Bidirectional Graphormer for Reactivity Understanding: neural network trained to reaction atom-to-atom mapping task

Ramil Nugmanov[1*], Natalia Dyubankova[1], Andrey Gedich[2] and Joerg Kurt Wegner[3]

[1] Janssen Pharmaceutica 30, Turnhoutseweg str., 2340 Beerse, Belgium

[2] Arcadia Inc. 28 k2, Bolshoy Sampsonievskiy pr., St. Petersburg, 194044, Russia

[3] Janssen Research & Development, LLC, 255 Main St, Cambridge, MA 02142, USA

*e-mail: rnugmano@its.jnj.com

Abstract

This work introduces *GraphormerMapper* – a new algorithm for reactions atom-to-atom mapping (AAM) based on a distance-aware BERT neural network. In benchmarking studies with IBM RxnMapper[1], the best AAM algorithm according to our previous study[2], we demonstrate that our AAM algorithm is superior on our "Golden" benchmarking dataset[2]. The mapper is implemented in Chython [https://github.com/chython/chython] and Chytorch [https://github.com/chython/chytorch, https://github.com/chython/chytorch-rxnmap] Python packages which are freely available for out-the-box use. Chython is a cheminformatics library with a simple interface for processing reaction and molecular data. The key features of Chython are: chemical functional groups standardization, checking atom valence errors, substructure search, and advanced reaction manipulation, for example, generating products from reactants and reaction atom-to-atom mapping. Chytorch provides a PyTorch-like[3] interface for graph-based neural networks developed specifically for chemical tasks.

Introduction

Reaction atom-to-atom mapping (AAM)[4,5] is a procedure that establishes a correspondence between the atoms of reactants and products. AAM allows identifying reaction centers (RC) which can be used for a reaction templates extraction serving multiple downstream tasks like automatized forward/retrosynthesis planning[6–10], reaction classification[11], and reaction substructure and similarity searching[12–15].

Several publicly and commercially available AAM tools are currently available[1,16–19]. A recent benchmarking studying[2] showed mediocre performance of various AAM algorithms, mostly expert systems. The best tested AAM is a data analytics/AI approach being called RxnMapper from IBM assigning 1565 of 1851 reactions in the "Golden" dataset correctly[2].

In this work, we implemented a new neural network inspired by representing compounds (from SMILES) as a graph[20] instead of sequences[21] and improved the attention-guided AAM algorithm[1]. The new algorithm has improved inference stability as it does not depend on the order of molecules in the reactants and products set and does also not depend on the order of tokens in a SMILES string, which is a typical challenge for ensuring a smooth space embedding[22,23] and a better generalization ability.

Methods

Condensed Graph of Reaction for AAM validation

A Condensed Graph of Reaction (CGR)[24,25] encodes a chemical reaction as a single graph, where edges and nodes may obtain dynamical properties corresponding to chemical transformations, for example, a bond creation or breaking, an atom gaining or losing a charge, etc (see Figure 1).

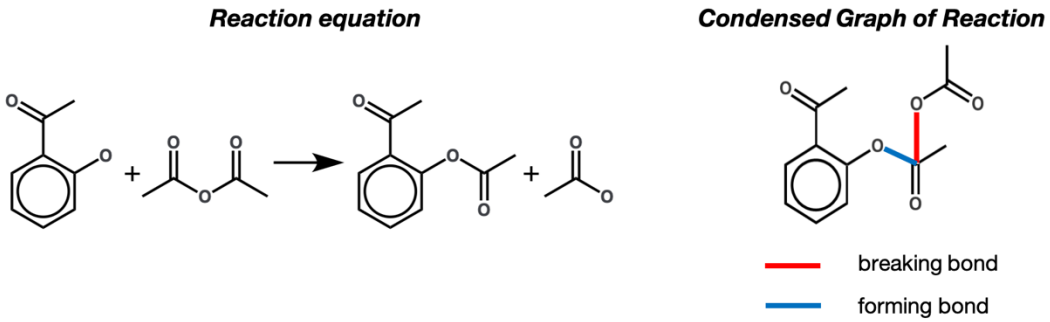


Figure 1. Depiction of a reaction equation (*on the left*) and the corresponding Condensed Graph of Reaction (*on the right*) for an esterification reaction. In addition to conventional (single, double, etc.) bonds, CGR contains dynamical bonds (forming and breaking during the transformation, colored here in blue and red, respectively).

We used CGR's SMILES[26,27] representations for a simple comparison of AAM obtained by RxnMapper, present algorithm, and manually curated. The workflow was the same as in the previous work[2].

Data sets

The benchmarking study has been performed on the "Golden" dataset[2], which contains 1851 reactions in total, obtained by merging curated Jaworski's[28] dataset with 1382 entries and 469 manually mapped USPTO[29] records.

The model is trained on the combined open-source reaction dataset USPTO[29] and commercial reaction dataset Pistachio[19]. The data normalization followed the process described in[30] and duplicated entries were removed.

Graph-based transformer neural network

Our neural network is inspired by a graphormer architecture[20]. We changed the centrality encoder to total neighbors encoder, which includes the count for explicit atom neighbors and implicit hydrogens. An edge encoder is not present as a combination of centrality encoder with an atom type encoder can encode hybridization of atoms implicitly. Moreover, atoms' formal charges

and radical states are also skipped. This approach allows us to code resonance structures as a single form, and skip the “aromatic” bonds concept frequently used for arenes.

