

Solvation Parameters of Solutes: An Attempt of Improvements in Characterization and Determination

P. Laffort* and P. Héricourt

CNRS, Centre des Sciences du Goût UMR 5170, 15 rue Hugues Picardet, F.21000 Dijon, France

Abstract: We are presenting here an updated version of our predictive model of solvation parameters of solutes, based on a simplified molecular topology (J. Chem. Inf. Model, 2006, 46, 1723-1734). The learning experimental database of this model is presently larger than in our first version and includes more compounds with more than one functional group of the same type. This experimental database is also surer, in the sense that it only includes compounds in liquid state at room temperature, when the polarizability of a compound in solution is established *via* its refractive index in gas or crystalline state. Indeed, we demonstrate in this paper, a bias in this case.

INTRODUCTION

The conceptual definition of solvation parameters or descriptors can be expressed as follows:

If a matrix SP of a given solubility property can be expressed as a product of matrices $A*B$, then A and B are respectively matrices of solute and solvent solvation parameters. The first tool needed in order to characterize the solvation parameters of solutes is a solid database SP of a solubility property. Recently, Laffort *et al.* [1] used a very accurate matrix of retention indices in GLC (gas-liquid chromatography) for 133 solutes and 10 stationary phases, by Kováts and co-authors [2-6].

The second tool needed is a suitable statistical analysis. Laffort *et al.* [1] applied the original MMA algorithm (as Multiplicative Matrix Analysis) that the principle is reproduced in Fig. (1).

The first interest of this tool, compared with more classical ones such as the step-wise multi linear regression analysis (MLRA), is that it provides a test for each of the elements of the matrix, independently of their weight in the given sample of experimental data. A second interest of the MMA algorithm is that the output parameters values should be more precise than the input parameters values, on the condition that experimental data are overabundant and precise enough. The MMA algorithm is free available on <http://paul.laffort.free.fr>.

There is an agreement between the authors presently involved in solvation parameters, to consider that five solute parameters and five solvent parameters are needed and sufficient to take into account the solubility phenomena. The nature of one of the five solute parameters is related to the molar volume (it is called APOLAR), whereas the four other ones are independent of the molar volume. These last four parameters are respectively named of ORIENTATION or

polarity, POLARIZABILITY or induction, ACIDITY (proton donor) and BASICITY (proton acceptor).

On this basis Laffort *et al.* [1] tested, among all the published values, only those concerning five solute parameters. In addition to the already mentioned good correlation between INPUT and OUPUT values using the MMA algorithm, two additional criteria have been considered by these authors: i) a good independence of the solute parameters (poor mutual correlation) and ii) an OUTPUT set of solvent parameters without negative values, difficult to understand in physicochemical terms.

From these compared tests, the published values according to Abraham [7] (314 solutes) appear as the most suitable, after an internal rearrangement of the original values for the apolar and the orientation parameters, *via* two simple equations. It also results from this enquiry a set of updated values of solvation parameters for 133 solutes, established using five out of the ten stationary phases studied by Kováts and co-authors [2-6]. Because these phases have been synthesized in limited quantity and are not commercially available, the method could be extended to any set of columns containing two apolar phases of very different molecular weight, a strongly fluorinated, a classical polyether and an alcoholic (e.g. diglycerol), after learning the data set for the above mentioned 133 solutes.

These two subsets of solute solvation parameters (314 + 133), having a total of 369 defined compounds, have been used as a learning material by Laffort and Héricourt [8] using a simplified molecular topology (SMT) which principally takes into account, for each atom of a molecule, its nature, the nature of its bonds and in some cases the nature of its first neighbors.

The stage presently reached could be improved on various aspects. This is the purpose of the present study:

- The apolar term established by Laffort *et al.* [1] using the partition coefficient air-hexadecane (L16) and two other polar terms according to Abraham [7], has been identified as the molar polarizability. We therefore call it a DISPERSION parameter. However, number of data published by Abraham and co-authors do not

*Address correspondence to this author at the CNRS, Centre des Sciences du Goût UMR 5170, 15 rue Hugues Picardet, F.21000 Dijon, France;
E-mail: laffort@cesg.cnrs.fr

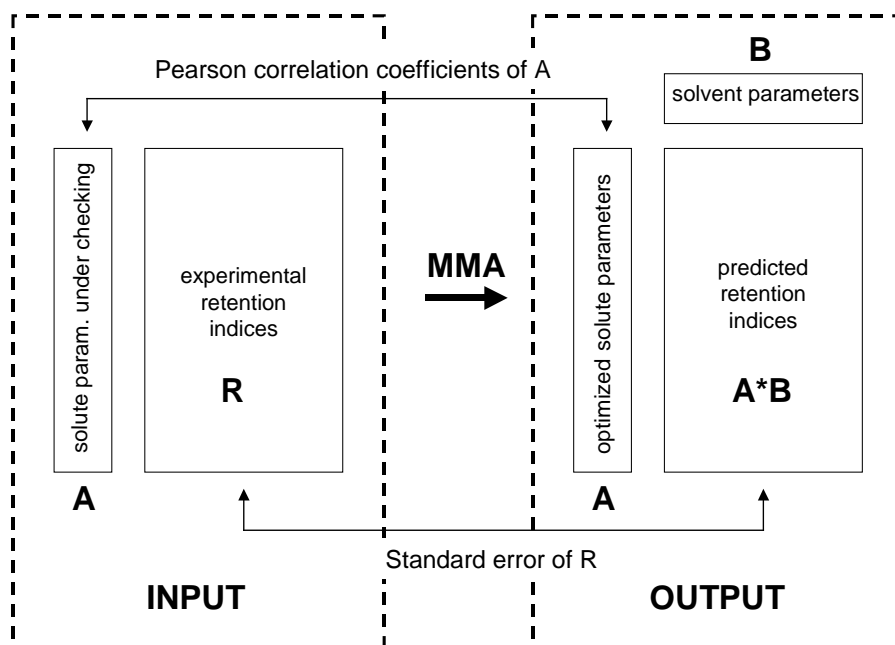


Fig. (1). Diagram of the INPUT/OUTPUT of the Multiplicative Matrix Analysis (MMA), according to Laffort *et al.* [1]. The correlations between input and output parameters A depend on the nature of the input parameters, for a given experimental matrix R, whereas the standard error on the reconstruction of matrix R only depends on the chosen number of parameters. The MMA algorithm can be considered as a tool to test theories (the input).

include L16 (or L in a more recent expression) but V_x (or V) which is the molar volume according to Abraham and McGowan [9]. The first purpose is therefore to check if the dispersion parameters values derived from L16 and V_x are equivalent or not.

