Supporting information

Intermediate Knowledge Enhanced the Performance of Amide Coupling Yield Prediction Model

Chonghuan Zhang[1‡], Qianghua Lin[1‡], Chenxi Yang[2], Yaxian Kong[2], Zhunzhun Yu[1]*, Kuangbiao Liao[1]*

[1] Guangzhou National Laboratory, No. 9 XingDaoHuanBei Road, Guangzhou International Bio Island, Guangzhou 510005, Guangdong, PR China

[2] AIChemEco Inc., Guangzhou 510005, Guangdong, PR China

Content

S1 General Information

Reagents and solvents were obtained from Energy Chemical, Aladdin, ACMEC or Bidepharm and used as received. Reactions were monitored on NuoTai GF254 TLC plates and detection was done by irradiation with UV light (254 nm or 366 nm). Reaction plate set up and sample preparation for TOF-DESI injection plate and LC-MS injection plate were performed using EVO-200 (Tecan, Swiss) Consisting of MCA-96 lead plane, an arm air LiHa first 8-channel, a worktable standard coated FDM. MCA-96 enabled the simultaneous transfer of 96 samples under the same conditions (speed of aspiration and dispensing, height of pipetting at source and destination positions, pattern of pipetting, etc.). Although the 8-channel head treated less sample simultaneously, it provided more flexibility in terms of volumes transferred, layout of source and destination plates, pipetting height, and speed. Ultra Performance Liquid Chromatography (UPLC) was performed on a Waters UPLC system consisting of a quaternary Solvent manager, a M5KEOE column oven, a PDA eλ UV/Vis detector, a BEH C18 column and HPLC grade solvents (acetonitrile, methanol and i-propanol) from Fisher Scientific. Desorption Electrospray Ionization Mass (DESI) mass spectra was measured on a XS+Xevo G2-XS Tof instrument. The found masses from high resolution measurements (DESI-TOF-MS) are reported in m/z units with M as the molecular ion. ^{1}H NMR and ^{13}C NMR spectra were measured on a Bruker AVANCE NEO 600 (600 MHz) and Bruker Avance Neo 151 MHz NMR spectrometer. Deuterated solvents were obtained from Energy Chemical and used without further purification. The chemical shifts are given in parts per million (ppm) on the delta scale (δ) relative to tetramethylsilane as external standard. The complete high-throughput experimental workflow is presented in Figure S1.

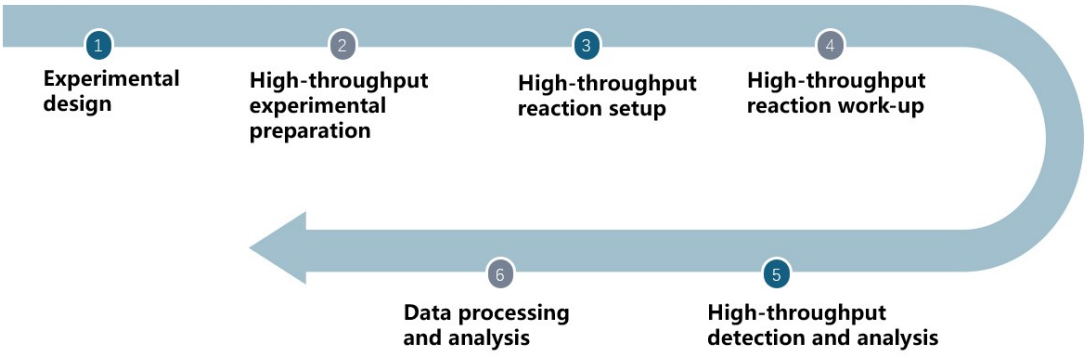


Figure S1. High-throughput experimental workflow

S2 High-throughput experimentation (HTE)

S2.1 Conditions list for HTE

We collected and prepared the conditions of acylation for high-throughput experimentation, the conditions list was illustrated as following (Table S1).

Table S1. Conditions list

Entry	Activation reagents	Reagents (Equiv)	Additives	Additive (Equiv)	Base	Base (Equiv)	Solvent
1	HATU	1.50			DIPEA	3	DMF
2	TFFH	1.50			DIPEA	3	DMF
3	BTFFH	1.50			DIPEA	3	DMF
4	2,4-Dichloro-6-Methoxy-1,3,5-Triazine	1.50			DIPEA	3	DMF
5	COMU	1.50			2,6-Lutidine	3	DMF
6	TBTU	1.50			DIPEA	3	DMF
7	HDMC	1.50			DIPEA	3	DMF
8	TCTU	1.50			DIPEA		DMF
9	TDBTU	1.50			DIPEA	3	DMF
10	TNTU	1.50			DIPEA	3	DMF
11	TPTU	1.50			DIPEA	3	DMF
12	TSTU	1.50			DIPEA	3	DMF
13	EDC-HCl	1.50	HOBt	1.50	DIPEA	3	DMF
14	BOP	1.50			DIPEA	3	DMF
15	PyClOP	1.50	HOBt	1.50	DIPEA	3	DMF
16	PyClOP	1.50			DIPEA	3	DMF
17	DEPC	1.50			DIPEA	3	DMF
18	FDPP	1.50			DIPEA	3	DMF
19	HATU	1.50	HOAT	3.00	DIPEA	3	DMF
	Control well (same with 1)						
20	HBTU	1.50	HOBT	1.50	DIPEA	3	DMF
21	HBTU	1.50			DIPEA	3	DMF

22	HCTU	1.50			DIPEA	3	DMF
23	TOTU	1.50			DIPEA	3	DMF
24	EDC-HCl	1.50	2,4,5-Trichlorophenol	1.50	DIPEA	3	DMF
25	EDC-HCl	1.50	Pentafluorophenol	1.50	DIPEA	3	DMF
26	EDC-HCl	1.50	N-Hydroxysuccinimide	1.50	DIPEA	3	DMF
27	EDC-HCl	1.50	N-Hydroxyphthalimide	1.50	DIPEA	3	DMF
28	DCC	1.50	2,4,5-Trichlorophenol	1.50	DIPEA	3	DMF
29	DCC	1.50	Pentafluorophenol	1.50	DIPEA	3	DMF
30	DCC	1.50	N-Hydroxysuccinimide	1.50	DIPEA	3	DMF
31	DCC	1.50	N-Hydroxyphthalimide	1.50	DIPEA	3	DMF
32	PyAOP	1.50			DIPEA	3	DMF
33	PyAOP	1.50	HOAt	1.50	DIPEA	3	DMF
34	PyBOP	1.50			DIPEA	3	DMF
35	PyBOP	1.50	HOBt	1.50	DIPEA	3	DMF
36	PyOxim	1.50			DIPEA	3	DMF
37	PyBrOP	1.50	HOBt	1.50	DIPEA	3	DMF
38	PyBrOP	1.50			DIPEA	3	DMF
39	PyClocK	1.50			DIPEA	3	DMF
40	EEDQ	1.50			DIPEA	3	DMF
41	DEPBT	1.50			DIPEA	3	DMF
42	DPPCl	1.50			DIPEA	3	DMF
43	DMTMM	1.50			DIPEA	3	DMF
44	PyCIU	1.50			DIPEA	3	DMF

45	BOPCl	1.50			DMAP	1.5	DMF
46	EDC-HCl	1.50			NMM	3	DMF
47	TFFH	1.50			NMM	3	DMF
48	BTFFH	1.50			NMM	3	DMF
49	2,4-Dichloro-6-Methoxy-1,3,5-Triazine	1.50			NMM	3	DMF
50	EDC-HCl	1.50	2,4,5-Trichlorophenol	1.50	NMM	3	DMF
51	EDC-HCl	1.50	Pentafluorophenol	1.50	NMM	3	DMF
52	EDC-HCl	1.50	N-Hydroxysuccinimide	1.50	NMM	3	DMF
53	EDC-HCl	1.50	N-Hydroxyphthalimide	1.50	NMM	3	DMF
54	DCC	1.50	2,4,5-Trichlorophenol	1.50	NMM	3	DMF
55	DCC	1.50	Pentafluorophenol	1.50	NMM	3	DMF
56	DCC	1.50	N-Hydroxysuccinimide	1.50	NMM	3	DMF
57	DCC	1.50	N-Hydroxyphthalimide	1.50	NMM	3	DMF
58	CITU	1.50			NMM	3	DMF
59	EDC-HCl	1.50	HOBt	1.50	NMM	3	DMF
60	PyAOP	1.50			NMM	3	DMF
61	TFFH	1.50	DMAP	**0.1**	DIPEA	3	DMF
62	CDI	1.10			DBU	0.75	DMF
63	CDMT	1.20			NMM	3	DMF

64	CIP	1.50	HOAt	1.00	DIPEA	3	DMF
65	DIC	1.50	Oxyma-B	1.10		3	DMF
66	EDC-HCl	1.50	Pentafluoro phenol	1.50	TEA	3	DMF
67	EDC-HCl	1.50	N-Hydroxysuc cinimide	1.50	TEA	3	DMF
68	EDC-HCl	1.50	N-Hydroxypht halimide	1.50	TEA	3	DMF
69	DCC	1.50	N-Hydroxysuc cinimide	1.50	TEA	3	DMF
70	DCC	1.50	Pentafluoro phenol	1.50	TEA	3	DMF
71	IBCF	1.00			NMM	3	DMF
72	DCC	1.50	N-Hydroxypht halimide	1.50	TEA	3	DMF
73	EDC-HCl	1.50	HOBt	1.50	TEA	3	DMF
74	TBDTU	1.50			DIPEA	3	DMF
75	TCFH	1.50			NMI	3	DMF
76	DCC	3.00					DMF
77	DCC	1.50	Oxyma	1.50			DMF
78	DCC	1.50	HOAT	1.50			DMF
79	DCC	1.50	HOBt	1.50			DMF
80	DCC	1.50	6-Cl-HOBT	1.50			DMF
81	DCC	3.00	DMAP	0.10			DMF
82	DIC	1.50	Oxyma	1.50			DMF
83	DIC	1.50	HOAT	1.50			DMF
84	DIC	1.50	HOBt	1.50			DMF
85	DIC	1.50	6-Cl-HOBT	1.50			DMF
86	EDC-HCl	1.50					DMF
87	EDC-HCl	1.50	Oxyma	1.50			DMF
88	EDC-HCl	1.50	HOAT	1.50			DMF
89	EDC-HCl	1.50	HOBt	1.50	2,6-Lutidine	3	DMF

90	EDC-HCl	1.50	6-Cl-HOBT	1.50			DMF
91	HATU	1.50					DMF
92	TPTU	1.50	HOBt	1.50	DIPEA	3	DMF
93	HBPYU	1.50					DMF
94	IIDQ	1.50			DIPEA		DMF
95	TOTU	1.50					DMF

S2.2 Substrates and products list

The list of substrates and products employed in the High-Throughput Experimentation (HTE) is presented on GitHub: https://www.github.com/aichemeco/amide_coupling/tree/main. Ultimately, more than 47,000 data points were collected via HTE. Given that these data pertain to the AIChemEco company, it is not feasible for us to incorporate all the data in the Supplementary Information (SI). Nevertheless, we have obtained authorization to disclose the data for six conditions, totaling approximately 3,000 data, on GitHub.

S2.3 Experimental design

First, test the solubility of each reagent and substrate to determine the solid/liquid packaging method. Then arrange the conditions in a 96-well reaction plate via a certain logic, and try to arrange the same reagents together to facilitate the addition of substrates, reagents and solvents. The following is an example of the arrangement of bases in a 96-well plate (Figure S2).

Amide Coupling screening set_Base Map

	1	2	3	4	5	6	7	8	9	10	11	12
A	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	NMM	NMM	DIPEA	TEA		
B	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	NMM	NMM		TEA		2,6-Lutidine
C	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	NMM	NMM	TEA	DIPEA		
D	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	NMM	NMM	TEA	NMI		
E	2,6-Lutidine	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	NMM	NMM	TEA			DIPEA
F	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	DMAP	NMM	DIPEA	TEA			
G	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	NMM	NMM	DBU	TEA			DIPEA
H	DIPEA	DIPEA	DIPEA	DIPEA	DIPEA	NMM	NMM	NMM	NMM			

Figure S2. Arrangement of bases in a 96-well plate

High-throughput experimental preparation

Before starting a high-throughput experiment, the following consumables were needed to be prepared in advance (taking the consumables required to start a HTE as an example, see Table S2).

Table S2. Experimental consumables

Experimental consumables	
Consumables Name	Quantity
96-well metal plate	1
750 μL glass tube	96
LC-MS plate	1
96-well deep well plate	1
96-well filter plate	1

DESI Sample plate	1
200 μL-MCA96 tips	400
50 μL-MCA96 tips	400

High-throughput reaction setup

Stock solutions of Carboxylic acid, amine, condensation reagents, additives and base were prepared in DMF. Firstly, 25 μL of a Carboxylic acid solution was dispense into a 96-well metal plate using EVO-200. Additionally, 25 μL of additive solution was added to each well. Lastly, 100 μL of condensation reagent in DMF was added to each well. The metal block was covered, tightened, and stirring at 25 °C for 0.5 h to activate carboxylic acid compounds using a Magnetic stirrer (for metal plate). After 0.5 h heating, unscrew the metal block cover, and amine in DMF were pipetted via EVO-200 into a 96-well metal plate. Then, 50 μL of base in DMF was added to each well. Lastly, 50 μL of DMF was added to each well, resulting in a final volume of 250 μL in each well. The metal block was covered, tightened, and stirring at 25 °C for 2 h. using a Magnetic stirrer (for HTE plate).

Reproducibility and Error Control in HTE Data

To ensure the accuracy and reproducibility of our high-throughput experimentation (HTE) data, several quality control measures were implemented during the experimental design and execution phases.

1. Control Condition Monitoring: In each 96-well HTE plate, condition C1 was designed to be identical to condition C20. These conditions should theoretically yield the same results. If the yield difference between C1 and C20 exceeded 15%, the data from the entire plate was discarded. This measure was employed to detect and eliminate potential errors arising from plate-specific variables.
2. Repeat Experiments: To further validate the experimental data, one HTE plate was randomly selected from each batch of 20 plates for repeat experimentation. The disparity in yield between the original and repeat experiments was required to be within 15% for at least 90% of the repeated conditions. If this criterion was not met, the entire batch was reinitialized. This procedure ensured that the data retained a high level of accuracy and consistency across different experimental runs.

These error control strategies were critical in maintaining the integrity of our dataset and ensuring that the prediction errors observed in our models could be accurately judged against the inherent experimental variability.

High-throughput reaction processing

4-Phenyltoluene (4.17 mmol) and Benzamide (4.17 mmol) were dissolved in 250 mL of acetonitrile to prepare the internal standard solution, see Table S3. After that, 150 μL of MeCN/H_2O (1:1) mixed solvent was dispensed separately into each plate followed by 150 μL of internal standard solution using the 96-tip head. Transfer the reaction plate to an ultrasonic cleaner and ultrasonicate for 5 minutes to ensure that the reaction solution and the internal standard solution are evenly mixed. 380 μL and 190 μL MeCN solvent were added to the transfer plate and filter plate respectively. 10

μL of solution was taken from the metal plate to the transfer plate, and then 10 μL solution was transferred from the transfer plate to the 96-well filter plate using EVO-200. After that, the 96-well filter plates were centrifuged for 2.5 min at 3000 RPM to prepare a plate for TOF-DESI injection. 180 uL MeCN solvent was moved to another filter plate using EVO-200. Finally, 20 μL solution from the transfer plate was taken to the 96-well filter plate using EVO-200. After that, the 96-well filter plates were centrifuged for 2.5 min at 3000 RPM to prepare another plate for LC-MS injection.

Table S3. Preparation of UPLC internal standard solution

IS	CAS	M.W. (g/mol)	Mol (mmol)	Mass (mg)	Solvent volume (mL)	UPLC sample concentration (μg/mL)	DESI sample concentration (μg/mL)
4-Phenyltoluene	644-08-6	168.23	4.17	**701.0**	250	1.96	0.98
Benzanilide	93-98-1	197.23	4.17	**821.8**	250	2.30	1.15

S2.5 TOF-DESI & UPLC analysis

The sample was spotted onto a 384-well glass plate using an automatic spotting tool (pin-tool). After the solvent evaporated naturally, the sample was injected into the TOF-DESI for MS analysis. Then the LC-MS plate was placed into the sample organizer of Waters UPLC system for UPLC analysis.

S2.6 NMR analysis

The wells with the highest yield were selected according to the TOF-DESI analyze data, and the mixture was concentrated by centrifugation at 35°C for 2 h. 1,1,2,2-tetrachloroethane (0.5 mmol) was dissolved in 50 mL of $CDCl_3$ to prepare the NMR internal standard solution, see Table S4. After that, 250 μL of internal standard solution was added to each well. After ultrasonic treatment for 5 min, the solution contained the internal standard was added to a 3 mm NMR tube using an 8-channel pipette and tested by 600M ^1H NMR.

Table S4. Preparation of NMR internal standard solution

IS	CAS	M.W.	eq.	Mol	Chemical / mg	solvent / mL	Solvent
1,1,2,2-tetrachloroethane	79-34-5	167.85	0.5	0.5	**83.9**	50	CDCl3

S2.7 Quantitative method

To guarantee the accuracy of our quantitative method using UPLC, we initially establish a standard curve of the standard sample. Dissolve Benzamide (0.1 mmol) and 4-Phenyltoluene (0.1 mmol) in 10 mL of acetonitrile to prepare IS1-a and IS2-a solutions respectively. Take 100 μL of IS1-a and IS2-a solutions respectively and add them to 1900 μL of acetonitrile to prepare IS1-b and

IS2-b internal standard solutions. Similarly, 100 μL of IS1-b and IS2-b solutions were added to 1900 μL of acetonitrile to prepare IS1 and IS2 internal standard solutions(Table S5). Next, prepare internal standard solutions with different concentration ratios for testing. Taking sample IS1-1 as an example, 100 μL IS1 and 100 μL IS2 internal standard solutions were mixed evenly to prepare the test sample IS1-1. The UV response obtained by UPLC test was 0.663(ratio of IS1 absorption area to IS2 absorption area = 79810.4/35171.38= 2.269), see Table S6. After testing the internal standard solution samples with different concentrations, the concentration ratio of IS1 and IS2 is used as the horizontal axis and the UV response is used as the vertical axis, those five points IS1-1, IS1-0_8, IS1-0_6, IS1-0_4, IS1-0_2 are plotted in the coordinate system, and the standard curve is Y=0.6468X (Figure S3). The R-Squared of the IS standard curve is 0.999(>0.99), indicating that there is a strong positive correlation between the variables. Benzamide and 4-Phenyltoluene can be used as UPLC internal standards for accurate quantification. Finally, draw the standard curve of the yield of different reactions, based on the UV Response (ratio of the UV absorption area of the product to the absorption area of the internal standard) and the NMR yield, a standard curve Y = kX (Y: NMR yield, X: UV response) is obtained, and the yield of the remaining wells were obtained by the standard curve. Randomly select 2-3 wells for NMR testing to verify the consistency between the NMR yield and the UV fitting yield.

Table S5. Preparation of UPLC internal standard solutions IS1 and IS2

No.	IS	CAS	M.W.	Mol	Sample Mass / mg	solvent / mL	Solvent	Concentration of IS-a(μg/mL)	Concentration of IS-b(μg/mL)	Concentration of IS(μg/mL)
IS1	BENZANILIDE	93-98-1	197.23	0.10	**19.7**	10	MeCN	1972.30	98.62	4.93
IS2	4-Phenyltoluene	644-08-6	168.23	0.10	**16.8**	10	MeCN	1682.30	84.12	4.21

Table S6. Preparation of UPLC internal standard test samples

Sample	Ratio (IS1/IS2)	IS1/μL	IS2/μL	MeOH/μL	IS1-area	IS2-area	UV response
IS1-1	1	100	100	0	359	542	0.663
IS1-0_8	0.8	80	100	20	298	582	0.512
IS1-0_6	0.6	60	100	40	225	583	0.386
IS1-0_4	0.4	40	100	60	140	598	0.235
IS1-0_2	0.2	20	100	80	74	599	0.124

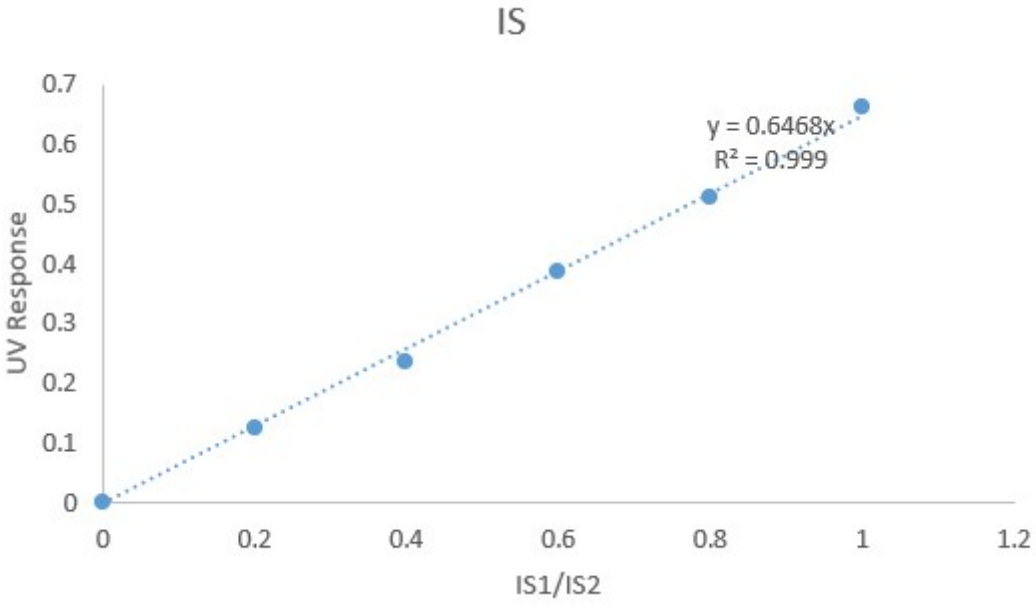


Figure S3. Standard curve of IS

An example for quantitative method

Herein, we will select a reaction as following (Figure S4) to present the quantitative process.

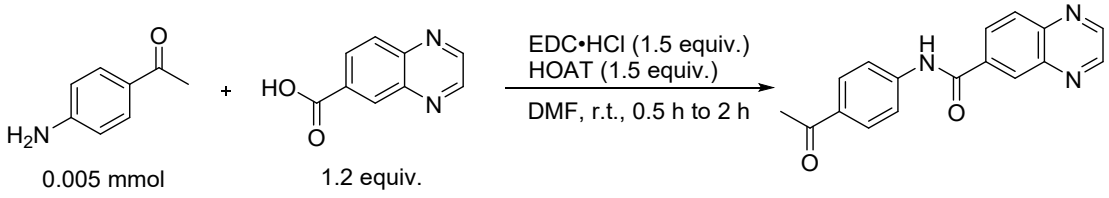


Figure S4. A selected reaction example

1) Samples injection from the plate for TOF-DESI were injected into the TOF-DESI for MS analysis, the automatic analysis program can realize the automatic analysis of the test samples. Subsequently, we could identify the success reaction quickly according the response of m/z (Figure S5).

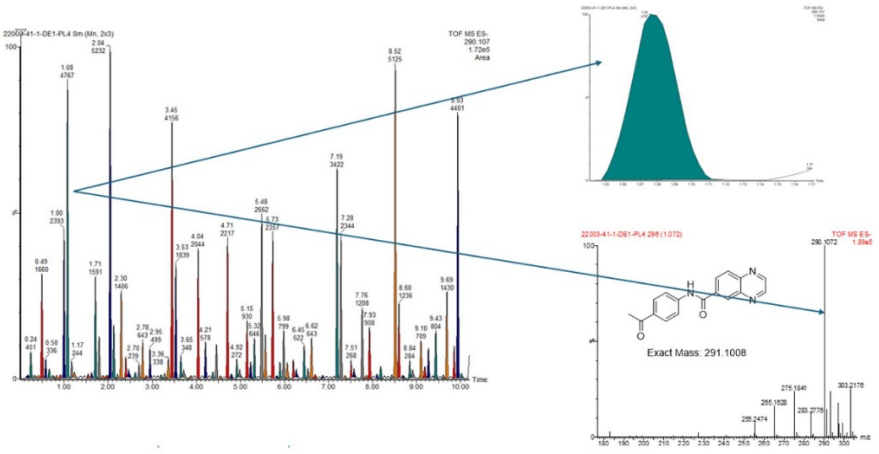


Figure S5. An example for DESI analysis

2) The samples with non-zero DESI response values were injected into UPLC, and randomly select 3 wells with zero DESI response values to verify whether UV Response is also zero. If they are inconsistent, inject the whole plate samples into UPLC. Taking a sample in W89 as an example (Figure S6), UV Response (Figure S7) is 2.269(ratio of product UV absorption area to internal standard absorption area = 79810.4/35171.38= 2.269)

	1	2	3	4	5	6	7	8	9	10	11	12
A	W1	W9	W17	W25	W33	W41	W49	W57	W65	W73	W81	W89
B	W2	W10	W18	W26	W34	W42	W50	W58	W66	W74	W82	W90
C	W3	W11	W19	W27	W35	W43	W51	W59	W67	W75	W83	W91
D	W4	W12	W20	W28	W36	W44	W52	W60	W68	W76	W84	W92
E	W5	W13	W21	W29	W37	W45	W53	W61	W69	W77	W85	W93
F	W6	W14	W22	W30	W38	W46	W54	W62	W70	W78	W86	W94
G	W7	W15	W23	W31	W39	W47	W55	W63	W71	W79	W87	W95
H	W8	W16	W24	W32	W40	W48	W56	W64	W72	W80	W88	W96

Figure S6. Well distribution of a HTE plate

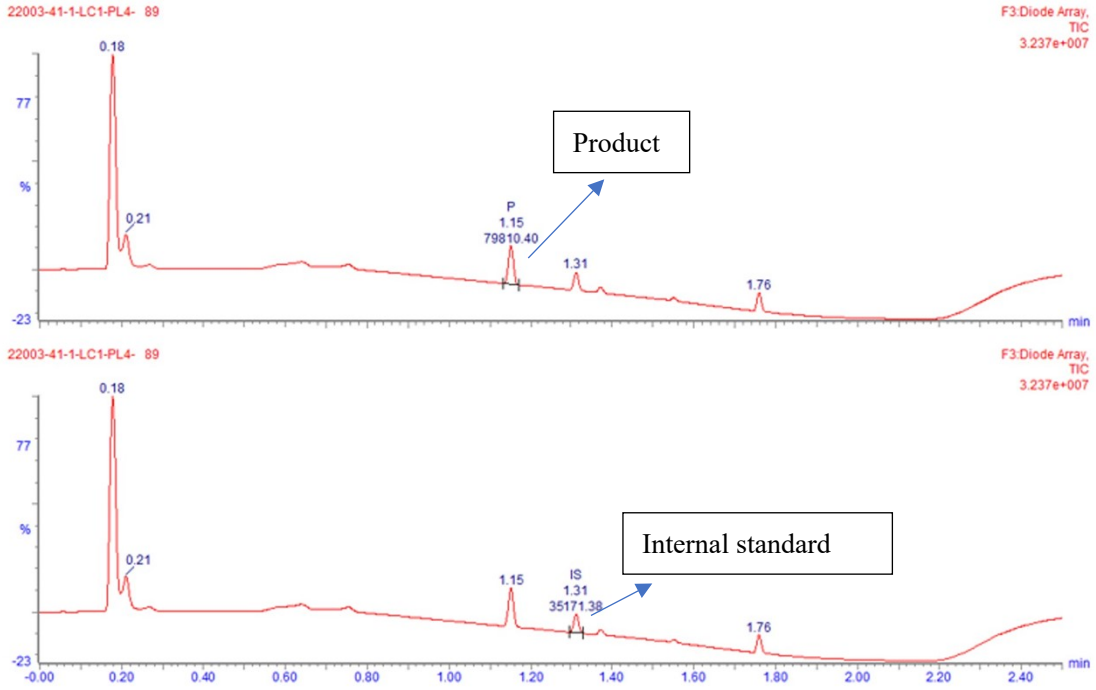


Figure S7. UPLC spectrum of W89

3) According to the UPLC analysis results, the wells with higher UV Response values were selected, centrifuged and concentrated at 35°C for 2 h, and 0.5 equiv. of NMR internal standard 1,1,2,2-tetrachloroethane was added. After ultrasonication for 5 min, the mixed solution contained the internal standard was added to a 3 mm NMR tube using an 8-channel pipette and tested by ^{1}H NMR (600 MHz). The characteristic peak of the internal standard 1,1,2,2-tetrachloroethane (δH = 5.96 ppm) was integrated as 1, and the yield of the product was equal to the number of integrated hydrogens / the actual number of hydrogens, the yield of PL4-W89 is 49% (Figure S8), and so on, the yields of the other wells are presented as following: PL4-W81= 42% (Figure S9), PL4-W83= 50% (Figure S10). Here, we also show the NMR spectrum of three other reactions:PL18-W83=76% (Figure S11), PL32-W10=81% (Figure S12), PL50-W64=80% (Figure S13).

