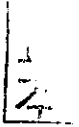


計畫編號：DOH92 -TD-1132



行政院衛生署九十二年度科技研究發展計畫

建構預測登革熱抑制劑之快速篩選系統

研究報告

執行機構：國立交通大學生物科技系

計畫主持人：楊進木

協同主持人：楊昫良

研究人員：陳俊辰、陳彥甫、林建宏、董其樺、沈再威

執行期間：92年1月1日至92年12月31日

本研究報告僅供參考，不代表本署意見

目 錄

Part A

1. 中文摘要 (3)
2. 英文摘要 (4)
3. 九十二年度計畫著作一覽表 (5)
4. 九十二年度計畫重要研究成果
 - 4.1 GEMDOCK 應用於預測 protein-ligand docking (6)
 - 4.2 GEMDOCK 應用於篩選藥物資料庫 (6)
 - 4.3 GEMDOCK 應用於 DenII NS3 serine protease 的前導藥物篩選 (8)
5. 九十二年度科技計畫重要研究成果產出統計表 (9)

Part B

1. 前言與背景分析 (11)
2. 研究目的與執行成果概要 (14)
3. 研究方法 (15)
4. 結果與討論
 - 4.1 GEMDOCK 應用於預測 protein-ligand docking (18)
 - 4.2 GEMDOCK 應用於篩選藥物資料庫 (20)
 - 4.3 GEMDOCK 應用於 DenII NS3 serine protease 及 SARS coronavirus
3C-like protease (21)
 - 4.4 CMC 藥物資料庫的篩選結果 (23)
 - 4.4.1 前 200 名化合物一覽表 (29)
 - 4.5 280 個蛋白水解酶抑制劑的篩選結果 (44)
 - 4.5.1 前 100 名化合物一覽表 (45)
5. 結論與建議 (57)
6. 審查意見之回覆 (57)
7. 參考文獻 (62)

行政院衛生署科技研究發展計畫原始數據資料庫
資料讀我檔案

計畫名稱：建構預測登革熱抑制劑之快速篩選系統
計畫編號：DOH92-TD-1132
執行機構：國立交通大學生物科技系
計畫主持人：楊進木
計畫主持人服務單位：國立交通大學生物科技系
計畫主持人職稱：助理教授

研究報告中文摘要：

目前藥物設計的主要方法之一為針對已知結構的蛋白質，應用電腦軟體工具在化合物資料庫中篩選潛在的抑制劑，此法可以有效縮短尋找潛在抑制劑的時間，並降低實驗所需的成本。目前較為廣泛運用的電腦輔助藥物設計軟體有 GOLD、DOCK 及 FlexX 等。GEMDOCK 為本計畫預計開發的一套輔助藥物設計軟體，可以計算小分子化合物與蛋白質鉗合(docking)的能量，準確地預測小分子與蛋白質活性區域(active site)結合的位置及位向。GEMDOCK 在預測已知結構之 protein-ligand docking 的 112 個預測結果中，計有 90 個的 RMSD ≤ 2.0 Å；另外在篩選藥物資料庫方面，將 GEMDOCK 應用於 TK (thymidine kinase) 上，已知 PDB 中與 TK 共同結晶的 ligand 有 10 個，再加上由藥物資料庫中隨機篩選的 1,000 個化合物，根據 docking 計算所得的分數做排序，排名前 5% 化合物中，有 10 個已知 PDB ligand 中的 9 個，尋回比率達 90%。排名前 5% 的化合物可分為三群：1. real ligand, 2. purine analog, 3. pyrimidine analog, 這三群均為核苷酸的相似物，可見 GEMDOCK 能有效地自 1,000 個化合物中篩選出真正會與 TK 結合的小分子，及與其結構相似的其他化合物。登革熱為台灣每年夏季常見的流行病，其蛋白水解酶(serine protease)可以用來做為藥物設計的標的。我們已以 GEMDOCK 應用於登革病毒的蛋白水解酶(DenII NS3 serine protease)，從 CMC 藥物資料庫(7,500 compounds)及已知結構的蛋白水解酶抑制劑(280 compounds)中篩選潛在的抑制劑，做為藥物設計的前導藥物。篩選的結果依照 docking 的位置可分為三群：1. 位於 binding site, 2. 位於 S1 pocket, 3. 位於 S1 pocket 及 catalytic site, 這三群化合物可以作為選擇前導藥物的參考。我們相信前的 2%(約 150 compounds)，極有可能成為登革病毒蛋白水解酶之抑制劑的前導藥物。部分 compounds 已由協同主持人楊昶良教授，進行細胞實驗中。

本計畫以 GEMDOCK 為工具，建構一套前導藥物(lead compound)的快速篩選系統。此系統若與傳統的藥物開發方式互補，將可以組合成一個更完善的藥物開發體系，並大大地縮短臨床實驗所需的時間與經費。

中文關鍵詞(至少三個)：登革熱，登革病毒蛋白水解酶，虛擬藥物篩選，鉗合，前導藥物

Research Data Archive, Department of Health, The Executive Yuan, R.O.C.

Readme file

Project Title: A high throughput system for identifying the inhibitors of dengue fever

Project Number: DOH92-TD-1132

Executing Institute: National Chiao-Tung University

Principal Investigator(P.I.): Jinn-Moon Yang

P.I. Position Title: Assistant Professor

P.I. Institute: National Chiao-Tung University

Abstract:

Virtual screening of chemical databases is now a well-established method for finding new lead compounds for rational drug design. When used prior to experimental screening, it can be considered as a powerful computational filter for reducing the size of a chemical library that will be further experimentally tested. Several docking programs are now available that generally are able to predict known protein-bound ligand poses with averaged accuracies of 1.5-2.0Å, such as GOLD, DOCK, and FlexX, etc. GEMDOCK is the program we developed in this project, and we apply the tool to establish a virtual screening system. Dengue fever, a graphically widespread epidemic caused by dengue virus type I-IV, annually causes many social and commercial damages in Taiwan. The NS3 serine protease is a good target for drug design. We had applied GEMDOCK to predict 112 protein-bound ligand poses, and there were 90 ligand poses with RMSD below 2.0Å. We also have applied GEMDOCK to virtual screening to TK (thymidine kinase). There were 10 known protein-ligand complexes of TK in the PDB (Protein Data Bank). For screening the dataset (1,055 compounds; 1,045 and 10 are from MDDR and PDB, respectively) to TK and scoring the compounds, we retrieved 9 of the 10 TK ligands from the top 5% scorer. The top 5% scorer could be clustered into three groups: 1) real ligands, 2) purine analogs, and 3) pyrimidine analogs, and all they were analogs of nucleoside. These experiments have improved that GEMDOCK could well discriminate real ligands and their analogs from other compounds. Finally, we screen CMC (Comprehensive Medicinal Chemistry) and protease inhibitors from PDB to DenII serine protease with GEMDOCK. The top 5% scorer of the docking results could be clustered to three groups by the docking positions: 1) in the binding site, 2) in the S1 pocket, and 3) in the S1 pocket and the catalytic site. The compounds of the three groups could be considered as possible lead compounds for rational drug design.

Keyword: dengue fever, DenII serine protease, virtual screening, protein-ligand docking, lead compound, GEMDOCK

九十二年度計畫著作一覽表

計畫名稱：建構預測登革熱抑制劑之快速篩選系統

主持人：楊進木 計畫編號：DOH92-TD-1132

列出貴計畫於本年度中所有計畫產出於下表，包含已發表或已被接受發表之文獻、已取得或被接受之專利、擬投稿之手稿 (manuscript) 以及專著等。「計畫產出名稱」欄位請依「臺灣醫誌」參考文獻方式撰寫；「產出形式」欄位則填寫該產出為期刊、專利、手稿或專著等，舉例如下：

序號	計畫產出名稱	產出形式	SCI*
1	J.-M. Yang and C.-C. Chen, "GEMDOCK: A generic evolutionary method for molecular docking", to appear in <i>Proteins: Structure, Function and Genetics</i>	期刊	√
2	J.-M. Yang "Development and evaluation of a generic evolutionary method for protein-ligand docking," to appear in <i>Journal of Computational Chemistry</i>	期刊	√
3	E.-S. Lin, J.-M. Yang , and Y.-S. Yang, "Modeling the binding and inhibition mechanism of nucleotide and sulfotransferase using molecular docking", <i>Journal of the Chinese Chemical Society</i> , vol. 50, pp. 655-663, 2003.	期刊	√
4			
5			
6			
7			
8			
9			
10			

* SCI: Science Citation Index，若發表之期刊為 SCI 所包含者，請打「√」。

九十二年度計畫重要研究成果

計畫名稱：建構預測登革熱抑制劑之快速篩選系統

主持人：楊進木 計畫編號：DOH92-TD-1132

1. 本計畫開發了一套輔助藥物計軟體 GEMDOCK，可用來計算小分子化合物與蛋白質活性區域間結合的能量，預測小分子化合物結合於活性區域最穩定的位置及位向。並以此為核心工具，建構出一套針對登革病毒蛋白水解酶的抑制劑快速篩選系統。GEMDOCK 可應用於各種藥物開發平台的前端開發(如 Figure 1 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。GEMDOCK 在預測的準確度方面，目前已有頗佳的之預測統計結果，預測結果依 RMSD 做分類，在 112 個預測結果中，有 90 個 $\leq 2.0 \text{ \AA}$ ，此意味著預測的 ligand 位置與原來結晶的位置幾乎重合 [1, 2]。

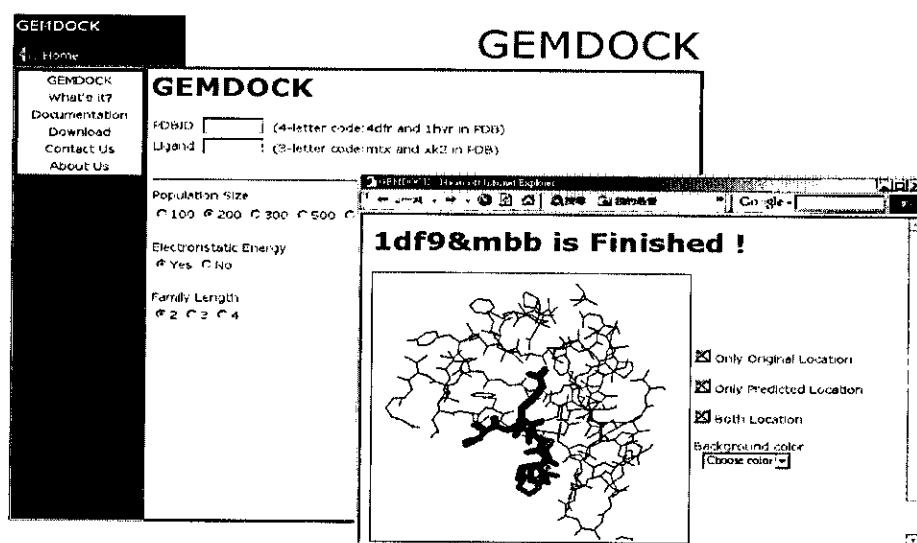


Figure 1 : GEMDOCK web based 遠端服務系統圖示

2. GEMDOCK 在篩選藥物資料庫方面，將 GEMDOCK 應用於 TK (thymidine kinase) 上，已知 PDB 中與 TK 共同結晶的 ligand 有 10 個，再加上由藥物資料庫中隨機篩選的 990 個化合物，根據 docking 計算所得的分數排序，排名前 5% 化合物中，有 10 個已知 PDB ligand 中的 9 個 (Figures 2 and 3)，尋回的機率達 90%。排名前 5% 的化合物可分為三群：1. real ligand，2. purine analog，3. pyrimidine analog，這三群均為核苷酸的相似物，可見 GEMDOCK 能有效地自 1,055 個化合物中篩選出真正會與 TK 結合的小分子，及與其結構相似的其他化合物。

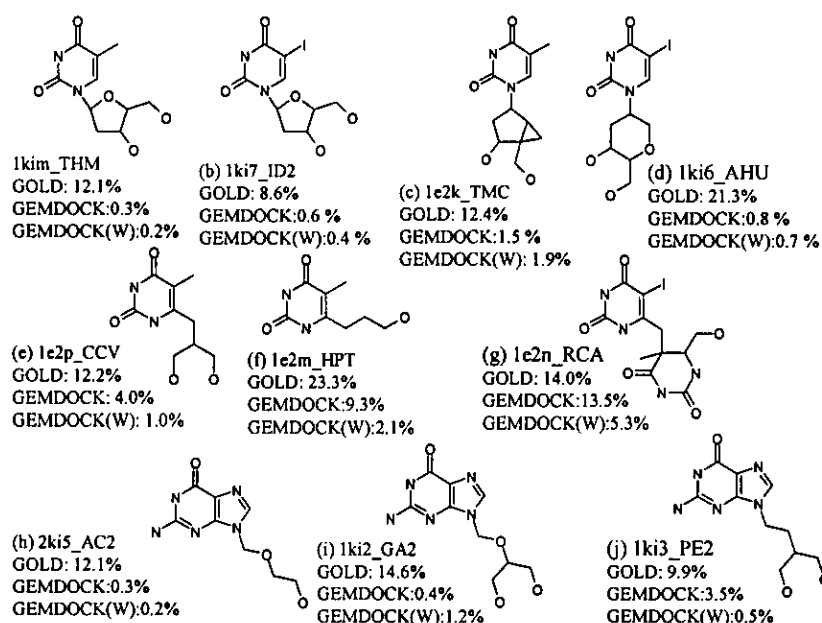


Figure 2: GEMDOCK 及 GOLD 在 thymidine kinase ligands 上的結果比較，以 1kim_THM 為例，THM在GOLD的排名為前 12.1%，在GEMDOCK的排名為前 0.3%，在蛋白質活性區域的重要氨基酸加重計分後 (GEMDOCK(W))，排名提昇為前 0.2%

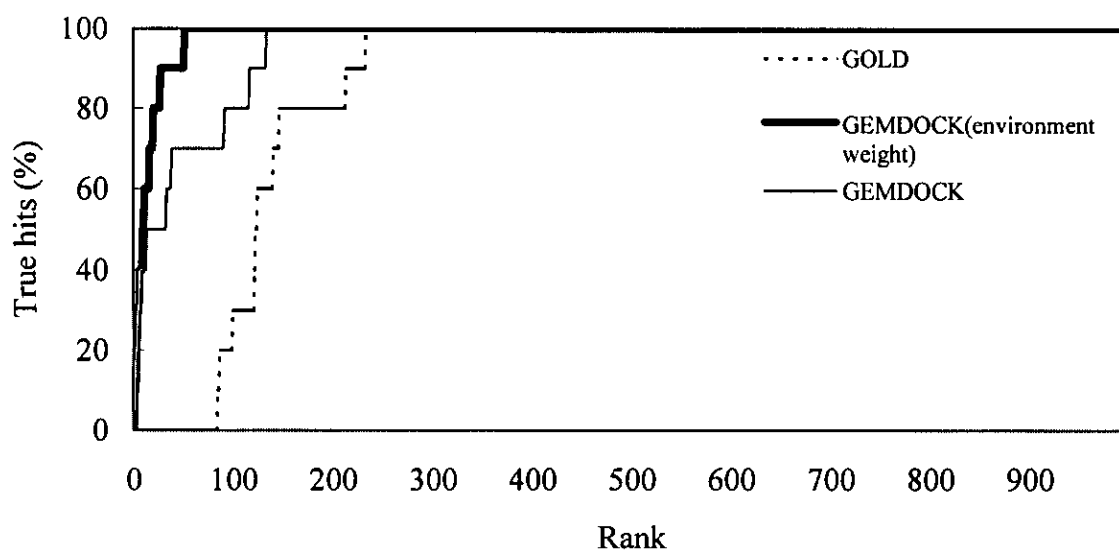


Figure 3: GEMDOCK 及 GOLD 在篩選 1000 個化合物上的確準率比較圖，GEMDOCK 在 200 名前準確率即已達到 100%，GOLD 則在 300 名前準確率才達到 100%

3. 將 GEMDOCK 應用於登革病毒蛋白水解酶 (DenII NS3 serine protease)上，自 CMC 藥物資料庫，做為藥物設計的前導藥物。篩選的結果依其 1) 分子與蛋白質嵌合之位置 2)

和蛋白質嵌合之結合作用區, 3) 分子基本骨架結構作為分類標準(Figure 4)。

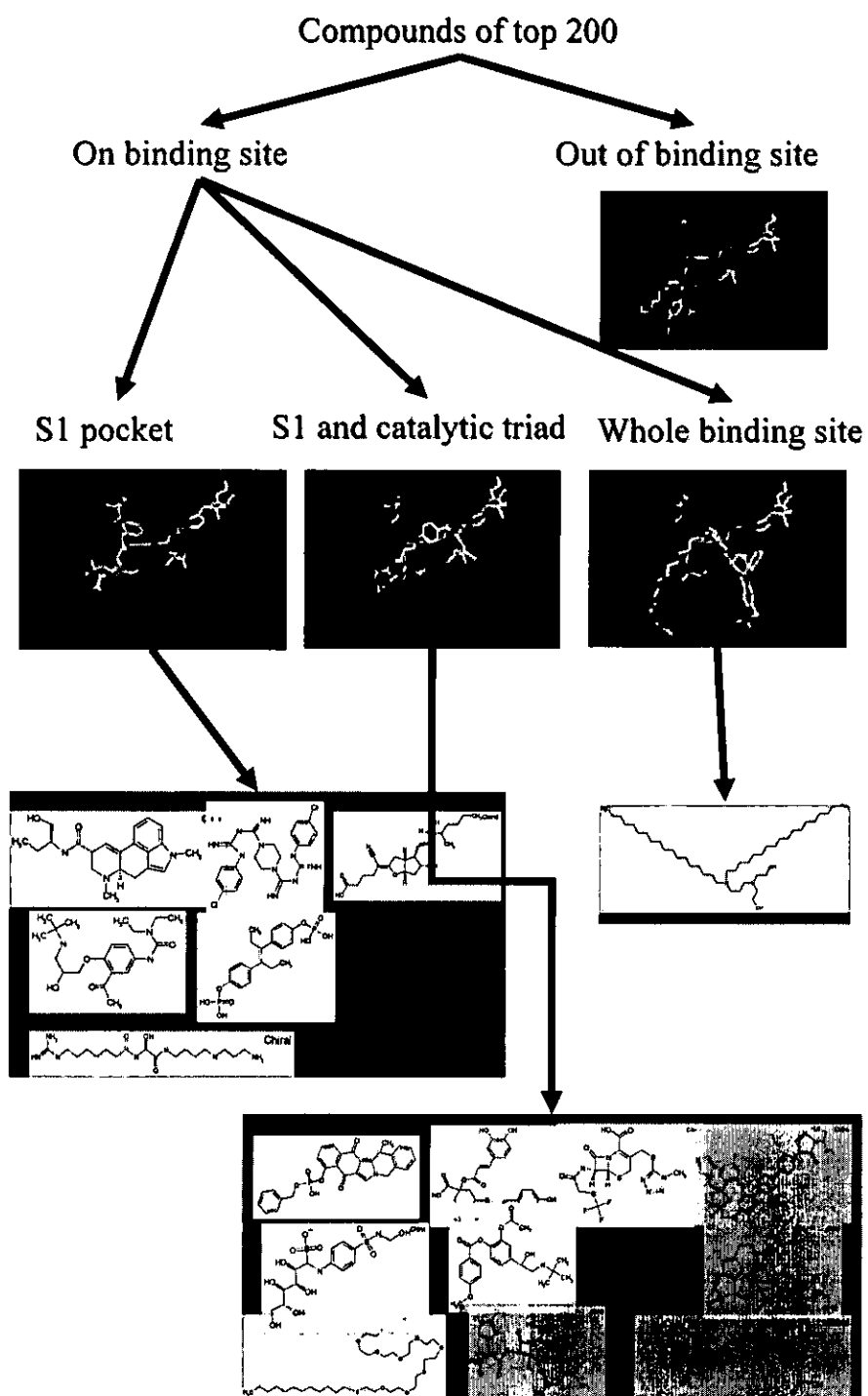


Figure 4 : 前兩百名之分子分群

4. 自己知結構的蛋白水解酶抑制劑(protease inhibitor)中篩選登革病毒蛋白水解酶潛在的抑制劑，做為藥物設計的前導藥物。篩選的結果依其 docking 的位置可分為三群：1) 位於 binding site (Figure 5)，2) 位於 S1 pocket (Figure 6)，3) 位於 S1 pocket 及 catalytic site (Figure 7)，這三群化合物可以作為選擇前導藥物的參考



