The Rockefeller University Compound Library

J. Fraser Glickman
Rockefeller University
High Throughput and Spectroscopy Resource Center
April 4th , 2014

Resources Available

Technologies

- Microplate assay and drug discovery technologies
- Automated Liquid Handling (pipetting and automated processes)
- HTS formatted Compound library (200,000) curated and annotated
- 14,000 Human ORF's for GOF studies
- ChemInformatics and lead optimization through inventory
- Spectroscopic Techniques for measuring structure and molecular Interactions
- Cytotoxicity assays
- Compound acquisition databases

Guidance/Training

- Drug /Tool Compound Discovery and Development
- Assay Development and optimization
- Screening Strategies
- Increasing Experimental Throughput

Examples of Research

- Potent and selective inhibitors of sonic hedgehog signaling to treat pancreatic cancer
- modulators of Aβ induced pathology for treating Alzheimer's disease
- Novel screening assays for identifying pro-and anti-apoptotic compounds for treatment of cancer
- Novel radiosensitizers for treatment of glioma
- Modulators of bacterial quorum sensing
- Novel natural products encoded on eDNA
- Novel rescue/protective approaches for heart attack
- Ultra high throughput methods for sampling autoantibodies in a patient populations.

Assay Technologies

- Fluorescence Polarization
- FRET (and Time-Resolved)
- AlphaLISA (LOCI)
- Scintillation Proximity
- Reporter Gene Assays
- High content Screening (microscopy and laser scanning)
- Bioluminescence resonance energy transfer
- In Cell Western/ELISA
- Mass Spectrometry
- Protein Complementation Assay/enzyme fragment complentation

- Enzymes
- Ion channels
- Receptors
- Protein-protein, protein-nucleic acid interactions
- Signaling pathways
- Cellular structure
- worms/flies/fish
- Small-molecule-protein interactions
- Cell survival/proliferation

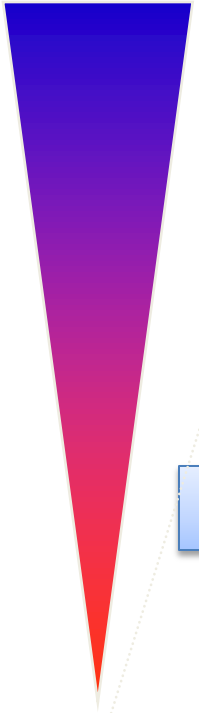
Biophysical/Spectroscopic Techniques

- Surface Plasmon Resonance
- Circular Dichroism
- Isothermal Titration Calorimetry
- Microscale Thermophoresis
- Nuclear Magnetic Resonance

- Small molecule-receptor interactions
- Protein-protein/nucleic acid interactions
- Protein structure (NYSBC)
- Enzyme function
- Small molecule structure

Low Selectivity
Potency
High Numbers

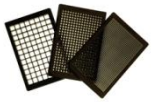
1 assay



High Selectivity
potency
Low Numbers
compounds
Several assays

Strategy: assay development
conversion to 384, 5-80 uL

HTS



Collaborative Drug Discovery

primary hit list

cherry picking

2‘/orthogonalassay

validation

Concentration Response Experiments

HPLC-MS Validation

Prelim. SAR by Inventory

in silico applications

Data normalized to intra-plate high and low controls

Percent inhibition reasonable, number of hits reasonable, quality of hits reasonable

Informatics: Prior to 2008

- Microsoft Excel
- “Cambridgesoft” Database
- In House Server
- Difficult to normalize data
- Server costs were very high
- processing speed low
- High software expertise needed to analyze data
- Could not easily share data
- library “unannotated”
- Users discouraged from using database

CDD and HTSRC

- Avoid need for in-house informatics expert
- No server maintenance
- Easy and Intuitive Data Uploads and downloads
- Annotation
- Cost effective
- flexible: willing to modify and expand utility
- We use *Pipeline Pilot* for compound purchasing and calculations
- We use *Dotmatics Vortex* for Publication graphics and some off-line analysis

