RCSB PDB
PROTEIN DATA BANK

Small molecule data and tools at RCSB

John Westbrook
Rutgers University

September 24, 2008

EMBO

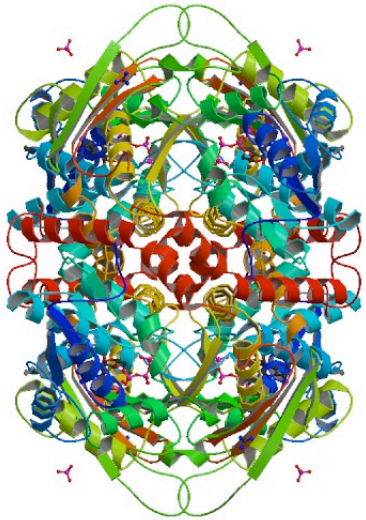
Computational Aspects of the Protein Target Selection, Protein Production Management and Structure Analysis Pipeline

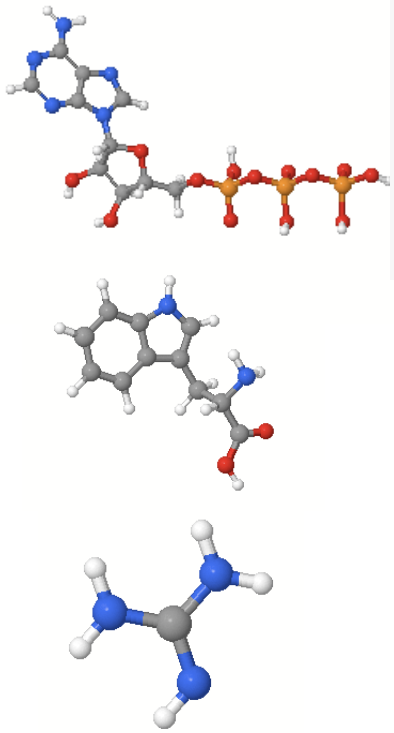
Outline

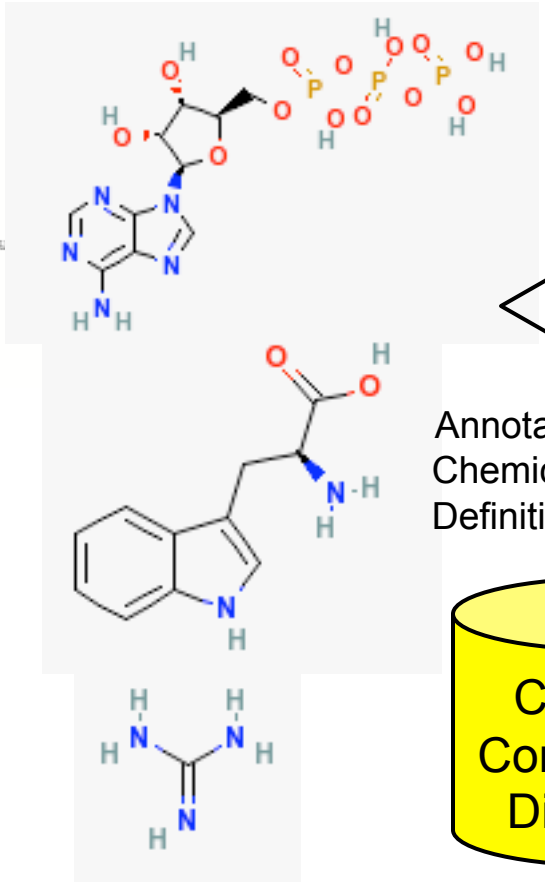
- Chemical Components Dictionary
- Ligand Expo
- RCSB PDB website

RCSB PDB
PROTEIN DATA BANK

How does new chemistry enter PDB?







New?

Standardize Residue/Atom Nomenclature

No

Yes

Annotate Chemical Definition

Chemical Components Dictionary

Deposited Coordinates

Chemical Components

Perceive Covalent Structure

Compare with Dictionary

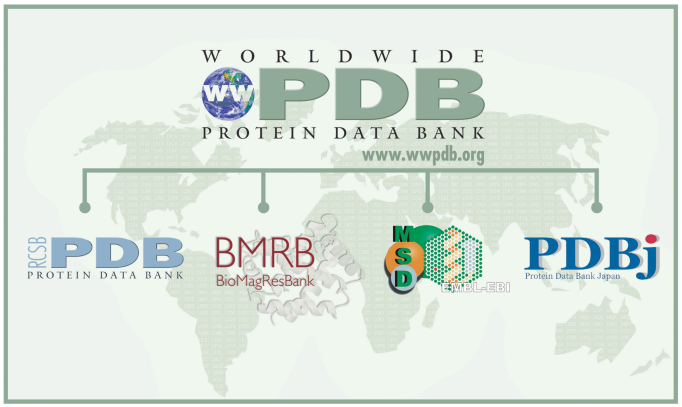
Process Deposited Entry

Chemical Components Dictionary

- Library of all polymer and non-polymer chemical components in PDB (~8700).
- Organized by PDB chemical component 3-letter identifier.
- Used to resolve residue and ligand identity and to standardize residue, ligand and atom nomenclature in PDB data files.
- Update and maintenance of the dictionary is shared by the wwPDB deposition and annotation sites.

Chemical annotation is a global effort

http://www.wwpdb.org/



The **RCSB PDB** is managed by Rutgers, The State University of New Jersey and the University of California, San Diego.

It is supported by grants from the NSF, NIGMS, Office of Science-DOE, NLM, NCI, NCRR, NIBIB, NINDS, NIDDKD.

The **BMRB** (Madison, Wisconsin) is supported by NIH grant LM05799 from the NLM.

PDBe/MSD (Cambridge, UK) is supported by grants from the Wellcome Trust, the EU (TEMBLOR, NMRQUAL and IIMS), CCP4, the BBSRC, the MRC and EMBL.

PDBj (Osaka, Japan) is supported by grant-in-aid from the Institute for Bioinformatics Research and Development, Japan Science and Technology Agency (BIRD-JST), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

RCSB PDB
PROTEIN DATA BANK

wwPDB Resources

http://www.wwpdb.org/

How is chemical information delivered by PDB?

Dictionaries and Data Files
3 Formats

PDB (ca. 1972)

mmCIF (ca. 1997)

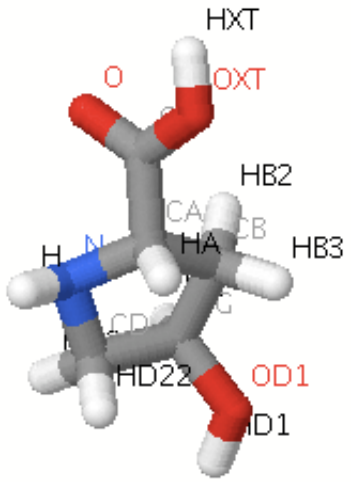
PDBML (ca. 2005)

PDB HET Dictionary Example

ftp://ftp.wwpdb.org/pub/pdb/data/monomers/

```
RESIDUE    HYP      18
CONECT       N       3 CA    CD    H
CONECT       CA      4 N     C     CB    HA
CONECT       C       3 CA    O     OXT
CONECT       O       1 C
CONECT       CB      4 CA    CG    HB2   HB3
CONECT       CG      4 CB    CD    OD1   HG
CONECT       CD      4 N     CG    HD22  HD23
CONECT       OD1     2 CG    HD1
CONECT       OXT     2 C     HXT
CONECT       H       1 N
CONECT       HA      1 CA
CONECT       HB2     1 CB
CONECT       HB3     1 CB
CONECT       HG      1 CG
CONECT       HD22    1 CD
CONECT       HD23    1 CD
CONECT       HD1     1 OD1
CONECT       HXT     1 OXT
END
```

Connectivity




```
HET    HYP             18
HETSYN     HYP HYDROXYPROLINE
HETNAM     HYP 4-HYDROXYPROLINE
FORMUL      HYP    C5 H9 N1 O3
```

Name & Formula

PDB Data File Example

```
HET    HYP  E  12       8
HETSYN     HYP HYDROXYPROLINE
HETNAM     HYP 4-HYDROXYPROLINE
FORMUL      HYP    C5 H9 N1 O3
ATOM   9302  N   HYP E  12      -6.687  -2.308  28.039  1.00 31.26           N
ATOM   9303  CA  HYP E  12      -7.953  -3.047  27.929  1.00 31.93           C
ATOM   9304  C   HYP E  12      -8.133  -3.707  26.555  1.00 32.45           C
ATOM   9305  O   HYP E  12      -7.784  -3.116  25.531  1.00 36.79           O
ATOM   9306  CB  HYP E  12      -9.041  -2.009  28.163  1.00 32.46           C
ATOM   9307  CG  HYP E  12      -8.383  -0.818  28.833  1.00 34.37           C
ATOM   9308  CD  HYP E  12      -6.885  -0.945  28.586  1.00 33.39           C
ATOM   9309  OD1 HYP E  12      -8.866   0.380  28.269  1.00 48.65           O
CONECT 9302 9303 9308
CONECT 9303 9302 9304 9306
CONECT 9304 9303 9305
CONECT 9305 9304
CONECT 9306 9303 9307
CONECT 9307 9306 9308 9309
CONECT 9308 9302 9307
CONECT 9309 9307
END
```

← Name & Formula

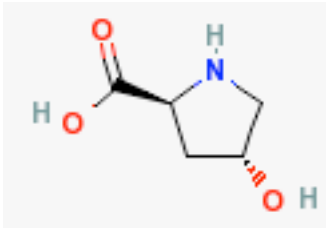
← Connectivity

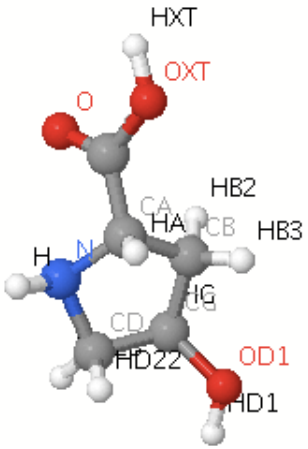
REMARK 600 – Chemical details
REMARK 610/615 – Missing & zero occupancy atoms
REMARK 620 – Metal coordination
REMARK 630 – Inhibitor description

mmCIF Component Definition Example

ftp://ftp.wwpdb.org/pub/pdb/data/monomers/components.cif

```
data_HYP
#
_chem_comp.id                                    HYP
_chem_comp.name                                  4-HYDROXYPROLINE
_chem_comp.type                                  "L-PEPTIDE LINKING"
_chem_comp.pdbx_type                             ATOMP
_chem_comp.formula                               "C5 H9 N O3"
_chem_comp.mon_nstd_parent_comp_id               PRO
_chem_comp.pdbx_synonyms                         HYDROXYPROLINE
_chem_comp.pdbx_formal_charge                    0
_chem_comp.pdbx_initial_date                     1999-07-08
_chem_comp.pdbx_modified_date                    2008-04-29
_chem_comp.pdbx_ambiguous_flag                   N
_chem_comp.pdbx_release_status                   REL
_chem_comp.pdbx_replaced_by                      ?
_chem_comp.pdbx_replaces                         ?
_chem_comp.formula_weight                        131.130
_chem_comp.one_letter_code                       ?
_chem_comp.three_letter_code                     HYP
_chem_comp.pdbx_model_coordinates_details        ?
_chem_comp.pdbx_model_coordinates_missing_flag   N
_chem_comp.pdbx_ideal_coordinates_details        Corina
_chem_comp.pdbx_ideal_coordinates_missing_flag   N
_chem_comp.pdbx_model_coordinates_db_code        1OB6
_chem_comp.pdbx_processing_site                  RCSB
#
```





