Electronic Supplementary Material (ESI) for Digital Discovery.

Predicting Ruthenium Catalysed Hydrogenation of Esters using Machine Learning

Challenger Mishra,*[a] Niklas von Wolff,*[b] Abhinav Tripathi,[c] Claire N Brodie,[c] Neil D. Lawrence,[a] Aditya Ravuri,[a] Eric Brémond,[d] Annika Preiss,[c] Amit Kumar*[c]

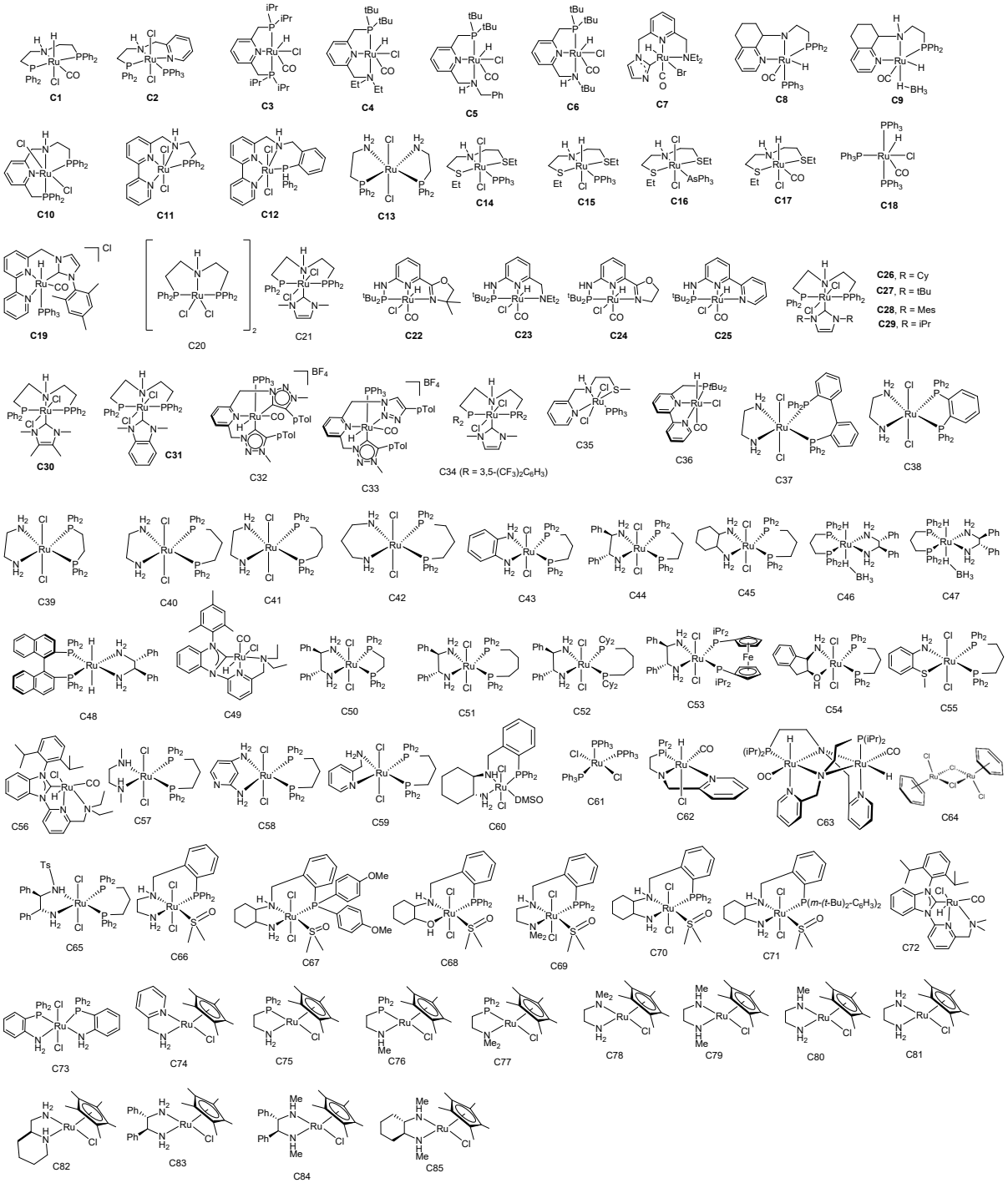
[a]Department of Computer Science and Technology, University of Cambridge, Cambridge, CB30FD, UK;
[b]Laboratoire d'Electrochimie Moléculaire, Université de Paris, CNRS, F-75006 Paris, France;
[c]School of Chemistry, University of St. Andrews, St. Andrews, KY169ST, UK;
[d]ITODYS, Université de Paris, CNRS, F-75006 Paris, France.

Table of Content

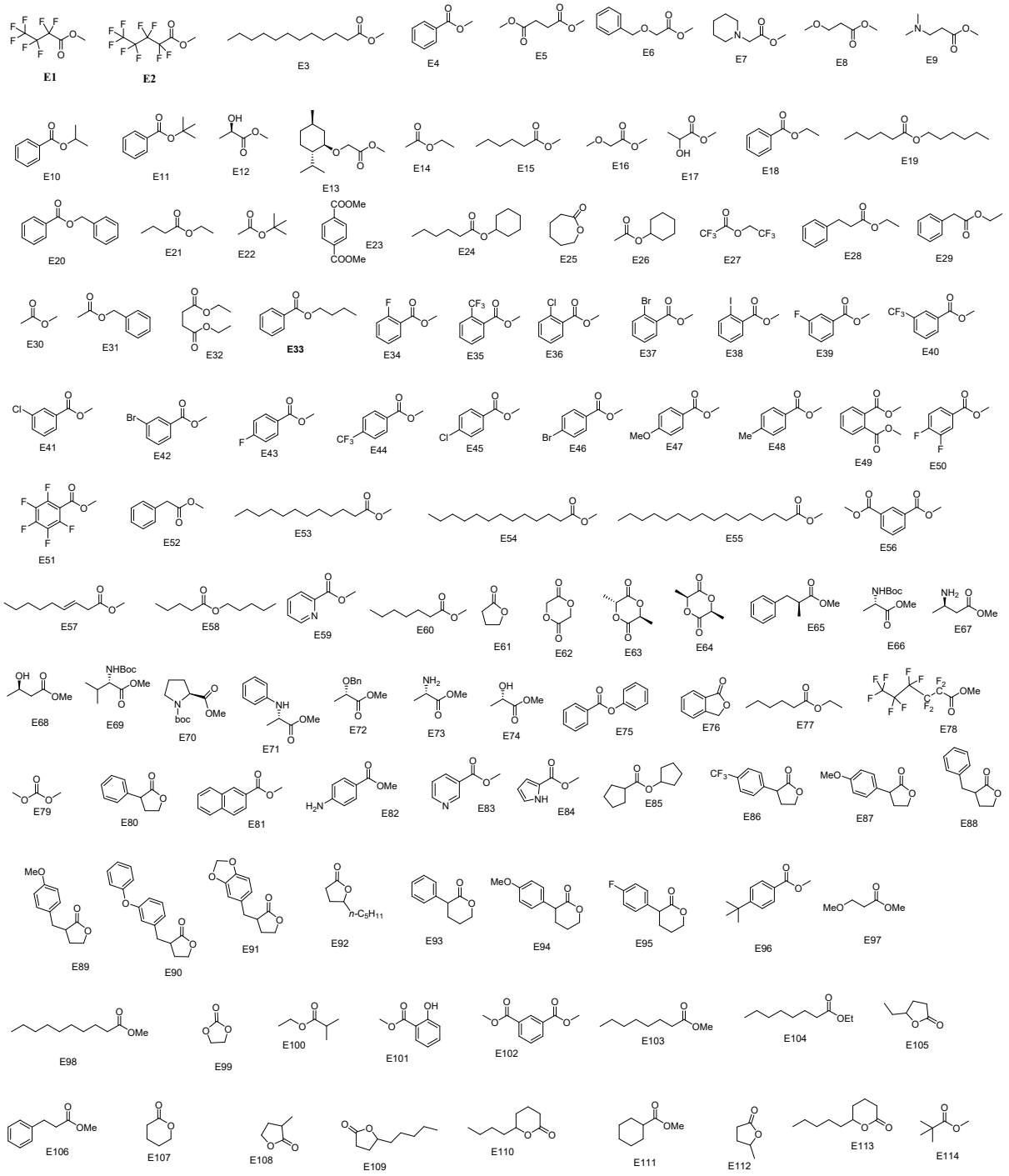
1. Details of the dataset

The data present in the dataset has been collected from the peer-reviewed literature. Each datapoint represents a reaction for the catalytic hydrogenation of an ester and is represented by nine parameters -catalyst structure, ester structure, catalyst amount (mol%), ester amount (mmol), base structure, base amount (mol%), temperature (°C), pressure (bar), solvent structure, and solvent amount. Their details along with corresponding references are given in the dataset file.

1.1 Structures of ruthenium catalysts used in the dataset



1.2 Structures of esters used in the dataset



2. Generation of chemical descriptors

2.1 DFT calculations

All geometries were fully optimized using the M06-L functional,[1] the def2-TZVP basis set[2] and W06 density fitting to improve the computational efficiency,[3] as well as the Grimme's empirical dispersion correction with zero damping.[4] Frequency calculations at the same level of theory confirmed stationary points and transition states. They were used to compute thermodynamic properties at

298.15K, if not stated otherwise. Then, single point energies of the optimized structures were computed using the ωB97M-V[5] range-separated hybrid exchange-correlation functional corrected with the VV10 non-local correlation,[6,7] together with the def2-TZVPP triple-ζ basis set[2] and the RIJCOSX resolution-of-the-identity using the auxiliary basis sets def2/J[3] and def2-TZVPP/C[8] to speed up computational time. The choice of the ωB97M-V functional was rationalized given its excellent results in a recent benchmark study on transition metal reactions.[9] Gibbs Free Energies, unless otherwise stated, were computed by adding the Free Energy correction terms from the frequency calculations to the single point energies at the ωB97M-V/def2-TZVPP level of theory in solution (SMD, THF) according to:

$$G_{SMD}^{\omega BP97M-V} = E_{el/SMD}^{\omega BP97M-V} + corr_{freq/T}^{M06-L},$$

where $corr_{freq/T}^{M06-L}$ is the thermal correction to the Gibbs Free Energy from the frequency calculation at temperature T.[10] All structures were optimized in solution (THF) using the integral equation formalism variant (IEFPCM) of the PCM model in the SMD variation of Truhlar and co-workers,[11] unless otherwise stated. Optimizations and frequency calculations were done using the Gaussian 16 software suite in the B.01 revision (Gaussian 16, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016). Single point energy calculations were performed using the ORCA software (4.1.1 release).[12] NBO charges were computed with the version 7 of the NBO software.

All structures and energies can be obtained free of charge via the ioChem-BD.org online data repository at the following adress:

https://doi.org/10.19061/iochem-bd-6-118

2.2 DFT-derived chemical descriptors

The optimized structures and energies from DFT-calculations were used to generate DFT-derived chemical descriptors, namely:

1) Autocorrelation functions
2) Sterics: Buried volume, Solvent accessible surface area and volume, topographic steric maps (%V_{free} and free volume of the four quadrants: NE, SE, SW, NW) and sterimol parameters.
3) Electronics: CO frequency, CO IR intensity, HOMO-LUMO gap, dipole moment and NBO charge.

Many of these features can be calculated directly from the xyz-files via suitable software packages, namely the Morfeus software package (kjelljorner.github.io/morfeus). Morfeus was thus used to calculate the autocorrelation functions, the buried volume, SASA area/volumes, as well as sterimol

parameters. The topographic map parameters were calculated manually from the xyz-files via the SambVca 2.1 web application tool.[13]

CO frequency, intensities, NBO charges, dipole moments and HOMO-LUMO gaps were obtained directly from the DFT calculations.

All descriptors are found in the corresponding descriptor data file.

2.3 Experimental chemical descriptors

In addition to the DFT-derived chemical descriptors, experimental descriptors were used, when available for all catalyst or substrates. Hence only for the ester substrate experimental descriptors were used, namely, the IR carbonyl stretch, the ^{13}C-NMR shift of the quaternary carbonyl carbon atom and the alpha-alkoxy carbon atom, as well as the ^{1}H-NMR chemical shifts of the alpha-alkoxy protons, where available.

2.4 Solvent and base descriptors

Solvents and bases were described by readily available experimental data: pK_a values, dielectric constants and Gutman donor numbers. If data was not available, reasonable approximations were used instead (marked in yellow in the tables below).

Base	$pK_a(H_2O)$	pK_a (DMSO)
NaOMe	15.5	27.9
KOMe	15.5	27.9
KOtBu	17.0	29.4
NaBH4	30.0*	30.0*
NaOEt	15.9	29.8
NaOtBu	17.0	29.4
LiHBEt3	30.0*	30.0*

***approximate values**

Table S1. pKa of bases

Solvent	Dielectric constant	Gutman DN
MeOH	32.7	19.0
EtOH	24.5	19.2
iPrOH	17.9	19.8
THF	7.58	20.0
Dioxane	2.25	14.3
Toluene	2.38	0.1
DCM	8.93	1.0
MeTHF	6.97	12.0
E4	6.60	15.0
E14	6.02	17.1
E3	3.86	16.0*
E22	1.97	35.9

* approximate values

Table S2. Dielectric constant and Gutman Donor Number of various solvents.

