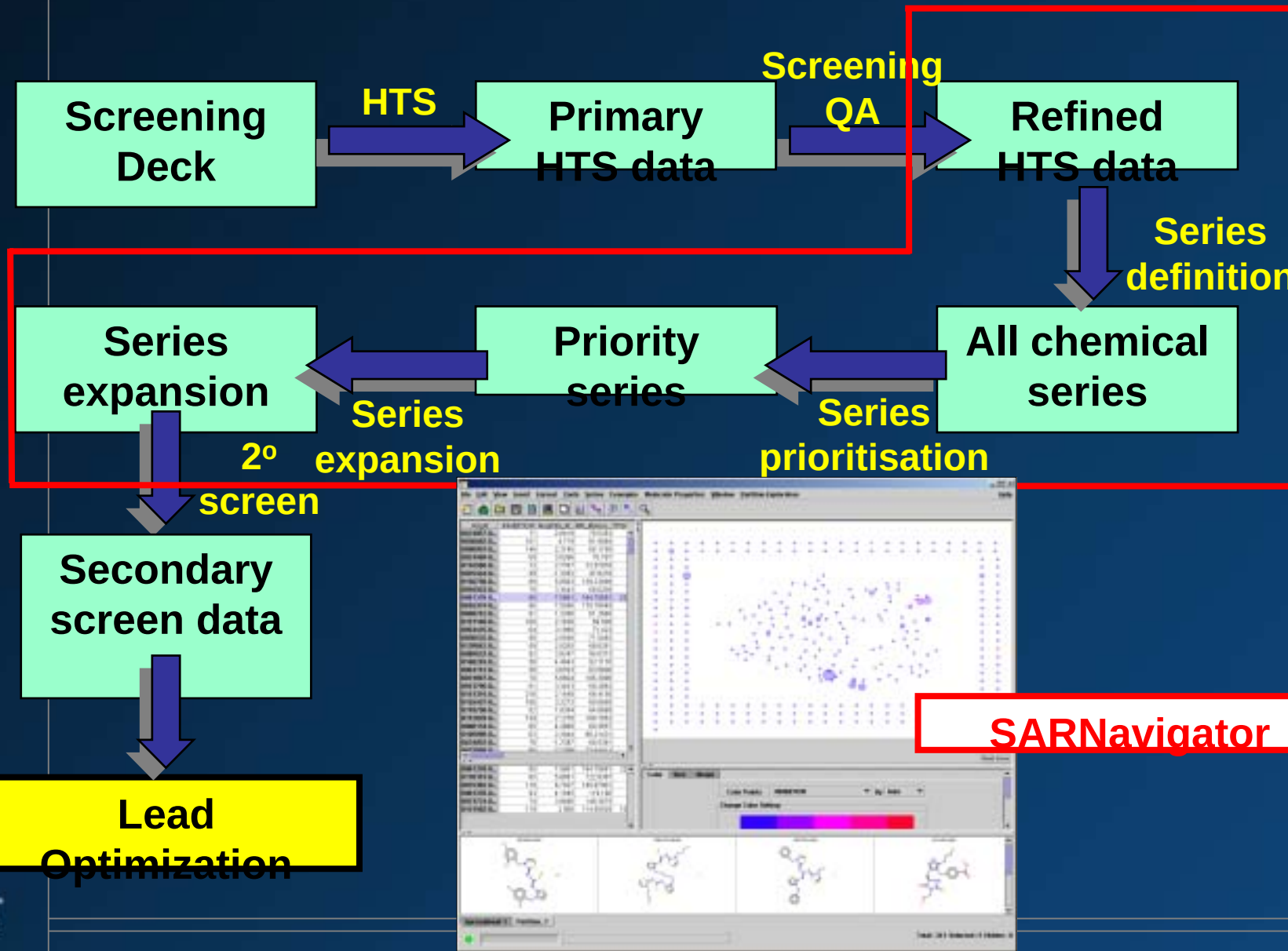
SARNavigator & HTS data analysis

Tripos Inc.

