

Jane Dannow Dyekjær · Kjeld Rasmussen  
Svava Ósk Jónsdóttir

## QSPR models based on molecular mechanics and quantum chemical calculations

### 1. Construction of Boltzmann-averaged descriptors for alkanes, alcohols, diols, ethers and cyclic compounds

Received: 7 June 2002 / Accepted: 16 July 2002 / Published online: 13 September 2002  
© Springer-Verlag 2002

**Abstract** Values for nine descriptors for QSPR (quantitative structure–property relationships) modeling of physical properties of 96 alkanes, alcohols, ethers, diols, triols and cyclic alkanes and alcohols in conjunction with the program Codessa are presented. The descriptors are Boltzmann-averaged by selection of the most relevant conformers out of a set of possible molecular conformers generated by a systematic scheme presented in this paper. Six of these descriptors are calculated with molecular mechanics and three with quantum chemical methods. Especially interesting descriptors are the relative van der Waals energies and the molecular polarizabilities, which correlate very well with boiling points. Five more simple descriptors that only depend on the molecular constitutional formula are also discussed briefly.

**Keywords** QSPR · Boiling point descriptors · Alcohols · Polarizabilities · Van der Waals energies

#### Introduction

Experimental work or thermodynamic calculations for industrial purposes, often require a value for a thermodynamic or physical property for a given compound. However, in many cases the value needed is missing in handbooks and databases. There may be numerous reasons for this, for example that the property of interest has for some reason not been measured experimentally. Hence, it is convenient to have simple models that provide estimates for missing experimental values. An interesting solution to this problem is found in QSPR (quantitative structure–property relationship) type models based on experimental values for structurally similar compounds.

A QSPR model may have the general form:

$$P = a + b\alpha + c\beta + d\gamma + \dots$$

This is a regression equation that originates from statistical fitting of experimental data to structural molecular parameters. It provides a relationship between a molecular structure and a physical property,  $P$ , for a given compound. The equation consists of regression coefficients,  $a, b, c, \dots$  and a set of molecular structure-dependent variables,  $\alpha, \beta, \gamma, \dots$  so-called descriptors. These descriptors are often divided into subgroups depending on their nature and origin, i.e. how they are evaluated from the molecular structure. Common subgroups are the constitutional, topological, geometric, electrostatic and quantum chemical descriptors. [1, 2] In most cases, difficulties in evaluating of these descriptors increase in the order of these subgroups mentioned.

Constitutional descriptors depend on the constitutional molecular formula, providing information about which atoms the molecule consists of. One such descriptor may, for instance, be the number of carbon or oxygen atoms in a molecule. More information can be achieved from topological descriptors, which include information about which atoms are bonded to one another. There are several schemes to calculate this type of descriptor, among which the more well known are the Kier and Hall indices [3, 4] and the Wiener index. [5] Geometrical descriptors provide information about the size and the spatial extent of the molecule, like the molecular volume, surface area and derivatives thereof. Dipole moments, charges and partial charges are electrostatic descriptors. Values for these descriptors are normally found by calculation with computational methods, like quantum chemical and molecular mechanics methods. This is also the case for the series of CPSA descriptors (Charged Partial Surface Area) proposed by Stanton and Jurs. [6] Finally, there is the group of descriptors that can only be evaluated by quantum chemical methods, like the ionization energy, HOMO–LUMO energy gap, etc. [7] These values are of course

J.D. Dyekjær · K. Rasmussen · S.Ó. Jónsdóttir (✉)  
Department of Chemistry, Technical University of Denmark,  
DK-2800 Kgs. Lyngby, Denmark  
e-mail: svava@kemi.dtu.dk

much more time consuming and difficult to obtain than those that can be obtained by inspection of the molecular structure. [2, 7, 8]

The purpose of this work is to improve the general model development by proposing new descriptors and by taking differences between molecular conformers (equilibrium conformations), i.e. changes of structurally important internal torsional angles within the same molecule, into account. We have done this to incorporate more physical reality in the descriptors used for QSPR modeling, because the thermodynamic properties we wish to model depend on the molecular conformers present. Our QSPR models [9, 10, 11] are developed with the Codessa program (COMprehensive DEscriptors for Structural and Statistical Analysis), [12] using the new descriptors along with the descriptors available in the program. Besides facilitating different statistical regression schemes for development of the QSPR models, this program also calculates a large number of different types of molecular descriptors when provided with information about a molecular structure.

Most molecules occur as many different conformers. An experimentally measured value of a physical property is not only measured for one single molecular conformer, but represents an average of all important conformers. Hence it gives physical meaning to incorporate this fact into the regression model, i.e. not only to use one molecular structure, but instead to use several significant molecular conformers, when developing the QSPR model. This approach is possible within the Codessa program, because Codessa allows statistical weighting of molecules – in this case, conformers – in the data regression procedure. Finally, it is possible to add more descriptors than the ones already available in Codessa. Therefore, the descriptors presented here are available in a local database and are used within the Codessa program.

In this paper a method for constructing and choosing the set of molecular conformers from a large pool of possible conformers is presented. We have calculated relative energies by molecular mechanics and electrostatic properties by quantum chemical methods. It is examined how well these properties correlate with an important physical property, the boiling point. The values presented here can therefore be used as descriptors in QSPR modeling for molecules within the compound classes alkanes, alcohols, diols, triols and ethers, also including cyclic alkanes, alcohols and diols. The purpose of this paper is to introduce and analyze the new additions to the pool of descriptors. QSPR models using this set of descriptors have been developed for a variety of properties, and are published in separate papers. These cover models for properties like boiling point, melting point, heat of fusion, heat of vaporization [9] and densities. [10] The work has recently been extended to monosaccharides, where the same set of descriptors is calculated, and QSPR models for a variety of properties have been developed. [11]

## Methods

### Considerations and computational details

In this study, 96 molecules including 16 alkanes, 39 alcohols, 23 diols, two triols, ten ethers and six other compounds, all containing up to 12 carbon atoms, were chosen. Most of the compounds are linear, but five- and six-membered rings are also included. The actual molecules considered are shown in Tables 1, 2 and 3.

Many molecules can occur as different conformers, and any solution property depends on this mixture of conformers. The idea is to include those conformers that contribute significantly to the thermodynamic properties of interest, and exclude those with smaller contributions.

A detailed study for the series of smaller molecules from butane to octane was carried out. In this study all possible conformers of each of the molecules were generated and energy minimized using the methodology described below. The relative probabilities at 298.16 K were calculated using Boltzmann statistics. From this study, it was seen that conformers with more than two “kinks”, as discussed below, did not contribute significantly, i.e. the relative probability approached zero. Based on these results and experience with a method developed for calculating interaction energies between pairs of molecules [13] the following procedure was established for selection of the significant conformers.

Linear carbon atom chains are not allowed to have more than two “kinks”, i.e. two *gauche*  $\pm 60^\circ$  torsional angles in an otherwise all-*anti* (all-*trans*) chain. This limitation to two “kinks” has been chosen to exclude unnecessary generation of conformers that are known beforehand not to give significant contributions. For instance, it is well known that the most preferred conformer for a linear carbon chain is the all-*anti* conformer. For the very flexible hydroxyl group torsional angles, i.e. torsional angles defined by the atoms C–C–O–H, the *anti* and the two *gauche* positions were taken into account. These selection criteria for generating the entire set of significant molecular conformers were applied to all the molecules mentioned, i.e. alkanes, alcohols and ethers, with the exception of cyclic molecules. In Fig. 1 the entire set of conformations generated for 2-butanol is shown, all having two *gauche* torsional angles at most. A somewhat similar method of averaging over a number of different conformers of chain molecules was proposed by Arteca and Shaw. [14]

Preliminary molecular mechanical studies of several ring conformations showed that it is reasonable to consider the chair conformer for the six-membered rings and the planar conformation as a starting point for the five-membered rings. In accordance with Dale, [15] large deviations from the planar conformation occur when substituents like methyl or alcohol groups are attached to the five-membered ring. In most cases, this results in the envelope conformer. Methyl and alcohol groups were placed on the ring in axial and equatorial positions, so that all possible isomers were accounted for. For each

**Table 1** Number of initial conformations generated ( $N_{\text{Initial}}$ ) for each individual molecule, and final number of conformers with relative contribution larger than 10% ( $N_{10\%}$ )

