



The NIH Roadmap: *New Pathways to Discovery*

Empowering Small
Molecule Research

Francis S. Collins, M.D., Ph.D.





New Pathways to Discovery

- ▶ [Building Blocks, Biological Pathways, and Networks](#)
- ▶ [Molecular Libraries and Imaging](#)
- ▶ [Structural Biology](#)
- ▶ [Bioinformatics and Computational Biology](#)
- ▶ [Nanomedicine](#)

Research Teams of the Future

- ▶ [High-Risk Research](#)
– [NIH Director's Pioneer Award](#)
- ▶ [Interdisciplinary Research](#)
- ▶ [Public-Private Partnerships](#)

Re-engineering the Clinical Research Enterprise

- ▶ [Re-engineering the Clinical Research Enterprise](#)

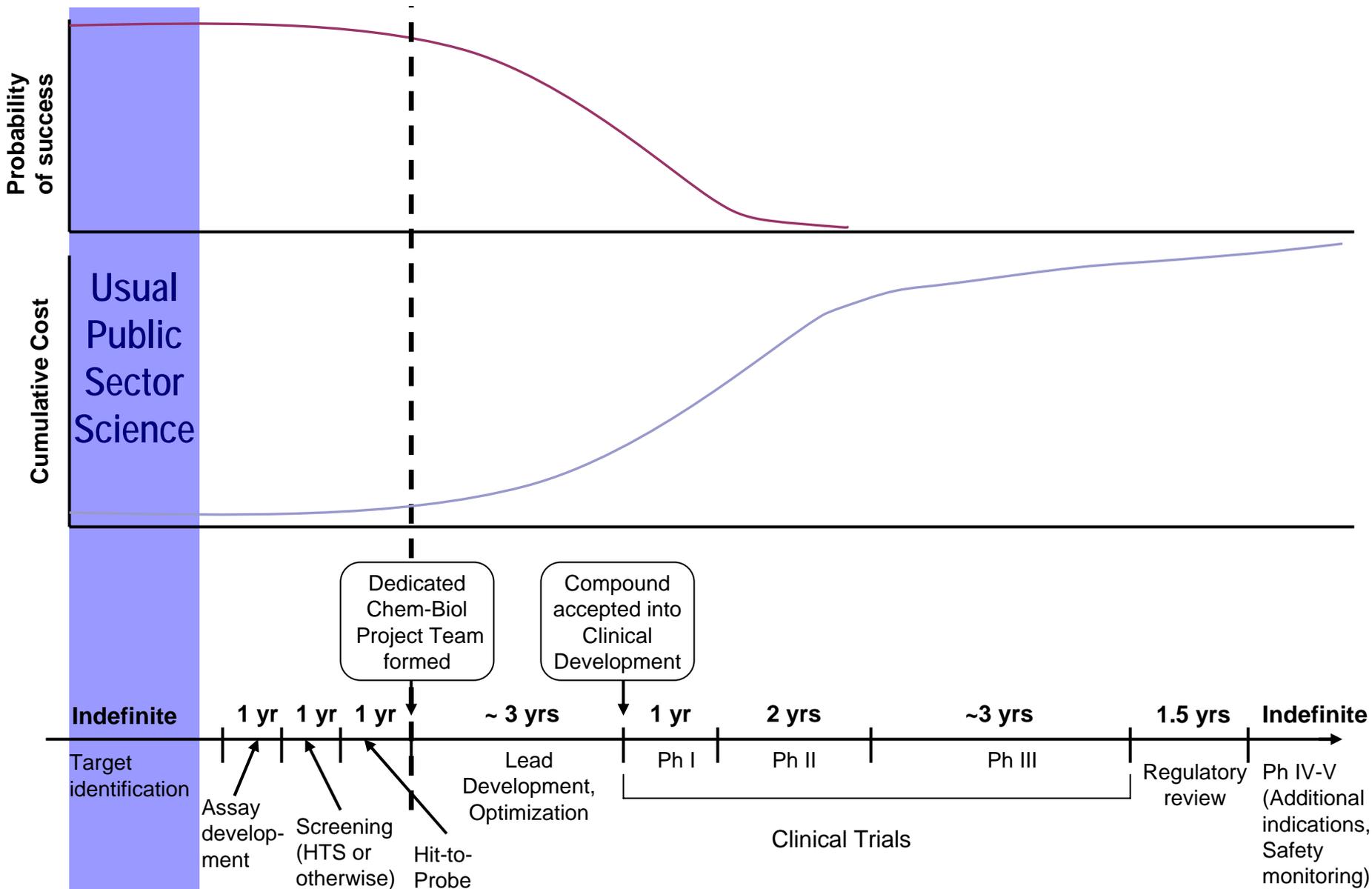
Purpose:

To empower the research community to use small molecule compounds in their research, whether as tools to perturb genes and pathways, or as starting points to the development of new therapeutics for human disease

