Journal of Computer-Aided Molecular Design, **12:** 573–581, 1998.
KLUWER/ESCOM

Multivariate analysis of experimental and computational descriptors of molecular lipophilicity

Raimund Mannhold[a,*], Gabriele Cruciani[b], Karl Dross[c] & Roelof Rekker[d]
[a]*Institut für Lasermedizin, Arbeitsgruppe Molekulare Wirkstoff-Forschung, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, Germany;* [b]*Dipartimento di Chimica, Laboratorio di Chemiometria, Università di Perugia, Via Elce di Sotto, 8, I-06123 Perugia, Italy;* [c]*Institut für Hirnforschung, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, Germany;* [d]*Vakgroep Farmacochemie, Faculteit Scheikunde Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands*

Received 16 March 1998; Accepted 2 July 1998

Key words: calculation procedures, diversity searching in databases, linear regression, octanol/water partitioning (log P_{oct}), principal component analysis, reversed-phase thin-layer chromatography (R_{Mw})

Summary

Two experimental (log P, R_{Mw}) and 17 calculation descriptors for molecular lipophilicity (fragmental, atom-based or based on molecular properties) were investigated by multivariate analysis for a database of 159 compounds including both simple structures as well as more complex drug molecules. Principal component analysis (PCA) of the entire database exhibits a clustering of chemical groups; preciseness of clustering corresponds to chemical similarity. Thus, diversity searching in databases might effectively be performed by PCA on the basis of calculated log P. The comparative validity check of experimental and computational procedures by regression analysis and PCA was performed with a chemically balanced, reduced data set ($n = 55$) representing 11 chemical groups with 5 members each. Regression of experimental descriptors (log P_{oct} versus R_{Mw}) proves that chromatographic data, obtained under well-defined experimental conditions, can be used as valid substitutes for log P. Regression of calculated versus experimental lipophilicity data shows a superiority of fragmental over atom-based methods and approaches based on molecular properties, as indicated by correlation coefficients, slopes and intercepts. In addition, PCA revealed that fragmental methods (Rekker-type, KOWWIN, KLOGP) sense the compound ranking in log P data to almost the same extent as experimental approaches. For atom-based procedures and CLOGP, both the comparability of absolute values and the sensing of the compound ranking in the database are slightly less. This trend is more pronounced for the methods based on molecular properties, with the exception of BLOGP.

Introduction

Lipophilicity is a molecular property which expresses the relative affinity of a solute for aqueous and organic phases, respectively. Accordingly, lipophilicity encodes most of the intramolecular forces that can take place between a solute and a solvent. Lipophilicity is a composite property and its physico-chemical nature can be derived from relations between log P and other molecular properties. From several lines of evidence it was shown that log P can be factorized into a bulk or volume term and an electronic or polar term.

The complex information circumscribed by log P might explain its successful application in countless QSAR studies. The breakthrough in defining the role of lipophilicity for biological properties of drug molecules at a molecular level is coupled with the names of Hansch, Fujita and Leo, who introduced substituent constants as lipophilicity descriptors and detected their additive, constitutive character. Their investigations made it possible to unravel the comprehensive impact of lipophilicity on almost all aspects of drug

*To whom correspondence should be addressed.

Table 1. Overview of calculation methods

Program	Method	Reference
Fragmental methods		
PROLOGP_cdr	original Rekker	7
SANALOGP_EO		
Σf-SYBYL	revised Rekker	8
SANALOGP_ER		
KOWWIN	Meylan-Howard	12
CLOGP 3.54	Leo-Hansch	9, 10
CLOGP 4.34		
KLOGP	Klopman	11
Atom-based methods		
MOLCAD	Ghose-Crippen	17
Tsar 2.2		13–16
PROLOGP_atomics		13–16
CHEMICALC2	Suzuki	18
SMILOGP	Dubost	19
Combined fragmental and atom-based method		
PROLOGP_comb	Rekker + Ghose-Crippen	
Methods based on molecular properties		
ASCLOGP	van de Waterbeemd	21
HINT	Abraham-Kellog	20
BLOGP	Bodor	

action including both their pharmacokinetic behaviour as well as the drug interaction with their biological macromolecular counterparts.

The outstanding importance of lipophilicity in QSAR and drug design makes it imperative that quick, precise and reproducible experimental approaches to quantify this physico-chemical property be available. The Hansch group introduced the determination of partition coefficients in the octanol-water system as the standard.

The need to derive lipophilicity data for steadily increasing numbers of compounds initiated the search for both experimental and computational alternatives to octanol-water partitioning. Computational approaches are either atom-based or use fragments; recently, attention has been paid to the impact of 3D aspects on lipophilicity.