2.5 Generation of Fingerprints (RDKit, Morgan, MCCAS) for esters, bases, and solvents

For the generation of fingerprints, the SMILES representation of esters, bases and solvents was used together with the RDKit python library[14] (https://rdkit.org/) using the CHEM module.

3. Results from the ML studies

3.1 Models and Methodologies

3.1.1. Linear models are a class of simple predictive models where the underlying relationship between inputs and outputs is assumed to be linear. In our experiments for predicting yield, we deploy *linear regression* and linear kernels for Gaussian Processes (which we describe in the following). We also employ linear models in predicting specific properties of catalysts in a given chemical reaction. We now proceed to describe our NN and GP models and the respective methodologies in making inference using them.

3.1.2. Neural Networks are universal function approximators, meaning, simple NN architectures can approximate a reasonably well-behaved function to an arbitrary degree of accuracy. Although NNs are good at making end-to-end predictions, and have thus found a plethora of applications, they are prone to overfitting. The size of our dataset of hydrogenation reactions is rather small in comparison

to the large number of parameters of a typical neural network. Therefore, particular care must be taken in employing NNs to predict yield and catalyst properties.

To start, we considered a class of simple *multilayer perceptron (MLP)* models. We specify them by the number of hidden layers, number of nodes in each layer, and activation functions in each layer. The training of the network effectively amounts to solving an optimization problem in a high-dimensional space of the networks parameters and is therefore a function of the optimization procedure. To further mitigate overfitting, we used dropout, which is a mechanism to randomly switch off hidden nodes with some probability *p* (*dropout probability*) during the network's training. This has been shown to yield more generalizable models. An infinite number of such models and optimization mechanisms are possible. Collectively, these are referred to as *hyperparameters* of the neural network.

For model selection, that is, to select ideal hyperparameters for our predictive models, we perform a *k-fold cross validation* with respect to the data we use for training (that is, the *training set*). This works by splitting our training set into k-folds and training on (k-1)-folds and validating on the remaining fold (the *validation set*). Model selection proceeds by computing errors on the k possible validation sets. These *cross-validation errors* are a proxy for how well the model generalizes on unseen data (that is, the *test se*t). Following model selection which returns the optimal hyperparameters associated with the architecture and training, we then train on the entire training set to fix the parameters (*weights* and *biases*) of the network, and report *test errors.*

3.1.3. Gaussian Processes are a class of non-parametric models in machine learning. A GP is a collection of random variables such that every finite subset of these has a joint Gaussian distribution. GPs are fully specified by their *mean* and *covariance functions*. The mean function is the expectation of the function (*f*) we are modelling. The covariance function is the covariance between pairs of random variables, that is the covariance of the underlying function values at two different inputs. The covariance function specifies the class of functions in play and is specified by a *kernel k.* It specifies the prior on noisy observations. The mean and covariance functions of a GP are therefore of the form:

$$m(x) = E(f(x)), \quad cov(f(x),f(y)) = k(x,y) + \sigma_n^2 I,$$

where σ_n controls the noise in the observations. In this work, we consider several kernels, namely, (a) the radial basis function (or RBF; also referred to as the squared exponential, or the exponentiated quadratic), (b) the Matérn kernel, (c) the Rational Quadratic kernel, and (d) a linear kernel. The linear kernel is special in this list, in the sense that it is *non-stationary* as its values depend on the actual input values as opposed to their difference. Below we write down the RBF kernel for one dimensional inputs x and y:

$$k_{RBF}(x,y) = \sigma^2 exp\left(-\frac{1}{2}|x-y|^2\right),$$

where σ is an overall scale. The above kernel generalizes to multiple input dimensions (corresponding, for example, to the different chemical descriptors) as

$$k_{RBF}(x,y) = \sigma^2 exp\left(-\frac{1}{2}(x-y)^T\Lambda^{-1}(x-y)\right),$$

where the diagonal matrix Λ contains the square of the length scales (λ_i) corresponding to the different input features. Different kernels can be combined, and the collective set of parameters associated with the combination is the set of hyperparameters for a GP model. For model selection, that is, to learn such hyperparameters from the data, one can either use the *marginal likelihood* or use cross-validation methods. We employ the marginal likelihood which is the probability of the data given the model. At the optimal value of the hyperparameters of the covariance function, the marginal likelihood achieves a trade-off between model complexity and the model fit.

Inference using a GP model is probabilistic, namely, the prediction itself is specified by a mean and variance of a Gaussian distribution. This contrasts with the neural network models we described above, where the predictions are point-predictions as opposed to Gaussian distributions. We further note that a GP does not scale well with dataset size, as it involves inverting matrices that are the size of the dataset. This is to our advantage here since our dataset is of a relatively small size. We now proceed to detail our experiments.

3.1.4 We also compared the outcomes of GP and NN with other models such as Decision Tree, Random Forest and K-Nearest Neighbour.

The baseline models (Decision Tree, Random Forest and K-nearest neighbors) were run on the whole dataset containing all features (Sterics, Autocorrelation, Electronics of esters and catalyst i.e. dataset DMAT_8_8). The training and testing was done on 100 random splits of training (70%) and testing (30%). The results are compared with GP and NN models which have been run on the similarly on DMAT_8_8 dataset with 100 random splits of training (70%) and testing(30%).

3.2 Experiments

In the experiments that follow we always use a train-test split of 70-30. A significant part of the dataset we have compiled comprises of cases where the experimental yield is relatively high. Low yield chemical reactions, although contain valuable information, may not be reported in the literature due a range of factors. This limits the predictive capabilities of machine learning as discussed in the main manuscript. To mitigate this, we conduct separate experiments on the full dataset of reactions as well as on those with reported yields greater than 0.2, and 0.5. In this sense, we propose different ML models when considering different regimes of yield.

The chemical descriptors for the catalyst and the ester molecules are highlighted in Figure 2 (main manuscript). For this work, we choose to work with autocorrelations (A), steric (S), and electronic (E) properties. To find importance of these various categories of descriptors, we also use their combinations. We also conduct experiments where we do not use any descriptors of either catalysts or esters, and instead encode them as unit vectors of size given by the number of esters or catalysts respectively in the dataset. We call such representations *one-hot* encoded.

To simplify a reading of what follows, we introduce the following notation to label all our experiments

$$E_{yield\_cut-off}\,[cat\_descriptors, ester\_descriptors, method]$$

In this notation, we can denote a GP experiment on data involving yields > 0.5 (equivalently 50%), using catalytic descriptors and descriptors for esters. The design matrices in the dataset have nomenclature 'DMAT_catalyst descriptor index_ ester descriptor index'. The indices for both catalyst and esters vary from 1 to 8. The indices have the following meaning:

Index	Corresponding Descriptors
1	One hot encoding
2	Autocorrelation
3	Sterics
4	Electronics
5	Autocorrelation and Sterics
6	Autocorrelation and Electronics
7	Sterics and Electronics
8	All descriptors

Table S3 : Descriptor indices in design matrices

For example, a design matrix – *DMAT_2_5* – would have Autocorrelation (index = 2) descriptors for catalyst, Autocorrelation and Electronics descriptors (index = 5) for esters.

Since there are three categories of descriptors for both catalysts and esters, there are a total of (2^{3x2})=64 combinations of feature categories for each yield cut-off. Although we conduct experiments on all these combinations, for the sake of brevity, we will only highlight our results for select combinations. These will include cases where we use the same descriptor combinations for both esters and catalysts.

All experiments involving neural networks were done on Mathematica 12. The GP and linear regression models were built using GPy.[15] All computations were carried out on a standard laptop computer with a 2.3 GHz 8-Core Intel Core i9 processor.

3.2.1. Predicting yield

To model the experimental yield of our class of reactions, we employ a linear model, a multi-layer perceptron model, and a Gaussian Process. We begin by presenting our neural network results. For each experimental configuration of the type:

$E_{yield\_cut-off}\,[cat\_descriptors, ester\_descriptors, NN]$

We perform a 5-fold cross-validation on the training set over the hyperparameters of activation function, number of hidden layers, number of nodes in each layer, and dropout amounts (p). These are tabulated in Table S4. Training is done using the ADAM optimizer. We find the best architectures for each yield cut-off $\gamma \in \{0, 0.2, 0.5\}$ and combination of chemical descriptors. We list the best architectures found in Table S5, while noting that in each case there were other architectures with similar cross-validation errors. We also note that several other combinations of descriptors are possible (as discussed above) and we have left them out for clarity.

Hyperparameter	Values
Activations (act)	ReLU, tanh, logistic sigmoid (log-sig)
# Hidden layers (N_H)	2, 3, 4
# Nodes in each hidden layer (n_H)	10, 20, 50, 200

Dropout amounts (p)	0, 0.1, 0.2, 0.5

Table S4. *Neural network hyperparameters for cross-validation.*

Descriptors	Optimal Hyperparameters: (act, $N_{H,}$ $n_{H,}$ p)		
	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.5$
None	log-sig, 3, 200, 0	tanh, 2, 200, 0.1	tanh, 2, 50, 0.2
Autocorrelation	log-sig, 4, 200, 0	tanh, 4, 50, 0.2	log-sig, 2, 200, 0.2
Steric	ReLU, 4, 20, 0.5	ReLU, 2, 20, 0.1	log-sig, 3, 200, 0
Electronic	ReLU, 3, 50, 0	log-sig, 3, 200, 0.1	ReLU, 2, 50, 0.2
Autocorrelation, Steric	ReLU, 4, 200, 0.5	ReLU, 2, 10, 0.1	ReLU, 2, 20, 0.5
Autocorrelation, Electronic	log-sig, 2, 200, 0	log-sig, 2, 50, 0	tanh, 3, 200, 0.5
Steric, Electronic	ReLU, 4, 200, 0.5	tanh, 4, 50, 0	tanh, 4, 200, 0.5
Autocorrelations, Steric, Electronic	ReLU, 3, 200, 0.2	ReLU, 4, 50, 0.1	log-sig, 3, 10, 0.2

Table S5*. Optimal neural network architectures found using 5-fold cross-validation for yield cut-off* $\gamma \in \{0,0.2,0.5\}$*. The first column lists the chemical descriptors used for both catalysts and esters.*

For the case of GPs, we use marginal likelihood to select hyperparameters of the kernels we use. In our experiments, we employ the RBF, Matérn, Rational Quadratic, and linear kernels. We employ automatic relevance determination (ARD) distance measures. The Gaussian noise variance and variance of the kernel (Mat52 variance, RatQuad variance, Linear Variance) of optimized GP models give a general indication of learning process. Gaussian noise variance parameter attributes noise in the data set which is expected in the dataset presented hereby. The kernel variance indicates how much the covariance matrix generated by the kernel rule contributes to the learning. The Gaussian noise variance and kernel variance are presented in Table S6. We present the RMSE test errors in Table S7 with different kernels.

The architectures for NNs and kernels for GPs were chosen to be the ones with the best cross-validation errors, and marginal likelihoods respectively.

Descriptors	Kernels	Optimal Hyperparameters (Kernel variance [KV], Gaussian variance [GV])					
		$\gamma = 0.0$		$\gamma = 0.2$		$\gamma = 0.5$	
		KV	GV	KV	GV	KV	GV
One hot encoding	Mat52	0.93	0.06	0.88	0.18	0.73	0.23
	Rat Q	1.02	0.04	1.2	0.11	0.96	0.19
	Linear	~0	1	~0	0.99	~0	0.99
Autocorrelation	Mat	0.91	0.09	0.61	0.44	0.44	0.54
	Rat Q	5.12	0.15	0.70	0.47	0.54	0.53
	Linear	2.8 e^-8	0.92	3.2e^-9	0.99	~0	0.87
Steric	Mat	0.96	0.14	0.78	0.39	0.74	0.39
	Rat Q	1.09	0.09	0.91	0.37	1.05	0.31
	Linear	~0	0.99	1.4e^-14	1.0	~0	1.0
Electronic	Mat	0.24	0.68	0.22	0.75	0.39	0.58
	Rat Q	1.25	0.16	0.57	0.47	1.97	0.31
	Linear	~0	1.00	7.9e^-22	0.99	2.7e^-5	0.88
Autocorrelation, Steric	Mat	0.91	0.10	0.54	0.46	0.42	0.55
	Rat Q	1.67	0.12	0.51	0.50	0.51	0.53
	Linear	2.39e^-8	0.93	~0	0.99	4.19e^-9	0.96
Autocorrelation, Electronic	Mat	0.89	0.11	0.55	0.47	0.44	0.54
	Rat Q	4.9	0.18	0.60	0.42	0.57	0.50

	Linear	2.64e^-8	0.93	~0	0.99	8.61e^-9	0.95
Steric, Electronic	Mat	0.89	0.10	0.49	0.51	0.44	0.54
	Rat Q	88.4	0.08	0.60	0.58	1.38	0.37
	Linear	5.7e^-7	0.97	~0	1.0	2.1e^-5	0.86
Autocorrelations, Steric, Electronic	Mat	0.89	0.10	0.55	0.46	0.44	0.54
	Rat Q	3.04	0.14	0.59	0.44	0.49	0.47
	Linear	2.2e^-8	0.94	~0	1.00	5.837e^-9	0.96

