

行政院國家科學委員會專題研究計畫 期末報告

位向距離結構模版應用於蛋白質功能之預測與註解

計畫類別：個別型
計畫編號：NSC 101-2221-E-216-041-
執行期間：101年08月01日至102年07月31日
執行單位：中華大學生物資訊學系

計畫主持人：董其樺

計畫參與人員：碩士班研究生-兼任助理人員：顏子維
碩士班研究生-兼任助理人員：莊利偉

報告附件：出席國際會議研究心得報告及發表論文

公開資訊：本計畫可公開查詢

中華民國 102 年 10 月 31 日

中文摘要： 分子間交互作用的研究對於了解生物機制非常重要，同時也有助於系統生物學和功能性基因體學領域的發展。目前的研究通常以蛋白質功能區域(domain)來研究蛋白質功能及其間的交互作用，蛋白質功能區域通常由結構核心(structure core)和分子交互介面(interface)所構成。然而蛋白質一般是藉由分子交互介面與其他分子（如蛋白質、小分子、金屬離子、DNA)產生交互作用來執行其功能。因此研究蛋白質分子介面可直接了解蛋白質的功能、交互作用及其機制。在後基因體時代，大部分基因或蛋白質的功能仍不明確，探討蛋白質交互作用機制是了解功能基因體學的關鍵，因為蛋白質分子間介面的研究，可用來註記蛋白質功能、訊息傳遞路徑、調控網路及代謝路徑。

本計畫將探討蛋白質與配體之間交互作用環境之局部結構模版在蛋白質功能扮演的角色。我們提出一個考慮空間方向及距離資訊之結構模版(structure-oriented distance motif)之概念，用於描述一個蛋白質與其他分子間(如藥物、蛋白質、金屬離子、DNA等)的交互作用環境。並且利用模版進行搜尋比對時，加強了立體結構疊合的確認，使該結構模版更具正確性。不同於蛋白質功能區域，位向距離結構模版著重於結合介面，而結合介面是執行蛋白質功能的核心所在。研究結果顯示，我們確實能夠為蛋白質結合部位的環境中定位出關鍵而重要的模版。未來我們將應用此一技術，進行蛋白質片段-藥物-疾病之三相網路的建立，探討蛋白質結構、藥物與疾病之關聯性。

中文關鍵詞： 結構模版、蛋白質-配體結合環境、交互作用介面

英文摘要： Studying molecular interactions is useful to understand biological mechanisms and is essential for studying systems biology and functional genomics. Currently, studies of protein domains, which are typical units of protein structures and functions, are derived for elucidating protein functions and interactions. A protein domain is collinear in sequence and often consists of two parts: structural core and molecular interfaces. However, the biological functions of a protein are often performed by interacting with other molecules (such as proteins, small compounds, metals, and DNA) via its interfaces. Studying molecular interfaces of a protein will provide clues to understand protein functions. In addition, functions of most

genes/proteins in genomic projects remain unknown, leading a great challenge to realize molecular interacting mechanisms. Therefore, understanding protein-molecule interfaces will annotate functions of proteins and play the core role to reveal the molecular mechanisms of signaling pathways, regulatory networks, and metabolic pathways. The scientific objective of this project is to address what is the role of structural motif in molecular interface for protein functions. In this project, we propose a new concept of Structure-Oriented Distance motif (SOD motif), which describes the binding environment between a protein and its interacted partners. The SOD motifs contain a set of spatially discontinuous peptide segments, whereas protein domains are continuous in sequences and structures, most of which do not directly involve functional sites. We also enhanced the ability of structure superimposition and refined the searching results of SOD motif. Different from protein domains, molecular interface motifs emphasize on binding interfaces, the cores to perform protein functions. Furthermore, our results showed that SOD motifs can identify the proteins with similar binding interfaces as well as annotate functions. We will apply this technique into research of segment-drug-disease network.

英文關鍵詞： structural motif, protein-ligand binding environment, protein interface

位向距離結構模版應用於蛋白質功能之預測與註解

Structure-Oriented Distance Motif for Protein Functions Prediction and
Annotation

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC 101 - 2221 - E - 216 - 041 -

執行期間：101年8月1日至102年7月31日

執行機構及系所：中華大學 生物資訊系

計畫主持人：董其樺 助理教授

共同主持人：

計畫參與人員：碩士班研究生-兼任助理人員：顏子維

碩士班研究生-兼任助理人員：莊利偉

本計畫除繳交成果報告外，另含下列出國報告，共 1 份：

移地研究心得報告

出席國際學術會議心得報告

國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權，一年二年後可公開查詢

中 華 民 國 102 年 10 月 31 日

研究計畫期末報告目錄

中文摘要.....	I
英文摘要.....	II
報告內容.....	1
前言.....	1
文獻探討.....	1
研究目的.....	1
研究方法.....	2
結果與討論.....	4
REFERENCES.....	5
國科會補助專題研究計畫成果報告自評表.....	13

圖表目錄

表一、69 對 SOD MOTIF 測試資料及其 SOD-SCORE 分數計算.....	8
圖一、SOD MOTIF 之概念圖.....	9
圖二、研究方法架構圖.....	10
圖三、片段決定向量示意圖.....	11
圖四、1AX9_A 及 1AQL_A 之 SOD MOTIF 結構疊合.....	12

中文摘要

分子間交互作用的研究對於了解生物機制非常重要，同時也有助於系統生物學和功能性基因體學領域的發展。目前的研究通常以蛋白質功能區域(domain)來研究蛋白質功能及其間的交互作用，蛋白質功能區域通常由結構核心(structure core)和分子交互介面(interface)所構成。然而蛋白質一般是藉由分子交互介面與其他分子(如蛋白質、小分子、金屬離子、DNA)產生交互作用來執行其功能。因此研究蛋白質分子介面可直接了解蛋白質的功能、交互作用及其機制。在後基因體時代，大部分基因或蛋白質的功能仍不明確，探討蛋白質交互作用機制是了解功能基因體學的關鍵，因為蛋白質分子間介面的研究，可用來註記蛋白質功能、訊息傳遞路徑、調控網路及代謝路徑。