Application of calculation approaches demands a validity check with experimental data. In this study 2 experimental (log P_{Oct}, R_{Mw}) and 17 calculated

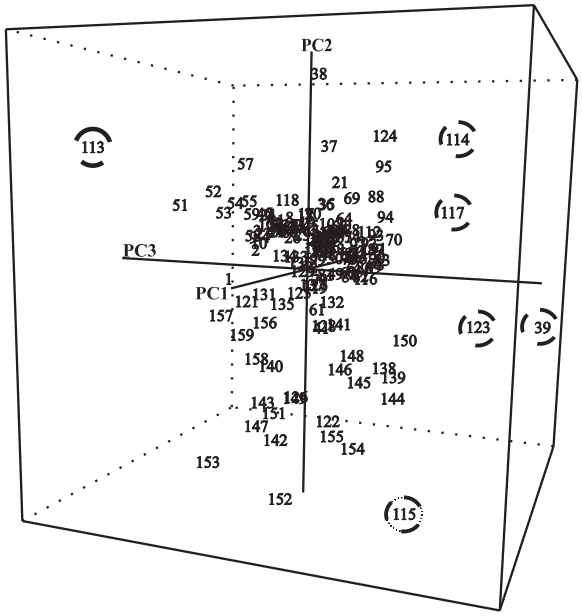

Figure 1. Score plot of the initial PCA with the entire data set ($n = 159$). It is apparent that 4-Br-acetophenone (no. 39), ethmozine (no. 115), propafenone (no. 123), carocainide (no. 113), disopyramide (no. 114) and indecainide (no. 117) are outliers, for further details see text.

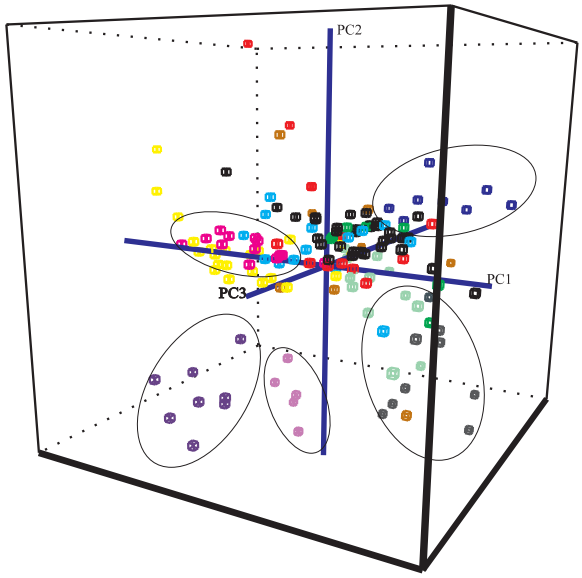

Figure 2. Score plot of PCA with the reduced set ($n = 153$). Chemical groups are colour coded and one can easily detect a clustering, most pronounced for imidazoles (dark blue), halo-benzenes (magenta), benzamides (grey) and phenothiazines; for the latter even a subgroup containing a piperazine ring (dark violet) is separated from those lacking this ring (light violet). Aromatic acids (black) or neutral aromates (yellow) show less pronounced clustering, since they are more broadly defined.

descriptors are compared by means of multivariate analysis.

Materials and methods

Experimental data. The octanol/water partition coefficients used in the present study are taken from the tabulation of Hansch and coworkers [1]; in some cases they stem from Mannhold et al. [2, 3] or Taylor and Cruickshank [4]. Chromatographic lipophilicity data were derived by reversed-phase thin-layer chromatography (RP-TLC), as described in detail by Dross et al. [5]. For the simple compounds they were published by Dross et al. [6].

Calculated data. Computerized lipophilicity calculations were done with 17 programs (Table 1) covering the main methodological approaches currently in use: Σ*f*-SYBYL [7, 8], CLOGP [9, 10], PROLOGP_cdr, SANALOGP_EO, SANALOGP_ER, KLOGP [11] and KOWWIN [12] represent fragmental methods; atom-based approaches [13–19] are PROLOGP_atomics and Tsar 2.2 [13–16], MOLCAD (17), CHEMICALC-2 [18] and SMILOGP [19]; methods based on molecular properties comprise HINT [20], BLOGP and ASCLOGP [21]. A special case is PROLOGP_comb, which combines a fragmental (Rekker) and an atom-based approach (Ghose/Crippen).

Statistical analysis. Principal component analysis [22] was performed with the GOLPE [23] software, version 3.1, on a Silicon Graphics workstation. The MREG option of the SIMCA 3B program [24] was used for regression analysis.

Results

Experimental lipophilicity data. Experimental log P data and R_{Mw}-values are given in Table 2. Both sets are not complete; some compounds are either not accessible to RP-TLC or log P data are missing in the literature. In total, 141 R_{Mw} and 140 log P values are available. R_{Mw} data range from 6.94 (9, 10-di-phenyl-anthracene) to −0.13 (imidazole); the corresponding log P values range from 6.01 (9-phenyl-anthracene) to −0.22 (acetic acid). Lack of coincidence is due to missing data.

