

ORAL PRESENTATION

Open Access

Text-based similarity searching for hit- and lead-candidate identification

Volker Hähnke

From 7th German Conference on Chemoinformatics: 25 CIC-Workshop
Goslar, Germany. 6-8 November 2011

The Pharmacophore Alignment Search Tool (PhAST) is a string-based approach to virtual screening. Molecules are represented by linear sequences which describe their respective pattern of interaction possibilities. The problem of molecule linearization is tackled by applying Minimum Volume Embedding in combination with a Diffusion Kernel to the molecular graph [1,2]. Linear representations are compared using global pairwise sequence alignment [3]. PhAST exhibited enrichment capabilities comparable or superior to most common virtual screening approaches. Compound rankings were proven to be dissimilar to those of other virtual screening techniques. It was shown that emphasis on key interactions through the application of position specific weights in the alignment process significantly increases enrichment.

Significance of chemical similarity was determined in form of p-values of global alignment scores, calculated in an approach that was adapted from its original application to local sequence alignments of protein sequences utilizing Markov chain Monte Carlo simulation [4]. Bonferroni correction was used to correct p-values with respect to the size of the screening library [5].

PhAST was employed in two prospective applications: A screening for non-nucleoside analogue inhibitors of bacterial thymidine kinase yielded a hit with a distinct structural framework but only weak activity. Screenings for drugs that are not members of the NSAID (non-steroidal anti-inflammatory drug) class as modulators of gamma secretase resulted in a potent modulator with clear structural distinction from the reference compound.

Published: 1 May 2012

Correspondence: volker.hahnke@nih.gov
National Center for Biotechnology Information, National Library of Medicine,
National Institutes of Health, Department of Health and Human Services,
8600 Rockville Pike, Bethesda, MD 20894, USA

References

1. Shaw R, Jebara T: **Minimum Volume Embedding**. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics: 21-24 March 2007; San Juan (Puerto Rico)*. Omnipress; Meila M, Shen X 2007:460-467.
2. Smola AJ, Kondor RI: **Kernels and Regularization on Graphs**. In *Proceedings of the 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop: 24-27 August 2003; Washington DC*. Springer; Schölkopf B, Warmuth, M 2003:144-158.
3. Durbin R, Eddy S, Krogh A, Mitchison G: **Alignment with affine gap scores**. *Biological Sequence Analysis* Cambridge University Press; 1998, 29-31.
4. Hartmann AK: **Sampling rare events: statistics of local sequence alignments**. *Phys Rev E Stat Nonlin Soft Matter Phys* 2002, **65**(5 Pt 2):056102.
5. Bonferroni CE: **Teoria statistica delle classi e calcolo delle probabilità**. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936, 8:3-62.

doi:10.1186/1758-2946-4-S1-O12

Cite this article as: Hähnke: Text-based similarity searching for hit- and lead-candidate identification. *Journal of Cheminformatics* 2012 **4**(Suppl 1):O12.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>


ChemistryCentral