Table S6*. Gaussian noise variance and Kernel variance of optimized GP models with different kernels.*

Descriptors	Test RMSE errors(%)								
	$\gamma = 0$			$\gamma = 0.2$			$\gamma = 0.5$		
	Mat52	Linear	RatQuad	Mat52	Linear	RatQuad	Mat52	Linear	RatQuad
One Hot Encoding	23.9	28.8	26.1	18.9	20.9	20.1	12.7	12.6	12.5
Autocorrelation	28.4	29.9	25.9	20.8	20.8	20.1	12.9	12.7	12.5
Steric	26.8	28.8	23.7	20.7	20.8	19.3	12.5	12.6	12.8
Electronic	27.4	28.8	24.9	20.6	20.8	19.5	12.7	12.9	12.9
Autocorrelation, Steric	26.3	28.8	26.8	20.9	20.8	20.7	13.0	12.7	12.5
Autocorrelation, Electronic	28.4	29.9	25.2	20.8	20.8	20.8	12.9	12.6	12.6
Steric, Electronic	26.4	29.9	27.5	20.8	20.8	19.4	13	12.6	12.6
Autocorrelations, Steric, Electronic	26.5	29.9	25.8	20.8	20.8	20.7	13	12.7	12.8

Table S7*. Test RMSE errors for GP models for a variety of kernels. For each kernel we list above, we also add a white noise kernel. We use cut-off yield* $\gamma \in \{0, 0.2, 0.5\}$*. The first column lists the chemical descriptors used for both catalysts and esters.*

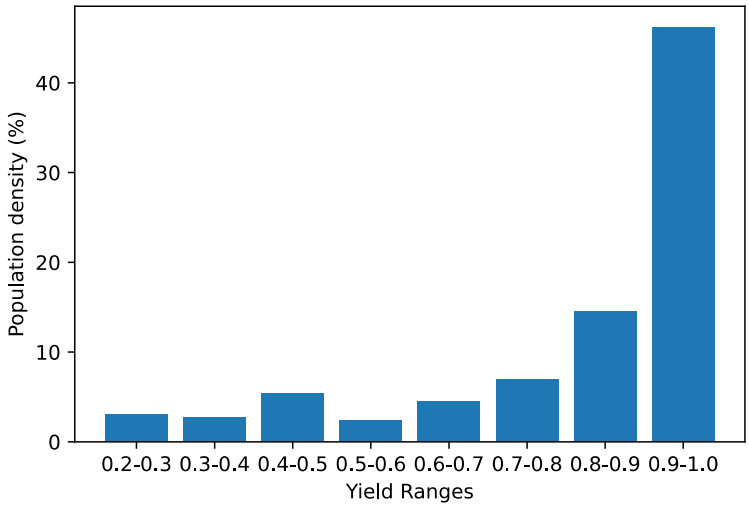


Figure S1. Plot of population density (%) against yield range in the dataset.

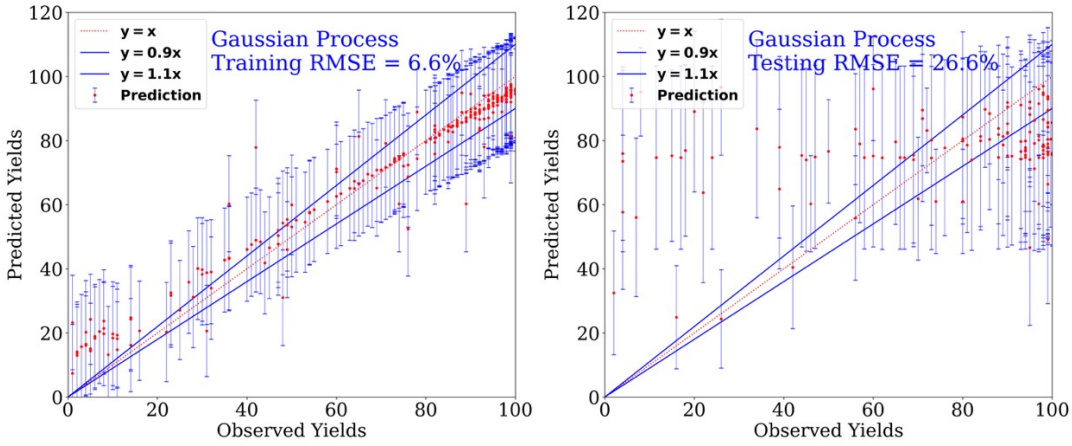


Figure S2. Prediction of yields for the catalytic hydrogenation of esters using GP architecture for the complete dataset.

To check the stability and generalizability, the model was used for 100 different test-train random splits. The variation of test and train data (Figure S3) shows that the model is able to generalize, and the issue of overtraining is not as severe as expected from a small dataset.

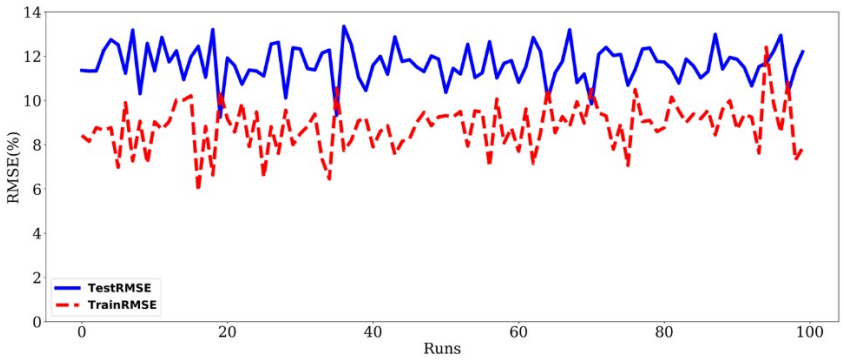


Figure S3. Plot of RMSE for training and test set for 100 random runs.

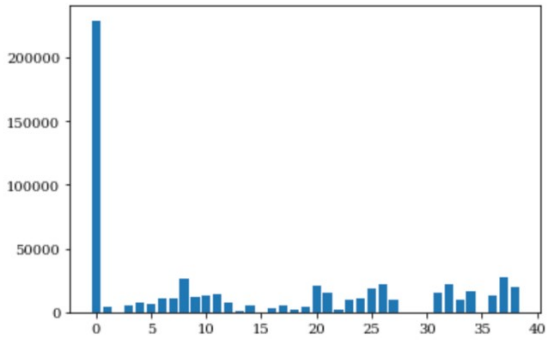


Figure S4. Length-scales of different features using Matern52 kernel for prediction of yields for the catalytic hydrogenation of esters using GP architecture for yields>50%.

Comparison of errors in the prediction of yields using various models

Yield > 0. 0

Model	Train RMSE (%)	Test RMSE (%)	Train R2	Test R2
Gaussian Processes	15.35 ± 2.12	25.46 ± 1.66	0.83± 0.14	0.19±0.10
Decision Trees	26.28 ± 3.46	31.49 ± 2.62	-0.07± 0.54	-1.49±1.27
Random Forest	10.73 ± 0.23	23.02 ± 0.50	0.82± 0.01	-1.25± 0.79
K Nearest Neighbours	24.44 ± 0.39	30.02 ± 3.5	-2.42± 0.53	-3.42±0.80
Mean Model	29.22 ± 0.70	29.58 ± 1.61	0.00 ± 0.00	-0.01 ± 0.018

Yield > 50%

Model	Train RMSE (%)	Test RMSE (%)	Train R2	Test R2
Gaussian Processes	8.55 ± 1.05	11.76 ± 0.84	0.57± 0.08	0.09±0.09
Decision Trees	11.55 ± 0.31	12.02 ± 0.81	-8.55± 2.24	-9/75±4.45
Random Forest	4.18± 0.01	13.60 ± 0.01	0.75± 0.02	-2.26± 0.80
K Nearest Neighbours	11.50 ± 0.39	12.04 ± 0.95	-10.47± 2.26	-12.34±3.78
Mean Model	12.11 ± 0.35	12.38 ± 0.79	0.00 ± 0.00	-0.02 ± 0.022

Table S8: Comparison of errors in the prediction of yields using various models for yields >0 and >50%.

3.3 Optimisation of models using various descriptors

3.3.1 Results from using combination of descriptors: In this study, models (using GPy) were optimised using various combination of descriptors of catalysts and esters (autocorrelation, sterics, electronics, and one-hot encoding). The use of Matern52 kernel showed that a combination of autocorrelation and steric descriptors for esters, and autocorrelation descriptors of catalysts result in the most accurate predictive models for yield prediction (Figure 4 of the main paper). However, when a Rational Quadratic kernel is employed, the best results vary marginally from the Matern52 case and are obtained when using autocorrelation parameters of esters and catalysts (Figure S4). These results

suggest that the predictions on the importance of descriptors are dependent on the choice of kernels we use.

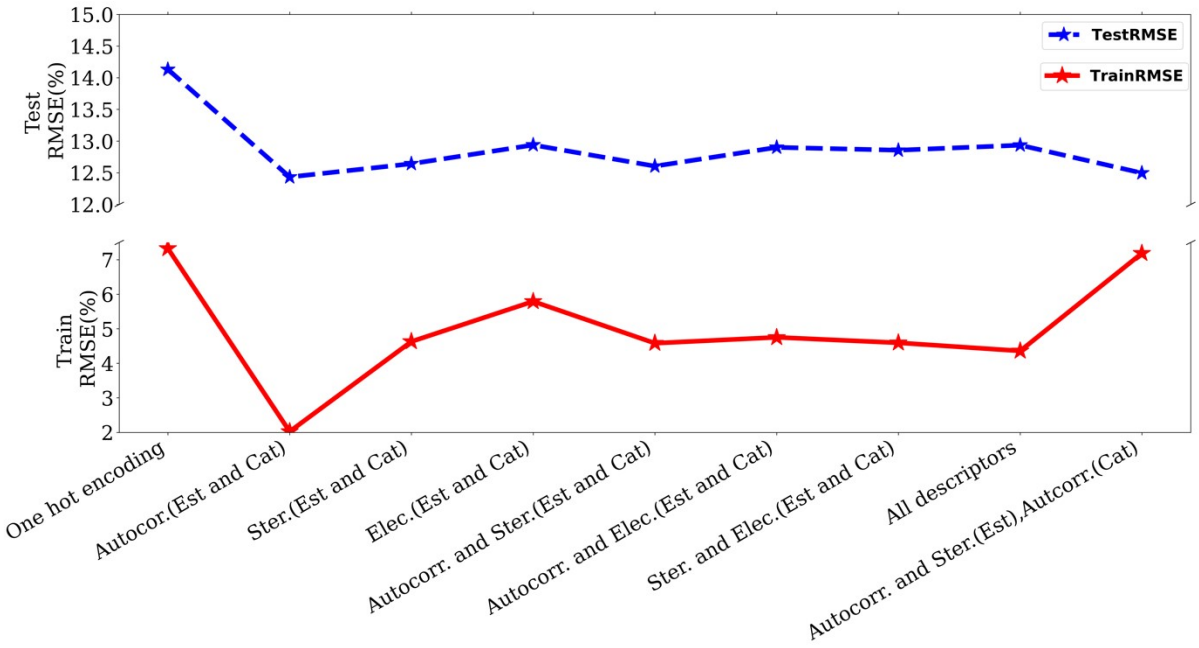


Figure S5. Test and Training RMSE errors with different sets of descriptors. (Autocorr.: Autocorrelation, Ster.: Sterics, Elec: Electronics, Est.: Esters, Cat.: Catalyst).

3.3.2 Use of SMILES as descriptors for catalysts and esters

Datapoints containing entries 167-197 from the dataset file (see the uploaded dataset file with the ESI) was used for the ML studies using GP with Weisfeiler-Lehman graph kernel. The Figure S5 shows results when SMILES were used as descriptors for catalysts and esters. Poor predictions were obtained when other descriptors were used.

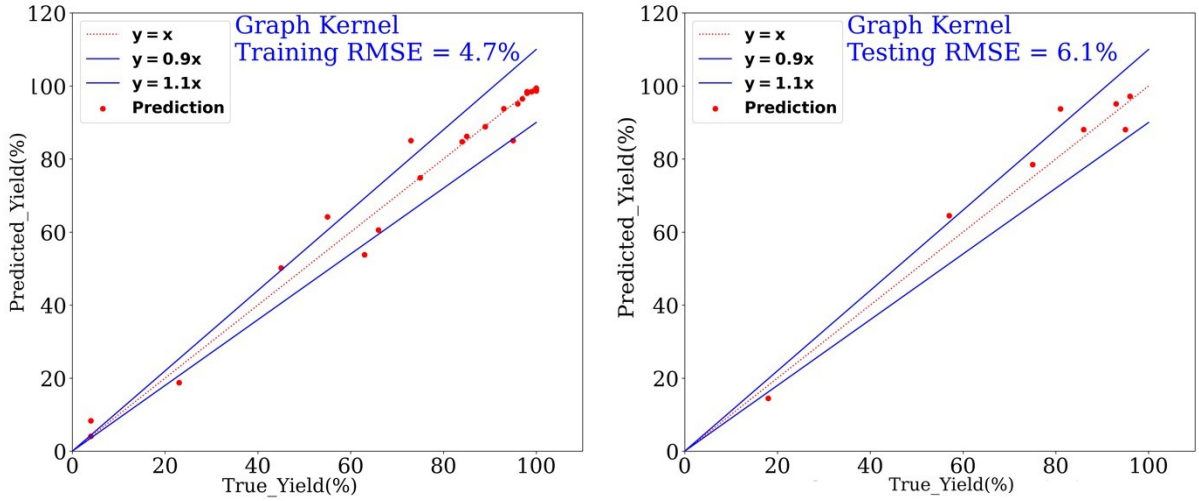


Figure S6. Prediction of yields for the catalytic hydrogenation of esters using GP architecture with Weisfeiler-Lehman graph kernel.

