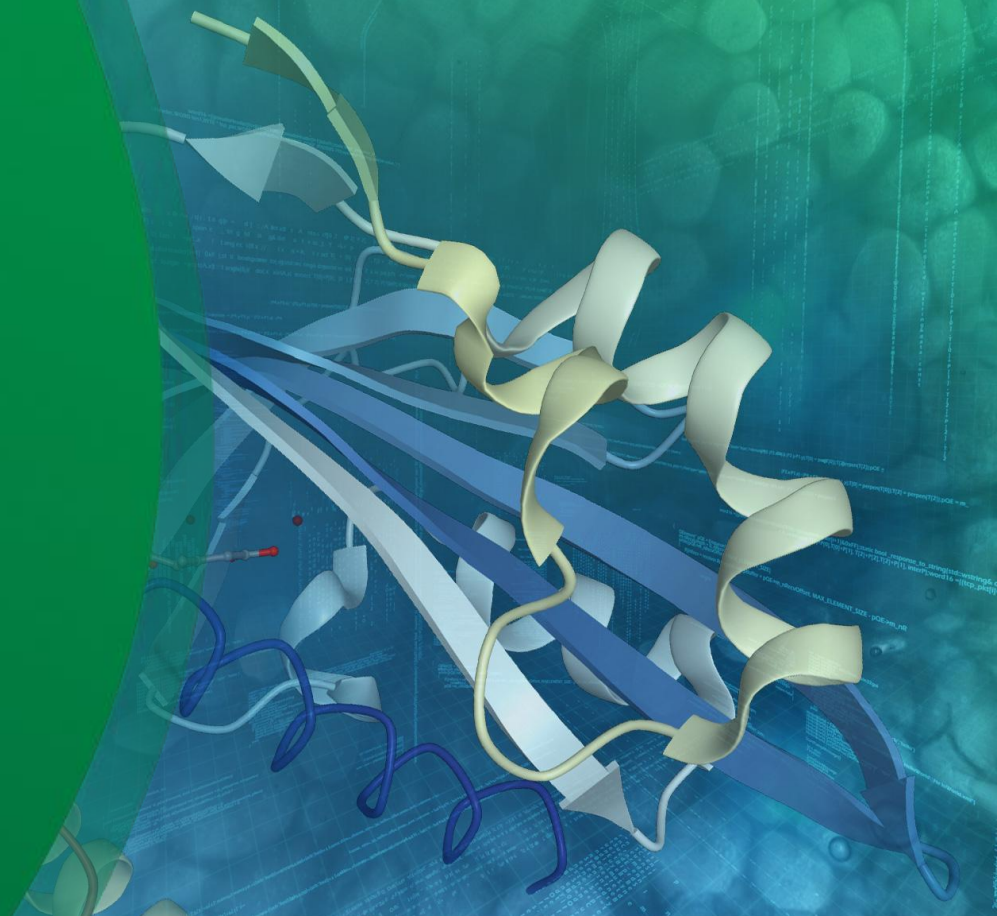


Introduction to the Protein Data Bank



Dr. Deborah Harrus

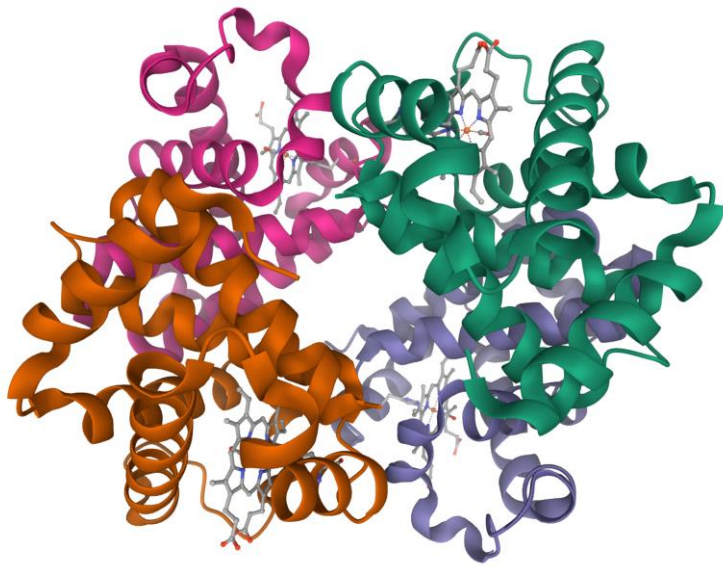
PDBe Archive Project Leader

CCDC Virtual Workshop