09/02/03

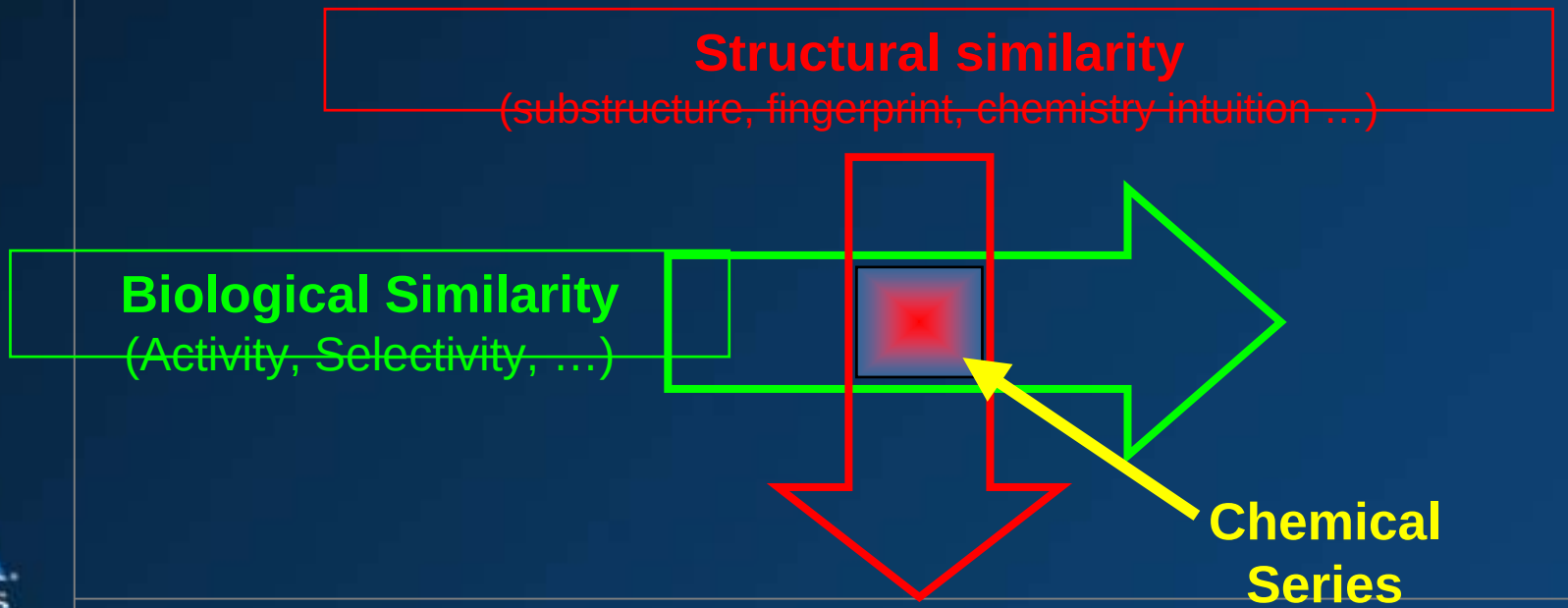
HTS Workflow



Tripos

Chemical Series

- **Goal: definition of chemical series from HTS data for continued research**
  - Secondary screening, follow up chemistry
  - First stage of hit ⟹ lead
- **What defines a chemical series?**
  - common core/synthetic pathway/what I say, …?



Tripos

Exploratory Data Analysis

"I don't know exactly what I'm looking for - but I'll know when I see it!"

- *… chemist (reviewing HTS results)*

- **EDA Methods:**
  - **visual analysis of hit structures**
    - chemists with print-outs of hit structures
    - subjective/slow/irreproducable
  - **clustering and cluster analysis**
    - hierarchical/non-hierarchical clustering clustering
    - data quality is paramount/use of negative data/chemical series?
  - **structural property/activity visualization**
    - multidimensional graphs
    - relationship to actual structures/view of all relevant data?
  - **Visualisation of high-D spaces**
    - NLM/MDS to reduce dimensionality

