Journal of Computer-Aided Molecular Design, **16:** 785–793, 2002.
KLUWER/ESCOM

Similarity based SAR (SIBAR) as tool for early ADME profiling

Christian Klein[1], Dominik Kaiser[1], Stephan Kopp[2], Peter Chiba[2] & Gerhard F. Ecker[1,*]
[1]*Institute of Pharmaceutical Chemistry, University of Vienna, Althanstrasse 14, A-1090 Wien, Austria;* [2]*Institute of Medical Chemistry, University of Vienna, Waehringerstrasse 10, A-1090 Wien, Austria*

Received 24 September 2002; Accepted 7 January 2003

Key words: ADME, P-glycoprotein, similarity based descriptors, propafenone

Summary

Estimation of bioavailability and toxicity at the very beginning of the drug development process is one of the big challenges in drug discovery. Most of the processes involved in ADME are driven by rather unspecific interactions between drugs and biological macromolecules. Within the past decade, drug transport pumps such as P-glycoprotein (Pgp) have gained increasing interest in the early ADME profiling process. Due to the high structural diversity of ligands of Pgp, traditional QSAR methods were only successful within analogous series of compounds. We used an approach based on similarity calculations to predict Pgp-inhibitory activity of a series of propafenone analogues. This SIBAR approach is based on selection of a highly diverse reference compound set and calculation of similarity values to these reference compounds. The similarity values (denoted as SIBAR descriptors) are then used for PLS analysis. Our results show, that for a set of 131 propafenone type compounds, models with good predictivity were obtained both in cross validation procedures and with a 31-compound external test set. Thus, these new descriptors might be a versatile tool for generation of predictive ADME models.

Introduction

Approximately 40% of compounds under development as drugs fail due to improper ADME parameters. Thus, to decrease this high failure rate, more informations on ADME parameters are required at an earlier stage of the design cycle. With respect to this demand, several in vitro and in vivo approaches were proposed. These include immobilised artificial membrane chromatography (IAM-HPLC) [1], parallel artificial membrane permeation assays (PAMPA) [2], Caco-2 permeation assays in a 96-well format or simultaneous application of up to 10 compounds in one animal (multi-cassette dosing). Within the last decade, *in silico* methods gained increasing interest due to their cost-effectiveness and the possibility of performing ultra-high throughput calculations. The methods applied include both classification systems for filtering of compound selections [3, 4] and regression models for prediction of absorption and permeation values [5]. Recently, descriptors calculated on basis of molecular interaction fields were also introduced and shown to give excellent results [6]. A completely different approach to predict ADME properties may be to use descriptors based on similarity values. Several QSAR-approaches based on similarity measures were reported in the literature, mostly using N × N similarity matrices [7, 8, 9]. Very interesting seems the approach introduced by Ghuloum et al., who used molecular hashkeys based on molecular surface similarity. However, each hashkey computation requires 30–45 processor minutes [10]. In this paper we propose the use of a similar approach, which is based on descriptors which require much less computational time for calculation.

*To whom correspondence should be addressed.
E-mail: gerhard.f.ecker@univie.ac.at

Materials and methods

Data set

Almost all models proposed for *in silico* prediction of absorption and blood-brain barrier permeation solely

Table 1. Chemical structure and PGP-inhibitory activity (in μM) of propafenone derivatives.