- The pool of 369 solutes used to calibrate the SMT algorithm by Laffort and Héricourt [8] does not include too much compounds with more than one functional groups. One of the purposes of the present study is to extend in this way the learning material.
- Once the two previous points overcome, the third purpose of the present study is to check – and if possible to improve – the SMT algorithm suggested by Laffort and Héricourt [8], in order to determine new values on the basis of a larger experimental data set.

MATERIALS AND METHODS

Statistical Tools

In addition to the EXCEL Windows facilities for drawing diagrams and handling data sets, the SYSTAT[®] 10.2 for Windows has been applied for stepwise MLRA (Multidimensional Linear Regression Analysis).

Molecular Topology Applied in the Present Study

Its principle is the same as in our first version [8]. It takes into account, for each atom of a molecule, its nature, the nature of its bonds, and in some cases the nature of its first neighbors. In other words, the applied strategy is a compromise between the assignments of fixed atomic increments, and the consideration of all first and second neighbors of each structural element, as used in more sophisticated topological approaches.

The topological features retained are also very similar. Each atom is provided with an index, comprising a series of digits. Their sum is at most equal to its valence. The value of the digits define the type of bonds (1 for a single, 2 for a double bond, etc.), but the bonds with hydrogen are not indicated. In addition, the immediate neighboring is in part considered for oxygen, nitrogen and halogens. So, the possibilities for oxygen, for example, are the following: O0, O1, O11, O2, with six subcategories for O1 (linked to C1, C11, C111, C1111, C12 and C112) and three subcategories for O2 (linked to C12, C112 and other cases).

In addition to the 44 atom characteristics and their immediate environment finally kept in the present study, we also consider two additional topological features trying to account for spatial proximities between proton donors and proton acceptors existing in some ortho derivatives of phenols and in some amides, which we respectively call POSPA and NCO. We also consider a connectivity parameter due to Zamora [10] called the “smallest set of smallest rings” (SSSR), and applied in one of our previous studies [11]. According to this concept, for the naphthalene for example, which contains two individual C-6 rings and one C-10 ring embracing them, only the two six numbered rings are considered. Two six numbered rings corresponding to 12 carbon atoms, the SSSR value of naphthalene is therefore be taken equal to 12. The definitions of these 47 features are summarized in Table 1. It should be noted that they are limited to not ionized compounds.

Molecular topological features are generated using the MarvinSketch program and other Java functionalities of ChemAxon Ltd [12]. In a first step the molecules under study are drawn using the MarvinSketch program. In a second step, the Java library of ChemAxon allows to scroll

Table 1. Nature of Atomic and Structural Elements Used in the Molecular Topology of the Present Study (Not Ionized Substances). See Explanations in the Text. Highlighted Elements are not Involved in the Present Study but are Often Present in Organic Chemistry

Structural Element	Bonds	Topological Features	Subcategories
BASIC			
Carbon	≤ 4	C0 , C1, C11, C111, C1111, C2, C12, C112, C22, C3, C13	
Oxygen	≤ 2	O0, O11	
Oxygen	≤ 2	O1	linked to C1, C11, C111, C1111, C12, C112
Oxygen	≤ 2	O2	linked to C12, C112, others
Nitrogen trivalent	≤ 3	N0	
Nitrogen trivalent	≤ 3	N1	linked to C12, C112, others
Nitrogen trivalent	≤ 3	N11	linked to C12, others
Nitrogen trivalent	≤ 3	N111, N12, N3	
Nitrogen pentavalent	≤ 5	N122	
Phosphor pentavalent	≤ 5	P1112	
Fluorine	= 1	F1	linked to C1111, others
Chlorine	= 1	Cl1	linked to C111, C1111, others
Bromine	= 1	Br1	
Iodine	= 1	I1	
Sulfur divalent	≤ 2	S0 , S1, S11, S2	
Sulfur hexavalent	≤ 6	S1122	
Silicon	≤ 4	Si1111	
Tin	≤ 4	Sn1111	
Hydrogen	= 1	H1 = sum (maximal bonds - explicit bonds)	
ADDITIONAL			
POSPA		Phenol O rtho Substituted with P roton Acceptor (N122, O11, Cl, Br, I, CO)	
NCO		N 1 or N11 or N111 linked to (a Carbon linked to O 2)	
SSSR		Smallest Sum of the S mallest R ings	

through the molecular structures generated in the first step, in order to establish the needed structural parameters.

Learning Experimental Data Set

The data bank of solute solvation parameters permitting to establish the rules involving the retained topological features, results from the pooling of various experimental published data sets which will be specified in the Results section.

RESULTS

Recovery of the Published Parameters According to Abraham and Co-Authors

The recovery of the parameters already published by Abraham and colleagues for a new definition of the dispersion parameter was limited, until now, to the solutes for which these authors published the partition coefficient air-hexadecane L16. The applied equation was as follows, δ standing for the dispersion parameter established using the solute parameters of Abraham [7]:

$$\delta = \log L16 - 0.532\pi_2^H - 0.894R_2 - 0.115 \quad (1)$$

in which π_2^H (or S) and R_2 (or E) are respectively the parameters of orientation and polarizability according to Abraham and co-authors. For the publications in which only values of the molar volume V_x according to the definition of Abraham and McGowan [9] are provided, in place of L16, the following equation is suggested:

$$\delta' = 1.238R_2 + 3.506 \frac{V_x}{100} - 0.875 \quad (2)$$

The remaining question was to check if these two definitions are equivalent or not on the basis of experimental published data. In this purpose, we have compared the dispersion parameters expressed using eqs. 1 and 2, for two data sets: i) 314 solutes studied by Abraham [7] for which L16 values are provided and for which we have established the corresponding V_x values using the rules of Abraham and McGowan [9]; ii) 93 solutes studied by Abraham and Acree

[13] for which both $L16$ and Vx are provided. The results are summarized in Table 2 and Fig. (2).

