Supplementary Information: Prediction of Chemical Reaction Yields using Deep Learning

Philippe Schwaller

IBM Research – Europe, Säumerstrasse 4, 8803 Rüschlikon, Switzerland
Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

E-mail: phs@zurich.ibm.com

Alain C Vaucher, Teodoro Laino

IBM Research – Europe, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

Jean-Louis Reymond

Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

Contents

1. Detailed results on Buchwald Hartwig reactions

Figure S1-S14 show the correlation between the measured yields and the predicted yields for the different splits published by Sandfort et al. [1]. Moreover, the root mean squared error (RMSE) and the mean average error (MAE) are shown in the figures.

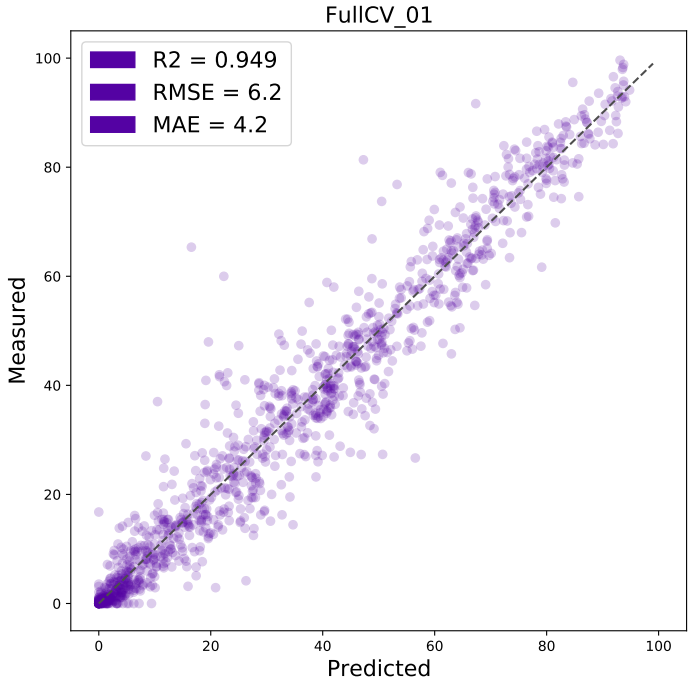

