A Supplementary Information

A.1 Model and training hyperparameters.

To find well-performing hyperparameters we employ sparse random grid searches (RGS) for all models we consider in this work. We train 30 models with hyperparameters sampled uniformly at random from a defined search space. Each of the models is trained on the same training set and evaluated on the same test set. We pick the best model according to its mean squared error (MSE) on the test set. To report fair error estimates, we exclude all data points of this test set from the 10-fold cross validation. Since the sample size of 30 models is very small compared to the size of the search space, one might be able to further improve on our results by extending the RGS.

We schedule the learning rate either using a constant schedule or one out of nine schedules. All schedules $S_i \in \mathscr{S}$ (except for the constant one) start from an initial learning rate ε that is decreased by a factor of (maximally) 10^{-1} (multiple times) during training. After some initial warm-up steps T_w the Robbins-Monro Learning Rate Scheduler[1] computes the learning rate at epoch t from an initial learning rate of ε_0 as follows.

$$\varepsilon_t = \varepsilon_0 \,/\, \left(\frac{t - T_w}{b} + a\right)^{\gamma} \tag{1}$$

The three parameter combinations we consider are shown in Table 1. The Cyclical Learning Rate Scheduler[2] linearly increases the learning rate from an initial $\varepsilon_- = 10^{-1}\varepsilon_0$ to $\varepsilon_+ = \varepsilon_0$ and back to ε_-. This is done T_c number of times during training. We consider $T_c \in \{1,2,4\}$ The Step Learning Rate Scheduler decays an initial learning rate ε_0 by γ every T_e epochs. Consequently, at epoch t the learning rate equals $\varepsilon_0^{\gamma\lfloor t/T_e \rfloor}$. We search over three combinations of parameters that are depicted in Table 2.

T_w	γ	a	b
$1.5 \cdot 10^3$	0.5	1.0	150
$1.5 \cdot 10^3$	0.6	1.0	300
$1.5 \cdot 10^3$	0.8	1.0	900

Table 1 Combinations of parameters for the RM scheduler.

γ	T_e
0.8	1500
0.45	3750
0.1	7500

Table 2 Combinations of parameters for the Step scheduler.

The ELBO (that is maximized) can be formulated in various ways[3], two of which we include in the grid search: ELBO-KL maximizes the data log-likelihood under the variational distribution and simultaneously minimizes a KL divergence between the variational posterior distribution and the prior distribution. ELBO-Entropy maximizes the data log-likelihood as well as the log-prior under the variational distribution while simultaneously maximizing the entropy of the variational distribution.

We apply skip connections (skip con. $\neq$ none) either every layer or every second layer. Additionally in the GNN MCM model, we experiment with mean aggregation besides the sum aggregation.

Table 3 lists hyperparameter values we search over for the GNN MCM model. Table 4 lists hyperparameter values we search over for the MoFo MCM model. Table 5 lists hyperparameter values we search over for the MoFo MCM (MLE) model.

parameter	value
loss	ELBO-KL, ELBO-Entropy
learning rate	0.005, 0.001, 0.0005, 0.0001
K	4, 8, 16
lr-scheduler	0, 1, 2, 3, 4, 5, 6, 7, 8
dropout prob.	0.0, 0.1
representation dim.	16, 32, 64, 128
aggregation	sum, mean
L	1, 2, 4, 6, 8
skip con.	none, every {1, 2}-th layer
bias	true, false

Table 3 Hyperparameters of grid search for GNN MCM models.

parameter	value
loss	ELBO-KL, ELBO-Entropy
learning rate	0.005, 0.001, 0.0005, 0.0001
K	4, 8, 16
lr-scheduler	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
dropout prob.	0.0, 0.1
representation dim.	16, 32, 64, 128
L	1, 2, 4, 6, 8
skip con.	none, every {1, 2}-th layer

Table 4 Hyperparameters of grid search for MoFo MCM models.

parameter	value
learning rate	$10^{-\{2,3,4\}}$,$5 \cdot 10^{-\{4,5\}}$
K	4,8,16
lr-scheduler	0,1,2,3,4,5,6,7,8
dropout prob.	0.0,0.1
representation dim.	16, 32, 64, 128
L	1,2,4,6,8
skip con.	none, every $\{1,2\}$-th layer

Table 5 Hyperparameters of grid search for MoFo MCM (MLE) models.

GNN MCM (in-domain) ELBO-Entropy, 0.005, 16, 0, 0.1, 64, sum, 2, none, true

GNN MCM (out-of-domain) ELBO-Entropy, 0.0005, 8, 7, 0.0, 64, mean, 8, 1, true

Medina *et al.*[4] ELBO-Entropy, 0.001, 16, 7, 0.1, 16, sum, 6, none, false

MoFo MCM (in-domain) ELBO-Entropy, 0.0005, 8, 4, 0.1, 16, 1, none

MoFo MCM (out-of-domain) ELBO-Entropy, 0.001, 8, 0, 0.1, 16, 6, 1

MoFo MCM (MLE) (in-domain) 10^{-2}, 8, 1, 0.0, 64, 8, none

abbr.	description
AN	alkanes (including compounds with long alkyl groups $\geq$ 10 C-atoms)
EN	alkenes and dienes
IN	alkynes
C_ARO	aromatic compounds without heteroatoms
POL_ARO	aromatic compounds with substituent heteroatoms, pi-systems with inductive and mesomeric effects
H_ARO	aromatic compounds with substituent heteroatoms that can build stable hydrogen bonds
HET_ARO	heteroarenes
XALK	chlorine, bromine, and iodine alkanes
OL	alcohols
COO	esters
NHR	amines
ON	ketones
AL	aldehydes
COC	ethers
CN	nitrils
NO2	nitro compounds
COOH	short carboxylic acids
FF	perfluorinated compounds
CONR	amides
S_NPOL	weakly polar sulfurous compounds
S_POL	strongly polar sulfurous compounds
H2O	water and heavy water

Table 6 Manually assigned chemical categories for solutes and solvents.

A.2 Solute and Solvent Categories

Table 6 defines the chemical categories used in the "Comparison by Chemical Structure" of the main text. We assigned each solute and solvent to one of these categories manually based on their chemical structure formula using human expert knowledge. These assignments were only used in the evaluation; the model is unaware of our assignments. Due to data licensing, we do not provide the solutes and solvents that fall into the chemical categories listed in Table 6 but only provide the total number of compounds in each category.

Notes and references

1 H. Robbins and S. Monro, *The annals of mathematical statistics*, 1951, 400–407.

2 L. N. Smith, 2017 IEEE winter conference on applications of computer vision (WACV), 2017, pp. 464–472.

3 M. D. Hoffman and M. J. Johnson, Workshop in Advances in Approximate Bayesian Inference, NIPS, 2016.

4 E. I. S. Medina, S. Linke, M. Stoll and K. Sundmacher, *Digital Discovery*, 2022.