Perspectives in Drug Discovery and Design, **18:** 1–18, 2000.
KLUWER/ESCOM

The hydrophobic fragmental constant approach for calculating log P in octanol/water and aliphatic hydrocarbon/water systems

RAIMUND MANNHOLD[a,*] and ROELOF F. REKKER[b]
[a] *Department of Laser Medicine, Molecular Drug Research Group, Heinrich-Heine-Universität, Universitätsstrasse 1, D-40225 Düsseldorf, Germany*
[b] *Leiden/Amsterdam Center for Drug Research, Department of Pharmacochemistry, Faculty of Chemistry, Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands*

Received 17 June 1999; Accepted 22 June 1999

Summary. Lipophilicity is a prime physicochemical parameter in describing both pharmacodynamic and pharmacokinetic aspects of drug action. Its quantitative descriptor (log P) is measured in or calculated for different model solvent systems depending on the respective scientific interest. The first part of this article focuses on the hydrophobic fragmental constant approach (Σf system) for octanol/water partitioning. The original approach and its revisions are presented, followed by recipes for the practical procedure of Σf calculations and a detailed description of how to apply the correction factor C_M. A Σf system for application in the study of aliphatic hydrocarbon/water (ahc/w) partitioning is described in the second part of the article. The current version is based on an optimized coupling of the ahc/w system with f constants for octanol/water partitioning. A total of 70 f_{ahc} values are now available.

Key words: aliphatic hydrocarbon/water partitioning, calculation procedures, fragmental methods, lipophilicity, octanol/water partitioning

Introduction

No other physicochemical property has attracted as much interest in QSAR studies as lipophilicity. Its quantitative descriptor, the partition coefficient P, expresses the ratio of monomeric, neutral solute concentrations in the organic and aqueous phase of a two-component system under equilibrium conditions. Experimental determination of log P [1,2] is tedious, time-consuming and demands a high purity of the solute. The disadvantages in experimental log P provoked an intensive search for alternative lipophilicity descriptors [2].

* To whom correspondence should be addressed.
E-mail: Raimund.Mannhold@uni-duesseldorf.de

Here, we describe the hydrophobic fragmental constant approach (Σf system) for octanol/water partitioning [3–8] and a Σf system for application in the study of aliphatic hydrocarbon/water (ahc/w) partitioning [9].

Fragmental constant approach for calculating log P_{oct}

Development of the system

A data set of more than 1000 experimental log P_{oct} values of simple organic compounds was used to derive fragmental values by regression analysis; hence, this approach has been labelled 'reductionistic'. This system is based on the relation

$$\log P = \sum f = \sum_{i=1}^{n} a_i \cdot f_i + \sum_{i=1}^{m} k_i \cdot C_M \quad (1)$$

where f is the fragmental constant, a the incidence of a fragment in a molecule, C_M the correction factor, and k_i the frequency of C_M.

The development of the Σf system comprised three main phases.

First phase

Reliable experimental log P for about 100 simple organic structures (mainly selected from literature sources) served as an initial data set to derive fragmental constants by means of Free-Wilson-type regression analyses, continuously fine-tuned by a stepwise enlargement of the data sets. The first period resulted in a valuable system for log P calculation based on 126 fragment values. Fragmentation is performed in such a way that functional groups with recognizable direct resonance interaction are left intact. Fragments range from atoms over substituents to complicated, in particular, heterocyclic ring structures; fragments are differentiated according to aliphatic or aromatic attachment.

An important outcome of the regression analyses was the detection of systematic differences between experimental log P and log P calculations based on the summation of fragment values. Differences between measurement and calculation could be attributed to chemical characteristics of the molecules, which in turn allowed the definition of correction rules for log P calculation. Among the prime rules to be detected was the so-called proximity effect, which describes the presence of electronegative centres in a molecule separated by one or two carbons. Later on correction rules were extended to other chemical features, such as aromatic condensation, cross-conjugation or hydrogen bonding. A closer inspection of the correction values revealed that

they represent multiples of a constant value of 0.289, known as the 'magic constant' (C_M); it proves to be of great importance in restoring imbalances between experimental log P and calculation done by merely adding fragmental values. This approach is known in the literature as the *original* Σf *system* [3–6].

Second phase

Some intriguing points initiated a thorough revision of the original Σf system: (a) the bad fit of aliphatic hydrocarbon log P_{oct} with Σf values; (b) the irregular fit of log P_{oct} for simple halo-alkanes with calculated data; and (c) the correction factor of –0.46 for structures with electronegativity facing alkyl bulk and the impossibility of connecting this correction with multiples of C_M. To treat these problems, the further revision of the Σf system was focused on a correct tracing of C_M. A set of 15 structure pairs (C_6H_5-CH_2-X versus C_6H_5-X), with X representing an electronegative fragment, revealed a fairly constant difference between f X_{ar} and f X_{al} of 0.87 (±0.06), close to 3 times the original C_M of 0.289. Regression analyses revealed, however, that a C_M value of 0.87/4 was adequate for renewing the f system and the value of C_M was revised to 0.219. Details of the *revised* Σf *system* are given in Reference 7.

