Supporting Information

Retrosynthetic reaction prediction using neural sequence-to-sequence models

Bowen Liu[1], Bharath Ramsundar[2], Prasad Kawthekar[2], Jade Shi[1], Joseph Gomes[1], Quang Luu Nguyen[1], Stephen Ho[1], Jack Sloane[1], Paul Wender[1,3], Vijay Pande[1,2,4]

[1] Department of Chemistry, Stanford University, Stanford, CA 94305, USA

[2] Department of Computer Science, Stanford University, Stanford, CA 94305, USA

[3] Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA

[4] Department of Structural Biology, Stanford University, Stanford, CA 94305, USA

Email: pande@stanford.edu

The neural network model, processed datasets and evaluation code will be made available at: https://github.com/pandegroup/reaction_prediction_seq2seq.git

Table S1: Key hyperparameters of the seq2seq model

General	
Batch size	32
Optimizer	Adam
Learning rate	0.0001
Max gradient norm	5.0
Max sequence length	140
Attention dim	512
Bidirectional LSTM encoder	
# layers	2
Embedding dim	512
Dropout (keep)	0.8
LSTM decoder	
# layers	4
Embedding dim	512
Dropout (keep)	0.8
Max decode length	140

Table S2: Beam search statistics. A completed candidate sequence contains an end of sequence character. The seq2seq beam search inference was performed on a system containing an Intel i7-7700K CPU and a NVIDIA GTX 1080 Ti GPU, using the GPU

	beam size				
	3	5	10	20	50
average # of complete candidates	2.93	4.84	9.56	19.26	48.34
# examples with all candidates complete	4700	4330	3679	3511	3314
# examples with no candidates complete	5	3	3	0	0
inference time gpu (s)	443	465	487	691	712

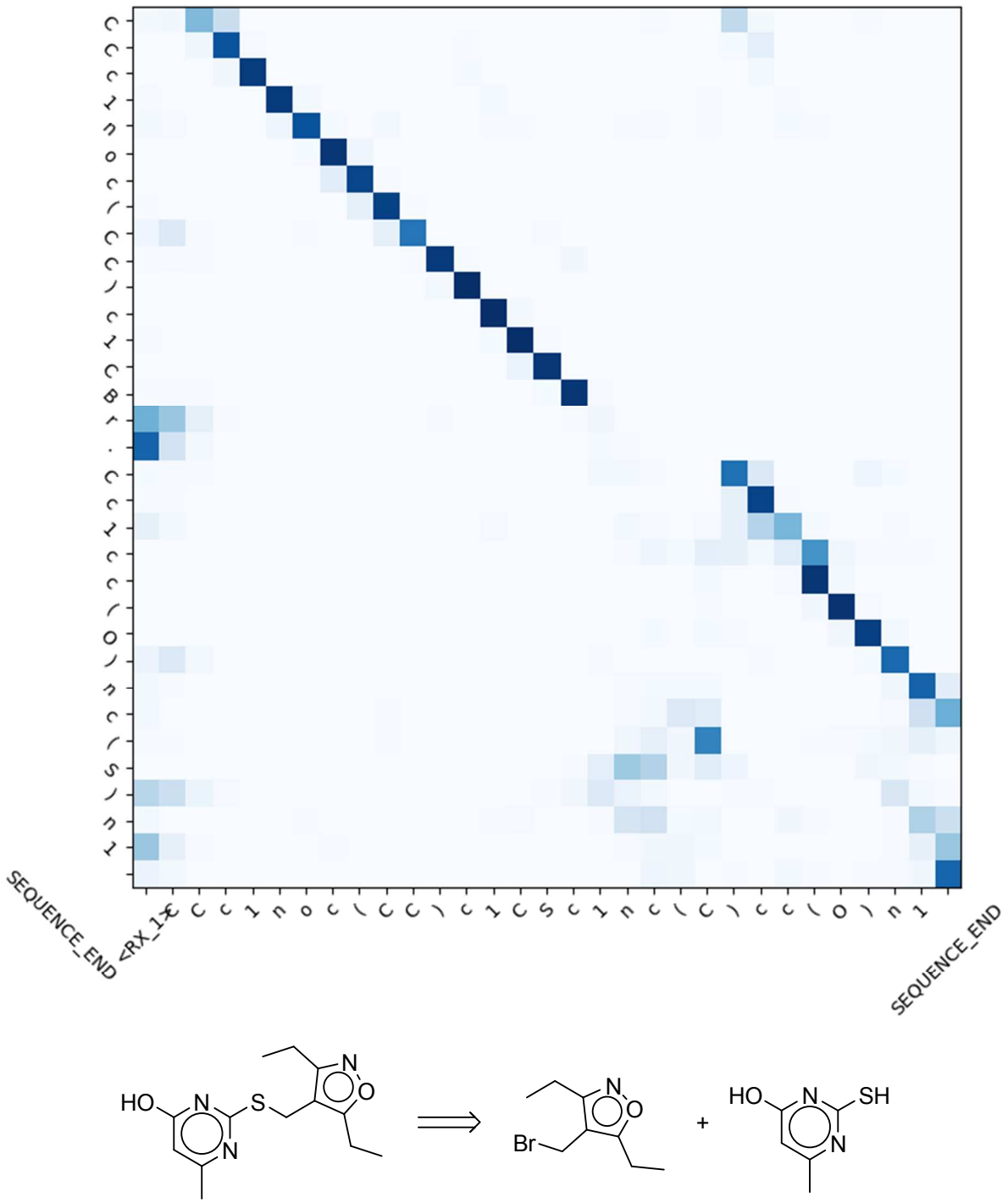


Figure S1: Attention weights for a reaction class 1 (Heteroatom alkylation and arylation) example. Note that for Figures S1-S10, darker blue represents larger attention weights. The input sequence characters (reaction type and target molecule) are along the x axis, and the output sequence characters (predicted reactants) are along the y axis