Figure 5

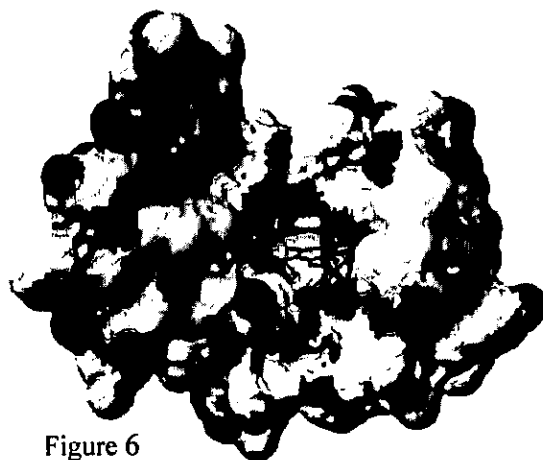


Figure 6



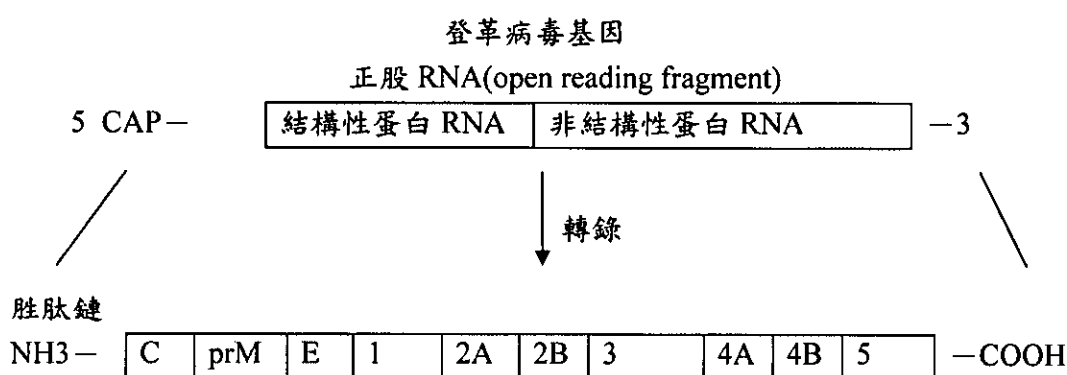
Figure 7

1. 前言與背景分析

登革病毒(Dengue virus)分類上屬於黃質病毒科(Flaviviridae)中黃質病毒屬(Flavivirus)內的登革病毒亞屬，共分四個血清型，廣泛分布於包括菲律賓、中國、越南、印尼、斯里蘭卡、千里達與哥倫比亞等二十餘個亞熱帶、熱帶國家，主要傳染媒介為白線斑蚊與埃及斑蚊。台灣曾於 1915、1931、1942、2002 年爆發四次嚴重的登革熱大流行，其中尤以 1942 年為最，估計約 80% 人口罹患，至今每年夏秋兩季仍有輕重不等的疫情傳出。典型原發性登革熱症狀劇烈，雖然致死率 < 1%，但不同血清型交叉感染衍生的登革出血熱(Dengue hemorrhagic fever, DHF)或登革休克症候群(Dengue shock syndrome, DSS)則有驚人的 15 ~ 50% 死亡率，實為台灣地區民眾健康的一大威脅。根據國外流行病學的統計顯示，登革熱的擴散與發生機率，會隨著居住人口稠密度與病媒蚊的密度而提高。台灣近年來都會區擴大，人口稠密度上升，但是環境衛生並未跟上人口與經濟發展的脚步，因此每逢兩季病媒蚊的孳生源(積水容器)就會大幅增加，使得病媒蚊大量孳生，實有爆發登革熱大流行的隱憂。

在登革病毒感染初期，會出現諸如發燒、頭痛和骨頭疼痛等類似於感冒的症狀，因此容易受病人忽視而延誤就醫，加重病情。甚者在感染未癒期間再度受到感染，引發出血性登革熱，導致死亡。針對這樣的情形，治本的方式是教育民眾自覺，加強環境的衛生與研發有效的疫苗來減少感染的可能。對於已經不幸感染的，希望能夠找到有效的藥物來加以治療，縮短病程或加以治癒。整體而言，目前世界上尚無穩定、安全的疫苗可供防治登革熱，亦無治療藥物。因此若能夠針對登革熱開發出可用的藥物，那麼不論對於台灣或是全世界熱帶地區的民眾，都將是一項極具價值的貢獻。

登革病毒為 RNA 病毒，遺傳物質是單一的正股 RNA (single-stranded RNA, (+) strand)，經轉譯生成一條多蛋白鏈(polyprotein)。如下圖，這條多肽鏈是登革病毒蛋白質的前驅物，經寄主的酵素(signal peptidase)切割後，才具有個別的功能，主要的切割位置分別位於 C-prM、prM-E、E-NS1、NS4A-NS4B，切割後生成核蛋白(core protein)、膜蛋白(pre-membrane protein)、套膜蛋白(envelope protein)及其他非結構性蛋白。

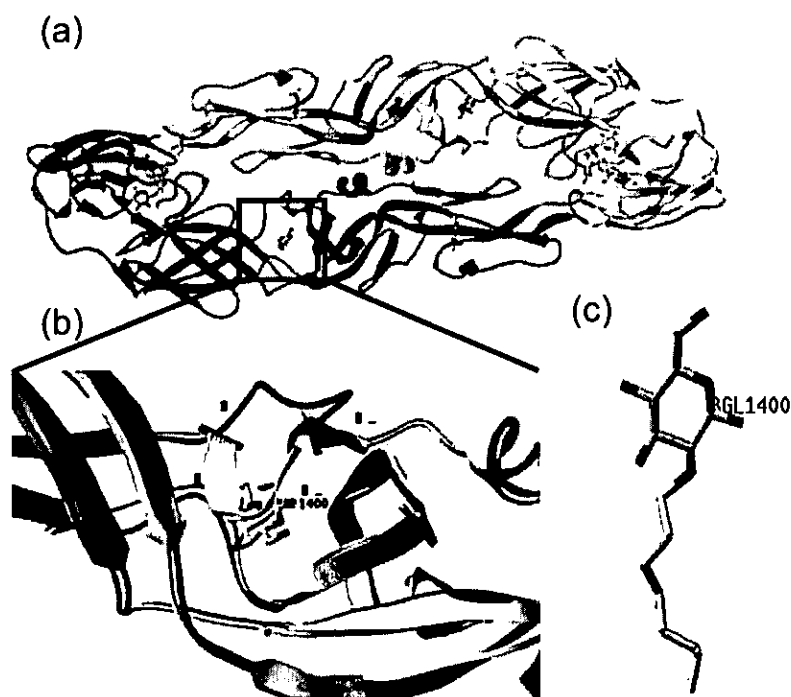


摘自 <http://www.science.mcmaster.ca/Biology/Virology/23/dengue.htm>

登革病毒基因體所轉譯出的非結構性蛋白質 NS3 (nonstructural protein 3) 雖不參與病毒套膜的組合過程，但仍與病毒的成熟有相當大的關係。NS3 為一多功蛋白(multifunctional protein)，屬 serine protease 之一種，與非結構性蛋白 NS2B 結合後方具活性及專一性，並共同作用於病毒 RNA 所轉譯出的蛋白質前驅物(polyprotein precursor)使其成熟。另外，NS3

還具有三磷酸腺苷酶(nucleoside triphosphatases)、解螺旋酶(helicases)、五端 RNA 三磷酸酶(5'-RNA triphosphatase)等功能，前兩者作用於病毒 RNA 複製過程；五端 RNA 三磷酸酶則作用於 RNA 複製後，在五端的部位加上 cap，使 RNA 結構穩定而不會被分解。這三種功能皆於病毒 RNA 成熟過程中扮演相當重要的角色。而且 NS3 的結構已被解出，我們可利用結構上的資訊發展可能的抑制劑，以治療登革熱或其他相關疾病[3, 4]。

登革病毒的套膜蛋白(envelope protein)亦為一可能的藥物標的蛋白質。此蛋白質為構成病毒外殼的主要蛋白，是主要的抗原決定部位，宿主的免疫系統即針對此蛋白質產生抗體。目前研究顯示，virus-antibody complex 進入細胞的能力遠高於單獨存在的病毒，這亦是不同血清型登革熱交叉感染後會引發登革出血熱的可能機制之一。此蛋白質目前已有兩個



X-ray 結晶結構，分別是 IOAM 及 IOAN，其中 IOAM 為化合物與套膜蛋白共同結晶，此化合物為 n-octyl- β -D-glucoside(β -OG)，它會穩定地結合在胺基酸序列 262-280 所形成的忌水性凹槽中，將 k1 loop 撐開，形成 open form；IOAN 則沒有化合物的結晶在凹槽內，k1 loop 會擋住凹槽的開口，並轉換成 k1 sheet，形成 close form。登革病毒進入細胞前，尚未經過酸性環

境的刺激，套膜蛋白為 close form，進入細胞之後，經過低 pH 值的刺激，套膜蛋白會轉變成 open form，下一步則會與微粒體的膜進行融合的作用，將 RNA 釋放到細胞質中。 β -OG 與凹槽的結合會影響套膜蛋白結構的轉變，可能可抑制套膜蛋白與微粒體間的融合作用。另外，當這個凹槽內的某些胺基酸發生點突變時，包括 Gln52、Phe193、Lys204、Thr268、Ile270、Leu277、Gly275、Leu277 及 Phe279，會使得套膜蛋白的結構更加穩定(close form)，此時則需要更低的 pH 值才會引起套膜蛋白的結構發生改變，才會發生融合的作用。由此可知，k1 loop 確實是影響套膜蛋白結構的重要區域，若能針對此結合區域(binding site)設計潛在的抑制劑，則極有可能達到抑制登革病毒的目的。套膜蛋白結構如圖所示：(a)登革病毒套膜蛋白(IOAM)，由兩條 chain 組成；(b) k1 loop 的局部放大，藍色為 IOAM 的結構(open form)，淡黃色為 IOAN 的結構(close form)；(c) n-octyl- β -D-glucoside 之結構[5]。

開發能有效治療登革熱的藥物，對於現階段國人的醫療保健極為重要。本計畫蒐集目

前已知結構的登革病毒蛋白，結合網路上可供搜索的藥物分子資料庫，利用本實驗室已開發出來的軟體，可快速並準確地尋找可能與登革病毒蛋白質結合的藥物，並配合生體實驗確認藥物的藥效及副作用，相信有可能發展出可治療登革熱的藥物，有利於我國民眾健康。傳統開發藥物的方法耗費大量時間與金錢，發展一種新藥平均約需十二年，其成本約為 4.5 億美金(超過一百億新台幣)，如此耗時及高成本的發展方法，無法滿足人類對抗疾病所需新藥的需求。幸運地，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。

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。

Ligand-protein inverse docking 技術應用於藥物開發初期，可以低成本且快速地預測藥物的可能副作用 (side effect) 以及毒性 (toxicity)，因之可大幅縮減後續的生體實驗以及藥物毒性測試的時間及金錢開銷。Ligand-protein inverse docking 技術的觀念在於搜尋可與單一 ligand 結合的多個蛋白質，與 ligand-protein docking 著重由單一蛋白質尋找可與該蛋白質結合的多個 ligands 恰好相反，而依照該 ligand 可與哪些蛋白質標的 (protein target) 結合，結合力強或弱，我們即可依之推測該藥物可能產生的副作用或毒性。

簡言之，Ligand-protein docking 就是以蛋白質活性區域與 ligand 的化學互補交互作用來辨識候選藥物 (candidate)，而 ligand-protein inverse docking 也是以交互作用為基準來辨識候選藥物對酵素或蛋白質的副作用以及毒性。綜上所述此整合工具有四項潛在功能：一、快速篩選潛在藥物；二、確認該藥物(或天然物)針對的目標 (target)，包含目前未知及次級 (secondary) 的 targets；三、預測該藥物可能造成的副作用及毒性；四、預測該藥物對生化反應途徑的可能干擾。

2. 研究目的與執行成果概要

Figure 3-1 所示為整個藥物開發系統的流程，包含了一套系統化的藥物搜尋步驟，其中涵蓋快速 ligand-protein docking (ligand 為剛體)、flexible ligand docking、及潛力藥物最佳化 (ligand optimization，此步驟以 ligand-protein inverse docking 技術為之，意在取得藥效最大，副作用及毒性最小的 ligands)，並可與傳統的藥物開發方式互補，組成一個更完善的藥物開發體系，可以有效地縮短臨床實驗所需的時間與經費。本計畫的研究目的在於發展一個能預測前導藥物 (lead compound) 的雛形系統，此系統應用 flexible ligand-protein docking 的核心技術，計算 ligand 可與 protein 結合的最佳位置，希望能準確地預測可能的前導藥物，以提供臨床實驗參考。在軟體的發展方面，已成功地利用高斯演化方法 (Gauss Evolutionary Method; GEMDOCK) 發展自動化高速藥物篩選，目前已有初步成果，在預測已知結構之 protein-ligand docking 的 112 個預測結果中，計有 90 個的 $RMSD \leq 2.0 \text{ \AA}$ 。另外在篩選藥物資料庫方面，將 GEMDOCK 應用於 TK (thymidine kinase) 上，已知 PDB 中與 TK 共同結晶的 ligand 有 10 個，再加上由藥物資料庫中隨機篩選的 1,045 個化合物，根據 docking 計算所得的分數做排序，排名前 5% 化合物中，有 10 個已知 PDB ligand 中的 9 個，尋回 (recall) 率達到 90%。

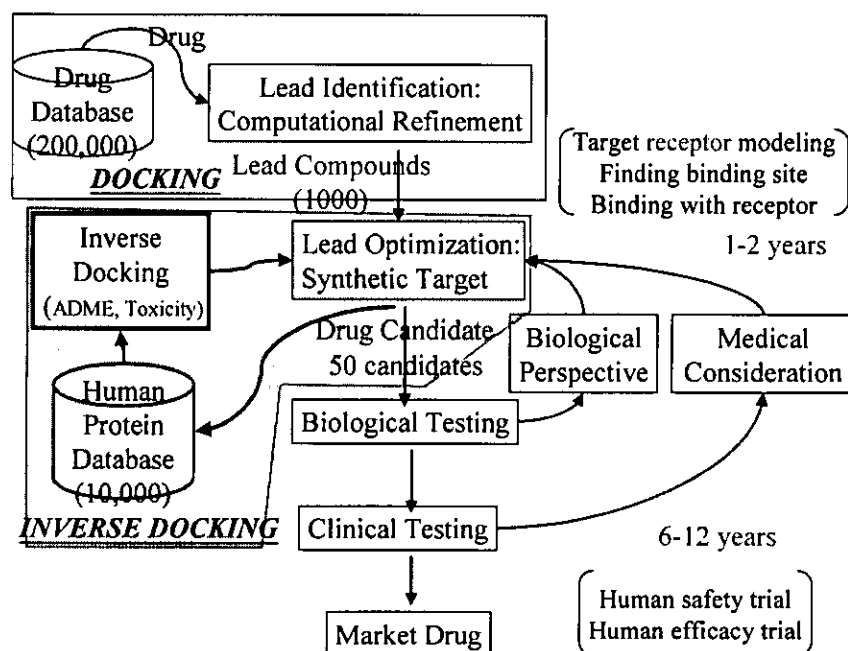


Figure 3-1. The framework of our rational drug design system

在預測登革病毒潛在抑制劑方面，以 DenII NS3 Serine protease 為標的蛋白質，對 CMC (Comprehensive Medicinal Chemistry, 5,742) 藥物資料庫及已有結構的蛋白酶抑制劑 (protease inhibitor, 280)，共計 6,022 個化合物進行 protein-ligand docking 篩選，並將 docking 的結果分為三群 1. binding site, 2. S1 pocket 及 3. S1 pocket and catalytic triad，預測可能的前導藥物 (lead compound)。

3. 研究方法

如 Figure 3-1 所示，研究方法主要有三個步驟 1. 準備 Docking 知識庫 2. 執行 Docking 工具庫 3. 分析 Docking 的結果及數據，分述如後：

1. 準備 Docking 知識庫：

(1) 藥物資料庫(drug database)：

- a. CMC (Comprehensive Medicinal Chemistry) 為藥物化合物資料庫，截至目前為止共收錄 7,937 筆化合物資料，內容包括化合物的立體結構、藥物種類、logP 值和 pKa 值等，且每年會依據美國藥典(the US Pharmacopoeia)認可的藥物名稱 (USAN, the United States Approved Names) 作更新。由 CMC 資料庫中截取分子量 200 到 750 之間的化合物來進行 docking，共計 5,742 個結構。
- b. 從蛋白質結構的資料庫 (Protein Data Bank) 中，搜尋與蛋白質共同結晶的蛋白酶抑制劑(protease inhibitor)，主要可分為三類，分別是 Aspartate protease、Cysteine protease 及 Serine protease 的抑制劑，共計 280 個可用來 docking 的結構。

(2) 確認標的蛋白與小分子化合物結合的活性區域(binding site)：

本實驗用作 docking 的標的蛋白為 DenII NS3 Serine protease(PDB code: 1DF9)[6]，這個結構是由登革病毒 NS3 protease 和 mung-bean Bowman-Birk inhibitor(MbBBI)共同結晶而成，但是由於 MbBBI 的分子量太大，與 NS3 protease 間的交互作用已屬 protein-protein interaction 範疇，故無法直接以 MbBBI 為 ligand 來做 docking，但卻可以利用 MbBBI 與 protease 間的作用關係，作為決定活性區域及篩選可能抑制劑的參考。如 Figure 3-2，與抑制劑或受質結合的位置位於圖中標示 P2、P1、P1 及 P2 之處，Table 3-1 條列了每個位置重要的胺基酸，及與抑制劑間形成的作用力。在對整個化合物資料庫作篩選時，將以蛋白質上這個位置周圍 10 Å 為半徑的範圍(461 個原子)做為計算的空間，預測化合物與蛋白質最穩定的結合位置。

