ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 479–490

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Subcellular pharmacokinetics and its potential for library focusing

Stefan Balaz*, Viera Lukacova

Department of Pharmaceutical Sciences, College of Pharmacy, North Dakota State University, Fargo, ND 58105, USA

Abstract

Subcellular pharmacokinetics (SP) optimizes biology-related factors in the design of libraries for high throughput screening by defining comparatively narrow ranges of properties (lipophilicity, amphiphilicity, acidity, reactivity, 3D-structural features) of the included compounds. The focusing ensures appropriate absorption, distribution, metabolism, excretion, and toxicity (ADMET) in those test biosystems, which are more complex than isolated receptors, and in humans. The SP deploys conceptual models that include transport and accumulation in a series of membranes, protein binding, hydrolysis, and other reactions with cell constituents. The kinetics of drug disposition is described as a non-linear disposition function of drug structure and properties. The SP capabilities are illustrated here using a model-based quantitative structure-activity relationship of toxicity of phenolic compounds against *Tetrahymena pyriformis* as dependent on lipophilicity and acidity. The resulting SP models clearly outperform empirical models in predictive ability outside the parameter space, as revealed by the leave-extremes-out cross-validation technique with omission of compounds beyond pre-defined lipophilicity and acidity ranges. The SP models do not change substantially if the parameters space is shrunk within some limits. In contrast, the shapes of empirical models vary widely depending upon the fraction of the data set used for their optimization. Once calibrated for a given biosystem, the SP models provide a detailed recipe for tailoring the drug properties to ensure optimum ADMET. The focusing is more accurate than with traditional empirical QSAR studies, assessment of drug-likeness, or the rules for identification of compounds with permeability problems.

Keywords: Lipophilicity; Amphiphilicity; Reactivity; Disposition function; Distribution

1. Introduction

Multiple factors that are optimized in design of compound libraries for high throughput screening deal with either the chemistry of the library itself or with the interactions of the library compounds with biological systems. Chemistry-related factors include coverage of chemical space(s) [1], synthetic feasibility and cost [2], novelty of compounds as compared to available libraries [3], and analytical requirements [4]. Biology-related factors essentially target three processes: binding of compounds to the tested receptor, pharmacokinetics of compounds in humans (absorption, distribution, metabolism, and excretion, and toxicity the latter two processes abbreviated as ADMET).

Optimization of structure-specific drug-receptor binding utilizes various representations of 3D-structure of compounds ranging from molecular structures that are used in free energy calculations [5], docking to known receptors [6–9], and pharmacophore identification [10], through interatomic distance matrices [11], steric/electrostatic fields [12], molecular electrostatic potential grids [13], topomer shapes [14], four-point pharmacophores [15], and pharmacophore fingerprints [16], to more abstract molecular hashkeys representing molecular surface properties [17], topological descriptors [18], and presence of molecular fragments [19]. ADMET processes are usually less structure-specific and depend on molecular properties like partition coefficients in a reference system, polar surface area [20], ionization [21], and reactivity. For mixtures, a simplified characterization was attempted using the descriptor average of individual component compounds [22]. The optimization methods used in library design include binning schemes [23], clustering [14,18], pattern recognition, [20] machine learning [17], genetic algorithms [4,24], and neural networks [19].

The ADMET properties represent a particularly attractive target for library design because the underlying processes mostly do not depend on detailed molecular structure, but rather on physicochemical properties of compounds. Therefore, the ADMET models, once calibrated for the given biosystem, can be applied to different libraries. Development of the SP models for human ADMET is more difficult than that for test biosystems because of complexity and variability of ADMET processes in the population. The ADMET properties of compounds in libraries can be improved using extrapolations of animal data for selected compounds

* Corresponding author. Tel.: +1-701-231-7749; fax: +1-701-231-7606.
E-mail address: stefan_balaz@ndsu.nodak.edu (S. Balaz).

1093-3263/02/$ – see front matter
PII: S1093-3263(01)00149-8

via allometric relations, [25] physiology-based simulations [26], restriction of ranges of molecular properties in the libraries [27], assessment of drug-likeness [28], predictions of cytochrome P450-mediated metabolism [29,30], and quantitative structure-activity relationships (QSAR).

The QSAR approaches can be classified as empirical and semi-empirical (model-based). The difference is in the optimization technique that is used to correlate ADMET properties with structure/property descriptors. While empirical approaches frequently use many descriptors and model-free methods (multiple linear regression on polynomials with cross-terms, partial least squares, neural nets), semi-empirical methods rely on the use of accumulated knowledge for selection of a few descriptors and derivation of model-based non-linear descriptions. The descriptions are usually obtained within the frame of subcellular pharmacokinetics (SP).

In this communication, we briefly review the current status of SP, demonstrate the use of SP models for prediction of toxicity, and compare the derived SP models with empirical models. Predictive ability of the models is also examined by the leave-extremes-out (LEO) cross-validation technique where only the data points outside a pre-defined parameter space are omitted. Potential applications of the presented SP model in library focusing are described.

2. Subcellular pharmacokinetics

The SP aims at understanding of drug disposition in membranes, and intracellular and extracellular aqueous phases of biosystems in terms of the drug structure and properties. Drug disposition is interplay of various processes including transport through (and accumulation in) membranes, covalent and non-covalent binding to proteins and other cell constituents, spontaneous and enzymatically catalyzed metabolic reactions, and excretion (in higher organisms).

2.1. Drug transport

Drugs, as man-made chemicals, usually cross cell membranes by passive diffusion through the phospholipid bilayer. Drugs that resemble physiologic molecules are, in some membranes (usually in specialized cells like epithelia of absorption or excretory organs and tumors), also transported by protein carriers either actively or passively. Relative contributions of passive diffusion and protein-mediated transport to the overall transport vary between drugs and also between membranes.

In the body, diffusion can either bypass the cells or proceed through the interior of cells. Although the paracellular route may prevail in partial steps of distribution in some structures of higher organisms (e.g. distribution of drugs from blood to the extracellular space of liver through sinusoids or to renal filtrate via glomerular filtration), in the vast majority of tissues, the transcellular route of crossing biological membranes is used for drug distribution. In addition to crossing the cytoplasmic membrane, intracellular distribution proceeds by diffusion through the membranes of many organelles including the endoplasmic reticulum.

Water and possibly other small hydrophilic molecules can cross some membranes through the hydrophilic protein channels. However, most drugs use the passive diffusion through phospholipid bilayer of the membrane as the means of transport through the majority of membranes.

In summary, passive trans-bilayer diffusion is one of the most important processes governing the distribution of drugs in biological systems. Important progress has been made in describing phospholipid bilayers at atomic and molecular resolution by application of molecular dynamics simulations and Monte Carlo methods [31,32]. It is now possible to simulate not only the permeation of small particles like protons [33], inorganic ions [34], oxygen and water [35], but also of molecules like benzene, nifedipine, adamantine [36], cholesterol [37,38], and dihydropyridines [39]. These methods are expected to contribute significantly to our understanding of membrane transport. However, they are currently limited to a patch of one phospholipid bilayer and cannot be used in simulation of the distribution of drugs in multi-membrane biosystems. We have to resort to phenomenological models.

A survey of published data on trans-bilayer transport of chemicals [40] indicates that there are two drug properties which determine the overall transport rate of a molecular species: lipophilicity—the tendency of the molecule to accumulate in the bilayer core and amphiphilicity—the tendency of the molecule to adsorb to the interface. A tentative model that includes transport mechanisms for various ranges of lipophilicity and amphiphilicity is shown in Fig. 1. Transport is slowed down by extreme values of lipophilicity and is significantly hindered by high amphiphilicity [40].

2.2. Models of subcellular pharmacokinetics

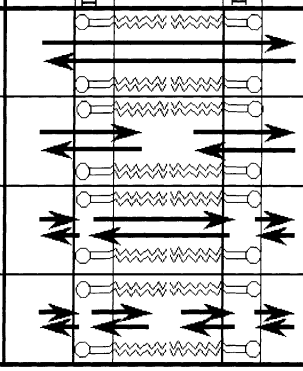
A biosystem is represented as a series of alternating aqueous phases and membranes. The transport mechanisms (Fig. 1) are a crucial part of the model. A portion of a SP model valid for lipophilic and amphiphilic compounds (row 4 in Fig. 1) is depicted in Fig. 2.