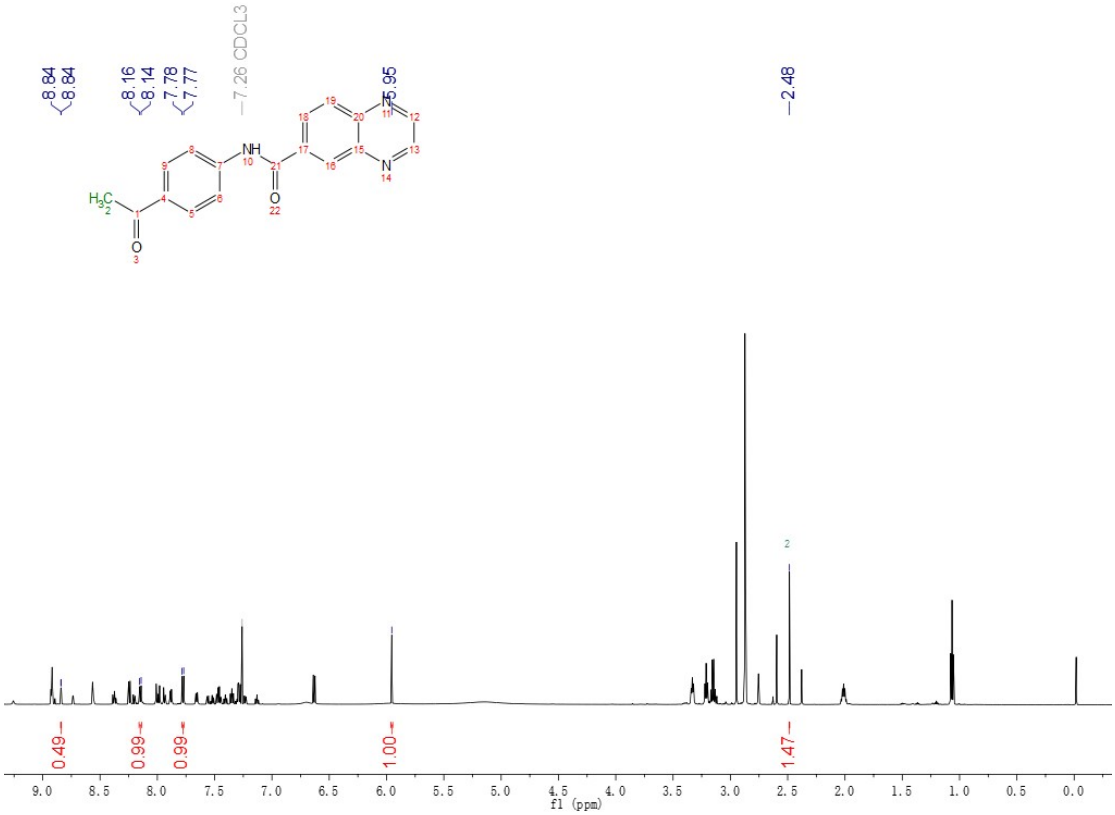


Figure S8. NMR spectrum of PL4-W89

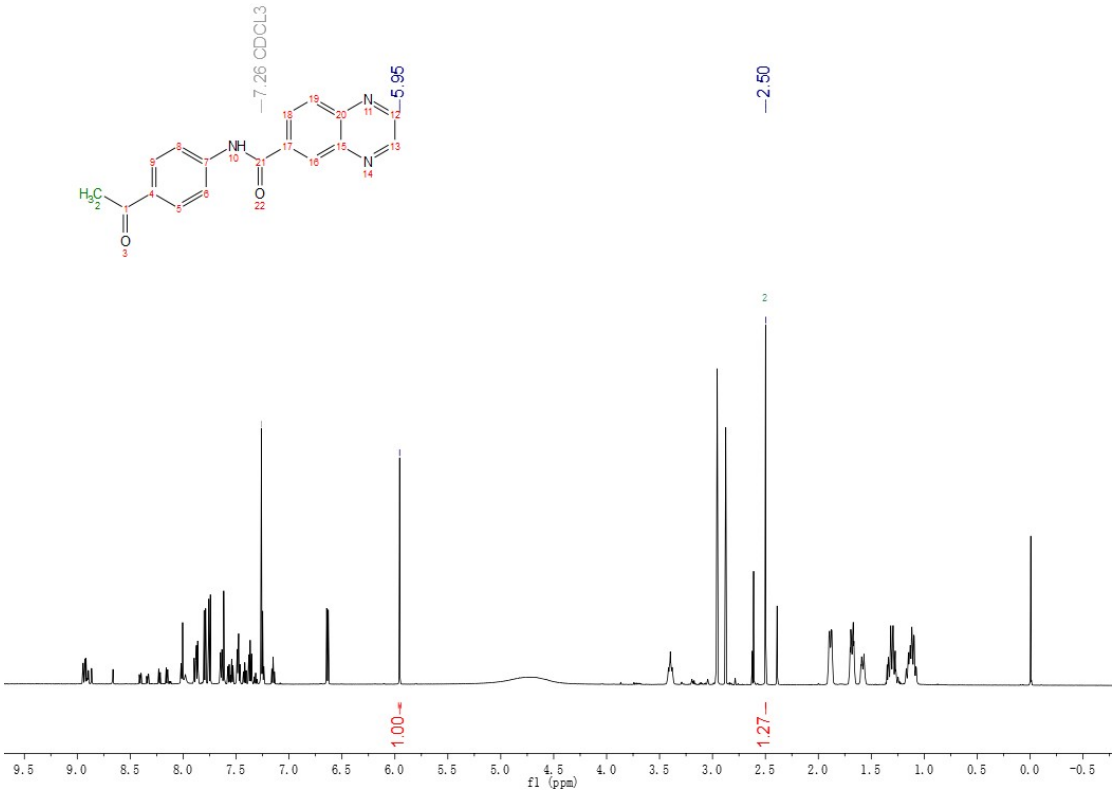


Figure S9. NMR spectrum of PL4-W81

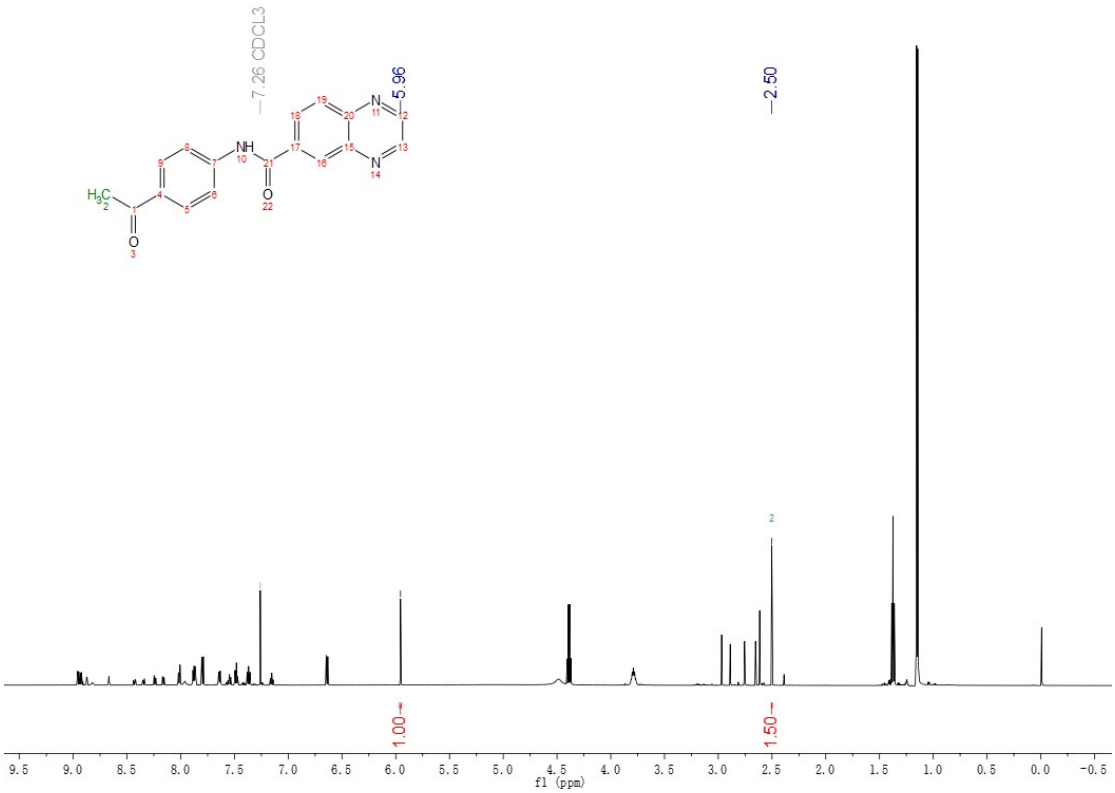


Figure S10. NMR spectrum of PL4-W83

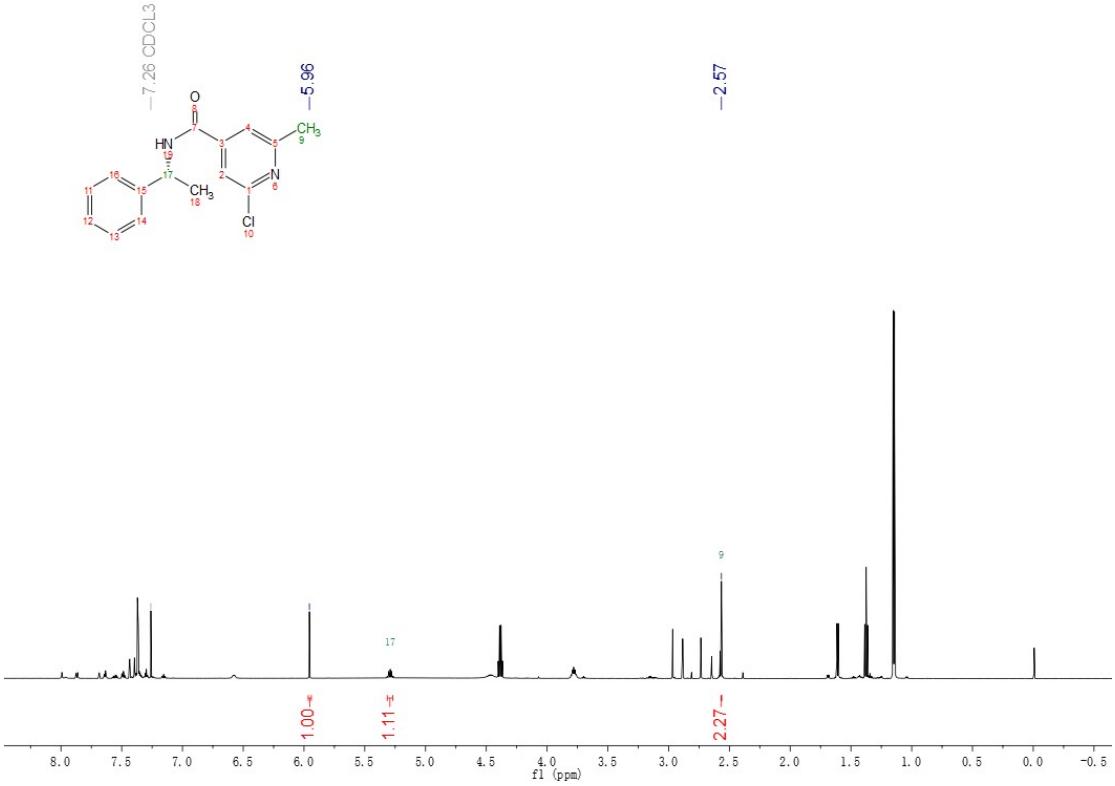


Figure S11. NMR spectrum of PL18-W83

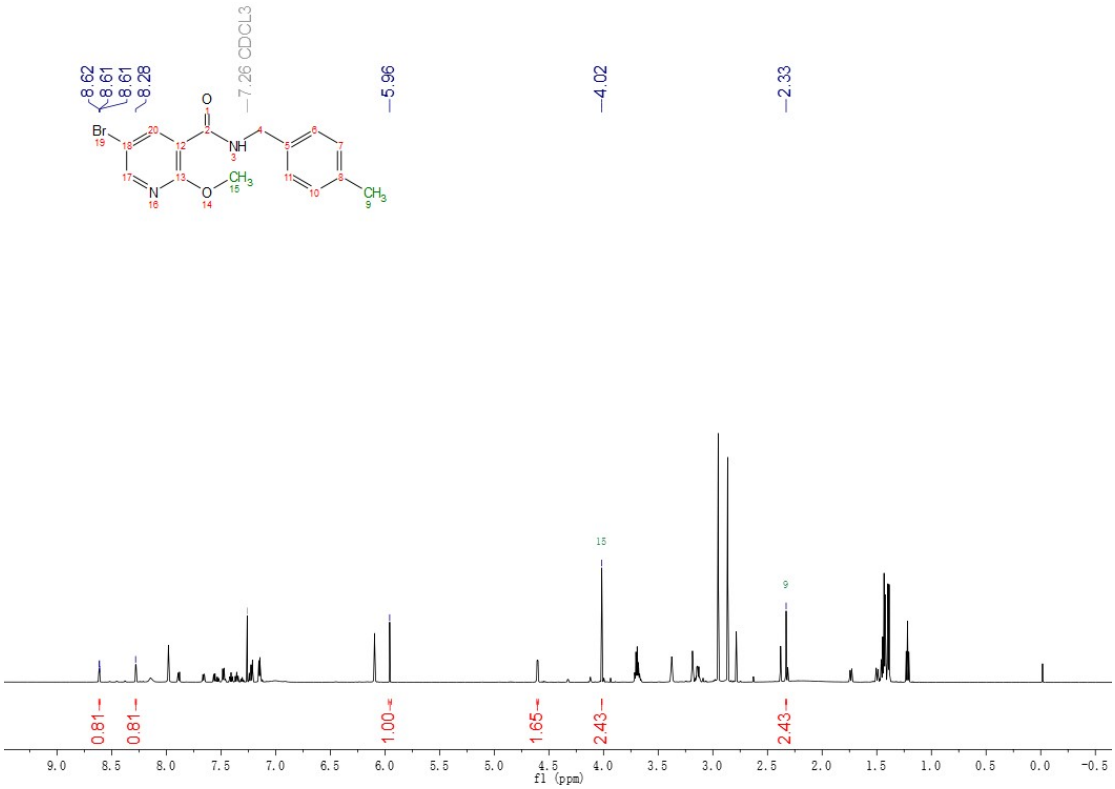


Figure S12. NMR spectrum of PL32-W10

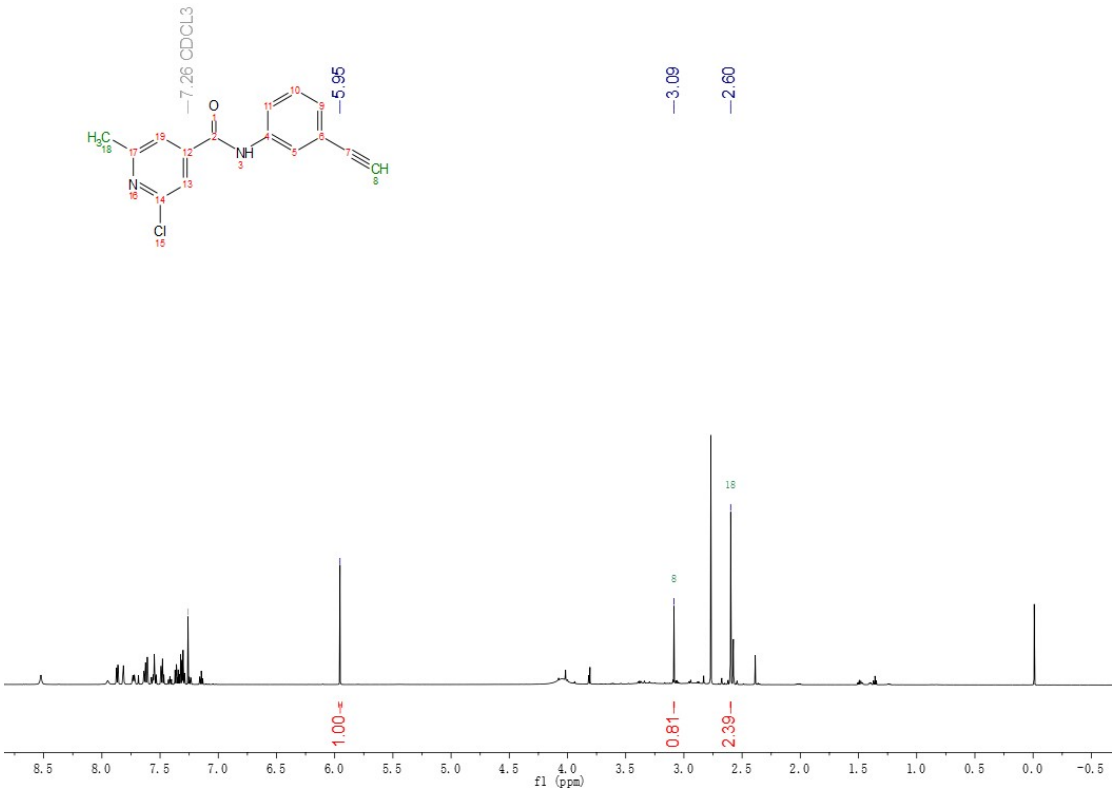


Figure S13. NMR spectrum of PL50-W64

4) Based on the UV Response (ratio of product UV absorption area to internal standard absorption area) and NMR yield, the standard curve Y = kX (Y is NMR yield, X is UV response) is obtained. For example, With UV response as the horizontal axis and NMR yield as the vertical axis, the three points PL4-W81, PL4-W83, and PL4-W89 are plotted in the coordinate system, and the standard curve of PL4 is Y=0.2176X (Figure S14).

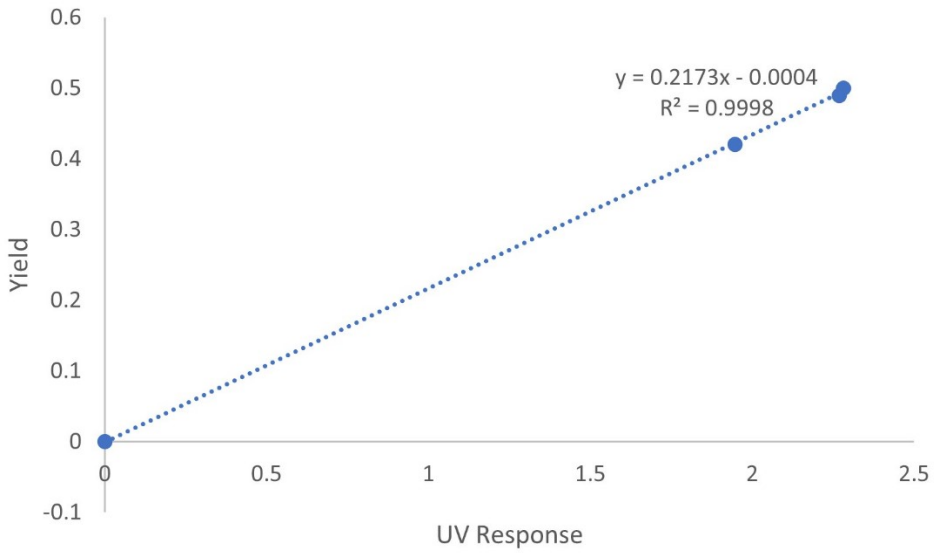


Figure S14. Standard curve of PL4

5) Finally, the yield values of the other wells are fitted by the standard curve, see Table S7 (W1 is a standard reaction which is performed to ensure the consistency between one plate with others, besides, W9 andeW75 are parallel reactions).

Table S7. Yield distribution

	1	2	3	4	5	6	7	8	9	10	11	12
A			0%	0%	41%	6%	3%	0%	0%	0%	42%	49%
B	0%	0%	0%	0%	46%	7%	0%	0%	67%	12%	0%	14%
C	1%	5%	16%	0%	9%	0%	0%	0%	0%	5%	50%	43%
D	0%	0%	66%	0%	11%	0%	0%	10%	0%	0%	50%	2%
E	18%	5%	21%	0%	66%	14%	0%	74%	0%	2%	8%	5%
F	15%	6%	16%	0%	0%	0%	0%	0%	0%	51%	41%	0%
G	30%	0%	29%	0%	3%	5%	0%	0%	0%	63%	15%	0%
H	19%	6%	22%	0%	30%	0%	0%	40%	0%	11%	60%	1%

S2.8 Yield distribution of HTE data and predicted results

After HTE, we collected approximately 47,000 reaction data and statistically analyzed the yield distribution (Figure S15)：

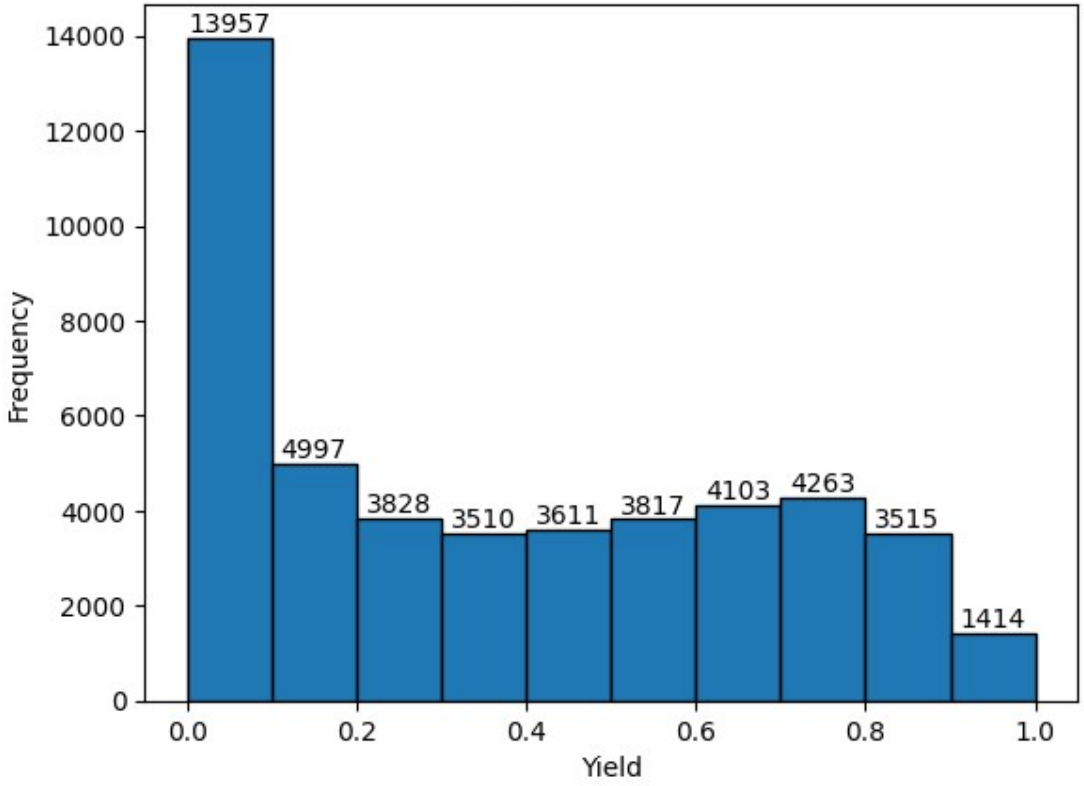


Figure S15. Yield distribution of HTE data

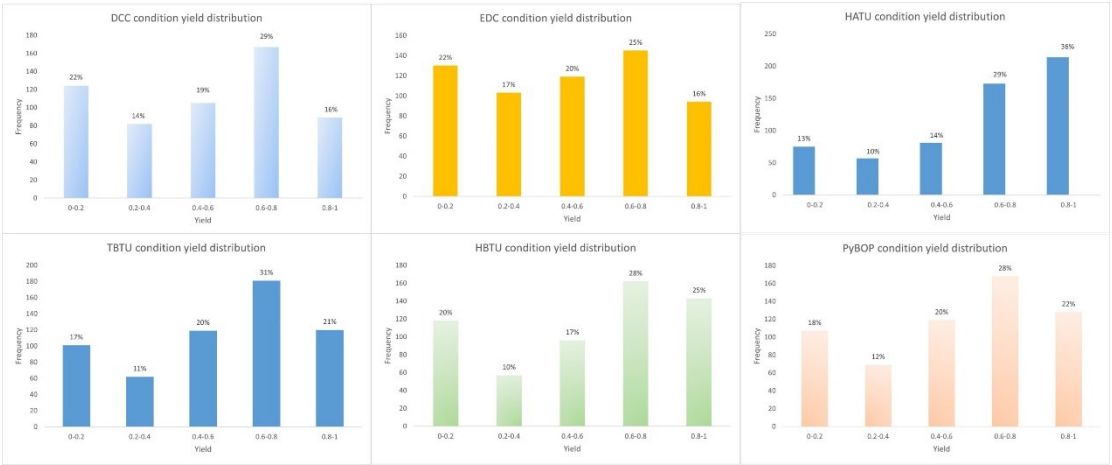


Figure S16. Yield distribution of HTE under six conditions

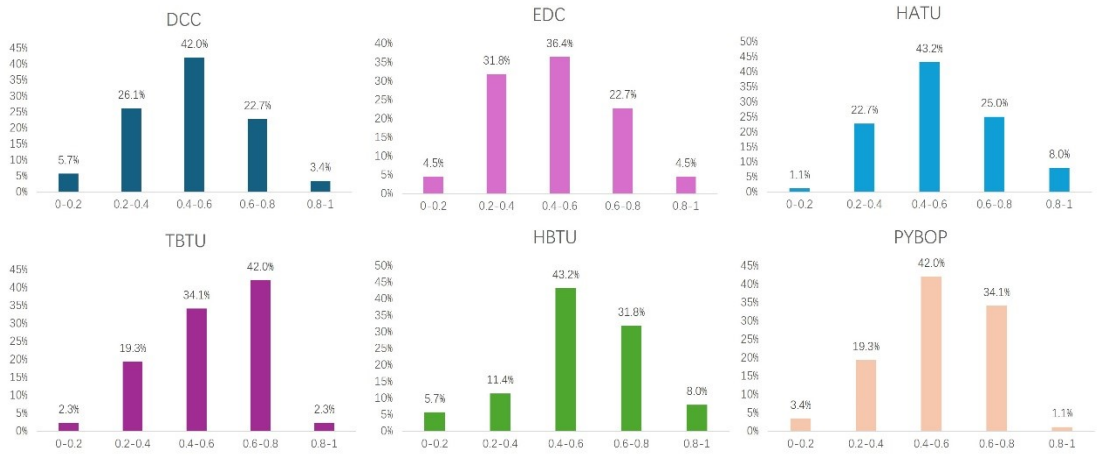


Figure S17. Yield distribution of predicted results under six conditions

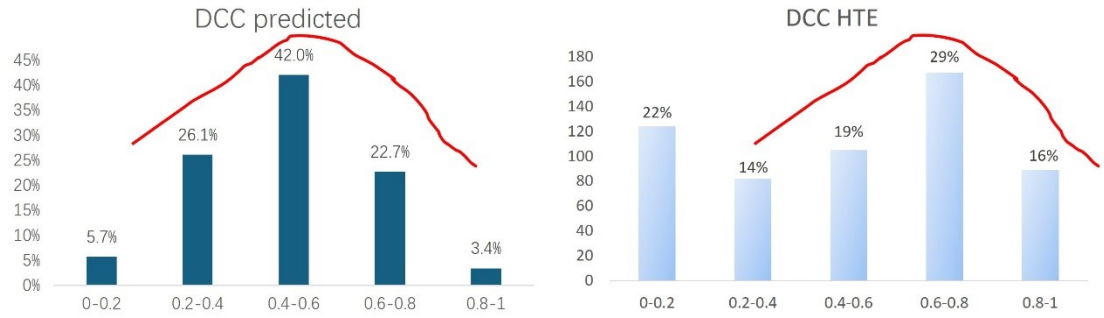


Figure S18. Comparison of yield distribution between predicted results and HTE

S3 Machine Learning Model Details

S3.1 Dimensionality reduction

Three dimensionality reduction techniques (PCA, t-SNE, UMAP) were used to reduce the 1,024 bits morgan fingerprints of the USPTO products and the virtual products into two dimensions. While PCA gives uniform visualization, as shown in Figure 2b, t-SNE and UMAP both illustrate globular structures after several parameter tunings, as shown in Figure S16 and Figure S17. The major parameters used in the t-SNE figure are perplexity of 10, numbers of iterations of 5000 and learning rate of 200. The major parameters used in the UMAP figure are numbers of neighbors of 15, minimum distance of 0.5 and metric of cosine.

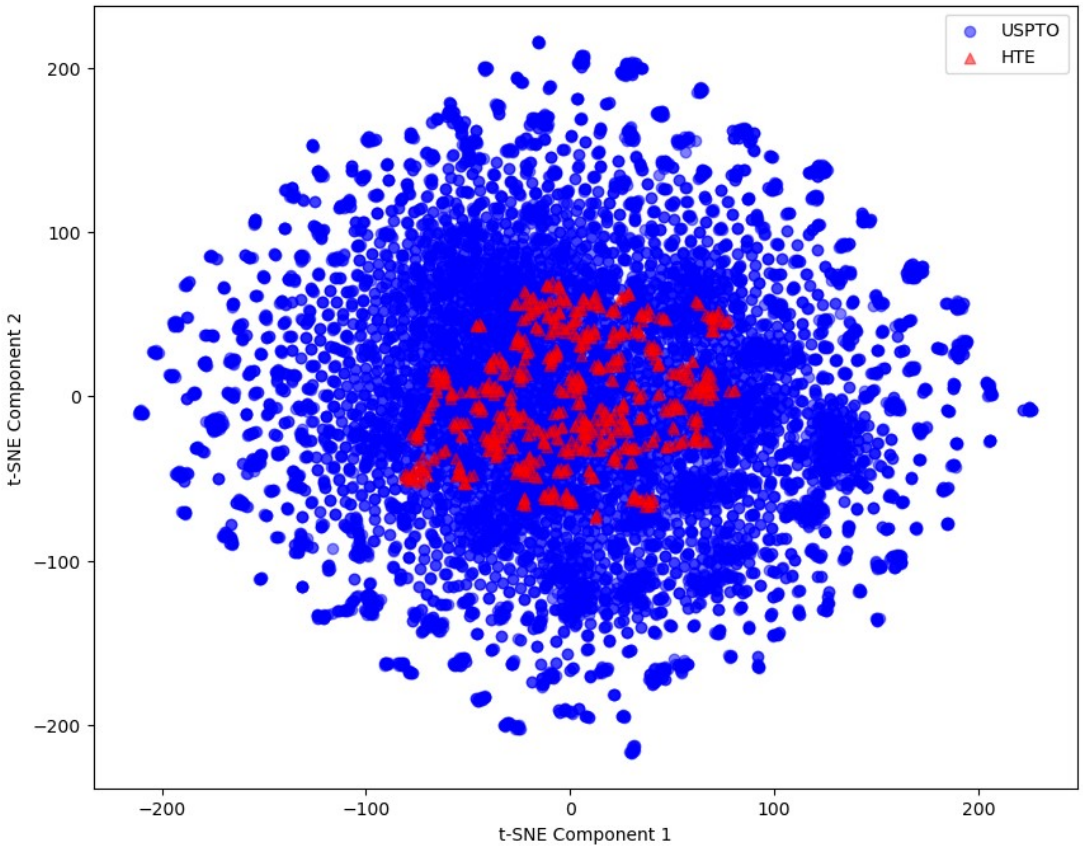


Figure S19. The chemical space of USPTO amide coupling products and products from commercially available substrates reduced from t-SNE.

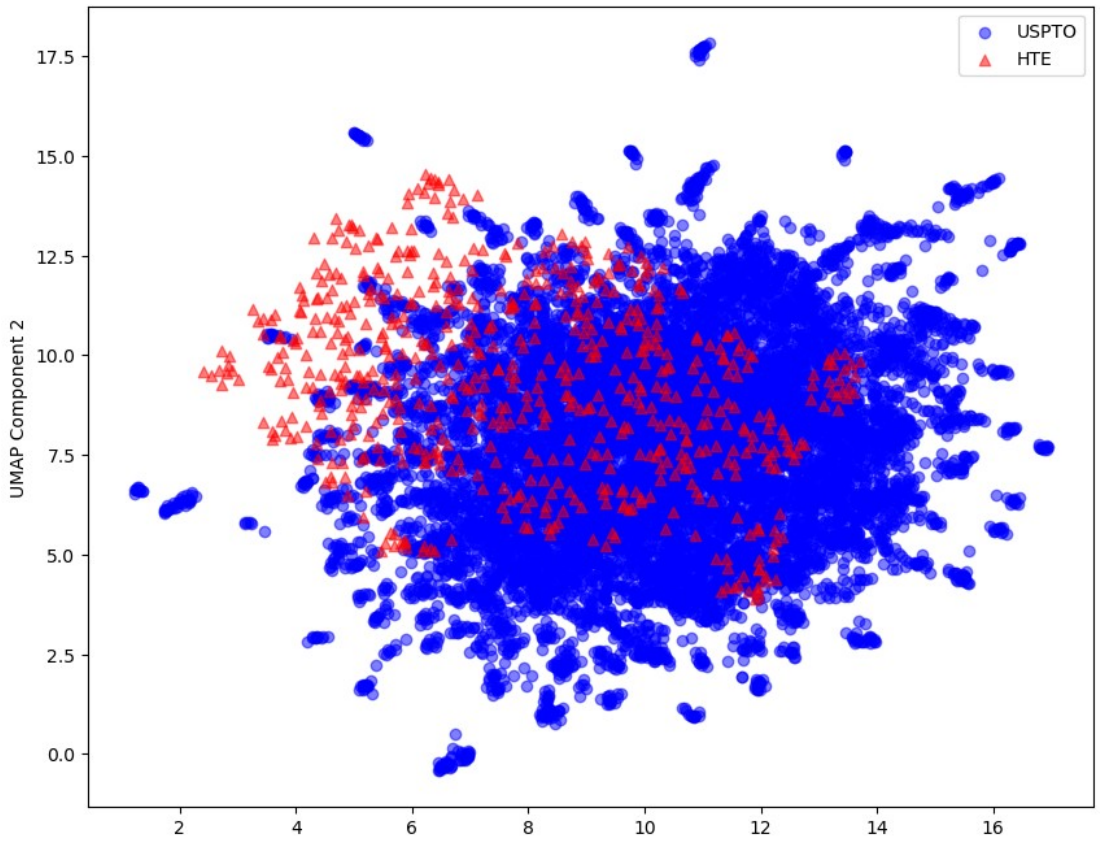


Figure S20. The chemical space of USPTO amide coupling products and products from commercially available substrates reduced from UMAP.

S3.2 Algorithms Used

BERT (Bidirectional Encoder Representations from Transformers): BERT is a pre-trained transformer model designed for natural language understanding. It uses bidirectional context to capture deeper language semantics. BERT is often used for tasks like text, classification, question answering, and named entity recognition.

XGBoost (Extreme Gradient Boosting): XGBoost is an optimized gradient boosting algorithm that is highly efficient and flexible. It uses gradient boosting framework for classification and regression tasks. XGBoost is widely used in structured data problems such as predictive modeling, ranking, and recommender systems.

SVM (Support Vector Machines): SVM is a supervised learning model used for classification and regression analysis. It finds the hyperplane that best separates classes in an n-dimensional space. SVM is effective in high-dimensional spaces and is popular in text classification, image classification, and bioinformatics.

Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes or mean prediction of the

individual trees. It is used for both classification and regression tasks, known for handling large datasets with high dimensionality.

T5 (Text-To-Text Transfer Transformer): T5 is a transformer model that converts all NLP tasks into a text-to-text format. It achieves state-of-the-art results by fine-tuning on downstream tasks. T5 can be used for text summarization, translation, question answering, and more by adapting to various input-output formats.

S3.3 Descriptor Sets

No Intermediate Descriptor Set: this set focuses on representing molecules based on their initial or final forms without considering any intermediate stages. This set typically includes descriptors that characterize the structural and chemical properties of the starting materials (e.g., amines or acids) or the final products after the reaction is complete. By omitting intermediates, this descriptor set simplifies the representation of molecules, focusing solely on endpoints that are directly relevant to the application or study.

Amine & Acid & Intermediate Descriptor Set: this descriptor set is particularly useful in studies where understanding the transformation process between reactants and products is crucial. It provides a more comprehensive representation of molecules across different stages of chemical reactions, aiding in the analysis of reaction mechanisms and predicting reaction outcomes.

Amine & Intermediate Descriptor Set: this set excludes descriptors related to acids or other reactants not directly involved in the initial stage of the reaction. It provides a more targeted approach to understanding molecular transformations during chemical reactions while simplifying the representation by focusing on key stages.

S3.4 Procedures of training and parameters

When SVM and XGBoost, RF, AutoGluon are used, molecular fingerprints of the compounds are utilized as inputs to the model. These fingerprints capture the structural features of molecules in a format that is suitable for machine learning algorithms, enabling effective classification or regression tasks based on their chemical properties. On the other hand, when employing T5 and BERT models, the input to the model consists of SMILES representations of the molecules. SMILES (Simplified Molecular Input Line Entry System) is a textual representation of a molecule's structure, where atoms and bonds are encoded into a string format. T5, a text-to-text transformer model, and BERT, a bidirectional transformer model, are adept at processing natural language and textual data. By utilizing SMILES as input, these models can effectively handle tasks such as molecular property prediction, virtual screening, and chemical reaction prediction by learning from the sequential and contextual information encoded in the SMILES strings. Here are some hyperparameters that can be tuned to optimize these models (Table S8):

Table S8. Hyperparameters and Candidate Values

SVM (Support Vector Machine)	
Hyperparameters	Candidate Values
Regularization parameter (C)	[0.1, 1, 10, 100]
Kernel type	[‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’]
Degree of the polynomial kernel function (degree)	[2, 3, 4]
Kernel coefficient (gamma)	[’scale’, ’auto’, 0.1, 0.01, 0.001]
Independent term in kernel function (coef0)	[0.0, 0.1, 0.5, 1.0]
Tolerance for stopping criteria (tol)	[1e-3, 1e-4, 1e-5]
Random Forest	
Hyperparameters	Candidate Values
Number of trees in the forest (n_estimators)	[100, 200, 500, 1000]
Maximum depth of the tree (max_depth)	[None, 10, 20, 30]
Minimum number of samples required to split an internal node (min_samples_split)	[2, 5, 10]
Minimum number of samples required to be at a leaf node (min_samples_leaf)	[1, 2, 4]
Number of features to consider when looking for the best split (max_features)	[’auto’, ’sqrt’, ’log2’, None, int]
XGBoost	
Hyperparameters	Candidate Values
Number of boosting rounds (n_estimators)	[100, 200, 500, 1000]
Step size shrinkage (learning_rate)	[0.01, 0.1, 0.2, 0.3]
Maximum depth of a tree (max_depth)	[3, 6, 9, 12]
Minimum sum of instance weight needed in a child (min_child_weight)	[1, 3, 5]
Minimum loss reduction required to make a further partition on a leaf node (gamma)	[0, 1, 5]
BERT (Bidirectional Encoder Representations from Transformers)	
Hyperparameters	Candidate Values
Size of the encoder and pooler layers (hidden_size)	[768, 1024, 2048]
Number of attention heads in each attention layer (num_attention_heads)	[12, 16, 24]
Size of the intermediate (feed-forward) layer in the encoder (intermediate_size)	[3072, 4096, 8192]
Number of hidden layers in the transformer encoder (num_hidden_layers)	[12, 24]
Maximum length of the input sequences	[512, 1024]

(max_position_embeddings)	
T5 (Text-To-Text Transfer Transformer)	
Hyperparameters	Candidate Values
Size of the encoder layers and the decoder layers (d_model)	[512, 768, 1024]
Dimensionality of the feed-forward layers (d_ff)	[2048, 3072, 4096]
Number of encoder layers (num_layers)	[6, 12, 24]
Number of decoder layers (num_decoder_layers)	[6, 12, 24]
Dropout rate for the attention and fully connected layers (dropout_rate)	[0.1, 0.2, 0.3]

S3.5 BERT model construction

Bidirectional Encoder Representations from Transformers (BERT) is a revolutionary model in the field of natural language processing (NLP). BERT is based on the Transformer architecture, which is known for its effectiveness in capturing dependencies in sequences, making it particularly suited for tasks involving text. One of the most promising models in this research work is the Yield-BERT model. While the hyperparameters of the model is given in Table S8, below is the step-by-step construction details of our BERT model, and the details scripts can be found in the GitHub repository.

1. SMILES Tokenization:

The input to the model is a reaction SMILES string, representing the chemical reaction. This string is tokenized into meaningful units for the model to process. This tokenization can be done using domain-specific rules to extract atomic symbols, bond types, and reaction components. For instance:

Example 1: The SMILES string C(=O)(C)C is tokenized into ['C', '(', '=', 'O', ')', '(', 'C', ')', 'C'].

Example 2: For a complex reaction SMILES CC(C)[C@@H](C)CCBr.[Na]C#N>>CC([C@@H](C)CCC#N)C, the tokenization might yield ['C', 'C', '(', 'C', ')', '[C@@H]', '(', 'C', ')', 'C', 'C', 'Br', '.', '[Na]', 'C', '#', 'N', '>>', 'C', 'C', '(', '[C@@H]', '(', 'C', ')', 'C', 'C', 'C', '#', 'N', ')', 'C'].

Custom tokenizers like SMILES-BPE (Byte Pair Encoding) or ChemBERTa tokenizer can be used to handle complex reaction formats efficiently by breaking them into substructures and functional groups.

2. Subword Tokenization:

To capture patterns in rarely encountered or out-of-vocabulary chemical symbols and structures, we implement a subword tokenization approach. This enables the model to decompose unknown or complex tokens into smaller, known subunits (e.g., C[C@H] -> [C], [C@@H], [=O]). This ensures that the model can generalize across unseen reactions or novel compounds by learning meaningful chemical fragments.

3. Special Tokens:

As in traditional BERT models, Yield-BERT adds special tokens to indicate the structure of the input sequence. Specifically:

[CLS]: A special token added at the beginning of the SMILES sequence. This token aggregates the

overall representation of the sequence and is used as the input for the yield prediction.

[SEP]: This token marks the end of a reaction or is used to separate reactants from products in the SMILES sequence. For example, C1=CC=CC=C1.O>>C1=CC(O)=CC=C1 becomes ['[CLS]', 'C1', '=', 'CC', '...', '[SEP]', 'O', '...', '[SEP]'].

4. Token Embeddings:

Each token from the SMILES string is converted into a fixed-dimensional vector, known as a token embedding. These embeddings are learned representations where each chemical token (e.g., atoms, bonds, or reaction fragments) is mapped to a dense vector that captures its chemical properties and context.

Pre-trained vs. Random Embeddings: Depending on the model configuration, the token embeddings can either be initialized randomly or using pre-trained embeddings from models such as **MolBERT** or **ChemBERTa**. Pre-trained embeddings offer a head start by encoding chemical knowledge learned from large datasets of molecular structures.

Embedding Size: The embedding vector's dimension (e.g., 256, 512) is a hyperparameter that can influence how much information each token can carry. Larger embeddings capture more detail about each token but increase computational cost.

Embedding Layers: Each token embedding layer is trained jointly with the rest of the model to refine the chemical understanding based on the target task, allowing the model to learn which aspects of a token (e.g., atom type, bond type, stereochemistry) are most important for yield prediction.

5. Positional Encoding:

Since the Transformer architecture does not inherently recognize the order of tokens in a sequence, **positional encoding** is crucial for SMILES strings, where the order of atoms and bonds is essential for accurately representing a molecule's structure.

Positional Encoding in Yield-BERT: A unique positional vector is added to each token embedding to provide the model with information about the relative position of tokens within the SMILES sequence. This is particularly important for preserving the sequential nature of chemical structures and reaction components, such as distinguishing between a methyl group bonded to different parts of a molecule.

Custom Positional Patterns: In the Yield-BERT model, the positional encodings could be designed to capture both local (atom-to-atom) and global (entire molecule) spatial relationships. This helps the model better understand which tokens are adjacent or distant within the chemical reaction, influencing yield predictions.

6. Segment Embeddings:

In tasks where the model needs to differentiate between multiple entities in the input (such as separating reactants from products or solvents), **segment embeddings** are used. Segment embeddings are added on top of the token and positional embeddings to inform the model which part of the input a token belongs to.

Reactant-Product Segmentation: For a reaction SMILES string that includes both reactants and products, the Yield-BERT model applies different segment IDs to tokens representing reactants and those representing products. For example, all reactants are assigned a segment embedding of 0, while all products receive a segment embedding of 1. This allows the model to distinguish between the two parts of the reaction.

Reaction Components: Segment embeddings can also be used to encode different components of

the reaction environment, such as catalysts or solvents. By adding a distinct segment embedding for these auxiliary elements, Yield-BERT can recognize their role and contribution to the overall reaction yield.

7. **Transformer Layers**:

The tokenized and embedded input passes through multiple layers of the Transformer architecture. Each layer consists of:

Multi-Head Self-Attention: This mechanism allows each token to attend to all other tokens in the sequence, enabling the model to capture long-range dependencies between molecular fragments, reactants, and products.

Feed-Forward Neural Networks: Following the self-attention mechanism, feed-forward layers provide additional nonlinear transformations, helping the model better learn complex chemical relationships. The number of attention heads and transformer layers can be tuned based on the task and dataset size (see Table S8).

8. **Linear Prediction Layer**:

After passing through the Transformer layers, the sequence representation corresponding to the [CLS] token (which encodes the overall reaction context) is fed into a fully connected linear layer. This layer is responsible for the regression task, predicting the reaction yield. The output from this linear layer is a continuous value between 0 and 100, representing the predicted yield percentage.

9. **Training and Output:**

The model is trained using a mean squared error (MSE) loss function to minimize the difference between predicted and actual reaction yields. The optimization is performed using AdamW or similar optimizers, with learning rate schedules (e.g., cosine decay) to improve convergence.

Data Augmentation: To increase model robustness, data augmentation techniques are applied. This includes random SMILES sampling, where the same reaction is represented by different equivalent SMILES strings during training.

10. **Evaluation**:

Yield-BERT is evaluated on standard regression metrics such as mean absolute error (MAE), R^2 score, and root mean square error (RMSE). Cross-validation is applied to ensure that the model generalizes well to unseen reaction data. Hyperparameters, including the number of Transformer layers, attention heads, batch size, and learning rate, are optimized through grid search (see Table S8).

In summary, a reaction SMILES string entering the Yield-BERT model undergoes tokenization, embedding, multi-layered self-attention, and yield regression, to predict the yield of the reaction substrates under different reaction conditions.

S3.6 Evaluation Metrics

Mean Squared Error (MSE): Mean Squared Error (MSE) is a widely used metric for regression tasks. It calculates the average squared difference between predicted values and actual values.

Mean Absolute Error (MAE): Mean Absolute Error (MAE) is another regression metric that calculates the average absolute difference between predicted values and actual values. It is less sensitive to outliers compared to MSE.

Root Mean Squared Error (RMSE): Root Mean Squared Error (RMSE) is the square root of MSE and is often used to provide an interpretable measure of the average magnitude of error.

R-squared (Coefficient of Determination): R-squared (R^2) is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variables. It is a relative measure of model fit and is often used to assess how well the model explains the variability of the data.

R-squared (Coefficient of Determination): R-squared (R^2) is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variables. It is a relative measure of model fit and is often used to assess how well the model explains the variability of the data.

S3.7 Molecular similarity between molecules from training and test datasets

The objective is to compute and compare the average molecular similarity between molecules (substrate 1 – amine, substrate 2 – acid and product) in the training dataset and those in the test dataset using the **Tanimoto coefficient**. The Tanimoto coefficient is a widely used similarity metric for comparing molecular fingerprints in cheminformatics.

The Tanimoto coefficient $T(A,B)$ between two molecules A and B, based on their binary fingerprints, is computed as:

$$T(A,B) = \frac{|A \cap B|}{|A| + |B| - A \cap B}$$

Where:

- $|A \cap B|$ is the number of bits from morgan fingerprints set to 1 in both fingerprints
- $|A|$ and $|B|$ are the number of bits from morgan fingerprints set to 1 in molecules A and B, respectively.

The Tanimoto coefficient ranges from 0 (no similarity) to 1 (identical molecules). Compute the Tanimoto coefficient between each pair of molecules, where one molecule is from the substrate 1 – amine or substrate 2 – acid, or product of **training dataset** and the other is from the that of **test dataset**, and the values are averaged to calculate **average similarity** in terms of Tanimoto coefficient. The results are shown in Table S9:

Table S9. Molecular similarity between molecules (substrate 1 – amine, substrate 2 – acid and product molecules) from training and test datasets (random split, partial substrate novelty and full substrate novelty test sets)

	Substrate 1 - amine	**Substrate 2 - acid**	**Product**
Random split	0.17	0.25	0.20
Partial substrate novelty	0.16	0.26	0.20
Full substrate novelty	0.10	0.24	0.17

The resulting average similarity scores indicate the structural distinctiveness between the training and test datasets in terms of Tanimoto coefficient.

S3.8 Model performance

The performance under 95 conditions: With HTE data set in hand, we first trained the model under 95 conditions and evaluate the performance of model. The metrics were illustrated in Table S10.

Table S10. The performance of model in test set under 95 conditions

Splitting	Metrics	XGBoost	SVM	RF	AutoGluon	Yield-BERT	T5-Chem
Random split	R^2	0.323	0.253	0.35	0.551	0.66	0.53
	MAE	0.18	0.186	0.173	0.152	0.15	0.22
	RMSE	0.222	0.234	0.218	0.204	0.10	0.16
Partial novelty	R^2	0.258	0.228	0.255	0.663	0.68	0.58
	MAE	0.038	0.162	0.143	0.134	0.14	0.20
	RMSE	0.197	0.212	0.191	0.180	0.10	0.15
Full novelty	R^2	0.247	0.22	0.258	0.420	0.63	0.58
	MAE	0.199	0.216	0.19	0.170	0.15	0.22
	RMSE	0.237	0.265	0.233	0.224	0.11	0.17

The performance under selected conditions: We selected six conditions from 95 conditions according their frequency in literature (Table S11) and extracted the corresponding data from HTE dataset. The size of training set, random split set, partial novelty set and full novelty set was illustrated in Table S12. It should be noted that random split set, partial novelty set and full novelty shared the same training set. The metrics of models under six conditions were shown in Table S13, including the models enhanced by intermediate knowledge. The results indicated the intermediate knowledge was indeed play a key role in elevating the performances of models.

