PRECLINICAL DEVELOPMENT HANDBOOK

ADME and Biopharmaceutical Properties

SHAYNE COX GAD, PH.D., D.A.B.T.
Gad Consulting Services
Cary, North Carolina

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

PRECLINICAL DEVELOPMENT HANDBOOK

ADME and Biopharmaceutical Properties

PRECLINICAL DEVELOPMENT HANDBOOK

ADME and Biopharmaceutical Properties

SHAYNE COX GAD, PH.D., D.A.B.T.
Gad Consulting Services
Cary, North Carolina

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data is available.

ISBN: 978-0-470-24847-8

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTRIBUTORS

Adegoke Adeniji, Philadelphia College of Pharmacy, University of the Sciences in Philadelphia, Philadelphia, Pennsylvania, *Chemical and Physical Characterizations of Potential New Chemical Entity*

Adeboye Adjare, Philadelphia College of Pharmacy, University of the Sciences in Philadelphia, Philadelphia, Pennsylvania, *Chemical and Physical Characterizations of Potential New Chemical Entity*

Zvia Agur, Institute for Medical Biomathematics (IMBM), Bene-Ataroth, Israel; Optimata Ltd., Ramat-Gan, Israel, *Mathematical Modeling as a New Approach for Improving the Efficacy/Toxicity Profile of Drugs: The Thrombocytopenia Case Study*

Melvin E. Andersen, The Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina, *Physiologically Based Pharmacokinetic Modeling*

Joseph P. Balthasar, University of Buffalo, The State University of New York, Buffalo, New York, *Pharmacodynamics*

Stelvio M. Bandiera, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada, *Cytochrome P450 Enzyme*

Ihor Bekersky, Consultant, Antioch, Illinois, *Bioavailability and Bioequivalence Studies*

Marival Bermejo, University of Valencia, Valencia, Spain, *How and Where Are Drugs Absorbed*?

Jan H. Beumer, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, *Mass Balance Studies*

Prasad V. Bharatam, National Institute of Pharamaceutical Education and Research (NIPER), Nagar, India, *Modeling and Informatics in Drug Design*

Deepa Bisht, National JALMA Institute for Leprosy and Other Myobacterial Diseases, Agra, India, *Accumulation of Drugs in Tissues*

Scott L. Childs, SSCI, Inc., West Lafayette, Indiana, *Salt and Cocrystal Form Selection*

Harvel J. Clewell III, The Hamner Institutes for Health Sciences, Research Triangle Park, North Carolina, *Physiologically Based Pharmacokinetic Modeling*

Brett A. Cowans, SSCI, Inc., West Lafayette, Indiana, *Salt and Cocrystal Form Selection*

Dipankar Das, University of Alberta, Edmonton, Alberta, Canada, *Protein–Protein Interactions*

A. G. de Boer, University of Leiden, Leiden, The Netherlands, *The Blood–Brain Barrier and Its Effect on Absorption and Distribution*

Pankaj B. Desai, University of Cincinnati Medical Center, Cincinnati, Ohio, *Data Analysis*

Merrill J. Egorin, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, *Mass Balance Studies*

Julie L. Eiseman, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, *Mass Balance Studies*

Moran Elishmereni, Institute for Medical Biomathematics (IMBM), Bene-Ataroth, Israel, *Mathematical Modeling as a New Approach for Improving the Efficacy/Toxicity Profile of Drugs: The Thrombocytopenia Case Study*

Dora Farkas, Tufts University School of Medicine, Boston, Massachusetts, *Mechanisms and Consequences of Drug–Drug Interactions*

Sandrea M. Francis, National Institute of Pharamaceutical Education and Research (NIPER), Nagar, India, *Modeling and Informatics in Drug Design*

Shayne Cox Gad, Gad Consulting Services, Cary, North Carolina, *Regulatory Requirements for INDs/FIH (First in Human) Studies*

P. J. Gaillard, to-BBB Technologies BV, Leiden, The Netherlands, *The Blood–Brain Barrier and Its Effect on Absorption and Distribution*

Srinivas Ganta, University of Auckland, Auckland, New Zealand, *Permeability Assessment*

Sanjay Garg, University of Auckland, Auckland, New Zealand, *Permeability Assessment*

Isabel Gonzalez-Alvarez, University of Valencia, Valencia, Spain, *How and Where Are Drugs Absorbed*?

Eric M. Gorman, The University of Kansas, Lawrence, Kansas, *Stability: Physical and Chemical*

Luis Granero, University of Valencia, Valencia, Spain, *Absorption of Drugs after Oral Administration*

David J. Greenblatt, Tufts University School of Medicine, Boston, Massachusetts, *Mechanisms and Consequences of Drug–Drug Interactions*

Ken Grime, AstraZeneca R&D Charnwood, Loughborough, United Kingdom, *Utilization of* In Vitro *Cytochrome P450 Inhibition Data for Projecting Clinical Drug–Drug Interactions*

William L. Hayton, The Ohio State University, Columbus, Ohio, *Allometric Scaling*

William W. Hope, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, *Experimental Design Considerations in Pharmacokinetics Studies*

Eugene G. Hrycay, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada, *Cytochrome P450 Enzymes*

Teh-Min Hu, National Defense Medical Center, Taipei, Taiwan, *Allometric Scaling*

Subheet Jain, Punjabi University, Patiala, Punjab, India, *Dissolution*

Izet M. Kapetanovic, NIH NCI Division of Cancer Prevention, Chemoprevention Agent Devleopment Research Group, Bethesda, Maryland, *Analytical Chemistry Methods: Developments and Validation*

Kamaljit Kaur, University of Alberta, Edmonton, Alberta, Canada, *Protein–Protein Interactions*

Jane R. Kenny, AstraZeneca R&D Charnwood, Loughborough, United Kingdom, *Utilization of* In Vitro *Cytochrome P450 Inhibition Data for Projecting Clinical Drug–Drug Interactions*

Masood Khan, Covance Laboratories Inc., Immunochemistry Services, Chantilly, Virginia, *Method Development for Preclinical Bioanalytical Support*

Smriti Khanna, National Institute of Pharamaceutical Education and Research (NIPER), Nagar, India, *Modeling and Informatics in Drug Design*

Yuri Kheifetz, Institute for Medical Biomathematics (IMBM), Bene-Ataroth, Israel, *Mathematical Modeling as a New Approach for Improving the Efficacy/Toxicity Profile of Drugs: The Thrombocytopenia Case Study*

Yuri Kogan, Institute for Medical Biomathematics (IMBM), Bene-Ataroth, Israel, *Mathematical Modeling as a New Approach for Improving the Efficacy/Toxicity Profile of Drugs: The Thrombocytopenia Case Study*

Niels Krebsfaenger, Schwarz Biosciences, Monheim, Germany, *Species Comparison of Metabolism in Microsomes and Hepatocytes*

Thierry Lave, F. Hoffman-LaRoche Ltd., Basel, Switzerland, *Physiologically Based Pharmacokinetic Modeling*

Albert P. Li, In Vitro ADMET Laboratories, Inc., Columbia, Maryland, In Vitro *Evaluation of Metabolic Drug–Drug Interactions: Scientific Concepts and Practical Considerations*

Charles W. Locuson, University of Minnesota, Minneapolis, Minnesota, *Metabolism Kinetics*

Alexander V. Lyubimov, University of Illinois at Chicago, Chicago, Illinois, *Analytical Chemistry Methods: Developments and Validation; Dosage Formulation; Bioavailability and Bioequivalence Studies*

Dermot F. McGinnity, AstraZeneca R&D Charnwood, Loughborough, United Kingdom, *Utilization of* In Vitro *Cytochrome P450 Inhibition Data for Projecting Clinical Drug–Drug Interactions*

Peter Meek, University of the Sciences in Philadelphia, Philadelphia, Pennsylvania, *Computer Techniques: Identifying Similarities Between Small Molecules*

Donald W. Miller, University of Manitoba, Winnipeg, Manitoba, Canada, *Transporter Interactions in the ADME Pathway of Drugs*

Mehran F. Moghaddam, Celegne, San Diego, California, *Metabolite Profiling and Structural Identification*

Guillermo Moyna, University of the Sciences in Philadelphia, Philadelphia, Pennsylvania, *Computer Techniques: Identifying Similarities Between Small Molecules*

Eric J. Munson, The University of Kansas, Lawrence, Kansas, *Stability: Physical and Chemical*

Ann W. Newman, SSCI, Inc., West Lafayette, Indiana, *Salt and Cocrystal Form Selection*

Mohammad Owais, Aligarh Muslim University, Aligarh, India, *Accumulation of Drugs in Tissues*

Brian E. Padden, Schering-Plough Research Institute, Summit, New Jersey, *Stability: Physical and Chemical*

Sree D. Panuganti, Purdue University, West Lafayette, Indiana, *Drug Clearance*

Jayanth Panyam, University of Minnesota, Minneapolis, Minnesota, *Distribution: Movement of Drugs through the Body*

Yogesh Patil, Wayne State University, Detroit, Michigan, *Distribution: Movement of Drugs through the Body*

James W. Paxton, The University of Auckland, Auckland, New Zealand, *Interrelationship between Pharmacokinetics and Metabolism*

Olavi Pelkonen, University of Oulu, Oulu, Finland, In Vitro *Metabolism in Preclinical Drug Development*

Vidmantas Petraitis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, *Experimental Design Considerations in Pharmacokinetics Studies*

