CHAPTER 2
METHODOLOGY

- Quantum Mechanical Parameters
- Description of Software Used
  - MOPAC
  - PCMODEL Software
  - CACHe Software
- Geometry Optimization
- Experimental
- References
**METHODOLOGY**

**Quantum Mechanical Parameters:**

Quantum chemical methods and molecular modeling techniques enable the definition of a large number of molecular and local quantities characterizing the reactivity, shape and binding properties of a complete molecule as well as of molecular fragments and substituents. Because of the large well-defined physical information content encoded in many theoretical descriptors, their use in the design of a training set in a QSAR study.

Quantum chemical descriptors [1,2,3] such as net atomic charges, highest occupied molecular orbital/lowest unoccupied molecular orbital (HOMO-LUMO) energies, frontier orbital densities, and superdelocalizabilities have been shown to correlate well with various biological activities [4].

We have based our QSAR study of β-Carboline derivatives on the basis of following quantum mechanical parameters:

1. Heat of formation ($\Delta H_f$)
2. Molecular weight (Mw)
3. Total energy ($E_T$)
4. HOMO energy ($\epsilon_{\text{HOMO}}$)
5. LUMO energy ($\epsilon_{\text{LUMO}}$)
6. Absolute Hardness ($\eta$)
7. Electronegativity ($\chi$)

The evaluation of these parameters is given as below:

The heat of formation is defined as:

$$\Delta H_f = E_{\text{elect}} + E_{\text{nuc}} - E_{\text{isol}} + E_{\text{atom}}$$  \hspace{1cm} \text{Equ.(1)}

where $E_{\text{elect}}$ is the electronic energy, $E_{\text{nuc}}$ is the nuclear-nuclear repulsion energy, $E_{\text{isol}}$ is the energy required to strip all the valence electrons of all the atoms in the system, and $E_{\text{atom}}$ is the total heat of atomization of all the atoms in the system.

Total energy ($E_T$) of a molecular system is the sum of the total electronic energy, $E_{\text{ee}}$ and the energy of internuclear repulsion, $E_{\text{in}}$. The total electronic energy of the system is given by [5]

$$E = P (H + F) / 2$$  \hspace{1cm} \text{Equ.(2)}

where $P$ is the density matrix and $H$ is the one-electron matrix.

Parr et al. [6] defined electronegativity as the negative of chemical potential:

$$\chi = -\mu = - \left( \frac{\partial E}{\partial N} \right)_{\nu(r)}$$  \hspace{1cm} \text{Equ. (3)}

The absolute hardness, $\eta$, is defined as [7]

$$\eta = 1/2. \left( \frac{\delta \mu}{\delta N} \right)_{\nu(r)}$$

$$= 1/2. \left( \frac{\delta^2 E_T}{\delta N^2} \right)_{\nu(r)}$$  \hspace{1cm} \text{Equ. (4)}
where $E_T$ is the total energy, $N$ the number of electrons of the chemical species, and $v(r)$ the external potential.

The operational definition of absolute hardness and electronegativity \[^8\] is defined as:

$$
\chi = -\mu = (\text{IP} + \text{EA}) / 2 \quad \text{Equ. (6)}
$$

where IP and EA are the ionization potential and electron affinity respectively, of the chemical species.

According to the Koopman’s theorem, the IP is simply the eigen value of the HOMO with change of sign \[^9\] and the EA is the eigen value of the LUMO with change of sign hence the equations 5 and 6 can be written as:

$$
\eta = (\varepsilon_{\text{LUMO}} - \varepsilon_{\text{HOMO}})/2 \quad \text{Equ. (7)}
$$

$$
\chi = (\varepsilon_{\text{LUMO}} + \varepsilon_{\text{HOMO}})/2 \quad \text{Equ. (8)}
$$

Finally a more general but important property of a molecular system is the molecular weight (Mw) that has been tested as descriptor.

**Description of Software Used for Evaluation of Above Parameters:**

The values of above parameters have been evaluated by following softwares:

1- **MOPAC:**

MOPAC is a general-purpose semiempirical molecular orbital package for the study of solid state and molecular structures and reactions.

The semiempirical Hamiltonians MNDO \[^{10}\], MNDO/3 \[^{11}\], AM1 \[^{12}\], PM3 \[^{13}\], and MNDO-$d$ \[^{14-15}\] are used in the electronic part of the calculation to obtain molecular orbitals, the heat of formation and its derivative with respect to molecular geometry. Using these results MOPAC calculates the vibrational spectra, thermodynamic quantities, isotopic substitution effects and force constants for molecules, radicals, ions, and polymers. For studying chemical reactions, a transition state location routine \[^{16}\] and two transition state optimizing routines \[^{17-18}\] are available. For users to get the most out of the program, they must understand how the program works, how to enter data, how to interpret the results, and what to do when things go wrong.

