

從蛋白質結構資料庫看同步輻射 光源在結構生物技術的發展與影響

Overview on the Development of Protein Structure Determination by Using Synchrotron X-Ray Crystallography from a Viewpoint of the Growth of Protein Structural Data

潘冠宇、詹鎮熊

Kuan-Yeu Pan, Chen-hsiung Chan

X 光晶體繞射是決定生物巨分子結構的主要技術，蛋白質結構資料庫中有 86% 的分子結構是透過這門技術而決定出來的。本文扼要地介紹蛋白質結構資料庫和新穎結構的重要性、資料量的變化趨勢與技術發展的關聯，以及同步輻射光源技術發展在結構生物領域的影響。

Most of the macromolecular structures deposited in Protein Data Bank (PDB) are determined by X-ray crystallography. In recent years, the growth rate of protein structural data has slowed, implying that the enhancement of techniques for novel protein structure determination is desired. The integrated technology platform for biocrystallography developed by advanced synchrotron radiation facilities provides the advantage of exploring innovative methods for novel structures discovery to enhance the growth rate of PDB.

一、前言

當實驗技術有了突破，往往讓該領域的科學進入新的發展，隨著技術的逐漸成熟而被大量引用後，加速了研究成果的累積，使該領域的科學呈現嶄新的風貌；相對地，當實驗技術到達瓶頸時，研究成果的新穎性將受到限制。因此從研究成果的資料庫數據的變化趨勢，可以一窺研究技術的進展，同時可以推測未來實驗技術與儀器需求的輪廓。

1958 年 John C. Kendrew 和 Max Perutz 突破技術瓶頸，首度利用 X 光繞射技術解出蛋白分子的三維立體結構—肌紅蛋白 (myoglobin)，並於 1962 年獲得諾貝爾化學獎。此研究開啟了人類藉由研究生物分子結構去瞭解生命現象的大門，而「結構生物學」這門跨領域的學科也就此孕育而生。

隨著 X 光光源技術與蛋白質結晶技術逐漸發展成熟，越來越多的蛋白質結構被解出來，所累積的結構資訊被轉換成數位資料存放於資料庫中，透

過資訊技術，研究者可以從開放的資料庫中取得檔案，利用軟體在電腦上呈現出分子的立體結構，進行比對、分析和模擬等工作。蛋白質分子都有其特定結構，而且結構與功能之間的關係密切，藉由對結構的了解，可以明瞭生物分子的運作機制，進而能夠控制、修改和設計其功能。這方面的知識是藥物設計、蛋白工程、蛋白質體學等學科的基礎。

結構生物學的發展不僅讓生物學家可以從分子層次去研究生命機制，蛋白質分子如何從一維的氨基酸序列生成獨特而具備功能的三維立體結構，也就是蛋白質分子的「摺疊 (folding)」機制，也讓其他領域的科學家為之著迷，特別是物理學家，紛紛投入相關的研究工作，希望能夠達到從氨基酸序列上的資訊就可以準確預測出三維立體結構，而從蛋白質結構資料庫中所提供的資訊去歸納出蛋白質摺疊法則，是預測結構的重要方法之一，因此資料庫的質與量扮演著關鍵的角色，而具有新穎摺疊模式的蛋白質分子結構資訊，更是研究焦點。

蛋白質三維結構的檔案量逐年快速成長 (圖 1)，至 2009 年 3 月為止，蛋白質資料庫⁽²⁾ (Protein Data Bank, PDB) 已累積了超過 56000 筆的生物分子三維結構檔案。儘管如此，新穎性的蛋白質結構發現卻遇上了瓶頸⁽¹⁾，這樣的結果有兩種可能原因：一是幾乎所有的蛋白質摺疊模式都已被發現，因此很難再發現新穎的蛋白質結構；另一個是由於實驗技術未能突破某些限制而造成的現象。目前決定蛋白質三維結構的技術主要有三種：X 光晶體繞射、核磁共振和低溫電子顯微術，每種技術都各有其限制，而且蛋白質純化與製備技術仍是針對水溶性較佳的球形蛋白 (global protein) 而設計，因此造成 PDB 的資料絕大部分都是球形蛋白質，在細胞生物功能上扮演重要角色的膜蛋白 (membrane protein) 和細胞骨架蛋白 (cytoskeleton) 卻極少有完整的結構資訊。由此可知，若要在新穎蛋白質結構資料上有更大的躍進，需要實驗技術上有開創性的突破。

