Chemical Genomics 101

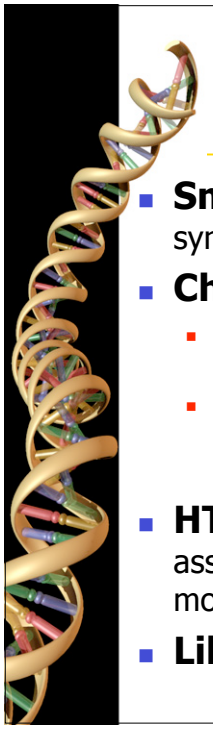
Christopher Austin, M.D.
James Inglese, Ph.D.

NIH Chemical Genomics Center
National Human Genome Research Institute
National Institutes of Health

National Toxicology Program
April 20, 2006

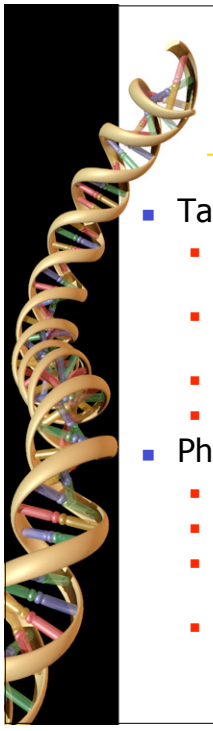
Course Outline

9:30 - 11:30 am	Introduction to Chemical Genomics • What is "Chemical Genomics"? • What is the mission of the NCGC, and the Molecular Libraries Roadmap? • How does the NCGC do screening?	Dr. Austin
	What is meant by "assay" in HTS parlance • What kinds of assays can be run via HTS generally? • What kinds of assays can be run at the NCGC?	Dr. Inglese
11:30 - 12:30 pm	*Lunch*	
12:30 - 1:30 pm	Nuts and bolts of how to convert a lab-based assay to HTS format	Dr. Inglese
	Caveats in interpreting HTS data	Dr. Inglese
	Profiling assays at the NCGC	Dr. Austin
	The tox profiling collaboration with NIEHS/NTP/EPA	Dr. Austin



Definitions

- **Small molecules** a diverse group of natural and synthetic substances having MW 200-1000
- **Chemical Genomics**
  - use of small molecules to understand gene, pathway, cell functions
  - general principles governing
    - interaction of small molecules with their targets
    - biological activity profiles of small molecule compounds
- **HTS** High Throughput Screening; the test of an assay in miniaturized format for activity of small molecules in a library.
- **Library** a collection of 10^2 to 10^7 small molecules

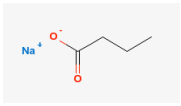


Screens can be target-based or phenotype-based

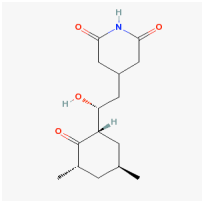
- Target-based
  - "target" refers to what you are trying to affect with the small molecule – usually a protein
  - may use purified protein or engineered cells – or even native cells if target expressed in very high quantity
  - *binding* assay: e.g., ligand displacement
  - *functional* assay: usually artificial reporter (e.g., luciferase)
- Phenotype-based
  - cells or whole organisms
  - knowledge of molecular target not required
  - phenotype being assayed should be close proxy for disease or process under study; ideally is *causally* related
  - readout is an image of varying degree of complexity

"Small" molecules are generally >1000 MW

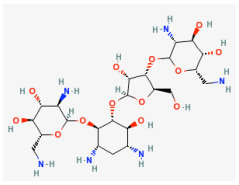
Lithium chloride
MW 42



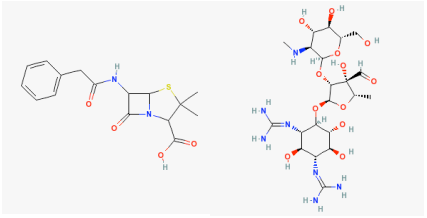
Butyrate
MW 110



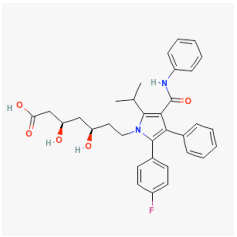
Cycloheximide
MW 281



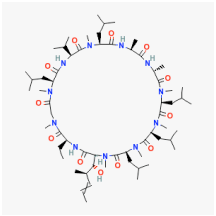
Neomycin
MW 713



Penicillin/streptomycin
MWs 334/582

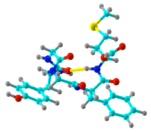


Lipitor
MW 559

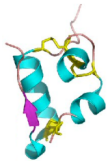


Cyclosporine
MW 1203

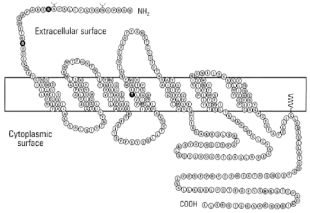
Small molecules are "small" compared to proteins



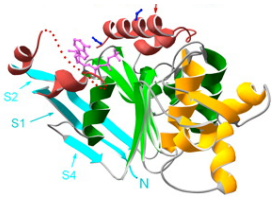
Met-enkephalin
(Tyr-Gly-Gly-Phe-Met)
MW 573



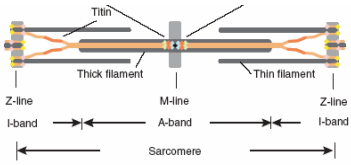
Insulin
MW 2,384



β2 adrenergic receptor
MW 46,426

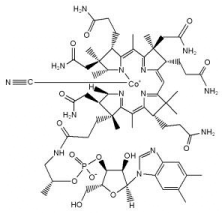


CFTR
MW 167,986

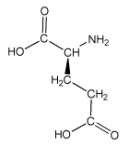


Titin
MW 2,992,867

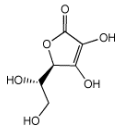
Small molecules can be naturally occurring signalling molecules…



