Using Machine Learning to Predict Suitable Conditions for Organic Reactions

Hanyu Gao, Thomas J. Struble, Connor W. Coley, Yuran Wang, William H. Green, Klavs F. Jensen*

Department of Chemical Engineering, Massachusetts Institute of Technology; 77 Massachusetts Avenue, Cambridge, MA 02139

E-mail: kfjensen@mit.edu

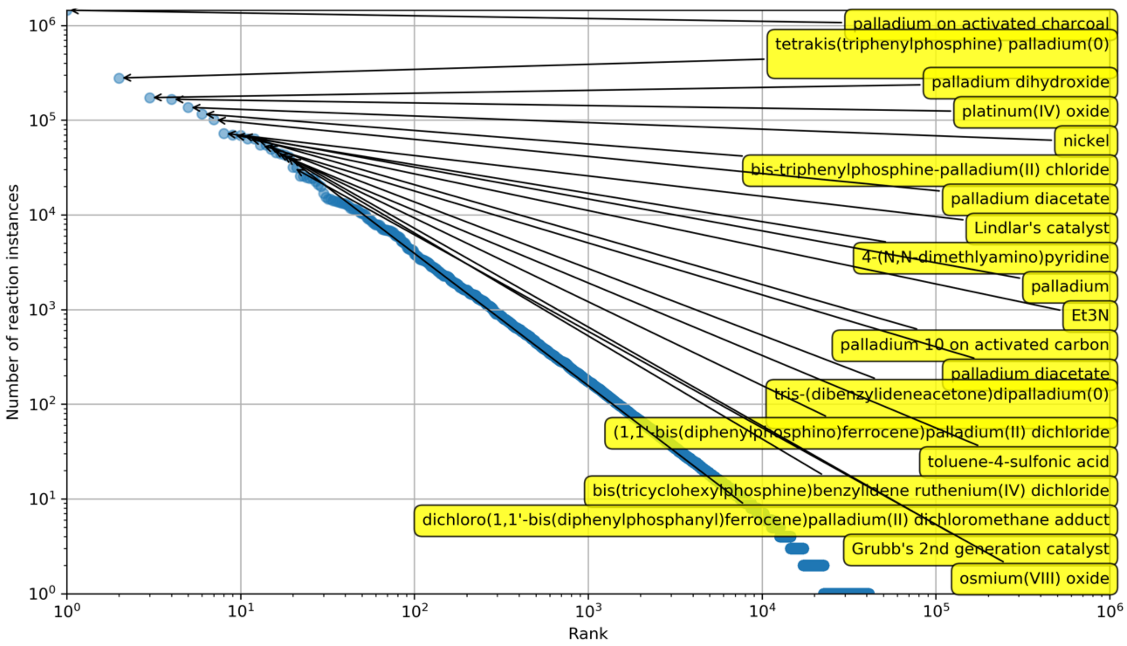
Supplementary Information

Computational methods

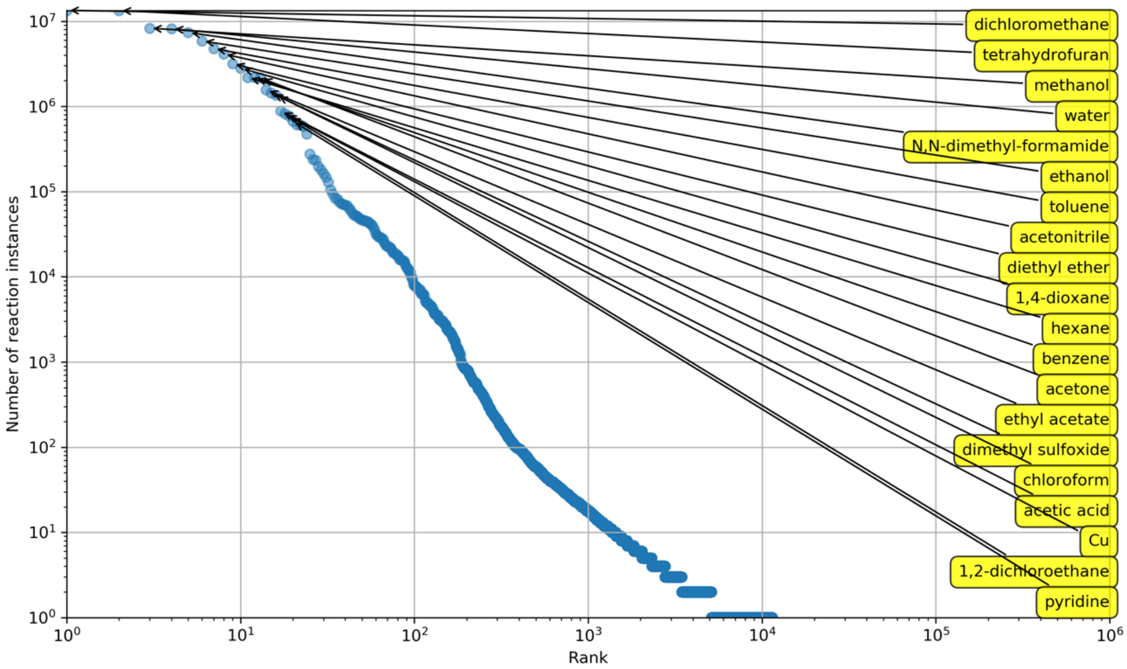
Neural network models are developed in Keras (Version 2.0.2) and trained using the Theano backend (Version 1.0.0). All trainings are performed on a single NVIDIA GeForce GTX 1080 GPU. The code is developed in Python 2.7.

Frequency vs rank plots for catalysts, reagents and solvents

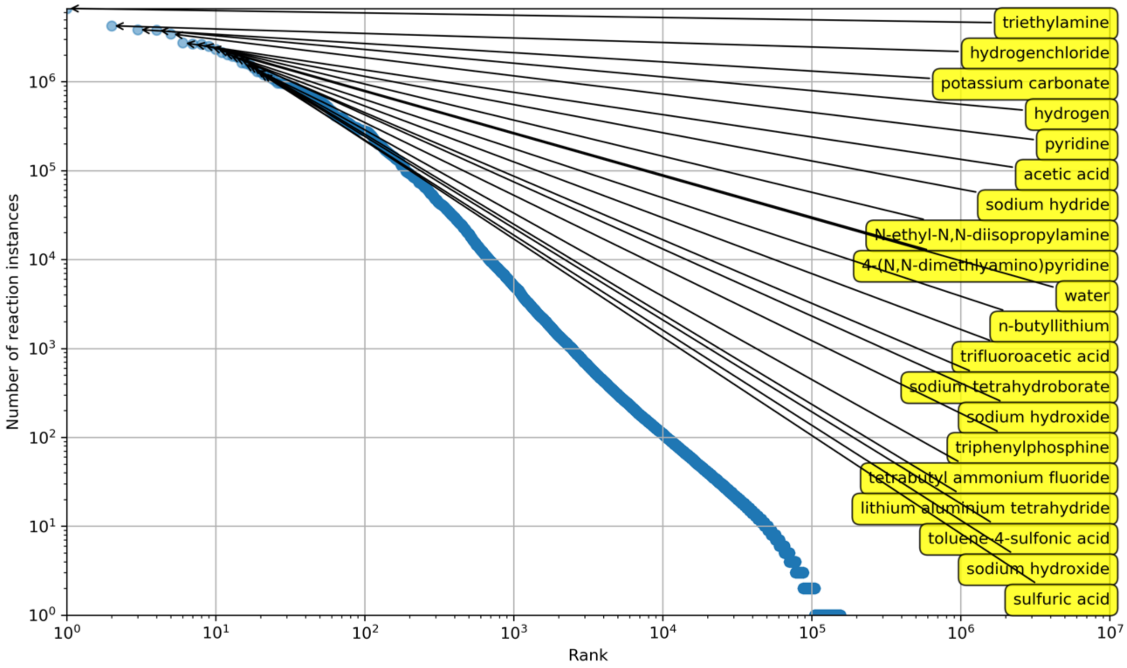
Supplementary Figures 1-3 are the frequency vs. rank plots for catalyst, solvent and reagent in Reaxys. It can be seen some duplicated records of the same chemical exists (e.g. palladium diacetate as catalyst and sodium hydroxide as reagent). We kept these as different classes as there is not a good way to systematically identify and curate this issue as it is not clear what is the pattern of these duplicated chemicals with different ids. However, during the training, the model learns that they are very similar entities because one or another of these duplicated chemicals are used in the same type of reactions.



Supplementary Figure 1. Frequency vs. rank plot for catalyst in Reaxys (with the top ones labeled in yellow)



Supplementary Figure 2. Frequency vs. rank plot for solvent in Reaxys (with the top ones labeled in yellow)



Supplementary Figure 3. Frequency vs. rank plot for reagent in Reaxys (with the top ones labeled in yellow)

Comparison of prediction accuracies with a null model

A null model is defined to always give the same prediction of top ten combinations chosen based on the frequencies of the catalysts, solvents and reagents. The top 10 combinations of the null model are listed below in Supplementary Table 1.

Supplementary Table 1. The top-ten combinations used in the null model

Rank	Catalyst	Solvent 1	Solvent 2	Reagent 1	Reagent 2
1					
2		DCM[a]			
3		THF			
4				TEA	
5				K_2CO_3	
6		DCM		TEA	
7		THF		TEA	
8		DCM		K_2CO_3	
9		THF		K_2CO_3	
10	Pd on activated charcoal				

[a]: DCM: dichloromethane, THF: tetrahydrofuran, TEA: triethylamine

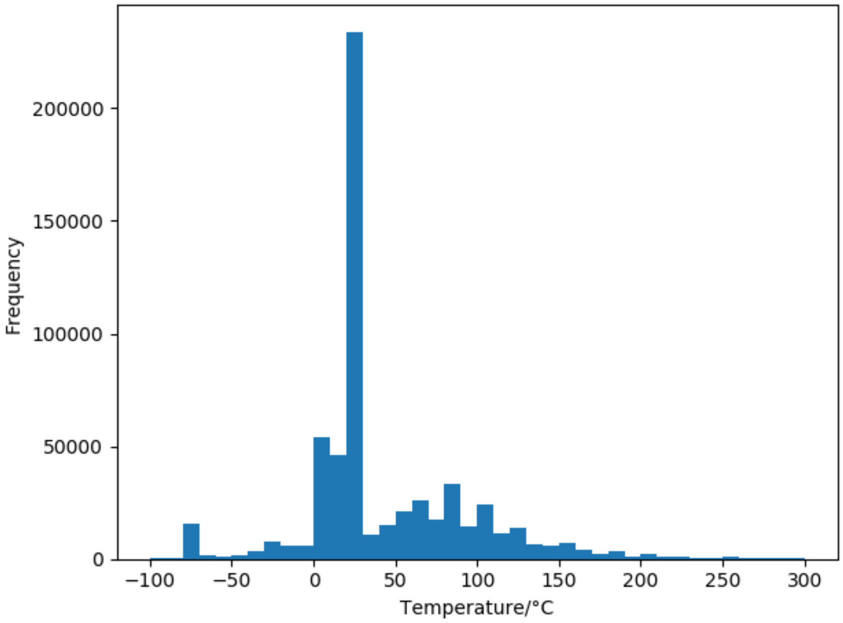
The comparison of prediction accuracies are shown in Supplementary Table 2. The accuracy values shown are for top-three exact matches. In general, it can be seen that the accuracy values are much lower for the null model. The accuracy of c, s2 and r2 predictions for the null model are high, since a majority of reactions do not use a catalyst, a second solvent or a second reagent, but the trained model is still better than the null model by a large margin.

