

行政院國家科學委員會專題研究計畫 成果報告

演化式方法快速藥物篩選與毒性預測

計畫類別：個別型計畫

計畫編號：NSC91-2320-B-009-001-

執行期間：91年08月01日至92年07月31日

執行單位：國立交通大學生物科技學系

計畫主持人：楊進木

報告類型：精簡報告

處理方式：本計畫可公開查詢

中華民國92年9月22日

行政院國家科學委員會專題研究計畫成果報告

演化式方法快速藥物篩選與毒性預測

High throughput screen and drug toxicity prediction using evolutionary approaches

計畫編號：NSC91-2320-B-009-001

執行期限：91 年8 月1 日至 92 年7 月31 日

主持人：楊進木 助理教授 交通大學生物科技學系暨生物資訊所
e-mail: moon@cc.nctu.edu.tw

一、中文摘要

傳統開發藥物的方法發展一種新藥，平均約需十二年，其成本約為 4.5 億美金(超過一百億台幣)，因人類基因體計畫及合理化藥物設計的發展，預期可縮短藥物開發時間及降低其開發成本，提供新藥開發的契機，而 protein-ligand docking 及 lead optimization 的技術是合理化藥物設計的核心技術，此技術能應用於辨識潛在藥物，預測這些藥物的副作用(side effects)以及毒性(toxicity)，因此在藥物發展的初期可作為低成本、快速度藥物測試的技術。在本計畫中我們已完成了三項主要工作：一.發展出一套完整的 protein-ligand docking 及藥物篩選工具 (drug-screening tool),此工具之核心演算法為知識為基之演化式方法，簡稱為 GEMDOCK，其網頁位置為 <http://gemdock.life.nctu.edu.tw>;二. 實際用於發展登革熱蛋白水解酶及 SARS coronavirus 3C-like protease 抑制劑快速篩選。三. 經由 cytochrome P450 蛋白的電腦藥物毒性預測來早期預測藥物毒性。

關鍵詞：藥物篩選、演化式方法、protein-ligand docking、合理化藥物設計、藥物毒性

Abstract

The average design cycle of traditional methods used for drug discovery is often over 12 years, which is too slow to meet today's ever increasing and urgent demand to fight various diseases. The cost of discovering a new drug is often over 450 million US dollars. Fortunately, human genomic project and rational computer-aided drug design are promising directions for shortening the time and reducing the cost for new

drug discovery. The protein-ligand docking and inverse protein-ligand problem is one of core problems of rational computer-aided drug design. Integrating protein-ligand docking and lead optimization techniques can be used to identify drug candidates and investigate unwanted side effects and toxicity with low-cost and fast-speed drug testing in the early stages of development. In this project, we achieved three main results: 1) developing a robust and flexible protein-docking tool, called GEMDOCK (Yang and Kao 2000; Yang and Kao 2000; Lin, Yang et al. 2003; Yang 2003), the URL is position at <http://gemdock.life.nctu.edu.tw>, 2) applying GEMDOCK to two practical applications that are to develop the inhibitors for the proteases of dengue virus and SARS virus, 3) Predicting the drug toxicity by measuring the binding affinity to P450.

Keywords: drug screening , evolutionary approach, protein-ligand docking, rational drug design, toxicity prediction

