

## CHAPTER 5

# Computational Chemistry

### Contents

5.1	Introduction	151
5.2	Computer Use in Chemistry	152
5.2.1	Visualization	153
5.2.1.1	<i>Structure drawing</i>	153
5.2.1.2	<i>3D visualization</i>	154
5.2.1.3	<i>Visualization of ligand–receptor interactions</i>	155
5.2.2	Calculation and simulation	155
5.2.3	Analysis and storage of data	156
5.3	Conformational Analysis and Energy Minimization	158
5.3.1	The concept	158
5.3.2	Conformational search	160
5.3.3	Minimization of energy	160
5.4	Molecular Mechanics	165
5.5	Molecular Dynamics	170
5.5.1	Definition	170
5.5.2	Development and components	171
5.5.3	The algorithm	172
5.6	Quantum Mechanics	173
5.6.1	The Born–Oppenheimer approximation	179
5.6.2	The Hartree–Fock approximation	179
5.6.3	Density functional theory	180
5.6.4	Semiempirical analysis	181
5.6.4.1	<i>Concept</i>	181
5.6.4.2	<i>Developmental background</i>	182
5.6.4.3	<i>Modified neglect of diatomic overlap</i>	182
5.6.4.4	<i>Austin model 1</i>	184
5.6.4.5	<i>Parametric method 3</i>	184
5.6.4.6	<i>PDDG/PM3 and PDDG/MNDO</i>	185
5.7	Overview and Conclusion	185
	References	188

## 5.1 INTRODUCTION

Advancement in any scientific discipline proves fruitful when it is supplied with suitable media of operations that make the science unambiguous to all its users. The development of various computer applications has enriched the operations involved in different technical and scientific processes. As in other scientific disciplines, the deployment of computer

workstations has instituted a revolution in the research and development of chemistry. Modern medicinal chemistry employs several computational applications in solving simple to complex problems with a high level of accuracy. Nowadays, computers have become an inevitable tool in drug discovery and drug development. The use of desktop computers has enabled the solving of complex chemical problems while requiring little specialist expertise at the laboratory-scale operations. The computer technology has enhanced the research in theoretical chemistry by allowing not only a platform for solving complex chemical equations and models, but also suitable graphical interfaces for a better understanding of the involved chemical phenomena. The quantitative structure–activity relationship (QSAR) methodology employs the quantification of molecular structures followed by the development of mathematical correlation with the aim of predicting the activity/property/toxicity of untested or new chemicals. The entire operation, starting from encoding of structural information to model development, and validation as well as prediction, employs a significant amount of data that can be reliably operated using suitable computer programs. At times, the predicted features have been found to be more accurate and reliable than experiments considering a higher degree of error involved in experimental studies. There are different technical aspects that must be considered while using computers as an essential component in the molecular modeling epitome, and it is thus obvious that the assistance of an expert in computational chemistry would be necessary while dealing with such *in silico* tools. However, it will be very helpful to have some basic insight regarding the fundamental theory involved in such operations based on how developers design their products. Hence, we shall highlight the basic theories involved in defining some of the chemical attributes, like molecular orbital (MO) theory, principles of molecular mechanics (MM), and quantum mechanics along with a focus on the avenues of their implementation in various types of algorithms.

## 5.2 COMPUTER USE IN CHEMISTRY

Discoveries and developments in chemistry have come a long way since the time of the Greek philosopher Thales (sixth century BC), who hypothesized that simple elements (earth, water, air, and fire) form chemical substances. With the passage of time, knowledge of the molecular composition, structural features, arrangement, and properties of chemicals have been developed, and computers have become an essential part of it by acting as a tool of documentation, analysis, and representation. It is not preposterous to state that the incorporation of computer technology has added significant momentum to the research of theoretical chemistry. However, it should be remembered that discovery and development of the theoretical bases are of primary importance, and computers act as an implementation tool. By the use of suitable software and hardware technologies, developers encode the theoretical information by applying definite algorithms, which makes the theory easily applicable to the users. Hence, any

limitations or approximations of a software platform entirely depend upon the developer, not the theory itself. Now, with the incorporation of various applications of computers in solving chemical problems, several terms are used. *Theoretical chemistry* broadly corresponds to the mathematical depiction of chemical information, while *computational chemistry* is used to denote a collection of techniques employed to solve problems of chemistry using computers.

Now, it is evident that computers have played an essential part in almost all types of technical operations, including various chemical analysis systems like spectroscopy and chromatography, which provide analog support to the system. However, in this chapter, we shall focus on the use of computers in enhancing the theoretical aspect of chemistry. Another term, *molecular modeling*, is also used to define the same formalism encompassed by computational chemistry. Technically, the word *model* refers to an idealized depiction of a system, and if we consider mathematical formalism, a model is a way of providing the calculations and predictions of a system [1]. Hence, molecular modeling covers several aspects of theoretical chemical computation and the reasonable prediction thereof. The sphere of theoretical chemistry can be computational as well as noncomputational, depending upon the mathematical basis developed to encode a chemical problem. Hence, the noncomputational part concerns the formulation of analytical expressions for molecular properties, and when sufficient mathematical background is developed in addressing a chemical problem, the algorithm is executed using computational tools. Therefore, we can observe an interdisciplinary embellishment of chemical theory put into strong mathematical formalism, followed by computational encoding. The implementation of computational technology in theoretical chemistry can be broadly viewed in three aspects: namely, visualization, computation, and analysis. In the next sections, we shall present a brief overview of different molecular modeling operations performed with the assistance of computers.

## 5.2.1 Visualization

Computers strengthen general understanding of chemicals by providing a suitable graphical visualization interface. This encompasses the visualization of single chemical structures to the hypothetical interaction pattern between the receptor and ligand.

### 5.2.1.1 Structure drawing

Such tools allow the drawing of a chemical structure in a workspace within the desktop. By using the sketching tool, a user can draw structures employing various chemical bond tools, atoms, chains, ring-template, and other items. The structures are encoded in the form of coordinates that are graphically converted into images on the computer screen. The same structure can be displayed in different graphical forms, some of which include Corey–Pauling–Koltun (CPK), stick, ball-and-stick, space fill, mesh, and ribbon [2]. Of these, the ribbon type of representation is usually

employed for large molecules like proteins and nucleic acids. Figure 5.1 shows several graphical depictions of the paracetamol (*p*-acetamidophenol) molecule. Most of the commercial structure-drawing packages assign a specific color coding to each type of atom, a system that is editable by the user. Examples of some chemical-drawing packages are ChemDraw ([http://www.cambridgesoft.com/Ensemble\\_for\\_Chemistry/ChemDraw/](http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemDraw/)), IsisDraw (a product of MDL Information Systems, Inc.), ChemSketch ([http://www.acdlabs.com/products/draw\\_nom/draw/chemsketch/](http://www.acdlabs.com/products/draw_nom/draw/chemsketch/)), and MarvinSketch (<https://www.chemaxon.com>). Many of these packages also allow the computation of various properties along with the drawing facility. Not only that, most of the software packages provide other tools that perform functions such as checking the structure (valence, bond order, etc.) and generating IUPAC names.

### 5.2.1.2 3D visualization

By using a suitable graphical conversion package, the user can obtain a three-dimensional (3D) visualization of the structure. Many of the molecular modeling packages allow the simple and direct conversion of a two-dimensional (2D) structural format into 3D when a 2D structure is opened or pasted into its drawing area. Examples of some software platforms that allow 3D viewing of molecular structures include Chem3D ([http://www.cambridgesoft.com/Ensemble\\_for\\_Chemistry/ChemDraw/](http://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemDraw/)), Discovery Studio (<http://accelrys.com/products/discovery-studio/>), Sybyl (<http://tripos.com/index.php>), Hyperchem

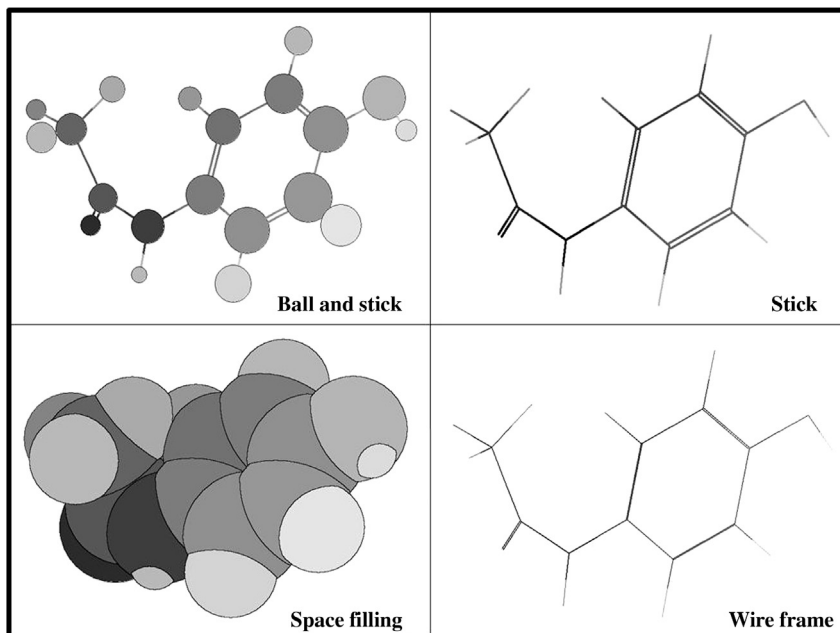


Figure 5.1 Different model representations of the paracetamol molecule.

(<http://www.hyper.com/>), and Maestro (<http://www.schrodinger.com/Maestro/>). However, it may be observed that the visualization plane of the computer is 2D, and hence a 3D molecular view is generated by making some modifications to the axes. This is achieved by making less dense coloring of the atoms/bonds away from the viewer or by reducing their visualization bond length.

### **5.2.1.3 Visualization of ligand–receptor interactions**

Visualization of ligand–receptor interactions is a very useful functionality that provides a 3D graphical presentation of a hypothetical interaction between the ligand and the receptor molecule to the viewer. Since the biological system encountered by a query molecule is entirely 3D in nature, such visualization helps in predicting the nature of molecular interaction at the receptor binding site. Hence, by providing a suitable interface to the user, computers allow the storage of compound information in terms of their structure. Various formats are available to save the molecules, including .cdx, .mol, .skc, and .rxn. Drawing chemical structures is the beginning of any molecular modeling operation, since the drawings are used as input queries to carry out further analyses. It can be observed that the visualization of a chemical structure in a computer window is entirely a part of graphical conversion. The structures are represented by Cartesian/polar coordinates, which are being processed by software algorithms to be portrayed in different forms and shapes. Moreover, the user must understand that any kind of visualization under the molecular modeling paradigm is basically a form of representation for designing and communication purposes only, not a realistic depiction.

