BindingDB SDfile Format

July 11, 2009

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. 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 web-site.

A model extract from a BindingDB SDfile follows; it lists one compound, 2-ketobenzothiazole 34, and two associated affinity measurements. The data fields are defined following the extract. Text formatting (color, bold) is used only for the sake of clarity. The PDB file links are artificial and used only for this formatting example.

```
34
  Marvin  03060720322D

 42 46  0  0  0  0            999 V2000
    2.5950    2.5375    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.5950    3.3625    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    3.3095    2.1250    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.0721    1.0294    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.5704    1.6869    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.3923    0.2691    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.2109    0.1662    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.7092    0.8237    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    5.3890    1.5841    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.0991    2.3640    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    3.2928    1.3002    0.0000 S   0  0  0  0  0  0  0  0  0  0  0  0
    0.3411    2.5375    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.0714    1.8230    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.0714    3.2520    0.0000 C   0  0  2  0  0  0  0  0  0  0  0  0
    0.3411    3.9664    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.8964    4.0770    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.1819    4.4895    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.8964    3.2520    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   -1.6109    3.6645    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.6109    4.4895    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -2.3253    3.2520    0.0000 C   0  0  2  0  0  0  0  0  0  0  0  0
   -3.0398    3.6645    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
   -3.7543    3.2520    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.3253    2.4270    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.0398    2.0145    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.7543    2.4270    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -4.4687    2.0145    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -4.4687    1.1895    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.7543    0.7770    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -3.0398    1.1895    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.8806    2.1250    0.0000 C   0  0  1  0  0  0  0  0  0  0  0  0
    1.1661    2.5375    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    1.8806    1.3000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.1661    0.8875    0.0000 C   0  0  2  0  0  0  0  0  0  0  0  0
    0.4516    1.3000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
```

```
   -0.2628    0.8875    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.2628    0.0625    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.1661    0.0625    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.4516   -0.3500    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    0.4516   -1.1750    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.2628   -1.5875    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    1.1661   -1.5875    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
 31  1  1  0  0  0  0
  1  3  1  0  0  0  0
  1  2  2  0  0  0  0
 10  3  2  0  0  0  0
  3 11  1  0  0  0  0
 10  5  1  0  0  0  0
 11  4  1  0  0  0  0
  4  6  1  0  0  0  0
  5  9  1  0  0  0  0
  6  7  2  0  0  0  0
  7  8  1  0  0  0  0
  8  9  2  0  0  0  0
  4  5  2  0  0  0  0
 31 32  1  0  0  0  0
 32 12  1  0  0  0  0
 12 14  1  0  0  0  0
 12 13  2  0  0  0  0
 18 14  1  0  0  0  0
 14 15  1  1  0  0  0
 18 16  1  0  0  0  0
 16 17  1  0  0  0  0
 17 15  1  0  0  0  0
 18 19  1  0  0  0  0
 19 21  1  0  0  0  0
 19 20  2  0  0  0  0
 21 22  1  0  0  0  0
 22 23  1  0  0  0  0
 21 24  1  1  0  0  0
 24 25  1  0  0  0  0
 30 25  1  0  0  0  0
 26 25  2  0  0  0  0
 26 27  1  0  0  0  0
 27 28  2  0  0  0  0
 28 29  1  0  0  0  0
 29 30  2  0  0  0  0
 31 33  1  6  0  0  0
 33 34  1  0  0  0  0
 34 38  1  1  0  0  0
 35 34  1  0  0  0  0
 35 36  1  0  0  0  0
 36 37  1  0  0  0  0
 37 39  1  0  0  0  0
 39 38  1  0  0  0  0
 39 40  1  0  0  0  0
 40 42  2  0  0  0  0
 40 41  1  0  0  0  0
M  END
> <From>
www.bindingDB.org

> <BindingDB monomerid>
```

14094

> <BindingDB monomer Link>
http://www.bindingdb.org/bind/chemsearch/marvin/MolStructure.jsp?monomerid=14094

