PubChem: An Open Repository for Chemical Structure and Biological Activity Information

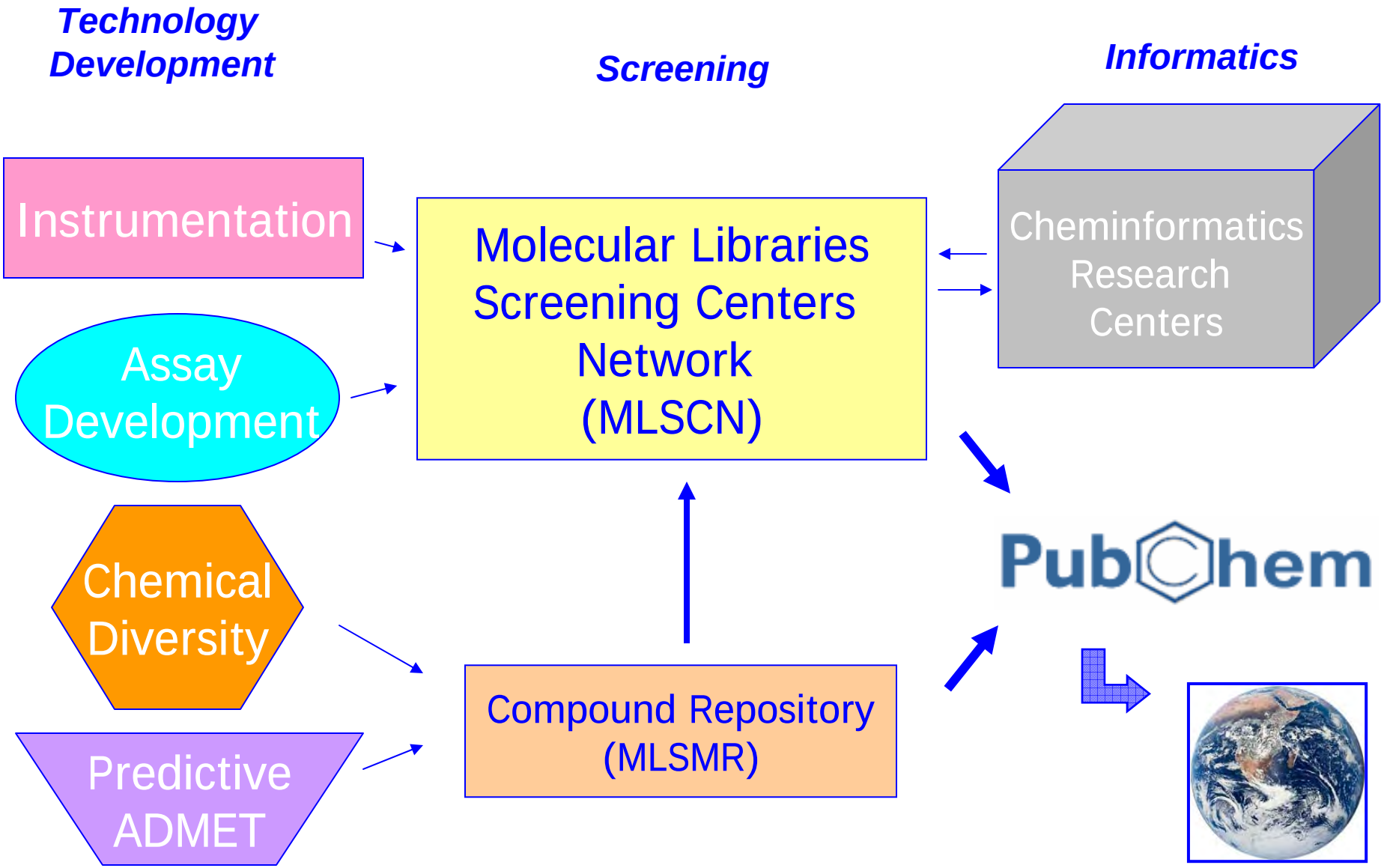
Stephen H. Bryant

Forum on Computational Toxicology

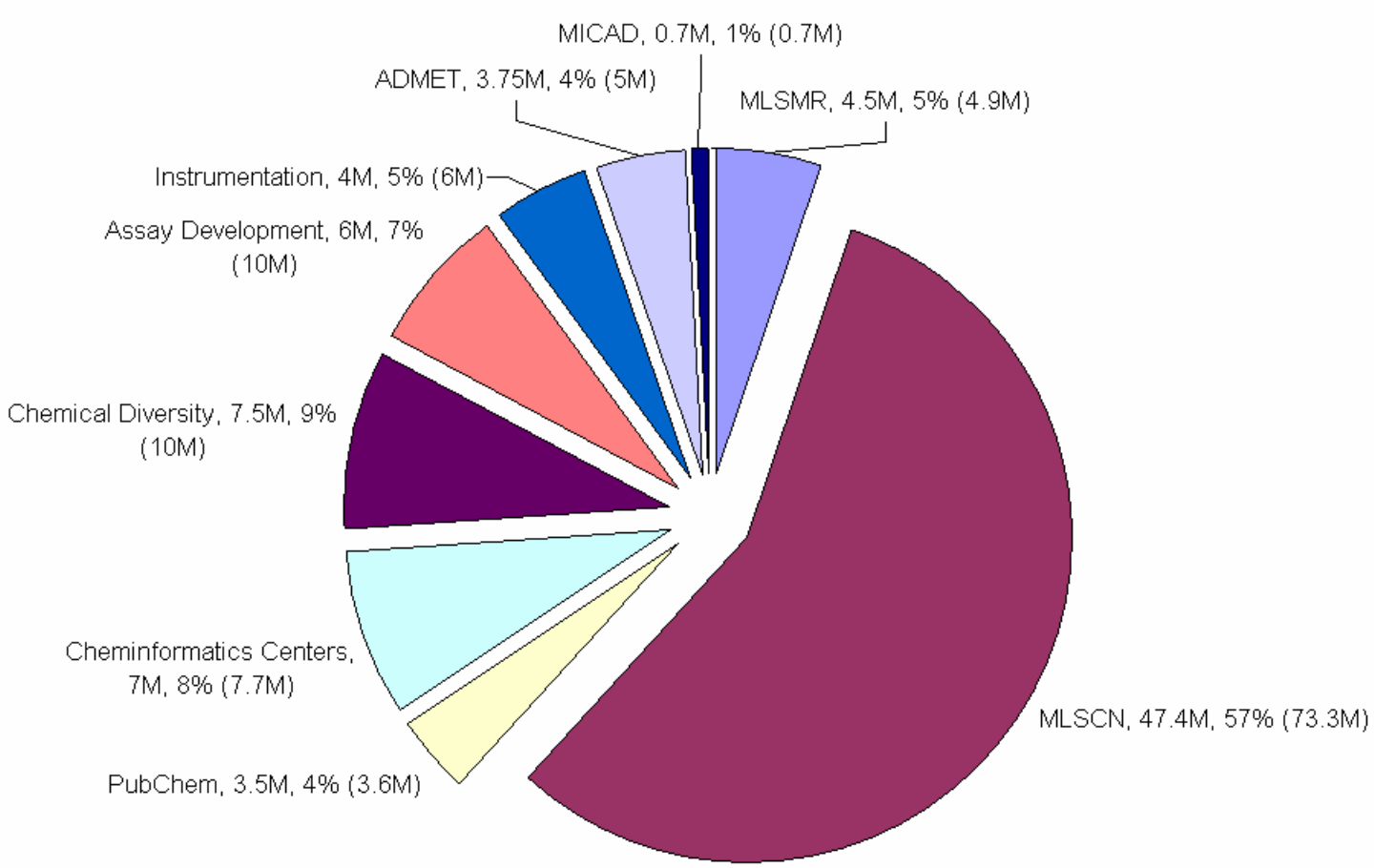
Research Triangle Park

May 21-23, 2007

The Molecular Libraries Program: An Integrated Initiative



Molecular Libraries Components …



PubChem Goals …

… Archive molecular structure and bioassay data from the Molecular Libraries Screening Center Network

… Provide search, retrieval and data analysis tools to optimize utility

PubChem Goals …

… Further optimize research utility by including other public sources of chemical structure and bioactivity information

… and by integration with other NIH Biomedical information resources whenever possible

PubChem Approach …

… “GenBank model”

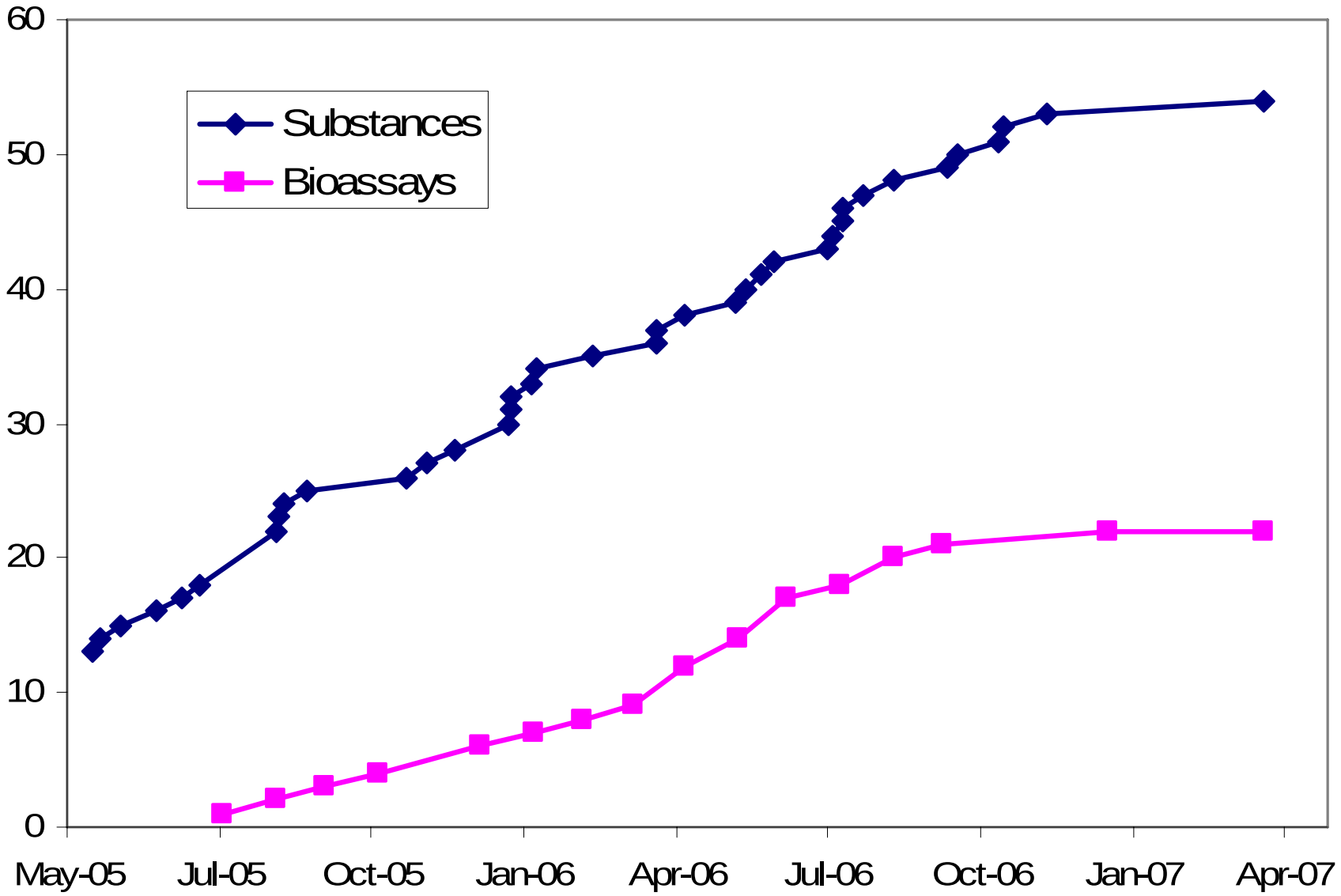
… direct depositions by investigators

… highly automated (low database cost)

… 25 year precedents in biology

… less precedent in chemistry

Growth In PubChem Contributing Organizations



PubChem Contents …

… Contributed substance records

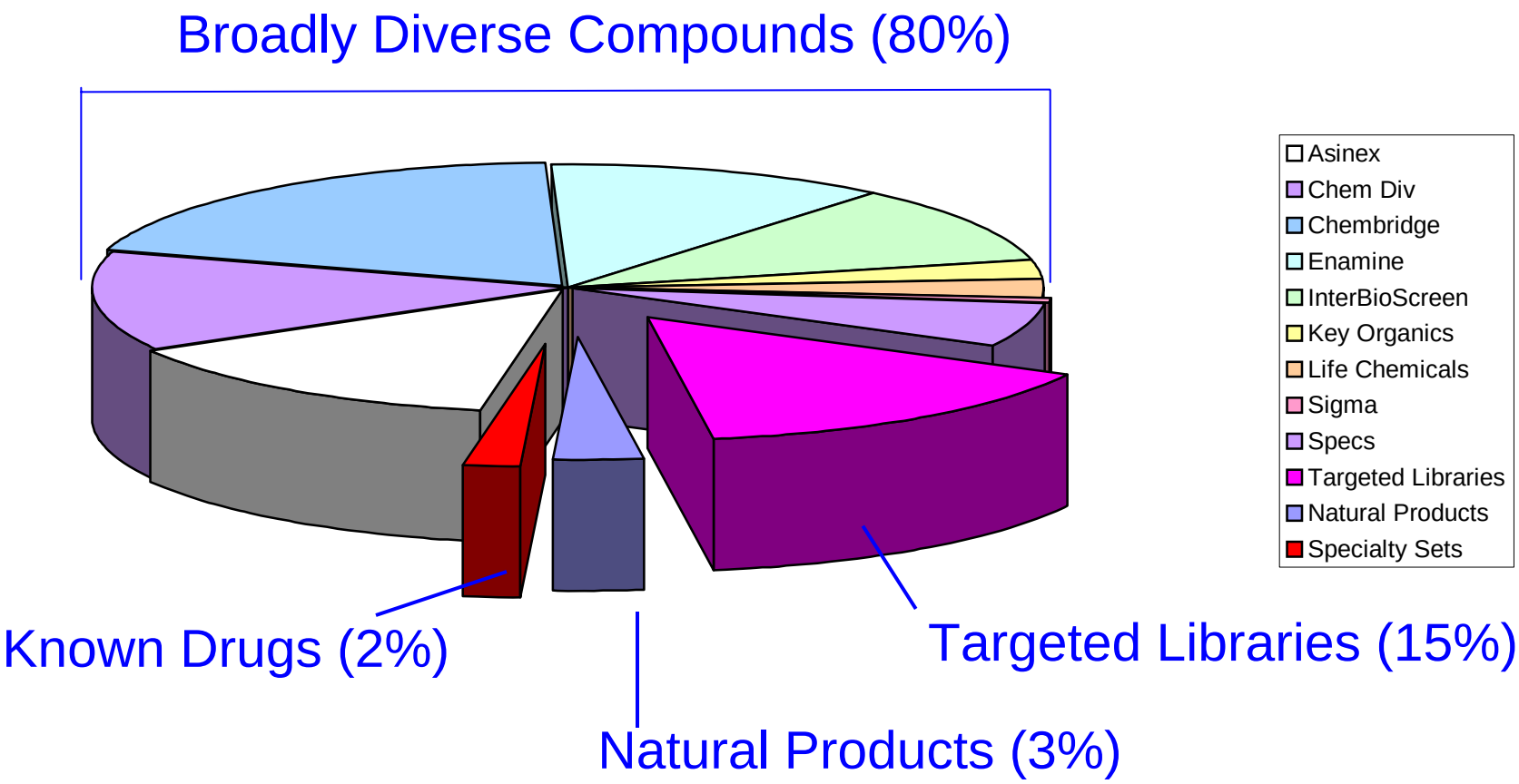
… with chemical structure

… chemical names and comments

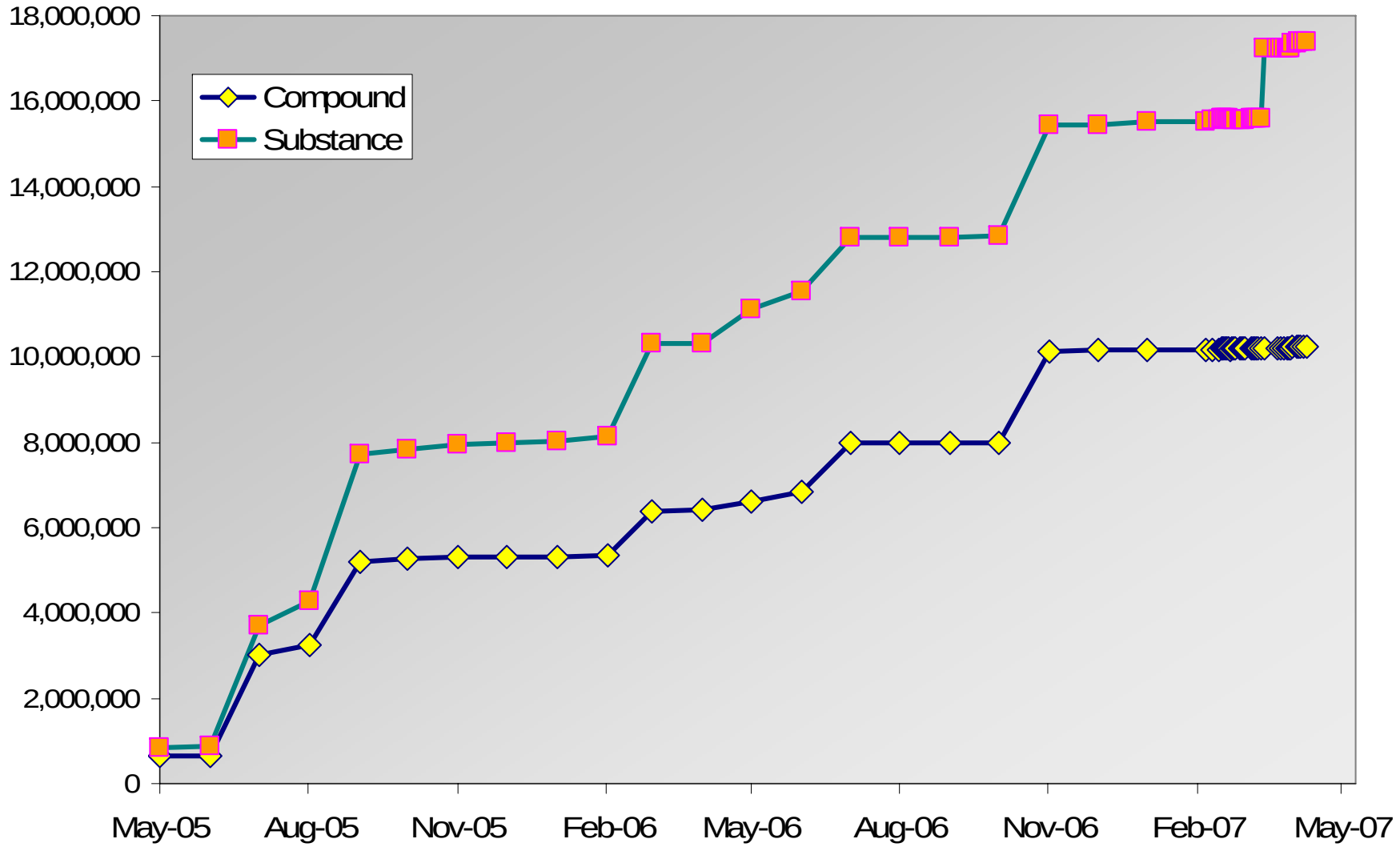
… links to contributor web sites

… contributed links to other NCBI biomedical databases

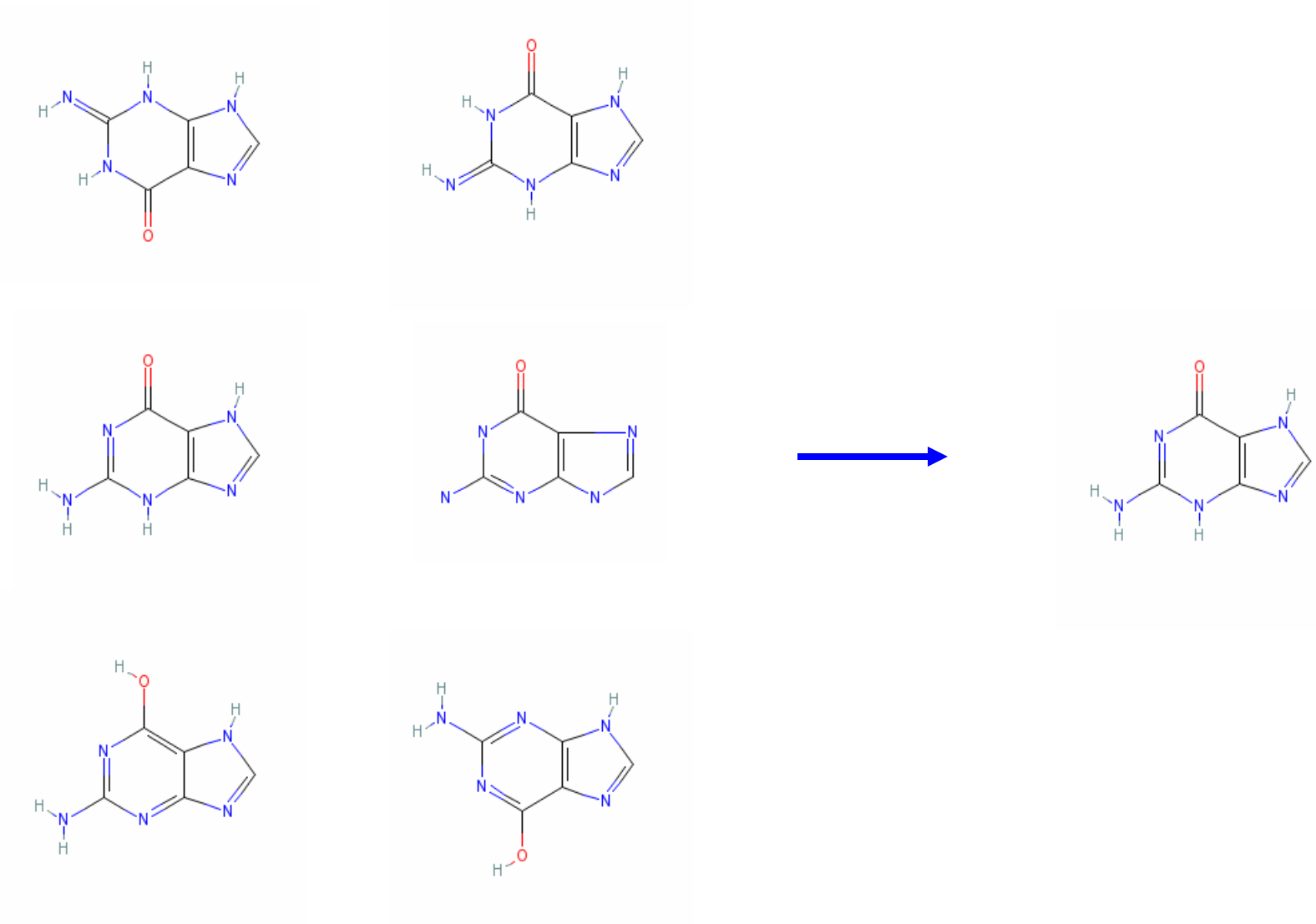
To Date 203,554 Molecular Libraries Compounds in PubChem



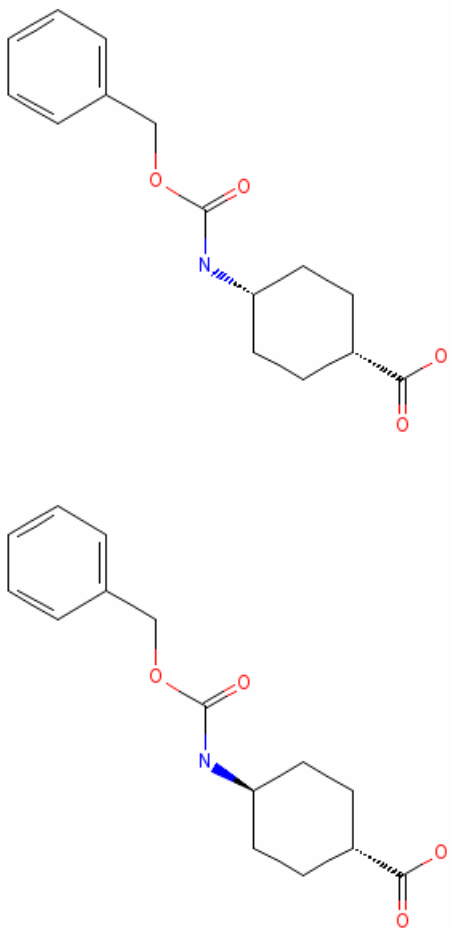
Growth In PubChem Substances / Compounds

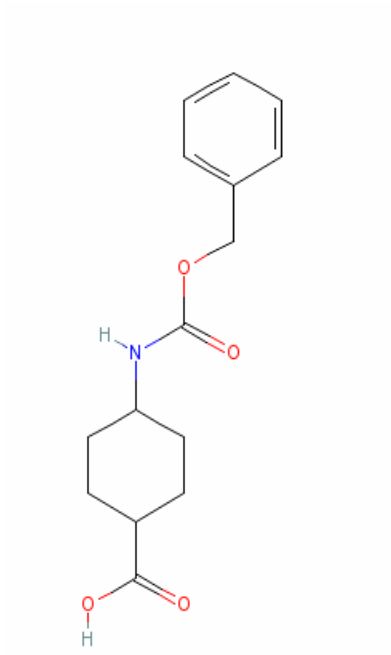


Compound Challenges ...



