Electronic Supplementary Material (ESI) for Energy & Environmental Science.
This journal is © The Royal Society of Chemistry 2018

This version of the ESI published on the 15th of May 2018 replaces the original version published on the 8th of December 2017.

SUPPORTING INFORMATION

Universal Kinetic Solvent Effects in Acid-catalyzed Reactions of Biomass-derived Oxygenates

Theodore W. Walker[(a)†], Alex K. Chew[(a)†], Huixiang Li[(b, c)], Benginur Demir[(a,d)], Prof. Z. Conrad Zhang[(b)], Prof. George W. Huber[(a)], Prof. Reid C. Van Lehn[(a)*], and Prof. James A. Dumesic[(a,d)*]

(a) Department of Chemical and Biological Engineering, University of Wisconsin – Madison, Madison, WI, 53706 (United States)
(b) Dalian National Laboratory for Clean Energy, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian 116023, Liaoning, China
(c) University of Chinese Academy of Sciences, Beijing 100049, China
(d) DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, 53706, United Sates

† These authors contributed equally

*send correspondence to: vanlehn@wisc.edu and jdumesic@wisc.edu

TABLE OF CONTENTS

Section S1: Experimental Details 2
S1.1 Materials 4
S1.2 Methods 4
S1.3 Reaction kinetics modeling 5
Section S2: Experimental Results 9
S2.1 Effects of temperature and reactor headspace on the apparent rate constants 14
Section S3: Molecular Dynamics Simulation Details 15
S3.1 System preparation 15
S3.2 Calculation of number of molecules within system 15
S3.3 Calculation of accessible hydroxyl fraction (δ) 16
S3.4 Calculation of radial distribution function 16
S3.5 Estimation of cutoff between local and bulk domains 16
S3.6 Robustness of the preferential exclusion coefficient (Γ) 16
S3.7 Hydrogen bonding lifetimes (τ) 18
S3.8 Rescaling of multidescriptor correlation model 19
S3.9 Simulation size, cutoff radii for local domain, and data for multidescriptor correlation model 19
Section S4: Molecular dynamic simulation additional results 23
S4.1 Two-descriptor correlation model for DIO 23
S4.2 Multidescriptor correlation model for GVL and THF 24
S4.3 Coefficients of multidescriptor correlation model without re-scaling procedure 24
S4.4 Error in multidescriptor correlation coefficients 25
S4.5 Correlation between Γ and τ 26

S4.6 Comparison of alternative models using the Akaike Information Criterion 26
S4.7 Predictability of multidescriptor correlation model 28
References 30

Section S1: Experimental Details

Table S1 presents the kinetic solvent parameters that describe the rates of the seven Brønsted-acid-catalyzed reactions considered in this study, including those data reported in our prior work,[1] in aqueous mixtures of γ-valerolactone (GVL), 1,4-dioxane (DIO), and tetrahydrofuran (THF). These kinetic descriptors capture changes in the apparent rate constant for each reaction as a function of solvent composition *at a fixed temperature,* compared to the same rate constant in pure water ($k^i_{H_2O}$)*:*

$$\log_{10}\left(\frac{k^i_{org,j}}{k^i_{H_2O}}\right) = \sigma^i_{org,j} \tag{S1}$$

where ($k^i_{org,j}$) is the apparent rate constant for the ith reaction, and the subscript denotes the identity and composition (in j^{th} mass fraction) of the organic phase. Herein, we demonstrate that these descriptors are themselves weak functions of temperature and reactor headspace, which allows for direct comparisons to be made between reactions that occur at different temperatures, and in solvents that exert significantly different vapor pressures. Confidence intervals were calculated at the 95% confidence level based on the propagation of error resulting from two known sources of uncertainty, which are discussed in the sections that follow. Details relating to experimental protocols and reaction kinetics modeling are also discussed.

Table S1. Kinetic solvent parameters describing the rates of the Brønsted-acid-catalyzed reactions considered in this study as a function of solvent composition. *Reaction conditions: see Table S2.*

Xylitol (XYL) dehydration to yield 1,4-anhydroxylitol

$$\boldsymbol{k^{XYL}_{H_2O} = 1.04x10^{-4}\ L\ mol^{-1}s^{-1}}$$

$\boldsymbol{T_{rxn}}$ **= 403 K**

Mass fraction (*j*) Of the organic phase	$\sigma^{XYL}_{GVL,j}$	$\sigma^{XYL}_{DIO,j}$	$\sigma^{XYL}_{THF,j}$
90 wt%	2.05 +/- 0.07	1.80 +/- 0.08	1.85 +/- 0.10
75 wt%	1.02 +/- 0.08	1.02 +/- 0.08	0.74 +/- 0.13
50 wt%	0.41 +/- 0.09	0.50 +/- 0.08	N/A
25 wt%	0.11 +/- 0.10	0.18 +/- 0.08	0.23 +/- 0.07

Fructose (FRU) dehydration to yield 5-hydroxymethylfurfural (HMF) [(a)]

$$\boldsymbol{k^{FRU}_{H_2O} = 1.95\ x\ 10^{-4}\ L\ mol^{-1}s^{-1}}$$

$\boldsymbol{T_{rxn}}$ **= 373 K**

Mass fraction (*j*) Of the organic phase	$\sigma^{FRU}_{GVL,j}$	$\sigma^{FRU}_{DIO,j}$	$\sigma^{FRU}_{THF,j}$

90 wt%	2.05	1.60	1.46
75 wt%	1.00 +/- 0.10[b]	0.89 +/- 0.10[b]	0.78
50 wt%	0.41	0.39	N/A
25 wt%	0.19	0.17	0.20

Cellobiose (CEL) hydrolysis to yield glucose (GLU)

$$k_{H_2O}^{CEL} = 1.44x\ 10^{-2}\ L\ mol^{-1}s^{-1}$$

T_{rxn} = 403 K

Mass fraction (*j*) Of the organic phase	$\sigma_{GVL,j}^{CEL}$	$\sigma_{DIO,j}^{CEL}$	$\sigma_{THF,j}^{CEL}$
90 wt%	1.60 +/- 0.07	1.15 +/- 0.08	1.00 +/- 0.08
75 wt%	0.72 +/- 0.07	0.84 +/- 0.08	0.60 +/- 0.07
50 wt%	0.23 +/- 0.08	0.21 +/- 0.08	N/A
25 wt%	0.08 +/- 0.08	0.05 +/- 0.08	0.08 +/- 0.08

1,2-propanediol (PDO) dehydration to yield propanal [(a)]

$$k_{H_2O}^{PDO} = 1.26\ x\ 10^{-4}\ L\ mol^{-1}s^{-1}$$

T_{rxn} = 393 K

Mass fraction (*j*) Of the organic phase	$\sigma_{GVL,j}^{PDO}$	$\sigma_{DIO,j}^{PDO}$	$\sigma_{THF,j}^{PDO}$
90 wt%	1.70	0.50	2.25
75 wt%	N/A	N/A	1.34
50 wt%	0.18	-0.03	N/A
25 wt%	0.06	-0.21	0.52

Levoglucosan (LGA) hydration to yield GLU

$$k_{H_2O}^{LGA} = 2.27\ x\ 10^{-2}\ L\ mol^{-1}s^{-1}$$

T_{rxn} = 403 K

Mass fraction (*j*) Of the organic phase	$\sigma_{GVL,j}^{LGA}$	$\sigma_{DIO,j}^{LGA}$	$\sigma_{THF,j}^{LGA}$
90 wt%	0.88 +/- 0.11	0.50 +/- 0.09	0.55 +/- 0.09
75 wt%	0.51 +/- 0.09	0.32 +/- 0.09	0.18 +/- 0.10
50 wt%	0.18 +/- 0.09	0.01 +/- 0.12	N/A
25 wt%	0.04 +/- 0.10	0.02 +/- 0.11	0.03 +/- 0.09

***tert*-butanol (TBA) dehydration to yield isobutene [(a)]**

$$k_{H_2O}^{TBA} = 1.39\ x\ 10^{-3}\ L\ mol^{-1}s^{-1}$$

T_{rxn} = 363 K

Mass fraction (*j*) Of the organic phase	$\sigma_{GVL,j}^{TBA}$	$\sigma_{DIO,j}^{TBA}$	$\sigma_{THF,j}^{TBA}$
90 wt%	0.16	-0.60	-0.72
75 wt%	-0.10 +/- 0.10[b]	-1.10 +/- 0.10[b]	-0.61

50 wt%	-0.3	-0.74	N/A
25 wt%	-0.14	-0.35	-0.41

Ethyl-*tert*-butyl ether (ETBE) hydrolysis to yield TBA and ethanol

$$k_{H_2O}^{ETBE} = 1.04\ x\ 10^{-4}\ L\ mol^{-1}s^{-1}$$

T_{rxn} = 343 K

Mass fraction (j) Of the organic phase	$\sigma_{GVL,j}^{ETBE}$	$\sigma_{DIO,j}^{ETBE}$	$\sigma_{THF,j}^{ETBE}$
90 wt%	0.25 +/- 0.09	-0.46 +/- 0.09	-0.41 +/- 0.11
75 wt%	-0.36 +/- 0.10	-0.40 +/- 0.09	-0.62 +/- 0.13
50 wt%	-0.21 +/- 0.19	-0.16 +/- 0.08	-0.67 +/- 0.09
25 wt%	-0.05 +/- 0.08	-0.39 +/- 0.08	-0.87 +/- 0.08

a) taken from reference[1] b) this study

S1.1 Materials

Water (Fisher W5; HPLC grade), gamma-valerolactone (Sigma-Aldrich; >98%), 1,4-dioxane (Sigma-Aldrich, anhydrous; >99.8%), and tetrahydrofuran (Acros; anhydrous, 99.9%) were used as received. D-xylitol (Acros Organics; 99+%), β-D-fructose (Sigma; >99%), β-D-cellobiose (Sigma-Aldrich; 99%), levoglucosan (Sigma-Aldrich; 99%), *tert*-butanol (Sigma-Aldrich; >99.7%), ethyl-*tert*-butyl ether (Sigma-Aldrich; 99%), 1,4-D-anhydroxylitol (ChemCruz; >99%), 5-hydroxymethylfurfural (Sigma-Aldrich; 99%), β-D-glucose (Sigma-Aldrich; ACS reagent grade), and ethanol (Sigma-Aldrich; anhydrous, >99.5%) were used as reactants, and/or as standards for developing calibration curves. Trifluormethane sulfonic (triflic) acid (99%; extra pure) was obtained from Acros Organics.