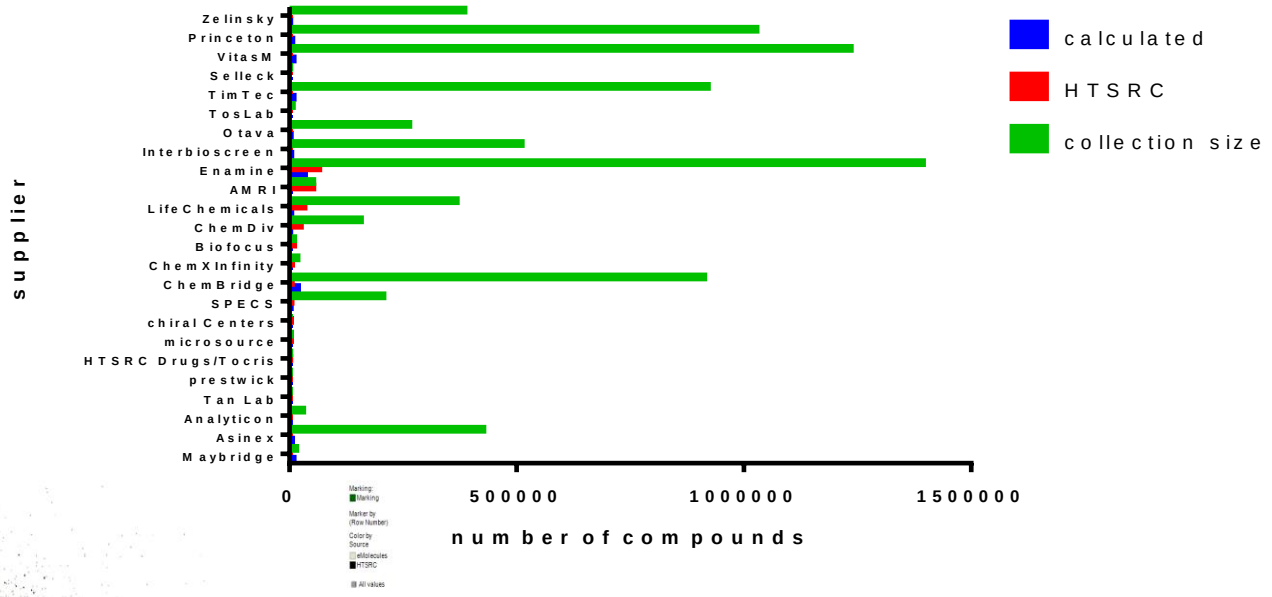
Compound Purchase History

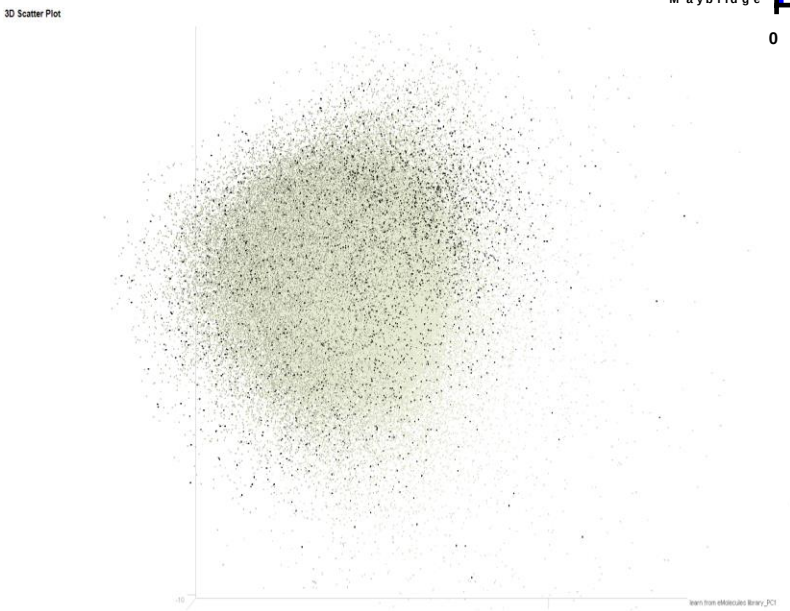
- Up to 2008:“Ready-made” libraries (about 60,000): AMRI, ChemDiv, Cerep, known drugs
- 2008-present: Custom cherry-picked (about 150,000) using Pipeline Pilot (from vendor .sd files)

