DS QSAR and Library Design Webinar

Practical QSAR and Library Design: Advanced tools for research teams

- Reservationless-Plus Dial-In Number (US): (866) 519-8942
- Reservationless-Plus International Dial-In Number: (770) 790-2186
- Conference Code: 8587995760
This presentation and/or any related documents contains statements regarding our plans or expectations for future features, enhancements or functionalities of current or future products (collectively "Enhancements"). Our plans or expectations are subject to change at any time at our discretion. Accordingly, Accelrys is making no representation, undertaking no commitment or legal obligation to create, develop or license any product or Enhancements. The presentation, documents or any related statements are not intended to, nor shall, create any legal obligation upon Accelrys, and shall not be relied upon in purchasing any product. Any such obligation shall only result from a written agreement executed by both parties. In addition, information disclosed in this presentation and related documents, whether oral or written, is confidential or proprietary information of Accelrys. It shall be used only for the purpose of furthering our business relationship, and shall not be disclosed to third parties.
DS QSAR+

- Property calculation protocols
  - Calculate Properties
  - Calculate all calculable properties in Pipeline Pilot
  - Calculate semi-empirical QM descriptors (VAMP)
  - Calculate density functional QM descriptors (DMOL3)

- Structure-activity model building protocols
  - Create Bayesian Model
  - Create Multiple Linear Regression Model
  - Create Partial Least Squares Model
  - Create Back Propagation Neural Network (BPNN) models
  - Create Genetic Function Approximation (GFA) model

Includes multi-objective Pareto optimiser
Calculate Molecular Properties

- All calculable properties in Pipeline Pilot
  - Lipinski and Molecular Properties
  - Element and Molecular Property Counts
  - Surface Area and Volume, and Jurs
  - Fingerprints
    - Extended connectivity
    - Functional class
    - Path-based
    - Atom Environment
    - MDL Public Keys
- Topological
- Estate Keys
- Shadow Indices
- Semi-empirical QM descriptors (VAMP)
  - energy, orbitals (HOMO/LUMO), multipoles
- Density functional QM descriptors (DMOL3)
- Statistics
Appropriate Descriptors are Critical

- About 20 descriptors in C2.QSAR+
- Over 100 descriptors in C2.Descriptor+
- Over 600 descriptors in DS
Genetic Function Approximation

- Evolutionary algorithm generates population of statistically valid SAR models, rather than single model
  - Multiple models provide different insights into system
  - Ensemble model also included
- Friedman's lack-of-fit (LOF) error measure to control number of terms in model whilst minimising least squares error
- Include higher order polynomials and spline functions
  - Creation of non-linear models
  - Automatic outlier removal (spline)
- Statistic
  - Monitor variable usage
Bayesian and Back-Propagation Neural Network

- Builds models with non-linear relationships
- More tolerant of noise and outliers
- BNN allows descriptors to be chosen as both response and predictor variables
  - Replicator architecture allows non-linear modes of PCA and PLS
- Sensitivity analysis for determining the importance of each variable
- Pre- and post processing features for inputs and outputs; permits splits of fingerprint keys
- Categorical, numeric and 2D fingerprint data handled
  - Any number of targets (multiple Y), and allows numerical Y variables
QSAR Modeling and Deployment Capabilities

• Integrated work environment
• Validated Statistical Methods
  - GFA, BNN, multiple linear regression, Partial Least Squares (PLS), Bayesian modeling
• Compound Clustering and 3D graphing
• QSAR models added to descriptor list
This presentation and/or any related documents contains statements regarding our plans or expectations for future features, enhancements or functionalities of current or future products (collectively “Enhancements”). Our plans or expectations are subject to change at any time at our discretion. Accordingly, Accelrys is making no representation, undertaking no commitment or legal obligation to create, develop or license any product or Enhancements. The presentation, documents or any related statements are not intended to, nor shall, create any legal obligation upon Accelrys, and shall not be relied upon in purchasing any product. Any such obligation shall only result from a written agreement executed by both parties. In addition, information disclosed in this presentation and related documents, whether oral or written, is confidential or proprietary information of Accelrys. It shall be used only for the purpose of furthering our business relationship, and shall not be disclosed to third parties.
The Compound Acquisition Challenge

- **Novelty**
  - Identical (or similar) compounds should not exist in the corporate collection

- **“Lead-likeness” or “Drug-likeness”**
  - Characterize the ‘hit’
  - Identify the good physiochemical properties
  - Absence of reactive functionalities

- **Procurement**
  - Availability from a reliable supplier at an appropriate price and quantity

- **Series**
  - Compounds can be acquired in series to establish SAR

- **Medicinal Chemists Review**
  - Compounds required for inspection
An Example Compound Acquisition Protocol in Pipeline Pilot

1. Read Vendor SD File(s)
2. Lipinski Filter
3. Eliminate Reactive Functionality
4. Calculate Unit Prices
5. Standardize Molecule
6. Canonical Tauberian
7. Merge Molecules
8. Select Unique Vendor Compounds
9. Select Preferred Vendors, Filter on Price
10. Cluster Similar Compounds
11. Visual molecule selector

Write Results To Excel

Write Flows for Order System
Library Design & Analysis

- Need to compare and prioritize various chemical libraries
  - Which of the possible chemical libraries best meets the requirements for the current project?
  - When considering multiple libraries, which set will provide the most information about the system?

