

Supporting Information

The Challenge of Balancing Model Sensitivity and Robustness in Predicting Yields: A Benchmarking Study of Amide Coupling Reactions

Zhen Liu¹, Yurii S. Moroz^{2,3,4}, Olexandr Isayev^{1*}

¹Department of Chemistry, Mellon College of Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

²*Enamine Ltd, Kyiv, 02660, Ukraine*

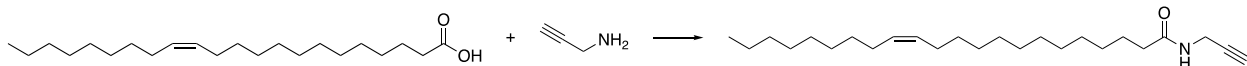
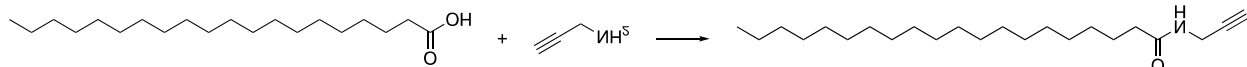
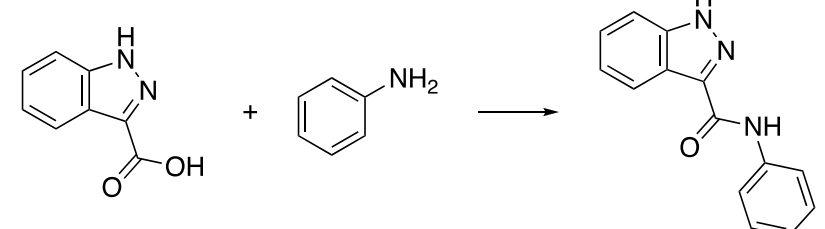
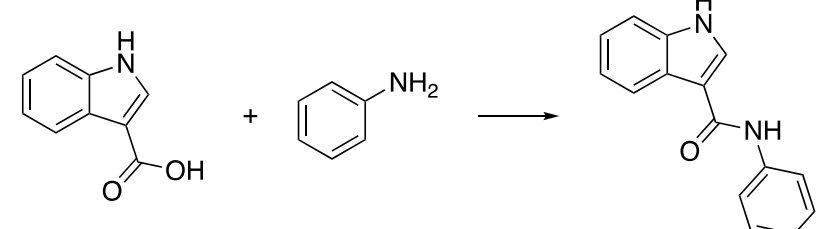
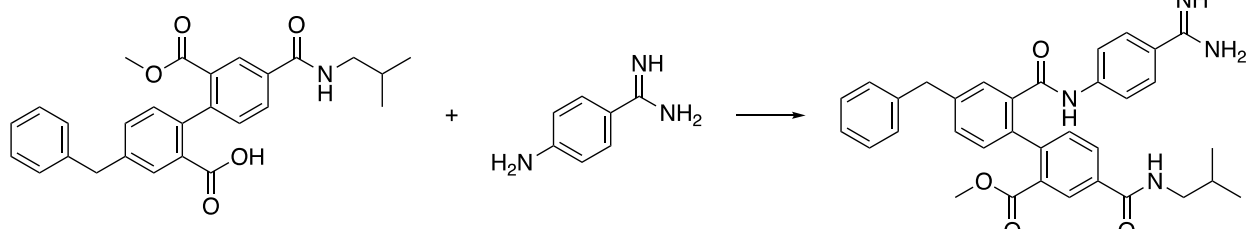
³*Chemspace LLC, Kyiv, 02094, Ukraine*

⁴*Taras Shevchenko National University of Kyiv, Kyiv, 01601, Ukraine*

* Correspondence: olexandr@olexandrisayev.com (O.I.)