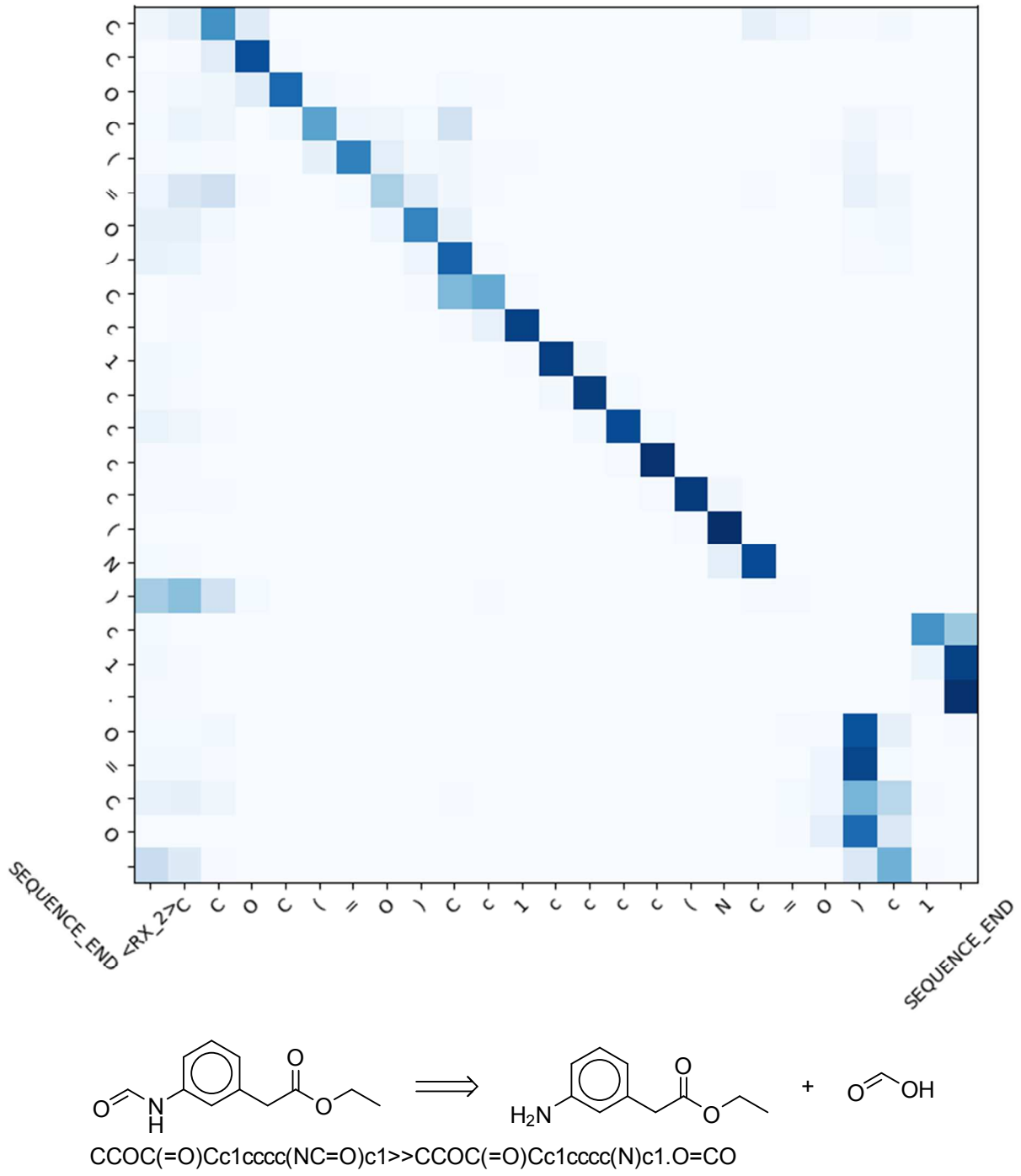
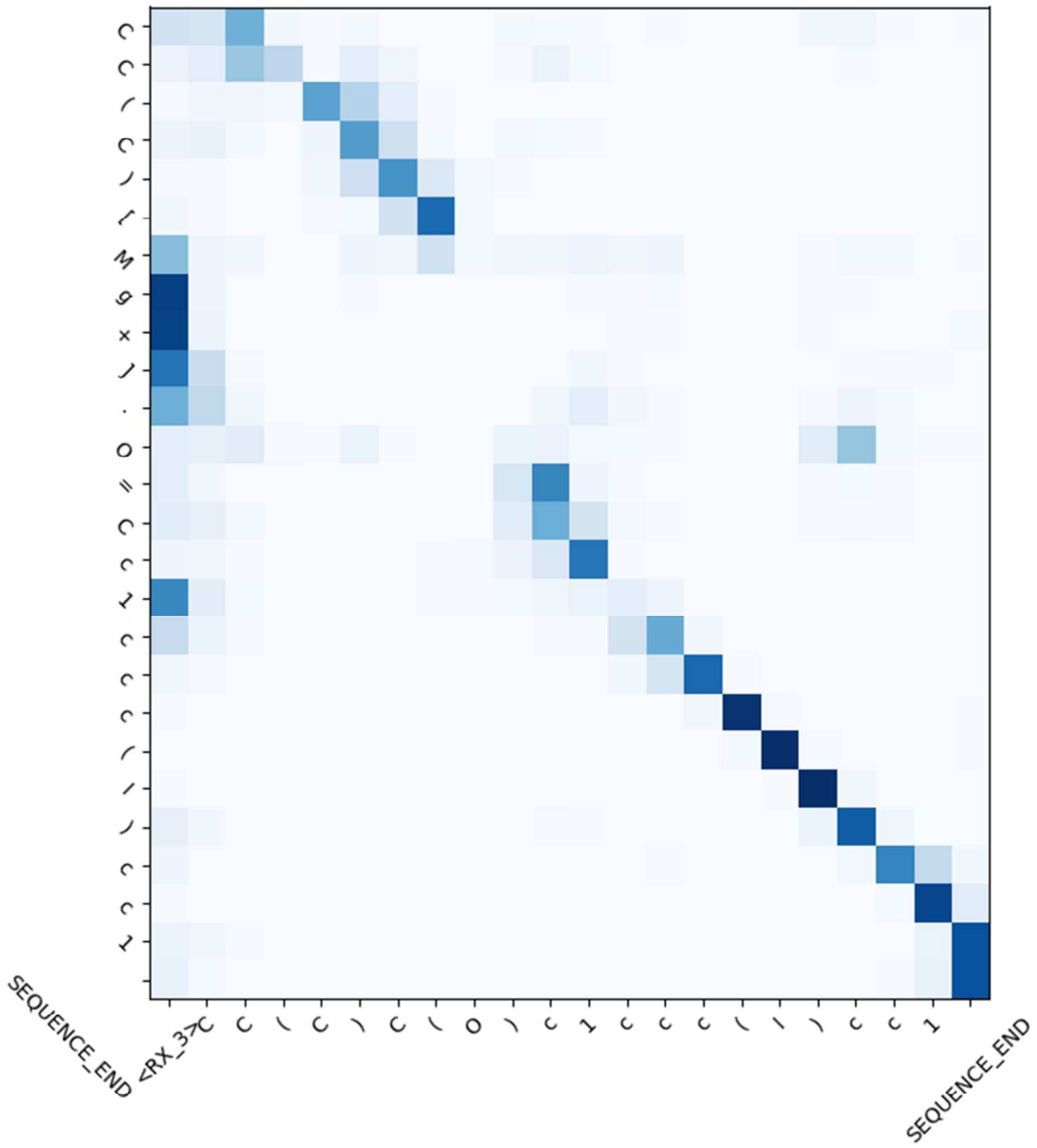
