PERGAMON

Chemosphere 53 (2003) 1155–1164

CHEMOSPHERE

www.elsevier.com/locate/chemosphere

Predicting log *P* of pesticides using different software

E. Benfenati [a,*], G. Gini [b], N. Piclin [a,c], A. Roncaglioni [a], M.R. Varì [a]

[a] *Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri, Laboratory of Environmental Chemistry and Toxicology, via Eritrea 62, 20157 Milan, Italy*

[b] *Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milan, Italy*

[c] *Laboratory of Chemometrics and Bioinformatics, University of Orléans, B.P. 6759, 45067 Orléans Cedex 2, France*

Received 1 August 2002; received in revised form 22 May 2003; accepted 9 June 2003

Abstract

We compared experimental and calculated log *P* values using a data set of 235 pesticides and experimental values from four different sources: The Pesticide Manual, Hansch Manual, ANPA and KowWin databases. Log *P* were calculated with four softwares: HyperChem, Pallas, KowWin and TOPKAT. Crossed comparison of the experimental and calculated values proved useful, especially for pesticides. These are harder to study than simpler organic compounds. Structurally they are complex, heterogeneous and similar to drugs from a chemical point of view. They offer an interesting way to verify the goodness of the different methods.

Other studies compared several log *P* predictors using a single set of experimental values taken as a reference. Here we discuss the utility of the different log *P* predictors, with reference to experimental data found in *different* databases. This offers three advantages: (1) it avoids bias due to the assumption that one single data set is correct; (2) a given predictor can be developed on the same data set used for evaluation; (3) it takes account of experimental variability and can compare it with the predictor's variability. In our study Pallas and KowWin gave the best results for prediction, followed by TOPKAT.

Keywords: Log *P*; Octanol–water partition coefficient; Prediction software; Pesticides; Atom/fragment contributions; Database

1. Introduction

Lipophilicity is a very important molecular descriptor that often correlates well with the bioactivity of chemicals (Leo and Hansch, 1999; Cronin et al., 2002; Sverdrup et al., 2002). In studies of the environmental fate of organic chemicals, log *P*, the logarithm of the partition coefficient between *n*-octanol and water, correlates with water solubility (Ran et al., 2002), soil/sediment adsorption coefficients (Sabljic et al., 1995) and bioconcentration factors for aquatic organisms (Lyman et al., 1990).

Lipophilicity can therefore be measured by log *P*, which reflects the equilibrium partitioning a molecule between an apolar and a polar (aqueous) phase. Partition coefficients can be measured experimentally by several methods, ranging from the simple "shake flask" technique (Lodge, 1999; Edelbach and Lodge, 2000) to popular chromatographic methods (Eadsforth and Moser, 1983; Finizio et al., 1997; Edelbach and Lodge, 2000). However, experimental determination of partition coefficients is time- and material-consuming, it can only be done if the compound is available, and in many cases only an estimate of the lipophilicity of the chemicals is required anyway. Thus the literature provides

* Corresponding author. Tel.: +39-02-39014420; fax: +39-02-39001916.

E-mail address: benfenati@marionegri.it (E. Benfenati).

0045-6535/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0045-6535(03)00609-X

various figures for the same compound which can differ by more than one order of magnitude (Fielding et al., 1992). It is therefore much easier, faster and cheaper to predict $\log P$ using just the chemical structure.

Not only are there different methods for experimental measurements of $\log P$, but several approaches exist for its theoretical calculation too (Buchwald and Bodor, 1998). The most common ones are classified as "fragment constant" methods in which a structure is divided into fragments (atom or larger functional groups) and the values for each group are summed together (sometimes with structural correction factors) to yield the $\log P$ estimate. However, routine application of these approaches requires a continuous check of their validity by comparison with experimental data.

In the present study we assessed four softwares for predicting $\log P$. Similar studies have been done in the past, using a single list of values as a reference (Mannhold and Dross, 1996; Mannhold and van de Waterbeemd, 2001; Tetko et al., 2001). We believe that the source of data may be important parameter for this evaluation. If a single data set is used it must be assumed to be the "true" one. However, as we said, experimental determination of $\log P$ is complex and experimental values reported for the same compound differ widely. For this reason we evaluated the experimental data for the same compound in different databases. The aim was to avoid potential bias arising from considering one data set as true. We also avoided the risk of obtaining better results by using a predictor simply because it has been built on the basis of the same data set used for evaluation. Finally, we can take into account the variability of experimental data, and compare it with the predictor variability.

2. Methods

We created a database of 235 different pesticides (Benfenati et al., 1999) using experimental data from various sources. Most values were from The Pesticide Manual (218 compounds) (Tomlin, 1997) and KowWin database (207 compounds) (http://esc.syrres.com/interkow/kowdemo.htm). Other values were collected from a compilation of the Hansch (Hansch et al., 1995) (144 compounds) and ANPA databases (139 compounds) (Finizio, 1999). Two separate regression analyses were done. The first related the experimental $\log P$ values from different databases. In the second, experimental $\log P$ from different databases were related to the calculated/predicted $\log P$. Linear regression analyses were carried out using Excel. Calculated/predicted $\log P$ were obtained with four softwares, HyperChem (HyperChem 5.1, Hypercube Inc., Gainesville, Florida, USA), Pallas (Pallas 2.0, CompuDrug Chemistry Ltd), KowWin (http://esc.syrres.com/interkow/kowdemo.htm) and TOPKAT (TOPKAT 5.0, Health Design, Inc., Rochester, New York, USA), using either atom-based or different fragments approaches.

