

Searching the NCBI Databases

Text

Entrez

Sequence

BLAST

Structure

VAST

Small Molecule Structure

PubChem Structure Search

Entrez, The Life Sciences Search Engine

How to Query a Database

(term1[tag delimiter] op term2[tag delimiter] op ...)

op = AND, OR, NOT
** Boolean operators MUST be in ALL CAPS!*

tag delimiter = Entrez indexing field

Organism	}	<i>Examples of tag delimiters</i>
Journal		
User compounds		
Author		

Organism
Journal
User compounds
Author

Query Translation:
 Brauninger a[Author] AND (c-src kinase[Protein Name] OR (c-src[All Fields] AND kinase[All Fields]))

Searching the NCBI Databases

Text

Entrez

Sequence

BLAST

Structure

VAST

Small Molecule
Structure

PubChem
Structure
Search

BLAST

The screenshot shows the NCBI homepage with the 'BLAST' link in the 'All Databases' menu highlighted by a red arrow. The page includes a search bar, a 'What do NCBI do?' section, and a 'Hot Spots' section with links to Assembly Archive, Clusters of orthologous groups, Coffee Break, Genes & Disease, NCBI Handbook, Electronic PCR, Entrez Home, Entrez Tools, Gene expression omnibus (GEO), and Human genome.

Searching with Sequences

Why do we need sequence similarity searching?

- To identify and annotate sequences with...
 - incomplete (or no) annotations (GenBank)
 - incorrect annotations
- To assemble genomes
- To explore evolutionary relationships by...
 - finding homologous molecules
 - developing phylogenetic trees

NOTE: Similar sequences may NOT have similar function!

EXAMPLE: Searching with BLAST

"How do I find out which bacteria is causing the infection with an amplified 16S rRNA sequence?"

Nucleotide > Somewhat similar sequences (BLASTn)
 Query: the 16S rRNA sequence DB: "Others"
 Entrez Query: Bacteria[organism] AND biomol_rRNA[properties]
 Program: Somewhat similar sequences (BLASTn)
(You can also go to the [microbe-specific Genomic BLAST page!](#))

"How do I find out if my PCR primers are specific for my human gene?"

Nucleotide > Somewhat similar sequences (BLASTn)
 Query: primer1-(N₁₀₋₁₅)-primer2 DB: "human genomic + transcript"
 Program: Somewhat similar sequences (BLASTn)
(You can also go to the [organism-specific Genomic BLAST page!](#))

"How do I find out what protein sequence my mRNA might encode?"

BLASTx: "Search protein database using a translated nucleotide query"

"How do I find out what functional domains (proposed functions) my protein sequence might have?"

Conserved Domain Search (CD-Search/RPS-BLAST):
 "Find conserved domains in your sequence (cds)"

Searching the NCBI Databases

Text: Entrez
 Sequence: BLAST
 Structure: VAST
 Small Molecule Structure: PubChem Structure Search

in-lieu-of BLAST

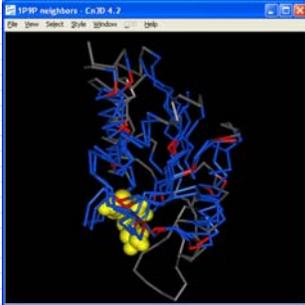
Precomputed BLASTn&p Services

***If you don't want to run a BLAST search...
 We've already done some for you!***

- Precomputed Entrez Links:
 "Related Sequences"
- The Transcript Cluster Database:
 "UniGene"
- The Protein Homolog Database:
 "HomoloGene"
- The "Conserved Domain" Database &
 CDART a.k.a. "Domain Relatives"
- The Multifunctional Blast Link:
 "BLink"

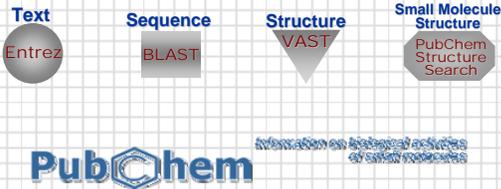
EXAMPLE: Using VAST to find binding sites

"I have a methyltransferase with a weird fold bound to the substrate s-adenosylmethionine. Are there any other proteins that have this fold?"



The Crystal Structure Of
E.coli
m1G37 tRNA
Methyltransferase
(1P9P)
&
A protein of unknown
function from
Thermotoga maritima
(1O6D)
2.4Å rmsd
11% sequence identity

Searching the NCBI Databases



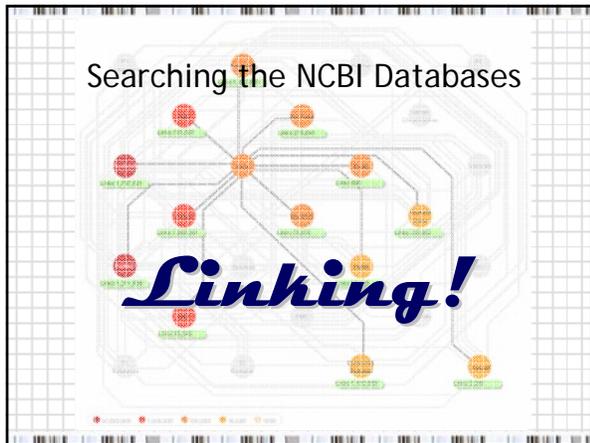
EXAMPLE: Using PubChem Structure Search

"We isolated a compound from grape skins. What is it and has it been found to have any biological activity?"

Compound Summary:
Medical Subject Annotations: (Total:2)
resveratrol
CID: 449154
BioActivity Summary: all: 155 Links, Active: 10 Links, Inactive: 136 Links, Inconclusive: 9 Links

Pharmacological Action:
Anti-Inflammatory Agents, Non-Steroidal
Antineoplastic Agents, Phytoantic
Antioxidants
Enzyme Inhibitors
Platelet Aggregation Inhibitors
Antimutagenic Agents
Anticarcinogenic Agents

#	ASD	Active	Inactive	Total	Tested	Data	Name
1	<input type="checkbox"/>	641	1	1	1	Data	Allosteric Modulators of DL Receptors: Primary Screen
2	<input type="checkbox"/>	629	1	1	1	Data	HTS for Estrogen Receptor-beta Coactivator Binding inhibitors
3	<input type="checkbox"/>	629	1	1	1	Data	HTS for Estrogen Receptor-alpha Coactivator Binding inhibitors
4	<input type="checkbox"/>	618	1	1	1	Data	Luminescent HTS for small molecule inhibitors of HT1-48P transcription
5	<input type="checkbox"/>	549	1	1	1	Data	Cell Viability - H4-B-E
6	<input type="checkbox"/>	541	1	1	1	Data	Cell Viability - N9M 3T3



Following Links

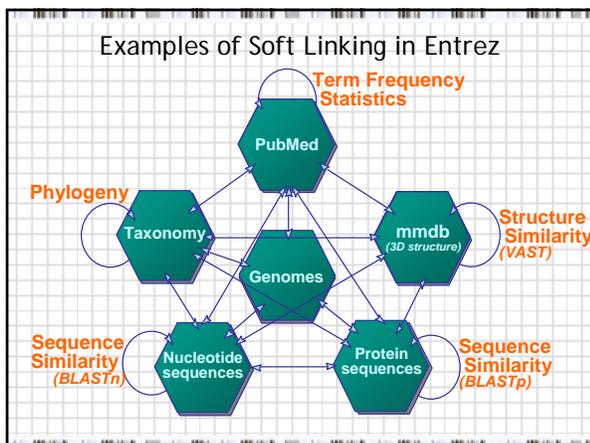
Display Summary Show 20 Send to Text

Links	Links	Links
Books	Gene	GEO
LinkOut	Full text in PMC	Nucleotide
	Related Sequences	OMIM
	Domain Relatives	Protein
	Map Viewer	PubMed
	Nucleotide	SNP
	OMIM	Taxonomy
	PubMed	UniSTS
	SNP	Map Viewer
	Taxonomy	LocusID: 3077
LinkOut	LinkOut	HGSO
		HomoloGene
		GDB
		HGMD
		Ensembl
		UCSC
		GeneTests-GeneClinics
		UniGene

Follow links to related data in the same database or in others!

“Hard” Links: Curated links based on biology
for example:
 nucleotide → taxonomy (based on organism identifier)
 protein → domain relatives (based on domain assignment)
 domains → pubmed (based on supporting literature)

“Soft” Links: Pre-computed analyses
for example:
 nucleotide → related sequences (BLASTn neighbors)
 protein → conserved domains (RPS-BLAST search)
 gene → map viewer (map position of annotated gene)



The Goal @ NCBI

Not only to provide access to the data with a way to find it, but also to help users answer a question/solve a puzzle!

-In many cases....
the data is already there.-

For More Information...

E-mail addresses:

General Help	info@ncbi.nlm.nih.gov
BLAST Help	blast-help@ncbi.nlm.nih.gov

Help Line: (301) 496-2475 (8³⁰am - 5³⁰pm EST)

The (free!) NCBI Newsletter
<http://www.ncbi.nih.gov/About/newsletter.html>

The NCBI Handbook
Follow the link from the NCBI Home Page under "Hot Spots"

The NCBI Education Page
<http://www.ncbi.nih.gov/Education/index.html>
