Small molecule data and tools at RCSB

John Westbrook
Rutgers University

September 24, 2008
Outline

• Chemical Components Dictionary
• Ligand Expo
• RCSB PDB website
How does new chemistry enter PDB?

Deposited Coordinates

Chemical Components

Perceive Covalent Structure

Compare with Dictionary

Process Deposited Entry

New?

Yes

No

Chemical Components Dictionary

Standardize Residue/Atom Nomenclature

Annotate Chemical Definition
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.
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.

PDE/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).
The Worldwide Protein Data Bank (wwPDB) consists of organizations that act as deposition, data processing and distribution centers for PDB data. The founding members are RCSB PDB (USA), PDBj (Europe) and PDBj (Japan). The BMRB (USA) group joined the wwPDB in 2006. The mission of the wwPDB is to maintain a single Protein Data Bank Archive of macromolecular structural data that is freely and publicly available to the global community.

This site provides information about services provided by the individual member organizations and about projects undertaken by the wwPDB.

Please note: ftp://ftp.rcsb.org is no longer updated. Please access the PDB archive using one of the FTP sites listed in the left menu.

15-September-2008

Announcement: Comprehensive Format Guide Version 3.2

During the past year, the wwPDB annotators have collaborated on a project to clarify the details and procedures related to data processing and annotation. The result is a PDB Contents Guide Version 3.2 that more fully describes the PDB file format. This document is available as a PDF and in HTML, and is accompanied by a document highlighting these clarifications.

In the coming months, all files released by the wwPDB will follow the format as described in this document. Details will be made available on this website and at www.wwpdb.org.

14-August-2008

IUCr: wwPDB Exhibition Stand and Presentations

The wwPDB partners will be exhibiting at the XXI Congress & General Assembly of the International Union of Crystallography (IUCr; August 23 – 31 in Osaka, Japan) at booth #14. Please stop by for website demonstrations and to meet with wwPDB members from around the globe.
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

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

HET | HYP | 18
--- | --- | ---
HETSYN | HYP | HYDROXYPROLINE
HETNAM | HYP | 4-HYDROXYPROLINE
FORMUL | HYP | C5 H9 N1 O3
PDB Data File Example

Name & Formula

<table>
<thead>
<tr>
<th>Name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>HETSYN</td>
<td>HYP HYDROXYPROLINE</td>
</tr>
<tr>
<td>HETNAM</td>
<td>HYP 4-HYDROXYPROLINE</td>
</tr>
</tbody>
</table>

ATOMIC COORDINATES AND CONNECTIVITY TABLE

<table>
<thead>
<tr>
<th>Atom</th>
<th>Type</th>
<th>Label</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>TempFactor</th>
<th>Occupancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>9302</td>
<td>N</td>
<td>HYP E</td>
<td>-6.687</td>
<td>-2.308</td>
<td>28.039</td>
<td>1.00</td>
<td>31.26</td>
</tr>
<tr>
<td>9303</td>
<td>CA</td>
<td>HYP E</td>
<td>-7.953</td>
<td>-3.047</td>
<td>27.929</td>
<td>1.00</td>
<td>31.93</td>
</tr>
<tr>
<td>9304</td>
<td>C</td>
<td>HYP E</td>
<td>-8.133</td>
<td>-3.707</td>
<td>26.555</td>
<td>1.00</td>
<td>32.45</td>
</tr>
<tr>
<td>9305</td>
<td>O</td>
<td>HYP E</td>
<td>-7.784</td>
<td>-3.116</td>
<td>25.531</td>
<td>1.00</td>
<td>36.79</td>
</tr>
<tr>
<td>9306</td>
<td>CB</td>
<td>HYP E</td>
<td>-9.041</td>
<td>-2.009</td>
<td>28.163</td>
<td>1.00</td>
<td>32.46</td>
</tr>
<tr>
<td>9307</td>
<td>CG</td>
<td>HYP E</td>
<td>-8.383</td>
<td>-0.818</td>
<td>28.833</td>
<td>1.00</td>
<td>34.37</td>
</tr>
<tr>
<td>9308</td>
<td>CD</td>
<td>HYP E</td>
<td>-6.885</td>
<td>-0.945</td>
<td>28.586</td>
<td>1.00</td>
<td>33.39</td>
</tr>
<tr>
<td>9309</td>
<td>OD1</td>
<td>HYP E</td>
<td>-8.866</td>
<td>0.380</td>
<td>28.269</td>
<td>1.00</td>
<td>48.65</td>
</tr>
</tbody>
</table>

