Molecule Structure Causal Modelling (SCM) of Choline Chloride Based Eutectic Solvents

Ž. Kurtanjek*
University of Zagreb, Faculty of Food Technology and Biotechnology (retired)

doi: https://doi.org/10.15255/CABEQ.2022.2104

Original scientific paper
Received: June 13, 2022
Accepted: September 2, 2022

manuscript dedicated to the late Prof. Paolo Alessi

This work applies the concept of structural causal modelling (SCM) for the prediction of eutectic temperatures of choline chloride based deep eutectic solvents (DES). Two SCM models were developed, one based on molecular descriptors (MD), and the other based on molecular fingerprints (MF). The models are presented in the form of directed acyclic graphs (DAG). The SCM-MD model shows that the chi simple cluster connectivity descriptor (SC.5) and a number of hydrogen atoms (nH.1) are the key causal variables. The causal relations between the model variables and eutectic temperature were determined after performing *d*-separation to block the variable confounding interference. The corresponding nonlinear causal relations were modelled by Bayes neural network with a single inner layer. Based on the SCM-MD model, a decision tree is proposed for the prediction of eutectic temperatures. Model performances were tested on a literature dataset of eutectic temperatures of ChCl based DESs. The SCM-MD model provided the most accurate prediction with an error of 7.5 °C.

Keywords:
DES, eutectic temperature, causal AI, molecular descriptors, molecular fingerprints

Introduction

The application of ionic liquids (IL) and similar deep eutectic solvents (DES) is a part of the development of new green technologies. Their specific properties, such as low vapour pressure, thermal stability, high solvating power, and catalytic intermediation, make them a promising choice for the efficient use of energy and minimisation of technology impact on human health and the environment. Compared to IL, the application of DES, belonging to a class of green solvents, is more attractive to industry due to their low cost, nontoxicity, and biodegradability. DESs are eutectic mixtures of molecules with hydrogen bond donor and acceptor properties.[1]

Their potential industrial applications are the desulfurisation of fuels and dissolution of CO_2 (carbon capture).[2–4] Potentially important is the structural transformation of wood lignin by DES treatment.[5] A broad area of application is in the harnessing of the variety of natural plant-originated compounds. For example, DESs are used as alternative solvents for the extraction of phenolic compounds from olive leaf and phenolic acids, flavonols, and flavan-3-ols from muscadine grape skins and seeds.[6,7] The mixture of choline chloride/urea (ChCl/urea) is among the first studied DESs, due to its broad availability, high biodegradability, biocompatibility, and low toxicity. There are numerous studies focused on the application of ChCl-based DESs in pharmacy. To design the appropriate DES for a particular application, computer-assisted models for the prediction of numerous thermodynamic and other physicochemical properties are needed. The properties include chemical potentials, solid-liquid equilibrium parameters, solubilities, eutectic temperatures and compositions, viscosities, etc. In most of those studies the COSMO-RS (COnductor like Screening MOdel for Real Solvents) and the corresponding software is used for quantum-chemistry-based equilibrium thermodynamics calculations.[8–15]

The objective of this work was to investigate the application of artificial intelligence (AI) algorithms for causal prediction of eutectic temperatures of ChCl-containing DESs, based on molecular descriptors and, alternatively, on molecular fingerprints. The elucidation of causal AI models is expected to provide transparent decision rules for designing of new appropriate DESs from large sets of potentially applicable molecules.

Modelling

A dataset from literature was used, collected in a previous COSMO-RS study[8] aimed at designing ChCl-based eutectic solvents for pharmaceutical applications. The dataset included 34 DESs composed

*Corresponding author: zelimir.kurtanjek@gmail.com

of ChCl and common chemicals, and a few common pharmaceuticals, and contained experimental eutectic temperatures, as well as those predicted by the simple ideal solution model and by the two alternative COSMO-RS models. In this work, 1875 molecular descriptors (1444 1D,2D descriptors, and 431 3D descriptors) of the 34 chemicals involved in forming ChCl-based DESs were calculated by the Padel software.[16] The R package RCDK based on PubChem list of 880 molecular substructures was used for calculating molecular substructures named molecular fingerprints.[16–19] All numerical and statistical evaluations were accomplished by the R software for statistics.[20] Due to a high level of multicollinearity between the molecular descriptors, the regularisation by the so-called Least Absolute Shrinkage and Selection Operator (LASSO) was applied as an elastic net, which accounted for the magnitude and number (L1 and L2) of the model parameters. The objective function F, defined by experimental and predicted data y_i and $\hat{y}_i$, model parameters β_i, and elasticity coefficients α and λ, was minimised and cross-validated:

$$F = \sum_i^N \left(y_i - \hat{y}_i \right)^2 + \lambda \sum_i^N |\beta_i| + \alpha \sum_i^N \beta_i^2 \qquad (1)$$