2. Docking 工具庫：以演化式方法(evolutionary approaches)為核心演算法的 GEMDOCK。

3. Docking 結果及數據分析：排名並分群。

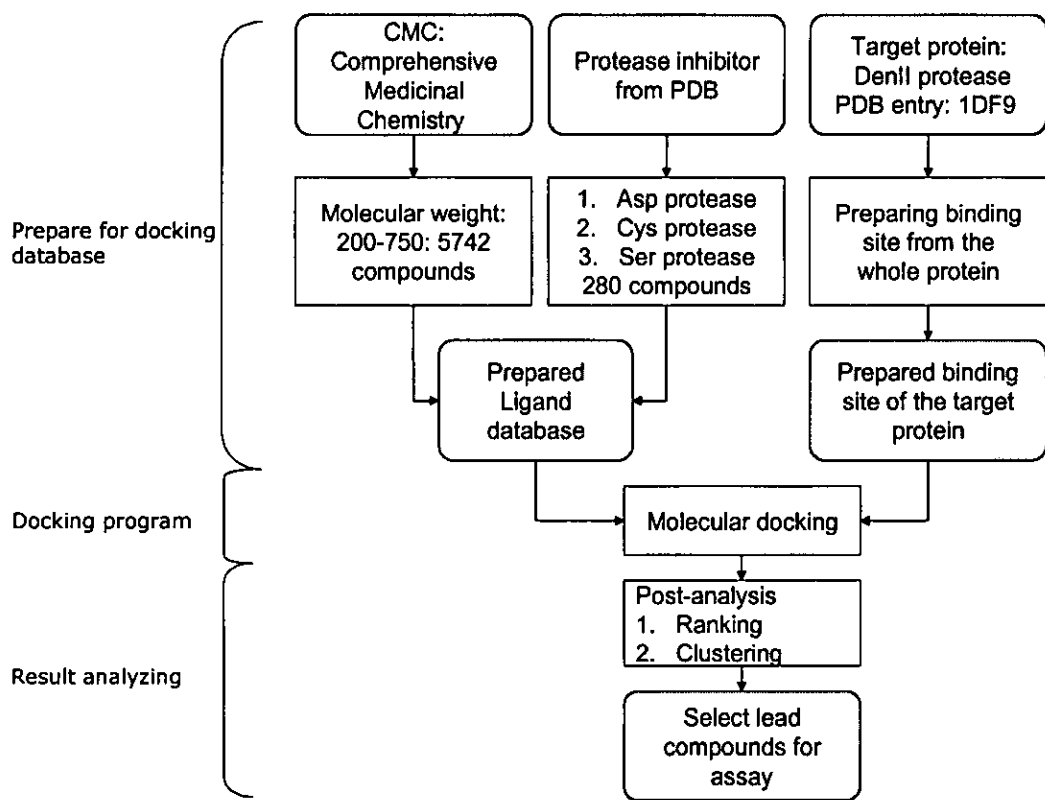


Figure 3-1

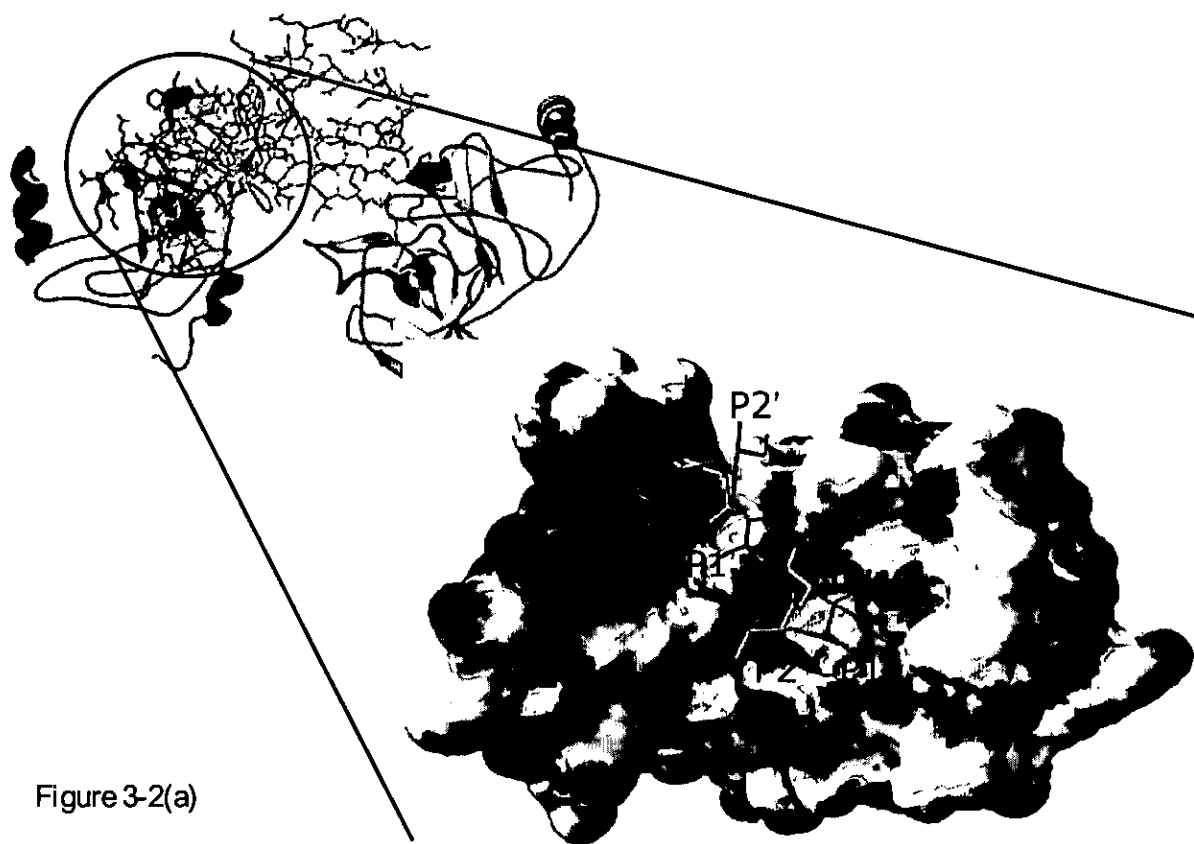


Figure 3-2(a)

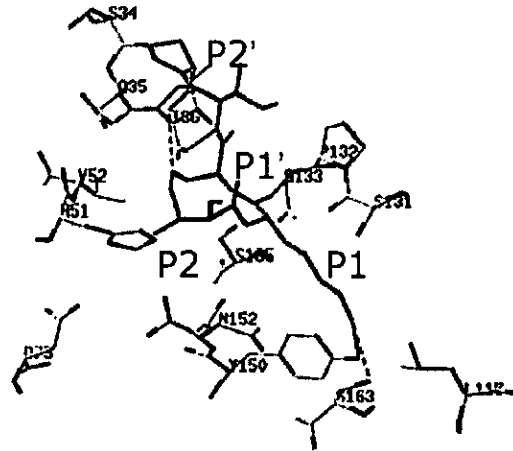


Figure 3-2(b)

Figure 3-2(a)：左圖為 1DF9 完整蛋白質結構，包括 chain A、B，右圖為 chain A 的表面圖，橘色為 MbBBI 的胺基酸片段，其序列為 Thr-Lys-Ser-Ile。

Figure 3-2(b)：活性區域與抑制劑間形成的作用力，綠色虛線為可能形成的氫鍵。

Table 3-1：活性區域的重要胺基酸與抑制劑間的作用關係。

Site	Residue	Binding pocket
P2	Thr	H51(E) ^a , N152(vW)
P1	Lys	L115(vW), S131(E), P132(vW), <u>G133</u> ^b (E), <u>S135</u> (E), Y150(E), <u>G151</u> (E), S163(E)
P1	Ser	I36(vW), H51(E), S135(E), V52(vW)
P2	Ile	S34(vW), Q35(vW), I36(E), P132(vW), G133(vW)

a. 括號內標示的是蛋白質活性區域的胺基酸與抑制劑間形成的作用力，E 為靜電力，包括氫鍵及一般的靜電力，vW 為凡得瓦力。

b. 標有底線者為構成 oxyanion hole 的胺基酸。

4. 結果與討論

以下簡述我們在本計畫達成的具體成果暨在藥物開發領域中研究獲得的相關成果：

4.1 GEMDOCK 應用於預測 protein-ligand docking

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，並已建構完成一通用的電腦輔助藥物快速篩選系統，簡稱為 GEMDOCK，網頁位置為 <http://gemdock.life.nctu.edu.tw>，此系統具備更快且精確描述 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 Figure 4-1 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。

在預測的準確度方面，目前已有頗佳的初步成果: Figure 4-2(a)、(b) 為 GEMDOCK 之預測結果，紅色為預測之 ligand 的結構，而白色為原 ligand 之 X-ray 結構；(c)、(d) 為 GEMDOCK 針對 112 個 protein-ligand docking 之預測統計結果，預測結果依 RMSD 做分類，在 112 個預測結果中，有 90 個 $\leq 2.0 \text{ \AA}$ ，此意味著預測的 ligand 位置與原來結晶的位置幾乎重合。

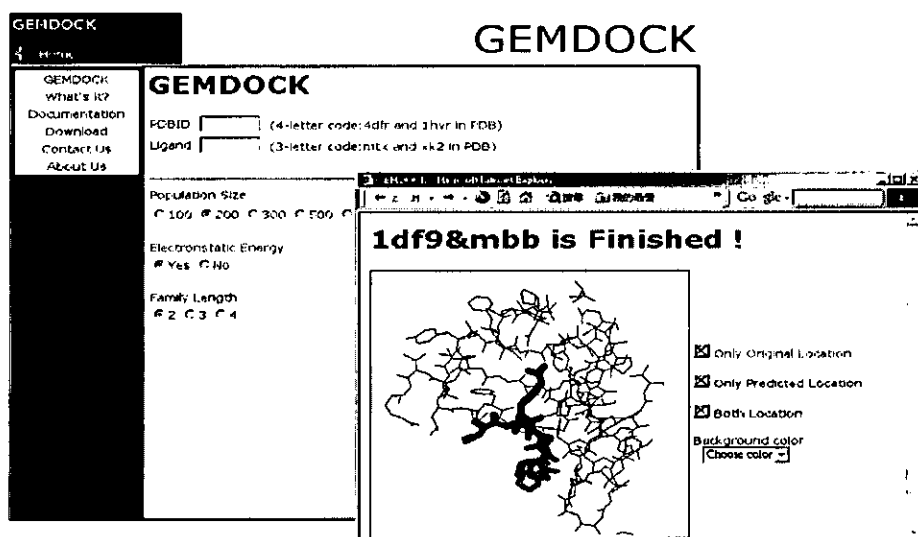
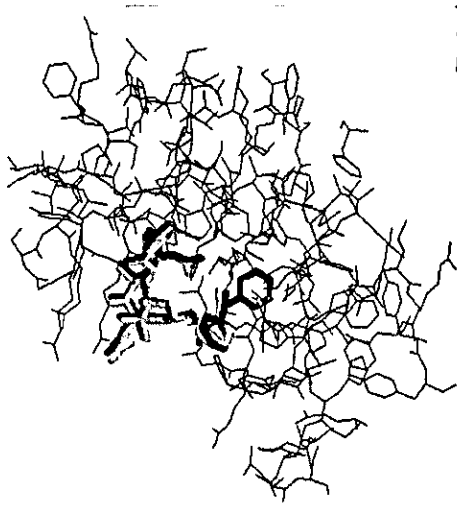


Figure 4-1 : GEMDOCK web based 遠端服務系統圖示

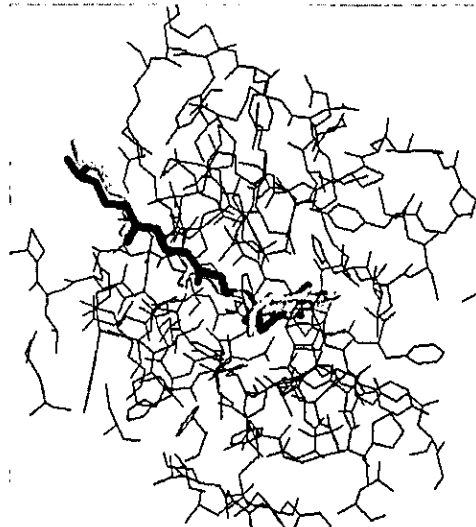
Figure 4-2：快速藥物篩選之初期成果: (a)、(b): GEMDOCK 之預測結果，紅色為預測 Ligand 之結構而白色為原 ligand 之 X-ray 結構；(c)、(d)為 GEMDOCK 針對 112 個 protein-ligand docking 之預測統計結果。



(a) 1jff-TAI

1jff: refine structure of alpha-beta tubulin from zinc-induced.

TAI: avoiding tubulin separation

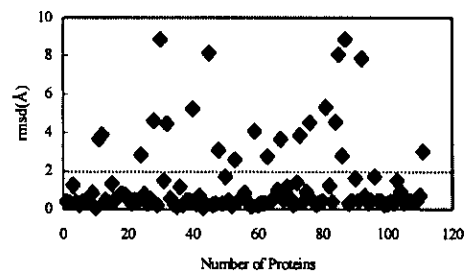


(b) 1fe7/VIT

1fe7: first structural evidence of anti-inflammatory action of vitamin

VIT: Vitamin E

rmsd value between X-ray and predicated structures of Ligand	# of Ligand	Range of # of atoms	Range of # of single bonds
$\leq 0.5 \text{ \AA}$	59	9~62	0~32
$\leq 1.0 \text{ \AA}$ and $> 0.5 \text{ \AA}$	20	9~53	0~22
$\leq 1.5 \text{ \AA}$ and $> 1.0 \text{ \AA}$	8	6~41	0~24
$\leq 2.0 \text{ \AA}$ and $> 1.5 \text{ \AA}$	3	12~44	0~17
$\leq 2.5 \text{ \AA}$ and $> 2.0 \text{ \AA}$	0	0	0
$\leq 3.0 \text{ \AA}$ and $> 2.5 \text{ \AA}$	5	9~60	0~16
$> 3.0 \text{ \AA}$	17	11~58	0~30
total	112		



(d): 將(c)之預測結果依 RMSD 繪出，112 個預測結果中有 90 個 $\leq 2.0 \text{ \AA}$

(c): 預測結果依 RMSD 做分類，112 個預測結果中有 90 個 $\leq 2.0 \text{ \AA}$

4.2 GEMDOCK 應用於篩選藥物資料庫

將 GEMDOCK 應用於藥物資料庫的篩選上，以 TK(thymidine kinase)為標的蛋白(PDB code: 1KIM)[7, 8]，已知 PDB 中與 TK 共同結晶的 ligand 有 10 個(Figure 4-3)，再由藥物資料庫中隨機挑選出 1,045 個化合物(分子量 200-750 間)，以 GEMDOCK 預測這 1,055 個化合物與標的蛋白結合的位置，並計算與標的蛋白結合的能量，以此能量由低到高排序，能量越低則表示此化合物與蛋白質間的結合越穩定，如 Figure 4-4。結果在排名前 5% 的化合物中，已知的 10 個 PDB ligand 出現了 9 個。將前 5% 的化合物(共 50 個)區分為三群：1. real ligand (9 個)，2. purine analog (26 個)，3. pyrimidine analog (15 個)，這三群均為核苷酸的相似物，可見 GEMDOCK 能有效地自 1,055 個化合物中篩選出真正會與 TK 結合的小分子，及與其結構相似的其他化合物。

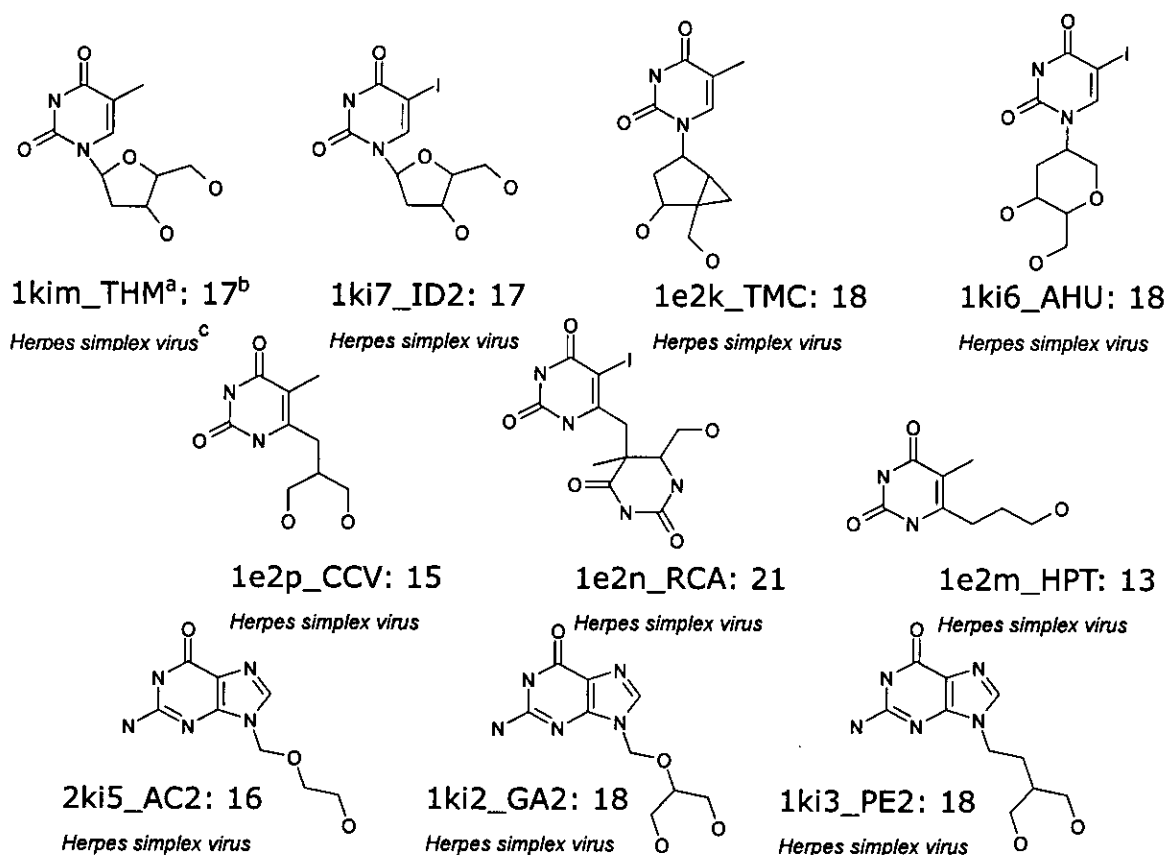


Figure 4-3 : PDB 中與 thymidine kinase 共同結晶的 10 個 ligand

- PDB code_ligand name
- Atom number
- source

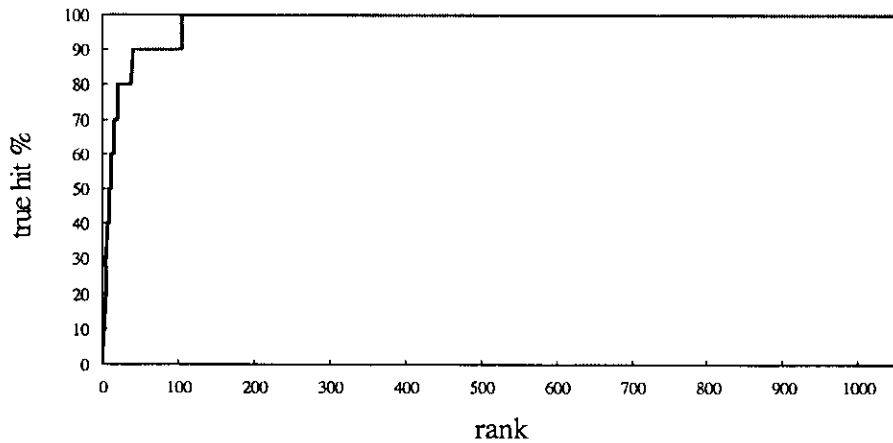


Figure 4-4：依 GEMDOCK 計算出的能量做排序，結果在前 100 名內即找到所有已知會與 TK 結合的 ligand。

4.3 GEMDOCK 應用於 DenII NS3 serine protease 及 SARS coronavirus 3C-like protease

將 GEMDOCK 實際用於發展登革病毒蛋白水解酶 (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 Table 4-1 及 Figure 4-5)，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。Table 4-1 為以 substrate 的片段(序列: KRSW)做為模擬的化合物做嵌合預測的結果。

Table 4-1：登革病毒蛋白水解酶活性區域的重要胺基酸與受質間的作用關係

Site	Substrate (1BEF)	Specific residues of 1BEF
P2	LYS	H51, D75, G151(H), N152(H), G153(H)
P1	ARG	L115(vW), S131, S135(H), G136, Y150, S163
P1'	SER	Q35(H), I36(H), H51, V52, S135
P2'	TRP	Q35(H), I36(H), P132, G133(H)

