <sup>+</sup>	5//3	0.608//0.512	0.999

Both	30/20	$1/r^2$	1.75	$O_{sp^3}^-$	6//3	0.603//0.468	0.998
Both	30/20	$1/r^2$	1.50	$O_{sp^3}^-$	5//3	0.680//0.573	0.998
Both	30/20	$1/r^2$	1.00	$O_{sp^3}^-$	5//3	0.658//0.521	0.997
Both	30/20	$1/r$	2.00	$O_{sp^3}^-$	5//3	0.589//0.506	0.999
Both	30/20	$1/r$	1.75	$O_{sp^3}^-$	6//3	0.614//0.485	0.999
Both	30/20	$1/r$	1.50	$O_{sp^3}^-$	5//3	0.680//0.565	0.999
Both	30/20	$1/r$	1.00	$O_{sp^3}^-$	5//3	0.646//0.523	0.999

(a) Fields are Steric(S) and Electrostatic (E) or Both

(b) Steric cutoff listed first/Coulomb cutoffs listed second, values in Kcal/mol

(c) Angstrom

(d) Optimum/Minimum number of components

(e) These values are for three latent variables only

(f) Column filtering set to 2.

1) *SUB-SECTION (A)*: - In this sub-section, we studied the effect of the variation in fields (steric / electrostatic), their energy cutoff varying from 30-0 for both steric and electrostatic respectively, dielectric function is  $1/r^2$ ,  $C_{sp^3}^+$  probe is used and column filter is *OFF*. In this sub-section, the best model is the one where the steric cutoff and electric cutoff are 30 Kcal/mol and 20 Kcal/mol, respectively. A  $q^2$  value of 0.785 for 6 latent variables, 0.653 for 3 latent variables and  $r^2$  value of 0.998 is obtained.

2) *SUB-SECTION (B)*: - In this sub-section we studied all the above parameters except the column filter is *ON*. The best model is the one where the steric / electric cutoff are 30 Kcal/mol and 20 Kcal/mol, respectively. A  $q^2$  value of 0.764 for 6 latent variables, 0.662 for 3 latent variables and  $r^2$  value of 0.998 is obtained. For rest of the study we considered column filter *ON*.



3) SUB-SECTION (C): - In this subsection, we considered dielectric function as  $1/r$ . We studied the effect of various force fields and their cutoff.  $C^+ sp^3$  probe is used for this study. The best results are obtained for the case where we had both the fields and their energy cutoff are 30/10 respectively. A  $q^2$  value of 0.792 for 5 variables, 0.692 for 3 latent variables and  $r^2$  value of 0.997 is obtained.

4) SUB-SECTION (D): - In this case we kept the original conditions of SUB-SECTION (B) and studied the effect of grid spacing, for both the cases of dielectric  $1/r^2$  and dielectric  $1/r$ .

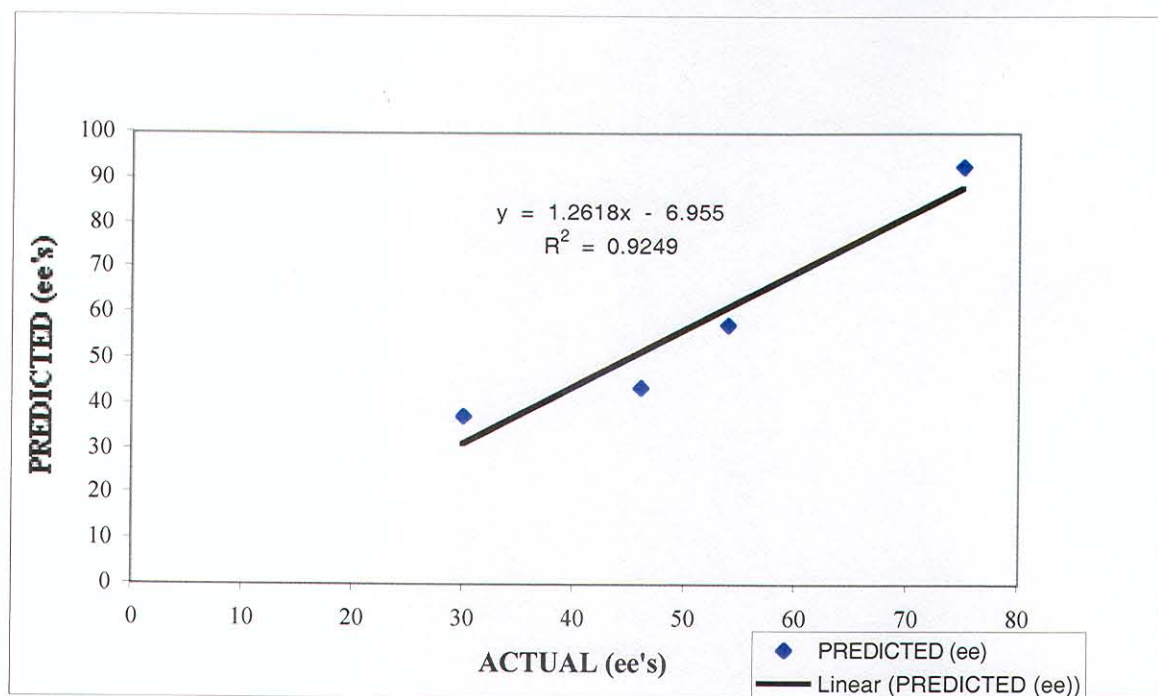
a. For dielectric function  $1/r^2$ - we get the best result when both the fields are considered, their energy cutoff are 30/20 respectively and the grid spacing of 2.0. A  $q^2$  value of 0.785 for 6 latent variables, 0.653 for 3 latent variables and  $r^2$  value of 0.998 is obtained.