## **5.2.2 Calculation and simulation**

Following the drawing/sketching of chemical structures, the next most important part of molecular modeling analysis involves the quantification of chemical information. These include a number of simple to complex theoretical analyses. Simply stated, with the aim of exploring the chemical attributes at the atomic and electronic levels, scientists have developed different theories on molecular environment by combining their knowledge of physics, chemistry, and mathematics. Such theories involve lengthy and intricate molecular mechanical and quantum chemical calculations to characterize the chemical nature at the electronic level. The use of computational programs has enabled such calculations to be done in less time with a good amount of reliability and accuracy. In other words, computational chemistry tools allow users to perform complex analysis just by providing some simple commands and other interfaces. Following the analysis, the derived features can be exploited in several aspects of analysis, like in a chemical characterization project, or as independent variables in QSAR analysis. With the progress of research in chemistry, various other theories of quantitative structural depiction have evolved. By the use of a suitable logical algorithm, computers can easily calculate many such molecular attributes—namely, molecular

topology, geometric features, thermodynamic properties, and others. One more potential application of using computer technology to explore the chemical field is in drug–receptor studies. A number of techniques [namely, molecular docking studies, pharmacophore development, 3D-QSAR like comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), etc.] attempt to find the essential structural requirements necessary for eliciting the biological activity of a ligand in a hypothetical 3D environment. Hence, the calculation of chemical features by employing computer applications can be viewed in the following four divisions:

1. Conformational analysis and energy minimization
2. MM- and quantum mechanics–based calculations
3. Miscellaneous molecular feature determination (charge, electrostatic potential, topological properties, thermodynamic properties, etc.)
4. Exploring drug–receptor interaction studies and structure–activity relationships (docking, pharmacophore development, CoMFA, etc.)

It can be observed that the abovementioned operations are linked to each other; for example, the computation of properties like charge, electrostatic potential, and molar refractivity is done after the energy minimization operation. In this chapter, we shall discuss the different aspects of conformational analysis and energy minimization using molecular and quantum mechanical calculations.

### 5.2.3 Analysis and storage of data

Computers are well known for their ability to process and analyze a large amount of data using a provided mathematical algorithm. Quantification of chemical information and the analysis thereof yield a significant amount of mathematical data, and these operations are properly done using various computational applications. Before the 1960s, the documentation and analyses of the QSAR technique was practiced without the use of computers. However, with the passage of time, the number of chemicals, end points, and descriptors has increased and needed a suitable platform for (i) storage, (ii) processing, and (iii) online availability of data. In order to pursue the analysis of chemical data, several mathematical techniques and suitable statistical measures are combined and used as *in silico* computational tools. Many of these tools allow the development of quantitative mathematical models employing techniques such as multiple linear regression (MLR), partial least squares (PLS), linear discriminant analysis (LDA), and genetic function approximation (GFA), followed by statistical validation; that is, computation of different validation metrics to judge the quality of the developed equation. Hence, the support provided by computers toward the analysis of chemical data can be categorized in the following ways:

1. Processing of data
2. Development of predictive models

3. Numerous statistical analyses
4. Storage of data

Many commercial software packages (e.g., Accelrys software) combine different features of molecular modeling like computation of chemical properties (descriptors), followed by an option for model development and validation, and finally storage of the analyzed data in a specified format. Currently, with the availability of suitable computers, it is possible to carry out virtual screening of library data that contains several thousands to millions of chemicals.

Finally, we would like to comment that by the use of suitable computer programs, it is possible to carry out various molecular modeling operations in a more rational and reliable way, eliminating possible sources of error. Using an efficient computer facility, a large amount of data can easily be processed in a short period of time. However, remember that computers do not and cannot replace any chemical experimentation or synthesis methodology; rather, it promotes molecular modeling analysis by providing a suitable platform to the experimental results toward the design and development of a desired, better chemical entity. Figure 5.2 depicts the privileges provided by various computational tools to facilitate computational chemistry analysis.

While performing complex chemical calculations, such as molecular mechanical, quantum chemical, and semiempirical, remember that the numerical solution provided by any platform cannot be assumed to be the exact or ultimate one, since each of these operations is being executed under several postulates or approximations. Hence, it is wise to decipher the results (i.e., extracted chemical information) with a

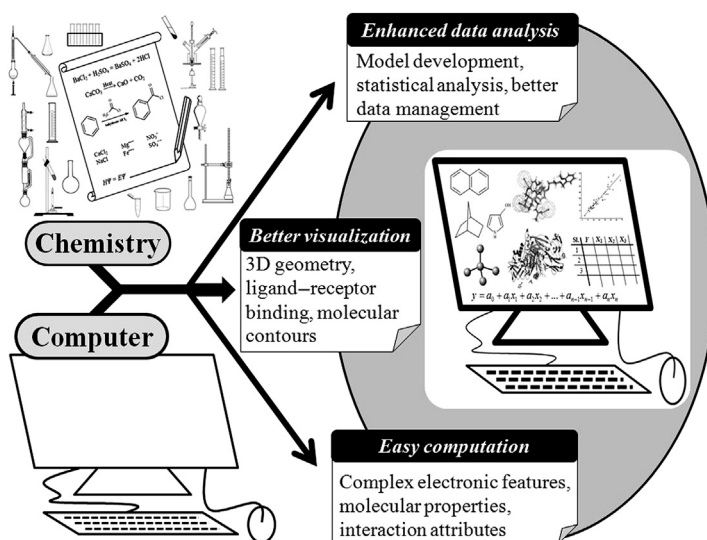


Figure 5.2 Computer technology enhancing the study of chemistry.

consideration of the assumptions involved during the calculations. If the chemist identifies any error in the calculation, that fact can be attributed to the algorithm employed or, more specifically, to the platform used but not to computers, which are blind boxes that process information based on the instruction fed to them.

## 5.3 CONFORMATIONAL ANALYSIS AND ENERGY MINIMIZATION

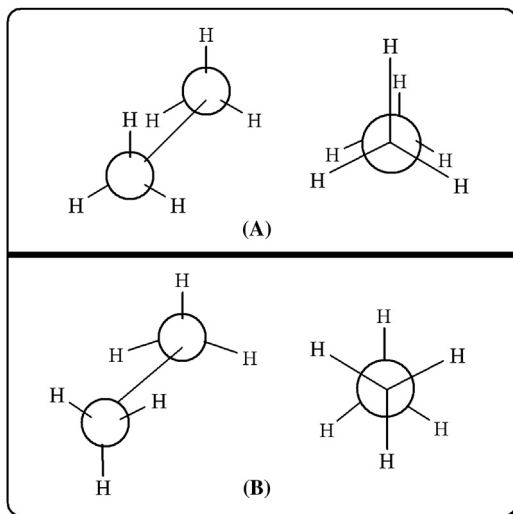
### 5.3.1 The concept

The term *conformation* refers to an arrangement of atoms in a molecule, which is interconvertible by rotation about single bonds. The nature of a covalent bond depends upon the state of orbital hybridization between the participating atoms. Since we know that different types of chemical bonds are characterized by specific energy values, free rotation about the sigma bond ( $\sigma$ ), such as the carbon–carbon single bond, might induce changes in energy of the whole molecule, leaving it in a favorable or unfavorable energy condition. Now, if a molecule undergoes various molecular arrangements in space, questions might be asked regarding the actual structure, which basically contains all the possible conformers. Considering the ethane molecule, for example, the sigma bond ( $\sigma$ ) between the two  $sp^3$ -hybridized carbon atoms is considered to be cylindrically symmetrical about a line joining both nuclei, and the bond strength is also expected to be same in all different arrangements.

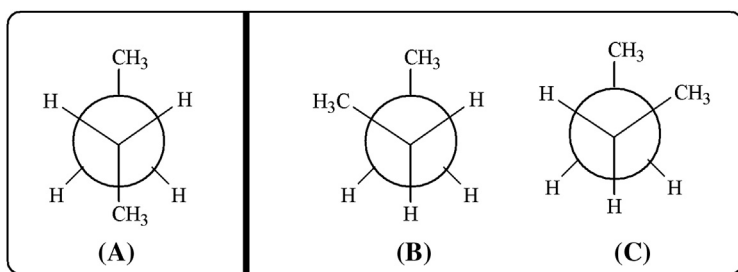
The molecule can perform a free change of conformation from one to another only if the energy of different arrangements is the same or similar. However, considering certain physical properties, it has been observed that the rotation about the carbon–carbon sigma bond is monitored by an energy barrier (about 3 kcal/mole for ethane), giving evidence that the potential energy of a molecule changes in different conformations. One of the most widely used methods for the representation of structural conformers is called *Newman projection*, which is named after its developer, M.S. Newman.

In [Figure 5.3](#), two different conformers of ethane are shown, in which the form *a* is termed *staggered conformation*, while the name of representation *b* is known as *eclipsed conformation*. In between these two, an infinite number of conformations can theoretically exist, which are called *skew conformations*. The staggered conformation is characterized by minimal potential energy (hence more stability for a given molecule); the energy rises following free rotation and reaches a maximal value at the eclipsed conformation. The energy required for conversion from one conformer to another is defined as *torsional energy*, and relative instability of any conformation is depicted in terms of *torsional strain*. It may be noted that a molecule in the eclipsed conformation faces instability owing to other forces besides torsional strain. These include van der Waals force, dipole–dipole interaction, and hydrogen bonding contributed by neighboring atoms or groups (this does not apply in the case of ethane). We would like to





**Figure 5.3** Staggered and eclipsed conformations of an ethane molecule. (A) Staggered conformations; (B) eclipsed conformations.



**Figure 5.4** Staggered and eclipsed conformations of an *n*-butane molecule.

elaborate this point with one additional example, concerning *n*-butane. Unlike ethane, *n*-butane contains two terminal methyl groups attached to each of the carbon atoms number 2 and 3. As a result, *n*-butane is characterized by three staggered conformations (Figure 5.4). Here, the arrangement *i* is termed *anti*, as both the methyl substituents are placed farthest apart, while the rest two conformers are named *gauche*, where the substituent methyl groups are close to each other. The anti conformer has been observed to be more stable because of the absence of the steric van der Waals repulsion caused by the closely placed methyl groups of the *gauche* form [3]. Note that the energy required for interconversion of various conformers in case of ethane or butane is easily supplied from the collisions among the molecules at room temperature, and the individual conformers cannot be isolated.

The behavioral manifestation of chemicals has long been observed to depend upon a suitable molecular arrangement in 3D space; hence conformation is considered to play an important role in modulating the activity/property/toxicity of chemicals. Barton [4] is considered as one of the pioneering contributors to the exploration of conformational analysis about the search for most reactive species. Barton defined conformational analysis as the nonsuperimposable arrangements of the atoms of a molecule in space and is known for depicting the impact of equatorial and axial orientation of substituents in monitoring the reactivity of substituted cyclohexanes. Hence, conformational analysis is aimed in finding out the energetically stable form; that is, the minimum energy structures or conformers of a molecule. The principal objective of conformational analysis is to gather data about the conformational features of flexible bioactive molecules (including drugs) and then assessing the correlation of the conformational flexibility with the activity of the analyzed molecules. Therefore, conformational analysis plays a pivotal role in various computational chemistry operations like molecular docking, library screening, and the optimization and design of lead molecules.

### 5.3.2 Conformational search

Conformational analysis involves a *search* for the identification of suitable molecular conformers that define the actual behavior of the molecule and are present at minimum points at the energy surface. The identification of a low-energy conformer of a molecule can be achieved by employing different search algorithms, which involve systemic variation of torsion angle, stochastic variation of torsion angle, stochastic variation of Cartesian coordinates, stochastic variation of internuclear distances and methods, which use molecular dynamics (MD); and the flipping, flapping, and flexing of rings or mapping of the rings onto generic shapes. [5]. Table 5.1 gives a representative overview of various “search” methods [6] usually employed for the purpose of conformational analysis, while different experimental techniques (noncomputational) used to characterize conformational analysis [7] are briefly presented in Table 5.2.

