



Biotechnology

生物技術

Hong, Tang-Yao assistant professor

洪堂耀 助理教授

任課時間：2009.09~2010.01



人類基因體定序計畫

✚ 國際人類基因組團隊版

(International Human Genome Consortium)

● 國家衛生研究院 NIH (National Institute of Health)

● 美國國家能源部

● 英國 Sanger 中心 (衛康信託基金會贊助)

● 德國、法國、日本、中國.....等十五個國家

✚ 負責總單位美國國家人類基因研究院

(National Human Genome Research Institute)



約翰 薩爾斯頓
John E. Sulston



➤ 人類染色體的結構

■ 23對染色體

■ 染色體的組成

- DNA：雙股線狀

- 蛋白質：蛋白質包括組織蛋白(histone)與非組織蛋白(non-histone)，其中DNA與蛋白質重量約相等

- RNA



2000年所公布人類基因組 草圖結果

	塞雷拉版	國際人類基因組團隊
基因組樣本	5位不同種族人體	數名匿名捐贈者
定序鹼基長度	99%連續	90% (43%不連續)
定序基因數目	26,383	23,147
主要定序方法	散彈定序法	大片段逐步定序 傳統 + 物理性圖譜
SNPs數量	約兩百一十萬	約一百四十萬
耗時	9個月	14年
計劃主持人	Craig Venter	Francis Collins

凡特與柯林斯



2003年 人類基因組計畫宣布完成

標記	目標	結果	完成日期
遺傳圖譜	2 to 5cM (600 – 1,500標記)	1-cM (3,000 標記)	1994.9
物理圖譜	30,000 STSs	52,000 STSs	1998.10
DNA序列	95% gene	finished to 99.99%	2003.4
耗費	< \$0.25 base	<\$0.09 base	2002.11
DNA變異	100,000 SNPs	3.7 million SNPs	2003.1
基因	Full-length cDNAs	15,000 cDNAs	2003.3



人類真染色體序列定序完成

articles

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

** A list of authors and their affiliations appears in the Supplementary Information*

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

NATURE | VOL 431 | 21 OCTOBER 2004 | www.nature.com/nature

©2004 Nature Publishing Group



人類基因組資訊

✚ 基因組含有大約30.8億個核苷酸

- 真染色體(euchromatic)約有28.8億個核苷酸
- 整個基因組含有341空缺
 - 異染色質區域有33 空缺(1.98億個核苷酸)
 - 真染色質區域有308 空缺(2800萬個核苷酸)

✚ 人類基因數

- 20,000-25,000 (含22,287 基因座)
- 34,214 轉錄產物
- 外顯子有3400萬個核苷酸，佔 1.2%
- 非轉譯區2100萬個核苷酸，佔 0.7%
- 19,599 蛋白質編碼基因
- 2,188 DNA 片段預測為蛋白質編碼基因

真染色質與異染色質

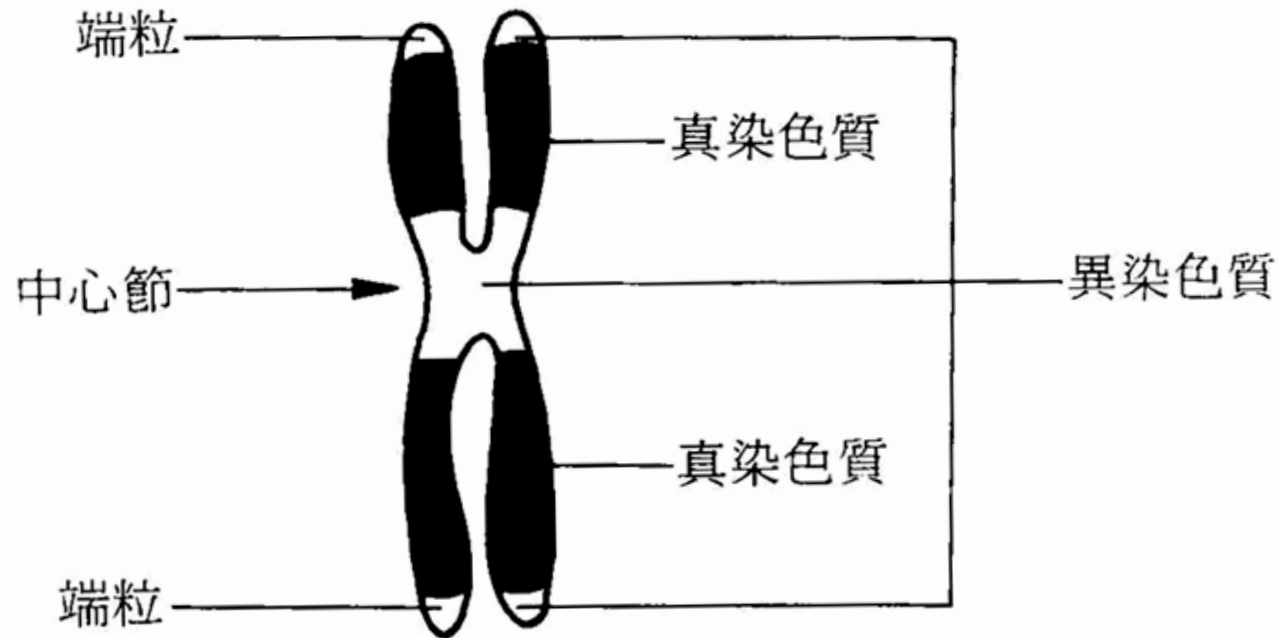
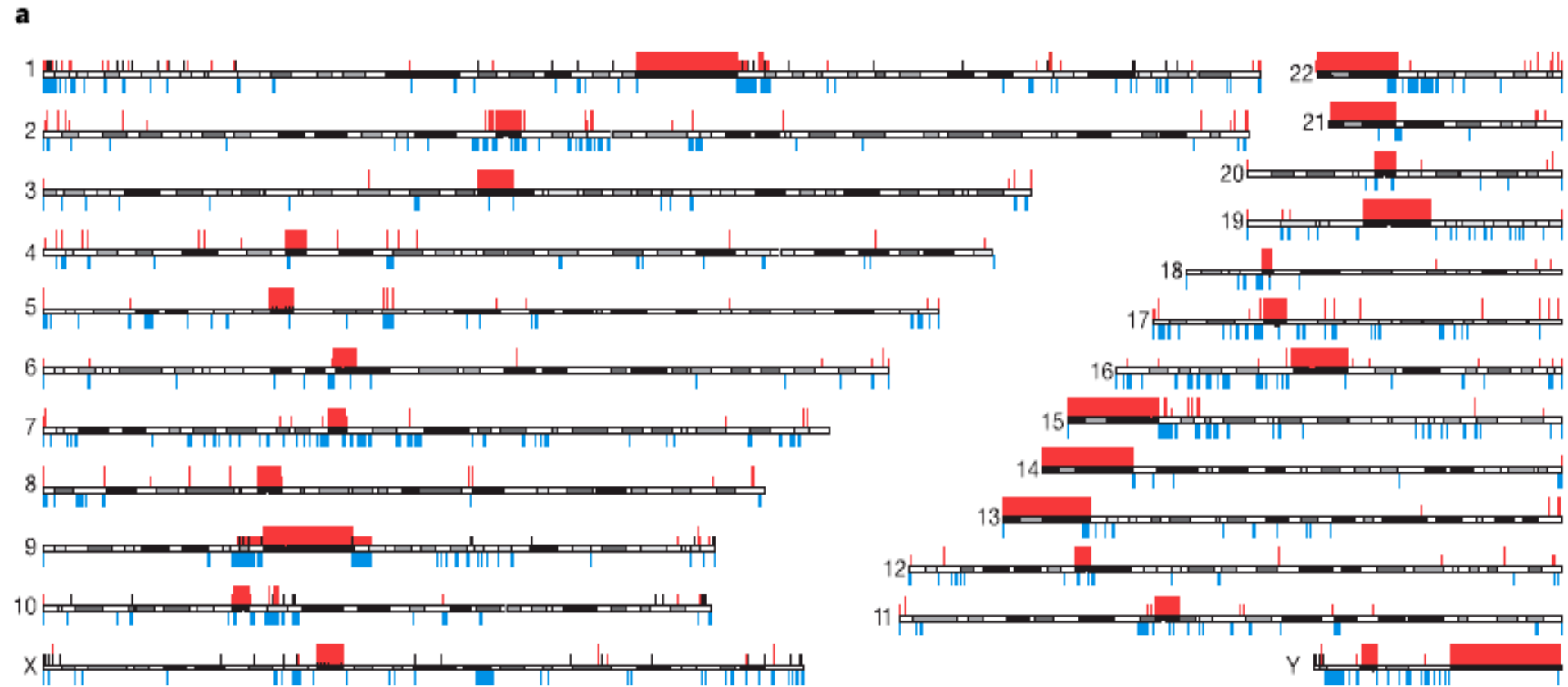


圖 6-2 染色體上中心節(centromere)，端粒(telomere)，真染色質(euchromatin)，異染色質(heterochromatin)的分布位置

(蔡嘉芬, 2006)

