ELSEVIER

European Journal of Pharmaceutical Sciences 17 (2002) 183–193

EUROPEAN JOURNAL OF PHARMACEUTICAL SCIENCES

www.elsevier.nl/locate/ejps

Congress Report

In silico prediction of ADME and pharmacokinetics Report of an expert meeting organised by COST B15☆

Alan Boobis[a], Ursula Gundert-Remy[b], Pierre Kremers[c], Panos Macheras[d], Olavi Pelkonen[e,*,1]

[a]*Section on Clinical Pharmacology, Imperial College, London, UK*
[b]*Assessment of Chemicals, Bundesinstitut für Gesundheitlichen Verbraucherschutz und Veterinärmedizin, Berlin, Germany*
[c]*Advanced Technology Corporation, University Hospital, Institute of Pathology, B23, University of Liege, B-4000 Sart Tilman, Belgium*
[d]*School of Pharmacy, University of Athens, Panepistimiopolis, Athens 15771, Greece*
[e]*Department of Pharmacology and Toxicology, University of Oulu, PO Box 5000, FIN-90014 Oulu, Finland*

Received 13 June 2002; received in revised form 8 August 2002; accepted 30 August 2002

Abstract

The computational approach is one of the newest and fastest developing techniques in pharmacokinetics, ADME (absorption, distribution, metabolism, excretion) evaluation, drug discovery and toxicity. However, to date, the software packages devoted to ADME prediction, especially of metabolism, have not yet been adequately validated and still require improvements to be effective. Most are 'open' systems, under constant evolution and able to incorporate rapidly, and often easily, new information from user or developer databases. Quantitative in silico predictions are now possible for several pharmacokinetic (PK) parameters, particularly absorption and distribution. The emerging consensus is that the predictions are no worse than those made using in vitro tests, with the decisive advantage that much less investment in technology, resources and time is needed. In addition, and of critical importance, it is possible to screen virtual compounds. Some packages are able to handle thousands of molecules in a few hours. However, common experience shows that, in part at least for essentially irrational reasons, there is currently a lack of confidence in these approaches. An effort should be made by the software producers towards more transparency, in order to improve the confidence of their consumers. It seems highly probable that in silico approaches will evolve rapidly, as did in vitro methods during the last decade. Past experience with the latter should be helpful in avoiding repetition of similar errors and in taking the necessary steps to ensure effective implementation. A general concern is the lack of access to the large amounts of data on compounds no longer in development, but still kept secret by the pharmaceutical industry. Controlled access to these data could be particularly helpful in validating new in silico approaches.

Keywords: Pharmacokinetics; Metabolism; ADME; In silico; Prediction

1. General introduction

Drug discovery and development are expensive undertakings. The research costs for a compound increase dramatically as it progresses through clinical development, and therefore there are economic reasons for identifying and discontinuing the development of poor drug candidates at the earliest possible time. Even compounds that do

☆Based on a COST B15 meeting held in Berlin, 30 November–1 December, 2001. COST is the acronym for European Cooperation in the Field of Scientific and Technical Research. COST Action B15 was set up at the end of 1998 and is entitled 'Modelling during drug development'. Other members of the COST Action B15 Working Group 1 are: Julio Benitez (Spain), Kim Brosen (Denmark), Maria Durisova (Slovak Republic), Jaime Kapitulnik (Israël), Jirina Martinkova (Czech Republic), José A. Guimares Morais (Portugal) and Gian Maria Pacifici (Italy). The invited experts were: Professor Per Artursson (SE), Professor Bas Blaauboer (NL), Professor Matt Bogdanffy (USA), Dr Thierry Lavé (CH), Dr Phil. Lowe (CH), Dr Neil Parrott (CH), Dr Patrick Poulin (CH), Professor Matt Segall (UK), Professor Bernard Testa (CH), Dr Peter Theil (CH), and Professor Nico Vermeulen (NL). Furthermore, the following software experts participated in the meeting: Dr Alan Beresford (UK; Camitro), Dr Mike B. Bolger (USA; Simulations Plus), Dr Jan Langowski (UK; Lhasa), Dr Laura Robinson and Glen Leesman (USA; Lion Bioscience), Dr Akos Papp (Hu; Comgenex), Dr Kristina Racz (HU; Comgenex), Dr Mariana Vaschetto (UK; Accelrys). Dr Patricia Crivori (I) and Dr Andreas Reichel (DE) participated as observers.

*Corresponding author. Tel.: +358-8-537-5230; fax: +358-8-537-5247.
E-mail address: olavi.pelkonen@oulu.fi (O. Pelkonen).

[1]On behalf of all of the members of Working Group 1 of COST Action B15 and of the invited COST B15 experts.

0928-0987/02/$ – see front matter
PII: S0928-0987(02)00185-9

eventually reach the market sometimes have less than ideal characteristics, complicating patient management. Poor pharmacokinetic properties are one of the main reasons for terminating the development of drug candidates (Prentis et al., 1988; Kennedy, 1997; Venkatesh and Lipper, 2000). The application of computational technology during drug discovery and development offers considerable potential for reducing the number of experimental studies required for compound selection and development and for improving the success rate. In this context, in silico approaches are being used today in drug discovery to assess the ADME (absorption, distribution, metabolism, excretion) properties of compounds at the early stages of discovery/development. The need for early consideration of ADME properties is also increasingly urgent because of the implementation of combinatorial chemistry and high-throughput screening, since this can generate vast numbers of potential lead compounds (Hodgson, 2001). The early assessment of ADME properties will help pharmaceutical scientists to select the best candidates for development as well as to reject those with a low probability of success. This report deals with the state of the art of the theoretical and experimental aspects of determining ADME properties, as well as in silico methods for their prediction, either described in the literature and/or currently available commercially.