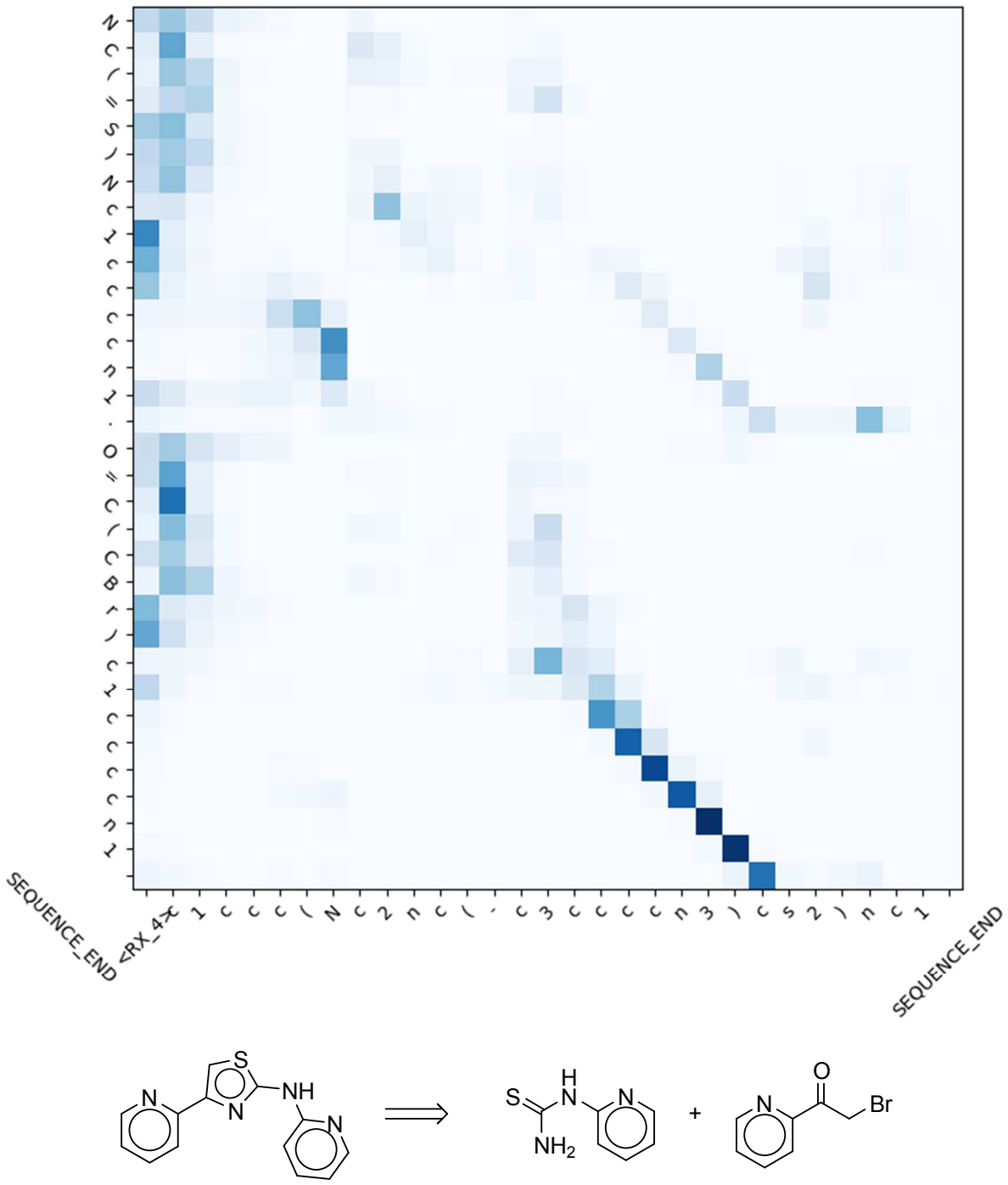