In the regularisation process, the importance of individual molecular descriptors was determined by the gradient boosted random forest.[20,21] The regularised set of the descriptors and the corresponding eutectic temperature data were analysed for statistical conditional independence and inference of the Bayes network model. Bayes network (BN) is a graphical presentation of statistical interdependencies between molecular descriptors and eutectic temperature. The network nodes or vertices are molecular descriptors and corresponding significant causal relations are given as unidirectional arrows or edges. In view of nonlinear interdependencies between the descriptors and eutectic temperatures and a presumed non-Gaussian distribution of unobserved random effects, the non-parametric Hilbert-Schmidt independence criteria (HSIC) were applied to two data distributions.[21,22] Thus, a set of interconnected significant conditional dependencies was obtained to create so-called stochastic unsupervised Bayes network (BN) models.[23] Two such models were created, the first one with molecular descriptors, and the second one with molecular fingerprints as the network nodes X_i; both models include eutectic temperature as the node Y. The network nodes were connected with unidirectional arrows forming a directional acyclic graph (DAG). The Markovian property of BN greatly simplifies the evaluation of the joint probability distribution P, which relates probability distributions of a node with probability distributions of the corresponding ancestors (parents, *par*). For example, the joint probability distribution with N molecular descriptors and eutectic temperature $Y = T_e$ is given by:

$$P\left(Y, X_1, X_2, \cdots X_N\right) = P\left(Y \mid par\left(Y\right)\right) \prod_{i=1}^{i=N} P\left(X_i \mid par\left(X_i\right)\right) \qquad (2)$$

For causality analysis of BN, the joint probability distribution P must be transformed to block variable confounding and/or "back door" counter causal interference. A causal BN is obtained by the application of Pearl's *do*(x) operator.[24,25] A causal effect of variable X on variable Y is determined by the so-called *d*-separation and denoted by $P(Y|do(X=x))$, which stresses the effect of setting the random variable X to a deterministic value x. The application of *d*-separation results in a corresponding adjustment set Z of unconfounded variables. The statistics of a causal effect Y upon the intervention of the variable X is given by the "truncated" probability distribution:

$$P\left(Y \middle| do(X = x)\right) = \int P\left(Y\left(x, Z = z\right)\right) \mathrm{d}P\left(z\right) \qquad (3)$$

Eq. (3) for the unknown probability distributions can be approximately evaluated from the experimental data. Commonly, the functional form of the marginal distribution of causality is depicted as a partial dependency plot. A Bayes neural network (BNN) model can be applied to approximate the underlying causal probability distribution.[26–28]

$$Y\left(x\right) = \frac{1}{N} \sum_{k=1}^{N} \mathrm{BNN}\left(x, Z_k\right) \qquad (4)$$

Having the model causal structure determined, the machine learning (ML) models, such as neural networks or decision trees, can be applied for unconfounded prediction of variables of interest, in this case of T_e. More importantly, it may provide rules for intervention policies, in this case in search for or possibly the design of new DES with desired properties. Usually, a data set is randomly five time folded, modelling is based on the four and evaluation on the fifth subset. After the procedure is repeated five times average validation metrics are void of sampling bias. In case of data sets with relatively small number of samples compared to number of model features, the data set is usually hundred times resampled, i.e. bootstrapped, and validation metrics is evaluated on synthetic data sets. After validation procedure the average absolute error e is commonly calculated as a measure of model accuracy:

$$e = \frac{1}{N} \sum_i^N \left| y_i - \hat{y}_i \right| \qquad (5)$$

Results and discussion

Molecular descriptors SCM

The causal analysis was based on the experimental data available in the literature, and the list of molecules is given in Table 1.[8] It contains 34 molecules, with three common pharmaceutical products (ibuprofen, ketoprofen, and paracetamol). It contains the experimental eutectic temperatures, those predicted by the ideal solution model as well as those predicted by the two models, ideal solution and multiplicative based on COSMO-RS. It is important to stress that ideal solution and COSMO-RS model predictions are taken from the literature as well.[8] To create the new models, the molecules were

Table 1 – *Molecule dataset used for model training and the predicted eutectic temperatures*

