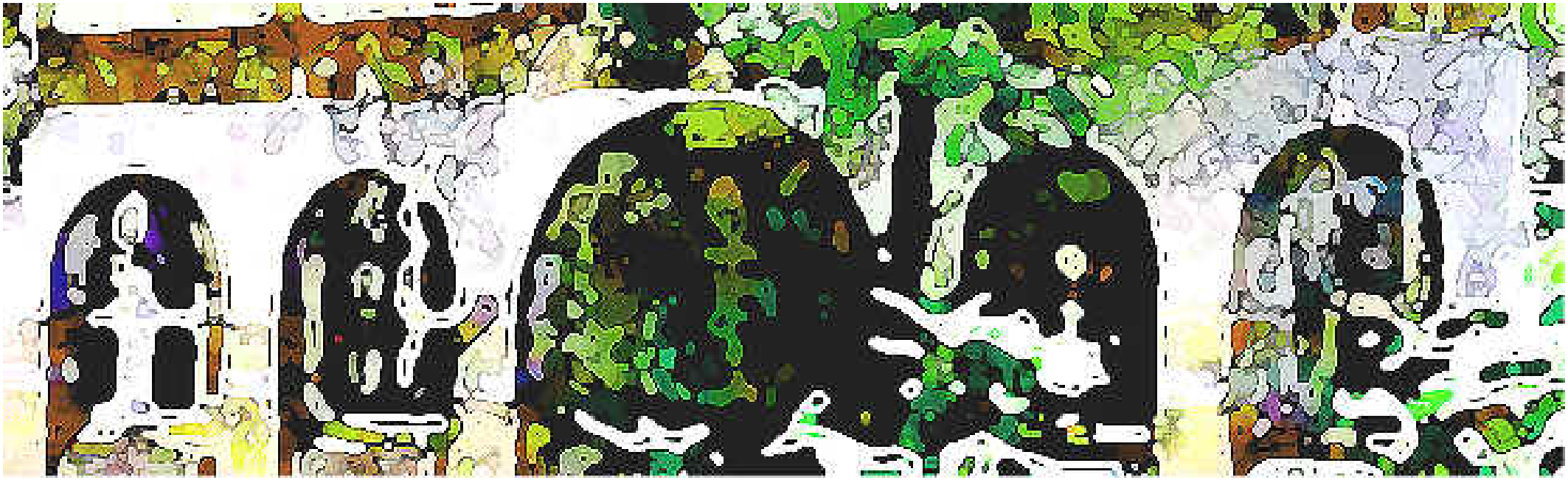
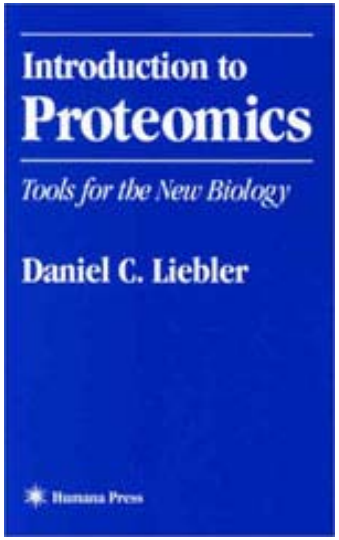
Introduction to Proteomics

新生物學之強大工具



國立台灣大學 生化科技學系

莊 榮 輝



Introduction to Proteomics: Tools for the New Biology

by Daniel C. Liebler (Editor) The University of Arizona, Tucson, AZ
2002 [ISBN: 0-89603-992-7]

I. Proteomics and the Proteome

1. Proteomics and the New Biology
2. The Proteome

II. Tools of Proteomics

3. Overview of Analytical Proteomics
4. Analytical Protein and Peptide Separations
5. Protein Digestion Techniques
6. Mass Spectrometers for Protein and Peptide Analysis
7. Protein Identification by Peptide Mass Fingerprinting
8. Peptide Sequence Analysis by Tandem Mass Spectrometry
9. Protein Identification with Tandem Mass Spectrometry
10. SALSA: An Algorithm for Mining Specific Features of Tandem MS Data

III. Applications of Proteomics

11. Mining Proteomes
12. Protein Expression Profiling
13. Identifying Protein-Protein Interactions and Protein Complexes
14. Mapping Protein Modifications
15. New Directions in Proteomics

http://www.amazon.com/gp/product/0896039919/sr=8-1/qid=1151672583/ref=pd_bbs_1/103-6419439-3067862?ie=UTF8

Proteomics and the Proteome

Central Dogma in macroscopic

Genome

↓

Transcriptome

↓

Proteome

↓

Metabolome

DNA

轉 錄
Transcription

↓

RNA

轉 譯
Translation

↓

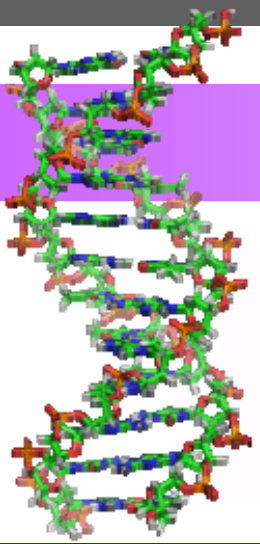
Protein

↓

Function

Interactome

Historical review for proteomics



Protein chemistry & enzymology

Gene bank
EST bank
Protein database

Human genome project

Microarray

Klause Biemann
Mass spectrometer
Amino acid analysis

Marc Wilkins:
"*Proteomics*"

1958 → **1981** → **1987** → **1990** → **2000**

Ionization technology
Fast atom bombardment (FAB)
Peptide sequence analysis

MALDI ⇢ MALDI-TOF
ESI ⇢ ESI-MS/MS
Biomolecule analysis
Protein 2-DE analysis

Francis William Aston
Invented mass spectrograph
Cavendish Laboratory (1919)
Cambridge University, England

Computer science
(Swiss-Prot, PC/GENE, GCG, NCBI, *etc*.)

Adapted from Liebler (2002) Introduction to Proteomics, Foreword

Is proteomics just what we used to call *protein chemistry* [5]

Protein chemistry	*Proteomics*
Individual proteins	Complex mixtures
Complete sequence analysis	Partial sequence analysis
Emphasis on structure and function	On identification from database
Structural biology	Systems biology
Fixed protein target	Dynamic proteomic targets
Traditional technology	High-through put technology

陸軍步兵之巷戰佔領　　空軍之地毯式轟炸

Genomics and proteomics have challenged biologists to think **Big**

Adapted from Liebler (2002) Introduction to Proteomics, p.7

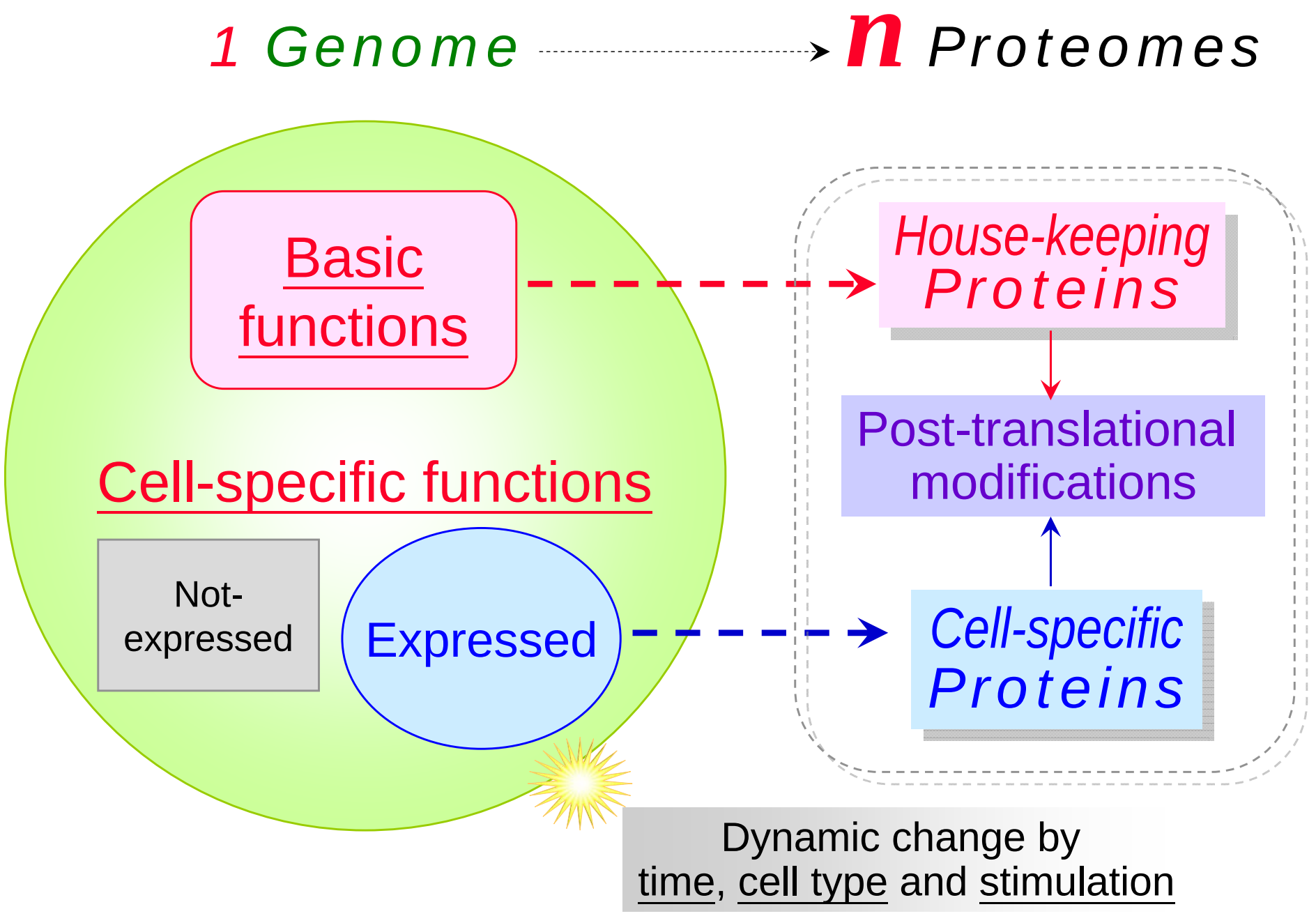
How to look at a proteome?

如何看待蛋白質體？

(1) One genome expresses many proteomes
(2) All proteins have life and death cycle
(3) Proteins are composed of modular structures
(4) Proteins belong to functional families
(5) Deducing the proteome from the genome
(6) Gene expression, codon bias and protein level

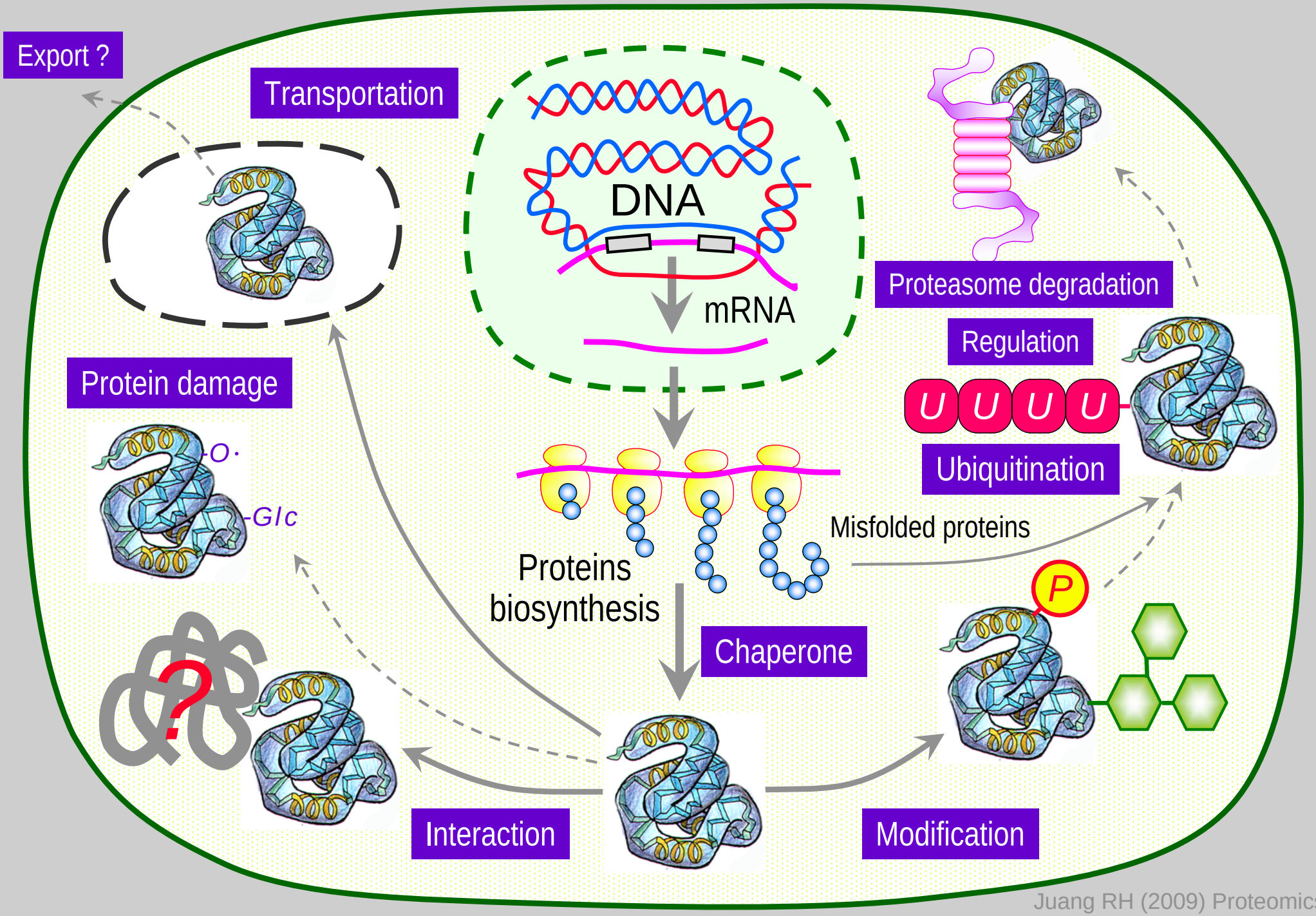
Adapted from Liebler (2002) Introduction to Proteomics, Ch.2

