Supporting Information for Neural networks for the prediction of organic chemistry reactions

Jennifer N. Wei,[†] David Duvenaud,[‡] and Alán Aspuru-Guzik[*,†]

Department of Chemistry and Chemical Biology, Harvard University, Cambridge MA 02138, USA, and Department of Computer Science, Harvard University, Cambridge MA 02138, USA

E-mail: aspuru@chemistry.harvard.edu

Table 1 shows the Tanimoto similarity of the training set to problems Wade 8-47 and 8-48 used in the main article. The code and all data used in our paper can all be found at `https://github.com/jnwei/neural_reaction_fingerprint.git`

[*]To whom correspondence should be addressed
[†]Department of Chemistry and Chemical Biology, Harvard University, Cambridge MA 02138, USA
[‡]Department of Computer Science, Harvard University, Cambridge MA 02138, USA

Supporting Information Table 1: Similarity between selected textbook questions and training set questions.

Problem number	Average Training Set Tanimoto Similarity	Highest Training Set Tanimoto Similarity
8-47a	0.30	0.94
8-47b	0.42	0.74
8-47c	0.47	0.86
8-47d	0.41	0.76
8-47e	0.47	0.88
8-47f	0.47	0.88
8-47g	0.35	0.65
8-47h	0.42	0.75
8-47i	0.43	0.80
8-47j	0.48	0.76
8-47l	0.44	0.77
8-47m	0.43	0.82
8-47n	0.45	0.75
8-47o	0.44	0.75
8-47p	0.44	0.76
8-48a	0.42	0.78
8-48b	0.42	0.77
8-48c	0.31	1.00
8-48d	0.34	0.54
8-48e	0.19	0.71
8-48f	0.20	0.66
8-48g	0.33	0.48