1

INTRODUCTION

1.1 A DIFFICULT PROBLEM

Let us take an incredibly simplified view of the statistics of drug design. There are an estimated 35,000 open reading frames in the human genome, which in turn generate an estimated 500,000 proteins in the human proteome. About 10,000 of those proteins have been characterized crystallographically. In the simplest terms, that means that there are 490,000 unknowns that may potentially foil any scientific effort.

The previous paragraph is far from being a rigorous analysis. However, it does illustrate the fact that drug design is a very difficult task. A pharmaceutical company may have from 10 to 100 researchers working on a drug design project, which may take from 2 to 10 years to get to the point of starting animal and clinical trials. Even with every scientific resource available, the most successful pharmaceutical companies have only one project in ten succeed in bringing a drug to market.

Drug design projects can fail for a myriad of reasons. Some projects never even get started because there are not adequate assays or animal models to test for proper functioning of candidate compounds. Some diseases are so rare that the cost of a development effort would never be covered by product sales. Even when the market exists, and assays exist, every method available may fail to yield compounds with sufficiently high activity. Compounds that are

Computational Drug Design. By David C. Young

active against the disease may be too toxic, not bioavailable, or too costly to manufacture. In all fairness, we should note that high manufacturing cost is seldom a sufficient deterrent in the pharmaceutical industry.

Sometimes the only compounds that work are already the competitor's intellectual property. This book will not be addressing intellectual property law, but we shall point out the following thumb rule of commercial product development:

> *A product does not have to be better than the competitor's product. It has to be about as good as the competitor's product and patentable under your own name.*

Biological systems are probably one of the most complex systems under study on the planet. Not surprisingly, drugs are seldom simple molecules. Most are heterocyclic, are of moderate molecular weight, and contain multiple functional groups. As such, the challenges of organic synthesis are sometimes as great as the challenge of determining what compounds should be synthesized. In the pharmaceutical industry, the answer is often to synthesize all possible derivatives within a given family of compounds.

In the course of computational drug design, researchers will find themselves tasked with solving a whole range of difficult problems, including efficacy, activity, toxicity, bioavailability, and even intellectual property. With the total drug development process costing hundreds of millions of dollars, and enormous amounts of money being spent daily, drug design chemists can be under incredible pressure to produce results. As such, it is necessary to effectively leverage every computational tool that can help to achieve successful results. This book has been written to give a solid understanding of the whole range of available computational drug design tools.

1.2 AN EXPENSIVE PROBLEM

There have been a number of published estimates of how much it costs to bring a drug to market. Recent estimates have ranged from $300 million to $1.7 billion. A single laboratory researcher's salary, benefits, laboratory equipment, chemicals, and supplies can cost in the range of $200,000 to $300,000 per year. Some typical costs for various types of experiments are listed in Table 1.1.

Owing to the enormous costs involved, the development of drugs is primarily undertaken by pharmaceutical companies. Indeed, the dilution of investment risk over multiple drug design projects pushes pharmaceutical companies to undertake many mergers in order to form massive corporations.

TABLE 1.1 Typical Costs of Experiments

Experiment	Typical Cost per Compound ($)
Computer modeling	10
Biochemical assay	400
Cell culture assay	4,000
Rat acute toxicity	12,000
Protein crystal structure	100,000
Animal efficacy trial	300,000
Rat 2-year chronic oral toxicity	800,000
Human clinical trial	500,000,000

Only rarely are drugs taken all the way through the approval process by academic institutions, individuals, government laboratories, or even small companies. In 1992, out of the 100 most prescribed drugs, 99 were patented by the pharmaceutical industry.

1.3 WHERE COMPUTATIONAL TECHNIQUES ARE USED

There is no one best computational drug design technique. Many techniques are used at various stages of the drug design project. At the beginning of a project, cheminformatics techniques are used to select compounds from available sources to be assayed. Once some marginally active compounds are found, relatively broad similarity searching techniques are used to find more compounds that should be assayed. As larger collections of more active compounds are identified, the computational chemists will shift to successively more detailed techniques, such as QSAR, pharmacophore searching, and structure-based drug design tools such as docking. A computational chemist may make their reputation by being a world-class expert at the design or use of one of these techniques. However, a functional knowledge of how to work with many of them is usually necessary in order to be successful as a computational chemist in the pharmaceutical industry.