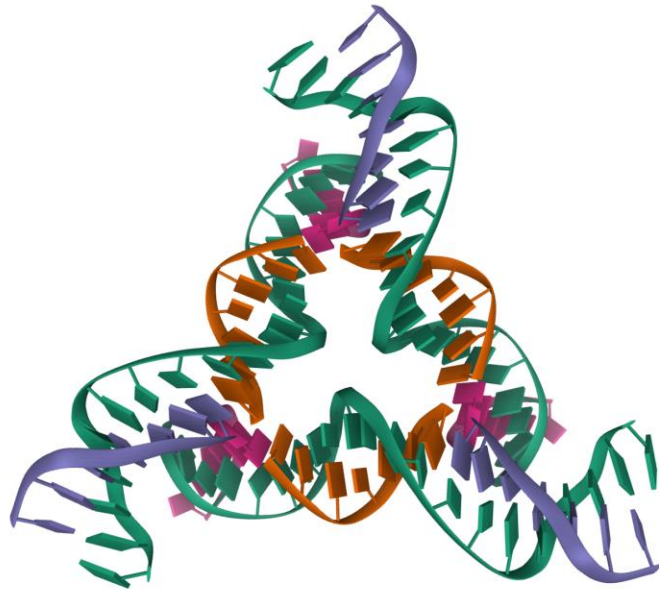
5th of November 2024



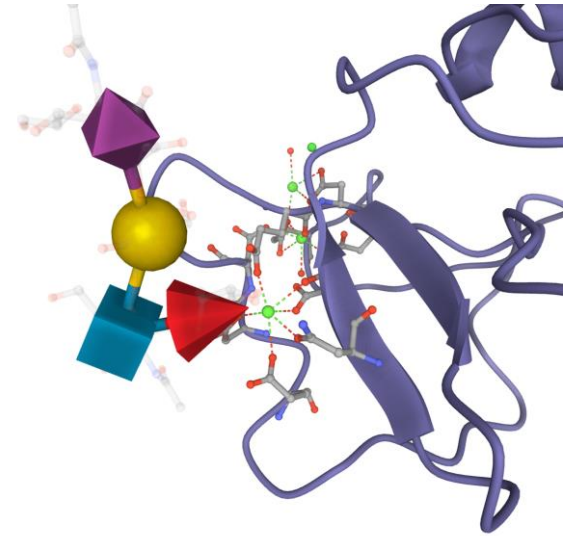
The Protein Data Bank (PDB) is an archive of **experimentally determined** 3-dimensional structures of biological macromolecules