本計畫將探討蛋白質與配體之間交互作用環境之局部結構模版在蛋白質功能扮演的角色。我們提出一個考慮空間方向及距離資訊之結構模版(structure-oriented distance motif)之概念，用於描述一個蛋白質與其他分子間(如藥物、蛋白質、金屬離子、DNA等)的交互作用環境。並且利用模版進行搜尋比對時，加強了立體結構疊合的確認，使該結構模版更具正確性。不同於蛋白質功能區域，位向距離結構模版著重於結合介面，而結合介面是執行蛋白質功能的核心所在。研究結果顯示，我們確實能夠為蛋白質結合部位的環境中定位出關鍵而重要的模版。未來我們將應用此一技術，進行蛋白質片段-藥物-疾病之三相網路的建立，探討蛋白質結構、藥物與疾病之關聯性。

關鍵字：結構模版、蛋白質-配體結合環境、交互作用介面

Abstract

Studying molecular interactions is useful to understand biological mechanisms and is essential for studying systems biology and functional genomics. Currently, studies of protein domains, which are typical units of protein structures and functions, are derived for elucidating protein functions and interactions. A protein domain is collinear in sequence and often consists of two parts: structural core and molecular interfaces. However, the biological functions of a protein are often performed by interacting with other molecules (such as proteins, small compounds, metals, and DNA) via its interfaces. Studying molecular interfaces of a protein will provide clues to understand protein functions. In addition, functions of most genes/proteins in genomic projects remain unknown, leading a great challenge to realize molecular interacting mechanisms. Therefore, understanding protein-molecule interfaces will annotate functions of proteins and play the core role to reveal the molecular mechanisms of signaling pathways, regulatory networks, and metabolic pathways.

The scientific objective of this project is to address what is the role of structural motif in molecular interface for protein functions. In this project, we propose a new concept of Structure-Oriented Distance motif (SOD motif), which describes the binding environment between a protein and its interacted partners. The SOD motifs contain a set of spatially discontinuous peptide segments, whereas protein domains are continuous in sequences and structures, most of which do not directly involve functional sites. We also enhanced the ability of structure superimposition and refined the searching results of SOD motif. Different from protein domains, molecular interface motifs emphasize on binding interfaces, the cores to perform protein functions. Furthermore, our results showed that SOD motifs can identify the proteins with similar binding interfaces as well as annotate functions. We will apply this technique into research of segment-drug-disease network.

Keyword: structural motif, protein-ligand binding environment, protein interface

前言

基因體計畫進行至目前為止已經將超過 160 種原核或真核生物的全基因體定序解碼完成。在這些已定序的基因體中,有很大一部份(百分比約三十至五十)的基因,所表現的蛋白質依舊未能得知其功能為何。為了註解這些未知功能的蛋白質,結構基因體學迫切需要強而有力的研究方法,來進行全基因體之蛋白質的全面性註解或功能預測 (1-3)。結構基因體學計畫的蓬勃發展,產生出越來越多新的蛋白質結構,但卻未能對其功能一一加以註解。而這種蛋白質在蛋白質結構資料庫 Protein Data Bank (4)中所累積的數量,已日益漸增。

另一方面,透過藥物配體結合至蛋白質,可觀察蛋白質內部進行交互作用的胺基酸位置,進而對蛋白質進行功能註解。因此,對蛋白質標的物辨認出其配體結合部位之介面(interface)或模版(motif),再透過相似模版的搜尋比對,將能夠找出可能亦受到相同配體所結合的其他標的物。如此,了解配體與蛋白質之間的結合模版與作用機制,並進一步註解蛋白質功能,甚至能進行舊藥新用之藥物開發研究,實為一個極有研究價值的議題。

文獻探討

針對上述議題,傳統的解決方案通常利用序列或結構搜尋工具,針對要查詢的蛋白質進行相似度比對 (5-7)。根據比對結果,再進一步辨識出結合部位附近的序列組成或局部結構片段相似性。通常結合相同配體的蛋白質多具有相似的結合型樣(pattern) (7)或結構片段。隨著蛋白質結構數量日漸倍增,越來越多的結構證據可用於分析結合部位的模版形狀,並描述結合環境 (8,9)。此外,也有許多相關研究(10-14)針對結合部位的局部結構進行搜尋比對。但這些研究只針對單一一個結合部位進行探索,並未同時考慮多個環繞在配體周遭的局部結構,而這些局部結構往往都是在序列上位於不連續的位置,並且可能因局部結構相似但全域結構較不相似時,使得搜尋上產生偽陰性的結果。此外,已有前人的研究 SRPMOTIF (15),證實利用空間上不連續的局部結構組成模版,並透過模版做快速搜尋比對能夠找出潛在的藥物或舊藥新用。然而,該研究在進行結構比對時,仍需結合額外的結構比對軟體 DALI (16),以程式獨立性來看,恐怕有所受限。

研究目的

本計畫中,我們提出以空間及距離資訊為導向之結構模版(structure-oriented distance motif, SOD motif),用於描述一個蛋白質與其他配體間(如蛋白質、金屬離子、藥物、DNA 等)的交互作用環境。不同於蛋白質功能區域(domain),SOD motif 著重於結合介面,而結合介面是執行蛋白質功能的核心所在。SOD motif 包含一組序列及空間上不連續的胜肽片段(peptide segment)。我們利用在蛋白質與配體結合部位裡的 SOD motif,表現其環繞在結合部位的結構表面之多樣性。進而瞭解蛋白質與配體之間交互作用的機制與蛋白質之功能。

SOD motif 的搜尋，乃是根據多個結合部位延伸出的局部結構，轉換成一維結構字元序列。再透過先前研究 3D-BLAST (17,18)，快速找出符合多個序列上不連續的結構字元序列之候選蛋白質。接著利用我們新開發的多點結構比對演算法，從而決定其中哪些結構字元序列在空間排列之位向相似，以致進一步決定其候選蛋白質之結合環境是否符合原 SOD motif 的形狀。本研究成果可助於我們進一步瞭解蛋白質活性中心之介面，並且深入研究蛋白質功能之註解，我們相信這些資訊對於結構基因體學及生醫製藥應用有相當的貢獻。