本文將從 PDB 資料庫的變化，一窺結構生物學上的實驗技術發展，並介紹同步加速光源設施為求在生命科學領域有所突破，擬以結構生物學技術方面的發展為研發重點。

二、蛋白質結構資料庫與結構預測

生物資料庫的建立，可以讓生物學家藉由資訊技術對其研究成果進行分析、比對和模擬等工作，擺脫見樹不見林的窘境。結構資料庫中最具代表性

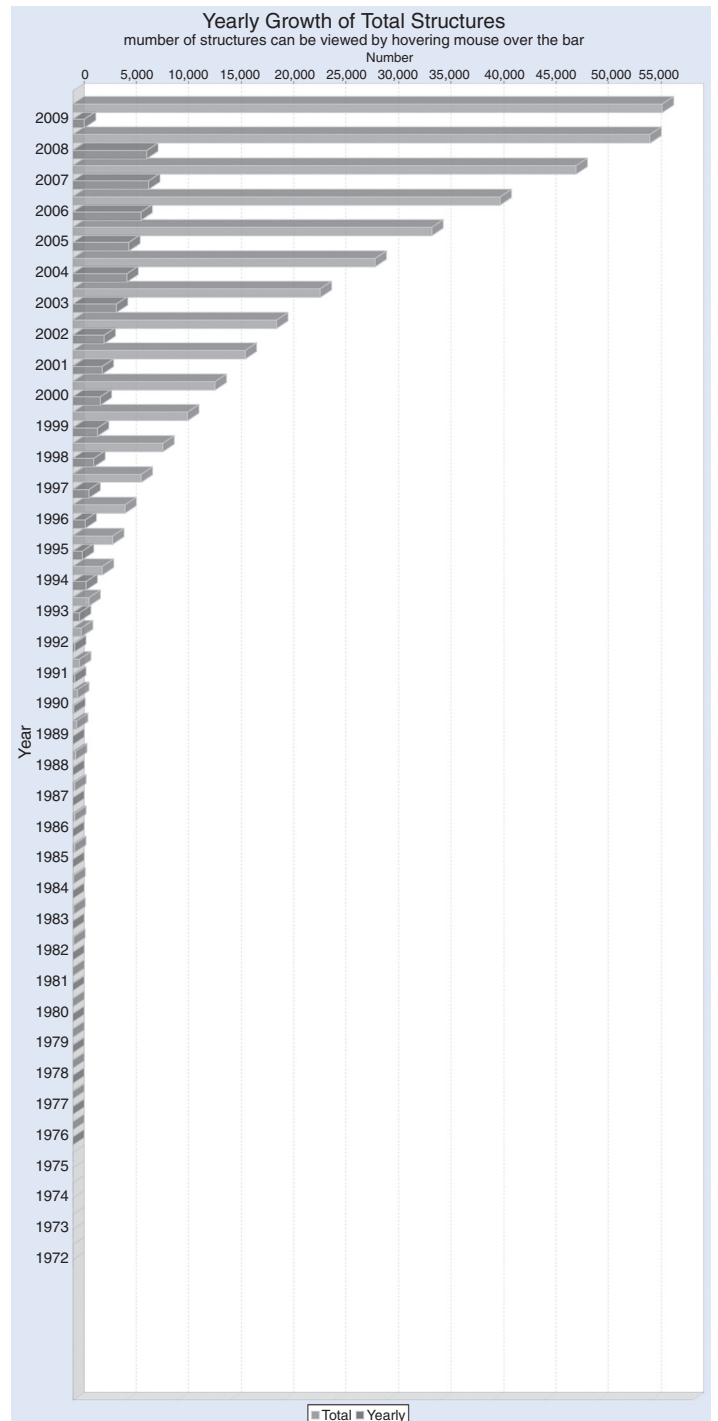


圖 1. PDB 成長趨勢圖 (摘自 <http://www.rcsb.org>)。

的就是 PDB，它是由結構生物資訊研究團隊 (Research Collaboratory for Structural Bioinformatics, RCSB) 維護的，其中包含蛋白質、核酸、醣類的三維立體結構資料，供研究者存取使用。這些結構是利用 X 光結晶繞射 (X-ray crystallography) 或核磁共振光譜 (NMR spectroscopy) 決定出來的，主要資訊包括：原子的空間座標、蛋白質的一級結構與二級結構，以及引用的文獻出處。PDB 本身是其他許多資料庫的原始資料來源，例如分子模型資料庫 (molecular modeling database, MMDB) 就利用 PDB 的資料建立分子的三維立體影像。