COST is the acronym for European Cooperation in the Field of Scientific and Technical Research. COST Action B15 was launched at the end of 1998 and is entitled 'Modelling during drug development' (Aarons et al., 2001). To help realise its objectives, COST Action B15 organised an expert meeting on 'In silico prediction of ADME and pharmacokinetics' on November 30 and December 1st, 2001 in Berlin.

The purpose of this meeting was to examine the possibilities offered by computer assisted modelling to predict absorption, distribution and metabolism of drugs and new chemical entities. A fourth session was devoted to PBPK modelling, in order to evaluate the possibilities of integrating all of these different aspects to predict pharmacokinetic outcomes de novo. Several well-known experts in the field joined software producers to review the state of the art in this fast evolving subject. The questions examined ranged from ' where we are now? ' to 'where do we want to go?', with the objective of analysing the failures as well as the successes, the advantages and the difficulties generated by these new approaches, and the reliability and robustness of selected commercial programs in predicting the ADME characteristics of chemicals. Producers of pertinent software were asked to give brief presentations on their products and to describe clearly their scientific basis. Live demonstrations were organised allowing all participants to interact directly with these models. Taking advantage of this presentation, members of COST B15 were able to produce a documented evaluation of these software packages.

The ultimate goal of the in silico prediction of ADME properties is the accurate prediction of the in vivo pharmacokinetics of a potential drug molecule in man, whilst it exists as only a virtual structure. This requires an integrated suite of models covering each of the processes involved and their incorporation into a full 'drug design' software package which combines ADME predictions with those for pharmacological properties, stability, chemical tractability, etc., to produce a molecule with the optimal combination of properties. This ideal, complete package does not yet exist. However, many of its component parts are already available and in some circumstances used for lead optimisation, screening, drug design and development.

2. Absorption

Drug absorption from the gastrointestinal (GI) tract is very complex. A large number of factors, which can be classified into three categories, i.e. physicochemical, physiological, and formulation related, affect GI absorption. Since formulation related factors are usually optimised experimentally while physiological factors cannot be controlled, prediction interests are centered on the extent of absorption as a function of physicochemical properties of the compounds.

Due to the complexity of the GI tract both in structure and function, its correct representation in the models utilised to predict GI absorption was and still is one of the major obstacles. The first and simpler models used (Suzuki et al., 1970a,b; Ho et al., 1972; Dressman et al., 1985; Macheras and Symillides, 1989; Boxenbaum, 1999) ignore the structure of the GI tract, assume a 'pseudo-steady state' and utilise the physicochemical properties of the compound in conjunction with the 'pH-partition' hypothesis (Hogben et al., 1959) to predict the fraction of dose absorbed. However, the quantitative and mechanistic interpretation of drug absorption was fostered, when mass balance approaches in a homogeneous (Sinko et al., 1991; Oh et al., 1993; Amidon et al., 1995; Yu and Amidon, 1999; Norris et al., 2000; Stoll et al., 2000) or heterogeneous tube (Kalampokis et al., 1999a,b) mimicking the GI tract were implemented. The use of these models for the analysis of transit, dissolution and drug uptake processes in the GI tract revealed that the solubility and the intestinal permeability of drug are the most important properties that determine absorption after oral administration (Sinko et al., 1991; Oh et al., 1993). This realisation led to the development of a biopharmaceutics classification scheme (Amidon et al., 1995) in which substances are classified into four categories according to their high or low solubility and permeability.

Although experimental and computational screening models are available for the calculation or prediction of solubility and intestinal permeability (Bergstrom et al., 2002; Stenberg et al., 2001), their use in predicting oral

absorption is a difficult task. This is because both properties are pH dependent and on top of that they should be considered in the dynamically changing and complex environment of the GI tract. Almost all of the computational approaches currently used to predict absorption are based on the assumption that absorption is passive, and can be predicted from molecular descriptors of the compound. No account is taken of active transport processes, including both uptake and efflux transporters, and currently it is not known how many compounds are actually actively processed in the gut. A recent review lists 35 substrates, 20 inhibitors, and 12 inducers for the human P-glycoprotein (Kim, 2002). Whilst, for most compounds, this is not thought to represent an important limitation, ignoring it will inevitably lead to some anomalous predictions. Two software packages, namely GastroPlus™ and iDEA™ are available commercially for predicting human intestinal fraction absorbed based on estimates of solubility and intestinal permeability (Table 1).

GastroPlus™ utilises inputs obtained from a parallel computer program QMPRPlus™ which generates estimates for a new chemical entity for lipophilicity ($\log P$), effective permeability, apparent permeability, diffusivity, and water solubility from its chemical structure. The estimates of QMPRPlus™ are derived from correlation models utilising a variety of data from human and in vitro studies in conjunction with primary molecular descriptors of the chemical structures. The GastroPlus™ program simulates oral absorption on the basis of an advanced version (Agoran et al., 2001) of the compartmental absorption-transit model (Yu and Amidon, 1999) and provides estimates of the fraction of dose absorbed. The model takes into account pH-dependency of basic parameters solubility and permeability. Additional modules of the GastroPlus™ computer system allow pharmacokinetic simulation and model fitting.

The iDEA™ predictive ADME simulation system has been developed using a proprietary database of clinical trial results coupled with in vitro data characterising the drug substance (solubility, Caco-2 permeability, protein binding and metabolic stability in human cryo-preserved hepatocytes). The iDEA™ program is described in the literature (Grass, 1997; Norris et al., 2000) and is based on the STELLA (Structural Thinking Experimental Learning Laboratory with Animation) simulation software. A physiologically based absorption model defining each intestinal segment as a separate compartment is utilised for the description of fluid movement in the GI tract with a calculation of drug absorption (flux) in each intestinal segment over time. The summation of the flux calculations in each segment gives the total absorption rate. The absorption model is coupled with a physiological metabolism model, which provides estimates for the rate and extent of first pass metabolism in humans. The combined system allows the prediction of bioavailability for a compound from in vitro data.