研究方法

利用全域結構比對軟體試圖找出結構模版，是需要耗費大量的運算時間。同時，同源結構的數量在不足的情況下，亦不容易決定出結構上具保留性的模版為何。此外，功能性結構模版不但需要是一長段連續的序列，也必須是折疊成片段靠近的局部結構。因此，這類模版的結構容易傾向表現成二級結構的型樣，而導致失去其辨認性。本研究計畫成果裡，我們提出了創新的概念，新定義一種環繞在結合部位周圍的功能性結構模版，命名為 SOD motif。此 SOD motif 的定義如圖一所示。

本計畫之研究步驟以圖二闡述之。首先，第一步為 SOD motif 建立階段。根據蛋白質內某一配體結合之部位為中心，向其序列上兩端延伸，決定數個局部片段。集合起各片段後，利用 3D-BLAST 之方法論，將對應之結構轉換成隱含結構資訊的一級序列。3D-BLAST 之結構字元集，乃是使用 DSSP (19)中所定義的 κ 和 α 角，將結構片段分類並給予特定的結構字元符號。每一個蛋白質結構片段，都可對應到 23 個結構字元的其中一個。此外，3D-BLAST 亦提出了類似於 BLOSUM62 (20)的計分矩陣，用於結構序列比對時，計算局部結構相似程度的分數依據。

第二步，則為 SOD motif 搜尋階段。將數個結構字元片段同時透過 3D-BLAST 進行快速結構資料庫比對，找出同樣擁有絕大部分相似結構片段的其它蛋白質，成為候選蛋白質。3D-BLAST 因結合了 BLAST (21,22)高速比對資料庫的能力，可達到可從結構資料庫中快速搜尋同源蛋白質結構。第三步驟為立體結構疊合確認階段。此步驟利用本研究中新開發的結構疊合演算法，計算並確認在三維空間中，候選蛋白質中各結構字元片段是否與查詢蛋白質的各結構片段結構相似、片段間距離相似、且位向一致。挑選出正確疊合的片段並輸出之，表現其查詢蛋白質與搜尋蛋白質之間的 SOD motif 相對應的結構，藉此說明搜尋蛋白質亦可能結合查詢蛋白質中的配體。

結構疊合演算法，乃是將各個在空間上獨立的結構片段，根據其上每一 C_{α} 之座標平均，決定其質心、以及各質心的質心後，計算出用以代表該 SOD motif 的向量。將查詢蛋白質與候選蛋白質分別計算出代表向量後，以向量作平移與旋轉，以三維方式疊合兩蛋白質的 SOD motif。最後計算其 SOD-score 及其統計數

值，用以決定兩蛋白質之間的 SOD motif 如何挑選正確、相似的片段及位向。

要判斷兩個位於不同空間位置與方向的結構是否相似，必須將其中一個結構旋轉成與另一結構相同的方向，再重疊起來計算其 RMSD 值(Root Mean Square Deviation)，其值愈小，相似度愈高，透過旋轉矩陣公式可以完成結構疊合動作(23)。

如圖三所示，先定義兩結構的旋轉支點與進行外積的兩向量，兩向量經過三次外積後會產生三個互相垂直的單位向量，由出現的先後分別為 X vector、Y vector、Z vector。參考文獻(23)中是以 v_i 外積 v_p ，但在本研究過程中發現 v_p 外積其 v_i 產生的三個垂直向量所旋轉的結果全部優於前人研究結果。外積的第一步如下：

$$X \text{ vector} = ((v_p \times v_i) / (|v_p \times v_i|)) \quad (1)$$

接著以 v_i 和 X vector 進行外積，產生出 Y vector，公式如下：

$$Y \text{ vector} = ((v_i \times X \text{ vector}) / (|v_i \times X \text{ vector}|)) \quad (2)$$

最後以 X vector 與 Y vector 外積出 Z vector，Z vector 會與 v_i 重疊，公式如下：

$$Z \text{ vector} = ((X \text{ vector} \times Y \text{ vector}) / (|X \text{ vector} \times Y \text{ vector}|)) \quad (3)$$

由於形成肽平面(peptide surface)的原子(碳 C、氮 N、氧 O)與此連線有著一定的結構關係才會照成這種結果。

這三個垂直的單位向量是描述結構片段在空間中的方向，當兩相異方向結構要重疊時，只要將其中一個結構的方向旋轉成另一個結構的方向即可，此過程依旋轉矩陣設計的公式進行運算，才能求得結果，公式如下：

$$R_x(\theta) * \text{Vector1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} X1 \\ Y1 \\ Z1 \end{bmatrix} = \begin{bmatrix} X2 \\ Y2 \\ Z2 \end{bmatrix} = \text{Vector2} \quad (4)$$

$$R_y(\phi) * \text{Vector2} = \begin{bmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{bmatrix} \begin{bmatrix} X2 \\ Y2 \\ Z2 \end{bmatrix} = \begin{bmatrix} X3 \\ Y3 \\ Z3 \end{bmatrix} = \text{Vector3} \quad (5)$$

$$R_z(\varphi) * \text{Vector3} = \begin{bmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X3 \\ Y3 \\ Z3 \end{bmatrix} = \begin{bmatrix} X4 \\ Y4 \\ Z4 \end{bmatrix} = \text{Vector4} \quad (6)$$

此時將三個垂直的單位向量(X vector、Y vector、Z vector)先後套入(4)~(6)，當 X vector 經(4)~(6)往另一組的 X vector 旋轉時，其餘的向量包含 Y vector、Z vector 以及所有原子的向量皆須跟著旋轉，接著 Y vector 與 Z vector 經(4)~(6)

往另一組的 Y vector 與 Z vector 旋轉時也是如此。旋轉結束後，將結構片段上的所有原子平移到另一組結構片段上，算出每一組對應原子的距離平方和，接著除以對應原子的組數再全部開根號，其值為 RMSD。再經過轉換，換算成為一個用以衡量兩組 SOD motif 的空間上相似程度的指標，稱做 SOD-score。其轉換公式如下，其值愈大代表相似度愈高：

$$\text{SOD-score} = \frac{1}{1+\text{RMSD}} \quad (7)$$

此外，統計資料庫中 SOD-score 平均值及標準差後，則可將 SOD-score 代入 Z-score 之公式，以求得統計上的可信程度。Z-score 公式如下：

$$\text{Z-score} = \frac{x-\bar{x}}{\sigma} \quad (8)$$

接著要我們欲驗證 Z-score 的門檻閾值(threshold)為何，以及 SOD motif 結構疊合之演算法的有效性，我們自 FDA 資料庫挑選已上市之藥物分子，並找出有紀錄在 PDB 資料庫中之藥物-蛋白質複合物。測試資料中，挑選 12 組藥物-蛋白質複合物共 69 對 SOD motif (如表一所示)。