Tripos

Visualisation of HTS Data

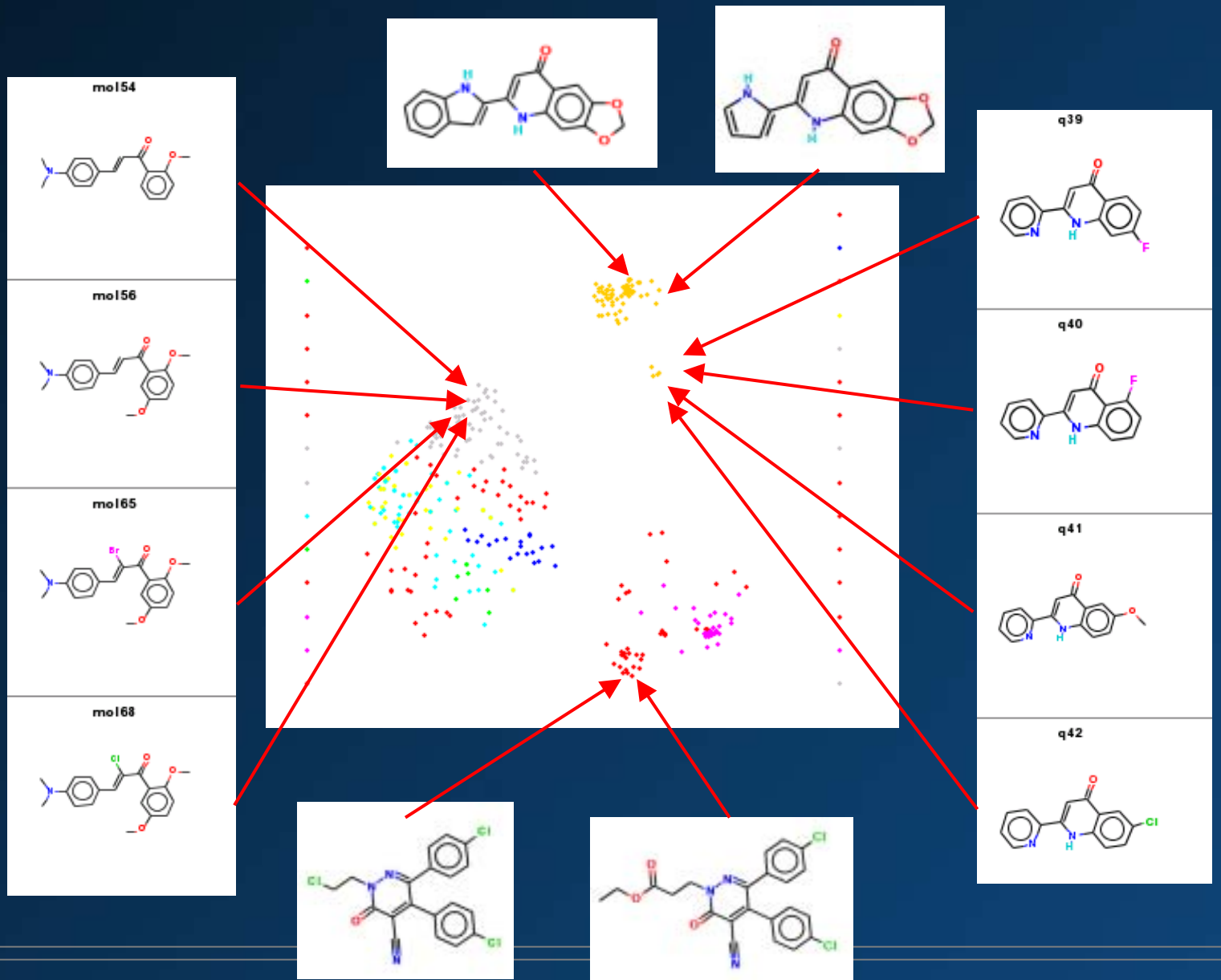
- **View entire SAR landscape**
  - **2D plot of HTS data**
    - either clusters or compounds
  - **such that similar compounds are close on the plot**
  - **map activity and other data onto plotted points**

- **Uses combination of clustering and PCA/NLM**
  - **select diverse, representative compounds (signposts)**
  - **other compounds partitioned into signpost clusters**
  - **run PCA/NLM on signpost properties and plot**
  - **similar clusters will be close to each other**
  - **user can drill down into clusters**