3.3.3 Comparison of error in the prediction of yields using various fingerprints of esters (using GPy with Mat52 kernels):

Yield > 0. 0

Entry	Featurization	Train RMSE (%)	Test RMSE (%)
1	DFT (all descriptors used as described in Table S3)	15.35 ± 2.12	25.46 ± 1.66
2	RDKit (esters+bases represented by RDKit, remaining as entry 1)	11.77 ± 2.97	26.64 ± 1.74
3	Morgan (esters represented by Morgan, remaining as entry 1)	13.64 ± 3.63	26.62± 1.65
4	MCCAS (esters represented by MCCAS, remaining as entry 1)	13.06 ± 3.87	26.70 ± 1.39
5	RDKit (esters+bases+solvents represented by RDKit, remaining as entry 1)	12.32 ± 3.25	27.12 ± 1.50

Yield > 0. 5

Entry	Featurization	Train RMSE (%)	Test RMSE (%)
1	DFT (all descriptors used as described in Table S3)	8.55 ± 1.05	11.76 ± 0.84
2	RDKit (esters+bases represented by RDKit, remaining as entry 1)	9.03 ± 1.75	12.12 ± 1.83
3	Morgan (esters represented by Morgan, remaining as entry 1)	9.09 ± 1.52	12.12 ± 0.75
4	MCCAS (esters represented by MCCAS, remaining as entry 1)	8.95 ± 1.06	11.82 ± 0.83
5	RDKit (esters+bases+solvents represented by RDKit, remaining as entry 1)	8.75 ± 1.62	11.95 ± 0.79

Table S9. Comparison of error in the prediction of yields using various fingerprints of esters

3.3.4 Leaving one feature out experiment

In this experiment (using GPy with Mat52 kernels) one feature was left, and the results (training/testing) was observed. This experiment was performed to evaluate the importance of absence/presence of individual features.

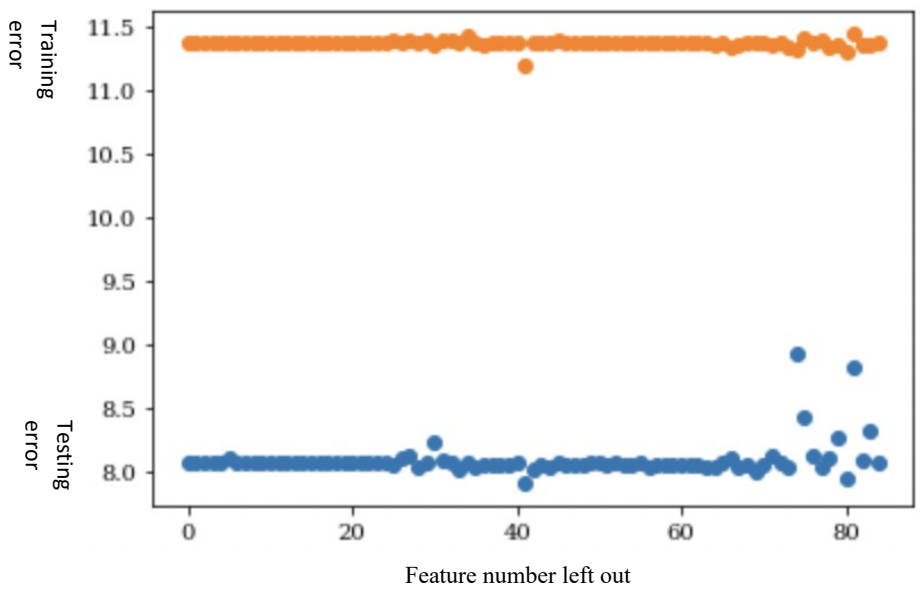


Figure S7. Plot of training and test errors when a particular feature was left out (see Table S10 for feature #).

3.3.5 Importance of individual features

The length scale of trained Gaussian Processes model has been used to determine the relative importance of the features. The ARD (Automatic Relevance Determination) feature of GPy library allows assigning different length-scales to different features. The length scale of various features is given as follows (higher the length scale, lower the feature importance):

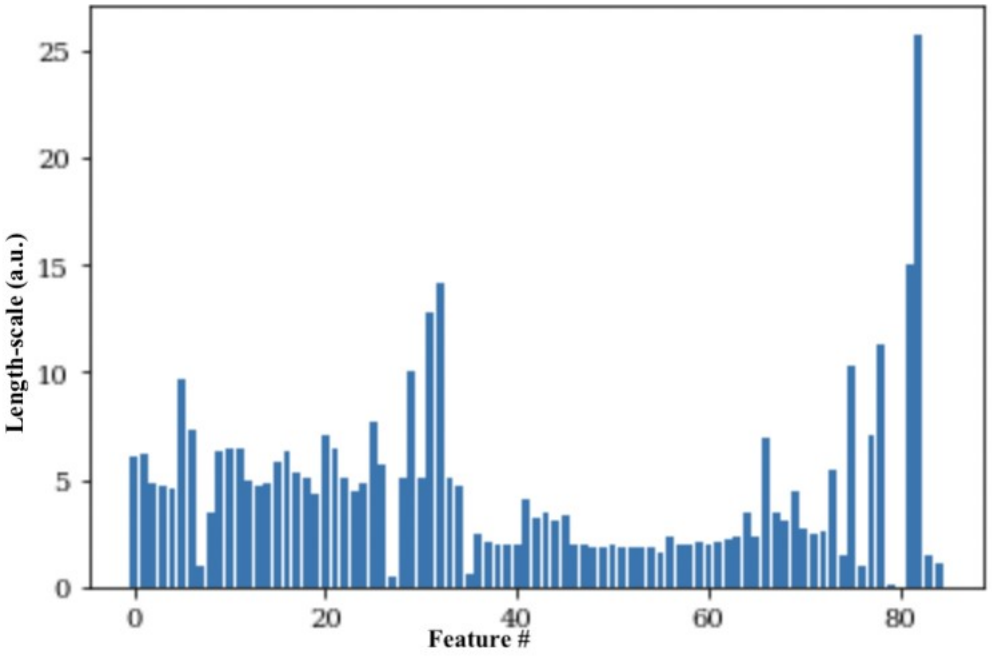


Figure S8. Plot of length scale vs feature number (see Table S10 for feature #).

Feature # :

#	Feature name	Lengthscale
0	Catalyst_f-chi-0-all	5.28
1	Catalyst_f-chi-1-all	5.16
2	Catalyst_f-chi-2-all	4.21
3	Catalyst_f-chi-3-all	3.76
4	Catalyst_f-chi-4-all	4.32
5	Catalyst_f-Z-0-all	5.36
6	Catalyst_f-Z-1-all	4.78
7	Catalyst_f-Z-2-all	3.54
8	Catalyst_f-Z-3-all	2.91
9	Catalyst_f-Z-4-all	3.22
10	Catalyst_f-I-0-all	5.36
11	Catalyst_f-I-1-all	5.15
12	Catalyst_f-I-2-all	4.10
13	Catalyst_f-I-3-all	3.49
14	Catalyst_f-I-4-all	4.44
15	Catalyst_f-T-0-all	4.74
16	Catalyst_f-T-1-all	4.75
17	Catalyst_f-T-2-all	3.94
18	Catalyst_f-T-3-all	3.92
19	Catalyst_f-T-4-all	3.92
20	Catalyst_f-S-0-all	5.47
21	Catalyst_f-S-1-all	5.18
22	Catalyst_f-S-2-all	3.83
23	Catalyst_f-S-3-all	3.32
24	Catalyst_f-S-4-all	3.37
25	Catalyst_buried volume	5.96

26	Catalyst_SW	4.84
27	Catalyst_NW	0.88
28	Catalyst_NE	5.50
29	Catalyst_SE	5.19
30	Catalyst_%Vfree	5.85
31	Catalyst_SASA area	10.22
32	Catalyst_SASA volume	10.73
33	Catalyst_ "HOMO-LUMO gap (kcal/mol)"	5.30
34	Catalyst_ "dipole (debye)"	2.54
35	Catalyst_ "NBO charge"	2.03
36	Ester_f-chi-0-all	1.33
37	Ester_f-chi-1-all	1.53
38	Ester_f-chi-2-all	1.58
39	Ester_f-chi-3-all	1.42
40	Ester_f-chi-4-all	1.29
41	Ester_f-Z-0-all	5.63
42	Ester_f-Z-1-all	3.82
43	Ester_f-Z-2-all	3.89
44	Ester_f-Z-3-all	3.54
45	Ester_f-Z-4-all	5.40
46	Ester_f-I-0-all	1.49
47	Ester_f-I-1-all	1.49
48	Ester_f-I-2-all	1.56
49	Ester_f-I-3-all	1.56
50	Ester_f-I-4-all	1.66
51	Ester_f-T-0-all	1.53
52	Ester_f-T-1-all	1.54
53	Ester_f-T-2-all	1.59
54	Ester_f-T-3-all	1.64

55	Ester_f-T-4-all	1.49
56	Ester_f-S-0-all	2.28
57	Ester_f-S-1-all	1.68
58	Ester_f-S-2-all	1.78
59	Ester_f-S-3-all	1.79
60	Ester_f-S-4-all	1.70
61	Ester_SASA area	1.93
62	Ester_SASA volume	2.01
63	Ester_L	1.94
64	Ester_B1	4.76
65	Ester_B5	2.03
66	Ester_1H (OCH)	4.99
67	Ester_13C (OCH)	2.14
68	Ester_13C (carbonyl)	2.80
69	Ester_IR (carbonyl stretch)	6.11
70	Ester_"freq CO (cm-1)"	1.90
71	Ester_ "Int IR CO"	3.00
72	Ester_ "HOMO-LUMO gap (kcal/mol)"	3.60
73	Ester_ "dipole (debye)"	12.79
74	Ester_mmol of Esters	1.42
75	Ester_catalyst mole	10.50
76	Base_pKa(H2O)	0.45
77	Base_pKa(DMSO)	0.00
78	base mol%	0.44
79	temp (oC)	0.03
80	pressure (bar)	0.12
81	time (h)	11.61
82	Solvent_Dielectric constant	18.67
83	Solvent_Gutman DN	0.69

84	solvent amount (mL)	1.38

Table S10. Feature numbers their description and length scales.

Features with decreasing importance:

mmol of ester
"Catalyst ""dipole (debye)"".1"
time (h)
Solvent Amount
Base pKa(H2O)
"Catalyst ""NBO charge"""
Catalyst f-T-4-all
Catalyst f-S-4-all
Catalyst f-Z-4-all
Catalyst SE
Catalyst f-Z-3-all
Catalyst f-S-0-all
Catalyst f-S-3-all
Catalyst f-S-2-all
Catalyst f-S-1-all
Catalyst f-I-0-all
Catalyst f-chi-0-all
Catalyst f-I-1-all
Catalyst f-chi-1-all
Catalyst f-T-3-all
Catalyst f-I-3-all
Catalyst f-chi-4-all
Catalyst f-I-2-all
Catalyst f-T-0-all
Catalyst f-chi-3-all
Catalyst f-Z-2-all
Solven Dielectric constant
Catalyst f-Z-1-all
Catalyst f-T-1-all
Catalyst f-I-4-all
Catalyst f-Z-0-all
Catalyst f-T-2-all
Catalyst f-chi-2-all
"Ester ""HOMO-LUMO gap (kcal/mol)"""
Ester 13C (carbonyl)
temp (oC)
Solvent Gutman DN
Ester f-T-2-all
Ester f-chi-3-all
Ester B5
Ester f-S-3-all
Ester f-I-3-all
Ester f-chi-4-all
Ester SASA area.1

Ester f-S-2-all
Ester f-T-3-all
Ester SASA volume.1
Ester f-T-1-all
Ester f-I-4-all
Ester f-S-0-all
Ester f-S-1-all
Ester f-I-2-all
Ester f-chi-0-all
Ester f-chi-2-all
Ester f-T-0-all
Catalyst SASA volume
Ester f-T-4-all
Ester f-S-4-all
Ester f-chi-1-all
Ester f-I-1-all
Ester f-I-0-all
Ester L
Catalyst SW
Catalyst SASA area
Ester f-Z-2-all
Ester 1H (OCH)
Catalyst NW
Ester f-Z-1-all
Catalyst buried volume
Ester f-Z-3-all
Ester f-Z-0-all
"Ester ""freq CO (cm-1)"""
Ester f-Z-4-all
Catalyst NE
Ester B1
Ester IR (carbonyl stretch)
Ester 13C (OCH)
"Ester ""Int IR CO"""
Catalyst %Vfree
Base pKa(DMSO)
"Catalyst ""HOMO-LUMO gap (kcal/mol)"".1"
"Ester ""dipole (debye)"""
pressure (bar)
base mol%
catalyst mol%

3.3.6 Studies by the addition of Artificial Random Descriptors

Random features were added to the dataset. Total 170 random features were added one by one and results on testing/training errors were observed.