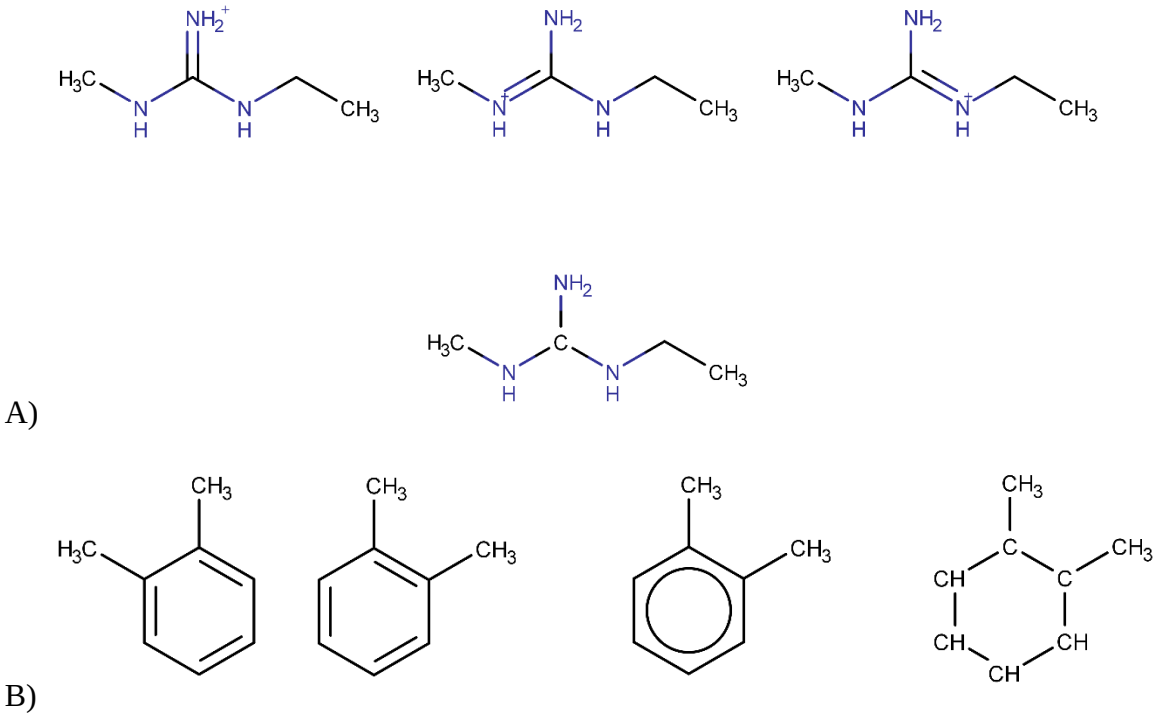


Figure 2. Multiple resonance structures folded into a single form (A), different aromatic ring representations folded into a single form (B)

The spatial encoder in our implementation is limited by a configurable maximum distance threshold. Values encoded by elements above a spatial threshold are considered equal. Short-range distance limits boosts attention to close atoms, while transformer architectures their multiple layers can extract features between long-range atoms. Though, long-range distances can lead to ambiguous predictions on molecules larger than in train set.