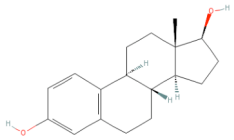
Vitamin B12
MW 1355



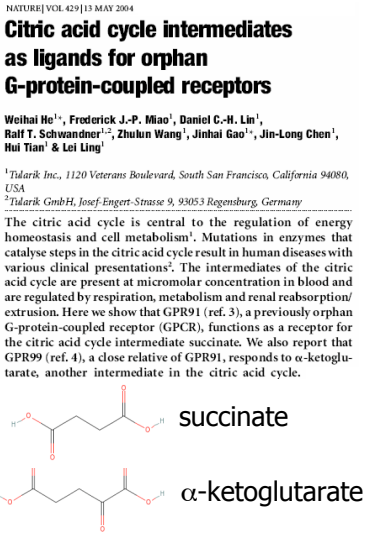
Glutamate
MW 147



Vitamin C
MW 176



Estradiol
MW 272



NATURE | VOL 429 | 13 MAY 2004

Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors

Weihai He[1]*, Frederick J.-P. Miao[1], Daniel C.-H. Lin[1], Ralf T. Schwandner[1,2], Zhulun Wang[1], Jinhai Gao[1]*, Jin-Long Chen[1], Hui Tian[1] & Lei Ling[1]

[1]*Tularik Inc., 1120 Veterans Boulevard, South San Francisco, California 94080, USA*
[2]*Tularik GmbH, Josef-Engert-Strasse 9, 93053 Regensburg, Germany*

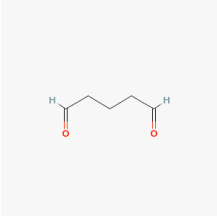
The citric acid cycle is central to the regulation of energy homeostasis and cell metabolism[1]. Mutations in enzymes that catalyse steps in the citric acid cycle result in human diseases with various clinical presentations[2]. The intermediates of the citric acid cycle are present at micromolar concentration in blood and are regulated by respiration, metabolism and renal reabsorption/extrusion. Here we show that GPR91 (ref. 3), a previously orphan G-protein-coupled receptor (GPCR), functions as a receptor for the citric acid cycle intermediate succinate. We also report that GPR99 (ref. 4), a close relative of GPR91, responds to α-ketoglutarate, another intermediate in the citric acid cycle.

succinate

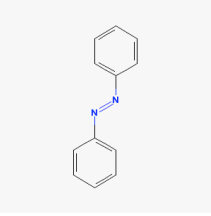
α-ketoglutarate

(when these come from non-mammals, they are called "natural products", studied in "metabonomics")

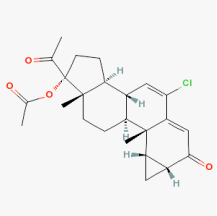
…but they may also be toxins



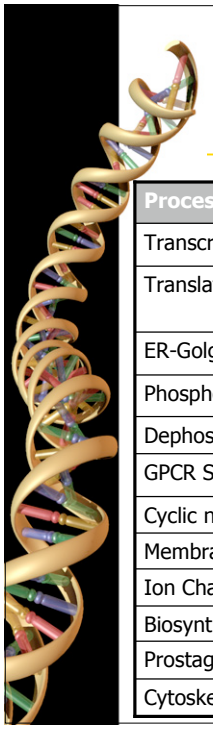
Glutaraldehyde
MW 100



Azobenzene
MW 182



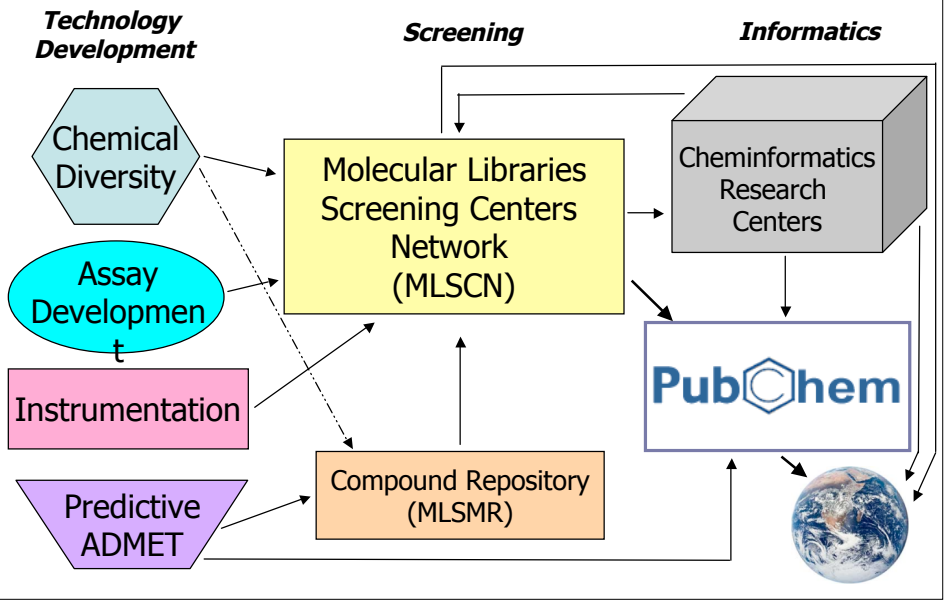
Cyproterone acetate
MW 417



Small molecules have an established record of affecting gene/cell functions

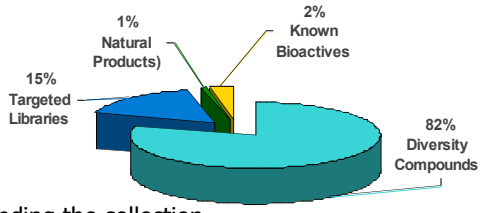
Process	Target	Small molecule
Transcription	PPARγ	troglitazone
Translation	30S bacterial ribosomal subunit Peptide chain elongation	G418 cycloheximide
ER-Golgi trafficking	ND	brefeldin A
Phosphorylation	Bcr-Abl; GSK3β	*imatinib; Li⁺*
Dephosphorylation	Calcineurin (PP2B); IMPase	*cyclosporine; Li⁺*
GPCR Signaling	β-adrenergic receptor	*isoproterenol/propranolol*
Cyclic nucleotide sig.	cAMP, cGMP	forskolin, *silfenadil*
Membrane transport	Serotonin, Glycine transporter	*fluoxetine*, NFPS
Ion Channel flux	GABA-R; L-type Ca^{2+} channel	*diazepam; nifedipine*
Biosynthesis	HMGCoA Reductase	*lovastatin*
Prostaglandin synth.	Cyclooxygenase-2	rofecoxib
Cytoskeleton	Microtubule subunits	*paclitaxel*, cytochalasin B

