STUDIA UBB CHEMIA, LXII, 2,Tom I, 2017 (p. 33-48)
(RECOMMENDED CITATION)
DOI:10.24193/subbchem.2017.2.03

Dedicated to Professor Costel Sârbu on the Occasion of His 65th Anniversary

PROPERTY-PROPERTY RELATIONSHIPS FOR MONOSACCHARIDES

LORENTZ JÄNTSCHI[a]

ABSTRACT. The enormous diversity and complexity of polysaccharides resides in the large number of anomeric positions, diversity in the size of the rings as well as of the large number of the linkage positions of the monosaccharides. By taking this fact into account, in order to provide useful knowledge for insight on polysaccharides, a more comprehensive study of the polymer units, the monosaccharides, is required. A study for relating experimentally measured properties - melting points and solubilities with other properties accessible by calculations was conducted for monosaccharides from trioses to hexoses.

Keywords: *Monosaccharides; Property-property relationships; Melting points; Solubilities*

INTRODUCTION

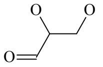
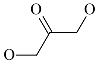
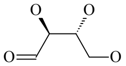
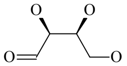
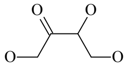
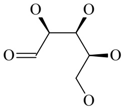
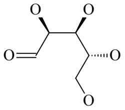
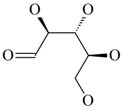
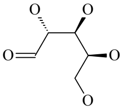
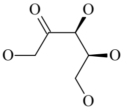
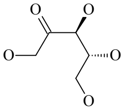
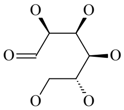
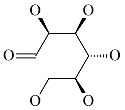
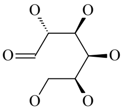
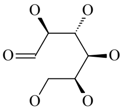
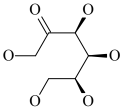
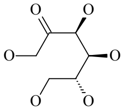
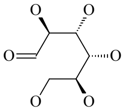
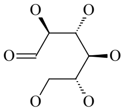
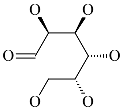
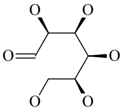
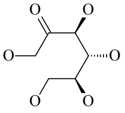
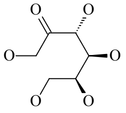
Carbohydrates are structural components of cell walls in plant and algae (cellulose [1]), of DNA - deoxyribonucleic acid (deoxyribose [2]) or RNA - ribonucleic acid (ribose [2]), or of tissues (lyxose [3]). Sugars are short chain carbohydrates, their molecule consisting of carbon (C), hydrogen (H) and oxygen (O) atoms with the general formula $C_m(H_2O)_n$ where $2 \le m$ (and usually $3 \le m \le 7$) and $n \le m$ (and usually $n = m$ or $n = m-1$).

---

[a] *Technical University of Cluj-Napoca, Faculty of Materials and Environmental Engineering, Dept. of Physics & Chemistry, 28 Memorandumului str., RO-400114, Cluj-Napoca, Romania; Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, Doctoral School of Chemistry, 11 Arany János str., RO-400028, Cluj-Napoca, Romania.*
E-mail: lorentz.jantschi@gmail.com

The simplest carbohydrate is the monosaccharide with general formula $(CH_2O)_n$, where *n* ranges from 2 (diose, H-(C=O)-(CH_2)-OH) to usually 7 (*n* = 3 for trioses, *n* = 4 for tetroses, *n* = 5 for pentoses, *n* = 6 for hexoses and *n* = 7 for heptoses). There are 23 monosaccharides (see Table 1) from trioses (*n* = 3) to hexoses (*n* = 6). The monosaccharides with lower number of atoms (e.g. *n* = 3 and *n* = 4) may cyclize by dimerization leading to cyclic monosaccharides with *n* = 6 and *n* = 8, respectively as the monosaccharides can join together to form disaccharides. A disaccharide is formed whenever two monosaccharides (identical or not) joined. Since two identical monosaccharides can form up to eleven different disaccharides [4], and the number increases even more abruptly when different monosaccharides are connected (in [5] were counted 720 trisaccharides, 34560 tetrasaccharides and 2144640 pentasaccharides) the consequence is an enormous diversity and complexity in carbohydrate structure and chemistry.

Table 1. Monosaccharides from trioses to hexoses in open-chain (acyclic) form

n=	Formula	Aldoses				Ketoses	
3	$C_3H_6O_3$	 D-glyceraldehyde				 D-dihydroxyacetone	
4	$C_4H_8O_4$	 D-erythose		 D-threose		 D-erythrulose	
5	$C_5H_{10}O_5$	 D-ribose	 D-arabinose	 D-xylose	 D-lyxose	 D-xylulose	 D-ribulose
6	$C_6H_{12}O_6$	 D-talose	 D-gulose	 D-altrose	 D-glucose	 D-psicose	 D-tagatose
		 D-galactose	 D-idose	 D-mannose	 D-allose	 D-sorbose	 D-fructose
When applies, the names are *ose for acyclic and *opyranose for cylic forms (e.g. Glucose - acyclic → Glucopyranose - cyclic)							

The main problem in studies relating the experimental measurements on carbohydrates is the scarcity of structural information from combined factors (difficulties to crystallize and the limitations in NMR analysis [6]). Another challenge is the fact that usually the researchers studying structural aspects are not the same with the ones conducting the property measurements, and thus the reliability of the data sources being reduced, since very easily during the experimental treatment, monosaccharides may switch from the acyclic to cyclic form as well as the cyclic forms can undergo mutarotation.

The data about melting points and solubilities of 23 monosaccharides were considered in this study to derive property-property relationships.