The observed scattering of the points concerning solids, in Fig. (2), could be explained by the values of the molecular refractive index in crystalline form, which can differ from their values in solutions (the molecular refractive indices are strongly involved in the calculation of the $R2$ parameter, and therefore in the calculation of δ_2 and δ_2' using eqs.1 and 2). The observed scattering for gases, slightly less dramatic, could be due to a lower experimental accuracy in the determination of the experimental data.

The principal conclusion of these results in the present study, concerns the building of a learning set of solubility parameters in order to calibrate the SMT (Simplified Molecular Topology). Only the data in liquid state at room temperature will be kept when a molecular refractive index is applied to establish the $R2$ parameter.

Building an Experimental Learning Data Set

As in our previous study [8], a homogeneous set of solutes solvation parameters has been established pooling various published data sets for which the consistency has been verified for the overlapping compounds:

Sub-Set A for 128 Solutes

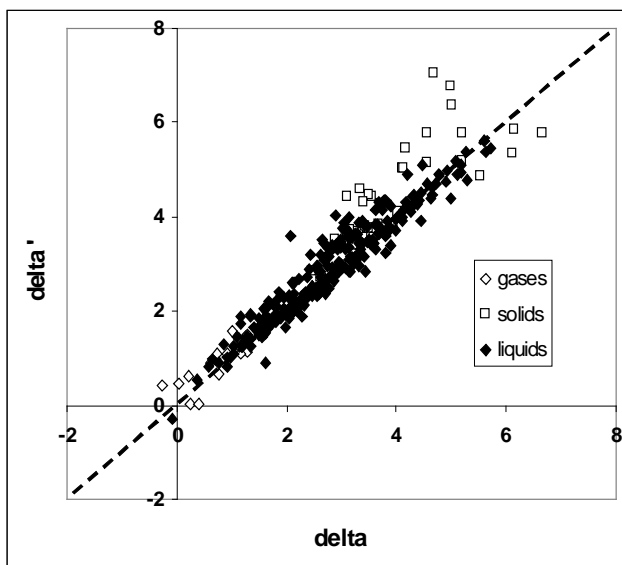
The values have been derived by Laffort *et al.* [1] from retention indices established by Kováts and co-authors [2-6] on five selected stationary phases in gas-liquid chromatography. From the data concerning 133 solutes, we discarded, for a better consistency, five solutes using retention chromatographic indices having a value < 500 (i.e. lower than that of pentane). This first sub-set is identical to that already applied in our previous topological study [8].

Sub-Set B for 185 Solutes

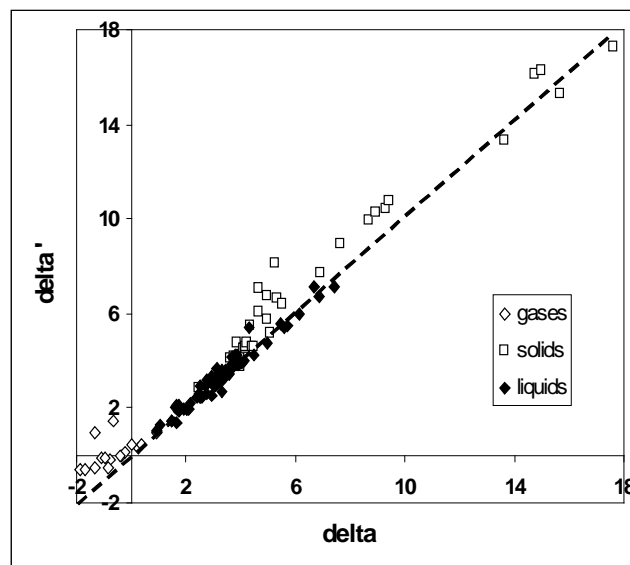
The values are derived from the solvation parameters published by Abraham [7] for 314 compounds for which log

Table 2. Comparison of the Dispersion Parameters d and d' According to Eqs.1 and 2, for 314 Solutes Studied by Abraham [7] and for 93 Solutes Studied by Abraham and Acree [13]. The Equivalence Between Both Expressions is Principally Observed for Liquids at Room Temperature (N Stands for the Number of Solutes, r for the Pearson Correlation Coefficient and SEE for the Standard Error of Estimate)

	GASES (BP < 20°C)	LIQUIDS (MP < 25°C; BP > 25°C)	SOLIDS (MP > 25°C)
Abraham, 1993	N = 17 $r = 0.81$ SEE = 0.29	N = 253 $r = 0.96$ SEE = 0.30	N = 44 $r = 0.82$ SEE = 0.69
Abraham and Acree, 2004	N = 18 $r = 0.75$ SEE = 0.53	N = 57 $r = 0.97$ SEE = 0.29	N = 42 $r = 0.98$ SEE = 0.75



Abraham, 1993



Abraham and Acree, 2004

Fig. (2). Correlogram for the same data as in Table 1. Most of the outliers concern the solutes in solid or gas state at room temperature.

L16 is provided, according to eq. 1 already mentioned and eqs.3:

$$\begin{aligned}
 \text{ORIENT} \quad \omega &= 1.523\pi_2^H - 0.538\Sigma\beta_2^H - 0.837R_2 \\
 \text{POLARIZ} \quad \varepsilon &= R_2 \\
 \text{ACID} \quad \alpha &= 2.825\Sigma\alpha_2^H \\
 \text{BASIC} \quad \beta &= 0.728\Sigma\beta_2^H
 \end{aligned} \quad (3)$$

in which π_2^H , R_2 , $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ are respectively the parameters of orientation, polarizability, acidity and basicity, according to Abraham and co-authors.