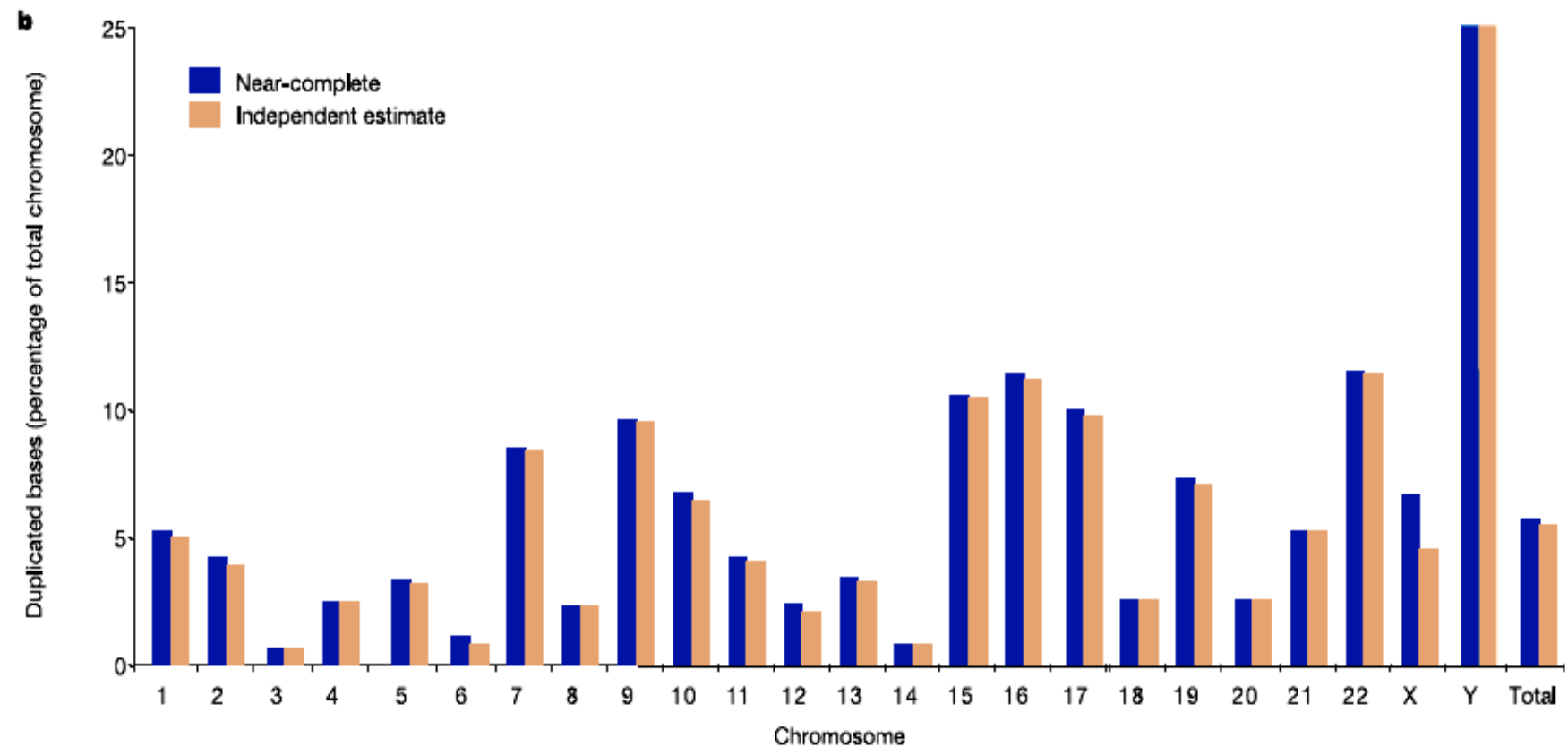
人類基因組定序的空缺位置



藍色：人類基因體中重複性片段在不同染色體中的分佈

紅色：人類基因體目前定序的空缺部分(gap)

各染色體重複性片段所佔的百分比圖



藍色：HGP的結果

橘色：散彈槍定序的結果

生物的基因體大小一

物種	百萬核苷酸
阿米巴原蟲類 <i>Amoeba dubia</i>	670,000
阿米巴原蟲類 <i>Amoeba proteus</i>	270,000
狹葉瓶爾小草 <i>Ophioglossum petiolatum</i>	160,000
非洲肺魚 <i>Protopterus aethiopicus</i>	139,000
川貝母 <i>Fritillaria assyriaca</i>	124,900
鐵炮百合 <i>Lilium longiflorum</i>	90,000
雙指兩棲蟾 <i>Amphiuma means</i>	84,000
泥鰍 <i>Necturus maculosus</i>	81,300
蘇苔松類 <i>Pinus resinosa</i>	68,000
台灣百合 <i>Lilium formosanum</i>	36,000
圓篩藻屬 <i>Coscinodiscus</i>	25,000
蟋蟀 <i>Triturus cristatus</i>	20,600

生物的基因體大小二

物種	百萬核苷酸
洋蔥 <i>Allium cepa</i>	18,000
沙漠蝗蟲 <i>Schistocerca gregaria</i>	9,300
草履蟲 <i>Paramecium caudatum</i>	8600
盤古蟾蜍 <i>Bufo bufo</i>	6900
貓鯊屬 <i>Scyliorhinus</i> [軟骨魚綱]	6000
土豚 <i>Orycteropus afer</i>	5763
鼠類 <i>Peromyscus eremiticus</i>	5295
原猴科 <i>Tarsius syrichta</i>	5152
猴科 <i>Cercopithecus cepus</i>	5142
猴科 <i>Cercopithecus nigroviridis</i>	5117
玉米 <i>Zea mays</i>	5000
人類 <i>Homo sapiens</i>	3080



不同生物體預估的基因數

1995 嗜血桿菌 1.83 Mb, 1740 基因

 徽漿菌 0.58 Mb, 517 基因

1996 酵母菌 12 Mb, 6000 基因

 古生菌 1.7 Mb, 1738 基因

1997 大腸桿菌 4.6 Mb, 4289 基因

2001 線蟲 97 Mb, 20262 基因(草圖 1998)

2002 河豚 365 Mb, 30000 基因

 小鼠 2500 Mb, 25000 基因(草圖 2002)

2003 水稻 430 Mb, 50000 基因

2003 果蠅 128 Mb, 13600 基因(草圖 2000)

2003 人類 3000 Mb, 25000 基因(草圖 2000)



功能基因體學

■ 基因與基因、身體結構與運作、人與生活環境

■ 研究目的

- 轉錄因子與基因

- 基因 where? when? 有作用

- 基因在生物體的功能

■ 最佳方式

- 疾病和發病過程基因型的變化

個人基因體定序

- HGP計畫花費3億美金

- 第一位個人基因體定序-華生

 - 2007.5.29 454生命科學公司 454 Life Sciences

 - 花費100萬美金

 - 近程目標降至10萬美金

 - 終極目標降至1000美元

- 第一位雙套染色體定序者J. Craig Venter J. Craig Venter

The Diploid Genome Sequence of an Individual Human

Samuel Levy^{1*}, Granger Sutton¹, Pauline C. Ng¹, Lars Feuk², Aaron L. Halpern¹, Brian P. Walenz¹, Nelson Axelrod¹, Jiaqi Huang¹, Ewen F. Kirkness¹, Gennady Denisov¹, Yuan Lin¹, Jeffrey R. MacDonald², Andy Wing Chun Pang², Mary Shago², Timothy B. Stockwell¹, Alexia Tsiamouri¹, Vineet Bafna³, Vikas Bansal³, Saul A. Kravitz¹, Dana A. Busam¹, Karen Y. Beeson¹, Tina C. McIntosh¹, Karin A. Remington¹, Josep F. Abril⁴, John Gill¹, Jon Borman¹, Yu-Hui Rogers¹, Marvin E. Frazier¹, Stephen W. Scherer², Robert L. Strausberg¹, J. Craig Venter¹

1 J. Craig Venter Institute, Rockville, Maryland, United States of America, 2 Program in Genetics and Genomic Biology, The Hospital for Sick Children, and Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada, 3 Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, United States of America, 4 Genetics Department, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

Presented here is a genome sequence of an individual human. It was produced from ~32 million random DNA fragments, sequenced by Sanger dideoxy technology and assembled into 4,528 scaffolds, comprising 2,810 million bases (Mb) of contiguous sequence with approximately 7.5-fold coverage for any given region. We developed a modified version of the Celera assembler to facilitate the identification and comparison of alternate alleles within this individual diploid genome. Comparison of this genome and the National Center for Biotechnology Information human reference assembly revealed more than 4.1 million DNA variants, encompassing 12.3 Mb. These variants (of which 1,288,319 were novel) included 3,213,401 single nucleotide polymorphisms (SNPs), 53,823 block substitutions (2–206 bp), 292,102 heterozygous insertion/deletion events (indels)(1–571 bp), 559,473 homozygous indels (1–82,711 bp), 90 inversions, as well as numerous segmental duplications and copy number variation regions. Non-SNP DNA variation accounts for 22% of all events identified in the donor, however they involve 74% of all variant bases. This suggests an important role for non-SNP genetic alterations in defining the diploid genome structure. Moreover, 44% of genes were heterozygous for one or more variants. Using a novel haplotype assembly strategy, we were able to span 1.5 Gb of genome sequence in segments >200 kb, providing further precision to the diploid nature of the genome. These data depict a definitive molecular portrait of a diploid human genome that provides a starting point for future genome comparisons and enables an era of individualized genomic information.