結果與討論

為了求出最佳組合，挑選正確的局部結構位向進行兩 SOD motif 的疊合，我們目前使用貪婪演算法，針對所有可能的組合分別計算其 SOD-score，再按照其 Z-score 分數差異，進行辨識。表一顯示 69 對 SOD motif 的測試資料，及其原始 SOD-score 為何。以目前研究結果顯示，在 69 對 SOD motif 測試資料中， \bar{x} 為 0.268、 σ 為 0.287，而我們將 Z-score 分數訂定 0.5，表示 Z-score 門檻閾值需大於或等於 0.5，方辨認為正確的結構位向與模版組合。以測試資料 PDB:1AX9_A 與 PDB:1AQL_A 之 SOD motif 的比對為例子，合理的位向組合結構疊合後，所計算出來的 SOD-score 為 0.426，換算成 Z-score 為 0.552。而不正確的局部結構位向組合之下，其 Z-score 則出現 -0.502 至 -0.026。

以圖四為例，1AX9_A (24) 中有一藥物分子 EDR。以此藥物配體分子為中心，決定其蛋白質-配體結合環境周圍的局部結構共有五段。而個別利用 3D-BLAST 去針對 1AQL_A (25) 做快速結構字元比對後，則找回五個局部結構片段。隨後利用貪婪演算法，將所有可能的位向進行組合並且疊合結構，計算出其中第一段局部結構片段，在兩蛋白質之間的結構位向是不合理的，其 Z-score 分別為 -0.434、-0.502、-0.026、-0.498 等值，皆低於閾值。從圖中可觀察到兩蛋白質第一段的局部結構在空間上不吻合，1AQL_A 中第 65 號胺基酸至第 79 號胺基酸之結構片段並未位於結合部位附近，距離 EDR 最近的距離也已超過 12Å，該結構明顯與藥物沒有關聯或鍵結作用力。而其他四段局部結構片段則可完整疊合，且在空間上

彼此距離相近，位向皆圍繞在藥物分子的結合部位附近。此結果顯示，本研究方法可針對 3D-BLAST 的搜尋結果做進一步的分析，確認其局部結構片段在空間中的方位，進而決定 SOD motif 的正確辨認。

然而，由於結構疊合的方法採用貪婪演算法，進行結構比對時的運行時間上，本研究所提出的技術較 DALI 程式慢了五至十倍。在準確度不受影響之下，未來仍需繼續研究，採用更有效率的演算法與運作策略，以改善執行的效率。

最後，本研究能夠針對查詢蛋白質，迅速自蛋白質結構資料庫中搜尋出具有可能的局部結構片段，在經由實際結構疊合，確認其局部片段組合，以建立出以結構位向為基礎的結合環境之模版 SOD motif。研究結果顯示，我們確實能夠為蛋白質結合部位的環境中定位出關鍵而重要的模版。進而用於未知蛋白質的功能確認，以及重要活性中心的註解。未來我們將應用此一技術，進行蛋白質片段-藥物-疾病之三相網路的建立，及其應用。

References

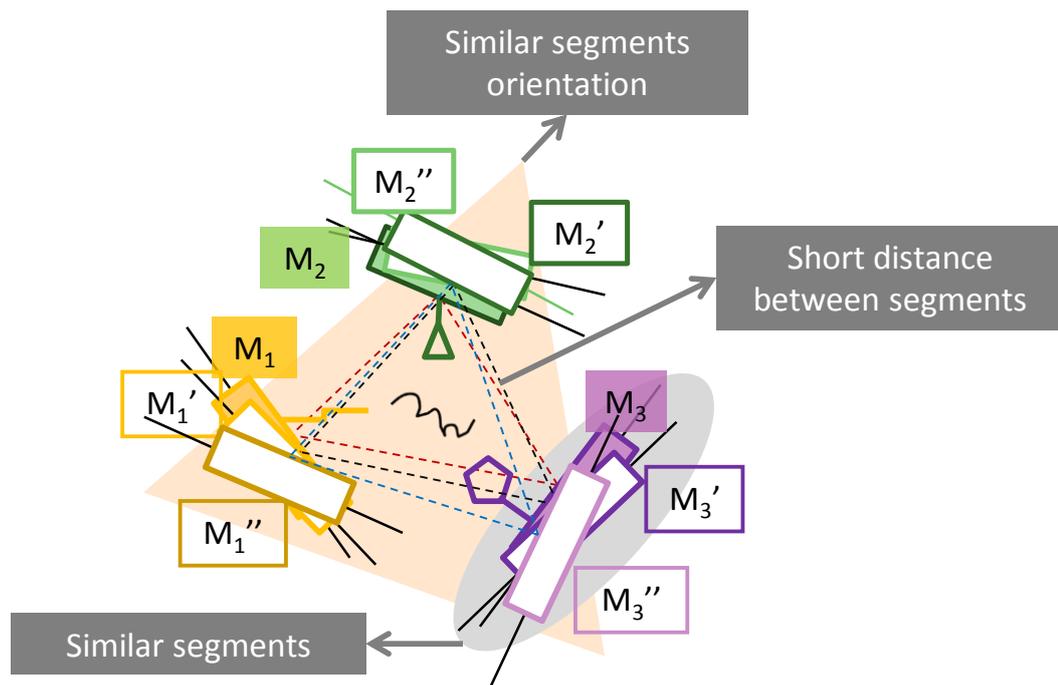
1. Todd, A.E., Marsden, R.L., Thornton, J.M. and Orengo, C.A. (2005) Progress of structural genomics initiatives: an analysis of solved target structures. *J Mol Biol*, **348**, 1235-1260.
2. Burley, S.K. and Bonanno, J.B. (2002) Structural genomics of proteins from conserved biochemical pathways and processes. *Current opinion in structural biology*, **12**, 383-391.
3. Burley, S.K., Almo, S.C., Bonanno, J.B., Capel, M., Chance, M.R., Gaasterland, T., Lin, D., Sali, A., Studier, F.W. and Swaminathan, S. (1999) Structural genomics: beyond the human genome project. *Nat Genet*, **23**, 151-157.
4. Deshpande, N., Address, K.J., Bluhm, W.F., Merino-Ott, J.C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z. *et al.* (2005) The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res*, **33**, D233-237.
5. Ku, S.Y. and Hu, Y.J. (2012) Structural alphabet motif discovery and a structural motif database. *Comput Biol Med*, **42**, 93-105.
6. Pugalenti, G., Suganthan, P.N., Sowdhamini, R. and Chakrabarti, S. (2008) MegaMotifBase: a database of structural motifs in protein families and superfamilies. *Nucleic Acids Res*, **36**, D218-221.
7. Sigrist, C.J., Cerutti, L., de Castro, E., Langendijk-Genevaux, P.S., Bulliard, V., Bairoch, A. and Hulo, N. (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res*, **38**, D161-166.
8. Kinnings, S.L., Liu, N., Buchmeier, N., Tonge, P.J., Xie, L. and Bourne, P.E. (2009)

- Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput Biol*, **5**, e1000423.
9. Weber, A., Casini, A., Heine, A., Kuhn, D., Supuran, C.T., Scozzafava, A. and Klebe, G. (2004) Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J Med Chem*, **47**, 550-557.
 10. Bauer, R.A., Bourne, P.E., Formella, A., Frommel, C., Gille, C., Goede, A., Guerler, A., Hoppe, A., Knapp, E.W., Poschel, T. *et al.* (2008) Superimpose: a 3D structural superposition server. *Nucleic Acids Res*, **36**, W47-54.
 11. Debret, G., Martel, A. and Cuniasse, P. (2009) RASMOT-3D PRO: a 3D motif search webserver. *Nucleic Acids Res*, **37**, W459-464.
 12. Madsen, D. and Kleywegt, G.J. (2002) Interactive motif and fold recognition in protein structures. *J Appl Crystallogr*, **35**, 137-139.
 13. Shulman-Peleg, A., Shatsky, M., Nussinov, R. and Wolfson, H.J. (2008) MultiBind and MAPPIS: webserver for multiple alignment of protein 3D-binding sites and their interactions. *Nucleic Acids Res*, **36**, W260-264.
 14. Wu, C.Y., Chen, Y.C. and Lim, C. (2010) A structural-alphabet-based strategy for finding structural motifs across protein families. *Nucleic Acids Res*, **38**, e150.
 15. Chiu, Y.Y., Lin, C.Y., Lin, C.T., Hsu, K.C., Chang, L.Z. and Yang, J.M. (2012) Space-related pharma-motifs for fast search of protein binding motifs and polypharmacological targets. *BMC genomics*, **13 Suppl 7**, S21.
 16. Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233**, 123-138.
 17. Tung, C.H., Huang, J.W. and Yang, J.M. (2007) Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for rapid search of protein structure database. *Genome Biol*, **8**, R31.31-R31.16.
 18. Yang, J.M. and Tung, C.H. (2006) Protein structure database search and evolutionary classification. *Nucleic Acids Research*, **34**, 3646-3659.
 19. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.
 20. Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 10915-10919.
 21. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.
 22. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and

- Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.
23. Milik, M., Kolinski, A. and Skolnick, J. (1997) Algorithm for rapid reconstruction of protein backbone from alpha carbon coordinates. *Journal of Computational Chemistry*, **18**, 80-85.
24. Ravelli, R.B., Raves, M.L., Ren, Z., Bourgeois, D., Roth, M., Kroon, J., Silman, I. and Sussman, J.L. (1998) Static Laue diffraction studies on acetylcholinesterase. *Acta crystallographica. Section D, Biological crystallography*, **54**, 1359-1366.
25. Wang, X., Wang, C.S., Tang, J., Dyda, F. and Zhang, X.C. (1997) The crystal structure of bovine bile salt activated lipase: insights into the bile salt activation mechanism. *Structure*, **5**, 1209-1218.