Third phase

The application of the rules postulated for Σf corrections as proposed in Reference 7 raised some problems in the calculation of halogenated structures. To tackle the complex pattern of halogen presence in aliphatic hydrocarbon structures, we assembled a series of log P_{oct} data from the available literature and elaborated the following updating of the system [8]: up-corrections of 0.219 were undertaken for Cl, Br and I, leaving F unchanged. The presence of two halogens on the same C requires an extra C_M to be added in the calculation. The presence of three halogens on the same C demands for an up-correction with four extra C_M. In contrast, satisfactory rules for C_M application are currently not available for per-halogenated compounds.

A complete tabulation of the current version of revised f values, comprising 169 fragments, is given in Table 1. It includes 14 new heterocyclic fragments as well as two- and threefold halogenated methyls.

Practical procedure of Σf *calculations*

We recommend starting with the gross formula and denoting the functional groups fg1, fg2, . . ., fgn together with their f values (Table 1). The atomic composition is subtracted from the gross formula leaving C_x H_y as the residue. The lipophilicity contribution of C_x H_y is obtained from x f (C) + y f (H). An

Table 1. Fragmental constants in the octanol/water (f_{oct}) and in the aliphatic hydrocarbon/water system (f_{ahc})

Fragment	f_{oct}	f_{ahc}	Fragment	f_{oct}	f_{ahc}
With C, with H					
C_6H_5	1.903	1.987 (1)	**ar** CONH	–1.559	–4.689 (11)
C_6H_4 phenyl-	1.698	1.745 (1)	**ar** NHCO•	–1.559	–4.689 (11)
C_6H_3 type	1.494	1.504 (1)	$OOCNH_2$	–1.405	–3.730 (8)
C_6H_2 fragments	1.289	1.263 (1)	**ar** $OOCNH_2$	–0.967	–2.954 (7)
C_6H	1.085	1.021 (1)	OOCNH	–1.829	–4.230 (8)
CH_3	0.724	0.854 (0)	**ar** OOCNH	–1.391	–3.454 (7)
CH_2	0.519	0.613 (0)	**ar** NHCOO	–0.734	
CH	0.315	0.371 (0)	CONHNH	–3.348	
CH_2=CH	0.834	0.984 (0)	**ar** CONHNH	–2.253	
CH≡C	0.425	0.501 (0)	$NHCONH_2$	–1.860	
Naphthalenyl	3.191	3.248 (2)	**ar** $NHCONH_2$	–0.984	
OCH_3	–0.821	–0.969 (0)	NHCONH	–2.284	
ar OCH_3	0.274	0.323 (0)	**ar** NHCONH	–1.408	–5.287 (14)
COOH	–0.942	–3.961 (11)	$NCONH_2$	–2.708	
ar COOH	–0.066	–2.668 (10)	NCONH	–3.132	
CO•H	–0.991	–1.946 (3)	**ar** CH=CH-NO_2	0.153	–0.337 (2)
ar CO•H	–0.334	–0.912 (2)	**ar** CH=CH-CONH	–1.367	–3.944 (2)
O–CH_2-COOH	–1.044		$CONHCONH_2$	–1.602	
ar O-CH_2-COOH	–0.606		$NHNHCONH_2$	–2.850	
ar CH=CH-COO	–0.132	–0.156 (0)	CH=N-NOH	–0.798[a]	
ar CH=CH-CO•H	–0.141	–0.684 (2)	**ar** CH=N-NOH	–0.141[a]	
ar CH=CH-CO•	–0.565	–1.185 (2)	SCH_3	0.166	
Benzoquinonyl	–0.020	–0.542 (2)	**ar** SCH_3	0.823	
Naphthoquinonyl	1.486	0.976 (3)	$NHCSNH_2$	–1.409	
Anthraquinonyl	3.211	2.753 (4)	**ar** $NHCSNH_2$	–1.190	
ar C=NH	–1.500		NHCSNH	–1.833	
NH-C(NH_2)=N-C≡N	–1.573		**ar** NHCSNH	–1.614	
ar NH-C(NH_2)=N-C≡N	–0.916[a]		$NCSNH_2$	–2.257	
$CONH_2$	–2.011	–5.481 (12)	NCSNH	–2.681	
ar $CONH_2$	–1.135	–4.188 (11)	**ar** $NHSO_2CF_3$	1.254	
CONH	–2.435	–5.981 (12)			

Table 1. (continued)

