SARvision

ESSENTIAL TOOLS FOR SAR
SARvision

- SARvision | is part of a product line for SAR analysis
  - SARvision | SM
  - SARvision | Biologics
  - AMEDEO
- In use for over 10 years
  - Japan, Europe and North America
  - Academic, small and large corporations
- Different departments
  - HTS, Chemistry, Computer Aided Design, Legal
  - Service providers
SAR Analysis Tool

Scaffold Perception

Data Filtering

Molecule Tables
R-group Tables
Gridviews
Parent Child Chemotype Organization

- Automated scaffold perception algorithm
- Knowledge based tree pruning: not MCS
- Tree Organization: selected chemotype subsetting
- User override
  - Generalized atom types (A,Q,X)
  - Generalized bond types
- User defined parent chemotypes
  - Enumerate children from a parent chemotype
Data Views

- Molecule Table
- Grid View
- R-group Table
  - Rearrangeable R definitions
  - Rename substituents (Me, CH3…)
- Edit, format data, add rows or columns
- Rich Calculator
  - Arithmetic and logic operations
- Heat maps
Graphs

**Bar Charts**
Multivariable

**Scatter Plots**
Color or shape by property or scaffold

**3D Scatter Plots**
Rotatable
What is highlighted in a panel is highlighted (filtered or marked) on the others.
ChemSABRE: In-depth R-group Analysis

- R-group based property analysis
  - Calculate properties for pendent groups
  - Pendent group property plots
- RxR Tables
  - Data as lists or averages, min-max, etc.
  - Heat map cells by property
  - Filter unwanted substituents
- Bioisosteres
  - Literature database
- Missing molecules enumeration (Holes)
  - Exhaustive or limited
Comparative Substituent Analysis
Scaffold Hopping
Additional Tools and Features

- **Simple Property Calculator:** MW, PSA, HBD, HBA, logP, nrot, SMILE strings
- **Library Design**
  - Scan SD files: multiple queries at once
  - Scaffold based diversity selection
  - Comparison of lists of scaffolds (Union, Intersection, OR)
- **Chemotype diversity compound selection**
  - For a collection of chemotypes select a compound subset that maximizes the number of chemotypes present.
- **Compound enumeration**
  - Simple compound enumeration tool, given two libraries with linkage positions, a new file is generated containing molecules that result from making all possible combinations.
  - No chemical intuition or knowledge.
Runs under Windows OS
- A chemotype generation tool is available for LINUX without interface
- Newer than Windows 7 64 bit is recommended, but 32 bits are supported.

Reads in SD files, SMILE strings, Comma Delimited data
- Components can be generated for workflow programs
- Handles millions of molecules and data points depending on RAM
- Chemotype lists can be exported

Output
- Data Views into MS Word, Excel, SD file
- Internal format to save and share projects

Integrates with other chemoinformatics or visualization tools
- Can be customized or called from other applications
Use examples

**Lead Optimization**
- Carry out RxR analysis
- Enumerate missing molecules
- Reorganize set according to chemotypes: select diverse subset
- Export via SDF for property calculation, acquisition, docking, etc.
- Matched Molecular Pairs

**Hit to Lead**
- From HTS results (or docking results) identify active chemotypes faster
- Scaffold hopping

**Library Design**
- Enumerate chemotypes in vendor catalogs – Work with hundreds of thousands of compounds
- Compare sets at the scaffold level
- Select diversity based on chemotypes
- Create libraries of unwanted chemotypes and remove them automatically

**IP protection**
- Organization of chemotypes into trees facilitates claim writing: close to Markush
- From patent literature database, identify missing molecules or chemotypes
SARvision | Biologics

A tool to organize and perform SAR analysis on biopolymers
SARvision | Biologics

- Research informatics solution for Biologics
  - Tool for SAR analysis
    - Desktop runs under Windows
  - Primary Sequence to Activity
    - Large datasets are becoming commonplace
    - 3D and bioinformatics solutions are available
- Three broad development considerations
  - Data handling
  - Data organization and visualization
  - Primary structure analytics
- Integration with other applications
Tools to correlate sequence to activity are not available

- Many large companies deal with Excel macros

Broad range of therapeutic molecules need to be handled

- Peptides, Antibodies, Protein Engineering, Polynucleotides...
- Unnatural amino acids
- Modified peptides and proteins