- **Only care about local similarity**

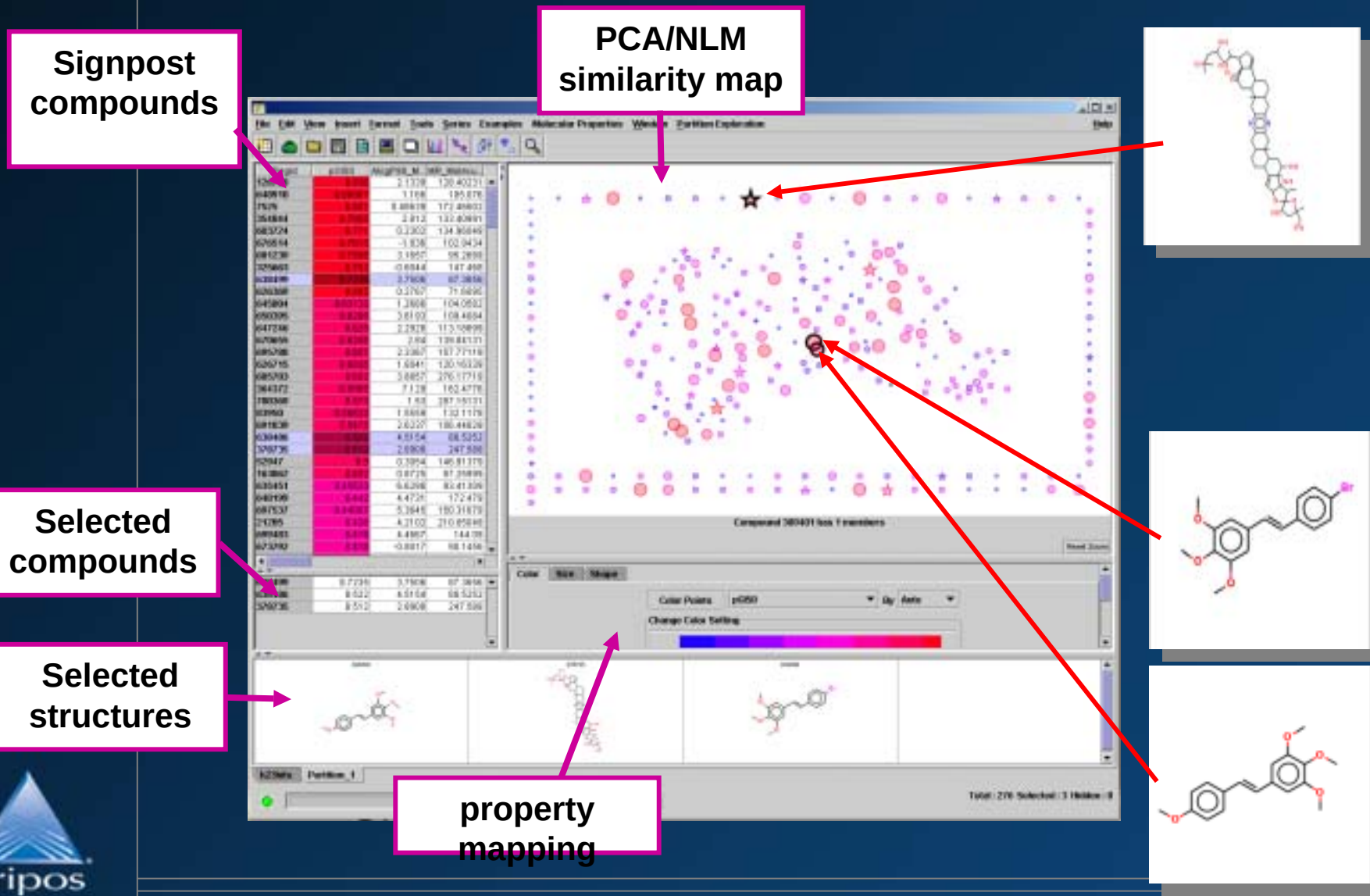
Tripos

PCA/NLM Plots with Horizon

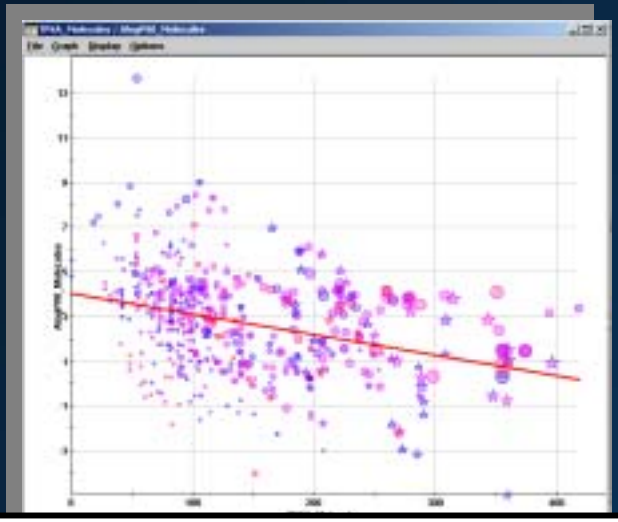


Tripos

SAR Landscape view

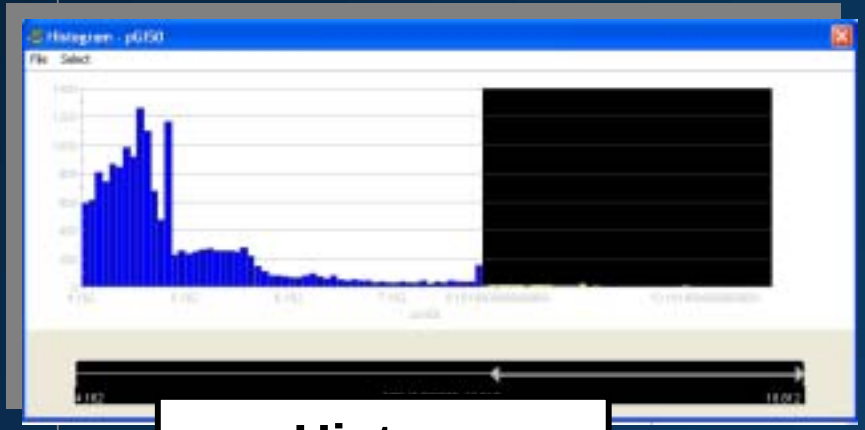


SARNav data visualisations



Scatter plots with data mapping

Bullseye plot
- single compound similarity neighbourhood

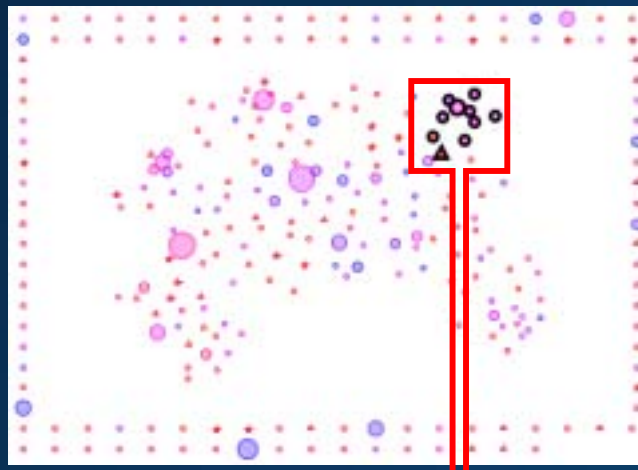