Citation: Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, et al. (2007) The diploid genome sequence of an individual human. PLoS Biol 5(10): e254. doi:10.1371/journal.pbio.0050254



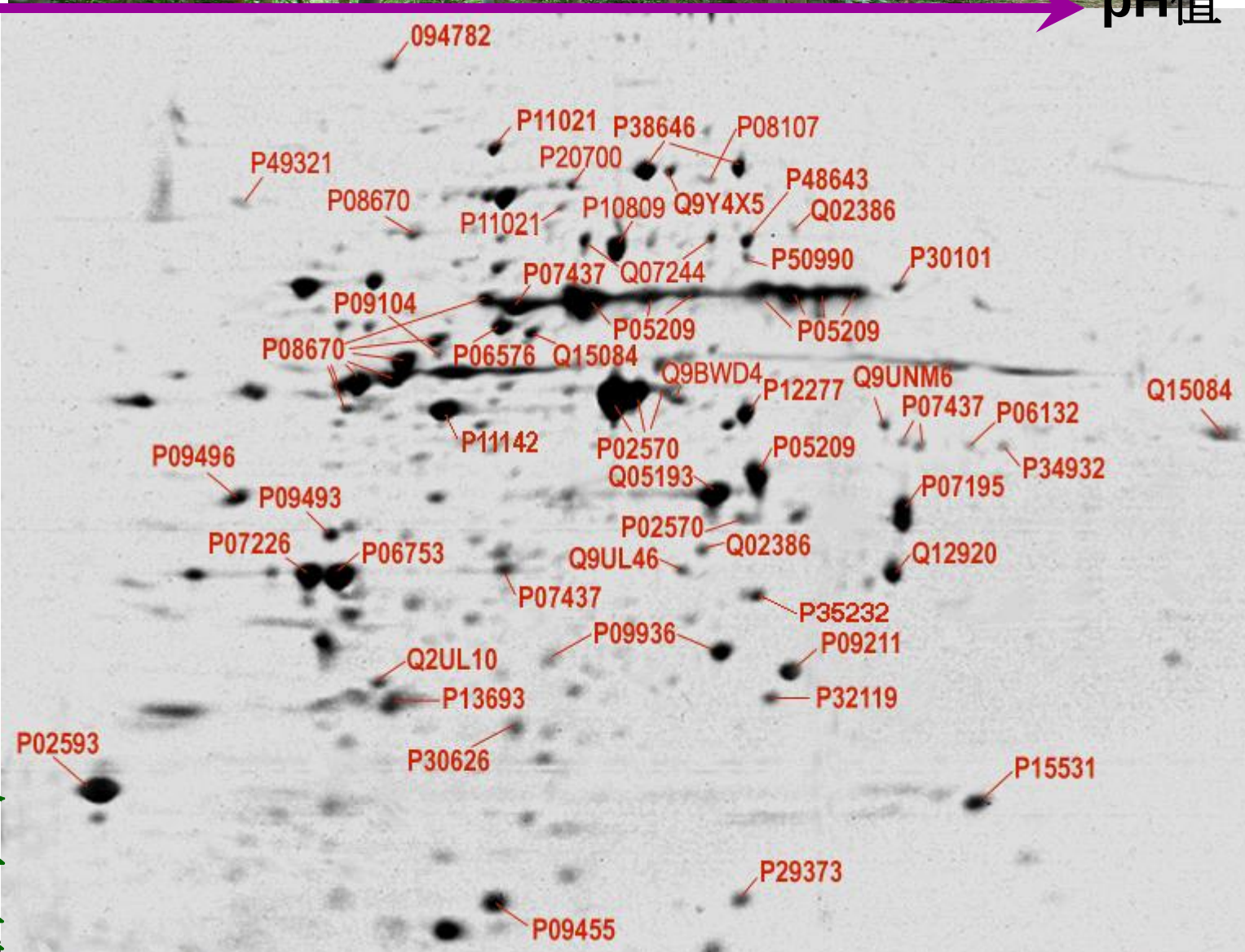
蛋白質體學

- ✚ 定義：細胞內所有的蛋白質
- ✚ 目標：DNA-胺基酸序列→電腦→結構與功能
- ✚ 蛋白質表現的預測？
 - 修飾、加工
- ✚ 人類蛋白質體組織(2001)
- ✚ 範疇
 - 表現蛋白質體學(圖譜)
 - 細胞圖譜蛋白質體學
 - 確定位置
 - 質譜儀→鑑定其組成→交互作用

二維蛋白質電泳圖

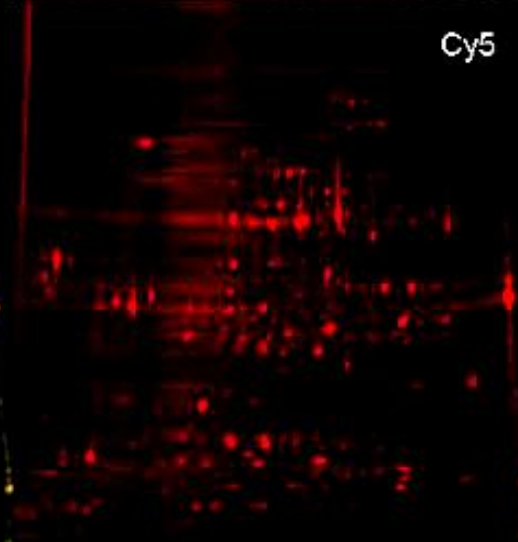
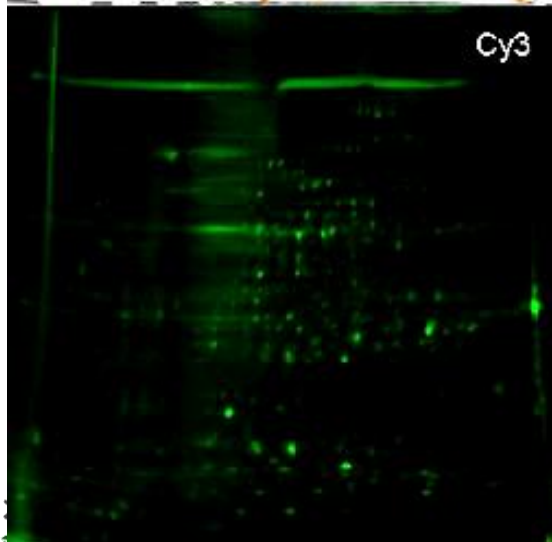
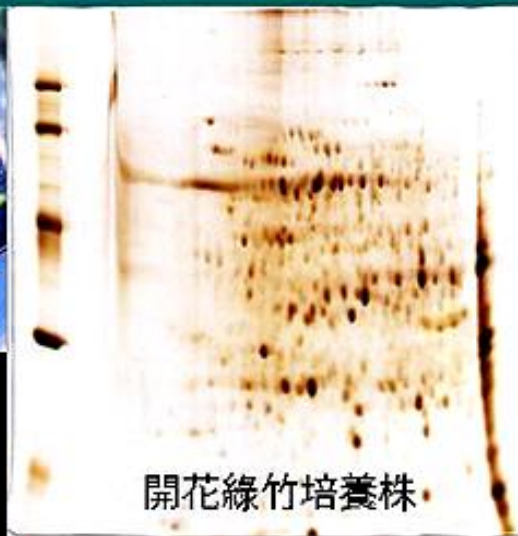
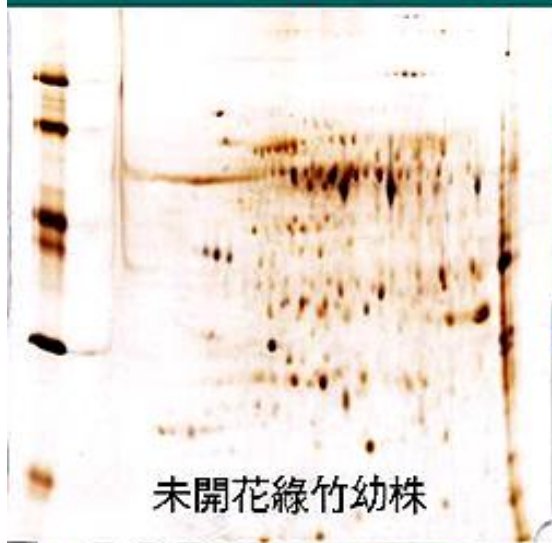
pH值

分子量



不同螢光區別蛋白質體圖譜

綠竹幼株開花前後的蛋白質表現差異





質譜術

質譜儀

化學分子離子化



帶電荷



外加電場(動能)



真空管中飛行



撞擊偵測標靶產生訊號
(質量愈大飛行速度越慢)