Code	Core	R1 - R4	L	A	EC_{50}
GPV 0001	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NH(*n*Pr)	0.329
GPV 0002	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-$N(C_2H_5)_2$	0.907
GPV 0005	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A1	0.599
GPV 0009	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-N(*i*Pr)$_2$	0.377
GPV 0012	A	R1 = $COCH_2CH_3$	-$CH_2CH(OH)CH_2$−	A1	14.31
GPV 0017	A	R1 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A1	31.96
GPV 0019	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A3	0.610
GPV 0021	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A4	0.230
GPV 0023	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A5	0.678
GPV 0025	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A6	0.256
GPV 0027	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A7	0.027
GPV 0029	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A10	0.659
GPV 0031	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A8	0.070
GPV 0045	A	R1 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A8	2.087
GPV 0046	A	R1 = $COCH_2CH_3$	-$CH_2CH(OH)CH_2$−	A2	207.2
GPV 0048	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-$N(CH_3)_2$	2.479
GPV 0049	A	R1 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A2	67.32
GPV 0050	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-$N(CH_3)$*n*Pr	0.416
GPV 0051	A	R1 = COPh	-$CH_2CH(OH)CH_2$−	-$N(C_2H_5)_2$	2.322
GPV 0057	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A2	3.645
GPV 0062	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A12	0.058
GPV 0073	A	R3 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A1	0.972
GPV 0088	A	R1 = $CH(OH)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A1	0.901
GPV 0090	A	R1 = $CH(OCH_3)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A1	0.173
GPV 0091	B	R1 = CH_2CH_2Ph	-$(CH_2)_2$−	A1	0.980
GPV 0092	B	R1 = CH_2CH_3	-$CH(OH)CH_2$−	-NH*n*Pr	11.52
GPV 0093	B	R1 = CH_2CH_2Ph	(S) -$CH(OH)CH_2$−	-NH*n*Pr	1.059
GPV 0094	B	R1 = CH_2CH_2Ph	(R) -$CH(OH)CH_2$−	-NH*n*Pr	1.103
GPV 0095	B	R1 = CH_2CH_2Ph	-$CH(OH)CH_2$−	A4	0.177
GPV 0128	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A11	0.259
GPV 0129	A	R1 = $COCH_2CH_2Ph$, R3 = OH	-$CH_2CH(OH)CH_2$−	-NH*n*Pr	3.021
GPV 0134	A	R3 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A8	2.535
GPV 0135	A	R2 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A1	0.424
GPV 0149	A	R3 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A2	6.875
GPV 0155	A	R1 = $CH(OH)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A8	0.671
GPV 0156	A	R1 = $CH(OCH_3)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A8	0.226
GPV 0157	A	R2 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A8	1.120
GPV 0159	A	R3 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-N(*i*Pr)$_2$	0.915
GPV 0163	A	R1 = $CH(OH)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-N(*i*Pr)$_2$	1.741
GPV 0164	A	R1 = $CH(OCH_3)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-N(*i*Pr)$_2$	0.658
GPV 0180	A	R1 = $COCH_2CH_2$-(1-naphthyl)	-$CH_2CH(OH)CH_2$−	A1	0.172
GPV 0181	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-$NHCH_2CH(Ph)_2$	0.797
GPV 0182	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-$NHCH(Ph)_2$	10.40
GPV 0184	A	R1 = $COCH_2CH_2$-(1-naphthyl)	-$CH_2CH(OH)CH_2$−	-$NHCH_2CH_2CH(Ph)_2$	0.723
GPV 0186	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_3$−	A1	0.651
GPV 0189	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_2$−	A1	1.452
GPV 0195	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_4$−	A1	0.533
GPV 0201	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_5$−	A1	0.242
GPV 0206	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_6$−	A1	0.203
GPV 0211	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_7$−	A1	0.179

Table 1 continued. Chemical structure and PGP-inhibitory activity (in μM) of propafenone derivatives.

