

# Application Integration

## Three Aspects of Integration at AstraZeneca

# Integration Definition

Connecting information and systems

Across Technical, Scientific, and  
Organizational Differences

So the whole is more valuable than the  
sum of the parts

# Challenges in Technology Integration

- Systems from different vendors
- Internal and external systems
- “Compute engines” and Data sources
- Limbic System



# Technology Integration Example

## *Library Design Application*

- Many Vendor Components
  - MDL; Daylight; Spotfire; ACD; etc.
- Internal Components
  - Descriptor Calculations
- Multiple Data Sources
  - Internal Oracle DB for Reactions
  - ACD for Reagents
  - Afferent DB for Library Information

# Library Design – Integrating Components

Current User: rlawson Enumerator Help

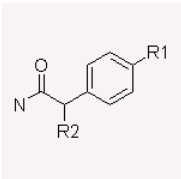
**Select Scaffold + RGroups**

Virtual Library Name:  Previous Results:

Number of RGroups:  Delete Previous Results: [\[Delete Previous Results\]](#)

Clear Form: [\[Clear Form\]](#)

Core Structure



**NOTE:** The enumerator will fail and generate errors when:

- The number of RGroups selected exceeds the number of RGroups in the Core Structure.
- If the opposite is the case, then any additional RGroups on the Core Structure will be ignored.
- The RGroups are not numbered sequentially beginning with R1.

Generate Unique IDs:  Generates a numeric ID for each enumerated product (required by certain registration systems)

Generate Plate Layout:

---

Source SD File

[\[UPLOAD a FILE\]](#) to your home directory on the SciTegic server  
Or browse for a file already located on the SciTegic server  
File locations must be specified with a UNC path, such as \\usbodap03\SciTegic\Users

[\[Reset File\]](#)

Compound ID Field:  [\[Display Fields\]](#)

---

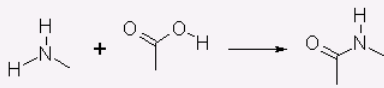
Current User: rlawson Select Reaction + Reagents

Virtual Library Name:  Previous Results:

Number of Reagents:  Delete Previous Results: [\[Delete Previous Results\]](#)

Clear Form: [\[Clear Form\]](#)

Reaction Structure



[\[Load Reaction Template\]](#)

Multiple Mapping Options:

Limit Reagent Mapping:  Match the reagent SD files with the reactant substructures in the same order as in the reaction.

Generate Unique IDs:  Generates a numeric ID for each enumerated product (required by certain registration systems)

Generate Plate Layout:

---

**Reagent 1**

Source SD File

**Product or Reaction Based Enumeration**

**ISIS Draw, Afferent, Multiple Enumerators Supported  
Able to Save and Retrieve Cores, R-groups, Libraries**

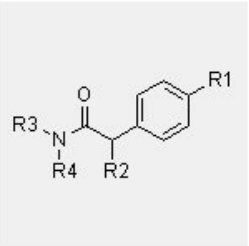
# Library Design – Integrating Components

Select Scaffold + RGroups

Virtual Library Name:

Number of RGroups:

Core Structure



**NOTE:** The enumerator will fail and generate selected is greater than the number of RGroup. If the opposite is the case, then any additions will be ignored.

RGroup 1

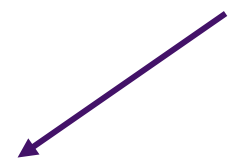
Source SD File

[UPLOAD a FILE](#) to your home directory on the SciTeGic server  
Or select a file already located on the SciTeGic server  
File locations must be specified with a UNC path