The Molecular Libraries Roadmap: An Integrated Initiative



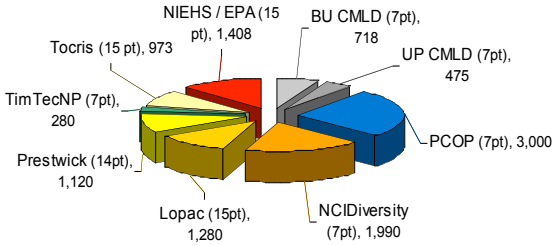
Molecular Libraries Compound Collection

- Housed at Discovery Partners International
- Initial set of ~67,000 compounds purchased from commercial vendors
  - Chosen by external advisors + DPI + NIH
  - >90% purity, >10mg, ±RO5, solubility >20ug/ml, all QCed

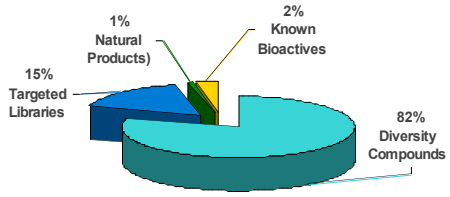


- Expanding the collection
  - Purchase of next 100,000 ongoing; 500,000 at maturity
    - Less stringent property requirements, filling out SAR clusters of 3-5
  - Molecular Libraries Roadmap Chemical Diversity initiatives
    - *Pilot scale libraries for HTS*
  - Centers for Methodology in Library Development
    - Boston U., Harvard, Pitt, U. Kansas
  - Solicitation of compounds from academia, biotech, pharma

NCGC Compound Libraries

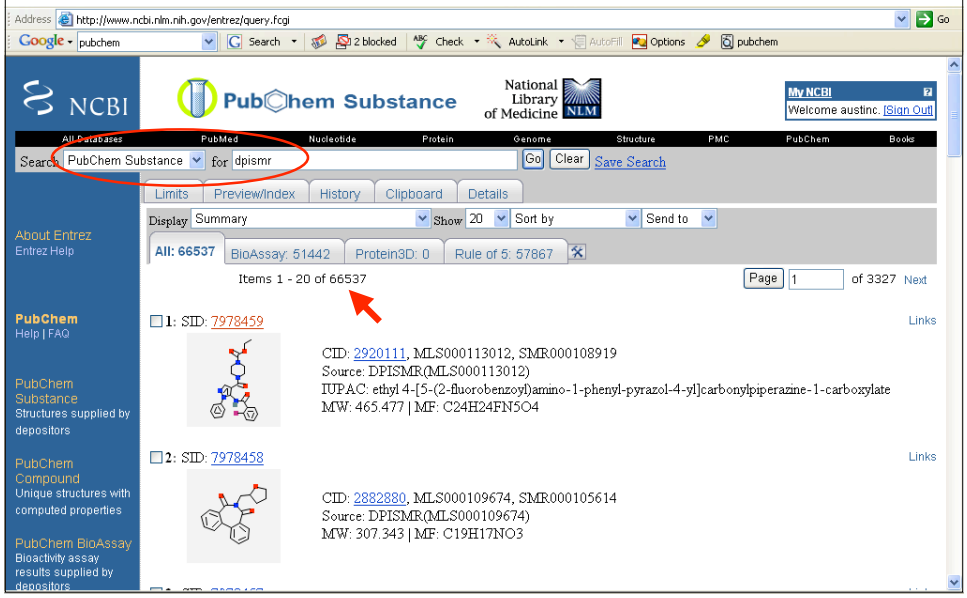


NCGC exploratory library
- ~11,000 cpds
- 10mM – 0.1μM in DMSO
- Primarily biologically active substances



MLSCN SMR
- ~100K cpds
- ~500K samples
- Primarily uncharacterized substances

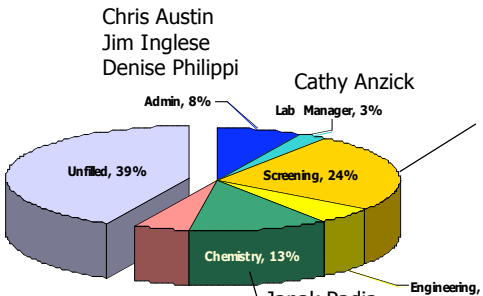
The NIH compound collection is available in PubChem



NCGC's Mission

- Biological activity profiles of chemical compounds
- Chemical probes for currently "non-druggable" targets
- New paradigms for screening and probe development
- "Chemical genomics"
  - General principles of small molecule interactions with biological targets

NCGC Divisions and Personnel



Chris Austin
Jim Inglese
Denise Philippi

Cathy Anzick

Doug Auld
Wei Zheng,
Ron Johnson
Anton Simeonov
Menghang Xia
Ya-Qin Zhang
Adam Yasgar
Pingjun Zhu
Henrike Veith
Steve Titus