2.3. The disposition function

The drug concentration in the receptor surroundings [D] is related to the initial concentration c_0 via the disposition function [41] $A(p_i, t)$ with physicochemical properties p_i and the exposure time t as variables:

$$[D] = A(p_i, t) \times c_0 \quad (1)$$

SP provides the forms of the disposition function for biosystems ranging from subcellular particles through cells and tissues to organisms. The disposition function represents either a numerical or an explicit solution to the system of

Lipophilicity	Amphiphilicity	Number of Compartments	Number of Rate Constants*	Overall Transport Duration	Compartment Model
low	low	2	1	minutes to hours	
medium/ high	low	3	2	seconds/ hours	
low	high	4	3	seconds to days	
high	high	5	4	seconds to days	

*valid for symmetric membrane

Fig. 1. Mechanisms of passive transport as dependent on drug properties. The data on overall transport duration are summaries of experimental data [40]. Each arrow symbolizes a transfer between two molecular states with the rate characterized by a transfer rate constant. The sketches of phospholipid molecules indicate the relation between intra-bilayer compartments and phospholipid structure.

differential equations describing the fate of drug molecules in the given biosystem (Fig. 2). In the equations, the rate/equilibrium parameters are substituted by the extrathermodynamic relationships to other physicochemical parameters of drug molecules that are easier to measure (e.g. lipophilicity or acidity). The biosystem is usually kept as invariant as possible during the experiments so that the parameters characterizing biological properties can be collected in adjustable coefficients.

2.3.1. Numerical simulations

In most cases, the differential equations describing the drug fates in the body (Fig. 2) cannot be solved explicitly and a numerical simulation is the only means of analyzing

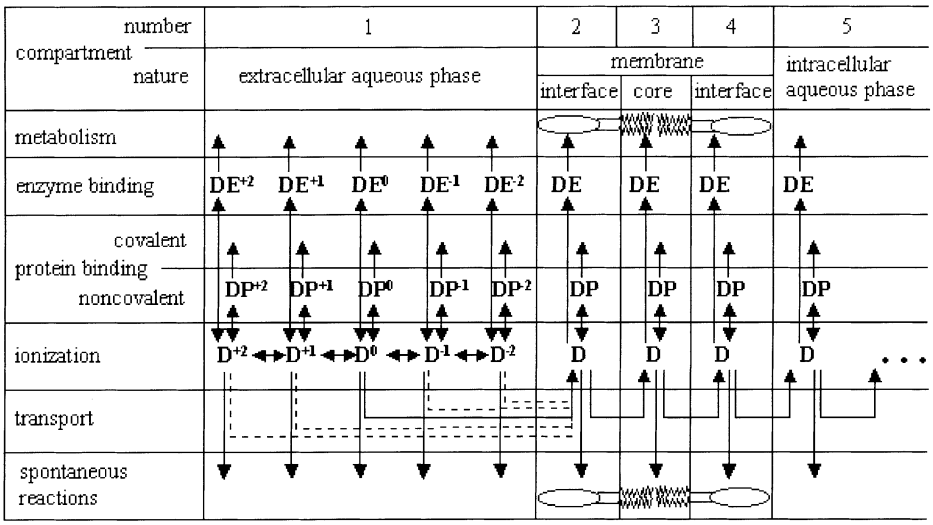


Fig. 2. A typical part of subcellular pharmacokinetics model consisting of alternating aqueous phases and membranes (extracellular phase, the first membrane and the first intracellular phase shown). The drug molecules (D) can ionize to arbitrary degree in any compartment (for simplicity, ionization is shown only in the extracellular phase). All molecular species can bind to proteins (P) and enzymes (E), be transported and metabolized either enzymatically or by spontaneous reactions. The sketch of phospholipid molecules indicates the correspondence between intra-membrane compartments and the phospholipid structure. The number of intra-membrane compartments can be reduced if the drugs do not bind to the interface or accumulate in the core (rows 1–3 in Fig. 1).

the drug behavior. The disposition function is then obtained by fitting an arbitrary function of physicochemical properties and time to the simulated time courses of drug concentrations in individual compartments of an SP model (Fig. 2). Specific non-linear functions [42] perform much better than polynomials and similar expressions. A detailed review of the results in the area of numerical simulations of drug transport is available [42]. The milestones are briefly reviewed here.

The roots of SP and the first form of the disposition function can be traced back to the origin of the QSAR field [43]. The logarithm of the disposition function was expressed as a parabolic dependence on lipophilicity that was parameterized by $\log P$, P being the partition coefficient in the reference (usually 1-octanol/water) system. Intuitive considerations about the influence of lipophilicity on drug transport through a series of membranes were later confirmed by simulations of pure transport [44] (row transport in Fig. 2, no adsorption to the interfaces). Treatment of drug transport as a stochastic process [45] led to bilinear form of the disposition function with lipophilicity as the only variable [46]. The observation of the relationship between the rate constants of transport from the aqueous phase into the core of a non-polar phase and backwards versus the partition coefficient [47,48] provided a basis for realistic SP simulations of non-amphiphilic drugs. The simulations led to refinement of the bilinear disposition function by addition of more segments [49]. The slopes of linear segments are related to the compartment number. This fact was used to locate the receptor compartment and its nature using the activity data of a series of compounds [50]. Reactivity was included into the disposition function in a non-linear form [42].

2.3.2. Explicit solutions: Reduced scenarios and pseudo-equilibrium models

The set of differential equations describing the drug fate in a biosystem is simple enough to be solved explicitly only if a simple biosystem is modeled or if the description is simplified using the time hierarchy of the fate-determining processes (Fig. 2).

Analyzed reduced scenarios include reversible transport of non-amphiphilic compounds through one membrane [51], two membranes [52], and uni-directional transport in a series of membranes as combined with protein binding and elimination [53].

Simplified pseudo-equilibrium models take advantage of the fact that elimination processes are usually slower than transport. The first model-based descriptions were produced for pseudo-equilibrium distribution of ionizable compounds [54]. The time of exposure and chemical reactivity were introduced into model-based QSAR of alkylating agents [55]. The pseudo-equilibrium models have been extended to describe the kinetics of drug disposition during the elimination period [56] and applied to QSAR of compounds with multiple ionization centers [57]. These models describe the time course of the concentration of non-ionized molecules in the aqueous phases of the studied biosystem during the elimination period using Eq. (1) and the disposition function in the form [56]

$$\mathrm{A}(p_i, t) = \frac{1}{A \times P^{\beta} + B} \times \mathrm{e}^{-(C \times P^{\beta} + D)t/(A \times P^{\beta} + B)t} \qquad (2)$$

where P is the reference partition coefficient, β an empirical coefficient, and t the time. The terms A, B, C, D characterize individual processes the drug molecules undergo in the biological system: A, accumulation in the membranes and protein binding, B, distribution in aqueous phases, C, hydrophobicity-dependent elimination, and D, hydrophobicity-independent elimination. For compounds which ionize to the Mth degree, the model suggests that the coefficients A, B, C, D (Y) from Eq. (2) be expanded as [57]

$$Y = Y_0 + \sum_{j=1}^{M} Y_j \times 10^{\mathrm{sgn} \times \sum_{k=1}^{j} \mathrm{p}K_{\mathrm{a}k}} \qquad (3)$$

The subscripts 0 and j indicate the quantities associated with non-ionized molecules and with the species ionized to the jth degree, respectively. The term sgn $= 1$ for bases and sgn $= -1$ for acids (like the sign of the charge of the resulting ion). Sometimes in vitro biological testing is performed under the conditions of varying acidity of the external medium, $\mathrm{pH_e}$. In this case, the terms Y_j associated with the molecular species ionized to the jth degree in Eq. (3) can be further deconvoluted [57]. To keep the number of optimized coefficients at an unavoidable minimum, individual coefficients are tested for the contribution to the fit quality and, depending on the context, fixed to either 0 or 1, if the optimization of the coefficient is not supported by experimental data.

