Available online at www.sciencedirect.com

SCIENCE DIRECT®

Carbohydrate Research 339 (2004) 269–280

Carbohydrate
RESEARCH

QSPR models for various physical properties of carbohydrates based on molecular mechanics and quantum chemical calculations

Jane Dannow Dyekjær and Svava Ósk Jónsdóttir*

Department of Chemistry, Technical University of Denmark, Building 207, DK-2800 Kgs. Lyngby, Denmark

Received 27 May 2003; revised 5 September 2003; accepted 16 September 2003

Abstract—Quantitative Structure–Property Relationships (QSPR) have been developed for a series of monosaccharides, including the physical properties of partial molar heat capacity, heat of solution, melting point, heat of fusion, glass-transition temperature, and solid state density. The models were based on molecular descriptors obtained from molecular mechanics and quantum chemical calculations, combined with other types of descriptors. Saccharides exhibit a large degree of conformational flexibility, therefore a methodology for selecting the energetically most favorable conformers has been developed, and was used for the development of the QSPR models. In most cases good correlations were obtained for monosaccharides. For five of the properties predictions were made for disaccharides, and the predicted values for the partial molar heat capacities were in excellent agreement with experimental values.

Keywords: QSPR model; Physical properties; Monosaccharides; Disaccharides

1. Introduction

QSPR models are empirical equations, used for estimating various physical or thermodynamic properties of molecules. A QSPR model has the form

$$P = a + b \cdot D_1 + c \cdot D_2 + d \cdot D_3 + \cdots, \qquad (1)$$

where P is the physical property of interest, $a, b, c, \ldots$ are regression coefficients, and $D_1, D_2, D_3, \ldots$ are parameters derived from the molecular structure, so-called descriptors. A variety of different types of descriptors can be used.[1] The simplest types are constitutional or topological descriptors, such as the number of carbon atoms, and parameters describing types and order of the chemical bonds in the molecules. Various geometrical descriptors, including the principal moment of inertia, can also be used. The most important and also the most complicated descriptors, are electrostatic and quantum chemical descriptors. The electrostatic descriptors are parameters, which depend on the charge distribution within the molecule, including the dipole moment. An example of a quantum chemically derived descriptor are the HOMO and LUMO energies.[2]

QSPR models describing saccharide properties have not been found in literature up to now. Several papers concerning the properties of interest to this work, but for other classes of compounds, have been published. Liu et al.[3] and Ivanciuc et al.[4] have developed models for the heat capacity of alkanes at 300 K. QSPR models for the melting points for a number of classes of organic compounds can be found in the literature.[5–10] Katritzky et al.[11] have developed QSPR models for the glass-transition temperatures of large industrial polymers, MW 400–9500. QSPR models for liquid densities of several classes of organic compounds have been published.[7,12–14]

We have previously described the construction of a database of novel molecular descriptors for compounds, which may be considered as carbohydrate substructures, that is, alkanes, alcohols, diols, ethers,

* Corresponding author. E-mail: svava@kemi.dtu.dk

0008-6215/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2003.09.025

and oxyalcohols,[15] and QSPR models for boiling and melting points, heat of evaporation and fusion, and liquid densities for these classes of compounds.[16]

In the present work, models for estimation of six different physical properties of monosaccharides are given. These properties are partial molar heat capacities, heats of solution of saccharides in aqueous solution, and melting points, enthalpies of fusion, glass-transition temperature, and densities of pure saccharides. They are based on monosaccharide structures, and both furanosides and pyranosides have been included in the study. In order to develop physically sound models, which account for the conformational flexibility of the compounds, a methodology for selecting the most energetically favorable monosaccharide conformers has been established. Similar methodology has previously been developed for a series of organic compounds.[15]

2. Computational methods

2.1. Computational details

The QSPR models were developed with the program Codessa[17] (COmprehensive DEscriptors for Structural and Statistical Analysis), using a heuristic method.[18,19] Use of this method makes it possible to develop regression equations with statistical weighting of each molecular conformer (equilibrium conformation). Furthermore, the method uses a statistical selection scheme, which selects the most appropriate descriptors efficiently.

As the first step Codessa calculates a large number of descriptors from Cartesian coordinates for the molecular conformers. All descriptors are verified prior to the model development, that is, descriptors that have missing values for some conformers, or are equal for a number of the molecular structures given, are discarded. After this, one parameter regression equations are developed with each of the remaining descriptors, and those descriptors that do not satisfy predefined statistical criteria are discarded. These criteria concern the value of the F-test, which is a measure of the variance caused by the model compared to the variance in experimental data, and a correlation coefficient, R^2 that tells how good the correlation is. Finally, a t-test was used to measure the importance of the descriptors used in the correlation.[18] These statistical criteria together enable a useful judgment of how well the descriptors perform individually, and if the descriptors are intercorrelated. For each of the one-parameter equations passing the statistical tests, another descriptor is added in turn and a new regression equation is calculated. The descriptor added in the resulting two-parameter equation is kept if the applied statistical conditions are fulfilled. This procedure is continued until a QSPR model with the desired number of descriptors is obtained.[5,18–21] Two other statistical parameters are of importance, the standard deviation, s, which is a measure of how much the calculated values differ from the experimental values, and the cross-correlation coefficient, R^2_{cv}. The latter is obtained by leaving out one experimental data point, develop a new QSPR model for the remaining compounds, and then use the QSPR model for prediction of the property for the compound that has been left out. This was done for every compound included in the study, the resulting differences between predicted and experimental values are summed and squared, and a value for R^2_{cv} is calculated. To ensure stability and predictability of the QSPR models, the R^2_{cv} value should preferably be as high as possible and have values comparable to the value of the correlation coefficient, R^2.[6,7]

A summary of the computational methodology is shown in Figure 1. For each molecule all possible conformers, discussed in a later section, have been minimized with the molecular mechanics program Consistent Force Field (CFF),[22–24] using the parameter set PEF95SAC,[25–27] optimized for carbohydrate structures. The potential energy functions treat bonded interactions with Morse functions and non-bonded interactions with a Lennard–Jones 12-6 potential and a Coulomb term.[27] Hence, the potential energy function is as follows:

$$E_{\mathrm{Total}} = E_{\mathrm{Bonded}} + E_{\mathrm{Non\text{-}bonded}} + E_{\mathrm{Correction}},$$

where

$$E_{\mathrm{Bonded}} = \sum_{\mathrm{Bonds}} D_e[\mathrm{e}^{-2\alpha(b-b_0)} - 2\mathrm{e}^{-\alpha(b-b_0)}],$$

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

and

$$E_{\mathrm{Correction}} = \sum_{\mathrm{Torsions}} \frac{1}{2} K_{\Phi}(1 + \cos k\Phi) + \sum_{\mathrm{Angles}} \frac{1}{2} K_{\theta}(\theta - \theta_0)^2.$$