Proteins



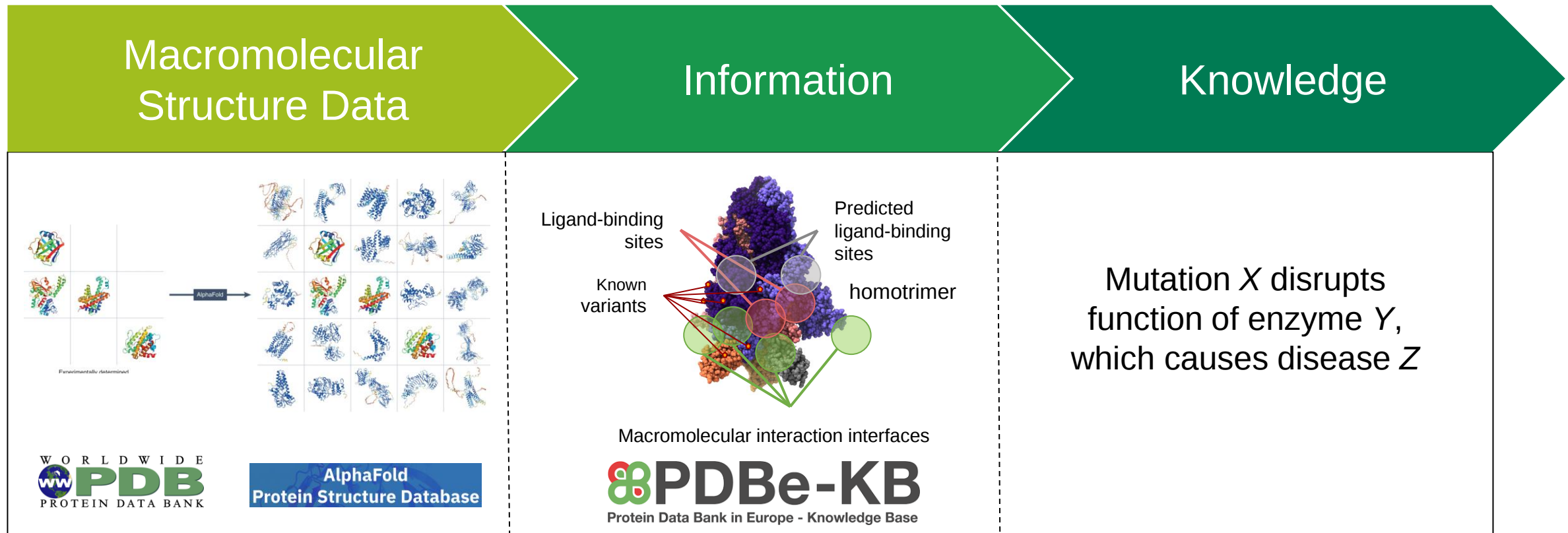
DNA/RNA



Bound ligands

Our mission...

Establish a **high-quality, up-to-date** and **integrated infrastructure** and innovative **data analysis tools** for **3D macromolecular structures** to support basic and applied **research** and **education** across the sciences



How chemical information is stored in PDB

CCD (chemical component dictionary) definitions (reference dictionary)

Ideal coordinates (*Experimental coordinates*)

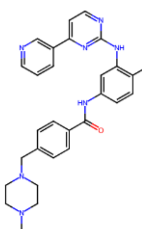
Name, synonyms, formula, weight

Atom-level info: unique ids, element, charge, R vs S

Bond-level info: aromaticity, stereochemistry, bond order

Descriptors: SMILES, InChI, InChIKey

```
#
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_component_atom_id
  _chem_comp_atom.pdbx_component_comp_id
  _chem_comp_atom.pdbx_ordinal
STI C1 C1 C 0 1 Y N N N N 16.356 100.406 50.614 2.740 -0.022 8.889 C1 STI 1
STI C6 C6 C 0 1 Y N N N N 15.218 100.012 51.414 1.914 0.207 7.802 C6 STI 2
```



CCD ID: STI
Mol. Name: Imatinib

https://ftp.ebi.ac.uk/pub/databases/msd/pdbechem_v2/ccd/S/STI/STI.cif

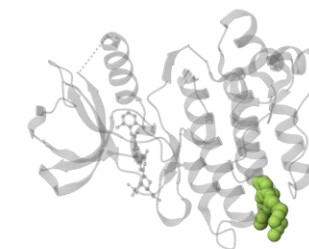
PDB entries (experimental 3D structures)

Experimental coordinates

Number of instances observed

Subject of investigation

```
#
loop_
  _atom_site.group_PDB
  _atom_site.id
  _atom_site.type_symbol
  _atom_site.label_atom_id
  _atom_site.label_alt_id
  _atom_site.label_comp_id
  _atom_site.label_asym_id
  _atom_site.label_entity_id
  _atom_site.label_seq_id
  _atom_site.pdbx_PDB_ins_code
  _atom_site.Cartn_x
  _atom_site.Cartn_y
  _atom_site.Cartn_z
  _atom_site.occupancy
  _atom_site.B_iso_or_equiv
  _atom_site.pdbx_formal_charge
  _atom_site.auth_seq_id
  _atom_site.auth_comp_id
  _atom_site.auth_asym_id
  _atom_site.auth_atom_id
  _atom_site.pdbx_PDB_model_num
ATOM 1 N N . ASP A 15 ? -22.963 -26.563 -47.248 1.000 77.262 ? 252 ASP A N 1
ATOM 2 C CA . ASP A 15 ? -22.572 -26.099 -45.878 1.000 75.723 ? 252 ASP A CA 1
.
.
.
HETATM 5887 C C1 . STI D 2 . ? -17.422 -2.386 -7.775 1.000 44.873 ? 601 STI A C1 1 ?
HETATM 5888 C C6 . STI D 2 . ? -16.730 -3.540 -7.470 1.000 45.283 ? 601 STI A C6 1 ?
```