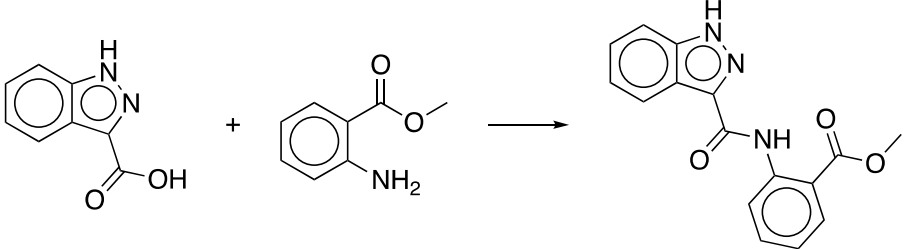
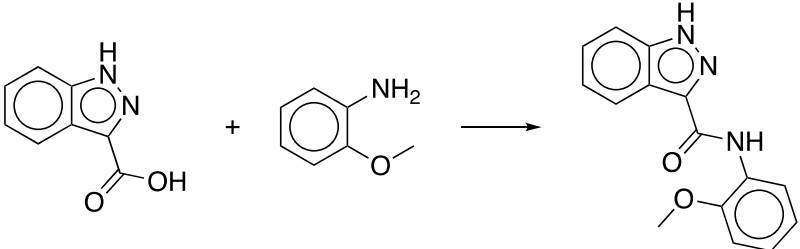
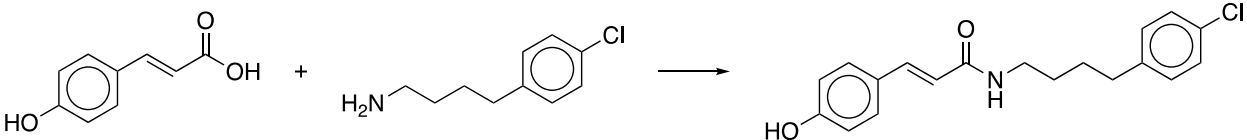
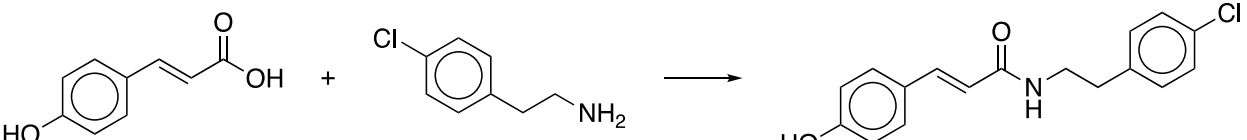
Examples of reactivity cliffs

The similarity of reactions was determined using the reaction fingerprint, and two reactions were deemed similar if their cosine similarity exceeded 0.90. Reactions were considered "cliffs" when their similarity surpassed 0.9, yet the yield difference was greater than 30. In our amide coupling datasets, we identified 6,365 reactivity cliffs. These examples serve to illustrate that even a minor alteration in molecular structure can result in a substantial variation in reaction yield. This observation underscores that the yield prediction surface is non-smooth, highlighting the challenges for accurately predicting reaction yields.

Reaxys ID	Reaction	Yield (%)	Condition
37377749		61	18 °C, 20 h; DMAP, DIC, DCM
37377774		13	18 °C, 20 h; DMAP, DIC, DCM
28563907		81	100 h; benzotriazol-1-ol, EDC, DMF
42264425		13	10 °C, 100 h; benzotriazol-1-ol, EDC, DMF
28683797		20	16.0 °C, 20 h; Inert atmosphere, pyridine, DCC, DMF

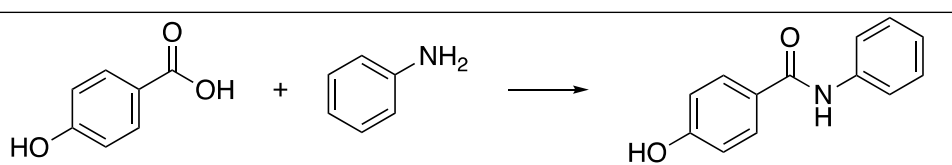
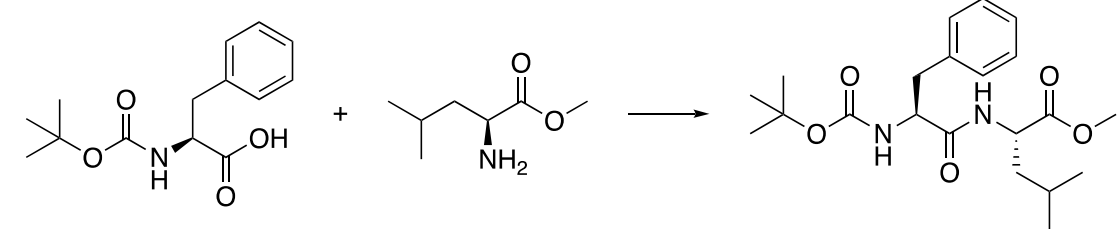
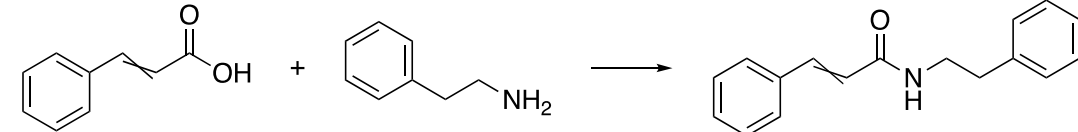
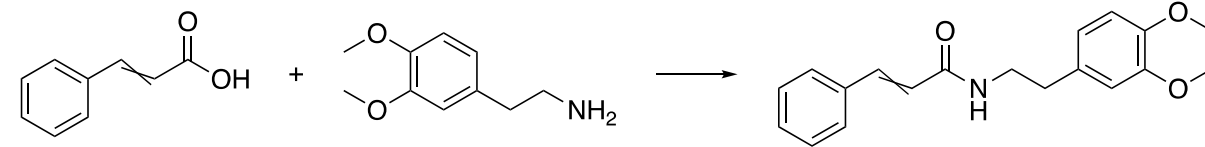
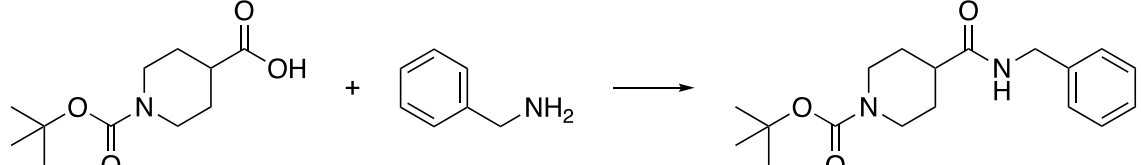
28683737		96	16 °C, 20 h; Inert atmosphere, pyridine, DCC, DMF
51739779		2	1 °C, 25 h; DMAP, EDC, DCM
52011231		83	2 °C, 20 h; DMAP, EDC, DCM
42724738		97	25 h; 2,6-dimethylpyridine; 1-hydroxy-7-aza- benzotriazole; EDC, DMF
42724744		17	25 h; 2,6-dimethylpyridine; 1-hydroxy-7-aza- benzotriazole; EDC, DMF
36270730		4	45 h; Inert atmosphere, DMAP; EDC, DCM

