ELSEVIER

Journal of Fluorine Chemistry 115 (2002) 119–128

JOURNAL OF FLUORINE CHEMISTRY

www.elsevier.com/locate/jfluchem

Statistical and theoretical studies of fluorophilicity

Farah T.T. Huque, Kyle Jones, Robert A. Saunders, James A. Platts*

Department of Chemistry, Cardiff University, P.O. Box 912, Cardiff CF10 3TB, UK

Received 26 October 2001; received in revised form 9 January 2002; accepted 26 January 2002

Abstract

Linear free energy relation (LFER) models of the partition of 91 organic compounds between organic and fluorous solvents are presented. By reducing organic/fluorous partition to a sum of solvent–solute interaction terms we have developed accurate predictive models, with a standard deviation for prediction less than three times the estimated experimental error. Moreover, the physicochemical factors which affect a molecule's fluorophilicity are identified: fluorine content, dispersion and hydrogen bond acidity factors are particularly important, with polarity, hydrogen bond basicity and size effects playing a much smaller role. The predictive ability of multiple linear regression is confirmed for several test sets, and factors affecting fluorophilicity for different classes of molecules discussed. *Ab initio* methods were used to calculate the stabilisation energies for different conformations of $(F_2)_2$ dimers. Correlation consistent basis sets at the MP2 level were used since split valence basis sets were found to be inadequate in their description of the system. The potential energy surface (PES) of the most stable conformer (L-shaped) was analysed and calculated stabilisation energies were used to successfully fit parameters for potential energy functions. A quadrupole term was added to the Buckingham function yielding an excellent fit.

Keywords: Fluorine; Fluorous partition coefficient; Structure–activity relations; Ab initio calculations

1. Introduction

The use of fluorinated solvents in synthesis and catalysis has become increasingly popular in recent years (see [1–4] for recent reviews). In particular, the ability of fluorous solvents such as perfluoro(methylcyclohexane), $CF_3C_6F_{11}$, to preferentially dissolve heavily fluorinated solutes in a 'like dissolves like' manner has been exploited to perform reactions in such media [5], or to separate desired products from reaction mixtures [6]. The fluorination of molecules often takes the form of addition of long perfluorinated-alkyl chains, or 'ponytails', to reactants or catalysts [7]. Biphasic reactions take advantage of the fact that organic and fluorous phases are typically immiscible at ambient temperatures, but may homogenise at elevated temperatures [3]. One can therefore expose reactants in an organic phase to a catalyst in a fluorous phase simply by heating, and separate products (into an organic phase) from the catalyst (into the fluorous phase) on cooling.

Clearly, the rational design of such experiments would be substantially aided by the ability to predict *a priori* a given molecule's tendency to dissolve in fluorous media. This is most commonly measured by the molecule's partition coefficient, P, between fluorous and organic layers [8]. This is usually converted onto a free energy scale by taking its natural logarithm, and the quantity $\ln P$ is referred to as the 'fluorophilicity'. Throughout this work we have used the standard system proposed by Rocaboy et al. [9], namely the partition of molecules between $CF_3C_6F_{11}$ and toluene, as shown in Eq. (1).

$$\ln P = \ln\left[\frac{c(CF_3C_6F_{11})}{c(CH_3C_6H_5)}\right], \qquad T = 298\,\mathrm{K} \tag{1}$$

Several empirical 'rules of thumb' for the design of fluorophilic molecules have been proposed [2,10], including a minimum fluorine content of 60%, the presence of one or more fluorous ponytails, and the absence of hydrogen bonding or polar groups which may interact with the organic phase.

Kiss et al. [10] have recently shown that the fluorophilicity of 59 fluorinated organic molecules may be estimated using a neural network (NN) combination of eight descriptors chosen from a pool of almost 100 possible descriptors. Descriptors tested for significance included simple structural properties such as fluorine content, surface area and volume, electronic properties such as HOMO and LUMO energies and electrostatic potentials, and 'weighted holistic invariant molecular' (WHIM) descriptors. They found that the surface area of the molecule and the distribution of fluorine atoms

* Corresponding author. Tel.: +44-29-2087-4950; fax: +44-29-2087-4030.
E-mail address: platts@cardiff.ac.uk (J.A. Platts).

0022-1139/02/$ – see front matter
PII: S0022-1139(02)00034-9

within the molecule were particularly important, indicating that fluorines should be on the exterior of the molecule, able to interact with the fluorinated solvent. The final NN model uses the Hildebrand solubility parameter (δ) which describes a substance's cohesive energy density and correlates strongly with polarity. This may be used to give some approximation of fluorophilicity. Parallels may be drawn between the Hildebrand parameter and the polarity/polarisability, ***S***, along with ***F***, the fluorine content, used in this study to describe fluorophilic interactions. They also reported that fluorine content was not significant in their NN analysis. However, they assigned this to a dearth of molecules with low fluorine content in their dataset, rather than to a conflict with the proposed rule.

Although Kiss et al.'s work represents a very impressive model of fluorophilicity, its reliance on WHIM descriptors makes it somewhat difficult to interpret in simple chemical terms. Also, NN methods inherently hide the working of the model from view, so interested readers cannot apply their methodology to their own compounds. We have therefore analysed fluorophilicity from a different standpoint, namely the linear free energy relations (LFERs) developed by Abraham and co-workers [11]. The LFER approach aims to describe solvation and related processes as a combination of specific solute–solvent interactions, namely hydrogen bonding, electrostatic, induction and dispersion effects, and cavity formation. To this end, Abraham has developed his 'general solvation equation', given here as Eq. (2):

$$\ln \mathrm{SP} = e\boldsymbol{E} + s\boldsymbol{S} + a\boldsymbol{A} + b\boldsymbol{B} + v\boldsymbol{V} + c \quad (2)$$

where ln SP is the set of solvation properties to be modelled. The solute descriptors ***E***, ***S***, ***A***, ***B***, and ***V*** characterise the ability of a solute to interact with solvents in specific ways, and are discussed in more detail (see Table 1).

E is the solute's 'excess molar refraction', defined as the molar refraction of the solute minus that of an alkane of equivalent volume [12]. In many cases, this is a measure of the number of n- and π-electron pairs in the solute, and their ability to interact via dispersion forces with solvents. Heavily fluorinated molecules (i.e. $F(CF_2)_n$-substituted ones), however, have lower molar refraction than their equivalent alkanes and negative ***E*** values, and so do not interact strongly via dispersion forces. ***S*** is the solute's polarity/polarisability, again taken relative to an equivalent alkane. The ***S*** scale was originally set-up from retention times on polar, non-hydrogen bonding GC columns [13], with the retention of alkanes set to zero. As with ***E***, perfluorinated molecules are less polar/polarisable that alkanes, and consequently have negative ***S*** values.

