Zpracování dat a vizualizace v chemii C2150

- Zdeněk Kříž a Martin Prokop

- Co se naučíte
  - Získávání strukturních dat z databází
  - Validace získaných dat
  - Příprava pro molekulové modelování
  - Extrahování dat z modelování
  - Vizualizace dat a tvorba prezentací

Zpracování dat a vizualizace v chemii C2150

- Jaký software budeme používat
  - Web browser
  - What If
  - Triton
  - VMD
  - Open Office

Jak můžu získat strukturní data o molekule

- Pro malé molekuly:
  - Cambridge Structural Database
  - Glycan Structural Database
  - 3Dchem database
  - Modelování od začátku
- Velké biomolekuly
  - Protein Structural Database – PDB
  - Nucleic Acids Database – NDB
  - Modelování na základě homologie

Jak hledat v databázi

- Znám název molekuly, nebo její vzorec
- U malých molekul znám jejich SMILES kód (CCO – ethanol, c1CCCCc1 – cyklohexan a pod).
- CSD umí hledat podle funkčních skupin (umožňuje přes grafické rozhraní nakreslit vzorec)
- NDB databáze – hledání podle sekvence bazí
- PDB umožňuje několikastupňové prohledávání

Cambridge structural database

- Co ukládá CSD
  - Data o organických molekulách
  - Data o komplexech kovů s organickými molekulami
- Jakými metodami jsou získávána data
  - RTG krystalografie
    - Monokrystalů
    - Prášková difrakce
- Co v CSD nenajdeme
  - Data o polypeptidech, polysacharidech a oligonukleotidech
  - Data o anorganických molekulách

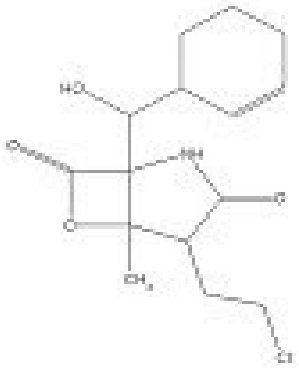
Cambridge structural database

1D Bibliographic Information

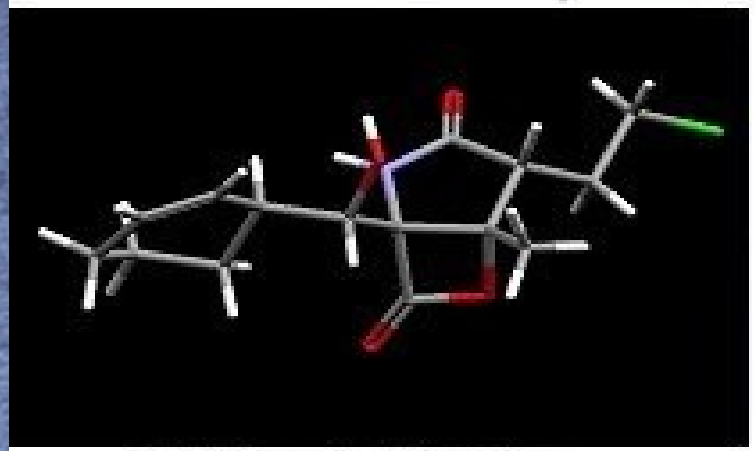
TUSJIA
4-(2-chloroethyl)-1-[cyclohex-2-en-1-yl(hydroxy)methyl]-5-methyl-6-oxa-2-azabicyclo[3.2.0]heptane-3,7-dione
Synonym: Salinosporamide A
Formula: C15 H20 Cl1 N1 O4
Source: Salinospora strain CNB-392 isolated from heat treated marine sediment
Extra info: absolute configuration; potent cancer cell cytotoxicity through inhibition of the 20S proteasome
R.H.Feling, G.O.Buchanan, T.J.Mincer, C.A.Kauffman, P.R.Jensen, W.Fenical

Angew.Chem., Int.Ed.Engl., **42**, 355, 2003

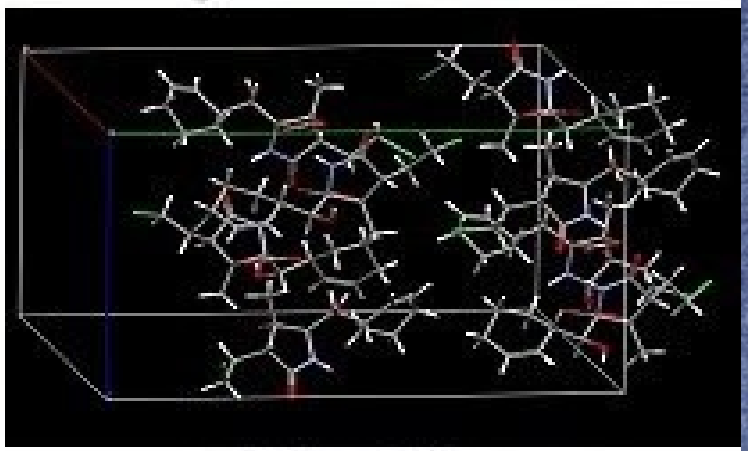
2D Chemical Connectivity



CSD



3D Molecular Structure



3D Crystal Structure

Cambridge structural database

CSD Entries: Summary Statistics (Historical Snapshots) - Mozilla Firefox

File Edit View History Bookmarks ScrapBook Tools Help

http://www.ccdc.cam.ac.uk/products/csd/stati

Google

	31 Dec 1983		*31 Dec 1990*		*30 Oct 2001*		*31 Dec 2006*	
	Structures	*%*	*Structures*	*%*	*Structures*	*%*	*Structures*	*%*
Total structures	52 363	100.0	104 380	100.0	251 515	100.0	400 977	100.0
Organic structures	28 995	55.4	52 450	50.3	112 113	44.6	172 583	43.0
Transition metal present	20 439	38.9	45 588	43.7	120 638	48.0	210 700	52.5
Li – Fr or Be – Ra present	2 887	5.5	5 299	5.1	13 471	5.4	21 219	5.3
Main group metal present	2 206	4.2	5 024	4.8	16 171	6.4	25 790	6.4
3D coordinates present	37 318	71.0	83 884	80.4	223 920	89.0	365 075	91.0
Error-free coordinates	35 032	93.9†	80 372	95.8†	219 864	98.2†	357541	97.9†
Neutron studies	567	1.08	786	0.8	1 062	0.4	1 331	0.3
Low/high temp. studies	3 275	6.2	9 943	9.5	55 752	22.2	133 471	33.3
Abs. config. determined	1 330	2.5	2 344	2.2	4 924	2.0	6 291	1.6
Disorder present	4 943	9.4	13 594	13.0	45 728	18.2	81 770	20.4
Polymorphic structures	3 231	6.1	4 618	4.4	7 892	3.1	13 411	3.3
R-factor < 0.100	37 190	70.8	85 389	81.8	227 181	90.3	370 220	92.3
R-factor < 0.075	29 937	57.0	73 424	70.3	202 848	80.7	333 501	83.2
R-factor < 0.050	15 974	30.4	42 996	41.2	125 112	49.7	209 929	52.4
R-factor < 0.030	2 231	4.2	7 150	6.8	22 346	8.9	39 602	9.9
n(atoms)/structure *	44	-	54	-	73	-	74	-
Mb data added in year+	14	-	29	-	86	-	139	-