RESULTS AND DISCUSSION

The reader would expect to have $H/T \sim S$ from $H = E + pV$ and $G = E + pV - TS$ (for other derived equations, see ref. [7]) but is not the case since here the substances are different.

It should be noted that for all acyclic forms listed in Table 3, the number of conformers is given by the formula 9^{n-1} where n comes from the molecular formula of monosaccharides, $(CH_2O)_n$; for cyclic forms, there is no general formula, but all hexoses have $3^6 \cdot 2^2$ conformers (column cf. Table 3).

In order to proceed to assignments (the alternatives from Table 5) between the chemical structures (Table 3) and melting points (Table 2) firstly a relationship between the melting points (column MP (K) in Table 5) and the other properties listed in Table 3, Table 4 and Table 5 inside of the assigned group (the first 13 entries in Table 5) is to be checked.

By keeping in mind that a linear model is significant only when the coefficient of the independent variable is significant enough [8], and aiming to identify at least one principal component [9] in a multiple linear regression [10], the researcher would seek for possible explanation of variables for a given dependent variable (see Table 6).

Table 6. Results of simple linear regressions

Dependent variable	Independent variable	p (%)	
MPK	C_v^{MP}	≈ 4.82	*
	S^{MP}	≈ 4.94	*
	all others	> 5.00	
1/MPK	all	> 5.00	
ln(MPK)	all	> 5.00	
$MPK*S^0$	H_E, L_E, DM	> 5.00	
	all others	< 5.00	*

Dependent variable	**Independent variable**	**p (%)**	
MPK*S^{MP}	H_E, L_E, DM	> 5.00	
	all others[a]	< 5.00	*
MPK*S^{0K}	H_E, L_E	> 5.00	
	DM	≈ 4.00	*
	all others[b]	< 5.00	*
H^0/MPK	H_E, L_E, DM	> 5.00	
	all others	< 5.00	*
G^0/MPK	H_E, L_E, DM	> 5.00	
	all others[c]	< 5.00	*
H^{MP}/MPK	H_E, L_E, DM	> 5.00	
	all others[d]	< 5.00	*
G^{MP}/MPK	H_E, L_E, DM	> 5.00	
	all others[e]	< 5.00	*

Notes on data in Table 6:

- p - Probability to cancel the effect of the independent variable; MPK - melting point in K.
- The alternatives for independent variables are: Conf(=9^{n-1}), L_E, H_E, DM, ZPE, Energy, Energy aq., Solv_E (data in Table 3), n (data in Table 2), ln(n), H^0, G^0, S^0, $C_v{}^0$, S^{0K}, $C_v{}^{0K}$(data in Table 4), S^{MP}, $C_v{}^{MP}$, H^{MP}, G^{MP} (data in Table 5).
- The asterisk (*) indicate statistical significant linear associations.
- Note [a]: even if all others give linear associations with the dependent variable, however, the associations provided by MPK*S^0 are stronger than the associations provided by MPK*S^{MP} (in all cases).
- Note [b]: with two exceptions (Conf and ZPE when the associations given by MPK*S^0 are stronger than the associations provided by MPK*S^{0K}) the associations given by MPK*S^{0K} are stronger than the associations provided by MPK*S^0.
- Note [c]: associations given by H^0/MPK are stronger than the associations given by G^0/MPK.
- Note [d]: associations provided by H^0/MPK and G^0/MPK are stronger than the associations provided by H^{MP}/MPK.
- Note [e]: associations provided by G^{MP}/MPK are stronger (in all cases) than the associations provided by H^0/MPK (and H^{MP}/MPK and G^0/MPK).

The analysis of simple linear associations (Table 6) revealed that:

- There is a very little chance to obtain an equation with good estimating capacity when the melting point (either MP or MPK) is used alone as dependent variable. Practically only two predictors barely qualifies to be considered statistically significant: $C_V{}^{MP}$ and S^{MP}.
- Good chances appear when a molar heat quantity (MPK*S^0, MPK*S^{MP}, or MPK*S^{0K}) is used in place of the melting temperature. Of course, the convenience is to use one of the MPK*S^0 and MPK*S^{0K} alternatives, because doesn't require the knowledge on

the melting point and can then be used for predictions. Fortunately, MPK*S^{MP} performs the worst, but is no a clear indication till this point which of the MPK*S^0 and MPK*S^{0K} alternatives is the best to be used in building of a property-property relationship.

÷ Good chances appear also when a heat transfer quantity (H^0/MPK, G^0/MPK, H^{MP}/MPK, or G^{MP}/MPK) is used in place of the melting temperature. Again, the convenience is to use one of the H^0/MPK and G^0/MPK alternatives, because doesn't require knowing of the melting point and can then be used for predictions. Unfortunately G^{MP}/MPK performs the best, but also H^0/MPK comes as the second best alternative.

Even considering the information provided by data in Table 6, to proceed in deriving a property-property relationship is not correct, because the assigned group contains 13 paired determinations (first 13 entries in Table 5) and the pool of possible independent predictors contains 18 variables and is no recipe from which to select a part of them other than trying any possible association [11]; it is still convenient to reduce their number. Of a particular interest is the group formed by Conf ($=9^{n-1}$), n and ln(n) since they provide the shape of the association with n. By conducting a multiple linear regression with all of them included, the hope is that the survival of the fittest [12] will emerge one or two of them.