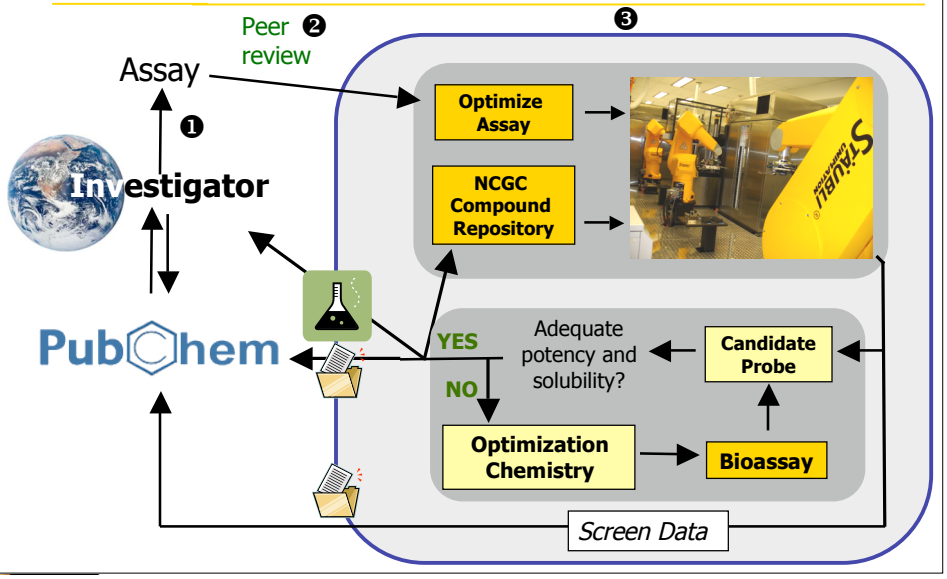
Ajit Jadhav
Yuhong Wang
Noel Southhall

Janak Padia
Michael Nelson
Teresa Phillips
Jeimin Lu
Karina Zuck

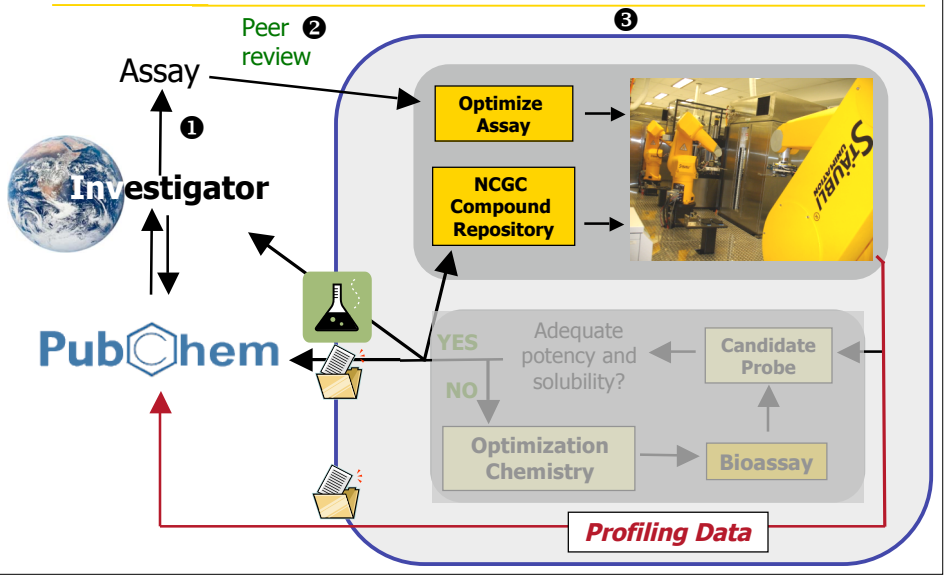
Sam Michael
Carleen Klumpp

- Recruited principally from Pharma and Biotech
  - **Merck, Pfizer, Pharmacopeia, Caliper, Human Genome Sciences, Amphora, Amgen, Avalon Pharma., Targacept, Pall Corp., NHGRI, NCI, NIAAA, Naval Network Warfare Command, Northrop Grumman, BD Biosciences, Celera**

NCGC Operation



NCGC Operation



Quantitative HTS (qHTS) Paradigm

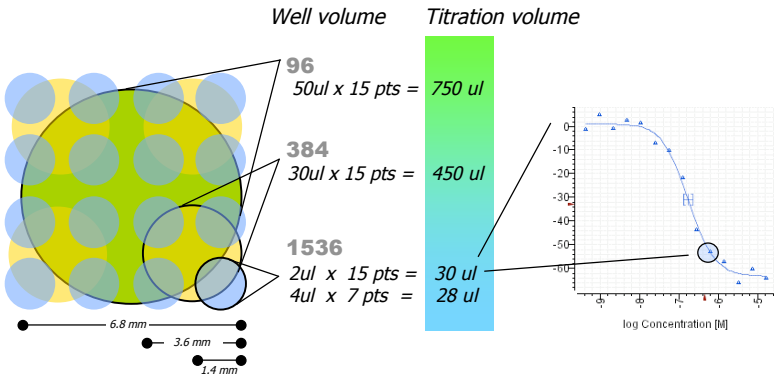
- Conventional HTS done at single [ ]
- But [ ] critical to small molecule effects
  - *"All things are poisons, for there is nothing without poisonous qualities. It is only the dose which make a thing a poison…a lot kills, a little cures."* - **Paracelsus (1493-1521)**

[and less does nothing]

- qHTS tests all compounds in titrations in primary screen
  - HTS as high-throughput pharmacology
  - Rich positive and negative activity data on chemical libraries

