BIOINFORMATICS LETTER TO THE EDITOR Vol. 21 no. 9 2005, pages 1751–1753
doi:10.1093/bioinformatics/bti295

Databases and ontologies

SuperDrug: a conformational drug database

Andrean Goede†, Mathias Dunkel†, Nina Mester, Cornelius Frommel and Robert Preissner*

Berlin Center of Genome Based Bioinformatics, 3D Datamining Group, Institute of Biochemistry, Charité, Monbijoustrasse 2, 10117 Berlin, Germany

Received on June 11, 2004; revised on January 25, 2005; accepted on January 26, 2005
Advance Access publication February 2, 2005

ABSTRACT

Motivation: Different resources exist for experimentally determined and computed three-dimensional (3D)-structures of low molecular weight structures but for approved drugs, no free, publicly accessible source of 3D-structures and conformers is available. Furthermore, for selection purposes or for correlation of structural similarity with medical application, the assignment of the Anatomical Therapeutic Chemical (ATC) classification codes to each structure according to the WHO-scheme would be desirable.

Results: The database contains ∼2500 3D-structures of active ingredients of essential marketed drugs. To account for structural flexibility they are represented by 10^5 structural conformers. Here we present a web-query system enabling searches for drug name, synonyms, trade name, trivial name, formula, CAS-number, ATC-code etc. 2D-similarity screening (Tanimoto coefficients) and an automatic 3D-superposition procedure based on conformational representation are implemented. Drug structures above a similarity threshold as well as superimposed conformers can be retrieved in the mol-file format via a graphical interface.

Availability: For academic use the system is accessible at http://bioinf.charite.de/superdrug. The retrieval system requires the free browser-plugin 'chime' from MDL for visualization.

Contact: robert.preissner@charite.de

Different resources exist for experimentally determined three-dimensional (3D)-structures of low molecular weight structures (Allen, 2002) and biological macromolecules. Furthermore, computed structures exist for millions of chemical compounds (Bradley, 2002). In a comparison of large chemical databases the publicly accessible NCI database (Ihlenfeldt *et al*., 2002) came out to have, by far, the highest number of compounds that are unique to it (Voigt *et al*., 2001). Commercial databases are often intended to cover a broad range of bioactive compounds, development drugs or patented compounds [WDI: 58 000 (http://www.derwent.com/products/lr/wdi/); CMC: 7500 (http://www.mdl.com/products/knowledge/medicinal_chem/); MDDR: 106 000 (http://www.prous.com/product/electron/mddr.html)]. Our approach excludes drugs which are entire plants, extracts, mixtures, colloids or (to some extent) biopolymers (see the list at 'statistics' on the SuperDrug website). Computed drug structures are available via commercial interfaces but the SuperDrug database is the first exhaustive free resource for WHO-classified drugs.

The Chemical Abstracts (CA, http://www.cas.org) provide information on drugs including the CAS-number, useful as cross-reference to other databases and the 2D-structure. The latter was used to generate 3D-structures (Discovery Studio, Accelrys, http://www.accelrys.com/dstudio/). Subsequently, fingerprints (MACCS keys, 960 bit, http://www.lib.uchicago.edu/cinf/221nm/talks/221nm069.pdf), Chime strings and Tanimoto coefficients were computed with ISIS database tools (Durant *et al*., 2002). The superposition of drugs requires the consideration of their flexibility and can be approached by the generation of conformers. For better coverage of the low-energy conformational space the algorithm of Smellie *et al*. (2003) was applied (MedChem Explorer, Accelrys, http://www.accelrys.com/dstudio/ds_medchem/) and a total of 110 000 conformers was computed (47 per drug). We encountered two limitations of the conformer generation, which will not affect most of the entries but should be addressed in a next database release: larger compounds (>8 rotatable bonds) could not be handled adequately; ions had to be ignored. The 3D-superposition algorithm was developed in our group and compares all conformers of two compounds to find the best structural alignment (Thimm *et al*., 2004). This algorithm can roughly be sketched by the following steps: (1) superposition of the centers of mass, (2) orientation according to principal moments of inertia, (3) atom pair assignment and (4) improvement. The data including precomputed Tanimoto coefficients are stored in a MySQL database on a web-server allowing convenient access via browser. The molecular visualization is performed by the free Chime-Plugin, MDL (available for Windows, SGI, Mac). This allows saving of the atomic coordinates in the MOL-format of one drug structure or the superimposed conformers.

