

基於幾何形狀分析的蛋白質接合研究

學生：黃滄智

指導教授：荊宇泰 博士

國立交通大學資訊科學與工程研究所



摘要

在儀器進步下，愈來愈多的蛋白質結構被解釋出來，也提供了解其間作用與藥物設計的另一途徑。在我的論文中使用幾何形狀互補程度當接合評選的好壞，不考慮化學作用。首先對產生的蛋白質網格表面作形狀分析、分類，在所找出的互補表面上計算其曲率當作轉換座標的依據，轉換後再經一次 ICP 校準，套用形狀互補計分函數後，以突顯較好的接合結果。

Protein Docking Based on Geometry Shape Analysis

Student : Tsang-Chih Huang

Advisor : Dr. Yu-Tai Ching

Department of Computer and Information Science

National Chiao Tung University



Abstract

More and more proteins structures are constructed in advanced equipments. It provides the path to understand reactions between proteins and structure drug design. In my thesis, first, analyzing the generating proteins meshes and calculating the classifying surfaces' principal curvatures for two proteins coordinate transformation. Then doing ICP registration and applying shape complementary scoring function to show the better docking results.

誌謝

感謝老師、委員、實驗室同學、朋友與家人、不小心被忘了的，與每個人背後的那一個更偉大的人。

除目前收穫外，我知道這段日子將為未來生活帶來勇氣與希望。

謝謝大家。



目錄

中文摘要.....	i
英文摘要.....	ii
誌謝.....	iii
目錄.....	iv
圖目錄.....	v
1. 緒論.....	1
1.1 前言.....	1
1.2 研究動機與目的.....	2
1.3 論文架構.....	3
2. 相關研究.....	4
2.1 表示.....	5
2.2 比對.....	7
2.3 計分.....	9
3. 產生蛋白質三維網格資料與接合.....	11
3.1 產生與處理網格資料.....	11
3.2 接合.....	14
3.2.1 網格形狀分析.....	14
3.2.2 表面匹配.....	15
3.2.3 主曲率方向.....	16
3.2.4 座標轉換.....	19
3.2.5 轉換座標後的表面匹配.....	20
3.2.6 計分.....	20
4. 實驗結果與討論.....	22
4.1 實驗流程與參數設定.....	22
4.2 實驗結果.....	27
4.3 討論.....	32
5. 結論、未來工作與展望.....	33

圖目錄

圖 2-1 年份與所解出蛋白質結構數量統計圖	2
圖 3-1 分子凡得瓦力表面與使用水分子大小的球所接觸出的區域示意圖	5
圖 3-2 對分子表面區分出凸、凹與鞍部	5
圖 3-3 檢查球與分子重疊體積，留下極大、極小值當作形狀代表關鍵點	6
圖 3-4 對分子表面的凸、凹與鞍部產生關鍵點	6
圖 3-5 PDB 1FKF 的原子分佈與 alpha shape 網格表面結果	7
圖 3-6 使用網格結構（右）可以用表面距離圈選相鄰點	7
圖 3-7 對一個兩關鍵點產生識別簽名	8
圖 3-8 Geometry Hashing 演算法兩步驟示意圖	8
圖 3-9 按接觸表面與距離表面不同的程度來計分	10
圖 4-1 產生蛋白質網格與接合流程圖	11
圖 4-2 蛋白質網格重建結果	12
圖 4-3 從原始網格資料藉由貼補前處理，找出最大的元件	13
圖 4-4 Half-Edge 資料結構示意圖	13
圖 4-5 CGAL 中利用 Half-Edge 產生網格結構的規定順序	14
圖 4-6 給一個點 v （圖中藍色點），FLOOD ($v, 2$) 結果	14
圖 4-7 分析點集法向量與其重心以區分形狀	15
圖 4-8 以表面中心點為原點，法向量與兩主曲率方向所成的座標系統	16
圖 4-9 對一個表面計算正剖面（normal section）與正曲率（normal curvature）	16
圖 4-10 利用中心點上的法向量與兩個主軸方向，建立四個右手則的座標系統	17
圖 4-11 (a) torus.obj (b) 主曲率較小方向（平行圓周）(c) 主曲率較大方向（垂直圓周）	19
圖 4-12 藉由法向量與主軸構成的座標系統轉換配體座標到受體座標系統中 ..	20
圖 4-13 兩種接合重疊原子數目相近，但應對較靠近內部原子（紅色部分）的相交扣分較多	21
圖 5-1 實驗流程使用輔助參數	24
圖 5-2 對 2TGP 化合物中兩個蛋白質，分別標示出接合位置的形狀匹配表面 ..	28
圖 5-3 放大觀察 2TGP 化合物配體與受體不同[參數]的表面構成點集	29
圖 5-4 放大觀察 2MHB 化合物配體與受體不同[參數]的表面構成點集	30
表 4-1 實驗之化合物資料	22
表 4-2 程式相關參數	23

表 4-3 (MIN_SHP_VERTICES = 20, ICP_INIT_POS_FLOOD_N = 1)27
表 4-4 (MIN_SHP_VERTICES = 80, ICP_INIT_POS_FLOOD_N = 1)28



第一章 緒論

1.1 前言

二〇〇三年四月，人類基因體圖譜解碼完成，宣告後基因時代的來臨，生物科技的探索從基因序列移向更高層次的蛋白質體學（proteomics）研究，雖然每個細胞上都帶有成為完整個體的遺傳基因訊息，但並非所有基因都會在細胞中表現出來，且就算相同基因，也可能因細胞的不同而有相異的蛋白質形態與彼此間不同的交互作用，例如：肝臟和腦部細胞的蛋白質便有不同，這是因為大部份蛋白質在不同細胞內會經過不同的轉錄、轉譯與修飾，使它們的結構、功能有所改變，細胞內的蛋白質間交互作用若失去平衡、被破壞，將造成細胞病變。所以更多樣性的蛋白質體研究比基因體更為複雜。

受益於分析儀器與電腦計算能力的進步，實驗過程產生的大量生物資訊資料可以應用電腦科技去搜尋、整理出有用的資訊，藉以加快研究腳步，於是也就有了各種跨領域的資源加入，各種整合性技術的使用，除了能更有效率的完成蛋白質功能分析外，也可以應用到藥物設計或疾病診斷的領域。

蛋白質體的研究應用很廣泛，預計在農、漁、牧與醫學、製藥、生技業都會產生變革性的影響，將是繼工業革命與電子資訊革命後的另一次重大改變，所以有能力發展的國家莫不投入相當大的人力、物力發展。

1.2 研究動機與目的

下圖示取自 PDB (The RCSB Protein Data Bank) 網站上的所解出化合物的三維結構的統計圖，可以觀察出在近十年內因為儀器的進步，讓成果數目成長的很快，若再加上一些電腦結構模擬軟體的幫助，相信可以被用來當作電腦模擬的蛋白質結構是超出這個數目。

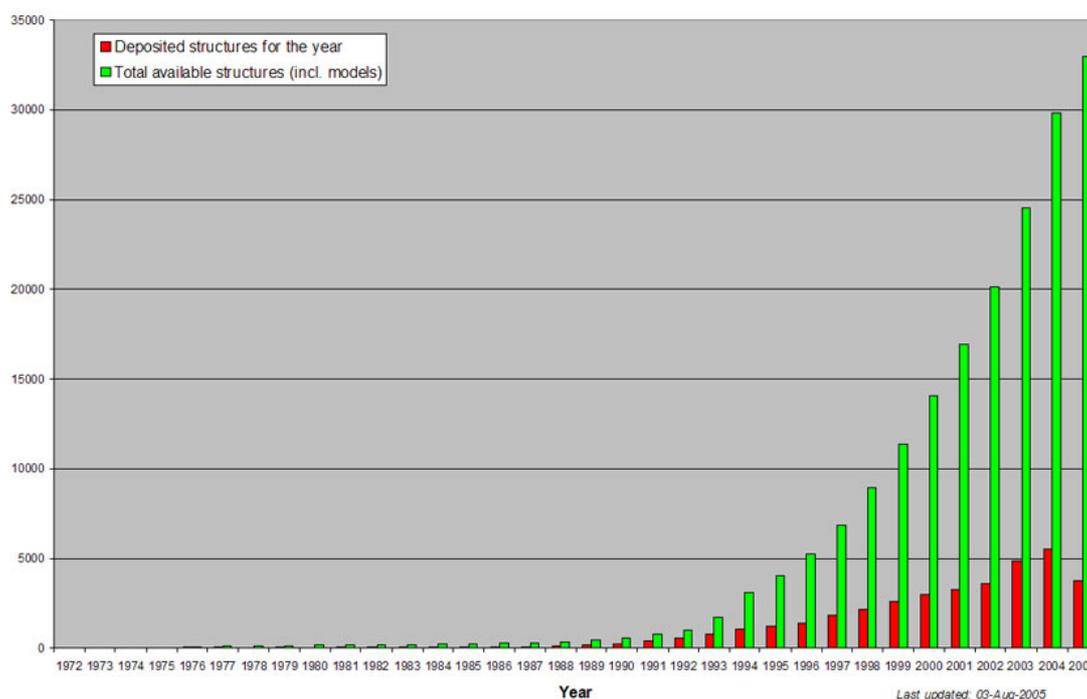


圖 1-1 年份與所解出蛋白質結構數量統計圖

在獲知蛋白質結構後，藉由觀察其在原子層次的互相作用情形，除了在了解蛋白質功能，與相較於傳統從事藥物設計的方法外，預計也可替藥物設計提供另一條尋找先導化合物 (lead compound) 的途徑，我們希望可以提出一個演算方法，替生物學家在利用蛋白質接合以達成上述工作中，篩選一些比較可能或排除不可能的結果，利用電腦適合執行重複性大量運算任務的優勢，加速整個過程的處理時間。