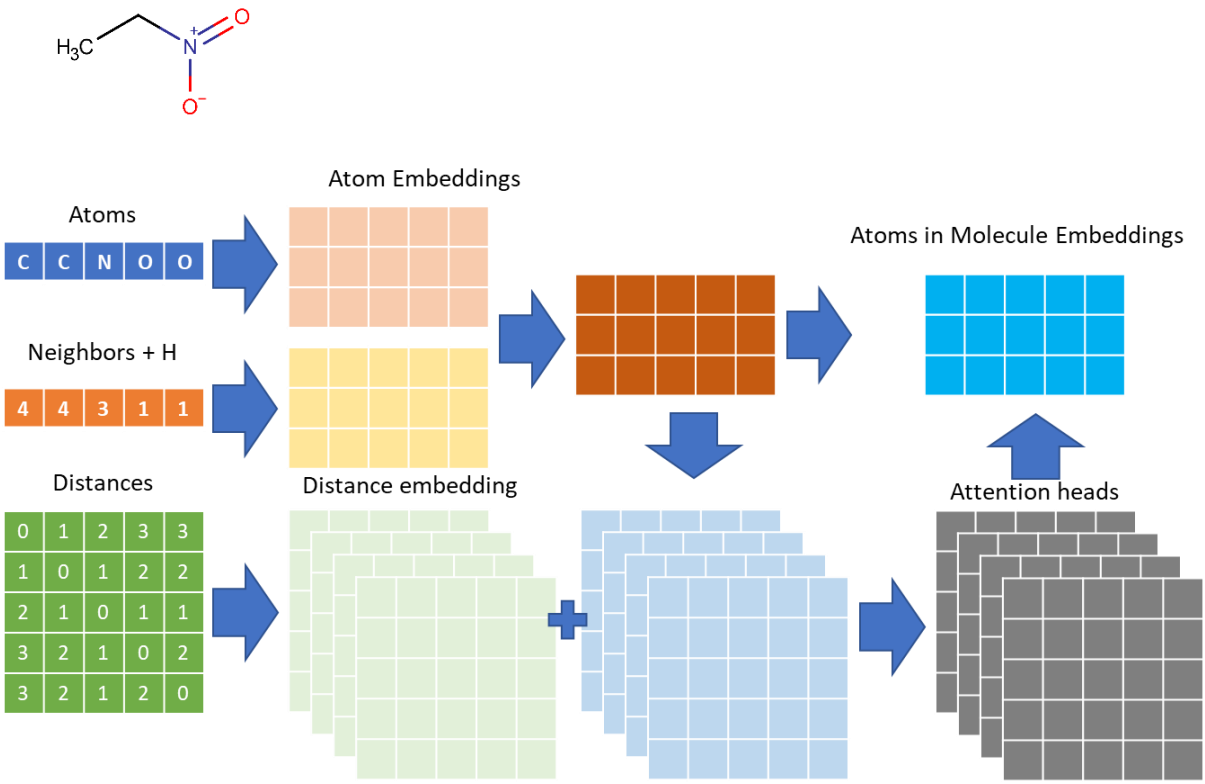


Figure 3. Modified Graphormer architecture.

A Graphormer has a classic transformer-style layer design and is applicable only for a single molecule. We used weight sharing similar to ALBERT[31] to prevent overfitting and to decrease model size. For multicomponent molecules (salts, complexes, etc) disconnected parts in a distance matrices were encoded as a new special token. This gives the possibility to detect relationships between ligands, metals or ions instead of enforcing hardcoded bond orders as input.

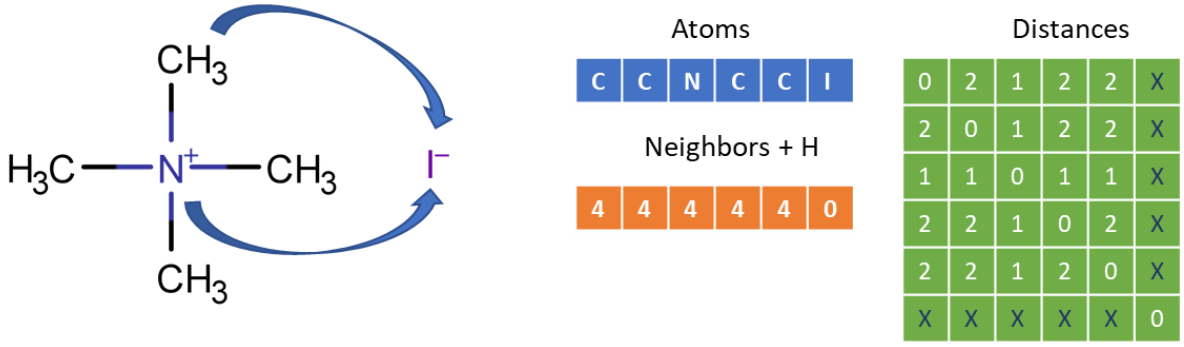


Figure 4. Salts and complexes coding technique.

Graph-based BERT neural network

In order to learn a reactants-to-products relationships we used techniques from BERT[32]. The architecture consists of two parts: the first part is the modified graphormer, used to generate atom-in-molecule embedding, the second part is a BERT-like network with sentence encoding, used to mark reactants and products. Atom embeddings from the first part of the network contain full information about a molecule, thus the second part can find the required reactant-to-product relationship. We didn't use positional encodings because the order of atoms and molecules in reactions is arbitrary, changing molecules' order does not change reaction outcome.

The model was trained via masked-learning-model (MLM) tasks. For MLM in reaction equation part of atoms and neighbors randomly replaced with "[mask]" tokens. Each epoch MLM is independent, like in RoBERTa[33], so each epoch is trained on "different" data.

Attention guided atom mapping

Inspired by SMILES-based attention guided AAM described by Schwaller et all[1] we developed a SMILES-independent approach using graph-based neural networks. In SMILES-based implementation, mapping is being done by iterative searching of maximal probability between reactants and products' atoms with independent pre-normalization of products to reactants weights on each iteration. Only one head attention from penultimate layer is used to extract the mapping. The choice of the head can lead to overfitting and provides dataset dependent model. and whose choice is not obvious and supervised despite the stated unsupervised learning. We used the averaged weights from all heads of the last layer, we also removed the weights normalization step. Additionally, enumeration in the new algorithm is done in a breadth-first search (BFS) manner. Each next mapped atom should be bond-connected to already mapped product atoms. The first atom in each product molecule or in each connected subgraph is chosen without restrictions.

Results and discussion

Model training

The first part of the model is pretrained on the original graphormer PubChemQC HOMO-LUMO gap data[34,35] in order to learn reasonable atoms in molecule embeddings. Training and

validation MAE was 0.0729 and 0.0978 respectively. The reported in Graphormer MAE 0.0253 and 0.0865 correspondingly, which indicates overfitting.

Pretrained model stacked with reaction-level part and trained on MLM task on all reaction data except “Golden” dataset. After 36 hours of training on 8 NVIDIA A100 GPU with 40 CPU cores and 100 GB RAM (5 epochs, batch size 10, lr $2.5*10^{-5}$) total loss (sum of cross-entropy losses of atom and neighbors MLMs) decreased to 0.003. Additional validation was not carried out due to the dynamic input randomization.

Attention guided atom mapping

Adopting the original AAM algorithm to our network[1] provided us with the tool, which fails to map symmetric structures. A strategy with selecting the highest attention leads to the independent mapping of substructures in a molecule. This results in a possible switch of substructures, what looks like pseudo-rearrangement reactions. For solving this we implemented BFS-like mapping (breadth-first search).