3. Results and discussion

We compared the availability of experimental data for the lipophilicity of our compounds in the four databases. At least two values were found only for 218 compounds. Regression analysis was applied for these correlations and the results are reported in Tables 1 and 2. Figs. 1 and 2 show the best and worst regres-

Table 1
R^2 of correlation among experimental $\log P$ from different sources

Square regression coefficient values R^2				
KowWin database	Hansch Manual	ANPA database	Average	
0.86 (196)[a]	0.86 (134)[a]	0.89 (137)[a]	0.87	The Pesticide Manual
	0.90 (139)[a]	0.79 (131)[a]	0.85	KowWin database
		0.85 (106)[a]	0.87	Hansch Manual
			0.84	ANPA database

[a] Number of compounds.

Table 2
Regression equations of experimental $\log P$ from different sources

Regression equations			
KowWin database	Hansch Manual	ANPA database	
$y = 0.8725x + 0.5254$	$y = 0.8439x + 0.5704$	$y = 0.9692x + 0.1006$	The Pesticide Manual
	$y = 0.9514x + 0.0876$	$y = 0.9812x - 0.0643$	KowWin database
		$y = 0.879x + 0.4121$	Hansch Manual

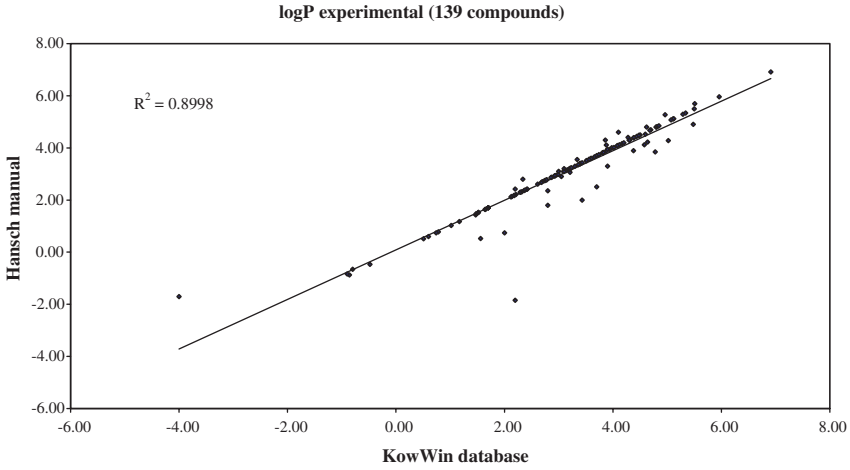

Fig. 1. Experimental log P values from the Hansch Manual and KowWin databases.

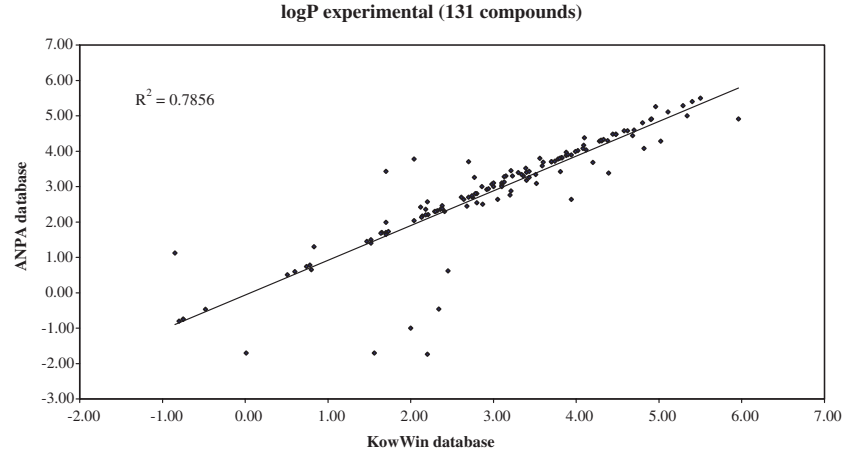

Fig. 2. Experimental log P values from the ANPA and KowWin databases.

sions. Some cases have a y intercept of more than 0.5 log-unit. For several compounds the same log P value was given in two different databases but this apparently high correlation could simply mean that the two databases refer to the same source. This may mean these sources are more reliable, but not necessarily. We therefore did a second regression analysis after deleting all identical values. Tables 3 and 4 shows these results, reporting the number of compounds considered. The determination coefficient values (R^2) are obviously lower in Table 3 than in Table 1, and the y intercept values are higher.

Table 3
R^2 of correlation among experimental log P from different sources (without the common log P values)

Square regression coefficient values R^2				
KowWin database	Hansch Manual	ANPA database	Average	
0.79 (104)[a]	0.84 (112)[a]	0.81 (83)[a]	0.82	The Pesticide Manual
	0.81 (38)[a]	0.64 (79)[a]	0.75	KowWin database
		0.79 (75)[a]	0.81	Hansch Manual
			0.75	ANPA database

[a] Number of compounds.