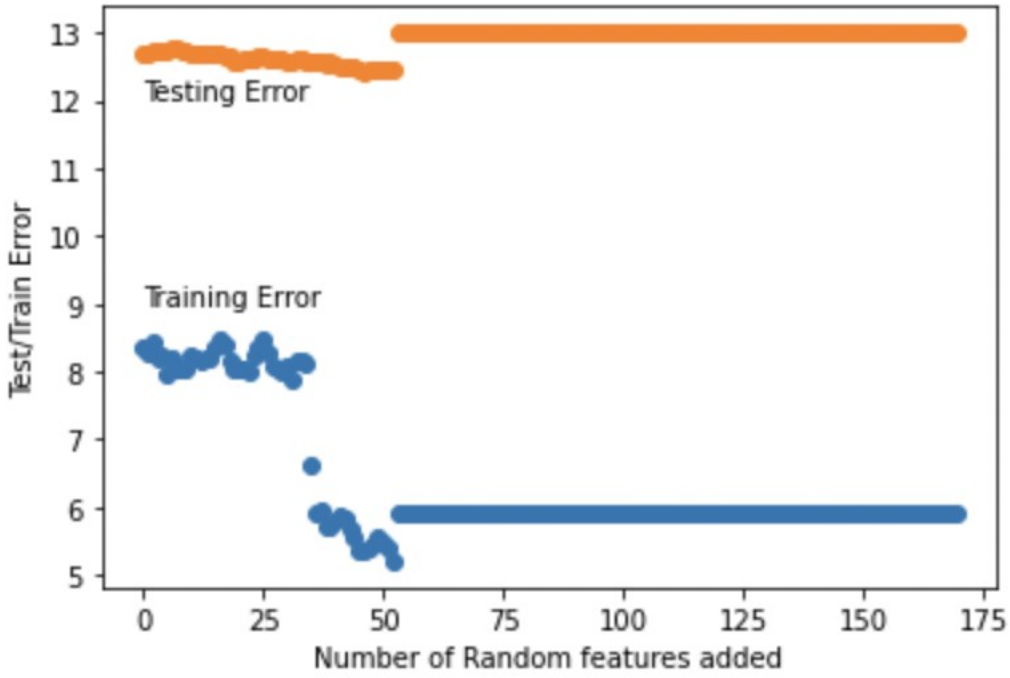


Figure S9. Plot of Test/train errors relative to the number of random features added.

3.3.7 Leave features of one parameter out of experiment

In this case, all the features of certain parameter (catalyst, ester, base, solvent, and experimental conditions) were left out individually while keeping other features, and the model was optimised using GPy with Mat52 kernels. As mentioned below the results are very similar to the case when all descriptors were used (Training RMSE: 7.5%, Testing RMSE: 12.1%, Yield > 0.5, Figure 3, main manuscript).

Yield > 0.0		
Left out features/data	Train RMSE (%)	Test RMSE (%)
Catalyst features (left out)	18.34±1.5	25.43±1.76
Esters features (left out)	11.64±1.10	24.75±2.00
Solvents features (left out)	8.23±2.94	24.57±1.74
Bases features (left out)	10.20±2.47	25.52±2.14
Experimental features (left out)	17.41±1.59	26.44±2.30

Yield > 0.5		
Left out features/data	Train RMSE (%)	Test RMSE (%)
Catalyst features	8.30±0.82	11.86±0.94
Esters features	8.16±0.71	11.55±0.97
Solvents features	6.93±1.17	11.82±0.97
Bases features	6.63±1.41	11.87±1.11
Experimental features	9.42±0.93	12.15±0.87

Table S11. Training and testing RMSE for experiments when one set of features were left out.

3.3.8
Use features of one parameter experiment

In this case, all the features of only one parameter (catalyst, ester, base, solvent, and experimental conditions) was used while other parameters were left out. The model was optimised using GPy with Mat52 kernels. As mentioned below the results are very similar to the case when all descriptors were used (Training RMSE: 7.5%, Testing RMSE: 12.1%, Yield > 0.5 Figure 3, main manuscript).

Yield > 0.0		
Select features/data	Train RMSE (%)	Test RMSE (%)
Catalyst features (only catalyst features were used)	23.43±0.84	27.55±1.61
Esters features	27.26±0.65	28.17±1.5
Solvents features	29.25±0.68	29.46±1.58
Bases features	28.91±0.84	29.65±1.8
Experimental features	21.72±1.47	25.71±1.54

Yield > 0.5		
Select features/data	Train RMSE (%)	Test RMSE (%)
Catalyst features	10.97±0.46	11.95±0.78
Esters features	11.44±0.50	12.31±0.81
Solvents features	12.22±0.36	12.07±0.82
Bases features	11.9±0.46	12.09±0.80
Experimental features	9.01±0.56	11.50±0.76

Table S12. Training and testing RMSE for experiments when only one set of features were used.

3.4 Experiments using balanced dataset

We acknowledge that our dataset has a large number of data points in high yield region in compared to low yield region. We therefore hypothesized that using a more balanced dataset containing datapoints from both high and low yield region (e.g. 30-70% or 40-80%) could lead to a higher accuracy. In case of using a dataset of the yield region 40-80%, the results (Training RMSE: 7.43±3.76 %, Testing RMSE: 12.31±1.06%) were similar to our optimised result (Training RMSE: 7.5%, Testing RMSE: 12.1%, Yield > 0.5). In other cases, the accuracy was worse than this as mentioned in Table S13.

Yield Cutoffs	Training RMSE (%)	Testing RMSE (%)
30-70	3.47±3.97	12.64±1.38
30-80	5.08±3.70	15.22±1.24
30-90	7.38±2.54	17.01±1.49
40-70	3.22±3.35	10.02±1.17
40-80	7.43±3.76	12.31±1.06
40-90	7.74±3.19	14.97±1.32

Table S13. Training and Testing errors in the selected region of the dataset.

We also created a more balanced dataset where 50% of the data was randomly picked from the yield ≤60% and 50% from the yield ≥50%. However, this led to poorer accuracy in the prediction of yield (16.59±6.63(Train), 31.16±3.11(Test)).

3.5 Predicting Catalyst and catalyst properties

In the above we showed how simple ML models can determine the outcome of hydrogenation reactions of ester molecules. In keeping with our broader goal of determining new effective catalysts for hydrogenation reactions, we now proceed to lay out our strategy for solving the inverse problem. To simplify the problem and demonstrate a proof of concept for an ML approach towards catalyst prediction, we turned this into a multi-channel classification problem asking our model to predict a particular catalyst given the reaction conditions and yields from the dataset. Catalysts in this study are represented as unit vectors using one-hot encoding. Architecture details: {activation function->Tanh, training fraction->0.9, number of layers->5, number of nodes->100, yield fraction used->0.5}. Gratifyingly, our model using the NN architecture predicted the corresponding catalysts (one-hot-encoding) with an accuracy of 81% (Figure 5, main manuscript).

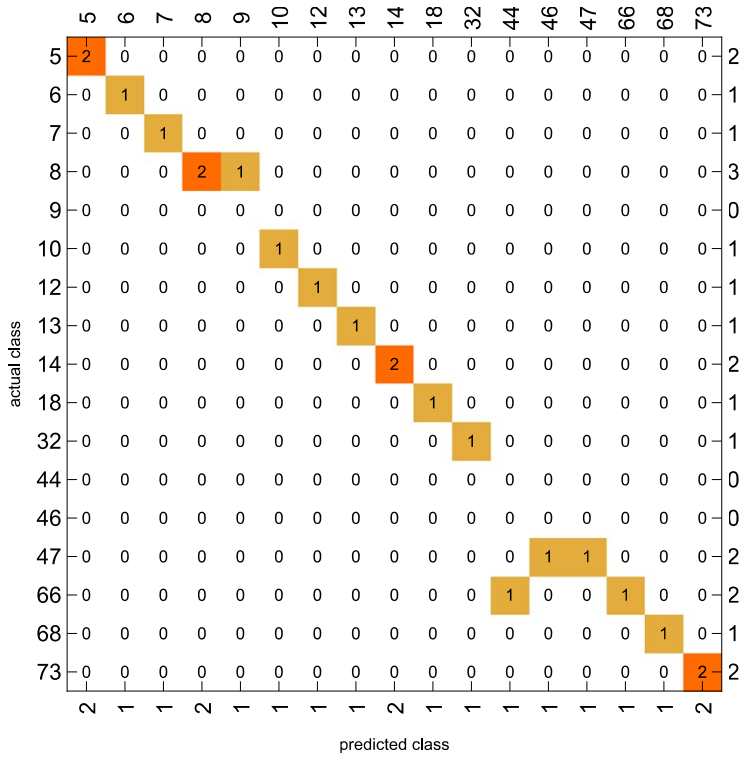


Figure S10A. Confusion matrix for the prediction of catalysts (corresponding to the histogram in Figure 5 of the main manuscript). Left vertical and top horizontal are the catalyst numbers; bottom horizontal and right vertical are sum of columns and rows, respectively.

Additionally, we have predicted various properties of Ruthenium based catalysts using our dataset as a testbed. We formulate the machine learning approach as follows. Using the data set we have acquired, we use the reaction conditions, all chemical descriptors of esters, solvents, and bases as inputs to our ML architectures, to predict steric and electronic descriptors of the catalysts.

We employ a simple MLP architecture and linear regression to predict the various catalyst descriptors. In total we have 8 steric descriptors, and three electronic descriptors. Since neural networks are good at making end-to-end predictions, we attempt to predict all these features simultaneously. We conduct two different sets of experiments. In the first set of experiments [Expt. A], we divide our

dataset into a train-test split of 80-20 and build an MLP model that aims to predict all such catalyst descriptors simultaneously. We compare our outcomes against linear models (realized through linear regression) for each of these features. In the second set of experiments [Expt. B], we divide the dataset into two disjoint parts such that each catalyst features in exactly one of the sets. This is in line with our goal of predicting new catalysts or their properties.

We state at the outset that our initial efforts in this direction have not yielded good results. This is largely in part due to the small amount of data available to us. Secondly, our efforts in this direction would likely benefit from incorporating domain expertise as priors to our ML models, which we have thus far not incorporated. We also note that a GP model for such predictions would likely shed more light, and we intend to return to this in a future publication.

		Buried Volume (Å^3)			SASA area (Å^2)			SASA volume (Å^3)			HOMO-LUMO gap (kcal/mol)		
		$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.5$
Expt. A	Linear	3.5	2.7	2.9	9.7	13.0	10.8	11.9	14.5	11.8	4.3	4.1	3.6
	NN	4.4	4.4	4.7	11.4	13.5	13.1	13.8	16.2	16.2	5.3	6.4	5.5
Expt. B	Linear	3.3	6.5	4.7	16.0	13.1	18.2	24.6	12.1	39.1	4.8	4.2	4.4
	NN	6.4	3.9	4.8	16.5	11.4	26.0	19.6	12.7	34.1	5.3	5.0	8.4

Table S14. *Test errors for our models of catalytic property prediction. Experiments A and B are described in Section 3.2.2.*

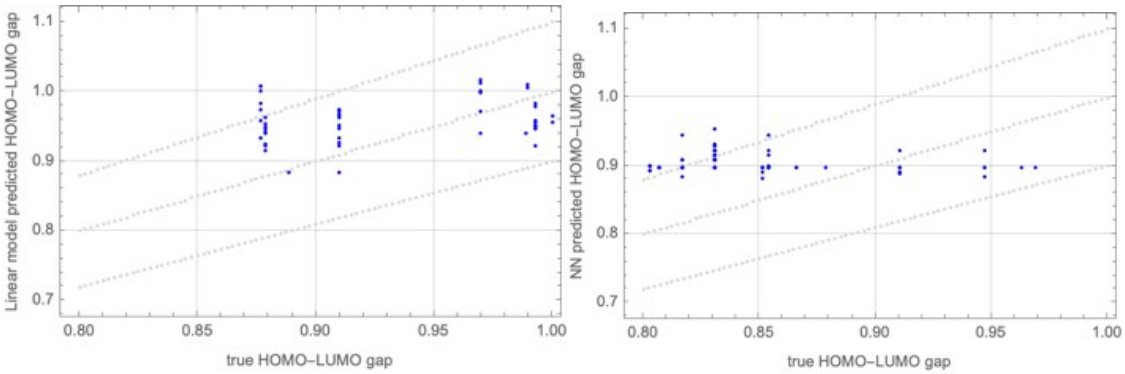


Figure S10. Prediction of HOMO-LUMO gap (unit: kcal/mol) of Ru-catalysts by a (left) linear model and a (right) multi-layer perceptron model. This experimental set up corresponds to Expt. B detailed in Section 3.2.2 for cut-off yield $\gamma = 0.5$. The dashed line in the middle is the line x=y. The other dashed lines reflect a 10% margin.