mmCIF Definition Atom List Example

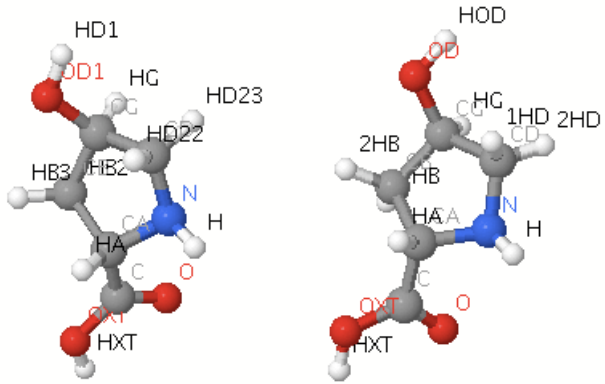
```
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.alt_atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.charge
_chem_comp_atom.pdbx_align
_chem_comp_atom.pdbx_aromatic_flag
_chem_comp_atom.pdbx_leaving_atom_flag
_chem_comp_atom.pdbx_stereo_config
_chem_comp_atom.model_Cartn_x
_chem_comp_atom.model_Cartn_y
_chem_comp_atom.model_Cartn_z
_chem_comp_atom.pdbx_model_Cartn_x_ideal
_chem_comp_atom.pdbx_model_Cartn_y_ideal
_chem_comp_atom.pdbx_model_Cartn_z_ideal
_chem_comp_atom.pdbx_ordinal
HYP N    N    N 0 1 N N N -3.366 16.585 44.188  0.168  1.360 -0.282 1
HYP CA   CA   C 0 1 N N S -2.955 15.768 43.044 -0.384 -0.003 -0.493 2
HYP C    C    C 0 1 N N N -1.447 15.609 43.030 -1.811 -0.072 -0.013 3
HYP O    O    O 0 1 N N N -0.722 16.484 43.503 -2.233  0.764  0.750 4
HYP CB   CB   C 0 1 N N N -3.408 16.578 41.829  0.515 -0.924  0.359 5
HYP CG   CG   C 0 1 N N R -4.437 17.482 42.330  1.847 -0.159  0.505 6
HYP CD   CD   C 0 1 N N N -4.068 17.803 43.753  1.640  1.159 -0.271 7
HYP OD1  OD   O 0 1 N N N -5.693 16.815 42.294  2.917 -0.911 -0.071 8
HYP OXT  OXT  O 0 1 N Y N -0.976 14.502 42.469 -2.614 -1.063 -0.433 9
HYP H    H    H 0 1 N Y N -3.980 16.047 44.765 -0.107  1.981 -1.028 10
HYP HA   HA   H 0 1 N N N -3.385 14.756 43.068 -0.325 -0.278 -1.546 11
HYP HB2  1HB  H 0 1 N N N -2.567 17.141 41.398  0.066 -1.092  1.337 12
HYP HB3  2HB  H 0 1 N N N -3.790 15.930 41.026  0.678 -1.873 -0.153 13
HYP HG   HG   H 0 1 N N N -4.508 18.399 41.726  2.052  0.048  1.555 14
HYP HD22 1HD  H 0 0 N N N -4.956 18.005 44.370  2.018  1.065 -1.289 15
HYP HD23 2HD  H 0 0 N N N -3.457 18.713 43.848  2.132  1.985  0.243 16
HYP HD1  HOD  H 0 1 N N N -5.999 16.666 43.181  3.780 -0.479 -0.009 17
HYP HXT  HXT  H 0 1 N N N -0.027 14.511 42.499 -3.520 -1.066 -0.098 18
#
```

Atom names

Stereochemistry & aromaticity

Model coordinates

Ideal coordinates



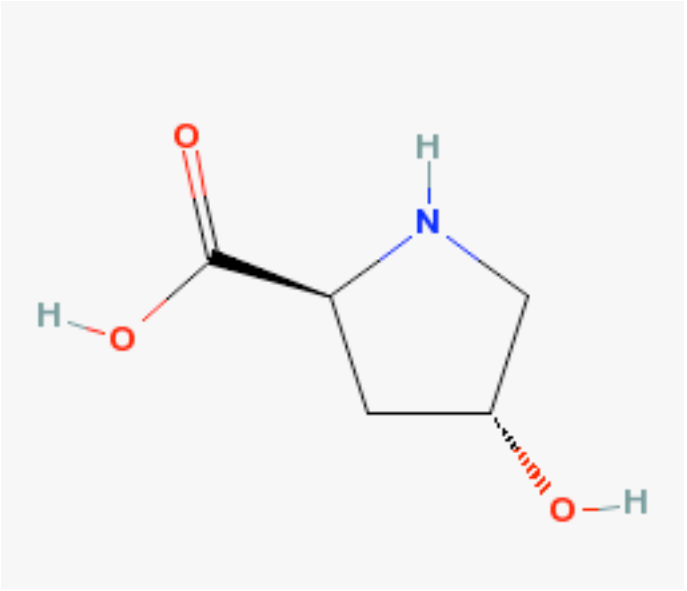
mmCIF Definition Connectivity Table

```
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.value_order
_chem_comp_bond.pdbx_aromatic_flag
_chem_comp_bond.pdbx_stereo_config
_chem_comp_bond.pdbx_ordinal
HYP N    CA    SING N N 1
HYP N    CD    SING N N 2
HYP N    H     SING N N 3
HYP CA   C     SING N N 4
HYP CA   CB    SING N N 5
HYP CA   HA    SING N N 6
HYP C    O     DOUB N N 7
HYP C    OXT   SING N N 8
HYP CB   CG    SING N N 9
HYP CB   HB2   SING N N 10
HYP CB   HB3   SING N N 11
HYP CG   CD    SING N N 12
HYP CG   OD1   SING N N 13
HYP CG   HG    SING N N 14
HYP CD   HD22  SING N N 15
HYP CD   HD23  SING N N 16
HYP OD1  HD1   SING N N 17
HYP OXT  HXT   SING N N 18
```

Connected atoms

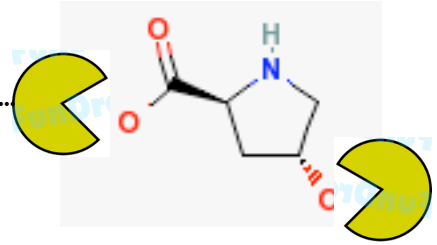
Bond type

Stereochemistry & aromaticity



Descriptors and Identifiers

…O=C(O)C1NCC(O)C1 …



…InChI=1/C5H9NO3…

```
#
loop_
_pdbx_chem_comp_descriptor.comp_id
_pdbx_chem_comp_descriptor.type
_pdbx_chem_comp_descriptor.program
_pdbx_chem_comp_descriptor.program_version
_pdbx_chem_comp_descriptor.descriptor
HYP SMILES           ACDLabs               10.04 O=C(O)C1NCC(O)C1
HYP SMILES_CANONICAL CACTVS                3.341 O[C@H]1CN[C@@H](C1)C(O)=O
HYP SMILES           CACTVS                3.341 O[CH]1CN[CH](C1)C(O)=O
HYP SMILES_CANONICAL "OpenEye OEToolkits" 1.5.0 C1[C@H](CN[C@@H]1C(=O)O)O
HYP SMILES           "OpenEye OEToolkits" 1.5.0 C1C(CNC1C(=O)O)O
HYP InChI            InChI                1.02b "InChI=1/C5H9NO3/c7 …. (H,8,9)/t3-,4+/m1/s1/f/h8H"
HYP InChIKey         InChI                1.02b PMMYEEVYMWASQN-FJFIVYRMDE
#
loop_
_pdbx_chem_comp_identifier.comp_id
_pdbx_chem_comp_identifier.type
_pdbx_chem_comp_identifier.program
_pdbx_chem_comp_identifier.program_version
_pdbx_chem_comp_identifier.identifier
HYP "SYSTEMATIC NAME" ACDLabs               10.04 (4R)-4-hydroxy-L-proline
HYP "SYSTEMATIC NAME" "OpenEye OEToolkits" 1.5.0 "(2S,4R)-4-hydroxypyrrolidine-2-carboxylic acid"
#
```

SMILES

InChI

Systematic Names

Descriptors

- SMILES - Simplified molecular input line entry specification
  - Weininger, D. (1988), SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28, 31-36
  - http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html.
- InChI - IUPAC International Chemical Identifier
  - Stephen E. Stein, Stephen R. Heller, and Dmitrii Tchekhovskoi, An Open Standard for Chemical Structure Representation: The IUPAC Chemical Identifier, in Proceedings of the 2003 International Chemical Information Conference (Nimes), pp. 131-143.
  - http://old.iupac.org/inchi/
- InChiKey – InChI digital signature
  - http://old.iupac.org/inchi/release102.html

Other Features

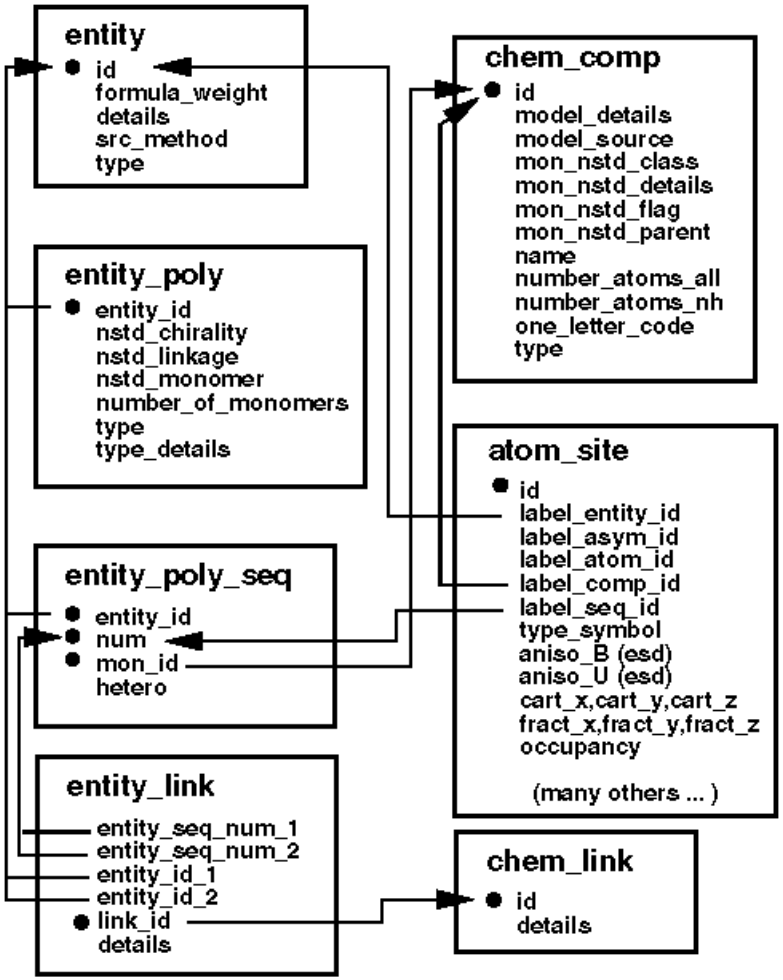
New in V3.2

```
loop_
_pdbx_chem_comp_feature.comp_id
_pdbx_chem_comp_feature.type
_pdbx_chem_comp_feature.value
_pdbx_chem_comp_feature.source
00X   'ENZYME INHIBITED'    'ASPARTIC PROTEINASE'                    PDB
00X   'FUNCTION'            'Transition-state analogue inhibitor'    PDB
00X   'STRUCTURE IMAGE URL' 'http://journals.iucr.org/00X.jpg'       IUCR
```

Key Categories in mmCIF Data Files

- entity – unique polymer and non -polymer molecules
- entity_poly, entity_poly_seq – polymer features and sequence
- chem_comp – chemical component names, identifiers and features
- atom_site – Cartesian coordinates
- struct_asym – list of molecular instances
- struct_conn – intermolecular linkages

Dictionary details are not reproduced in each data file!



PDBML Schemas

http://pdbml.pdb.org/

- Schemas provided for the PDB Exchange dictionary and component dictionaries
- Schemas and data dictionaries are updated synchronously

PDBML: the representation of archival macromolecular structure data in XML. John Wesbrook, Nobutoshi Ito, Haruki Nakamura, Kim Henrick and Helen M. Berman, Bioinformatics, 21(7), 988-992, 2005.