Molecule	$N_{\text{Initial}}$	$N_{10\%}$	Molecule	$N_{\text{Initial}}$	$N_{10\%}$
Methane	1	1	Decanol	54	2
Ethane	1	1	Undecanol	60	2
Propane	1	1	Dodecanol	58	2
Butane	2	2	2-Propanol	3	1
Pentane	4	2	2-Butanol	9	3
Cyclopentane	3	1	2-Pentanol	19	3
1,1-Dimethylcyclopentane	15	1	3-Pentanol	19	4
Hexane	7	2	2-Hexanol	33	2
Cyclohexane	6	1	3-Hexanol	33	4
1,1-Dimethylcyclohexane	36	1	2-Heptanol	51	2
Heptane	11	2	2-Octanol	73	2
Octane	16	1	3-Octanol	73	3
Nonane	19	1	4-Octanol	73	2
Decane	26	1	2-Decanol	129	2
Undecane	33	1	4-Decanol	129	2
Dodecane	42	1	1,2-Ethandiol	19	2
Methanol	1	1	1,3-Propanediol	33	4
Ethanol	2	2	1,4-Butanediol	47	3
Propanol	4	2	1,5-Pentanediol	61	3
Butanol	11	3	1,6-Hexanediol	87	4
Pentanol	28	2	1,7-Heptanediol	105	4
Hexanol	40	2	1,8-Octanediol	139	4
Heptanol	33	2	1,9-Nonanediol	172	4
Octanol	40	2	1,10-Decanediol	203	4
Nonanol	48	2	1,12-Dodecanediol	279	4
1,2-Propanediol	19	2	2-Ethoxyethanol	33	2
1,2-Butanediol	33	3	1-(2-hydroxy-1-propoxy)-2-propanol	55	1
1,3-Butanediol	33	1	3-Ethoxy-1,2-propanediol	73	4
1,2-Pentanediol	51	4	3-Methoxy-1,2-propanediol	51	2
1,4-Pentanediol	51	2	Isopropoxy-2-propanol	33	2
2,4-Pentanediol	33	2	Cyclopentanol	90	1
1,2-Hexanediol	73	4	Cyclohexanol	216	2
1,5-Hexanediol	73	2	3-Methylcyclopentanol	60	4
2,5-Hexanediol	51	4	1-Methylcyclopentanol	30	2
1,2-Octanediol	129	3	1-Methylcyclohexanol	36	3
1,2-Dodecanediol	289	1	Z-2-Methylcyclohexanol	36	3
1,2,4-Butanetriol	73	1	E-2-Methylcyclohexanol	36	1
1,2,6-Hexanetriol	129	3	Z-3-Methylcyclohexanol	36	2
Butyl-ethyl-ether	33	4	E-3-Methylcyclohexanol	36	4
Dimethylether	1	1	Z-4-Methylcyclohexanol	36	3
Dipentyl-ether	97	3	E-4-Methylcyclohexanol	36	2
Ethyl-methyl-ether	3	1	3,5-Dimethylcyclohexanol	48	1
Ethyl-propyl-ether	19	1	3,4-Dimethylcyclohexanol	48	4
Methyl-propyl-ether	9	1	2,6-Dimethylcyclohexanol	48	3
Dibutyl-ether	73	3	1,4-Cyclohexandiol	72	4
Tertbutyl-methyl-ether	3	1	<i>trans</i> -1,2-Cyclohexanediol	36	2
Diisopropyl-ether	9	1			
Diisobutyl-ether	33	2			
2-Butoxy-ethanol	73	3			
1-Butoxy-2-propanol	73	3			

ring conformer, the hydroxyl group torsional angles were arranged to give the same three possibilities as for the linear chains. Finally, the energetic differences that may arise due to ring puckering will also have to be accounted for. Ring puckering in a substituted five-membered ring may be viewed as placing the substituents successively on the individual carbon atom in the ring, where one fixed atom deviates from the planar geometry. Having one fixed atom moved slightly to create the envelope conformer, and placing substituents on each of the five ring carbon atoms in turn, thus corresponds to ring puckering. This is illustrated in Fig. 2. Hence, substituents

were placed on all possible sites in the ring, except for those equivalent under symmetry.

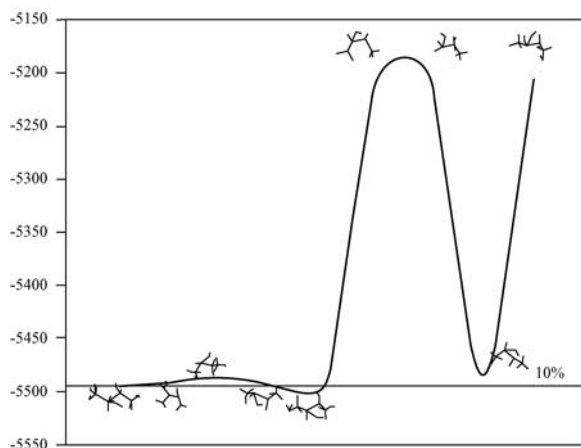
All of the above mentioned molecules and molecular conformers were generated and energy minimized under similar conditions with the molecular mechanics program CFF (consistent force field). [16, 17, 18] The parameter set PEF95SAC [19] was used for these calculations. In this force field, bonded interactions are treated with Morse functions and non-bonded interactions with the Lennard-Jones 12–6 potential, and a Coulomb term. [20] Hence, the potential energy function is as follows:

**Table 2** Energetic descriptors. Non-bonded, Coulombic, van der Waals, the total potential and Gibbs free energies calculated with CFF using the PEF95SAC parameter set. The B3LYP/6-31++g using Gaussian 98 are also listed. The energy unit for this is Hartree

