Application integration: Providing coherent drug discovery solutions

Mitch Miller, Manish Sud, LION bioscience, American Chemical Society 22 August 2002
Overview

- Introduction: exploring ‘application integration’
- Our integration platform
- Compound prioritization — one example
- Information traversal — another example
Introduction

- Application Integration
  - Great idea!
  - What does it mean?

- ‘The act or process of integrating’
  - ‘To make into a whole by bringing all parts together; unify’

- Mathematical
  - ‘A number computed by a limiting process in which the domain of a function, often an interval or planar region, is divided into arbitrarily small units, the value of the function at a point in each unit is multiplied by the linear or areal measurement of that unit, and all such products are summed.’
  - The area under a curve!
Which curve?

The area under the information-access *learning* curve!

- Learning curves go down, according to Gary Anthes in ComputerWorld, July 2, 2001
- We want to *reduce* the area under the learning curve
Application integration:
A situation in which different applications share data and user interface elements, allowing the user to move quickly and efficiently between tasks and *work more productively*. 
What are we integrating?

- **Vertical applications:**
  - Chemical database searching
  - Compound selection
  - Compound analysis

- **Horizontal integration:**
  - Assay data browsing and searching
  - Relationship browsing
Introducing LION DiscoveryCenter

Integration framework for compound discovery
- Supports all the applications listed

Guiding principal: ‘Show me everything we know about X’
- uniting information about drug-discovery topics
- variety of sources — public and private
- provide access to specialist tools
- capture knowledge of individual scientists
- jump to related Xs

Bioinformatics products released last year

Chemistry: from vision in April to functioning prototype
- release scheduled for early 2003
Summary sheet — key to ‘show me everything’

- variety of information on a single topic in a single screen
- described by XML -- easy to extend and customize
- addressed by URL -- can be transmitted

Session paradigm:

- search -> browse -> drill down to summary sheets -> traverse

‘Favorites’ provide pointers to individual summary sheets

- find out what’s happening to the entities you’re interested in

Links between different types of data

- build on natural relationships between items
Annotations capture user knowledge
  - pertain to specific sections or to entire entities
  - shared or private

Follow-up searching integrated into content display

Fully-functional Java client
  - includes IceStorm browser to render Web content
  - domain-specific viewers for scientific content

Component-based — easy to extend
Change in Session Paradigm for Chemistry

- Search -> browse -> **analyze** -> drill down to summary sheets -> traverse
  - Use computational services to enhance data from databases
  - Possibly load set of compounds from disk file to initiate analysis
- LDC has compute services API to facilitate hooking in computational tools to better understand our compounds
- Examine one case in point:
  - Compound Prioritization
An application currently in use at LION, predating LDC

A means of assigning a score to a compound
  - how useful will this compound be for us?

Based on existing models for biological activity and ADME
  - Biological activity models use Receptor Relevant ChemSpaces, generated using DiverseSolutions from R. S. Pearlman and K. M. Smith, University of Texas, distributed by Tripos
  - Predictive ADME using in-house models
Chemical Database(s) \(\rightarrow\) Generate BCUT Descriptors

BCUT Descriptors:
- H donor, H Acceptor, atomic polarizabilities and charges
- 29+ descriptors

5 - 6 Orthogonal descriptors represent chemistry space

Choose Chemistry Space
Known Ligands

Generate BCUT Descriptors

Develop Receptor-Relevant Chemistry Space

Only the chemistry space BCUT descriptors

Subset of chemspace BCUT descriptors relevant for a specific receptor
We have generated Receptor-Relevant Chemspaces for about 18 different receptors

Used to estimate receptor specificity

- suggest when a compound is similar to known ligands for our receptors
- Is a proposed compound for one receptor likely to interact with others?