### 5.3.3 Minimization of energy

Energy minimization is essential to determining the proper molecular arrangement in space since the drawn chemical structures are not energetically favorable. The potential energy of a molecule contains different energy components like stretching, bending, and torsion; hence, when an energy minimization program is run, it will immediately reach a minimum local energy value, and it might stop if the employed program is not exhaustive. In other words, an energy minimization might stop after it finds the first stable conformer that is structurally closest to the starting molecular arrangement. At this point, identified as the *local energy minimum*, structural variation

**Table 5.1** An overview of different conformational search methods

Sl. No.	Name of the search method	Brief description
1	Systematic search	<p><i>Grid search:</i> Considering a dihedral angle to be the dominating parameter for differentiating conformers, conformers are generated by systematically varying the dihedral angle by some increment while keeping bond angle and length fixed, thereby obtaining all combinatorial possibilities of dihedral angles for the molecule. This type of systemic searching of conformers, termed as <i>grid search</i> or <i>grid scan</i>, generates an intractable number of conformers without identifying the unique low-energy local minima on the conformational hypersurface.</p> <p><i>Custom search:</i> Here, specific values are assigned to the torsion angles. Such approach is advantageous if favorable states of torsion angles are known from a previous knowledge (study) so that one can limit the systemic search. Furthermore, this method can operate simultaneous changes in several torsion angles.</p>
2	Model-building method	This method uses molecular fragments or larger building blocks, considering each fragment to be independent of the other. Hence, it is a substructure-based method and is applicable to molecules in which fragments are available.
3	Random approach	Following the generation of an initial structure, a random movement in Cartesian space occurs, leading to minimization. This minimized conformer is added to a list, the operation moves to the next starting structure, and the method stops after obtaining desired structures or completion of sampling of all conformers or the finish of a predefined number of steps.
4	Distance geometry	In this method, a matrix of all pairwise atomic distance values in a molecule is used to form a series of Cartesian coordinates. Standard geometries are used for some of the distances, while the others are gathered from experimental data and random number generation, as provided by the upper- and lower-bound range of the known distance. This method is suitable for the search of both small molecules and macromolecules.

(Continued)

**Table 5.1** (Continued)

Sl. No.	Name of the search method	Brief description
5	Monte Carlo method	Here, simulation of the dynamic behavior of a molecule is done by randomly making changes to the system like rotation of dihedral angles or displacement of atoms. The newly generated atomic configuration is accepted if its energy is less than the previous one. However, in the case of a higher energy value of the conformer, acceptance is made using an algorithmic probability, such as the Metropolis algorithm. The probability is defined by the Boltzmann distribution as follows: $\exp\left(-\frac{\Delta E_i}{kT}\right)$ , where $k$ is the Boltzmann constant and $T$ is the absolute temperature. If a randomly generated number is smaller than the Boltzmann factor, the configuration is accepted.

**Table 5.2** Some representative experimental methods employed for characterizing conformational analysis

Sl. No.	Experimental techniques for conformational analysis	Brief notes
1	Gas-phase electron diffraction (GED)	A powerful old technique. Limitation includes interpretation of the experimental data. GED has been used by the researchers for the conformational analysis of ethane, cyclohexane, chlorocyclohexane, etc.
2	Electronic circular dichroism (ECD)	A chiroptical method. Useful in analyzing conformers that are different in the relative character of the chromophoric parts, giving different ECD spectra.
3	Vibrational circular dichroism (VCD)	Uses the infrared (IR) technique and is more advantageous than ECD since conformational changes are more sensibly reflected by IR bands and it does not necessitate chromophoric groups.
4	Raman optical activity (ROA)	Rarely used. Studies have been carried out on chiral deuterated [ $^2\text{H}_1$ , $^2\text{H}_2$ , $^2\text{H}_3$ ]-neopentane.
5	Dynamic nuclear magnetic resonance (DNMR)	Allows wide signal dispersion and high sensitivity of chemical shifts toward conformational change. Can be operated at different NMR frequencies (e.g., $^1\text{H}$ , $^{19}\text{F}$ , $^{13}\text{C}$ , $^{29}\text{Si}$ , $^{31}\text{P}$ ), as well as in a wide temperature range.
6	Nuclear overhauser effect (NOE)	Proton–proton intramolecular NOE determines the H-atoms in a molecule that are reciprocally close. It has been employed for determining stereochemical features of taxane derivatives.

yields a low change of energy; hence, minimization can stop. However, this may not be (and usually is not) the most stable conformer since the structural minimization stops before an energy barrier. This encumber of energy can be overcome by the use of suitable algorithms, which can increase the strain energy of the structure and finally lead to the most stable conformer, called *global energy minimum*. Hence, the identification of the energy minima (i.e., the potential energy hypersurface of a stable molecule) is crucial to determining its behavior. Molecular modeling operation (namely, MD) allows achieving the most stable conformational stage. Figure 5.5 shows different possible phases of an energy minimization operation. Now, since we are interested in finding the behavior of bioactive molecules, our intention will be finding the *bioactive conformer*. Although the most active conformer seems to be biologically potent, studies have shown that the bioactive conformer might differ from it. However, the bioactive conformer remains in a zone close to the most active conformer. Usually, if the cocrystal geometry of a molecule is present (i.e., geometry a ligand bound to a receptor pocket determined by experimental study like X-ray crystallographic analysis), that conformation of the ligand is considered as the bioactive molecular arrangement or conformation. In the absence of any cocrystal geometric structure, one can consider the most stable conformer as the bioactive conformer. It might be interesting to note that various studies have focused on the determination of the global energy minimum even if the cocrystal geometry of the molecule is present, thereby allowing a comparative assessment of molecular geometry.

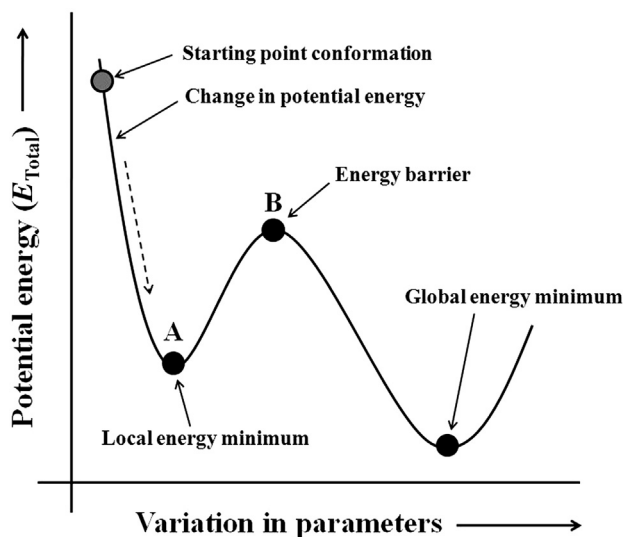
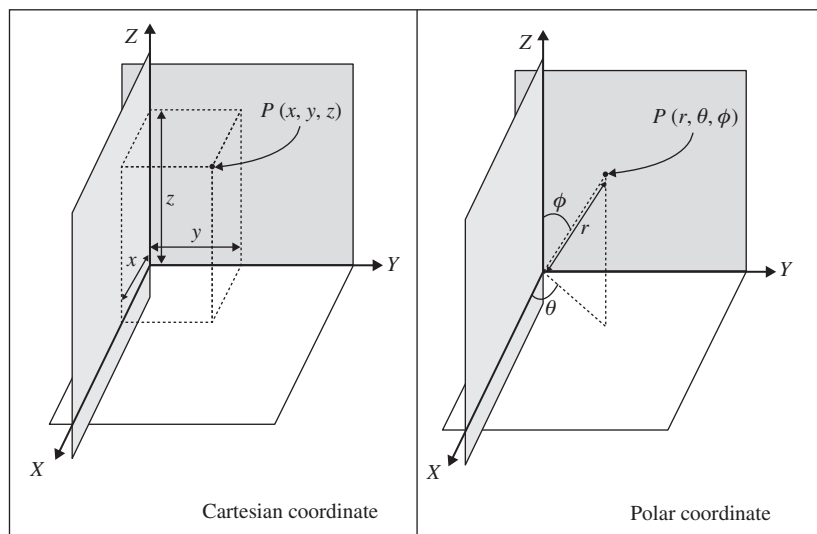


Figure 5.5 Different phases of a molecule during minimization of its energy.

The energy minimization methodology needs to involve identification of the point closest to the starting structure. It might involve a separate algorithmic support for the generation of initial starting structures toward ensuing minimization. It is necessary to understand the difference between conformation searches and other simulation operations like MD and Monte Carlo simulation. Conformational analysis aims to identify minimum energy structures, while simulation operations give an assembly of states that includes structures not at energy minima. However, both MD and Monte Carlo methods can be deployed as part of the conformational search mechanism. Therefore, minimization of energy of 3D structures is crucial to identifying the molecular behavior, but the level of analysis purely depends on the employed algorithm.

Two most important methods facilitating the computational aspects of theoretical chemistry are calculations involving MM and quantum mechanics. The approaches are directed toward the development of energy equations for the total structure of a molecule under investigation. One of the important aspects of these analyses is the position of the atoms in a molecular structure, which is defined by Cartesian or polar coordinates. The initial values of the coordinates can either be set by the modeler or can be obtained from preexisting structural fragments where computer programs set up the coordinates from the program database. Computer programs also can adjust the coordinates if additional fragments are added considering their relative positions. After running a job of MM or quantum mechanics (i.e., the establishment of an energy equation), a final set of coordinates for the minimized structure is calculated by computer. This final coordinate set is converted by using a suitable graphics package for the visualization of the energy-minimized structure. [Figure 5.6](#) presents



**Figure 5.6** Representation of Cartesian and polar coordinates of an arbitrary point.

arbitrary Cartesian and polar coordinates in a 3D plane. Now, it should be noted that although calculations made using computer programs are likely to yield precise results, the conditions or constraints at which computations are made need to be considered. In many cases, the default calculations are based on a molecule defined at 0 K in vacuum instead of considering the actual influences of molecular vibration or the effects of the medium. Quantum mechanical calculations consume considerably more computation power than molecular mechanical calculations [8]. The choice of method depends upon the desire of the modeler and (more often) the available information.

## 5.4 MOLECULAR MECHANICS

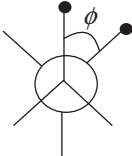
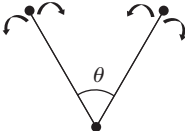
MM assumes that the relative positions of the nuclei of the atoms forming a structure are a function of operating attractive and repulsive forces [9]. Different types of bond stretching, angle bending, torsional energy, and other nonbonded attributes are computed by employing equations of classical physics, giving various interactions and energies also known as *force fields*. Thus, the total potential energy of a molecule is expressed as the sum of all different types of attractive and repulsive forces between the atoms in the structure, considering the impact of nuclei and avoiding the impact of electrons. A hypothetical mechanical model is employed that considers spheres joined by mechanical springs, where the sphere and spring represent the atom and covalent bond, respectively, thereby allowing the application of laws of classical mechanics. A general form of the equation showing components of the total energy can be represented as follows:

$$E_{\text{Total}} = \sum E_{\text{Stretching}} + \sum E_{\text{Bend}} + \sum E_{\text{van der Waals}} + \sum E_{\text{Coulombic}} + \sum E_{\text{Torsion}} \quad (5.1)$$

It is beyond the scope of this chapter to provide detailed discussions on all different energy terminologies. A brief explanation on the force and energy terminologies presented in Eq. (5.1) has been provided in Table 5.3.