b. For dielectric function  $1/r$ - we get the best results for grid spacing 2.0, when both the fields are considered their energy cutoff's are 30/20, respectively. A  $q^2$  value of 0.789 for 5 latent variables, 0.653 for 3 latent variables and  $r^2$  value of 0.998 is obtained.

5) SUB-SECTION (E):- We considered the variation with respect to probe type. We considered various combinations of fields as given in the above table and their cutoffs. We studied both condition when the dielectric are  $1/r^2$  and  $1/r$ . The best results are obtained when we considered both the fields and their cutoffs are 30/20 Kcal/mol, respectively. The dielectric function of  $1/r^2$  and  $H^+ sp^3$  probe is used. A  $q^2$  value of 0.682 for 5 latent variables, 0.569 for 3 latent variables and the  $r^2$  value of 0.999 is obtained.

On comparing the above two tables we see that although we have eliminated only 4 compounds for validation there is a difference in the analysis with respect to variation in the parameters. Because fewer compounds are being used in this CoMFA one anticipates and one sees a degradation of the quality of the models derived. Still, the results are extremely good with  $r_{cv}^2$  values approaching 0.8 for five latent variables (LVs) and 0.7 for three latent variables (LVs). Again, we emphasize that these are predictive models but they are validated internally only. Our main aim was to do EXTERNAL VALIDATION. Using the best model we then predicted the ees of the four compounds omitted from the training set. A plot of compound ee versus experimental ee gave a linear relationship with a squared

correlation coefficient,  $r^2$ , of 0.93. Most importantly, in concordance with the criteria highlighted by Golbraikh and Tropsha, we find a slope near unity (1.23) and an intercept of 7.0, as shown in the Figure 19.<sup>23</sup> While the latter is somewhat elevated we find that the model derived from the test set satisfies well Tropsha's arguments for a statistically valid QSAR.<sup>23</sup>

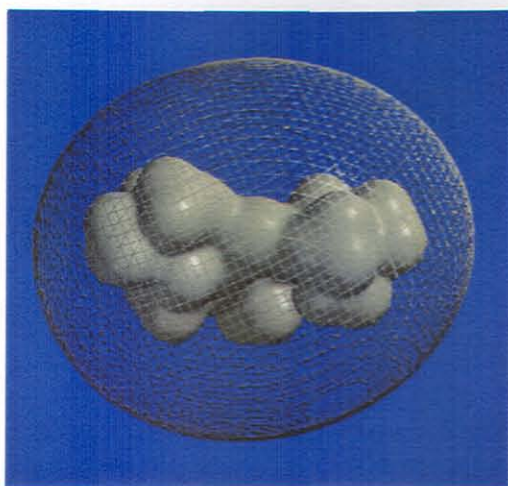


**Figure 19: Graph of predicted (ee's) vs actual (ee's)**

Given a relatively small set of catalysts (19), we are thus able to predict the ee of new catalysts. Hence we have shown here that commercially available software can be readily implemented for this purpose. This is important because this size of data set and this range of ees are about what one would find from an initial combi-chem evaluation of chiral catalysts. From this limited data one can then make some predictions about how good or how poor a given catalyst will be that has yet to be made.

### PART III: Understanding how the catalysts work

The third part of this research project is meant to help explain how these catalysts work. In particular we want to know, in a quantitative fashion, what role steric and electrostatic factors play in asymmetric induction. Inspection of mechanical models provide no quantifiable information about this. Computational models are better but they suffer from being difficult to derive quantitative information. For example, in Figure 20 are plots of van der Waals and electrostatic surfaces of the  $\text{Cu}^{2+}$  bis-oxazoline, 23.



