

SUPPORTING INFORMATION

Fast Predictions of Liquid-Phase Acid-Catalyzed Reaction Rates Using Molecular Dynamics Simulations and Convolutional Neural Networks

Alex K. Chew^{1,2}, Shengli Jiang¹, Weiqi Zhang¹, Victor M. Zavala¹, and Reid C. Van Lehn^{1,2*}

¹Department of Chemical and Biological Engineering, University of Wisconsin – Madison, Madison, WI, 53706, USA.

²DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, 53706, USA.

*send correspondence to: vanlehn@wisc.edu

TABLE OF CONTENTS

Section S1: Parameters for simulations and experiments.....	2
Section S2: Voxel representations and data augmentation.....	6
S2.1 Data augmentation.....	6
S2.2 Additional examples of voxel representations.....	7
Section S3: Extent of simulation data required to train SolventNet.....	9
Section S4: Three-dimensional convolutional neural network models.....	11
S4.1 VoxNet and ORION architecture	11
S4.2 Learning curves	13
Section S5: Alternative voxel representations as input for 3D CNNs	15
S5.1 Varying input channels of voxel representations	15
S5.2 Varying size of volume elements in voxel representations	16
Section S6: Comparison between 3D and 2D CNNs	19
Section S7: Propagation of error in the reaction rate predictions	20
References.....	21

Section S1: Parameters for simulations and experiments

Table S1 lists the experimental conditions, molecular descriptors (Γ, τ, δ), and experimental kinetic solvent parameters (σ_{exp}) for the 76 reactant-solvent systems used for training the models in the main text, obtained from Ref. 1. Molecular descriptors were re-scaled between 0 to 1 by min-max scaling prior to training linear and neural network models, as discussed in the main text.

Table S2 lists the number of cosolvent and water molecules used for the classical molecular dynamics (MD) simulations to generate the test set. The table also lists the experimental kinetic solvent parameters (σ^r ; the superscript indicates the reactant). These parameters were calculated using reaction rates obtained from Refs. 2 and 3 and were used to test the generalizability of SolventNet to new solvents and reactants (Fig. 5 of the main text).

Table S1. Parameters used for model training. Molecular descriptors (Γ, τ, δ) used for training linear and neural network models and experimental kinetic solvent parameters (σ_{exp}) used as labels for 76 reactant-solvent systems, obtained from Ref. 1. Molecular simulations consist of a single reactant molecule in aqueous mixtures with cosolvents at a specified organic mass fraction (m_{org}) and temperature, consistent with the experiments performed. For total number of solvent molecules in each system, please refer to the Supplementary Information of Ref. 1.

Reactant	Cosolvent	m_{org}	Temperature (K)	Γ	τ	δ	σ_{exp}
ETBE	DIO	0.90	343.15	-2.68	5.98	0.00	-0.46
ETBE	DIO	0.75	343.15	-9.06	4.76	0.00	-0.40
ETBE	DIO	0.50	343.15	-4.51	6.38	0.00	-0.16
ETBE	DIO	0.25	343.15	-1.32	3.06	0.00	-0.39
TBA	DIO	0.90	363.15	-0.83	3.40	0.17	-0.60
TBA	DIO	0.75	363.15	-6.03	2.47	0.17	-1.10
TBA	DIO	0.50	363.15	-3.40	1.23	0.17	-0.74
TBA	DIO	0.25	363.15	-0.84	1.03	0.17	-0.35
LGA	DIO	0.90	403.15	7.03	2.51	0.41	0.50
LGA	DIO	0.75	403.15	-0.69	2.74	0.41	0.32
LGA	DIO	0.50	403.15	-1.90	1.69	0.41	0.01
LGA	DIO	0.25	403.15	-0.69	1.19	0.41	0.02
PDO	DIO	0.90	433.15	5.52	2.31	0.42	0.50
PDO	DIO	0.50	433.15	-0.16	1.55	0.42	-0.03
PDO	DIO	0.25	433.15	-0.40	1.19	0.42	-0.21
FRU	DIO	0.90	373.15	14.77	3.11	0.62	1.60
FRU	DIO	0.75	373.15	2.80	2.28	0.62	0.89
FRU	DIO	0.50	373.15	-0.79	1.75	0.62	0.39
FRU	DIO	0.25	373.15	-0.55	1.36	0.62	0.17
CEL	DIO	0.90	403.15	14.71	2.05	0.63	1.15
CEL	DIO	0.75	403.15	7.59	1.73	0.63	0.84
CEL	DIO	0.50	403.15	-0.30	1.58	0.63	0.21
CEL	DIO	0.25	403.15	-0.78	1.21	0.63	0.05
XYL	DIO	0.90	403.15	9.05	2.88	0.66	1.80
XYL	DIO	0.75	403.15	4.68	1.70	0.66	1.02
XYL	DIO	0.50	403.15	0.13	1.39	0.66	0.50
XYL	DIO	0.25	403.15	-0.38	1.28	0.66	0.18
ETBE	GVL	0.90	343.15	2.47	3.70	0.00	0.25
ETBE	GVL	0.75	343.15	-5.80	4.51	0.00	-0.36
ETBE	GVL	0.50	343.15	-11.34	5.50	0.00	-0.21
ETBE	GVL	0.25	343.15	-4.37	2.25	0.00	-0.05
TBA	GVL	0.90	363.15	1.71	4.61	0.17	0.16
TBA	GVL	0.75	363.15	-1.56	2.59	0.17	-0.10
TBA	GVL	0.50	363.15	-4.81	1.39	0.17	-0.30
TBA	GVL	0.25	363.15	-3.22	1.20	0.17	-0.14
LGA	GVL	0.90	403.15	1.49	4.81	0.41	0.88

Reactant	Cosolvent	m_{org}	Temperature (K)	Γ	τ	δ	σ_{exp}
LGA	GVL	0.75	403.15	1.11	1.97	0.41	0.51
LGA	GVL	0.50	403.15	-1.17	1.19	0.41	0.18
LGA	GVL	0.25	403.15	-0.66	1.14	0.41	0.04
PDO	GVL	0.90	433.15	2.60	2.30	0.42	1.70
PDO	GVL	0.50	433.15	-0.03	1.50	0.42	0.18
PDO	GVL	0.25	433.15	-0.58	1.19	0.42	0.06
FRU	GVL	0.90	373.15	6.08	4.09	0.62	2.05
FRU	GVL	0.75	373.15	4.77	2.13	0.62	1.00
FRU	GVL	0.50	373.15	0.08	1.62	0.62	0.41
FRU	GVL	0.25	373.15	-0.33	1.21	0.62	0.19
CEL	GVL	0.90	403.15	5.19	2.27	0.63	1.60
CEL	GVL	0.75	403.15	5.88	1.70	0.63	0.72
CEL	GVL	0.50	403.15	0.39	1.48	0.63	0.23
CEL	GVL	0.25	403.15	-0.46	1.18	0.63	0.08
XYL	GVL	0.90	403.15	2.89	3.51	0.66	2.05
XYL	GVL	0.75	403.15	2.98	2.07	0.66	1.02
XYL	GVL	0.50	403.15	0.94	1.48	0.66	0.41
XYL	GVL	0.25	403.15	-0.20	1.19	0.66	0.11
ETBE	THF	0.90	343.15	-6.62	12.43	0.00	-0.41
ETBE	THF	0.75	343.15	-59.47	2.49	0.00	-0.62
ETBE	THF	0.50	343.15	-67.40	4.33	0.00	-0.67
ETBE	THF	0.25	343.15	-37.84	4.69	0.00	-0.87
TBA	THF	0.90	363.15	0.10	5.17	0.17	-0.72
TBA	THF	0.75	363.15	-21.64	2.68	0.17	-0.61
TBA	THF	0.25	363.15	-18.53	1.51	0.17	-0.41
LGA	THF	0.90	403.15	14.36	7.71	0.41	0.55
LGA	THF	0.75	403.15	13.31	1.98	0.41	0.18
LGA	THF	0.25	403.15	-3.73	1.00	0.41	0.03
PDO	THF	0.90	433.15	13.37	2.69	0.42	2.25
PDO	THF	0.75	433.15	24.42	1.39	0.42	1.34
PDO	THF	0.25	433.15	-1.69	1.19	0.42	0.52
FRU	THF	0.90	373.15	59.63	2.64	0.62	1.46
FRU	THF	0.75	373.15	122.45	1.59	0.62	0.78
FRU	THF	0.25	373.15	0.09	1.39	0.62	0.20
CEL	THF	0.90	403.15	73.56	1.73	0.63	1.00
CEL	THF	0.75	403.15	151.78	1.42	0.63	0.60
CEL	THF	0.25	403.15	-0.34	1.16	0.63	0.08
XYL	THF	0.90	403.15	38.53	2.14	0.66	1.85
XYL	THF	0.75	403.15	84.05	1.40	0.66	0.74
XYL	THF	0.25	403.15	-0.06	1.09	0.66	0.23