Molecule	Eutectic temperature/°C				
	Experiment	Ideal solution model prediction	COnductor like Screening MOdel for Real Solvents (COSMO-RS) prediction	Structural causal model with molecular descriptors (SCM-MD) prediction	Structural causal model with molecular fingerprints (SCM-MF) prediction
Urea	16	73.00	14.90	18.72	20.51
Methylurea	31	56.30	23.60	39.07	53.17
1,1-Dimethylurea	119	126.30	130.90	118.01	96.74
1,3-Dimethylurea	56	54.90	47.70	50.78	53.17
Thiourea	100	95.20	NA	18.72	20.51
Decanoic acid	34	17.80	–6.00	37.32	50.41
Dodecanoic acid	39	32.10	34.90	39.72	50.41
Tetradecanoic acid	52	41.20	47.50	43.14	50.41
Hexadecanoic acid	52	51.90	61.20	43.73	50.41
Octadecanoic acid	66	59.70	67.60	53.36	50.41
Tetradecanol	30	29.00	37.80	38.75	43.23
Hexadecanol	55	34.00	48.40	52.88	43.23
Octadecanol	56	43.40	56.90	54.71	43.23
Benzoic acid	78	72.00	–45.20	73.56	58.16
Salicylic acid	61	104.60	–107.0	62.00	58.16
5-Phenylvaleric acid	20	37.10	–11.60	44.40	50.41
D-Mannitol	112	134.20	100.30	110.82	96.94
Meso-erythritol	78	91.90	32.30	75.87	65.40
D(+)-Glucose	55	108.40	–7.40	59.74	75.53
D(–)-Fructose	53	79.90	–119.7	55.19	75.53
D(+)-Sucrose	58	139.40	–68.70	63.52	85.13
D(+)-Xylose	47	104.70	49.60	56.27	59.13
Malonic acid	12	89.93	NA	16.13	19.01
Citric acid	79	106.00	NA	76.88	58.17
Succinic acid	66	132.70	NA	50.69	58.17
Choline acetate	35	30.80	27.00	39.60	59.87
[N4444] Cl	47	41.30	60.00	49.95	69.18
[C2mim] Cl	30	23.30	26.00	36.11	49.73
Acetamide	48	43.40	32.20	40.67	49.06
Benzamide	89	87.40	71.20	89.00	88.45
Acetylsalicylic acid	63	NA	NA	62.98	58.16
Ibuprofen	51	NA	NA	61.26	50.41
Ketoprofen	75	NA	NA	73.23	50.41
Paracetamol	39	NA	NA	40.43	48.13
Mean absolute error		22.90	9.57*	7.50	15.35

Table 2 – *The most important molecular descriptors accounting for 95 % of the variance*

Mark	Name
SC.5	Chi simple cluster of 5th order
nH.1	Number of hydrogen atoms
BIC4	Bond Information Content index (neighbourhood symmetry of 4th order)
CIC4	Complementary Information Content index (neighbourhood symmetry of 4th order)

transformed into SMILES code, and PADEL online calculator was used for the determination of the 1875 molecular descriptors. The LASSO regularisation was applied to extract the most important descriptors accounting for 95 % of the total variance in the prediction of the eutectic temperatures. The results showed that only four descriptors, those given in Table 2, were responsible for this.

The datasets of the most important molecular descriptors and corresponding eutectic temperatures were subject to HSIC independence (nullity) test, of Schmidt norm of covariance matrix in Hilbert space of features. The Gaussian kernel with the test significance level of $\alpha = 0.05$ was applied to infer the causal structure, i.e., the model DAG as presented in Fig. 1. The DAG model predicts that SC.5 and nH1 descriptors are the parent variables for the eutectic temperature. Hence, SC.5 and nH.1 are to be considered as the key factors in screening molecule database for the design of new DES systems with targeted eutectic temperature. The CIC4 and BIC4 are the second level ancestor descriptors, and may also be accounted for in the design of a new DES.

Since the descriptors are statistically interrelated, the BN network (Fig. 1) is deconfounded by the *d*-separation for causal analysis. The potential adjustment sets for unconfounded inference of causal functional dependencies, i.e., the corresponding marginal distributions, are given in Table 3.

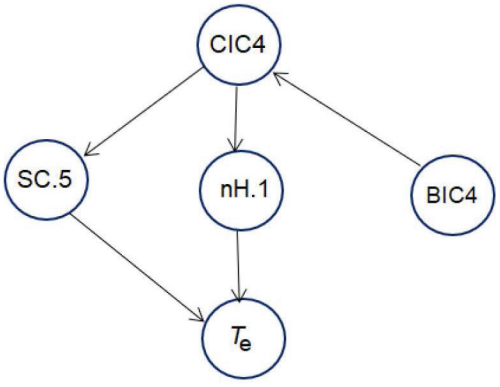

Fig. 1 – *Structural model of the causal dependence of eutectic temperature on the molecular descriptors (SCM-MD)*

Table 3 – *Adjustment Z sets for the causal determination of the eutectic temperature dependence on SC.5 and nH.1 molecular descriptors*