氣相質譜儀



飛行時間質譜儀 MALDI-TOF MS

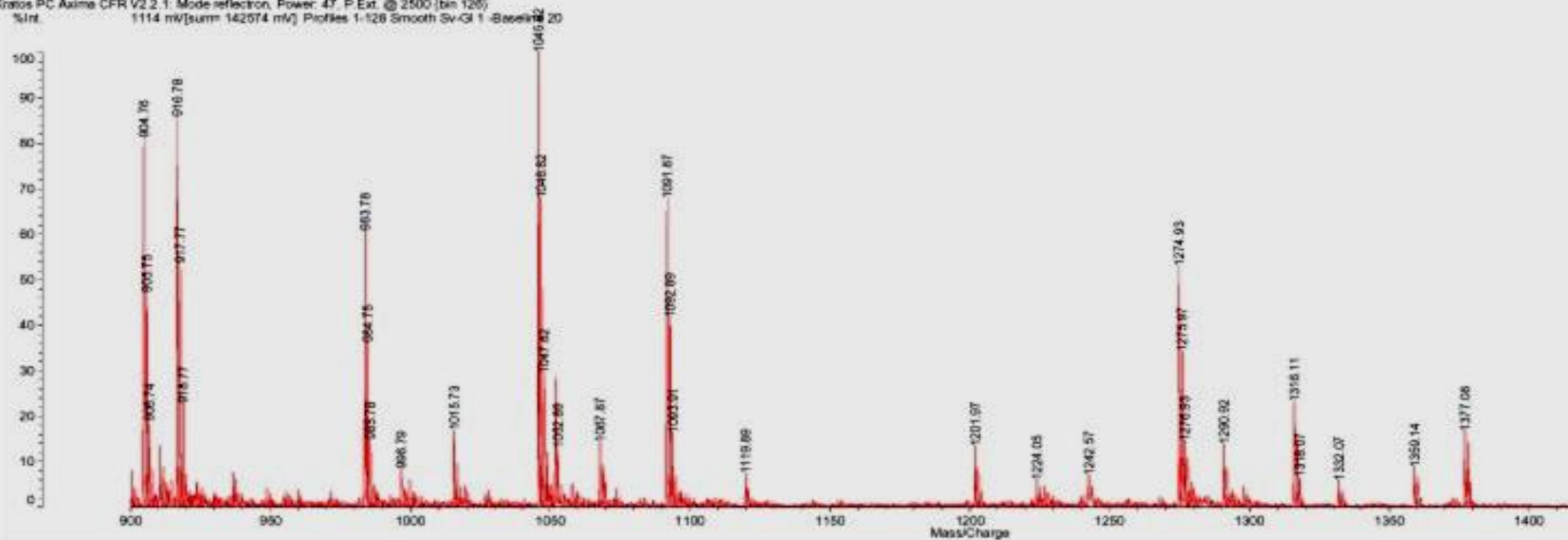
✚ 可獲得胜肽質量的資訊



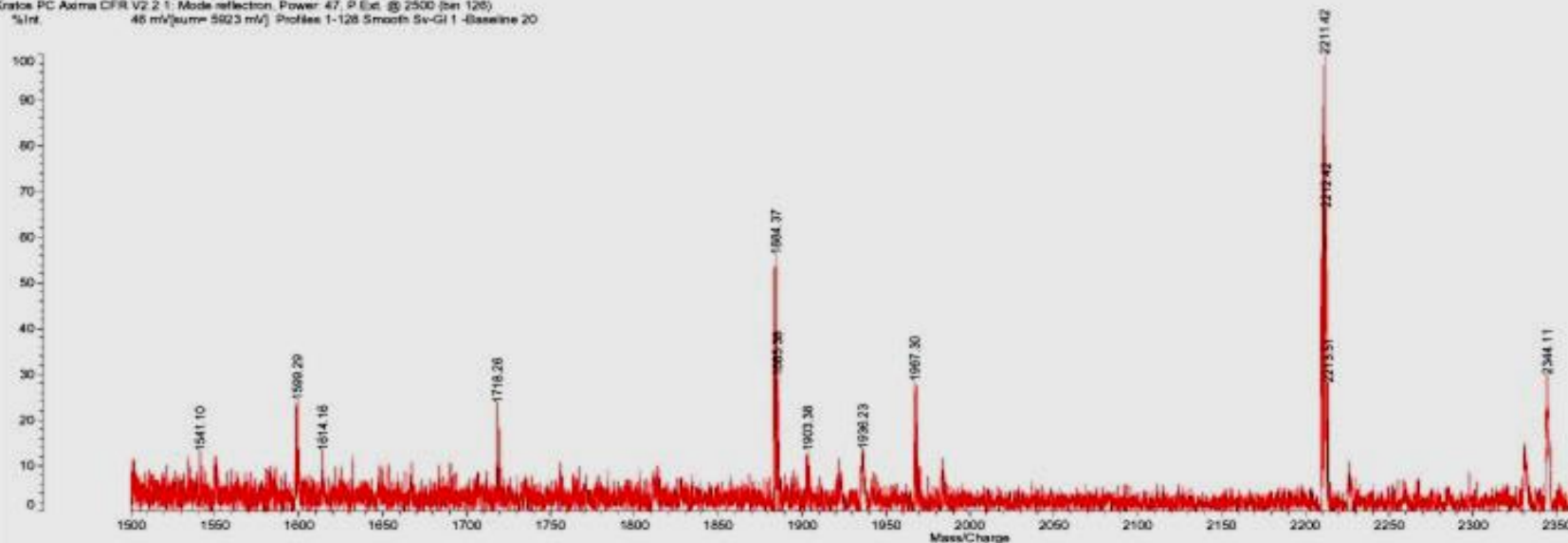
Applied Biosystems 4700 MALDI TOF-TOF
參考資料：<http://proteomics.missouri.edu>

飛行時間質譜儀資料比對

030304
Kyusyu a49+spot 6
Data: 1000117 4 Mar 2003 19:34 Cal: 030304 4 Mar 2003 19:33
Kratos PC Axima CFR V2.2.1: Mode: reflectron, Power: 47, P Ext. @ 2500 (bin 126)
%Int. 1114 mV[sum: 142674 mV] Profiles 1-126 Smooth Sv-Gl 1-Baseline 20



030304
Kyusyu a49+spot 6
Data: 1000117 4 Mar 2003 19:34 Cal: 030304 4 Mar 2003 19:33
Kratos PC Axima CFR V2.2.1: Mode: reflectron, Power: 47, P Ext. @ 2500 (bin 126)
%Int. 46 mV[sum: 5923 mV] Profiles 1-126 Smooth Sv-Gl 1-Baseline 20