Table S2. Parameters used to test generalizability of SolventNet. Simulation parameters include the number of organic cosolvent molecules (N_{org}) and water molecules ($N_{\text{H}_2\text{O}}$) for each mass fraction of the organic component (m_{org}). Values are listed for aqueous mixtures containing dimethyl sulfoxide (DMSO), acetonitrile (MeCN), or acetone (ACE). All simulations included one reactant molecule. Experimental kinetic solvent parameters (σ^r) are listed for *tert*-butanol (TBA), 1,2-propanediol (PDO), fructose (FRU), and glucose (GLU), obtained from Refs. 2 and 3.

Cosolvent	m_{org}	Sim. parameters		Experimental kinetic solvent parameters			
		N_{org}	$N_{\text{H}_2\text{O}}$	σ^{TBA}	σ^{PDO}	σ^{FRU}	σ^{GLU}
DMSO ^a	0.90	816	402	-1.04	1.13	2.15	-
	0.75	617	889	-0.91	0.73	1.44	-
	0.50	357	1525	-0.60	0.17	0.61	-
	0.25	152	2027	-0.33	-0.07	0.35	-
MeCN ^b	0.90	2104	526	0.85	1.06	1.73	-
	0.75	1459	1101	0.06	0.35	1.13	-
	0.50	744	1737	-0.36	-0.04	0.69	-
	0.25	316	2118	-0.58	-0.34	0.38	-
ACE ^c	0.88	1077	484	-	-	1.17	0.23
	0.65	689	1174	-	-	0.65	0.13
	0.44	415	1661	-	-	0.16	0.00
	0.25	201	2041	-	-	-0.10	-0.07

^a The reaction temperature in DMSO-water mixtures is 363.15 K for TBA, 433.15 K for PDO, and 373.15 K for FRU.

^b The reaction temperature in MeCN-water mixtures is 363.15 K for TBA, 393.15 K for PDO, and 373.15 K for FRU.

^c The reaction temperature in ACE-water mixtures is 393.15 K for FRU and GLU. Note that for ACE-water mixtures, experimental kinetic solvent parameters are measured in the presence of hydrochloric acid, whereas all other kinetic solvent parameters are measured in triflic acid.

Section S2: Voxel representations and data augmentation

S2.1 Data augmentation

All voxel representations were augmented by rotations of 90° to train rotationally-invariant 3D CNNs. Fig. S1 shows all twenty-four unique rotations performed for a single voxel representation using xylitol in 90 wt% dioxane as an example. Voxel representations are rotated 90° along either the “x” or “y” axis. Multiple rotations are denoted by a sequence of these axes; for instance, “xy” would mean rotate 90° counter-clockwise along the x-axis, then 90° counter-clockwise along the y-axis.

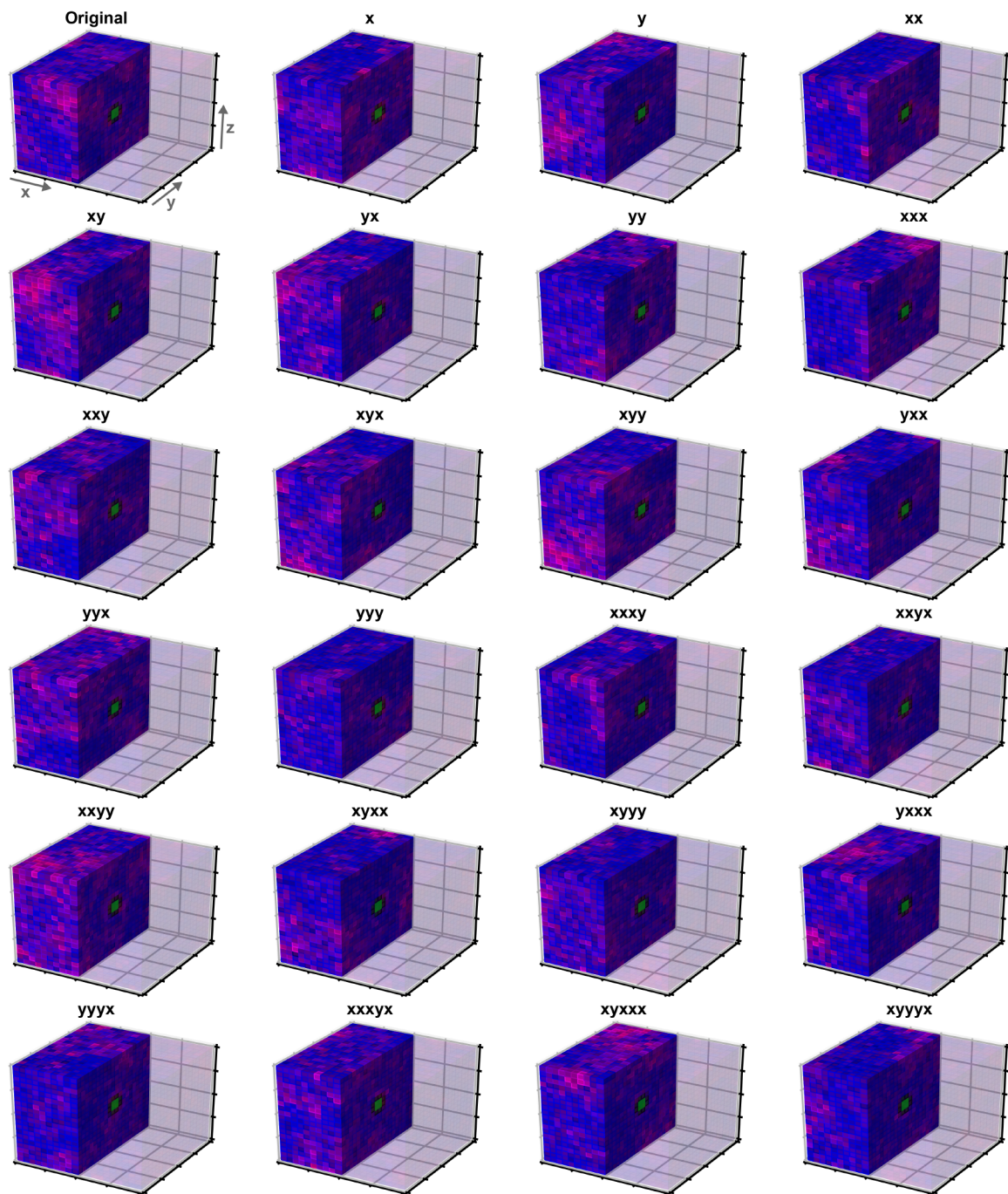


Fig. S1. Data augmentation for voxel representations. Twenty-four unique rotations performed for each voxel representation using $t = 0-2$ ns xylitol in 90 wt% dioxane as an example. The original voxel representation is shown in the top left. 90° rotations are performed along either the x or y axis in sequential order. For example, “xy” means rotate 90° counter-clockwise along the x-axis, then 90° counter-clockwise along the y-axis.