Neither the GastroPlus™ nor the iDEA™ system has been rigorously evaluated, nor can they be. This is because a large, well-validated data bank for fraction of dose absorbed is not currently available; besides, most of the variability, which is observed in the existing data originates from physiological factors. Despite the importance of physiology in drug absorption, the physiologically relevant part of the models in both systems is kept to a minimum because of the complexity of drug processes in the varying physiological milieu. However, from the limited and unpublished evaluation of the systems one can infer some conclusions. Their predictive ability on qualitative grounds (pure in silico classification as low, medium and high fraction of dose absorbed) for 28 drugs, was found to be similar, i.e. 68–79% of correct classification and RMSE (root-mean-square error) values ranging from 19 to 24% depending on the nature of input data. The RMSE values are high when compared to the 14–16% found in relevant QSAR studies involving large numbers of compounds (Wessel et al., 1998; Zhao et al., 2001). As far as the user interface and functionality of the two systems is concerned, these are as follows. The GastroPlus™ is well designed and suitable for expert users, allows comprehensive access to model parameters, good combination with structure-based prediction in QMPRPlus™ and has batch capabilities for handling of multi-structures (up to thousands). The iDEA™ system has a simple structure and is too simple for expert users. Once the server is installed, corporate-wide deployment is simple and training is not a big issue, a structure-based prediction capability is not available, structure import/export capabilities need to be improved, batch import capability must be simplified and addition of batch export is essential.

Overall, in silico approaches for predicting oral drug absorption are in their infancy. The animal models currently used for the preliminary assessment of drug absorption in humans cannot be replaced with in silico approaches in their present state. The use of animal data in conjunction with in silico approaches for quantitative prediction, after appropriate optimisation, of oral drug absorption in humans is one of the most promising avenues in this field of research. Advances along this line will certainly facilitate the development and enhance the validity of in silico methods utilising pure in vitro data for predicting oral drug absorption in humans.

3. Prediction of distribution

Tissue distribution is an important determinant of the pharmacokinetic (PK) profile of a drug. Hence, in drug development, the prediction of tissue distribution would help predict the in vivo PK of a compound prior to any experiments in animals or man. An understanding of tissue distribution, particularly when coupled with knowledge of the in vitro effects of the compound on biological targets

Table 1
In silico prediction of ADME and pharmacokinetics: software evaluation

	Accelrys, Cambridge (UK) Dr M. Vaschetto www.accelrys.com	Camitro, Cambridge (UK) Dr A. Beresford www.camitro.com	Comgenex, Budapest (HU) Dr A. Papp www.comgenex.com	Lhasa, Ltd Leeds (UK) Dr J. Langowski www.chem.leeds.ac.uk/luk/meteor	LionBioscience, Inc., San Diego (USA) Dr G. Leesman www.lionbioscience.com	Simulations Plus, Inc., Lancaster (USA) Dr M. Bolger www.simulations-plus.com
Program(s)	Suit of in silico tools	In-house server based software suite	Suit of in silico tools, including MetabolExpert	Meteor	iDEA™ Predictive ADME Simulation System	QMPRPlus GastroPlus
Purpose and/or function	In silico prediction of ADME properties for designing drug-like virtual libraries	Computational models for prediction of solubility ADME: intestinal absorption, blood–brain barrier penetration, CYP metabolism (currently by 3A4, 2D6 and 2C9) and CYP inhibitory potential (currently K_i values for 2D6 and 2C9)	Predicting ADME/Tox for lead optimisation (MetabolExpert is described below)	Prediction of the metabolic fate of a query chemical structure (real or virtual)	The Absorption Module: predicts permeability and fraction dose absorbed (FDp) and FDp over time. The Metabolism Module: predicts bioavailability, linked to the Absorption Module for inputs.	QMPRPlus: Biopharmaceutical property estimation: solubility, permeation, absorption and distribution. GastroPlus: Simulations and predictions of GI dissolution, transit, absorption, bioavailability, and PD. Predictions of first-pass effects in gut and liver and plasma conc-time profiles.
Scientific basis	Correlations of some basic physicochemical properties (log *P*, PSA) with absorption, blood–brain barrier penetration and aqueous solubility	Surface properties and the electronic properties of the molecule taking into account theoretical energy differences by the reaction/diffusion	A knowledge base of structure-metabolism rules	A knowledge base of structure-metabolism rules together with a reasoning engine	A knowledge based model based on over 70 substances from 30 different therapeutic classes	QMPRPlus: Ensemble of artificial neural networks, and PLS models based on literature and proprietary data. GastroPlus: Physiologically-based mechanistic advanced compartmental absorption and transit model.
Nature of the software	Models (passive intestinal absorption model, blood–brain barrier penetration model, solubility model) based on correlations	Models. Absorption: diffusion model Blood–brain barrier: partition coefficient. Metabolism: potential sites of metabolism by CYP 3A4 (currently no binding affinity for this enzyme), metabolism and inhibitory constants for/by CYP 2D6 and 2C9	Rules based on examples from the scientific literature and on the basis of possible sites and restrictions from the compound under study (for animals: ~180 rules, for plants: ~240 rules, photodegradation: over 300 rules)	The reasoning model takes into account the lipophilicity (log *P* estimate) and the most likely metabolites generated	The iDEA chemical structure absorption model utilises chemical structure to predict various absorption properties The iDEA Physiological Absorption Model uses in vitro data to construct a physiological model to predict FDp over time, mass absorbed, soluble mass, insoluble mass, absorption rate, and intestinal drug concentration. The iDEA Physiological Metabolism Model predicts bioavailability from metabolic turnover and protein binding data	The computer applications QMPRPlus™ and GastroPlus™ are products of Simulation Plus Inc., designed to run under all Windows operating systems. QMPRPlus: Biopharmaceutical property estimation GastroPlus: GI Simulation linked to PK/PD models.