Out of the original 314 compounds, have been discarded those in solid or gas state at room temperature and also those already present in the sub-set A.

Sub-Set C for 143 Solutes

Zissimos *et al.* [14] have reported solute solvation parameters according to the Abraham presentation for little less than 470 compounds (some are duplicated). After discarding those in solid or gas state at room temperature and also those already present in the sub-sets A and B, data for 143 new substances are kept, using the rules of transformation given by eq. 2 for the dispersion parameter and eqs.3 for the four other ones.

The homogeneous learning set finally obtained is more satisfactory than that used in our previous study [8]:

- It includes only liquids at room temperature when the polarizability-induction parameter is obtained *via* the refractive index.
- It includes more compounds (456 instead of 369).
- It includes more compounds with more than one functional group (79 instead of 62).

This learning set is reported in Table SI (Supporting Information).

Polar Surface Areas

In addition to the solvation parameters of solutes strictly speaking, we also reported in Table SI the values of polar surface areas (PSA) according to Ertl [15] and Ertl *et al.* [16]. The polar surface area has been defined by Palm *et al.* [17], as “the area occupied by nitrogen and oxygen atoms, and hydrogen atoms attached to these heteroatoms”. Ertl *et al.* [16] also take into consideration slightly polar heteroatoms as sulfur and phosphor, but no one author has included the contributions of halogens, strongly polar heteroatoms. The PSA concept is quite fascinating on various aspects:

- Its definition is chemically very simple and precise, but its justification is pharmacological (reflecting very well the molecular transport properties, particularly intestinal absorption and blood-brain barrier penetration). The discarding of the halogens contribu-

tion is due to their absence of contribution to the molecular transport properties.

- PSA values have been established using sophisticated programs, taking into account the molecular three-dimensional shape and flexibility [17].
- A topological method using summation of surface contributions of polar fragments (termed TPSA) has been applied by Ertl *et al.* [16], exhibiting an excellent correlation with theoretical PSA values ($r = 0.991$, $N = 34\,810$ substances).

Because the present study is not only dedicated to the solubility phenomena, but also in a future perspective of QSAR (Quantitative Structure Activity Relationship), the TPSA values according to Ertl *et al.* [16] have been included in Table SI, in order to also check our SMT procedure for this molecular polar property.

Learning the Experimental Data Set and Calibration of the SMT (Simplified Molecular Topology)

A step wise multi linear regression analysis has been applied to the 456 compounds listed in Table SI, each of the five solubility parameters and the polar surface area (PSA) being successively the dependent variables, and the 47 molecular features listed in Table 1 the independent variables (taken alone or sometimes grouped: e.g., O0+O1). The first used rule has been to have not more than 14 final independent variables in each regression (which represent 3% of the total observations). The second rule has been that each partial F ratio associated with the independent variables finally kept has to be at least equal to 25 (the partial F ratios reflect the degree of involvement of each feature). The selected molecular features permitting to predict the five solubility parameters and the PSA, their respective coefficients and their partial F ratios are listed in Table 3. The corresponding correlograms are drawn in Fig. (3).

Before entering a general discussion, the results as exhibited in Fig. (3) suggest two principal comments:

Polar Surface Areas

The theoretical values of PSA being unavailable for the 456 compounds of Table SI, we have compared two topological approaches, one of them (Ertl *et al.* [16]) supposed to be very close to the theoretical values. The two topological methods appear equivalent, except for eight slightly outliers which are all 5-ring aromatic molecules (pyrrole, furfural, furan, methylfuran, benzofuran, thiophene, 3-methylthiophene, thiazole). The explanation is very clear: in the Ertl *et al.* [16] procedure, the heteroatoms of these compounds are linked to carbon atoms *via* aromatic bonds, whereas in the present work *via* single bonds. This problem does not occur with 6-ring aromatic substances, since in our approach we have one single and one double bond, which is equivalent to two aromatic bonds. It would be interesting to know the theoretical values for these eight 5-ring compounds, in order to introduce or not an additional feature of aromatic bond in our model. It is of interest to underline that in the Ertl *et al.* [16] procedure are involved 43 fragments (32 not ionized and 11 ionized) and in the present study only 14 non ionized molecular features.

Table 3. Molecular Features and their Corresponding Coefficients Involved in a Simplified Topology Defining the Five Solvation Parameters of Dispersion, Polarizability/Induction, Acidity, Basicity and Orientation, as Well as the Polar Surface Area (See Text) for the 456 Compounds Reported in Table SI*

Features	Coefficients	Partial F Ratios
Constant	-0.612	
C total	0.413	3818
O -(O0+O 1)	0.250	339
O0+O1	0.449	194
N-N1	0.420	209
N1	0.698	117
C11	0.628	907
Br1	0.942	929
I1	1.264	373
S1+S11+S2	0.958	491
Si1111	0.492	38
Sn1111	1.204	29
H1	0.044	183
DISPERSION		

Features	Coefficients	Partial F Ratios
C2+C12+C13	0.073	636
O1	0.180	135
O2	0.096	68
N1	0.250	67
F1	-0.084	147
C11	0.149	280
Br1	0.315	546
I1	0.657	503
S1	0.365	119
S11	0.299	115
S2	0.426	86
Si1111	-0.420	152
SSSR	0.053	758
POLARIZABILITY		

Features	Coefficients	Partial F Ratios
C12	0.037	37
O11	0.115	30
O0+O1	0.167	26
O2	0.503	769
N3	0.944	286
F1	0.133	92
C11 nlk C1111	0.173	57
Br1	0.146	27
NCO	0.405	41
ORIENTATION		