Indeed, the multiple regression analysis conducted with MPK·S^0 as dependent variable (Y) and Conf ($=9^{n-1}$), n and ln(n) as independent variables (X's) when the condition that all coefficients of the model to be statistically significant was imposed, resulted in selection of only one predictor (ln(n), eq.1 below) while analysis conducted with MPK·S^{0K} as dependent variable selected another one predictor (n, eq.2 below).

$$\text{MPK*}S^0 = Y \sim \hat{Y} = 105337(\pm 4803)_{p=4e-15}\cdot\ln(n),\ r^2_{adj} = 0.68 \quad (1)$$

$$\text{MPK*}S^{0K} = Y \sim \hat{Y} = 6637(\pm 532)_{p=4e-12}\cdot n,\ r^2_{adj} = 0.67 \quad (2)$$

The results in eq.1 and eq.2 are consistent with the experimental measurements, since the variation of the entropy, at low temperatures, increases its slope (see the data in ref. [13] as an example).

The entropies and the heat capacities at constant volume at 0 K (S^{0K} and C_v^{0K} - see the columns in Table 4) are all multipliers of R/2 and R respectively (where R is the gas constant). Since $S = k_B\cdot\log(\Omega)$, where Ω is the number of microscopic configurations [14] and molar S is $S_{molar} = N_A\cdot S = R\cdot\log(\Omega)$, the values obtained for S^{0K} (Table 4) can be used to obtain the number of microscopic configurations for monosaccharides (by inversing the logarithms: $\Omega = \log^{-1}(S^{0K}/R)$); analogously, for C_v^{0K} (Table 4), their number of energy components ($J = 2C_V/R$, see [15] for derivation of the energy components and [16] as an example of the calculation from C_p for Hydrogen) at 0 K. Even more, the data in Table 4 reveals a relationship between S^{0K} and C_v^{0K} (eq.3).

$$S^{0K}\ Y \sim \hat{Y} = C_v^{0K} + 4.1561,\ r^2 = 1.00 \quad (3)$$

It is no big guess that 4.1561 (both S^{0K} and C_v^{0K} have $J\cdot mol^{-1}\cdot K^{-1}$ as measurement unit, the same as the gas constant R and the same as 4.16212 in eq.3) is actually R/2 (R/2 = 4.1572) because the standard error of the difference, SE(S^{0K}-C_v^{0K}-R/2) is $7.3\cdot 10^{-4}$ and the probability to be $S^{0K}-C_v^{0K}-R/2 \neq 0$ from Student t distribution is 0.08% (< 5.00%).

There are a series of predictors that are linearly related (Energy, Energy aq., and Solv_E) in eq.4, (S^0, H^0 and G^0) in eq.5, (MPK, S^{MP}, H^{MP} and G^{MP}) in eq.6.

There are a series of predictors that are linearly related (Energy, Energy aq., and Solv_E) in eq.4, (S^0, H^0 and G^0) in eq.5, (MPK, S^{MP}, H^{MP} and G^{MP}) and in eq.6.

$$Solv\_E = Y \sim \hat{Y} = 2625\cdot(Energy\ aq. - Energy),\ r^2 = 1.00 \quad (4)$$

$$S^0 = Y \sim \hat{Y} = 8806\cdot(H^0 - G^0),\ r^2 = 1.00 \quad (5)$$

$$MPK*S^{MP} = Y \sim \hat{Y} = 2625670\cdot(H^{MP} - G^{MP}),\ r^2 = 1.00 \quad (6)$$

The eqs.4 to 6 are just expected results, since 1 Hartree = 2625.499 kJ/mol (see the coefficient in eq.4 and Solv_E is expressed in kJ/mol and Energy aq. and Energy in Hartrees), 1 Hartree/298.15K = 8806 J/mol (see the coefficient in eq.5 and Solv_E is expressed in J/mol and H^0 and G^0 in Hartrees) and 1 Hartree = 2625499 J/mol (see the coefficient in eq.6 and S^{MP} is expressed in J/mol/K and H^0 and G^0 in Hartrees).

Even more, eq.6 is nothing else than a well known relation among the state parameters, G = H – TS, rewritten now as TS = H – G. Somebody may say that this is wonderful, but it is not. It cannot be used for predictions of the melting points (MP), because in order to obtain H^{MP} and G^{MP} first MP should be known.

Other two predictors proved to be highly correlated with other relations among quantities: C_v^0 with Energy - H^0, and C_v^{MP} with Energy – H^{MP}.

After removal of the dependent predictors from the pool of potential descriptors, 11 still remained: n, ln(n), ZPE, DM, H^0, G^0, H^{MP}, G^{MP}, C_v^{0K}, Energy, and Energy aq. At this point, a step-by-step strategy of removal for the not statistically significant predictors was applied based on the likelihood of their coefficients.

For MPK*S^{0K} as dependent variable, Energy was first to be removed, G^0 the second, H^{MP} the third, DM the fourth, Energy aq. the fifth. At this point two variables (n and ZPE) had a probability of non-null effect between 5% and 1% (2.4% for ZPE and 1.1% for n). It was decided to continue the removal by removing ZPE. After this removal, other two were eliminated at 5% risk being in error: n and ln(n). The obtained equation is eq.7.

$$\mathrm{MPK^*S^{0K}} = Y \sim \hat{Y} = -14804(\pm 1495) + 589191(\pm 46615)\cdot H^{0} - 589157(\pm 46610)\cdot G^{MP} + 383(\pm 21)\cdot C_{v}^{0K} \quad \text{with } r_{adj}^{2} = 0.999 \tag{7}$$

The same procedure was applied to MPK*S^{MPK}, excepting in this case H^{MP} was removed implicitly (see eq.6 and its comments for the reason). ZPE was the first removed, C_v^{0K} the second, G^0 the third, n the fourth, Energy aq. the fifth. At this point one variable (DM) had a very little probability of non-null effect when compared to the rest of the variables in the model (10^{-2} vs. 10^{-6}). It was decided to continue the removal with DM. The obtained model contained two variables with high contribution to the explained variance (H^0 and G^{MP}) and other two with much less (ln(n) and Energy). It was decided to keep only first two variables. The obtained equation is eq.8.