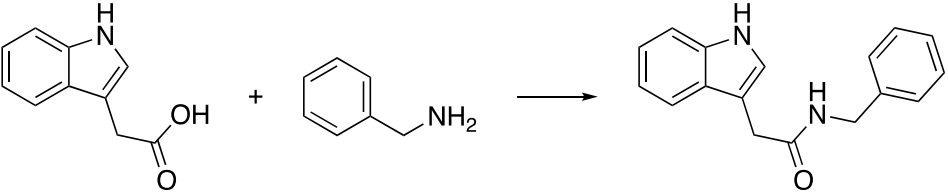
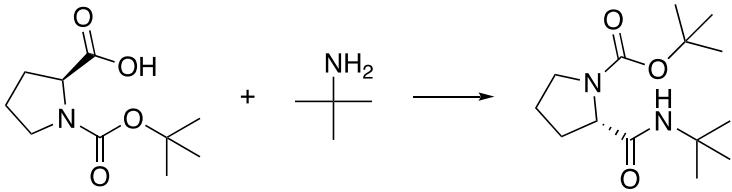
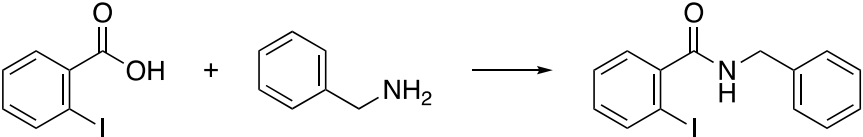
36270732		93	45 h; Inert atmosphere, DMAP; EDC, DCM
9190843		0.23	benzotriazol-1-ol; EDC, DCM
2107062		93	20 h; 4-methyl-morpholine; benzotriazol-1-ol; EDC, DMF
51739741		29	2 °C, 25 h; DMAP; EDC, DCM
51739653		100	2 °C, 20 h; DMAP; EDC, DCM

42264445		6	<p>10 °C, 100 h; benzotriazol-1-ol; EDC, DMF</p>
42264442		82	<p>10 °C, 100 h; benzotriazol-1-ol; EDC, DMF</p>
5267967		9	<p>12 °C, 80 h; benzotriazol-1-ol; DCC, DMF</p>
4832409		96	<p>12 °C, 80 h; benzotriazol-1-ol; DCC, DMF</p>

Examples of uncertain reactions

The reactions were deemed to have "uncertain" yields when the carboxylic acid, amine, and product were the same, yet the yields (obtained from the Reaxys database) differed by at least 30. A group of uncertain reactions share the same catalyst, despite slight variations in the experimental context. Within our amide coupling dataset, we identified 649 reactions with uncertain yields. These instances emphasize that reproducibility poses an additional challenge when constructing accurate yield prediction models.

Reaxys ID	Reaction	Yield records in the Reaxys database (%)
712688		29, 71
1889311		66, 70, 85, 87,
5211116		84, 96
44962717		32.5, 97.5
28286573		23, 62, 62, 99

1789336	 <p>Indole-3-acetic acid + benzylamine → N-benzyl-L-tryptophan</p>	35, 85, 90
1820687	 <p>Cyclic amide with tert-butyl ester + primary amine → bicyclic urea derivative</p>	40, 40, 93
3747743	 <p>2-iodobenzoic acid + benzylamine → N-benzyl-2-iodobenzamide</p>	60, 60, 99