S2.2 Additional examples of voxel representations

Fig. S2 shows voxel representation examples XYL and *tert*-butanol (TBA) in aqueous mixtures with 25, 50, 75, and 90 wt% dioxane (DIO), 25 and 90 wt% γ -valerolactone (GVL) or tetrahydrofuran (THF). These examples show that increasing wt% of the cosolvent results in voxel representations with more blue voxels. Furthermore, water-enrichment could be observed for larger wt% cosolvent with red voxels near the green voxels. While some solvent features could be observed from the voxel representations in Fig. S2, it is generally challenging to visually resolve the complex solvent arrangements that leads to differences in experimental reaction rates.

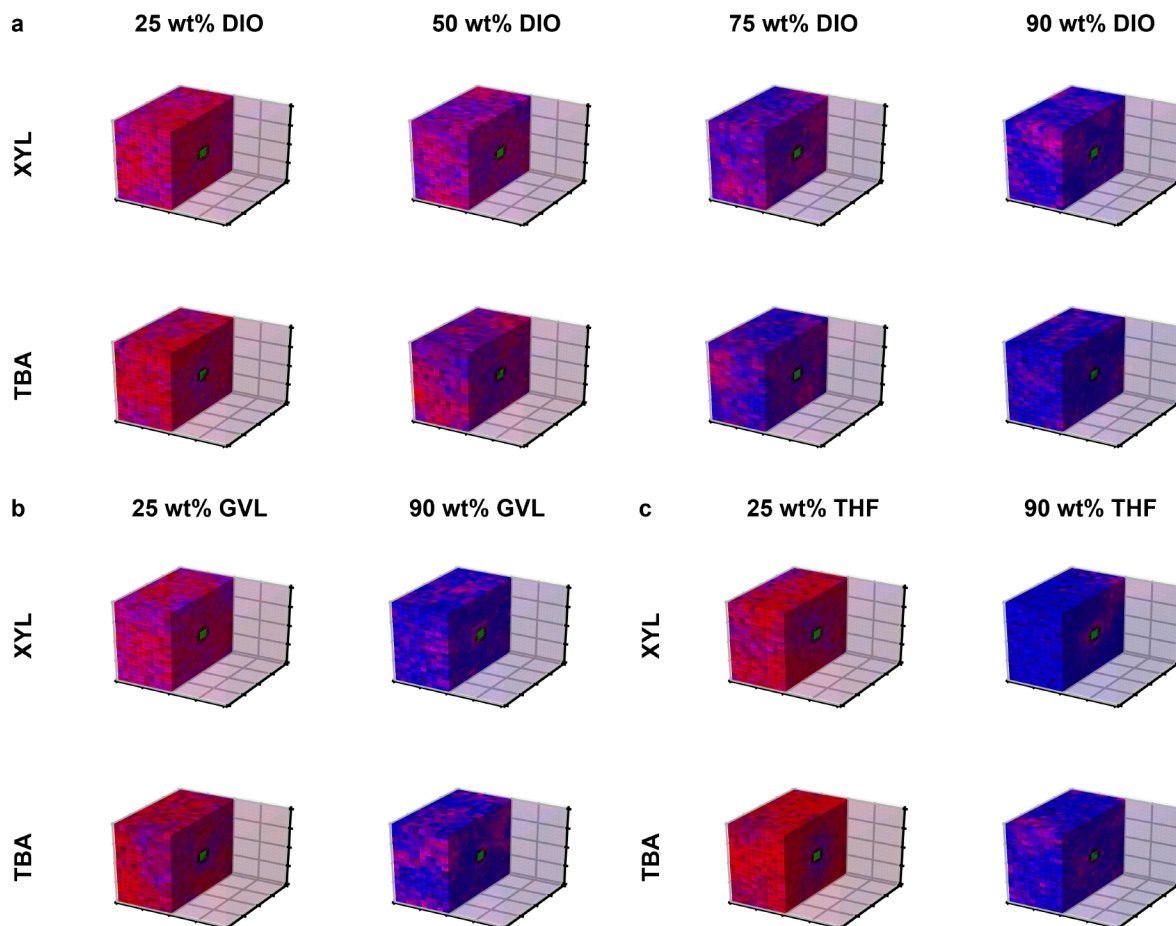


Fig. S2. Voxel representation examples. (a) Example voxel representations of xylitol (XYL) and *tert*-butanol (TBA) in aqueous mixtures with 25, 50, 75, and 90 wt% dioxane (DIO). Example voxel representations of XYL and TBA in aqueous mixtures with 25 and 90 wt% γ -valerolactone (GVL) (b) or tetrahydrofuran (THF) (c). Voxel representations are illustrated as described in Fig. 3 in the main text.

Section S3: Extent of simulation data required to train SolventNet

We trained SolventNet with varying amounts of simulation data to determine the effect on model accuracy. Simulation and experimental data for the 76 reactant-solvent combinations studied in Ref. 1 were used to train the model. This data set includes a 200-ns MD simulation trajectory for each reactant-solvent combination. To determine the amount of MD data needed for accurate predictions, we divided the first 110 ns of each trajectory into 11 equal 10-ns partitions of consecutive MD configurations. We created a collection of training sets by varying two parameters: the number of training partitions used to generate voxel representations (which affects the size of the training set) and the amount of simulation time per partition used to generate each voxel representation in both the training and test set (which affects the number of MD configurations that were averaged together per voxel representation). As described in the main text, each training partition was used to generate 24 voxel representations: one using the original atomic positions and 23 using atomic positions from data augmentation (Section S2.1). For each training set, we trained SolventNet using the same parameters and 5-fold cross validation procedure described in the main text. We assessed the accuracy of each training set based on the predicted kinetic solvent parameters for the validation set of each fold.

Fig. S3a shows the root-mean-squared error (RMSE) between the predicted (σ_{pred}) and experimental (σ_{exp}) kinetic solvent parameters when varying the number of partitions and the simulation time per partition and used to train SolventNet. The total simulation time required for model training is the product of these two quantities, whereas the simulation time required to test a new reactant-solvent combination is equal to the simulation time per partition. Increasing the number of training partitions generally decreases the RMSE, as this increases the amount of training data, but the difference is minor for greater than 2 training partitions and the difference plateaus for greater than 7 training partitions. There is a non-monotonic dependence of the RMSE on the amount of MD simulation data per partition depending on the number of partitions – large amounts of simulation time (>5 ns) tend to increase the RMSE, likely because solvent arrangements average out and precise features are not resolved. We chose to use 10 training partitions with 2 ns simulation time per partition (indicated by the black circle in Fig. S3a) because these parameters lead to a low RMSE while requiring minimal simulation data, with 2 ns representing a simulation time that consistently leads to a low RMSE for multiple different numbers of training partitions. For these parameters, a total of 20 ns of simulation data are required to train and test SolventNet while only 2 ns of data are needed to test new reactant-solvent combinations.

We next determined if SolventNet accuracy was affected by which simulation configurations were used for model training and testing. We extracted 20 ns blocks from the first 110 ns of each simulation trajectory and divided the block into 10, 2-ns training partitions (based on the choice from Fig. S3a) and used these to generate augmented voxel representations to train SolventNet. We repeated this process for a collection of blocks with the starting time of each block offset from the start of the trajectory. Fig. S3b shows the validation set RMSE between σ_{pred} and σ_{exp} for SolventNet trained with each 20-ns block. The RMSE does not substantially vary for any of the blocks. We thus chose to train all 3D CNNs using the first 20-ns simulation block, marked by the asterisk in Fig. S3b, to minimize the total amount of simulation time necessary for training.