Molecule	$E_{\text{Non-bonded}}$ (kJ mol <sup>-1</sup> )	$E_{\text{Coulomb}}$ (kJ mol <sup>-1</sup> )	$E_{\text{vdW}}$ (kJ mol <sup>-1</sup> )	$E_{\text{Total}}$ (kJ mol <sup>-1</sup> )	$G_{\text{Total}}$ (kJ mol <sup>-1</sup> )	$E_{\text{B3LYP}}$ (a.u.)
Methane	0.000	0.000	0.000	-1,700.378	-1,641.063	-40.51299
Ethane	47.250	45.554	1.696	-2,872.013	-2,744.784	-79.81608
Propane	20.107	13.448	6.659	-4,118.011	-3,922.100	-119.12129
Butane	31.870	21.285	10.585	-5,324.413	-5,059.613	-158.42622
Pentane	35.734	21.238	14.496	-6,539.439	-6,205.734	-197.73138
Cyclopentane	16.489	6.437	10.052	-6,005.103	-5,776.746	-196.50936
1,1-Dimethylcyclopentane	-116.495	-148.049	31.554	-8,593.357	-8,126.260	-275.12762
Hexane	39.673	21.330	18.343	-7,754.164	-7,351.524	-237.03639
Cyclohexane	39.096	9.803	29.292	-7,271.645	-6,904.751	-235.82911
1,1-Dimethylcyclohexane	-95.719	-144.301	48.581	-9,836.817	-9,328.796	-314.43903
Heptane	43.620	21.446	22.175	-8,968.945	-8,497.377	-276.34153
Octane	47.368	21.330	26.038	-10,184.309	-9,643.868	-315.64675
Nonane	51.267	21.402	29.865	-11,399.147	-10,789.779	-354.95189
Decane	55.149	21.458	33.691	-12,614.003	-11,935.708	-394.25699
Undecane	59.021	21.506	37.516	-13,828.869	-13,081.646	-433.56208
Dodecane	62.884	21.544	41.339	-15,043.745	-14,227.595	-472.86717
Methanol	36.429	34.928	1.500	-2,033.302	-1,964.313	-115.69101
Ethanol	-16.450	-20.552	4.102	-3,305.095	-3,168.986	-155.00271
Propanol	38.817	31.564	7.253	-4,467.970	-4,263.452	-194.30793
Butanol	26.178	14.898	11.280	-5,699.001	-5,425.438	-233.61273
Pentanol	29.935	14.662	15.273	-6,914.423	-6,571.965	-272.91817
Hexanol	33.824	14.696	19.128	-8,129.266	-7,717.887	-312.22331
Heptanol	37.527	14.579	22.948	-9,344.291	-8,863.990	-351.52842
Octanol	41.310	14.536	26.774	-10,559.240	-10,010.015	-390.83353
Nonanol	45.050	14.459	30.590	-11,774.233	-11,156.084	-430.13864
Decanol	48.825	14.414	34.411	-12,989.190	-12,302.117	-469.44375
Undecanol	52.586	14.358	38.228	-14,204.163	-13,448.165	-508.74885
Dodecanol	56.367	14.319	42.048	-15,419.117	-14,594.193	-548.05394
2-Propanol	-138.118	-150.823	12.704	-4,645.642	-4,440.197	-194.31372
2-Butanol	-41.360	-52.670	11.310	-5,765.382	-5,492.401	-233.61885
2-Pentanol	-61.857	-77.655	15.797	-7,005.850	-6,663.459	-272.92393
3-Pentanol	48.984	35.413	13.571	-6,892.464	-6,550.358	-272.92323
2-Hexanol	-58.071	-77.898	19.827	-8,221.142	-7,809.742	-312.22867
3-Hexanol	27.714	10.562	17.152	-8,132.588	-7,721.338	-312.22830
2-Heptanol	-54.615	-78.272	23.657	-9,436.457	-8,956.101	-351.43429
2-Octanol	-51.022	-78.474	27.452	-10,651.599	-10,102.331	-390.83942
3-Octanol	35.504	10.868	24.636	-10,562.872	-10,013.795	-390.83876
4-Octanol	12.149	-12.623	24.772	-10,587.450	-10,038.260	-390.83875
2-Decanol	-43.791	-78.877	35.085	-13,081.849	-12,394.723	-469.44966
4-Decanol	19.520	-12.892	32.412	-13,017.619	-12,330.576	-469.44902
1,2-Ethanediol	58.244	54.313	3.932	-3,599.918	-3,454.007	-230.18790
1,3-Propanediol	-17.059	-23.360	6.301	-4,893.719	-4,679.698	-269.49566
1,4-Butanediol	25.115	14.681	10.434	-6,068.537	-5,786.145	-308.79959
1,5-Pentanediol	28.726	12.744	15.983	-7,284.452	-6,933.251	-348.10453
1,6-Hexanediol	30.088	9.959	20.129	-8,501.882	-8,081.568	-387.40983
1,7-Heptanediol	33.266	9.115	24.151	-9,717.471	-9,228.174	-426.71491
1,8-Octanediol	36.103	8.186	27.918	-10,933.386	-10,375.167	-466.02015
1,9-Nonanediol	39.579	7.770	31.810	-12,148.647	-11,521.488	-505.32514
1,10-Decanediol	42.884	7.295	35.589	-13,364.077	-12,667.998	-544.63042
1,12-Dodecanediol	50.004	6.778	43.226	-15,794.419	-14,960.493	-623.24065
1,2-Propanediol	26.587	18.363	8.223	-4,849.589	-4,635.509	-269.50020
1,2-Butanediol	94.970	85.416	9.554	-5,997.590	-5,714.039	-308.80441
1,3-Butanediol	-124.662	-135.496	10.834	-6,220.293	-5,937.424	-308.80786
1,2-Pentanediol	72.640	59.345	13.295	-7,238.404	-6,885.565	-348.10920
1,4-Pentanediol	-59.636	-75.469	15.833	-7,372.770	-7,021.864	-348.11100
2,4-Pentanediol	-228.897	-250.823	21.926	-7,538.162	-7,182.556	-348.11743
1,2-Hexanediol	75.132	58.112	17.020	-8,454.710	-8,032.917	-387.41448
1,5-Hexanediol	-58.509	-78.206	19.696	-8,590.693	-8,170.761	-387.41639
2,5-Hexanediol	-150.352	-170.238	19.885	-8,681.303	-8,260.456	-387.42406
1,2-Octanediol	81.358	56.679	24.716	-10,886.287	-10,326.831	-466.02487
1,2-Dodecanediol	95.045	54.946	40.099	-15,748.813	-14,913.734	-623.24605
1,2,4-Butanetriol	17.716	9.274	8.441	-6,446.321	-6,153.373	-383.99389
1,2,6-Hexanetriol	77.645	60.230	17.415	-8,820.063	-8,392.078	-462.60094
Butyl-ethyl-ether	1.723	-15.926	17.649	-8,168.438	-7,761.530	-312.21757
Dimethylether	31.605	22.638	8.966	-3,265.037	-3,131.198	-154.98544
Dipentyl-ether	50.953	19.324	31.635	-12,994.426	-12,312.128	-469.43817

**Table 2** (continued)

Molecule	$E_{\text{Non-bonded}}$ (kJ mol <sup>-1</sup> )	$E_{\text{Coulomb}}$ (kJ mol <sup>-1</sup> )	$E_{\text{vdW}}$ (kJ mol <sup>-1</sup> )	$E_{\text{Total}}$ (kJ mol <sup>-1</sup> )	$G_{\text{Total}}$ (kJ mol <sup>-1</sup> )	$E_{\text{B3LYP}}$ (a.u.)
Ethyl-methyl-ether	-0.104	-10.219	10.115	-4,515.583	-4,314.077	-194.29694
Ethyl-propyl-ether	14.252	0.422	13.831	-6,938.429	-6,600.681	-272.91314
Methyl-propyl-ether	53.765	40.833	12.932	-5,679.897	-5,409.901	-233.60196
Dibutyl-ether	43.742	19.693	24.049	-10,564.181	-10,019.742	-390.82778
Tertbutyl-methyl-ether	-267.080	-296.911	29.831	-7,205.963	-6,864.244	-272.91292
Diisopropyl-ether	-266.059	-288.825	22.767	-8,426.412	-8,016.640	-312.22569
Diisobutyl-ether	51.365	23.493	27.871	-10,555.133	-10,009.541	-390.83034
2-Butoxy-ethanol	65.601	47.672	17.929	-8,474.811	-8,057.936	-387.40290
1-Butoxy-2-propanol	32.398	12.298	20.100	-9,725.978	-9,239.550	-426.71519
2-Ethoxyethanol	27.129	15.823	11.306	-6,076.770	-5,797.550	-308.79314
1-(2-hydroxy-1-propoxy)-2-propanol	31.785	14.715	17.070	-8,876.425	-8,450.565	-462.60167
3-Ethoxy-1,2-propanediol	78.508	64.206	14.302	-7,624.787	-7,254.212	-423.29542
3-Methoxy-1,2-propanediol	119.444	105.960	13.485	-6,353.193	-6,023.333	-383.98448
Isopropoxy-2-propanol	-114.480	-134.759	20.352	-8,648.769	-8,232.244	-387.41425
Cyclopentanol	-20.001	-32.687	12.686	-6,436.493	-6,141.718	-271.70739
Cyclohexanol	-2.481	-32.801	30.320	-7,681.766	-7,306.513	-311.02269
3-Methylcyclopentanol	-75.435	-93.032	17.597	-7,709.519	-7,342.518	-311.01522
1-Methylcyclopentanol	-140.625	-159.644	19.019	-7,772.453	-7,406.692	-311.01899
1-Methylcyclohexanol	-124.759	-162.247	37.488	-9,019.807	-8,575.240	-350.33170
Z-2-Methylcyclohexanol	56.686	25.277	31.409	-8,839.043	-8,394.392	-350.32727
E-2-Methylcyclohexanol	50.328	14.634	35.695	-8,846.038	-8,401.186	-350.32864
Z-3-Methylcyclohexanol	-59.739	-96.740	37.001	-8,957.592	-8,513.031	-350.32942
E-3-Methylcyclohexanol	-56.621	-91.352	34.731	-8,951.847	-8,506.753	-350.32738
Z-4-Methylcyclohexanol	-30.625	-65.353	34.728	-8,926.302	-8,481.179	-350.32774
E-4-Methylcyclohexanol	-35.701	-72.407	36.705	-8,933.325	-8,488.932	-350.32941
3,5-Dimethylcyclohexanol	-124.263	-166.128	41.866	-10,235.002	-9,720.421	-389.63261
3,4-Dimethylcyclohexanol	-48.474	-92.530	44.056	-10,153.257	-9,636.762	-389.62980
2,6-Dimethylcyclohexanol	101.973	66.390	35.583	-10,005.538	-9,491.134	-389.63080
1,4-Cyclohexanediol	-23.931	-53.196	29.265	-8,071.914	-7,687.845	-386.21519
trans-1,2-Cyclohexanediol	108.815	81.260	27.555	-7,938.434	-7,553.894	-386.21594