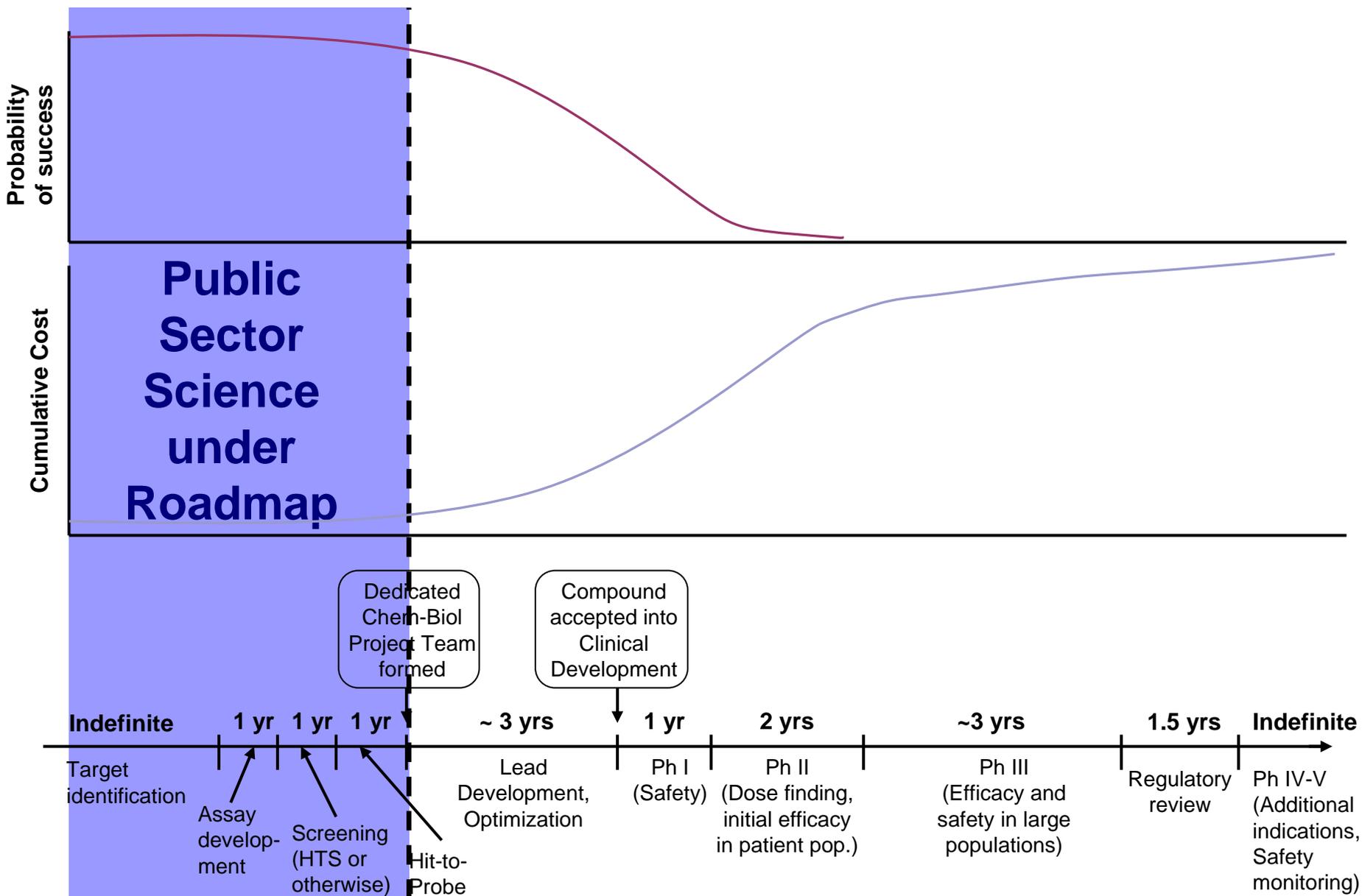
■ <http://nihroadmap.nih.gov/>



How does Molecular Libraries relate to drug development?

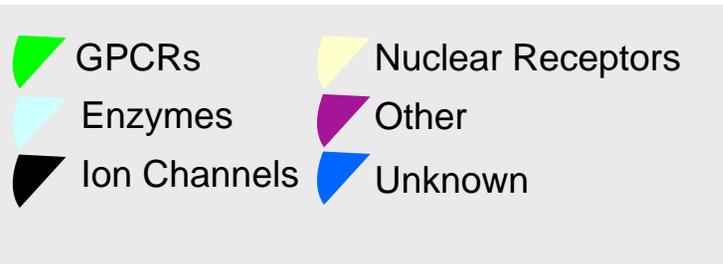
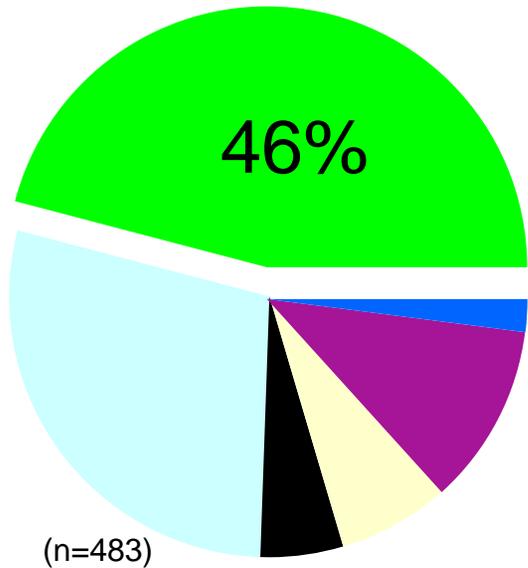


How does Molecular Libraries relate to drug development?

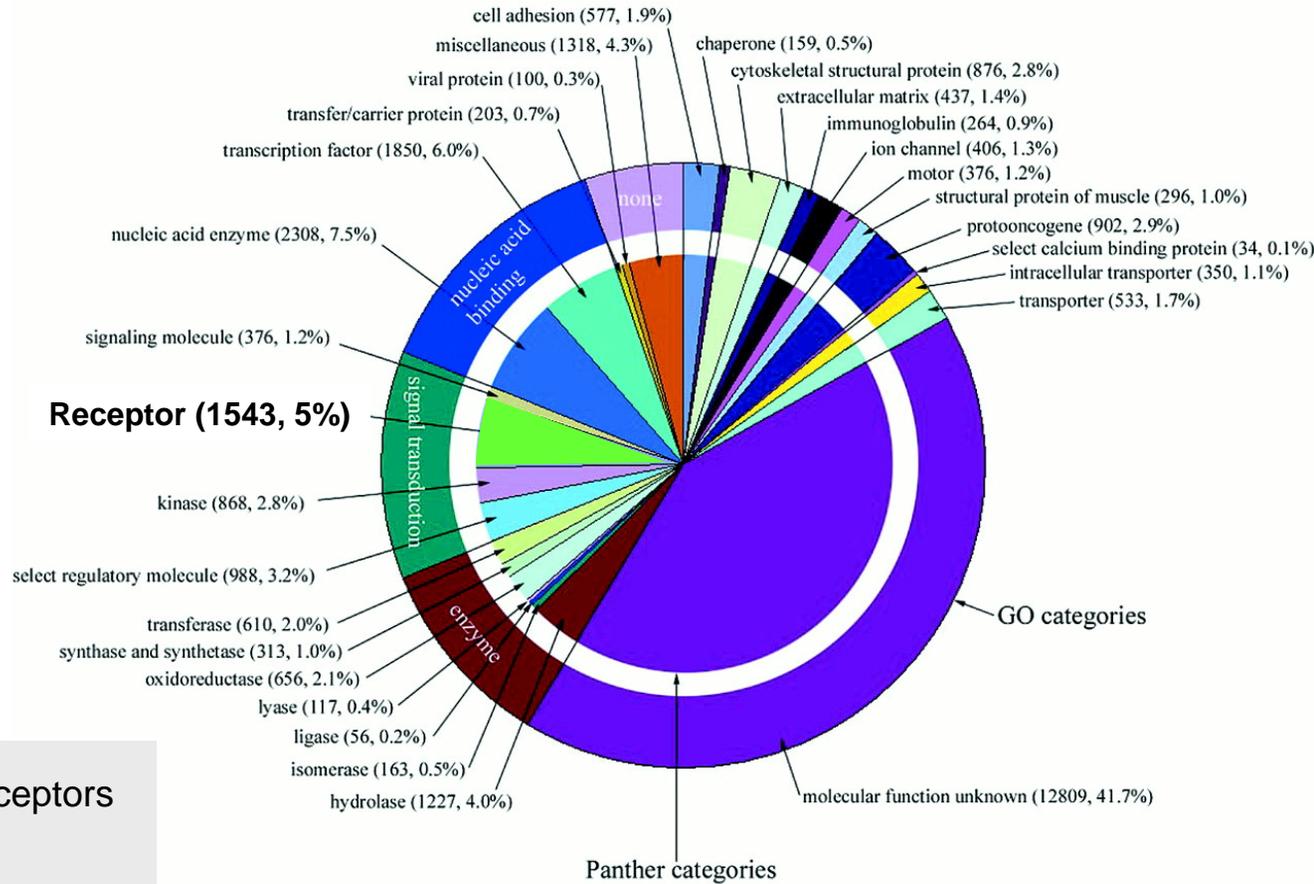


Targets Based on Current Therapies^a and their relationship to the Human Genome^b

