

Poster presentation

Analyzing molecular landscapes using random walks and information theory

M Emmerich*, BVY Li, A Bender, E Faddiev, J Kruisselbrink, AH Deutz, E van der Horst, A IJzerman and Th Bäck

Address: LIACS, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

* Corresponding author

from 4th German Conference on Chemoinformatics
Goslar, Germany. 9–11 November 2008

Published: 5 June 2009

Chemistry Central Journal 2009, 3(Suppl 1):P20 doi:10.1186/1752-153X-3-S1-P20

This abstract is available from: <http://www.journal.chemistrycentral.com/content/3/S1/P20>

© 2009 Emmerich et al; licensee BioMed Central Ltd.

Search heuristics for the *in silico* discovery of drug candidates have recently received increased attention. Most of these heuristics such as Simulated Annealing and Evolutionary Algorithms gradually improve molecules by exploiting that the similarity in the molecular structure often relates to the similarity in the properties of molecules such as activity against a target. However, it often remains unproven whether such continuity assumptions actually hold. Generally speaking, there is a need to better understand and assess the properties of molecular search landscapes in order to design/choose appropriate optimization methods to search these spaces.

The theory of *combinatorial landscape analysis* aims to provide such analysis tools. However, many of the methods proposed in this field require the complete knowledge of the landscape and thus are inappropriate for analyzing the huge search spaces of chemical structures. If the size of the search space forbids enumeration, statistical landscape analysis methods are the only available tool.

Following the approach of Vassilev et al. [1] we propose to estimate landscape properties from random walks using the variation operator of the search heuristic. Once a search heuristic is built, these random walks can be generated with little extra effort. The precision of the obtained results scales with the number and length of the random walks available. Given the data from random walks the following analysis methods can be used: (1) Correlation Length Analysis which reveals the validity of the continu-

ity assumption; (2) Information Complexity which reveals the structural diversity of the search landscape; (3) Multimodality measures which estimate the frequency of local optima for different neighbourhood radii, and finally (4) Neutrality measures which account for the size distribution of plateaus in the landscape. Each measure indicates difficulties for optimization routines encountered when optimizing the objective function.

We apply random-walk based landscape analysis for two search spaces in the context of de-novo drug design: Firstly, we study the properties of the search space induced by the mutation operators of the Molecule Evuator™ [2] with an activity model as an objective function. Secondly, we study the properties of a peptide design problem using the software MOE. Here a ligand that binds tightly to a 14-3-3 isoform is searched for.

References

1. Vassilev V, Miller J, Fogarty T: **Information characteristics and the structure of landscapes.** *Evol Comp* 2000, **8(1)**:31-60.
2. Lameijer E-W, Kok JN, Bäck Th, IJzerman AP: **The molecule evaluator. An interactive evolutionary algorithm for the design of drug-like molecules.** *J Chem Inf Modeling* 2006, **46(2)**:545-552.