Eq. (2) describes the concentration of the free non-ionized molecules in the aqueous phases. Concentrations of the other molecular species (ionized to arbitrary degree, bound to proteins, accumulated in membranes) can be calculated for any moment using multiplication of the free aqueous non-ionized concentration by a factor containing a drug property ($\mathrm{p}K_{\mathrm{a}}$ values, lipophilicity).

3. Results and discussion

Effectivity of a drug development program is tightly associated with understanding of the ADMET properties of the tested compounds that can be achieved using SP models. The SP models for human ADMET eliminate the compounds with adverse ADMET properties early, ideally before the synthesis on purchase. Since ADMET processes mostly depend on physicochemical properties, rather than on exact 3D-structure, chemical libraries can be optimized for ADMET properties by focusing physicochemical properties. In a high throughput setting, the most frequently used methods include the restriction of ranges of physicochemical and molecular properties in the libraries [27], assessment of drug-likeness [28], and empirical QSAR. Model-based QSAR utilizing SP is another alternative.

3.1. Model-based QSAR for toxicity

Focusing of physicochemical properties using SP is illustrated using a published set of toxicities against *Tetrahymena pyriformis* of 129 phenolic compounds (Table 1). A subset of the data has been analyzed before [58]. However, availability of new data and improvement in the prediction methods for the pK_a values allow for extension of the previously published model. For a biological effect that is an immediate consequence of a 1:1, fast, and reversible drug-receptor interaction, the model based QSAR equation is simply [41]

$$1/c_X = KXA(p_i, t) \quad (4)$$

Here, c_X is the isoeffective concentration, K the drug-receptor association constant and SA(p_i, t) the disposition function as defined in Eq. (1).

Table 1
Structure (substituents on the benzene ring of phenol), CAS registry numbers, 1-octanol/water partition coefficients P, the pK_a values, experimental and calculated (Eq. (5)) toxicities against *Tetrahymena pyriformis* (T is 1/IGC50 and IGC50 is the concentration in mmol/l causing the 50% growth reduction after 96 h exposure) of phenolic compounds. The terms in brackets were used in the situation when IGC_{50} was not reached below solubility limit and log(1/S) is given instead (S is the solubility in mmol/l). Unless specifically stated otherwise the data are taken from shown references

No.	Substituents	CAS No.	log P	pK_a	Log T (log(1/S))		Reference
					Exp.	Calc.	
1	–	108-95-2	1.50[a]	9.98[b]	−0.321	−0.275	[60]
2	2-F	367-12-4	1.76[a]	8.73	0.185	0.115	[60]
3	2-Cl	95-57-8	2.15[a]	8.53[b]	0.277	0.419	[61]
4	2-Br	95-56-7	2.33[a]	8.43[b]	0.504	0.563	[61]
5	2-NO_2	88-75-5	1.77[a]	7.24[b]	0.670	0.384	[61]
6	2-CH_3	95-48-7	1.98[a]	10.26	−0.274	−0.004	[61]
7	2-CH_2CH_3	90-00-6	2.47[a]	10.27[b]	0.176	0.339	[61]
8	2-$CH_2CH{=}CH_2$	1745-81-9	2.64	10.29[b]	0.346	0.457	[61]
9	2-$CH(CH_3)_2$	88-69-7	2.88[a]	10.49[b]	0.803	0.591	[61]
10[f]	2-$C(CH_3)_3$	88-18-6	3.31[a]	11.34[b]	1.239	0.735	[61]
11	2-C_6H_5	90-43-7	3.09[a]	10.00[c]	1.094	0.823	[61]
12	2-CN	611-20-1	1.61[a]	7.17[b]	0.031	0.290	[61]
13	2-CHO	90-02-8	1.81[a]	8.37[b]	0.483	0.212	[61]
14	2-$C(O)CH_3$	118-93-4	1.92[a]	10.17[b]	0.078	−0.029	[61]
15	2-$C(O)NH_2$	65-45-2	1.28[a]	8.37[b]	−0.242	−0.075	[61]
16[f]	2-CH_2OH	90-01-7	0.73[a]	9.93[b]	−0.953	−0.129	[61]
17	2-CH=NOH	94-67-7	1.10	9.08[b]	−0.252	−0.362	[61]
18[f]	2-$NHC(O)CH_3$	614-80-2	0.72[d]	9.86[c]	(−0.765)[d]	−0.708	[61]
19	3-F	372-20-3	1.93[a]	9.29	0.473	0.132	[61]
20	3-Cl	108-43-0	2.50[a]	9.13[b]	0.871	0.561	[60]
21	3-I	626-02-8	2.93[a]	9.05[b]	1.118	0.879	[61]
22	3-NO_2	554-84-7	2.00[a]	8.38[b]	0.506	0.340	[61]
23	3-CH_3	108-39-4	1.98[a]	10.10[b]	−0.062	0.024	[61]
24	3-CH_2CH_3	620-17-7	2.50[a]	10.07	0.229	0.396	[61]
25	3-$CH(CH_3)_2$	618-45-1	3.05	10.03[c]	0.609	0.790	[61]
26	3-$C(CH_3)_3$	585-34-2	3.30[a]	10.08[b]	0.730	0.950	[61]
27	3-C_6H_5	580-51-8	3.23[a]	9.74[b]	1.351	0.963	[61]
28	3-$(CH_2)_{14}CH_3$	501-24-6	9.53[a]	10.36[c]	(5.472)[e]	1.678	[61]
29	3-CN	873-62-1	1.70[a]	8.59[b]	−0.065	0.100	[61]
30	3-CHO	100-83-4	1.38[a]	9.01[b]	0.085	−0.180	[61]
31	3-$C(O)CH_3$	121-71-1	1.39[a]	9.25[b]	−0.381	−0.216	[61]
32	3-OCH_3	150-19-6	1.58[a]	9.65	−0.145	−0.165	[61]
33[f]	3-CH_2OH	620-24-6	0.50[a]	9.83	−1.043	−0.813	[61]
34	3-$C(O)OCH_3$	19438-10-9	1.89[a]	9.15[b]	−0.046	0.129	[61]
35	3-$C(O)OCH_2CH_3$	7781-98-8	2.47[a]	9.10[b]	0.478	0.544	[61]
36	4-F	371-41-5	1.77[a]	9.91[b]	0.017	−0.085	[61]
37	4-Cl	106-48-9	2.39[a]	9.43	0.545	0.430	[61]
38	4-Br	106-41-2	2.59[a]	9.34	0.681	0.588	[61]
39	4-I	540-38-5	2.90[a]	9.32[b]	0.854	0.811	[61]
40	4-NO	104-91-6	1.29[a]	6.33[b]	0.654	0.237	[61]
41[g]	4-NO_2	100-02-7	1.91	7.15	1.926	0.495	[61]
42	4-CH_3	106-44-5	1.97[a]	10.26	−0.192	−0.011	[62]
43	4-CH_2CH_3	123-07-9	2.50[a]	10.21[b]	0.206	0.371	[61]
44	4-$(CH_2)_2CH_3$	645-56-7	3.20[a]	9.99[b]	0.635	0.899	[61]
45	4-$CH(CH_3)_2$	99-89-8	3.05	10.19[c]	0.473	0.762	[61]
46	4-$CH(CH_3)CH_2CH_3$	99-71-8	3.58	10.20[c]	0.980	1.109	[61]
47	4-$C(CH_3)_3$	98-54-4	3.31[a]	10.23	0.913	0.930	[61]

Table 1 (*Continued*)