In these equations, the bond lengths, b, the interatomic distances, r, the valence angles, θ, and the torsional angles, Φ, were optimized during energy minimization, while the remaining parameters were fixed as a part of the parameter set PEF95SAC. The dielectric constant, D, was set to a standard value of 2, as described in more detail by Engelsen et al.[27]

Following energy minimization, thermodynamic properties were calculated using standard statistical mechanics formalism, assuming ideal gas behavior, where the vibrational contribution to the energy is calculated with the Einstein relation. The molecules were treated as rigid rotors and as coupled harmonic oscillators.[24] Using the calculated Gibbs free energies, G_i, at 298.16 K, the relative probabilities, p_i, of all the conformers could be calculated with Boltzmann statistics:

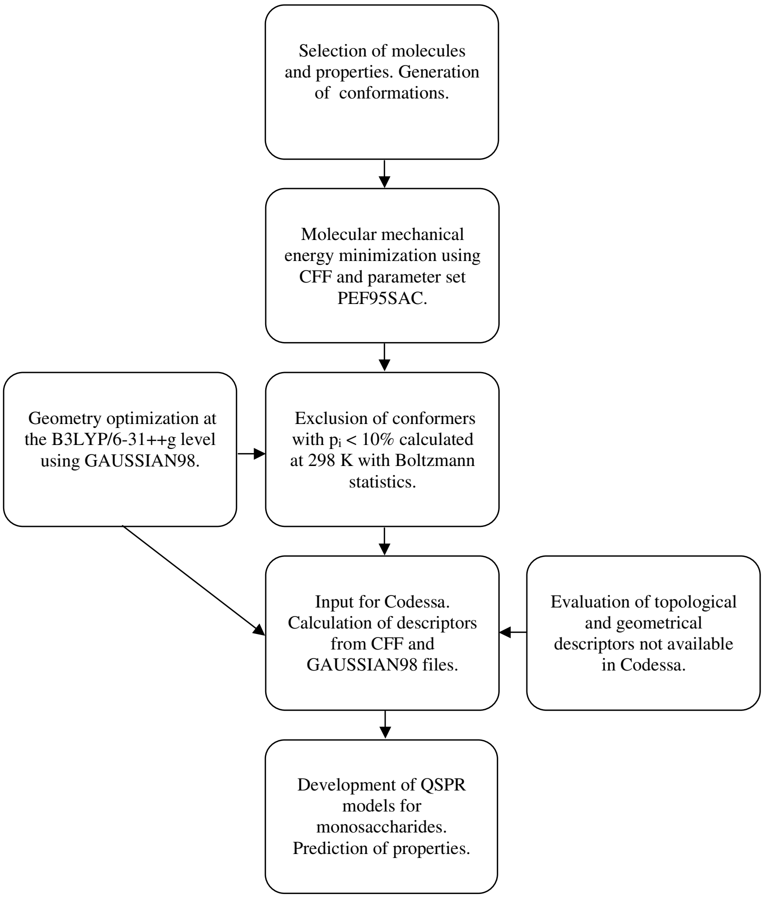
Selection of molecules and properties. Generation of conformations.
Molecular mechanical energy minimization using CFF and parameter set PEF95SAC.
Geometry optimization at the B3LYP/6-31++g level using GAUSSIAN98.
Exclusion of conformers with pi < 10% calculated at 298 K with Boltzmann statistics.
Input for Codessa. Calculation of descriptors from CFF and GAUSSIAN98 files.
Evaluation of topological and geometrical descriptors not available in Codessa.
Development of QSPR models for monosaccharides. Prediction of properties.

Figure 1. Schematic presentation of the methodology for development of the QSPR models.

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

where $\Delta G_i = G_i - G_1$, T is the temperature and R is the gas constant. The conformers with a relative probability above 10% are kept, all other conformers are excluded from further analysis. This limit was chosen to include conformational flexibility in the QSPR models, and at the same time ensure a reasonable computational time. The 10% limit of selection of conformers and the importance of Boltzmann averaging is discussed in further detail in Dyekjær et al.[15]

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[28] with the Gaussian98 program.[29] The Cartesian coordinates of the energy minimized molecular conformers calculated by CFF and selected using the 10% limit, were used as input to these calculations and optimized further until convergence was reached.

In this work we have both used descriptors available within the Codessa program and added several new descriptors. These are the total energy (E_{Total}), the van der Waals energy (E_{vdW}), the Coulomb energy (E_{Coulomb}), dipole moments, as well as the total Gibbs free energy (G_{Total}) calculated at 298.16 K. An additional descriptor, the total non-bonded energy ($E_{\mathrm{Non\text{-}bonded}}$) can easily be evaluated by summing up the van der Waals and the Coulomb contributions. All these descriptors are extracted from the molecular mechanical calculations carried out with the CFF program, as described above. The molecular energies (E_{B3LYP}) and dipole moments obtained from the B3LYP/6-31++g quantum chemical calculation are also used. For each of the descriptors, Boltzmann averaged values according to the selection scheme discussed above, were used. A number of new constitutional and geometric descriptors were also added, as discussed previously.[15]

2.2. Conformational analysis of carbohydrates

To incorporate the effect of the large conformational flexibility monosaccharides exhibit in solution, a method was developed for selection of the most favorable molecular conformers. For each saccharide the hydroxyl groups can adopt many different conformers due to the rotation of the H–O–C–C torsional angles to each of the three positions, 180° (*anti*, indicated by a) and the two *gauche* positions ±60 ° (indicated as g for the positive torsional angle and g′ for the negative one). It is well-known that of the possible pyranoside conformations, the chair conformer has the lowest energy, and is the only conformer with significant probability.[30,31] Due to the ring oxygen and the anomeric carbon atom, an equilibrium between two different chair conformers exists, as shown in Figure 2a. Both chair conformers need thus to be accounted for. For five-membered furanosides, the envelope conformer is predominant, but here, ring-puckering has to be accounted for instead. For each molecule, ring-puckering has been modeled by placing the substitutes on the appropriate carbon atoms with respect to a fixed atom. This fixed atom is moved out of a plane in order to form the envelope conformer, such that all five possible envelope conformers are formed. This approach is shown in Figure 2b and is also discussed in more detail by Dyekjær et al.[15]

To generate all these conformations in a systematic way, a set of Cartesian coordinates for each of the the atoms forming the ring in each of the structures A and B for pyranosides, and A, B, C, D, and E for each of the furanosides, shown in Figure 2, have been obtained from the molecular mechanical program CFF. These fundamental ring Cartesian coordinates have been obtained from the generation of the two glucose chair conformers, in the case of six-membered rings, and the five ring-puckering ribose conformers for the five-membered cases. OH groups are initially placed in their proper positions on the ring atoms, but after energy minimization all the OH groups and the H atoms have been removed to obtain the set of basic ring Cartesian coordinates, as shown in Figure 2. This set of Cartesian coordinates forms a basis for generation of all desired monosaccharides and their molecular conformations considered in the study. To these basic conformer structures, hydroxyl groups and hydrogen atoms have been added on each axial/equatorial position, and in such a way that all relevant hydroxyl torsional angles a, g, g′ are taken into account. The Cartesian coordinates have been tabulated.[32] Atomic labels used in the following description are shown in Figure 3, using the same numbering system as described by Pérez et al.[33]