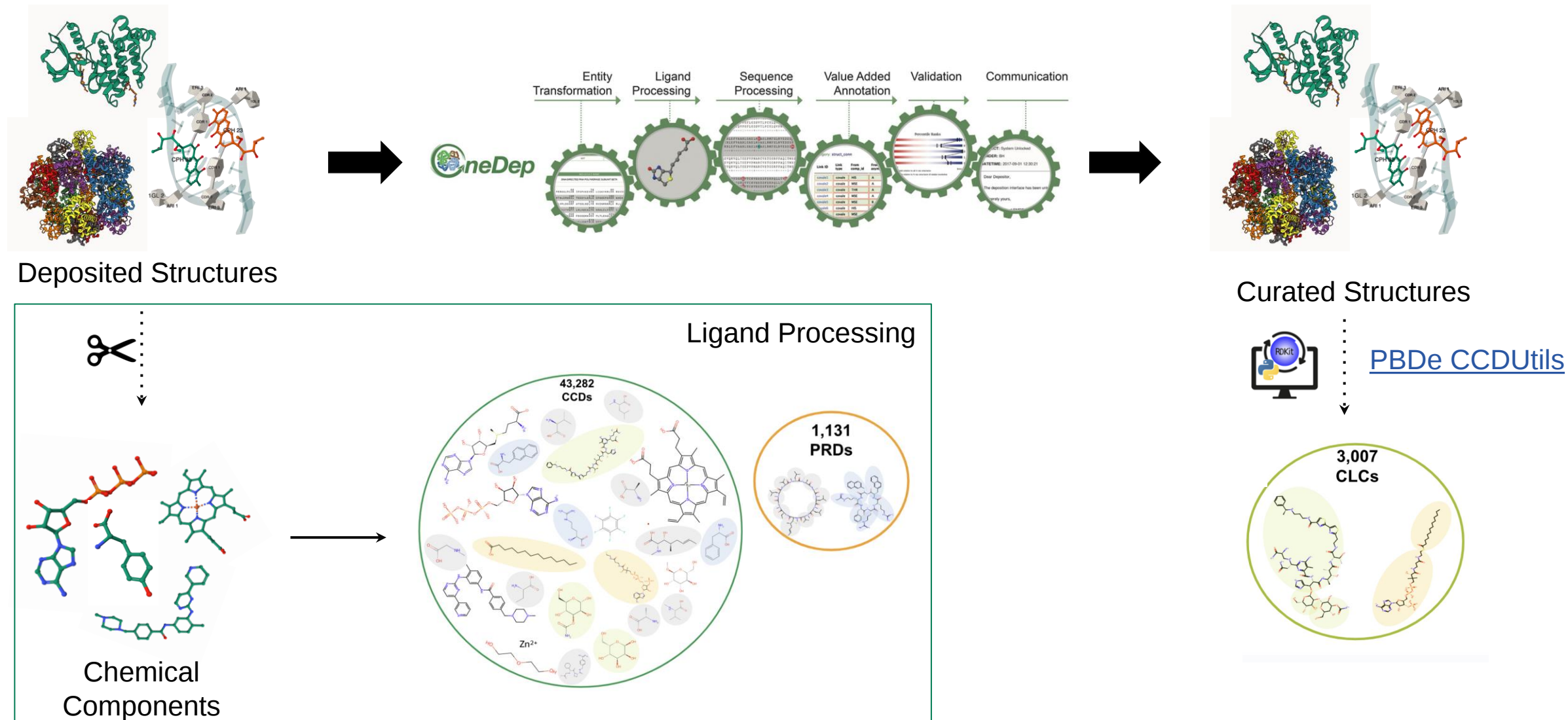


PDB ID: 7n9g
Protein Name: Tyrosine-protein kinase ABL1

https://www.ebi.ac.uk/pdbe/entry-files/7n9g_updated.cif

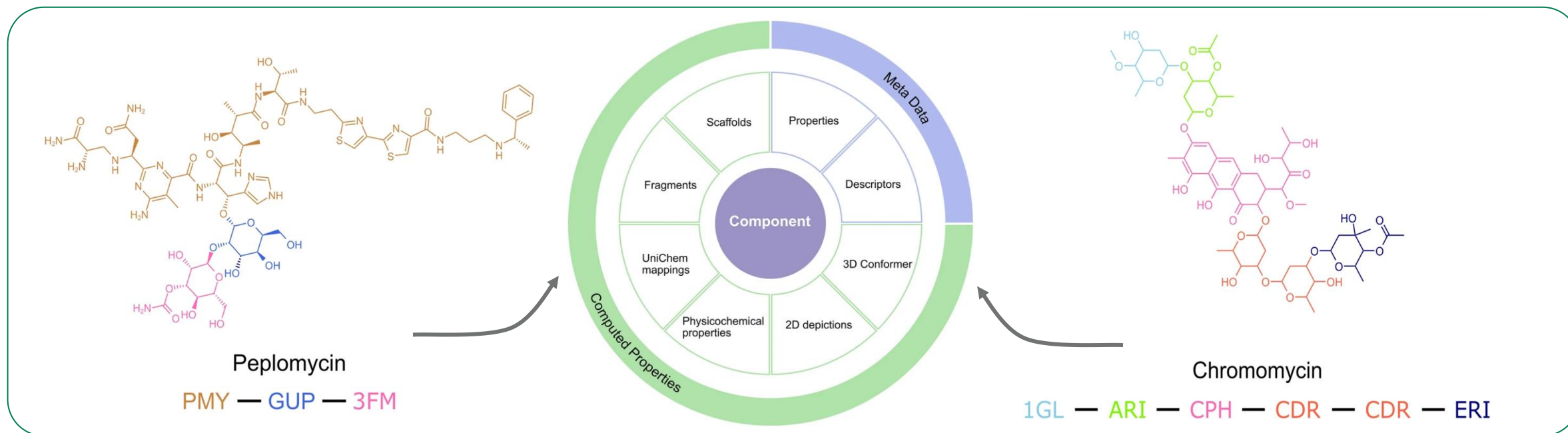
https://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Index/index.html

Small molecules annotation



CCDUtils: Structural analysis of small molecules in the PDB

Toolbox for chemoinformatic visualization, property calculations, and CLC identification.

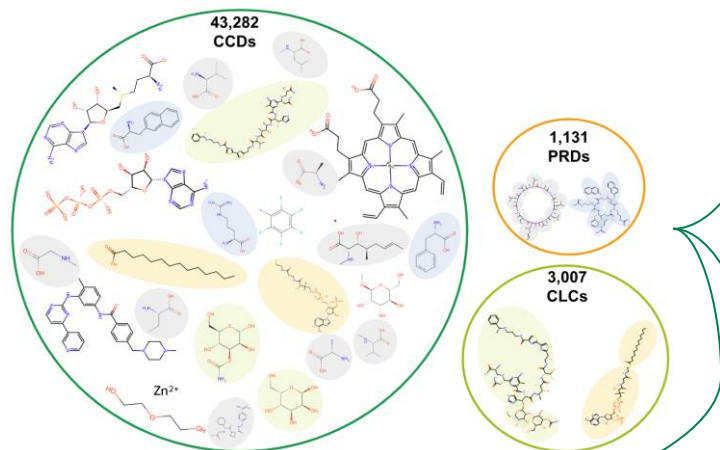


Kunnakkattu et al. (2023)

doi.org/10.1186/s13321-023-00786-w

<https://github.com/PDBEurope/ccdutils>

PDBe Ligand Pages



PDBe CCDUtils

PDBe-KB Ligands

Search in PDBe and PDBe-KB

Examples: STI, QLC, XIS, NAG, HEM

Description

Physicochemical properties

Bond structures

Interaction statistics

Related ligands

Same scaffold

Similar ligands

Ligand-specific databases

STI [Drug-like](#)

4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE

[View 3D](#) [Download as](#)

Description

Synonyms

Imatinib, MInab, Enliven, Celonib, STI-STI, Imatinib, IMATIN...