PDBML

PDBML Component Definition Example

```
<PDBx:datablock datablockName="HYP" xsi:schemaLocation=
   http://deposit.pdb.org/pdbML/pdbx.xsd pdbx.xsd>
<PDBx:chem_compCategory>
  <PDBx:chem_comp id="HYP">
    <PDBx:name>4-HYDROXYPROLINE</PDBx:name>
    <PDBx:type>L-peptide linking</PDBx:type>
    <PDBx:pdbx_type>ATOMP</PDBx:pdbx_type>
    <PDBx:formula>C5 H9 N O3</PDBx:formula>
    <PDBx:mon_nstd_parent_comp_id>PRO</PDBx:mon_nstd_parent_comp_id>
    <PDBx:pdbx_synonyms>HYDROXYPROLINE</PDBx:pdbx_synonyms>
    <PDBx:pdbx_formal_charge>0</PDBx:pdbx_formal_charge>
    <PDBx:pdbx_initial_date>1999-07-08</PDBx:pdbx_initial_date>
    <PDBx:pdbx_modified_date>2008-04-29</PDBx:pdbx_modified_date>
    <PDBx:pdbx_ambiguous_flag>N</PDBx:pdbx_ambiguous_flag>
    <PDBx:pdbx_release_status>REL</PDBx:pdbx_release_status>
    <PDBx:formula_weight>131.130</PDBx:formula_weight>
    <PDBx:three_letter_code>HYP</PDBx:three_letter_code>
    <PDBx:pdbx_model_coordinates_missing_flag>N</PDBx:pdbx_model_coordinates_missing_flag>
    <PDBx:pdbx_ideal_coordinates_details>Corina</PDBx:pdbx_ideal_coordinates_details>
    <PDBx:pdbx_ideal_coordinates_missing_flag>N</PDBx:pdbx_ideal_coordinates_missing_flag>
  </PDBx:chem_comp>
</PDBx:chem_compCategory>
```

PDBML Data File Example

```
loop_
_entity_poly.entity_id
_entity_poly.type
_entity_poly.nstd_linkage
_entity_poly.nstd_monomer
_entity_poly.pdbx_seq_one_letter_code
_entity_poly.pdbx_seq_one_letter_code_can
1 polypeptide(L) no no
;DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADG
VPSRFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWSTPRTFGGGTKLEIK
;
;DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADG
VPSRFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWSTPRTFGGGTKLEIK
;
```

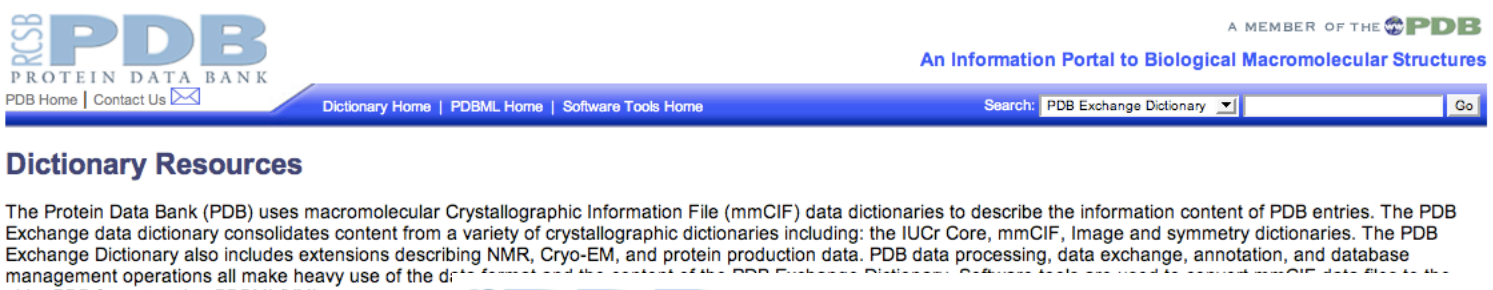
```
  <PDBx:entity_polyCategory>
       <PDBx:entity_poly entity_id="1">
          <PDBx:type>polypeptide(L)</PDBx:type>
          <PDBx:nstd_linkage>no</PDBx:nstd_linkage>
          <PDBx:nstd_monomer>no</PDBx:nstd_monomer>
          <PDBx:pdbx_seq_one_letter_code>
           DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADG
           VPSRFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWSTPRTFGGGTKLEIK
          </PDBx:pdbx_seq_one_letter_code>
          <PDBx:pdbx_seq_one_letter_code_can>
           DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADG
           VPSRFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWSTPRTFGGGTKLEIK
          </PDBx:pdbx_seq_one_letter_code_can>
       </PDBx:entity_poly>
    </PDBx:entity_polyCategory>
```

Standard container for each category

Keys mapped as attributes

Dictionary and Schema Resources

http://mmcif.pdb.org & http://pdbml.pdb.org



RCSB PDB PROTEIN DATA BANK

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

PDB Home | Contact Us

Dictionary Home | PDBML Home | Software Tools Home

Search: PDB Exchange Dictionary Go

Dictionary Resources

The Protein Data Bank (PDB) uses macromolecular Crystallographic Information File (mmCIF) data dictionaries to describe the information content of PDB entries. The PDB Exchange data dictionary consolidates content from a variety of crystallographic dictionaries including: the IUCr Core, mmCIF, Image and symmetry dictionaries. The PDB Exchange Dictionary also includes extensions describing NMR, Cryo-EM, and protein production data. PDB data processing, data exchange, annotation, and database management operations all make heavy use of the d
older PDB format and to PDBML/XML.

- Data files in mmCIF format can be downloadec
- Software tools are available for preparing and
- Software tools are available for converting mm
- A complete list of PDB software tools for mana

Dictionary Content and Representation

- Background and Introduction about mmCIF
- The Macromolecular Crystallographic Informat
- STAR/mmCIF: An Extensive Ontology for Macr
- mmCIF Software Developers Workshop 1997
- mmCIF Dictionary Templates
- mmCIF Examples
- References

Data Dictionaries

RCSB PDB PROTEIN DATA BANK

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

PDB Home | Contact Us

Dictionary Home | PDBML Home | Software Tools Home

PDBML

PDBML Resources

PDBML

The Protein Data Bank Markup Language (PDBML) provides a representation of PDB data in XML format. The description of this format is provided in XML schema of the PDB Exchange Data Dictionary. This schema is produced by direct translation of the mmCIF format PDB Exchange Data Dictionary Other data dictionaries used by the PDB have been electronically translated into XML/XSD schemas and these are also presented in the list below.

- PDBML data files are provided in three forms:
  - fully marked-up files,
  - files without atom records
  - files with a more space efficient encoding of atom records
- Data files in PDBML format can be downloaded from the RCSB PDB website or by ftp.
- Software tools for manipulating PDB data in XML format can be found here.
- An article describing PDBML is available.
  PDBML: the representation of archival macromolecular structure data in XML.
  John Wesbrook, Nobutoshi Ito, Haruki Nakamura, Kim Henrick and Helen M. Berman,
  Bioinformatics, 21(7), 988-992, 2005.

PDBML Schema

- **PDB Exchange Dictionary** | | current version | previous versions | alternative atom record markup ||
  XML Schema for Exchange Data dictionary developed as a collaboration
  between MSD-EBI, PDBj and RCSB and used by wwPDB members for data exchange.

Other Supporting Software Tools

http://sw-tools.rcsb.org

- Validating Parsers for Files and Dictionaries
- Dictionary access and presentation tools
- File format translation tools
- Database Builder and Loader
- XML translation tool for data files and dictionaries
- Others: BioPerl, BioPython, mmLib (py), CCP4

RCSB PDB
PROTEIN DATA BANK

Ligand Expo
http://ligand-expo.rcsb.org

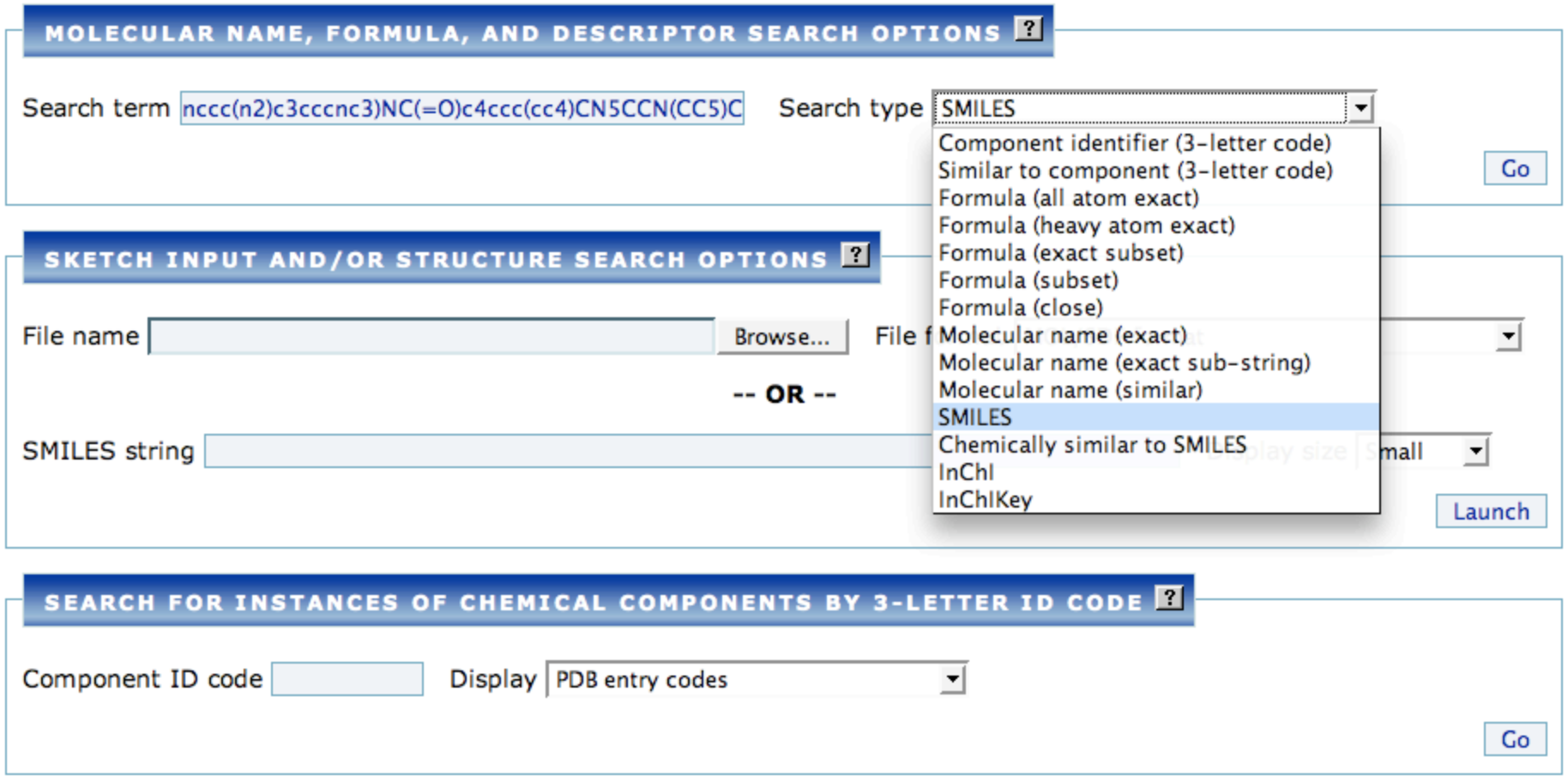
Ligand Expo presents a view of PDB data focusing chemical and structural information about small molecules within the structure entries of the Protein Data Bank. Tools are provided to browse and search the PDB dictionary of chemical components, to identify structure entries containing particular small molecules, and to download the 3D structures of the small molecule components within PDB entries. Tools are also provided to build new chemical definitions for PDB chemical dictionary.

Chemical Dictionary Search

- Molecular Name
- Formula
- SMILES
- InChI/InChIKey
- PDB Component Identifier
- Chemical substructure

RCSB PDB
PROTEIN DATA BANK

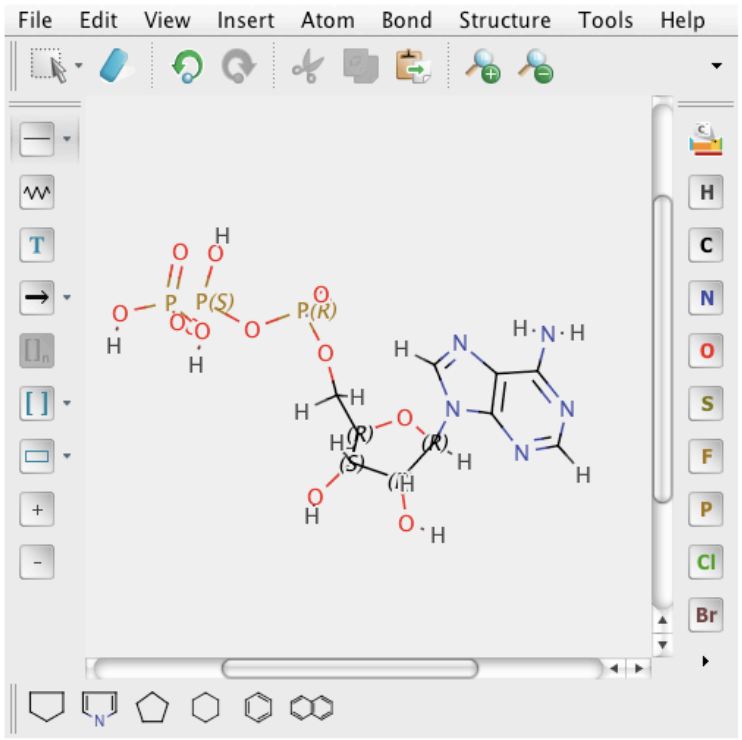
Search Options



RCSB PDB
PROTEIN DATA BANK

Chemical Component Sketch Tool

Component Building & Substructure Search



SUBSTRUCTURE SEARCH OPTIONS

Search type: Strict (all atom/exact bond types)