† Percentage of structures for which coordinates are present in the CSD

Done

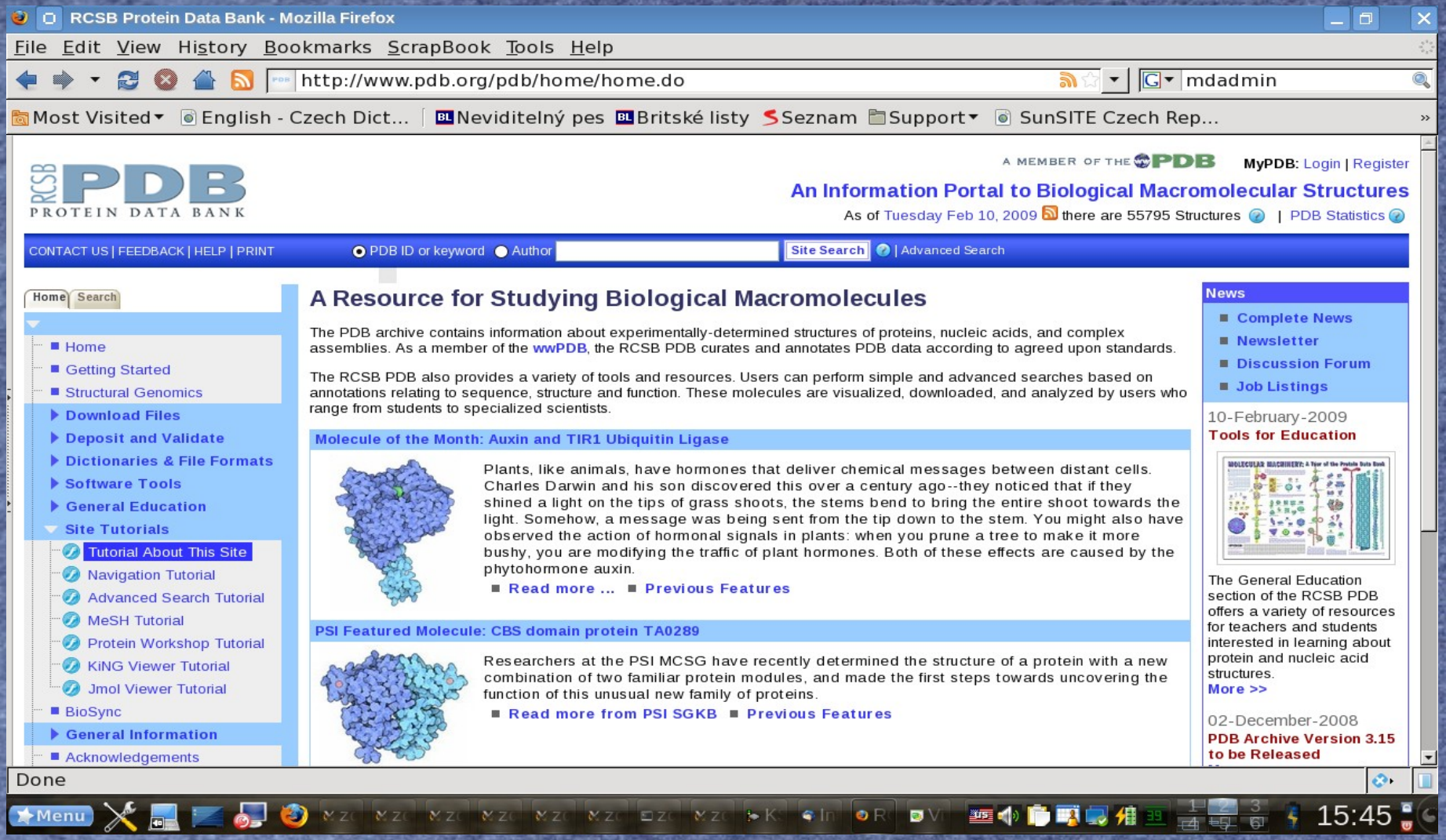
ICSD databáze

- **http://www.fiz-karlsruhe.de/icsd_content.html**

- Co najdeme v ICSD
  - Krystalové struktury prvků
  - Krystalové struktury binárních sloučenin
  - Krystalové struktury složitějších anorganických sloučenin
- ICSD obsahuje více než 108000 záznamů