Principal component analysis (PCA) of the entire data set. A first PCA of the entire set revealed

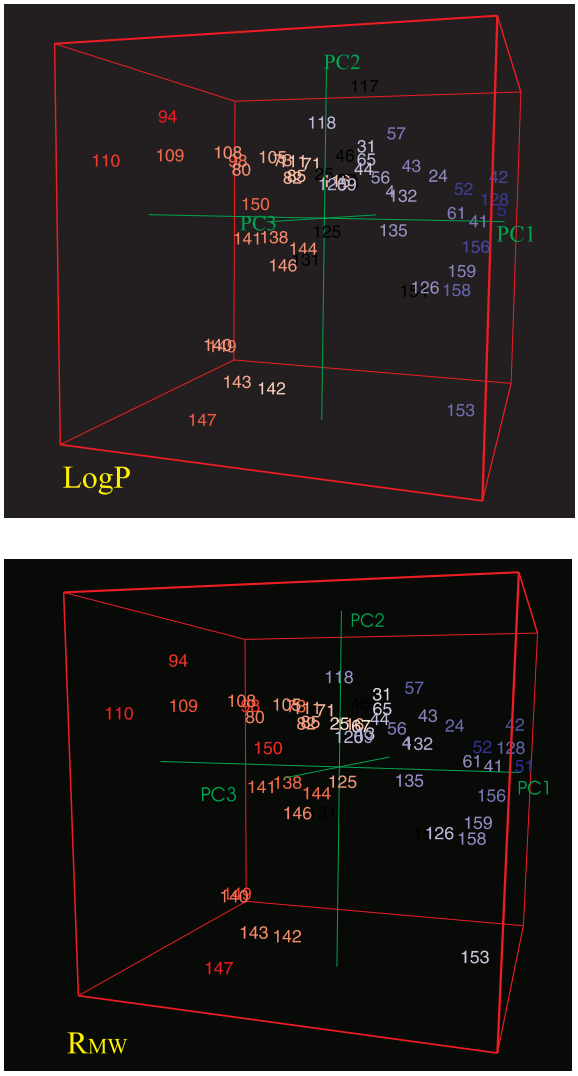

Figure 3. Score plot of the PCA with the reduced, balanced set ($n = 55$), shown for log P- and R_{Mw}-data. Colour coding reflects a continuous spectrum of high (red) to low log P (blue) and indicates the very similar pattern of test compounds for the two experimental procedures, thus substantiating their mutual exchangeability.

that the data can be described with a 3-component model explaining about 95% of the variance in X. From the score plot (Figure 1) it is obvious that 4-Br-acetophenone (no. 39), ethmozine (no. 115), propafenone (no. 123), carocainide (no. 113), disopyramide (no. 114) and indecainide (no. 117) are outliers. For the first 3 compounds this is due to more than 50% missing data. Carocainide dominated the third component of the initial PCA due to its anomalous ASCLOGP value. Disopyramide and indecainide contain aliphatic carbonamide moieties, obviously mis-

Table 2. Chromatographic and partitioning lipophilicity descriptors for the entire data set