Search

SAVE CHEMICAL COMPONENT

Export format: mmCIF component definition

Optional Component ID (3-letter-code): ATP

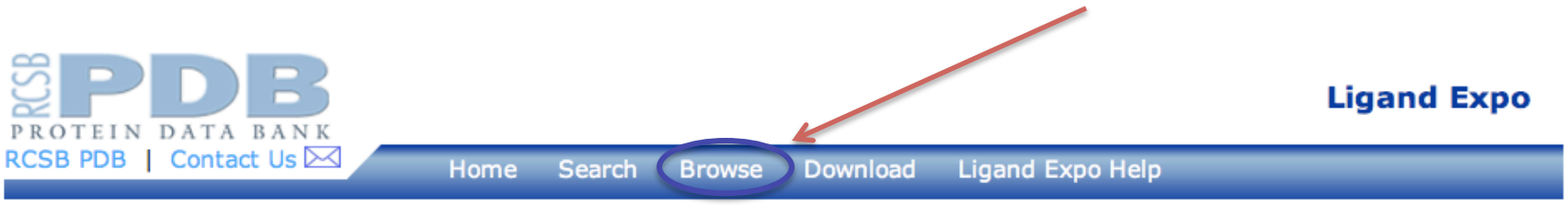
Save file

Browsing Options

Predefined queries based on SMILES patterns are provided in menus to permit convenient browsing of molecules containing:

- Amino acids
- Nucleotides
- Selected top-selling pharmaceuticals
- Common aromatic ring systems

Example: Browsing Popular Drugs



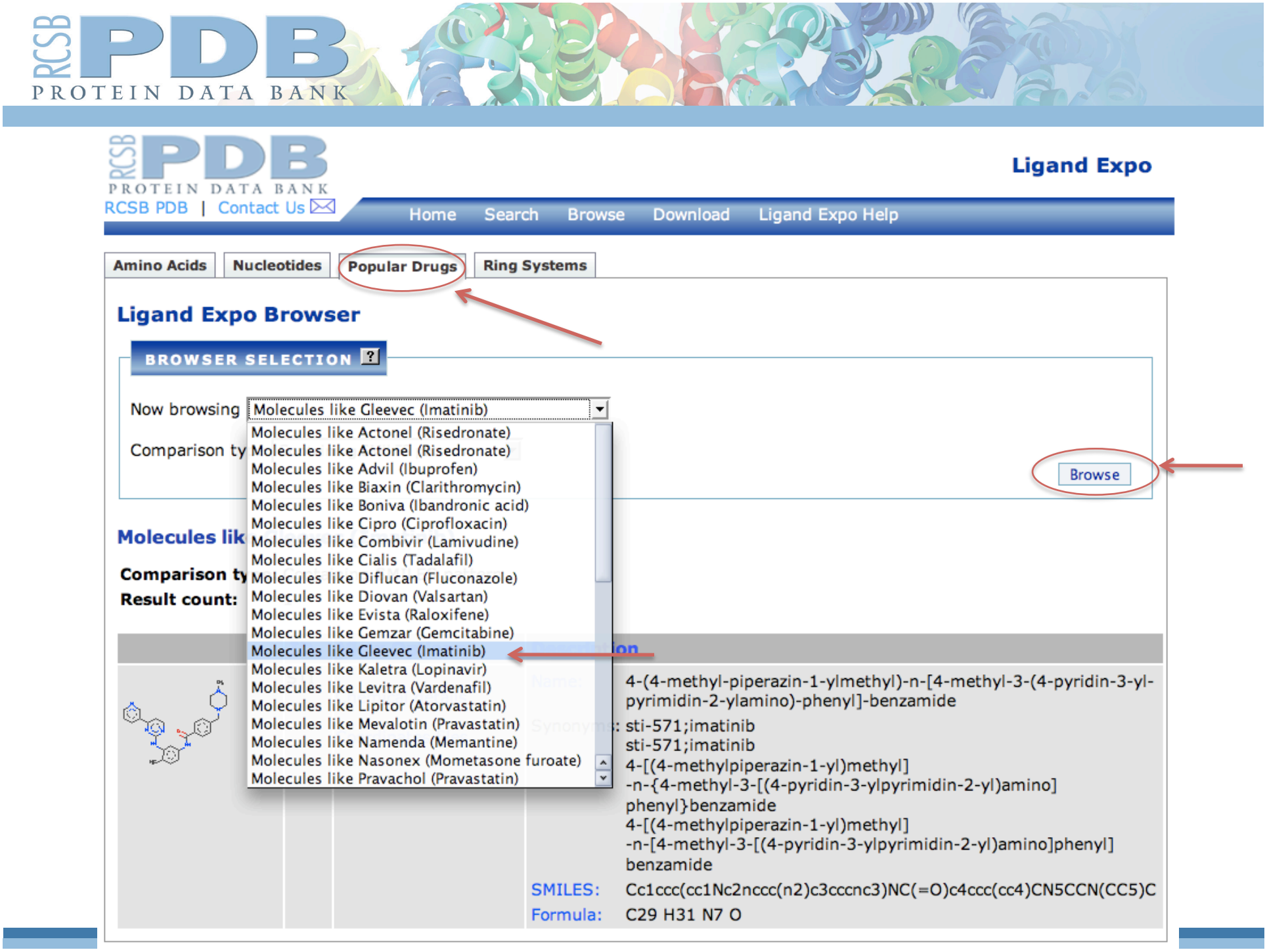
Ligand Expo

Version 1 Updated: 2-May-2008

Ligand Expo (formerly Ligand Depot) provides chemical and structural information about small molecules within the structure entries of the Protein Data Bank. Tools are provided to search the PDB dictionary for chemical components, to identify structure entries containing particular small molecules, and to download the 3D structures of the small molecule components in the PDB entry. A sketch tool is also provided for building new chemical definitions from reported PDB chemical components.

Getting Started:

The Ligand Expo Tutorial contains a summary of query, report and download features.
Select the Search menu to find a chemical component or build a new chemical component.
Select the Browse menu to view tabulations of modified amino acids and nucleotides.
Select the Download for links to chemical dictionaries and resource files.

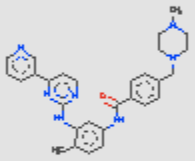

RCSB PDB
PROTEIN DATA BANK
RCSB PDB
PROTEIN DATA BANK
RCSB PDB | Contact Us
Ligand Expo
Home
Search
Browse
Download
Ligand Expo Help
Amino Acids
Nucleotides
Popular Drugs
Ring Systems
Ligand Expo Browser
BROWSER SELECTION
Now browsing
Molecules like Gleevec (Imatinib)
Comparison ty
Molecules like Actonel (Risedronate)
Molecules like Actonel (Risedronate)
Molecules like Advil (Ibuprofen)
Molecules like Biaxin (Clarithromycin)
Molecules like Boniva (Ibandronic acid)
Molecules like Cipro (Ciprofloxacin)
Molecules like Combivir (Lamivudine)
Molecules like Cialis (Tadalafil)
Molecules like Diflucan (Fluconazole)
Molecules like Diovan (Valsartan)
Molecules like Evista (Raloxifene)
Molecules like Gemzar (Gemcitabine)
Molecules like Gleevec (Imatinib)
Molecules like Kaletra (Lopinavir)
Molecules like Levitra (Vardenafil)
Molecules like Lipitor (Atorvastatin)
Molecules like Mevalotin (Pravastatin)
Molecules like Namenda (Memantine)
Molecules like Nasonex (Mometasone furoate)
Molecules like Pravachol (Pravastatin)
Browse
Molecules lik
Comparison ty
Result count:
4-(4-methyl-piperazin-1-ylmethyl)-n-[4-methyl-3-(4-pyridin-3-yl-pyrimidin-2-ylamino)-phenyl]-benzamide
sti-571;imatinib
sti-571;imatinib
4-[(4-methylpiperazin-1-yl)methyl]
-n-{4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]
phenyl}benzamide
4-[(4-methylpiperazin-1-yl)methyl]
-n-[4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]phenyl]
benzamide
SMILES:
Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C
Formula:
C29 H31 N7 O

RCSB PDB
PROTEIN DATA BANK

Search report summary

Molecules like Gleevec (Imatinib)

Comparison type: Containing SMILES pattern
Result count: 1

	ID	View Options	Description
	STI	Chemical details Coordinates files	Name: 4-(4-methyl-piperazin-1-ylmethyl)-n-[4-methyl-3-(4-pyridin-3-yl-pyrimidin-2-ylamino)-phenyl]-benzamide Synonyms: sti-571;imatinib sti-571;imatinib 4-[(4-methylpiperazin-1-yl)methyl] -n-{4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino] phenyl}benzamide 4-[(4-methylpiperazin-1-yl)methyl] -n-[4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]phenyl] benzamide SMILES: Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C Formula: C29 H31 N7 O

Chemical Details

RCSB PDB
PROTEIN DATA BANK

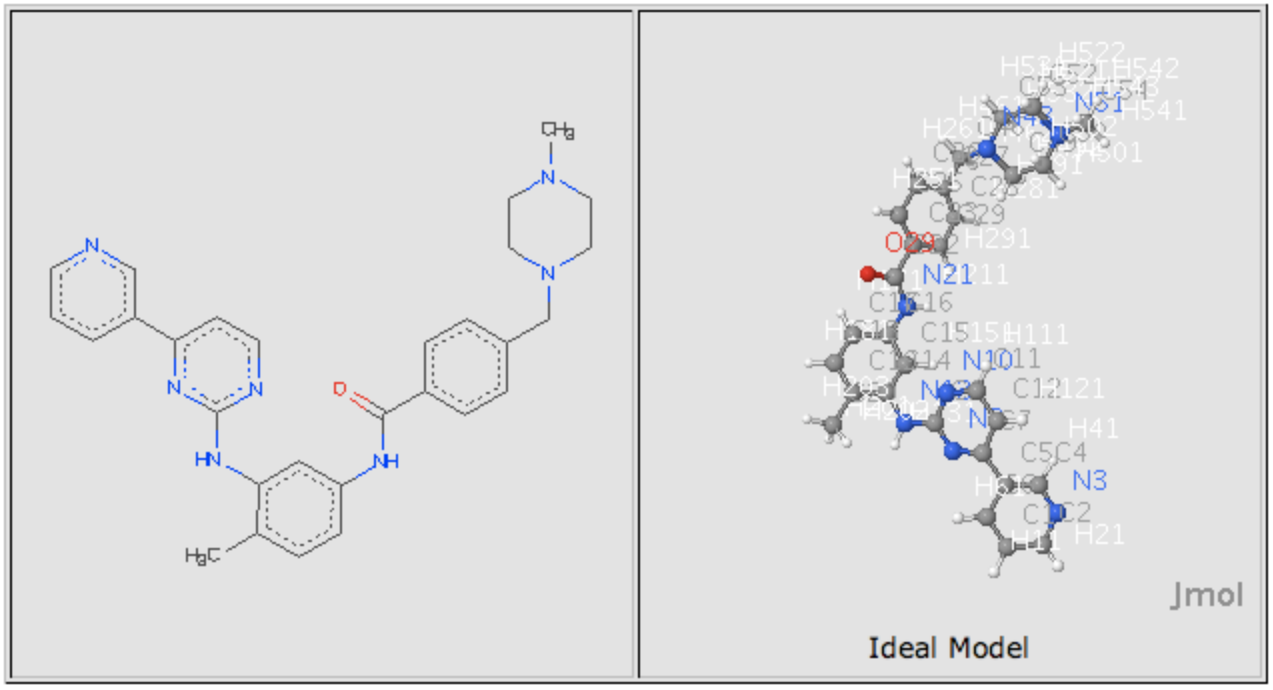
Ligand Expo

RCSB PDB | Contact Us

Home | Search | Browse | Download | Ligand Expo Help

Chemical Details | Geometry | Atom Nomenclature | Downloads | Related Resources

PDB Chemical Component STI



RCSB PDB
PROTEIN DATA BANK

Chemical Details

Chemical Description

Name	4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE
Synonyms	STI-571;IMATINIB
Formula	C29 H31 N7 O
Formal charge	0
Molecular weight	493.603 g/mol
Component type	NON-POLYMER