Figure S1. Measured vs predicted yields [%] - FullCV_01

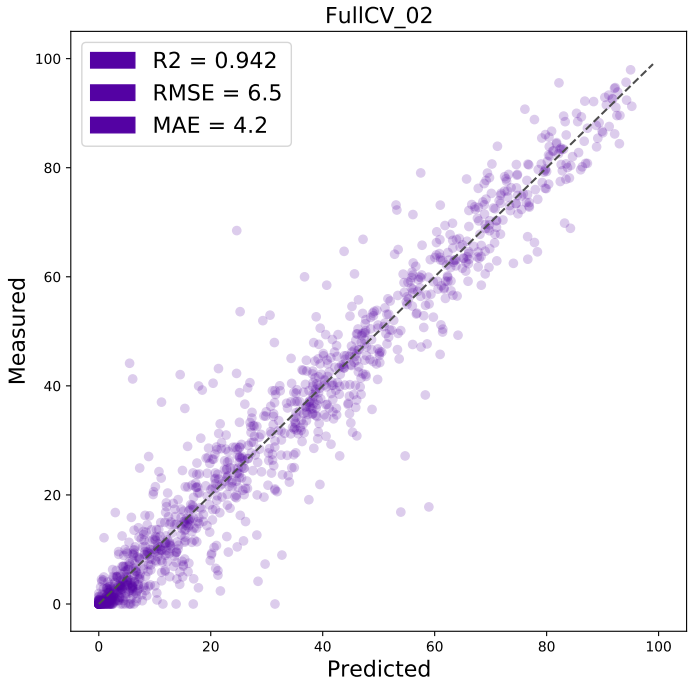

Figure S2. Measured vs predicted yields [%] - FullCV_02

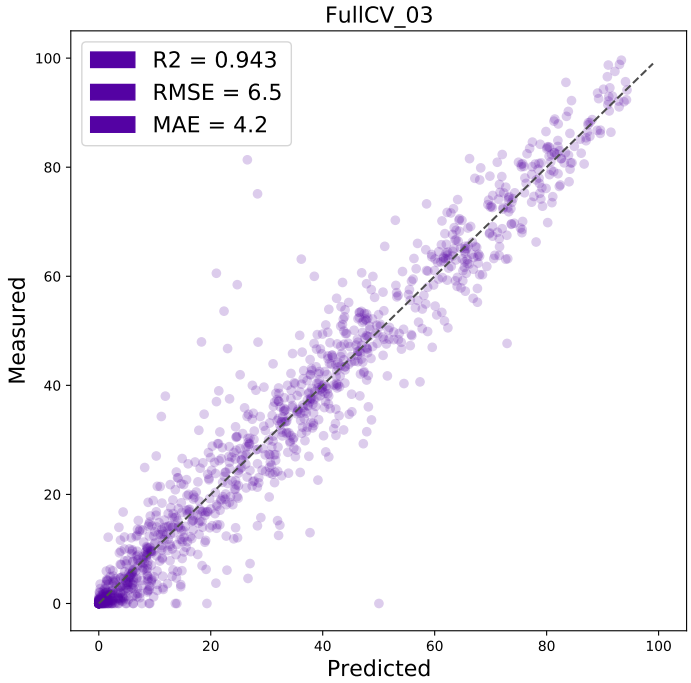

Figure S3. Measured vs predicted yields [%] - FullCV_03

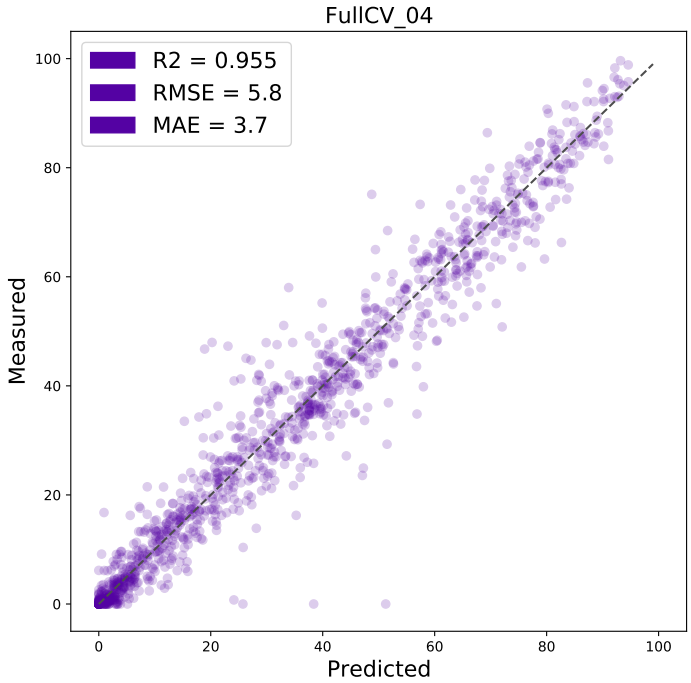

Figure S4. Measured vs predicted yields [%] - FullCV_04

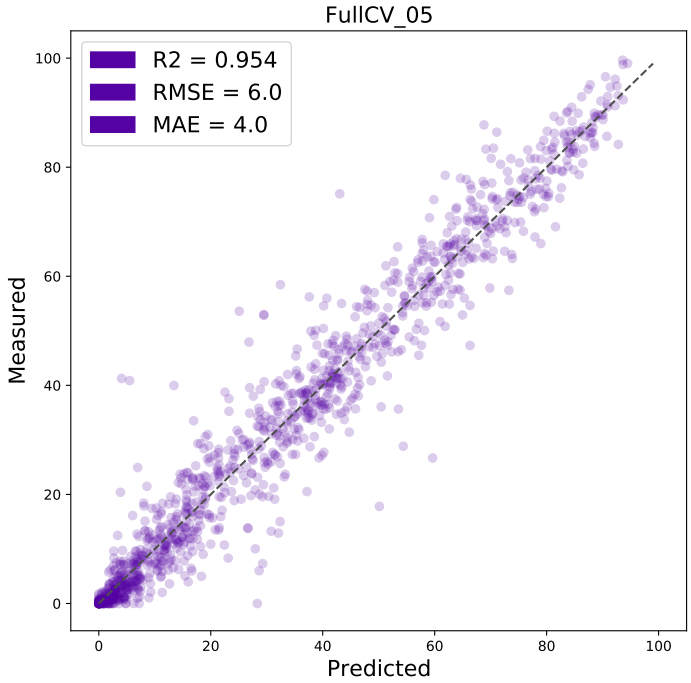

Figure S5. Measured vs predicted yields [%] - FullCV_05

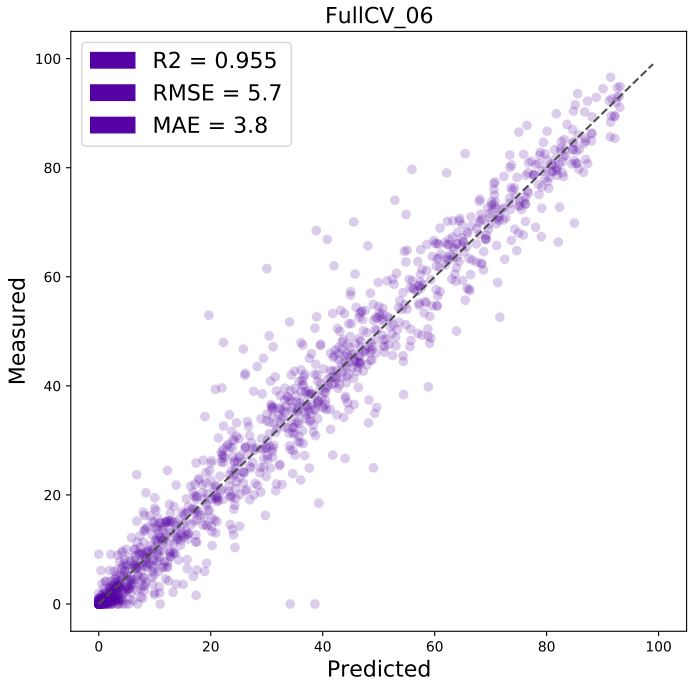

Figure S6. Measured vs predicted yields [%] - FullCV_06

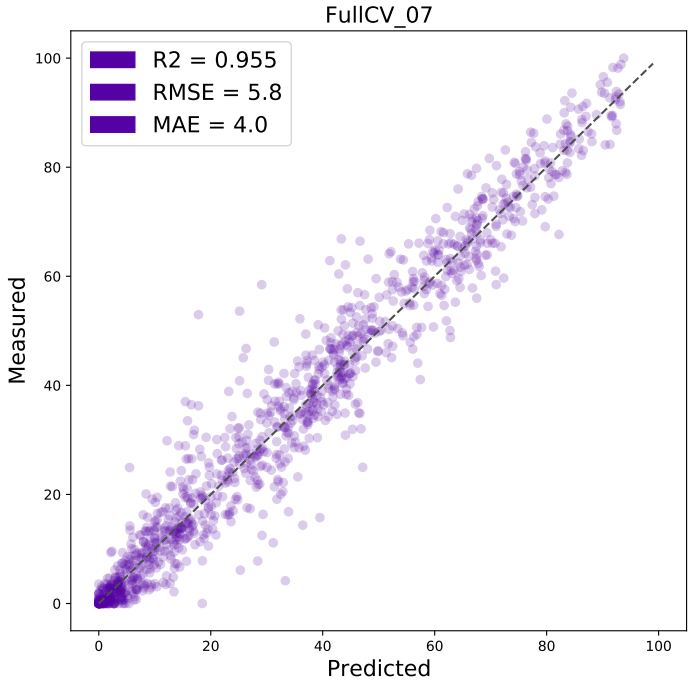

Figure S7. Measured vs predicted yields [%] - FullCV_07

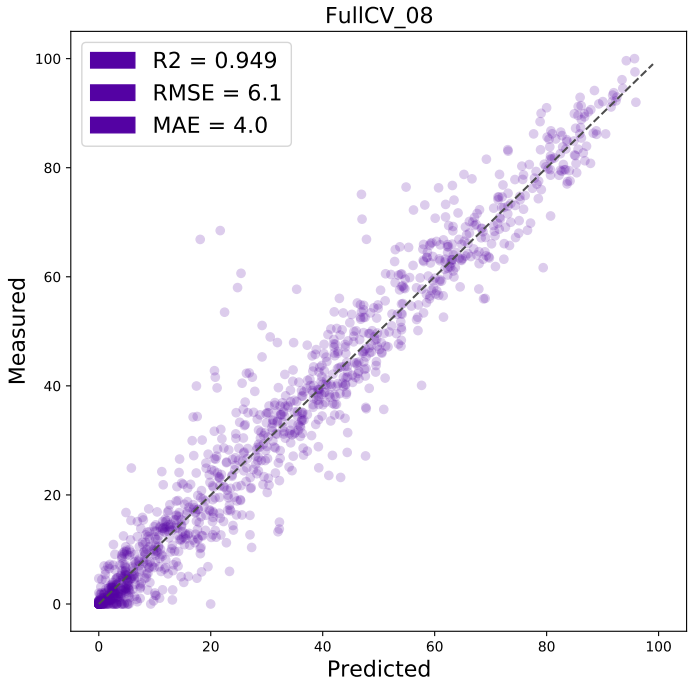

Figure S8. Measured vs predicted yields [%] - FullCV_08