A and ***B*** are the solute hydrogen bond acidity and basicity, respectively, and characterise these specific interactions [14]. ***V*** is McGowan's characteristic molecular volume [15], calculated from a sum of atomic and bond contributions. Examples of typical values of descriptors for some common organic compounds are given in Table 1. In this work, a sixth descriptor ***F***, the fluorine content of the molecule, was also considered. It has been suggested that a fluorine content of 0.60 is required for a molecule to be 'fluorous', although Kiss et al. suggest the detailed structure of this fluorous content is also important, with long perfluorinated 'ponytails' particularly favoured.

The coefficients *e*, *s*, *a*, *b*, *v* and *c* in Eq. (2) determine the relative importance of each descriptor in modelling a given solvation property, and are typically found by multiple linear regression (MLR) for a set of molecules with known ln SP and descriptor values. In the current case of fluorous biphase partition, they reflect how strongly and in what manner the solute interacts with the fluorous and organic phases. For instance, a negative *a* coefficient will indicate that acidic molecules interact more strongly with toluene than with perfluoro(methylcyclohexane). The sixth descriptor ***F*** lies on dissimilar scale from the others and cannot therefore be compared directly with these.

In order to investigate the factors that cause heavily fluorinated solutes to dissolve in fluorous solvents, we have investigated the interaction between fluorine molecules in the F_2 dimer. *Ab initio* calculations were used to derive stabilisation energies for the fluorine dimer in various conformations, and at a range of separations. Recent studies by Parra and Zeng [16] focused on the interaction between two molecules of CF_4 at the MP2 level. They found that varying their split valence basis sets considerably affected the calculated interaction energies (1.1 ± 0.4–2.2 ± 0.4 kJ/mol). A study of N_2 dimers by Couronne and Ellinger [17] found that, at the DFT level, the B3LYP functional was unable to properly describe weak intermolecular interactions, whereas the PW91 exchange-correlation functional gives results closer to experimental values. BSSE corrected results from MP2 to MP4 with a 6-31G* basis set was found to satisfactorily describe the N_2–N_2 system, finding T-shaped and skewed conformations as the only ones stable. Studies on rare gas—F_2 complexes [18] suggested that correlation consistent basis sets such as those developed by Woon and Dunning [19], at a high level of theory were required to adequately describe such systems.

We have addressed similar issues of theoretical method and basis set for specific case of the F_2 dimer, and selected an

Table 1
Typical descriptors for some organic compounds

Name	*E*	*S*	*A*	*B*	*V*
Hexane	0.00	0.00	0.00	0.00	0.954
Ethanol	0.25	0.42	0.37	0.48	0.449
Ethylamine	0.24	0.35	0.16	0.61	0.490
γ-Pentanolactone	0.34	1.26	0.00	0.55	0.779
Toluene	0.60	0.52	0.00	0.14	0.857
Benzoic acid	0.73	0.90	0.59	0.40	0.932
3-Nitrophenol	1.05	1.57	0.79	0.23	0.949
$R_{f7}C(O)NHMe$	−0.71	0.53	0.37	0.62	1.758
$R_{f6}(CH_2)_3OH$	−0.68	0.00	0.35	0.33	1.666
Tetrafluoromethane	−0.55	−0.25	0.00	0.00	0.320

appropriate compromise between accuracy and computational resources. From this we have generated potential energy surfaces (PES) of the fluorine dimer, from which Buckingham and Lennard-Jones potential energy functions will be derived. The contribution of the quadrupole–quadrupole term to the intermolecular fluorine–fluorine interactions was also be considered.

2. Results and discussion

2.1. Fluorous phase partition

Initial regression of the entire dataset against the five Abraham descriptors (model 1) gave a reasonable linear model ($R^2 = 0.85$, RMS = 1.02). Eight molecules were omitted from the final model as outliers: see below for discussion and justification of these. The remaining 90 data were modelled using the standard Abraham approach (Eq. (2)), which yielded $R^2 = 0.86$, RMS = 0.93 (model 2). A sixth descriptor, ***F***, was added to the Abraham descriptors and the total data set of 98 points was modelled giving $R^2 = 0.91$, RMS = 0.79 (model 3).

Removing the eight outliers resulted in considerable improvement the fit: this is shown as Eq. (3). Hydrogen bond basicity, ***B***, was insignificant at the 95% level, and was therefore excluded (model 4).

$$\ln P = -2.130 - 1.462\boldsymbol{E} - 1.030\boldsymbol{S} - 2.680\boldsymbol{A} - 0.398\boldsymbol{V} + 4.632\boldsymbol{F}, \quad n = 90, \quad R^2 = 0.949, \quad R^2_{CV} = 0.940, \quad \text{S.D.} = 0.568, \quad F(90,5) = 261.9 \tag{3}$$

Statistics for all models are given in Table 3.

The *t*-ratios indicate that all five remaining descriptors are significant in regression against ln *P*. Observed against calculated ln *P* values are shown in Fig. 1 and are reported in Table 2. Eq. (3) represents impressive accuracy for a large and varied set of fluorophilicity data, with a standard deviation less than three times the estimated experimental error of 0.2 [10]. The accuracy of this fit is not as impressive as for Kiss et al.'s NN model, although it should be noted that Eq. (3) is trained on almost twice as many data gathered from several sources, and covers a much wider range of functionality and fluorine content. Table 4 reports correlation coefficients between the descriptors used in Eq. (3), showing that ***E*** is somewhat correlated to other descriptors, especially ***S*** and ***V***, but that no other descriptors show any significant correlation with each other. We therefore have confidence in the coefficients shown in Eq. (3).

Eight molecules were omitted from Eq. (3) as outliers, which fall into two specific classes of molecule. Firstly, three closely related aromatic esters containing R_{f7} chains (**63**, **64**, and **65**) show a great deal of scatter, for instance, **63** and **64** differ only by a single CH_2 group, but their ln *P* values differ by more than 1.5 units. Any model based on structure alone will fail in such circumstances, and we have omitted these three from all further analysis. Secondly, a set of five molecules (**82** and **93–96**) containing three fluorous chains on a single phenyl ring were all predicted to have much lower fluorophilicities than is experimentally observed. It appears that the fragmental approach does not describe the effect of multiple fluorous substitution on one ring, such that calculated descriptors for such structures are in error. Again, these five molecules were omitted from Eq. (3). Exclusion of these eight outliers significantly improved fit, giving a final R^2 of 0.95. The order of significance of the descriptors, as indicated by *t*-ratios (see Table 3), was not affected by the removal of these eight points. This indicates that, overall, the descriptors can describe fluorophilicity well, but these outliers could not be described by the current fragmental method. Non-linearity between fluorophilicity and the descriptors may mean that some inaccuracy may have been incurred on imposition of a linear fit. Such non-linearity was found between fluorophilicity and the top eight descriptors in the NN approach of Kiss et al. [10].