Chemical features

Atom count	68
Chiral atom count	0
Bond count	72
Aromatic bond count	24

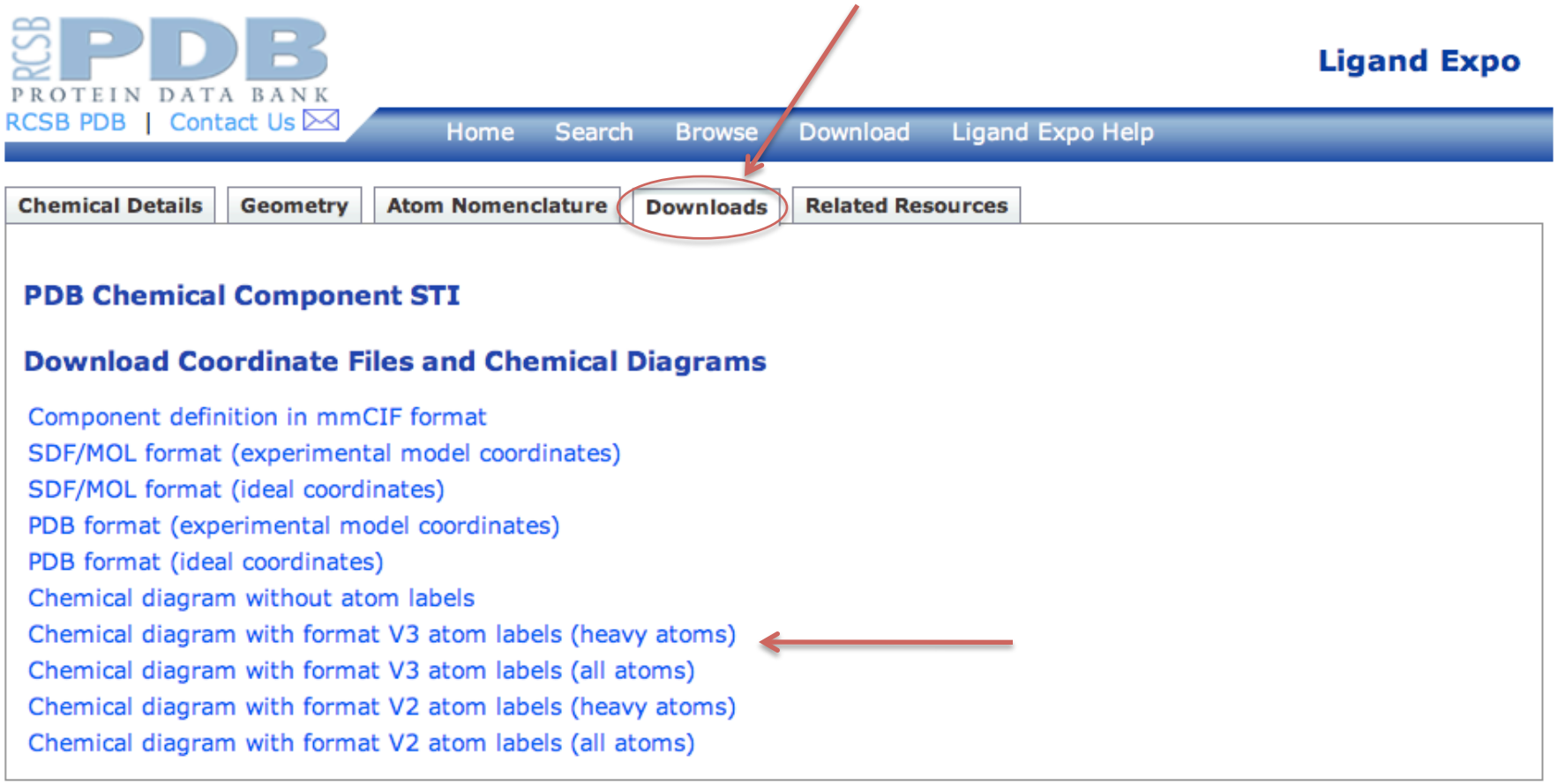
Chemical Identifiers

Systematic name (ACDLabs)	4-[(4-methylpiperazin-1-yl)methyl]-N-{4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]phenyl}benzamide
Systematic name (OpenEye OEToolkits)	4-[(4-methylpiperazin-1-yl)methyl]-N-[4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]phenyl]benzamide

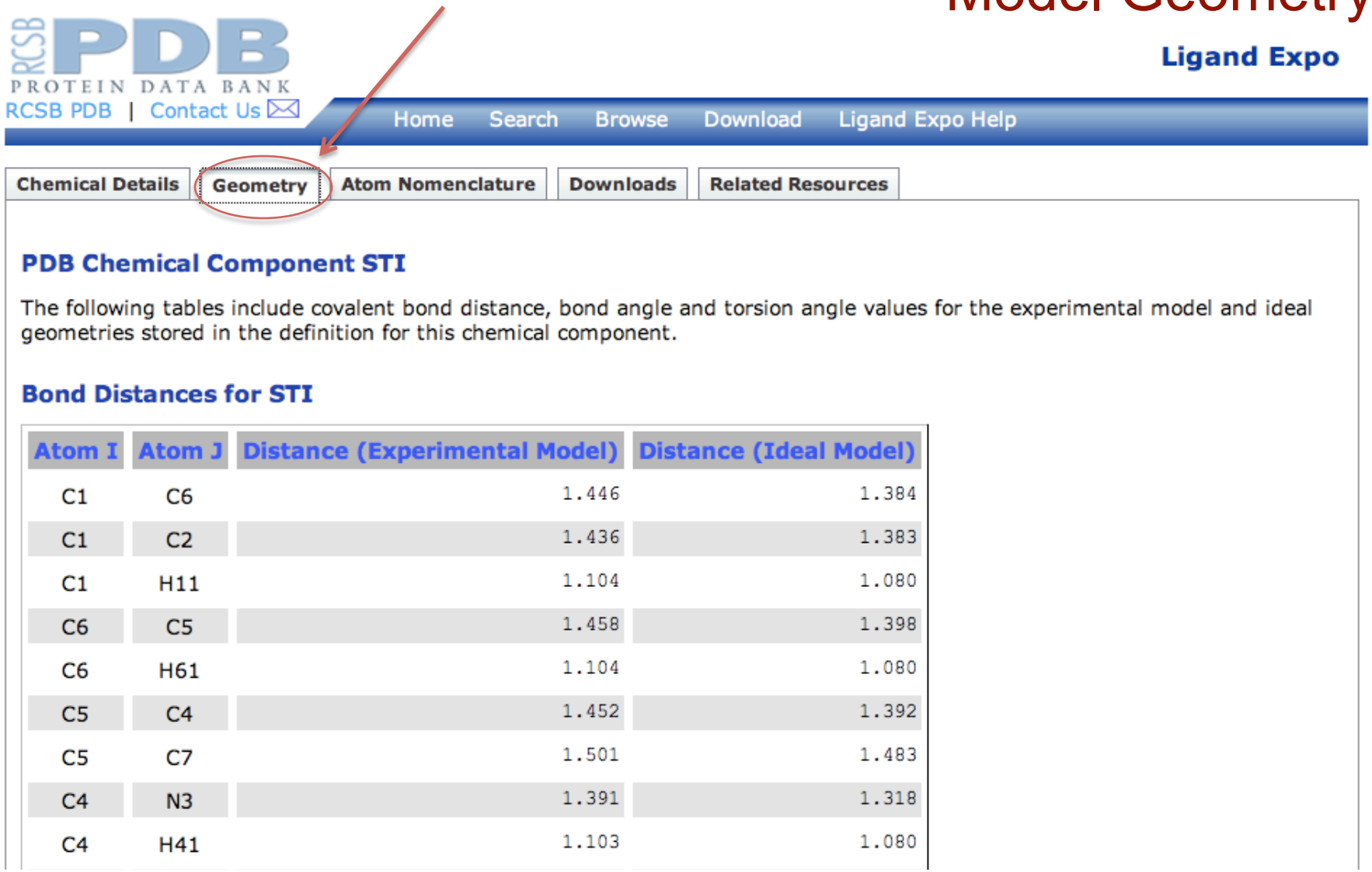
Chemical Descriptors

Stereo SMILES (CACTVS)	CN1CCN(CC1)Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4nccc(n4)c5cccnc5)c3
SMILES (CACTVS)	CN1CCN(CC1)Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4nccc(n4)c5cccnc5)c3
Stereo SMILES (OpenEye)	Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C
InChI descriptor	InChI=1/C29H31N7O/c1-21-5-10-25(18-27(21)34-29-31-13-11-26(33-29)24-4-3-12-30-19-24)32-28(37)23-8-6-22(7-9-23)20-36-16-14-35(2)15-17-36/h3-13,18-19H,14-17,20H2,1-2H3,(H,32,37)(H,31,33,34)/f/h32,34H
InChIKey descriptor	KTUFNOKKBVMGRW-RPGFEBOUCX

Downloads


RCSB PDB
PROTEIN DATA BANK
RCSB PDB | Contact Us
Ligand Expo
Home
Search
Browse
Download
Ligand Expo Help
Chemical Details
Geometry
Atom Nomenclature
Downloads
Related Resources
PDB Chemical Component STI
Download Coordinate Files and Chemical Diagrams
Component definition in mmCIF format
SDF/MOL format (experimental model coordinates)
SDF/MOL format (ideal coordinates)
PDB format (experimental model coordinates)
PDB format (ideal coordinates)
Chemical diagram without atom labels
Chemical diagram with format V3 atom labels (heavy atoms)
Chemical diagram with format V3 atom labels (all atoms)
Chemical diagram with format V2 atom labels (heavy atoms)
Chemical diagram with format V2 atom labels (all atoms)

Model Geometry



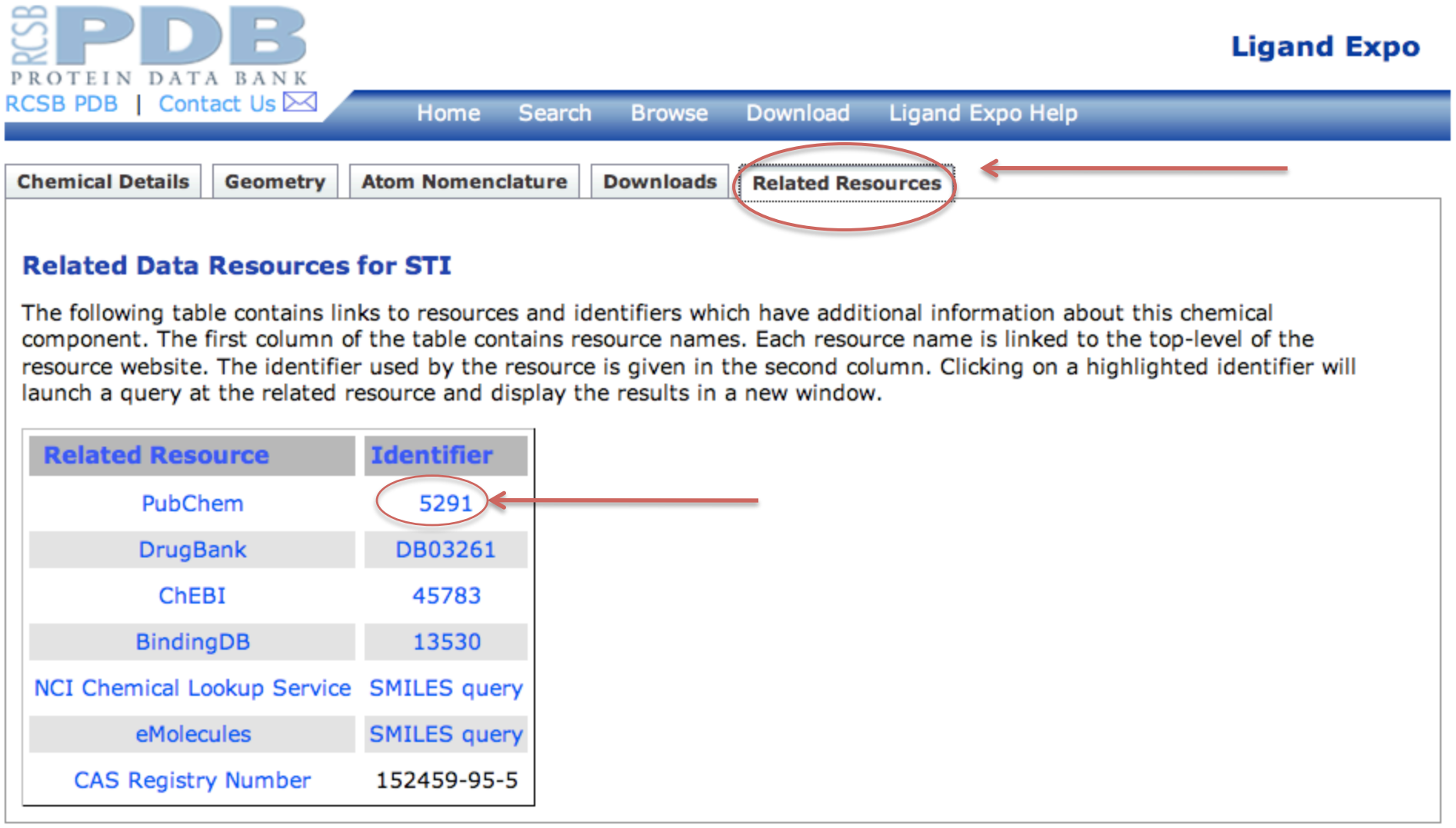
PDB Chemical Component STI

The following tables include covalent bond distance, bond angle and torsion angle values for the experimental model and ideal geometries stored in the definition for this chemical component.