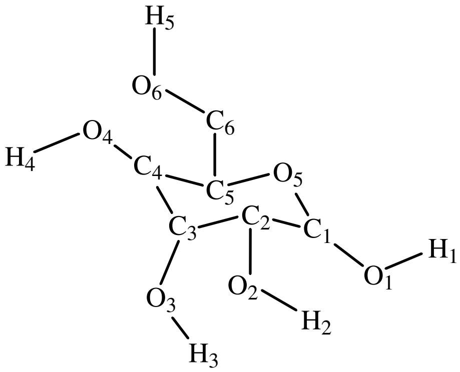

Figure 3. Labeling of atoms to illustrate the torsional angles.

An initial study was made, using six-membered rings as an example, generating both of the conformers A and B, and in each case the α and the β anomer. For all these four cases, the following procedure was carried out, where all possible combinations of conformers for the CH_2–OH group attached to the sugar ring were studied. All the remaining OH groups were placed in the a orientation. The torsional angle O6–C6–C5–O5 was varied to each of the three positions a, g, and g′ for each of the mentioned conformers, while at the same time orienting the H5–O6–C6–C5 torsional angle in the same manner (36 possible combinations in total). All these conformations were energy minimized, and if one of the resulting conformers stood out as being more energetically favorable than the other conformers, the remaining conformations to be generated were based on that conformer by fixing the O–C–C–O torsional angle and the ring conformer and varying the other H–O–C–C torsional angles. This means that only one-third of the possible conformations need to be generated and energy minimized. A similar approach has been used by Kirschner and Woods.[34] In some cases it was not possible to

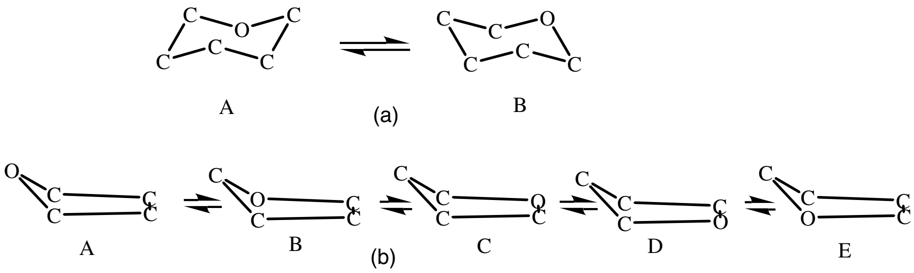

Figure 2. (a) Equilibrium chair conformers for pyranoside rings. The A and B notation distinguishes between these two conformers. (b) For the envelope conformer of furanosides, ring-puckering is modeled using each of these conformers denoted A, B, C, D, and E, respectively.

decide if one torsional angle O6–C6–C5–O5 was preferable over another, and in those cases, all possible angles were investigated.

The generated conformations were subjected to energy minimization followed by Boltzmann averaging as described in the previous section. In Table 1, the number of conformations initially generated are shown for all compounds included in the study. Also, the number of conformers remaining after Boltzmann weighting and exclusion of conformers with a probability less than 10% are shown. Furthermore, the actual contributions for each conformer, having a relative probability above 10%, are shown as well.

Table 1 shows that three conformers are selected for galactose. These are the aaaaa-(a)-B, aaaag-(a)-B, and ag′gaa-(a)-B conformers, respectively. In each case the first five letter stands for the H–O–C–C torsional angles in the following order: O5–C1–O1–H1, H2–O2–C2–C1, H3–O3–C3–C2, H4–O4–C4–C3, and H5–O6–C6–C5. The (a) denotes that the torsional angle O6–C6–C5–O5 is app. 180°. The last conformer, ag′gaa-(a), has thus an H2–O2–C2–C1 torsional angle of −60° and H3–O3–C2–C2 of 60°, with the remaining torsional angles in the a position. The last letter B stands for the ring conformer B, see Figure 2a. The numbering of torsional angles shown in Figure 3 is valid for all the compounds listed in Table 1. For three compounds one of the OH groups is replaced by a O–CH_3 group, but the same numbering system is maintained. Four compounds have no CH_2OH side group, and thus not any torsional angles

Table 1. The calculated relative contributions of the most energetically favorable conformers of each compound

3-*O*-Methylglucose, $N_{\text{Initial}} = 115$, $N_{10\%} = 4$			
agag′a-(a)-B	ggg′gg-(g′)-B	g′aaaa-(a)-B	g′g′g′g′g-(g′)-B
0.2292	0.2348	0.1901	0.3460
D-(+)-Galactose, $N_{\text{Initial}} = 153$, $N_{10\%} = 3$			
aaaaa-(a)-B	aaaag-(a)-B		ag′gaa-(a)-B
0.5095	0.3169		0.1736
D-Mannose, $N_{\text{Initial}} = 153$, $N_{10\%} = 4$			
gaaaa-(a)-B	g′aaag′-(a)-B	g′gg′gg-(g′)-B	g′g′g′g′g′-(g′)-B
0.3291	0.1709	0.3088	0.1911
α-D-(+)-Arabinose, $N_{\text{Initial}} = 110$, $N_{10\%} = 1$			
aaaa-A			
1.0000			
α-D-(+)-Glucose, $N_{\text{Initial}} = 153$, $N_{10\%} = 2$			
ggggg′-(g′)-B			g′gg′gg′-(g′)-B
0.3856			0.6144
α-D-Lyxose, $N_{\text{Initial}} = 110$, $N_{10\%} = 2$			
aaga-B			aag′a-B
0.4737			0.5263
α-D-Xylose, $N_{\text{Initial}} = 110$, $N_{10\%} = 4$			
aaga-B	agg′a-B	gag′a-B	g′ag′a-B
0.3500	0.2326	0.1467	0.2708
α-Methylglucoside, $N_{\text{Initial}} = 115$, $N_{10\%} = 6$			
ggggg-(g′)-B	ggg′gg-(g′)-B		ggg′gg′-(g′)-B
0.1768	0.2212		0.1641
g′gg′gg-(g′)-B	g′g′gg′g′-(g′)-B		g′g′g′g′g′-(g′)-B
0.1325	0.1400		0.1654
α-Methylmannopyranoside, $N_{\text{Initial}} = 115$, $N_{10\%} = 1$			
ggg′g′g-(g)-A			
1.0000			
β-D-(−)-Arabinose, $N_{\text{Initial}} = 110$, $N_{10\%} = 1$			
aaaa-B			
1.0000			
β-D-Ribose, $N_{\text{Initial}} = 325$, $N_{10\%} = 2$,			
aaag-(a)-E			ag′ag′-(a)-E
0.7250			0.2750
β-L-(−)-Fucose, $N_{\text{Initial}} = 220$, $N_{10\%} = 3$			
aaaa-A	aaag-A		agaa-A
0.2043	0.5879		0.2079

All conformers with a contribution larger than 10% ($N_{10\%}$) are listed, and the number of initially generated number of conformations N_{Initial} is also given. The names of the conformers describe the conformational arrangement of the molecules. a, g, g′ denotes hydroxyl group torsional angles, (a), (g) and (g′) is the orientation of the OCCO torsional angle and A, B, etc. denote the ring conformer as shown in Figure 2.

corresponding to H5–O6–C6–C5 and O6–C6–C5–O5. For ribose similar numbering system is used, but with one less H–O–C–C torsion.