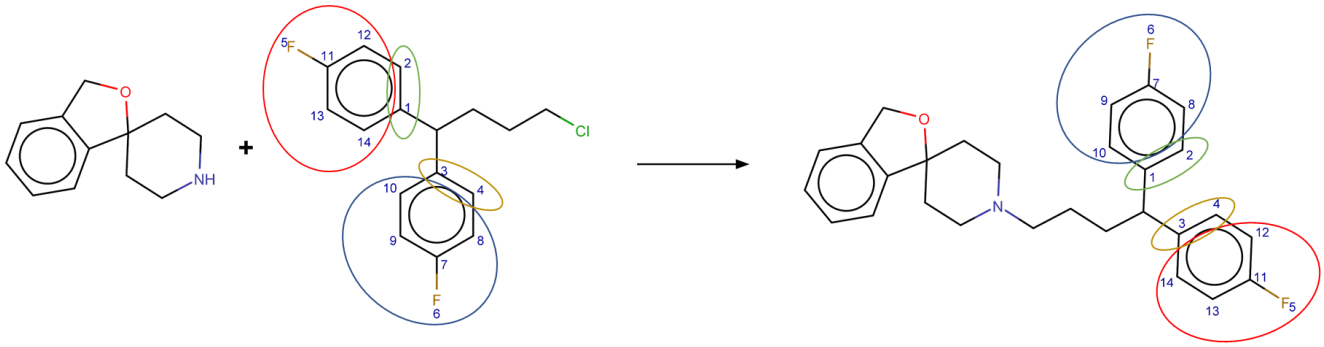
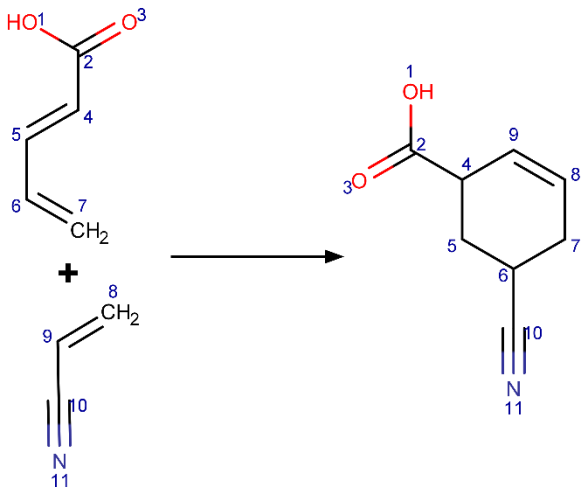


Figure 5. Example of the mapping of symmetric compounds

The normalization of attention weights in the original algorithm[1] depends on the number of atoms of each type. For this reason, unique and rare atoms are always mapped first, however frequent atoms can have higher correspondence, as in the example below. In the current work, the

normalization step was removed, which allowed us to increase the number of correctly mapped reactions in the benchmarked dataset.



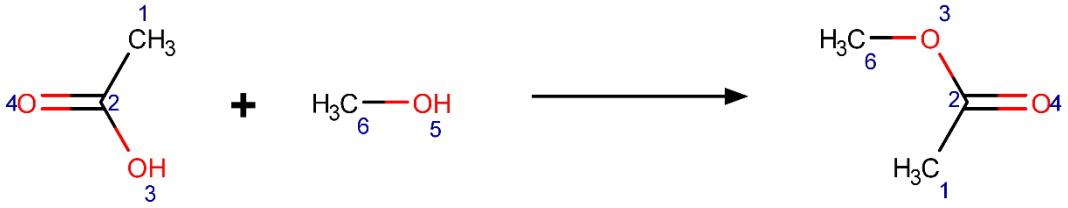
	1	2	3	4	5	6	7	8	9	10	11
1	0.15	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.26	0.00	0.06	0.05	0.04	0.05	0.06	0.04	0.07	0.00
3	0.09	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.05	0.00	0.14	0.08	0.05	0.07	0.07	0.08	0.04	0.00
5	0.00	0.03	0.00	0.10	0.08	0.05	0.12	0.14	0.09	0.03	0.00
6	0.00	0.04	0.00	0.10	0.09	0.08	0.09	0.07	0.12	0.05	0.00
7	0.00	0.04	0.00	0.09	0.09	0.06	0.16	0.08	0.09	0.05	0.00
8	0.00	0.05	0.00	0.07	0.10	0.15	0.10	0.08	0.10	0.05	0.00
9	0.00	0.05	0.00	0.07	0.12	0.08	0.10	0.07	0.09	0.05	0.00
10	0.00	0.08	0.00	0.04	0.05	0.04	0.04	0.04	0.03	0.23	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10

Figure 6. Attention weights for Diels-Alder reaction. (Top) Numbers correspond to atoms identification number (not AAM); (Attention weights) Rows – product atom IDs, Columns – reactants atom IDs.

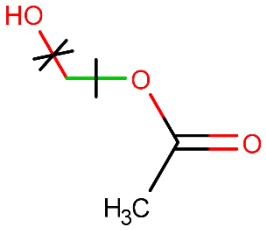
Heuristic mapping for difficult cases

Due to the unsupervised nature of model learning, for some reactions the obtained AAM is mechanistically incorrect, e.g. esterification; or not aligned with chemical sense, e.g. acids to alcohols reduction (see figure below E and F). Therefore, we fixed in our earlier work these reactions mapping by predefined heuristic rules.[2] In the new algorithm, we extract only reaction centers (RC) and convert them into the CGR SMILES at the first step from a prepared CGR. From the list of heuristic rules, we select rules with the same RC CGR SMILES. At the next step, by

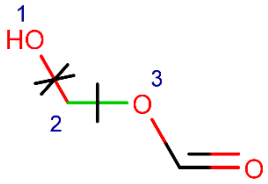
isomorphism search, we find matched pair of rules and CGR. Rule-to-CGR mapping we use for remapping the wrong reaction by the given fix. For example, rule on the figure code changing of atom number 3 to 1, so, matched by isomorphism corresponding atoms in the reaction are 3 and 5. Changing the product atom number from 3 to 5 gives the correct AAM.



