

Supporting Information for: Assessment of the PC-SAFT model accuracy: from sensitivity analysis to property prediction

Benoit Creton ^a, Chakib Agoudjil ^a, Jean-Charles de Hemptinne ^{a,*}

^a IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

* E-mail: jean-charles.de-hemptinne@ifpen.fr

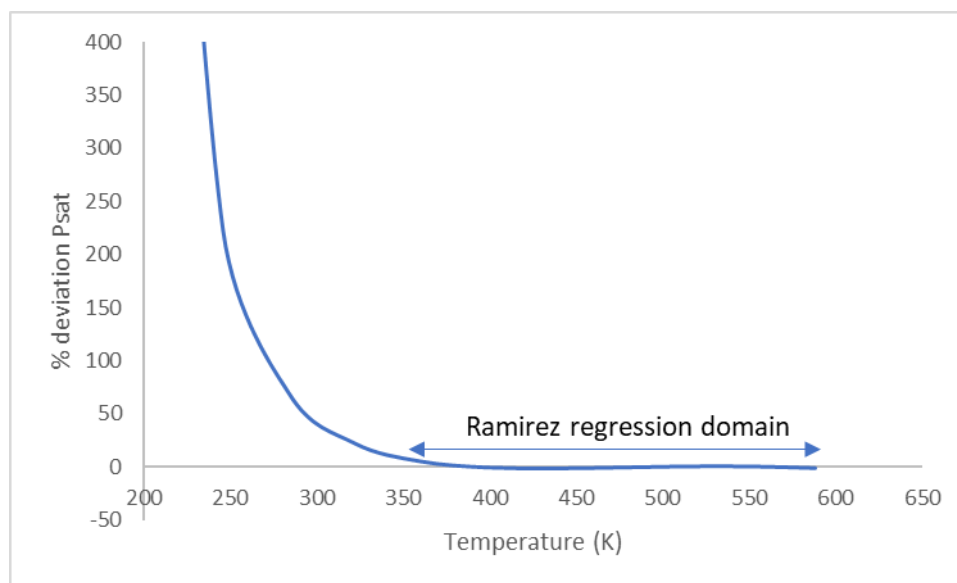


Figure S1. Deviations as a function of temperature for 3-methylpentan-1-ol and the regressed parameters by Ramírez-Vélez et al.

Some examples of molecular construction with GC-PPC-SAFT

(extracted from a technical note)

1.1 Excel file construction

An excel file is attached that contains all validated parameters. It contains following tabs:

1.1.1 PCP

This tab contains some pure component parameters. The significance of the parameter list is provided in Table S1.

Table S1. Significance of the parameters after the key-letter 'm'.

Parameter number	meaning	Unit	Equation number in GC-PPC-SAFT
1	Molar mass	(kg/mol)	indicative
2	Critical Temperature	(K)	Used only for initialization with Wilson equation
3	Critical Pressure	(Pa)	Used only for initialization with Wilson equation
4	Acentric factor	(-)	Used only for initialization with Wilson equation
5	eps/k	(K)	(4), (12), (13), (14)
6	sigma	(A)	(4), (12), (13),
7	Softness	(-)	(4)
8	Chain length m	(-)	(3), (6), (12), (13),
9	Number of associating sites	(-)	(16), (17)
10	type of association *	(-)	
11	epsAB/k	(K)	(18), (20)
12	kappaAB (-)	(-)	(18), (21)
13	dipole moment	(D)	(26-35)
14	quadrupole moment	(B)	(26-35)
15	xpm (dipole fraction)	(-)	(26-35)
16	xpm (quadrupole fraction)	(-)	(26-35)
17	Ionisation energy	(eV)	(15)
18	Charge	(-)	Only for ions; not used in this work
19	m_Gass	(-)	Id of the associating group (so as to apply UIJ and WIJ interaction parameters)

*: The code for the association type is defined as follows: each site carries a charge (positive, negative or zero). If the charge is zero, the site can associate with any other one (for acids for example). Positive charges can only associate with negative ones and vice-versa. Here are the meanings of the codes:

- Type = -1 : this choice leads to the selection of only negative sites, allowing to create species that cross-associate only (no self-association is possible). It may also be used for ionic species where all sites bear the sign of the ionic charge.
- type = 0 : all sites carry zero charge. We use this type to identify association of type A according to the Huang and Radosz [20] classification.
- type = 1: the first site carries a positive site and all others a positive one. This corresponds to the type B association type according to Huang et Radosz[20].
- type = 2: the two first sites bear a negativ charge while all other are positive, as in the 4C scheme of Huang et Radosz [20].

1.1.2 PGP

The PGP tab contains group parameters. For each group (the id if the groups are listed in table S3), following parameters are provided:

Table S2. Significance of the parameters after the key-letter 'g' (for groups).