蛋白質有著複雜緊密的三維結構，其構形 (conformation) 可分為二級、三級和四級結構。蛋白質主要是由氨基酸以簡單的線性共價鍵結所組成，二級結構是該序列在空間中生成的規律性區域構形，一般有摺版 (sheet)、圈線 (coil)、轉折 (loop) 及螺旋 (helix) 等形式，而大部分蛋白質包含數種二級結構的組合。蛋白質的三級結構描述分子在三度空間中的摺疊方式。四級結構是描述蛋白質的複合構形，也就是蛋白質如何在三度空間中和其他分子作用。

利用視覺化的方式檢視蛋白質的三維結構，藉以找出可能可以用藥物調節分子活性的位置，是蛋白質體研究中最主要的步驟之一。然而目前能解出蛋白質三維結構的技術或方法仍然繁複耗時，例如蛋白質純化或 X 光結晶繞射，即使有自動化技術的協助，在沒有突破性的新技術被引入之前，要對蛋白質結構進行分類、解讀和分析往往需要很長的時間。蛋白質和核酸序列不同，核酸序列是單純的一維空間資訊，蛋白質結構則是三維空間資訊，而且是動態的，會隨著溫度、化學反應、酸鹼度和環境中的其他因素而改變形狀以及與其他分子的作用關係。要決定出蛋白質體裡約 30000 個蛋白質分子的靜態結構就已經非常費工耗時了，更遑論這些蛋白有近乎無窮的動態結構。因此利用資訊技術模擬和預測蛋白質結構的計算方法，被認為是快速決定新蛋白質結構的一條捷徑⁽³⁾。

目前有兩種主要的計算方式用來從氨基酸序列預測蛋白質的二級或三級結構。第一種是從頭演算法 (*ab initio*) 的方法，從基本分子物理原理出發，

不管分子和其他蛋白質之間的關連，演算出可能的結構。第二種方法通常稱為 heuristic，是利用從已知蛋白質結構獲得的資訊與知識來預測新蛋白質的結構。

從頭演算法的方法可以單由序列資料就計算出結構，但是除了最小的蛋白質以外，這個方法所需要的計算量實在過於龐大，並不實用。Heuristic 方法的前提是：大部分新的蛋白質序列都可以從目前已有結構資料庫中找到結構相似的蛋白質，可以利用這些結構相似的蛋白質做為模版，預測出新蛋白質的結構。另一個假設是：蛋白質序列中部分氨基酸的改變，對於蛋白質結構不會造成顯著的變化，也就是說當序列相似時結構也可能相似。由於一定需要一個模版做為起點，因此這個方法受限於沒有辦法建立新結構的模型。Heuristic 是目前普遍運用於蛋白質結構預測的方法，能決定出越多的新穎蛋白質結構，就越能增加預測的準確性。

三、蛋白質結構的分類

如何判定一個結構是否為新穎蛋白質結構呢？要回答這個問題，必須參考結構分類的準則，也就是分子摺疊方式或拓樸結構相同的蛋白質會被歸為同類，如果無法歸類為已有的結構，才算是新穎的結構。所謂相同的摺疊方式或拓樸結構，表示結構中包含數目近乎相同的二級結構單元 (secondary structure elements, SSE)，而且二級結構單元彼此的連結方法相同，相對的空間分布位置也相似。

SCOP (structural classification of proteins)⁽⁴⁾ 是常用的結構分類依據，以樹狀結構的分類方式將蛋白質分成 class、fold、superfamily、family、protein 及 species 六個層次，希望將蛋白質的三級結構與其演化之間的關係連結起來⁽⁵⁾。

一般而言，新穎的摺疊結構會隨著 PDB 資料量的累積而增加，但是這趨勢從 2004 年後有了改變，PDB 資料的增加並沒有讓 SCOP 新的摺疊類型隨之增加 (圖 2)，意味著蛋白質的摺疊類型可能是有限的，當然也非常可能是實驗技術已經遇到了瓶頸。在 Michael Levitt 的研究中發現 PDB 從 1995 年開始就已經不是指數型成長了，而且 PDB 資料