3. Results and discussion

The new descriptors proposed in the present work are listed in Table 2. For each compound, the given value is obtained by Boltzmann averaging over the calculated values for the conformers listed in Table 1. All the energy values given are absolute energies, but they should only be considered on relative terms, that is, how they differ from one molecule to another. An individual value for the total energy or the total Gibbs free energy of a particular compound is thus only useful in relation to the corresponding values for the other compounds. As discussed in the following section, the non-bonded energy values, van der Waals, and Coulomb, are of particular importance, as they give an estimate of the strength of the non-bonded interactions between the atoms in the molecule. There is a significant difference between the quantum chemical and molecular mechanical calculated dipole moments, where the quantum chemical are considered to be more accurate.

Several QSPR models for the partial molar heat capacities at 298.15 K were developed by fitting to experimental data for 10 different monosaccharides in aqueous solution.[35] The best QSPR model allowing use of up to five descriptors is given in Eq. 2. Several excellent correlations were obtained, with correlation coefficients ranging from 0.996 to 0.992. Furthermore, the descriptors used in different models were similar, which is a good indication of the stability of the QSPR models.

$$\begin{aligned} C_p\ (\mathrm{J\,K^{-1}\,mol^{-1}}) = {} & -2020.6 - 2076.6 \cdot D_{\mathrm{RNCG}} \\ & + 9778.0 \cdot D_{\mathrm{Rel.C}} - 38.186 \\ & \cdot D_{\mathrm{HOMO-1}} - 129.91 \cdot D_{XZ/XZ} \\ & - 11.410 \cdot D_{\mathrm{HOMO}}, \end{aligned} \tag{2}$$

$$N_{\mathrm{Molec}} = 10, \quad N_{\mathrm{Conf}} = 30, \quad R^2 = 0.9961,$$
$$F = 1235, \quad s = 4.401\,\mathrm{J\,K^{-1}\,mol^{-1}}, \quad R^2_{\mathrm{cv}} = 0.9936.$$

The model uses a descriptor accounting for the distribution of negative charge in the molecule, the relative number of carbon atoms, which can be seen as a measure of the size of the molecule, a so-called shadow index, accounting for the shape of the molecule, and the quantum chemically obtained HOMO and HOMO − 1 energies. See Table 3 for detailed information about the descriptors.

The results obtained with Eq. 2 are shown in Table 4 and Figure 4, including predicted values of the same property for three disaccharides. This model gives results in excellent agreement with experimental values, not only for monosaccharides, but also for disaccharides. It is therefore possible to extrapolate a model developed for monosaccharides to predict this property for disaccharides. An attempt to develop QSPR models containing constitutional and topological descriptors only, gave good correlation for the monosaccharides, but poor predictions for the disaccharides. It is thus clearly beneficial to use more physically sound descriptors for predictive purposes. All the experimental data used are from Galema et al.[35]

QSPR models for heat of solution at 298.15 K for monosaccharides dissolved in water to a dilute solution of 10^{-5} M were developed, using experimental data from Jasra and Ahluwalia.[36] The best model developed in this

Table 2. Energetic and electronic descriptors

Molecule	E_{Coulomb} (kJ mol^{-1})	E_{vdW} (kJ mol^{-1})	E_{Total} (kJ mol^{-1})	G_{Total} (kJ mol^{-1})	E_{B3LYP} (a.u.)	μ_{CFF} (Debye)	μ_{G98} (Debye)
3-*O*-Methylglucose	223.928	24.905	−10,501.469	−10,014.237	−726.27738	3.08	2.09
D-(+)-Galactose	211.354	21.293	−9301.629	−8881.334	−686.98308	1.48	1.29
D-Mannose	208.954	21.434	−9301.636	−8880.989	−686.98998	3.54	2.31
α-D-(+)-Arabinose	170.941	25.925	−7747.851	−7406.316	−572.48471	2.26	1.93
α-D-(+)-Glucose	217.714	21.325	−9293.657	−8876.102	−686.98686	1.05	1.14
α-D-Lyxose	171.727	23.417	−7750.409	−7409.266	−572.48659	1.52	1.40
α-D-Xylose	176.759	25.361	−7742.751	−7403.927	−572.48283	2.72	2.29
α-Methylglucoside	261.639	21.355	−10,472.955	−9987.287	−726.27889	2.83	2.49
α-Methylmannopyranoside	253.956	18.417	−10,486.238	−10,486.239	−276.27649	3.25	2.50
β-D-(−)-Arabinose	170.941	25.925	−7747.851	−7406.316	−572.48471	2.26	1.93
β-D-Ribose	154.432	8.291	−7734.950	−7397.615	−572.48318	2.66	2.70
β-L-(−)-Fucose	151.466	23.829	−8988.427	−8576.908	−611.79582	3.17	2.01

Coulomb, van der Waals, the total potential, and Gibbs free energies calculated with CFF using the PEF95SAC parameter set.

The B3LYP/6-31G++ energies using Gaussian98 are also listed. The energy unit for this is Hartree. Dipole moments (μ) are calculated with CFF and Gaussian98, respectively. An additional descriptor, the total non-bonded energy, can easily be evaluated by summing together the Coulomb and the van der Waals energy. It is important to mention that these energy values are absolute, and should only be considered relative to one another.

