Needles in hayfields: strategies for rapid HTS triage analysis

Jack Andrew Bikker, James B. Dunbar Jr., Dirk Bornemeier, David Wild, Alain Calvet, Christine Humblet

Pfizer Global Research and Development,
Ann Arbor Laboratories
“HTS Triage facilitator”
PGRD-AA

- **Computer assisted drug discovery**
  - Jim Dunbar
  - Dirk Bornemeier
  - Mark Snow
  - Eric Gifford

- **Chemistry**
  - Alex Bridges
  - Chuck Stankovic
  - Ellen Dobrusin
  - Joe Warmus
  - Andy White
  - BK Trivedi
  - Denton Hoyer
  - Bob Kennedy

- **Scientific Computing**
  - Sherry Marcy
  - David Wild
  - Alain Calvet
  - George Cowan
  - Rob Goulet
  - John Blankley

- **Discovery Research Informatics**
  - Christine Humblet
  - Heather Barbre
  - Chandana Kathuria

- **HTS screening**
  - David Mack
  - Jim Marks
  - Maxine Santoro
Others

- Tripos Associates
  - Bob Clark
  - Stephan Reiling
  - Fred Soltanshahi
  - Roman Dorfman
  - Dick Cramer
  - Rob Jilek
  - Roman Dorfman
  - Trevor Heritage

- PGRD- elsewhere
  - Alex Alex (Sandwich)
  - Mike Miller (Groton)
  - Cris Hageman (La Jolla)

- Bioreason
- Leadscope
Outline

• HTS triage process
• Software
  - Suite of tools
  - SARNavigator
• Examples
  - A kinase
  - A fatty acid synthase
Question: What is HTS triage?

Assumptions:
1 million compounds/ screen
Hit rate of 0.3%  \rightarrow  3000 Hits/screen
"Many" screens per year
Quality filters

Challenge: chemically diverse screening library contains undesirable ligands

Solution: Defined SMARTS filters for
  Reactive: chemically reactive groups
  Risky: known toxicophores, etc.

Culled entire database
Electronic and physical elimination of reactives
Database alert on risky compounds
Series perception

Traditional: chemist organizes hits into series

Computational: computer groups similar compounds for chemist

Screened Compounds

Filter for Quality ➔ Identify hits ➔ Organize hits into series ➔ Mine legacy data ➔ Identify Related Compounds

SARNavigator: non-linear mapping
Bioreason: Kohonen maps
Hitsifter: clustering of compound fingerprints
Leadscope: sorts into pre-defined chemical bins
Distill: sorts by maximal common substructure
Series perception

Combinatorial: core, sidechains, protocol available from database

Caveat: If it's not in the database, it doesn't exist...
High throughput decision support
- series history (name that compound)
- series ennui
- toxicity alerts
physico-chemical: MW, RO5
- novelty

Low throughput decision support
- new PK studies
- patent searches
- research reports
- analytical analysis
- resynthesis

Decision support tools

Screened Compounds
Filter for Quality -> Identify hits -> Organize hits into series

ADME/Phys Properties
Screening History
Previous Decisions
Pfizer-wide Databases
Mine legacy data
Identify Related Compounds

Series Selection Decisions
Characteristics of analysis

• Many many decisions, made very quickly
  - First cut is the deepest
  - Yes (10%), Yuck (30%), Don't know (the rest)

• Compound interrelationships
  - Series: synthetic feasibility
  - Pharmacophore: common chemical elements
Outline

• HTS triage process
• **Software**
  - Suite of tools
  - SAR Navigator
• **Examples**
  - A kinase
  - A fatty acid synthase
Tripos SARNavigator
Collaboration with Tripos (St. Louis)  
Non-linear mapping

Tripos SARNavigator
Collaboration with Tripos (St. Louis)  
Non-linear mapping

Pfizer

Hitsifter
Hierarchical clustering of Daylight fingerprints  
Web-based actives-only tool

Bioreason DrugPharmer
Product of prior collaboration  
Fragment perception and tree building

Leadscope
Product of prior collaboration  
Predefined chemical bins
SARNavigator

Screened Compounds

Filter for Quality → Identify hits → Organize hits into series → Mine legacy data → Identify Related Compounds

ADME/Phys Properties → Screening History → Previous Decisions

Pfizer-wide Databases

“Knowledge base” decision support tool

SARNavigator
SAR Navigator - Main Screen

Data Visualization

Structure Depiction

Spreadsheet (Linked to Knowledge-Base System)

Color/Size Widget
Another way of looking at it

Print off hits, one sheet at a time. Sort similar compounds together. Leave singletons on edge.

Find inactives that are similar to hits. Place under closest signpost. If <30% similar, discard. Repeat for ~300 000 compounds.

Look at the results from a distance

NLM!!
SAR Navigator - Main Screen

Data Visualization

Structure Depiction

Spreadsheet (Linked to Knowledge-Base System)

Color/Size Widget
SAR Navigator - Subset Drilldown Screen

Second linked session (see actives and inactives)

Multiple types of structure visualization capabilities 1X2, 2X5, 1X10, etc.

Sliderbars linked to spreadsheet

Colors linked to original or independent
SAR Navigator - Full Subset Visualization

Structure Depiction of Entire List in This Subset
Outline

• HTS triage process
• Software
  - SAR Navigator
• Examples
  - A kinase
  - A fatty acid synthase
The process: kinase assay

- 1040 structures 60% inhibition or greater at screening concentration
  - 306 combinatorial
  - 734 med chem
  - 188 “reactive”
  - 19 “risky”
- Coloured by %inh@20µM
  - Red 85-110
  - Magenta 110+
  - Green 50-85
  - Blue 0-50
  - Grey <0 or not tested
- 428 compounds in centre, 612 at edges

Centre: At least one other “signpost” 70% similar

No other “signpost” 70% similar
Combi signposts: 306 compounds

Red 85-110
Magenta 110+
Green 50-85
Blue 0-50
Grey <0 or not tested

Compound 0006008-0000 has 2 members
1040 out of 1040 compounds visible, 398 selected
Excluded signposts

Distance: 0.40

Red 85-110
Magenta 110+
Green 50-85
Blue 0-50
Grey <0 or not tested

188: “reactive”
330: JW/ED/JAB
Outcome

• 10 series identified and prioritized
• Change in methodology: compounds from series of interest with lower %inhibition ("activity") sent to secondary testing
• Combinatorial follow-up ongoing
• Team input to triage process

• 20+ enhancements to software suggested
Outline

• HTS triage process
• Software
  - Suite of tools
  - SAR Navigator
• Examples
  - A kinase
  - A fatty acid synthase
Fatty acid synthase

- Triage of collection made up entirely from combinatorial libraries
- Chemistry, core, reagents, series available from database
Collection screened contained only compounds from combinatorial libraries

Combinatorial core, sidechains, chemistry ID available from database

Triaged using parallel coordinates view within Spotfire

330 hits defined
Fatty acid synthase

predefined series projected in NLM
Fatty acid synthase: by chemistry

Some favorable groups

Range of inhibition

% INH

RX A  RX B  RX C
Pattern - reactive or rewarding?
Outcomes

• 330 hits triaged to prioritized series in one 2-hour session
• Pre-organization greatly facilitated analysis
• Follow-up ongoing
The near future

Highly summarized data

Providing data contributing to a decision

Capable of challenging prejudice