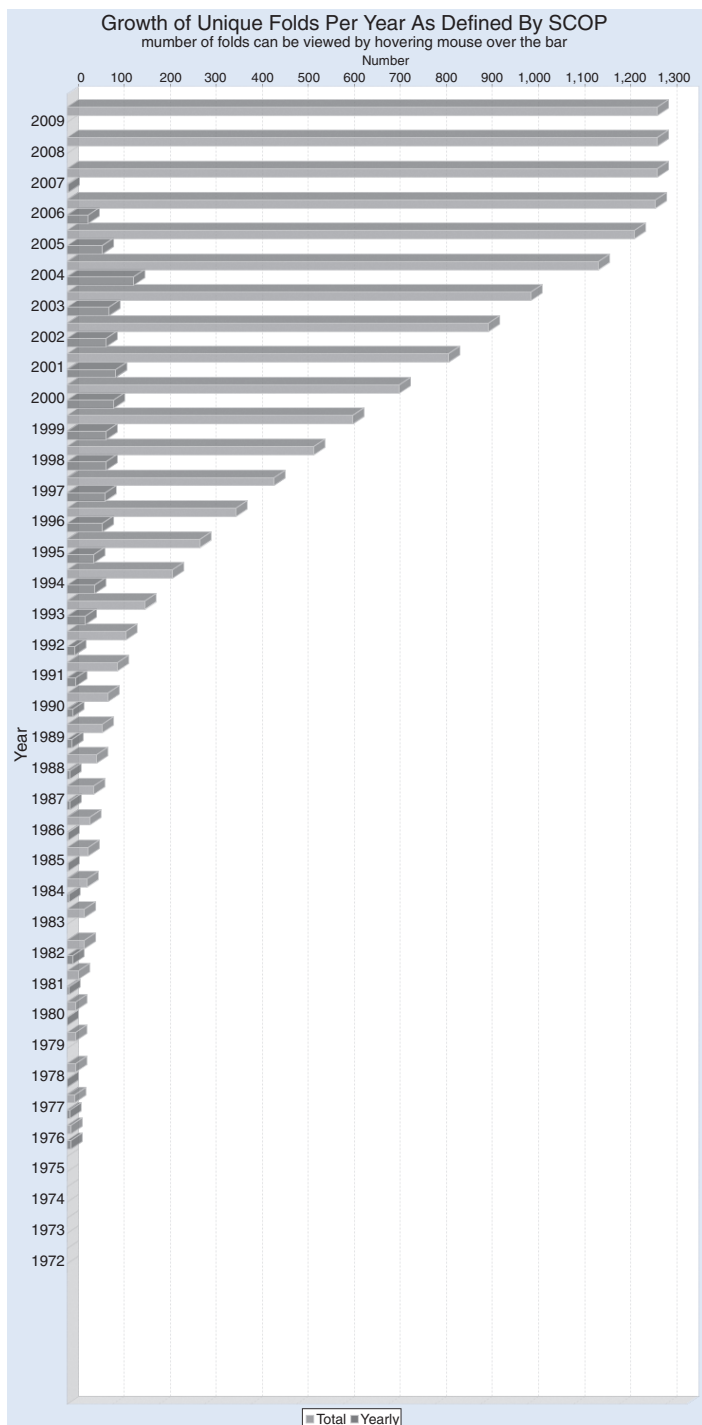


圖 2. SCOP 成長趨勢圖 (摘自 <http://www.rcsb.org>)。

量增加速率的變化與新技術導入息息相關⁽¹⁾。圖 3 是 PDB 檔案的年增率變化圖。圖中的第一個高峰發生在 1972 至 1976 年間，是蛋白質結晶學的濫觴時期；第二個高峰發生的時間似乎和 Digital