Parameter number	Group Property	Unit	Equation number in GC-PPC-SAFT
1	Molar mass (MM)	kg/mol	
2	Contribution to the critical Temperature (Tc)*	K	Used only for initialization with Wilson equation
3	Contribution to the critical pressure (Pc)*	Pa	Used only for initialization with Wilson equation
4	Contribution to the acentric factor (ω)*		Used only for initialization with Wilson equation
5	Epsilon (ϵ/k)	K	(33), (18), (20)
6	Sigma (σ)	A	(34)
7	Softness (λ)	(-)	(36)
8	Dipole moment (μ)	Debye	(26-35)
9	Quadrupole moment (Q)	Buckingham	(26-35)
10	Dipole fraction ($m x_k^{p\mu}$)		(26-35)
11	Quadrupole fraction ($m x_k^{pQ}$)		(26-35)
12	Pseudo-ionization energy (J)	eV	(44)
13	Charge		Not used
14	Number of associating sites		(16)
15	Charge site 1		(18)
16	Association energy site 1 (ϵ^{A_1}/k)	K	(18), (20)
17	Association volume site 1 (κ^{AB})		(18), (21)
18	Charge site 2		(18)
19	Association energy site 2 (ϵ^{A_2}/k)	K	(18), (20)
20	Association volume site 2 (κ^{AB})		(18), (21)
21	Charge site 3		(18)
22	Association energy site 3 (ϵ^{A_3}/k)	K	(18), (20)
23	Association volume site 3 (κ^{AB})		(18), (21)
24	Parameter μ_1 of the dipole moment **		(26-35)

25	Parameter μ_2 of the dipole moment **		(26-35)
26	Chain length contribution for position p=1 ($R_{[1]}$)		(42)
27	Chain length contribution for position p=2 ($R_{[2]}$)		(42)
28	Chain length contribution for position p=3 ($R_{[3]}$)		(42)
29	Chain length contribution for position p=4 ($R_{[4]}$)		(42)
30	Chain length contribution for position p=5 ($R_{[5]}$)		(42)
31	Chain length contribution for position p=6 ($R_{[6]}$)		(42)
32	Chain length contribution for position p=7 ($R_{[7]}$)		(42)
33	Chain length contribution for position p=8 ($R_{[8]}$)		(42)
34	Chain length contribution for position p=9 ($R_{[9]}$)		(42)
35	Chain length contribution for position p=10 ($R_{[10]}$)		(42)

* Numerical values for the critical parameters are needed for initializing the flash (Wilson method). Yet, for as much as the flash algorithm is stable, these values should have no effect on the final solution of the PC-SAFT computation. As a consequence, a simplified group contribution method (inspired by Marero Gani) is used here.

1.1.3 CROSS

The tab 'CROSS' contains binary interaction parameters, that may be between molecules, between groups (in case of association), or between groups and molecules.

1.1.4 mg

This tab contains the group contributions used for various molecules. The description of the numbers is provided in section 1.2.

1.2 Available groups and examples

Using the SAFTBANK database, GC-PPC-SAFT is capable of treating compounds from any of the following families (see the references in table S1) :

- Linear, cyclic and methyl- alkanes (groups 1, 2, 3, 4, 401, 11, 1101, 1102, 1103)
- Alkenes (groups 5 and 6)

- 1, 2, 3, 4 alcohols (groups 14, 15 and 1401)
- Aromatic alcohols (groups 1402, 1403, 1404 and 1405)
- Primary (group 33 or 36), secondary (group 32) and tertiary (group 34 or 35) amines
- Alkyl formates (group 24)
- Alkyl esters (group 77)
- Alkyl aldehydes (group 20) or aromatic aldehydes (group 2001)
- Alkyl ketones (2, 3, 4, 5 ketones, group 112) and aromatic ketones (group 113)
- Alkyl ethers (2, 3, 4, 5 éthers, group 166) and aromatic ethers (group 167)
- Alkyl benzenes (groups 9, 10, 101 and second order groups)
- Polyaromatics (groups 102 and second order groups)

In some cases the group position is of importance. Two integers are available for this purpose, named p1 and p2. They stand for the distance of the group to the end of chain that is closest, in number of carbon. Only the ester group (77) uses p2 (see below section 5.1.5). The position p1 is often used to make it possible to adapt the chain length contribution (R) of the group.

Table S3 shows the groups that are available in the SAFTBANK database, with their identifier. You may note that some identifiers are fixed because they are recognized in the code for specific treatments (see comments):

Table S3: identifiers of the groups in the 'SAFTBANK.TXT' database, the origin of the parameters and comments on their use.

Group Identifier	Type of group	Reference	Comment
Alkanes			
1	CH3	[18,21]	P1 used to adapt Ri
2	CH2	[18,21]	
3	CH	[18,21]	
4	CH2(cyc)	[22]	
401	CH(cyc)	[22]	
11	Cquat	[23]	
1101	two neighbours are CH2	[23]	second order correction for Cquat
1102	one neighbour is end of chain CH3 and other is ternary CH	[23]	second order correction for Cquat
1103	one neighbour is end of chain CH3 and other is Cquat	[23]	second order correction for Cquat
! Olefins (count twice the group number 6 or once 6 and once 5)			
6	(CH=) olef	[24]	P1 used to adapt Ri
5	(CH2=)olef	[24]	