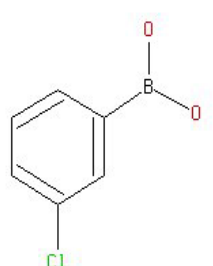
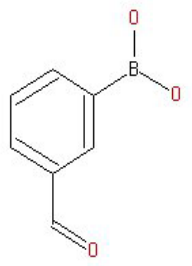
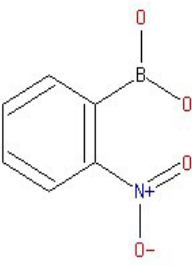
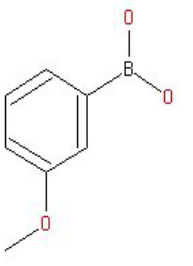
Number of Reagents to Use:

Leave blank to use all reagents

**Potential Reagents from  
ACD Profiled with ACD  
LogP and Other  
Descriptors**

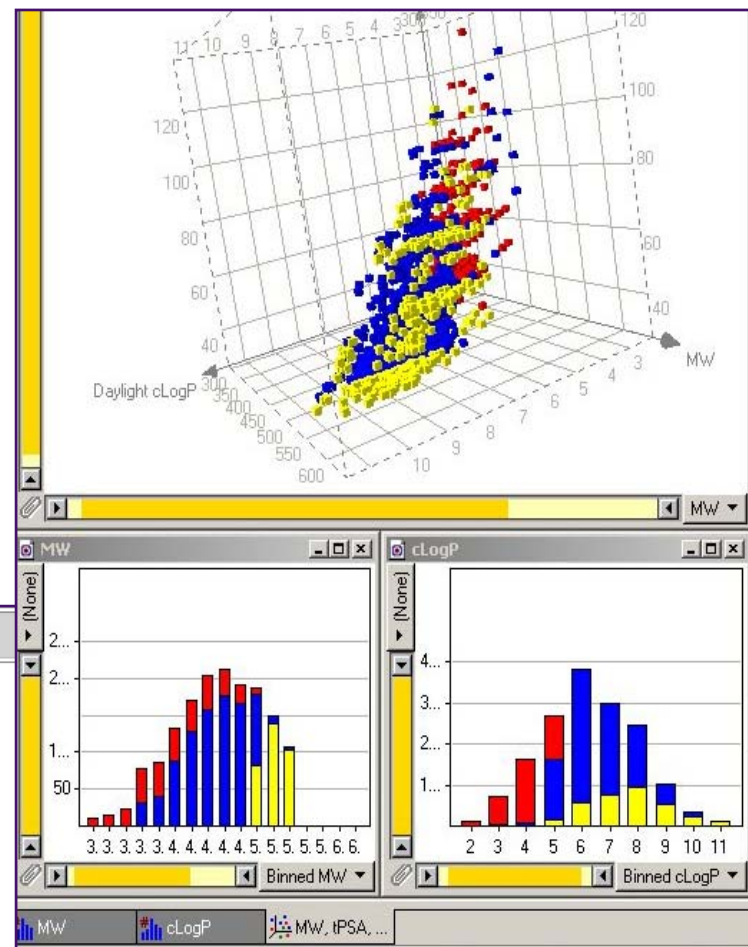


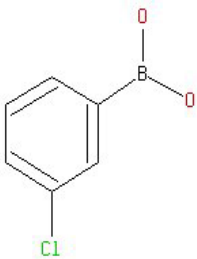
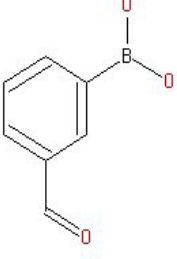
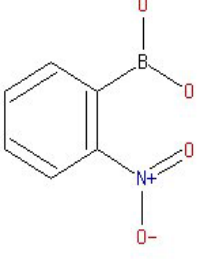
LipinskiScore: 0
 Select Sect

			
<input type="checkbox"/> 111985 ACD ID      MFC000161354 SMART ID    SS00002988 MW            156.38	<input type="checkbox"/> 111987 ACD ID      MFC000161356 SMART ID    SS00003680 MW            149.94	<input type="checkbox"/> 111989 ACD ID      MFC000161358 SMART ID    N/A MW            166.93	<input type="checkbox"/> 111990 ACD ID      MFC000161359 SMART ID    SS00005318 MW            151.96

