

BindingDB SDfile Format

July 16, 2007

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 research groups may have published measurements, or a single research 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 BindingDB web-site.

A sample extract from a BindingDB SDfile follows; it lists one compound, Tacrine Dimer 3b, 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.

Tacrine Dimer 3b
Marvin 05020610432D

```
37 42 0 0 0 0          999 V2000
-1.6125 2.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.0250 2.7145 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.8500 2.7145 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-3.2625 2.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.0250 1.2855 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.8500 1.2855 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.6125 0.5711 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.0250 -0.1434 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.8500 -0.1434 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.6125 -0.8579 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-3.2625 -0.8579 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.8500 -1.5723 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-2.0250 -1.5723 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-3.2625 0.5711 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.7875 0.5711 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.0730 0.9836 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0.6415 0.5711 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 1.3559 0.9836 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 2.0704 0.5711 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 2.7849 0.9836 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 3.4994 0.5711 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 4.2138 0.9835 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 4.9283 0.5710 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5.7533 0.5710 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.1658 1.2542 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.9908 1.2542 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 7.4033 1.9687 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.9908 2.6832 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.1658 2.6832 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5.7533 1.9687 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.9908 -0.1122 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.1658 -0.1122 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 7.4033 -0.8267 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 5.7533 -0.8267 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.1658 -1.5411 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 6.9908 -1.5411 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 7.4033 0.5710 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 1 2 1 0 0 0 0
 1 5 2 0 0 0 0
 2 3 2 0 0 0 0
 3 4 1 0 0 0 0
```

```
4 6 2 0 0 0 0
7 5 1 0 0 0 0
5 6 1 0 0 0 0
6 14 1 0 0 0 0
7 8 2 0 0 0 0
14 9 2 0 0 0 0
10 8 1 0 0 0 0
8 9 1 0 0 0 0
9 11 1 0 0 0 0
10 13 1 0 0 0 0
11 12 1 0 0 0 0
12 13 1 0 0 0 0
7 15 1 0 0 0 0
15 16 1 0 0 0 0
16 17 1 0 0 0 0
17 18 1 0 0 0 0
18 19 1 0 0 0 0
19 20 1 0 0 0 0
20 21 1 0 0 0 0
21 22 1 0 0 0 0
22 23 1 0 0 0 0
23 24 1 0 0 0 0
25 24 2 0 0 0 0
24 32 1 0 0 0 0
37 26 2 0 0 0 0
37 31 1 0 0 0 0
30 25 1 0 0 0 0
25 26 1 0 0 0 0
27 26 1 0 0 0 0
27 28 1 0 0 0 0
28 29 1 0 0 0 0
29 30 1 0 0 0 0
33 31 2 0 0 0 0
31 32 1 0 0 0 0
32 34 2 0 0 0 0
33 36 1 0 0 0 0
34 35 1 0 0 0 0
35 36 2 0 0 0 0
```

M END

> <From>

www.bindingDB.org

> <BindingDB monomer Link>

<http://www.bindingdb.org/bind/chemsearch/marvin/MolStructure.jsp?monomerid=8963>

> <HET ID>

AA7

> <BindingDB Monomer Display Name>

Tacrine Dimer 3b

> <TARGET Biomolecule 1>

Acetylcholinesterase (AChE)

> <TARGET Source Organism 1>

Rattus norvegicus

> <TARGET Sequence 1>

```
MRPPWYPLHT PSLASPLLFL LLSLLGGGAR AEGREDPQLL VRVRGGQLRG
IRLKAPGGPV SAFLGIPFAE PPVGSRRFMP PEPKRPWSGI LDATTFQNV
YQYVDTLYPG FEGTEMWNPV RELSEDCLYL NVWTPYPRPT SPTPVLIIY
GGGFYSGASS LDVYDGRFLA QVEGTVLVSM NYRVGTFGFL ALPGSREAPG
NVGLLDQRLA LQWVQENIAA FGGDPMSVTL FGESAGAASV GMHILSLPSR
```

SLFHRAVLQS GTPNGPWATV SAGEARRRAT LLARLVGCPP GGAGGNDTEL
ISCLRTRPAQ DLVDHEWHVL PQESIFRFSF VPVVDGDFLS DTPDALINTG
DFQDLQVLVG VVKDEGSYFL VYGVPGFSKD NESLISRAQF LAGVRIGVPQ
ASDLAAEAVV LHYTDWLHPE DPAHLRDAMS AVVGDHNVVC PVAQLAGRLA
AQGARVYAYI FEHRASLTW PLWVGVPVPHGY EIEFIFGLPL DPSLNYTVEE
RIFAQRLMQY WTNFARTGDP NDPRDSKSPR WPPYTAAQQ YVSLNLKPLE
VRRGLRAQTC AFWNRFLPKL LSATDTLDEA ERQWKAEFHR WSSYVMVHWKN
QFDHYSKQER CSDL

> <PDB ID 1>

n/a

> <TARGET Monomer-Polymer Link: 1>

http://www.bindingdb.org/jsp/dbsearch/PrimarySearch_ki.jsp?energyterm=kJ/mole&tag=r21&monomerid=8963&enzyme=Acetylcholinesterase+%28AChE%29&column=ic50&startPg=0&Increment=50&submit=Search

> <Enzymologic: Ki nM 1>

n/a

> <Enzymologic: IC50 nM 1>

0.2

> <ITC: Delta_G0 kJ/mole 1>

Not search

> <pH 1>

7.4

> <T: C 1>

37

> <TARGET Biomolecule 2>

Butyrylcholinesterase (BuChE)

> <TARGET Source Organism 2>

Homo sapiens

> <TARGET Sequence 2>

mshkvtiici rflfwflllc mligkshted diiiatknkgk vrgmnlvtfv
gtvtaflgip yaqpplgrlr fkkpqsitkw sdiwnatkya nsccqnidqs
fpgfhgsemw npntdlsecd lylnvwipap kpknatvliw iygggfgtgt
sslhvydgkf larvervivv smnyrvgalg flalpgnpea pgnmgldfqq
lalqvwqkni aafggnpksv tlfgesagaa svslhlhllspg shslftrail
qsgsfnapwa vtslyearnr tlnlakltgc sreneteiik clrnkdpqei
llneafvvpv gtpslsvnfgp tvdgdfltdm pdillelgqf kktqilvgvn
kdegtaflvy gapgfskdn siitrkefge glkiffpgvs efgkesilfh
ytdwvddqrp enyrealgdv vgdynficpa leftkkfsew gnaaffyyfe
hrssklpwe wmgvmhgyei efvfglpler rdnytkaei lrsivkrwa
nfakygnpne tqnnstswpv fksteqkylt lntestrimt klraqqcrfw
tsffpkvlem tgnidaeawe wkagfhrwnn ymmdwknqfn dytskkescv
gl

> <PDB ID 2>

n/a

> <TARGET Monomer-Polymer Link: 2>

http://www.bindingdb.org/jsp/dbsearch/PrimarySearch_ki.jsp?energyterm=kJ/mole&tag=r21&monomerid=8963&enzyme=Butyrylcholinesterase+%28BuChE%29&column=ic50&startPg=0&Increment=50&submit=Search

> <Enzymologic: Ki nM 2>

n/a

> <Enzymologic: IC50 nM 2>

54

> <ITC: Delta_G0 kJ/mole 2>

Not search

> <pH 2>

7.4

> <T: C 2>

37

\$\$\$\$

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 monomer Link>: The hyperlink to this compound in BindingDB

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

<HET ID>: The PDB HET ID for this compound, if available.

<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 of the biomolecular target.

<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.

<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.

<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.

<T: C 1>: The temperature (degrees C) at which the first measurement was carried out.

<TARGET Biomolecule 2>: Name of the biomolecular target (usually a protein) for which the second 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.

<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 second measurement is an enzymologic inhibition constant, this field provides the dissociation constant in nM.

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

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

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

<T: C 2>: The temperature (degrees C) at which the second measurement was carried out.

N.B. If the SDfile included additional measurements for this compound (Monomer ID=8963), the additional affinity data would be provided with numbers 3, 4, etc.