Formula

C29H31N7O

IUPAC InChI

InChI=1S/C29H31N7O/c1-21-5-10-25(18-27)134-29-31-13-11-26(33-29)24-4-3-12-10-19-24(32-28)37(23-8-6-22)7-9-23(20-36)-16--

IUPAC InChIkey

KTUFNOKBVMGRW-UHFFFAOYSA-N

SMILES

Cc1ccc(cc1N2ccc(cc2)N3C(=O)C(=O)C4=CC=CC=C4)C5=CC=CC=C5C

Source [OpenEye](#)

[View Atoms](#) [View Bonds](#)

Overall view, and highlighted scaffolds and fragments

Displayed: 1 / 10

Atom Labelled STI

Related Ligands [Download related ligands](#)

Same scaffold - 1

Search by name or PDB ID

[MP2](#) [97.3% similar](#) [Bound in 1 PDB Entry](#)

Similar ligands - 4

Search by name or PDB ID

Filter by similarity between and

[PRC](#) [73.7% similar](#) [Bound in 1 PDB Entry](#)

[406](#) [61.2% similar](#) [Bound in 1 PDB Entry](#)

[748](#) [62% similar](#) [Bound in 1 PDB Entry](#)

[CQO](#) [60.9% similar](#) [Bound in 1 PDB Entry](#)

PDBe Arpeggio

PBDe RelLig

(Relevant
Ligands in
proteins)

Bound structures [Download structures \(csv\)](#) [Download coordinates \(mmCIF\)](#)

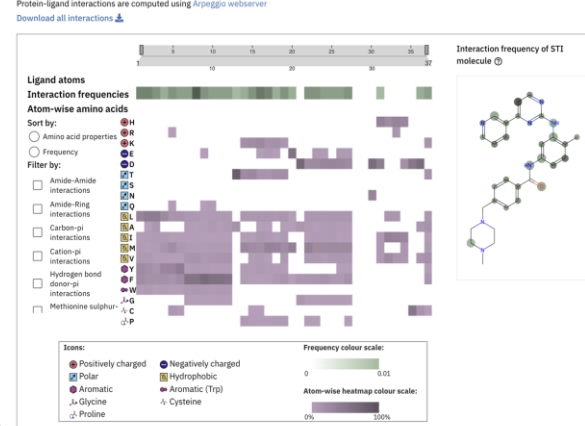
Found in 15 Proteins and 28 PDB Structures. Group data by: ☐ Proteins ☒ Structures

Protein name	PDBe-KB link	PDB ID and Chain	Organism	EC number	Ligand function
Tyrosine-protein kinase ABL1	P00519	6npe_B	Homo sapiens (human)	2.7.10.2	Drug-like
Tyrosine-protein kinase ABL1	P00519	6npe_A	Homo sapiens (human)	2.7.10.2	Drug-like
Tyrosine-protein kinase ABL1	P00519	6npv_A	Homo sapiens (human)	2.7.10.2	Drug-like
Tyrosine-protein kinase ABL1	P00519	3pyy_B	Homo sapiens (human)	2.7.10.2	Drug-like
Tyrosine-protein kinase ABL1	P00519	6npv_B	Homo sapiens (human)	2.7.10.2	Drug-like
Tyrosine-protein kinase ABL1	P00519	6npv_A	Homo sapiens (human)	2.7.10.2	Drug-like
Tyrosine-protein kinase ABL1	P00519	2hyy_A	Homo sapiens (human)	2.7.10.2	Drug-like
Tyrosine-protein kinase ABL1	P00519	7n9g_A	Homo sapiens (human)	2.7.10.2	Drug-like

Page Size: 1 to 8 of 8 < > Page 1 of 1

Interaction statistics

Interaction statistics shows the frequency of interaction of each atom of small molecule with any amino acid in all the structures to which it is bound. This information was computed using 32 ligand instances in 16 PDB structures and 28 PDB chains. Protein-ligand interactions are computed using [Arpeggio webserver](#)