Connectivity

REMARK 600 – Chemical details
REMARK 610/615 – Missing & zero occupancy atoms
REMARK 620 – Metal coordination
REMARK 630 – Inhibitor description
data_HYP
  #
  _chem_comp.id
  _chem_comp.name
  _chem_comp.type
  _chem_comp.pdbx_type
  _chem_comp.formula
  _chem_comp.mon_nstd_parent_comp_id
  _chem_comp.pdbx_synonyms
  _chem_comp.pdbx_formal_charge
  _chem_comp.pdbx_initial_date
  _chem_comp.pdbx_modified_date
  _chem_comp.pdbx_ambiguous_flag
  _chem_comp.pdbx_release_status
  _chem_comp.pdbx_replaced_by
  _chem_comp.pdbx_replaces
  _chem_comp.formula_weight
  _chem_comp.one_letter_code
  _chem_comp.three_letter_code
  _chem_comp.pdbx_model_coordinates_details
  _chem_comp.pdbx_model_coordinates_missing_flag
  _chem_comp.pdbx_ideal_coordinates_details
  _chem_comp.pdbx_ideal_coordinates_missing_flag
  _chem_comp.pdbx_model_coordinates_db_code
  _chem_comp.pdbx_processing_site
  #

HYP
4-HYDROXYPROLINE
"L-PEPTIDE LINKING"
ATOMP
"C5 H9 N O3"
PRO
HYDROXYPROLINE
0
1999-07-08
2008-04-29
N
REL
?
?
131.130
?
HYP
?
N
Corina
N
1OB6
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 -3.855 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
#```
<table>
<thead>
<tr>
<th>Connected atoms</th>
<th>Bond type</th>
<th>Stereochemistry &amp; aromaticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYP N   CA   SING N N 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP N   CD   SING N N 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP N   H    SING N N 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CA  C    SING N N 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CA  CB   SING N N 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CA  HA   SING N N 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP C   O    DOUB N N 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP C   OXT  SING N N 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CB  CG   SING N N 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CB  HB2  SING N N 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CB  HB3  SING N N 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CG  CD   SING N N 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CG  OD1  SING N N 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CG  HG   SING N N 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CD  HD22 SING N N 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP CD  HD23 SING N N 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP OD1 HD1 SING N N 17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYP OXT HXT SING N N 18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Descriptors and Identifiers

- **Descriptors**
  - **SMILES**
    - ACDLabs
      - 10.04 \(O=C(O)C1NCC(O)C1\)
    - CACTVS
      - 3.341 \([O(C@H)]1CN[C@H](C1)C(O)=O\)
    - OEToolkits
      - 1.5.0 \([C@H](CN[C@@H]1C(=O)O)O\)
  - **InChI**
    - ACDLabs
      - 1.02b \(=1/C5H9NO3/c7 … (H,8,9)/t3-,4+/m1/s1/f/h8H\)
    - CACTVS
      - 1.02b \(PMMYEEVYMwasQN-FJFIVYRMDE\)
  - **Systematic Names**
    - ACDLabs
      - (4R)-4-hydroxy-L-proline
  - **Systematic Names**
    - OEToolkits
      - (2S,4R)-4-hydroxypyrrolidine-2-carboxylic acid

### Identifiers

- **Identifiers**
  - **SMILES**
    - ACDLabs
      - 10.04 (4R)-4-hydroxy-L-proline
    - OEToolkits
      - 1.5.0 "(2S,4R)-4-hydroxypyrrolidine-2-carboxylic acid"
Descriptors