No.	Substituents	CAS No.	log *P*	pK_a	Log *T* (log(1/*S*))		Reference
					Exp.	Calc.	
48	4-$C(CH_3)_2CH_2CH_3$	80-46-6	3.98	10.14[c]	1.233	1.355	[61]
49	4-C_5H_9	1518-83-8	3.69	10.24[c]	1.292	1.170	[61]
50	4-C_6H_5	92-69-3	3.20[a]	9.82[c]	1.383	0.929	[61]
51	4-CH_2-C_6H_5	101-53-1	3.69	10.05[b]	1.195	1.203	[61]
52[f]	4-$C(CH_3)_2CH_2C(CH_3)_3$	140-66-9	5.31	10.15[c]	2.092	1.898	[61]
53	4-$C(C_6H_5)_3$	978-86-9	7.36[a]	9.49[c]	(3.673)[e]	2.167	[61]
54	4-CF_3	402-45-9	2.88	8.68[b]	0.618	0.909	[61]
55	4-CN	767-00-0	1.60[a]	7.97[b]	0.516	0.143	[61]
56	4-CHO	123-08-0	1.44	7.62	0.266	0.101	[61]
57[f]	4-CH_2OH	623-05-2	0.25	9.82	(−1.732)	−0.927	[61]
58	4-$C(O)CH_3$	99-93-4	1.35[a]	8.05	−0.093	−0.030	[61]
59	4-$C(O)CH_2CH_3$	70-70-2	2.03[a]	8.21[b]	0.056	0.391	[61]
60	4-$C(O)C_6H_5$	1137-42-4	3.07[a]	7.95[b]	1.024	1.169	[61]
61[f]	4-CH_2CN	14191-95-8	0.90	9.63[c]	−0.384	−0.572	[61]
62	4-OCH_3	150-76-5	1.34[a]	10.21[b]	−0.143	−0.416	[61]
63	4-OCH_2CH_3	622-62-8	1.81[a]	10.13[b]	0.013	−0.097	[61]
64	4-$O(CH_2)_3CH_3$	122-94-1	2.90[a]	10.35[c]	0.702	0.629	[61]
65	4-$OCH_2C_6H_5$	103-16-2	3.14	10.29[c]	1.038	0.806	[61]
66	4-$O(CH_2)_6CH_3$	18979-55-0	4.22	10.36[c]	1.648	1.444	[61]
67	4-$O(CH_2)_7CH_3$	13037-86-0	4.75	10.35[c]	2.033	1.681	[61]
68[f]	4-$NHC(O)CH_3$	103-90-2	0.46[a]	9.86[c]	−0.819	−0.838	[61]
69[f]	4-$C(O)NH_2$	619-57-8	0.33[a]	8.56[b]	−0.780	−0.670	[61]
70[f]	4-$(CH_2)_2OH$	501-94-0	0.67	10.17[b]	−0.828	−0.789	[61]
71	4-$C(O)OCH_3$	99-76-3	1.96[a]	8.47[b]	0.084	0.297	[63]
72	4-$C(O)OCH_2CH_3$	120-47-8	2.47[a]	8.50[b]	0.572	0.650	[61]
73	4-N=N-C6H5	1689-82-3	3.18[a]	8.36[b]	1.655	1.172	[63]
74	4-OC_6H_5	831-82-3	3.75	10.07[c]	1.355	1.235	[64]
75	2,3-diCH_3	526-75-0	2.77	10.54[b]	0.122	0.505	[65]
76	2,3-diCl	576-24-9	3.04[a]	7.70[b]	1.271	1.192	[65]
77	2,4-diCH_3	105-67-9	2.30	10.59[b]	−0.029	0.162	[62]
78	2-Br, 4-CH_3	6627-55-0	2.91	8.73[c]	0.789	0.921	[65]
79	2,4-diCl	120-83-2	3.17[a]	7.89[b]	1.036	1.248	[62]
80	2-CH_3, 4-Cl	1570-64-5	2.78[a]	9.71[b]	0.700	0.658	[65]
81	2,4-diBr	615-58-7	3.25[a]	7.79[b]	1.403	1.319	[65]
82[f]	2-$C(CH_3)_3$, 4-CH_3	2409-55-4	4.10	11.64[c]	1.297	1.156	[65]
83[f]	2,4-diNO_2	51-28-5	1.54[a]	4.08[b]	1.096	0.978	[63]
84	2-NO_2, 5-F	446-36-6	1.91	6.18[c]	1.123	0.667	[60]
85	2,5-diCH_3	95-87-4	2.34[a]	10.40[b]	0.009	0.224	[62]
86	2-Cl, 5-CH_3	615-74-7	2.85	8.60[c]	0.640	0.902	[65]
87	2,5-diCl	583-78-8	3.14[a]	7.53[c]	1.128	1.291	[65]
88	2,5-diNO_2	329-71-5	1.86	5.27[b]	0.929	0.803	[66]
89	2,6-diF	28177-48-2	1.65	7.45[c]	0.396	0.267	[65]
90	2,6-diNO_2	573-56-8	1.38[a]	3.71	0.573	1.101	[66]
91	2,6-diC_6H_5	12432-11-3	5.25	10.02[c]	2.113	1.905	[65]
92	3,4-diCH_3	95-65-8	2.23[a]	10.36[b]	0.122	0.153	[65]
93	3-Cl, 4-F	2613-23-2	2.78[a]	9.01[c]	0.842	0.781	[65]
94	3-CH_3, 4-Cl	59-50-7	3.10[a]	9.55	0.795	0.909	[63]
95	3,4-diCl	95-77-2	3.33[a]	8.63	1.745	1.225	[67]
96	3,5-diCH_3	108-68-9	2.35	10.19[b]	0.113	0.268	[65]
97	3,5-diCl	591-35-5	3.61[a]	8.18	1.562	1.483	[65]
98	2,3,5-triCH_3	697-82-5	3.33[a]	10.67[b]	0.360	0.866	[65]
99	2,3,5-triCl	933-78-8	3.84[a]	6.57[c]	2.373	1.903	[67]
100	2,3,6-triCH_3	2416-94-6	2.67[a]	10.77[c]	0.418	0.394	[62]
101[f]	2,4,6-tri$C(CH_3)_3$	732-26-3	6.06[d]	12.20[d]	(2.710)[e]	1.676	[61]
102	2,4,6-triC_6H_5	3140-01-0	7.36[a]	9.94[c]	(3.783)[e]	2.088	[61]
103	2,4,5-triCl	95-95-4	3.72[a]	7.10[c]	2.097	1.740	[67]
104	2-$CH(CH_3)_2$, 4-Cl, 5-CH_3	89-68-9	3.92[a]	10.16[c]	1.862	1.319	[65]
105	2,4,6-triCl	88-06-2	3.69	6.23[b]	1.695	1.873	[65]
106	2,4,6-triBr	118-79-6	4.08	6.10[b]	2.050	2.118	[62]
107	3-CH_3, 4-Cl, 6-NO_2	7147-89-9	2.93	6.51[c]	1.635	1.325	[60]
108	2,6-diCH_3, 4-Br	2374-05-2	3.93	9.81[b]	1.278	1.386	[65]

Table 1 (*Continued*)

No.	Substituents	CAS No.	log *P*	pK_a	Log *T* (log(1/*S*))		Reference
					Exp.	Calc.	
109	2-CH_3, 4-Br, 6-Cl	7530-27-0	3.87[a]	8.28[c]	1.277	1.621	[65]
110	2,6-diCl, 4-Br	697-86-9	3.84	6.21[b]	1.779	1.966	[65]
111[f]	2,6-di$C(CH_3)_3$, 4-CH_3	128-37-0	4.17[a]	12.76[c]	1.788	0.996	[65]
112[f]	2,4-diCH_3, 6-$C(CH_3)_3$	1879-09-0	4.75	12.00[c]	1.245	1.391	[65]
113[f]	2,6-diNO_2, 4-CH_3	609-93-8	2.47	4.23[b]	1.230	1.402	[66]
114[f]	2,6-diNO_2, 4-$C(CH_3)_3$	4097-49-8	3.92	4.07[c]	1.802	2.211	[60]
115[f]	2,6-diBr, 4-NO_2	99-28-5	3.06[a]	3.38[b]	1.357	1.615	[66]
116[f]	2,4-diBr, 6-C_6H_5	55815-20-8	5.33	7.90[c]	2.208	2.299	[60]
117[f]	2,4-diCl, 6-NO_2	609-89-2	3.21	4.64[c]	1.750	1.812	[66]
118[f]	2,6-diI, 4-NO_2	305-85-1	3.76	3.32[b]	1.812	1.813	[66]
119[f]	2,4,6-triNO_2	88-89-1	1.82	0.37[b]	−0.155	−0.610	[66]
120[f]	2-CH_3, 4,6-diNO_2	534-52-1	2.56	4.34[b]	1.729	1.438	[63]
121	3,4,5-triCH_3	527-54-8	3.42	10.37[c]	0.930	0.978	[60]
122	3,5-diCH_3, 4-Cl	88-04-0	3.78	9.70[b]	1.203	1.319	[65]
123	2,3,4,5-tetraCl	4901-51-3	4.21[a]	6.15[b]	2.699	2.177	[67]
124	2,3,5,6-tetraF	769-39-1	2.07	5.46[c]	1.167	0.907	[66]
125	2,3,5,6-tetraCl	935-95-5	3.88[a]	5.09[c]	2.222	2.168	[67]
126[f]	2-CH_3, 3,4,5,6-TetraBr	576-55-6	5.70	6.31[c]	2.573	2.658	[66]
127	pentaF	771-61-9	3.23[a]	5.50[b]	1.631	1.703	[66]
128[f]	pentaCl	87-86-5	5.18[a]	4.82[b]	2.523	2.765	[67]
129	pentaBr	608-71-9	5.69	4.43[c]	2.664	2.917	[66]