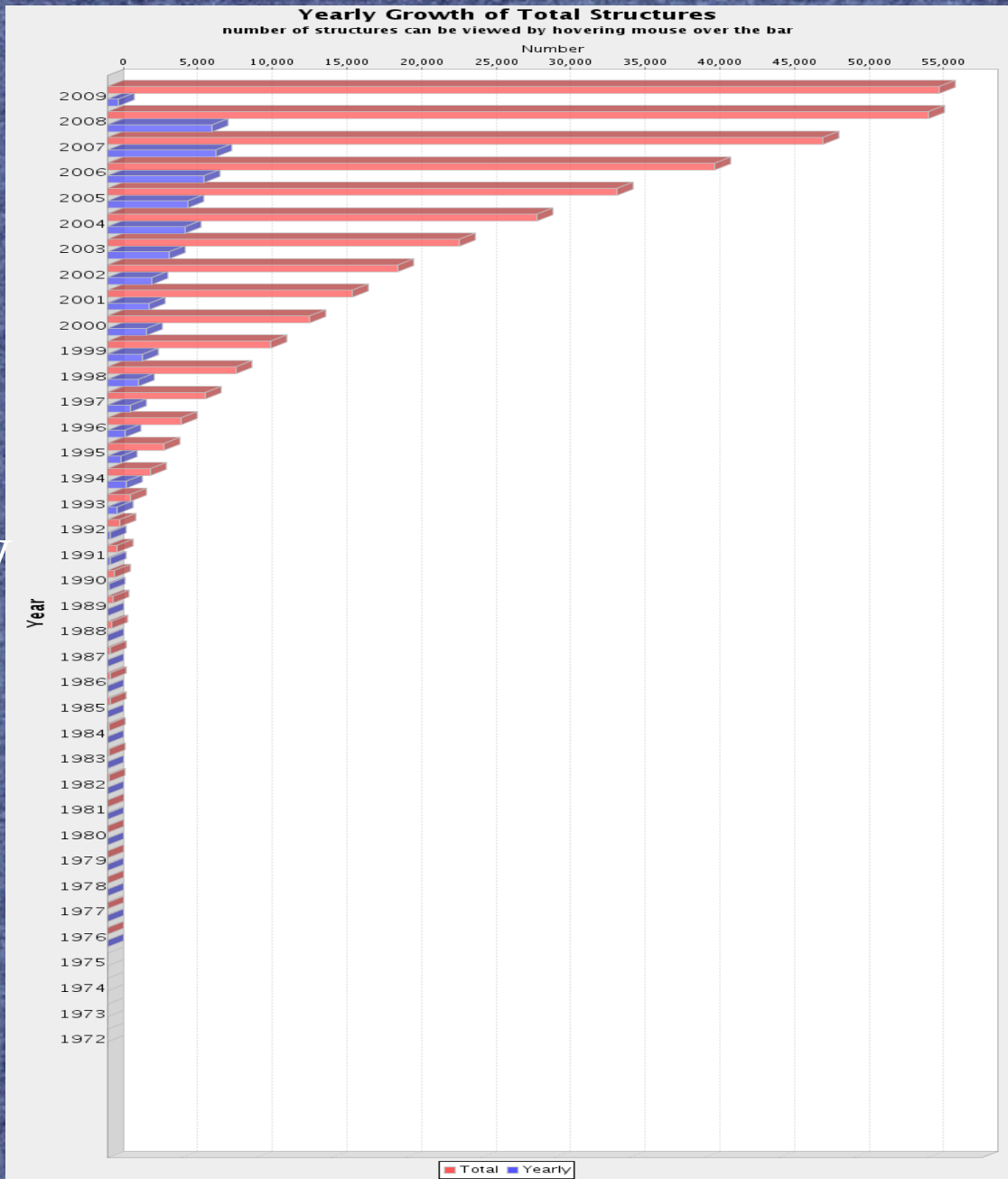
Protein Data Bank - PDB

- **http://www.pdb.org**

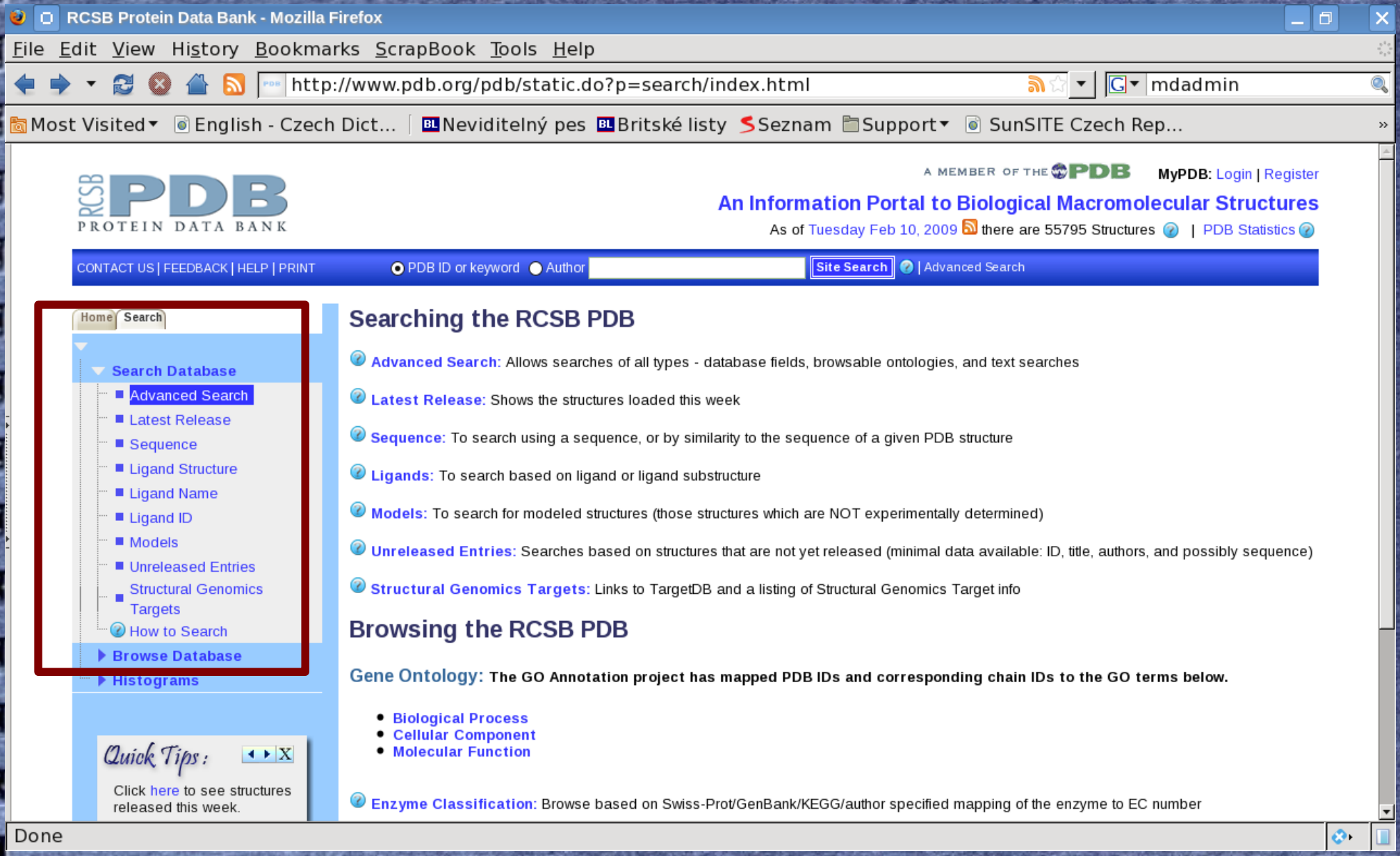


Protein Data Bank - PDB

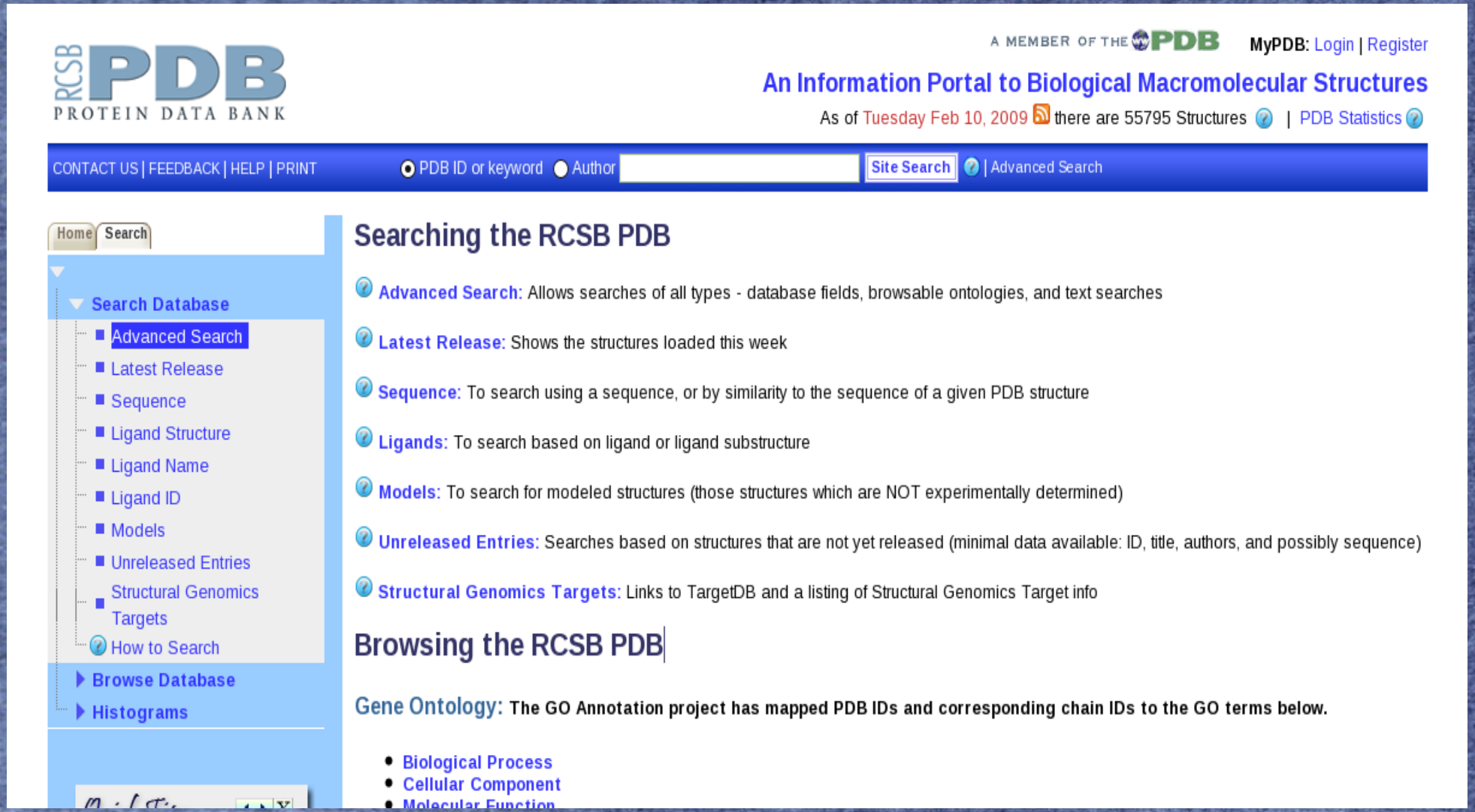
- 55795 struktur
  - 47777 – RTG
  - 7696 – NMR
  - 217 – EM
- proteiny/NA/komplexy
  - 51535 – protein
  - 1173 – DNA
  - 701 – RNA
  - 2255 – komplexy