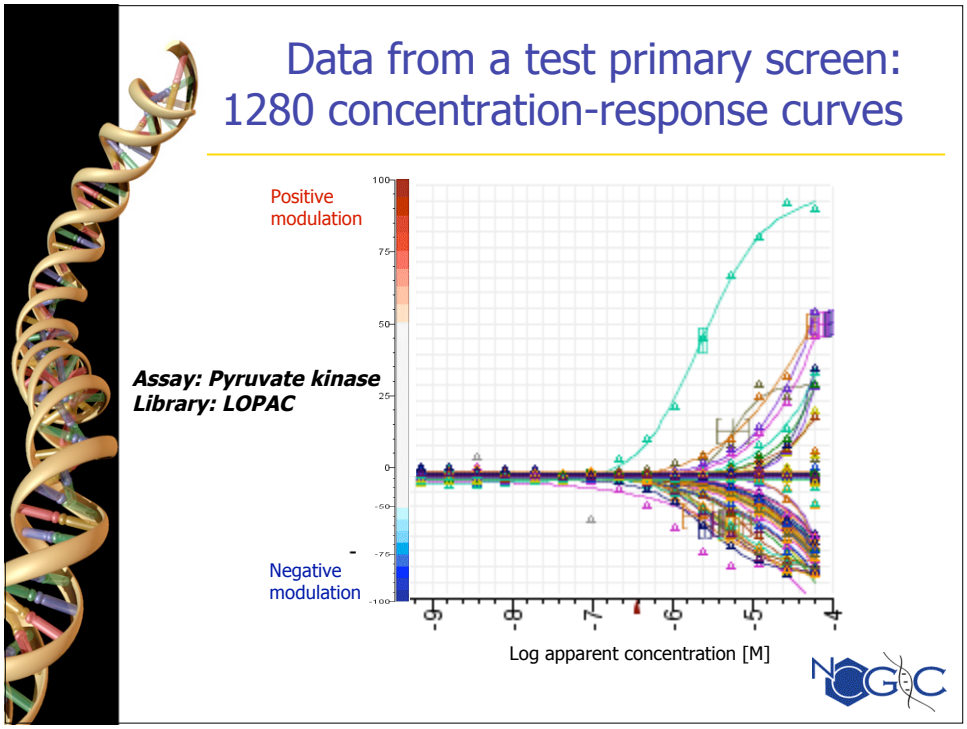
NIH Chemical Genomics Center

Miniaturization and automation enable qHTS

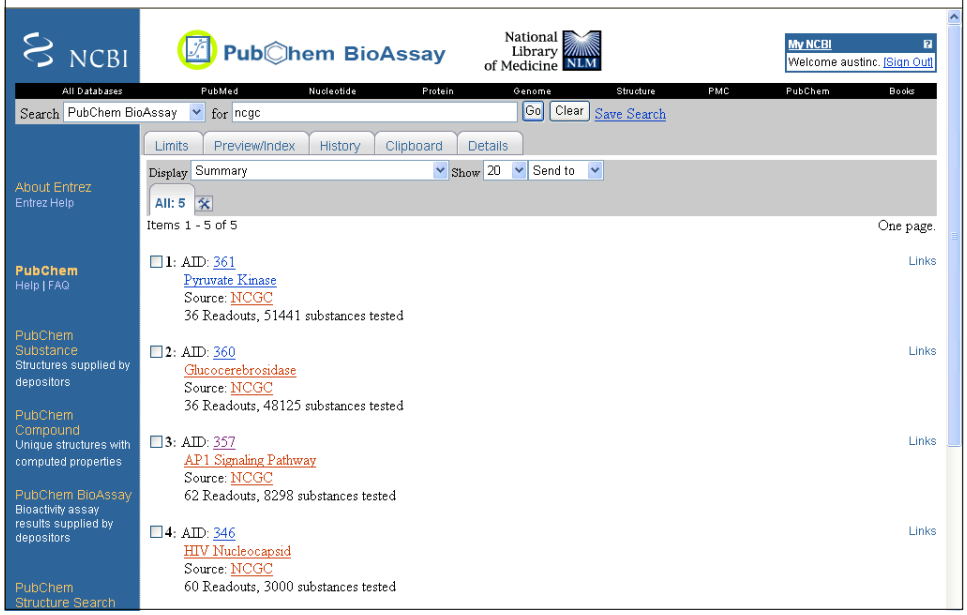


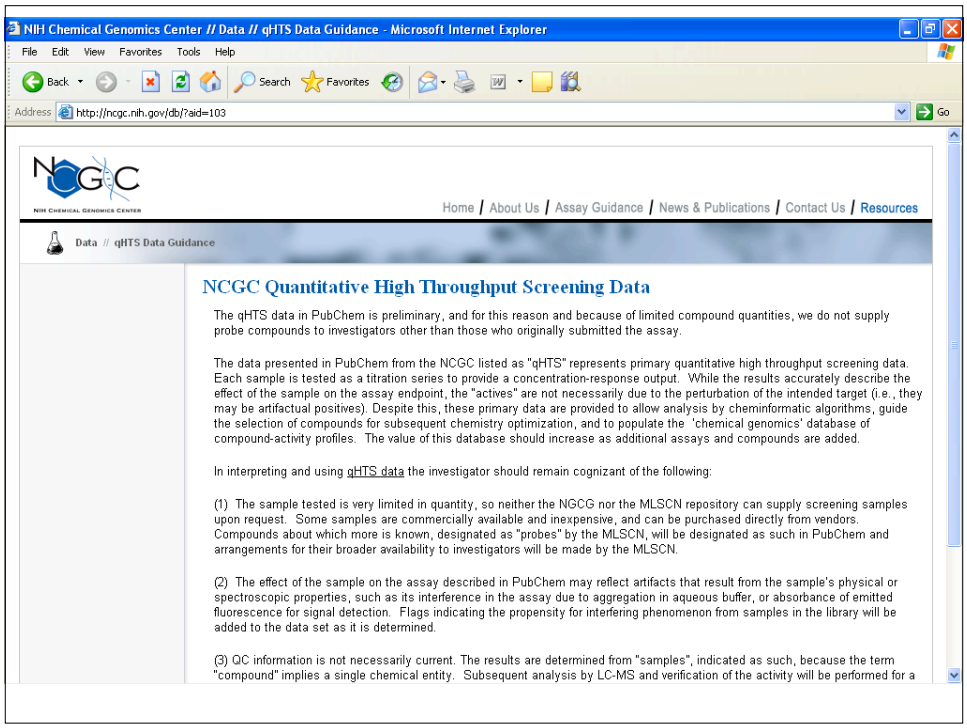
- Enabled by 1536-well microtiter plates and low volume liquid handling
- 'Front-loading' titrations on screening platform (inter-plate titrations)
- Integrate all components with robust automation
- Recruit scientists with experience in assay development and HTS technology