二、緣由與目的

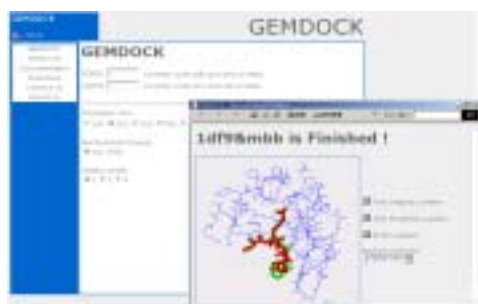
合理化藥物設計的核心技術乃是 ligand-protein docking 及 ligand-protein inverse docking 兩者。整合此兩者的工具是現代藥物發展的關鍵技術之一，此整合技術可應用於辨識潛在藥物，預測這些藥物的副作用以及毒性，因此在藥物發展的初期可作為低成本、快速度藥物測試的技術。Ligand-protein docking 技術的中心概念在於通過電腦高速運算，由已知立體構形(conformation)的彈性配體(flexible ligand)與蛋白質巨分子兩者來預測其 ligand-protein 複合體(complex)的結構。Docking process 以分子間交互作用能量的極小化(minimizing)為依歸。近年來由於高解析度的蛋白質結構較早年易於取得，以及 docking process 可藉由電腦模擬(computer-based simulation)進行全自動化演算之故，Ligand-protein docking 技術愈來愈受重視，應用範圍亦愈加廣泛。此技術可用於篩選或設計能與蛋白質上特定位置達成良好的結構上與化學上互補(complementarity)的 ligand，此即能幫助我們尋得具備藥物潛力的小分子物質。而過去的研究顯示，ligand-protein docking 程式所預測的 ligand-protein complex 構形實際上已相當接近實驗所得的 complex 真實結晶結構。過去傳統藥物發展過程中，經常使用實驗高速篩選(*in vitro* high-throughput screening)平台來尋找可能的候選藥物。這樣的過程不僅耗費大量金錢與時間，而且所找出的候選者(candidates)也無法應用實驗過程獲得的數據，輔助後續設計藥物結構的最佳化與評估。1990 年代的研究發現，許多藥物開發案例在最後階段的失敗常肇因於藥物動力特性不良與其毒性(van de Waterbeemd and Gifford 2003)。因此目前藥物發展的趨勢傾向利用電腦高速運算特性，加速藥物開發與篩選。其中組合性化學(combinatorial chemistry)與電腦高速篩選(*in silico* high-throughput screening)技術(Dearden 2003)，在近幾年經常被應

用到藥物開發的早期階段上，評估潛在藥物的 ADMET(absorption, distribution, metabolism, elimination and toxicity)特性是否符合，進一步提高與加速實驗高速篩選的效率與準確率(van de Waterbeemd and Gifford 2003)

三、結果、討論與成果評估

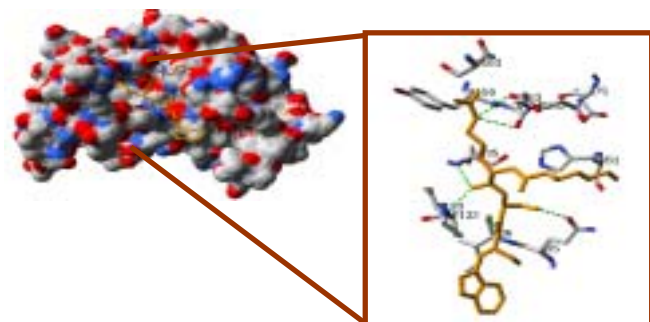
我們在本計畫中達到下列多項成果:

1. 建構完成通用的電腦輔助藥物快速篩選系統，簡稱為 GEMDOCK(Yang and Kao 2000; Yang and Kao 2000; Lin, Yang et al. 2003; Yang 2003)，其網頁位置為 <http://gemdock.life.nctu.edu.tw>，此系統具更快且精確描述 ligand-protein interaction 能量變化的計分函式。GEMDOCK 可應用於各種藥物開發平台的前端開發(如下圖一所示：綠色為原來 ligand 位置，紅色為預測之 ligand 位置)。此系統可提供 web based 的遠端服務。