Table 3. Abbreviations and overview of the descriptors used

Abbreviation	Descriptors	Unit	Ref.
Constitutional			
$D_{\mathrm{Rel.\,C}}$	Relative number of carbon atoms		1
Topological			
$D_{\mathrm{Kier\,Flex}}$	Kier flexibility index		40
Geometric			
$D_{XZ/XZ}$	*XZ* shadow/*XZ* rectangle		37
D_{YZ}	*YZ* shadow		37
$D_{YZ/YZ}$	*YZ* shadow/*YZ* rectangle		37
$D_{\mathrm{Mol.vol.}/XYZ}$	Molecular volume/*XYZ* box		37
$D_{\mathrm{Principal}\,I_{\mathrm{C}}/\mathrm{Atoms}}$	Principal moment of Inertia$_{\mathrm{C}}$/number of atoms		1
$D_{\mathrm{Principal}\,I_{\mathrm{A}}/\mathrm{Atoms}}$	Principal moment of Inertia$_{\mathrm{A}}$/number of atoms		1
Electrostatic			
D_{Polarity}	Polarity parameter ($q_{\mathrm{max}} - q_{\mathrm{min}}$)		41
D_{PPSA1}	Partial positive surface area	Area in Å^2	42,43
$D_{\max\, q^{\mathrm{O}}}$	Maximum partial charge for an O atom		1,44
$D_{\mathrm{Polarity}/r^2}$	Polarity parameter divided by the square distance		1
D_{RNCG}	Relative negative charge		42,43
D_{RPCS}	Relative positive charges SA (SAMPOS∗RPCG)	Area in Å^2	42,43
Quantum chemical			
$D_{\mathrm{DPSA1(QC)}}$	Difference in positively and negatively charged partial surface areas (quantum chemical)	Area in Å^2	42,43
$D_{\mathrm{PPSA3(QC)}}$	Atomic charges weighted PPSA (quantum chemical)	Area in Å^2	42,43
$D_{\mathrm{HASA1/TMSA(QC)}}$	Surface charge of hydrogen bond acceptor divided by total molecular surface area (quantum chemical)	Area in Å^2	42,43
$D_{\mathrm{RPCS(QC)}}$	Relative positive charges SA (SAMPOS∗RPCG) (quantum chemical)	Area in Å^2	42,43
$D_{\mathrm{Maxsigma-sigma}}$	Max sigma–sigma bond order		2
$D_{\mathrm{HOMO-LUMO}}$	HOMO–LUMO energy gap	Hartree	2
D_{FHBCA}	Fractional HBSA (HBSA/TMSA)	Area in Å^2	42,43
$D_{\mathrm{Tot.\,hyb.}\,\mu}$	Total hybridization component of the molecular dipole		
D_{HOMO}	HOMO (highest occupied molecular orbital) energy	Hartree	2
$D_{\mathrm{HOMO-1}}$	HOMO − 1 energy	Hartree	2
D_{LUMO}	LUMO (lowest unoccupied molecular orbital) energy	Hartree	2
Others			
$D_{E_{\mathrm{vdW}}}$	Van der Waals energy obtained from CFF	kJ mol^{-1}	15

work is a five descriptor model with a correlation coefficient of 0.894 and is shown in Eq. 3.

$$\Delta H_{\mathrm{sol}}\ (\mathrm{kJ\,mol^{-1}}) = -5.4911 - 0.85479 \cdot D_{YZ} + 0.020964 \cdot D_{\mathrm{PPSA3(QC)}} + 107.89 \cdot D_{YZ/YZ} - 0.011286 \cdot D_{\mathrm{Mol.Vol.}/XYZ} + 1.9541 \cdot D_{\mathrm{RPCS(QC)}}, \tag{3}$$

$$N_{\mathrm{Molec}} = 10, \quad N_{\mathrm{Conf}} = 30, \quad R^2 = 0.8940,$$
$$F = 40.49, \quad s = 1.276\,\mathrm{kJ\,mol^{-1}}, \quad R^2_{\mathrm{cv}} = 0.8497.$$

The descriptors used are geometric shadow indices of the molecule oriented in different directions,[37] and descriptors concerning the molecular charge distribution. The QSPR models for the heat of solution have turned out to be very dependent on the number of descriptors, and the models are thus in general not very stable. However, all models contain geometrical shadow indices, which are very dependent on the molecular conformation. Even so, the QSPR models presented in Eq. 3 correlate heats of solution within a reasonable accuracy as seen in Table 4. Comparable data for disaccharides are not available, and thus predictions for these compounds have not been made.

QSPR models for melting points of saccharides were developed based on experimental data from Dean[38] for 11 compounds and in total 33 conformers. The best QSPR model obtained using five descriptors is

$$\mathrm{MP}\ (^{\circ}\mathrm{C}) = 12.416 + 0.54037 \cdot D_{\mathrm{PPSA1}} + 3.9642 \cdot D_{E_{\mathrm{vdW}}} - 6541.2 \cdot D_{\mathrm{Polarity}/r^2} - 362.98 \cdot D_{XZ/XZ} + 331.14 \cdot D_{YZ/YZ}, \tag{4}$$

$$N_{\mathrm{Molec}} = 11, \quad N_{\mathrm{Conf}} = 33, \quad R^2 = 0.9242,$$
$$F = 65.88, \quad s = 8.586\,^{\circ}\mathrm{C}, \quad R^2_{\mathrm{cv}} = 0.8289.$$

As seen in Eq. 4, this QSPR model depends on the distribution of positive atomic charges in the molecule,

Table 4. Partial molar heat capacity, heat of solution, melting point, and heat of fusion

	C_p (J K^{-1} mol^{-1})		ΔH_{sol} (kJ mol^{-1})		MP (°C)		ΔH_m (J g^{-1})	
	Exp (Ref. 35)	Calcd (Eq. 2)	Exp (Ref. 36)	Calcd (Eq. 3)	Exp (Ref. 38)	Calcd (Eq. 4)	Exp (Ref. 39)	Calcd (Eq. 5)
3-*O*-Methylglucose	425.4	425.40	8.27	7.90	168	163.0		
D-(+)-Galactose	346.7	349.20	17.20	16.13	167	165.1	243	241.8
D-(+)-Mannose	345.9	350.60	6.86	7.00	129	138.6	137	140.9
Arabinose	279.5	279.99	13.24	13.42	161.5	156.6	238	237.3
α-D-(+)-Glucose	363.3	356.74	10.70	11.58	146	151.0	179	176.8
α-D–Lyxose	281.7	280.23	10.10	10.79	106.5	118.1		
α-D-(+)-Xylose	294.4	290.33	11.98	10.80	144.5	138.5	211	212.3
α-Methylglucoside	435.1	437.54	3.67	4.98	168	171.5		
α-Methyl-D-mannopyranoside	438.0	435.56	9.09	7.90	194.5	191.8		
β-D-(+)-Arabinose					163	157.2		
β-D-Ribofuranose	271.5	275.27	13.04	13.74	87	82.3	146	145.6
β-L-(−)-Fucose					151.5	153.3		
Cellobiose*					225	310.8		
Lactose*	679.9	671.36			253	321.4		
Sucrose*	662.0	667.52			160–165	224.9	135	194.3
Maltose*	674.2	674.33			185–186	310.0		
RMS deviation		3.413		0.872		5.91		1.97

The QSPR models are based on experimental data from the cited references, respectively. The disaccharides marked with an asterisk have not been included in the development of the models.