[a] The partition coefficient was calculated by the CLOGP program [69].
[b] Experimental ionization constant (temperature 25 °C, ionic strength 0.0; if several values were available, the latest value was used) [70].
[c] The ionization constant was estimated [70].
[d] Experimental log *P*, pK_a or solubility [71].
[e] Predicted solubility [72].
[f] Omitted in the LEO cross-validation.
[g] Excluded from correlation due to a different mechanism [68].

Combination of Eqs. (2)–(4) results in the following description of toxicity T ($\sim 1/c_X$) on lipophilicity ($\log P$) and acidity (pK_a):

$$\log T = -\log(A_0 \times P^{\beta} + B_1 \times K_a + 1) - \frac{D_0 + D_1 \times K_a}{A_0 \times P^{\beta} + B_1 \times K_a + 1} - E \times pK_a + F \quad (5)$$

where toxicity T is expressed as the reciprocal concentration (mmol/l) of the tested phenolic compounds causing 50% reduction in the growth of the protozoa after 96 h exposure. The first two terms in Eq. (5) come from the disposition function as given by Eqs. (2) and (3); the third term represents non-specific binding to the receptors (K in Eq. (4)). The non-linear regression analysis started with Eq. (5) having the first two terms complete as in Eqs. (2) and (3). Some processes did not contribute sufficiently to the observed effects, as manifested by high standard deviations of related coefficients and no improvement in the quality of the fit. The related coefficients were fixed as follows: $B_0 = 1$ and $A_1 = C_0 = C_1 = 0$. The values of individual adjustable coefficients were optimized as: $A_0 = 0.1830 \pm 0.1033$, $\beta = 0.2665 \pm 0.0631$, $B_1 = 6.407 \pm 3.805$, $D_0 = 5.471 \pm 0.929$, $D_1 = 16018 \pm 8630$, $E = 0.1758 \pm 0.0143$, and $F = 5.393 \pm 0.330$. Statistical indices of the fit for $n = 122$ experimental points (one compound omitted due to an additional mechanism of action, six inactive compounds) were: the correlation coefficient $r = 0.953$, the fit standard error S.E. $= 0.257$, and the value of the Fisher test $F = 191.0$. The agreement between experimental and calculated toxicities (Fig. 3, panel A1) is satisfactory. Previous data set [58] did not allow optimization of the coefficient B_1. The values of the coefficients that are not associated with the K_a values did not change significantly in the new model. This fact indicates another advantageous feature of semi-empirical models as compared to empirical models: the model can be conceptually extended when new data become available.

Toxicity of the compounds is determined by interplay of distribution as described by the first term in Eq. (5), metabolism (the second term), and receptor binding (the third term). The model described by Eq. (5) can be interpreted as follows. The differences in toxicity are essentially ascribable to distribution of free non-ionized molecules, although a small effect of receptor binding (the decrease in toxicity with increasing pK_a values as described using the term E in Eq. (5)) is apparent. Significantly lower toxicity of two groups of compounds (the first group has $pK_a > 6$ and $\log P < 4$ and the second group has $pK_a < 2$ as seen in Fig. 3, panel A1) is due metabolism (degradation). This process takes place in aqueous phases of the cells and proceeds with the approximately equal rate (micro)constants for all the tested substances. The hydrophilic compounds

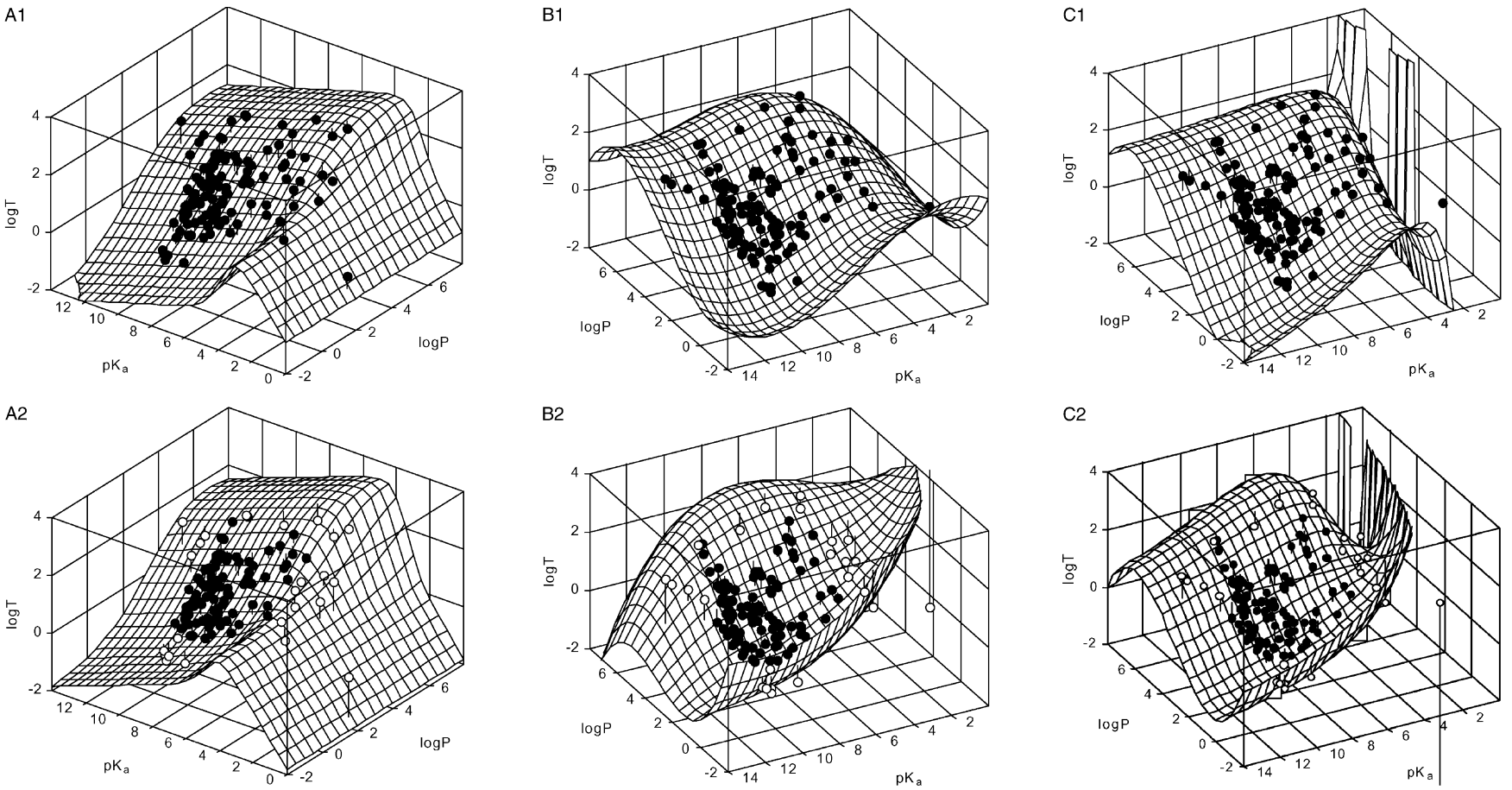


Fig. 3. Surfaces generated by the model-based Eq. (5) (A) and empirical Models 1 (B) and 2 (C) for the complete set of compounds (1) and for the reduced set of compounds (2). The omitted points in the reduced set are indicated by empty circles. The polynomial Models 1 and 2 are specified in the text.

