Incorrect T_g values in Table S4

G. Armeli, J.-H. Peters, and T. Koop

Faculty of Chemistry, Bielefeld University, 33615 Bielefeld, Germany

Major comment

The T_g values presented in Table S4 of Chen et al. 2026,[1] which are stated to have been calculated using the machine-learning model of Armeli et al. 2023,[2] do not agree with the values actually obtained through the public website of the model[3] when used correctly. As explained in the article by Armeli et al.[2] and on the corresponding website[3], the model consists of four different modes, depending upon which input parameters of a particular compound are known: one can either use the *Functional Group Mode* (FGM) or the *SMILES Mode* (SM). For both modes, there is a version for which the melting temperature T_m of the compound can be added as an additional feature (FGM with T_m; SM with T_m), and a version for which T_m is not required (FGM no T_m; SM no T_m). When the respective input parameters from our dataset[4] for the five compounds of Table S4 are entered, the following T_g values are obtained:

Table A: Comparison of T_g values

	T_g / K					
Compound	**Incorrect values given in Table S4**[1]	**Literature values**[1,4]	**Correct values from model by Armeli et al.**[2]			
			FGM with T_m	**FGM no T_m**	**SM with T_m**	**SM no T_m**
glucose	188.4	303	295.8	300.2	275.3	282.3
sucrose	230.5	334	344.0	342.4	342.5	339.1
glycerol	156.6	186	187.4	187.5	187.4	187.7
citric acid	186.3	284.35	284.4	284.4	283.6	283.2
1,2,6-hexanetriol	167.7	204.15	202.3	201.5	203.6	202.7
Mean Absolute Error / K	76.4		4.0	3.1	7.8	6.0

It is obvious that the T_g values given in Table S4 of Chen et al. 2026[1] (red values in the Table A above) differ very strongly from those actually calculated with our model (green values), in several cases by more than 100 K. Similarly, the mean absolute error (MAE) calculated by the authors (76.4 K) is about one order of magnitude larger than the actual MAE (about 3–8 K, see last row of Table A). We do not know how the authors arrived at the incorrect values because they neither explain which model mode they used for their calculations nor which input parameters they applied. For reference, we provide the correct input parameters in Table B below.

We disagree with the representation of our model as given in Table S4[1] and strongly request a revision. Moreover, the input parameters and the model (mode) used should be specified, too.

We further note that all of the five compounds in Table S4 are part of the training dataset used in our study and, therefore, show a higher accuracy. As outlined in our paper, we performed a cross-validated

model evaluation and provided the corresponding full data set MAE values for each mode (typically 12–13 K),[2] which is a far more representative way of evaluating a model's accuracy than testing only five selected compounds. Therefore, these MAE values should be considered.

Table B: Input parameters

Compound	FG Mode								Optional*	SMILES Mode
	$-CH_3$	$-CH_2$	-CH	-C-	-OH	-O-	=O	DBE	T_m / K	SMILES String
glucose	0	1	4	1	5	0	1	1	420	C(C1C(C(C(C(O1)O)O)O)O)O
sucrose	0	3	8	1	8	3	0	2	457	C(C1C(C(C(C(O1)OC2(C(C(C(O2)CO)O)O)CO)O)O)O)O
glycerol	0	2	1	0	3	0	0	0	291	OCC(O)CO
citric acid	0	2	0	4	4	0	3	3	432	C(C(=O)O)C(CC(=O)O)(C(=O)O)O
1,2,6-hexanetriol	0	5	1	0	3	0	0	0	300	C(CCO)CC(CO)O

* T_m is an optional input parameter for both modes.

Additional comments

(1) On page 6 of the main text the authors state "Since inorganic salts involved in this study do not exhibit T_g due to crystallization [...]". However, even if T_g of pure inorganic salts such as sodium nitrate cannot be measured directly due to crystallization, such salts have been shown to influence the T_g of organic/inorganic mixtures. For example, depending upon the T_g of the pure organic, inorganic salts have shown to increase or decrease the T_g of the mixture (cf. Figure 2 and 4 in Dette and Koop, 2015).[5] Therefore, it appears to us that ignoring the effect of the inorganic salt on T_g by representing it only by the fractional T_g of the organic phase may lead to a misrepresentation. In some cases, for example the binary mixture of sodium nitrate and sucrose, the T_g of the mixture over nearly the entire composition range is known and could be used instead, or at least used as a test case for comparison.[5]

(2) It is unclear to us why the authors used a training set and a test set because, as far as we understand, they perform a least-squares regression. This splitting procedure is followed in ML algorithms to avoid overfitting. However, in our opinion, overfitting in least-squares regression can be avoided by using a simple regression formula with not too many parameters, as it is done in the current procedure of Chen et al. 2026[1]. If the authors prefer to use a training set and a test set, would a cross-validation procedure be advantageous when splitting up the dataset, as it would allow for a more robust assessment of the model's generalizability than a single training-test split?

(3) A comparison of Figures 2 and S1 reveals that the Glucose/AS mixture is listed twice in the legend of Figure 2, and the Sucrose/$NaNO_3$ mixture is missing.

(4) In equation (5) and Figure 2, the authors report a threshold viscosity of $4.76 \cdot 10^2$ Pa · s. Could they also provide the 95% confidence intervals and the prediction intervals for this value?

(5) In lines 139–140, the authors state that our T_g model is a T_m-based method. This statement is incorrect because our ML model is a chemical structure-based model and does not require information of T_m. However, it allows T_m to be included as an additional input parameter, which slightly improves its predictions (see Table A for comparison).

References

[1] Chen, S.; Huang, Q.; Li, Y.; Pang, S.; Liu, P.; Zhang, Y.-H. Predicting Organic–Inorganic Aerosol Efflorescence Using Thermodynamically Modeled Viscosity. Preprint, *Egusphere* **2026**, 1–22. https://doi.org/10.5194/egusphere-2026-117.

[2] Armeli, G.; Peters, J.-H.; Koop, T. Machine-Learning-Based Prediction of the Glass Transition Temperature of Organic Compounds Using Experimental Data. *ACS Omega* **2023**, *8* (13), 12298–12309. https://doi.org/10.1021/acsomega.2c08146.

[3] TgML-Website (https://tgml.chemie.uni-bielefeld.de/), last accessed: 25.02.2026.

[4] Armeli, G.; Peters, J.-H.; Koop, T. Bielefeld Molecular Organic Glasses (BIGMOG) Database. *Zenodo*, **2023**. https://doi.org/10.5281/zenodo.7319484.

[5] Dette, H. P.; Koop, T. Glass Formation Processes in Mixed Inorganic/Organic Aerosol Particles. *J. Phys. Chem. A* **2015**, *119* (19), 4552–4561. https://doi.org/10.1021/jp5106967.