1.3 論文架構

我們論文共分為五章。第一章簡介生物資訊在生物科技中所扮演的角色和研究動機與目的。在第二章中，介紹蛋白質接合的相關研究與方法分類。第三章是對提出方法的基本理論與演算法做詳細說明。第四章實驗步驟則顯示實驗結果與說明使用參數意義。第五章為結論、未來工作與展望。



第二章 相關研究

蛋白質接合實驗中常見有幾種分類，分別是依處理資料的解出結構的時機可分成 bound 或 unbound，若是依資料大小可分成 protein-protein 與 protein-ligand，依形狀的變異程度可分為 rigid 與 flexible。

bound 與 unbound 的分別在於形成化合物時的解出結構與個別在形成化合物前的解出結構，拿形成化合物後的蛋白質來作接合實驗，因為形狀比較接近真正作用的狀態，所以比 unbound 資料更容易得到好結果，而 protein 與 ligand 通常是指蛋白質資料量的大小，flexible 與 rigid 則是蛋白質結構上的鍵結是否可以依照某個規則轉動。

是否形成一個好的接合結果，通常會從兩個蛋白質的幾何與化學性質上來探討，幾何上要求兩者形狀盡量互補，化學則是經由一些如：氫鍵 (Hydrogen bonds)、電荷 (Electrostatic interactions) 與水解 (Solvation energy) 等特性強弱來決定，本篇論文的範圍在處理 bound 資料、rigid 接合與 protein-protein 資料實驗。

蛋白質接合主要可區分成下面三個子問題：

- 蛋白質表示方式。
- 在蛋白質上找出可能產生接合的位置。
- 如何對產生的接合結果計分排名。

2.1 表示

在解出結構，有了原子座標後，首先必須先為這團原子定義出一個表示區域，根據分子表面定義為：溶劑分子可以與之接觸的區域 (Solvent Accessible Area)，由 Lee 與 Richards 在 [1] 中提出，並使用一個半徑接近水分子大小的球 (probe sphere) 在分子表面滾動出表面輪廓 (圖 2-1) 用以描述蛋白質分子表面。

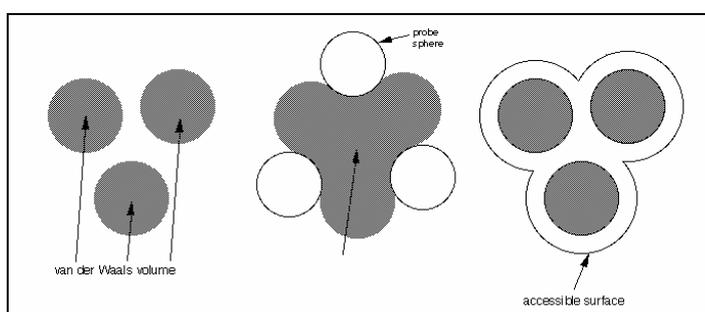


圖 2-1 分子凡得瓦力表面與使用水分子大小的球所接觸出的區域示意圖

Connolly [3] 將此概念延伸，替分子表面依與水分子不同接觸原子數目情況分出三種表面形狀：凸面 (convex)、凹面 (concave) 與鞍部 (saddle) (圖 2-2)，稍後又提出一個可以用較少的關鍵點來描述表面的方法 [4]，在每個點上擺一個球，計算這個球與分子所重疊的體積，然後在一個給定範圍內把各點的球與分子體積重疊的極大 (凹處明顯) 與極小值 (凸處明顯) 留下當作關鍵點 (圖 2-3)，以盡量少的點來保留分子表面的凹凸特徵處。

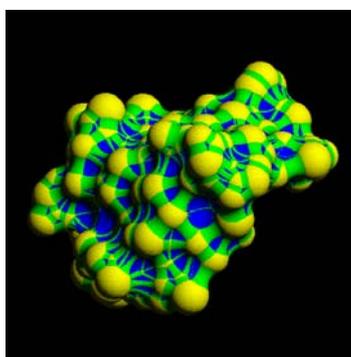


圖 2-2 對分子表面區分出凸、凹與鞍部

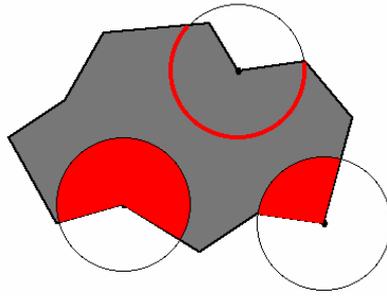


圖 2-3 檢查球與分子重疊體積，留下極大、極小值當作形狀代表關鍵點

Lin [9] 則提出了另外的關鍵點表示法，其作法是對 Connolly 所分類的不同形狀表面都產生一個代表它的關鍵點。這些關鍵點為 Connolly 三種形狀表面的重心對這個表面投影所交出的點，投影方向則依不同種類表面而異，如果是一個凸面是用原本形成這個凸面形狀的原子重心指向構成凸面點集的重心方向當投影方向，如果是凹面則對當初形成這個形狀所掠過的球的重心朝構成凹面點集的重心為方向作投影方向，若是一個鞍面則是先利用任意一個構成此鞍面的甜甜圈形狀 (torus) 的外圓半徑軸加上鞍面形狀重心計算出位於甜甜圈外圓上的投影中心，再用這個中心內插出位於甜甜圈內圓上位於鞍面形狀上的關鍵點。

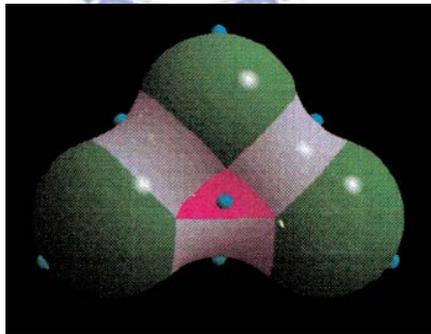


圖 2-4 對分子表面的凸、凹與鞍部產生關鍵點

另外還有用網格 (mesh surface) 的結構表示 Connolly [3] 的 MS Dot 分子表面產生程式除了產生 dot surface 外也可以產生 mesh surface、Peters [8] 則是用 alpha shape 來產生不同精細度 (level of detail) 的蛋白質形狀。使用網格描述表面，可以比較容易處理點在表面距離 (Geodesic) 上的遠近關係，把某個距離內的點包含進來，若使用歐式空間距離算法可能會造成點在表面上其實是沒有接

續在一起 (圖 2-6)，而用表面距離計算，因為把點與點間是否相鄰的資訊也考慮在內，所以能避免上述問題。



圖 2-5 PDB 1FKF 的原子分佈與 alpha shape 網格表面結果

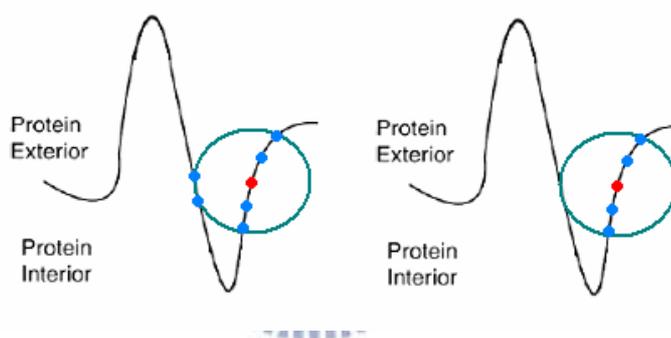


圖 2-6 使用網格結構 (右) 可以用表面距離圈選相鄰點

2.2 比對

在建立蛋白質的表示結構後，我們必須在這兩個結構間搜尋、比對出一個可能可以接合的位置組合在稍後計分用。

Connolly 先是把找出的關鍵點以凹對凸作配對接合，後來又提出四個關鍵點的比對，來降低處理時間，但如果遇到接合位置比較平坦或接合位置上的關鍵點次序不同就會造成較差的結果，Norel et al. [9]、[10] 則用自己的所算出的關鍵點加上點的法向量資訊來處理，對一個距離內的兩關鍵點以其連線 (\overline{AB}) 與此線段與兩端法向量的角度 ($\angle a$ 、 $\angle b$)，再加上兩法向量各自與關鍵點連線所

張平面的夾角 ($\angle c$) 這些資訊來當作兩個關鍵點的識別簽名 (signature)。比對時，對蛋白質上的識別簽名項目加上一些限制，用來過濾一些形狀差異過大的組合。

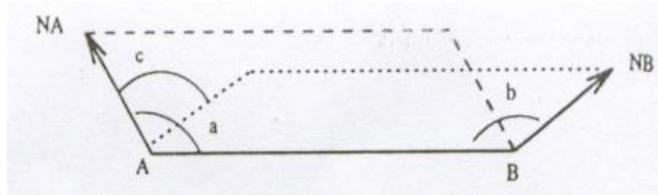


圖 2-7 對一個兩關鍵點產生識別簽名

Fischer [6] 則是使用 Geometric Hashing 來加快互補形狀的尋找速度，其作法分前處理 (pre-processing) 與識別 (recognition) 兩部份。前處理時，先對兩個要接合的蛋白質上的每個凹凸形狀都先各別產生一個參考格 (reference frame)，再從各形狀點集中，每次取兩點當基底，計算每個點在一開始所產生參考格中的座標值，並在參考格中填入基底的值，如此就有各點相對於每個基底在轉換到初始參考格中的座標位置，在識別步驟，把要比對的座標點拿到互補形狀產生的參考格中查詢，查詢過程是拿座標點投票給初始參考格中相同座標內的基底，最後把獲票最多的基底當作最好的比對結果，重複上面步驟把所有形狀的最佳比對結果找出，再用來作座標轉換計分用。