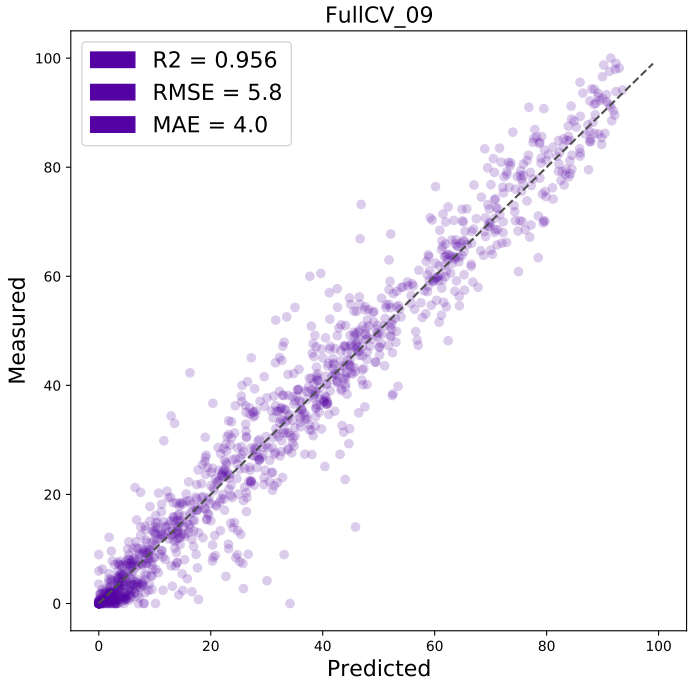

Figure S9. Measured vs predicted yields [%] - FullCV_09

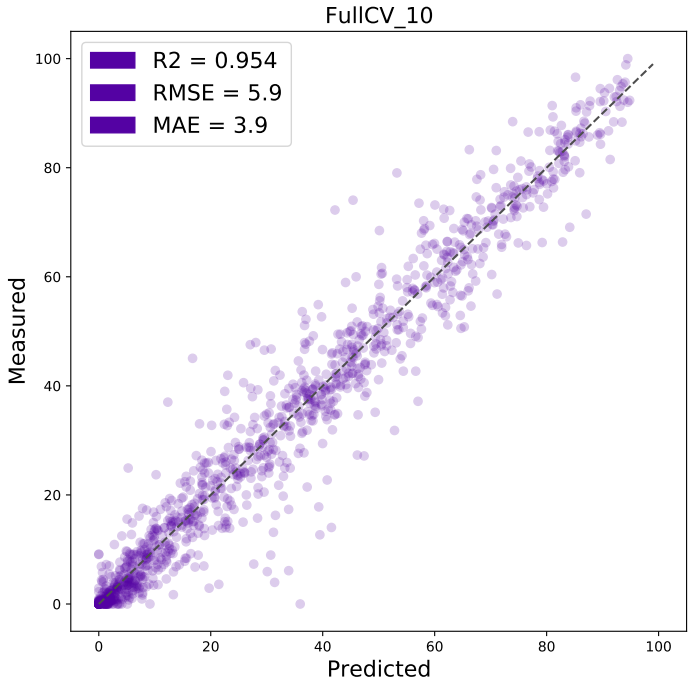

Figure S10. Measured vs predicted yields [%] - FullCV_10

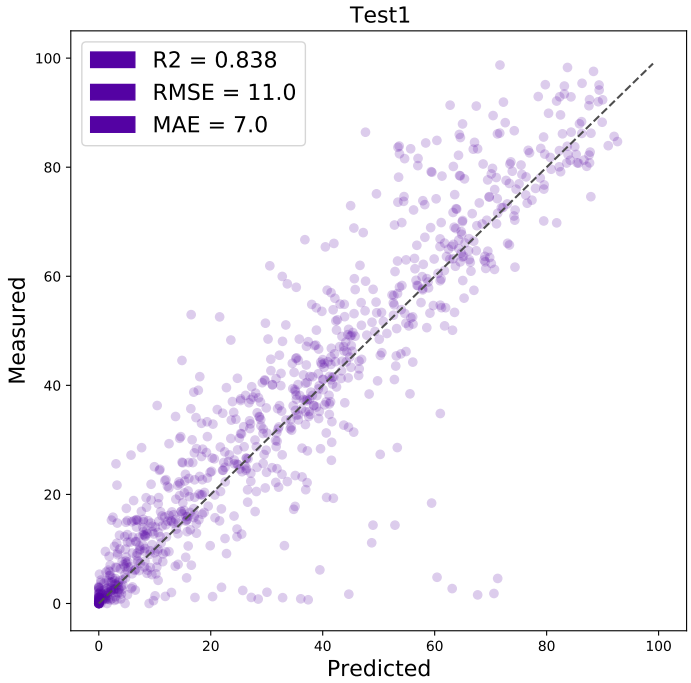

Figure S11. Measured vs predicted yields [%] - Test1

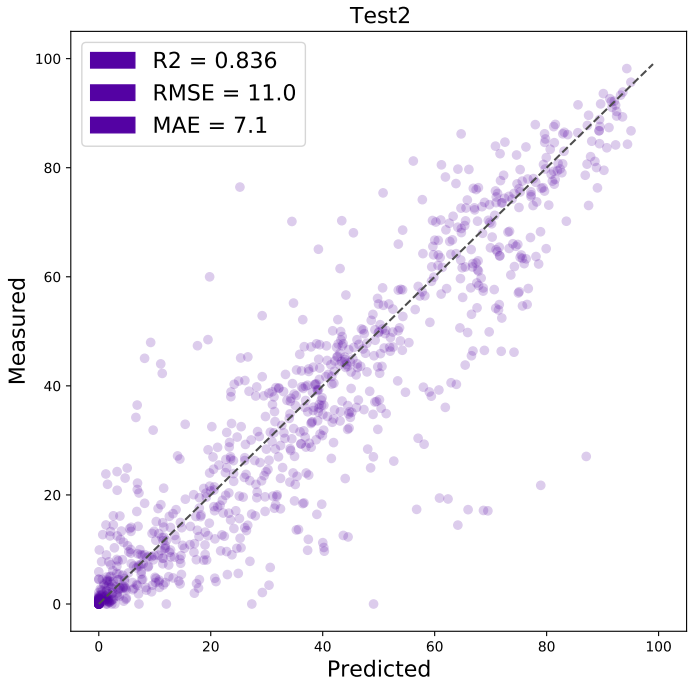

Figure S12. Measured vs predicted yields [%] - Test2

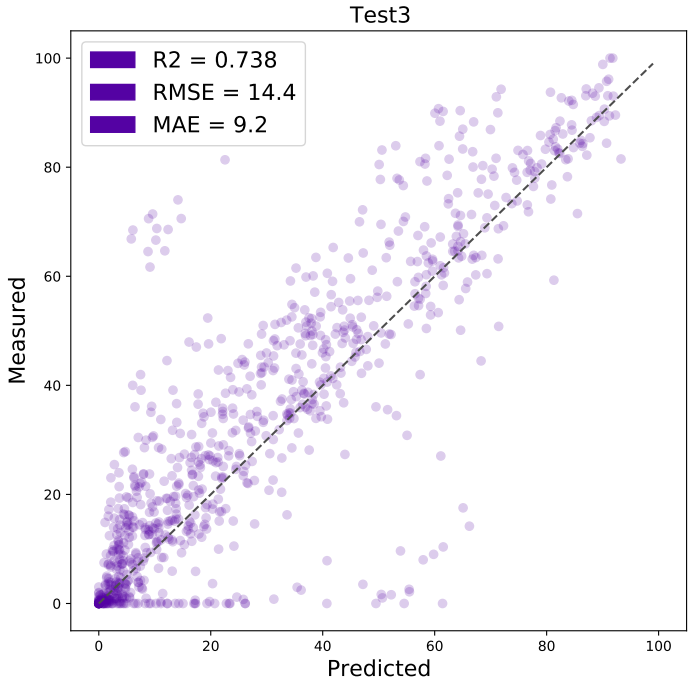

Figure S13. Measured vs predicted yields [%] - Test3

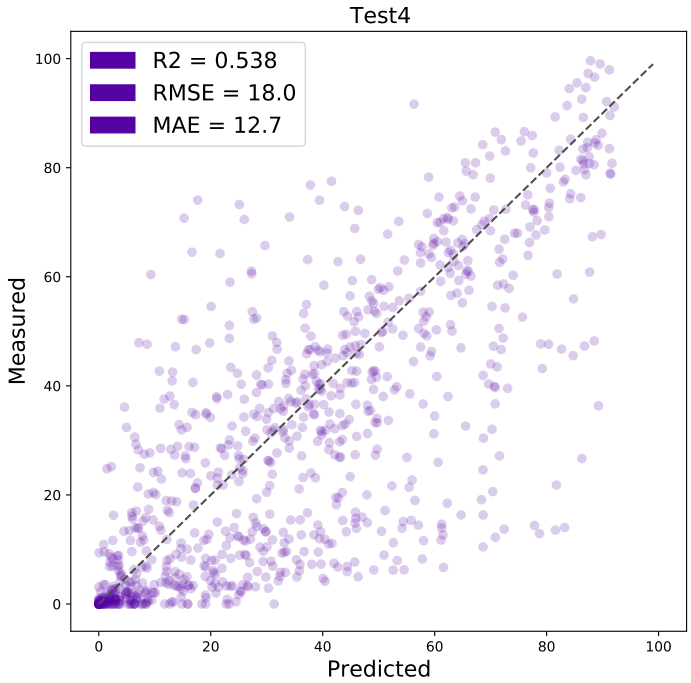

Figure S14. Measured vs predicted yields [%] - Test4

2. Detailed results on Suzuki-Miyaura reactions

Figure S15-S24 show the correlation between the measured yields and the predicted yields for model with the *rxnfp ft* base encoder on the 10 random splits.