Histograms

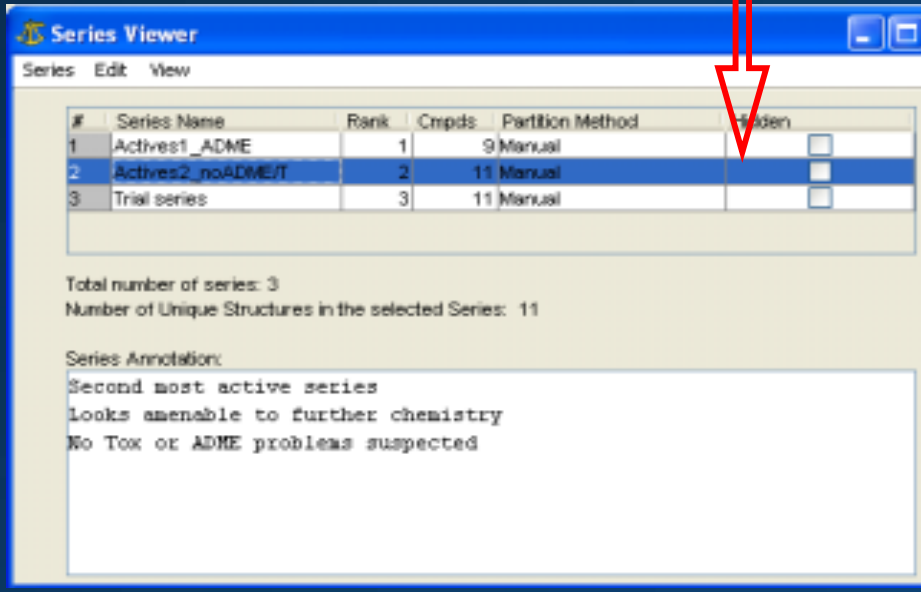
Correlation coefficients

Tripos

Series Definition



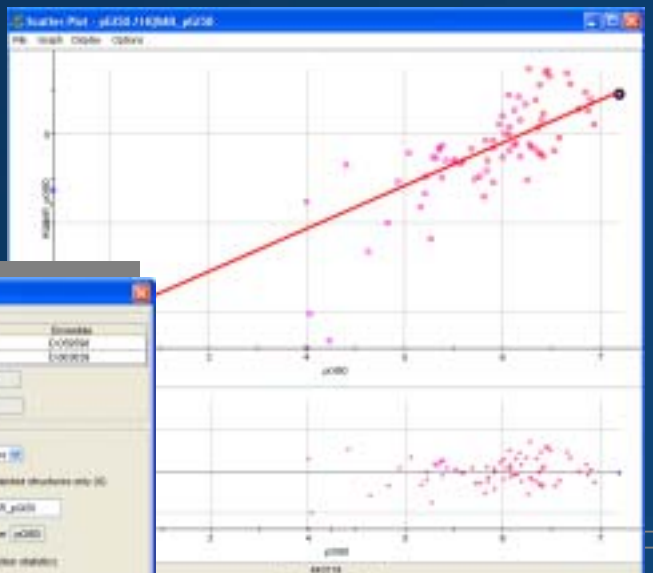
- **Interactive series definition**
  - selected cluster(s)
  - selected compounds
  - from substructure search
  - from property ranges
- **Series tracking**
  - series viewer
  - series annotation
  - series ranking
- ***Connection to SeriesBase***