Drug Target Classes



Human Genome



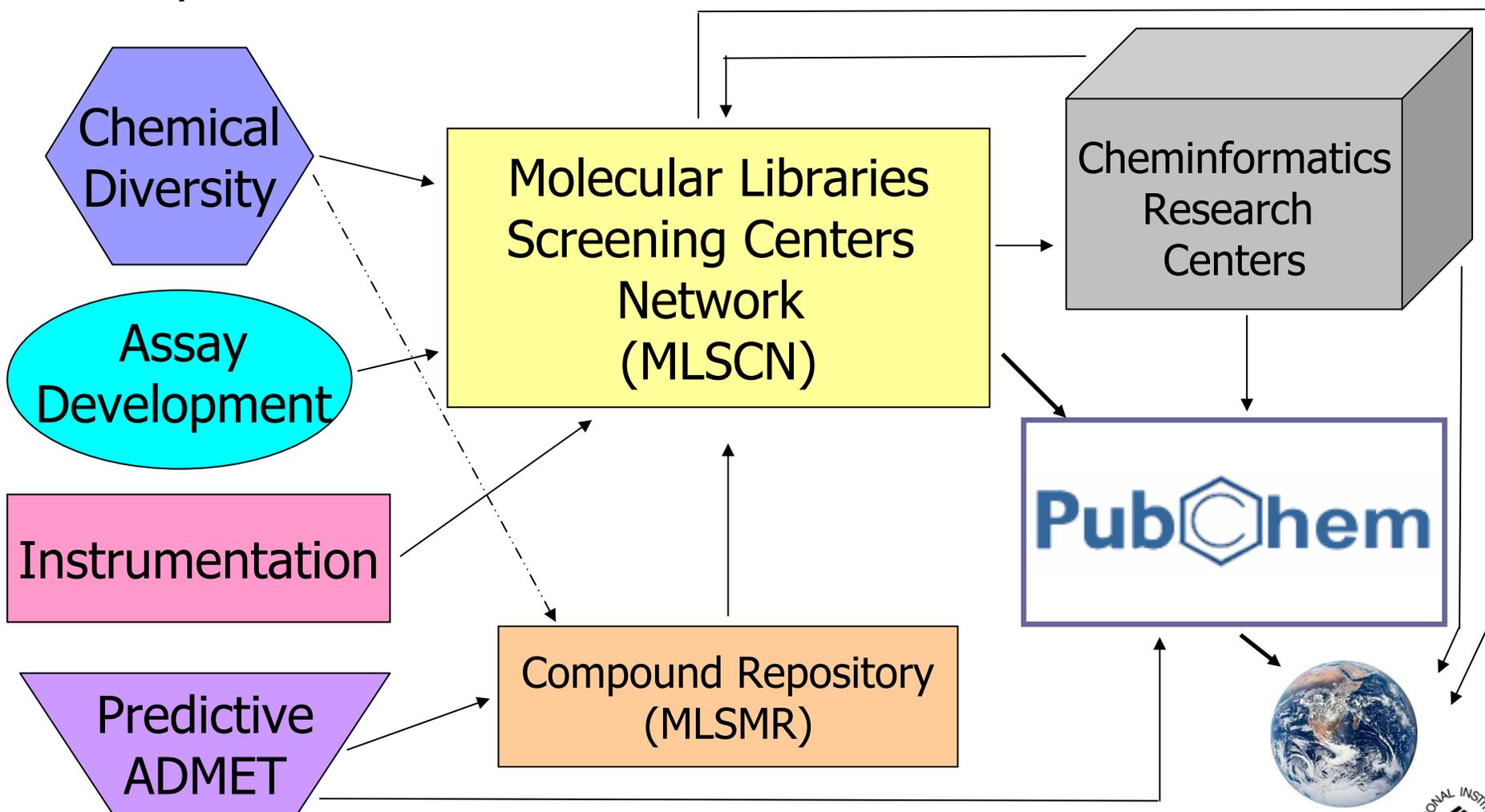
^aScience 287:1962 (2000); ^bScience 291:1304 (2001)

The Molecular Libraries Roadmap: An Integrated Initiative

*Technology
Development*

Data Production

Data Analysis/Dissemination





▶ [Home Page](#)

Molecular Libraries and Imaging

- ▶ [Overview](#)
- ▶ [Implementation Group Members](#)
- ▶ [Funding Opportunities](#)
- ▶ [Funded Research](#)
- ▶ [Related Activities](#)
- ▶ [Meetings](#)
- ▶ [PubChem](#)