**Figure 20: Electrostatic potential (grid) surrounding the van der Waals surface of catalyst 23**

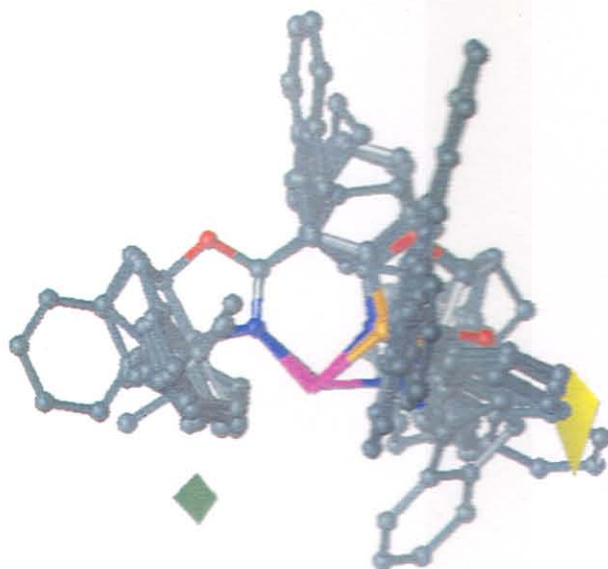
Although the electrostatic potential surrounding the catalyst in Figure 20 looks symmetric, it isn't (like the van der Waals surface it too is chiral) and we want to know what percent of the asymmetric induction for such catalysts can be attributed to these subtle, chiral electrical effects. The van der Waals surface has better defined grooves and cavities in which stereoinduction can arise but, as with the electrostatic plot, quantifying this is difficult. What is needed is information concerning

steric and electrostatic modes of stereoiduction. Specifically we want to know how much of the stereoiduction arises from steric effects and how much comes from electrical effects.

CoMFA can provide insights concerning these issues. Moreover, one can also use the CoMFA model as a guide for synthesis. In this study, as in QSAR, we want to find a linear relationship that relates an activity (in this case the ability of a catalyst to induce asymmetry during a Diels-Alder reaction) to the intensity of the surrounding fields. From our best models we find that the steric field can describe 60-70% of the variance in the data. The remaining 30%-40% of the variance is attributed to the electrostatic field. This can be loosely interpreted as meaning that most of the stereoiduction originates from steric effects of the ligands surrounding the catalyst.

While this seems intuitive, and while it is fully consonant with our perceptions of how these particular catalysts work, we point out that we are able to provide a quantitative assessment of how much each effect contributes to the stereoiduction. An intuitive evaluation for other catalysts may not be so obvious, however, when electrical effects are as important as (or more so) than steric effects. Hence models like ours have the potential for assisting in the construction of improved catalysts where both steric bulk and electronics can be modulated via synthesis.

Another advantage of using a QSAR like that presented here is that one can visualize the large number of computed descriptor coefficients by making iso-value contour maps of those coefficients at grid points surrounding the aligned data set. Rather than using the coefficients themselves we present a more common "standard deviation times coefficient" plot ( $STDV * COEFFICIENTS$ ), where, at each grid point the standard deviation of the energies for all compounds is multiplied by the PLS coefficient. Plotted in figure 21 are regions of space where steric bulk should enhance or destroy stereoiduction.



**Figure 21: CoMFA steric STDEV\*COEFF contour plot. Shown inside the field is the aligned set of 23 chiral catalysts with hydrogen atoms removed for clarity. Placement of bulky groups near the green region (contoured at contribution level = 93) and/or removal of steric bulk near the yellow region (contoured at contribution level = 7) should increase ee for those catalysts that are not very stereoselective.**

We focus here on steric influences because most of the variance is explained with the steric field. In figure 21 is the aligned data set of 23 catalysts. To increase enantioselectivity the models indicate that more bulk is to be placed in the space encapsulated in green and steric bulk is to be removed from the region encapsulated in yellow. If these plots are meaningful one should find that highly efficient catalysts already have steric bulk in the green region and are already devoid of bulk in the yellow region. In figure 22 we present one such efficient catalyst (3) where this clearly seen.

Furthermore one would expect that inefficient chiral catalysts either lack steric bulk in the green region and /or have too much steric bulk in the yellow region. In figure 22 we show catalysts 13 (ee =96). In this figure it appears that the phenyl group on the left-hand side of the diagram is too small to be effective while the phenyl group on the right is too large. Modulating bulk such that more steric pressure exists on the left-hand side and less steric pressure exists on the right-hand side is being suggested by our model as a means of enhancing stereoselectivity.

