Supporting Information For: Closed-Loop Electrolyte Design for Lithium-Mediated Ammonia Synthesis

Dilip Krishnamurthy,[†,¶] Nikifar Lazouski,[‡,¶] Michal L. Gala,[‡] Karthish Manthiram,[‡] and Venkatasubramanian Viswanathan[*,†,§]

†*Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA*

‡*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*

¶*Equally contributed authors*

§*Corresponding Author*

E-mail: venkvis@cmu.edu

Materials

Tetrahydrofuran (THF, 99+%, stabilized with BHT), molecular sieves (3Å, 4-8 mesh), 1-propanol (99+%, extra pure), L-alanine (99%), and 2-methyl-1-propanol (isobutanol, ACS reagent, spectro grade, 99+%) were purchased from Acros Organics. Lithium tetrafluoroborate ($LiBF_4$, 98%), tert-butyl alcohol (99%), 2-butanol (>99%), 2-ethyl-1-butanol (98%), 1-pentanol (ACS reagent, $\geq$99%), 1-hexanol (reagent grade, 98%), 1-heptanol (98%), 1-nonanol (98%), benzyl alcohol (99.8%, anhydrous), phenol (unstabilized, $\geq$99%), 1-phenylethanol (98%), 2-phenylethanol (99%), 2-chloroethanol (99%), 2,2,2-trifluoroethanol (ReagentPlus, $\geq$99%), hexafluoro 2-propanol ($\geq$99%), ethylene glycol (anhydrous, 99.8%), 1,3-butanediol ($\pm$, 99%, anhydrous), glycerol (Reagent-

Plus, ≥99%), triethyleneglycol (ReagentPlus, 99%), 1,5-pentanediol (≥97%), acetic acid (ReagentPlus, ≥99%), hexanoic acid (≥99%), allyl alcohol (99%), 2-methoxyethanol (99.8%, anhydrous), 1-propanethiol (99%), hydrochloric acid (HCl, ACS Reagent, 37%), sodium salicylate (ReagentPlus, ≥99.5%), and sodium hypochlorite (NaOCl, 10-15%) were purchased from Sigma-Aldrich. Methanol (anhydrous, 99.9%), cyclohexanol (99%), 3-butene-1-ol (98+%), sodium nitroprusside (99-102%), and ammonium chloride (NH_4Cl, anhydrous, 99.99%) were purchased from Alfa Aesar. Ethanol (Koptec, anhydrous, 200 proof), 2-propanol (Semi grade, BDH), sodium hydroxide (NaOH, Macron Fine Chemicals, pellet form), and acetone (ACS, BDH Chemical) were purchased from VWR International. 1-butanol (Certified ACS), 3-methyl-1-butanol (isoamyl alcohol, for molecular biology), dichloromethane (DCM, 99.5%), and hexanes (C_6H_14) were purchased from Fisher Scientific. Formic acid (ACS Reagent, 98-100%) and ethanolamine were purchased from EMD Millipore. Milli-Q water was obtained by filtering deionized (DI) water through a Milli-Q purification system (Merck, Millipore Corporation). Platinum foil (Pt, 0.025 mm thick, 99.99%, trace metals basis) and 1-octanol (99%) were purchased from Beantown Chemical. Argon gas (UHP, 5.0 grade) was purchased from Airgas. Nitrogen gas was available in-house; it is generated by boil-off of liquid nitrogen from Airgas. Steel foil (cold-worked 304 stainless steel, 0.002" thick) was purchased from McMaster-Carr. Polyporous Daramic 175 separators were received as a sample from Daramic (Charlotte, NC).

Electrolyte preparation

Dry molecular sieves were prepared by washing as-purchased or previously used molecular sieves with acetone and drying in a muffle furnace at 300 °C for 5 hours. The sieves were added as 20% by volume to as-purchased THF in a round-bottom flask. The flask was sealed from the atmosphere with a rubber septum are dried for at least 96 hours before use.

As purchased $LiBF_4$ was dissolved in dry THF to obtain a 1 M $LiBF_4$ in THF electrolyte solution. The $LiBF_4$ must be sufficiently pure for successful ammonia production; we found that

salt purchased from Sigma-Aldrich is sufficiently pure for these experiments, while other vendors' may require additional purification; one potential purification procedure is given in prior work.[1] The solution was centrifuged at 6000 rpm (4430 rcf) to remove insoluble precipitates. The clear solution was transferred to oven-dried vials, stored in a desiccator, and used within 12 hours of preparation. The solution transfer operations can be performed in the ambient atmosphere; the solutions should not be stored open to the atmosphere, however, as the electrolyte solution can absorb a significant amount of water from ambient air.

The proton donor was added to the electrolyte immediately prior to experiments. The total volume of proton donor-containing electrolyte solution prepared for each experiment is 4 mL. If the volume of proton donor that needs to be added to obtain the desired concentration is <100 μL, then the proton donor was added to 4 mL of electrolyte directly. If the volume required is >100 μL, then the proton donor added to a smaller volume of electrolyte that was rounded to the nearest 0.1 mL, so that the final volume of the proton donor in electrolyte solution would equal 4 mL. For example, to prepare 0.2 M ethanol, 47 μL of ethanol were added to 4 mL of electrolyte solution, while to prepare 0.6 M 1-butanol, 220 μL of 1-butanol were added to 3.8 mL of electrolyte solution.

Nitrogen reduction experiments

Polished stainless steel electrodes were used as the cathode stainless steel shims were cut into 2x2 cm pieces, wet with DI water, and polished with 400 grit followed by 1500 grit sandpaper thoroughly. The polished foils were rinsed thoroughly with DI water and dried in air at 80 °C. Stainless steel cathodes were used in a single experiment before discarding.

Parallel plate cells described in prior work were used to perform nitrogen reduction experiments.[1] Briefly, a polished steel foil was used as the cathode, a platinum foil was used as the anode, a piece of Daramic was used as a separator, and machined polyether ether ketone (PEEK) cell parts were used for the cell body (Fig. S1). All cell parts were dried in air at 80°C for at least

20 minutes prior to use.

Nitrogen (or argon, in control experiments) gas was flowed at 10 standard cubic centimeters per minute (sccm) through a vial containing THF and molecular sieves to saturate the feed gas with THF. The THF-saturated feed gas was then flowed to an assembled 2-compartment cell. The proton-donor electrolyte was added first to the anode compartment, then to the cathode compartment. 1.75 mL of electrolyte was added to each compartment; note that this is the volume added to each compartment, and may not be the final volume in each compartment at the conclusion of the experiment due to solvent evaporation. The feed gas was flowed through the electrolyte for 10 minutes at open circuit to saturate the electrolyte with gas and to strip oxygen from the solution.

After saturating the solution with the feed gas, a constant current of 20 mA was applied for 6 minutes using a Tekpower 5003 DC power source, for a total of 7.2 coulombs of charge passed. In some experiments (see Supporting Information Table 2), the potential required to applied 20 mA exceeded 50 V. In experiments where an excess of 50 V was required to apply 20 mA, a constant potential of 50 V was applied across the cell for 6 minutes, and the total charge passed was quantified by measuring the potential drop across a resistor in series with the cell (Fig. S6). As the electrolyte resistance does not significantly change with concentration of most proton donors, the higher voltage required is likely due to changes in SEIs at the cathode or anode in these experiments.

Following application of current, the catholyte was immediately removed from the cell and diluted in water. In most experiments, the electrolyte was used directly to prepare samples for ammonia quantification. In these cases, the samples were made as follows: one by adding 200 μL of catholyte to 1800 μL of Milli-Q water (10-fold dilution), and another by adding 100 μL of catholyte to 1900 μL of water (20-fold dilution), to be able to accurately quantify ammonia at both lower and higher Faradaic efficiencies. In some cases, the proton donor can affect the colorimetric

assay negatively by either phase separating with water (e.g. octanol), leading to higher spurious absorbances, or chemically (e.g. ethyl acetate, thiols), leading to lower or shifted absorbances (Fig. S3). In these cases, the proton donor was extracted from the ammonia-containing samples. To extract the proton donor, 500 μL of electrolyte were added to 4.5 mL of 0.05 M H_2SO_4 in water. The proton donor in resulting acidified solution was extracted with 3 mL of either DCM or hexanes three times. Milli-Q water was then added to the aqueous phase to a final volume of 5 mL if the volume decreased, which may occur if the THF was extracted into the organic phase. The aqueous phase was centrifuged at 6000 rpm (4430 rcf) for 10 minutes to promote complete phase separation. The aqueous phase was then quenched with base by adding 1500 μL of the acidified solution to 500 μL of 0.4 M NaOH, or by adding 750 μL of the acidified solution to 250 μL of 0.4 M NaOH and 1000 μL of Milli-Q water.

After experiments, Daramic separator pieces were rinsed with acetone and soaked in DI water for at least 10 minutes to remove traces of solvent and ammonia. Cell parts and platinum anodes were rinsed with acetone and washed thoroughly with DI water. All cell parts, electrodes, and separators were dried at 80 °C in air prior to use in further experiments.

Choosing proton carrier concentrations

In order to determine whether a proton donor can be used to produce ammonia using the lithium-mediated approach, a range of proton concentrations had to be efficiently screened for activity. From prior work,[1] it is known that ammonia yields depend on the concentration of ethanol, the proton donor. At low concentrations, no ammonia is formed and a large amount of lithium remains on the cathode, while at high concentrations, no ammonia is formed due to competition from the hydrogen evolution reaction. We posited that this behavior is not unique to ethanol and can be observed for various proton donors. From this hypothesis, we developed a heuristic to rapidly screen the concentration range. Initially, electrolyte containing 0.2 M of the proton donor was

used to test for ammonia production. If a significant amount of lithium metal was found to remain on the cathode or in solution, the concentration of the proton donor was increased for the next run. Typically, the concentration was increased 2- or 3-fold, depending on the extent of lithium coverage on the surface. If the surface is clean and the steel cathode is visible, the concentration of the proton donor was decreased by a similar amount. Every proton donor was tested at three concentrations, with a maximum concentration tested of 1 M. The 1 M cutoff is arbitrary, and was chosen as it is a concentration that results in fairly large volume fractions of proton donor in electrolyte for most proton donors. Using this heuristic, we typically obtained runs with a high lithium coverage after the experiment, a low lithium coverage, and an intermediate coverage, except for cases where 1 M of proton donor did not decrease lithium coverage; in these cases, all experiments had a large lithium coverage. We believe that approach allowed us to probe a large compound-concentration phase space efficiently to determine which compounds are capable of promoting lithium-mediated nitrogen reduction.

Ammonia quantification

The amount of ammonia in samples produced in nitrogen reduction experiments was quantified by using the salicylate assay according to a procedure described in earlier work.[1] Briefly, 280 μL of 1% NaOCl in 0.4 M NaOH solution was added to 2 mL of ammonia sample solution, followed by 280 μL of 2.5 M sodium salicylate, 3.5 mM sodium nitroprusside solution. The resulting solution was mixed vigorously and left to evolve color in the dark for at least 2 hours. The absorbance spectrum of the resulting solution was measured using an Ocean Optics Flame-S UV-Vis spectrophotometer. The relevant signal for ammonia quantification was taken to be the difference in absorbance values at 650 nm and 475 nm to avoid overestimating the amount of ammonia produced.[1]

A fresh ammonia calibration curve was made for each quantification batch. Calibration curves were made by adding 100 μL of electrolyte to solutions of known ammonium sulfate concentra-

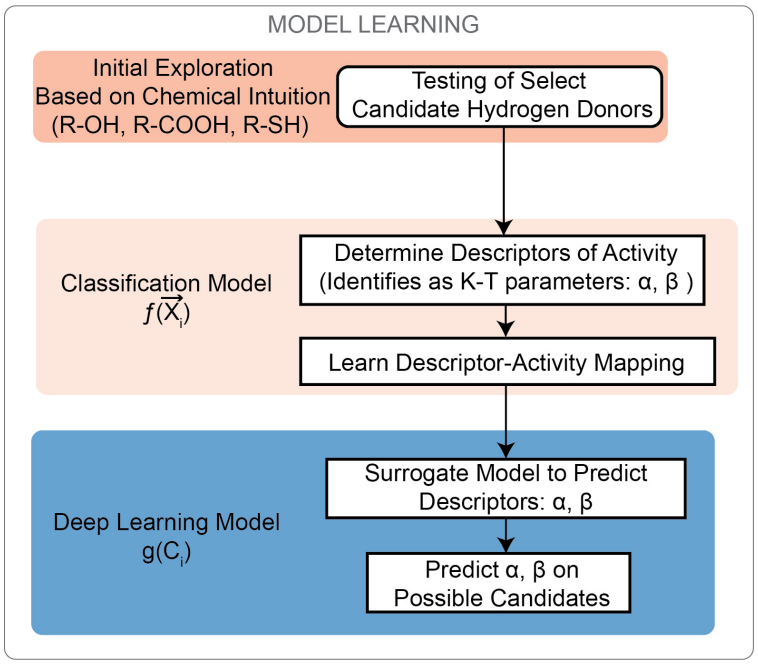
tions, ranging from 0 to 80 μM. A typical calibration curve and absorbance spectra can be seen in Fig S9. Addition of most proton donors to the electrolyte used do not change the calibration curve significantly; proton donors which may affect the quantification (such as thiols or long chain alcohols) were typically removed by extraction prior to quantification (Supporting Information Table 2). The lowest accurately quantifiable concentration of ammonia in the solutions was typically 2 μM, computed from the error in the intercept of the calibration curve. Assuming a ten-fold dilution of the electrolyte solution to make a sample, the minimum quantifiable ammonia FE from a typical experiment is 0.1%.