We then tested if increasing the simulation data from 20 ns to 200 ns would affect the accuracy using SolventNet and the 5-fold cross validation training procedure. Fig. S3c shows the same parity plot as Fig 4b in the main text, which used 20 ns of MD data for each reactant-solvent

system, resulting in 10 unaugmented voxel representations per label (for each fold, a total of 14,640 augmented voxel representations is used for training). Fig. S3d shows a parity plot between predicted and experimental kinetic solvent parameters when using 200 ns of MD data to train SolventNet, which results in 100 unaugmented voxel representations per label (for each fold, a total of 146,400 augmented voxel representations is used for training). We find that using either 20 ns or 200 ns of MD data results in similar model accuracies. Thus, we chose to use 20 ns of MD data to minimize the amount of simulation data required for training.

In summary, based on these tests we decided to train the 3D CNNs using: (i) 10 training partitions, (ii) 2 ns of simulation data per partition, and (iii) the first 20-ns block in each simulation trajectory. These choices yield reasonable model accuracy with minimal input simulation data. Unless otherwise noted, all 3D CNNs were trained using these parameters.

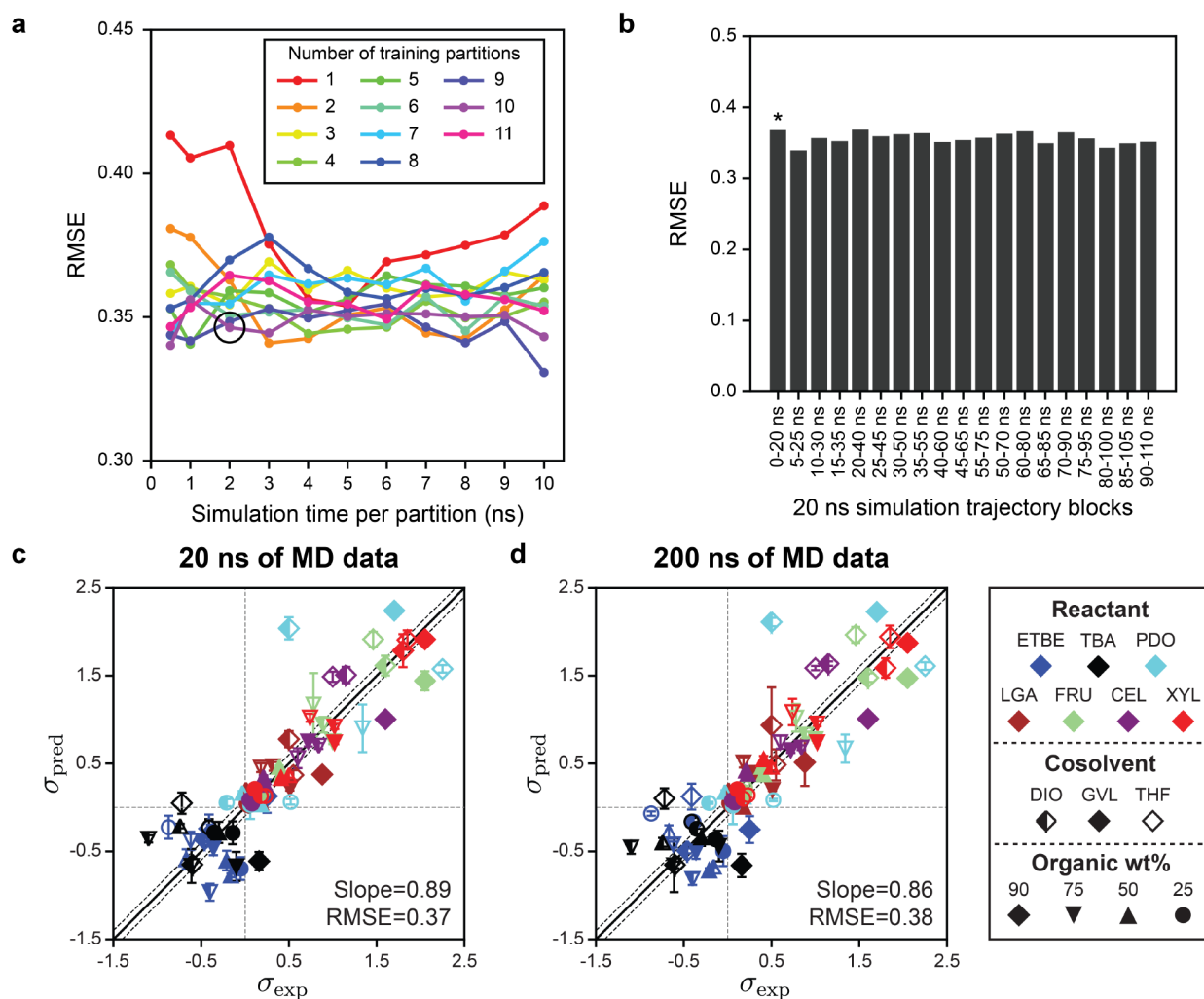


Fig. S3. Effect of varying the amount of simulation time on SolventNet accuracy. (a) Root-mean-squared error (RMSE) between σ_{pred} and σ_{exp} as a function of the number of training partitions and the simulation time per partition. SolventNet was separately trained for each set of parameters using all 76 reactant-solvent combinations and 5-fold cross validation training procedure (Figure 2b of main text). The black circle indicates the parameters used to generate voxel representations in the main text (2 ns voxel representations and 10 training partitions). (b)

RMSE between σ_{pred} and σ_{exp} as a function of the specific 20-ns block of simulation data used to generate voxel representations. Since the RMSE is similar across different 20 ns-block of simulation data, we selected the first simulation block for model training in the main text, denoted by the asterisk. Parity plot between predicted (σ_{pred}) and experimental (σ_{exp}) kinetic solvent parameters using (c) 20 ns or (d) 200 ns of MD data for each reactant-solvent system. (c) is the same as Fig. 4b of the main text. The 200 ns MD data was equally partitioned into 100 2 ns voxel representations for each reactant-solvent system and trained with SolventNet using the 5-fold cross validation training procedure.

Section S4: Three-dimensional convolutional neural network models

S4.1 VoxNet and ORION architecture

VoxNet is a shallow 3D CNN designed for real-time 3D object recognition, schematically illustrated in Fig. S4a.⁴ VoxNet consists of two convolutional layers: the first convolutional layer has 32 $5\times5\times5$ filters with a stride of 2 and the second convolutional layer has 32 $3\times3\times3$ filters with a stride of 1. A $2\times2\times2$ max-pooling layer 2 is after the convolutional layers. The results from the max-pooling layer are passed to a fully connected layer with 128 nodes. The final layer in the original model is a softmax layer for classification, which we replaced with a linear activation function for regression. VoxNet has a total of 150,689 parameters.

Orientation-boosted Voxel Nets for 3D Object Recognition (ORION) is a 3D CNN that uses the basic structure of VoxNet, schematically illustrated in Fig. S4b.⁵ ORION adds an additional auxiliary task of orientation estimation based on the assumption that the trained network induces different features with objects under rotation. We chose ORION to test against other 3D CNN models because it can reflect if our data set contains information in the azimuthal direction. We modified the number of filters and number of strides in the original ORION network because the size of the voxel representation used in this work is smaller ($20\times20\times20$) than that of the object classification data set used in the original work ($30\times30\times30$). The modified ORION network has 4 convolutional layers with 4, 8, 16, and 32 $3\times3\times3$ filters, respectively. Each convolutional layer is followed by a batch normalization layer. After the 4 convolution layers, a $2\times2\times2$ max pooling layer is applied, followed by a fully connected layer consisting of 2 dense layers with 128 and 40 nodes, respectively. The ReLU activation function is used for the fully connected layers. The modified ORION network has a total of 908,833 parameters.