Features	Coefficients	Partial F Ratios
C112	0.047	65
(O0+O1) nlk C112	0.344	448
O11	0.131	147
O2	0.213	291
N1	0.452	229
N11 nlk C112	0.416	134
N111	0.360	132
N12	0.381	330
N3	0.267	74
N122	-0.250	38
BASICITY		

Features	Coefficients	Partial F Ratios
C13	0.139	44
O0	2.317	344
sqrt [O1 x (C1+C11)]	1.091	1773
O1 x C111	0.932	677
O1 x C1111	0.655	80
O1 x C12	2.034	265
O1 x C112	1.705	2503
N1 x C12	1.751	196
N11 x C12	1.144	166
(N1+N11) x C112	0.631	74
sqrt [(N1+N11) nlk C12, nlk C112]	0.314	92
F x C1111	0.121	72
Clx C111	0.139	37
POSPA	-0.795	151
ACIDITY		

Features	Coefficients	Partial f Ratios
C22	31.715	8871
O0	29.270	3839
O1	20.227	95500
O11	9.386	32650
O2	17.017	101600
N1	26.027	38170
N11	12.351	7262
N111	3.366	588
N12	13.109	19850
N122	11.789	45980
N3	23.792	28980
P1112	9.407	690
S1	38.800	71050
S11	25.873	43730
POLAR SURFACE AREA		

* The term *linked to* is marked x, and nlk stands for *not linked to*.

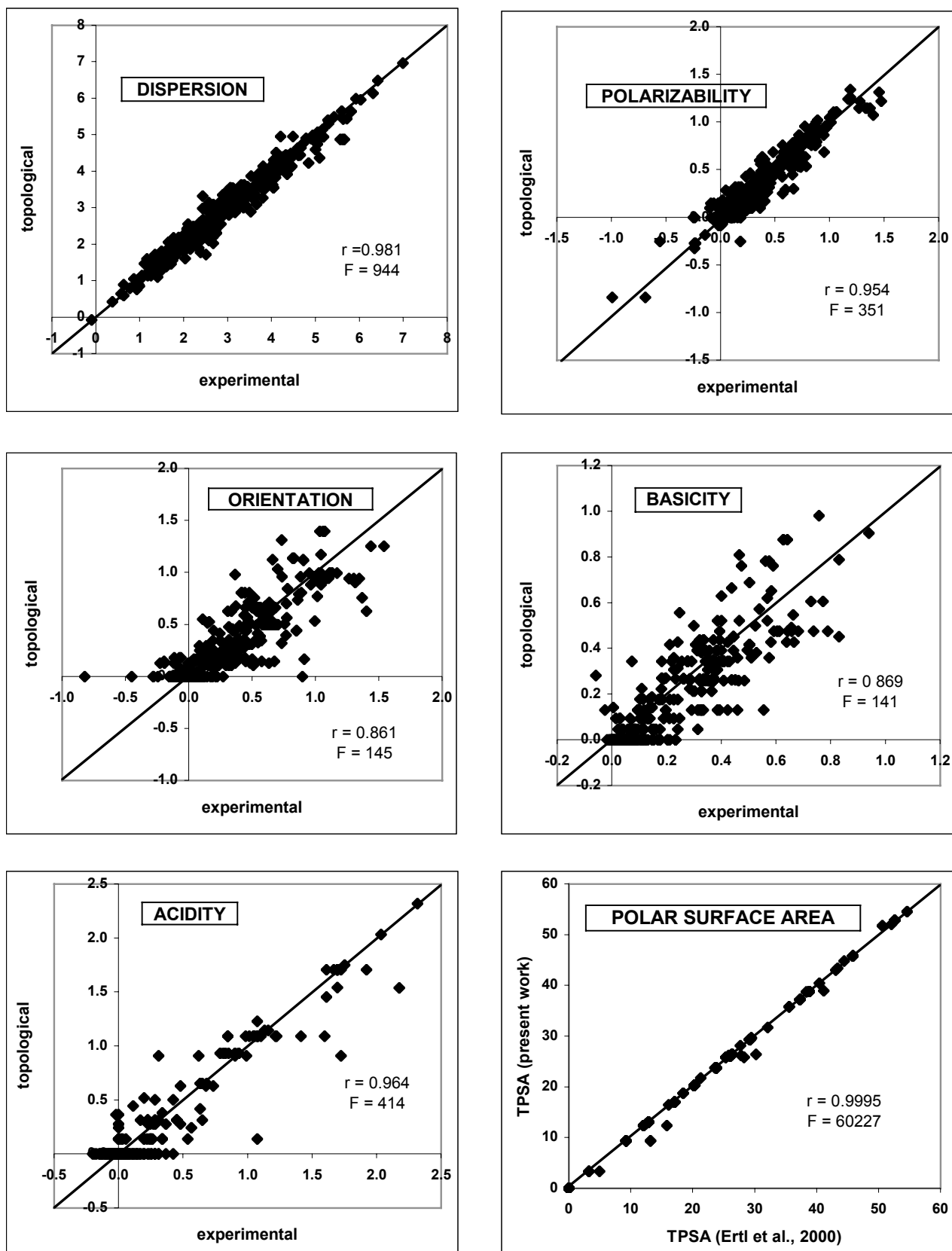


Fig. (3). Correlograms of the solvation parameters topologically defined (Table 3) versus experimental (Table SI), and the polar surface area topologically defined in Table 3 and by Ertl *et al.* [16].

Solvation Parameters Strictly Speaking

The general trends are the same as those observed in Laffort and Héricourt [8]: rather satisfactory results for the dispersion and the polarizability/induction parameters, disappointing results for the orientation and the basicity parameters, and in between for the acidity parameter. A deeper investigation is developed in the Discussion section, in order to see if the challenge of the present study announced in the title is at least partially reached or not.

GENERAL DISCUSSION

Expression of the Solutes Solvation Parameters

A controversy has been developed by Mintz *et al.* [18], on the respective advantages and disadvantages in the expression of solvation parameters by Abraham and co-authors in one side, and as we have recently proposed [1,8] on the other side. These authors agree that there are advantages associated with having an orthogonal set of solute descriptors, but that this is only observed for the data set of 369 compounds studied by Laffort and Héricourt [8]. We observe that this fact is also verified for the data set reported in Table SI, as it can be shown in Fig. (4).