A missing rational drug classification according medical and chemical criteria is a problem of chemical databases like ChemID and NIH, http://chem.sis.nlm.nih.gov/chemidplus/. Recently, the recommendations of the WHO Expert Committee responsible for updating the WHO Model List of Essential Medicines were published (WHO, 2004). For the first time, a list of all items on the Model List sorted according to their 5-level Anatomical Therapeutic Chemical (ATC) classification codes was given. The WHO-list will be updated annually and the SuperDrug database will follow this schedule. The therapeutic subgroup is determined by the second level and the chemical component describes the lower level(s) of classification useful for analyses correlating structural similarity with similar therapeutic

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

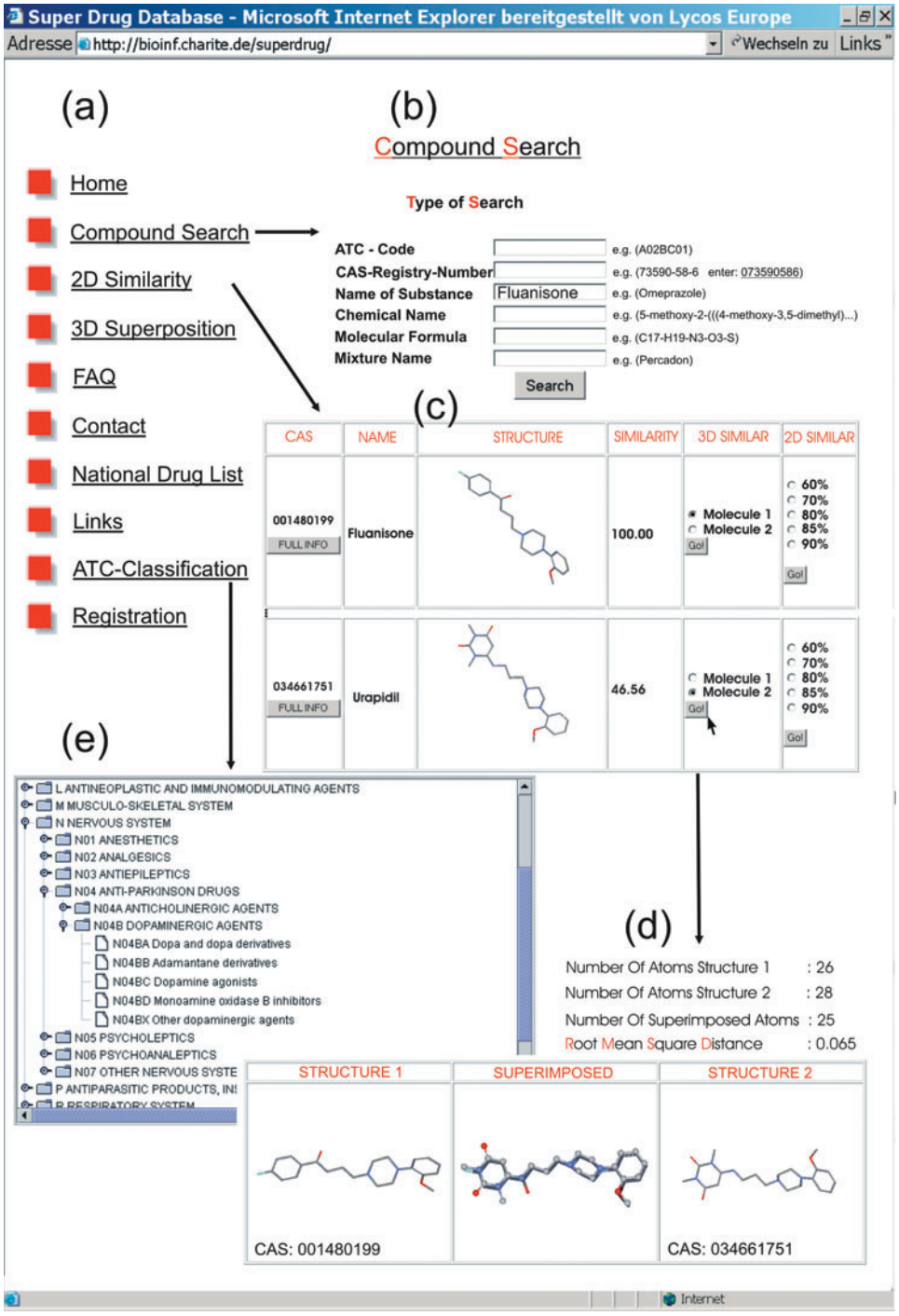

Fig. 1. Data and results of the SuperDrug Web-Interface. (**a**) Navigation frame. (**b**) Text query options. (**c**) Query results with search options for 2D similarity and 3D superposition. Screenshot of a 2D similarity query: Zoledronic acid (CAS 118072-93-8) and Risedronic acid (CAS 105462-24-6) have a Tanimoto coefficient of only 74.21 but belong to the same ATC-class (M05BA07/8-M, Musculo-skeletal system; M05, Bone diseases; M05B, Bone mineralization; M05BA, Bisphosphonates). The compounds can be rotated (left mouse button), different display styles are available (right mouse button) and more detailed information, such as matching ATC-codes, synonymous names and access to the conformers can be obtained by use of the FULL INFO button. (**d**) Results of the 3D superposition of the most similar conformations of both structures, showing the very high similarity of the structures indicating similar action. The table below separately depicts the structures and the superposition of the corresponding conformations in the middle. The picture of the superposition was slightly revised to make it clearer as to which atom belongs to which structure. The (superimposed) 3D structures can be saved by right clicking on the molecule. (**e**) Part of the ATC-tree, showing the structure of the classification: Anatomical (N, nervous system), Therapeutic (N04, anti-Parkinson drugs; N04B dopaminergic agents) and Chemical (N04BA dopa derivatives, N04BB adamantane derivatives, etc.). Clicking on the five-letter-ATC-code triggers a database search in a separate window.

action. Therefore we included a mapping of ATC-codes and active agents in the SuperDrug database. For access to ATC-codes of certain therapeutic or chemical subclasses of the drugs we have constructed a clickable java-tree giving the descriptions up to the fourth level (Fig. 1). This feature requires the installation of the java2 runtime environment on the client.

