SARNavigator & HTS data analysis

Tripos Inc. 09/02/03





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?
 - o common core/synthetic pathway/what I say, ...?





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:

0

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



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
 - o map activity and other data onto plotted points

Uses combination of clustering and PCA/NLM

- select diverse, representative compounds (signposts)
- o other compounds partitioned into signpost clusters
- o 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

PCA/NLM Plots with Horizon



Tripos

SAR Landscape view



SARNav data visualisations



Scatter plots with data mapping





Bullseye plot - single compound similarity neighbourhood

| | 0.04 | 100 | 112 | 100 | 22 | 22 | | | | | | | | |
|---------------------------------|-------|-------|---------|---------|-------|----------|-------|-------|-------|-------|-------|-------|-------|----------|
| er ment call be a date and take | | - | let you | 1.0 | in a | 1 | | | | | | | | |
| about "N" means that there is | - | - | for 14 | ineed o | NR 00 | Linn. | | | | | | | | |
| Statistical states and | 14 | 12 | 18 | 1. | 1 | 1.4 | T | | 1.1 | 10 | 141 | 12 | 15 | 144 |
| (pORD) | | 0.80 | 8.27 | 0.23 | 8.71 | \$ 28 | ate | 0.27 | 8.29 | 0.37 | 0.36 | 8.08 | 18:13 | 0.21 |
| Allephin selecutes | 40.08 | 100 | 000 | 4.33 | 6.30 | -8.00 | 10.28 | -0.11 | -8.00 | 0.08 | 6.1 | -2.00 | 100 | -0.83 |
| MP Molecules | 0.37 | 8.00 | | 0.1 | | 1.00 | 6.57 | 1.00 | 100 | 0.94 | 167 | 8.05 | 8:50 | 0.741 |
| TTLA (Manules | 0.22 | 1133 | 110 | 1.0 | | 1.00 | 0.78 | 1.84 | 1.00 | 0.53 | 1.41 | 8.06 | 8.47 | 3.75 |
| 1-logil/CSOL/HDROMAN | 0.21 | 1.30 | | 1120 | | 111 | 0.27 | 9.62 | 1.17 | 0.55 | 2.80 | 8109 | 10.41 | 10.00 |
| AMM, Mulecialent | 0.29 | -0.83 | | | 471 | 101 | 0.51 | 1.10 | 1.00 | 0.58 | 10.00 | -0.05 | 4.51 | 10.00 |
| HE Dovies Manuality | 0.18 | 11.22 | 1107 | 10.00 | IL TT | 2.18 | | 10.0 | 1 | 0.01 | 10.00 | 8.02 | 8.76 | 11 M T |
| HE Acceptors Molecules | 037 | -0.17 | 1.1 | 10.54 | 842 | 100 | 0.78 | | | 0.53 | 0.58 | 8.09 | 2.45 | 0.74 |
| Herevy Molecules | 0.25 | -0.85 | | 11.1 | 1.75 | 100 | D M | | | 0.51 | 9.67 | -0.01 | 4.50 | 0.75 |
| C Potesste (Refessere | 10.27 | 8.00 | 2.18 | 11.82 | 111 | 8.50 | 21 | 8.81 | | | 11.22 | 8.00 | 8.33 | STATE OF |
| TERMINE_ARCHICLARD | 0.25 | -0.87 | 202 | 0.4 | 2.00 | 104 | 0.25 | 3.92 | 117 | -6-02 | | -0.00 | 4.9 | 0.49 |
| 2) Shuthame, Malancier | 0.81 | -0.85 | 0005 | 10.88 | 5.00 | 1-8.00 | 0.82 | 3.84 | -8.84 | 0.07 | -0.00 | 100 | 4.02 | -0.82 |
| TJ L'hoore_Motessare | 4113 | 0.81 | 18.80 | 10.45 | 41.81 | CR. R.C. | 6.34 | 10.88 | 0.40 | 1.11 | 11.27 | 487 | 1.1 | D.68 |
| ALLEWISH AND MONTHLESS | 0.21 | -0.83 | 10.04 | 11.78 | 1.11 | 1210 | D-84 | 10.04 | 4.12 | 0.91 | 0.45 | -0.62 | 1.41 | |