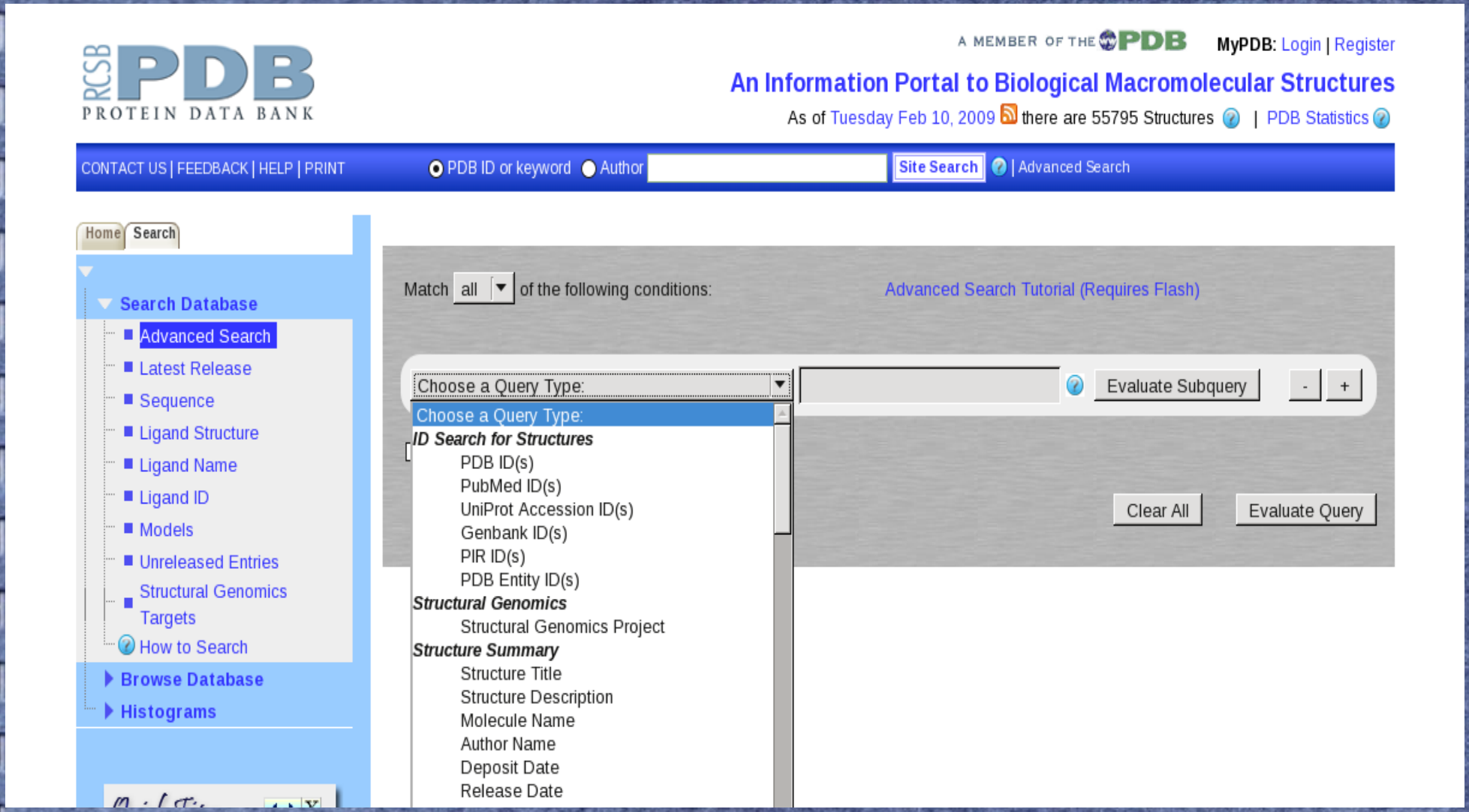
Protein Data Bank - PDB



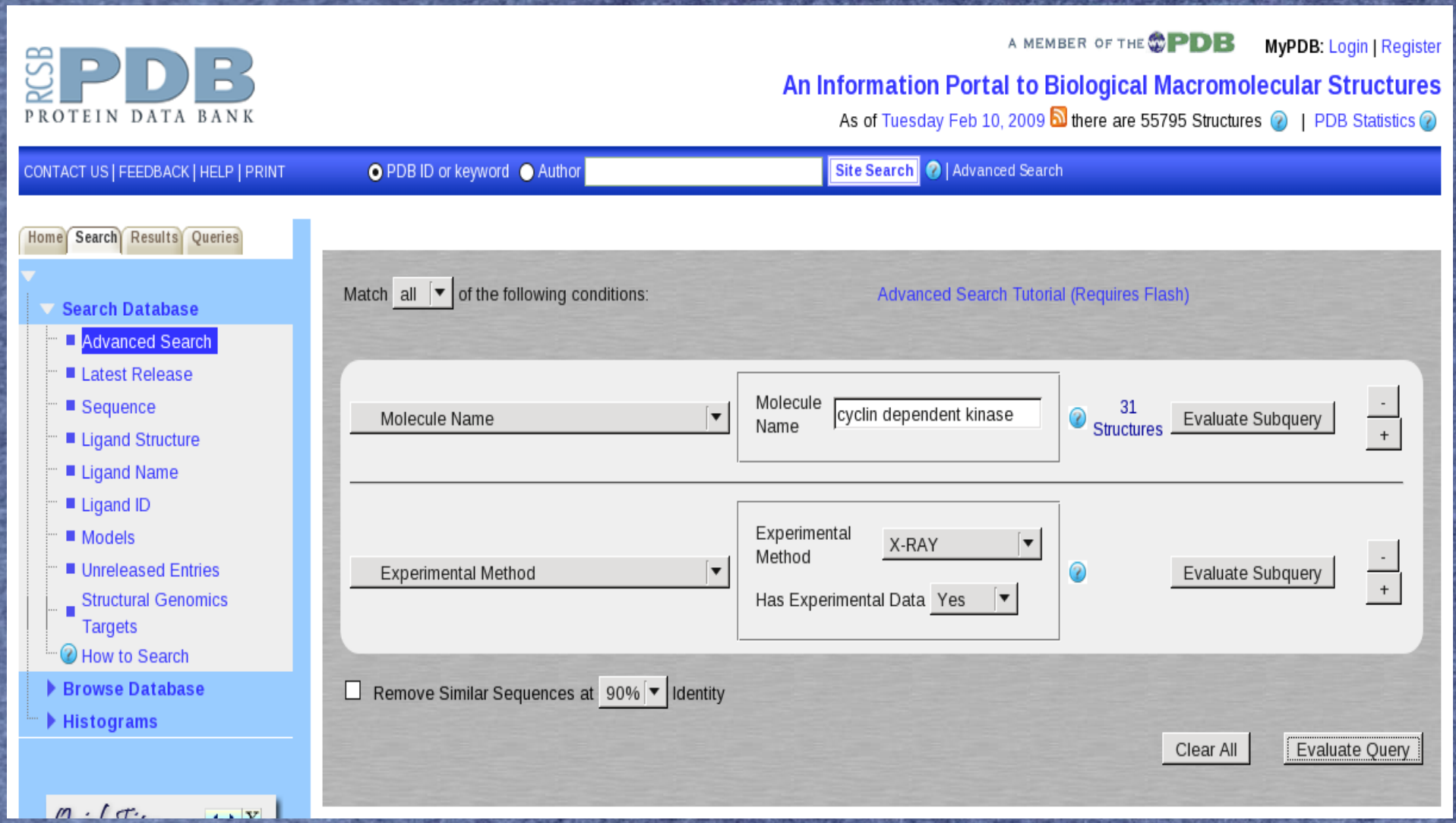
PDB – prohledávání databáze



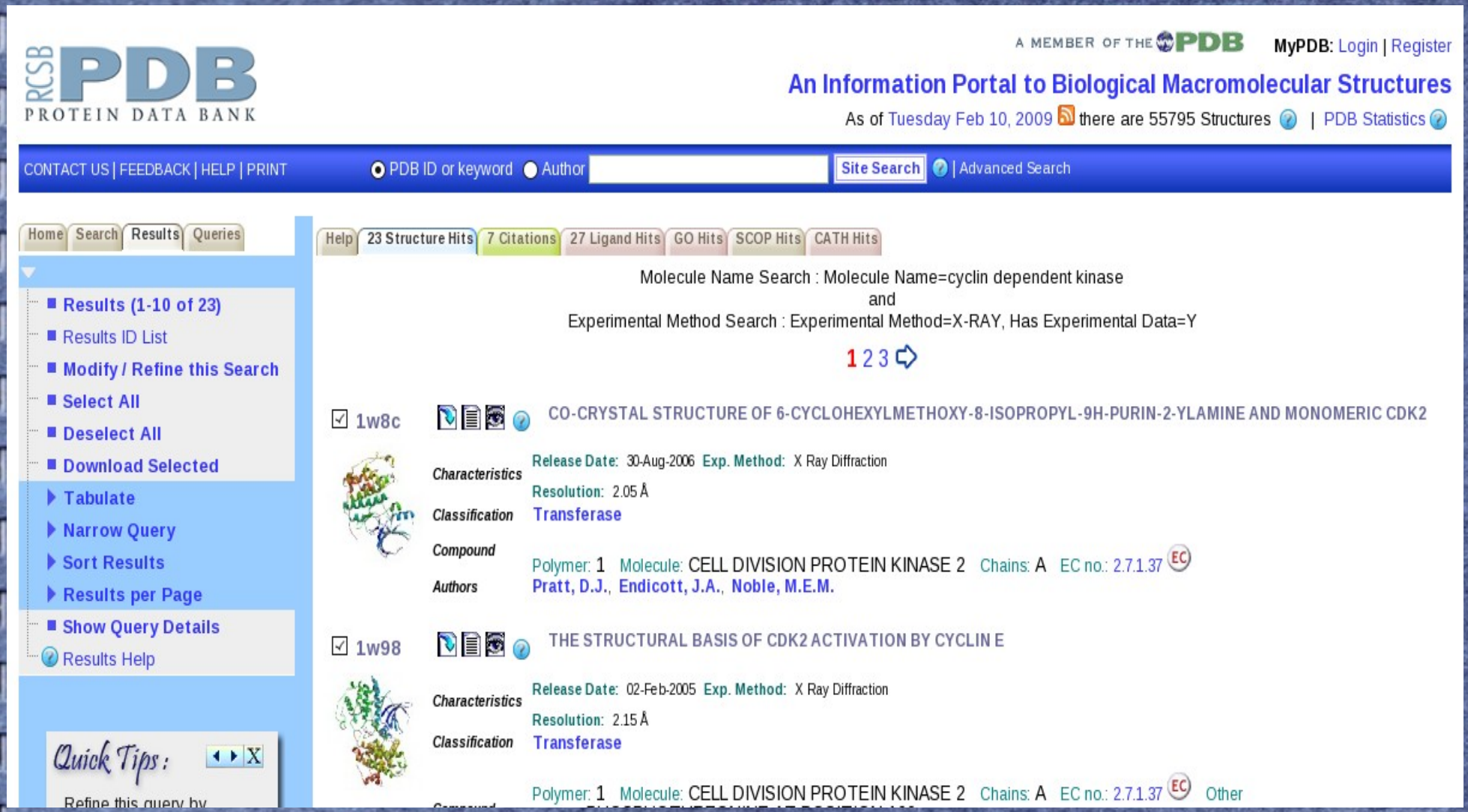
PDB – prohledávání databáze



PDB – prohledávání databáze



PDB – prohledávání databáze



PDB – prohledávání databáze

PROTEIN DATA BANK

As of Tuesday Feb 10, 2009 there are 55795 Structures | PDB Statistics

CONTACT US | FEEDBACK | HELP | PRINT PDB ID or keyword Author Site Search | Advanced Search

Home | Search | Structure | Results | Queries

Help | Structure Summary | Sequence Details | Biology & Chemistry | Materials & Methods | Geometry | External Links

- 1W98
- Download Files
- FASTA Sequence
- Download Original Files
- Display Files
- Display Molecule
- Structural Reports
- External Links
- Structure Analysis
- Help

Quick Tips :

To view **sequence details** of this structure click on the *Sequence Details* tab above the summary page.