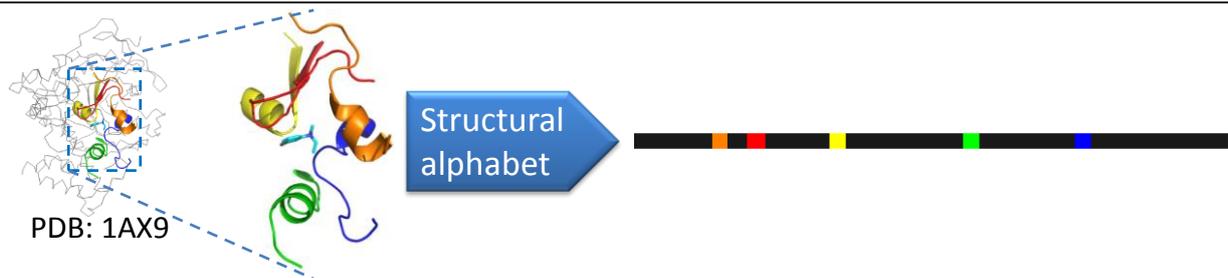
表一：69 對 SOD motif 測試資料及其 SOD-score 分數計算

FDA Testing set	SOD motif pairs								
	SOD-score								
1C3S_A	1C3S_A	1C3R_B	1C3R_A	1C3P_A					
	1.000	0.864	0.864	0.876					
1BCU_H	1SC8_U	3KHV_A	2BDI_L	2BDI_B	2BDI_M				
	0.116	0.109	0.302	0.301	0.375				
1A4G_B	1NSD_B	1NSC_B	1NSB_B	1NSB_A	1INV_A				
	0.917	0.924	0.898	0.899	0.876				
1ACJ_A	2WFZ_A	2ACE_A	1EA5_A	1E66_A	1LLF_B				
	0.943	0.404	0.841	0.542	0.231				
1ACL_A	1W4L_A	1VXR_A	2WG0_A	2VQ6_A	2VT7_B				
	0.786	0.887	0.919	0.778	0.875				
1AJ6_A	1KZN_A	1EI1_A	3FV5_B	3G7B_B	3G7B_A				
	0.638	0.496	0.346	0.872	0.877				
1AQU_B	1AQU_A	1AQY_B	1AQY_A	1G3M_A	1Q20_A				
	0.939	0.884	0.817	0.296	0.439				
1AX9_A	2ACE_A	1VOT_A	2VIC_B	2VJB_B	1AQL_A				
	0.937	0.908	0.915	0.909	0.444				
1AZM_A	1AZM_A	1BZM_A	1HCB_A	1HUG_A	1JV0_B	2IT4_B			
	1.000	0.974	0.971	0.985	0.411	0.894			
1B3N_A	1B3N_A	2GFW_A	3G0Y_A	2GFY_A	3HNZ_A	2GQD_A			
	1.000	0.836	0.868	0.886	0.795	0.823			
1A27_A	1FDS_A	1FDU_C	1FDV_B	1FDV_A	1FDV_C	3HB5_X	3HB4_X	1AQU_B	3HIC_A
	0.206	0.475	0.153	0.266	0.474	0.250	0.250	0.130	0.106
1BKF_A	1BKF_A	2FKE_A	1FKD_A	1FKJ_A	1FKF_A	2FAP_A	1NSG_A	1N1A_A	2VN1_B
	1.000	0.598	0.594	0.593	0.593	0.608	0.608	0.585	1.000

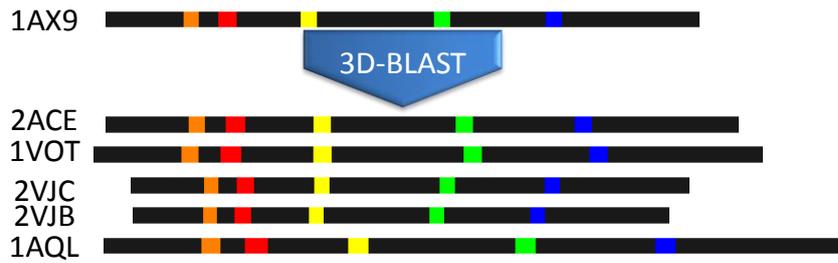


圖一: SOD motif 之概念圖。M1-M2-M3 是一組胜肽片段，環繞在某一特定蛋白質與配體結合部位之介面上。若有另外兩個蛋白質，其中分別包含 M1'-M2'-M3' 或 M1''-M2''-M3'' 等三個局部結構片段，其個別結構、彼此之間距離、以及結構位向，皆符合 M1-M2-M3 所呈現的模版，則可推測此二蛋白質亦可能結合該配體。

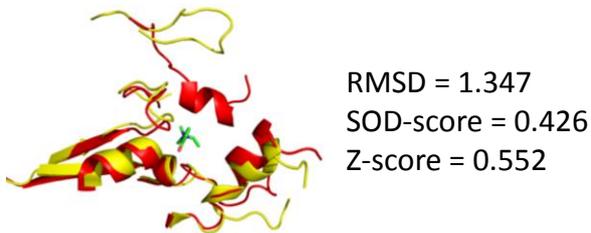
1) 建立SOD motif



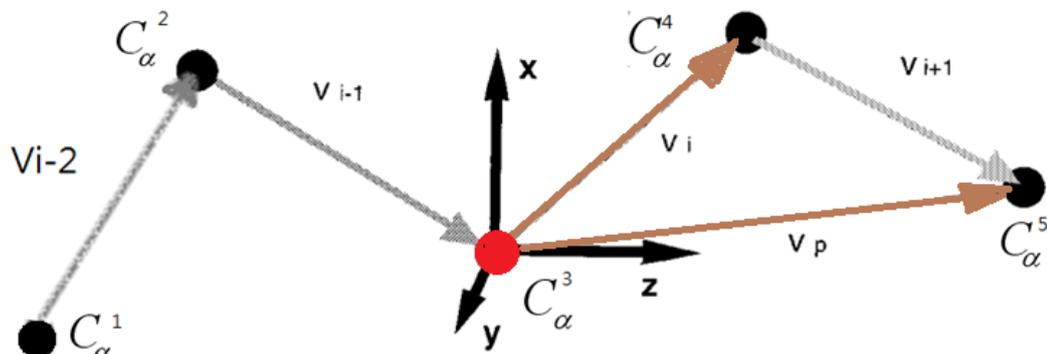
2) 搜尋SOD motif



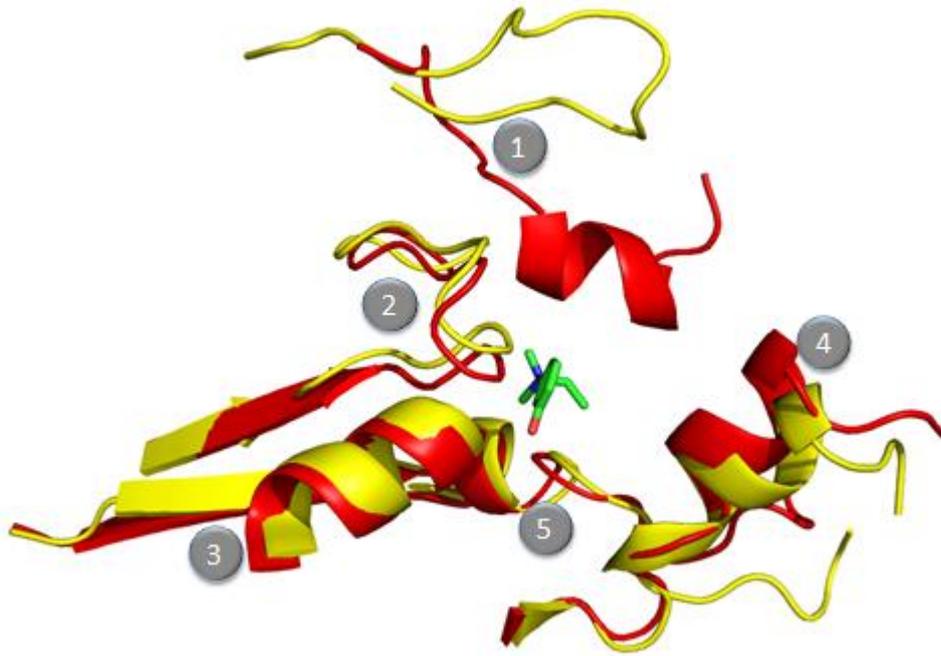
3) 立體結構疊合確認



圖二: 研究方法架構圖。1)為 SOD motif 建立階段。根據蛋白質內某一配體結合之部位為中心,向其序列上兩端延伸,決定數個局部結構字元片段。2)為 SOD motif 搜尋階段。將數個結構字元片段同時透過 3D-BLAST 進行快速結構資料庫比對,找出同樣擁有絕大部分相似結構片段的其他蛋白質,成為候選蛋白質。3)為立體結構疊合確認階段。此步驟利用結構疊合演算法,計算並確認在三維空間中,候選蛋白質中各結構字元片段是否與查詢蛋白質的各結構片段結構相似、片段間距離相似、且位向一致。