S1.2 Methods

Reaction kinetics measurements for *tert*-butanol (TBA), and fructose (FRU) dehydration were carried out using methods described in in prior work.[1] All other reactions were carried out in closed, 10 mL thick-walled glass batch reactors. In a typical experiment, 2 mL of solution, consisting of an appropriate amount of reactant (*e.g.*, xylitol, XYL) and triflic acid catalyst (TfOH), in a solvent system of desired composition (*e.g.*, 90 wt% GVL in water) were filtered using a 0.2 μm membrane (VWR International; PTFE) prior to being added into closed batched reactors. Owing to the limited solubility of some reactants in organic solvents (*e.g.*, XYL), the initial filtration step ensured that only fully solubilized materials were present under reaction conditions, eliminating the possibility of transport effects confounding the kinetic analyses. Concentrations of TfOH in each experiment were varied to maintain consistent kinetic profiles in all solvent systems with respect to reaction time. For example, the concentration of TfOH used for XYL dehydration in pure water was 1.3 M. In contrast, the concentration of TfOH used for the same reaction in 90 wt% GVL/water was 0.037 M. Reactors were placed in an oil bath at the desired temperature and stirred at 500 rpm using magnetic stir bars. The reactors were allowed to

equilibrate to the desired reaction temperature for a period of ~15 minutes, after which the reactors were removed at intervals corresponding to the desired reaction times, and quenched in an ice bath at 273 K.

After quenching the reactors, the contents were diluted 9:1 by mass with HPLC grade water, neutralized with sodium bicarbonate to a pH of 7.0, and stirred vigorously for 30 seconds. The neutralized solution was then filtered using a 0.2 μm membrane (VWR International; PTFE) prior to analysis. Reaction products for ethyl-*tert*-butyl ether (ETBE) hydrolysis in water and in GVL/water mixtures were analyzed using a gas chromatograph (Shimadzu GC-2010) equipped with a flame ionization detector. All other analyses were performed using a high-performance liquid chromatograph (Shimadzu LC-20AD) equipped with a differential refractometer (Shimadzu RID-20A) and a photodiode array (Shimadzu SPD-M20A). Concentrations of all analytes were quantified based on calibration curves using external standards. Separation of XYL, 1,4-anhydroxylitol, ETBE, ethanol and TBA was achieved using two Bio-Rad Aminex HPX-87P HPLC ion-exclusion columns in series (7.8 x 300 mm, 7μm). A mobile phase of HPLC grade water at a flow rate of 0.6 mL min^{-1} was used. Separation of cellobiose, glucose, and levoglucosan was achieved with an ion-exclusion HPLC column (Bio-Rad; Aminex HPX-87H; 7.8 x 300 mm, 5 μm). A mobile phase of 5 mM sulfuric acid aqueous solution (Ricca Chemical; HPLC grade) at a flow rate of 0.6 mL min^{-1} was used. Separation of ETBE, ethanol and TBA in water in GVL/water mixtures was achieved in a gas chromatography column (Shimadzu; SHRXI-5MS; 30 m, 0.25mm ID, 0.25 mm DF) to establish carbon balances. To minimize exposure of the GC column to mineral acids and other ions, all other reactions of ETBE were monitored using the HPX-87P HPLC system, and the reaction was assumed to be quantitative in these cases as well.

S1.3 Reaction kinetics modeling

Rate constants were estimated in MATLAB (nlinfit function; Levenberg-Marquardt nonlinear least squares algorithm) by optimizing their values to minimize the residuals between experimental and model-predicted time courses in each experiment. Model-predictive expressions in the MATLAB routine took the form of rate laws based on transition state theory, formulated as described below.

We assume that the acid-catalyzed reactions considered in this study occur *via* a sequence of elementary steps, one of which is rate limiting and passing through a protonated transition state. Under these assumptions,[2] the forward rate of each reaction per unit volume is equal to the concentration of the activated complex (or the transition state) in the rate-limiting step (C_{TS}), multiplied by the frequency of vibration along the reaction coordinate (ν_{RC}):

$$r = \nu_{RC} C_{TS} \tag{S2}$$

Transition state theory treats the activated complex as being in equilibrium with the solvated reactants and proton.[3] Accordingly, the concentration of the activated complex can be written in terms of the activities of the free proton (a_{H+}), and the reactants and/or products ($a_{R/P}$) that are consumed or produced in the elementary steps prior to the rate-limiting step:

$$r = \frac{k_B T}{h} K \frac{a_{H^+} \prod a_{R/P}^n}{\gamma_\ddagger} \tag{S3}$$

where k_B is the Boltzmann constant, T is the temperature, h is Planck's constant, and $\gamma_\ddagger$ is the activity coefficient of the activated complex. K is the product of all equilibrium constants for the elementary steps prior to the rate limiting step, and n denotes the stoichiometry of the species that participate in those elementary steps (being positive for reactants and negative for products). The leading rational term in Equation S3 is the frequency factor that results from standard statistical mechanical treatments of the vibrational mode along the reaction coordinate.[4] While the reactant is consumed over the course of the reaction, it remains approximately constant and dilute (~ 0.1 M) throughout the experiments conducted in this study. The activity coefficients of all species represented in Equation S3 may thus be approximated as constants, meaning that the rate expression may be recast as a function of molar concentrations:

$$r = k_{org,j}^i C_{H^+} \prod C_{R/P}^n \tag{S4}$$

where $k_{org,j}^i$ is the apparent rate constant for the ith reaction containing all invariant terms (including activity coefficients) in Equation S3, and the subscript denotes the identity and composition (in j^{th} mass percent) of the organic phase. The rate expression for each reaction in the present study thus takes the form of Equation S4, with as many $C_{R/P}$ terms as there are elementary steps involving the consumption of a reactant, or generation of a product prior to the rate-limiting step.

<u>For acid-catalyzed dehydration reactions</u>

We have shown that the removal of water in the carbenium ion formation step is rate-limiting for reactions of TBA and PDO.[1] It has been shown elsewhere that the first water elimination step is rate-limiting in the acid-catalyzed dehydration of FRU,[5, 6] and XYL dehydration involves only a single water removal.[7] The rate expressions for these four dehydration reactions are thus:

$$r_{dehydration}^{XYL} = k_{org,j}^{XYL} C_{XYL} C_{H^+} \tag{S5}$$

$$r_{dehydration}^{FRU} = k_{org,j}^{FRU} C_{FRU} C_{H^+} \tag{S6}$$

$$r_{dehydration}^{PDO} = k_{org,j}^{PDO} C_{PDO} C_{H^+} \tag{S7}$$

$$r_{dehydration}^{TBA} = k_{org,j}^{TBA} C_{TBA} C_{H^+} \tag{S8}$$

<u>For acid-catalyzed hydrolysis reactions</u>

It is generally accepted [8, 9] that the addition of water in Brønsted-acid-catalyzed hydrolysis of ethers is fast compared to the formation of the carbocation that follows removal of the leaving group. As such, the concentration of water should not appear in the forward rate expression for Brønsted-acid catalyzed hydrolysis reactions. Accordingly, the forward rate expressions for each of the hydrolysis reactions in this study are:

$$r_{hydrolysis}^{CEL} = k_{org,j}^{CEL} C_{CEL} C_{H^+} \tag{S9}$$

$$r_{hydrolysis}^{ETBE} = k_{org,j}^{ETBE} C_{ETBE} C_{H^+} \tag{S10}$$

$$r_{hydrolysis}^{LGA} = k_{org,j}^{LGA} C_{LGA} C_{H^+} \tag{S11}$$

<u>Regarding the reversibility of the LGA hydrolysis reaction</u>

In solvent systems containing less than 50 wt% water, glucose undergoes dehydration to afford LGA. LGA hydrolysis (Equation S11) is thus an equilibrium-limited reaction in such solvent systems, and the reverse rate of this reaction must be accounted for in the kinetic modeling scheme. Accordingly, the forward rate of LGA hydration is measured by accounting for the reaction affinity (or the extent of the system's departure from its equilibrium state) in the rate expression.[10] We thus consider the overall rate of the LGA hydrolysis reaction as the difference in the rates of LGA hydrolysis and glucose (GLU) dehydration:

$$r_{hydrolysis}^{LGA} - r_{dehydration}^{GLU} = k_{hydrolysis}^{LGA} C_{LGA} C_{H^+} - k_{dehydration}^{GLU} \frac{C_{GLU} C_{H^+}}{C_{H_2O}} \tag{S12}$$

where again, the apparent rate constants contain the activity coefficients for the kinetically relevant species, which are approximated as constant. The equilibrium constant for this reversible reaction is:

$$K_{eq} = \frac{a_{GLU}}{a_{LGA} a_{H_2O}} = \frac{C_{GLU}^o}{C_{LGA}^o C_{H_2O}^o} \frac{\gamma_{GLU}}{\gamma_{LGA} \gamma_{H_2O}} \tag{S13}$$

where C_i^o is the concentration of the ith species at equilibrium. Even in solvent systems consisting of 90 wt% of the organic phase, the concentration of water is in significant excess compared to LGA (~5.6 M compared to ~0.1 M, respectively). As such, C_{H_2O} is approximately equal to $C_{H_2O}^o$, and combining Equations S12 and S13 affords:

$$r_{hydrolysis}^{LGA} - r_{dehydration}^{GLU} = k_{hydrolysis}^{LGA} C_{LGA} C_{H^+} - k_{dehydration}^{GLU} K_{eq} \frac{C_{LGA}^o}{C_{GLU}^o} C_{GLU} C_{H^+} \tag{S14}$$