**Fig. 1** Energetic differences among generated conformations of 2-butanol. The 10% line indicates the selection criteria of conformers. The Gibbs free energy values shown are in kJ mol<sup>-1</sup>

$$E = E_{\text{bonded}} + E_{\text{non-bonded}} + E_{\text{correction}}$$

where

$$E_{\text{bonded}} = \sum_{\text{bonds}} D_c \left[ e^{-2\alpha(b-b_0)} - 2e^{-\alpha(b-b_0)} \right]$$

$$E_{\text{non-bonded}} = \sum_{i < j} \left[ \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6} + \frac{e_i e_j}{Dr} \right]$$

**Fig. 2** Approach used to model ring puckering in 5-membered rings

and

$$E_{\text{correction}} = \sum_{\text{torsions}} \frac{1}{2} K_{\Phi} (1 + \cos k\Phi) + \sum_{\text{angles}} \frac{1}{2} K_{\theta} (\theta - \theta_0)^2$$

In these equations, the bond lengths,  $b$ , the inter-atomic distances,  $r$ , the valence angles,  $\theta$ , and the torsional angles,  $\phi$ , are to be optimized during energy minimization, while the other variables are fixed parameters in the parameter set PEF95SAC.  $D$  is the dielectric constant, set to a constant value of 2. The use of  $D=2$  is a pragmatic choice that introduces an averaged medium, where electrostatic interactions give rise to significant contributions without being too dominant. [20] For minimization, we perform up to 1,000 iterations with the BFGS [21] variable metric method, followed by up to 50 Newton-Raphson steps until a strict convergence criterion of  $10^{-4}$  kJ mol<sup>-1</sup> Å<sup>-1</sup> is reached in the maximum force component. If convergence was not reached at this stage, the minimized molecular structure was carefully inspected and either subjected to further minimization or, in case the structure had changed into another conformer, discarded. Dipole moments and thermodynamic properties,



**Table 3** Electronic descriptors. Dipole moments ( $\mu$ ) calculated with CFF and Gaussian 98, and polarizabilities  $\alpha$  calculated with PQS

Molecule	$\mu$ CFF (Debye)	$\mu$ Gaussian (Debye)	$\alpha$ PQS (a.u.)
Methane	0.000	0.000	13.947
Ethane	0.000	0.000	25.563
Propane	0.084	0.082	36.538
Butane	0.037	0.013	48.687
Pentane	0.116	0.074	60.233
Cyclopentane	0.000	0.000	55.310
1,1-Dimethylcyclopentane	0.290	0.060	77.780
Hexane	0.016	0.005	71.835
Cyclohexane	0.000	0.000	66.264
1,1-Dimethylcyclohexane	0.297	0.140	88.615
Heptane	0.134	0.073	83.453
Octane	0.000	0.000	95.217
Nonane	0.123	0.076	106.941
Decane	0.000	0.000	118.707
Undecane	0.127	0.076	130.483
Dodecane	0.000	0.000	142.272
Methanol	2.745	2.307	17.866
Ethanol	2.684	2.219	29.467
Propanol	2.419	2.154	40.861
Butanol	2.483	2.167	52.278
Pentanol	2.366	2.152	63.894
Hexanol	2.579	2.193	75.493
Heptanol	2.587	2.152	87.152
Octanol	2.907	2.195	98.869
Nonanol	3.040	2.152	110.600
Decanol	3.422	2.196	122.375
Undecanol	3.637	2.152	134.145
Dodecanol	4.048	2.196	145.938
2-Propanol	2.214	2.242	40.936
2-Butanol	2.590	2.122	52.163
2-Pentanol	2.661	2.080	63.679
3-Pentanol	2.498	2.089	63.370
2-Hexanol	3.021	2.034	75.312
3-Hexanol	2.590	2.043	74.909
2-Heptanol	3.219	1.999	86.952
2-Octanol	3.561	2.022	98.657
3-Octanol	3.098	2.009	98.299
4-Octanol	2.618	2.004	98.320
2-Decanol	4.120	2.019	122.141
4-Decanol	2.964	1.996	121.813
1,2-Ethandiol	0.447	0.554	32.904
1,3-Propanediol	1.846	1.264	44.196
1,4-Butanediol	1.711	1.618	55.855
1,5-Pentanediol	4.118	2.547	67.524
1,6-Hexanediol	1.205	1.334	79.139
1,7-Heptanediol	4.054	2.644	90.813
1,8-Octanediol	1.252	1.370	101.285
1,9-Nonanediol	4.076	2.661	114.191
1,10-Decanediol	1.267	1.377	126.229
1,12-Dodecanediol	1.271	1.377	149.628
1,2-Propanediol	1.171	0.648	44.377
1,2-Butanediol	1.618	0.320	55.520
1,3-Butanediol	1.008	0.423	55.637
1,2-Pentanediol	2.376	0.355	66.914
1,4-Pentanediol	3.111	3.058	67.424
2,4-Pentanediol	3.473	2.358	67.102
1,2-Hexanediol	3.137	0.389	78.489
1,5-Hexanediol	3.272	3.064	79.020
2,5-Hexanediol	1.629	1.847	78.679
1,2-Octanediol	4.780	0.377	101.922
1,2-Dodecanediol	8.263	0.437	149.116
1,2,4-Butanetriol	1.859	1.839	59.040
1,2,6-Hexanetriol	3.888	1.955	82.320
Butyl-ethyl-ether	3.629	1.526	76.048
Dimethylether	3.206	1.864	29.563
Dipentyl-ether	3.473	1.313	123.018
Ethyl-methyl-ether	3.473	1.702	41.266
Ethyl-propyl-ether	3.492	1.456	64.590

Table 3 (continued)

Molecule	$\mu$ CFF (Debye)	$\mu$ Gaussian (Debye)	$\alpha$ PQS (a.u.)
Methyl-propyl-ether	3.405	1.600	52.794
Dibutyl-ether	3.691	1.441	99.516
Tertbutyl-methyl-ether	2.746	1.709	63.485
Diisopropyl-ether	3.202	1.621	75.390
Diisobutyl-ether	3.686	1.206	98.809
2-Butoxy-ethanol	2.314	0.997	79.700
1-Butoxy-2-propanol	2.318	0.829	91.128
2-Ethoxyethanol	1.295	1.016	56.511
1-(2-hydroxy-1-propoxy)-2-propanol	1.767	2.755	82.914
3-Ethoxy-1,2-propanediol	2.700	2.782	71.512
3-Methoxy-1,2-propanediol	2.892	2.381	59.821
Isopropoxy-2-propanol	1.221	0.795	79.042
Cyclopentanol	2.473	2.042	59.021
Cyclohexanol	2.253	2.239	70.067
3-Methylcyclopentanol	2.769	2.437	70.451
1-Methylcyclopentanol	2.378	2.048	70.068
1-Methylcyclohexanol	2.348	2.008	80.954
Z-2-Methylcyclohexanol	2.845	1.978	80.550
E-2-Methylcyclohexanol	2.081	2.309	81.273
Z-3-Methylcyclohexanol	2.270	2.243	81.614
E-3-Methylcyclohexanol	2.378	2.010	81.302
Z-4-Methylcyclohexanol	2.796	2.147	80.858
E-4-Methylcyclohexanol	1.909	2.177	81.548
3,5-Dimethylcyclohexanol	1.885	2.107	93.845
3,4-Dimethylcyclohexanol	1.891	2.291	91.885
2,6-Dimethylcyclohexanol	2.521	1.917	92.442
1,4-Cyclohexandiol	2.370	2.239	73.431
<i>trans</i> -1,2-Cyclohexanediol	2.171	1.715	73.154

including the Gibbs free energies at 298.16 K, were calculated. The thermodynamic properties were calculated using standard statistical mechanics formulae, assuming ideal gas behavior, where the vibrational contribution to the energy is calculated with the Einstein relation. The molecules are treated as rigid rotors and as coupled harmonic oscillators. [18]

Using the calculated Gibbs free energies,  $G_i$ , the relative probabilities,  $p_i$ , of all the conformers within a molecule can be calculated with the Boltzmann distribution at a given temperature:

$$p_i = \frac{e^{-\frac{\Delta G_i}{RT}}}{\sum_i e^{-\frac{\Delta G_i}{RT}}}$$

where  $\Delta G_i = G_i - G_1$ ,  $T$  is the temperature and  $R$  is the gas constant.