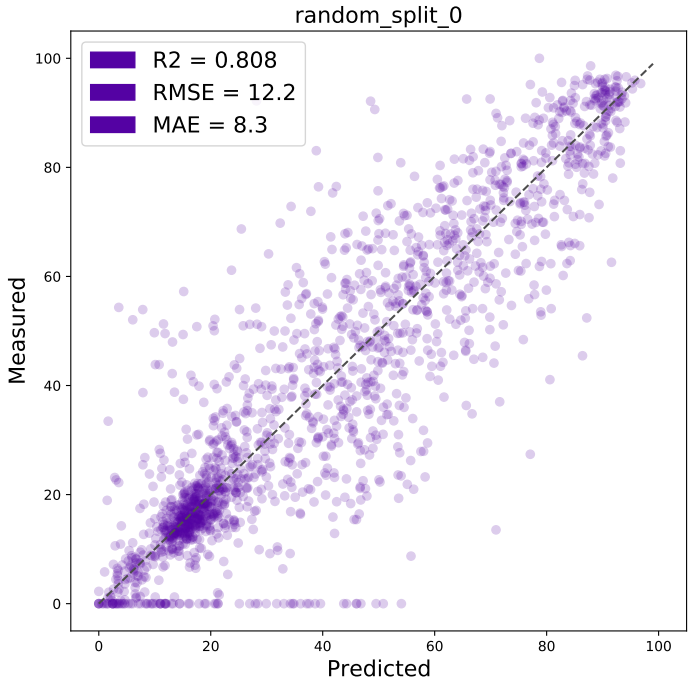

Figure S15. Measured vs predicted yields [%] - random_split_0

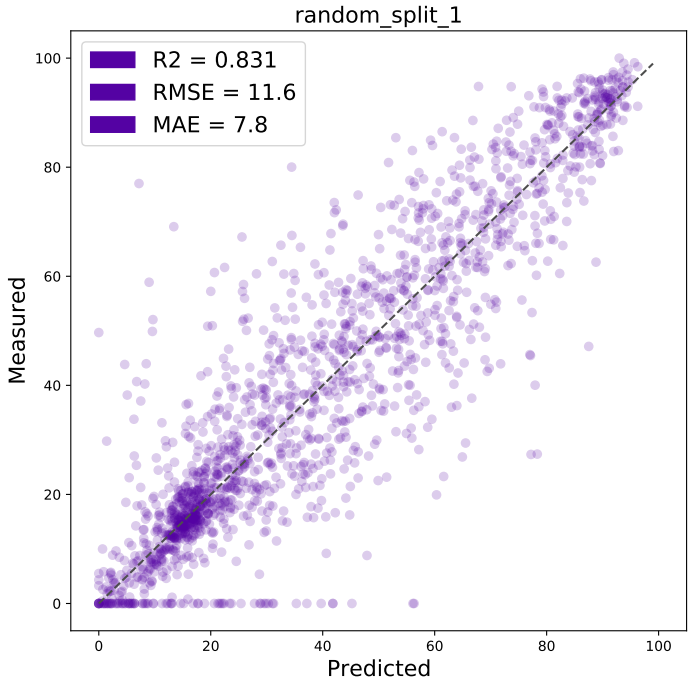

Figure S16. Measured vs predicted yields [%] - random_split_1

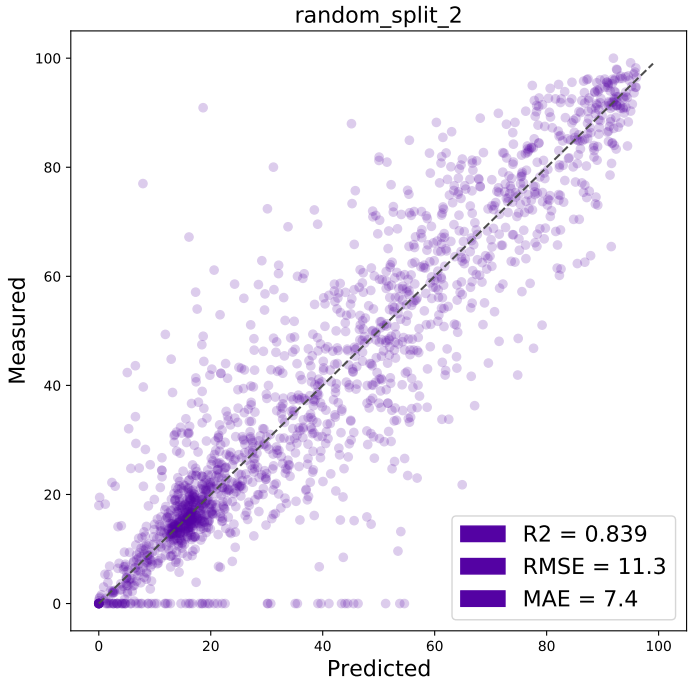

Figure S17. Measured vs predicted yields [%] - random_split_2

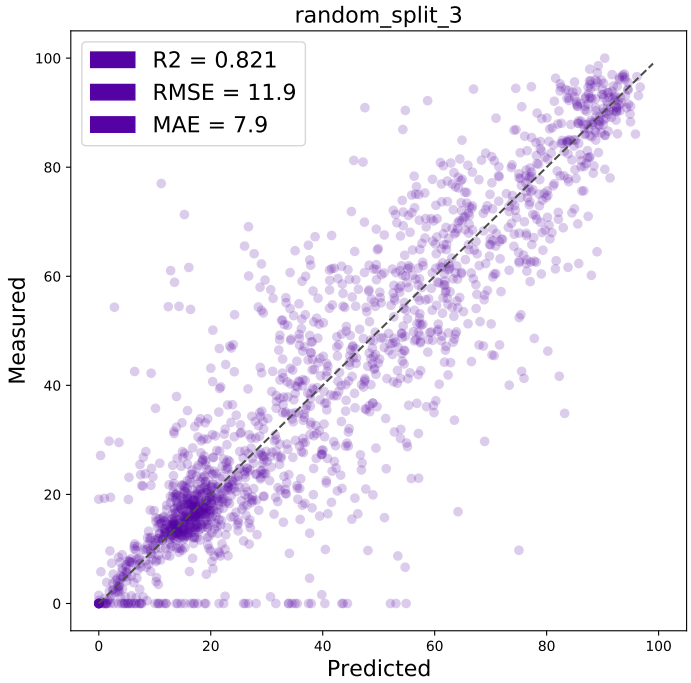

Figure S18. Measured vs predicted yields [%] - random_split_3

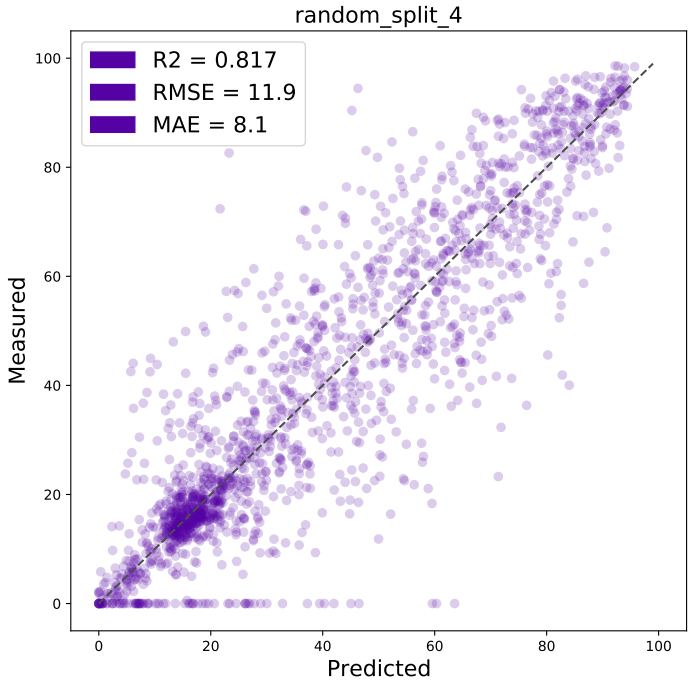

Figure S19. Measured vs predicted yields [%] - random_split_4

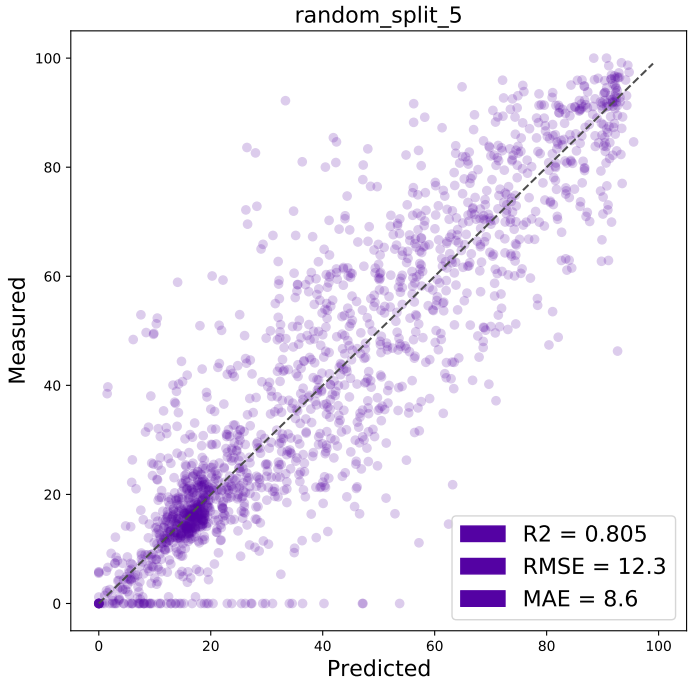

Figure S20. Measured vs predicted yields [%] - random_split_5

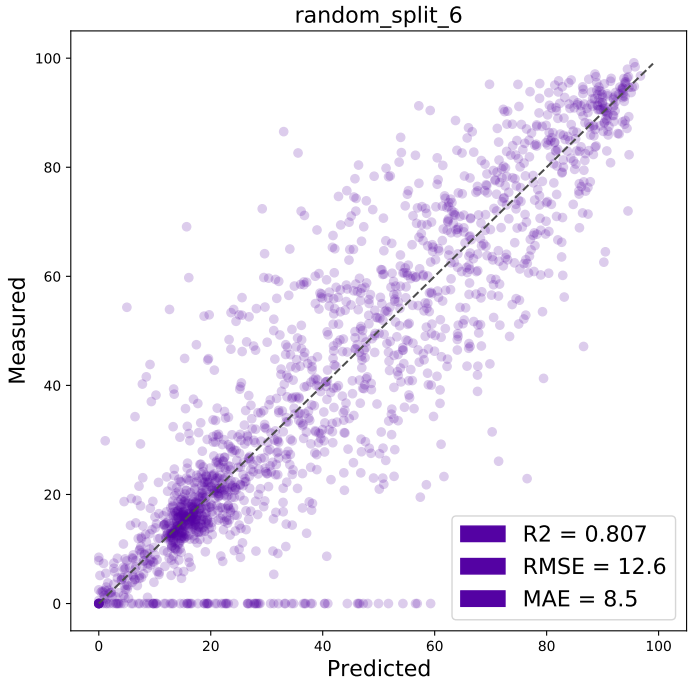

Figure S21. Measured vs predicted yields [%] - random_split_6

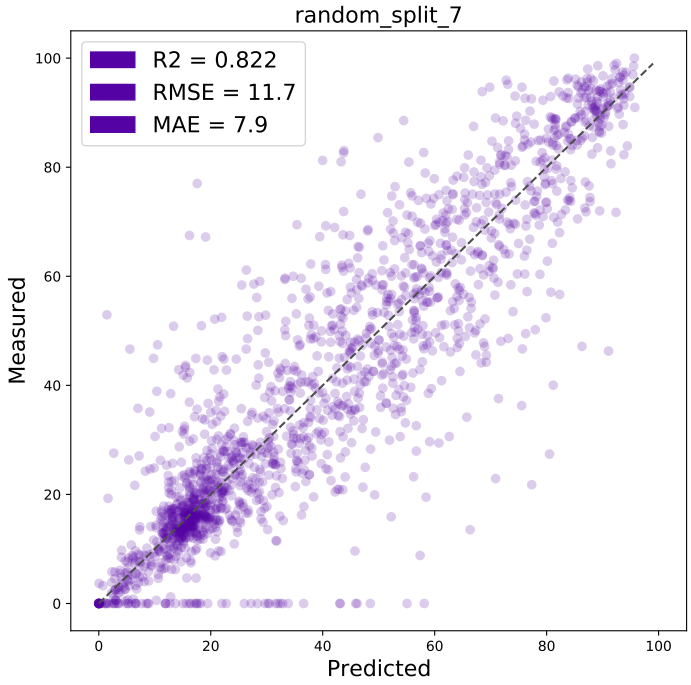

Figure S22. Measured vs predicted yields [%] - random_split_7

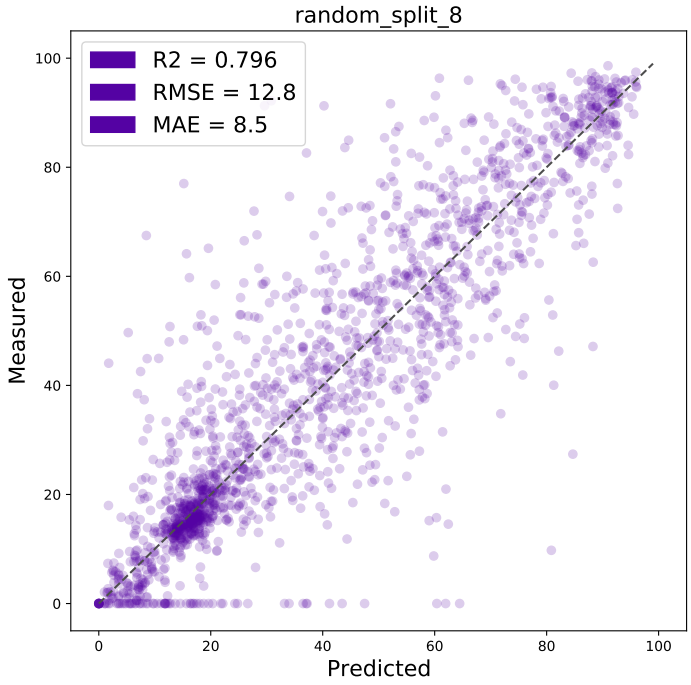

Figure S23. Measured vs predicted yields [%] - random_split_8

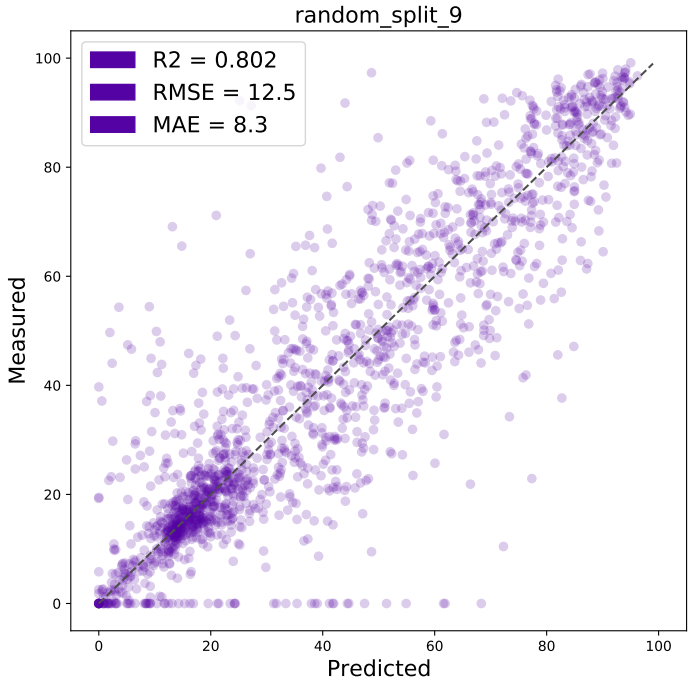

Figure S24. Measured vs predicted yields [%] - random_split_9

3. Detailed analysis of USPTO yields data

Table S1 show global statistics on the gram scale and sub-gram scale USPTO yields data sets.