No.	Compound	R_{Mw}	log P	No.	Compound	R_{Mw}	log P
1	acetic acid	–	−0.17	∗51	imidazole	−0.13	−0.08
2	propionic acid	–	0.33	∗52	2-methyl-imidazole	0.01	0.24
3	butyric acid	–	0.79	53	2-ethyl-imidazole	0.16	–
∗4	benzoic acid (ba)	1.65	1.87	54	2-propyl-imidazole	0.31	–
5	2-methyl-ba	1.97	2.46	55	2-butyl-imidazole	0.61	–
6	3-methyl-ba	2.21	2.37	∗56	2-phenyl-imidazole	1.17	1.88
7	4-methyl-ba	2.22	2.27	∗57	benzimidazole	0.82	1.32
8	3,4-dimethyl-ba	2.67	–	58	2-methyl-benzimidazole	0.92	–
9	3-methoxy-ba	1.80	2.02	∗59	5,6-di-methyl-benzimidazole	1.69	2.35
10	4-methoxy-ba	1.95	1.96	60	aniline	–	0.90
11	3-F-ba	1.76	2.15	∗61	4-nitro-aniline	1.42	1.39
12	4-F-ba	1.80	2.07	62	4-Cl-aniline	1.69	1.88
∗13	3-Cl-ba	2.11	2.68	63	4-Br-aniline	1.99	2.26
14	4-Cl-ba	2.19	2.65	64	2,5-di-tert.butyl-aniline	4.38	–
15	3-Br-ba	2.27	2.87	∗65	2-naphthylamine	2.20	2.28
16	4-Br-ba	2.37	2.86	66	4-Br-1-naphthylamine	3.34	–
17	3-I-ba	2.54	3.13	∗67	2-amino-biphenyl	2.99	2.84
18	4-I-ba	2.63	3.02	68	2-amino-fluorene	3.42	3.14
19	4-butyl-ba	3.94	3.97	69	2-amino-7-Br-fluorene	4.11	3.92
20	4-pentyl-ba	4.43	–	70	2-amino-1,3-di-Br-fluorene	4.87	–
21	4-heptyl-ba	5.44	–	∗71	1-amino-anthracene	3.56	3.69
22	2-OH-ba	1.17	2.26	72	3-amino-fluoranthene	4.59	4.20
23	4-OH-ba	1.07	1.58	∗73	1-amino-pyrene	4.66	4.31
∗24	2,4-di-OH-ba	0.87	1.63	74	acridine	3.40	3.40
∗25	1-naphthalenecarbox.acid	3.05	3.10	75	4-nitrotoluene	2.61	2.37
26	phenylacetic acid	–	1.41	76	4-Cl-nitro-benzene	2.70	2.39
27	3-methyl-phenylacetic acid	2.05	1.95	77	4-Br-nitro-benzene	2.76	2.55
28	3-F-phenylacetic acid	1.64	1.65	78	1-nitro-naphthalene	3.25	3.19
29	4-F-phenylacetic acid	1.65	1.55	79	benzene	–	2.13
30	4-Cl-phenylacetic acid	2.17	2.12	∗80	penta-methyl-benzene	4.35	4.56
∗31	4-Br-phenylacetic acid	2.31	2.31	81	toluene	–	2.73
32	3-phenylpropionic acid	2.10	1.84	∗82	biphenyl	3.92	3.98
33	4-phenylbutyric acid	2.53	2.42	83	bibenzyl	4.68	4.79
34	benzophenone (bp)	3.36	3.18	84	naphthalene	3.17	3.30
35	2,6-di-methyl-bp	4.04	–	85	2-methyl-naphthalene	3.75	3.86
36	2,2-di-methyl-bp	4.12	–	∗86	2,6-di-methyl-naphthalene	4.29	4.31
37	2,2,6,6-tetra-methyl-bp	4.46	–	87	1-phenyl-naphthalene	4.53	–
38	2,2,6,6-tetra-ethyl-bp	6.13	–	88	2,6-di-tert.butyl-naphthalene	6.42	–
39	4-Br-acetophenone	2.94	2.43	89	anthracene	4.23	4.45
40	benzamide	–	0.64	90	2-methyl-anthracene	4.62	5.00
∗41	2-OH-benzamide	1.38	1.28	91	2-ethyl-anthracene	5.09	5.85
∗42	4-OH-benzamide	0.46	0.33	92	2-Cl-anthracene	4.75	–
∗43	phenol	1.28	1.46	93	9-Br-anthracene	4.96	–
∗44	4-Cl-phenol	2.03	2.39	∗94	9-phenyl-anthracene	5.16	6.01
45	4-Br-phenol	2.22	2.59	95	9,10-di-phenyl-anthracene	6.94	–
46	1-naphthol	2.57	2.84	96	fluoranthene	–	5.16
∗47	2-naphthol	2.58	2.70	97	pyrene	–	4.88
48	benzylalcohol	–	1.10	∗98	2-methyl-phenanthrene	5.16	5.15
49	4-methyl-benzylalcohol	1.92	1.58	99	Cl-benzene	–	2.89
50	4-Cl-benzylalcohol	2.12	1.96	100	1,2-di-Cl-benzene	–	3.43

Table 2 continued.

No.	Compound	R_{Mw}	log P	No.	Compound	R_{Mw}	log P
101	1,3-di-Cl-benzene	–	3.53	∗131	metipranolol	2.75	2.66
102	1,4-di-Cl-benzene	–	3.44	132	metoprolol	1.84	1.88
103	1,2,3-tri-Cl-benzene	–	4.05	∗133	oxprenolol	2.20	2.10
104	1,2,4-tri-Cl-benzene	–	4.02	134	penbutolol	4.19	4.15
∗105	1,3,5-tri-Cl-benzene	4.05	4.19	∗135	pindolol	1.45	1.75
106	1,2,3,4-tetra-Cl-benzene	–	4.64	136	propranolol	2.99	2.98
107	1,2,3,5-tetra-Cl-benzene	–	4.66	137	sotalol	0.56	0.59
∗108	1,2,4,5-tetra-Cl-benzene	4.52	4.60∗	138	alimemazine	4.73	4.71
∗109	penta-Cl-benzene	4.90	5.18	139	chlorpromazine	5.38	5.19
∗110	hexa-Cl-benzene	5.36	5.73	∗140	fluphenazine	4.50	4.36
∗111	1,4-di-Br-benzene	3.88	3.79	∗141	levomepromazine	4.59	4.68
112	aprindine	4.75	4.86	∗142	perazine	4.19	3.61
113	carocainide	2.00	1.38	∗143	perphenazine	4.21	4.20
114	disopyramide	2.42	2.58	∗144	promazine	4.62	4.55
115	ethmozine	3.73	2.98	145	promethazine	4.74	4.81
∗116	flecainide	3.64	3.78	∗146	sulforidazine	4.25	4.45
∗117	indecainide	2.38	3.11	∗147	thiethylperazine	5.40	5.41
∗118	lidocaine	1.48	2.26	148	thioridazine	5.76	5.90
119	lorcainide	4.76	4.85	∗149	trifluoperazine	5.10	5.03
∗120	mexiletine	1.93	2.15	∗150	trifluopromazine	5.22	5.19
121	nicainoprol	2.40	1.63	151	alizapride	1.86	1.79
122	procainamide	0.98	0.88	152	alpiropride	1.91	1.69
123	propafenone	4.50	4.63	∗153	amisulpride	2.11	1.10
124	quinacainol	3.99	3.63	∗154	bromopride	2.50	2.83
∗125	quinidine	3.96	2.88	155	metoclopramide	2.39	2.62
∗126	acebutolol	1.98	1.71	∗156	sulpiride	1.08	0.42
127	alprenolol	3.00	3.10	157	sultopride	1.85	1.06
∗128	atenolol	0.74	0.16	∗158	tiapride	1.38	0.90
129	bunitrolol	1.84	1.91	∗159	veralipride	1.43	1.47
130	bupranolol	2.86	2.80				