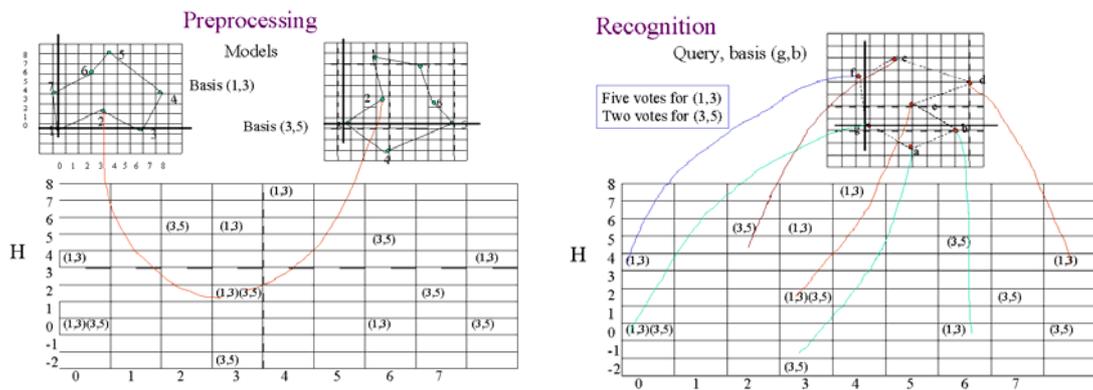


圖 2-8 Geometry Hashing 演算法兩步驟示意圖

另有不明顯針對特定形狀或位置比對的基因演算法 Gardiner [5]，主要概念是把物體剛體轉換的六個自由度 (degree of freedom) 的值，編碼成位元字

串 (bit string)，在競爭選擇 (selection) 過程中利用這些字串換位 (mutation)、交配 (crossover) 以產生新的編碼，重複到到達某個條件後，再還原字串當作比對目的轉換位置，並加入是否有產生氫鍵判斷。Yang [7] 則是把蛋白質的 flexibility 也考慮在內。

2.3 計分

在比對階段產生可能接合的位置與角度後，接著必須靠計分函數將認為可能產生接合的組合篩選出來，因為產生出來的組合數量相當多，所以計分函數的好壞、適合與否影響排名甚鉅。

計分主要是從幾何互補與化學鍵結力量的強弱兩方面去計算，兩個蛋白質的幾何互補程度通常是用彼此間的接觸面積減去重疊體積，詳細作法常會因資料表示方式不同而異，不過概念都是相近，幾何互補的加、扣分評量可分成三種：接觸表面積 (surface contact)、分子間體積重疊 (intermolecular overlap) 與分子內體積重疊 (intramolecular overlap)，最後一項是當蛋白質要處理可折疊特性 (flexibility) 才需要加入考慮。

我們用 Norel et al. [11] 的方法來說明 (圖 2-9)：先把原子依在使用 Connolly 方法下，是否會產生表面點來分成外部與內部原子，如果該原子會在該方法下產生表面點則視為一個外部原子，否則就當成內部原子，再區分出內、外部原子後，把較大的蛋白質擺到一個解析度為 1 埃的三維格子裡，把每個內部原子除模擬凡得瓦半徑外，再加上一個探索球半徑下，整個體積所佔的格子標示成 I，接著在把外部原子僅用凡得瓦半徑模擬大小，把所佔的表格標示成 E，若遇上已經標示為 I 的就改成 E，在把由 Connolly 方法產生的表面點所佔格子標示為 S，遇 E 改 S，計分時，把要接合蛋白質轉換到位置上，套用 $S-4E-10I$ 算出彼此接觸面積減去距離分子表面不同程度重疊的扣分當分數。

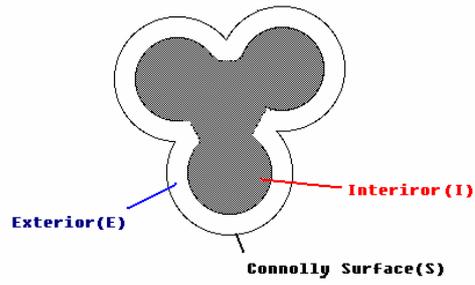


圖 2-9 按接觸表面與距離表面不同的程度來計分

而分子內體積重疊扣分用於在處理 flexibility 會把蛋白質拆成一段一段 (fragments)，需要對這些片段的折疊過程判斷是否彼此間有重疊狀況而另設。



第三章 產生蛋白質網格資料與接合

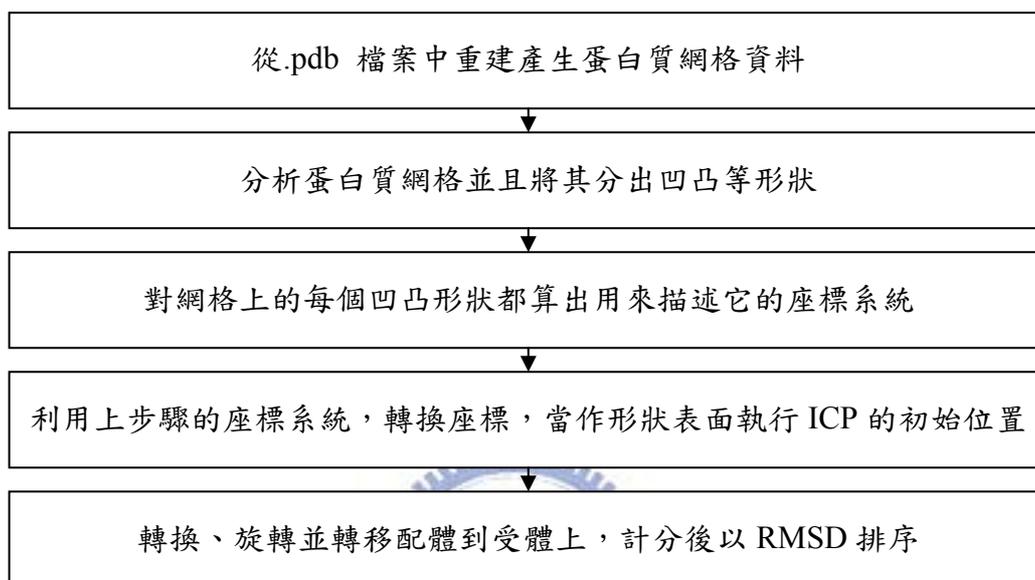


圖 3-1 產生蛋白質網格與接合流程圖

在本章將介紹我們採用的方法，首先產生蛋白質網格資料並作初步處理，再對網格的表面形狀區分成凸凹兩類後對其計算出曲率最大與最小方向所在的兩軸加上形狀表面中心點的法向量，以此三正交向量當作轉換座標系的基底，把轉換後的座標當作 ICP (Iterative Closet Point) 找表面匹配 (surface matching) 的初始位置與角度，經由這一連串的轉換後對每組可能的接合作計分，比對原始接合算出 RMSD (Root Mean Square Deviation) 並且排序結果。

3.1 產生與處理網格資料

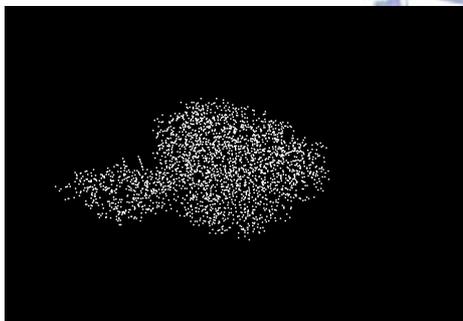
我們所使用的蛋白質網格資料是由 PDB 上蒐集、發布的分子生物三維結構資料檔案所擷取產生。PDB 是一個專門收集並提供網路下載服務的蛋白質資料庫，網站上面的蛋白質結構檔都是由各實驗室以結晶繞射或核磁共振的儀器所產

生出來的原子三維座標資料，並紀錄成.pdb 標準格式，.pdb 檔除了記載蛋白質內含原子的座標值外（ATOM 欄位）還有與其相關實驗細節說明（REMARK 欄位）等，以下是 PDB 與其檔案格式說明的網址：

<http://www.rcsb.org/pdb/>

http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html

這部分的網格資料是取自學長〔13〕先前的成果。首先把.pdb 檔案內的原子座標轉換到一個 128x128x128 的三維立體空間（volume）中，由於所用的實驗資料沒有長、寬、高是大於這個空間大小的，所以不會有建立不完整的狀況，再為每個原子給個模擬凡德瓦半徑（不管原子種類皆為 2Å）作用力的亮度值，亮度由原子中心最大值為 255 往外成距離反比遞減，把所有原子都依照這個亮度計算公式填入三維立體空間後並正規化至 0~255，最後再用電腦圖學中的方塊行進法（marching cube algorithm），施以亮度門檻值為 30，把位於三維空間中的立體資料用三角形集合的方式表示出來。