SC.5 → T_e	nH.1 → T_e
CIC4	CIC4
BIC4, CIC4	BIC4, CIC4
nH.1	SC.5
BIC4, nH.1	BIC4, SC.5
CIC4, nH.1	CIC4, SC.5
BIC4, CIC4, nH.1	BIC4, CIC4, SC.5

For the prediction of eutectic temperatures based on the four most important molecular descriptors, a decision tree model was created, as presented in Fig. 2. The decision tree depicts that SC.5 is the key factor in determining the eutectic temperature. It shows that, for SC.5 < 0.167 and number of hydrogen atoms (nH.1) < 5, the eutectic temperature is in the range of room temperatures. For molecules with more than five hydrogen atoms and SC.5 < 0.167, the eutectic temperature is governed by BIC4 descriptor. For BIC4 < 0.53, the eutectic temperature is about 60 °C, while for BIC4 > 0.53, it is about 40 °C. For SC.5 >0.167, the eutectic temperatures are in the range above 60 °C.

The causal partial dependency models were developed as Bayes Neural Networks (BNNs) with a single inner layer. Parental molecular descriptors were selected as inputs, as given in Table 2. The effects are shown in Fig. 3. The most "deterministic" (least variance) direct effect is due to SC.5. The eutectic temperature linearly increases with SC.5 in the range of 0.1 < SC.5 < 0.4, and is at saturation levels for lower and higher values. In the linear range, the causal sensitivity coefficient is 150 °C per unit of SC.5. The dependency of the eutectic temperature on the number of hydrogen atoms (nH.1) is nonlinear; however, for the molecules with up to 8 atoms, a proportional increase is observed. The sensitivity is about 2 °C per one H atom. However, the dependency is pronouncedly dispersed due to the influence of other descriptors. The second-level ancestor descriptors (CC4 and BIC4) show a minimal direct causal effect on temperature. However, they contribute to the prediction model, as depicted by the decision tree presented in Fig. 2.

Molecular fingerprints SCM

The molecular substructures, known as molecular fingerprints, were determined by the RCDK database with 880 substructures accounted by PubChem. The LASSO regularisation was applied to extract important molecular fingerprints. In order to

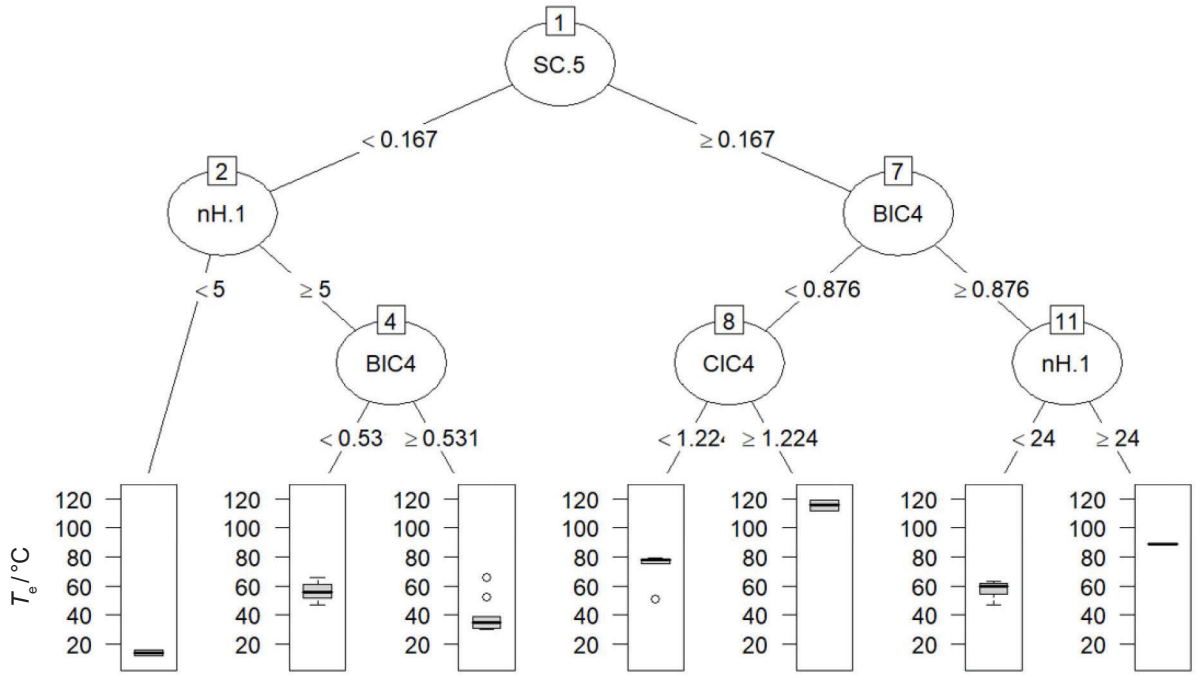

Fig. 2 – *Decision tree model for the prediction of eutectic temperatures, T_e, based on molecular descriptors*