Ana Polache, University of Valencia, Valencia, Spain, *Absorption of Drugs after Oral Administration*

Elizabeth R. Rayburn, University of Alabama at Birmingham, Birmingham, Alabama, *Linkage Between Toxicology of Drugs and Metabolism*

Micaela B. Reddy, Roche Palo Alto LLC, Palo Alto, California, *Physiologically Based Pharmacokinetic Modeling*

Robert J. Riley, AstraZeneca R&D Charnwood, Loughborough, United Kingdom, *Utilization of* In Vitro *Cytochrome P450 Inhibition Data for Projecting Clinical Drug–Drug Interactions*

Sevim Rollas, Marmara University, Istanbul, Turkey, In Vivo *Metabolism in Preclinical Drug Development*

Bharti Sapra, Punjabi University, Patiala, Punjab, India, *Dissolution*

Richard I. Shader, Tufts University School of Medicine, Boston, Massachusetts, *Mechanisms and Consequences of Drug–Drug Interactions*

Dhaval Shah, University of Buffalo, The State University of New York, Buffalo, New York, *Pharmacodynamics*

Puneet Sharma, University of Auckland, Auckland, New Zealand, *Permeability Assessment*

Beom Soo Shin, University of Buffalo, The State University of New York, Buffalo, New York, *Pharmacodynamics*

Meir Shoham, Optimata Ltd., Ramat-Gan, Israel, *Mathematical Modeling as a New Approach for Improving the Efficacy/Toxicity Profile of Drugs: The Thrombocytopenia Case Study*

Mavanur R. Suresh, University of Alberta, Edmonton, Alberta, Canada, *Protein–Protein Interactions*

Craig K. Svensson, Osmetech Molecular Diagnostics, Pasadena, California, *Drug Clearance*

A. K. Tiwary, Punjabi University, Patiala, Punjab, India, *Dissolution*

Ari Tolonen, Novamass Analytical Ltd., Oulu, Finland; University of Oulu, Oulu, Finland, In Vitro *Metabolism in Preclinical Drug Development*

Timothy S. Tracy, University of Minnesota, Minneapolis, Minnesota, *Metabolism Kinetics*

Miia Turpeinen, University of Oulu, Oulu, Finland, In Vitro *Metabolism in Preclinical Drug Development*

Jouko Uusitalo, Novamass Analytical Ltd., Oulu, Finland, In Vitro *Metabolism in Preclinical Drug Development*

Vladimir Vainstein, Institute for Medical Biomathematics (IMBM), Bene-Ataroth, Israel; Optimata Ltd., Ramat-Gan, Israel, *Mathematical Modeling as a New Approach for Improving the Efficacy/Toxicity Profile of Drugs: The Thrombocytopenia Case Study*

Krishnamurthy Venkatesan, National JALMA Institute for Leprosy and Other Myobacterial Diseases, Agra, India, *Accumulation of Drugs in Tissues*

Lisa L. von Moltke, Tufts University School of Medicine, Boston, Massachusetts, *Mechanisms and Consequences of Drug–Drug Interactions*

Jayesh Vora, PRTM Management Consultants, Mountain View, California, *Data Analysis*

Thomas J. Walsh, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, *Experimental Design Considerations in Pharmacokinetics Studies*

Naidong Weng, Johnson & Johnson Pharmaceutical Research & Development, Bioanalytical Department, Raritan, New Jersey, *Method Development for Preclinical Bioanalytical Support*

Randy Zauhar, University of the Sciences in Philadelphia, Philadelphia, Pennsylvania, *Computer Techniques: Identifying Similarities Between Small Molecules*

Ruiwen Zhang, University of Alabama at Birmingham, Birmingham, Alabama, *Linkage Between Toxicology of Drugs and Metabolism*

Yan Zhang, Drug Metabolism and Biopharmaceutics, Incyte Corporation, Wilmington, Delaware, *Transporter Interactions in the ADME Pathway of Drugs*

Irit Ziv, Optimata Ltd., Ramat-Gan, Israel, *Mathematical Modeling as a New Approach for Improving the Efficacy/Toxicity Profile of Drugs: The Thrombocytopenia Case Study*

CONTENTS

PREFACE

This *Preclinical Development Handbook: ADME and Biopharmaceutical Properties* continues and extends the objective behind the entire *Handbook* series: an attempt to achieve a through overview of the current and leading-edge nonclinical approaches to evaluating the pharmacokinetic and pharmacodynamic aspects of new molecular entity development for therapeutics. The 38 chapters cover the full range of approaches to understanding how new molecules are absorbed and distributed in model systems, have their biologic effects, and then are metabolized and excreted. Such evaluations provide the fundamental basis for making decisions as to the possibility and means of pursuing clinical development of such moieties. Better performance in this aspect of the new drug development process is one of the essential keys to both shortening and increasing the chance of success in developing new drugs.

The volume is unique in that it seeks to cover the entire range of available approaches to understanding the performance of a new molecular entity in as broad a manner as possible while not limiting itself to a superficial overview. Thanks to the persistent efforts of Mindy Myers and Gladys Mok, these 38 chapters, which are written by leading practitioners in each of these areas, provide coverage of the primary approaches to the problems of understanding the mechanisms that operate in *in vivo* systems to transfer a drug to its site of action and out.

I hope that this newest addition to our scientific banquet is satisfying and useful to all those practitioners working in or entering the field.

1

MODELING AND INFORMATICS IN DRUG DESIGN

Prasad V. Bharatam,* Smriti Khanna, and Sandrea M. Francis
National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar, India

Contents

*Corresponding author.

Preclinical Development Handbook: ADME and Biopharmaceutical Properties,
edited by Shayne Cox Gad

1.1 INTRODUCTION

Modeling and informatics have become indispensable components of rational drug design (Fig. 1.1). For the last few years, chemical analysis through molecular modeling has been very prominent in computer-aided drug design (CADD). But currently modeling and informatics are contributing in tandem toward CADD. Modeling in drug design has two facets: modeling on the basis of knowledge of the drugs/leads/ligands often referred to as ligand-based design and modeling based on the structure of macromolecules often referred to as receptor-based modeling (or structure-based modeling). Computer-aided drug design is a topic of medicinal chemistry, and before venturing into this exercise one must employ computational chemistry methods to understand the properties of chemical species, on the one hand, and employ computational biology techniques to understand the properties of biomolecules on the other. Information technology is playing a major role in decision making in pharmaceutical sciences. Storage, retrieval, and analysis of data of chemicals/biochemicals of therapeutic interest are major components of pharmacoinformatics. Quite

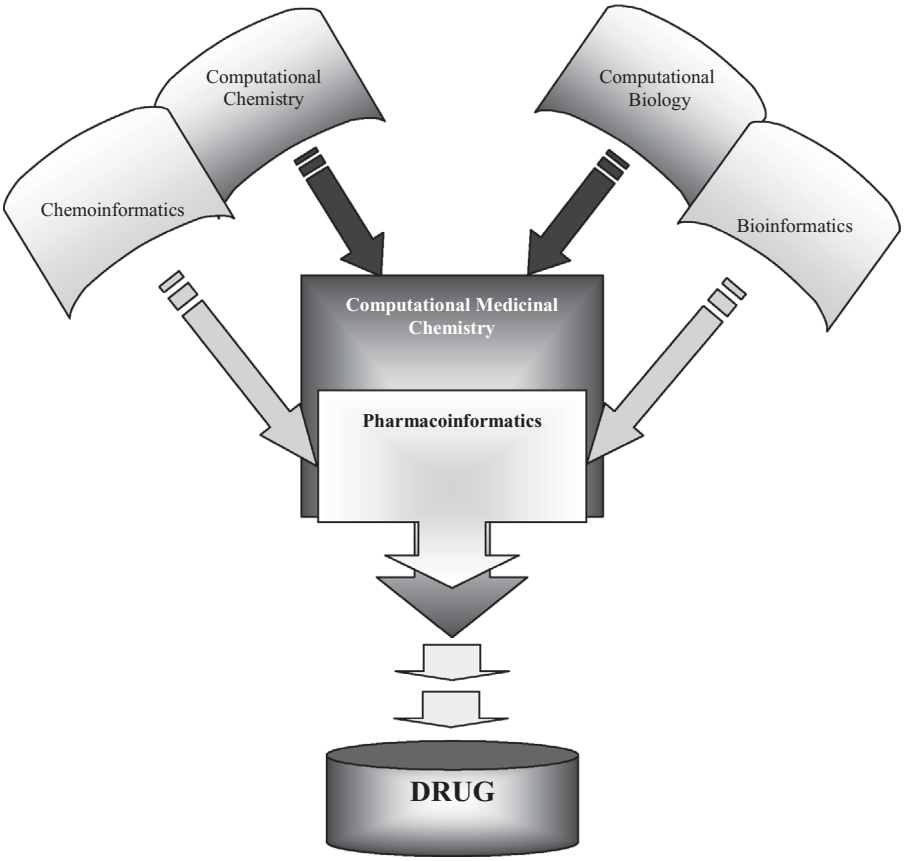

FIGURE 1.1 A schematic diagram showing a flowchart of activities in computer aided drug development. The figure shows that the contributions from modeling methods and informatics methods toward the drug development are parallel and in fact not really distinguishable.

often, the efforts based on modeling and informatics get thoroughly integrated with each other, as in the case of virtual screening exercises. In this chapter, the molecular modeling methods that are in vogue in the fields of (1) computational chemistry, (2) computational biology, (3) computational medicinal chemistry, and (4) pharmacoinformatics are presented and the resources available in these fields are discussed.