! Alcohol : ###			
(OH) ### Position			
1 – 2 – 3 – 4			
14	(OH)1	[24,25]	
15	(OH)1 ethanol	[3]	
1401	(OH)i	[25]	P1 used to adapt Ri
1402	(OH)aro 3B	[26]	
1403	(OH)aro 3B pos 2	[26]	In case of two branches on the aromatic ring
1404	(OH)aro 3B pos 3	[26]	In case of two branches on the aromatic ring
1405	OH aro Guaiacol	[27]	OH and –O– are neighbouring on the aromatic ring
!Amines			
32	(NH)	[28]	
33	(NH ₂) 2B	[28]	
34	(NCH ₃)	[28]	
35	(N(CH ₃) ₂)	[28]	
36	(NH ₂) 3B	[28]	
! Formates : ###			
(HCOO) ###			
24	(O-CH=O)formates	[29]	
!Esters : ### (COO)			
### Position 2 – 3 – 4			
77	COO	[30]	P1 and P2 used to adapt the dipole moment and the R value
! Aldehydes : ###			
(CHO) ###			
20	(CHO) aldehyde	[29]	P1 used to adapt the dipole moment
2001	(CHO) aro	[26]	
2002	(CHO)furfural	[31]	
!Ketones : ###			
(CO) ### Position =			
2 – 3 – 4 – 5			
112	(C=O)linéaire	[29]	P1 may not be larger than 5
113	(C=O)aro	[26]	
!Ethers : ### (O)			
### Position = 2 – 3 – 4 – 5			
166	-O-	[29]	P1 may not be larger than 5
167	(-O-) aro	[26]	
168	(-O-) 5-ring aro	[31]	Specific for furane
169	(-O-) 5-ring aro	[31]	Specific for Furfural
165	(-O-) aro guaiacol	[27]	
acids COOH			
42	COOH acid	[24]	
! Alkyl benzenes :			
### ©AB ### ###			

(CH=)BZ ### (C)Pa
 ### ## (CH3)P=1-
 8 des
 naphthalenes

10	(C-)aro	[32]
9	(CH)aro	[32]
901	(C)polyaro	[33]
902	(C)polyaro next to O	[31]
101	(CH3)aro	[34]
102	(CH3)polyaro	[33]
103	(CH2)aro	[26]

Ex: benzofurane
 P1 may not be larger than 5
 P1 may not be larger than 7

! Second order
 corrections for
 aromatic
 molecules

10003	Mono aromatic ring	[3]
10013	Several groups on the multi-branched aromatic ring (xylenes +)	[34]
10021	Naphthalene	[33]
10022	phenanthrene	[33]
10023	anthracene	[33]
10024	benzanthracene	[33]
10025	chrysene	[33]
10026	naphtacene	[33]
10027	triphenylene	[33]
10028	phenol	[26]
10029	benzaldehyde	[26]
10030	anisole	[26]
10031	methyl benzoate	[26]
10032	acetophenone	[26]
10033	Branched phenols	[31]
11003	benzoFurane	[31]
11004	dibenzoFurane	[31]
11005	Furfural	[31]
11006	guaiacol	[27]

The molecules are then constructed with a list of quadruplets that finishes with the number '-1'. The significance of the four numbers are:

- Number of occurrences
- Id of the group
- P1
- P2

1.2.1 alkanes

a. n-alkanes :

This family of molecules is rather simple to implement, as it contains only two type of groups : (1) – CH₃ and (2) – CH₂.

Example :

n-pentane is described with following line:

109660 mg 2 1 1 0 3 2 1 0 -1

The group position p₁ is always 1 for the n-alkanes.

b. Methyl-alkanes:

In order to distinguish among isomers, de Hemptinne et al. [19] suggested to modify the value of the parameter R as a function of the position p of the methyl group. The value of p is defined as the shortest distance to the chain end. Specific parameters for R exist until position 5. Beyond that position, following linear relationship is used:

$$R = R_6 p + R_7$$

As an example, 3-methyl pentane is described as:

96140 mg 2 1 1 0 1 1 3 0 2 2 1 0 1 3 1 0 -1

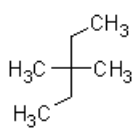
When the methyl is in position 2, the two end groups cannot be distinguished, so both are given in that position.

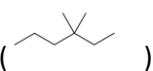
Example: 2-methyl hexane

107835 mg 1 1 1 0 2 1 2 0 3 2 1 0 1 3 1 0 -1

c. Branched alkanes with quaternary carbon group (group 11):

These alkanes have been worked on by Bui [23], and they must combine a quaternary group with second order groups. Three different second order groups may exist:

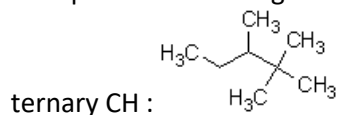
- Group 1101: one neighbour of the quaternary carbon is a CH₂: 


Example: 3,3 dimethyl hexane ()

563166 mg 4 1 1 0 3 2 1 0 1 11 1 0 2 1101 1 0 -1

Note that group 1101 comes in twice because it has two 'CH₂' neighbours

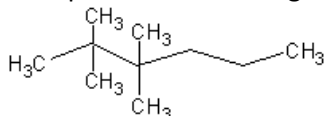
- Group 1102 : one neighbour of the quaternary carbon is an end of chain CH₃, another is a