1w98

DOI 10.2210/pdb1w98/pdb

Red - Derived Information

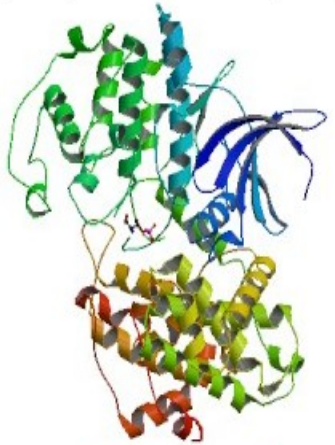
Title	THE STRUCTURAL BASIS OF CDK2 ACTIVATION BY CYCLIN E
Authors	Lowe, E.D., Honda, R., Dubinina, E., Skamnaki, V., Cook, A., Johnson, L.N.
Primary Citation	Honda, R., Lowe, E.D., Dubinina, E., Skamnaki, V., Cook, A., Brown, N.R., Johnson, L.N. (2005) The structure of cyclin E1/CDK2: implications for CDK2 activation and CDK2-independent roles. *Embo J.* ***24***: *452-463* [Abstract] PubMed
History	Deposition 2004-10-07 Release 2005-02-02 Last Modified (REVDAT) 2005-06-15
Experimental Method	Type X-RAY DIFFRACTION Data

Parameters	Resolution[Å]	R-Value	R-Free	Space Group
	2.15	0.185 (obs.)	0.246	$P\,4_1\,2_1\,2$

Length [Å]	a	99.62	b	99.62	c	149.99

Images and Visualization

<< Biological Molecule >>

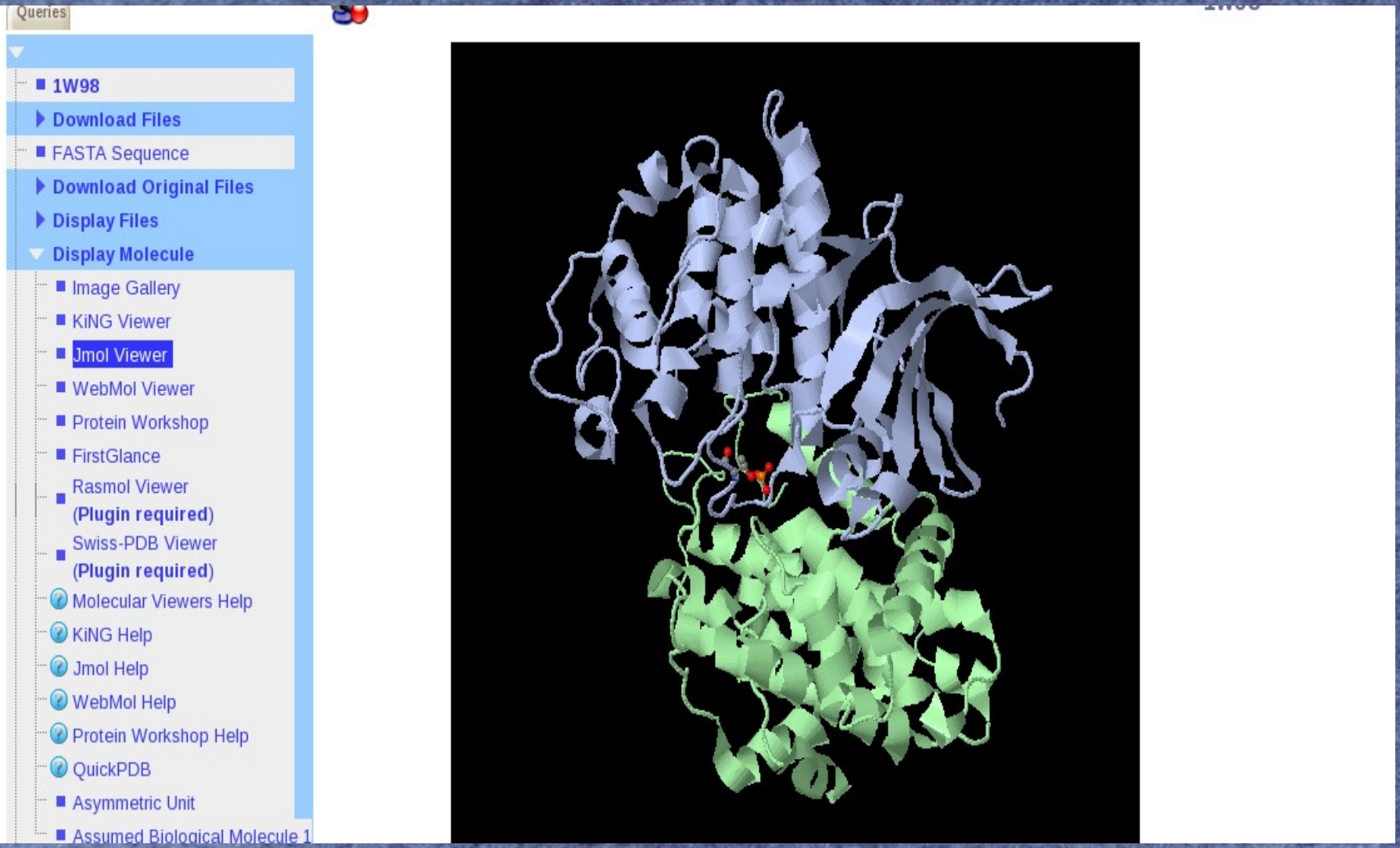


Display Options

KiNG
Jmol
WebMol
MBT SimpleViewer*
MBT Protein Workshop
QuickPDB
All Images

* Capable of displaying biological molecules.

PDB – prohledávání databáze



PDB – prohledávání databáze

Nucleic Acid Database

http://ndbserver.rutgers.edu

Nucleic Acid Database (NDB) - Mozilla Firefox

File Edit View History Bookmarks ScrapBook Tools Help

http://ndbserver.rutgers.edu/

NDBDatabase

Most Visited English - Czech Dict... Neviditelný pes Britské listy Seznam Support SunSITE Czech Rep...

ndb

Site Index

WELCOME TO THE

NUCLEIC ACID DATABASE

a repository of three-dimensional structural information about nucleic acids

- Atlas
- Deposit Data
- Download Data
- Search
- Reports
- Education
- Standards
- Tools
- Links

Number of Released Structures:
4089 Structures
Last Update: 15-Jan-2009

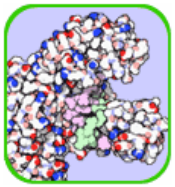
Search the NDB by ID

Enter an NDB ID or PDB ID

Search

Search for Released Structures

Nucleic Acids Highlight



About NDB

News

The NDB follows the dictionaries and formats used by the Worldwide Protein Data Bank. Please see www.wwpdb.org for format announcements and documentation.