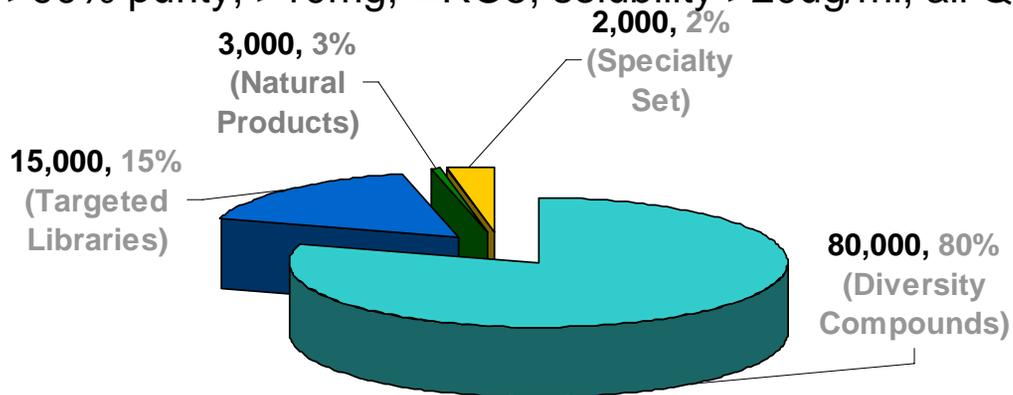


Molecular Libraries Screening Centers Network (MLSCN) RFA-04-017

PI Name	Institution Name	Title
AUSTIN, CHRIS	NHGRI	NIH Chemical Genomics Center (NCGC) <ul style="list-style-type: none"> ■ Additional Information
DIAMOND, SCOTT	UNIVERSITY OF PENNSYLVANIA	The Penn Center for Molecular Discovery <ul style="list-style-type: none"> ■ Additional Information ■ Abstract (from CRISP)
DINGLEDINE, RAYMOND	EMORY UNIVERSITY	Emory Chemistry-Biology Center in the MLSCN <ul style="list-style-type: none"> ■ Abstract (from CRISP)
LAZO, JOHN	UNIVERSITY OF PITTSBURGH AT PITTSBURGH	University of Pittsburgh Molecular Libraries Screening Center <ul style="list-style-type: none"> ■ Additional Information ■ Abstract (from CRISP)
PIAZZA, GARY	SOUTHERN RESEARCH INSTITUTE	Southern Research Molecular Libraries Screening Center <ul style="list-style-type: none"> ■ Abstract (from CRISP)
REED, JOHN	THE BURNHAM INSTITUTE	San Diego Center for Chemical Genomics <ul style="list-style-type: none"> ■ Abstract (from CRISP)
ROSEN, HUGH	THE SCRIPPS RESEARCH INSTITUTE	Scripps Research Institute Molecular Screening Center <ul style="list-style-type: none"> ■ Abstract (from CRISP)
ROTHMAN, JAMES	COLUMBIA UNIVERSITY MEDICAL CENTER	MLSCN Center at Columbia University <ul style="list-style-type: none"> ■ Abstract (from CRISP)
SKLAR, LARRY	UNIVERSITY OF NEW MEXICO ALBUQUERQUE	New Mexico Molecular Libraries Screening Center <ul style="list-style-type: none"> ■ Additional Information ■ Abstract (from CRISP)
WEAVER, C. DAVID	VANDERBILT UNIVERSITY	Vanderbilt Screening Center for GPCRs, Ion Channels, and Transporters <ul style="list-style-type: none"> ■ Additional Information ■ Abstract (from CRISP)