Analysis of water content

A possible explanation for the differences in activity between various compound could the difference in water content between various proton donors. However, even the most water-rich proton donor tested, triethylene glycol, contained only 2600 parts per million (ppm) water, which corresponds to a water concentration of 130 mM in the pure proton donor. Considering that the maximum concentration of proton donor used was 1 M (Supporting Information Table 2), the amount of water added by the proton donor is at most 17 mM. While this is a non-negligible amount, it is similar to the amount of water present in the electrolyte initially, as measured by Karl Fischer titration. In addition, an increase in the water content by 17 mM is predicted to decrease the ammonia yields of an active proton donor, such as ethanol, by 20-30%,[1] not eliminate it entirely. As mentioned in the main text, there is not a strong correlation between the water content of the pure proton donor and its activity toward LM-NRR (Fig. S8). All this supports the notion that the differences in activity observed between proton donors are not simply due to a difference in water content, but are instead related to the chemical structure and properties of the proton donors.

Modeling approach

The following figure shows the computational modeling workflow. The three parts (in the order of construction): initial exploration, classification model construction and the deep learning model building are discussed in detail below.



Initial exploration: This is the phase of data generation for the discovery of descriptors for activity towards ammonia production. Data emerged from experimental testing of candidates within the families of largely alcohols, thiols and carboxylic acids based on chemical intuition and promising classes discussed in the literature.

The *Classification model* details and the *Deep Learning Model* details are listed below in separate sections. Please refer to the methods section of the main manuscript in conjunction with these details.

Classification Model

Several models including linear and non-linear supervised learning models, regression models and decision tree were constructed. For results pertaining to decision tree, support vector ma-

chine (Gaussian and polynomial kernels), diagonal quadratic model, linear and quadratic kernels, please see Fig. S10 and S11. We find that the classification tree is high performing while being highly interpretable based on mechanistic insights. Hence, in the following section focus on the classification tree model.

Classification trees: This class of machine-learning methods are typically for constructing models to partition data into different classes. Classification models are constructed by recursively partitioning the data space and fitting a simple prediction model within each partition. The partitioning can be represented graphically as a decision tree as shown in the main manuscript (Fig. 3). The prediction error of classification trees is typically measured in terms of misclassification cost. Classification trees are designed for dependent variables that take a finite number of unordered values. A key advantage of the tree structure is its applicability to any number of variables. In this work, we employ a binary classification tree for experimental activity classification since it is particularly well suited for this application.

Training data: In the activity classification problem at hand, we have a training sample of many observations on a class variable Y (for activity towards ammonia production) that takes values 0 or 1 (inactive or active respectively), and p predictor variables, X_1,..., X_p. Our goal is to find a model for predicting the values of Y from new X values. In theory, the solution is simply a partition of the X space into disjoint sets, A_1, A_2,..., A_k, such that the predicted value of Y is j if X belongs to A_j, for $j = 0, 1$. In our case, $p = 8$ with candidate descriptors of activity towards ammonia production being acid dissociation constant (pKa), donor number (DN), Dielectric Constant (ϵ_r), Kamlet-Taft parameters (α, β, π), highest occupied molecular orbital level (HOMO), lowest unoccupied molecular orbital level (LUMO), band gap (BG), Bader volume (BV).

Learning algorithm: The classification tree learning algorithm is carried out within MATLAB (R2017a). The Classfication tree methods yield rectangular sets A_j of the predictor variable by recursively partitioning the data set one X variable at a time. Several classification tree algorithms, abbreviated as C4.5,[2] CART,[3] CHAID,[4] CRUISE,[5,6] GUIDE[7] and QUEST[8] have been proposed since the first published classification tree algorithm, THAID.[9,10] The typical algorithm

(pseudocode) for the construction of classification trees is:[11]

1. Start at the root node
2. For each ordered variable X, convert it to an unordered variable X' by grouping its values in the node into a small number of intervals. If X is unordered, set $X' = X$.
3. Perform a chi-squared test of independence of each X' variable versus Y (activity classification, in this case) on the data in the node and compute its significance probability.
4. Choose the variable $X*$ associated with the X' that has the smallest significance probability.
5. Find the split set $\{X* \in S*\}$ that minimizes the sum of Gini indexes and use it to split the node into two child nodes. The Gini index is a generalization of the binomial variance, which is used as an impurity index. Other algorithm use entropy as the impurity index.
6. If a stopping criterion is reached, exit. Otherwise, apply steps 2–5 to each child node.
7. Prune the tree with the CART method.[3]

Tree architecture optimization: Decision trees have to be optimized before being used for classification of new data because the highest accuracy model could be highly complex and consist of hundreds of levels. Therefore, tree optimization implies choosing the right size of tree, which involves cutting off insignificant nodes and even subtrees. Cross-validation is a typical pruning algorithm used in practice.[12]

Cross validation: The process of cross-validation is based on optimal proportion to strike the general trade-off between tree complexity and misclassification error. With an increase in size of the tree, misclassification error typically decreases and in the case of maximum tree size, misclassification error equals 0. On the other hand, complex decision trees poorly perform on generalizability towards independent data, which is termed *true predictive power* of the tree. Therefore, the primary task is to find the optimal proportion between the tree complexity and misclassification

error, which is achieved through a cost-complexity function:

$$R_\alpha(T) = R_\alpha(T) + \alpha(\tilde{T}) \longrightarrow \min_T$$

where $R(T)$ is the misclassification error of the tree T, $\alpha(\tilde{T})$ is complexity measure which depends on $\tilde{T}$, the total sum of terminal nodes in the tree. The α parameter is found through the sequence of in-sample testing when a part of learning sample is used to build the tree, the other part of the data is taken as a testing sample.

The process repeated several times for randomly selected learning and testing samples. Although cross-validation does not require adjustment of any parameters, this process is time consuming since the sequence of trees is constructed. Because the testing and learning sample are chosen randomly, the final tree may differ from time to time. The classification tree (model) reported in this work is well-converged and validated through several runs with different starting points.

Physical reasoning of the decision tree model: The LR decision tree, shown in Figure 4a of the main manuscript, identifies a rationalizable criterion for above-threshold activity towards electrochemical ammonia production, which can be written down as $\alpha > \alpha_t = 0.795$ and $\beta > \beta_t = 0.635$. We rationalize the optimal classification tree based on the fact that promising proton carriers should exhibit both high acidity ($\alpha > \alpha_t$) and high basicity ($\beta > \beta_t$). This is in accordance with the fact that the key displacement reaction for ammonia production, $Li_3N+3HA \rightarrow NH_3 +3LiA$, not only involves the release of a proton by the hydrogen carrier but also requires abstraction by the carrier of a Li^+, the closest chemical analogue to a proton.

Deep Learning Model

Refer to the methods section of the main manuscript in conjunction with the details here.

Model input: The deep learning model's input is the molecular features of candidates obtained

from the simplified molecular-input line-entry system (SMILES).

Featurization: Out of several molecular featurization approaches on SMILES representations, we find that the Weave featurization coupled with a Weave model (deep neural network)[13] yielded the most accurate (RMSE of $\approx$0.016 for both α and β) and generalizable predictions with low cross-validation error). The Weave featurization[14] method encodes both local chemical environment and connectivity of atoms in a molecule. The Weave featurization is similar to graph convolution in the atomic feature vectors, whereas in terms of encoding the connectivity it uses more detailed pair-wise features instead of just neighbor listing by means of "weaving" atom and pair features (Fig. 4). The Weave featurization computes a feature vector for each pair of atoms in the molecule, including bond properties, graph distance and ring information, giving rise to a feature matrix. This approach supports graph-based models that make use of properties of both atoms and bonds.

Weave model details: The molecular features from the weave featurization method were input into a neural network with a converged architecture, which was found to consist of two weave layers and a fully connected layer in regression mode to predict α and β. For the training process, the learning rate was selected to be 0.001, batch size was set to 50 and number of epochs were 100 for the 222 data points available of experimentally obtained α and β values. Version and model specifications are provided in the following table:

DeepChem version	’2.3.0’
Tensorflow version	’1.14.0’
Python version	Python 3.7.7
Featurizer	WeaveFeaturizer, DeepChem RDKit
Dataset Split details	frac_train=0.8, frac_valid=0.1, frac_test=0.1, with random seed
Weave Model Details (several default)	n_tasks (number of tasks): 2, mode=’regression’
n_atom_feat (number of atom features)	75
n_pair_feat (number of pair features)	14
n_hidden	5
n_graph_feat	128
n_weave	int = 2
fully_connected_layer_sizes	[2000, 100]
weight_init_stddevs	[0.01, 0.04]
bias_init_consts	[0.5, 3.0]
weight_decay_penalty	0.0
weight_decay_penalty_type	"l2"
dropouts	0.25
activation_fns	Tensorflow relu
batch_normalize	bool = True
batch_normalize_kwargs	Dict = "renorm": True,"fused": False
gaussian_expand	True
compress_post_gaussian_expansion	False

Sources of Uncertainty

The differences between experimentally measured and predicted proton donor activity could arise from a few potential sources. One source could be changes in the proton donors’ structures under the very reductive conditions of the experiment (in the presence of lithium metal).

For instance, halogenated compounds could easily decompose to form lithium halide salts on the electrode, preventing further electrochemical reactions, as well as changing the structure of the proton donor significantly. Analyzing the effect of these decomposition pathways is outside the scope of the current work and will be analyzed in a subsequent study. Another source of deviation between predictions and experiment could be the fact that the inputs to the classification model are bulk KT parameters, while the effective KT parameters for the electrolyte composition close to the electrode may have some deviation from the bulk value. The agreement between model predictions and experimental activity could be enhanced through a more expressive descriptor-to-activity mapping with additional descriptors for specific material classes of interest (Fig. 3). However, it is worth highlighting that such approaches that maximize predictability over subsets of materials are expected to have modest generalizability. Therefore, we anticipate the current approach coupled with additional experimental testing data to be an effective direction towards other promising proton donors for electrochemical ammonia production.

While most of the experimental and modeling studies in this work are focused on the cathode reaction of nitrogen reduction to ammonia, it is important to briefly discuss possible anode reactions to understand the overall reaction in the system. In the current setup with a flooded platinum anode, likely anode reactions include solvent (THF) and proton donor oxidation. Both of these reactions are undesired in a practical system, but are acceptable in the present study as the anode and cathode compartments are separated by a polyporous separator, which prevents products of oxidation at the anode from affecting the cathode reaction. A more practical anode reaction is hydrogen oxidation to protons, which has been demonstrated in nonaqueous systems.[15] In systems where hydrogen oxidation is the anode reaction, THF oxidation can be avoided. However, proton donor oxidation may still occur if the proton donor is sufficiently electron-rich and its structure allows for oxidation reactions. For example, 2-propanol could be oxidized to acetone under mild conditions, while tert-butanol would be significantly more stable to oxidation. The oxidative stability of proton donors is a direction of future work.

While the approach identified descriptors based on strong correlations with desired outputs,

sufficiently strong correlations help build hypotheses for mechanistic understanding of complex chemical processes. By using a limited set of experimental data, we are able to determine experimental parameters than can predict and affect future experiments (Fig. 3). An automated method for predicting the values of relevant parameters allows for rapid identification of potential leads. By testing novel leads, the model and understanding of the process can be significantly improved when compared to a conventional, intuition-driven experimental approach.