電噴灑游離串聯式質譜儀 ESI-Tandem MS

✚ 可獲得胜肽詳細的資料（可知胺基酸序列）



（參考資料：<http://www.rzuser.uni-heidelberg.de>）

分析蛋白質種類的方法

- 二維電泳

- 胰蛋白酶切

- 質譜儀分析

- 資料庫比對

- 利用氣體分子撞擊胺基酸片段

➤ 研究蛋白質體學的新興技術

- 成本高、普及率低



蛋白質結構之基礎

► 蛋白質性質

- 胺基酸性質

- 蛋白質組成

- 蛋白質功能

 - 生物反應的催化作用、儲存、運輸

 - 傳遞訊息、抗體、調控、結構

► 蛋白質分離與純化

- 分離與純化流程、樣品備製、蛋白質溶解性

- 蛋白質沈澱、蛋白質變性、蛋白質電泳分離

- 酵素水解、蛋白質層析



➤ 蛋白質鑑定

- ✚ 蛋白質定序
- ✚ 蛋白質基因選殖
- ✚ 蛋白質體學
- ✚ MAS 鑑定蛋白質法
- ✚ 蛋白質序列與功能分析

➤ 蛋白質結構

- ✚ 蛋白質結構介紹、蛋白質結構及功能分析
- ✚ 蛋白質結構決定法: NMR方法、X-ray 繞射方法
- ✚ 蛋白質結構模擬、蛋白質晶体學的應用



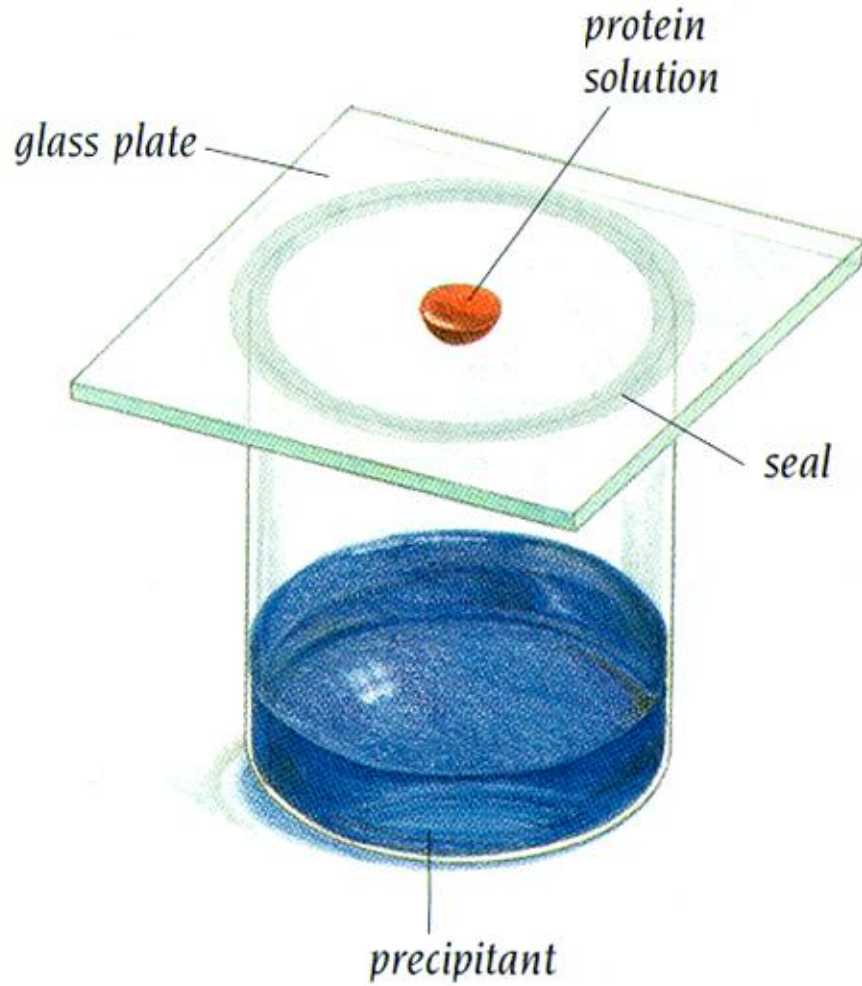
➤ 蛋白質的純化

✚ 防止蛋白質變性失去活性

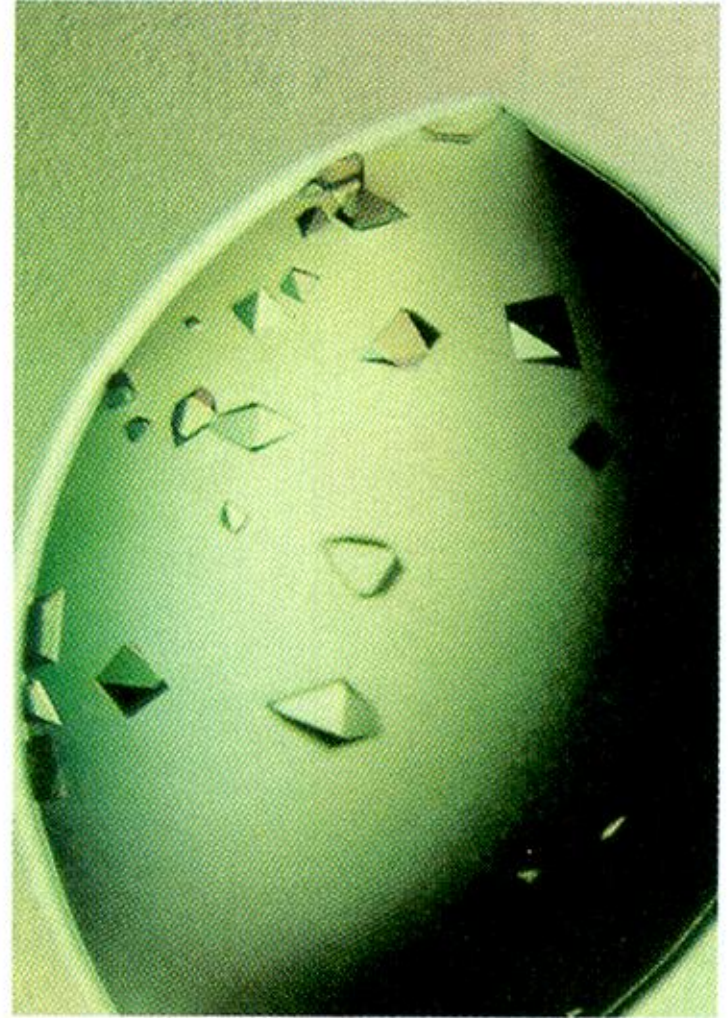
- 大部分於低溫操作（4°C 冷房）
- pH 4-10之間操作
- 減少泡沫的產生
- 注意重金屬、有機溶劑、微生物污染、
- 蛋白質水解酵素

✚ 選擇有效的純化方法

✚ 純化過程中持續追蹤蛋白質活性

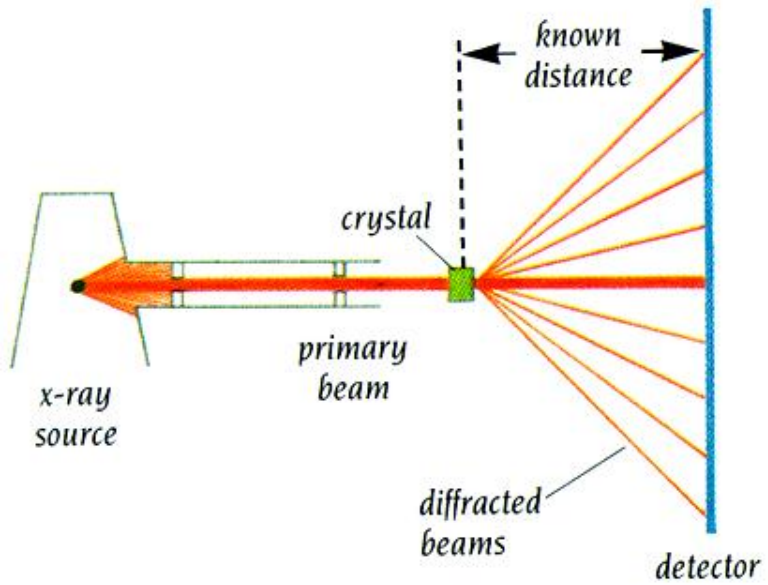


(a)

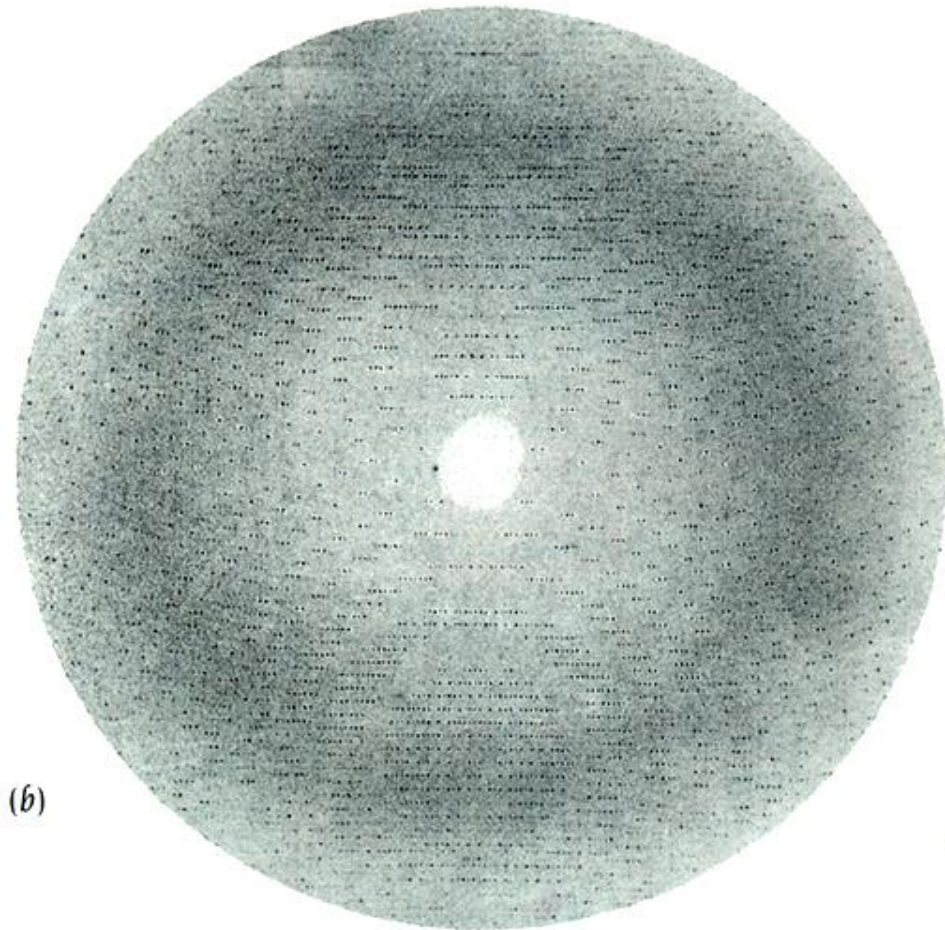


(b)

資料來源Petsko ,2003



(a)



(b)

