GEMDOCK 之驗證與應用以及資料融合在虛擬藥物篩選之應用

學生：陳彥甫　　指導教授：楊進木 博士

國立交通大學生物科技學系（研究所）碩士班

摘　　要

我們已經發展了一個以演化式方法為基的彈性配體嵌合（flexible ligand docking）程式。我們的方法稱之為 GEMDOCK，結合了基因演算法為核心的分子嵌合（molecular docking）技術與經驗式計分方程式。我們的演算法結合了分離的與連續性的總體搜尋策略合併區域搜尋策略，來加速收斂；也就是達成文後結果所述的快速潛在可能分子辨識。在我們之前的研究中，GEMDOCK 已經被測試在 100 個蛋白質與配體的複合物上，並且達到 79%的成功率。在本研究中，GEMDOCK 再次被重新測試在一個更大而且更具有歧異度的測試組上，這個測試組包含了 305 個由蛋白質資料庫（PDB）取得的蛋白質與配體的複合物。測試結果有 78%的最低能量嵌合構形和結晶資料差距在 2.0A 個方均根差（RMSD）以內。當考慮結構水的時候成功率更可以達到 83%。我們根據這個測試資料探討 GEMDOCK 在分子嵌合實驗上的影響因子，並且透過修正這些因子來提升 GEMDOCK 的表現。此外我們也評估測試了 GEMDOCK 在人類的 α-thrombin 上的虛擬篩選能力。人類 α-thrombin 是一個絲胺酸蛋白水解酶，在人類的凝血作用中扮演著重要的起始調控角色。在人類 α-thrombin 的虛擬藥物篩選測試結果顯示，在所有已知的 thrombin 抑制劑都被找回時，enrichment factor 達到 20，而且偽陽性比率只有 3.4%。我們分析並且驗證了 GEMDOCK 在原始的經驗式計分方程與 pharmacophore-based 計分方程。我們發現 pharmacophore 的知識的確有助於降低在虛擬藥物篩選時的偽陽性機率。這樣的結果顯示 GEMDOCK 是強健的，並且可以是有效的分子嵌合與虛擬篩選的工具。在這個研究的另外一部份是將資料融合運用在虛擬藥物篩選上。我們透過結合不同的計分方程產生的排名來達到結合不同計分方程的效果。結果顯示這樣的方式的確可以提升命中準確率與降低偽陽性的機率。這個研究採用了三個具有藥物發展潛力的目標蛋白質虛擬藥物篩選結果作排名與分數結合的分析，他們分別是 thymidine kinase, dihydrofolate reductase 和 estrogen receptor 。這三個資料分析的結果皆顯示這樣的結合可以提升準確度並降低偽陽性比率。我們的資料分析顯示，透過排名結合不同的計分方程式在偽陽性比率與 enrichment 上可以產生比單獨的計分方程式要好的虛擬藥物篩選表現

Validation and Application of GEMDOCK and Application of Data Fusion in Virtual Screening

Student : Yan-Fu Chen Adviser : Dr. Jinn-Moon Yang

Department (Institute) of Biological Science and Technology
National Chiao Tung University

ABSTRACT

We have developed an evolutionary approach for flexible ligand docking. Our approach, GEMDOCK, used a Generic Evolutionary Method for molecular DOCKing and an empirical scoring function. The former combines both discrete and continuous global search strategies with local search strategies to speed up convergence, whereas the latter results in rapid recognition of potential ligands. In our pervious works, GEMDOCK has been tested on 100 protein-ligand complexes with 79% successful rate. In this study, GEMDOCK was tested on a large diverse data set of 305 protein-ligand complexes from the Protein Data Bank. In 78% of these complexes, the docked lowest energy structures of ligands had root-mean-square derivations (RMSDs) below 2.0? with respect to the corresponding crystal structures. The successful rate increased to 83% if the structures water molecules were retained. According to this test, we explored and investigated influence factors of GEMDOCK on molecular docking and improved the performance of GEMDOCK by revising these factors. We also evaluated GEMDOCK on virtual screening for human a-thrombin which is a serine protease and plays a central role in the initiation of blood coagulation. The enrichment factor was 20 for the screening validation result of a-thrombin and the false positive rate was 3.4% when the true positive rate was 100%. We analyzed and validated GEMDOCK with origin scoring functions and our pharmacophore-based scoring function and found that the pharmacophore knowledge indeed could reduce the false positive rates on virtual screening. These results suggest that GEMDOCK is robust and can be a useful tool for molecular docking and virtual database screening. In the meanwhile we present the results of a computational study in which we show that combining scoring functions by their rankings results in an improvement in the ability to increase hit rate and reduce false positive rate. The analysis for three virtual screening results of pharmaceutical interest: thymidine kinase, dihydrofolate reductase, and estrogen receptor provides the differences of combinations between ranking and scoring. Both combinations of ranking and scoring enhance the enrichment factors and reduce the false positive rate. Our data shows that the combination of ranking from different scoring functions farther provides an exciting reduction in false positive rates distinguished by individual scoring functions and contributes the increasing to enrich the active compounds in virtual screening.