Supplementary Table 2. Comparison of accuracy for the true context to be in the top-3 predictions to a null model

	Trained neural network model	Null model
c	93.6%	87.3%
s1	75.8%	49.4%
s2	90.1%	85.6%
r1	73.2%	22.0%
r2	89.3%	82.3%
c, s1, r1	57.3%	5.7%
c, s1, s2, r1, r2	50.1%	4.7%

[a] c, s1, s2, r1, r2 refer to catalyst, solvent 1, solvent 2, reagent 1 and reagent 2, respectively;

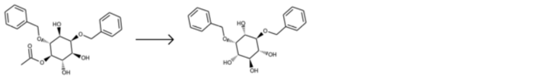
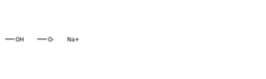
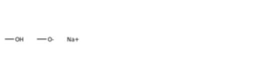
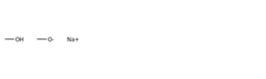
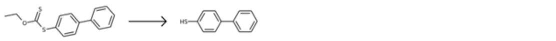
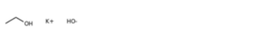
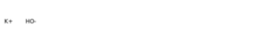
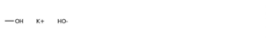
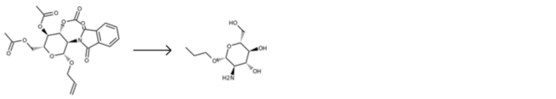
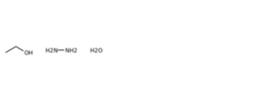
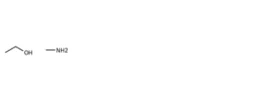
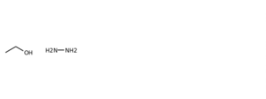
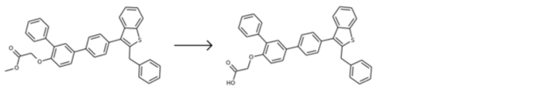
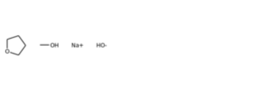
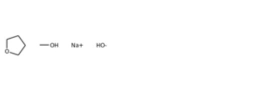
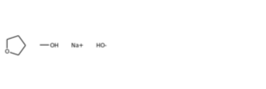
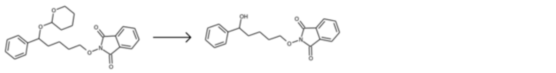
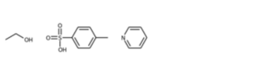
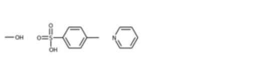
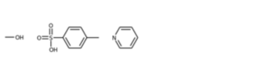
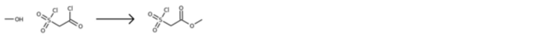
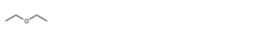
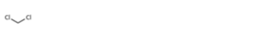
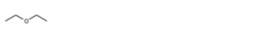
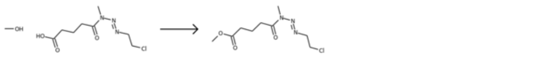
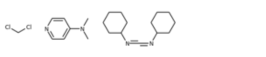
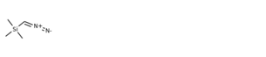
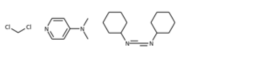
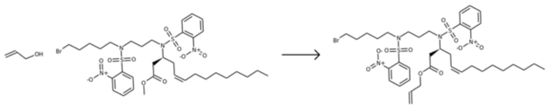
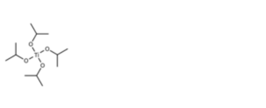
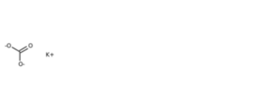
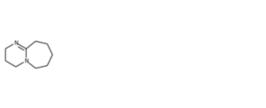
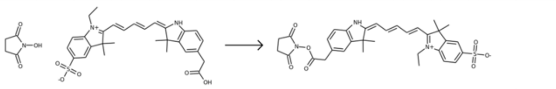
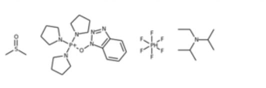
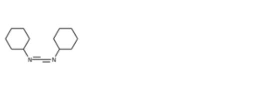
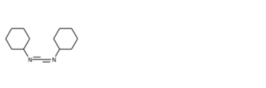
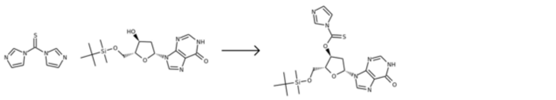
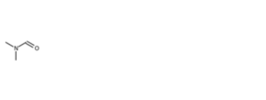
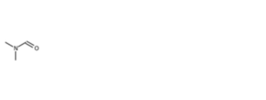
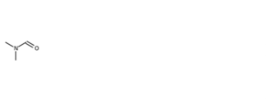
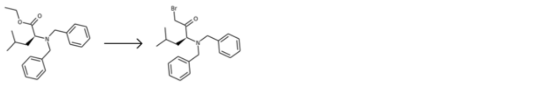
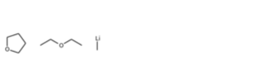
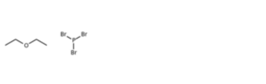
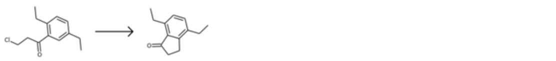
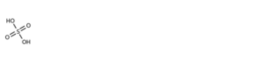
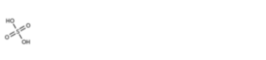
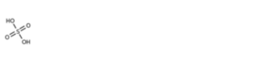
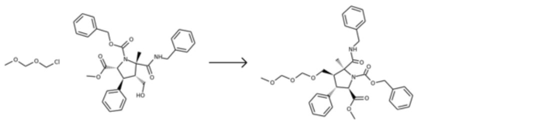
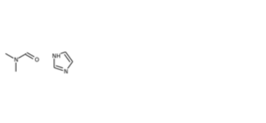
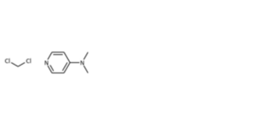
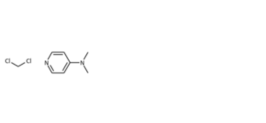
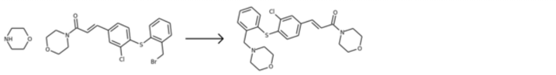
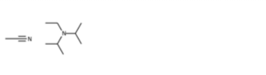
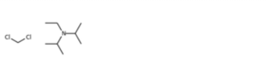
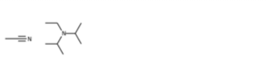
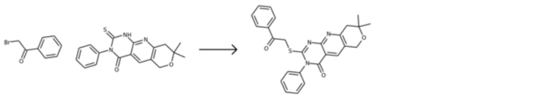
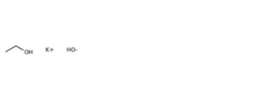
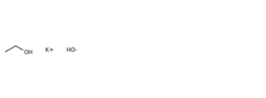
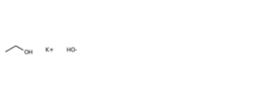
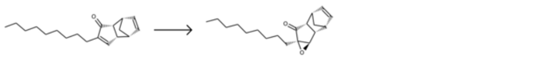
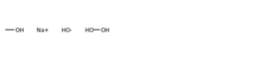
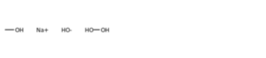
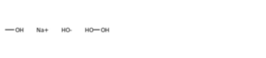
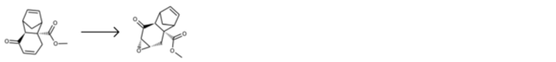
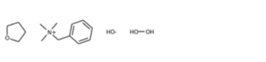
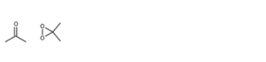
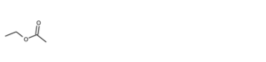
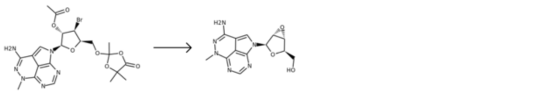
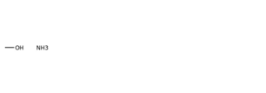
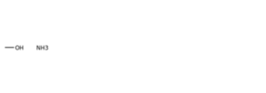
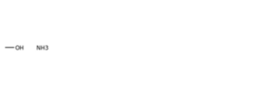
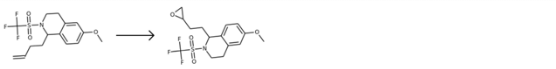
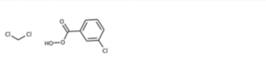
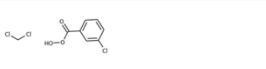
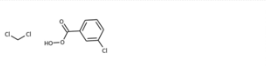
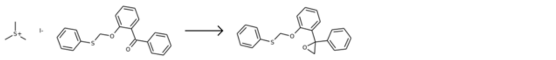
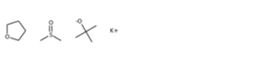
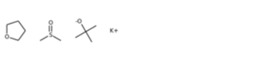
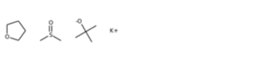
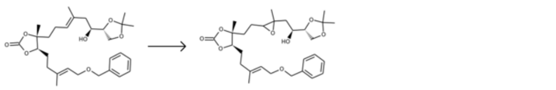
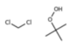
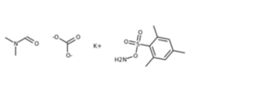
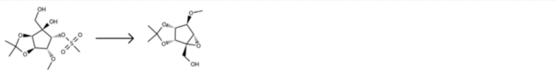
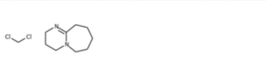
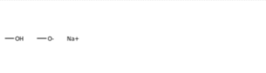
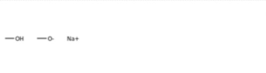
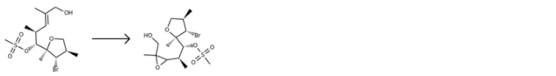
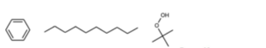
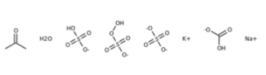
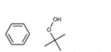
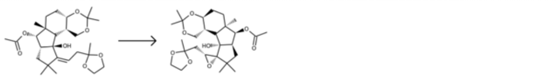
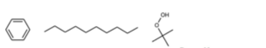
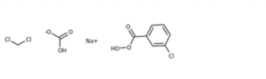
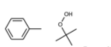
For temperature prediction, we analyzed a baseline model that predicts the most frequently used temperature for all reactions. Supplementary Figure 4 shows the distribution of temperature for reactions in the test set. The most frequently used temperature is the room temperature (20 °C) which covers a majority of reaction, and the accuracy of the predicted temperature by the baseline model (which is always 20 °C) being within the ±10°C or ±20°C range of the recorded temperature are 40.0% and 49.4%. In the meantime the distribution spans a wide range. Simply predicting the room temperature (20 °C) will result in a mean absolute error of 35.3 °C, which is significantly larger than prediction given by the trained model and would be misleading for reactions that require high or low temperatures.

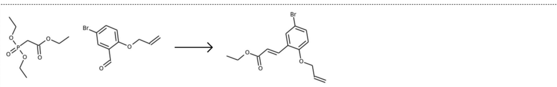
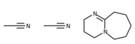
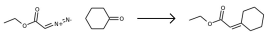
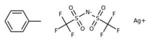
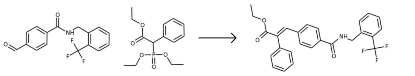
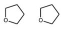
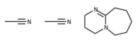
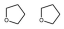
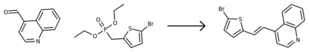
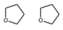
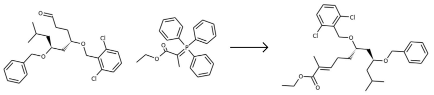
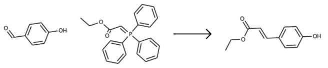
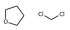
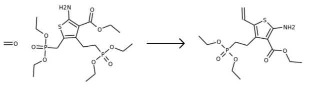
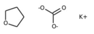
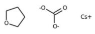
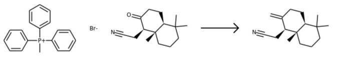
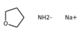
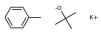
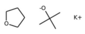
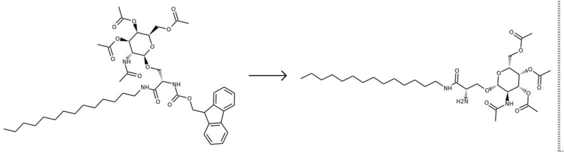
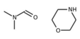
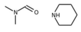
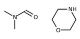
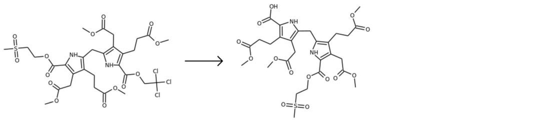
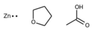
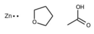
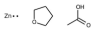
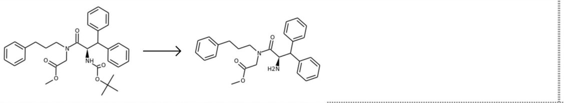
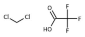
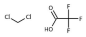
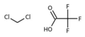
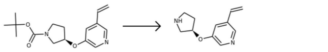
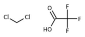
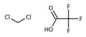
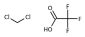
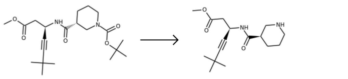
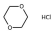
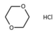
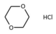
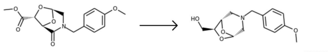
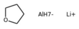
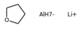
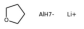
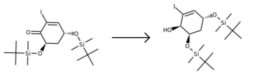
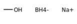
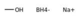
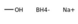
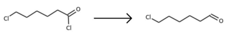
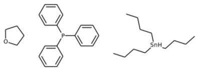
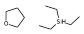
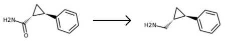
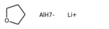
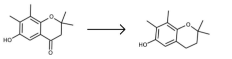
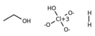
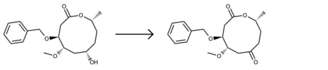
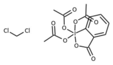
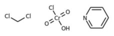
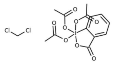
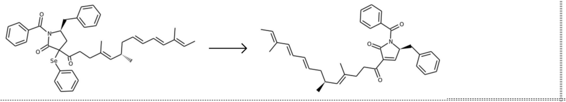
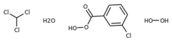
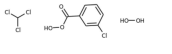
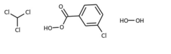
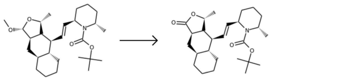
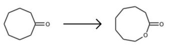
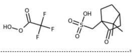
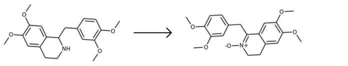
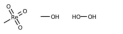
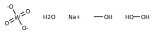
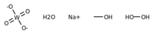
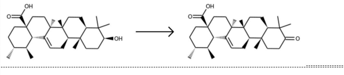
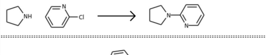
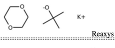
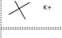
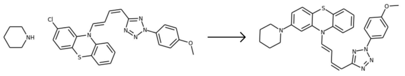
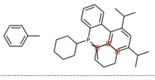
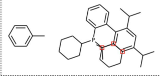