(a) 5.0 Å

(b) 3.0 Å

(c) 1.5 Å

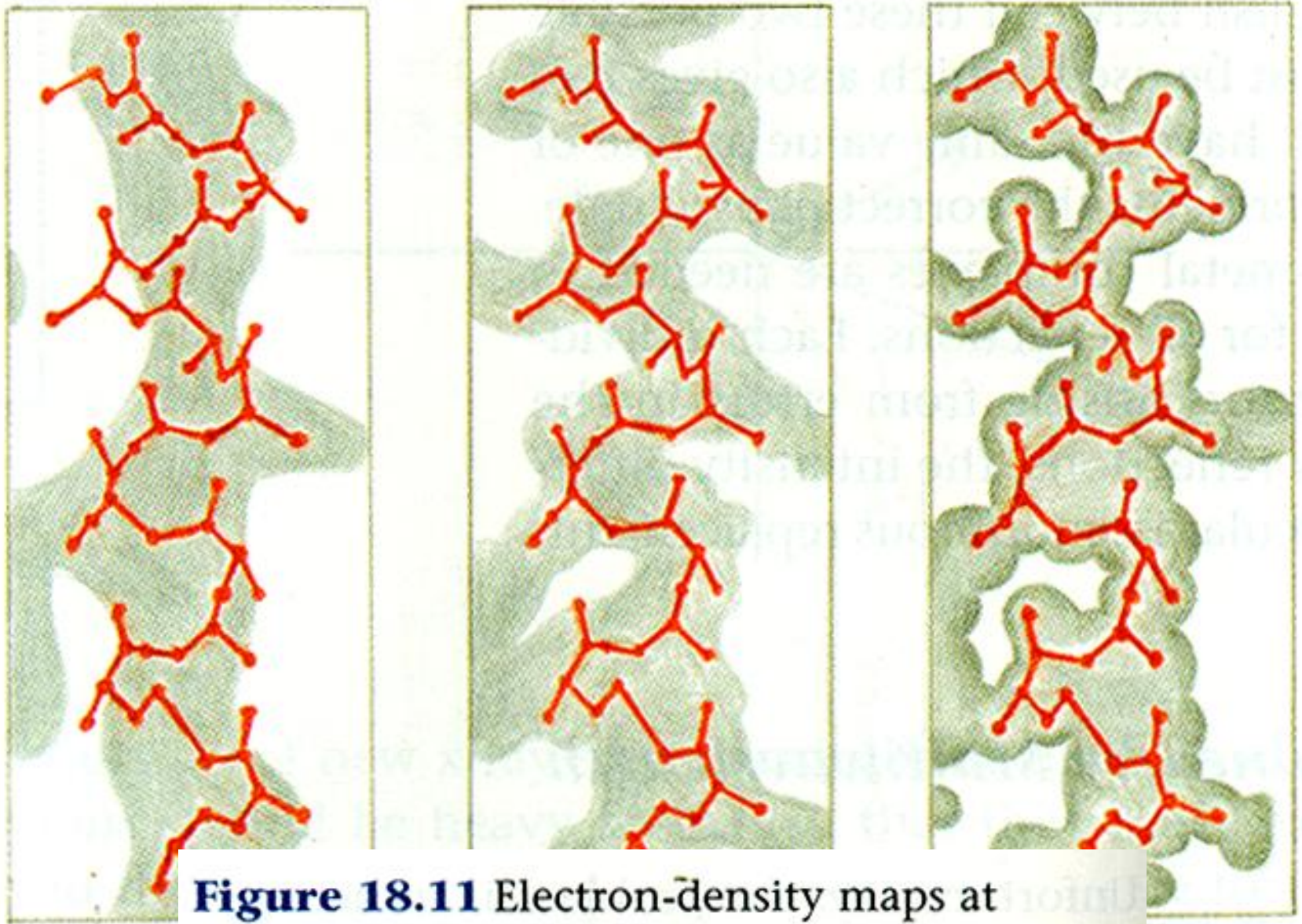
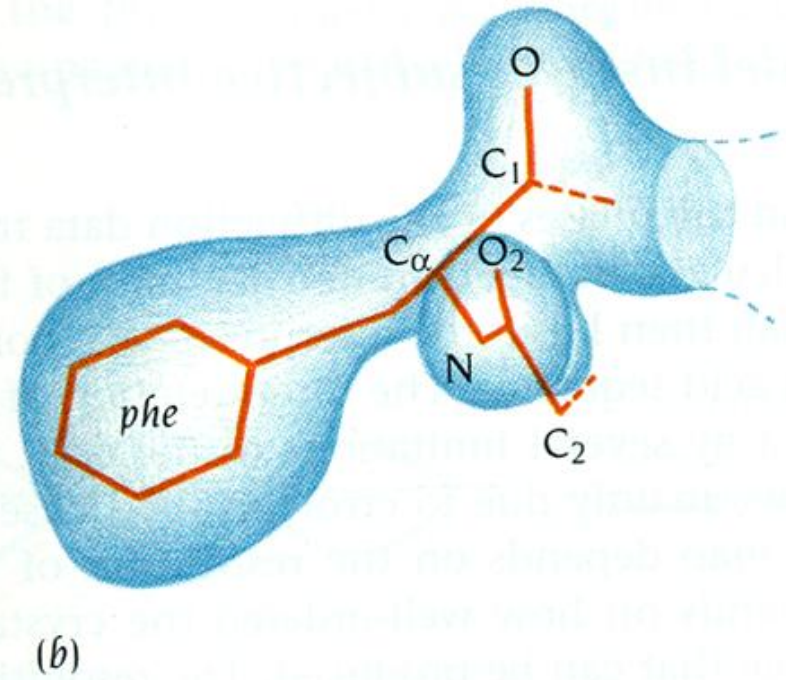
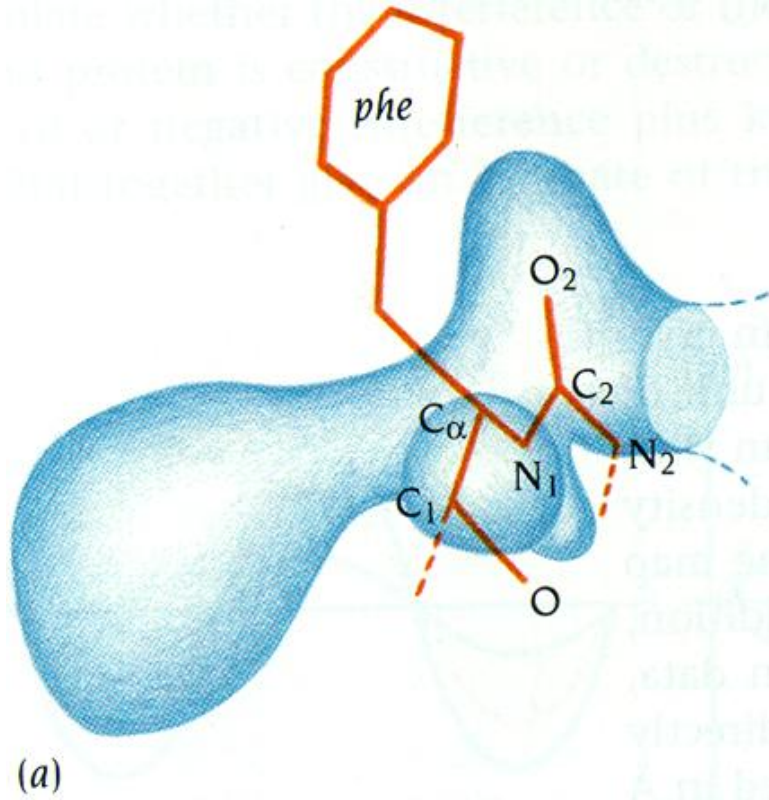


Figure 18.11 Electron-density maps at different resolution show more detail at higher resolution. (a) At low resolution (5.0 Å)

資料來源Petsko, 2003



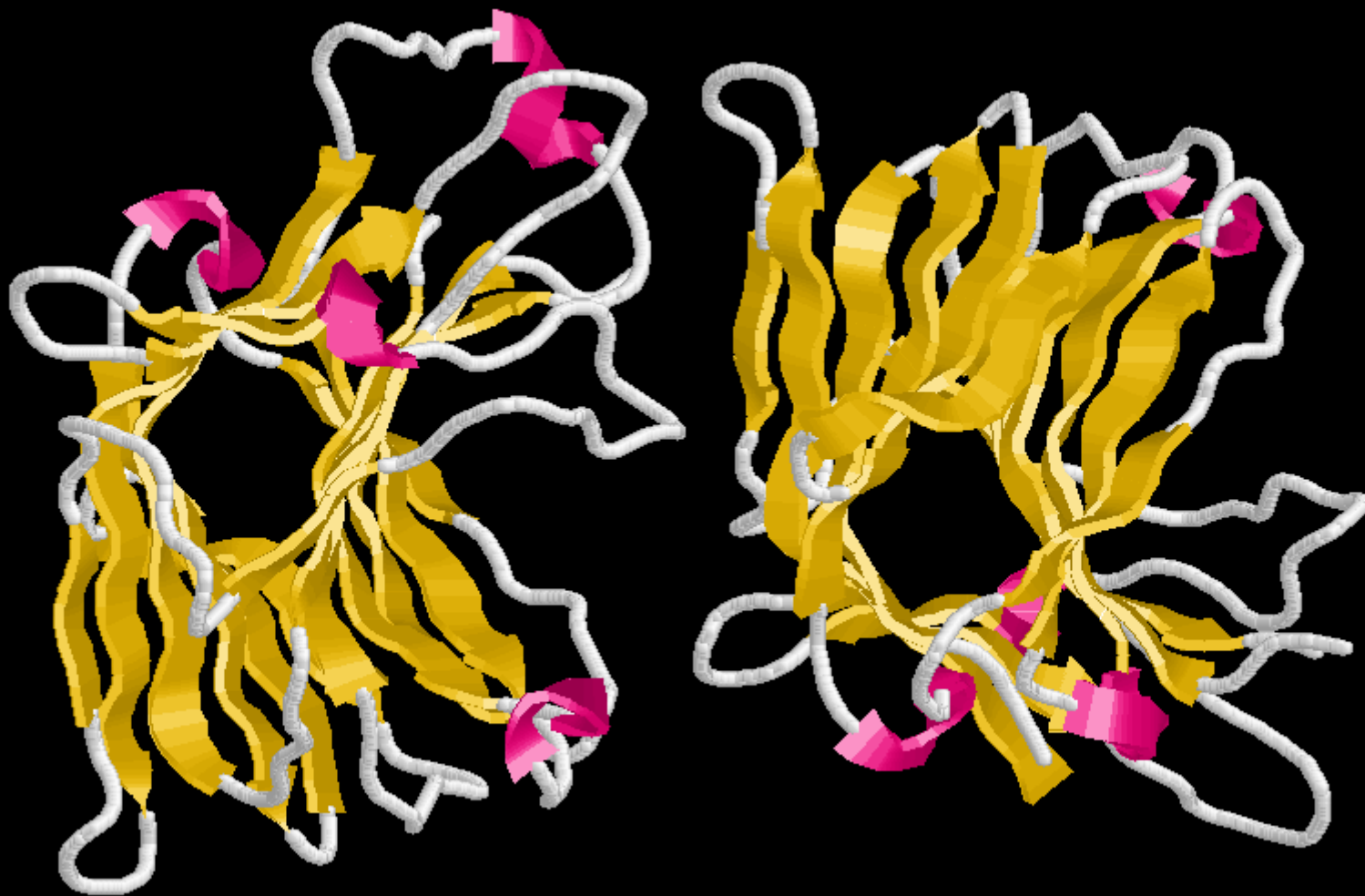
資料來源Petsko ,2003

觀察蛋白質結構軟體

➤ 免費軟體

- **Rasmol**: 顯示速度快，但較難學。
- **Cn3D**: 閱讀 MMDB 資料時，可直接顯示預設之觀察模型
- **Swiss PDB Viewer**: 可將多個分子重疊比較，而且可配合 ExPASy 上建立模型的工具使用
- **Chemscape ChimeTM**: 瀏覽器上直接觀察
- **WebLab Viewer?**: 雖較慢，可是較易學習，而且又能將模型存為各種不同的模式，適合一般生物學者使用