PubChem Standardization ...





PubChem Contents …

… Contributed bioassay records

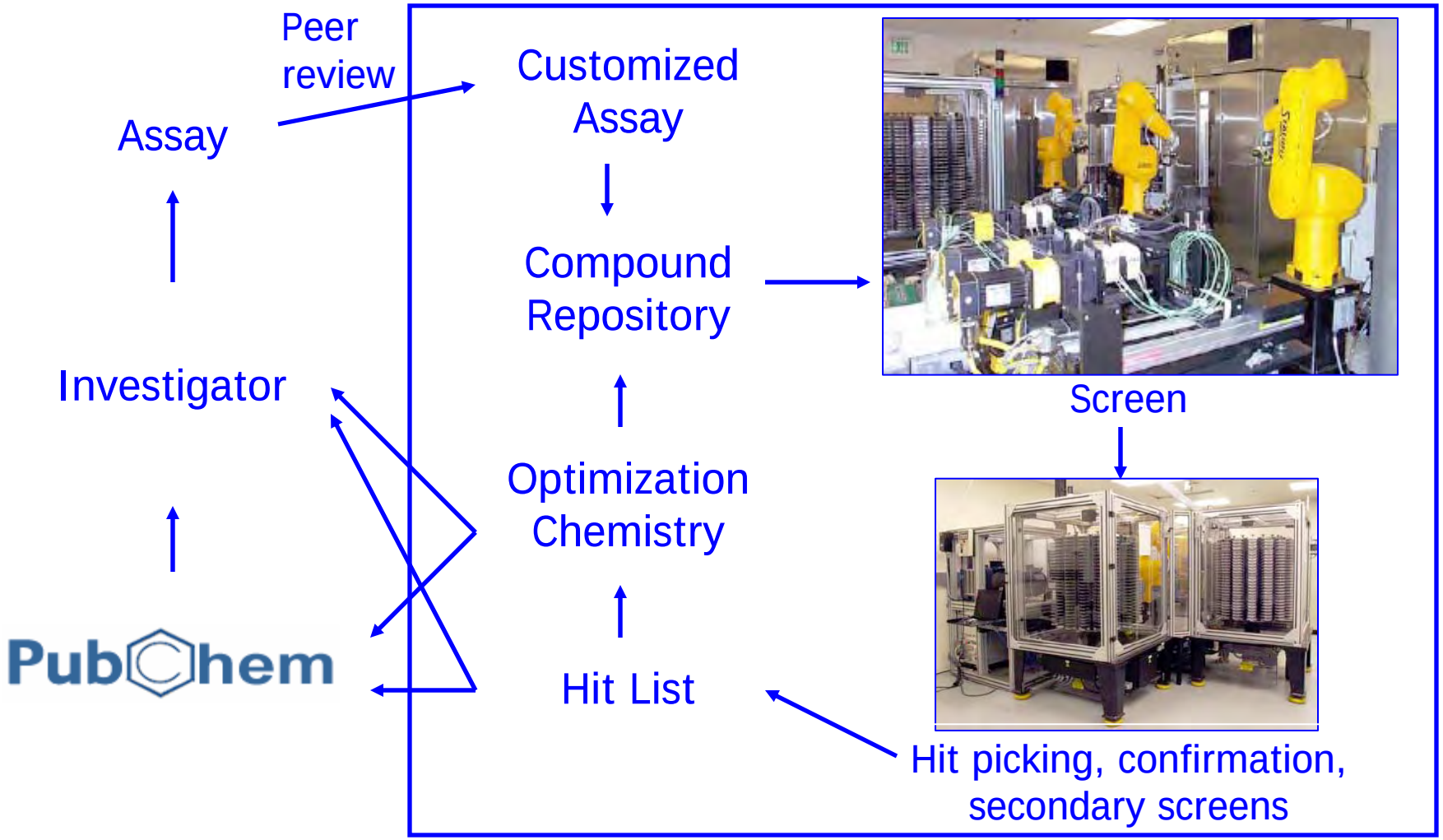
… with assay description / protocol

… links to tested substances

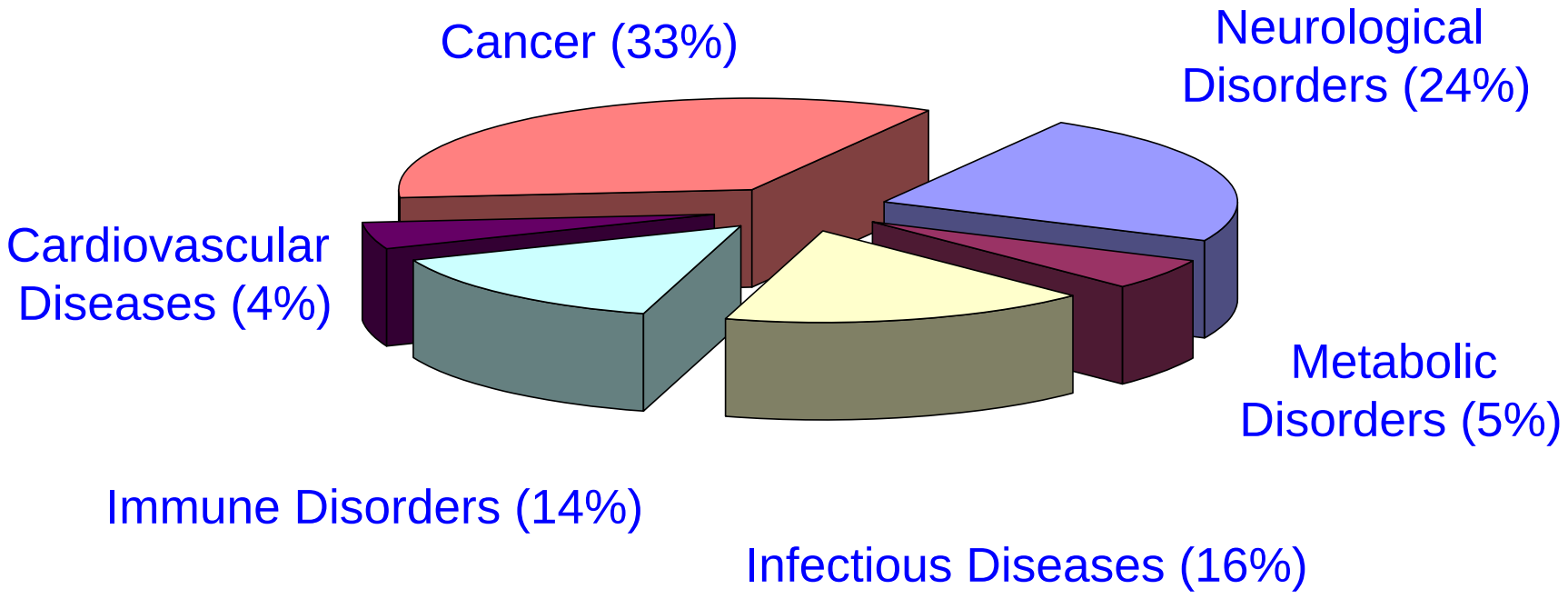
… summary and detailed test results

… links to contributor web sites and other NCBI databases

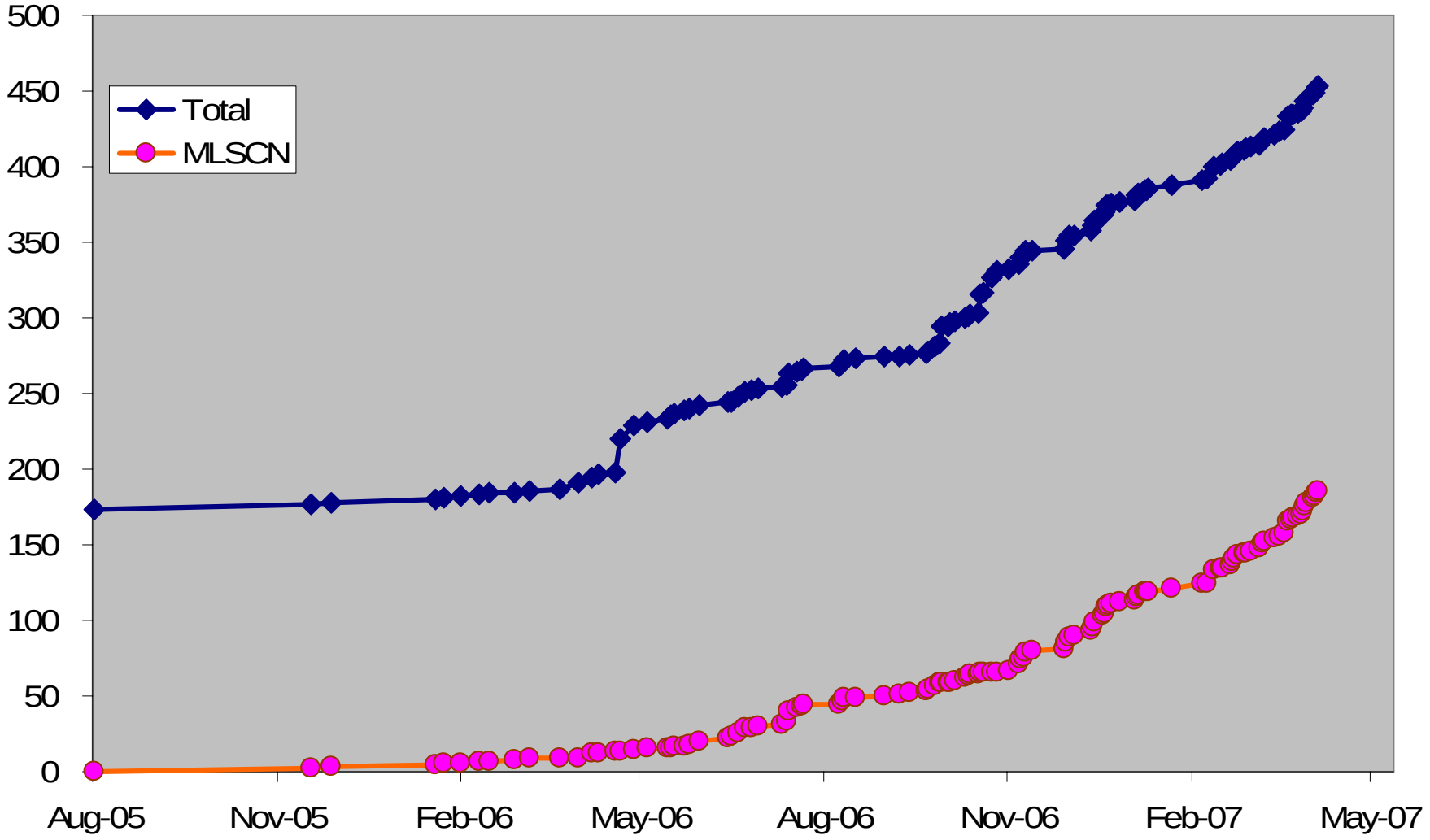
Molecular Libraries BioAssays …



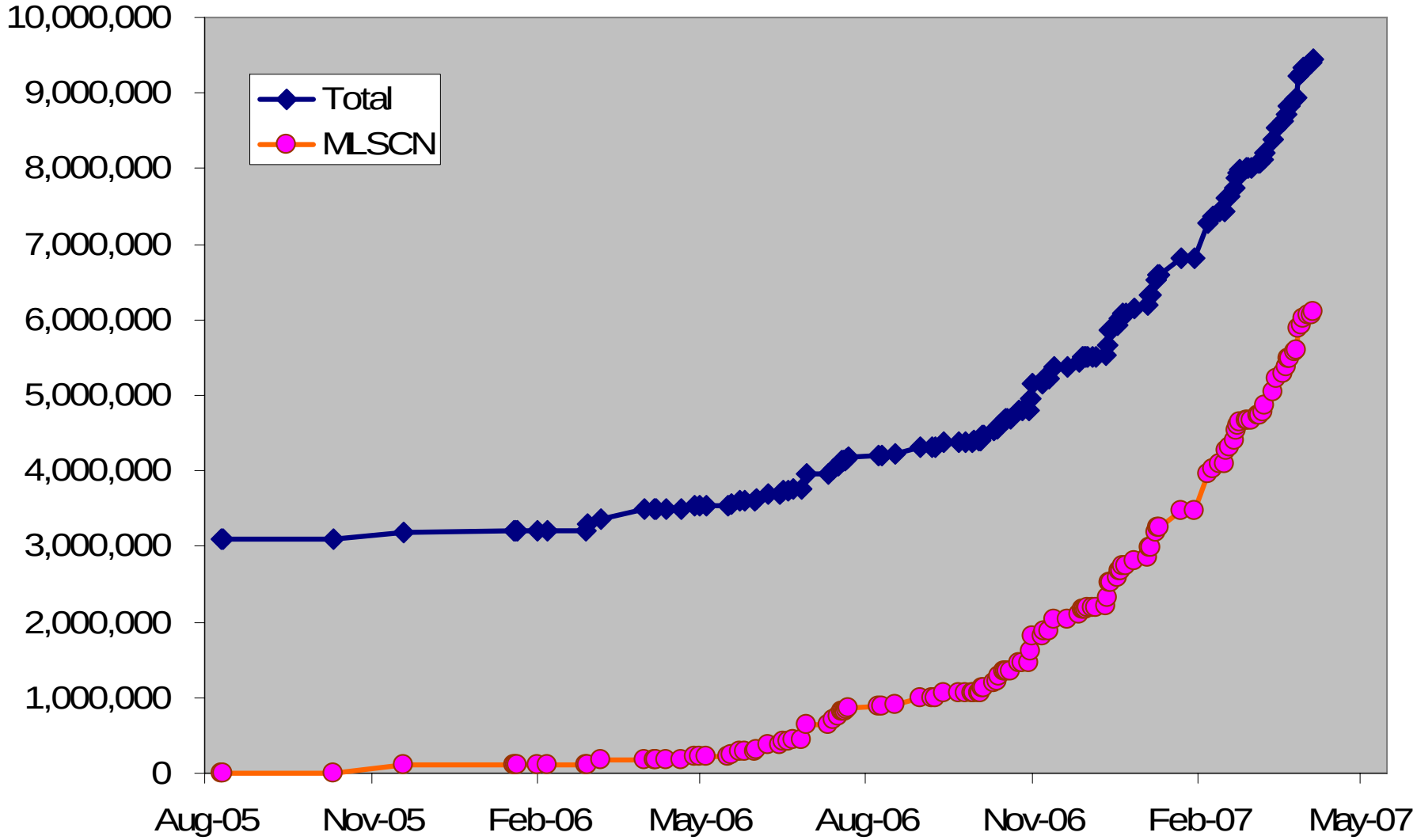
To Date 248 Molecular Libraries BioAssays in PubChem