The steric energy of the molecules is first computed using force fields, followed by adjustment of the conformation for the minimization of steric energy. One of the methods of calculation in MM employs atom types for the determination of the functions and parameters that comprise the force field. A single element such as a carbon atom can be defined by different MM atom types, the selection of which depends upon various features such as hybridization and chemical environment. Examples of some MM force fields include MM2, MM3, MMFF, Amber, Dreiding, and UFF, all of which are implemented in different software packages [10–19]. Table 5.4 lists the basic features that are implemented in various force fields.

**Table 5.3** Formal definition of different energy terms that define the total energy of a molecule

Phenomenon	Equation	Brief explanation
Torsion	$E_{\text{Torsion}} = \frac{1}{2}k_{\phi}[1 + \cos m(\phi + \phi_{\text{offset}})]$ <p>where <math>\phi_{\text{offset}}</math> is the ideal torsion angle relative to a staggered conformation of two atoms and <math>k_{\phi}</math> represents the energy barrier for rotation about the torsion angle <math>\phi</math>. The periodicity of rotation is denoted by <math>m</math>.</p>	<p>Torsional energy presents the energy required for the free rotation of a sigma bond. Torsion is about the atoms that are separated by three bonds from each other. The torsion angle represents the dihedral angle defining the relative orientation of the atoms. The following figure shows the torsion angle <math>\phi</math> between two sample atoms in a staggered conformation.</p> 
Bond stretching	$E_{\text{Stretching}} = \frac{1}{2}k_{\text{stretch}} \times (r - r_0)^2$ <p>where the ideal and stretched bond lengths are denoted by <math>r_0</math> and <math>r</math>, respectively, and <math>k_{\text{stretch}}</math> is a force constant giving a measure of the strength of the spring; that is, the bond. Hence, a double bond will have a larger value of <math>k_{\text{stretch}}</math> than a single bond.</p>	<p>Considering a covalent bond made up of a spring, Hooke's law can be employed for the computation of bond stretching energy. However, the Morse function containing complex mathematical terms also allows computation of bond stretching.</p>
Angle bending	$E_{\text{Bend}} = \frac{1}{2}k_{\theta} \times (\theta - \theta_0)^2$ <p>where the ideal bond angle is denoted by <math>\theta_0</math> and <math>\theta</math> is the bond angle in the bending position.</p>	<p>The ideal bond angle corresponds to the angle formed by three consecutive atoms at their minimum energy position. Bending angle <math>\theta</math> can be represented as follows, where the arrows show movement of atoms:</p> 



van der Waals force

$$E_{vdW} = \varepsilon \times \left[ \left( \frac{r_{\min}}{r} \right)^{12} - 2 \times \left( \frac{r_{\min}}{r} \right)^6 \right]$$

Here, at minimum energy value  $\varepsilon$ ,  $r_{\min}$  presents the distance between atoms  $i$  and  $j$  while the actual distance between the atoms is  $r$ .

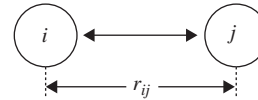
Coulombic force

$$E_{\text{Coulombic}} = \frac{q_i \times q_j}{D \times r_{ij}}$$

where  $q_i$  and  $q_j$  represent the point charges on atoms  $i$  and  $j$ , respectively, with  $r_{ij}$  being the distance between them.  $D$  denotes the dielectric constant of the medium.

The van der Waals force of interaction can be represented by the Lennard–Jones potential equation, where the first term bearing power 6  $\{()\^6\}$  represents forces of attraction, and the term with 12th power  $\{()\^{12}\}$  denotes the short-range repulsive forces involved.

Measures the effect of a charge between two points. The attractive or repulsive interaction between two atoms  $i$  and  $j$  separated by distance  $r_{ij}$  can be represented as



**Table 5.4** Representative examples of various force fields that are employed in MM/dynamics studies

Sl. No.	Name	Characteristics
1	MM2	It is applicable for simple molecular structures containing common functional groups like ketones, ethers, and aromatic compounds. However, MM2 is more applicable to nonheteroatom-containing organic compounds. In order to characterize the real potential function of a chemical bond, MM2 adds terms to the bonded interaction considering anharmonic breakage of bonds. The MM2 force field considers the hybridization pattern and bonding partners in depicting the change in equilibrium bond lengths and angles. One more important fact about the MM2 force field is that it employs the Buckingham equation instead of the Lennard–Jones equation for the computation of van der Waals interactions.
2	MM3	It is a more sophisticated version of the MM2 force field incorporating complex potential functions. It considers several corrections and modifications over the MM2 algorithm, like correction of high rotational barriers in congested hydrocarbons, changes in the van der Waals parameter to circumvent strong H/H nonbonded repulsion when placed at a short distance, torsion–stretch interaction differentiating bond length between eclipsed and staggered conformations, and application of bond dipole moment correction to define crystal packing in benzene.
3	MM4	It performs improved calculation of vibrational frequencies, rotational barriers, etc. for compounds like alkanes and cycloalkanes, excluding small ring systems. It includes special interactions like torsion–bend and the bend–torsion–bend.
4	MMFF	Merck molecular force field (MMFF) is a force field comprising of high-quality data of wide range employed for MM/dynamics simulation operation. MMFF is supposed to present the structures of organic compounds in the Merck index or the Fine Chemicals Directory. This force field comprises several updated versions; however, the basic parameters include attributes of bond stretching, angle bending, stretch–bend interactions, out-of-plane bending at tri-coordinate centers, van der Waals force, torsion, and electrostatic interactions.
5	AMBER	Assisted model building with energy refinement (AMBER) algorithm uses an empirical energy approach, allowing the modeling of small molecules and polymers. AMBER comprises of various subunits, namely, PREP (residue preparation), LINK (residue joining), EDIT (structural modification, change in charges, etc.), PARM (adds parameter), MINM (energy partitioning minimization), ANAL (comparison of rms), etc. for the effective processing of the data.

(Continued)

**Table 5.4** (Continued)

Sl. No.	Name	Characteristics
6	DREIDING	<p>It aims at the use of general force constants and geometry parameters considering the state of hybridization instead of the information derived from combination of atoms. DREIDING uses atomic radii to compute all bond distances, as well as a single force constant, to denote each bond, angle and inversion accompanied by six values for the torsional strain. Atom types defined by a five-character mnemonic label are used as the components of the DREIDING force field, and the potential energy is considered to be summation of valence (<math>E_{val}</math>) and nonbonded (<math>E_{nb}</math>) interaction that depends on atomic distance. The bond stretch is defined either considering harmonic oscillator or using the Morse function.</p>
7	UFF	<p>Universal force field (UFF) focuses on the element, its state of hybridization, and the connectivity possessed by it. UFF allows large amplitude displacements for the functional forms that define angular distortion. Apart from being a molecular mechanical force field, UFF can be employed in an MD energy computation algorithm. Atomic bond radii dependent on the state of hybridization, hybridization angles, parameters defining van der Waals interaction, torsional and inversion barriers, and a set of effective nuclear charges are used as parameters in the UFF formalism.</p>
8	CHARMM	<p>Chemistry at HARvard Macromolecular Mechanics (CHARMM) presents a suitable simulation program allowing a versatile suit application for conformational and path sampling methods, free energy estimates, molecular minimization, MD, analysis techniques, as well as model-building capabilities involving many-particle systems. CHARMM can be employed for the study of biomolecules like peptides, proteins, prosthetic groups, small molecule ligands, nucleic acids, lipids, and carbohydrates. It can be used involving various energy functions and models.</p>
9	OPLS	<p>Optimized potentials for liquid simulations (OPLS) atomic nucleus are appended with interaction site with the exception of <math>CH_n</math> groups that are considered to be united atoms centered on the carbon. Special functions are used to denote H-bonding, and standard combination rules are employed for the Lennard-Jones interaction potential.</p>
10	ECEPP	<p>Empirical conformational energy program for peptides (ECEPP) defines the geometry of amino acid residues and the functions for interatomic interaction by employing a set of internally consistent and standardized parameters. ECEPP is characterized by experimental data and is updated following the development of new data.</p>

The development of a molecular model by using MM force field can be achieved by two means: namely, (i) employment of a commercial force field program in a computer and (ii) use of a database of a molecular modeling program to assemble suitable structural fragments. In the first case, users can select the appropriate molecular mechanical force field from the available packages. In this case, the relevant values of the force field equation are provided as input. Computers calculate an initial value of  $E_{\text{Total}}$  for the model, which undergoes energy minimization, and then a final set of coordinates corresponding to the minimized structure is calculated. Coordinates of this final structure undergo suitable graphical conversion for the visualization of the energy minimized structure.

In the second method, the user collects fragments of desired configuration (hybridization, etc.) from the available database of a suitable molecular modeling program, which are assembled in a form that allows no steric hindrance. Now, the whole structure may not be at the minimized energy conformation that can be processed further to reach the state. It may be noted that during the process of minimization, any molecule is twisted, allowing steric hindrance, and the coordinates are changed accordingly. The graphical packages are designed such that they show the entire process of energy minimization (i.e., the twisting of the molecule in a computer screen), and some of these packages allow the user to record a video of the entire phenomenon.

## 5.5 MOLECULAR DYNAMICS

The foundation of life involves the dynamic evolution of a complex network of chemicals at the molecular level: namely, folding of proteins, nucleic acids, transport of ions through membrane, catalysis of biochemical reactions by enzymes, etc. In order to address the complexity and the dynamic nature of the biological systems, computational simulation methodologies have become increasingly important with the growth and development of powerful *in silico* workstations. The previous techniques discussed so far consider a static molecule during investigation; for example, MM calculations are performed at 0 K considering a frozen molecule. Hence, the natural motion of the atoms in a molecule is not considered when these studies are initiated. Hence, there is a need of a theoretical simulation system that can provide us with a hypothetical dynamic behavior of chemicals and biomolecules.

### 5.5.1 Definition

MD can be defined as a computer simulation technique that permits the prediction of time evolution of an interacting particular system involving the generation of atomic trajectories of a system using numerical integration of Newton's equation of motion for a specific interatomic potential defined by an initial condition and boundary condition. The dynamic simulation also provides information on molecular kinetics and

thermodynamics. The determination of time-dependent motion of individual particles of a system allows quantification of the properties of the given system on a definite time scale that is otherwise unattainable.

### 5.5.2 Development and components

The base concept of MD emerged from the experiments carried out by theoretical physicists. Alder and Wainwright [20] are considered to be the first to study the dynamics of liquid using the “hard-sphere” model where atomic interaction took place through perfect collisions. This study was followed by Rahman’s [21] simulation experiment impersonating the real atomic interactions by the use of smooth and continuous potential. One more important aspect about the development of MD simulation can be attributed to the revolutionary advancements of computational algorithm and technology, which actually allowed the application of MD in several areas of chemistry and physics. From 1970 onward, MD simulation has become a widely practiced simulation method for the study of structure and dynamics of macromolecules: namely, protein and nucleic acids among various research groups.