Code	Core	R1 - R4	L	A	EC_{50}
GPV 0216	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_8$−	A1	0.137
GPV 0220	A	R1 = $CH(OCH_3)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-$NHCH_2CH_2CH(Ph)_2$	0.752
GPV 0226	A	R1 = $CH(OH)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A2	9.540
GPV 0227	A	R1 = $CH(OCH_3)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A2	1.807
GPV 0231	A	R1 = $COCH_2CH_2Ph$, R3 = OBz	-$CH_2CH(OH)CH_2$−	-NH*n*Pr	0.110
GPV 0232	A	R1 = $COCH_2CH_2Ph$, R3 = OBz	-$CH_2CH(OH)CH_2$−	A1	0.172
GPV 0233	A	R1 = $COCH_2CH_2Ph$, R3 = OH	-$CH_2CH(OH)CH_2$−	A1	1.727
GPV 0238	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-$NHCH_2CH_2CH(Ph)_2$	0.717
GPV 0240	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NHPh	6.467
GPV 0242	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NHcyclohexyl	0.341
GPV 0245	A	R1 = $COCH_2CH_2Ph$, R3 = OH	-$CH_2CH(OH)CH_2$−	-N(*i*Pr)$_2$	1.036
GPV 0253	A	R1 = $COCH_2CH_2Ph$, R3 = OBz	-$CH_2CH(OH)CH_2$−	A10	0.116
GPV 0264	A	R1 = $CH(iPr)CH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A1	0.460
GPV 0317	A	R1 = COPh	-$CH_2CH(OH)CH_2$−	A12	0.307
GPV 0319	A	R1 = COPh	-$CH_2CH(OH)CH_2$−	A8	0.147
GPV 0321	A	R1 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A11	0.357
GPV 0323	A	R1 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A12	3.197
GPV 0334	A	R1 = $COCH_2CH_2$(4-Cl-Ph)	-$CH_2CH(OH)CH_2$−	A7	0.019
GPV 0335	A	R1 = $COCH_2CH_2$(4-CH_3-Ph)	-$CH_2CH(OH)CH_2$−	A7	0.018
GPV 0336	A	R1 = $COCH_2CH_2$(4-OCH_3-Ph)	-$CH_2CH(OH)CH_2$−	A7	0.014
GPV 0338	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OAc)CH_2$−	A1	0.336
GPV 0339	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NH(Ph-4-$COOCH_3$)	1.536
GPV 0354	A	R1 = $COCH_2CH_2Ph$, R3 = Cl	-$CH_2CH(OH)CH_2$−	A7	0.059
GPV 0356	A	R1 = $COCH_2CH_2Ph$, R3 = OCH_3	-$CH_2CH(OH)CH_2$−	A7	0.179
GPV 0357	A	R1 = $COCH_2CH_2Ph$, R3 = CH_3	-$CH_2CH(OH)CH_2$−	A7	0.026
GPV 0358	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NH(Ph-4-CF_3)	8.490
GPV 0359	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NH(Ph-4-NO_2)	2.568
GPV 0360	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-N($COCH_2CH_3$)Bz	1.649
GPV 0361	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_4$−	A2	1.006
GPV 0363	A	R1 = $COCH_2CH_2Ph$	-$(CH_2)_4$−	A8	0.192
GPV 0366	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-N(*n*Pr)COPh	1.686
GPV 0374	A	R1 = $COCH_2CH_2$-(1-naphthyl)	-$CH_2CH(OH)CH_2$−	A2	0.730
GPV 0376	A	R1 = $COCH_2CH_2$-(1-naphthyl)	-$CH_2CH(OH)CH_2$−	-N(*i*Pr)$_2$	0.465
GPV 0381	A	R1 = $COCH_2CH_3$	-$CH_2CH(OH)CH_2$−	A12	0.299
GPV 0382	A	R1 = $COCH_2CH_2$-(1-naphthyl)	-$CH_2CH(OH)CH_2$−	A12	0.075
GPV 0384	A	R2 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A2	128.4
GPV 0385	A	R2 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A1	9.073
GPV 0386	A	R2 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A8	10.07
GPV 0388	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	A13	0.130
GPV 0389	A	R3 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A1	48.97
GPV 0390	A	R3 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A8	11.89
GPV 0391	A	R3 = $COCH_3$	-$CH_2CH(OH)CH_2$−	A2	302.1
GPV 0470	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NHCH(*i*Pr)Ph	0.210
GPV 0472	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NHCH(cyclohexyl)Ph	0.145
GPV 0476	A	R1 = $COCH_2CH_2Ph$, R3 = Cl, R4 = Cl	-$CH_2CH(OH)CH_2$−	A7	0.017
GPV 0479	A	R1 = $COCH_3$, R3 = CH_3	-$CH_2CH(OH)CH_2$−	A1	5.470
GPV 0485	E	R1 = H, R2 = O, R3 = H,H	-$CH_2CH(OH)CH_2$−	A1	9.684
GPV 0491	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$−	-NHCH(tBu)Ph	0.292
GPV 0512	E	R1 = H, R2 = H,H, R3 = O	-$CH_2CH(OH)CH_2$−	A1	11.19

Table 1 continued. Chemical structure and PGP-inhibitory activity (in μM) of propafenone derivatives.

Code	Core	R1 - R4	L	A	EC_{50}
GPV 0515	E	R1 = CH_2Ph, R2 = O, R3 = H,H	-$CH_2CH(OH)CH_2$–	A1	0.212
GPV 0522	E	R1 = CH_2Ph, R2 = H,H, R3 = O	-$CH_2CH(OH)CH_2$–	A1	0.142
GPV 0523	E	R1 = CH_2Ph, R2 = O, R3 = H,H	-$CH_2CH(OH)CH_2$–	A9	0.167
GPV 0543	D	———-	-CH_2–	A1	176.3
GPV 0570	A	R1 = $COCH_2CH_2Ph$	-CH_2COOEt	———	23.18
GPV 0571	A	R1 = $COCH_2CH_2Ph$	-CH_2COOH	———	1588
GPV 0574	A	R3 = $COCH_3$	-$CH_2CH(OH)CH_2$–	A9	1.349
GPV 0576	A	R1 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$–	A9	0.006
GPV 0577	A	R3 = $COCH_3$	-$CH_2CH(OH)CH_2$–	A7	1.550
GPV 0579	A	R2 = $COCH_3$	-$CH_2CH(OH)CH_2$–	A9	0.379
GPV 0590	C	R1 = CH_2CH_2Ph	-CO-	A1	6.658
GPV 0591	C	R1 = CH_3	-CO-	A1	103.6
GPV 0594	C	R1 = CH_2CH_2Ph	-CH_2–	A1	1.619
GPV 0595	C	R1 = CH_3	-CH_2–	A1	48.92
GPV 0596	A	R2 = $COCH_3$	-$CH_2CH(OH)CH_2$–	A7	0.537
GPV 0598	A	R1 = $COCH_3$	-$CH_2CH(OH)CH_2$–	A9	0.200
GPV 0600	A	R3 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$–	A9	0.227
GPV 0608	A	R1 = $COCH_2CH_2Ph$, R4 = OCH_3	-$CH_2CH(OH)CH_2$–	A7	0.296
GPV 0610	A	R1 = $COCH_2CH_2$-(4-$N(CH_3)_2$-Ph)	-$CH_2CH(OH)CH_2$–	A7	0.013
GPV 0613	A	R1 = $COCH_3$, R3 = OCH_3	-$CH_2CH(OH)CH_2$–	A7	0.859
GPV 0615	A	R1 = $COCH_3$, R3 = CH_3	-$CH_2CH(OH)CH_2$–	A7	0.396
GPV 0616	A	R1 = $COCH_3$, R3 = $COCH_3$, R4 = CH_3	-$CH_2CH(OH)CH_2$–	A7	0.211
GPV 0626	A	R1 = $COCH_3$	-$CH_2CH(OH)CH_2$–	A7	0.249
GPV 0633	A	R1 = $COCH_3$, R4 = CH_3	-$CH_2CH(OH)CH_2$–	A7	0.162
GPV 0636	A	R1 = COCH3, R3 = Cl, R4 = Cl	-$CH_2CH(OH)CH_2$–	A7	0.125
GPV 0643	A	R1 = $COCH_2CH_2$-(4-$N(CH_3)_2$-Ph), R4 = OCH_3	-$CH_2CH(OH)CH_2$–	A7	0.040
GPV 0645	A	R3 = $COCH_3$	-$CH_2CH(OH)CH_2$–	-$NHCH_2CH_2CH(Ph_2)$	0.178
GPV 0647	A	R2 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$–	A7	0.074
GPV 0649	A	R2 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$–	-$NHCH_2CH_2CH(Ph_2)$	0.190
GPV 0651	A	R2 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$–	A9	0.137
GPV 0653	A	R2 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$–	A2	3.475
GPV 0655	A	R3 = $COCH_2CH_2Ph$	-$CH_2CH(OH)CH_2$–	-$NHCH_2CH_2CH(Ph_2)$	0.505