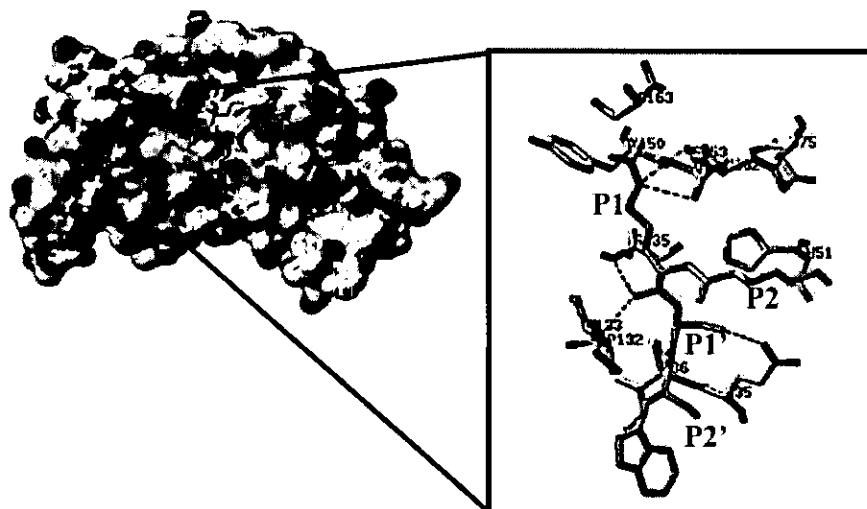


Figure 4-5：登革病毒 NS3 蛋白水解酶與受質之交互作用示意圖

4.4 CMC 藥物資料庫的篩選結果

CMC 資料庫全名為 Comprehensive Medicinal Chemistry 資料庫，存放具有生理活性之藥物分子結構與活性數值，目前資料庫內共收集 7,937 個藥物分子資料。我們採用分子量 200-750 之間的藥物分子(共 5,742 個)作為 virtual screening 的標的分子。由 5,742 個分子中取 virtual screening 分數在前 200 名的結構作更進一步之人工分析與分類，希望能夠藉此發現適合與 DenII NS3 Protease 結合之潛在藥物分子，作為進一步實驗分析的依據。



Figure 4-8 : NS3 protease 結合區表面電性分佈圖，紅色為負電性，藍色為正電性。

DenII NS3 protease 在 binding site 部分之蛋白質表面具有大量負電性區域(如 Figure 4-8 所示，圖中分子為編號 MFCD00006589 之化合物用以標示 Substrate recognition site 1 位置。中央凹槽部分即為 DenII NS3 protease 之 binding site)，此外在 Substrate recognition site 1 處為一穿透蛋白質分子表面之隧道，而非一般認為之洞穴型構造。此種蛋白質特性會相當程度地影響 virtual screening 之排名與結果。



Figure 4-9：NS3 protease 重要位置之胺基酸的凡得瓦表面分色標示 Catalytic triad(黃色, Ser165)與 Substrate recognition site 1(綠色, Ser131、Tyr150)之位置。

我們對於 CMC 資料庫對 Dengue Virus NS3 Protease 之 virtual screening 前 200 名之分子分類階層與標準如下所示：

5. 分子與蛋白質嵌合之位置
6. 和蛋白質嵌合之結合作用區
7. 分子基本骨架結構

首先我們先將 docking 結果分為在蛋白質結合區與不在蛋白質結合區兩大類，藉此找出真正與目標區域有交互作用之分子。嗣後再以此分子是位於 Substrate recognition site 1(S1)、Catalytic triad 與 Substrate recognition site 1(S1)或 polypeptide binding site(S2-Sn)三大區域中之何者，做進一步的交互作用力與分子本身大小的分群，希望能夠分析出分子群的共通性，找出能夠產生潛在抑制劑可能之分子群及其共通特性。最後再以分子本身的基本骨架結構做更細部的分群，藉此區分出最能夠和目標產生良好作用的基本骨架作為以後 lead compound 修飾之依據。

Figure 4-10 即為前兩百名之分子分群樹狀圖。圖上之分類階層如上所述，最後一層為分子骨架分群，每一個結構為群代表之結構。S1 pocket 群，是屬於在 Substrate recognition site 1(以下簡稱 S1)者，為位於 S1 但未與 catalytic triad 有交互作用的 docking 結果。此一分群裡的分子又以骨架分為 7 類。歸類於 Catalytic triad 與 S1 群者屬於和 S1 中的胺基酸與 catalytic triad(主要以 Ser135 為準)有交互作用者，此一群共分為 11 類，前 6 類的基本骨架與 S1 群的相同，不同處在於構成化合物骨架的原子特性不同。最後一群則是與整個 binding site 有交互作用者，此一群只有一群，包含六個不同分子。此群的共通處在於均具有長鏈碳鏈，利用長碳鏈與 binding site 的凹槽產生凡得瓦作用力而相互結合。在此三大群中以第二群(Catalytic triad 與 S1)最能夠產生符合 serine protease 抑制劑設計的理想交互作用力與構形。

S1 pocket 群，此群內包含 7 類，如 Figure 4-11 所示。第一類以 METHYSERGIDE (MCMC00001002) 為代表。第二類以 PICLOXYDINE (MCMC00002335) 為代表。第三

類以 NILEPROST (MCMC00004953) 為代表，第四類以 CELIPROLOL (MCMC00004392) 為代表，第五類以 DIETHYLSTILBESTROL DIPHOSPHATE (MCMC00001235) 為代表，前五類都是將環的部分 docking 到 S1 的深處，分支的長鏈伸出和 S1 部位重要的胺基酸側鏈做交互作用。這樣的結果和 S1 區域的特性有著密切的關係。第六類以 DEOXYSPERGUALIN (MCMC00005734) 為代表，此一類完全由可動性高的長鏈構成，因此可以穿過 S1 pocket 到另外一邊的表面。

S1 and catalytic triad 群共有 11 類，如 Figure 4-12 所示。前六類與 S1 pocket 群狀況大致相同，不同點在於分支的長鏈上具有可以和 Ser135 形成氫鍵或交互作用的原子。此六類的代表為，FOSQUIDONE (MCMC00006071)、CYNARINE (MCMC00001729)、CEFAZAFUR (MCMC00004475)、SAPERCONAZOLE (MCMC00005779)、GLUCOSULFAMIDE (MCMC00006589)、NISBUTEROL (MCMC00004574)，它們都能夠和 S1 部位的胺基酸形成足夠的穩定力量(以氫鍵或靜電力)，同時和負責催化剪斷 peptide bond 的 Ser135 產生交互作用力或空間上的阻擋，干擾催化的發生。CLINPROST (MCMC00006282)、CEFTEZOLE (MCMC00003420)、TRIMETREXATE (MCMC00004094)、LAURETH 10S (MCMC00002590) 這四類是屬於和 S1 群不同的結構，但是它們 docking 的行為仍然是以環的部分深埋在 S1 部位裡，以極性長鏈分支和 binding site 裡的胺基酸產生氫鍵，構成良好的結合。這樣的結合方式有助於外來分子穩固的停留在 binding site 的特定位置，並且干擾到 catalytic triad 的催化機制運作，讓這一大群的分子較有機會成為潛在抑制劑。尤其以 CEFAZAFUR (MCMC00004475)、GLUCOSULFAMIDE (MCMC00006589)、CEFTEZOLE (MCMC00003420)、TRIMETREXATE (MCMC00004094) 這四類的鍵結方式最為理想。LAURETH 10S (MCMC00002590) 這一類由於是可動性大的長鏈結構，因此產生鍵結的方式與數量變動性較大，所以較不適合作為潛在抑制劑的設計依據。

Whole binding site 群，此群內只有一類，六種分子，均為長鏈狀分子結構，富可動性，這一類的分子均以長碳鏈和 binding site 結合，如 Figure 4-13。但是這樣的結合方式並不能提供穩定而強力的結合作用力，因此此類分子結構並不適合作為我們設計潛在抑制劑的方向。

在分類樹上的另一分支為不在 binding site 的分子，此一分支因為在 docking 位置上均位於 S1 口袋的另外一側，並沒有跟 binding site 的任何重要胺基酸產生作用力，暫不列入分析考量之中。

由 Serine protease 抑制劑的設計概念觀之，良好的抑制劑必須和 binding site 有足夠的結合與辨識能力。此外還必須能夠和 Serine 之 OH 基團產生交互作用，阻斷催化發生。因此在設計抑制劑時首要便是找出有可能和 binding site 的專一性控制區產生影響的分子結構，而後再以此一分子結構作為修飾的基礎，發展新的抑制劑。在我們 docking 的結果分析裡，的確找到具有此種能力之分子與其結構共通特性。這樣的分類與分群結果將可提供後續實驗工作努力的方向。

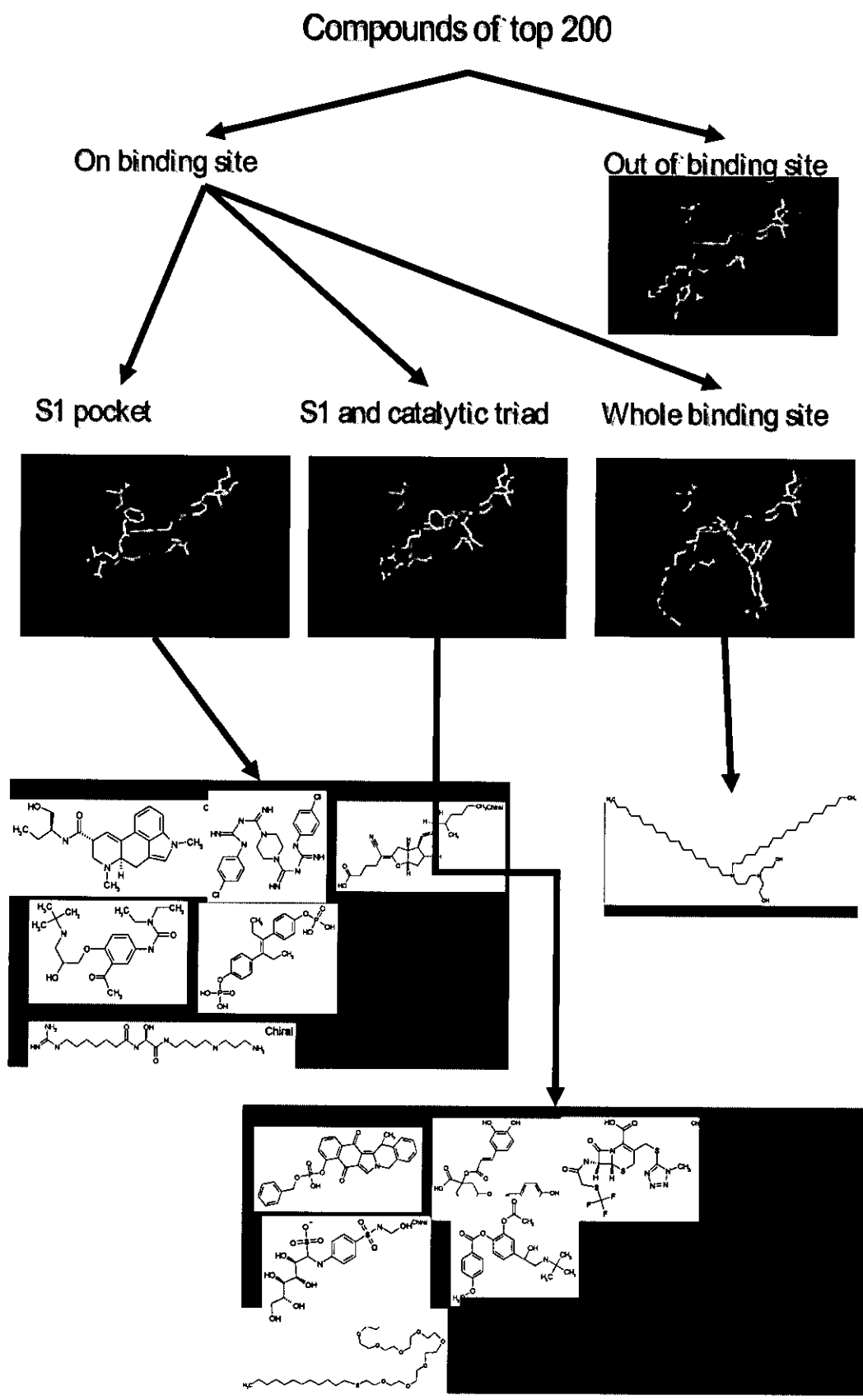
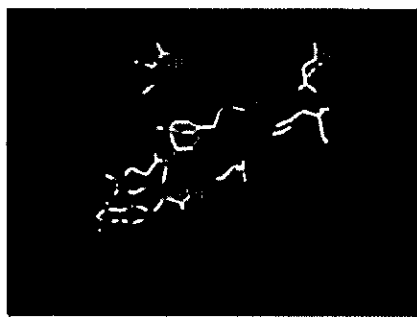
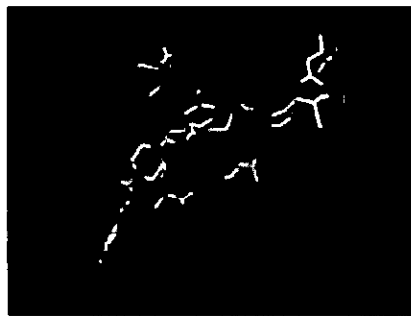


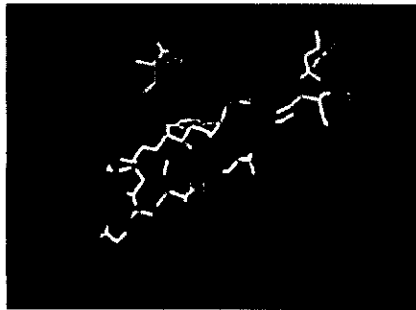
Figure 4-10 : 前兩百名之分子分群樹



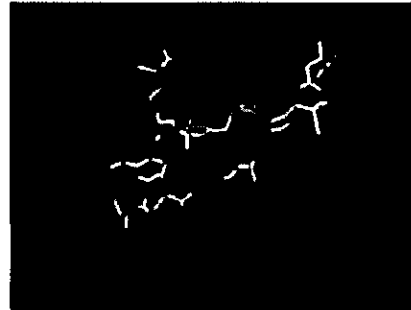
MCMC00001002



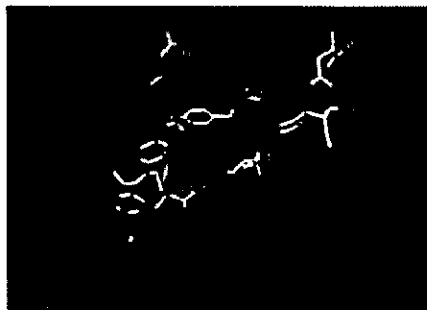
MCMC00002335



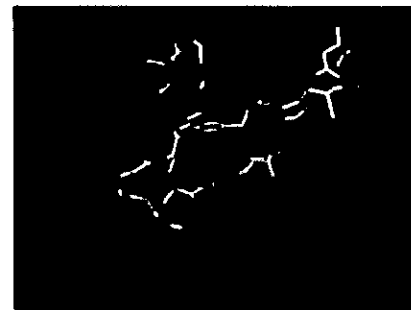
MCMC00004953



MCMC00004392



MCMC00001235



MCMC00005734

Figure 4-11: S1 pocket 群，此群內包含 7 類(第 7 類為其他，不在圖中)，圖中所示為每群之代表與結合區交互作用圖。藍色為蛋白質胺基酸，黃色為 catalytic triad，綠色為 S1 pocket 重要胺基酸，CPK 標示者為 docking 結果。

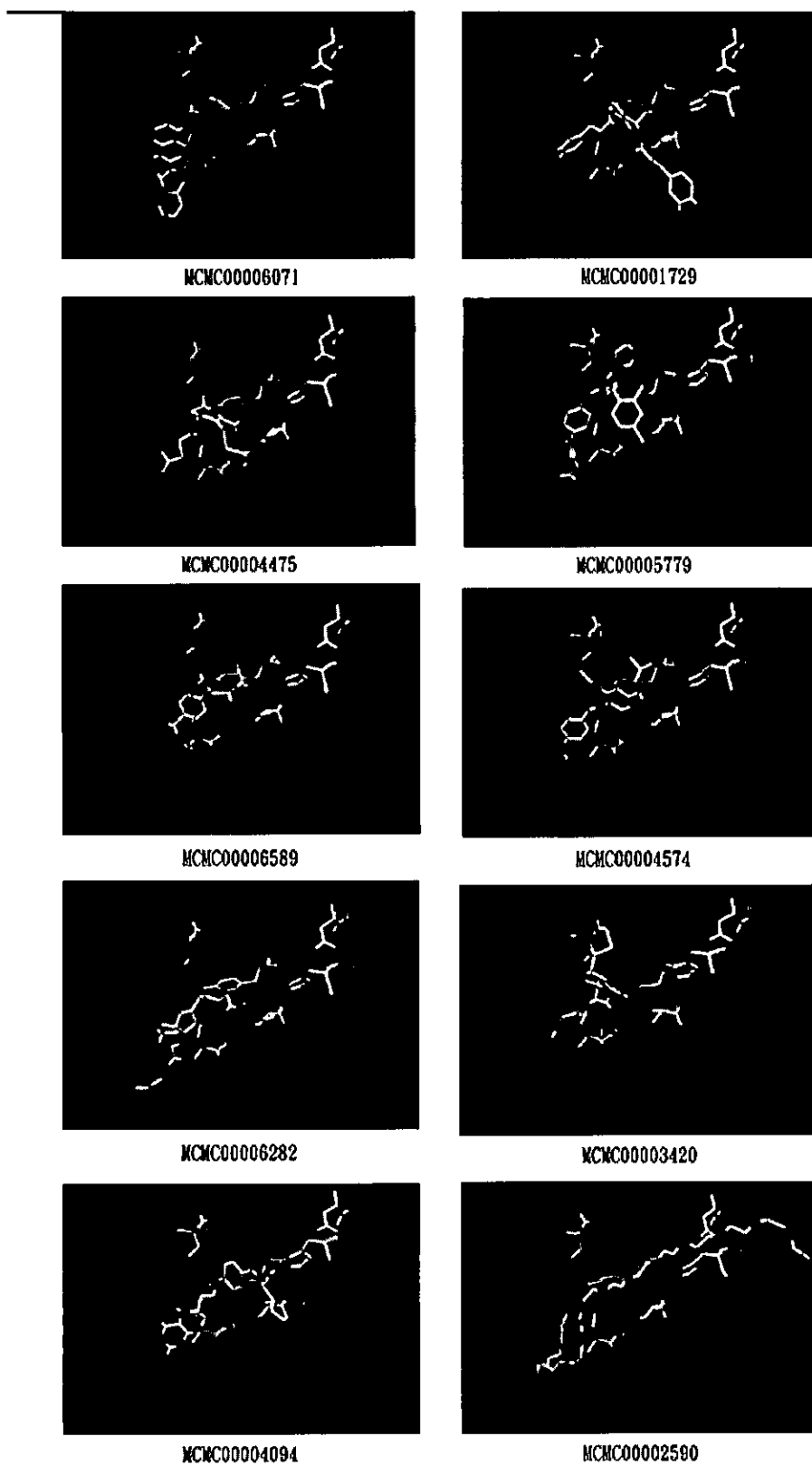


Figure 4-12 S1 and catalytic triad 群共有 11 類(第 11 類為其他，不在圖中)，圖中所示為每群之代表與結合區交互作用圖。藍色為蛋白質胺基酸，黃色為 catalytic triad，綠色為 S1 pocket 重要胺基酸，CPK 標示者為 docking 結果。