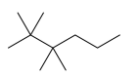


Example: 2,2,3 trimethyl pentane ()

564023 mg 5 1 1 0 1 2 1 0 1 3 1 0 1 4 1 0 1 1102 1 0 -1

- Group 1103: one neighbor of the quaternary carbon is another quaternary carbon:



Example 2,2,3,3 tetramethyl hexane ():

13475815 mg 6 1 1 0 2 2 1 0 2 11 1 0 1 1103 1 0 1 1101 1 0 -1

Note that in this last case two second order groups must be considered. Once the group 1103, because the neighbour of the quaternary carbon in position 3 is another quaternary carbon (i.e. position 2), and another (group 1101) because the quaternary carbon in position two has three CH₃ neighbours.

d. Olefins :

There must always be two olefin groups in a given molecule. For alfa olefins, we have groups 5 and 6.

Example : 1-pentene

109671 mg 1 1 1 0 2 2 1 0 1 5 1 0 1 6 1 0 -1

When the olefinic groups are internal to the molecule, we have twice group 6. The position p is always the number of carbon closest to the end of the chain.

Example : n-trans-2 pentene

646048 mg 2 1 1 0 1 2 1 0 2 6 2 0 -1

It is impossible at this stage to distinguish among cis- and trans-olefins.

e. Cyclo-alkanes :

Cyclohexane and methylcyclohexane are considered as specific compounds. Beyond these, the alkyl-cyclo hexanes are treated using 5 cyclic carbons (group 4) and one (CH<)c group which is bonded to an alkyl chain [22].

Example : ethyl-cyclohexane

1678917 mg 5 4 1 0 1 401 1 0 1 1 1 0 1 2 1 0 -1

1.2.2 alcohols

Here we discuss the -OH (1-alcohols, group n°14 et 2,3,4-alcohols, group n°1401). They are all treated as 3B associating.

a. 1-alcohols :

They are treated with a 3B-type association (group 14). The position (p1 and p2) is not used (their value is 1).

example : 1-hexanol

111273 mg 1 1 1 0 5 2 1 0 1 14 1 0 -1

Note that ethanol is a special case (see section 5.6)

b. i-alcohols :

When the OH group is not at the end of a chain (as for 2-, 3-, ...alcohols), we use the group 1401, for which the association volume is smaller [25]. The position of the OH group in the molecule must be indicated since specific R parameters exist until position 4. Beyond that point, the R value remains at R_[4].

Exemple : 3-hexanol:

623370 mg 2 1 1 0 3 2 1 0 1 3 1 0 1 1401 3 0 -1

1.2.3 aldehydes

Aldehydes (R-CHO) are described using group 20. This group is always at the end of a chain (i.e. position 1).

The first member of this family, formaldehyde, is treated as a specific molecule (i.e. no group contributions).

Example: n-pentanal:

110623 mg 1 1 1 0 3 2 1 0 1 20 1 0 -1

The dipole moment of the aldehydes depend on the chain length as follows [29]:

$$\mu = \mu_0 + \frac{\mu_1}{m}$$

1.2.4 formates

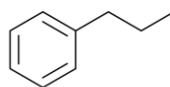
Formates (R-OCH=O) require group 24. It is always at the end of the chain.

638493 mg 1 1 1 0 4 2 1 0 1 24 1 0 -1

1.2.5 Esters (-COO-, 77) , ketones (-CO-, 112), ethers (-O-, 166)

These groups have a negatively charged association site with the alcohol parameters (see Nguyen-Huynh [24]). This makes it possible to take into account cross-association with proton donor sites (e.g. [25]).

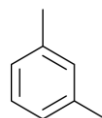
For the aromatic molecules, an additional, second order group, must be added to the molecular group description. The most recent description of these molecules is proposed by Grandjean [34].



As an example, here is the description of propyl benzene :

103651 mg 5 9 10 1 10 10 1 103 10 1 2 1 0 1 1 10 1 10003 10 -1

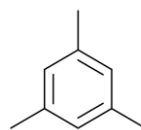
It contains 5 aromatic CH groups (9), one aromatic >C- (10), one aromatic CH₂ (103), one linear CH₂ (2) and one linear CH₃ (1). In addition, the aromatic group is treated as a second order contribution using the default aromatic group (10003).



The second example is that of m-xylene (1,3 dimethyl benzene) :

108383 mg 4 9 10 2 10 10 1 101 10 1 101 3 0 1 10013 10 -1

It contains 4 aromatic CH groups (9), 2 aromatic >C- (10), one aromatic CH₃ (101) in position 1, and another on position 3. In addition, the aromatic group is that which is recommended for xylenes (10013).



The second example is that of mesitylene (1,3,5 trimethyl benzene) :

108678 mg 3 9 10 3 10 10 1 101 10 1 101 3 0 1 101 5 0 1 10013 10 -1

It contains 3 aromatic CH groups (9), 3 aromatic >C- (10), three aromatic CH₃ (101) in positions 1, 3 and 5 (here each group must be described). In addition, the aromatic group is that which is recommended for xylenes (10013).