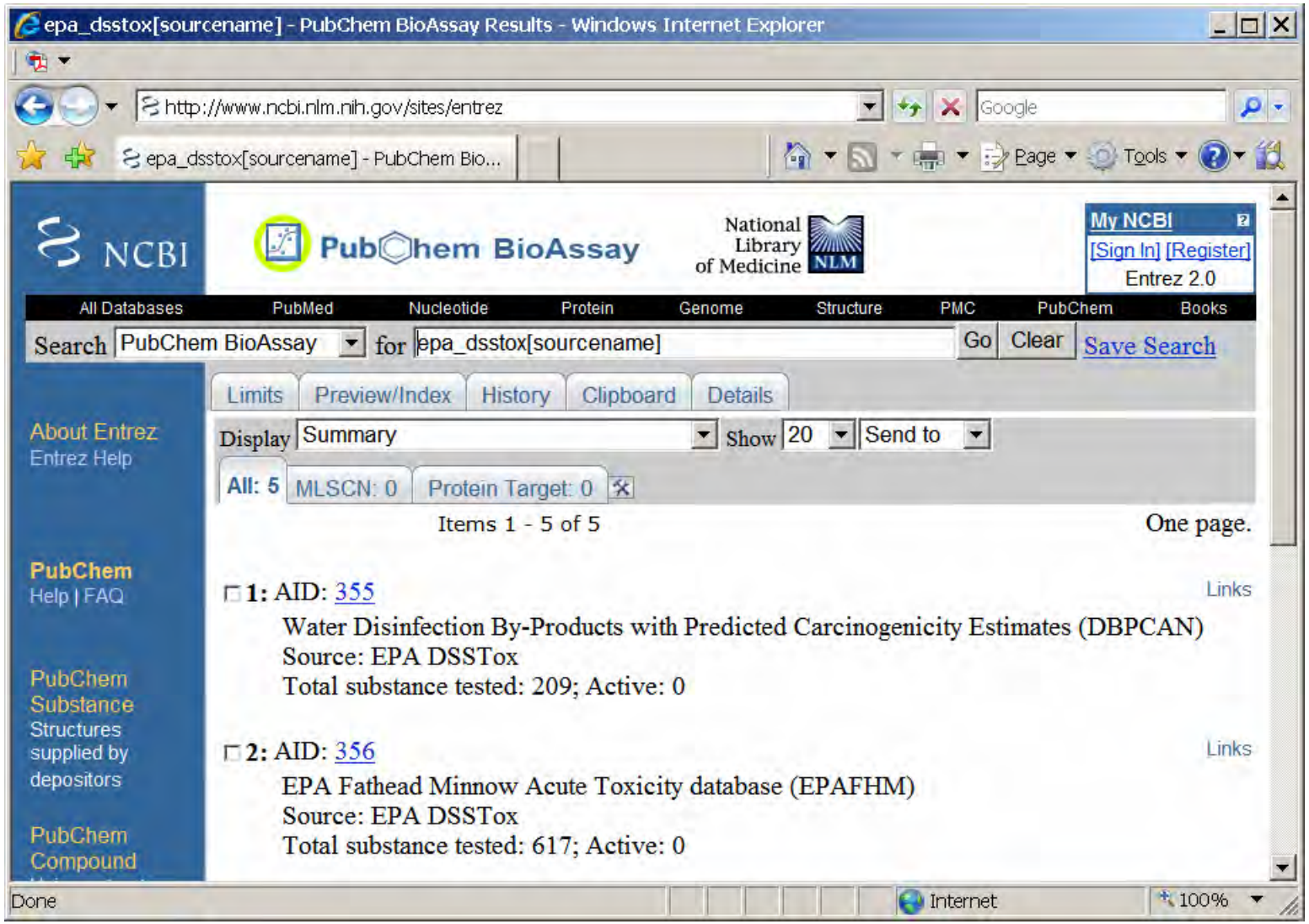
Growth In PubChem BioAssays



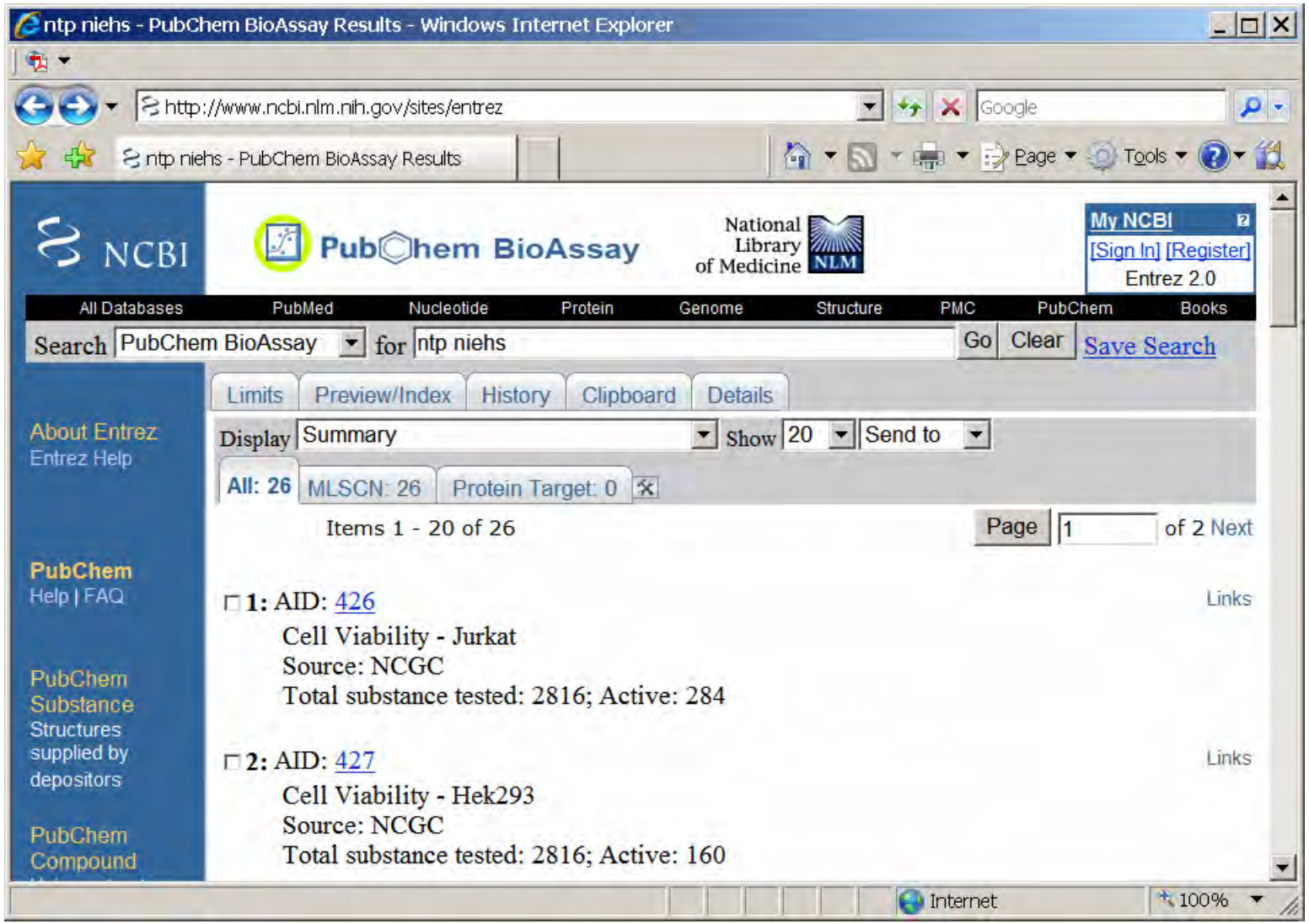
Growth In PubChem Tested Substances



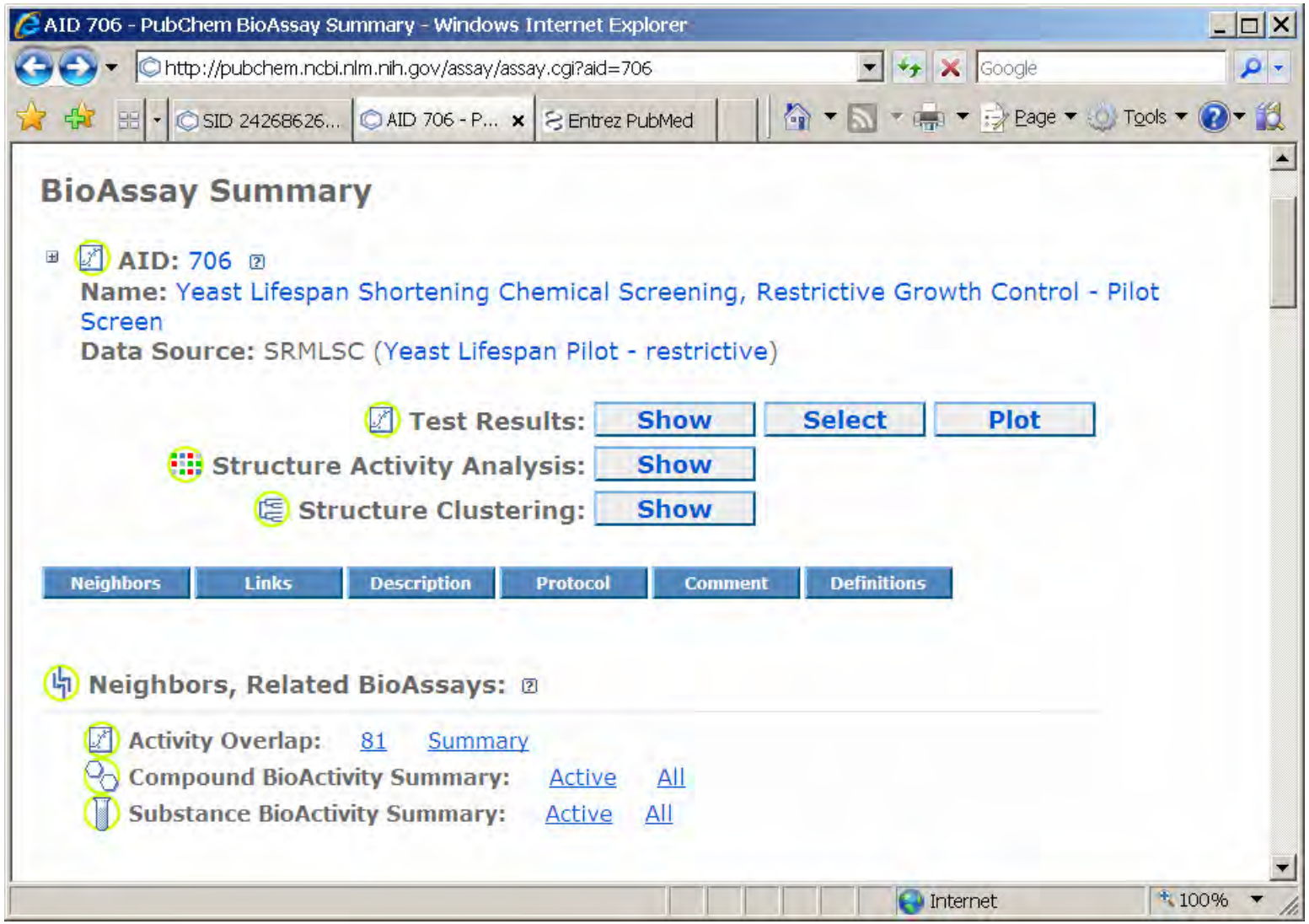
EPA “DSSTox” BioAssays ...



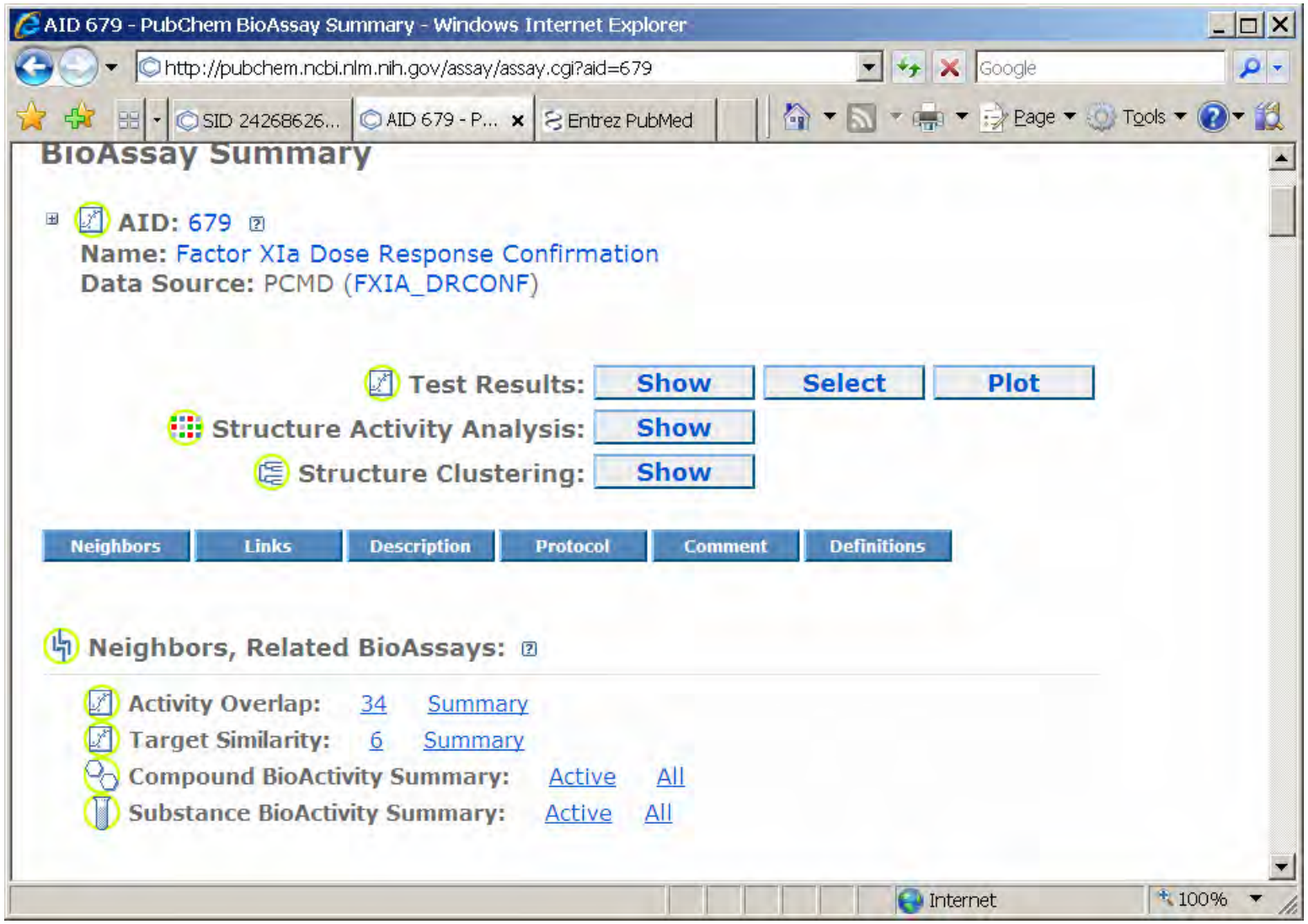
NCGC / NTP / NIEHS Screens ...



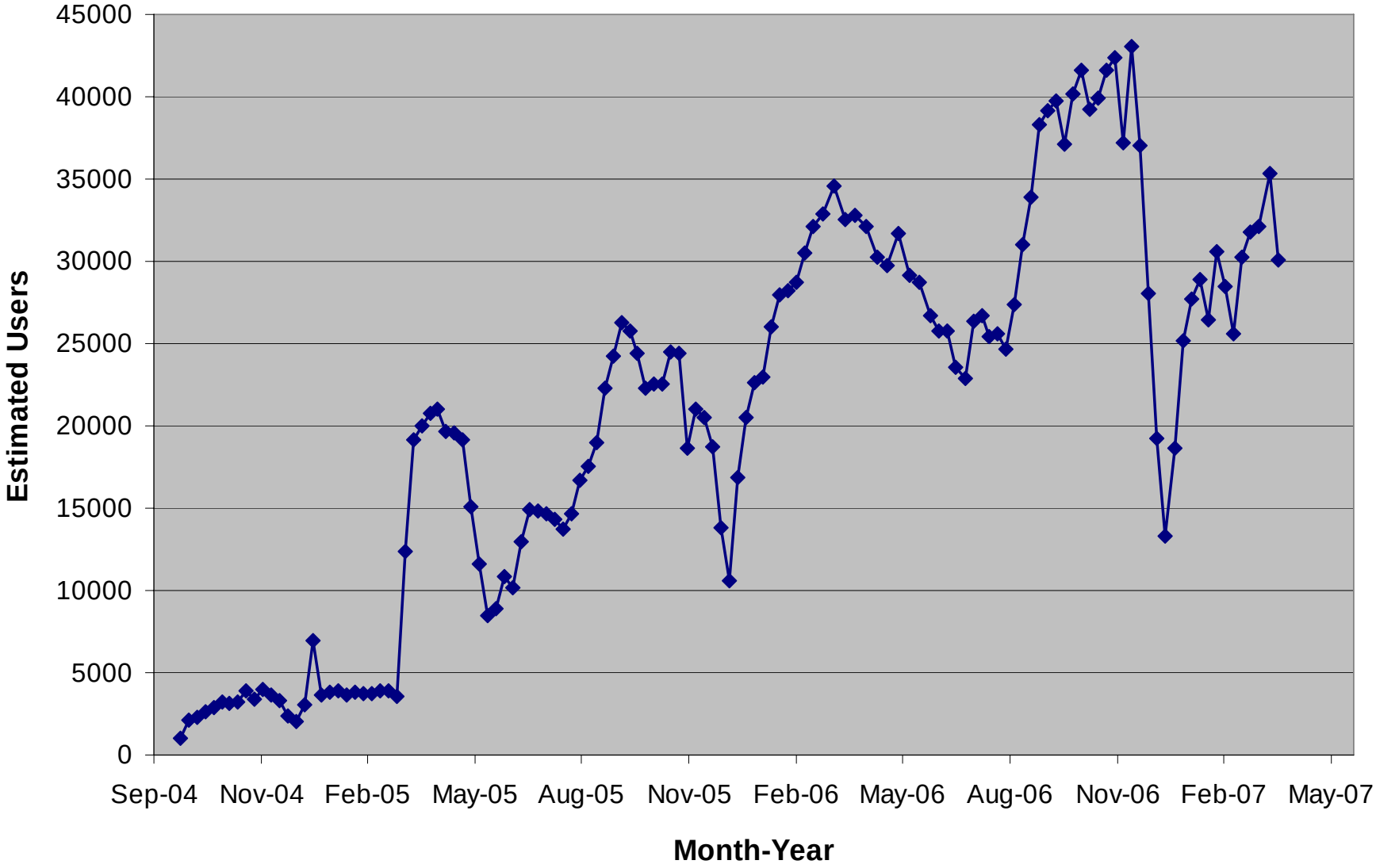
BioAssay Challenges ...



BioAssays are Diverse ...



Engineering Challenges …



PubChem Retrieval System …

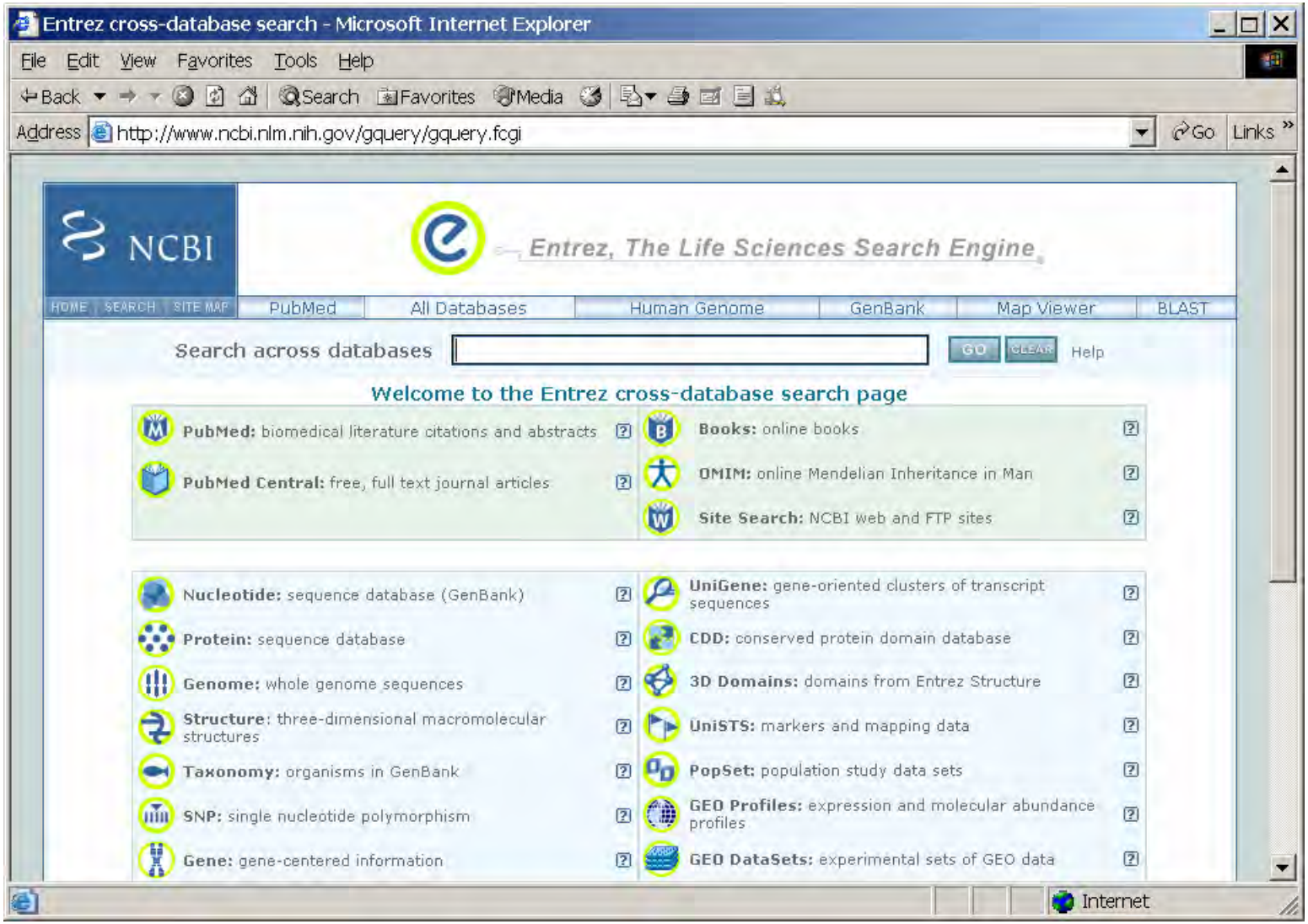
… Optimize “discoverability” for molecular biologists by integrating PubChem into NCBI’s Entrez / PubMed Search Engine

… Chemical structure search

… Bioassay result search

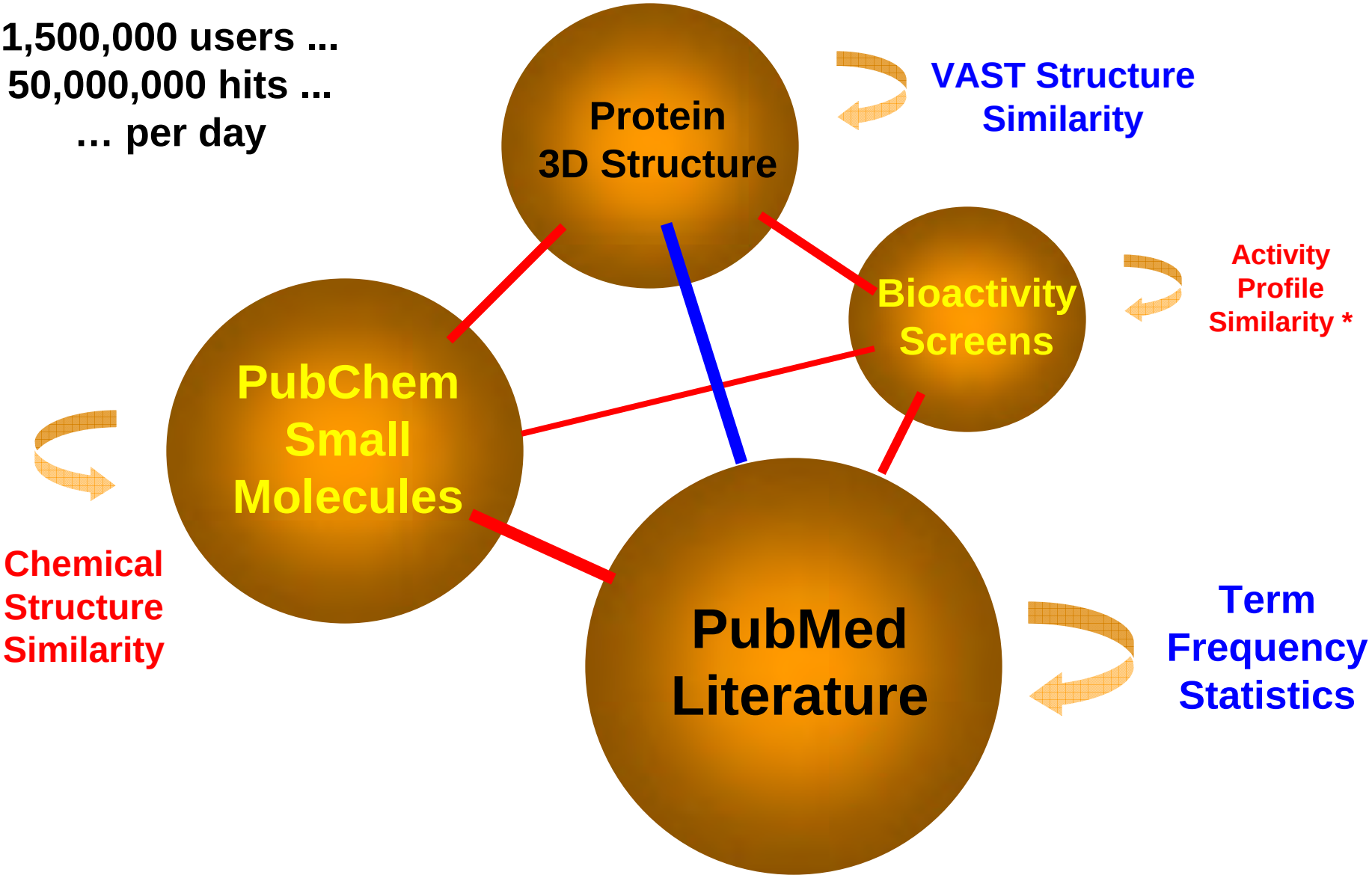
… Exploratory structure-activity tools

NCBI's Entrez Search Engine ...

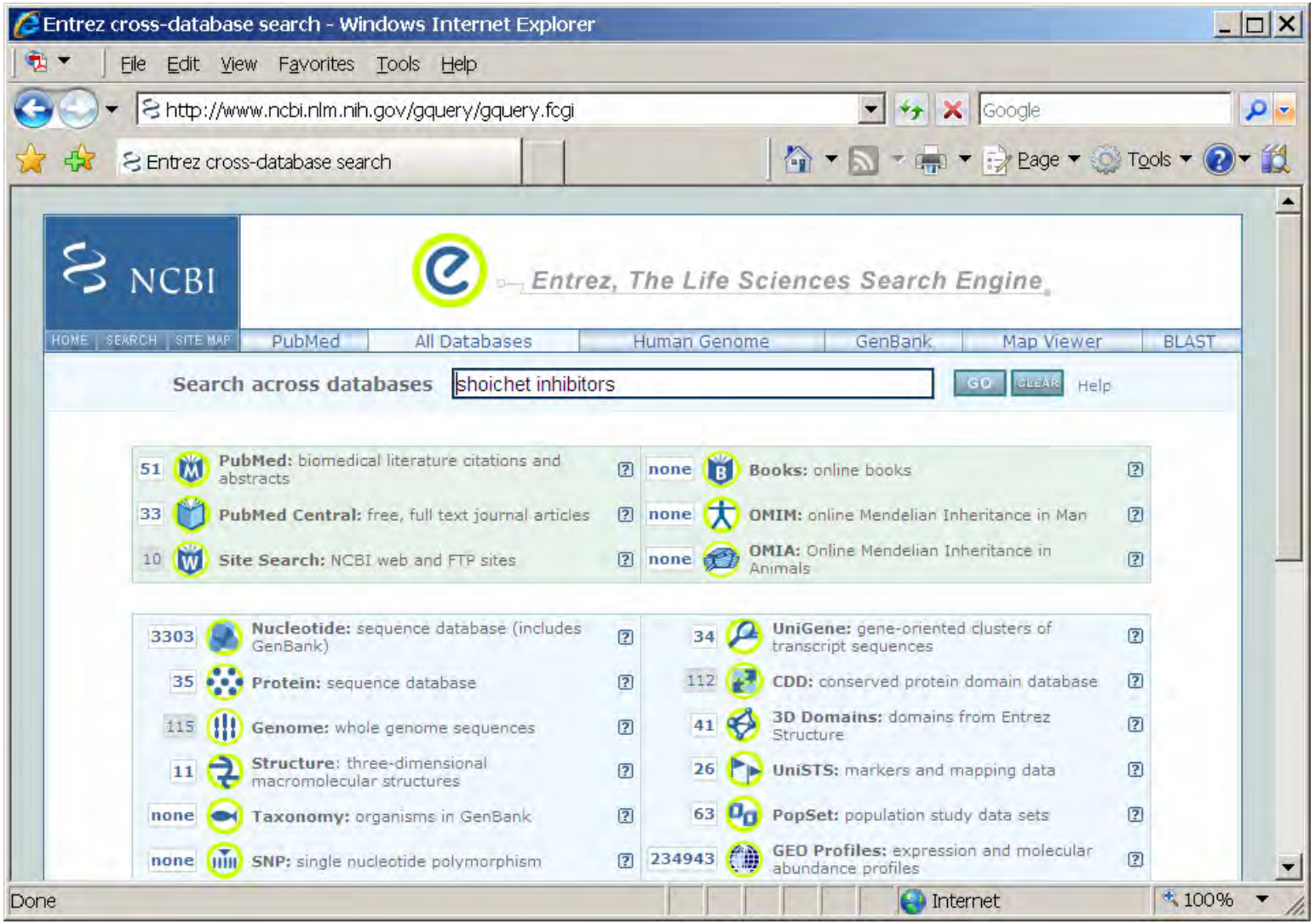


Entrez Links and Neighbors ...