While MOPAC calls upon many concepts in quantum theory and thermodynamics and uses some fairly advanced mathematics, the user need not be familiar with these specialized topics. MOPAC is written with the non-theoretician in mind. The input data are kept as simple as possible, so users can give their attention to the chemistry involved and not concern themselves with quantum and thermodynamic exotica.

The simplest description of how MOPAC works is that the user creates a data-file, which describes a molecular system and specifies what kind of calculations and output are desired. The user then commands MOPAC to carry out the calculation.
using that data-file. Finally the user extracts the desired output on the system from the
output files created by MOPAC. The name MOPAC should be understood to mean
"Molecular Orbital Package". The origin of the name is somewhat unusual, and might
be of general interest: The original program was written in Austin, Texas. One of the
roads in Austin is unusual in that the Missouri-Pacific railway runs down the middle
of the road. Since this railway was called the MO-PAC, when names for the program
were being considered, MO-PAC was an obvious contender.

Most of the theory used in MOPAC is in the literature, so that in principle one
could read and understand the algorithm. However, it is convenient to have the theory
gathered together in one document. The theory given here is intended for users who
(a) want to modify MOPAC, or (b) want to understand how MOPAC does what it
does. All derivations can be followed with a little patience, and, as far as possible, "It
can be shown that" jumps in logic have been avoided.

There are five distinct methods available within MOPAC: MINDO/3, MNDO,
AM1, PM3, and MNDO-d. All are semiempirical, and have roughly the same
structure. A complete knowledge of these methods is not necessary in order to use
MOPAC; however, a superficial understanding of these methods and their
relationship to \textit{ab initio} methods is important for using MOPAC and particularly for
interpreting the results.

The five methods within MOPAC have many features in common. They are
all self-consistent field (SCF) methods, they take into account electrostatic repulsion
and exchange stabilization, and, in them, and all calculated integrals are evaluated by
approximate means. Further, they all use a restricted basis set of one s orbital and
three p orbitals \((p_x, p_y, \text{and } p_z)\) per atom (except MNDO-d, which has five d orbitals in
addition to the s, p basis set) and ignore overlap integrals in the secular equation.
Thus, instead of solving

\[ | H \cdot ES | = 0, \]

The expression

\[ | H \cdot E | = 0, \]

in which \( H \) is the secular determinant, \( S \) is the overlap matrix, and \( E \) is the set
of eigenvalues, is solved. These approximations considerably simplify quantum
mechanical calculations on systems of chemical interest. As a result, larger systems
can be studied. Computational methods are only models, and there is no advantage in
rigorously solving Schrödinger's equation for a large system if that system has had to
be abbreviated in order to make the calculations tractable. Semiempirical methods are
thus seen to be well balanced: they are accurate enough to have useful predictive
powers, yet fast enough to allow large systems to be studied.

All five semiempirical methods contain sets of parameters. For MINDO/3
atomic and diatomic parameters exist, while MNDO, AM1, PM3, and MNDO-d use
only single-atom parameters. Not all parameters are optimized for all methods; for
example, in MINDO/3, MNDO and AM1 the two electron one center integrals are
normally taken from atomic spectra. In the list given in following Table, parameters
optimized for a given method are indicated by '•'. A '+' indicates that the value of
the parameter was obtained from experiment (not optimized). Where either symbol is
given, the associated parameter is not used in that method.
All five semiempirical methods also use two experimentally determined constants per atom: the atomic mass of the most abundant isotope and the heat of atomization.

2- PCM ODEL Software:

PCM ODEL Software has been obtained from Indian Institute of Science, Bangalore, by the courtesy of Prof. Chandrasekharan and has been installed in a P.C. in our Computer Laboratory. This software is used for molecular modeling and for geometry optimization to create data file for MOPAC input.

3. CACHe Software:

CACHe is a computer-aided molecular design (CAMD) modeling tool for the Microsoft® Windows ME, Microsoft® Windows 98 or Microsoft® Windows NT 4.0 operating systems. CACHe enables us to draw and model molecules and perform calculations on a molecule to discover molecular properties and energy values.

CACHe uses computational chemistry as an essential part of computer-aided chemistry. By applying computational molecular models derived from mathematical equations to a chemical sample we create, we can calculate molecular properties such as the position and behavior of electrons, and nuclei in certain conditions. The computational chemistry tools that CACHe uses are derived from classical mechanics and quantum mechanics, and are applied to the chemical sample by a number of computational applications that perform calculations based on specified parameters.

GEOMETRY OPTIMIZATION:

Many systematic mathematical procedures (algorithms) exist to find a local minimum of a function of several variables. These procedures will find a local minimum in E in the neighborhood of initially assumed geometry. The process of finding such a minimum is called Geometry Optimization or energy minimization. For a molecule with several conformations, one must repeat the local minimum search procedure for each possible confirmation, so as to locate the Global Minimum. For a large molecule, there may be too many confirmations for all of them to be examined. Moreover, the true-global minimum equilibrium geometry might correspond to a highly unconventional structure that might not think to be considered.