In this work, the relative probabilities were calculated at 298.16 K. From the initial set of conformers constructed, the conformers that either did not converge or had a relative contribution of less than 1% were excluded. After renormalization, only conformers with a weight of more than 10% were considered important for further use in the QSPR model development. In order to verify this, a number of QSPR models for a range of different compounds were developed by excluding all conformers having relative contributions of less than 1% and 10%, respectively. It was seen that the relative values, i.e. the differences in values calculated with the QSPR models versus experimental values were similar for every compound. The change in the absolute values from using the

10% cut instead of 1% therefore disappeared in the regression of the QSPR model. Although 10% is a somewhat arbitrary choice, it is obvious that choosing an even higher cutoff value would exclude conformers which contribute significantly to the physical property of interest. This is also illustrated in Fig. 1 where all conformations of 2-butanol generated initially are shown with respect to the calculated Gibbs free energy after energy minimization. As seen from Fig. 1 and Table 1, nine conformations were generated initially. After energy minimization it turned out that one of the conformers had changed into another. Furthermore, five of the conformers generated had a very high Gibbs free energy, rendering their relative contributions negligible. Hence three conformers remained when using 10% as the cut-off value. From the set of remaining conformers, no further exclusion was done.

The Codessa program can use Gaussian output files as input for calculation of several quantum chemical descriptors like ionization energies, HOMO–LUMO gaps etc. Therefore, each of the selected conformers of the final Boltzmann-weighted set was geometry optimized, using a hybrid functional, on the B3LYP/6-31++G level [22] with the Gaussian 98 program. [23] As input to these calculations, the Cartesian coordinates of the energy-minimized molecular conformers calculated by CFF were used and optimized further until convergence was reached. Besides calculation of the vibrational frequencies, population analysis, including NBO (natural bond orbital) analysis, was also carried out such that as many descriptors as possible could be calculated with the Codessa program.

Finally, calculations of molecular polarizabilities were performed. Bulk properties, like density or boiling point, are known to depend significantly on London dispersion forces, when hydrogen bonding is not taken into account. Molecular dipole moments and polarizabilities have a direct influence on this type of interaction. [4, 24] The polarizability can be calculated in Gaussian 98. However, these calculations have several deficiencies regarding the polarizability. The basis set used is reasonable with respect to geometry optimization, but with respect to polarizabilities it is too small. Calculation of polarizabilities requires a large basis set with many diffuse functions. [25, 26] Furthermore, these calculations are performed with the DFT method, which gives polarizabilities that in general are too high compared to experiment. [26, 27] According to these references, the reason is that common DFT functionals give an incorrect description of the asymptotic behavior of the virtual orbitals. Perdew et al. [28] have explained this disability to be caused by improper description of the discontinuities in the exchange-correlation potential. [27] Therefore, we have calculated static polarizabilities at the Hartree–Fock level as the second derivatives of the SCF energy with respect to the electric field components. The PQS (parallel quantum solutions) program [29] was used in these calculations. The polarizability is calculated analytically using CPHF (coupled perturbed Hartree–Fock) theory. In CPHF theory, each of the terms in the HF equation is subjected to a perturbation, giving a system of equations that is solved iteratively until self-consistency is reached. [30] The second derivative of the energy with respect to the field (the perturbed orbitals) results in a polarizability tensor, the trace of which gives the molecular polarizability:

$$\alpha = \frac{1}{3} (\alpha_{XX} + \alpha_{YY} + \alpha_{ZZ})$$

As input to these calculations, the Cartesian coordinates from the previous DFT calculations in Gaussian 98 were used. Since the geometry was already optimized using the electron-correlated DFT method, it would be pointless to optimize the geometry further at the more inaccurate HF level. Therefore, only single point calculations on the RHF/6-311++G\*\* level were performed. In a few cases, especially for the large linear and cyclic molecules, like 1,12-dodecanediol and 2,6-dimethylcyclohexanol, the convergence criterion of  $10^{-5}$  for the first order density could not be reached. To obtain convergence in these cases, a lighter criterion of  $10^{-4}$  was used. Preliminary test calculations proved that this criterion did not alter the polarizability tensor and hence the molecular polarizability significantly.

#### Other descriptors added to the local database

The UNIFAC [31, 32] group contributions method uses parameters that depend on the van der Waals volume,  $r$ , and van der Waals surface area,  $q$ . The UNIFAC param-

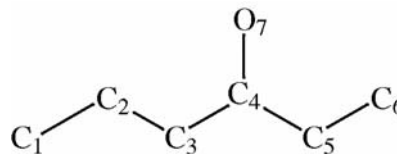


Fig. 3 Key atoms of the 3-hexanol molecule

eters are commonly used in calculation of phase equilibria, and are based on a method proposed by Bondi. [33] Preliminary studies have shown that using  $q$  and  $r$  as descriptors in QSPR models results in regression models that are in good agreement with experimental data. However, one major disadvantage is that these parameters fail to represent branching properly, due to the group additivity principle. As an example of this, the van der Waals volume or surface area for the molecules 2-pentanol and 3-pentanol would be equal, i.e.  $r_{\text{UNIFAC}}=4.5979$  and  $q_{\text{UNIFAC}}=4.204$ . Even despite the obvious difference between the much more compact molecule 1,1-dimethylcyclohexane and the stretched-out molecule octane, these values are still fairly equal as their respective surface areas are  $q_{\text{UNIFAC}}=4.936$  and  $q_{\text{UNIFAC}}=4.396$ . The  $r$  and  $q$  parameters can be obtained by consulting the UNIFAC parameter table. [31]

In the same preliminary study, this problem has been diminished by incorporating the topological descriptors  $P_2$ ,  $P_3$  and  $P_4$  into the regression model. For these descriptors, only carbon and oxygen atoms are accounted for, and all hydrogen atoms are neglected.  $P_2$  represents the number of C–C (or C–O) bonds in the molecule, and  $P_3$  is the number of times a line can be drawn between three connected carbon (or oxygen) atoms. Finally,  $P_4$  is the number of times a line can be drawn between four connected carbon or oxygen atoms, as explained in [5] and [34]. An illustrative example of this is shown in Fig. 3.

It is readily seen that the number of bonds in the hydrogen-depleted molecule is 6, i.e.  $P_2=6$ .  $P_3$  also equals 6 as there are six possibilities of connecting three atoms; C<sub>1</sub>–C<sub>3</sub>, C<sub>2</sub>–C<sub>4</sub>, C<sub>3</sub>–C<sub>5</sub>, C<sub>4</sub>–C<sub>6</sub>, O<sub>7</sub>–C<sub>3</sub> and O<sub>7</sub>–C<sub>5</sub>. Similarly,  $P_4=5$  as there are five ways of connecting four atoms, C<sub>1</sub>–C<sub>4</sub>, C<sub>2</sub>–C<sub>5</sub>, C<sub>3</sub>–C<sub>6</sub>, O<sub>7</sub>–C<sub>2</sub> and O<sub>7</sub>–C<sub>6</sub>. These topological values can easily be evaluated for other molecules. The descriptors are mentioned here because they have been added to the existing database of descriptors in Codessa. The  $P_2$ ,  $P_3$  and  $P_4$  have been shown to be useful as secondary parameters in QSPR modeling of boiling points, as they account for branching in alkanes and alcohols. [35]

These descriptors depend on molecular constitution only, meaning that considering several molecular conformers will not have any effect on their value for the individual molecule. They can easily be evaluated by inspecting the molecular structure.