are metabolized faster because they are present in the aqueous phases in higher concentrations than more hydrophobic derivatives, which are prevented from metabolism by accumulation in the membranes and by non-specific binding to proteins. Both non-ionized and ionized species are metabolized as expressed by the coefficients D_0 and D_1 in Eq. (5), comprising the rate (micro)constants of metabolism for non-ionized and ionized species, respectively. Faster metabolism of ionized species ($D_1 \gg D_0$) causes a rapid decrease in toxicity of very acidic derivatives ($pK_a < 2$) regardless their hydrophobicity (Fig. 3, panel A1). Fixing the coefficient A_1 to zero indicates that the processes of the ionized species that are represented by the coefficient (membrane accumulation and protein binding) do not contribute significantly to the total distribution of the compounds in the protozoan suspension. Eq. (5) contains the term $E \times pK_a$ that reflects the association of the non-ionized phenol molecules with the receptors. The sign of the optimized value of the coefficient E is negative, i.e. the strength of the association grows with the increase in the electron withdrawal from the phenol group. According to the model, the receptors seem to be localized in the aqueous phases, not in the membranes.

The SP model (Eq. (5)) describes toxicity as a process that is based mainly on disposition and, to a much lesser extent, on receptor binding. Therefore, a part of Eq. (5) (the first term describing distribution and the second term describing metabolism) can also be used to assess disposition of compounds in the studied biosystem. If one is interested in compounds with low toxicity or high metabolism at the tested exposure time, Eq. (5) suggests that the library be focused on compounds with $pK_a > 6$ and $\log P < 4$ or with $pK_a < 2$ (Fig. 3, panel A1).

The SP models for those test biosystems, which are more complex than isolated receptors, may help correctly rank receptor affinities of the tested compounds. For instance, a compound may elicit low effects in test biosystems due to metabolism or toxicity. This compound may be very effective in humans where ADMET may differ and an SP model may help to keep this compound in the pipeline. In this particular case, if any of the quickly metabolized compounds (with $pK_a > 6$ and $\log P < 4$ or with $pK_a < 2$—Fig. 3, panel A1) would show some activity in a related test biosystem, the activity could be adjusted for metabolism using the SP model described by Eq. (5).

The kinetics of disposition was not specifically incorporated into Eq. (5) because the data were only available for the 96 h exposure. However, the kinetics can be inferred from the comparison of Eq. (5) with Eq. (2) that was used in its derivation: D_0 and D_1 are the only coefficients, which contain the time of exposure, t. Both coefficients can be replaced in Eq. (5) by $D_i \times t/96$ ($i = 0$ or 1) to explicitly express the kinetics of disposition. Using modified Eq. (5), we can perform a very flexible library focusing by specifying the borderline values of $\log P$ and pK_a for any chosen exposure period.

3.2. Comparison of subcellular pharmacokinetics with empirical approaches

Model-free QSAR methods (e.g. neural networks and linear methods sometimes using polynomials with or without cross-terms, with the regression coefficients optimized by linear regression analysis, partial least squares or genetic algorithms) can also be used for library focusing, therefore we tested how they would perform in a similar situation. Polynomials with cross-terms were shown to perform equally well or better than several neural network architectures with a higher number of adjustable weights [59]. Hence, we decided to use polynomials with cross-terms of $\log P$ and pK_a, and their inverse values (up to 10 optimized coefficients) as representative empirical models. The software TableCurve 3D [73] generated 113 empirical models that had better statistics (r, S.E., F) than the model-based Eq. (5). The best two models, Models 1 and 2, are presented graphically in Fig. 3, panels B1 and C1, respectively. For Model 1, the independent variables were ($x = \log P$, $y = pK_a$) $x, y, x^2, y^2, x^3, y^3, xy, x^2y, xy^2$ and the statistical indices $n = 122$ (one compound omitted due to an additional mechanism of action, six compound inactive), $r = 0.963$, $s = 0.233$ and $F = 157.4$. Model 2 contained the same variables as Model 1 except that $1/y$ was used instead of y. The statistical indices for Model 2 were $n = 122$, $r = 0.962$, $s = 0.235$ and $F = 155.9$. The surfaces generated by the 113 models had widely differing shapes. All of them closely copied the surface generated by the data but adopted various shapes in the areas outside the parameter space. In this case, the parameter space is defined by the ranges of the two used independent variables, $\log P$ and pK_a (i.e. a square area marked by the lowest and highest $\log P$ and the lowest and highest pK_a).

Predictive ability of the models can be tested by prediction of toxicity of compounds that were inactive in concentrations equal to solubilities (Table 1, compounds 18, 28, 53, 57, 101, and 102). All three descriptions, the model-based Eq. (5), and empirical Models 1 and 2, incorrectly predicted the isoeffective concentrations IGC_{50} of two compounds (Table 1, nos. 18 and 57) as being lower than solubility and, thus, measurable. The two compounds have very low toxicities that are difficult to measure precisely. All three models also grossly underestimated toxicity of 4-nitrophenol (Table 1, no. 41) that was excluded from all correlations. This compound is known to exhibit a pro-electrophile mechanism of toxicity, in addition to non-specific toxicity [68].

In absence of additional experimental data, predictive ability of models is usually tested by cross-validation. The data set is divided into groups containing one or more data points. A group is omitted from the set and the correlation is recalculated. The new model is used to calculate the predicted magnitudes of the dependent variable for the omitted points. This procedure is repeated so that each group is omitted once and the predictive sum of squares of deviations (PRESS) for all points is calculated. The points are usually assigned to the groups based on the experimentally

Table 4
Cross-validation of the QSAR models for toxicity of phenolic compounds against *Tetrahymena pyriformis*[a]

Method	# Groups	PRESS		
		Eq. (5)	Empirical models	
			Model 1	Model 2
Leave-one-out	122	9.577	7.981	8.020
Leave-several-out	20	9.458	7.486	7.670
Leave-several-out	7	9.821	7.242	7.404
Leave-extremes-out	1	6.055	30.40	36.83

[a] PRESS is the predictive sum of squares of deviations between the calculated and experimental values of the omitted points. The optimized coefficients of the subcellular pharmacokinetics-based Eq. (5) are given in the text and the corresponding surface is shown in Fig. 3, panels A1/A2. The empirical Models 1 and 2 are described in the text and the corresponding surface is shown in Fig. 3, panels B1/B2 and C1/C2, respectively.

determined dependent variable or randomly. In both cases, the assignment is random with respect to the independent variables. When a group is omitted during cross-validation, the ranges of parameter values do not change substantially and the ($\log P$, pK_a) coordinates of the majority of omitted points remain inside the parameter space. As a result, this type of cross-validation examines the models inside the area defined by the parameter space and effectively eliminates too flexible models. Table 2 summarizes the results of the leave-one-out cross-validation and two types of leave-several-out cross-validations. The picture was consistent for all three cross-validations: the model-based Eq. (5) (Fig. 3, panel A1) had slightly higher PRESS values than the empirical Models 1 and 2 (Fig. 3, B1 and C1, respectively).