PDB 化合物 ITC 原子分布



重建網格著色結果

圖 3-2 蛋白質網格重建結果

因為蛋白質內部原子可能會排列出空洞，在套用方塊行進法後所產生的網格會造成存在多於一個的連結元件（connected component）的結果，所以在前述資料產生後會稍作處理找出最大的網格集（max connected component），以便接下來的接合判斷使用。

找出最大網格集系先從原始產生網格集合內任意取出一個三角形，再以它的三個邊為初始邊界開始向外貼補，直到目的網格沒有可以再需要再被貼補，或是原

始網格集合內的三角形全部貼補完畢停止，最後再把所得到的最大的網格集合當作接合的網格資料用。

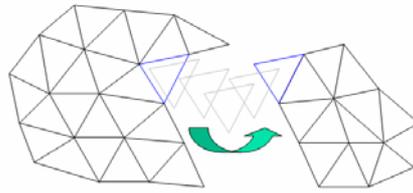


圖 3-3 從原始網格資料藉由貼補前處理，找出最大的元件

儲存、操作網格資料所使用的是 CGAL (Computational Geometry Algorithms Library) [16] 框架函式庫所提供的資料結構，CGAL 是一個科學計算框架，主要可以用來處理幾何演算法上的問題，相關參考網址如下：

<http://www.cgal.org/>

我們使用其中的 Half-Edge 資料結構來表示網格結構，它使用一組 Half-Edge 來表示一條線段，可以藉由一個已知的 Half-Edge 來快速存取同組的反向 Half-Edge 與各自所指向的點，即形成一條線段的兩個端點，也可以在給定一個點的情況下，把與這個點相臨的點按一個次序找出，這個方向特性將有助於我們對網格資料存取操作的方便性，這個特性的建立是在以 Half-Edge 為基礎的網格結構的建構過程，每個面必須按照固定的順序 (compatible) 加到多面體的資料結構裡。

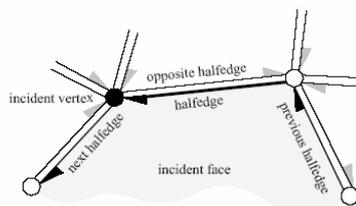


圖 3-4 Half-Edge 資料結構示意圖

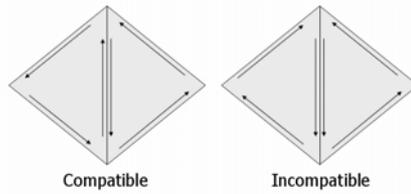


圖 3-5 CGAL 中利用 Half-Edge 產生網格結構的規定順序

在此，我們先定義一些在使用方法與實驗過程中，所使用且與這個資料表示方法有關的名詞與操作。

點 (vertex)：構成蛋白質表面點座標值中的任何其一。

網格表面距離 (mesh geodesic distance)：網格上兩點，從一端到達另一端所必須經過的最少點數。

FLOOD (vertex, N)：給一個點，把網格表面距離小於某個 N 值內的點收集起來成一集合。

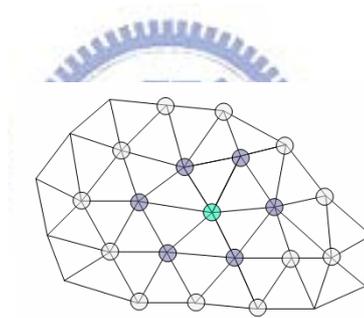


圖 3-6 給一個點 v (圖中藍色點)，FLOOD ($v, 2$) 結果

3.2 接合

3.2.1 網格形狀分析

從.pdb 檔案的化合物擷取出大小蛋白質網格資料(受體 receptor、配體 ligand)後，我們會先把網格上這些點區分成三類：Knob、Hole 與不屬於這兩類的一類，所用的分類概念先對網格表面上每點去 FLOOD 一個給定的值範圍內的點集，統計點集的法向量(從方塊行進法離散化計算所得到)方向與其重心的關係來決定為何種形狀類別，若點的法向量大部分指離重心則為紀錄成一個 Knob，若大部

分指向重心則紀錄為一個 Hole 形狀，否則歸類成不用來做互補形狀匹配的第三類。

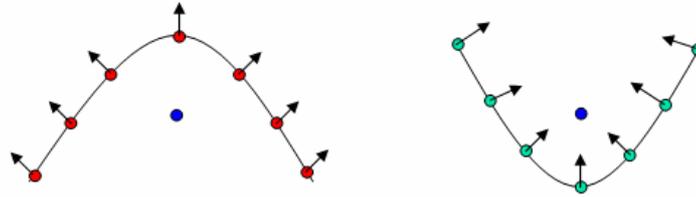


圖 3-7 分析點集法向量與其重心以區分形狀

在幾何的完整定義上，應該會有凸部(convex)凹部(concave)與鞍部(saddle)三類形狀被找出，在我們的形狀分類下，凸部會對應到 Knob，凹部會對應到 Hole，而鞍部則是會對應到不屬於這二類或是 Hole 的機會比較高，接合時為分別列舉配體上的 Knob 與 Hole 形狀對應受體的 Hole 與 Knob 作互補形狀匹配計算的，不屬於 Knob 與 Hole 形狀的表面則不列入計算範圍內。

3.2.2 表面匹配



在受體與配體的網格找出這兩類的形狀後，我們用 ICP 演算法〔12〕來作網格上互補形狀表面匹配，ICP 是一個用來作兩物體間校準 (registration) 與對齊 (alignment) 的方法，藉尋找兩物體間距離最近的對應關係，計算出一組最小平方的轉換矩陣，迭代縮小兩物體間的 MSD (mean square distance)，最後算出最好的轉換矩陣。

雖然 ICP 會計算出一個收斂到最佳值的轉換矩陣，但這個轉換矩陣是一個區域最佳值 (local minimum)，並不保證是整體最佳值 (global minimum)。所以在 ICP 原始論文中建議，可以對所要校準的物體作不同角度的旋轉與轉移後再執行校準，把這些經過不同幾何轉換後的物體拿來作 ICP 計算後，所以可以得出最小 MSD 值的解，就是所要逼近整體最佳值的解。

為了減少替所要教準物體產生不同的旋轉與轉移的數目，所以我們先為受體

與配體中每個分析出的 Knob 與 Hole 形狀，以每個形狀表面的中心點為原點並在其上算出一個以法向量和主曲率方向兩軸所構成的座標系統，利用這個座標系統來轉換蛋白質的原子座標，當作原本配體表面在執行 ICP 教準前，所要設定的不同角度旋轉和轉移。

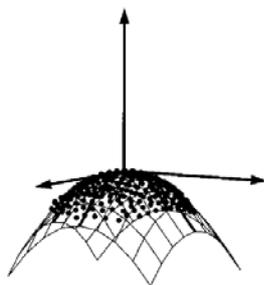


圖 3-8 以表面中心點為原點，法向量與兩主曲率方向所成的座標系統

3.2.3 主曲率方向

用來轉換蛋白質原子位置的座標系統是以從方塊行進法所到的法向量，加上形狀表面上主曲率所在的兩軸所構成，假設表面中心點的法向量為 n (圖 3-9)，經過中心點的切平面 (tangent plane)，則會有無數個包含向量 n 且垂直於切平面的法平面 (normal plane) 會對這個表面交出無限多的正剖面 (normal section)，而對這些正剖面計算出正曲率 (normal curvature)，最大和最小值就是這個表面的主曲率 (principal curvatures)，切出這兩個曲率的剖面方向就是我們用來建立座標系統的兩軸。

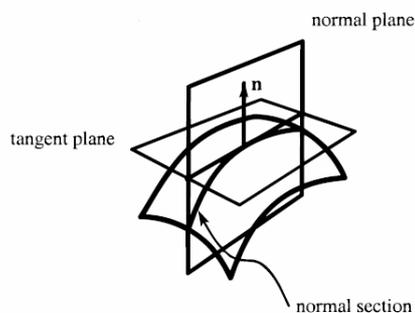


圖 3-9 對一個表面計算正剖面 (normal section) 與正曲率 (normal curvature)

藉由產生主曲率的這兩個正剖面可以建出四個右手則系統，先找出一個右手

則的座標系統，假設為 (p, q, n) ，依序可以找出另外三個右手則座標系為 $(p, -q, n)$ 、 $(-p, -q, n)$ 與 $(-p, q, n)$ (圖 3-10)。

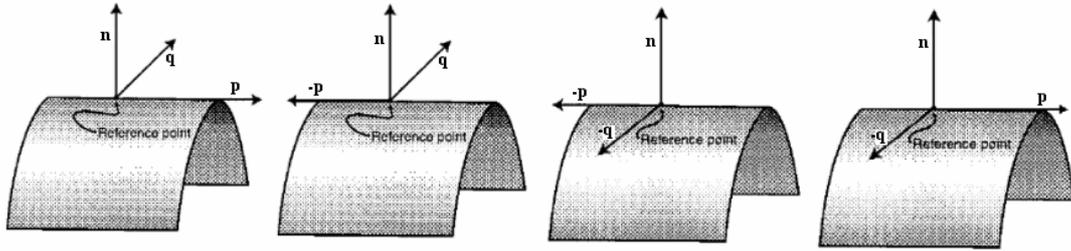


圖 3-10 利用中心點上的法向量與兩個主軸方向，建立四個右手則的座標系統

計算的方法是參考 Goldman [2]，在表面的中心點計算一個 Hessian matrix 以描述在此一表面的區域曲率變化狀況，以下先對 Hessian matrix 作簡單介紹。