ω	-0.12			
ϵ	0.14	-0.05		
α	-0.19	0.11	-0.04	
β	-0.04	0.30	0.06	0.17
	δ	ω	ϵ	α

Table SI (N = 456)
Laffort *et al.* expression

π_2^H	0.25			
R_2	0.44	0.56		
$\Sigma\alpha_2^H$	-0.17	0.10	-0.04	
$\Sigma\beta_2^H$	0.05	0.48	0.06	0.17
	$\log L_{16}$	π_2^H	R_2	$\Sigma\alpha_2^H$

Table SI (N = 456)
Abraham *et al.* expression

Fig. (4). Comparison of mutual independence (low values of correlation coefficients) in the two expressions of the solvation parameters reported in Table SI

Over statistical considerations, Abraham and co-authors consider that our expression provides solute parameters less chemically independent than their own expression, particularly the orientation/polar parameter. As an objection to this point of view, it can be recalled that our ω (orientation) parameter has been obtained using together the dipole moment and the MMA algorithm applied to experimental chromatographic data (Laffort *et al.* [1], footnote 3, page 95). The fact that the parameters appeared more orthogonal using the consequent transformation of data - as well as in the set of 314 compounds according to Abraham [7] as in the set of 125 compounds according to Kováts *et al.* [2-6] - was observed after the event.

Another objection of Mintz *et al.* [18] against our expression, is that it generates solvent parameters (called equation coefficients) with, on their point of view, unrealistic values. This problem can certainly be solved using relative values of solvent parameters to a hydrocarbon solvent with infinite number of carbon atoms as a reference. However, it is not the aim of the present study to pursue this controversy, and anyway, it is not very complicate to transform a data set expressed in one way into its expression *via* the other one.

The Learning Data Set of 456 Compounds (Table SI)

As we saw, this learning data set of solute solvation parameters includes more compounds than in our previous study

[8], and more compounds with more than one functional group. In addition, it only includes liquids at room temperature when the polarizability-induction parameter is obtained *via* the refractive index, in order to increase, as suggested by Fig. (2), the validity of pooling the three data subsets. The pooled set presents however a difficulty, in the sense that among the data published by Abraham and co-authors, not all are from experimental origin, with the risk to build prediction rules based on already predicted values. As a proof of this assertion, it can be easily verified in Table SI that the values for the acidity and basicity parameters from sources B and C, are always exactly the same along various homologous series, except the first terms. The same phenomenon is observed for the polarity parameter π_2^H (or S) in the original tables of these authors.

These absolutely identical values are not observed when the procedures applied are 100% experimental (see source A).

Therefore, Table SI can be considered as a compromise with a reasonably good amount of compounds of a reasonably diversity of types, but with data of not entirely experimental nature. Improvements in this field should very probably need more experimental data.

A Possible Improvement of Source A

Kováts and Kresz [19] have recently published updated values of the retention indices in gas-liquid chromatography previously published by Dallos *et al.* [20] and Kováts *et al.* [2-6], the later being applied in defining the source A of solvation parameters by Laffort *et al.* [1]. A corrected version of source A on these updated retention indices has not yet been achieved, but in order to refine more and more the values of solvation parameters established on 100% experimental basis, this will be done soon and published.

Comparison Between the Topological Predictive Tools Applied in 2006 and in the Present Study

As we already saw, the results observed in Fig. (3) are not very satisfactory for the basicity and the orientation/polar parameters. In order to evaluate in which proportion the challenge of the present study announced in the title is reached, we have compared the topological model previously published [8] and that of the Table 3 in the present study, to the same two data sets as reported in Table 4. These two data sets respectively are the entire Table SI and the part of this table from source A.

The results exhibited in Table 4 are interesting and surprising. For the data from Table SI as a whole (top of the table), the results are very similar using one or another group of predictive equations, with a slight advantage using the most recent ones for the acidity and the dispersion parameters. For the subset of 128 compounds of source A (bottom of the table), the two groups of equations give also very similar results, but with comparable values of correlation coefficients for the four "polar" parameters (i.e. around 0.94). Three explanations of this phenomenon can be done, but there are not presently sufficient arguments to decide in which proportion each of these three causes is involved.

- The data subset from source A could be differently distributed than the Table SI taken as a whole. It can be shown in Fig. (5) that this assertion is partially true.

Table 4. Comparisons of the Topological Model Proposed in Our Previous Study [8] and in Table 3 of the Present Publication, Respectively Applied to the 456 Compounds of Table SI and to the 128 Compounds of the Source A of the Same Table SI. See Comments in Text

	Dispersion	Orientation	Polarizability	Acidity	Basicity
N = 456 2006 model	r = 0.97 SEE = 0.26	r = 0.85 SEE = 0.19	r = 0.95 SEE = 0.10	r = 0.92 SEE = 0.18	r = 0.87 SEE = 0.11
N = 456 present model	r = 0.98 SEE = 0.22	r = 0.86 SEE = 0.18	r = 0.95 SEE = 0.10	r = 0.96 SEE = 0.12	r = 0.87 SEE = 0.09
N = 128 2006 model	r = 0.98 SEE = 0.18	r = 0.92 SEE = 0.12	r = 0.93 SEE = 0.13	r = 0.94 SEE = 0.11	r = 0.94 SEE = 0.08
N = 128 present model	r = 0.98 SEE = 0.19	r = 0.94 SEE = 0.11	r = 0.94 SEE = 0.12	r = 0.94 SEE = 0.11	r = 0.93 SEE = 0.08

- Some types of molecules, present in the Table SI taken as a whole, could not be present in the subset from source A. This assertion is true as we already saw. It can however be underlined that the equations of Table 3 have been optimized on the basis of the all Table SI.
- The experimental data from the subset A could be more precise than from the subsets B and C...