A deep learning based model (material-descriptor mapping) was developed based on molecular features to predict α and β values for any given compound, and was used in conjunction with the classification model (descriptor-activity mapping) to perform a vast search for promising candidates from about 1M compounds. Through a closed-loop approach, candidates were proposed for experimental testing with the primary goal of stress-testing the model and learning the delineating surface (in $\alpha - \beta$ space) between active and inactive candidates. The loop between computation and experiments was closed by data augmentation after every batch of experimental testing. After the initial experimentation phase four loops were carried out (Fig. 5) with batches of experiments performed each time towards learning the material-activity relationship. The closed-loop approach between experiments and theory has enabled an increase in the fraction of tested active candidates from 30% during the initial exploration to 65% during the combined effort. In the process, several novel active proton donors were discovered, demonstrating the robustness and power of the coupled experimental-data driven approach to studying complex systems.

Supporting Information Figures

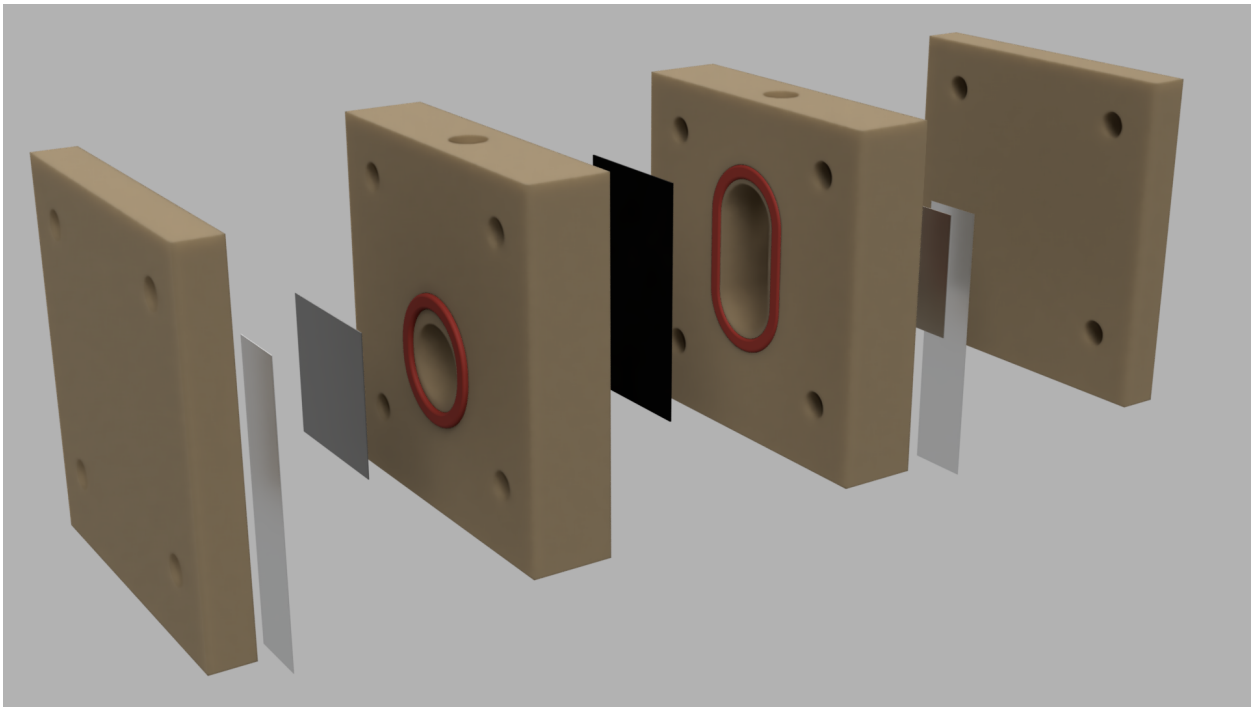


Fig. S1. A depiction of the 2-compartment cell used in electrochemical experiments. The cell body is made of polyether ether ketone (PEEK) polymer. Platinum and polished stainless steel foils were used as the anode and cathode, respectively. The anode and cathode compartments were separated by a piece of polyporous Daramic 175 separator. IDEX fittings were used to feed gases and plug unused holes.

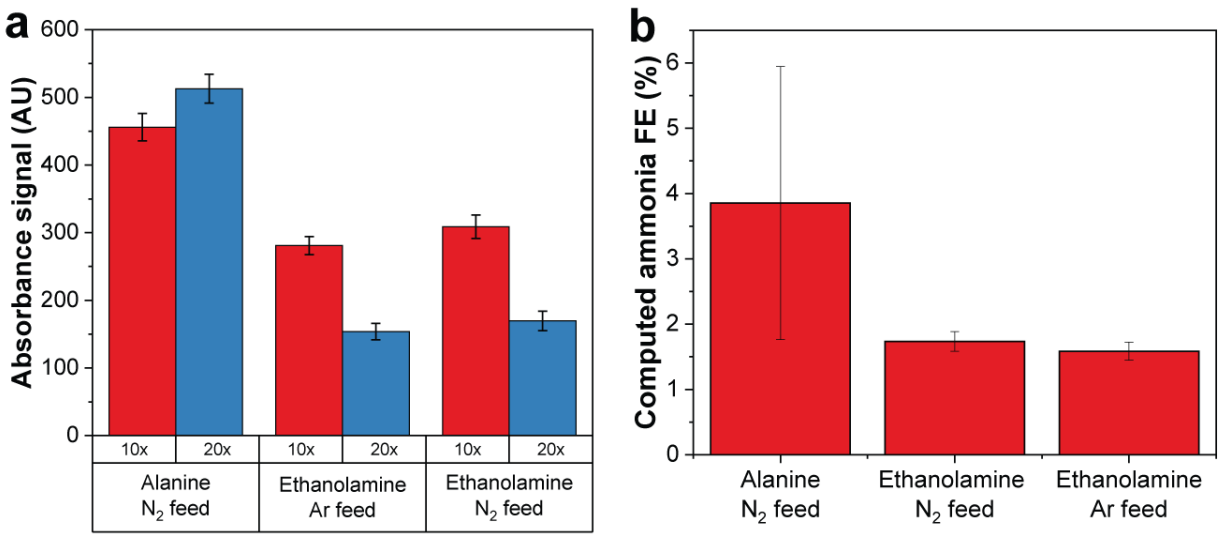


Fig. S2. Nitrogen reduction experiments run with nitrogen-containing proton donors. (**a**) Raw absorbance signal data for runs using 0.2 M ethanolamine or saturated alanine as a proton donor in LM-NRR. Note that alanine is not readily soluble in the electrolyte, so a saturated solution with a concentration <0.1 M was used. In these experiments, significant absorbance signals were detected in the ethanolamine argon blank solutions and alanine-containing solutions. (**b**) NH_3 FE values computed from the absorbance signals in (**a**). Note that a non-zero FE is computed even in the ethanolamine argon blank. The FE value computed for the alanine experiment has significant uncertainty.

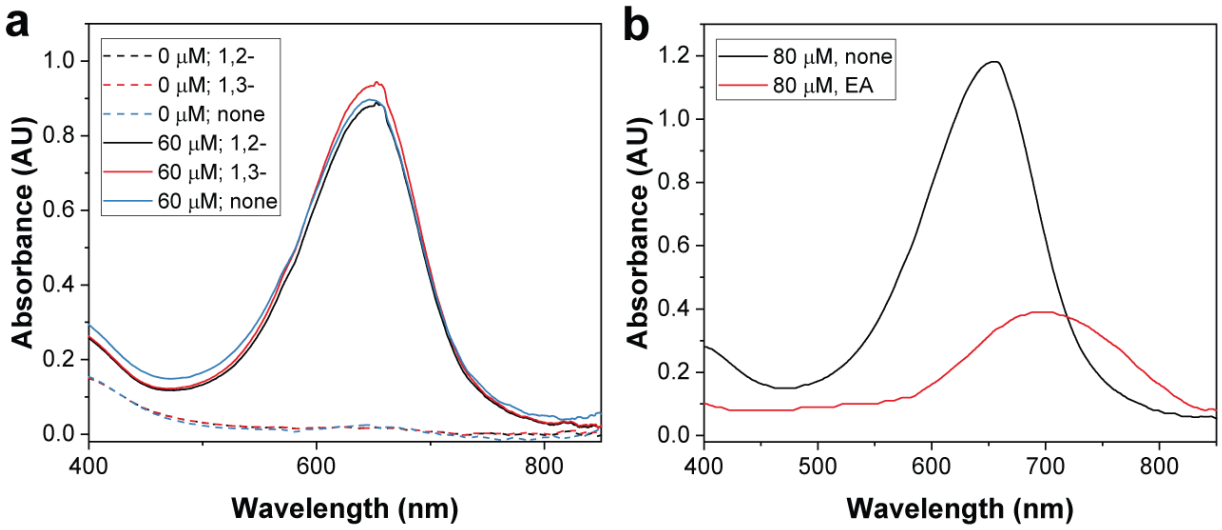


Fig. S3. Various effects of proton donors on the salicylate quantification assay. (**a**) Effect of addition of 0.5 M of 1,2-propanediol and 1,3-propanediol to the electrolyte used to make a 10 v/v% electrolyte in water solution containing a known concentration of ammonium. Note that the addition of the proton donors does not significantly alter the peak shape or signal magnitude. (**b**) Effect of addition of 0.2 M of ethyl acetate (EA) to the electrolyte used to make a 10 v/v% electrolyte in water solution containing 80 μM ammonium. Note that the shape of the peak changes significantly, rendering it useless for quantification of ammonia in the solution. Samples containing protons sources such as these were typically extracted with dichloromethane or hexanes prior to quantification.

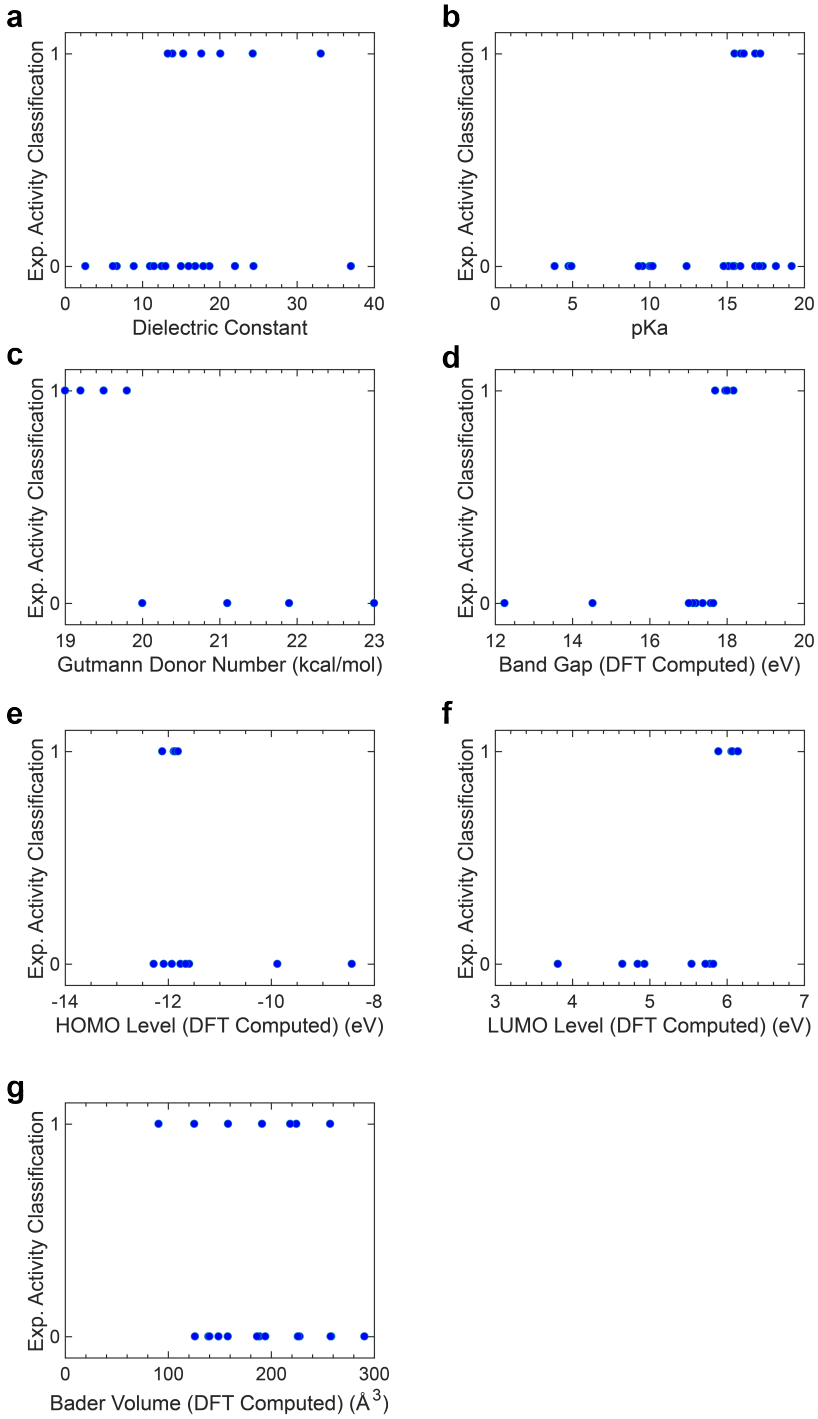


Fig. S4. Simple refuted hypotheses for explaining experimentally proton donor activity trends. (**a**) The binary activity of the proton donor plotted against the proton donor's pKa in DMSO. Note that while there is no direct correlation between the proton donor pKa and activity. (**b**) The binary activity of the hydrogen source plotted against the proton donor's donor number.