對一個有 n 維變數 $(x_1 \sim x_n)$ 的函式 f ，可以定義其 Hessian matrix 如下：

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

其中 $\frac{\partial f}{\partial x_i}$ 表 f 在 x_i 方向上的偏微分。

假設 $\mathbf{A} \in F^{n \times n}$ ， $\lambda \in F$ ，若 $\exists \mathbf{X} \in F^{n \times 1}$ ， $\mathbf{X} \neq \mathbf{0}$ 使得 $\mathbf{A}\mathbf{X} = \lambda\mathbf{X}$ ，則稱 λ 為 \mathbf{A} 相對於 \mathbf{X} 的特徵根 (characteristic value or eigenvalue) 且稱 \mathbf{X} 為 \mathbf{A} 相對於 λ 的一個特徵向量 (characteristic vector or eigenvector)。

Hessian matrix 中所計算出的特徵根以及特徵向量有以下的幾何意義：

特徵根的大小代表其對應特徵向量方向的曲率，所以絕對值最小 (大) 的特徵根所對應的特徵向量的方向有最小 (大) 曲率。

這些特徵根被稱為主曲率 (principal curvature)，而對應的特徵向量被稱為主軸方向 (principal direction)。

主軸方向兩兩垂直 (orthogonal)。

也就是說，一函數上某個點的 Hessian matrix，其特徵根與特徵向量可以描述該點的局部形狀 (local shape)。

以我們所要解決的問題為例，給定一個表面要描述出它的形狀，可以定義它的 Hessian matrix 如下：

$$\mathbf{H} = \begin{bmatrix} \partial^2 w / \partial u^2 & \partial^2 w / \partial u \partial v \\ \partial^2 w / \partial u \partial v & \partial^2 w / \partial v^2 \end{bmatrix}$$

w 是 u 、 v 座標系中計算一個點到切平面上的高度函數 (height function)，我們參考 [2]，取用 w 的二次泰勒展開式 (Taylor expansion) 中，計算 Hessian matrix 裡所需要的項次，再用表面上的點契合 (fitting) 算出各項的值。

$$\frac{1}{2} \left(\frac{\partial^2 w}{\partial u^2} u^2 + 2 \frac{\partial^2 w}{\partial u \partial v} uv + \frac{\partial^2 w}{\partial v^2} v^2 \right)$$

以最小平方契合法，將上述式子轉換成解下列代數系統的解：

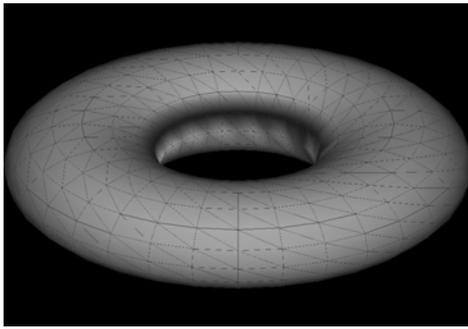
$$y = \mathbf{X}\beta$$

$$\beta = \begin{bmatrix} \partial^2 w / \partial u^2 \\ \partial^2 w / \partial u \partial v \\ \partial^2 w / \partial v^2 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \frac{1}{2} u_1^2 & u_1 v_1 & \frac{1}{2} v_1^2 \\ \vdots & \vdots & \vdots \\ \frac{1}{2} u_n^2 & u_n v_n & \frac{1}{2} v_n^2 \end{bmatrix}, \quad \text{且 } y = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

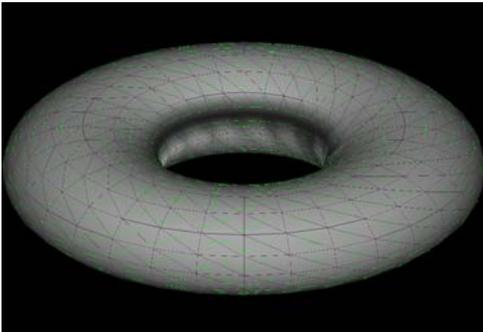
\mathbf{X} 與 y 是以表面中心點為原點，相對 u 、 v 座標系中的轉移座標與相對於此中心點切平面的高度，皆為已知，乘上轉置矩陣並求出反矩陣，最後算出 β 。

$$\beta = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'y$$

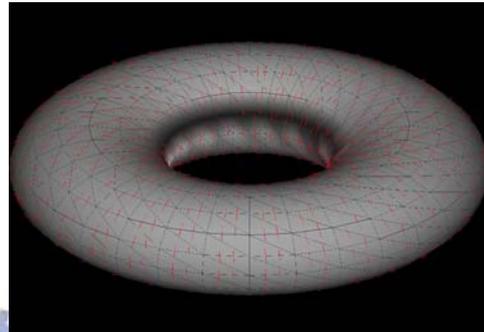
算出 Hessian matrix 的特徵根與特徵向量後，再從 u 、 v 轉換回原座標系中，得到的大、小曲率所對應的兩軸，加上法向量即可用來建立作執行 ICP 前的四個轉換座標系統。



(a)



(b)



(c)

圖 3-11 (a) torus.obj (b) 主曲率較小方向 (平行圓周) (c) 主曲率較大方向 (垂直圓周)

3.2.4 座標轉換

為了節省電腦計算的時間，所以轉換時都是把資料量較小的配體當轉換目標，所以會為配體上分析出的形狀都建立四個座標系統，而受體只會隨選一個座標系統當作被轉換目的，如果受體也建立多個座標系統來轉換，將會產生額外的計算。

假設所要轉換的配體與受體形狀的法向量與主軸各為 $(\vec{p}_1, \vec{q}_1, \vec{n}_1)$ 與 $(\vec{p}_2, \vec{q}_2, \vec{n}_2)$ 則在目標座標系任何一點要變換到目的座標系統中可藉由座標變換公式達成，其一般化的定義如下：

假設 $\alpha = [\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n]$ 與 $\beta = [\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n]$ 為空間 V 中兩組有序基底，則

$\forall \vec{v} \in V, [\vec{v}]_{\alpha} = [I_v]_{\alpha}^{\beta} [\vec{v}]_{\beta}$ ，其中 $[I_v]$ 為 n 階方陣且等於 $[[\vec{a}_1]_{\beta}, \dots, [\vec{a}_n]_{\beta}]$ 稱從 α 到 β 的

轉移矩陣， $[\bar{a}_i]_\beta$ 為把 α 中第 i 個基底寫成 β 基底的線性組合的係數擺成行。

上述提到我們只建立一個受體的座標系統，而要在配體形狀上建立四個提供轉換的座標系，在真正轉換之前必須先把其配體形狀上的法向量反向 (\bar{n}_1')，在維持右手則下依序產生 $(\bar{p}_1, \bar{q}_1', \bar{n}_1')$ 、 $(\bar{q}_1', \bar{p}_1', \bar{n}_1')$ 、 $(\bar{p}_1', \bar{q}_1, \bar{n}_1')$ 與 $(\bar{q}_1, \bar{p}_1, \bar{n}_1')$ 才是轉換的目標基底。

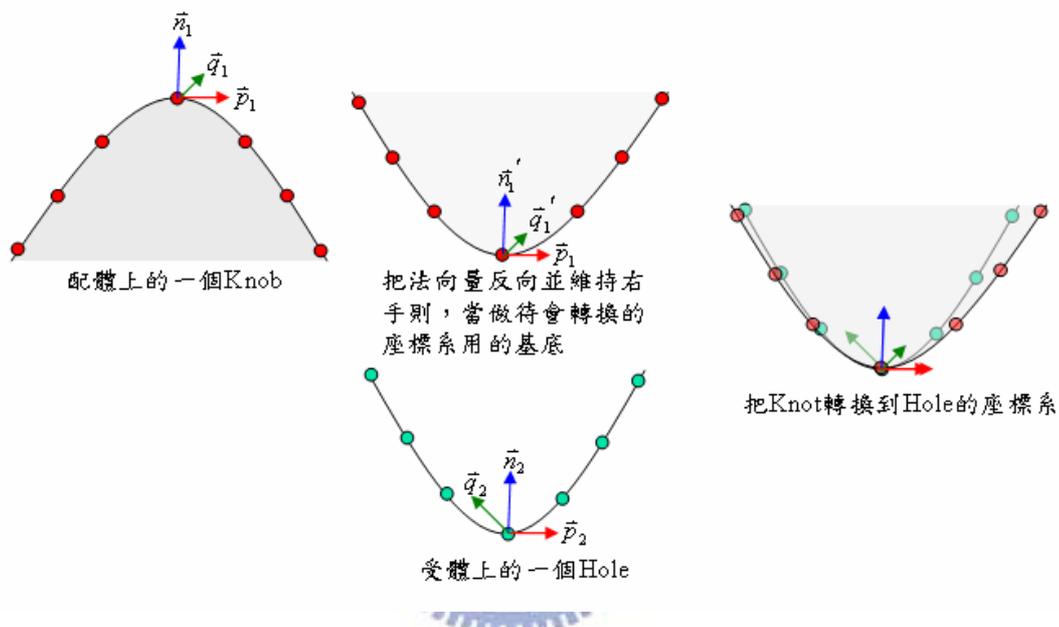


圖 3-12 藉由法向量與主軸構成的座標系統轉換配體座標到受體座標系統中

3.2.5 轉換座標後的表面匹配

每一回拿配體的形狀與受體上的互補形狀做過座標轉換後，我們會以這個位置當作 ICP 校準目標的初始角度和位置，再執行 ICP，執行後所得出的旋轉和轉移矩陣加上先前步驟所計算出的座標轉換套用回配體上所有資料的座標點，把配體旋轉、搬移到算出的可能接合位置，再用計分函數來算分。