Data from a test primary screen: 1280 concentration-response curves

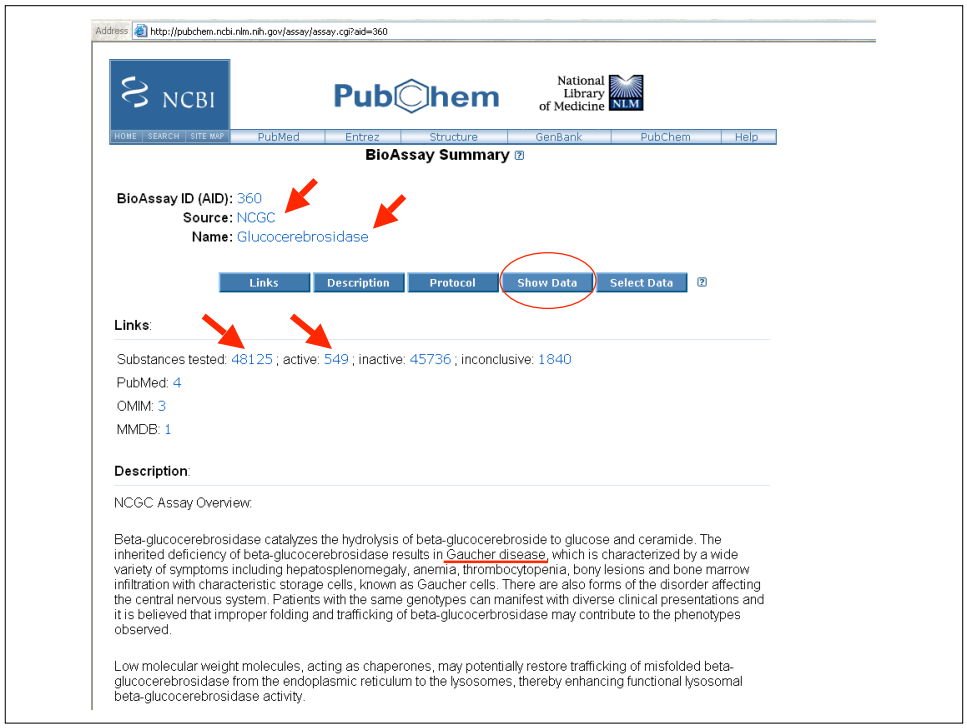


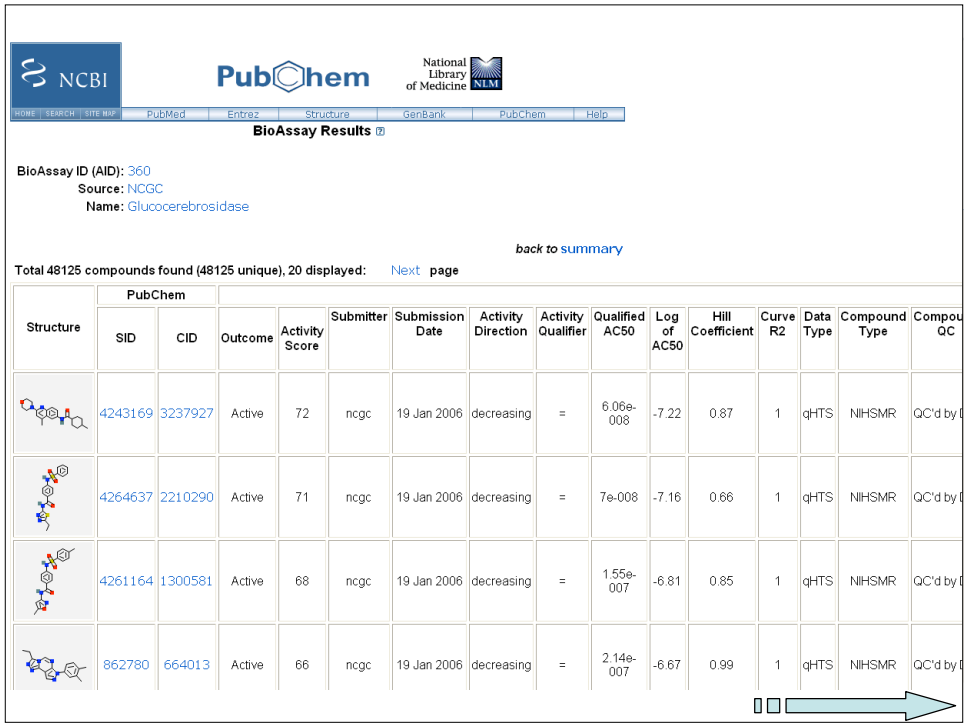
NCGC Data in PubChem




NIH Chemical Genomics Center // Data // qHTS Data Guidance - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://ncgc.nih.gov/db/?aid=103
Go
NCGC
NIH Chemical Genomics Center
Home / About Us / Assay Guidance / News & Publications / Contact Us / Resources
Data // qHTS Data Guidance
NCGC Quantitative High Throughput Screening Data
The qHTS data in PubChem is preliminary, and for this reason and because of limited compound quantities, we do not supply probe compounds to investigators other than those who originally submitted the assay.
The data presented in PubChem from the NCGC listed as "qHTS" represents primary quantitative high throughput screening data. Each sample is tested as a titration series to provide a concentration-response output. While the results accurately describe the effect of the sample on the assay endpoint, the "actives" are not necessarily due to the perturbation of the intended target (i.e., they may be artifactual positives). Despite this, these primary data are provided to allow analysis by cheminformatic algorithms, guide the selection of compounds for subsequent chemistry optimization, and to populate the 'chemical genomics' database of compound-activity profiles. The value of this database should increase as additional assays and compounds are added.
In interpreting and using qHTS data the investigator should remain cognizant of the following:
(1) The sample tested is very limited in quantity, so neither the NGCG nor the MLSCN repository can supply screening samples upon request. Some samples are commercially available and inexpensive, and can be purchased directly from vendors. Compounds about which more is known, designated as "probes" by the MLSCN, will be designated as such in PubChem and arrangements for their broader availability to investigators will be made by the MLSCN.
(2) The effect of the sample on the assay described in PubChem may reflect artifacts that result from the sample's physical or spectroscopic properties, such as its interference in the assay due to aggregation in aqueous buffer, or absorbance of emitted fluorescence for signal detection. Flags indicating the propensity for interfering phenomenon from samples in the library will be added to the data set as it is determined.
(3) QC information is not necessarily current. The results are determined from "samples", indicated as such, because the term "compound" implies a single chemical entity. Subsequent analysis by LC-MS and verification of the activity will be performed for a


NCBI
HOME SEARCH SITE MAP
PubChem
National Library of Medicine
NLM
PubMed Entrez Structure GenBank PubChem Help
PubChem Text Search
PubChem BioAssay
Gaucher
GO
PubChem provides information on the biological activities of small molecules. It is a component of NIH's Molecular Libraries Roadmap Initiative. If you would like to learn more about how to use the PubChem resources, please go to our help page.
New BioAssay data from NCGC (NIH Chemical Genomics Center) is now available in PubChem.
New Structures from NCGC (NIH Chemical Genomics Center) are now available in PubChem.
More PubChem announcements ...
PubChem Compound: Search unique chemical structures using names, synonyms or keywords. Links to available biological property information are provided for each compound.
PubChem Substance: Search deposited chemical substance records using names, synonyms or keywords. Links to biological property information and depositor web sites are provided.
PubChem BioAssay: Search bioassay records using terms from the bioassay description, for example "cancer cell line". Links to active compounds and bioassay results are provided.
Structure Search: Search PubChem's Compound database using a chemical structure as the query. Structures may be sketched or specified by SMILES, MOL files, or other formats.