1,500,000 users ...
50,000,000 hits ...
... per day



Search for “Shoichet inhibitors” ...



PubMed Article Retrieved ...

Entrez PubMed - Windows Internet Explorer

File Edit View Favorites Tools Help

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=

Entrez PubMed

NCBI

PubMed www.pubmed.gov

A service of the National Library of Medicine and the National Institutes of Health

My NCBI [Sign In] [Register]

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for Go Clear

Limits Preview/Index History Clipboard Details

Display AbstractPlus Show 20 Sort by Send to

All: 1 Review: 0

1: Nat Chem Biol. 2005 Aug;1(3):146-8. Epub 2005 Jul 3.

Comment in:
Nat Chem Biol. 2005 Aug;1(3):125.

High-throughput assays for promiscuous inhibitors.

Feng BY, Shelat A, Doman TN, Guy RK, Shoichet BK.

Department of Pharmaceutical Chemistry & Graduate Group in Chemistry and Chemical Biology, 1700 4th St., University of California San Francisco, San Francisco, California 94143-2550, USA.

Links
BioAssay
Substance via MeSH
Substance (Publisher)
Cited in PMC
LinkOut

Related Links

A common mec promiscuous inh [J Med Chem. 2002]

Identification and prediction of promiscuous aggre [J Med Chem. 2003]

Virtual screening to enrich hit lists from high-throughput screer [Proteins. 2003]

Kinase inhibitors: not just for kinases anymore. [J Med Chem. 2003]

Internet 100%

Link to PubChem Records ...

PubChem Substance - Windows Internet Explorer

File Edit View Favorites Tools Help

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?itool=pubmed_DocSum&db=pubm

PubChem Substance

Limits | Preview/Index | History | Clipboard | Details

Display Summary | Show 20 | Sort by | Send to

All: 29 | BioAssay: 0 | Protein3D: 0 | Rule of 5: 24

Items 1 - 20 of 29 Page 1 of 2 Next

About Entrez
Entrez Help

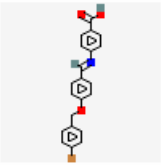
PubChem
Help | FAQ

PubChem Substance
Structures supplied by depositors

PubChem Compound
Unique structures with computed properties

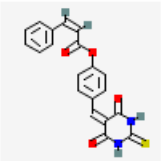
PubChem BioAssay

1: SID: 855510 Related Structures, Other Links



CID: 657179, nchembio718-comp29, (E)-4-(4-(4-bromobenzyloxy) benzylideneamino)benzoic acid
Source: Nature Chemical Biology(nchembio718-comp29)
IUPAC: 4-[[4-[(4-bromophenyl)methoxy]phenyl]methylideneamino] benzoic acid
MW: 410.261 | MF: C21H16BrNO3

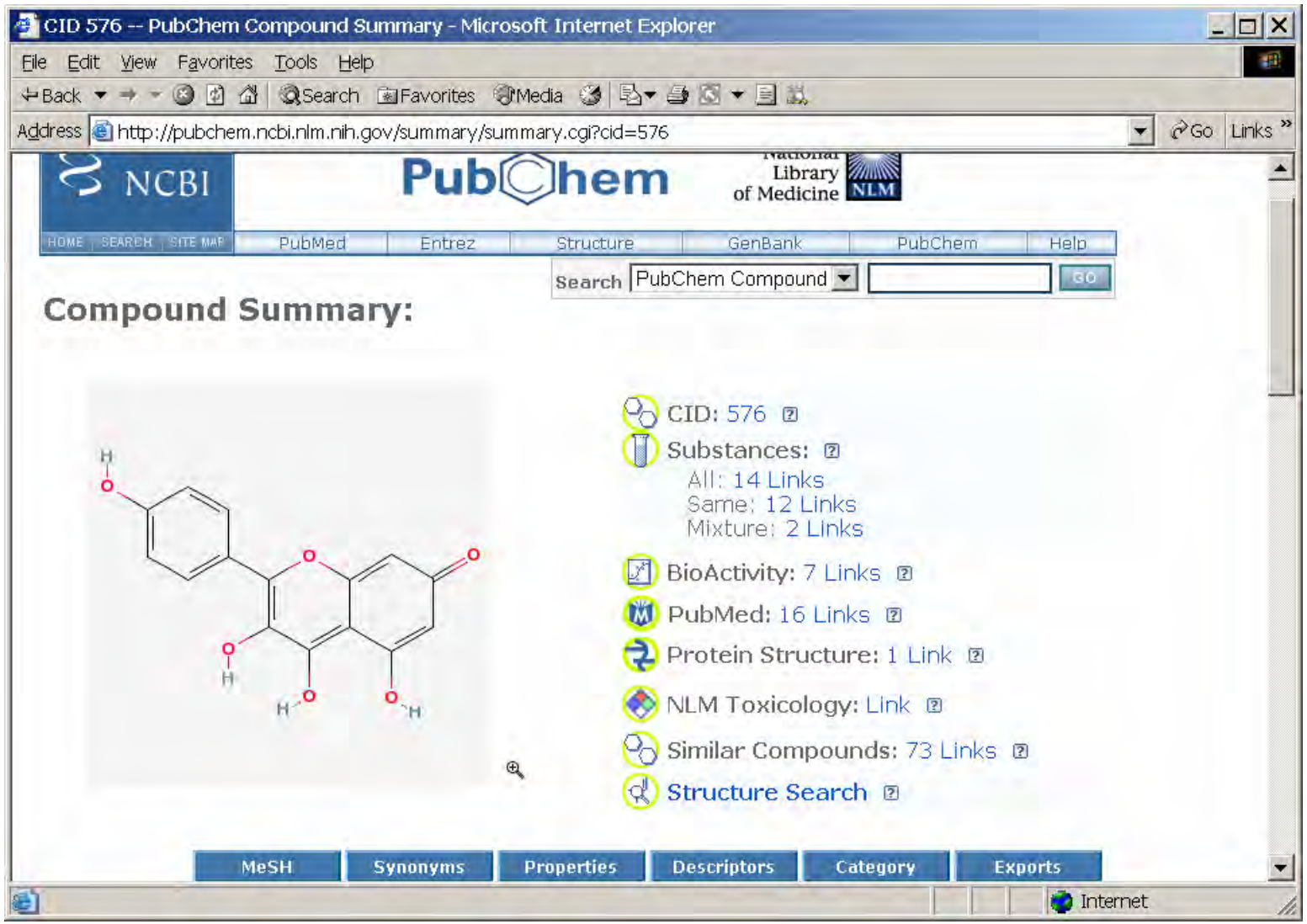
2: SID: 855509 Related Structures, Other Links



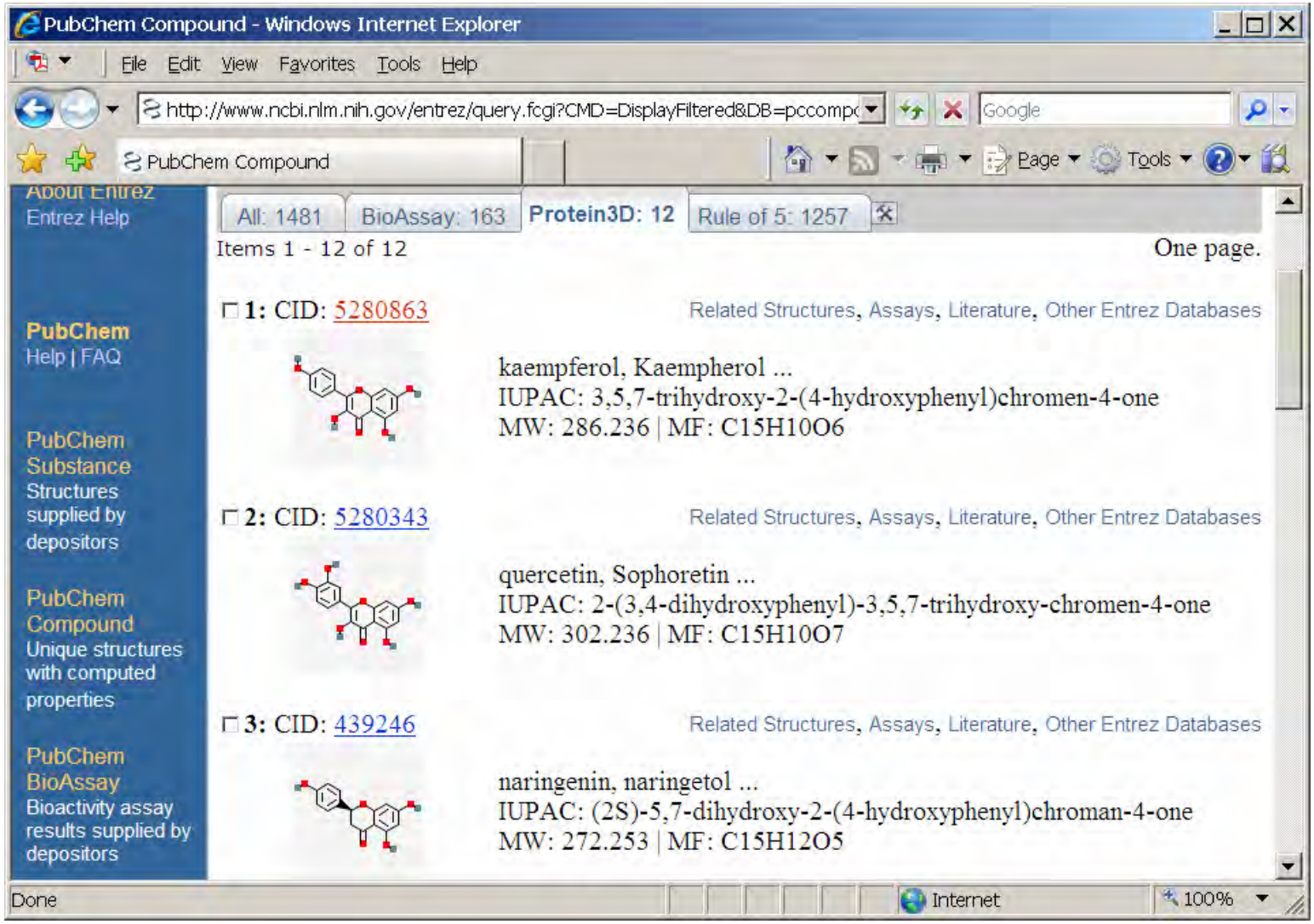
CID: 5388956, nchembio718-comp28, (Z)-4-((4,6-dioxo-2-thioxo-tetrahydropyrimidin-5(6H)-ylidene)methyl)phenyl 3-phenylacrylate
Source: Nature Chemical Biology(nchembio718-comp28)
IUPAC: [4-[(4,6-dioxo-2-sulfanylidene-1,3-diazinan-5-ylidene) methyl]phenyl] (Z)-3-phenylprop-2-enoate
MW: 378.402 | MF: C20H14N2O4S

Internet 100%

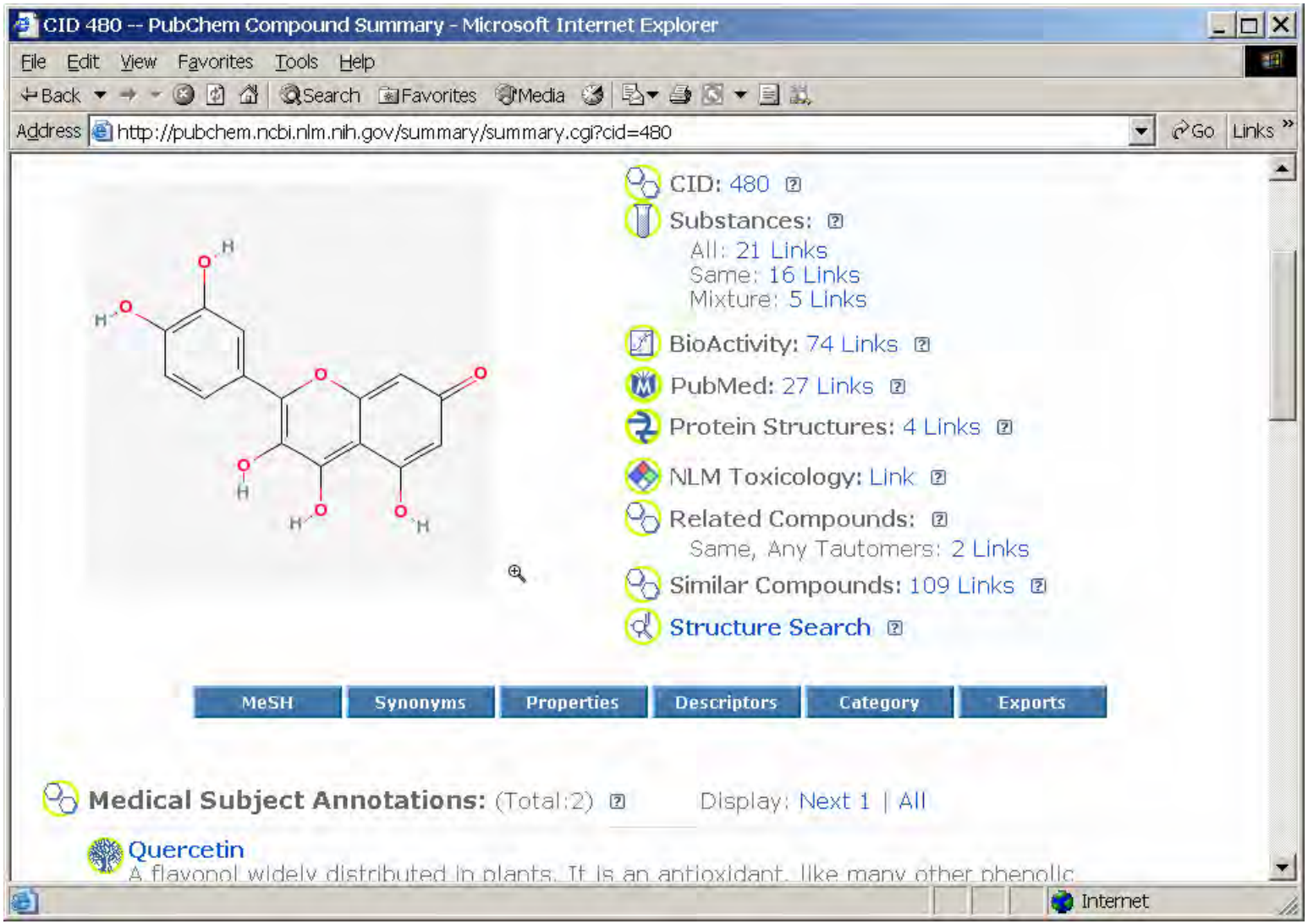
“Kaempferol” in PubChem ...



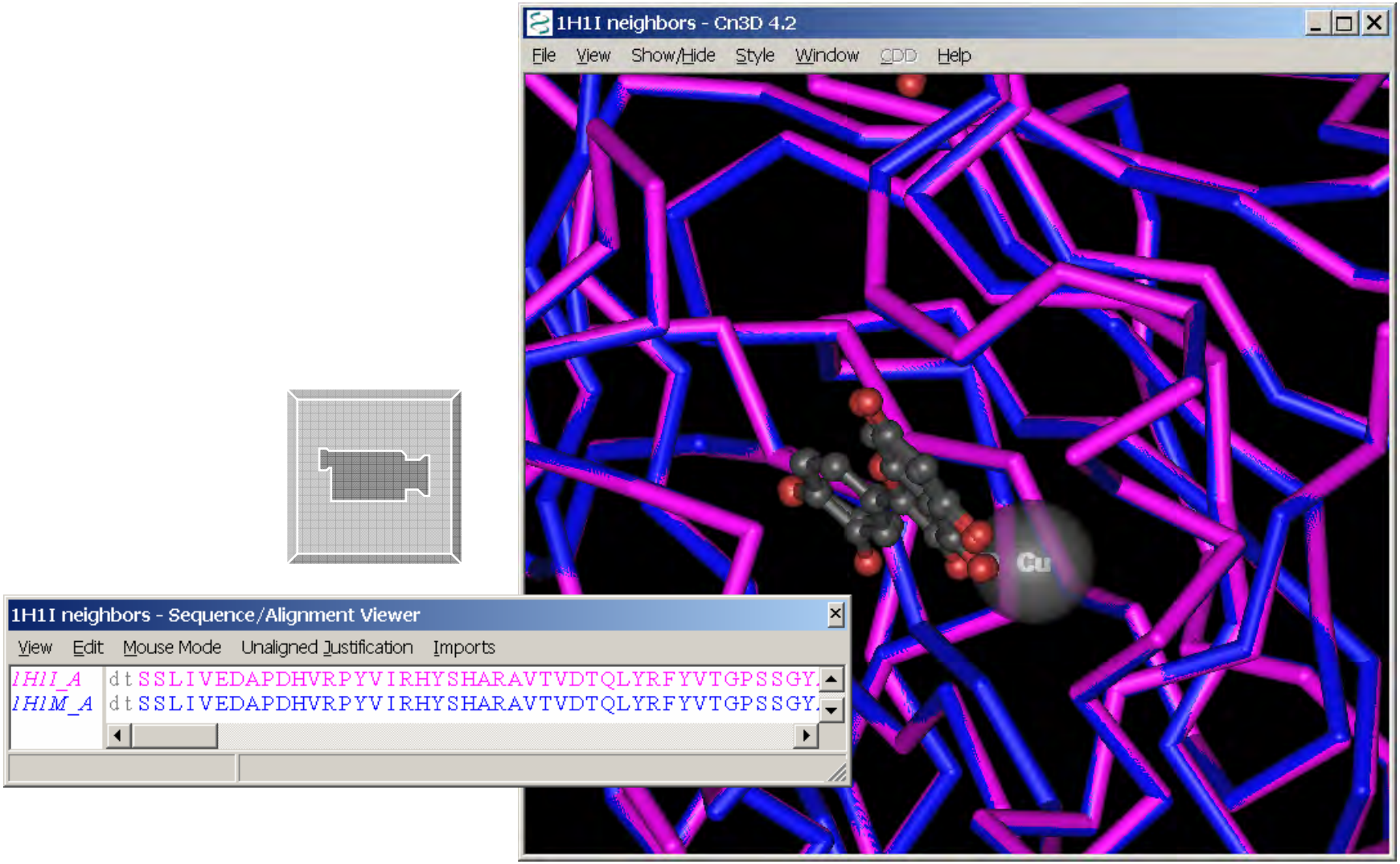
Similar Compounds in PubChem ...



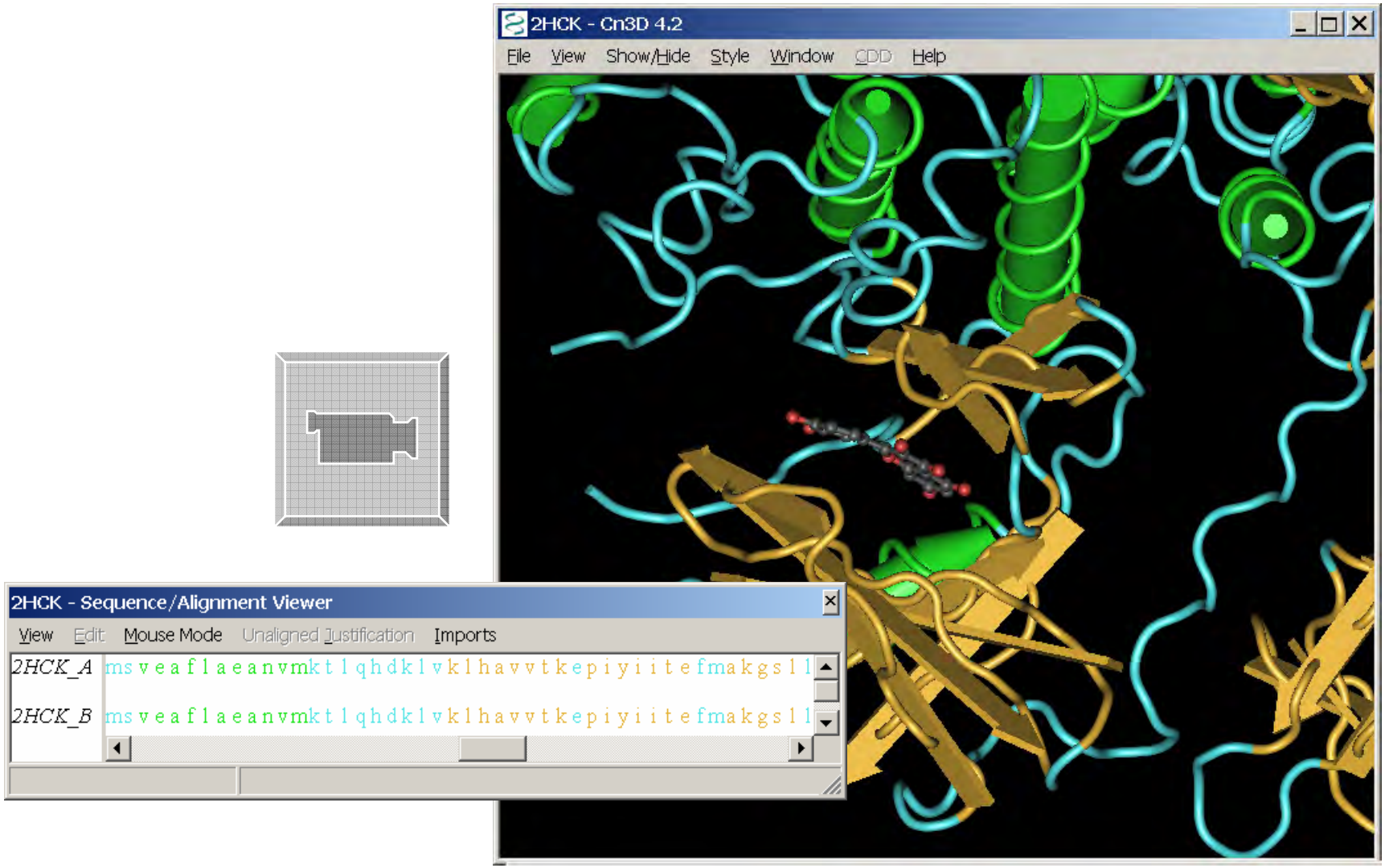
“Quercetin” in PubChem ...