Table S11. Statistics of 25 conditions

No.	Frequency	Activation reagents	Additives or catalysts	Base	Solvent
1	1430	HATU	□	DIPEA	DMF

2	932	EDC-HCl	HOBt	DIPEA	DCM
3	892	EDC-HCl	HOBt	TEA	DCM
4	635	EDC-HCl	HOBt	DIPEA	DMF
5	623	EDC-HCl	HOBt		DCM
6	604	EDC-HCl	HOBt		DMF
7	549	EDC-HCl	HOBt	TEA	DMF
8	523	EDC-HCl	DMAP		DCM
9	515	HBTU	□	DIPEA	DMF
10	514	DCC	HOBt	□	DMF
11	513	DCC	DMAP	□	DCM
12	498	HATU		DIPEA	DCM
13	464	DCC			DCM
14	340	PyBOP	□	DIPEA	DMF
15	320	EDC-HCl	HOBt	NMM	DMF
16	306	DCC	HOBt	TEA	DMF
17	275	DCC	HOBt		DCM

18	266	CDI	□	□	DMF
19	262	EDC-HCl			DCM
20	221	PyBOP		DIPEA	DCM
21	193	CDI			THF
22	190	DCC	HOBt		THF
23	187	IBCF		NMM	THF
24	186	DCC	HOBt	NMM	THF
25	183	TBTU	□	DIPEA	DMF

Table S12. The size of training set, random split set, partial novelty set and full novelty set

Data Set	TBTU	HATU	PyBOP	DCC	HBTU	EDC
Training set	477	443	502	455	430	439
Random split set	144	157	89	112	146	152
Partial novelty	60	82	31	66	71	67
Full novelty	11	11	8	6	13	13

Table S13. The performance under selected conditions

TBTU					
Splitting	Metrics	BERT	Embedded BERT	XGBT	Embedded XGBT
Random split	MAE	10%	7.0%[1] (7.0%[2])	8.0%	8.0%[1] (9.0%[2])
	RMSE	13%	10.0%[1] (10.0%[2])	12%	12%[1] (13%[2])
	R^2	0.71	0.83[1] (0.84[2])	0.76	0.76[1] (0.70[2])
	MAE	12%	9.0%[1] (8.0%[2])	10%	10%[1] (12%[2])

Partial novelty	RMSE	16%	12%[1] (12%[2])	13%	13%[1] (14%[2])
	R^2	0.57	0.76[1] (0.77[2])	0.72	0.71[1] (0.65[2])
Full novelty	MAE	11%	11%[1] (7.0%[2])	12%	12%[1] (14%[2])
	RMSE	13%	13%[1] (9.0%[2])	13%	12%[1] (12%[2])
	R^2	0.66	0.67[1] (0.85[2])	0.62	0.68[1] (0.57[2])
Splitting	Metrics	SVM	Embedded SVM	RF	Embedded RF
Random split	MAE	17%	17%[1] (17%[2])	12%	12%[1] (12%[2])
	RMSE	21%	21%[1] (21%[2])	14%	14%[1] (13%[2])
	R^2	0.34	0.34[1] (0.34[2])	0.73	0.76[1] (0.79[2])
Partial novelty	MAE	19%	19%[1] (19%[2])	15%	15%[1] (16%[2])
	RMSE	24%	24%[1] (24%[2])	21%	20%[1] (19%[2])
	R^2	0	0[1] (0.02[2])	0.62	0.65[1] (0.68[2])
Full novelty	MAE	10%	10%[1] (10%[2])	12%	12%[1] (12%[2])
	RMSE	12%	12%[1] (12%[2])	15%	15%[1] (14%[2])
	R^2	0.82	0.80[1] (0.80[2])	0.79	0.80[1] (0.81[2])
HATU					
Splitting	Metrics	BERT	Embedded BERT	XGBT	Embedded XGBT
Random split	MAE	10%	8.0%[1] (6.0%[2])	7.0%	9.0%[1] (7.0%[2])
	RMSE	14%	11%[1] (9.0%[2])	11%	12%[1] (11%[2])
	R^2	0.69	0.79[1] (0.86[2])	0.81	0.74[1] (0.78[2])
Partial novelty	MAE	13%	10%[1] (8.0%[2])	10%	10%[1] (10%[2])
	RMSE	17%	14%[1] (12%[2])	13%	13%[1] (13%[2])
	R^2	0.53	0.70[1] (0.78[2])	0.72	0.70[1] (0.70[2])
Full novelty	MAE	11%	9.0%[1] (6.0%[2])	8.0%	11%[1] (10%[2])
	RMSE	14%	13%[1] (7.0%[2])	12%	15%[1] (14%[2])
	R^2	0.39	0.51[1] (0.84[2])	0.58	0.47[1] (0.52[2])
Splitting	Metrics	SVM	Embedded SVM	RF	Embedded RF
Random split	MAE	18%	17%[1] (17%[2])	14%	14%[1] (14%[2])
	RMSE	21%	21%[1] (21%[2])	16%	16%[1] (14%[2])
	R^2	0.30	0.32[1] (0.32[2])	0.68	0.72[1] (0.74[2])
	MAE	19%	18%[1] (18%[2])	17%	16%[1] (16%[2])

Partial novelty	RMSE	23%	22%[1] (23%[2])	23%	21%[1] (18%[2])
	R^2	0	0.03[1] (0.05[2])	0.62	0.65[1] (0.68[2])
Full novelty	MAE	10%	10%[1] (10%[2])	15%	14%[1] (14%[2])
	RMSE	12%	13%[1] (13%[2])	18%	18%[1] (17%[2])
	R^2	0.78	0.76[1] (0.76[2])	0.67	0.68[1] (0.71[2])
PyBOP					
Splitting	Metrics	BERT	Embedded BERT	XGBT	Embedded XGBT
Random split	MAE	8.0%	6.0%[1] (5.0%[2])	6.0%	6.0%[1] (7.0%[2])
	RMSE	11%	8.0%[1] (8.0%[2])	10%	10%[1] (10%[2])
	R^2	0.80	0.88[1] (0.90[2])	0.88	0.88[1] (0.86[2])
Partial novelty	MAE	11%	7.0%[1] (7.0%[2])	8.0%	8.0%[1] (7.0%[2])
	RMSE	14%	11%[1] (10%[2])	10%	10%[1] (11%[2])
	R^2	0.63	0.77[1] (0.82[2])	0.81	0.80[1] (0.78[2])
Full novelty	MAE	12%	6.0%[1] (9.0%[2])	7.0%	7.0%[1] (9.0%[2])
	RMSE	18%	8.0%[1] (12%[2])	11%	11%[1] (12%[2])
	R^2	0.40	0.89[1] (0.74[2])	0.81	0.80[1] (0.76[2])
Splitting	Metrics	SVM	Embedded SVM	RF	Embedded RF
Random split	MAE	17%	17%[1] (17%[2])	16%	15%[1] (14%[2])
	RMSE	21%	21%[1] (21%[2])	16%	16%[1] (14%[2])
	R^2	0.34	0.34[1] (0.34[2])	0.69	0.74[1] (0.76[2])
Partial novelty	MAE	19%	19%[1] (19%[2])	18%	17%[1] (14%[2])
	RMSE	24%	24%[1] (241%[2])	24%	23%[1] (14%[2])
	R^2	0	0[1] (0[2])	0.54	0.74[1] (0.76[2])
Full novelty	MAE	10%	10%[1] (10%[2])	8.0%	8.0%[1] (7.0%[2])
	RMSE	12%	12%[1] (12%[2])	10%	10%[1] (10%[2])
	R^2	0.82	0.80[1] (0.80[2])	0.80	0.87[1] (0.85[2])
DCC					
Splitting	Metrics	BERT	Embedded BERT	XGBT	Embedded XGBT
Random split	MAE	8.0%	7.0%[1] (7.0%[2])	7.0%	7.0%[1] (7.0%[2])
	RMSE	11%	10%[1] (9.0%[2])	8.0%	8.0%[1] (8.0%[2])
	R^2	0.80	0.85[1] (0.86[2])	0.84	0.84[1] (0.84[2])

Partial novelty	MAE	9.0%	8.0%[1] (8.0%[2])	8.0%	8.0%[1] (8.0%[2])
	RMSE	13%	11%[1] (11%[2])	12%	12%[1] (11%[2])
	R^2	0.74	0.79[1] (0.81[2])	0.77	0.77[1] (0.79[2])
Full novelty	MAE	10%	5.0%[1] (7.0%[2])	6.0%	7.0%[1] (7.0%[2])
	RMSE	12%	7.0%[1] (10%[2])	10%	11%[1] (11%[2])
	R^2	0.1	0.67[1] (0.41[2])	0.61	0.56[1] (0.56[2])
Splitting	Metrics	SVM	Embedded SVM	RF	Embedded RF
Random split	MAE	19%	18%[1] (18%[2])	16%	15%[1] (14%[2])
	RMSE	23%	23%[1] (22%[2])	16%	16%[1] (14%[2])
	R^2	0.24	0.26[1] (0.27[2])	0.66	0.71[1] (0.73[2])
Partial novelty	MAE	17%	17%[1] (17%[2])	18%	17%[1] (16%[2])
	RMSE	20%	20%[1] (21%[2])	24%	23%[1] (21%[2])
	R^2	0.33	0.33[1] (0.31[2])	0.49	0.52[1] (0.53[2])
Full novelty	MAE	10%	11%[1] (10%[2])	18%	18%[1] (17%[2])
	RMSE	13%	13%[1] (13%[2])	17%	18%[1] (17%[2])
	R^2	0.77	0.75[1] (0.76[2])	0.58	0.57[1] (0.59[2])
HBTU					
Splitting	Metrics	BERT	Embedded BERT	XGBT	Embedded XGBT
Random split	MAE	8.0%	6.0%[1] (6.0%[2])	8.0%	8.0%[1] (8.0%[2])
	RMSE	11%	10%[1] (9.0%[2])	6.0%	6.0%[1] (6.0%[2])
	R^2	0.83	0.85[1] (0.89[2])	0.85	0.84[1] (0.84[2])
Partial novelty	MAE	10%	8.0%[1] (7.0%[2])	8.0%	8.0%[1] (8.0%[2])
	RMSE	13%	12%[1] (11%[2])	12%	11%[1] (12%[2])
	R^2	0.72	0.76[1] (0.81[2])	0.75	0.76[1] (0.75[2])
Full novelty	MAE	8.0%	9.0%[1] (7.0%[2])	8.0%	8.0%[1] (8.0%[2])
	RMSE	14%	17%[1] (10%[2])	17%	17%[1] (18%[2])
	R^2	0.68	0.54[1] (0.83[2])	0.67	0.66[1] (0.63[2])
Splitting	Metrics	SVM	Embedded SVM	RF	Embedded RF
Random split	MAE	18%	17%[1] (17%[2])	19%	19%[1] (18%[2])
	RMSE	21%	21%[1] (21%[2])	18%	18%[1] (17%[2])
	R^2	0.3	0.32[1] (0.32[2])	0.51	0.53[1] (0.56[2])

Partial novelty	MAE	19%	18%[1] (18%[2])	19%	18%[1] (18%[2])
	RMSE	23%	22%[1] (23%[2])	22%	22%[1] (22%[2])
	R^2	0	0.03[1] (0.04[2])	0.39	0.40[1] (0.43[2])
Full novelty	MAE	10%	10%[1] (10%[2])	14%	13%[1] (13%[2])
	RMSE	12%	13%[1] (13%[2])	16%	16%[1] (16%[2])
	R^2	0.78	0.76[1] (0.76[2])	0.45	0.52[1] (0.50[2])
EDC					
Splitting	Metrics	BERT	Embedded BERT	XGBT	Embedded XGBT
Random split	MAE	7.0%	5.9%[1] (6.1%[2])	9.0%	9.0%[1] (9.0%[2])
	RMSE	11%	8.0%[1] (8.0%[2])	10%	11%[1] (11%[2])
	R^2	0.82	0.89[1] (0.89[2])	0.78	0.76[1] (0.76[2])
Partial novelty	MAE	8.0%	6.0%[1] (7.0%[2])	10%	10%[1] (10%[2])
	RMSE	12%	9.0%[1] (9.0%[2])	12%	12%[1] (12%[2])
	R^2	0.79	0.88[1] (0.87[2])	0.66	0.65[1] (0.64[2])
Full novelty	MAE	13%	8.0%[1] (11%[2])	12%	13%[1] (13%[2])
	RMSE	18%	12%[1] (14%[2])	15%	16%[1] (15%[2])
	R^2	0.46	0.75[1] (0.67[2])	0.63	0.59[1] (0.62[2])
Splitting	Metrics	SVM	Embedded SVM	RF	Embedded RF
Random split	MAE	19%	19%[1] (19%[2])	17%	16%[1] (16%[2])
	RMSE	22%	22%[1] (22%[2])	18%	17%[1] (17%[2])
	R^2	0.27	0.28[1] (0.28[2])	0.58	0.62[1] (0.65[2])
Partial novelty	MAE	21%	20%[1] (20%[2])	19%	19%[1] (17%[2])
	RMSE	24%	24%[1] (24%[2])	24%	24%[1] (22%[2])
	R^2	0.19	0.22[1] (0.22[2])	0.33	0.37[1] (0.38[2])
Full novelty	MAE	11%	11%[1] (11%[2])	13%	14%[1] (13%[2])
	RMSE	13%	14%[1] (14%[2])	15%	15%[1] (15%[2])
	R^2	0.75	0.74[1] (0.74[2])	0.72	0.72[1] (0.73[2])
[1]Intermediate knowledge was embedded into the model; [2]Intermediate knowledge was embedded into the model but the representation (SMILES or ECFP) on acid was removed.					

Performance comparison of acylation yield prediction models trained using the BERT algorithm under different scenarios (Figure S18 - Figure S23): (a) Model trained without intermediate

information, (b) Model trained with both amine, acid, and intermediate information for predicting amide formation, and (c) Model trained with amine and intermediate information only for predicting amide formation. Each scenario is evaluated under three different data splitting strategies: Random split, Partial substrate novelty, and Full substrate novelty.

For each plot, predicted yield is plotted against experimental yield, with key performance metrics (MSE, MAE, RMSE, and R^2) listed on each subplot. The red lines represent ideal predictions (diagonal line) and ±10% deviations. In general, the models show robust performance under Random split, while performance degrades under the more challenging Partial and Full substrate novelty scenarios. Notably, the model in amine with intermediate exhibits the best overall performance across all scenarios, particularly in Full substrate novelty, indicating the potential importance of intermediate information for accurate yield prediction.

The model performance figures using other algorithm can be found in GitHub: https://www.github.com/aichemeco/amide_coupling/tree/main.

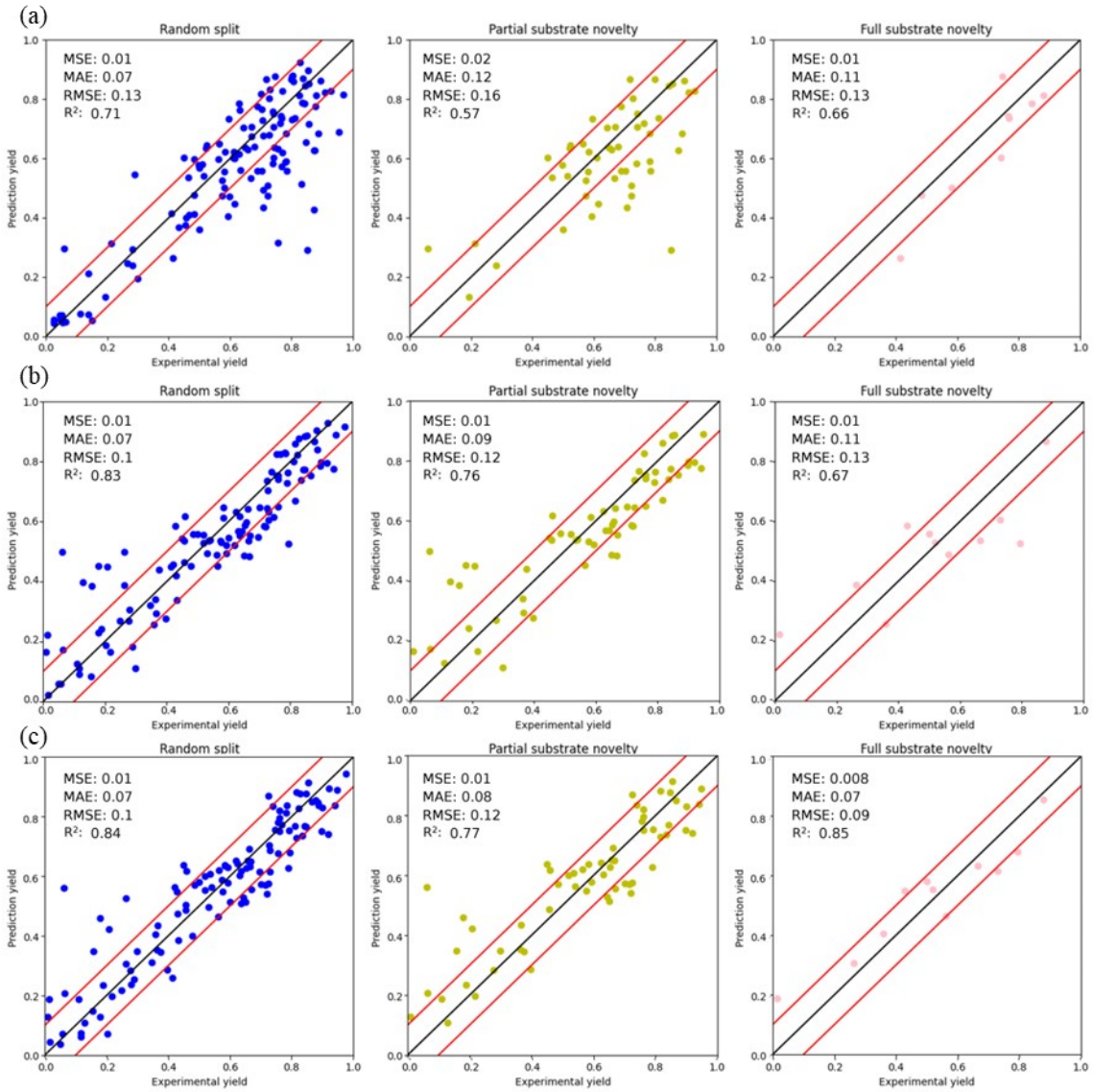


Figure S21. Model performance of TBTU condition using the BERT algorithm, trained under three different scenarios: (a) without intermediate, (b) amine + acid + intermediate → amide (c) amine + intermediate → amide. Each scenario is evaluated using three data splitting strategies: Random split, Partial substrate novelty, and Full substrate novelty. The red lines indicate the ideal prediction line (center) and ±10% error margins. Performance metrics (MSE, MAE, RMSE, R^2) are provided for each case.

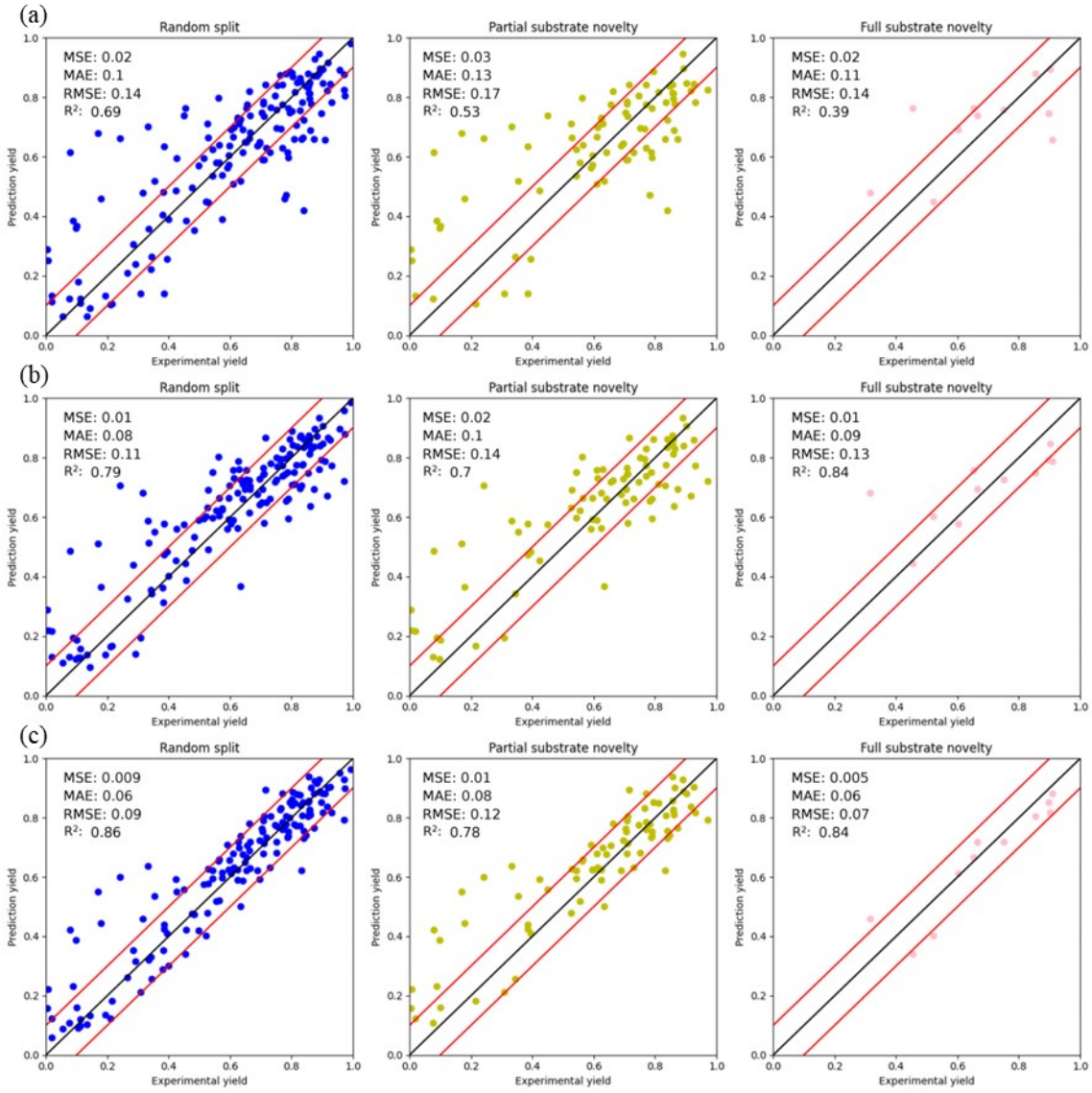


Figure S22. Model performance of HATU condition using the BERT algorithm, trained under three different scenarios: (a) without intermediate, (b) amine + acid + intermediate → amide (c) amine + intermediate → amide. Each scenario is evaluated using three data splitting strategies: Random split, Partial substrate novelty, and Full substrate novelty. The red lines indicate the ideal prediction line (center) and ±10% error margins. Performance metrics (MSE, MAE, RMSE, R^2) are provided for each case.

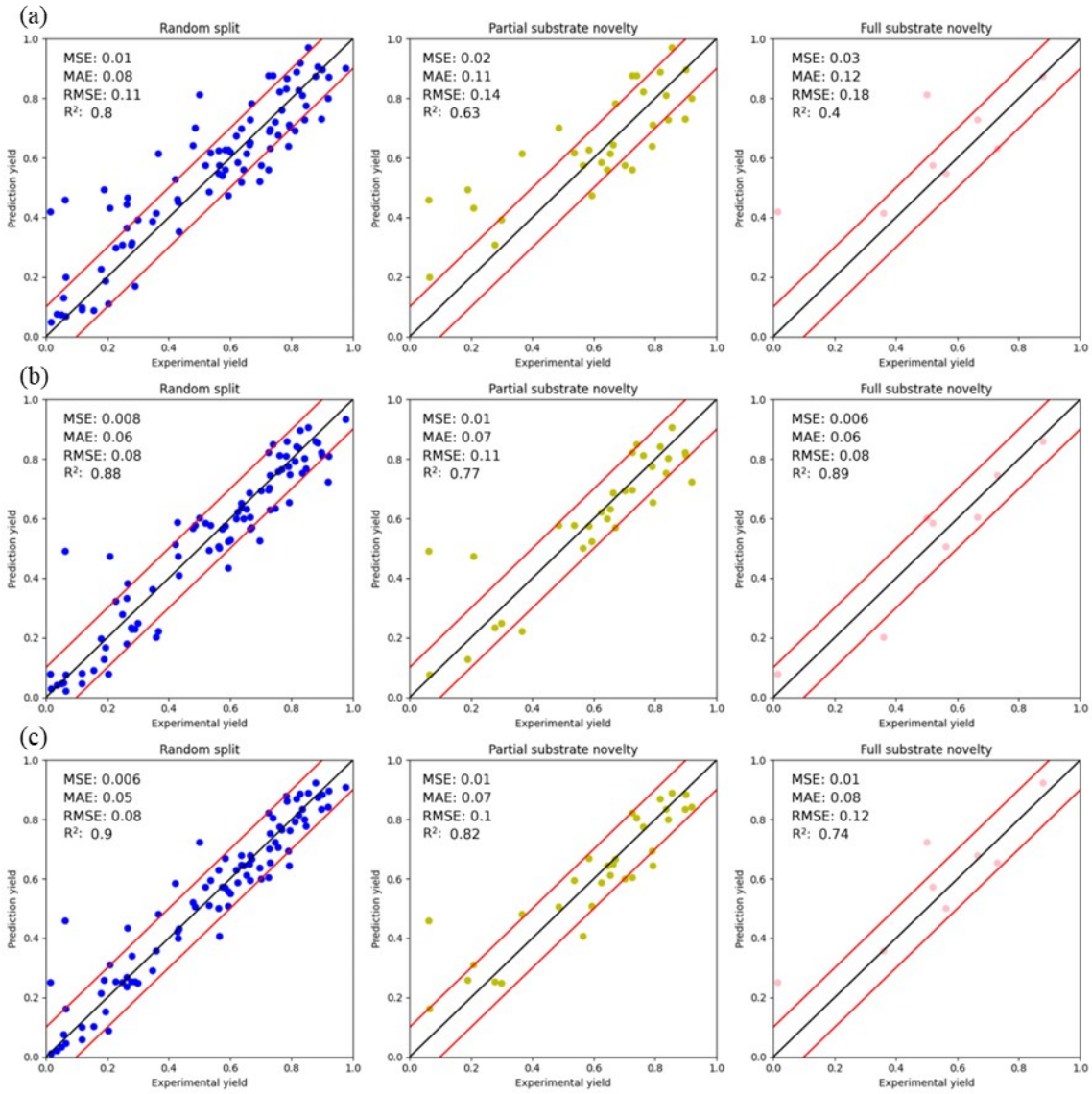


Figure S23. Model performance of PyBOP condition using the BERT algorithm, trained under three different scenarios: (a) without intermediate, (b) amine + acid + intermediate → amide (c) amine + intermediate → amide. Each scenario is evaluated using three data splitting strategies: Random split, Partial substrate novelty, and Full substrate novelty. The red lines indicate the ideal prediction line (center) and ±10% error margins. Performance metrics (MSE, MAE, RMSE, R^2) are provided for each case.

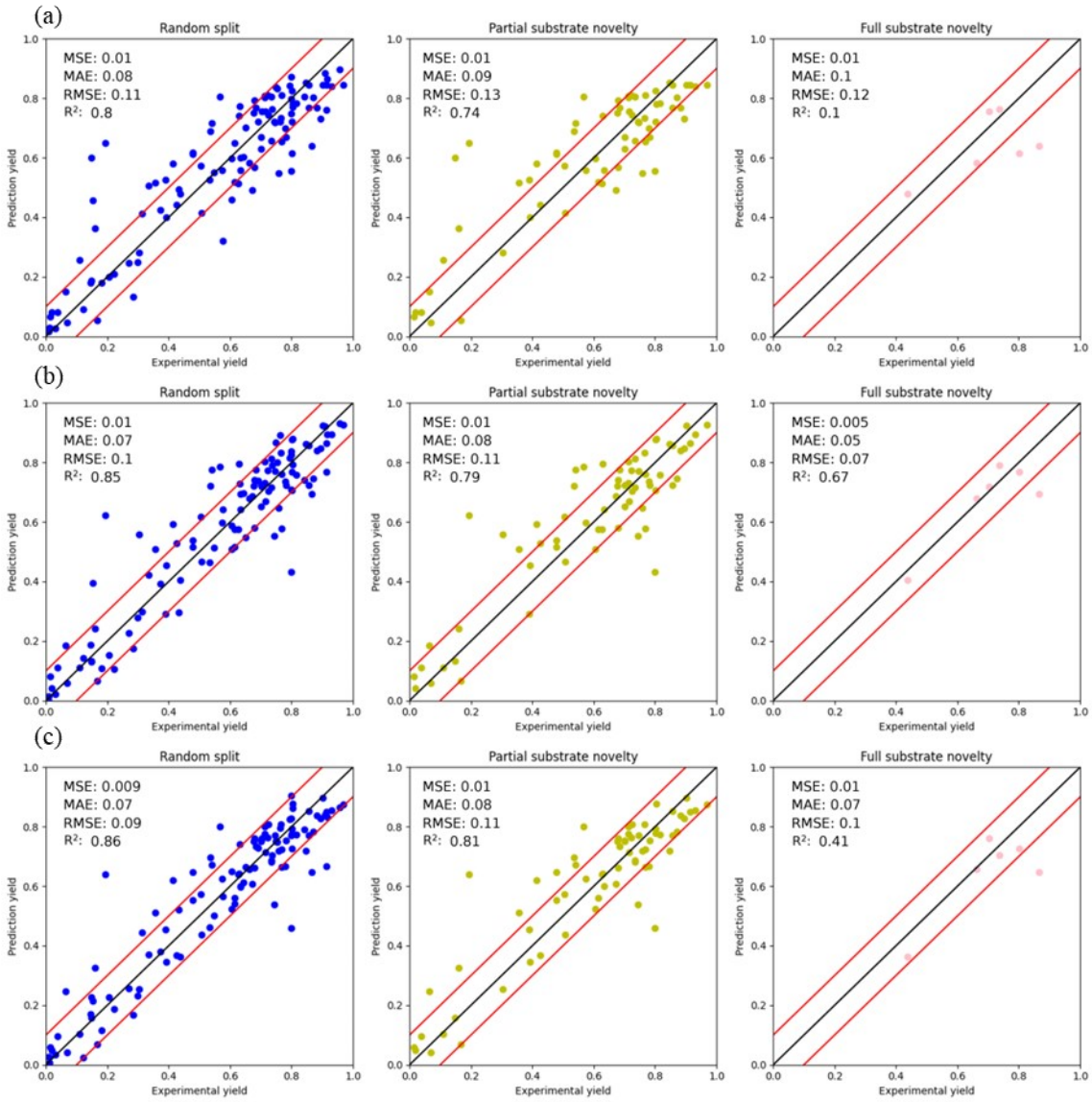


Figure S24. Model performance of DCC condition using the BERT algorithm, trained under three different scenarios: (a) without intermediate, (b) amine + acid + intermediate → amide (c) amine + intermediate → amide. Each scenario is evaluated using three data splitting strategies: Random split, Partial substrate novelty, and Full substrate novelty. The red lines indicate the ideal prediction line (center) and ±10% error margins. Performance metrics (MSE, MAE, RMSE, R^2) are provided for each case.

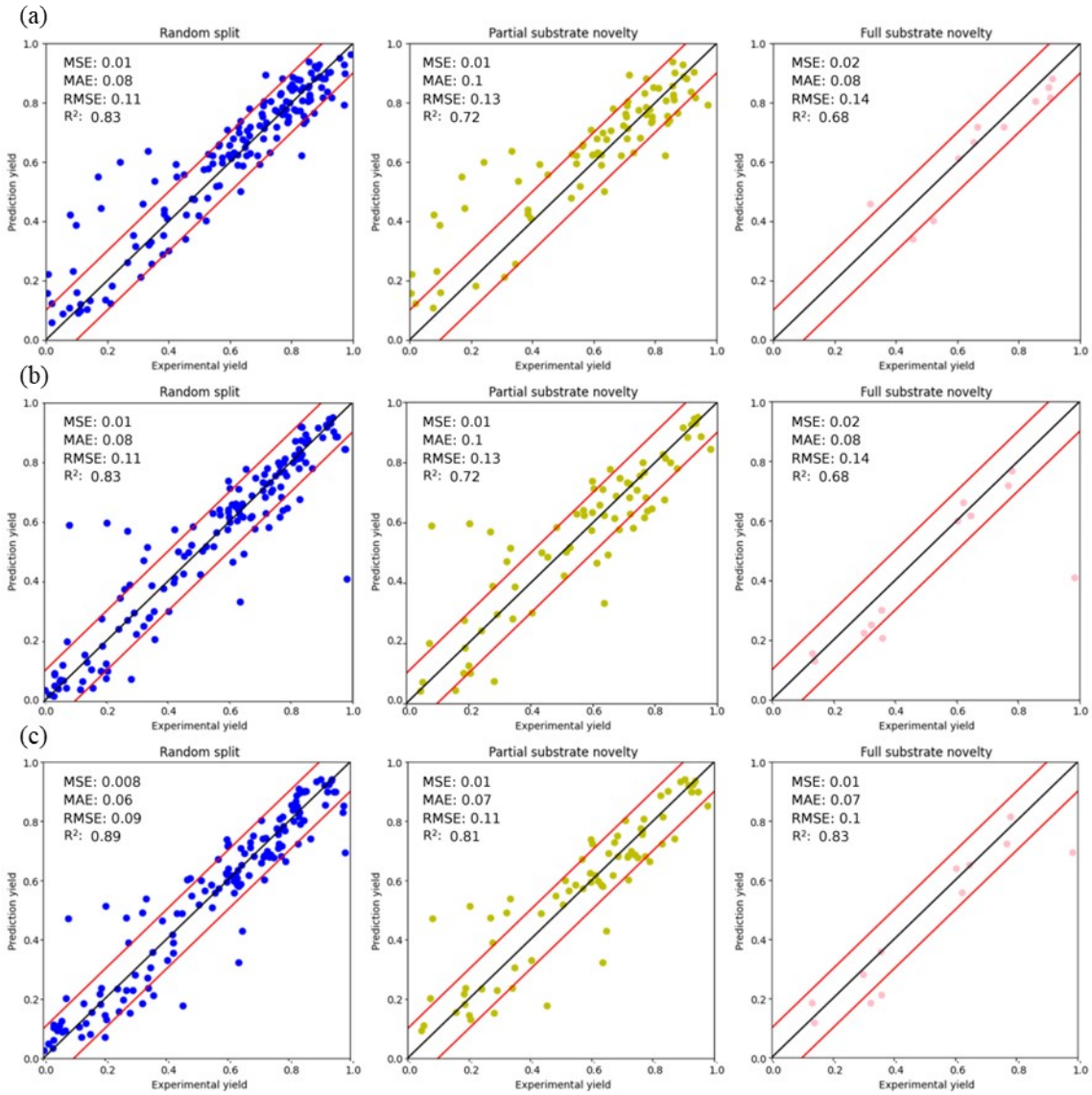


Figure S25. Model performance of HBTU condition using the BERT algorithm, trained under three different scenarios: (a) without intermediate, (b) amine + acid + intermediate → amide (c) amine + intermediate → amide. Each scenario is evaluated using three data splitting strategies: Random split, Partial substrate novelty, and Full substrate novelty. The red lines indicate the ideal prediction line (center) and ±10% error margins. Performance metrics (MSE, MAE, RMSE, R^2) are provided for each case.

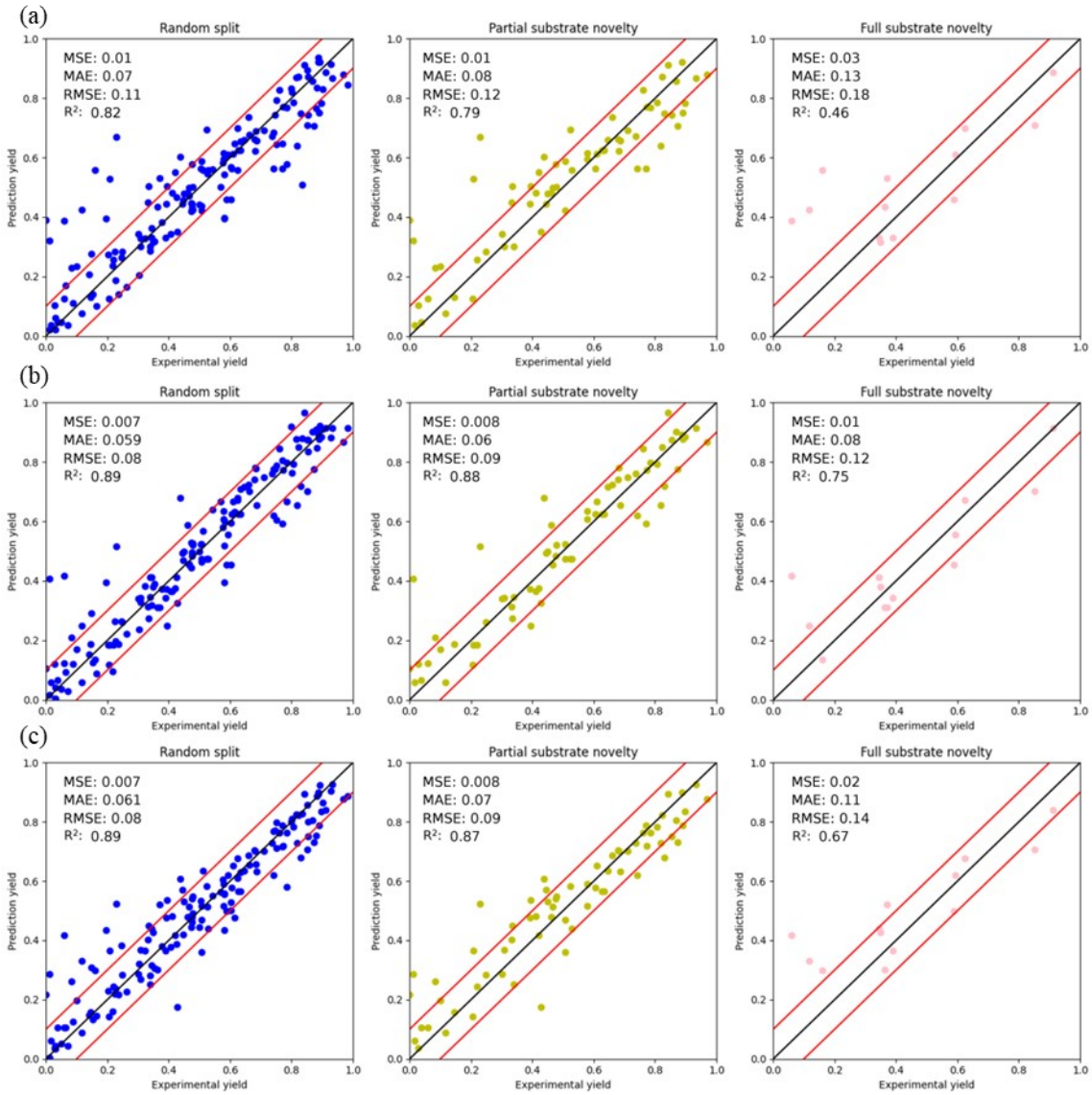


Figure S26. Model performance of EDC condition using the BERT algorithm, trained under three different scenarios: (a) without intermediate, (b) amine + acid + intermediate → amide (c) amine + intermediate → amide. Each scenario is evaluated using three data splitting strategies: Random split, Partial substrate novelty, and Full substrate novelty. The red lines indicate the ideal prediction line (center) and ±10% error margins. Performance metrics (MSE, MAE, RMSE, R^2) are provided for each case.