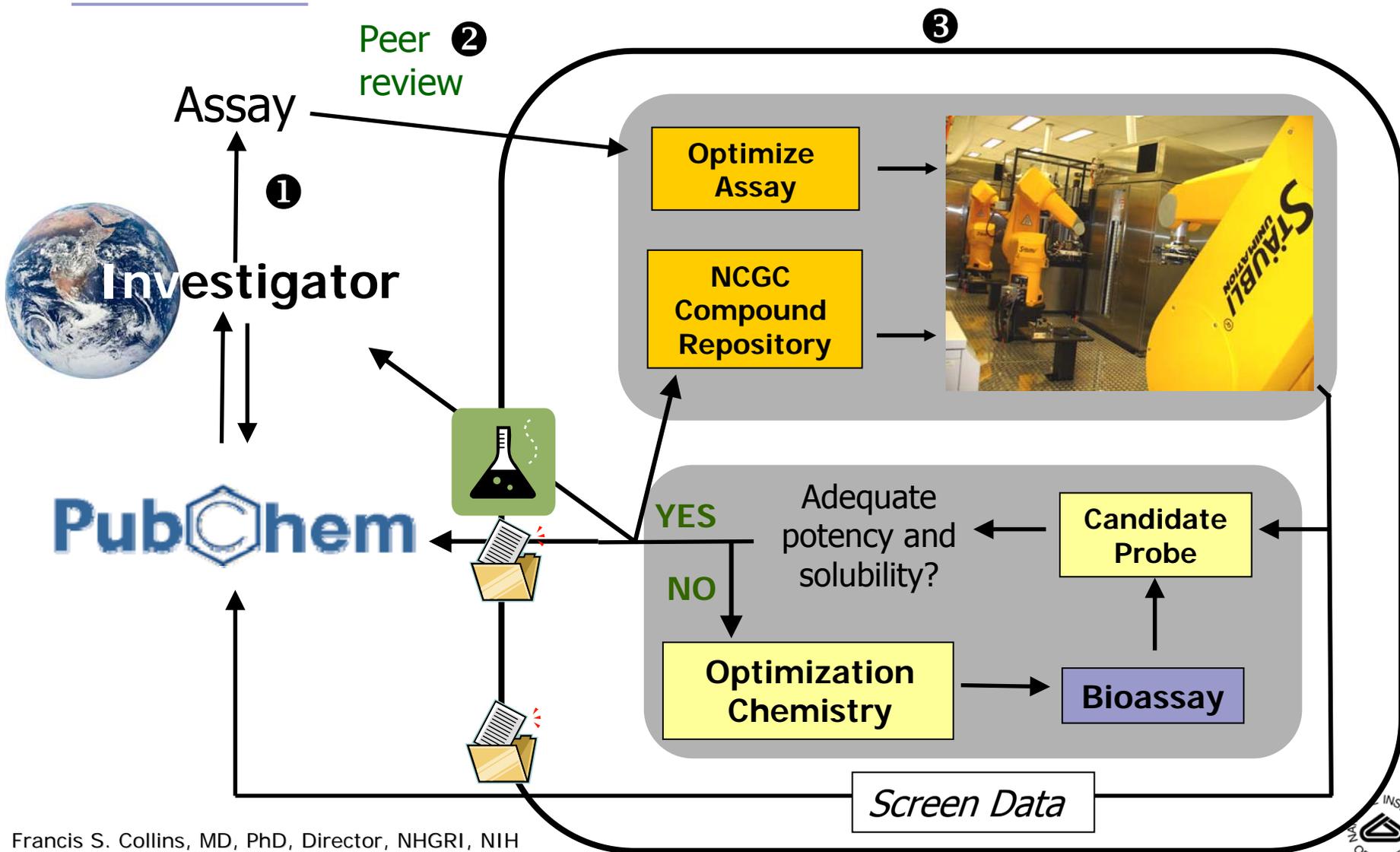
Molecular Libraries Compound Collection

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



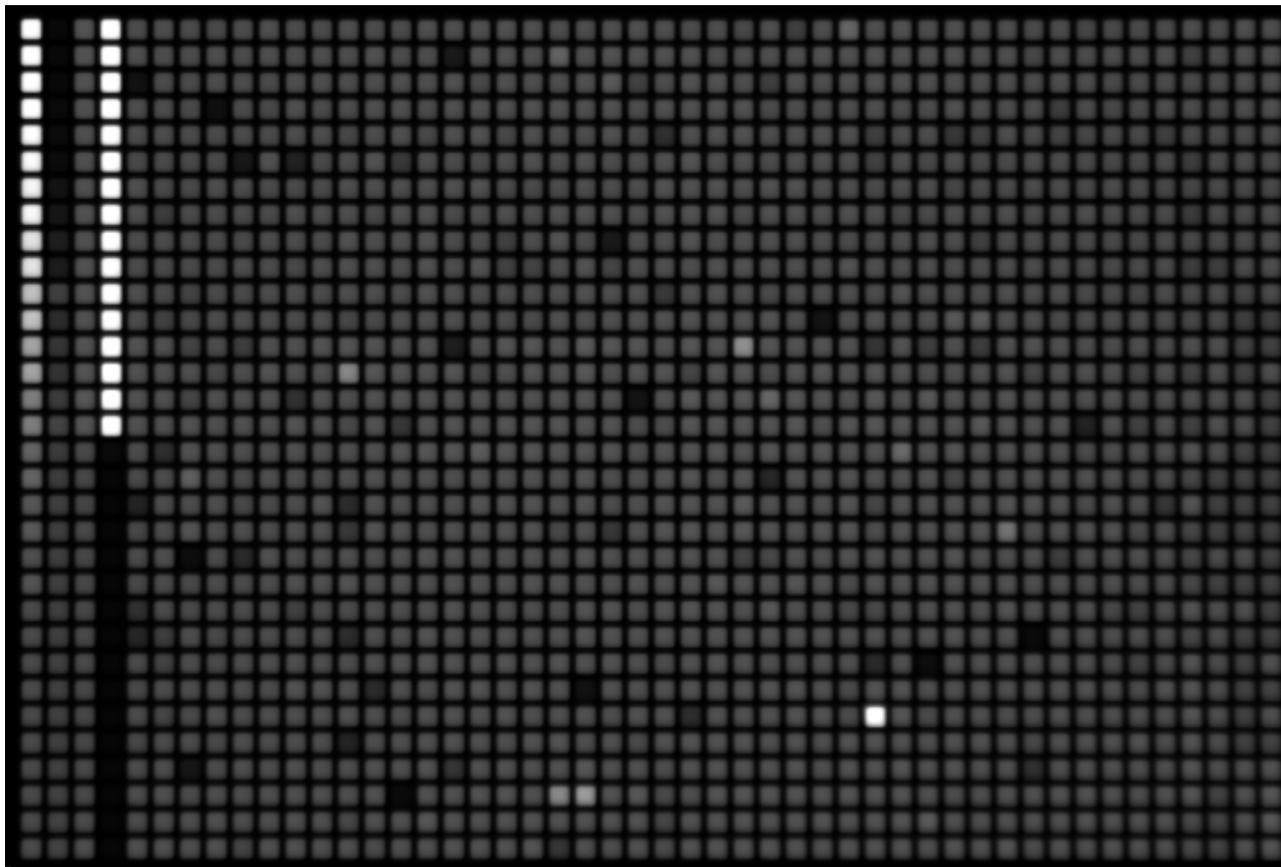
- Expanding the collection
 - Purchase of next 100,000 ongoing; 500,000 at maturity
 - Less stringent property requirements, filling out SAR clusters of 3-5
 - Molecular Libraries Roadmap Chemical Diversity initiatives
 - *Pilot scale libraries for HTS*
 - Centers for Methodology in Library Development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

NCGC Operation

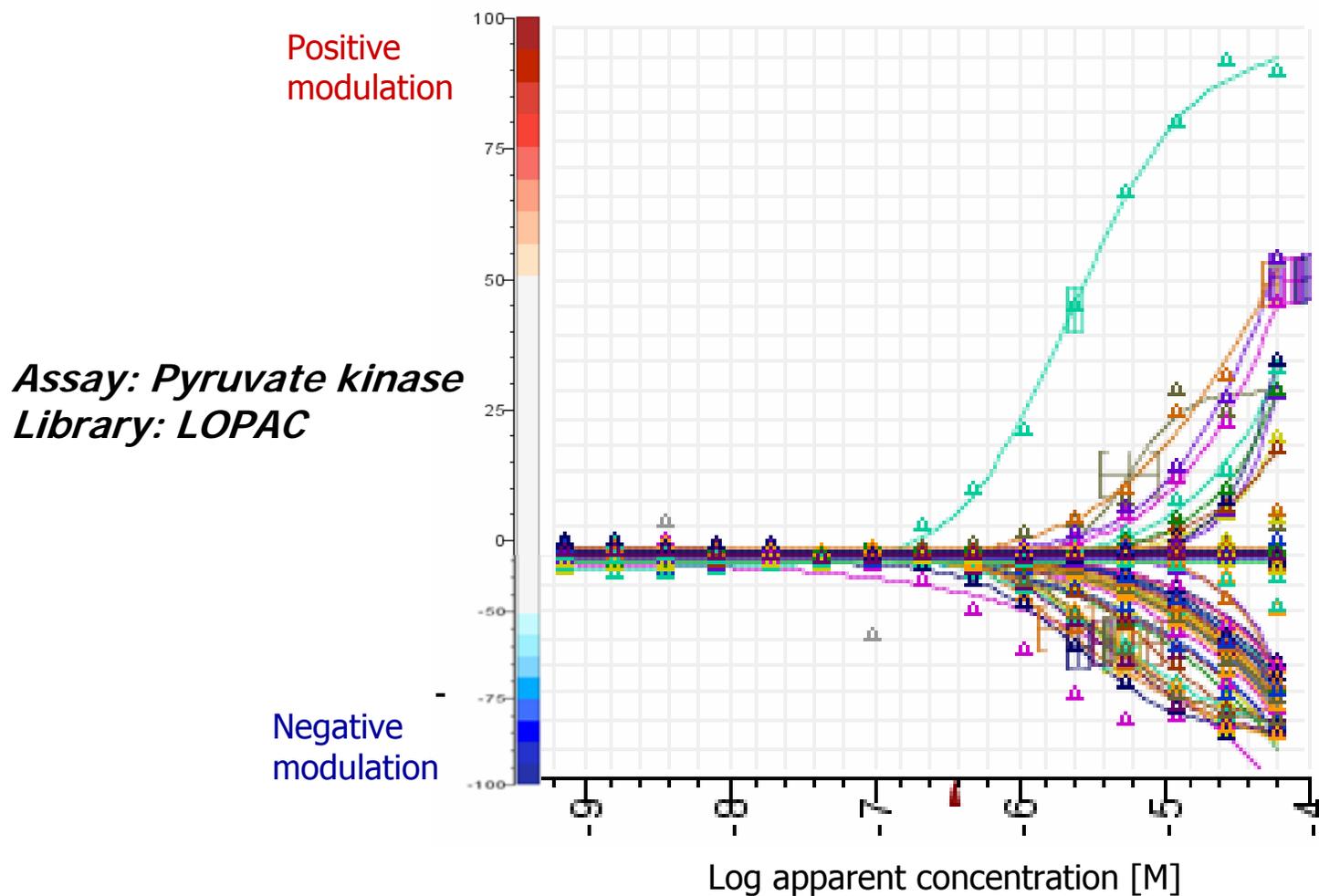


Titration screening: Raw luminescence data

57 μ M



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



Products of the MLSCN

- Chemical probes of gene, pathway, and cell functions
- Optimized only for potency ($\leq 1\mu\text{M}$) and aqueous solubility
 - SAR ideally also present
- No IP obtained on any probes identified by the MLSCN
 - Maximal freedom of operation for
 - Basic research
 - Target validation
 - Use of results as starting points for further optimization