Compound Prioritization user can select:

- whether compounds should be inside or outside each of the various RRCs
- cutoff for RRC
- weighting
We predict a variety of ADME parameters:

- Caco-2 cell permeability ($P_{\text{eff}}$)
- %Fractional dose in portal vein (FDP)
- Blood-brain barrier penetration (BBB)
- CYP2D6
- CYP3A4

Similar to iDEA product
LION Structure-Based ADMET Predictions: Methodology

Assay Data

Data Quantitization

Structure Data

Molecular Descriptor Calculation/Selection

Descriptors
1D: MW etc
2D: b_single etc

0.91

Methods
- PCA
- PLS
- PLS DA
- Neural Networks
- GA
- SA

Machine Learning/
Pattern Recognition

Activity Model
Make Predictions using Structure-Based ADME and Receptor-Relevant Chemspace models

Compound Score: 0 or 1

Consensus Score: 0 or 1

Consensus Scoring:
Consensus Score:
\[
\frac{\sum_{\text{models}} (\text{score} \times \text{weight})}{\sum_{\text{models}} \text{maxscore} \times \text{weight}}
\]
where score = 1 or 0 and maxscore = 1

Structure-Based Models
ADME: Caco-2 Peff, %FDP, BBB CYP2D6, CYP3A4
Selectivity: Receptor-Relevant Chemspace models

Example: Caco-2 Peff
Acceptable: High or Medium
Unacceptable: Low
Weight: 2.0

Scoring
0: Fail criterion
1: Pass Criterion

Ranked Compounds List using Consensus Score

Set of Input Compounds
Previous Compound Prioritization Workflow

- Perform search in ISIS/Base
- Export SD file
- Start web application
- Load SD file
- Specify parameters
- View results
Compounds Prioritization

Input Type: SD

Input File: F:\Projects\Sonoma\acs_mddr3.sdf

Available Models: select all

- Caco-2 Permeability: High or Medium
  Weight: 1.0

- %FDp: High or Medium
  Weight: 1.0

- BBB Penetration: No
  Weight: 1.0

- CYP2D6 Activity: No
  Weight: 3.0

- CYP3A4 Activity: No
  Weight: 3.0

- ERRa Chemspace: 
  Distance: 1.5
  Weight: 1.0

- ERa Chemspace: 
  Distance: 1.5
  Weight: 1.0
Making Visions Work.

Prioritization

- **Compounds Prioritization:** File, IESketcher or Sketcher output

Assessment

- **Compounds Assessment:** File, IESketcher or Sketcher output

AR Analysis

- **R Table Prioritization & Assessment:** File input

Note: The computation process encompasses three primary steps: transfer of data from browser to server, structure-based model calculations, and prioritization. Based on types of model(s) you plan to use for prioritization, it can take anywhere from few minutes to few hours! Specify your e-mail address to receive job completion notification by mail.

(C) LION Bioscience Inc. All rights reserved.
Compounds Prioritization Results

Compounds Prioritization calculation succeeded.

Results:

To download the results file, click the right mouse button on the download link and select Save Target/Link As option.

View:

- Compounds Models Criteria Distribution
- Compounds Prioritization Results
- Compounds Prioritization Results with Structures
- Compounds Prioritization Results with Background Highlighting
- Compounds Prioritization Results with Structures and Background Highlighting
- Compounds Prioritization Results with Text Highlighting
- Compounds Prioritization Results with Structures and Text Highlighting
- Caco-2 Peff Predictions
- Caco-2 Peff Predictions with Structures
- %FDp Predictions
- %FDp Predictions with Structures
- BBB Penetration Predictions
- BBB Penetration Predictions with Structures
- CYP2D6 Activity Predictions
- CYP2D6 Activity Predictions with Structures
- CYP2D6 Activity Predictions with Background Highlighting
- CYP2D6 Activity Predictions with Structures and Background Highlighting
- CYP2D6 Activity Predictions with Text Highlighting
- CYP2D6 Activity Predictions with Structures and Text Highlighting
### Compounds Prioritization Results