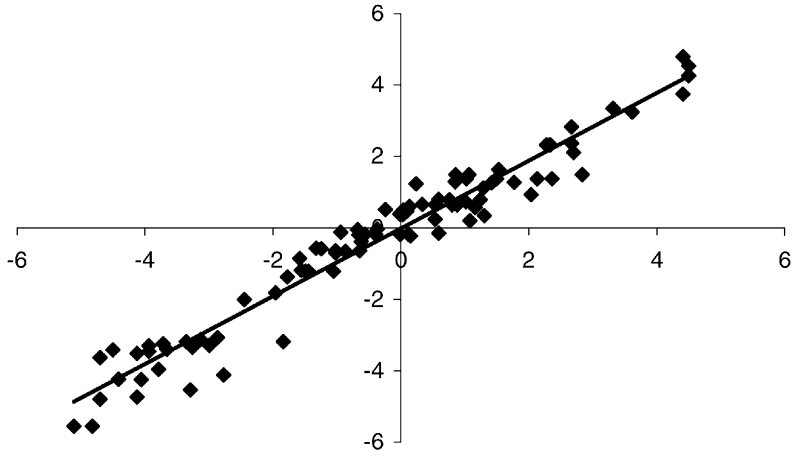

Fig. 1. Observed vs. calculated ln *P* from Eq. (3).

Table 2
Observed and calculated ln P from Eq. (3)

No.	Name[a]	ln P	
		Observed	Calculated
1	Decane	−2.86	−3.07
2	Undecane	−3.13	−3.13
3	Dodecane	−3.35	−3.19
4	Tridecane	−3.71	−3.24
5	Tetradecane	−3.94	−3.3
6	Hexadecane	−4.50	−3.41
7	Dec-1-ene	−2.99	−3.29
8	Undec-1-ene	−3.26	−3.34
9	Dodec-1-ene	−3.66	−3.40
0	Tridec-1-ene	−3.94	−3.46
11	Tetradec-1-ene	−4.12	−3.51
12	Hexadec-1-ene	−4.70	−3.62
13	$R_{f8}CH{=}CH_2$	2.67	2.82
14	Cyclohexanone	−3.79	−3.96
15	Cyclohexenone	−4.06	−4.25
16	Cyclohexanol	−4.12	−4.74
17	Trifluoroethanol	−1.77	−1.37
18	$(CF_3)_2CHOH$	−1.02	−0.70
19	$R_{f6}(CH_2)_2OH$	0.10	0.47
20	$R_{f6}(CH_2)_3OH$	−0.24	0.50
21	$R_{f8}(CH_2)_2OH$	1.02	0.72
22	$R_{f8}(CH_2)_3OH$	0.59	0.80
23	$R_{f10}(CH_2)_3OH$	1.42	1.25
24	Pentafluorobenzene	−1.24	−0.58
25	Hexafluorobenzene	−0.94	−0.12
26	Ethylbenzene	−4.41	−4.23
27	Dodecylbenzene	−4.70	−4.79
28	$R_{f8}(CH_2)_3C_6H_5$	−0.02	0.38
29	o-$R_{f6}(CH_2)_3C_6H_4(CH_2)_3R_{f6}$	1.03	1.37
30	o-$R_{f8}(CH_2)_3C_6H_4(CH_2)_3R_{f8}$	2.34	2.32
31	o-$R_{f10}(CH_2)_3C_6H_4(CH_2)_3R_{f10}$	3.62	3.23
32	m-$R_{f8}(CH_2)_3C_6H_4(CH_2)_3R_{f8}$	2.28	2.32
33	p-$R_{f8}(CH_2)_3C_6H_4(CH_2)_3R_{f8}$	2.33	2.32
34	$R_{f8}(CH_2)_3Cl$	0.03	0.37
35	$R_{f8}(CH_2)_3NH_2$	0.85	1.29
36	$R_{f8}(CH_2)_3NH(CH_2)_3R_{f8}$	3.32	3.34
37	$(R_{f6}(CH_2)_2)_3P$	4.41	3.75
38	$(R_{f8}(CH_2)_3)_3P$	4.41	4.79
39	$(R_{f8}(CH_2)_4)_3P$	4.50	4.53
40	$(R_{f8}(CH_2)_5)_3P$	4.50	4.27
41	$(R_{f6}(CH_2)_2)_2PC_{10}H_{19}$ (menthyl)	1.29	1.11
42	$(R_{f8}(CH_2)_2)_2P$ $C_{10}H_{19}$ (menthyl)	2.70	2.10
43	$(p$-$R_{f6}C_6H_4)_3P$	−1.32	−0.57
44	$(p$-$R_{f8}C_6H_4)_3P$	0.76	0.78
45	$Ph(CH_2)_2SiH$	−3.29	−4.53
46	$Ph(CH_2)_2SiOC_8H_{13}$	−5.11	−5.56
47	$Ph(CH_2)_2SiOC_6H_{11}$ (cyclohexyl)	−4.82	−5.56
48	$R_{f6}I$	1.31	0.34
49	$R_{f8}I$	2.04	0.93
50	$R_{f10}I$	2.84	1.48
51	$R_{f8}CH{=}CH_2$	2.67	2.82
52	$R_{f8}(CH_2)_3SH$	0.24	1.23
53	$R_{f8}N(CH_2CH_2)_2O$	0.86	1.48
54	$R_{f6}S(CH_2)_2CO_2Et$	−0.67	−0.05
55	$R_{f8}S(CH_2)_2CO_2Et$	0.04	0.49
56	CF_3SPh	−2.45	−2.01
57	m-$CF_3SC_6H_4CF_3$	−1.58	−0.85
58	$R_{f8}SPh$	0.59	−0.15
59	$R_{f7}CH_2NHMe$	1.07	1.49
60	$R_{f7}CH_2NMe_2$	1.53	1.63
61	$R_{f7}CH_2N(CH_2CH_2)_2O$	0.14	0.60
62	$R_{f7}CH_2NHCH(Me)Ph(+)$	−0.87	−0.65
63	$R_{f7}C(O)Oph$	0.48	b
64	$R_{f7}C(O)OCH_2Ph$	2.14	b
65	p-$R_{f7}C(O)OCH_2CH_4OCF_3$	3.15	b
66	$R_{f7}C(O)Sme$	1.16	0.57
67	$R_{f7}C(O)NHMe$	0.15	−0.23
68	$R_{f7}C(O)NMe_2$	0.34	0.66
69	$R_{f7}C(O)N(CH_2CH_2)_2O$	−0.62	−0.38
70	$R_{f7}C(S)Ome$	1.08	0.19
71	$R_{f7}C(S)NMe_2$	−0.66	−0.20
72	$R_{f7}C(S)N(CH_2CH_2)_2O$	−1.56	−1.18
73	$R_{f7}C(S)NHCH(Me)Ph(+)$	−1.84	−3.18
74	C_6H_6	−2.77	−4.12
75	CF_3Ph	−1.96	−1.82
76	$R_{f6}Ph$	0.54	0.24
77	$R_{f8}Ph$	1.24	0.78
78	$R_{f10}Ph$	1.77	1.28
79	o-$R_{f8}C_6H_4CF_3$	1.50	1.37
80	m-$R_{f8}C_6H_4CF_3$	2.37	1.37
81	p-$R_{f8}C_6H_4CF_3$	2.13	1.37
82	p-$R_{f8}C_6H_4R_{f8}$	4.98	b
83	$[p$-$CF_3C_6H_4(CF_2)_4]_2$	−0.56	−0.18
84	o-$R_{f6}(CH_2)_2C_6H_4Cl$	−0.64	−0.63
85	p-$R_{f6}(CH_2)_2C_6H_4Cl$	−1.02	−0.63
86	p-$R_{f8}(CH_2)_2C_6H_4CI$	−0.37	−0.04
87	o-$R_{f6}(CH_2)_2C_6H_4Br$	−1.05	−1.22
88	m-$R_{f6}(CH_2)_2C_6H_4Br$	−1.44	−1.22
89	p-$R_{f6}(CH_2)_2C_6H_4Br$	−1.49	−1.22
90	o-$R_{f8}C_6H_4CO_2Me$	−0.39	−0.18
91	m-$R_{f8}C_6H_4CO_2Me$	0.12	−0.18
92	p-$R_{f8}C_6H_4CO_2Me$	−0.01	−0.18
93	1,3,5-$R_{f8}C_6H_3(CF_3)_2$	4.05	b
94	1,3,5-$(R_{f8})_2C_6H_3CO_2Me$	4.41	b
95	1,3,5-$(R_{f8})_2C_6H_3CH_2OH$	3.62	b
96	1,3,5-$(R_{f8})_2C_6H_3CHO$	4.25	b
97	2-$R_{f8}C_5H_4N$ (pyridine)	0.54	0.64
98	3-$R_{f8}C_5H_4N$ (pyridine)	0.88	0.64
99	4-$R_{f8}C_5H_4N$ (pyridine)	0.80	0.64