The entire database consists of 159 molecules comprising 111 'simple' organics and 48 complex drugs including 14 antiarrhythmics, 12 β-blockers, 13 phenothiazines and 9 benzamides. Most octanol/water partition coefficients are taken from Hansch and coworkers [1]; in some cases they stem from Mannhold et al. [2, 3] or Taylor and Cruickshank [4]. Chromatographic lipophilicity data were derived by reversed-phase thin-layer chromatography, as described in detail by Dross et al. [5]. For the simple compounds they were published by Dross et al. [6]. Asterisked compounds were selected for the reduced, balanced data set ($n = 55$).

calculated by most of the procedures. When these 6 outliers were eliminated and PCA was repeated, the reduced set of 153 compounds (Figure 2) exhibited a pattern quite similar to that of the entire set. In Figure 2 chemical groups are colour coded and one can easily detect a clustering due to chemical similarity, most pronounced for imidazoles, halo-benzenes, benzamides and phenothiazines; in the latter group, compounds with a piperazine ring are distinguishable from compounds lacking this moiety. The proximity effects between the two piperazine nitrogens are not sufficiently reflected by several computational procedures. Clustering in chemically less homogeneous groups such as aromatic acids or neutral aromates is less pronounced.

Comparison of regression analysis and PCA for a balanced, reduced data set. Chemical clustering of the entire data set allowed to select a balanced set ($n = 55$) containing 11 chemical groups with 5 members each (aromatic acids, aromatic alcohols, imidazoles, amines, neutral aromates, halobenzenes, antiarrhythmics, β-blockers, phenothiazines with or without piperazine, benzamides). These data were

compared by means of regression analysis and PCA. Regression of log P_{Oct} versus R_{Mw} data gives a good correlation, with a slope near 1.0 and an intercept near 0.0, indicating that properly measured R_{Mw} values can be used as substitutes for log P_{oct}:

$$R_{mw} = 0.994(\pm 0.03)\log P_{oct} - 0.009(\pm 0.09), \quad (1)$$
$$n = 55;\ r = 0.995;\ s = 0.169;\ F = 5707$$

Regression analysis of calculated versus experimental lipophilicity data (Table 3) shows in general a closer fit of calculation data to log P_{oct} than to R_{Mw}. Comparing calculated and experimental log P, statistical data (correlation coefficients, slopes and intercepts) indicate the following quality ranking: fragmental methods > atom-based methods > approaches based on molecular properties. SMILOGP may serve to exemplify the detailed information obtained from slope and intercept; the rather high intercept of −0.661 indicates some constant error in this approach, while the slope, very close to 1.0, proves that the method adequately senses the compound ranking in the database.

Comparing calculated log P with chromatographic data substantiates the superiority of fragmental methods over the other approaches. Atom-based methods and approaches based on molecular properties exhibit equivalent validity in this case.

The same data set was investigated with the PCA option of the GOLPE program. The score plot is comparatively shown in Figure 3 for log P- and R_{Mw}-data. Colour coding reflects a continuous spectrum of high (red) to low lipophilicity (blue), indicating the very similar pattern of test compounds for the two experimental approaches.

The loading plot of the first versus the second principal component (Figure 4) gives a comparative overview of the quality of the respective calculation procedures. The information content of this figure is twofold: (1) the distance between the projection of the data points for experimental and calculation procedures onto the first component indicates the similarity in absolute values for experiment and calculation; (2) the deviation of the loading direction of a given variable from the direction of the first PC indicates the similarity between calculation and experiment in reflecting the compound ranking in log P-data within a database.