<table>
<thead>
<tr>
<th>Structure</th>
<th>MOL NAME</th>
<th>Rank Score</th>
<th>Caco2 Peff (High or Medium)</th>
<th>%FDp (High or Medium)</th>
<th>BBB Penetration (No)</th>
<th>CYP2B6 Activity (No)</th>
<th>CYP3A4 Activity (No)</th>
<th>ERRa Distance (&gt;= 1.5)</th>
<th>EURa Distance (&gt;= 1.5)</th>
<th>FXR Distance (&gt;= 1.5)</th>
<th>GR Distance (&gt;= 1.5)</th>
<th>LY Distance (&gt;= 1.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="RO-25-6833" /></td>
<td>RO-25-6833</td>
<td>0.926</td>
<td>Low</td>
<td>Low</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>10.082</td>
<td>6.464</td>
<td>2.918</td>
<td>7.587</td>
<td>4.0</td>
</tr>
<tr>
<td><img src="image" alt="ME-3277" /></td>
<td>ME-3277</td>
<td>0.026</td>
<td>Low</td>
<td>Medium</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>9.295</td>
<td>3.819</td>
<td>1.037</td>
<td>7.851</td>
<td>3.3</td>
</tr>
<tr>
<td><img src="image" alt="LOMETREXOL" /></td>
<td>LOMETREXOL</td>
<td>0.778</td>
<td>Low</td>
<td>High</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>6.374</td>
<td>3.419</td>
<td>2.389</td>
<td>8.768</td>
<td>3.5</td>
</tr>
</tbody>
</table>
- Perform search in LDC Search Guide
- Browse results
- Select ‘Compound Prioritization’ from application menu
- Fill in parameters
- Add results to form
## Compound Prioritization

### Available Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Criterion</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caco-2 Peff</td>
<td>High or Medium</td>
<td>2.0</td>
</tr>
<tr>
<td>%FDp</td>
<td>High or Medium</td>
<td>1.0</td>
</tr>
<tr>
<td>BBB Penetration</td>
<td>No</td>
<td>1.0</td>
</tr>
<tr>
<td>CYP2D6 Activity</td>
<td>No</td>
<td>3.0</td>
</tr>
<tr>
<td>CYP3A4 Activity</td>
<td>No</td>
<td>3.0</td>
</tr>
<tr>
<td>ERRα Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>ERα Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>FXR Chemospace</td>
<td>&lt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>GR Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>LXR Chemospace</td>
<td>&lt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>PPARα Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>PPARδ Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>PPARγ Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>PXR Chemospace</td>
<td>&lt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>RARα Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>RAR0 Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>RARγ Chemospace</td>
<td>&lt;=</td>
<td>1.5</td>
</tr>
<tr>
<td>ROR Chemospace</td>
<td>&gt;=</td>
<td>1.5</td>
</tr>
</tbody>
</table>

### Actions

- Deselect all models
- Select all models
- Prioritize
- Cancel
Search Guide - Chemical Structure: Substructure search


Search terms: Substructure search

Chemical Results 1 to 29.

<table>
<thead>
<tr>
<th>Row</th>
<th>Structure</th>
<th>Molecular Weight</th>
<th>Molecular Formula</th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure Image" /></td>
<td>379.5450</td>
<td>C₂₄H₃₃N₃O</td>
<td>TR02405000002</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure Image" /></td>
<td>442.0850</td>
<td>C₂₄H₃₁ClN₄O₂</td>
<td>TR02405000013</td>
</tr>
</tbody>
</table>

Search completed. 29 results.
Search Guide - Chemical Structure: Substructure search

Address: http://molvel.sdlionбиосience.com/NetGenics/DiscoveryCenter/lau...SEARCH_TYPE%3DSubstructureSearch&AUTO_START%3DV8CRITERIA%3Dabcdef&PROF...
Starting from one compound of interest, we can proceed in many directions:

- Find other, similar compounds
- Find assays of interest
### Similar Structures from MDDR 1 to 41:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>190292</td>
<td><img src="#" alt="Structure 1" /></td>
<td>69.98</td>
<td>358.4880</td>
<td>C24 H26 N2 0</td>
</tr>
<tr>
<td>2</td>
<td>295061</td>
<td><img src="#" alt="Structure 2" /></td>
<td>65.60</td>
<td>460.5897</td>
<td>C29 H33 F N2 02</td>
</tr>
<tr>
<td>3</td>
<td>168806</td>
<td><img src="#" alt="Structure 3" /></td>
<td>63.59</td>
<td>422.6092</td>
<td>C27 H38 N2 02</td>
</tr>
<tr>
<td>4</td>
<td>163304</td>
<td><img src="#" alt="Structure 4" /></td>
<td>63.42</td>
<td>329.4649</td>
<td>C20 H31 N3 0</td>
</tr>
</tbody>
</table>

**Older women’s experiences of living alone after heart surgery.**

Robinson AW.

Anne White Robinson, RN, DNS, College of Nursing and Allied Health Professions, University of Louisiana at Lafayette, Lafayette, LA.

An increasing number of women who have coronary artery disease have undergone heart surgery and are living alone. This qualitative descriptive study identified four primary themes that constituted essential structures of the post-recovery experience for 12 older women living alone after coronary artery bypass surgery. Themes included (1) survival relief, experienced as awe and gratitude, (2) going on, described as an obligation, (3) living within a contracted world, structured by a sense of vulnerability, and (4) regained independence, characterized by affirmation of self-worth and personal freedom. Post-recovery meant restoring the identity the women had known before surgery together with actualizing the value of self-reliance. Copyright 2002, Elsevier Science (USA). All rights reserved.

PMID: 12173163 [PubMed - in process]
### Bio. Assay Results

#### Bio. Activity Results 1 to 5:

<table>
<thead>
<tr>
<th>Row</th>
<th>Compound ID</th>
<th>Lot Number</th>
<th>Salt Code</th>
<th>Assay Code</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TR0240500055</td>
<td>1</td>
<td>HF</td>
<td>MC1BD</td>
<td>22.504</td>
</tr>
<tr>
<td>2</td>
<td>TR0240500055</td>
<td>1</td>
<td>HF</td>
<td>MC1BD</td>
<td>28.445</td>
</tr>
<tr>
<td>3</td>
<td>TR0240500055</td>
<td>1</td>
<td>HF</td>
<td>MC1BHM</td>
<td>45.028</td>
</tr>
<tr>
<td>4</td>
<td>TR0240500055</td>
<td>1</td>
<td>HF</td>
<td>MC4BD</td>
<td>17.520</td>
</tr>
<tr>
<td>5</td>
<td>TR0240500055</td>
<td>1</td>
<td>HF</td>
<td>MC4BHM</td>
<td>75.343</td>
</tr>
</tbody>
</table>
### Biological Assay Protocol 1 to 2:

<table>
<thead>
<tr>
<th>Row</th>
<th>Assay D</th>
<th>Version</th>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MCIHBM</td>
<td>2</td>
<td>Binding high throughput screen; no replicates; no replicate</td>
<td>HTS</td>
<td>Mast80 to HScR01 for b-HTS</td>
</tr>
<tr>
<td>2</td>
<td>MCIHBM</td>
<td>1</td>
<td>Binding high throughput screen version not in use</td>
<td>HTS</td>
<td>Mast80 to HScR01 for b-HTS</td>
</tr>
</tbody>
</table>
**Name and Description**

**Name:** AF153431

**Description:** Homo sapiens melanocortin 1 receptor (MC1R) gene, complete cds.

**Source:** GENBANK

**ID:** AF153431
Integration decreases the learning curve for a set of applications by providing a unified interface and allowing them to share information.

LION DiscoveryCenter provides integration:

- between chemical applications
- between chemistry and related drug-design disciplines
Acknowledgements

- John Harby
- Ty Jacobs
- John Jaeger
- Lewis Jardine
- Chris Knauft
- Anatoli Krassavine
- Darryl Leòn
- Scott Markel
- Stephen Porter
- Adrian Smith
- Manish Sud
- IT Groups
Questions?

mitchell.miller@lionbioscience.com