The simplest form of drug design is to start with a marginally active compound, and then make slightly modified derivatives with slightly different functional groups. However, this type of trial-and-error modification of molecules is a "blind man's bluff" game, until you see how those molecules fit in the active site and interact with the protein residues. Thus, the majority of the time that researchers are designing structures "by hand" today, they do so by examining the way that the compounds fit in the target's active site as displayed through three-dimensional computer rendering. Once a compound has been built within such computer programs, it is easy to subsequently test how strongly it will bind in the active site using computational techniques such as docking.

Computational techniques provide other options for understanding chemical systems, which yield information that is difficult, if not nearly impossible, to obtain in laboratory analysis. For example, quantum mechanically computed reaction coordinates can show the three-dimensional orientation that species adopt at each step of a reaction mechanism. Likewise, they can show exactly where the unpaired spin density is located at each point along a reaction coordinate. This is of particular concern in drug design, since enzymes often catalyze reactions by holding species in the preferred orientation, and sometimes include a mechanism to provide for necessary electron or hydrogen transfer.

Computer simulations are less costly per compound than any laboratory test, as illustrated in Table 1.1. Because of this cost-efficiency, large databases of compounds are often tested in software. Many of these compounds will never see laboratory testing of any sort. Indeed, many compounds are designed and tested in software, but never synthesized at all, owing to poor results *in silico* (computer calculations). Likewise, the use of computational techniques to choose compounds for testing results in an enrichment, meaning that a higher percentage of the compounds that are tested are active.

In today's world of mass synthesis and screening, the old practice of sitting down to stare at all of the chemical structures on a single sheet of paper is hopeless. Drug design projects often entail having data on tens of thousands of compounds, and sometimes hundreds of thousands. Computer software is the ideal means for sorting, analyzing, and finding correlations in all of this data. This has become so common that a whole set of tools and techniques for handling large amounts of chemical data have been collectively given the name "cheminformatics."

The problems associated with handling large amounts of data are multiplied by the fact that drug design is a very multidimensional task. It is not good enough to have a compound that has the desired drug activity. The compound must also be orally bioavailable, nontoxic, patentable, and have a sufficiently long half-life in the bloodstream. The cost of manufacturing a compound may also be a concern—less so for human pharmaceutics, more so for veterinary drugs, and an extremely important criterion for agrochemicals, which are designed with similar techniques. There are computer programs for aiding in this type of multidimensional analysis, optimization, and selection.

Most importantly, drug design projects may fail without the efforts of experts in computational modeling. Drug design is such a difficult problem that every relevant technique is often utilized to its best advantage. Computational modeling techniques have a long history of providing useful insights, new suggestions for molecular structures to synthesize, and cost-effective (virtual) experimental analysis prior to synthesis.

Thus, computational drug design techniques play a valuable role in pharmaceutical research. This role makes computational techniques an important part of a successful and profitable drug design process.

BIBLIOGRAPHY

Cost of Drug Development

DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: New estimates of drug development costs. J Health Econ 2003; 22: 151–185.

Gilbert J, Henske P, Singh A. Rebuilding Big Pharma's business model. In Vivo: The Business & Medicine Report 2003; 21(10).

Klaassen CD. Principles of toxicity. In: Klaassen CD, Amdur MO, Doull J, eds. The Basic Science of Poisons. Columbus, OH: McGraw-Hill; 1986. p. 11–32.

Tufts Center for the Study of Drug Development. Total cost to develop a new prescription drug, including cost of post-approval research, is $897 million. Available at http://csdd.tufts.edu/NewsEvents/RecentNews.asp?newsid=29. Accessed 2006 Apr 13.

Computational Drug Design Process in General

Boyd DB. Drug design. In: von Ragué Schleyer P et al., editors. Encyclopedia of Computational Chemistry. New York: Wiley; 1998. p. 795–804.

Additional references are contained on the accompanying CD.