

ORAL PRESENTATION

Open Access

# Representation and searching of biomolecules

Joseph L Durant\*, WL Chen, BD Christie, DL Grier, BA Leland, JG Nourse

From 5th German Conference on Cheminformatics: 23. CIC-Workshop  
Goslar, Germany. 8-10 November 2009

Biomolecules present challenges to chemical information systems designed for small molecules. Their sizes, up to tens of thousands of atoms, overwhelm representation/storage/searching solutions built on explicit chemical representation of the structures. But biomolecules are largely made up of many repeats of a limited number of building-block molecules, a fact which has been used to provide a compressed representation for biomolecules using templates for the building blocks.

We have adopted a modified template-based representation for biomolecules. Our primary interest is in the chemically modified portions of biomolecules, for which we choose to use explicit chemistry. These areas of explicit chemistry are then embedded in the template-compressed, unmodified portions of the full biomolecule.

The regions containing explicit chemistry are indexed, and thus can be structure searched with good performance. A limited number of residues surrounding explicit chemistry regions are included in the index for searching the context of these explicit regions. By using explicit chemistry to represent modified regions we can search across classes of modifications for common features. For example a single substructure search query will find green fluorescent protein, and its histidine, phenylalanine and tryptophan analogs.

Templates are stored with the structure providing a self-contained file format. The use of NEMA keys allows templates from different structures to be compared, and allows storage of structures containing a canonical list of templates. The residues have defined attachment points, allowing automated traversal of a protein backbone, or location of non-backbone bonds to residues.

We will present example structures and structural queries highlighting capabilities of our representation.

Published: 4 May 2010

doi:10.1186/1758-2946-2-S1-O4

Cite this article as: Durant et al.: Representation and searching of biomolecules. *Journal of Cheminformatics* 2010 **2**(Suppl 1):O4.



Publish with **Chemistry**Central 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/>



Chemistry Central