Moreover, by definition [11]:

$$K_{eq} = \frac{k_{hydrolysis}^{LGA}}{k_{dehydration}^{GLU}} \tag{S15}$$

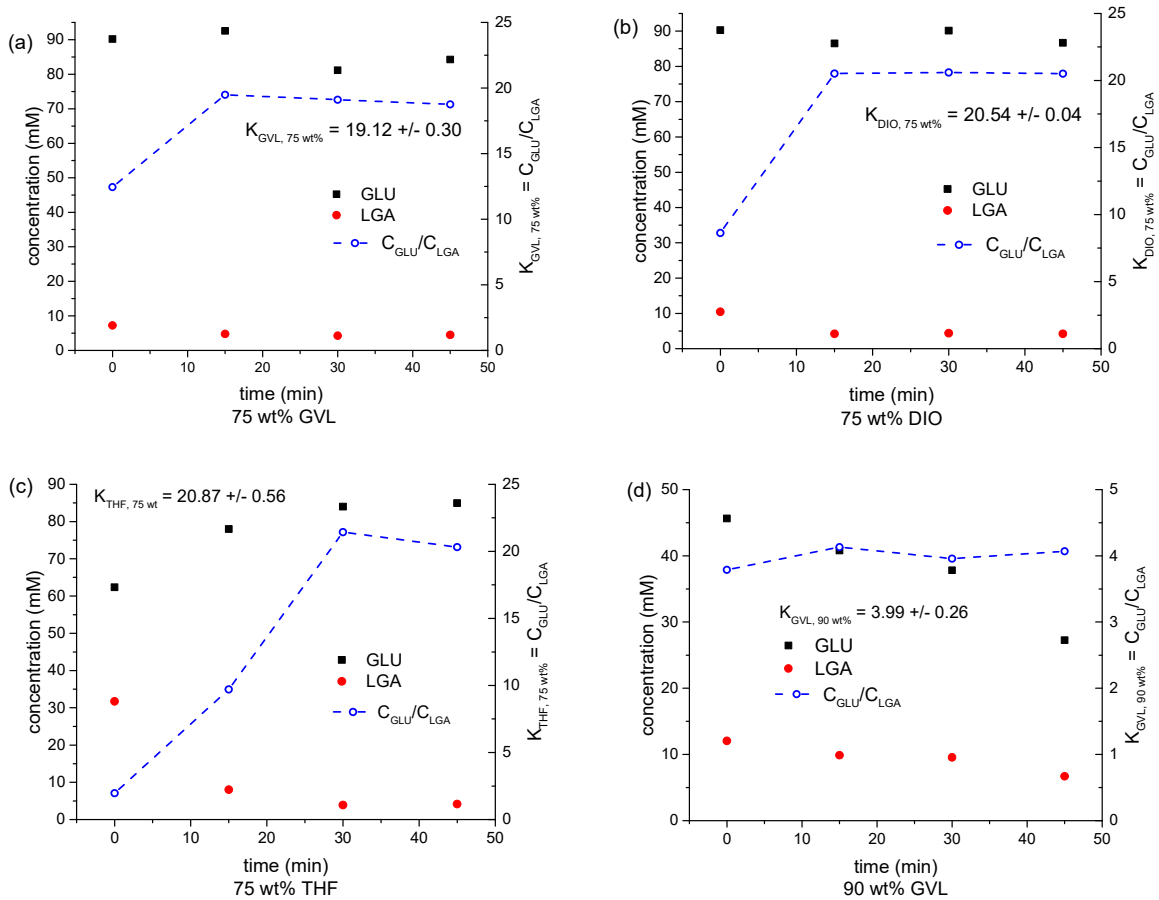
So that, finally, Equation (S11) can be written as:

$$r_{hydrolysis}^{LGA} - r_{dehydration}^{GLU} = -\frac{dC_{LGA}}{dt} = k_{org,j}^{LGA} C_{LGA} \left(1 - \frac{1}{K_{org,j}} \frac{C_{GLU}}{C_{LGA}}\right) C_{H^+} \tag{S16}$$

where $K_{org,j}$ is the *equilibrium ratio of the concentrations of glucose and LGA:*

$$K_{org,j} = \frac{C^o_{GLU}}{C^o_{LGA}} \tag{S17}$$

The value of $K_{org,j}$ was measured independently of the kinetic experiments in this study by allowing a 0.1 M solution of LGA to equilibrate, in the presence of a triflic acid catalyst, in each of the six solvent systems where the rate of the glucose dehydration reaction competes with the forward rate of LGA hydrolysis. Figure S1 presents the reaction time courses for these six experiments, and captures the relationship between LGA and glucose in terms of the pseudo-equilibrium constant, $K_{org,j}$ at 403 K.



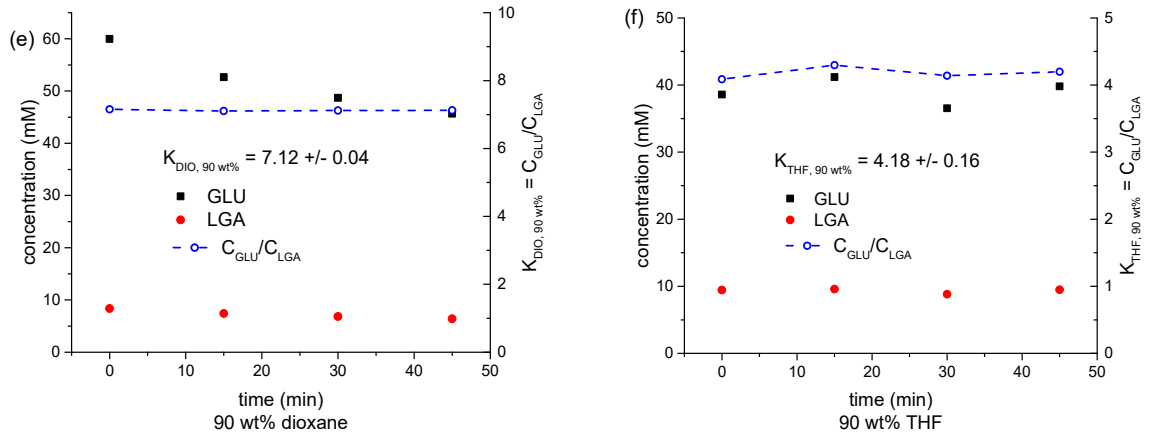
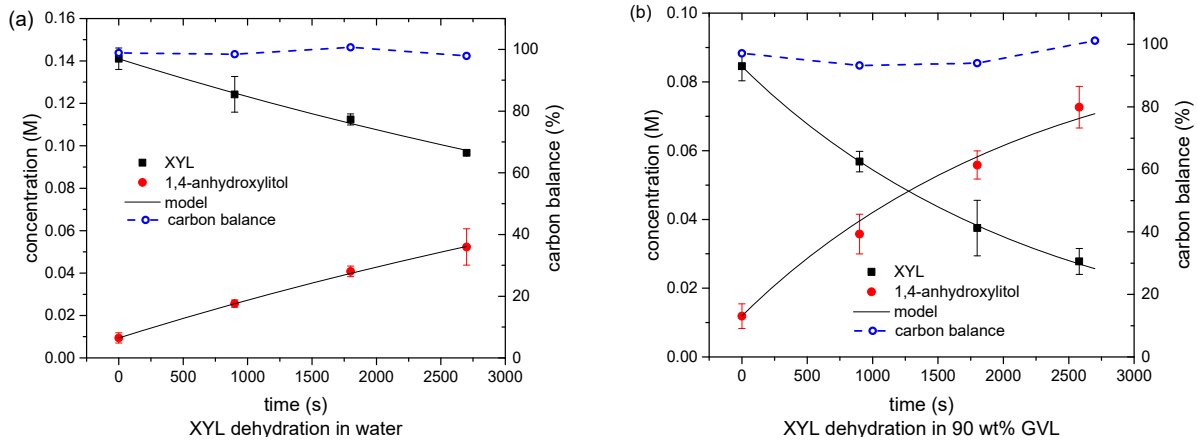


Figure S1. Time courses for LGA hydrolysis to afford GLU in the presence of a triflic acid catalyst to demonstrate the equilibrium relationship between reactant and product as a function of solvent system. *Reaction conditions: 0.3 – 0.15 M TfOH; 403 K. Error in the apparent equilibrium ratio of glucose to LGA represents the 95% confidence interval, based on the variability in the two to four data points in each graph after the two species have equilibrated.*

Section S2: Experimental Results

Having measured the values of $K_{org,j}$, experimental reaction time courses for LGA hydrolysis were fit to Equation S17 in the MATLAB routine, yielding a set of values for the apparent forward rate constant ($k_{org.j}^{LGA}$) in each solvent system. In a similar fashion, Equations S5-S10 were fit to experimental time courses to afford the apparent rate constants for the Brønsted-acid-catalyzed reactions for XYL, CEL and ETBE as a function of solvent composition. Figure S2 contains exemplary reaction time courses for each reaction in pure water, and in 90 wt% GVL, along with the associated model fits and carbon balances.