$$\mathrm{MPK^*S^{0K}} = Y \sim \hat{Y} = -38044(\pm 4736) + 3950677(\pm 144780)\cdot H^{0} - 3950629(\pm 144769)\cdot G^{MP} \quad \text{with } r_{adj}^{2} = 0.999 \tag{8}$$

Analogously was proceeded for H^0/MPK, and the resulted equation is eq.9.

$$H^{0}/\mathrm{MPK} = Y \sim \hat{Y} = 48.7319(\pm 2.6944)\cdot H^{MP} - 48.7288(\pm 2.6944)\cdot H^{0} \quad \text{with } r_{adj}^{2} = 0.998 \tag{9}$$

The model eq.9 was used to do the assignments in relation: measured melting points and chemical structures. The entries in Table 5 from 'Alternate assignments: first option' and 'Alternate assignments: second option' were one by one alternatively joined with the main group (of first 13 entries). The decision which assignment to be kept is based on the standard error (SE) since it is an unbiased estimator of the population variance [17]. The step-by-step results of this analysis are given in Table 7.

Table 7. Assignments between chemical structures and melting points

Sample size	**Added CID**	**MPK (MP in K)**	**Standard error (SE)**	**MPK (MP in K)**	**Standard error (SE)**	**Selected**
		first option		second option		option
13	-	-	0.011047	-	0.011047	Eq.9
14	751	405.15	0.011408	418.15	0.011339	second
14	751	-	0.011339	-	0.011339	-
15	79014	418.15	0.011707	405.15	0.011846	none
14	-	-	0.011339	-	0.011339	-
15	66308	436.15	0.055598	429.15	0.010967	second

Sample size	**Added CID**	**MPK (MP in K)**	**Standard error (SE)**	**MPK (MP in K)**	**Standard error (SE)**	**Selected**
15	66308	-	0.010967	-	0.010967	-
16	65550	379.65	0.010797	352.65	0.011143	first
16	65550	-	0.010797	-	0.010797	-
17	5311110	368.15	0.010497	355.15	0.010649	first
17	5311110	-	0.010497	-	0.010497	
18	644160	416.15	0.116208	363.65	0.122194	none
17	-	-	0.010497	-	0.010497	
18	5289590	343.65	0.012235	324.15	0.017894	first
18	5289590	-	0.012235	-	0.012235	
19	102288	414.15	0.011956	401.15	0.012163	first
19	102288	-	0.011956	-	0.011956	
20	167792	424.15	0.011620	333.15	0.015747	first
20	167792	-	0.011620	-	0.011620	
21	92092	407.65	0.011487	404.15	0.011526	first
21	92092	-	0.011487	-	0.011487	

As can be seen in Table 7, the difference between the initial (from 13 paired values) estimation of the standard error (0.11047) and the final (from 21 paired values) estimation of the standard error (0.11487) is negligible – less than 5%, which indicates that the sample of 13 paired data and the sample of 21 paired data belongs to the same population (see for further details of this type of analysis [18] and [19]). Based on the results from Table 7, the following assignments have been made: MP(CID_751) = 145 °C (418.15 K), MP(CID_66308) = 156 °C (429.15 K), MP(CID_65550) = 106.5 °C (379.65 K), MP(CID_5311110) = 95 °C (368.15 K), MP(CID_5289590) = 70.5 °C (343.65 K), MP(CID_102288) = 141 °C (414.15 K), MP(CID_167792) = 151 °C (424.15 K), MP(CID_92092) = 134.5 °C (407.65 K).

Due to the lack of assignment between the structural data and the solubilities, it is even more difficult to derive a relationship able to express the solubility as a function of other properties. Actually, by using all data in Table 2 combined with the data in Table 3 and Table 4, no linear relationship can be derived to express the solubility. The main reason is the fact that actually the same solubility is assigned to two cyclic forms in three out of eight cases. Actually these forms are in slow equilibrium with each other and with the acyclic form in aqueous solution [20]. By using the square root transformation [21] applied to the solubility, a relationship were derived and in eq.10.

$$\sqrt{\text{Solubility}} = Y \sim \hat{Y} = -37(\pm 35) - 4.4(\pm 4.2)\cdot\ln(\text{Conf}) + 0.079(\pm 0.072)\cdot S^0 + 0.076(\pm 0.075)\cdot \text{ZPE} \quad \text{with } r_{adj}^2 = 0.3 \tag{10}$$

Unfortunately, even if all the coefficients of eq.10 are statistically significant, qualifying all predictors (ln(Conf), S^0 and ZPE) to belong to the model, due to the small number of measurements (11 paired data from which only 8 distinct) the eq.10 failed to provide a reliable model (probability to reject the model from Fisher's distribution is 15%).

CONCLUSIONS

The study revealed that is very difficult to derive reliable property-property relationships when the structural determinations of the substances subjected to property measurements are scarce. Therefore more structural and property determinations are essential for the advance of the knowledge in this field.

By involving statistical analysis, in this study were assigned the melting points for a number of eight monosaccharides to the corresponding deposited PubChem structure information files by constructing property-property relationships on their pool of chemical structures.

EXPERIMENTAL SECTION

The available data about melting points and solubilities of 23 monosaccharides (listed in Table 1) were collected from the literature.

The structural information as 3D geometries was taken from PubChem database. As can be seen in Table 2 the assignments between a certain monosaccharide conformation and its measured properties were not guarantied by the source of the data. Thus excepting D-glucose of which melting points are available in all three listed conformations, when the literature provided more than one melting point was necessary to conduct an assignment (alternatives listed in Table 5).

