Supporting Information for the Paper Exploring Chemical Reaction Space With Reaction Difference Fingerprints and Parametric t-SNE

Mikhail Andronov,[†] Maxim V. Fedorov,[‡,¶,§] and Sergey Sosnin[*,§,¶]

†*Faculty of Fundamental Physical and Chemical Engineering, Lomonosov Moscow State University, Leninskie gory, 1 119991 Moscow, Russian Federation*

‡*Sirius University of Science and Technology, Olimpiysky ave. b.1, 354000 Sochi, Russian Federation*

¶*Syntelly LLC, Bolshoy Boulevard 30, bld. 1, 121205 Moscow, Russian Federation*

§*Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, 121205 Moscow, Russian Federation*

E-mail: sergey.sosnin@skoltech.ru

Phone: +7 (926)6556761

The Architecture of the Artificial Neural Network

The architecture of our neural network is given in Table S1. It is a fully-connected network with a batch normalization layers and Rectified Linear Unit (ReLU) activation functions. To train our network we used Adam optimizer with learning rate = 0.002.

Our experiments revealed that four layers demonstrated optimal speed-vs-performance balance and for this reason, we used this architecture in all experiments.

Table S1: The architecture of the neural network

Layer	Neurons	Batch Normalization
Input	2048	Yes
Hidden 1	2048	Yes
Hidden 2	2048	Yes
Output	2	No

Network training

Fig. S1 shows the learning curves for our models. These models were trained on BERT reaction fingerprints and reaction difference fingerprints. Difference fingerprints are based on MorganFP, AtomPairFP, and Topological Torsion descriptors. Each model was trained several times with different perplexities. Solid and dashed lines indicate the training and validation loss, respectively. One can see that there is no overfitting and early stopping is not necessary.

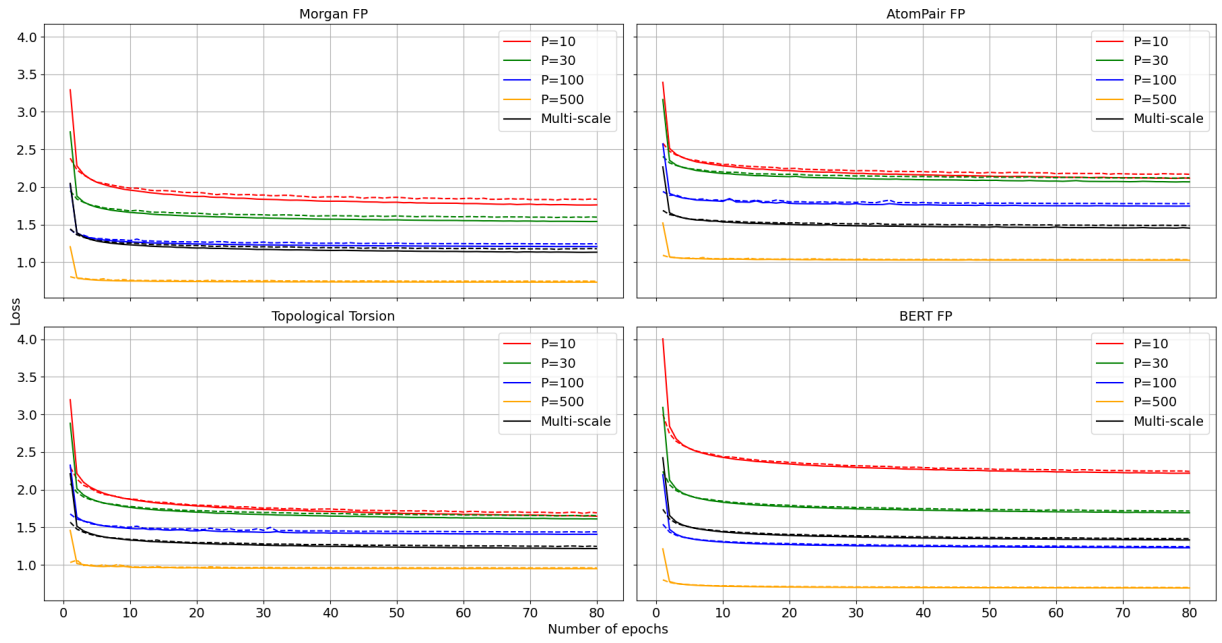

Figure S1: Learning curves of the models trained with various perplexities on different vector representations of reactions. Continuous and dashed lines indicate the training and validation loss, respectively.