(1) The proteome and the genome



Juang RH (2009) Proteomics

(2) The life and death of a protein

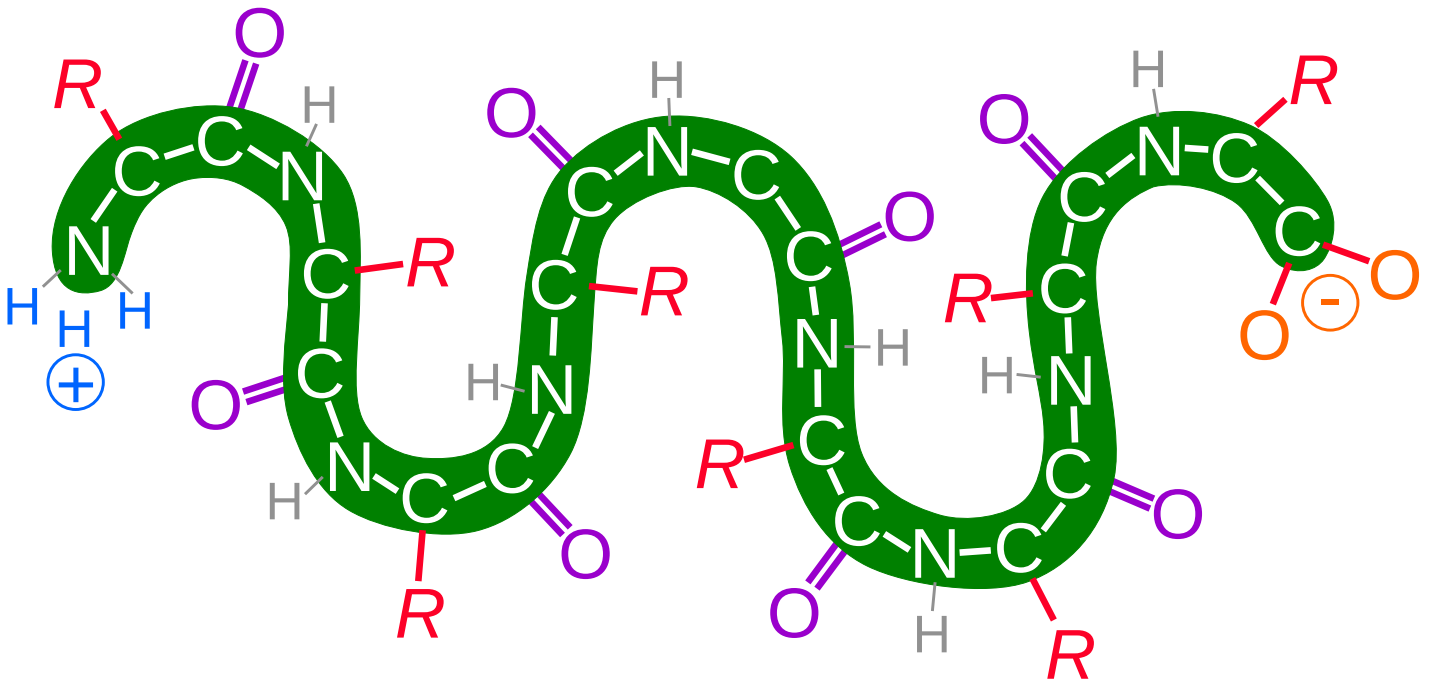


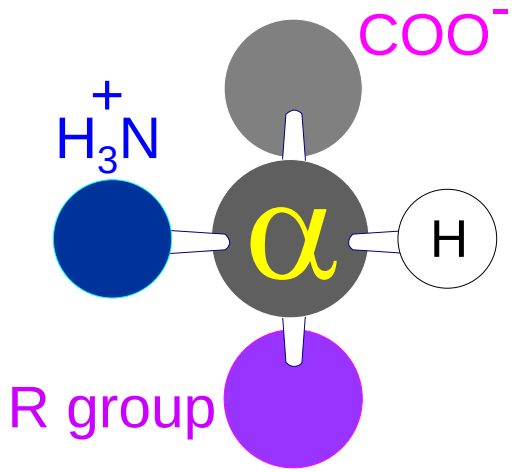
Juang RH (2009) Proteomics

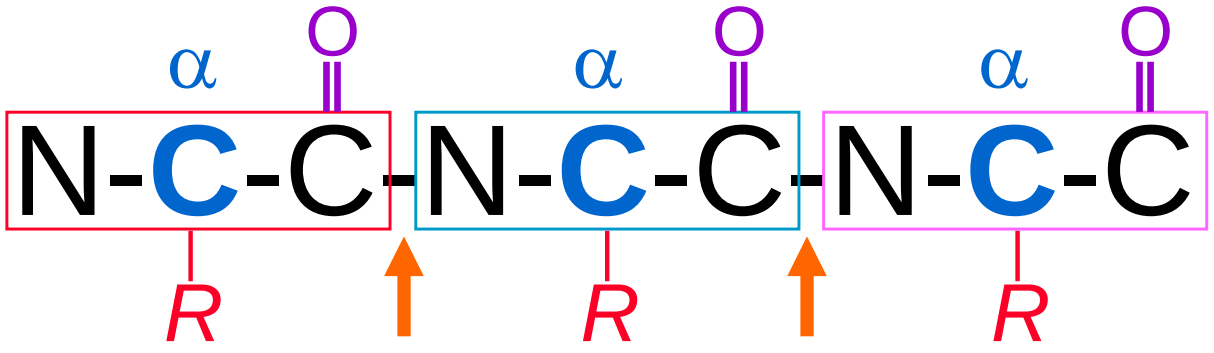
The backbone of protein molecule

Constant

Variable



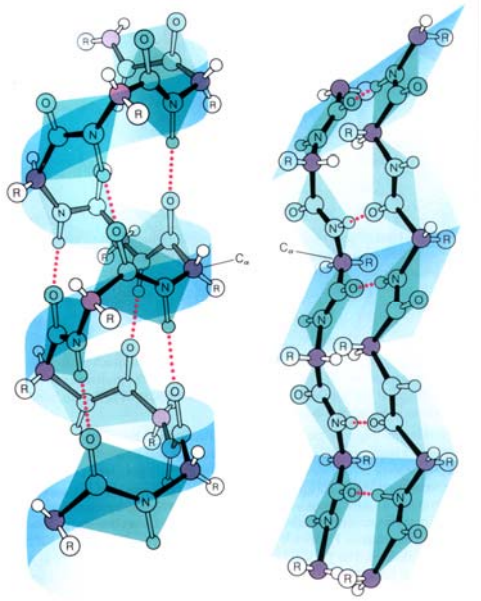


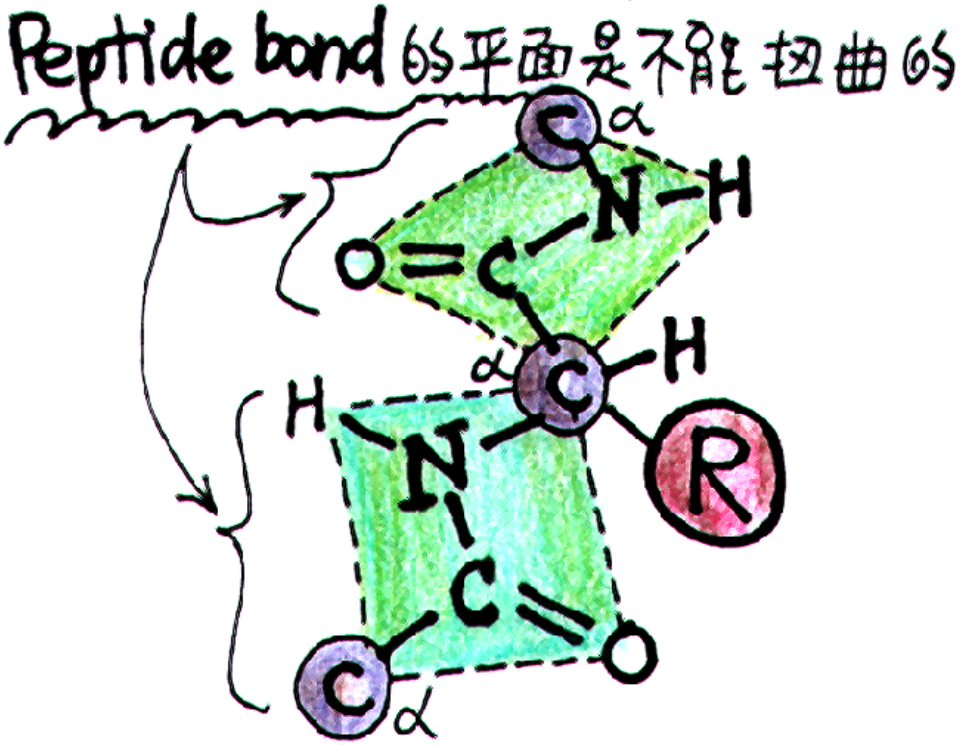


Juang RH (2009) Proteomics

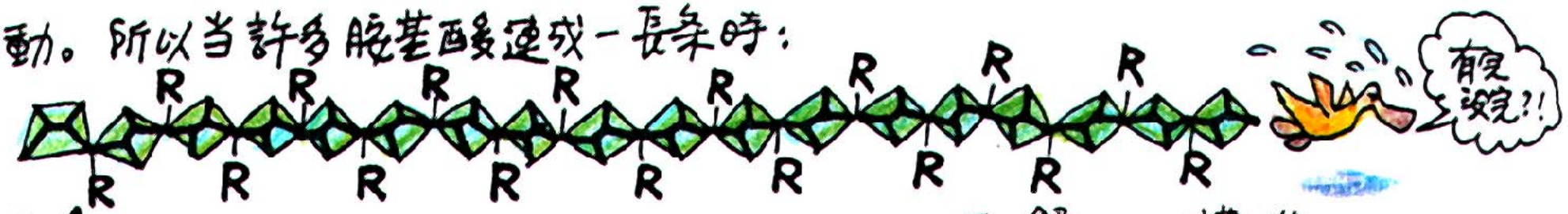
How protein folded into secondary structure

Mathews et al (2000) Biochemistry (3e) p.164





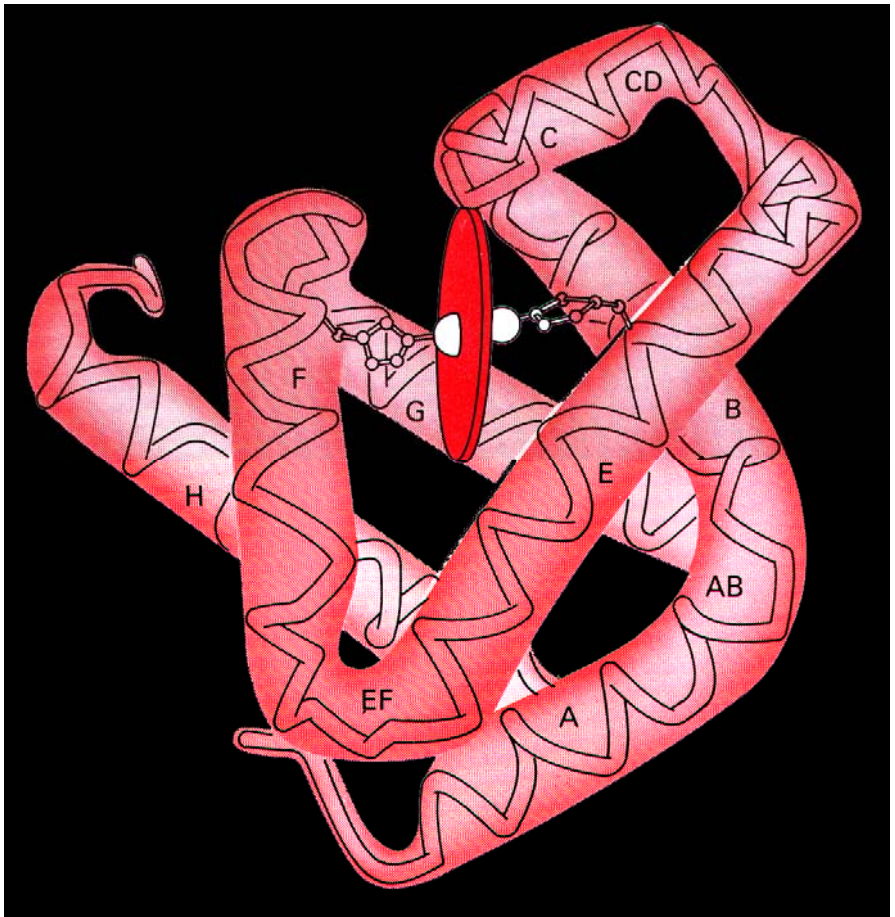
这兩個平面又因為 R group 的
関係，只能在一定範圍的角度内
活動。所以当許多胺基酸連成一長条時：

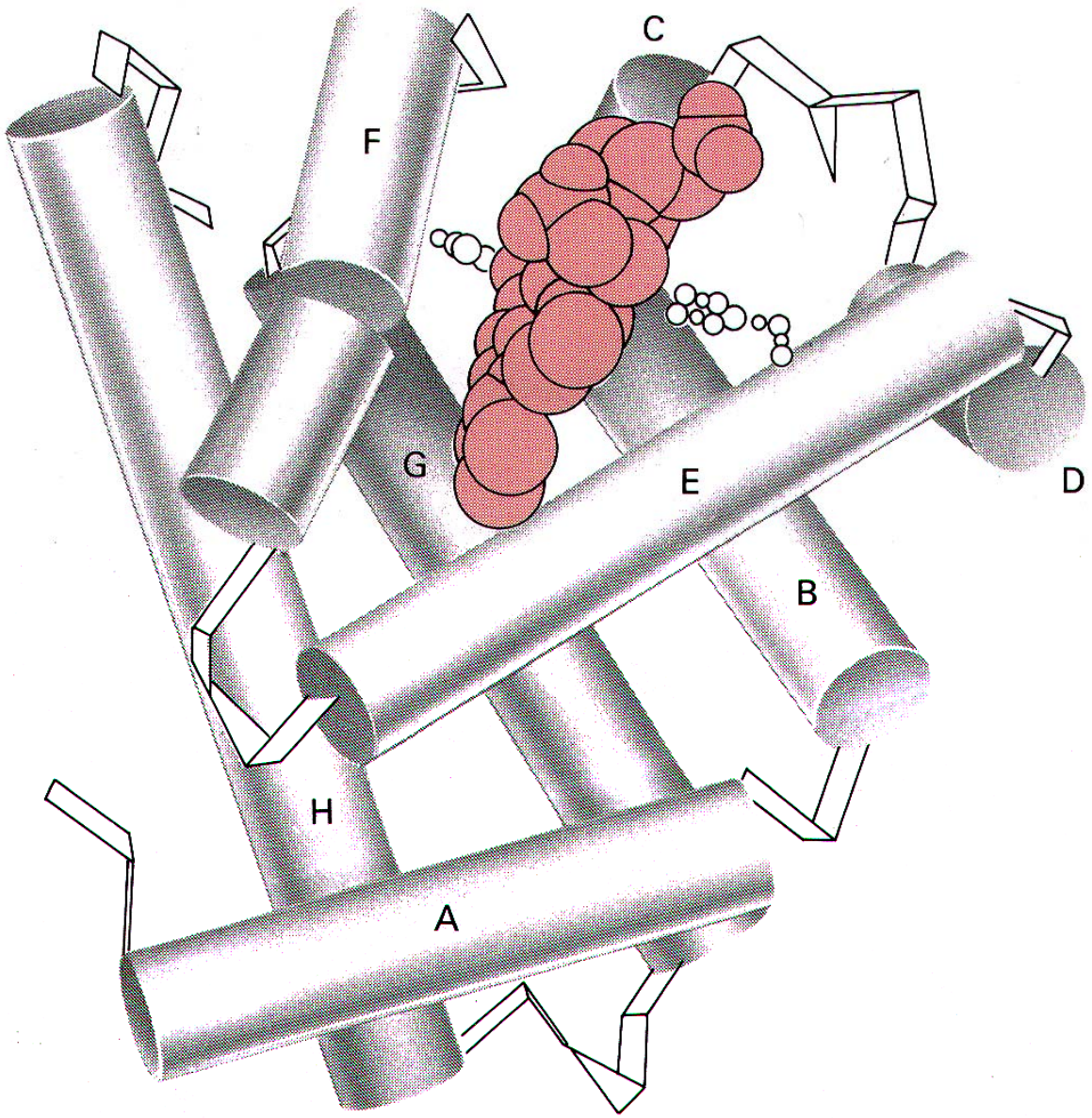


因為↑的関係，会自動捲曲成一定的構造（Secondary 二級 Structure 構造），
大略說來有：兩种形式 Helix  或 Sheet 

Myoglobin is built by eight α helices

Kendrew & Perutz (1962)
Cambridge University

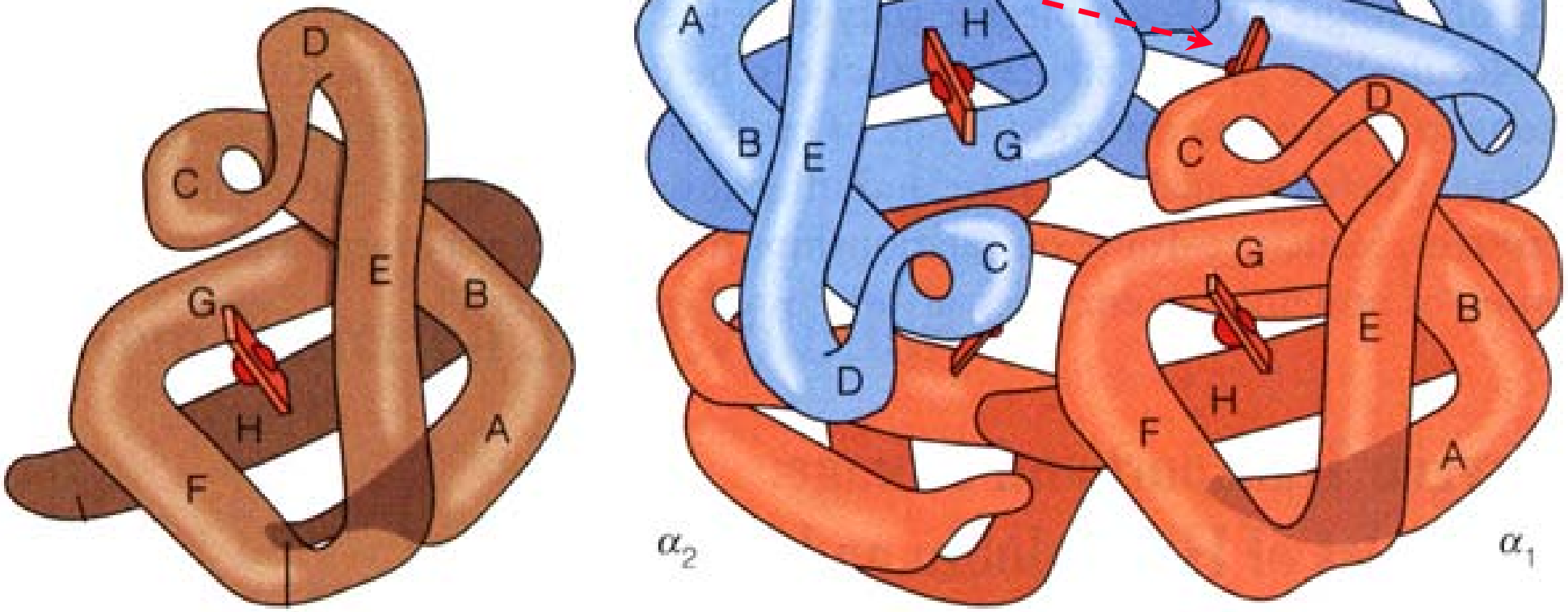




α helix is a solid structure

Zubay (1988) Biochemistry (2e) p.64

Hemoglobin has quaternary structure



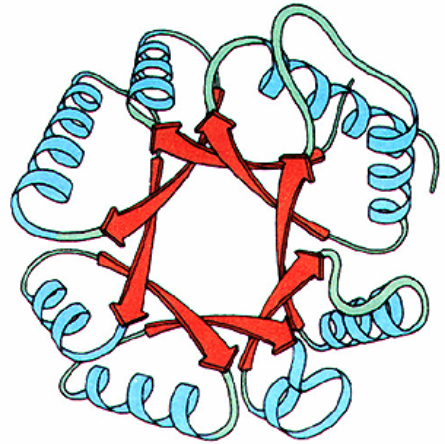
Myoglobin **Hemoglobin**

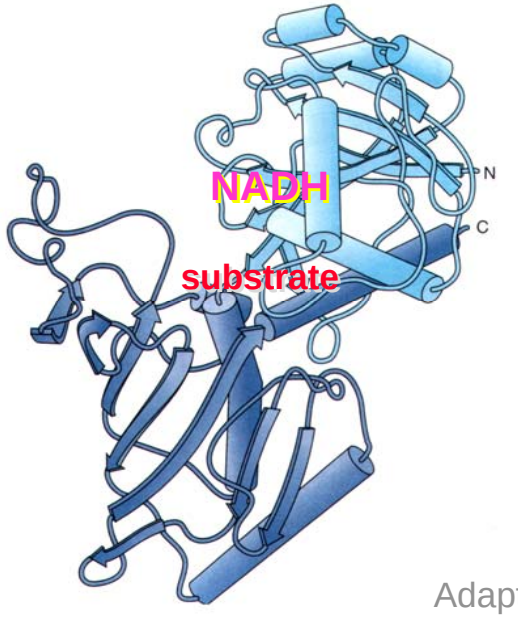
Mathews et al (2000) Biochemistry (3e) p.214