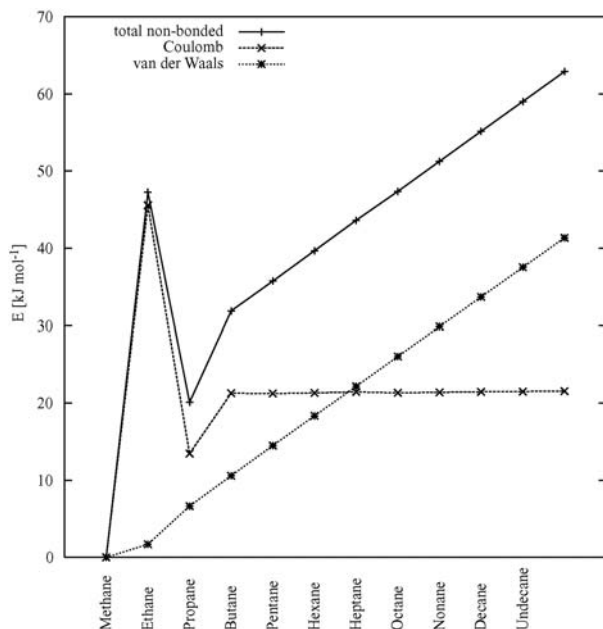
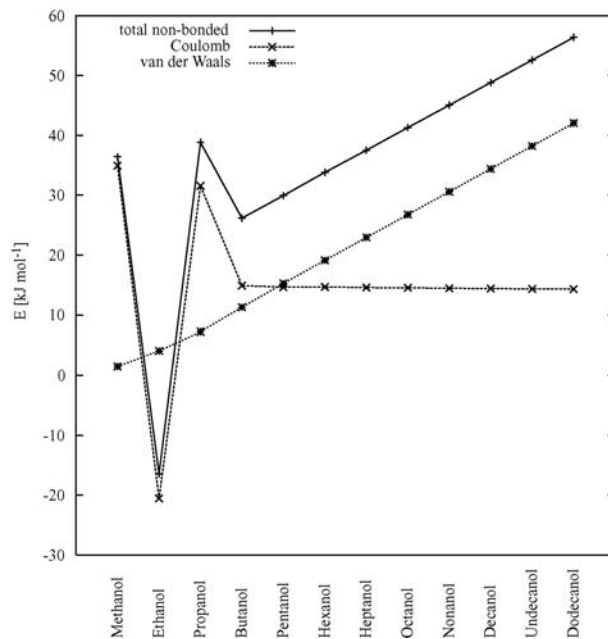
## Results and discussion

In Table 2, Boltzmann-averaged descriptor values obtained from the molecular mechanical calculations using



**Table 4** Total non-bonded energies and van der Waals energies (in  $\text{kJ mol}^{-1}$ ) for all conformers with relative contributions larger than 10% for 2-butanol, 2-pentanol and butyl-ethyl-ether

2-Butanol			2-Pentanol			Butyl-ethyl-ether		
	$E_{\text{non-bonded}}$	$E_{\text{vdW}}$		$E_{\text{non-bonded}}$	$E_{\text{vdW}}$		$E_{\text{non-bonded}}$	$E_{\text{vdW}}$
aa	-40.159	10.957	aaa	-61.351	14.937	aaaa	1.554	17.761
ag	-41.898	11.290	gaa	-62.146	17.676	gaaa	0.721	17.647
ag'	-42.507	11.957	ag'g'	-62.207	15.070	agaa	4.666	15.780
						aaga	1.317	18.780

**Fig. 4** Van der Waals, Coulomb and total non-bonded energies of alkanes plotted against increasing molecular size**Fig. 5** Van der Waals, Coulomb and total non-bonded energies of alcohols plotted against increasing molecular size

CFF are given. Among these data are the calculated non-bonded, Coulomb and van der Waals energies. The total potential energies and the Gibbs free energies calculated with CFF at 298.16 K, as well as the Boltzmann-averaged total energies calculated with Gaussian 98 at the B3LYP/6-31++G level are also listed.

In Table 3 the dipole moments calculated using CFF and Gaussian 98, respectively, are shown. The dipole moments obtained with Gaussian 98 are automatically used as descriptors in Codessa, provided the output files from Gaussian 98 are available. The polarizabilities calculated with PQS using the CPHF method mentioned above, are also listed in Table 3.

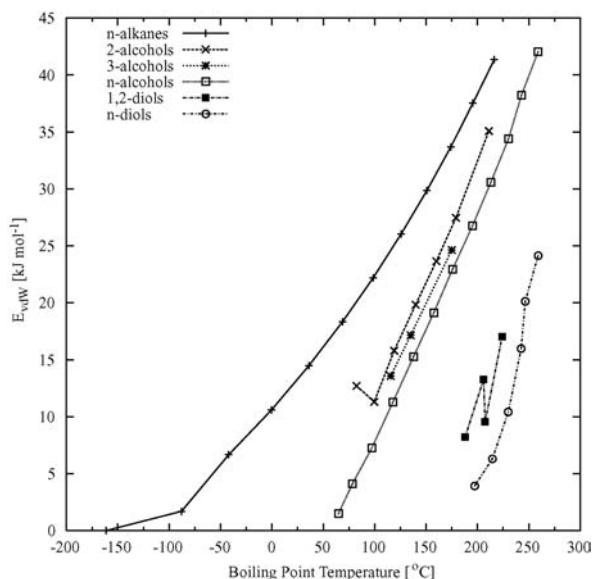
To illustrate the effect of considering more than one conformer for each compound, the total non-bonded energies and van der Waals energies are used as an example. In Table 4 the calculated values of these properties are given for all conformers with contribution larger than 10% for three arbitrary molecules: 2-butanol, 2-pentanol and butyl-ethyl-ether. It can be seen that variation in the van der Waals energies is indeed significant and the difference between conformers of one molecule ranges from approximately 3 to 17% of the total energy for the

compounds shown. The internal difference between conformers is also significant compared to the difference in energy when an extra carbon atom is added, as can be seen for the compounds 2-butanol and 2-pentanol.

The van der Waals energy has proved to be a particularly good descriptor in QSPR modeling as it has been used as a descriptor for modeling of melting point [9, 11] and heat of fusion. [11] Also the total non-bonded energy appears to be a useful descriptor. [9]

In the process of developing an actual QSPR model, the Boltzmann-averaged values are not used directly. Instead the descriptor value for each conformer is taken directly into account with its relative probability during the statistical fitting process in Codessa. This also allows for statistical weighting of the properties calculated from the Gaussian output files.

The total non-bonded, Coulomb and van der Waals energies for alkanes and alcohols are shown in Figs. 4 and 5, respectively. With respect to the Coulomb energies, it is seen in Fig. 4 for the non-polar alkanes, which do not have the ability to form hydrogen bonds, the energy increases greatly for ethane. This increase is inherent in the model used within the CFF program, as only 1–4

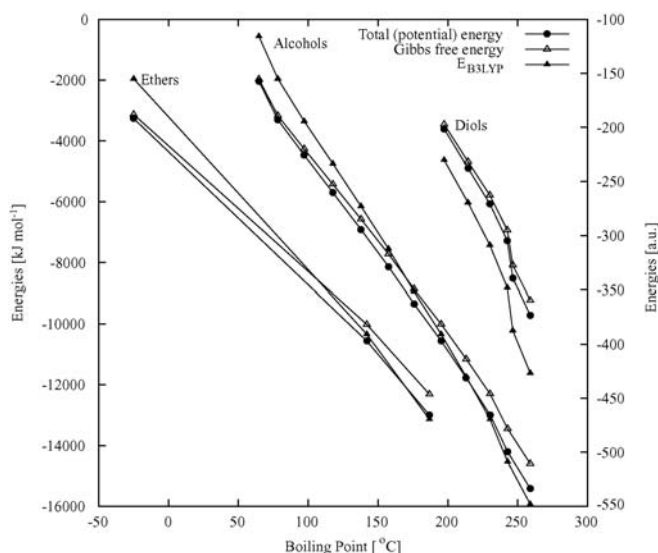


**Fig. 6** Van der Waals energies of *n*-alkanes (methane–dodecane), *n*-alcohols (methanol–dodecanol), 2-alcohols (2-propanol–2-decanol, except 2-nonanol), 3-alcohols (3-pentanol, 3-hexanol and 3-octanol), *n*-diols (1,2-ethanediol–1,7-heptanediol) and 1,2-diols (1,2-ethanediol–1,2-hexanediol), plotted against the respective experimental boiling points

non-bonded interactions and higher are treated. [20] For 1–3 interactions (valence angles), non-bonded forces are not taken into account, and therefore the non-bonded energy is zero. Therefore, due to the small size of the ethane molecule, the hydrogen atoms on each of the carbon atoms give the only Coulombic interactions in this molecule, i.e. only H···H interactions are treated. When more carbon atoms are added to the chain, this effect is seen to fall to a constant level, where the Coulomb energy is independent of the chain length.

The van der Waals energies are seen to increase constantly as the chain length increases. Since the total non-bonded energy is the sum of the Coulomb and van der Waals energies, the total non-bonded energy shows a large increase for ethane. While the Coulomb energy is constant, the non-bonded energy increases with the van der Waals energy.

