CHEMICAL SIMILARITY SEARCHING ON FPGA

Peter Szántó, Béla Fehér, Attila Katona, Tibor Fortenbacher, Ivan Bangov, Csaba Mészáros, Attila Bérces* (Budapest University of Technology and Economics, evopro Kft. and Chemistry Logic Ltd., email: attila.berces@chemistrylogic.com, tel: +36 96 769305, Chemistry Logic Kft. Szűk utca 9, Nyúl, 9082, Hungary)

Chemical similarity searching involves the calculation of a similarity measure between the query molecules and all molecules contained in databases. On its own, this is a fast computation but the computational demand multiplies when users search against multiple queries as well as consider an entire database as a query set against another database. We have developed highly efficient implementations of chemical similarity searching on FPGA.

The molecular structures are represented by a hashed binary fingerprint generated from the structural graph. The similarity measure is the Tanimoto coefficient. Our implementation involves 128 parallel processing elements on a singel Virtex 4 LX200 FPGA on the Silicon Graphics RC100 platform. The performance of this implementation is limited by the data bandwidth. We compared the performance of our implementation with that of a commercial software from Chemaxon. Our reference calculation consists of a query containing 128 molecules searched in a database of 10 million compounds. While the commercial software running on a single core 3.2 GHz CPU took 30 minutes, our implementation took 14 seconds.