It is generally accepted that similar compounds having Tanimoto coefficients >0.85 tend to exhibit similar biological activity (Matter, 1997). Similarity searches based on fingerprints and Tanimoto coefficients are standard (von Grotthuss *et al*., 2003) and were implemented in the SuperDrug database. As a fragment- or topomer-based 3D screening was shown to be more selective than 2D similarity (Cramer *et al*., 2002) we were interested in implementing a fast automatic conformer-based superposition algorithm for the SuperDrug database. This enables a comparison of 2D- and 3D-similarity between drugs of different indication classes elucidating structural reasons for adverse effects that might be neglected by exclusive consideration of their 2D-resemblence (Thimm *et al*., 2004). An example illustrating the detection of such a case of diverging 2D- and 3D-similarity is presented in Figure 1.

REFERENCES

Allen,F.H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B*, **58**, 380–388.

Bradley,M.P. (2002) An overview of the diversity represented in commercially-available databases. *Mol. Divers.*, **5**, 175–183.

CA, Chemical Abstracts Service, American Chemical Society.

ChemID, National Library of Medicine, National Institutes of Health.

CMC, Comprehensive Medicinal Chemistry, MDL Information Systems, Inc.

Cramer,R.D. *et al*. (2002) Dbtop: topomer similarity searching of conventional structure databases. *J. Mol. Graph. Model.*, **20**, 447–462.

Discovery Studio, Accelrys, Inc.

Durant,J.L. *et al*. (2002) Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.*, **42**, 1273–1280.

Ihlenfeldt,W.D. *et al*. (2002) Enhanced CACTVS browser of the open nci database. *J. Chem. Inf. Comput. Sci.*, **42**, 46–57.

MACCS keys, MDL Information Systems, Inc.

Matter,H. (1997) Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. *J. Med. Chem.*, **40**, 1219–1229.

MDDR, MDL Drug Data Report, Prous Science, Inc.

MedChemExplorer, Accelrys, Inc.

Smellie,A., Stanton,R., Henne,R. and Teig,S. (2003) Conformational analysis by intersection: CONAN. *J. Comput. Chem.*, **24**, 10–20.

Thimm,M. *et al*. (2004) Comparison of 2D similarity and 3D superposition: application to searching a conformational drug database. *J. Chem. Inf. Comp. Sci.*, **44**, 1816–1822.

Voigt,J.H. *et al*. (2001) Comparison of the NCI open database with seven large chemical structural databases. *J. Chem. Inf. Comput. Sci.*, **41**, 702–712.

von Grotthuss,M. *et al*. (2003) Ligand-Info, searching for similar small compounds using index profiles. *Bioinformatics*, **19**, 1041–1042.

WDI, World Drug Index, Thomson, Inc.

WHO (2004) The selection and use of essential medicines. *World Health Organ. Tech. Rep. Ser.*, **920**, 1–127.