- **SMILES** - Simplified molecular input line entry specification

- **InChI** - IUPAC International Chemical Identifier

- **InChIKey** – InChI digital signature
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 Component Definition Example

```xml
<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: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
VPSRFSGSGGTQYSLKINSLQPEDFGSYCYQHQFWSTPRTFGGTKLEIK
;
;DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADG
VPSRFSGSGGTQYSLKINSLQPEDFGSYCYQHQFWSTPRTFGGTKLEIK
;

<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
      VPSRFSGSGGTQYSLKINSLQPEDFGSYCYQHQFWSTPRTFGGTKLEIK
    </PDBx:pdbx_seq_one_letter_code>
    <PDBx:pdbx_seq_one_letter_code_can>
      DIVLTQSPASLSASVGETVTITCRASGNIHNYLAWYQQKQGKSPQLLVYYTTTLADG
      VPSRFSGSGGTQYSLKINSLQPEDFGSYCYQHQFWSTPRTFGGTKLEIK
    </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

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 older PDB format and to PDBML/XML.

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

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

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 RCSS 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 Westbrook, Nobutoshi Ito, Haruki Nakamura, Kim Henrick and Helen M. Berman,

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 RCSS 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
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
**Search Options**

### Molecular Name, Formula, and Descriptor Search Options

**Search term:** `ncc(n2)c3cc(c3)NC(=O)c4cc(c4)CN5CCN(CC5)C`

**Search type:**
- SMILES
- Component identifier (3-letter code)
- Similar to component (3-letter code)
- Formula (all atom exact)
- Formula (heavy atom exact)
- Formula (exact subset)
- Formula (subset)
- Formula (close)
- Molecular name (exact)
- Molecular name (exact sub-string)
- Molecular name (similar)
- Chemically similar to SMILES
- InChI
- InChIKey

**Go**

### Sketch Input and/or Structure Search Options

**File name:**
- **Browse...**

**SMILES string:**

**-- OR --**

**Launch**

### Search for Instances of Chemical Components by 3-Letter ID Code

**Component ID code:**

**Display:**
- PDB entry codes

**Go**
Component Building & Substructure Search
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.
### Molecules like Gleevec (Imatinib)

**Comparison type:** Containing SMILES pattern  
**Result count:** 1

<table>
<thead>
<tr>
<th>ID</th>
<th>View Options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>STI</td>
<td>Chemical details</td>
<td>Name: 4-(4-methyl-piperazin-1-ylmethyl)-n-[4-methyl-3-(4-pyridin-3-ylpyrimidin-2-ylamino)-phenyl]-benzamide</td>
</tr>
</tbody>
</table>
|     | Coordinates files  | 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)c3cccn3)NC(=O)c4ccc/cc4CN5CCN(CC5)C |
|     |                    | Formula: C29 H31 N7 O |

---

Search report summary
### Chemical Details

#### Chemical Description

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

#### Chemical features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atom count</td>
<td>68</td>
</tr>
<tr>
<td>Chiral atom count</td>
<td>0</td>
</tr>
<tr>
<td>Bond count</td>
<td>72</td>
</tr>
<tr>
<td>Aromatic bond count</td>
<td>24</td>
</tr>
</tbody>
</table>

#### Chemical Identifiers

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic name (ACDLabs)</td>
<td>4-[[4-methyl[piperazin-1-y]methyl]-N-4-methyl-3-[[4-pyridin-3-y]pyrimidin-2-y] amino]phenyl]benzamide</td>
</tr>
<tr>
<td>Systematic name (OpenEye OEToolkits)</td>
<td>4-[[4-methyl[piperazin-1-y]methyl]-N-4-methyl-3-[[4-pyridin-3-y]pyrimidin-2-y] amino]phenyl]benzamide</td>
</tr>
</tbody>
</table>