Table 1. Continued

Required data	Physicochemical properties	Structure of the molecule	Chemical structure of the compound	Chemical structure of the compound in any one of several possible formats	For structure-based absorption model: chemical structure or MOL file. For physiological absorption model: chemical structure, dosage, solubility, permeability (predicted or measured in Caco-2 cells).	Depends on mode of operation: (1) Purely in silico: (QP+GP): 2D or 3D structures in any one of multiple formats; (2) In vitro: permeability, solubility, pK_a, K_m and V_{max} for metabolism and/or transport and efflux. (3) In vivo: same as in vitro plus Cp vs. time and/or pharmacological response vs. time required.
Performance	Can be used for ADME-properties based selection of compounds in virtual libraries	Gives relevant physicochemical properties, predicts absorption, BBB, metabolism and inhibition by 2D6 and 2C9 and metabolism by 3A4	Gives potential phase I and II metabolites	Single metabolites and a metabolic tree	Used to identify ADME liabilities early in the drug discovery process. Compound training set consists of compounds representing a large number of therapeutic classes; iDEA has a very large database generated for predictive ADME	Purely in silico: ADME-based selection of compounds in virtual libraries. In vitro: modeling of pre-clinical animal data and scale-up to human. Predicts bioavailability. In vivo: modeling of formulation changes, SUPAC, iviv correlations, BCS classification, controlled release.
Predictive power	Not (extensively?) validated	Not (extensively?) validated	Not (extensively?) validated	Not (extensively?) validated	Absorption model has been externally validated	QMPRPlus™ and GastroPlus have been extensively validated against external test sets.
User friendliness	Needs considerable training?	Is used only by the developer to whom the chemical structures are to be sent	User has several options to interact with model	User has several options to interact with model	Easy to use; improvements have been made to the user interface	Easy for the user to learn and use. Incorporates a good graphical user interface and extensive help.
Flexibility	ADME descriptors can be employed at will in the selection of libraries	Not applicable	Linked to several tools for prediction of physicochemical parameters and to a in-house produced 'ex silico' approach	Can be linked to DEREK (toxicity prediction system). Extensible by linking to other software packages	iDEA modules are integrated to be a system for predicting ADME liabilities, but modules can be used separately	Flexible, can be used from early discovery through clinical trials with inputs from purely in silico, in vitro, and in vivo data.
Evolution possibilities	?	Yes	New rules on the basis of the literature or from other sources can be added to the model	New rules on the basis of the literature or from other sources can be added to the model	Continuous upgrades to the models and to the user interface	QMPRPlus is customisable outside the executable code. GastroPlus: User can recalibrate with in-house data.

of toxicological or pharmacological relevance, may also help in predicting the pharmacodynamic or toxicodynamic effects of a drug in specific tissues. Thus, in silico/in vitro tools enabling the prediction of the PK profile would be invaluable for high-throughput screening and selection of compounds for in vivo testing.

Currently, there are several methods available to predict tissue distribution. These predict either tissue:plasma ratios or the volume of distribution at steady state (V_{ss}). All are based on the assumption of passive diffusion between tissue compartment. Whilst this might be true for many compounds, there are exceptions, where active influx or efflux transport can play an important role in determining V_{ss}.

3.1. Tissue composition based approaches

These are based on mechanistic principles and make use of both in silico and in vitro outputs. From molecular descriptors (or data from in vitro experiments, e.g. $P_{oct-water}$ or $P_{olive\ oil-water}$) relevant physicochemical properties are estimated, such as lipophilicity/hydrophobicity (Poulin et al., 2001) and pK_a. In some approaches, plasma protein binding (determined in vitro) is also taken into consideration (Lombardo et al., 2002). Physiological information on tissue composition (lipid/water/protein fraction), the blood composition (lipid/water/protein) and blood flow to the tissues is utilised to develop a partitioning model (Poulin and Theil, 2000, 2002a,b).

Such approaches have been evaluated in the pharmaceutical industry and have been shown to be of considerable value in early drug discovery for predicting V_{ss} and tissue distribution. Using this type of approach on a data set comprising 23 substances (MW between 180 and 630 and $\log P$ between -0.02 and 8.84), V_{ss} for all of the compounds, with one exception, was predicted within a factor of 2-fold of the V_{ss} derived from in vivo single dose pharmacokinetic testing. V_{ss} of the one outlier was only slightly outside the 2-fold limit. In a study of 123 structurally unrelated compounds, with a broad range of physicochemical properties, there was an 80% success rate (i.e. within 0.5–2.0-fold of the correct value, as determined in vivo) in predicting V_{ss}. This could be improved by including additional considerations for certain classes of compound (Lombardo et al., 2002). In another evaluation, of toxic compounds, tissue concentrations were predicted reasonably well.

3.2. Semi-empirical/in vivo based approach

This method utilises in vivo kinetic data. The principle is to partition the volume of distribution at steady state, as determined by in vivo experimentation, to predict individual tissue:plasma ratios. In the experience of scientists in industry this approach has medium usefulness in early drug discovery.

3.3. QSAR/rule based approaches

These incorporate rules, derived from the analysis of the relationship between physicochemical data/properties/structures and experimental data on V_{ss} or tissue:plasma ratios from in vivo studies with specific groups of compounds. Experience in the pharmaceutical industry to date has shown these approaches to be of very limited value in early drug discovery.

By comparing the predictions from such in silico/in vitro approaches with the results of in vivo testing, it is possible to increase knowledge of the behaviour of the compound and to enhance understanding of its properties. In this context, drug distribution is an important component of a mechanistic PBPK model. Among the programs assessed in this occasions, those of Camitro, Lion-Bioscience and Simulations Plus have some features to predict drug distribution (Table 1).

4. In silico prediction of drug metabolism

Within the ADME processes, M, which stands for metabolism, certainly covers the largest, and still poorly understood, aspect and consequently the most difficult to evaluate and to predict. In fact, the metabolic fate of a compound depends on a large number of variables related to both the chemical itself (chemical structure, physicochemical properties, etc.) and the biological system (enzyme and its environment) (Kumar and Surapaneni, 2001). In this field, it is particularly difficult to develop reliable prediction software packages. The existing attempts are still relatively ‘crude’ products under constant evolution and needing continuous refinement (Darvas et al., 1999). There are several aspects of metabolic behaviour that one might wish to predict in silico, the most relevant ones being outlined in the table below (Ekins et al., 2000; Ekins and Wrighton, 2001).