A)

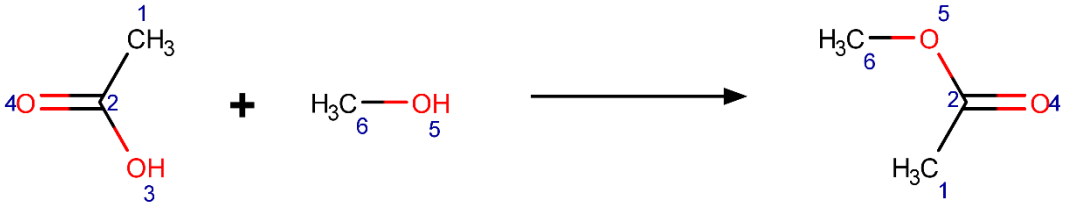


B)

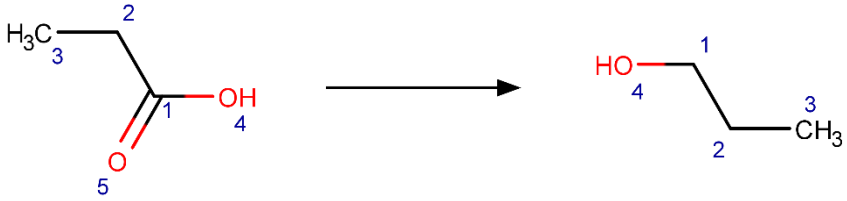


C)

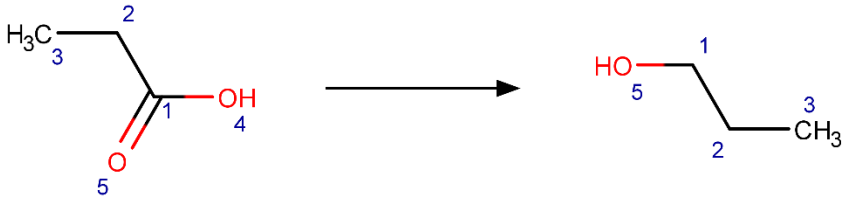
RC SMILES	O[.>-]C[->.]O
Fix	3 > 1
Changed AAM	3 > 5



D)



E)



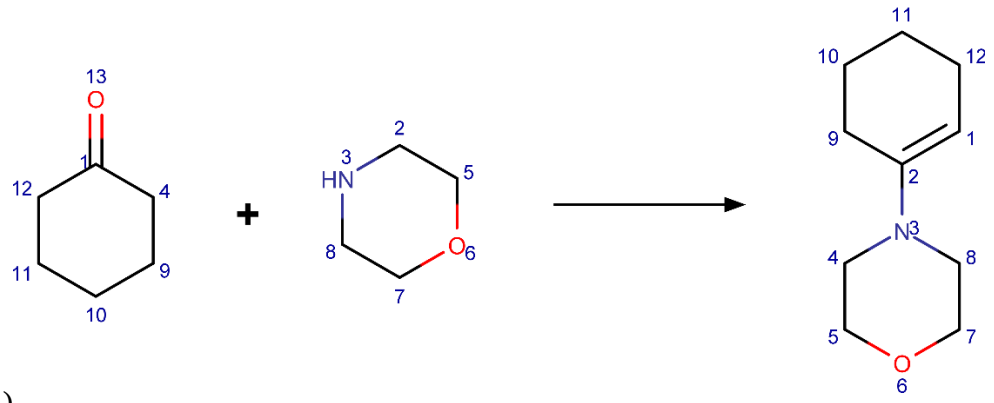
F)

Figure 7. Invalid AAM (A), CGR (B) heuristic rule (B), and Correct AAM (D); reduction not aligned with chemical sense (E), preferrable reduction AAM (F).

Benchmarking studies

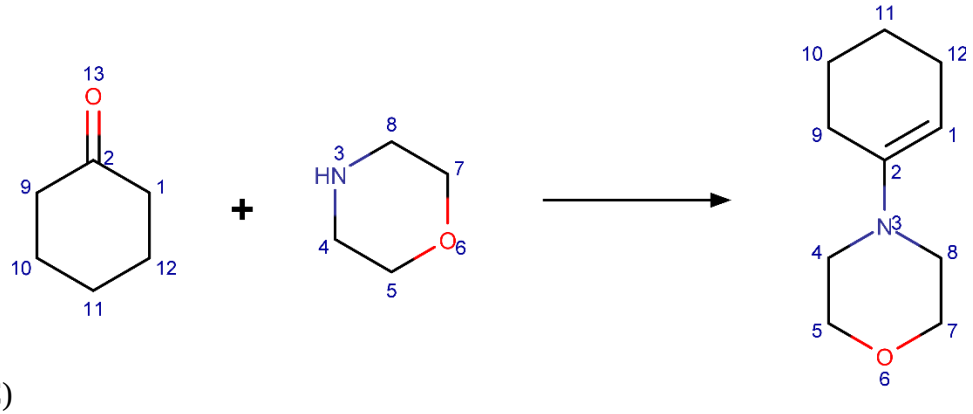
We evaluated the quality of our AAM algorithm with the previously published "Golden" dataset[2]. We compared AAM with IBM RxnMapper[1]. For 1851 records new algorithm gives 1656 correct AAM (5.0 % better) and 187 unique correct.