Table S1. USPTO yield statistics

	gram scale	subgram scale
count	197619	302040
mean	73.2	56.8
std	20.9	26.6
min	0.0	0.0
25%	60.2	35.5
50%	78.0	58.9
75%	90.3	79.5
max	100.0	100.0

Tables S2 and S3 show the yields average in the random split test set for the different reaction superclasses.

Figure S25 shows the distributions of the smoothed yields. To smooth the yields of the USPTO data set [2, 3] we calculated the average of the 3 nearest-neighbours of the reaction, computed using the *rxnfp ft* [4] and faiss [5], and twice the own reaction yield.

Table S2. Test set sub-gram scale. Average and standard deviation per class.

Class	Name	Mean [%]	Std	Count
0	Unrecognised	52.1	26.8	12359
1	Heteroatom alkylation and arylation	53.3	25.8	12995
2	Acylation and related processes	54.8	25.6	10583
3	C-C bond formation	53.2	25.6	5111
4	Heterocycle formation	48.0	25.1	2043
5	Protections	69.8	22.3	527
6	Deprotections	68.7	25.2	8542
7	Reductions	67.5	26.1	3528
8	Oxidations	63.4	25.3	1078
9	Functional group interconversion (FGI)	62.3	25.2	2779
10	Functional group addition (FGA)	56.2	25.1	863