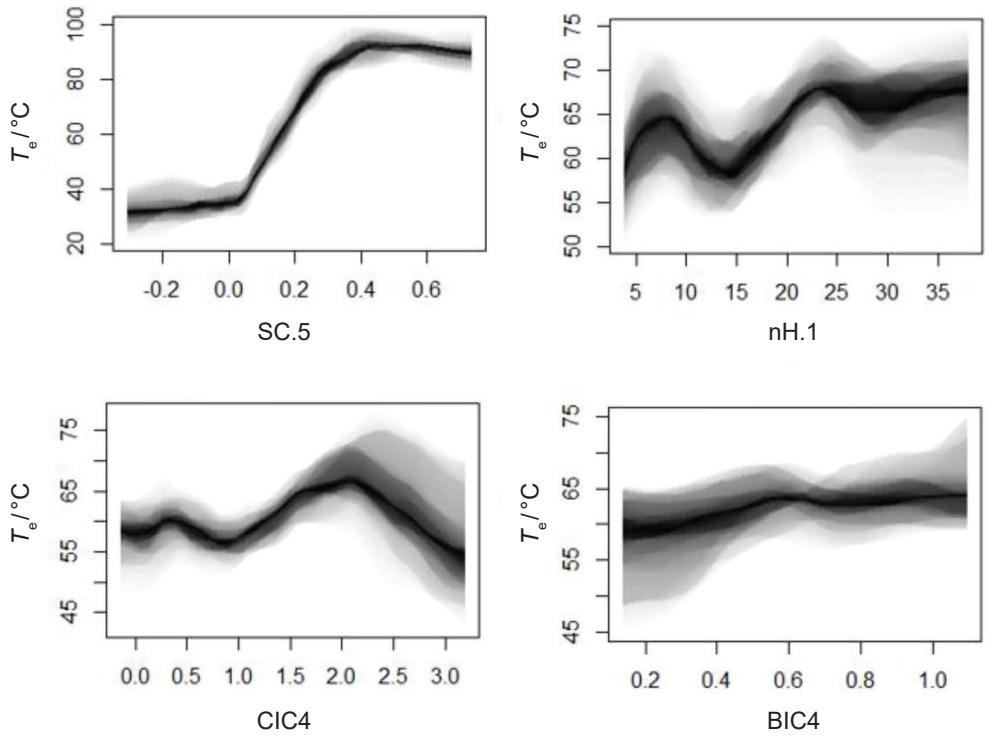

Fig. 3 – *Causal plots (partial dependency plots) of the eutectic temperature, T_e, on the molecular descriptors*

