# Molecular Transformer Unifies Reaction Prediction and Retrosynthesis across Pharma Chemical Space

Alpha A. Lee, Qingyi Yang, Vishnu Sresht, Peter Bolgar, Xinjun Hou, Jacquelyn L Klug-McLeod, and Christopher R. Butler

2021.6.18

# Outlines

# Methods to Synthesise Complex Molecules

- template-based
  - Handcrafting/ automatic template extraction algorithms: rely on pre-existing atom mapping that maps atoms in the reactants to atoms in the product.
- template-free
  - graph-based: require atom-mapped datasets to generate the ground truth for training, and atom mapping algorithms make use of reaction templates.
  - sequence-based: use a text representation of the reactants, reagents and products (usually SMILES), and treat reaction prediction as machine translation from one language (reactants-reagents) to another language (products).

# Problems in Existing Methods

- The Molecular Transformer model outperforms all methods in the literature in reaction prediction, but retrosynthesis has not been described.
- Beyond accuracy, dataset bias is a major challenge in validating machine learning models in chemistry applications. The training and test sets are harvested from medicinal chemistry patents and are very similar.

# Contribution

- This is the first validation of reaction prediction tools with industrial pharma data.
- Molecular Transformer is the only framework that accurately tackles both reaction prediction and retrosynthesis.

# Molecular Transformer

The transformer is a step-wise autoregressive encoder-decoder model comprised of a combination of multi-head attention layers and positional feed forward layers.

# Encoder and Decoder of Molecular Transformer

- In the encoder, the multi-head attention layers attend the input sequence and encode it into a hidden representation.
- The decoder consists of two types of multi-head attention layers.
  - The first is masked and attends only the preceding outputs of the decoder.
  - The second multi-head attention layer attends encoder outputs, as well as the output of the first decoder attention layer.
  - Multi-head attention allows the encoder and decoder to peek at different tokens simultaneously.

# Datasets

- USPTO: contains stereochemical information.
- PF-ELN: a set of proprietary electronic lab notebook data from internal medicinal chemistry projects in Pfizer. This dataset contains 147,392 reactions with 2 reactants.
- USPTO−R: 50,000 reactions drawn from diverse reaction classes in USPTO as a benchmark for retrosynthesis models.

# Test of Generalisability: Reaction Prediction

Baseline: Weisfeiler-Lehman Difference Network (WLDN)

| Model | Training set | Test set | Top-1 [% acc.] | Top-2 [% acc.] | Top-3 [% acc.] |
|---|---|---|---|---|---|
| Molecular Transformer | PF – ELN | PF – ELN | 97.0% | 98.5 % | 98.8% |
| Molecular Transformer | USPTO | PF – ELN | 69.0% | 80.3% | 82.9% |
| WLDN[11] | USPTO | PF – ELN | 50.4% | 51.7% | 52.2% |

**Table 1** Molecular Transformer accurately predicts products of reactions given reactants and is generalisable across chemical space. The table show the top-$k$ accuracy of Molecular Transformer tested on proprietary electronic lab notebook data for different training sets.

# Test of Accuracy: Retrosynthesis Prediction

| Model | Training set | Test set | Top-1 [% acc] | Top-2 [% acc] | Top-3 [% acc] |
|---|---|---|---|---|---|
| Molecular Transformer | USPTO – R | USPTO – R | 43.8% | 56.0% | 60.5% |
| Coley et al.[5] | USPTO – R | USPTO – R | 37.3% | – | 54.7% |
| Liu et al.[14] | USPTO – R | USPTO – R | 37.4% | – | 52.4% |

**Table 2** Molecular Transformer for retrosynthesis accurately predicts reactants of reactions given products and outperforms the state-of-the-art. The table show the top-$k$ accuracy of Molecular Transformer trained and tested on USPTO – R. The top-2 accuracy is not reported in refs[5,14].

The Molecular Transformer outperforms the benchmark model, which uses a template-based approach.

# Test of Generalisability: Retrosynthesis Prediction

| Model | Training set | Test set | Top-1 [% acc] | Top-2 [% acc] | Top-3 [% acc] |
|---|---|---|---|---|---|
| Molecular Transformer | USPTO − R | PF − ELN | 31.5% | 40.5% | 44.3% |
| Molecular Transformer | PF − ELN | PF − ELN | 91.0% | 94.5% | 94.5% |

**Table 3** Molecular Transformer for retrosynthesis is accurate for chemistries that are industrially relevant, and generalises across chemical space. The table show the top-$k$ accuracy of Molecular Transformer with different training set-test set combinations.

# Conclusions

- Molecular Transformer is a versatile framework that tackles both reaction prediction and retrosynthesis. It is
- A key advantage of our framework is that it can be easily extended to tackle problems such as predicting reaction condition, reagents and yield.