Over statistical considerations showing a slight advantage of the present version of the SMT model over the 2006 version (when applied to the 456 solutes of Table SI), we also prefer the new one for reasons of chemical consistency, as it appears on the regression equations of Table 3:

- **Dispersion parameter δ .** Univocal coefficients, for each atom type (including hydrogen), except acidic features of oxygen and nitrogen.
- **Polarizability parameter ϵ .** It depends on the polarizability contribution of each atom, of the multiple bonds and of the molecular connectivity.
- **Acidity parameter α .** It principally depends on hydroxyl groups and primary and secondary amines features, in various types of environments.

- **Basicity parameter β .** It also principally depends on oxygen and nitrogen features in various types of environments.
- **Orientation parameter ω .** Its principal features are O2 (involved in nitro compounds, ketones, aldehydes, amides, lactones, carboxylic acids, esters), N3 (nitriles) and F1 (fluoro compounds).

All these trends were present in our 2006 version, but perhaps less clearly.

Comparisons with Other Predictive Models

Klamt [21,22] has developed a novel method, called COSMO-RS technique, for predicting the thermodynamic properties of pure and mixed fluids, particularly the vapor-liquid equilibria. The thermodynamic data are established from molecular surface polar distributions, resulting from quantum chemical calculations of the individual compounds in the mixture. Zissimos *et al.* [14] have compared the five solvation parameters according to Abraham and co-authors to five COSMO-based descriptors, for 467 solutes. The results using a global MLRA (multi linear regression analysis) appeared excellent for the molar volume ($r = 0.99$), and very good for the acidity and the basicity parameters (respectively $r = 0.96$ and 0.94). They were by contrast disappointing for the polarity and

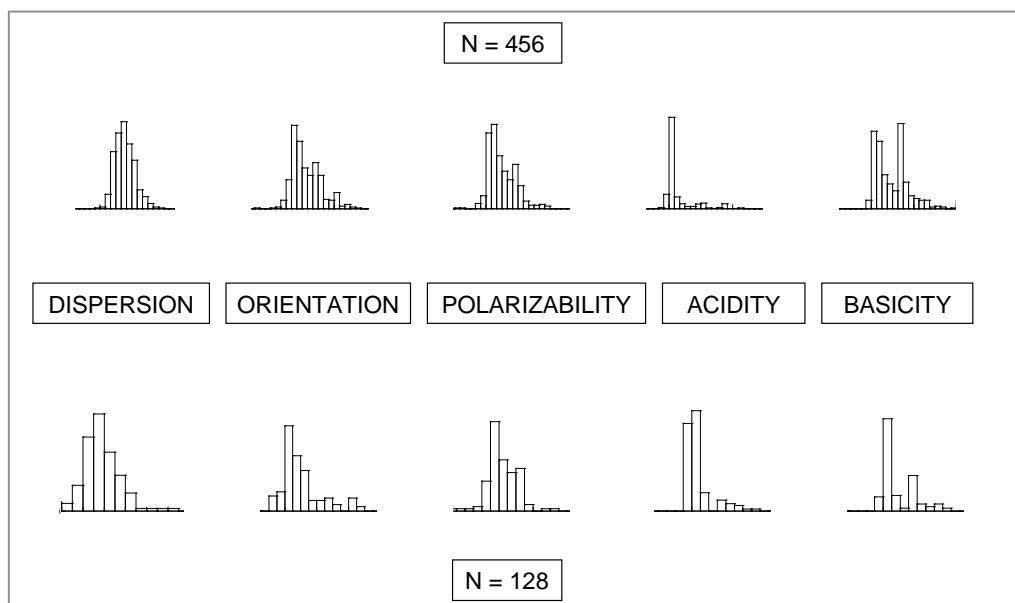


Fig. (5). Histograms of the five solvation parameters according to Table SI, for the subset of source A ($N = 128$), and for the entire table itself ($N = 456$). See comment in text.

the polarizability parameters (respectively $r = 0.88$ and 0.71). It is however difficult to compare these results to those of Table 4, the overlapping of substances under study being partial (around 2/3).

Our previous study [8] on the present field included a comparison with the results obtained by Oliferenko *et al.* [23] using a theoretical approach, and by Jover *et al.* [24] using an empirical approach of neural networks. The comparison, limited to the acidity and the basicity parameters, showed similar performances of the three approaches. Because this paper is principally devoted to an updating of the SMT procedure, we have not extended the comparison of its performances with those of other empirical [25, 26] or theoretical [27-29] models.

CONCLUSION

In spite of limited improvements in the prediction of solvation parameters as it appears in Table 4, we will apply from now on the second version of the SMT model as it is described on Table 3. Firstly, because based on a larger and surer database of experimental values. Secondly, because more chemically consistent. This updated model will soon take the place of the former SMT model in the site: <http://paul.laffort.free.fr/>

ACKNOWLEDGEMENTS

The authors warmly thank Ervin sz. Kováts (Ecole Polytechnique Fédérale de Lausanne) and Andreas Klamt (University of Regensburg, and COSMOlogic Company, Leverkusen, Germany) for fruitful discussions and critical reading. They also sincerely thank the Chemaxon Company (Budapest) for the MarvinSketch facilities.