# Library Design – Integrating Components

*Reagent Space  
Visualized and Selected  
within Spotfire*



LipinskiScore: 0		
		
<input type="checkbox"/> 111985	<input type="checkbox"/> 111987	<input type="checkbox"/> 111989
ACD ID	ACD ID	ACD ID
MFCD00161354	MFCD00161356	MFCD00161358
SMART ID	SMART ID	SMART ID
SS00002988	SS00003680	N/A
MW	MW	MW
156.38	149.94	166.93
<input type="checkbox"/> 111990		
ACD ID	ACD ID	
MFCD00161359		
SMART ID	SMART ID	
SS00005318		
MW	MW	
151.96		

# Technology Integration Foundations



- SOAP and Web Services
  - Neutralizes Proprietary and Vendor Interface Issues
  - Increasingly Common Skills and Availability
- Framework for Sharing Components
  - Still a Work in Progress
- Evangelism
  - With Customers and Solution Partners
  - Share Information; Sell the Vision



# Challenges in Scientific Integration

- System Diversity
  - Unfamiliar Vendors
  - Different Favorite Languages, Methods, etc.
  - Indirect mapping of identifiers or ontologies
- Understanding
  - Terminology
  - Requirements
  - Opportunities



# Scientific Integration Example

## Kinase DB – Targets through Leads

- Genes and Gene Sequences
- Expression Levels
- Protein Structures
- Small Molecule Structures
- Biological Screening Data

# Global Kinase Database

**KINASE DATABASE - Microsoft Internet Explorer provided by AstraZeneca**

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media History Mail Print Edit Discuss

Address <http://kinase.rd.astrazeneca.net/kinase/default.asp>

Links AP Bos Lead Inf CLASS DIPP GCF GPCR DB ICU Kinase DB Mölndal Montréal SPTMNG LGG PC

Google Search Web Search Site News PageRank Page Info Up Highlight

**kinase db** News Search Links  
Further evidence in IRESSA's favour

Gene | Assay | Toxicity | Compound | Download | SRS | Help

Today is: WEDNESDAY, 16 JULY 2003

**+ INTERNAL LINKS**

- × AZ Kinase Projects
- × AZ Kinase web
- × Kinase Update
- × Kinase Initiative
- × ASSETS
- × Cancer Discovery

**+ EXTERNAL LINKS**

- × Protein Kinase Resource
- × Human Kinome

**+ KINOME ANALYSIS**

- × Fold change analysis of kinome in 4 major tumor types
- × Amplification analysis of kinome in tumors by CGH analysis

**+ HELP**

- × Help Pages
- × AZK# History

**+ ABOUT**

**Quick Gene Search:**

**news**

[05.19.03] New features in this release:

- × IC50 data for all protein kinase assays within AZ
- × Substructure search for all compounds assayed against kinases
- × First release of Kinase Toxicity DB, which is in the process of being populated
- × Kinome analysis: Kinases that are over-expressed, amplified, or contain activation mutations in major tumor type
- × Patent information for Tyrosine Kinases
- × Improved homepage design and many enhanced functionalities

Here are some of the things we are currently working on and they will be available soon:

**search options**

please make a selection


- × If you want to add annotations to Kinase db please [read this](#).