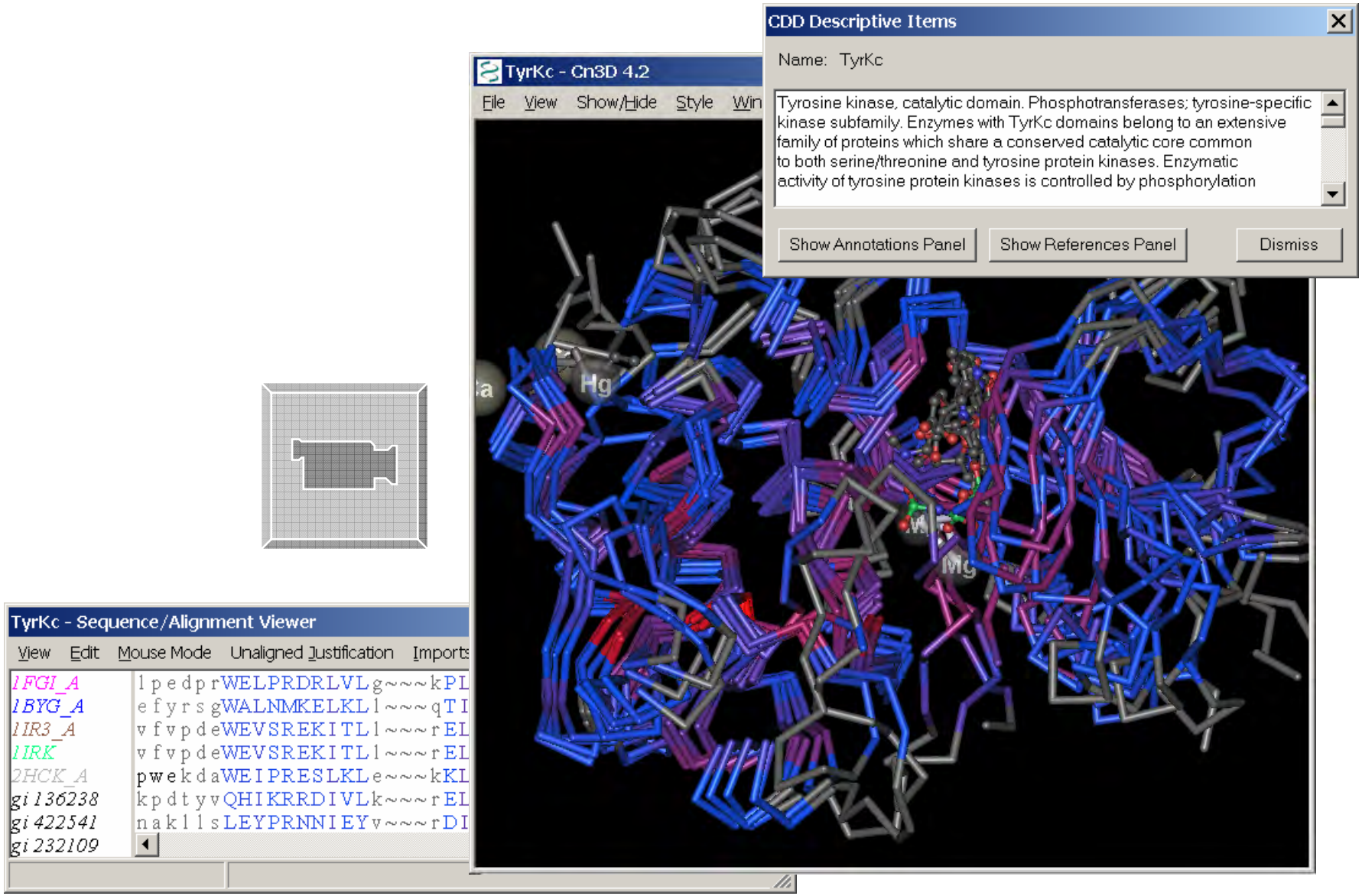
Compare Protein / Ligand Complexes ...



Link to Another Structure ...



Tyrosine Kinase Family Member ...



Links from “Quercetin” to PubMed ...

PubChem Compound Summary - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=480 Go Links

MeSH | Synonyms | Properties | Descriptors | Category | Exports

Medical Subject Annotations: (Total:2) Display: Next 1 | All

Quercetin
A flavonol widely distributed in plants. It is an antioxidant, like many other phenolic heterocyclic compounds. Glycosylated forms include RUTIN and quercetrin.

Show MeSH Tree Structure

PubMed via MeSH Choose by Subheadings:

administration and dosage	adverse effects	analogs and derivatives
analysis	antagonists and inhibitors	biosynthesis
blood	chemical synthesis	chemistry
diagnostic use	genetics	immunology
isolation and purification	metabolism	pharmacokinetics
pharmacology	physiology	radiation effects
secretion	standards	therapeutic use
toxicity	urine	

Depositor-Supplied Synonyms: (Total: 74)

Internet

Links from Compounds to PubMed ...

Entrez PubMed - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=pubmed&term=%22Quercetin%2fantagonists%20and

Go Links

PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI (Cubby)

Related Resources
Order Documents
NLM Mobile
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

2: Lin HC, Cheng TH, Chen YC, Juan SH. Related Articles, Links

Mechanism of heme oxygenase-1 gene induction by quercetin in rat aortic smooth muscle cells.
Pharmacology. 2004 Jun;71(2):107-12.
PMID: 15118350 [PubMed - indexed for MEDLINE]

3: Ko WC, Chen MC, Wang SH, Lai YH, Chen JH, Lin CN. Related Articles, Links

3-O-methylquercetin more selectively inhibits phosphodiesterase subtype 3.
Planta Med. 2003 Apr;69(4):310-5.
PMID: 12709896 [PubMed - indexed for MEDLINE]

4: Revuelta MP, Cantabrana B, Hidalgo A. Related Articles, Links

Mechanisms involved in kaempferol-induced relaxation in rat uterine smooth muscle.
Life Sci. 2000 Jun 8;67(3):251-9.
PMID: 10983869 [PubMed - indexed for MEDLINE]

5: Ciolino HP, Daschner PJ, Yeh GC. Related Articles, Links

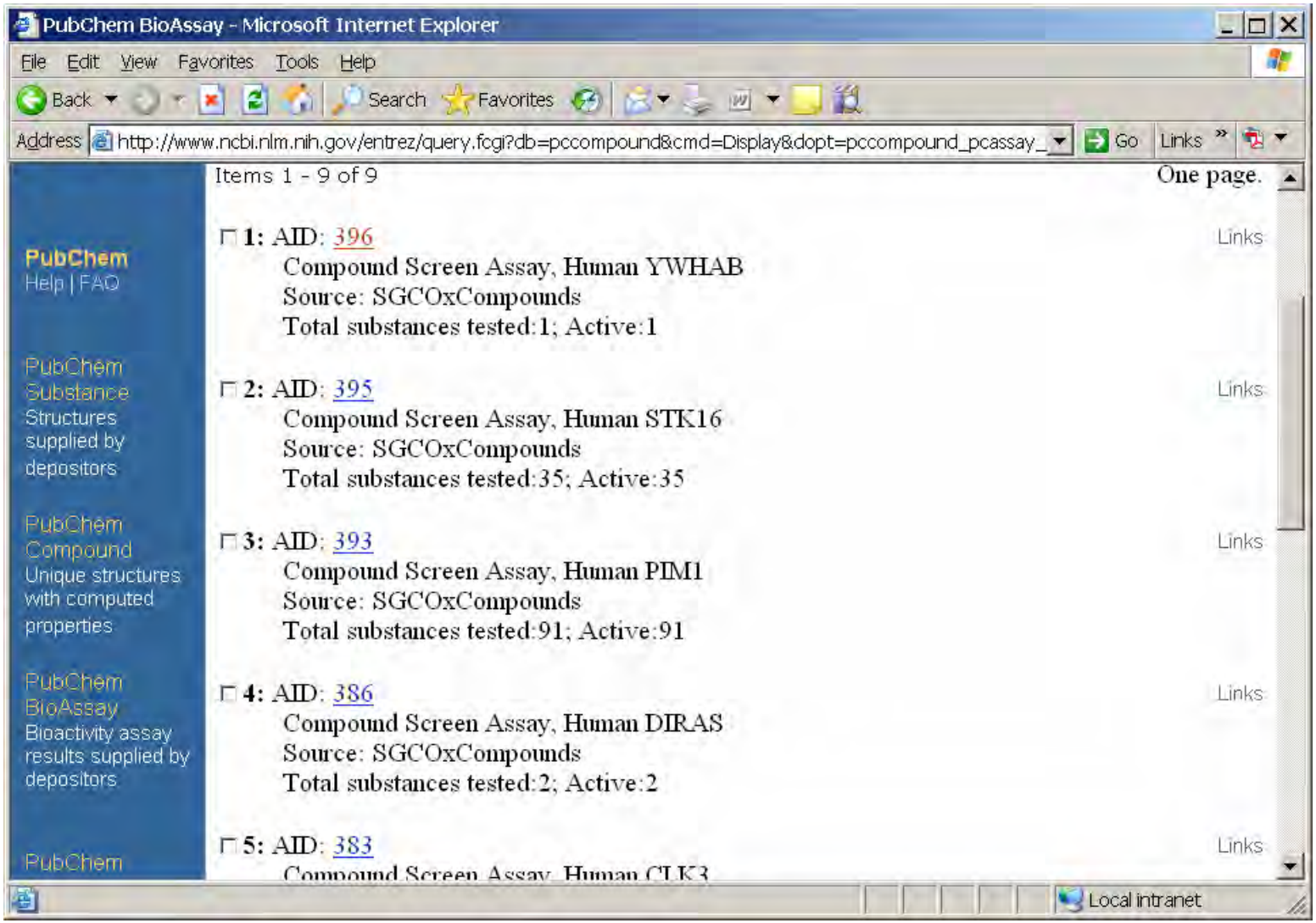
Dietary flavonols quercetin and kaempferol are ligands of the aryl hydrocarbon receptor that affect CYP1A1 transcription differentially.
Biochem J. 1999 Jun 15;340 (Pt 3):715-22.
PMID: 10359656 [PubMed - indexed for MEDLINE]

6: Lagarrigue S, Chaumontet C, Heberden C, Martel P, Gaillard-Sanchez I. Related Articles, Links

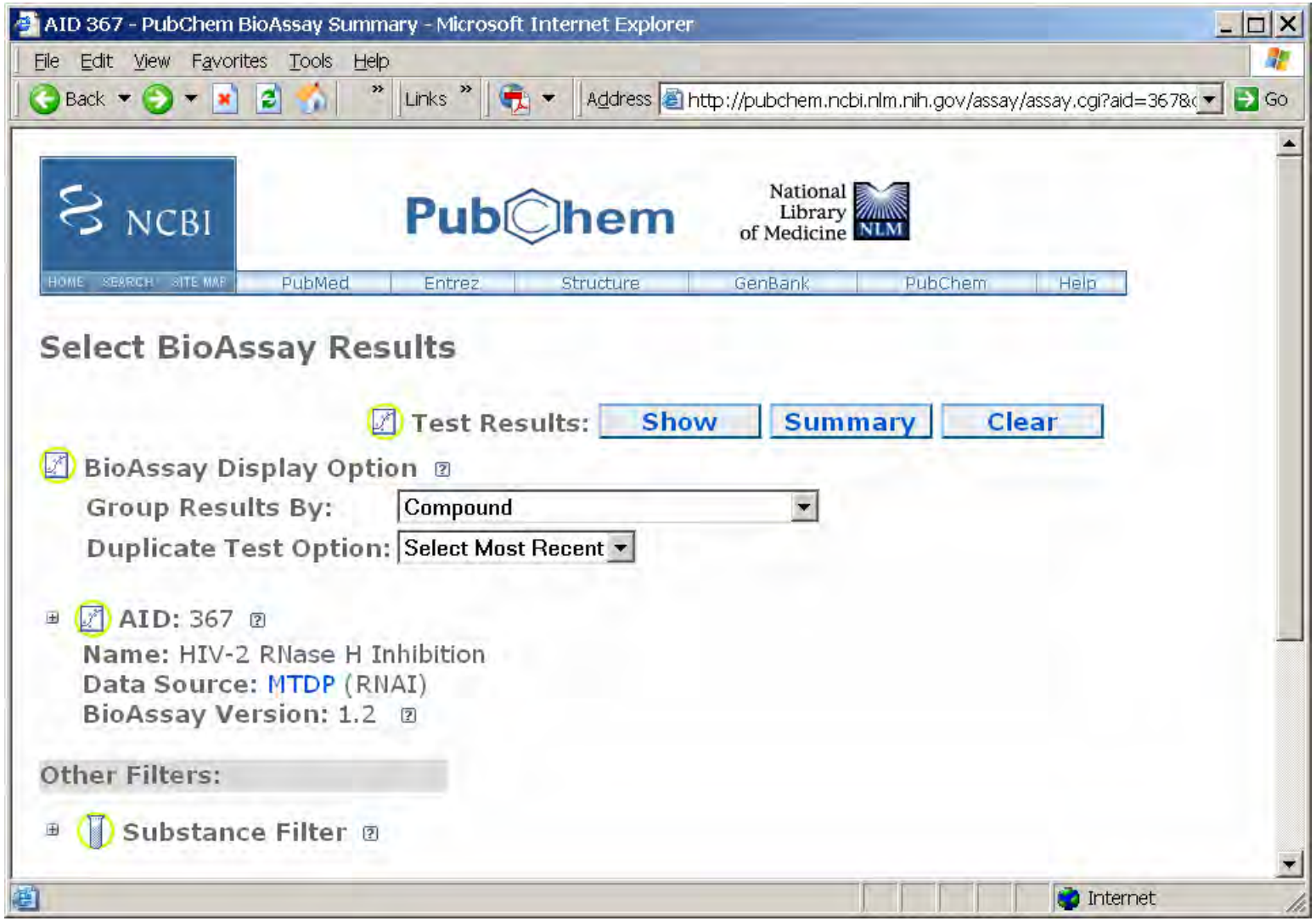
Suppression of oncogene-induced transformation by quercetin and retinoic acid in rat liver epithelial cells.
Cell Mol Biol Res. 1995;41(6):551-60.

Internet

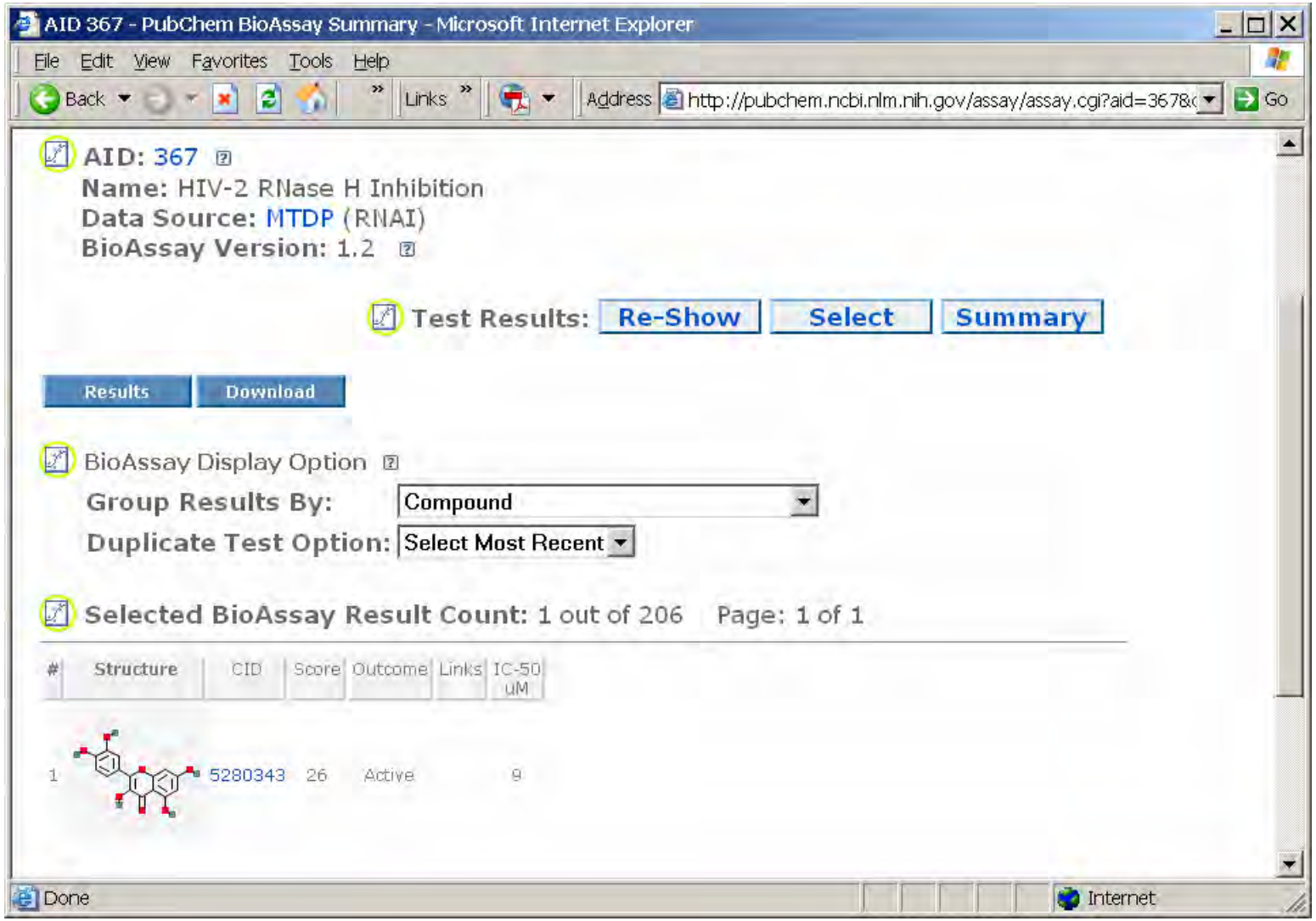
Link to BioAssays where “Active” ...



BioAssay where “Active” ...

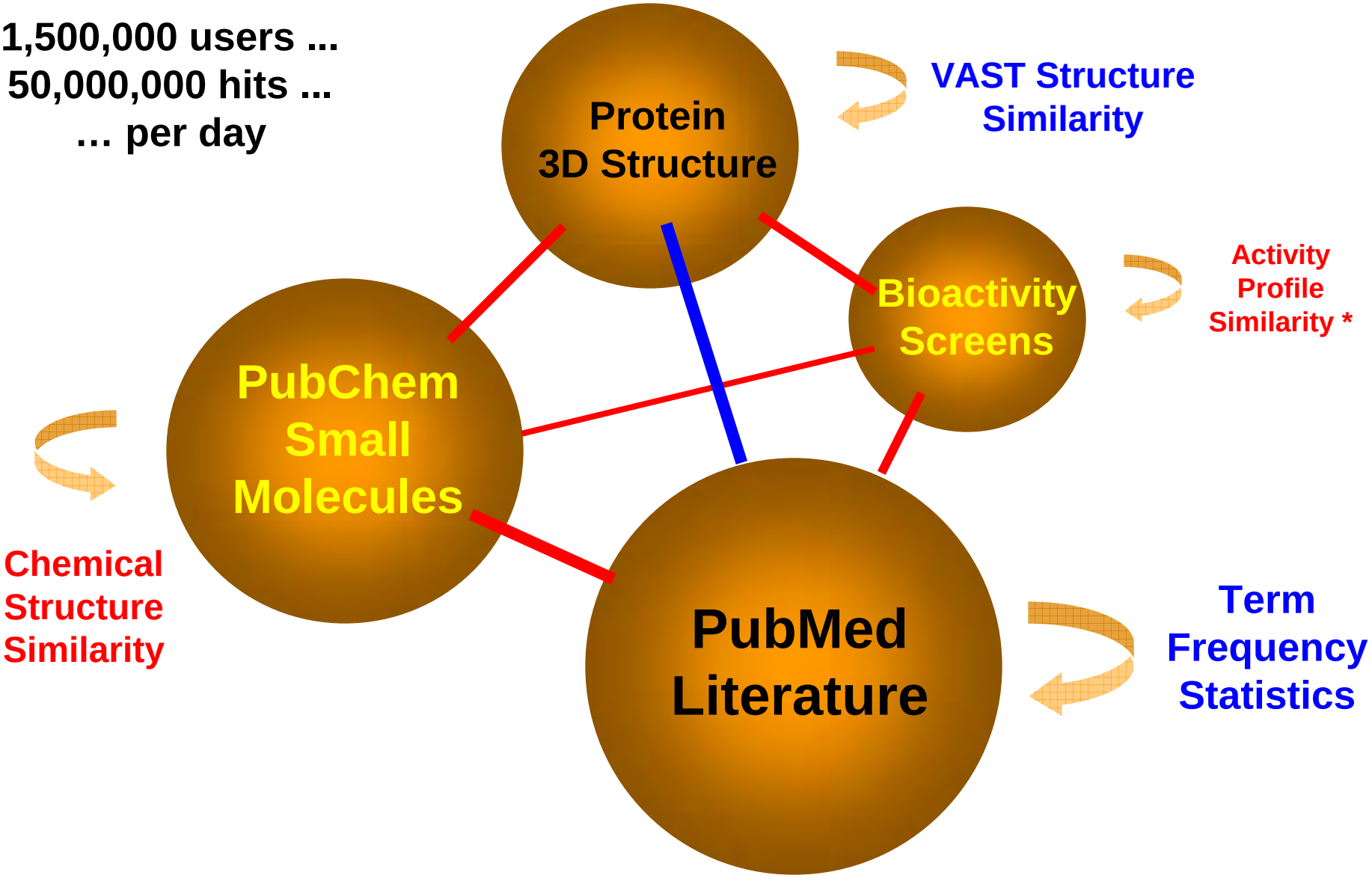


BioAssay where “Active” ...



Entrez Links and Neighbors ...

1,500,000 users ...
50,000,000 hits ...
... per day



PubChem Retrieval System …

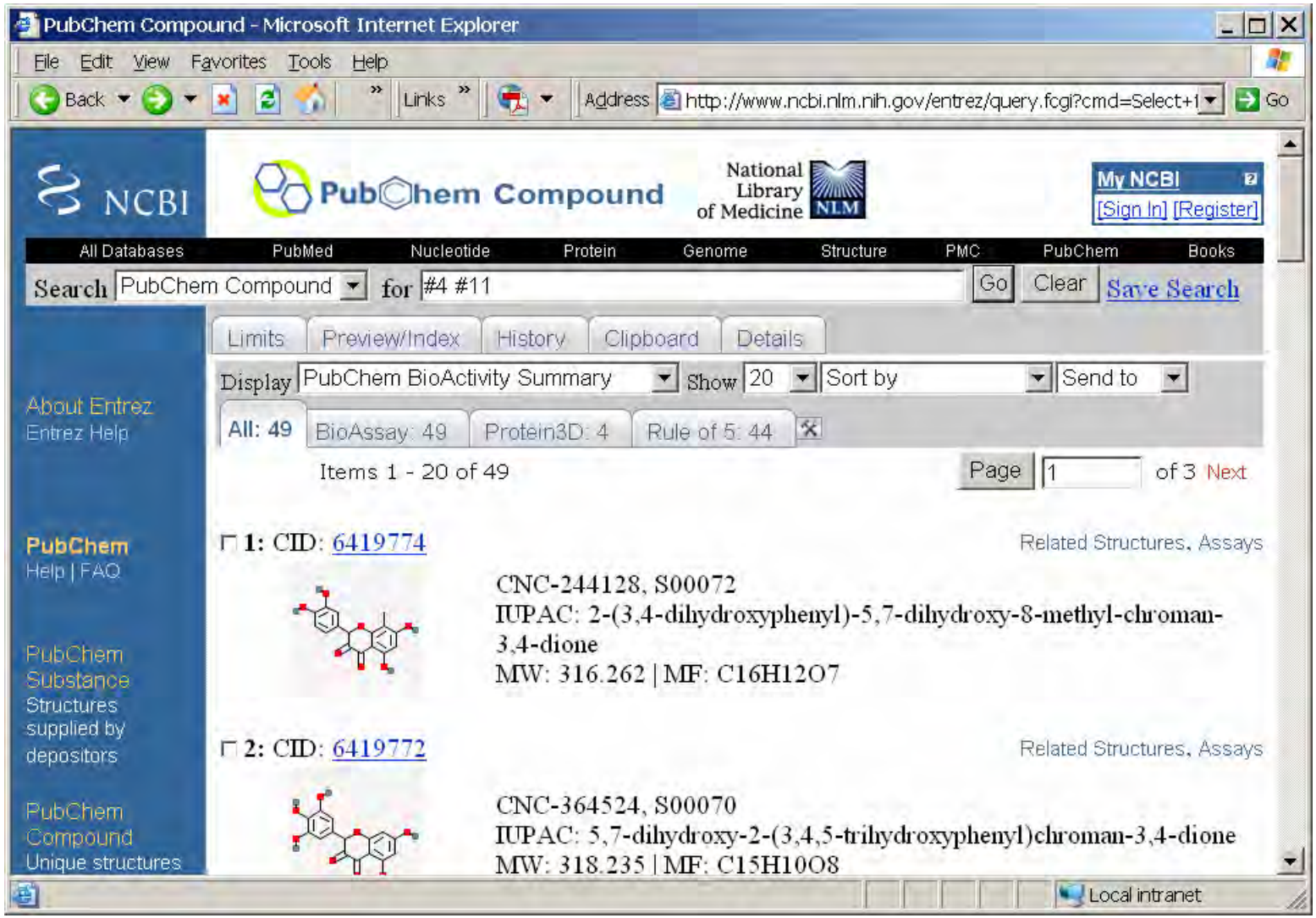
… Optimize “discoverability” for molecular biologists by integrating PubChem into NCBI’s Entrez / PubMed Search Engine

… Chemical structure search

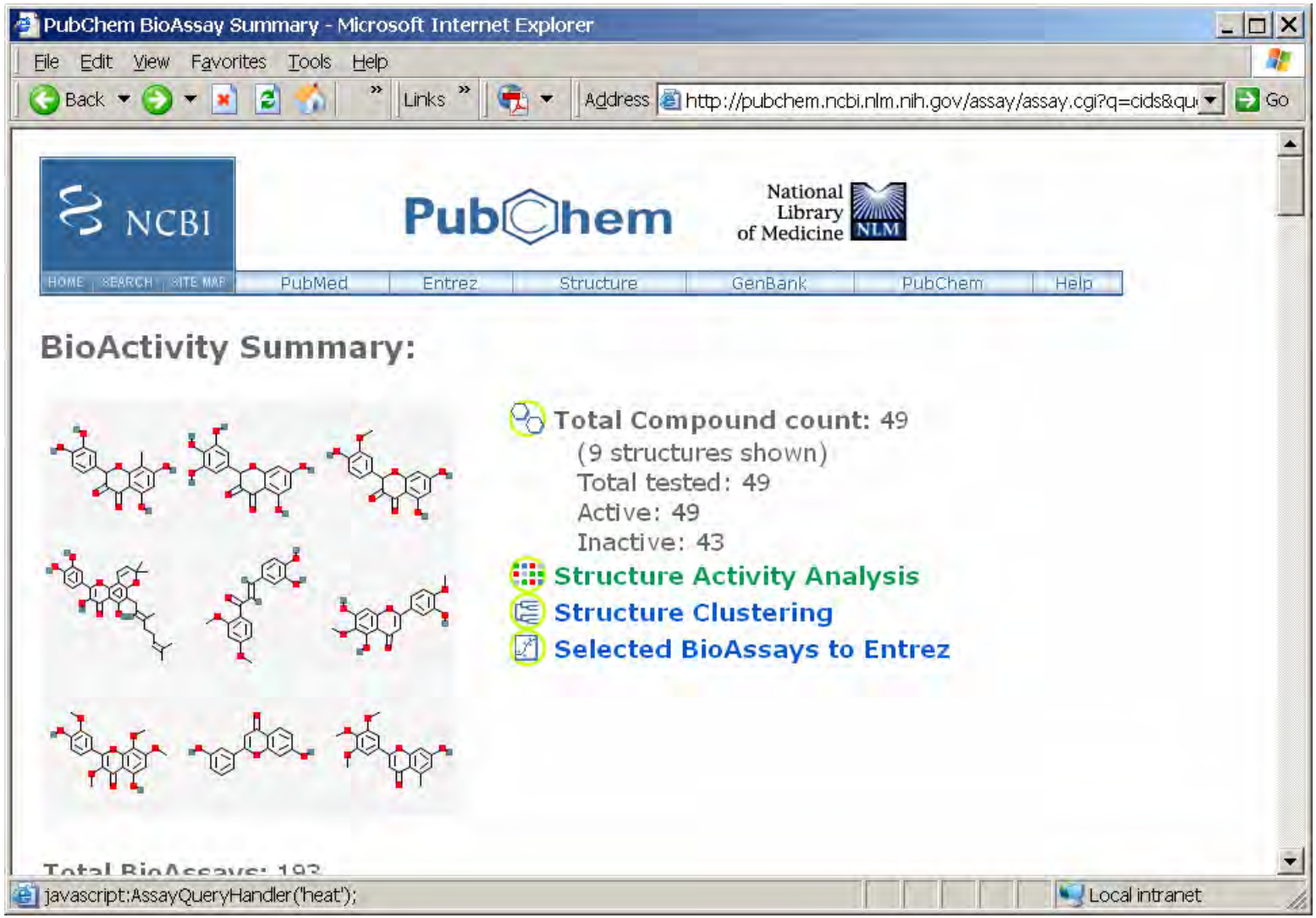
… Bioassay result search

… Exploratory structure-activity tools

Quercetin Structure-Activity ...



PubChem Bioactivity Summary ...



PubChem Bioactivity Summary ...

PubChem BioAssay Summary - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Links Address http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?q=cids&qu Go

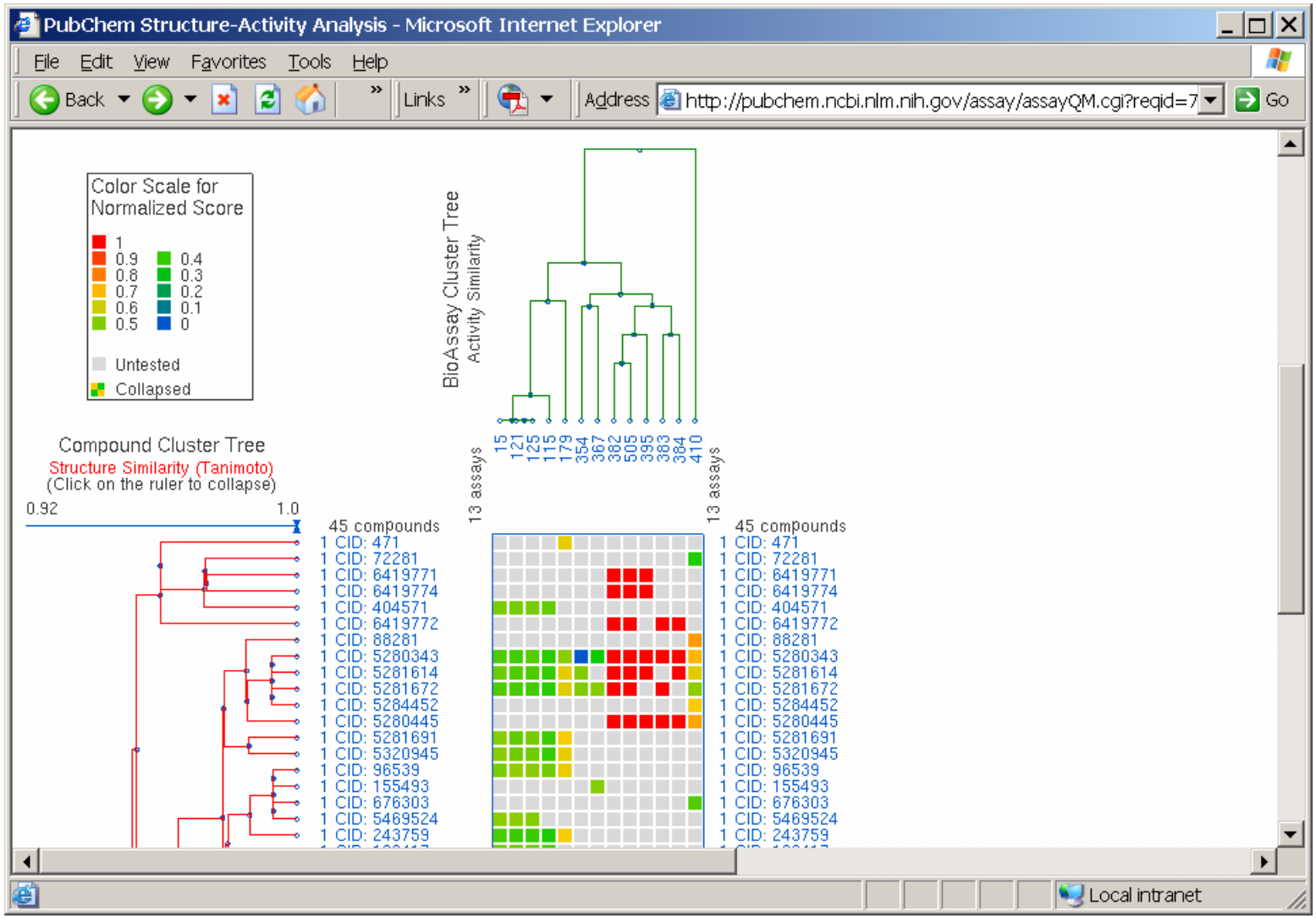
Total BioAssays: 193
Total Pages: : 2

Display: 100 Go To Page 1

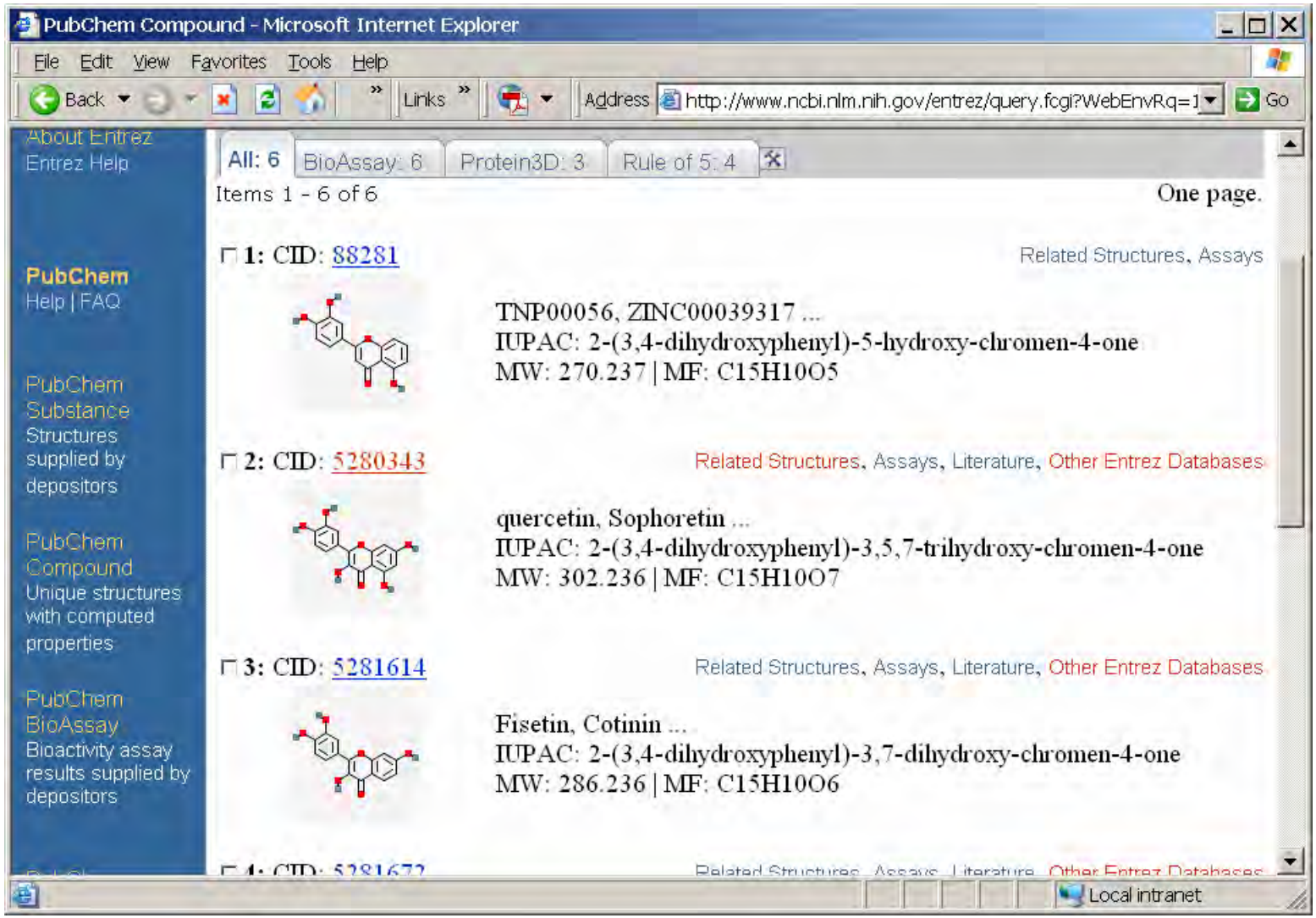
#	AID	Active	Inactive	Total Tested	Data	Name
1	410	22	1	23	Data	p450-cyp1a2
2	382	11		11	Data	Compound Screen Assay, Human CLK1
3	505	9		9	Data	Compound Screen Assay, Human PIM2
4	395	8		8	Data	Compound Screen Assay, Human STK16
5	354	7	2	9	Data	NCTR Estrogen Receptor Binding Database (NCTRER)
6	383	5		5	Data	Compound Screen Assay, Human CLK3
7	115	5	12	17	Data	NCI human tumor cell line growth inhibition assay. Data for the SR Leukemia cell line
8	384	4		4	Data	Compound Screen Assay, Human CSNK1G2
9	367	4		4	Data	HIV-2 RNase H Inhibition
10	179	4	15	19	Data	NCI AIDS Antiviral Assay
11	125	4	16	20	Data	NCI human tumor cell line growth inhibition assay. Data for the HL-60(TB) Leukemia cell line

Local intranet

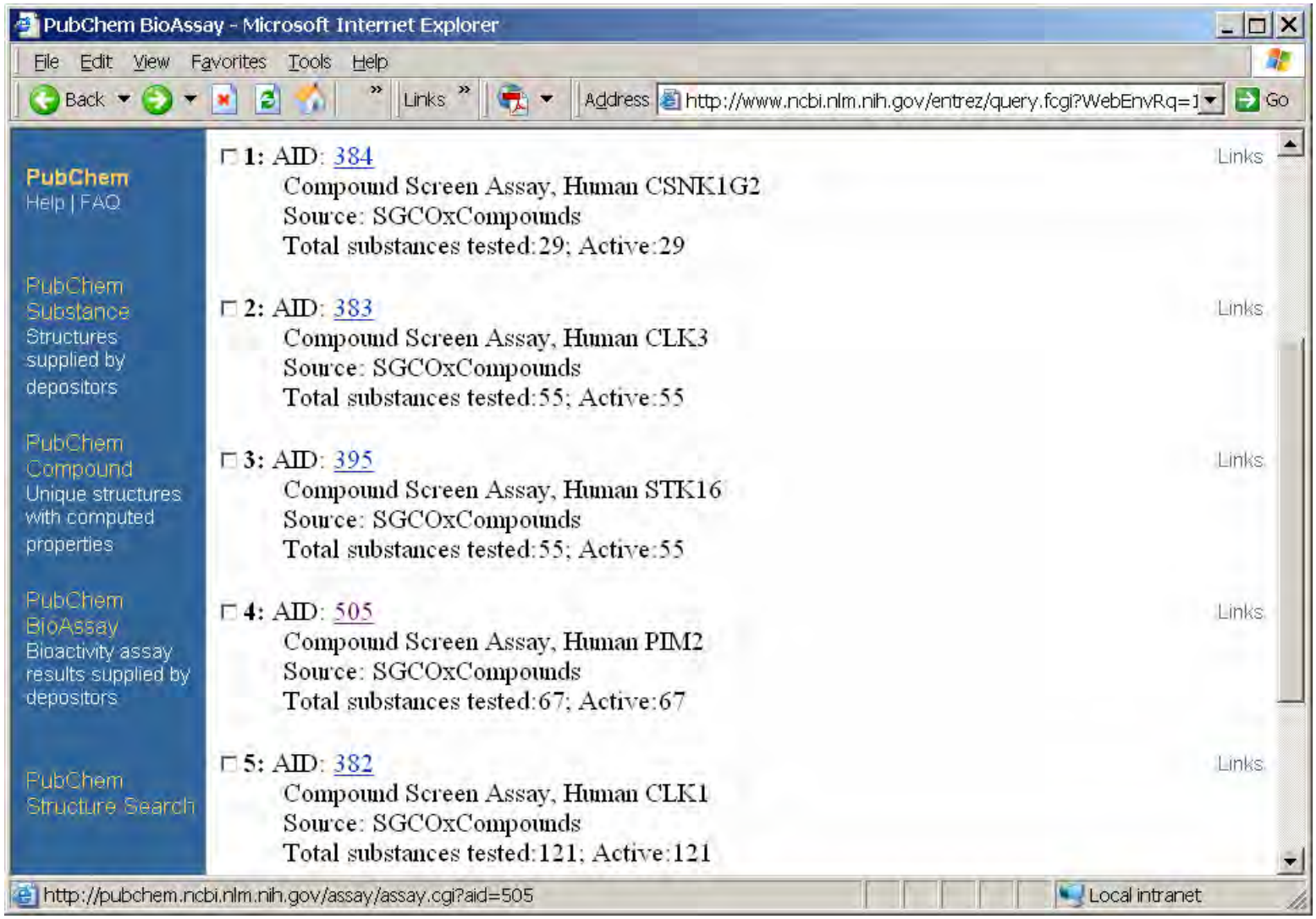
PubChem Structure-Activity ...



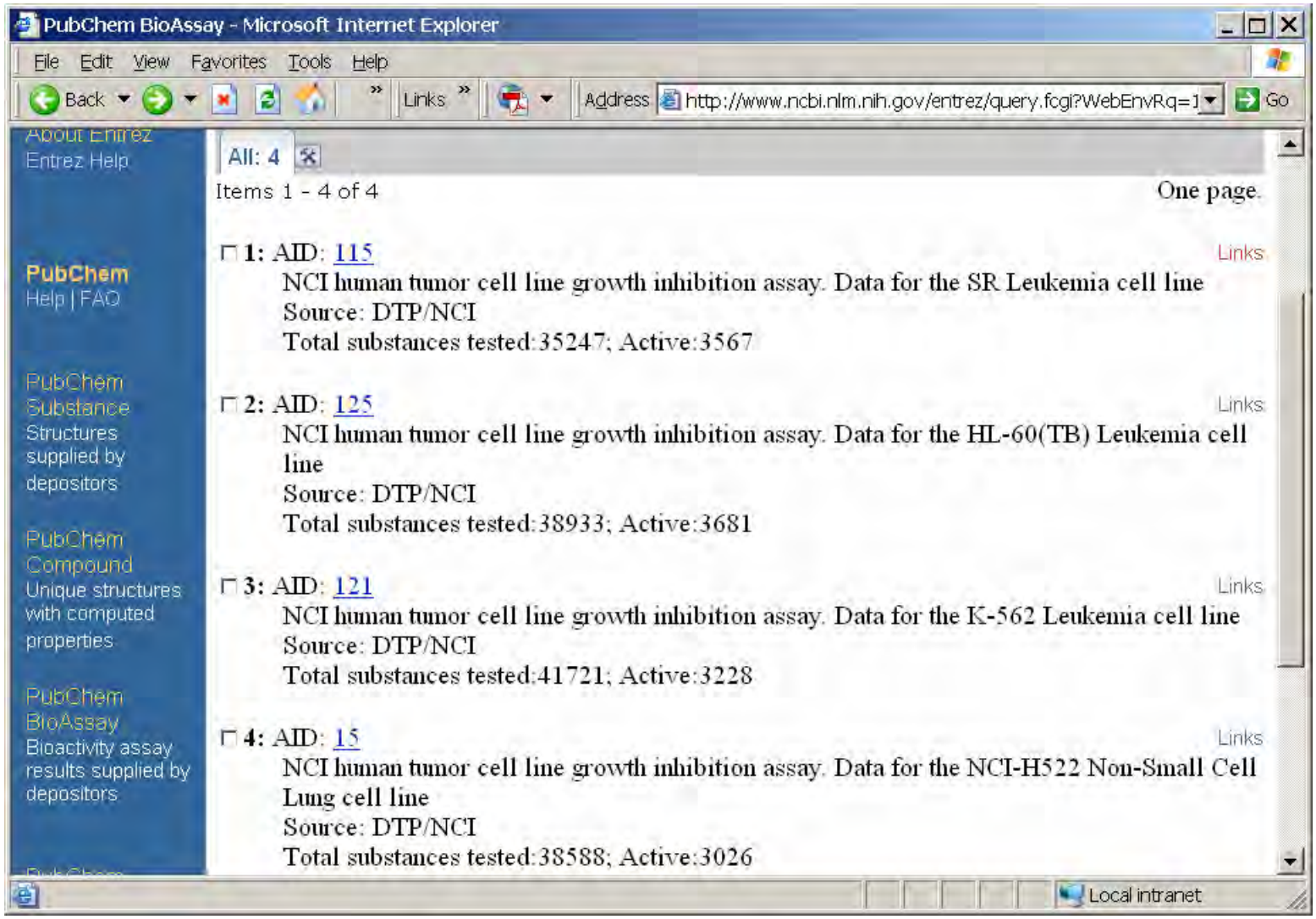
Active Compound Cluster ...



BioAsay Cluster ...

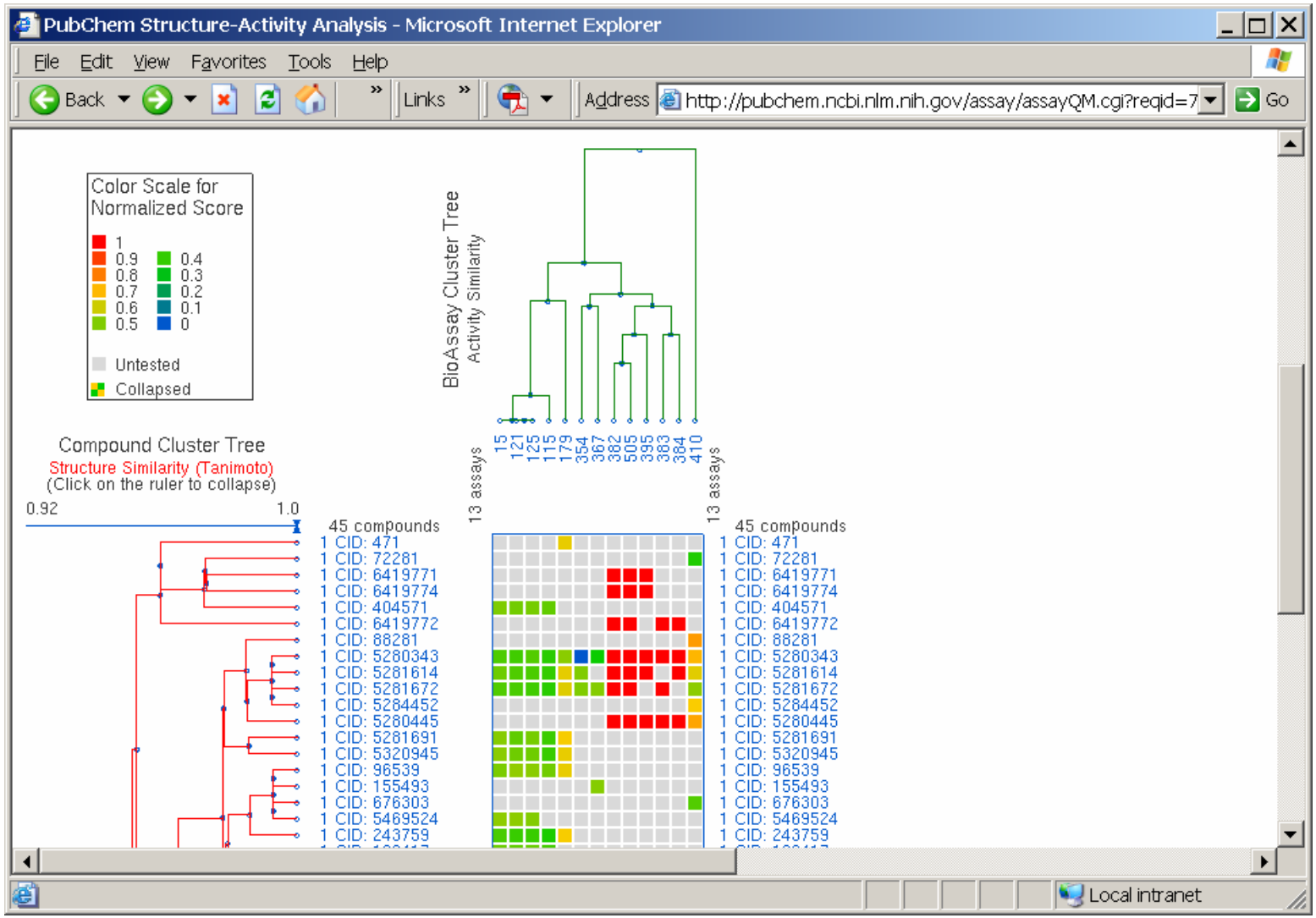


Another BioAssay Cluster ...