4.1. Expected predictions

Aspects of metabolic behaviour which would be useful to predict in silico, are as follows:

Biotransformation	Chemical structure of single metabolites Metabolic tree Most probable metabolic route Warnings for possible ‘toxic’ intermediates, adduct formation
Binding to enzymes	Identification of concerned enzymes (CYP) Inhibition of these enzymes Induction of these enzymes
Catalytic reaction	Rate of metabolism Extent of metabolism Mechanism of reaction
Possible drug–drug interactions	Inhibition Induction Competition for a receptor

There have been two main scientific approaches to the in silico prediction of drug metabolism. The first is based on a consideration of physicochemical properties of the molecule itself, often utilising structure–activity relationships. The second is based on knowledge of the structure of the enzyme and/or its mechanism of action. Most recently, approaches are being developed that incorporate aspects of both.

1. Descriptors of the chemical properties of a compound include:

- Molecular sites sensitive to oxidation or conjugation reactions
- 3-D structure of the chemical, steric hindrance, etc.
- Molecular surface properties
- Electronic structure (distribution of electric charges on the molecule)
- Quantum mechanics
- Polarity
- Hydrophobicity
- Lipophilicity (log *P,* log *D*)
- Hydrogen bonding capacities
- 3D–molecular interaction fields (electrostatic, spin, molecular surface, molecular hydrogen binding potentials, etc.) (Testa and Cruciani, 2001)

The importance of defining these different concepts correctly and precisely was emphasised. Confusion in their definition may lead to incorrect prediction or misinterpretation of the outcomes.

2. Biological or biochemical parameters include:

- Protein structure, 3D-structure of the binding or active site
- Specificity and regioselectivity of the enzyme
- Accessibility of the binding site
- Activity of the enzyme
- Reaction mechanism

4.2. Energy levels involved in the process leading to metabolism

Two different molecular interactions are involved in the processes leading to a metabolic reaction: one of these represents binding events (membrane crossing, binding to the active site) and requires low energy (5–10 kcal/mol), while the other represents the catalytic reaction and requires ten times more energy (50–100 kcal/mol).

This important difference between the required energy levels allows selection of the most relevant process to model in the prediction of metabolism; the electronic properties of the molecule are clearly important and quantum mechanical modelling can be used to predict the chemical structure of the metabolites.

Based on such considerations, Camitro have developed a quantum chemical electronic model for P450 activity. The model is based on the evaluation of the energy necessary to abstract a hydrogen atom from different groups, enabling the ease of, for example, aromatic oxidation or S-oxidation to be calculated (Jones et al., 2002).

4.3. Cytochrome P450 dependent reactions

Existing in silico models are based mainly on a knowledge base of structure-metabolism rules found in the literature, essentially taking into account the physicochemical properties of the molecule (electronic density, vulnerability of certain chemical bonds or functions, lipophilicity, and so on). They are never completely set and offer large interaction possibilities to the user in order to introduce new rules or to modulate their application. All of them are equipped with ‘filters’ to avoid improbable reactions as much as possible. Often they are able to take into account some biological conditions like strain, tissue, sex, and other biological and physiological characteristics. Some of these models incorporate experimental in vitro and/or in vivo data in order to increase the score of the prediction.

Numerous SAR and QSAR studies have been performed and are exploited to set up prediction conditions (Ekins et al., 2000). Based on the compilation of information on the shape, electronic properties and conformation of substrates, inhibitors and metabolic product it has been possible to create pharmacophore models for some enzymes (ter Laak and Vermeulen, 2001). From such models it is possible to infer a structure of the enzyme active site, to predict the metabolism and/or substrate/inhibitor selectivity. For instance, all CYP2D6 substrates have a very similar 3D structure, they are oxidised on a position situated at 5 or 7 Å from a nitrogen group that is probably locked in the active site by electrostatic binding to Asp 301 amino acid.

A majority of the initial metabolic reactions to which drugs are subjected are catalysed by cytochrome P450 enzymes. Consequently, a precise understanding of the catalytic mechanism at a molecular level should enable rules for metabolism by this class of enzyme to be devised.

The broad specificity and regioselectivity of P450 enzymes means that the electronic structure of the substrate molecule plays an important role in determining the site of metabolism. Ab initio methods can be used to identify the properties that will predict the vulnerable sites on a molecule. In trials, performed by M. Segall using semi-empirical quantum mechanical approaches the electronic model could explain more than 65% of the observed sites of metabolism due to CYP3A4. When steric influences on the active site were included in the model, this improved the accuracy of prediction to more than 80%.

Two important issues for future prediction of P450 supported biotransformations are: the presence of labile sites on the molecule and the spin distribution on this molecule and its interaction with the spin state of the catalytic site (heme).

Studies are progressing rapidly toward 3D modelling of the active site of each P450 enzyme. This could rapidly emerge as a means to predict accurately the specific enzyme that will be responsible for the metabolism of a given compound and so open a new field in the prediction of drug–drug interactions.

4.4. Strategy in software and model development

The type of model needed will depend on the questions to be resolved and therefore on the drug development stage at which the software will be used. In lead discovery and optimisation a rough evaluation of susceptibility to metabolism, the potential for enzyme inhibition, the nature of the principal metabolites and the enzymes involved is often largely sufficient. For these purposes qualitative, or at most semi-quantitative, data are sufficient. However, in the preclinical and clinical phases, more information is needed on the nature of the main and minor metabolites, their distribution and elimination, their toxicity, the enzymes and tissue involved, and the influence of genetic and other factors. Often, quantitative information is required. Therefore the approach is likely to be different from one stage to the another.