1.2 COMPUTATIONAL CHEMISTRY

Two-dimensional (2D) structure drawing and three-dimensional (3D) structure building are the important primary steps in computational chemistry for which several molecular visualization packages are available. The most popular of these are ChemDraw Ultra and Chem3D Pro, which are a part of the ChemOffice suite of software packages [1]. ACD/ChemSketch [2], MolSuite [3], and many more of this kind are other programs for the same purpose. Refinement has to be carried out on all the drawings and 3D structures so as to improve the chemical accuracy of the structure on the computer screen. Structure refinement based on heuristic rules/cleanup procedures is a part of all these software packages. However, chemical accuracy of the 3D structures still remains poor even after cleanup. Further refinement can be carried out by performing energy minimization using either molecular mechanical or quantum chemical procedures. By using these methods, the energy of a molecule can be estimated in any given state. Following this, with the help of first and second derivatives of energy, it can be ascertained whether the given computational state of the molecules belongs to a chemically acceptable state or not. During this process, the molecular geometry gets modified to a more appropriate, chemically meaningful state – the entire procedure is known as geometry optimization. The geometry optimized 3D structure is suitable for property estimation, descriptor calculation, conformational analysis, and finally for drug design exercise [4–6].

1.2.1 Ab Initio Quantum Chemical Methods

Every molecule possesses internal energy (U), for the estimation of which quantum chemical calculations are suitable. Quantum chemical calculations involve rigorous mathematical derivations and attempt to solve the Schrödinger equation, which in its simplest form may be written as

$$H\Psi = E\Psi \tag{1.1}$$

$$\hat{H}_{\mathrm{el}} = \sum\left(-\frac{1}{2}\nabla_i^2\right) - \sum_i\sum_a \frac{Z_a}{|r_i - d_a|} + \frac{1}{2}\sum_i\sum_{j\neq i}\frac{1}{|r_i - r_j|} + \frac{1}{2}\sum_a\sum_{b\neq a}\frac{Z_a Z_b}{|d_a - d_b|} \tag{1.2}$$

where ψ represents the wavefunction, E represents energy, ∇ represents the kinetic energy operator for electrons, r_i defines the vector position of electron i with vector components in Bohr radii, Z_a is the charge of fixed nucleus a in units of the elementary charge, and d_a is the vector position of nucleus a with vector components in Bohr radii.

Exact solutions to Schrödinger equation cannot be provided for systems with more than one electron. Several *ab initio* molecular orbital (MO) and *ab initio* density functional theory (DFT) methods were developed to provide expectation value for the energy. This energy can be minimized and thus the geometry of any molecule can be obtained, with high confidence level, using quantum chemical methods. During this energy estimation, the wavefunctions of every molecule can be defined, which possess all the information related to the molecule. Thus, properties like relative energies, dipole moments, electron density distribution, charge distribution, electron delocalization, molecular orbital energies, molecular orbital shapes, ionization potential, infrared (IR) frequencies, and chemical shifts can be estimated using *ab initio* computational chemistry methods. For this purpose, several quantum chemical methods like Hartree–Fock (HF), second order Moller–Plesset perturbation (MP2), coupled cluster (CCSD), configuration interaction (QCISD), many-body perturbation (MBPT), multiconfiguration self-consistent field (MCSCF), complete active space self-consistent field (CASSCF), B3LYP, and VWN were developed. At the same time to define the wavefunction, a set of mathematical functions known as *basis set* is required. Typical basis sets are 3-21G, 6-31G*, and 6-31+G*. Combination of the *ab initio* methods and basis sets leads to several thousand options for estimating energy. For reliable geometry optimization of drug molecules, the HF/6-31+G*, MP2/6-31+G*, and B3LYP/6-31+G* methods are quite suitable. When very accurate energy estimation is required, G2MP2 and CBS-Q methods can be employed. Gaussian03, Spartan, and Jaguar are software packages that can be used to estimate reliable geometry optimization and very accurate energy estimation of any chemical species. In practice, quantum chemical methods are being used to estimate the relative stabilities of molecules, to calculate properties of reaction intermediates, to investigate the mechanisms of chemical reactions, to predict the aromaticity of compounds, and to analyze spectral properties. Medicinal chemists are beginning to take benefit from these by studying drug–receptor interactions, enzyme–substrate binding, and solvation of biological molecules. Molecular electrostatic potentials, which can be derived from *ab initio* quantum chemical methods, provide the surface properties of drugs and receptors and thus they offer useful information regarding complementarities between the two [7–10].

1.2.2 Semiempirical Methods

The above defined *ab initio* methods are quite time consuming and become prohibitively expensive when the drugs possess large number of atoms and/or a series of calculations need to be performed to understand the chemical phenomena. Semiempirical quantum chemical methods were introduced precisely to address this problem. In these methods empirical parameters are employed to estimate many integrals but only a few key integrals are solved explicitly. Although these calculations do not provide energy of molecules, they are quite reliable in estimating the heats of formation. Semiempirical quantum chemical methods (e.g., AM1, PM3, SAM1) are very fast in qualitatively estimating the chemical properties that are of interest to a drug discovery scientist. MOPAC and AMPAC are the software packages of choice; however, many other software packages also incorporate these methods. Qualitative estimates of HOMO and LUMO energies, shapes of molecular orbitals, and reaction mechanisms of drug synthesis are some of the applications of

semiempirical analysis [5, 10]. When the molecules become much larger, especially in the case of macromolecules like proteins, enzymes, and nucleic acids, employing these semiempirical methods becomes impractical. In such cases, molecular mechanical methods can be used to estimate the heats of formation and to perform geometry optimization.

1.2.3 Molecular Mechanical Methods

Molecular mechanical methods estimate the energy of any drug by adding up the strain in all the bonds, angles, and torsions due to the energy of the van der Waals and Coulombic interactions across all atoms in the molecule. It reflects the internal energy of the molecule; although the estimated value is nowhere close to the actual internal energy, the relative energy obtained from these methods is indicative enough for chemical/biochemical analysis. It is made up of a number of components as given by

$$E_{\text{mm}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{vdw}} + E_{\text{torsion}} + E_{\text{charge}} + E_{\text{misc}} \tag{1.3}$$

Molecular mechanical methods are also known as force field methods because in these methods, the electronic effects are estimated implicitly in terms of force fields associated with the atoms. In Eq. 1.3, the energy (E) due to bonds, angles, and torsional angles can be estimated using the simple Hooke's law and its variations, whereas the van der Waals (vdw) interactions are estimated using the Lennard-Jones potential and the electrostatic interactions are estimated using Coulombic forces. The energy estimation, energy minimization, and geometry optimization using these methods are quite fast and hence suitable for studying the geometries and conformations of biomolecules and drug–receptor interactions. Since these methods are empirical in nature, parameterization of the force fields with the help of available spectral data or quantum chemical methods is required. AMBER, CHARMM, UFF, and Tripos are some of the force fields in wide use in computer-aided drug development [5, 10].

1.2.4 Energy Minimization and Geometry Optimization

Drug molecules prefer to adopt equilibrium geometry in nature, that is, a geometry that possesses a stable 3D arrangement of atoms in the molecule. The 3D structure of a molecule built using a 3D builder does not represent a natural state; slight modifications are required to be made on the built 3D structure so that it represents the natural state. For this purpose, the following questions need to be addressed: (1) Which minimal changes need to be made? (2) How much change needs to be made? (3) How does one know the representation at hand is the true representation of the natural state? To provide answers to these questions we can depend on energy, because molecules prefer to exist in thermodynamically stable states. This implies that if the energy of any molecule can be minimized, the molecule is not in a stable state and thus the current representation of the molecule may not be the true representation of the natural state. This also implies that we can minimize the energy and the molecular structure in that energy minimum state probably represents a true natural state. Several methods of energy minimization have been developed by

computational chemists, some of which are nonderivative methods (simplex method) but many of which are dependent on derivative methods (steepest descent, Newton–Raphson, conjugate gradient, variable metrics, etc.) and involve the estimation of the gradient of the potential energy curve [4–6]. The entire procedure of geometry modification to reach an energy minimum state with almost null gradient is known as geometry optimization in terms of the structure of the molecule and energy minimization in terms of the energy of the molecule. All computational chemistry software packages are equipped with energy minimization methods—of which a few incorporate energy minimization based on *ab initio* methods while most include the semiempirical and molecular mechanics based energy minimization methods.