focus on description of passive diffusion. It is widely accepted, that active efflux of drugs caused predominantly by P-glycoprotein (Pgp) plays a major role both for gastrointestinal uptake and for blood–brain barrier permeation [11]. Pgp is a membrane-bound ATP-dependent xenotoxin efflux pump and thus – despite its disputed physiological role – responsible for development of multidrug resistance in tumour therapy [12]. Pgp transports a wide variety of functionally and structurally diverse anticancer drugs, such as anthracyclines, podophyllotoxines, vinca alkaloids, taxanes, actinomycin D and colchicine. Identical structural diversity is found in the group of inhibitors, which comprises e.g. verapamil, amiodarone, dihydropyridines, acridonecarboxamides, steroids, cyclosporins and even detergents [13]. This promiscuity in binding represents a demanding challenge for the design of specific inhibitors, which might be useful as resistance modifiers in tumour therapy. In our studies, we focused on analogues of the class Ic antiarrhythmic agent propafenone. 2D- and 3D-QSAR studies revealed structural and conformational requirements for high ligand affinity [14]. Additionally evidence has been obtained, that propafenones are acting as substrates and thus block the pump via an extrusion/rediffusion cycle [15]. Therefore, our *in house* data set of 131 propafenone derivatives represents a versatile tool for exploration and validation of new