Fig. S4 further shows parity plots between predicted and experimental kinetic solvent parameters using VoxNet (Fig. S4c) and ORION (Fig. S4d). Both models were trained with input data for the 76 reactant-solvent compositions from Ref. 1 using the same voxel representation and training procedure that were used to train SolventNet as described in the main text (Fig. 4b). We find that 3D CNNs perform comparably using 5-fold cross validation, with SolventNet having a higher best-slope of 0.89 compared to VoxNet and ORION, both models having a slope of 0.86.

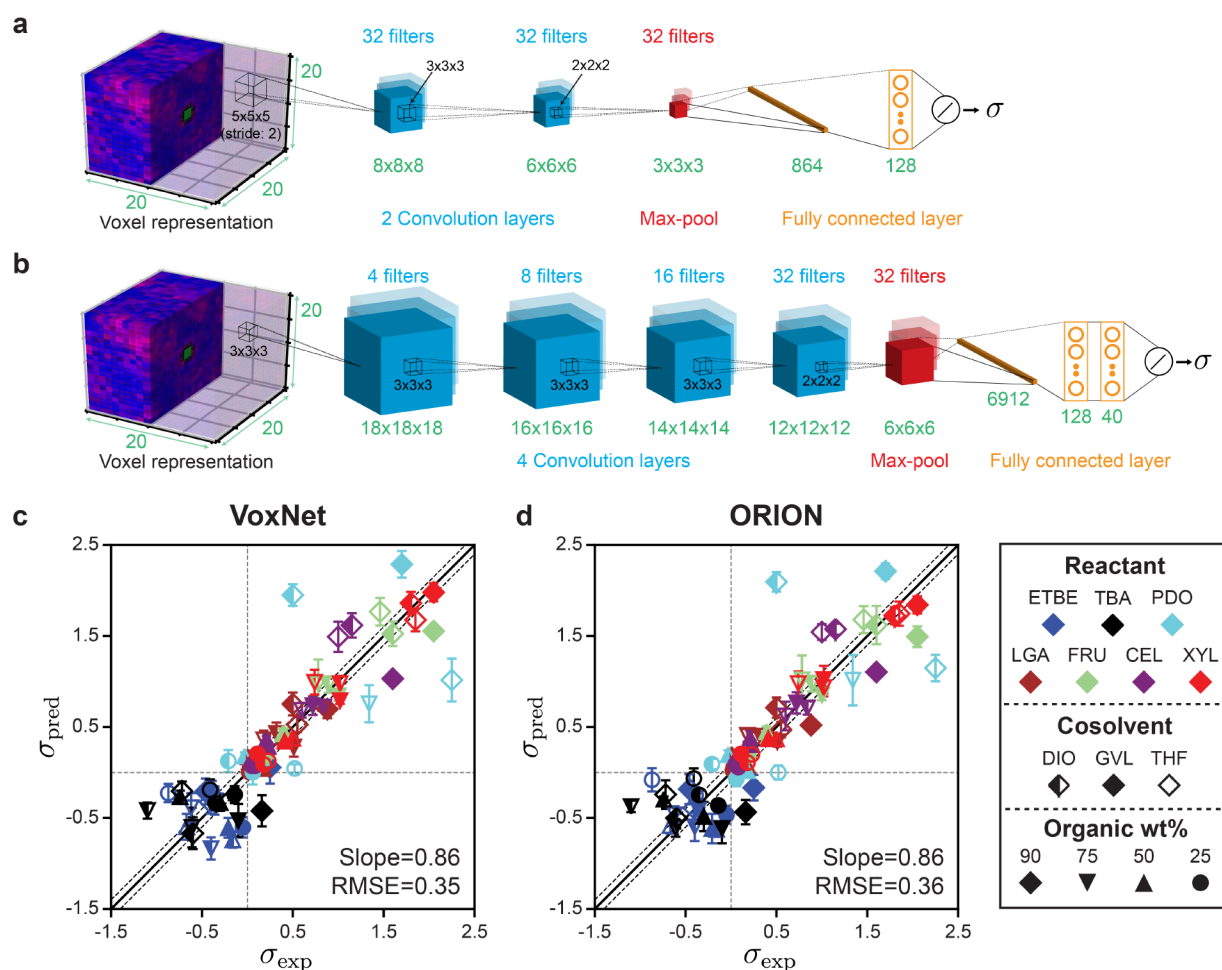


Fig. S4. Architecture and prediction accuracy of VoxNet and ORION. 3D CNN architectures for (a) VoxNet and (b) ORION. Parity plots between predicted (σ_{pred}) and experimental (σ_{exp}) kinetic solvent parameters using (c) VoxNet and (d) ORION. The models were trained for 500 epochs using simulation and experimental data for 76 reactant-solvent combinations following the 5-fold cross validation procedure described in the main text. The best-fit slope and root-mean-squared error (RMSE) between σ_{pred} and σ_{exp} are shown within the plot. The solid black line indicates perfect correlation ($\sigma_{\text{pred}} = \sigma_{\text{exp}}$), the dashed black lines show approximate experimental error, and the dashed gray lines are drawn at $\sigma_{\text{exp}} = 0$ and $\sigma_{\text{pred}} = 0$ as a guide to the eye.

S4.2 Learning curves

Learning curves were analyzed to ensure training convergence for the supervised learning task of predicting kinetic solvent parameters using the 3D CNNs. Fig. S5 shows the training and validation loss for ORION, VoxNet, and SolventNet when training with input data from all 76 reactant-solvent combinations as described in the main text using a 5-fold cross validation training procedure. The training loss (filled lines) for all 3D CNNs reaches zero after 500 epochs, indicating that the models were fully trained. The validation loss (dashed lines) all converge after ~400, indicating that the models are well-converged.

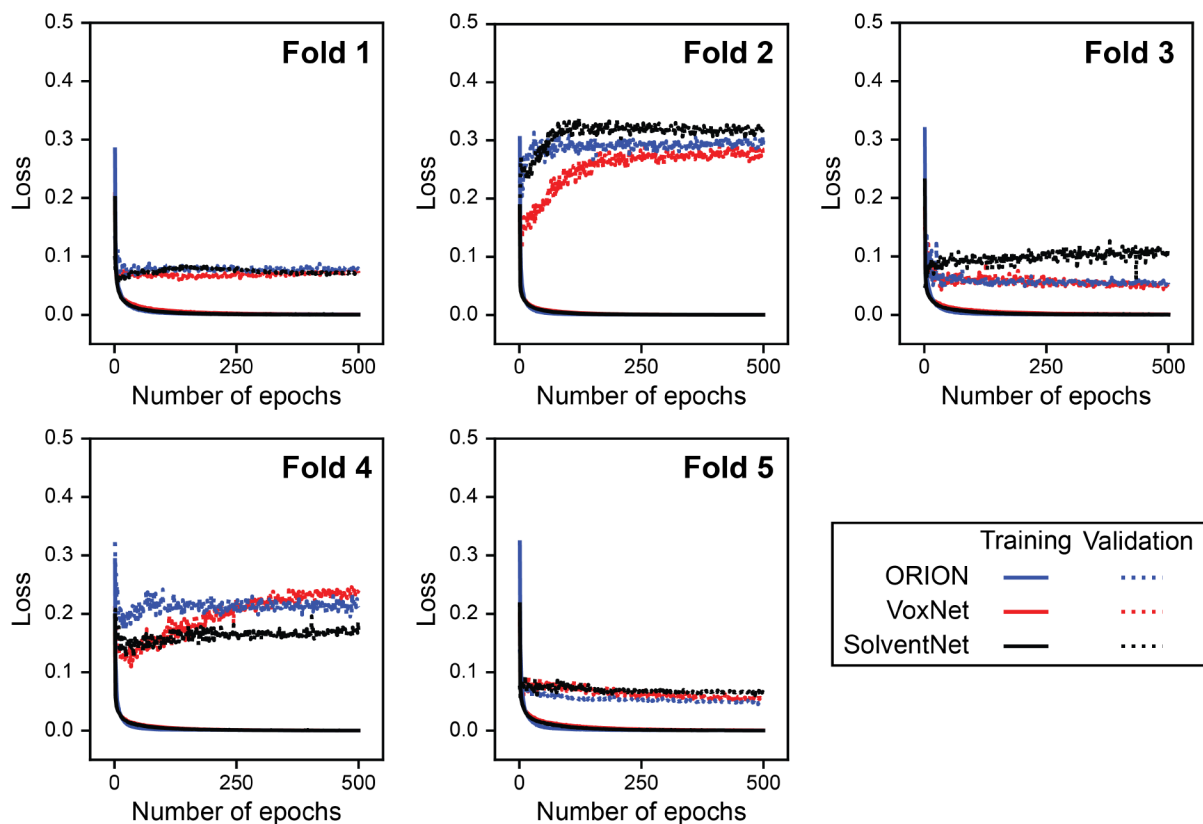


Fig. S5. Learning curves for ORION, VoxNet, and SolventNet. Training (filled lines) and validation loss (dashed lines) are shown across 500 epochs when training with input data from all 76 reactant-solvent combinations. Each of the models obtained from the 5-fold cross validation training procedure are shown.