Figure S2: Attention weights for a reaction class 2 (Acylation and related processes) example



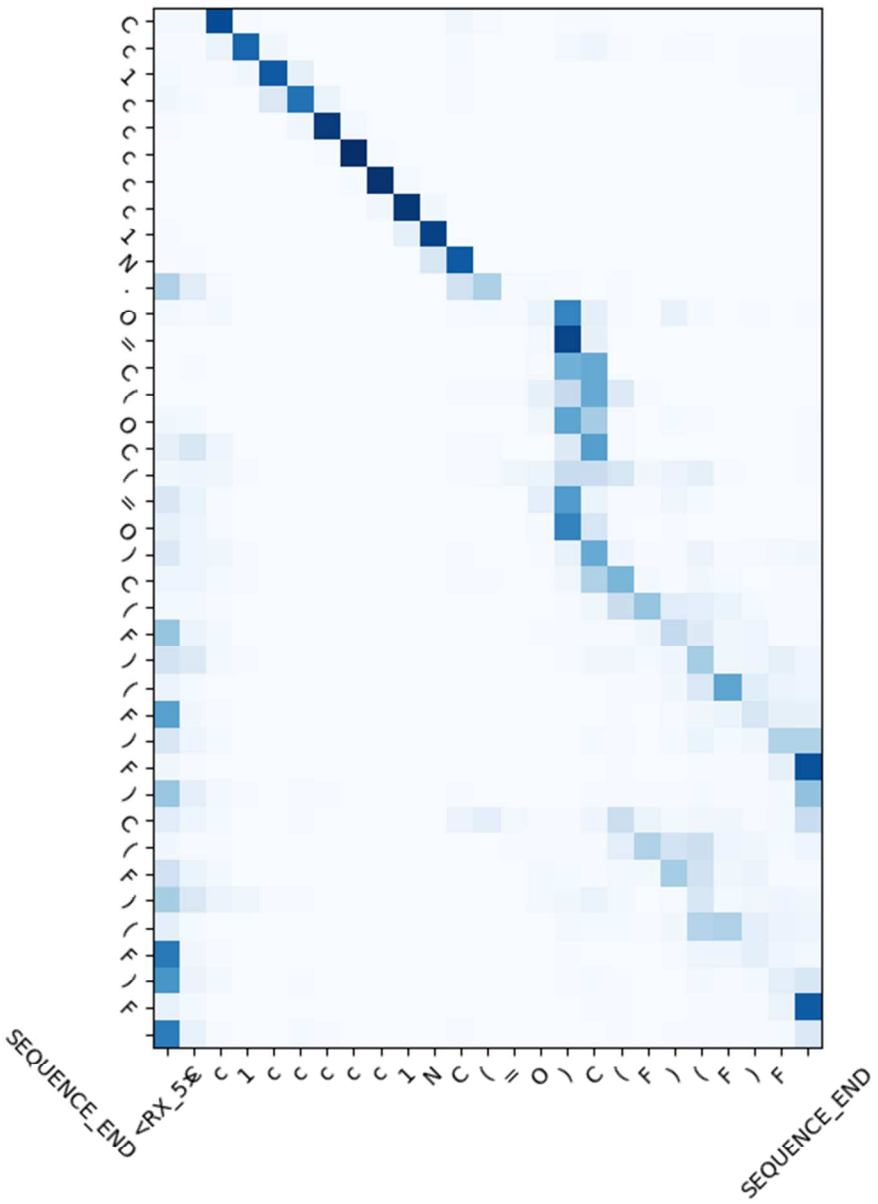
CC(C)C(O)c1ccc(I)cc1>>CC(C)[Mg+].O=Cc1ccc(I)cc1