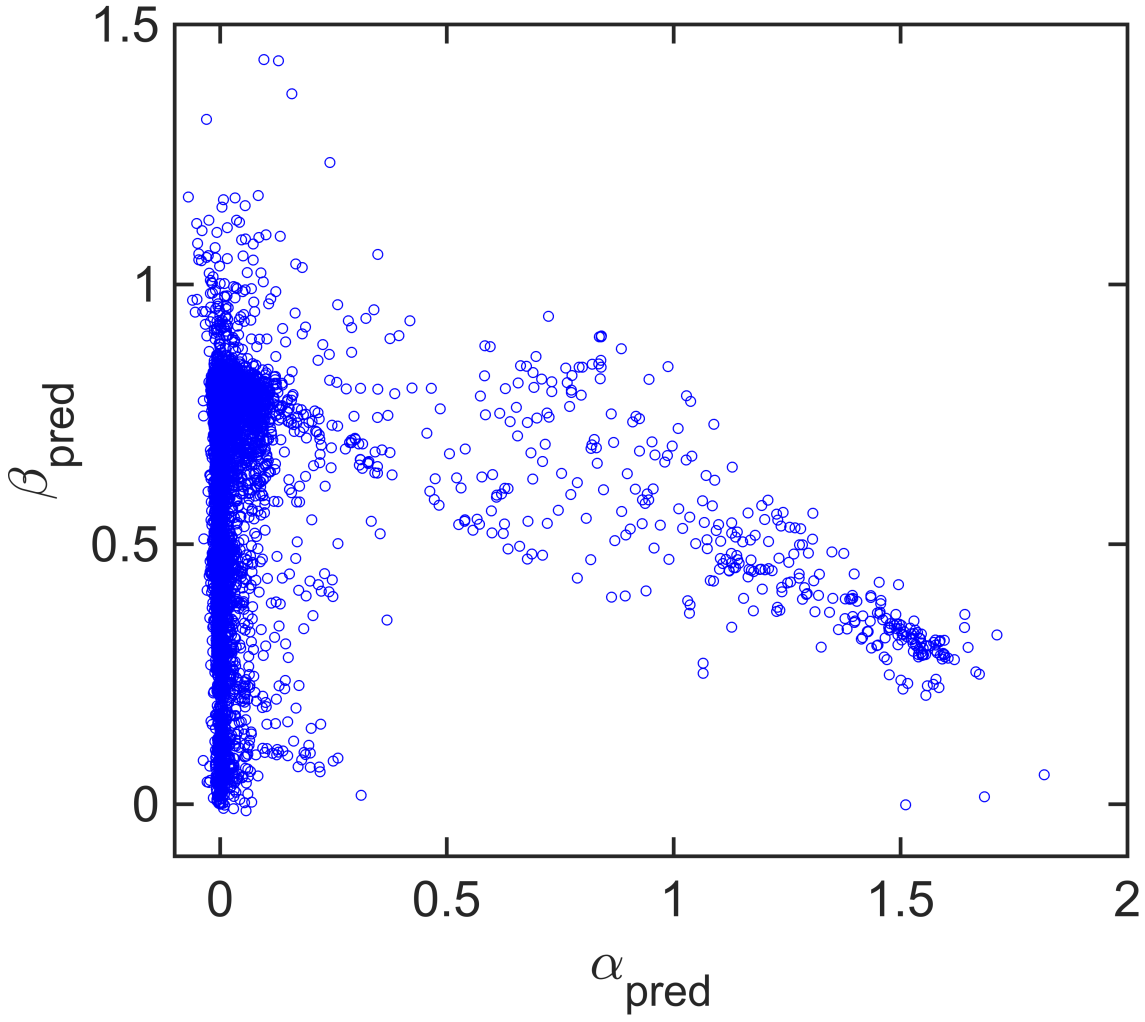


Fig. S5. Predicted average Kamlet-Taft parameters for a subset of proton donors. Proton donors with Pubchem IDs below 10,000 and with above-threshold (standard deviation ≤ 0.2) agreement between models in the ensemble are plotted; candidates with higher standard deviations are considered to have uncertain predictions. The inherent trade-off emerges in the $\alpha - \beta$ space as shown whereby achieving high α and β is challenging, which can rationalized based on the trade-off between proton donating and accepting tendencies of molecules represented by the K-T parameters.

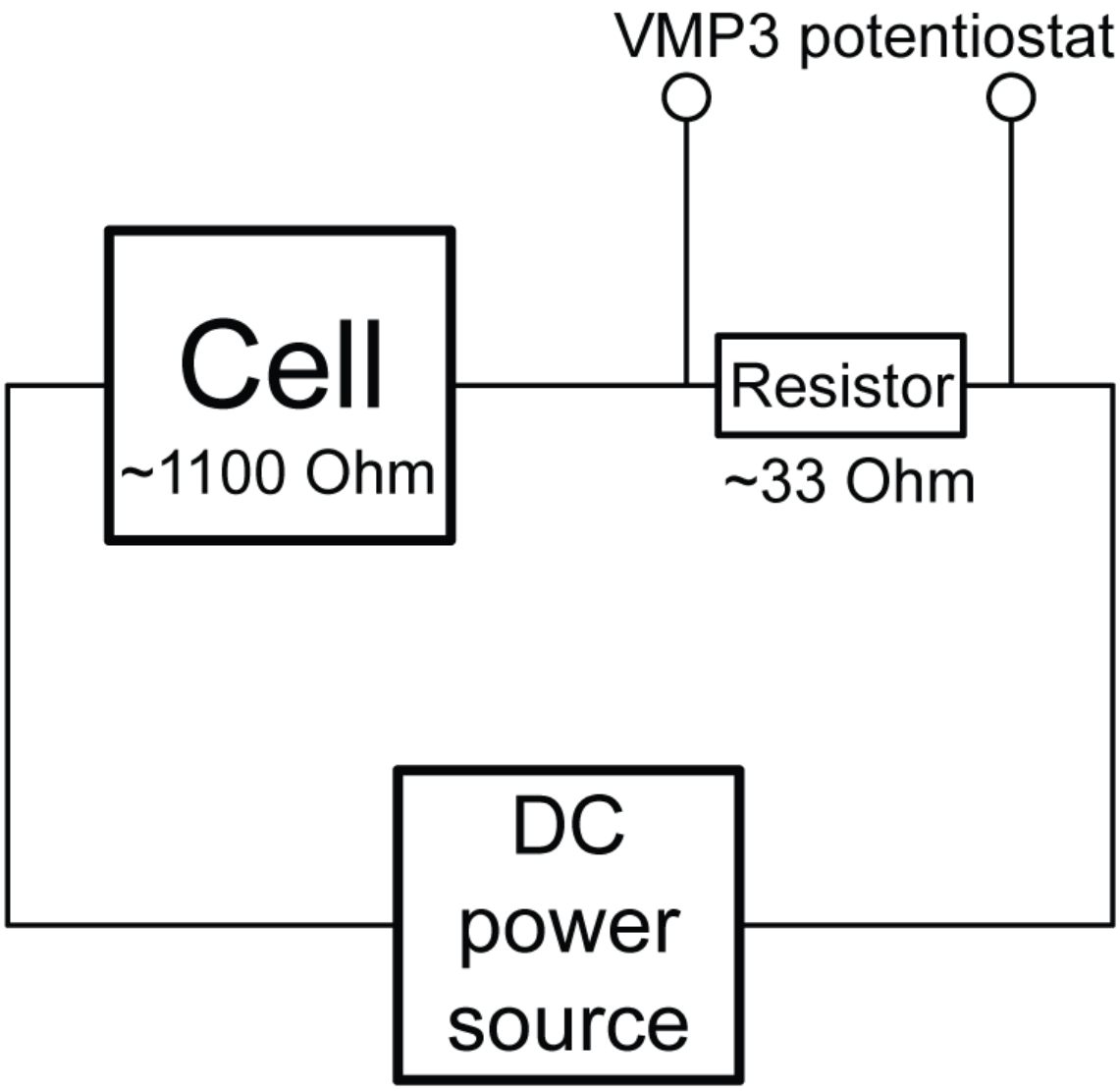


Fig. S6. A diagram of the wiring scheme used to measure charge passed in experiments. As the DC power source cannot independently quantify charge, the current passed through the circuit was measured with an accurate VMP3 potentiostat.

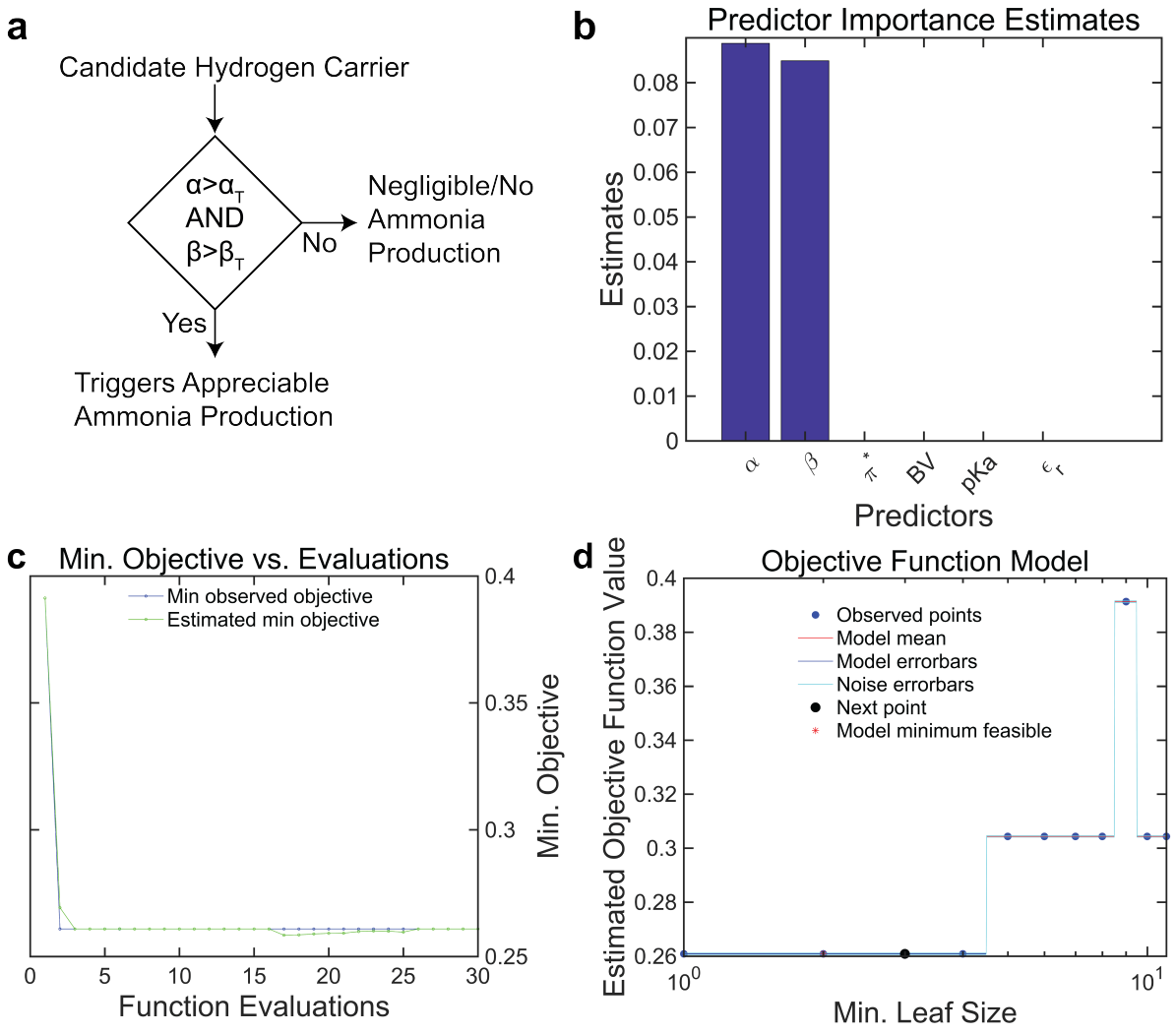


Fig. S7. Details of the activity classification model based on decision trees. **(a)** Identified classification model with α_t and β_t equal to 0.815 and 0.59 respectively. **(b)** Predictor importance estimates by summing up changes in the risk due to splits on every predictor and dividing the sum by the number of branch node. **(c)** Minimum of the objective function, normalized misclassification rate, as a function of functional evaluations. **(d)** Evolution of the estimated objective function value with the associated minimum leaf size (classification model complexity).