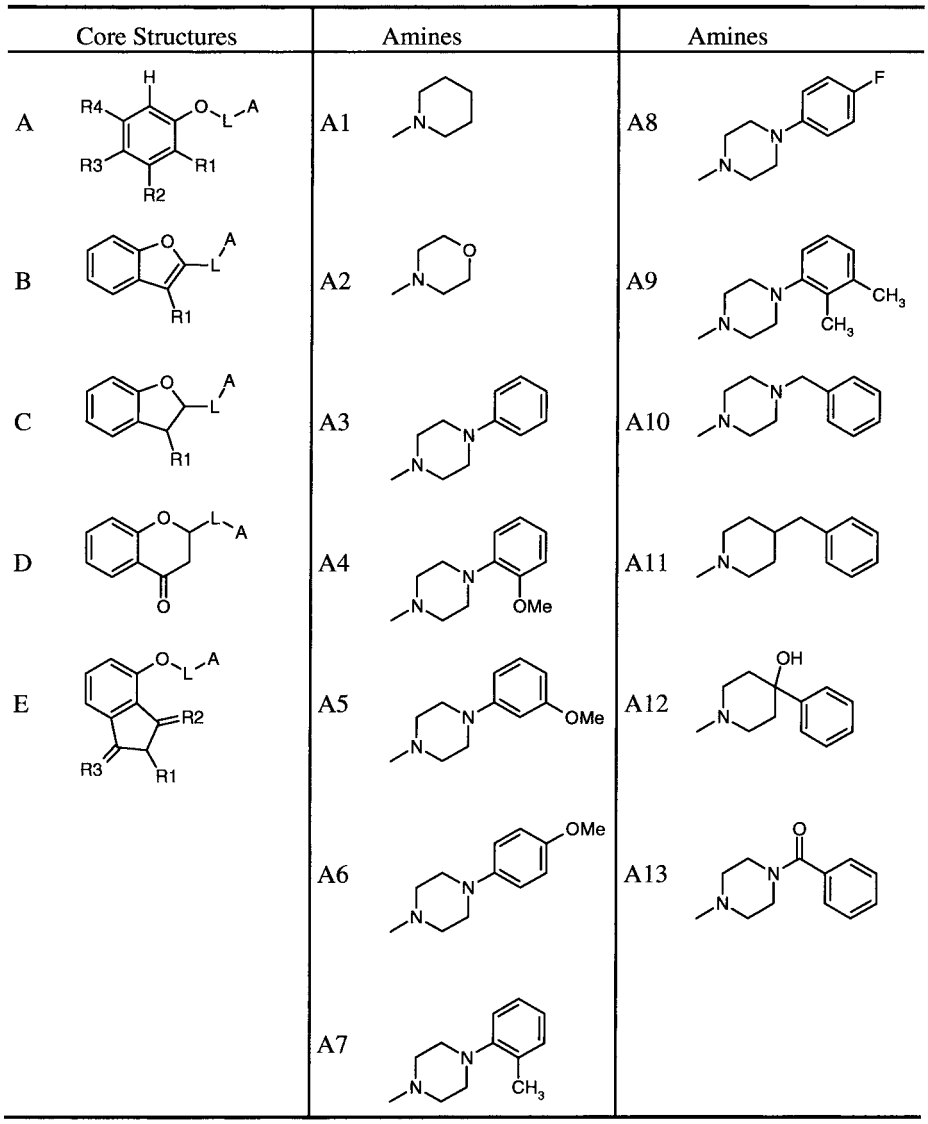

Figure 1. Core structures for propafenone derivatives.

decriptors for ADME profiling with a focus on efflux phenomena.

Table 1 gives the chemical structure and Pgp-inhibitory potency of the propafenone analogues. The corresponding core structures A-E and amine moieties A1-A13 are given in Figure 1. Compounds were synthesised in analogy to previously described procedures [16, 17, 18, 19]. Pharmacological testing was performed using a daunomycine efflux assay protocol [20].

SIBAR-descriptors

The approach for calculation of the SIBAR-descriptors is outlined as follows (Figure 2):
(1) selection of a reference compound set on basis of maximum diversity; (2) calculation of a set of descriptors for both the training set and the reference set; (3) calculation of similarity values for each compound of the training set to each compound of the reference set; this leads to a given number of similarity values (equal to the number of reference compounds used) for each compound of the training set, which are assigned as SIBAR-descriptors; (4) PLS analysis of the training

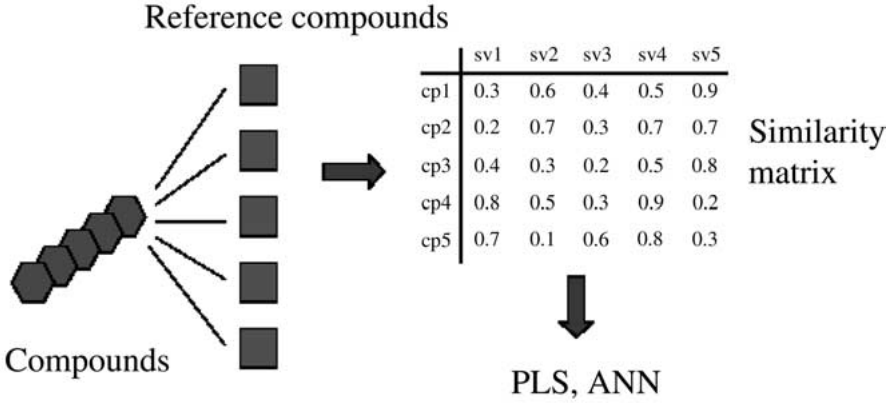

Figure 2. Workflow for the SIBAR-approach; cp: training set compound; sv: similarity value; PLS: partial least squares; ANN: artificial neural network.

Table 2. Models with the highest predictivity, as judged by SDEP, obtained with 39 molecular descriptors and different number of SIBAR-descriptors; NOD: number of descriptors; NOC: number of components in the PLS model.

	NOD	NOC	Q^2	SDEP
Molecular Descriptors	39		0.513	0.507
	10	4	0.660	0.434
	15	7	0.671	0.406
No of reference compounds	20	7	0.650	0.402
	25	7	0.657	0.407
	30	7	0.674	0.416
	40	7	0.653	0.423

set data matrix; (5) validation of the model using cross validation and an external test set.