[a] R_{fn} is shorthand for $CF_3(CF_2)_{n-1}$.

[b] Omitted from final model, see text for discussion.

This model may therefore be used to predict new ln P values with an approximate error of 0.57, with the following caveats: firstly, Eq. (3) is only valid within the range of descriptors spanned by the molecules used in training it. In fact this is a rather wide range, and should not overly restrict application of the model. Secondly, it is important to note that this method cannot be applied to transition metal-containing compounds, not due to failure of the LFER method itself but because of an inability to calculate descriptors for such species. This is unfortunate, since much of the interest in fluorous biphase partition involves metal-based catalysts—subsequent work will attempt to address this problem.

Eq. (3) not only represents a good statistical model of fluorophilicity, but also indicates the physicochemical factors that affect it. A t-score of 11.9 indicates that the most

Table 3
Coefficients and statistics for the four models of fluorophilicity

Model[a]	R^2	RMS	***E***		***S***		***A***		***B***		***V***		***F***	
			Coefficient	*t*-ratio	Coefficient	*t*-ratio	Coefficient	*t*-ratio	Coefficient	*t*-ratio	Coefficient	*t*-ratio	Coefficient	*t*-ratio
1	0.851	1.02	−5.106	−12.88	3.237	4.85	−2.275	−0.26	−2.690	−5.21	−0.851	−5.35	–	–
2	0.863	0.93	−4.465	−11.37	2.176	3.29	−2.294	−2.74	−1.889	−3.76	−0.856	−5.87	–	–
3	0.910	0.79	−2.700	−6.16	−0.848	−1.40	−2.641	−3.84	−1.379	−3.16	−0.432	−3.19	4.087	7.76
4	0.949	0.57	−1.462	−6.40	−1.030	−3.68	−2.680	−5.16	–	–	−0.398	−4.02	4.632	13.26

[a] Models defined as 1: standard Abraham 5-descriptor, no outliers removed; 2: Abraham 5-descriptor, eight outliers removed; 3: 6-descriptor (Abraham plus fluorine content), no outliers removed; 4: 6-descriptor, eight outliers removed.

significant factor in this model is ***F***, the fluorine content. Thus, it appears that Kiss et al.'s findings that ***F*** was not significant in their model is indeed an artefact of their restricted dataset, and is not an inherent property of fluorous solute–solvent systems. This is consistent with the empirical rule noted above, i.e. that fluorous solvents preferentially dissolve heavily fluorinated solutes. The individual interactions which give rise to this 'like dissolves like' preference will be investigated further in Section 2.

E, the excess molar refraction, has a large negative coefficient ($e = -1.651$, $t = -4.87$), which indicates that n- and π-electron pairs act to strongly reduce fluorophilicity, pulling the molecule into the organic phase. This is a result of relatively strong dispersion interactions between such electron pairs and the π-system of toluene. The polarity/polarisability descriptor, ***S***, has a small, negative coefficient and *t*-ratio, indicating that polar and/or polarisable molecules have a slightly greater tendency to partition into the organic phase. However, this is a small effect, indicating that there is little difference in electrostatic or induction forces between organic and fluorous phases.

Hydrogen bond acidity, ***A***, has a rather large negative coefficient ($a = -2.66$, $t = -5.21$) and hence acts to strongly reduce fluorophilicity. That H-bond or Lewis acidity reduces ln *P* is perhaps unsurprising, since it is well known [14] that π-systems such as the benzene ring of toluene can act as H-bond bases. However, Eq. (3) represents the first time this specific effect has been quantified, such that the effect of inclusion of various H-bonding functional groups [28] can be predicted with confidence. The insignificance of ***B*** in Eq. (3) might also be expected, since neither toluene or perfluoro(methylcyclohexane) has any hydrogen bond acidity, and hence no preferential interaction with H-bond bases. Its non-significance must have implications for the design of new fluorous molecules, since it suggests that inclusion of such groups would not affect fluorophilicity, contrary to previously reported empirical rules. It should be noted that ***B*** is often related to ***E*** and ***S***, since all are commonly associated with excess electron density: in this case though, ***B*** shows no correlation with $r > 0.5$ with any other descriptor.