Table 4
Regression equations of experimental log P from different sources (without the common log P values)

Regression equations			
KowWin database	Hansch Manual	ANPA database	
$y = 0.8347x + 0.6905$	$y = 0.8333x + 0.5924$	$y = 0.9479x + 0.1577$	The Pesticide Manual
	$y = 0.8883x + 0.0816$	$y = 0.9576x - 0.0745$	KowWin database
		$y = 0.8231x + 0.5956$	Hansch Manual

The coefficient average is also listed in Tables 1 and 3. The column averages in Table 1 (all values) are similar for all databases. In Table 3, the averages for some databases are lower but this does not mean they are wrong. Table 5 sets out the mean and standard deviation calculated from experimental data for each pesticide. We used values from the four databases, when available, but always used at least two. The experimental measurements can be affected by several factors like chemical purity, experimental protocol, ionizable compounds, poor solubility in water, and typing errors. In fact, in Table 5 13 compounds have a standard deviation of one

Table 5
Mean and standard deviation (SD) of experimental log P for the 216 pesticides for which the four databases provided more than one value

CAS	Compound	n	Mean	SD	CAS	Compound	n	Mean	SD
15165-67-0	Dichlorprop-*P*[*]	3	2.20	2.12	1582-09-8	Trifluralin	4	5.13	0.26
74223-64-6	Metsulfuron-methyl[*]	4	−0.78	1.99	74738-17-3	Fenpiclonil	3	4.01	0.25
78-48-8	Tribufos[*]	2	4.47	1.75	16752-77-5	Methomyl	4	0.47	0.25
25057-89-0	Bentazone[*]	4	1.36	1.49	15972-60-8	Alachlor	4	3.30	0.25
64902-72-3	Chlorsulfuron[*]	4	0.19	1.46	17109-49-8	Edifenphos	2	3.66	0.25
79277-27-3	Thifensulfuron-methyl[*]	4	0.15	1.36	43121-43-3	Triadimefon	4	2.98	0.25
122931-48-0	Rimsulfuron[*]	2	−0.59	1.24	67747-09-5	Prochloraz	4	4.30	0.24
1071-83-6	Glyphosate[*]	3	−3.03	1.19	71283-80-2	Fenoxaprop-*p*-ethyl	4	4.47	0.23
1918-00-9	Dicamba[*]	4	1.62	1.18	62-73-7	Dichlorvos	4	1.56	0.23
30560-19-1	Acephate[*]	3	−0.21	1.15	467-69-6	Flurenol	2	1.16	0.23
81335-37-7	Imazaquin[*]	2	1.10	1.07	133-06-2	Captan	4	2.62	0.22
83055-99-6	Bensulfuron-methyl[*]	3	1.84	1.06	34123-59-6	Isoproturon	4	2.69	0.21
94593-91-6	Cinosulfuron[*]	3	2.62	1.00	330-55-2	Linuron	4	3.04	0.21
111991-09-4	Nicosulfuron	3	−0.68	0.90	834-12-8	Ametryn	4	2.92	0.19
16672-87-0	Ethephon	2	−1.58	0.88	7287-19-6	Prometryn	4	3.37	0.19
23103-98-2	Pirimicarb	4	2.13	0.87	24151-93-7	Piperophos	2	4.17	0.18
13071-79-9	Terbufos	4	4.05	0.85	10071-13-3	Maleic hydrazide	3	−0.77	0.18
55285-14-8	Carbosulfan	2	2.75	0.78	950-37-8	Methidathion	4	2.35	0.18
82558-50-7	Isoxaben	3	3.51	0.75	298-02-2	Phorate	4	3.71	0.18
59756-60-4	Fluridone	3	2.73	0.74	2310-17-0	Phosalone	4	4.27	0.18
1689-83-4	Ioxynil	4	3.07	0.72	919-86-8	Demeton-*S*-methyl	3	1.12	0.17
41483-43-6	Bupirimate	4	3.25	0.64	1897-45-6	Chlorothalonil	4	2.87	0.17
88283-41-4	Pyrifenox	3	3.20	0.62	21087-64-9	Metribuzin	4	1.75	0.17
2104-64-5	EPN	3	4.55	0.62	330-54-1	Diuron	4	2.67	0.16
3383-96-8	Temephos	4	5.44	0.61	51218-45-2	Metolachlor	4	3.11	0.16
14816-18-3	Phoxim	4	3.89	0.58	5915-41-3	Terbuthylazine	4	3.09	0.16
99-30-9	Dicloran	4	2.55	0.50	54-11-5	Nicotine	3	1.09	0.14
74115-24-5	Clofentezine	4	3.38	0.49	41198-08-7	Profenofos	4	4.56	0.14
42509-80-8	Isazofos	3	3.54	0.48	2642-71-9	Azinphos-ethyl	4	3.29	0.13
51235-04-2	Hexazinone	2	1.53	0.46	759-94-4	EPTC	4	3.27	0.12
17804-35-2	Benomyl	4	2.01	0.45	23950-58-5	Propyzamide	3	3.30	0.12

Table 5 (*continued*)