1.2.5 Conformational Analysis

Molecules containing freely rotatable bonds can adopt many different conformations. Energy minimization procedures lead the molecular structure to only one of the chemically favorable conformations, called the local minimum. Out of the several local minima on the potential energy (PE) surface of a molecule, the lowest energy conformation is known as the global minimum. It is important to note all the possible conformations of any molecule and identify the global minimum before taking up a drug design exercise (Fig. 1.2). This is important because only one of the possible conformations of a drug, known as its bioactive conformation, is responsible for its therapeutic effect. This conformation may be a global minimum, a local minimum, or a transition state between local minima. As it is very difficult to identify

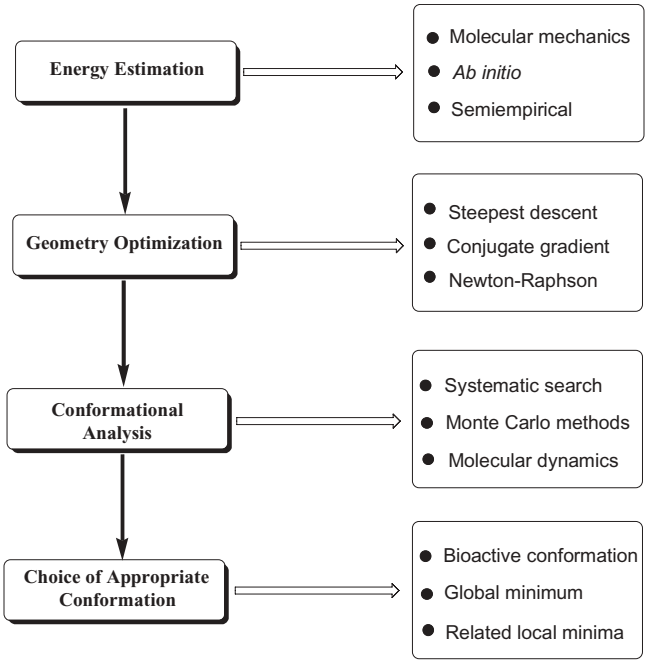

FIGURE 1.2 Flowchart showing the sequence of steps during molecular modeling of drug molecules.

the bioactive conformation of many drug molecules, it is common practice to assume the global minima to be bioactive. The transformation of drug molecules from one conformer to another can be achieved by changing the torsional angles. The computational process of identifying all local minima of a drug molecule, identifying the global minimum conformation, and, if possible, identifying the bioactive conformation is known as conformational analysis. This is one of the major activities in computational chemistry.

Manual conformational search is one method where the chemical intuition of the chemist plays a major role in performing the conformational analysis. Here, a chemist/modeler carefully chooses all possible conformations of a given drug molecule and estimates the energy of each conformation after performing energy minimization. This procedure is very effective and is being widely used. This approach allows the application of rigorous quantum chemical methods for the conformational analysis. The only limitation of this method arises from the ability and patience of the chemist. There is a possibility that a couple of important conformations are ignored in this approach. To avoid such problems, automated conformational analysis methods were introduced.

Various automated methods of conformational analysis include systematic search, random search, Monte Carlo simulations, molecular dynamics, genetic algorithms, and expert systems (Table 1.1) [4, 5]. The systematic conformational search can be performed by varying systematically each of the torsion angles of the rotatable bonds of a molecule to generate all possible conformations. The step size for torsion angle change is normally 30–60°. The number of conformations across a C—C single bond would vary between 6 and 12. With an increase in the number of rotatable bonds, the total number of conformations generated becomes quite large. The "bump check" method reduces the number of possibilities; still, the total number of conformations generated can be in the tens of thousands for drug molecules. Obviously, most of the conformations are chemically nonsignificant.

The random conformational search method employs random change in torsional angle across rotatable bonds and performs energy minimization each time; thus, a handful of chemically meaningful conformations can be generated [11–18].

Molecular dynamics is another method of carrying out conformational search of flexible molecules. The aim of this approach is to reproduce time-dependent motional behavior of a molecule, which can identify bound states out of several possible

TABLE 1.1 Different Methods of Conformational Analysis

Methods for Conformational Analysis	Remarks
Systematic search	Systematic change of torsions
Random search	Conformations picked up randomly
Monte Carlo method	Supervised random search
Molecular dynamics	Newtonian forces on atoms and time dependency incorporated in conformational search
Genetic algorithm	Parent–child relationship along with survival of the fittest techniques employed
Expert system	Heuristic methods based on rules and facts employed

states. The user needs to define step size, time of run, and the temperature supplied to the system at the beginning of the computational analysis. A simulated annealing method allows "cooling down" of the system at regular time intervals by decreasing the simulation temperature. As the temperature approaches 0 K, the molecule is trapped in the nearest local minimum. It is used as the starting point for further simulation and the cycle is repeated several times [19].

1.3 COMPUTATIONAL BIOLOGY

Computational biology is a fast growing topic and it is really not practical to distinguish this topic from bioinformatics. However, we may broadly distinguish between the two topics as far as this chapter is concerned. Molecular modeling aspects of computational biology, which lead to structure prediction, may be discussed under this heading, whereas the sequence analysis part, which leads to target identification, may be discussed under the section of pharmacoinformatics. Structure prediction of biomolecules (often referred to as "structural bioinformatics") adopts many aspects of computational chemistry. For example, energy minimization of protein receptor structure is one important step in computational biology. Molecular mechanics, molecular simulations, and molecular dynamics are employed in performing conformational analysis of macromolecules.

A rational drug design approach is very much dependent on the knowledge of receptor protein structures and is severely limited by the availability of target protein structure with experimentally determined 3D coordinates. Proteins exhibit four tiered organization: (1) primary structure defining the amino acid sequence, (2) secondary structure with α-helical and β-sheet folds, (3) tertiary structure defining the folding of secondary structure held by hydrogen bonds, and (4) quaternary structure involving noncovalent association between two or more independent proteins. Methods for identifying the primary amino acid sequence in proteins are now well developed; however, this knowledge is not sufficient enough to understand the function of the proteins, the drug–receptor mutual recognition, and designing drugs. Various experimental techniques like X-ray crystallography, nuclear magnetic resonance, and electron diffraction are available for determining the 3D coordinates of the protein structure; however, there are many limitations. It is not easy to crystallize proteins and even when we succeed, the crystal structure represents only a rigid state of the protein rather than a dynamic state. Thus, the reliability of the experimental data is not very high in biomolecules. Computational methods provide the alternative approach—although with equal uncertainty but at a greater speed. Homology modeling and *ab initio* methods are being employed to elucidate the tertiary structure of various biomolecules. The 3D structures of proteins are useful in performing molecular docking, *de novo* design, and receptor-based pharmacophore mappings. The computational methods of biomolecular structure prediction are discussed next (Fig. 1.3) [20, 21].

1.3.1 Ab Initio Structure Prediction

This approach seeks to predict the native conformation of a protein from the amino acid sequence alone. The predictions made are based on fundamental understanding of the protein structure and the predictions must satisfy the requirements of free-

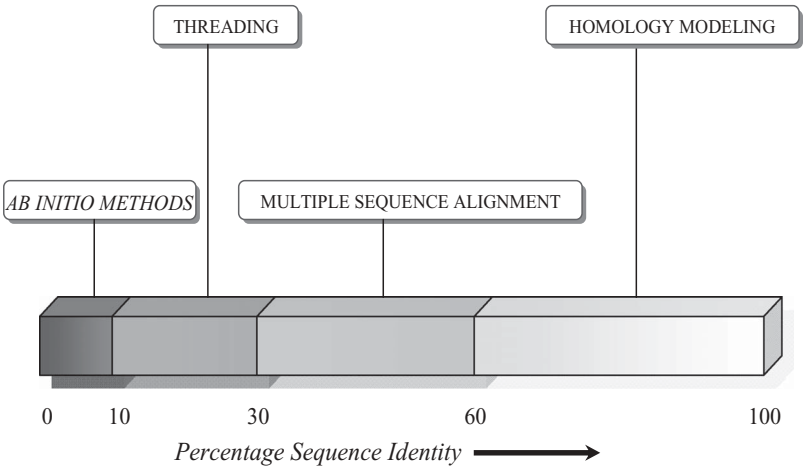

FIGURE 1.3 A list of computer-aided structure prediction methods with respect to their suitability to the available sequence similarity.

energy function associated with lowest free-energy minima. The detailed representation of macromolecules should include the coordinates of all atoms of the protein and the surrounding solvent molecules. However, representing this large number of atoms and the interactions between them is computationally expensive. Thus, several simplifications have been suggested in the representations during the *ab initio* structure prediction process. These include (1) representation of side chains using a limited set of conformations that are found to be prevalent in structures from the Protein Data Bank (PDB) without any great loss in predictive ability [22] and (2) restriction of the conformations available to the polypeptides in terms of phi–psi (ϕ–ψ) angle pairs [23]. Building the protein 3D structure is initiated by predicting the structures of protein fragments. Local structures of the protein fragments are generated first after considering several alternatives through energy minimization. A list of possible conformations is also extracted from experimental structures for all residues. Protein tertiary structures are assembled by searching through the combinations of these short fragments. During the assembling process, bump checking and low energy features (hydrophobic, van der Waals forces) should be incorporated. The final suggested structure is subjected to energy minimization and conformational analysis using molecular dynamics simulations. *Ab initio* structure prediction can be used to guide target selection by considering the fold of biological significance. The *ab initio* macromolecular structure prediction methods, if successful, are superior to the widely used homology modeling technique because no *a priori* bias is incorporated into the structure prediction [24].

1.3.2 Homology Modeling

Homology or comparative modeling uses experimentally determined 3D structure of a protein to predict the 3D structure of another protein that has a similar amino acid sequence. It is based on two major observations: (1) structure of a protein is uniquely determined by its amino acid sequence and (2) during evolution, the structure is more conserved than the sequence such that similar sequences adopt

practically identical structure and distantly related sequences show similarity in folds. Homology modeling is a multistep method involving the following steps: (1) obtaining the sequence of the protein with unknown 3D structure, (2) template identification for comparative analysis, (3) fold assignment based on the known chemistry and biology of the protein, (4) primary structure alignment, (5) backbone generation, (6) loop modeling, (7) side chain modeling, (8) model optimization, and (9) model validation.