Thus, the fragmental methods (Rekker-type, KOWWIN, KLOGP) compare most favorably to experimental log P and also capture the compound ranking in log P data to almost the same extent as the experimental approaches. For atom-based procedures and CLOGP both the comparability of absolute values and in compound ranking within the database is slightly less. This tendency is even more pronounced for the methods based on molecular properties except for BLOGP, which captures the compound ranking similarly to fragmental methods, while the comparability in absolute values is limited, indicating some systematic error in this approach.

Discussion

The central role of lipophilicity for various aspects of research in medicinal chemistry is undisputed [25, 26]. Thus, the availability of precise approaches to quantify lipophilicity with experimental and calculation methods is of continuing interest [27–34]. In the present study, we compared experimental and computational lipophilicity descriptors by means of regression analysis and PCA.

Experimental lipophilicity data. Octanol/water partition coefficients or corresponding chromatographic data are commonly used as experimental descriptors of molecular lipophilicity. While the scientific community prefers HPLC as the chromatographic approach, we have elaborated experimental conditions for RP-TLC (5) which yield reliable R_{Mw} values. Log P_{Oct} data from the literature (1–4) are compared in this study with own RP-TLC data for 159 compounds. The range of data comprises 7 log units; compounds are appropriately diverse in structure, including both simple organics and complex drug molecules. Regression analysis and PCA clearly show that the RP-TLC data, obtained under proper experimental conditions, can be used as valid alternatives to log P_{oct}.

Log P as a diversity descriptor in database searching. Large databases can be synthesized by means of combinatorial chemistry or automated synthesis technologies. The information hidden in a database resides in its molecular diversity rather than in its size. Definition of diversity remains difficult. Molecular diversity for instance depends on the molecular properties used in the description phase: using too few descriptors makes the clustering of compounds hard to recognize, while including too many descriptors can result in a biased model. Thus, diversity can depend on the selection of appropriate molecular descriptors. While there is no strict rule for their selection, a small number of highly informative descriptors is clearly superior to

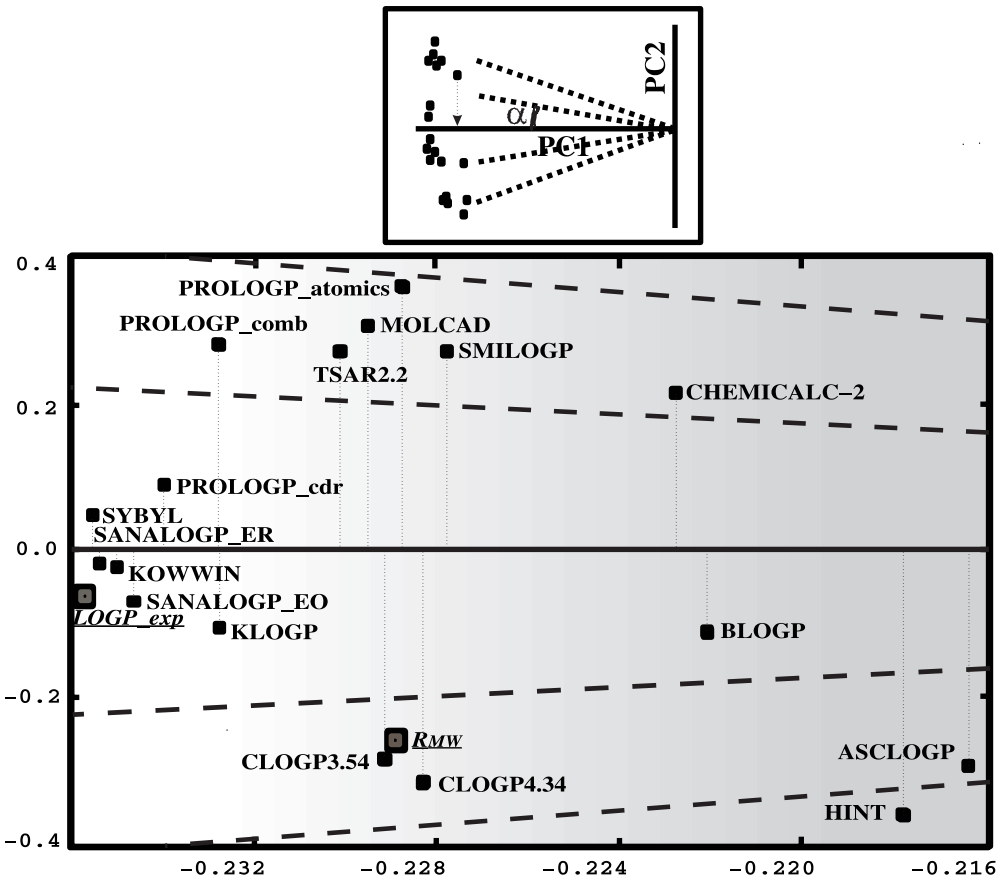

Figure 4. PCA with the reduced, balanced set ($n = 55$): loading plot of the first versus the second component. The information content of this figure is twofold: (1) The distance between the projection of the data points (arrow in the upper scheme) for experimental and computational procedures onto the first component indicates the similarity in absolute values for experiment and calculation. (2) The deviation of the loading direction of a given variable from the direction of the first PC (angle α in the upper scheme) reflects the similarity between calculation procedure and the experimental approach in sensing the compound ranking in the database.

a large number of less informative ones. In addition, the chemical meaning and the interpretation of the molecular descriptors should be taken into account, as should their accuracy and the time needed for their determination.