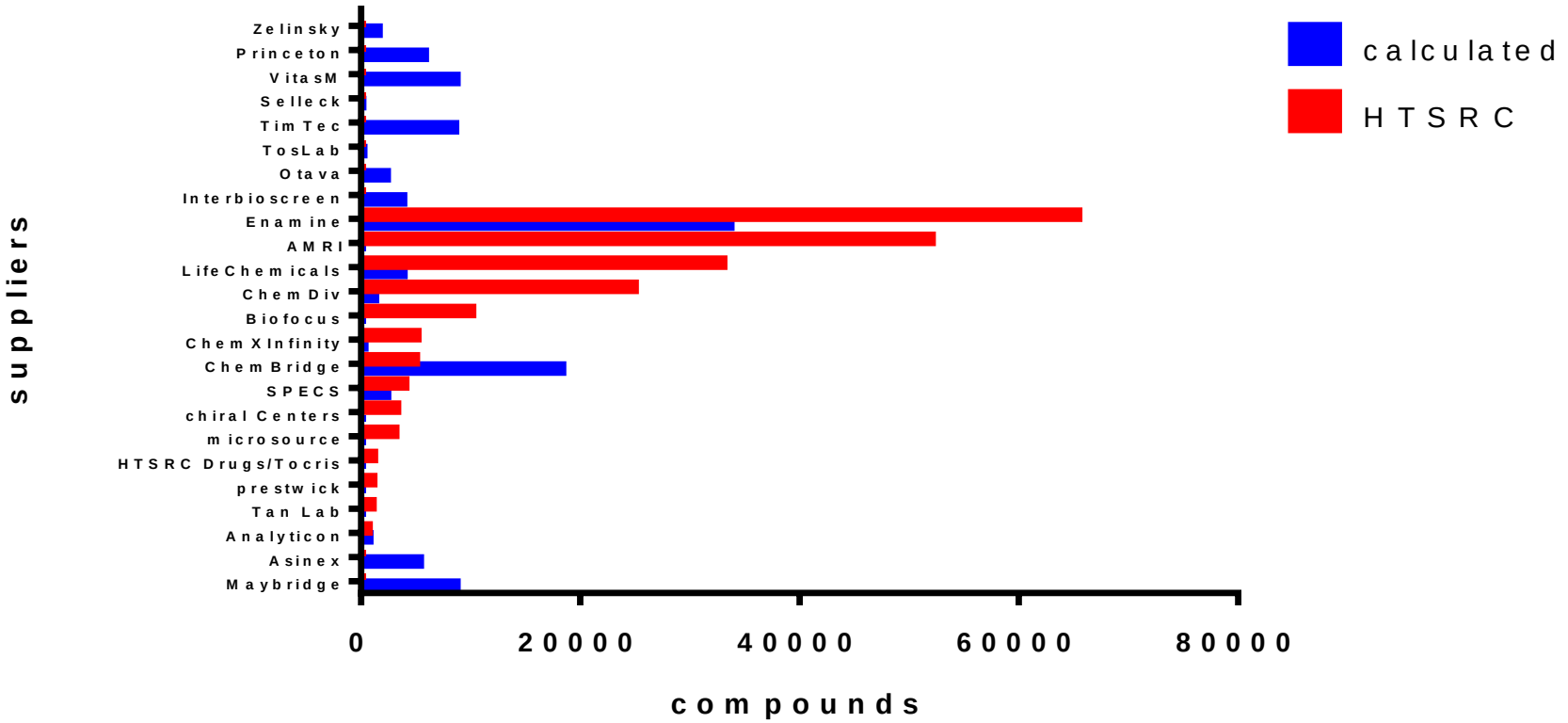
HTSRC diversity approach

- Compile DB of 17 vendors, 4.4 million unique vendor compounds, ready made, re-supply with vendor ID's
- Filter based on QED score (Bickerton et. al. Nat. Chem 2012)"Quantifying the chemical beauty of drugs"
- Choose "affordable" number of diverse compounds based on ECFP's (Accelrys Pipeline Pilot)
- Bin among vendors, approach large numbers first
- Hit "analoging" from dB

Diversity of Rockefeller's Screening Compounds






suppliers
Zelinsky
Princeton
VitasM
Selleck
TimTec
TosLab
Otava
Interbioscreen
Enamine
AMRI
LifeChemicals
ChemDiv
Biofocus
ChemXInfinity
ChemBridge
SPECS
chiral Centers
microsource
HTSRC Drugs/Tocris
prestwick
Tan Lab
Analyticon
Asinex
Maybridge
0
20000
40000
60000
80000
compounds
calculated
HTSRC

PAINS" Pan-Assay Interference Compounds
Baell et. al. J. Med. Chem. 2009 (cited 235 times)

- 6 alphascreen assays
- 93,000 compounds
- Defined hits based on individual assay cutoff criteria
- 4689 PAINS found hitting 33% of the time
- Structural themes

HTSRC Frequent hitters

- 7-16 assays depending on sub-library
- large variety of formats (ion channel, protein-protein ,enzymes, etc)
- 169,000 compounds
- Defined hits based on statistically significant % inhibition in primary screen normalized to plate based internal controls
- Defined frequent hitter as 3 or more times (20% of screens) for follow-up by completely unrelated targets (projects)
- 217 compounds found, none hit more than 4 times, no structural themes
- 36 were "PAINS" as reported by Baell et. al.