1851 reactions	This work	IBM RxnMapper
Total correct	1656	1565
Unique correct	187	96
Percent correct	89.5	84.5



A)

B) C1COCCN1.O=C1CCCCC1>>C1=C(N2CCOCC2)CCCC1



C)

D)

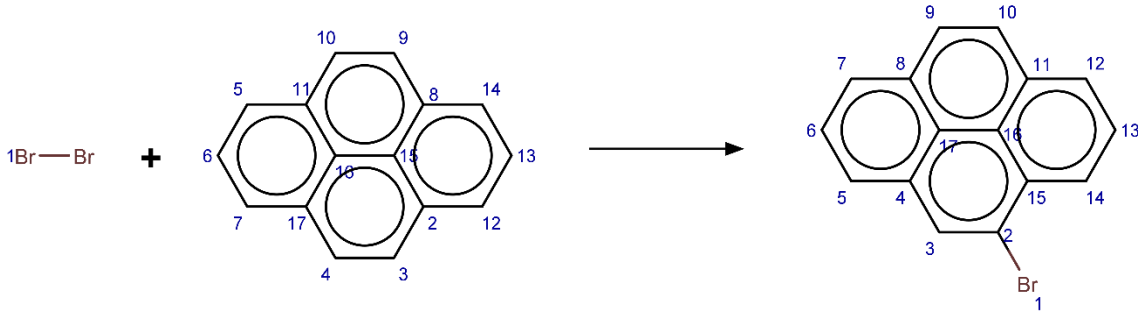
	13	2	1	12	11	10	9	8	7	6	5	4	3
1	0.00	0.11	0.14	0.07	0.07	0.07	0.14	0.07	0.06	0.00	0.06	0.07	0.00
2	0.00	0.13	0.07	0.05	0.05	0.05	0.07	0.06	0.05	0.00	0.05	0.06	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12
4	0.00	0.05	0.05	0.04	0.04	0.04	0.05	0.08	0.05	0.00	0.05	0.08	0.00
5	0.00	0.05	0.04	0.03	0.04	0.03	0.04	0.06	0.07	0.00	0.07	0.06	0.00
6	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00
7	0.00	0.05	0.04	0.03	0.04	0.03	0.04	0.06	0.07	0.00	0.07	0.06	0.00
8	0.00	0.05	0.05	0.04	0.04	0.04	0.05	0.08	0.05	0.00	0.05	0.08	0.00
9	0.00	0.06	0.08	0.06	0.07	0.06	0.08	0.05	0.05	0.00	0.05	0.05	0.00
10	0.00	0.04	0.05	0.09	0.07	0.09	0.05	0.04	0.04	0.00	0.04	0.04	0.00
11	0.00	0.04	0.06	0.07	0.12	0.07	0.06	0.04	0.04	0.00	0.04	0.04	0.00
12	0.00	0.05	0.06	0.09	0.06	0.09	0.06	0.04	0.04	0.00	0.04	0.04	0.00

Figure 8. Invalid AAM from IBM RxnMapper (A), B) SMILES used by mapper, C) Correct AAM, D) Graph attention. Rows – product atom numbers, Columns – reactants atom numbers.

As the AAM in RxnMapper is based on SMILES sequences does the order of the sequence influence the errors. Interestingly, the mapper pays high attention to '=C' substrings on both sides of the reaction and marks these atoms as the same. In opposite, the graph-based attention is invariant to the atom order and does not use bond order information, and works with full neighbors context, which can be seen in the table above. For example, atom 1 (rows) in the product has high and correct equal attention to atoms 1 and 9 (columns). A similar pattern is observed for all atoms.

For multicondensed rings, RxnMapper also returns an invalid mapping due to drastic differences in canonical SMILES of products and reactants. However, by disabling of automated

canonicalization of SMILES string (especially handwritten), with the same order of atoms, the returned AAM is correct. Our algorithm in this case returns correct AAM.



A)

B) BrBr.c1cc2ccc3cccc4ccc(c1)c2c34>>Brc1cc2cccc3ccc4cccc1c4c32

C) BrBr.c1cc2ccc3cccc4ccc(c1)c2c34>>c1cc2c(Br)cc3cccc4ccc(c1)c2c34

Figure 9. Invalid AAM in condensed rings (A), canonicalized SMILES (B), handwritten SMILES (C)

Additionally, we compared AAM on USPTO dataset for top-100 correct reaction centers. Obtained by RxnMapper centers contains 303457 records. Our algorithm generates 310570 records (2.3 % better).

Conclusions

Here we present a new algorithm *GraphormerMapper* for reaction atom-to-atom mapping based on attention neural network and heuristics expansion. This algorithm outperforms IBM RxnMapper, currently the best freely available atom-to-atom mapper. Graph attention implemented in Python neural network library – Chytorch, that is easy to use and modify for specific goals, such as reaction yield prediction or conditions prediction. Atom-to-atom mapping implemented in Chython Python library – a framework for reaction and molecule processing.