Bond Distances for STI

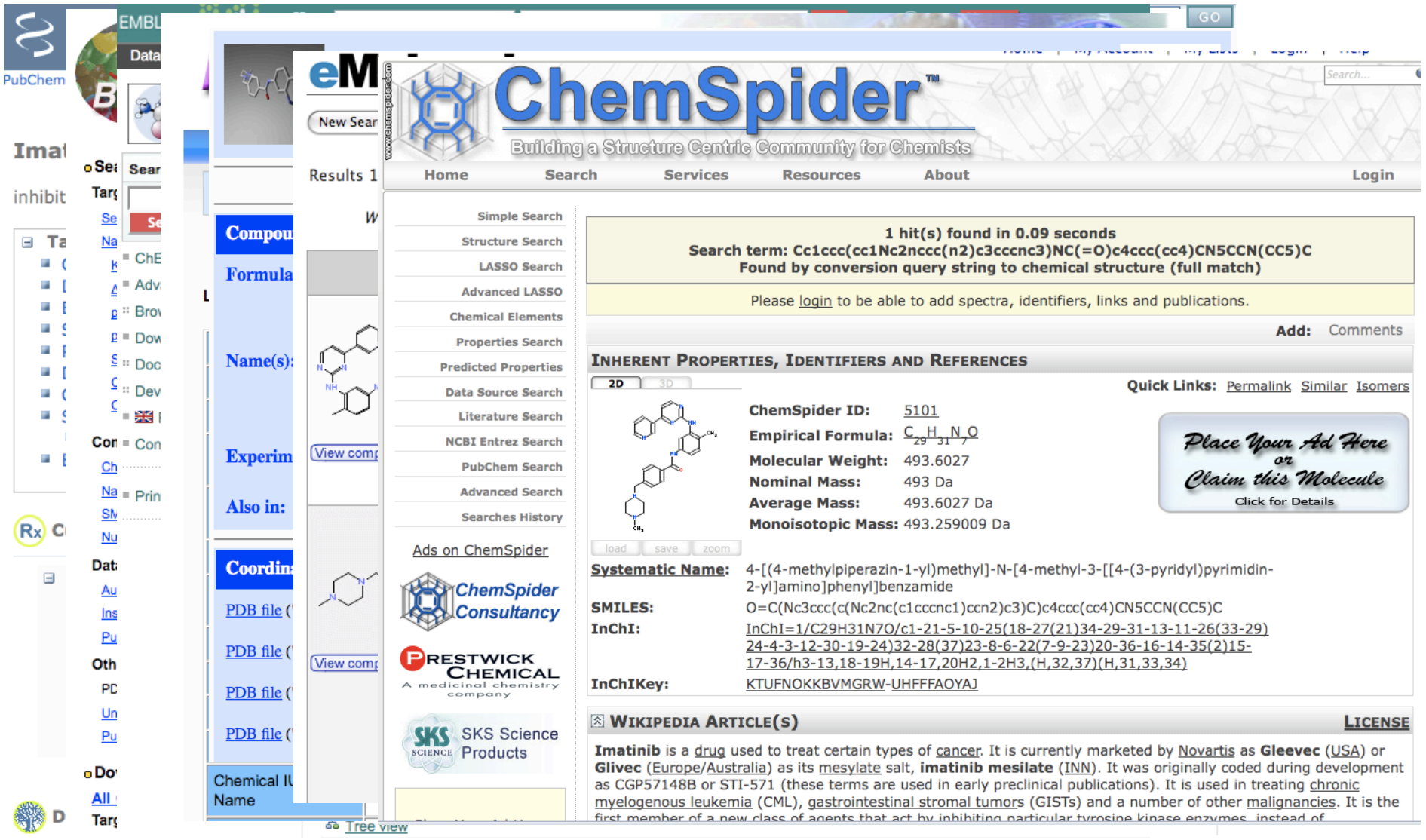
Atom I	Atom J	Distance (Experimental Model)	Distance (Ideal Model)
C1	C6	1.446	1.384
C1	C2	1.436	1.383
C1	H11	1.104	1.080
C6	C5	1.458	1.398
C6	H61	1.104	1.080
C5	C4	1.452	1.392
C5	C7	1.501	1.483
C4	N3	1.391	1.318
C4	H41	1.103	1.080

Related Resources



RCSB PDB
PROTEIN DATA BANK

Related Resources



RCSB PDB
PROTEIN DATA BANK

Experimental Examples

Search Result Summary

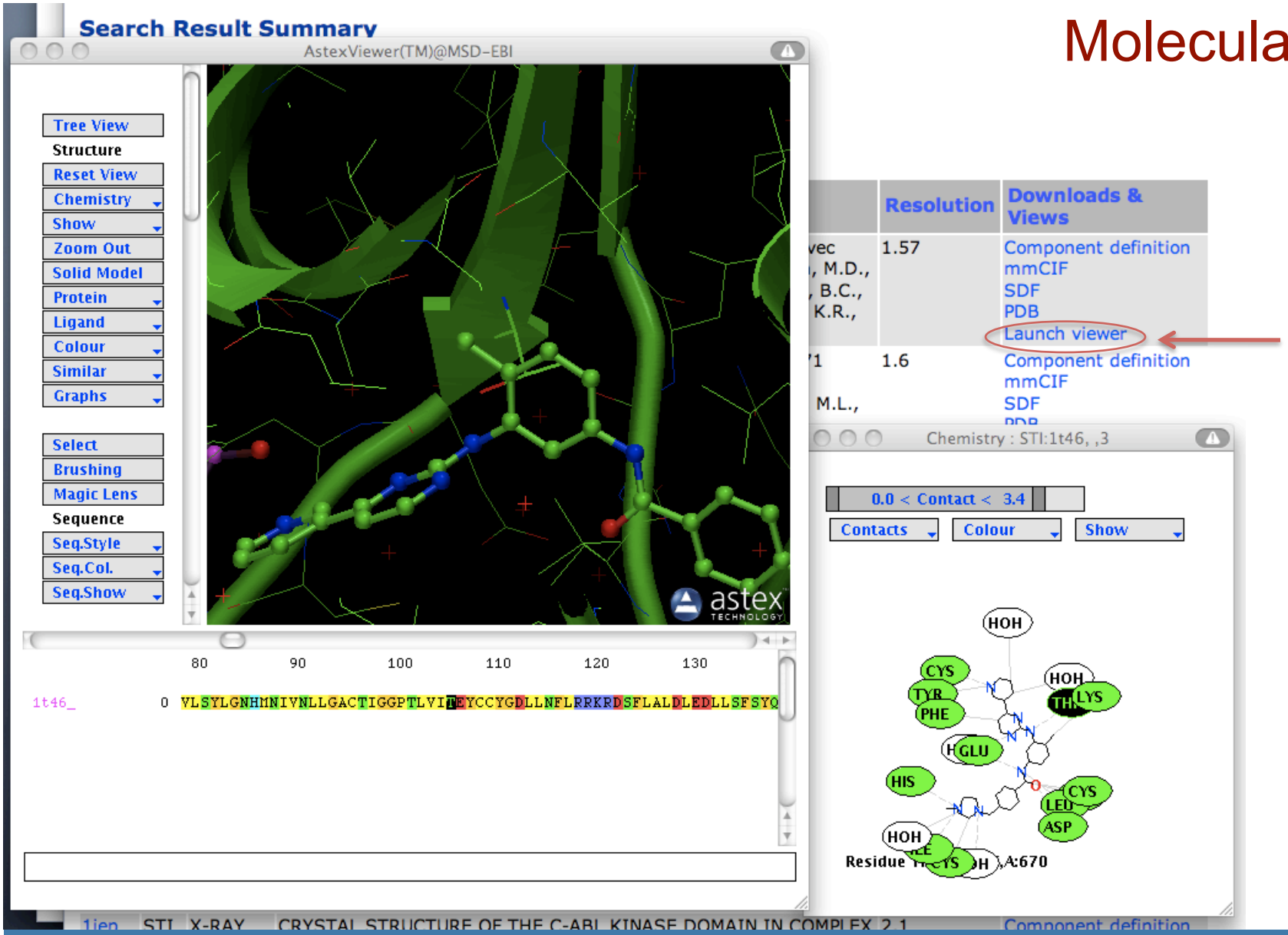
Target ID: STI
Count in released entries: 12

Instances of STI in Released Entries

PDB ID	CC ID	Method	Title/Authors	Resolution	Downloads & Views
1xbb	STI	X-RAY	Crystal structure of the syk tyrosine kinase domain with Gleevec Nienaber, V.L., Atwell, S., Adams, J.M., Badger, J., Buchanan, M.D., Feil, I.K., Froning, K.J., Gao, X., Hendle, J., Keegan, K., Leon, B.C., Muller-Deickmann, H.J., Noland, B.W., Post, K., Rajashankar, K.R., Ramos, A., Russell, M., Burley, S.K., Buchanan, S.G.	1.57	Component definition mmCIF SDF PDB Launch viewer
1t46	STI	X-RAY	STRUCTURAL BASIS FOR THE AUTOINHIBITION AND STI-571 INHIBITION OF C-KIT TYROSINE KINASE Mol, C.D., Dougan, D.R., Schneider, T.R., Skene, R.J., Kraus, M.L., Scheibe, D.N., Snell, G.P., Zou, H., Sang, B.C., Wilson, K.P.	1.6	Component definition mmCIF SDF PDB Launch viewer
1opj	STI	X-RAY	Structural basis for the auto-inhibition of c-Abl tyrosine kinase Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.	1.75	Component definition mmCIF SDF PDB Launch viewer
1opj	STI	X-RAY	Structural basis for the auto-inhibition of c-Abl tyrosine kinase Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.	1.75	Component definition mmCIF SDF PDB Launch viewer
2oiq	STI	X-RAY	Crystal Structure of chicken c-Src kinase domain in complex with the cancer drug imatinib.	2.07	Component definition mmCIF