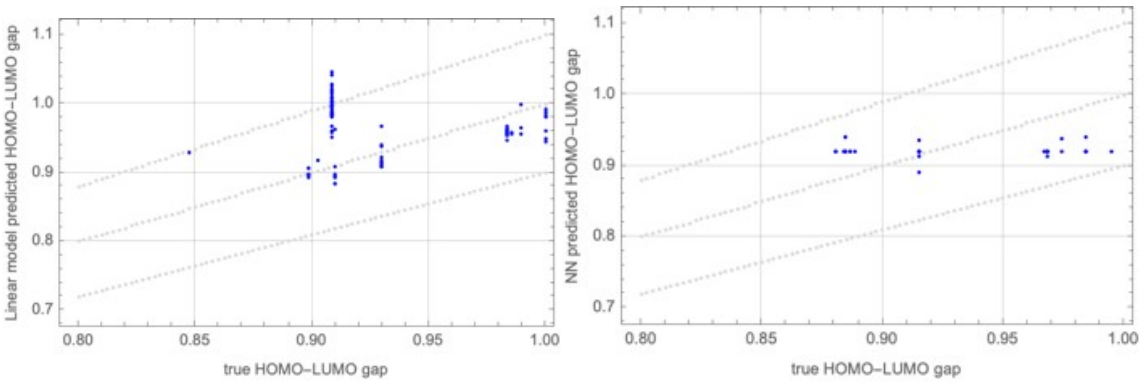


Figure S11. Prediction of buried volume of Ru-catalysts by a (left) linear model and a (right) multi-layer perceptron model. This experimental set up corresponds to Expt. B detailed in Section 3.2.2 (SI) for cut-off yield $\gamma = 0.5$. The dashed line in the middle is the line x=y. The other dashed lines reflect a 10% margin.

3.6 Out of Sample Studies

To mimic the use of the model in real conditions, the out-of-sample test set was constituted of the reaction containing a catalyst or an ester that is not in the training set (as it has been done for the prediction of catalyst properties Expt B, section 3.5). We carried out this study on the complete dataset and the RMSE as described below is similar to those without carrying out the out-of-sample analysis (Table S15). This suggests that our model would have similar accuracy on an unknown data point.

Left out Catalyst/Ester	RMSE_train(%)	RMSE_test(%)	RMSE_val(%)	R^2_train	R^2_test	R^2_val
C1	12.90	26.91	28.52	0.78	0.36	0.11
C2	13.81	25.85	26.80	0.77	0.27	-9.31
C3	11.60	24.91	77.49	0.83	0.34	-1185.18
C4	16.80	21.82	34.17	0.68	0.39	-0.06
C5	16.29	22.20	15.60	0.72	0.32	-1.38
C6	12.26	22.70	13.43	0.84	0.23	
C7	10.73	24.29	23.87	0.88	0.11	-51.01
C8	14.25	24.63	23.16	0.77	0.33	0.01
C9	14.11	24.42	31.30	0.76	0.37	-0.16
C10	13.93	25.60	24.12	0.78	0.24	-13.41
C11	12.49	24.04	32.44	0.83	0.19	
C12	13.73	22.49	42.43	0.78	0.23	-0.20

C13	12.92	24.43	29.61	0.83	0.17	-0.19
C14	15.05	25.05	21.78	0.74	0.25	0.01
C15	13.81	24.97	27.87	0.78	0.25	-20.58
C16	15.22	24.97	15.57	0.73	0.29	-1.99
C17	11.29	23.96	21.72	0.86	0.19	-3.28
C18	12.24	26.66	47.17	0.83	0.10	-0.30
C19	14.99	24.48	29.81	0.75	0.24	0.05
C20	17.46	22.39	33.90	0.65	0.38	-0.03
C21	16.67	26.61	37.26	0.65	0.30	
C22	12.42	22.78	5.32	0.84	0.16	
C23	12.57	22.02	3.82	0.84	0.17	
C24	12.75	23.03	11.32	0.83	0.23	
C25	14.73	25.20	18.48	0.74	0.29	-1.02
C26	14.04	27.03	9.46	0.76	0.23	
C27	9.36	27.32	45.66	0.89	0.24	
C28	13.25	21.48	36.26	0.81	0.32	
C29	13.32	24.02	56.21	0.80	0.24	
C30	10.78	26.34	25.84	0.86	0.24	
C31	12.98	24.37	15.85	0.82	0.17	
C32	17.91	22.48	16.37	0.65	0.33	-0.02
C33	14.73	24.20	16.92	0.75	0.28	
C34	13.63	23.00	41.48	0.79	0.32	
C35	12.51	26.46	14.15	0.80	0.31	-4.56
C36	14.49	25.57	12.79	0.75	0.28	-0.62
C37	14.86	23.35	56.85	0.74	0.37	
C38	16.56	23.16	16.37	0.70	0.27	
C39	14.81	22.26	42.42	0.75	0.36	
C40	15.26	22.93	12.15	0.74	0.30	-589.56
C41	12.58	25.55	25.36	0.81	0.29	

C42	16.20	22.37	37.47	0.70	0.37	-0.03
C43	14.49	24.48	30.45	0.76	0.22	-56.94
C44	15.67	24.85	34.73	0.71	0.33	-0.13
C45	12.97	25.71	39.63	0.81	0.20	
C46	14.87	25.10	38.98	0.74	0.26	-0.07
C47	14.37	24.43	26.46	0.76	0.32	0.33
C48	13.37	25.78	23.40	0.79	0.23	-2.15
C49	14.57	24.63	24.79	0.76	0.24	0.01
C50	9.54	23.74	0.40	0.90	0.30	
C51	15.30	24.83	17.96	0.72	0.32	
C52	14.38	23.81	36.92	0.77	0.24	
C53	14.33	25.23	61.15	0.74	0.37	
C54	12.90	25.28	31.01	0.81	0.21	
C55	17.04	23.31	72.65	0.65	0.38	
C56	13.57	25.15	38.61	0.78	0.28	0.12
C57	13.12	26.69	26.86	0.79	0.21	
C58	14.31	24.39	38.00	0.75	0.35	
C65	12.98	23.10	51.81	0.80	0.40	
C59	14.61	24.06	47.64	0.76	0.24	
C66	15.66	19.96	46.11	0.75	0.23	-0.03
C67	12.50	24.53	17.11	0.82	0.29	
C68	15.81	23.82	16.54	0.71	0.33	0.00
C69	14.37	25.63	40.13	0.76	0.26	
C70	14.42	23.79	28.70	0.76	0.30	-130.77
C71	15.11	24.23	44.21	0.73	0.34	
C72	13.51	24.25	60.23	0.79	0.26	-35.27
C60	15.38	28.32	21.95	0.69	0.26	0.09
C61	14.45	22.45	27.78	0.77	0.27	
C62	17.99	21.94	27.83	0.64	0.37	

C63	14.15	25.12	22.40	0.77	0.28	-3.02
C64	14.34	25.70	78.94	0.76	0.26	0.00
C74	13.00	25.40	10.52	0.81	0.28	-2.85
C75	13.56	23.74	21.43	0.79	0.33	
C76	9.82	25.64	27.47	0.89	0.19	
C77	13.59	27.88	57.75	0.77	0.21	
C78	12.74	24.31	40.75	0.82	0.25	
C79	14.99	24.66	9.29	0.75	0.20	
C80	15.84	25.66	11.49	0.70	0.27	
C81	17.18	21.62	8.93	0.68	0.32	
C82	13.76	25.15	37.36	0.77	0.30	
C83	16.61	27.70	47.47	0.65	0.22	
C84	11.69	25.00	24.10	0.84	0.25	
C85	13.82	24.75	1.34	0.77	0.34	
C73	13.82	24.98	18.67	0.79	0.27	-0.16
E1	10.48	24.35	26.66	0.87	0.38	-0.04
E3	14.34	26.60	24.99	0.76	0.22	-0.40
E4	10.33	25.58	25.24	0.83	0.25	-0.07
E5	15.21	23.24	16.78	0.74	0.33	-3.40
E6	13.89	22.88	25.52	0.78	0.32	0.26
E7	13.72	24.53	21.55	0.77	0.35	
E8	13.33	24.65	49.98	0.80	0.21	-0.35
E9	12.34	23.31	40.47	0.83	0.28	
E10	15.92	26.27	22.25	0.69	0.28	-0.16
E12	14.65	23.83	44.21	0.76	0.31	-65.13
E14	15.76	23.89	13.74	0.70	0.39	-1.49
E15	12.84	24.62	17.88	0.81	0.29	-0.25
E16	17.86	26.15	18.62	0.65	0.10	-1559.43
E17	14.73	23.68	17.38	0.75	0.32	-31.66

E18	14.96	25.18	31.18	0.75	0.22	0.04
E19	12.51	23.56	9.52	0.83	0.27	-0.28
E20	14.55	24.84	12.46	0.76	0.24	-1.46
E21	11.10	25.36	15.99	0.85	0.29	-1022.27
E23	13.72	24.83	11.12	0.78	0.27	-18.80
E24	13.17	24.00	10.03	0.81	0.27	
E25	11.14	28.81	42.14	0.85	0.10	-0.20
E26	13.19	24.67	66.41	0.81	0.17	-1.29
E27	13.98	23.34	61.09	0.77	0.38	-2.07
E28	16.16	24.01	9.66	0.69	0.36	
E29	13.99	26.39	8.46	0.77	0.20	
E30	11.57	21.24	9.57	0.86	0.21	
E31	12.16	26.52	15.27	0.82	0.28	
E32	14.94	24.66	13.14	0.74	0.32	
E33	17.47	25.06	8.40	0.65	0.26	-10.29
E34	11.97	27.61	20.10	0.82	0.27	
E35	15.46	20.92	18.00	0.74	0.40	-6.66
E36	12.83	27.53	9.90	0.80	0.20	0.37
E37	12.98	25.05	14.94	0.80	0.28	-0.06
E38	16.71	22.85	0.31	0.67	0.40	
E39	12.06	22.56	10.09	0.84	0.29	0.08
E40	15.60	26.38	22.84	0.70	0.29	0.10
E41	14.37	24.99	5.77	0.75	0.30	-32.25
E42	12.90	23.77	12.18	0.82	0.25	-8.15
E43	11.28	25.44	46.90	0.83	0.13	-0.22
E44	10.81	27.87	22.57	0.86	0.19	-0.62
E45	14.97	26.90	13.01	0.73	0.26	-2.73
E46	15.20	26.40	11.81	0.70	0.33	-4.04
E47	12.19	22.42	13.66	0.84	0.27	-0.25

E48	13.70	22.07	20.61	0.79	0.36	0.29
E49	13.00	28.48	25.66	0.79	0.19	0.18
E50	14.75	24.98	27.43	0.73	0.36	
E51	14.94	24.88	15.21	0.74	0.28	
E52	12.68	24.74	13.50	0.82	0.26	-1.01
E53	14.29	25.14	12.86	0.76	0.31	-40.36
E54	12.75	23.60	18.30	0.81	0.35	
E55	15.45	24.73	7.50	0.72	0.31	
E56	11.93	28.52	8.90	0.81	0.24	-5.46
E57	14.73	24.94	6.18	0.76	0.22	
E58	16.03	23.04	11.56	0.71	0.30	-533.54
E59	14.33	23.81	53.27	0.78	0.18	-0.34
E112	9.46	27.40	13.10	0.89	0.17	0.37
E60	17.31	23.13	18.95	0.66	0.35	0.00
E61	11.51	23.99	14.52	0.86	0.21	-0.44
E62	13.20	26.05	32.70	0.79	0.28	-0.07
E63	12.35	24.38	7.02	0.83	0.23	
E64	17.74	24.94	7.11	0.65	0.22	-0.33
E65	15.62	23.95	5.61	0.71	0.37	-34.46
E66	14.35	25.96	13.93	0.76	0.23	-0.49
E67	13.35	27.26	85.13	0.78	0.24	
E68	12.68	25.16	77.74	0.82	0.16	
E69	14.15	26.55	8.31	0.74	0.32	
E70	17.17	23.13	3.58	0.69	0.20	
E71	15.38	24.17	1.12	0.74	0.26	
E72	13.08	26.65	8.57	0.80	0.21	-10.74
E73	15.50	23.60	58.74	0.72	0.35	0.00
E74	11.17	26.11	45.81	0.86	0.15	-36.30
E85	11.98	24.28	19.73	0.84	0.25	

E76	13.59	25.21	32.03	0.78	0.28	-0.14
E77	15.14	23.12	23.94	0.74	0.37	
E81	11.72	24.11	12.95	0.84	0.32	-9.49
E82	15.91	25.73	88.82	0.69	0.30	
E83	15.27	25.59	14.39	0.74	0.17	-50.74
E84	11.74	26.30	90.76	0.84	0.23	
E78	10.72	25.62	26.50	0.86	0.29	0.00
E79	13.10	23.01	23.64	0.81	0.28	-0.21
E80	14.22	27.99	1.57	0.73	0.27	
E86	13.84	23.72	0.97	0.78	0.35	
E87	14.05	24.92	2.16	0.76	0.36	
E88	15.50	26.87	7.02	0.70	0.26	
E89	14.35	25.99	5.12	0.75	0.29	
E90	14.23	24.40	12.24	0.77	0.27	
E91	16.04	24.22	4.40	0.69	0.37	
E92	10.57	26.00	5.47	0.87	0.26	
E93	13.24	25.11	50.20	0.77	0.35	-2.31
E94	14.74	23.33	3.44	0.76	0.31	
E95	11.91	25.18	6.95	0.83	0.29	
E106	18.45	24.88	5.87	0.58	0.37	0.64
E111	15.48	24.00	3.17	0.72	0.34	-39.25
E103	15.78	23.37	11.28	0.71	0.38	-9.38
E109	15.75	22.53	7.98	0.74	0.19	0.30
E113	13.43	25.42	9.69	0.79	0.29	-6.67
E114	12.09	27.13	33.55	0.82	0.25	
E100	13.77	26.50	10.92	0.76	0.29	
E101	11.47	26.68	60.43	0.84	0.25	
E102	11.67	24.28	44.78	0.85	0.23	
E104	12.74	23.57	19.70	0.82	0.22	

E105	16.69	26.06	41.78	0.65	0.32	
E108	15.53	26.89	39.06	0.70	0.25	
E107	13.35	24.32	15.06	0.78	0.37	
E110	15.85	23.10	20.88	0.73	0.20	
E96	8.00	28.20	2.62	0.92	0.22	
E97	12.74	24.07	16.79	0.82	0.28	
E98	12.21	23.94	14.92	0.83	0.24	
E99	14.70	25.42	15.43	0.74	0.30	

Table S15. RMSE and R^2 values for the out of sample analysis. Test and Train sets are made after leaving out a Catalyst/Ester Val (validation set) consists of datapoints pertaining to the leftout catalyst/ester. R^2_val score is not provided if the number of datapoints is less than 2.

