Automated Pharmacophore Extraction

FlexX Goes SQL

Holger Claußen, Volker Apelt, Marcus Gastreich, Sally A. Hindle, Christian Lemmen

Jonathan Greene

Introduction

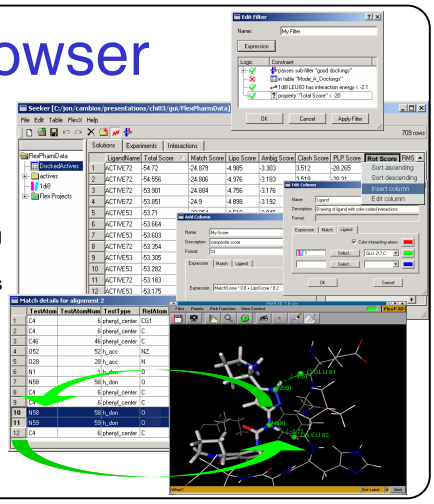
Scoring is still an open problem for docking and an all-purpose scoring function is unlikely to be found at all. Thus, target specific scoring functions are quite attractive, especially as the available data for particular targets and target families grows. In order to gain sufficient insight and create such target specific scoring functions, iterative and interactive detailed analysis of docking solutions will be helpful.

We present an integrated docking workflow that facilitates such analysis and also rapid prototyping of novel scoring functions. All docking information is stored in a database. The data can be analyzed by interactive spread sheets, from which 2D and 3D viewers can easily be launched. New scoring functions as well as types of filters can be defined and tested in this environment.

As a proof of concept for this approach, we performed a docking study with a set of known CDK2 inhibitors with standard FlexX[1] using ScreenScore[2]. We derived pharmacophore constraints by statistical analysis of the docking data. Applying these in FlexX-Pharm[3] led to significantly improved enrichments in screening.

Solution Browser

- **Interactively navigate**
  - browsing and sorting
  - all raw docking data avail.
  - launch 2D / 3D viewer
  - bidirectional selection
- **Rapid prototyping of scoring**
  - user-defined columns
  - interactive definition of filters
- **Database front end**
  - project data manager
  - data import and export
  - SQL GUI
- **Docking GUI**
  - assembly of experiments
  - visible job control

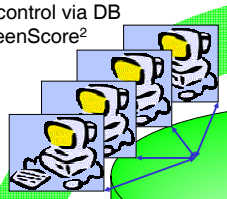


Preparation

- **Prepare input and store in database**
  - protein: define active site, assign AA templates etc.
  - ligands: assign formal charges, minimize, collect activity data
  - FlexX: set up script, parameters etc.
- **Docking of training data set**
  - small set containing active and inactive compounds
  - learn about target
  - adjust parameters
  - create data for detailed analysis

CDK2 experiment

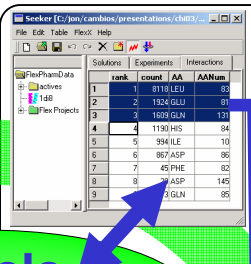
- 72 known actives[6] are docked
- all scoring terms stored in DB
- parallel docking on cluster
- job control via DB
- ScreenScore[2]



target | known actives | screening compounds

Oracle DB

- Contains
  - input and output
  - parameters, scripts, configuration
  - all docking data for all solutions
- Advantages
  - structured storage ⇨ no parsing
  - no constraint on top-x solutions
  - machine learning interface
  - target tailored scoring
  - reproducibility



Easy interactive pharmacophore definition in GUI

Pharmacophore Constraints

```
@interact
essential   _n leu   A   83   h_don
optional    _o glu   A   81   h_acc
optional    _o leu   A   83   h_acc

@partial 1 2
```

Analysis

interactively in browser

- sorting by anything
- define and test filters
- statistical analysis
- visual inspection
- pharmacophore extraction

CDK2 experiment

- count hydrogen bridges
  - ⇨ 3 AA top-ranking
- use two buried interactions as pharmacophore for virtual screening