Fragment	f_{oct}	f_{ahc}	Fragment	f_{oct}	f_{ahc}
With C, no H					
C	0.1102	0.1300 (0)	**ar** C≡N	–0.155	–0.960 (3)
C_6 (phenyl-skeleton)	0.880	0.780 (1)	**ar** C=N	–1.930	
CBr_3	2.417		COO	–1.200	–1.934 (2)
CCl_3	1.814		**ar** COO	–0.543	–0.641 (0)
CF_3	0.566		**ar** OOC	–0.981	–1.676 (2)
ar CF_3	1.223		COO^-	–4.967	
CI_2	1.907		**ar** COO^-	–4.091	
CBr_2	1.283		CO•	–1.633	–2.704 (3)
CCl_2	0.881		**ar** CO•	–0.976	–1.670 (2)
CF_2	–0.097		CON	–2.859	–6.482 (12)
CBrCl	1.082		**ar** CON	–1.983	–5.189 (11)
CBrF	0.812		**ar** NCO•	–1.544	–4.930 (11)
CClF	0.611		NCS•	0.471	
CCl F_2	0.836		**ar** NCS•	1.347	
CCl_2F	1.325		SCN	–0.405	
C≡N	–1.031	–1.994 (3)			
No C, with H					
H	0.2045	0.2413 (0)	**ar** NH	–0.938	–1.625 (2)
H (neg)	0.424	0.500 (0)	SH	–0.046	–0.054 (0)
OH	–1.448	–3.522 (7)	**ar** SH	0.611	
ar OH	–0.353	–2.748 (9)	**ar** SO_2NH_2	–1.440	
NH_2	–1.340	–3.135 (6)	**ar** SO_2NH	–1.864	
ar NH_2	–0.902	–2.100 (4)	**ar** $NHSO_2$	–1.645	
NH	–1.814	–2.918 (3)			
No C, no H					
Br	0.477		NNO	–2.063	–3.211 (3)
ar Br	1.134	1.079 (1)	O	–1.545	–2.082 (1)
Cl	0.276		**ar** O	–0.450	–0.531 (0)
ar Cl	0.933	0.842 (1)	S	–0.558	
F	–0.213		**ar** S	0.099	
ar F	0.444	0.265 (1)	S-S	0.320	
I	0.789		SO•	–2.79	
ar I	1.446	1.447 (1)	**ar** SO•	–2.13	
N	–2.074	–2.965 (2)	SO_2	–2.83	
ar N	–0.979	–1.414 (1)	**ar** SO_2	–2.07	
NO_2	–0.915	–1.598 (2)	**ar** SO_2N	–2.288	
ar NO_2	–0.039	–0.564 (2)			

Table 1. (continued)

Fragment	f_{oct}	f_{ahc}	Fragment	f_{oct}	f_{ahc}
Heterocycles					
Imidazolyl	–0.046		Cinnolinyl	0.605[a]	
Pyrrolyl	0.615		Pyrido (2,3) pyrazinyl	–0.251[a]	
Pyridinyl	0.534	–0.665 (5)	Phenazinyl	2.550[a]	
1,2,4-Triazolyl	–0.937[a]		Furyl	1.086	
1,2,3-Triazolyl	–0.499[a]		Benzofuryl	2.374	
Tetrazolyl	–0.917[a]		Dibenzofuryl	3.839[a]	
Benzimidazolyl	1.241		Thienyl	1.613	
Uracilyl	–1.297		Benzothienyl	2.901	
Barbituryl	–1.500		Dibenzothienyl	4.388[a]	
Indolyl	1.902		Oxazolyl	–0.250[a]	
Carbazolyl	3.570[a]		Benzoxazolyl	1.257[a]	
Quinolinyl	1.821[a]	0.857 (5)	Isoxazolyl	–0.250[a]	
Isoquinolinyl	1.821[a]	0.857 (5)	Benzisoxazolyl	1.257[a]	
Acridinyl	3.110[a]		Thiazolyl	0.300[a]	
Benzotriazolyl	1.227[a]		Benzthiazolyl	1.807[a]	
Pyrimidinyl	–0.683[a]		Benzoxdiazolyl	1.557[a]	
Pyrazinyl	–0.464[a]		Phenothiazinyl	3.665	
Pyridazinyl	–0.902[a]		Phenylaminophenyl	3.319	
Quinazolinyl	0.824[a]		Phenyloxyphenyl	4.026	
Quinoxalinyl	1.043[a]		Phenylthiophenyl	4.190	
Phthalazinyl	0.386[a]				
C_M	0.219	0.259	C_M	0.219	0.259

ar: Aromatic fragment; the left-hand atom indicates the site of its attachment on the aromatic structure. O● and S●: double-bonded O and S, respectively. The tabulated f_{oct} values for CH fragments were obtained from the four-decimal f C_{oct} and f H_{oct} values. The tabulated f_{ahc} values were calculated by $f_{ahc} = 1.18\ f_{oct} - 0.259\ k_n$. Numbers in parentheses after f_{ahc} are the k_n's applied for transfer of the octanol–water into the aliphatic hydrocarbon–water system.