MD techniques are pursued in two major families to address a physical system considering the nature of the model and the mathematical formalism involved. The approaches include classical mechanics and quantum chemical formalisms, as described here:

1. *Classical mechanics approach*: In this treatment, molecules are considered as classical objects resembling that of the ball-and-stick model, where atoms denote soft balls and the bonds represent elastic sticks. The dynamics of a given system here is judged by the laws of classical mechanics.
2. *Quantum mechanics approach*: This is also termed the *first-principles* MD simulation and originated from the pioneering studies of Car and Parinello, who considered the quantum nature of the chemical bonds. The bonding in a system as defined by the electron density function of the valence electrons is determined employing quantum equations while the dynamics of ions (nuclei with their inner electrons) is subjected to classical treatment. Quantum MD simulations are the necessary enhancement of the classical formalism and they provide valuable information on several biological problems at the cost of consuming more computational resources.

The core requirement for MD simulation is actually simple; it involves a set of conditions defining the initial positions and velocities of all particles and the interaction potential that defines the forces among all the particles. Second, the determination of the evolution of the system in time is done by solving a set of equations of motion for all particles considered in the system. In the case of classical mechanics, Newton’s law is applied to define the motion of classical particles. It may be noted that even a classical MD simulation for biomolecular systems consisting of thousands

of atoms over a nanosecond time scale consumes a significant amount of computational resources [22].

The MD simulation formalism may be considered to be comprised of five conditions: namely, boundary condition, initial condition, force calculation, integrator/ensemble, and property calculation.

### 5.5.3 The algorithm

Considering a classical mechanics approach, the force  $F_i$  acting upon the  $i$ th particle possessing mass  $m_i$  at time  $t$  among a set of interacting particles will be given as follows:

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \quad (5.2)$$

where  $r_i(t)$  is the position vector of the  $i$ th particle and can be represented as  $r_i(t) = \{x_i(t), y_i(t), z_i(t)\}$ . Here, the term *particles* usually refer to atoms though distinct entity; for example, chemical groups can also be represented. Now, Eq. (5.2) is a differential equation of second order that can be integrated by providing specific values of the initial position of particles, their velocities, and the instantaneous force acting on them. The equation of motion is discretized followed by numerical solution because of the many-body system comprised of the particles. The trajectories in MD simulation are defined by position and velocity vector components, and the time evolution of the system is depicted in phase space. The position and velocities components are promulgated with a finite time interval by employing numerical integrators. The examples include the Verlet algorithm. The position of each particle in space is designated by  $r_i(t)$ , while the kinetic energy and temperature of the system is determined by velocity  $v_i(t)$ . The specialty of MD simulation is that it allows a direct tracing of the dynamic events that might be influential to the functional properties of the system.

The integration of Newton's force equation is performed to obtain an expression that gives the position  $r_i(t + \Delta t)$  at time  $t + \Delta t$  in terms of the already-known positions at time  $t$ . By employing the Taylor series, the mentioned position can be mathematically represented as follows [Eq. (5.3)]:

$$r_i(t + \Delta t) \cong 2r_i(t) - r_i(t - \Delta t) + \frac{F_i(t)}{m_i} \Delta t^2 \quad (5.3)$$

The calculation of velocity can be done using the positions or by the use of explicit methods as implemented in systems like alternative *leapfrog* and *velocity* Verlet scheme. It may be noted that an infinitesimally small integration step is obtained by the trajectories, although it is necessary to have larger time steps for sampling longer

trajectories. In reality, determination of  $\Delta t$  involves the fastest motion. For example, the  $\Delta t$  value resides in a sub-femtosecond scale while simulating bonds bearing light atoms, ensuring the stability of the integration. However, coarse-grained simulations use atoms of larger mass, thereby leading to an increased integration time, as well as the trajectory length. One more important aspect of MD simulation is its behavioral nature with statistical mechanics, thereby allowing averaging of values obtained at a microscopic level. The Newtonian dynamics follows the conservation of energy, and MD trajectories give a set of microcanonical ensemble distribution of configurations. This permits the measurement of physical quantities by taking arithmetic average over instantaneous values from the trajectories in a MD simulation job. MD allows simulation of a wide range of experimental conditions; for example, simulation of protein in vacuum, explicit water environment, and crystal environment. Furthermore, the efficiency of MD simulation can be enhanced by the incorporation of improvements in algorithms. Examples of such improvements include RESPA, SHAKE, RATTLE, and LINCS. A larger time step ( $\Delta t$ ) without any significant degradation in the trajectory can be obtained by the use of RESPA coupled with a fixed bond length involving H-atoms with SHAKE, RATTLE, or LINCS. Among other methods, adiabatic mapping is an example of studying motion in proteins.

## 5.6 QUANTUM MECHANICS

With the progress of scientific research, Erwin Schrödinger [23] developed the theory of quantum mechanics in 1926 while studying on the mathematical expression of the motion of electron in terms of its energy. Before going into the details of the quantum mechanical formalism, we would like to present a necessary look back at the electronic picture of atoms and molecules so that readers can have an overview of the essential theories and assumptions involved in formulating electronic models (Box 5.1).

The electron was assumed to depict the property of waves along with being particles and hence the mathematical equations developed by Schrödinger were termed as *wave equations*. A wave equation possesses a series of solutions termed as *wave functions*, each of which depicts a different energy level for the electron. A wave function is designated as a time-dependent state function since it defines the nature as well as the properties of the system. Even for a simple system, solving all of the wave equations is cumbersome. The basic principle of the Schrödinger wave equation can be mathematically depicted as given in Eq. (5.4):

$$H\psi = E\psi \quad (5.4)$$

### BOX 5.1 A Necessary Look at the Assumptions and Theories That Define the Electronic Attributes of Atoms (and Molecules)

Theories and explanations on the fundamental basis of the chemical structures allow the building of a framework of ideas regarding the arrangement of atoms, their order, compatibility, electronic configuration, and possible interaction with neighboring moieties. Electrons are characterized by both the wave and particle nature. However, the wave equations are unable to show the exact position of an electron at a particular moment or the exact velocity at which it is moving; instead, they depict the probability of finding the electron at any particular space.

- *Different hypothesis/models/rules characterizing the nature of electrons*
  - *Thompson*: Proposed that electrons and protons are uniformly mixed throughout an atom and represent a “plum pudding” arrangement.
  - *Rutherford*: Proposed that the nucleus at the center of an atom carries a positive charge and contains the maximum mass, while electrons having minimum mass and negative charge orbit the nucleus at a certain distance. This is also termed the *planetary model*.
  - *Bohr*: Proposed that electrons are moving around the nucleus in specific circular quantized orbits having a definite angular momentum. The angular momentum of electrons is quantized (using Planck’s theory), and the amount of energy depends upon the size of orbit. Electrons can absorb (gain) or release (lose) energy in the form of defined quanta while moving in between orbits.
  - *De Broglie*: Postulated a *wave-particle* duality of matter; that is, particles could have properties of a wave. The de Broglie wavelength was proposed to show that the wavelength is inversely proportional to the momentum of a particle using Planck’s constant.
  - *Heisenberg uncertainty principle*: It is impossible to know the position and the momentum of an electron simultaneously at a given time.
  - *Aufbau principle*: Electrons occupy the lowest-energy orbital available.
  - *Hund’s rule*: The spin of electrons filling orbitals of the same energy level remains parallel (aligned) until the formation of electron pairs. Aligned spins represent more stable forms considering quantum mechanical reasons.
  - *Pauli exclusion principle*: An orbital can only be filled with two electrons possessing opposite spin. Electrons of opposite spin are termed “paired” while those having “like” spin tend to get as far as possible and constitute the basis for shape and other molecular properties.
- *Wave nature of electrons*

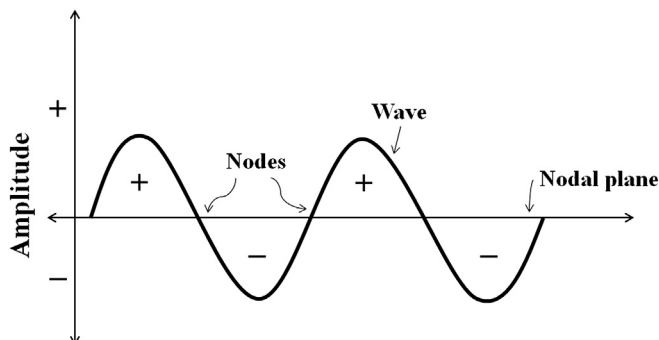
The wave property that describes an electron cloud is very similar to that of a standing or stationary wave. A sample stationary wave generated by the vibration of a string secured at both ends (for example) can be characterized by several segments when observed horizontally along its length as shown in [Figure B5.1](#).

The amplitude of the wave increases in one direction and passes through a maximum, followed by a gradual decrease into zero. Then it again increases in the reverse direction and follows a similar path. The places depicting zero amplitude are termed

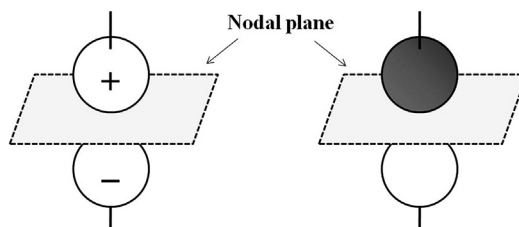
(Continued)



### BOX 5.1 A Necessary Look at the Assumptions and Theories That Define the Electronic Attributes of Atoms (and Molecules)—cont'd



**Figure B5.1** Example of a sample stationary wave generated by the vibration of a string that is secured at both ends.



**Figure B5.2** Representation of a  $p$ -orbital with the lobes being placed above and below the nodal plane.

*nodes*, and the corresponding plane is referred to as the *nodal plane*, while the upward and downward displacements denote opposite wave phases. A wave equation is a differential equation that can be solved in terms of the amplitude  $\psi$  as a function of the distance  $x$  along the wave; that is,  $f(x)$ . The amplitude ( $\psi$ ) for the wave of an electron is represented by a function of three coordinates providing a 3D view of the motion. An electron wave is also characterized by nodes; that is, zero amplitude zone accompanied with upward and downward displacements. In the case of  $p$ -orbitals, these displacements are denoted by lobes (using + and - or shades) being above and below the nodal plane as given in [Figure B5.2](#).

The wave function  $\psi$  represents an MO and is defined by the specific energy value required to move that electron from the molecule. The orbitals are polycentric in nature and considering a normalized solution of  $\psi$ , square of the amplitude ( $\psi^2$ ) corresponds to the probability of finding an electron at any particular point in space. This allows a pictorial visualization of contours of constant probability of finding electron and regions with high and low probability.

- *Orbital*

*Orbital* represents the region in space where an electron is likely to be observed. Different kinds of orbitals ( $s$ ,  $p$ ,  $d$ ,  $f$ ) possess characteristic shapes and sizes and correspond

(Continued)

### BOX 5.1 A Necessary Look at the Assumptions and Theories That Define the Electronic Attributes of Atoms (and Molecules)—cont'd

to specific energy levels according to which the electrons are filled. Theoretically, an orbital cannot be defined by a specific boundary since it has a small probability of finding an electron essentially separated from the atom or may be in some other atom. Nevertheless, such probability is diminished beyond a certain distance from the nucleus. Here,  $1s$  is the spherical orbital of the lowest energy level with the center being placed at the atomic nucleus. Next is the  $2s$  orbital, a sphere with the center being at the atomic nucleus possessing higher energy than  $1s$  (and hence less stability). At the next energy level are three dumbbell-shaped orbitals of equal energy named  $2p$  ( $2p_x$ ,  $2p_y$ , and  $2p_z$ ). Here, the atomic nucleus lies between two lobes, and the atomic nucleus lies between two lobes.

- *Molecular orbital*

MOs represent centered orbitals about the nuclei of the molecule rather the individual nuclei. This theory is exhaustively mathematical and employs less pictorial depiction. In the case of MOs, each pair of electrons remains localized near the two nuclei and their shapes and disposition are related to those of the atomic orbitals of component atoms.

- *Valence bond theory*

The valence bond theory describes a molecular structure as the weighted contribution of numerous possible structures possessing whole numbers of electrons. The depiction of larger molecules with less symmetry using this theory is complex. For instance, for simple molecules, benzene comprises 6 structures, while the number is 42 for naphthalene and 429 for anthracene. Valence bond theory allows a pictorial depiction of the most probable contributions.

- *LCAO approximation*

The linear combination of atomic orbital (LCAO) method employs a linear mathematical relationship of combining atomic orbitals. For a molecule comprising atoms  $A$  and  $B$ , the MO  $\psi$  can be represented as the summed contribution of the atomic orbitals  $\psi_A$  and  $\psi_B$ , respectively, where  $c_a$  and  $c_b$  are the coefficients denoting the weights of the atomic orbitals  $A$  and  $B$ , respectively, as follows:  $\psi = c_a\psi_a + c_b\psi_b$

An MO  $\psi$  is considered to be more stable than the atomic orbitals  $\psi_A$  and  $\psi_B$  if the latter overlap to a considerable extent, possess comparable energy, and have symmetry about the bond axis.

- *Bonding and antibonding orbitals*

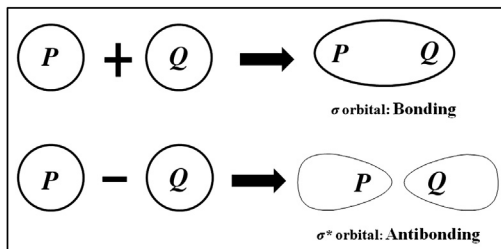
According to the rules of quantum mechanics, a linear combination of two functions yields two combinations instead of one. Hence, combination of two atomic orbitals will give two MO. Bonding orbitals are more stable than the component atomic orbitals, while antibonding orbital are less stable. In other words, a bonding orbital tends to stabilize a molecule, whereas an antibonding orbital tends to destabilize it. Mathematically, they can be represented as follows:

$$\text{Bonding orbital: } \psi_+ = \psi_A + \psi_B$$

$$\text{Antibonding orbital: } \psi_- = \psi_A - \psi_B$$

(Continued)

### BOX 5.1 A Necessary Look at the Assumptions and Theories That Define the Electronic Attributes of Atoms (and Molecules)—cont'd



**Figure B5.3** Two sample *s* orbitals *P* and *Q* forming bonding and antibonding orbitals.

An example of bonding and antibonding orbital formation by two *s* orbitals *P* and *Q* can be represented as shown in [Figure B5.3](#).

where  $\psi$  denotes the time dependent wave function,  $H$  is the Hamiltonian operator, and  $E\psi$  represents the total potential and kinetic energy of all the particles belonging to the molecular structure. Considering the movement in a 3D space bound by the  $x$ -,  $y$ -, and  $z$ -axes, the wave equation is represented by the following differential form:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{8\pi^2 m}{h^2} (E - V)\psi = 0 \quad (5.5)$$

where  $m$  is the mass,  $h$  denotes the Planck's constant term, and  $E$  and  $V$  represent the total and potential energy, respectively. [Equation \(5.5\)](#) can be represented in a shorter form by using a Laplacian operator  $\nabla^2$  for the partial differentials as follows:

$$\nabla^2 \psi + \frac{8\pi^2 m}{h^2} (E - V)\psi = 0 \quad (5.6)$$

However, we shall stick to [Eq. \(5.4\)](#) in order to maintain simplicity. The quantum mechanical calculations are performed employing theories of quantum physics, which account the interaction between nuclei and electrons. The approximations considered during quantum mechanical calculations are provided here:

1. Allowing the fast motion of electrons, nuclei are considered motionless, thereby differentiating nuclear energy from the energy of the electrons.
2. The movement of electrons is assumed to be independent, considering the influence of other electrons and nuclei as average.

The depiction of many-particle systems remains an essential and crucial task for scientists. The exact analytical solution to the Schrödinger equation can be formulated for only a very few simple systems comprising a small number of atoms. The

deployment of the Hamiltonian model, however, provides an option of finding a solution to the equation. The Schrödinger equation actually defines a fundamental relationship of logical coherence to the experimental observations by providing a conceptual scaffold to theoretical chemistry.

The complexity of the molecular structure being analyzed defines the specific mathematical form of  $E$  and  $\psi$ . The Hamiltonian operator  $H$  is defined by interaction terms between (i) electron–electron, (ii) electron–nucleus, and (iii) nucleus–nucleus, accounting for all possible energy components.

Since summation of kinetic and potential energy terms defining the electron and nuclei in a structure constitutes the ideology of the Schrödinger equation, it can be represented as follows:

$$H\psi = (K + U) \times \psi \quad (5.7)$$

Here,  $E$  is defined as the summed contribution of kinetic energy term  $K$  and potential energy  $U$ . The  $H$  term for a simple molecule bearing two electron and two nuclei, such as hydrogen ( $H_2$ ), will be represented by eight different terms, as depicted in Eq. (5.8):

$$H = -\frac{1}{2} \times \overline{V}_1^2 - \frac{1}{2} \times \overline{V}_2^2 + \frac{1}{R_1R_2} - \frac{1}{R_1r_1} - \frac{1}{R_1r_2} - \frac{1}{R_2r_1} - \frac{1}{R_2r_2} + \frac{1}{r_1r_2} \quad (5.8)$$

where  $\frac{1}{2}\overline{V}_1^2$  and  $\frac{1}{2}\overline{V}_2^2$  represent the kinetic energies of electrons 1 and 2, respectively. The positions of two electrons 1 and 2 are represented by  $r_1$  and  $r_2$ , while  $R_1$  and  $R_2$  denote the positions of the nuclei 1 and 2, respectively. Computation of  $H$  becomes complex with the increased number of atoms in a molecule, and it becomes economically less viable for compounds possessing more than 50 atoms.

Different approaches for performing quantum mechanical calculations include the *ab initio* method, density function theory (DFT) technique, and semiempirical analysis. The quantum chemical *ab initio* (i.e., from the beginning) methods aim at providing the absolute solution by employing a convergent approach that gives high-quality, accurate results. However, considering the high operation cost and time consumption, such methods are limited to small molecules. Methods other than *ab initio* avoid many less important terms and attempts to hasten the computational procedure by applying several assumptions. The DFT approach provides a favorable performance, considering the cost, and gives reasonably accurate results for medium-sized molecules, while semiempirical computations are very efficient and applicable to large systems, although the accuracy is hindered owing to integral parameterizations. In the pursuit of finding a suitable solution of the Schrödinger wave equation, scientists have developed different approximations in order to reduce the computational burden. In the following sections, we shall highlight some of these approximations, as well as potential techniques that are widely used in quantum mechanical computations.

### 5.6.1 The Born–Oppenheimer approximation

In order to reduce the computation burden while dealing with various wave functions, the famous Born–Oppenheimer (BO) approximation (named after the contributors Max Born and J. Robert Oppenheimer) is employed to provide a solution to the time-dependent Schrödinger equation [24]. The BO approximation assumes the nucleus to be stationary with respect to the electrons and thereby leaves the kinetic energy of the nuclei out of the relationship, making it simpler. In other words, BO approximation separates the electronic motion and nuclear motion, considering the electronic wave functions to be dependent on the position of nucleus and not its velocity, and a tarnished potential of the speedy electrons is observed by the nuclear motion. Mathematically, the following equation can be written, where  $r_i$  denotes position of the electron and  $R_j$  represents nuclear position:

$$\psi_{\text{molecule}}(r_i, R_j) = \psi_{\text{electrons}}(r_i, R_j) \cdot \psi_{\text{nuclei}}(R_j) \quad (5.9)$$

In order to derive the energy and molecular wave functions employing the Schrödinger equation for the benzene molecule, which comprises 42 electrons and 12 nuclei, a partial differential eigenvalue equation of 162 variables has to be considered. The BO approximation can make the computation process less demanding, so it is important while dealing with quantum chemical problems.

### 5.6.2 The Hartree–Fock approximation

The Hartree–Fock (HF) approximation, also known as the *self-consistent field (SCF)* method, is attributed to the seminal work by Hartree [25] and Fock [26]. Considering an interacting particle system, a many-electron wave function can be defined as  $\psi(r_1, r_2, \dots, r_n)$ , in which  $r_i$  represents the coordinates and spins of the particles. Hartree has provided a useful approximation of a many-electron wave function in terms of the product of single-particle functions, which can be represented as in Eq. (5.10):

$$\psi(r_1, r_2, \dots, r_n) = \phi_1(r_1) \times \phi_2(r_2) \times \dots \times \phi_n(r_n) \quad (5.10)$$

where each function  $\phi_i(r_i)$  corresponds to a one-electron Schrödinger equation with a potential term belonging to the average field of the other electrons. The following equation for the one-particle function  $\phi_i(r_i)$  was proposed by Hartree:

$$\left[ -\frac{1}{2}\Delta + \nu(r) + \sum_{j=1, j \neq i}^N \int \frac{|\phi_j(r')|^2}{|r - r'|} dr' \right] \phi_i(r) = E_i \phi_i(r) \quad (5.11)$$

where  $N$  represents the total number of electrons. The term  $\nu(r)$  is related to a nuclear charge parameter  $Z$  as follows:  $\nu(r) = -Z/r$ . Here, an electron is considered to be

under the SCF at the  $i$ th state, which is determined by all electrons but the  $i$ th one. The two main deviations of the HF approximation from the Schrödinger equation include the presence of nonlinear and nonlocal parameters in the former. Because of the nonorthogonal nature of the functions  $\phi_i(r)$ , the HF equation suffers from the violation of exclusion principle, characterized by the nonorthogonal nature of the functions  $\phi_i(r)$  since the self-consistent potential of the  $i$ th electron depends upon  $i$ . Various modifications and extensions have been made by researchers to eliminate the problems of the HF equation: namely, antisymmetrized modification, Fermi-statistics inclusion, configuration interaction (CI), etc. The HF equation is iteratively solved by employing a suitable computational platform. Investigation of dynamic properties of multielectron objects (such as atoms, molecules, clusters, and fullerene) by computing ground state energy and employing methods such as random phase approximation (RPA) and the random phase approximation with exchange (RPAE) uses HF as a basis [27].