The methodology adopted in homology modeling of proteins can be described as follows. The target sequence is first compared to all sequences reported in the PDB using sequence analysis. Once a template sequence is found in the data bank, an alignment is made to identify optimum correlation between template and target. If identical residues exist in both the sequences, the coordinates are copied as such. If the residues differ, then only the coordinates of the backbone elements (N, Cα, C, and O) are copied. Loop modeling involves shifting all insertions and deletions to the loops and further modifying them to build a considerably well resembling model. Modeling the side chains involves copying the conserved residues, which also includes substitution of certain rotamers that are strongly favored by the backbone. Model optimization is required because of the expected differences in the 3D structures of the target and the template. The energy minimizations can be performed using molecular mechanics force fields (either well defined and/or self-parameterizing force fields). Molecular dynamic simulations offer fast, more reliable 3D structure of the protein. Model validation is a very important step in homology modeling, because several solutions may be obtained and the scientific user should interfere and make a choice of the best generated model. Often, the user may have to repeat the process with increased caution [20, 24].

1.3.3 Threading or Remote Homology Modeling

Threading (more formally known as "fold recognition") is a method that may be used to suggest a general structure for a new protein. It is mainly adopted when pairwise sequence identity is less than 25% between the known and unknown structure. Threading technique is generally associated with the following steps: (1) identify the remote homology between the unknown and known structure; (2) align the target and template; and (3) tailor the homology model [24].

1.4 COMPUTATIONAL MEDICINAL CHEMISTRY

Representation of drug molecular structures can be handled using computational chemistry methods, whereas that of macromolecules can be handled using computational biology methods. However, finding the therapeutic potential of the chemical species and understanding the drug–receptor interactions *in silico* requires the following well developed techniques of computational medicinal chemistry.

1.4.1 Quantitative Structure–Activity Relationship (QSAR)

QSAR is a statistical approach that attempts to relate physical and chemical properties of molecules to their biological activities. This can be achieved by using easily

TABLE 1.2 Different Dimensions in QSAR

1D QSAR: Affinity correlates with pK_a, log P, etc.
2D QSAR: Affinity correlates with a structural pattern.
3D QSAR: Affinity correlates with the three-dimensional structure.
4D QSAR: Affinity correlates with multiple representations of ligand.
5D QSAR: Affinity correlates with multiple representations of induced-fit scenarios.
6D QSAR: Affinity correlates with multiple representations of solvation models.

calculatable descriptors like molecular weight, number of rotatable bonds, and log P. Developments in physical organic chemistry over the years and contributions of Hammett and Taft in correlating the chemical activity to structure laid the basis for the development of the QSAR paradigm by Hansch and Fujita. Table 1.2 gives an overview of various QSAR approaches in practice. The 2D and 3D QSAR approaches are commonly used methods, but novel ideas are being implemented in terms of 4D–6D QSAR. The increased dimensionality does not add any additional accuracy to the QSAR approach; for example, no claim is valid which states that the correlation developed using 3D descriptors is better than that based on 2D descriptors.

2D QSAR Initially, 2D QSAR or the Hansch approach was in vogue, in which different kinds of descriptors from the 2D structural representations of molecules were correlated to biological activity. The basic concept behind 2D QSAR is that structural changes that affect biological properties are electronic, steric, and hydrophobic in nature. These properties can be described in terms of Hammett substituent and reaction constants, Verloop sterimol parameters, and hydrophobic constants. These types of descriptors are simple to calculate and allow for a relatively fast analysis.

Most 2D QSAR methods are based on graph theoretical indices. The graph theoretical descriptors, also called the molecular topological descriptors, are derived from the topology of a molecule, that is, the 2D molecular structure represented as graphs. These topological connectivity indices representing the branching of a molecule were introduced by Randíc [25] and further developed by Kier and Hall [26, 27]. The graph theoretical descriptors include mainly the Kier–Hall molecular connectivity indices (chi) and the Weiner [28, 29], Hosoya [30], Zagreb [31], Balaban [32], kappa shape [33], and information content indices [32]. The electrotopological state index (E-state) [34] combines the information related to both the topological environment and the electronic character of each skeletal atom in a molecule. The constitutional descriptors are dependent on the constitution of a molecule and are numerical descriptors, which include the number of hydrogen bond donors and acceptors, rotatable bonds, chiral centers, and molecular weight (1D) [35]. Apart from that, several indicator descriptors, which define whether or not a particular indicator is associated with a given molecule, are also found to be important in QSAR. The quantum chemical descriptors include the molecular orbital energies (HOMO, LUMO), charges, superdelocalizabilities, atom–atom and molecular polarizabilities, dipole moments, total and binding energies, and heat of formation. These are 3D descriptors derived from the 3D structure of the molecule and are electronic in nature [36]. These parameters are also often clubbed with the 2D QSAR analysis.

Statistical data analysis methods for QSAR development are used to identify the correlation between molecular descriptors and biological activity. This correlation may be linear or nonlinear and accordingly the methods may be divided into linear and nonlinear approaches. The linear approaches include simple linear regression, multiple linear regression (MLR), partial least squares (PLS), and genetic algorithm–partial least squares (GA-PLS). Simple linear regression develops a single descriptor linear equation to define the biological activity of the molecule. MLR is a step ahead as it defines a multiple term linear equation. More than one term is correlated to the biological activity in a single equation. PLS, on the other hand, is a multivariate linear regression method that uses principal components instead of descriptors. Principal components are the variables found by principal component analysis (PCA), which summarize the information in the original descriptors. The aim of PLS is to find the direct correlation not between the descriptors and the biological activity but between the principal component and the activity. GA-PLS integrates genetic algorithms with the PLS approach. Genetic algorithms are an automatic descriptor selection method that incorporates the concepts of biological evolution within itself. An initial random selection of descriptors is made and correlated to the activity. This forms the first generation, which is then mutated to include new descriptors, and crossovers are performed between the equations to give the next generation. Equations with better predictability are retained and the others are discarded. This procedure is continually iterated until the desired predictability is obtained or the specified number of generations have been developed. The nonlinear approaches include an Artificial Neural Network (ANN) and machine learning techniques. Unlike the linear approaches, nonlinear approaches work on a black box principle; that is, they develop a relation between the descriptors and the activity to predict the activity, but do not give the information on how the correlation was made or which descriptors are more contributing. The ANN algorithm uses the concept of the functioning of the brain and consists of three layers. The first layer is the input layer where the structural descriptors are given as an input; second is the hidden layer, which may be comprised of more than one layer. The input is processed in this part to give the predicted values to the third output layer, which gives the result to the user. The user can control the input given and the number of neurons and hidden layers but cannot control the correlating method [37–40].

The QSAR model developed by any statistical method has to be validated to confirm that it represents the true structure–activity relationship and is not a chance correlation. This may be done by various methods such as the leave-one-out and leave-multiple-out cross-validations and the bootstrap method. The randomization test is another validation approach used to confirm the adequacy of the training set. Attaching chemical connotation to the developed statistical model is an important aspect. A successful QSAR model not only effectively predicts the activity of new species belonging to the same series but also should provide chemical clues for future improvement. This requirement, as well as the recognition that the 3D representation of the chemicals gives more detailed information, led to the development of 3D QSAR.

3D QSAR 3D QSAR methods are an extension of the traditional 2D QSAR approach, wherein the physicochemical descriptors are estimated from the 3D struc-

tures of the chemicals. Typically, properties like molecular volume, molecular shape, HOMO and LUMO energies, and ionization potential are the properties that can be calculated from the knowledge of the 3D coordinates of each and every atom of the molecules. When these descriptors of series of molecules can be correlated to the observed biological activity, 3D QSAR models can be developed. This approach is different from the traditional QSAR only in terms of the descriptor definition and, in a sense, is not really 3D in nature.

Molecular fields (electrostatic and steric), which can be estimated using probe-based sampling of 3D structure of molecules within a molecular lattice, can be correlated with the reported numeric values of biological activity. Such methods proved to be much more informative as they provide differences in the fields as contour maps. The widely used CoMFA (comparative molecular field analysis) method is based on molecular field analysis and represents real 3D QSAR methods [41]. A similar approach was adopted in developing modules like CoMSIA (comparative molecular similarity index analysis) [42], SOMFA (self-organizing molecular field analysis) [43], and COMMA (comparative molecular moment analysis) [44]. Utilization and predictivity of CoMFA itself has improved sufficiently in accordance with the objectives to be achieved by it [45]. Despite the formal differences between the various methodologies, any QSAR method must include some identifiers of chemical structures, reliably measured biological activities, and molecular descriptors. In 3D QSAR, alignment (3D superimposition) of the molecules is necessary to construct good models. The main problems encountered in 3D QSAR are related to improper alignment of molecules, greater flexibility of the molecules, uncertainties about the bioactive conformation, and more than one binding mode of ligands. While considering the template, knowledge of the bioactive conformation of any lead compound would greatly help the 3D QSAR analysis. As discussed in Section 1.2.5, this may be obtained from the X-ray diffractions or conformation at the binding site, or from the global minimum structure. Alignment of 3D structures of molecules is carried out using RMS atoms alignment, moments alignment, or field alignment. The relationship between the biological activity and the structural parameters can be obtained by multiple linear regression or partial least squares analysis. Given next are some details of the widely used 3D QSAR approach CoMFA.