PubChem

- PubChem is a free, publicly available database that provides information about potential starting points for the development of new medications.
- PubChem connects chemical information with biomedical research and clinical information in a connect-the-dots fashion.
- PubChem is a critical part of the NIH Molecular Libraries initiative.
- PubChem is the latest member of the powerful family of integrated databases operated by the National Library of Medicine.
- The integration of these databases makes the whole much greater than the sum of its parts

PubChem Contents ...

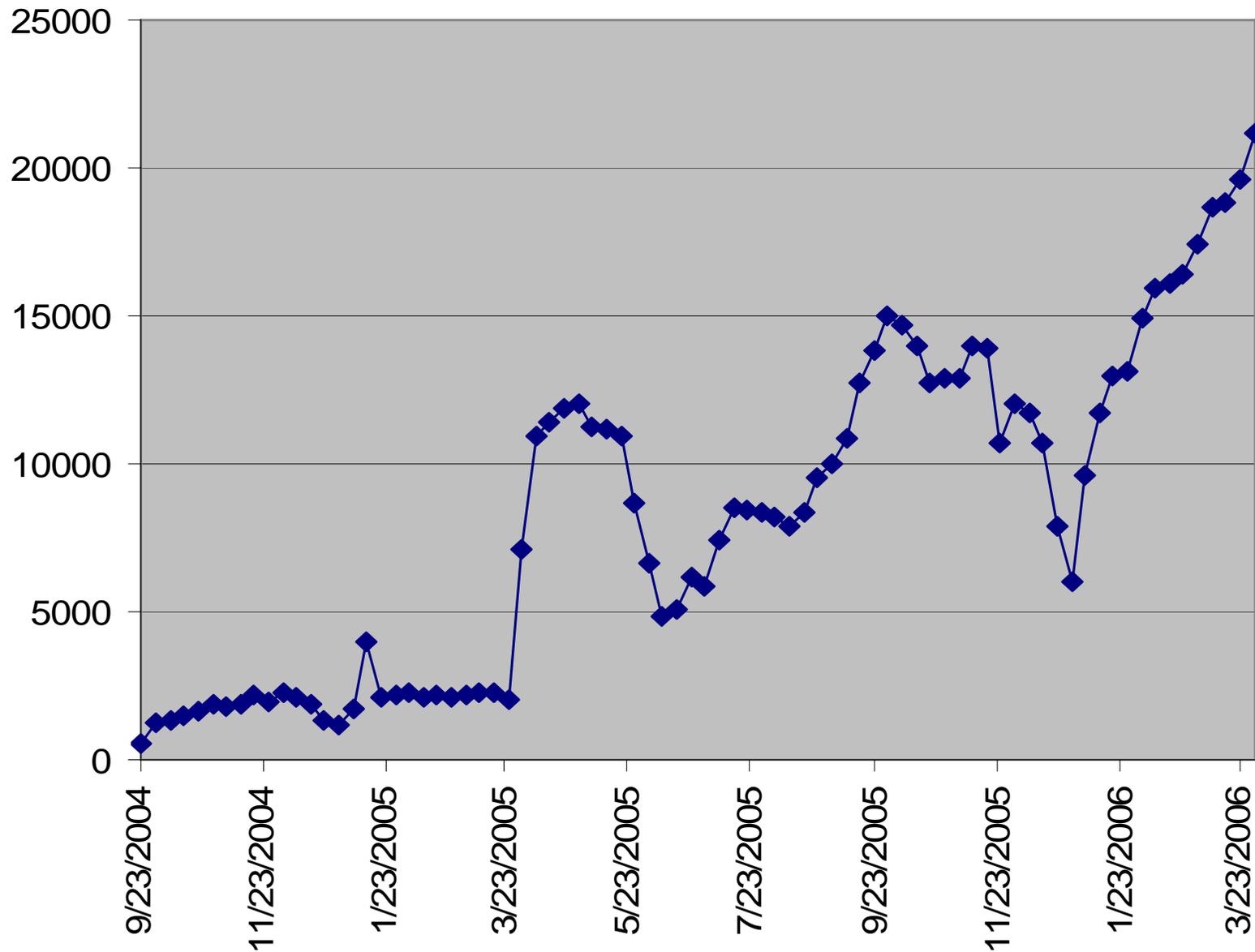
... 194 Bioassays Contributed

... 10,316,814 Substances Contributed

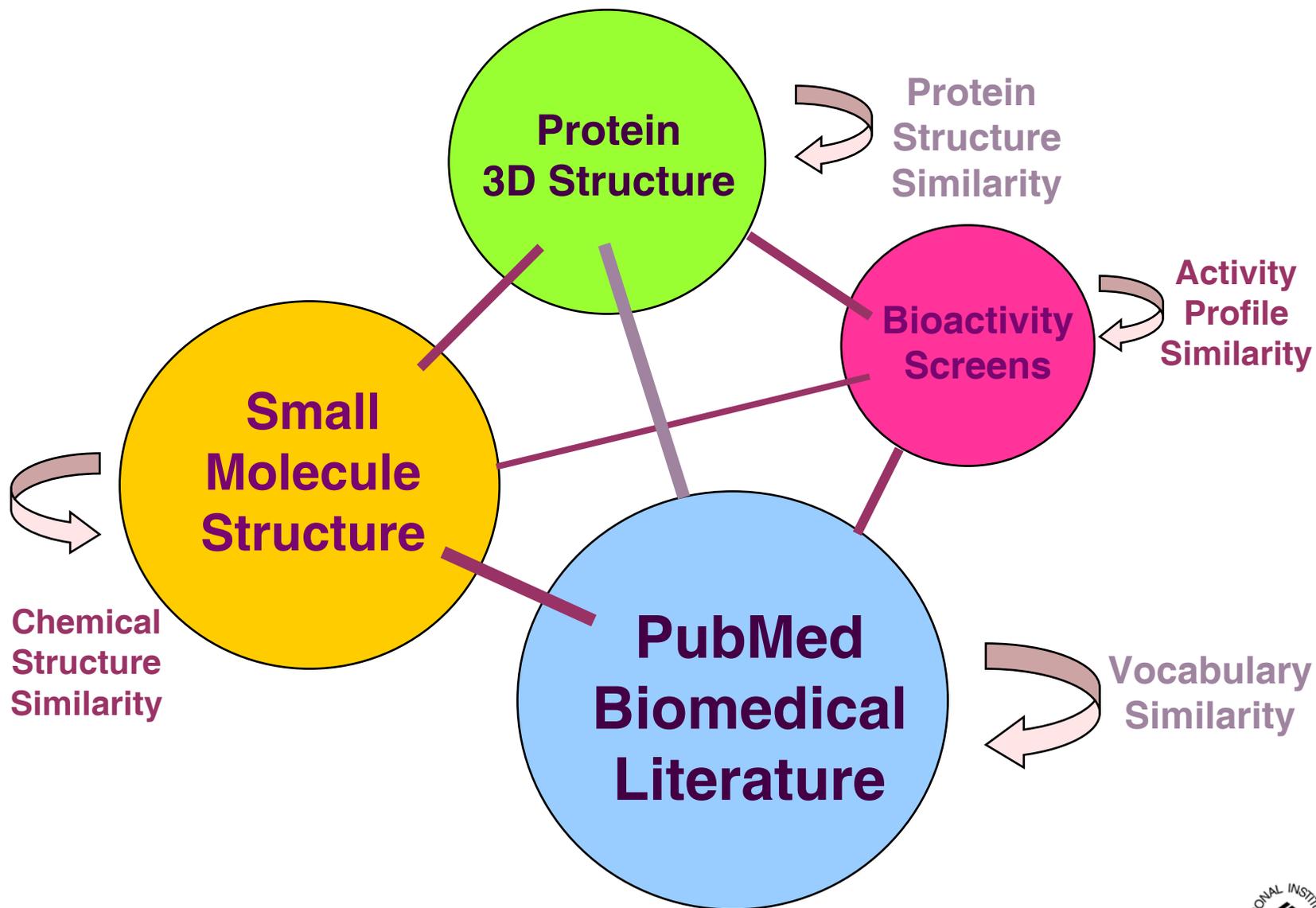
... 5,338,430 Unique Compound Structures

... 41 Depositing Organizations

Growth in PubChem Users per Day



PubChem Database Integration



Search for “Gaucher” ...

Entrez cross-database search - Microsoft Internet Explorer

File Edit View Favorites Tools Help

← Back → Search Favorites Media

Address <http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi> Go Links »

NCBI Entrez, The Life Sciences Search Engine

HOME SEARCH SITE MAP PubMed All Databases Human Genome GenBank Map Viewer BLAST

Search across databases GO CLEAR Help

3577		PubMed: biomedical literature citations and abstracts	?	57		Books: online books	?
244		PubMed Central: free, full text journal articles	?	45		OMIM: online Mendelian Inheritance in Man	?
2		Site Search: NCBI web and FTP sites	?	2		OMIA: Online Mendelian Inheritance in Animals	?
118		Nucleotide: sequence database (GenBank)	?	5		UniGene: gene-oriented clusters of transcript sequences	?
226		Protein: sequence database	?	none		CDD: conserved protein domain database	?
none		Genome: whole genome sequences	?	41		3D Domains: domains from Entrez Structure	?
13		Structure: three-dimensional macromolecular structures	?	20		UniSTS: markers and mapping data	?
none		Taxonomy: organisms in GenBank	?	2		PopSet: population study data sets	?
170		SNP: single nucleotide polymorphism	?	468		GEO Profiles: expression and molecular abundance profiles	?
10		Gene: gene-centered information	?	none		GEO DataSets: experimental sets of GEO data	?

Internet

OMIM link to Gaucher disease ...