Some procedures to find local minimum in E require only repeated calculations of E at various values of its variables, but these procedures are very insufficient. More efficient procedures require repeated calculation of both E and its derivatives. The set of (3N-6) first partial derivatives of E with respect to each variable constitutes a vector (in a "space" of 3N-6 dimensions) called a gradient of E. at a local minimum; the gradient must be equal to zero. A point on the PES where the gradient is zero is called a stationary point. A stationary on the PES may be a minimum or maximum or a SADDLE POINT.

Analytical calculation of the gradient is key to efficient geometry optimization. The SCF energy expression is given as

\[ E_{\text{HF}} = \sum_{i=1}^{N/2} E_i + 1/2 \sum_{r=1}^{b} \sum_{s=1}^{b} P_{rs} H_{rs}^{\text{core}} + V_{N,N} \]

And its derivative w.r.t. nuclear coordinates would seem to involve the derivatives of the Hrs and (rs/tu) integrals (which occur in Ei), the derivatives of VNN. The
derivatives of integrals are readily calculated, since the derivative of a Gaussian type function with respect to nuclear coordinate is another GTF. The derivatives of VNN are trivial. Thus an analytical formula for the gradient of SCF energy is known. Once the SCF energy E and wave function have been found for some chosen geometry, the time needed to analytically calculate the energy gradient is roughly equal to the time needed to do the SCF wave function and energy calculation.

As well as using the energy gradient (the 3N-6 first partial derivatives \( \partial E/\partial q_1, \partial E/\partial q_2 \ldots \)), some energy minimization methods also use the second derivatives of E. The set of second derivatives

\[
\partial^2 E/\partial q_1^2, \quad \partial^2 E/\partial q_1 \partial q_2, \quad \partial^2 E/\partial q_2^2
\]

When arranged in square array from a matrix called the HESSIAN or the force constant matrix (since the second derivatives of E are force constants). An efficient way to find local minimum of a function of several variables is the Newton Method, which approximates the function by Taylor–series expressions, that is terminated after quadratic terms, and uses accurately evaluated first and second partial derivatives of the function (which occur in the linear and quadratic terms of the Taylor–series), because the analytical calculation of the second derivatives in an SCF calculation is very costly in computer time. The SCF-geometry optimization commonly uses a modification of Newton procedure called quasi-Newton method.

The most frequently employed geometry optimization schemes are those of Mc Iver and Komornicki\(^{19,20}\) and force method due to Pulay.\(^{21-22}\) The Mc Iver and Komornicki method utilizes the minimization algorithm of Murgath and Sargent, a quadratically convergent method, which preserves the molecular symmetry during calculations.

The optimization method developed by Pulay is based on Newton–Raphson minimization algorithm. This method requires some approximation to force constant matrix \( \kappa \), which is assumed to be diagonal. From this matrix and the forces calculated at some initial geometry, a new geometry is generated and the whole process is repeated until the final forces become negligibly small. Force constant calculations require the determination of forces in the neighborhood of the equilibrium. If the ith force \( F_i \) is changed by \( \Delta F_i \) arising from internal coordinate change \( \Delta q_j \), then

\[
K_{ij} = - \Delta F_i / \Delta q_j
\]

This is known as the force method and was used within the CNDO/2 approximation by Pulay et al.\(^{23,24}\). There are several applications of this method at ab initio level\(^{25-31}\).

The force method appears the far more accurate than the straight-forward calculation of force constants via numerical differentiation to the total energy. The latter method, which is widely employed for large molecules, requires the calculation of total energy of selected geometries. Then either, directly or indirectly, some polynomial expression to the energy hyper surface is performed and corresponding second order derivatives are obtained. The analytical form of the polynomial is sensitive both to the assumed grid. The accuracy of polynomial approximation depends also on the precise q dependence on energy. However, the method is computationally simple, only requiring standard energy calculations to be repeated several times. As long as the cost of a single calculation is fairly low the method will
be sufficiently inexpensive. For this reason, this brute force scheme is frequently employed within PM3 calculations.

Stretching force constants (for which the internal vibrational energy coordinates is the charged in bond length) and bending force constants (for which the internal coordinate is the change in valance angle) are most commonly calculated and the diagonal force constants, $K_{ij}$, are usually computed. However, some attempts at a complete prediction of the whole set of the force constants and vibration frequencies do exist.

**Experimental:**

In this research work we have taken 46 derivatives of β-carbolines to propose new QSAR models for the prediction of their activities. The observed biological activities of these compounds have been taken from literature.

For QSAR study, the 3D modeling and geometry optimization of all the derivatives have been carried out on CAChe software by applying semiemperical method using MOPAC 2002. Result of these compounds by the software is given in appendix for sample, and all the values required for the determination of the value of heat of formation, molecular weight, total energy, energy of HOMO, energy of LUMO, absolute hardness and electronegativity have been obtained from this software by solving the equations described earlier and the result are reported in respective tables.

Multilinear regression analysis have been made by Project Leader software associated with CAChe, using the above descriptors in different combinations
References:


