SECTION I

PRINCIPLES
AND METHODS
MODELING IN THE PHARMACEUTICAL INDUSTRY

CONTENTS

1.1 Introduction ........................................ 3
1.2 Modeling Approaches .................................. 4
1.3 Steps Needed to Maximize Effective Integration of Models into R&D Workflow .................................. 7
1.4 Scope of the Book ..................................... 8
    Keywords ............................................. 10
    Reference ............................................. 12

1.1 INTRODUCTION

In an effort to reduce the attrition rates of drugs, pharmaceutical companies are constantly looking to improve and understand compound behavior through the use of novel tools. Modeling is one such tool that has gradually gained recognition in the pharmaceutical industry, over the last couple of decades, as a means of achieving quality, efficiency, and significant cost savings. Modeling...
and simulation methods have played a crucial role in the pharmaceutical industry in identifying and validating target, predicting the efficacy, absorption, distribution, metabolism, excretion, toxicity (ADMET), and safety of drug candidates, aiding a better understanding of data through effective integration and extraction of knowledge, predicting the human dose, developing new formulations, designing safety and efficacy trials, and guiding regulatory decisions. Most models are used in a build–validate–learn–refine cycle in which all available knowledge that can aid prediction of a property of interest is initially captured during model building. It is then used for predicting observations (validation phase), and any discrepancies observed predicted from observed is then understood on a scientific basis (learning phase) and appropriately incorporated in the model (refine phase). Once a model has been tested to provide satisfactory results, it can be used on a routine basis, reserving animal studies and other resource-intensive experiments for confirmation only.

The use of *in silico* technologies can reduce the cost of drug development by up to 50% according to some analysts. The impact of integrating modeling into the research and development (R&D) workflow has been so encouraging that many companies have increased their investments in this sector. The Food and Drug Administration (FDA) “critical path” document recommends model-based drug development for improved knowledge management and decision making. The key elements of such a model-based drug development and how they fit together to aid strategy and decision making in drug development is outlined by Lalonde et al.

### 1.2 MODELING APPROACHES

From understanding a disease to bringing a safe and effective new treatment to patients, it takes about 10–15 years for a pharmaceutical company to discover a potential drug (drug discovery) and to develop it as a final product (drug development). A schematic of a drug discovery and development pipeline is shown in Figure 1.1. Advances in genomics and proteomics and an increase in computational power have contributed to increasing our knowledge of disease at the level of genes, proteins, and cells. This understanding leads to the identification of proteins, which are involved in a disease of interest. A single protein/gene that has been validated to be relevant in a disease and to be druggable is chosen as the target. Hits to this target are identified through virtual screening and high-throughput screening (HTS) assays. Compounds that can best modulate the target are chosen as hits. Hits are classified into a small set of lead series (*lead generation*). The most promising series showing potential drug activity, reduced off-target toxicity, and with physicochemical and metabolic profiles that are compatible with acceptable *in vivo* bioavailability progress into the *lead optimization* stage. The objective at this stage is to select a candidate drug that meets predefined criteria with respect to efficacy,
pharmacokinetics, and safety. The candidate drug is then developed to a final drug product, after sufficient testing in animals (preclinical development) and humans (clinical development) to confirm the efficacy and safety of the drug. The modeling methods along the drug discovery and development value chain are indicated in Figure 1.1.

Models allow us to understand how complex interactions and processes work. Sometimes, modeling provides a unique way to understanding a system. Figure 1.2 summarizes the reasons for employing models in the pharmaceutical industry. It is important to be aware that all models are only approximations of the system they represent. Underlying assumptions should be carefully weighed to get the best benefits from a model. In addition, different modeling approaches differ in their strengths and limitations. Quantitative structure–activity relationships (QSAR) and quantitative structure property relationships (QSPR) models rely on combining appropriate descriptors for compounds in a training set. Their key strengths are simplicity and ease of use. However, the predictive power of these models is restricted to compounds within the same chemical space as that of the training set. Empirical or data-driven models are built and refined only after the experimental data is collected (cannot be prespecified) and its parameters lack physical/physiological/biochemical interpretation. They are best employed for exploratory data analysis. On the other hand, mechanistic models are prespecified and capture the underlying mechanisms of the system they represent to the extent known, with parameters corresponding to some physical entities of the system. These models can, therefore, be used to predict the next set of data. An example of this is physiologically-based models.

Pharmacokinetic (PK) modeling provides information about processes that affect the kinetics of a compound in a species, such as absorption, distribution, metabolism, and excretion using the concentration–time profile. Traditionally, this has been done with compartmental PK modeling, a data-driven approach in which the model structure is defined by the data (therefore empirical). The fall in concentration with respect to time is fitted to a series of exponential terms whose decay constants and pre-exponents are related to the rates of absorption, distribution, metabolism, and elimination. The model that best fits the in vivo PK data, according to some defined statistical criterion, is chosen as the final model. Empirical models are case specific and the potential for credible extrapolations using these models is limited. Since the pharmacodynamic response of a drug need not necessarily parallel its pharmacokinetics, PK models are combined with pharmacodynamic (PD) effect–concentration profiles at different doses. This data-driven, exploratory PK/PD modeling has long been used in drug development, for getting a continuous description of the effect–time course resulting directly from the administration of a certain dose. Physiologically-based pharmacokinetic (PBPK) modeling offers a mechanistic approach to predicting the disposition of a drug, which can then be combined with a PD model (PBPK/PD). Although the principles behind a PBPK approach has long been known through the work of Teorell in 1937, the mathematical complexity of the model and the lack of physiological data needed for the
Figure 1.1. Modeling at various stages of drug discovery and development.
The tremendous increase in computational power at relatively low cost paved the way for complex PBPK models to be built. PBPK models help simulate the concentration–time profile of a drug in a species by integrating the physicochemical properties of the compound with the physiology of the species. Being mechanistic, PBPK models can be used to simulate and to predict the next set of data and to plan the next experiment.

1.3 STEPS NEEDED TO MAXIMIZE EFFECTIVE INTEGRATION OF MODELS INTO R&D WORKFLOW

Although PBPK models were developed for cancer drugs even during the 1960s and 1970s by Bischoff et al. and Bischoff and Dedrick, the pharma industry has been slow to exploit the power of PBPK. While the importance of integrating modeling, simulation, and other in silico technologies in the R&D workflow is clearly acknowledged by leaders in the industry and by regulatory authorities, practical implementation has been slow especially in some areas of modeling. A number of reasons have been identified. The lack of trained/skilled scientists, sceptical attitude from project teams, and lack of commitment on the
part of leadership to implement are the most important among them. In this, the role of management in driving the integration is seen as key to bringing about a change in the workflow and mindset of the scientists as well as to allocate resources for training scientists. Gaining acceptance among project teams is vital to ensure that modeling results are seriously considered and incorporated in decisions, thus paving the way for cost-effective and efficient drug discovery and development.

1.4 SCOPE OF THE BOOK

Physiologically-based pharmacokinetic modeling for the discovery and development of small-molecule and biological drugs will be the main focus of the book, as applications of PBPK in environmental toxicology and human health risk assessment have already been the subject of a previous publication. The chapters in the first section will cover the basics of PBPK modeling and simulation, while the second section will deal with its applications in drug discovery and development.

Chapters 2–6 will elaborate on the principles essential for integrating species physiology with compound-dependent properties. Chapter 7 will put together all of the absorption, distribution, metabolism, and excretion (ADME) physiological models for small-molecule drugs.

Physiologically-based PK modeling involves the use of a number of compound-dependent and physiology-dependent parameters. Being a parameter-intensive model, the predicted outcome could be associated with a high level of uncertainty. It is, therefore, important to consider the propagation of error arising from the uncertainties in input parameters. These uncertainties can be modeled using the Monte Carlo approach, which forms the subject of Chapter 8.

As late failures in the drug development process become more costly, the desire to evaluate the potential for risks earlier in the drug discovery process has become a growing industry trend. An early assessment of the potential for drug–drug interactions (DDI) with co-medications mediated by inhibition/induction of cytochrome P450 (CYP) enzymes or from transporters is, therefore, seen as imperative even in the lead optimization stage. PBPK models provide a mechanistic approach to integrating relevant information on a potential inhibitor and a substrate for the prediction of DDI risk. Chapter 9 details the differential equations that describe the mutually dependent kinetics of co-administered drugs and wraps up with a discussion on the advantages of physiological models over static models in the evaluation of drug–drug interactions.

Biologics (or biologics) are fast emerging as alternative therapeutics to small molecules. Biologics are proteins such as monoclonal antibodies, cytokines, growth factors, enzymes, and thrombolytics that can treat a variety of diseases. Since the launch of Eli Lilly’s recombinant human insulin in 1982, more than 100 biologics have received marketing approval in the United States, highlighting their importance as a source of new drugs and new revenues. With an increasing fraction of pharmaceutical R&D devoted to
biologicals, it is expected to have a significant role in drug development in the future. Chapter 10 is devoted to examining the differences between biologicals and small molecules with respect to PK behavior and how these differences can be accommodated within PBPK models.

Section II of the book will cover applications of PBPK modeling in drug discovery and development with examples. Applications in the pharmaceutical sector will be the main focus. PBPK modeling can be used as a prediction, simulation, or as an extrapolation tool. PK properties such as absorption, distribution, and elimination of compounds are influenced not only by compound properties but also by the physiology of the species in which they are observed. PBPK modeling attempts to integrate available structural, in silico or in vitro physicochemical, and human-specific biochemical compound data in a physiological context for the predictions of PK parameters such as absorption and distribution or time profiles of plasma concentrations of drugs. Chapter 11 describes how PBPK models provide an excellent framework for enabling data integration and human PK predictions. Chapter 11 also describes the applications of parameter sensitivity analysis for optimizing lead compounds during drug discovery. In the lead optimization stage, understanding the effects of modulating key ADME-determining compound-dependent properties on a desired PK outcome is often needed in order to optimize the physicochemical space. The PK outcome could be metabolic liability, absorption, distribution, or bioavailability of compounds. The effects of modulation depend very much on the physicochemical space chosen initially.

The value of a PBPK model as a prediction tool is sometimes limited by the lack of reliable input parameters especially for clearance, where the in vitro measurements for intrinsic clearance rarely match up to the in vivo. The mechanistic structure of PBPK models can be better exploited when it is used as a simulation tool. In a simulation, the focus is not on quantitative predictions. Instead, the emphasis is on gaining valuable insights into processes driving the pharmacokinetics of a compound, through hypothesis generation and testing. This neglected area, holding the promise of improving the quality of selected leads, reducing animal studies and cost, is the subject of Chapter 12. The mechanistic basis of PBPK models makes them ideal for extrapolation.

The structure of PBPK models allows the prediction of tissue concentrations, which can be valuable in human health risk assessment or for linking with pharmacodynamic models. PBPK models when combined with PD models can be powerful in predicting the time-course of drug effects under physiological and pathological conditions. The integration of PBPK models with PD models and a robust design of clinical trials and is covered in Chapter 13. PBPK–based predictions aid the optimal use of all available compound information within a physiological context, making experiments confirmatory rather than exploratory. These have a tremendous impact in reducing preclinical and clinical studies thereby reducing costs.

Applications of PBPK in population modeling form the subject of Chapter 14. Drug failures can sometimes result from considering only an average person and neglecting physiological and genomic variability that can lead to a spread in
both plasma drug concentrations and drug response. Chapter 14 describes how targeted therapy and personalized medicine can be achieved with PBPK/PD modeling.

Chapter 15 aims to seamlessly integrate all the applications of PBPK along the drug discovery and development value chain.

**KEYWORDS**

**Binding Site Analysis:** Use of computational tools for the prediction of potential ligand-binding active sites in a target protein, given its three-dimensional structure. This is achieved through searching for surface features of the protein (geometry and functional groups) that provide the best shape complementarity and interactions with a set of known ligands.

**Biological Systems Modeling:** Involves computer simulations of biological systems to analyze and visualize the complex connections of cellular processes such as the networks of metabolites and enzymes that comprise metabolism, signal transduction pathways, and gene regulatory networks.

**Clinical Trial Simulation:** Combining structural and stochastic elements of pharmacokinetic and pharmacodynamic models to produce a data set that will resemble the results of an actual trial.

**Compartmental PK Modeling:** Use kinetic models to describe the concentration–time profile. The compartments do not relate to meaningful physiologic spaces.