Table 4
Correlations between descriptors used in Eq. (3)

	E	***S***	***A***	***B***	***V***	***F***
E	1.00					
S	0.83	1.00				
A	0.00	0.06	1.00			
B	−0.49	−0.02	0.04	1.00		
V	−0.81	−0.59	−0.23	0.53	1.00	
F	−0.66	−0.45	0.18	0.28	0.36	1.00

The volume term, ***V***, is also small and negative, but with $t = -4.16$ this is clearly an important effect. It is apparent that larger molecules are more likely to partition into the organic phase. Molecular volume describes both the energetically favourable dispersion interactions between solute and solvent and also the energy required to create a cavity in the solvent to accommodate the solute. In this case, it seems that the attractive dispersion effects between solutes and toluene lead to the negative coefficient, although it is difficult to estimate the contribution of cavity effects to this value.

The R^2_{CV} in Eq. (3) indicates to some extent the predictive ability of the model, as individual molecules are omitted from the regression and the resulting model used to predict values of ln *P*. However, a much more rigorous test of the predictive power of such a model, analogous to how the model may be used in 'real' applications, is to split the total data set into training and test sets. Accordingly, we have omitted one-half of the data points, trained a model on the remainder, and applied this model to the omitted data. This process yields $R^2 = 0.946$, S.D. = 0.576 for the training set and $R^2 = 0.945$, S.D. = 0.561 for the test set. Reversing these test and training sets now gives $R^2 = 0.957$, S.D. = 0.499 for the training and $R^2 = 0.937$, S.D. = 0.616 for the test set. Thus, the model's accuracy applies not only to the data to which it was fitted, but can be used to predict new values within an S.D. of about 0.6.

A comparison of the factors affecting fluorophilicity for a series of related molecules is informative: Table 5 reports the individual contributions to fluorophilicity for four alcohols. Cyclohexanol has large ***E***, due mainly to the lone pairs on O, along with significant hydrogen bond acidity and basicity, and is rather large, all of which act to reduce fluorophilicity. There is no fluorine present in the molecule, so there is no

Table 5
Contributions to ln *P* for four alcohols

	E	***S***	***A***	***B***	***V***	***F***	Calculated	Observed
Cyclohexanol	0.50	0.56	0.35	0.40	0.904	0.00	−4.74	−4.12
CF_3CH_2OH	0.01	0.32	0.50	0.28	0.502	0.57	−1.37	−1.77
$R_{f8}(CH_2)_2OH$	−0.96	−0.12	0.35	0.34	1.887	0.70	0.36	0.44
$R_{f8}(CH_2)_3OH$	−0.96	−0.12	0.35	0.34	2.018	0.65	0.31	0.26

Fig. 2. Conformations of $(F_2)_2$ used and point group symmetries applied.

preference for the fluorous phase, and the partition is well modelled at roughly 98:2 organic:fluorous. $(CF_3)CH_2OH$ partitions slightly more into the fluorous phase (experimental value 85:15), and the LFER model predicts this. The fluorine content (57%) increases the partitioning to the fluorous phase although the effect of the other factors is also important. ***E*** and ***S*** are reduced substantially by the non-polar CF_3 group, which also affects ***A***. Overall, the presence of fluorine and the reduction of ***E*** are the dominant factors in explaining the increased fluorophilicity.

The related molecules $R_{f8}(CH_2)_2OH$ and $R_{f8}(CH_2)_3OH$ differ only by a CH_2 group, having the same fluorous ponytails, which act to greatly increase fluorophilicity, attached. Consequently, they have very similar partition values (27:73 and 36:64, respectively). The OH group is 'insulated' from the fluorous group by the CH_2, so ***A*** has values typical of primary alcohols. However, the R_{f8} group has very low ***E*** and ***S*** values due to its very low polarisability. These effects are also of great significance to the calculated partition values. The difference between the two molecules' descriptors lies in the volume due to the extra CH_2 group. The presence of this group also reduces the percentage fluorine content from 70 to 65% therefore acting to very slightly reduce fluorophilicity.

2.2. Intermolecular interactions between fluorines

2.2.1. Selection of theoretical model

In order to explore further the strength and nature of intermolecular interactions between fluorine containing species, we have analysed the dimer of F_2 using correlated *ab initio* and density functional methods. These calculations were performed primarily to make at least an estimate of the stabilisation due to the specific $F \cdots F$ interactions responsible for fluorophilicity. They will also serve as a preliminary study prior to more in-depth computational studies of such interactions. Although $(F_2)_2$ is perhaps not the most realistic model of the larger systems considered above, its small size allows us to perform much more rigorous calculations than would be feasible for larger models, and hence to carry the lessons from this through to subsequent studies.

The first step in such a study is to select an appropriate theoretical method and basis set for the task in hand. To do this, calculations were performed on five distinct conformations of $(F_2)_2$ (Fig. 2) using a wide range of basis sets and treatments of correlation (see Supplementary Information)[1]. The point group symmetry noted in Fig. 2 was applied throughout. Following a previous study [16] of $(CF_4)_2$, we initially used MP2 methods with basis sets based around Poplés [29] 6-31G and 6-311G type basis sets, varying in the number of polarisation and diffuse functions used. No overall pattern of stability was evident, though 6-311G type basis sets consistently predicted the L-shaped conformer to be the minimum, but there was disagreement as to which was the most stable geometry when 6-31G basis sets were employed. Discrepancies were also found regarding whether the linear and skewed conformers were minima or transition states. Single point energy calculations at the MP3 and CISD levels also gave little consistency as to which conformation was the most stable, each calculation predicting differently.

Due to the inconsistencies obtained using these split valence basis sets, we turned our attention to Dunning's correlation consistent basis sets which, by definition, include polarisation functions. This type of basis set has been shown to describe weak intermolecular interactions well for complexes of F_2 with rare gases [18]. Geometry optimisations on the five conformers shown in Fig. 2 were carried out at the

[1] Supplementary information is available on request from Corresponding Author.

MP2/cc-pVTZ level [19]. Harmonic frequency calculations confirmed that the L-shaped, skewed and cross conformers were true minima, with the L-shaped being the global minimum. This is an interesting contrast to Couronne and Ellinger's [17] finding that $(N_2)_2$ has T-shaped and skewed minima of roughly equal energy. In the case of $(F_2)_2$ the T-shape has one imaginary frequency, even though a comparatively large stabilisation energy was observed.

