

AutoBind: automatic extraction of protein–ligand-binding affinity data from biological literature

Darby Tien-Hao Chang¹, Chao-Hsuan Ke², Jung-Hsin Lin^{3,4} and Jung-Hsien Chiang^{2,*}

¹ Department of Electrical Engineering, ² Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 70101, ³ School of Pharmacy, National Taiwan University, Taipei 10051 and ⁴ Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan

jchiang@mail.ncku.edu.tw

[Bioinformatics, Vol. 28 No.16, pp. 2162-2168 \(2012\)](#)

Determination of the binding affinity of a protein-ligand complex is important to quantitatively specify whether a particular small molecule will bind to the target protein. In the past decades, several databases of protein-ligand binding affinities have been created via visual extraction from literature. However, such approaches are time-consuming and most of these databases are updated only a few times per year. Hence, there is an immediate demand for an automatic extraction method with high precision for binding affinity collection.



Recently, we have created a new database of protein-ligand binding affinity data, AutoBind, based on automatic information retrieval. We first compiled a collection of 1586 articles where the binding affinities have been marked manually. Based on this annotated collection, we designed four sentence patterns that are used to scan full-text articles as well as a scoring function to rank the sentences that match our patterns. The proposed sentence patterns can effectively identify the binding affinities in full-text articles. Our assessment shows that AutoBind achieved 84.22% precision and 79.07% recall on the testing corpus. Currently, 13616 protein-ligand complexes and the corresponding binding affinities have been deposited in AutoBind from 17221 articles.

Table 1. Comparison of AutoBind to other databases

	Protein Complex	Number of entries (Affinity Data)	Last updated date	First published year
AutoBind	37929	13616	February 13, 2013	This work
PDBBind	7986	7986	September 22, 2011	2004
Binding MOAD	16955	5630	2010	2005
BindingDB	3056	1817	March, 2010	2001
AffinDB	474	748	2006	2006
PLD	149	149	2003	2003
LPDB	195	195	2001	2001

Reference:

1. Wang R., Fang X., Lu Y., Wang S., The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. *Journal of Medicinal Chemistry* 2004; 47: 2977-2980.
2. Hu L., Benson M.L., Smith R.D., Lerner M.G., Carlson H.A., Binding MOAD (Mother Of All Databases). *Proteins: Structure, Function, and Bioinformatics* 2005; 60: 333-340.
3. Liu T., Lin Y., Wen X., Jorissen R.N., Gilson M.K., BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Research* 2007; 35: D198-

D201.

4. Block P., Sotriffer C.A., Dramburg I., Klebe G., AffinDB: a freely accessible database of affinities for protein–ligand complexes from the PDB. *Nucleic Acids Research* 2006; 34: D522-D526.
5. Puvanendrapillai D., Mitchell J.B.O., Protein Ligand Database (PLD): additional understanding of the nature and specificity of protein–ligand complexes. *Bioinformatics* 2003; 19: 1856-1857.

Copyright 2013 National Cheng Kung University