1.2.7 Poly-aromatics

The recommended description of poly-aromatic compounds is that proposed by Tran [33]. A new group was designed which is the (C)polyaro (group 901), and consists in a carbon that is shared among two aromatic rings. Its chain length parameter depends on the type of condensation. This is why the use of group 901 comes with the definition of a position parameter, as further detailed in table 5. In addition, second order groups were defined that depend on the number of condensed rings. The branches are described in the same way as for mono-aromatic molecules, with the groups (CH₃)aro (group 101) or (CH₂)aro (group 103). The position of the branches must also be specified as shown in figure 2.

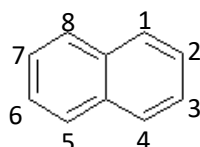
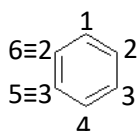


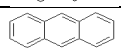
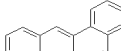
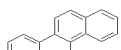
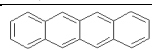

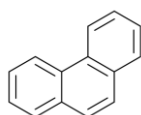


Figure 2. Definition of the position of the CH₃ group that is branched on the aromatic group, for alkyl-benzenes and alkyl-naphthalenes.

Table S25 shows how each polyaromatic structure requires two identifications that must be used together : a position parameter that comes with the polyaromatic group (901) and one second order group, the list of which is also in table S2 .

Table S2. Use of the polyaromatic group (C)polyaro (n° 901) as defined in by Tran et al. [33]

Compound	Chemical formula	Position of group 901	Second order group
Naphthalene		1	10021
Phenanthrene		2	10022
Anthracene		3	10023
Benzanthracene		4	10024
Chrysene		5	10025
Naphtacene		6	10026
Triphenylene		7	10027

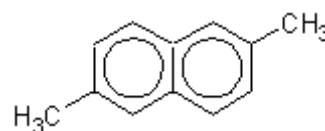


Example : Phenanthrene

85018 mg 10 9 10 4 901 2 0 1 10022 10 -1

This compound is composed of 10 aromatic CH groups (9), 4 polyaromatic C groups (901) located on position 2 according to table 5, and one second order group (10022).

Example of a branched polyaromatic: 2-6 di methyl naphthalene:



581420 mg 6 9 1 0 2 10 1 0 2 901 1 0 1 102 2 00 1 102 6 0 1 10021 1 0 -1

This compound is composed of 6 aromatic CH groups (9), 2 aromatic (C) groups (10), 2 polyaromatic C groups (901) located on position 1 according to table xx, two aromatic CH3 groups (102), one on position 2 and the other on position 6, and one second order group for naphthalene (10021).

1.2.8 Oxygenated aromatics

This family of compounds is described at length by Nguyen-Huynh et al. [26]. Specific treatment was found to be needed for oxygenated groups attached to an aromatic ring. Following groups were then introduced (table S3)

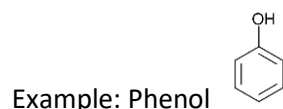
Table S3: Groups for oxygenated aromatics

Group Id	Type of group	occurrence
1402	(OH)aro 3B	Phenol
1403	(OH)aro 3B pos 2	In case of two branches on the aromatic ring
1404	(OH)aro 3B pos 3	In case of two branches on the aromatic ring
2001	(CHO) aro	Aldehydes
113	(C=O)aro	Ketones
167	(-O-) aro	Esters

In addition, second order groups must be used on the aromatic ring (table S4).

Table S4: second order groups defined for oxygenated aromatics

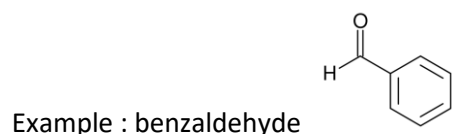
Group Id	occurrence
10028	phenol
10029	benzaldehyde
10030	anisole
10031	methyl benzoate
10032	acetophenone



This molecule [26] requires a specific second order group (10028):

108952 mg 5 9 10 1 10 10 1 1404 10 1 10028 10 -1

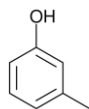
Otherwise, the composition is described straightforwardly as 5 aromatic C (9), one branched aromatic C (10) and one aromatic OH.



100527 mg 1 10 10 5 9 10 1 2001 10 1 10029 10 -1

The specific second order group is here 10029, and the aromatic aldehyde group 2001 is used.

Note that when several substitutes are present on the aromatic ring, the group 10013 (xylenes) remains.

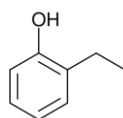


Example: m-cresol = 3-methyl phenol

The alkylphenols have been revisited by Grandjean [31] after the publication by Nguyen-Huynh [26] who contained an inversion in the position of the CH₃ groups: the all alkylphenols now use the 1402 aromatic OH, except the 2-alkylphenols who use 1403.

108394 mg 4 9 10 2 10 10 1 101 3 0 1 1404 1 0 1 10033 1 0 -1

It is composed of four aromatic CH groups (9), 2 aromatic (C) groups (10), one aromatic CH₃ (101) and one aromatic OH that has no close neighbour on the aromatic ring (1402). In addition, a second order group must be used for the aromatic ring, which is the one for oxygenated aromatics (10033).