CAS	Compound	*n*	Mean	SD	CAS	Compound	*n*	Mean	SD
82-68-8	Quintozene	3	4.65	0.44	86-50-0	Azinphos-methyl	4	2.79	0.12
1563-66-2	Carbofuran	4	2.12	0.40	72490-01-8	Fenoxycarb	4	4.24	0.11
63-25-2	Carbaryl	4	2.17	0.39	120928-09-8	Fenazaquin	3	5.57	0.11
115-32-2	Dicofol	4	4.47	0.37	13360-45-7	Chlorbromuron	3	3.03	0.11
55-38-9	Fenthion	4	4.30	0.36	1134-23-2	Cycloate	4	3.96	0.11
66063-05-6	Pencycuron	4	4.60	0.35	51338-27-3	Diclofop-methyl	4	4.65	0.11
117-18-0	Tecnazene	2	4.14	0.35	24017-47-8	Triazophos	4	3.39	0.10
35400-43-2	Sulprofos	3	5.29	0.33	122-14-5	Fenitrothion	4	3.37	0.10
112281-77-3	Tetraconazole	2	3.33	0.33	1918-16-7	Propachlor	4	2.26	0.09
79983-71-4	Hexaconazole	4	3.75	0.30	84332-86-5	Chlozolinate	3	3.20	0.09
2921-88-2	Chlorpyrifos	4	5.05	0.27	23505-41-1	Pirimiphos-ethyl	3	4.90	0.09
41394-05-2	Metamitron	3	0.99	0.27	77732-09-3	Oxadixyl	3	0.75	0.09
333-41-5	Diazinon	4	3.59	0.26	732-11-6	Phosmet	4	2.83	0.08
29232-93-7	Pirimiphos-methyl	4	4.07	0.26	15545-48-9	Chlorotoluron	4	2.41	0.08
298-00-0	Parathion-methyl	4	2.93	0.08	56-38-2	Parathion-ethyl	4	3.83	0.01
2032-65-7	Methiocarb	4	2.96	0.08	1113-02-6	Omethoate	3	−0.75	0.01
84-65-1	Anthraquinone	3	3.43	0.08	301-12-2	Oxydemeton-methyl	3	−0.75	0.01
98730-04-2	Benoxacor	2	2.65	0.07	35367-38-5	Diflubenzuron	4	3.88	0.01
71626-11-4	Benalaxyl	4	3.44	0.07	76674-21-0	Flutriafol	4	2.30	0.00
10265-92-6	Methamidophos	4	−0.77	0.07	29091-05-2	Dinitramine	4	4.30	0.00
13684-56-5	Desmedipham	4	3.42	0.06	84496-56-0	Clomeprop	3	4.80	0.00
61-82-5	Amitrole	3	−0.90	0.06	2439-01-2	Chinomethionat	3	3.78	0.00
2136-79-0	Chlorthal-dimethyl	4	4.31	0.06	13457-18-6	Pyrazophos	3	3.80	0.00
10605-21-7	Carbendazim	4	1.49	0.06	73250-68-7	Mefenacet	3	3.23	0.00
36734-19-7	Iprodione	4	3.05	0.06	18691-97-9	Methabenzthia-zuron	3	2.64	0.00
50471-44-8	Vinclozolin	4	3.05	0.06	137-26-8	Thiram	3	1.73	0.00
65907-30-4	Furathiocarb	4	4.65	0.06	120923-37-7	Amidosulfuron	2	1.63	0.00
19666-30-9	Oxadiazon	4	4.83	0.05	33089-61-1	Amitraz	4	5.50	0.00
60168-88-9	Fenarimol	4	3.65	0.05	64249-01-0	Anilofos	3	3.81	0.00
57646-30-7	Furalaxyl	4	2.66	0.05	35575-96-3	Azamethiphos	2	1.05	0.00
1746-81-2	Monolinuron	4	2.28	0.05	25059-80-7	Benazolin-ethyl	2	2.50	0.00
57837-19-1	Metalaxyl	4	1.69	0.05	1861-40-1	Benfluralin	4	5.29	0.00
25311-71-1	Isofenphos	4	4.08	0.05	18181-80-1	Bromopropylate	3	5.40	0.00
122-34-9	Simazine	4	2.17	0.04	23184-66-9	Butachlor	2	4.50	0.00
886-50-0	Terbutryn	4	3.71	0.04	95465-99-9	Cadusafos	2	3.90	0.00
2303-17-5	Tri-allate	3	4.58	0.04	54593-83-8	Chlorethoxyfos	2	4.59	0.00
22224-92-6	Fenamiphos	4	3.27	0.04	122453-73-0	Chlorfenapyr	2	4.83	0.00
52-68-6	Trichlorfon	4	0.49	0.04	81777-89-1	Clomazone	2	2.50	0.00
60-51-5	Dimethoate	4	0.76	0.04	99607-70-2	Cloquintocet mexyl	2	5.03	0.00
3060-89-7	Metobromuron	4	2.41	0.04	56-72-4	Coumaphos	2	4.13	0.00
298-04-4	Disulfoton	4	4.00	0.04	57966-95-7	Cymoxanil	2	0.67	0.00
7085-19-0	Mecoprop	4	3.15	0.04	121552-61-2	Cyprodinil	2	4.00	0.00
33693-04-8	Terbumeton	4	3.07	0.03	66215-27-8	Cyromazine	2	−0.06	0.00
15299-99-7	Napropamide	4	3.33	0.03	50-29-3	DDT	2	6.91	0.00
5598-13-0	Chlorpyryfos-methyl	4	4.29	0.03	1014-69-3	Desmetryn	3	2.38	0.00
19937-59-8	Metoxuron	4	1.64	0.03	1085-98-9	Dichlofluanid	3	3.70	0.00
59669-26-0	Thiodicarb	3	1.68	0.03	83164-33-4	Diflufenican	3	4.90	0.00
731-27-1	Tolylfluanid	3	3.92	0.03	34205-21-5	Dimefuron	2	2.51	0.00
5234-68-4	Carboxin	4	2.16	0.03	50563-36-5	Dimethachlor	2	2.17	0.00
114-26-1	Propoxur	4	1.53	0.03	106325-08-0	Epoxiconazole	2	3.44	0.00
944-22-9	Fonofos	4	3.93	0.02	55283-68-6	Ethalfluralin	4	5.11	0.00