### 5.6.3 Density functional theory

The ideological root of the DFT stems from the hypothesis of Thomas [28] and Fermi [29], who considered the employment of electron density to characterize the many-particle systems. It assumes the electronic motions to be uncorrelated, and the kinetic energy of the electrons is depictable by using a local approximation on the free electrons. The Thomas–Fermi equation presents a primitive approach to the density function-based theories and can be described by the following integral:  $n(r) = N \int dr_2 \cdots \int dr_N \psi^*(r, r_2, \dots, r_N) \times \psi(r, r_2, \dots, r_N)$ , where  $n(r)$  presents density of the electron. Following Thomas–Fermi, various developments in the DFT theory took place through the notable contributions of Dirac, Slater, and Gáspár. It is interesting to note that the explorations in the DFT formalism took place by filling the loopholes of the HF formalism. For example, Gáspár obtained better values of the HF eigenfunctions while studying the  $\text{Cu}^+$  ion. Slater showed that the approximation of exchange potential in a system of variable density can be performed by incorporating a term possessing local dependence ( $[n(r)]^{1/3}$ ) on the density. Such dependence on density relishes an idea of exchange known as *Fermi-hole* representing a region near an electron that is being avoided by the electron of the same spin and not on the exchange potential in a homogeneous system [30]. The simple local density (LD) approximation has been a very useful tool for the study of solids. Hohenberg and Kohn [31] provided the actual theorem for the DFT in 1964, which was later simplified and modified by Levy [32]. The Hamiltonian operator for  $N$  electrons moving in an external potential  $V_{\text{ext}}(r)$  can be represented as follows:

$$H = T + V_{ee} + \sum_{i=1}^N V_{\text{ext}}(r_i) \quad (5.12)$$

where  $T$  and  $V_{ee}$  respectively denote the kinetic and electron–electron interaction operators. Considering  $\psi_{\text{GS}}$  as the wave-function and  $n_{\text{GS}}(r)$  as the density, the ground-state energy  $E_{\text{GS}}$  can be represented as follows:

$$\begin{aligned} E_{\text{GS}} &= \int dr V_{\text{ext}}(r) n_{\text{GS}}(r) + \langle \psi_{\text{GS}} | T + V_{ee} | \psi_{\text{GS}} \rangle \\ &= \int dr V_{\text{ext}}(r) n_{\text{GS}}(r) + F[n_{\text{GS}}] \end{aligned} \quad (5.13)$$

where  $V_{\text{ext}}(r)$  is the external potential and the term  $F[n]$  represents a density that is functionally independent of any specific system or the external potential. It is interesting to observe that the Kohn–Sham theory presents another famous derivation in the realm of DFT for solving the Schrödinger equation for a fictitious system of noninteracting particles. Kohn and Sham depicted the application of an LD approximation to the limiting case of a slowly varying density using an exchange and correlation energy term (Eq. (5.14)):

$$E_{\text{xc}}^{\text{LD}} = \int dr n(r) \varepsilon_{\text{xc}}[n(r)] \quad (5.14)$$

where  $\varepsilon_{\text{xc}}[n]$  denotes the exchange and correlation energy per particle of a homogeneous electron gas characterized by density  $n$ . It was observed that use of the Kohn–Sham theory for the HF-like calculation of finite electron system leads to a ground energy value smaller than the actual HF method. The drawbacks of the HF theory are actually solved in the DFT formalism; for example, the nonlocality of single particle exchange potential in HF is overcome by LD approximation in the KS theory.

Many molecular attributes (namely, vibrational frequencies, atomization energies, ionization energies, electric and magnetic properties, reaction paths, etc.) are computable using the DFT.

## 5.6.4 Semiempirical analysis

### 5.6.4.1 Concept

Technically the term *semiempirical* refers to methods or techniques that employ assumptions, generalizations, or approximation in order to simplify complex calculations. Semiempirical analysis of the quantum chemical methods uses integral approximations and parameterizations to simplify large calculations of solving the Schrödinger wave equation. Because of the incorporation of several assumptions, the results obtained from semiempirical analysis can be characterized as less accurate, although it attempts to provide a realistic strategy for dealing with large molecules. The semiempirical methods begin with the *ab initio* formalism, followed by assumptive avoidance of several less important terms for the sake of speeding up the calculations. However, empirical parameters are utilized in the formalism with calibration against reliable theoretical or experimental data as a measure of compensation toward the errors of assumptions.

Some *ab initio* and DFT methods employ a number of empirical assumptions, and hence can be categorized into semiempirical analysis.

#### 5.6.4.2 Developmental background

The parameterization of the MO-based valence electron methods are widely employed to address the semiempirical analysis with the aim of enhancing the accuracy of the *ab initio* HF results by the use of a minimal basis set. The primitive semiempirical approach to address electronic structure of chemical compounds includes the Hückel MO (HMO) method [33], which involves a  $\pi$ -electronic formalism for the generation of MO values of unsaturated molecules using a connectivity matrix. A citable enhancement of this method is Hoffman's [34] extended Hückel theory, which uses all valence electrons for computation. These methods make a significant contribution toward the development of qualitative MO theory, which accounts for orbital interactions. It should be noted here that Hückel methods are noniterative in nature, involving one-electron integrals. The Pariser–Parr–Pople [35] formalism describes the electronic spectra of unsaturated molecules that use antisymmetrized products of quantitative atomic orbital integrals possessing the core Hamiltonian and introduces an approximation of zero differential overlap, along with an optional, uniformly charged sphere depiction of atomic orbitals. This theory also allows the incorporation of  $\sigma$ -electron adjustment to the  $\pi$ -electronic distribution. Pople et al. [36] did bring in a hierarchy of integral approximations satisfying various consistency criteria, including the rotational invariance. They showed that the results obtained by neglecting a differential overlap in electron interaction integral without further adjustments are not constant to simple transformation of the atomic orbital basis set, such as the *s*, *p* orbital replacement by hybrids or the rotation of axes. This study led to the development of two schemes that are invariant to transformation among atomic orbitals: namely, the complete neglect of differential overlap (CNDO) and the neglect of diatomic differential overlap (NDDO). Modern semiempirical analyses largely employ the formalism of NDDO and INDO (intermediate neglect of differential overlap) in order to get successful computational results. Exploration of this differential overlap concept has led to the development of a number of schemes, the most frequently employed of which are discussed next, while Table 5.5 gives an overview of different such models.

#### 5.6.4.3 Modified neglect of diatomic overlap

The modified neglect of diatomic overlay (MDNO) approach was proposed and developed by Dewar and Thiel [37] and is based on the NDDO algorithm for the parameterization of one-center, two-electron integrals from the spectroscopic data for isolated atoms. It is aimed at the estimation of other two-electron integrals using the formalism of multipole–multipole interactions from classical electrostatics. Classical MNDO models employ *s* and *p* orbitals as basis sets, while *d* orbitals are added in the



**Table 5.5** A representative view of different schemes which have been implemented in semiempirical/self-consistent quantum chemical calculations

Sl. No.	Abbreviated name	Full form of the formalism	Notes
1	LCAOSCF	LCAO self-consistent function	It provides self-consistent function approximation using the LCAO method. Here energy minimization is facilitated by the coefficient of the orbitals. Application is limited due to computational difficulty.
2	CNDO	Complete neglect of differential overlap	CNDO and NDDO represent the simplification of LCAOSCF by employing the approximation of neglecting differential overlap. CNDO does not consider any differential overlap in all the basis sets. Here, a product of two different atomic orbitals corresponding to a specific electron is always "neglected" in electron interaction integrals.
3	NDDO	Neglect of diatomic differential overlap	This corresponds to the product of pairs of atomic orbitals of different atoms that have been neglected in certain electron repulsion integral. For a specific electron, the product of atomic orbitals will be neglected if they are on separate centers.
4	INDO	Intermediate neglect of differential overlap	This corresponds to the neglect of the differential overlap in the integral of all electron interaction except those using one center only; that is, the retention of a one-center product of different atomic orbitals in only one-center integral. It presents an intermediate complexity between the CNDO and NDDO methods.
5	MINDO	Modified intermediate neglect of differential overlap	This algorithm considers a common value in order to represent the two-center electron repulsion integral between the atomic orbitals of a chosen atomic pair.

(Continued)

**Table 5.5** (Continued)

Sl. No.	Abbreviated name	Full form of the formalism	Notes
6	MNDO	Modified neglect of diatomic overlap	It gives better results in depicting the heat of formation of hydrocarbons, as well as radicals. Here, the approximation has been applied to the closed-shell molecules and their valence electrons, which are assumed to move in a constant core-field composed of the nuclei and inner shell electrons. It improves the computational results obtained from MINDO/3.

latest MNDO/d method for the depiction of hypervalent sulfur species and transition metals. The drawbacks of the MNDO formalism include its inability in describing H-bonding caused by strong intermolecular repulsion and poor consistency in predicting heats of formation. By applying the MNDO method, instability in prediction is shown for highly substituted stereoisomers with respect to the linear isomers, which can be attributed to the overestimation of repulsive forces in a sterically crowded system.

#### 5.6.4.4 Austin model 1

Dewar et al. [38] developed the parametric Austin model 1 (AM1) as an approximation of the NDDO algorithm. Unlike MNDO, the two-electron integrals have been approximated here by using a modified version of the nuclear–nuclear core repulsion function (CRF) that mimics the van der Waals interactions as a nonphysical attraction force. This modification reparameterized the model by instituting changes in dipole moments, ionization potentials, and molecular geometries. The problem of reproducing hydrogen bonds in the MNDO scheme has been overcome in the AM1 method without any increase in computing time. Other advantages include the improvement of computation of some properties, such as the heats of formation with respect to the MNDO method. The disadvantages of the AM1 method include systemic overestimation of basicities and incorrect prediction of the lowest-energy geometry of water dimer.

#### 5.6.4.5 Parametric method 3

The parametric method 3 (PM3) has been developed by Stewart [39] by using a similar Hamiltonian operator like that of AM1, but a separate parameterization strategy.

The PM3 formalism is parameterized in reproducing a large number of molecular properties,, unlike the AM1 strategy, which uses a relatively small number of atomic data. Hydrogen bonds are well assessed in PM3 method due to its specific parameterization protocol and nuclear repulsion treatment. However, the nonphysical hydrogen–hydrogen attraction forces are sometimes expressed, leading to trouble while computing intermolecular interactions. For example, methane is falsely predicted as a strongly bound dimer; and the determination of conformers of flexible molecules such as a hydroxyl group of 1-pentanol is strongly attracted to the methyl group. The PM3 method has a wider application for the computation of electronic attributes and yields more accurate thermochemical data than AM1, and the recent extended versions allow the inclusion of transition metals.