> <HET ID>
n/a

> <BindingDB Display Name>
2-ketobenzothiazole 34

> <TARGET Biomolecule 1>
Thrombin

> <TARGET Source Organism 1>
Homo sapiens (human)

> <TARGET Sequence 1>
MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQARSLLQR VRRANTFLEE
VRKGNLEREC VEETCSYEEA FEALESSTAT DVFWAKYTAC ETARTPRDKL
AACLEGNCAE GLGTNYRGHV NITRSGIECQ LWRSRYPHKP EINSTTHPGA
DLQENFCRNP DSSTTGPWCY TTDPTVRRQE CSIPVCGQDQ VTVAMTPRSE
GSSVNLSPPL EQCVPDRGQQ YQGRLAVTTH GLPCLAWASA QAKALSKHQD
FNSAVQLVEN FCRNPDGDEE GVWCYVAGKP GDFGYCDLNY CEEAVEEETG
DGLDEDSDRA IEGRTATSEY QTFFNPRTFG SGEADCGLRP LFEKKSLEDK
TERELLESYI DGRIVEGSDA EIGMSPWQVM LFRKSPQELL CGASLISDRW
VLTAAHCLLY PPWDKNFTEN DLLVRIGKHS RTRYERNIEK ISMLEKIYIH
PRYNWRENLD RDIALMKLKK PVAFSDYIHP VCLPDRETAA SLLQAGYKGR
VTGWGNLKET WTANVGKGQP SVLQVVNLPI VERPVCKDST RIRITDNMFC
AGYKPDEGKR GDACEGDSGG PFVMKSPFNN RWYQMGIVSW GEGCDRDGKY
GFYTHVFRLK KWIQKVIDQF GE

> <PDB ID 1>
1NAV(100%)

> <UniProtKB Accession Number 1>
P00734

> <TARGET Monomer-Polymer Link: 1>
http://www.bindingdb.org/jsp/dbsearch/PrimarySearch_ki.jsp?energyterm=kJ/mole&tag=r21&monomerid=14094&enzyme=Thrombin&column=ki&startPg=0&Increment=50&submit=Search

> <Enzymologic: Ki nM 1>
 1.6

> <Enzymologic: IC50 nM 1>
 51

> <Enzymologic: Kd nM 1>
n/a

> <Enzymologic: EC50/IC50 nM 1>
n/a

> <ITC: Delta_G0 kJ/mole 1>
n/a

> <pH 1>
7.4

> <temp 1>
310.15

> <TARGET Biomolecule 2>
Trypsin

> <TARGET Source Organism 2>
Bos taurus (bovine)

> <TARGET Sequence 2>
FIFLALLGAA VAFPVDDDDK IVGGYTCGAN TVPYQVSLNS GYHFCGGSLI
NSQWVVSAAH CYKSGIQVRL GEDNINVVEG NEQFISASKS IVHPSYNSNT
LNNDIMLIKL KSAASLNSRV ASISLPTSCA SAGTQCLISG WGNTKSSGTS
YPDVLKCLKA PILSDSSCKS AYPGQITSNM FCAGYLEGGK DSCQGDSGGP
VVCSGKLQGI VSWGSGCAQK NKPGVYTKVC NYVSWIKQTI ASN

> <PDB ID 2>
1E2J(>99%) 1KIM(100%) 1P7C(100%) 2VTK(100%)

> <UniProtKB Accession Number 2>
P00760

> <TARGET Monomer-Polymer Link: 2>
http://www.bindingdb.org/jsp/dbsearch/PrimarySearch_ki.jsp?energyterm=kJ/mole&tag=r21&monomerid=14094&enzyme=Trypsin&column=ki&startPg=0&Increment=50&submit=Search

> <Enzymologic: Ki nM 2>
 8

> <Enzymologic: IC50 nM 2>
n/a

> <Enzymologic: Kd nM 2>
n/a

> <Enzymologic: EC50/IC50 nM 2>
n/a

> <ITC: Delta_G0 kJ/mole 2>
n/a

> <pH 2>
7.4

> <temp 2>
310.15

$$$$

Molecule Block

Compound name and chemical description are highlighted in **blue** in the sample file. For detailed information on the format specification, see http://www.mdl.com/solutions/white_papers/ctfile_formats.jsp.