For the 23 monosaccharides 45 different geometries (3D) were available in the PubChem database. For one monosaccharide (CID 111123 corresponding to D-idose) the 3D geometry was build from its 2D geometry. On the 46 files containing different geometries of monosaccharides, a series of properties were calculated (listed in Table 3) using Spartan'14 software in the following configuration: energy calculation with Hartree-Fock (HF) method, 6-31G* dual basis, with computing of infra-red (IR) parameters and deriving of thermodynamic entities (listed in Table 4).

Table 2. Monosaccharides experimental data

Name	**$(CH_2O)_n$**	**n**	**MP/ref**		**Solubility/ref**	
D-dihydroxyacetone	$C_3H_6O_3$	3	90	1		
D-glyceraldehyde			132	11		
D-glyceraldehyde			145	2		
D-erythrose	$C_4H_8O_4$	4	129	2		
D-erythrulose						
D-threose			115	6		
D-arabinose	$C_5H_{10}O_5$	5	163	5	0.0816	12
D-arabinose			156	2		
D-lyxose			106.5	4		
D-lyxose			79.5	1		
D-ribose			95	1		
D-ribose			82	2		
D-ribulose			85	4		
D-xylose			143	4	0.12953	12
D-xylose			90.5	1		
D-xylulose			70.5	4		
D-xylulose			51	4		
D-allose	$C_6H_{12}O_6$	6	141	8	0.04489	14
D-allose			128	8		
D-altrose			104	1		
D-fructose			103	10	0.0735	15
D-galactose			164	8	0.0432	12
D-glucose			83	1	0.09447	12
D-glucose			146	5		
D-glucose			149	5		
D-gulose			151	5		
D-gulose			60	7		
D-idose			168.5	11		
D-mannose			132	3	0.25884	12
D-psicose			165	3	0.2266	13
D-sorbose			165	9		
D-tagatose			134.5	1		
D-tagatose			131	8		
D-talose			126.5	8		

Notes on data in Table 2:

- MP: melting points, in °C;
- MP/refs: 1: Chem.nlm.nih.gov; 2: ChemSpider.com; 3: CompTox CompTox.epa.gov; 4: [22]; 5: SigmaAldrich.com; 6: [23]; 7: [24]; 8: [25]; 9: [26]; 10: [27]; 11: http://drugfuture.com/chemdata/;
- Solubility: in mole fraction (mol/mol), at 25°C;
- Solubility/refs: 12: [28]; 13: converted from [29], 2.93g/g; 14: converted from [30], 47%wt; 15: extrapolated from [31] at 79.3%wt.

Table 3. Calculated molecular properties

	CID	Name (D-…)	F	Conf	L_E	H_E	DM	ZPE	Energy	Energy aq.	Solv_E
1	670	dihydroxyacetone	A	81	2.67	-12.25	3.28	244.43	-341.62377	-341.64382	-52.645
2	751	glyceraldehyde	A	81	2.62	-12.23	3.14	244.26	-341.61839	-341.63820	-52.017
3	79014	glyceraldehyde	A	81	2.72	-12.39	3.88	246.46	-341.62398	-341.63983	-41.617
4	94176	erythrose	A	729	2.37	-12.57	2.85	325.24	-455.49945	-455.51811	-48.998
5	5460177	erythrulose	A	729	2.43	-12.06	5.55	322.53	-455.49497	-455.52238	-71.961
6	439665	threose	A	729	2.07	-12.49	3.37	325.25	-455.49416	-455.51614	-57.712
7	66308	arabinose	A	6561	2.13	-12.53	2.28	401.87	-569.37840	-569.40304	-64.690
8	7044039	arabinose	C	324	4.31	-12.56	4.91	390.66	-569.38571	-569.41441	-75.327
9	65550	lyxose	A	6561	2.57	-12.32	2.65	401.39	-569.38046	-569.40478	-63.846
10	439240	lyxose	C	324	4.63	-12.38	3.93	402.92	-569.39170	-569.41829	-69.811
11	5311110	ribose	A	6561	2.47	-12.41	1.81	405.34	-569.38097	-569.40334	-58.726
12	10975657	ribose	C	324	4.21	-12.07	3.30	402.93	-569.39333	-569.41408	-54.473
13	151261	ribulose	A	6561	2.28	-12.48	2.00	400.21	-569.37434	-569.40118	-70.455
14	439203	ribulose	C	1944	3.83	-12.63	4.34	384.39	-569.38383	-569.41247	-75.173
15	644160	xylose	A	6561	2.41	-12.62	4.19	401.19	-569.37312	-569.39904	-68.046
16	89398913	xylose	C	324	4.62	-12.26	3.4	403.45	-569.38100	-569.41096	-78.662
17	444173	xylose	C	324	4.04	-12.43	3.81	393.20	-569.39701	-569.42139	-63.999
18	6027	xylose	C	324	4.68	-12.46	0.98	401.23	-569.39506	-569.41926	-63.537
19	5289590	xylulose	A	6561	2.64	-12.20	5.54	401.61	-569.38225	-569.40690	-64.704
20	439204	xylulose	C	1944	4.20	-12.69	5.53	381.81	-569.38713	-569.41472	-72.455
21	102288	allose	A	59049	2.29	-12.30	5.05	475.89	-683.23768	-683.26573	-73.659
22	439507	allose	C	2916	3.96	-12.65	2.43	456.74	-683.26202	-683.29304	-81.437
23	94780	altrose	A	59049	1.93	-12.16	5.53	474.85	-683.22915	-683.26127	-84.332
24	441032	altrose	C	2916	4.32	-11.94	2.36	477.72	-683.26918	-683.29816	-76.073
25	5984	fructose	A	59049	2.46	-12.36	7.42	470.36	-683.24467	-683.28345	-101.82
26	2723872	fructose	C	2916	4.48	-12.55	2.11	465.13	-683.27403	-683.30156	-72.267
27	3037556	galactose	A	59049	2.34	-12.53	2.62	476.41	-683.24712	-683.27771	-80.301
28	439357	galactose	C	2916	4.03	-12.44	4.35	468.31	-683.26896	-683.29681	-73.121
29	6036	galactose	C	2916	4.35	-12.51	3.65	466.37	-683.27512	-683.30107	-68.126
30	107526	glucose	A	59049	2.24	-12.46	3.06	472.20	-683.23648	-683.27206	-93.406
31	79025	glucose	C	2916	3.62	-12.58	3.04	460.59	-683.26300	-683.29710	-89.536
32	64689	glucose	C	2916	4.04	-12.54	2.29	459.19	-683.26572	-683.29809	-84.990
33	167792	gulose	A	59049	2.13	-12.74	2.77	473.93	-683.24025	-683.27051	-79.442
34	441033	gulose	C	2916	4.52	-12.25	0.54	467.88	-683.27228	-683.30047	-74.005