Expanding the basis set for single point calculations at the MP2 level to cc-pVQZ and cc-pV5Z made little difference to the calculated energies. This intimates that the triple zeta basis set is sufficient for describing the system in question. Adding diffuse (aug) functions to the cc-pVTZ basis set consistently increased stabilisation energies by ~0.5 kJ/mol for all conformers, but this is a much smaller variation than those seen for split valence basis sets. Single point energy calculations were also carried out at higher levels, including MP3, MP4, QCISD, and CCSD(T), using the cc-pVTZ basis set. Interaction energies fall very slightly (around 0.2 kJ/mol) with these higher levels, but overall the results are remarkably stable using the correlation consistent basis sets.

The MP2/cc-pVTZ counterpoise correction for the L-shaped F_2 dimer was calculated to be 0.65 kJ/mol, or approximately one-third of the overall stabilisation energy, similar to results for $(N_2)_2$ [18]. It is important to note, however, that the counterpoise correction is only an approximation to the true basis set superposition error (BSSE). The zero point energy (ZPE) correction to the stabilisation energy is rather large at 1.02 kJ/mol, but still considerably smaller than the overall interaction energy.

The popular density functional method B3LYP [30] failed completely to describe the intermolecular interactions present in $(F_2)_2$, predicting that all conformations are minima with little or no energetic preference for any. That the B3LYP method does not properly describe weak interactions such as van der Waals forces has also been observed in studies of $(N_2)_2$ interactions [18]. It has been suggested that the PW91 functional [31] performs better for such weak interactions. Our results for $(F_2)_2$ with PW91/cc-pVTZ go some way to support this interpretation, but also indicate that stabilisations are considerably over-estimated compared with *ab initio* methods.

2.2.2. Intermolecular potential energy surface

Based on the results presented above, the MP2/cc-pVTZ method was selected to investigate the PES around the most stable L-shaped conformation. The closest intermolecular $F\cdots F$ contact distance was varied from 2.3 to 5.3 Å at intervals of 0.02 Å giving 150 points on the PES. The stabilisation relative to two independent F_2 molecules was calculated and plotted to give Fig. 3a. The energy well was deepest (−1.84 kJ/mol) at an intermolecular $F\cdots F$ distance of 2.84 Å. The calculated stabilisations were then used to fit parameters for the Buckingham and Lennard-Jones energy functions (Eqs. (4) and (5), respectively). The relevant parameters were optimised for each function, including all four intermolecular $F\cdots F$ distances (optimised parameters are shown in Table 7). While both functions appeared to describe the PES well, the Buckingham potential ($R^2 = 0.995$, RMS error = 0.069 kJ/mol) had a far superior fit to the Lennard-Jones potential ($R^2 = 0.987$), which also displayed considerably larger scatter in the predicted data. Hence, it was decided to concentrate on the Buckingham function to describe this system.

Table 6
Selected stabilisation energies for two conformations of $(F_2)_2$

Level	Stabilisation energy (kJ/mol)	
	Cross	L-shape
MP2/cc-pVTZ	−1.456	−1.843
MP3/cc-pVTZ	−1.151	−1.303
MP4/cc-pVTZ	−1.464	−1.865
CCSD/cc-pVTZ	−1.141	−1.334
CCSD(T)/cc-pVTZ	−1.357	−1.672
CISD/cc-pVTZ	−0.825	−0.928
QCISD/cc-pVTZ	−1.170	−1.384
MP2/cc-pVQZ	−1.302	−1.652
MP2/cc-pV5Z	−1.241	−1.614
MP2/aug-cc-pVTZ	−1.793	−2.329
MP2/aug-cc-pVQZ	−1.683	−2.071
PW91/cc-pVTZ	−2.943	−3.856

$$E_R = Ae^{-BR} - CR^{-6} \tag{4}$$

$$E_R = \varepsilon\left[\left(\frac{R_o}{R}\right)^{12} - 2\left(\frac{R_o}{R}\right)^{6}\right] \tag{5}$$

Interestingly, it proved possible to fit this data with slightly better accuracy ($R^2 = 0.997$, RMS = 0.055) using only the closest $F\cdots F$ contact distance. This could mean that this close contact dominates the interaction, and hence that just one distance must be considered in describing fluorous interactions, for example in simulation of solvation in fluorous environments. The main source of error in this fit was found for the longer $F\cdots F$ distances where the potential energy curve tails off toward zero. In this region the Buckingham fit slightly underestimated stabilisation energy and returned to zero too quickly.

As well as general van der Waals interactions, F_2 also has a permanent quadrupole moment, which will give rise to electrostatic forces between molecules. To account for this, in a qualitative sense at least, we have added an extra term to

Table 7
Optimised Buckingham and Lennard-Jones parameters for $(F_2)_2$

Buckingham	Lennard-Jones
$A = 190369$	$\varepsilon = 0.80$
$B = 3.912$	$R_0 = 2.89$
$C = -3369.6$	
$D = 1053.9$	