**questions & comments?**

- × Please send your feedback to the [Kinase Db Team](#)

# Gene Information

## Links to Internal and External Sources

 GSP-L IHL - E H F E E D K Y I K Y N S N S G - - F V S N - A A R E T P Q A F S R E T P E R S C H O M M V Y D I G C V G D L Y S S P S S N G Q P I L C S I - E P L L V G E R F S N S N S Y G - - A V L T - N - R S T R Q A F S R E T P Y E L S N K O M I V D I G C V D D L Y S S P S


Gene | Assay | Toxicity | Compound | Download | SRS | Help

**Gene:** AZK00457

<b>Name</b>	EGFR
<b>Synonym</b>	EGFR, ERBB1, ERBB, HER1
<b>HUGO</b>	EGFR <a href="#">E</a>
<b>Description</b>	Epidermal growth factor receptor precursor (EC 2.7.1.112).
<b>GO</b>	Function <a href="#">+</a> , Process <a href="#">+</a> , Component <a href="#">+</a>
<b>Kinase Class</b>	EGFR: epidermal growth factor receptor <a href="#">+</a>
<b>LocusLink</b>	1956 <a href="#">E</a>
<b>Unigene</b>	Hs.77432 <a href="#">E</a>
<b>Chromosome/Cytogenetic Band</b>	7p11.2 <a href="#">+</a> ;7p12 <a href="#">+</a>
<b>Protein Structure</b>	6 External <a href="#">+</a>
<b>Patent</b>	IDDB <a href="#">+</a> MicroPatent <a href="#">+</a> GeneSeqP <a href="#">+</a>
<b>AZ Activity</b>	ASSETS (Target System) <a href="#">E</a>
<b>Screening Results</b>	EGFR_L <a href="#">+</a> EGFR_E <a href="#">+</a> EGFR_I <a href="#">+</a> EGFR_C <a href="#">+</a> EGFR_F <a href="#">+</a> EGFR_G <a href="#">+</a> E



















Gene Expression	
Human: HG 095	4507 <a href="#">+</a> <a href="#">E</a>  2707 <a href="#">+</a> <a href="#">E</a>  1574 <a href="#">+</a> <a href="#">E</a>  1005 <a href="#">+</a> <a href="#">E</a>

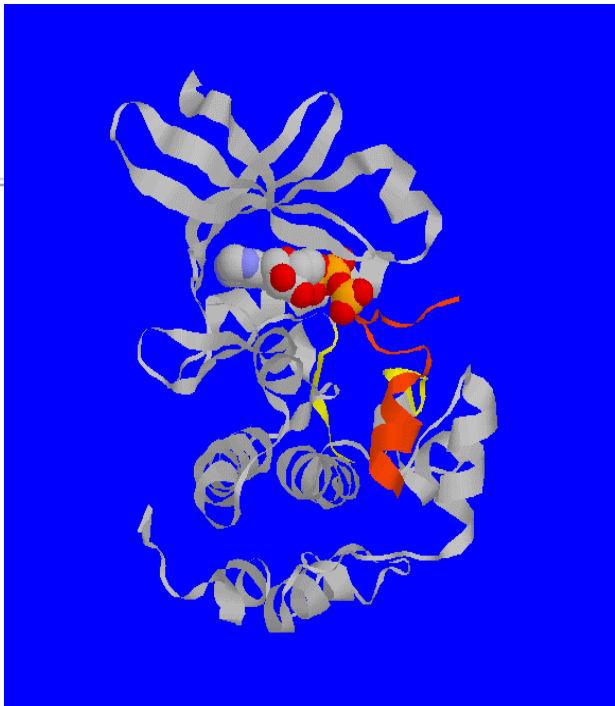
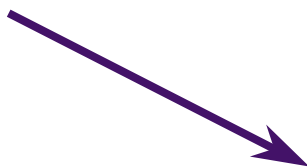
# Protein Structure

 kinase db

Gene | Assay | Toxicity | Compound | Download | SRS | Help

Gene: AZK0C157

Name	EGFR
<b>Protein Structure</b>	
1IVO    (PDB)	
1NQL    (PDB)	
1M14    (PDB)	
1M17    (PDB)	
1DNR    (PDB)	
1DNQ    (PDB)	