Table S3. Test set gram scale. Average and standard deviation per class.

Class	Name	Mean [%]	Std	Count
0	Unrecognised	69.4	22.0	10327
1	Heteroatom alkylation and arylation	71.9	20.9	7912
2	Acylation and related processes	74.5	19.7	4745
3	C-C bond formation	70.7	20.0	2547
4	Heterocycle formation	67.1	22.9	1417
5	Protections	79.9	18.5	1154
6	Deprotections	82.2	16.9	3332
7	Reductions	81.2	18.2	3105
8	Oxidations	76.0	18.8	742
9	Functional group interconversion (FGI)	74.9	20.1	2751
10	Functional group addition (FGA)	71.7	21.7	1491

4. Hyperparameter tuning

The two hyperparameters we tuned were dropout rate (between 0.05 and 0.8) and learning rate (between 1e-6 and 1e-4). For the *rxnfp pretrained* model on the Buchwald-Hartwig reactions a learning rate of 9.659e-05 and dropout probability of 0.7987 led to the highest validation R^2 score. We observe high R^2 scores for a wide range of dropout probabilities. The hyperparameter tuning was performed on a single *Nvidia RTX 2070 super* GPU and the optimal hyperparameters were found in less than 12 hours. A typical training run (10 epochs) on the same hardware takes 4 minutes and 30 seconds. We trained the final models for 15 epochs.