1q21'. A section titled 'TEXT' contains a paragraph: 'A number sign (#) is used with this entry because Gaucher disease is caused by mutation in the gene encoding acid-beta glucosidase (GBA; [606463](#)). Mutation in the same gene causes Gaucher disease type II ([230900](#)) and type III ([231000](#)).'. A section titled 'CLINICAL FEATURES' contains a paragraph: 'The cardinal features of type I Gaucher disease are hematologic abnormalities with hypersplenism, bone lesions, skin pigmentation, and pingueculae (brown spots of Gaucher cells at corneoscleral limbus). The disorder is particularly frequent in Ashkenazi Jews. The several forms of Gaucher disease are cerebroside lipidoses. The disease has been diagnosed as early as the first week of life and as late as 86 years. Although the disorder is clearly autosomal recessive in most cases, a dominant form was suggested by [Hsia et al. \(1959\)](#) on the basis of affected father and son. The father was German-Jewish and the mother Swedish-English. Even'. On the left side of the browser window, there is a navigation menu with the NCBI logo and various links like 'MIM #230800', 'Text', 'Clinical', 'Features', 'Clinical Management', 'Population Genetics', 'Molecular Genetics', 'References', 'Contributors', 'Creation Date', 'Edit History', 'Clinical Synopsis', 'Gene map', and 'Entrez Gene' with sub-links for 'Nomenclature', 'RefSeq', 'GenBank', and 'Protein'."/>

OMIM - GAUCHER DISEASE, TYPE I - Microsoft Internet Explorer

Address <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=230800>

NCBI

MIM #230800

Text
Clinical
Features
Clinical Management
Population Genetics
Molecular Genetics
References
Contributors
Creation Date
Edit History

• Clinical Synopsis
• Gene map

Entrez Gene
Nomenclature
RefSeq
GenBank
Protein

GD I
GAUCHER DISEASE, NONCEREBRAL JUVENILE
GLUCOCEREBROSIDASE DEFICIENCY
ACID BETA-GLUCOSIDASE DEFICIENCY
GBA DEFICIENCY

Gene map locus [1q21](#)

TEXT

A number sign (#) is used with this entry because Gaucher disease is caused by mutation in the gene encoding acid-beta glucosidase (GBA; [606463](#)). Mutation in the same gene causes Gaucher disease type II ([230900](#)) and type III ([231000](#)).

CLINICAL FEATURES

The cardinal features of type I Gaucher disease are hematologic abnormalities with hypersplenism, bone lesions, skin pigmentation, and pingueculae (brown spots of Gaucher cells at corneoscleral limbus). The disorder is particularly frequent in Ashkenazi Jews. The several forms of Gaucher disease are cerebroside lipidoses. The disease has been diagnosed as early as the first week of life and as late as 86 years. Although the disorder is clearly autosomal recessive in most cases, a dominant form was suggested by [Hsia et al. \(1959\)](#) on the basis of affected father and son. The father was German-Jewish and the mother Swedish-English. Even

PubChem BioAssay Record ...