Tripos

Series Prioritisation

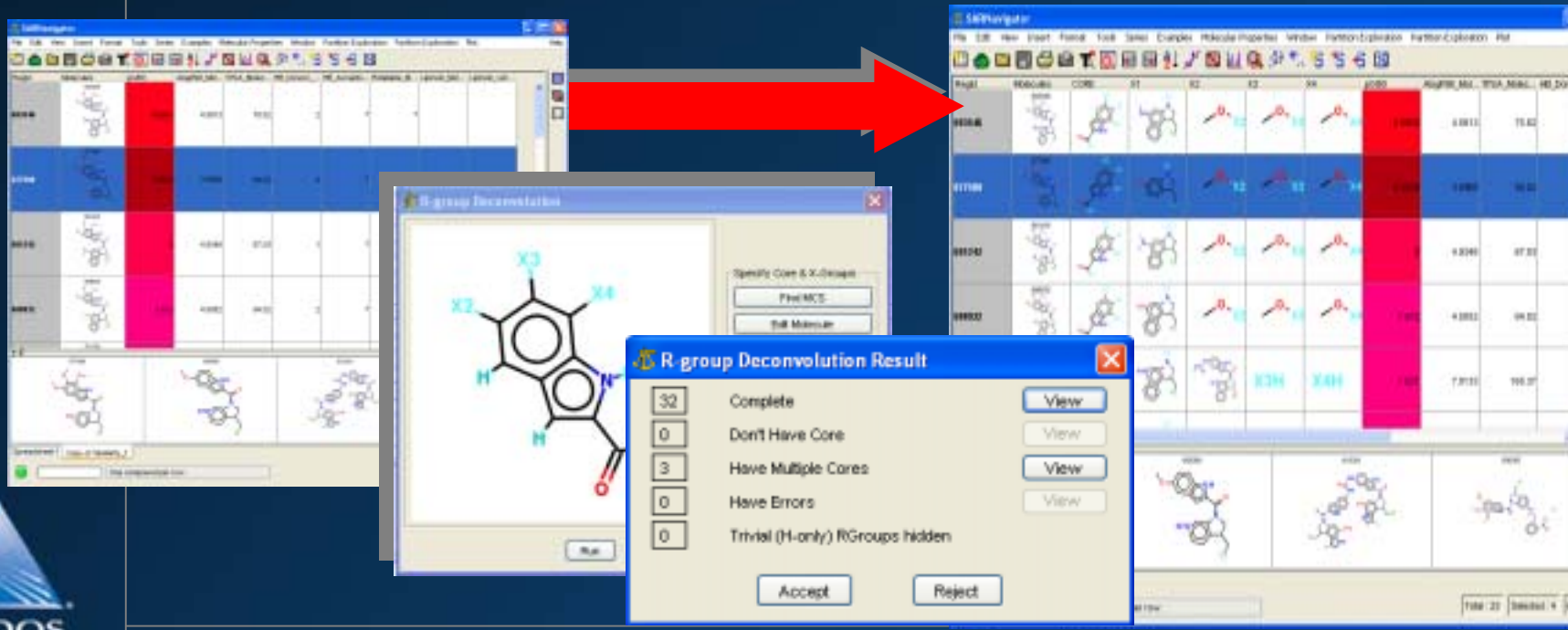
- **Visual analysis**
  - **biologist**
    - might these compounds have interfered with the screen?
  - **chemist**
    - what are the prospects for further chemical modification?
- **HQSAR**
  - **can a QSAR model be produced for the series**
  - **does HQSAR tell us anything about the SAR**
  - **rapid predictive QSAR**
    - based on molecular holograms



Tripos

Series Prioritisation

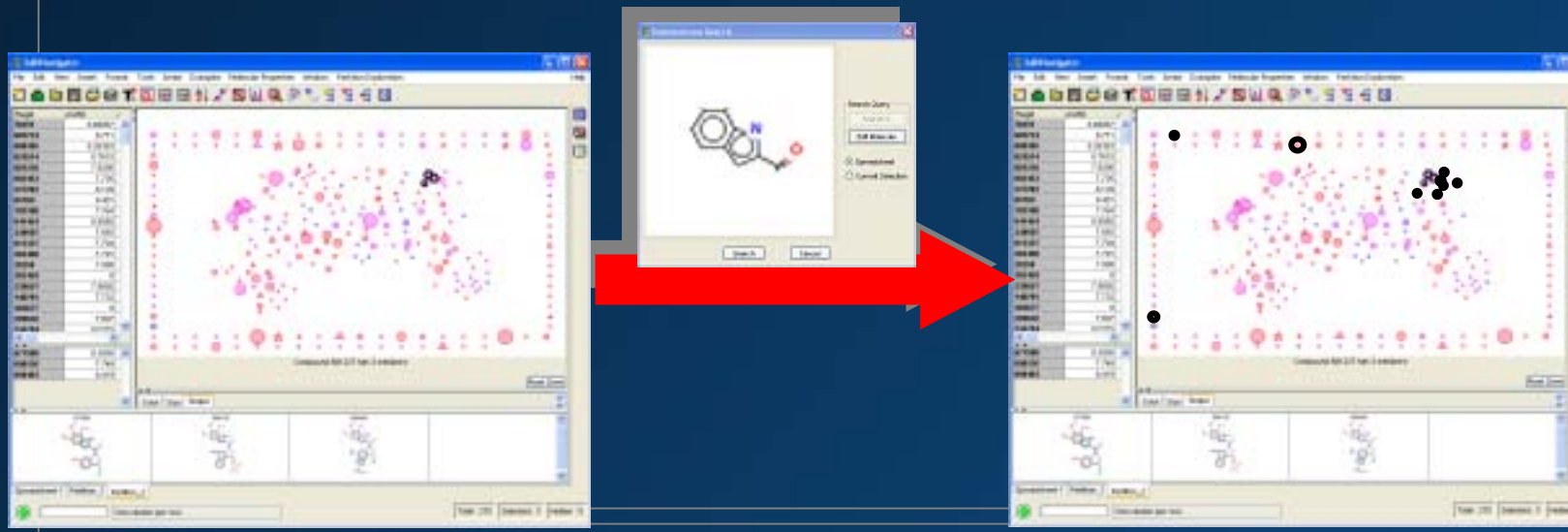
- **Is there any obvious SAR**
  - if so this may be a good series for 2º screening
- **R-group deconvolution**
  - determines scaffold through MCS calculation
  - splits structures into scaffold + R-groups