S3.9 Five-fold cross validation of model

Due to the out-of-sample issue of the full novelty substrate data, we were unable to provide multiple partial and full substrate novelty test sets. Nevertheless, we conducted 5-fold cross-validation to

create five different out-of-sample randomly split test datasets under six different conditions. These results were illustrated in Figure S24 to Figure S29

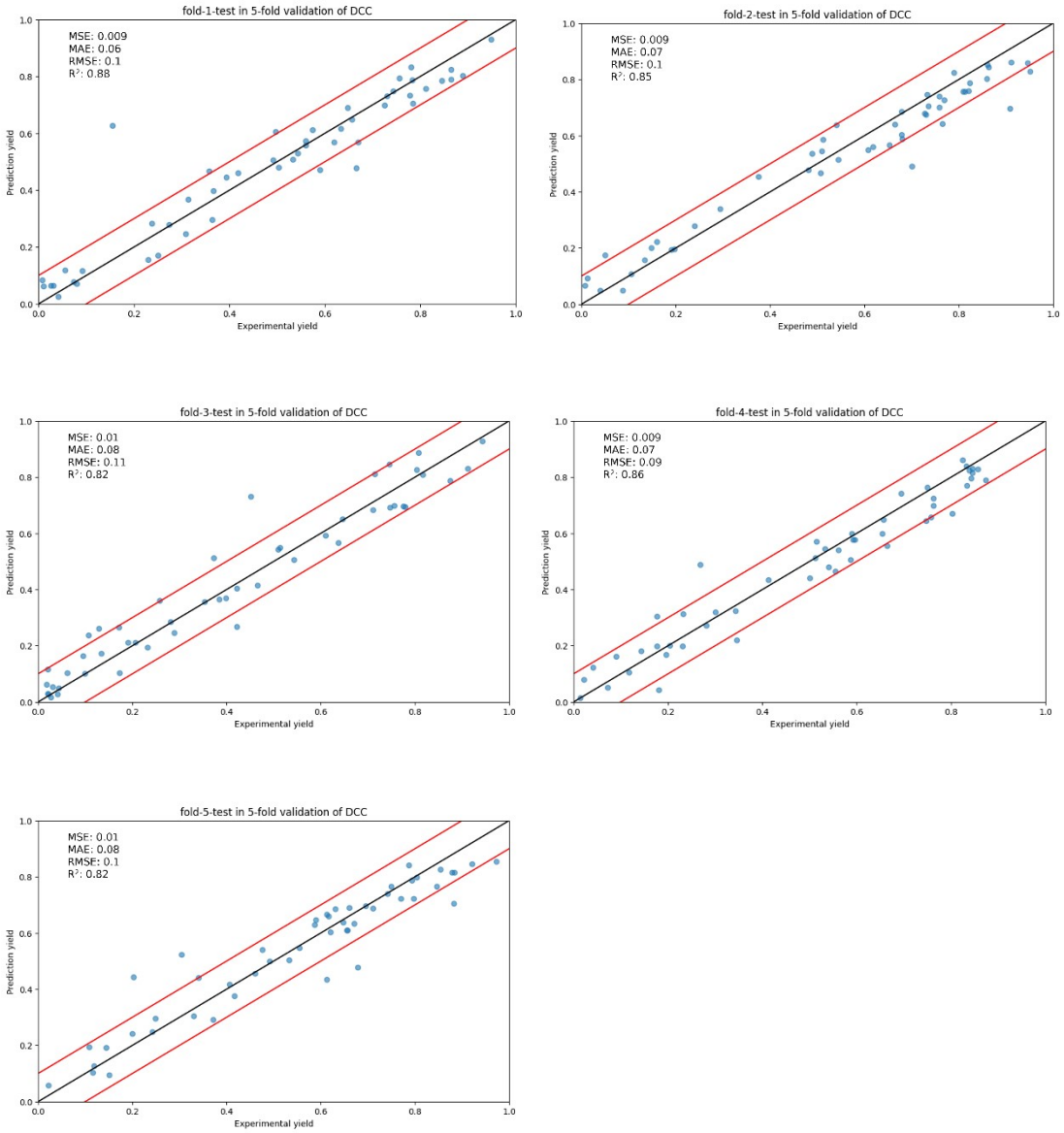
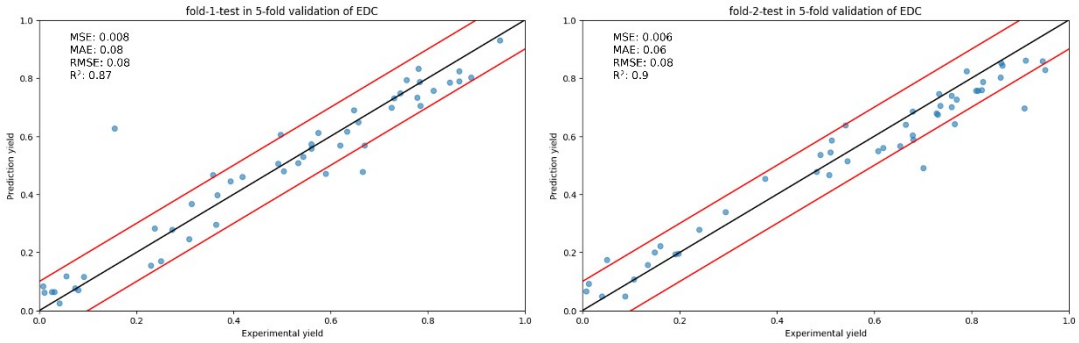


Figure S27. 5-fold cross validation of DCC condition.



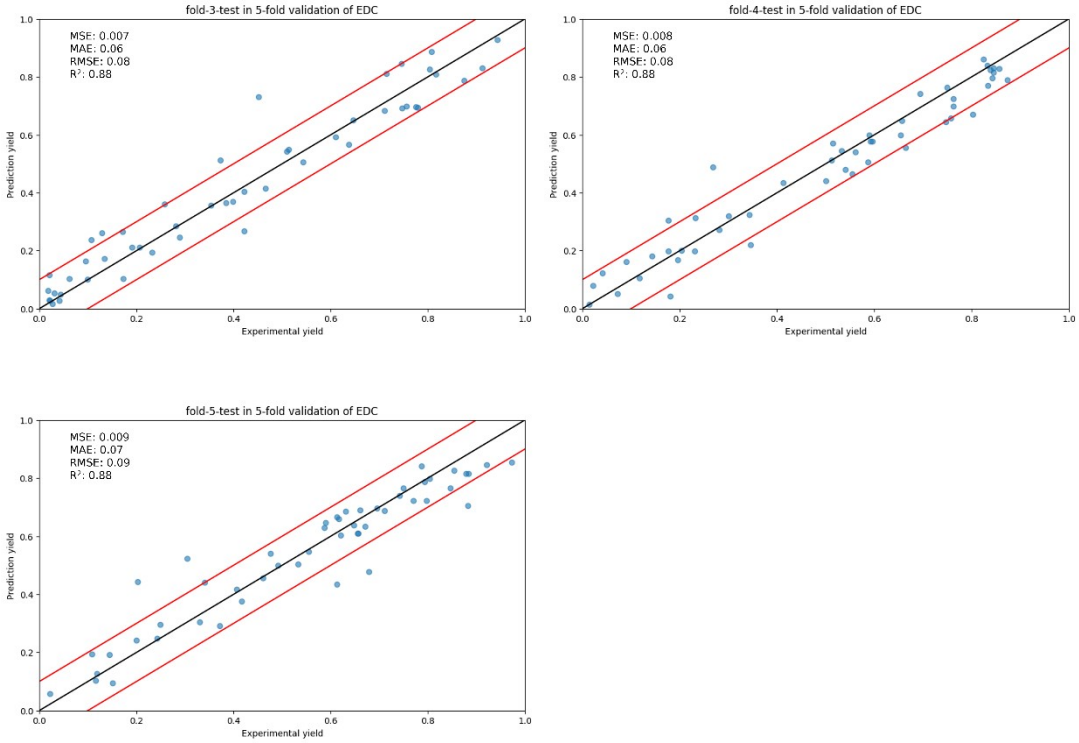


Figure S28. 5-fold cross validation of EDC condition.

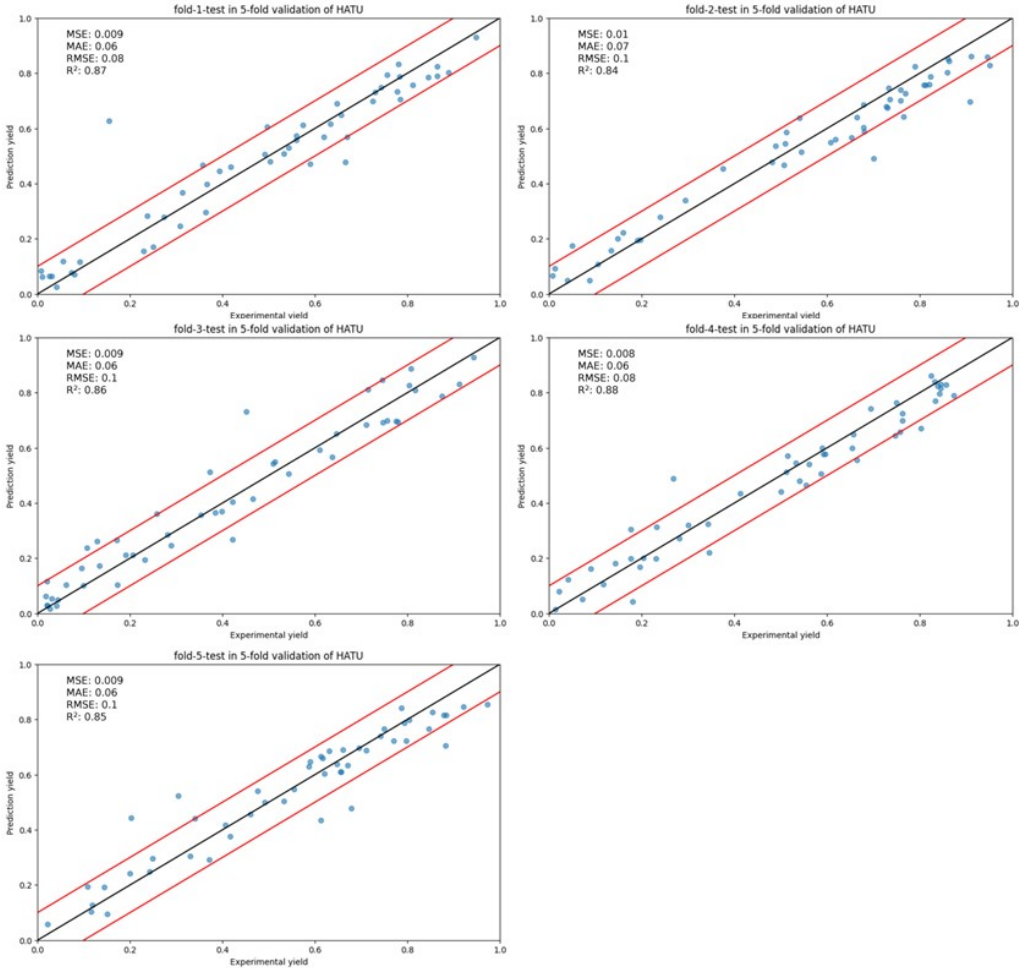
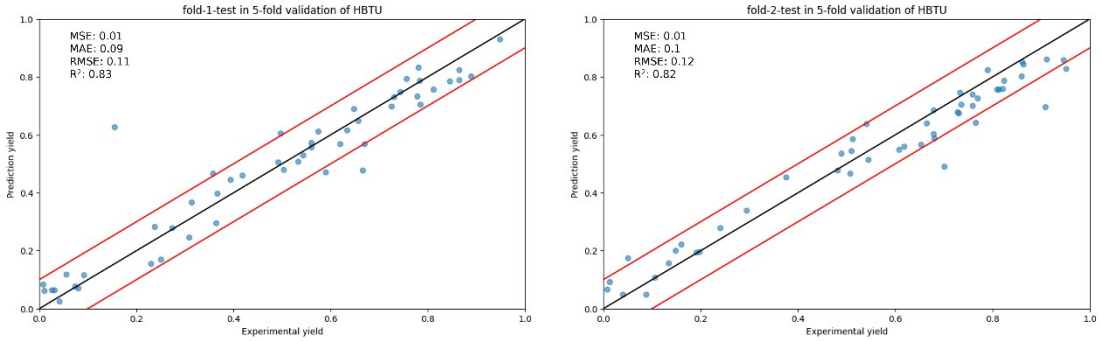


Figure S29. 5-fold cross validation of HATU condition.



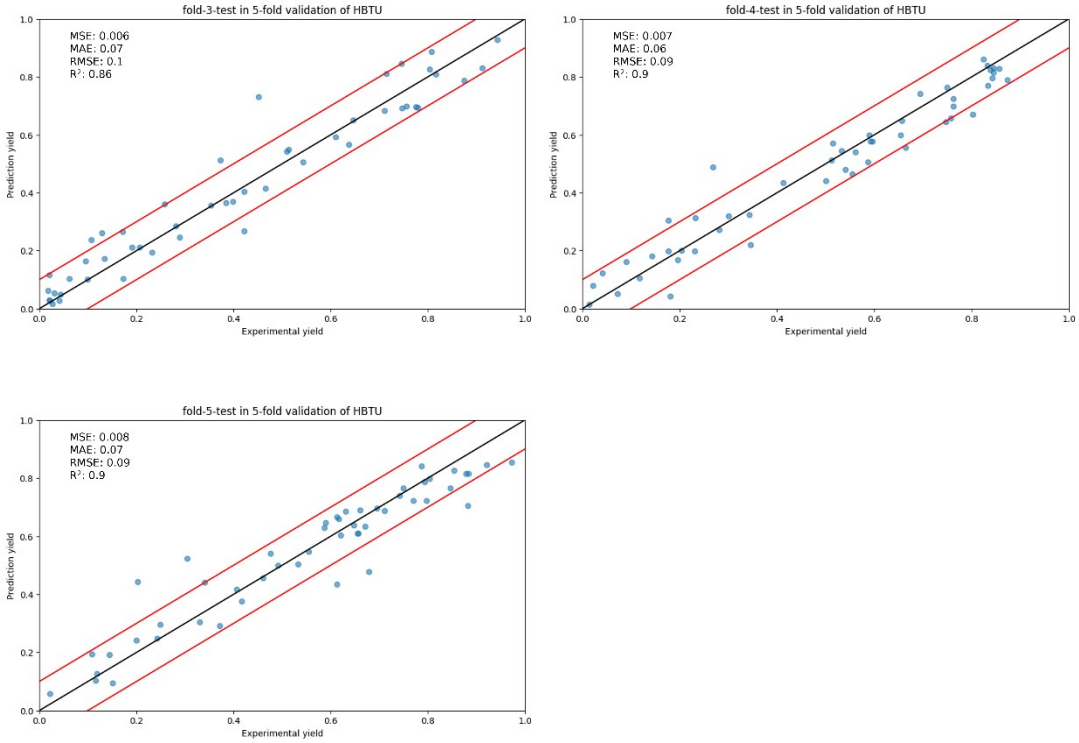
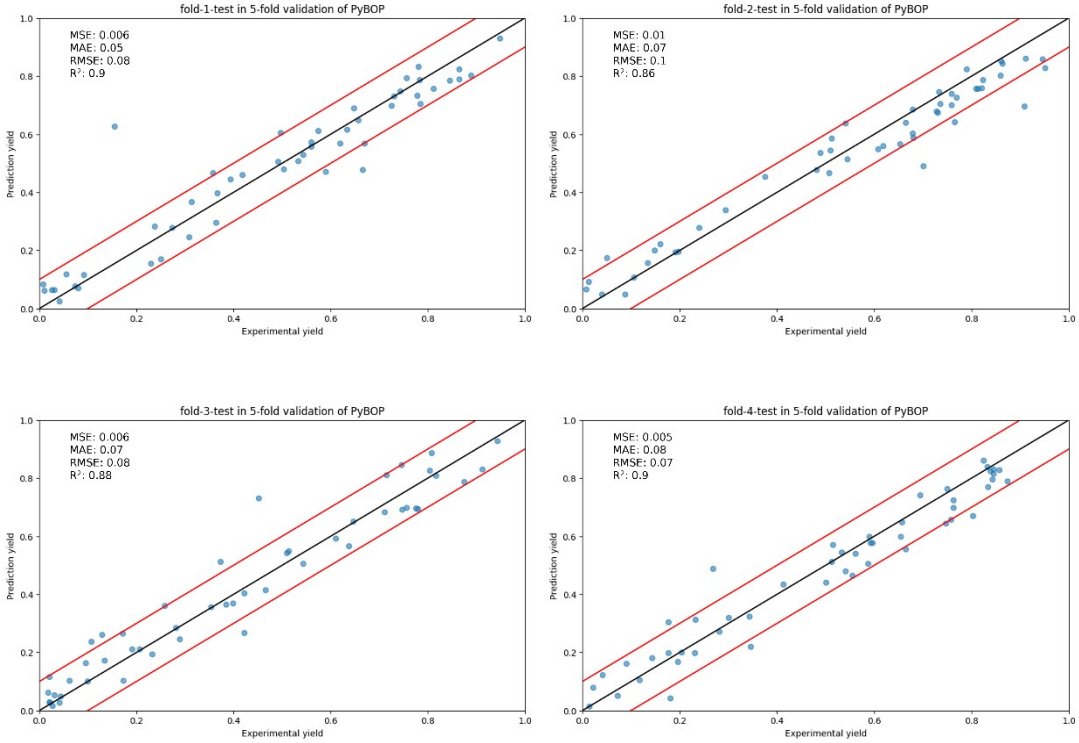


Figure S30. 5-fold cross validation of HBTU condition.



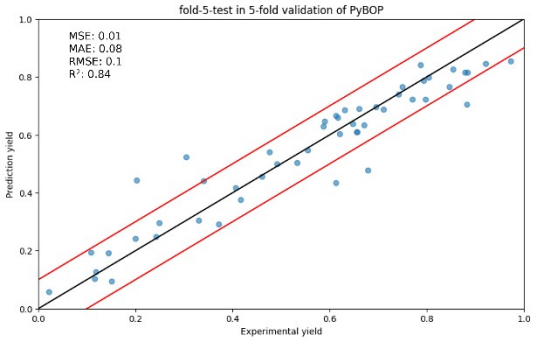


Figure S31. 5-fold cross validation of PyBOP condition.

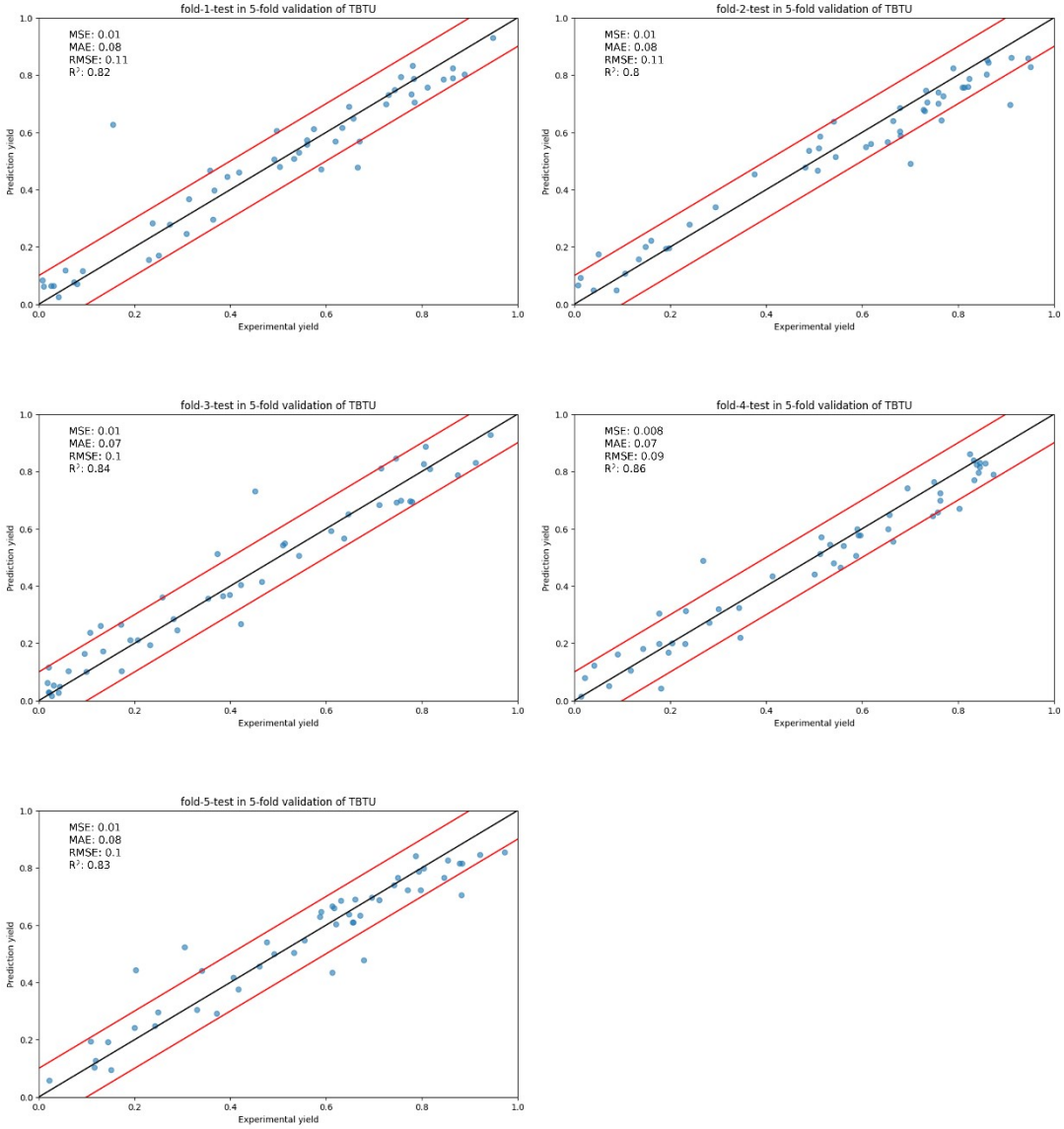
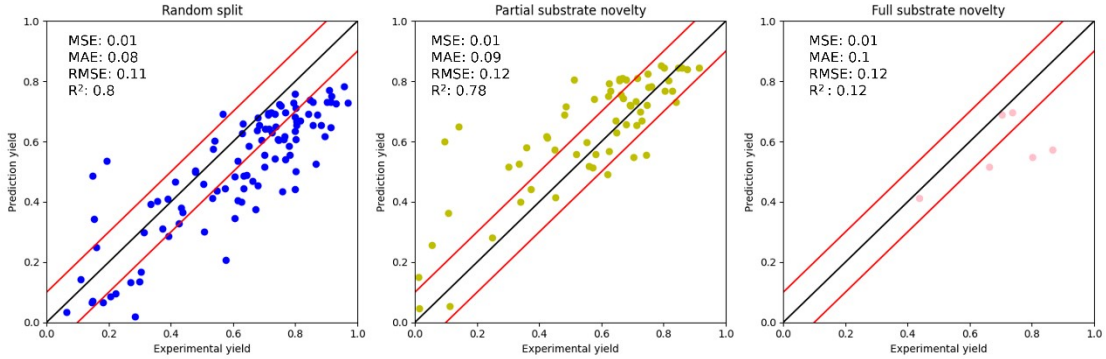


Figure S32. 5-fold cross validation of TBTU condition

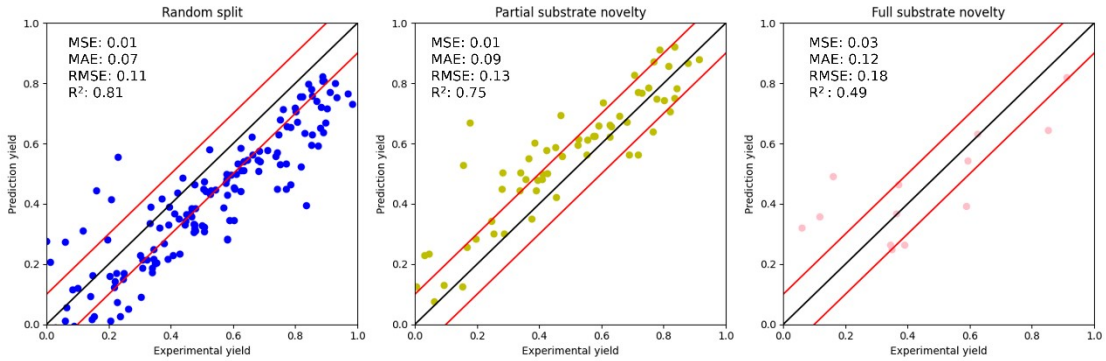
S3.10 Performance of conditions encoded with one-hot

We also evaluated the performance of BERT yield prediction model when the conditions were encoded with one-hot, which did not contain physical organic chemistry knowledge related to conditions. The performance of model toward the data set of full substrate novelty decreased dramatically (Figure S30), indicating the key role of intermediate knowledge in elevating the performance of model.

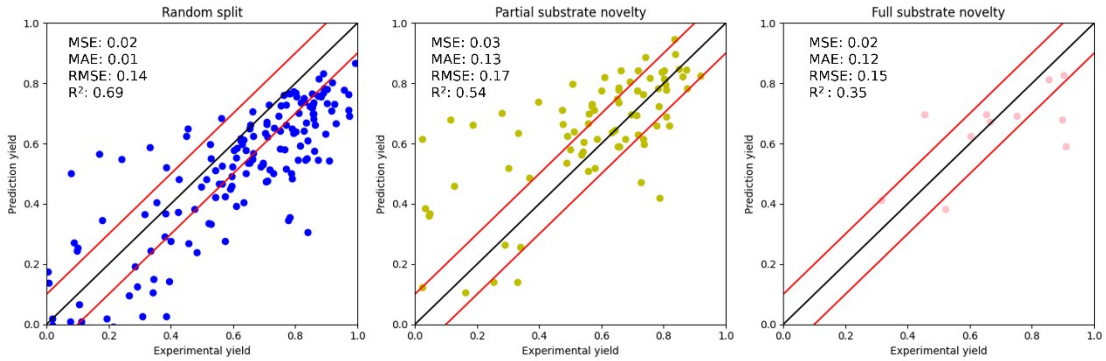
DCC



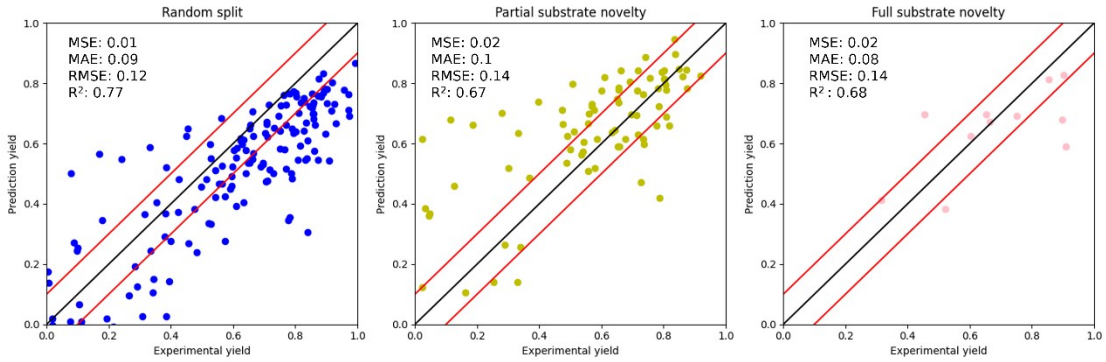
EDC



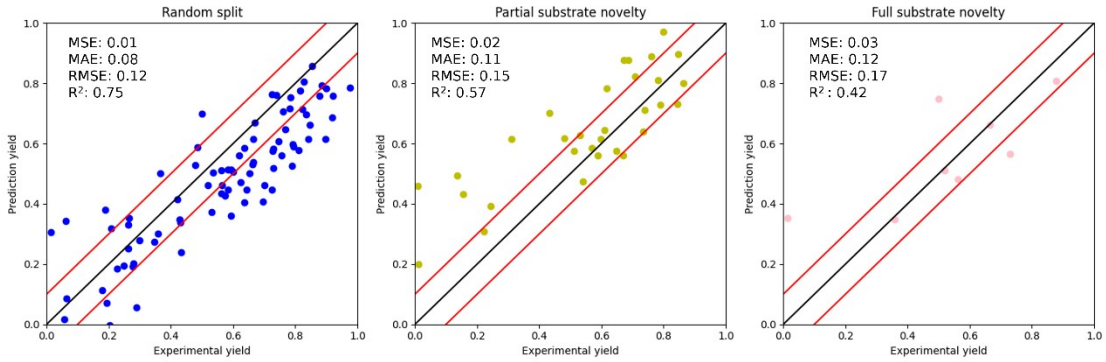
HATU



HBTU



PyBOP



TBTU

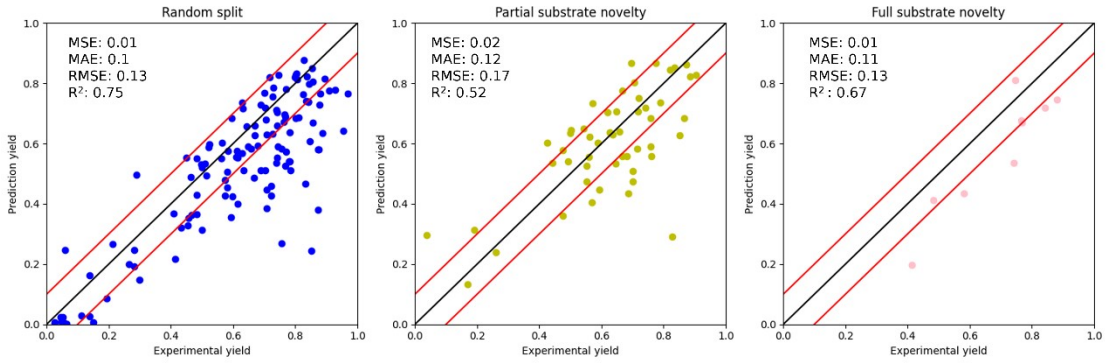


Figure S33. Performance of conditions encoded with one-hot

Predict the average yield in the training set. To verify the accuracy of the model, we also attempted to predict the average yield in the training set. However, the performance metrics for these test data splits would be extremely poor. For all six conditions, the R^2 metrics are nearly zero. Please find the full metric results for the six conditions in Figure S31 and Table S14 below.

Table S14. Metric results of predicting the average yield in the training set

Metrics\Conditions	**DCC**	**EDC**	**HATU**	**HBTU**	**PyBOP**	**TBTU**
R^2	0.00	0.00	0.00	0.00	0.00	0.00
MSE	0.08	0.08	0.08	0.09	0.08	0.08
RMSE	0.28	0.28	0.28	0.30	0.29	0.28
MAE	0.25	0.25	0.23	0.26	0.24	0.24

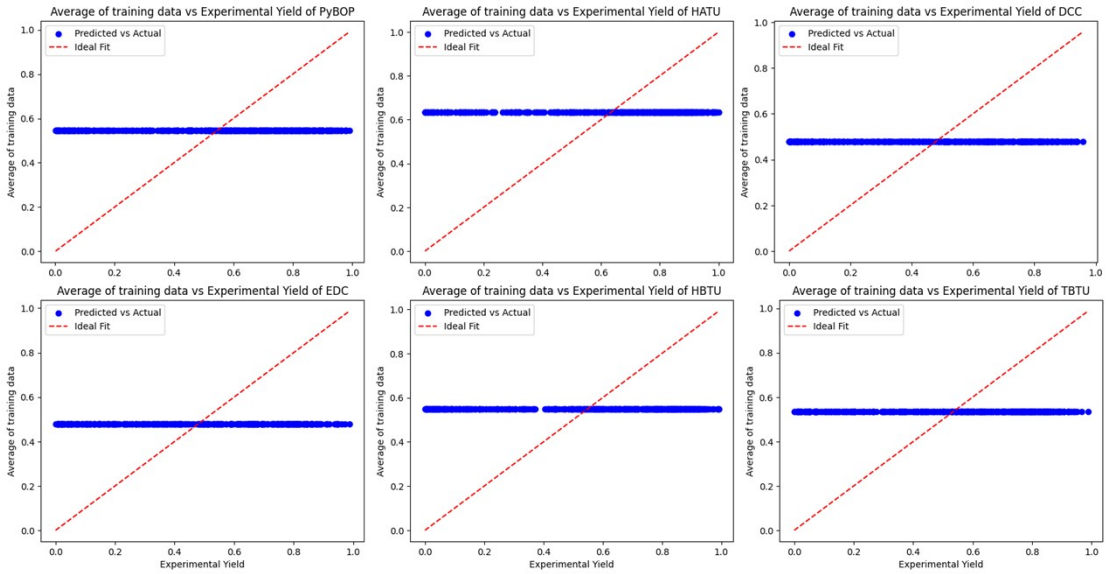


Figure S34. Model performance by predicting the average yield in the training set

Besides, we also studied the effect of activation function. Thus, for BERT model, we changed the activation function of output layer into Sigmoid. However, similar or lower performance was obtained in the most cases (Table S15).

Table S15. The effect of Sigmoid activation function

Conditions	**TBTU**			**HATU**	
Splitting	Metrics	BERT	Embedded BERT	BERT	Embedded BERT
Random split	MAE	10%	7.0%[1] (8.0%[2])	10%	9.0%[1] (6.0%[2])
	RMSE	13%	10%[1] (11%[2])	13%	11%[1] (9.0%[2])
	R^2	0.70	0.83[1] (0.81[2])	0.71	0.77[1] (0.87[2])
Partial novelty	MAE	12%	11%[1] (9.0%[2])	13%	11%[1] (8.0%[2])
	RMSE	16%	14%[1] (12%[2])	16%	14%[1] (11%[2])
	R^2	0.59	0.64[1] (0.73[2])	0.54	0.65[1] (0.81[2])

Full novelty	MAE	11%	11%[1] (7.0%[2])	12%	10%[1] (6.0%[2])
	RMSE	13%	13%[1] (9.0%[2])	14%	13%[1] (7.0%[2])
	R^2	0.62	0.62[1] (0.82[2])	0.34	0.49[1] (0.84[2])
Conditions	**PyBOP**			**DCC**	
Splitting	Metrics	BERT	Embedded BERT	BERT	Embedded BERT
Random split	MAE	9.0%	6.0%[1] (7.0%[2])	8%	6.0%[1] (8.0%[2])
	RMSE	11%	8.0%[1] (9.0%[2])	11%	11%[1] (10%[2])
	R^2	0.76	0.88[1] (0.84[2])	0.81	0.88[1] (0.84[2])
Partial novelty	MAE	12%	8.0%[1] (7.0%[2])	9%	8.0%[1] (8.0%[2])
	RMSE	15%	11%[1] (9.0%[2])	13%	13%[1] (12%[2])
	R^2	0.58	0.74[1] (0.80[2])	0.73	0.74[1] (0.79[2])
Full novelty	MAE	12%	6.0%[1] (9.0%[2])	10%	7.0%[1] (7.0%[2])
	RMSE	18%	8.0%[1] (12%[2])	12%	7.0%[1] (11%[2])
	R^2	0.42	0.90[1] (0.73[2])	0.11	0.63[1] (0.39[2])
Conditions	**HBTU**			**EDC**	
Splitting	Metrics	BERT	Embedded BERT	BERT	Embedded BERT
Random split	MAE	8.0%	6.0%[1] (6.0%[2])	8.0%	6.0%[1] (6.0%[2])
	RMSE	11%	8.0%[1] (10%[2])	11%	9.0%[1] (8.0%[2])
	R^2	0.83	0.82[1] (0.87[2])	0.83	0.84[1] (0.89[2])
Partial novelty	MAE	12%	9.0%[1] (7.0%[2])	8.0%	6.0%[1] (8.0%[2])
	RMSE	12%	12%[1] (11%[2])	10%	9.0%[1] (9.0%[2])
	R^2	0.70	0.74[1] (0.81[2])	0.82	0.88[1] (0.83[2])
Full novelty	MAE	9.0%	10%[1] (8.0%[2])	12%	8.0%[1] (12%[2])
	RMSE	14%	18%[1] (10%[2])	17%	11%[1] (15%[2])
	R^2	0.66	0.48[1] (0.81[2])	0.54	0.76[1] (0.60[2])

[1]Intermediate knowledge was embedded into the model; [2]Intermediate knowledge was embedded into the model but the representation (SMILES or ECFP) on acid was removed.

The performance under six conditions: After realizing the high performance of model enhanced by intermediate knowledge in the selected conditions, we also intend to know whether the performance of an embedded model could also be elevated when the data from six conditions were combined. With this notion in mind, we subsequently examined the performance of the embedded BERT model under six conditions, and the results are illustrated in Table S16.

Table S16. The performance under six conditions

Splitting	R^2	RMSE	MAE
Random split	0.77[a]	12%[a]	9.0%[a]
Partial novelty	0.71[a]	14%[a]	10%[a]
Full novelty	0.62[a]	10%[a]	8.0%[a]
Random split	0.85[b]	10%[b]	7.0%[b]
Partial novelty	0.80[b]	11%[b]	8.0%[b]
Full novelty	0.65[b]	9.0%[b]	8.0%[b]

S3.11 The performance toward external data set from literatures

To rigorously evaluate the performance of our prediction model, we conducted tests on reactions sourced from the medicinal chemistry literature. The selection process was as follows:

Literature Search: We initiated our search by drawing the general formula for amide coupling reactions using SciFinder (Figure S32). We then applied the "Structure Match" filter to select reactions based on substructure similarity.

Focus on Medicinal Chemistry: To ensure relevance to medicinal chemistry, we screened literature from drug- and biologically-related journals, including but not limited to Journal of Medicinal Chemistry, European Journal of Medicinal Chemistry, Bioorganic & Medicinal Chemistry, Organic & Biomolecular Chemistry, and ACS Medicinal Chemistry Letters.

Randomized Selection: Given the extensive amount of reported data (over 19,000 reactions), we randomly selected five reactions from each identified literature source for every reaction condition tested.

Diversity and Independence from Training Data: Importantly, all selected substrate combinations were distinct from those in our HTE dataset, ensuring that the test reactions were entirely independent of the training data. To further ensure robustness, we expanded our test set by selecting an additional ten reactions per condition for external validation.

This strategy allowed us to assess the model's generalizability and performance on a diverse set of reactions representative of medicinal chemistry applications. HATU condition results in Scifinder in shown in Figure S33. As a result, the BERT model performed quite well, achieving a MAE of 10% and the prediction results of 163 reactions were less than or equal to 5%. The prediction results are illustrated in Figure S34 and Table S17.