**Druggable:** A druggable target is a protein whose activity can be modulated by a small molecule drug. A druggable target is crucial in determining the progression of a drug discovery project to the lead generation stage.

**hERG Modeling:** The human ether-a-go-go related gene (hERG) codes for the potassium ion channel **Kv11.1**, a protein that mediates the repolarizing **IKr** current in the cardiac action potential. Drugs inhibiting the channel can cause a potentially fatal QT prolongation with a concomitant risk of sudden death. In computational drug design, there are 2 main approaches to hERG modeling. Pharmacophoric or ligand-based modeling relies on determining the physicochemical features associated with the channel block to predict the hERG blocking potential of compounds. Target-based partial homology models of the hERG channel have also been built to interpret electrophysiological and mutagenesis studies.

**Homology Modeling:** Involves taking a known sequence with an unknown structure and mapping it against a known structure of one or more homologous proteins in an effort to gain insights into three-dimensional structure of the protein.

**Lead Generation:** A phase in drug discovery in which the objective is to identify one or more chemical series with potential drug activity, reduced off-target toxicity, and with physicochemical and metabolic profile that are compatible with acceptable in vivo bioavailability.
Lead Optimization: Phase in drug discovery, following lead generation, in which the objectives are to optimize the PK and PD (efficacy, selectivity and safety) in the screening stage and to select for drug development, a high-quality candidate drug that satisfies a preset target profile in the drug candidate selection stage.

Model-Based Drug Development: Statistical and mathematical modeling that allows for quantitative and effective use of prior information (preclinical efficacy and safety models) and clinical data (information across drugs, end points, trials and doses) for improved data analysis, clinical study design, knowledge management and decision making in clinical drug development.

PBPK/PD Modeling: Linking a physiologically-based pharmacokinetic (PBPK) model, which relates a drug’s exposure to its dose with a pharmacodynamic (PD) model, which relates the pharmacological response to exposure.

Pharmacophore Modeling: A ligand-based approach to virtual screening, which makes use of two- or three-dimensional pharmacophores generated from a set of known active compounds to the selected target.

PK/PD Modeling: Linking a pharmacokinetic (PK) model, which relates a drug’s exposure to its dose with a pharmacodynamic (PD) model, which relates the pharmacological response to exposure.

Population Modeling: Seeks to identify and quantify the pathophysiological factors that cause changes in the dose–concentration relationship, so that any resulting clinically significant shifts in the therapeutic index can be addressed through appropriate dose adjustments.

Prediction of PK Properties: Generally QSPR models that employ the structure-dependent properties of a compound to arrive at pharmacokinetic properties such as fraction of drug unbound in plasma, volume of distribution, renal elimination, fraction of compound absorbed, or bioavailability. Physiological models are also used to predict PK properties.

Protein Modeling is the prediction of the three-dimensional (secondary, tertiary, and quaternary) structure of a protein from its amino acid sequence.

Reactive Metabolite Prediction: Using the chemical structure of a compound and a database of known reactive metabolites to predict the likelihood that a compound of interest might produce reactive metabolites.

QSAR and QSPR Models: Quantitative structure–activity relationships (QSAR) are mathematical equations relating pharmacological activity to chemical structure for a series of structurally related compounds. Quantitative structure–property relationships (QSPR) relate physicochemical properties of compounds to their structures. QSARs are derived using regression and pattern recognition techniques.
**Scaffold Hopping:** Computational approaches that use a set of known active compounds to find structurally novel compounds with chemically completely different core structures, and yet binding to the same receptor by modifying the central core structure of the molecule.

**Site of Metabolism Prediction:** Given the structure of a lead compound, to predict the sites that are prone to metabolic activity. Databases of known reactivity and/or principles of chemical reactivity are employed to predict sites of metabolism. If metabolizing enzyme is identified, then its protein structure is also used for getting poses of a compound of interest at the active site of the enzyme. Machine learning and semiempirical quantum chemical calculation can also be incorporated into prediction models.

**Virtual Screening:** A computational technique used in early drug discovery for rapid *in silico* assessment of large libraries of chemical structures against three-dimensional structure of a target protein in order to identify structures that are most likely to bind to a target protein.

**REFERENCES**