Tripos

Series Expansion

- **Once series have been defined and prioritised**
  - find other compounds that may be suitable for further evaluation
- **Expand series with**
  - clusters close to/between defined series
  - compounds with similar/R-groups scaffolds
  - compounds from outside the screening deck



Tripos

SARNavigator Workflow

Import data

Filter data

Cluster hits

Initial series exploration

Add non hits to clusters

Reexamine series with inactives

Define series

Prioritise series

SAR, QSAR property data

Expand series

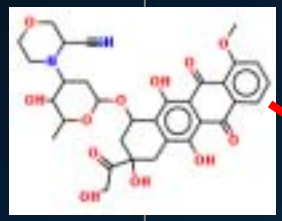
Substructure searching

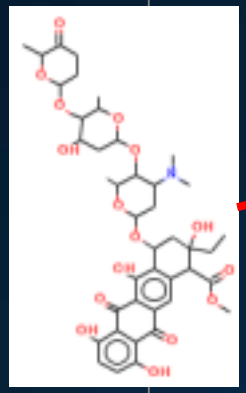
Tripos

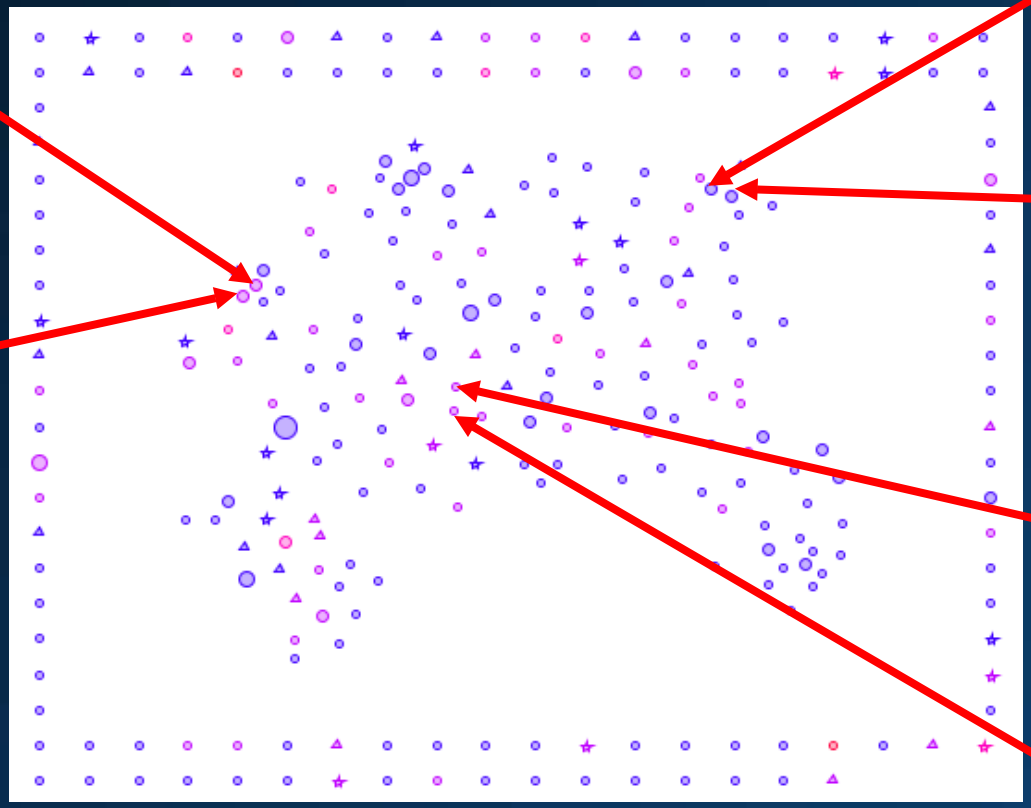
NCI H23 screen - example

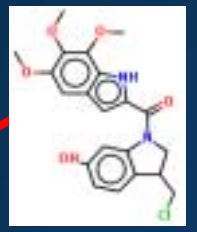
- National Cancer Institute Human Tumor data set
- Only one cell line used
  - NCI-H23 (non small cell lung)
- Activity data pGI50
  - -log concentration at 50% growth inhibition
- 35000+ compounds
- Activity threshold: pGI50 $\geq$ 6 (1848 actives)
- All compounds used