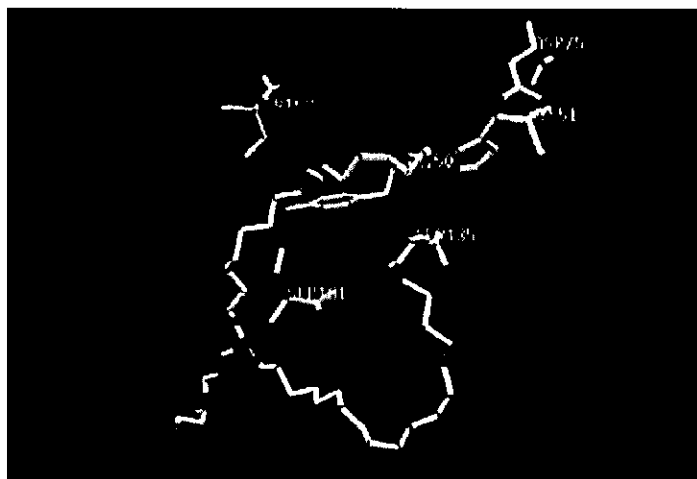
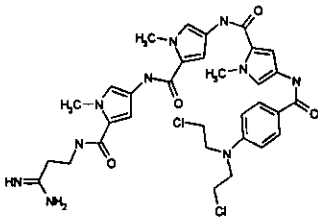
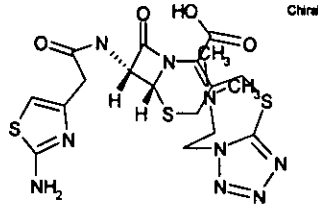
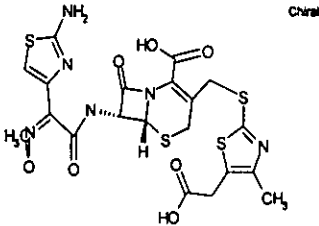
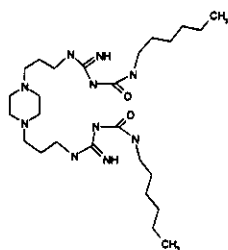


Figure 4-13 Whole binding site 群，此群內只有一類，藍色為蛋白質胺基酸，黃色為 catalytic triad，綠色為 S1 pocket 重要胺基酸，CPK 標示者為 docking 結果。

前 200 名化合物一覽表(不包括 docking 在活性區域外的化合物)

Compound name	Structure	MDL number	Molecular weight	Drug class
TALLIMUSTINE		MCMC00006385	697.63	Antineoplastic
CEFOTIAM	 Chiral	MCMC00004603	525.637	Antibacterial
CEFODIZIME	 Chiral	MCMC00004909	584.677	Antibacterial