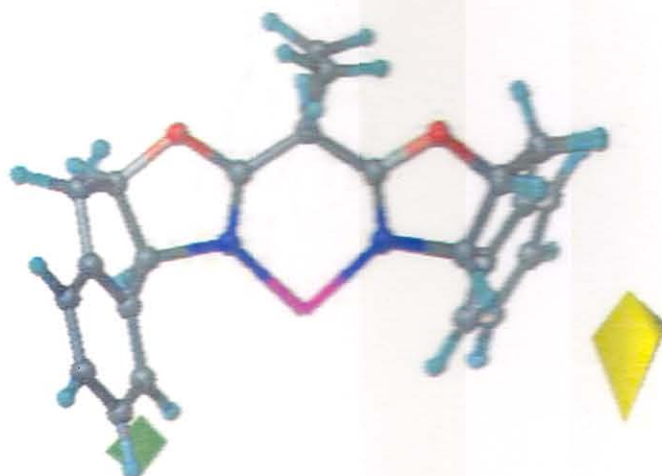


Figure 22: CoMFA steric  $STDEV \times COEFF$  contour plot. Shown inside the field is the highly efficient catalyst 3 (ee = 96%). It is to be noted that significant steric bulk lies in the green region while the yellow region is devoid of steric bulk confirming the model.

Figure 23 gives the interaction of the electrostatic field with the set of 23 aligned catalysts. To increase enantioselectivity the models indicate that more charge is to be placed in the space encapsulated in blue and electrostatic charge is to be removed from the region encapsulated in red. If these plots are meaningful one should find that highly efficient catalysts already have electrostatic charge in the blue region and are already devoid of charge in the red region.

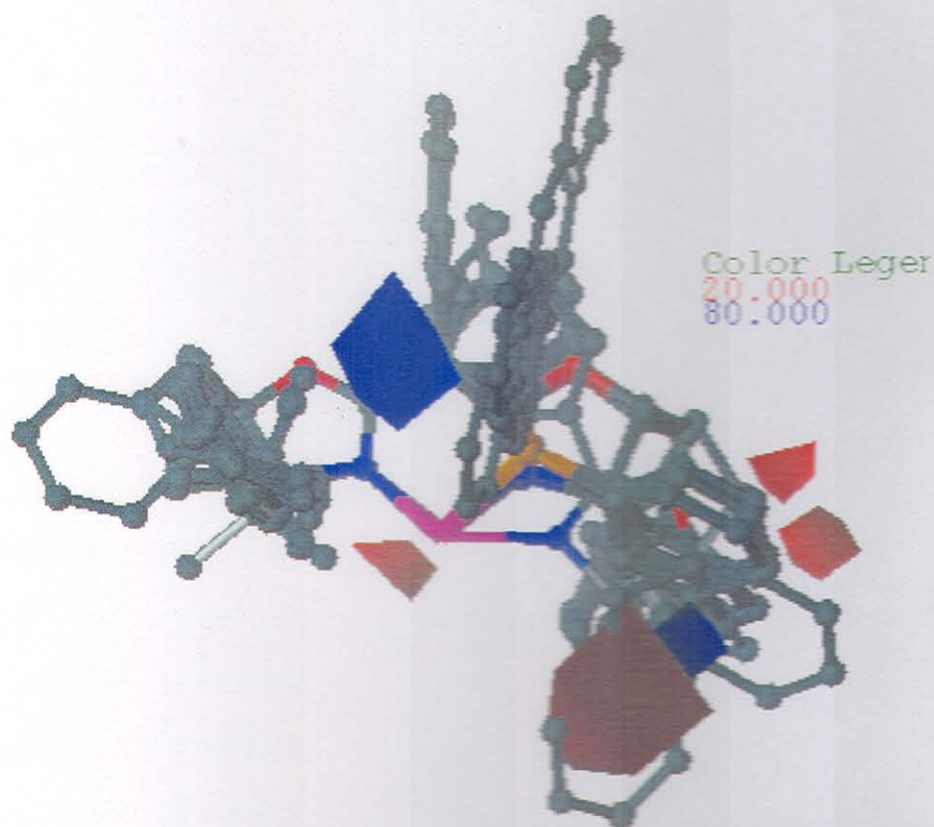


Figure 23: CoMFA electrostatic  $STDEV \times COEFF$  contour plot. Shown is the aligned set of 23 chiral catalyst with hydrogen atom removed for clarity. Placement of groups near blue gives more electrostatic and / or removal near red gives more negative electrostatic.