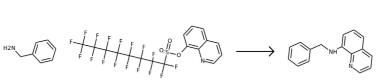
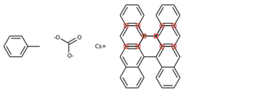
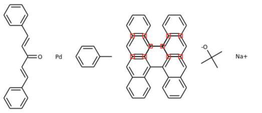
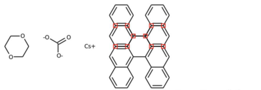
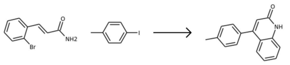
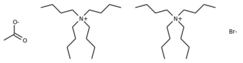
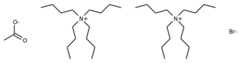
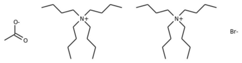
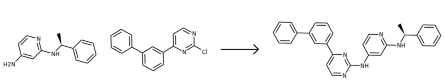
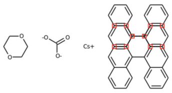
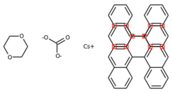
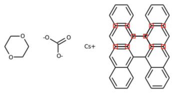
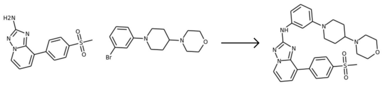
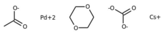
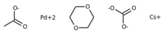
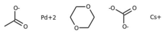
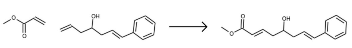
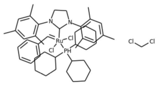
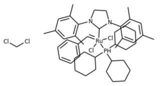
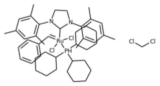
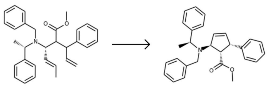
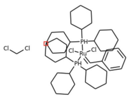
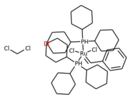
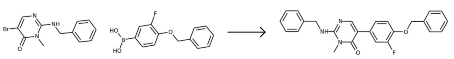
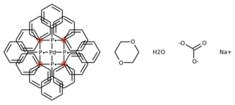
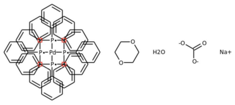
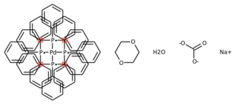
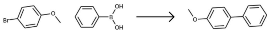
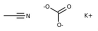
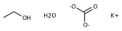
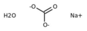
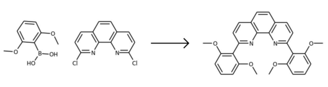
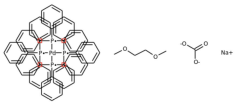
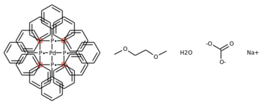
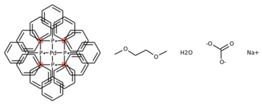
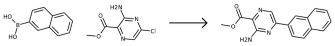
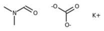
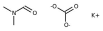
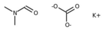
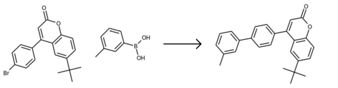
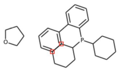
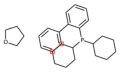
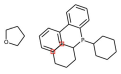
Supplementary Figure 4. Temperature distribution for reactions in the test set

Full list of evaluation of reaction examples

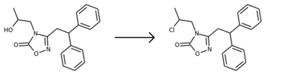
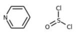
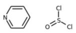
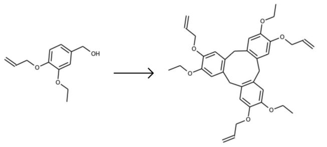
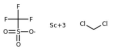
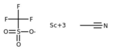
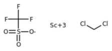
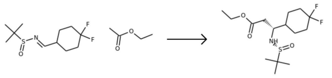
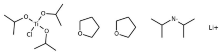
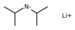
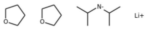
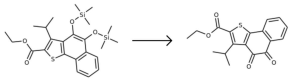
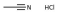
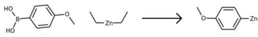
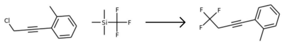
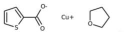
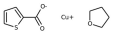
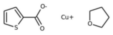
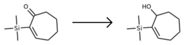
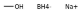
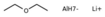
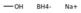
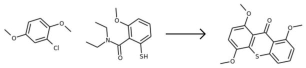
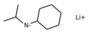
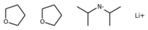
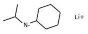
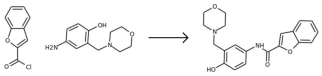
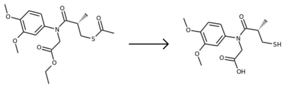
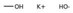
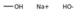
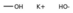
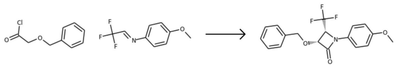
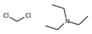
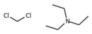
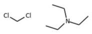
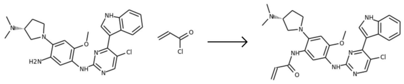
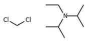
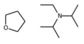
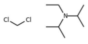
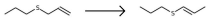
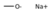
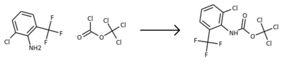
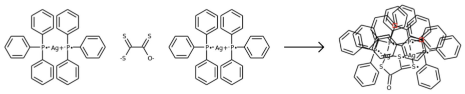
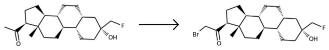
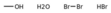
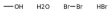
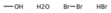
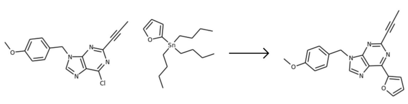
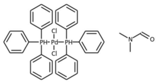
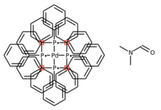
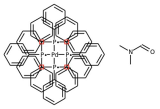
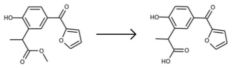
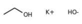
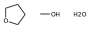
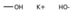
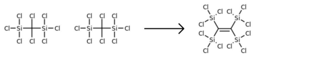
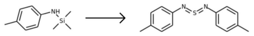
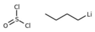
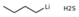
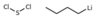
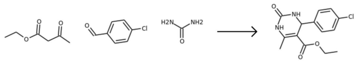
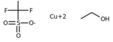
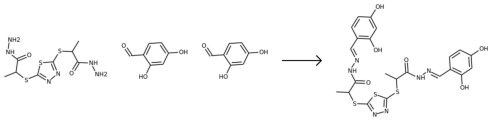
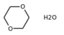
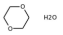
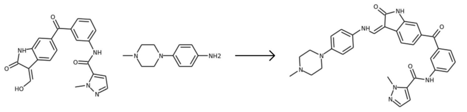
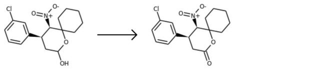
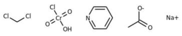
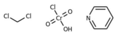
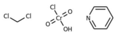
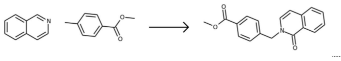
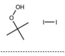
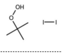
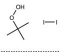
Supplementary Table 3. 62 reactions from eleven reaction types randomly chosen from the test dataset

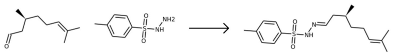
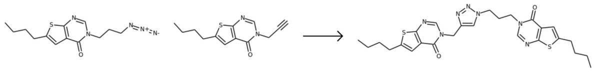
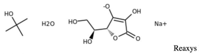
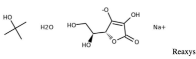
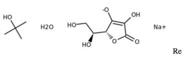
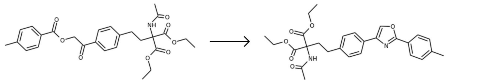
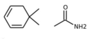
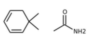
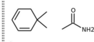
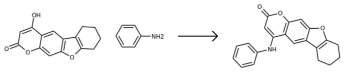
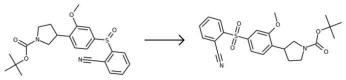
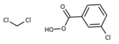
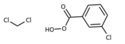
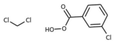
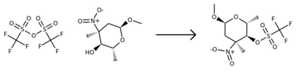
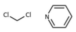
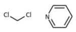
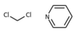
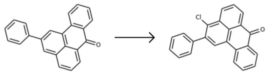
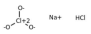
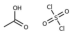
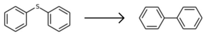
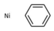
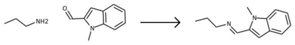
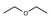
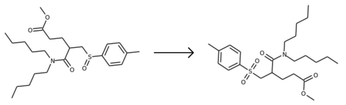
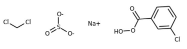
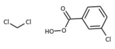
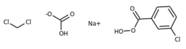
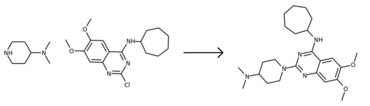
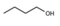
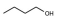
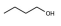
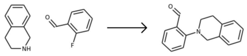
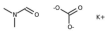
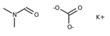
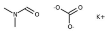
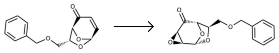
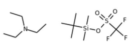
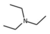
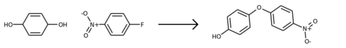
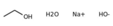
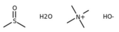
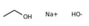
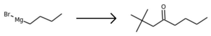
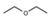
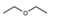
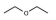
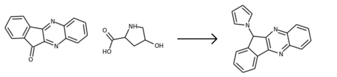
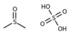
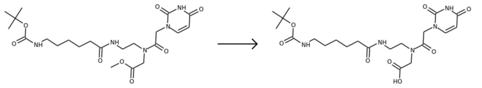
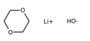
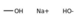
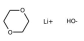
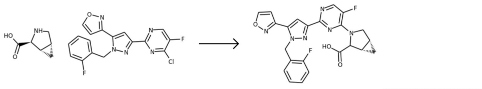
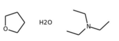
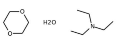
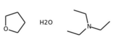
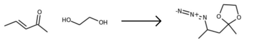
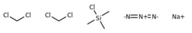
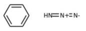
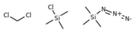
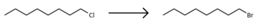
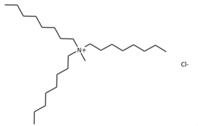
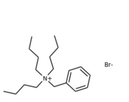
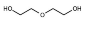
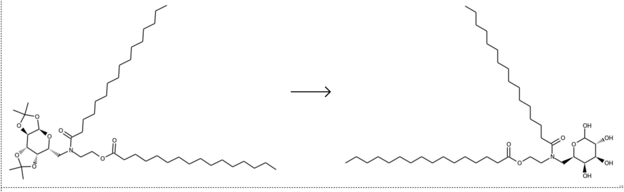
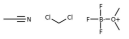
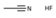
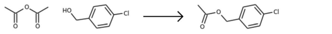
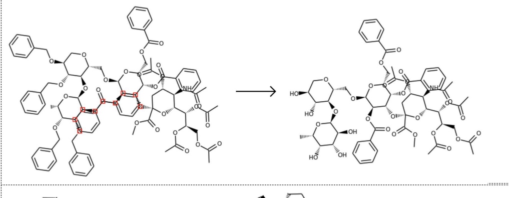
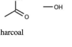
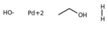
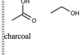
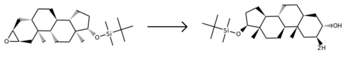
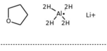
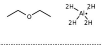
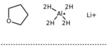
Reaxys ID	Reaction	True Context	Top Prediction	Closest Prediction	True Temperature/ °C	Top Predicted Temperature/ °C	Closest Predicted Temperature/ °C	Reaction type
5301921					20.0	19.3	19.3	Hydrolysis
8712792					N/A	104.4	59.7	Hydrolysis
5261303					N/A	92.9	88.2	Hydrolysis
8655717					N/A	30.2	30.2	Hydrolysis
5302180					57.5	35.1	35.1	Hydrolysis
53162					0.0	12.2	5.5	Esterification
5185761					10.0	15.7	11.2	Esterification
8698819					95.0	51.6	59.7	Esterification
8669084					N/A	15.0	15.0	Esterification
5336683					20.0	48.2	48.2	Esterification
5220300				Reaxys Name lithium bromide	-78.0	5.7	13.6	Alkylation
2820838					54.5	71.6	71.6	Alkylation
8568283					N/A	18.4	18.4	Alkylation
8738792					20.0	23.1	47.6	Alkylation
8574982					40.0	28.2	28.2	Alkylation
8591626					10.0	15.1	15.1	Epoxidation
8647861					0.0	5.4	47.8	Epoxidation
8655813					20.0	22.5	22.5	Epoxidation
8720054					20.0	15.2	15.2	Epoxidation
8560867					0.0	20.7	20.7	Epoxidation
8598928		 Reaxys Name bis(acetylacetonate)oxovanadium		Reaxys Name sodium hydride	0.0	12.7	20.9	Epoxidation
5229646					0.0	35.8	35.8	Epoxidation
10069687		 Reaxys Name bis(acetylacetonate)oxovanadium		 Reaxys Name bis(acetylacetonate)oxovanadium	25.0	15.2	48.5	Epoxidation
9925682		 Reaxys Name bis(acetylacetonate)oxovanadium		 Reaxys Name bis(acetylacetonate)oxovanadium	N/A	6.0	34.3	Epoxidation