[a] Fragment values, calculated according to Reference 5.

evaluation of correction factors ($n \cdot C_M$) completes the calculation; details are given in Reference 5. We recommend carrying out the calculations in three decimals, with the final result rounded to two decimals.

Successful Σf calculations require a correct application of correction factors (Table 2). The numbering of the correction rules as used in Table 2 appears here in parentheses:

Table 2. Correction factors for application in Σf calculations

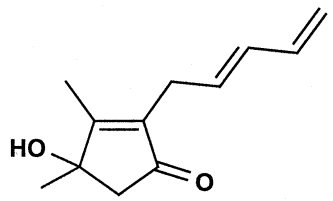
No.	Correction factor	Multiples of C_M
1	Saturated aliphatic hydrocarbon chains	
1.1	General	+2
1.2	Exception: methane	+1
2	Saturated aliphatic hydrocarbon rings	
2.1	General	+2
2.2	Exception: cyclopropane	+1
3	Unsaturation	
3.1	Double bonds	–
3.2	Triple bonds	–1
4	Extended chain conjugation	+2
5	Aromatic hydrocarbons	
5.1	Benzene	+1
5.2	Condensation in aromatics	+1
5.3	Aryl–aryl conjugation	+1
5.4	Cross-conjugation	+1
6	Proximity effects	
6.1	1C separation	+3
6.2	2C separation	+2
7	H attached to electronegative groups	+1
8	Electronegativity facing alkyl bulk	
8.1	Bulk involving quaternary carbon	–2
8.2	Bulk involving tertiary carbon	–1
9	Oxygen bound to aromatics via 1 carbon	+1
10	Hydrogen bonding	+3
11	Decoupling of resonance interaction	–1 to –5
12	Resonance interaction	+2

(1) Saturated aliphatic hydrocarbon chains
Saturated hydrocarbons in general (1.1) need the application of 2 C_M. Methane represents the only exception (1.2) and needs a correction with 1 C_M.

(2) Saturated aliphatic hydrocarbon rings
Saturated aliphatic hydrocarbon rings need 2 C_M for correction (2.1); cyclopropane (2.2) represents an exception with 1 C_M.

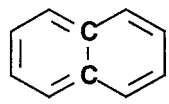
(3) Unsaturation
A double bond needs no correction (3.1), while a triple bond requires one negative C_M (3.2).

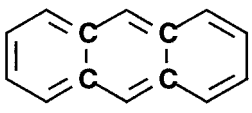
(4) Extended chain conjugation
In cases of extended chain conjugation, 2 C_M should be added to the normal f summation; C=C-C=O, as present in *pyrethrolone*, is also an example of this type.



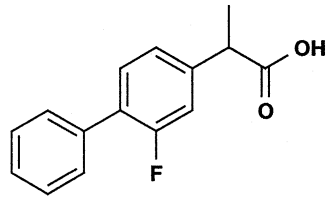
(5) Aromatic hydrocarbons
(5.1) Benzene requires 1 C_M.

(5.2) Condensation in aromatics calls for 1 C_M per condensation site, indicated by 2 Cs in naphthalene and correspondingly 4 Cs in anthracene.

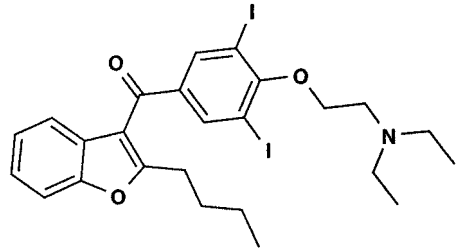




(5.3) Aryl-aryl conjugation as exemplified by flurbiprofen requires 1 C_M.



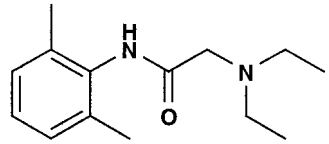
(5.4) Cross-conjugation calls for 1 C_M as shown here for amiodarone.



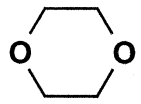
(6) Proximity effects

Proximity effects, although rather simple in the conceptual approach, appear not so easy to handle in an unambiguous way. 1C separations X-CH_2-X (6.1), with X representing an electronegative group, obtain an upward correction of 3 C_M. With 2C separations X-CH_2-CH_2-X (6.2) the correction amounts to 2 C_M.

(6.1) 1C separations X-CH_2-X like lidocaine.

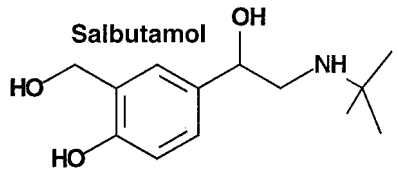


(6.2) 2C separations X-CH_2-CH_2-X like dioxane.