The 3D-structure of a molecule is insufficient to predict its interaction with biological targets (receptors, enzymes, membranes and other proteins). Each molecule has its own 3D molecular interaction fields, resulting from interaction with its direct surroundings. This has a major influence on some molecular properties and hence affects the site and type of metabolic reactions. For instance, comparative molecular field analysis (CoMFA), which takes into account steric, electrostatic and lipophilic forces, most correctly predicts the binding of ligands to AGP (alpha$_1$-glycoprotein) (Ekins et al., 2000; Testa and Cruciani, 2001).

The need to consider the reliability of data for analysis is emphasised. Too often, a phenomenon is described being linearly related to a parameter, whereas the relationship is, in fact, curvilinear because the original conclusions were based on an homologous series of compounds. Incorrect or inappropriate statistical analysis may lead to false interpretation. Elimination of outliers from a data series may lead to fallacious predictions.

The main problem in developing a global expert system based on metabolic rules is the large number of false positives. To avoid this as much as possible, filters are introduced into the program. These take into account parameters such as steric hindrance, lipophilicity, H-bonding, biological factors, and so on. The main disadvantage of this is is the introduction of unanticipated false negatives, and achieving the correct balance is one of the major barriers to the full implementation of such approaches at present.

4.5. Computer systems for the prediction of the metabolic pattern of a product

Two companies demonstrated computer systems to predict metabolites: Lhasa Ltd has developed METEOR and Comgenex-Compudrug METABOLEXPERT. Both systems are based on a knowledge base of structure-metabolism rules to predict the metabolic fate of a compound. The knowledge base and metabolic rules are essentially based on several extensive books of Testa and Jenner (Testa and Jenner, 1976; Jenner and Testa, 1981; Testa, 1995), and other publications. The reasoning engine in METEOR takes into account some knowledge of chemical reaction mechanisms, lipophilicity, competition between possible reactions, etc.

Both packages are easy to use, give rapid answers and are linked to a toxicity prediction system. They allow interaction of the user in the generation of the metabolic tree, namely to distinguish between phase I and phase II reactions. Introduction of proprietary rules, biotransformations, examples and literature references is possible. They usually predict many more metabolites than observed experimentally. Like most in silico systems, these packages are constantly improved and evolving.

The chemical structures of original molecules were submitted to those responsible for the presentations. Very rapidly, a metabolic tree was proposed. The score were relatively satisfactory. Nevertheless some metabolites observed in vivo were not predicted. Furthermore, the selective chemical rules applicable to steroid molecules are only very imperfectly taken into account by these software packages. Prediction of many more metabolic possibilities than actually observed is a common drawback of these packages. Developers are well aware of it and are constantly introducing new search criteria and rules allowing the use of restriction filters. The difference between the two packages analysed is essentially at this level. Filters may be based on physico-chemical properties, on species differences, on implementation of new reasoning rules and on input from the possibility of linkage with other software.

A scientist with a wide experience in drug metabolism will reach the same conclusions very easily and rapidly. Nevertheless, for a first and rapid overview of the metabolic fate of a new compound such software is a powerful tool. Moreover, principally with METEOR, a direct explanation of the underlying rule and appropriate bibliographic references are provided for each biotransformation. This

constitutes an important advantage, avoiding long bibliographic inquiries.

5. Prediction of excretion

To date, there has been very little work on the in silico modelling or prediction of excretion. Although most existing pharmaceuticals are excreted to a variable extent as unchanged compounds via the kidneys or the bile, for only a few is urinary or biliary excretion a major route of elimination. Those that are eliminated to an appreciable extent by these routes, are usually quite hydrophilic and ionisable at physiological pH. Many serve as substrates for tubular or biliary epithelial transporters. In addition, highly charged conjugate metabolites are eliminated in urine or bile, usually following excretion via active transport mechanisms. However, because such phase II metabolites are often not pharmacodynamically active, or because excretion is not rate-limiting, there has been very little interest focussed on them.

Passive excretion can theoretically be predicted using some of the approaches described above for the prediction of tissue distribution, as it is determined by similar physicochemical and physiological properties (blood flow, protein binding, lipophilicity, pK_a), possibly with different limits, e.g. glomerular filtration and molecular weight. However, in practice metabolic stability in vitro/in vivo and first animal pharmacokinetic studies would give first indications about potential significance of renal excretion route. No adequate in silico studies were known to the participants of the meeting.

6. Physiologically-based pharmacokinetic modelling

Even after predictive modelling of absorption, distribution, metabolism and excretion, there is a need to integrate this information into a coherent and predictive model of the complete behaviour of the substance under study and to perform more advanced predictions of, for example, route-dependent differences, species differences, factors contributing to variability in disposition and to interindividual variability. PBPK modeling is a promising tool for these purposes, because—in contrast to classical compartmental analysis—PBPK modeling attempts to describe the system in physiological terms that have relevance to chemical distribution, mode of action and underlying biochemical processes. Chemical-specific data, be it experimental or in silico produced (see table below), can be incorporated into PB-based mathematical descriptions of kinetic processes in a way that yields reliable predictions. In essence, PBPK modeling offers a scientifically-defensible way—instead of an educated guess—to integrate various pieces of information from in silico models, in vitro studies and other preclinical information, to evaluate the outcome under various assumptions (Oliver et al., 2001).

Basic data needs for PBPK models

Chemical-specific data	Biological data
Partition coefficients	Anatomical dimensions
Metabolic rate constants (V_{max}, K_m, K_i,	Organ blood flows
elimination rate constants)	Organ volumes
Molecular weight	Cardiac output
Aqueous solubility data	Ventilation rate
Vapour pressure	Body mass
Permeability coefficients	Level of physical activity
Diffusion coefficients	Age
Protein binding constants	Gender

A particularly useful application of PBPK modelling is in the risk assessment of industrial or environmental chemicals, where ethical considerations frequently prevent the collection of pharmacokinetic data from human subjects (Bogdanffy et al., 2001; Clewell et al., 2001). Here, risk assessment is based mostly on animal data and this has to be extrapolated to human risk assessment. PBPK modelling can incorporate experimental animal data with in silico-derived and in vitro human data into a coherent framework, from which meaningful and reliable assessments could be made.