Code availability

Described Python packages are freely available at GitHub: https://github.com/chython/chython, https://github.com/chython/chytorch and https://github.com/chython/chytorch-rxnmap. The repositories contain the source code, as well as the data processing protocol.

References

(1) Schwaller, P.; Hoover, B.; Reymond, J.-L.; Strobelt, H.; Laino, T. Extraction of Organic Chemistry Grammar from Unsupervised Learning of Chemical Reactions. *Science Advances* **2021**, *7* (15), eabe4166. https://doi.org/10.1126/sciadv.abe4166.

(2) Lin, A.; Dyubankova, N.; Madzhidov, T. I.; Nugmanov, R. I.; Verhoeven, J.; Gimadiev, T. R.; Afonina, V. A.; Ibragimova, Z.; Rakhimbekova, A.; Sidorov, P.; Gedich, A.; Suleymanov, R.; Mukhametgaleev, R.; Wegner, J.; Ceulemans, H.; Varnek, A. Atom-to-atom Mapping: A Benchmarking Study of Popular Mapping Algorithms and Consensus Strategies. *Molecular Informatics* **2021**, 2100138. https://doi.org/10.1002/minf.202100138.

(3) PyTorch https://pytorch.org/ (accessed 2022 -02 -01).

(4) Fontain, E. The Problem of Atom-to-Atom Mapping. An Application of Genetic Algorithms. *Analytica Chimica Acta* **1992**, *265* (2), 227–232. https://doi.org/10.1016/0003-2670(92)85028-5.

(5) Chen, W. L.; Chen, D. Z.; Taylor, K. T. Automatic Reaction Mapping and Reaction Center Detection. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3* (6), 560–593. https://doi.org/10.1002/WCMS.1140.

(6) Law, J.; Zsoldos, Z.; Simon, A.; Reid, D.; Liu, Y.; Knew, S. Y.; Johnson, A. P.; Major, S.; Wade, R. A.; Ando, H. Y. Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated Retrosynthetic Rule Generation. *Journal of Chemical Information and Modeling* **2009**, *49* (3), 593–602. https://doi.org/10.1021/CI800228Y.

(7) Bøgevig, A.; Federsel, H. J.; Huerta, F.; Hutchings, M. G.; Kraut, H.; Langer, T.; Löw, P.; Oppawsky, C.; Rein, T.; Saller, H. Route Design in the 21st Century: The IC SYNTH Software Tool as an Idea Generator for Synthesis Prediction. *Organic Process*

Research and Development **2015**, *19* (2), 357–368. https://doi.org/10.1021/OP500373E/SUPPL_FILE/OP500373E_SI_001.PDF.

(8) Segler, M. H. S.; Waller, M. P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chemistry – A European Journal* **2017**, *23* (25), 5966–5971. https://doi.org/10.1002/CHEM.201605499.

(9) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Accounts of Chemical Research* **2018**, *51* (5), 1281–1289. https://doi.org/10.1021/ACS.ACCOUNTS.8B00087.

(10) Fortunato, M. E.; Coley, C. W.; Barnes, B. C.; Jensen, K. F. Data Augmentation and Pretraining for Template-Based Retrosynthetic Prediction in Computer-Aided Synthesis Planning. *Journal of Chemical Information and Modeling* **2020**, *60* (7), 3398–3407.
https://doi.org/10.1021/ACS.JCIM.0C00403/SUPPL_FILE/CI0C00403_SI_001.PDF.

(11) Hendrickson, J. B.; Chen, L. Reaction Classification. In *Encyclopedia of Computational Chemistry*; John Wiley & Sons, Ltd: Chichester, UK, 2002. https://doi.org/10.1002/0470845015.cca022.

(12) Chen, L.; Nourse, J. G.; Christie, B. D.; Leland, B. A.; Grier, D. L. Over 20 Years of Reaction Access Systems from MDL: A Novel Reaction Substructure Search Algorithm. *Journal of Chemical Information and Computer Sciences* **2002**, *42* (6), 1296–1310. https://doi.org/10.1021/CI020023S.

(13) Lin, A. I.; Madzhidov, T. I.; Klimchuk, O.; Nugmanov, R. I.; Antipin, I. S.; Varnek, A. Automatized Assessment of Protective Group Reactivity: A Step toward Big Reaction Data Analysis. *Journal of Chemical Information and Modeling* **2016**, *56* (11), 2140–2148. https://doi.org/10.1021/ACS.JCIM.6B00319/SUPPL_FILE/CI6B00319_SI_001.PDF.

(14) Sidorov, P.; Varnek, A.; Gimadiev, T.; Nugmanov, R.; Batyrshin, D.; Madzhidov, T.; Maeda, S. Combined Graph/Relational Database Management System for Calculated Chemical Reaction Pathway Data. *Journal of Chemical Information and Modeling* **2021**, *61* (2), 554–559. https://doi.org/10.1021/ACS.JCIM.0C01280/SUPPL_FILE/CI0C01280_SI_001.PDF.

(15) Gimadiev, T.; Nugmanov, R.; Khakimova, A.; Fatykhova, A.; Madzhidov, T.; Sidorov, P.; Varnek, A. CGRdb2.0: A Python Database Management System for Molecules,

Reactions, and Chemical Data. *Journal of Chemical Information and Modeling* **2021**. https://doi.org/10.1021/ACS.JCIM.1C01105/SUPPL_FILE/CI1C01105_SI_001.PDF.