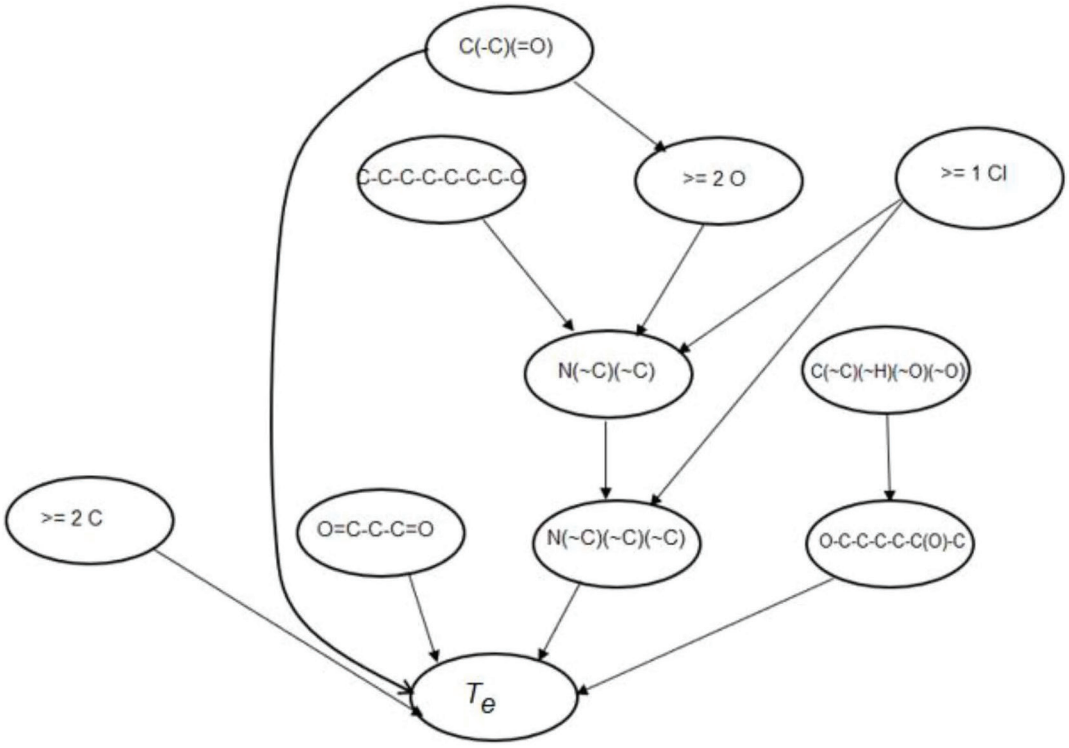

Fig. 4 – *Structural causal model (SCM) of the eutectic temperature, T_e, based on the molecular fingerprints*

obtain "first view" of DAG, only the ten most important fingerprints were selected, which accounted for 70 % of the total variance in the prediction of the eutectic temperatures. The molecular fingerprints data and the corresponding eutectic temperatures were analysed by the HSIC criteria with the significance level of $\alpha = 0.05$ to infer the DAG depicted in Fig. 4.

The network shows that the eutectic temperature is directly causally determined by five molecular substructures. Since the model with 10 variables and 34 molecules has a relatively low cross-validated accuracy (70 %), the individual causal relations were not considered.

Comparison with literature models

Performances of the developed molecular descriptor (SCM-MD) and molecular fingerprint (SCM-MF) models were compared with those reported in the literature: the ideal solution model and the COSMO-RS model, both of which assuming that choline chloride molecule was undissociated, i.e., treated as a paired ion.[8] The model predictions are inserted in Table 1.

The corresponding model predictions are compared in Fig. 5. The population of the model errors can be inferred from the boxplot and whiskers presentation. It uses descriptive statistics to plot the distributions of the numerical values, outliers as "whiskers", and skewness by displaying the data quartiles (or percentiles). All four models have negligible biases but differ significantly in the error distributions and the average absolute errors. The error distribution of the ideal solution model is skewed to negative values as it overpredicts the eutectic temperatures. The COSMO-RS errors are skewed to positive values as it underpredicts the eutectic temperatures. It is interesting to note that COSMO-RS

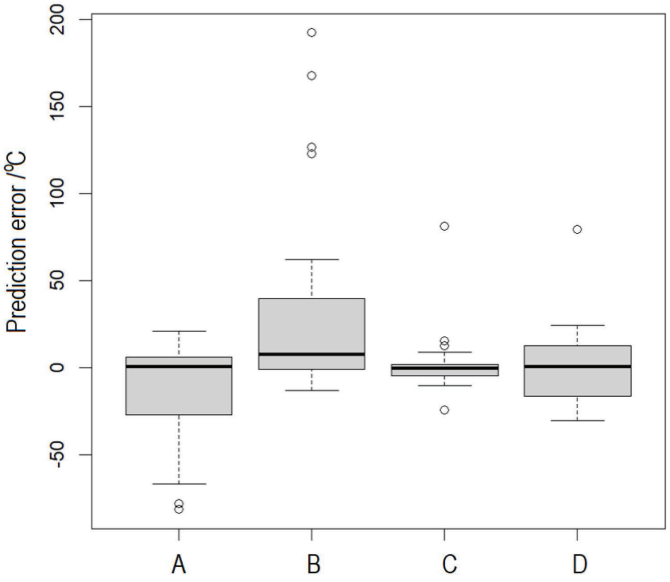

Fig. 5 – *Box plots of the errors in predictions of the eutectic temperatures by A) ideal solution model, B) COnductor like Screening MOdel for Real Solvents (COSMO-RS), C) structural causal model with molecular descriptors (SCM-MD), and D) structural causal model with molecular fingerprints (SCM-MF)*

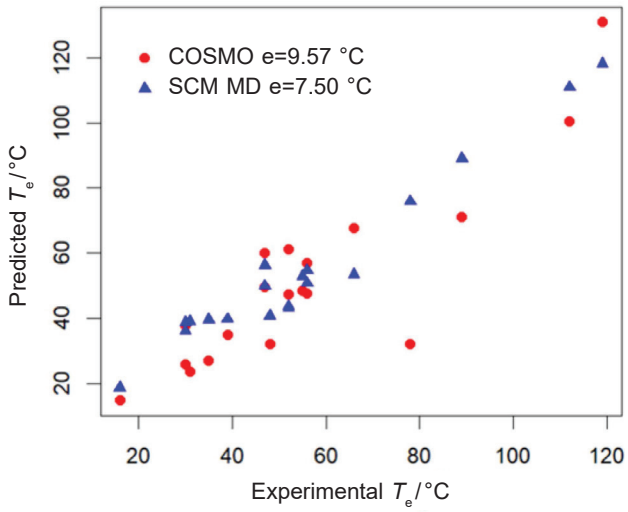

Fig. 6 – *Comparison of experimental eutectic temperatures, T_e, with the predictions with COnductor like Screening MOdel for Real Solvents (COSMO-RS), and structural causal model with molecular descriptors (SCM-MD)*