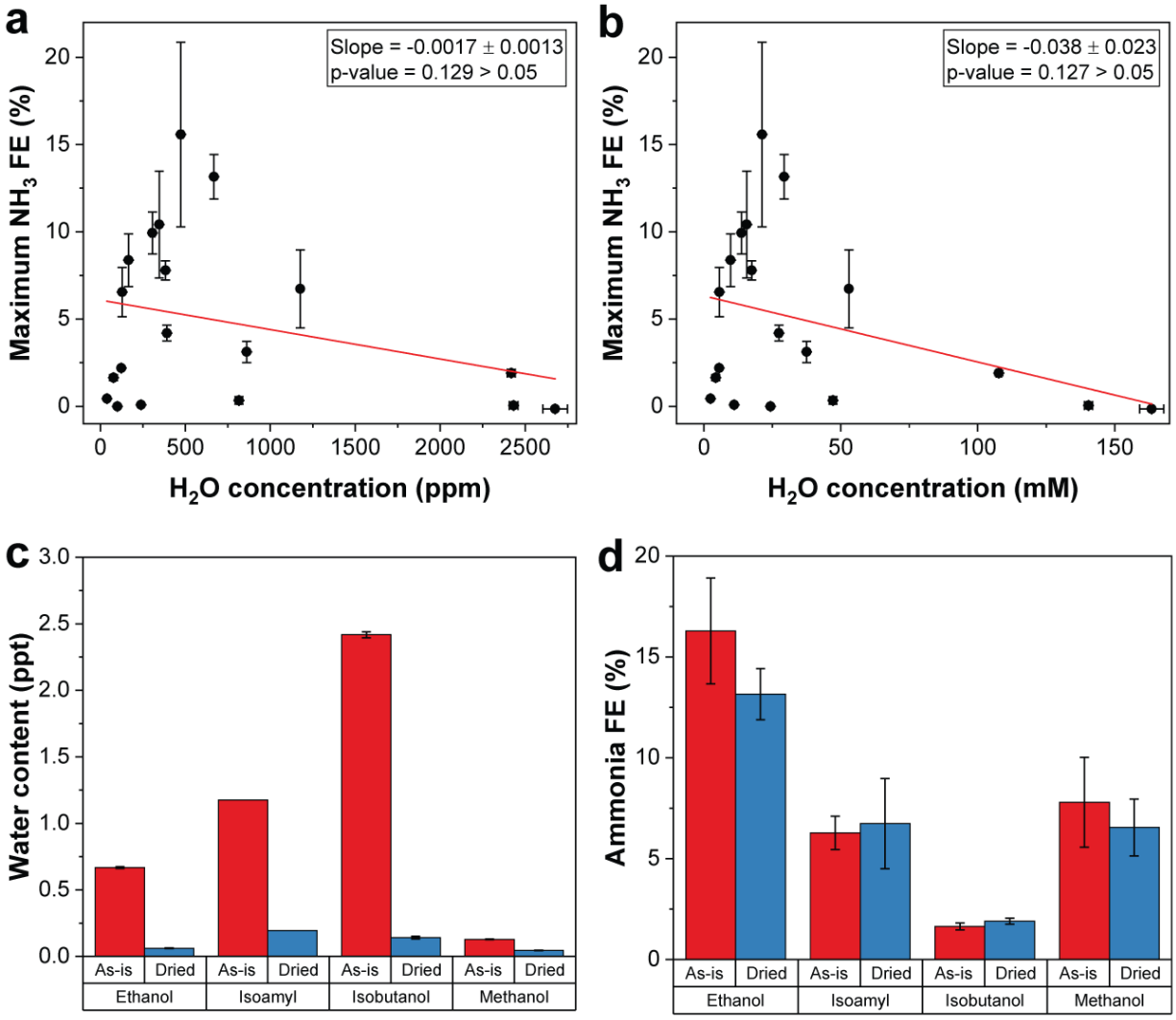


Fig. S8. Effect of water in the proton donor on nitrogen reduction activity (a, b) Dependence of the maximum ammonia FE obtained with a single proton donor on water content **(a)** parts per million (ppm) and **(b)** millimolar units. Note the maximum FE correlates poorly with water concentration ($p \approx 0.12$ for a non-zero slope), suggesting that the water content of the proton donor is a poor predictor of nitrogen reduction activity. **(c)** The water content in parts per thousand (ppt) of various proton donors before drying with molecular sieves and after. **(d)** The NH_3 FE for several proton donors before and after drying the proton donors with molecular sieves. Note that the FE does not change in a predictable manner after drying with sieves.

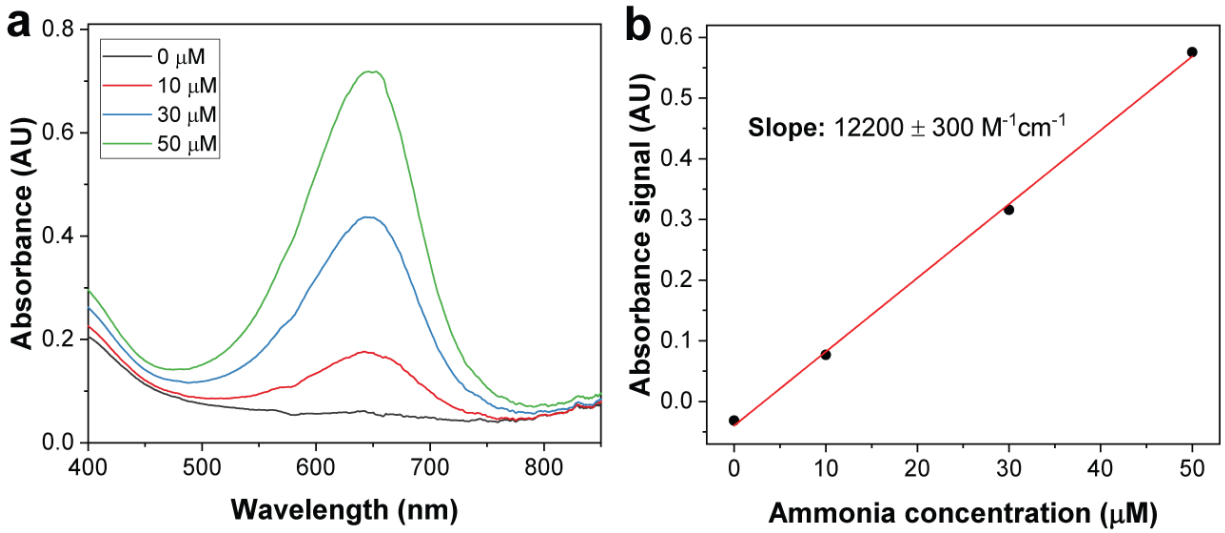


Fig. S9. Typical ammonia quantification calibration curves. (**a**) Absorbance spectra obtained from solutions containing known concentrations of ammonium ions in 10 v/v% 1 M $LiBF_4$ in THF electrolyte in water. (**b**) A typical calibration curve made from the spectra in (**a**). Note that the absorbance signal is taken to be the difference between the absorbance at 650 and 475 nm.[1]

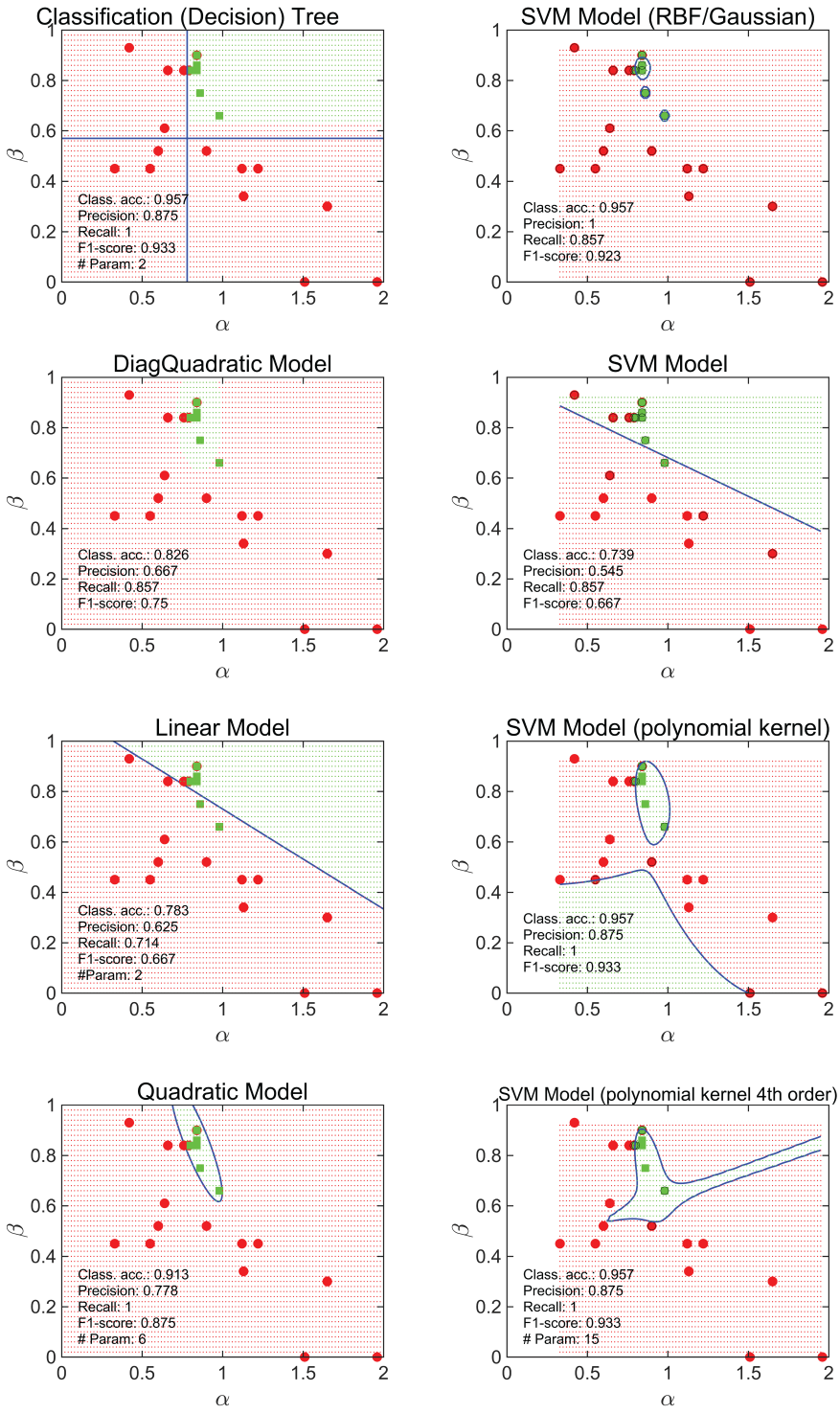


Fig. S10. Classification models trained on the initial set of data (post initial experimental testing) to delineate active candidates from inactive candidates.