On the Suzuki-Miyaura reactions, we selected a learning rate of 5.812e-05 and

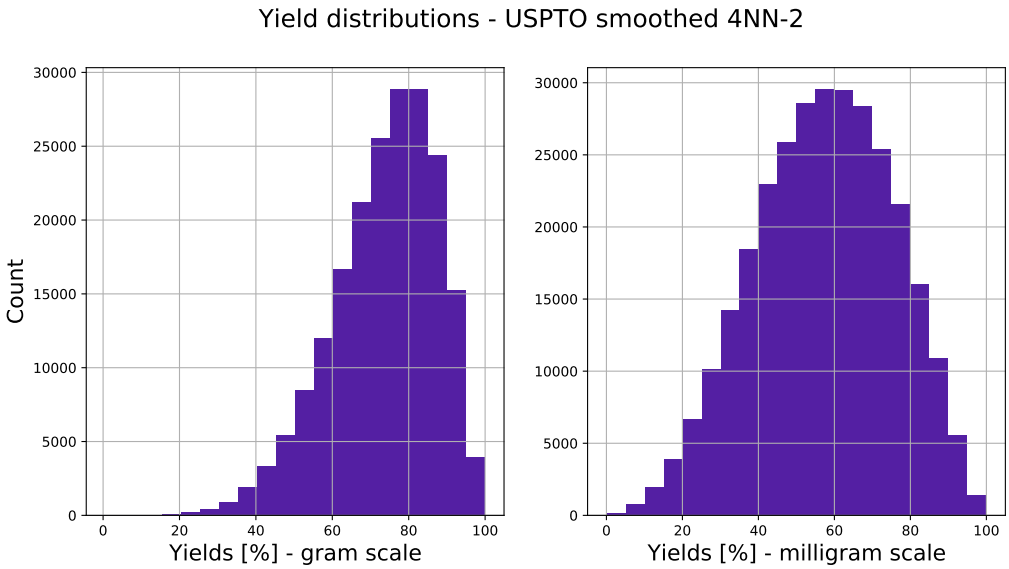

Figure S25. Smoothed USPTO yields distribution separated in gram and sub-gram scale

dropout probability of 0.5848 for the *rxnfp pretrained* base encoder and a learning rate of 9.116e-05 and dropout probability 0.7542 for the *rxnfp ft* base encoder model.

On the USPTO data we performed a hyperparameter search using a reduced training set of 50k reactions and only 3 epochs. We selected a learning rate of 1.562e-05 and dropout probability of 0.5237 for the gram scale and 2.958e-05 and 0.5826 respectively, for the sub-gram scale. The final models were trained for 2 epochs on the complete training data, as an evaluation showed signs of over-fitting from the third epochs on.

Figure S26 – S30 show the hyperparameters with the corresponding R^2 values on the validation set. The validation was made on subsplit of the training set of the first random split for all three data sets. Overall, the learning rate seemed to be more important to tune than the dropout probability.

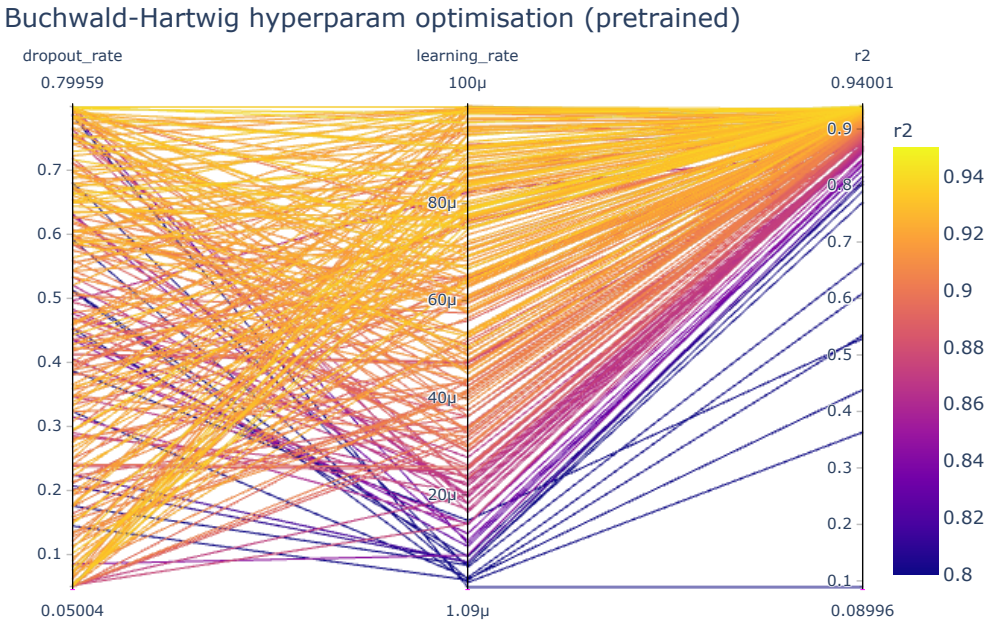

Figure S26. Hyperparameter optimisation on Buchwald-Hartwig data set (pretrained base encoder)

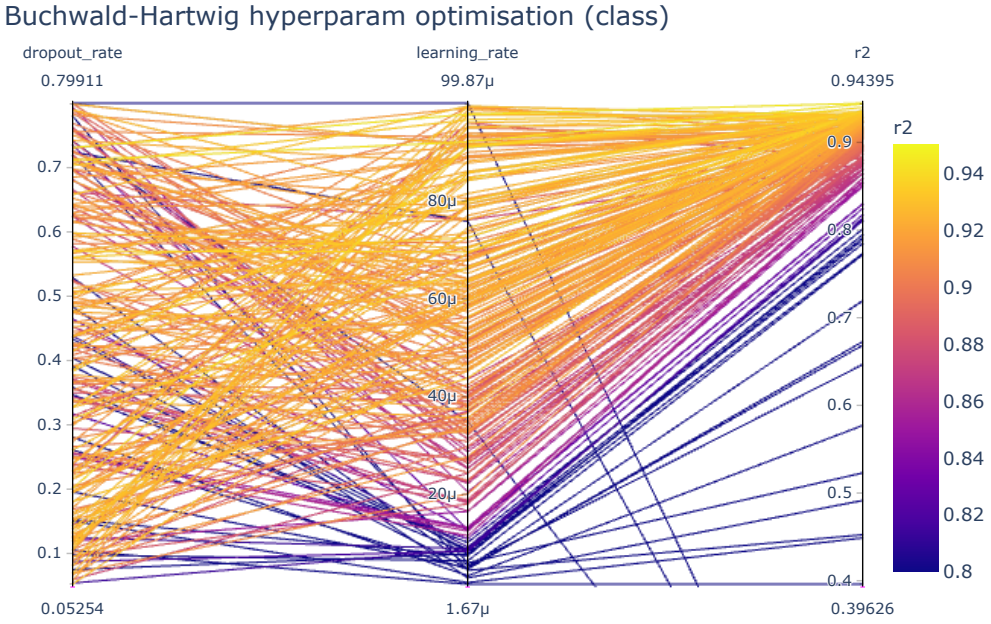

Figure S27. Hyperparameter optimisation on Buchwald-Hartwig data set (class base encoder)

References

[1] Sandfort, F., Strieth-Kalthoff, F., Kühnemund, M., Beecks, C. & Glorius, F. A structure-based platform for predicting chemical reactivity. *Chem* (2020).