for acids (decanoic, benzoic, salicylic, 5-phenylvaleric) and sugars [D(+)-glucose, D(–)-fructose, D(+)-sucrose erroneously predicts too low eutectic temperatures, below 0 °C. Both SCM models yield correct prediction trends; however, the model with molecular descriptors has the highest accuracy (the minimal bias and variance). The average absolute errors of all the models were calculated; however, the values of the low-temperature outliers for acids and sugars were omitted from calculating the COSMO-RS errors. The SCM-MD was found to be the best model with the prediction error of 7.5 °C as compared to COSMO-RS with 9.6 °C.

A predicted vs. measured plot for two of the models is shown in Fig. 6. Standard deviations of the COSMO-RS and SCM-MD models are 13.2 °C and 6.3 C, respectively, and the Pearson correlation coefficients are $R = 0.88$ and $R = 0.97$, respectively.

Conclusions

This work introduces the concept of structural causal modelling for the prediction of eutectic temperatures for DESs based on choline chloride. Two sets of molecular predictors were considered: a set of 1875 molecular descriptors generated by the PaDEL-Descriptor software, and a set of 880 molecular fingerprints generated by the RCDK and PubChem. The causal analysis was based on the LASSO regularisation and HSIC test for the discovery of the causal structures as plotted in directed acyclic graphs (DAG). Four molecular descriptors were identified that account for 95 % of the total variance, and 10 molecular fingerprints were found which account for 70 % of the total variance of eutectic temperature prediction. Thus, two structural causal models (SCM) were created. The DAG network of the SCM based on molecular descriptors was *d*-separated to deconfound the evaluation of descriptor partial dependences. Chi simple cluster of 5th order (SC.5) and the number of hydrogen atoms (nH.1) were identified as the key molecular descriptors.

Eutectic temperature predictions by the new models were compared to the predictions by the ideal solution model and COSMO-RS model. The results showed significantly improved prediction accuracy of SCM-MD of 7.5 % when compared with that of COSMO-RS of 9.5 %. A rather low accuracy of SCM-MF model clearly showed that it is not appropriate to approximate a molecule with a sum of its substructures in an attempt to determine the eutectic temperature of DESs based on choline chloride.

In conclusion, powerful AI algorithms enable fast and accurate predictions of eutectic temperatures of DES systems to be used in scanning large databases for the potentially important molecules. A rather low predictive potential of molecular fingerprints proves the concept that a molecule property is not a sum of its substructure properties. However, the joint causal analysis of molecular descriptor and fingerprint based models has an interventional potential to be used in scanning large molecule databases in tailoring new DES systems.

Literature

1. *Cysewski, P.*, Editorial: Eutectic solvents, Crystals **10** (2020) 932. doi: https://doi.org/10.3390/cryst10100932
2. *Makoš, P., Boczkaj, G.,* Deep eutectic solvents based highly efficient extractive desulfurization of fuels – Eco-friendly approach, J. Mol. Liq. **296** (2019) 111916. doi: https://10.1016/j.molliq.2019.111916
3. *Salehi, H., Hens, R., Maulatos, O. A., Vlugt, T. J. H.*, Computation of gas solubilities in choline chloride urea and choline chloride ethylene glycol deep eutectic solvents using Monte Carlo simulations, J. Mol. Liq. **316** (2020) 113729. doi: https://doi.org/10.1016/j.molliq.2020.113729
4. *Luo, F., Liu, X., Chen, S., Song, Y., Yi, X., Xue, C., Sun, L.*, Comprehensive evaluation of a deep eutectic solvent based CO_2 capture process through experiment and simulation, ACS Sustainable Chem. Eng. **9** (2021) 30 10250. doi: https://doi.org/10.1021/acssuschemeng.1c02722
5. *Wang, S.,* Unraveling the structural transformation of wood lignin during deep eutectic solvent treatment, Front. Energy Res. **8** (2020) 48. doi: https://doi.org/10.3389/fenrg.2020.00048
6. *Alañón, M. E., Ivanović, M., Carvatza, A. M. G.,* Choline chloride derivative-based deep eutectic liquids as novel green alternative solvents for extraction of phenolic compounds from olive leaf, Arab. J. Chem. **13** (2020) 1685. doi: https://doi:10.1016/j.arabjc.2018.01.003A