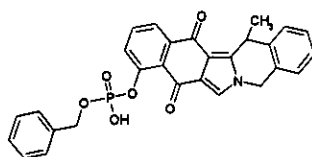
IPEXIDINE



MCMC00004882

538.786 Dental caries prophylactic

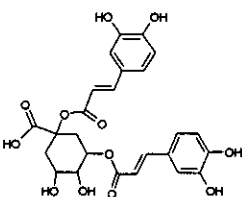
FOSQUIDONE



MCMC00006071

499.465 Antineoplastic

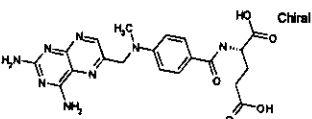
CYNARINE



MCMC00001729

516.463 Choleric

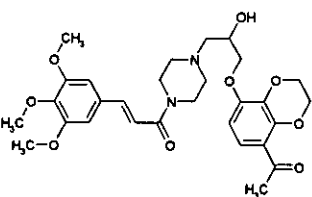
METHOTREXATE



MCMC00006960

454.449 Antineoplastic

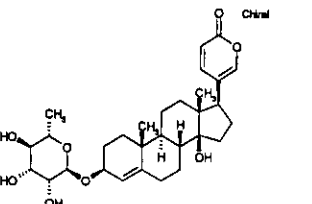
CINEPAXADIL



MCMC00004885

556.618

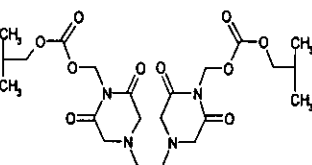
PROSCILLARIDIN



MCMC00007049

530.664 Cardiotonic

SOBUZOXANE



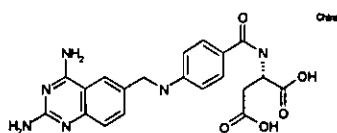
MCMC00005987

514.537 Antineoplastic

QUINASPAR

MCMC00009899

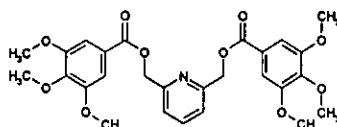
424.415 Antineoplastic



PIROZADIL

MCMC00004223

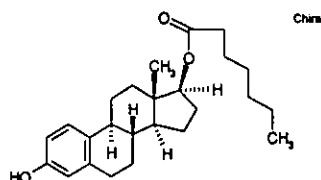
527.533 Antihyperlipoproteinemic



ESTRADIOL ENANTHATE

MCMC00002210

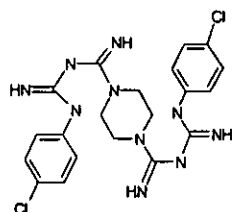
384.556 Estrogen



PICLOXYDINE

MCMC00002335

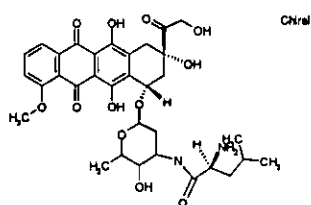
475.387 Disinfectant



LEURUBICIN

MCMC00006250

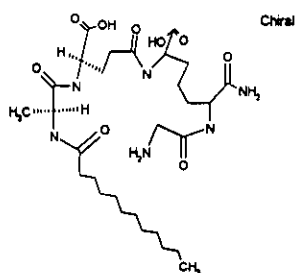
656.693 Antileukemic



PIMELAUTIDE

MCMC00005207

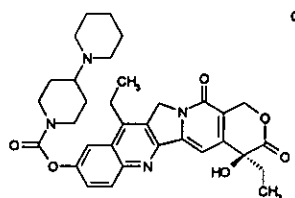
628.763 Immunostimulant



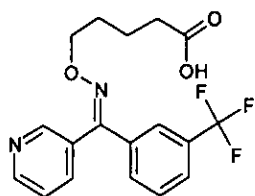
IRINOTECAN

MCMC00006298

586.694 Antineoplastic



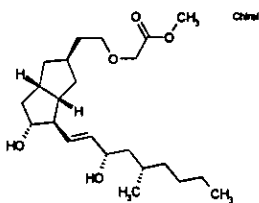
RIDOGREL



MCMC00006042

366.343 Antithrombotic

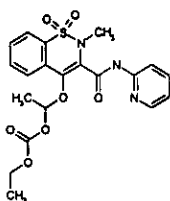
PIMILPROST



MCMC00007525

396.572 Peripheral vascular disease (therapeutic)

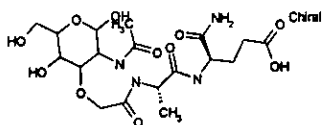
AMPIROXICAM



MCMC00005685

447.47 Antiinflammatory

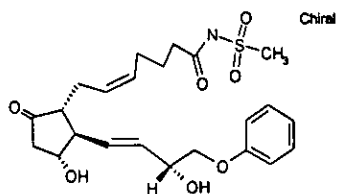
ALMURTIDE



MCMC00007785

478.452

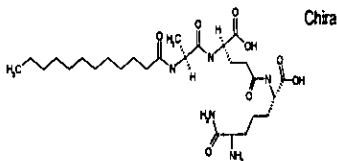
SULPROSTONE



MCMC00004555

465.564 Abortifacient

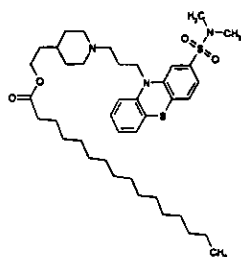
TABILAUTIDE



MCMC00006258

571.72 Immunogenic

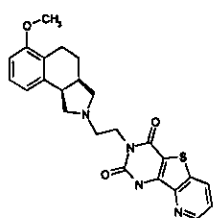
PIPOTIAZINE PALMITATE



MCMC00003807

714.0941 Antipsychotic

ALPHA-1-A
ADRENORECEPTOR
ANTAGONIST

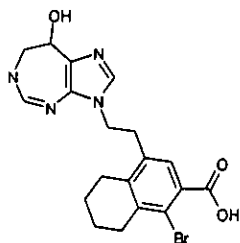


Chiral

MCMC00008029

448.545 Antiadrenergic

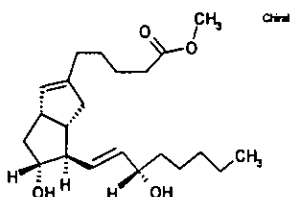
COFORMYCIN
AGLYCON
DERIVATIVE-3



MCMC00008060

433.304 AMPDA
antagonist

CLINPROST

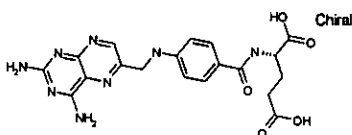


Chiral

MCMC00006282

364.53 Antithrombotic

AMINOPTERIN

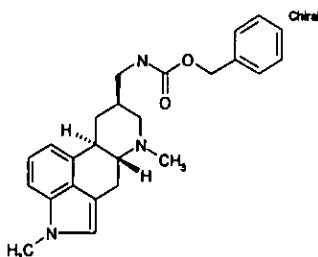


Chiral

MCMC00006938

440.422 Enzyme
inhibitor

METERGOLINE

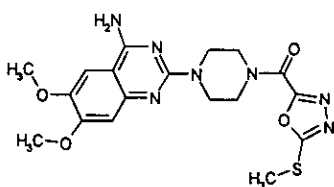


Chiral

MCMC00007176

403.529 Analgesic

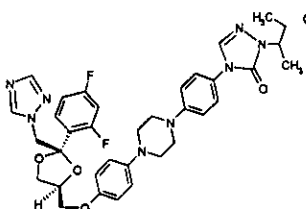
TIODAZOSIN



MCMC00004816

431.477 Antihyperten-
sive

SAPERCONAZOLE



Chiral

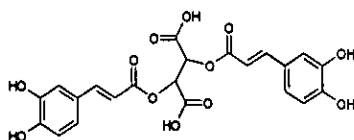
MCMC00005779

672.733 Antifungal

CICHORICACID

MCMC00007713

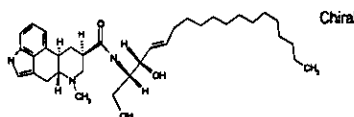
474.372 Antiasthmatic



DOSERGOSIDE

MCMC00005494

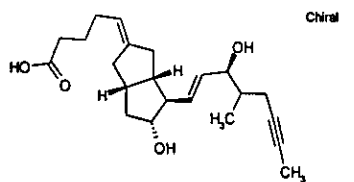
551.811



ILOPROST

MCMC00005048

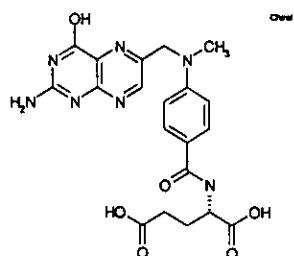
360.491 Antihypertensive



METHOPTERINE

MCMC00009906

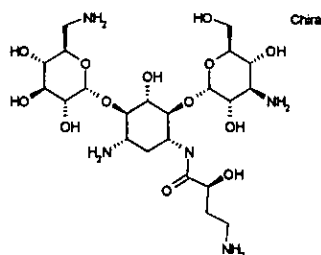
455.429 Antimetabolite



AMIKACIN

MCMC00007221

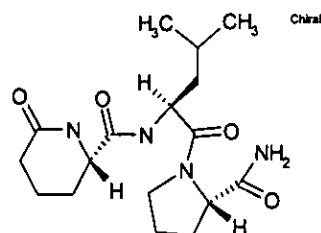
585.604 Antibacterial



POSATIRELIN

MCMC00005901

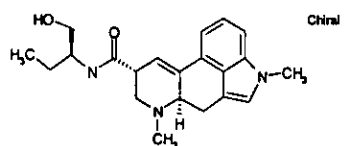
352.432 Anticatalytic



METHYSERGIDE

MCMC00001002

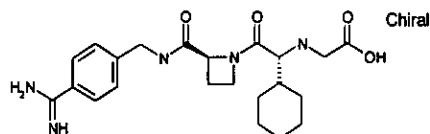
353.463 Antimigraine



MELAGATRAN

MCMC00007753

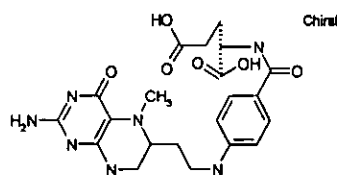
429.518 Anticoagulant



KETOTREXATE

MCMC00004097

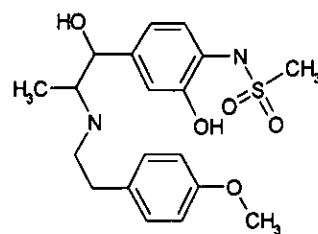
473.487 Antineoplastic



MESUPRINE

MCMC00002547

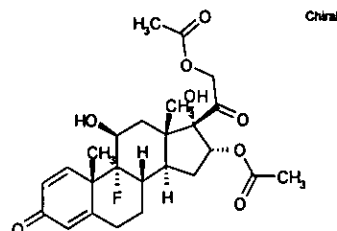
394.493 Muscle relaxant



TRIAMCINOLONE
DIACETATE

MCMC00006975

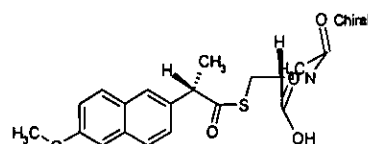
478.519 Glucocorticoid



CINAPROXEN

MCMC00005541

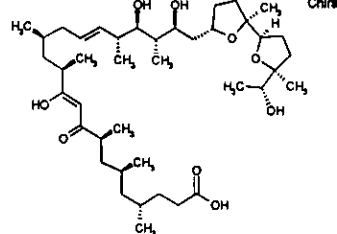
375.443



IONOMYCIN

MCMC00010295

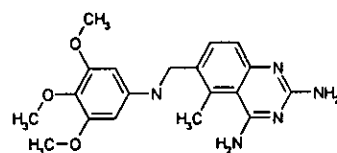
709.011 Antibiotic



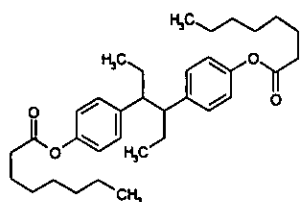
TRIMETREXATE

MCMC00004094

369.427 Antineoplastic



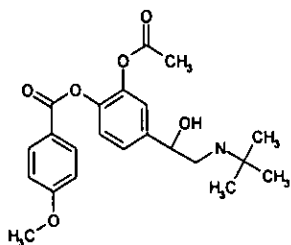
HEXESTROL
DICAPRYLATE



MCMC00010262

522.765 Synthetic
estrogen

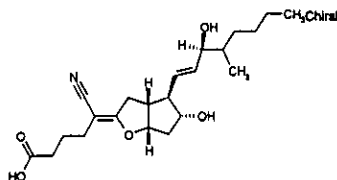
NISBUTEROL



MCMC00004574

401.464 Bronchodilator

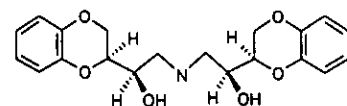
NILEPROST



MCMC00004953

391.505 Antihyperten-
sive

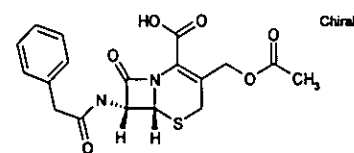
BENDACALOL



MCMC00005911

373.403 Antihyperten-
sive

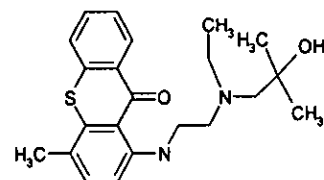
CEFALORAM



MCMC00001497

390.414 Antibacterial

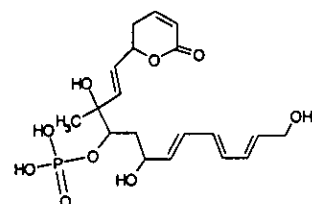
BECANTHONE



MCMC00002852

384.545 Antischistosomal

FOSTRIECIN



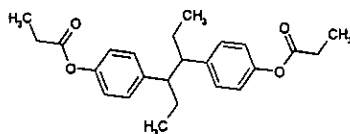
MCMC00005511

430.395 Antineoplastic

HEXESTROL
DIPROPIONATE

MCMC00010123

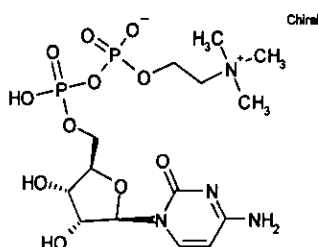
382.497 Synthetic
estrogen



CITICOLINE

MCMC00001543

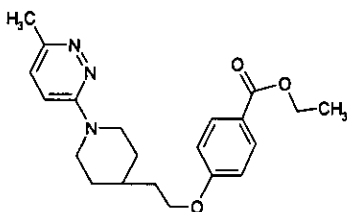
488.324 Cerebral
stimulant



PIRODAVIR

MCMC00006125

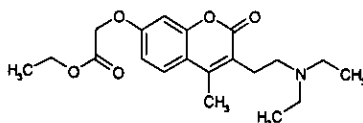
369.468 Antiviral



CHROMONAR

MCMC00001470

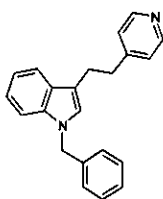
361.442 Coronary
vasodilator



BENZINDOPYRINE

MCMC00002963

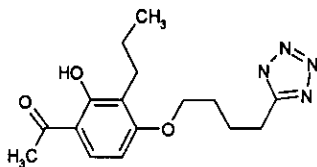
312.418 Antipsychotic



TOMELUKAST

MCMC00005519

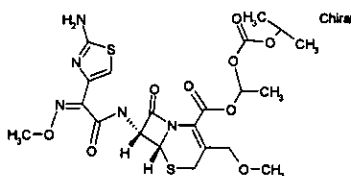
318.379 Antiasthmatic



CEFPODOXIME
PROXETIL

MCMC00005496

557.602 Antibacterial

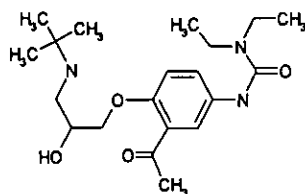


LEFRADAFIBAN

MCMC00007873

439.465 Fibrinogen antagonist

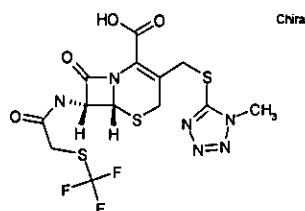
CELIPROLOL



MCMC00004392

379.504 Antiadrenergic

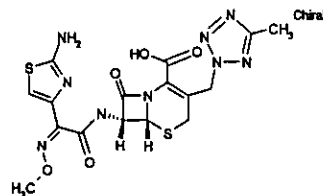
CEFAZAFLUR



MCMC00004475

470.476 Antibacterial

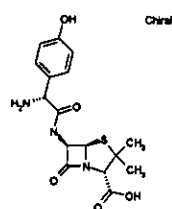
CEFTERAM



MCMC00005348

479.5 Antibacterial

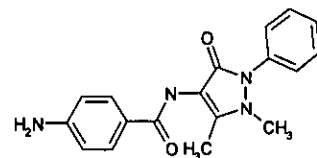
AMOXICILLIN



MCMC00007203

365.411 Antibacterial

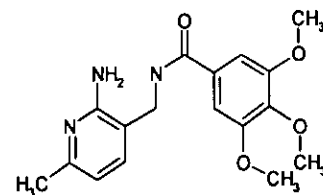
AMINO BENZAMIDO
PHENAZONE



MCMC00009775

322.366 Analgesic

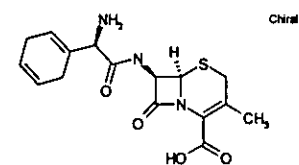
TRIMETHAMIDE



MCMC00002364

331.375

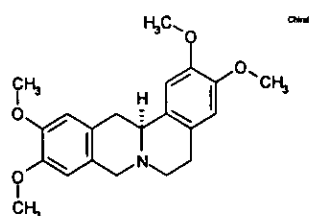
CEPHRADINE



MCMC00007223

349.412 Antibacterial

NORCORALYDINE

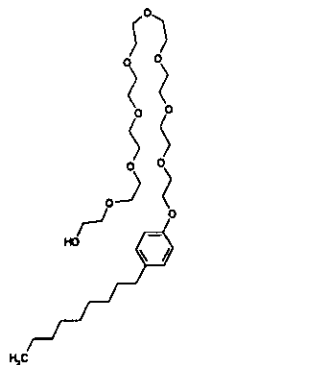


OH

MCMC00009996

355.431 Protoberberine alkaloid

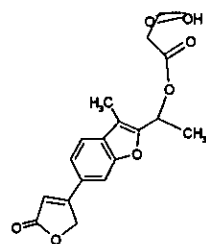
NONOXYNOL 9



MCMC00003375

616.84 Spermicide

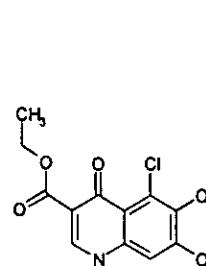
BENFURODIL HEMISUCCINATE



MCMC00001964

358.351 Cardiotonic

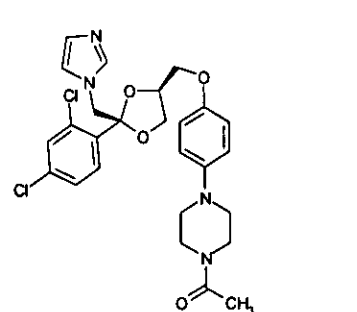
QUINCARBATE



MCMC00004234

367.789 Diuretic

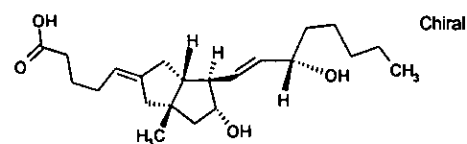
KETOCONAZOLE



MCMC00004734

531.4371 Antifungal

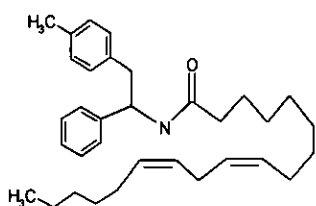
CIPROSTONE



MCMC00005326

364.522 Anti-thrombotic

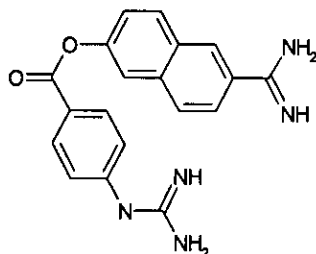
MOCTAMIDE



MCMC00003518

473.749 Anti-hyperlipoproteinemic

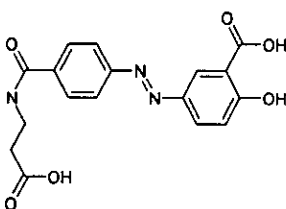
NAFAMOSTAT



MCMC00005316

347.38 Anti-inflammatory

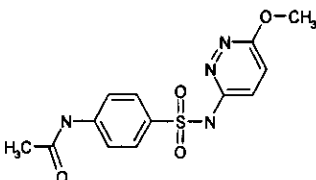
BALSALAZIDE



MCMC00005278

357.326 Anti-ulcerative colitis

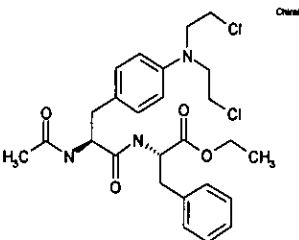
SULFAMETHOXYPYRIDAZINE ACETYL



MCMC00001992

322.345 Antibacterial

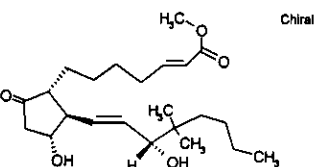
ASAFAN



MCMC00010156

522.47 Antineoplastic

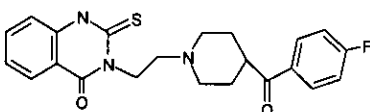
GEMEPROST



MCMC00004697

394.548 Prostaglandin

ALTANSERIN

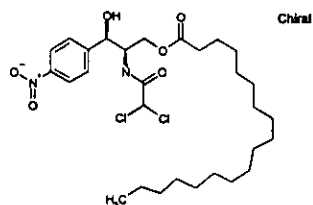


MCMC00005124

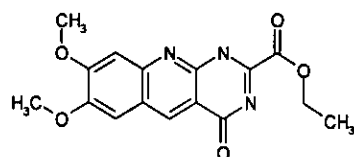
411.502 Serotonin antagonist

DEOXYSPERGUALIN
MCMC00005734 387.525 Antineoplastic

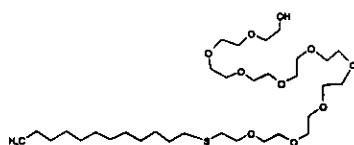
CHLORAMPHENICOL STEARATE
MCMC00010216 589.596 Antibiotic



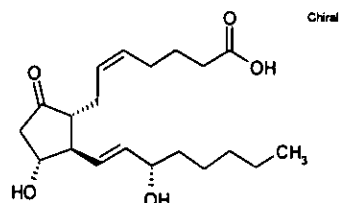
PIROLATE
MCMC00004275 329.315 Antiasthmatic



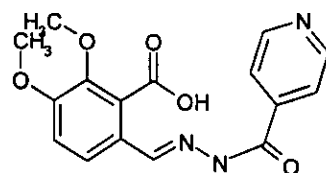
LAURETH 10S
MCMC00002590 642.941 Spermicide



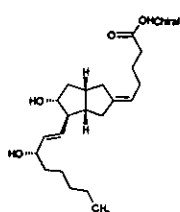
DINOPROSTONE
MCMC00001007 352.468 Oxytocic



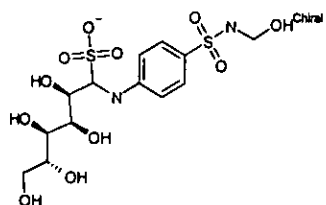
OPINIAZIDE
MCMC00001880 329.315 Antibacterial



CARBACYCLIN
MCMC00010028 350.496 Antithrombotic



GLUCOSULFAMIDE
MCMC00006589 445.447



4.5 280 個蛋白水解酶抑制劑的篩選結果

PDB 中有許多 protease 與其抑制劑共同結晶的結構(protein-ligand complex)，這些抑制劑多半是藥物發展過程中的前導藥物，若再經過一些最佳化的動作，即有可能成為正式的商品化藥物。如 aspartate protease 這一類的結晶結構，就有許多包含有目前已用於臨床治療的抗 HIV 藥物，主要的作用即在抑制 HIV aspartate protease 的活性。這一類抑制劑大部分都屬於 peptide drug，模擬原本受質主鏈(backbone)的結構，與活性區域主鏈形成氫鍵，以固定化合物的位置，再以特定的官能基擋住原本會與受質作用的 pocket，以達到抑制的效果。

DenII NS3 serine protease 亦屬蛋白酶，因此我們假設 protease inhibitor 較有可能成為前導藥物。在這個假設下，由 PDB 中已知的 protease inhibitor 挑選出 280 個化合物結構，與登革病毒的蛋白水解酶進行 docking，分析前 100 名的結果，依 docking 的位置可分為三群：



Figure 4-14

1. 位於 binding site：13 個

如 Figure 4-14，綠色的化合物為 GEMDOCK 預測的結果，其位置擋住了 binding site 原本會與受質結合的位置。



Figure 4-15

2. 位於 S1 pocket：23 個

如 Figure 4-15，GEMDOCK 預測的結果 docking 在 S1 pocket 內，僅與 S1 pocket 內的胺基酸間有交互作用。



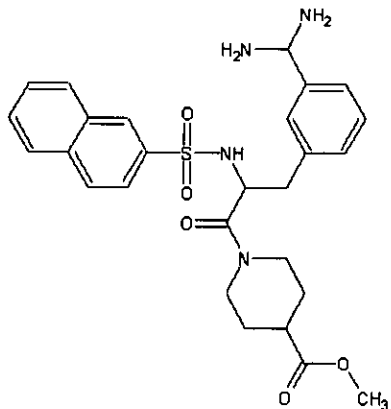
Figure 4-16

3. 位於 S1 pocket 及 catalytic site : 13 個
 如 Figure 4-16, GEMDOCK 預測的化合物
 位置除了與 S1 pocket 的重要胺基酸有交互
 作用外, 與 catalytic triad (His51, Asp75,
 Ser135)間也有交互作用。

前 100 名的結果中, 第一類及第三類較有可能成為前導藥物, 因這兩類化合物與蛋白質結合的位置恰好阻擋了受質與活性區域結合的位置, 而且這些位置在演化上有高度的保留性, 如 catalytic triad: His51、Asp75、Ser135, 是受質水解發生的位置; S1 pocket 內的重要胺基酸: Ser131、Tyr150、Ser163, 這幾個胺基酸使得 S1 pocket 具有辨識受質的專一性, 皆是維持 NS3 蛋白水解酶功能的重要胺基酸。若是針對這幾個位置設計潛在藥物, 有兩個優勢: 1. 能確實阻擋受質與蛋白質間的作用關係, 2. 因這幾個位置不易發生突變, 因此針對這些位置設計的藥物, 較不易因病毒的突變而失去效用。除了以上三類之外, 其餘的 51 個化合物 docking 的位置均位於活性區域之外, 較無成為潛在抑制劑的可能。

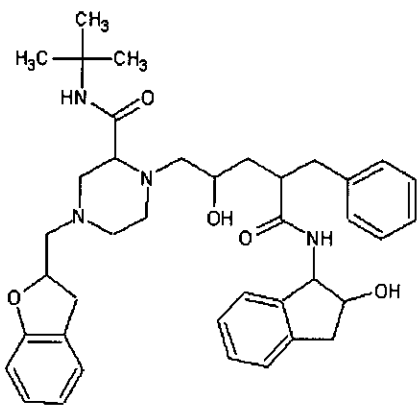
前 100 名化合物分群一覽表

Group 1: in the binding site	PDB code	Description
	1a8g_CBZ	<i>HIV-1 protease in complex with sdz283-91</i>
	11OD_PPP	<i>Crystal structure of the complex between the coagulation factor x binding protein from snake venom and the gla domain of factor</i>



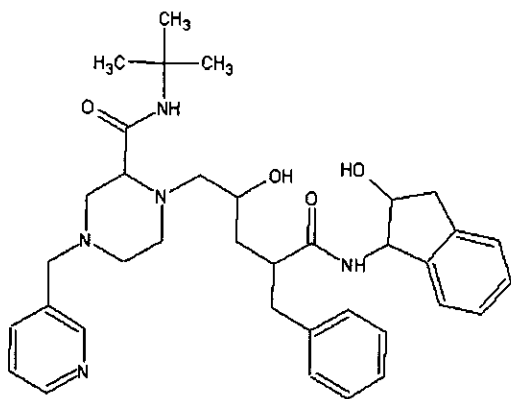
1k1j_FD2

Bovine trypsin-inhibitor complex



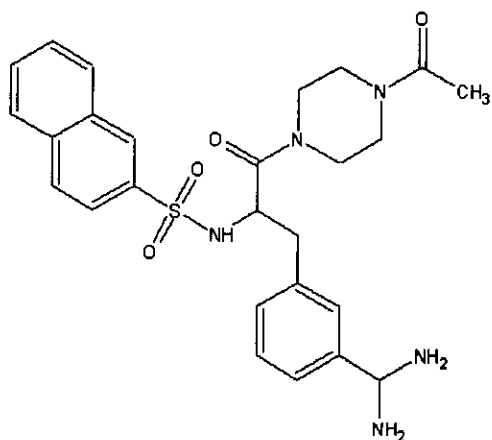
1c70_L75

Alternate binding site for the p1-p3 group of a class of potent HIV-1 protease inhibitors as a result of concerted structural change in 80's loop



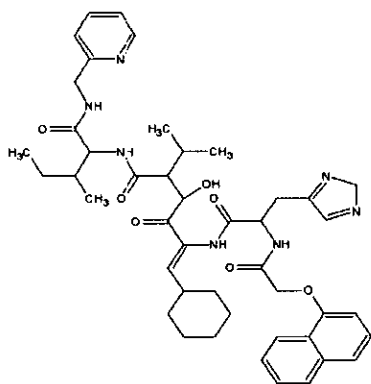
1c6y_MK1

Alternate binding site for the p1-p3 group of a class of potent HIV-1 protease inhibitors as a result of concerted structural change in 80's loop

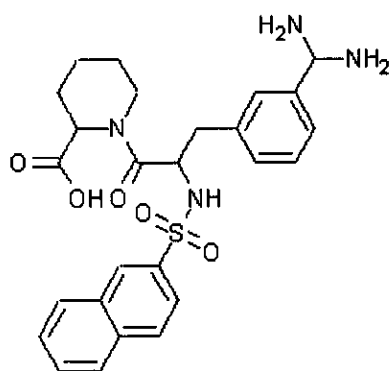


1k1m_FD4

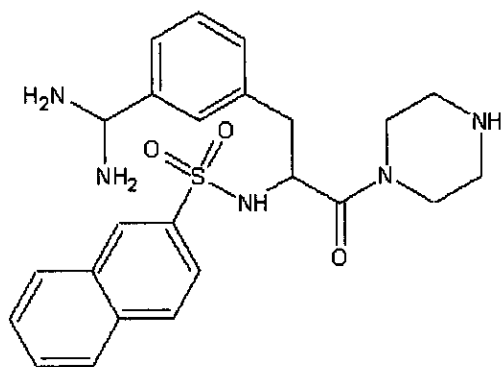
Bovine trypsin-inhibitor complex



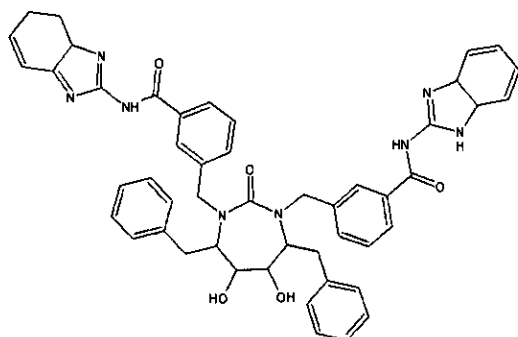
1ivp_NOA *Human immunodeficiency virus type 2 protease mutant with lys 57 replaced by leu (K57L) complex with u75875 (noa-his-cha-psi[ch(oh)ch(oh)]val-ile-apy)*



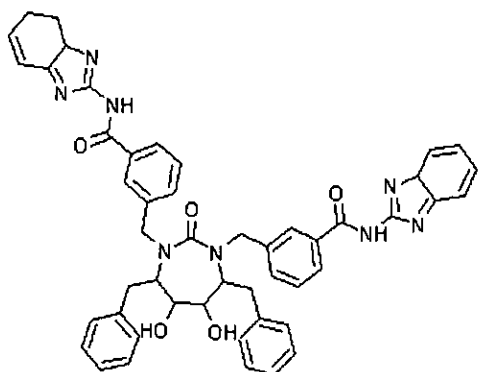
1k1i_FD1 *Bovine trypsin-inhibitor complex*



1k1l_FD3 *Bovine trypsin-inhibitor complex*

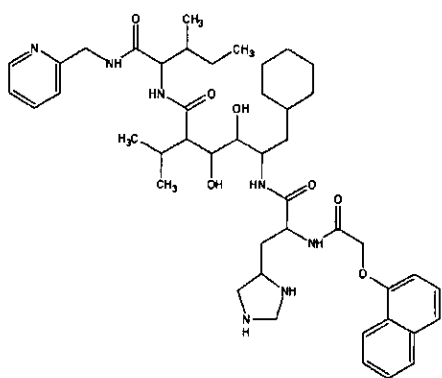


1bwb_146 *HIV-1 protease (v82f/i84v) double mutant complexed with sd146 of dupont pharmaceuticals*



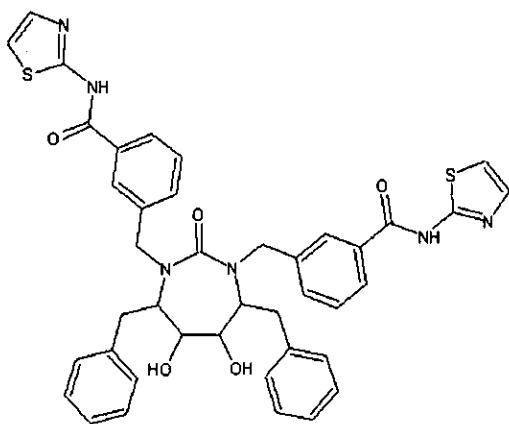
1qbt_146

HIV-1 protease inhibitors with low nanomolar potency



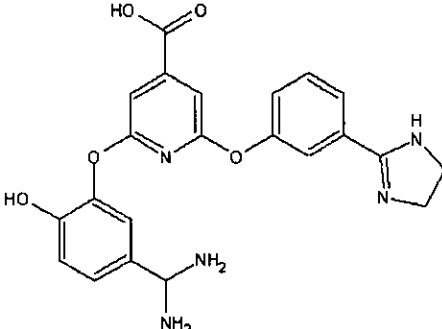
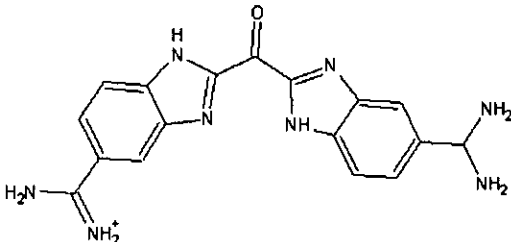
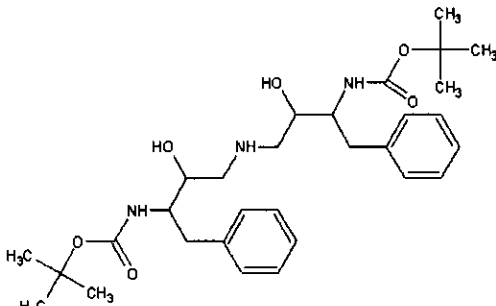
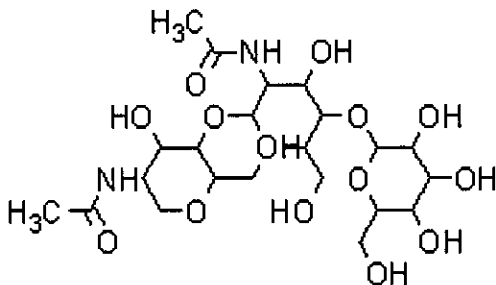
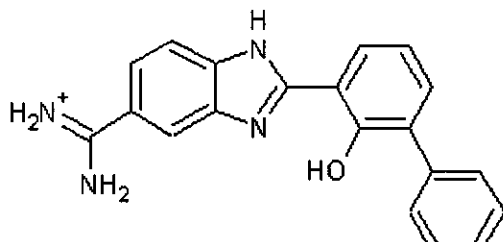
1hiv_NOA

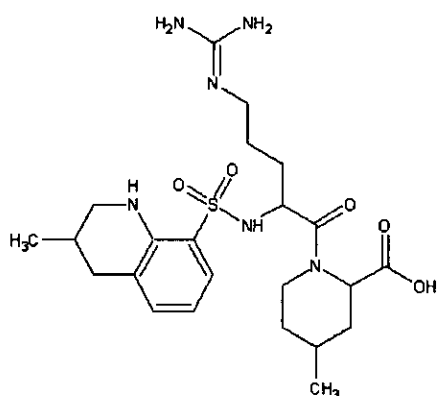
HIV-1 protease (HIV-1 pr) complex with u75875 (noa-his-cha-psi[ch(oh)ch(oh)]val-ile-apy)



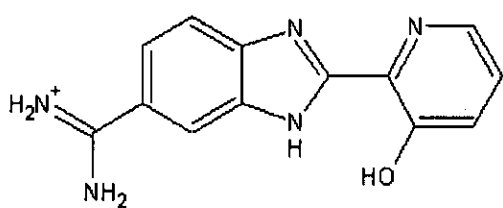
1bv7_XV6

Counteracting HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with xv638 and sd146 cyclic urea amides with broad specificities