## Internal and External Sources

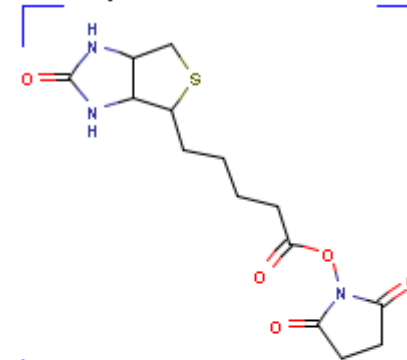
# Small Molecule Perspective

## plC50 selectivity

2262 hits found for TEK\_b   [Select columns](#)

[why are there now fewer columns?](#)

Compound No: M24



compound_no	TEK_b	ERBB2_c	TEK_c	IGFR_b	SYK_b	ZAP70_b	FAK	MAPKAPK	GSK3	PDPK1	IGFR	LCK	Aurora-a1	CDK4	MET_b	CHK1	TEK	TRKB_d	CDK2	KDR_b	FAK_b	FLT1_b	TRKB_e	ZAP70	JAK3	PLK	AKT1_d	SRC	Aurora-a2	P38a
<a href="#">M2</a>																														
<a href="#">M2</a>																														
<a href="#">M4</a>																														
<a href="#">M3</a>																														
<a href="#">M3</a>																														
<a href="#">M3</a>																														
<a href="#">M1</a>																														
<a href="#">AR-CC0000081</a>																														

Assay	IC50 (nM)	plC50	N
<a href="#">IGFR_b</a>	20.9	<a href="#">7.68</a>	1
<a href="#">ERBB2_c</a>	26.9	<a href="#">7.57</a>	1
<a href="#">TEK_b</a>	33.1	<a href="#">7.48</a>	1

Gene	A	18	+
Alias	E1		
Description	Dose response for inhibit		
Purpose	To identify novel inhibitors of ERK2 kinase using an		
Project	EFT		
Bioscientist	W	M	
Chemist	Peters JW		
Registration Date	25-JAN-2000		
Type	HTS		
Last Update	25-OCT-2000		
Location	Alderley Park		



# Scientific Integration Foundations

- “Internally” Agreed, Standards
  - Ontology – With Appropriate Scope
  - Identifiers and Relationships (Compounds, Samples, Targets, etc.)
- Agreed Application Interfaces
  - Web Services
  - Owned, Maintained by right people
- Strong Sponsor
  - Focus on Requirements and Solutions, not history or turf

# Challenges in Organizational Integration

- Geography
  - 8 Major Sites in 4 Countries
  - No global coffee room
- Departments
  - Ownership, Control, Resources
  - Decision Making Authority
- Limbic Systems





# Library Design Tools - Again

***We Value Diversity!***

***“SaSA & Excel  
Parallel Synthesis”***

***Pipeline Pilot Library Tool***

***“Lund Tool”***

***Accelrys Components***

***“Product Based Design”***

***Triplos Tools***



# Organizational Integration Principles



- **Single, Global Project**
  - Clear Leadership – Project Manager + Sponsor
  - Make Carrots and Sticks Obvious
  - Involve Right People – Scientists, Technical, etc.
- **Integrate Appropriately**
  - Core Computing Components – Critical, Expensive
  - Flexible UI – Cheap, Local Accents
- **Communicate Big Picture**
  - On This Project and Overall

# Communities of Practice

- Routine Virtual Meetings Across All Sites
  - NetMeeting & Phone + Web Repository
- Means to Share Information, Solutions, Best Practice
  - Screening Informatics
  - Chemical Computing
  - Electronic Lab Notebook