7. *lrugaibah, M., Washington, W. L., Yagiz, Y., Gu, L.,* Ultrasound-assisted extraction of phenolic acids, flavonols, and flavan-3-ols from muscadine grape skins and seeds using natural deep eutectic solvents and predictive modelling by artificial neural networking, Ultrason. Sonochem. **79** (2021) 105773.
doi: https://doi.org/10.1016/j.ultsonch.2021.10577
8. *Abranches, D. O., Larriba, M., Silva, L. P., Melle-Franco, M., Palomar, J. F., Pinho, S. P., Coutihno, J. P.,* Using COSMO-RS to design choline chloride pharmaceutical eutectic solvents, Fluid Ph. Equilibria **497** (2019) 71.
doi: https://doi.org/10.1016/j.fluid.2019.06.005
9. *Salehi, H. S., Hens, R., Moultos, O. A., Vlugt, T. J. H.,* Computation of gas solubilities in choline chloride urea and choline chloride ethylene glycol deep eutectic solvents using Monte Carlo simulations, J. Mol. Liq. **316** (2020) 113729.
doi: https://doi.org/10.1016/j.molliq.2020.113729
10. *Wang, J., Song, Z., Chen, L., Xu, T., Deng, Q. Z.,* Prediction of CO_2 solubility in deep eutectic solvents using random forest model based on COSMO-RS-derived descriptors, GreenChE **2** (2021) 431.
doi: https://doi.org/10.1016/j.gce.2021.08.002
11. *Song, Z., Wang, J., Sundmacer, K.,* Evaluation of COSMO-RS for solid-liquid equilibria prediction of binary eutectic solvent systems, Green Energy & Environment **6** (2021) 371.
doi: https://doi.org/10.1016/j.gee.2020.11.020
12. *Mahmoudabadi, S. Z., Pazuki, G.,* Investigation of COSMO-SAC model for solubility and cocrystal formation of pharmaceutical compounds, Sci. Rep. **10** (2020) 19789.
doi: https://doi.org/10.1038/s41598-020-76986-3
13. *Tak, S. S., Kundu, D.,* A-priori modeling of density of deep eutectic solvent with cohesion based cubic equation of state, Chemical Thermodynamics and Thermal Analysis **5** (2022) 100026.
doi: https://doi.org/10.1016/j.ctta.2021.100026
14. *Bergua, F., Castro, M., Munoz-Embid, J., Lafuente, C., Artal, M.,* Hydrophobic eutectic solvents: Thermophysical study and application in removal of pharmaceutical products from water, Chem. Eng. J. **411** (2021) 128472.
doi: https://doi.org/10.1016/j.cej.2021.128472
15. *Nowosielski, B., Jamrógiewicz, M., Łuczak, J., Śmiechowski, M., Warmińska, D.*, Experimental and predicted physicochemical properties monopropanolamine-based deep eutectic solvents, J. Mol. Liq. **309** (2020) 113110.
doi: https://doi.org/10.1016/j.molliq.2020.113110
16. *Yap, C. W.,* PaDEL-descriptor: An open-source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. **32** (2011) 1466.
doi: https://doi.org/10.1002/jcc.21707
17. *Guha, R.*, 'Chemical Informatics Functionality in R', J. Stat. Softw. **6** (2007) 18.
doi: https://doi.org/10.18637/jss.v018.i05
18. PubChem System http://pubchem.ncbi.nlm.nih.gov
19. R version 4.0.3 (2020) https://www.r-project.org
20. *Chen, T.*, *Tong He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Jiaming, Y.*, Extreme Gradient Boosting. R package version 1.6.0.1.
https://CRAN.R-project.org/package=xgboost
21. *Gretton, A.*, *Gretton, A., Herbrich, R., Smola, A., Bousquet, O., Sc, B.,* Kernel methods for measuring independence, JMLR **6** (2005) 2075.
22. *Verbyla, P.*, *Bertille Desgranges, N. I., Wernisch, L.*, Kernel PC algorithm for causal structure detection. 2017.
https://CRAN.R-project.org/package=kpcalg
doi: https://doi.org/10.1101/138669
23. *Nagarajan, R., Scutari, M., Lèbre, S.*, Bayesian Networks in R, Springer, NY, (2013).
24. *Pear, J., Glymour, M., Jewell, N. P.*, Causal Inference in Statistics: A Primer, Wiley, New York, USA, (2021).
25. *Pearl, J.*, Causality, 2nd edition, Cambridge University Press, UK, 2021.
26. *Zhao, Q., Hastie, T.*, Causal interpretations of black-box models, J. Bus. Econ. Stat. **39** (2021) 272.
doi: https://doi.org/10.1080/07350015.2019.1624293
27. https://mran.microsoft.com/snapshot/2020-02-28/web/packages/BNN/index.html
28. https://cran.r-project.org/web/packages/BoomSpikeSlab/index.html