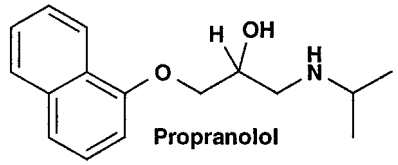


(7) H attached to electronegative groups
In structures like **H**COOH, the normal f_H value should be replaced by 0.2045 + 0.219 = 0.424, i.e. the direct connection of these **H** atoms gains in lipophilicity by the electronegative character of the rest of the molecule.

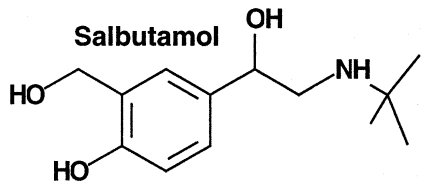
(8) Electronegativity facing alkyl bulk
Electronegativity facing alkyl bulk necessitates a negative correction of $2k_n$ or $1k_n$, depending on whether the bulk arises from a quaternary or a tertiary C centre. This rule does not hold for functional groups like -COO-; the oxygen is far enough from the alkyl bulk to avoid this effect.
(8.1) Electronegativity facing quaternary C



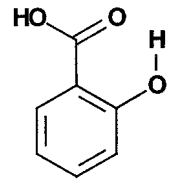
(8.2) Electronegativity facing tertiary C



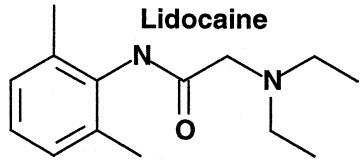
(9) Oxygen connected to aromatics
Oxygen connected to aromatics via 1 C atom is an example of an orbital overlap between the O atom and the π-electronic system of the phenyl ring.



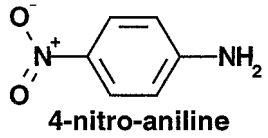
(10) Hydrogen bonding
Intramolecular hydrogen bonding increases lipophilicity with 3 C_M as exemplified here for salicylic acid.



(11) Decoupling of resonance
Neutral moieties (e.g. alkyl) in ortho-position to another substituent that is able to undergo resonance interaction (both attached to an aromatic ring) may perform a decoupling of resonance with regard to the aromatic system. This will convert the lipophilicity contribution of the aromatic substituent to a more aliphatic value. The difference between aliphatic and aromatic f values is connected with multiples of C_M depending on the resonance power of the substituent: OCH_3: –5; COOH, $CONH_2$: –4; C=O, CONH, NHCO: –3; NH_2: –2.



(12) Resonance interaction
The combination of two groups like nitro, carboxyl or carbonamide on a phenyl ring in para or meta position gives rise to a resonance interaction resulting in increased log P (1 to 3 C_M). Subrules have not been developed so far; for a practical approach we propose an averaged correction of 2 C_M.



The fragmental constant approach for calculating log P_{ahc}

An inspection of partition data [10] reveals that most log P_{ahc} values are independent of chain length, cyclization or branching of the hydrocarbon partition partner. In particular, for phenol 32 partition data are found in the

Table 3. Overview of the data selection for aliphatic alcohols

Alcohol	a	b	c	Δ	Alcohol	a	b	c	Δ
Methanol	3	3	3	0.02	1-Pentanol	7	1	6	0.08
Ethanol	6	2	4	0.08	1-Hexanol	3	0	3	–0.05
1-Propanol	7	1	5	0.05	1-Heptanol	4	1	4	0.03
1-Butanol	9	2	7	0.00					

a = Number of data included in the final correlation.
b = Number of rejected data points.
c = Number of ahc/water systems considered.
$\Delta = \Sigma f_{ahc} - \log P_{ahc}$ (average); these values stem from the final correlation study comprising 365 data.

literature, measured in 10 ahc/water systems: pentane (1), hexane (7), cyclohexane (12), heptane (5), octane (2), iso-octane (1), nonane (1), decane (1), dodecane (1) and decahydronaphthalene (1). After the omission of six evidently outlying data, an average $\log P_{ahc}$ value of 0.79 (±0.089) was calculated for phenol. Analogous treatments of other $\log P_{ahc}$ data resulted in a data pool of sufficient size to start our f_{ahc} derivations.