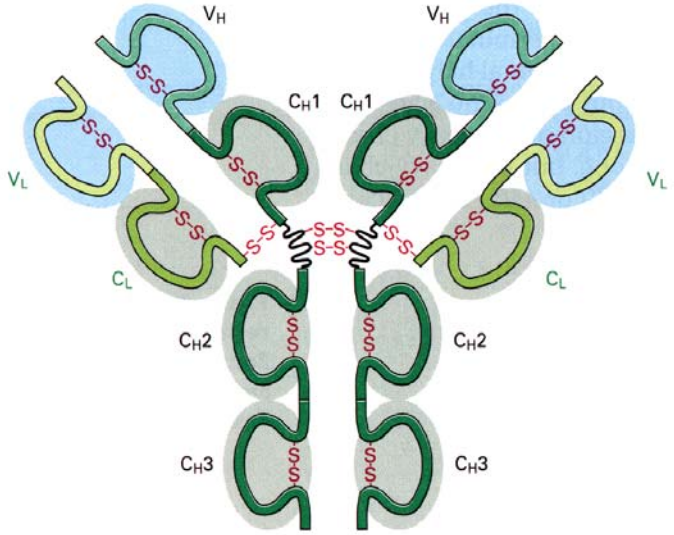
(3) Proteins as modular structures

Protein is composed of *modular* or mosaic structures in several levels
蛋白質是由各種層次的『模組』所組合成的 (像積木一樣的組合單位)

Module	*Example*	*Function*
Amino acid motif	-Asn-Xaa-Ser/Thr-	N-glycosylation
Secondary structure	α helix, β sheet	Building blocks
Supersecondary	$\alpha_8\beta_8$	Barrel structure
Domain	Dehydrongenase	NADH binding domain
Subunit	Ab	$V_H + V_L \rightarrow$ Ag binding site



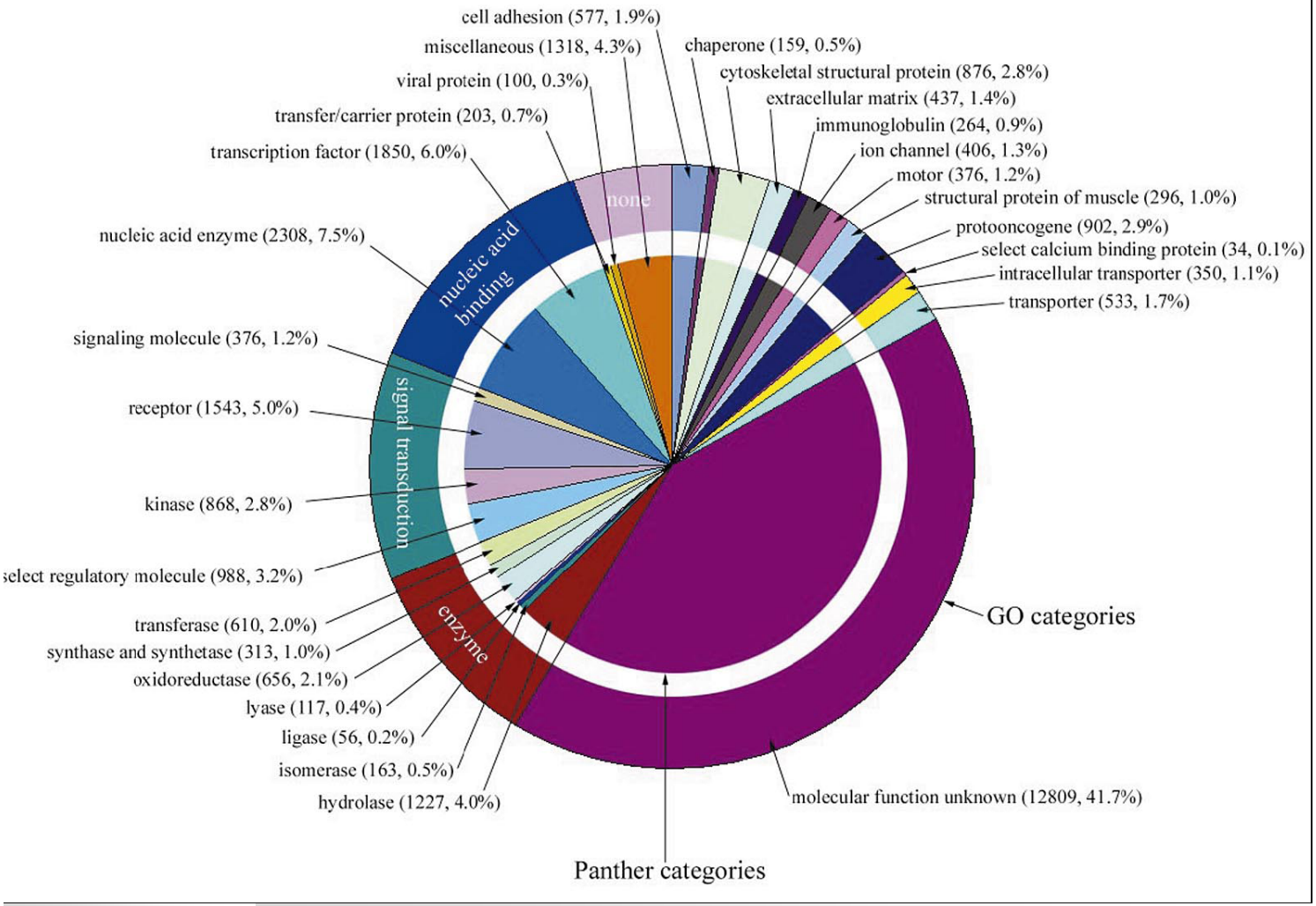




Module → *Function*

Adapted from Liebler (2002) Introduction to Proteomics, p.18

(4) Functional protein families (human)



信息傳導 (11%)

Signal transduction

- 激脢 kinase (2.8%)

核酸結合 (13.5%)

Nucleic acid binding

其他酵素 (10.3%)

Other enzymes

- 水解脢最多 (4%)

Hydrolyzing enzymes

病毒蛋白 (0.3%)

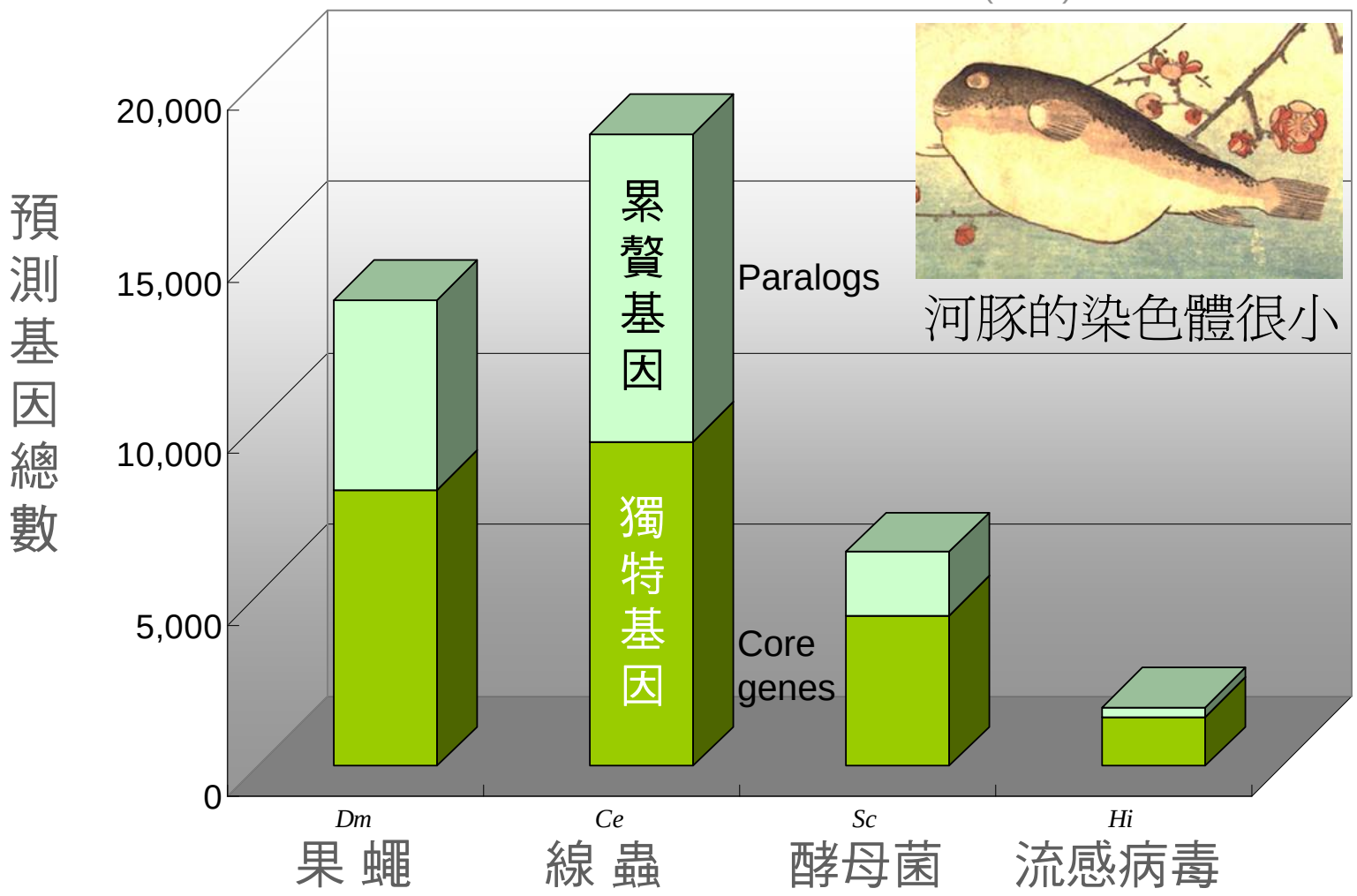
Viral proteins

15%	Intermediary and nucleic acid metabolism	**基本代謝**
15~20%	Structure and protein synthesis/turnover	**構造**
20~25%	Signaling proteins and DNA binding proteins	**核酸信息**
40%	Unknown	

Venter et al. (2001) Science 291: 1335

(5) Predicted protein products from genes

Rubin et al. (2000) Science 287: 2204



(a) 基因數目的多寡，與生物之複雜度不一定成正比關係。

(b) 線蟲有許多累贅基因 (paralogs)，可能是因於基因重複之故。

(c) 人類基因數目並不很多，因此人類的複雜度可能源自蛋白質之修飾調控。

Adapted from Liebler (2002) Introduction to Proteomics, p.21

(6) Gene expression not correlates with protein

The abundance of cellular protein is affected by:

(a) DNA transcription rate (codon bias)

(b) mRNA translation rate

(c) Protein degradation rate

Codon bias

Ser = UCA, UCC, UCG, UCU, AGG, AGU

有些胺基酸擁有一個以上的基因密碼 (如 Ser 就有六個密碼)。

有些生物會偏好其中的某些密碼 (稱爲 high codon bias value)。

因此那些含 low codon bias value 的基因就比較不容易被表現。

Activity regulation

具有活性的蛋白質 (例如酵素) 其最終表現就更複雜。

表現出來的蛋白質不一定有活性，還要受到很多調控。

Adapted from Liebler (2002) Introduction to Proteomics, p.22

Genome

基因表現不一定完全反映在蛋白質

由基因體較難預測蛋白質的修飾及調控

也無法預測蛋白質間的交互作用

Proteome

Juang RH (2009) Proteomics

Protein repertoire not equivalent to gene expression

One Gene, One Protein?