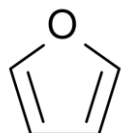
Example: 2-ethyl phenol:

90006 mg 4 9 10 2 10 10 1 1403 1 0 1 103 2 0 1 1 1 0 1 10033 0 0 -1

It is composed of four aromatic CH groups (9), 2 aromatic (C) groups (10), one aromatic CH₂ (103) on position 2, one normal CH₃ (1) and one aromatic OH that has a close neighbor on the aromatic ring (1403). In addition, a second order group must be used for the aromatic ring, which is the one for oxygenated aromatics (10033).

1.3 Specific molecules to be constructed using group contribution

When several functional groups appear on the same molecule, there is no choice but using group contribution to define them.

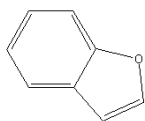


1.3.1 Furane :

This molecule is discussed by [31]. The -O- group is described using the specific group numbered 168. In addition, a second order group has been designed to carry the quadrupole.

110009 mg 1 168 1 0 4 9 10 1 11001 1 0 -1

1.3.2 2,3-benzofurane

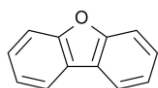


This compound is discussed by [31]. It is described as follows:

271896 mg 6 9 1 0 1 901 3 0 1 902 2 0 1 168 1 0 1 11003 1 0 -1

Since it contains 6 aromatic CH (group9), one polyaromatic CH (901) that is on position 3 with respect to the O group, one polyaromatic CH neighboring the oxygen (902) that is on position 2 with respect to the O group, the furanic O group (168) and the second order group for benzofuran.

1.3.3 Dibenzofurane

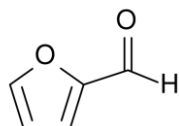


This molecule is discussed by [31]. It is described as follows:

132649 mg 8 9 1 0 2 902 2 0 2 901 3 0 1 168 1 0 1 11004 1 0 -1

It contains 8 aromatic CH (group 9), two polyaromatic CH (901) that are on position 3 with respect to the O group, two polyaromatic CH neighboring the oxygen (902) that are on position 2 with respect to the O group, the furanic O group (168) and the second order group for dibenzofuran.

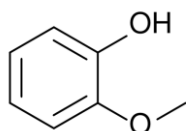
1.3.4 Furfural



This molecule is discussed by [31]. It has two specific groups: the -O- group in the aromatic ring, 169 (O in the ring, which includes the quadrupole parameters) and 2002 (aldehyde group with a modified dipole moment) :

98011 mg 3 9 1 0 1 10 1 0 1 2002 2 0 1 169 1 0 -1

1.3.5 Guaiacol



This molecule is discussed by [27]. It is described as follows:

90051 mg 1 1405 1 0 4 9 1 0 2 10 1 0 1 165 1 0 1 1 1 0 1 11006 1 0 -1

This molecule has many specific groups. This is due to the fact that these groups neighbor each other, such that a screening effect is observed. The -O- group is attached to an aromatic ring, but neighbors

the OH group. It was thus found to require a specific groups (165). The OH group similarly neighbors the –O- group. It is therefore described with 1405. The aromatic ring, is conventionally described using 4 aromatic CH groups (9) and 2 aromatic C groups (10). The methyl end group is (1). The aromatic cycle is also affected by the proximity of two oxygenated groups and requires therefore a specific group (11006).

1.4 Some pure compounds

As described above, a number of low molecular weight pure compounds are directly available in the database and are recognized using their CAS number. No additional description is needed to identify them. Table S5 provides the CAS numbers to be used for each of these compounds.

Table S5: list of pure compounds available in the database 'SAFTBANK.TXT'

name	CAS (Id)	Ref
H2	1333740	[33]
N2	7727379	[35]
H2S	7783064	[35]
water	7732185	[36]
water	7732186	[9]
Methane	74828	[35]
Methanol	67561	[21]
CO	630080	[35]
O2	7782447	[35]
CO2	124389	[9]
Ethane	74840	[35]
ethanol_3B	64175	[37]
cyclohexane	110827	[33]
benzene	71432	[33]
Toluene	108883	[33]
Methyl cyclohexane	108872	[33]
Acetic acid	64197	[38]
Ammonia	7664417	[34]
Tetrahydrofuran	109999	[39]
Acetonitrile	75058	[39]
DMSO	67685	Never published
DME (dimethyl ether)	115106	Never published
Glycerol	56815	[40]
Methyl Oleate	112629	[40]
Ethyl benzene	100414	[36]

Note that two descriptions of water exist. One is given with CAS number 7732185 and the other with 7732186. The first [36] has been extensively used and tested with many compounds. Its main drawback is a relatively bad representation of the water density at room temperature. This is why recently a second parameterization was performed [9] which improves significantly the results. However, the number of tests with this water is much lower (only equilibria with alkanes and mono-functional molecules were tested).

➔ The two representations of water can be compared using the file water.dat.