We selected the following starting series of simple, mono-functional structures: (a) aliphatic alcohols, (b) pyridines, and (c) aliphatic carboxylic acids. They fulfilled our demands of presenting themselves with at least six homologues per series and a sufficient variation in the applied hydrocarbon of the solvent system. To indicate what we were aiming for, we refer to Table 3. We correlated the measured $\log P_{ahc}$ values for these structures with calculated Σf_{oct} and with the measured $\log P_{oct}$. Relating $\log P_{ahc}$ to Σf_{oct} leads to better results than relating it to $\log P_{oct}$. Values in parentheses are 95% confidence limits. $|\overline{\Delta}|$ represents the averaged absolute residuals:

Aliphatic alcohols:

$$\log P_{ahc} = 1.205\ (\pm 0.024)\ \Sigma f_{oct} - 1.900\ (\pm 0.054) \quad (2)$$

$$n = 7;\ r = 0.9996;\ s = 0.043;\ F = 5840;\ |\overline{\Delta}| = 0.031$$

$$\log P_{ahc} = 1.061\ (\pm 0.080)\ \log P_{oct} - 1.862\ (\pm 0.120) \quad (3)$$

$$n = 7;\ r = 0.9978;\ s = 0.098;\ F = 1150;\ |\overline{\Delta}| = 0.073$$

Pyridines:

$$\log P_{ahc} = 1.190\ (\pm 0.021)\ \Sigma f_{oct} - 1.261\ (\pm 0.078) \quad (4)$$

$n = 11;\ r = 0.9997;\ s = 0.058;\ F = 15200;\ |\overline{\Delta}| = 0.047$

$$\log P_{ahc} = 0.995\ (\pm 0.038)\ \log P_{oct} - 0.959\ (\pm 0.137) \quad (5)$$

$n = 10;\ r = 0.9989;\ s = 0.101;\ F = 3730;\ |\overline{\Delta}| = 0.056$

Aliphatic carboxylic acids:

$$\log P_{ahc} = 1.192\ (\pm 0.012)\ \Sigma f_{oct} - 2.832\ (\pm 0.059) \quad (6)$$

$n = 13;\ r = 0.9999;\ s = 0.061;\ F = 48100;\ |\overline{\Delta}| = 0.038$

$$\log P_{ahc} = 1.265\ (\pm 0.184)\ \log P_{oct} - 2.916\ (\pm 0.461) \quad (7)$$

$n = 8;\ r = 0.9896;\ s = 0.333;\ F = 283;\ |\overline{\Delta}| = 0.057$

Continuing our developments, we chose a *stepwise* procedure in order to make correct choices of the necessary k_n values. *Step I* forms the combination of Equations 2, 4 and 6 into one single equation:

$$\log P_{ahc} =$$

$$1.199\ (\pm 0.010)\ \Sigma f_{oct} - 0.259\ (\pm 0.009)\ k_n - 0.021\ (\pm 0.078) \quad (8)$$

$n = 31;\ r = 0.9998;\ s = 0.067;\ F = 32400;\ |\overline{\Delta}| = 0.053$

The applied k_n values 7, 5 and 11, respectively, were estimated via a carefully controlled trial-and-error fitting. The slope (ρ) of Equation 8 lies close to the average slopes of Equations 2, 4 and 6, and is suited to transform octanol/water details into ahc/water analogues; for instance, C_M (ahc) = $\rho \times 0.219$.

In *step II* the aliphatic carboxylic acid group was connected with eight aliphatic esters and nine dialkyl-nitrosamines applying $k_n = 2$ and $k_n = 3$ for the new participating groups:

$$\log P_{ahc} =$$

$$1.189\ (\pm 0.012)\ \Sigma f_{oct} - 0.252\ (\pm 0.007)\ k_n - 0.045\ (\pm 0.047) \quad (9)$$

$n = 30;\ r = 0.9997;\ s = 0.070;\ F = 21100;\ |\overline{\Delta}| = 0.049$

The $|\overline{\Delta}|$ values for the three types of structures were 0.037 (acids), 0.050 (esters) and 0.067 (nitrosamines).

In *step III* the five series were united in one regression equation using the k_n values denoted in the two preceding steps:

$$\log P_{ahc} =$$

$$1.196\ (\pm 0.010)\ \Sigma f_{oct} - 0.256\ (\pm 0.006)\ k_n - 0.040\ (\pm 0.043) \quad (10)$$

$$n = 48;\ r = 0.9996;\ s = 0.071;\ F = 31400;\ |\overline{\Delta}| = 0.055$$

The stepwise procedure was ended with the incorporation of seven aromatic hydrocarbons applying $k_n = 1$ for the benzene derivatives and $k_n = 2$ for naphthalene. In this way, a correct correlation procedure was achieved:

$$\log P_{ahc} =$$

$$1.194\ (\pm 0.009)\ \Sigma f_{oct} - 0.254\ (\pm 0.006)\ k_n - 0.051\ (\pm 0.039) \quad (11)$$

$$n = 55;\ r = 0.9996;\ s = 0.073;\ F = 32200;\ |\overline{\Delta}| = 0.056$$

The procedure can be completed by adding small groups of compounds with identical structural pattern or by adding the available compounds separately. In the final setup, we obtain a $\log P_{ahc}$ system based on the averaging of 799 partition data; especially the neglectable intercept value is worth mentioning:

$$\log P_{ahc} = 0.995\ (\pm 0.005)\ \Sigma f_{ahc} - 0.003\ (\pm 0.014) \quad (12)$$