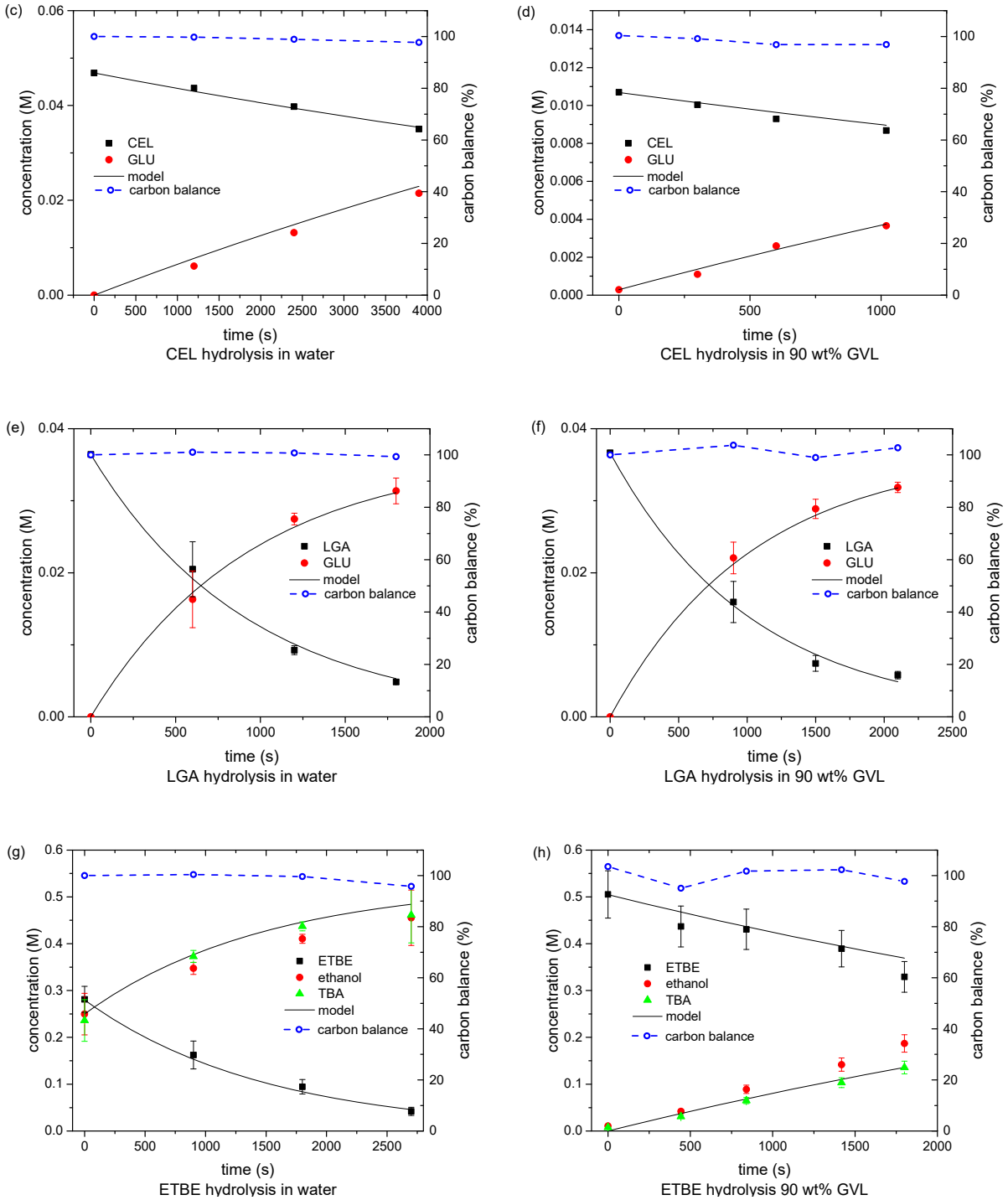


Figure S2. Exemplary experimental and model-predicted time courses for reactions of XYL, CB, LGA, and ETBE in water, and in aqueous mixtures of 90 wt% GVL. *Reaction conditions: see Table S2. Error bars represent 95% confidence intervals on the data based on triplicate experiments.*

Table S2 presents the apparent rate constants measured in this study, along with the reaction conditions for each experiment. Selectivities differing from unity generally reflect uncertainties in quantification of reactants and products, which is captured in the 95% confidence intervals reported along with each rate constant. In calculating the 95% confidence intervals, we account for the propagation of errors resulting from two sources of uncertainty: uncertainty in the quantification of reactants and products (based on triplicate experiments) and the residuals between experimental and model-predicted time courses (MATLAB; nlparci function).

Table S2. Apparent rate constants as a function of solvent system for the four acid-catalyzed reactions examined in this study. *Reaction conditions: 2 mL of solution in closed, 10 mL thick-walled glass reactors stirred at 500 rpm.*

XYL dehydration to afford 1,4-anhydroxylitol

Solvent	Wt% organic phase	triflic acid (M)	Temperature (K)	Conversion (mole %)	Selectivity (mole %)	k_{app} (L mol^{-1} s^{-1})	+/- (95% confidence interval)
Water	--	1.300	403	33%	98%	1.04e-4	1.38e-5
GVL	25%	0.987	403	25%	99%	1.34e-4	2.60e-5
GVL	50%	0.650	403	35%	100%	2.67e-4	4.06e-5
GVL	75%	0.338	403	50%	95%	1.10e-3	1.47e-4
GVL	90%	0.037	403	70%	94%	1.18e-2	1.22e-3
DIO	25%	0.981	403	35%	96%	1.59e-4	2.16e-5
DIO	50%	0.659	403	43%	100%	3.26e-4	4.59e-5
DIO	75%	0.333	403	65%	94%	1.10e-3	1.47e-5
DIO	90%	0.033	403	45%	95%	6.60e-3	7.60e-4
THF	25%	1.022	403	40%	106%	1.80e-4	3.65e-5
THF	75%	0.323	403	37%	110%	5.71e-4	1.51e-4
THF	90%	0.016	403	30%	97%	7.40e-3	7.41e-4

CEL hydrolysis to afford GLU

Solvent	Wt% organic phase	triflic acid (M)	Temperature (K)	Conversion (mole %)	Selectivity (mole %)	k_{app} (L mol^{-1} s^{-1})	+/- (95% confidence interval)
Water	--	0.005	403	25%	90%	0.014	0.002
GVL	25%	0.005	403	43%	90%	0.017	0.002
GVL	50%	0.005	403	32%	86%	0.024	0.003
GVL	75%	0.003	403	30%	85%	0.080	0.011
GVL	90%	0.003	403	21%	80%	0.577	0.084
DIO	25%	0.008	403	28%	78%	0.016	0.003
DIO	50%	0.007	403	30%	85%	0.023	0.004
DIO	75%	0.006	403	45%	91%	0.099	0.014
DIO	90%	0.001	403	27%	77%	0.205	0.031
THF	25%	0.005	403	78%	96%	0.017	0.003
THF	75%	0.049	403	55%	84%	0.058	0.006
THF	90%	0.001	403	42%	89%	0.145	0.021

LGA hydration to afford GLU

Solvent	Wt% organic phase	triflic acid (M)	Temperature (K)	Conversion (mole %)	Selectivity (mole %)	k_{app} (L mol^{-1} s^{-1})	+/- (95% confidence interval)
Water	--	0.048	403	86%	99%	0.023	0.0039
GVL	25%	0.067	403	95%	95%	0.025	0.0039
GVL	50%	0.038	403	79%	82%	0.035	0.0040
GVL	75%	0.010	403	38%	92%	0.075	0.0086
GVL	90%	0.005	403	81%	103%	0.174	0.0303
DIO	25%	0.036	403	92%	91%	0.024	0.0040
DIO	50%	0.050	403	89%	99%	0.023	0.0041
DIO	75%	0.013	403	53%	104%	0.048	0.0049
DIO	90%	0.009	403	52%	104%	0.073	0.0075
THF	25%	0.068	403	96%	100%	0.024	0.0027
THF	75%	0.013	403	60%	98%	0.034	0.0054
THF	90%	0.005	403	52%	96%	0.080	0.0105

ETBE hydrolysis to afford TBA and ethanol

Solvent	Wt% organic phase	triflic acid (M)	Temperature (K)	Conversion (mole %)	Selectivity (mole %)	k_{app} (L mol^{-1} s^{-1})	+/- (95% confidence interval)
Water	--	0.951	343	85%	93%	6.70e-4	6.20e-5
GVL	25%	0.406	343	24%	104%	6.08e-4	6.10e-5
GVL	50%	0.660	343	50%	92%	4.10e-4	1.74e-5
GVL	75%	0.495	343	29%	103%	2.89e-4	5.23e-5
GVL	90%	0.142	343	23%	98%	1.20e-3	1.90e-4
DIO	25%	1.020	343	48%	-- [a]	2.70e-4	2.10e-5
DIO	50%	0.703	343	60%	-- [a]	4.62e-4	3.00e-5
DIO	75%	0.351	343	22%	-- [a]	2.69e-4	3.40e-5
DIO	90%	0.155	343	12%	-- [a]	2.31e-4	2.50e-5
THF	25%	1.02	343	22%	-- [a]	9.10e-5	9.00e-6
THF	50%	0.661	343	21%	-- [a]	1.38e-4	1.50e-5
THF	75%	0.321	343	14%	-- [a]	1.62e-4	4.90e-5
THF	90%	0.153	343	9%	-- [a]	2.59e-4	2.50e-5

a) *Due to experimental limitations (see text above), selectivities for ETBE hydrolysis were not quantified in these solvent systems. Product selectivities in GVL/water mixtures were assumed to be representative of these cases. Accordingly, rate constants for ETBE hydrolysis in these solvent systems were estimated based on the rate of ETBE disappearance only.*

S2.1 Effects of temperature and reactor headspace on the apparent rate constants

Control experiments were conducted to assess the sensitivity of the kinetic measurements to reaction temperature. Figure S3 demonstrates the weak temperature dependence of the kinetic solvent parameters associated with XYL and PDO dehydration in GVL/water mixtures, indicating that these kinetic descriptors may be compared between reactions that take place at different temperatures.

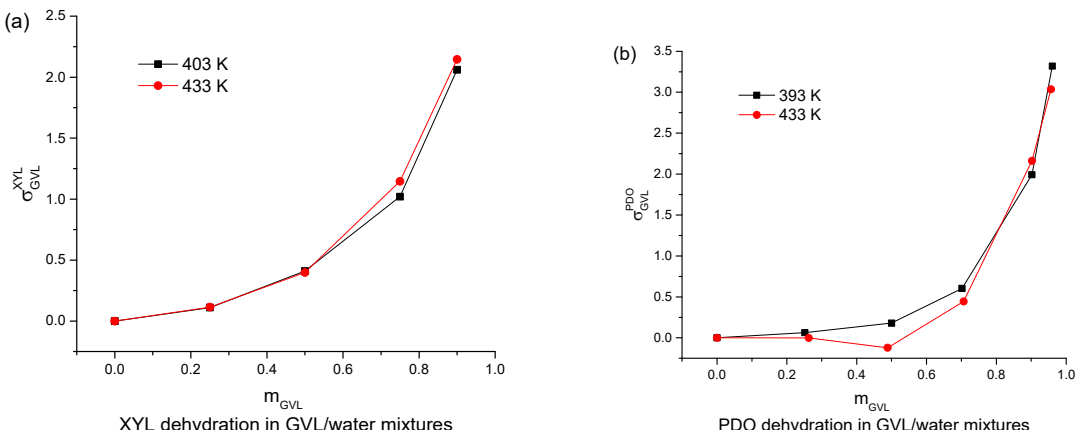


Figure S3. Effect of temperature on the apparent rate constants for XYL and PDO dehydration in GVL/water mixtures, as expressed by their kinetic solvent parameters (σ_{GVL}^{i}).