The most valuable predictions are those outside the parameter space, at least for structure-non-specific properties where the boundaries of the parameter space are known. The reason is that predictions for new compounds which have the parameter values in the ranges defined by the parameter values of tested compounds can essentially be done using various interpolations. Cross-validation based on random omission of data points does not deal with predictions outside the tested parameter space. To address this issue, we devised a cross-validation technique that we call leave-extremes-out (LEO cross-validation). As the name indicates, the data points with high and low values of both independent variables were omitted. The data set then included 97 compounds having $\log P$ between 1.0 and 5.3 and pK_a between 5.0 and 11.0. These limits were chosen arbitrarily to leave out at least four compounds at each side of the property rectangle. For the reduced data set, the TableCurve 3D software [73] generated 105 empirical models of various shapes that had better statistical indices (r, S.E., F) than the model-based Eq. (5). The calculated PRESS values (Table 2, LEO) for 25 omitted compounds (Table 1, footnote g) show that the LEO cross-validation provides a significant discrimination among the models. The model-based Eq. (5) predicts outside the parameter space much better than the empirical Models 1 and 2. As seen in Fig. 3, omission of compounds with extreme $\log P$ and pK_a values causes the empirical equations to dramatically change the surface shape in the region of omission. In contrast, the model-based Eq. (5) maintains the shape of the associated surface even if optimized for reduced parameter ranges.

Predictions outside the parameter space are important for correct focusing of physicochemical properties of chemical libraries. Conclusions regarding property magnitudes outside the parameter space can be grossly erroneous if based on empirical models (Fig. 3).

4. Conclusions

The SP creates conceptual models of drug disposition in various biological systems as dependent on physicochemical properties. The models can be calibrated for a given biological system using a representative set of compounds and used for a detailed focusing of physicochemical properties of compounds in libraries. Predictive ability of the SP models outside the parameter space is much better than that of empirical models as shown by the leave-extremes-out cross-validation technique. In summary, library focusing using SP models is more precise than those based on ad hoc restriction on properties, drug-likeness, and empirical QSAR models.

Acknowledgements

Although the research described in this article has been funded in parts by US EPA (R82-6652-011) and NIH (1 P20 RR15566-01), it has not been subjected to the agencies' peer and policy reviews and, therefore, does not necessarily reflect the views of the agencies and no official endorsement should be inferred.

References

[1] D.K. Agrafiotis, J.C. Myslik, F.R. Salemme, Advances in diversity profiling and combinatorial series design, Mol. Divers. 4 (1998) 1–22.
[2] E.J. Martin, R.E. Critchlow, Beyond mere diversity: tailoring combinatorial libraries for drug discovery, J. Comb. Chem. 1 (1999) 32–45.
[3] W. Zheng, S.T. Hung, J.T. Saunders, G.L. Seibel, PICCOLO: A tool for combinatorial library design via multicriterion optimization. Pac. Symp. Biocomput. (2000) 588–99.
[4] R.D. Brown, Y.C. Martin, Designing combinatorial library mixtures using a genetic algorithm, J. Med. Chem. 40 (1997) 2304–2313.
[5] M.L. Lamb, J. Tirado-Rives, W.L. Jorgensen, Estimation of the binding affinities of FKBP12 inhibitors using a linear response method, Bioorg. Med. Chem. 7 (1999) 851–860.
[6] M.L. Lamb, K.W. Burdick, S. Toba, M.M. Young, K.G. Skillman, X.Q. Zou, J.R. Arnold, I.D. Kuntz, Design, docking, and evaluation of multiple libraries against multiple targets, Prot. Struct. Funct. Gen. 42 (2001) 296–318.
[7] D.L. Kirkpatrick, S. Watson, S. Ulhaq, Structure-based drug design: combinatorial chemistry and molecular modeling, Comb. Chem. High Throughput Screen. 2 (1999) 211–221.

[8] H.J. Bohm, M. Stahl, Structure-based library design: molecular modeling merges with combinatorial chemistry, Curr. Opin. Chem. Biol. 4 (2000) 283–286.
[9] J.W. Godden, F. Stahura, J. Bajorath, Evaluation of docking strategies for virtual screening of compound databases: cAMP-dependent serine/threonine kinase as an example, J. Mol. Graph. Model. 16 (1998) 139–143, and 165.
[10] X. Chen, A. Rusinko, A. Tropsha, S.S. Young, Automated pharmacophore identification for large chemical data sets, J. Chem. Inf. Comput. Sci. 39 (1999) 887–896.
[11] C.A. Pepperrell, P. Willett, Techniques for the calculation of three-dimensional structural similarity using inter-atomic distances, J. Comput. Aid. Mol. Des. 5 (1991) 455–474.
[12] J. Mestres, D.C. Rohrer, G.M. Maggiora, A molecular-field-based similarity study of non-nucleoside HIV-1 reverse transcriptase inhibitors. 2. The relationship between alignment solutions obtained from conformationally rigid and flexible matching, J. Comput. Aid. Mol. Des. 14 (2000) 39–51.
[13] D.A. Thorner, P. Willett, P.M. Wright, R. Taylor, Similarity searching in files of three-dimensional chemical structures: representation and searching of molecular electrostatic potentials using field-graphs, J. Comput. Aid. Mol. Des. 11 (1997) 163–174.
[14] K.M. Andrews, R.D. Cramer, Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries, J. Med. Chem. 43 (2000) 1723–1740.
[15] J.S. Mason, D.L. Cheney, Library design and virtual screening using multiple four-point pharmacophore fingerprints. Pac. Symp. Biocomput. (2000) 576–87.
[16] M.J. McGregor, S.M. Muskal, Pharmacophore fingerprinting. 2. Application to primary library design, J. Chem. Inf. Comput. Sci. 40 (2000) 117–125.
[17] A.M. Ghuloum, C.R. Sage, A.N. Jain, Molecular hashkeys: a novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules, J. Med. Chem. 42 (1999) 1739–1748.
[18] R.D. Brown, Y.C. Martin, An evaluation of structural descriptors and clustering methods for use in diversity selection, SAR QSAR Environ. Res. 8 (1998) 23–39.
[19] Ajay, G.W. Bemis, M.A. Murcko, Designing libraries with CNS activity, J. Med. Chem. 42 (1999) 4942–4951.
[20] W.J. Egan, K.M. Merz, J.J. Baldwin, Prediction of drug absorption using multivariate statistics, J. Med. Chem. 43 (2000) 3867–3877.
[21] F. Darvas, G. Dorman, A. Papp, Diversity measures for enhancing ADME admissibility of combinatorial libraries, J. Chem. Inf. Comput. Sci. 40 (2000) 314–322.
[22] R.P. Sheridan, The centroid approximation for mixtures: calculating similarity and deriving structure—activity relationships, J. Chem. Inf. Comput. Sci. 40 (2000) 1456–1469.
[23] M.J. Bayley, P. Willett, Binning schemes for partition-based compound selection, J. Mol. Graph. Model. 17 (1999) 10–18.
[24] S.J. Cho, W. Zheng, A. Tropsha, Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches, J. Chem. Inf. Comput. Sci. 38 (1998) 259–268.
[25] I. Mahmood, J.D. Balian, The pharmacokinetic principles behind scaling from preclinical results to phase I protocols, Clin. Pharmacokinet. 36 (1999) 1–11.
[26] G.M. Grass, Simulation models to predict oral drug absorption from in vitro data, Adv. Drug Deliv. Rev. 23 (1997) 199–219.
[27] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25.
[28] Ajay, W.P. Walters, M.A. Murcko, Can we learn to distinguish between drug-like and non-drug-like molecules? J. Med. Chem. 41 (1998) 3314–3324.
[29] K.R. Korzekwa, J. Grogan, S. DeVito, J.P. Jones, Electronic models for cytochrome P450 oxidations, Adv. Exp. Med. Biol. 387 (1996) 361–369.
[30] M.J. De-Groot, M.J. Ackland, V.A. Horne, A.A. Alex, B.C. Jones, Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6, J. Med. Chem. 42 (1999) 1515–1524.
[31] T.E. Thompson, Lipids: social order among lipids in membranes, Curr. Opin. Struct. Biol. 7 (1997) 509–510.
[32] K.M. Merz Jr., Molecular dynamics simulations of lipid bilayers, Curr. Opin. Struct. Biol. 7 (1997) 511–517.
[33] S.J. Marrink, F. Jahnig, H.J. Berendsen, Proton transport across transient single-file water pores in a lipid membrane studied by molecular dynamics simulations, Biophys. J. 71 (1996) 632–647.
[34] M.A. Wilson, A. Pohorille, Mechanism of unassisted ion transport across membrane bilayers, J. Am. Chem. Soc. 118 (1996) 6580–6587.
[35] S.J. Marrink, H.J.C. Berendsen, Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations, J. Phys. Chem. 100 (1996) 16729–16738.
[36] D. Bassolino, H. Alper, T.R. Stouch, Drug-membrane interactions studied by molecular dynamics simulation: size dependence of diffusion, Drug. Des. Discov. 13 (1996) 135–141.
[37] K. Tu, M.L. Klein, D.J. Tobias, Constant-pressure molecular dynamics investigation of cholesterol effects in a dipalmitoylphosphatidylcholine bilayer, Biophys. J. 75 (1998) 2147–2156.
[38] A.M. Smondyrev, M.L. Berkowitz, Structure of dipalmitoyl-phosphatidylcholine/cholesterol bilayer at low and high cholesterol concentrations: molecular dynamics simulation, Biophys. J. 77 (1999) 2075–2089.
[39] M. Aiello, O. Moran, M. Pisciotta, F. Gambale, Interaction between dihydropyridines and phospholipid bilayers: a molecular dynamics simulation, Eur. Biophys. J. 27 (1998) 211–218.
[40] S. Balaz, Lipophilicity in trans-bilayer transport and subcellular pharmacokinetics, Perspect. Drug Discov. Des. 19 (2000) 157–177.
[41] S. Balaz, E. Sturdik, M. Tichy, Hansch approach and kinetics of biological activities, Quant. Struct. Act. Relat. 4 (1985) 77–81.
[42] S. Balaz, Subcellular pharmacokinetics and drug properties: numerical simulations in multicompartment systems, Quant. Struct. Act. Relat. 13 (1994) 381–392.
[43] C. Hansch, T. Fujita, ρ–σ–π analysis. A method for the correlation of biological activity and chemical structure, J. Am. Chem. Soc. 86 (1964) 1616–1626.
[44] J.T. Penniston, L. Beckett, D.L. Bentley, C. Hansch, Passive permeation of organic compounds through biological tissue: a non-steady-state theory, Mol. Pharmacol. 5 (1969) 333–341.
[45] J.W. McFarland, On the parabolic relationship between drug potency and hydrophobicity, J. Med. Chem. 13 (1970) 1192–1196.
[46] H. Kubinyi, Quantitative structure-activity relationships. 7. The bilinear model: A new model for non-linear dependence of biological activity on hydrophobic character, J. Med. Chem. 20 (1977) 625–629.
[47] H. Kubinyi, Drug partitioning: relationships between forward and reverse rate constants and partition coefficient, J. Pharm. Sci. 67 (1978) 262–263.
[48] H. van de Waterbeemd, H. van Bakel, A. Jansen, Transport in quantitative structure-activity relationships VI: relationship between transport rate constants and partition coefficients, J. Pharm. Sci. 70 (1981) 1081–1082.
[49] S. Balaz, E. Sturdik, M. Hrmova, M. Breza, T. Liptaj, Kinetics of drug partitioning in closed four-compartment systems and its application to QSAR, Eur. J. Med. Chem. 19 (1984) 167–171.
[50] S. Balaz, E. Sturdik, I. Dibus, L. Ebringer, L. Stibranyi, M. Rosenberg, Quantitative relationships between lipophilicity and mutagenic effects of *N*-substituted amides of 3-(5-nitro-2-furyl)-acrylic acid on *Salmonella typhimurium*, Chem. Biol. Interact. 55 (1985) 93–108.
[51] S. Balaz, E. Sturdik, Kinetics of passive transport in water/membrane/water system. A mathematical description, Gen. Physiol. Biophys. 4 (1985) 105–108.