#### 5.6.4.6 PDDG/PM3 and PDDG/MNDO

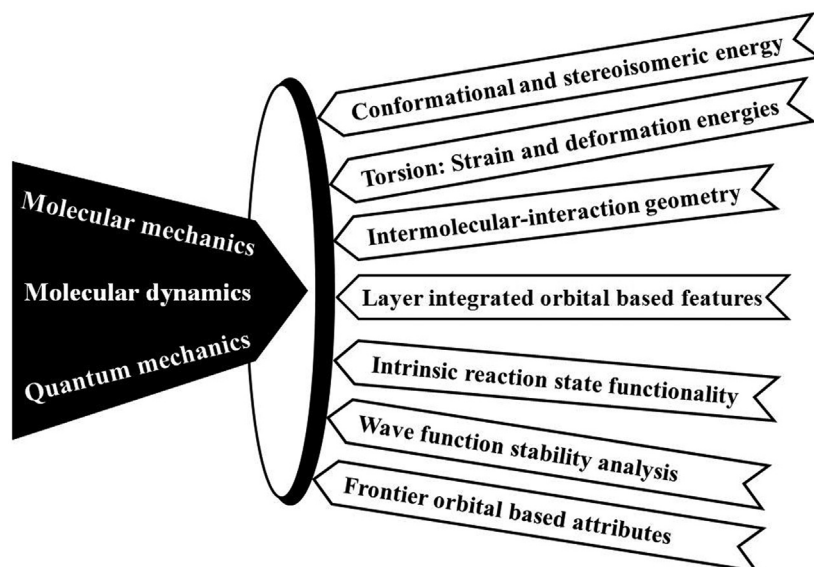
William Jorgensen and coworkers [40] developed two new formalisms in the realm of semiempirical quantum chemical calculations by using a pairwise distance directed Gaussian (PDDG) modification of the existing MNDO and PM3 methods. The PDDG/PM3 and PDDG/MNDO methods use reparameterized functional group-specific enhancement of the CRF and improve the accuracy of the previous NDDO methods (namely, PM5, PM3, AM1, and MNDO formalisms). Incorporation of the PDDG modification has led to improved computation of van der Waals attraction between atoms, accurate estimation of heat of formation values, and trustworthy calculation of intermolecular complexes, overcoming relative stability of hydrocarbon isomers, energetics of small rings and molecules containing multiple heteroatoms, and other issues. Improvement of the internal consistency of PDDG isomerization energy enjoys better results than the B3LYP/6-31G\* method, a hybrid level (density functional) in which Gaussian computations are carried out depending upon the usage.

## 5.7 OVERVIEW AND CONCLUSION

The application of computer technology is a very significant and crucial element contributing to the exploration of theoretical chemistry. Computers have provided the necessary momentum to pursue studies in theoretical chemistry at a potentially higher level. Moving from visualization to computation is easily done using suitable *in silico* platforms. Moreover, the promising features of MD simulation were derivable only after the development of higher-capacity computer systems. Today, several commercial and open-access molecular modeling software/packages provide users with encouraging, user-friendly interfaces for carrying out modeling operations. Most of these packages offer tutorials describing the operational characteristics of the platform in order to run a specific molecular modeling job (computation or visualization), thereby making the operations easier. A user can perform various such analyses using the

default specified values without using so-called expert knowledge. However, it is always helpful to gather some theoretical knowledge behind a major operation. It is also obvious that the basis of several computational chemistry operations stem from the depth of classical physics, mechanics, knowledge on mathematics, biology, and, of course, chemistry. Hence, instead of giving any cumbersome description of theory, we have attempted to provide some basic information defining various molecular modeling operations so that readers can get an overview of the process constraints. Figure 5.7 presents different facets of operations attained by molecular modeling algorithms involving MM, MD, and quantum mechanics.

Molecular mechanical methods are not concerned with the properties and distribution of electrons. Quantum mechanical methods should be used for computation of electron density at various atoms and energies of the highest occupied MO (HOMO) and lowest unoccupied MO (LUMO) and for understanding of the possible orientation of the transition state geometries during reaction. These calculations also provide information on possible reaction pathway and thermodynamic data like heat of formation. The choice of using a suitable molecular modeling technique depends upon the nature of the chemicals and also on the objective of the analysis. Although *ab initio* models provide a theoretical possibility of obtaining most accurate results, they are practiced far less owing to the unfeasible computation power required. In spite of the limitations and accuracy problems, semiempirical methods are quite often used to address the computation of quantum mechanical electronic attributes of chemicals



**Figure 5.7** Different molecular attributes determined by MM, MD, and quantum chemical calculations.

possessing a relatively large structure, which are impractical to model using more accurate methods. Actually, large systems are either subjected to MM operation or semiempirical analysis for the structural optimization, as well as computation of conformational energy.

Molecular mechanical methods are comparatively faster and generate accurate (or nearly accurate) molecular geometries and conformational energies. Sometimes the prediction of thermochemical data of stable chemical species is obtained with appreciable reliability employing the MM3 or MM4 method. On the other hand, semiempirical analysis becomes valuable while addressing systems like reactive intermediates or transition states owing to the absence of any suitable force field program. Furthermore, a compromise is made between choosing semiempirical and *ab initio* methods while dealing with small molecules. We would like to add that the selection of a suitable technique also depends on the knowledge of the modeler regarding the nature of chemicals employed—for example, semiempirical methods elicit good results for chemical systems that are similar to those used in the parameterization set. Similarly, care should be taken where semiempirical analysis is prone to failure like the prediction of activation barriers. Finally, we would like to say that it is better to study a little regarding the pros and cons and applicability of a particular energy optimization scheme instead of placing blind faith in it; for example, the semiempirical method may not be applicable to a particular large chemical system, although they are known to produce good results in many such other systems. In this context, we would like to mention that among various molecular modeling platforms, Gaussian software (<http://www.gaussian.com/>) is one of the oldest, and it allows meticulous calculation involving *ab initio* formalism (HF, MP2, etc.), density functional theory (HFB, PW91, PBE, G96, LYP, VWN5, etc.), semiempirical techniques (AM1, MNDO, PM3, PM6, etc.), MM (Amber, Dreiding, UFF), and other hybrid methods (G1, G2, G2MP2, G3, G3B3, G4, G4MP2, MPW1PW91, B2PLYP, B3LYP, etc.). Gaussian software also allows the use of various set of functions in the form of basis sets (namely, STO-3G, 3-21G, 6-21G, 4-31G, 6-31G, etc.) to characterize the wave-function using the approximations obtained from different research outcomes. The development of this chemometric program is credited to John Pople and his research group, which released the first version, Gaussian70, in 1970. This software has undergone continual updates since then, and Gaussian 09 is the latest one [41].

One last issue to be discussed here is the fact that the study of molecules using suitable computer platforms and programs provides the user with only a hypothetical overview of the actual phenomenon involved. Hence, we prefer to use the term *simulation* since it is not a real-time situation and in fact may not represent the reality if all the constraints are not addressed properly. Instinctively, it may be observed that with the increased complexity of the system, less accurate approximations become foreseeable. Consideration of all real-time constraints, or even most of them, has yet not

been feasible for large systems. However, computers facilitate considerably less drastic approximations, thereby allowing an association of experimental and theoretical insights that give a reasonable, realistic representation of problems.

## REFERENCES

- [1] Lewars EG. Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics. Heidelberg, Germany: Springer; 2003.
- [2] Corey RB, Pauling L. Molecular models of amino acids, peptides, and proteins. *Rev Sci Instrum* 1975;24:621–7.
- [3] Morrison RT, Boyd RN. Organic chemistry. Upper Saddle River, NJ: Prentice-Hall Inc; 2002.
- [4] Barton DHR. Principles of conformational analysis. Nobel lectures, chemistry 1963–1970. Amsterdam, the Netherlands: Elsevier Publishing Company; 1972.
- [5] Kolossvary I, Guida WC. Low mode search. An Efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. *J Am Chem Soc* 1996;118:5011–19.
- [6] Leech A. In: Lipkowitz KB, Boyd DB, editors. Reviews in computational chemistry, vol. 2. New York, NY: VCH; 1991.
- [7] Mazzanti A, Casarini D. Recent trends in conformational analysis. *WIREs Comput Mol Sci* 2012;2:613–41.
- [8] Golebiewski A, Parczewski A. Theoretical conformational analysis of organic molecules. *Chem Rev* 1974;74:519–30.
- [9] Hendrickson JB. Molecular geometry. I. Machine computation of the common rings. *J Am Chem Soc* 1961;83:4537–47.
- [10] Allinger NL. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing  $V_1$  and  $V_2$  torsional terms. *J Am Chem Soc* 1977;99:8127–34.
- [11] Allinger NL, Yuh YH, Lii J-H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J Am Chem Soc* 1989;111:8551–66.
- [12] Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 1996;17:490–519.
- [13] Allinger NL, Chen K, Lii J-H. An improved force field (MM4) for saturated hydrocarbons. *J Comput Chem* 1996;17:642–68.
- [14] Mayo SL, Olafson BD, Goddard III WA. DREIDING: a generic force field for molecular simulations. *J Phys Chem* 1990;94:8897–909.
- [15] Rappe AK, Casewit CJ, Colwell KS, Goddard III WA, Skiff WM. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc* 1992;114:10024–35.
- [16] Weiner PK, Kollman PA. AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J Comput Chem* 1981;2:287–303.
- [17] Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 1988;110:1657–66.
- [18] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 2010;31:671–90.
- [19] Nemethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, et al. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J Phys Chem* 1992;96:6472–84.
- [20] Alder BJ, Wainwright TE. Phase transition for a hard sphere system. *J Chem Phys* 1957;27:1208–9.
- [21] Rahman A. Correlations in the motion of atoms in liquid argon. *Phys Rev* 1964;136:A405–11.

- [22] Petrenko R, Jaroszaw M. Molecular dynamics. Encyclopedia of life sciences (ELS). Chichester: John Wiley & Sons; 2010.
- [23] Schrödinger E. Quantisierung als eigenwertproblem (erste Mitteilung). *Ann Phys* 1926;79:361–76.
- [24] Born M, Oppenheimer R. Zur quantentheorie der molekeln. *Ann Phys* 1927;389:457–84.
- [25] Hartree DR. The wave mechanics of an atom with a non-Coulomb central field. I. Theory and methods. *Proc Cambridge Philos Soc* 1928;24:89–110.
- [26] Fock V. Näherungsmethode zur Lösung des Quantenmechanischen Mehrkörperproblems. *Z Phys* 1930; 61:126–48.
- [27] Amusia MY. Atomic photoeffect. New York, NY: Plenum Press; 1990.
- [28] Thomas LH. The calculation of atomic fields. *Proc Cambridge Philos Soc* 1927;23:542–8.
- [29] Fermi E. *Z Phys* 1928;48:73–9.
- [30] Jones RO, Gunnarsson O. The density functional formalism, its applications and prospects. *Rev Mod Phys* 1989;61:689–746.
- [31] Hohenberg P, Kohn W. Inhomogeneous Electron Gas. *Phys Rev* 1964;136:B864.
- [32] Levy M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the  $v$ -representability problem. *Proc Natl Acad Sci (USA)* 1979;76:6062–5.
- [33] Hückel E. Quantum contributions to the benzene problem. *Z Phys* 1931;70:204–86.
- [34] Hoffmann R. An extended Hückel theory. I. Hydrocarbons. *J Chem Phys* 1963;39:1397–412.
- [35] Pariser R, Parr RG. A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. *J Chem Phys* 1953;21:466–71.
- [36] Pople JA, Santry DP, Segal GA. Approximate self consistent molecular orbital theory. I. Invariant procedures. *J Chem Phys* 1965;43:S129–35.
- [37] Dewar MJS, Thiel W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J Am Chem Soc* 1977;99:4899–907.
- [38] Dewar MJS, Zoebisch E, Healy EF, Stewart JJP. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J Am Chem Soc* 1985;107:3902–9.
- [39] Stewart JJP. Optimization of parameters for semiempirical methods I. Method. *J Comput Chem* 1989;10:209–20.
- [40] Repasky MP, Chandrasekhar J, Jorgensen WL. PDDG/PM3 and PDDG/MNDO: improved semi-empirical methods. *J Comput Chem* 2002;23:1601–22.
- [41] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09, Revision D.01. Wallingford, CT: Gaussian, Inc.; 2009.