Screening with FlexX-Pharm[3]

Checking constraints ***during*** docking

⇩

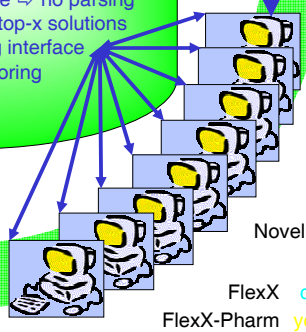
Rejection of wrong solutions ***during*** docking

⇩

More directed search in conformational space

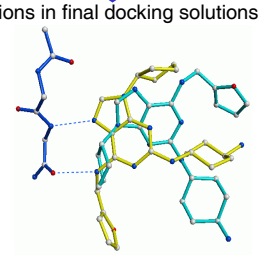
⇩

Novel docking solutions in final docking solutions set



FlexX cyan -38.73

FlexX-Pharm yellow -40.86

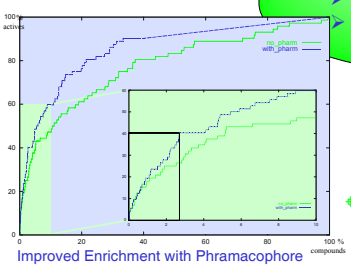


FlexX-Pharm[3] finds a solution with *smaller* RMSD and *higher* score on top rank[4]

Results

CDK2 experiment

- Virtual screening of Bionet[5] data set ~34,000 cmpds
- Application of statistically derived pharmacophore
- Enrichment factor of up to 15



Improved Enrichment with Phramacophore

⇨ Simple statistical analysis led to new insights / pharmacophore

⇨ Structured storage of docking information allows for easy analysis

⇨ Integrated docking environment facilitates rapid experimentation

⇨ Target-tailored pharmacophores and scoring improves enrichment significantly

⇨ Straightforward interface for machine learning algorithms

Technical Details

- Test data set
  - CDK2 structure 1di8 from PDB
  - 72 known CD2 inhibitors from literature [6]
  - 34,000 Bionet [5] screening compounds
- FlexX version 1.12.0, BioSolveIT GmbH
  - ScreenScore [2]
  - FlexX-Pharm module [3]
  - PyFlexX (Python Interface)
- Seeker version 2.3 , Cambios Computing, LLC
- FlexV version 1.6.1, BioSolveIT GmbH
- Oracle version 9i, Oracle Corporation

PyFlexX: FlexX Python Interface

- Python[7] is an open-source, full-featured language
  - variables, loops, if-clauses, functions
  - file IO, additional modules
- Interface to
  - molecular modeling tools (MMTK, MDTools, PyMol)
  - data analysis (Excel, ChemDraw & other COM apps)
  - database (Oracle, Daylight, ReliBase)
  - cheminformatics toolkits (PyDaylight, OElib)
- Implementation
  - menus become modules
  - C functions wrapped, shared memory C / Python

⇨ fast computation, rapid prototyping

References

[1] Rarey et al., JMB, 261, 470–489, 1996
[2] Stahl & Rarey, JMC, 44(7), 1035-1042, 2001
[3] Hindle et al., JCAMD, 16, 129–149, 2002
[4] M. Stahl, personal communications
[5] http://www.keyorganics.ltd.uk/screenin.htm
[6] A. Steffen, personal communications
[7] http://www.python.org/

Project Information

Developments in collaboration with:

Bayer CropScience

Cambios Computing, LLC
1481 Pitman Avenue, Palo Alto, CA 94301, USA, ☎ +1 650 327 9056
info@cambioscomputing.com, http://www.cambioscomputing.com

BioSolveIT GmbH
An der Ziegelei 75, D-53757 Sankt Augustin, Germany
☎ +49 2241 25 25 - 0, **FAX** +49 2241 25 25 - 525
contact@BioSolveIT.de, http://www.BioSolveIT.de