[52] R. Dvorsky, S. Balaz, R.J. Sawchuk, Kinetics of subcellular distribution of compounds in simple biosystems and its use in QSAR, J. Theor. Biol. 185 (1997) 213–222.
[53] S. Balaz, E. Sturdik, J. Augustin, Subcellular distribution of compounds in biosystems, Bull. Math. Biol. 50 (1988) 367–378.
[54] Y.C. Martin, J.J. Hackbarth, Theoretical model-based equations for the linear free energy relationships of the biological activity of ionizable substances. 1. Equilibrium-controlled potency, J. Med. Chem. 19 (1976) 1033–1039.
[55] S. Balaz, M. Rosenberg, M. Tichy, E. Sturdik, J. Augustin, Kinetics of biological activities. Inhibition of respiration in *Escherichia coli* by 2-furylethylenes, in: D. Hadzi, B. Jerman-Blazic (Eds.), QSAR in Drug Design and Toxicology, Elsevier, Amsterdam, 1987, pp. 325–330.
[56] S. Balaz, M. Wiese, J.K. Seydel, A time hierarchy-based model for kinetics of drug disposition and its use in quantitative structure-activity relationships, J. Pharm. Sci. 81 (1992) 849–857.
[57] S. Balaz, K. Pirselova, T.W. Schultz, J. Hermens, Kinetics of subcellular distribution of multiply ionizable compounds: A mathematical description and its use in QSAR, J. Theor. Biol. 178 (1996) 7–16.
[58] K. Pirselova, S. Balaz, T.W. Schultz, Model-based QSAR for ionizable compounds: toxicity of phenols against *Tetrahymena pyriformis*, Arch. Environ. Contam. Toxicol. 30 (1996) 170–177.
[59] B. Lucic, D. Amic, N. Trinajstic, Non-linear multivariate regression outperforms several concisely designed neural networks on three QSPR data sets, J. Chem. Inf. Comp. Sci. 40 (2000) 403–413.
[60] T.W. Schultz, A.P. Bearden, J.S. Jaworska, A novel QSAR approach for estimating toxicity of phenols, SAR QSAR Environ. Res. 5 (1996) 99–112.
[61] T.W. Schultz, D.T. Lin, S.K. Wesley, QSARs for monosubstituted phenols and the polar narcosis mechanism of toxicity, Quality Assur. Good Pract. Regul. Law 1 (1992) 132–143.
[62] T.W. Schultz, G.W. Riggin, Predictive correlations for the toxicity of alkyl- and halogen-substituted phenols, Toxicol. Lett. 25 (1985) 47–54.
[63] T.W. Schultz, G.W. Holcombe, G.L. Phipps, Relationships of quantitative structure-activity to comparative toxicity of selected phenols in the Pimephales promelas and *Tetrahymena pyriformis* test systems, Ecotoxicol. Environ. Safety 12 (1986) 146–153.
[64] J.S. Jaworska, T.W. Schultz, Quantitative relationships of structure-activity and volume fraction for selected non-polar and polar narcotic chemicals, SAR QSAR Environ. Res. 1 (1993) 3–19.
[65] T.W. Schultz, L.M. Arnold, T.S. Wilke, M.P. Moulton, Relationships of quantitative structure-activity for normal aliphatic alcohols, Ecotoxicol. Environ. Safety 19 (1990) 243–253.
[66] M. Cajina-Quezada, T.W. Schultz, Structure-toxicity relationships for selected weak acid respiratory uncouplers, Aquat. Toxicol. 17 (1990) 239–252.
[67] S.E. Bryant, T.W. Schultz, Toxicological assessment of biotransformation products of pentachlorophenol—*Tetrahymena* population growth impairment, Arch. Environ. Contam. Toxicol. 26 (1994) 299–303.
[68] D.W. Roberts, An analysis on published data on fish toxicity of nitrobenzene and aniline derivatives, in: K.L.E. Kaiser (Ed.), QSAR in Environmental Toxicology, Reidel, Dordrecht, 1987, pp. 295–308.
[69] ClogP for Windows, Version 1.0.0, 1996, BioByte, Claremont, CA.
[70] pK_a DB, Version 4.59, 2001, Advanced Chemistry Development Inc., Toronto, Canada.
[71] SRC PhysProp database, 2001, Syracuse Research Corporation, Syracuse, NY (http://esc.syrres.com/interkow/physdemo.htm).
[72] WS-KOW, Version 1.26, 1996, Syracuse Research Corporation, Syracuse, NY.
[73] TableCurve 3D, Version 2.05, 1996, Jandel Scientific, San Rafael, CA.