圖一、GEMDOCK web based 遠端服務系統圖示

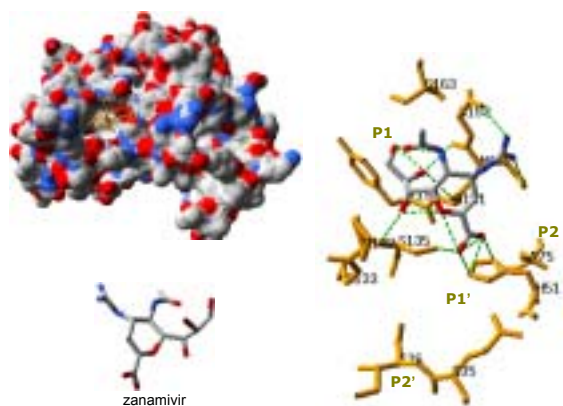
2. 已將第一項之工具實際用於發展登革熱蛋白水解酶 NS3(PDB Code: 1BEF, 1DF9)抑制劑快速篩選之專屬系統，此系統會針對這類蛋白水解酶的重要胺基酸，如活性區域中的 Ser 135、Asp 75、His 51(catalytic triad)等(詳見下表一及圖二)，這些胺基酸會與潛在的抑制劑間形成作用力，如靜電力、氫鍵(H)及凡得瓦爾力(vW)等，系統會自動偵測並加重其權重，以提昇藥物快速篩選系統的專一性及正確性



圖二、登革熱蛋白水解酶 NS3 與受質之交互作用示意圖

3 以藥物資料庫 CMC(Comprehensive medicinal chemistry : about 7,000 Compounds)及 MDD(MDL drug data report : about 200,000 Compounds) 為藥物

篩選的資料庫，分別針對登革熱蛋白水解酶 NS3(1BEF 及 1DF9)的活性區域進行嵌合預測。目前已完成的部分為 CMC 資料庫對 1BEF 的活性區域所作的潛在藥物篩選，初步篩選出 23 個 Compounds 對蛋白水解酶 NS3 的潛在抑制劑，我們目前以化合物和 Ser135 作用的官能基作為分類依據，可將這 23 個 Compounds 粗分為四大類：磷酸根，碳酸根，亞硫酸根及其他(Prepared for publications)。我們以 zanamivir 為例說明 Docking 結果，zanamivir 屬於碳酸根類，本來用作為治療感冒。圖三是化合物 zanamivir 對 1BEF 所做的嵌合預測結果，zanamivir 穩定地結合在 1BEF 活性中心的位置上，且與重要的胺基酸均有形成氫鍵。

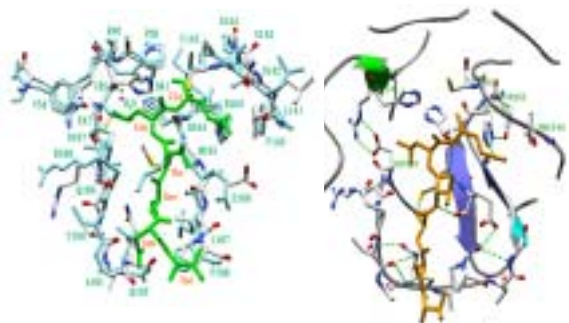


圖三、zanamivir 對 1BEF 所做的嵌合預測結果

- 我們也將第一項之工具實際用於發展 SARS coronavirus 3C-like protease 抑制劑快速篩選之專屬系統，在 SARS 的實際成果如下(Prepared for publications)：已初步成功證實本實驗室發展之 docking 工具 GEMDOCK 可應用於 SARS coronavirus 同一 Family 之 3C-like proteases (1P9U, 1P9S)之 docking 模擬。成功以 Homology Model 及分子模擬方法 modeling SARS 3CL^{pro}。此蛋白質模擬結構已登錄在 PDB，其 code 為 1Q1X，經與嗣後發表的結晶結構比較後證實兩者甚為相似，此或可作為本實驗室結構模擬策略正確之一佐證。此電腦模擬結構之最大貢獻在包含一抑制劑，較適合應用於藥

物篩選。我們已利用 3CL^{pro} 切割蛋白鏈後的殘剩 peptide fragment (VNSTLQ 與 TSAVLQ)及 AG7088(用於治療 rhinovirus 引起的感冒症狀, 已進入臨床實驗)作為基礎, 其 docking 結果如圖四所示, 印證我們工具之正確性。本期計畫將發展 SARS 3CL^{pro} 專屬抑制劑篩選系統。

5.