	CID	Name (D-…)	F	Conf	L_E	H_E	DM	ZPE	Energy	Energy aq.	Solv_E
35	111123	idose	A	59049	2.24	-12.56	4.37	484.82	-683.25735	-683.28439	-70.993
36	441034	idose	C	2916	4.21	-12.04	2.69	471.29	-683.27792	-683.30248	-64.465
37	12305800	mannose	A	59049	2.27	-12.39	4.74	473.14	-683.24174	-683.27724	-93.210
38	18950	mannose	C	2916	4.14	-12.62	2.26	458.14	-683.27324	-683.30087	-72.558
39	90008	psicose	A	59049	2.43	-12.15	5.56	472.52	-683.24826	-683.28337	-92.166
40	441036	psicose	C	2916	4.43	-12.34	2.01	463.92	-683.29092	-683.30944	-48.621
41	107428	sorbose	A	59049	2.37	-12.40	5.76	471.30	-683.24880	-683.28402	-92.456
42	439192	sorbose	C	2916	4.39	-12.32	4.86	466.68	-683.26798	-683.29185	-79.204
43	92092	tagatose	A	59049	2.20	-12.37	5.61	467.84	-683.24604	-683.28039	-90.195
44	439312	tagatose	C	2916	4.61	-12.39	4.43	463.69	-683.26650	-683.29556	-76.297
45	99459	talose	A	59049	2.21	-12.3	4.86	474.96	-683.23829	-683.26867	-79.783
46	441035	talose	C	2916	4.11	-12.27	6.09	466.81	-683.26205	-683.29147	-77.247

Notes on data in Table 3:

÷ CID - compound identifier from PubChem; F - form (A: acyclic, C - cyclic); L_E: LUMO energy (in eV); H_E: HOMO energy (in eV); DM: Dipole moment (in Debye); ZPE: Zero point energy (in kJ/mol); Energy: total energy (in a.u.; a.u. = Hartrees); Energy aq.: total energy in solvated form at infinite dilution (in a.u.; a.u. = Hartrees); Solv_E: solvation energy (in kJ/mol).

Table 4. Calculated molecular thermodynamic properties

	H^0 (a.u.)	G^0 (a.u.)	S^0 ($J\cdot mol^{-1}\cdot K^{-1}$)	C_v^0 ($J\cdot mol^{-1}\cdot K^{-1}$)	S^{0K} ($J\cdot mol^{-1}\cdot K^{-1}$)	C_v^{0K} ($J\cdot mol^{-1}\cdot K^{-1}$)
1	-341.52336	-341.56082	329.83	101.2	45.73	41.57
2	-341.51804	-341.55559	330.66	100.2	45.73	41.57
3	-341.52291	-341.56019	328.23	98.60	37.41	33.26
4	-455.36621	-455.40860	373.31	142.3	62.36	58.20
5	-455.36255	-455.40540	377.30	145.8	70.67	66.52
6	-455.36090	-455.40336	373.82	142.5	54.04	49.89
7	-569.21371	-569.26121	418.31	188.0	78.99	74.83
8	-569.22494	-569.27270	420.52	195.2	78.99	74.83
9	-569.21592	-569.26352	419.12	188.3	78.99	74.83
10	-569.22690	-569.27333	408.87	184.2	78.99	74.83
11	-569.21506	-569.26235	416.42	185.1	78.99	74.83
12	-569.22854	-569.27490	408.17	184.5	70.67	66.52
13	-569.20997	-569.25802	423.13	191.4	95.62	91.46
14	-569.22476	-569.27392	432.88	204.2	103.9	99.77