(*continued on next page*)

Table 5 (*continued*)

CAS	Compound	*n*	Mean	SD	CAS	Compound	*n*	Mean	SD
2164-17-2	Fluometuron	3	2.41	0.02	29973-13-5	Ethiofencarb	3	2.04	0.00
92-52-4	Biphenyl	2	4.00	0.02	563-12-2	Ethion	2	5.07	0.00
1194-65-6	Dichlobenil	4	2.73	0.02	26225-79-6	Ethofumesate	3	2.70	0.00
13593-03-8	Quinalphos	4	4.45	0.02	13194-48-4	Ethoprophos	4	3.59	0.00
23135-22-0	Oxamyl	4	−0.47	0.02	3740-92-9	Fenclorim	3	4.17	0.00
5259-88-1	Oxycarboxin	4	0.75	0.02	120068-37-3	Fipronil	2	4.00	0.00
87-86-5	Pentachlorophenol	3	5.11	0.01	63729-98-6	Flamprop-*M*-methyl	3	3.00	0.00
42576-02-3	Bifenox	4	4.49	0.01	79622-59-6	Fluazinam	2	3.56	0.00
22781-23-3	Bendiocarb	4	1.71	0.01	86811-58-7	Fluazuron	2	5.10	0.00
131341-86-1	Fludioxonil	2	4.12	0.00	51218-49-6	Pretilachlor	4	4.08	0.00
62924-70-3	Flumetralin	2	5.45	0.00	1610-18-0	Prometon	2	2.99	0.00
56425-91-3	Flurprimidol	2	3.34	0.00	139-40-2	Propazine	2	2.93	0.00
2540-82-1	Formothion	2	1.48	0.00	86763-47-5	Propisochlor	2	3.50	0.00
98886-44-3	Fosthiazate	2	1.68	0.00	52888-80-9	Prosulfocarb	2	4.65	0.00
87237-48-7	Haloxyfop etotyl	3	4.33	0.00	34643-46-4	Prothiofos	2	5.67	0.00
78587-05-0	Hexythiazox	2	2.53	0.00	84087-01-4	Quinclorac	2	−1.15	0.00
35554-44-0	Imazalil	4	3.82	0.00	3689-24-5	Sulfotep	3	3.99	0.00
86598-92-7	Imibenconazole	2	4.94	0.00	162320-67-4	SZI-121	2	3.30	0.00
138261-41-3	Imidacloprid	2	0.57	0.00	107534-96-3	Tebuconazole	4	3.70	0.00
2164-08-1	Lenacil	2	2.31	0.00	112410-23-8	Tebufenozide	2	4.25	0.00
55814-41-0	Mepronil	3	3.66	0.00	34014-18-1	Tebuthiuron	2	1.79	0.00
67129-08-2	Metazachlor	3	2.13	0.00	116-29-0	Tetradifon	2	4.61	0.00
139528-85-1	Metosulam	2	3.08	0.00	640-15-3	Thiometon	2	3.15	0.00
88671-89-0	Myclobutanil	4	2.94	0.00	112143-82-5	Triazamate	2	2.69	0.00
116714-46-6	Novaluron	2	5.27	0.00	101200-48-0	Tribenuron methyl	2	−0.44	0.00
58810-48-3	Ofurace	2	1.39	0.00	64628-44-0	Triflumuron	3	4.91	0.00
19044-88-3	Oryzalin	2	3.73	0.00					

*Compounds with a SD greater than one.

log *P* unit or more. These compounds show acid–basic properties, so the experimental measurements may not have been made under the right conditions to ensure neutrality. Moreover, our aim was not to establish the "right" log *P* for each compound, but to highlight the variability of experimental measurements even in very common, "official" sources. For several compounds, the values we found were identical in the different databases, because they all refer to the same source. In this case SD = 0.0. In summary, the quality of experimental data is not optimal, and several problems arise.

Tables 6 and 7 show the results of regression analysis for the experimental versus predicted values. The different programs to predict log *P* use calculation methods including atom-based and fragmental approaches. The structure of the molecule is broken down into fragments of one or two atoms and these are checked one by one. The software does this breakdown in various ways, producing all possible combinations in which the molecule can be assembled on the basis of the fragment database in use, and applying correction rules coupled with the molecular connectivity. In contrast, atom-based

Table 6
Square regression coefficient values R^2 of correlation among predicted and experimental log *P* and root mean squared error (RMS)

	The Pesticide Manual (218)[a]		KowWin database (207)[a]		Hansch Manual (144)[a]		ANPA database (139)[a]		Average R^2
	R^2	RMS	R^2	RMS	R^2	RMS	R^2	RMS	
HyperChem	0.50	1.32	0.63	0.99	0.57	1.06	0.50	1.24	0.55
KowWin software	0.66	1.05	0.90	0.52	0.79	0.74	0.65	1.00	0.75
Pallas	0.69	1.03	0.82	0.75	0.81	0.72	0.74	0.88	0.77
TOPKAT	0.64	1.03	0.84	0.63	0.78	0.72	0.56	1.08	0.71
Average R^2	0.62		0.80		0.74		0.61		

[a] Number of compounds.

