

Oral presentation

Open Access

Seamless integration of the PubChem database into an universal scriptable chemical information processing environment

W-D Ihlenfeldt

Address: Xemistry GmbH Auf den Stieden 8, D-35094 Lahntal, Germany
from 4th German Conference on Chemoinformatics
Goslar, Germany. 9–11 November 2008

Published: 5 June 2009

Chemistry Central Journal 2009, 3(Suppl 1):O16 doi:10.1186/1752-153X-3-S1-O16

This abstract is available from: <http://www.journal.chemistrycentral.com/content/3/S1/O16>

© 2009 Ihlenfeldt; licensee BioMed Central Ltd.

The PubChem database has quickly become a premier information source for the lookup of information on chemical structures and their testing results in biological assays. While the primary access route is a Web interface, which is designed for human interaction, PubChem is unique among the chemistry Web databases in that it provides an API which in theory allows programmatic interaction with the database by custom clients. However, the various disconnected HTML, XML- and SOAP-based APIs are complex and hardly usable by chemists who are capable of writing minor scripting solutions, but do not intend to spend a long time to become experts in the intricacies and limitations of the various access methods.

We have now released an update to the Cactvs Chemoinformatics Toolkit which for the first time allows access to PubChem as if one were to interact with a local structure file. All details of the access methods are hidden to the user and optimized behind the scenes. Supported features include:

- Direct I/O of the native ASN.1 encoding of the PubChem structure and bioassay data, preserving all information
- Presentation of the PubChem compound database as a virtual file, allowing reading, positioning, scanning as if were a local SD-type file
- Support of the complete feature set of the toolkit structure search, with transparent, optimized and automatic offloading of those parts of a query which can be executed by the PubChem servers to these
- Download of structures and assays by CID and AID with full data retention

- Reverse loop-up of PubChem CIDs and SID sets from arbitrary structures
- Name and CAS number lookup and reverse structure instantiation from PubChem references
- Interface to the PubChem structure standardization algorithm to obtain common structure representations for comparability or lookup
- Retrieval of inter-database link data, such as MeSH terms, literature references, bioassay associations and similar references
- Output of bioassay data with augmented information, such as structure depictions, for further processing in formats such as MS Excel files on any platform