4. Experimental details for the hydrogenation of esters

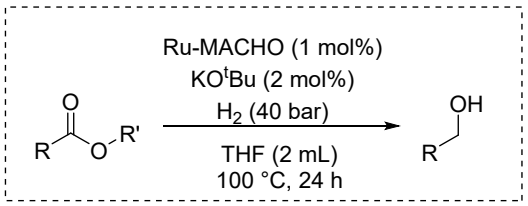
4.1 General considerations

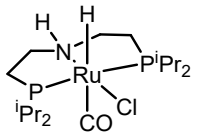
Ru-MACHO complex (**C1**), diphenylethylene (internal standard), $CDCl_3$, and esters (E59, E82, E84 and E105) were purchased from Alfa or Sigma Aldrich and used as received without further purification. KO^tBu was dried before use. THF was dried using a Grubbs type solvent purification system and degassed before use.

GC-MS spectra were collected using an Agilent 8860 GC System fitted with an HP-5ms Ultra Inert column (30 m × 250 µm × 0.25 µm) coupled with an Agilent 5977B GC/MSD. Samples were injected at an inlet temperature of 250 °C to ensure all volatiles entered the gas phase. GC conditions as follows: 50 °C isothermal for 4 minutes, ramp 10 °C/min to 250 °C, 250 °C isothermal for 15 minutes.

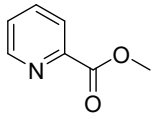
NMR spectra were recorded on a Bruker AVIII-HD 500 MHz NMR spectrometer at 298 K unless otherwise specified. Residual protio solvent was used as reference for ^1H spectra in deuterated solvent samples.

4.2 General procedure for the hydrogenation of esters

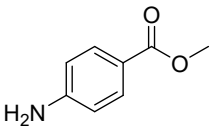




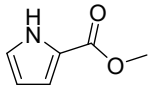
Ru-Macho (C1)



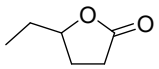
E59



E82



E84



E105

The reaction was performed using standard Schlenk techniques using an argon-vacuum double manifold.

To a microwave vial was added: Ru-MACHO (12.1 mg, 20 µmol), KO^{t}Bu (4.4 mg, 40 µmol), ester (2 mmol) and THF (2 mL). The microwave vial septum was pierced with two needles, which remained throughout the reaction, and placed in a stainless-steel autoclave equipped with a pressure gauge and partially filled with stainless-steel beads (to allow effective heat transfer). The atmosphere within the autoclave was replaced with H_2 and the vessel pressurised to 40 bar before sealing and heating to 100 °C for 24 h. After cooling to room temperature, excess H_2 is vented and the reaction mixture interrogated by GC-MS with the addition of a known quantity of internal standard: 1,1'-diphenylethylene.

Entry	Ester	Conversion[a] /%	Yield[b] /%	Product	Literature Yield/%	Predicted Yield/%
1	**E59**	>99	96	2-pyridinemethanol	3[16]	95.4
2	**E82**	63[c]	63	4-aminobenzylalcohol	0[17]	86
3	**E84**	<1	<1	2-methylpyrrole	0[17]	84
4	**E105**	>99	99	1,4-hexanediol	32[18]	96

[a] Determined by GC-MS. [b] NMR yield, *c.f.* internal [1,1'-diphenylethylene]. [c] Determined by ^{1}H NMR due to poor peak shapes in the GC-MS obtained.

Table S16. Observed yield (from experiments carried out here), literature yield and predicted yield (from ML model) for the hydrogenation of esters.

4.3 GC-MS Chromatograms and Mass Spectra of the crude reaction mixture after hydrogenation of esters

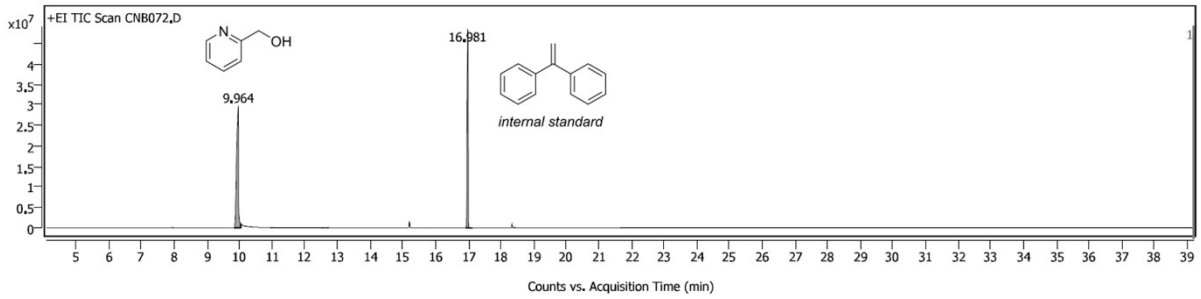


Figure S12 GC-MS chromatogram obtained of the reaction mixture from the hydrogenation of E59.

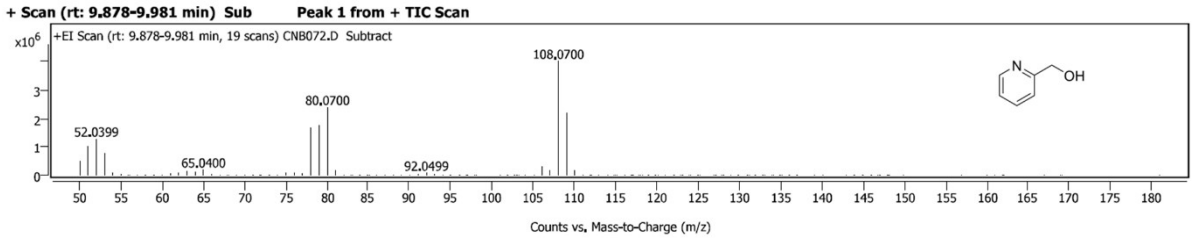


Figure S13 MS spectra corresponding to 2-pyridinemethanol obtained from the reaction mixture from the hydrogenation of E59.

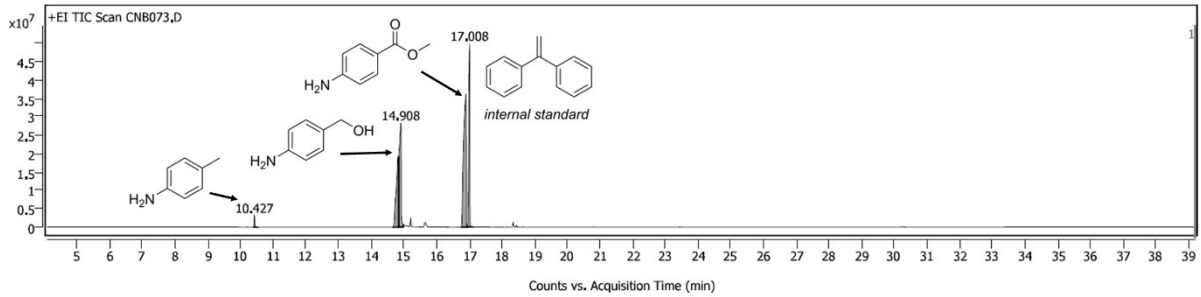


Figure S14 GC-MS chromatogram obtained of the reaction mixture from the attempted hydrogenation of E82.

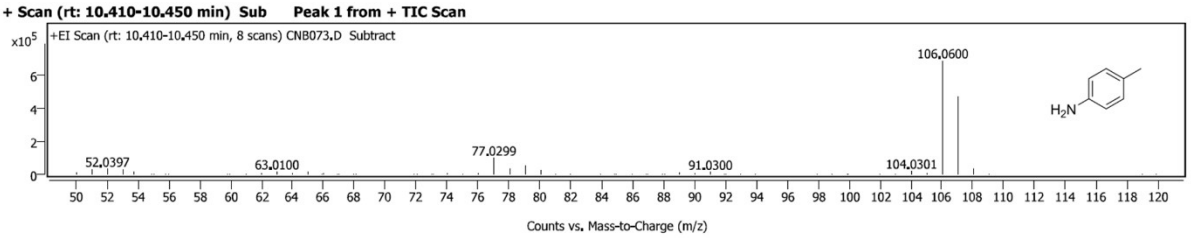


Figure S15 MS spectra corresponding to 4-methylaniline obtained from the reaction mixture from the attempted hydrogenation of E82.

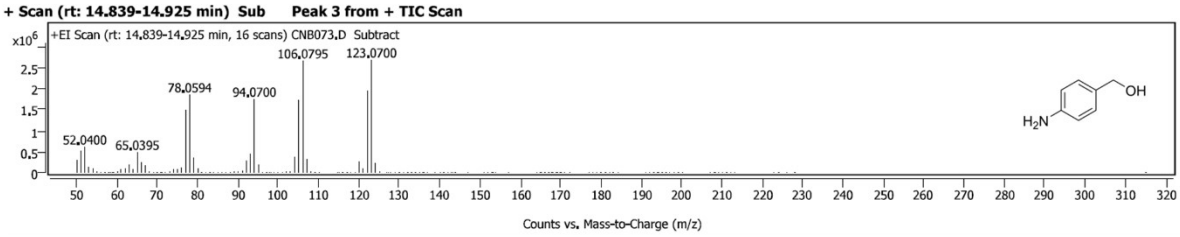


Figure S16 MS spectra corresponding to 4-aminobenzyl alcohol obtained from the reaction mixture from the attempted hydrogenation of E82.

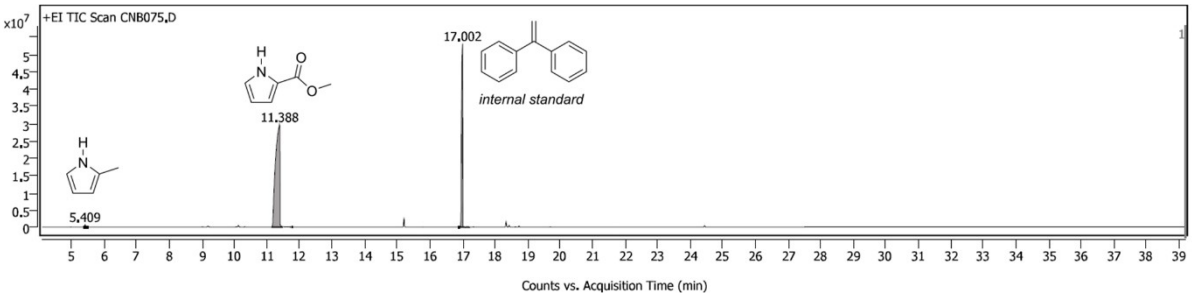


Figure S17 GC-MS chromatogram obtained of the reaction mixture from the attempted hydrogenation of E84.

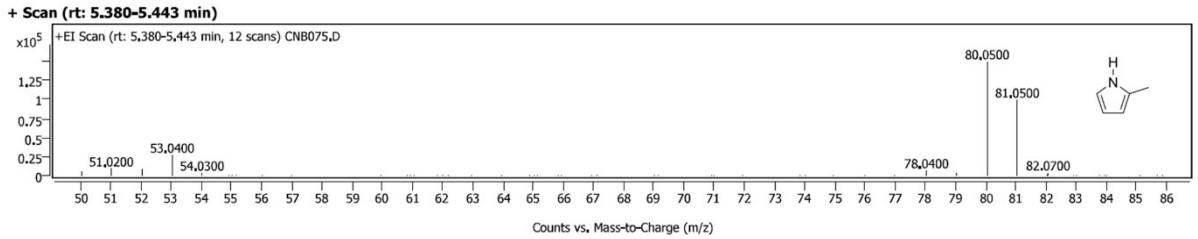


Figure S18 MS spectra corresponding to 2-methylpyrrole obtained from the reaction mixture from the attempted hydrogenation of E84.