	H^0 (a.u.)	**G^0 (a.u.)**	**S^0 ($J\cdot mol^{-1}\cdot K^{-1}$)**	**C_V^0 ($J\cdot mol^{-1}\cdot K^{-1}$)**	**S^{0K} ($J\cdot mol^{-1}\cdot K^{-1}$)**	**C_V^{0K} ($J\cdot mol^{-1}\cdot K^{-1}$)**
15	-569.28865	-569.25624	419.08	188.2	78.99	74.83
16	-569.21608	-569.26216	405.82	183.5	70.67	66.52
17	-569.23537	-569.28291	418.65	193.2	87.30	83.14
18	-569.23080	-569.27752	411.47	185.6	70.67	66.52
19	-569.21749	-569.26524	420.54	189.9	87.30	83.14
20	-569.22899	-569.27818	433.18	206.0	112.2	108.1
21	-683.04234	-683.09508	464.50	236.5	95.62	91.46
22	-683.07285	-683.12749	481.15	253.3	112.2	108.1
23	-683.03418	-683.08687	464.02	237.3	95.62	91.46
24	-683.07345	-683.12530	456.59	232.1	95.62	91.46
25	-683.05095	-683.10463	472.74	242.8	112.2	108.1
26	-683.08219	-683.13587	472.70	246.2	112.2	108.1
27	-683.05159	-683.10428	463.96	236.3	87.30	83.14
28	-683.07611	-683.12922	467.70	242.4	112.2	108.1
29	-683.08298	-683.13636	470.08	243.2	112.2	108.1
30	-683.04232	-683.09538	467.29	240.0	103.9	99.77
31	-683.07265	-683.12685	477.27	249.8	112.2	108.1
32	-683.07587	-683.13011	477.65	249.9	112.2	108.1
33	-683.04548	-683.09850	466.87	238.3	103.9	99.77
34	-683.07967	-683.13273	467.31	241.9	103.9	99.77
35	-683.05924	-683.11040	450.51	224.9	78.99	74.83
36	-683.08423	-683.13684	463.34	238.9	103.9	99.77
37	-683.04731	-683.10033	466.88	238.6	95.62	91.46
38	-683.08370	-683.13821	479.99	250.8	120.6	116.4
39	-683.05385	-683.10740	471.59	241.0	112.2	108.1
40	-683.09941	-683.15337	475.09	247.8	120.6	116.4
41	-683.05475	-683.10841	472.61	242.0	112.2	108.1
42	-683.07571	-683.12909	470.13	244.0	112.2	108.1
43	-683.05310	-683.10723	476.66	245.3	120.6	116.4
44	-683.07505	-683.12875	472.97	248.1	112.2	108.1
45	-683.04323	-683.09607	465.35	237.4	95.62	91.46
46	-683.06980	-683.12283	467.00	242.8	103.9	99.77

Notes on data in Table 4:

÷ The thermodynamic quantities from Table 4 are given for standard conditions (T_0 = 298.15 K): H^0, G^0, S^0, C_V^0 and at 0K: S^{0K}, C_v^{0K}.

Table 5. Calculated molecular thermodynamic properties for acyclic forms at their melting points

	Name	CID	F	MP (K)	S^{MP}	C_v^{MP}	H^{MP}	G^{MP}
3	D-dihydroxyacetone	670	A	363.15	346.47	116.72	-341.520858	-341.568781
4	D-erythrose	94176	A	402.15	407.12	176.02	-455.360734	-455.423093
5	D-threose	439665	A	388.15	402.46	171.90	-455.356251	-455.415750
19	D-ribulose	151261	A	358.15	447.69	215.97	-569.206107	-569.267187
25	D-altrose	94780	A	377.15	501.43	277.25	-683.027889	-683.099919
45	D-fructose	5984	A	376.15	509.55	281.01	-683.044734	-683.117736
34	D-galactose	3037556	A	437.15	527.01	303.34	-683.039957	-683.127706
27	D-glucose	107526	A	356.15	495.00	269.52	-683.037758	-683.104905
37	D-idose	111123	A	441.65	517.27	295.25	-683.047430	-683.134443
39	D-mannose	12305800	A	405.15	516.81	291.30	-683.038595	-683.118346
30	D-psicose	90008	A	438.15	533.21	306.21	-683.042179	-683.131162
43	D-sorbose	107428	A	438.15	536.19	307.44	-683.043023	-683.132504
21	D-talose	99459	A	399.65	512.83	287.75	-683.035010	-683.113071
Alternate assignments: first option								
1	D-glyceraldehyde	751	A	405.15	357.24	125.28	-341.513873	-341.569000
2	D-glyceraldehyde	79014	A	418.15	357.24	126.72	-341.518231	-341.575127
9	D-arabinose	66308	A	436.15	449.85	221.91	-569.208455	-569.273418
15	D-lyxose	65550	A	379.65	450.95	222.72	-569.210598	-569.275806
7	D-ribose	5311110	A	368.15	444.24	214.67	-569.210632	-569.272924
11	D-xylose	644160	A	416.15	464.37	236.51	-569.200737	-569.274341
17	D-xylulose	5289590	A	343.65	439.48	209.07	-569.214618	-569.272142
41	D-allose	102288	A	414.15	517.60	293.47	-683.032798	-683.114444
23	D-gulose	167792	A	424.15	525.01	299.42	-683.035059	-683.119874
32	D-tagatose	92092	A	407.65	525.84	297.25	-683.044164	-683.125809
Alternate assignments: second option								
1	D-glyceraldehyde	751	A	418.15	360.22	128.12	-341.513322	-341.570673
2	D-glyceraldehyde	79014	A	405.15	354.30	123.86	-341.518795	-341.573468
9	D-arabinose	66308	A	429.15	468.38	240.79	-569.204757	-569.281317
15	D-lyxose	65550	A	352.65	440.36	211.75	-569.212468	-569.271616
7	D-ribose	5311110	A	355.15	439.26	209.36	-569.211508	-569.270926
11	D-xylose	644160	A	363.65	444.85	216.08	-569.204485	-569.266100
17	D-xylulose	5289590	A	324.15	431.37	200.99	-569.215879	-569.269137
41	D-allose	102288	A	401.15	511.86	287.66	-683.033978	-683.112185
23	D-gulose	167792	A	333.15	484.19	256.56	-683.042845	-683.104284
32	D-tagatose	92092	A	404.15	524.34	295.74	-683.044475	-683.125189

Notes on data in Table 5:

÷ Not for all monosaccharides are available the melting points; were included in Table 5 only the ones with available data.