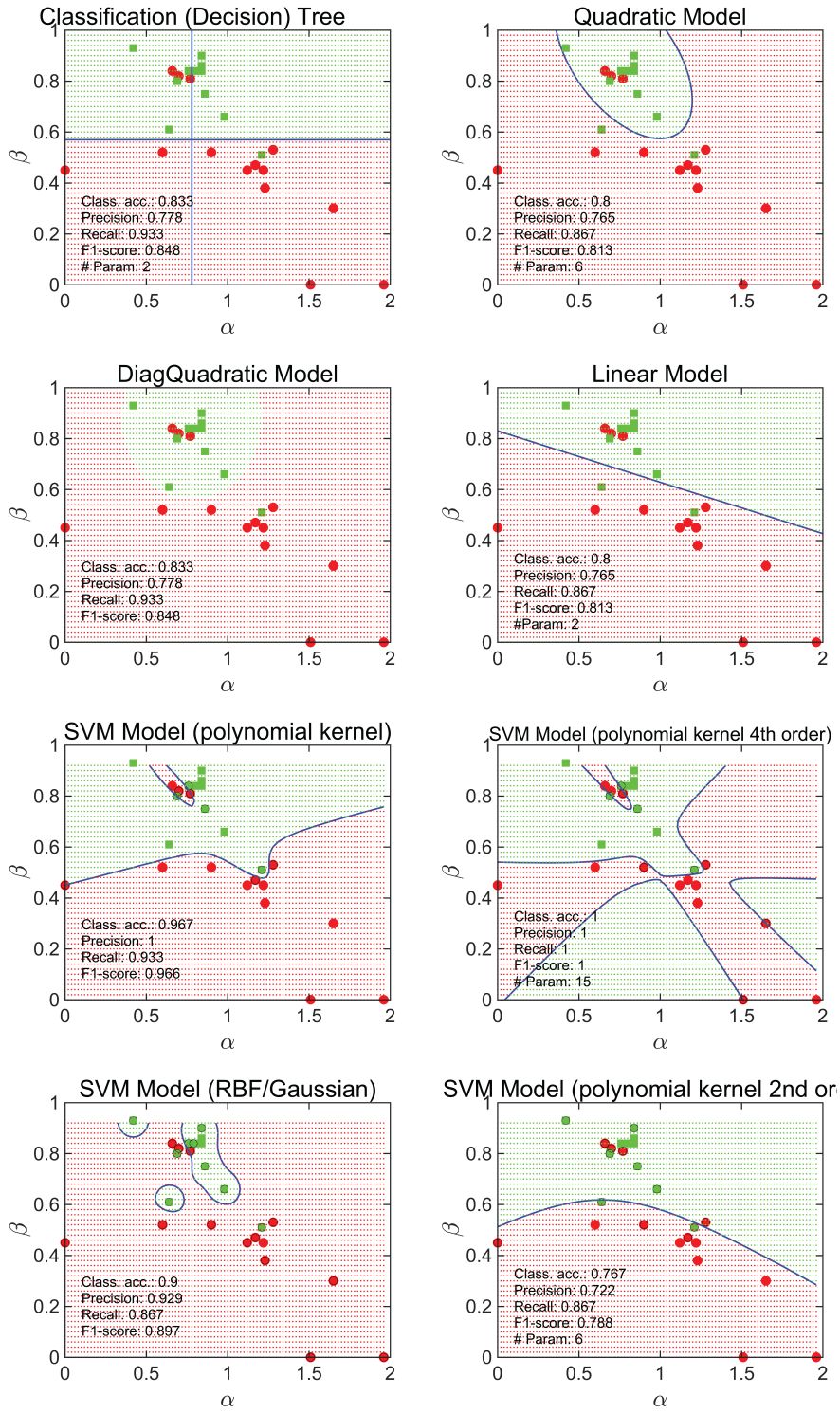


Fig. S11. Classification models trained on the final set of data (after all experimental testing) to delineate active candidates from inactive candidates.

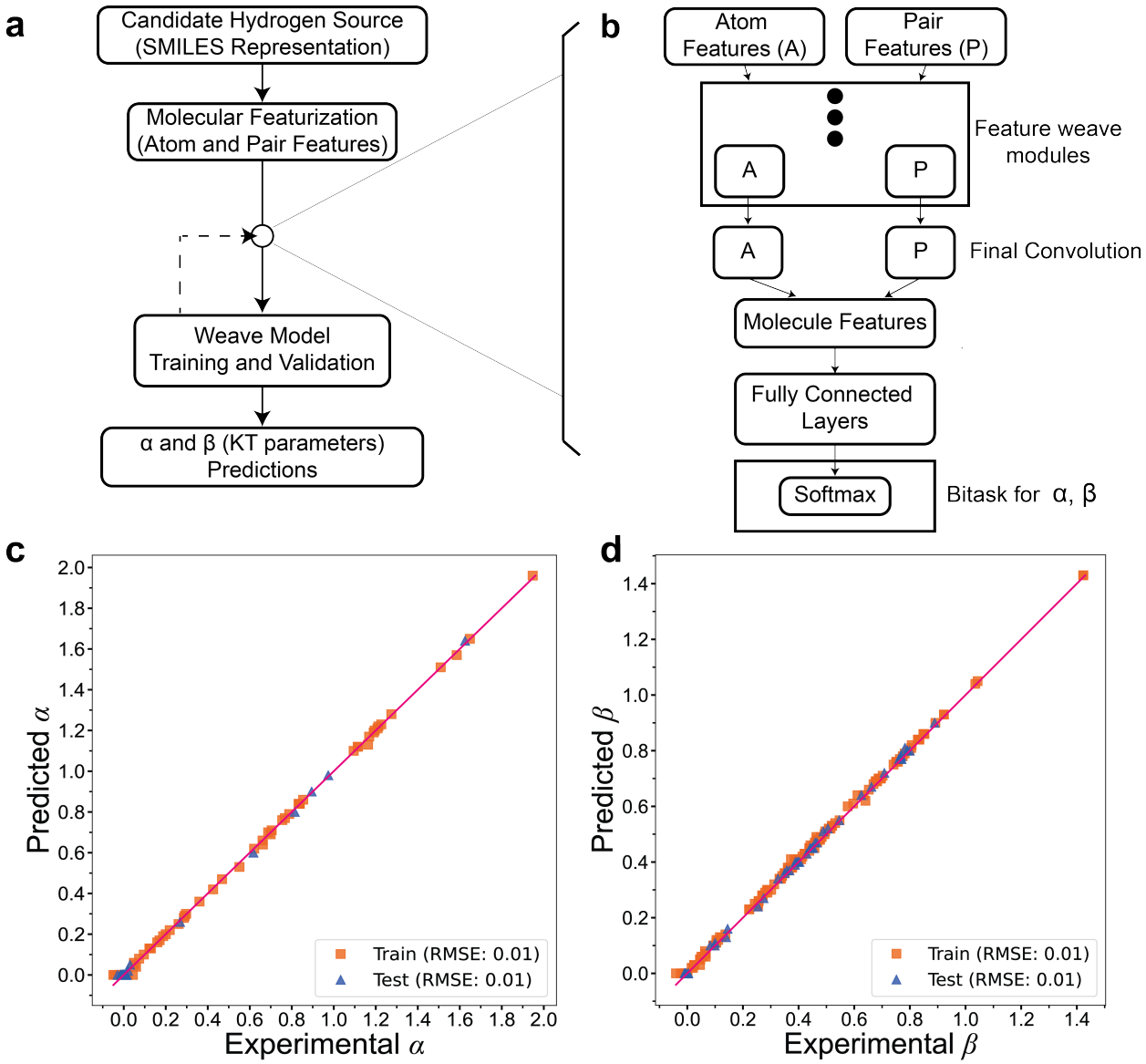


Fig. S12. A deep learning model to predict Kamlet-Taft parameters. (**a**) Steps involved in the approach to predicting α and β (KT) parameters. (**b**) The weave featurization technique and deep learning framework involving an ensemble of models for robust predictions.[14] Parity plots for (**c**) α and (**d**) β values obtained from the developed deep-learning model. Note that the predictions on test set after cross-validation have comparable performance to that on the training set indicating generalizability of the model.

Supporting Information Tables

Table S1. Measured water content of pure proton donors.

Proton donor	Water content (ppm)	Water concentration (mM)
1,2-propanediol	2430 ± 30	140 ± 1
1,3-butanediol	78 ± 9	4.4 ± 0.5
1,3-propanediol	166 ± 2	9.7 ± 0.1
1-butanol	470 ± 10	21.3 ± 0.6
1-heptanol	123.6 ± 0.9	5.62 ± 0.04
1-hexanol	383 ± 5	17.5 ± 0.2
1-octanol	239 ± 1	11.04 ± 0.05
1-pentanol	346 ± 8	15.7 ± 0.4
1-propanol	310 ± 10	13.7 ± 0.6
2-propanol	860 ± 5	37.6 ± 0.2
Benzyl alcohol	820 ± 10	47.1 ± 0.6
Cyclohexanol	455 ± 8	24.3 ± 0.4
Ethanol	668 ± 6	29.3 ± 0.3
Ethylene glycol	39 ± 1	2.42 ± 0.07
Glycerol	390 ± 10	27.4 ± 0.8
Isoamyl alcohol	1177 ± 1	52.97 ± 0.06
Isobutanol	2420 ± 20	108 ± 1
Methanol	128 ± 2	5.64 ± 0.09
Triethylene glycol	2680 ± 70	164 ± 4

Table S2. Experimentally tested proton donors and details of experimental results. Proton donors were tested at several concentrations (see Supporting Information methods); the concentration at which the highest FE was obtained is reported along with the FE. At the concentration that led to the highest FE, three repeat experiments were conducted and the resulting standard deviation is listed below as the Max FE error. If ammonia quantification solutions required extraction (see Supporting Information methods), it is reported below. The binary activity classification is also given.

Compound Name	Max FE (%)	Max FE error (%)	Conc. at max FE (M)	Charge (C)	Solvent used for extraction	Experimental activity classification
1,2-propanediol	0.04	0.02	0.2	7.2	None	FALSE
1,3-butanediol	1.65	0.17	0.2	7.2	None	TRUE

Compound Name	**Max FE (%)**	**Max FE error (%)**	**Conc. at max FE (M)**	**Charge (C)**	**Extraction solvent**	**Experimental activity classification**
1,3-propanediol	8.38	1.51	0.1	7.2	None	TRUE
1,4-cyclohexane dimethanol	0.02	0.05	0.2	7.2	None	FALSE
1,5-pentanediol	4.43	1.39	0.2	1.5	None	TRUE
1-butanol	15.58	5.29	0.1	7.2	None	TRUE
1-decanol	0.09	0.05	1	7.2	Hexane	FALSE
1-heptanol	2.19	0.08	1	7.2	Hexane	TRUE
1-hexanol	7.79	0.55	0.6	7.2	Hexane	TRUE
1-nonanol	0.18	0.08	0.6	7.2	Hexane	FALSE
1-octanol	0.08	0.05	0.2	7.2	Hexane	FALSE
1-pentanol	10.42	3.06	0.2	7.2	None	TRUE
1-phenylethanol	1.02	0.20	0.8	7.2	None	TRUE
1-propanol	9.93	1.20	0.1	7.2	None	TRUE
2,2,2-trifluoroethanol	0.02	0.06	0.4	7.2	None	FALSE
2,2-difluoroethanol	0.02	0.00	0.5	7.2	None	FALSE
2,2-dimethyl-1,3-propanediol	0.84	0.17	0.4	7.2	None	TRUE
2-butanol	1.36	0.06	1	7.2	None	TRUE
2-chloroethanol	0.06	0.02	0.2	7.2	None	FALSE
2-ethyl-1-butanol	3.62	0.59	0.2	7.2	None	TRUE
2-phenylethanol	1.64	0.26	0.9	7.2	None	TRUE
3-butene-1-ol	1.94	0.08	0.6	7.2	None	TRUE

Compound Name	**Max FE (%)**	**Max FE error (%)**	**Conc. at max FE (M)**	**Charge (C)**	**Extraction solvent**	**Experimental activity classification**
4-methoxybutan-1-ol	0.34	0.02	0.4	7.2	None	FALSE
Acetic acid	0.19	0.07	0.07	7.2	None	FALSE
Allyl alcohol	0.69	0.14	0.6	7.2	None	TRUE
Benzyl alcohol	0.34	0.21	0.8	7.2	None	FALSE
Cyclohexanol	0.00	0.04	0.6	7.2	None	FALSE
Ethanol	13.16	1.27	0.1	7.2	None	TRUE
Ethyl acetate	0.15	0.06	0.2	7.2	DCM	FALSE
Ethylene glycol	0.44	0.03	0.4	4.1	None	FALSE
Formic acid	0.00	0.07	0.2	7.2	None	FALSE
Glycerol	4.20	0.45	0.2	4.6	None	TRUE
Hexafluoro iso-propyl alcohol	0.03	0.09	0.4	7.2	None	FALSE
Hexanoic acid	0.00	0.07	0.2	7.2	None	FALSE
Isoamyl alcohol	6.74	2.24	0.6	7.2	None	TRUE
Isobutanol	3.09	0.43	0.4	7.2	None	TRUE
Isopropyl alcohol	3.12	0.61	0.2	7.2	None	TRUE
Lactic acid	-0.03	0.22	0.1	7.2	None	FALSE
Methanol	6.55	1.41	0.2	7.2	None	TRUE
Phenol	-0.05	0.06	0.2	7.2	None	FALSE
Propanethiol	0.03	0.03	0.1	7.2	DCM	FALSE
t-butyl alcohol	0.57	0.08	0.6	7.2	None	FALSE
Triethylene glycol	-0.15	0.03	0.2	7.2	None	FALSE

Compound Name	Max FE (%)	Max FE error (%)	Conc. at max FE (M)	Charge (C)	Extraction solvent	Experimental activity classification
Water	-0.06	0.02	0.2	7.2	None	FALSE

Table S3. Experimentally measured initial and final cell voltages for selected proton donors. While recording voltage-time curves was infeasbile due to setup limitations (see Fig. S6), the initial and final cell voltages could be recorded and used for qualitatively analyzing experimental results. In general, the cell voltage stays constant or increases with time. The increase in cell voltage is larger at higher concentrations for certain proton donors. For certain proton donors, particularly diols, the cell voltage necessary to apply a constant 20 mA current exceeded 50 V, which is the largest voltage that the power source could apply. In these cases, the current flowing through the cell was lower. No direct correlation between cell voltages and ammonia FEs were observed, though in experiments where ammonia was produced, a moderate increase in potential was observed.