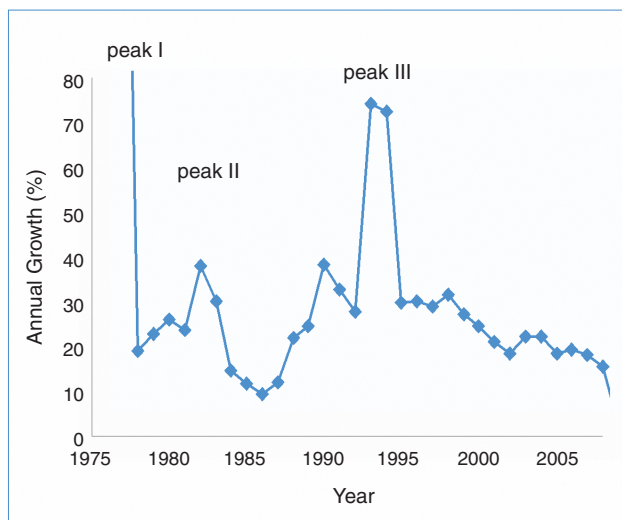


圖 3. PDB 年增率變化圖。

Equipment Corporation 的 VAX 780 電腦在 1978 年問世相關，也就是說電腦運算能力的提昇增加了蛋白質結構資料的產出；第三個高峰是在 1991 至 1997 年間，這段期間同步輻射光源的 X 光及液態氮低溫處理晶體樣品的技術被引入結構生物學的研究。這項資料顯示實驗技術的突破會反映在研究成果的資料量上。而 SCOP 成長的停滯，則暗示著目前的結構生物實驗技術亟待有新的突破。

四、X 光晶體繞射技術在結構生物學上的限制

目前 X 光晶體繞射技術是決定蛋白質三維結構的最主要方法，在 PDB 中有 86% 是利用此技術決定出來的，其餘的 14% 則是利用 NMR 技術決定出來的 (圖 4)。顧名思義，X 光晶體繞射實驗的先決條件是必須培養出蛋白質的晶體，而且晶體的質與量必須達到一定的標準，才有機會收到較佳 X 光繞射訊號，進行後續的結構決定工作。如果不能製備出適當的蛋白質晶體，就無法藉由 X 光繞射獲得分子立體結構的訊息。但是分子量低於 25 kD 的蛋白質則還有另一個機會，就是利用核磁共振技術解出結構。

PDB 檔案資料中包含決定結構的方法。進行 PDB 搜尋時，會發現有些分子量較小的蛋白質分

子，同一個分子會有由 X 光晶體繞射技術解出一個 PDB 檔案，也會有核磁共振技術解出的另一個 PDB 檔案。前者是由繞射點數據分析出晶體中電子密度的分布，再推算出晶體內的分子結構；後者則是從多維核磁共振光譜中取得分子在溶液環境中鍵結和空間的結構關係，再以這些結構限制條件，計算模擬出蛋白質最有可能的動態結構，而動態結構正是藥物設計和蛋白質分子動力的關鍵。X 光晶體繞射雖然仍是決定生物分子結構的主流方法，但受限於晶體的製備工作，以及不易測動態結構，這些都是目前有待突破的技術瓶頸。

五、技術困局的曙光—先進的同步輻射光源設施

生物醫學領域是世界各國的科技產業發展重點，大型的先進同步輻射光源設施紛紛積極投入相關的研發與應用，結構基因體計畫的啟動為結構生物學的技術突破帶來契機。大型的同步加速器設施提供高亮度的 X 光源，讓結構生物學家從蛋白質晶體中取得結構資訊，但是目前的設施並不是為了分析大量新蛋白分子的結構而設計，不足以應付高通量 (high throughput) 實驗的需求。Bioxhit (biocrystallography X on a highly integrated technology) 就是為了因應這個問題而發起的計畫。

Bioxhit 是統合歐洲所有的同步輻射光源設施共同發展 X 光結晶學的技術整合平台，其目標在於設定分析蛋白質三維結構的最佳標準步驟，垂直整合技術平台：從樣品的製備、晶體的培養到 X 光繞射數據的蒐集，包括生物資訊工具的整合，提供標準化、自動化的研究工具，預定在四年半的時間 (從 2004 年 1 月至 2008 年 6 月) 決定出近萬個近年來發現的新蛋白質。

此外，德國聯邦教育研究部 (German Federal Ministry for Education and Research, BMBF) 已經核撥八百八十萬歐元給歐洲分子生物實驗室 (European Molecular Biology Laboratory, EMBL)，在德國同步加速器研究中心 (German Synchrotron Research Centre DESY) 的 PETRA-III 儲存環建構一個全新的結構生物學整合研究設施，命名為