A similar energetic pattern is observed for these three types of energies for the alcohols in Fig. 5. The anomaly in Coulomb energy for ethanol is due to inherent properties of the model selected, i.e. only 1–4 interactions and higher are included. This causes a drop in the Coulombic energy for ethanol, and hence the total non-bonded energy. For this molecule, as opposed to methanol, the non-bonded energy accounts for the interactions between oxygen and the methyl hydrogens, and between the hydroxyl hydrogen attached to oxygen and all other hydrogens and the methyl carbon atoms in the molecule. The hydroxyl hydrogen is highly positively charged. At the same time, the *gauche* conformer contributes appreciably to the population, so that the distances between the hydroxyl hydrogen and the remaining hydrogen at-



**Fig. 7** Total potential energies and Gibbs free energies at 298.16 K calculated with CFF (left axis) and B3LYP/6–31++g energies calculated with Gaussian 98 (right axis) versus the experimental boiling point for alcohols (methanol–decanol), *n*-diols (1,2-ethanediol–1,7-heptanediol) and ethers (dimethyl-, dibutyl- and dipentylether)

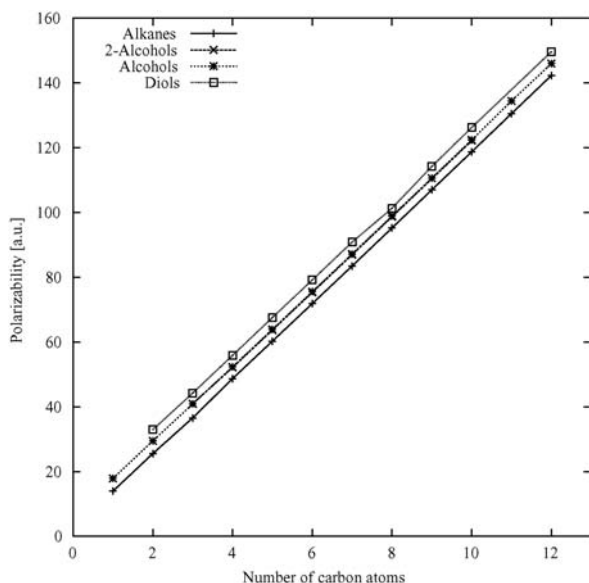
oms are rather short, leading to a large electrostatic contribution to the Coulomb energy for the ethanol molecule. The same behavior is observed for diols and ethers, see Table 2.

Since the van der Waals energies are seen to increase almost linearly for several classes of compounds, they can be a useful descriptor in QSPR modeling. As an example, the different descriptors have been compared to the boiling points, see Figs. 6, 7, 8 and 9, a well-studied physical property as far as QSPR is concerned. These graphs may be useful in elucidating why these descriptors may prove useful in subsequent QSPR modeling. Several articles have been published concerning QSPR modeling of boiling points. [1, 4, 8, 34, 36, 37, 38, 39, 40, 41, 42, 43, 44]

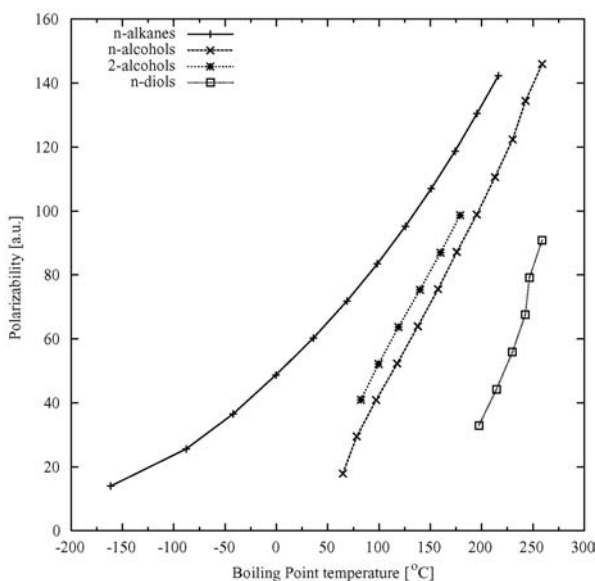
Generally, the boiling point increases as the length of the carbon atom chain increases. The possibilities and occurrence of hydrogen bonding raise, while side chains and branching tend to lower, the boiling point of a molecule, compared to a non-branched molecule with the same constitutional formula.

This pattern is also seen in Fig. 6, where the van der Waals energies of several classes of compounds are plotted against the boiling points of the molecules. For each individual class of compounds, it is seen that as the boiling point increases there is a similar increase in the van der Waals energies. This illustrates that, when more carbon atoms are added to the chain, the van der Waals energies increase in a similar manner to the boiling point.

It is observed that the van der Waals energies increase from alkanes to diols when comparing molecules with



**Fig. 8** Polarizabilities calculated by PQS with increasing molecular size for alkanes (ethane–dodecane), alcohols (methanol–dodecanol), 2-alcohols (2-propanol–2-decanol, except 2-nonanol) and *n*-diols (1,2-ethanediol–1,12-dodecanediol, except 1,11-undecanediol), respectively



**Fig. 9** Polarizabilities of the molecules shown in Fig. 8 plotted against the experimental boiling points

the same number of carbon atoms. For instance, considering the  $C_6$  compounds, the tabulated boiling points [45] are 68.7 °C for hexane, 157.5 °C for hexanol and 246.5 °C for 1,6-hexanediol. The corresponding van der Waals values are 18.344 kJ mol<sup>-1</sup>, 19.128 kJ mol<sup>-1</sup> and 20.129 kJ mol<sup>-1</sup>, showing a good correlation between the van der Waals energies and the boiling points.

Furthermore, in the cases of branched alcohols and diols it is observed that the van der Waals energies are slightly lower than for the corresponding *n*-alcohols and

*n*-diols, respectively. This tendency is also in good accord with the boiling points. For instance, 3-hexanol has a boiling point of 135 °C and a van der Waals energy of 17.152 kJ mol<sup>-1</sup>. This fits well into the pattern of the  $C_6$  compounds mentioned above.

In Fig. 7 the total energies from Table 2 for *n*-alcohols, diols and ethers are shown. Three types of energies are included, the Gibbs free energy calculated by molecular mechanics, the total molecular energy (potential energy) calculated by molecular mechanics and the electron-correlated energy calculated by the quantum chemical DFT method, respectively. These different types of energies follow the same pattern, but of course the numerical values differ. This shows that the molecular mechanics and quantum chemical methods' energy values scale well with one another, which validates the molecular mechanical calculations. The B3LYP energy also includes the atomic energies, whereas in molecular mechanical methods only the binding energy is calculated. In recent work, it has been shown that the difference in molecular and atomic energies correlate well with energies obtained with molecular mechanics. [46] The calculated Gibbs free energies follow a similar pattern to the total potential energies, and some entropic effects are seen. With respect to the boiling point, it is seen that here also the values in general become more negative as the boiling point increases.

In Table 3, calculated molecular dipole moments and polarizabilities are shown. As already mentioned, descriptors like these can provide information about the intermolecular dispersion interactions that have significant importance with respect to the numerical size of bulk physical properties. Dipole moments have been calculated using both molecular mechanical and quantum chemical methods, while polarizabilities are determined by a quantum chemical method only. When comparing the values of the dipole moments, it is seen that there are some differences between the values calculated with the two methods. Obviously, this is due to the theoretical differences of these two methods, and the more empirical molecular mechanics method is of course less accurate in comparison to the electron-correlated quantum chemical calculations. The reason for this is mainly the use of atomic point charges in the molecular mechanical calculations, which is a crude approximation.

In Fig. 8, calculated polarizabilities for several classes of compounds are depicted as a function of the molecular size. For all four classes of compounds, this results in a straight line, which means there is a correlation between polarizability and size and type of molecule.