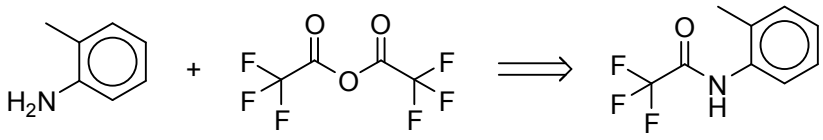
Figure S3: Attention weights for a reaction class 3 (C-C bond formation) example



c1ccc(Nc2nc(-c3ccccn3)cs2)nc1>>NC(=S)Nc1ccccn1.O=C(CBr)c1ccccn1

Figure S4: Attention weights for a reaction class 4 (Heterocycle formation) example





Cc1ccccc1N.O=C(OC(=O)C(F)(F)F)C(F)(F)F>>Cc1ccccc1NC(=O)C(F)(F)F

Figure S5: Attention weights for a reaction class 5 (Protections) example

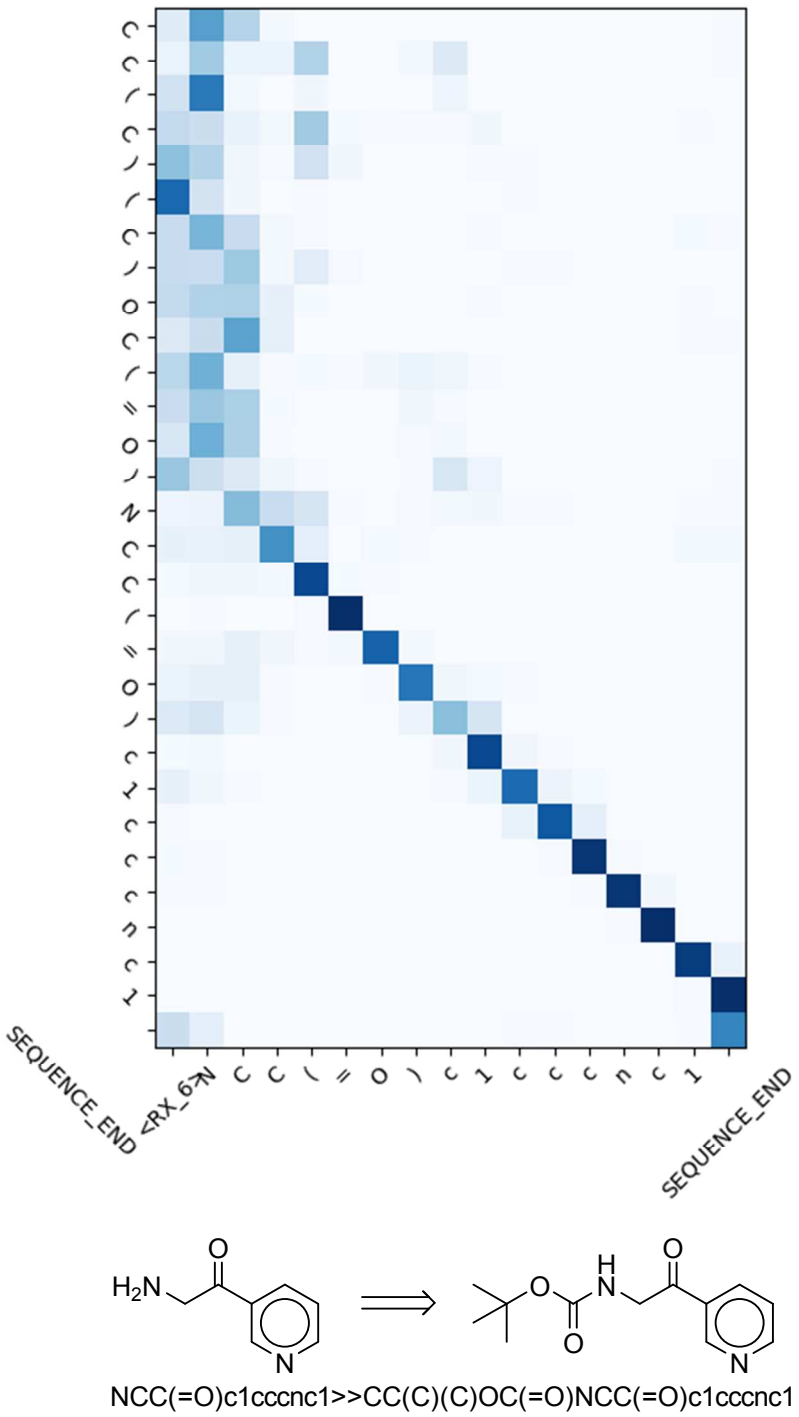
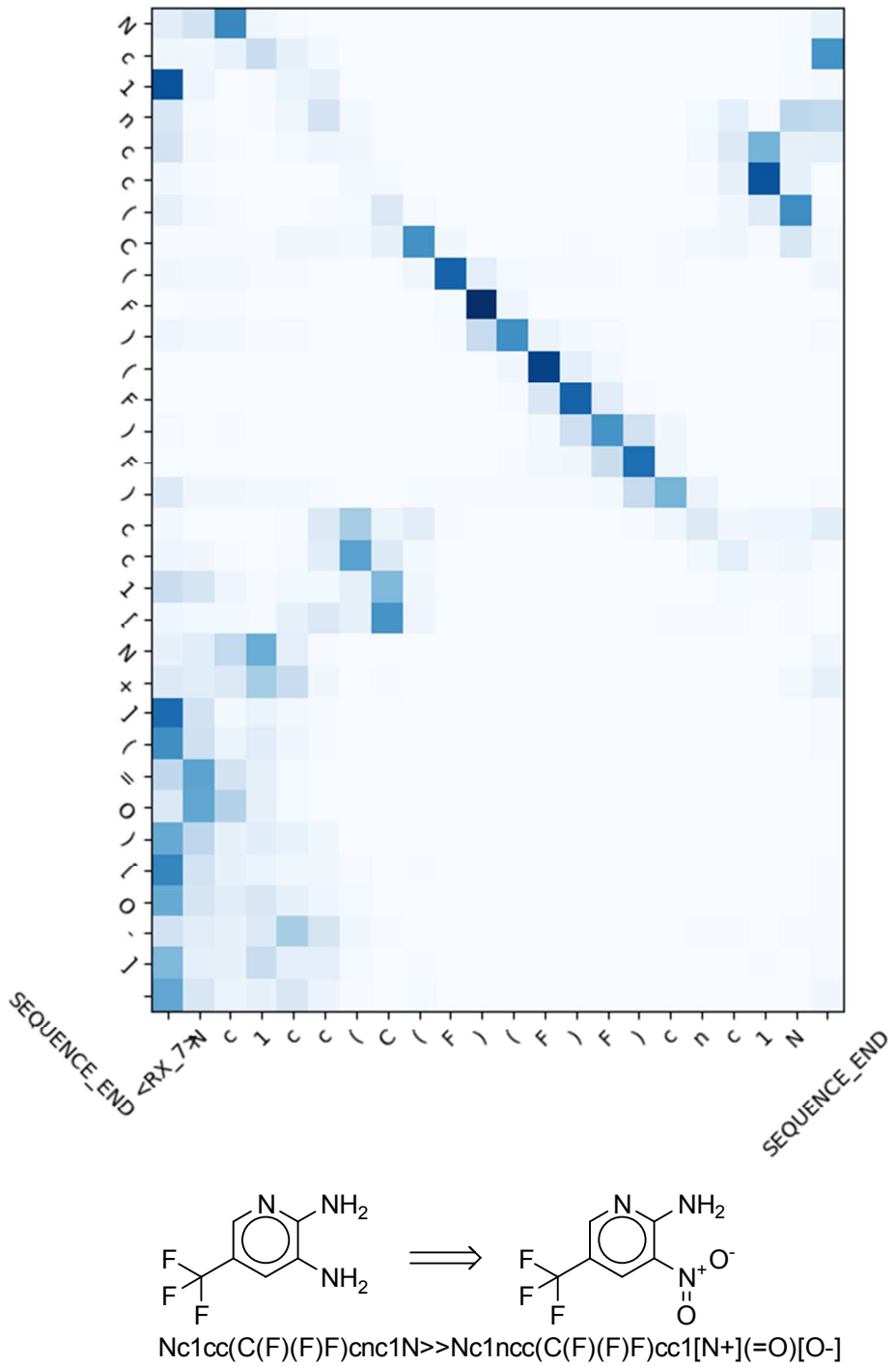
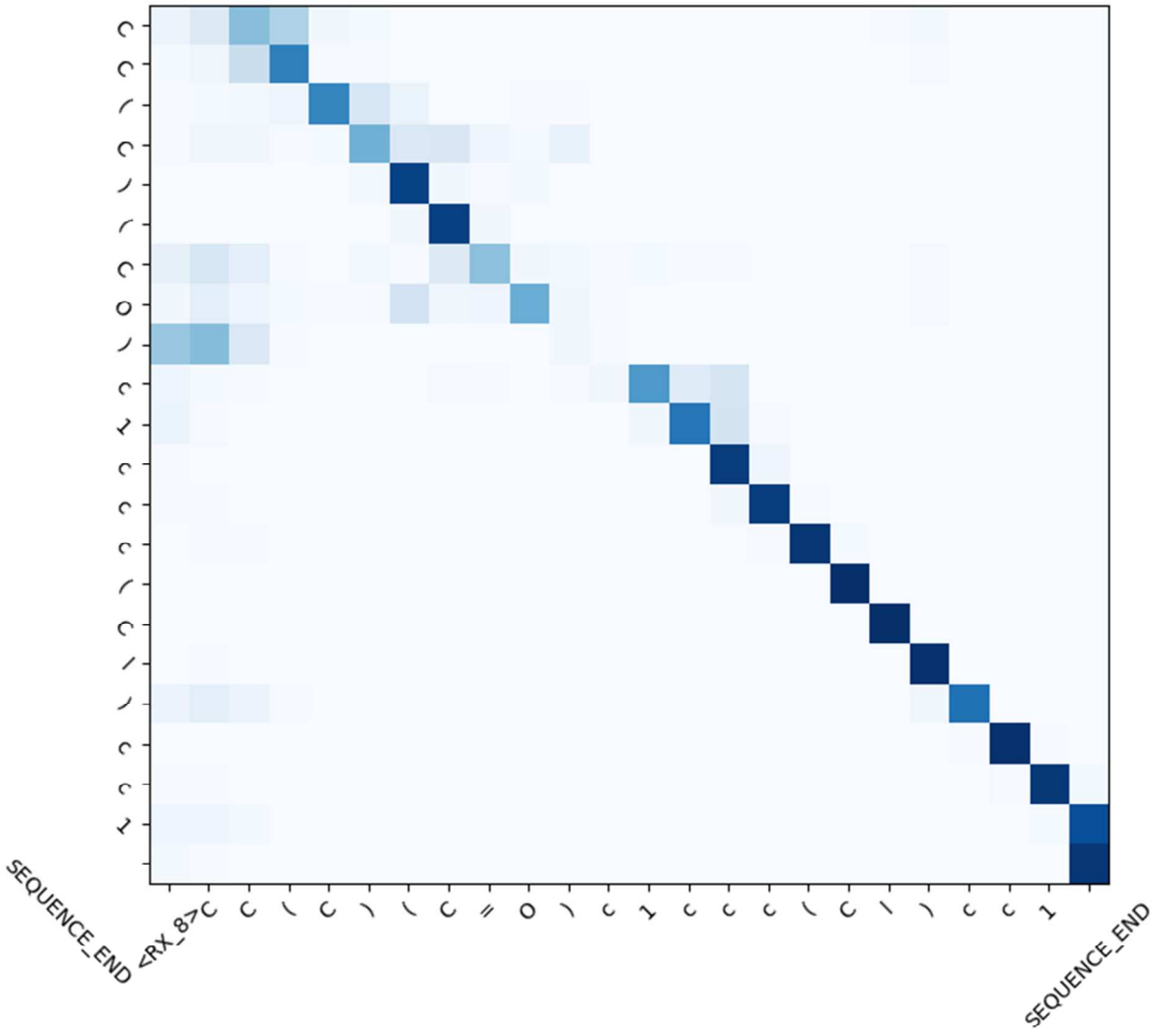