We have shown here that computational methods for log P can be useful to address the molecular diversity problem. Log P itself is a 'latent variable' comprising a variety of molecular poperties. In addition, different calculation methods are able to highlight different molecular properties. The most important features of chemical libraries can thus be discovered by combining a few different calculation methods as descriptors for the X-space. Moreover, log P calculations are in general fast and reliable.

Our present data indicate that diversity searching in databases could effectively be performed by PCA on the basis of log P, as shown by the nice clustering of chemical groups due to this parameter. Not surprisingly, precision of clustering corresponds to the similarity of chemical structures within a group. The loading plot in Figure 4 indicates that an appropriate set of calculation procedures to perform diversity searching could consist of KOWWIN, CLOGP, HINT and MOLCAD.

Validity check of calculation methods by regression analysis and PCA. Regression analysis of calculated versus experimental data shows that in general fragmental methods are superior to atom-based and 3D-related approaches. These results are in accord with our earlier analysis with a smaller dataset [35]. A limited applicability is often attributed to fragmental methods due to missing fragment values. This is

Table 3. Overview of regression data

Program	log P_{Oct}			R_{Mw}		
	r	*a*	*i*	*r*	*a*	*i*
PROLOGP_cdr	0.961	1.027	−0.045	0.927	1.001	0.101
SANALOGP_EO	0.964	1.035	0.024	0.936	1.015	0.152
Σ*f*-SYBYL	0.969	1.058	−0.119	0.927	1.023	0.057
SANALOGP_ER	0.969	1.060	−0.096	0.934	1.032	0.058
KOWWIN	0.984	1.019	−0.111	0.959	1.003	0.006
CLOGP 3.54	0.962	1.110	−0.211	0.949	1.105	−0.121
CLOGP 4.34	0.958	1.084	−0.168	0.950	1.085	−0.097
KLOGP	0.978	0.940	0.216	0.954	0.928	0.319
MOLCAD	0.939	0.786	0.288	0.883	0.746	0.456
Tsar 2.2	0.943	0.813	0.156	0.892	0.777	0.316
PROLOGP_at.	0.946	0.838	0.145	0.883	0.790	0.340
CHEMICALC-2	0.921	0.901	−0.171	0.877	0.867	−0.009
SMILOGP	0.942	1.014	−0.661	0.900	0.978	−0.487
PROLOGP_comb	0.960	0.888	0.094	0.906	0.846	0.277
ASCLOGP	0.903	0.921	0.282	0.901	0.928	0.326
HINT	0.899	1.041	−0.040	0.890	1.041	0.033
BLOGP	0.918	0.926	0.311	0.921	0.938	0.340

Correlation coefficients (r), slopes (a) and intercepts (i) for regression analyses of the computational procedures versus experimental partitioning (log P_{Oct}) and chromatography data (R_{Mw}).

true for CLOGP and Rekker-type methods, but not for KOWWIN.

Information obtained by PCA on the same dataset, in general parallels the regression data, but unravels more precisely the comparability in absolute values and the reflection of compound ranking in the database. Accordingly, the ranking of calculation methods observed in regression analysis is confirmed, exceptional behaviour of CLOGP and BLOGP, however, is only detected by PCA.

Conclusions

Results of this investigation are derived with a database of 159 compounds. Enlarging the database would significantly increase the statistical robustness of the analysis. Present data indicate it worthwhile to derive separate quality rankings for various chemical classes of interest; e.g. KOWWIN, shown to be excellent for drug molecules, exhibits a reduced predicitivity for simple organics. Future refinements of the 3D related approaches will show if they are able to compete with fragmental methods. A special attention should be devoted to demonstrating the value of log P as a discriminative parameter in database searching.

Acknowledgements

Part of this work was done, while R.M. was a guest of the Laboratory of Chemometrics, University of Perugia, Italy; he wants to express his gratitude to Sergio Clementi and Gabriele Cruciani for their cooperation and hospitality. We thank the following colleagues for providing log P data: Z. Bencz and F. Csizmadia (PROLOGP), M. Bohl (Σ*f*-SYBYL), C. Cook and G.E. Kellogg (HINT), J.P. Dubost (SMILOGP), M. Kansy (ASCLOGP), G. Klopman (KLOGP), W. Meylan (KOWWIN), D. Petelin (SANALOGP), E.E. Polymeropoulos (MOLCAD), A. ter Laak (CHEMICALC-2) and H. van de Waterbeemd (CLOGP and Tsar 2.2).