Structure	PubChem															
	SID	CID	Outcome	Activity Score	Submitter	Submission Date	Activity Direction	Activity Qualifier	Qualified AC50	Log of AC50	Hill Coefficient	Curve R2	Data Type	Compound Type	Compou QC
	4243169	3237927	Active	72	ncgc	19 Jan 2006	decreasing	=	6.06e-008	-7.22	0.87	1	qHTS	NIHSMR	QC'd by [
	4264637	2210290	Active	71	ncgc	19 Jan 2006	decreasing	=	7e-008	-7.16	0.66	1	qHTS	NIHSMR	QC'd by [
	4261164	1300581	Active	68	ncgc	19 Jan 2006	decreasing	=	1.55e-007	-6.81	0.85	1	qHTS	NIHSMR	QC'd by [
	862780	664013	Active	66	ncgc	19 Jan 2006	decreasing	=	2.14e-007	-6.67	0.99	1	qHTS	NIHSMR	QC'd by [

NCBI PubChem National Library of Medicine

HOME | SEARCH | SITE MAP | PubMed | Entrez | Structure | GenBank | PubChem | Help

BioAssay Results

BioAssay ID (AID): 360
Source: NCGC
Name: Glucocerebrosidase

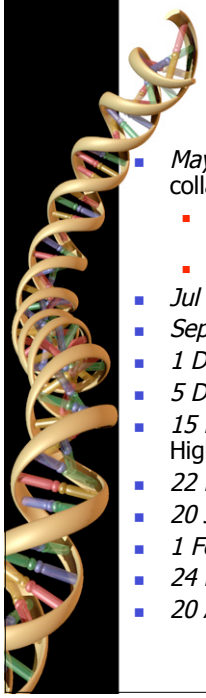
back to summary

Full CRC

Total 48125 compounds found (48125 unique), 20 displayed: Next page

AID: 360

Compound Type	Compound QC	Curve Fit Model	Hill S0	Hill Sinf	Hill dS	Log AC50 Std Error	Curve Chi2f	Excluded Points	Number of Points	Activity at 4.925nM (%)	Activity at 24.623nM (%)	Activity at 0.123uM (%)	Activity at 0.615uM (%)	Activity at 3.077uM (%)	Activity at 15.386uM (%)	Activity at 0.077mM (%)
NIHSMR	QC'd by DPI	4pHill (AC50,n,S0,Sinf)	-1.23	-100.1	98.92	0.02	0.5	{}	7	-11.3	-31.9	-65.8	-88.4	-96.4	-99.6	-100.2
NIHSMR	QC'd by DPI	4pHill (AC50,n,S0,Sinf)	11.41	-107.2	118.6	0.1	2.04	{}	7	-5.5	-30.2	-56.9	-84.4	-100	-103.7	-105
NIHSMR	QC'd by DPI	4pHill (AC50,n,S0,Sinf)	2.97	-100.6	103.5	0.04	1.4	{}	7	-2.4	-14.2	-44.7	-74.8	-94.6	-98.5	-99.4
NIHSMR	QC'd by DPI	4pHill (AC50,n,S0,Sinf)	-1.83	-95.57	93.74	0.04	1.8	{}	7	-2.9	-13.4	-35.5	-70.9	-90.7	-94.9	-93.7



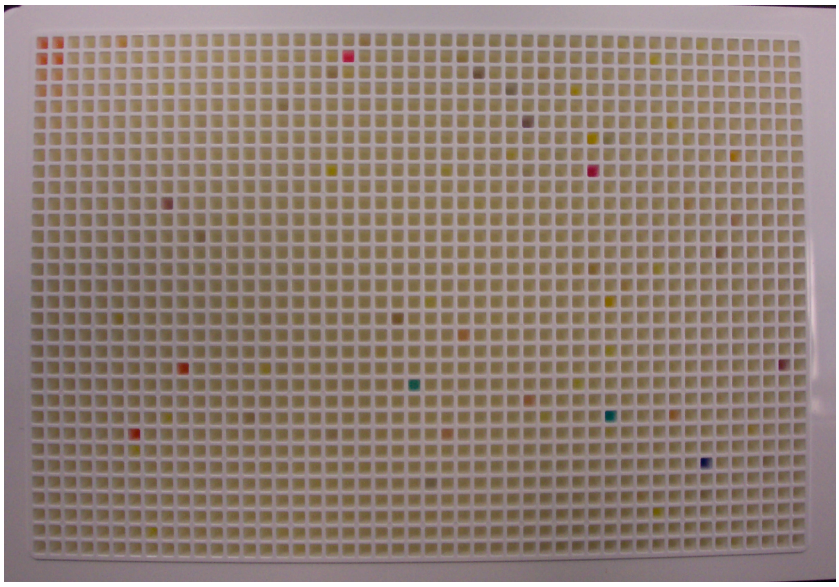
NTP-NCGC Collaboration

- *May 2005:* David Schwartz, Chris Portier initiatiate NTP-NCGC collaboration
  - Getting collection of NTP identified toxic compounds into NCGC screening collection
  - Getting assays of interest to NTP to NCGC to screen
- *Jul 2005:* First tox compounds received at NCGC
- *Sept 2005:* First tox assays received from NIEHS
- *1 Dec 2005:* Chemical Prioritization Community of Practice initiated
- *5 Dec 2005:* CA gives talk at NIEHS HTS Faculty seminar
- *15 Dec 2005:* Austin, Inglese, Jadhav help organize, and talk at, NTP High Throughput Screening Assays Workshop
- *22 Dec 2005:* Final tox compounds received at NCGC to complete 1408
- *20 Jan 2006:* Tox compound set 1408 plated in dilutions at NCGC
- *1 Feb 2006:* First tox assay run on tox 1408
- *24 Mar 2006:* Sixth tox assay run on tox 1408
- *20 Apr 2006:* Chemical Genomics 101 and project update

NCGC NIH Chemical Genomics Center

NIEHS 1408 Compound Set

92 uM concentration



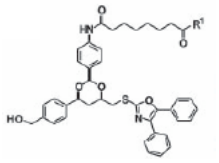
What is a "chemical probe"?