Table S3. Model parameters and prediction accuracy of all models. Number of parameters, convolutional layers, max-pooling layers, and fully connected (FC) layers for all models studied. Input data type (either descriptors or voxel representations) and array shape used for each model is shown. The total number of arrays for either 5-fold cross validation training or using 76 reactant-solvent combinations is shown with numbers before and after augmentation (aug., see Section S2.1). The best-fit slope and root-mean-squared error (RMSE) between σ_{pred} and σ_{exp} are listed based on training and evaluating each model using all 76 reactant-solvent combinations, denoted as “Model accuracy”. Multilinear regression with molecular descriptors and using all 76 reactant-solvent compositions is denoted as “Linear (Regression)”. All other models use the 5-fold cross validation procedure discussed in the main text. The time required to train the 3D CNNs with all training set data is reported using a workstation with one GPU (NVIDIA GeForce RTX 2080 Ti) and one core. The best-fit slope and RMSE between σ_{pred} and σ_{exp} for predicting the test set consisting of 32 reactant-solvent combinations is shown for all 3D CNNs.

		Linear (Regression) ^a	Linear (5-fold)	NN	ORION	VoxNet	SolventNet	VGG16
Model details	# parameters	3	3	271	908,833	150,689	172,417	33,601,345
	# conv. layers	-	-	-	4	2	4	13
	# max-pool layers	-	-	-	1	1	2	5
	# FC layers	-	-	-	2	1	3	2
Input data	Type	Descriptors	Descriptors	Descriptors	Voxel rep.	Voxel rep.	Voxel rep.	2D Voxel rep.
	Array shape	3	3	3	20×20×20×3	20×20×20×3	20×20×20×3	32×32×3
Input data for training each fold (5-fold)	# training arrays (before aug.)	-	61	61	610	610	610	1,830
	# training arrays (after aug.)	-	61	61	14,640	14,640	14,640	7,320
	# validation arrays	-	-	15	150	150	150	450
	# parameters / # training data	-	0.0	4.4	62.1	10.3	11.8	4590.3
Input data training using all data	# training arrays (before aug.)	76	-	-	760	760	760	-
	# training arrays (after aug.)	76	-	-	18,240	18,240	18,240	-
Model accuracy	Slope	0.51	0.49	0.46	0.86	0.86	0.89	0.82
	RMSE	0.52	0.58	0.62	0.36	0.35	0.37	0.44
Training time (hr)		-	-	-	2.4	1.6	2.0	-
Test set prediction accuracy	Slope	-	-	-	0.71	0.76	0.72	-
	RMSE	-	-	-	0.48	0.57	0.48	-

^aWhen separately performing multilinear regression separately for each cosolvent-water system:

Using 27 combinations of DIO-water mixtures: Slope = 0.89, RMSE = 0.23

Using 27 combinations of GVL-water mixtures: Slope = 0.71, RMSE = 0.36

Using 22 combinations of THF-water mixtures: slope = 0.51, RMSE = 0.59

Section S5: Alternative voxel representations as input for 3D CNNs

The voxel representations used for all 3D CNNs in the main text stored the normalized occurrence of water atoms, reactant oxygen atoms, and cosolvent atoms within $(0.2\text{ nm})^3$ volume elements. We tested the effect of alternative voxel representations on SolventNet accuracy by varying the atomic positions stored in each voxel and varying the size of each volume element.

S5.1 Varying input channels of voxel representations

We tested the effect of storing normalized occurrences of different atom types within $(0.2\text{ nm})^3$ volume elements. Fig. S6 summarizes the five different voxel representations tested using xylitol (XYL) in 90 wt% DIO as an example. The 2-channel voxel representation stores the normalized occurrences of water and cosolvent atomic positions. The 3-channel voxel representations all store the normalized occurrences of water and cosolvent atomic positions in 2 channels. The third channel stored either the normalized occurrence of oxygen atoms of the reactant, atoms within hydroxyl groups on the reactant, or all reactant atoms. The 4-channel voxel representation stores the normalized occurrences of water atoms, cosolvent atoms, reactant oxygen atoms, and all reactant atoms.

Table S4 compares the best-fit slope and RMSE between σ_{pred} and σ_{exp} calculated when training and testing SolventNet using all 76 reactant-solvent combinations and five-fold cross validation procedure for each voxel representation shown in Fig. S6. The 2-channel representation has the lowest training accuracy with a slope of 0.82 and RMSE of 0.34. All other representations with $(0.2\text{ nm})^3$ volume elements have a higher training accuracy with slope ≥ 0.85 and comparable RMSEs. Of the voxel representations in Fig. S6, the 3-channel representation that stores the normalized occurrences of reactant oxygen atoms has one of the highest slopes, although we note no substantial difference between the different representations. We use this voxel representation for all results in the main text and mark it with an asterisk in Table S4.