Furthermore, PBPK modelling is an excellent tool for simulating variability at different levels, organ, organism (interindividual) and population (interethnic). Simulation tools available today allow for the incorporation of variability into different factors in the model and the predicted outcome can then serve as a framework in the design of appropriate experiments or in building safeguards for clinical trials (Jonsson and Johanson, 2001).

7. Conclusions

· Considerable progress has been made in the last few years in the development of computational approaches for the prediction of absorption, distribution and metabolism. There has been little work on the prediction of excretion per se.

· Different software tools are applicable to different stages of the drug discovery/development process and in the risk assessment of other chemicals. This needs to be recognised in assessing a package for ‘fitness for purpose’.

· Whilst several software packages have been developed for ‘in-house’ use, others are being developed for commercial purposes. Most of the latter have not yet been adequately validated, particularly those for the prediction of metabolism, and they require further improvement.

· Many metabolic packages are 'open', and are able to incorporate rapidly and often relatively easily, new information from published sources, from the user or from the developer.

· Accurate in silico prediction of oral absorption and distribution of a number of compounds is now possible. Such predictions are probably not worse than those derived from studies in vitro, with the decisive advantages of increased speed and fewer resources. Some packages can handle thousands of molecules in a few hours or even less.

· Whilst not as advanced as those for oral absorption, methods are being developed for the in silico prediction of absorption following dermal and inhalational exposure.

· Methods for the prediction of absorption and distribution generally do not take account of active transport. Hence, identification of any classes of compound for which prediction is not yet reliable would be helpful.

· In silico approaches will continue to evolve rapidly, just as in vitro methods did during the last decade. Experience with the latter showed that failure to adhere to best practice, such as elaboration of a prediction model and adequate validation, resulted in loss of confidence in their reliability and applicability, even when this was not warranted. It will be important to learn from this experience, and avoid repeating similar errors.

· Simulation is widely used in other disciplines, such as engineering. Wherever possible, in silico prediction of ADME should learn from these other disciplines.

· Pharmaceutical companies hold large amounts of data, much of it on compounds no longer of commercial interest, to which there is no access outside of the company. Progress in the development of effective in silico approaches would be greatly enhanced should controlled access to such data be made available. Some limited initiatives in this direction are already underway.

· Common experience shows that, even when reliable, there is a general lack of confidence in relying on such approaches, for essentially irrational reasons. To try to overcome this, software producers need to improve transparency and state clearly the assumptions underlying their predictive approach. Published examples of the successful application of in silico techniques would also be very helpful in promoting acceptance.

References

Aarons, L., Balant, L., Boobis, A.R., 2001. Cost B15: modelling in drug development. Br. J. Clin. Pharmacol. 52, 118–119.

Agoran, B., Woltosz, W.S., Bolger, M.B., 2001. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Adv. Drug Deliv. Rev. 50, S41–S67.

Amidon, G.L., Lennernas, H., Shah, V.P., Crison, J.R., 1995. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. Pharm. Res. 12, 413–420.

Bergstrom, C.A.S., Norinder, U., Luthman, K., Artursson, P., 2002. Experimental and computational screening models for the prediction of aqueous drug solubility. Pharm. Res. 19, 182–188.

Bogdanffy, M.S., Plowchalk, D.R., Sarangapani, R., Starr, T.B., Andersen, M.E., 2001. Mode-of-action-based dosimeters for interspecies extrapolation of vinyl acetate inhalation risk. Inhal. Toxicol. 13, 377–396.

Boxenbaum, H., 1999. Absorption potential and its variants. Pharm. Res. 16, 1893.

Clewell, H.J., Gentry, P.R., Gearhart, J.M., Allen, B.C., Andersen, M.E., 2001. Comparison of cancer risk estimates for vinyl chloride using animal and human data with a PBPK model. Sci. Total Environ. 274, 37–66.

Darvas, F., Markohazy, S., Kormos, P., Kulkami, G., Kalasz, H., Papp, A., 1999. MetabolExpert: Its use in metabolism research and in combinatorial chemistry. In: Erhardt, P.W. (Ed.), Drug Metabolism: Databases and High-throughput Screening Testing During Drug Design and Development, pp. 237–270.

Dressman, J.B., Amidon, G.L., Fleisher, D., 1985. Absorption potential: estimating the fraction absorbed for orally administered compounds. J. Pharm. Sci. 74, 588–589.

Ekins, S., Wrighton, S.A., 2001. Application of in silico approaches to predicting drug–drug interactions. J. Pharmacol. Toxicol. Methods 45, 65–69.

Ekins, S., Waler, C.L., Swaan, P.W., Cruciani, G., Wrighton, S.A., Wikel, J.H., 2000. Progress in predicting human ADME parameters in silico. J. Pharmacol. Toxicol. Methods 44, 251–272.

Grass, G.M., 1997. Simulation models to predict oral drug absorption from in vitro data. Adv. Drug Deliv. Rev. 23, 199–219.

Ho, N.F., Higuchi, W.I., Tun, J., 1972. Theoretical model studies of drug absorption and transport in the GI tract (3). J. Pharm. Sci. 61, 192–197.

Hodgson, J., 2001. ADMET—turning chemicals into drugs. Nat. Biotechnol. 19, 722–726.

Hogben, C.A.M., Tocco, D.J., Brodie, B.B., Schanker, L.S., 1959. On the mechanism of intestinal absorption of drugs. J. Pharmacol. Exp. Ther. 125, 275–282.

Jenner, P., Testa, B. (Eds.), 1981. Concepts in Drug Metabolism. Marcel Dekker, New York.