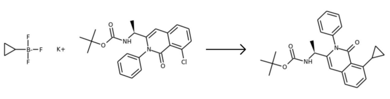
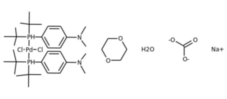
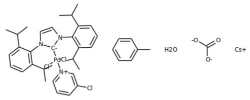
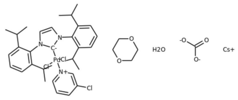
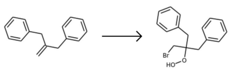
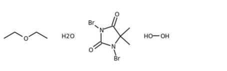
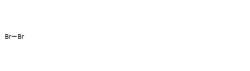
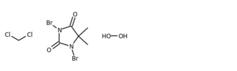
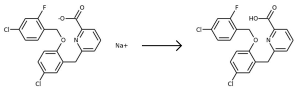
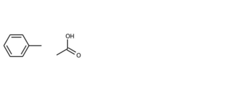
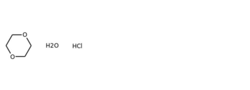
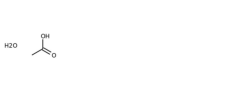
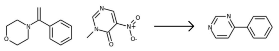
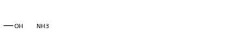
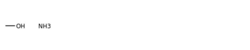
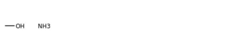
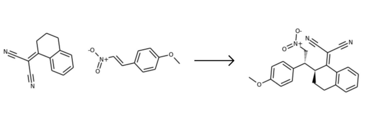
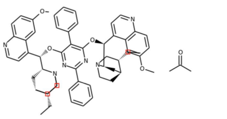
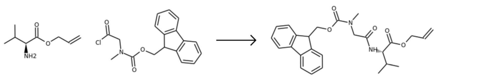
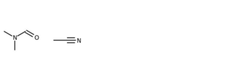
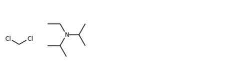
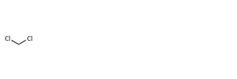
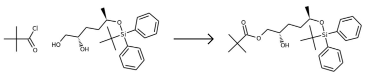
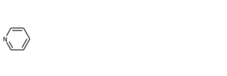
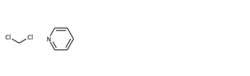
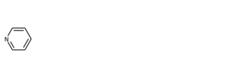
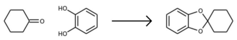
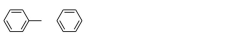
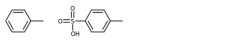
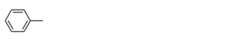
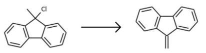
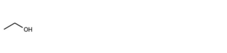
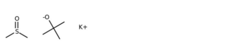
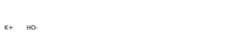
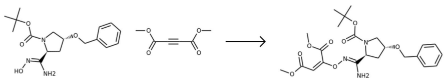
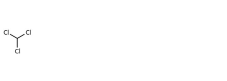
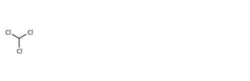
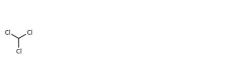
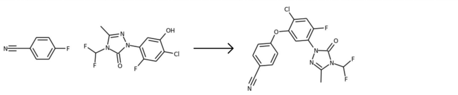
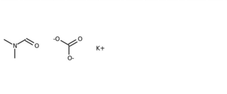
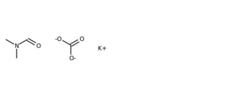
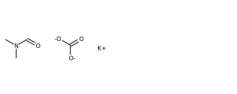
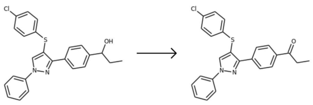
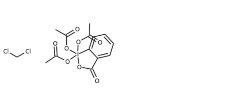
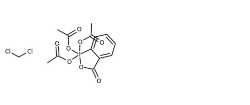
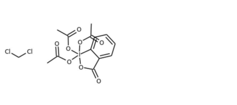
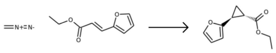
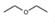
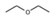
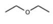
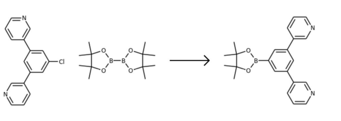
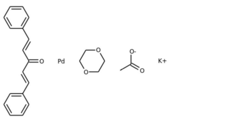
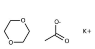
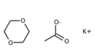
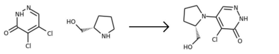
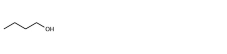
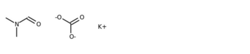
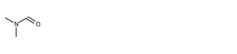
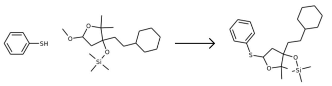
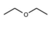
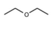
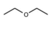
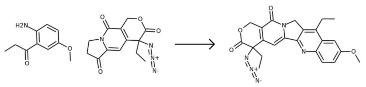
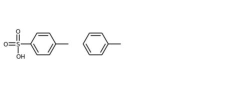
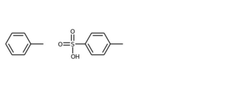
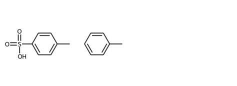
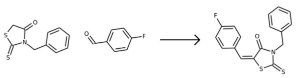
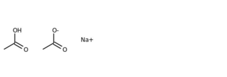
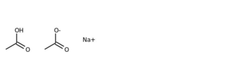
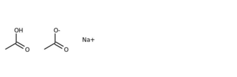
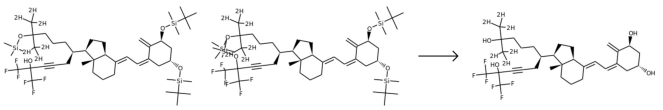
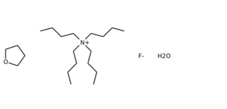
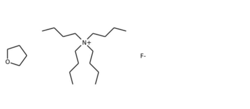
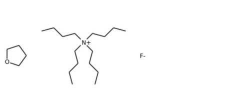
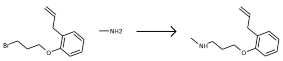
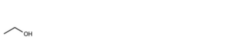
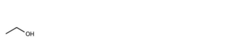
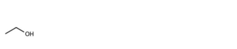
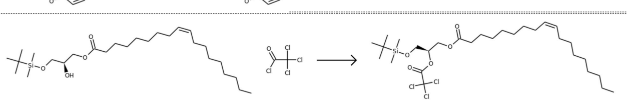
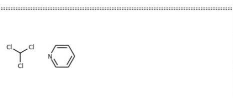
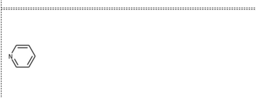
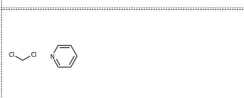
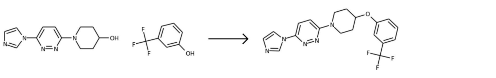
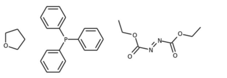
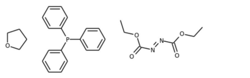
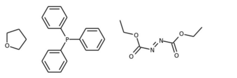
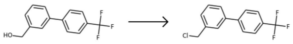
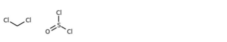
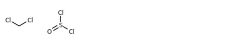
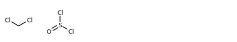
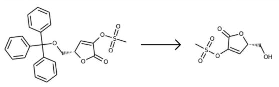
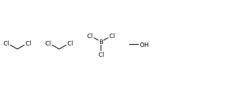
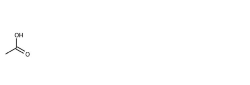
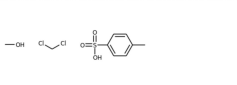
40053563		 Reaxys Name lithium bromide	Reaxys Name sodium hydride	Reaxys Name sodium hydride	20.0	15.4	15.4	Wittig
1558712					110.0	44.0	80.1	Wittig
41951969		 Reaxys Name sodium hydride	 Reaxys Name lithium chloride	 Reaxys Name sodium hydride	0.0	10.7	9.5	Wittig
9299119		 Reaxys Name sodium hydride	Reaxys Name sodium hydride	Reaxys Name sodium hydride	N/A	16.6	16.6	Wittig
28171183					10.0	47.6	19.9	Wittig
28476112					20.0	53.2	53.2	Wittig
9240363					20.0	30.3	26.4	Wittig
2287762					20.0	42.3	9.4	Wittig
8659689					20.0	21.5	18.4	Deprotection
5357436					20.0	17.8	17.8	Deprotection
5304351					N/A	16.9	16.9	Deprotection
8588894					N/A	15.9	15.9	Deprotection
5351204					N/A	20.0	20.0	Deprotection
5297919					N/A	8.7	8.7	Reduction
31266961		 Reaxys Name cerium(III) chloride heptahydrate	 Reaxys Name cerium(III) chloride	 Reaxys Name cerium(III) chloride	-78.0	-16.2	-16.2	Reduction
5265062		 Reaxys Name indium(III) chloride	Reaxys ID 11378932.Reaxys Name xylene	 Reaxys Name palladium on activated charcoal	-30.0	124.2	16.9	Reduction
8619786		AlH7- Li+		AlH7- Li+	N/A	20.8	26.9	Reduction
5261651		 Reaxys Name palladium on activated charcoal	Hg•• Zn•• HCl	Hg•• Zn•• HCl	20.0	75.4	75.4	Reduction
28315260					10.0	19.5	18.2	Oxidation
5307293					-25.0	3.1	3.1	Oxidation
9228133		Reaxys Name jones reagent	Reaxys Name jones reagent	Reaxys Name jones reagent	20.0	1.7	1.7	Oxidation
116049					N/A	46.2	46.2	Oxidation
5081868					20.0	20.6	20.6	Oxidation
339468		Reaxys Name Jone's reagent.Reaxys Name silica gel	Reaxys Name Jone's reagent	Reaxys Name Jone's reagent	0.0	12.2	12.2	Oxidation
4977260		 Reaxys ID 21565323			80.0	176.7	130.2	Buchwald-Hartwig
34039114					95.0	42.6	70.1	Buchwald-Hartwig

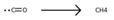
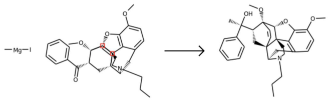
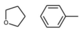
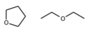
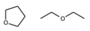
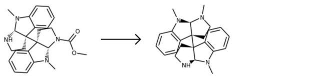
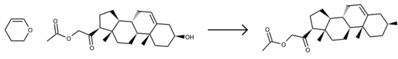
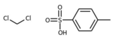
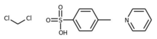
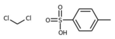
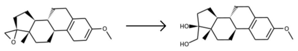
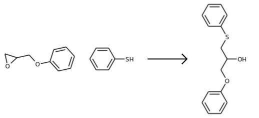
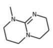
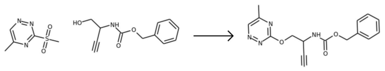
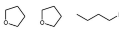
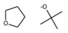
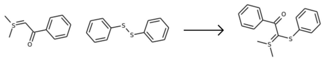
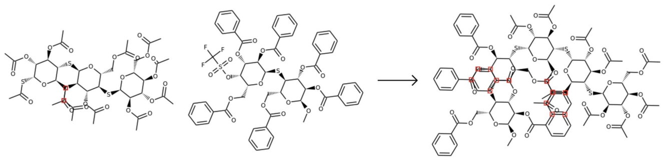
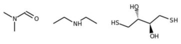
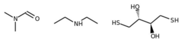
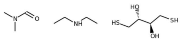
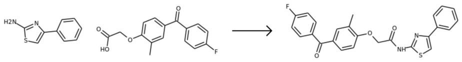
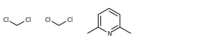
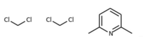
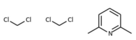
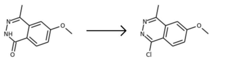
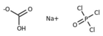
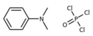
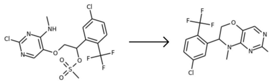
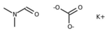
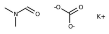
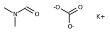
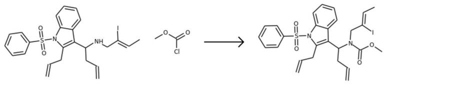
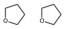
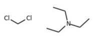
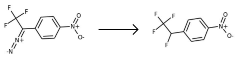
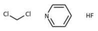
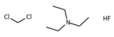
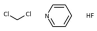
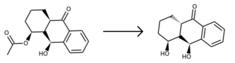
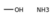
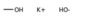
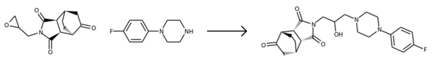
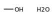
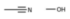
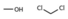
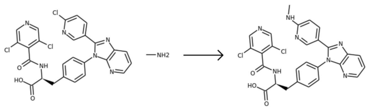
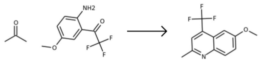
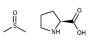
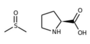
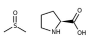
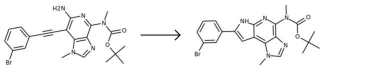
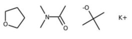
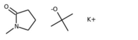
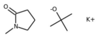
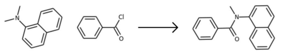
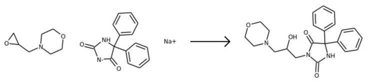
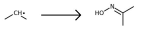
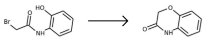
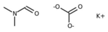
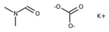
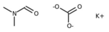
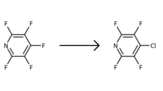
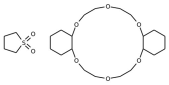
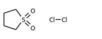
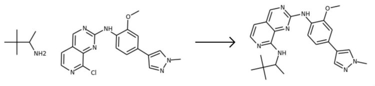
11260415		 Reaxys Name palladium diacetate		 Reaxys Name palladium diacetate	90.0	94.9	94.7	Buchwald-Hartwig
11077029		 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	120.0	120.5	120.5	Buchwald-Hartwig
26046174		 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	90.0	104.3	104.3	Buchwald-Hartwig
33823237					80.0	113.5	113.5	Buchwald-Hartwig
9603192					19.0	31.5	39.1	Grubbs
36032266					40.0	35.6	22.6	Grubbs
25840737		 Reaxys Name lithium chloride	 Reaxys Name lithium chloride	 Reaxys Name lithium chloride	100.0	101.8	101.8	Suzuki
4033276		 Reaxys Name palladium diacetate			8.0	51.5	81.1	Suzuki
9208506					N/A	84.1	84.1	Suzuki
28885714		 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	N/A	95.3	95.3	Suzuki
10040528		 Reaxys Name palladium diacetate.Reaxys Name potassium fluoride	 Reaxys Name palladium diacetate.Reaxys Name potassium fluoride	 Reaxys Name palladium diacetate.Reaxys Name potassium fluoride	20.0	52.2	52.2	Suzuki