Bound structures [Download structures \(csv\)](#) [Download coordinates \(mmCIF\)](#)

Found in 15 Proteins and 28 PDB Structures. Group data by: ☒ Proteins ☐ Structures

Protein name	PDBe-KB link	Total structures	Species	EC number	Ligand function
Tyrosine-protein kinase ABL1	P00520	7	Mus musculus	2.7.10.2	Unannotated
Tyrosine-protein kinase ABL1	P00519	6	Homo sapiens	2.7.10.2	Drug-like
Proto-oncogene tyrosine-protein kinase Src	P00523	2	Gallus gallus	2.7.10.2	Unannotated
Mitogen-activated protein kinase 14	Q16539	1	Homo sapiens	2.7.11.24	Unannotated
Tyrosine-protein kinase ABL2	P42684	1	Homo sapiens	2.7.10.2	Unannotated
Deoxycytidine kinase	P27707	1	Homo sapiens	2.7.1.74, 2.7.1.113, 2.7.1.76	Unannotated
Epithelial discoidin domain-	Q08345	1	Homo sapiens	2.7.10.1	Unannotated

Page Size: 1 to 10 of 16 < > Page 1 of 2

PDBE Ligand Pages

Keyed on CCD/PRD/CLC IDs

Description

Physicochemical properties

Structures

Interaction statistics

Related ligands

Same scaffold

Similar ligands

Ligand-specific databases

STI

4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE

Drug-like

View 3D Download as

Description

Synonyms Imatib, Mitinab, Enliven, Celonib, STI-571, Imatinib, IMATIN...
Show more

Formula C29H31N7O

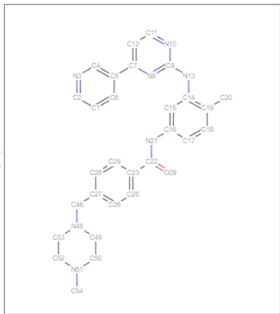
IUPAC InChI InChI=1S/C29H31N7O/c1-21-5-10-25(18-27(21)34-29-31-13-11-26(33-29)24-4-3-12-30-19-24)32-28(37)23-8-6-22(7-9-23)20-36-16-...
Show more

IUPAC InChIKey KTUFNOKKBVMGRW-UHFFFAOYSA-N

SMILES Cc1ccc(cc1Nc2cccc(n2)c3ccccc3)NC(=O)c4ccc(cc4)CN5CCN(CC5)C
Source [OpenEye](#)

View Atoms View Bonds

Overall view, and highlighted scaffolds and fragments



Displayed: 1 / 10
Atom Labelled STI

Physicochemical properties [Download physicochemical properties](#)

Molecular properties

Molecular weight	493.259 Da
Labute accessible surface area	260.626 Å ²
Heavy atoms	37
Heteroatoms	8
Carbon SP3 value	0.241
Wildman-Crippen molar refractivity	142.302
Wildman-Crippen Log P	2.904

Conformational properties

Rotatable bonds	0
-----------------	---

Ring properties

Aromatic rings	4
Rings	5
Aliphatic rings	1
Heterocycles	3
Saturated rings	1
Aromatic heterocycles	2
Saturated heterocycles	1
Aliphatic heterocycles	1
Spiro atoms	0

Surface properties

Topological surface area	86.28
--------------------------	-------

Functional group properties

Hydrogen bond donors	2
Hydrogen bond acceptors	7
Amide bonds	1

Stereochemical properties

Stereocenters	0
---------------	---

Releasing soon (this November)

These webpages will replace [pdbechem service](#)

Following data will be accessible

Basic physicochemical properties

Cross-references to other resources

Functional annotations

Descriptors, synonyms

All the structure to which it binds

Similar ligands

Ligands with same scaffold

Stereoisomers

All the interacting proteins, enzymes

2D and 3D views of ligand/protein

<https://wwwdev.ebi.ac.uk/pdbe/connect/ligands/STI>

Acknowledgements



The Protein Databank in Europe team



Funded by
the European Union

Google DeepMind