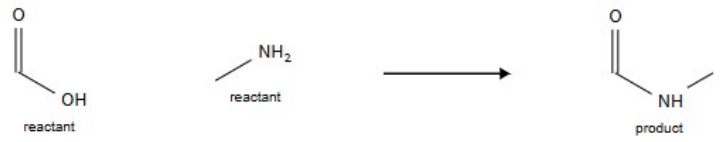


Figure S35. Search formula for Amide Coupling

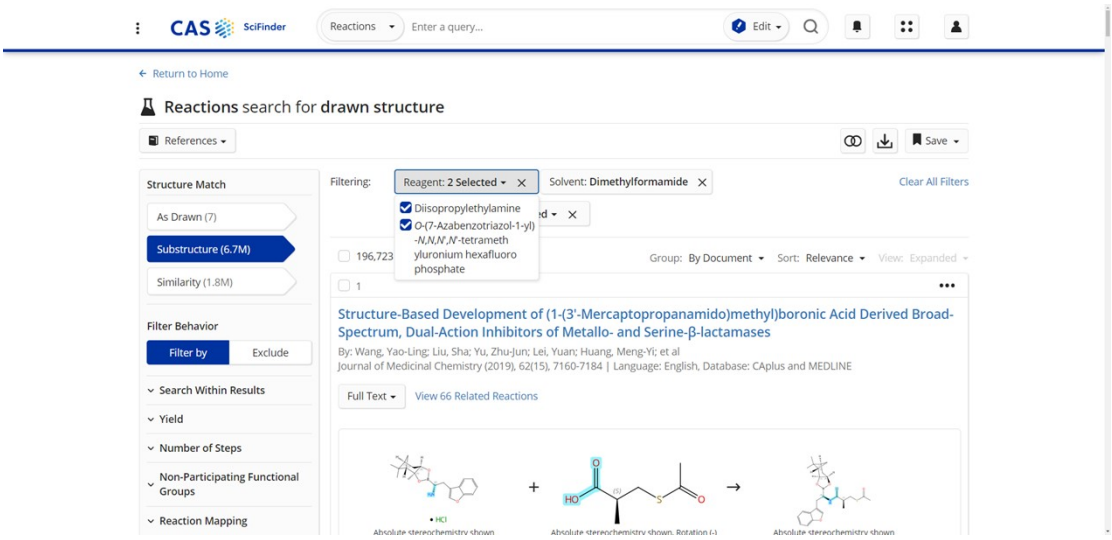


Figure S36. HATU results in Scifinder

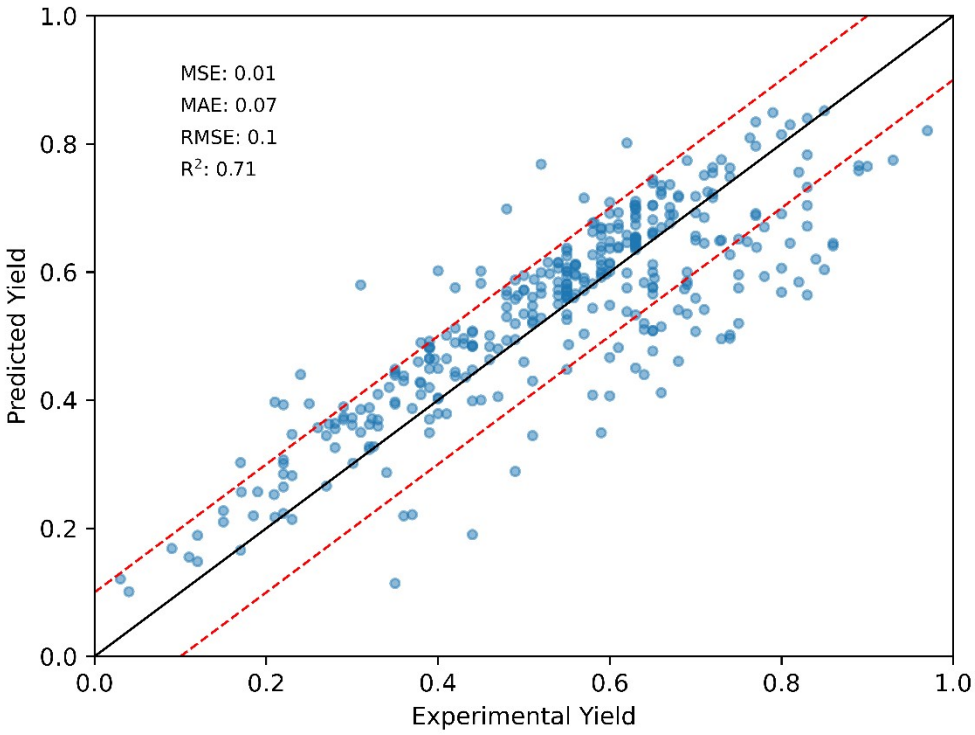


Figure S37. Performance in reactions related to medicinal chemistry

Table S17. The performance toward reactions related to medicine chemistry

Condition	Subtrate1	Substrate2	Product	Yield	Ref.	Prediction	Error
HATU	C(OC(N[C@@H](CCCCNC(OCC1C=2C(C=3C1=CC=CC3)=CC=CC2)=O)C(O)=O)=O)C4C=5C(C=6C4=CC=CC6)=CC=CC5	C#CCN	C(OC(N[C@@H](CCCCNC(OCC1C=2C(C=3C1=CC=CC3)=CC=CC2)=O)C(NCC#C)=O)=O)C4C=5C(C=6C4=CC=CC6)=CC=CC5	0.86	Bioorganic & Medicinal Chemistry (2022), 57, 116646	0.645	0.215
HATU	O=C(O)CNC	NC1=CC=CC(=C1C)C	O=C(NC1=CC=CC(=C1C)C)CNC	0.85	International Journal of Molecular Sciences (2022), 23(20), 12675	0.8518	0.0018
HATU	COc1ccc(C)c2c(=O)cc(C(=O)O)[nH]c12	Nc1ccc(-c2ccc(Cl)cc2)cc1	COc1ccc(C)c2c(=O)cc(C(=O)Nc3ccc(-c4ccc(Cl)cc4)cc3)[nH]c12	0.4	Medicinal Chemistry Research (2022), 31(3), 485-496	0.4495	0.0495
HATU	COc1ccc(C)c2c(=O)cc(C(=O)	Nc1cccc(Cl)c1Cl	COc1ccc(C)c2c(=O)cc(C(=O)	0.43	Medicinal Chemistry Research	0.4963	0.0663

	O)[nH]c12		Nc3cccc(Cl)c3Cl)[nH]c12		(2022), 31(3), 485-496		
HATU	COc1ccc2oc(-c3ccc(OCc4ccccc4)cn3)c(C(=O)O)c2c1	C1CNC1	COc1ccc2oc(-c3ccc(OCc4ccccc4)cn3)c(C(=O)N3CCC3)c2c1	0.73	Journal of Medicinal Chemistry (2022), 65(1), 409-423	0.7754	0.0454
HATU	[3H]C([3H])([3H])N[C@H](C(=O)N[C@H](C(=O)N(C)[C@@H]([C@@H](C)CC)[C@@H](CC(=O)N1CCC[C@H]1[C@H](OC)[C@@H](C)C(=O)N[C@H](C)[C@@H](O)c1ccccc1)OC)C(C)C)C(C)C	O=C(O)CCCCN1C(=O)C=CC1=O	[3H]C([3H])([3H])N(C(=O)CCCCN1C(=O)C=CC1=O)[C@H](C(=O)N[C@H](C(=O)N(C)[C@@H]([C@@H](C)CC)[C@@H](CC(=O)N1CCC[C@H]1[C@H](OC)[C@@H](C)C(=O)N[C@H](C)[C@@H](O)c1ccccc1)OC)C(C)C)C(C)C	0.32	Journal of Medicinal Chemistry (2022), 6953-6968	0.3235	0.0035
HATU	[N-	FC1=CC=	[N-	0.65	Russian	0.6119	0.038

	]=[N+]=NC1=CC=C(C=C1)C(=O)O	C(C=C1)CN	]=[N+]=NC1=CC=C(C=C1)C(=O)NCC2=CC=C(F)C=C2		Journal of General Chemistry (2022), 92(10), 2119-2131		1
HATU	O=C(O)C(F)(C)C	OC1=CC=CC(=C1)CCN	O=C(NCCC=1C=CC=C(O)C1)C(F)(C)C	0.48	European Journal of Medicinal Chemistry (2021), 226, 113870	0.6986	0.2186
HATU	O=C(O)CCCCCCCCCC(=O)OCc1ccccc1	CC(C)(C)OC(=O)Nc1ccc(-c2cccs2)cc1NC(=O)c1ccc(N)cc1	CC(C)(C)OC(=O)Nc1ccc(-c2cccs2)cc1NC(=O)c1ccc(NC(=O)CCCCCCCCCC(=O)OCc2ccccc2)cc1	0.58	Journal of Medicinal Chemistry (2022), 65(7), 5642-5659	0.6268	0.0468
HATU	O=C(O)COCCOCCOCC(=O)OCc1ccccc1	CC(C)(C)OC(=O)Nc1ccc(-c2cccs2)cc1NC(=O)c1ccc(N)cc1	CC(C)(C)OC(=O)Nc1ccc(-c2cccs2)cc1NC(=O)c1ccc(NC(=O)COCCOCCOCC(=O)OCc2ccccc2)cc1	0.58	Journal of Medicinal Chemistry (2022), 65(7), 5642-5659	0.663	0.083

HATU	O=C(O)c1ccc(CN2CC/C(=C\c3ccccn3)C2=O)cc1	Nc1ccc(F)cc1N	Nc1cc(F)ccc1NC(=O)c1ccc(CN2CC/C(=C\c3ccccn3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6381	0.0081
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccccc3)C2=O)cc1	Nc1ccc(F)cc1N	Nc1cc(F)ccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccccc3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6892	0.0592
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccccc3Cl)C2=O)cc1	Nc1ccc(F)cc1N	Nc1cc(F)ccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccccc3Cl)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.7105	0.0805
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccccn3)C2=O)cc1	Nc1ccc(F)cc1N	Nc1cc(F)ccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccccn3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6981	0.0681
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3cccnc3)C2=O)cc1	Nc1ccc(F)cc1N	Nc1cc(F)ccc1NC(=O)c1ccc(CN2CCC/C(=C\c3cccnc3)C2=O	0.63	European Journal of Medicinal Chemistry (2022), 229,	0.6998	0.0698

			)cc1		114049		
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccco3)C2=O)cc1	Nc1ccc(F)cc1N	Nc1cc(F)ccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccco3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6488	0.0188
HATU	O=CC1=CC=C(C=C1)C(=O)O	O(C1=CC=C(C=C1)CN)C	O=CC1=CC=C(C=C1)C(=O)NCC2=CC=C(OC)C=C2	0.6	Journal of Medicinal Chemistry (2021), 64(12), 8423-8436	0.709	0.109
HATU	O=C(O)CCOCCOCCOCCNC(=O)CCC(=O)N1C=2C=CC=CC2C#CC=3C=CC=CC3C1	C(CNC(OCC1=CC=C(NC([C@@H](NC([C@H](C(C)C)N)=O)CCCNC(N)=O)=O)C=C1)=O)(=O)[C@]23[C@]4(C)[C@@](C[C@]2(OC(CCC)O3)[H])([C@]5([C@](F)([C@@H](O)C4)[C@]6(C)C([C@@H](	C(CNC(OCC1=CC=C(NC([C@@H](NC([C@@H](NC(COCCOCCOCCNC(CCC(=O)N2C=3C(C#CC=4C(C2)=CC=CC4)=CC=CC3)=O)=O)C(C)C)=O)CCCNC(N)=O)=O)C=C1)=O)(=O)[C@]56[C@	0.23	Journal of Medicinal Chemistry (2021), 64(16), 11958-11971	0.3468	0.1168

		F)C5)=CC(=O)C=C6)[H])[H]	]7(C)[C@@](C[C@]5(OC(CCC)O6)[H])([C@]8([C@](F)([C@@H](O)C7)[C@]9(C)C([C@@H](F)C8)=CC(=O)C=C9)[H])[H]				
HATU	O=C(O)CCCCCCC(=O)OC	N#CCC1=CC=C(N)C=C1	N#CCC1=CC=C(C=C1)NC(=O)CCCCCCC(=O)OC	0.45	Journal of Medicinal Chemistry (2021), 64(22), 16573-16597	0.4003	0.0497
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccncc3)C2=O)cc1	Nc1ccc(F)cc1N	Nc1cc(F)ccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccncc3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.7043	0.0743
HATU	O=C(Cc1ccccc1)Nc1nnc(N2CCC(CCOCc3ccc(C(=O)O)cc3)CC2)s1	Cc1ccccc1CN1CCC(CN)CC1	Cc1ccccc1CN1CCC(CNC(=O)c2ccc(COCCC3CCN(c4nnc(NC(=O)Cc5ccccc5)	0.55	European Journal of Medicinal Chemistry (2022), 243, 114686	0.6251	0.0751

			s4)CC3)cc2)CC1				
HATU	O=C(Cc1ccccc1)Nc1nnc(N2CCC(CCOCc3cccc(C(=O)O)c3)CC2)s1	Cc1ccccc1CN1CCC(CN)CC1	Cc1ccccc1CN1CCC(CNC(=O)c2cccc(COCCC3CCN(c4nnc(NC(=O)Cc5ccccc5)s4)CC3)c2)CC1	0.55	European Journal of Medicinal Chemistry (2022), 243, 114686	0.6331	0.0831
HATU	O=C(Cc1ccccc1)Nc1nnc(N2CCC(OCc3ccc(C(=O)O)cc3)CC2)s1	CCN1CCC(N)CC1	CCN1CCC(NC(=O)c2ccc(COC3CCN(c4nnc(NC(=O)Cc5ccccc5)s4)CC3)cc2)CC1	0.56	European Journal of Medicinal Chemistry (2022), 243, 114686	0.5658	0.0058
HATU	O=C(O)C1=NC=C(Br)C=C1OC	Cl.FC(F)(F)CN	O=C(NCC(F)(F)F)C1=NC=C(Br)C=C1OC	0.83	Journal of Medicinal Chemistry (2024), 67(7), 5233-5258	0.7832	0.0468
HATU	O=C(O)CN(C(=O)CN(C(=O)OCC1C=2C=CC=CC2C=3C=CC=CC31	NCCCCCCCCCCCCCC	O=C(OCC1C=2C=CC=CC2C=3C=CC=CC31)N(CC#C)CC(=O)N(CC	0.5	Journal of Medicinal Chemistry (2024), 67(7), 5945-5956	0.4599	0.0401

	)CC#C)CC#C		#C)CC(=O)NCCCCCCCCCCCCCCC				
HATU	c1cnc2[nH]ccc2c1	COc1cc(C(=O)O)cc(OC)c1OCc1ccccc1	COc1cc(C(=O)n2ccc3cccnc32)cc(OC)c1OCc1ccccc1	0.49	European Journal of Medicinal Chemistry (2022), 242, 114682	0.4948	0.0048
HATU	Cc1ccc(N2CCN(C(=O)OC(C)(C)C)CC2)c2c1CC[C@H](N)C2	Cc1cc(Cl)ccc1-c1ccc(C(=O)O)nc1	Cc1cc(Cl)ccc1-c1ccc(C(=O)N[C@H]2CCc3c(C)ccc(N4CCN(C(=O)OC(C)(C)C)CC4)c3C2)nc1	0.53	European Journal of Medicinal Chemistry (2022), 229, 114059	0.5976	0.0676
HATU	Cc1ccc(N2CCN(C)CC2)c2c1CC[C@H](N)C2	Cc1cc(Cl)ccc1-c1ccc(C(=O)O)nc1	Cc1cc(Cl)ccc1-c1ccc(C(=O)N[C@H]2CCc3c(C)ccc(N4CCN(C)CC4)c3C2)nc1	0.55	European Journal of Medicinal Chemistry (2022), 229, 114059	0.5717	0.0217
HATU	Cc1ccc(N2CCN(C)CC2)c2c1	Cc1cc(Cl)ccc1-c1ncc(C(=	Cc1cc(Cl)ccc1-c1ncc(C(=	0.56	European Journal of Medicinal	0.6111	0.0511

	CC[C@@H](N)C2	O)O)cn1	O)N[C@@H]2CCc3c(C)ccc(N4CCN(C)CC4)c3C2)cn1		Chemistry (2022), 229, 114059		
HATU	O=C(O)CC1SCCNC1=O	OC1=CC=C(N)C=C1	O=C(NC1=CC=C(O)C=C1)CC2SCCNC2=O	0.57	ACS Medicinal Chemistry Letters (2021), 12(2), 302-308	0.7158	0.1458
HATU	Cc1ccc(N2CCN(C)CC2)c2c1CC[C@H](N)C2	Cc1cc(Cl)ccc1-c1ncc(C(=O)O)cn1	Cc1cc(Cl)ccc1-c1ncc(C(=O)N[C@H]2CCc3c(C)ccc(N4CCN(C)CC4)c3C2)cn1	0.54	European Journal of Medicinal Chemistry (2022), 229, 114059	0.6151	0.0751
HATU	O=C(O)c1ccc(CN2CC/C(=C\c3ccccn3)C2=O)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CC/C(=C\c3ccccn3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6453	0.0153
HATU	O=C(O)c1ccc(CN2CC/C(=C\c3ccc(Cl)c(C(F)(F)F)	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccc(C	0.63	European Journal of Medicinal Chemistry (2022),	0.6843	0.0543

	c3)C2=O)cc1		l)c(C(F)(F)F)c3)C2=O)cc1		229, 114049		
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccccc3)C2=O)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccccc3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6542	0.0242
HATU	O=C(O)C1(C2=NC=C(Br)C=C2)COC1	FC1=CC=C(N)C=C1	O=C(NC1=CC=C(F)C=C1)C2(C3=NC=C(Br)C=C3)COC2	0.86	ACS Medicinal Chemistry Letters (2020), 11(4), 550-557	0.6412	0.2188
HATU	O=C(O)C1(C2=CC=C(Br)C=C2)COC1	FC1=CC=C(N)C=C1	O=C(NC1=CC=C(F)C=C1)C2(C3=CC=C(Br)C=C3)COC2	0.84	Journal of Medicinal Chemistry (2022), 65(8), 6001-6016	0.6201	0.2199
HATU	O=C(O)C=1C=CC=C(Br)C1C	O=CC1=CC=C(N)C=C1	O=CC1=CC=C(C=C1)NC(=O)C=2C=CC=C(Br)C2C	0.44	Journal of Medicinal Chemistry (2023), 66(24), 16807-16827	0.4471	0.0071
HATU	O=C(O)CCCCCCCCCCCCC	O=C1C=CC(=O)N1CCCN	O=C1C=CC(=O)N1CCCNC(=O)CCCC	0.37	Journal of Medicinal Chemistry (2020),	0.2214	0.1486

			CCCCCC CCCCC		63(19), 10782-10795		
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccccn3)C2=O)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccccn3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.635	0.005
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3cccnc3)C2=O)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CCC/C(=C\c3cccnc3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.652	0.022
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccco3)C2=O)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccco3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6739	0.0439
HATU	O=C(O)c1ccc(CN2CCC/C(=C\c3ccncc3)C2=O)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CCC/C(=C\c3ccncc3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6554	0.0254
HATU	O=C(O)C1=CN(N=C1C(F)F)	S(C[C@@H](C(OC)=O)N)C	C(N[C@@H](CSC1=C(Cl)C	0.58	Journal of Agricultural and	0.4083	0.1717

	C	1=C(Cl)C=C(C(F)(F)F)C=N1	=C(C(F)(F)F)C=N1)C(OC)=O)(=O)C=2C(C(F)F)=NN(C)C2		Food Chemistry (2021), 69(38), 11470-11484		
HATU	C(C(O)=O)(C([2H])([2H])[2H])(C([2H])([2H])[2H])O	FC=1C=C(F)C(N)=CC1Cl	N(C(C(C([2H])([2H])[2H])(C([2H])([2H])[2H])O)=O)C1=C(F)C=C(F)C(Cl)=C1	0.36	Journal of Labelled Compounds and Radiopharmaceuticals (2020), 63(10), 434-441	0.2195	0.1405
HATU	O=C(O)c1ccc(CN2CCC3=C(C2)C(=O)N(Cc2cccc(F)c2)C2=NCCN23)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CN2CCC3=C(C2)C(=O)N(Cc2cccc(F)c2)C2=NCCN23)cc1	0.59	European Journal of Medicinal Chemistry (2022), 238, 114049	0.6692	0.0792
HATU	COc1ccc(/C=C/C(=O)Nc2ccccc2C(=O)O)cc1OC	CN	CNC(=O)c1ccccc1NC(=O)/C=C/c1ccc(OC)c(OC)c1	0.61	European Journal of Medicinal Chemistry (2022), 242, 114685	0.675	0.065
HATU	Cc1ccc(/C=C2\CCCN(Cc3ccc(	Nc1ccc(F)cc1N	Cc1ccc(/C=C2\CCCN(Cc3ccc(	0.63	European Journal of Medicinal	0.6429	0.0129

	C(=O)O)cc3)C2=O)cc1		C(=O)Nc4ccc(F)cc4N)cc3)C2=O)cc1		Chemistry (2022), 229, 114049		
HATU	COc1ccc(/C=C/C(=O)Nc2ccc(C(=O)O)cc2)cc1OC	NC1CC1	COc1ccc(/C=C/C(=O)Nc2ccc(C(=O)NC3CC3)cc2)cc1OC	0.66	European Journal of Medicinal Chemistry (2022), 242, 114685	0.726	0.066
HATU	COc1ccc(/C=C/C(=O)Nc2ccc(C(=O)O)cc2)cc1OC	NC1CCC1	COc1ccc(/C=C/C(=O)Nc2ccc(C(=O)NC3CCC3)cc2)cc1OC	0.72	European Journal of Medicinal Chemistry (2022), 242, 114685	0.7632	0.0432
HATU	COc1ccc(/C=C/C(=O)Nc2cccc(C(=O)O)c2)cc1OC	NC1CCC1	COc1ccc(/C=C/C(=O)Nc2cccc(C(=O)NC3CCC3)c2)cc1OC	0.77	European Journal of Medicinal Chemistry (2022), 242, 114685	0.7968	0.0268
HATU	COc1ccc(/C=C2\CCCN(Cc3ccc(C(=O)O)cc3)C2=O)cc1	Nc1ccc(F)cc1N	COc1ccc(/C=C2\CCCN(Cc3ccc(C(=O)Nc4ccc(F)cc4N)cc3)C2=O)cc1	0.63	European Journal of Medicinal Chemistry (2022), 229, 114049	0.6407	0.0107
HATU	C=CC(=O)Nc1cccc(	NCCCCCCNc1cccc	C=CC(=O)Nc1cccc(	0.65	European Journal of	0.7446	0.0946

	Nc2nc(Nc3ccc(N4CCN(CC(=O)O)CC4)cc3OC)ncc2C(F)(F)F)c1	2c1C(=O)N(C1CCC(=O)NC1=O)C2=O	Nc2nc(Nc3ccc(N4CCN(CC(=O)NCCCCCCNc5cccc6c5C(=O)N(C5CCC(=O)NC5=O)C6=O)CC4)cc3OC)ncc2C(F)(F)F)c1		Medicinal Chemistry (2022), 238, 114455		
HATU	C(N[C@H](CC1=CC=C(Cl)C=C1)C(O)=O)(=O)C2=CC(NCC3=C(C)N=CN3)=CC(Br)=C2	NC1CC1	[C@H](CC1=CC=C(Cl)C=C1)(NC(=O)C2=CC(NCC3=C(C)N=CN3)=CC(Br)=C2)C(NC4CC4)=O	0.83	Journal of Medicinal Chemistry (2024), 67(2), 1079-1092	0.5643	0.2657
HATU	O=C(O)CN1CCN(CC1)S(=O)(=O)N2CCC(C)CC2	FC(F)(F)C1=CC=C(OC)C(N)=C1	O=C(NC1=CC(=CC=C1OC)C(F)(F)F)CN2CCN(CC2)S(=O)(=O)N3CCC(C)CC3	0.64	Bioorganic & Medicinal Chemistry Letters (2022), 76, 129013	0.5836	0.0564
HATU	O=C(O)C1=CC(Cl)=CN1	O=C(O)C(N)CCC(C)(C)C	C(N[C@@H](CCC(C)(C)C)C(O)=O)(=	0.22	Bioorganic & Medicinal Chemistry	0.2647	0.0447

			O)C1=CC(Cl)=CN1		Letters (2020), 30(17), 127403		
HATU	O=C(Cc1ccccc1)Nc1nnc(N2CCC(COCc3ccc(C(=O)O)cc3)CC2)s1	CCCCN	CCCCNC(=O)c1ccc(COCC2CCN(c3nnc(NC(=O)Cc4ccccc4)s3)CC2)cc1	0.46	European Journal of Medicinal Chemistry (2022), 243, 114686	0.4635	0.0035
HATU	O=C(Cc1ccccc1)Nc1nnc(N2CCC(CCOCc3cccc(C(=O)O)c3)CC2)s1	NCCN1CCCC1	O=C(Cc1ccccc1)Nc1nnc(N2CCC(CCOCc3cccc(C(=O)NCCN4CCCC4)c3)CC2)s1	0.56	European Journal of Medicinal Chemistry (2022), 243, 114686	0.6128	0.0528
HATU	O=C(Cc1ccccc1)Nc1nnc(N2CCN(Cc3ccc(C(=O)O)cc3)CC2)s1	NCCN1CCCC1	O=C(Cc1ccccc1)Nc1nnc(N2CCN(Cc3ccc(C(=O)NCCN4CCCCC4)cc3)CC2)s1	0.56	European Journal of Medicinal Chemistry (2022), 243, 114686	0.5968	0.0368
HATU	O=C(Cc1ccccc1)Nc1nnc(N2CCC(CCOCc3cccc(C(=O)O)c3)C	CCCCN1CCC(CN)CC1	CCCCN1CCC(CNC(=O)c2ccc(COCCC3CCN(c4nnc(NC	0.63	European Journal of Medicinal Chemistry (2022), 243,	0.6529	0.0229

	C2)s1		(=O)Cc5ccccc5)s4)CC3)c2)CC1		114686		
HATU	COc1ccc(C)c2c(=O)cc(C(=O)O)[nH]c12	COc1ccc(C)cc1N	COc1ccc(C)cc1NC(=O)c1cc(=O)c2c(C)ccc(OC)c2[nH]1	0.51	Medicinal Chemistry Research (2022), 31(3), 485-496	0.3449	0.1651
HATU	COc1ccc(C)c2c(=O)cc(C(=O)O)[nH]c12	Nc1ccccc1C(F)(F)F	COc1ccc(C)c2c(=O)cc(C(=O)Nc3ccccc3C(F)(F)F)[nH]c12	0.6	Medicinal Chemistry Research (2022), 31(3), 485-496	0.5484	0.0516
HATU	COc1ccc(C)c2c(=O)cc(C(=O)O)[nH]c12	Nc1ccc(-c2cc(F)c(F)c(F)c2)cc1	COc1ccc(C)c2c(=O)cc(C(=O)Nc3ccc(-c4cc(F)c(F)c(F)c4)cc3)[nH]c12	0.39	Medicinal Chemistry Research (2022), 31(3), 485-496	0.3493	0.0407
HATU	O=C(O)c1cn2cc(-c3ccc(-n4cccn4)nc3)ccc2n1	Nc1ccc(Cl)cc1	O=C(Nc1ccc(Cl)cc1)c1cn2cc(-c3ccc(-n4cccn4)nc3)ccc2n1	0.71	Medicinal Chemistry Research (2021), 30(1), 74-83	0.5416	0.1684
HATU	O=C(O)c1cn2cc(-c3ccc(-n4cccn4)n	N#Cc1ccc(N)cc1	N#Cc1ccc(NC(=O)c2cn3cc(-c4ccc(-	0.83	Medicinal Chemistry Research (2021),	0.7042	0.1258

	c3)ccc2n1		n5cccn5)nc4)ccc3n2)cc1		30(1), 74-83		
HATU	O=C(O)c1cn2cc(-c3ccc(-n4cccn4)nc3)ccc2n1	COc1cc(N)cc(OC)c1	COc1cc(NC(=O)c2cn3cc(-c4ccc(-n5cccn5)nc4)ccc3n2)cc(OC)c1	0.62	Medicinal Chemistry Research (2021), 30(1), 74-83	0.5383	0.0817
HATU	O=C(O)CCCCCCCC(=O)OCc1ccccc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(N)cc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(NC(=O)CCCCCCCCC(=O)OCc2ccccc2)cc1	0.65	Journal of Medicinal Chemistry (2022), 65(7), 5642-5659	0.5083	0.1417
HATU	CC(C)[C@@H]1CC[C@@H](C)C[C@H]1OCC(=O)O	CC(C)(C)OC(=O)N1CCNCC1	CC(C)[C@@H]1CC[C@@H](C)C[C@H]1OCC(=O)N1CCN(C(=O)OC(C)(C)C)CC1	0.68	Journal of Medicinal Chemistry (2022), 65(16), 11034-11057	0.5409	0.1391
HATU	O=C(O)CCCCCCCCCCCCC(=O)OCc1ccccc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(N)cc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(NC(=O)CCCC	0.61	Journal of Medicinal Chemistry (2022), 65(7), 5642-5659	0.4822	0.1278

			CCCCCCCCC(=O)OCc2ccccc2)cc1				
HATU	O=C(O)c1ccc2cc1OCCOCCNc1ccn3ncc-2c3n1	NCc1ccccc1	O=C(NCc1ccccc1)c1ccc2cc1OCCOCCNc1ccn3ncc-2c3n1	0.75	Journal of Medicinal Chemistry (2022), 65(11), 7799-7817	0.6514	0.0986
HATU	O=C(O)COCCOCCOCC(=O)OCc1ccccc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(N)cc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(NC(=O)COCCOCCOCC(=O)OCc2ccccc2)cc1	0.63	Journal of Medicinal Chemistry (2022), 65(7), 5642-5659	0.5333	0.0967
HATU	O=C(O)COCCOCCOCCOCC(=O)OCc1ccccc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(N)cc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(NC(=O)COCCOCCOCCOCC(=O)OCc2ccccc2)cc1	0.65	Journal of Medicinal Chemistry (2022), 65(7), 5642-5659	0.5659	0.0841
HATU	Cn1c(=O)[nH]c(Cl)c(N)c1=O	Cc1ccc(CCC(=O)O)cc1	Cc1ccc(CCC(=O)Nc2c(Cl)[nH]c(=O)n(	0.71	Journal of Medicinal Chemistry (2022),	0.6452	0.0648

			C)c2=O)cc1		65(19), 12747-12780		
HATU	O=C(O)c1cccc(Br)c1	CC1CCNCC1	CC1CCN(C(=O)c2cccc(Br)c2)CC1	0.82	Journal of Medicinal Chemistry (2022), 65(4), 3266-3305	0.7561	0.0639
HATU	CC(C)CC(NC(=O)OC(C)(C)C)C(=O)O	C=Cc1cc(-c2cnco2)c(OC)cc1N	C=Cc1cc(-c2cnco2)c(OC)cc1NC(=O)C(CC(C)C)NC(=O)OC(C)(C)C	0.34	Journal of Medicinal Chemistry (2022), 65(5), 4121-4155	0.2872	0.0528
HATU	O=C(O)C1CN(C2=CC=C(I)C=C2)C1	Cl.O1CCC(N)CC1	O=C(NC1CCOCC1)C2CN(C3=CC=C(I)C=C3)C2	0.61	Bioorganic & Medicinal Chemistry (2022), 64, 116763	0.6378	0.0278
HATU	C(O)(=O)[C@@H]1[C@H](CC(CCC=C(C)C)=CC1)C2=CC=CC=C2	O(C1=CC=C(N)C=C1)C	C(NC1=CC=C(OC)C=C1)(=O)[C@@H]2[C@H](CC(CCC=C(C)C)=CC2)C3=CC=CC=C3	0.61	Bioorganic & Medicinal Chemistry Letters (2020), 30(7), 127003	0.6684	0.0584
HATU	Cc1cc(NS(=O)(=O)c	CC1CCNCC1	Cc1cc(NS(=O)(=O)c	0.62	Journal of Medicinal	0.5753	0.0447

	2cc(-c3nc(C)c(C(=O)O)s3)n(C)c2C)no1		2cc(-c3nc(C)c(C(=O)N4CCC(C)CC4)s3)n(C)c2C)no1		Chemistry (2022), 65(4), 3266-3305		
HATU	O=C(O)c1cc2ccccc2[nH]1	NCCO	O=C(NCCO)c1cc2ccccc2[nH]1	0.7	Journal of Medicinal Chemistry (2022), 65(13), 9376-9395	0.6565	0.0435
HATU	O=C(O)COCCOCC(=O)OCc1ccccc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(N)cc1	CC(C)(C)OC(=O)Nc1ccccc1NC(=O)c1ccc(NC(=O)COCCOCC(=O)OCc2ccccc2)cc1	0.58	Journal of Medicinal Chemistry (2022), 65(7), 5642-5659	0.5438	0.0362
HATU	O=C(O)c1ccnnc1	Cc1cncc(C2=CC[C@H]3[C@@H]4CC=C5C[C@@H](N)CC[C@]5(C)[C@H]4CC[C@]23C)c1	Cc1cncc(C2=CC[C@H]3[C@@H]4CC=C5C[C@@H](NC(=O)c6ccnnc6)CC[C@]5(C)[C@H]4CC[C@]23C)c1	0.71	Journal of Medicinal Chemistry (2022), 65(18), 12460-12481	0.6853	0.0247
HATU	O=C(O)C	CC(C)(C)	CC(C)(C)	0.55	Journal of	0.534	0.016

	OCCCCCCOCC(=O)OCc1ccccc1	OC(=O)Nc1ccccc1NC(=O)c1ccc(N)cc1	OC(=O)Nc1ccccc1NC(=O)c1ccc(NC(=O)COCCCCCCOCC(=O)OCc2ccccc2)cc1		Medicinal Chemistry (2022), 65(7), 5642-5659		
HATU	O=C(O)C1CCN(c2cncnc2-c2ccccc2)CC1	Nc1ccccc1	O=C(Nc1ccccc1)C1CCN(c2cncnc2-c2ccccc2)CC1	0.7	Journal of Medicinal Chemistry (2022), 65(4), 3343-3358	0.6925	0.0075
HATU	O=C(O)c1cnn2cccnc12	COC(=O)c1cc(N)cc(C(=O)OC)c1	COC(=O)c1cc(NC(=O)c2cnn3cccnc23)cc(C(=O)OC)c1	0.17	Journal of Medicinal Chemistry (2022), 65(4), 3518-3538	0.1659	0.0041
HATU	COc1ccc(-c2oc3ccc(O)cc3c2C(=O)O)cc1	C1CNC1	COc1ccc(-c2oc3ccc(O)cc3c2C(=O)N2CCC2)cc1	0.72	Journal of Medicinal Chemistry (2022), 65(1), 409-423	0.7165	0.0035
HATU	N[C@H]1CC[C@H](O)CC1	CNC(=O)c1cc(C(=O)O)c(C(C)c2cccc(Cl)c2)o1	CNC(=O)c1cc(C(=O)N[C@H]2CC[C@H](O)CC2)c(C(C)c2cccc(Cl)	0.75	Journal of Medicinal Chemistry (2021), 64(15), 10772-10805	0.5752	0.1748

			c2)o1				
HATU	CCCn1nc c2c(C(=O) O)cc(C3C C3)nc21	CN(C)CC N	CCCn1nc c2c(C(=O) NCCN(C) C)cc(C3C C3)nc21	0.83	Journal of Medicinal Chemistry (2021), 64(12), 8755-8774	0.6717	0.158 3
HATU	O=C(O)c1 cccc([N+] (=O)[O-])c1	COc1ccc(N)cn1	COc1ccc(NC(=O)c2 cccc([N+] (=O)[O-])c2)cn1	0.93	Journal of Medicinal Chemistry (2021), 64(16), 12003-12021	0.7749	0.155 1
HATU	CNC(=O) c1cc(C(= O)O)c(C(C)c2ccccc 2)o1	NC1CC1	CNC(=O) c1cc(C(= O)NC2CC 2)c(C(C)c 2ccccc2)o 1	0.69	Journal of Medicinal Chemistry (2021), 64(15), 10772-10805	0.535	0.155
HATU	CNC(=O) c1cc(C(= O)O)c([C @@H](C) c2ccccc2) o1	N[C@H]1 CC[C@H] (O)CC1	CNC(=O) c1cc(C(= O)N[C@ H]2CC[C @H](O)C C2)c([C@ @H](C)c2 ccccc2)o1	0.75	Journal of Medicinal Chemistry (2021), 64(15), 10772-10805	0.5959	0.154 1
HATU	COCOc1c cc(C(=O) O)cc1-	CCN	CCNC(=O)c1ccc(OC OC)c(-	0.66	Journal of Medicinal Chemistry	0.5148	0.145 2

	c1cn(C)nn1		c2cn(C)nn2)c1		(2021), 64(6), 3249-3281		
HATU	C[C@@H](Nc1nc2c(cnn2C2CCCC2)c(=O)[nH]1)C(=O)O	C1CN(C2CC2)CCN1	C[C@@H](Nc1nc2c(cnn2C2CCCC2)c(=O)[nH]1)C(=O)N1CCN(C2CC2)CC1	0.65	Journal of Medicinal Chemistry (2021), 64(13), 9537-9549	0.5088	0.1412
HATU	O=C(O)C=CCC	NCC1CCCCC1	O=C(C=CCC)NCC1CCCCC1	0.89	World Intellectual Property Organization, WO2018050631 A1 2018-03-22	0.7583	0.1317
HATU	O=C(O)C=CC=CCCCCCC	NCC=1C=CC=CC1	O=C(C=CC=CCCCCCCC)NCC=1C=CC=CC1	0.62	Proceedings of the National Academy of Sciences of the United States of America (2017), 114(25), E5006-E5015	0.8017	0.1817

HATU	C[C@H]1C[C@@H]1N	CNC(=O)c1cc(C(=O)O)cc2c1OC[C@@]2(C)c1ccccc1	CNC(=O)c1cc(C(=O)N[C@H]2C[C@@H]2C)cc2c1OC[C@@]2(C)c1ccccc1	0.6	Journal of Medicinal Chemistry (2021), 64(15), 10711-10741	0.4669	0.1331
HATU	O=C(O)c1ccc2c(c1)nc(Nc1cccc(Cl)c1)c1ccncc12	NCCN1CCCCC1	O=C(NCCN1CCCCC1)c1ccc2c(c1)nc(Nc1cccc(Cl)c1)c1ccncc12	0.771	Journal of Medicinal Chemistry (2021), 64(8), 5082-5098	0.6388	0.1322
HATU	CNC(=O)c1cc(C(=O)O)c([C@@H](C)c2ccccc2)o1	NCCO	CNC(=O)c1cc(C(=O)NCCO)c([C@@H](C)c2ccccc2)o1	0.64	Journal of Medicinal Chemistry (2021), 64(15), 10772-10805	0.5105	0.1295
HATU	Cc1ccc(C(=O)O)cc1C#Cc1cc(-c2cnn(C)c2)cnc1N	COc1ccc(N)cc1OC	COc1ccc(NC(=O)c2ccc(C)c(C#Cc3cc(-c4cnn(C)c4)cnc3N)c2)cc1OC	0.89	Journal of Medicinal Chemistry (2021), 64(18), 13588-13603	0.766	0.124
HATU	O=C(O)c1ccc2c(c1)nc(Nc1cccc(Cl)c1)c1ccncc12	NCCN1CCOCC1	O=C(NCCN1CCOCC1)c1ccc2c(c1)nc(Nc1cccc(Cl)	0.687	Journal of Medicinal Chemistry (2021), 64(8),	0.5736	0.1134

			c1)c1ccnc c12		5082-5098		
HATU	CNC(=O) c1cc(C(= O)O)cc(C c2cccc3c2 CCN3C(= O)OC(C)(C)C)n1	NC1CC1	CNC(=O) c1cc(C(= O)NC2CC 2)cc(Cc2c ccc3c2CC N3C(=O) OC(C)(C) C)n1	0.74	Journal of Medicinal Chemistry (2021), 64(15), 10742-10771	0.6274	0.112 6
HATU	Cn1ncc2c(C(=O)O)c c(C3CC3) nc21	CN(C)CC N	CN(C)CC NC(=O)c1 cc(C2CC2)nc2c1cnn 2C	0.76	Journal of Medicinal Chemistry (2021), 64(12), 8755-8774	0.6477	0.112 3
HATU	O=C1Cc2 cc(C(=O) O)ccc2N1	Cc1ccc([C @@H](C) N)cc1	Cc1ccc([C @@H](C) NC(=O)c2 ccc3c(c2) CC(=O)N 3)cc1	0.69	Journal of Medicinal Chemistry (2021), 64(1), 566-585	0.5854	0.104 6
HATU	CCOc1ccc (C(=O)O) nc1	Nc1ccccc 1	CCOc1ccc (C(=O)Nc 2ccccc2)n c1	0.77	Journal of Medicinal Chemistry (2021), 64(24), 17936-17949	0.6882	0.081 8
HATU	O=C(O)c1 ccc2c(c1) nc(Nc1ccc c(Cl)c1)c1 ccncc12	NCCN1C CCC1	O=C(NCC N1CCCC 1)c1ccc2c (c1)nc(Nc 1cccc(Cl)c	0.728	Journal of Medicinal Chemistry (2021), 64(8),	0.649	0.079