Compound Name	Exp. Conc. (M)	Init. voltage (V)	Final Voltage (V)	FE (%)
Diols and polyols				
1,2-propanediol	0.2	23.8	24	0.04
1,2-propanediol	0.4	22.2	22.6	0.04
1,2-propanediol	0.6	21.1	21.3	-0.03
1,3-butanediol	0.1	24.6	25	1.35
1,3-butanediol	0.2	24.5	26.1	1.65
1,3-butanediol	0.4	22.4	>50	0.25
1,3-propanediol	0.1	25.3	26.2	8.38
1,3-propanediol	0.2	23.4	30	1.42
1,3-propanediol	0.3	25	>50	0.40
1,4-cyclohexanedimethanol	0.2	26.8	27.9	0.07
1,4-cyclohexanedimethanol	0.4	29.3	>50	0.84
1,4-cyclohexanedimethanol	0.6	30	>50	0.72

Compound Name	**Exp. Conc. (M)**	**Init. voltage (V)**	**Final Voltage (V)**	**FE (%)**
1,5-pentanediol	0.05	27.3	27.3	0.21
1,5-pentanediol	0.1	27.5	28.5	1.66
1,5-pentanediol	0.2	27	>50	4.43
Ethylene glycol	0.1	24.5	24.8	-0.05
Ethylene glycol	0.2	24.6	24.2	0.00
Ethylene glycol	0.4	20	>50	0.44
2,2-dimethyl-1,3-propanediol	0.2	25.3	25.5	0.02
2,2-dimethyl-1,3-propanediol	0.4	23.5	25.6	-0.07
2,2-dimethyl-1,3-propanediol	0.6	23.1	23.1	-0.03
Glycerol	0.07	27.9	28.3	0.04
Glycerol	0.1	26.6	28.1	0.39
Glycerol	0.2	42	>50	4.20
Linear aliphatic alcohols				
1-butanol	0.05	25	25.1	0.08
1-butanol	0.1	25.3	26	15.58
1-butanol	0.2	26.2	27.23	7.82
1-decanol	0.2	28.1	28.2	0.13
1-decanol	0.6	31.5	31.8	0.11
1-decanol	1	39.6	40	0.08
1-heptanol	0.2	27.2	27.3	0.06
1-heptanol	0.6	29.2	29.8	0.23
1-heptanol	1	30.5	31.4	2.19
1-hexanol	0.2	26.85	27	0.77
1-hexanol	0.6	27.2	28.45	7.79
1-hexanol	1	29.9	31.2	5.82

Compound Name	**Exp. Conc. (M)**	**Init. voltage (V)**	**Final Voltage (V)**	**FE (%)**
1-octanol	0.2	27.6	27.8	0.08
1-octanol	0.6	33.6	31.5	0.05
1-octanol	1	32.8	33.8	0.01
Ethanol	0.1	24.5	25.2	13.16
Ethanol	0.2	25.5	26.8	6.59
Ethanol	0.4	23.7	25.4	2.64
Methanol	0.1	25.7	25.8	1.55
Methanol	0.2	24	25	6.55
Methanol	0.4	23	27.1	2.28
Other alcohols				
1-phenylethanol	0.2	28	27.8	0.09
1-phenylethanol	0.6	30.2	30.8	0.07
1-phenylethanol	1	32.8	48	1.02
2-ethyl-1-butanol	0.2	27.7	28.2	3.62
2-ethyl-1-butanol	0.4	29.4	29.8	1.21
2-ethyl-1-butanol	0.8	31.3	32.4	1.05
2-phenylethanol	0.2	27.3	27.4	0.10
2-phenylethanol	0.5	28.1	30	0.47
2-phenylethanol	0.9	28.9	38.7	1.64
4-methoxybutan-1-ol	0.1	24.8	24.9	-0.01
4-methoxybutan-1-ol	0.2	24.2	24.6	0.34
4-methoxybutan-1-ol	0.4	22.8	23.5	0.05
Allyl alcohol	0.2	28.6	28.6	0.02
Allyl alcohol	0.4	26.3	27.3	-0.03
Allyl alcohol	0.6	26.7	32	0.69

Compound Name	**Exp. Conc. (M)**	**Init. voltage (V)**	**Final Voltage (V)**	**FE (%)**
Benzyl alcohol	0.2	26.4	27.3	0.20
Benzyl alcohol	0.5	27	31.3	0.29
Benzyl alcohol	0.8	27.9	40	0.34
Cyclohexanol	0.2	24.8	24.65	0.00
Cyclohexanol	0.6	27.3	27.5	0.00
Cyclohexanol	1	28.9	28.8	-0.14
Isobutanol	0.2	26.4	26.7	1.24
Isobutanol	0.4	26	28.4	3.09
Isobutanol	0.6	26.1	33	1.64
Isopropyl alcohol	0.2	25.7	26.2	3.12
Isopropyl alcohol	0.4	26.2	26.9	1.47
Isopropyl alcohol	0.6	26.2	27.3	1.34
t-butyl alcohol	0.2	26.4	26.4	0.40
t-butyl alcohol	0.6	28.2	27.6	0.57
Triethylene glycol	0.2	22	22.3	-0.15
Triethylene glycol	0.5	18.1	21.3	-0.16
Triethylene glycol	0.8	17.5	19.2	-0.17
Halogenated alcohols				
2,2,2-trifluoroethanol	0.1	24.9	26.4	-0.01
2,2,2-trifluoroethanol	0.4	24.2	24.6	0.02
2-chloroethanol	0.2	23.9	24	0.06
2-chloroethanol	0.6	23.6	23.7	-0.02
2-chloroethanol	1	21.6	22	0.01
Carboxylic acids				
Acetic acid	0.07	26.2	26.8	0.19

Compound Name	**Exp. Conc. (M)**	**Init. voltage (V)**	**Final Voltage (V)**	**FE (%)**
Acetic acid	0.12	24.7	26.7	-0.07
Formic acid	0.1	25	25.6	0.00
Formic acid	0.2	24.9	25.6	0.00
Formic acid	0.3	23	23.9	-0.05
Hexanoic acid	0.2	27.8	28.8	0.00
Hexanoic acid	0.4	30.7	34.9	-0.08
Hexanoic acid	0.6	31.8	36.6	-0.16
Acetic acid	0.2	24.8	26.5	0.04
Other				
Ethyl acetate	0.2	27	27	0.15
Ethyl acetate	0.6	27.9	27.8	0.13
Ethyl acetate	1	29.2	29.5	0.12
Phenol	0.6	26.5	27.1	-0.13
Phenol	1	27.3	28	-0.17
Propanethiol	0.1	26	26.5	0.03
Propanethiol	0.2	26.5	26.7	-0.03
Propanethiol	0.4	27	27.3	-0.04
Water	0.2	22	22	-0.06
Water	0.4	20.5	20.7	-0.07
Water	0.6	20.2	20.3	-0.09

Table S4. A subset of the initial input data for the activity classification model. Here, Kamlet-Taft (α, β, π^*) parameters obtained from experimental reports in the literature [16–18] are given and the corresponding experimental activity classification obtained in the current work. Additional parameters were also fed to the initial classification model to determine their important (Fig. S1, Supporting Information methods).

Compound Name	**PubChem CID**	α	β	π^*	**Experimental Activity Classification**
1,3-butanediol	7966	0.66	0.84	0.45	0
1,5-pentanediol	244	0.6	0.52	0.98	0
1-butanol	263	0.84	0.84	0.47	1
1-heptanol	8129	0.33	0.45	0.4	0
1-hexanol	8103	0.8	0.84	0.4	1
1-octanol	957	0.77	0.81	0.4	0
1-pentanol	6276	0.84	0.86	0.4	1
1-phenylethanol	8892	1.22	0.45	0.52	0
1-propanol	1031	0.84	0.9	0.52	1
2-phenylethanol	11005	0.55	0.45	0.36	0
2,2,2-trifluoroethanol	6409	1.51	0	0.73	0
Acetic acid	176	1.12	0.45	0.64	0
Allyl alcohol	7858	0.84	0.9	0.52	0
Benzyl alcohol	996	1.65	0.3	0.72	0
Benzyl alcohol	244	0.6	0.52	0.98	0
Cyclohexanol	342	1.13	0.34	0.68	0
Ethanol	702	0.86	0.75	0.54	1
Ethylene glycol	174	0.9	0.52	0.92	0
Ethylene glycol	174	0.9	0.52	0.92	0
Hexafluoro isopropyl alcohol	13529	1.96	0	0.65	0

Compound Name	PubChem CID	α	β	π^*	Experimental Activity Classification
Hexanoic acid	8892	1.22	0.45	0.52	0
Isoamyl alcohol	31260	0.84	0.86	0.4	1
Isobutanol	6560	0.79	0.84	0.4	0
Isopropyl alcohol	3776	0.76	0.84	0.48	0
Lactic acid	612	-	-	-	0
Methanol	887	0.98	0.66	0.6	1
Phenol	6054	0.64	0.61	0.88	0
Propanethiol	7848	-	-	-	0
t-butyl alcohol	6386	0.42	0.93	0.41	0
Water	962	1.17	0.47	1.09	0

Table S5. The final input data for the activity classification model. Here, Kamlet-Taft (α, β) parameters obtained from experimental reports in the literature[16–18] are given for a larger set of compounds and the corresponding experimental activity classification obtained in the current work. For compounds for which experimentally measured KT parameters are not known, the values predicted from the deep-learning. The set of compounds here include compounds that were initially tested and compounds which were suggested by the data-driven approach for testing. Note that experimental values for KT parameters for some compounds were not known, so predicted values from the deep learning model were given in italics for these compounds.

Compound Name	PubChem CID	α	β	Experimental Activity Classification
1,2-propanediol	1030	-	-	0
1,3-butanediol	7966	0.66	0.84	0
1,3-butanediol	7896	-	-	1
1,3-propanediol	10442	-	-	1
1,5-pentanediol	244	0.6	0.52	0
1-butanol	263	0.84	0.84	1
1-decanol	8174	0.7	0.82	0

Compound Name	PubChem CID	α	β	Experimental Activity Classification
1-heptanol	8129	0.33	0.45	0
1-hexanol	8103	0.8	0.84	1
1-nonanol	8914	-	-	0
1-octanol	957	0.77	0.81	0
1-pentanol	6276	0.84	0.86	1
1-phenylethanol	8892	1.22	0.45	0
1-propanol	1031	0.84	0.9	1
2-butanol	6568	0.69	0.8	1
2-chloroethanol	34	1.28	0.53	0
2-methoxyethanol	8019	-	-	0
2-phenylethanol	11005	0.55	0.45	0
2,2,2-trifluoroethanol	6409	1.51	0	0
3-butene-1-ol	69389	-	-	1
Acetic acid	176	1.12	0.45	0
Allyl alcohol	7858	0.84	0.9	0
Benzyl alcohol	996	1.65	0.3	0
Benzyl alcohol	244	0.6	0.52	0
Cyclohexanol	342	1.13	0.34	0
Ethanol	702	0.86	0.75	1
Ethylene glycol	174	0.9	0.52	0
Ethylene glycol	174	0.9	0.52	0
Glycerol	753	1.21	0.51	1
Hexafluoro isopropyl alcohol	13529	1.96	0	0
Hexanoic acid	8892	1.22	0.45	0
Isoamyl alcohol	31260	0.84	0.86	1
Isobutanol	6560	0.79	0.84	0

Compound Name	PubChem CID	α	β	Experimental Activity Classification
Isopropyl alcohol	3776	0.76	0.84	0
Lactic acid	612	-	-	0
Methanol	887	0.98	0.66	1
Phenol	6054	0.64	0.61	0
Propanethiol	7848	-	-	0
t-butyl alcohol	6386	0.42	0.93	0
Water	962	1.17	0.47	0

Table S6. Input Data for the Deep Learning Prediction Model Kamlet-Taft (α and β) parameters obtained from experimental reports in the literature.[16–18]