CoMFA (Comparative Molecular Field Analysis) DYLOMMS (dynamic lattice-oriented molecular modeling system) was one of the initial developments by Cramer and Milne to compare molecules by aligning in space and by mapping their molecular fields to a 3D grid. This approach when used with partial least squares based statistical analysis gave birth to the CoMFA approach [46]. The CoMFA methodology is a 3D QSAR technique that allows one to design and predict activities of molecules. The database of molecules with known properties is suitably aligned in 3D space according to various methodologies. After consistently aligning the molecules within a molecular lattice, a probe atom (typically carbon) samples the steric and electrostatic interactions of the molecule. Charges are then calculated for each molecule using any of the several methods proposed for partial charge estimation. These values are stored in a large spreadsheet within the module (SYBYL software) and are then accessed during the partial least squares (PLS) routine, which attempts to correlate these field energy terms with a property of interest by the use of PLS with cross-validation, giving a measure of the predictive power of the model.

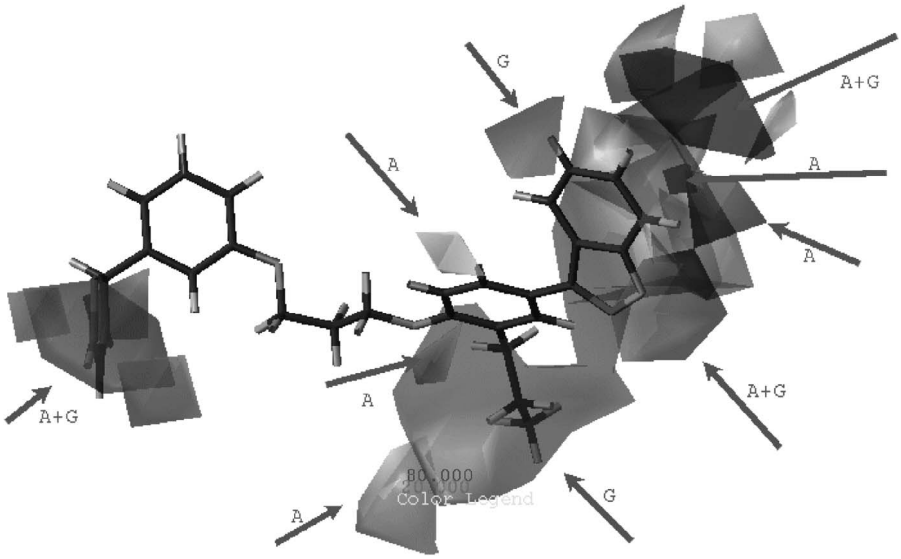

FIGURE 1.4 Steric and electrostatic contour map for the dual model showing the contributions from each model. "A" depicts the contributions made by the α-model and "G" depicts the contributions made by the γ- and model. (Reproduced with permission from The American Chemical Society; S. Khanna, M. E. Sobhia, P. V. Bharatam *J Med Chem* 2005;48:3015.)

Electrostatic maps are generated, indicating red contours around regions where high electron density (negative charge) is expected to increase activity, and blue contours where low electron density (partial positive charge) is expected to increase activity. Steric maps indicate areas where steric bulk is predicted to increase (green) or decrease (yellow) activity [41, 45]. Figure 1.4 shows a typical contour map from CoMFA analysis. CoMSIA [42], CoMMA [44], GRID [47], molecular shape analysis (MSA) [48], comparative receptor surface analysis (CoRSA) [49], and Apex-3D [50] are other 3D QSAR methods that are being employed successfully.

4D QSAR 4D QSAR analysis developed by Vedani and colleagues incorporates the conformational alignment and pharmacophore degrees of freedom in the development of 3D QSAR models. It is used to create and screen against 3D-pharmacophore QSAR models and can be used in receptor-independent or receptor-dependent modes. 4D QSAR can be used as a CoMFA preprocessor to provide conformations and alignments; or in combination with CoMFA to combine the field descriptors of CoMFA with the grid cell occupancy descriptors (GCODs) of 4D QSAR to build a "best" model; or in addition to CoMFA because it treats multiple alignments, conformations, and embedded pharmacophores, which are limitations of CoMFA [51].

5D QSAR The 4D QSAR concept has been extended by an additional degree of freedom—the fifth dimension—allowing for multiple representations of the topology of the quasi-atomistic receptor surrogate. While this entity may be generated using up to six different induced-fit protocols, it has been demonstrated that the

simulated evolution converges to a single model and that 5D QSAR, due to the fact that model selection may vary throughout the entire simulation, yields less biased results than 4D QSAR, where only a single induced-fit model can be evaluated at a time (software Quasar) [52, 53].

6D QSAR A recent extension of the Quasar concept to sixth dimension (6D QSAR) allows for the simultaneous consideration of different solvation models [54]. This can be achieved explicitly by mapping parts of the surface area with solvent properties (position and size are optimized by the genetic algorithms) or implicitly. In Quasar, the binding energy is calculated as

$$E_{\text{binding}} = E_{\text{ligand-receptor}} - E_{\text{desolvation,ligand}} - T\,\Delta S - E_{\text{internal strain}} - E_{\text{induced fit}} \qquad (1.4)$$

1.4.2 Pharmacophore Mapping

A pharmacophore may be defined as the spatial arrangement of a set of key features present in a chemical species that interact favorably with the receptor leading to ligand–receptor binding and which is responsible for the observed therapeutic effect. It is the spatial arrangement of key chemical features that are recognized by a receptor and are thus responsible for biological response. Pharmacophore models are typically used when some active compounds have been identified but the 3D structure of the target protein or receptor is unknown. It is possible to derive pharmacophores in several ways, by analogy to a natural substrate, by inference from a series of dissimilar biologically active molecules (active analogue approach) or by direct analysis of the structure of known ligand and target protein.

A pharmacophoric map is a 3D description of a pharmacophore developed by specifying the nature of the key pharmacophoric features and the 3D distance map among all the key features. Figure 1.5 shows a pharmacophore map generated from the DISCO software module of SYBYL. A pharmacophore map may be generated from the superimposition of the active compounds to determine their common features. Given a set of active molecules, the mapping of a pharmacophore involves two steps: (1) analyzing the molecules to identify pharmacophoric features, and (2) aligning the active conformations of the molecules to find the best overlay of the corresponding features. Various pharmacophore mapping algorithms differ in the way they handle the conformational search, feature definition, tolerance definition, and feature alignment [55]. During pharmacophore mapping, generation and optimization of the molecules and the location of ligand points and site points (projections from ligand atoms to atoms in the macromolecule) are carried out. Typical ligand and site points are hydrogen bond donors, hydrogen bond acceptors, and hydrophobic regions such as centers of aromatic rings. A pharmacophore map identifies both the bioactive conformation of each active molecule and how to superimpose and compare, in three dimensions, the various active compounds. The mapping technique identifies what type of points match in what conformations of the compounds.

Besides ligand-based automated approaches, pharmacophore maps can also be generated manually. In such cases, common structural features are identified from a set of experimentally known active compounds. Conformational analysis is carried out to generate different conformations of the molecules and interfeature distances

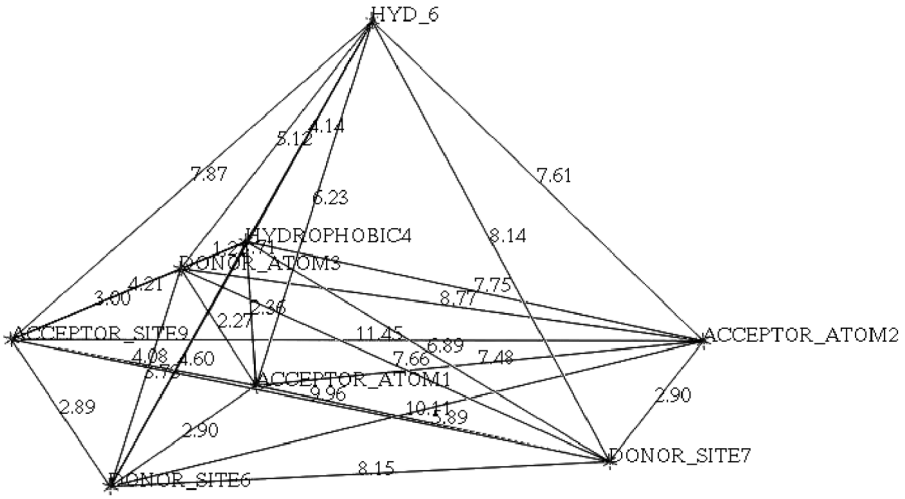

FIGURE 1.5 A pharmacophore map developed from a set of GSK3 inhibitors. The pharmacophore features include hydrogen bond acceptor atoms, hydrogen bond donor atoms, hydrogen bond donor site, hydrogen bond acceptor site, and hydrophobic centers. This 3D picture also shows the distance relationship between various pharmacophoric features present in the map.

are inferred to develop the final models. The receptor mapping technique is also currently in practice to develop pharmacophore models. The important residues required for binding the pharmacophores are identified, which are employed for generating the receptor-based pharmacophores. The structure of protein can be used to generate interaction sites or grids to characterize favorable positions for ligands.