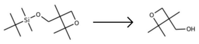
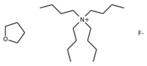
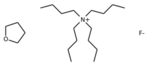
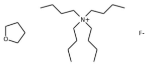
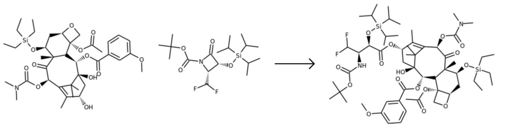
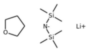
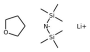
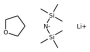
Supplementary Table 4. 100 reactions randomly chosen from the test dataset

Reaxys ID	Reaction	True Context	Top Prediction	Closest Prediction	True Temperature/ °C	Top Predicted Temperature/ °C	Closest Predicted Temperature/ °C
1383215					N/A	40.3	32.1
10061846					60.0	29.4	23.8
44110175					-78.0	-65.6	-66.6
10261736		 Reaxys Name ammonium cerium(IV) nitrate	 Reaxys Name ammonium cerium(IV) nitrate	 Reaxys Name ammonium cerium(IV) nitrate	0.0	15.1	15.1
40887565					60.0	54.3	54.3
36028195		 Reaxys Name potassium fluoride	 Reaxys Name potassium fluoride	 Reaxys Name potassium fluoride	60.0	49.1	49.1
2176355		 Reaxys Name cerium(III) chloride		 Reaxys Name cerium(III) chloride	0.0	-18.7	-4.2
2069147					N/A	-42.6	-13.1
1720472					N/A	54.0	54.0
3204380					N/A	26.2	24.5
4822624					40.0	5.7	5.7
34927996					N/A	14.5	12.1
1013620					N/A	263.4	263.4
3928353					60.0	66.7	66.7
26314583					N/A	25.8	25.8
38789242					25.0	36.3	36.3
25852492					90.0	88.2	88.2
3665938			 Reaxys Name lithium hydroxide hydrate		N/A	28.7	36.7
3958431		Cu	Cu	Cu	190.0	254.5	254.5
2215973					N/A	-15.5	-47.1
4940666					100.0	85.4	60.6
5310995					70.0	42.1	44.6
25900877					65.0	46.3	46.3
35559718					30.0	21.7	21.7
42570631					120.0	104.7	104.7

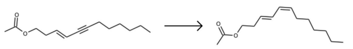
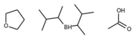
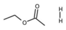
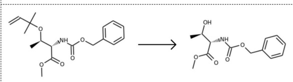
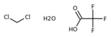
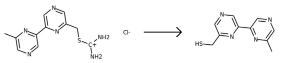
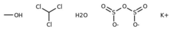
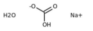
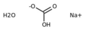
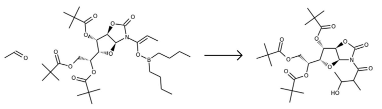
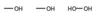
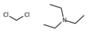
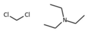
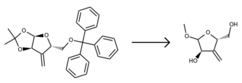
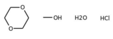
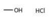
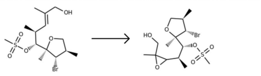
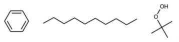
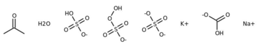
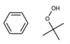
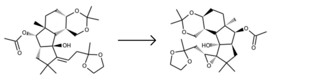
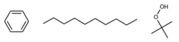
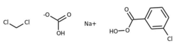
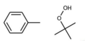
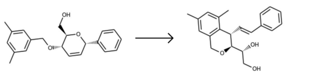
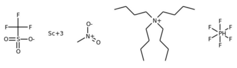
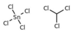
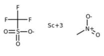
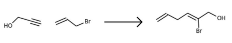
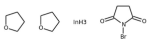
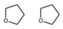
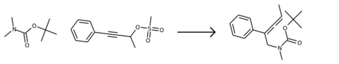
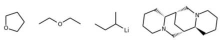
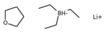
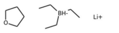
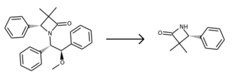
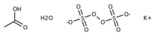
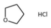
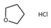
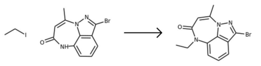
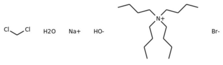
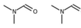
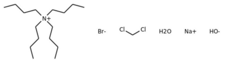
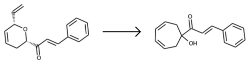
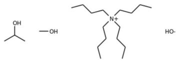
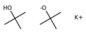
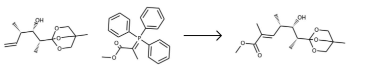
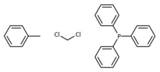
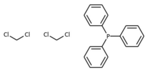
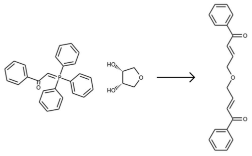
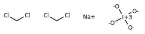
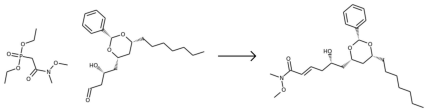
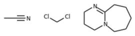
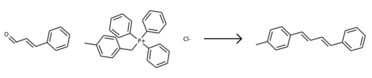
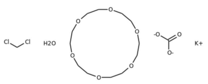
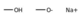
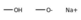
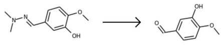
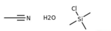
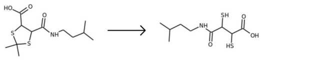
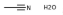
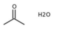
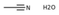
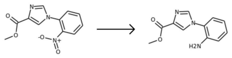
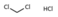
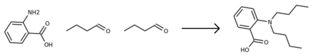
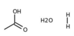
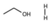
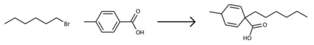
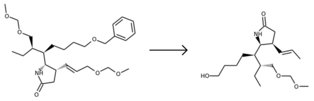
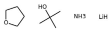
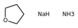
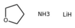
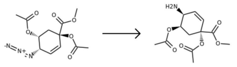
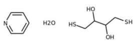
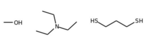
9336462					20.0	32.6	33.0
41425736		 Reaxys ID 13154734	 Reaxys ID 13154734	 Reaxys ID 13154734	20.0	33.5	33.5
38554472		 Reaxys Name boron trifluoride diethyl etherate	 Reaxys Name boron trifluoride diethyl etherate	 Reaxys Name boron trifluoride diethyl etherate	N/A	123.1	123.1
9671652					N/A	189.9	152.5
27763214					20.0	7.2	7.2
5052676					0.0	-7.8	-7.8
297394					N/A	78.1	77.4
281216				Ni	200.0	259.9	233.1
5186063					N/A	36.4	36.4
24973931					N/A	16.9	16.9
27977943					N/A	149.4	149.4
1654607					N/A	123.7	123.7
8841633		Na+ HO-			N/A	-48.9	26.0
23314056					N/A	37.4	49.0
6963834					N/A	-12.7	-12.7
9899404		 Reaxys Name silica gel	Reaxys ID 19278319		N/A	105.5	96.6
5127331					0.0	27.4	30.9
38768506					70.0	45.6	34.0
31607336			 Reaxys Name boron trifluoride diethyl etherate		20.0	18.9	22.8
1852491		 Reaxys Name lithium bromide		 Reaxys Name potassium bromide	98.0	83.4	169.9
11221187			Reaxys Name boron trifluoride diethyl etherate		20.0	-2.0	12.5
2383147					20.0	42.3	42.3
4492881		 Reaxys Name palladium on activated charcoal		 Reaxys Name palladium on activated charcoal	40.0	18.8	21.1
4233762					N/A	2.5	11.3

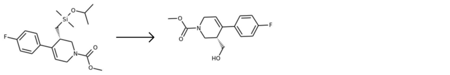
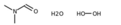
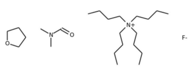
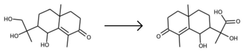
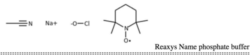
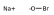
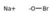
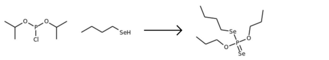
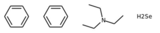
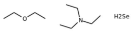
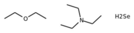
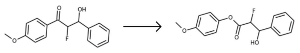
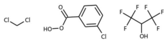
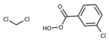
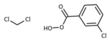
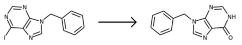
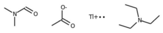
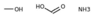
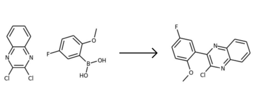
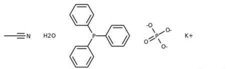
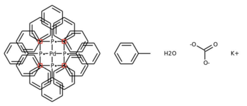
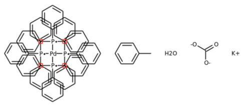
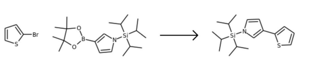
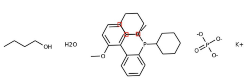
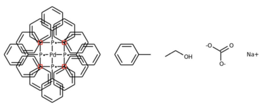
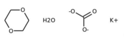
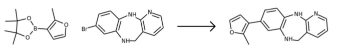
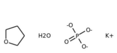
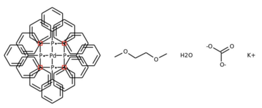
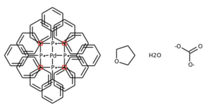
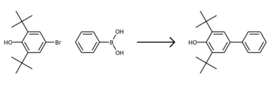
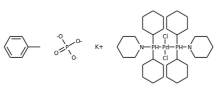
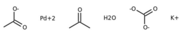
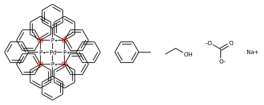
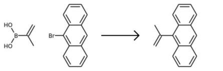
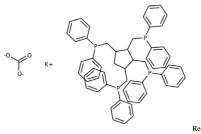
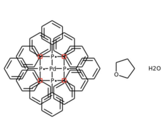
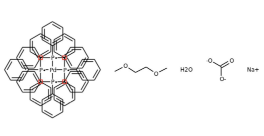
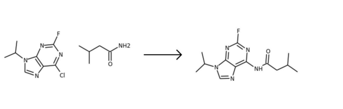
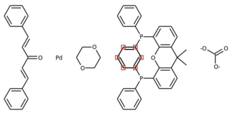
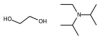
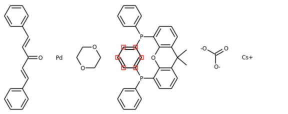
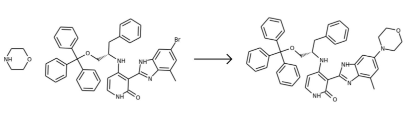
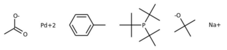
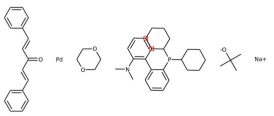
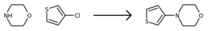
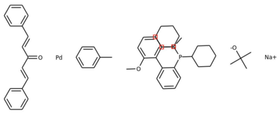
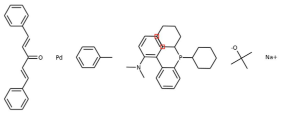
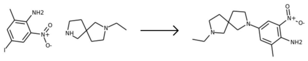
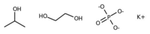
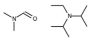
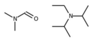
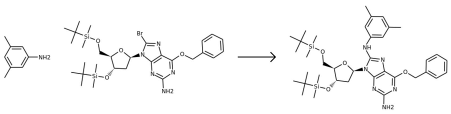
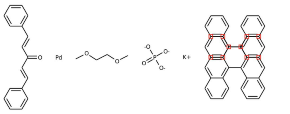
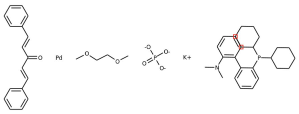
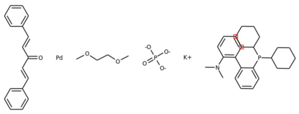
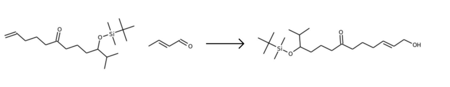
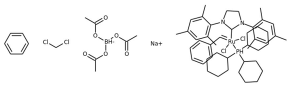
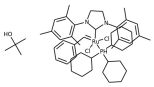
36324962					85.0	98.8	101.2
1986429					0.0	10.6	-1.9
23851486					N/A	42.9	68.6
5337948					100.0	23.6	23.6
10505025					-40.0	75.7	75.7
37060180					20.0	10.9	15.3
5327471					N/A	8.7	13.5
554534					N/A	115.6	122.6
461707					N/A	35.1	146.1
11242864					N/A	29.7	29.7
24960905					N/A	80.0	80.0
30084950					20.0	16.7	16.7
4504557		 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	N/A	2.8	2.8
30102370			 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	110.0	95.8	95.8
3816753					N/A	82.7	87.5
5210772		 Reaxys Name boron trifluoride diethyl etherate	 Reaxys Name boron trifluoride diethyl etherate	 Reaxys Name boron trifluoride diethyl etherate	0.0	11.4	11.4
2176468					N/A	99.7	98.1
5139901					N/A	100.6	100.6
29034611					20.0	22.4	22.4
1387456					N/A	62.1	62.1
10149615					20.0	21.0	13.9
29237461					20.0	15.6	15.6
22880776					20.0	14.6	14.6
32792368					-30.0	52.5	16.2

