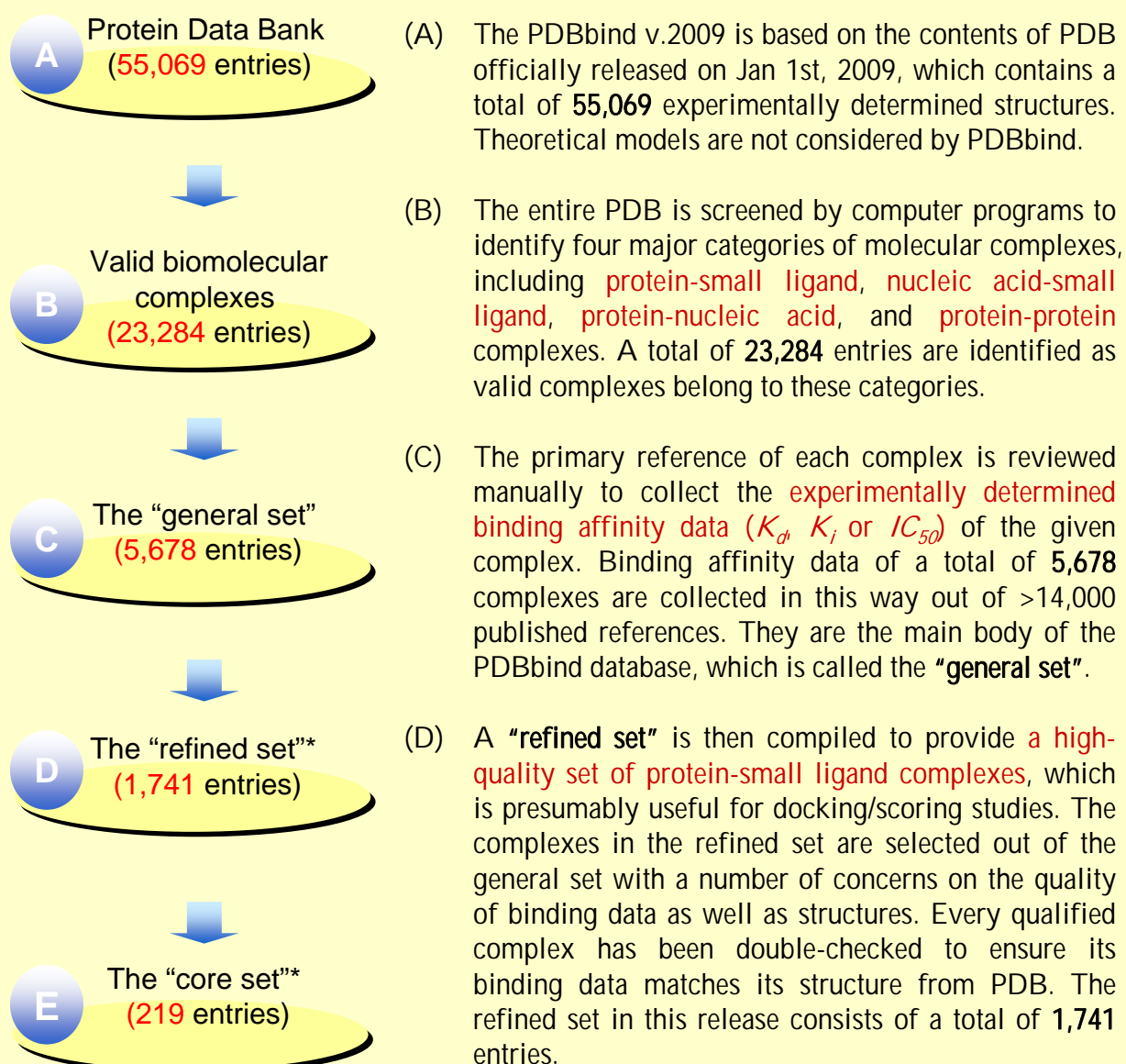


## A Brief Introduction to the PDBbind Database v.2009

The PDBbind database provides a collection of experimentally measured binding affinity data exclusively for the biomolecular complexes available in the Protein Data Bank (PDB). It thus represents a linkage between energetic and structural information of these complexes, which may be used in various molecular recognition studies. Since its first public release in May 2004, over 1,400 users from 40 countries around the world have already registered to use this database. The PDBbind database is now updated annually to keep up with the growth of PDB. The current release is **version 2009**.

The PDBbind database is compiled through a hierarchical process.



\* Only protein-ligand complexes are considered in this data set.

- (E) A “core set” is further compiled as a benchmark with an appropriate and relatively stable size. **The core set is selected through a systematic, non-redundant sampling of the refined set with special focus on the diversity on structures and binding data.** Briefly, the refined set is clustered by protein sequence similarity using a cutoff of 90%. For each cluster containing at least four members, the one with the highest binding affinity, the one with the lowest binding affinity, and the one with a medium binding affinity are selected as the representatives of this cluster. The core set in this release consists of a total of **73** clusters (**219** protein-ligand complexes) in total.

### History of the PDBbind Database

Version	Entries in PDB	Biomolecular Complexes	Complexes with binding data		
			General Set	Refined Set**	Core Set**
2002	19621	5671	1446	800	---
2003	23790	5897	1763	900	---
2004	28991	6847	2276	1091	231
2005	34338	9775	2756	1296	288
2006*	34338	9775	2632	1122	234
2007	40876	11822	3124	1300	195
2008	48092	18211	4300	1401	210
2009	55069	23284	5678	1741	219

\*: V.2006 is in fact a correction to v.2005, both of which were based on the same release of PDB.

\*\* : Only the complexes formed between proteins and small-molecule ligands are considered in these sets.

### Features

**New**

- ❑ The PDBbind database originally considered only the complexes formed between proteins and small-molecule ligands. Since version 2008, other types of biomolecular complexes in PDB are covered by PDBbind as well. This release contains binding data for protein-ligand (**4277**), protein-protein (**1053**), protein-nucleic acid (**304**), and nucleic acid-ligand (**44**) complexes.
- ❑ For users' convenience, PDBbind also provides processed “clean” structural files in addition to binding data for the protein-ligand complexes in the general set. In brief, the biological unit of each complex is split into a protein molecule saved in the PDB format, and a ligand molecule saved in the Mol2 format and the SD format. Atom types and bond types on the ligand molecule are assigned carefully as appropriate. These structural files can be readily utilized by most molecular modeling software.

### References

The PDBbind database is developed by a collaboration between Prof. Renxiao Wang's group at the Shanghai Institute of Organic Chemistry and Prof. Shaomeng Wang's group at the University of Michigan. To cite the PDBbind database, please refer to:

1. Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. *J. Med. Chem.* **2005**, *48*, 4111-4119.
2. Wang, R.; Fang, X.; Lu, Y.; Wang, S. *J. Med. Chem.* **2004**, *47*, 2977-2980.