Ethanol can be described in two distinct ways: in a molecular way as proposed by Mourah [41] or using the group contribution approach as done by Ahmed [37]. The molecular model is obtained by using the CAS number (64175); the group contribution approach is found using (the CAS number was modified so as to avoid interference with the existing compound):

64176 mg 1 1 1 0 1 2 1 0 1 15 1 0 -1

➔ The file etol.dat makes it possible to compare the two methods.

List of references

- [1] J. Gross, G. Sadowski, *Fluid Phase Equilib.* 168 (2000) 183–199.
- [2] J. Gross, G. Sadowski, *Ind.Eng.Chem.Res.* 40 (2001) 1244–1260.
- [3] D. Nguyen-Huynh, J.-P. Passarello, P. Tobaly, J.-C. de Hemptinne, *Fluid Phase Equilib.* 264 (2008) 62–75.
- [4] P.K. Jog, W.G. Chapman, *Mol.Phys.* 97 (1999) 307–319.
- [5] T.-K.-H. Trinh, J.-P. Passarello, J.-C. de Hemptinne, R. Lugo, *Fluid Phase Equilibria* 429 (2016) 177–195.
- [6] T. Boublik, *J.Chem.Phys.* 53 (1970) 471–473.
- [7] G.A. Mansoori, N.F. Carnahan, K.E. Starling, T.W.J. Leland, *J.Chem.Phys.* 54 (1971) 1523–1525.
- [8] S.S. Chen, A. Kreglewski, *Ber.Bunsenges.Phys.Chem* 81 (1977) 1048–1052.
- [9] S. Ahmed, N. Ferrando, J.-C. de Hemptinne, J.-P. Simonin, O. Bernard, O. Baudouin, *J. Chem. Eng. Data* 61 (2016) 4178–4190.
- [10] W.G. Chapman, K.E. Gubbins, G. Jackson, M. Radosz, *Fluid Phase Equilibria* 52 (1989) 31–38.
- [11] W.G. Chapman, K.E. Gubbins, G. Jackson, M. Radosz, *Ind.Eng.Chem.Res.* 29 (1990) 1709–1721.
- [12] J.A. Barker, D. Henderson, *J.Chem.Phys.* 47 (1967) 2856–2861.
- [13] J.A. Barker, D. Henderson, *J CHEM PHYS* 47 (1967) 4714–4721.
- [14] D. Nguyen-Huynh, J.P. Passarello, P. Tobaly, J.C. de Hemptinne, *Ind.Eng.Chem.Res.* 47 (2008) 8847–8858.
- [15] G. Jackson, W.G. Chapman, K.E. Gubbins, *Mol.Phys.* 65 (1988) 1–31.
- [16] S.O. Derawi, G.M. Kontogeorgis, M.L. Michelsen, E.H. Stenby, *Ind.Eng.Chem.Res.* 42 (2003) 1470–1477.
- [17] G. Stell, J.C. Rasaiah, H. Narang, *Mol.Phys.* 27 (1974) 1391–1414.
- [18] S. Tamouza, J.P. Passarello, P. Tobaly, J.C. de Hemptinne, *Fluid Phase Equilib.* 222-223 (2004) 67–76.
- [19] de Hemptinne, J. C, P. Mougine, A. Barreau, L. Ruffine, S. Tamouza, R. Inchekel, *Oil & Gas Sci.Tech.- Rev.IFP Energies nouvelles* 61 (2006) 363–386.
- [20] S.H. Huang, M. Radosz, *Ind.Eng.Chem.Res.* 29 (1990) 2284–2294.
- [21] M. Mourah, D. NguyenHuynh, J.P. Passarello, J.C. de Hemptinne, P. Tobaly, *Fluid Phase Equilib.* 298 (2010) 154–168.
- [22] T.K.S. TRAN, J.-P. Passarello, P. Tobaly, De Hemptinne J.-C., N. Ferrando, *Etude des paramètres d'interaction binaires kij des mélanges d'hydrocarbures et d'hydrogène par l'équation d'état SAFT.* master, LIMHP, 2007.
- [23] H.-V. Bui, *Utilisation d'une équation d'état issue de la thermodynamique statistique pour décrire des alcanes ramifiés.* Rapport Master Catalyse et Procédés, IFP-School, 2013.
- [24] D. Nguyen-Huynh, *Modélisation thermodynamique des mélanges syématriques et asymétriques de composés polaires oxygénés et/ou aromatiques par GC-SAFT.* PhD, 2008.
- [25] D. Nguyen-Huynh, J.P. Passarello, J.C. de Hemptinne, F. Volle, P. Tobaly, *Journal of Supercritical Fluids* 95 (2014) 146–157.

