Available online at www.sciencedirect.com
ScienceDirect
Advanced Drug Delivery Reviews 58 (2006) 1409–1430
Advanced DRUG DELIVERY Reviews
www.elsevier.com/locate/addr

Application of data mining approaches to drug delivery ☆

Sean Ekins [a,b,*], Jun Shimada [c], Cheng Chang [b]

[a] *ACT LLC, 1 Penn Plaza-36th Floor, New York, NY 10119, USA*
[b] *Department of Pharmaceutical Sciences, University of Maryland, 20 Penn Street, Baltimore, MD 21201, USA*
[c] *Columbia University, Department of Chemistry, 3000 Broadway, New York, NY 10027, USA*

Received 15 August 2006; accepted 4 September 2006
Available online 22 September 2006

Abstract

Computational approaches play a key role in all areas of the pharmaceutical industry from data mining, experimental and clinical data capture to pharmacoeconomics and adverse events monitoring. They will likely continue to be indispensable assets along with a growing library of software applications. This is primarily due to the increasingly massive amount of biology, chemistry and clinical data, which is now entering the public domain mainly as a result of NIH and commercially funded projects. We are therefore in need of new methods for mining this mountain of data in order to enable new hypothesis generation. The computational approaches include, but are not limited to, database compilation, quantitative structure activity relationships (QSAR), pharmacophores, network visualization models, decision trees, machine learning algorithms and multidimensional data visualization software that could be used to improve drug delivery after mining public and/or proprietary data. We will discuss some areas of unmet needs in the area of data mining for drug delivery that can be addressed with new software tools or databases of relevance to future pharmaceutical projects.

Keywords: Data mining; Drug delivery; QSAR; Pharmacophore; Networks; Transporters; Modeling; Predictions; Systems biology; Databases

Contents

☆ This review is part of the *Advanced Drug Delivery Reviews* theme issue on "Computational Drug Delivery", Vol. 58/12-13, 2006.

* Corresponding author. ACT LLC, 1 Penn Plaza-36th Floor, New York, NY 10119, USA. Tel.: +1 269 930 0974; fax: +1 269 983 7654.
E-mail addresses: ekinssean@yahoo.com, sekins@arnoldllc.com (S. Ekins).

0169-409X/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.addr.2006.09.005

1. Introduction

Drug discovery and development are said to be composed of distinct decision gates [1,2]. Key questions can be asked of a drug candidate and answers are usually provided by pivotal experimental studies. It is well known that the metabolic transformations of pharmaceuticals and other xenobiotics profoundly impact bioavailability, efficacy, chronic toxicity, excretion rate and route. Both the parent molecule and the products of metabolic transformations may also interfere with other co-administered compounds. At the molecular level, a coordinated system of transporters, channels, receptors and enzymes acts as gatekeepers to foreign molecules. This may affect the absorption, metabolism, excretion and toxicology (ADME/Tox) properties of a molecule in humans, ultimately influencing decisions for optimal drug delivery [3] and achieving clinical success. Whether for individual therapeutic areas [4] or in total, preclinical studies investigating a molecule's physiochemical properties [2] can facilitate the exploration of medicinal chemistry space for optimal ADME/Tox properties that are compatible with available drug delivery options. The routes of drug delivery whether oral, buccal, intravenous, subcutaneous, inhalation, transdermal, intramuscular, ocular or suppository are selected primarily based on the physicochemical properties of the molecule and the site of action.

With advances in computational chemistry and the rapid accumulation of empirical data, these properties can now be predicted computationally [5–8]. Computational methods for drug discovery range from molecule docking to virtual screening methods that may use molecular descriptors and machine learning algorithms [9–11]. Computational predictions provide a way to rapidly make decisions, prototype, innovate and, possibly most important of all, learn from failure. Currently, within the pharmaceutical industry, computers are used primarily for bioinformatics, systems biology, chemoinformatics, drug design, drug metabolism and toxicology assessment, pharmacokinetics/

Table 1
Examples of aspects of pharmaceutical research and development where computer algorithms can be readily used to make predictions or be used for data mining

Bioinformatics	Financial/Business case	Computer-aided drug design	Molecule selection	Clinical trials and post-marketing
Prior art publications/ patent	Drug life cycle	Ease of X-ray structure determination	Dosing route preferred	Label warnings
Homologous targets	Population demographics	Binding site characteristics	Whole cell/ organ effects	Time for review
Public/private Target discovery	Marketing	Protein folding	Novelty of molecule	Pharmacogenetics
Complex systems analysis	Blockbuster status	Flexibility of protein	Synthesizability/ cost of goods	Regulatory response
Target expression site	Urgency/disease projection	Disordered regions	Formulation/ polymorph	Clinical response
Public/private genome/ protein databases	Pharmacoeconomics	Ligand refinement	ADME/Tox/biopharm properties	Project planning/ organization
		Target QSAR	Target affinity	Adverse Response/DDI
		Size/shallowness of binding site	Drug likeness	PK/PD and trial simulation
				Dosing level
				Pharmacoeconomics