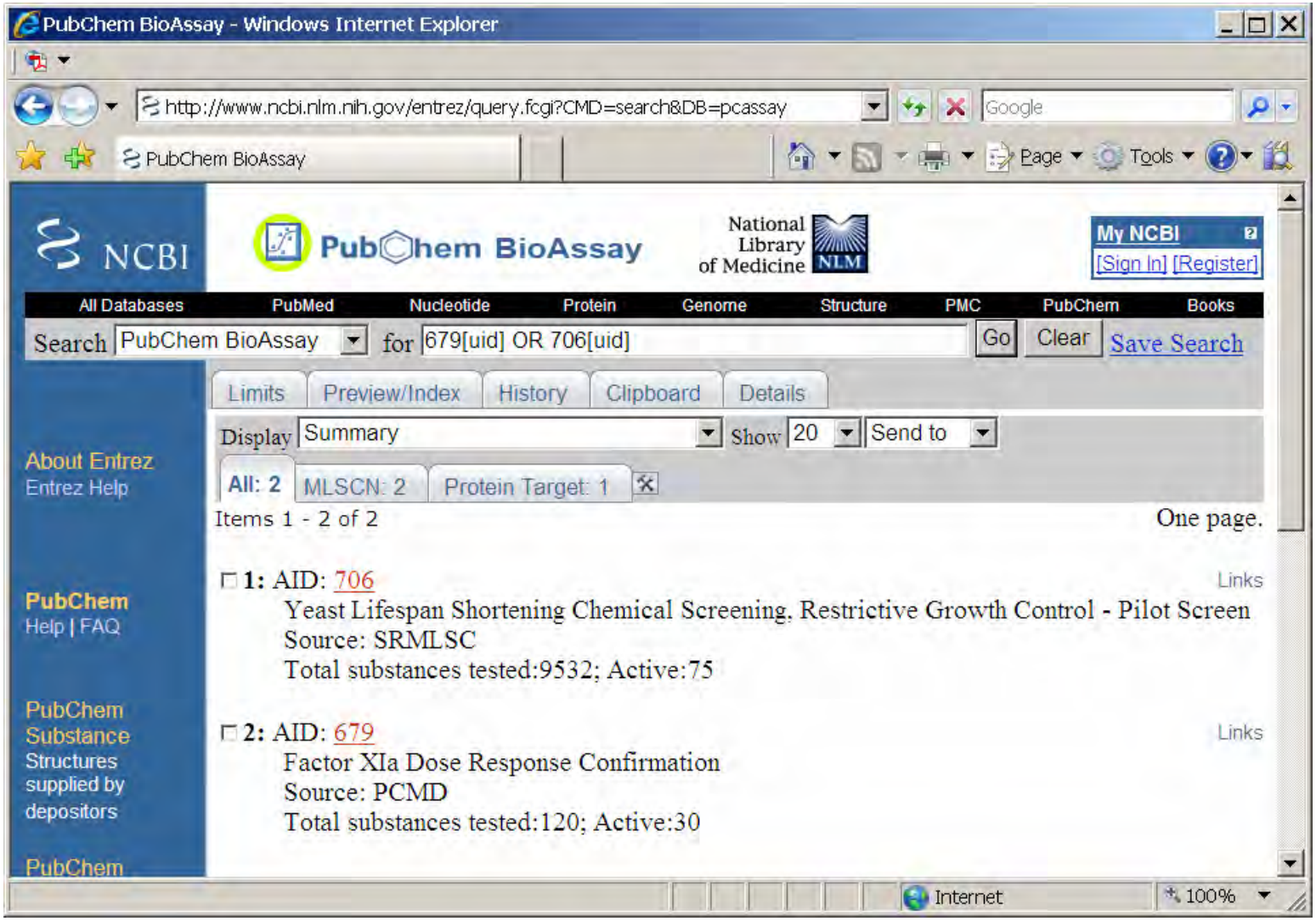


PubMed Connection ...

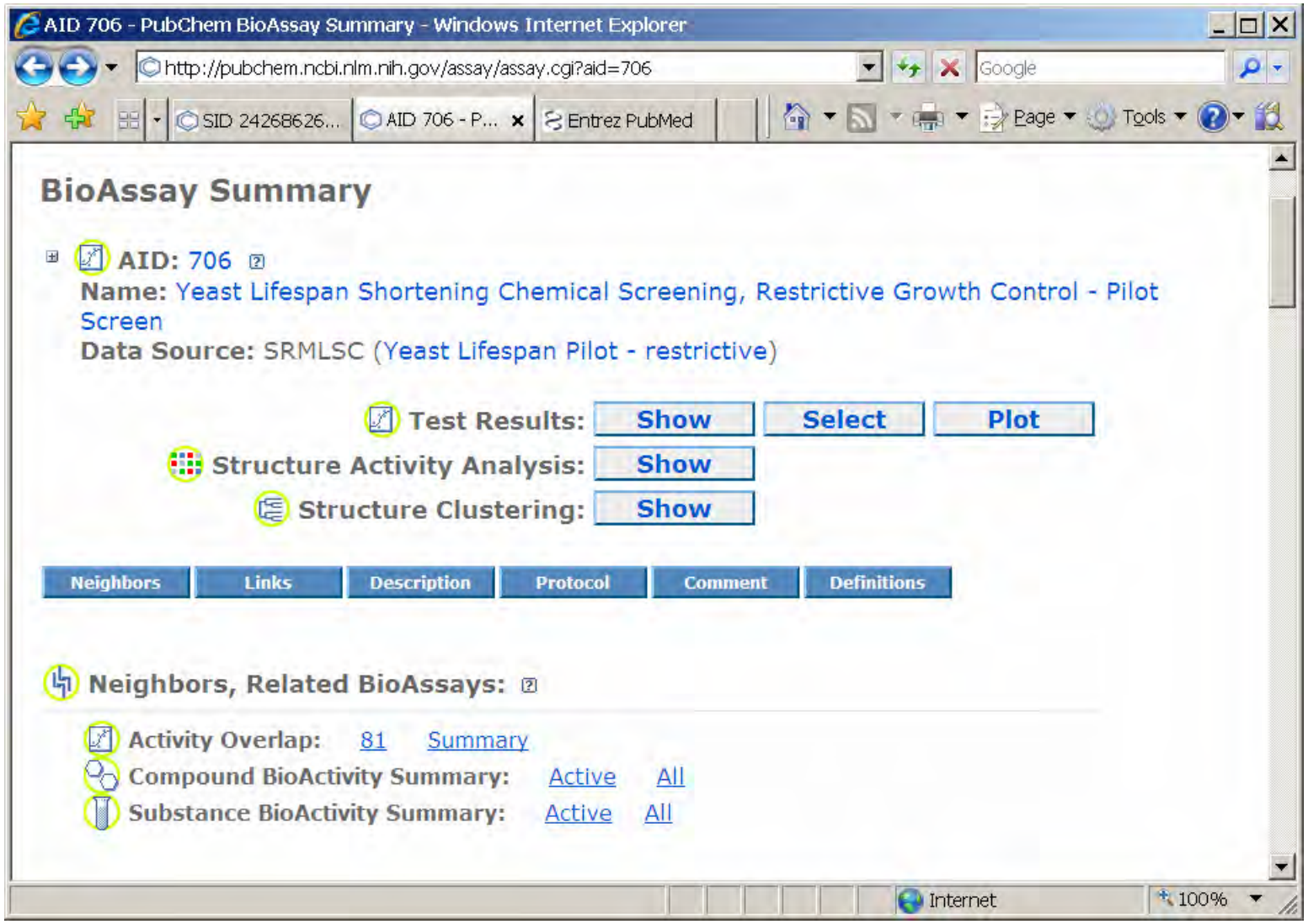
PubChem Structure-Activity ...



PubChem BioAssay Tools Demo ...



Yeast Replicative Lifespan Assay ...



BioAssay Description ...

AID 706 - PubChem BioAssay Summary - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=706

SID 24268626... | AID 706 - P... | Entrez PubMed

Description:

Southern Research Molecular Libraries Screening Center (SRMLSC)
Southern Research Institute (Birmingham, Alabama)
NIH Molecular Libraries Screening Centers Network (MLSCN)
Assay Provider: Dr. David S. Goldfarb, University of Rochester
Award: R03 MH076395-01

There is now solid evidence for the early evolution of conserved pathways for aging. These pathways may allow eukaryotic cells and animals to postpone reproduction in unfavorable environmental conditions. Key elements of public cellular mechanisms that extend or shorten lifespan, including the role of sirtuins in lifespan, were first discovered in Saccharomyces cerevisiae, which provides a strong rationale for the use of yeast as a model genetic system for studying aging. For example, lifespan extension by caloric restriction occurs in both yeast and rodents.

Yeast replicative lifespan is the number of times a mother cell replicates before she senesces and dies. The replicative lifespan of a yeast strain is described by the mean or median lifespan of a cohort of mother cells, which can vary widely among laboratory strains, but is normally between 20-25 generations. The clock for daughters is generally reset to zero although daughters of older mothers, which replicate more slowly, have reduced lifespans. The genetic program(s) that sets the clock, and the cellular mechanisms that respond to environmental cues to extend lifespan, such as caloric restriction, are poorly understood.

Here we use a genetically modified version of a high throughput replicative lifespan assay called the DeaD assay (1). Under permissive conditions, in galactose-containing medium, the culture divides exponentially. Under restrictive conditions, in glucose medium, the daughters show a great propensity to die, and the saturation point of the culture is limited by the lifespan of the of mother cells rather than nutrient limitation. Compounds that

Internet | 100%

PubMed search for “Sirtuins” ...

Entrez PubMed - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=

NCBI PubMed www.pubmed.gov

A service of the National Library of Medicine and the National Institutes of Health

My NCBI [Sign In] [Register]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals | Books

Search PubMed for Go Clear

Limits | Preview/Index | History | Clipboard | Details

Display AbstractPlus Show 20 Sort by Send to

All: 1 Review: 1

1: Trends Mol Med. 2007 Feb;13(2):64-71. Epub 2007 Jan 4. ELSEVIER FULL-TEXT ARTICLE Links

SIR2: a potential target for calorie restriction mimetics.

Chen D, Guarente L.

Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.

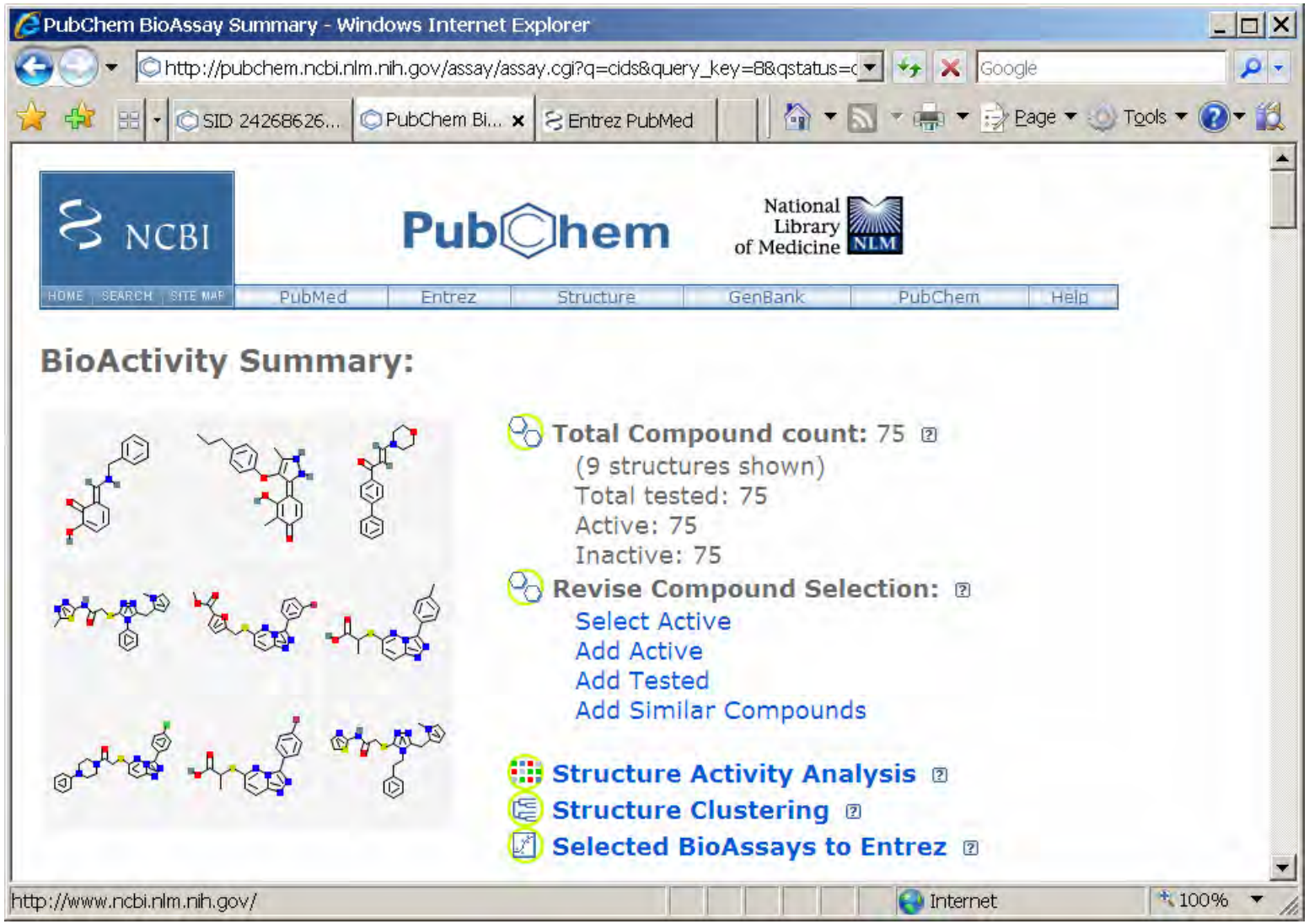
Calorie restriction (CR) extends lifespan in a wide variety of species and mitigates diseases of aging in mammals. Here, we describe the evidence that the silent information regulator 2 (SIR2) gene, which encodes a nicotinamide adenine dinucleotide (NAD)-dependent deacetylase, regulates lifespan and mediates CR in lower species such as Saccharomyces cerevisiae and Caenorhabditis elegans. We discuss

Related Links

- Sirtuin-independent effects of nicotinamide on lifespa [Aging Cell. 2006]
- HST2 mediates SIR2-independent life-span extension by calori [Science. 2005]
- Small molecule activators of sirtuins extend Saccharomyces c [Nature. 2003]
- Sirtuin activators mimic caloric restriction and delay age [Nature. 2004]
- Nicotinamide and PNC1 govern lifespan extension by calorie rest

Internet 100%

Bioactivity Summary for Actives ...



Bioactivity Summary for Actives ...

PubChem BioAssay Summary - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?q=cids&query_key=8&qstatus=c

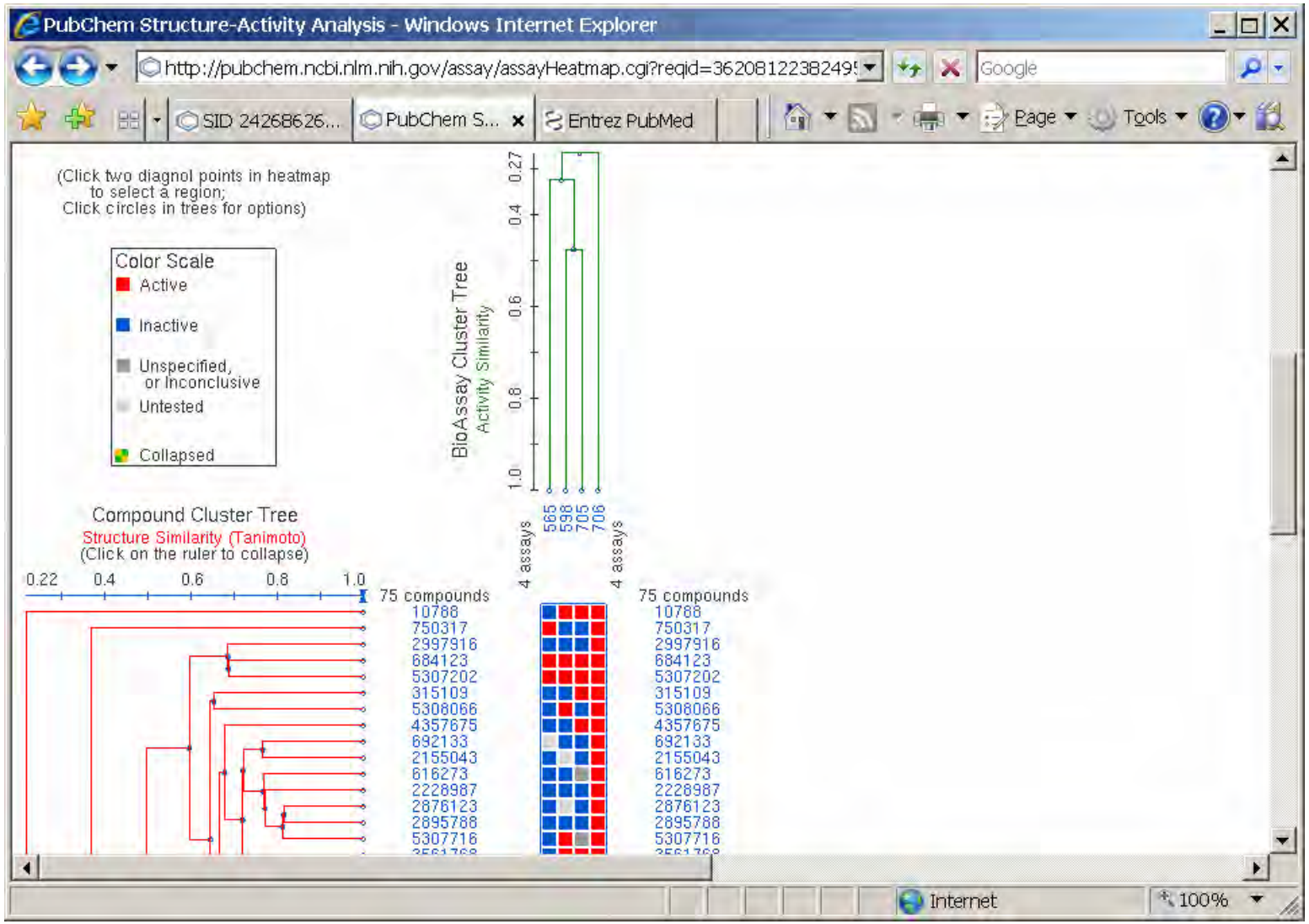
Total BioAssays: 203
Total Pages: : 3

Display: 100 Go To Page 1

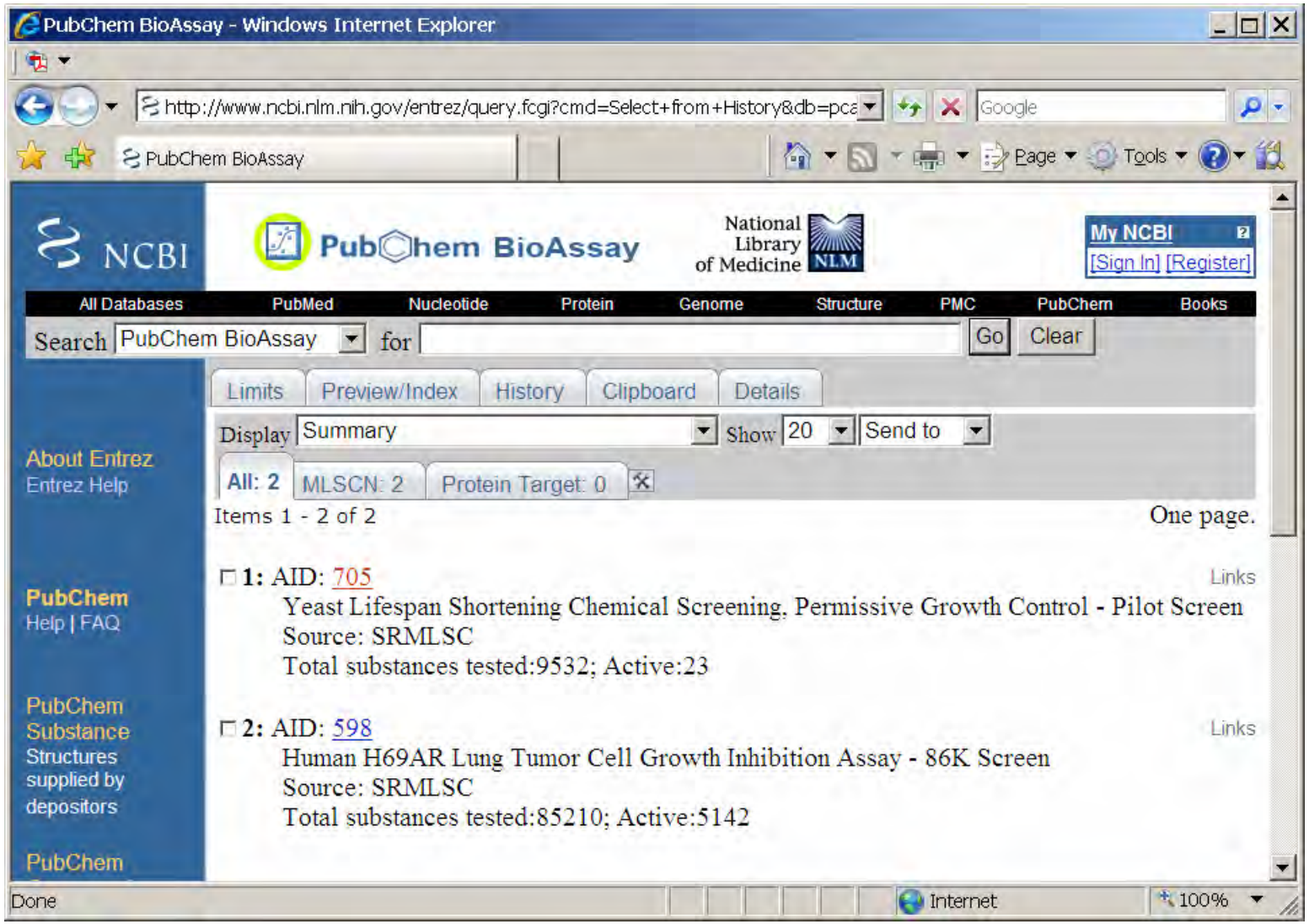
#	AID	Active	Inactive	Total Tested	Data	Name
1	706	75		75	Data	Yeast Lifespan Shortening Chemical Screening, Restrictive Growth Control - Pilot Screen
2	705	22	48	75	Data	Yeast Lifespan Shortening Chemical Screening, Permissive Growth Control - Pilot Screen
3	565	20	54	74	Data	HIV-1 RT-RNase H MLSCN HTS MH077605
4	598	19	47	66	Data	Human H69AR Lung Tumor Cell Growth Inhibition Assay - 86K Screen
5	641	11	59	70	Data	Allosteric Modulators of D1 Receptors: Primary Screen
6	642	8		8	Data	Allosteric Modulators of D1 Receptors: Confirmation Screen
7	647	8		8	Data	Allosteric Modulators of D1 Receptors: Secondary Assay 2
8	571	8	64	72	Data	Primary Cell Based High Throughput Screening Assay for Antagonists of the 5-