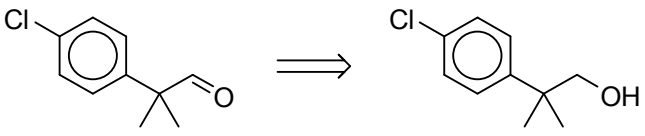
Figure S6: Attention weights for a reaction class 6 (Deprotections) example



Nc1cc(C(F)(F)F)cnc1N>>Nc1ncc(C(F)(F)F)cc1[N+](=O)[O-]

Figure S7: Attention weights for a reaction class 7 (Reductions) example





CC(C)(C=O)c1ccc(Cl)cc1>>CC(C)(CO)c1ccc(Cl)cc1

Figure S8: Attention weights for a reaction class 8 (Oxidations) example

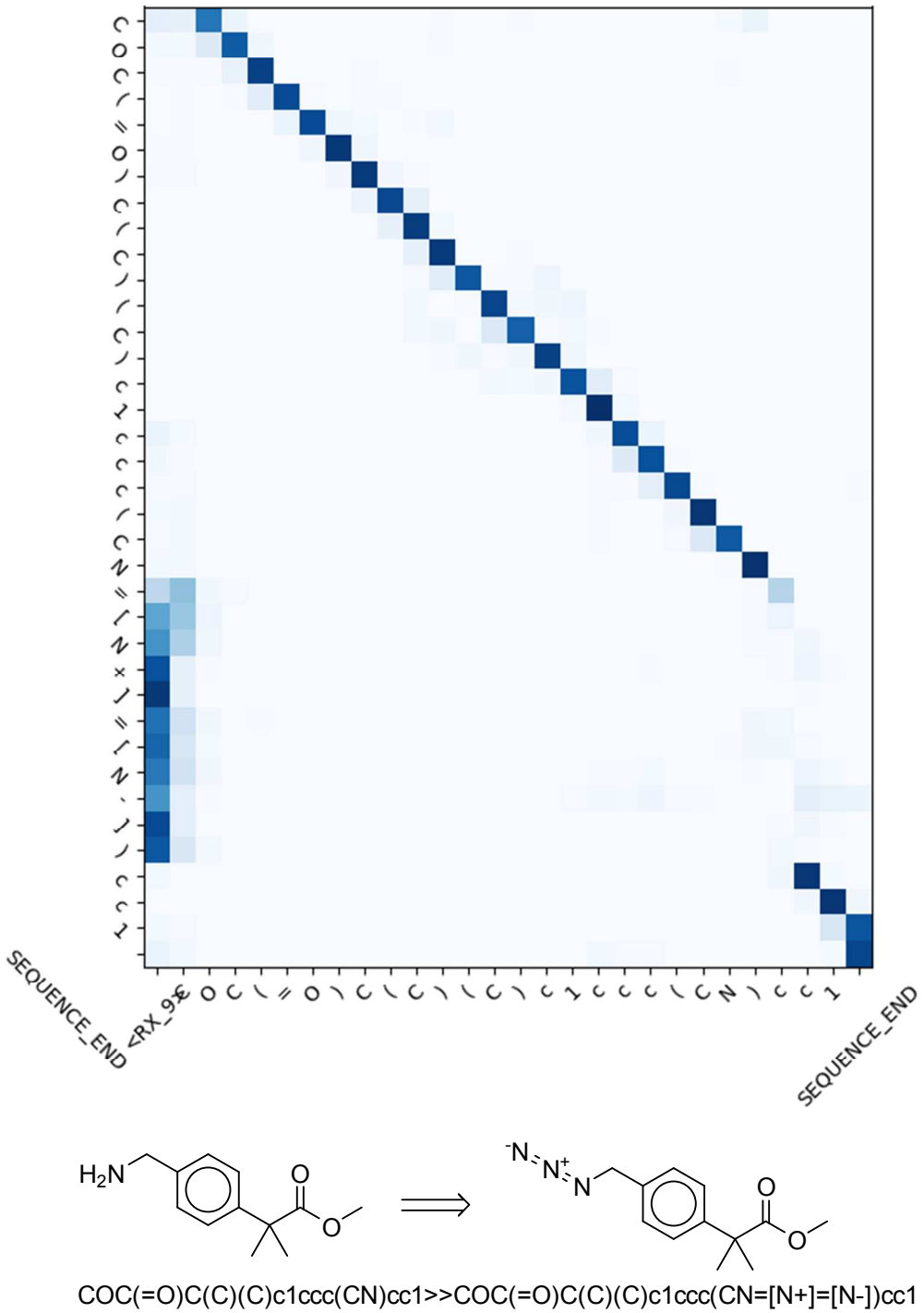
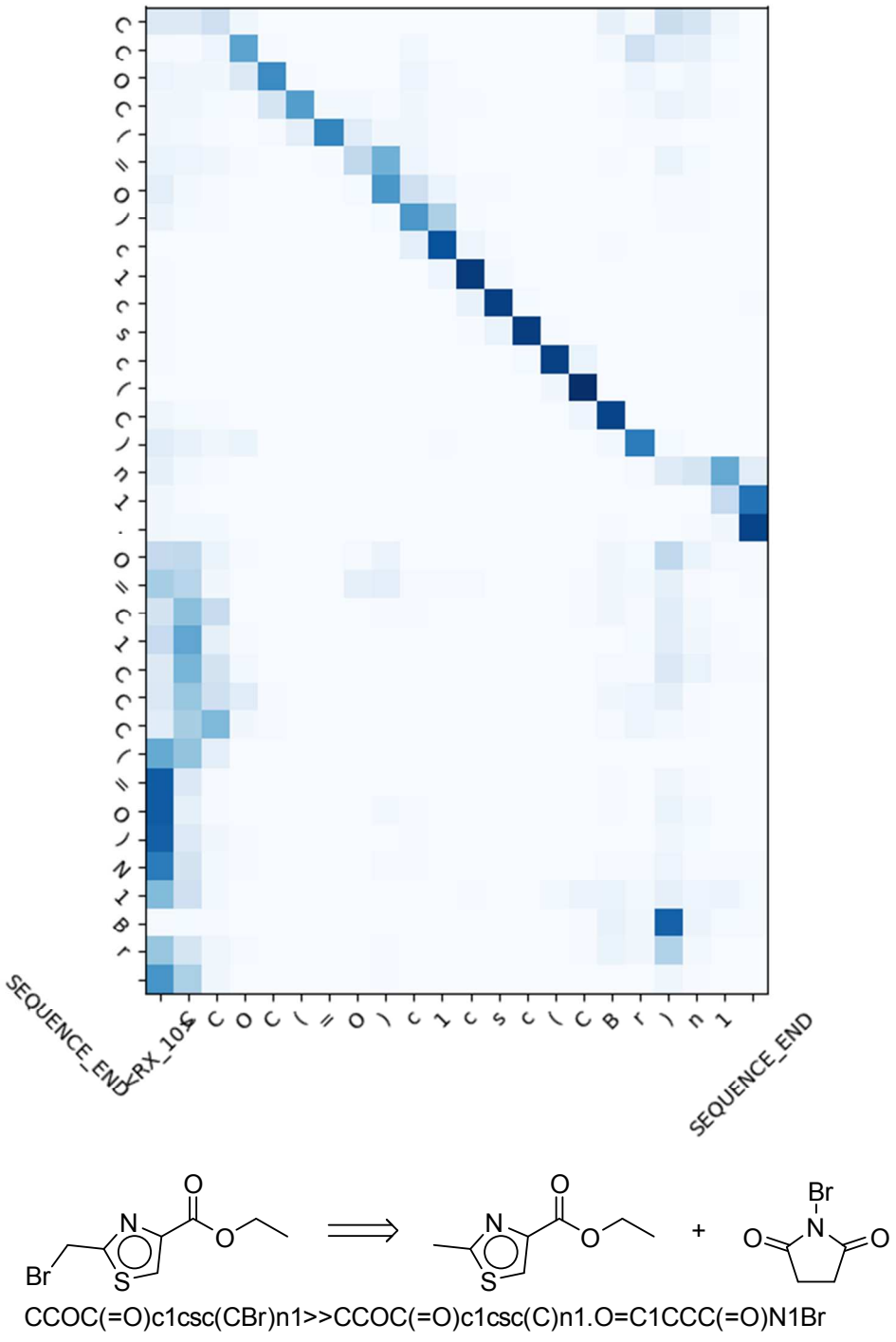


Figure S9: Attention weights for a reaction class 9 (Functional group interconversion) example



CCOC(=O)c1csc(CBr)n1>>CCOC(=O)c1csc(C)n1.O=C1CCC(=O)N1Br

Figure S10: Attention weights for a reaction class 10 (Functional group addition) example