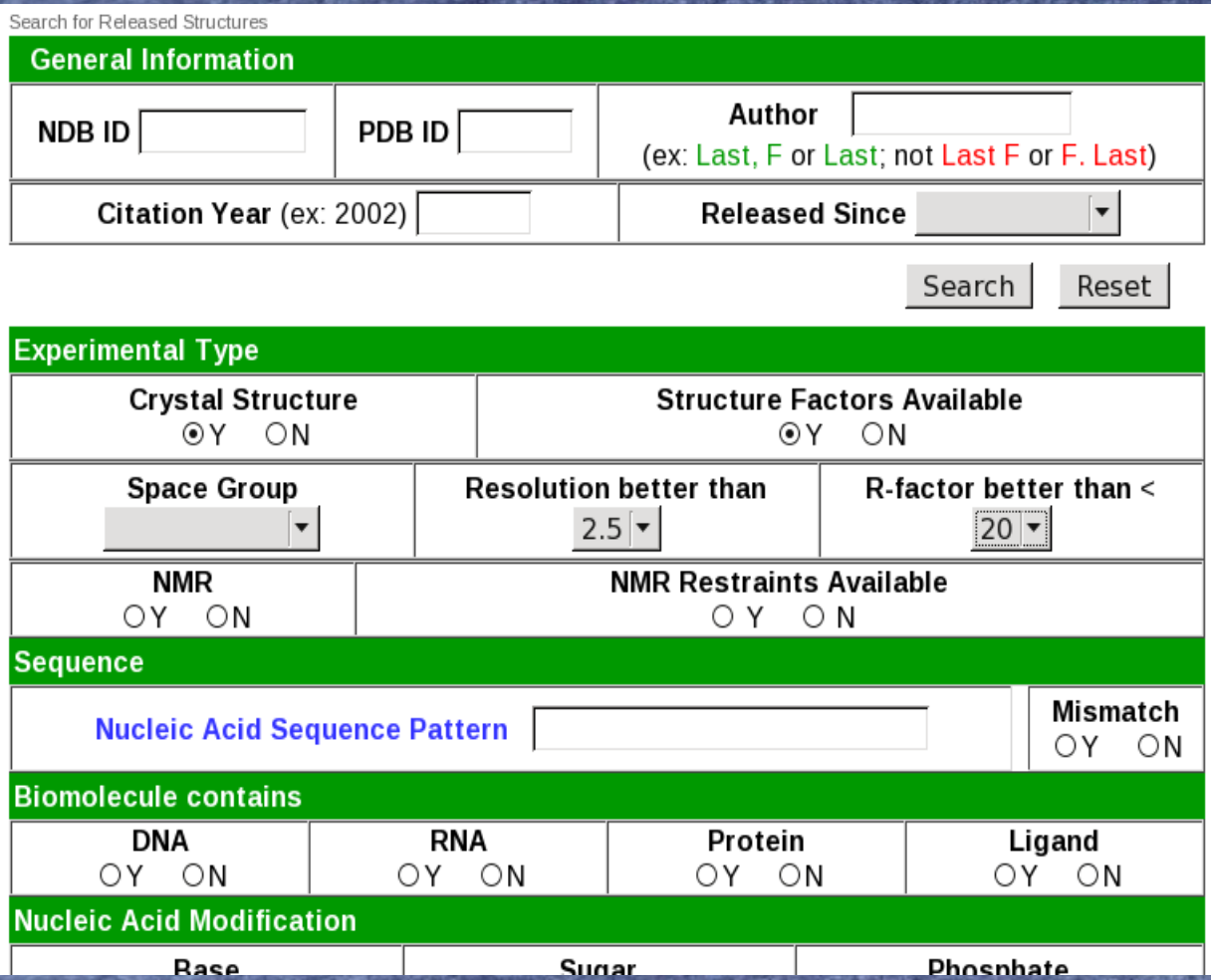
Archive of NDB newsletters

The NDB is supported by funds from the National Science Foundation and the Department of Energy.

In citing the NDB please refer to: H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. (1992) The Nucleic Acid Database: A Comprehensive Relational Database of Three

http://ndbserver.rutgers.edu/download_data/index.html

NDB - prohledávání



NDB - prohledávání

NUCLEIC ACID DATABASE

Results **1** - **20** of **374** records in **NDB** Search took **0.595** seconds.

There are **19** pages in total. Help Close

1. Display Dynamically Generated Atlas of _____ Checked Structures Display Atlas

2. View NDB Status report of View Report

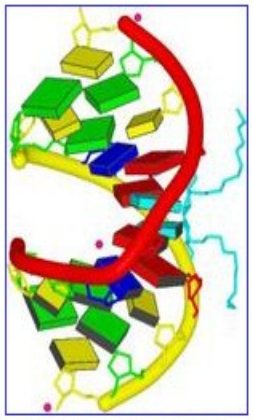
#	ID	DESCRIPTION
☐ All		
☐ 1	AD0007	A-DNA DECAMER GCGTA(T23)TACGC WITH INCORPORATED 2'-METHOXY-3'-METHYLENEPHOSPHATE-THYMIDINE
☐ 2	AD0010	ALTERNATION OF DNA AND SOLVENT LAYERS IN THE A FORM OF D(GGCGCC) OBTAINED BY ETHANOL CRYSTALLIZATION
☐ 3	AD0011	1.6 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-AMINOOXYETHYL THYMINE IN PLACE OF T6, BA-FORM
☐ 4	AD0012	1.45 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-FLUOROETHYL THYMINE IN PLACE OF T6, MEDIUM NA-SALT
☐ 5	AD0013	1.06 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-METHYL-3'-METHYLENEPHOSPHONATE (T23) THYMINE IN PLACE OF T6, HIGH CS-SALT
☐ 6	AD0014	1.05 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-METHYL-[TRI(OXYETHYL)] THYMINE IN PLACE OF T6, MEDIUM CS-SALT
☐ 7	AD0015	1.05 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-FLUOROETHYL THYMINE IN PLACE OF T6, HIGH RB-SALT
☐ 8	AD0016	1.3 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-METHYL-[TRI(OXYETHYL)] THYMINE IN PLACE OF T6, MEDIUM RB-SALT
☐ 9	AD0018	1.3 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-METHYL-[TRI(OXYETHYL)], MEDIUM K-SALT

NDB - prohledávání

NDB ID: AD0014

Home
NMR Atlas X-Ray Atlas

Title:	1.05 A STRUCTURE OF THE A-DECAMER GCGTATACGC WITH A SINGLE 2'-O-METHYL-[TRI(OXYETHYL)] THYMINE IN PLACE OF T6, MEDIUM CS-SALT
Molecular Description:	5'-D(GpCpGpTpAp(126)pApCpGpC)-3'
Structural Features:	A DOUBLE HELIX
Nucleic Acid Sequence:	ChainsA, B: (DG) (DC) (DG) (DT) (DA) (126) (DA) (DC) (DG) (DC)
Primary Citation:	Tereshko, V., Wilds, C.J., Minasov, G., Prakash, T.P., Maier, M.A., Howard, A., Wawrzak, Z., Manoharan, M., Egli, M. Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. *Nucleic Acids Res.* , **29**, pp. 1208 - 1215, 2001.

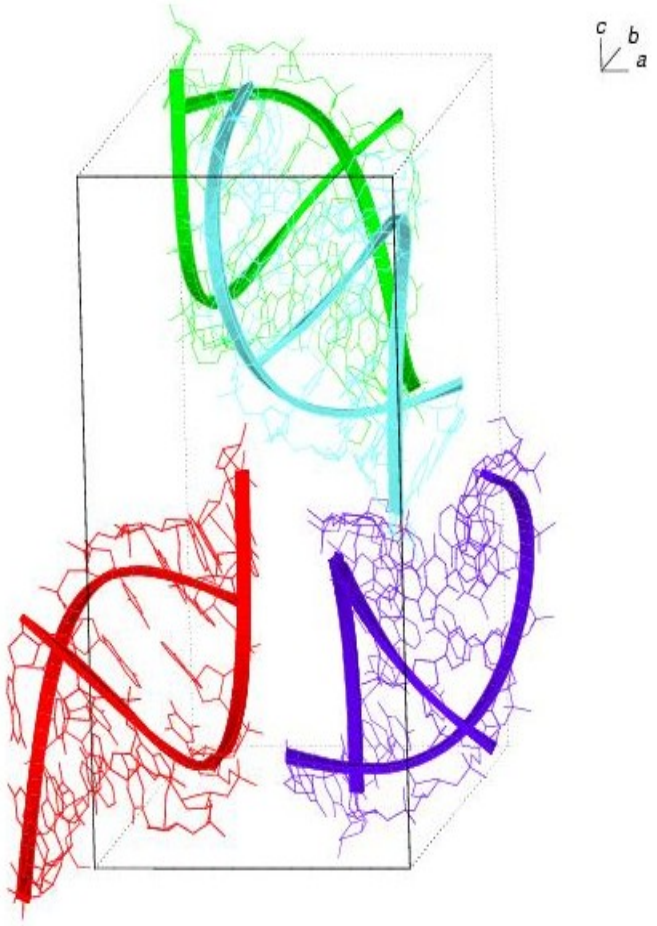


Biological Unit 1

Other Views

Asymmetric Unit
Crystal Packing
Enlarge Biological Unit 1

NDB - prohledávání



Glyco3D databáze

- **http://www.cermav.cnrs.fr/glyco3d/index.php**

Cermav

Home
- Home page

Databases
- Monosaccharides
- Disaccharides
- Oligosaccharides
- Polysaccharides
- Lectines
- Glycosyltransferases
- Gag binding proteins