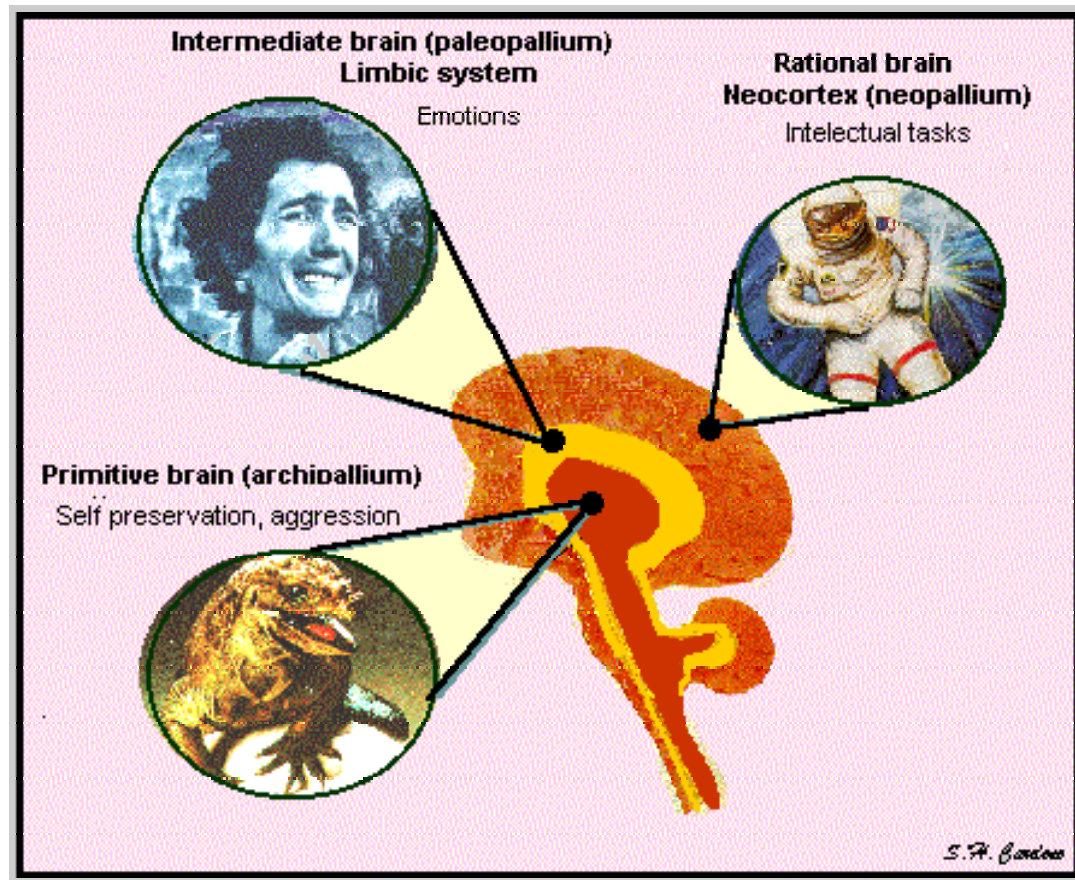
The screenshot shows the 'Screening Informatics Community' website. At the top, there is a navigation bar with 'DISCOVERY PROJECT HOME PAGE' and a search box. Below this, the page is divided into several sections: 'PROJECT MENU' (listing navigation options like 'Back to homepage', 'Add Event', etc.), 'QUICK LINKS' (listing 'Home', 'Search', 'Directory', etc.), 'MY PROJECTS' (listing 'ACD Labs and AZ', 'ActivityBase', etc.), 'DETAILS' (showing project information for 'Screening Informatics Community'), 'NEWS' (listing recent news items with dates and titles), 'LINKS' (listing various links with dates and titles), 'MEMBERS' (listing community members with their roles and access), and 'LATEST DOCUMENTS' (listing recent documents with dates and titles). The interface is colorful and includes various icons and a logo for 'Our Discovery'.

Rich Lawson

ACS Meeting NY

# Evangelism and Human Factors

*The Greatest Challenges!*



# Conclusions

- **Technical Integration Easier than Ever**
  - SOAP, Web Services, etc. Enabling
- **Domain Integration Valuable + Improving**
  - Standard Internal Ontology & Business Rules
  - Agreed ID's and Interfaces
- **Human Factors are Critical**

# Questions or Comments?