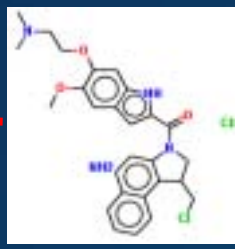
H23 - Actives Partitioned

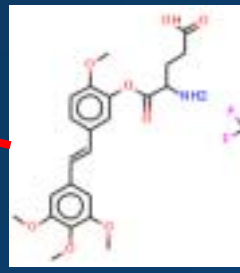


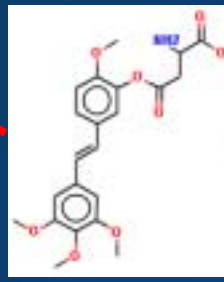












Mean cluster activity

Low — High

Mean cluster Lipinski violations

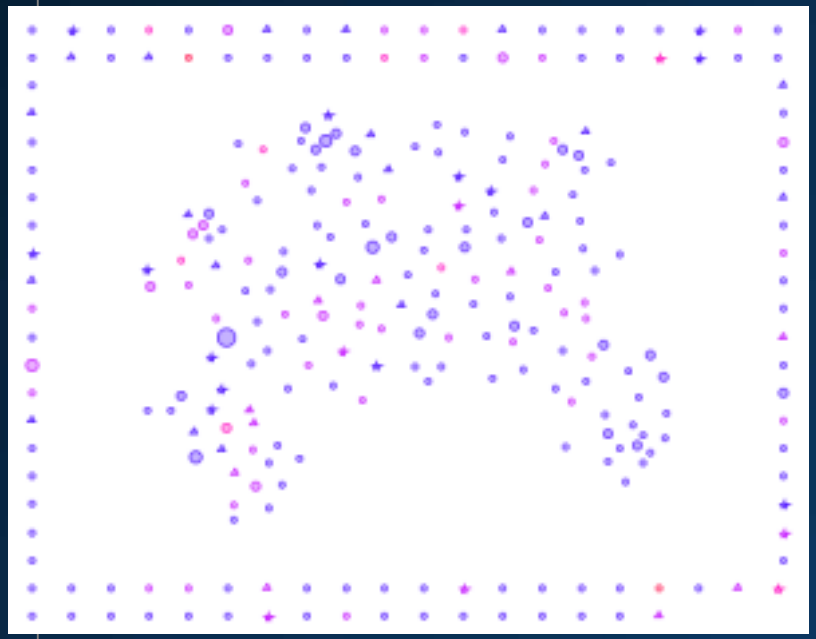
● <2 ▲ 2 - 3 ★ >3

Points sized by cluster size

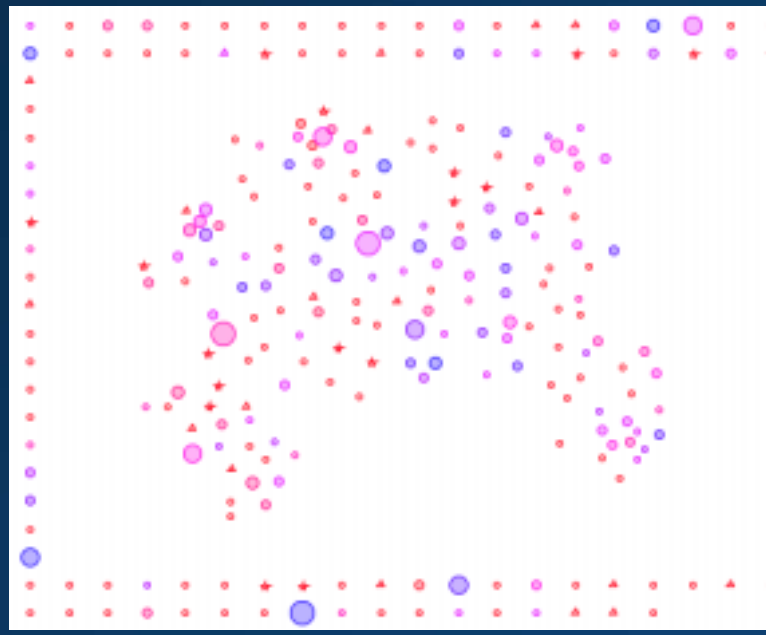
Tripos

Including the inactives

- partition actives into clusters and project
- add inactives to active clusters
- compare active *vs.* active & inactive projections



Before
Actives Only



After
Actives & Inactives
colour - %active cmpds in cluste

Tripos

Summary

- **SARNavigator**
  - **Support for HTS data analysis**
  - **Raw HTS data ⇨ prioritised chemical series**
  - **Modification and extension**