# Reaction Classification by Reaction Vectors

G. Ghiandoni [1], B. Chen [2], M. J. Bodkin [3], V. J. Gillet [1]

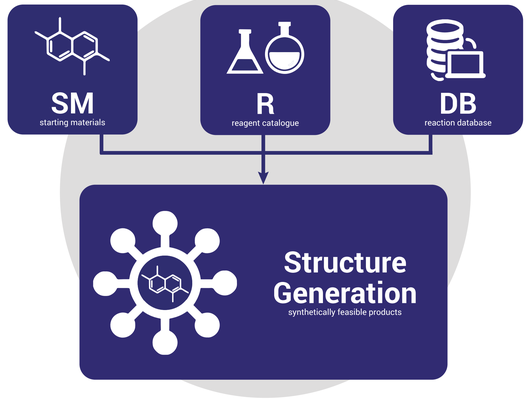
[1] Information School, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP, United Kingdom, [2] Chemistry Department, University of Sheffield, Dainton Building, Brook Hill, Sheffield, S3 7HF, United Kingdom
[3] Evotec (U.K.) Ltd, 114 Innovation Drive, Milton Park, Abingdon, OX14 4RZ, United Kingdom

## Reaction-vector-based Drug Design

The potential number of chemical structures, that meet Lipinski's Rule-of-Five (RO5) requirements for drug-likeness, has been estimated at $10^{60}$ molecules. [1]

Fragment-based de novo design techniques, that exploit databases of real reaction examples such as ELNs (electronic laboratory notebooks), have drastically reduced this huge space into a smaller number of structures. [2]

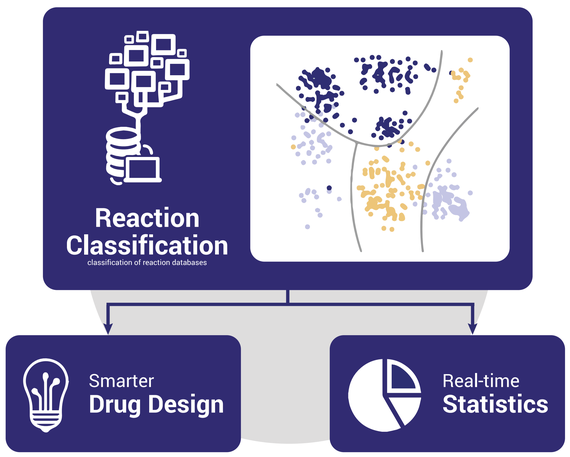
In reaction vector de novo design, the structure generation process occurs through the combination of a specified set of starting materials (SM), a reagent catalogue (R) and a reaction database (DB): [3, 4]




## Reaction Classification in Drug Design

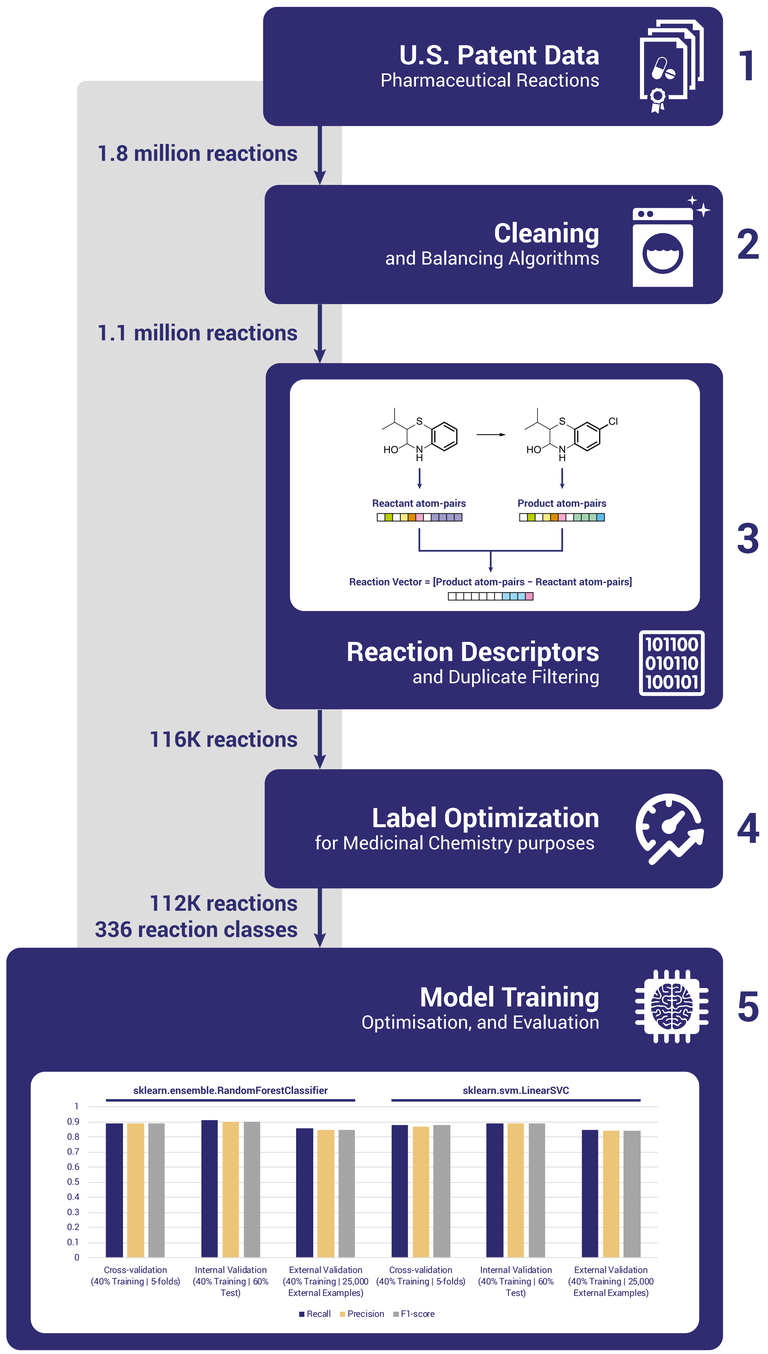
Reaction classification is intended as a labelling technique that will enable the classification of a large number of apparently different objects (i.e. reactions) into a smaller number of classes (e.g. Bond Formations, Rearrangements, etc.).