After a pharmacophore map has been derived, there are two ways to identify molecules that share its features and thus elicit the desired response. First is the *de novo* drug design, which seeks to link the disjoint parts of the pharmacophore together with fragments in order to generate hypothetical structures that are chemically novel. Second is the 3D database searching, where large databases comprising 3D structures are searched for those that match to a pharmacophoric pattern. One advantage of the second method is that it allows the ready identification of existing molecules, which are either easily available or have a known synthetic route [56, 57]. Pharmacophore mapping methods are described next.

Distance Comparison Method (DISCO) The various steps involved in DISCO-based generation of a pharmacophore map are conformational analysis, calculation of the location of the ligand and site points, finding potential pharmacophore maps, and graphics analysis of the results. In the process of conformational search, 3D structures can be generated using any building program like CONCORD, from crystal structures, or from conformational searching and energy minimization with any molecular or quantum mechanical technique. Comparisons of all the duplicate conformations are excluded while comparing all the conformations. If each corresponding interatomic distance between these atoms in the two conformations is less

than a threshold (0.4 Å), then the higher energy conformation is rejected. DISCO calculates the location of site points, which can be the location of ligand atoms, or other atom-based points, like centers of rings or a halogen atom, which are points of potential hydrophobic groups. The other point is the location of the hydrogen bond acceptors or donors. The default locations of site hydrogen bond donor and acceptor points are based on literature compilations of observed intermolecular crystallographic contacts in proteins and between the small molecules. Hydrogen bond donors and acceptors such as OH and NH_2 groups can rotate to change the locations of the hydrogen atom.

During the process of performing pharmacophore mapping in DISCO, the user may input the tolerance for each type of interpoint distance. The user may direct the DISCO algorithm to consider all the potential points and to stop when a pharmacophore map with a certain total number of points is found. Alternatively, the user may specify the types of points, and the maximum and minimum number of each, that every superposition must include. It can also be directed to ignore specific compounds if they do not match a pharmacophore map found by DISCO. The user may also specify that only the input chirality is used for certain molecules and that only certain conformations below a certain relative energy should be considered.

The DISCO algorithm involves finding the reference molecule, which is the one with the fewest conformations. The search begins by associating the conformations of each molecule with each other. DISCO then calculates the distances between points in each 3D structure. Then it prepares the corresponding tables that relate interpoint distances in the current reference conformation and distances in every other 3D structure. Distances correspond if the point types are the same. These distances differ by no more than the tolerance limits. The clique-detection algorithm then identifies the largest clique of distances common between the reference XYZ set and every other 3D structure. It then forms union sets for the cliques of each molecule. Finally, the sets with cliques that meet the group conditions are searched [58, 59].

CATALYST According to the pharmacophore mapping software CATALYST, a conformational model is an abstract representation of the accessible conformational space of a ligand. It is assumed that the biologically active conformation of a ligand (or a close approximation thereof) should be contained within this model. A pharmacophore model (in CATALYST called a hypothesis) consists of a collection of features necessary for the biological activity of the ligands arranged in 3D space, the common ones being hydrogen bond acceptor, hydrogen bond donor, and hydrophobic features. Hydrogen bond donors are defined as vectors from the donor atom of the ligand to the corresponding acceptor atom in the receptor. Hydrogen bond acceptors are analogously defined. Hydrophobic features are located at the centroids of hydrophobic atoms. CATALYST features are associated with position constraints that consist of the ideal location of a particular feature in 3D space surrounded by a spherical tolerance. In order to map the pharmacophore, it is not necessary for a ligand to possess all the appropriate functional groups capable of simultaneously residing within the respective tolerance spheres of the pharmacophoric features. However, the fewer features an inhibitor maps to, the poorer is its fit to them and the lower is its predicted affinity [60–63].

1.4.3 Molecular Docking

There are several possible conformations in which a ligand may bind to an active site, called the binding modes. Molecular docking involves a computational process of searching for a conformation of the ligand that is able to fit both geometrically and energetically into the binding site of a protein. Docking calculations are required to predict the binding mode of new hypothetical compounds. The docking procedure consists of three interrelated components—identification of the binding site, a search algorithm to effectively sample the search space (the set of possible ligand positions and conformations on the protein surface), and a scoring function. In most docking algorithms, the binding site must be predefined, so that the search space is limited to a comparatively small region of the protein. The search algorithm effectively samples the search space of the ligand–protein complex. The scoring function used by the docking algorithm gives a ranking to the set of final solutions generated by the search. The stable structures of a small molecule correspond to minima on the multidimensional energy surface, and different energy calculations are needed to identify the best candidate. Different forces that are involved in binding are electrostatic, electrodynamic, and steric forces and solvent related forces. The free energy of a particular conformation is equal to the solvated free energy at the minimum with a small entropy correction. All energy calculations are based on the assumption that the small molecule adopts a binding mode of lowest free energy within the binding site. The free energy of binding is the change in free energy that occurs upon binding and is given as

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}) \tag{1.5}$$

where G_{complex} is the energy of the complexed protein and ligand, G_{protein} is the free energy of noninteracting separated protein, and G_{ligand} is the free energy of noninteracting separated ligand.

The common search algorithms used for the conformational search, which provide a balance between the computational expense and the conformational search, include molecular dynamics, Monte Carlo methods, genetic algorithms, fragment-based methods, point complementary methods, distance geometry methods, tabu searches, and systematic searches [64].

Scoring functions are used to estimate the binding affinity of a molecule or an individual molecular fragment in a given position inside the receptor pocket. Three main classes of scoring functions are known, which include force field-based methods, empirical scoring functions, and knowledge-based scoring functions. The force field scoring functions use molecular mechanics force fields for estimating binding affinity. The AMBER and CHARMM nonbonded terms are used as scoring functions in several docking programs. In empirical scoring functions, the binding free energy of the noncovalent receptor–ligand complex is estimated using chemical interactions. These scoring functions usually contain individual terms for hydrogen bonds, ionic interactions, hydrophobic interactions, and binding entropy, as in the case of SCORE employed in DOCK4 and Böhm scoring functions (explained in detail in Section 1.4.4) used in FlexX. In empirical scoring functions, less frequent interactions are usually neglected. Knowledge-based scoring functions try to capture the knowledge about protein–ligand binding that is implicitly stored in the Protein Data Bank by means of statistical analysis of structural data, for example, PMF and

DrugScore functions, Wallqvist scoring function, and the Verkhivker scoring function [5, 37, 65–67]. Various molecular docking software packages are available, such as FlexX [68], Flexidock [58], DOCK [69], and AUTODOCK [70].

FlexX FlexX is a fragment-based method for docking which handles the flexibility of the ligand by decomposing the ligand into fragments and performs the incremental construction procedure directly inside the protein active site. It allows conformational flexibility of the ligand while keeping the protein rigid. The base fragment or the ligand core is selected such that it has the most potential interaction groups and the fewest alternative conformations. It is placed into the active site and joined to the side chains in different conformations. Placements of the ligand are scored on the basis of protein–ligand interactions and ranked after the estimation of binding energy. The scoring function of FlexX is a modification of Böhm's function developed for the *de novo* design program LUDI. Figure 1.6 shows details of the interaction between a ligand and a receptor, obtained from FlexX molecular docking.

DOCK DOCK is a simple minimization program that generates many possible orientations of a ligand within a user selective region of the receptor. DOCK is a program for locating feasible binding orientations, given the structures of a "ligand" molecule and a "receptor" molecule [69]. DOCK generates many orientations of one ligand and saves the best scoring orientation. The docking process is handled

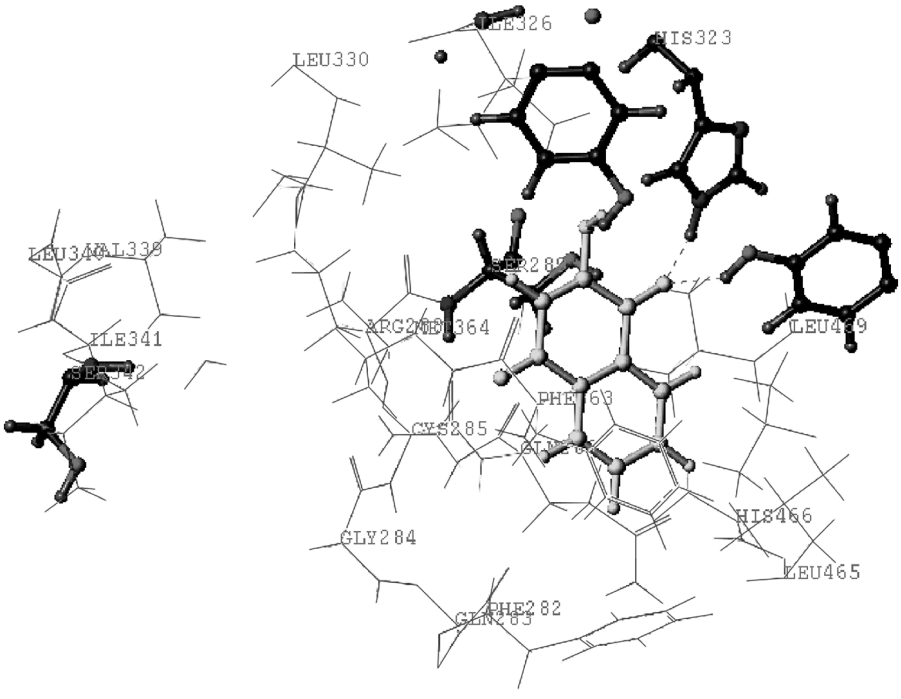

FIGURE 1.6 The result of docking a ligand in the active site of PPARγ. The ligand has a hydrogen bonding interaction with histidine and tyrosine.

in four stages—ligand preparation, site characterization, scoring grid calculation, and finally docking. Site characterization is carried out by constructing site points, to map out the negative image of the active site, which are then used to construct orientations of the ligand. Scoring grid calculations are necessary to identify ligand orientations. The best scoring poses may be viewed using a molecular graphics program and the underlying chemistry may be analyzed.