圖四、電腦模擬之蛋白質結構(1Q1X)與結晶結構(1Q2W)之比較示意圖。經比較後發現兩者在 binding site 區域甚為相似, CPK 著色部分為 1Q2W, 淺藍色者為 1Q1X, peptide VNSTLQ 標示為綠色、電腦模擬預測之 Docking Results

5. 我們已利用高斯演化方法(Gauss Evolutionary Method; GEMDOCK)發展自動化高速藥物篩選, 成果已發表於(Yang 2003)。

6. 結合演化式方法(evolutionary approaches)與 neural network 並應用於分類(classification)以及智慧型代理人(agent)等問題上已取得相當成果(Yang, Horng et al. 2000; Yang and Kao 2001)。此技術為發展 4D/5D QSAR 的重要技術。

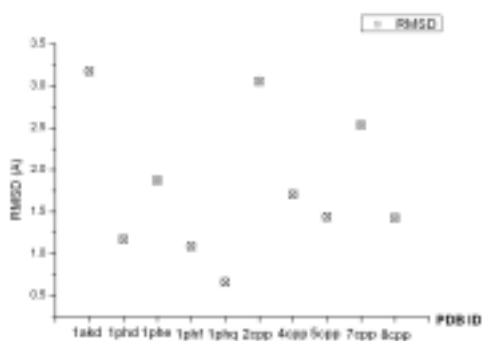
7. 我們在 cytochrome P450 上的研究已達成兩個具體成果, 簡述如下

- 受質結合區的分析: 我們由 SCOP 資料庫中找出 P450 monooxygenase family 中的 8 個 domain 針對其結構與序列特性加以分析比對。發現 P450 monooxygenase 在序列上相似度不高(31.3-12.4%), 但結構上卻相當相似(RMSD: 4.3-2.3)(圖五)。在 binding site 附近的氨基酸也具有高度保留性。
- GEMDOCK 實際應用測試: 我們採用了 Gyorgy M. Keseru(Keseru 2001)所使用的測試組, 以我們的工具(GEMDOCK)做分子嵌合測試(ligand docking test), 結果如圖六所示。我們的工具和 Gyorgy M. Keseru 所使用的 FlexX 結

合 consences scores 的結果要好。



圖五、在 1AKD 結合區中重要氨基酸位置標示



圖六、GEMDOCK 測試結果

Reference

Dearden, J. C. (2003). "In silico prediction of drug toxicity." J Comput Aided Mol Des **17**: 119-127.

Keseru, G. M. (2001). "A virtual high throughput screen for high affinity cytochrome P450cam substrates. Implications for in silico prediction of drug metabolism." J. Comput. Aided Mol. Des. **15**(7): 649-57.

Lin, E. S., J. M. Yang, et al. (2003). "Modeling the binding and inhibition mechanism of nucleotide and sulfotransferase using molecular docking." Chinese Chemical Society.

van de Waterbeemd, H. and E. Gifford (2003). "ADMET in silico modeling: towards prediction paradise." Nature reviews drug discovery **2**: 192-204.

Yang, J. M. (2003). "An evolutionary approach for molecular docking." Lecture Notes in Computer Science **2724**: 2372-2383.

Yang, J. M., J. T. Horng, et al. (2000). "A genetic algorithm with adaptive mutations and family competition for training neural networks." International Journal of Neural Systems **10**(5): 333-352.

Yang, J. M. and C. Y. Kao (2000). "A Family competition evolutionary algorithm for automated docking of flexible ligands to Proteins." IEEE Transaction on Information Technology in Biomedicine **4**(3): 225-237.

Yang, J. M. and C. Y. Kao (2000). "Flexible ligand docking using a robust evolutionary algorithm." Journal of Computational Chemistry **21**(11): 988-98.

Yang, J. M. and C. Y. Kao (2001). "A robust evolutionary algorithm for training neural networks." Neural Computing and Application **10**(3): 214-230.