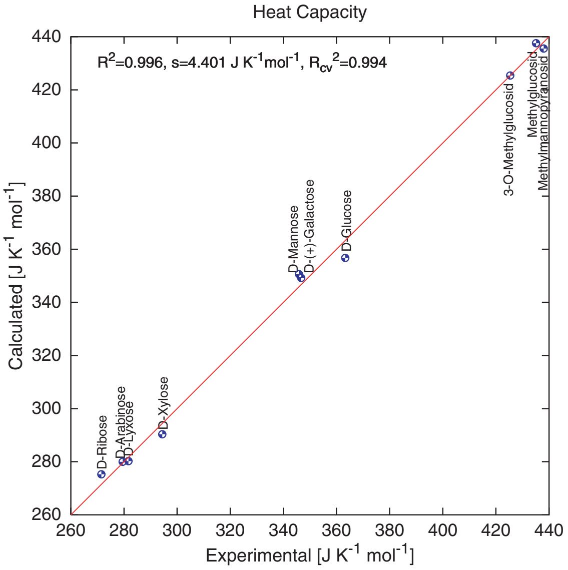

Figure 4. QSPR model for the partial molar heat capacity.

the van der Waals energy, and shadow indices. The obtained cross-correlation coefficient in Eq. 4 is too low to expect good predictability.

Using this QSPR model, melting points for three disaccharides have been predicted and compared to values from Dean.[38] In all cases melting points for anhydrous sugars are used. From Table 4, it is seen that the predicted values are considerably larger than the experimental values. This difference can be caused by the more complicated structure and hydrogen bond pattern of the disaccharides. Furthermore, it should be kept in mind that melting points for carbohydrates themselves vary within experimental data in the literature. Also melting points for simpler compounds have shown to be difficult to model.[16] Better models for the melting points could probably be developed if descriptors like lattice energies, or interaction energies between pairs of molecules could be used. It is, however, very computationally demanding to calculate such descriptors and is beyond the scope of this paper.

The QSPR models presented here for the heat of fusion have been developed for seven different monosaccharides based on 17 conformers. Experimental data have been obtained from Roos.[39] The best model using five descriptors is shown in Eq. 5.

$$\begin{aligned}\Delta H_m\ (\mathrm{J\,g^{-1}}) = &-736.75 + 7.0843 \cdot D_{E_{vdW}} + 1200.7 \\ &\cdot D_{YZ/YZ} - 80.492 \cdot D_{RPCS} + 67.018 \\ &\cdot D_{\mathrm{Kier\,Flex}} - 1725.0 \\ &\cdot D_{\mathrm{HASA1/TMSA(QC)}}, \end{aligned} \tag{5}$$

$$N_{\mathrm{Molec}} = 7,\quad N_{\mathrm{Conf}} = 17,\quad R^2 = 0.9781,$$
$$F = 98.31,\quad s = 7.671\,\mathrm{J\,g^{-1}},\quad R^2_{cv} = 0.8958.$$

The most important descriptor used is the van der Waals energy. The other descriptors are the shadow index, the Kier flexibility index, and two charge dependent descriptors. As seen in Table 4 a very good agreement between experimental and calculated values is obtained. This is also illustrated in Figure 5. Heat of fusion has been predicted for sucrose, but in this case the predictability of the model is not satisfactory. Unfortunately, data was only available for seven monosaccharides, and the cross-correlated correlation coefficient

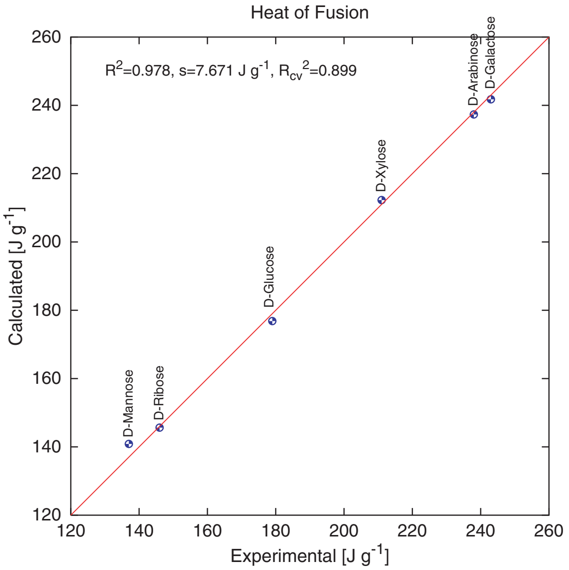

Figure 5. QSPR model for the heat of fusion.

is too low compared to the correlation coefficient. Thus good predictions are difficult to obtain with this model.

QSPR models for the glass-transition temperature for a range of monosaccharides were developed based on the experimental data from Roos.[39] Very good correlations were obtained both by using five descriptors, as well as three descriptors. All the models developed have correlation coefficients around 0.99, and standard deviations around 0.5 °C. The best model using five descriptors is seen shown Eq. 6.

$$T_g\ (^\circ\mathrm{C}) = -10,828 - 65,234 \cdot D_{\mathrm{Principal}\, I_c/\mathrm{Atoms}} - 52,919 \cdot D_{\max\, q^{\mathrm{O}}} + 4.7112 \cdot D_{\mathrm{LUMO}} + 0.094079 \cdot D_{\mathrm{DPSA1(QC)}} + 2438.5 \cdot D_{\mathrm{Maxsigma-sigma}}, \qquad (6)$$

$$N_{\mathrm{Molec}} = 6, \quad N_{\mathrm{Conf}} = 17, \quad R^2 = 0.9994,$$
$$F = 3513.42, \quad s = 0.560\,^\circ\mathrm{C}, \quad R^2_{\mathrm{cv}} = 0.9975.$$

It is seen that the glass-transition temperature depends highly on the principal moment of inertia divided by the number of atoms. This descriptor can be considered as a measure of how extended the molecule is, and is thus a reasonable descriptor for the glass-transition temperature. The model also includes a charge related descriptors, and the quantum chemically determined LUMO energy and maximum sigma–sigma bond order, see Table 3 for details.