Propafenones were built and minimised within the SYBYL program package [21]. Charges were calculated employing the semiempirical AM1 method [22]. For generation of the reference compound set, the SPECS library [23] was used. Compounds were treated in their uncharged forms and all metal complexes and compounds with reactive groups were removed from the library prior to diversity selection. For generation of maximally diverse subsets, the SELECTOR module implemented in SYBYL was used. Selection was performed on basis of Tanimoto indices using UNITY fingerprints [24]. Chemical structures of the compounds are given in Figure 3.

(2) In total 39 'classical' molecular descriptors, such as molecular volume, logP, molecular refractivity, dipole moment, connectivity indices ($^{i}\chi_{p}$, $\chi_{p/c}$, $^{i}\chi_{ch}$, $^{i}\chi_{p}^{v}$, $\chi_{p/c}^{v}$, $^{i}\chi_{ch}^{v}$) [25], shape indices (κ_i, $^{\alpha}\kappa_i$) [25], the sum of electrotopological indices [26], the Balaban topological index [27], and the number of hydrogen bond donors and hydrogen bond acceptors were calculated using the TSAR package [28].

(3) SIBAR-descriptors D were obtained *via* calculation of euclidian distances between the i reference compounds and the j compounds of interest (training set and test set) using k molecular descriptors, X_k:

$$D(i,j) = \sqrt{\sum_{k} \left(X_{ik} - X_{jk}\right)^2} \qquad (1)$$

(4) Partial least squares analysis (PLS), as implemented in TSAR, was used to correlate the calculated properties to the biological activity, which was expressed as $\log(1/EC_{50})$ values.

(5) The propafenone data set was divided into two subsets via random selection: a 100 compound training set and a 31 compound test set. The quality of the models obtained was estimated from statistical parameters as the squared correlation coefficient (r^2), the cross-validated correlation coefficient (Q^2), and the standard deviation of errors of prediction between predicted and actual log(potency) values (SDEP). In the cross-validation process groups of 10 compounds (i.e. leave 10% out procedure) were left out.

Results

Models using 10, 20, 30 and 40 highly diverse reference compounds chosen from the SPECS library were