÷ It is known that in the absence of the water monosaccharides have the tendency to take the acyclic form, while in water exists mainly in the cyclic form (see pentoses and hexoses as typical examples); therefore, the melting points were assigned to the acyclic forms (F='A' in Table 5).

÷ The melting points were converted to Kelvin scale (MP (K) in Table 5) and were used to obtain the thermodynamic quantities at the melting point S^{MP} (in $J \cdot mol^{-1} \cdot K^{-1}$), C_v^{MP} (in $J \cdot mol^{-1} \cdot K^{-1}$), H^{MP} (in a.u.), and G^{MP} (in a.u.).

ACKNOWLEDGMENTS

This paper is dedicated to Prof. Costel SÂRBU with the occasion of his 65th birthday (on January 12, 2016).

REFERENCES

[1]. D.M. Updegraff, *Analytical Biochemistry*, **1969**, *32(3)*, 420.
[2]. C.M. Hogan, "Deoxyribonucleic acid". In: Encyclopedia of Earth. National Council for Science and the Environment (Eds.: S. Draggan and C. Cleveland). Washington DC, **2010**.
[3]. D.B. Dickinson, *Phytochemistry*, **1982**, *21(4)*, 843.
[4]. A. Paulus, A. Klockow-Beck, "Structures and properties of carbohydrates". In: Analysis of Carbohydrates by Capillary Electrophoresis (Eds.: A. Paulus, A. Klockow-Beck), Vieweg Teubner Verlag, Wiesbaden, **1999**, p. 28.
[5]. R.R. Schmidt, *Angewandte Chemie International Edition in English*, **1986**, *25*, 212.
[6]. C. Zwahlen, S.J.F. Vincent, *Journal of the American Chemical Society*, **2002**, *124(24)*, 7235.
[7]. L. Jäntschi, S. D. Bolboacă, *Journal of Computational Science*, **2014**, *5*, 597.
[8]. L. Jäntschi, *Leonardo Journal of Sciences*, **2002**, *1(1)*, 31.
[9]. C. Sârbu, H.F. Pop, *Talanta*, **2005**, *65(5)*, 1215.
[10]. L. Jäntschi, *Applied Medical Informatics*, **2004**, *15(3-4),* 48.
[11]. L. Jäntschi, S.D. Bolboacă, *Leonardo Electronic Journal of Practices and Technologies*, **2007**, *6(11)*, 163.
[12]. L. Jäntschi, *Genetic algorithms and their applications*, PhD Thesis in Horticulture (PhD Advisor: R. E. Sestraş), University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, Cluj-Napoca, 2010.
[13]. E.F. Jr. Westrum, J.B. Hatcher, D.W. Osborne, *The Journal of Chemical Physics*, **1953**, *21(3)*, 419.
[14]. L. Boltzmann, *Wiener Berichte*, **1868**, *58*, 517.
[15]. L. Jäntschi, *Physical chemistry - theory, applications and problems*, Academic Direct, Cluj-Napoca, **2013**, p. 24.
[16]. M.P. Fodor, S.D. Bolboacă, L. Jäntschi, *Bulletin UASVMCN. Horticulture*, **2013**, *70(1)*, 10.
[17]. R.A. Fisher, *Metron*, **1925**, *5*, 90.
[18]. L. Jäntschi, S.D. Bolboacă, *The Scientific World Journal*, **2010**, *10*, 865.

[19]. L. Jäntschi, S.D. Bolboacă, M. V. Diudea, R. E. Sestraş, *Bulletin UASVMCN.* Agriculture, **2010**, *67(1)*, 169.

[20]. F.W. Lichtenhaler, S. Ronninger, *Journal of the Chemical Society, Perkin Transactions 2*, **1990**, *8*, 1489.

[21]. S.A. Alber, D.W. Schaffner, *Applied and Environmental Microbiology*, **1992**, *58(10)*, 3337.

[22]. P.A. Levene, R. Stuart Tipson, *The Journal of Biological Chemistry*, **1936**, *115*, 731.

[23]. G.G.S. Dutton, K.N. Slessor, *Canadian Journal of Chemistry*, **1964**, *42(3)*, 614.

[24]. R. Kosler, P. Idelmann, W.F. Dahlhof, *Angewandte Chemie International Edition in English*, **1980**, *19(7)*, 547.

[25]. M.L. Wolfrom, J.N. Schumacher, H.S. Isbell, F.L. Humoller, *Journal of the American Chemical Society*, **1954**, *76(22)*, 5816.

[26]. J.V. Karabinos, *Advances in Carbohydrate Chemistry*, **1952**, *7*, 99.

[27]. C.S. Hudson, D.H. Brauns, *Journal of the American Chemical Society*, **1916**, *38(6)*, 1216.

[28]. M.C. Gray, A.O. Converse, C.E. Wyman, *Applied Biochemistry and Biotechnology*, **2003**, *105-108*, 179.

[29]. K. Fukada, T. Ishii, K. Tanaka, M. Yamaji, Y. Yamaoka, K.-I. Kobashi, K. Izumori, *Bulletin of the Chemical Society of Japan*, **2010**, *83(10)*, 1193.

[30]. T. Kozakai, K. Fukada, R. Kuwatori, T. Ishii, T. Senoo, K. Izumori, *Bulletin of the Chemical Society of Japan*, **2015**, *88(3)*, 465.

[31]. A.E. Flood, J. Addai-Mensah, M. R. Johns, E. T. White, *Journal of Chemical & Engineering Data*, **1996**, *41(3)*, 418.