細胞中

(1) 細胞內的每個基因不一定都在表現

(2) 各器官、組織的基因表現都不相同

(3) 基因表現隨著生長時期而有改變

(4) 蛋白質表現後有進一步的修飾與調控

(5) 蛋白質在細胞內的代謝速率不同

(6) 隨著外界刺激與養分有不同表現

試管中

(a) 蛋白質的水溶性會影響抽取效率

(b) 蛋白質的含量差異很大

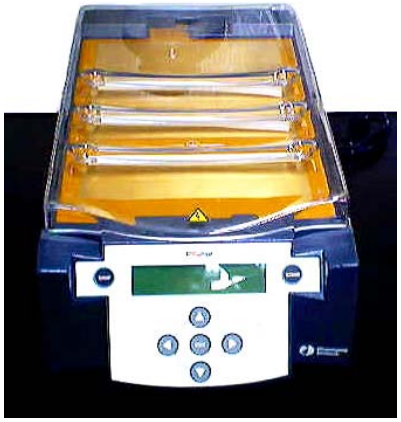
(c) 蛋白質在抽取後的安定性與半衰期不同

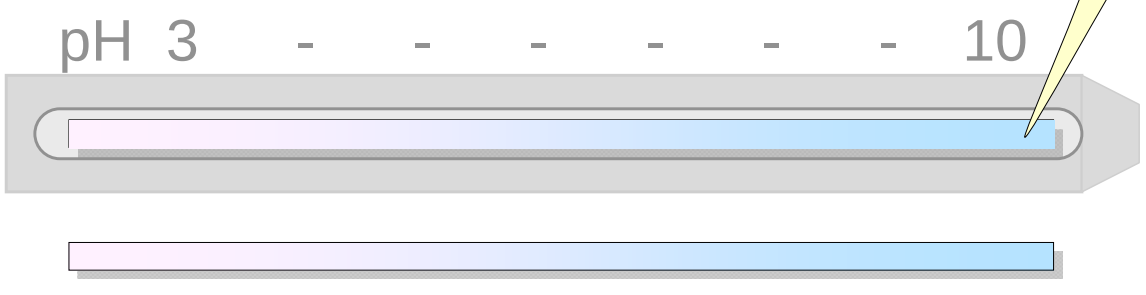
(d) 胞器破壞後許多物質混雜在一起

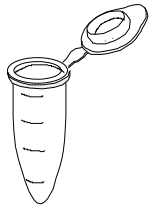
Juang RH (2009) Proteomics

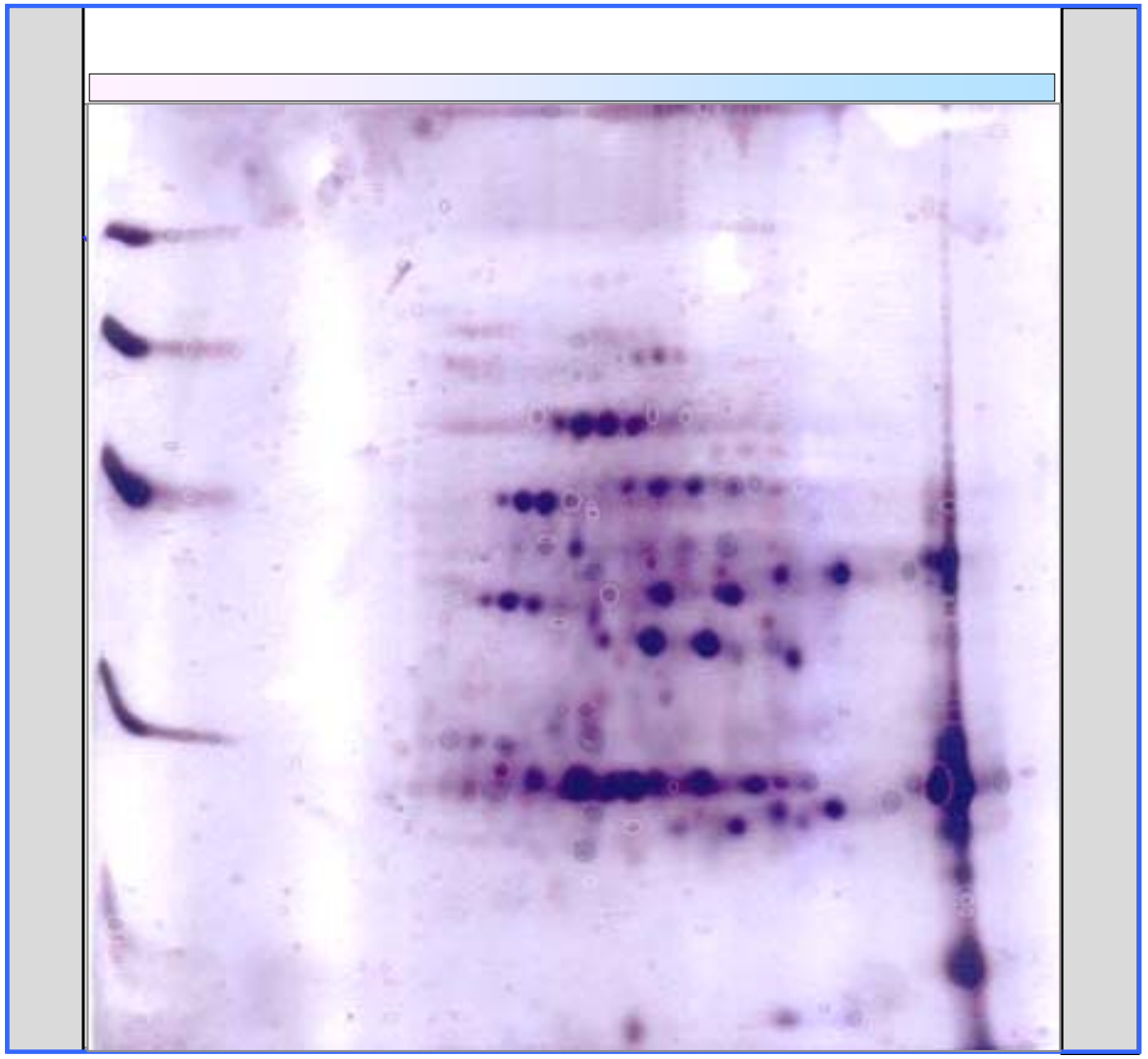
2-DE analysis is not perfect, but useful

(1) IEF
等電焦集電泳





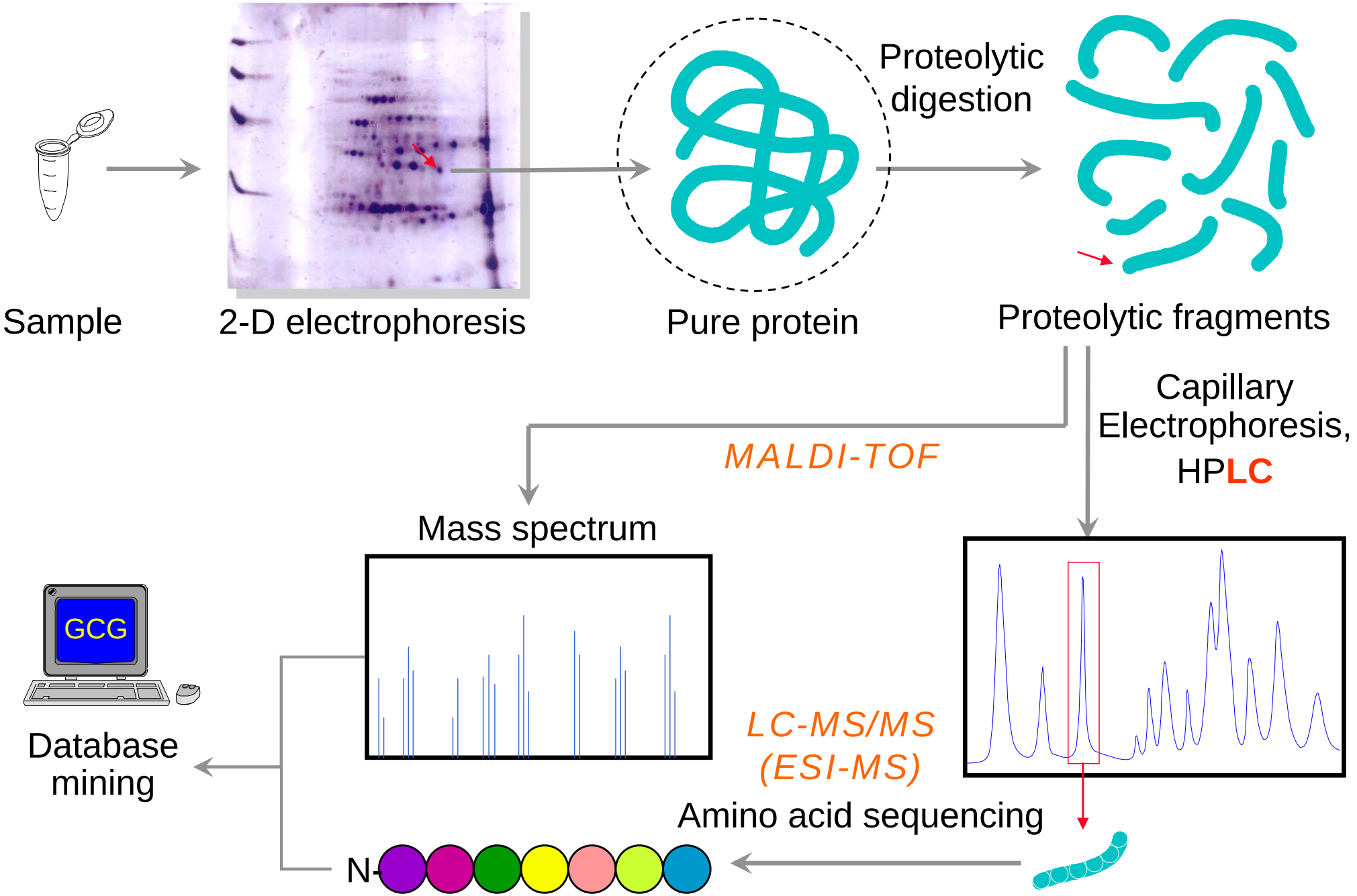




(2)
SDS-PAGE
分離膠體

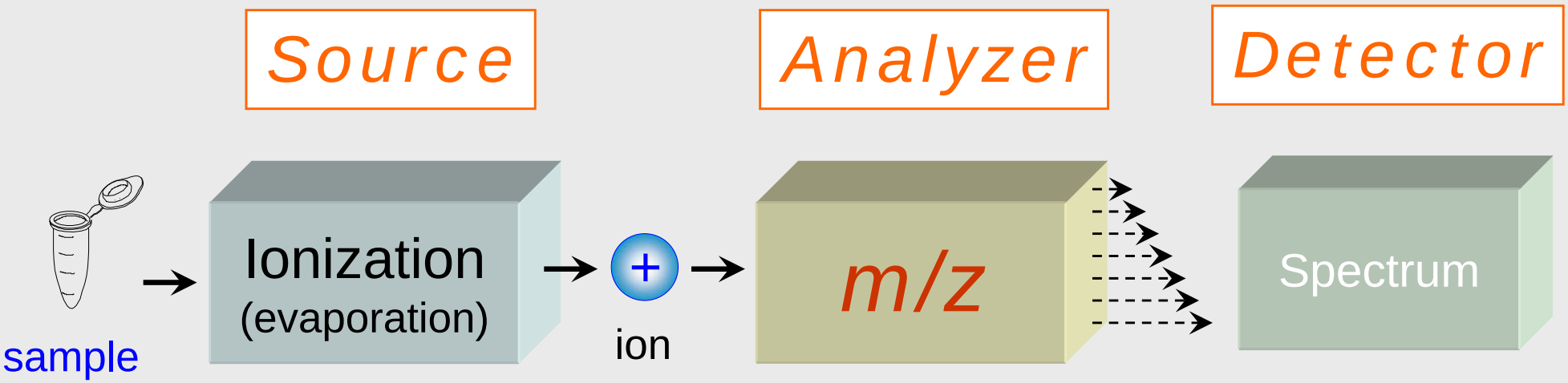
(3) Staining
染色脫色

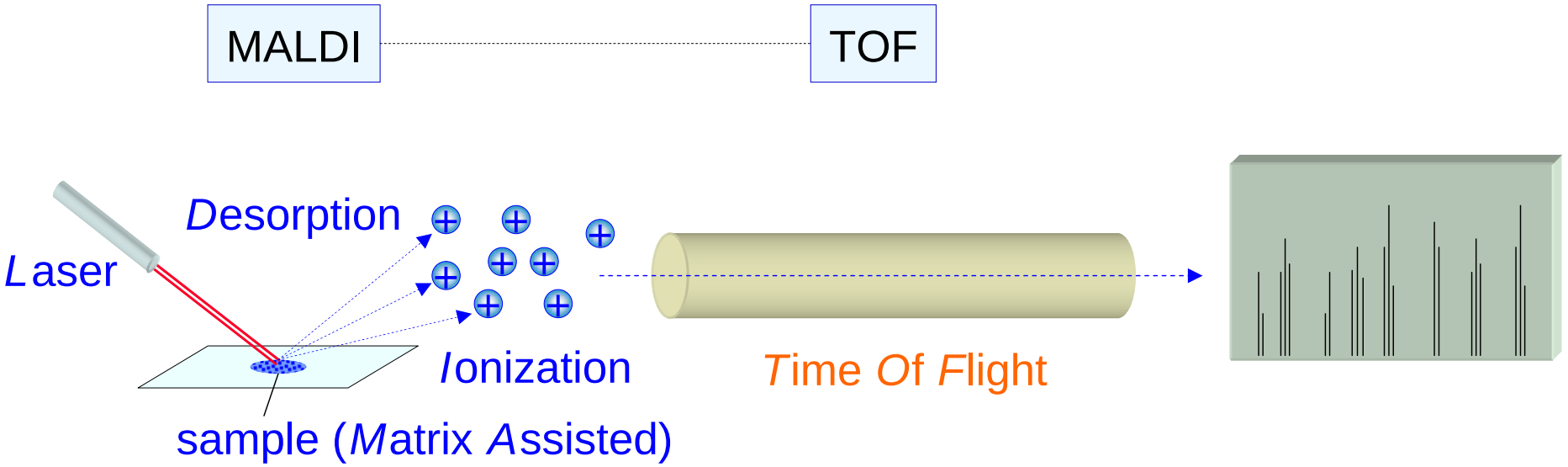
Quick identification of an unknown protein



Juang RH (2009) Proteomics

Mass spectrometer has three essential parts





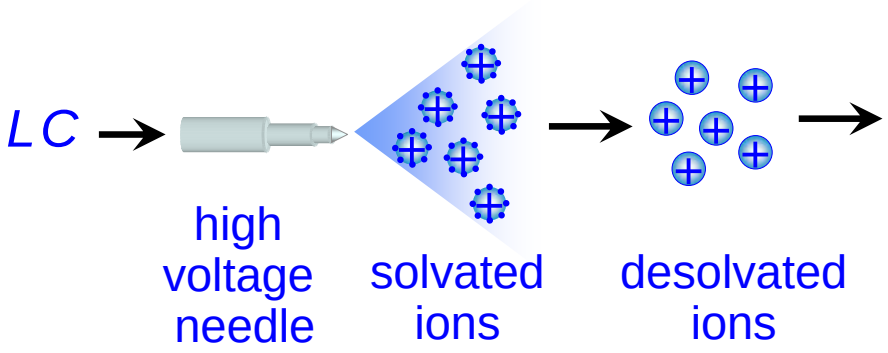
影片

ESI-MS/MS can analyze smaller fragments

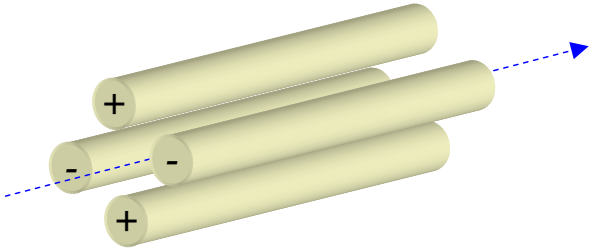
Source | Analyzer | Detector

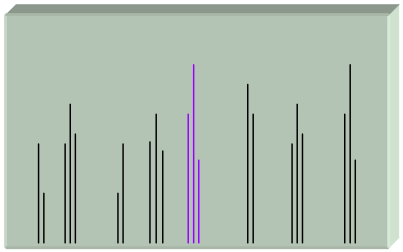
LC — ESI — MS/MS (or Ion-trap, *Q-TOF*)

*E*lectro*S*pray *I*onization



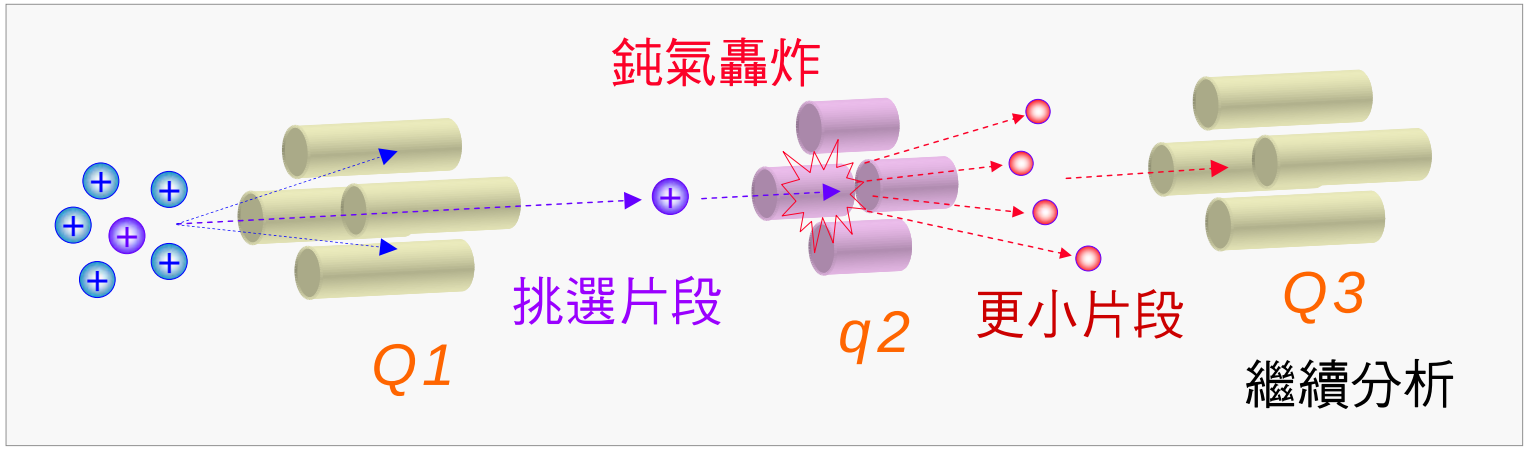
Quadrupole 四極棒





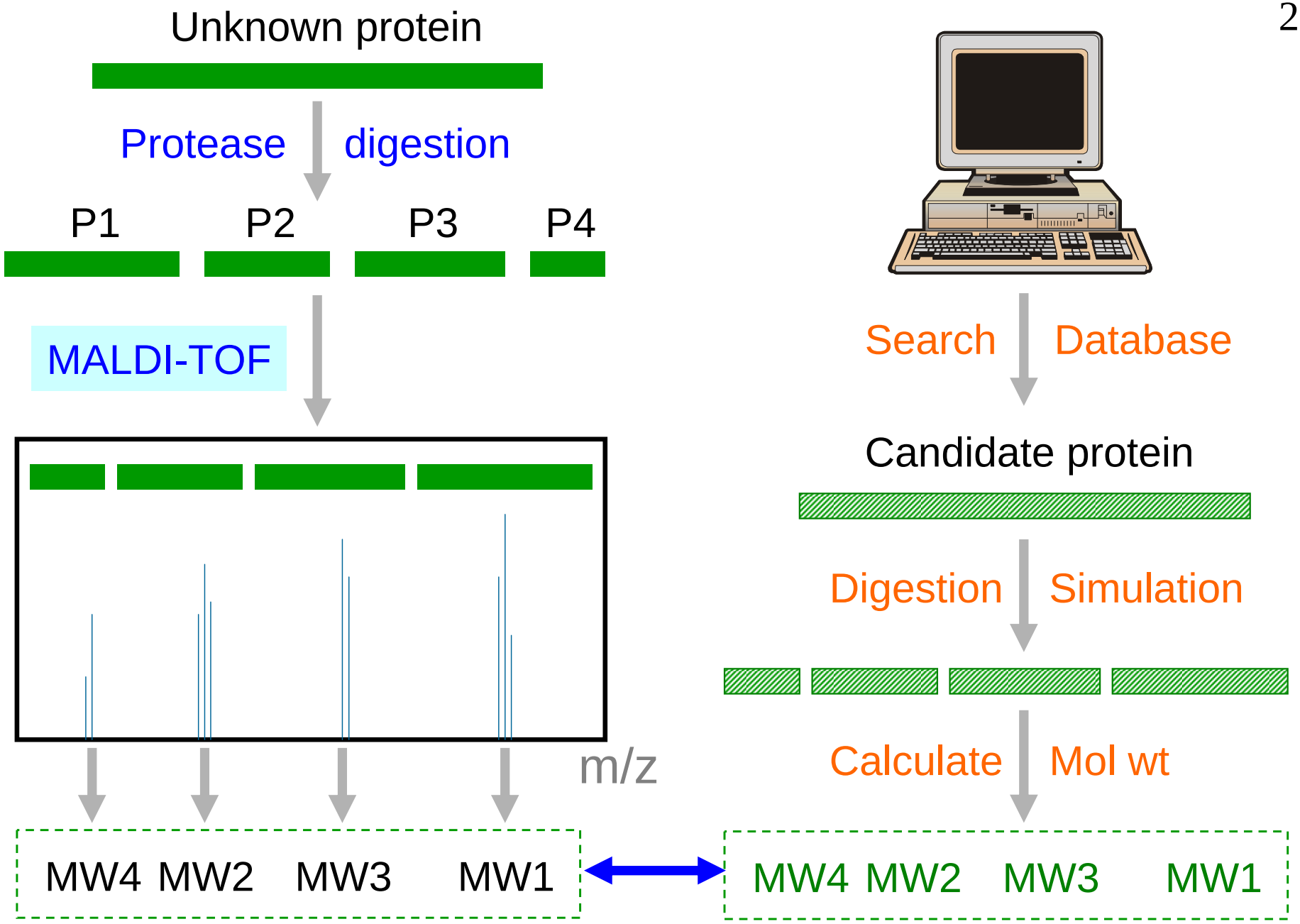
Triple Quadrupole

也可以加上 *TOF*



Adapted from Liebler (2002) Introduction to Proteomics, p.68 ,70

質譜儀可檢定蛋白質身分

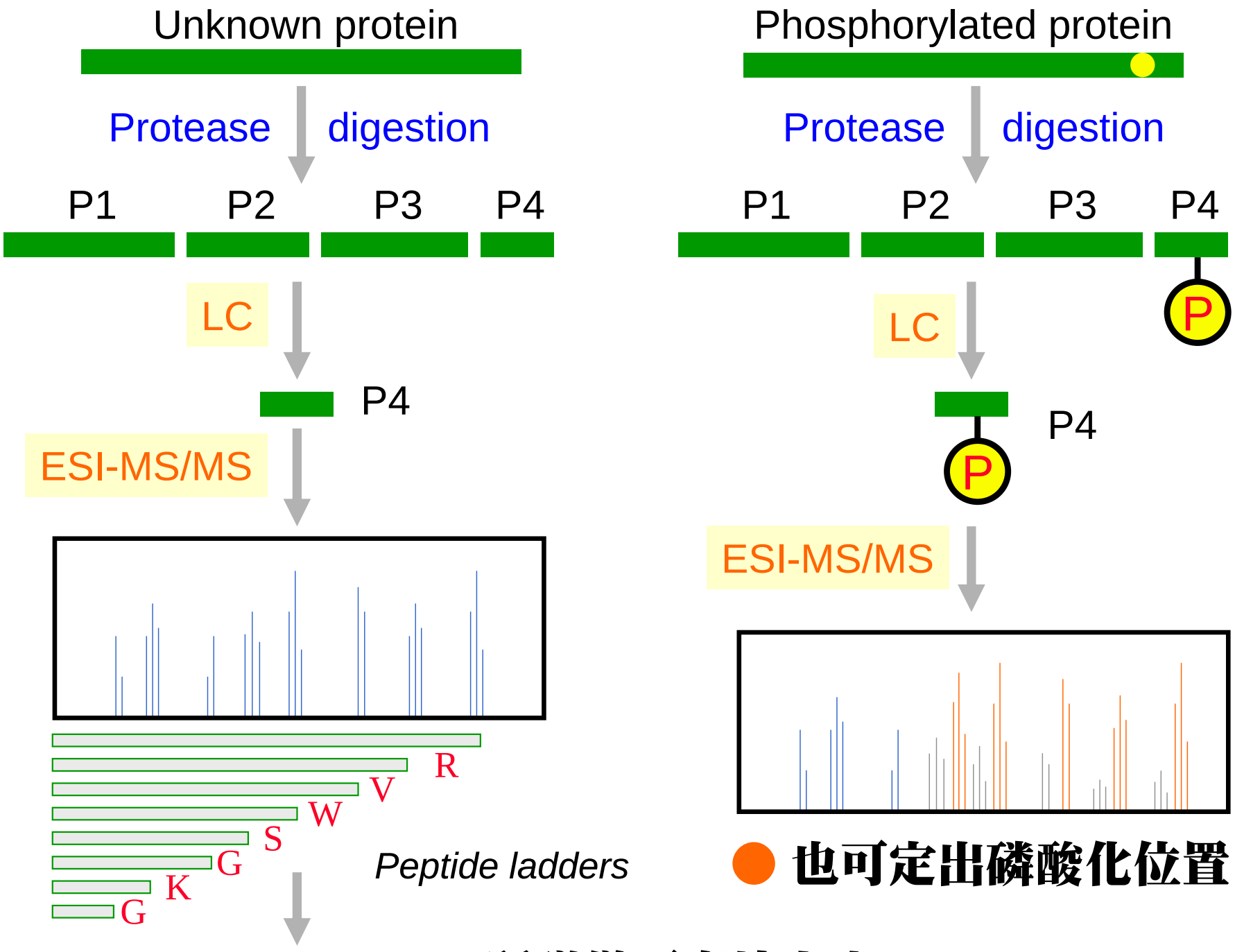


- **比對各片段分子量可確定該蛋白質身分**

MALDI-TOF is a quick tool for identifying proteins in a decoded genome

Juang RH (2009) Proteomics

質譜儀可進行胺基酸序列分析

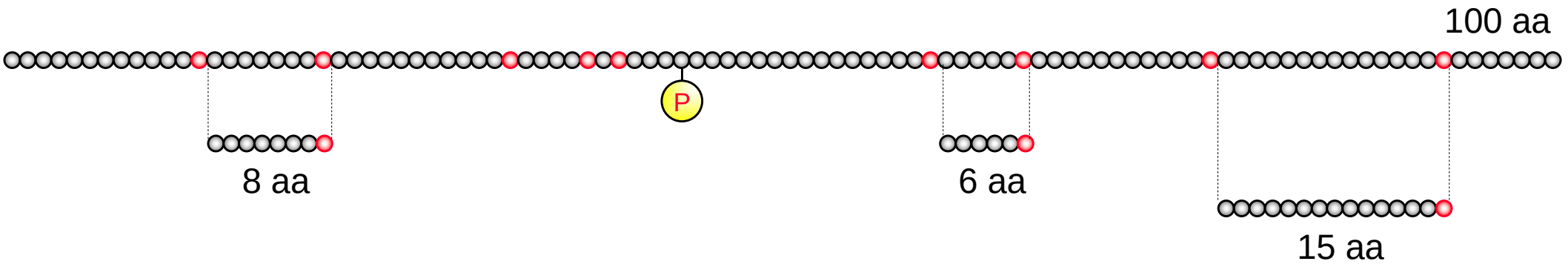


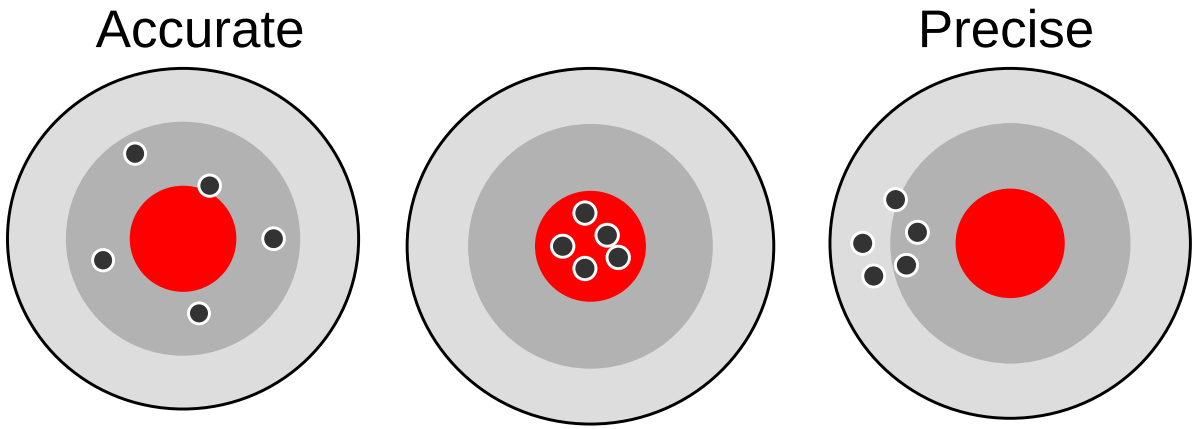
Amino acid sequence of a protein can be analyzed by ESI-MS/MS

Juang RH (2009) Proteomics

Access the mass data

- Sensitivity 靈敏度 ~ femtomole (10^{-15} mole)
- Precise 精確度 ~ 1 Dalton (Resolution)
- Accuracy 正確性 Correct identification?
- Coverage 涵蓋度 > 15%

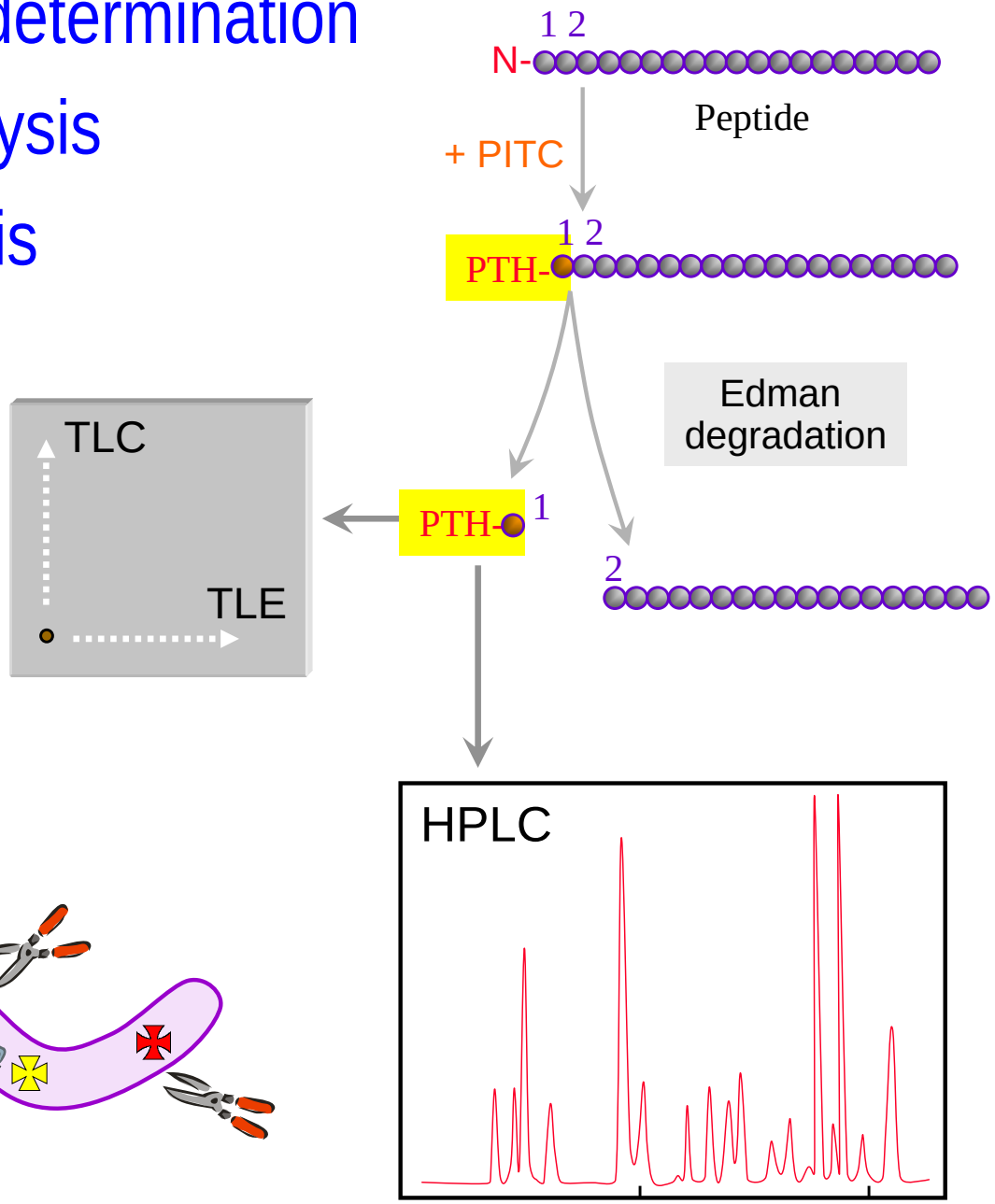


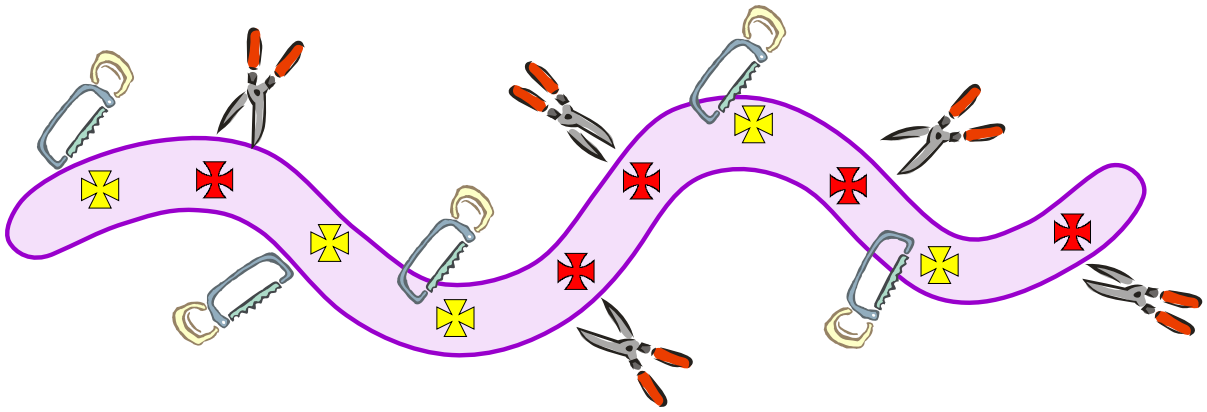


Juang RH (2009) Proteomics

Basic principles for protein analysis unchanged

- N- or C-terminal amino acid determination
- Amino acid composition analysis
- Amino acid sequence analysis
  - Edman degradation
  - cDNA open reading frame
  - Sequencing by LC-MS/MS
- Peptide mapping (Trypsin)
  - TLE/TLC, HPLC, SDS-PAGE, MS