Table 7
Regression equations between predicted and experimental log P values

	Regression equations			
	The Pesticide Manual	KowWin database	Hansch Manual	ANPA database
HyperChem	$y = 0.5995x + 1.6543$	$y = 0.7367x + 1.0858$	$y = 0.6942x + 1.1864$	$y = 0.5893x + 1.5514$
KowWin software	$y = 0.7947x + 0.8193$	$y = 1.0332x - 0.1496$	$y = 0.9346x + 0.2566$	$y = 0.7323x + 0.9239$
Pallas	$y = 0.8716x + 0.5124$	$y = 1.0512x - 0.215$	$y = 0.9563x + 0.17$	$y = 0.8747x + 0.4028$
TOPKAT	$y = 0.6675x + 1.0754$	$y = 0.8495x + 0.3535$	$y = 0.7687x + 0.674$	$y = 0.5813x + 1.2199$

procedures avoid correction factors and define huge numbers of atom-types; lipophilicity is quantified simply by summing atom-type values.

Both procedures show various inefficiencies, like oversimplification of steric and conformational effects of complex structures and the inability to estimate log P for uncorrelated or unknown fragments. Log P estimates based on atom values alone needed to be improved by inclusion of substructures layer or more complex than "atom". Moreover log P estimates based

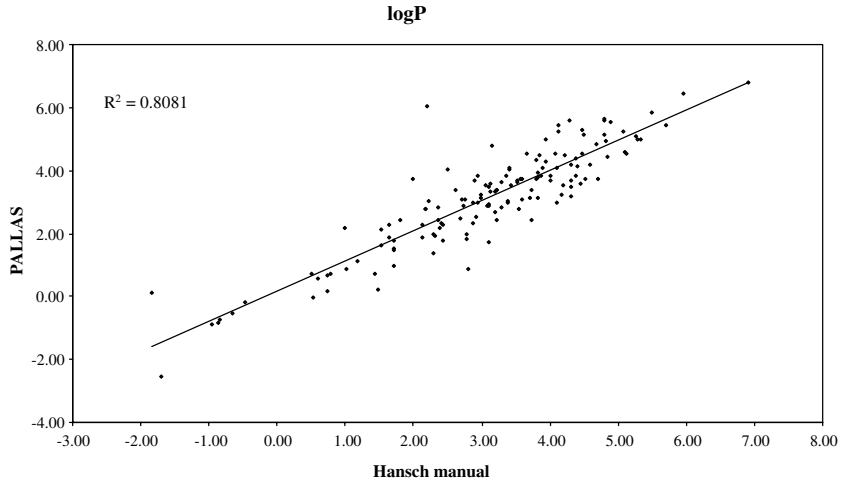

Fig. 3. Predicted (Pallas) and experimental (Hansch Manual) log P values.

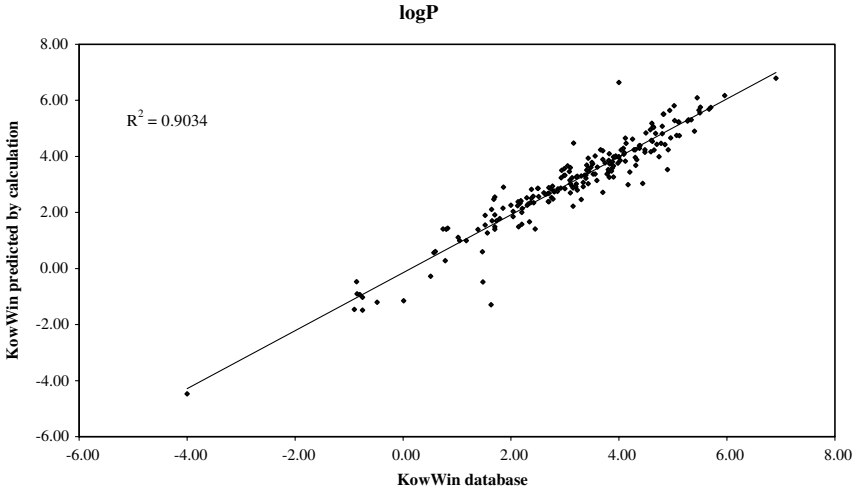

Fig. 4. Predicted (KowWin) and experimental (KowWin database) log P values.

on atom/fragment values, can usually be improved by inclusion of correction factors able to take into account steric interaction, ionization form, hydrogen bondings, effects from polar functional substructures, and so on.