圖三: 片段決定向量示意圖。此處以四個局部結構片段為例，每個結構片段根據 C_α 位置計算其質心（以黑點表示之）。四個質心的質心作為該 motif 的旋轉支點（以紅點表示之）。根據其中兩個進行外積的向量（棕色箭頭），產生出以紅點為中心，三個垂直的單位向量（黑色箭頭）。



圖四: 1AX9_A 及 1AQL_A 之 SOD motif 結構疊合。紅色標示之蛋白質結構為 1AX9_A；黃色標示者則為 1AQL_A。1AX9_A 之結合配體為 EDR，位於結構中心位置，以綠色表示之。此 SOD motif 乃是以編號第 2 至第 4 段局部結構，圍繞在蛋白質-配體結合環境周遭。然而，兩蛋白質中第一段局部結構片段，在空間位向上並未一致，1AQL_A 之局部結構偏離配體結合位置甚遠，不構成 SOD motif 的一部份。

國科會補助專題研究計畫成果報告自評表

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

- 達成目標
- 未達成目標（請說明，以 100 字為限）
- 實驗失敗
- 因故實驗中斷
- 其他原因

2. 研究成果在學術期刊發表或申請專利等情形：

論文：已發表 未發表之文稿 撰寫中 無

專利：已獲得 申請中 無

技轉：已技轉 洽談中 無

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本研究內容與結果依照原計畫行程表執行，本研究皆相符於原計畫內容：完成發展 SOD motif、設計新的結構疊合方法與適當的門檻閾值。最後研究結果顯示，SOD motif 定義合理且具生物意義，驗證了原計畫之預期想法：「為蛋白質結合部位的環境中定位出關鍵而重要的模版，以進行功能註解。」。未來我們將進一步發展此技術，進行蛋白質片段-藥物-疾病之三相網路的建立，並探討蛋白質結構、藥物與疾病之關聯性。

本研究計畫指導碩士級研究生共兩位，期間訓練學生程式設計、文獻搜尋與整理等資訊相關技能，並培養學生探討生物意義之生物相關能力。此亦符合原計畫預期目標。本研究成果正著手準備發表期刊論文，將投稿於 *Computers in Biology and Medicine* 或 *BMC Bioinformatics* 等國際期刊。

國科會補助專題研究計畫出席國際學術會議心得報告

日期：102 年 10 月 30 日

計畫編號	NSC 101 - 2221 - E - 216 - 041		
計畫名稱	位向距離結構模版應用於蛋白質功能之預測與註解		
出國人員姓名	董其樺	服務機構及職稱	中華大學生物資訊系 助理教授
會議時間	102 年 6 月 8 日至 102 年 6 月 9 日	會議地點	馬來西亞 吉隆坡
會議名稱	生物資訊暨生物醫學國際研討會 2013 2nd International Conference on Bioinformatics and Biomedical Science		
發表題目	複雜網路理論應用於蛋白質結構字元相似度分析 A Complex Network Approach for the Analysis of Protein Units Similarity Using Structural Alphabet		

一、參加會議經過

此國際會議場地舉辦在飯店 Concorde Inn 內，靠近吉隆坡機場不遠。因此交通方便（有接駁專車）之下，前往會議場地沒有遇到令人困擾的問題。會議第一天，我隻身前往會場報到。向櫃台報到過後，領取相關資料後，便在飯店準備的場地裡準備演講的資料與講稿。

上台報告當天，我被安排在上午十點五十分開始的場次。我的報告順序為第四位，可以好整以暇地聆聽前面講者所帶來的研究成果，以及後面講者豐富的演講（[照片一、二](#)）。尤其是在我所發表的這場次裡，有位印度的學者以穩健的台風與生動的講解，給現場所有優秀研究人員們一個精彩的報告，令我收穫良多。而該位講者，在會後也如預期獲得了大會評選 Best Papers 的榮耀。

現場大會除了提供中餐外，上、下午也有提供 coffee break 的時間，這時候可以看到所有從世界各地來的學者，集結在大廳裡，一邊端著咖啡、拿著小點心，一邊討論剛剛在演講廳裡所聽到的最新研究資訊。而我則與同樣來自台灣的教授與學生們，討論並交換各自己的心得與經驗（[照片三](#)）。會議結束於傍晚六點多，頒發最

佳論文獎之後，所有人也一同合影留念（[照片四](#)），為這次的國際會議留下完美的句點。

二、與會心得

出國參加國際會議，這已不是第一次。但馬來西亞吉隆坡則是我第一次踏入。心得感想中，覺得最衝擊的地方，正是馬來西亞多種族和諧融合在一個國家的這種文化風情。

會議之中，報告者除了來自日本、印度、印尼、埃及等國家外，最多的即是馬來西亞當地的研究學者。就我觀察，在這次會議中，許多馬來西亞的研究成果與能量，一點也不比台灣的研究差，甚至更勝之。這讓我覺得，若只是待在台灣獨自做研究，不常出去到別的國家看看，與其他國家的研究學者接觸與交流，實在是把自己的格局做小了，成了井底之蛙。甚至是以往只知道美國或日本等國的生物資訊研究領先全球，殊不知原來還有印尼、泰國、馬來西亞等地，他們的生物資訊相關研究，也是有其獨到之處。這次的會議，讓我會想要在未來更積極爭取研究經費，多多參與亞洲各地所舉辦的國際會議。

三、發表論文全文或摘要

Abstract— In this paper, we present a network approach based on the recent developed 3D-BLAST method of rapid protein structure search. We defined new local segments that represent structural feature of proteins named units of structural alphabet (USA). Each USA is composed of two protein secondary structures, and one loop located between these two secondary structures. We performed all-against-all structural comparison of USA and recognized the USA-based similarity network. The analytical result shows that the network with a power degree distribution is called scale free. These results not only suggest the existence of organizing principles in the local protein structure but also allow us to identify potential key fragments that could be useful for future new drug development and design.