2255356					299.9	263.1	263.1
4667439					N/A	-17.4	-17.4
39121363			Reaxys Name sodium bis(2-methoxyethoxy)aluminium dihydride	Reaxys Name sodium bis(2-methoxyethoxy)aluminium dihydride	N/A	65.8	65.8
42279161					20.0	22.3	19.9
1821342		H2O K+ HO-	AlH7- Li+	K+ HO-	100.0	10.5	103.5
2170049					30.0	35.0	35.0
28774608			NaH	 K+	-78.0	56.2	16.7
1320851					N/A	51.6	32.7
9372184					20.0	21.2	21.2
40224906		 Reaxys Name O-(benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium tetrafluoroborate	 Reaxys Name O-(benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium tetrafluoroborate	 Reaxys Name O-(benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium tetrafluoroborate	27.5	21.1	21.1
29311031					120.0	98.3	97.4
40807619					120.0	64.6	64.6
27931253		 Reaxys Name sodium hydride		Reaxys Name sodium hydride	-20.0	10.4	9.4
39062813			 Reaxys Name dirhodium tetraacetate		0.0	19.8	6.1
4410590			AlH7- Li+		N/A	12.1	36.3
27978248					N/A	56.1	45.2
27784122					130.0	47.3	76.0
28190848					20.0	39.7	39.7
24898802					70.0	87.0	87.0
609622			Na+ HO-	Na+ HO-	180.0	41.4	41.4
884394					70.0	62.9	62.9
7050517					255.0	49.4	49.4
5021035					80.0	49.3	49.3
26014672		 Reaxys Name calcium chloride			N/A	180.7	113.6
37452339					140.0	125.2	125.2

9905540					20.0	17.1	17.1
28176521					-40.0	-25.5	-25.5

Supplementary Table 5. 40 reactions from eleven reaction types randomly chosen from the test dataset that has the least number of correctly or similarly predicted elements

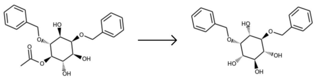
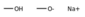
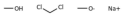
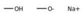
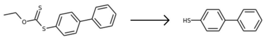
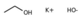
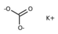
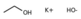
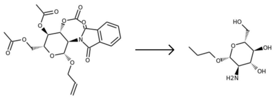
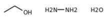
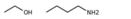
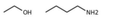
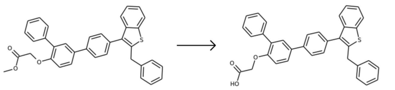
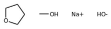
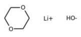
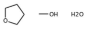
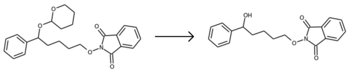
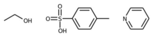
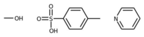
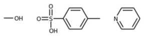
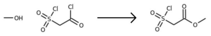
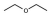
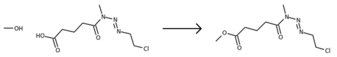
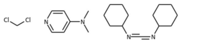
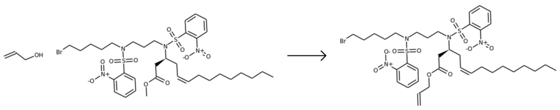
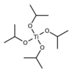
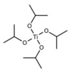
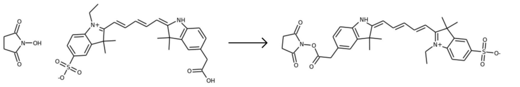
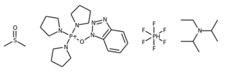
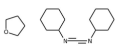
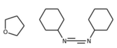
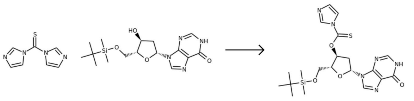
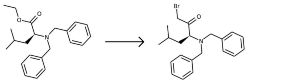
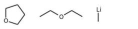
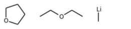
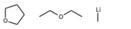
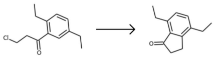
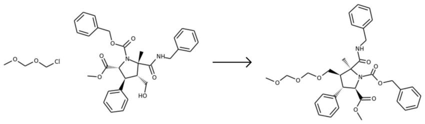
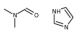
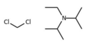
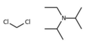
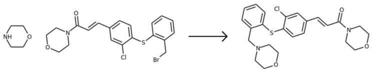
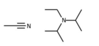
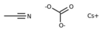
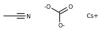
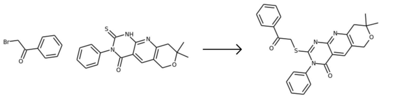
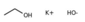
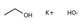
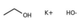
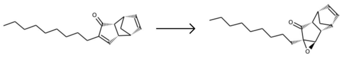
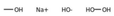
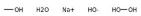
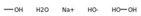
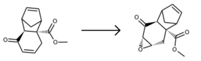
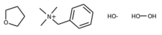
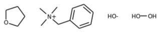
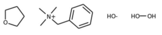
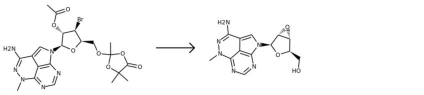
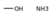
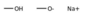
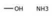
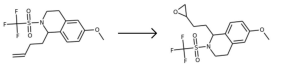
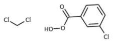
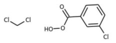
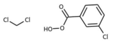
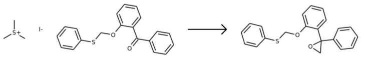
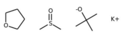
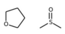
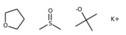
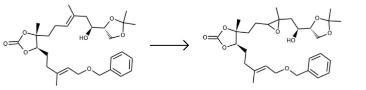
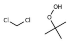
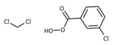
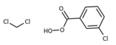
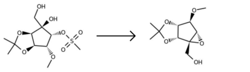
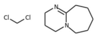
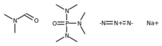
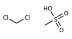
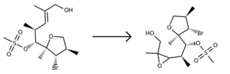
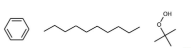
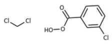
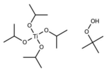
Reaxys ID	Reaction	True Context	Top Prediction	Closest Prediction	True Temperature/°C	Top Predicted Temperature/°C	Closest Predicted Temperature/°C	Reaction type
8542026			Reaxys Name Lindlar's catalyst		-5.0	25.0	29.1	Reduction, Hydrolysis
5342174			BH4- Na+ Reaxys Name ozone	BH4- Na+ Reaxys Name ozone	20.0	-65.7	-65.7	Hydrolysis
5240460					N/A	48.4	48.4	Hydrolysis
8538458		 Reaxys Name phosphate buffer			-39.0	14.4	14.4	Condensation, Hydrolysis, Oxidation
8651198					N/A	25.5	27.1	Hydrolysis
10069687		 Reaxys Name bis(acetylacetonate)oxovanadium		 Reaxys Name bis(acetylacetonate)oxovanadium	25.0	15.2	48.5	Epoxidation
9925682		 Reaxys Name bis(acetylacetonate)oxovanadium		 Reaxys Name bis(acetylacetonate)oxovanadium	N/A	6.0	34.3	Sharpless epoxidation
29729359					10.0	25.7	60.8	Friedel-Crafts alkylation
8529950			Reaxys Name copper(I) chloride	 Reaxys Name copper(I) chloride	20.0	52.8	7.6	Alkylation, allylindation, Bromination
8526713					-78.0	-14.9	-14.9	deprotonation, Addition, Alkylation
8592120					90.0	28.3	28.3	dealkylation
8629688			 Reaxys Name sodium hydride		20.0	13.4	24.6	Alkylation
34446920			Reaxys Name samarium diiodide		20.0	-9.7	26.5	[2,3]-Wittig Rearrangement
33420713		 Reaxys Name ozone	Reaxys Name ozone	 Reaxys Name ozone	-78.0	-58.5	-70.5	Wittig reaction
28838907		 Reaxys Name silica gel		Reaxys Name 4 A molecular sieve	20.0	21.2	8.3	Wittig reaction
29349499		 Reaxys Name lithium chloride	Reaxys Name sodium hydride	Reaxys Name sodium hydride	20.0	7.4	7.4	Wittig-Horner reaction
35093977					20.0	17.6	17.6	Wittig Olefination
5266589		 Reaxys Name sodium iodide	Reaxys Name silica gel.Reaxys ID 13273590	Reaxys Name tin(II) chloride.Reaxys Name silica gel	N/A	46.2	40.0	deprotection
8780005		 Reaxys Name mercury dichloride		 Reaxys Name mercury dichloride	22.0	32.3	38.4	deprotection
8585244		 Reaxys Name tin(II) chloride	Reaxys Name palladium 10 on activated carbon	Reaxys Name palladium 10 on activated carbon	20.0	24.8	24.8	Reduction
8692625		 Reaxys Name palladium on activated charcoal		 Reaxys Name palladium on activated charcoal	100.0	87.6	42.0	Formylation, reduction
29313449		H2O H2O NH3 LiH	LiH	NH3 LiH	-60.0	-3.6	-60.5	Birch reduction
42522603					-63.0	-68.8	-58.8	Birch reduction
8646319					20.0	25.3	25.3	Reduction