RCSB PDB
PROTEIN DATA BANK

Molecular View



RCSB PDB
PROTEIN DATA BANK

Experimental Examples

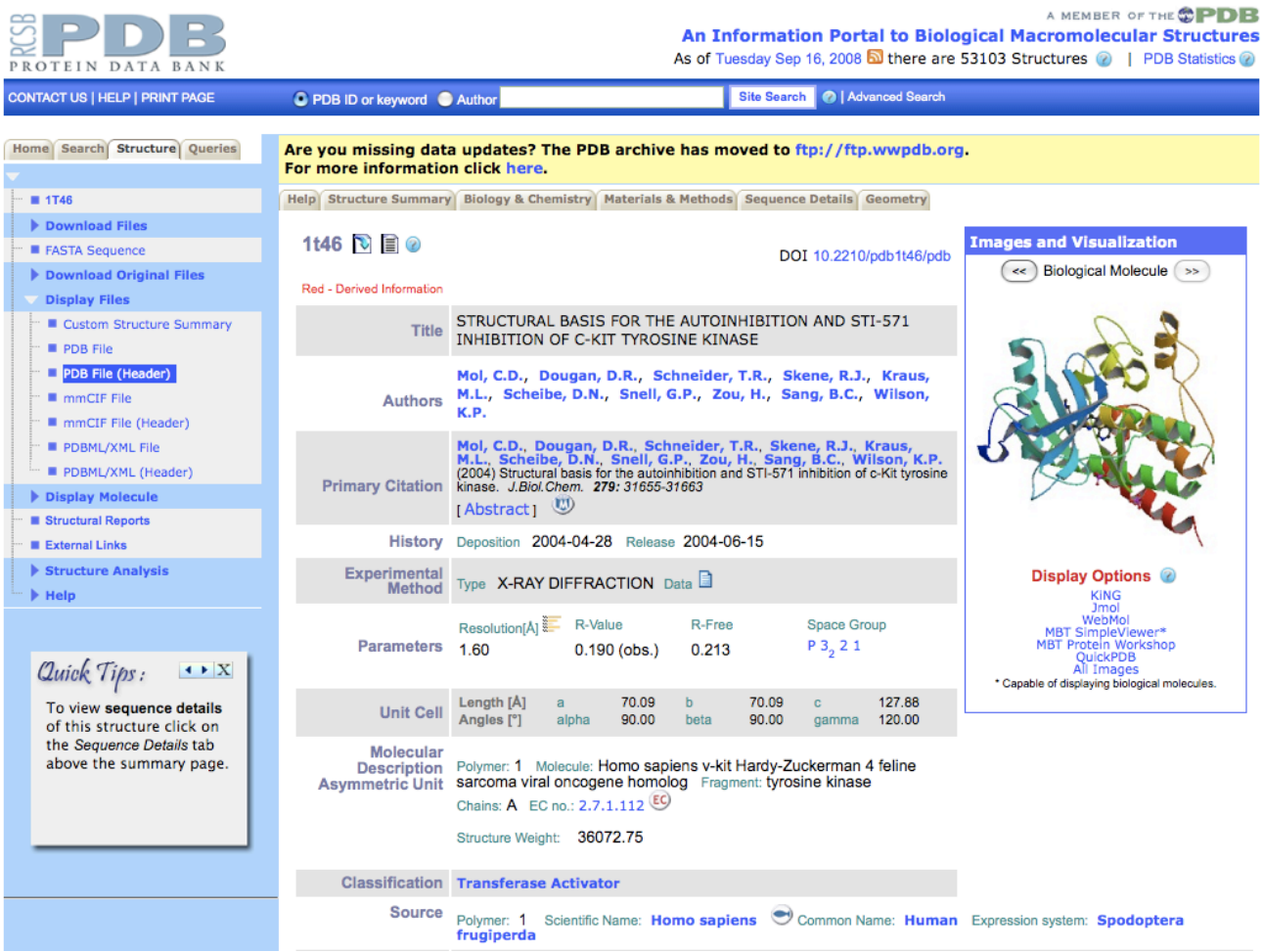
Search Result Summary

Target ID: STI
Count in released entries: 12

Instances of STI in Released Entries

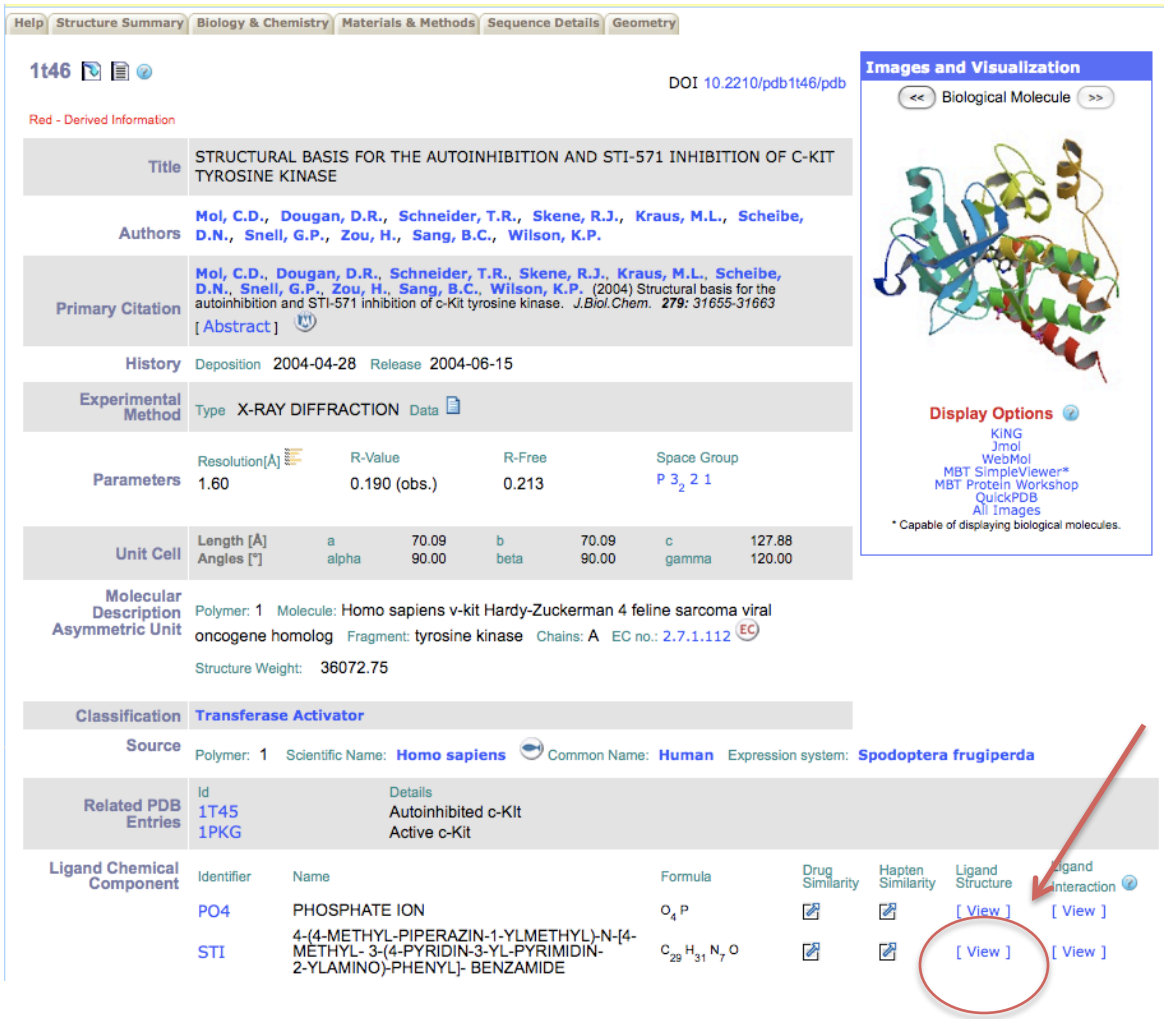
PDB ID	CC ID	Method	Title/Authors	Resolution	Downloads & Views
1xbb	STI	X-RAY	Crystal structure of the syk tyrosine kinase domain with Gleevec Nienaber, V.L., Atwell, S., Adams, J.M., Badger, J., Buchanan, M.D., Feil, I.K., Froning, K.J., Gao, X., Hendle, J., Keegan, K., Leon, B.C., Muller-Deickmann, H.J., Noland, B.W., Post, K., Rajashankar, K.R., Ramos, A., Russell, M., Burley, S.K., Buchanan, S.G.	1.57	Component definition mmCIF SDF PDB Launch viewer
1t46	STI	X-RAY	STRUCTURAL BASIS FOR THE AUTOINHIBITION AND STI-571 INHIBITION OF C-KIT TYROSINE KINASE Mol, C.D., Dougan, D.R., Schneider, T.R., Skene, R.J., Kraus, M.L., Scheibe, D.N., Snell, G.P., Zou, H., Sang, B.C., Wilson, K.P.	1.6	Component definition mmCIF SDF PDB Launch viewer
1opj	STI	X-RAY	Structural basis for the auto-inhibition of c-Abl tyrosine kinase Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.	1.75	Component definition mmCIF SDF PDB Launch viewer
1opj	STI	X-RAY	Structural basis for the auto-inhibition of c-Abl tyrosine kinase Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.	1.75	Component definition mmCIF SDF PDB Launch viewer
2oiq	STI	X-RAY	Crystal Structure of chicken c-Src kinase domain in complex with the cancer drug imatinib.	2.07	Component definition mmCIF

RCSB PDB Structure Explorer Page



RCSB PDB
PROTEIN DATA BANK

Structure Explorer Page for PDB 1T46



RCSB PDB
PROTEIN DATA BANK

RCSB Chemical Details

Chemical Component Summary — 1T46

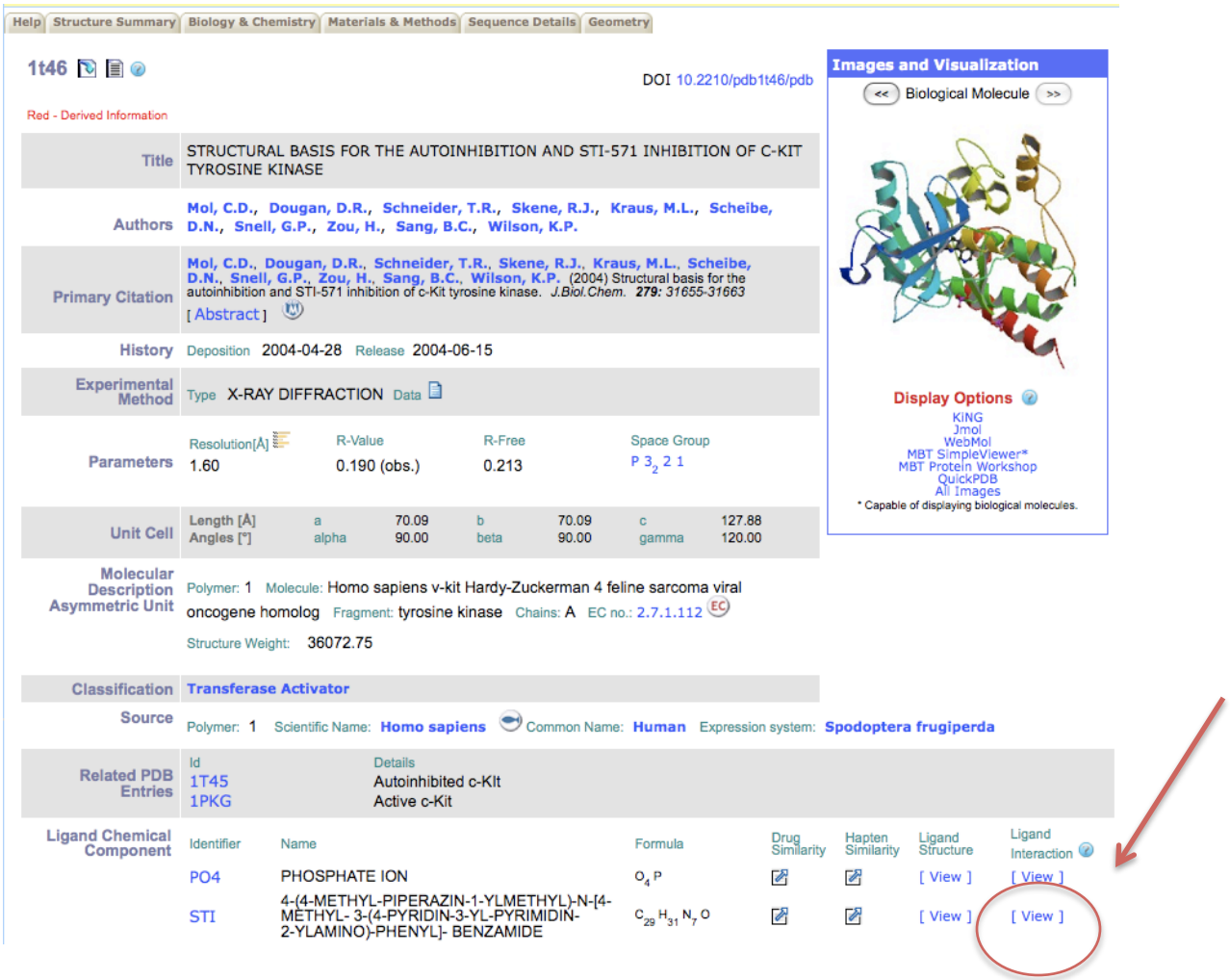
Name	4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE 4-[(4-methylpiperazin-1-yl)methyl]-N-[4-methyl-3-[(4-pyridin-3-ylpyrimidin-2-yl)amino]phenyl]benzamide
ID	STI
Synonyms	STI-571;IMATINIB
Formula	$C_{29}H_{31}N_7O$
Molecular Weight	493.603 g/mol
Formal Charge	0
Type	NON-POLYMER
Atom Count	68
Chiral Atom Count	0
Chiral Atoms	
Bond Count	72
Aromatic Bond Count	24