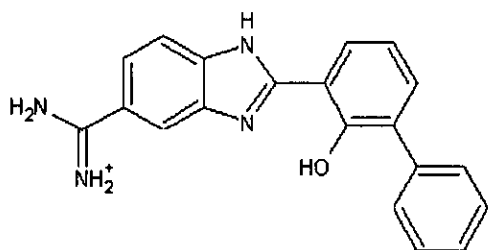
Group 2: in the S1 pocket	PDB code	Description
 <p>The structure shows a pyridine ring with a carboxylic acid group at the 4-position and a phenoxy group at the 2-position. The phenoxy group is further substituted with a 1H-imidazol-2-yl group. The pyridine ring also has a hydroxy group at the 6-position and a 2-aminoethyl group at the 3-position.</p>	1qbn_688	<i>Bovine trypsin</i> <i>2-[amino(imino)methyl]-2-hydroxyphenoxy]-6-[3-(4,5-dihydro-1h-imidazol-2-yl)phenoxy]pyridine-4-carboxylic acid (zk-806688) complex</i>
 <p>The structure features two benzimidazole rings connected by a central amide bond. Each benzimidazole ring is substituted with a 2-aminoethyl group on the benzene ring and an iminoamino group (H2N+=C-NH2) on the imidazole ring.</p>	1c2h_BAK	<i>Recruiting zinc to mediate potent, specific inhibition of serine proteases</i>
 <p>The structure is a complex molecule with multiple amide and hydroxyl groups. It includes a central chain with a hydroxyl group, a secondary amine, and a tertiary amine. There are also two tert-butyl ester groups and two phenyl rings attached to the chain.</p>	1odx_OTB	<i>HIV-1 proteinase mutant a71t, v82a</i>
 <p>The structure shows a complex oligosaccharide chain with multiple hydroxyl groups and methyl groups. It appears to be a branched structure with several sugar units linked together.</p>	1nb3_NAG	<i>Crystal structure of stefin a in complex with cathepsin h: n-terminal residues of inhibitors can adapt to the active sites of endo-and exopeptidases</i>
 <p>The structure consists of a benzimidazole ring system. One benzimidazole ring is substituted with an iminoamino group (H2N+=C-NH2) on the imidazole ring and a 2-aminoethyl group on the benzene ring. The other benzimidazole ring is substituted with a hydroxyl group and a phenyl group on the benzene ring.</p>	1gj5_130	<i>Selectivity at s1, h2o displacement, upa, tpa, ser190/ala190 protease, structure-based drug design</i>



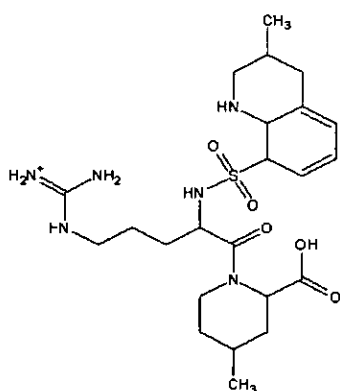
1dwc_MIT *Alpha-thrombin complex with (des-amino asp 55) hirudin (residues 55 - 65) and md-805 (argatroban)*



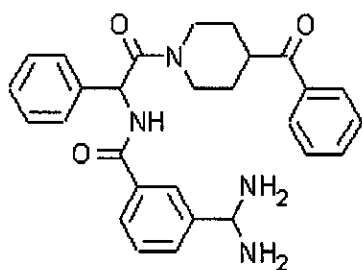
1ghy_121 *A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site*



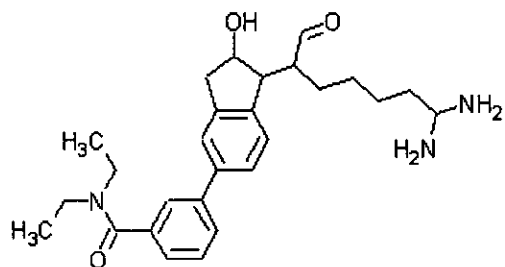
1GJB_130 *Engineering inhibitors highly selective for the S1 sites of ser190 trypsin-like serine protease drug targets*



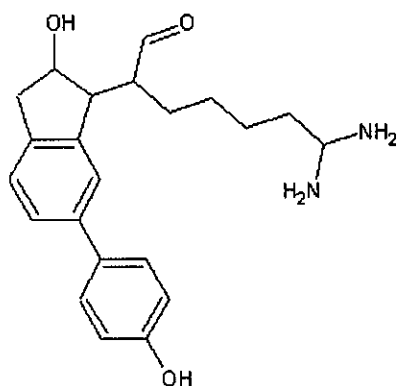
1etr_MQI *Epsilon-thrombin non-covalent complex with mqpa*



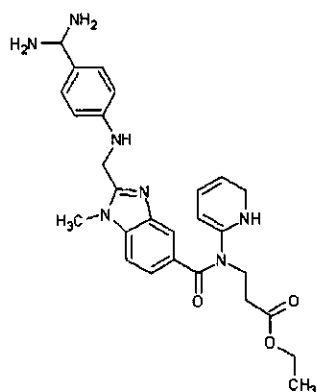
1eb2_BPO *Trypsin inhibitor complex (bpo)*



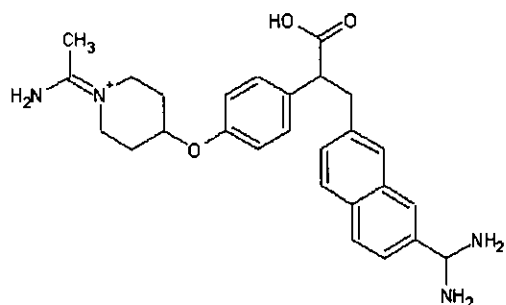
1qj7_GR1 *Novel covalent active site thrombin inhibitors*



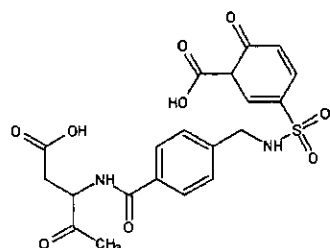
1qj1_166 *Novel covalent active site thrombin inhibitors*



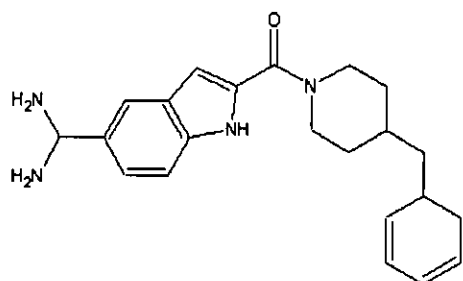
1kts_C24 *Thrombin inhibitor complex*



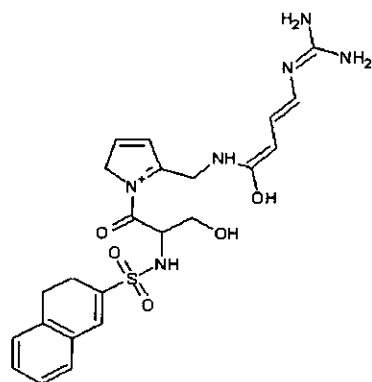
1mtv_BX3 *Factor xa specific inhibitor in complex with bovine trypsin*



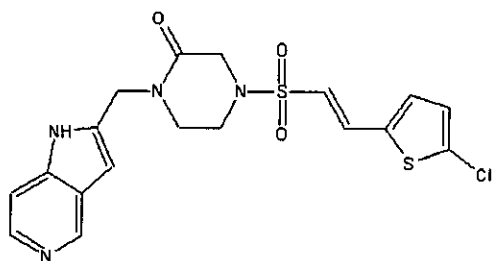
1EZQ_RPR *Crystal structure of human coagulation factor xa complexed with rpr128515*



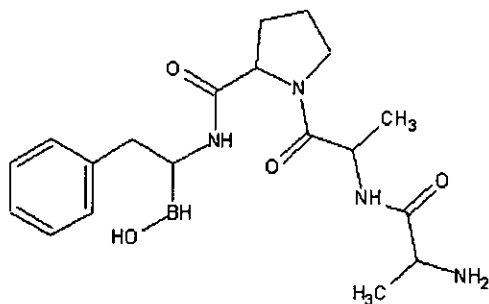
1d4p_BPP *Crystal structure of human alpha thrombin in complex with 5-amidinoindole-4-benzylpiperidine inhibitor*



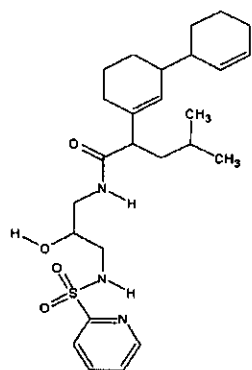
1bmm_BM2 *Human alpha-thrombin complexed with [s-(r,R)]-4-[(Aminoiminomethyl)amino]-n-[[1-[3-hydroxy-2-[(2-naphthalenylsulfonyl)amino]-1-oxopropyl]-2-pyrrolidinyl]methyl]butanamide (bms-186282)*



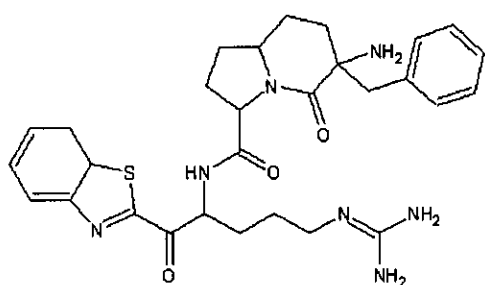
1NFW_RRR *Crystal structure of human coagulation factor xa complexed with rpr209685*



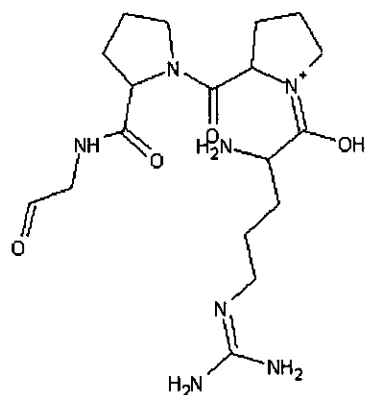
1P06_PPP *Alpha-lytic protease complex with methoxysuccinyl- Ala- Ala- Pro- Phenylalanine boronic acid*



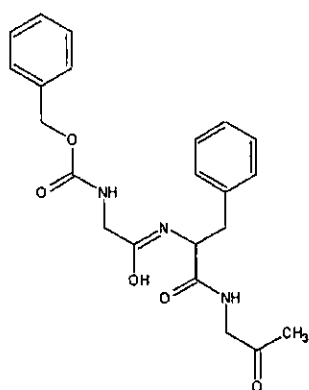
1bgo_I10 *Crystal structure of cysteine protease human cathepsin k in complex with a covalent peptidomimetic inhibitor*



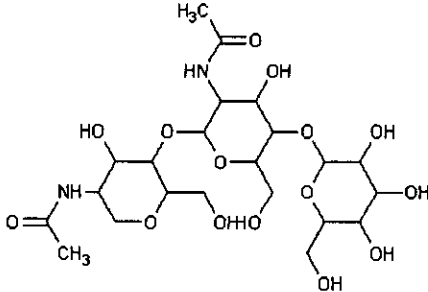
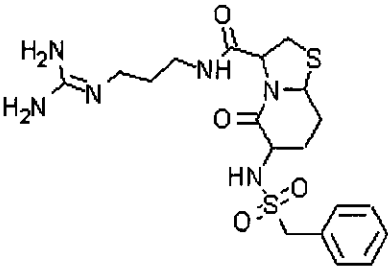
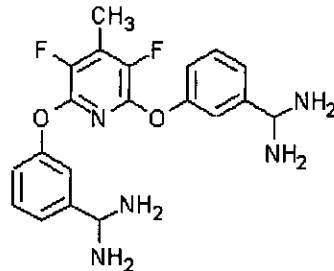
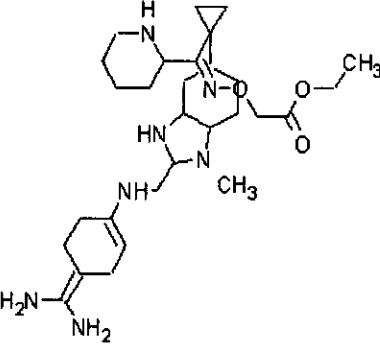
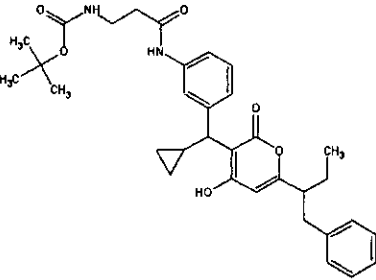
1b5g_BCC *Human thrombin complexed with novel synthetic peptide mimetic inhibitor and hirugen*

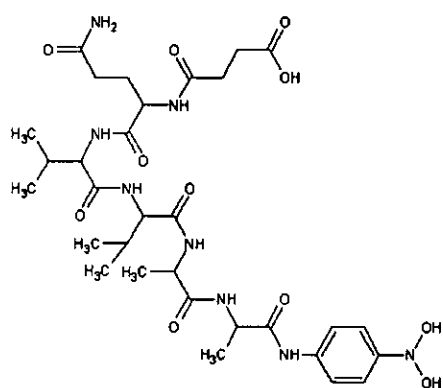


1ny2_PPP *Human alpha thrombin inhibited by rppgf and hirugen*



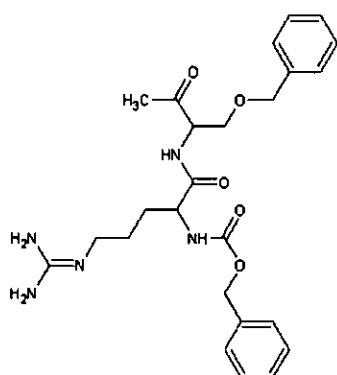
5pad_BZO *Papain -benzyloxycarbonyl-glycyl-phenylalanyl-methylenylglycyl derivative (/zgpgck)*

Group 3: in the S1 pocket and catalytic site	PDB code	Description
	1nb5_NAG	<i>Crystal structure of stefin a in complex with cathepsin h</i>
	1bhx_R56	<i>X-ray structure of the complex of human alpha thrombin with the inhibitor sdz 229-357</i>
	1qb6_623	<i>Bovine trypsin 3,3'-[3,5-difluoro-4-methyl-2, 6-pyridinediylbis(oxy)]bis(benzenecarboximide) (zk-805623 complex)</i>
	1G2L_T87	<i>Factor xa inhibitor complex</i>
	2upj_U02	<i>HIV-1 protease complex with u100313 ([3-[[3-[cyclopropyl [4-hydroxy-2oxo-6-[1-(phenylmethyl)propyl]-2 h-pyran-3-yl] methyl]phenyl]amino]-3-oxo-propyl]carbamic acid tert-butyl ester)</i>



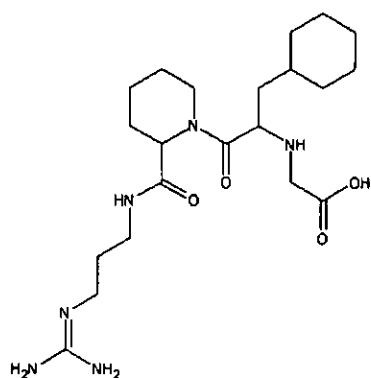
1pip_SIN

Papain complex with succinyl-gln-val-val-ala-ala-p-nitroanilide



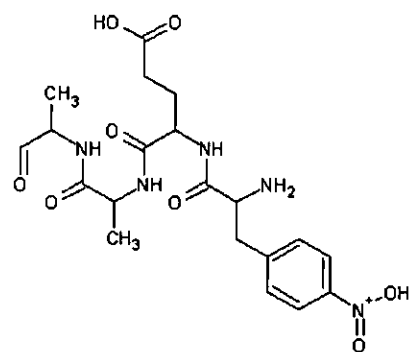
1the_CBZ

Crystal structures of recombinant rat cathepsin b and a cathepsin b-inhibitor complex: implications for structure-based inhibitor design



1k1o_IGN

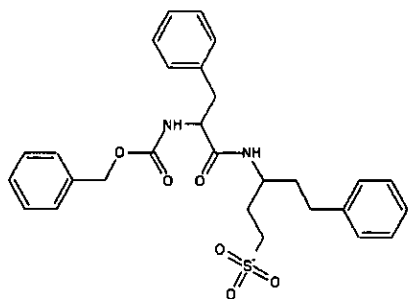
Bovine trypsin-inhibitor complex



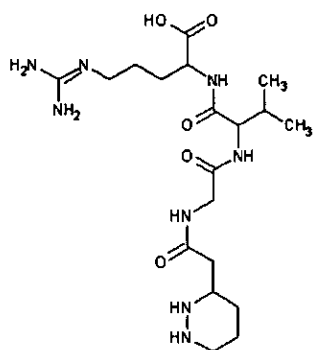
1ytj_PPJ

Siv protease crystallized with peptide product

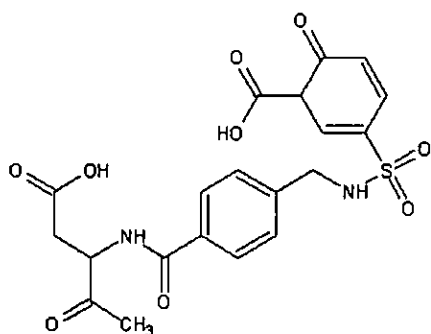
1ewo_VSC *The cysteine protease cruzain bound to
wrr-204*



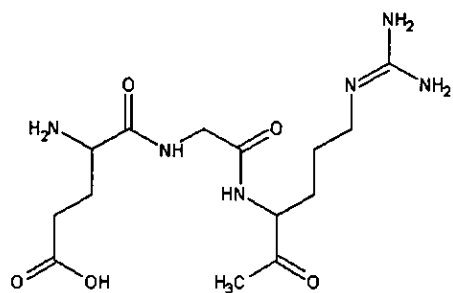
3hat_RNG *Alpha-thrombin complexed with hirugen and
fpam (fibrinopeptide a mimic)*



1nms_161 *Caspase-3 tethered to irreversible inhibitor*



1dx5_PPP *Crystal structure of the
thrombin-thrombomodulin complex*



5. 結論與建議

由 Serine protease 抑制劑的設計概念觀之，良好的抑制劑必須和 binding site 有足夠的結合與辨識能力。此外還必須能夠和 Serine 之 OH 基團產生交互作用，阻斷催化發生。因此在設計抑制劑時首要便是找出有可能和 binding site 的專一性控制區產生影響的分子結構，而後再以此一分子結構作為修飾的基礎，發展新的抑制劑。除此之外，因這幾個重要位置的胺基酸不易發生突變，因此針對這些位置設計的藥物，較不易因病毒的突變而失去效用。在我們 docking 的結果分析裡，docking 位置在 binding site 裡的，及 docking 位置在 S1 pocket 及 catalytic triad 的化合物較符合成為前導藥物的條件，可以做為實驗時的優先選擇。

6. 審查意見之回覆

6.1 審查意見

本研究在預測藥物之副作用及毒性方面之預期目標，未在報告中有實驗結果或結論之呈現，請補充修正。

6.2 回覆意見

有關毒性預測方面，我們已嘗試以下方法：

1. Building P450 methods

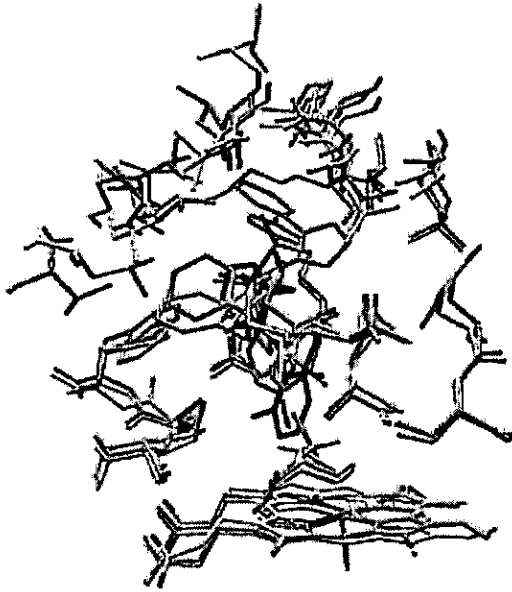
我們採用人類 P450 2C9 蛋白質作為預測系統之發展基礎。利用 GEMDOCK 將已知 P450 2C9 之受質或抑制劑嵌合(docking)回人類 P450 2C9 蛋白質(PDB code: 10G5, 10G2)，並將 dock 的結構與位置做成 profiles，希望藉由找出這些已知受質與抑制劑 profiles 的差異，並加以應用達成毒性預測的目標。