- Need to select library members to optimize various properties within a proposed chemical library.
  - Of the thousands of compounds that could go into a proposed library, which will:
    - be the most representative?
    - have the best combination of properties?
Library Design in DS 2.0

1. Select a Library Enumeration protocol
2. Set the library parameters
3. Create the library
Library Design Book-Keeping

Complete, accurate, and automatic logging of all Discovery Studio experiments in a standard easy to search and access format.
Tools for Library Optimization

• Common selection tools in DS Library Design:
  - Compare libraries
  - Compute diversity metrics on a single library
  - Cluster molecules
  - Find a diverse set of molecules
  - Find similar molecules
  - Perform a principle component analysis
  - Perform R Group sub-setting

• Need to optimize multiple properties simultaneously
Optimizing Multiple Library Properties

• Weighted Sum Optimization
  - Create a function as a weighted sum of properties:
    \[ O_{library} = w_u \cdot uniqueness / s_u - w_{sol} \cdot Opt.Sol. / s_{sol}. \]
  - Drawbacks
    • Provides only a single solution
      - Would just a small change in optimal solubility, have a large effect on uniqueness?
    • Weighting values are largely arbitrary
    • Small changes in weights may dramatically change results

There must be a better way
Vilfredo Pareto: Italian economist (1848-1923)

- Discovered 80/20 rule (80% of land in Italy owned by 20% of people) - since generalized to many other contexts
- His idea of “efficiency” forms basis of Pareto optimization
  - Pareto optimal solution: A set of solutions where any change in allocation of solutions will result in other solutions becoming worse
Pareto Optimization

- Puts all goals on an equal footing (no need to assign weights)
- Discard potential solutions that can’t possibly be of interest (given the goals)
- Yields a family of solutions on the \textit{Pareto-optimal surface}

1,575 compounds
- Create libraries of 40 compounds
- What is the best library?
Specific Library Design Functions Based on QSAR

- Pareto optimization and sorting creates libraries with optimal sampling of multiple properties

- DS Library design provides multiple methods for calculating diversity metrics to guide the prioritization of libraries

- Integrated clustering, diversity and similarity tools help select specific compounds from within the Discovery Studio work environment

DS Library Design implements advanced descriptors and models to create optimal libraries
Additional New Functionality

- **Enhanced 3D graphing**
  - Better visualization
  - Interactive manipulation of 3D graphs
  - Instant access to information behind graph

- **Perl scripting capabilities**
  - Enhanced flexibility, supported worldwide

- **Support for clusters, multiple parallel processing (MPI)**
  - Faster performance, more options, innovative software design

- **Integration with Pipeline Pilot**
  - Much greater flexibility, customizability than any other molecular modeling software product
All tools Integrated into Discovery Studio

• Single integrated work environment improves productivity
  - Ligand-based and Structure-based design tools are together in a powerful environment
  - The ability to use Pipeline Pilot protocols provides unprecedented capabilities to customize and automate workflows.
  - Seamlessly integrate the steps of the drug development process.

• Just a few of the new capabilities added to Discovery Studio 2.0
  - Flexible docking
  - Loop modeling for creating better protein models; protein pK prediction
  - De Novo Ligand Builder for creating new compounds from pharmacophores.
  - Ligand Profiler for understanding the environment of a drug.
  - CAESAR for improved modeling of bioactive geometries
  - Pareto Optimization for creating ideal libraries

“Our modeling experts were pleased with the time savings realized by having a single environment in which to generate conformers, dock poses, and model pharmacophores.”
-- Dr. Jeff Zablocki (Head of Chemistry, CV Therapeutics)
Integrated Work Environment Improves Productivity

Legacy Tools for Drug Discovery

• Many disjointed programs
• Time consuming to change applications.

New Accelrys

• Integrated interface
• High quality, new science!
Data Exchange Among Scientist is Critical

- Example: Deploy pharmacophore and ADMET screening tools so that chemists have the ability to easily add, predict, and visualize new compounds.

Create a workflow around your tools

Allow the Chemists to use it
Conclusion

- DS QSAR includes an extensive descriptor set and powerful tools for building predictive models
  - QM descriptors, GFA
- DS Library Design provides the ability to quickly and easily generate targeted chemical libraries
  - R-group and reaction-based library enumeration
  - Pareto optimization
- Discovery Studio integrates these applications into a complete drug discovery solution
Acknowledgements

- Tien Luu
- Shikha Varma-O’Brien
- Dana Honeycutt
- Jon Sutter and the R&D Team

Thank You