References

1. Hansch, C., Leo, A. and Hoekman, D., Exploring QSAR. Hydrophobic, Electronic, and Steric Constants. American Chemical Society, Washington, DC, 1995.
2. Mannhold, R., Dross, K. and Rekker, R.F., Quant. Struct.-Act. Relat., 9 (1990) 21.
3. Mannhold, R., Rekker, R.F., Sonntag, C., ter Laak, A.M., Dross, K. and Polymeropoulos, E.E., J. Pharm. Sci., 84 (1995) 1410.
4. Taylor, P.J. and Cruickshank, J.M., J. Pharm. Pharmacol., 37 (1985) 143.

5. Dross, K., Sonntag, Ch. and Mannhold, R., J. Chromatogr., A 638 (1993) 287.
6. Dross, K., Sonntag, Ch. and Mannhold, R., J. Chromatogr., A 673 (1994) 113.
7. Rekker, R.F., The Hydrophobic Fragmental Constant. Pharmacochemistry Library, Vol. 1, Elsevier, Amsterdam, 1977.
8. Rekker, R.F. and Mannhold, R., Calculation of Drug Lipophilicity, VCH, Weinheim, 1992.
9. Leo, A., Jow, P.Y.C., Silipo, C. and Hansch, C., J. Med. Chem., 18 (1975) 865.
10. Hansch, C. and Leo, A.J., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, NY, 1979.
11. Klopman, G., Li, J.-Y., Wang, S. and Dimayuga, M., J. Chem. Inf. Comput. Sci., 34 (1994) 752.
12. Meylan, W. and Howard, P., J. Pharm. Sci., 84 (1995) 83.
13. Ghose, A.K. and Crippen, G.M., J. Comput. Chem., 7 (1986) 565.
14. Ghose, A.K. and Crippen, G.M., J. Chem. Inf. Comput. Sci., 27 (1987) 21.
15. Ghose, A.K., Pritchett, A. and Crippen, G.M., J. Comput. Chem., 9 (1988) 80.
16. Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
17. Brickmann, J. and Waldherr-Teschner, M., Informationstechnik, 33 (1991) 83.
18. Suzuki, T. and Kudo, Y., J. Comput.-Aided Mol. Design, 4 (1990) 155.
19. Convard, T., Dubost, J.-P., Le Solleu, H. and Kummer, E. Quant. Struct.-Act. Relat., 13 (1994) 34.
20. Kellogg, G.E. and Abraham, D., J. Comput.-Aided Mol. Design, 5 (1991) 545.
21. Van de Waterbeemd, H., Karajiannis, H., Kansy, M., Obrecht, D., Mueller, K. and Lehmann, Ch., Conformation-lipophilicity relationships of peptides and peptide mimetics. In Sanz, F. (Ed.) Trends in QSAR and Molecular Modeling 94, Prous, Barcelona (1996).
22. Wold, S., Esbensen, K. and Geladi, P., Chem. Intell. Lab. System, 2 (1987) 37.
23. Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R. and Clementi, S., Quant. Struct.-Act. Relat., 12 (1993) 9.
24. Wold, S., University of Umeå, 1983.
25. Lipophilicity in Drug Action and Toxicology (eds. V. Pliska, B. Testa and H. van de Waterbeemd), in: Methods and Principles in Medicinal Chemistry, VCH Publishers, Weinheim, 1996.
26. Leo, A., Chem. Rev., 93 (1993) 1281.
27. Pixner, P., Heiden, W., Merx, H., Moeckel, G., Möller, A. and Brickmann, J., J. Chem. Inf. Comput. Sci., 34 (1994) 1309.
28. Heiden, W., Moeckel, G. and Brickmann, J., J. Comput.-Aided Mol. Design, 7 (1993) 503.
29. Gaillard, P., Carrupt, P.-A., Testa, B. and Boudon, A., J. Comput.-Aided Mol. Design, 8 (1994) 83.
30. Richards, N.G.J. and Williams, P.B., Chem. Design Automat. News, 9 (1994) 1.
31. Kim, K. H., J. Comput.-Aided Mol. Design, 9 (1995) 308.
32. Kellogg, G.E., Joshi, G.S. and Abraham, D.J., Med. Chem. Res., 1 (1992) 444.
33. Abraham, D. and Kellogg, G.E., J. Comput.-Aided Mol. Design, 8 (1994) 41.
34. Cruciani, G., Clementi, S., Fravolini, A., Cecchetti, V., Tabarrini, O. and Filipponi, E., XIII meeting of the Italian Pharm. Chemistry Society, Paestum, Italy, 1996, abstr. book P 126.
35. Mannhold, R. and Dross, K., Quant. Struct.-Act. Relat., 15 (1996) 403.