

Selecting druglike pieces for the virtual chemistry jigsaw puzzle: towards optimal fragment spaces

Christof Gerlach¹

Jörg Degen², Matthias Rarey² and Andrea Zaliani²

¹*Bayer-Schering Pharma AG, Med Chem VII-Comp. Chemistry, 13342 Berlin, Germany*

²*Universität Hamburg, Zentrum für Bioinformatik, 20146 Hamburg, Germany*

Fragment-based approaches have become very popular within the lead finding phase of a drug design project. Different experimental techniques such as X-ray and NMR-supported protocols have been developed to detect and applied to successfully novel lead structures [1]. In addition, *in silico* approaches considering either descriptor- [2], ligand- [3] or structure-based [4] information for navigating within chemical fragment spaces have been established.

Still the question remains, how does a typical ‘druglike’ fragment looks like and from which source it should be derived?

We present our results from a comparison of two retrosynthetic sets of rules for the generation of fragment spaces. RECAP [5] was checked against our newly developed procedure BRICS (Breaking of Retrosynthetically Interesting Chemical Substructures). Within BRICS the shredding of molecules to fragments tries to mirror retrosynthetical concepts in a more elaborate way considering, for instance, bioisosteric replacements for cyclic and acyclic systems separately and also differentiates between activated and inactivated heterocyclic systems.

For a detailed analysis, fragment spaces from WDI [6] and from the ZINC database [7] were derived by both sets of rules. These datasets were characterized with respect to the number of generated fragments, connection points and size of the fragments. Finally, identical fragments which occur in both datasets have been compiled to generate an optimal fragment space consisting of approximately 5000 fragments.

The performance of these sets (generated by RECAP and BRICS) were evaluated by means of multiple FTree-FS searches using very large and diverse query sets. The BRICS-generated fragment space was able to exactly rebuild more than the double amount of query molecules in comparison with the RECAP-generated fragment space. Thus, the better performing BRICS-generated fragment space have been further enriched with fragments from ZINC having a reasonably high similarity to the WDI fragments. This led to two larger fragment spaces showing further improvements with respect to exact rebuilding of the query

In conclusion, our analysis underlines that the performance of a fragment space derived from ‘druglike’ molecules can be improved, using fragments which are originally derived from vendor catalogues. Thus, it seems that a high-performance set of fragments does not have to be derived solely from databases of drug molecules. Based on our findings three new fragment sets have been compiled, with different optimized performances in retrieving random sets of queries from different sources We also plan to make them publicly available in the near future. These can be used for further fragment-based searches to identify chemical probes for a given protein binding assay.

1. Hajduk, P. J.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **2007**, 6, (3), 211-9.
2. Pärn, J.; Degen, J.; Rarey, M., Exploring fragment spaces under multiple physicochemical constraints. *J Comput Aided Mol Des* **2007**, 21, (6), 327-340.

3. Rarey, M.; Stahl, M., Similarity searching in large combinatorial chemistry spaces. *J Comput Aided Mol Des* **2001**, 15, (6), 497-520.
4. Degen, J.; Rarey, M., FlexNovo: structure-based searching in large fragment spaces. *ChemMedChem* **2006**, 1, (8), 854-68.
5. Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M., RECAP--retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci* **1998**, 38, (3), 511-22.
6. *World Drug Index, Version 2004*, Thomson: Philadelphia PA, 2004.
7. Irwin, J. J.; Shoichet, B. K., ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* **2005**, 45, (1), 177-82.