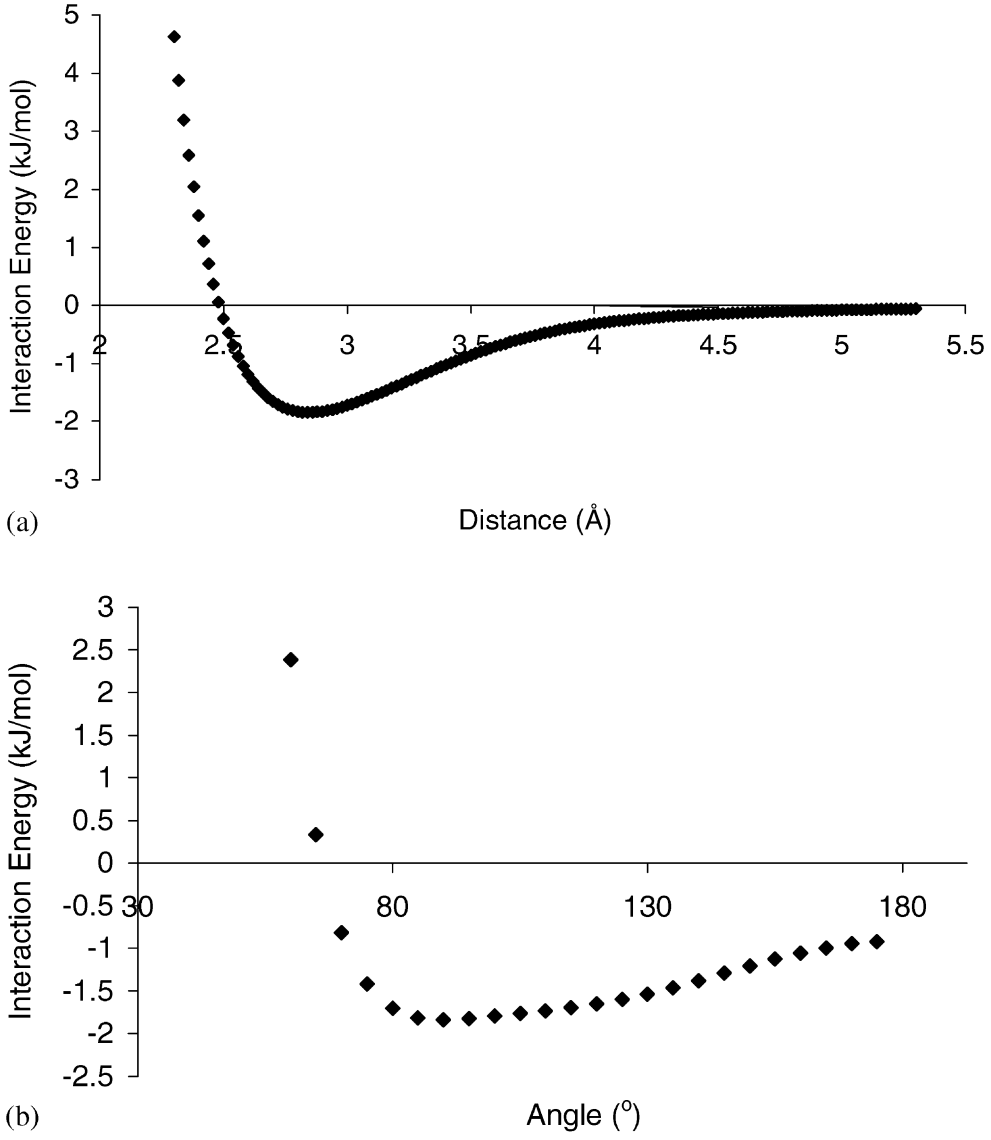

Fig. 3. PES scan for L-shaped $(F_2)_2$: (a) varying *F–F* distance; (b) varying angle between *F–F* vectors.

the Buckingham potential. Since the interaction of two quadrupoles varies with $1/R^5$, we have simply given this term the form D/R^5 and added the coefficient D to those optimised in the non-linear fitting. The overall fit is improved to an impressive $R^2 = 0.999$ and the RMS error halved to 0.024 kJ/mol, which is just 1.3% of the maximum well depth. It is apparent that quadrupole–quadrupole as well as dispersion interactions form a significant part of the fluorine–fluorine interaction.

The rather flat nature of the PES of $(F_2)_2$ around the global minimum is demonstrated by Fig. 3b, in which the angle between the two F_2 molecules is varied from 45 to 180° (90° corresponds to the L-shaped minimum). The dimer is stable across a wide range of angles, and only becomes unstable when this angle drops below 70°, where $F \cdots F$ contacts become very close and the energy rises steeply. The linear conformation at 180° shows some stabilisation, despite being characterised as a transition state by harmonic frequency calculation.

Finally, we have applied the same MP2/cc-pVTZ method to the $(CF_4)_2$ dimer previously studied by Parra and Zeng [16]. The size of this system prevented us from testing other methods, as in Table 6, or from studying the entire PES—instead, we have concentrated on the D_{3d} minimum energy conformation. We calculate the maximum stabilisation of $(CF_4)_2$ to be 2.80 kJ/mol at a $C \cdots C$ separation of 3.63 Å, considerably stronger and shorter than previous results (i.e. 2.12 kJ/mol at 4.20 Å) with smaller basis sets suggest. Thus, the Buckingham and Lennard-Jones parameters developed by Parra and Zeng probably underestimate the binding of fluorous molecules—this issue will be addressed in subsequent work.

Thus, it is evident that there are small but important stabilising interactions between fluorines in these model systems, which are apparently responsible for the preference of fluorous molecules for fluorous solvents. Attractive intermolecular interactions which act in a stabilising capacity are extremely low in fluorous solvents since they arise merely from fluorine–fluorine dispersion forces and quadrupolar interactions. This is manifested by the fact that such solvent systems possess negative values for the Abraham descriptors ***E*** and ***S*** as well as having amongst the lowest Hildebrand solubility values. It has been demonstrated here that the magnitude of the small stabilising fluorine–fluorine interactions is distance and orientation dependent, a fact which can be of use in the purposeful design of novel fluorophilic entities.

3. Conclusions

Accurate predictive models using LFER have been used to model the organic/fluorous partition of 91 organic compounds, with an error of prediction less than three times the estimated experimental error. The partition can be reduced to a sum of solvent–solute interaction terms and furthermore, the physicochemical factors which affect a molecule's fluorophilicity are identified. Important properties in determining fluorophilicity are the molecule's overall fluorine content, its ability to participate in dispersion interactions, and its hydrogen bond acidity. Perhaps more importantly, it is shown that hydrogen bond basicity and molecular polarity are much less important in determining fluorophilicity, contrary to some previous models.

Ab initio methods have been used to study different conformations of $(F_2)_2$, with an L-shaped structure found to be the most stable. Thus, the $F_2 \cdots F_2$ interaction is orientation dependent and this may facilitate the design of more complex fluorophilic molecules such as $[F(CF_2)_n]_x$. Correlation consistent basis sets were used since split valence basis sets were found to be unreliable, while higher levels of theory give similar results to MP2. The PES around the L-shaped conformer was calculated at the MP2/cc-pVTZ level, and stabilisation energies used to fit parameters for Buckingham and Lennard-Jones potential energy functions. The addition of a quadrupole term to the Buckingham function yielded an excellent fit. The quadrupole and dispersion interactions are both significant in the fluorine–fluorine interaction.

4. Computational methods

A total of 98 data was collected from Gladysz's online database [20], and transformed into ln P as set out in Eq. (1). These molecules and their partition coefficients are reported in Table 2: full details, including calculated descriptors and SMILES strings (see below) have been deposited as supporting information. structures for this set of 98 molecules were encoded as SMILES strings [21], which were then passed to our fragmental descriptor calculation program [22]. This program breaks down a molecule into its constituent fragments, and interactions between fragments, and sums contributions from these fragments to each of the molecular descriptors ***E***, ***S***, ***A***, and ***B*** (***V*** is calculated exactly according to McGowan's formula [15]). A sixth descriptor, ***F***, the fluorine content by percentage of molecular weight was added, also calculated from SMILES strings. Considerably more data was available from this source, including many transition metal and tin compounds. However, the fragment method used to calculate descriptors was not trained for such species, and cannot reliably predict values for them. We have therefore omitted all such species from the current analysis.