A chemical probe is a member of a chemical series displaying a structure activity relationship (SAR) over minimally one order of magnitude, having aqueous solubility equal to or better than 5 ug/ml in a buffered aqueous solution, and active at concentration of low-micromolar or below in an in vitro (cellular or cell-free) assay. Ideally the probes series will include inactive isomers and another distinct active chemical series.

Note: potency is dependent on context and supporting data. For example, millimolar potency would be acceptable for highly soluble compounds showing efficacy and low cellular toxicity.

Example 1: Distinguishing effects of a multidomain protein

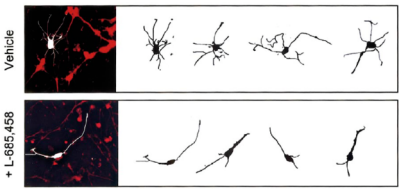
- Histone deacetylases (HDACs) are zinc-dependent hydrolases that mediate chromatin remodeling and gene expression
- small-molecule histone deacetylase inhibitors are of interest for cancer and neurodegenerative diseases.
- "HDAC inhibitors" also affect microtubule acetylation.
- Haggarty et al. carried out a chemical genetic screen of 7,392 small molecules and identified "tubacin".

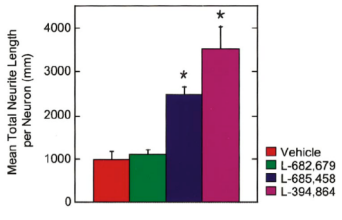


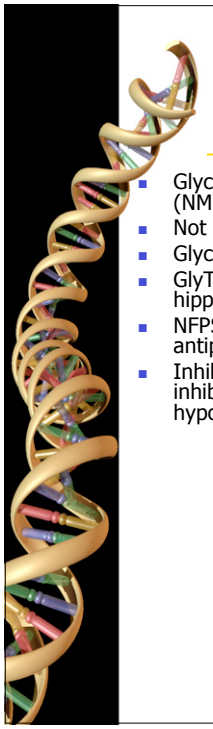
- "Tubacin inhibits tubulin deacetylation but does not affect the histone acetylation, gene-expression patterns, or cell-cycle progression.
- Class II histone deacetylase 6 (HDAC6) was identified as the intracellular target of tubacin.
- Only one of the two catalytic domains of HDAC6 possesses tubulin deacetylase activity, and only this domain is bound by tubacin.
- Tubacin treatment did not affect the stability of microtubules but did decrease cell motility.

Example 2: Identifying pleiotropic effects of a multisubstrate enzyme

- Cleavage of both Amyloid Precursor Protein (APP) and Notch are dependent on presenilin-1 (PS1)
- Notch cleavage may be affected by APP PS1/γ-secretase inhibitors under development for treatment of Alzheimer's disease
- Using chemical inhibitors of PS1/γ-secretase, Figueroa et al demonstrated that intracellular trafficking of Notch in human CNS neurons is altered by inhibition of PS1 and is accompanied by dramatic changes in neurite morphology, c/w inhibition of Notch
- Notch dysregulation may contribute to the neuritic dystrophy characteristic of AD brain; inhibition of γ-secretase/PS1 may have clinically beneficial effects on the neuritic pathology of AD, in addition to its expected effect to reduce amyloid burden.

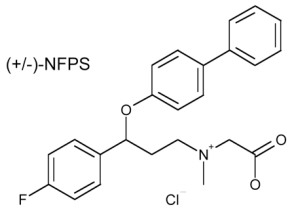


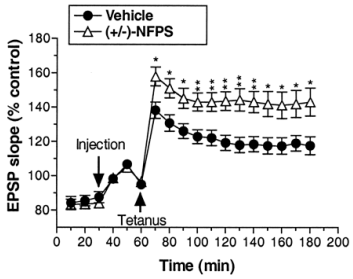




Example 3: Testing a hypothesis based on genetic evidence

- Glycine acts as a necessary coagonist for glutamate at the NMDA receptor (NMDAR) complex
- Not clear if glycine normally saturates NMDAR-containing synapses *in vivo*
- Glycine transporter type 1 (GlyT1) may keep glycine sub-saturated at synapses
- GlyT1 inhibitor NFPS significantly enhanced long-term potentiation in the hippocampal dentate gyrus
- NFPS induced a pattern of c-Fos immunoreactivity comparable with the atypical antipsychotic clozapine and enhanced prepulse inhibition in mice
- Inhibition of GlyT1 can enhance NMDAR-sensitive activity *in vivo* and GlyT1 inhibition may represent a novel target for disorders associated with NMDAR hypofunction such as schizophrenia





Update on Tox Assay qHTS

CellTiter-Glo® Luminescent Cell Viability Assay

- Description
  - Method of measuring number of viable cells in culture
  - Based on quantitation of ATP, an indicator of metabolic activity
  - Luminescent signal proportional to amount of ATP present
- Applications
  - Cell proliferation
  - Cytotoxicity
  - Cell viability
- Six cell lines qHTS for 1408 compounds from NIEHS
  - HepG2 (human hepatocytes, hepatocellular carcinoma)
  - Jurkat (Clone E6-1, human T lymphocytes, T cell leukemia)
  - HEK293 (human embryonic kidney cells, transformed with adenovirus)
  - SK-N-SH (human neuroblastoma)
  - MRC-5 (normal human lung fibroblasts)
  - BJ (normal human foreskin fibroblasts)
- Data analysis
  - Robust Z'
  - Heat maps
  - Complete data analysis pending
- Application of Acea technology for follow up toxicology study in HepG2 cells
  - Tamoxifen
  - Doxorubicin