我們測試了 GEMDOCK 對已知的人類 P450 蛋白質與 ligand 結晶，是否能夠成功的嵌合(docking)回原結晶位置。圖一為測試結果，GEMDOCK 可以成功將其受質嵌合回原位。

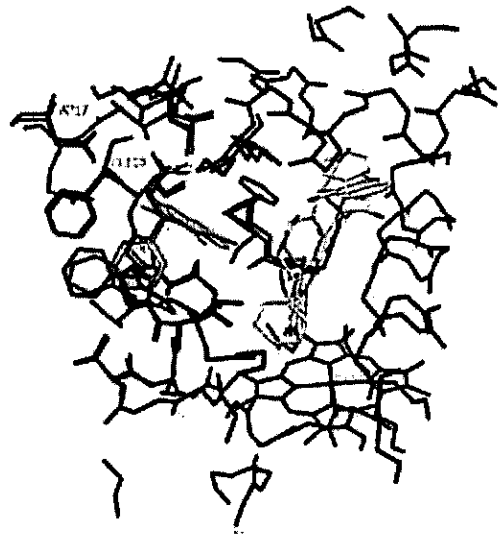


圖一、對10G5與其分子SWF的分子嵌合結果。藍色部分為10G5之結合區域，下方CPK顏色者為HEME，左方標示橘色者為原結晶之分子，和橘色位置重和的是分子嵌合結果，在左側標示CPK顏色的是與原結晶分子有氫鍵之氨基酸。

但由文獻得知此一位置並不是P450正常受質結合與催化位置，一般認為正常的位置應如圖二所示，受質位於heme之上方。因此我們對有受質的10G5與沒有受質的10G2分別加以測試，得到結果如圖三所示，對10G2，SWF並沒有回到10G5受質的位置，反而是偏好heme上方的位置。因此我們用來作分子嵌合的蛋白質仍然選擇有受質的10G5，但是對結合區作適當的修飾，使得SWF能夠回到於接近催化狀態的位置，採用這樣的活性區域作為作docking profile的目標(docking target)。



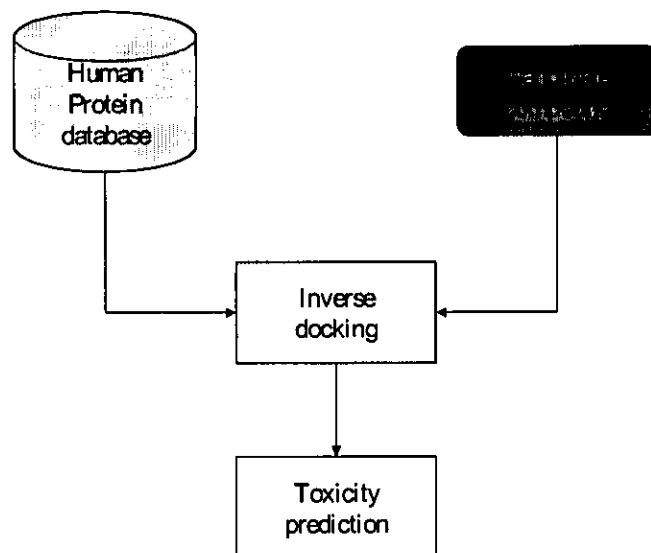
圖二、兔子之P450 2C5的結構。淡綠色部分為1DT6，淡黃色部分為1N6B，橘色者為其受質DMZ，淡紅色部分為1NR6，紅色者為其受質DIF。此三者為同一蛋白質之結晶，受質通常位於Heme之上方，且受質不同造成結合區域之結構也不同。



圖三、10G5_swf與10G2的分子嵌合結果與結合區的差異。藍色與粉紅色者是10G5，橘色是結晶位置，青色是嵌合結果。CPK顏色者則是swf對10G2的嵌合結果，兩者有相當大的差異。綠色部分是10G2與10G5在結合區的蛋白質結構不同處。

我們將已經修飾過的10G5的結合區和已知的2C9之抑制劑與非抑制劑作分子嵌合(molecular docking)實驗，並將其嵌合結果之最低能量構形與位置作成profiles，但是目前實際效果並不是很好。因此我們正在思考引入其他的資訊，例如分子物化性質與特性，使毒性預測準確率達到預期效果。

2. Inverse docking



圖四、Inverse docking 流程圖

Inverse docking 是以已經篩選出的可能抑制劑為標的，針對人類蛋白質資料庫中的蛋白質做鉗合預測 (docking) (圖四)，目的在找出是否還有其他的蛋白質可能會與篩選出的抑制劑結合，而影響其活性，即預測此抑制劑可能造成的副作用。在做鉗合預測的時候，由於蛋白質的活性區域目前被程式認為是剛體，所以不會因與小分子的結合而改變 (induce-fit)。這在鉗合預測時是比較不合理的，目前的解決方法就是以蛋白質和抑制劑共同結晶 (protein-ligand complexes) 的結構為鉗合預測的標的，以減少因為活性區域的結構改變 (induce-fit) 造成的誤差。所以在建立蛋白質資料庫的時候，主要是以與小分子結合的蛋白質為主，但是因為目前已知的蛋白質和抑制劑共同結晶的結構數目有限，所以無法對人類的所有蛋白質做完整的毒性預測。在 PDB 中人類的蛋白質 X-ray 結晶結構共有 4,219 個，其中與抑制劑的共同結晶結構有 981 個 (表一)。且目前我們的程式仍然將蛋白質的部分當作剛體，在諸多限制之下，inverse docking 的效果可能不是很理想，所以未在結報中提出。

1HNE	1E99	1E8Z	1BB0	1FTA	1HNI	1SLM	1BB5	1SLN	1IKN	1E9A	1E9B	1E9C	1E9D	1AE8	1E9E	1E9F	1Q4L
1HKW	1G3N	1E9H	1DZH	1Q4N	1IKX	1IKY	1HNV	1OYN	1QSA	1H0V	1OYQ	1H0W	1UGH	1G49	1OY7	1BBS	1LD7
1SHVP	1FUJ	1G4K	1HIP	1IJ9	1IJZ	1HIQ	1HIR	1HIS	1OYN	1H24	1HP7	1BCD	1PW2	1H25	1H26	1OZ1	1H27
1PW6	1OYT	1CA0	1MAR	1CA2	1BCU	1CA8	1MAZ	1BD8	1JK3	1H2K	1BD9	1HPO	1H2L	1H2M	1H2N	1RQQ	1QTN
1HPS	1PWM	1HPT	1HPV	1GSS	1BDA	1R42	1HPX	1S18	1G55	1HPZ	1FVT	1FVV	1F92	1F8U	1UPJ	1BDL	1JKH
1BDQ	1BDR	1K10	1F9E	1AGW	1PXH	1PAH	1R4L	1PXI	1PXJ	1PKX	1PXL	1HQU	1G73	12GI	1JLA	1XAN	1JLB
1JLD	1JLE	1JLF	1EZX	1JLG	1CC0	1NA7	1LG1	1LG2	1UK0	1UK2	1JLQ	1GUH	1AHT	1UK3	1G7F	1UK4	1G7G
1GUL	1TMU	1A18	1S2Q	1HIR	1RT1	1RT2	1RT3	1PYN	1RT4	1PYO	1RT5	1H52	1ACH	1RT6	1H56	1H53	1RT7
1H4W	1HMB	1A1D	1QVY	1FXY	1AII	1HMI	1S3B	1NB3	1NB5	1PHV	1S3E	1AIN	1HSG	1RTH	1HSH	1JMO	1HSI
1R6K	1AIX	1KJY	1R6N	1Q9M	1S3U	1S3V	1S3W	1LHC	1LHD	1LHE	1LHF	1R78	1LHG	1G96	1DAN	1VU	1MEM
1DB2	1MER	1DB4	1BGO	1MES	1DB5	1MET	1HTE	1MEU	1HTF	1HTG	1NOS	1TOM	1DAZ	1L4	1HT1	1AJV	1QXK
1RV1	1R80	1R81	1R82	1R7T	1R7U	1UMA	1R7V	1R7X	1BHF	1R7Y	1KLM	1MFU	1R1	1MFV	1HUG	1GXD	1LIT
1UN3	1BHX	1B17	1UN4	1B18	1UN5	1I40	1I51	1GY3	1KMC	1CA2	1BK	1BL	1BM	1STF	1HVC	1HVD	1HVE
1DCY	1LJT	1HVH	1HVI	1KMS	1HVJ	1KMT	1HVK	1BTV	1CAB	1HVL	1KMV	1GSS	1AM6	1JQ6	1JQ7	1ISR	1HVS
1CGE	1EAK	1CAN	1CGF	1KPI	1HW8	1HW9	1CGH	1CGI	1GZ8	1CGJ	1OC0	1CGL	1LKK	1EAW	1DE4	1K15	1PAH
1HWI	1HWJ	1KNT	1HWK	1HWL	1UOU	1YDA	1ITQ	1YDB	130C	1YDC	1YDD	1ITU	1HX1	1I73	1I76	1EBK	1NFI
1EC2	1ANG	1EC3	1UPJ	1HXB	1EBW	1EBY	1HXE	1EBZ	1HXF	1A0L	1H4N	1ANT	1BL6	1JRR	1B17	1ANW	1KIT
1RVD	1AID	1GDS	1HXW	1HY7	1KPA	1KPB	1KPC	1K2B	1PAU	1KPF	1K2C	1CIL	1CIM	1CIN	1ECV	1ODW	
1ODY	1CIZ	1LMW	1BLX	1IVP	1IVQ	1JSU	1FAK	1MKD	1OEM	1OEO	1A2C	1CIL	1OES	1FAV	1OET	1BMK	1OEU
1OEV	1BMM	1BMM	1BMQ	1NHX	1BN1	1BN3	1BN4	1AQ1	1A30	1KME	1LO6	1KQU	1PCG	1FC0	1EEN	1EEO	1A3B
1EET	1A3E	1PD8	1BMM	1CKP	1PD9	1FBZ	1BNN	1VPP	1BNQ	1USN	1DI8	1BNT	1DI9	1BNU	1BNV	1BNW	1PDB
1AQW	1AQX	1DIA	1XKA	1A46	1XKB	1DIB	1K58	1MMB	1L59	1CA2	1DIF	1BOA	1DIG	1LPB	1OAT	1K5A	1L2E
1OGU	1B1I	1OGS	1NIS	1MMP	1B1J	1INT	1MMQ	1DIT	1CAA	1NIU	1MMR	1LQ0	1GSS	1JVP	1UJ3	1AS4	1GAG
1U08	1A4W	1A4Y	1TLH	1OHJ	1OHK	1MOB	1QBR	1QBS	1QBT	1NKK	1QBU	1LQF	1CMI	1QBV	1NKM	1CMK	1OHR
1PF8	1A5G	1A5H	1EH6	1KNT	1EH7	1EH8	1HZH	1NL6	1JZI	1VRT	1NL9	1VRU	1M14	1K6P	1AT3	1A61	1QCF
1K6T	1JWT	1K6V	1FEJ	1GBN	1ANG	1NLJ	1MIP	1F0	1UJZ	1B3D	1OIQ	1OIR	1OIT	1ATH	1BQM	1ATK	1B3K
1BQQ	1B30	1AU0	1C0T	1CNW	1C0U	1CNX	1UVR	1KV1	1AU2	1CNV	1QDD	1KV2	1AU3	1JXQ	1ATU	1AU4	1AID
1A7I	1UW7	1AU8	1FFF	1THN	1FFT	1C1B	1GDO	1BRC	1C1C	1UWB	1PH0	1QDU	1A7C	1DL0	1FG6	1GCZ	1NMS
1MQ0	1CP3	1KVO	1AUT	1HAG	1JYQ	1HAH	1JYR	1AV5	1FGC	1HAK	1JYU	1A85	1A86	1B59	1FGI	1HAO	1HAP
1OKM	1OKN	1EJN	1FH0	1CPI	1OL0	1GR1	1OL1	1OL2	1B5G	1AVH	1OKU	1A8G	1OKV	1OKW	1KME	1KWM	1A8K
1CPU	1NNY	1AVR	1A94	1QFK	1HBT	1HVP	1R4D	1B6A	1HBV	1HBY	1M4H	1AWF	1C3I	1ZUS	1N1M	1LUJ	1AWH
1HCB	1M51	1A9M	1MS6	1MRW	1MRX	1HCK	1A9U	1NPA	1DOA	1CRA	1PK	1SAV	1PJP	1AXA	1AXC	1CAA	1GFW
1MSM	1L8G	1MSN	1M5K	1M50	1HVT	1NPV	1M5P	1AXN	1NPW	1LW0	1SBG	1LW2	1MT7	1MT8	1M5V	1N39	1MT9
1B7X	1B1	1O0D	1N3A	1IAS	1MTB	1ONI	1IAU	1HDT	1LWC	1REV	1LWE	1LWF	1CSG	1EMR	1PL6	1PL7	1PKX
1FK9	1BPW	1KZK	1BPK	1MTR	1BPK	1O08	1ONY	1BPK	1ONZ	1AYP	1QIA	1QIB	1M73	1AZ1	1QIC	1AZ2	1HEF
1CSW	1AYU	1CSX	1AYV	1CSY	1AYW	1CSZ	1B8Y	1FKG	1FKH	1FKI	1UPJ	1D3D	1FKN	1BWC	1FKO	1FKP	1WPO
1G18	1G19	1D3P	1PAH	1NRS	1D3Q	1M7N	1AZM	1C70	1D3T	1M7Q	1HFC	1API	1QJB	1C6X	1C6Y	1PM6	1C6Z
1AZX	1G1I	1GU	1TAW	1OPH	1ID4	1HFS	1PMN	1NSI	1PMQ	1D4H	1D4I	1D4J	1G17	1PMU	1G18	1PMV	1G19
1IDA	1IDB	1E1V	1D4S	1DS6	1E1X	1C83	1GJA	1GJB	1C84	1G1C	1FGI	1D4Y	1C85	1G1D	1C86	13987	1HTC
1JAO	1C88	1JAP	1JAQ	1RHO	1E2D	1E2E	1E2F	1E2G	1TBZ	1HH4	1PO1	1N6J	1BYG	1PO2	1C2P	1D5M	1Q3P
1M9K	1E+31	1FN7	1O41	1O42	1MWP	1M9M	1O43	1O44	1MX1	1D5S	1IEC	1O45	1MX2	1O46	1M9Q	1LZQ	1O47
1O48	1O49	1MX6	1CVW	1C8T	1D5X	1GKC	1IEI	1GKD	1D5Z	1HVP	1O4A	1O4B	1O4C	1JBU	1O4D	1O4E	1O4F
1O4G	1O4H	1D6E	1O4I	1BZC	1O4J	1O4K	1N7I	1O4L	1O4M	1N7J	1BZH	1O4N	1O4O	1BZJ	1OS2	1O4P	1D6N
1O4R	1BZM	1OS9	1HCK	1BZS	1QMB	1HHI	1JCN	1HHJ	1P2A	1BZY	1JQC	1PPF	1PPG	1QMN	1GLO	1CAC	1F1J
1KAK	1NW9	1D7X	1FPC	1KAV	1DVA	1PAI	1RKO	1RKP	1MZC	1D8F	1FQ1	1NWL	1DVM	1DVN	1KBC	1D8M	1G05
1FPZ	1H8B	1DVU	1TFG	1DW6	1HKB	1KBO	1AAP	1UPJ	1KBQ	1AAQ	1DWB	1HKN	1DWC	1AAX	1DWD	1GNM	1H8
1GNM	1HHT	1GNO	1JF7	1TFX	1QIM	1OUK	1PRO	1PAH	1RM8	1P4R	1CZM	1API	1FOX	1CZQ	1UDI	1OUY	1DX5
1P5E	1OVE	1IIQ	1PSI	1HLT	1FRO	1UBA	1Q2U	1NZ7	1HAT	1RNE	1VBS	1Q3A	1VBT	1THP	1THR	1JH1	1BA8
1G2K	1HMR	1HMS	1CPU	1F5K	1H00	1HMT	1F5L	1FSO	1H01	1FT0	1FT1	1FT2	1H06	1H07	1G35	1H08	1NZQ
1E98																	

表一、與抑制劑的共同結晶結構 981 個

3. 類神經網路毒性預測系統

我們擷取已知為 P450 抑制劑與非抑制劑之化合物 2 維平面結構與化學式的特徵值。然後將化合物分為兩組，抑制劑組與非抑制劑組，並將每一個化合物的特徵值之統計量送入類神經網路，作為訓練組。利用訓練組，訓練類神經網路系統分辨，並找出抑制劑與非抑制劑之重要不同與相同之特徵值。然後再將訓練好的類神經網路應用到分辨具有 P450 抑制效果的化合物之辨認。

H-ar.	H-na.	C-H.	C=O.	C-qj.	N sp2.	N-na.	N=C.	O-na.	O=x.
Psp2.	F-x.	Cl-x.	Br-x.	I-x.	Metal.	#cir1.	#atom.	chrg	

上表為我們選出作為化合物性質代表之特徵值，其中 ar 代表芳香環，na 代表非芳香環，C-H 代表接碳的氫，C=O 則是酮基，sp2 為 sp2 軌域，N=C 則是接氮而且具有雙鍵的碳，x 是任意原子，#cir1 是環數，#atom 是原子總數，chrg 是帶電量。

我們希望上表所分類與擷取的化合物特徵能夠和化合物毒性有相關性，而後藉著類神經網路系統的幫助加以找出真正與毒性相關的特徵，對 P450 抑制劑與非抑制劑得到適當的分辨率。未來除了類神經網路以外，我們也將嘗試使用 SVM(support vector machine)，作為分類的工具。

目前我們已經有了毒性預測系統之原型(prototype)，但是實際效果並不如預期。因此我們未來在計畫中將會對毒性預測系統之原型繼續加以發展，希望能夠達到理想的毒性預測能力。

7. 參考文獻

1. Yang, J.M., et al., *GEM: A gaussian evolutionary method for predicting protein side-chain conformations*. *Protein Science*, 2002. 11: p. 1897~1907.
2. Yang, J.M., *An evolutionary approach for molecular docking*. *Lecture Notes in Computer Science*, 2003. 2724: p. 2372-2383.
3. Murthy, H.M., S. Clum, and R. Padmanabhan, *Dengue virus NS3 serine protease. Crystal structure and insights into interaction of the active site with substrates by molecular modeling and structural analysis of mutational effects*. *J Biol Chem*, 1999. 274(9): p. 5573-80.
4. Bartelma, G and R. Padmanabhan, *Expression, purification, and characterization of the RNA 5'-triphosphatase activity of dengue virus type 2 nonstructural protein 3*. *Virology*, 2002. 299(1): p. 122-32.
5. Modis, Y., et al., *A ligand-binding pocket in the dengue virus envelope glycoprotein*. *Proc Natl Acad Sci USA*, 2003. 100(12): p. 6986-91.
6. Murthy, H.M., et al., *Crystal structure of Dengue virus NS3 protease in complex with a Bowman-Birk inhibitor: implications for flaviviral polyprotein processing and drug*

- design*. J Mol Biol, 2000. 301(4): p. 759-67.
7. Bissantz, C., G. Folkers, and D. Rognan, *Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations*. J Med Chem, 2000. 43: p. 4759-4767.
 8. Champness, J.N., et al., *Exploring the Active Site of Herpes Simplex Virus Type-1 Thymidine Kinase by X-Ray Crystallography of Complexes With Aciclovir and Other Ligands*. Proteins, 1998. 32: p. 350-361.