			1)c1ccncc12		5082-5098		
HATU	Cc1cc(C(=O)O)c2cnn(C(C)C)c2n1	CN(C)CCN	Cc1cc(C(=O)NCCN(C)C)c2cnn(C(C)C)c2n1	0.77	Journal of Medicinal Chemistry (2021), 64(12), 8755-8774	0.6913	0.0787
HATU	CNC(=O)c1cc(C(=O)O)c(C(C)c2cccc(F)c2)o1	N[C@H]1CC[C@H](O)CC1	CNC(=O)c1cc(C(=O)N[C@H]2CC[C@H](O)CC2)c(C(C)c2cccc(F)c2)o1	0.65	Journal of Medicinal Chemistry (2021), 64(15), 10772-10805	0.5759	0.0741
HATU	CCN(c1ccc(CC(=O)O)cc1)S(=O)(=O)c1cc(Cl)ccc1Cl	CN1CCNCC1	CCN(c1ccc(CC(=O)N2CCN(C)CC2)cc1)S(=O)(=O)c1cc(Cl)ccc1Cl	0.57	Journal of Medicinal Chemistry (2021), 64(15), 10951-10966	0.5036	0.0664
HATU	C=CCCC(=O)O	C=CCCOc1cccc(CC(=O)Nc2nnc(CCCCc3nnc(N)s3)s2)c1	C=CCCOc1cccc(CC(=O)Nc2nnc(CCCCc3nnc(NC(=O)CCC=C)s3)s2)c1	0.552	Journal of Medicinal Chemistry (2021), 64(8), 4588-4611	0.4868	0.0652
HATU	C=CCCOc1cccc(CC(=O)O)c	CCOC(=O)CCCCc1nnc(N)s1	C=CCCOc1cccc(CC(=O)Nc2	0.652	Journal of Medicinal Chemistry	0.5907	0.0613

	1		nnc(CCCCC(=O)OCC)s2)c1		(2021), 64(8), 4588-4611		
HATU	O=C(O)C1C2CCC#CCCC12	NCC=1C=CC=CC1	O=C(NCC=1C=CC=CC1)C2C3CCC#CCCC23	0.83	RSC Advances (2021), 11(58), 36777-36780	0.7326	0.0974
HATU	O=C(O)C(=O)C=CC=1C=CC=CC1	OCCN	O=C(C=CC=1C=CC=CC1)C(=O)NCCO	0.45	Bioorganic & Medicinal Chemistry (2011), 19(13), 4067-4074	0.5824	0.1324
HATU	O=C(O)C(=O)C=CC=1C=CN=CC1	NC(C)(C)C	O=C(C=CC=1C=CN=CC1)C(=O)NC(C)(C)C	0.39	Bioorganic & Medicinal Chemistry (2011), 19(13), 4067-4074	0.4921	0.1021
HATU	NC	O=C(O)C1=CN(C=2SC(Br)=CC2)C=3C=CC(O)=CC13	O=C(NC)C1=CN(C=2SC(Br)=CC2)C=3C=CC(O)=CC13	0.7	European Journal of Medicinal Chemistry (2025), 283, 117148	0.65	0.05
HATU	NC1=CC=C(C=C1)CN2CCCCC2	O=C(O)C=1C=CC=C(C#CC2=CN=C(N	O=C(NC1=CC=C(C=C1)CN2CCCCC2)	0.65	European Journal of Medicinal Chemistry	0.6	0.05

		=C2)NC=3C(=CC=CC3C)N(=O)=O)C1	C=3C=CC=C(C#CC4=CN=C(N=C4)NC=5C(=CC=CC5C)N(=O)=O)C3		(2025), 284, 117206		
HATU	N=1C=CC=C(N)C1	O=C(O)C1=CNC=2N=CC(Br)=CC21	O=C(NC=1C=NC=CC1)C2=CNC=3N=CC(Br)=CC32	0.51	European Journal of Medicinal Chemistry (2025), 285, 117236	0.43	0.08
HATU	ClC=1N=CC=2SC=C(C2N1)C=3C=CC=C(N)C3	[C@H](C(O)=O)(C)N1CCN(C(OC(C)(C)C)=O)CC1	ClC=1N=C2C(=CSC2=CN1)C3=CC(NC([C@@H](C)N4CCN(C(OC(C)(C)C)=O)CC4)=O)=CC=C3	0.64	European Journal of Medicinal Chemistry (2025), 286, 117308	0.52	0.12
HATU	OC1=CC=C(N)C=C1	O=C(O)CN1CCN(C=2N=C3C(OCC(=O)NCC=4C=CC=CC4)=CC=CC3=CC2)CC1	O=C(NCC=1C=CC=CC1)COC2=CC=CC3=CC=C(N=C23)N4CCN(CC(=O)NC5=CC=C(O)C=C5)CC4	0.81	Bioorganic & Medicinal Chemistry Letters (2025), 118, 130081	0.83	0.02
EDC	O=C(O)C	Cl.NC	O=C(NC)	0.39	Journal of	0.3703	0.019

	1=NC=CC(Cl)=C1		C1=NC=CC(Cl)=C1		Medicinal Chemistry (2021), 64(21), 15651-15670		7
EDC	O=C(OC(C)(C)C)NCC(=O)O	C([C@H](C(N)=O)N)C1=CC=CC=C1.Cl	C([C@@H](NC(CNC(OC(C)(C)C)=O)=O)C(N)=O)C1=CC=CC=C1	0.63	Bioorganic & Medicinal Chemistry Letters (2020), 30(14), 127117	0.4503	0.1797
EDC	COc1cc(C=C(C(=O)O)c2ccc(OC)c(OC)c2)ccc1O	CCN(CC)c1ccc2cc(C(=O)NCCN)c(=O)oc2c1	CCN(CC)c1ccc2cc(C(=O)NCCNC(=O)C(=Cc3ccc(O)c(OC)c3)c3ccc(OC)c(OC)c3)c(=O)oc2c1	0.379	Bioorganic Chemistry (2022), 127, 106037	0.4285	0.0495
EDC	COc1ccc(C(=Cc2ccc(O)c(OC)c2)C(=O)O)cc1	CCN(CC)c1ccc2cc(C(=O)NCCN)c(=O)oc2c1	CCN(CC)c1ccc2cc(C(=O)NCCNC(=O)C(=Cc3ccc(O)c(OC)c3)c3ccc(OC)cc3)c(=O)oc2c1	0.343	Bioorganic Chemistry (2022), 127, 106037	0.4203	0.0773
EDC	Cc1ccc(C	CCN(CC)	CCN(CC)	0.396	Bioorgani	0.4643	0.068

	=C(C#N)C(=O)O)s1	c1ccc2cc(C(=O)NCCN)c(=O)oc2c1	c1ccc2cc(C(=O)NCCNC(=O)C(C#N)=Cc3ccc(C)s3)c(=O)oc2c1		c Chemistry (2022), 127, 106037		3
EDC	COc1cc2cc(C(=O)CCC(=O)O)sc2cc1OC	CNOC	COc1cc2cc(C(=O)CCC(=O)N(C)OC)sc2cc1OC	0.51	European Journal of Medicinal Chemistry (2022), 241, 114627	0.5197	0.0097
EDC	NS(=O)(=O)c1ccc(-c2cccc(C(=O)O)c2)cc1	CC1CCNCC1	CC1CCN(C(=O)c2cccc(-c3ccc(S(N)(=O)=O)cc3)c2)CC1	0.32	Journal of Medicinal Chemistry (2022), 65(4), 3266-3305	0.3884	0.0684
EDC	Cc1ccc(CC(=O)O)cc1	Cc1ccc2oc(-c3cccc(N)c3)nc2c1	Cc1ccc(CC(=O)Nc2cccc(-c3nc4cc(C)ccc4o3)c2)cc1	0.54	European Journal of Medicinal Chemistry (2022), 227, 113933	0.5634	0.0234
EDC	O=C(O)CCCCCC=1C=CC=CC1	N[C@H]1CC[C@H](O)CC1	N(C(CCCCCC1=CC=CC=C1)=O)[C@@H]2CC[C@@H](	0.78	Journal of Medicinal Chemistry (2020), 63(13), 7033-7051	0.6705	0.1095

			O)CC2				
EDC	C(=C/C(O)=O)\C1=CC=C(O)C=C1	N=1C(N)=CC=CC1CN	C(NC(/C=C/C1=CC=C(O)C=C1)=O)C=2N=C(N)C=CC2	0.9	Drug Development Research (2020), 81(2), 206-214	0.7651	0.1349
EDC	Cc1ccc(CC(=O)O)cc1	Nc1ccc(-c2cc3ccccc3o2)cc1	Cc1ccc(CC(=O)Nc2ccc(-c3cc4ccccc4o3)cc2)cc1	0.55	European Journal of Medicinal Chemistry (2022), 227, 113933	0.5589	0.0089
EDC	O=C(CCCC[C@@H]1SC[C@@H]2NC(=O)N[C@@H]21)NCCCC[C@H](NC(=O)OCC1c2ccccc2-c2ccccc21)C(=O)O	COc1ccc(CC(CCN)c2cc(OC)c(OC)c(OC)c2)cc1O	COc1ccc(CC(CCNC(=O)[C@H](CCCCNC(=O)CCCC[C@@H]2SC[C@@H]3NC(=O)N[C@@H]32)NC(=O)OCC2c3ccccc3-c3ccccc32)c2cc(OC)c(OC)c(OC)c2)cc1O	0.55	Journal of Medicinal Chemistry (2022), 65(1), 460-484	0.5623	0.0123
EDC	O=C(CCC	COc1ccc(	COc1ccc(	0.55	Journal of	0.5593	0.009

	C[C@@H]1SC[C@@H]2NC(=O)N[C@@H]21)NCCCC[C@H](NC(=O)OCC1c2ccccc2-c2ccccc21)C(=O)O	CCc2cc(OC)c(OC)c(OC)c2OCCN)cc1O	CCc2cc(OC)c(OC)c(OC)c2OCCNC(=O)[C@H](CCCCNC(=O)CCCC[C@@H]2SC[C@@H]3NC(=O)N[C@@H]32)NC(=O)OCC2c3ccccc3-c3ccccc32)cc1O		Medicinal Chemistry (2022), 65(1), 460-484		3
EDC	O=C(O)CC(CC(=O)NOC1CCCCO1)c1ccc(Cl)cc1Cl	Cc1ccccc1CN	Cc1ccccc1CNC(=O)CC(CC(=O)NOC1CCCCO1)c1ccc(Cl)cc1Cl	0.27	ACS Medicinal Chemistry Letters (2021), 12(8), 1318-1324	0.3449	0.0749
EDC	O=C(O)C1=CC=C(Br)C=C1	[C@@H](C(OC)=O)(CO)N.Cl	C(N[C@H](C(OC)=O)CO)(=O)C1=CC=C(Br)C=C1.Cl	0.74	European Journal of Medicinal Chemistry (2022), 233, 114195	0.4967	0.2433
EDC	O=C(O)C1=CC=C(C#C)C=C1	[C@@H](C(OC)=O)([C@@H](C)O)N.Cl	C(N[C@H](C(OC)=O)[C@@H](C)O)(=O)C1=CC	0.55	ACS Infectious Diseases (2021), 7(9),	0.4481	0.1019

			=C(C#C)C=C1		2755-2763		
EDC	CC(C)(C)OC(=O)N1CCC[C@@H](Nc2ccn3ncc(C(=O)O)c3n2)C1	COc1cc(N(C)CCN(C)C)ccc1N	COc1cc(N(C)CCN(C)C)ccc1NC(=O)c1cnn2ccc(N[C@@H]3CCCN(C(=O)OC(C)(C)C)C3)nc12	0.48	Bioorganic and Medicinal Chemistry (2021), 48, 116422	0.5302	0.0502
EDC	O=C(O)CC(CC(=O)NOC1CCCCO1)c1ccc(Cl)cc1Cl	Cc1ccc(CN)cc1	Cc1ccc(CNC(=O)CC(CC(=O)NOC2CCCCO2)c2ccc(Cl)cc2Cl)cc1	0.48	ACS Medicinal Chemistry Letters (2021), 12(8), 1318-1324	0.5455	0.0655
EDC	O=C(O)Cc1ccc2ccccc2c1	Cn1cccc(N)c1=O	Cn1cccc(NC(=O)Cc2ccc3ccccc3c2)c1=O	0.55	Organic and Biomolecular Chemistry (2021), 19(28), 6244-6249	0.581	0.031
EDC	O=C(O)COc1ccc(C(=O)c2cc(Cl)ccc2Br)cc1	Nc1cc(Cl)cc(C(=O)NCC2CC2)c1	O=C(COc1ccc(C(=O)c2cc(Cl)ccc2Br)cc1)Nc1cc(Cl)cc(C(=O)NCC2C	0.11	European Journal of Medicinal Chemistry (2021), 212, 113033	0.1553	0.0453

			C2)c1				
EDC	O=C(O)CC(CC(=O)NOC1CCCCO1)c1ccc(Cl)cc1Cl	Cc1cccc(CN)c1	Cc1cccc(CNC(=O)CC(CC(=O)NOC2CCCCO2)c2ccc(Cl)cc2Cl)c1	0.51	ACS Medicinal Chemistry Letters (2021), 12(8), 1318-1324	0.561	0.051
EDC	CCn1c(=O)c(C(=O)O)c(O)c2ccccc21	CCCCc1nnc(N)s1	CCCCc1nnc(NC(=O)c2c(O)c3ccccc3n(CC)c2=O)s1	0.185	European Journal of Medicinal Chemistry (2020), 188, 112022	0.2197	0.0347
EDC	O=C(O)CCCCCc1ccccc1	N[C@H]1CC[C@@H](O)CC1	O=C(CCCCCc1ccccc1)N[C@H]1CC[C@@H](O)CC1	0.674	Journal of Medicinal Chemistry (2020), 63(13), 7033-7051	0.6899	0.0159
EDC	CCn1c(=O)c(C(=O)O)c(O)c2ccccc21	CCCc1nnc(N)s1	CCCc1nnc(NC(=O)c2c(O)c3ccccc3n(CC)c2=O)s1	0.209	European Journal of Medicinal Chemistry (2020), 188, 112022	0.2528	0.0438
EDC	CCCn1c(=O)c(C(=O)O)c(O)c2ccccc21	Nc1nncs1	CCCn1c(=O)c(C(=O)Nc2nncs2)c(O)c2ccccc21	0.171	European Journal of Medicinal Chemistry (2020), 188,	0.257	0.086

					112022		
EDC	O=C(O)C1(C(=O)Nc2cccc(Br)c2)CC1	Nc1ccc(N)cc1	Nc1ccc(NC(=O)C2(C(=O)Nc3cccc(Br)c3)CC2)cc1	0.58	European Journal of Medicinal Chemistry (2019), 181, 111541	0.5819	0.0019
EDC	COc1ccc(/C=C/C(=O)O)cc1	CCN(CC)c1ccc2cc(C(=O)N3CCNCC3)c(=O)oc2c1	CCN(CC)c1ccc2cc(C(=O)N3CCN(C(=O)/C=C/c4ccc(OC)cc4)CC3)c(=O)oc2c1	0.28	European Journal of Medicinal Chemistry (2019), 170, 45-54	0.3554	0.0754
EDC	COc1cc(/C=C/C(=O)O)cc(OC)c1OC	CCN(CC)c1ccc2cc(C(=O)NCCN)c(=O)oc2c1	CCN(CC)c1ccc2cc(C(=O)NCCNC(=O)/C=C/c3cc(OC)c(OC)c(OC)c3)c(=O)oc2c1	0.432	European Journal of Medicinal Chemistry (2019), 170, 45-54	0.436	0.004
EDC	O=C(O)CCSSC(C)(C)C	OCC(N)C(O)C	O=C(NC(CO)C(O)C)CCSSC(C)(C)C	0.8	Molecules (2012), 17, 10026-10045	0.6062	0.1938
EDC	O=C(O)C1=CC=C(SSC2=CC=C(C=C2)C(=O)O)C=C1	OCCN	O=C(NCCO)C1=CC=C(SSC2=CC=C(C=C2)C(=O)NCCO)C	0.65	ChemBioChem (2020), 21(5), 656-662	0.4768	0.1732

			=C1				
EDC	CCCCCn1c(C(=O)O)cc2cccnc21	NC12CC3CC(CC(C3)C1)C2	CCCCCn1c(C(=O)NC23CC4CC(CC(C4)C2)C3)cc2cccnc21	0.59	European Journal of Medicinal Chemistry (2019), 180, 291-309	0.5973	0.0073
EDC	CCCCCn1nc(C(=O)O)c2cccnc21	NC12CC3CC(CC(C3)C1)C2	CCCCCn1nc(C(=O)NC23CC4CC(CC(C4)C2)C3)c2cccnc21	0.52	European Journal of Medicinal Chemistry (2019), 180, 291-309	0.5993	0.0793
EDC	O=C(O)c1ccc(OC2CCCCO2)cc1	Clc1cccc(CNCCc2cc(OC3CCCCO3)cc(OC3CCCCO3)c2)c1	O=C(c1ccc(OC2CCCCO2)cc1)N(CCc1cc(OC2CCCCO2)cc(OC2CCCCO2)c1)Cc1cccc(Cl)c1	0.6	ACS Medicinal Chemistry Letters (2019), 10(4), 615-620	0.6372	0.0372
EDC	C=Cc1c(OC)cc(/C=C/c2ccc(C(=O)O)cc2)cc1OC	NCCc1ccc(O)cc1	C=Cc1c(OC)cc(/C=C/c2ccc(C(=O)NCCc3ccc(O)cc3)cc2)cc1OC	0.72	European Journal of Medicinal Chemistry (2019), 184, 111733	0.755	0.035
EDC	CC(=O)N	Nc1nc2c(s	CC(=O)N	0.26	European	0.3573	0.097

	1CCCC1c1cc(C(=O)O)c(C)s1	1)CCC2	1CCCC1c1cc(C(=O)Nc2nc3c(s2)CCC3)c(C)s1		Journal of Medicinal Chemistry (2018), 156, 269-294		3
EDC	Cc1ccc(C(=O)O)cc1	NC1=NC2(CCNCC2)N(c2cccc(Cl)c2)C(N)=N1	Cc1ccc(C(=O)N2CCC3(CC2)N=C(N)N=C(N)N3c2cccc(Cl)c2)cc1	0.33	European Journal of Medicinal Chemistry (2018), 155, 229-243	0.4094	0.0794
EDC	CCCCCC(=O)O	Nc1ccc(N)cc1	CCCCCC(=O)Nc1ccc(N)cc1	0.6	Journal of Medicinal Chemistry (2018), 61(7), 3166-3192	0.6011	0.0011
EDC	Cc1ccc(-c2cc(-c3ccccc3)nc(SCC(=O)O)c2C#N)cc1	COC(=O)C[C@H](NC(=O)[C@@H](N)C(C)C)C(=O)N[C@@H](CCSC)C(=O)OC	COC(=O)C[C@H](NC(=O)[C@@H](NC(=O)CSc1nc(-c2ccccc2)cc(-c2ccc(C)cc2)c1C#N)C(C)C)C(=O)N[C@@H](CCSC)C(=O)OC	0.41	European Journal of Medicinal Chemistry (2018), 157, 743-758	0.5013	0.0913

EDC	O=C(O)C1=CC=C(C=C1)C	NC	O=C(NC)C1=CC=C(C=C1)C	0.41	European Journal of Medicinal Chemistry (2018), 143, 390-401	0.3789	0.0311
EDC	O=C(OC(C)(C)C)NC(C(=O)O)CSC(C=1C=CC=CC1)(C=2C=CC=CC2)C=3C=CC=CC3	NC	O=C(NC)C(N)CS	0.74	Journal of the American Chemical Society (2013), 135(15), 5839-5847	0.5019	0.2381
EDC	O=C(O)CCC=C	C=CCN	O=C(NCC=C)CCC=C	0.78	Chemistry Letters (2020), 49(1), 71-74	0.5933	0.1867
EDC	O=C(O)C=1C=CN=C(N)C1	Cl.NC	O=C(NC)C=1C=CN=C(N)C1	0.64	Journal of Medicinal Chemistry (2022), 65(21), 14366-14390	0.44	0.2
EDC	O=C(O)C1(C(=O)Nc2ccc(Cl)c(C(F)(F)F)c2)CC1	Nc1ccc(N)cc1	Nc1ccc(NC(=O)C2(C(=O)Nc3ccc(Cl)c(C(F)(F)F)c3)CC2)cc	0.29	Bioorganic and Medicinal Chemistry Letters (2017),	0.374	0.084

			1		27(15), 3231-3237		
EDC	O=C(O)C1(C(=O)Nc2cccc(F)c2)CC1	Nc1ccc(N)cc1	Nc1ccc(NC(=O)C2(C(=O)Nc3cccc(F)c3)CC2)cc1	0.39	Bioorganic and Medicinal Chemistry Letters (2017), 27(15), 3231-3237	0.4651	0.0751
EDC	O=C(O)C1(C(=O)Nc2cccc(O)c2)CC1	Nc1ccc(N)cc1	Nc1ccc(NC(=O)C2(C(=O)Nc3cccc(O)c3)CC2)cc1	0.51	Bioorganic and Medicinal Chemistry Letters (2017), 27(15), 3231-3237	0.5235	0.0135
EDC	O=C(O)C1(C(=O)Nc2ccccc2F)CC1	Nc1ccc(N)cc1	Nc1ccc(NC(=O)C2(C(=O)Nc3ccccc3F)CC2)cc1	0.54	Bioorganic and Medicinal Chemistry Letters (2017), 27(15), 3231-3237	0.5498	0.0098
EDC	COc1cc(C(=O)O)ccc1-c1c[nH]c2ccccc12	Nc1ccc(F)cc1	COc1cc(C(=O)Nc2ccc(F)cc2)ccc1-c1c[nH]c2ccccc12	0.377	European Journal of Medicinal Chemistry (2017), 139, 644-656	0.4601	0.0831
EDC	COc1cc(C	CN1CCC	COc1cc(C	0.596	European	0.5971	0.001

	(=O)O)ccc1-c1c[nH]c2ccccc12	NCC1	(=O)N2CCCN(C)CC2)ccc1-c1c[nH]c2ccccc12		Journal of Medicinal Chemistry (2017), 139, 644-656		1
EDC	O=C(O)c1cc(Cl)c(O)cc1O	Nc1ccccc1	O=C(Nc1ccccc1)c1cc(Cl)c(O)cc1O	0.37	European Journal of Medicinal Chemistry (2016), 124, 1069-1080	0.3872	0.0172
EDC	Cc1ccc(-c2cc(C(=O)O)c3cc(Cl)ccc3n2)cc1	NCCN1CCOCC1	Cc1ccc(-c2cc(C(=O)NCCN3CCOCC3)c3cc(Cl)ccc3n2)cc1	0.22	Journal of Medicinal Chemistry (2016), 59(21), 9672-9685	0.2848	0.0648
EDC	COc1cc(OC)c(C=C2SC(=O)N(CC(=O)O)C2=O)c(OC)c1	C1CCC(N2CCNCC2)CC1	COc1cc(OC)c(C=C2SC(=O)N(CC(=O)N3CCN(C4CCCCC4)CC3)C2=O)c(OC)c1	0.44	European Journal of Medicinal Chemistry (2015), 90, 507-518	0.4841	0.0441
EDC	Cc1nc(-n2cnn(Cc3ccc(F)cc3)c2=O)sc1C(=O)O	Cc1ncc(CN)s1	Cc1ncc(CNC(=O)c2sc(-n3cnn(Cc4ccc(F)cc4)c3=O)nc	0.68	Bioorganic and Medicinal Chemistry (2015), 23(3),	0.7196	0.0396

			2C)s1		455-465		
EDC	Cc1nc(-n2cnn(Cc3ccc(F)cc3)c2=O)sc1C(=O)O	Cn1cnc(CN)c1	Cc1nc(-n2cnn(Cc3ccc(F)cc3)c2=O)sc1C(=O)NCc1cn(C)cn1	0.57	Bioorganic and Medicinal Chemistry (2015), 23(3), 455-465	0.5857	0.0157
EDC	Cc1nc(-n2cnn(Cc3ccc(F)cc3)c2=O)sc1C(=O)O	NCc1ccccn1	Cc1nc(-n2cnn(Cc3ccc(F)cc3)c2=O)sc1C(=O)NCc1ccccn1	0.46	Bioorganic and Medicinal Chemistry (2015), 23(3), 455-465	0.501	0.041
EDC	Cc1nc(-n2cnn(Cc3ccc(F)cc3)c2=O)sc1C(=O)O	NCc1nccs1	Cc1nc(-n2cnn(Cc3ccc(F)cc3)c2=O)sc1C(=O)NCc1nccs1	0.66	Bioorganic and Medicinal Chemistry (2015), 23(3), 455-465	0.6694	0.0094
EDC	COc1cc(OC)c(C=C2SC(=O)N(CC(=O)O)C2=O)c(OC)c1	C1CNCCN1	COc1cc(OC)c(C=C2SC(=O)N(CC(=O)N3CCNCC3)C2=O)c(OC)c1	0.42	European Journal of Medicinal Chemistry (2015), 90, 507-518	0.444	0.024
EDC	O=C(O)c1ccc2[nH]ccc2c1	c1ccc2c(c1)-c1ccccc1C2N1CCN	O=C(c1ccc2[nH]ccc2c1)N1CCN(C2c3cc	0.49	European Journal of Medicinal Chemistry	0.5195	0.0295

		CC1	ccc3-c3ccccc32)CC1		(2015), 101, 218-235		
EDC	O=C(O)C1=CNC2C=CC=CC12	COc1cc2c(cc1OC)CN(CC1CCNCC1)CC2	COc1cc2c(cc1OC)CN(CC1CCN(C(=O)C3=CNC4C=CC=CC34)CC1)CC2	0.39	Organic and Biomolecular Chemistry (2014), 12(5), 783-794	0.4199	0.0299
EDC	O=C(O)c1cc2cc(I)ccc2o1	COc1cc2c(cc1OC)CN(CC1CCNCC1)CC2	COc1cc2c(cc1OC)CN(CC1CCN(C(=O)c3cc4cc(I)ccc4o3)CC1)CC2	0.42	Organic and Biomolecular Chemistry (2014), 12(5), 783-794	0.49	0.07
EDC	O=C(O)c1cc2cc(Br)ccc2o1	COc1cc2c(cc1OC)CN(CCC1CCNCC1)CC2	COc1cc2c(cc1OC)CN(CCC1CCN(C(=O)c3cc4cc(Br)ccc4o3)CC1)CC2	0.51	Organic and Biomolecular Chemistry (2014), 12(5), 783-794	0.5339	0.0239
EDC	O=C(O)c1cc2cc(I)ccc2o1	COc1cc2c(cc1OC)CN(CCC1CCNCC1)CC2	COc1cc2c(cc1OC)CN(CCC1CCN(C(=O)c3cc4cc(I)ccc4o3)CC1)CC2	0.5	Organic and Biomolecular Chemistry (2014), 12(5),	0.535	0.035

					783-794		
EDC	C[C@H](c1cccc2ccccc12)N1CCC(C(=O)O)CC1	CC(=O)Nc1ccc(CN)cc1	CC(=O)Nc1ccc(CNC(=O)C2CCN([C@H](C)c3cccc4ccccc34)CC2)cc1	0.23	Journal of Medicinal Chemistry (2014), 57(6), 2393-2412	0.2823	0.0523
EDC	O=C(O)/C=C/c1cnc2c(c1)CCC(=O)N2	c1ccc(C2CNC2)cc1	O=C1CCc2cc(/C=C/C(=O)N3CC(c4ccccc4)C3)cnc2N1	0.5	European Journal of Medicinal Chemistry (2014), 84, 382-394	0.5948	0.0948
EDC	O=C(O)CCC(F)(F)F	c1csc(-c2noc([C@@H]3CCNC3)n2)n1	O=C(CCC(F)(F)F)N1CC[C@@H](c2nc(-c3nccs3)no2)C1	0.33	Journal of Medicinal Chemistry (2012), 55(1), 68-83	0.3694	0.0394
EDC	O=C(O)CCC(F)(F)F	c1csc(Cc2noc(C3CCNCC3)n2)c1	O=C(CCC(F)(F)F)N1CCC(c2nc(Cc3cccs3)no2)CC1	0.46	Journal of Medicinal Chemistry (2012), 55(1), 68-83	0.4838	0.0238
EDC	Nc1cc(C(=O)O)cc(C(F)(F)F)c1	Nc1cccc(-c2nnn[nH]2)c1	Nc1cc(C(=O)Nc2cccc(-c3nnn[nH]3)c2)cc(C(	0.52	Journal of Medicinal Chemistry (2012), 55(5),	0.5284	0.0084

			F)(F)F)c1		2163-2172		
EDC	N#Cc1ccc(CC(=O)O)cc1	NC(=O)c1ccsc1N	N#Cc1ccc(CC(=O)Nc2sccc2C(N)=O)cc1	0.15	Bioorganic and Medicinal Chemistry (2011), 19(8), 2582-2588	0.2276	0.0776
EDC	COc1cc([C@@H]2c3cc4c(cc3[C@H](OC(=O)CCC(=O)O)[C@H]3COC(=O)[C@H]23)OCO4)cc(OC)c1OC	COc1ccc2c(c1)c(CCN)c(C)n2Cc1ccc(Br)cc1	COc1ccc2c(c1)c(CCNC(=O)CCC(=O)O[C@H]1c3cc4c(cc3[C@@H](c3cc(OC)c(OC)c(OC)c3)[C@H]3C(=O)OC[C@@H]31)OCO4)c(C)n2Cc1ccc(Br)cc1	0.65	Bioorganic and Medicinal Chemistry Letters (2010), 20(5), 1787-1791	0.7044	0.0544
EDC	COc1cc([C@@H]2c3cc4c(cc3[C@H](OC(=O)CCC(=O)O)[C@H]3COC(=O)[C@H]23)OCO4)cc(OC)c1OC	C[C@@H](CN)c1ccccc1	COc1cc([C@@H]2c3cc4c(cc3[C@H](OC(=O)CCC(=O)N[C@H](C)c3ccccc3)[C@H]3COC(=O)[C@H]23)	0.66	Bioorganic and Medicinal Chemistry Letters (2010), 20(5), 1787-1791	0.7354	0.0754

			OCO4)cc(OC)c1OC				
EDC	Cc1ccc(C(=O)NC2CC2)cc1Nc1ncnn2cc(C(=O)O)c(C)c12	N[C@@H](CO)c1ccccc1	Cc1ccc(C(=O)NC2CC2)cc1Nc1ncnn2cc(C(=O)N[C@@H](CO)c3ccccc3)c(C)c12	0.59	Journal of Medicinal Chemistry (2010), 53(18), 6629-6639	0.5921	0.0021
EDC	C=C(C)[C@@H]1CC[C@]2(C(=O)O)CC[C@]3(C)[C@H](CC[C@@H]4[C@@]5(C)CC[C@H](O)C(C)(C)[C@@H]5CC[C@]43C)[C@@H]12	NC(CO)CO	C=C(C)[C@@H]1CC[C@]2(C(=O)NC(CO)CO)CC[C@]3(C)[C@H](CC[C@@H]4[C@@]5(C)CC[C@H](O)C(C)(C)[C@@H]5CC[C@]43C)[C@@H]12	0.48	Journal of Medicinal Chemistry (2010), 53(1), 178-190	0.5707	0.0907
EDC	C=C(C)[C@@H]1CC[C@]2(C(=O)O)CC[C@]3(C)[C@H](CC[C@@H]4[C@@]5(C)C	NCCOCCO	C=C(C)[C@@H]1CC[C@]2(C(=O)NCCOCCO)CC[C@]3(C)[C@H](CC[C@@H]4[C	0.48	Journal of Medicinal Chemistry (2010), 53(1), 178-190	0.5641	0.0841

	C[C@H](O)C(C)(C)[C@@H]5CC[C@]43C)[C@@H]12		@@]5(C)CC[C@H](O)C(C)(C)[C@@H]5CC[C@]43C)[C@@H]12				
EDC	CC=C(NC(=O)c1ccccc1)C(=O)O	NCCc1ccc(O)cc1	CC=C(NC(=O)c1ccccc1)C(=O)NCCc1ccc(O)cc1	0.49	Journal of Organic Chemistry (2003), 68(26), 10098-10102	0.5578	0.0678
EDC	O=C(O)C=1C=CC=CC1N	FC(F)(F)CN	O=C(NCC(F)(F)F)C=1C=CC=CC1N	0.4	Journal of Medicinal Chemistry (2023), 66(5), 3540-3565	0.6022	0.2022
EDC	C#CCN	O=C(O)CCNC(=O)C(O)C(C)(C)CO	O=C(NCC#C)CCNC(=O)C(O)C(C)(C)CO	0.38	Journal of the American Chemical Society (2006), 128(37), 12174-12184	0.4269	0.0469
EDC	O=C(O)CP(=O)(OCC)OCC	NCCCCCCCCCCCC	O=C(NCCCCCCCCCCCC)CP(=O)(O)O	0.29	ChemMedChem (2008), 3(12), 1936-1945	0.3898	0.0998

EDC	NC=1C=CC=CC1N	O=C(O)C1=CC=C(C=C1)CNC(=O)C2=CN=C(N2)C=3C=CC=CC3	O=C(NC=1C=CC=CC1N)C2=CC=C(C=C2)CNC(=O)C3=CN=C(N3)C=4C=CC=CC4	0.69	Journal of Medicinal Chemistry (2025), 68(3), 3048-3064	0.6	0.09
HBTU	O=C(O)C=1C=CC(=CC1)C=2C=CC=CC2	NCC1=CC=C(C=C1)C	O=C(NCC1=CC=C(C=C1)C)C=2C=CC(=CC2)C=3C=CC=CC3	0.97	Journal of Medicinal Chemistry (2021), 64(9), 5447-5469	0.8209	0.1491
HBTU	C#CCN	O=C(O)C1=CC=C(C=C1)S(=O)(=O)N	O=C(NCC#C)C1=CC=C(C=C1)S(=O)(=O)N	0.44	Bioconjugate Chemistry (2008), 19(8), 1614-1624	0.3989	0.0411
HBTU	O=C(O)CCCCCSC	O=C(O)C(N)C(OP(=O)(O)O)C	O=C(O)C(NC(=O)CCCCCSC)C(OP(=O)(O)O)C	0.38	JBIC, Journal of Biological Inorganic Chemistry (2002), 7(4-5), 500-513	0.4901	0.1101
HBTU	CCOc1ccc2c(c1)CC(C(=O)O)CO2	CNCc1cc(C(=O)OC)c(C)o1	CCOc1ccc2c(c1)CC(C(=O)N(C)Cc1cc(C(=O)OC)c(	0.72	ACS Medicinal Chemistry Letters (2022),	0.7222	0.0022

			C)o1)CO2		13(8), 1286-1294		
HBTU	O=C(O)CCCCCCNc1cccc2c1CN(C1CCC(=O)NC1=O)C2=O	NCc1ccc(Nc2nc(N)n(-c3ccc(-c4ccccc4)nn3)n2)cc1	Nc1nc(Nc2ccc(CNC(=O)CCCCCCCNc3cccc4c3CN(C3CCC(=O)NC3=O)C4=O)cc2)nn1-c1ccc(-c2ccccc2)nn1	0.528	European Journal of Medicinal Chemistry (2022), 234, 114253	0.6155	0.0875
HBTU	CCOC(=O)CCCCCCOc1ccc(-c2nc3ccc(C(=O)O)cc3[nH]2)cc1	Nc1ccccc1N	CCOC(=O)CCCCCCOc1ccc(-c2nc3ccc(C(=O)Nc4ccccc4N)cc3[nH]2)cc1	0.6	Journal of Medicinal Chemistry (2022), 65(4), 3667-3683	0.6687	0.0687
HBTU	CCOC(=O)CCCCCCOc1cccc(-c2nc3ccc(C(=O)O)cc3[nH]2)c1	Nc1ccccc1N	CCOC(=O)CCCCCCOc1cccc(-c2nc3ccc(C(=O)Nc4ccccc4N)cc3[nH]2)c1	0.59	Journal of Medicinal Chemistry (2022), 65(4), 3667-3683	0.657	0.067
HBTU	CC(=O)Nc1nc(-c2cccc(NC(=O)CC	C#CC[C@H](N)C(=O)N[C@@H](Cc1c	C#CC[C@H](NC(=O)CCC(=O)Nc1cc	0.55	ACS Medicinal Chemistry Letters	0.6071	0.0571

	C(=O)O)c2)cs1	n(C(c2ccccc2)(c2ccccc2)c2ccccc2)cn1)C(N)=O	cc(-c2csc(NC(C)=O)n2)c1)C(=O)N[C@@H](Cc1cn(C(c2ccccc2)(c2ccccc2)c2ccccc2)cn1)C(N)=O		(2021), 12(6), 899-906		
HBTU	CS(=O)(=O)N1CCN(Cc2cc3c(N4CCOCC4)nc(-c4ccc(NC(=O)Nc5ccc(C(=O)O)cc5)cc4)nn3c2)CC1	CN(C)C1CCNCC1	CN(C)C1CCN(C(=O)c2ccc(NC(=O)Nc3ccc(-c4nc(N5CCOCC5)c5cc(CN6CCN(S(C)(=O)=O)CC6)cn5n4)cc3)cc2)CC1	0.39	European Journal of Medicinal Chemistry (2021), 209, 112913	0.4825	0.0925
HBTU	Cc1cc(C(=O)O)cc(C)n1	CCCCNc1ncc2c(C3CCNCC3)cn([C@H]3CC[C@H](O)CC3)c2n1	CCCCNc1ncc2c(C3CCN(C(=O)c4cc(C)nc(C)c4)CC3)cn([C@H]3CC[C@H](O)CC3)c2n1	0.19	European Journal of Medicinal Chemistry (2021), 220, 113534	0.2572	0.0672
HBTU	CC(C)(C)OC(=O)N	C1CCNCC1	CC(C)(C)OC(=O)N	0.79	European Journal of	0.849	0.059

	[C@@H](Cc1ccccc1)C(=O)O		[C@@H](Cc1ccccc1)C(=O)N1CCCCC1		Organic Chemistry (2021), 25		
HBTU	COc1cccc2c1C(=O)c1c(O)c3c(c(O)c1C2=O)C[C@@](O)(C(=O)O)C[C@@H]3O[C@H]1C[C@H]2[C@H](O[C@@H]3[C@@H](OC)OCCN32)[C@H](C)O1	NCCc1cccc(N)c1	COc1cccc2c1C(=O)c1c(O)c3c(c(O)c1C2=O)C[C@@](O)(C(=O)NCCc1cccc(N)c1)C[C@@H]3O[C@H]1C[C@H]2[C@H](O[C@@H]3[C@@H](OC)OCCN32)[C@H](C)O1	0.65	Bioorganic and Medicinal Chemistry Letters (2020), 30(24), 127640	0.6581	0.0081
HBTU	O=C(O)c1ccc(CCc2ccccc2NS(=O)(=O)c2ccccc2)cc1	Nc1ccccc1N	Nc1ccccc1NC(=O)c1ccc(CCc2ccccc2NS(=O)(=O)c2ccccc2)cc1	0.6	European Journal of Medicinal Chemistry (2020), 192, 112158	0.6145	0.0145
HBTU	Cc1nc(C(C)(C)C)sc1C(=O)O	NC12CC3CC(CC(C3)C1)C2	Cc1nc(C(C)(C)C)sc1C(=O)NC12CC3CC(CC(C3)C1)C2	0.582	European Journal of Medicinal Chemistry (2019), 180, 154-	0.6776	0.0956