We conducted control experiments to probe the sensitivity of the kinetic measurements to differences in the partitioning of water and the organic cosolvent between the liquid and gas phases. Table S3 demonstrates that even when the most volatile cosolvent (THF) is used, the apparent rate constant for XYL dehydration is essentially invariant with respect to reactor headspace. As such, differences in volatility between the three cosolvents results do not lead to significantly different concentrations of water in the liquid phase under reaction conditions.

Table S3. Reactor head space versus apparent rate constants for XYL dehydration in 90 wt% THF and water. *Reaction conditions: 0.50 TfOH; 500 rpm; 403 K.*

Solution volume (mL)	Reactor head space (mL)	$k_{THF,90wt\%}^{XYL}$ (L mol^{-1} s^{-1})
2.0	8.0	0.00210
4.0	6.0	0.00208
8.0	2.0	0.00204

Section S3: Molecular Dynamics Simulation Details

Unbiased molecular dynamics (MD) simulations were performed using the CHARMM36 all-atom force field [12] with the TIP3P water model. Molecular parameters not included within CHARMM36 were generated using the CGenFF force field,[13, 14] which is fully compatible with CHARMM36. Molecular dynamics was performed using a leapfrog integrator with a 2 femtosecond time step. Verlet lists were generated using a 1.2 nm neighbor list cutoff. Van der Waals interactions were modeled with a Lennard-Jones potential using a 1.2 nm cutoff that was smoothly shifted to zero between 1.0 nm and 1.2 nm. Electrostatic interactions were calculated using the Smooth Particle Mesh Ewald method with a short-range cutoff of 1.2 nm, grid spacing of 0.12 nm, and 4th order interpolation. Bonds were constrained using the LINCS algorithm. All simulations were carried out in the *NPT* ensemble using Gromacs 2016 in a cubic simulation cell with periodic boundaries conditions in all directions [15]. All thermostats used a 1.0 ps time constant and all barostats used a 5.0 ps time constant with an isothermal compressibility of 5.0×10^{-5} bar^{-1} (specific thermostats/barostats are detailed in Section S3.1). Simulation configurations were output every 10 ps and the final 190 ns of each production trajectory were used for analysis.

S3.1 System preparation

Each reactant/cosolvent system was prepared using the workflow schematically illustrated in Figure S4. First, water and cosolvent molecules were added to a cubic simulation box with 6 nm box vectors at the desired mass fraction. The method used to calculate the number of cosolvent and water molecules in each system is described in Section S3.2. The solvent mixture was equilibrated for 5 ns at T = 300 K and P = 1 bar with a velocity-rescale thermostat and Berendsen barostat. The reactant was then added to the system and the system was equilibrated for 500 ps at the reaction temperature (see Scheme 1 in main paper) at 1 bar using the same thermostat and barostat. The system was then simulated for 200 ns of production at the same temperature and pressure using the Nose-Hoover thermostat and Parrinello-Rahman barostat.

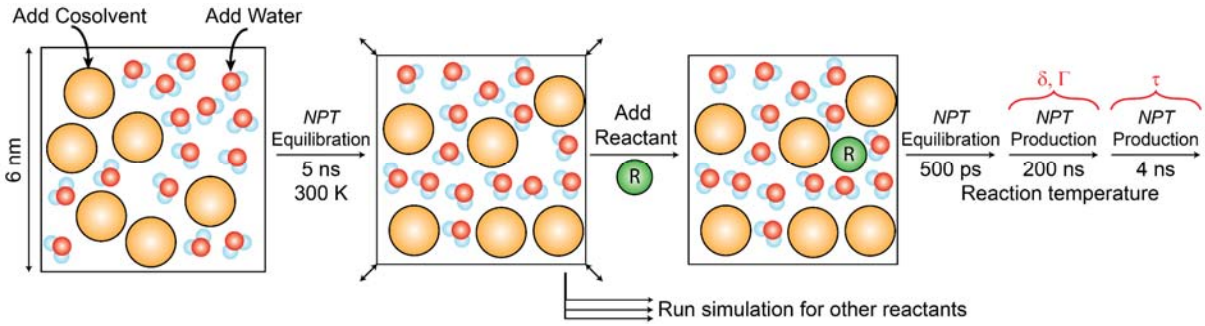


Figure S4. Schematic depiction of mixed-solvent system preparation. R denotes the reactant. Note that the second production trajectory, used to calculate hydrogen bonding lifetimes, was not always 4 ns; some reactants required a longer simulation time to obtain accurate hydrogen-bonding lifetime data.

S3.2 Calculation of number of molecules within system

The number of water and cosolvent molecules required for each solvent mixture was determined based on the total volume of the simulation box. The molecular volume of each cosolvent species was estimated by calculating the volume occupied by a single molecule in a cubic box, then

multiplying the volume by two (1.5 for water) to ensure the molecules appropriately fit when added to the system. The estimated molecular volumes for water, DIO, GVL, and THF were 0.09 nm^3, 0.24 nm^3, 0.30 nm^3, and 0.16 nm^3, respectively. The molecular weights for DIO, GVL, and THF were 88.11, 100.067, and 72.11 g/mol, respectively, which were used to convert weight fractions to mole fractions. To find the number of water and cosolvent molecules to add to the system given the mixed cosolvent composition, we performed a volume balance to obtain:

$$N_1 = \frac{V}{\bar{V_1} + \left(\frac{1 - y_1}{y_1}\right)\bar{V_2}} \tag{S18}$$

$$N_2 = \frac{N_1(1 - y_1)}{y_1} \tag{S19}$$

where N_1 and N_2 are the number of molecules of component 1 (water) and 2 (cosolvent), $\bar{V_1}$ and $\bar{V_2}$ are molecular volumes of component 1 and 2, y_1 is the mole fraction of component 1, and V is the total volume.

S3.3 Calculation of accessible hydroxyl fraction (δ)

The accessible surface area (ASA) of each reactant was calculated using the Shrake and Rupley algorithm [16] as implemented in the MDTraj package. The default 0.14 nm probe radius, which is the van der Waals radius of water, and 960 points were used. The van der Waals radii of all atoms were adjusted according to the values from Bondi.[17] The final 190 ns of production data from trajectories containing reactant in pure water were used to compute the hydroxyl fraction (see Figure S4).

S3.4 Calculation of radial distribution function

The reactant-water radial distribution function (RDF) was calculated using the "compute_rdf" function available in the MDTraj package using a bin width of 0.02 nm. The center of mass of each molecule was used to compute intermolecular distances. Periodic boundary conditions were accounted for in all calculations.

S3.5 Estimation of cutoff between local and bulk domains

The cutoff radius between local and bulk solvent domains, $r_{cutoff,}$ was defined for each system as the distance from the reactant where the radial distribution function between the reactant and water molecules reaches unity. These cutoffs were determined by calculating the running average of the radial distribution function starting from r=2 nm, which is a distance large enough for all RDFs to plateau at unity (*i.e.*, bulk behavior). We define r_{cutoff} as the radius for which the difference between the running average and the next point (*i.e.*, r_{i-1}) is greater than 0.015.

S3.6 Robustness of the preferential exclusion coefficient (Γ)

The preferential exclusion coefficient, Γ, is plotted in Figure S5 as a function of the cutoff radius between the local and bulk solvent domains for XYL and TBA in 90% and 50% wt%DIO/water mixtures. For comparison, the cutoff radius for XYL in 90% and 50% DIO/water mixtures is calculated as 1.51 and 0.83 nm, respectively, using the approach in Section S3.5, while the cutoff radius for TBA in 90% and 50% DIO/water mixture is 1.67 nm and 1.37 nm, respectively. The

plot illustrates that the Γ is sensitive to the cutoff radius, but for each system the Γ plateaus within the error bars by the value of the cutoff radius used in the main manuscript. In Figure S5(b), Γ for TBA in various 90 wt% organic cosolvents are plotted to show that Γ converges within the error bars for different cosolvent systems. The cutoff radius for TBA in 90 wt.% in DIO, GVL, and THF is 1.67, 1.57, and 1.73 nm, respectively.

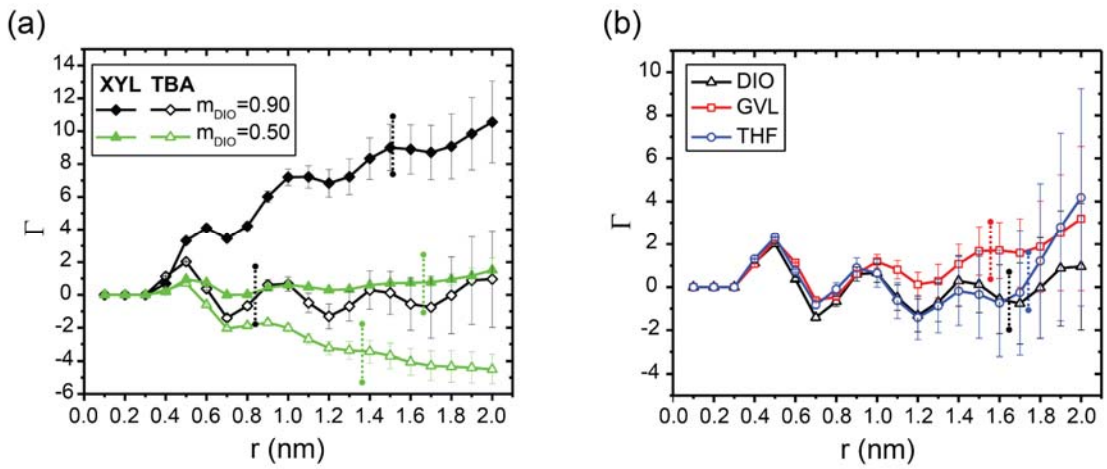


Figure S5. Γ as a function of cutoff radius for (a) XYL and TBA in 90% and 50% DIO/water mixtures. (b) Γ for TBA in different 90% organic cosolvent mixtures is plotted to show convergence across cosolvents. The dashed vertical lines indicate the cutoff radius used for the calculation of the corresponding Γ in the main manuscript.

Figure S6 presents Γ as a function of the amount of simulation time used to calculate its value (sampling time) and the initial size of the simulation box. The dependence of Γ on sampling time was calculated by partitioning the production trajectory in increasing increments of 0.5 ns from the end to the beginning of the trajectory and calculating Γ for each partition. We find that 95 ns of simulation time is sufficient to determine accurate Γ values for both TBA in 50 wt% DIO and XYL in 50 wt% GVL. In addition, we find that varying the initial simulation box size does not significantly change the Γ values. Therefore, we selected 6 nm as the initial box length for all simulations and 95 ns as the simulation block required to calculate reliable Γ values; note that the box length decreases in each trajectory due to the barostat.