In Fig. 9 the polarizabilities are plotted with respect to the boiling point, showing that there is also good correlation between polarizability and boiling point. The polarizability increases with increasing carbon atom number within each class of compound, but also an increase in polarizability is seen when oxygen atoms are added to the chain. Finally a slight decrease in polarizability is observed when branching occurs. This overall trend can be accompanied by a few examples. In Table 3 it is seen

that polarizabilities for pentanols decrease in the following order 1-pentanol>2-pentanol>3-pentanol. Similarly, the respective boiling points of these compounds are 137.8 °C, 119 °C and 115 °C. This is also the case for hexanols and octanols, i.e. 1-hexanol>2-hexanol>3-hexanol having boiling points of 157.5 °C, 139.9 °C and 135 °C, and finally 1-octanol>2-octanol>3-octanol with boiling point values of 195 °C, 179 °C and 175 °C. The tendency shown in these examples is not typically reproduced by other descriptors that may correlate well with the boiling point, for instance the molecular size and volume. It should be noticed that the slope for the alkanes differs slightly from the slopes for the alcohols in Fig. 9. This implies that a better QSPR fit could be achieved if alkanes and alcohols were treated in two separate QSPR models, as there are some inherent differences between these two classes of compounds. Despite the deficiencies, the descriptors presented here are physically understandable, and considering the relatively good correlation with the boiling point, they should also be useful and give a good description of other properties.

In a recent paper [47] it was stated that solvent-dependent Boltzmann averaging based on gas-phase calculations would not be appropriate. However, we would like to add to that statement that the usefulness of Boltzmann averaging may be governed by the specific problem and the accuracy one wants to achieve. For instance, Boltzmann weighting of conformers has proved successful in calculation of interaction parameters for use in the UNIQUAC equation for phase equilibria calculations. [13, 48] Due to otherwise tremendous computational costs, these parameters could not have been obtained if Boltzmann averaging over gas-phase conformers had been replaced by modeling of the liquid phase. Hence, we believe that Boltzmann averaging is appropriate in QSPR modeling. The QSPR model is a fitted equation, and a full, perfect description of the system of interest cannot be obtained. Our goal is to use physically sound parameters as descriptors, which relate well to the physical properties of interest.

## Conclusions

In this work additional descriptors for use in the program Codessa have been presented. These descriptors are based on Boltzmann averaging over several molecular conformers. A method of selecting the most important conformers from a large number of possible conformations has been established. These new descriptors are relative energies of different kinds, calculated with molecular mechanics, and Boltzmann-weighted dipole moments and polarizabilities, calculated with both molecular mechanics and quantum chemical methods. The systematic nature of these descriptors is investigated and their general trend is compared to the boiling point of several classes of compounds. Two of these descriptors are particularly interesting, the van der Waals

energy, which accounts for non-ionic repulsive and attractive non-bonded interactions, and the molecular polarizability. These two descriptors generally follow the pattern of the boiling points; in consequence they may be useful in QSPR modeling of other physical and thermodynamic properties.

**Acknowledgments** We would like to thank the Danish Technical Research Council for the financial support to this project and the Danish Natural Science Research Council for providing financial support for a Parallel Quantum Station. Special thanks to Peter Pulay for help concerning calculations of polarizabilities and useful comments on this manuscript. We should also like to thank Krzysztof Wolinski, Jon Baker and Helge Johansen for useful discussions and help concerning computational problems.

## References

- Murugan R, Grendze MP, Toomey Jr JE, Katritzky AR, Karelson M, Lobanov V, Rachwal P (1994) *CHEMTECH* 24:17–23
- Katritzky AR, Lobanov VS, Karelson M (1995) *Chem Soc Rev* 24:279–287
- Kier LB, Hall LH, Frazer JW (1993) *J Chem Inf Comput Sci* 33:143–147
- Kier LB, Hall LH (1976) *Molecular connectivity in chemistry and drug research*. Academic Press, New York
- Wiener H (1947) *J Am Chem Soc* 69:17–20
- Stanton DT, Jurs PC (1990) *Anal Chem* 62:2323–2329
- Karelson M, Lobanov VS, Katritzky AR (1996) *Chem Rev* 96:1027–1044
- Katritzky AR, Mu L, Lobanov VS, Karelson M (1996) *J Phys Chem* 100:10400–10407
- Dyckjær JD, Jónsdóttir SÓ (2002) submitted for publication
- Dyckjær JD, Jónsdóttir SÓ (2002) submitted for publication
- Dyckjær JD, Jónsdóttir SÓ (2002) in preparation
- Katritzky A, Karelson M (2000) Codessa. Semichem Inc, 7204 Mullen, Shawnee, KS 66216, U.S.A.
- Jónsdóttir SÓ, Rasmussen K, Fredenslund A (1994) *Fluid Phase Equilibria* 100:121–138
- Arteca GA, Shaw MD (2001) *J Mol Struct Theochem* 537:181–187
- Dale J (1978) *Stereochemistry and conformational analysis*. Universitetsforlaget, Oslo
- Rasmussen K (1998) CFF – Consistent Force Field, version 17.21. Department of Chemistry, Building 207, Technical University of Denmark, DK-2800 Lyngby, Denmark
- Rasmussen K (1985) *Potential energy functions in conformational analysis*. Lecture notes in chemistry, vol 37. Springer, Berlin Heidelberg New York
- Rasmussen K, Engelsen SB, Fabricius J, Rasmussen B (1993) In Fausto R (ed) *Recent experimental and computational advances in molecular spectroscopy*. NATO-ASI series C, vol 406. Kluwer, Dordrecht, pp 381–419
- Fabricius J, Engelsen SB, Rasmussen K (1997) *J Carbohydr Chem* 16:751–772
- Engelsen SB, Fabricius J, Rasmussen K (1994) *Acta Chem Scand A* 48:548–552
- Stewart JJP (1989) *Comput Chemistry* 13:157–158
- Becke AD (1993) *J Chem Phys* 98:5648–5652
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratman RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu C, Liashenko A, Piskorz P, Komaromi, I, Gomperts R,



- Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson BG, Chen W, Wong MW, Andres JL, Gonzales C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98. Gaussian, Pittsburgh, Pa.
24. Atkins PW (1992) Physical chemistry, 4th edn. Oxford University Press, Oxford
  25. Werner HJ, Meyer W (1976) Mol Phys 31:855–872
  26. McDowell SAC, Amos RD, Handy NC (1995) Chem Phys Lett 235:1–4
  27. Tozer DJ, Handy NC (1998) J Chem Phys 109:10180–10189
  28. Perdew JP, Parr RG, Levy M, Balduz JL, Jose Jr L (1982) Phys Rev Lett 49:1691–1694
  29. PQS (2000) PQS version 2.3, Parallel Quantum Solutions. 2013 Green Acres, Suite A, Fayetteville, AK 72703, U.S.A.
  30. Gerratt J, Mills IM (1968) J Chem Phys 49:1719–1729
  31. Fredenslund A, Sørensen JM (1994) In Sandler SI (ed) Models for thermodynamic and phase equilibria calculations. Chemical industries, vol 52. Marcel Dekker, New York, pp 287–361
  32. Fredenslund A, Gmehling J, Rasmussen P (1977) Vapor–liquid equilibria using UNIFAC. Elsevier, Amsterdam
  33. Bondi A (1968) Physical properties of molecular crystals, liquids and glasses. Wiley, New York
  34. Basak SC, Gute BD, Grunwald GD (1996) J Chem Inf Comput Sci 36:1054–1060
  35. Jónsdóttir SÓ, Dyekjær JD (2002) in preparation
  36. Estrada E (1995) J Chem Inf Comput Sci 35:708–713
  37. Estrada E (1998) J Chem Inf Comput Sci 38:23–27
  38. Hall LH, Kier LB, Murray WJ (1975) J Pharm Sci 64:1974–1977
  39. Ivanciuc O, Ivanciuc T, Balaban AT (1998) Tetrahedron 54:9129–9142
  40. Tetteh J, Suzuki T, Metcalfe E, Howells S (1999) J Chem Inf Comput Sci 39:491–507
  41. Goll ES, Jurs PC (1999) J Chem Inf Comput Sci 39:974–983
  42. Stanton DT, Egolf LM, Jurs PC (1992) J Chem Inf Comput Sci 32:306–316
  43. Koziol J (2001) J Quantum Chem 84:117–126
  44. Katritzky AR, Lobanov VS, Karelson M (1998) J Chem Inf Comput Sci 38:28–41
  45. Dean JA (1985) Lange’s Handbook of Chemistry, 13th edn. McGraw-Hill
  46. Rasmussen K (2002) to be published
  47. Basma M, Sundara S, Çalgan D, Vernali T, Woods RJ (2001) J Comp Chem 22:1125–1137
  48. Jónsdóttir SÓ, Klein RA, Rasmussen K (1996) Fluid Phase Equilibria 115:59–72