Correlation coefficients

Series Definition

Interactive series definition

- selected cluster(s) 0
- selected compounds 0
- from substructure search 0

lariac.

from property ranges 0



Series tracking

- series viewer 0
- series annotation 0
- series ranking 0

Total number of series: 3 Number of Unique Structures in the selected Series: 11

Series Annotation: Second nost active series Looks amenable to further chemistry No Tox or ADME problems suspected



Connection to SeriesBase

Series Prioritisation

Visual analysis

- biologist
 - might these compounds have interfered with the screen?

• chemist

what are the prospects for further chemical modification?

• HQSAR

o can a QSAR model be produced for the series

BUS - HERES STREET, HERES

- o does HQSAR tell us anything about the SAR
- rapid predictive QSAR
 - based on molecular holograms



Series Prioritisation

- Is there any obvious SAR
 - o if so this may be a good series for 2^o screening

R-group deconvolution

- determines scaffold through MCS calculation
- splits structures into scaffold + R-groups

| Te lä de bes fors lab ine Langle deutsfragen bids fasteliginge februigingen be | | | | | | | | | The Lat. | Alfrenzigator 18 view part fond took land Complex Mikola Hopeful Vinder Hantonbylinkko hattonbykokon Hat | | | | | | | | | | |
|--|-----------|------|-------|-------|-----|----------------|-------------------|--------|----------------------------|---|--------|--------|---------|-------|--------|---------|-------|------|------------|------------------|
| 201 | BCOT | | Map | 1.356 | B | No. | | | | | | | a to | 目目目 | JOU | Q. 例代 | | 5 69 | | |
| 1 | al fil | | ten . | 1 | - | ene per larie, | | | | | and a | or All | the set | 183 | NC _0. | 10 | | 1000 | AG490, MA. | 11.42 |
| | 100 | | - | - | 100 | Fgrang Dec | restation | _ | | X | | 200 | A | 104 | - | 1 | * | 10 | 1,000 | |
| **** | an fil | 1.00 | (min) | 39 | • | | xp | | · Specify Core & X. Donard | | | 14 Pa | Ja. | 188 | 10. | 10. | 10. | | 1210 | en |
| | est. frit | | ** | 3 | | ×2. | A | 6 | Processo The Marcular | | ***** | Ly'FR | J. | 83 | _0. | ~ | 1 | | | *** |
| ** | i Ch | 1 | | -0 | | | V. | 🖧 R-gr | oup Deconvolutio | n Result | | | | -00 | 190 | | | | L. | |
| | -day | | | 3 | ž. | 3 | JOJ | 32 | Complete | | | Vie | w. | 80 | .8. | 8.389 J | 2.411 | | 7,958 | 14.37 |
| - | -04 | | | - 36- | _ | | - | | Don't Have Core | | | Vie | nw - | mist | | | | | - | in second in the |
| | | | | | 4 | 3 | Have Multiple Cor | es | | Vie | :w | 14 | See. | | - | Ger. | | - | | |
| C. | | | | | | | | 0 | Have Errors | | | Vie | nw - | | age - | | 00 | R. | - | pa. 10 |
| | | | | | | | |] | Trivial (H-only) R | Groups hidd | en . | | | | °GG | | -95- | 1 | | 0, |
| | | | | | | | | | Accept | | Reject |] | | a 194 | | 3 | | | Tuta | 27 Desited 4 |

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





SARNavigator Workflow



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 ≥ 6 (1848 actives)



ightarrow

All compounds used

H23 - Actives Partitioned



Including the inactives

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



Before Actives Only After Actives & Inactives colour - %active cmpds in cluste



Summary

• SARNavigator

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