Descriptors

SMILES (OpenEye OEToolkits)	Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C
SMILES_CANONICAL (OpenEye OEToolkits)	Cc1ccc(cc1Nc2nccc(n2)c3cccnc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C
SMILES (CACTVS)	CN1CCN(CC1)Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4nccc(n4)c5cccnc5)c3
SMILES_CANONICAL (CACTVS)	CN1CCN(CC1)Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4nccc(n4)c5cccnc5)c3
InChI (InChI)	InChI=1/C29H31N7O /c1-21-5-10-25(18-27(21)34-29-31-13-11-26(33-29)24-4-3-12-30-19-24)32-28(37)23-8-6-22(7-9-23)20-36-16-14-35(2)15-17-36 /h3-13,18-19H,14-17,20H2,1-2H3,(H,32,37)(H,31,33,34)
SMILES (ACDLabs)	O=C(Nc3ccc(c(Nc2nc(c1cccnc1)ccn2)c3)C)c4ccc(cc4)CN5CCN(CC5)C

RCSB PDB PROTEIN DATA BANK

Structure Explorer Page for PDB 1T46



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

1t46

DOI 10.2210/pdb1t46/pdb

Red - Derived Information

Title	STRUCTURAL BASIS FOR THE AUTOINHIBITION AND STI-571 INHIBITION OF C-KIT TYROSINE KINASE
Authors	Mol, C.D., Dougan, D.R., Schneider, T.R., Skene, R.J., Kraus, M.L., Scheibe, D.N., Snell, G.P., Zou, H., Sang, B.C., Wilson, K.P.
Primary Citation	Mol, C.D., Dougan, D.R., Schneider, T.R., Skene, R.J., Kraus, M.L., Scheibe, D.N., Snell, G.P., Zou, H., Sang, B.C., Wilson, K.P. (2004) Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J.Biol.Chem.* **279:** 31655-31663 [Abstract]
History	Deposition 2004-04-28 Release 2004-06-15
Experimental Method	Type X-RAY DIFFRACTION Data

Parameters	Resolution[Å]	R-Value	R-Free	Space Group
	1.60	0.190 (obs.)	0.213	$P\,3_2\,2\,1$

Unit Cell						
Length [Å]	a	70.09	b	70.09	c	127.88
Angles [°]	alpha	90.00	beta	90.00	gamma	120.00

Molecular Description Asymmetric Unit

Polymer: 1 Molecule: Homo sapiens v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog Fragment: tyrosine kinase Chains: A EC no.: 2.7.1.112

Structure Weight: 36072.75

Classification Transferase Activator

Source Polymer: 1 Scientific Name: **Homo sapiens** Common Name: **Human** Expression system: **Spodoptera frugiperda**

Related PDB Entries	Id	Details
	1T45	Autoinhibited c-Kit
	1PKG	Active c-Kit

Ligand Chemical Component	Identifier	Name	Formula	Drug Similarity	Hapten Similarity	Ligand Structure	Ligand Interaction
	PO4	PHOSPHATE ION	$O_4\,P$			[View]	[View]
	STI	4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL- 3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]- BENZAMIDE	$C_{29}\,H_{31}\,N_7\,O$			[View]	[View]

Images and Visualization

<< Biological Molecule >>

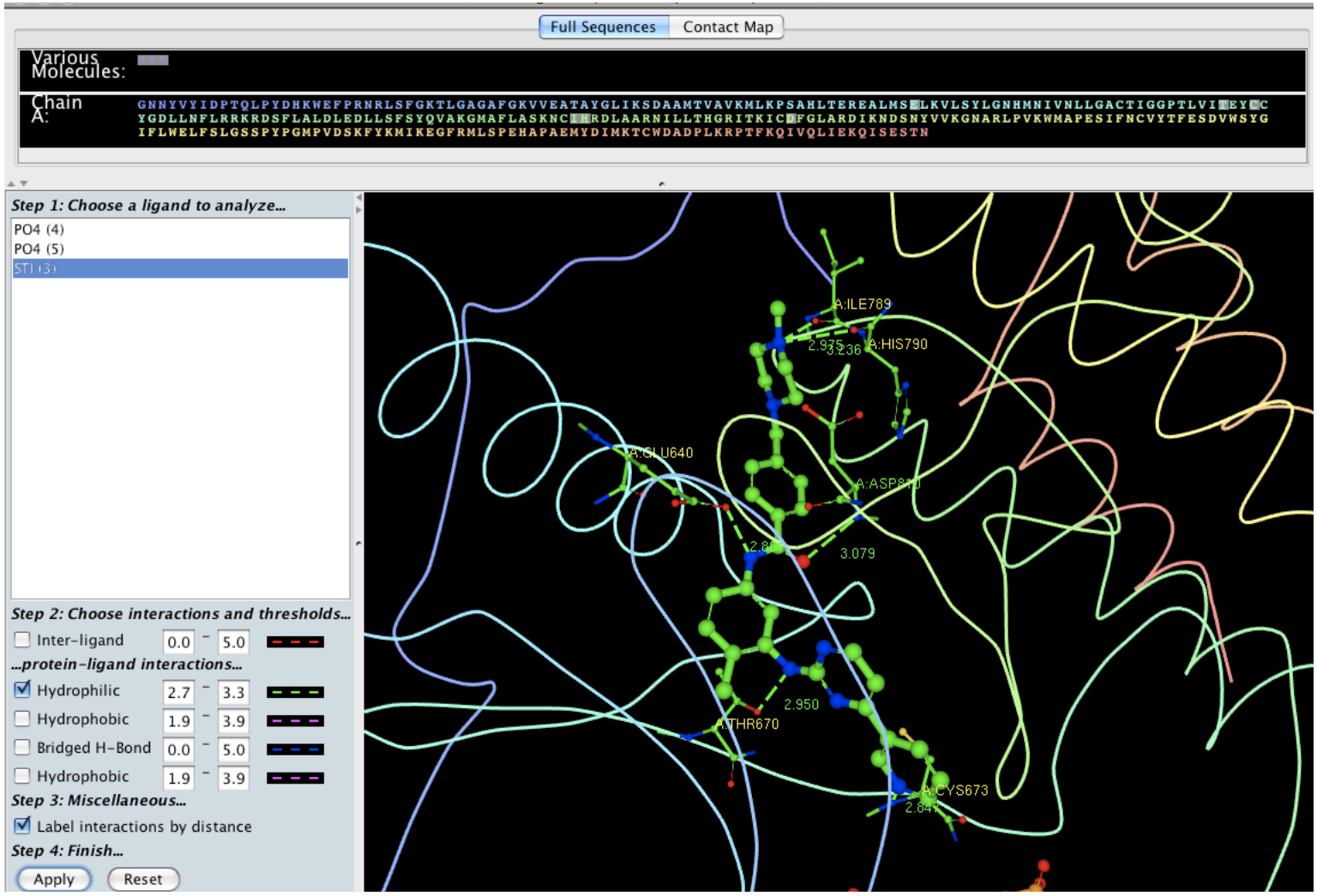
Display Options

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

* Capable of displaying biological molecules.

RCSB PDB
PROTEIN DATA BANK

Ligand Explorer View



RCSB PDB
PROTEIN DATA BANK

Dictionary Downloads

Ligand Expo

Ligand Expo Downloads

Chemical component dictionaries

mmCIF format dictionary (text | gzip)
mmCIF format variants dictionary (text | gzip)
SDF/MOL format file of dictionary ideal geometries (gzip)
PDBML format dictionary (tar.gz)
PDBML format variants dictionary (tar.gz)

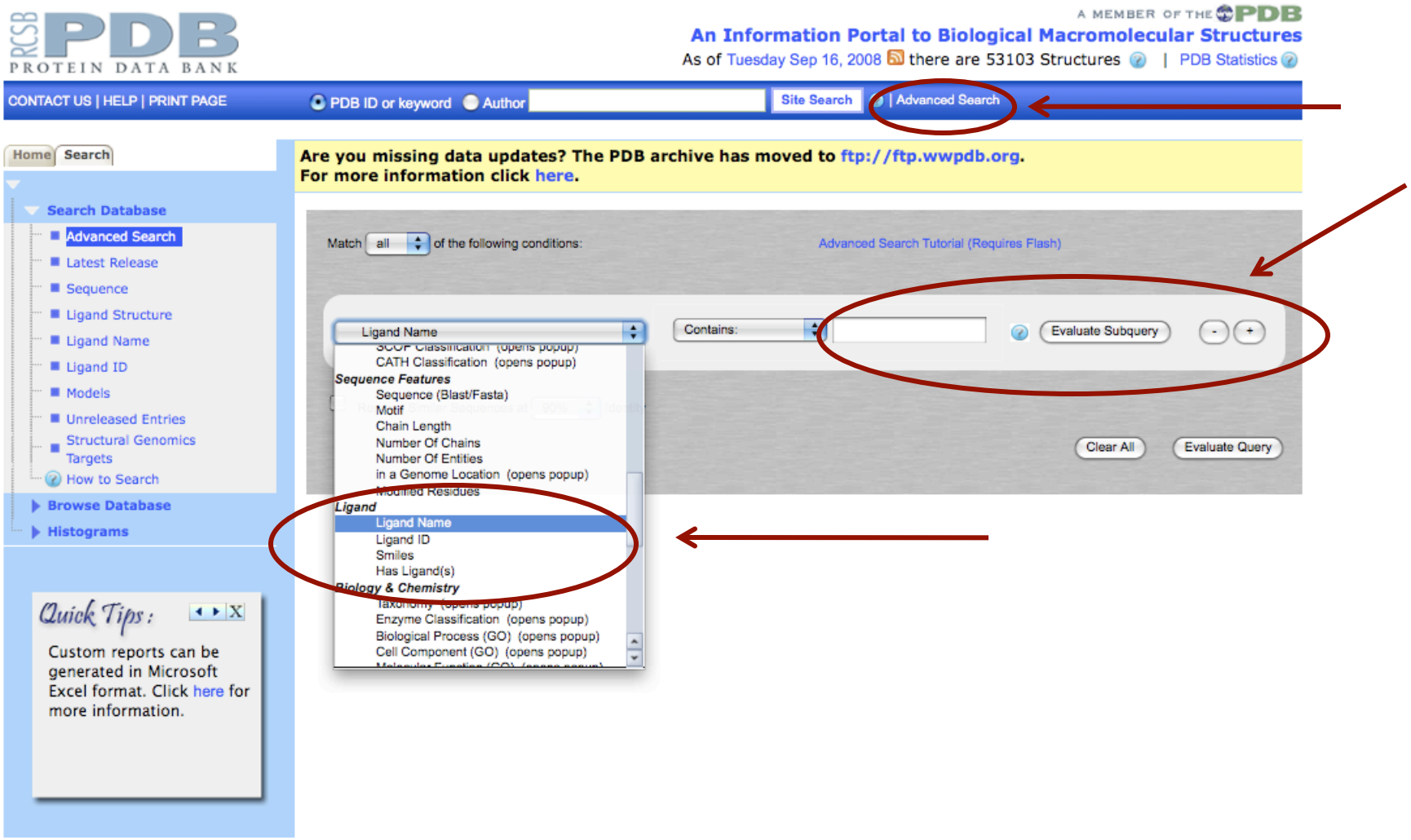
SMILES/InChI data files:

SMILES (OpenEye with stereo) (tab delimited text)
SMILES (OpenEye) (tab delimited text)
SMILES (CACTVS with stereo) (tab delimited text)
SMILES (CACTVS) (tab delimited text)
InChI (tab delimited text)
InChIKey (tab delimited text)

Chemical component coordinate data files[1]:

Experimental coordinates for non-polymer molecules and non-standard amino acids and nucleotides are bundled in single data files
(SDF/MOL (gz) (>45MB) | mmCIF (gz) (>98MB))

RCSB PDB Ligand Search



Access

- RCSB Protein Data Bank Site
  - http://www.rcsb.org/pdb/
- wwPDB Site
  - http://www.wwpdb.org/
- Dictionary Resource Site
  - http://mmcif.pdb.org/ & http://pdbml.pdb.org /
- Ligand Expo
  - http://ligand-expo.rcsb.org
- RCSB Software Download Site
  - http://sw-tools.pdb.org /
  - CVS server rcsb-cvs.rcsb.org - anonymous access

RCSB PDB
PROTEIN DATA BANK

WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB retreat September 2007

RCSB PDB
PROTEIN DATA BANK

Acknowledgements

Operated by two members of the RCSB:

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

UCSD
SDSC • SKAGGS SCHOOL of PHARMACY and PHARMACEUTICAL SCIENCES

The RCSB PDB is a member of the

WORLDWIDE PDB
PROTEIN DATA BANK

Supported by:

NSF

NIGMS

Office of Science
U.S. DEPARTMENT OF ENERGY

NATIONAL LIBRARY OF MEDICINE

NATIONAL CANCER INSTITUTE

National Center for Research Resources

NIBIB

NATIONAL INSTITUTE OF NEUROLOGICAL DISORDERS AND STROKE

National Institute of Diabetes & Digestive & Kidney Diseases