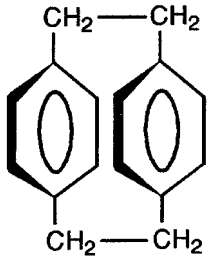
$$n = 365;\ r = 0.9984;\ s = 0.118;\ F = 113000;\ |\overline{\Delta}| = 0.10$$

Conclusions

The *hydrophobic fragmental constant* or Σf *system for octanol/water partitioning* [3–8] represents the first fragmental system for calculating log P. Fragmental methods splice molecules into adequate fragments and apply correction rules coupled with the molecular connectivity. The attractivity of the Σf system is its ease of application; due to the rather limited number of correction rules implemented in the system, it is the only fragmental approach that allows a log P calculation by hand. Consequently, this approach is of high didactic value; lipophilicity contributions of molecular moieties can simply be attributed and correction rules are simple to apply. Ever since the first presentation of the Σf system, criticisms of the constant C_M were rife. In the course of years, however, a growing amount of evidence became available

due to the increasing number of log P data and the possibility of tracing outliers. The applicability of C_M in the derivation of an adequate Σf_{ahc} system significantly contributed to the acceptability of the constant. The value of C_M is at the moment involved in an extended confirmation procedure with HPLC data [11].

A number of shortcomings in the system as it operates at present are evident and they demand for future studies: regarding the rules given in the 'Practical procedure of Σf calculations' section for the application of C_M corrections, it should be stressed that a good knowledge about chemical structure in connection with the physicochemical properties is required. This is especially true for a further development of the Σf system. Difficulties appear to arise for corrections coupled with proximity effects (in particular polyhalogenation). Both resonance interaction and its decoupling by steric factors will need careful evaluation of the total impact on lipophilic behaviour. The observation of UV absorption spectra will be helpful in case of doubt.



Conformational aspects need a lot of attention. Correct rules are not available, as exemplified in the paracyclophane structure. The log P* value is 3.70 [10]; the Σf calculation results in 5.47. It seems adequate to suggest that the facing sides of the two phenyl rings prevent the development of the full hydrophobic interaction pattern favouring a down-correction by one f C_6H_4 (= 1.697) and resulting in a final Σf value of 3.77.

For industrial purposes, log P data have to be calculated for huge numbers of compounds, which inevitably demands the availability of computerized versions of a calculation method. Computerized versions of the Σf approach that are available are shown in Table 4.

The routine application of calculation procedures requires the continual comparison of their results with experimental data [14–16]. Useful sets of experimental log P data are the Hansch/Leo listings of log P* values [10,17] or the collection provided by Meylan and Howard as part of the KOWWIN software.

We evaluated 14 commercially available calculation programs for a comparative test of their predictive power [14]. The entire database included

Table 4. Software to calculate log P with the Σf approach

Program	Method	References
PROLOGP_cdr	Rekker, original version	Rekker [5], Darvas et al. [12]
SANALOGP_EO	Rekker, extended original version	Petelin et al. [13]
Σf SYBYL	Rekker, revised version	Rekker and Mannhold [7]
SANALOGP_ER	Rekker, extended revised version	Petelin et al. [13]

simple organic structures and more complicated drugs. The validity of the calculation programs was far better for the simple organic structures than for the drug molecules. Fragmental methods yielded the best results, followed by atom-based approaches and procedures based on molecular properties. However, we did not make a comprehensive study of the methods based on molecular properties and included only a limited data set (n = 170). Thus, the comparative validity of calculation procedures remains to be clarified in future investigations.

The Σ*f system for aliphatic hydrocarbon/water partitioning* [9] has its special merits in the modelling of blood–brain barrier crossings. Among the chemical descriptors that capture the penetration of drug molecules into the brain, a key role is ascribed to lipophilicity. A decade ago, the Ganellin group [18] showed that Δ log P (i.e. log P_{oct} minus log P_{ahc}) was useful in predicting the brain penetration of H_2 antagonists.

As an initial, tentative test, we applied our Σf_{ahc} system to a set of 24 compounds, for which logarithms of equilibrium brain/blood concentration ratios (log BB) were available [18,19]. Selection of the compounds was due to their accessibility to calculation with our Σf_{ahc} system:

$$\log BB = 0.23\ (\pm 0.08)\ \Sigma f_{oct} - 0.22\ (\pm 0.25) \tag{13}$$

$$n = 24;\ r = 0.778;\ s = 0.321;\ F = 34$$

$$\log BB = 0.18\ (\pm 0.03)\ \Sigma f_{ahc} + 0.04\ (\pm 0.09) \tag{14}$$

$$n = 24;\ r = 0.935;\ s = 0.181;\ F = 153$$

$$\log BB = -0.22\ (\pm 0.10)\ \Delta\ \Sigma f + 0.52\ (\pm 0.17) \tag{15}$$

$$n = 24;\ r = 0.699;\ s = 0.365;\ F = 21$$

We surprisingly found a better fit with Σf_{ahc} than with $\Delta\ \Sigma f$ data, thereby contradicting the above cited preference to use Δ log P for modelling brain

penetration. Obviously, one swallow does not make a summer. Thus, this single example and the limited data do not allow generalized statements; further and more detailed investigations are warranted.

