

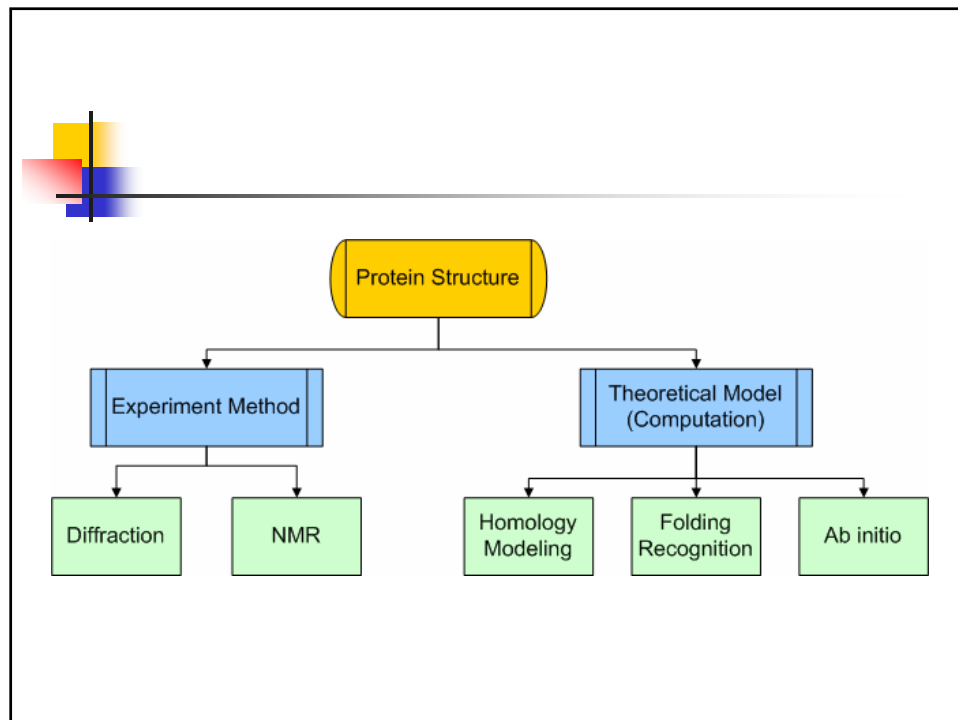
蛋白質結構模擬

Protein Structure Modeling

Ping-Chiang Lyu

Institute of Bioinformatics and Structural Biology,
Department of Life Science,
National Tsing Hua University

97/07/22





Experimental Method

目前利用實驗方法得到蛋白質的結構有下列兩種方法，但是這些方法有個共同的缺點，就是需要較多的時間，才能得到結構。

方法	原理	缺點
X-光繞射 (X-ray Diffraction)	利用X-光線繞射的特性，對已經結晶的蛋白質進行繞射實驗，然後再將所得到的數據加以分析，就可以得到結構。	其缺點是要得到蛋白質的結晶是實驗中最重要也是最困難的。
核磁共振方法(NMR)	利用核磁共振現象得到蛋白質的結構。並不是所有原子核都能產生核磁共振的現象。一般常用來偵測的對象包含了 ^1H 、 ^{13}C 和 ^{15}N 。	其一方法的缺點是只能對小的蛋白質來作分析。



Theoretical Model

- 因此，科學家就想利用電腦計算方式來獲得蛋白質結構，以加快定出蛋白質三度空間結構的速率。
- 由於目前已知資料庫中，完整地蛋白質序列數量已遠遠超過所解出的蛋白質結構數量，而且利用實驗方法解出蛋白質結構不是太耗時間，就是受到一些限制而無法定出結構，因此，利用電腦計算的方式來計算出蛋白質結構已逐漸受到矚目。大致可分為下列三種方式：



Theoretical Model

方法	原理	缺點
同源模擬法 Homology Modeling	將未知的目標序列跟蛋白質結構資料庫(PDB)作序列比對，來尋找出最好的模板(Template)，來當作模型將序列穿進去，然後再作最佳化的運算，來獲的目標序列的蛋白質結構。	遇到序列相似度很低的時候，就無法預測或是預測的結果可信度很低。
摺疊辨識法 Folding Recognition	直接將目標序列套上已知結構的相似序列，然後作能量的計算 (MD, SA, Energy Minimize)，然後依照序列順序跟結構排列的法則，來找出符合要求的分子。	需要大量的計算量，且只能針對蛋白質的核心作預測。
重頭起算法 Ab initio	利用熱力學的原理，考慮氨基酸和溶液的所有交互作用力，從一級結構開始，來計算出蛋白質的三級結構。	此方法完全靠電腦計算，而且需要大量的計算。

接下來我們將對理論模型中最可靠的方法——同源模擬法——來做進一步的講解。



Homology Modeling (同源模擬法)

- 同源性模擬法(Homology modeling)又可稱為比較性模擬法(comparative modeling)或知識基礎性模擬法(knowledge-base modeling)。因為這一個方法可以快速模擬出蛋白質的三度空間結構。因此，目現已有利用這個方法模擬出整個酵母菌(yeast)基因體的所有蛋白質結構。
- 同源性模擬法的基本假設是特定的胺基酸序列會構成特定的蛋白質結構。主要是利用現存已解出的結構為模板(由核磁共振(NMR)或是X光繞射(X-ray diffraction)所解出的結構)，模擬出未知結構蛋白質序列 (protein sequence)的三度空間結構。一般來說，當所要解的蛋白質序列(target protein)和模板(template)之間的序列相似度越高，所模擬出來的結構越正確也越可信。



合理的蛋白質分子模型必須滿足下列條件：

- 與X光晶體學或多維度NMR光譜實驗數據相互比較後，可相互驗證。
- 和蛋白質分子之生物特性實驗結果相符。
- 提供詳細的資料，以供分子結構和功能相關研究用。



Homology Modeling (同源模擬法)

- 一旦模擬出蛋白質結構並建立合理的分子模型後，進一步利用結構來探討和功能相關的胺基酸，再配合實驗方法加以驗證，不但可以大幅縮減實驗嘗試的時間，也可以大幅降低實驗成本。

Homology Modeling 分成六個步驟：

Step1:

選擇參考的蛋白質分子(Template)。

Step2:

目標分子(Target)和參考分子的胺基酸序列比對。

Step3:

建立目標分子核心部份的分子骨幹。

Step4:

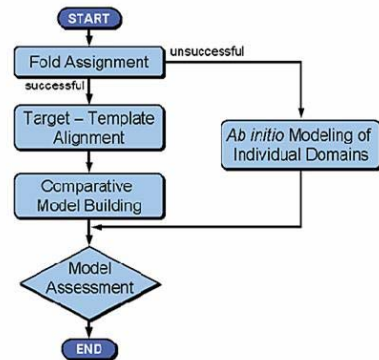
產生各結構守恆區域之間鬆散分子鏈的結構。

Step5:

目標分子結構修正微調。

Step6:

三維分子結構的檢驗和證實。



Nat. Struct. Biol. **7**, 986-990, 2000

Step1:

選擇和目標蛋白質分子相關的參考蛋白質分子

- 考慮兩個條件：
 1. 目標分子和參考分子間序列相似的程度
 2. 參考蛋白質分子已知結構的準確性
- 若在目標蛋白質所屬的家族中，有一個以上的蛋白質結構已被解出，且目標蛋白質和參考蛋白質序列相似程度在25%以上時，一般而言，此方法的成功機率相當的高。
- 若沒有精確的蛋白質結構可供參考，或序列相似度低於25%，此方法的成功機率就不高，此時需利用其他實驗數據的協助，例如是否參考和蛋白目標蛋白的功能是否類似。

例如 HIV-1 (HIV-RT) 中 reverse transcriptase, (反轉錄)的 polymerase 區域，唯一已知結構的參考蛋白質是 E.Coli的 DNA polymerase (Pol-I)，雖然兩者序列類比程度低於 25%，但兩者具有極相近的功能而且由 affinity-labelling chemical modification 實驗中確認出關鍵的對應胺基酸lysine，提供重要的證據支持以 Pol-I 的結構做為建立HIV-RT模型的參考分子，因為目前只有Pol-I的 α 碳原子座標，因此HIV-RT模型的準確度受到 Pol-I α 碳原子準確度的影響。

Step2:比對目標和參考蛋白質分子的胺基酸序列

- 整個程序中最關鍵性的步驟，胺基酸序列比對即決定目標分子序列中對應於核心分子骨幹及鬆散分子鏈的各個段落，對產生正確三維結構有關鍵性的影響。
- 當序列相似程度大於50%時，很容易將兩序列對齊。相似程度低於50%時，則需一些額外的資料才能得到可靠的對齊序列，困難之處在於確認序列中適當的基準點。這基準點必須都存在於參考和目標蛋白質分子胺基酸序列之中，並具存在結構和功能上的重要性，這些基準點提供兩蛋白質分子正確的結構對應關係，找尋基準點的方法有幾種：
 1. 將參考及目標分子與同一系統內其他蛋白質分子序列排列比對，以各分子序列對應相同序列段落做為基準點
 2. 將數個可能的參考分子的結構重疊，確認出雙硫鍵cystein的位置
進行特定位置的胺基酸更換 (site-directed mutagenesis) 在參考和目標蛋白質分子做 affinity-labelling chemical modification 實驗等等，都有助於基準點的確證。
- 在對齊兩胺基酸序列時，只有在轉折 (turn) 及鬆散分子鏈 (loop) 的區域內才可進行胺基酸段落的插入、刪除和更換，厭水性的胺基酸應限制在蛋白質分子核心部份的分子骨幹上。

Step3:產生目標分子核心部份的分子骨幹

- 利用參考分子的核心部份分子骨幹的結構為基礎，將胺基酸換成目標分子對應位置的胺基酸。
- 核心部份分子骨幹的胺基酸的更換只會改變二級結構元素的相對位置和取向，而不致於破壞這些結構元素或三級摺疊的一般特性，而且在這個區域要做的胺基酸插入或刪除的機會很小，修改後可再利用位能函數來調整各胺基酸的位置。為了避免造成分子核心部份分子骨幹結構的重大改變，這部份的修改是逐步進行的，一個一個胺基酸分別處理。在需要進行胺基酸插入或更換時，同時要決定該胺基酸支鏈部份的分子構形。做胺基酸更換時，被更換胺基酸支鏈部份的一個或多個扭角 χ 的值是採用原參考分子對應胺基酸支鏈部份的對應扭角值，支鏈中其他部份的扭角值則由蛋白質結構資料庫中找尋最合理的值代入。在插入額外胺基酸時，胺基酸支鏈部份的所有扭角都要指定適當的數值，當所有原子都有指定的座標值之後，再以分子力學方法能量最小化程序消除可能的原子重疊的情況。

Step4: 產生鬆散分子鏈部份的結構

兩種方法

1. 資料庫裏搜尋具有相同胺基酸序列的片段並借用其座標資料直接代人
2. 利用能量計算方法直接產生合理的原子座標。

- **資料庫搜尋方法**，能產生短分子鏈（4-6 個胺基酸）或由重複出現的基本結構式樣構成的中等大小的分子鏈之結構。適用範圍受限於蛋白質結構資料庫內已知結構的數量、種類和多樣性。小段的分子序列本身即具較小的分子構型空間，並且出現的機會較多，比較適合這種方法。
- **能量計算方法**直接估算這分子片段可用的分子構形空間中重要的區域，以系統式或隨機式的程序進行分子構形搜尋，用這種方式可處理到長度為9的胺基酸序列，並且可產生和已知相似結構以外的分子構形，增加結構的多樣性。處理較長的分子鏈時，由於系統化的分子構形搜尋中各自由度排列組合產生極大的構形空間，能量計算方法需要很大的計算資源。為了降低計算量，這部份的**計算可分成兩個階段**，

能量計算方法

- 第一個階段中先忽略不計胺基酸的支鏈部份（glycine 和 proline 的支鏈部份保留到B-碳原子為止），即皆當作 alanine 處理，先在這分子骨幹的分子構形空間內做能量最小化計算，找到合理的分子骨幹構形後，
- 第二階段中將支鏈部份結構考慮進來，以能量最小化的方式計算各支鏈部分的分子構形。在做能量最小化計算時，為了考慮鄰近支鏈部份的影響，需將所有支鏈同時處理。為保證得到適當的低能量分子構形，可用模擬退火法（Simulated Annealing），
- 因蛋白質分子較鬆散部份的分子鏈出現在分子的表層部份，在決定這部份分子構形時要考慮和溶劑分子的作用。
- 實際上，資料庫搜尋和能量計算方法是合併使用的。先由資料庫提供近似的結構再以能量最小化或分子運動方法做調整。在搜尋的過程中也可加入能量的計算以增加搜尋的效率。



Step5: 修正最後的分子結構

- 在產生鬆散部份的分子鏈結構階段可能得到數個合理的分子構形，若不止一段的分子鏈各具幾種可能的分子構形，則要考慮各段落分子構形的組合構形，再進行整體的結構調整和修正。若鬆散部份的結構來自資料庫，資料庫現存的資料可能無法適當代表可用的分子構形空間，因此要在這個階段利用分子運動學方法來修改分子結構，並借此辨出能量最低分子構形。



Step6: 蛋白質分子結構的驗證

- 所建立的分子模型正確與否必須和該蛋白質的實驗數據做比較。例如溶劑分子的可接觸性，胺基酸堆積密度、厭水性或帶電荷或極性胺基酸的位置，分子靜電位場分佈和溶解自由能等等實驗測量值。
- 另外用分子運動學方法來探測是否所得到的分子構形是停留在穩定的能量最小的區域或是停留在很淺的局部能量最小區域，由 epitope mapping, site-specific mutagenesis, affinity-labelling, chemical modification 等實驗所得的實驗數據也有助於檢驗預測的分子模型是否合理。
- 以上這六個步驟是一個週期，一般而言需要反覆幾次的預測和校正檢驗的程序，最後才得到可靠的分子模型。



結論：

- 同源性建模方法預測蛋白質的三維結構，有其限制：
 - 蛋白質分子中可供作參考的蛋白質結構及式樣的多樣性很有限。
 - 在同源性很低時，兩胺基酸序列比對，以及在鬆散分子鏈部份的三維結構模擬都有待進一步改進和發展。
 - 在簡化位能函數及提供計算效率及處理溶劑效應方面仍有待努力。



應用：

- 此一個方法可以應用在蛋白質突變研究 (mutation)、活性位置研究 (active site) 與藥物設計等。能夠減少在實驗上、製藥上所須的時間。



軟體：

SWISS-MODEL	Server	http://www.expasy.org/swissmod/SWISS-MODEL.html http://www.expasy.ch/swissmod/SWISS-MODEL.html
WHAT IF	Program	http://www.cmbi.kun.nl/whatif/
MODELLER	Program	http://guitar.rockefeller.edu/modeller/modeller.html
InsightII	Program	http://www.accelrys.com/about/msi.html
SDSC1	Server	http://cl.sdsc.edu/hm.html
PredictProtein	Server	http://cubic.bioc.columbia.edu/predictprotein/ http://biobug.life.nthu.edu.tw/predictprotein/



SWISS-MODEL



<http://www.expasy.org/swissmod/SWISS-MODEL.html>



SWISS-MODEL

SWISS-MODEL 是網路上提供以Homology Modeling的原理來預測蛋白質結構的網路服務，起源自一個瑞士日內瓦所進行的實驗計劃。

模式(Mode)：

MENU

Modelling requests:

- [First Approach mode](#)
- [Optimise \(project\) mode](#)
- [Oligomer modelling](#)
- [GPCR mode](#)

First Approach mode :	輸入序列，經過伺服器的電腦搜尋、比對、計算等，自動由E-mail傳回結果
Optimise (Project) mode :	將 First Approach mode 的結果，用 Swiss-PdbViewer (觀看蛋白質的軟體) 加以檢視及修改的檔案，再做結果修正後的預測結構。
Oligomer modelling :	提供你如何針對多鍵的結構來利用上面兩種模式來做預測的步驟。
GPCR mode:	若預測的蛋白質為7個Helix的方式透過細胞膜的受體(Receptor)，便可透過此一方式，建立模型。

通常我們都是用第一個模式來獲的結構，再進行其他的處理，所以我們在此特別對First Approach mode來做範例：

SwissModel First Approach Mode

Please fill these fields:

Your Email address: (MUST be correct!)

Your Name:

Request title: Will be added to the results header.

[PDBViewer](#) (?) my name and email for next login.

Provide a sequence or a SWISS-PROT AC code:

NOTES: A SWISS-PROT AC code looks like this: **P04406**
Sequences can be provided in either RAW, SWISS-PROT, FASTA or GCG format

Now [Send Request](#) or [Reset Form](#)

輸入(Input)：

個人電子郵件(Your Email address)	最後得到的結果會傳到這個電子郵件，通常會寄給你三封信，分別是： <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>Welcome_to_SwissModel</td> <td>這封是他們所接受到的資料</td> </tr> <tr> <td>SwissModel-TraceLog-XO</td> <td>所選取的模板及所有的計算的過程</td> </tr> <tr> <td>SwissModel-Model-XO</td> <td>最後計算所得到的結果(PDB file)</td> </tr> </table>	Welcome_to_SwissModel	這封是他們所接受到的資料	SwissModel-TraceLog-XO	所選取的模板及所有的計算的過程	SwissModel-Model-XO	最後計算所得到的結果(PDB file)
Welcome_to_SwissModel	這封是他們所接受到的資料						
SwissModel-TraceLog-XO	所選取的模板及所有的計算的過程						
SwissModel-Model-XO	最後計算所得到的結果(PDB file)						
個人名字(Your Name)	沒有重要的用途，只是會在郵件上打上你的名字						
此次使用的標題(Request title)	這次你所預測結構的標題，通常是蛋白質的名字或是編號。						
提供的序列或是SWISS-PORT AC碼 (Provide a sequence or a SWISS-PORT AC code)	這裡有兩種方式來輸入你所想要預測的蛋白質結構： <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>Sequence</td> <td>未知結構的蛋白質序列</td> </tr> <tr> <td>SWISS-PORT AC code</td> <td>SWISS-PORT為一個序列的資料庫，可以直接利用它們的編號來當作序列輸入</td> </tr> </table>	Sequence	未知結構的蛋白質序列	SWISS-PORT AC code	SWISS-PORT為一個序列的資料庫，可以直接利用它們的編號來當作序列輸入		
Sequence	未知結構的蛋白質序列						
SWISS-PORT AC code	SWISS-PORT為一個序列的資料庫，可以直接利用它們的編號來當作序列輸入						

選項 (Options) :

Options:

- Define the lower BLAST P(N) limit for template selection:

Lower BLAST limit:

- Define the templates you wish to use for this request:

In some cases you may wish to define a set of template structures to be used for a modelling attempt. As an example, modelling a serpin (Serine esterase inhibitor similar to the plasminogen activator inhibitor, antithrombin III, etc) will generally fail, since these proteins have two distinct structures in the database:

- the activated form of all true serpins
- the precursor form as found in the serpin-analogue Ovalbumin.

It is thus best to chose the correct template(s) you wish to base your model on.
To do so you may use ExPDB templates, your own templates, or a combination of both.

- Using ExPDB templates:
 - Search for suitable templates in the ExNRL-3D database.
 - Select the entries you consider appropriate from the hit list and/or check if their codes exist in the ExPDB database.
 - List up to 5 entries in the window below, separated by space.

NOTE: The ExPDB database is derived from the PDB database. Each chain is in a separate file and the residues have been renumbered consecutively. The ExPDB codes are built from the PDB codes as explained [here](#).
- Provide one or more of your own templates:

Until further notice, modelling templates must comply with these guidelines:

- The format MUST be PDB.
- The files MUST contain only ONE chain.
- The files should contain no HETATM records.

Template file 1:

Template file 2:

Template file 3:

Template file 4:

Define the lower BLAST P(N) limit for template selection :

目標和模板序列相似度在低於 P(N) 值下的才考慮選做模板，P(N) 值越底代表此兩序列越相似。

Define the templates you wish to use for this request :

自己選取想要的模板，此時你可以選擇一個或一個以上它們所比對出來的或是自己擁有的蛋白質結構來當做模板，如果要它們所比對出來的模板，需要輸入ExPDB的碼，如果是用自己擁有的模版的話，需要是符合PDB的格式，且只能是單鏈(ONE chain)的不能有其他不是蛋白質的分子在裡面(HETATM)。

- Results options: (The SWISS-MODEL server will return all results via Email)

<input checked="" type="radio"/> Swiss-PdbViewer mode	will return the final model and the template(s) as a Swiss-PdbViewer project file and a log file tracing all actions taken by the server
<input type="radio"/> Normal mode	includes the final model co-ordinates file in PDB format and a log file tracing all actions taken by the server
<input type="radio"/> Short mode	will return only the final model co-ordinates file

Send the results as plain ASCII mail instead of email attachment.

Include a WhatCheck report of the final model.
- The SWISS-MODEL server can forward your sequence to the following services:
 - PHD Secondary structure prediction: at Columbia University
 - 1. Do not forward the sequence
 - 2. Forward the sequence only if no modelling templates can be found
 - 3. Forward the sequence in any case
 - Fold Recognition Server (3D-pssm) of the ICRF.
 - 1. Do not forward the sequence
 - 2. Forward the sequence only if no modelling templates can be found
 - 3. Forward the sequence in any case

Results options:

傳回結果的選項主要有三種：

Swiss-PdbViewer mode	以PdbViewer的檔案形式傳回來，包含預測及模板的結構
Normal mode	以PDB的檔案形式傳回來，只有預測的結構
Short mode	以每個氨基酸的座標傳回來

另外兩個額外的選項分別是以文字形式的郵件傳回結果，而不是以附加檔的郵件形式；另一個是會將你預測的結果用What If 這個軟體來做分析，並會多寄一封郵件來告訴你分析的結果。

The SWISS-MODEL server can forward your sequence to the following services:

還額外提供自動幫你的序列傳到別的结构預測網站：

PHD Secondary structure prediction	提供序列做預測二級結構的功能
Fold Recognition Server (3D-pssm)	提供以Fold Recognition的方法來預測三級結構

在這兩選項的下面分別是：

1. Do not forward the sequence	不要轉寄序列。
2. Forward the sequence only if no modelling templates can be found	當沒有模板時，轉寄序列。
3. Forward the sequence in any case	轉寄序列在任何狀況下。



利用 Swiss-PdbViewer 直接寄送任務到 SWISS-MODEL SERVER

Homology modeling 的實做方法：

Homology modeling可分成兩種：全自動及手動調整。

- 全自動Homology modelling: (前一章節)
直接連線到[SWISS-MODEL Server](#)，點選[First Approach mode](#)，將蛋白質序列貼上，並填上正確個人資料，便可以得到Modeling的結果。
- **手動調整Homology modelling：**
可以利用Swiss-Pdb viewer進行序列比對的調整或是進行site-directed mutagenesis 之後，再直接寄送到SWISS-MODEL Server，進行Homology modeling。

利用Swiss-Pdbviewer手動調整蛋白質序列後，
進行Homology modelling

在這個例子中，將以FAS antigen ligand為示範，建立其homology model:

啓動Swiss-Pdbviewer，並將蛋白質序列（本例使用FAS antigen ligand; SWISSPROT entry P41047）載入：

DeepView / Swiss-PdbViewer 3.7 (SP4)

File Edit Select Build Tools Fit Display Color Preferences SwissModel Window Help

Load New Sequence to Model... (highlighted with a red arrow)

Load FoldFit Alignment...

Save FoldFit Alignment...

Ignore Selected AA during modelling

Use Selected AA during modelling

Drag Residues to Ignore as "X"

Set current layer as reference

Move new sequence into structure

Move structure into raw sequence

Lock Selected Residues of Model

Unlock Selected Residues of Model

Homo Multimer Mode

Build Preliminary Model

Save Optimize Model Job

Update Threading Display Automatically

Update Threading Display Now

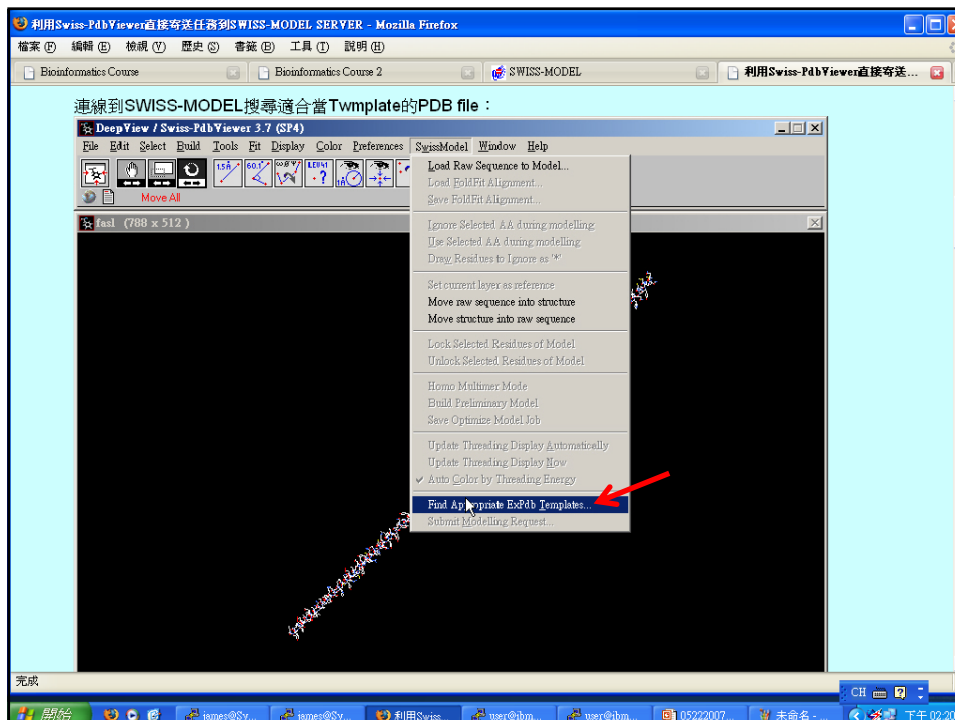
Auto Color by Threading Energy

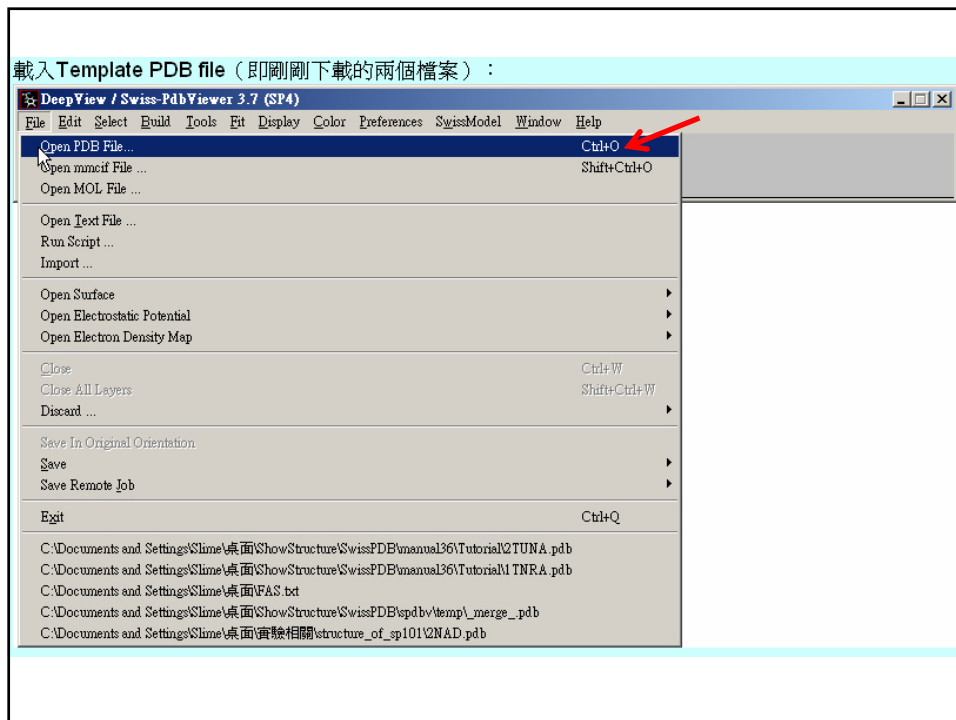
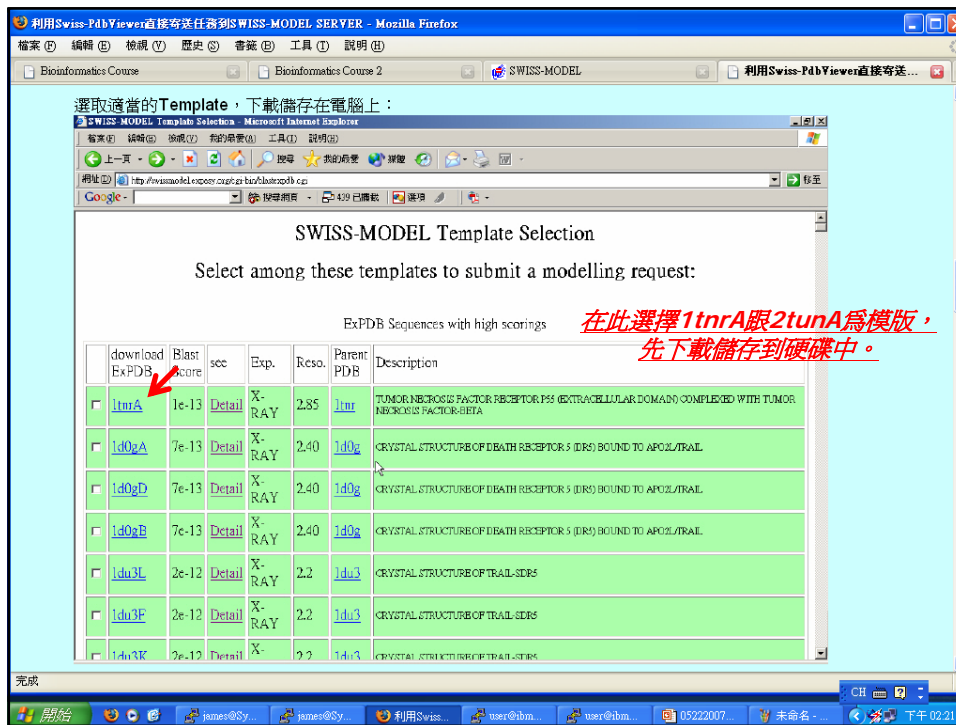
Find Appropriate EsPdb Templates...

Submit Modeling Request...

蛋白質序列的內容如下

```
>sp|P41047|TNFL6_MOUSE Tumor necrosis factor ligand superfamily member 6 (Fas antigen ligand) (Fas ligand) (CD178 antigen)
MQQPMMNYPCPQIIFWVDSSATSSWAPPGSVFPCPCSPGRGPDQRPPPPPPVWSPLPFPPSQ
PLPLPPLTLPLKKKDHNTNLWLPVVFVFLVALVGMGLGMYQLFHLQKELAELEPTNQSLS
KVSSFEKQIANPSTPSEKKEPRVAHLTGMPHSRSLPLEWEDTYGTALISGWKYYKGGGLV
INETGLYFVYSKYVFRGQCENQPLMHKVMRNSKYPEDLVLMEEKRLNVCYTGQIWAHS
SYLGAVFNLTSADHLVNI SQLSLINFEESKTFPGLYKL
```





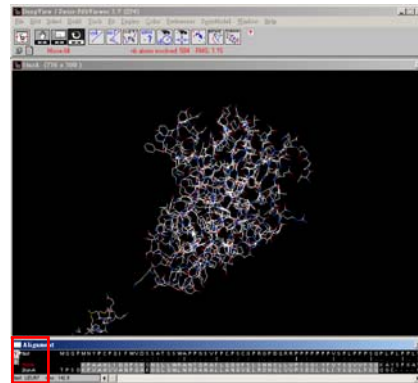
將兩條template蛋白質序列進行alignment：

Window → alignment → 點選2tnrA

Fit → Magic fit → 會出現一個對話視窗，告知以1tnrA 為 reference → 點選ok

Fit → Generate structural alignment

註：若出現錯誤訊息"These structure seems to be too badly....."，請先把fasl sequence layer關掉，再進行structure alignment，之後再將fasl sequence load 進來。



Target seq與template seq 序列比對

記得在進行原始序列跟 template 進行比對前，先勾選下列兩點：

- "SwissModel"->"Update Threading Display Automatically" 和 "SwissModel"->"Auto Color by Threading Energy"

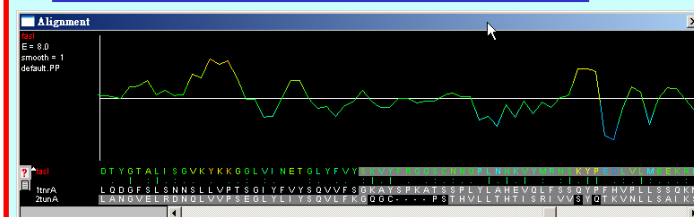


將原始序列跟template進行alignment:

- 點選fasl->"Fit"->"Magic Fit"

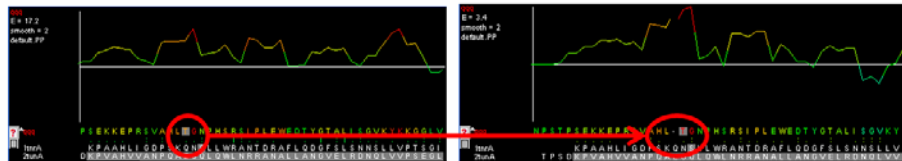
進行手動調整：

- "SwissModel"->"Update Threading Display now"
- 點選Alignment Window中，fasl旁的白色三角形符號
- 調整Mean force potential energy至綠色（紅色代表能量高，綠色代表能量低）





先選取一個胺基酸如 **L**，緊接著按“空白鍵”和“Backspace 鍵”就可調整，調整至能量降低為止。



確認修改過的alignment不會造成空間上的碰撞：

- "Select"->"aa making clash"
- "Tool"->"Fix selected sidechain"
- 重複以上兩個步驟，直到沒有互相碰撞的sidechain

設定電子信箱