Transformer fingerprints

Fig. S2 shows a reaction space map yielded by a four-layer parametric multi-scale t-SNE model trained on BERT fingerprints for 80 epochs.

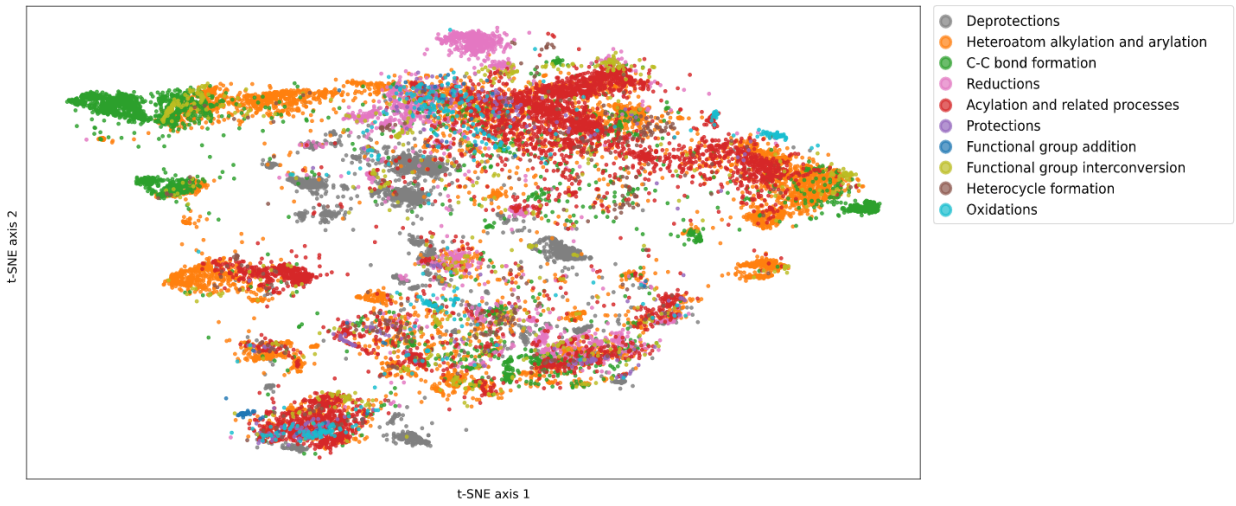

Figure S2: A map of chemical reaction space produced by a multi-scale model trained on BERT FP.

Structural fingerprints

Fig. S3 demonstrates a projection for reaction structural fingerprints. We used Morgan fingerprints with perplexity 30 and the Euclidean distance function. We noted that structural fingerprints could not provide distinct, separable clusters and, apparently, are not applicable for the exploration of reaction space. Our experiments revealed that other types of molecular fingerprints used with structural FP or different distance functions do not improve the picture.