Internet 100%

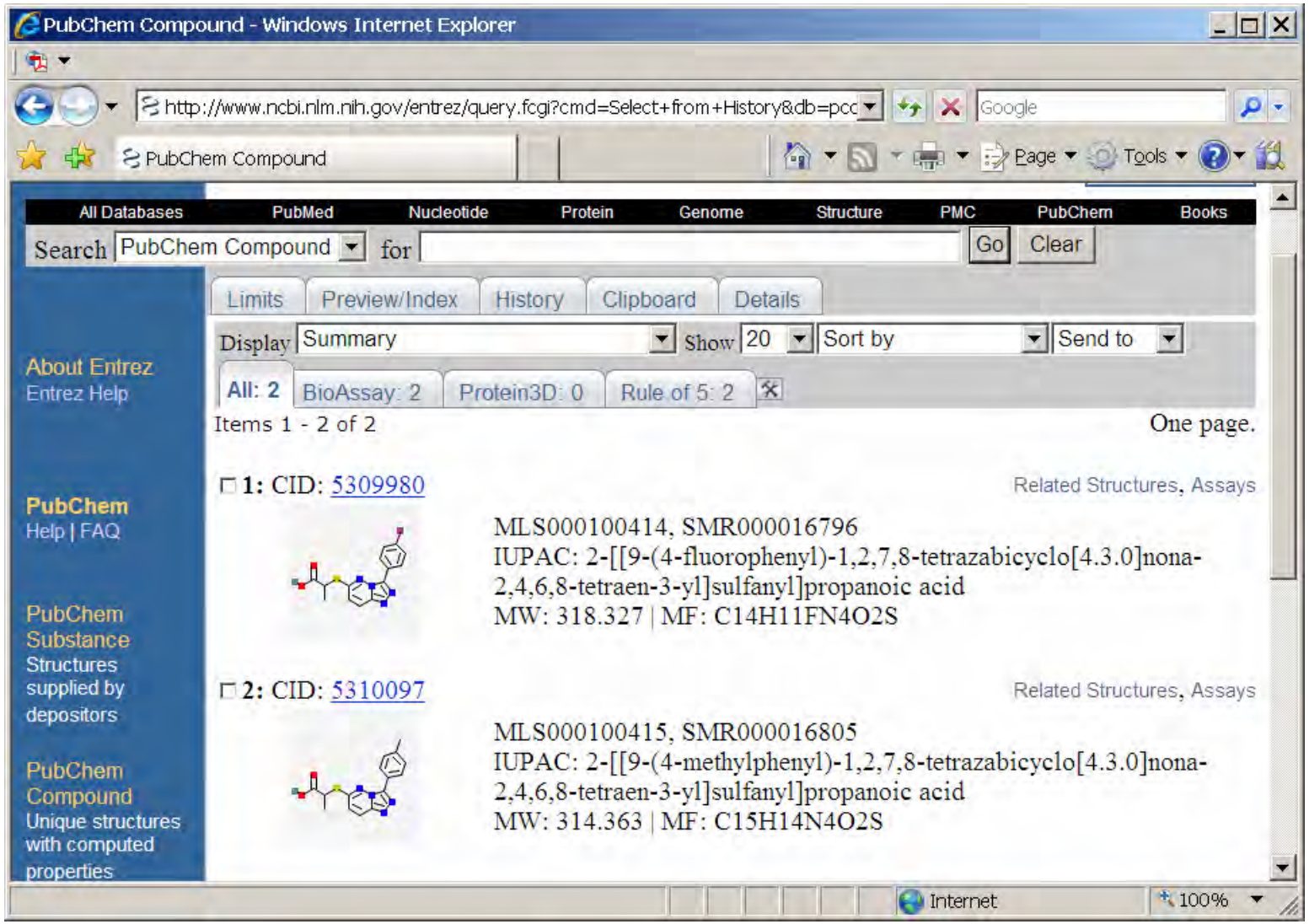
Structure-Activity Analysis ...



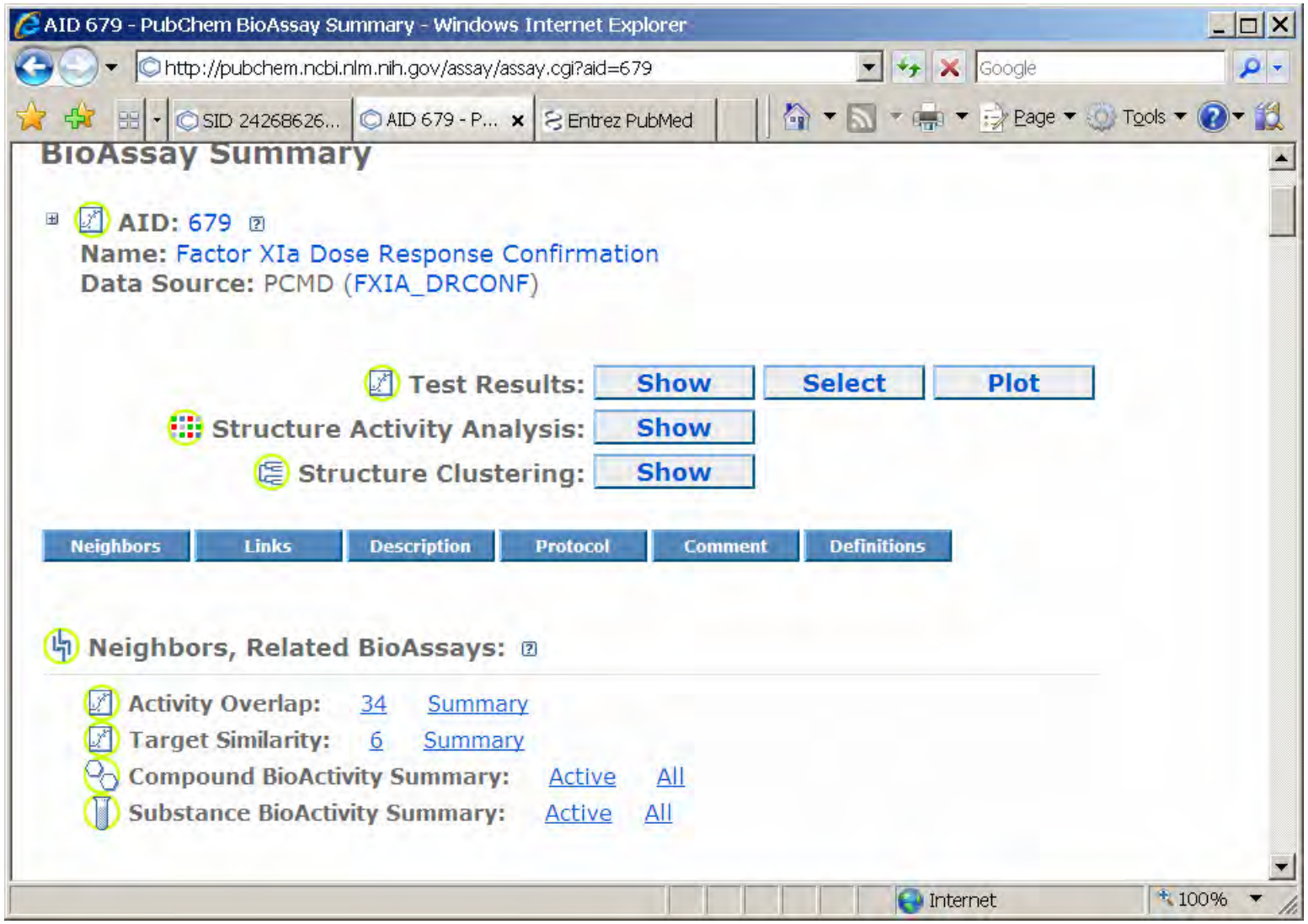
Growth Inhibition Assays Cluster ...



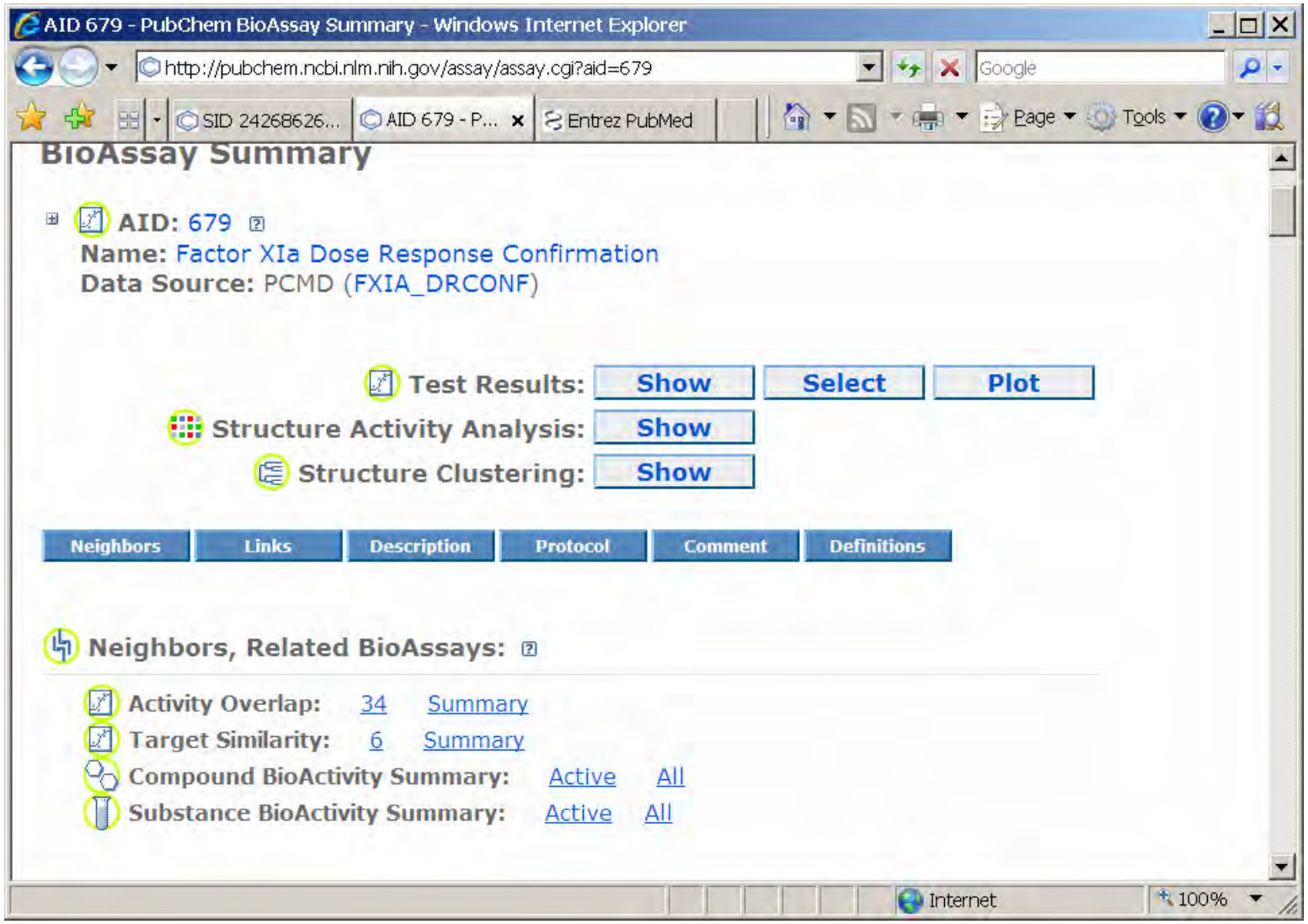
Specific Replicative Lifespan Probes ?



Factor XI Inhibition Assay ...



Factor XI Inhibition Assay ...



Factor XI Inhibition Assay ...

AID 679 - PubChem BioAssay Summary - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=679

SID 24268626... | AID 679 - P... | Entrez PubMed

Description:

Molecular Library Screening Centre Network (MLSCN)
Penn Center for Molecular Discovery (PCMD)
Assay Provider: Scott L. Diamond, University of Pennsylvania
MLSCN Grant: X01-MH076406-01

Target

Factor XI (FXI) circulates as a complex with high molecular weight kininogen (HK) in the plasma at a concentration of 5 ug/ml (equivalent to 31.3 nM, dimeric concentration) as a homodimeric glycosylated blood plasma zymogen of approximately 160 kDa, containing monomeric subunits of 80 kDa each (1). Thrombin (2, 3), factor XIa (FXIa) (3), and factor XIIa alpha (FXIIa) (4), all cleave an internal R369-I370 bond in each monomer of FXI, yielding the enzyme FXIa. The primary role of FXIa has been recognized to be part of a feedback loop resulting from thrombin activation of FXI after inhibition of the extrinsic pathway by tissue factor pathway inhibitor (TFPI).

After activation from FXI to FXIa, FXIa possesses a heavy chain of 369 residues and a light chain of 238 residues. The heavy chain consists of four apple domains (A1-A4) and the light chain represents a trypsin-like serine protease domain with active site residues at H413, D464, and S557 (1, 5-7). FXIa catalyzes FIX to FIXa activation by cleaving two scissile bonds at R145 and R180 (8). The FIXa generated can catalyze FXa formation on the platelet surface with the active cofactor factor VIIIa (FVIIIa), with FVIIIa increasing the Vmax of FX activation by FIXa by 100,000 fold (9). After activation of sufficient levels of FXa by the consolidation pathway, FXa can go on to form a ternary complex with FVa and prothrombin on the platelet surface, to give sufficient levels of thrombin for activation of fibrinogen to fibrin. Formation of this ternary prothrombinase complex in the presence of phospholipids has been shown to increase the rate of prothrombin to thrombin activation

Done | Internet | 100%

Bioactivities of Related Targets ...

Bioactivities of Related Targets ...

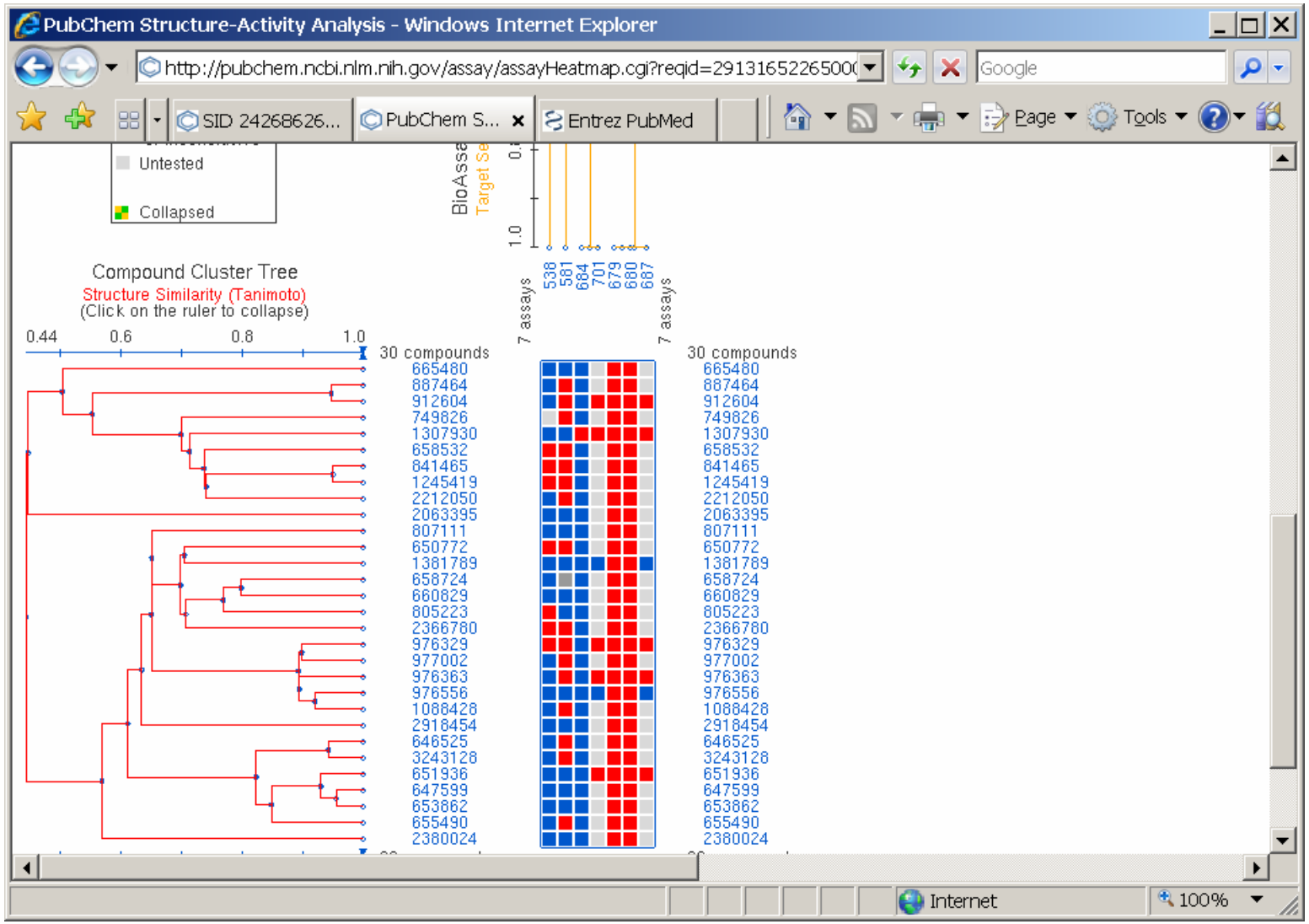
Pubchem BioAssay Neighbors - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay/assayHeatmap.cgi?reqid=7799455235327

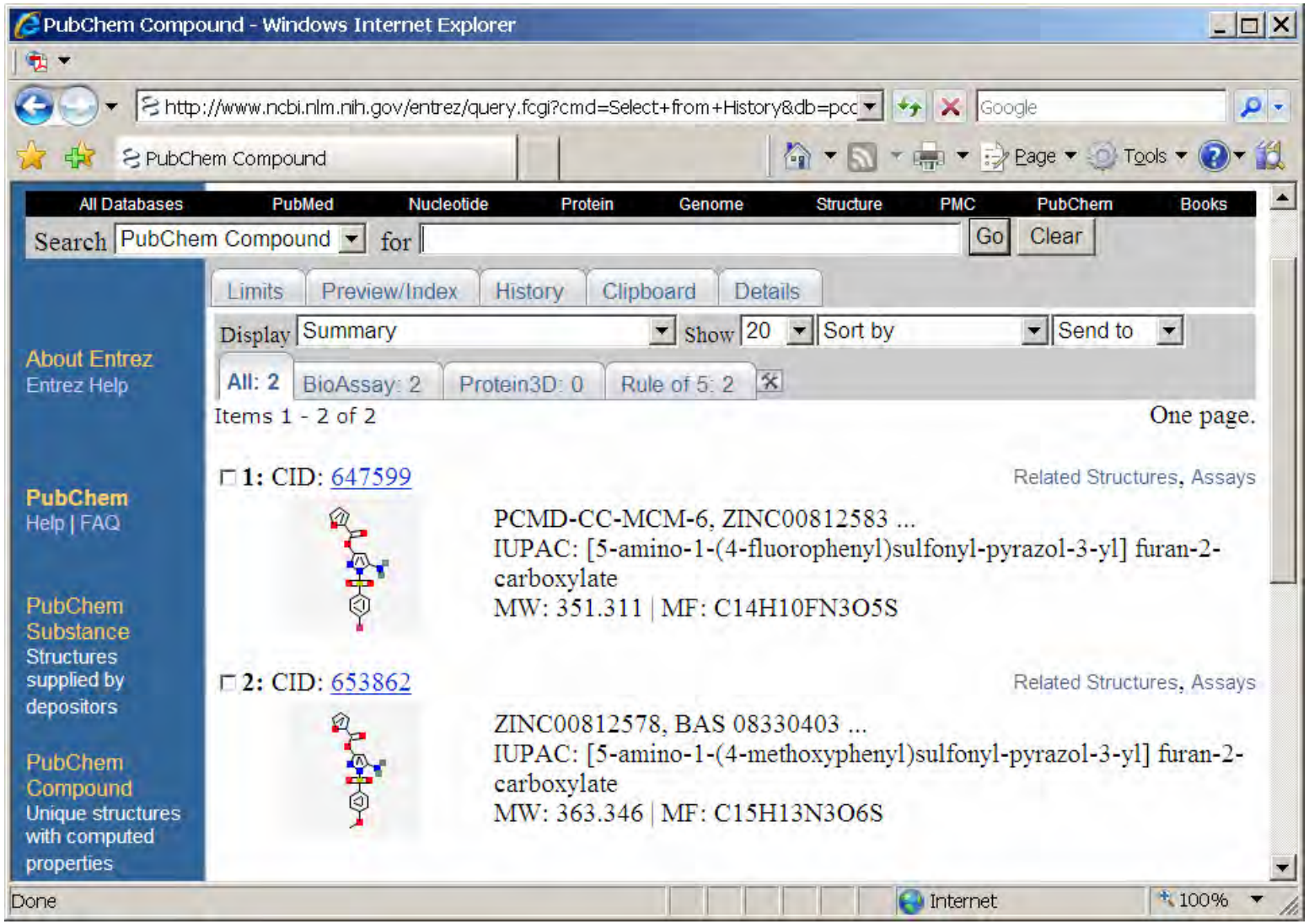
1	☑	680	1314	0e+00	1	180352	180352	coagulation factor XI
2	☑	687	1314	0e+00	1	180352	180352	coagulation factor XI
3	☑	684	155	1e-35	0.362	180352	119763	Coagulation factor XII precursor (Hageman factor) (HAF) [Contains: Coagulation factor XIIa heavy chain; Beta-factor XIIa part 1; Beta-factor XIIa part 2; Coagulation factor XIIa light chain]
4	☑	701	155	1e-35	0.362	180352	119763	Coagulation factor XII precursor (Hageman factor) (HAF) [Contains: Coagulation factor XIIa heavy chain; Beta-factor XIIa part 1; Beta-factor XIIa part 2; Coagulation factor XIIa light chain]
5	☑	538	109	1e-21	0.312	180352	4502495	complement component 1, s subcomponent [Homo sapiens]
6	☑	581	97	4e-18	0.312	180352	15680217	Cathepsin G [Homo sapiens]

Internet 100%

Structure-Activity Analysis ...



Specific Factor XI Probes ?



http://pubchem.ncbi.nlm.nih.gov

Evan Bolton

Jie Chen

Svetlana Dracheva

Lewis Geer

Lianyi Han

Jane He

Siqian He

Wolf-Dietrich Ihlenfeldt

Karen Karapetian

Vahan Simonyan

Ben Shoemaker

Wenyao Shi

Tugba Suzek

Paul Thiessen

Valery Tkachenko

Jiyao Wang

Yanli Wang

Jewen Xiao

Jian Zhang