BindingDB Data Blocks

<From>: Documents that the SDfile was obtained from BindingDB.

<BindingDB monomerid>: The identifier of this compound in BindingDB

<BindingDB Monomer Link>: The hyperlink for this compound in BindingDB

<BindingDB Monomer Display Name>: The compound name displayed in BindingDB's web-interface.

<HET ID>: The corresponding HET id used by PDB.

<TARGET Biomolecule 1>: Name of the biomolecular target (usually a protein) for which the first set of affinity data for this compound were obtained.

<TARGET Source Organism 1>: Name of the source organism the biomolecular target was originated.

<TARGET Sequence 1>: The primary sequence of the biomolecular target.

<PDB ID 1>: PDB ID of complex or complexes in PDB with this compound (exact match) & protein(s) with >85% sequence identity. The BLAST sequence identity of the protein is listed in parentheses.

<UniProtKB Accession Number 1>: The corresponding UniProt Accession Number.

<TARGET Monomer-Polymer Link: 1>: URL link to an on-the-fly search for BindingDB data on the present compound ("Monomer") and biomolecular target ("Polymer"). Note that if the biomolecular target were a multicomponent complex, such as a dimer, then this field would be called **<TARGET Monomer-Complex Link 1>**.

<Enzymologic: Ki nM 1>: If first measurement is an enzymologic inhibition constant, this field provides the dissociation constant in nM.

<Enzymologic: IC50 nM 1>: If first measurement is an enzymologic IC50, this field provides the IC50 in nM.

<Enzymologic: Kd nM 1>: If first measurement is an enzymologic Kd, this field provides the dissociation constant in nM.

<Enzymologic: EC50/IC50 nM 1>: If first measurement is an enzymologic IC50, this field provides the IC50 in nM.

<ITC: Delta_G0 kJ/mole 1>: If first measurement is calorimetric, this field provides the free energy of binding.

<pH 1>: The solution pH of the first measurement.

<temp 1>: The temperature (K) at which the first measurement was carried out.

<TARGET Biomolecule 2>: Name of the biomolecular target (usually a protein) for which the third set of affinity data for this compound were obtained.

<TARGET Source Organism 2>: Name of the source organism the biomolecular target was originated.

<TARGET Sequence 2>: The primary sequence of the biomolecular target.

<PDB ID 2>: PDB ID of complex or complexes in PDB with this compound (exact match) & protein(s) with >85% sequence identity. The BLAST sequence identity of the protein is listed in parentheses.

<UniProtKB Accession Number 2>: The corresponding UniProt Accession Number.

<TARGET Monomer-Polymer Link: 2>: URL link to an on-the-fly search for BindingDB data on the present compound ("Monomer") and biomolecular target ("Polymer"). (Again, if the biomolecular target were a multicomponent complex, such as a dimer, then this field would be called **<TARGET Monomer-Complex Link 2>.**

<Enzymologic: Ki nM 2>: If third measurement is an enzymologic inhibition constant, this field provides the dissociation constant in nM.

<Enzymologic: IC50 nM 2>: If third measurement is an enzymologic IC50, this field provides the IC50 in nM.

<Enzymologic: Kd nM 2>: If third measurement is an enzymologic Kd, this field provides the dissociation constant in nM.

<Enzymologic: EC50/IC50 nM 2>: If third measurement is an enzymologic EC50, this field provides the IC50 in nM.

<ITC: Delta_G0 kJ/mole 2>: If third measurement is calorimetric, this field provides the free energy of binding.

<pH 2>: The solution pH of the second measurement.

<temp 2>: The temperature (K) at which the second measurement was carried out.

N.B. If the SDfile included additional measurements for this compound ("Monomer"), the additional affinity data would be provided with numbers 2, 4, etc.