Compound Name	PubChem CID	α	β
1,1,1,3,3,3-hexafluoro-2-propanol	13529	1.96	0
1,1,1-trichloroethane	6278	0	0
1,1,2,2-tetrachloroethane	6591	0	0
1,1,3,3-tetramethylguanidine	66460	0	0.86
1,1-dichloroethane	6365	0.1	0.1
1,2,3-propanetriol	753	1.21	0.51
1,2-diaminoethane	3301	0.13	1.43
1,2-dibromoethane	7839	0	0
1,2-dichlorobenzene	7239	0	0.03
1,2-dichloroethane	11	0	0.1
1,2-dimethoxyethane	8071	0	0.41
1,3,5-trimethylbenzene	7947	0	0.13
1,3-dichlorobenzene	10943	0	0.03
1,3-dimethylbenzene	7929	0	0.11
1,3-dioxolane	12586	0	0.45

Compound Name	PubChem CID	α	β
1,4-difluorobenzene	10892	0	0.03
1,4-dimethylbenzene	7809	0	0.12
1-bromobutane	8002	0	0.13
1-chlorobutane	8005	0	0
1-iodobutane	10962	0	0.23
2,2,2-trifluoroethanol	6409	1.51	0
2,3,4-trifluoronitrobenzene	69871	0	0.24
2,3-difluoronitrobenzene	81335	0	0.26
2,6-dimethylpyridine	7937	0	0.76
2-bromoacetophenone	6259	0	0.45
2-bromopyridine	7973	0	0.53
2-butanol	6568	0.69	0.8
2-butanone	6569	0.06	0.48
2-chloroacetophenone	10757	0	0.45
2-chloroaniline	7240	0.25	0.4
2-chlorobenzaldehyde	6996	0	0.4
2-chloroethanol	34	1.28	0.53
2-cyanopyridine	7522	0	0.29
2-decanone	12741	0	0.48
2-fluoroacetophenone	96744	0	0.47
2-fluoronitrobenzene	73895	0	0.28
2-fluoropyridine	9746	0	0.51
2-heptanone	8051	0.05	0.48
2-methyl-1-propanol	6560	0.79	0.84
2-methyl-2-butanol	6405	0.28	0.93
2-methyl-2-propanol	6386	0.42	0.93

Compound Name	PubChem CID	α	β
2-methylnitrobenzene	6944	0	0.3
2-nonanone	13187	0	0.48
2-octanone	8093	0	0.48
2-pentanone	7895	0	0.52
2-phenylacetonitrile	8794	0	0.41
2-phenylethanol	6054	0.64	0.61
2-propanol	3776	0.76	0.84
2-propanone	180	0.08	0.43
2-propen-1-ol	7858	0.84	0.9
2-pyrrolidone	12025	0.36	0.77
3,4,5-trifluoronitrobenzene	2782793	0	0.24
3,4-difluoronitrobenzene	123053	0	0.26
3,4-dimethylpyridine	11417	0	0.78
3-bromoacetophenone	16502	0	0.45
3-bromopyridine	12286	0	0.6
3-chlorobenzaldehyde	11477	0	0.4
3-chlorophenol	7933	1.57	0.23
3-fluoroacetophenone	9967	0	0.47
3-fluoronitrobenzene	9823	0	0.28
3-methyl-1-butanol	31260	0.84	0.86
3-methylphenol	342	1.13	0.34
3-pentanone	7288	0	0.45
3-phenylpropanol	31234	0.53	0.55
3-trifluoromethylnitrobenzene	7386	0	0.25
4-chlorobenzaldehyde	7726	0	0.4
4-fluoronitrobenzene	9590	0	0.28

Compound Name	PubChem CID	α	β
4-methyl-2-oxo-1,3-dioxolane	7924	0	0.4
4-methyl-2-pentanone	7909	0.02	0.48
4-methylphenol	2879	1.64	0.34
4-methylpyridine	7963	0	0.67
acetic acid	176	1.12	0.45
acetic anhydride	7918	0	0.29
acetonitrile	6342	0.19	0.4
acetophenone	7410	0.04	0.49
aniline	6115	0.26	0.5
benzene	241	0	0.1
benzonitrile	7505	0	0.37
benzyl alcohol	244	0.6	0.52
bis(2-chloroethyl) ether	8115	0	0.4
bis(2-methoxyethyl) ether	8150	0	0.4
bromobenzene	7961	0	0.06
butane	7843	0	0
butanenitrile	8008	0	0.4
butanoic acid	264	1.1	0.45
butanol	263	0.84	0.84
butyl acetate	31272	0	0.45
butylamine	8007	0	0.72
carbon disulfide	6348	0	0.07
chloroacetonitrile	7856	0	0.34
chlorobenzene	7964	0	0.07
cis-decalin	7044	0	0.08
cyclohexane	8078	0	0

Compound Name	PubChem CID	α	β
cyclohexanol	7966	0.66	0.84
cyclohexanone	7967	0	0.53
cyclopentanone	8452	0	0.52
decane	15600	0	0
decanol	8174	0.7	0.82
diaminoethane	3301	0.13	1.43
dibenzyl ether	7657	0	0.41
dibromomethane	3024	0	0
dibutyl ether	8909	0	0.46
dichloromethane	6344	0.13	0.1
diethyl carbonate	7766	0	0.4
diethyl ether	3283	0	0.47
diethyl sulfide	9609	0	0.37
diethylamine	8021	0.3	0.7
diethylformamide	12051	0	0.79
diiodomethane	6346	0	0
diisopropyl ether	7914	0	0.49
diisopropyl sulfide	12264	0	0.38
dimethyl carbonate	12021	0	0.43
dimethyl phthalate	8554	0	0.78
dimethyl sulfate	6497	0	0.36
dimethyl sulfide	1068	0	0.34
dimethyl sulfoxide	679	0	0.76
dimethylamine	674	0	0.7
dimethylcyanamide	15112	0	0.64
di-n-butyl sulfide	11002	0	0.38

Compound Name	PubChem CID	α	β
dioxane	31275	0	0.37
dipentylamine	16316	0	0.7
dipentylether	12743	0	0.47
diphenyl ether	7583	0	0.13
dipropyl ether	8114	0	0.46
dodecane	8182	0	0
ethane	6324	0	0
ethanediol	174	0.9	0.52
ethanol	702	0.86	0.75
ethoxybenzene	7674	0	0.3
ethyl acetate	8857	0	0.45
ethyl benzoate	7165	0	0.41
ethyl chloroacetate	7751	0	0.35
ethyl formate	8025	0	0.36
ethyl trichloroacetate	10588	0	0.25
ethylene carbonate	7303	0	0.41
fluorobenzene	10008	0	0.07
formamide	713	0.71	0.48
formic acid	284	1.23	0.38
furan	8029	0	0.14
heptane	8900	0	0
heptanoic acid	8094	1.2	0.45
hexachlorobiphenyl	91635	0	0.03
hexadecane	11006	0	0
hexafluorobenzene	9805	0	0.02
hexamethylphosphoramide	12679	0	1.05

Compound Name	PubChem CID	α	β
hexane	8058	0	0
hexanoic acid	8892	1.22	0.45
hexanol	8103	0.8	0.84
iodobenzene	11575	0	0.06
methane	297	0	0
methanol	887	0.98	0.66
methoxybenzene	7519	0	0.32
methyl acetate	6584	0	0.42
methyl benzoate	7150	0	0.38
methyl formate	7865	0	0.37
methyl propanoate	11124	0	0.27
methylamine	6329	0	0.7
morpholine	8083	0.29	0.7
N,N,N',N'-tetramethylurea	12437	0	0.8
N,N-diethylacetamide	12703	0	0.78
N,N-dimethylacetamide	31374	0	0.76
N,N-dimethylaniline	949	0	0.43
N,N-dimethylbenzylamine	7681	0	0.64
N,N-dimethylcyclohexylamine	7415	0	0.84
N,N-dimethylformamide	6228	0	0.69
n-decylamine	8916	0	0.68
n-heptylamine	8127	0	0.69
nitrobenzene	7416	0	0.3
nitromethane	6375	0.22	0.06
N-methylacetamide	6582	0.47	0.8
N-methylaniline	7515	0.17	0.47

Compound Name	PubChem CID	α	β
N-methylformamide	31254	0.62	0.8
N-methylpyrrolidon	13387	0	0.77
n-nonylamine	16215	0	0.69
n-octylamine	8143	0	0.69
nonane	8141	0	0
octadecafluoronaphthalene	9386	0	-0.05
octane	356	0	0
octanol	957	0.77	0.81
oxane	8894	0	0.54
oxolan-2-one	7302	0	0.49
pentachlorobiphenyl	17348	0	0.06
pentadecane	12391	0	0
pentafluoropyridine	69690	0	0.16
pentane	8003	0	0
pentanoic acid	7991	1.19	0.45
pentanol	6276	0.84	0.86
perfluoro(methylcyclohexane)	9637	0	-0.06
phenol	996	1.65	0.3
piperidine	8082	0	1.04
propane	6334	0	0
propanenitrile	7854	0	0.39
propanoic acid	1032	1.12	0.45
propanol	1031	0.84	0.9
propyl acetate	7997	0	0.4
pyridine	1049	0	0.64
pyrrolidine	31268	0.16	0.7

Compound Name	PubChem CID	α	β
quinoline	7047	0	0.64
styrene	7501	0	0.12
sulfolane	31347	0	0.39
tetrachloroethene	31373	0	0.05
tetrachloromethane	5943	0	0.1
tetradecafluorohexane	9639	0	-0.08
tetradecane	12389	0	0
tetrahydrofuran	8028	0	0.55
tetramethylsilane	6396	0	0.02
thiane	15367	0	0.36
thiolane	1127	0	0.44
thiolane-1-oxide	1128	0	0.81
toluene	1140	0	0.11
trans-1,2-dichloroethene	638186	0	0
tribromomethane	5558	0.05	0.05
tributyl phosphate	31357	0	0.8
tributylamine	7622	0	0.62
trichloroethene	6575	0	0.05
trichloromethane	6212	0.2	0.1
tridecane	12388	0	0
triethyl phosphate	6535	0	0.77
triethylamine	8471	0	0.71
trimethyl phosphate	10541	0	0.77
undecane	14257	0	0
water	962	1.17	0.47

References

(1) Lazouski, N.; Schiffer, Z. J.; Williams, K.; Manthiram, K. Understanding Continuous Lithium-Mediated Electrochemical Nitrogen Reduction. *Joule* **2019**, *3*, 1127–1139.

(2) Quinlan, J. The Morgan Kaufmann series in machine learning. *San Mateo* **1993**,

(3) Breiman, L.; Friedman, J.; Stone, C. J.; Olshen, R. A. *Classification and regression trees*; CRC press, 1984.

(4) Kass, G. V. An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **1980**, *29*, 119–127.

(5) Kim, H.; Loh, W.-Y. Classification trees with unbiased multiway splits. *Journal of the American Statistical Association* **2001**, *96*, 589–604.

(6) Kim, H.; Loh, W.-Y. Classification trees with bivariate linear discriminant node models. *Journal of Computational and Graphical Statistics* **2003**, *12*, 512–530.

(7) Loh, W.-Y. Improving the precision of classification trees. *The Annals of Applied Statistics* **2009**, 1710–1737.

(8) Loh, W.-Y.; Shih, Y.-S. Split selection methods for classification trees. *Statistica sinica* **1997**, 815–840.

(9) Fielding, A.; O'Muircheartaigh, C. Binary segmentation in survey analysis with particular reference to AID. *Journal of the Royal Statistical Society: Series D (The Statistician)* **1977**, *26*, 17–28.

(10) Messenger, R.; Mandell, L. A modal search technique for predictive nominal scale multivariate analysis. *Journal of the American statistical association* **1972**, *67*, 768–772.

(11) Loh, W.-Y. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2011**, *1*, 14–23.

(12) Timofeev, R. Classification and regression trees (CART) theory and applications. *Humboldt University, Berlin* **2004**, 1–40.

(13) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design* **2016**, *30*, 595–608.

(14) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9*, 513–530.

(15) Lazouski, N.; Chung, M.; Williams, K.; Gala, M. L.; Manthiram, K. Non-aqueous gas diffusion electrodes for rapid ammonia synthesis from nitrogen and water-splitting-derived hydrogen. *Nature Catalysis* **2020**, *3*, 463–469.

(16) Marcus, Y. The properties of organic liquids that are relevant to their use as solvating solvents. *Chemical Society Reviews* **1993**, *22*, 409–416.

(17) Meyer, P.; Maurer, G. Correlation and prediction of partition coefficients of organic solutes between water and an organic solvent with a generalized form of the linear solvation energy relationship. *Industrial & engineering chemistry research* **1995**, *34*, 373–381.

(18) Stenutz, R. Kamlet-Taft solvent parameters. `http://www.stenutz.eu/chem/solv26.php`, accessed June 2020.