The KowWin and TOPKAT programs calculate log *P* values based on different atom/fragment contribution approaches (Meylan and Howard, 1995; Gombar and Enslein, 1996). TOPKAT also provides a warning about the reliability of the prediction but this was reported for only 17 compounds in our data set. HyperChem calculates the log *P* from atomic contributions (Ghose et al., 1988). Pallas, similar to KowWin and TOPKAT, calculates log *P* values based on atomic and fragmental contributions but uses a weight for the

Table 8
Overestimated, underestimated and well-predicted log *P* values

	HyperChem			Pallas			KowWin			TOPKAT		
	↑[a]	↓[b]	=[c]	↑	↓	=	↑	↓	=	↑	↓	=
The Pesticide Manual (218)[d]	58	32	10	45	38	17	42	33	25	36	41	23
KowWin database (207)[d]	50	36	14	37	40	23	30	34	36	25	46	29
Hansch Manual (144)[d]	53.5	38.2	8.3	37.5	39.6	22.9	33.3	35.4	31.3	27	44	29
ANPA database (139)[d]	54	33	13	42	37	21	39	34	27	31.7	41.7	26.6

[a] Percentage of overestimated predicted values.
[b] Percentage of underestimated predicted values.
[c] Percentage of similar predicted values (within 0.15 log-unit).
[d] Number of compounds.

Table 9
R^2 values of correlation between predicted and experimental log *P* values for anilines, carbamates, ureas, organophosphorus, heterocyclics and halogenated aromatics compounds

	Square regression coefficient values R^2								Chemical class
	The Pesticide Manual		KowWin database		Hansch Manual		ANPA database		
HyperChem	0.64	(36)[a]	0.68	(35)[a]	0.71	(22)[a]	0.66	(20)[a]	Anilines (39)[a]
KowWin software	0.86		0.88		0.88		0.88		
Pallas	0.72		0.74		0.78		0.86		
TOPKAT	0.71		0.75		0.81		0.86		
HyperChem	0.66	(23)[a]	0.69	(24)[a]	0.54	(20)[a]	0.65	(23)[a]	Carbamates
KowWin software	0.90		0.93		0.73		0.73		(26)[a]
Pallas	0.85		0.90		0.55		0.61		
TOPKAT	0.78		0.81		0.55		0.60		
HyperChem	0.42	(31)[a]	0.54	(26)[a]	0.50	(15)[a]	0.42	(22)[a]	Ureas (31)[a]
KowWin software	0.39		0.84		0.57		0.47		
Pallas	0.66		0.73		0.82		0.74		
TOPKAT	0.54		0.81		0.67		0.40		
HyperChem	0.67	(54)[a]	0.80	(53)[a]	0.82	(39)[a]	0.72	(34)[a]	Organophos-
KowWin software	0.88		0.96		0.91		0.91		phorus (59)[a]
Pallas	0.87		0.93		0.93		0.87		
TOPKAT	0.82		0.91		0.92		0.91		
HyperChem	0.35	(113)[a]	0.45	(100)[a]	0.44	(60)[a]	0.34	(67)[a]	Heterocyclics
KowWin software	0.51		0.80		0.63		0.51		(119)[a]
Pallas	0.58		0.73		0.69		0.63		
TOPKAT	0.50		0.77		0.64		0.33		
HyperChem	0.30	(78)[a]	0.37	(74)[a]	0.45	(50)[a]	0.35	(46)[a]	Halogenated
KowWin software	0.49		0.80		0.67		0.42		aromatics (83)[a]
Pallas	0.46		0.74		0.74		0.57		
TOPKAT	0.51		0.81		0.73		0.50		

[a] Number of compounds.

contributions of atom-based and fragmental parameters (Ghose and Crippen, 1986; Rekker, 1977). The weight used was: $\log P_{comb} = 0.733 \log P_{atomics} + 0.267 \log P_{cdr}$, where P_{comb} stands for combined value (atomic and fragment based), $P_{atomics}$ uses atomic fragments and P_{cdr} is based on Rekker's collection of hydrophobic fragmental constant (Rekker and de Kort, 1979).

Figs. 3 and 4 show two examples of correlations between experimental and predicted data. Table 6 shows R^2 and root mean squared error (RMS). The RMS values are quite high. Table 6 indicates a superiority of the fragmental over atom-based methods (HyperChem), in agreement with a previous study (Mannhold and Dross, 1996). Further inspection was done on the basis of the $\log P$ values. The fragmental method reportedly tends to overestimate $\log P$ and atom-based approaches underestimate it. This underestimation was significantly more pronounced for $\log P > 4$ (Mannhold and Dross, 1996).

Our results are shown in Table 8. All software gave overestimates and underestimates. HyperChem, an atom-based method, overestimated the $\log P$ values for each database, whatever the value, whereas TOPKAT, an atom/fragment method, tended to underestimate them. For KowWin and Pallas, the situation was variable.

Finally, to extract more information from our data, we divided the 235 pesticides into chemical classes: anilines, carbamates, ureas, organophosphorus, heterocyclics and halogenated aromatics. The similarities between several compounds allows some comparative considerations on how the different calculations process these structures (Table 9). The best correlations (R^2) for anilines and carbamates were given with the KowWin software. Experimental data for ureas and heterocyclics, for three of the four databases, were closely correlated with predicted values using Pallas software. For organophosphorus, the experimental data of three of the databases highly correlated with the values predicted by the KowWin software. For halogenated aromatics, the situation was unclear.

In conclusion, reliable $\log P$ values for pesticides are still limited and experimental values show a certain degree of variability. It is therefore questionable to rely on a single value. Predictors can be an alternative but, as shown in Table 6, some are better then others. In any case, it is necessary to use values predicted with the same software to have consistent results, because of the predictor variability.