The main drawback of the Σf_{ahc} system is the availability of quite a restricted number of f_{ahc} values, as easily proven by inspecting Table 1. Thus, the most important step forward in this system would be the derivation of additional f_{ahc} values, in particular for heterocyclic fragments.

Lipophilicity is clearly a major determinant of biological activity and its use in new approaches like diversity searching adds additional impact on this chemical descriptor. Thus, the need for quick and reliable methods to calculate lipophilicity is undisputed. Despite their obvious drawbacks, fragmental methods including the Σf system remain the preferential choice.

Acknowledgements

The valuable contributions of Greetje Bijloo (Amsterdam), Karl Dross (Düsseldorf) and Gerrit de Vries (Amsterdam) to various steps in the development of the Σf systems are gratefully acknowledged.

References

1. Taylor, P.J., In Hansch, C., Sammes, P.G. and Taylor, J.B. (Eds.) Comprehensive Medicinal Chemistry, Vol. 4, Pergamon Press, Oxford, 1990, pp. 241–294.
2. Sangster, J., Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry, Wiley Series in Solution Chemistry, Vol. 2, Wiley, New York, NY 1997.
3. Nys, G.G. and Rekker, R.F., Chim. Ther., 8 (1973) 521.
4. Nys, G.G. and Rekker, R.F., Chim. Ther., 9 (1974) 361.
5. Rekker, R.F., The Hydrophobic Fragmental Constant, Its Derivation and Application. A Means of Characterizing Membrane Systems, Pharmacochem. Library Vol. 1, Elsevier, Amsterdam, 1977.
6. Rekker, R.F. and de Kort, H.M., Eur. J. Med. Chem., 14 (1979) 479.
7. Rekker, R.F. and Mannhold, R., Calculation of Drug Lipophilicity. The Hydrophobic Fragmental Constant Approach, VCH, Weinheim, 1992.
8. Mannhold, R., Rekker, R.F., Dross, K., Bijloo, G. and de Vries, G., Quant. Struct.–Act. Relats., 17 (1998) 517.
9. Rekker, R.F., Mannhold, R., Bijloo, G., de Vries, G. and Dross, K., Quant. Struct.–Act. Relats., 17 (1998) 537.
10. Hansch, C. and Leo, A., The Pomona College Medicinal Chemistry Project, Pomona College, Claremont, CA, 1983.
11. Rekker, R.F. and de Vries, G., The Applicability of the Fragmentation Principle, as Operated in log P Studies, to HPLC Retention Values. Part I: HPLC Studies on Methylbenzenes; The Derivation of a Corrective Multiple, in preparation.
12. Darvas, F., Erdoz, I. and Teglas, G., In Hadzi, E. and Jerman-Blazic, B. (Eds.) QSAR in Drug Design and Toxicology, Elsevier, Amsterdam, 1987, pp. 70–73.

13. Petelin, D.E., Arslanov, N.A., Palyulin, V.A. and Zefirov, N.S., Extended parametrisation of Rekker's f system for drug lipophilicity calculation, 10th European Symposium on Structure–Activity Relationships, Barcelona, Abstract B263, 1994.
14. Mannhold, R. and Dross, K., Quant. Struct.–Act. Relats., 15 (1996) 403.
15. Mannhold, R., Rekker, R.F., Sonntag, C., ter Laak, A.M., Dross, K. and Polymeropoulos, E.E., J. Pharm. Sci., 84 (1995) 1410.
16. van de Waterbeemd, H. and Mannhold, R., In Pliska, V., Testa, B. and van de Waterbeemd, H. (Eds.) Lipophilicity in Drug Action and Toxicology. Methods and Principles in Medicinal Chemistry, Vol. 4, VCH, Weinheim, 1996, pp. 401–418.
17. Hansch, C., Leo, A. and Hoekman, D. (Eds.) Exploring QSAR; Hydrophobic, Electronic and Steric Constants, ACS Professional Reference Book, 1995.
18. Young, R.C., Mitchell, R.C., Brown, T.H., Ganellin, C.R., Griffiths, R., Jones, M., Rana, K.K., Saunders, D., Smith, I.R., Sore, N.E. and Wilks, T.J., J. Med. Chem., 31 (1988) 656.
19. Abraham, M.H., Chadha, H.S. and Mitchell, R.C., J. Pharm. Sci., 83 (1994) 1257.