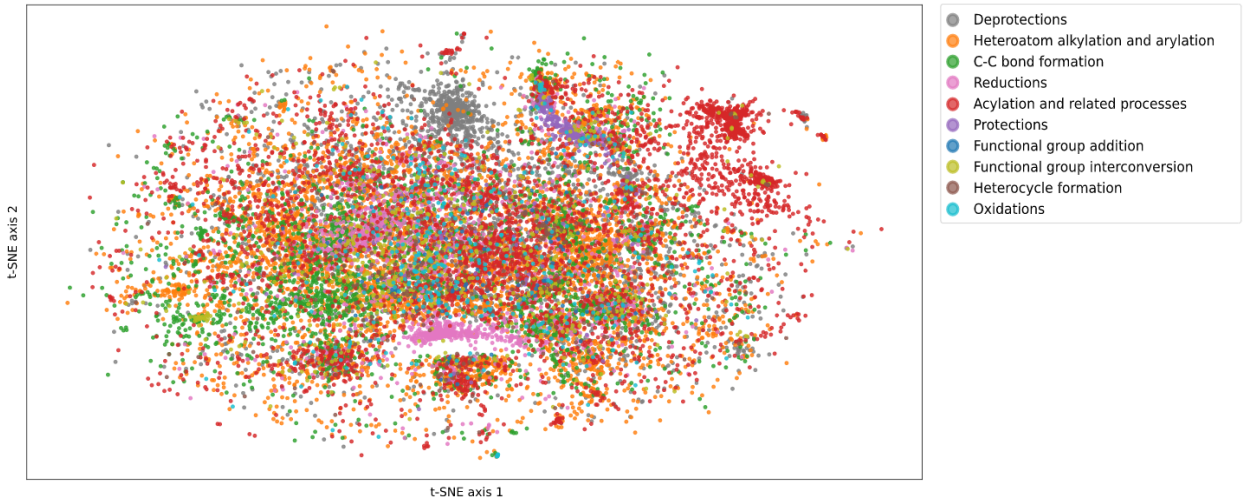

Figure S3: Reaction space map produced by a model built on structural Morgan fingerprints and trained for 80 epochs with perplexity 30.

Maps obtained with Jaccard dissimilarity as a distance function in the initial space

Fig. S4 shows a reaction space map yielded by a model built on reaction difference Morgan fingerprints and trained for 50 epochs wit Jaccard dissimilarity as the distance function in the higher-dimensional space. The classes separation accuracy for this map is 73%. There are not many distinct clusters and reaction points are mostly mixed up. Our experiments show that the pictures produced by other models using Jaccard dissimilarity do not differ significantly from Fig. S4.

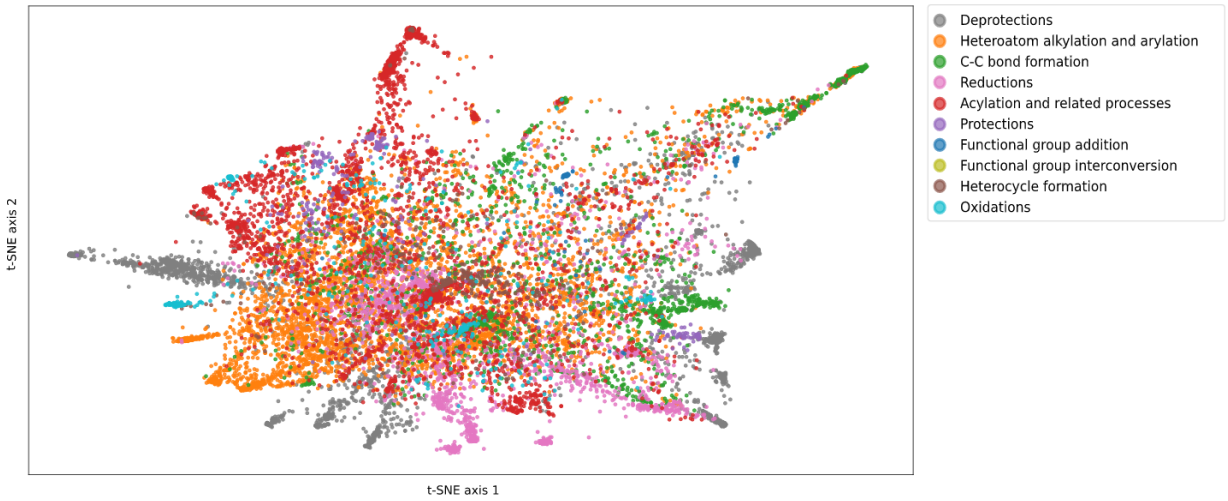

Figure S4: Reaction space map yielded by a model built on reaction difference Morgan fingerprints and trained for 50 epochs wit Jaccard dissimilarity as the distance function in the higher-dimensional space.