## CONCLUSIONS

There were two goals set forth at the beginning of our work. The first was to see if one could use off-the-shelf software to generate high quality QSAR's for a small data set of chiral catalysts. It was shown that extremely high cross-validated  $r^2$  values could be generated quickly, and with undue difficulty. The importance of this is that the set of compounds used here is comparable to a typical small combinatorial library that one might develop in an exploratory research endeavor. We have shown that such small libraries are amenable to such modeling. In this work we carried out two CoMFA's. One included all 23 catalysts with internal validation only. The second study divided the catalysts into a training set that was used to develop the mathematical model and a test set that was used to validate the model. Moreover, we demonstrated here that the external test set gave results for which a plot of predicted ees versus observed ees had a high coefficient of regression, a slope near unity, and an intercept that is reasonably close to zero, thus fulfilling the stringent requirements for a statistically valid mathematical model set forth by Golbraikh and Tropsha.<sup>23</sup>

The second goal of the work was to better understand, qualitatively and quantitatively, why some catalysts work efficiently at asymmetric induction for the reaction in figure 15 while others do not. Quantitatively we are able to predict with a high degree of accuracy which catalysts are effective at carrying out this stereoinduction and which are not. Quantitatively we were also able to show that approximately 70% of the variance in the model arises from the steric field while the remaining 30% is electrostatic in nature. Quantitatively, we were able to define regions in space where steric bulk will influence the outcome of the reaction. Using this model, then, we could now compute the shapes of other copper-coordinated ligands and predict their activity, i.e., we are in the position of doing statistically meaningful computer-aided molecular design (CAMD).



## REFERENCES

1. March, J. *Advanced Organic Chemistry: Reactions, Mechanisms and Structure*, fourth edition, New York: Wiley, **1992**.
2. <http://www.leffingwell.com/chirality/>
3. McMurry, J. *Organic Chemistry*, fifth edition, Pacific Grove, Calif: Brooks/Cole. Pub., **2000**.
4. Dias, C. L. *Braz. Chem. Soc.*, **1997**, Vol. 8, No. 4, 289-332.
5. Kumosinski, T.; Liebman, M. N. *ACS SYMPOSIUM SERIES*, **1994**.
6. Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1997**, 40, 1136-1148.
7. Sadler, R. B.; Cho, S. J.; Ishaq, K. S.; Chae, K.; Korach, S. K. *J. Med. Chem.* **1998**, 41, 2261-2267.
8. Leach, A. R. *Molecular Modelling: Principles and Applications*, Second edition, Addison Welsey Longman limited, **1996**.
9. Yao, S.; Roberson, M.; Reichel, F.; Hazell, R. G.; Jorgensen, K. A. *J. Org. Chem.*, **1999**, 64, 6677-6687.
10. Schefzick, S. *Study of Chirality with Computational Methods*, PhD Thesis, Purdue University, **2002**.
11. Tomooka, K.; Wang, L. F.; Komine, N.; Nakai, T. *Tetrahedron Lett.* **1999**, 40, 6813-6816.
12. Evans, D. A.; Rovis, T.; Johnson, J. S. *Pure Appl. Chem.* **1999**, 71, 1407-1415.
13. Cramer, R. D.III.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, 110, 5959-5967.
14. Tripos Associates Inc., 1699 South Hanley Road, St. Louis, MO, 63144 USA.
15. Ghosh, A. K.; Mathivanan, P.; Cappiello, J. *Tetrahedron Lett.* **1996**, 37, 3815.
16. Ghosh, A. K.; Mathivanan, P.; Cappiello, J.; Krishnan, P. *Tetrahedron: Asymmetry*, **1996**, 7, 2165.
17. Davies, I. W.; Gerena, L.; Cai, D.; Larsen, R. D.; Verhoeven, T. R.; Reider, P. J. *Tetrahedron Lett.* **1997**, 38, 1145.

18. Evans, D. A.; Bartroli, H.; Shih, T. L. *J. Am. Chem. Soc.* **1981**, 103, 2127.
19. Evans, D. A.; Miller, S. J.; Lectka, T.; Matt, P. V. *J. Am. Chem. Soc.* **1999**, 121, 7559-7573.
20. Helmchen, G.; Sagasser, I. *Tetrahedron Lett.* **1998**, 39, 261-264.
21. Evans, D. A.; Miller, S. J.; Lectka, T. *J. Am. Chem. Soc.* **1993**, 115, 6460.
22. Evans, D.A.; Miller, S. J.; Lectka, T. *J. Am. Chem. Soc.* **1993**, 34, 7027.
23. Golbraikh, A.; Tropsha, A. *J. Molec. Graphics and Modelling*, **2002**, 20, 269.