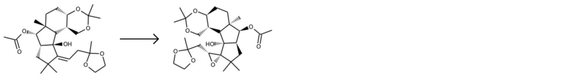
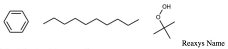
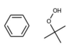
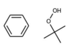
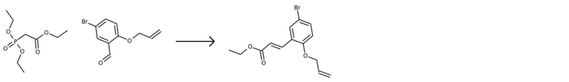
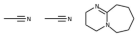
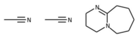
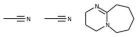
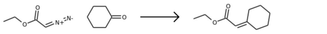
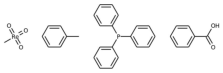
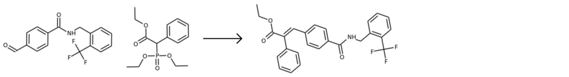
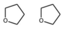
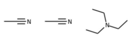
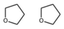
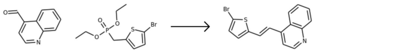
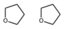
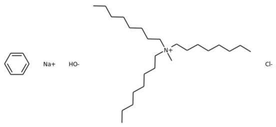
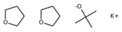
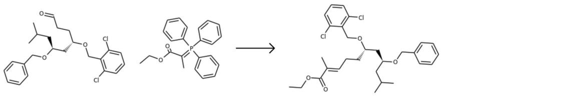
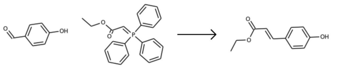
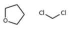
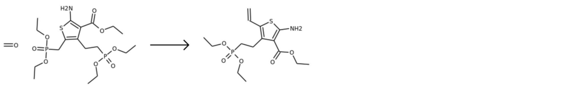
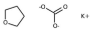
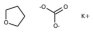
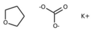
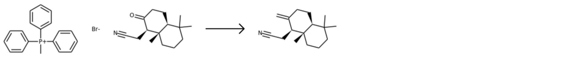
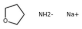
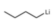
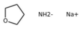
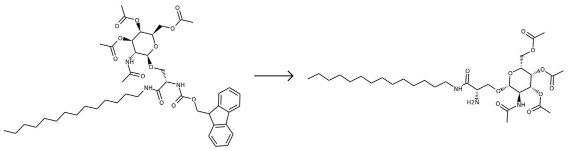
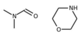
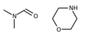
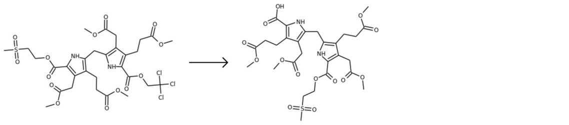
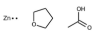
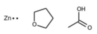
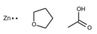
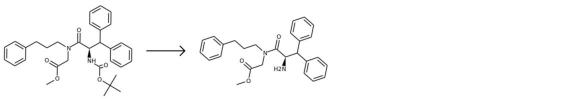
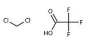
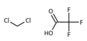
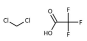
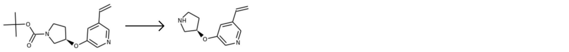
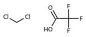
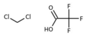
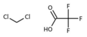
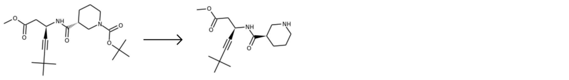
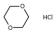
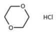
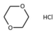
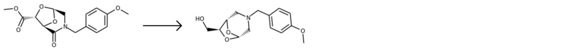
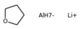
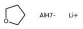
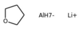
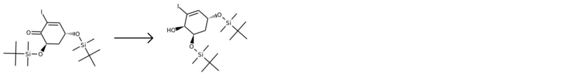
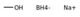
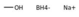
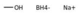
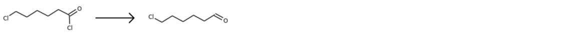
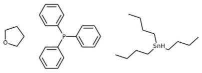
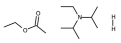
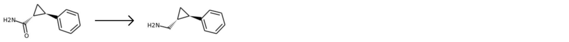
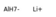
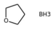
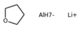
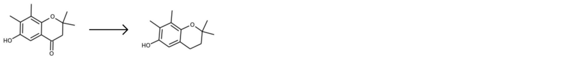
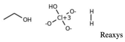
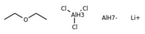
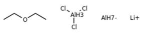
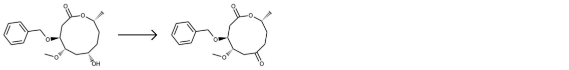
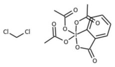
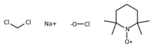
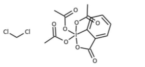
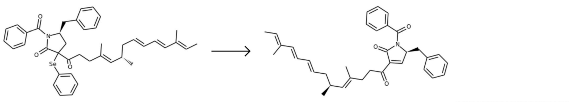
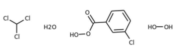
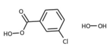
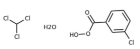
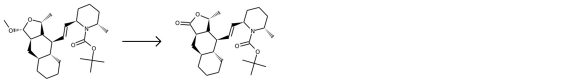
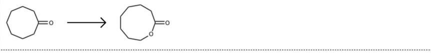
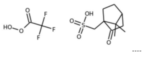
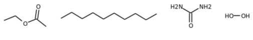
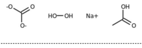
10255318		 Reaxys Name potassium fluoride		 Reaxys Name cesium fluoride	25.0	19.0	31.9	Tamao oxidation
5351285		 Reaxys Name phosphate buffer			55.0	17.0	17.0	Oxidation
8520314					20.0	20.5	20.5	Substitution, Oxidation
9273620		 Reaxys Name phosphate buffer	 Reaxys Name boron trifluoride diethyl etherate	 Reaxys Name boron trifluoride diethyl etherate	20.0	-4.6	-4.6	Baeyer-Villiger oxidation
3627506		 Reaxys Name trans-di(?-acetato)bis[o-(di-o-tolylphosphino)benzyl]dipalladium(II)	H2O	 Reaxys Name palladium on activated charcoal	80.0	89.1	44.4	Oxidation
26007660		 Reaxys Name palladium diacetate			60.0	95.2	95.2	Suzuki coupling
10342169		 Reaxys Name palladium diacetate		 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	100.0	92.5	90.7	Suzuki-Miyaura cross-coupling
42866790		 Reaxys ID 23164488			50.0	102.9	97.7	Suzuki coupling
29522485					80.0	50.3	88.1	Suzuki-Miyaura cross-coupling reaction
9645904		 Reaxys Name bis(?3-allyl-?-chloropalladium(II)).Reaxys Name xylene	 Reaxys Name sodium hydroxide		130.0	60.2	78.5	Suzuki reaction
11152198					110.0	134.9	111.2	Buchwald amidation
23567330					100.0	73.7	100.1	Buchwald-Hartwig coupling
9293801					100.0	126.4	95.9	Buchwald-Hartwig coupling
25897520		 Reaxys ID 11405916			90.0	92.9	92.9	Buchwald reaction
25936014					80.0	67.1	67.1	Buchwald-Hartwig reaction
36087581					N/A	-3.3	14.7	Grubbs Olefin Metathesis

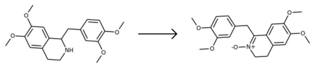
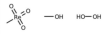
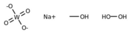
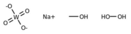
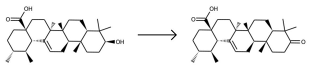
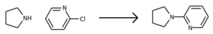
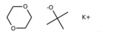
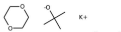
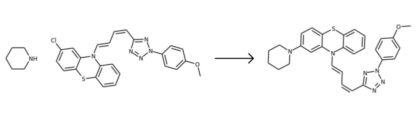
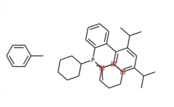
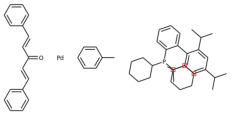
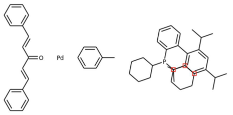
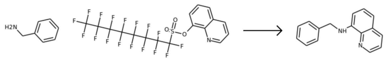
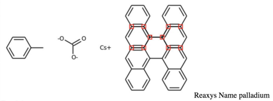
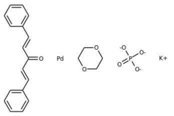
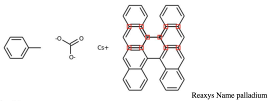
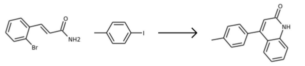
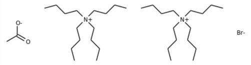
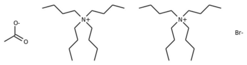
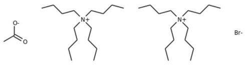
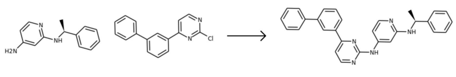
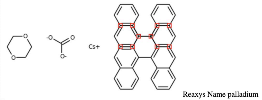
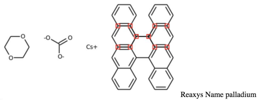
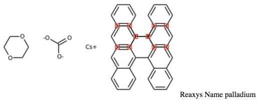
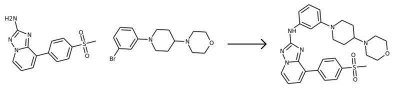
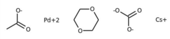
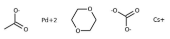
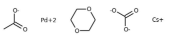
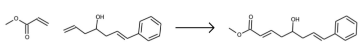
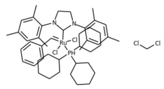
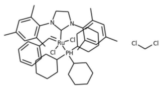
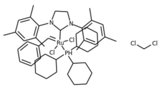
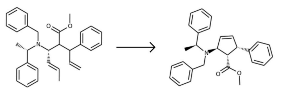
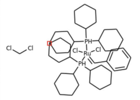
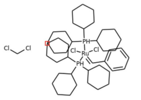
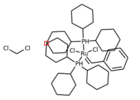
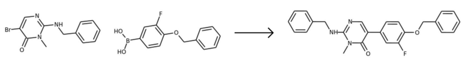
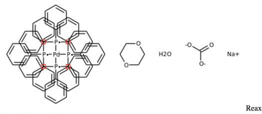
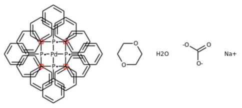
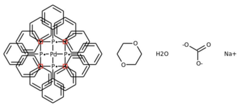
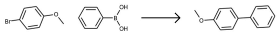
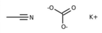
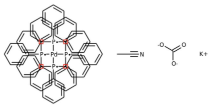
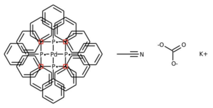
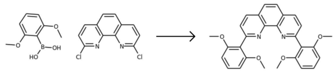
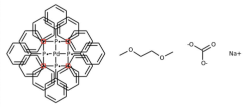
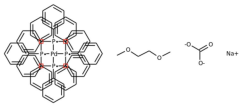
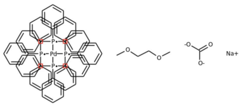
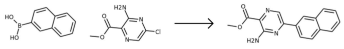
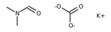
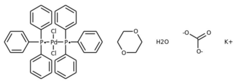
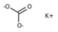
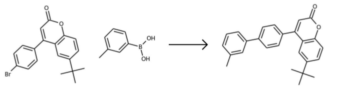
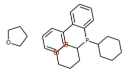
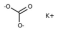
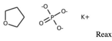
Comparison with the neighbor approach

The nearest-neighbor method searches the training set for the reaction with the maximum similarity to the reaction for prediction. Cosine similarity of the reaction fingerprint[1] is used to quantify reaction similarity. A thorough search is performed for the 62 reactions in Supplementary Table 3. The top-ten most similar reactions are retrieved, and the top-one prediction and the prediction within top-ten that has the maximum elements matching the true condition are listed in Supplementary Table 6. Good suggestions are found for a majority of the cases, and the overall accuracy for these 62 reactions is comparable to the neural network model. However, the search for one reaction takes ~40 minutes on a single intel® Xeon(R) CPU E5-2690 0 @2.90GHz, as compared to ~100ms for the neural network approach running on the same machine, which is over 10,000 times faster. Further, the nearest neighbor method proposes conditions by simply copy and paste, which does not have the ability to infer missing information.

Supplementary Table 6. Nearest neighbor predictions for the 62 reaction examples in Supplementary Table 3

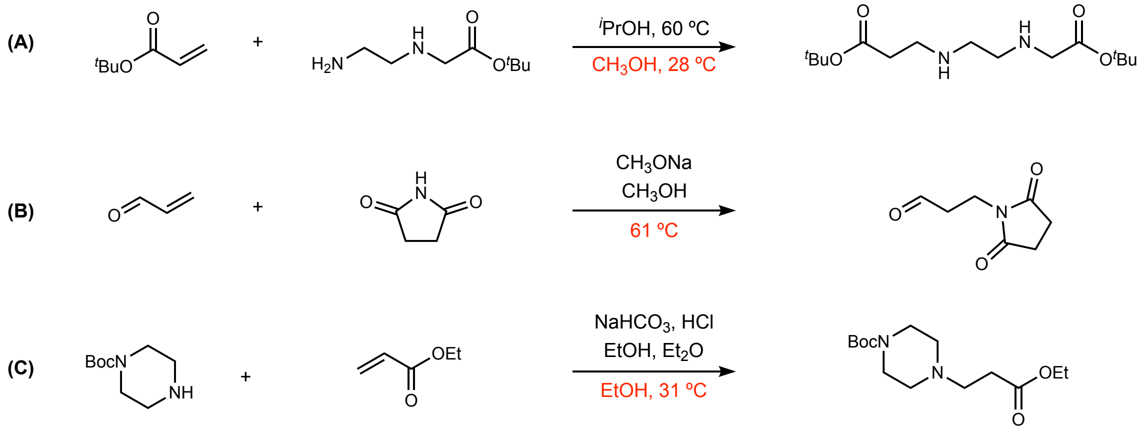
Reaxys ID	Reaction	True Context	Top Prediction	Closest Prediction	True Temperature/ °C	Top Predicted Temperature/ °C	Closest Predicted Temperature/ °C
5301921					20.0	20.0	12.0
8712792					N/A	N/A	80.0
5261303					N/A	N/A	N/A
8655717				 Reaxys Name lithium hydroxide monohydrate	N/A	20.0	20.0
5302180					57.5	N/A	N/A
53162					0.0	0.0	0.0
5185761					10.0	N/A	N/A
8698819					95.0	N/A	N/A
8669084					N/A	N/A	N/A
5336683					20.0	80.0	80.0
5220300					-78.0	-78.0	-78.0
2820838					54.5	95.0	95.0
8568283					N/A	0.0	0.0
8738792					20.0	35.0	35.0
8574982					40.0	40.0	40.0
8591626					10.0	0.0	0.0
8647861					0.0	0.0	0.0
8655813					20.0	N/A	20.0
8720054					20.0	20.0	20.0
8560867			 Reaxys Name sodium hydride		0.0	N/A	0.0
8598928		 Reaxys Name bis(acetylacetonate)oxovanadium			0.0	N/A	N/A
5229646					0.0	110.0	N/A
10069687		 Reaxys Name bis(acetylacetonate)oxovanadium			25.0	-10.0	N/A

9925682		 Reaxys Name bis(acetylacetonate)oxovanadium	 Reaxys Name bis(acetylacetonate)oxovanadium	 Reaxys Name bis(acetylacetonate)oxovanadium	N/A	N/A	N/A
40053563		 Reaxys Name lithium bromide		 Reaxys Name lithium bromide	20.0	20.0	20.0
1558712					110.0	80.0	110.0
41951969		 Reaxys Name sodium hydride	 Reaxys Name lithium chloride	 Reaxys Name sodium hydride	0.0	20.0	20.0
9299119		 Reaxys Name sodium hydride			N/A	N/A	-15.0
28171183					10.0	10.0	10.0
28476112					20.0	20.0	20.0
9240363					20.0	20.0	20.0
2287762					20.0	N/A	20.0
8659689					20.0	N/A	20.0
5357436					20.0	20.0	20.0
5304351					N/A	N/A	N/A
8588894					N/A	20.0	20.0
5351204					N/A	N/A	N/A
5297919					N/A	N/A	N/A
31266961		 Reaxys Name cerium(III) chloride heptahydrate	 Reaxys Name cerium(III) chloride heptahydrate	 Reaxys Name cerium(III) chloride heptahydrate	-78.0	-78.0	-78.0
5265062		 Reaxys Name indium(III) chloride	Reaxys Name palladium-barium carbonate.Reaxys Name xylene	 Reaxys Name palladium on activated charcoal	-30.0	150.0	25.0
8619786					N/A	35.0	N/A
5261651		 Reaxys Name palladium on activated charcoal			20.0	-5.0	-5.0
28315260					10.0	N/A	20.0
5307293			 Reaxys Name chloroform-d1		-25.0	N/A	N/A
9228133		Reaxys Name jones reagent	Reaxys Name jones reagent	Reaxys Name jones reagent	20.0	20.0	20.0
116049					N/A	20.0	12.5