(16) Indigo Toolkit https://lifescience.opensource.epam.com/indigo/index.html (accessed 2022 -02 -01).

(17) AutoMapper https://docs.chemaxon.com/display/docs/automapper-user-s-guide.md (accessed 2022 -02 -01).

(18) Rahman, S. A.; Torrance, G.; Baldacci, L.; Martínez Cuesta, S.; Fenninger, F.; Gopal, N.; Choudhary, S.; May, J. W.; Holliday, G. L.; Steinbeck, C.; Thornton, J. M. Reaction Decoder Tool (RDT): Extracting Features from Chemical Reactions. *Bioinformatics* **2016**, *32* (13), 2065–2066. https://doi.org/10.1093/bioinformatics/btw096.

(19) NextMove https://www.nextmovesoftware.com/ (accessed 2022 -02 -01).

(20) Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; Liu, T.-Y. Do Transformers Really Perform Badly for Graph Representation? In *Advances in Neural Information Processing Systems*; Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J. W., Eds.; 2021.

(21) Lunnon, W. F.; Brunvoll, J.; Cyvin, S. J.; Cyvin, B. N.; Balaban, A. T. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *Journal of Chemical Information and Computer Sciences* **1988**, *28* (1), 31–36. https://doi.org/10.1021/CI00057A005.

(22) Bjerrum, E. J.; Sattarov, B. Improving Chemical Autoencoder Latent Space and Molecular De Novo Generation Diversity with Heteroencoders. *Biomolecules* **2018**, 1–17. https://doi.org/10.3390/biom8040131.

(23) Berthelot, D.; Raffel, C.; Roy, A.; Goodfellow, I. Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer. **2018**.

(24) Hoonakker, F.; Lachiche, N.; Varnek, A.; Wagner, A. A Representation to Apply Usual Data Mining Techniques to Chemical Reactions — Illustration on the Rate Constant of SN2 Reactions in Water. *http://dx.doi.org/10.1142/S0218213011000140* **2012**, *20* (2), 253–270. https://doi.org/10.1142/S0218213011000140.

(25) Madzhidov, T. I.; Polishchuk, P. G.; Nugmanov, R. I.; Bodrov, A. v.; Lin, A. I.; Baskin, I. I.; Varnek, A. A.; Antipin, I. S. Structure-Reactivity Relationships in Terms of

the Condensed Graphs of Reactions. *Russian Journal of Organic Chemistry 2014 50:4* **2014**, *50* (4), 459–463. https://doi.org/10.1134/S1070428014040010.

(26) Bort, W.; Baskin, I. I.; Gimadiev, T.; Mukanov, A.; Nugmanov, R.; Sidorov, P.; Marcou, G.; Horvath, D.; Klimchuk, O.; Madzhidov, T.; Varnek, A. Discovery of Novel Chemical Reactions by Deep Generative Recurrent Neural Network. *Scientific Reports* **2021**, *11* (1). https://doi.org/10.1038/S41598-021-81889-Y.

(27) Nugmanov, R. I.; Mukhametgaleev, R. N.; Akhmetshin, T.; Gimadiev, T. R.; Afonina, V. A.; Madzhidov, T. I.; Varnek, A. CGRtools: Python Library for Molecule, Reaction, and Condensed Graph of Reaction Processing. *Journal of Chemical Information and Modeling* **2019**, *59* (6), 2516–2521. https://doi.org/10.1021/ACS.JCIM.9B00102/SUPPL_FILE/CI9B00102_SI_001.PDF.

(28) Jaworski, W.; Szymkuć, S.; Mikulak-Klucznik, B.; Piecuch, K.; Klucznik, T.; Kaźmierowski, M.; Rydzewski, J.; Gambin, A.; Grzybowski, B. A. Automatic Mapping of Atoms across Both Simple and Complex Chemical Reactions. *Nature Communications 2019 10:1* **2019**, *10* (1), 1–11. https://doi.org/10.1038/s41467-019-09440-2.

(29) Chemical reactions from US patents (1976-Sep2016) https://figshare.com/articles/dataset/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873 (accessed 2022 -02 -01).

(30) Gimadiev, T. R.; Lin, A.; Afonina, V. A.; Batyrshin, D.; Nugmanov, R. I.; Akhmetshin, T.; Sidorov, P.; Duybankova, N.; Verhoeven, J.; Wegner, J.; Ceulemans, H.; Gedich, A.; Madzhidov, T. I.; Varnek, A. Reaction Data Curation I: Chemical Structures and Transformations Standardization. *Molecular Informatics* **2021**, *40* (12), 2100119. https://doi.org/10.1002/MINF.202100119.

(31) Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R.; Research, G. ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations. **2019**.

(32) Devlin, J.; Chang, M. W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference* **2018**, *1*, 4171–4186.

(33) Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V.; Allen, P. G. RoBERTa: A Robustly Optimized BERT Pretraining Approach. **2019**.

(34) Nakata, M.; Shimazaki, T. PubChemQC Project: A Large-Scale First-Principles Electronic Structure Database for Data-Driven Chemistry. *Journal of Chemical Information and Modeling* **2017**, *57* (6), 1300–1308. https://doi.org/10.1021/ACS.JCIM.7B00083.

(35) Hu, W.; Fey, M.; Ren, H.; Nakata, M.; Dong, Y.; Leskovec, J. OGB-LSC: A Large-Scale Challenge for Machine Learning on Graphs. **2021**.

Table of Contents graphic