BindingDB SDfile Format

July 23, 2025

This document describes the format of an SDfile downloaded from BindingDB. Briefly, an SDfile contains one or many compound records, along with data fields for each compound (https://depth-first.com/articles/2020/07/13/the-sdfile-format). BindingDB uses the data fields to provide information on affinity measurements for each compound. Note that there may be multiple affinity measurements for a given compound and biomolecule: several different groups may have published measurements, or a single group may have repeated the measurement under different experimental conditions. Note, too, that a BindingDB SDfile does not include all the detailed information about each affinity measurement that BindingDB stores. More information on each measurement is available at the website, www.bindingdb.org.

The top of the file is a v2000 Ctab Molfile; see, e.g. https://depth-first.com/articles/2020/07/13/the-sdfile-format. This is followed by a series of Data Items, which are detailed below.

> <From>
www.bindingdb.org

- > <**BindingDB Reactant_set_id**> Stable internal BindingDB identifier for this binding reaction.
  Text.
- > <**Ligand SMILES>**

  Text.
- > <**Ligand InChI**>

  Text
- > <**Ligand InChI key**>

  Text
- > <**BindingDB MonomerID**> Internal BindingDB identifier for this ligand.

  Integer
- > <**BindingDB Ligand Name**> Human-readable name assigned to this Ligand by whoever curated the data

  Text
- > <**Target Name**> This is the standard UniProt protein name with custom BindingDB annotations, if needed, that indicate deletions (e.g., a truncated sequence) or mutations (e.g., Met residues changed for the sake of stability) in the protein used in the binding studies. For

example, [262-930, L692P] means the protein includes residue 262-930 of the full sequence and Leu 692 is replaced by Pro.

Text.

- > <**Target Source Organism According to Curator or DataSource**> Organism associated with the protein Target.

  Text

- > <**Ki (nM)>**

  Real. >0 (values are occasionally reported as “>X”, where X is Real, following the source format.)

- > <**IC50 (nM)>**

  Real. >0 (values are occasionally reported as “>X”, where X is Real, following the source format.)

- > <**Kd (nM)>**

  Real. >0 (values are occasionally reported as “>X”, where X is Real, following the source format.)

- > <**EC50 (nM)>**

  Real. >0 (values are occasionally reported as “>X”, where X is Real, following the source format.)

- > <**kon (M-1-s-1)>**

  Real. >0

- > <**koff (s-1)>**

  Real. >0

- > <**pH>**

  Real. >0

- > <**Temp (C)>**

  Real.

- > <**Curation/DataSource>** Typically one of the following: BindingDB, ChEMBL, PubChem, PDSP Ki, CSAR, PubChem AID, or Deposited by Author.

  Text

- > <**Article DOI>** Digital object identifier for the source document, usually a journal article, if available.

  Text

- > <**BindingDB Entry DOI>** Digital object identifier for BindingDB entry containing these data.

  Text

- > <**PMID**> PubMed ID of article, if available.

  Integer

- > <**PubChem AID>** PubMed Assay ID, for data drawn from PubChem.

  Text

- > <**Patent Number**> If applicable.

  Text

- > <**Authors>**

  Text

- > <**Date of publication>** Published date of article or patent.

  Text

- > <**Date of publication>** Curated date of article or patent.

  Text

- > <**Institution>** Where the measurement was made. Usually, a university or company.

  Text

- > <**Link to Ligand in BindingDB>** Preformatted URL to a query for data for this Ligand within BindingDB.

  Text.

- > <**Link to Target in BindingDB>** Preformatted URL to a query for data for this Target within BindingDB.

  Text.

- > <**Link to Ligand-Target Pair in BindingDB>** Preformatted URL to a query for data for the Ligand-Target pair in BindingDB.

  Text.

- > <**Ligand HET ID in PDB>** If available, the hetero group ID for this ligand in the PDB.

  Text

- > <**PDB ID(s) for Ligand-Target Complex>** If available. Criterion for protein match is 85% sequence identity.

  Text

- > <**PubChem CID>** Compound ID of this Ligand in PubChem.

  Integer

- > <**PubChem SID>** Substance ID of this Ligand in PubChem.

  Integer

- > <**ChEBI ID of Ligand>**

  Text

- > <**ChEMBL ID of Ligand>**

  Text

- > <**DrugBank ID of Ligand>**

  Integer

- > <**IUPHAR_GRAC ID of Ligand>**

  Text

- > <**KEGG ID of Ligand>**

  Text

- > <**ZINC ID of Ligand>**

  Text

- > <**Number of Protein Chains in Target** (N>1 implies a multichain complex). The following information will be provided, if available, for each chain in the protein.

  Integer

  - > <**BindingDB Target Chain Sequence i>** (i=1, …, N)

    Text

  - > <**PDB ID(s) of Target Chain i>** (i=1, …, N) Criterion for a match is 85% sequence identity.

    Text

  - > <**UniProt (SwissProt) Recommended Name of Target Chain i>** (i=1, …, N) Criteria for a UniProt match is 100% sequence identity and a matching Source Organism. However, it is not required that the full lengths of the chains match.

    Text

  - > <**UniProt (SwissProt) Entry Name of Target Chain i>** (i=1, …, N).

    Text

- > <**UniProt (SwissProt) Primary ID of Target Chain i>** (i=1, …, N).

  Text

- > <**UniProt (SwissProt) Secondary ID(s) of Target Chain i>** (i=1, …, N) Secondary IDs are obsolete or deprecated. They are provided for the sake of backward compatibility.

  Text

- > <**UniProt (SwissProt) Alternative ID(s) of Target Chain i>** (i=1, …, N) Alternative IDs are Primary or Secondary IDs for chains with 100% matches that are annotated as deriving from a different Source Organism.

  Text

- > <**UniProt (TrEMBL) Submitted Name of Target Chain i>** (i=1, …, N).

  Text

- > <**UniProt (TrEMBL) Entry Name of Target Chain i>** (i=1, …, N).

  Text

- **> <UniProt (TrEMBL) Primary ID of Target Chain i>** (i=1, …, N).

  Text

- **> <UniProt (TrEMBL) Secondary ID(s) of Target Chain i>** (i=1, …, N).

  Text

- **> <UniProt (TrEMBL) Alternative ID(s) of Target Chain i>** (i=1, …, N).

  Text