5081868					20.0	0.0	0.0
339468		Reaxys Name Jone's reagent.Reaxys Name silica gel	Reaxys Name Jone's reagent	Reaxys Name Jone's reagent	0.0	N/A	N/A
4977260		 Reaxys ID 21565323		 Reaxys ID 21565323	80.0	N/A	90.0
34039114					95.0	105.0	105.0
11260415		 Reaxys Name palladium diacetate		 Reaxys Name palladium diacetate	90.0	100.0	85.0
11077029		 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	120.0	120.0	120.0
26046174		 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	 Reaxys Name palladium diacetate	90.0	90.0	90.0
33823237					80.0	80.0	80.0
9603192					19.0	19.0	19.0
36032266					40.0	40.0	40.0
25840737		 Reaxys Name lithium chloride	 Reaxys Name lithium chloride	 Reaxys Name lithium chloride	100.0	100.0	100.0
4033276		 Reaxys Name palladium diacetate			8.0	120.0	120.0
9208506					N/A	N/A	N/A
28885714		 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride		 Reaxys Name (1,1'-bis(diphenylphosphino)ferrocene)palladium(II) dichloride	N/A	N/A	N/A
10040528		 Reaxys Name palladium diacetate.Reaxys Name potassium fluoride		 Reaxys ID 23164488	20.0	20.0	50.0

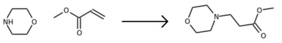
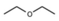
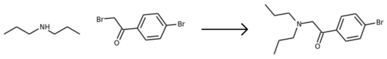
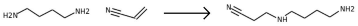
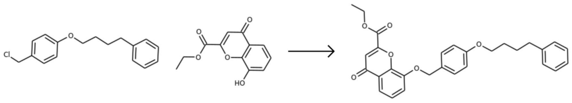
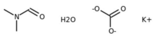
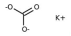
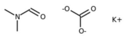
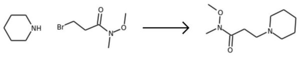
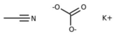
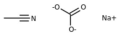
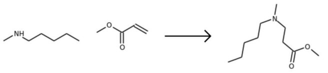
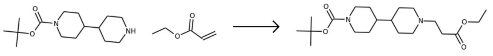
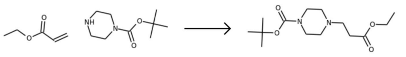
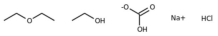
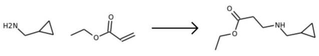
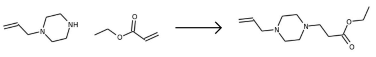
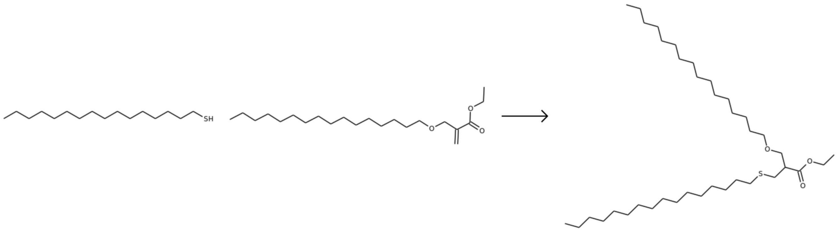
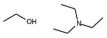
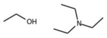
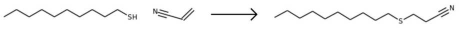
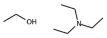
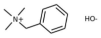
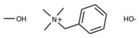
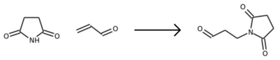
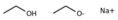
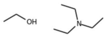
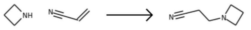
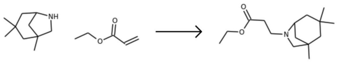
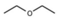
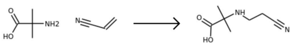
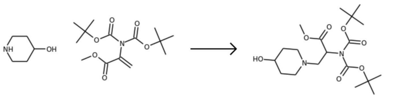
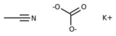
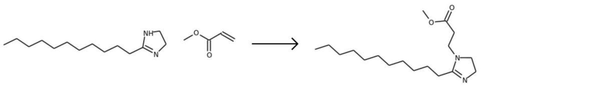
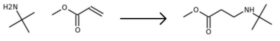
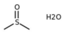
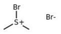
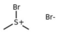
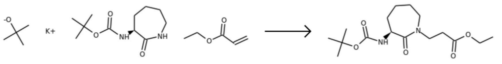
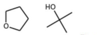
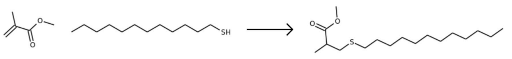
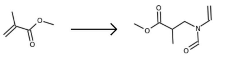
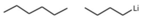
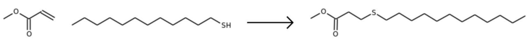
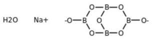
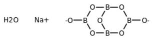
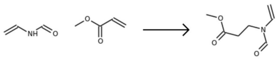
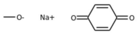
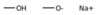
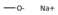
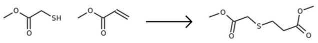
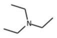
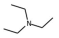
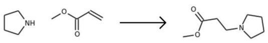
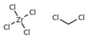
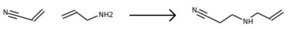
Evaluation of the Michael additions used by Marcou et al. [2]

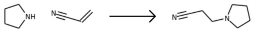
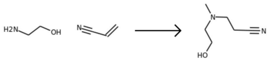
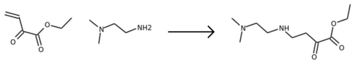
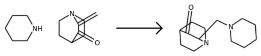
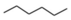
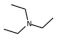
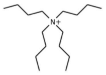
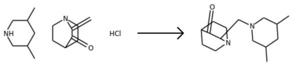
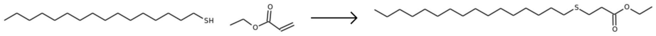
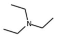
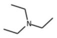
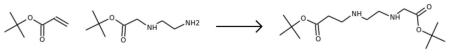
Among the 52 reactions, 34 of them are found in the final dataset, and we have detailed information about these reactions for evaluation. The model is retrained with these reactions excluded from the training set. The top-ten accuracy of these 34 reactions for similar matches is 47.0%, and the accuracy for similarly matching catalyst, solvent 1 and reagent 1 is 55.9%, lower than the overall accuracy of the entire dataset, yet a significant improvement over literature results.[2] Furthermore, it has been pointed out that many of these reactions can occur under different conditions, meaning some predictions are not necessarily wrong, even when they differ from the recorded context. For example, Supplementary Figure 5 shows reactions where the exact recorded condition is not predicted. The first example has a different solvent predicted, but it is similar to the recorded solvent (both are alcohols).[3] For the second example, although the model does not suggest the correct solvent and reagent in the first choice, it recognizes the need for basic conditions in the subsequent suggestions, with the second suggestion being piperidine and the third being triethylamine.[4] Supplementary Figure 5(C) is another example of data quality that complicates the analysis. Two reagents $NaHCO_3$ and HCl are not commonly used in the same reactions but in separate steps which is the case in the reported procedure.[5] Additionally, the true solvent used for the reaction is ethanol where diethyl ether is used as an extraction solvent in workup. The full prediction results are presented in Supplementary Table 7.



Supplementary Figure 5. Examples of Michael additions where none of the recorded c, s1 or r1 are predicted. A) Methanol is predicted, which is similar to the recorded solvent (isopropanol); B) the top prediction is incorrect, but subsequent predictions suggest bases as reagents (the second suggestion is piperidine and the third is triethylamine); C) the reaction has incompatible reagents that are used in different stages but not reflected in the reaction record.

Supplementary Table 7. 34 examples of Michael additions that are used for evaluation in the work of Marcou et al.

Reaxys ID	Reaction	True Context	Top Prediction	Closest Prediction	True Temperature/°C	Top Predicted Temperature/°C	Closest Predicted Temperature/°C
1493719					20.0	23.7	23.7
635835					N/A	34.6	52.9
1288921					20.0	29.9	29.9
24352854		 H2O K+	 K+ Reaxys Name potassium iodide	 K+	N/A	67.4	57.3
8861928		 K+		 Na+	N/A	18.6	29.6
23748241					10.0	26.3	26.3
25669226					N/A	57.2	57.2
23227051		 Na+ HCl			N/A	30.7	30.7
24273953					N/A	36.3	36.3
24717084					N/A	34.9	34.9
24971546					N/A	46.9	50.0
694909			 HO-	 HO-	30.0	35.6	26.5
46547		 Na+			N/A	60.8	60.8
1456140					20.0	37.4	37.4
24879120					N/A	79.5	79.5
2924844		H2O Na+ HO-		H2O Na+ HO-	0.0	51.9	16.7
22879782			 K+		N/A	43.4	48.4
24889073					N/A	43.7	43.7
1284594		 H2O	 Br-	 Br-	20.0	14.3	14.3
24576080		 Reaxys Name ammonium chloride			N/A	-3.6	-3.6
3363418					N/A	61.1	29.8
24576110					N/A	73.3	73.3
701198		 H2O Na+		 H2O Na+	20.0	47.0	27.6
24581594		 Na+	 Na+	 Na+	N/A	19.6	36.2
649317					20.0	25.9	25.9
4051					20.0	28.6	28.6
1297143					10.0	24.0	24.0

4046					90.0	33.8	33.8
35987442					50.0	57.8	57.8
24722182					N/A	32.4	26.5
24503535			H2O	H2O	N/A	62.4	62.4
23567748	H2O   Br-		—OH	H2O Na+ HO-	80.0	38.3	41.3
24960665					N/A	66.9	44.4
23665071			—OH	—OH	60.0	28.4	28.4

Feature definition used in the Morgan fingerprints

Supplementary Table 8. Feature definition as defined by Gobbi et al. [6]

Invariants	SMARTS
Hydrogen bond donor	[[N;!H0;v3],[N;!H0;+1;v4],[O,S;H1;+0],[n;H1;+0]]
Hydrogen bond acceptor	[$([O,S;H1;v2]−[!$(*4[O,N,P,S])]),[O,S;H0;v2],[O,S;−],$([N&v3;H1,H2]−[!$(*4[O,N,P,S])]), [N;v3;H0],[n,o,s;+0],F]
Basic group	[$([N;H2&+0][[C,a];!$([C,a](4O))]),$([N;H1&+0]([[C,a]; !$([C,a](4O))])[[C,a];!$([C,a](4O))]),$([N;H0&+0]([C;!$(C(4O))])([C;!$(C(4O))])[C;!$(C(4O))]),[N,n;X2;+0]]
Hydrophobic group	[$([C;H2,H1](!4*)[C;H2,H1][C;H2,H1][[C;H1,H2,H3];!$(C4*)]),$(C([C;H2,H3])([C;H2,H3])[C;H2,H3])]
Acidic group	[O;H1]−[C,S](4[O,S,P])
Halogen	[F,Cl,Br,I]
Attachment point to aliphatic ring	[$([A;D3](@*)(@*)~*)]
Attachment point to aromatic ring	[$([a;D3](@*)(@*)*)]
Any unusual atom (not H,C,N,O,F,S,Cl,Br,I)	[!#1;!#6;!#7;!#8;!#9;!#16;!#17;!#35;!#53]

Reference

(1) Schneider, N.; Lowe, D. M.; Sayle, R. A.; Landrum, G. A. Development of a Novel Fingerprint for Chemical Reactions and Its Application to Large-Scale Reaction Classification and Similarity. *J. Chem. Inf. Model.* **2015**, *55* (1), 39–53.

(2) Marcou, G.; Aires de Sousa, J.; Latino, D. A. R. S.; de Luca, A.; Horvath, D.; Rietsch, V.; Varnek, A. Expert System for Predicting Reaction Conditions: The Michael Reaction Case. *J. Chem. Inf. Model.* **2015**, *55* (2), 239–250.

(3) Alihodzic Suleijam; Lazarevski Gorjana. 4" Amino Linked Macrolides Useful for the Treatment of Microbial Infections. WO2006120545, 2006.

(4) Moe, O. A.; Warner, D. T. 1, 4 Addition Reactions. III. The Addition of Cyclic Imides to α-β-Unsaturated Aldehydes. A Synthesis of β-Alanine Hydrochloride. *J. Am. Chem. Soc.*

1949, *71* (4), 1251–1253.

(5) Burkholder, Timothy P.; Maynard, George D.; Kudlacz, E. M. . Carboxy Substituted Acylic Carboxamide Derivatives. 09/648,005, 1999.

(6) Gobbi, A.; Poppinger, D. Genetic Optimization of Combinatorial Libraries. *Biotechnol. Bioeng.* **1998**, *61* (1), 47–54.