#### Chemical Descriptors

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereo SMILES (CACTVS)</td>
<td>CN1CCN(CC1)Cc2cc(cc2)C(=O)Nc3ccc(C)c(Nc4ncccc(n4)c5cccn5)c3</td>
</tr>
<tr>
<td>SMILES (CACTVS)</td>
<td>CN1CCN(CC1)Cc2cc(cc2)C(=O)Nc3ccc(C)c(Nc4ncccc(n4)c5cccn5)c3</td>
</tr>
<tr>
<td>Stereo SMILES (OpenEye)</td>
<td>Ccc1cc(cc1Nc2ncc(n2)c3cccn3)NC(=O)c4ccccc4)CN5CCN(CC5)C</td>
</tr>
<tr>
<td>InChI descriptor</td>
<td>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,14-17,20H2,1-2H3,1b(h3,2,37)(16,H,31,33,34)/f/h32,34H</td>
</tr>
<tr>
<td>InChIKey descriptor</td>
<td>KTUFNOKKBVMGRW-RPGE6OUCX</td>
</tr>
</tbody>
</table>
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

#### Ligand Expo

**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**

<table>
<thead>
<tr>
<th>Atom I</th>
<th>Atom J</th>
<th>Distance (Experimental Model)</th>
<th>Distance (Ideal Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>C6</td>
<td>1.446</td>
<td>1.384</td>
</tr>
<tr>
<td>C1</td>
<td>C2</td>
<td>1.436</td>
<td>1.383</td>
</tr>
<tr>
<td>C1</td>
<td>H11</td>
<td>1.104</td>
<td>1.080</td>
</tr>
<tr>
<td>C6</td>
<td>C5</td>
<td>1.458</td>
<td>1.398</td>
</tr>
<tr>
<td>C6</td>
<td>H61</td>
<td>1.104</td>
<td>1.080</td>
</tr>
<tr>
<td>C5</td>
<td>C4</td>
<td>1.452</td>
<td>1.392</td>
</tr>
<tr>
<td>C5</td>
<td>C7</td>
<td>1.501</td>
<td>1.483</td>
</tr>
<tr>
<td>C4</td>
<td>N3</td>
<td>1.391</td>
<td>1.318</td>
</tr>
<tr>
<td>C4</td>
<td>H41</td>
<td>1.103</td>
<td>1.080</td>
</tr>
</tbody>
</table>
Related Data Resources for STI

The following table contains links to resources and identifiers which have additional information about this chemical component. The first column of the table contains resource names. Each resource name is linked to the top-level of the resource website. The identifier used by the resource is given in the second column. Clicking on a highlighted identifier will launch a query at the related resource and display the results in a new window.

<table>
<thead>
<tr>
<th>Related Resource</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubChem</td>
<td>5291</td>
</tr>
<tr>
<td>DrugBank</td>
<td>DB03261</td>
</tr>
<tr>
<td>ChEBI</td>
<td>45783</td>
</tr>
<tr>
<td>BindingDB</td>
<td>13530</td>
</tr>
<tr>
<td>NCI Chemical Lookup Service</td>
<td>SMILES query</td>
</tr>
<tr>
<td>eMolecules</td>
<td>SMILES query</td>
</tr>
<tr>
<td>CAS Registry Number</td>
<td>152459-95-5</td>
</tr>
</tbody>
</table>
## Experimental Examples

### Search Result Summary

**Target ID:** STI  
**Count in released entries:** 12

### Instances of STI in Released Entries

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

**Target ID:** STI  
**Count in released entries:** 12

**Instances of STI in Released Entries**

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>CC ID</th>
<th>Method</th>
<th>Title/Authors</th>
<th>Resolution</th>
<th>Downloads &amp; Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>1t46</td>
<td>STI</td>
<td>X-RAY</td>
<td>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.</td>
<td>1.6</td>
<td>Component definition mmCIF SDF PDB Launch viewer</td>
</tr>
<tr>
<td>1opj</td>
<td>STI</td>
<td>X-RAY</td>
<td>Structural basis for the auto-inhibition of c-Abl tyrosine kinase Nagar, B., Hantschel, O., Young, M.A., Schefzik, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.</td>
<td>1.75</td>
<td>Component definition mmCIF SDF PDB Launch viewer</td>
</tr>
<tr>
<td>1opj</td>
<td>STI</td>
<td>X-RAY</td>
<td>Structural basis for the auto-inhibition of c-Abl tyrosine kinase Nagar, B., Hantschel, O., Young, M.A., Schefzik, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., Kuriyan, J.</td>
<td>1.75</td>