					170		
HBTU	Cc1nc(C)c(C(=O)O)s1	NC1CCCc2ccccc21	Cc1nc(C)c(C(=O)NC2CCCc3ccccc32)s1	0.714	European Journal of Medicinal Chemistry (2019), 180, 154-170	0.7252	0.0112
HBTU	O=C(O)c1c[nH]nc1C1CCCCC1	CCN(CC)c1ccc(N)cc1	CCN(CC)c1ccc(NC(=O)c2c[nH]nc2C2CCCCC2)cc1	0.58	Journal of Medicinal Chemistry (2019), 62(17), 8284-8310	0.6766	0.0966
HBTU	Cc1ccccc1OCC(=O)O	Cn1ccc(N)n1	Cc1ccccc1OCC(=O)Nc1ccn(C)n1	0.59	MedChem Comm (2019), 10(8), 1361-1369	0.6674	0.0774
HBTU	Cc1ccccc1OCC(=O)O	Nc1ccc2cc[nH]c2c1	Cc1ccccc1OCC(=O)Nc1ccc2cc[nH]c2c1	0.65	MedChem Comm (2019), 10(8), 1361-1369	0.7389	0.0889
HBTU	Cc1ccccc1OCC(=O)O	Nc1ccccc1	Cc1ccccc1OCC(=O)Nc1ccccc1	0.83	MedChem Comm (2019), 10(8), 1361-1369	0.8403	0.0103
HBTU	[N-]=[N+]=NC(C)(C)C	[N-]=[N+]=NCCCN	[N-]=[N+]=NCCCNC(=	0.66	Chemistry - A European	0.4119	0.2481

	C(=O)O		O)CC(N=[N+]=[N-])(C)C		Journal (2019), 25(3), 754-758		
HBTU	[N-]=[N+]=NC(C(=O)O)CC(C)C	C#CCN	[N-]=[N+]=NC(C(=O)NCC#C)CC(C)C	0.68	Angewandte Chemie, International Edition (2009), 48(26), 4725-4729, S4725/1-S4725/28	0.461	0.219
HBTU	O=C(O)C(=O)CC(C)C	Cl.O=C(OC)CCCN	O=C(NCCCC(=O)OC)C(=O)CC(C)C	0.27	Proteins: Structure, Function, and Bioinformatics (2014), 82(9), 2067-2077	0.2662	0.0038
HBTU	O=C(O)c1ccc(-n2cccc2)cc1	Nc1ccccc1F	O=C(Nc1ccccc1F)c1ccc(-n2cccc2)cc1	0.59	Bioorganic Chemistry (2018), 81, 440-453	0.64	0.05
HBTU	O=C(O)c1ccc(-n2cccc2)cc1	COc1ccccc(N)c1	COc1ccccc(NC(=O)c2ccc(-n3cccc3)cc2)c1	0.69	Bioorganic Chemistry (2018), 81, 440-453	0.7741	0.0841
HBTU	O=C(O)C	Cc1sc(N)n	Cc1sc(NC	0.21	European	0.2175	0.007

	1(c2ccc3c(c2)OC(F)(F)O3)CC1	c1-c1ccc(Br)cc1	(=O)C2(c3ccc4c(c3)OC(F)(F)O4)CC2)nc1-c1ccc(Br)cc1		Journal of Medicinal Chemistry (2018), 144, 179-200		5
HBTU	O=C(O)C1(c2ccc3c(c2)OCO3)CC1	Cc1sc(N)nc1-c1ccccc1	Cc1sc(NC(=O)C2(c3ccc4c(c3)OCO4)CC2)nc1-c1ccccc1	0.51	European Journal of Medicinal Chemistry (2018), 144, 179-200	0.5476	0.0376
HBTU	C[C@H](NC(=O)OCC1c2ccccc2-c2ccccc21)C(=O)O	CNC(=O)[C@@H]1CCCN1	CNC(=O)[C@@H]1CCCN1C(=O)[C@H](C)NC(=O)OCC1c2ccccc2-c2ccccc21	0.65	Bioorganic and Medicinal Chemistry (2017), 25(3), 897-911	0.7041	0.0541
HBTU	O=C(O)c1ccc(CNC2=C(Cl)C(=O)c3ccccc3C2=O)cc1	Nc1ccccc1	O=C(Nc1ccccc1)c1ccc(CNC2=C(Cl)C(=O)c3ccccc3C2=O)cc1	0.7633	European Journal of Medicinal Chemistry (2017), 140, 84-91	0.8096	0.0463
HBTU	O=C(O)COc1ccc(Cl)cc1	O=[N+]([O-])c1ccc(N2CCNCC2)cc1	O=C(COc1ccc(Cl)cc1)N1CCN(c2ccc([N+](=O)[O-	0.12	Bioorganic and Medicinal Chemistry (2016),	0.1482	0.0282

			])cc2)CC1		24(19), 4660-4674		
HBTU	O=C(O)COc1ccc(Cl)cc1	O=[N+]([O-])c1ccccc1N1CCNCC1	O=C(COc1ccc(Cl)cc1)N1CCN(c2ccccc2[N+](=O)[O-])CC1	0.36	Bioorganic and Medicinal Chemistry (2016), 24(19), 4660-4674	0.4302	0.0702
HBTU	O=C(O)COc1ccc(Cl)cc1Cl	O=[N+]([O-])c1cccc(N2CCNCC2)c1	O=C(COc1ccc(Cl)cc1Cl)N1CCN(c2cccc([N+](=O)[O-])c2)CC1	0.35	Bioorganic and Medicinal Chemistry (2016), 24(19), 4660-4674	0.4484	0.0984
HBTU	O=C(O)COc1ccc(F)cc1	O=[N+]([O-])c1ccc(N2CCNCC2)cc1	O=C(COc1ccc(F)cc1)N1CCN(c2ccc([N+](=O)[O-])cc2)CC1	0.12	Bioorganic and Medicinal Chemistry (2016), 24(19), 4660-4674	0.189	0.069
HBTU	O=C(O)COc1cccc(Cl)c1	O=[N+]([O-])c1ccccc1N1CCNCC1	O=C(COc1cccc(Cl)c1)N1CCN(c2ccccc2[N+](=O)[O-])CC1	0.44	Bioorganic and Medicinal Chemistry (2016), 24(19), 4660-4674	0.4843	0.0443
HBTU	O=C(O)COc1ccccc1C(F)(F)F	O=[N+]([O-])c1cccc(N2CCNC	O=C(COc1ccccc1C(F)(F)F)N1CCN(c2cc	0.39	Bioorganic and Medicinal Chemistry	0.4832	0.0932

		C2)c1	cc([N+](=O)[O-])c2)CC1		(2016), 24(19), 4660-4674		
HBTU	O=C(OC(C)(C)C)NC(C(=O)O)CC(C)C	NCCCCCCCCCCCCCCC	O=C(NCCCCCCCCCCCCCCCC)C(N)CC(C)C	0.24	ACS Nano (2019), 13(8), 9292-9297	0.4401	0.2001
HBTU	O=C(OC(C)(C)C)NC(C(=O)O)C	NCCN(CC)CC	O=C(OC(C)(C)C)NC(C(=O)NCCN(C)C)C	0.22	Journal of Organic Chemistry (2003), 68(20), 7788-7794	0.393	0.173
HBTU	O=C(O)C=1C=CC=C(OCC=2C=CC=CC2C=3C=CC=CC3COC4=CC=CC(=C4)C(=O)O)C1	NCC=1C=CC=CC1	O=C(NCC=1C=CC=CC1)C=2C=CC=C(OCC=3C=CC=CC3C=4C=CC=CC4COC5=CC=CC(=C5)C(=O)NCC=6C=CC=CC6)C2	0.65	Arabian Journal of Chemistry (2019), 12(7), 1676-1683	0.5765	0.0735
HBTU	O=C(O)COc1ccccc1Cl	O=[N+]([O-])c1ccccc1N1CCNCC1	O=C(COc1ccccc1Cl)N1CCN(c2ccccc2[N+](=O)[O-])CC1	0.49	Bioorganic and Medicinal Chemistry (2016), 24(19),	0.5878	0.0978

					4660-4674		
HBTU	CCc1c(C(=O)O)cc(C)n1-c1ccccc1	Nc1ccc2ccccc2c1	CCc1c(C(=O)Nc2ccc3ccccc3c2)cc(C)n1-c1ccccc1	0.35	European Journal of Medicinal Chemistry (2016), 122, 619-634	0.3973	0.0473
HBTU	CCCCCn1cc(C(=O)O)c(=O)n2nc(-c3ccc(C)cc3)nc12	NC12CC3CC(CC(C3)C1)C2	CCCCCn1cc(C(=O)NC23CC4CC(CC(C4)C2)C3)c(=O)n2nc(-c3ccc(C)cc3)nc12	0.44	European Journal of Medicinal Chemistry (2016), 113, 11-27	0.5061	0.0661
HBTU	CCCCCn1cc(C(=O)O)c(=O)n2nc(C)nc12	NC12CC3CC(CC(C3)C1)C2	CCCCCn1cc(C(=O)NC23CC4CC(CC(C4)C2)C3)c(=O)n2nc(C)nc12	0.56	European Journal of Medicinal Chemistry (2016), 113, 11-27	0.6104	0.0504
HBTU	CCCCCn1cc(C(=O)O)c(=O)n2nc(N3CCOCC3)nc12	NC12CC3CC(CC(C3)C1)C2	CCCCCn1cc(C(=O)NC23CC4CC(CC(C4)C2)C3)c(=O)n2nc(N3CCOCC3)nc12	0.4	European Journal of Medicinal Chemistry (2016), 113, 11-27	0.4036	0.0036
HBTU	CCCCCn1	NC12CC3	CCCCCn1	0.55	European	0.5939	0.043

	cc(C(=O)O)c(=O)n2nc(SC)nc12	CC(CC(C3)C1)C2	cc(C(=O)NC23CC4CC(CC(C4)C2)C3)c(=O)n2nc(SC)nc12		Journal of Medicinal Chemistry (2016), 113, 11-27		9
HBTU	O=C(O)Nc1cccc(Cl)c1Cl	COC(=O)c1cccc(-c2ccc(CN)c3cnccc23)c1	COC(=O)c1cccc(-c2ccc(CNC(=O)Nc3cccc(Cl)c3Cl)c3cnccc23)c1	0.33	Journal of Medicinal Chemistry (2015), 58(23), 9345-9353	0.3594	0.0294
HBTU	Cc1ccccc1C(=O)O	Nc1cccc(O)n1	Cc1ccccc1C(=O)Nc1cccc(O)n1	0.61	Journal of Medicinal Chemistry (2014), 57(15), 6393-6402	0.693	0.083
HBTU	Cc1ccncc1C(=O)O	Nc1ccccn1	Cc1ccncc1C(=O)Nc1ccccn1	0.55	Journal of Medicinal Chemistry (2014), 57(15), 6393-6402	0.5715	0.0215
HBTU	CCc1ccccc1C(=O)O	Nc1ccccn1	CCc1ccccc1C(=O)Nc1ccccn1	0.56	Journal of Medicinal Chemistry (2014), 57(15), 6393-6402	0.5949	0.0349
HBTU	COc1ccccc(C(=O)O)c1C	Nc1ccccn1	COc1ccccc(C(=O)Nc2ccccn2)c	0.65	Journal of Medicinal Chemistry	0.6816	0.0316

			1C		(2014), 57(15), 6393-6402		
HBTU	O=C(O)c1cc(Cc2n[nH]c(=O)c3ccccc23)ccc1F	CNC1CC(=O)c2ccccc21	CN(C(=O)c1cc(Cc2n[nH]c(=O)c3ccccc23)ccc1F)C1CC(=O)c2ccccc21	0.68	Journal of Medicinal Chemistry (2013), 56(7), 2885-2903	0.7169	0.0369
HBTU	Cc1csc2[nH]cc(C(=O)O)c(=O)c12	NC12CC3CC(CC(C3)C1)C2	Cc1csc2[nH]cc(C(=O)NC34CC5CC(CC(C5)C3)C4)c(=O)c12	0.57	Journal of Medicinal Chemistry (2013), 56(3), 1098-1112	0.5908	0.0208
HBTU	Cc1sc2[nH]cc(C(=O)O)c(=O)c2c1C	NC12CC3CC(CC(C3)C1)C2	Cc1sc2[nH]cc(C(=O)NC34CC5CC(CC(C5)C3)C4)c(=O)c2c1C	0.55	Journal of Medicinal Chemistry (2013), 56(3), 1098-1112	0.6003	0.0503
HBTU	CCCCCn1cc(C(=O)O)c(=O)c2c(C)noc21	NC12CC3CC(CC(C3)C1)C2	CCCCCn1cc(C(=O)NC23CC4CC(CC(C4)C2)C3)c(=O)c2c(C)noc21	0.59	Journal of Medicinal Chemistry (2013), 56(3), 1098-1112	0.6182	0.0282
HBTU	CC1COc2cccc3c(=O)c(C(=O)	NC12CC3CC(CC(C3)C1)C2	CC1COc2cccc3c(=O)c(C(=O)	0.55	Journal of Medicinal Chemistry	0.6021	0.0521

	O)cn1c23		NC45CC6CC(CC(C6)C4)C5)cn1c23		(2012), 55(14), 6608-6623		
HBTU	Cc1ccc(C2COc3cccc4c(=O)c(C(=O)O)cn2c34)cc1	NC12CC3CC(CC(C3)C1)C2	Cc1ccc(C2COc3cccc4c(=O)c(C(=O)NC56CC7CC(CC(C7)C5)C6)cn2c34)cc1	0.59	Journal of Medicinal Chemistry (2012), 55(14), 6608-6623	0.6112	0.0212
HBTU	O=C(O)c1cc(Br)cc([N+](=O)[O-])c1	NCc1ccccc1	O=C(NCc1ccccc1)c1cc(Br)cc([N+](=O)[O-])c1	0.74	Bioorganic and Medicinal Chemistry (2011), 19(5), 1823-1838	0.7486	0.0086
HBTU	O=C(O)Cc1ccc(Br)cc1F	C1CCNC1	O=C(Cc1ccc(Br)cc1F)N1CCCC1	0.8	Journal of Medicinal Chemistry (2011), 54(1), 78-94	0.8146	0.0146
HBTU	O=C(O)CCCCN1C[C@H](O)[C@@H](O)[C@@H](O)[C@H]1CO	NCCNC(=O)CCCC[C@@H]1SC[C@@H]2NC(=O)N[C@@H]21	O=C(CCCCCN1C[C@H](O)[C@@H](O)[C@@H](O)[C@H]1CO)NCCNC(=O)CCC[C@	0.5	Bioorganic and Medicinal Chemistry Letters (2010), 20(14), 4077-4079	0.5721	0.0721

			@H]1SC[C@@H]2NC(=O)N[C@@H]21				
HBTU	O=C(O)CC(c1ccccc1)c1ccccc1	CNC(=O)[C@@H]1CCCN1	CNC(=O)[C@@H]1CCCN1C(=O)CC(c1ccccc1)c1ccccc1	0.7	Journal of Medicinal Chemistry (2009), 52(9), 2708-2715	0.7182	0.0182
HBTU	O=C(O)C1CC(=O)N(C2CCCCCC2)C1	CNCc1c2ccccc2cc2ccccc12	CN(Cc1c2ccccc2cc2ccccc12)C(=O)C1CC(=O)N(C2CCCCCC2)C1	0.629	Journal of Medicinal Chemistry (2006), 49(21), 6308-6323	0.7047	0.0757
HBTU	CN(CC(=O)O)C(=O)CN(C)C(=O)CCN1CCCCC1	CNCCN(C)Cc1ccccc1	CN(CCN(C)C(=O)CN(C)C(=O)CN(C)C(=O)CCN1CCCCC1)Cc1ccccc1	0.55	Journal of Medicinal Chemistry (2000), 43(25), 4822-4833	0.6376	0.0876
HBTU	CN(CC(=O)O)C(=O)CN(C)C(=O)CN1CCCCC1	CNCCN(C)Cc1ccccc1	CN(CCN(C)C(=O)CN(C)C(=O)CN(C)C(=O)CN1CCCCC1)Cc1ccccc1	0.71	Journal of Medicinal Chemistry (2000), 43(25), 4822-4833	0.7511	0.0411
HBTU	CN(CC(=O)O)C(=	c1ccc(CN2CCNCC	CN(CC(=O)N(C)C	0.65	Journal of Medicinal	0.6628	0.0128

	O)CN(C)C(=O)CN1CCCCC1	2)cc1	C(=O)N1CCN(Cc2ccccc2)CC1)C(=O)CN1CCCCC1		Chemistry (2000), 43(25), 4822-4833		
HBTU	[N-]=[N+]=NC1C(OC2OC(OC21)(C)C)CC(=O)O	NCC=1C=CC=CC1	[N-]=[N+]=NC1C(OC2OC(OC21)(C)C)CC(=O)NCC=3C=CC=CC3	0.62	RSC Advances (2015), 5(25), 19455-19464	0.6581	0.0381
HBTU	O=C(OC(C)(C)C)NC(C(=O)O)CSC(C=1C=CC=CC1)(C=2C=CC=CC2)C=3C=CC=CC3	NCC=1C=CC=CC1	O=C(OC(C)(C)C)NC(C(=O)NCC=1C=CC=CC1)CSC(C=2C=CC=CC2)(C=3C=CC=CC3)C=4C=CC=CC4	0.82	Bioconjugate Chemistry (2014), 25(2), 202-206	0.5845	0.2355
DCC	C(OC(=O)N1[C@@H](C(O)=O)CCC1)C2C=3C(C=4C2=CC=CC4)=CC=CC3	O(C)C1=C(OC)C=C(/C=C\C2=CC(N)=C(OC)C=C2)C=C1OC	C(OC(=O)N1[C@@H](C(NC2=C(OC)C=CC(/C=C\C3=CC(OC)=C(OC)C(OC)=C3)=C2)=O)CCC1)	0.73	Molecules (2020), 25(3), 660	0.6498	0.0802

			C4C=5C(C=6C4=CC=CC6)=CC=CC5				
DCC	O=C(O)C=1N=NN(C1)C2=CC=C(C=C2)C(O)(C(F)(F)F)C(F)(F)F	NC=1C=CC=CC1	O=C(NC=1C=CC=CC1)C=2N=NN(C2)C3=CC=C(C=C3)C(O)(C(F)(F)F)C(F)(F)F	0.55	Bioorganic & Medicinal Chemistry Letters (2021), 42, 127999	0.5258	0.0242
DCC	O=C(OC(C)(C)C)NCCCCCC(=O)O	FC1=CC=C(C=C1)CN	Cl.O=C(NCC1=CC=C(F)C=C1)CCCCCN	0.85	Chinese Chemical Letters (2010), 21(3), 257-260	0.6038	0.2462
DCC	O=C(OC(C)(C)C)NCC(=O)O	OCCN	O=C(OC(C)(C)C)NCC(=O)NCCO	0.35	Biopolymers (2006), 84(6), 605-614	0.4435	0.0935
DCC	Cc1cc(C)c(-n2cc(C(=O)O)nn2)c(C)c1	Nc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1	Cc1cc(C)c(-n2cc(C(=O)Nc3ccc(C(O)(C(F)(F)F)C(F)(F)F)cc3)nn2)c(C)c1	0.28	Bioorganic and Medicinal Chemistry Letters (2021), 42, 127999	0.3261	0.0461
DCC	COc1ccc(-n2cc(C(=O)O)nn2)	Nc1ccc(C(O)(C(F)(F)F)C(F)(F)	COc1ccc(-n2cc(C(=O)Nc3ccc	0.47	Bioorganic and Medicinal	0.4803	0.0103

	cc1OC	F)cc1	(C(O)(C(F)(F)F)C(F)(F)F)cc3)nn2)cc1OC		Chemistry Letters (2021), 42, 127999		
DCC	COc1ccc2[nH]nc(C(=O)O)c2c1	NCC1CCN(Cc2ccc(OCc3ccccc3)cc2)CC1	COc1ccc2[nH]nc(C(=O)NCC3CCN(Cc4ccc(OCc5ccccc5)cc4)CC3)c2c1	0.77	Journal of Medicinal Chemistry (2015), 58(22), 8920-8937	0.8345	0.0645
DCC	C[C@]12CC[C@H]3[C@@H]([C@H](C/C=C/C(=O)O)CC4=CC(=O)CC[C@@]43C)[C@@H]1CC[C@@H]2O	NCCO	C[C@]12CC[C@H]3[C@@H]([C@H](C/C=C/C(=O)NCCO)CC4=CC(=O)CC[C@@]43C)[C@@H]1CC[C@@H]2O	0.44	Medicinal Chemistry (2015), 11(6), 531-539	0.508	0.068
DCC	O=C(O)c1cnc(F)c(I)c1	CCN(CC)CCN	CCN(CC)CCNC(=O)c1cnc(F)c(I)c1	0.31	European Journal of Medicinal Chemistry (2015), 92, 818-838	0.3857	0.0757
DCC	O=C(O)CNC(=O)OCC1c2ccccc2-	COc1ccc([C@H]2[C@H](c3ccc(N)cc3)C	COc1ccc([C@H]2[C@H](c3ccc(NC(=O)	0.54	European Journal of Medicinal Chemistry	0.5845	0.0445

	c2ccccc21	(=O)N2c2cc(OC)c(OC)c(OC)c2)cc1	CNC(=O)OCC4c5ccccc5-c5ccccc54)cc3)C(=O)N2c2cc(OC)c(OC)c(OC)c2)cc1		(2013), 62, 705-721		
DCC	O=C(N[C@@H](Cc1ccccc1)C(=O)O)OCC1c2ccccc2-c2ccccc21	COc1ccc([C@H]2[C@H](c3ccccc3)C(=O)N2c2cc(OC)c(OC)c(OC)c2)cc1N	COc1ccc([C@H]2[C@H](c3ccccc3)C(=O)N2c2cc(OC)c(OC)c(OC)c2)cc1NC(=O)[C@H](Cc1ccccc1)NC(=O)OCC1c2ccccc2-c2ccccc21	0.54	European Journal of Medicinal Chemistry (2013), 62, 705-721	0.5723	0.0323
DCC	O=C(O)c1ccccc1	Cc1csc(-c2nc(CN)[nH]c2-c2ccc3c(c2)OCO3)n1	Cc1csc(-c2nc(CNC(=O)c3ccccc3)[nH]c2-c2ccc3c(c2)OCO3)n1	0.39	Bioorganic and Medicinal Chemistry Letters (2012), 22(5), 2024-2029	0.4651	0.0751
DCC	O=C(O)C(CCCCCCCCCC)P(=O)(OC)	OC1=CC=C(C=C1)CCN	O=C(NCCC1=CC=C(O)C=C1)C(CCCCC	0.62	Organic & Biomolecular Chemistry	0.5987	0.0213

	OC		CCCCC)P(=O)(OC)OC		(2009), 7(17), 3491-3498		
DCC	O=C(O)CCCN(CP(=O)(OC)OC)CP(=O)(OC)OC	NCCCN(CCCN)CCCCN(CCCN)CCCN	O=C(NCCCN(CCCNC(=O)CCCN(CP(=O)(OC)OC)CP(=O)(OC)OC)CCCCN(CCCNC(=O)CCCN(CP(=O)(OC)OC)CP(=O)(OC)OC)CCCNC(=O)CCCN(CP(=O)(OC)OC)CP(=O)(OC)OC)CCCN(CP(=O)(OC)OC)CP(=O)(OC)OC	0.69	Nature Communications (2015), 6, 7722	0.5805	0.1095
DCC	Cl.O=C(OC)C(N)CSSCC(N)C(=O)OC	O=C(O)C1CC1C	O=C(OC)C(NC(=O)C1CC1C)CSSCC(NC(=O)C2CC2C)C(=O)OC	0.44	Tetrahedron Letters (1995), 36(8), 1189-92	0.1903	0.2497
DCC	Cc1csc(-	CCCN	CCCNC(=	0.31	Bioorgani	0.35	0.04

	c2nc(C(=O)O)[nH]c2-c2ccc3c(c2)OCO3)n1		O)c1nc(-c2nc(C)cs2)c(-c2ccc3c(c2)OCO3)[nH]1		c and Medicinal Chemistry Letters (2012), 22(5), 2024-2029		
DCC	CC(C)Cc1ccc([C@@H](C)C(=O)O)cc1	NCCCCCO	CC(C)Cc1ccc([C@@H](C)C(=O)NCCCCCO)cc1	0.55	Journal of Medicinal Chemistry (2005), 48(13), 4312-4331	0.5752	0.0252
DCC	N#Cc1c(C(=O)O)cn2ccccc12	N#CN=C(NCCCN)NCCCOc1cccc(CN2CCCCC2)c1	N#CN=C(NCCCNC(=O)c1cn2ccccc2c1C#N)NCCCOc1cccc(CN2CCCCC2)c1	0.35	Bioorganic and Medicinal Chemistry (2004), 12(24), 6495-6503	0.3939	0.0439
DCC	C[C@H](NC(=O)[C@H](Cc1ccccc1)NC(=O)OCc1ccccc1)C(=O)O	COc1nsc(N)n1	COc1nsc(NC(=O)[C@H](C)NC(=O)[C@H](Cc2ccccc2)NC(=O)OCc2ccccc2)n1	0.67	Bioorganic and Medicinal Chemistry (2003), 11(24), 5529-5537	0.676	0.006
DCC	Nc1ccc(C(=O)O)cc1	CNCC(=O)N1CCC(Cn2c(C)nc3cnccc32)CC1	Cc1nc2cnccc2n1CC1CCN(C(=O)CN(C)C(=O)c2c	0.32	Journal of Medicinal Chemistry (2001), 44(18),	0.3276	0.0076

			cc(N)cc2)CC1		3001-3013		
DCC	O=C(O)CCCC(C1CC1)C1CC1	O=C([C@@H]1C[C@@H]2CCCC[C@@H]2N1)N1CCCC1	O=C([C@@H]1C[C@@H]2CCCC[C@@H]2N1C(=O)CCCC(C1CC1)C1CC1)N1CCCC1	0.39	Journal of Medicinal Chemistry (1996), 39(12), 2379-2391	0.481	0.091
DCC	O=C(O)C(C1=CC=C(C=C1)CC(C)C)C	O=C(O)CCN	O=C(O)CCNC(=O)C(C1=CC=C(C=C1)CC(C)C)C	0.3	Journal of Medicinal Chemistry (2005), 48(13), 4312-4331	0.3612	0.0612
DCC	O=C(O)CN(CC(=O)O)CC=1C=CC=CC1	NC(C)(C)C	O=C(NC(C)(C)C)CN(CC=1C=CC=CC1)CC(=O)NC(C)(C)C	0.49	Mendeleev Communications (2023), 33(2), 157-159	0.2889	0.2011
DCC	O=C(O)C1=CC=CN=C1N	NNC	O=C(C1=CC=CN=C1N)N(N)C	0.22	Chemical & Pharmaceutical Bulletin (1987), 35(1), 80-9	0.3013	0.0813
DCC	O=C(O)CNC(=O)C(	NC=1C=CC=CC1C	O=C(NC=1C=CC=C	0.4	Chemical Communi	0.3791	0.0209

	F)(F)F		C1C)CNC(=O)C(F)(F)F		cations (Cambridge, United Kingdom) (2022), 58(69), 9638-9641		
DCC	O=C(O)CCS	OCC(O)CN	O=C(NCC(O)CO)CCS	0.6	Polymer Preprints (American Chemical Society, Division of Polymer Chemistry) (2009), 50(1), No pp. given	0.6918	0.0918
DCC	O=C(O)CCCc1ccccc1	O=C([C@H]1NC2CCC1CC2)N1CCCC1	O=C([C@@H]1C2CCC(CC2)N1C(=O)CCCc1ccccc1)N1CCCC1	0.35	Journal of Medicinal Chemistry (1996), 39(12), 2379-2391	0.4391	0.0891
DCC	CC(C)C[C@H](N[C@H](CCN1C(=O)c2cc3ccccc3cc2C1=O)C(=O)OC(C)(C)C)C(=O)O	NCc1ccccc1	CC(C)C[C@H](N[C@H](CCN1C(=O)c2cc3ccccc3cc2C1=O)C(=O)OC(C)(C)C)C(=O)NCc1	0.5	Journal of Medicinal Chemistry (1994), 37(5), 674-688	0.5718	0.0718

			ccccc1				
DCC	CC(C)C[C@H](N[C@H](CCO[Si](C)(C)C(C)(C)C)C(=O)OC(C)(C)C)C(=O)O	CNC(=O)[C@@H](N)Cc1ccccc1	CNC(=O)[C@H](Cc1ccccc1)NC(=O)[C@H](CC(C)C)N[C@H](CCO[Si](C)(C)C(C)(C)C)C(=O)OC(C)(C)C	0.43	Journal of Medicinal Chemistry (1994), 37(5), 674-688	0.4884	0.0584
DCC	O=C(O)c1ccccc1	CC1(C)C(=O)c2ccc(C#N)cc2[C@H](N)[C@H]1O	CC1(C)C(=O)c2ccc(C#N)cc2[C@H](NC(=O)c2ccccc2)[C@H]1O	0.55	Journal of Medicinal Chemistry (1993), 36(15), 2121-2133	0.5818	0.0318
DCC	O=C(O)c1ccoc1	CC1(C)C(=O)c2ccc(C#N)cc2[C@H](N)[C@H]1O	CC1(C)C(=O)c2ccc(C#N)cc2[C@H](NC(=O)c2ccoc2)[C@H]1O	0.67	Journal of Medicinal Chemistry (1993), 36(15), 2121-2133	0.701	0.031
DCC	O=C(O)[C@H](Cc1ccccc1)NS(=O)(=O)N1CCOCC1	N[C@@H](Cc1csc(NC(=O)OCc2ccccc2)n1)C(=O)N[C@@H](CC1CCCCC1)[C	O=C(C[C@H](O)[C@H](CC1CCCCC1)NC(=O)[C@H](Cc1csc(NC(=O)OCc2ccc	0.42	Journal of Medicinal Chemistry (1992), 35(14), 2562-2572	0.4378	0.0178

		@@H](O)CC(=O)NCCN1CCOCC1	cc2)n1)NC(=O)[C@H](Cc1ccccc1)NS(=O)(=O)N1CCOCC1)NCCN1CCOCC1				
DCC	COc1ccc(C[C@H](NS(=O)(=O)N2CCOCC2)C(=O)O)cc1	CC(C)C[C@H](O)[C@H](O)[C@H](CC1CCCCC1)NC(=O)[C@@H](N)Cc1csc(NC(=O)OCc2ccccc2)n1	COc1ccc(C[C@H](NS(=O)(=O)N2CCOCC2)C(=O)N[C@@H](Cc2csc(NC(=O)OCc3ccccc3)n2)C(=O)N[C@@H](CC2CCCCC2)[C@@H](O)[C@@H](O)CC(C)C)cc1	0.44	Journal of Medicinal Chemistry (1992), 35(14), 2562-2572	0.4863	0.0463
DCC	O=C(O)C=CC=CC	OCC(N)C(O)C	O=C(C=CC=CC)NC(CO)C(O)C	0.7	Chemistry - A European Journal (2004), 10(1), 173-181	0.5594	0.1406
DCC	O=C(O)CCCCSSCCCCCC(	C#CCN	O=C(NCC#C)CCCCCSSCCC	0.73	Biomaterials Science	0.4959	0.2341

	=O)O		CCC(=O)NCC#C		(2020), 8(11), 3186-3192		
DCC	O=C(O)C1=CC=CN1	NC=1C=CC=CC1N	O=C(NC=1C=CC=CC1N)C2=CC=CN2	0.31	ACS Chemical Biology (2017), 12(6), 1644-1655	0.58	0.27
DCC	O=C(OC)CCN(CCC(=O)OC)CCCCC(C(=O)O)N(CCC(=O)OC)CCC(=O)OC	NCCCCCCCCCCCC	O=C(OC)CCN(CCC(=O)OC)CCCCC(C(=O)NCCCCCCCCCCC)N(CCC(=O)OC)CCC(=O)OC	0.7	European Journal of Medicinal Chemistry (2015), 105, 106-119	0.5075	0.1925
PyBOP	[N-]=[N+]=NCCCCCCCC(=O)O	O1C=CC=C1CN	[N-]=[N+]=NCCCCCCCC(=O)NCC=1OC=CC1	0.81	European Journal of Medicinal Chemistry (2023), 250, 115170	0.6451	0.1649
PyBOP	FC1=CC=C(C=C1)CN	O=C(O)C1=CC=C(C=C1)CNC(=O)OC(C)(C)C	O=C(NCC1=CC=C(F)C=C1)C2=CC=C(C=C2)CN	0.59	Journal of Medicinal Chemistry (2022), 65(10), 7246-7261	0.3494	0.2406
PyBOP	ClC1=CC=C(C=C1)	O=C(O)C=1C=CC=	O=C(NCCC(C=1C=	0.67	Journal of Medicinal	0.7368	0.0668

	C(C=2C=CC=CC2)CCN	C(I)C1	CC=CC1)C2=CC=C(Cl)C=C2)C=3C=CC=C(I)C3		Chemistry (2020), 63(20), 11498-11521		
PyBOP	CN(CCc1ccccc1)c1nc(C(=O)O)cc(N2CCOCC2)n1	NCCO	CN(CCc1ccccc1)c1nc(C(=O)NCCO)cc(N2CCOCC2)n1	0.59	Journal of Medicinal Chemistry (2021), 64(1), 481-515	0.6009	0.0109
PyBOP	O=C(O)c1nc2ccccc2[nH]1	COc1cccc(OCCN2CCNCC2)c1	COc1cccc(OCCN2CCN(C(=O)c3nc4ccccc4[nH]3)CC2)c1	0.61	Journal of Medicinal Chemistry (2020), 63(10), 5526-5567	0.6395	0.0295
PyBOP	O=C(O)/C=C/c1ccccc1	Cc1cc(NCCN)c2ccccc2n1	Cc1cc(NCCNC(=O)/C=C/c2ccccc2)c2ccccc2n1	0.57	Bioorganic Chemistry (2019), 93, 103310	0.5717	0.0017
PyBOP	O=C(O)c1cnn2cccnc12	CN1CCC2(CC1)Cc1cc(N)c(N3CCOCC3)cc1O2	CN1CCC2(CC1)Cc1cc(NC(=O)c3cnn4cccnc34)c(N3CCOCC3)cc1O2	0.4	Journal of Medicinal Chemistry (2019), 62(13), 6223-6240	0.4025	0.0025
PyBOP	COC(=O)c1ccc(C(=O)O)cc1	C1CCNC1	COC(=O)c1ccc(C(=O)N2CCCC2)cc1	0.67	Journal of the American Chemical Society	0.6898	0.0198

					(2018), 140(43), 14440-14454		
PyBOP	CC1(C)O[C@@H]2[C@H](O1)[C@@H](CO[Si](C)(C)C(C)(C)C)O[C@H]2n1c(SCC(=O)O)nc2c(N)ncnc21	NCCc1ccccc1	CC1(C)O[C@@H]2[C@H](O1)[C@@H](CO[Si](C)(C)C(C)(C)C)O[C@H]2n1c(SCC(=O)NCCc2ccccc2)nc2c(N)ncnc21	0.65	European Journal of Medicinal Chemistry (2016), 124, 1041-1056	0.6606	0.0106
PyBOP	COc1ccc(Cl)cc1C(=O)O	CC(C)(C)OC(=O)N1CCN(Cc2cccc(N)c2)CC1	COc1ccc(Cl)cc1C(=O)Nc1cccc(CN2CCN(C(=O)OC(C)(C)C)CC2)c1	0.66	ChemMedChem (2016), 11(3), 283-288	0.7213	0.0613
PyBOP	NC=1C=CC=CC1	O=C(O)C1=CC=2C=CC=CC2OS1(=O)=O	O=C(NC=1C=CC=CC1)C2=CC=3C=CC=CC3OS2(=O)=O	0.6	Molecules (2022), 27(13), 4076	0.4068	0.1932
PyBOP	NCCCCN	C(C1=CC=C(/C=C/C(O)=O)C=C1)(C2=CC=C(O)	C(C1=CC=C(/C=C/C(NCCCCNC(/C=C/C2=CC	0.75	European Journal of Medicinal Chemistry (2020),	0.52	0.23

		C=C2)=C3CCCCC3	=C(C(C3=CC=C(O)C=C3)=C4CCCCC4)C=C2)=O)=O)C=C1)(C5=CC=C(O)C=C5)=C6CCCCC6		192, 112191		
PyBOP	Cl.O1B(OC(C)(C)C1(C)C)CN	O=C(O)CC	O=C(NCB1OC(C)(C)C(O1)(C)C)CC	0.42	Journal of Medicinal Chemistry (2019), 62(15), 7160-7184	0.5758	0.1558
PyBOP	O=C(O)C1=C(SC2=C1CCN(C)C2)NC(=O)OC(C)(C)C	NCC=1C=CC=CC1	O=C(OC(C)(C)C)NC=1SC2=C(C1C(=O)NNC=3C=CC=CC3)CCN(C)C2	0.8	Bioorganic & Medicinal Chemistry (2009), 17(20), 7353-7361	0.6909	0.1091
PyBOP	COC(=O)c1c(Cc2ccc(C(=O)O)cc2)c(=O)c2cccnc2n1-c1ccccc1	NC1(CO)CC1	COC(=O)c1c(Cc2ccc(C(=O)NC3(CO)CC3)cc2)c(=O)c2cccnc2n1-c1ccccc1	0.59	ACS Medicinal Chemistry Letters (2012), 3(9), 764-768	0.6337	0.0437
PyBOP	CN(CCCC(=O)O)C	NCCCCO	CN(CCCC(=O)NC	0.6	Bioorganic and	0.6127	0.0127

	[C@H]1O[C@@H](n2cnc3c(N)ncnc32)[C@@H]2OC(C)(C)O[C@@H]21		CCCO)C[C@H]1O[C@@H](n2cnc3c(N)ncnc32)[C@@H]2OC(C)(C)O[C@@H]21		Medicinal Chemistry Letters (2012), 22(1), 278-284		
PyBOP	O=C(O)c1nc2cc(Cl)ccc2[nH]1	N[C@@H](Cc1ccccc1)C(=O)Nc1nccs1	O=C(N[C@@H](Cc1ccccc1)C(=O)Nc1nccs1)c1nc2cc(Cl)ccc2[nH]1	0.52	European Journal of Medicinal Chemistry (2012), 58, 624-639	0.5668	0.0468
PyBOP	O=C(O)/C=C\C(=O)NC1CCCCC1	Clc1ccc(C(c2ccccc2)N2CCCNCC2)cc1	O=C(/C=C\C(=O)N1CCCN(C(c2ccccc2)c2ccc(Cl)cc2)CC1)NC1CCCCC1	0.62	Bioorganic and Medicinal Chemistry (2010), 18(6), 2327-2336	0.646	0.026
PyBOP	O=C(O)C=CC=1C=C(OC)C(O)=C(C1)C=2C=CC=CC2	NCCCCCC	O=C(C=CC=1C=C(OC)C(O)=C(C1)C=2C=CC=CC2)NCCCCCC	0.39	RSC Advances (2015), 5(21), 15800-15811	0.4493	0.0593
PyBOP	O=C(OC(C)(C)C)NCCCCC(	O=C(O)C#C	O=C(C#C)NC(C(=O)NC(C(=O	0.21	Bioconjugate Chemistry	0.3971	0.1871