3.2.6 計分

幾何互補程度是使用 [11]，以大小蛋白質鍊間的原子接觸面積扣去重疊的原子來計分，並對內部與外部原子的相交，作不同程度的扣分方式。

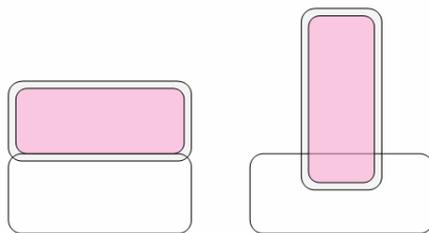


圖 3-13 兩種接合重疊原子數目相近，但應對較靠近內部原子（紅色部分）的相交扣分較多

首先對.pdb 檔案擷取出的每個原子依照距離表面程度標示出外部與內部兩類，標示的方法是利用方塊行進法所產生的網格表面協助標示，對每個.pdb 原子檢查如果有任何一個網格表面的點距離小於某個設定值，則把該原子標示成外部原子，如果全部檢查過的最小距離沒有小於這個設定值的就標示成內部原子，計分時對內、外部原子採不同比重的扣分。

由於計算受體與配體之間的接觸面積，是用方塊行進法產生處理過的表面點，取樣數大於.pdb 檔案所擷取的原子數，所以在計分時，我們會把用來計算重疊相交的.pdb 檔案原子乘上一個常數來調整兩者間的差距，以避免接合情況是重疊程度很嚴重的，卻產生較高的分數。

第四章 實驗結果與討論

本章將介紹程式撰寫與執行所用的環境、實驗資料、設定的參數與所產生的接合結果，並討論之。

電腦環境為 Intel 2.8GHz、Microsoft Windows XP Professional 作業系統，程式是以 C++ 寫成 (Microsoft Visual Studio .NET 編譯環境)，搭配 CGAL 函式庫作為蛋白質網格儲存、操作的資料結構，矩陣運算用 Newmat C++ matrix library [17]，並以 OpenGL 將結果繪出。

4.1 實驗流程與參數設定

所使用的實驗資料如下表所列：

表 4-1 實驗之化合物資料

	PDB	Receptor name	Ligand name	Res.(Å)
1	1cho	Alpha-chymotrypsin (E)	Turkey 2 ovomucoid third domain (I)	1.8
2	1tec	Thermitase eglin-c (E)	Leech (I)	2.2
3	1tgs	Trypsinogen (Z)	Pancreatic secretory trypsin inhibitor (I)	1.8
4	2ptc	Beta-trypsin (E)	Pancreatic trypsin inhibitor (I)	1.9
5	2sec	Subtilisin carlsberg (E)	Genetically-engineered N-acetyl eglin-c (I)	1.8
6	2sni	Subtilisin novo (E)	Chymotrypsin inhibitor (I)	2.1
7	2tgp	Trypsinogen (Z)	Pancreatic trypsin inhibitor (I)	1.9
8	4cpa	Carboxypeptidase	Potato carboxypeptidase a inhibitor (I)	2.5
9	4hvp	HIV-1 protease chain A	Chain B	2.3
10	4sgb	Serine proteinase (E)	Potato inhibitor pci-1 (I)	2.1
11	4tpi	Trypsinogen (Z)	Pancreatic trypsin inhibitor (I)	2.2
12	2mhb	Hemoglobin α chain (A)	B-chain (B)	2.0
13	2kai	Kallikrein a (AB)	Bovine pancreatic trypsin inhibitor (I)	2.5
14	1fdl	IG*G1 fab fragment (LH)	2-lysozyme (Y)	2.5
15	2hfl	IG*G1 fab fragment (LH)	Lysozyme (Y)	2.5
16	3hfm	IG*G1 fab fragment (LH)	Lysozyme (Y)	3.0

使用這 16 個化合物當實驗資料，受體為較大蛋白質鍊，配體為計算出轉換

與校準結果後要做旋轉與轉移的較小蛋白質鍊，括號內的英文單字為蛋白質鍊在 PDB 檔案中的代號，Res. 是當初化合物結晶繞射解出結構的解析度。

這些蛋白質鍊都是已經接合形成化合物的形式，相較於尚未形成化合物前的各單一蛋白質鍊，更接近真實接合的形狀，所以使用這些已形成化合物的蛋白質鍊通常可以算出較好的互補結果。實驗的流程是把化合物中的蛋白質鍵分開，建立網格資料後，用我們提出的方法將其接合並計分，再把這些結果拿來比照真正接合位置，算出 RMSD 排序後，最小值就是所得的最佳解。

實驗流程中所使用到的一些輔助參數：

表 4-2 程式相關參數

參數 (說明)	設定值
SHP_ANA_FLOOD_N (分析點形狀時，向外搜尋距離)	5
TO_KNOB_RATIO (視為一個凸點的法向量最小指離重心比率)	0.9
TO_HOLE_RATIO (視為一個凹點的法向量最小指向重心比率)	0.6
MIN_SHP_VERTICES (構成表面的最少點數)	20
ICP_INIT_POS_FLOOD_N (由表面中心，初始 ICP 轉移的距離的點)	1
OUTER_PENETRATE (相交一個外部原子的扣分數)	10
INNER_PENETRATE (相交一個內部原子的扣分數)	100

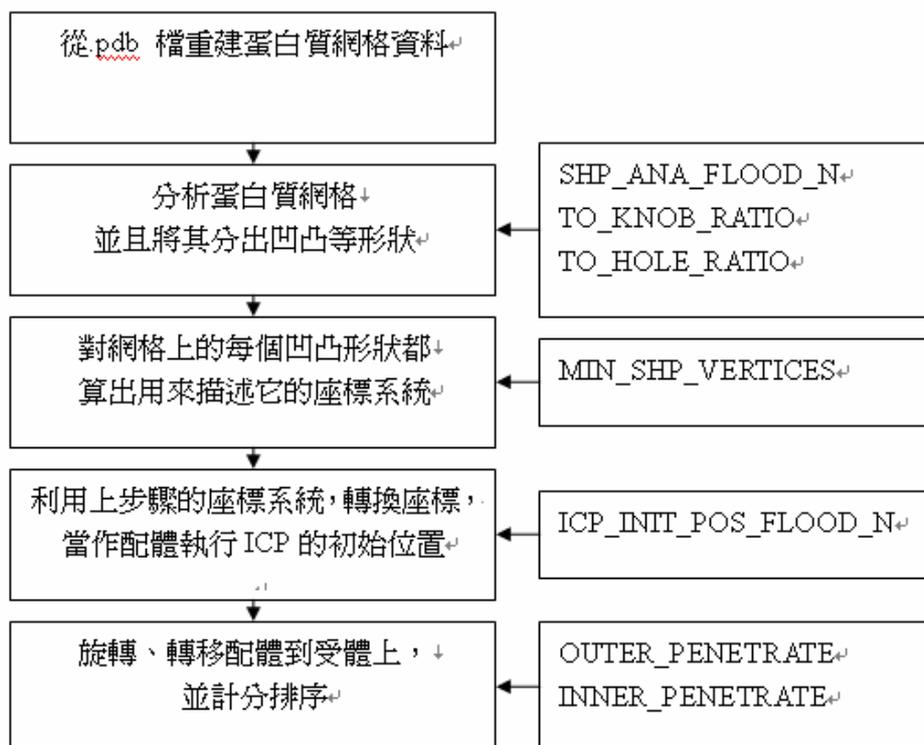


圖 4-1 實驗流程使用輔助參數

上圖是實驗流程中使用輔助參數的時機，我們利用 SHP_ANA_FLOOD_N、TO_KNOB_RATIO 與 TO_HOLE_RATIO 來控制在蛋白質網格上分析區域的大小與標示為一個凹或凸點的難易門檻值，再對每個找出的形狀引入 MIN_SHP_VERTICES 參數避免有些形狀包含的點數過少，ICP_INIT_POS_FLOOD_N 則是設定在受體上向外的一個範圍，可以讓配體有較多的初始轉移位置，計分時則使用 OUTER_PENETRATE 與 INNER_PENETRATE 對不同部位的原子相交作不同程度扣分。

在對重建好的網格做表面形狀分析時，是對網格上每個點去標示出它和附近點所構成的形狀是一個凸、凹或不屬於這兩類，對每個點做 FLOOD (vertex, SHP_ANA_FLOOD_N) 操作，使用 SHP_ANA_FLOOD_N 去調整分析範圍的大小，TO_KNOB_RATIO 與 TO_HOLE_RATIO 兩個門檻值判斷分析這範圍屬於那種形狀。演算法如下：

For each vertex on proteins

FLOOD (vertex , SHP_ANA_FLOOD_N) 得到點集

if(TO_HOLE_RATIO < 點集中的點，指向重心的比例)

把這個 vertex 標示為 hole

else if(TO_KNOB_RATIO < 點集中的點，指離重心的比例)