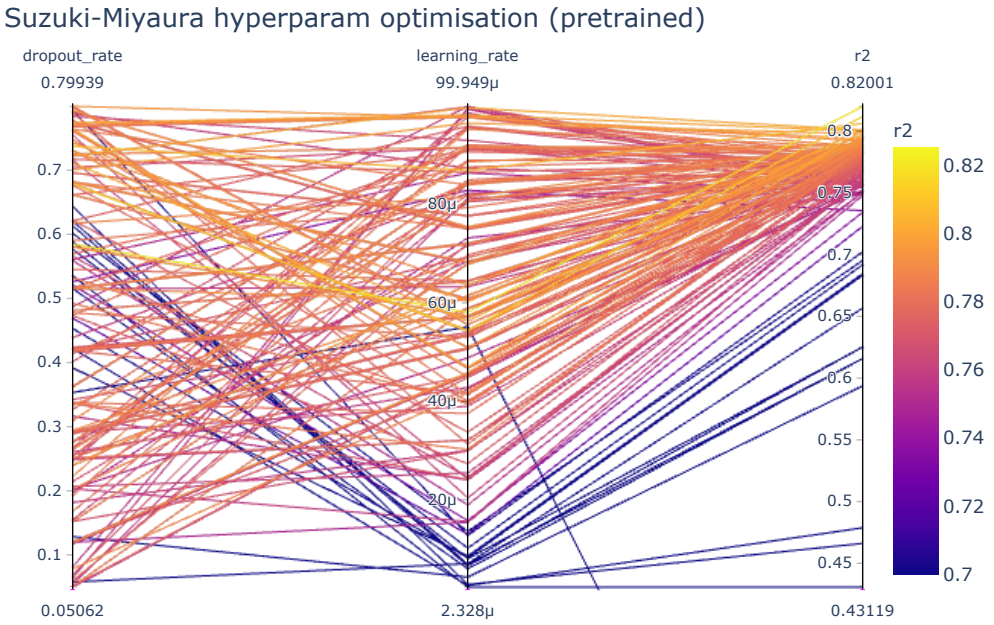

Figure S28. Hyperparameter optimisation on Suzuki-Miyaura data set (pretrained base encoder)

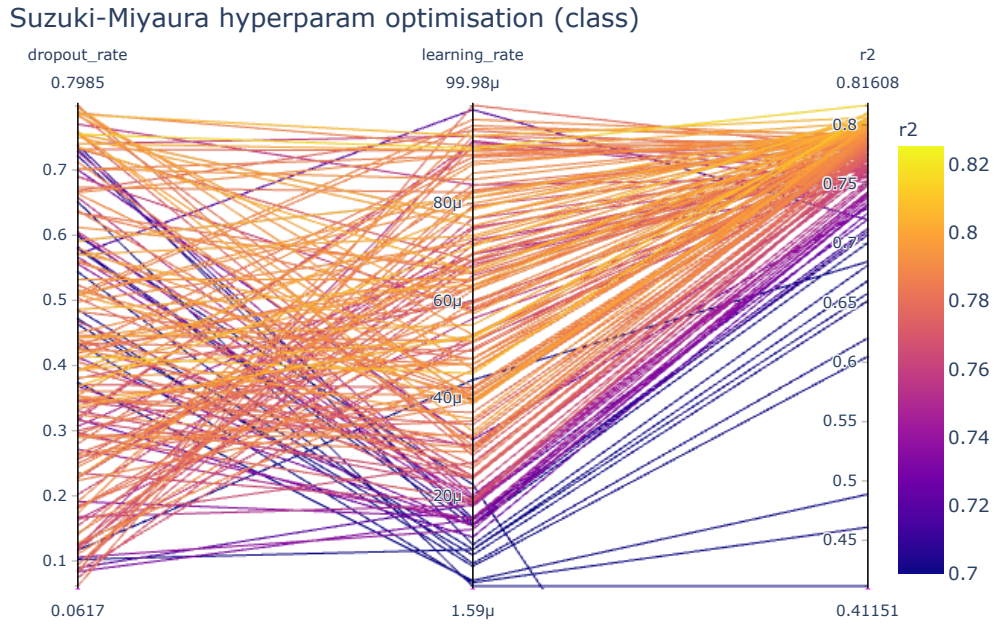

Figure S29. Hyperparameter optimisation on Suzuki-Miyaura data set (class base encoder)

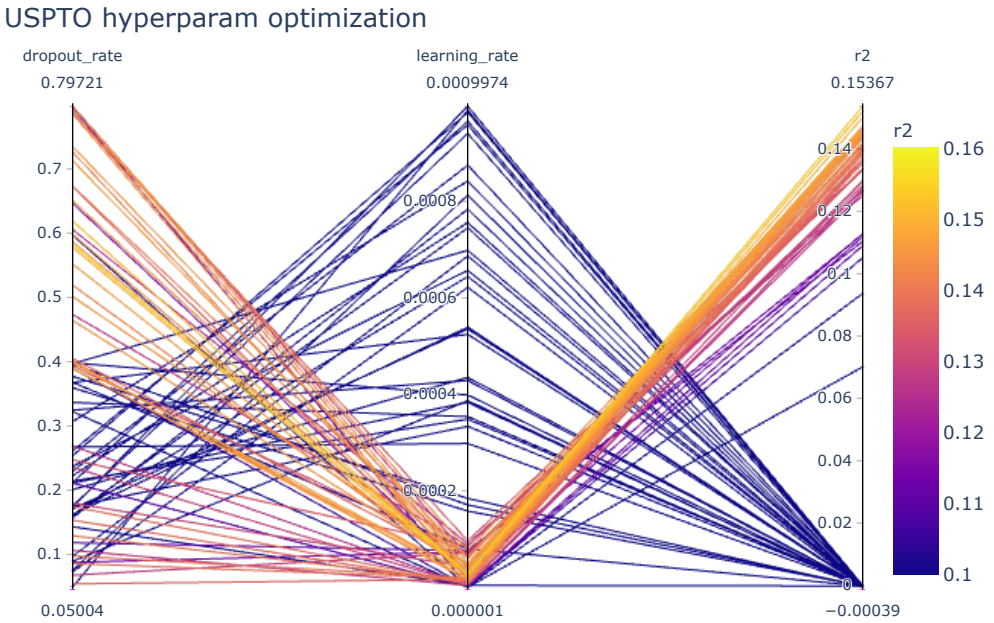

Figure S30. Hyperparameter optimisation on USPTO subgram data set (pretrained base encoder)

[2] Lowe, D. M. *Extraction of chemical structures and reactions from the literature.* Ph.D. thesis, University of Cambridge (2012).

[3] Lowe, D. Chemical reactions from US patents (1976-Sep2016) (2017).

[4] Schwaller, P. *et al.* Mapping the space of chemical reactions using attention-based neural networks. *ChemRxiv preprint* doi:10.26434/chemrxiv.9897365 (2019).

[5] Johnson, J., Douze, M. & Jégou, H. Billion-scale similarity search with gpus. *arXiv preprint arXiv:1702.08734* (2017).