	N)C(=O)NC(C(=O)NC(C(=O)NC(C(=O)NC(C(=O)O)C)CC=1C=CC=CC1)CC=2C=CC=CC2)C(C)C)CC(C)C		)NC(C(=O)NC(C(=O)NC(C(=O)NC(C(=O)O)C)CC=1C=CC=CC1)CC=2C=CC=CC2)C(C)C)CC(C)C)CCCCN		(2009), 20(11), 2123-2132		
PyBOP	O=C(O)CCC(=O)NC1CCCCC1	Clc1ccc(C(c2ccccc2)N2CCNCC2)cc1	O=C(CCC(=O)N1CCN(C(c2ccccc2)c2ccc(Cl)cc2)CC1)NC1CCCCC1	0.52	Bioorganic and Medicinal Chemistry (2010), 18(6), 2327-2336	0.5768	0.0568
PyBOP	O=C(O)/C=C\C(=O)NC1CCCCC1	Fc1ccc(C(c2ccc(F)cc2)N2CCNCC2)cc1	O=C(/C=C\C(=O)N1CCN(C(c2ccc(F)cc2)c2ccc(F)cc2)CC1)NC1CCCCC1	0.54	Bioorganic and Medicinal Chemistry (2010), 18(6), 2327-2336	0.6145	0.0745
PyBOP	COc1ccc2cc(C(=O)O)c(=O)oc2c1	Nc1ccc(O)cc1	COc1ccc2cc(C(=O)Nc3ccc(O)cc3)c(=O)oc2c1	0.74	Journal of Medicinal Chemistry (2007), 50(24), 6189-6200	0.7627	0.0227
PyBO	ClC1=CC	O=C(O)C	O=C(NC1	0.29	European	0.3698	0.079

P	=CC(=C1)C2=CN=C(N)N2C	1=NC=CN1	=NC=C(C=2C=CC=C(Cl)C2)N1C)C3=NC=CN3		Journal of Medicinal Chemistry (2022), 240, 114577		8
PyBOP	O=C(O)C(F)(F)F.[Se]1[Se]CCNCC1	O=C(OC(C)(C)C)NC(C(=O)O)C	O=C(OC(C)(C)C)NC(C(=O)N1CC[Se][Se]CC1)C	0.8	Organic Letters (2015), 17(14), 3636-3639	0.5688	0.2312
PyBOP	Cl.O=S(=O)(N)C1=CC=C(C=C1)CN	O=C(O)C=1C=CC=C(C1)B(O)O	O=C(NCC1=CC=C(C=C1)S(=O)(=O)N)C=2C=CC=C(C2)B(O)O	0.47	Analytical Chemistry (Washington, DC, United States) (2015), 87(8), 4231-4236	0.4057	0.0643
TBTU	O=C(O)C1=CC=C2C=C(OC)C=CC2=C1	O(C1=CC=C(C=C1OC)C=2C=CC(N)=CC2)C	O=C(NC1=CC=C(C=C1)C2=CC=C(OC)C(OC)=C2)C3=CC=C4C=C(OC)C=CC4=C3	0.36	Bioorganic & Medicinal Chemistry (2012), 20(4), 1557-1568	0.4382	0.0782
TBTU	O=C(O)C=CC(=O)NC1=CC=CC(=C1)C(=O)N	C=1C=CC(=CC1)C2CNCCC2	O=C(N)C=1C=CC=C(C1)NC(=O)C=CC(=O)N2C	0.25	Bioorganic & Medicinal Chemistry Letters	0.3947	0.1447

			CCC(C=3C=CC=CC3)C2		(2017), 27(13), 2907-2911		
TBTU	O=C(O)C=CC1=CC=C2OCOC2=C1	BrC1=CC=C(N)C=C1	O=C(C=CC1=CC=C2OCOC2=C1)NC3=CC=C(Br)C=C3	0.45	Pharmaceutical Chemistry Journal (2018), 51(11), 995-1004	0.6017	0.1517
TBTU	O=C1C=C(C(=O)NCC2=CC=C(C=C2)C(=O)O)C=3C=CC=CC3N1	N=1C=CC(N)=CC1	O=C1C=C(C(=O)NCC2=CC=C(C=C2)C(=O)NC=3C=CN=CC3)C=4C=CC=CC4N1	0.17	Frontiers in Chemistry (Lausanne, Switzerland) (2021), 9, 666122	0.3027	0.1327
TBTU	O=C(O)c1ccc(Cl)cn1	COC(=O)c1ccc(Br)cc1N	COC(=O)c1ccc(Br)cc1NC(=O)c1ccc(Cl)cn1	0.15	European Journal of Medicinal Chemistry (2021), 210, 112958	0.2099	0.0599
TBTU	O=C(O)c1ccco1	COC(=O)c1ccc(Br)cc1N	COC(=O)c1ccc(Br)cc1NC(=O)c1ccco1	0.03	European Journal of Medicinal Chemistry (2021), 210, 112958	0.1212	0.0912
TBTU	O=C(O)c1	Cc1cc(F)c	Cc1cc(F)c	0.273	Bioorgani	0.3632	0.090

	nc(N2CCOCC2)c2ccccc2n1	cc1N	cc1NC(=O)c1nc(N2CCOCC2)c2ccccc2n1		c Chemistry (2020), 105, 104394		2
TBTU	CC1CN(c2nc(C(=O)O)nc3ccccc23)CC(C)O1	COc1ccccc(CN)c1	COc1ccccc(CNC(=O)c2nc(N3CC(C)OC(C)C3)c3ccccc3n2)c1	0.493	Bioorganic Chemistry (2020), 105, 104394	0.5636	0.0706
TBTU	CC1CN(c2nc(C(=O)O)nc3ccccc23)CC(C)O1	Cc1cc(F)ccc1N	Cc1cc(F)ccc1NC(=O)c1nc(N2CC(C)OC(C)C2)c2ccccc2n1	0.301	Bioorganic Chemistry (2020), 105, 104394	0.3018	0.0008
TBTU	CC1CN(c2nc(C(=O)O)nc3ccccc23)CC(C)O1	NCc1ccc(N)cc1	CC1CN(c2nc(C(=O)NCc3ccc(N)cc3)nc3ccccc23)CC(C)O1	0.558	Bioorganic Chemistry (2020), 105, 104394	0.5599	0.0019
TBTU	Cl.O1B(OC(C)(C)C1(C)C)CN	O=C(O)CCC	O=C(NCB1OC(C)(C)C(O1)(C)C)CCC	0.42	Journal of Medicinal Chemistry (2019), 62(15), 7160-7184	0.5126	0.0926
TBTU	CC1CN(c2nc(C(=O)O)nc3ccccc23)CC(C)O1	NCc1cccc(N)c1	CC1CN(c2nc(C(=O)NCc3cccc(N)c3)nc3ccccc23)	0.512	Bioorganic Chemistry (2020), 105,	0.5885	0.0765

			CC(C)O1		104394		
TBTU	O=C(O)c1nc(N2CCOCC2)c2ccccc2n1	CCCCc1ccc(N)cc1	CCCCc1ccc(NC(=O)c2nc(N3CCOCC3)c3ccccc3n2)cc1	0.323	Bioorganic Chemistry (2020), 105, 104394	0.3715	0.0485
TBTU	O=C(O)c1ccccnc1O	CCN(C)C(=O)Nc1ccccc(N)c1	CCN(C)C(=O)Nc1ccccc(NC(=O)c2ccccnc2O)c1	0.09	European Journal of Medicinal Chemistry (2019), 174, 216-225	0.1687	0.0787
TBTU	O=C(O)c1ccccnc1O	CCN(C)C(=O)Oc1ccccc(CN)c1	CCN(C)C(=O)Oc1ccccc(CNC(=O)c2ccccnc2O)c1	0.38	European Journal of Medicinal Chemistry (2019), 174, 216-225	0.4088	0.0288
TBTU	O=C(O)C1CCCC1	Cn1cc(C2CCNCC2)c2cc(NC(=O)c3cc(C#N)ccn3)ccc21	Cn1cc(C2CCN(C(=O)C3CCCC3)CC2)c2cc(NC(=O)c3cc(C#N)ccn3)ccc21	0.04	Journal of Medicinal Chemistry (2018), 61(23), 10415-10439	0.1013	0.0613
TBTU	O=C(O)c1cccc(-c2cnc(-c3c[nH]c4ccccc34)[	Nc1ccccc1	O=C(Nc1ccccc1)c1cccc(-c2cnc(-c3c[nH]c4	0.55	European Journal of Medicinal Chemistry (2017),	0.5628	0.0128

	nH]2)c1		ccccc34)[nH]2)c1		125, 1213-1224		
TBTU	CC(C)CC(=O)O	COc1ccc2nc3cc(Cl)ccc3c(NCCCCN)c2c1	COc1ccc2nc3cc(Cl)ccc3c(NCCCCNC(=O)CC(C)C)c2c1	0.325	ChemMedChem (2015), 10(8), 1344-1349	0.3276	0.0026
TBTU	NS(=O)(=O)c1cccc(-c2n[nH]c3ccc(C(=O)O)cc23)c1	CNCc1ccccc1	CN(Cc1ccccc1)C(=O)c1ccc2[nH]nc(-c3cccc(S(N)(=O)=O)c3)c2c1	0.54	Bioorganic and Medicinal Chemistry (2014), 22(17), 4968-4997	0.6061	0.0661
TBTU	CC(C)c1ccc(C=CC(=O)O)cc1	NCCCCNc1c2ccccc2nc2ccccc12	CC(C)c1ccc(C=CC(=O)NCCCCNc2c3ccccc3nc3ccccc23)cc1	0.22	Bioorganic and Medicinal Chemistry Letters (2013), 23(3), 610-613	0.3072	0.0872
TBTU	N#Cc1ccc2cc(C(=O)O)ccc2c1	COc1cccc(-c2ccc(N)cc2)c1	COc1cccc(-c2ccc(NC(=O)c3ccc4cc(C#N)ccc4c3)cc2)c1	0.41	Bioorganic and Medicinal Chemistry (2012), 20(4), 1557-1568	0.4647	0.0547
TBTU	N#Cc1ccc2cc(C(=O)O)ccc2c1	COc1cc(N)cc(OC)c1	COc1cc(NC(=O)c2ccc3cc(C#N)ccc3c2)cc1	0.6	Bioorganic and Medicinal Chemistry	0.6477	0.0477

			cc(OC)c1		(2012), 20(4), 1557-1568		
TBTU	CC(C)c1ccc(C=CC(=O)O)cc1	NCCCCNc1ccnc2cc(Cl)ccc12	CC(C)c1ccc(C=CC(=O)NCCCCNc2ccnc3cc(Cl)ccc23)cc1	0.28	ChemMedChem (2012), 7(9), 1537-1540	0.3641	0.0841
TBTU	Cc1ccc(C=CC(=O)O)cc1	NCCCCNc1ccnc2cc(Cl)ccc12	Cc1ccc(C=CC(=O)NCCCCNc2ccnc3cc(Cl)ccc23)cc1	0.32	ChemMedChem (2012), 7(9), 1537-1540	0.3622	0.0422
TBTU	O=C(O)c1ccc2cc(Cl)ccc2c1	COC(=O)c1cccc(-c2ccc(N)cc2)c1	COC(=O)c1cccc(-c2ccc(NC(=O)c3ccc4cc(Cl)ccc4c3)cc2)c1	0.22	Bioorganic and Medicinal Chemistry (2012), 20(4), 1557-1568	0.2233	0.0033
TBTU	COc1ccc(C=CC(=O)O)cc1	NCCCCNc1ccnc2cc(Cl)ccc12	COc1ccc(C=CC(=O)NCCCCNc2ccnc3cc(Cl)ccc23)cc1	0.3	ChemMedChem (2012), 7(9), 1537-1540	0.3725	0.0725
TBTU	Cc1nc2c3c(c(C(=O)O)cn2c1C)CCC1(CCc2ccccc21)O3	C1CNC1	Cc1nc2c3c(c(C(=O)N4CCC4)cn2c1C)CCC1(CCc2ccccc21)	0.48	Bioorganic and Medicinal Chemistry (2009), 17(1),	0.4935	0.0135

			O3		368-384		
TBTU	O=C(O)C=CC=1C=CC=CC1	NC=1C=CC=CC1	O=C(C=CC=1C=CC=CC1)NC=2C=CC=CC2	0.52	Pharmaceuticals (2020), 13(7), 141	0.7686	0.2486
TBTU	O=C(O)C1=CC=CNC1=O	O=C(OC1=CC=CC(N)=C1)N(C)CC	O=C(OC1=CC=CC(=C1)NC(=O)C2=CC=CNC2=O)N(C)CC	0.35	European Journal of Medicinal Chemistry (2019), 174, 216-225	0.1141	0.2359
TBTU	O=C(O)C1=NC(Cl)=CC=C1	O=C(OC)C1=CC=C(Br)C=C1N	O=C(OC)C1=CC=C(Br)C=C1NC(=O)C2=NC(Cl)=CC=C2	0.23	European Journal of Medicinal Chemistry (2021), 210, 112958	0.214	0.016

Reaction conditions recommendation for low yield reactions. Our research shown that our BERT yield prediction model embedded with intermediate knowledge had quite good generalization ability on the literature data set, which was involved 94 reactions. However, there are some low yield reactions among them. As a result, we intended to recommend a condition to elevate the yield with the assistance our prediction model. We selected five reactions from above 94 reactions, whose yields were less than 40% and the starting materials are commercially available. Subsequently, we predicted the yield of these reaction under six different conditions with BERT model embedded with intermediate intermediate knowledge and prediction yields were shown in Table S18. Subsequently, we performed the reactions according the conditions with the highest prediction yield. To our pleasure, the yield of four reactions was increased dramatically under recommended conditions (Figure S33), indicating our model had ability to recommend suitable conditions for some low yield reactions. The yield of these reactions was determined by ^{1}H NMR, whose data was consistent with previous work.[1] Besides, all these reaction were detected by UPLC-MS and corresponding signal of mass spectrometry could be found.

Table S18. The prediction yield under six different conditions

Entry	sub1	sub2	product	yield	Prediction_HATU	Prediction_TBTU	Prediction_HBTU	Prediction_EDC	Prediction_PyBOP	Prediction_DCC
1	O=C(O)C1=NC=CC(Cl)=C1	Cl.NC	O=C(NC)C1=NC=CC(Cl)=C1	0.39	0.443	0.3609	0.2435	0.3703	0.3837	0.1872
2	O=C(O)C(=O)CC(C)C	Cl.O=C(OC)CCCN	O=C(NCCCC(=O)OC)C(=O)CC(C)C	0.27	0.3035	0.247	0.2662	0.3951	0.2956	0.1903
3	O=C(OC(C)(C)C)NCC(=O)O	OCCN	O=C(OC(C)(C)C)NCC(=O)NCCO	0.35	0.3428	0.5666	0.557	0.5616	0.4585	0.4585
4	O=C(O)C1=CC=CN1	NC=1C=CC=CC1N	O=C(NC=1C=CC=CC1N)C2=CC=CN2	0.31	0.4356	0.3161	0.7078	0.5415	0.726	0.5772
5	O=C(O)C1=NC(Cl)=CC=C1	O=C(OC)C1=CC=C(Br)C=C1N	O=C(OC)C1=CC=C(Br)C=C1NC(=O)C2=NC(Cl)=CC=C2	0.23	0.32	0.214	0.2293	0.157	0.3216	0.3023

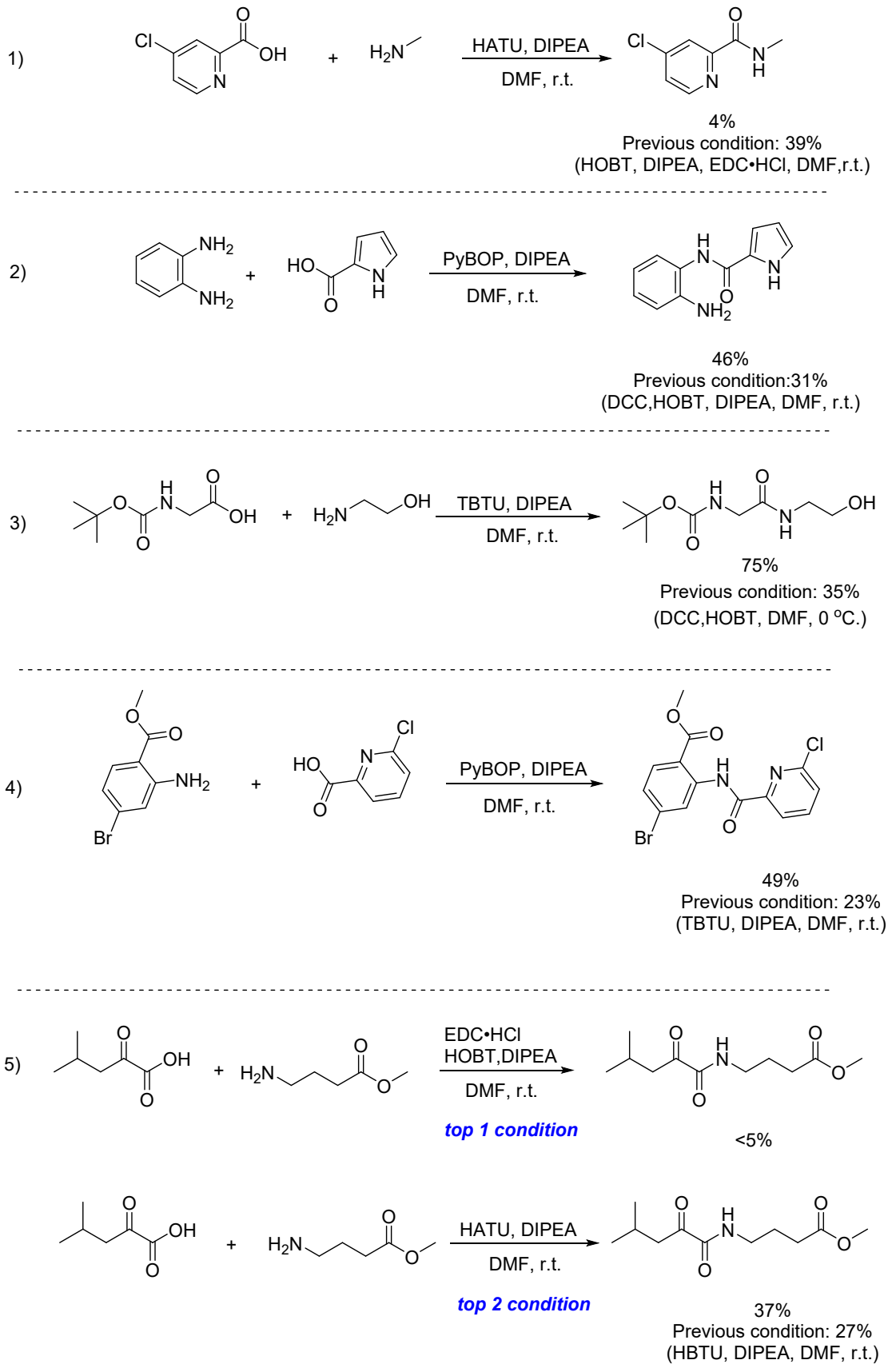


Figure S38. The yields of reactions under recommended conditions

The performance toward reactions with reactivity cliff. We also evaluated the performance of a model toward the reaction with reactivity cliff, which were presented in Isayev's work [2]. We used the data set from Isayev's work to train an embedded BERT model enhanced by intermediate knowledge. The prediction was illustrated in Table S19, the result indicated that the performance of the model in regression was not so good. To our excitement, the model has the ability to identify which reaction could obtain the desired product in a higher yield, achieving an accuracy of 0.73. Besides, the successful prediction examples on binary classification were highlighted in color.

Table S19. The performance toward reactions with reactivity cliff

Entry	Substrate_1	Substrate_2	Product	Yield	Prediction yield
1	CCCCCCCC/C=C\CCCCCCCCCCCC(O)=O	C#CCN	CCCCCCCC/C=C\CCCCCCCCCCC(NCC#C)=O	0.61	0.63
2	OC(CCCCCCCCCCCCCCCCCC)=O	C#CCN	CCCCCCCCCCCCCCCCCCCCCC(NCC#C)=O	0.13	0.57
3	O=C(O)C1=NNC2=C1C=CC=C2	C1C=C(C=CC=1)N	O=C(C1=NNC2=C1C=CC=C2)NC3=CC=CC=C3	0.81	0.65
4	O=C(O)C1=CNC2=CC=CC=C21	C1C=C(C=CC=1)N	O=C(NC1=CC=CC=C1)C2=CNC3=CC=CC=C32	0.13	0.52
5	C1=C(C=CC(C2=C(C(=O)OC)C=C(C=C2)C(N(CC(C)C))=O)C1C(=O)O)CC1=CC=CC=C1	NC1C=CC(=CC=1)C(N)=N	O=C(OC)C(C=C(C(NCC(C)C)=O)C=C1)=C1C2C=CC(CC3=CC=CC=C3)=CC2C(NC4=CC=C(C(N)=N)C=C4)=O	0.2	0.59

6	C1(=C(C2=C(C(=O)OCC3C=CC=CC=3)C=C(C=C2)C(NCC(C)C)=O)C=CC(CC=C)=C1)C(O)=O	C1=CC(=CC=C1C(=N)N)N	O=C(OCC1=CC=CC=C1)C(C=C(C(NCC(C)C)=O)C=C2)=C2C3=C(C(NC4=CC=C(C(N)=N)C=C4)=O)C=C(CC=C)C=C3	0.96	0.62
7	N12C(C3C(N=CC1CC1C=CC=CC2=1)=CC(=C(C=3)OC)OCCCC(O)=O)=O	NC1=CN(C)C(C(NC2=CN(C(C(NCC(C)(SSC)C)=O)=C2)C)=O)=C1	COC1=C(OCCCC(NC2=CN(C)C(C(NC3=CN(C)C(C(NCC(C)(C)SSC)=O)=C3)=O)=C2)=O)C=C4C(C(N5C(CC6=C5C=CC=C6)C=N4)=O)=C1	0.02	0.63
8	N12C(C3C(N=CC1CC1C=CC=CC2=1)=CC(=C(C=3)OC)OCCCC(O)=O)=O	NC1=CN(C)C(C(NC2=CC=C(SC(C(NCC(C)(SSC)C)=O)=C3)C3=C2)=O)=C1	COC1=C(OCCCC(NC2=CN(C)C(C(NC3=CC=C(SC(C(NCC(C)(C)SSC)=O)=C4)C4=C3)=O)=C2)=O)C=C5C(C(N6C(CC7=C6C=CC=C7)C=N5)=O)=C1	0.83	0.51
9	O=[N+]([O-])C1=CC=C(C(O)=O)C=C1C#CC2=CC=C(O)C=C2	CC(C)OC([C@@H](N)CCC1=CC=CC=C1)=O	O=[N+](C1=CC=C(C=C1C#CC2=CC=C(C=C2)O)C(N[C@H](C(OC(C)C)=O)CCC3=CC=CC=C3)=O)[O-]	0.97	0.5
10	O=[N+]([O-])C1=CC=C(C(O)=O)C=C1C#CC2=CC=C(O)C=C2	COC([C@@H](N)CCC1=CC=CC=C1)=O	O=[N+](C1=CC=C(C=C1C#CC2=CC=C(C=C2)O)C(N[C@H](C(OC)=O)CCC3=CC=CC=C3)=O)[O-]	0.17	0.34
11	OC(C1=CC=CO1)=O	FC1=CC=C(N)C=C1C2=NC3=NC=CC=C3O2	O=C(C1=CC=CO1)NC2=CC(C3=NC(N=CC=C4)=C4O3)=C(F)C=C2	0.04	0.4

12	C1=NC=C(C(O)=O)O1	FC1=CC=C(N)C=C1C2=NC3=NC=CC=C3O2	O=C(C1=CN=CO1)NC2=CC(C3=NC(N=CC=C4)=C4O3)=C(F)C=C2	0.93	0.54
13	O=C(O)[C@H](CC1=CC=CC=C1)NC(OCC2=CC=CC=C2)=O	COC(C[C@H](N)CCC1=CC=CC=C1)=O	O=C([C@@H](NC(OCC1=CC=CC=C1)=O)CC2=CC=CC=C2)N[C@@H](C(OC)=O)CCC3=CC=CC=C3	0.23	0.42
14	O=C([C@H](CC1=CC=CC=C1)NC(OCC2=CC=CC=C2)=O)O	N[C@H](C(OC)=O)CC1=CC=CC=C1	O=C([C@@H](NC(OCC1=CC=CC=C1)=O)CC2=CC=CC=C2)N[C@@H](C(OC)=O)CC3=CC=CC=C3	0.93	0.56
15	COc1cc(C(N2[C@@H](CC3=C2C=CC=C3)C=N4)=O)c4cc1OCCCC(O)=O	NC1=CN(C)C(C(NC2=CC=C(C3=CC(C(NCC(C)(C)SSC)=O)N(C)C3)C=C2)=O)=N1	COC1=C(OCCCC(NC2=CN(C)C(C(NC3=CC=C(C4=CN(C)C(C(NCC(C)(C)SSC)=O)=C4)C=C3)=O)=N2)=O)C=C5C(C(N6C(CC7=C6C=CC=C7)C=N5)=O)=C1	0.29	0.41
16	COc1cc(C(N2[C@@H](CC3=C2C=CC=C3)C=N4)=O)c4cc1OCCCC(O)=O	NC1=CN(C(C(NC2=CC=C(C=C2)C3=CN(C)C(C(NCC(C)(SSC)C)=O)=C3)=O)=C1)C	COC1=C(OCCCC(NC2=CN(C)C(C(NC3=CC=C(C4=CN(C)C(C(NCC(C)(C)SSC)=O)=C4)C=C3)=O)=C2)=O)C=C5C(C(N6C(CC7=C6C=CC=C7)C=N5)=O)=C1	1	0.59
17	O=C(O)C1=NNC2=CC=CC=C12	C1C=C(C(=CC=1)C(OC)=O)N	O=C(C1=NNC2=CC=CC=C12)NC3=C(C(OC)=O)C=CC=C3	0.06	0.62
18	O=C(O)C1=NNC2=CC=CC=C12	C1C=C(OC)C(=CC=1)N	O=C(C1=NNC2=CC=CC=C12)NC3=C(OC)C=CC=C3	0.82	0.42

19	C1(C=CC(/C=C/C(O)=O)=CC=1)O	ClC1=CC=C(CCCCN)C=C1	O=C(NCCCCC1=CC=C(Cl)C=C1)/C=C/C2=CC=C(O)C=C2	0.09	0.63
20	OC1=CC=C(/C=C/C(O)=O)C=C1	C1(C=CC(CCN)=CC=1)Cl	OC1=CC=C(C=C1)/C=C/C(NCCC2=CC=C(Cl)C=C2)=O	0.96	0.58
21	CC(C)(C)C1=CC=C(C(O)=O)C=C1	NC1=CC=C(C=CN2CC3=CC=C(C(OC)=O)C=C3)C2=C1	CC(C)(C)C1=CC=C(C(NC2=CC(N(CC3=CC=C(C(OC)=O)C=C3OC)C=C4)=C4C=C2)=O)C=C1	0.42	0.45
22	CC(C)(C)C1=CC=C(C(O)=O)C=C1	NC1=CC=C(C=CN2CC3=CC=C(C(OC)=O)C=C3OC)C2=C1	CC(C)(C)C1=CC=C(C(NC2=CC(N(CC3=CC=C(C(OC)=O)C=C3)C=C4)=C4C=C2)=O)C=C1	0.97	0.58

S3.12 Transfer learning among reaction conditions

We acknowledge the potential use of transfer learning to reduce the amount of data required for generalization to new reaction conditions. Transfer learning can be a powerful approach for leveraging knowledge from reactions with similar intermediates. We explored transfer learning using pretrained models across reaction conditions, as shown throughout the manuscript and the SI. The results in Table S demonstrate the outcomes of transfer learning in three scenarios:

1. **Pretrained EDC transferred to DCC:**
   The model pretrained on EDC conditions was fine-tuned for DCC reactions, yielding a Mean Absolute Error (MAE) of 0.1535, Root Mean Squared Error (RMSE) of 0.1987, and R^2 of 0.5253.

2. **Pretrained HATU transferred to DCC:**
   When the model was pretrained on HATU conditions and then fine-tuned on DCC, the performance improved slightly compared to EDC-to-DCC transfer, achieving an MAE of 0.145, RMSE of 0.2039, and R^2 of 0.5002.

3. **Pretrained HATU transferred to EDC:**
   Transferring the pretrained HATU model to EDC conditions resulted in an MAE of 0.1776, RMSE of 0.2217, and R^2 of 0.3053.

For comparison, we also included results for models trained directly on DCC and EDC conditions without any transfer learning:

- **DCC itself – no transfer:** MAE of 0.07, RMSE of 0.05, and R^2 of 0.67.
- **EDC itself – no transfer:** MAE of 0.14, RMSE of 0.11, and R^2 of 0.75.

These results indicate that while transfer learning improves generalization to new conditions compared to training from scratch on smaller datasets, the performance is still not as strong as models trained directly on large datasets for the specific reaction conditions. However, transfer learning remains a promising approach when data availability is limited, particularly for reactions with shared intermediate structures.

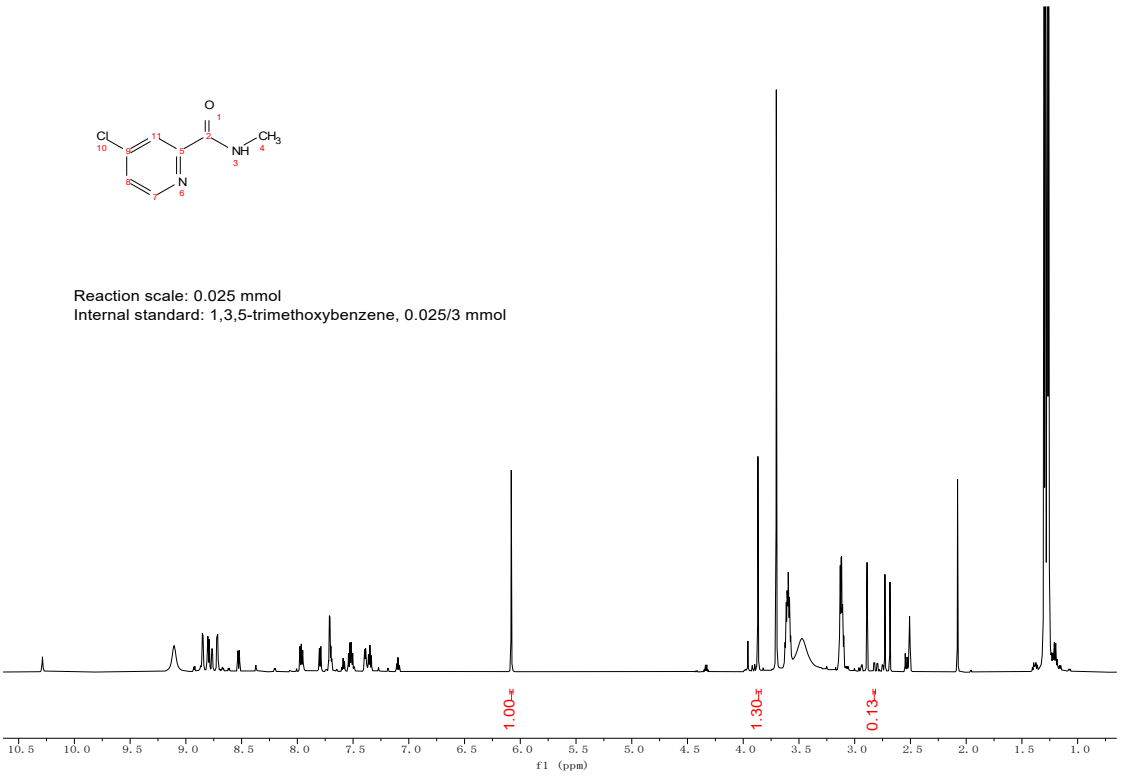
Table S20. The comparison of BERT model performance before and after reaction conditions

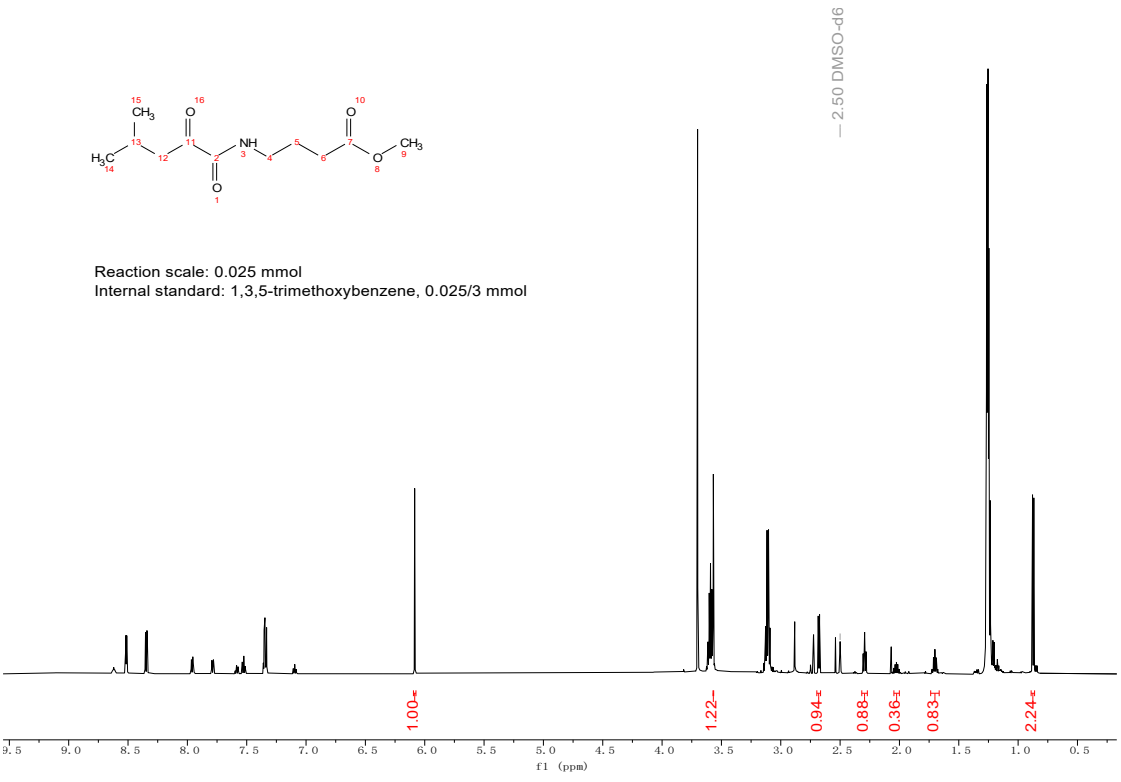
	MAE	**RMSE**	**R2**
Pretrained EDC - transferred to DCC	0.15	0.19	0.52
Pretrained HATU - transferred to DCC	0.15	0.20	0.50
Pretrained HATU- transferred to EDC	0.18	0.22	0.31
DCC itself – no transferred knowledge	**0.07**	**0.05**	**0.67**
EDC itself – no transferred knowledge	**0.14**	**0.11**	**0.75**

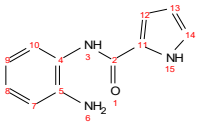
S4 Reference

1) a) Zengin K. B.; Ozturk C. D.; Cakmak, E. B.; Kolcuoglu, Y.; Senol, H.; Saglk O. B. N.; Dag, A.; Benkli, K. *Journal of Medicinal Chemistry* **2024**, *67*, 4463-4482; b) McClure, J. J.; Inks, E. S.; Zhang, C.; Peterson, Y. K.; Li, J.; Chundru, K.; Lee, B.; Buchanan, A.; Miao, S.; Chou, C. J. *ACS Chemical Biology* **2017**, *12*, 1644-1655; c) Antoni, F.; Wifling, D.; Bernhardt, G. European *Journal of Medicinal Chemistry* **2021**, *210*, 112958; d) Piasecki, S. K.; Zheng, J.; Axelrod, A. J.; Detelich, M. E.; Keatinge-Clay, A. T. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 2067-2077; e) Kuo, C.-H.; Hsieh, W.-T.; Yang, Y.-H.; Hwang, T.-L. ; Cheng, Y.-S.; Lin, Y. A *Journal of Organic Chemistry* **2024**, *89*, 4958-4970.
2) Liu, Z.; Moroz, Y. S.; Isayev, O. *Chemical. Science.* **2023**, *14*, 10835–10846.

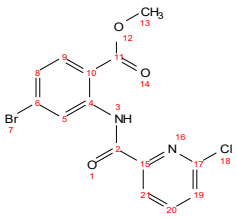
S5 Spectrum of crude reactions



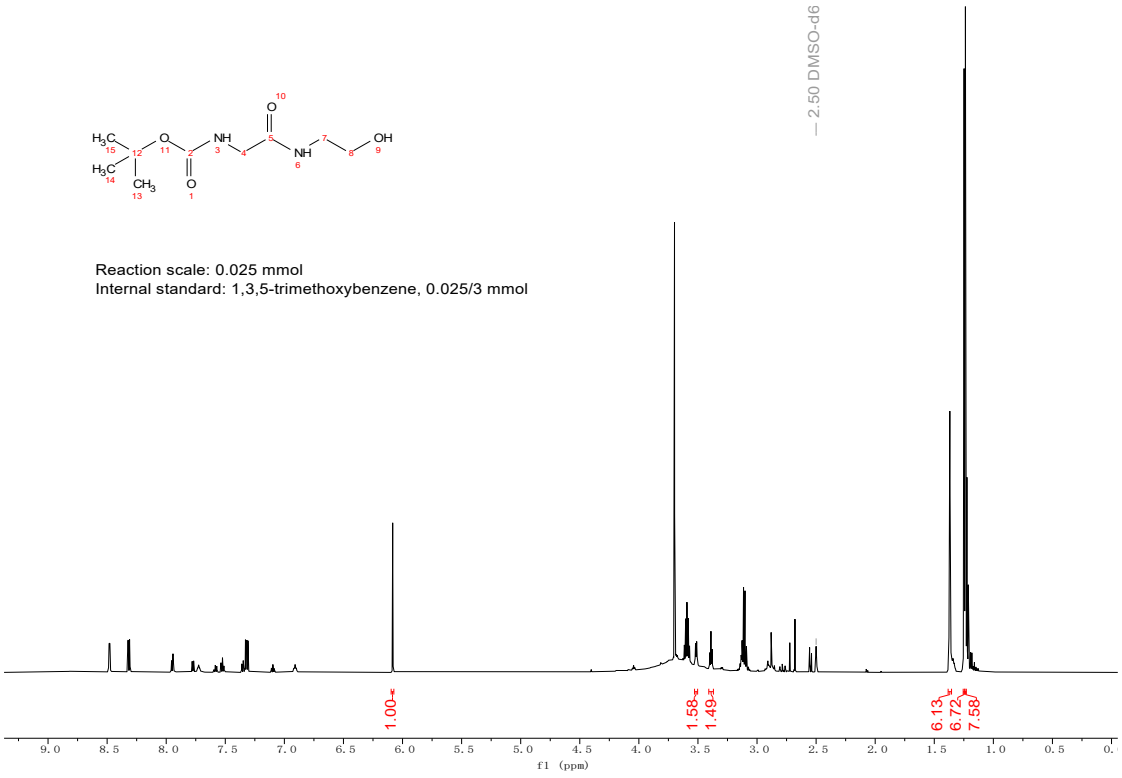




Reaction scale: 0.025 mmol
Internal standard: 1,3,5-trimethoxybenzene, 0.025/3 mmol



Reaction scale: 0.025 mmol
Internal standard: 1,3,5-trimethoxybenzene, 0.025/3 mmol


Reaction scale: 0.025 mmol
Internal standard: 1,3,5-trimethoxybenzene, 0.025/3 mmol
2.50 DMSO-d6
1.00
1.58
1.49
6.13
6.72
7.58
f1 (ppm)