Since this model is based on few data points, a QSPR model having three descriptors has also been developed, to ensure that no over-fitting is done.

$$T_g\ (^\circ\mathrm{C}) = -8484.4 - 70,359 \cdot D_{\mathrm{Principal}\, I_c/\mathrm{Atoms}} - 53,603 \cdot D_{\max\, q^{\mathrm{O}}} + 10.3048 \cdot D_{\mathrm{LUMO}}, \qquad (7)$$

$$N_{\mathrm{Molec}} = 6, \quad N_{\mathrm{Conf}} = 17, \quad R^2 = 0.9978,$$
$$F = 1990.12, \quad s = 0.960\,^\circ\mathrm{C}, \quad R^2_{\mathrm{cv}} = 0.9947.$$

The three descriptor model, Eq. 7, is seen to be very similar to the five descriptor QSPR model, verifying the stability of the model. In Table 5 and Figure 6 the good agreement between experimental and calculated glass-transition temperatures for monosaccharides using these two QSPR models is shown. Both of these models have been used for prediction of the glass-transition temperature for disaccharides, but as seen in Table 5, the predicted values are about 100 °C smaller than the experimental values. This is the case for both QSPR models, and is therefore not caused by over-fitting. This large deviation can be caused by the large regression coefficients, seen in both equations, which may make a small deviation in a descriptor value to a large deviation in the overall value of the predicted property. Also, only

Table 5. Glass-transition temperature and solid state density

	T_g (°C)			ρ (g cm^{-3})		
	Exp (Ref. 39)	Calcd (Eq. 6)	Calcd (Eq. 7)	Exp (Ref. 38)	Calcd (Eq. 8)	Calcd (Eq. 9)
D-(+)-Galactose	30	29.7	29.4	1.562	1.5618	1.5596
D-(+)-Mannose	25	25.1	25.6	1.540	1.5402	1.5425
Arabinose	−2	−2.2	−2.5			
α-D-(+)-Glucose	31	31.1	30.9			
α-D–Lyxose				1.545	1.5445	1.5417
α-D-(+)-Xylose	6	6.3	6.9	1.535	1.5366	1.5385
α-Methylglucoside				1.460	1.4600	1.4597
β-D-Ribofuranose	−20	−19.9	−19.8			
Cellobiose*	108.1	5.2	4.4			
Lactose*	112.3	4.6	4.0	1.59	1.437	1.468
Sucrose*				1.5805	1.386	1.422
Maltose*	100.6	1.1	1.7			
RMS deviation		0.20	0.54		0.0008	0.0026

The QSPR models are based on experimental data from the cited references, respectively. The disaccharides marked with an asterisk have not been included in the development of the models.

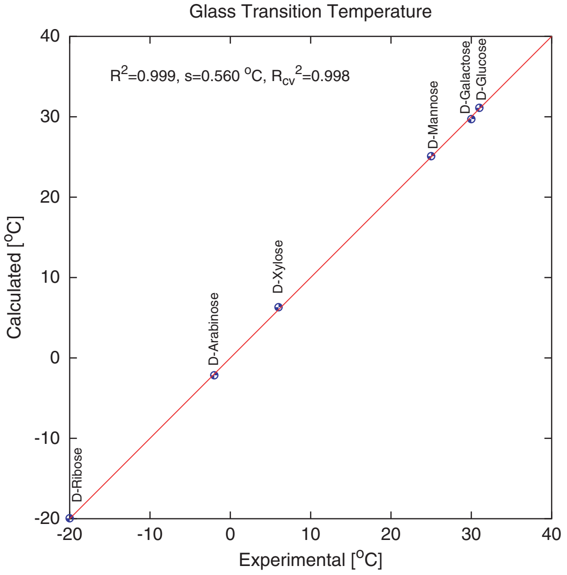

Figure 6. QSPR model for the glass-transition temperature.

one conformer of each of the disaccharides is used. Nonetheless, both QSPR models have high cross-correlation coefficients, indicating that predictions made for other monosaccharides not included in the fit, will probably be reasonable.

The QSPR models developed for the density of carbohydrates are based on experimental solid state data from Dean.[38] These models developed give results that are in very good agreement with experimental data, but unfortunately only five experimental data points for anhydrous monosaccharides were available. Similar descriptors appear in all models when using between five and two descriptors. The correlation coefficients for all the models developed for the density are about 0.99 and the standard deviation is 0.000. The cross-correlation coefficient is high, indicating good predictability for monosaccharides. In Eq. 8, the best model obtained using five parameters is given.

$$\begin{aligned}\rho\ (\mathrm{g\,cm^{-3}}) = {} & -3.6593 + 68.921 \cdot D_{\mathrm{Principal}\,I_{\mathrm{A}}/\mathrm{Atoms}} \\ & + 19.031 \cdot D_{\mathrm{Polarity}} + 0.020709 \\ & \cdot D_{\mathrm{HOMO-LUMO}} - 0.27324 \cdot D_{\mathrm{FHBCA}} \\ & + 0.00090161 \cdot D_{\mathrm{Tot.\,hybr.}\,\mu}, \end{aligned} \tag{8}$$

$$N_{\mathrm{Molec}} = 5, \quad N_{\mathrm{Conf}} = 19, \quad R^2 = 0.9987,$$
$$F = 2030.20, \quad s = 0.0000\,\mathrm{g\,cm^{-3}}, \quad R^2_{\mathrm{cv}} = 0.9964.$$

The descriptor used are the principal moments of inertia, in direction A, divided by the number of atoms, and the polarity parameter, which is the difference between the maximum and minimum atomic charge in the molecule under consideration. The HOMO–LUMO energy gap and a descriptor depending on the charge distribution in the molecule also appear. To avoid overfitting due to the few data points, a QSPR model using only two descriptors was developed, and is given in Eq. 9.

$$\begin{aligned}\rho\ (\mathrm{g\,cm^{-3}}) = {} & -2.6355 + 56.705 \cdot D_{\mathrm{Principal}\,I_{\mathrm{A}}/\mathrm{Atoms}} \\ & + 15.292 \cdot D_{\mathrm{Polarity}}, \end{aligned} \tag{9}$$

$$N_{\mathrm{Molec}} = 5, \quad N_{\mathrm{Conf}} = 19, \quad R^2 = 0.9896,$$
$$F = 764.77, \quad s = 0.0000\,\mathrm{g\,cm^{-3}}, \quad R^2_{\mathrm{cv}} = 0.9890.$$

Calculations made for monosaccharides using these two QSPR models, Eqs. 8 and 9, are shown in Table 5. Good agreement between experimental and correlated values is obtained. Both QSPR models have been used for prediction of the density for two disaccharides, see Table 5. The predicted values for the two disaccharides are in reasonable agreement with experiment, especially considering that those values are calculated with a model developed for monosaccharides.

In general, the models presented in this paper give good correlations for monosaccharides. In many of the models the shadow indices, which are very dependent on the molecular conformation, appear. This type of descriptor can account for and differentiate between the small conformational differences among monosaccharides. Monosaccharides are constitutionally similar, and in most case one compound only differs from the others by the orientation of the hydroxyl groups.

Predictions made for disaccharides are not satisfactory in some cases, except for the partial molar heat capacity for which the predicted values for disaccharides are in very good accord with experimental values. There are several reasons for the discrepancies. Firstly, the models are developed for monosaccharides, and therefore extrapolation to the much more complex compound disaccharides may be likely to break down. Secondly, due to lack of experimental data, the data sets used are in some cases are rather small. We have used each conformer as a data point, and therefore it has been assumed reasonable to use up to five descriptors. In cases that may be questionable, models developed using fewer descriptors are presented also, to ensure that overfitting of data has been avoided.