把這個 vertex 標示為 knob

else

把這個 vertex 標示為 null

為兩個蛋白質表示網格標示完每個點形狀後，再用廣度搜尋法 (BFS) 把同被標示為凹或凸相鄰點分出，相較於概念所述，原本形成一個凸或凹狀的邊界點，會因為接近形狀的分界處，較不容易被標示成一個凸或凹點，而成 null，所以這樣實際分出的形狀區域會較概念所述小的傾向。

在分類出來的形狀裡，凸的部分通常較凹的部分明顯，原因是凹處必須是由多個原子所排列出所形成的一個凹洞所產生，而這些構成這些原子的點在分析時也可能會被視為一個凸點，因此會擠壓到凹處產生的機會，而且就算形成凹處，所構成的點數通常也會比較少，所以我們使用 MIN_SHP_VERTICES 參數將原本形成凹處的點往外擴張加入一些在周圍的點，目的是使凹處更明顯，使接下流程有較多的點可以契合計算主曲率方向。

在計算出供轉換的座標系後，除了以兩形狀表面的中心點當作座標系統原點轉換外，另外引入參數 ICP_INIT_POS_FLOOD_N 在受體形狀表面的中心點作 FLOOD (vertex , ICP_INIT_POS_FLOOD_N) 操作，把這操作所收集到的點集也當轉換的目標原點，當作配體執行 ICP 運算前初使轉移的座標位置，以彌補形狀分析的不夠準確。

每個形狀有了轉換的座標系統與經過 ICP 計算的轉換矩陣，我們會把配體的所有資料座標 (計算接觸分數的網格表面與扣分的.pdb 原子)，也套用這些轉換

的方法把配體轉換到互補形狀的位置上，計算兩者間的互補程度為網格接觸點數減去.pdb 外部原子相交乘 OUTER_PENETRATE 與.pdb 內部原子相交乘 INNER_PENETRATE 的和，再以 RMSD 排序。

整個實驗流程的演算法如下：

Do shape analysis on proteins and BFS to cluster their shape sets

For each surface in receptor

Calculate it's principal directions and setup it's coord. system

For each complementary surface in ligand

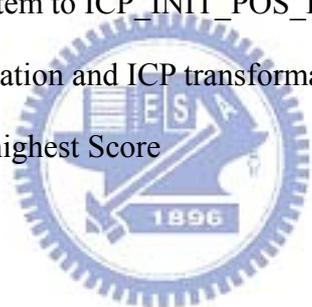
Calculate it's principal directions and setup it's coordinate systems

Translate coord. system to ICP_INIT_POS_FLOOD_N points and do ICP

Do coord. transformation and ICP transformation to ligand

Score and save the highest Score

Sort by RMSD



4.2 實驗結果

表 4-3 (MIN_SHP_VERTICES = 20, ICP_INIT_POS_FLOOD_N = 1)

PDB	排名	最小 RMSD	計算時間
1CHO	1 / 4296	1.796281	02:01:05
1TEC	13 / 5233	2.806800	02:49:56
1TGS	702 / 3249	3.827787	01:58:42
2PTC	546 / 3658	3.043766	01:56:07
2SEC	4 / 4893	2.324402	03:02:43
2SNI	301 / 5134	3.346792	02:46:49
2TGP	426 / 3113	4.023614	01:48:26
4CPA	77 / 3963	2.533531	01:48:55
4HVP	2 / 2781	2.274334	01:30:47
4SGB	1 / 2464	3.198337	00:55:50
4TPI	4 / 3394	2.011618	01:56:15
2MHB	42 / 5656	1.298471	03:52:08
2KAI	289 / 3152	4.630289	02:16:10
1FDL	1 / 12476	2.695780	13:15:08
2HFL	28 / 11963	3.461358	13:42:18
3HFM	2 / 12221	2.395157	14:50:28

我們發現在這組參數下，名次超出百名外的都是靠配體上一個比較明顯凸處嵌入在受體的凹處，把形成一個表面的最少點數參數取的太小，會造成受體與配體的匹配表面包含點數差距太大（因為我們所分類的凹處點數通常都比較少），在把 MIN_SHP_VERTICES 參數調大 4 倍後（20 調到 80），讓互補形狀上重疊點數變多且更接近真實接合狀況，結果就有明顯改善（表 4-4），但這樣會使每個表面點數變多，在計算契合運算因為需要求解較大維度的反矩陣，而讓整個執行時間變的較長。

表 4-4 (MIN_SHP_VERTICES = 80, ICP_INIT_POS_FLOOD_N = 1)

PDB	排名	最小 RMSD	計算時間
1TGS	1 / 3291	1.156368	05:13:19
2PTC	33 / 3789	1.340632	05:22:11
2SNI	227 / 5099	2.650388	07:04:54
2TGP	4 / 3332	1.767239	04:29:49
2KAI	9 / 3262	2.436961	04:34:09

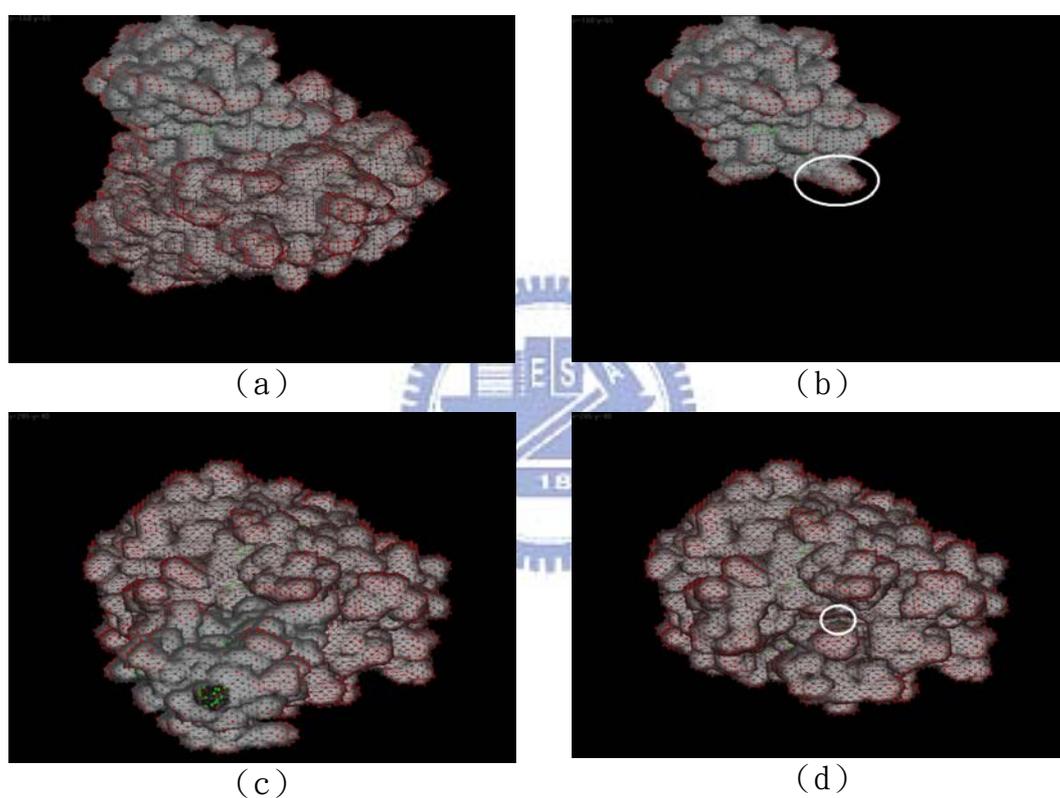


圖 4-2 對 2TGP 化合物中兩個蛋白質，分別標示出接合位置的形狀匹配表面

- (a) 從配體的角度來觀察接合形狀
- (b) 把受體從 (a) 移去，在配體上標示接合位置
- (c) 從受體的角度來觀察接合形狀
- (d) 把配體從 (c) 移去，在受體上標示接合位置