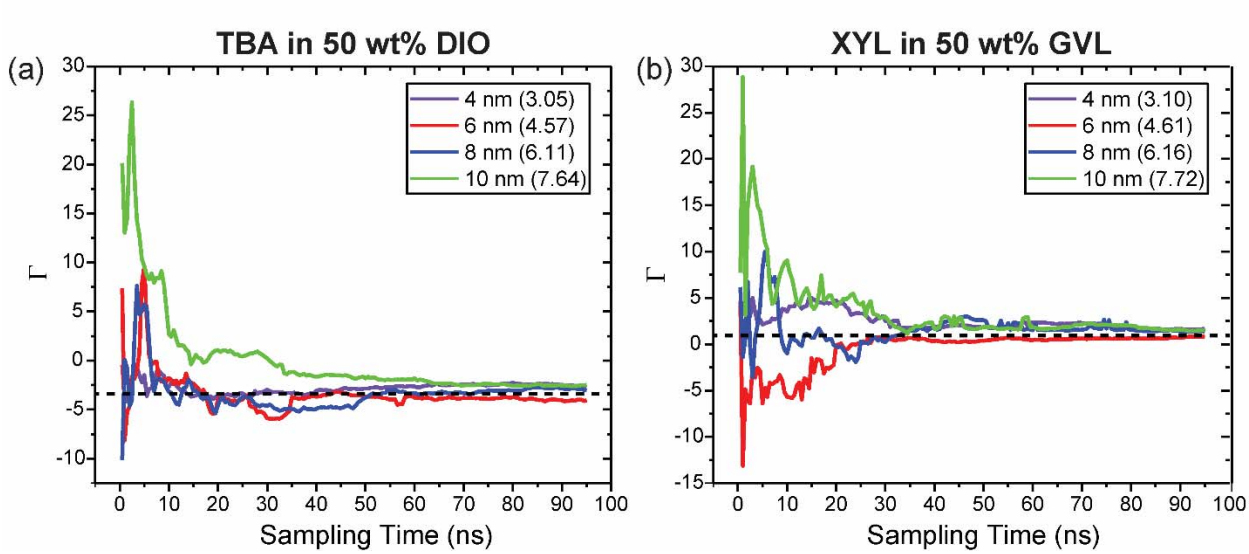


Figure S6. Γ as a function of sampling time for (a) TBA in 50 wt% DIO and (b) XYL in 50 wt% GVL for different initial box lengths. The sampling time is presented in increments of 0.5 ns. The values within the parentheses in each legend are the final box lengths after 95 ns of sampling due to the action of the barostat. The dashed line represents the value of Γ reported in the main manuscript (which is averaged between two independent 95 ns intervals).

S3.7 Hydrogen bonding lifetimes (τ)

After the 200-ns production trajectory, an additional 4 ns production trajectory was generated with configurations output every 0.1 ps to obtain hydrogen bonding lifetimes. Hydrogen bonding lifetimes were computed by splitting the 4 ns trajectory into two 2 ns trajectories and calculated using the hydrogen bonding lifetime analysis tool in Gromacs 5.0.1 [18-21]. Some systems required a longer production trajectory to obtain reliable hydrogen bonding lifetimes based on the standard deviation between the two trajectories. These systems are listed in Table S4. In particular, it was difficult to obtain accurate hydrogen bonding lifetimes for ETBE because it is only a hydrogen bond acceptor.

Table S4. Hydrogen bonding lifetime production trajectories that required more than 4 ns to obtain accurate hydrogen bonding lifetimes.

Reactant	Cosolvent	m_{org}	Hydrogen Bonding Lifetime Total Production Run (ns)
ETBE	DIO	0.90	30
ETBE	DIO	0.75	30
ETBE	DIO	0.50	30
ETBE	DIO	0.25	40
ETBE	GVL	0.90	30
ETBE	GVL	0.75	30
ETBE	GVL	0.50	80
ETBE	GVL	0.25	30
ETBE	THF	0.90	30
ETBE	THF	0.75	30
ETBE	THF	0.50	40
ETBE	THF	0.25	120*
ETBE	-	0.00	30
TBA	DIO	0.90	8
TBA	DIO	0.75	8
TBA	GVL	0.90	12
TBA	GVL	0.75	8
TBA	THF	0.90	8
LGA	THF	0.90	30

*Full 120 ns was used to calculate hydrogen bonding lifetime

S3.8 Rescaling of multidescriptor correlation model

Coefficients for Equations 7-9 in the main text were calculated using the "fitlm" function in MATLAB 2017a. The values for Γ, τ, and δ were rescaled between 0 and 1 using the following procedure:

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{S20}$$

where x is the original simulated descriptor and $\tilde{x}$ is the rescaled descriptor. This rescaling procedure allows us to compare the coefficients in the multidescriptor correlation model.

S3.9 Simulation size, cutoff radii for local domain, and data for multidescriptor correlation model

Table S5 presents the number of molecules, simulation box size, and the cutoff radius used to define the local solvent domain for each simulated system. Note that the cutoff radii for pure water are presented as a comparison to the cutoff radii in the cosolvent mixtures, but they were not used to calculate any simulation-derived parameters.

Table S5. Simulation size and number of molecules for the 91 systems studied. N_{org} and N_{H2O} are the number of organic cosolvents and water molecules, respectively. V is the average volume of the cubic simulation box. r_{cutoff} is the cutoff radius between the local and bulk solvent domains. Γ, τ, and δ are descriptors used in Equation 9 of the main text.

ETBE (δ=0.000)							
Cosolvent	m_{org}	N_{org}	N_{H2O}	V (nm^3)	r_{cutoff} (nm)	Γ	τ
DIO	0.90	748	403	123.8	1.77	-2.68	5.98
	0.75	558	911	110.5	1.71	-9.06	4.76
	0.50	317	1552	94.3	1.37	-4.51	6.38
	0.25	130	2050	81.8	0.91	-1.32	3.06
GVL	0.90	606	372	115.2	1.65	2.47	3.70
	0.75	461	858	105.6	1.43	-5.80	4.51
	0.50	266	1511	92.3	1.61	-11.34	5.50
	0.25	126	1978	82.2	1.37	-4.37	2.25
THF	0.90	1077	484	171.7	1.87	-6.62	12.43
	0.75	773	1025	142.9	1.93	-59.47	2.49
	0.50	415	1661	111.2	1.93	-67.40	4.33
	0.25	180	2078	89.9	1.87	-37.84	4.69
Pure Water	0.00	0	2400	74.5	0.89	-	-

TBA (δ=0.170)							
Cosolvent	m_{org}	N_{org}	N_{H2O}	V (nm^3)	r_{cutoff} (nm)	Γ	τ
DIO	0.90	748	403	126.5	1.67	-0.83	3.40
	0.75	558	911	113.6	1.65	-6.03	2.47
	0.50	317	1552	95.3	1.37	-3.40	1.23
	0.25	130	2050	84.2	0.93	-0.84	1.03
GVL	0.90	606	372	117.3	1.57	1.71	4.61
	0.75	461	858	107.5	1.33	-1.56	2.59
	0.50	266	1511	93.1	1.43	-4.81	1.39
	0.25	126	1978	84.9	1.35	-3.22	1.20
THF	0.90	1077	484	175.7	1.73	0.10	5.17
	0.75	773	1025	147.3	1.87	-21.64	2.68
	0.50	415	1661	112.8	1.87	-39.04	1.59
	0.25	180	2078	91.5	1.75	-18.53	1.51
Pure Water	0.00	0	2400	76.2	0.85	-	-

LGA (δ=0.412)							
Cosolvent	m_{org}	N_{org}	N_{H2O}	V (nm^3)	r_{cutoff} (nm)	Γ	τ
DIO	0.90	748	403	138.1	1.63	7.03	2.51

	0.75	558	911	118.6	1.35	-0.69	2.74
	0.50	317	1552	100.6	1.31	-1.90	1.69
	0.25	130	2050	87.5	0.91	-0.69	1.19
GVL	0.90	606	372	123.9	1.39	1.49	4.81
	0.75	461	858	113.0	1.15	1.11	1.97
	0.50	266	1511	98.5	1.23	-1.17	1.19
	0.25	126	1978	89.3	0.97	-0.66	1.14
THF	0.90	1077	484	190.6	1.59	14.36	7.71
	0.75	773	1025	159.8	1.57	13.31	1.98
	0.50	415	1661	121.8	1.25	-0.53	1.20
	0.25	180	2078	97.7	1.37	-3.73	1.00
Pure Water	0.00	0	2400	77.5	0.89	-	-

PDO (δ=0.425)

Cosolvent	m_{org}	N_{org}	N_{H2O}	V (nm^3)	r_{cutoff} (nm)	Γ	τ
DIO	0.90	748	403	140.1	1.59	5.52	2.31
	0.75	558	911	124.3	1.05	1.81	2.07
	0.50	317	1552	106.4	0.87	-0.16	1.55
	0.25	130	2050	92.4	0.87	-0.40	1.19
GVL	0.90	606	372	127.2	1.57	2.60	2.30
	0.75	461	858	118.0	1.09	1.28	1.75
	0.50	266	1511	101.9	0.89	-0.03	1.50
	0.25	126	1978	91.4	0.97	-0.58	1.19
THF	0.90	1077	484	206.7	1.99	13.37	2.69
	0.75	773	1025	167.6	1.75	24.42	1.39
	0.50	415	1661	126.2	1.05	1.53	1.23
	0.25	180	2078	100.3	1.23	-1.69	1.19
Pure Water	0.00	0	2400	81.3	0.81	-	-

FRU (δ=0.623)

Cosolvent	m_{org}	N_{org}	N_{H2O}	V (nm^3)	r_{cutoff} (nm)	Γ	τ
DIO	0.90	748	403	128.1	1.65	14.77	3.11
	0.75	558	911	114.7	1.15	2.80	2.28
	0.50	317	1552	97.4	1.27	-0.79	1.75
	0.25	130	2050	84.2	0.87	-0.55	1.36
GVL	0.90	606	372	118.6	1.59	6.08	4.09
	0.75	461	858	109.9	1.51	4.77	2.13
	0.50	266	1511	95.3	0.87	0.08	1.62
	0.25	126	1978	85.1	0.85	-0.33	1.21