There are many other widely used molecular docking software packages, like Flexidock (based on genetic algorithm), Autodock (based on Monte Carlo simulations and annealing), MCDOCK (Monte Carlo simulations), FlexE (ensemble of protein structures to account for protein flexibility), and DREAM++ (to dock combinatorial libraries).

1.4.4 De Novo Design

De novo design is a complementary approach to molecular docking: whereas in molecular docking already known ligands are employed, in *de novo* design, ligands are built inside the ligand binding domain. This is an iterative process in which the 3D structure of the receptor is used to build the putative ligand, fragment by fragment, within the receptor groove. Two basic types of algorithms are being widely used in *de novo* design. The first one is the "outside-in-method," in which the binding site is first analyzed to determine which specific functional groups might bind tightly. These separated fragments are then connected together with standard linker units to produce the ligands. The second approach is the "inside-out-method," where molecules are grown within the binding site so as to efficiently fit inside. *De novo* design is the only method of choice when the receptor structure is known but the lead molecules are not available. This method can also be used when lead molecules are known but new scaffolds are being sought. There are several programs developed by various researchers for constructing ligands *de novo*. GROW [71], GRID [72], CAVEAT [73], LUDI [74–77], LEAPFROG [58], GROUPBUILD [78], and SPROUT [79] are some of the *de novo* design programs that have found wide application.

GRID The GRID program developed by Goodford [72] is an active site analysis method where the properties of the active site are analyzed by superimposition with a regular grid. Probe groups like water, methyl group, amine nitrogen, carboxyl oxygen, and hydroxyl are placed at the vertices of the grid and its interaction energy with the protein is calculated at each point using an empirical energy function that determines which kind of atoms and functional groups are best able to interact with the active site. The array of energy values is represented as a contour, which enables identification of regions of attractions between the probe and the protein. It is not a direct ligand generation method but positions simple fragments.

LUDI LUDI, developed by Böhm [74–77], is one of the most widely used automated programs available for *de novo* design. It uses a knowledge-based approach based on rules about the energetically favorable interaction geometries of nonbonded contacts like hydrogen bonds and hydrophobic contacts between the functional groups of the protein and ligand. In LUDI the rules derived from statistical analysis of crystal packings of organic molecules are employed. LUDI is

fragment based and works in three steps. It starts by identifying the possible hydrogen bonding donors and acceptors and hydrophobic interactions, both aliphatic and aromatic, in the binding site represented as interaction site points. The site points are positions in the active site where the ligand could form a nonbonded contact. A set of interaction sites encompasses the range of preferred geometries for a ligand atom or functional group involved in the putative interaction, as observed in the crystal structure analyses. LUDI models the H-donor and H-acceptor interaction sites and the aliphatic or aromatic interaction sites. The interaction sites are defined by the distance R, angle α, and dihedral angle ω. The fragments from a 3D database of small molecules are then searched for positioning into suitable interaction sites such that hydrogen bonds can be formed and hydrophobic pockets filled with hydrophobic groups. The suitably oriented fragments are then connected together by spacer fragments to the respective link sites to form the entire molecule. Figure 1.7 shows LUDI generated fragment interaction sites inside the iNOS substrate binding domain.

An empirical but efficient scoring function is used for prioritizing the hit fragments given by LUDI. It estimates the free energy of binding (ΔG) based on the

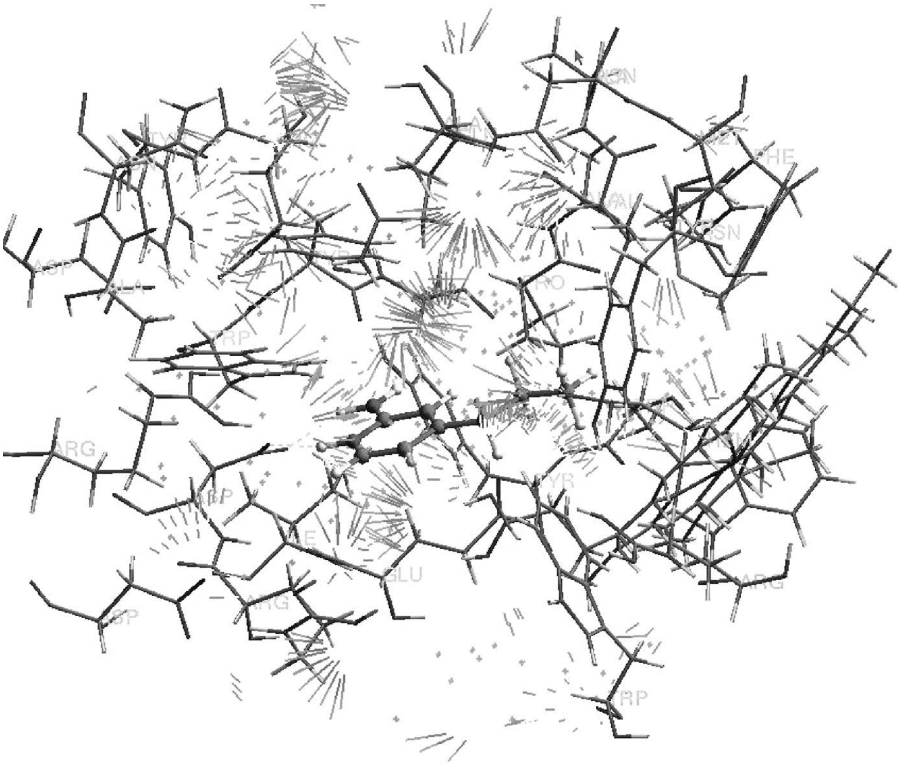

FIGURE 1.7 An example of *de novo* design exercise. In the substrate binding domain of inducible nitric oxide synthase, the stick representation shows the protein structure; ball-and-stick representation belongs to the designed ligand, and the gray sticks point out the interaction sites.

hydrogen bonding, ionic interactions, hydrophobic contact areas, and number of rotatable bonds in the ligand. The LUDI scoring function is given as

$$\Delta G = \Delta G_{\mathrm{o}} + \Delta G_{\mathrm{hb}} \sum f(\Delta R) f(\Delta\alpha) + \Delta G_{\mathrm{ion}} \sum f(\Delta R) f(\Delta\alpha) + \Delta G_{\mathrm{lipo}} A_{\mathrm{lipo}} + \Delta G_{\mathrm{rot}} NR + \Delta G_{\mathrm{aro/aro}} \Delta N_{\mathrm{aro/aro}} \quad (1.6)$$

ΔG_{o} represents the constant contribution to the binding energy due to loss of translational and rotational entropy of the fragment. ΔG_{hb} and ΔG_{ion} represent the contributions from an ideal neutral hydrogen bond and an ideal ionic interaction, respectively. The ΔG_{lipo} term represents the contribution from lipophilic contact and the ΔG_{rot} term represents the contribution due to the freezing of internal degrees of freedom in the fragment. *NR* is the number of acyclic sp^3-sp^3 and sp^3-sp^2 bonds.

1.5 PHARMACOINFORMATICS

Information technology provides several databases, data analysis tools, and knowledge extraction techniques in almost every facet of life. In pharmaceutical sciences, several successful attempts are being made under the umbrella of pharmacoinformatics (synonymously referred to as pharmainformatics) (Fig. 1.8). The scope and limitations of this field are not yet understood. However, it may be broadly defined as the application of information technology in drug discovery and development. It encompasses all possible information technologies that eventually contribute to drug discovery. Chemoinformatics and bioinformatics contribute directly to drug discovery through virtual screening. Topics like neuroinformatics, immunoinformatics, vaccine informatics, and biosystem informatics contribute indirectly by providing necessary inputs for pharmaceutical design in this area. Topics like metabolomics, toxicoinformatics, and ADME informatics are contributing to this field by providing information regarding the fate of a NCE/lead *in vitro* and *in vivo* conditions. In this chapter some important aspects of these topics are presented. It is not easy to offer a comprehensive definition of this field at this stage owing to the fact that several bold attempts are being made in this field and initial signals related to a common platform are only emerging. *Drug Discovery Today* made initial efforts in this area by bringing out a supplement on this topic in which it was mainly treated as a scientific discipline with the integration of both bioinformatics and chemoinformatics [80, 81]. Recent trends in this area include several service-oriented themes including healthcare informatics [82], medicine informatics [83], and nursing informatics [84]. Here we present an overview of the current status.

1.5.1 Chemoinformatics

Chemoinformatics deals with information storage and retrieval of chemical data. This has been pioneered principally by the American Chemical Society and Cambridge Crystallographic Databank. However, the term chemoinformatics came into being only recently when methods of deriving science from the chemical databases was recognized. The integration of back-end technologies (for storing and representing chemical structure and chemical libraries) and front-end technologies (for assessing and analyzing the structures and data from the desktop) provides