- [26] D. Nguyen-Huynh, de Hemptinne, Jean Charles, R. Lugo, J.-P. Passarello, P. Tobaly, *Chem. Eng. Res. Des.* 92 (2014) 1912–1935.
- [27] C.G. Pereira, C. Féjean, S. Betoulle, N. Ferrando, R. Lugo, J.C. de Hemptinne, P. Mougin, *Fluid Phase Equilibria* (2018).
- [28] J. Rozmus, I. Brunella, P. Mougin, J.-C. de Hemptinne, *J.Chem.Eng.Data* 57 (2012) 2915–2922.
- [29] D. Nguyen-Huynh, J.P. Passarello, J.C. de Hemptinne, P. Tobaly, *Fluid Phase Equilib.* 307 (2011) 142–159.
- [30] D. Nguyen-Huynh, A. Falaix, J.P. Passarello, P. Tobaly, J.C. de Hemptinne, *Fluid Phase Equilib.* 264 (2008) 184–200.
- [31] L. Grandjean, de Hemptinne, J. C., *Application du modèle thermodynamique GC-PPC-SAFT aux systèmes contenant des molécules oxygénées multifonctionnelles et issus de la liquéfaction du charbon ou de la biomasse lignocellulosique*, 2013.
- [32] D. Nguyen-Huynh, M. Benamira, J.P. Passarello, P. Tobaly, J.C. de Hemptinne, *Fluid Phase Equilib.* 254 (2007) 60–66.
- [33] T.K.S. Tran, D. NguyenHuynh, N. Ferrando, J.P. Passarello, J.C. de Hemptinne, P. Tobaly, *Energy & Fuels* 23 (2009) 2658–2665.
- [34] L. Grandjean, de Hemptinne, Jean Charles, R. Lugo, *Fluid Phase Equilib.* 367 (2014) 159–172.
- [35] D. Nguyen-Huynh, *Modélisation thermodynamique de mélanges symétriques et asymétriques de composés polaires oxygénés et/ou aromatiques par GC-SAFT*, 2008.
- [36] D. Nguyen-Huynh, J.C. de Hemptinne, R. Lugo, J.P. Passarello, P. Tobaly, *Ind.Eng.Chem.Res.* 50 (2011) 7467–7483.
- [37] S. Ahmed, N. Ferrando, J.-C. de Hemptinne, J.-P. Simonin, O. Bernard, O. Baudouin, *Fluid Phase Equilibria* 459 (2018) 138–157.
- [38] D. Nguyen-Huynh, J.C. de Hemptinne, R. Lugo, *Extension de l'équation GC-PPC-SAFT aux mélanges contenant de l'eau, des hydrocarbures et des composés oxygénés multifonctionnels*.
- [39] M.A. Varfolomeev, R.N. Nagrimanov, M.A. Stolov, N. Ferrando, R. Lugo, J.-C. de Hemptinne, *SI:John P O'Connell* 470 (2018) 91–100.
- [40] A. Barreau, I. Brunella, J.C. de Hemptinne, V. Coupard, X. Canet, F. Rivollet, *Ind.Eng.Chem.Res.* 49 (2010) 5800–5807.
- [41] M. Mourah, *Application de l'equation d'etat th'orique SAFT aux 'equilibres de phases intervenant dans la synthSse ou l'utilisation des biocarburants*.
- [42] T.X.N. Thi, S. Tamouza, P. Tobaly, J.P. Passarello, J.C. de Hemptinne, *Fluid Phase Equilib.* 238 (2005) 254–261.
- [43] S. Tamouza, J.P. Passarello, P. Tobaly, J.C. de Hemptinne, *Fluid Phase Equilib.* 228-229 (2005) 409–419.
- [44] C. Le Thi, S. Tamouza, J.P. Passarello, P. Tobaly, J.C. de Hemptinne, *Ind.Eng.Chem.Res.* 45 (2006) 6803–6810.
- [45] D. Nguyen-Huynh, T.K.S. Tran, S. Tamouza, J.P. Passarello, P. Tobaly, J.C. de Hemptinne, *Ind.Eng.Chem.Res.* 47 (2008) 8859–8868.
- [46] Mourah M., *Modélisation des équilibres liquide-liquide et liquide-vapeur des mélanges contenant de l'eau, des alcools et des hydrocarbures*, 2009.
- [47] T.B. Nguyen, J.C. de Hemptinne, B. Creton, G.M. Kontogeorgis, *Fluid Phase Equilib.* 372 (2014) 113–125.
- [48] J. Rozmus, J.C. de Hemptinne, P. Mougin, *Fluid Phase Equilib.* 303 (2011) 15–30.
- [49] J. Rozmus, J.C. de Hemptinne, N. Ferrando, P. Mougin, *Fluid Phase Equilibria* 329 (2012) 78–85.
- [50] TRINH Thi Kim Hoang, *Prediction of phase equilibria associated with hydrotreating process of biomass by GC-PPC-SAFT Equation of State*, Paris 13, 2015.
- [51] L. Grandjean, *Modélisation des équilibres liquide-vapeur de mélanges contenant des composés furaniques avec le modèle thermodynamique GC-PPC-SAFT. Cas du furane et du 2-méthyl furane.*, 2012.

- [52] C.G. Pereira, L. Grandjean, S. Betoulle, N. Ferrando, C. Fejean, R. Lugo, J.C. de Hemptinne, P. Mougin, *Fluid Phase Equilib.* 382 (2014) 219–234.
- [53] W.G. Chapman, S.G. Sauer, D. Ting, A. Ghosh, *Fluid Phase Equilib.* 217 (2004) 137–143.