

建立一個自動擷取蛋白質-配體結合親和力數據之資料庫

張天豪¹、柯兆軒²、林榮信³、蔣榮先^{2,*}

¹ 國立成功大學電機工程學系

² 國立成功大學資訊工程學系

³ 國立台灣大學藥學系

jchiang@mail.ncku.edu.tw

Bioinformatics, Vol. 28 No.16, pp. 2162-2168 (2012)

測

量蛋白質-配體的親和力的重要性在於測量一個特定的小分子是否可以有效的與目標蛋白質結合。在過去數十年間，雖然已有數個經由手工擷取的親和性數據資料庫被製作並發表，然而這是一個非常耗費時間的動作，而且這種方式無法讓資料庫的更新頻率跟上新發表文章的速度。因此開發一個自動化的擷取方式在此顯得相對重要。我們利用文獻擷取技術針對蛋白質親和性資訊擷取進行研究，目的在於開發一個自動從文獻中擷取有效的蛋白質親和性資訊系統。



我們開發了一個訊息自動檢索演算法擷取蛋白質-配體結合親和力數據後並建立一個資料庫，名為AutoBind。我們採用了1586篇經由PDBbind^[1]確認過具有親和性數據的文章設計出4個句型模板，並設計一個有效的評估函數來選出所有符合句型模板的句子中最有可能描述蛋白質親和性的句子。利用我們所提出的句型結合評估函數可以有效地識別的結合親和力文章全文。實驗結果顯示，AutoBind獲得了84.22%的準確度和79.07%召回率。目前共有13616的蛋白質 - 配體複合物和相應的結合親和力從17221篇文章中收錄出來並存放在AutoBind中。

表 1. AutoBind與其他相關資料庫的數據比較

	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

參考文獻：

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., Sottriffer 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