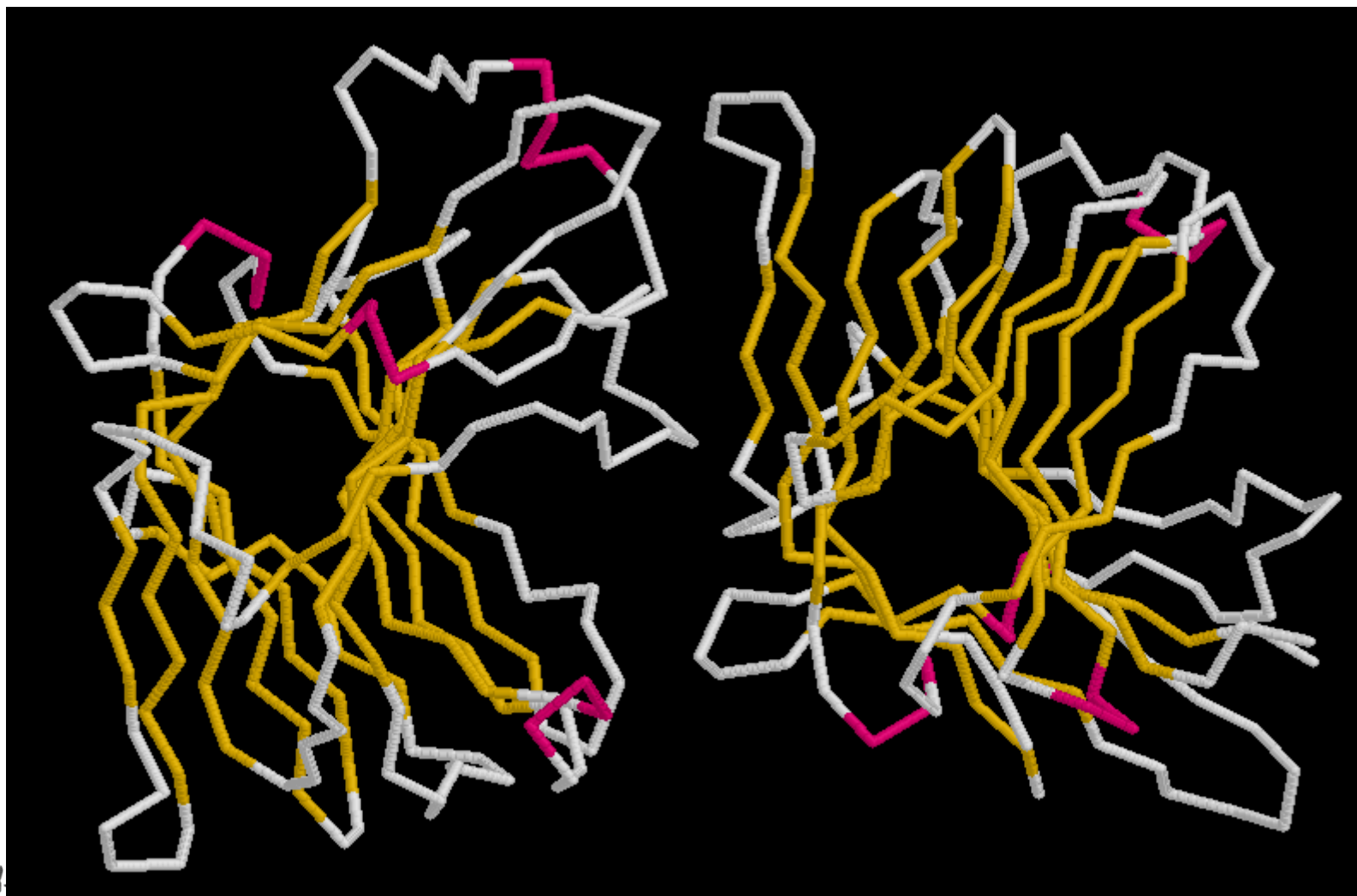
蛋白質的立體結構一



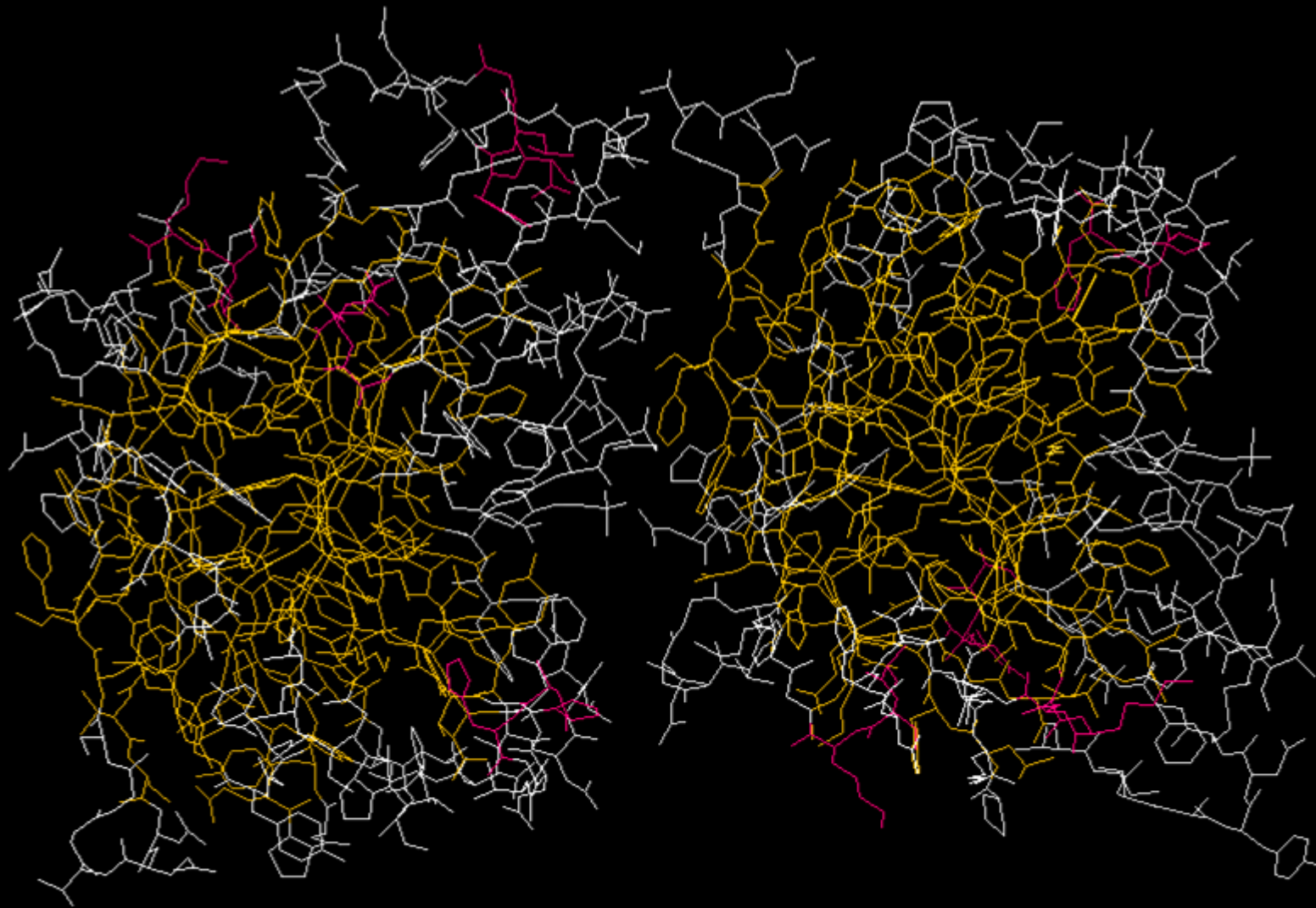
蛋白質的立體結構二



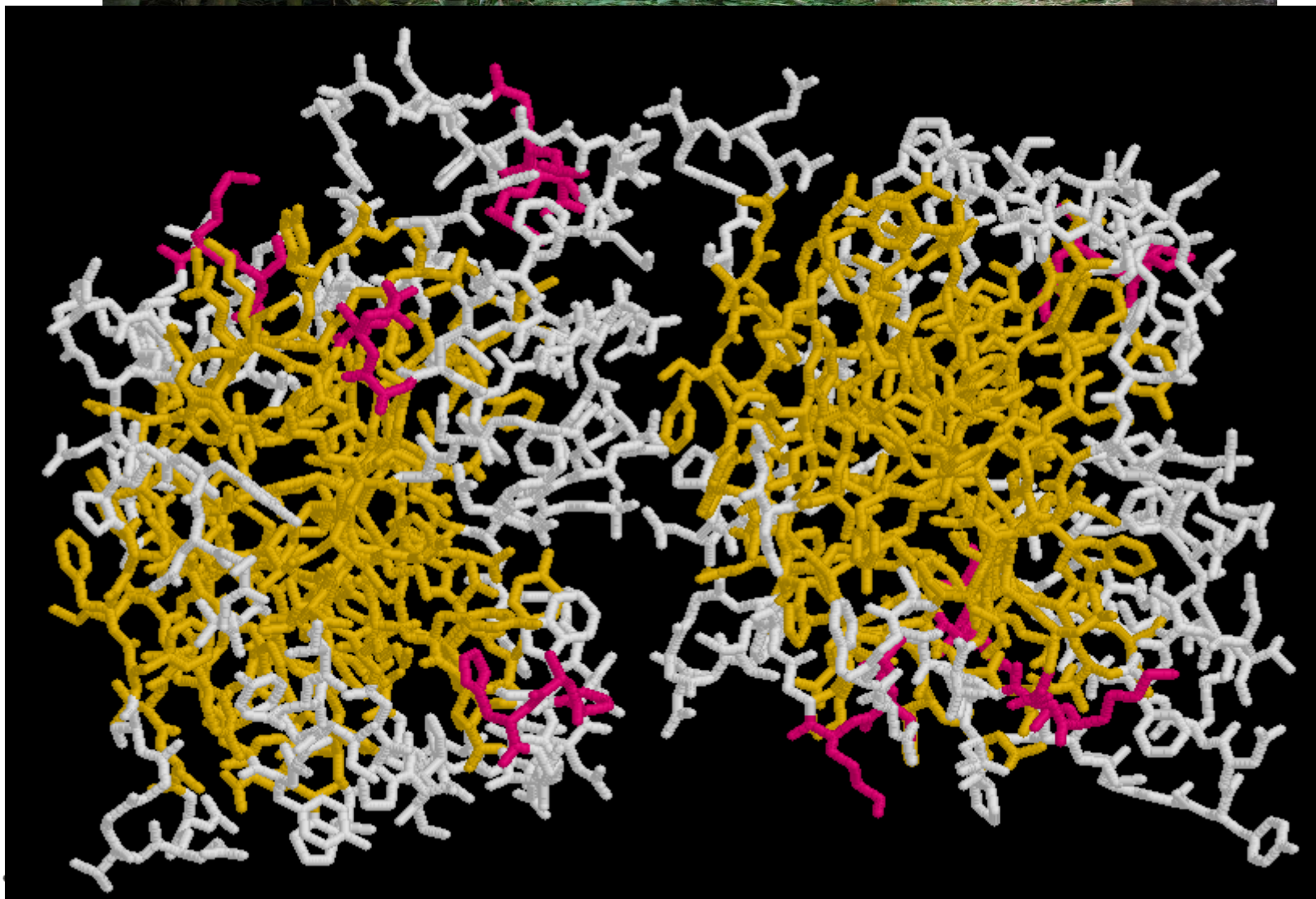
蛋白質的立體結構三



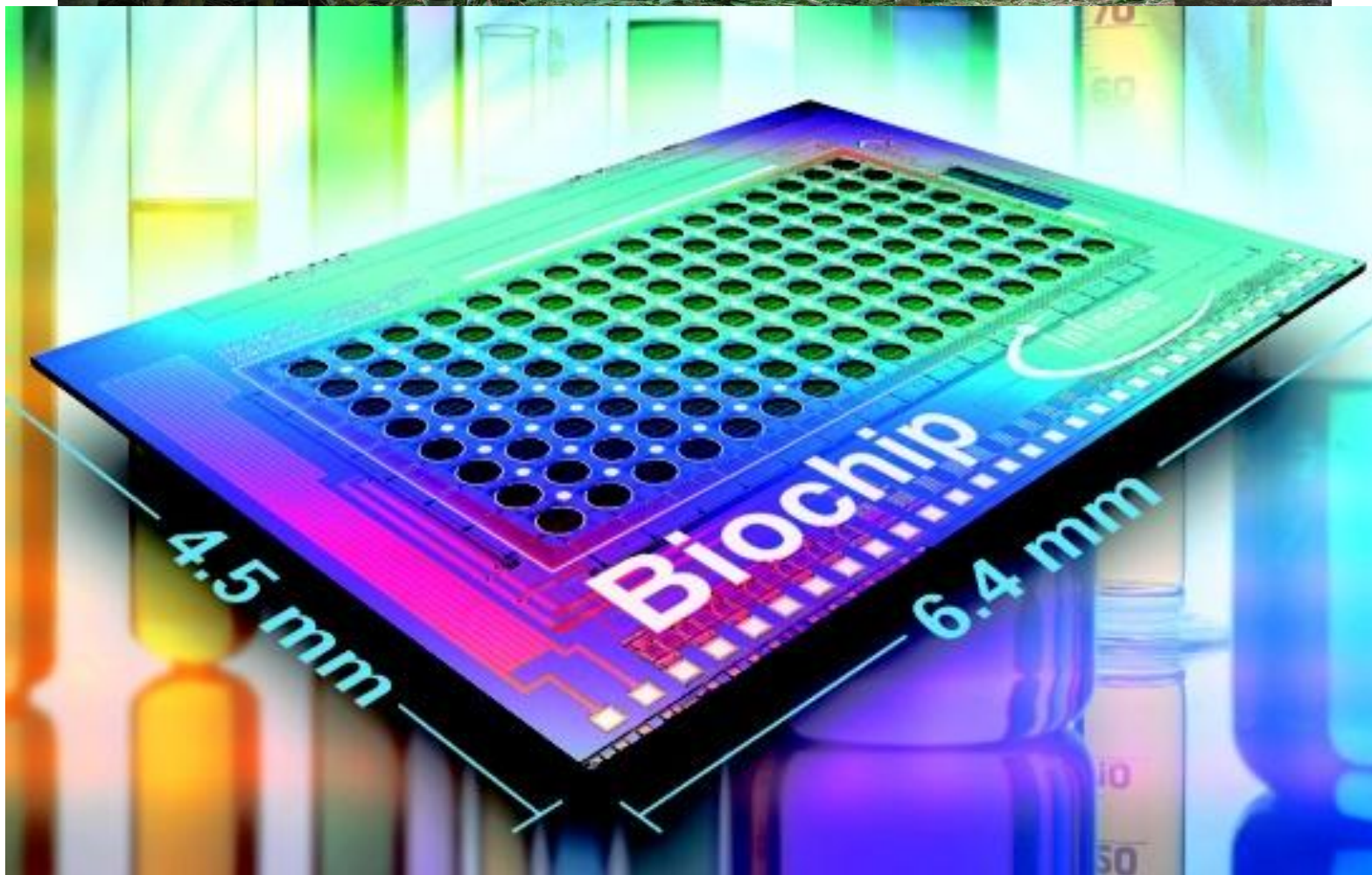
蛋白質的立體結構四



蛋白質的立體結構五

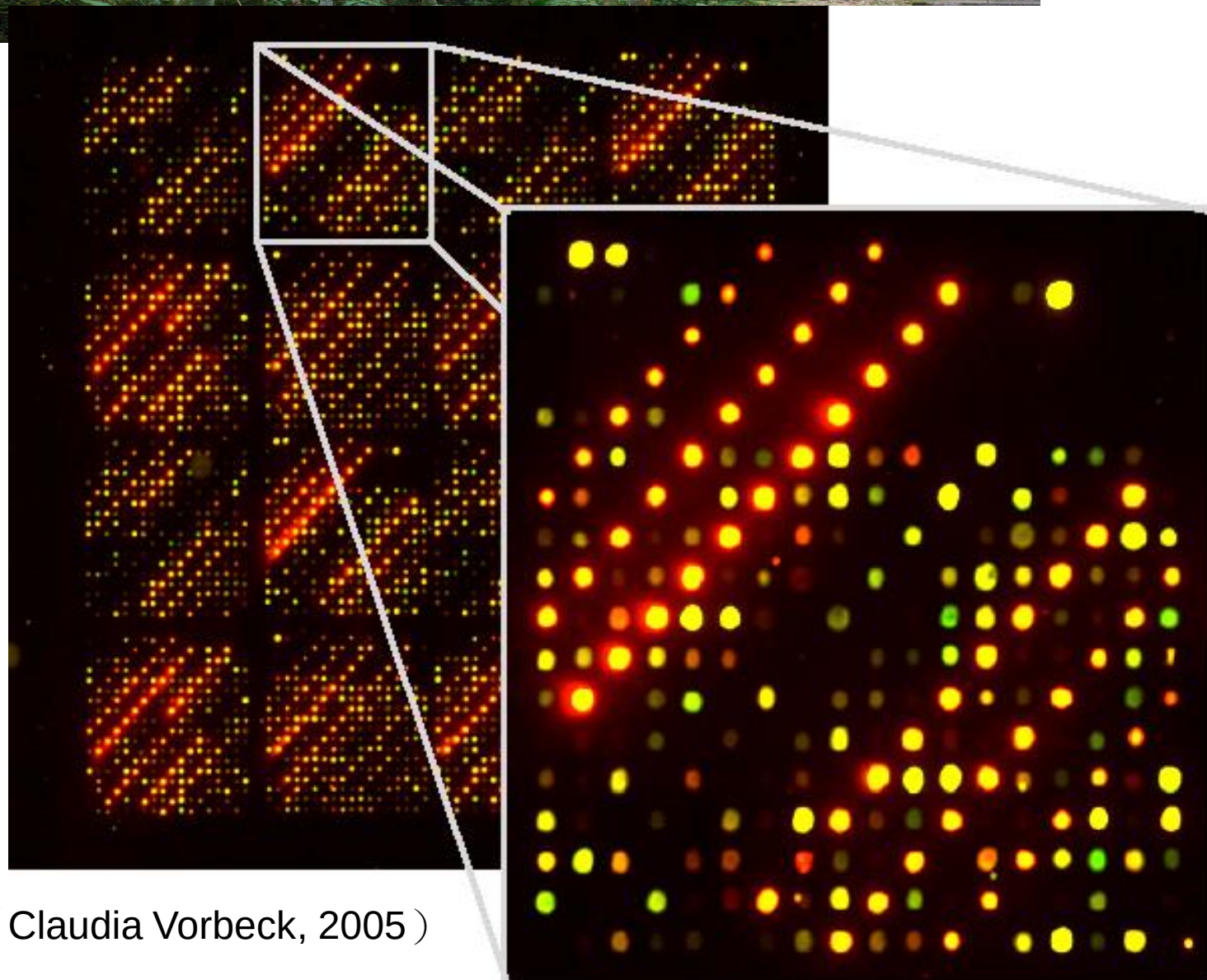


生物晶片圖例



參考資料：<http://www.fraunhofer.de>

生物晶片圖例



(Claudia Vorbeck, 2005)