THF	0.90	1077	484	178.6	1.83	59.63	2.64
	0.75	773	1025	150.7	1.93	122.45	1.59
	0.50	415	1661	115.0	1.83	36.86	1.01
	0.25	180	2078	92.8	0.89	0.09	1.39
Pure Water	0.00	0	2400	75.9	0.91	-	-

CEL (δ=0.632)							
Cosolvent	m_{org}	N_{org}	N_{H2O}	V (nm^3)	r_{cutoff} (nm)	Γ	τ
DIO	0.90	748	403	133.9	1.65	14.71	2.05
	0.75	558	911	118.9	1.53	7.59	1.73
	0.50	317	1552	101.2	1.03	-0.30	1.58
	0.25	130	2050	87.2	1.03	-0.78	1.21
GVL	0.90	606	372	125.2	1.53	5.19	2.27
	0.75	461	858	113.4	1.31	5.88	1.70
	0.50	266	1511	98.2	0.91	0.39	1.48
	0.25	126	1978	89.3	0.95	-0.46	1.18
THF	0.90	1077	484	190.6	1.87	73.56	1.73
	0.75	773	1025	158.0	1.99	151.78	1.42
	0.50	415	1661	122.9	1.81	36.00	1.12
	0.25	180	2078	96.9	0.91	-0.34	1.16
Pure Water	0.00	0	2400	79.7	0.83	-	-

XYL (δ=0.657)							
Cosolvent	m_{org}	N_{org}	N_{H2O}	V (nm^3)	r_{cutoff} (nm)	Γ	τ
DIO	0.90	748	403	135.9	1.51	9.05	2.88
	0.75	558	911	119.5	1.53	4.68	1.70
	0.50	317	1552	101.0	0.83	0.13	1.39
	0.25	130	2050	87.7	0.95	-0.38	1.28
GVL	0.90	606	372	124.0	1.39	2.89	3.51
	0.75	461	858	113.9	1.17	2.98	2.07
	0.50	266	1511	98.3	1.05	0.94	1.48
	0.25	126	1978	89.2	0.89	-0.20	1.19
THF	0.90	1077	484	188.4	1.87	38.53	2.14
	0.75	773	1025	158.6	1.89	84.05	1.40
	0.50	415	1661	120.3	1.85	39.19	1.22
	0.25	180	2078	97.0	0.79	-0.06	1.09
Pure Water	0.00	0	2400	78.6	0.75	-	-

Section S4: Molecular dynamic simulation additional results

S4.1 Two-descriptor correlation model for DIO

Figure S7 compares values of σ calculated using the two-descriptor correlation model (Equation 8 of the main text) to experimentally determined values for DIO/water mixtures. The two descriptors used in this model are the preferential exclusion parameter, Γ, and hydrogen bonding lifetime ratio, τ. This model is further improved by including the accessible hydroxyl fraction, δ, as a third descriptor, leading to the multidescriptor correlation model described in Equation 9 of the main text.

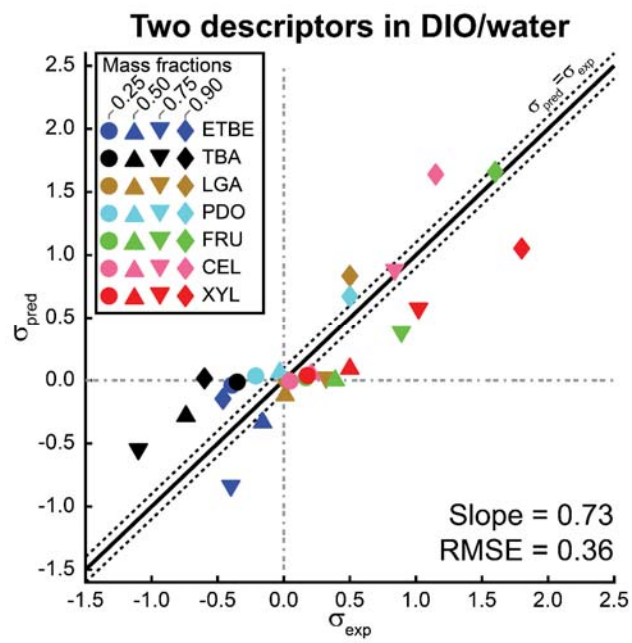


Figure S7. Comparison of kinetic solvent parameters calculated using the two-descriptor correlation model (σ_{pred}) to experimentally determined values (σ_{exp}) for seven reactants in DIO/water mixtures. Each reactant has four data points for 0.25, 0.50, 0.75, and 0.90 mass fractions of the organic phase, with the exception of PDO in a 0.75 mass fraction DIO mixture (see Table S1). The slope of the best-fit line for all the data points and the average root-mean-squared error (RMSE) between the values of σ_{pred} and σ_{exp} are shown at bottom right. The solid black line indicates a perfect correlation ($\sigma_{\text{pred}} = \sigma_{\text{exp}}$) and dotted lines are drawn at $\sigma_{\text{exp}} = 0$ and $\sigma_{\text{pred}} = 0$ to help visualize false positive/negative predicted values. Lines above and below the $\sigma_{\text{pred}} = \sigma_{\text{exp}}$ line are shifted by ±0.10, denoting the approximate experimental error.

S4.2 Multidescriptor correlation model for GVL and THF

Figure S8 compares values of σ calculated using the multidescriptor correlation model (Equation 9 of the main text) to experimentally determined values for (a) GVL/water and (b) THF/water mixtures, respectively. While these correlations are less accurate than the DIO/water systems, most of the predicted values lie in the expected quadrants, indicating few false positives/negatives.

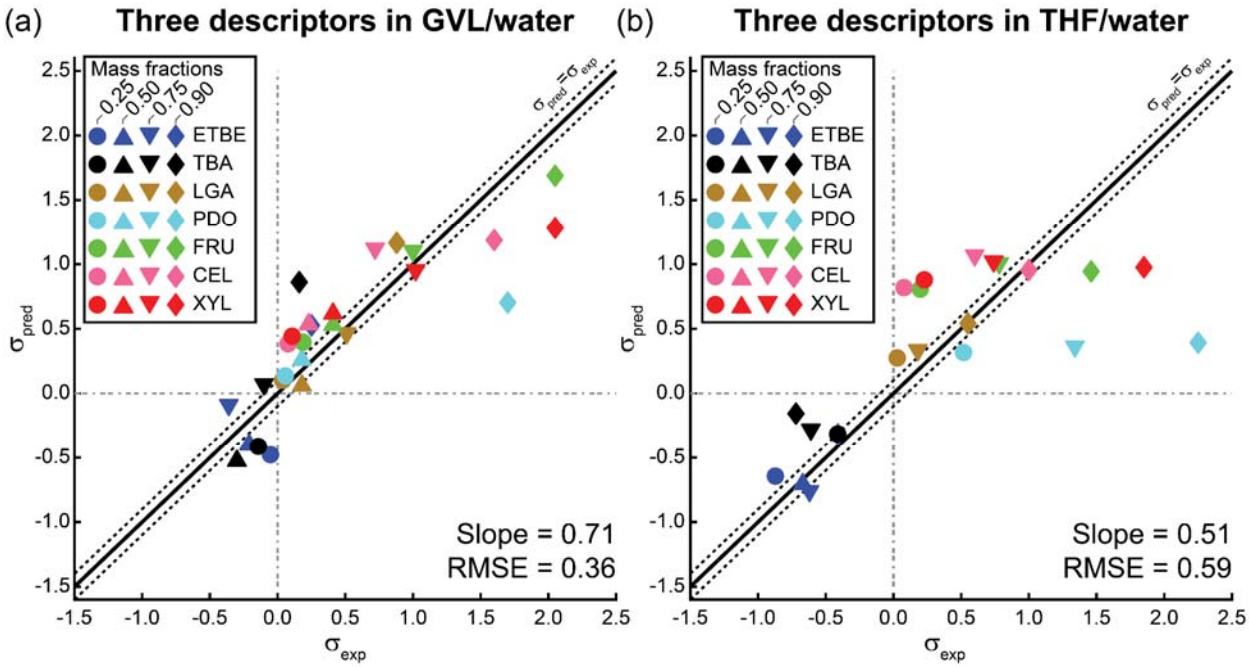


Figure S8. Comparison of kinetic solvent parameters calculated using the multidescriptor correlation model (σ_{pred}) to experimentally determined values (σ_{exp}) for seven reactants in (a) GVL/water mixtures and (b) THF/water mixtures. Each reactant has four data points for 0.25, 0.50, 0.75, and 0.90 mass fractions of the organic phase, with the exception of reactants in 0.50 mass fraction THF mixtures (see Table S2). The slope of the best-fit line for all the data points and the average root-mean-squared error (RMSE) between the values of σ_{pred} and σ_{exp} are shown at bottom right. The solid black line indicates a perfect correlation ($\boldsymbol{\sigma_{pred} = \sigma_{exp}}$) and dotted lines are drawn at $\boldsymbol{\sigma_{exp} = 0}$ and $\boldsymbol{\sigma_{pred} = 0}$ to help visualize false positive/negative predicted values. Lines above and below the $\boldsymbol{\sigma_{pred} = \sigma_{exp}}$ line are shifted by ±0.10, denoting the approximate experimental error.

S4.3 Coefficients of multidescriptor correlation model without re-scaling procedure

Coefficients for the multidescriptor correlation model (Equation 9 of the main text) calculated without rescaling molecular descriptors are shown for all cosolvent systems in

Table S6. Note that the re-scaling procedure (described in Section S3.8) does not change the accuracy of the multiparameter correlation model.

Table S6. Coefficients of the multidescriptor correlation model (Equation 9) calculated without re-scaling molecular descriptors.

$\sigma_{\text{pred}} = A + B(\Gamma) + C(\tau) + D(\delta)$						
Cosolvent	A	B	C	D	Slope	RMSE
DIO	-1.092	0.064	1.892	0.187	0.89	0.23
GVL	-0.579	0.097	1.156	0.234	0.71	0.36
THF	-0.754	0.002	2.421	0.036	0.51	0.59

S4.4 Error in multidescriptor correlation coefficients

The standard error of the best-fit coefficients used in Equation 9 was computed using the "fitlm" function in MATLAB 2017a, shown in Table S7.