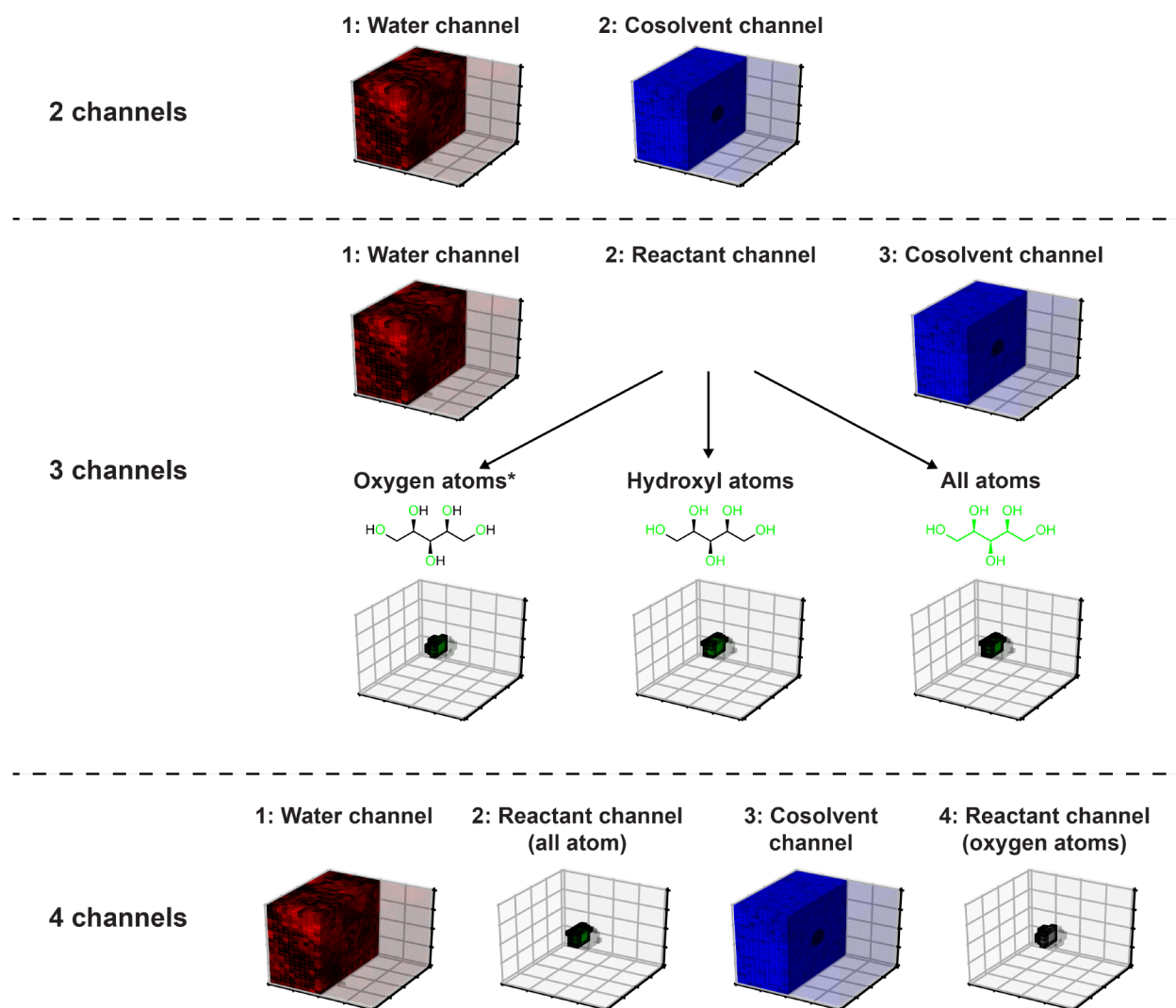


Fig. S6. Different voxel representations used as input to SolventNet. Voxel representations are visualized by coloring different channels using xylitol in 90 wt% dioxane as an example. For each MD configuration, a $20 \times 20 \times 20$ grid of $(0.2 \text{ nm})^3$ volume elements was centered on the reactant. All voxel representations include normalized occurrences averaged over 2 ns of MD data and are visualized with half the voxels transparent. The asterisk indicates the voxel representation used in the main text.

S5.2 Varying size of volume elements in voxel representations

We tested the effect of varying the volume element size on the prediction accuracy of SolventNet. Tests were performed using the 3-channel voxel representation that stores the oxygen atoms of the reactant. The voxel representation used in the main text divided a $(4 \text{ nm})^3$ cubic box into a $20 \times 20 \times 20$ grid of $(0.2 \text{ nm})^3$ volume elements. Using the same $(4 \text{ nm})^3$ cubic box, we generated voxel representations in which the box was divided into a $16 \times 16 \times 16$ grid of $(0.25 \text{ nm})^3$ and a $32 \times 32 \times 32$ grid of $(0.125 \text{ nm})^3$ volume elements. Fig. S7 shows examples of each voxel representation for XYL in 90 wt% DIO. The $32 \times 32 \times 32$ representation shows higher resolution

features compared to the other two representations, such as more clearly defined locations of reactant oxygen atoms.

Table S4 compares the effect of the volume element size on SolventNet predictions. Training SolventNet with $16 \times 16 \times 16$ or $32 \times 32 \times 32$ voxel representations lead to similar slope and RMSEs compared to $20 \times 20 \times 20$ voxel representations. These results suggest that changing the volume element size do not significantly influence model accuracy. Therefore, we select $20 \times 20 \times 20$ voxel representations for all the results in the main text.

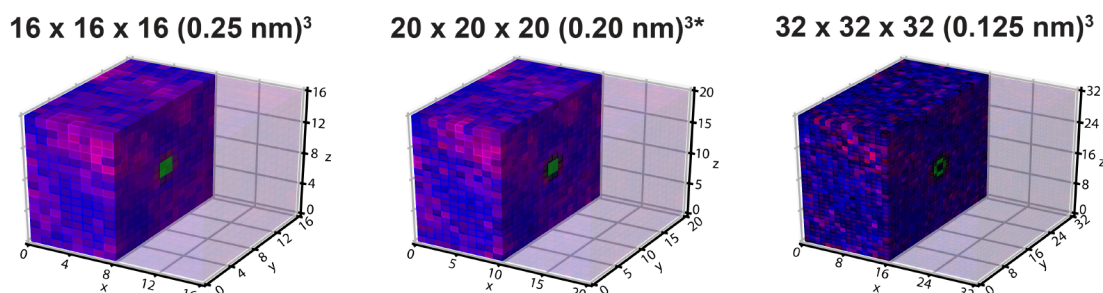


Fig. S7. Voxel representations with different volume element sizes. Example voxel representations of XYL in 90 wt% DIO when varying the volume element size and corresponding number of voxel elements. Each voxel representation is labeled with the grid dimensions and volume element size. Voxel representations are visualized by showing normalized occurrences of water atoms, cosolvent atoms, and reactant oxygen atoms in red, green, and blue, respectively. Half of the voxels are transparent to visualize the spatial distribution around the reactant. The asterisk indicates the voxel representation used in the main text.

Table S4. Effect of voxel representation on SolventNet predictions. Accuracy metrics are reported for all different voxel representations tested. The training accuracy reports the best-fit slope and root-mean-squared error (RMSE) between σ_{pred} and σ_{exp} obtained when training and evaluating SolventNet with 5-fold cross validation. The asterisk indicates the voxel representation used in the main text.

Voxel representation	Slope	RMSE
2 channel	0.82	0.34
3 channel with oxygens of reactant*	0.89	0.37
3 channel with hydroxyls of reactant	0.90	0.36
3 channel with all atom reactant	0.85	0.33
4 channel	0.88	0.36
3 channel with oxygens of reactant (16 x 16 x 16 x 3)	0.89	0.38
3 channel with oxygens of reactant (32 x 32 x 32 x 3)	0.87	0.36
VGG16 (32 x 32)	0.82	0.44

Section S6: Comparison between 3D and 2D CNNs

We also predicted σ using a 2D CNN to determine the effect of input data dimensionality on prediction accuracy. We selected the VGG16 network as a representative 2D CNN architecture since it was one of the top performers in the ImageNet Large-Scale Visual Object Challenge (Ref. 6). VGG16 has a deep architecture that has a total of 13 convolutional layers, 5 max-pooling layers, and 2 fully connected layers, totaling 33,601,345 parameters. We replaced the final softmax layer of VGG16 with a linear activation function for the regression task of predicting σ . We flattened the 3D voxel representations to provide 2D input images for VGG16. Since this network has a restriction that the smallest image is 32×32 , we convert $32 \times 32 \times 32$ voxel representations that store reactant oxygen atoms from Section S5.2 into 2D images (pixel fields). Fig. S8a illustrates the conversion from 3D to 2D for XYL in 90 wt% DIO. For each voxel representation, the data was flattened by averaging along either the x -, y -, or z -axis to produce three 32×32 images. Each image was augmented by rotating counterclockwise by 90° , 180° , and 270° . Therefore, for 10 training partitions, there are 120 images for a single reactant-solvent combination and 9,120 total images for all 76 reactant-solvent combinations. We trained the model following the 5-fold cross validation procedure described in the main text, in which each fold uses 20% of the 76 reactant-solvent combinations as validation data (450 validation samples) and the remaining reactant-solvent combinations were used as training data (7,320 training samples after augmentation). We trained VGG16 using the same training parameters as SolventNet (*e.g.* Adam optimizer, *etc.*).

Fig. S8b shows the parity plot between σ_{pred} and σ_{exp} using VGG16. The best-fit slope and RMSE for this data are 0.82 and 0.44, respectively. The larger RMSE for VGG16 compared to SolventNet (Table S4) suggests that this model may overfit the training examples, given that VGG16 has a large number of parameters. Fig. S8c shows the learning curve with the training and validation loss across 500 epochs when training VGG16. The learning curve shows convergence of training VGG16 within 500 epochs. Therefore, we find that the using a 3D CNN both provides a more accurate description of system geometry, reduces the number of parameters needed for the network compared to a top-performing 2D CNN, and generalizes more accurately. These findings agree with our hypothesis that 3D CNNs are more natural architectures to study the effect of solvent spatial distributions on reaction rates.