<td>Component definition mmCIF SDF PDB Launch viewer</td>
</tr>
<tr>
<td>2o1q</td>
<td>STI</td>
<td>X-RAY</td>
<td>Crystal Structure of chicken c-Src kinase domain in complex with the 2.07 cancer drug imatinib.</td>
<td>2.07</td>
<td>Component definition mmCIF</td>
</tr>
</tbody>
</table>
### Chemical Component Summary

**Name**: 4-(4-METHYL-PIPERAZIN-1-YL METHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL PYRIMIDIN-2-YL AMINO)-PHENYL]-BENZAMIDE  
4-(4-methylpiperazin-1-yl)methyl]-N-[4-methyl-3-(4-pyrimidin-2-yl)amino]phenyl]benzamide  

**ID**: STI  
**Synonyms**: STI-571; IMATINIB  
**Formula**: C_{23}H_{31}N_{5}O  
**Molecular Weight**: 493.603 g/mol  
**FormalCharge**: 0  
**Type**: NON-POLYMER  
**Atom Count**: 88  
**Chiral Atom Count**: 0  
**Chiral Atoms**: 0  
**Bond Count**: 72  
**Aromatic Bond Count**: 24

#### Descriptors

**SMILES (OpenEye OEToolkits)**: \(\text{Cc1ccc(c)c1Nc2ccccc(n2)c3ccc(c)c3} \text{NC(=O)c4ccc(c)c4} \text{CN5CCN(CC5)C}\)  
**SMILES_CANONICAL (OpenEye OEToolkits)**: \(\text{Cc1ccc(c)c1Nc2ccccc(n2)c3ccc(c)c3} \text{NC(=O)c4ccc(c)c4} \text{CN5CCN(CC5)C}\)  
**SMILES (CACTVS)**: \(\text{CN1CCN(CC1)Ge2ccc(cc2)C(=O)Nc3ccc(C)c4(Nc4nocc(n4)c5ccc(c)c5)C} \text{03}\)  
**SMILES_CANONICAL (CACTVS)**: \(\text{CN1CCN(CC1)Ge2ccc(cc2)C(=O)Nc3ccc(C)c4(Nc4nocc(n4)c5ccc(c)c5)C} \text{03}\)  
**InChI**: \(\text{InChI=1S/C20H31N7O} \)  
**InChI (inChI)**: \(\text{c1-21-5-10-25(18-27(21)34-29-31-13-11-26(33-29)24-4-3-12-30-19-24)32-28(37)23-8-6-22(7-9-23)20-36-16-14-35(2)15-17-36} \)  
\(\text{h13,16-19H1,14-17,20H2,1-2H3,(H,32,37)(H,31,33,34)} \)  
**SMILES (ACDLabs)**: \(\text{O=C(Nc3ccc(c)c2nc1ccnc2)(c3)c4ccc(c)=O} \text{CN5CCN(CC5)C}\)
Structure Explorer Page for PDB 1T46

<table>
<thead>
<tr>
<th>Title</th>
<th>STRUCTURAL BASIS FOR THE AUTOINHIBITION AND STI-571 INHIBITION OF C-KIT TYROSINE KINASE</th>
</tr>
</thead>
</table>

**Experimental Method**
- Type: X-RAY DIFFRACTION

**Parameters**
- Resolution: 1.00 Å
- R-Value: 0.190 (obs.), 0.213
- R-Free: 0.213
- Space Group: P 3 2 1

**Unit Cell**
- a: 70.09 Å
- b: 90.02 Å
- c: 70.09 Å
- alpha: 90.00°
- beta: 90.00°
- gamma: 127.88°

**Molecular Description**
- Asymmetric Unit: 1 molecule
- Homo sapiens v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (KIT)
- Fragment: tyrosine kinase
- Chain: A

**Classification**
- Transferase Activator

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

**Related PDB Entries**
- 1T45
- 1PKG

**Ligand Chemical Component**
- Identifier: PO4
- Name: PHOSPHATE ION
- Formula: O4P
- Drug Similarity: [ View ]
- Hetero Similarity: [ View ]
- Legend Interaction: [ View ]
- STI
- 4-(4-METHYL-Piperazin-1-yl)-methyl)-N-(4-METHYL-3-(4-PYRIDIN-2-YLAMINO)-PHENYL)-BENZAMIDE
- Formula: C22H31N3O2
- Drug Similarity: [ View ]
- Hetero Similarity: [ View ]
- Legend Interaction: [ View ]
Ligand Explorer View

Step 1: Choose a ligand to analyze...
- PO4 (4)
- PD4 (5)

Step 2: Choose interactions and thresholds...
- Check Inter-ligand (0.0 - 5.0)
- Check Hydrophobic (1.9 - 3.9)
- Check Bridged H-Bond (0.0 - 5.0)

Step 3: Miscellaneous...
- Check Label interactions by distance

Step 4: Finish...
Dictionary Downloads

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 files¹:

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)
Access

- RCSB Protein Data Bank Site
  - http://www.rcsb.org/pdb/
- wwPDB Site
  - http://www.wwpdb.org/
- Dictionary Resource Site
- Ligand Expo
  - http://ligand-expo.rcsb.org
- RCSB Software Download Site
  - http://sw-tools.pdb.org/
  - CVS server rcsb-cvs.rcsb.org - anonymous access
wwPDB retreat September 2007
Acknowledgements

Operated by two members of the RCSB:

The RCSB PDB is a member of the

Supported by: