

# 植物病原菌 *Xanthomonas* *Campestris* 的結構基因體研究

## The Structural Genomic Studies of *Xanthomonas* *Campestris*: A Plant Pathogenic Bacterium

周三和

Shan-Ho Chou

本文主要報導中興大學生化研究所在植物病原菌 *Xanthomonas campestris* 結構基因體學研究的最近進展。結構基因體是一門最近快速發展的跨領域科學，其中包括生物資訊學、核磁共振學、X 光晶體繞射學、功能基因體學及蛋白質體學等等。它的順利推展對現代生命科學、生物科技及藥物研發都將有極深遠的影響。目前本團隊已做出初步的成績。相信假以時日，台灣的微生物結構基因體學將在國際上佔一席之地。

This article mainly discusses the recent progress in structural genomics studies of the plant pathogen *Xanthomonas campestris* executed in the Institute of Biochemistry, NCHU. Structural genomics is a new and rapidly progressing research frontier including bioinformatics, NMR, X-ray diffraction, functional genomics, and proteomics, etc. If successfully executed, it can make a significant impact toward the modern life science, biotechnology and drug development. We have made a significant progress toward this goal, and believe that Taiwan Structural Genomics teams will make their its debut in future soon.

### 一、基因體計畫的重要性

人類基因體定序計畫的完成可說是廿一世紀初最具震撼性的成就。此一計畫的完成代表了人類的科學技術已邁入了一個全新的里程碑。生物科技無疑是廿一世紀影響人類最重要的科技之一，也是台灣急待技術升級的重要產業。但是生命的奧秘不只是存在於基因序列而已，其中還有一段很大的距離

尚待結構生物學家們努力，才能將生命現象用生物分子結構的語言來解釋。這也就是目前極為熱門的結構生物學 (structural biology) 領域。唯有在充分瞭解重要的生化高分子，如蛋白質、去氧核糖核酸及核糖核酸的三度空間結構後，我們才能明白它們之間的互動情形，或設計出一些所謂藥物分子來幫助或阻斷這些生化高分子之間的互動，以增加對生命現象的瞭解，並改善生活的品質。因此結構基因

體學之研究，即為強調針對某個基因體所表達的大部分蛋白質進行快速三維結構的解析，並進而探討其結構及其相互作用與功能等。

## 二、*Xanthomonas campestris* 結構基因體計畫

*Xanthomonas campestris* 為格蘭式陰性病原菌，為中興大學、陽明大學及清華大學一些系所為執行教育部卓越計畫所合力研究的一頗具學術性及應用性的菌種。它可攻擊十字花科植物，使之感染黑腐病，因而造成農業之巨大損失。它在學術上的特色為：

- (1) 它能分泌多種胞外蛋白，已成為研究革蘭氏陰性菌蛋白分泌之模式系統；
- (2) 它的轉錄單位多以單基因的方式為之，與熟知的大腸菌之多單基因調節方式不同；
- (3) 它不具 cyclic AMP (cAMP) 接受蛋白，但其多元調控蛋白 (cAMP receptor-like protein, Clp) 能夠協助執行此菌的調控功能。

在應用上此病菌可合成 Xanthan gum (為一多醣體)，可用在眾多工業如紡織、造紙等。此一黑腐病菌所造成的植物病徵請參閱圖 1。

## 三、結構基因體研究為一跨領域的高度整合性科學

首先本團隊基因定序組必須將 *Xanthomonas campestris* (Xcc) 約含五百萬對鹼基對的基因體定序完成。接著生物資訊組則將整個基因組作系統化的分析，以提供決定蛋白結構的依據。接著蛋白表達組利用生物資訊組所提供的基因註解，建立有效的蛋白載體體系，供蛋白質表達之用，並做蛋白純化及功能分析的工作。至今為止，超過一千個蛋白已被嘗試表達。接著結構基因組則利用多維核磁共振或 X 光繞射分析法解出它們的三維結構，以瞭解其可能的功能。為了執行此一跨領域的高度整合性計畫，我們已在中興大學建立一結構蛋白體中心，包括硬體 (一個約二百坪的三層建築，並容納世界



圖 1. 高麗菜受到 *Xanthomonas campestris* 十字花科黑腐病菌感染葉緣呈現的病徵。

先進的 800 MHz 及 600 MHz 核磁共振儀、X 光繞射儀及液態層析雙質譜蛋白分析儀) 及軟體。此計畫的最終目的為提供一個含有結構和功能註解的 *Xanthomonas campestris* 結構基因庫。相信本計畫的完成，不僅在學術上有極重要的意義 (可發現一些含有特殊折疊的蛋白質結構與功能以及未知功能的基因與 Xcc 所具特殊的生化調節現象的探討等)，在工業上及農業上也對十字花科植物黑腐病的預防及 Xanthan 的應用，有相當大的助益。我們目前已成功的整合所有的步驟，從基因的擴增、植株、蛋白表達、純化、結晶、數據收集、電腦計算等領域都已順利進行，並已成功解析出一些蛋白的結構。圖 2 表示一些目前已得的蛋白晶體，這些結晶大部分都得到高解析的 X 光繞射圖樣，有些結構並已順利解出。可喜的是，此一十字花科植物黑腐病菌的基因體蛋白似有較高的結晶成功率。利用中研院國家基因體 X 光繞射中心的自動晶體篩選機器，我們發現大部分 (約 60%) 可溶性蛋白都可得到結晶，其中並有約 20% 晶體可得到極佳的繞射圖樣供結構解析之用。

而結構基因體最大的貢獻之一為從解出的結構中，探討這些未知蛋白可能的功能<sup>(1-6)</sup>。從過去的經驗中，我們發現單從基因序列的比對中，無法得知許多未知蛋白的功能 (一般認為當基因序列比對相似度低於 30% 時，即無法判定類似功能)。但在執行本計畫後，我們發現事實上許多相似度低於甚至 20% 的蛋白，也具類似的結構及可能相似的功

能。所以結構的解析對於未知蛋白的功能認定扮演了一個相當重要的角色。

以下將報導此一跨領域計畫所解出的一些蛋白結構及其可能的生物功能。

### 1. XC1739：抗多重抗生素調節蛋白 (Multiple Antibiotic Resistance Repressor, MarR)

許多細菌的抗多重抗生素現象的出現已經造成世界上醫療單位的高度注意與關心。如果沒有儘早找出對策，到時可能出現可抵抗所有已知抗生素的超級致病菌，因而造成遭細菌感染時無法投藥的窘境，而造成人類大量死亡。細菌為了抵擋抗生素的攻擊，本身已發展出兩種機制，一為利用質體 (plasmid) 上的分解抗生素基因將外來的抗生素加以分解。由於質體可在細菌中輕易傳送，菌落中的每一個細菌都可獲得所需質體，以抵抗外來抗生素。但此一機制的缺點為它只能針對少數的抗生素

產生作用，不是一種普遍的方法。同時它是外來性的，不是內生性的，所以當細菌受到抗生素攻擊時，必須等到其他細菌的支援才能避免被抗生素消滅。另一細菌用來抵抗外來逆境，包括各種有害物質的機制為細菌的染色體本身即有一種有謂的 *marRAB* 的調控基因組存在。當細菌受到外來有害物質如有機溶劑、清潔劑及有毒物質等等時，此一調控基因就會被啟動，而表達出一些俱 pump 功能的蛋白，將這些有害物質輸送到細胞外，避免破壞細胞的完整及功能<sup>(7,8)</sup>。所以此種 *marRAB* 調控基因組為細菌能否適應環境變化的一個重要機制。而 MarR 蛋白即為 *marRAB* 調控基因組所表達的第一個蛋白，在沒有外來有害物質時，它會被順利表達，進而與 *marRAB* 調控基因組上游的操作子核酸結合，關閉其他有 pump 功能蛋白的表達。但當外來抗生素存在時，它會與此 MarR 蛋白結合，進而改變其構型，使得 MarR 蛋白無法持續與上游的操

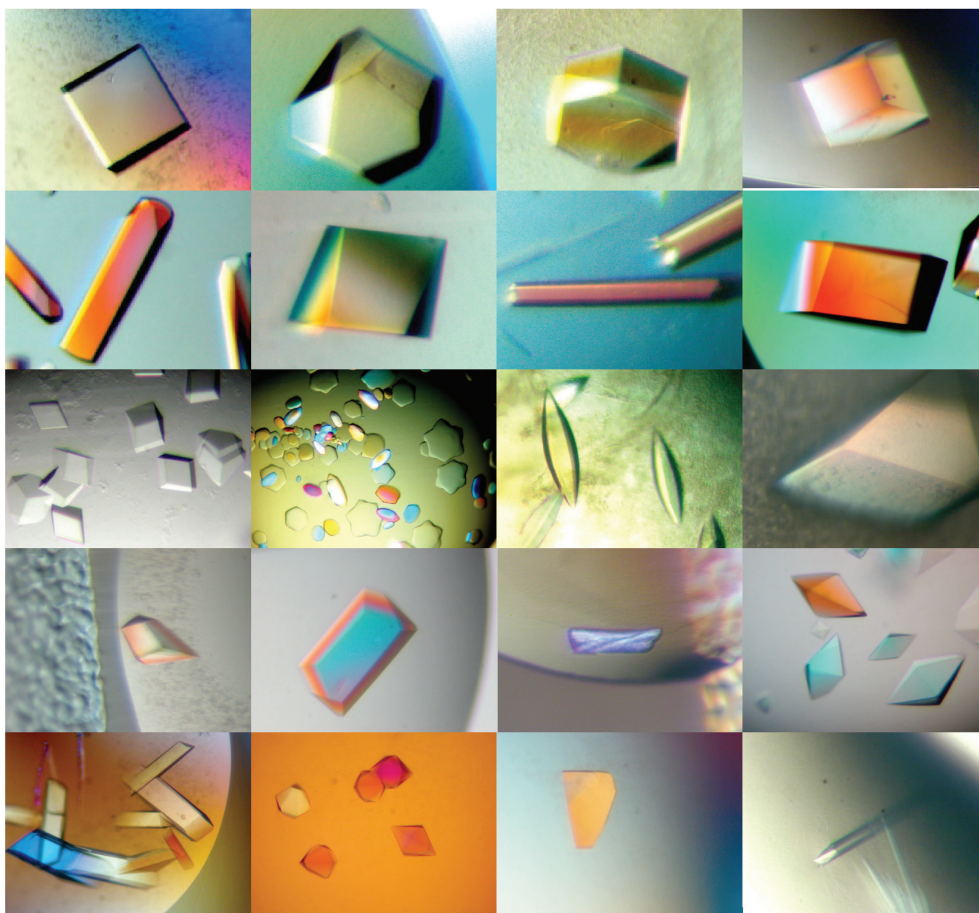


圖 2.

十字花科黑腐病菌結構基因體所得到的一些蛋白結晶。當一個蛋白分子有穩定的三度空間結構，並能有秩序的堆疊在一起時，在適當的過飽和溶液中，即可形成類似鑽石的晶體。當這些晶體長到 0.2 mm 左右的大小時，即可供 X 光繞射分析之用。

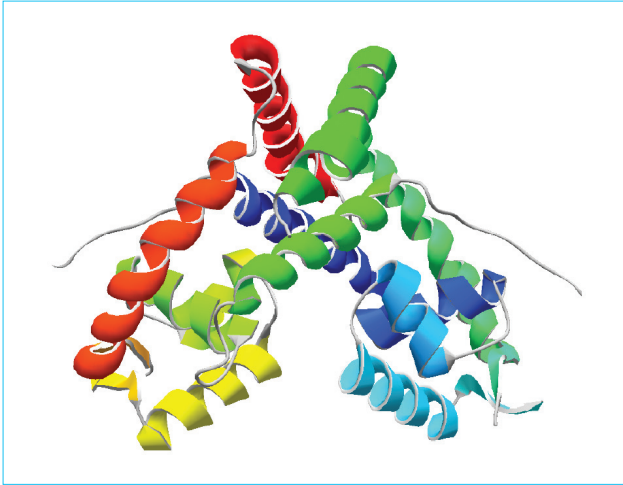


圖 3. 抗多重抗生素調節蛋白的三度空間結構。它為二個蛋白單體所組成，並且大部分為螺旋結構。下方的螺旋（黃色及青色）為與核酸結合之處。

作子核酸結合，進而啟動了其他蛋白如 MarA 與 MarB 等的表達，將細菌內的抗生素運送到細胞外，以保護細菌免受抗生素的攻擊。

顯然的，這個 *marRAB* 調控基因組蛋白的表達對研究細菌有關多重抗生素的抗藥性有很大的影響。我們已成功的表達、純化及結晶 Xcc MarR 蛋白，並利用 X 光繞射解析它的結構（如圖 3 所示）。當與已發表的 *E. coli* MarR/salicylate 複合體作

比較時<sup>(7,9)</sup>，發現 MarR 蛋白在與 salicylate 結合後，它的 DNA-binding domain 構型的確發生很大的變化<sup>(10)</sup>，造成它無法繼續與其操作子核酸結合（如圖 4 所示）。

## 2. XC6422 蛋白：屬於 $\alpha/\beta$ Hydrolase 折疊的蛋白水解酵素

此一蛋白在尚未解析其結構時，單從其一級核酸序列並無法得知其功能。目前已決定 XC 6422 之三級結構（圖 5），它係由 8 個  $\beta$ -strand 及 5 個  $\alpha$ -helix 所組成。由結構看來，此蛋白是屬於  $\alpha/\beta$  hydrolase 折疊的一員<sup>(11)</sup>。由於蛋白質的三級結構往往會顯露出許多關於其生化功能的線索，因此透過 DALI 工具<sup>(12)</sup> 將此蛋白質結構與已知結構之蛋白資料庫比對後發現此蛋白結構與 *Aeropyrum pernix* K1 之 acylpeptide hydrolase (apAPH)<sup>(13)</sup> 之相似性最高。這意味著 XC 6422 雖然在一級結構上與其他蛋白的相似性很小 (< 20%)，但由三級結構上來看它可能為一蛋白水解酵素。生物體中在合成多肽鏈時往往會在 N 端加入醯基團 (acetyl, formyl) 來防止其在胞內被降解<sup>(14)</sup>，而 APH 主要的功能就是將這些蛋白之醯化胺基酸移除，藉此調控蛋白質分子在細胞中之半衰期。ApAPH 在結構上包含了兩個部分，其主要的 apAPH 活性是位於  $\alpha/\beta$  hydrolase domain，也是 XC 6422 與其相似的區域。將此二

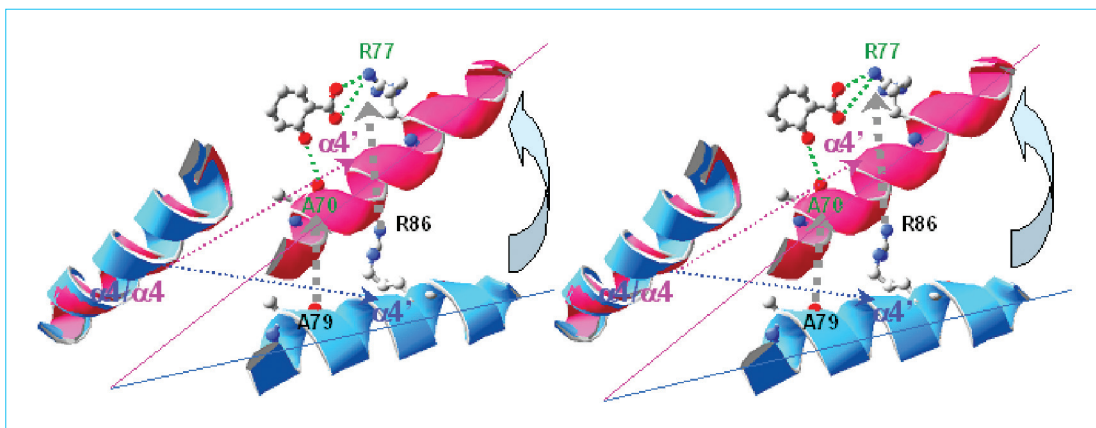


圖 4. 當藥物（以原子形式表示）進入抗多重抗生素調節蛋白內時，與  $\alpha 4'$  螺旋上的 A79 骨架及 R77 胺基酸側鏈以氫鍵結合（綠色虛線），而將原本與核酸結合的  $\alpha 4'$  螺旋往上移動約 15 Å，造成此一蛋白無法繼續再與操作子核酸結合。

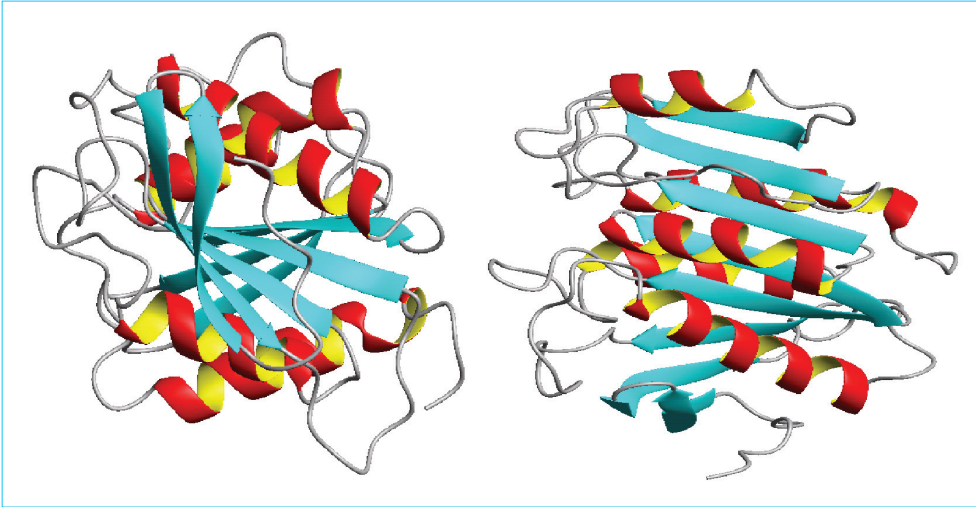


圖 5.  
XC 6422 之三度空間結構圖。其結構屬於  $\alpha/\beta$  hydrolase 折疊的一員。其特點為在蛋白中間俱一由八個  $\beta$ -strands 所組成的結構，其兩側則有二及三個  $\alpha$  螺旋，組成一含  $\alpha$  螺旋及  $\beta$  板塊的混合結構。

蛋白重疊後亦可以看出此二蛋白之三級結構相似性極高 (圖 6)。同時經過結構上的分析後，我們亦發現 XC 6422 與 apAPH 一樣亦具有 Ser-Asp-His 的催化三胺基酸組 (catalytic triad)，且其幾何位置幾乎完全相同 (圖 6)。進一步比較兩者之活性區也都有凹陷的區域，不過電荷分布有些不同。ApAPH 除了有帶負電荷之胺基酸分布外，尚有一些疏水性胺基酸分布，這也是其最高活性之受質為 Ac-Phe 與 Ac-Leu 的原因。而 XC 6422 則是以帶負電荷之胺

基酸分布為主。這似乎意味著 XC 6422 所需要的受質是與 apAPH 不一樣，推測其最高活性之受質應為帶正電之醃化胺基酸 (Ac-Lys、Ac-Arg 與 Ac-His)。儘管 XC 6422 之蛋白質序列與 apAPH 相似性很低 (identity < 17%)，但是 XC 6422 之三級結構卻與 apAPH 極為相似，且包含了 serine protease 在催化上的重要特徵，包括：Ser-Asp-His catalytic triad 與 Gly-X-Ser-X-Gly 之序列。雖然目前我們仍然沒有足夠的實驗數據來證明 XC 6422 在細胞中

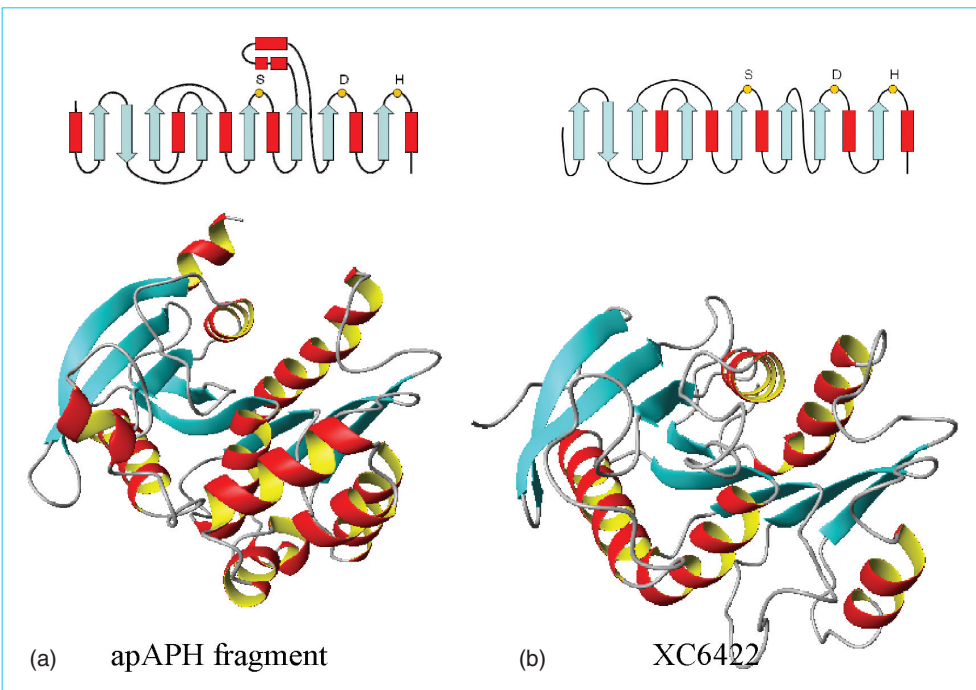


圖 6.  
apAPH fragment (a) 與 XC 6422 (b) 之二級結構 (上) 及三級結構比較。由此圖可知，除了 apAPH 多了兩個  $\alpha$  螺旋以外，兩者的三級結構極為類似。而且兩者的催化三胺基酸組 SDH 在空間的位置亦極為相似。由此可見此兩個蛋白的功能有異曲同工之妙。

所執行的功能為 acylpeptide hydrolase，不過可以確定的是 XC 6422 極有可能是以 serine protease 催化機制來執行其功能。另外 apAPH 除了包含  $\alpha/\beta$  hydrolase domain 外，尚包含  $\beta$ -propeller domain<sup>(13)</sup>，這是個非常值得討論的地方。在 apAPH 中要進入活性區必須先經過  $\beta$ -propeller domain，而其直徑約為 7 Å，這樣的大小只能容許小肽通過，因此可以避免大肽或蛋白被意外地水解；反觀 XC 6422 並沒有此構造，也就是說如果 XC 6422 也是一種 APH，則可能對於受質之大小較沒有選擇性。目前我們還不知道 XC 6422 之真正受質為何，但由其結構上做推測，可猜測其可能的受質，再進一步做其功能分析 (functional assay)，如此就能證明 XC 6422 在生物體中所執行的生化功能。

### 3. XC847 蛋白：3' 端至 5' 端之外切核糖核酸水解酵素

核糖核酸的正常合成及水解對細胞的存活扮演了一個重要的角色，人類的一些重要疾病如癌症、發炎、或阿滋海默症等都被發現與此一功能的不正常有關<sup>(15)</sup>。而此一功能需要各式各樣的核酸水解酵素來執行，包括內切及外切酵素等。從大量的數據搜尋，大致上外切核糖核酸水解酵素可以分為六類<sup>(16)</sup>，而 XC847 蛋白則被認為屬於 DEDDh 類的外切核糖核酸水解酵素，而且是唯一能分解短核酸 (2-5 個核苷酸長) 的外切酵素<sup>(16)</sup>。同時它也是維持生命所必需<sup>(17)</sup>，所以它們是一類極重要的蛋白質，值得仔細研究。

XC847 是由 194 個胺基酸所組成，分子量

為 21778.3 Da。由 Yuhong Zuo 所提出的資料，oligoribonuclease 含有 3 個功能區域。下列為以三個功能區域：功能區域 I：[LIMV](2)-w-[IV]-D-[LC]-E-M-T-G-L、功能區域 II：[IL]-A-G-N-S-[IV]-x(2)-D-x(2)-F-x(4)-M-P、功能區域 III：K-x(3)-H-x-A-x(2)-D-I-x-E-S-[VI]-x-E-L-x(2)-Y-[RK]<sup>(16)</sup>。保留的 DEDDh 胺基酸以粗體字表示。由序列的比對來看 XC847 可以確定為 oligoribonuclease。目前對此三個功能區域中的 DEDD 的了解為和金屬離子產生配位鍵來穩定反應過渡狀態。由於對反應來說此四個胺基酸和金屬產生配位鍵非常重要，所以此四個胺基酸被高度的保留。XC847 的二級結構的組成為  $\beta\beta\beta\alpha\alpha\alpha\beta\alpha\beta\alpha\alpha$ ，由 6 個  $\beta$ -strands 及 9 個  $\alpha$ -helix 所組成。功能區域 I 所形成的二級結構為  $\beta$ -strand。功能區域 II 為一個  $\beta\alpha$  形式。功能區域 III 主要為  $\alpha$ -helix。活性的四個胺基酸 D15、E17、D115 及 D166 分別在  $\beta$ -strand、 $\beta$ -strand、 $\alpha$ -helix 及  $\alpha$ -helix 上，而 H161 則是位在 loop 的位置上。圖 7 為 XC847 為整體結構。 $\beta 1$ 、 $\beta 2$ 、 $\beta 3$  及  $\beta 4$  四個 strands 形成 antiparallel  $\beta$ -sheet。 $\beta 5$  及  $\beta 6$  和  $\beta 1$  形成 parallel  $\beta$ -sheet。 $\beta$ -sheet 被 9 個 helix 所包圍。由於 DEDDh 家族都帶有相同的功能區域，因此可藉由比較結構來找尋活性位置。由 DALI 軟體做三級結構比較，發現同為 DEDDh 家族的 DNase 或 RN<sub>ase</sub> 蛋白 Pop2、ISG20 及  $\epsilon 186$  與 XC847 有最佳的相似度；Pop2 為一 3'-5' deadenylation RNase<sup>(18)</sup>，在 yeast 中被發現；ISG20 主要水解單股 RNA，對單股 DNA 會有次要的活性，在人類中被發現<sup>(19)</sup>；而  $\epsilon 186$  為 DNA polymerase III 的 subunit，為一

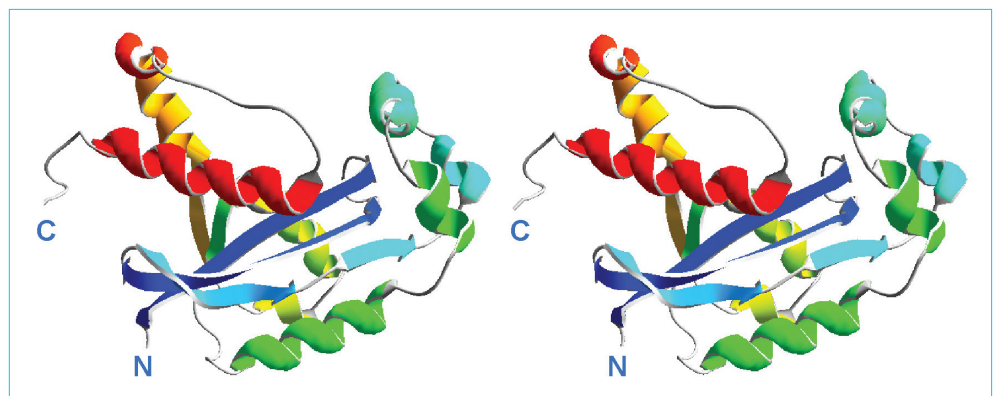


圖 7.  
XC847 的三級結構立體圖。

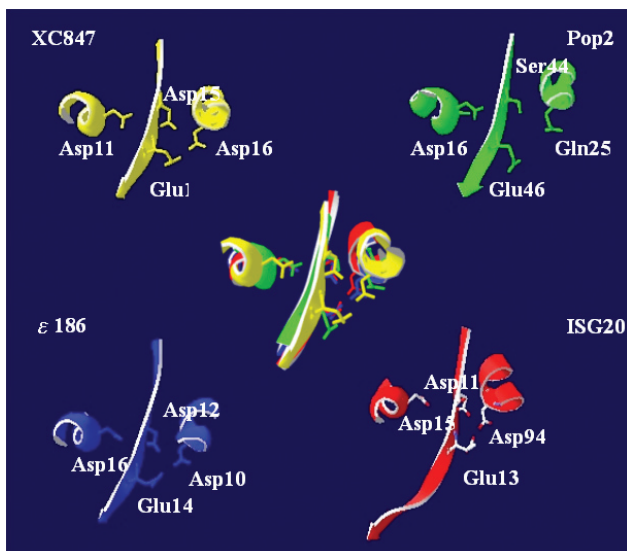


圖 8. XC847、Pop2、 $\epsilon$ 186 及 ISG20 活化中心四個保留胺基酸 DEDD 位置比對。黃色為 XC847，綠色為 Pop2，藍色為  $\epsilon$ 186，紅色為 ISG20，中間為四個蛋白質活性位置的重疊圖。

DNase<sup>(20)</sup>。圖 8 為四種類似核酸酵素 XC847、Pop2、ISG20 及  $\epsilon$ 186 活化中心位置的 DEDD 胺基酸分布位置圖。從活性位置比對可以推測 XC847 也是以相同的方法和金屬產生配位鍵，並具相似核酸水解機制。但是 XC847 為何只負責水解 2-5 個核苷酸長度的核糖核酸原因尚不太清楚，必須從結構上再做仔細的分析。

#### 4. XC5357 蛋白：屬於 Cupin Fold 的未知功能蛋白

XC5357 之基因產物的功能註解為 tetracenomycin polyketide synthesis protein。Tetracenomycin (Tcm) C 是一種由 *S. glaucescens* 所合成的抗生素並具有抗癌 (antitumor) 的活性<sup>(21)</sup>。*S. glaucescens* 上有一段約 12.6 Kb 的 DNA 片段，其上包含 12 個與 Tcm C 生合成 (biosynthesis) 有關的基因<sup>(22, 23)</sup>。由 Tcm JKLMN 這些蛋白所組成的 Tcm PKS 蛋白會藉由與脂肪酸的合成途徑相似的反應機構合成 Tcm C 的前驅物 (Tcm F2) (圖 4-1)，目前已知 Tcm J 是催化 Tcm C 形成環狀物以及增加 TCM F2 蛋白產量的蛋白之一，然而 Tcm J 蛋白至

目前為止尚未分析出其功能，仍是一未知功能蛋白<sup>(24)</sup>。XC5357 在 Pfam 中是屬於「cupin fold」superfamily<sup>(25)</sup> 的家族成員之一。經由序列比對得知此 superfamily 的主要特色是它們都擁有一段保守性的 barrel domain，而「cupa」這個詞彙在拉丁語中有「small barrel」之意，故而將此 superfamily 命名為「cupin fold」。Cupin 這個保守性 domain 分布的範圍極廣，自細菌、真菌至動植物皆可發現其蹤跡，包含不同的酵素、結合蛋白等<sup>(26,27)</sup>。也就是說 cupin 這個 domain 藉由一些  $\beta$ -strand 所產生的緊密的  $\beta$ -barrel 結構形成一個 cupin core，提供了一個穩定的三級結構，並在各種極端的環境下可以發揮其自身功用<sup>(28)</sup>。在已知的 PDB (protein data bank, <http://www.rcsb.org/pdb/>) 資料庫中，*Thermotoga maritime* 中的 TM1459 (PDB:1vj2)<sup>(29)</sup> 是一個 novel manganese-containing cupin、TM1287 (PDB:1o4t)<sup>(30)</sup> 其功能預測為 putative oxalate decarboxylase 以及尚未發表的 antibiotics synthesis protein (PDB:1v70) 都是屬於 cupin\_2 family 的一員，它們都擁有 cupin domain 特有的摺疊構造—以 8-10 個  $\beta$ -strand 圍繞成一個緊密的  $\beta$ -barrel，且以 N 端的  $\beta$ 1 strand 和另一個單體的  $\beta$ -strand 之間有交互作用存在，並形成雙聚體 (dimer)。

在 X 光的結晶結構中，XC5357 可看出自第 2 至第 111 個胺基酸的構形，由這 110 個胺基酸組成 9 個  $\beta$ -strand，以 antiparallel 的形式組成  $\beta$ -sheet，並由兩組  $\beta$ -sheet 包圍成一個  $\beta$ -sandwich 的構形，為 cupin domain 特有的摺疊構造。兩個單體之間以疏水性作用力 (hydrophobic interaction) 和同一側的  $\beta$ -sheet 以上下交錯的方式形成雙聚體，如圖 9 所示。將所得的結晶結構以 DALI server 作三級結構比對發現，1vj2 與 XC5357 的三級結構最相似。1vj2 與 XC5357 之間所比對到的 94 個胺基酸當中，其 RMSD 值為 2.3 Å，胺基酸序列比對則只有 16% 的 identity。1vj2 亦是 cupin\_2 family 的一員。由於屬於 cupin family 的蛋白功能涵蓋極廣，包括酵素及非酵素的蛋白在內，所以 XC5357 可能的活性區 (active site)、結合區 (binding site) 與其他蛋白質間的交互作用 (interaction)、單體之間作用的區域以及可能的功能，仍有待進一步的探究。

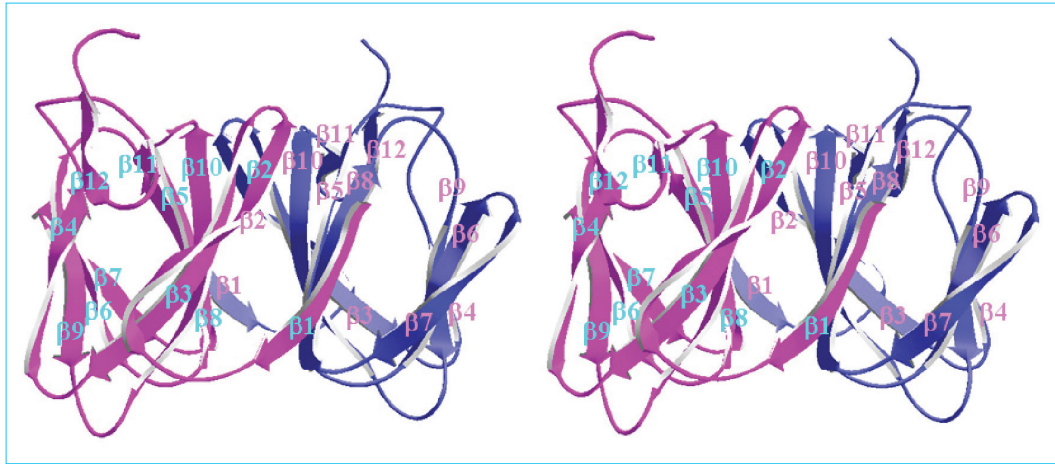


圖 9. XC5357 的 X 光三級結構。XC5357 的單體是由 9 個  $\beta$  strands 以 antiparallel 的形式組成  $\beta$ -sheet，並由兩組  $\beta$ -sheets 包圍成一個  $\beta$ -barrel 的構形，兩個單體之間以疏水性作用力 (hydrophobic interaction) 形成雙聚體 (以粉紅及藍色表示)；此外，在形成雙聚體之時， $\beta_1$  會與另一個單體的  $\beta_8$  產生 domain swapping 的現象，大大增加雙聚體的穩定性。

## 5. XC229 蛋白：屬於 Acyl-CoA Thioesterase 的四聚體蛋白

從基因序列比對，XC229 蛋白起初亦被歸屬於保守型功能未知蛋白。它含 134 個胺基酸，但是在純化過程中，發現它不僅在單體的位置中出現，亦在含四聚體的位置中出現，故估計它是一個同源四聚體的結構。圖 10 表示 XC229 的單體結構。它由九個  $\beta$ -strands 組成，並包住一個長的  $\alpha$  螺旋，而形成一個類似熱狗型的緊密結構。而四聚體的結構則是由雙聚體 (I and II、III and IV) 的雙聚體組成的四聚體。在 I 跟 II 及 III 跟 IV 的雙聚體結構中，單體 I 中的  $\beta_4$  strand 與單體 II 的  $\beta_4'$  strand 緊密的利用氫鍵結合而成一個含八個  $\beta$ -strands 的板塊。同時單體 I 中位於  $\alpha$  螺旋上的 E35 胺基酸側鏈亦與單體 II  $\alpha$  螺旋上的 N26 側鏈形成氫鍵，以穩定此一雙聚體的結構 (圖 11)。所以在此一雙體中，共有至少八個氫鍵參與穩定的力量。而單體 I 除了與單體 II 有許多的交互作用力外，與單體 III 及單體 IV 亦有許多氫鍵參與穩定四聚體結構。如單體 I 中的 D18 的 carboxylate 側鏈與單體 III 中的 R15' guanido 側鏈形成兩個氫鍵。同樣的單體 I 中的 R15 的 guanido 側鏈與單體 III 中的 D18' carboxylate 側鏈形成兩個氫鍵。同時單體 I 中 R40 的 guanido 側鏈及 D74 carboxylate 側鏈亦分別與單

體 III 中的 D74' carboxylate 側鏈及 R40' guanido 側鏈形成四個氫鍵 (圖 11 左下圖)；至於單體 I 與單體 IV 的結合則係藉單體 I 中的 M19 骨架與單體 IV 中的 R17' guanido 側鏈及單體 IV 中的 M19' 骨架與單體 IV 中的 R17 guanido 側鏈氫鍵所造成 (圖 11 右下圖)。所以總共在此一四聚體中，共至少有廿個氫鍵形成。難怪即使在 denatured PAGE 實驗中，亦可觀察到相當多的四聚體成分。

雖然 XC229 蛋白起初係被歸屬於保守型功能未知蛋白，但在決定其四聚體結構，並利用 DALI 軟體尋找類似蛋白結構時，卻發現有兩個在不同細菌中的 CoA-thioesterase 酵素，即在 *Pseudomonas* 的 4-hydroxybenzoyl CoA-thioesterase 及在 *Arthrobacter* 的 4-hydroxybenzoyl CoA-thioesterase 與 XC229 有相類似的結構<sup>(31,32)</sup> (如圖 12 所示)。此二類細菌可分解環境中的有毒物質 p-chlorobenzoic acid，並以此為生活唯一碳原。我們相信 XC229 亦為一 CoA-thioesterase，但對於其真正的受質為何則尚需仔細分析結構及做功能測試才能決定。即使如此，XC229 的結構決定已對它的真正功能邁出一大步。

## 6. XC2382 蛋白：可能參與合成細胞逆境所需的分子蛋白

ORPHEUS 與 Glimmer 兩基因預測軟體所得

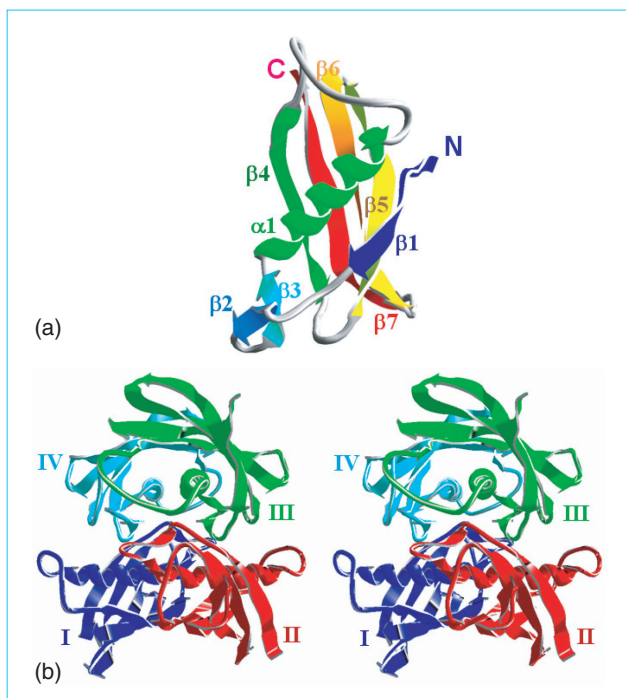


圖 10. XC229 的「熱狗型」單體結構 (a) 及「小狗型」四聚體結構 (b)。單體結構及四聚體結構都以絲帶方式表示。單體結構的顏色係從 N-端的藍色到 C-端的紅色，而四聚體的顏色則以藍、紅、綠及青色表示，以資分別。

XC2382 之基因長度均為 127 個胺基酸。此蛋白質序列經 BLASTp 所得之序列比對結果，發現具有相似胺基酸序列之物種分布自原核到真核。XC2382 之基因命名為 *apaG*，此基因位於 *ksgA-apaG-apaH* operon<sup>(33)</sup>。ApaH 基因之基因產物為 diadenosine tetraphosphatase，主要功能為參與 bis (5'-nucleosidyl) oligophosphates (Ap4N) (A 指 adenine，N 指任何 nucleoside) 之代謝，例如其基因產物可以將 Ap4A 代謝成兩個 ADP<sup>(34)</sup>。研究指出，*apaG* 基因表現常伴隨 *apaH*，推測 ApaG 可能與 ApaH 共同參與 Ap4N 之代謝<sup>(33)</sup>。因而具此段 DNA 序列的基因命名便依 *apaH* 而命名為 *apaG*。當此兩基因 *apaH* 與 *apaG* 發生突變時，會導致 Ap4N 的含量高於正常值 100 倍以上。細胞中，不正常之 Ap4N 含量可能會啟動細胞進行複製。若是給予細胞逆境，則亦會提高細胞中 Ap4N 之含量。

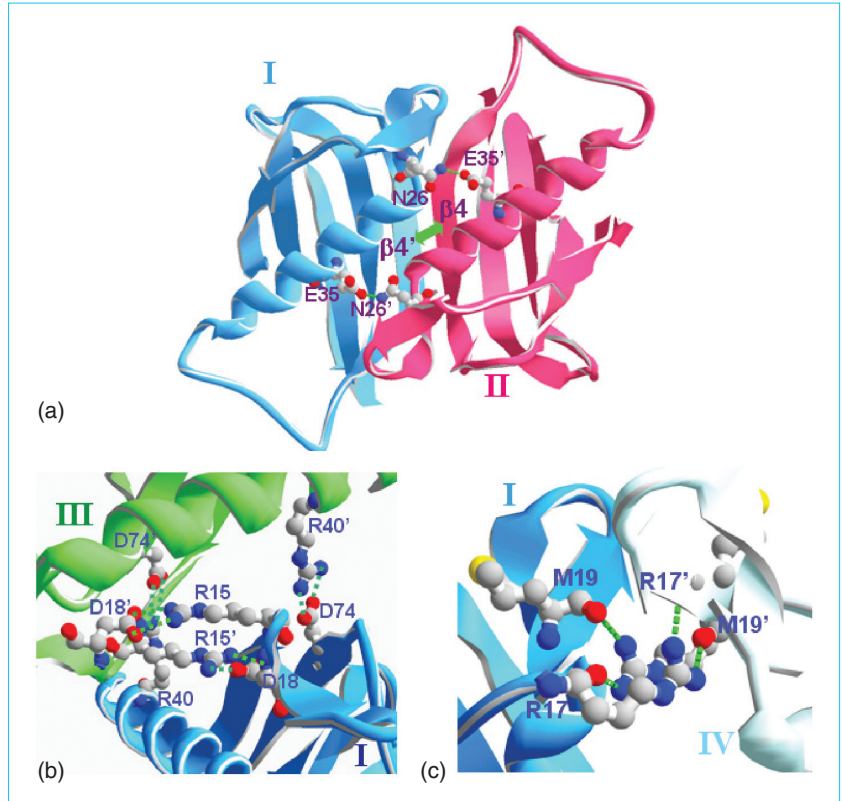
推測此兩基因 *apaH* 與 *apaG* 的基因產物可能參與合成細胞因應逆境所產生之訊息分子<sup>(34)</sup>，並且已知 Ap4N 的代謝存在於真核以及原核。雖然此基因已被命名且此基因很廣泛地存在原核與真核生物，但是在近十年來此基因的相關研究並沒有很大的進展；也就是說，此基因的功能至目前仍不清楚。但是可以肯定的是，此蛋白質應該在細胞中具有重有的生理功能。此蛋白質在 InterPro 資料庫的分類上屬於 UDF525 家族，此家族為未知功能之蛋白質家族。此家族除了 ApaG 外，Cyclin-like F-box 蛋白質之 C-端 domain 亦屬此家族。F-box 蛋白質之 C-端 domain 的受質為蛋白質，因此推測此蛋白質可能參與蛋白質與蛋白質間的作用。雖然 ApaG 的功能未知，但有研究指出，當 ApaG 的同源蛋白 CorD (*Salmonella typhimurium*) 之基因產生突變時，細胞對於環境中鈷 (cobalt) 含量的耐受性 (cobalt resistance) 會下降，且會降低細胞對於鎂離子的輸出量 (CorA transport system)<sup>(35)</sup>。

與 ApaG 之蛋白質序列相似的物種分布雖自原核到真核，但不同物種間的胺基酸序列長度相差頗多，且對於較大的蛋白質如人類與老鼠的 FBA 蛋白質 (371aa 或 367aa) 等，通常這些與 ApaG 具有相似序列的部分位於其蛋白質之 C 端，也就是說此 ApaG domain 多位於蛋白質之 C 端。而大部分之原核生物如 *Xylella fastidiosa*、*Ralstonia solanacearum* 等其蛋白質大小均約在 120 幾個胺基酸長度。XC2382 與原核及真核同源蛋白質序列比對結果所得的保守性序列 GXGVVGXXP 中，GXGXXG 此段保守性序列為能與 NAD 及 FAD 結合之蛋白質的序列特性 (NAD-and FAD-binding protein)<sup>(36,37)</sup>，此段序列為蛋白質參與與 NAD 及 FAD 之 ADP 部分結合之胺基酸序列。XC975 經由 PSORT<sup>(38)</sup> 預測結果為位於細胞之細胞質 (cytoplasm) 的蛋白質且此蛋白質無 transmembrane domain。

XC2382 的三度空間結構亦由 X 光繞射實驗得知，主要由七個  $\beta$ -strands 組成，形成一極度彎曲的筒狀結構，屬於 IgG-like  $\beta$ -sandwich 家族的一員。推測其受質結合區應位於此筒狀結構的頂端，但詳細的結構及功能尚待分析。

圖 11.

XC229 雙體，I/II (a)、I/III (b) 及 I/IV (c) 的作用力示意圖。I/II 的雙體係以  $\beta$ -strand 4 及 4' 以氫鍵結合形成一連續含有八條  $\beta$ -strand 的  $\beta$ -板塊。同時也有一些在  $\alpha$ -螺旋的胺基酸 (以分子結構形式表示) 形成氫鍵。在雙體 I/III 中則主要以單體 I 中 D18/D74 的 carboxylate 側鏈與單體 III 中的 R15/R40 的 guanido 側鏈形成氫鍵。在雙體 I/IV 則以 R17 的 guanido 側鏈與 M19 的骨架形成氫鍵。氫鍵都以綠色虛線表示。



## 7. XC975 蛋白：可能參與細胞遭遇逆境時改變細胞形態的蛋白

此蛋白質序列經 BLASTp 所得之序列比對結果，其同源蛋白質之物種分布自原核至真核，這表示在其他多數物種中都具此相似胺基酸序列。與 XC975 具高度相似胺基酸序列之其他物種所得的功能註解也多為未知功能蛋白 (hypothetical protein)。因此 XC975 之基因功能註解為未知功能蛋白質，且具序列之保守性。與 XC975 序列相近的其他基因產物有 BolA-like protein。在 Pfam 資料庫 (<http://www.sanger.ac.uk/Software/Pfam/>)，其歸屬於 PF01722 BolA 家族。*E. coli* 之 BolA 屬於形態蛋白質 (morphoprotein)，當其大量表現時會使 *E. coli* 之外形變成圓球狀<sup>(39)</sup>。BolA 蛋白質的表現受到 *E. coli* 生長速度的調節，並且在 *E. coli* 進入細胞週期之靜止期 (stationary phase) 時會誘發 BolA 蛋白質的表現。推測此蛋白質可能參與細胞分裂時 elongation 與 septation (為細胞質分裂的過程) 間之轉變。另有研究指出，*E. coli* 生長初期若受到 UV、氧化等逆境亦會誘發 BolA 蛋白質的表現。

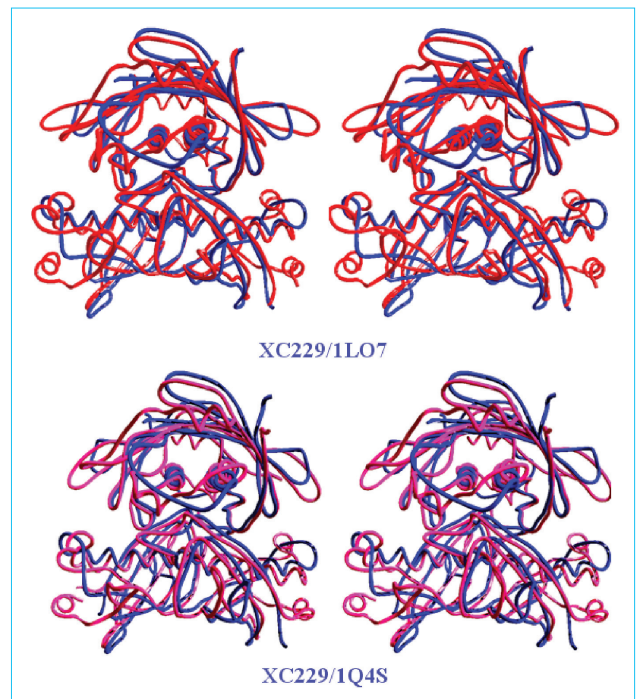


圖 12. XC229 的四聚體與 *Pseudomonas* 的 4-hydroxybenzoyl CoA-thioesterase 及 *Arthrobacter* 的 4-hydroxybenzoyl CoA-thioesterase 重疊圖。

故此蛋白質亦可能為細菌在逆境時，所誘發出相關的蛋白。BolA 同時也被推測能夠誘發 penicillin binding proteins 6 和 5 之轉錄反應<sup>(40)</sup>。依蛋白質於細胞中之功能予以分類的 COG (<http://www.ncbi.nlm.nih.gov/COG/xindex.html>) 資料庫，其 COG 命名為 stress-induced morphogen (COG0271)。

與 XC975 同屬 Pfam 之 BolA-like protein 家族的 gi | 26325949 (*Mus musculus*) 之基因產物的結構已被用 X 光繞射法解出<sup>(41)</sup>。此真核蛋白質的結構 topology 為  $\alpha\beta\beta\alpha\alpha 3^{10}\beta\alpha$ ，與 XC 975 所得之二級結構 topology( $\alpha\beta\beta\alpha\beta$ ) 類似。XC 975 與 *Mus musculus* 之 BolA-like protein 的蛋白質序列有約 30% 的相似度。

解析生化大分子的結構除了 X 光繞射法以外，利用高磁場核磁共振法亦為一常用之方法 (詳情可參閱本期陳金榜博士的專文報導)。利用高磁場核磁共振法解析生化大分子結構的最大的優點為：它可解析生化分子在水溶液中的結構，而不需像 X 光繞射法一樣，一定要有蛋白質的固態結晶，才能得到其繞射圖樣。同時它也可用來研究蛋白質分子在水溶液中的動態情形，得知蛋白質分子在不同部位的局部動態行為。一般蛋白分子的受質接受部位或催化部都較具動態，故利用此法也可猜測蛋白質分子可能的活化中心。但是利用高磁場核磁共振法解析生化大分子結構也有其缺點。最重要的是，它是利用核磁共振原理偵測質子、碳原子或氮原子的共振吸收位置，再利用所謂 nuclear overhauser effect (NOE) 現象，測量所有在 6 Å 以內的質子間的距離。有了質子間大約的距離，才能利用距離設限電腦軟體，將生化大分子所有原子的三度空間座標解出 (即所謂的生化分子結構)。所以利用高磁場核磁共振法解析生化大分子結構必須先行判讀所有質子的共振吸收位置，才能據以測量質子間的距離。對於一個動則上百或數百胺基酸數目的生化大分子，很顯然的，質子的吸收位置將產生很大的重疊現象，造成光譜判讀的不易。所有利用高磁場核磁共振法能解析的生化分子結構一般分子量都局限在 20 kD 以下。同時由於利用 NOE 現象測

量質子間的距離一般都不太準確，同時其數量也不太夠，所以利用核磁共振法所得的生化分子結構一般都比較不若 X 光繞射法精確，不能得到一『個』結構，而只能得到一『組』結構。

圖 14 即為利用高磁場核磁共振法解析出的 XC975 的蛋白結構。由圖可知它是一個由廿個分子所重疊而得的一『組』結構。如果蛋白質分子較小，而質子間的距離數目夠多，就可得到較佳的重疊結構。由於 XC975 只有八十個胺基酸，同時其核磁共振光譜品質極佳，所以我們得到一組其 rmsd 只有 1.2 Å 左右，解析度較高的結構。圖中所標示的胺基酸為在不同種生物中高度保留者，而其都位於蛋白質的內部中心位置，表示它們對於穩定此一蛋白結構具重要貢獻，不能輕易被改變。

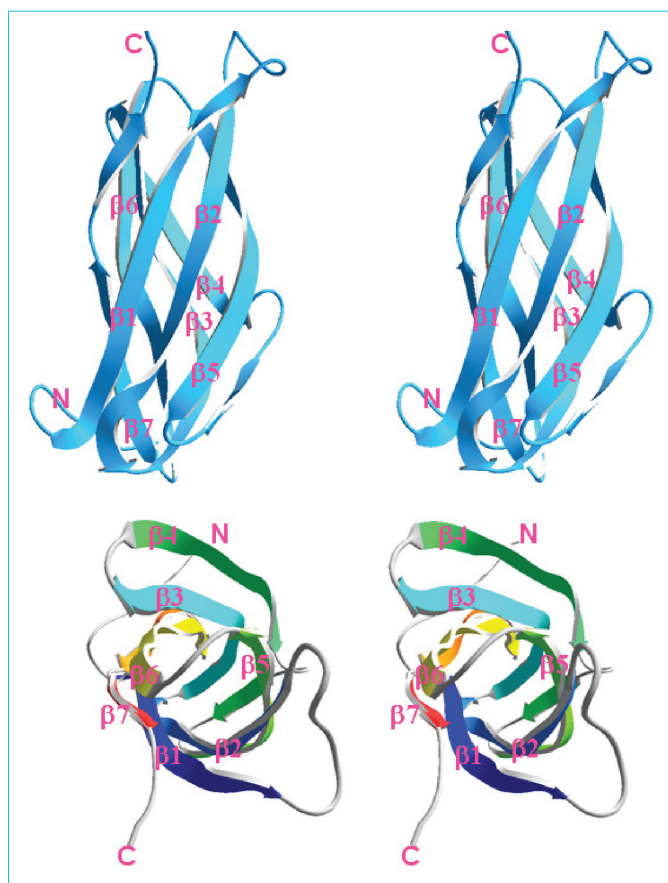


圖 13. XC2382 的三度空間結構圖。它主要由兩個高度彎曲的  $\beta$ -sheet 組成，這可由下圖 (由上往下俯視圖) 得知。

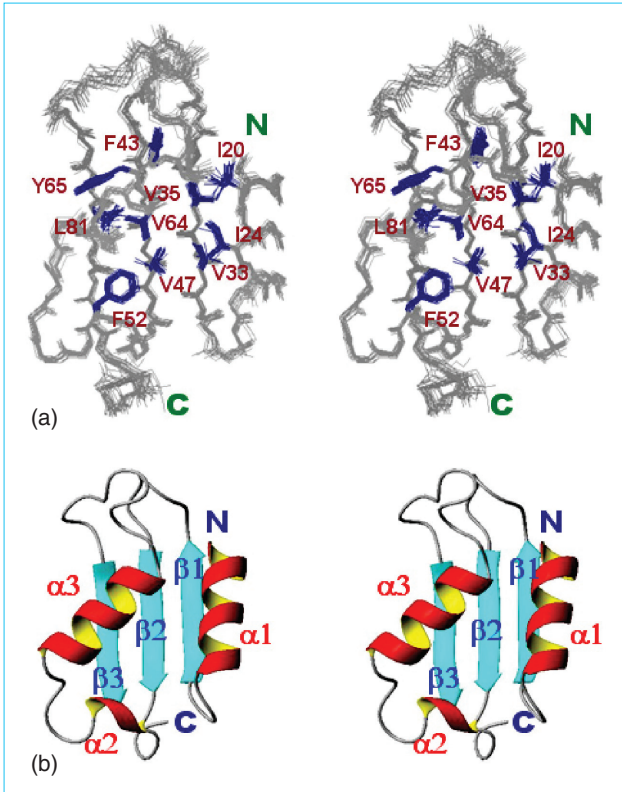


圖 14. 利用高磁場核磁共振法解析出的 XC975 的蛋白結構。(a) 為二十個結構所組成的重疊結構，表示此一蛋白結構俱有極佳的解析度。以藍色表示的胺基酸為高度保留的胺基酸，而它們都位在內部區域，表示它們對穩定此一蛋白結構具重要的價值，而不可取代。(b) 為一絲帶圖，表示此一蛋白俱二層次結構，由  $\beta 1-\beta 2-\beta 3$  組成的  $\beta$ -sheet 及  $\alpha 1$ 、 $\alpha 2$ 、 $\alpha 3$  的  $\alpha$  螺旋組成。

#### 四、結論

以上是本實驗室目前所解析出的蛋白質分子結構的一些說明。雖然本實驗室尚有許多蛋白分子結晶，但由於解析蛋白質分子結構大部分須依賴在日本或新竹的同步輻射儀，有其時間限制，不能像在實驗室內一樣可隨時使用，所以進度不易掌握。同時本校生化所的詹迺立教授也參與此一結構基因體計畫，也已解出一些具重要生化功能的 Xcc 蛋白分子結構。所以結構基因體為一可實現的目標，相信有朝一日台灣在結構基因體的研究將可在國際上佔一席之地。

#### 誌謝

感謝教育部追求卓越計畫的支持，使得台灣在此一重要領域不致缺席。也感謝合作夥伴中研院基因體中心 X 光核心設施的同仁及新竹國家同步輻射中心同仁們的大力協助，使得本大型計畫得以順利推動。當然本實驗室的成員們所付出的辛勞亦功不可沒。

#### 參考文獻

1. A. Yee, K. Pardee, D. Christendat, A. Savchenko, A. M. Edwards, and C. H. Arrowsmith, *Acc. Chem. Res.*, **36**, 183 (2003)..
2. E. A. Woestenenk, M. Hammarstrom, T. Hard, and H. Berglund, *Analytical Biochem.*, **318**, 71 (2003).
3. Y. Kyogoku, Y. Fujiyoshi, I. Shimada, H. Nakamura, T. Tsukihara, H. Akutsu, T. Odahara, T. Okada, and N. Nomura, *Acc. Chem. Res.* **36**, 199 (2003).
4. P. Walian, T. A. Cross, and B. K. Jap, *Genome Biology*, **5**, 215 (2004).
5. A. F. Edwards, A. Yee, A. Savchenko, A. M. Edwards, and C. H. Arrowsmith, *Cur. Opin. Chem. Biol.*, **8**, 42 (2004).
6. C. Zhang and S.-H. Kim, *Cur. Opin. Chem. Biol.*, **7**, 28 (2004).
7. A. S. Seoane and S. B. Levy, *J. Bacteriol.*, **177**, 3414 (1995).
8. R. G. Martin and J. L. Rosner, *Mol Microbiol.* **53**, 183 (2004).
9. M. N. Alekshun, S. B. Levy, T. R. Mealy, B. A. Seaton, and J. F. Head, *Nature Struct. Biol.*, **8**, 710 (2001).
10. R. G. Martin and J. L. Rosner, *Proc. Natl. Acad. Sci. USA* **92**, 5456 (1995).
11. M. Nardini and B. W. Dijkstra, *Cur. Opin. Struct. Biol.*, **9**, 732 (1999).
12. L. Holm and C. Sander, *TIBS* **20**, 478 (1995).
13. M. Bartlam, G. Wang, H. Yang, S. Gao, Y. Feng, and Z. Rao, *Structure*, **12**, 1481 (2004).
14. A. Hershko, H. Heller, E. Eytan, G. Kyaklij, and I. A. Rose, *Proc. Natl. Acad. Sci. USA*, **81**, 7021 (1984).
15. E. M. Hollams, K. M. Giles, A. M. Thomson, and P. J. Leedman, *Neurochemical Res.*, **27**, 957 (2002).
16. Y. Zuo and M. P. Deutscher, *Nucl. Acids Res.*, **29**, 1017 (2001).
17. S. Ghosh and M. P. Deutscher, *Proc. Natl. Acad. Sci. USA*, **96**, 4372 (1999).
18. S. Thore, F. Mauxion, B. Seraphin, and D. Suck, *EMBO Reports*, **4**, 1150 (2003).
19. T. Horio, M. Murai, T. Inoue, T. Hamasaki, T. Tanaka, and T. Ohgi, *FEBS Letters*, **577**, 111 (2004).
20. W. A. Breyer and B. W. Matthews, *Nature Struct. Biol.*, **7**, 1125 (2000).
21. W. Weber, H. Zahner, J. Siebers, K. Schroder, and A. Zeeck, *Arch Microbiol.*, **121**, 111 (1979).

22. H. Decker and C. R. Hutchinson, *J. Bacteriol.*, **175**, 3887 (1993).
  23. B. Shen and C. R. Hutchinson, *A. Science*, **262**, 1535 (1993).
  24. W. Bao, E. Wendt-Pienkowski, and C. R. Hutchinson, *Biochemistry*, **37**, 8132 (1998).
  25. J. M. Dunwell, *Biotechnol Genet Eng. Rev.*, **15**, 1 (1998).
  26. J. M. Dunwell, A. Purvis, and S. Khuri, *Phytochemistry*, **65**, 7 (2004).
  27. J. M. Dunwell, A. Culham, C. E. Carter, C. R. Sosa-Aguirre, and P. W. Goodenough, *TIBS*, **26**, 740 (2001).
  28. S. Khuri, F. T. Bakker, and J. M. Dunwell, *Mol. Biol. Evol.*, **18**, 593 (2001).
  29. L. Jaroszewski, *et al.*, *Proteins*, **56**, 611 (2004).
  30. R. Schwarzenbacher, *et al.*, *Proteins*, **56**, 392 (2004).
  31. M. M. Benning, G. Wesenberg, R. Liu, K. L. Taylor, D. Dunaway-Mariano, and H. M. Holden, *J. Biol. Chem.*, **273**, 33572 (1998).
  32. J. B. Thoden, Z. Zhuang, D. Dunaway-Mariano, and H. M. Holden, *J. Biol. Chem.*, **44**, 43709 (2003).
  33. S. Blachin-Roland, S. Blanquet, J.-M. Schmitter, and G. Fayat, *Mol. Gen. Genet.*, **205**, 515 (1986).
  34. F. Leveque, S. Blanchin-Roland, G. Fayat, P. Plateau, and S. Blanquet, *J. Mol. Biol.*, **20**, 319 (1990).
  35. M. M. Gibson, C. G. Bagga, C. G. Miller, and M. E. Maguire, *Mol. Microbiol.*, **5**, 2753 (1991).
  36. M. G. Rossmann, D. Moras, and K. W. Olsen, *Nature*, **250**, 194 (1974).
  37. R. K. Wierenga, P. Terpstra, and W. G. Hol, *J. Mol. Biol.*, **187**, 101 (1986).
  38. K. Nakai and P. Horton, *Trends in Biochem. Sci.*, **24**, 34 (1999).
  39. M. Aldea, T. Garrido, M. Hernandez-Chico, M. Vicente, and S. R. Kushner, *EMBO Reports*, **8**, 3923 (1989).
  40. J. M. Santos, P. Freire, M. Vicente, and C. M. Arraiano, *Molecular Microbiology*, **32**, 789 (1999).
  41. T. Kasai, M. Inoue, S. Koshiba, T. Yabuki, M. Aoki, E. Nunokawa, *et al.*, *Protein Sci.*, **13**, 545 (2003).
- 
- 周三和先生為美國西雅圖華盛頓大學博士，現任國立中興大學生化研究教授。
  - Shan-Ho Chou received his Ph.D. from the University of Washington in Seattle, USA. He is currently a professor in the Institute of Biochemistry at National Chung Hsing University.