Frequent Hitter Methods

- CDD Database with custom scripts: Search for all compounds with DRC data
- Choose Those appear in 3 or greater projects
- Draw all problem structures from Baell publication, saving as .mol file using Marvin Sketch
- Using pipeline pilot, run batch sub-structure searches against HTSRC library compound database
- Generate list of PAINS structures from HTSRC library, with substructure name.

Frequent hitters and PAINS

Library Name	# compounds	frequent hitter freq.	predicted "PAINS"	% True FH in library	% predicted PAINS	comment
Pharmakon-900	905	26	85	2.87	9.39	drugs
Spectrum	2000	56	149	2.80	7.45	drugs and NP's
Prestwick	1110	26	110	2.34	9.91	drugs
LOPAC	1280	25	117	1.95	9.14	drugs
Greenpharma	240	3	9	1.25	3.75	NP's
Enamine	35466	33	937	0.09	2.64	
ChemDiv	21976	11	1306	0.05	5.94	colorful
Chembridge	11638	2	246	0.02	2.11	
LifeChem	30243	1	750	0.00	2.48	
AMRI	50000	0	762	0.00	1.52	
Analyticon	700	0	5	0.00	0.71	semi-synthetic NP's
Biofocus	10150	0	7	0.00	0.07	
Cerep	4000	0	29	0.00	0.73	
TOTAL	**169708**	**183**	**4512**	**0.11**	**2.66**	**38/183 predicted by Baell**

New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays

Jonathan B. Baell*[,†,‡] and Georgina A. Holloway[†,‡]

[†] *The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3052, Australia and* [‡] *Cancer Therapeutics-CRC P/L, 4 Research Avenue, La Trobe R&D Park, Bundoora, Victoria 3086, Australia*

problematic structure	Freq. of this group	# found in Frequent hitters	% of total library	comment
4-hydroxy-phenyl-hydrazone	1141	5	0.66%	
1,2,3-aralkylpyrroles	852		0.49%	
alkylidene1	585	1	0.34%	
alkyl halides	475		0.27%	
phenolicMannichBase	452	2	0.26%	
rhodanines2	432	1	0.25%	
alkylidene barbiturate1	282		0.16%	
catechols	268	17	0.15%	drugs
azo	242		0.14%	
alkylidene barbiturate2	225		0.13%	
alkylidene5	224		0.13%	
divinylketone	199	5	0.11%	
alkylidene3	192		0.11%	
BETA LACTAM CORE	161	1	0.09%	drugs
rhodanines	135		0.08%	
Aldehyde	122	4	0.07%	
Disulfide	41	1	0.02%	
cyano-pyridones	39		0.02%	
fusedTetraHydroQuinolines	33		0.02%	
1,2,3-aralkylpyrroles2	30		0.02%	
alkylidene4	16		0.01%	
2-amino-3-carbonyl-thiophenes2	12		0.01%	
benzofurazans	9		0.01%	
Peroxide	8		0.00%	
quinones	7		0.00%	drugs
Anhydride	5	1	0.00%	
cyano-pyridones2	4		0.00%	

Conclusions

- “PAINS” predictions had little overlap with HTSRC frequent hitters
- Predictions would have led to removal of important lead compounds
- Frequent hitter rates are very low and not a cause for concern in HTSRC library
- Known drug collections and natural product libraries have the highest rate of PAINS compounds **and** HTSRC frequent hitters
- No apparent structural themes in HTSRC frequent hitters

Acknowledgments

- **Carolina Adura, Ph.D.** Research Support Specialist
- **Antonio "TJ" Luz, B.S**. Research Support Assistant
- **Jeanne Chiaravalli-Giganti**,**M.S.** Research Support Associate
- **Brendan Smith**, Medical Student, UMDNJ
- **Andrew Lane**, M.S., Research Support Assistant
- **Krishna Dole,** CDD
- **Anna Spector,** CDD