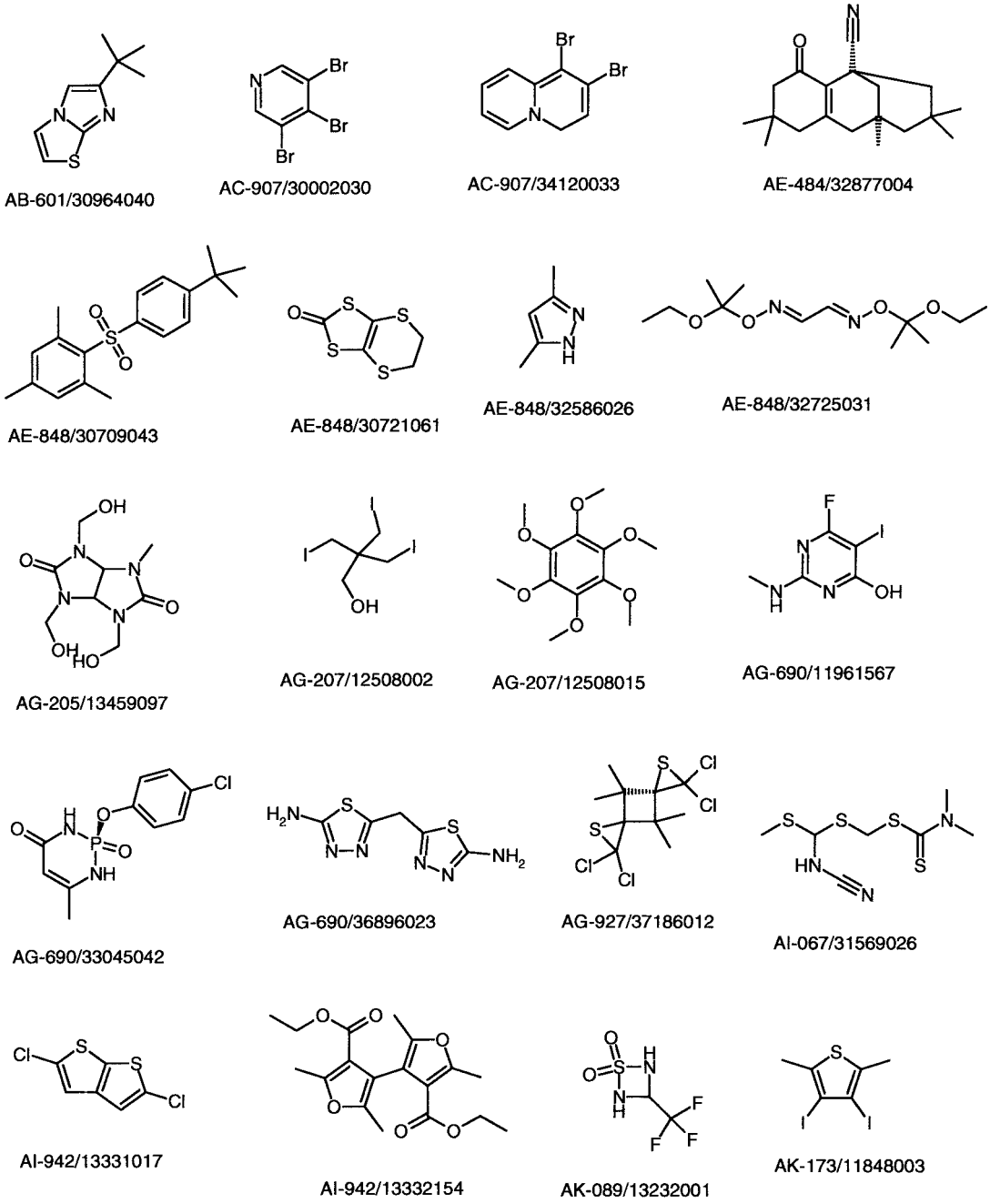

Figure 3. Chemical structures of the 20 maximally diverse compounds from the SPECS database.

generated. Best predictivity – as judged by the SDEP – was obtained with 20 compounds. Thus, also models with 15 and 25 compounds were derived to determine the optimum number of reference compounds more exactly. Table 2 shows the statistical parameters of the models obtained. Additionally, the corresponding statistical parameters for the best model obtained using the 39 molecular descriptors directly (i.e. lacking the additional calculation step for SIBAR-descriptors), is given. As can be seen, in all cases use of SIBAR-descriptors leads to models with higher predictive power than those obtained with molecular descriptors. This justifies the additional calculation step required for obtaining the SIBAR-descriptors.

Table 3. PLS models with different number of components obtained with the 39 molecular descriptors and 20 SIBAR-descriptors, respectively.

		Number of Components					
		1	2	3	4	5	6
39 Molecular descriptors	r^2	0.535	0.610	0.648	0.654	0.697	0.710
	Q^2	0.526	0.560	0.582	0.576	0.542	0.513
	SDEP	0.549	0.652	0.600	0.584	0.519	0.507
20 SIBAR descriptors	r^2	0.445	0.634	0.657	0.724	0.731	0.741
	Q^2	0.437	0.608	0.607	0.631	0.661	0.653
	SDEP	0.623	0.485	0.487	0.437	0.437	0.451

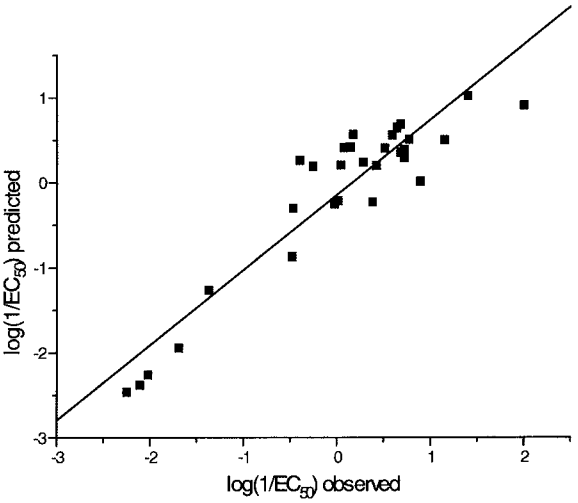

Figure 4. Plot of observed *vs.* predicted $\log(1/EC_{50})$ values for a test set of 31 propafenone analogues

The model obtained on basis of 20 reference compounds shows lowest SDEP. Table 3 shows in more detail a comparison of the model with the 39 molecular descriptors and those using 20 SIBAR-descriptors. With the exception of the single component model, the SIBAR-models generally showed higher predictive power.

Discussion

Development of *in silico* models for prediction of bioavailability is one the big challenges in drug development. While for prediction of passive diffusion through biological membranes several models have been published, active efflux phenomena have been disregarded. However, as recently outlined by H. van de Waterbeemd, bioavailability has to be treated as the combination of permeability + influx – efflux – gut wall metabolism – liver metabolism [29]. In this paper, we present first results on the use of a set of new descriptors (SIBAR) for prediction of interaction with the drug efflux pump P-glycoprotein, which plays a major role in the ADME scheme. The SIBAR descriptors are calculated on basis of similarity values to a set of highly diverse reference compounds. Thus, compounds are encoded via a similarity fingerprint, which undoubtedly reflects a more general description of the molecules rather than taking into account individual structural and conformational features. Our results demonstrate, that in the case of the rather 'unspecific' recognition pattern of Pgp, these descriptors give better results than the set of molecular descriptors used for calculation of the similarity values. Models were validated both *via* leave 10% out cross validation and via an external test set of 31 compounds (Figure 4). Almost identical results were obtained when analyzing the data set of Palm et al. [30], which comprizes intestinal absorption values for 20 diverse compounds. In this case, a predictive power (Q^2) of 0.74 was obtained [31]. However, applying the same method on the benchmark set of steroids, no statistically significant models were obtained (data not shown). This demonstrates that use of similarity fingerprints on basis of SIBAR-descriptors seems to be a versatile approach for description of biological processes in which a high promiscuity in ligand recognition takes place. This is especially the case when dealing with the influence of drug transport on ADME. Thus, use of SIBAR-descriptors seems to be a versatile new approach for development of predictive models for bioavailability. Future work will focus on improvements regarding the reference set (tailored for the specific problem) and the descriptors used for calculating the similarity values.