4.4 NMR spectra of the crude reaction mixture after hydrogenation of esters

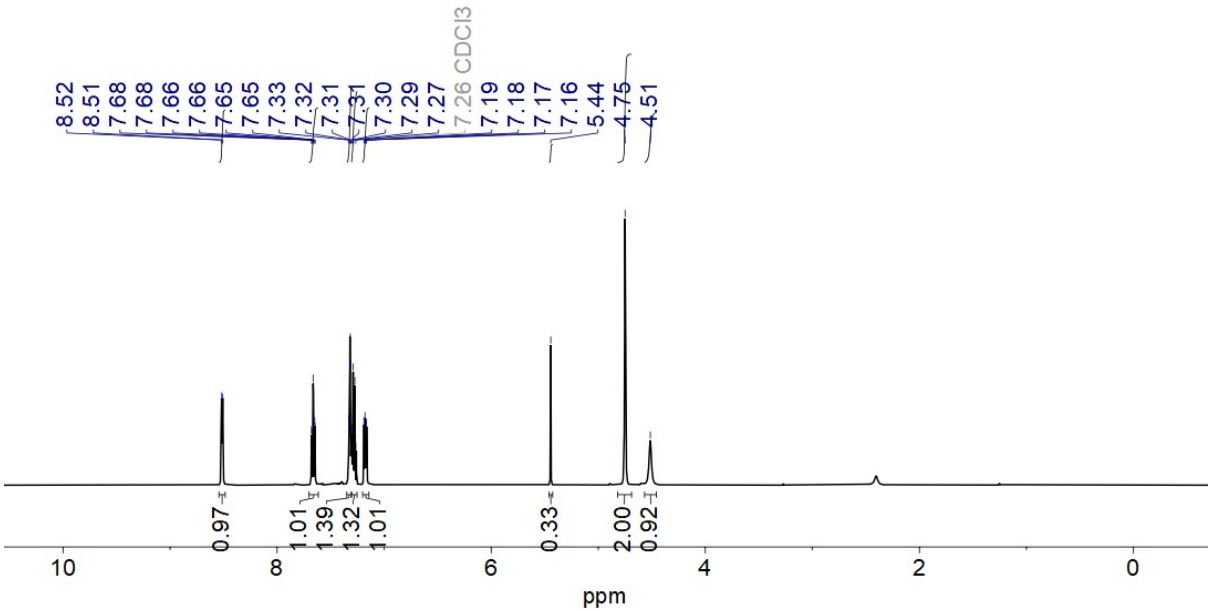


Figure S19 ^{1}H NMR (500 MHz, $CDCl_3$) spectrum of hydrogenation product of E59, with internal standard: 1,1'-diphenylethylene.

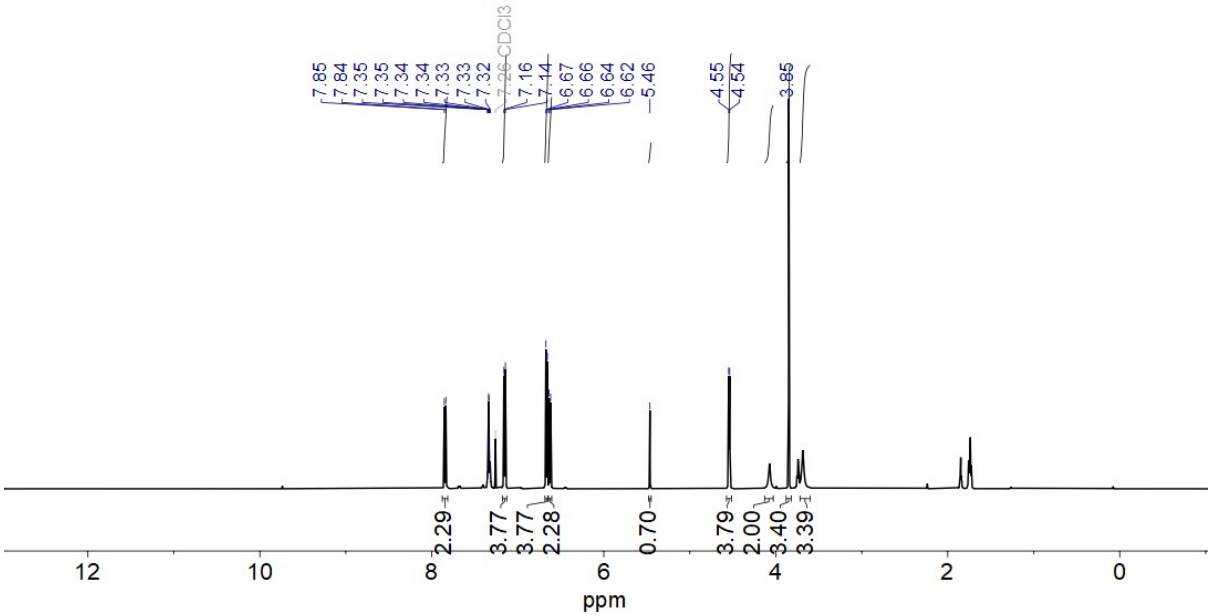


Figure S20 ^{1}H NMR (500 MHz, $CDCl_3$) spectrum of hydrogenation product of E82, with internal standard: 1,1'-diphenylethylene.

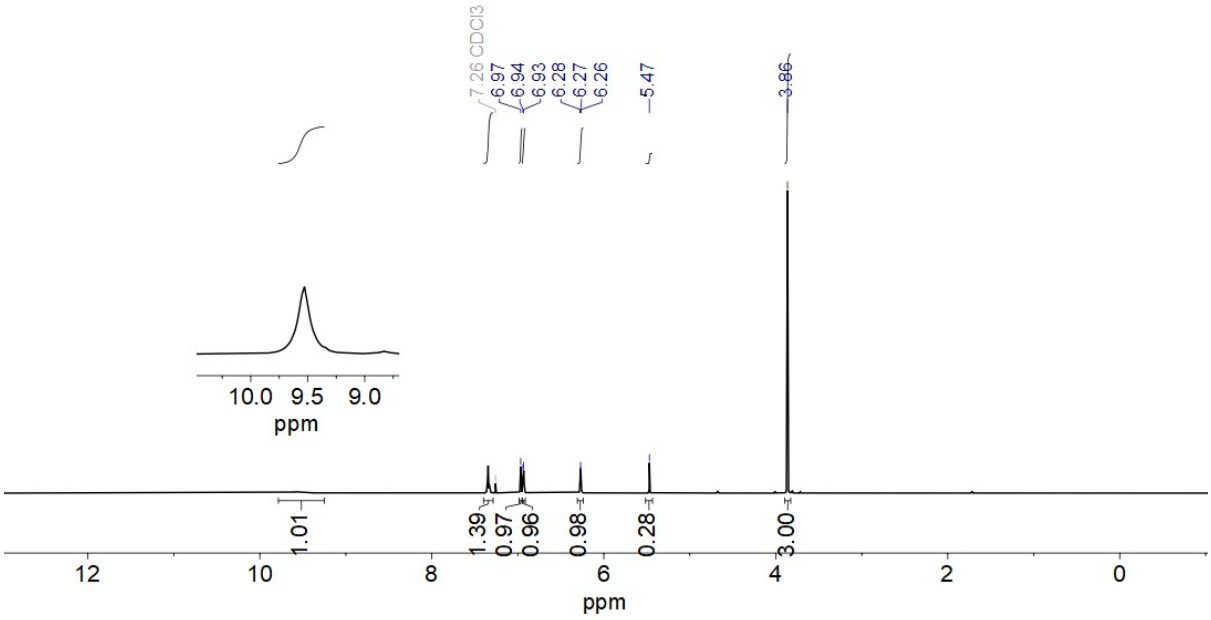


Figure S21 ^{1}H NMR (500 MHz, $CDCl_3$) spectrum of hydrogenation product of E84, with internal standard: 1,1'-diphenylethylene.

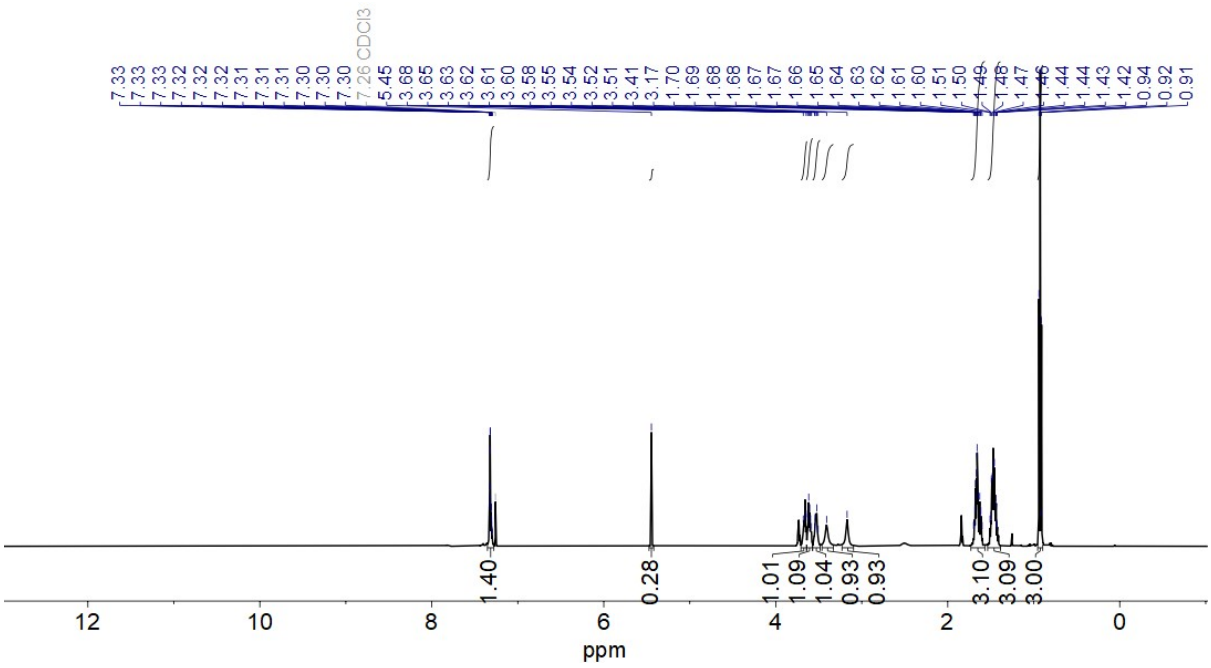


Figure S22 ^{1}H NMR (500 MHz, $CDCl_3$) spectrum of hydrogenation product of E105, with internal standard: 1,1'-diphenylethylene.

5. References

(1) Zhao, Y.; Truhlar, D. G. A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2006**, *125* (19), 194101. https://doi.org/10.1063/1.2370993.

(2) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305. https://doi.org/10.1039/B508541A.

(3) Weigend, F. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8* (9), 1057–1065. https://doi.org/10.1039/B515623H.

(4) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.

(5) Mardirossian, N.; Head-Gordon, M. ΩB97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation. *J. Chem. Phys.* **2016**, *144* (21), 214110. https://doi.org/10.1063/1.4952647.

(6) Vydrov, O. A.; Van Voorhis, T. Nonlocal van Der Waals Density Functional: The Simpler the Better. *J. Chem. Phys.* **2010**, *133* (24), 244103. https://doi.org/10.1063/1.3521275.

(7) Hujo, W.; Grimme, S. Performance of the van Der Waals Density Functional VV10 and (Hybrid)GGA Variants for Thermochemistry and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7* (12), 3866–3871. https://doi.org/10.1021/ct200644w.

(8) Hellweg, A.; Hättig, C.; Höfener, S.; Klopper, W. Optimized Accurate Auxiliary Basis Sets for RI-MP2 and RI-CC2 Calculations for the Atoms Rb to Rn. *Theor. Chem. Acc.* **2007**, *117* (4), 587–597. https://doi.org/10.1007/s00214-007-0250-5.

(9) Iron, M. A.; Janes, T. Evaluating Transition Metal Barrier Heights with the Latest Density Functional Theory Exchange–Correlation Functionals: The MOBH35 Benchmark Database. *J. Phys. Chem. A* **2019**, *123* (17), 3761–3781. https://doi.org/10.1021/acs.jpca.9b01546.

(10) Mondal, B.; Neese, F.; Ye, S. Control in the Rate-Determining Step Provides a Promising Strategy To Develop New Catalysts for CO2 Hydrogenation: A Local Pair Natural Orbital Coupled Cluster Theory Study. *Inorg. Chem.* **2015**, *54* (15), 7192–7198. https://doi.org/10.1021/acs.inorgchem.5b00469.
(11) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396. https://doi.org/10.1021/jp810292n.
(12) Software update: the ORCA program system, version 4.0 - Neese - 2018 - WIREs Computational Molecular Science - Wiley Online Library https://wires.onlinelibrary.wiley.com/doi/full/10.1002/wcms.1327 (accessed 2021 -10 -20).
(13) Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L. Towards the Online Computer-Aided Design of Catalytic Pockets. *Nat. Chem.* **2019**, *11* (10), 872–879. https://doi.org/10.1038/s41557-019-0319-5.
(14) https://github.com/rdkit/rdkit
(15) https://github.com/SheffieldML/GPy
(16) Ogata, O.; Nakayama, Y.; Nara, H.; Fujiwhara, M.; Kayaki, Y. *Org. Lett*. **2016**, *18*, 3894-3897.
(17) Carpenter, I.; Eckelmann, S. C.; Kuntz, M. T.; Fuentes, J. A.; France, M. B.; Clarke, M. L. *Dalton Trans.* **2012**, *41*, 10136-10140.
(18) O, W. W. N.; Morris, R. *ACS Catal*., **2013**, *3*, 32-40.