REFERENCES

- [1] P. Laffort, F. Chauvin, A. Dallos, P. Callegari, and D. Valentin, "Solvation parameters. Part I: Mutual improvements of several approaches and determination of two first sets of optimized values," *J. Chromatogr. A*, vol. 1100, pp. 90-107, 2005.
- [2] G. Défayes, D.F. Fritz, T. Gömer, G. Huber, C. Reyff De, and E. sz. Kováts "Organic solutes in paraffin solvents. Influence of the size of the solvent molecule on solution data." *J. Chromatogr.*, vol. 500, pp. 139-184, 1990.
- [3] K.S. Reddy, J.C. Dutoit, and E. Kováts sz. "Pair-wise interactions by gas chromatography. Part I. Interaction free enthalpies of solutes with non associated primary alcohol groups." *J. Chromatogr.*, vol. 609, pp. 229-259, 1992.
- [4] K.S. Reddy, R. Cloux, and E. Kováts sz. "Pair-wise interactions by gas chromatography. Part IV. Interaction free enthalpies of solutes with trifluoromethyl-substituted alkanes." *J. Chromatogr.*, vol. 673, pp. 181-209, 1994.
- [5] G. Défayes, K.S. Reddy, A. Dallos, and E. Kováts sz. "Pair-wise interactions by gas chromatography. Part V. Interaction free enthalpies of solutes with primary chloro- and bromo-alkanes." *J. Chromatogr.*, vol. 699, pp. 131-154, 1995.
- [6] K.S. Reddy, R. Cloux, and E. Kováts sz. "Pair-wise interactions by gas chromatography. Part VI. Interaction free enthalpies of solutes with primary methoxyalkane, cyanoalkane, and alkanethiol groups." *J. Chromatogr.*, vol. 704, pp. 387-436, 1995.
- [7] M.H. Abraham, "Scales of solute hydrogen-bonding: Their construction and application to physicochemical and biochemical processes." *Chem. Soc. Rev.*, pp. 73-83, 1993.
- [8] P. Laffort and P. Héricourt, "Solvation parameters.2. A simplified molecular topology to generate easily optimized values." *J. Chem. Inf. Model.*, vol. 46, pp. 1723-1734, 2006.
- [9] M.H. Abraham and J.C. McGowan, "The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography." *Chromatographia*, vol. 23, pp. 243-246, 1987.
- [10] A. Zamora, "An algorithm for finding the smallest set of smallest rings." *J. Chem. Inf. Comput. Sci.*, vol. 16, pp. 40-43, 1976.
- [11] F. Chauvin and P. Laffort, "Determination of a polarizability index by means of molecular topology." *J. Chim. Phys.*, vol. 94, pp. 1216-1233, 1997.
- [12] ChemAxon Ltd., [cited 2008]. Available from: <http://www.chemaxon.com/marvin/>
- [13] M.H. Abraham and W.E. Acree, "Correlation and prediction of partition coefficients between the gas phase and water, and the solvents dodecane and undecane." *J. Chem. New.*, vol. 28, pp. 1538-1543, 2004.
- [14] A.M. Zissimos, M.H. Abraham, A. Klamt, F. Eckert, and J. Wood, "A comparison between the two general sets of linear energy descriptors of Abraham and Klamt." *J. Chem. Inf. Comput. Sci.*, vol. 42, pp. 1320-1331, 2002.
- [15] Ertl P (Novartis), "JME editor". [cited 2008]. Available from: <http://www.molinspiration.com/cgi-bin/properties>
- [16] P. Ertl, B. Rohde, and P. Selzer, "Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties." *J. Med. Chem.*, vol. 43, pp. 3714-3717, 2000.
- [17] K. Palm, P. Stenberg, K. Luthman, and P. Artursson, "Polar molecular surface properties predict the intestinal absorption of drugs in humans." *Pharm. Res.*, vol. 14, pp. 568-571, 1997.
- [18] C. Mintz, W.E. Acree Jr, and M.H. Abraham, "Comments on Solvation Parameters. 2. A simplified molecular topology to generate easily optimized values." *J. Chem. Inf. Model.*, vol. 46, pp. 1879-1881, 2006.
- [19] E.S.Z. Kováts and R. Kresz, "Wrong gas/liquid partition data by gas chromatography." *J. Chromatogr. A*, vol. 1113, pp. 206-219, 2006.
- [20] A. Dallos, A. Sisak, Z. Kulcsár, and E. Kováts sz. "Pair-wise interactions by gas chromatography. VII. Interaction free enthalpies of solutes with secondary alcohol groups." *J. Chromatogr. A*, vol. 904, pp. 211-242, 2000.
- [21] A. Klamt, "COSMO and COSMO-RS. In Encyclopedia of Computational Chemistry;" Wiley. New York: vol. 2, P. v. R. Schleyer, L. Allinger, Eds., pp. 604-615, 1998.
- [22] A. Klamt, "COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design," Elsevier, Amsterdam, pp. 246, 2005
- [23] A.A. Oliferenko, P.V. Oliferenko, J.G. Hudleston, R.D. Rogers, V.A. Palyulin, N.S. Zefirov, and A.R. Katrizky, "Theoretical scales of hydrogen bond acidity and basicity for application in QSAR/QSPR studies and drug design. Partitioning of aliphatic compounds." *J. Chem. Inf. Comput. Sci.*, vol. 42, pp. 1320-1331, 2002.
- [24] J. Jover, R. Bosque, and J. Sales, "Determination of Abraham solute parameters from molecular structure." *J. Chem. Inf. Comput. Sci.*, vol. 44, pp. 1098-1106, 2004.
- [25] D. Svozil and J.G.K. Sevcik, "Neural network prediction of the solvatochromic polarity/polarisability parameter." *J. Chem. Inf. Comput. Sci.*, vol. 37, pp. 338-342, 1997.
- [26] J.A. Platts, D. Butina, M.H. Abraham, and A. Hersey, "Estimation of molecular linear free energy relation descriptors using a group contribution approach." *J. Chem. Inf. Comput. Sci.*, vol. 39, pp. 835-845, 1999.
- [27] J.A. Platts, "Theoretical prediction of hydrogen bond donor capacity." *Phys. Chem. Chem. Phys.*, vol. 2, no. 5, pp. 973-980, 2000.
- [28] J.A. Platts, "Theoretical prediction of hydrogen bond basicity." *Phys. Chem. Chem. Phys.*, vol. 2, no. 14, pp. 3115-3120, 2000.
- [29] O. Lamarche, J.A. Platts, and A. Hersey, "Theoretical prediction of the polarity/ polarizability parameter π_2^H ." *Phys. Chem. Chem. Phys.*, vol. 3, no. 14, pp. 2747-2753, 2001.