Acknowledgement

We gratefully acknowledge financial support from the Austrian Science Fund, grant # 13851-MOB and the Austrian National Bank grant # 8525. We are grateful to the ZID of the University of Vienna for providing us with several Software packages.

Acknowledgements

References

1. Kaliszan, R., Kaliszan, A. and Wainer, I.W., J. Pharm. Biomed. Anal., 11 (1993) 505.
2. Kansy, M., Senner, F. and Gubernator, K., J. Med. Chem., 41 (1998) 1007.
3. Sadowski, J. and Kubinyi, H., J. Med. Chem., 41 (1998) 3325.
4. Wagener, M. and van Geerestein, V. J., J. Chem. Inf. Comput. Sci., 40 (2000) 280.
5. Ertl, P., Rohde, B. and Selzer, P., J. Med. Chem., 43 (2000) 3714.
6. Cruciani, G., Pastor, M. and Guba, W., Eur. J. Pharm. Sci., 11 (2000) Suppl 2 S29.
7. Kubinyi, H., Hamprecht, F.A. and Mietzner, T., J. Med. Chem., 41 (1998) 2553.
8. Good, A.C., Peterson, S.J. and Richards, W.G., J. Med. Chem., 36 (1993) 2929.
9. Benigni, R., Cotta-Ramusino, M., Giorgi, F. and Gallo, G., J. Med. Chem., 38 (1995) 629.
10. Ghuloum, A.M., Sage, C.R. and Jain, A.N., J. Med. Chem., 42 (1999) 1739.
11. Ayrton, A. and Morgan, P., Xenobiotica, 31 (2001) 469.
12. Gottesma,, M.M., Fojo, T. and Bates, S.E., Nature Rev. Cancer, 2 (2002) 48.
13. Teodori, E., Dei, S., Scapecchi, S. and Gualtieri, F., Farmaco, 57 (2002) 385.
14. Ecker, G.F. and Chiba, P., Recent Res. Devel. Medicinal Chem., 1 (2001)121.
15. Schmid, D., Ecker, G., Richter, E., Hitzler, M. and Chiba, P., Biochem. Pharmacol., 58 (1999) 1447.
16. Chiba, P., Burghofer, S., Richter, E., Tell, B., Moser, A. and Ecker, G., J. Med. Chem., 38 (1995) 2789.
17. Ecker, G., Chiba, P., Hitzler, M., Schmid, D., Visser, K., Cordes, H.P., Csoellei, J., Seydel, J.K. and Schaper, K.J., J. Med. Chem., 39 (1996) 4767.
18. Salem, M., Richter, E., Hitzler, M., Chiba, P. and Ecker, G., Sci. Pharm., 66 (1998) 147.
19. Tmej, C., Chiba, P., Huber, M., Richter, E., Hitzler, M., Schaper, K.J. and Ecker, G., Arch. Pharm. Pharm. Med. Chem., 331 (1998) 233.
20. Ecker, G., Huber, M., Schmid, D. and Chiba, P., Mol. Pharmacol., 56 (1999) 791
21. SYBYL 6.7, Tripos Inc., Munich (Germany) 2001
22. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902
23. SPECS and BioSPECS B.V., Rijswijk, The Netherlands, 2001; http://www.specs.net
24. UNITY 4.3, Tripos Inc., Munich (Germany) 2001
25. Hall, L.H. and Kier, L.B., In Lipkowitz, K.B. and Boyd, D.B., (eds.) Reviews in Computational Chemistry Vol. 2, VCH Publishers, New York, NY, 1992, pp. 367–422.
26. Kier, L.B. and Hall, L.H., J. Chem. Inf. Comput. Sci., 31 (1991) 76.
27. Balaban, A.T., Chem. Phys. Lett., 89 (1982) 399.
28. TSAR 3.3, Oxford Molecular Ltd., The Medware Center, Oxford Science Park, Oxford 2000.
29. van de Waterbeemd, H., Drugs Fut., 27/Suppl. A (2002) 12 (L10)
30. Palm, K., Stenberg, P., Luthman, L., Artursson, P., Pharm. Res., 14 (1997) 568.
31. Ecker, G.F., unpublished results