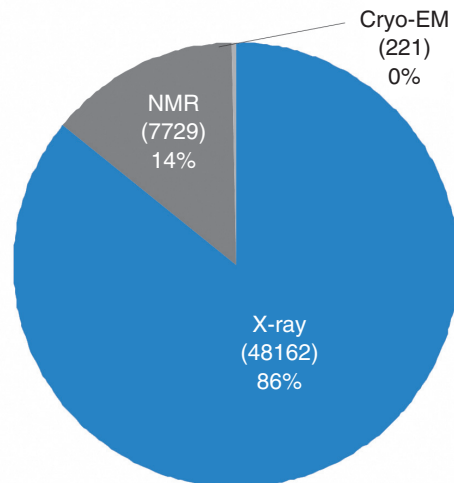


圖 4. PDB 檔案中 X 光晶體繞射技術、核磁共振技術與低溫電子顯微術解出結構的比例圖 (括號內為檔案數)。

EMBL@PETRA-III，預計於 2010 年開始運作。

在此類的大型計畫支持下，X 光晶體繞射技術已朝向更強而精細的光束線、更高的解析度、更有效率的樣品製備方法、更快的偵測技術、更快的數據處理能力、以及動態結構的決定技術而努力⁽⁶⁾，而且有了大幅的進展。X 光光學系統和設備的提升，使歐洲同步輻射設施 (European Synchrotron Radiation Facility, ESRF) 能將 X 光光束線控制在 5 μm 的大小，可以偵測極小的晶體或是纖維狀的樣品，降低晶體製備門檻⁽⁷⁾。利用流體技術探討蛋白質的結晶條件⁽⁸⁾，其成果可以運用於加速晶體的生成，提高培養晶體的成功率。在進行 X 光繞射的過程中，對生物巨分子晶體的實驗壽命、光束線的性質，以及輻射對生物巨分子損害的掌控技術，可以提升訊號偵測與數據蒐集的效能⁽⁹⁾。而動態結構的測定技術上也已經有了新突破，Bourgeois 和 Royant 利用即時解析勞厄繞射圖 (real-time-resolved Laue diffraction) 和冷凍捕捉法 (freeze-trapping methods) 發展出「運動結晶學 (kinetic crystallography)」，跳脫 X 光晶體繞射技術只能測定靜態結構的限制⁽¹⁰⁾。前述的技術將會日趨成熟，突破既有窠臼，可以預期將會為未來的結構生物學帶來嶄新的風貌。

將同步輻射光源技術應用在結構生物學上的趨勢已經在國際間風起雲湧的展開，國家同步輻射研

究中心也順應趨勢，積極地推廣同步輻射光源技術在生命科學領域的應用。在國家基因體計畫的支持下，中心在 X 光晶體繞射技術與設施已具備深厚的基礎，未來台灣光子源 (Taiwan Photon Source, TPS) 的運轉將會產生更先進的光源，輔以跨領域人才的培育與投入，將會是台灣在生物醫學科技領域做出重大發展的新契機。

參考文獻

1. M. Levitt, *Proc. Natl. Acad. Sci. USA*, **104**, 3183 (2007).
 2. <http://www.rcsb.org/pdb/home/home.do>
 3. B. Bergeron, *Bioinformatics Computing*, New Jersey: Pearson Education, Inc., 366 (2003).
 4. <http://scop.mrc-lmb.cam.ac.uk/scop>
 5. D. R. Westhead, J. H. Parish, and R. M. Twyman, *Instant Notes: Bioinformatics*, Oxford: BIOS Scientific Publishers Ltd., 123 (2002).
 6. K. Moffat and W. Chiu, *Curr. Opin. Struct. Biol.*, **15**, 535 (2005).
 7. C. Riek, M. Burghammer, and G. Schertler, *Curr. Opin. Struct. Biol.*, **15**, 556 (2005).
 8. B. Zheng, C. J. Gerdt, and R. F. Ismagilov, *Curr. Opin. Struct. Biol.*, **15**, 548 (2005).
 9. R. BG Ravelli and E. F. Garman, *Curr. Opin. Struct. Biol.*, **16**, 608 (2006).
 10. D. Bourgeois and A. Royant, *Curr Opin Struct Biol*, **15**, 538 (2005).
-
- 潘冠宇先生為國立清華大學生命科學博士，現任國家同步輻射研究中心助理工程師。
 - 詹鎮熊先生為國立清華大學生命科學博士，現任慈濟大學醫學資訊系助理教授。
 - Kuan-Yeu Pan received his Ph.D. in life science from National Tsing Hua University. He is currently an assistant engineer at National Synchrotron Radiation Research Center.
 - Chen-hsiung Chan received his Ph.D. in life science from National Tsing Hua University. He is currently an assistant professor in the Department of Medical Informatics at Tzu-Chi University.