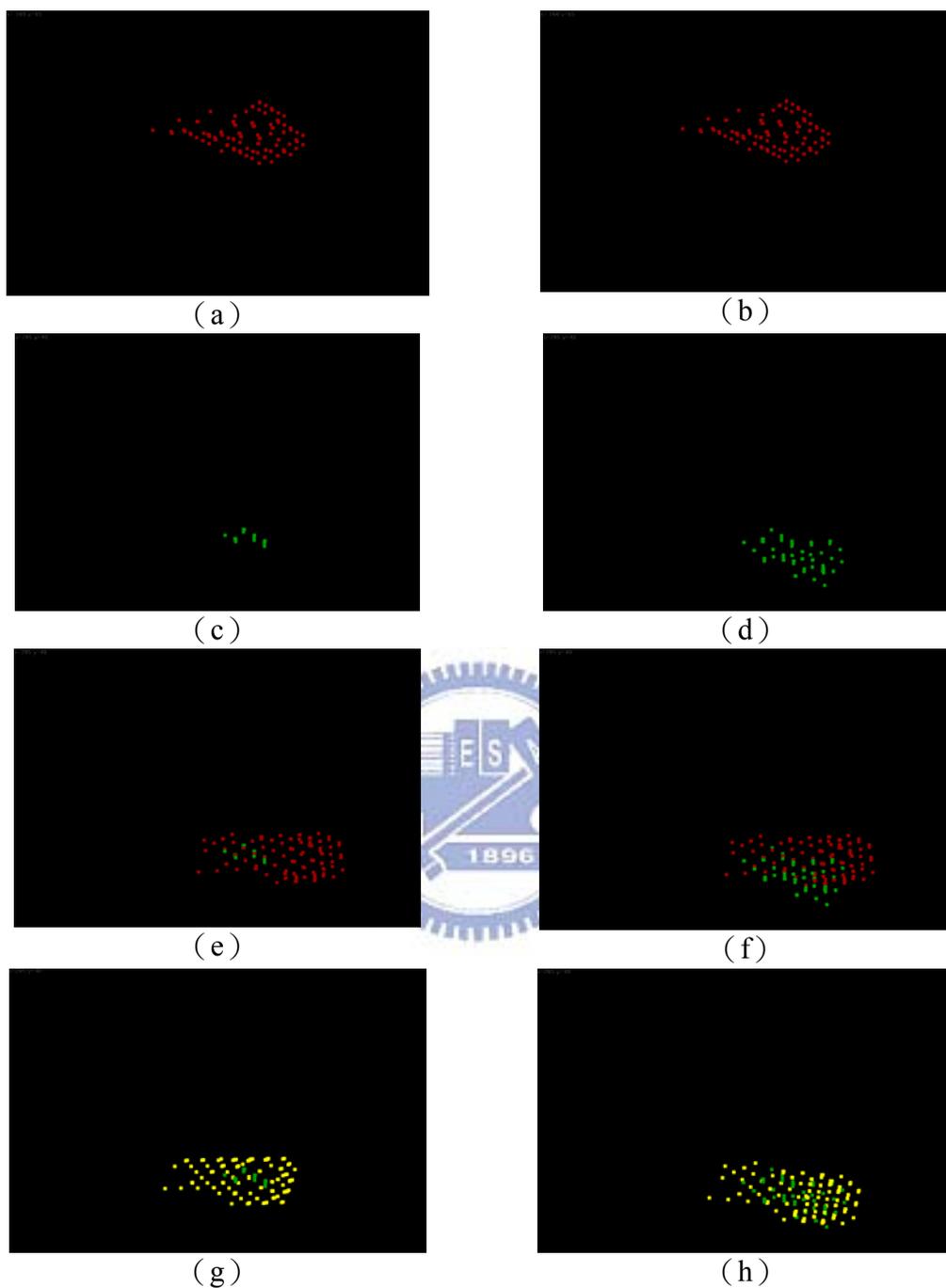


圖 4-3 放大觀察 2TGP 化合物配體與受體不同[參數]的表面構成點集

- (a) 配體在 $\text{MIN_SHP_VERTICES} = 20$ 時的接合位置點集
- (b) 配體在 $\text{MIN_SHP_VERTICES} = 80$ 時的接合位置點集
- (c) 受體在 $\text{MIN_SHP_VERTICES} = 20$ 時的接合位置點集
- (d) 受體在 $\text{MIN_SHP_VERTICES} = 80$ 時的接合位置點集
- (e) 配體、受體在 $\text{MIN_SHP_VERTICES} = 20$ 時的接合點集
- (f) 配體、受體在 $\text{MIN_SHP_VERTICES} = 80$ 時的接合點集

- (a) 配體、受體在 $\text{MIN_SHP_VERTICES} = 20$ 時的接合點集
- (b) 配體、受體在 $\text{MIN_SHP_VERTICES} = 80$ 時的接合點集
- (c) 在 $\text{MIN_SHP_VERTICES} = 20$ 時的接合點集做 ICP 校準後結果
- (d) 在 $\text{MIN_SHP_VERTICES} = 80$ 時的接合點集做 ICP 校準後結果



4.3 討論

在所做的實驗與調整參數中，我們發現形狀分析與決定一個形狀表面需要最少包含多少點會影響結果，假如形狀分析不夠好，連帶也會影響 ICP 應該轉移到的初始位置，透過 ICP_INIT_POS_FLOOD_N 參數雖然可以稍微調整這個轉移位置，可是卻也造成計算量的快速加大。造成這問題的原因，一是我們使用來分析形狀的法向量是方塊行進法離散化計算的結果，一般來說比較不那麼準確，二是對各點做形狀評估方法太過簡單，只是對接合互補形狀的粗略計算。至於應該怎麼決定至少該包含多少點才是最好的情況，這個問題就比較難決定，因為必須事先知道接合位置到底是由多大範圍的表面組合構成，如果接合位置的表面形狀是比較明顯的，我們可以把這個參數調大些，反之則小，以免多加或少給的點形成額外、不足的資訊而影響計算主軸方向與 ICP 匹配，這是這個方法上需要處理的一個問題：怎麼找到一個適當的最小點集構成點數的參數。

我們所用的計分函數，只從幾何互補的角度去計算，在處理幾何互補明顯的情況表現都還可以，在沒有考慮任何化學鍵結下，可以用來篩選一些幾何形狀上比較不可能或可能的組合，當作接下步驟化學鍵結力量強弱的更仔細計算。

第五章 結論、未來工作與展望

藉由各種創新儀器與電腦模擬演算法的幫助，藉由蛋白質結構來了解其功能無論是在藥物設計或疾病篩檢都將提供不少初步的幫助。

本文提出了一個以網格為基礎的接合方法，先是以產生網格演算法的法向量替網格表面分類出凸、凹形狀，並在這些形狀表面算出用來提供座標轉換的座標系統，由前述的法向量與主曲率兩軸所構成，轉換後用 ICP 對這些表面再做一次校準，套用上以幾何互補性為主的計分函數來評量接合程度的好壞，實驗過程中我們發現幾個接合位置由於凸與凹匹配表面構成點數差距過大，需要經由參數調整才能找到一個較好的排名結果，對於抗原抗體（Antibody-Antigen）因為整體上的形狀互補仍很高，所以排名結果也還不錯。

對於我們的方法還有幾點需要改進：

- 形狀分析，此步驟結果會影響後面操作的準確程度。
- 表面最少點數參數，因為我們的形狀分析事先對每個點標示為一個凸或凹點後在用 BFS 把相同且鄰近的點分成同一種形狀，雖然都是形狀表面，但構成點數目會有明顯差異，這個差異會使得轉換座標軸與 ICP 匹配結果與真實接合狀況可能產生較大的差距。
- 計分函數，加入化學計分條件，以符合真實狀況下的其他因素產生的作用力。
- 計算時間，整個程式最耗時的部分就是必須對配體做轉移的運算，轉換成網格表示雖然在執行一些操作相當方便，但整體計算量若沒有把網格做簡化以較少網格表示蛋白質結構，計算量仍過大。

參考文獻

- [1] B. Lee, and Richards, F. M. “Solvent accessibility of groups in proteins.”, *J. Mol. Biol.* 55, 379-400, 1971.
- [2] Brain B. Goldman, W. Todd Wipke “Quadratic Shape Descriptors. 1. Rapid Superposition of Dissimilar Molecules Using Geometrically Invariant Surface Descriptors”, *J. Chem. Inf. Comput. Sci.* 40, 644-658, 2000
- [3] Connolly, M. L. “Solvent-Accessible Surfaces of Proteins and Nucleic Acids.”, *Science* 221, 709-713, 1983
- [4] Connolly, M. L. “Shape complementarity at the hemoglobin $\alpha_1\beta_1$ subunit interface.”, *Biopolymers* 25, 1229-1247, 1986.
- [5] Eleanor J. Gardiner, Peter Willett, and Peter J. Artymiuk “Protein docking using a Genetic Algorithm.”, *Proteins: Struct. Funct. Genet.* 44, 44-56, 2001.
- [6] Fischer, D., Lin, S. L., Wolfson, H. J. & Nussinov, R. “A geometry-based suite of molecular docking processes.”, *J. Mol. Biol.* 248, 459-477, 1995.
- [7] Jinn-Moon Yang “Development and Evaluation of a Generic Evolutionary Method for Protein-Ligand Docking”, *Journal of Computational Chemistry*, 843-857, 2004
- [8] Klaus P. Peters, Jana Fauck and Cornelius Frömmel “The Automatic Search for Ligand Binding Sites in Proteins of Known Three-dimensional Structure Using only Geometric Criteria”, *J. Mol. Biol.* 256, 201-213, 1996.
- [9] Lin, S. L., Nussinov, R., Fischer, D. & Wolfson, H. J. “Molecular surface representation by sparse critical points.”, *Proteins: Struct. Funct. Genet.* 18, 94-101, 1994..
- [10] Norel, R., Lin, S. L., Wolfson, H. J. & Nussinov, R. “Shape complementarity at protien-protein interfaces.”, *Biopolymers* 34, 933-940 1994.

- [11] Norel, R., Lin, S. L., Wolfson, H. J. & Nussinov, R. “Molecular surface complementarity at protein-protein interfaces: The critical role played by surface normals at well placed, sparse, points in docking.”, *J. Mol. Biol.* 252, 263-273, 1995.
- [12] 蔡明倫,「三度空間腦部結構校準」, 交通大學資訊科學所碩士論文, 2002
- [13] 嚴國才,「根據幾何形狀之蛋白質接合研究」, 交通大學資訊科學所碩士論文, 2003
- [14] 葉書豪,「使用最小成本路徑搜尋演算法的血管擷取與分割」, 交通大學資訊科學所碩士論文, 2004
- [15] 林榮耀等編註, 後基因體時代之生物技術, 醫藥基因生物技術教學資源中心, 2003
- [16] “Computational Geometry Algorithms Library,” <http://www.cgal.org/>
- [17] Robert Davies “Newmat C++ matrix library,” <http://www.robertnz.net/>,