PubChem Substance And Assay Service at NCBI - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Mail

Address <http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=360> Go Links

HOME SEARCH SITE MAP PubMed Entrez Structure GenBank PubChem Help

BioAssay Summary

BioAssay ID (AID): [360](#)
Source: [NCGC](#)
Name: [Glucocerebrosidase](#)

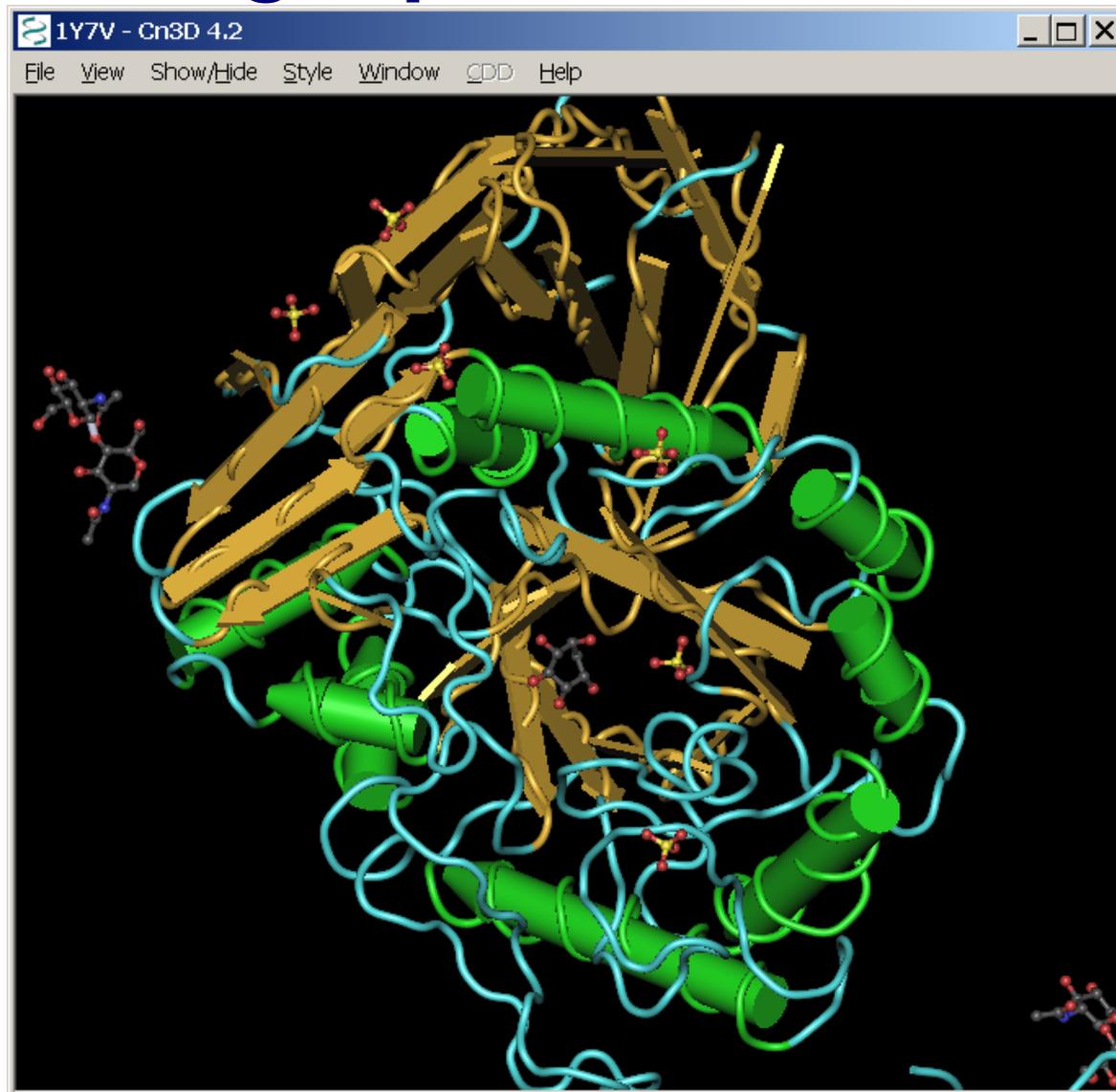
[Links](#) [Description](#) [Protocol](#) [Show Data](#) [Select Data](#)

Links:

Substances tested: [48125](#); active: [549](#); inactive: [45736](#); inconclusive: [1840](#)
PubMed: [4](#)
OMIM: [3](#)
MIMDIP: [1](#)

Done Local intranet

Link to target protein 3D structure ...



PubChem BioAssay Results ...

PubChem Substance And Assay Service at NCBI - Microsoft Internet Explorer

File Edit View Favorites Tools Help

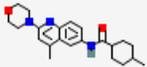
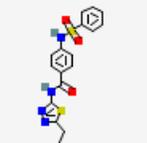
Back Forward Stop Home Search Favorites Media Mail Print View Source

Address nih.gov/assay/assay.cgi?&IDWV=true&ResultDisplay=true&aid=360&refresh_count=4&req_id=761592328923984462 Go Links »

Name: [Glucocerebrosidase](#)

[back to summary](#)

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

Structure	PubChem		Outcome	Activity Score	Submitter	Submission Date	Activity Direction	Activity Qualifier	Qualifi AC50
	SID	CID							
	4243169	3237927	Active	72	ncgc	19 Jan 2006	decreasing	=	6.06e-008
	4264637	2210290	Active	71	ncgc	19 Jan 2006	decreasing	=	7e-00
									

Done Local intranet



Fundamental Changes in Biomedical Research due to the Molecular Libraries Roadmap

- Small molecule high-throughput screening, chemistry, and informatics are available to the academic and nonprofit sectors on a scale previously available only to pharma and biotech
- Biological activities of small molecules are available to the research community and fully integrated with other medical informatics resources for the first time
- New resources for synthetic and natural products chemistry are being supported by NIH at an unprecedented level
- Researchers in the public and private sectors can get grant support to turn their basic discoveries into assays for small molecule high throughput screening for the first time
- Pharma and biotech have public sector screening data to advance their own drug discovery programs for the first time
 - Directly facilitates developing new drugs on new targets from the Human Genome Project
- Small molecules are being viewed as pre-competitive research tools rather than protected intellectual property due to the unprecedented sharing data sharing policies of the MLSCN





NIH Roadmap for Medical Research

IDEAS, PEOPLE, RESOURCES, LEADERSHIP