Jones, J.P., Mysinger, M., Korzekwa, K.R., 2002. Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction. Drug Metab. Dispos. 30, 7–12.

Jonsson, F., Johanson, G., 2001. A Bayesian analysis of the influence of GSTT1 polymorphism on the cancer risk estimate for dichloromethane. Toxicol. Appl. Pharmacol. 174, 99–112.

Kalampokis, A., Argyrakis, P., Macheras, P., 1999a. A heterogeneous tube model of intestinal drug absorption based on probabilistic concepts. Pharm. Res. 16, 1764–1769.

Kalampokis, A., Argyrakis, P., Macheras, P., 1999b. Heterogeneous tube model for the study of small intestinal transit flow. Pharm. Res. 16, 87–91.

Kennedy, T., 1997. Managing the drug discovery/development interface. Drug Discov. Today 2, 436–444.

Kim, R.B., 2002. Drugs as P-glycoprotein substrates, inhibitors and inducers. Drug Metab. Rev. 34, 47–54.

Kumar, G.N., Surapaneni, S., 2001. Role of drug metabolism in drug discovery and development. Med. Res. Rev. 21, 397–411.

Lombardo, F., Obach, R.S., Shalaeva, M.Y., Gao, F., 2002. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. J. Med. Chem. 45, 2867–2876.

Macheras, P., Symillides, M., 1989. Toward a quantitative approach for the prediction of the fraction of dose absorbed using the absorption potential concept. Biopharm. Drug Dispos. 10, 43–53.

Norris, D.A., Leesman, G.D., Sinko, P.J., Grass, G.M., 2000. Development of predictive pharmacokinetic simulation models for drug discovery. J. Controlled Release 65, 55–62.

Oh, D.M., Curl, R.L., Amidon, G.L., 1993. Estimating the fraction dose absorbed from suspensions of poorly soluble compounds in humans: a mathematical model. Pharm. Res. 10, 264–270.

Oliver, R.E., Jones, A.F., Rowland, M., 2001. A whole-body physiologically based pharmacokinetic model incorporating dispersion concepts: short and long time characteristics. J. Pharmacokinet. Biopharm. 28, 27–55.

Poulin, P., Theil, F.P., 2000. A priori prediction of tissue:plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery. J. Pharm. Sci. 89, 16–35.

Poulin, P., Theil, F.P., 2002a. Prediction of pharamcokinetics prior to in vivo studies. 1. Mechanism-based prediction of volume of distribution. J. Pharm. Sci. 91, 129–156.

Poulin, P., Theil, F.P., 2002b. Prediction of pharamcokinetics prior to in vivo studies. II. Generic physiologically based pharmacokinetic models of drug disposition. J. Pharm. Sci. 91, 1358–1370.

Poulin, P., Schoenlein, K., Theil, F.P., 2001. Prediction of adipose tissue:plasma partition coefficients for structurally unrelated drugs. J. Pharm. Sci. 90, 436–447.

Prentis, R.A., Lis, Y., Walker, S.R., 1988. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964–1985). Br. J. Clin. Pharmacol. 25, 387–396.

Sinko, P.J., Leesman, G.D., Amidon, G.L., 1991. Predicting fraction dose absorbed in humans using a macroscopic mass balance approach. Pharm. Res. 8, 979–988.

Stenberg, P., Norinder, U., Luthman, K., Artursson, P., 2001. Experimental and computational screening models for the prediction of intestinal drug absorption. J. Med. Chem. 44, 1927–1937.

Stoll, B.R., Batycky, R.P., Leipold, H.R., Milstein, S., Edwards, D.A., 2000. A theory of molecular absorption from the small intestine. Chem. Eng. Sci. 55, 473–489.

Suzuki, A., Higuchi, W.I., Ho, N.F., 1970a. Theoretical model studies of drug absorption and transport in the gastrointestinal tract I. J. Pharm. Sci. 59, 644–651.

Suzuki, A., Higuchi, W.I., Ho, N.F., 1970b. Theoretical model studies of drug absorption and transport in the gastrointestinal tract II. J. Pharm. Sci. 59, 651–659.

ter Laak, A.M., Vermeulen, N.P.E., 2001. Molecular modeling approaches to predict drug metabolism and toxicity: a summary. In: Testa, B., van de Waterbeemd, H., Folkers, G., Guy, R. (Eds.), Pharmacokinetic Optimization in Drug Research—Biological, Physicochemical and Computational Strategies. Wiley–VCH, Helvetica Chimica Acta, Zürich, pp. 551–588.

Testa, B., 1995. The Metabolism of Drugs and Other Xenobiotics–Biochemistry of Redox Reactions. Academic Press, London.

Testa, B., Cruciani, G., 2001. Structure–metabolism relations and the challenge of predicting biotransformation. In: Testa, B., van de Waterbeemd, H., Folkers, G., Guy, R. (Eds.), Pharmacokinetic Optimization in Drug Research—Biological, Physicochemical and Computational Strategies. Wiley–VCH, Helvetica Chimica Acta, Zürich, pp. 65–84.

Testa, B., Jenner, P., 1976. Drug Metabolism: Chemical and Biochemical Aspects. Marcel Dekker, New York.

Venkatesh, S., Lipper, R.A., 2000. Role of the development scientist in compound lead selection and optimization. J. Pharm. Sci. 89, 145–154.

Wessel, K., Jurs, P.C., Tolan, J.W., Muscal, S.M., 1998. Prediction of human intestinal absorption of drug compounds from molecular structure. J. Chem. Inf. Comput. Sci. 38, 726–735.

Yu, L.X., Amidon, G.L., 1999. A compartmental absorption and transit model for estimating oral drug absorption. Int. J. Pharm. 186, 119–125.

Zhao, Y.H., Le, J., Abraham, M.H., Hersey, A., Eddershaw, P.J., Luscombe, C.N., Boutina, D., Beck, G., Sherborne, B., Cooper, I., Platts,