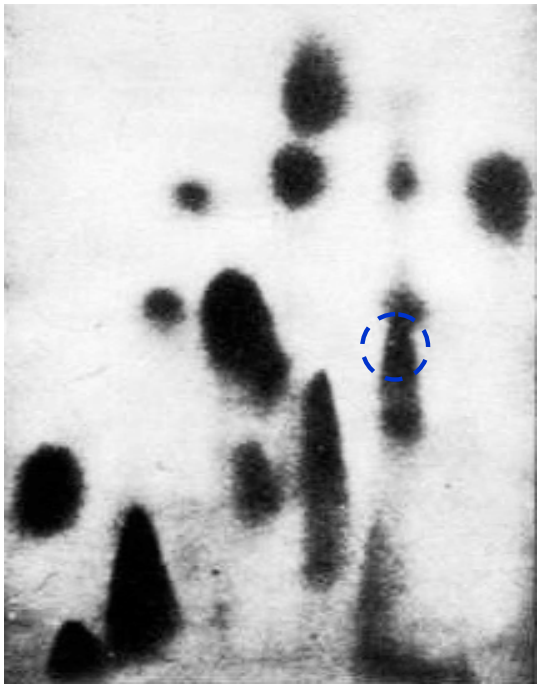


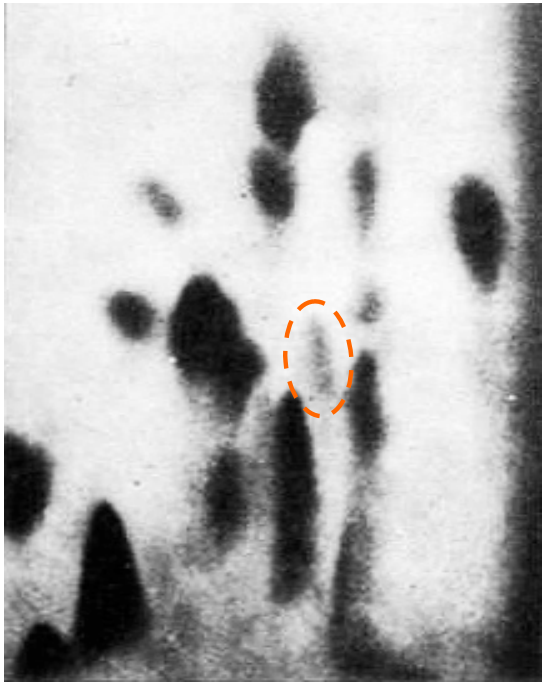


Juang RH (2009) Proteomics

2-DE has been applied for over half century

Hemoglobin A



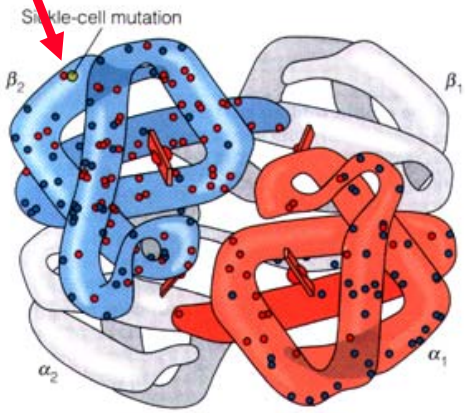


Hemoglobin S

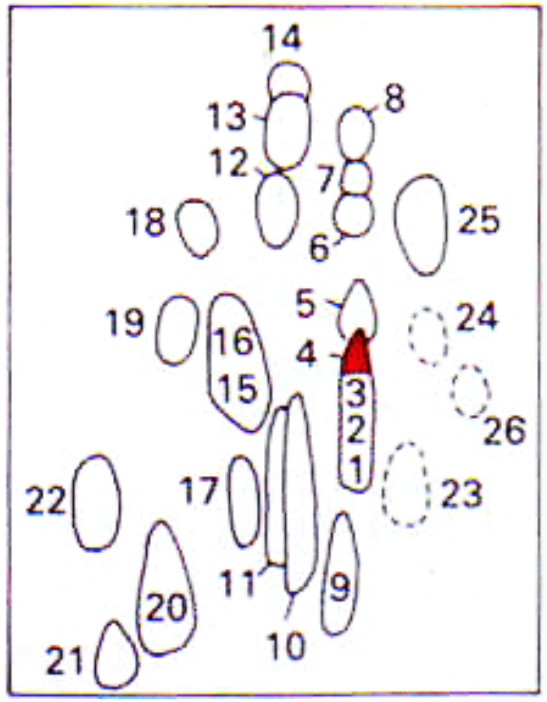
鐮型血球

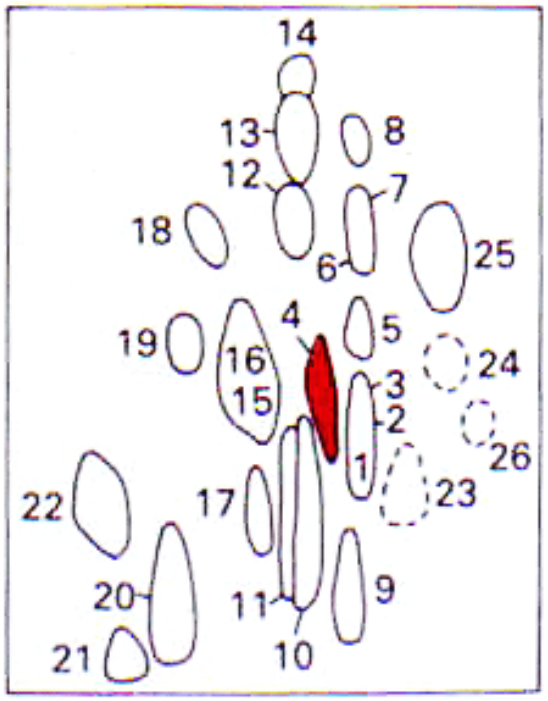
Linus Pauling

Glu → Val



血紅蛋白 4 號片段



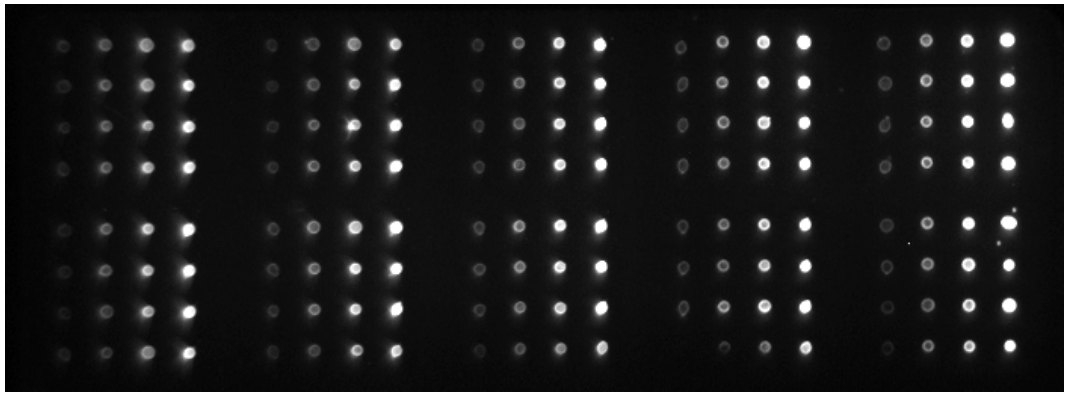


色析 TLC

電泳 TLE

The features of modern protein technology

- 高產能 High-through put
- 快 速 High-speed
- 微 量 Micro-scaled



Application of proteomics

(1) Mining Proteomes

蛋白質體的探勘與分離

(2) Protein Expression Profiling

探索蛋白質體表現之輪廓

(3) Identifying Protein-Protein Interactions

釐清蛋白質之間的交互作用

(4) Mapping Protein Modifications

鑑定蛋白質之修飾（磷酸化、醣化)

(5) New Directions in Proteomics

微小化、自動化、蛋白質晶片

Sampling proteome is critical

(a) Growth period

不同生長時期有不同蛋白質體表現

(b) Tissue or organ

各組織或器官的蛋白質體表現不同

(c) Cell compartmentalization

細胞內分隔或胞器各表現其特有蛋白質

(d) Solubility of the target proteome

蛋白質之極性或非極性

(e) Modification of the target proteome

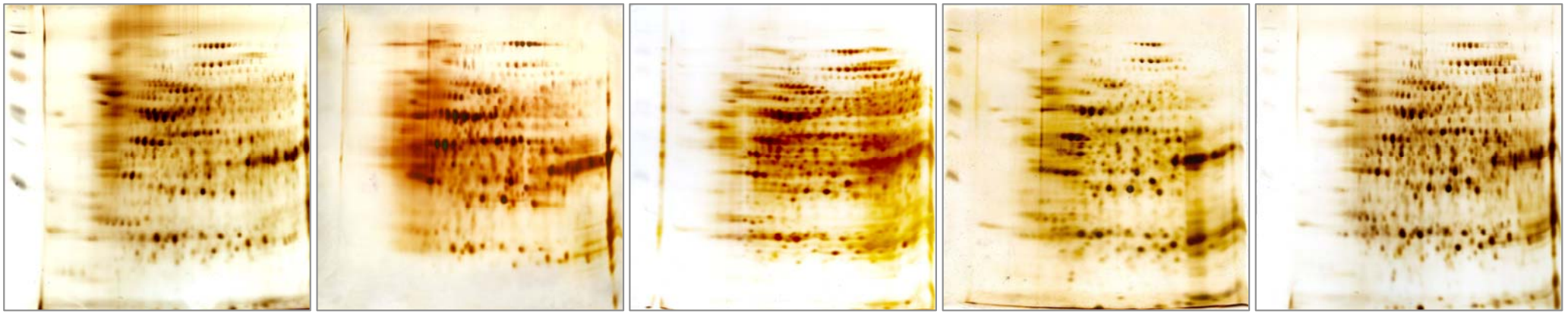
蛋白質可能醣化、脂化或磷酸化

(f) Stimulation or signaling

外加刺激或試劑對蛋白質表現的影響

Proteomic pattern changes during growth

0 cm | 10 cm | 20 cm | 40 cm | 60 cm



綠竹筍

Cellulose synthesis

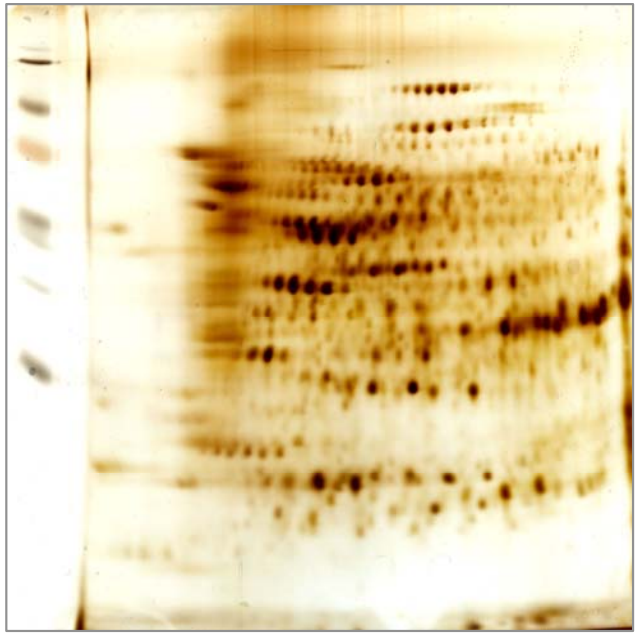
Juang RH (2007) Proteomics (with Wu YJ)

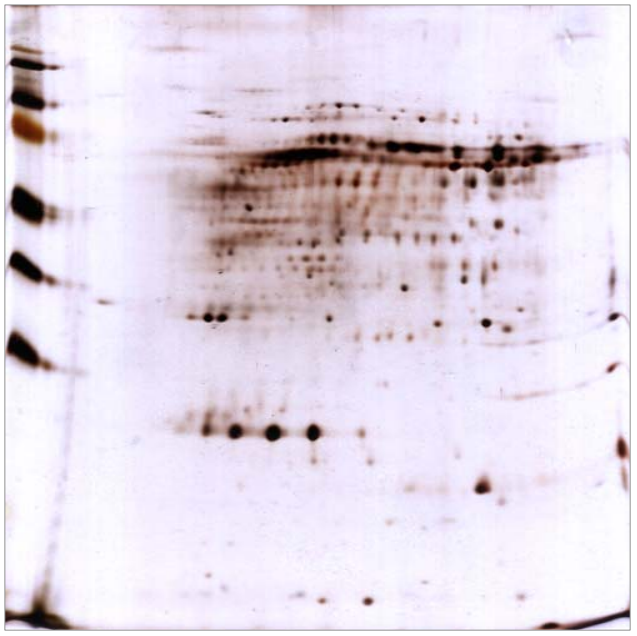
點	Protein ID	Accession no.	Calculated Mr (kD) / pI	Sequence coverage (%)	Score (MASCOT)	Match fragments
79	Sucrose synthase	AAV64256 (*Bambusa oldhamii*)	92.8 / 6.03	35	402	14
80	Sucrose synthase	AAV64256 (*Bambusa oldhamii*)	92.8 / 6.03	35	245	7
82	Sucrose synthase	AAV64256 (*Bambusa oldhamii*)	92.8 / 6.03	35	1112	45
8	UDP-glucose-pyrophosphorylase	BAB69069 (*Oryza sativa*)	51.6 / 5.4	18	302	26
9	UDP-glucose-pyrophosphorylase	BAB69069 (*Oryza sativa*)	51.6 / 5.4	17	359	20
10	UDP-glucose-pyrophosphorylase	BAB69069 (*Oryza sativa*)	51.6 / 5.4	21	408	38
11	UDP-glucose-pyrophosphorylase	BAB69069 (*Oryza sativa*)	51.6 / 5.4	20	377	35

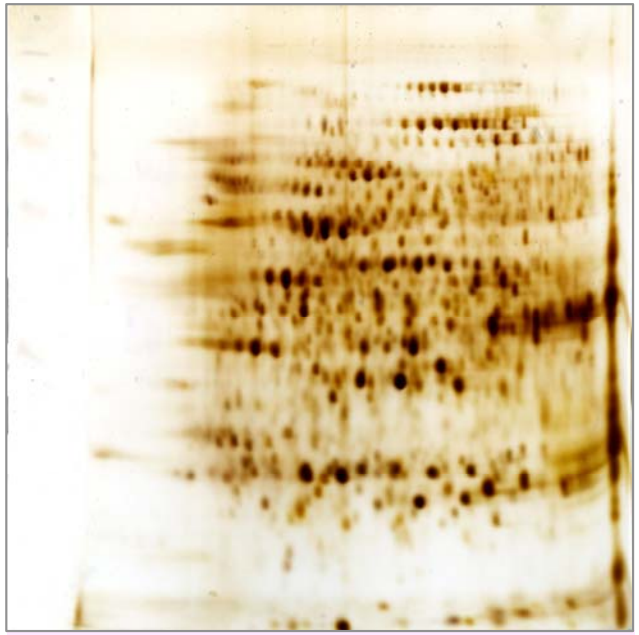
Total proteins extracted by different methods