Moreover, the different performances of the four softwares depend on the chemical class. Anilines, carbamates and organophosphorus seem to be predicted better using KowWin, whereas ureas and heterocyclics can be predicted better using Pallas. So the quality of results tends to differ depending on the software used and chemical classes of the compounds investigated.

Acknowledgements

We acknowledge financial support from The European Commission: project IMAGETOX, HPRN-CT-1999-00015. NP was recipient of a fellowship from IMAGETOX.

References

Benfenati, E., Pelagatti, S., Grasso, P., Gini, G., 1999. COMET: the approach of a project in evaluating toxicity. In: Gini, G.C., Katritzky, A.R. (Eds.), Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools. AAAI 1999 Spring Symposium Series. AAAI Press, Menlo Park, CA, pp. 40–43.

Buchwald, P., Bodor, N., 1998. Octanol–water partition: searching for predictive models. Curr. Med. Chem. 5, 353–380.

Cronin, M.T.D., Dearden, J.C., Duffy, J.C., Edwards, R., Manga, N., Worth, A.P., Worgan, A.D.P., 2002. The importance of hydrophobicity and electrophilicity descriptors in mechanistically-based QSARs for toxicological endpoints. SAR QSAR Environ. Res. 13, 167–176.

Eadsforth, C.V., Moser, P., 1983. Assessment of reverse-phase chromatographic methods for determining partition coefficients. Chemosphere 12, 1459–1475.

Edelbach, D.J., Lodge, K.B., 2000. Insights into the measurement of the octanol–water partition coefficient from experiments with acrylate esters. Phys. Chem. Chem. Phys. 2, 1763–1771.

Fielding, M., Barcelò, D., Helweg, A., Galassi, S., Torstensson, L., Van Zoonen, P., Wolter, R., Angeletti, G., 1992. Pesticides in ground and drinking water (Water pollution research reports, No. 27), Commission of the European Communities, Brussels.

Finizio, A., 1999. Valutazione del rischio per gli organismi non bersaglio: ANPA DATABASE: L' IMPATTO AMBIENTALE DEI PRODOTTI FITOSANITARI.

Finizio, A., Vighi, M., Sandroni, D., 1997. Determination of *N*-octanol/water partition coefficient (k_{ow}) of pesticide critical review and comparison of methods. Chemosphere 34, 131–161.

Ghose, A.K., Crippen, G.M., 1986. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure–activity relationships. I. Partition coefficients as a measure of hydrophobicity. J. Computat. Chem. 7, 565–577.

Ghose, A.K., Prichett, A., Crippen, G.M., 1988. Atomic physicochemical parameters for three dimensional structures directed quantitative structure–activity relationships. III. Modeling hydrophobic interactions. J. Comput. Chem. 9, 80–90.

Gombar, V.K., Enslein, K., 1996. Assessment of *n*-octanol/water partition coefficient: when is the assessment reliable? J. Chem. Inf. Comput. Sci. 36, 1127–1134.

Hansch, C., Leo, A., Hoekman, D., 1995. Exploring QSAR Hydrophobic, Electronic and Steric Constants. ACS, Washington, DC.

Leo, A., Hansch, C., 1999. Role of hydrophobic effects in mechanistic QSAR. Perspect. Drug Discov. 17, 1–25.

Lodge, K.B., 1999. Octanol–water partition coefficients of cyclic C-7 hydrocarbons and selected derivatives. J. Chem. Eng. Data 44, 1321–1324.

Lyman, W.J., Reehl, W.F., Rosenblatt, D.H., 1990. Handbook of Chemical Property Estimation Methods. ACS, Washington, DC.

Mannhold, R., Dross, K., 1996. Calculation procedures for molecular lipophilicity: A comparative study. Quant. Struct. –Act. Relat. 15, 403–409.

Mannhold, R., van de Waterbeemd, H., 2001. Substructure and whole molecule approaches for calculating log *P*. J. Comp. Aided-Mol. Des. 15, 337–354.

Meylan, W.M., Howard, P.H., 1995. Atom/fragment contribution method for estimating octanol–water partition coefficients. J. Pharm. Sci. 84, 83–92.

Ran, Y., He, Y., Yang, G., Johnson, J.L.H., Yalkowsky, S.H., 2002. Estimation of aqueous solubility of organic compounds by using the general solubility equation. Chemosphere 48, 487–509.

Rekker, R.F., 1977. The hydrophobic fragmental constant. In: Pharmacochemistry Library. Elsevier Amsterdam.

Rekker, R.F., de Kort, H.M., 1979. The hydrophobic fragmental constant; an extension to a 1000 data point set. Eur. J. Med. Chem. 14, 479–488.

Sabljic, A., Güsten, H., Verhaar, H., Hermens, J., 1995. Qsar modelling of soil sorption. Improvements and systematics of $\log K_{OC}$ vs. $\log K_{OW}$ correlations. Chemosphere 31, 4489–4514.

Sverdrup, L.E., Nielsen, T., Krogh, P.H., 2002. Soil ecotoxicity of polycyclic aromatic hydrocarbons in relation to soil sorption, lipophilicity and water solubility. Environ. Sci. Tecnol. 36, 2429–2435.

Tetko, I.V., Tanchuk, V.Y., Villa, A.E., 2001. Prediction of *n*-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. J. Chem. Inf. Comput. Sci. 41, 1407–1421.

Tomlin, C., 1997. The Pesticide Manual, eleventh ed. British Crop Protection Council, Farnham, UK.