Growing gap in research informatics

- Shift from small molecule therapeutics
Correlate Sequence to Activity

- **Tools to handle and organize sequences**
  - Sequence alignment techniques
    - You can perform sequence alignments using Clustal, Needleman-Wunsch, etc.
    - You can edit your sequence alignments by adding insertions or deletions.
    - You can read in alignments from other sources
    - You can define generalized substitution matrices (unnatural amino acids)
  - Phylogenetic trees
  - Color by property
- **Tools to handle and organize data**
  - Sorting, heatmaps, filters, operations on columns, etc.
  - Visualization tools: bar graphs, scatter plots, etc.
  - Simple calculators
Template Definition: Motif based analysis
Property based analysis
Subsets: Filtering by properties

<table>
<thead>
<tr>
<th>Mutations</th>
<th>delta_InfFdy</th>
<th>delta_Logit</th>
<th>delta_Hedge</th>
<th>delta_pval</th>
<th>RCL_3</th>
<th>Phen-Knowl Inv</th>
</tr>
</thead>
<tbody>
<tr>
<td>H75_Mut6</td>
<td>123.45</td>
<td>67.89</td>
<td>0.12</td>
<td>0.001</td>
<td>987</td>
<td>Inv 1</td>
</tr>
<tr>
<td>H75_Mut7</td>
<td>234.56</td>
<td>78.90</td>
<td>0.13</td>
<td>0.002</td>
<td>876</td>
<td>Inv 2</td>
</tr>
<tr>
<td>H75_Mut8</td>
<td>345.67</td>
<td>89.01</td>
<td>0.14</td>
<td>0.003</td>
<td>765</td>
<td>Inv 3</td>
</tr>
<tr>
<td>H75_Mut9</td>
<td>456.78</td>
<td>90.12</td>
<td>0.15</td>
<td>0.004</td>
<td>654</td>
<td>Inv 4</td>
</tr>
</tbody>
</table>
Graphing Capabilities
User generated properties can be used to color residue backgrounds.
Dendrogram
Subsets based on dendrograms
Motifs based analysis

- Antibody Engineering, CDRs

Antibodies from the Protein DB, CDRs shown colored by Hopps Wood scores
SAR for Biotherapeutics

- Identify features in the sequence related to activity
  - Logo Plots
    - How frequent are residues in critical positions for active peptides?
  - Phylogenetic trees
    - Do similar sequences result in similar activities or patterns of activity?
  - Residue properties
    - Can trends in residue properties be observed?
  - Mutation Cliffs
    - Are there residue substitutions that lead to large changes in activity?
  - Invariant Maps
    - Are there critical residues that should be kept in place?
Data Analysis and Visualization
<table>
<thead>
<tr>
<th>title</th>
<th>bioluminate</th>
<th>delta_Affinity</th>
<th>delta_Stability</th>
<th>ROI_0</th>
<th>ROI_1</th>
<th>ROI_2</th>
<th>ROI_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3HOT</td>
<td>NONE</td>
<td>23.1437</td>
<td>-15.9151</td>
<td>Y</td>
<td>Y</td>
<td>F</td>
<td>S</td>
</tr>
<tr>
<td>3HOT_Mut_1</td>
<td>A37Tyr=MET,B105Phe=ILE,B107Tyr=ARG</td>
<td>40.5989</td>
<td>-13.7951</td>
<td>Y</td>
<td>M</td>
<td>F</td>
<td>S</td>
</tr>
<tr>
<td>3HOT_Mut_10</td>
<td>A37Tyr=HIF,B105Phe=MET,B107Tyr=GLU</td>
<td>29.3437</td>
<td>-8.9170</td>
<td>Y</td>
<td>M</td>
<td>F</td>
<td>S</td>
</tr>
<tr>
<td>3HOT_Mut_100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ala-san plot
Invariant Maps

- Activity trends as a position in the sequence is kept invariant
Sequence mutations that elicit a significant change in activity
AMEDEO

tools in data science to speed up your lead optimization cycle
AMADEO will **LEARN** about your data, **BUILD** and **SELECT** the most predictive models, and make clear **RECOMMENDATIONS** as to what compounds to make next. AMADEO can take into account your multiple wishes as it recommends your next steps.
SaaS SARvision

Web based solution for library organization