Once descriptors were obtained for all 98 molecules, these were used as independent variables in MLR against the observed ln P values using the JMP statistical package [23]. The significance or otherwise of each descriptor in the resulting model of ln P was determined using the Student's t-test. The overall quality of fit was measured using the following statistical quantities: R^2, the square of the correlation coefficient, or the percentage of the overall variance explained by the model; S.D., the standard deviation between observed and calculated data; R^2_{CV}, the cross-validated or 'leave-one-out' R^2 value, calculated by omitting each data point in turn, and predicting a value from the resulting model, and F, Fischer's significance test. The original LFER descriptors are defined such that they all lie on a similar scale, so direct comparison between coefficients is valid. However, ***F*** does not lie on such a scale and cannot be directly compared to the other coefficients: instead, we use t-ratios to compare the relative significance of descriptors in models of ln P.

All *ab initio* calculations were performed using GAUSSIAN98 [24]. Geometries of $(F_2)_2$ dimers were optimized using MP2 [25] and DFT [26] theory and different basis sets. Harmonic frequency calculations were performed on the resulting structures to check whether they were minima or not. Higher level single-point calculations were also carried out at the MP3, MP4, CCSD, CCSD, CCSD(T), CISD and QCISD levels with a correlation consistent triple zeta basis set to compare stabilization energies of the chosen conformers. A total of five different conformations of $(F_2)_2$ were considered, and denoted linear, skew, cross, L- and T-shaped: appropriate point group symmetries were applied throughout. The effect of BSSE, the error in intermolecular complex calculations due to the use of finite basis sets, was estimated by the counterpoise procedure [27]. The ZPE was obtained from harmonic frequency calculations.

The PES around the global minimum L-shaped dimer was explored by varying the closest contact $F \cdots F$ distance from 2.3 to 5.3 Å, and the F–$F \cdots F$ angle from 45 to 180°. The stabilisation energies obtained from these PES scans were then used to fit Buckingham and Lennard-Jones potential

energy functions. This employed the non-linear fitting functions of the JMP statistical package.

Acknowledgements

The authors are grateful to Sirius Analytical Instruments and EPSRC for studentships (FTTH and RAS, respectively), and to the UKCCF for use of their central Columbus computing facility.

References

[1] I.T. Horváth, Acc. Chem. Res. 31 (1998) 641.
[2] E. de Wolf, G. van Kosten, B.-J. Deelman, Chem. Soc. Rev. 28 (1999) 37.
[3] L.P. Barthel-Rosa, J.A. Gladysz, Coord. Chem. Rev. 192 (1999) 587.
[4] E.G. Hope, A.M. Stuart, J. Fluor Chem. 100 (1999) 75.
[5] D. Rutherford, J.J.J. Juliette, C. Rocaboy, I.T. Horváth, J.A. Gladysz, Catal. Today 42 (1998) 381.
[6] I.T. Horvath, J. Rábai, Science 266 (1994) 72.
[7] P. Bhattacharyya, B. Croxtall, J. Fawcett, J. Fawcett, D. Gudmunsen, E.G. Hope, R.D.W. Kemmitt, D.R. Paige, D.R. Russell, A.M. Stuart, D.R.W. Wood, J. Fluorine Chem. 101 (2000) 247.
[8] L.E. Kiss, J. Rábai, L. Varga, I. Kövesdi, Synletters (1998) 1243.
[9] C. Rocaboy, D. Rutherford, B.L. Bennett, J.A. Gladysz, J. Phys. Org. Chem. 13 (2000) 596.
[10] L.E. Kiss, I. Kövesdi, J. Rábai, J. Fluorine Chem. 108 (2001) 95.
[11] (a) M.H. Abraham, H.S. Chadha, F. Martins, R.C. Mitchell, M.W. Bradbury, J.A. Gratton, Pesticide Sci. 55 (1999) 78;
(b) M.H. Abraham, Chem. Soc. Rev. 22 (1993) 73.
[12] I. Gutman, L. Popovic, E. Estrada, S.H. Bertz, ACH Models Chem. 135 (1998) 147.
[13] M.H. Abraham, G.S. Whiting, R.M. Doherty, W.J. Shuely, J. Chromatogr. 587 (1991) 213.
[14] (a) M.H. Abraham, P.L. Grellier, D.V. Prior, P.P. Duce, J.J. Morris, P.J. Taylor, J. Chem. Soc., Perkin Trans. 2 (1989) 699;
(b) M.H. Abraham, P.L. Grellier, D.V. Prior, J.J. Morris, P.J. Taylor, J. Chem. Soc., Perkin Trans. 2 (1990) 521.
[15] M.H. Abraham, J.C. McGowan, Chromatographia 23 (1987) 243.
[16] R.D. Parra, X.C. Zeng, J. Mol. Struct. Theochem. 503 (2000) 213.
[17] O. Couronne, Y. Ellinger, Chem. Phys. Lett. 306 (1999) 71.
[18] K.W. Chan, T.D. Power, J. Jai-nhuknan, S.M. Cybulski, J. Chem. Phys. 110 (1999) 860.
[19] D.E. Woon, T.H. Dunning, J. Chem. Phys. 103 (1995) 4572.
[20] http://www.organik.uni-erlangen.de/gladysz/research/partition.html.
[21] D. Weininger, J. Chem. Inf. Comput. Sci. 28 (1988) 31.
[22] JMP Rev. 4, SAS Institute Inc., Cary, NC, 2000.
[23] J.A. Platts, D. Butina, M.H. Abraham, A. Hersey, J. Chem. Inf. Comput. Sci. 38 (1999) 835.
[24] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, GAUSSIAN98 Rev. A.6, Gaussian Inc., Pittsburgh, PA, 1998.
[25] C. Møller, M.S. Plesset, Phys. Rev. 46 (1934) 618.
[26] (a) P. Hohenberg, W. Kohn, Phys. Rev. 136 (1964) B864;
(b) R.G. Parr, W. Yang, Density Functional Theory, Oxford University Press, Oxford, 1989.
[27] F.B. van Duijneveldt, J.G.C.M. van Duijneveldt-van de Rijdt, J.H. van Lenthe, Chem. Rev. 94 (1994) 1873.
[28] M.H. Abraham, J.A. Platts, J. Org. Chem. 66 (2001) 3484.
[29] (a) R. Ditchfield, W.J. Hehre, J.A. Pople, J. Chem. Phys. 54 (1971) 724;
(b) M.J.S. Dewar, C.H. Reynolds, J. Comput. Chem. 2 (1986) 140.
[30] (a) A.D. Becke, J. Chem. Phys. 98 (1993) 5648;
(b) C. Lee, W. Yang, R.G. Parr, Phys. Rev. B 37 (1988) 785.
[31] J.P. Perdew, K. Burke, Y. Wang, Phys. Rev. B 54 (1996) 16533.