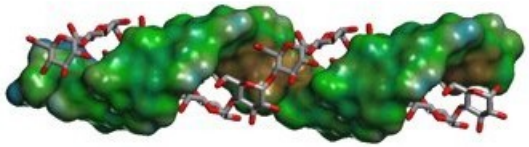
Book chapters
- Modeling
- Cellulose
- Starch

Gallery
- Models

Tools
- Useful links

GLYCO3D

A site for glycosciences



BP 53 - 38041 Grenoble Cedex 9 - France

Domaine Universitaire - 601 Rue de la Chimie
38400 St Martin d'Hères - France

tél : 33 (0)4 76 03 76 03 - fax : 33 (0)4 76 54 72 03

Glyco3D databáze

Mozilla Firefox <2>

http://www.cermav.cnrs.fr/cgi-bin/monos/monos.cgi

Google

MONOSACCHARIDES

- ABEQUOSE
- ALTROSE
- APIOSE
- ARABINOSE
  - Alpha
    - a-L-Arabinopyranose
  - Beta
- FRUCTOSE
- FUCOSE
- GALACTOSE
- GLUCOSE
- GULOSE
- IDOSE
- LYXOSE
- MANNOSE
- MISC
- RHAMNOSE
- RIBOSE
- SIALIC
- TALOSE
- XYLOSE

a-L-Arabinopyranose

Family : Arabinose
Anomery : Alpha

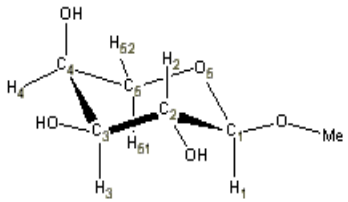
Carbons : 5

Cycle : Pyranose

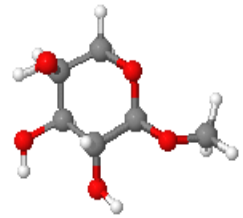
Configuration : Levogyre

Comment :

2D view



3D view



Jmol

PDB file

Glyco3D databáze

Mozilla Firefox <2>

http://www.cermav.cnrs.fr/cgi-bin/di/di.cgi

Google

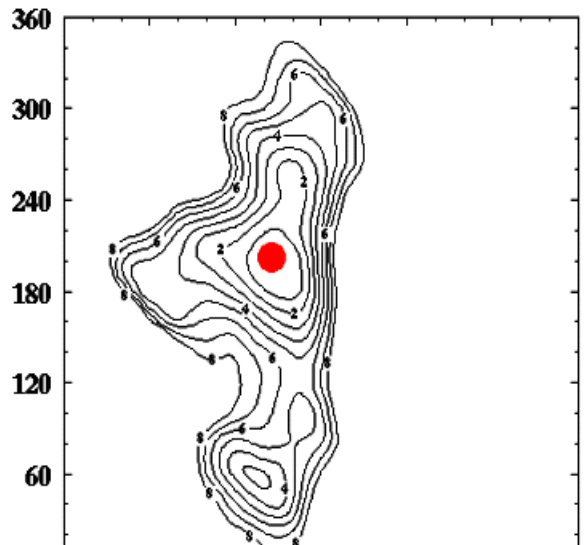
monosaccharide 1 | Carbon 1 | glycosidic link | Carbon 2 | monosaccharide 2

Fucose | 1 | ϕ O ψ | 4 | Glucose | search | Help and informations

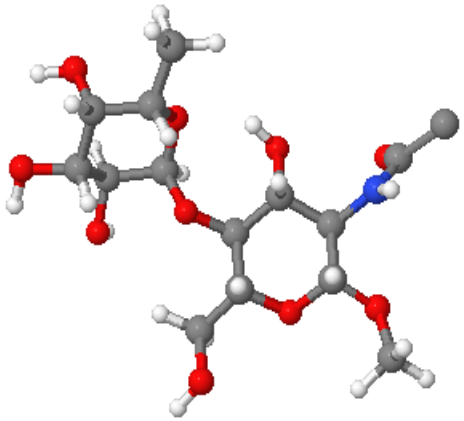
IUPAC name : a-L-Fucopyranose-(1-4)-2-acetamido-2-deoxy-b-D-glucopyranose | select

Monomer 1 :	*a-L-Fucopyranose*
Monomer 2 :	*2-Acetamido-2-deoxy-b-D-glucopyranose*
Abreviation :	*a-L-Fucp(1,4)b-D-GlcpNAc*

Iso-potential energy map



3D interactive stucture



Jmol

VALUES :

Energy :	*29.0400*
ϕ :	*99.99*
Ψ (n+1) :	*99.99*
Ψ (n-1) :	*81.40*
Force-field :	*MM3(92)*
Parameters :	*MM3(92)*
Dielectric constant :	*78.50*

Glyco3D databáze

Mozilla Firefox <2>

http://www.cermav.cnrs.fr/lectines/

Google

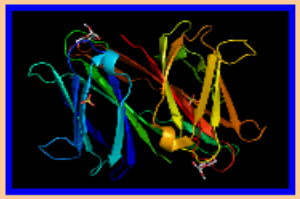
LECTINES

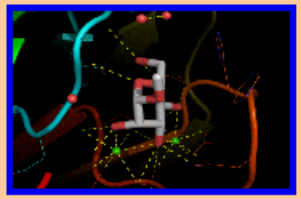
- Algal lectins
- Animal lectins
- Bacterial lectins
  - 1-Ca b-sandwich
  - 2-Ca b-sandwich
    - Burkholderia lectins
    - Chromobacterium CV-IIL
    - Pseudomonas PA-IIL
    - Ralstonia RS-IIL
  - ADP-ribosylating toxin (AB5 toxin)
  - b-propeller
  - b-trefoil hemagglutinin
  - Bacterial neurotoxin
  - cyanobacterial lectins
  - Pili adhesin
  - Staphylococcal toxin
  - Toxin repetitive domain

Bacterial lectins - 2-Ca b-sandwich

Burkholderia lectins

PDB code : 2VNV
Species : Burkholderia cenocepacia
Commentaire : Complexed with **aManOMe**
Resolution : 1.70
Reference : Lameignere E, Malinovská L, Sláviková M, Duchaud E, Mitchell EP, Varrot A, Sedo O, Imberty A, Wimmerová M., Structural basis for mannose recognition by a lectin from opportunistic bacteria Burkholderia cenocepacia:, *Biochem. J.*, **411**, 307-318, 2008





Medline

Link to Databases :
[PDB ENTRY (RCSB Site) / HSSP ENTRY / SCOP / RELIBASE]

PDB Files :
(RCSB (USA CA) / Rutgers University (USA NJ) / FRANCE / SINGAPORE / UK / JAPAN / BRAZIL)

Done

Practical structural databases

- **http://xray.bmc.uu.se/embo/structdb/links.html**