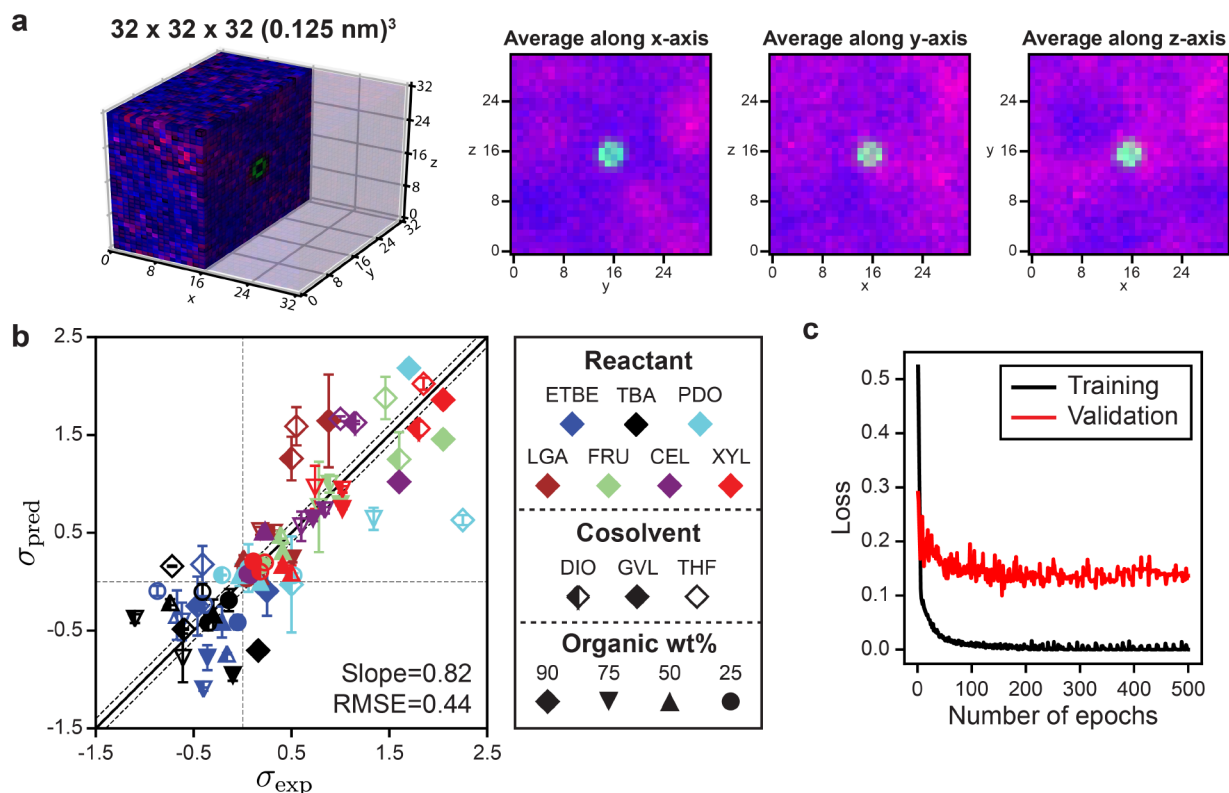


Fig. S8. Prediction of kinetic solvent parameters using the VGG16 model. (a) Example 32×32×32 3D voxel representation for xylitol in 90 wt% dioxane and corresponding 32×32 2D images used to train VGG16. 3D voxel representations were flattened to 2D images by averaging the voxel data along the *x*-, *y*-, and *z*-axes. (b) Parity plot between σ_{pred} and σ_{exp} using the VGG16 model when training and testing with 76 reactant-solvent combinations. The best-fit slope and root-mean-squared error (RMSE) between σ_{pred} and σ_{exp} are shown within the plot. (c) Learning curve with training and validation loss for training VGG16 with 2D images. Only one of five models are shown.

Section S7: Propagation of error in the reaction rate predictions

Since σ is defined as log-ratio of reaction rates, the error in the kinetic solvent parameter predictions may result in large errors in the actual reaction rates (k_{org}^r). Fig. S9 shows the predicted versus experimental reaction rates using the data from Fig. 4b of the main text. k_{org}^r is computed by rearranging Equation 1 of the main text, shown in Equation S1:

$$k_{org}^r = 10^\sigma \cdot k_{H_2O}^r \quad (S1)$$

$k_{H_2O}^r$ is the apparent rate constant for the reaction in pure water, taken from Ref. 2. The error of the reaction rates in mixed-solvent environments (Δk_{org}^r) was propagated using the total differential of k_{org}^r and assuming that the error associated with $k_{H_2O}^r$ is negligible, shown in Equation S2:

$$\Delta k_{org}^r = 10^\sigma \cdot \ln(10) k_{H_2O}^r \Delta \sigma \quad (S2)$$

$\Delta\sigma$ is the estimated predicted error. Fig. S9 shows that the predicted reaction rates in mixed-solvent environments are well-correlated with experimental reaction rates, with the exception of some outliers for CEL and LGA.

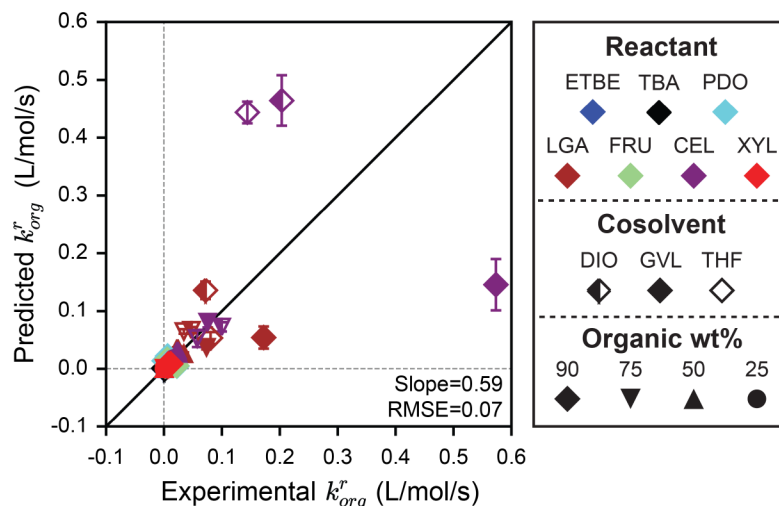


Fig. S9. Propagated error in the prediction of actual reaction rates. Parity plots between predicted and experimental reaction rates (k_{org}^r) using the SolventNet model described in Fig. 4b in the main text. Most points are overlapping along the main diagonal.

References

1. T. W. Walker, A. K. Chew, H. X. Li, B. Demir, Z. C. Zhang, G. W. Huber, R. C. Van Lehn and J. A. Dumesic, *Energ Environ Sci*, 2018, **11**, 617-628.
2. M. A. Mellmer, C. Sanpitakseree, B. Demir, P. Bai, K. W. Ma, M. Neurock and J. A. Dumesic, *Nature Catalysis*, 2018, **1**, 199-207.
3. A. H. Motagamwala, K. Huang, C. T. Maravelias and J. A. Dumesic, *Energ Environ Sci*, 2019, DOI: 10.1039/C9EE00447E.
4. D. Maturana and S. Scherer, *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, 922-928.
5. N. Sedaghat, M. Zolfaghari, E. Amiri and T. Brox, *arXiv preprint arXiv:1604.03351*, 2016.
6. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. H. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, *Int J Comput Vision*, 2015, **115**, 211-252.