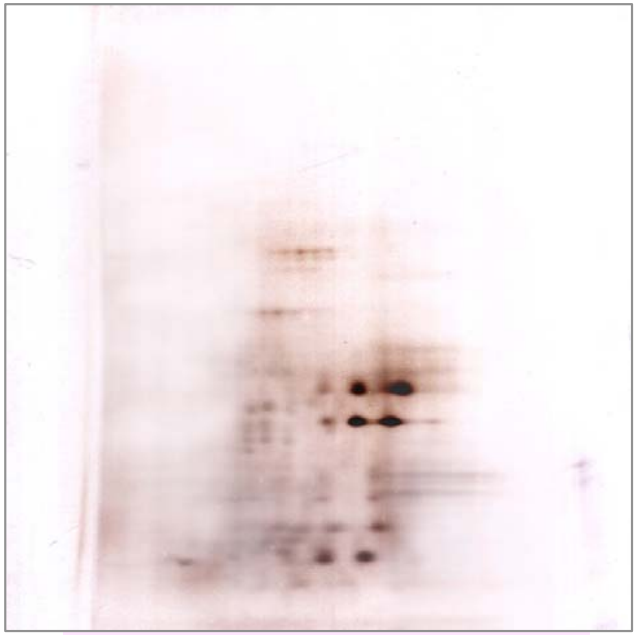
Underground shoot

60-cm shoot









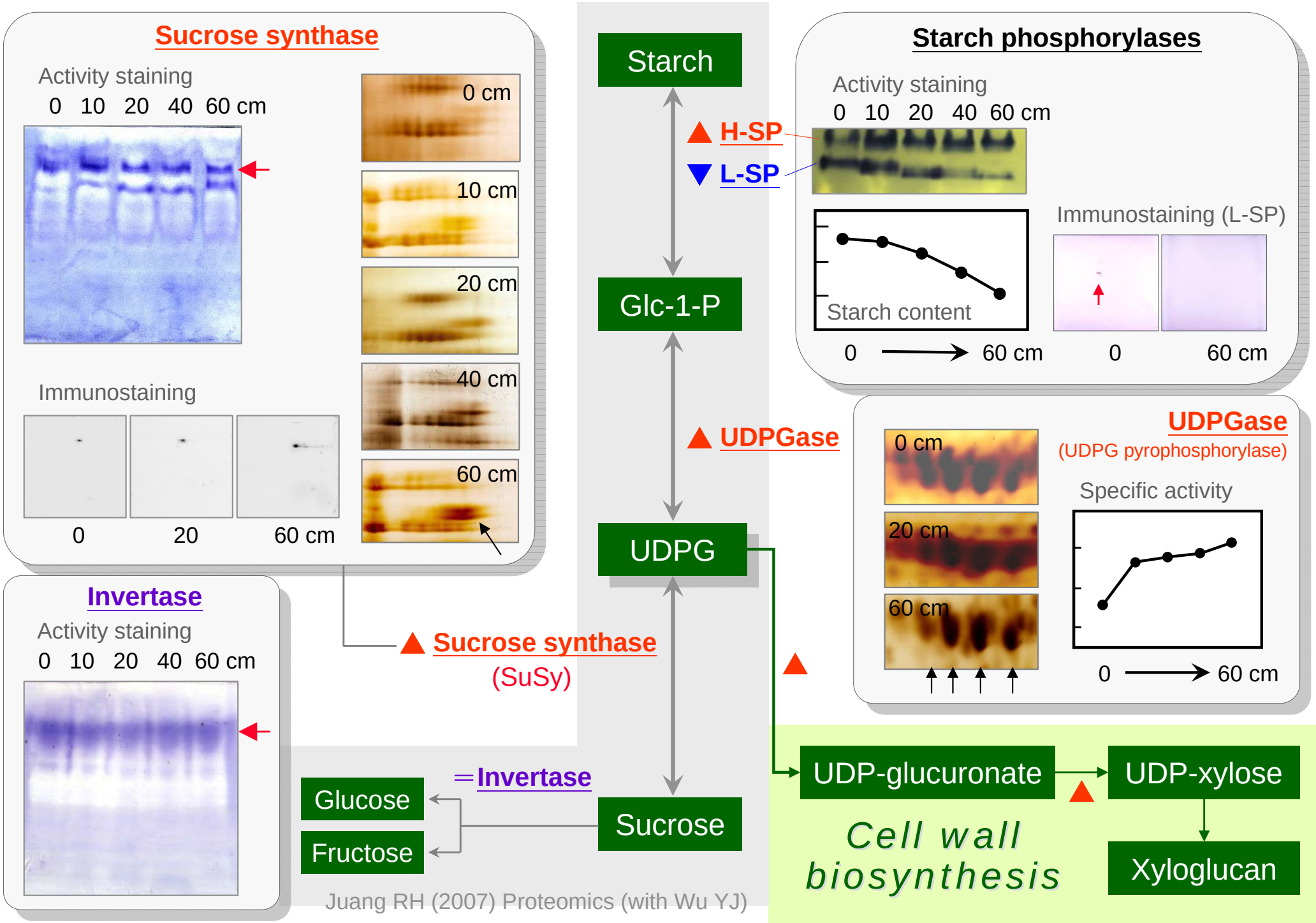
Water soluble proteins

Non-polar proteins

Glycoproteins

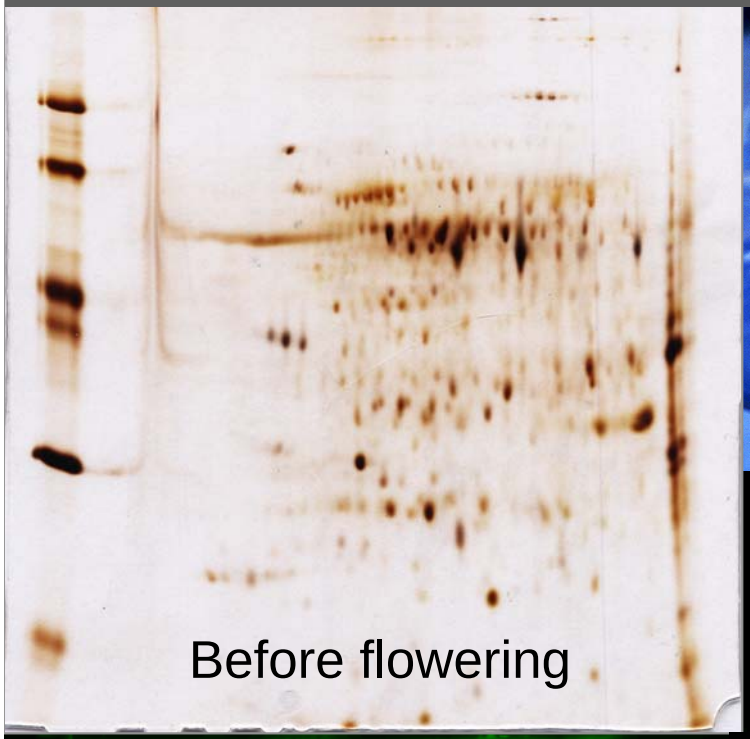
Juang RH (2007) Proteomics (with Wu YJ)

Connect 2-DE results to metabolic pathway

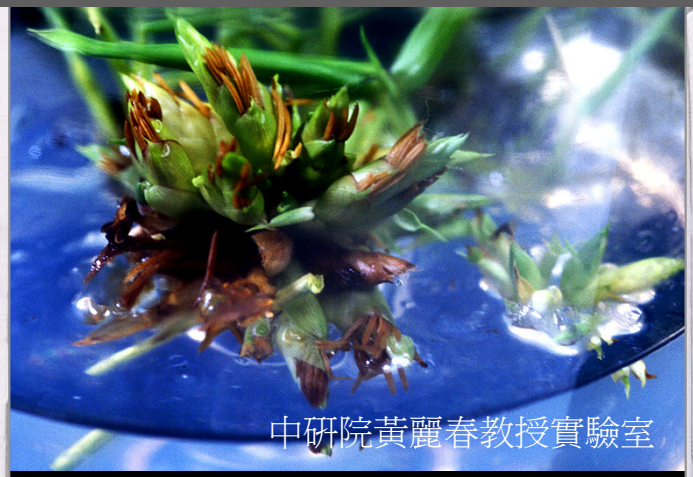


Juang RH (2007) Proteomics (with Wu YJ)

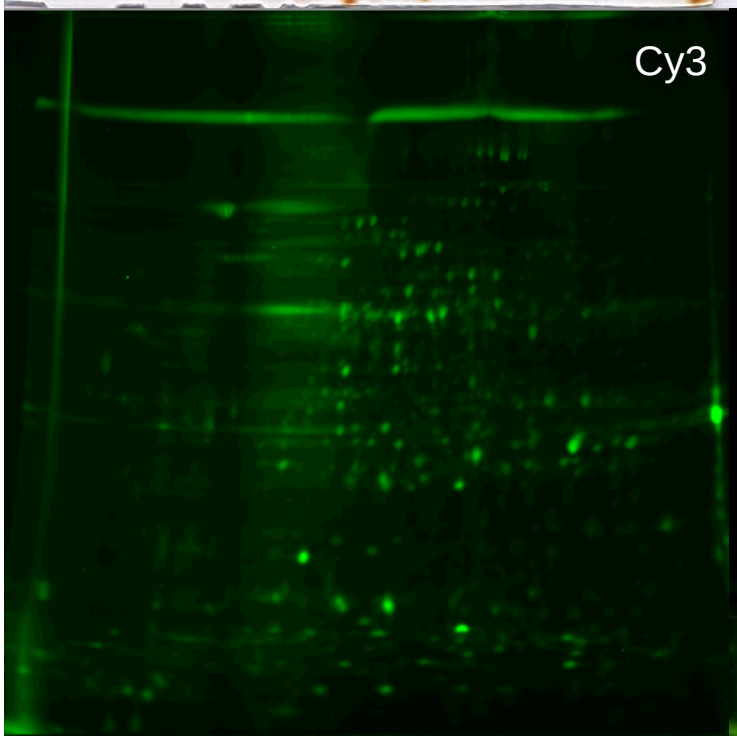
Proteomic change during bamboo flowering



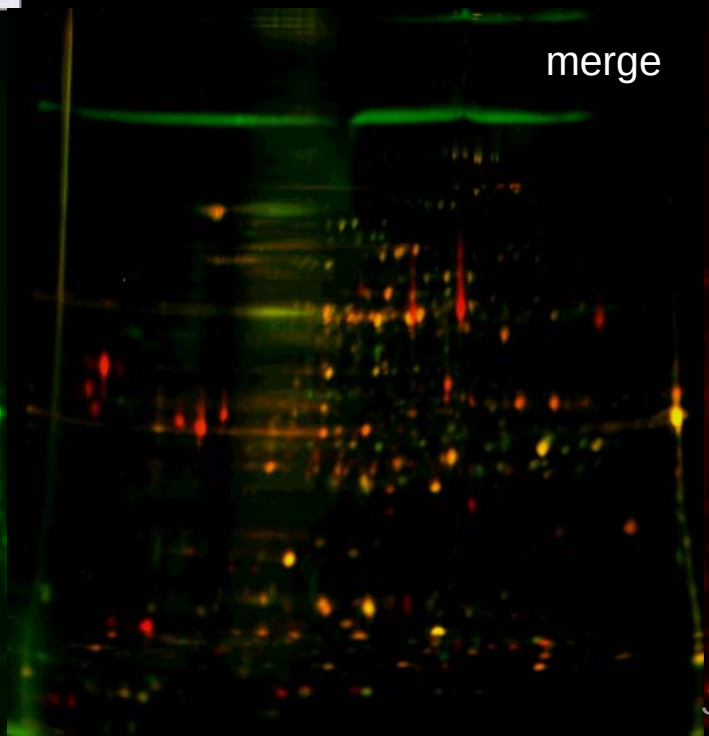
Before flowering



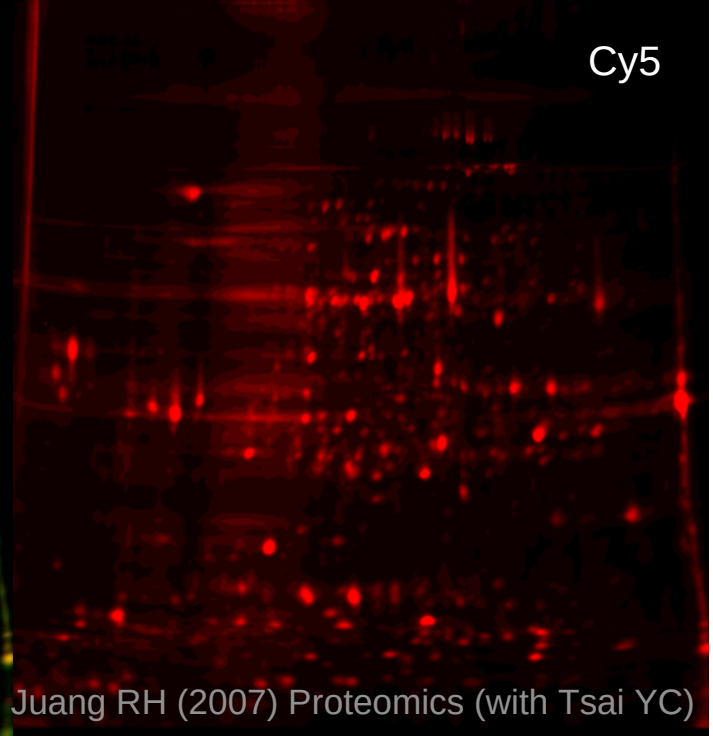
After flowering



Cy3



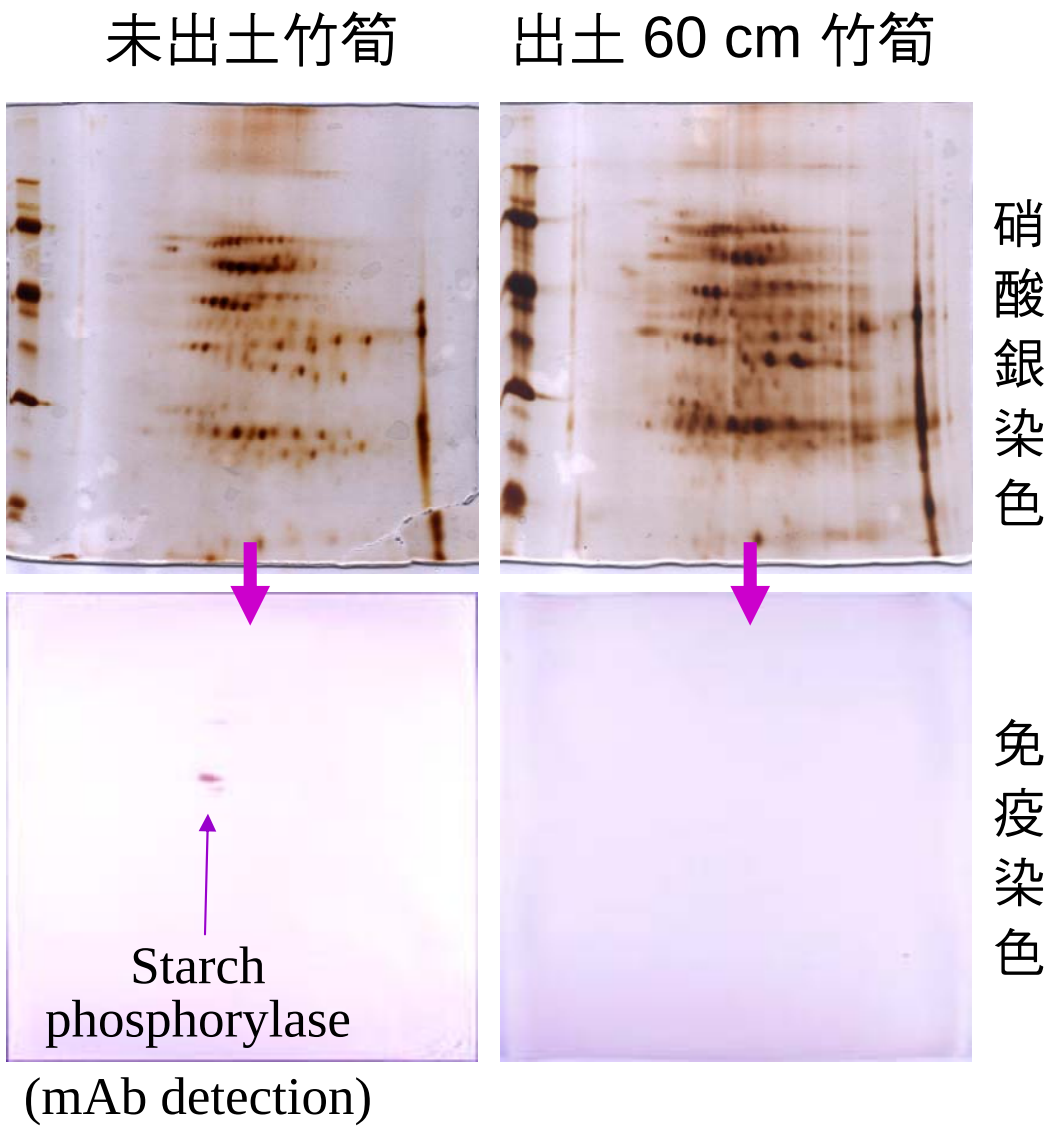
merge



Cy5

Juang RH (2007) Proteomics (with Tsai YC)