Finally, it should be noted that most of the models depend on the shadow indices or the moment of inertia, which show a strong dependence on molecular conformation. The predictions made for disaccharides are based on one conformer only. This may therefore lead to small deviations that may be particularly apparent in the cases where the regression coefficients connected to the shadow index descriptors are large. This is for instance the case for the predictions made for the glass-transition temperature, where the regression coefficient for the moment of inertia descriptor is large. A small deviation in this descriptor may therefore lead to a large deviation in the predicted values.

In particular, it is very interesting that the van der Waals energy appears in the QSPR models for the melting points and heats of fusion. This descriptor has previously also been used for modeling of melting points, but for the classes of compounds alkanes, alcohols, polyalcohols, ethers, and oxyalcohols. Its presence in the QSPR models for melting points and heats of fusion thus confirms the physical importance of this descriptor.

As discussed in a previous paper,[16] better descriptors for accurate predictions of properties relating to solid–liquid transitions like melting points and heats of fusion need to be involved. It is most likely that the used descriptors fail to reproduce lattice energies and molecular properties in general relating to the solid phase. This is understandable as all the descriptors used are calculated in the gas phase. Improvements were obtained by using the van der Waals energy as descriptor, but better descriptors need to be invented for obtaining good predictions. However, this work is the first attempt to develop QSPR models for carbohydrates, and in most cases good correlations were obtained.

Acknowledgements

The financial support for this project has been provided by the Danish Technical Research Council. We would also like to thank Kjeld Rasmussen for many fruitful discussions and comments on this manuscript.

References

1. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; John Wiley & Sons, New York, 2000.
2. Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96*, 1027–1044.
3. Liu, S.; Cai, S.; Cao, C.; Li, Z. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1337–1348.
4. Ivanciuc, O.; Ivancuic, T.; Cabrol-Bass, D.; Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 631–643.
5. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, *24*, 279–287.
6. Murugan, R.; Grendze, M.; Toomey, J. E.; Katritzky, A. R.; Karelson, M.; Lobanov, V.; Rachwal, P. *Chem. Tech.* **1994**, 17–23.
7. Katritzky, A. R.; Gordeeva, E. V. J. *Chem. Inf. Comput. Sci.* **1993**, *33*, 835–857.
8. Needham, D. E.; Wei, I.-C.; Seybold, P. G. J. *Am. Chem. Soc.* **1988**, *110*, 4186–4194.
9. Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 913–919.
10. Kozioł, J. *J. Quantum Chem.* **2001**, *84*, 117–126.
11. Katritzky, A. R.; Rachwal, P.; Law, K. W.; Karelson, M.; Lobanov, V. S. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 879–884.
12. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic press: New York, 1976.
13. Kier, L. B.; Murray, W. J.; Randić, M.; Hall, L. H. *J. Pharm. Sci.* **1976**, *65*, 1226–1230.
14. Karelson, M.; Perkson, A. *Comput. Chem.* **1999**, *23*, 49–59.
15. Dyekjær, J. D.; Rasmussen, K.; Jónsdóttir, S. Ó. *J. Mol. Model.* **2002**, *8*, 277–289.
16. Dyekjær, J. D.; Jónsdóttir, S. Ó. *Ind. Eng. Chem. Res.* **2003**, *42*, 4241–4259.
17. Codessa, Semichem, 7204 Mullen, Shawnee, KS 66216, USA.
18. Reference Manual (Codessa), Semichem, 7204 Mullen, Shawnee, KS 66216, USA.
19. User's Manual (Codessa), Semichem, 7204 Mullen, Shawnee, KS 66216, USA.
20. Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. *J. Phys. Chem.* **1996**, *100*, 10400–10407.
21. Ivanciuc, O.; Ivancuic, T.; Balaban, A. T. *Tetrahedron* **1998**, *54*, 9129–9142.
22. CFF—Consistent Force Field, version 17.21, Department of Chemistry, Building 207, Technical University of Denmark, DK-2800 Lyngby, Denmark.
23. Rasmussen, K. In: *Potential Energy Functions in Conformational Analysis. Lecture Notes in Chemistry*; Springer: Heidelberg, 1985; Vol. 37.
24. Rasmussen, K.; Engelsen, S. B.; Fabricius, J.; Rasmussen, B. In *Recent Experimental and Computational Advances in Molecular Spectroscopy*; Fausto, R., Ed.; *NATO-ASI series C*; Kluwer Academic Publishers: Dordrecht, 1993; Vol. 406, pp 381–419.
25. Fabricius, J.; Engelsen, S. B.; Rasmussen, K. *J. Carbohydr. Chem.* **1997**, *16*, 751–772.
26. Rasmussen, K. *Asian Chem. Lett.* **2000**, *4*, 33–44.
27. Engelsen, S. B.; Fabricius, J.; Rasmussen, K. *Acta Chem. Scand. A* **1994**, *48*, 548–552.
28. Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
29. *Gaussian98, Revision A.6*, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A. Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian Inc.: Pittsburgh, PA, USA, 1998.
30. Dale, J. *Stereochemistry and Conformational Analysis*; Verlag Chemie: New York and Weinheim, 1978.
31. Rees, D. A. *Polysaccharide Shapes*; John Wiley & Sons: New York, 1977.
32. Dyekjær, J. D. Ph.D. thesis, Technical University of Denmark, 2002.
33. Pérez, S.; Imberty, A.; Scaringe, R. P. In *Computer Modeling of Carbohydrate Molecules*; French, A. D., Brady, J. W., Eds.; *ACS Symposium Series*; American Chemical Society: Washington, DC, 1990; Vol. 430, pp 281–299.
34. Kirschner, K. N.; Woods, R. J. *PNAS* **2001**, *98*, 10541–10545.
35. Galema, S. A.; Engberts, J. B. F. N.; Høiland, H.; Førland, G. M. *J. Phys. Chem.* **1993**, *97*, 6885–6889.
36. Jasra, R. V.; Ahluwalia, J. C. *J. Sol. Chem.* **1982**, *11*, 325–338.

37. Rohrbaugh, R. H.; Jurs, P. C. *Anal. Chim. Acta* **1987**, *199*, 99.
38. Dean, J. A. *Lange's Handbook of Chemistry*. 13th ed. McGraw-Hill: New york, 1985.
39. Roos, Y. *Carbohydr. Res.* **1993**, *238*, 39–48.
40. Kier, L. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science Publishers: New York, 1990; p. 151.
41. Ośmiałowski, K.; Halkiewicz, J.; Kaliszan, R. *J. Chromatogr.* **1985**, *346*, 53–60.
42. Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323.
43. Stanton, D. T.; Egolf, L. M.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306–316.
44. Ośmiałowski, K.; Halkiewicz, J.; Kaliszan, R. *J. Chromatogr.* **1986**, *361*, 39–63.