REAL-ly large-scale Virtual Screening -
Traversing Enormous Regions of Chemical Space with the GPU and CPU

Gunther Stahl

OpenEye Scientific Software Ltd. Zweigniederlassung Deutschland

The sheer size of the universe of molecules (variously estimated to be between 10^{40} and 10^{60} [1]) is daunting. The number of molecules in a particular library or collection has, until recently, dictated the methods that can be used to search it; only very rapid graph-based methods were fast enough to search hundreds of millions of molecules in a reasonable time and with reasonable computational resources [2]. However, the advent of the cloud as a routine compute resource enables new approaches to the rapid searching of previously intractably large chemical spaces, as it allows cost-effective access to both CPU and GPU-enabled search.

In this presentation we will demonstrate shape similarity searching with FastROCS [3], the GPU version of the widely used lead discovery tool ROCS [4]. We will briefly present how porting to the GPU enabled us to accelerate shape searching by over 3 orders of magnitude and yet maintain identical virtual screening performance. The unprecedented speed of the current version of FastROCS (> 20 million molecules/GPU/minute) enables searching of enormous libraries in a matter of minutes. We will illustrate this capability by searching parts of the Enamine REAL library [5].

Searching the full Enamine REAL database (1.4 Billion enumerated molecules) requires the great parallelizability of computation on the cloud. We will present results on searching the entirety of the REAL library using both FastROCS and OpenEye's high-throughput docking tool [6], FRED. To our knowledge these are the largest scale virtual screens so far conducted, being over an order of magnitude larger than the largest screens reported to date [7].

[1] J.L. Reymond, M. Awale, *ACS Chem Neurosci*, **2012**, *3(9)*, 649-57
[2] D. Probst, J.L. Reymond, *J. Cheminform*, **2018**, *10(1)*, 66
[3] https://docs.eyesopen.com/toolkits/python/fastrocstk/introduction.html
[4] P.C.D. Hawkins, A.G. Skillman, A. Nicholls, *J Med Chem,* **2007**, *50(1)*, 74-82
[5] https://enamine.net/index.php?option=com_content&task=view&id=254
[6] M. McGann, *J Comput Aided Mol Des*, **2012**, *26(8)*, 897-906
[7] J. Lyu, S. Wang, et al, *Nature*, **2019**, *566(7743)*, 224-229