四、建議

本次國際會議的差旅費補助，只夠支付教師一人出席。學生沒有多餘的經費可一同前往，失去了讓他們遊歷國際研究的機會。在此建議學校或國科會可設置「出席國際會議獎助貸款」，一種提供類似助學貸款的專案。此專案可讓研究生以低利或甚至無息的方式，向學校或國科會申請借貸，以支付註冊費、交通費、生活費等高額費用。

出席國際會議，通常要讓學生在短時間內拿出大量的現金支付旅費，對私立學校的部分學生來說可能有點難度，但若是可用分期方式，將出國貸款按月償還，對每個月領有獎助學金的研究生來說，應該會輕鬆不少。這也能夠給很多因為一聽到出國就聯想到花大錢的學生們，一個吸引他們多多參與國際會議的誘因。

五、攜回資料名稱及內容

1. Conference Program 議程一本。



2. 與會及報告證書一紙。

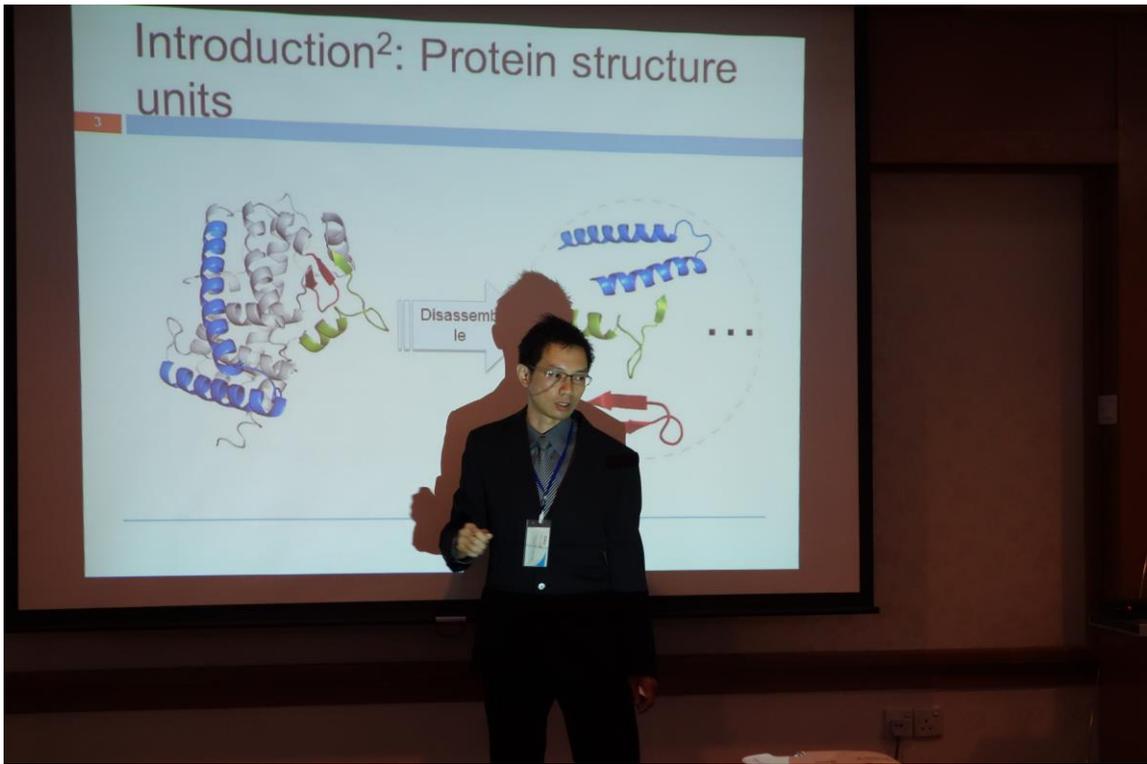


六、其他

出席會議時，以數位相機拍攝之照片數張。



照片一，口頭報告開始。



照片二，投影幕前解說研究內容。



照片三，與元智大學副教授李宗宜博士合影。



照片四，會後留影。

國科會補助計畫衍生研發成果推廣資料表

日期:2013/10/31

國科會補助計畫	計畫名稱: 位向距離結構模版應用於蛋白質功能之預測與註解
	計畫主持人: 董其樺
	計畫編號: 101-2221-E-216-041- 學門領域: 生物資訊與醫療資訊
無研發成果推廣資料	

101 年度專題研究計畫研究成果彙整表

計畫主持人：董其樺		計畫編號：101-2221-E-216-041-				計畫名稱：位向距離結構模版應用於蛋白質功能之預測與註解	
成果項目		量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）	
		實際已達成數（被接受或已發表）	預期總達成數（含實際已達成數）	本計畫實際貢獻百分比			
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 （本國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	1	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 （外國籍）	碩士生	2	2	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

<p>其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</p>	<p>無</p>
----------------------------------------------------------------------------------------	----------

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本研究內容與結果依照原計畫行程表執行，本研究皆相符於原計畫內容：完成發展 SOD motif、設計新的結構疊合方法與適當的門檻閾值。最後研究結果顯示，SOD motif 定義合理且具生物意義，驗證了原計畫之預期想法：「為蛋白質結合部位的環境中定位出關鍵而重要的模版，以進行功能註解。」。未來我們將進一步發展此技術，進行蛋白質片段-藥物-疾病之三相網路的建立，並探討蛋白質結構、藥物與疾病之關聯性。

本研究計畫指導碩士級研究生共兩位，期間訓練學生程式設計、文獻搜尋與整理等資訊相關技能，並培養學生探討生物意義之生物相關能力。此亦符合原計畫預期目標。本研究成果正著手準備發表期刊論文，將投稿於 Computers in Biology and Medicine 或 BMC Bioinformatics 等國際期刊。