University of Wisconsin

Department of Biostatistics

ISMB 2002

Department of Computer Science

Mining Three-dimensional Chemical Structure Data

Sean McIlwain & David Page

University of Wisconsin

Arno Spatola & David Vogel

University of Louisville

Slyvie Blondelle

Torey Pines Research Institute for Molecular Studies

University of Wisconsin

Department of Biostatistics

ISMB 2002

Department of Computer Science

Advantages of ILP for Pharmacophore Discovery

- Works with 3-dimensional databases without loss of information.
- Multi-relational.
- More comprehensive search of the space than typical greedy or hill-climbing in RP or ANNs.
- Pruning of search space can be achieved using appropriate scoring functions.

Methodology

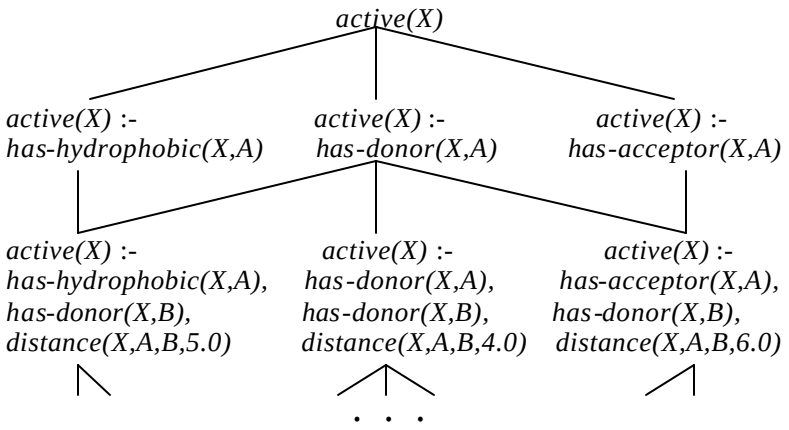
- Pick active/inactive molecules for system under study.
- Generate 3-dimensional structures via a conformational search (*Charmm*).
- Convert 3-dimensional results into Datalog format.
- Extract point groups from Datalog data.
- Perform ILP search of point groups for a pharmacophore clause.

University of Wisconsin
Department of Biostatistics
ISMB 2002
Department of Computer Science

Overview of Aleph ILP Search (Top-down)

- Saturates on 1st uncovered positive example.
- Performs top-down admissible search of the subsumption lattice above this example.
- Use of compression function to limit size of clauses and number of allowed negative example coverage.

Lattice of Clauses for the Given Hypothesis Language



University of Wisconsin
Department of Biostatistics
ISMB 2002
Department of Computer Science

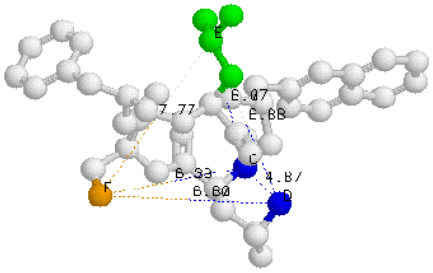
Pharmacophores Found in Pseudomonas Growth Inhibition

Active(A) ← molecule(A), positive(A,B), positive(A,C), hydrophobic(A,D), hydrogen-donor(A,E), distance(A,B,C,4.05), distance(A,B,D,7.45), distance(A,B,E,6.53), distance(A,C,D,6.93), distance(A,C,E,5.90), distance(A,D,E,7.88).

Cross-validation accuracy is 100%.

Example 4-point Pharmacophore Overlayed With an Active Molecule

- Green - hydrophobic
- Blue – positive charge
- Orange – hydrogen donor

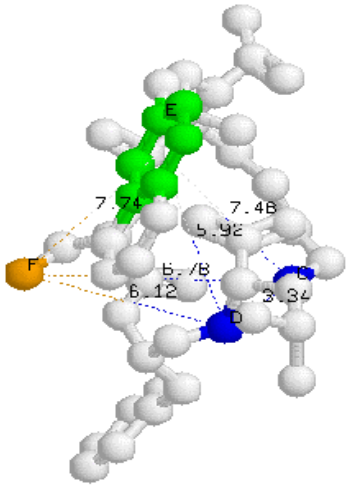


Distances are in Angstroms (Å)

University of Wisconsin
ISMB 2002
Department of Biostatistics
Department of Computer Science

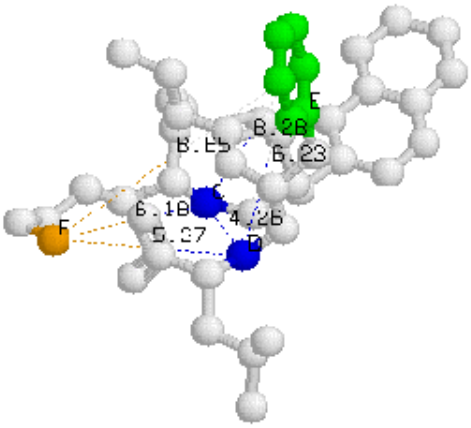
Example Pharmacophore

- Green – hydrophobic
- Blue – positive charge
- Orange – hydrogen donor



Example Pharmacophore

- Green - hydrophobic
- Blue – positive charge
- Orange – hydrogen donor



University of Wisconsin
Department of Biostatistics
ISMB 2002
Department of Computer Science

Conclusion

- Cross-validate results and pharmacophore found shows that ILP is well suited to mining 3-dimensional chemical structure data.
- Directly mines relational data with the use of feature vectors.
- Interacts well with scientists.
- Approach also repeated successfully by Marchand-Geneste, N. (2002).

Future Work

- Testing the proposed pharmacophore with new molecules.
- Application of ILP to related data (drug-discovery, proteomics, SNPs, etc.).

University of Wisconsin
Department of Biostatistics
ISMB 2002
Department of Computer Science

Bibliography

- Brooks, B. (1983). Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4:187-217.
- Finn, P., Muggelton, S., Page, D., and Srinivasan, A. (1998). Discovery of pharmacophores using Inductive Logic Programming. *Machine Learning*, 30:241-270.
- Marchand-Geneste, N. (2002). A new approach to pharmacophore mapping and qsar analysis using inductive logic programming. Application to thermolysin inhibitors and glycogen phosphorylase b inhbitors. *J. Med. Chem.*, 45:389-409.
- Ramakrishnan, R. (1999). Database Management Systems. McGraw-Hill Higher Education, Columbus, OH, 2nd edition.

Aleph

- Maintained by Ashwin Srinivasan publicily available at http://web.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/aleph.html.
- Runs using yap prolog compiler maintained by Vítor Santos Costa obtainable at http://www.ncc.up.pt/~vsc/Yap/ .