Table S7. Standard error in the coefficients for re-scaled and non-rescaled multidescriptor correlation models.

Re-scaled (Table 4)				
Cosolvent	A	B	C	D
DIO	0.159	0.271	0.218	0.249
GVL	0.268	0.489	0.323	0.293
THF	0.384	0.876	0.658	0.756

Not re-scaled (Table S6)				
Cosolvent	A	B	C	D
DIO	0.224	0.011	0.331	0.047
GVL	0.321	0.028	0.491	0.067
THF	0.509	0.004	1.001	0.066

S4.5 Correlation between Γ and τ

The linear correlation between Γ and τ is shown in Figure S9. The correlation is poor across the range of data, suggesting that the two parameters provide complementary contributions to the correlation. Note that since δ is a constant value for each reactant, it is also uncorrelated with Γ and τ.

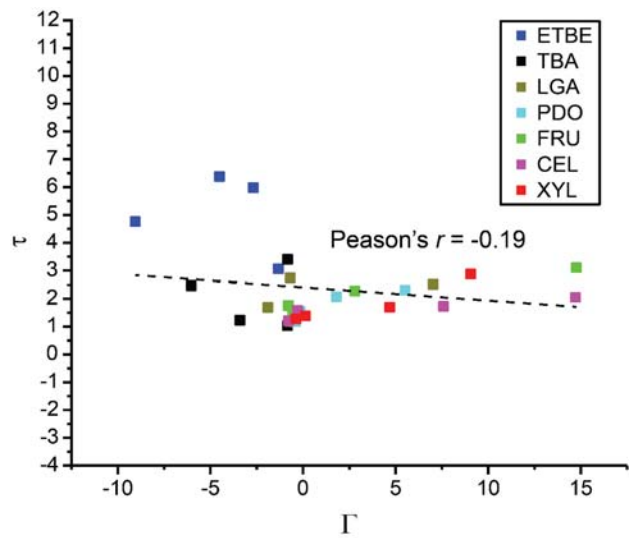


Figure S9. τ versus Γ for all seven substrates in DIO/water mixtures (0 wt%, 25 wt%, 50 wt%, and 90 wt %). The dashed line is the best-fit line between the two variables.

S4.6 Comparison of alternative models using the Akaike Information Criterion

When additional parameters are incorporated into a model-predictive framework, the ability of the model to predict experimental outcomes will improve in most cases. However, one generally assumes that the simplest model that is able to predict realistic behavior is the most physically meaningful.[22] When comparing alternative models, the question thus becomes whether: (1) the improved predictive power afforded by additional parameters is statistically significant compared to the alternative models, or (2) the improvements might be a product of random chance, and the additional parameters introduce an arbitrary level of complexity. Minimum information criteria (MIC) are statistical tests that allow the quality of alternative models to be compared on the basis of their predictive power versus their relative complexity.[23, 24] One popular MIC is the Akaike Information Criterion (AIC), which considers the number of parameters in a model (k), the number of experimental data points being fitted (n), and the maximum value of the model's log-likelihood function ($\ln P$), as shown in Equation (S21).[23]

$$\mathrm{AIC} = 2k - n \ln P \tag{S21}$$

When comparing alternative models, the model with the smallest AIC value is preferred. If the residuals in the model fits are assumed to be normally distributed with zero mean, this gives rise to the special case of least-squares model fitting. In this case, the log-likelihood function is:

$$\ln P = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(Var) - \frac{1}{2Var}RSS \tag{S22}$$

where Var is the variance and RSS is the residual sum of squares. The maximum value is found by differentiating Equation (S21) with respect to the variance, and equating to zero. The result is:

$$\ln P = \frac{RSS}{n} \tag{S23}$$

The AIC then becomes:

$$\mathrm{AIC} = 2k - n \ln\left(\frac{RSS}{n}\right) \tag{S24}$$

For sample sizes where $n/k < 40$, the AIC is corrected as follows:

$$\mathrm{AIC_c} = \mathrm{AIC} + \frac{2k(k+1)}{n-k-1} \tag{S25}$$

Corrected AIC ($\mathrm{AIC_c}$) values for four different correlation models are shown in Table S8. We find that the four-parameter fit (corresponding to the multidescriptor correlation model in Equation 9 of the main text) has the lowest $\mathrm{AIC_c}$ values for DIO and GVL (two out of three cases), indicating that inclusion of τ and δ improves the model to a statistically significant extent.

Table S8. Akaike information criteria with a correction ($\mathrm{AIC_c}$) for finite sample sizes.

		Cosolvent		
Number of parameters	Model	DIO	GVL	THF
4	$\sigma_{\mathrm{pred}} = A + B(\Gamma) + C(\tau) + D(\delta)$	19.1	44.3	55.4
3	$\sigma_{\mathrm{pred}} = A + B(\Gamma) + C(\tau)$	40.2	47.3	58.6
3	$\sigma_{\mathrm{pred}} = A + B(\Gamma) + D(\delta)$	30.6	52.9	52.7
2	$\sigma_{\mathrm{pred}} = A + B(\Gamma)$	37.7	50.8	56.7

S4.7 Assessing the predictive ability of the multidescriptor correlation model

Table S9 shows the results of leave-one-out validation performed by iteratively defining a single reactant as the test set and fitting the parameters of the multidescriptor correlation model (presented in Equation 9 of the main text) using the remaining six reactants as the training set. Validation was performed separately for DIO, GVL, and THF solvent systems. The accuracy of the model was assessed by calculating the RMSE between the predicted and experimentally determine values of the kinetic solvent parameters for the test set reactant for each cosolvent mass fraction. The slope of the best-fit line of the training set and the RMSE of all training set reactants is listed for each test set reactant. For comparison, the slope of the best-fit line and RMSE computed using all seven reactants in the training set (corresponding to the parameters in Table 4 of the main text) are included. These results show that the test set RMSE and best-fit parameters of the multiparameter correlation model are largely robust with respect to the selection of training set, particularly in DIO/water solvent systems, supporting the predictability of the modeling approach.

Table S9. Leave-one-out validation of multidescriptor correlation model.

		Training set		Test set
Cosolvent	Test set reactant	Slope	RMSE	RMSE
DIO	ETBE	0.90	0.21	0.40
	TBA	0.88	0.20	0.35
	LGA	0.89	0.24	0.16
	PDO	0.88	0.24	0.08
	FRU	0.86	0.24	0.12
	CEL	0.89	0.23	0.28
	XYL	0.89	0.20	0.38
	*	0.89	0.23	-
GVL	ETBE	0.71	0.37	0.46
	TBA	0.75	0.35	0.52
	LGA	0.71	0.39	0.20
	PDO	0.76	0.32	0.62
	FRU	0.64	0.38	0.27
	CEL	0.72	0.36	0.39
	XYL	0.71	0.35	0.49
	*	0.71	0.36	-
THF	ETBE	0.31	0.65	0.26
	TBA	0.44	0.62	0.49
	LGA	0.51	0.63	0.19
	PDO	0.74	0.38	1.37
	FRU	0.51	0.61	0.50
	CEL	0.55	0.60	0.64
	XYL	0.51	0.58	0.65
	*	0.51	0.59	-

*All reactants are part of the training set

References

1. M. A. Mellmer, C. Sanpitakseree, B. Demir, P. Bai, K. Ma, M. Neurock, and J. A. Dumesic, *Nature Catalysis*, 2018, **1**, 199-207.
2. M. A. Mellmer, C. Sener, J. M. R. Gallo, J. S. Luterbacher, D. M. Alonso and J. A. Dumesic, *Angewandte Chemie International Edition*, 2014, **53**, 11872-11875.
3. T. H. Lowry and K. S. Richardson, *Mechanism and Theory in Organic Chemistry*, HarperCollins*Pubilishers*, 3rd edn., 1987.
4. K. J. Laider, *Theories of Chemical Reaction Rates*, McGraw-Hill, New York, 1969.
5. T. D. Swift, C. Bagia, V. Choudhary, G. Peklaris, V. Nikolakis and D. G. Vlachos, *ACS Catalysis*, 2013, **4**, 259-267.
6. G. R. Akien, L. Qi and I. T. Horváth, *Chemical Communications*, 2012, **48**, 5850-5852.
7. J. U. Oltmanns, S. Palkovits and R. Palkovits, *Applied Catalysis A: General*, 2013, **456**, 168-173.
8. A. J. Bennet and M. L. Sinnott, *J. Am. Chem. Soc.;(United States)*, 1986, **108**.
9. J. K. Lee, A. D. Bain and P. J. Berti, *Journal of the American Chemical Society*, 2004, **126**, 3769-3776.
10. J. A. Dumesic, *Journal of Catalysis*, 1999, **185**, 496-505.
11. R. de Levie, *The Chemical Educator*, 2002, **7**, 132-135.
12. R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, *J Chem Theory Comput*, 2012, **8**, 3257-3273.
13. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. MacKerell, *Journal of Computational Chemistry*, 2010, **31**, 671-690.
14. W. B. Yu, X. B. He, K. Vanommeslaeghe and A. D. MacKerell, *Journal of Computational Chemistry*, 2012, **33**, 2451-2468.
15. S. Pall, M. J. Abraham, C. Kutzner, B. Hess and E. Lindahl, *Lect Notes Comput Sc*, 2015, **8759**, 3-27.
16. A. Shrake and J. A. Rupley, *Journal of Molecular Biology*, 1973, **79**, 351-371.
17. A. Bondi, *The Journal of Physical Chemistry*, 1964, **68**, 441-451.
18. H. J. C. Berendsen, D. Vanderspoel and R. Vandrunen, *Comput Phys Commun*, 1995, **91**, 43-56.
19. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J Chem Theory Comput*, 2008, **4**, 435-447.
20. E. Lindahl, B. Hess and D. van der Spoel, *J Mol Model*, 2001, **7**, 306-317.
21. D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *Journal of Computational Chemistry*, 2005, **26**, 1701-1718.
22. W. H. Jefferys and J. O. Berger, *American Scientist*, 1992, **80**, 64-72.
23. H. Akaike, *IEEE Transactions on Automatic Control*, 1974, **19**, 716-723.
24. S. Konishi and G. Kitagawa, *Information Criteria and Statistical Modeling*, 2008, 211-237.