This method permits efficient organisation and retrieval of the information from a given database, instead of simply applying all the reactions in one-shot, along with a real-time statistical evaluation of the reaction class composition.




## Reaction-vector-based Machine Learning

Reaction vectors are directly convertible into sets of reactions descriptors (RDs) that can be used to encode sets of already classified reactions. Descriptors and classes can be used to train supervised machine learning algorithms in order to build efficent, fast, and scalable classification models.



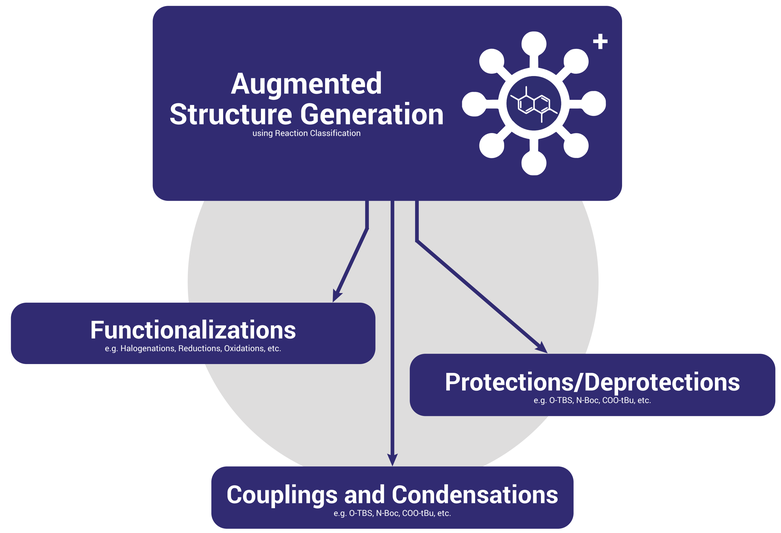

## MedChem-like Layered Labelling

The labelling system is based on a manually curated hierarchical classification tree, in order to provide a very flexible and detailed reaction cataloguing. The level of detail is distributed across 4 layers, ranging from general reaction categories to increasingly more specific reaction sub-classes and reactant descriptions:

| Layer 1 | Layer 2 | Layer 3 | Layer 4 |
|---|---|---|---|
| **C-C Bond Formation** | **Coupling** | **Suzuki** | **Bromo** |

## Fine-grained Synthesis Selection

Reaction classification can be applied to any de novo design context where reaction information is available. This methodology offers fine-grained selection of products from sets of specific reaction classes, in order to separate them through different data processing pathways:




## Acknowledgements

This project is funded by Evotec (U.K.) and the Engineering and Physical Sciences Research Council (EPSRC). Special thanks to Dr Daniel Lowe and the Sheffield Chemoinformatics Group.

## References

1. Bohacek, R.S., McMartin, C. & Guida, W.C., The art and practice of structure-based drug design: A molecular modeling perspective. Medicinal Research Reviews. 1996, 16, 3-50.

2. Hartenfeller, M., Reaction-Driven De Novo Design: a Keystone for Automated Design of Target Family-Oriented Libraries. In De novo Molecular Design; Schneider, G., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2013; 245-266.

3. Patel, H., Bodkin, M. J., Chen, B. & Gillet, V. J., Knowledge-Based Approach to De Novo Design Using Reaction Vectors. Journal of Chemical Information and Modeling. 2009, 49, 1163-1184.

4. Hristozov, D., Bodkin, M., Chen, B., Patel, H., & Gillet, V. J., Validation of Reaction Vectors for de Novo Design. Library Design, Search Methods, and Applications of Fragment-Based Drug Design, 2011, 29-43.