Antibody is specific probe against target protein



首次提出

高產能抗體製備計畫構想

(1) 建立高產能抗體製備之 標準流程

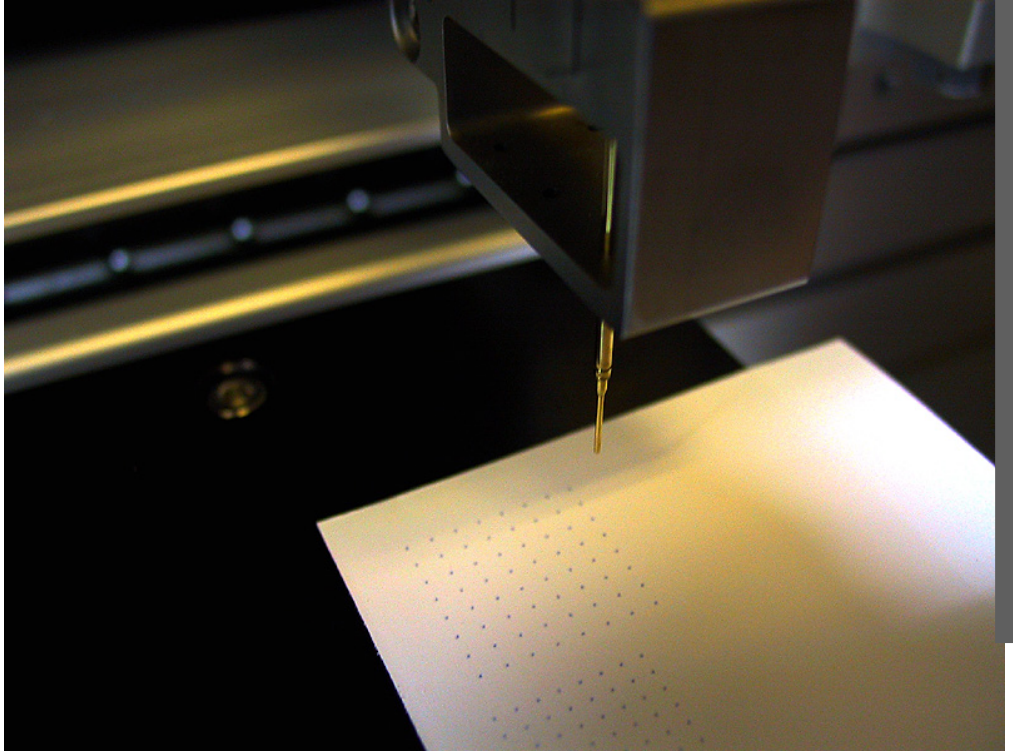
(2) 大量快速 生產有用的 單株抗體庫 mAb bank

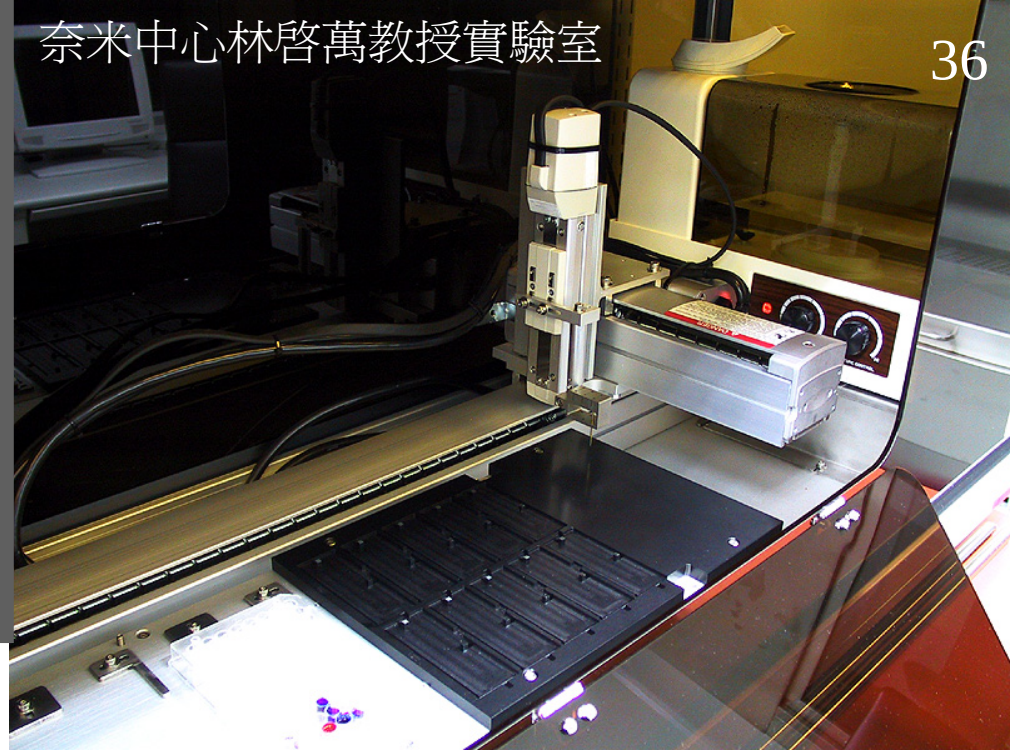
(3) 提供 抗體晶片 所需之抗體庫

均質抗原 → 單一抗體 → 整體抗原 → 全部抗體

Juang RH (2009) Proteomics (with Wu YJ)

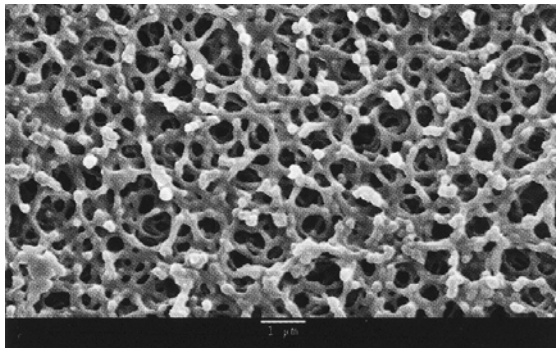
蛋白質晶片試製

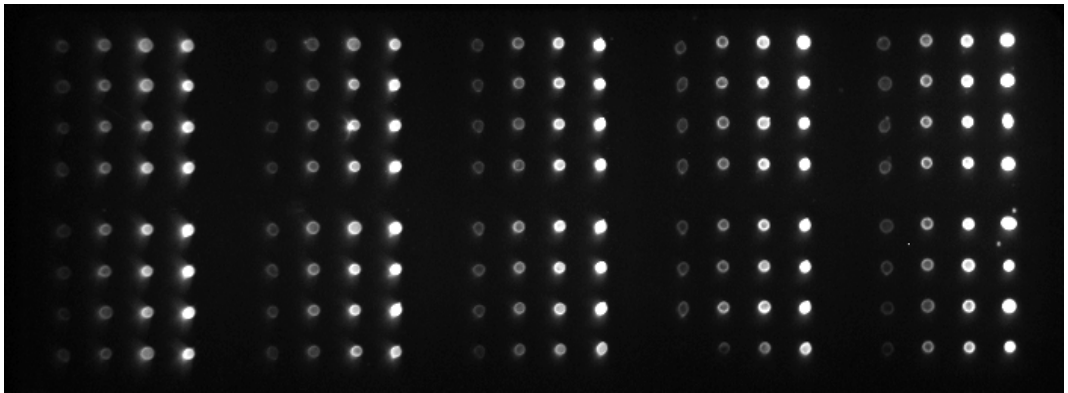




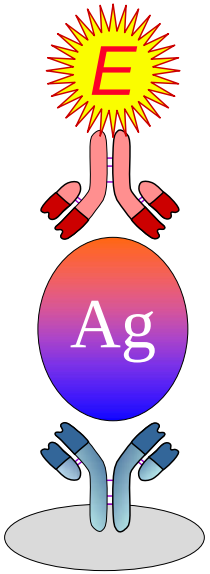
Schleicher & Schuell

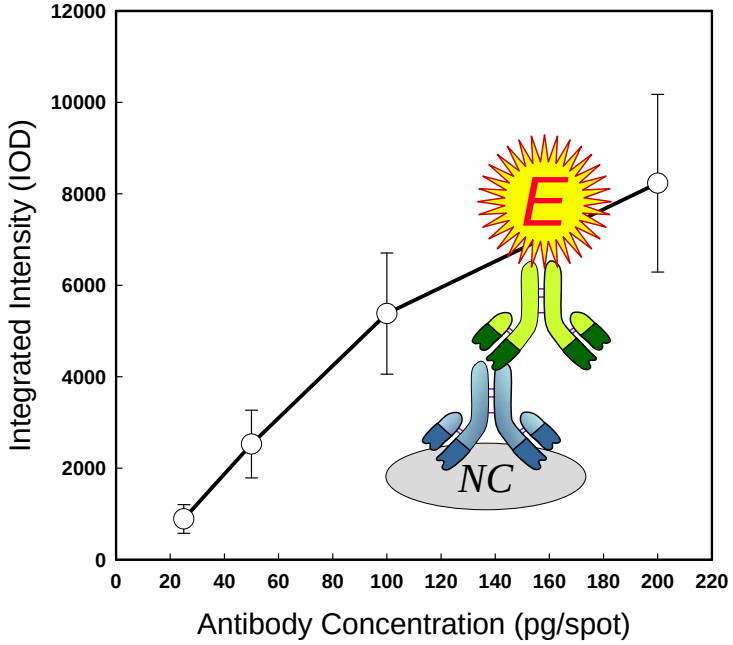
Nitrocellulose





三明治法

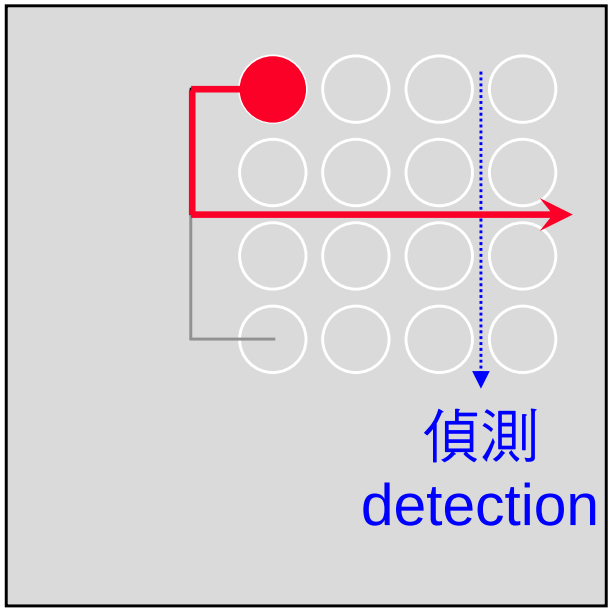




Juang RH (2007) Proteomics (with Ho JL)

Agilent 所有蛋白質純化與活性分析均予微小化

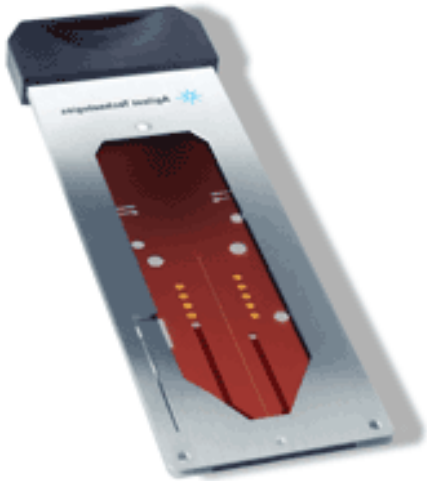
Agilent 2100 bioanalyzer

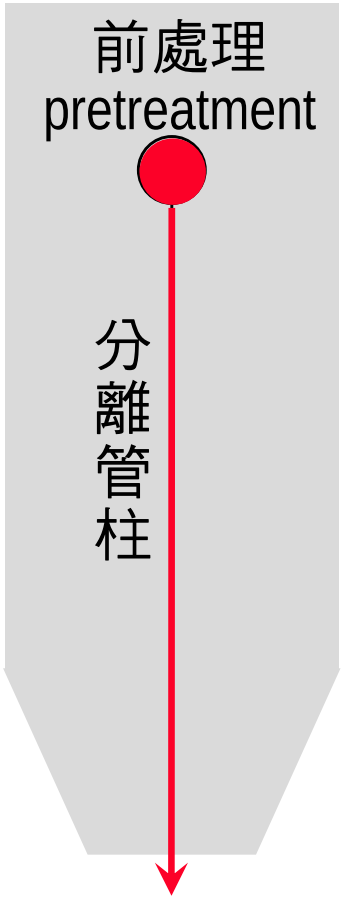


毛細管電泳
Capillary electrophoresis

http://www.chem.agilent.com/Scripts/Phome.asp

Agilent HPLC-Chip/MS

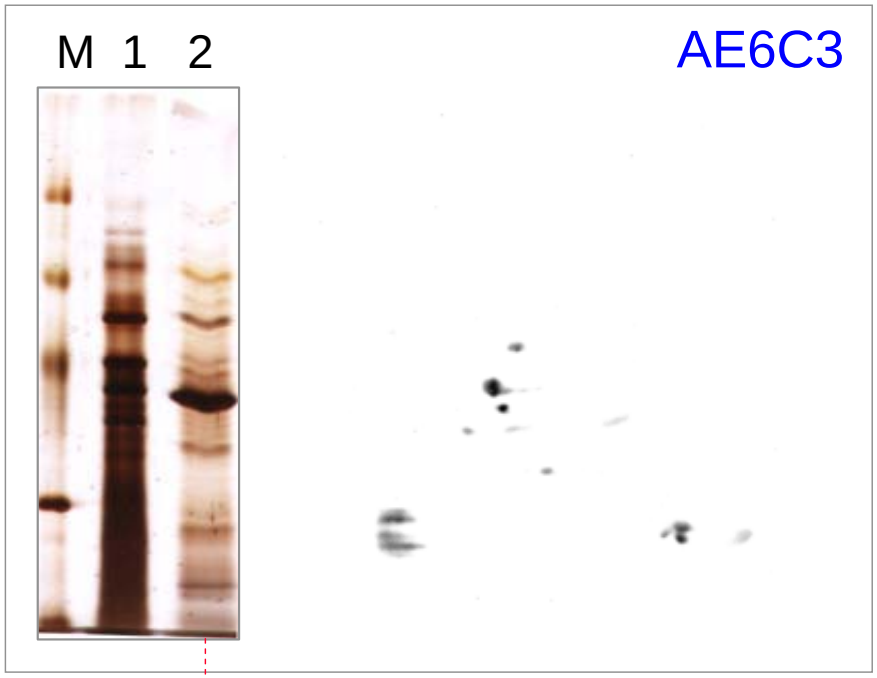


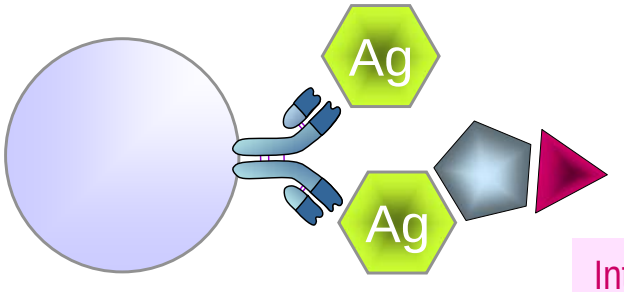


質譜儀分析
Mass analysis

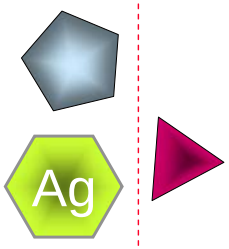
Minimized protein purification and analysis in one chip

Immunoprecipitation of interacted proteins





Interacted proteins



Interacted proteins

LC-MS/MS

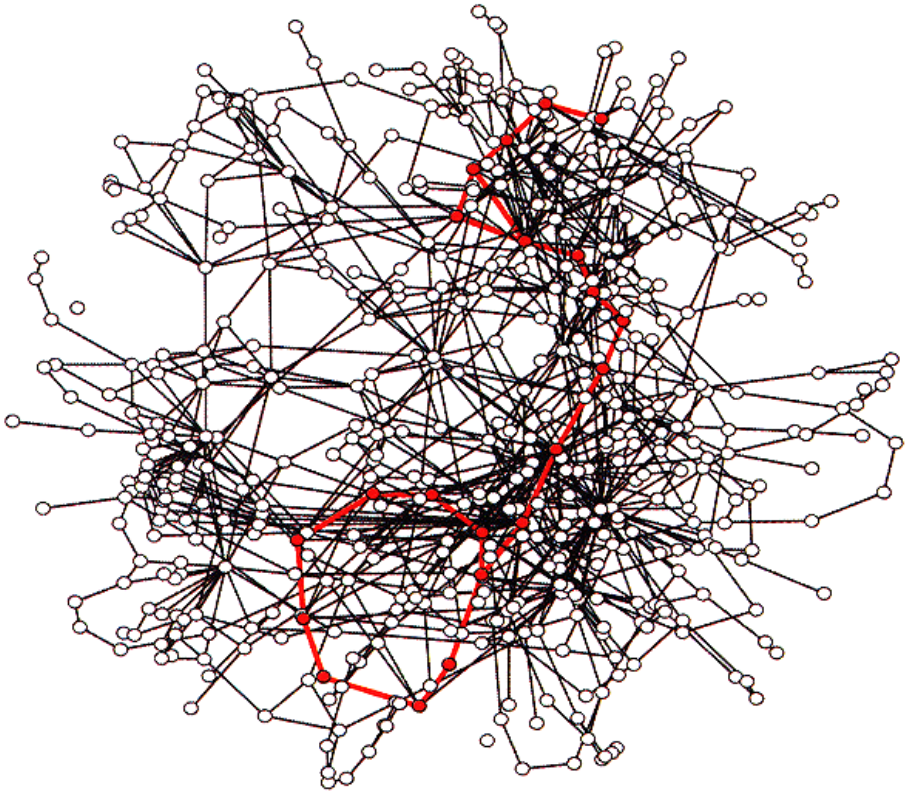
Protein ID	Match peptide
Adenosylhomocysteinase	IVLTIIR DSAAVFAWK HSLPDGLMR LVGVSEETTTGVK
Histone H4 (wheat)	IFLENVIR IDGLIYEETR TVRAMDVVYALKR
Fructose bisphophate aldolase	VTPEVIAEYTVR IGPNEPSQLAIDLNAQGLAR
Triosephosphate isomerase	TNVSPEVAESTR VIACVGETLEQR
NAD-dependent malate dehydrogenase	DDLFNINAGIVK
Histone H3	ASAPATGGVK
Putative lipase	DQVLEEVRR

Validation!

Juang RH (2007) Proteomics (with Wu YJ)

From proteomics to systems biology

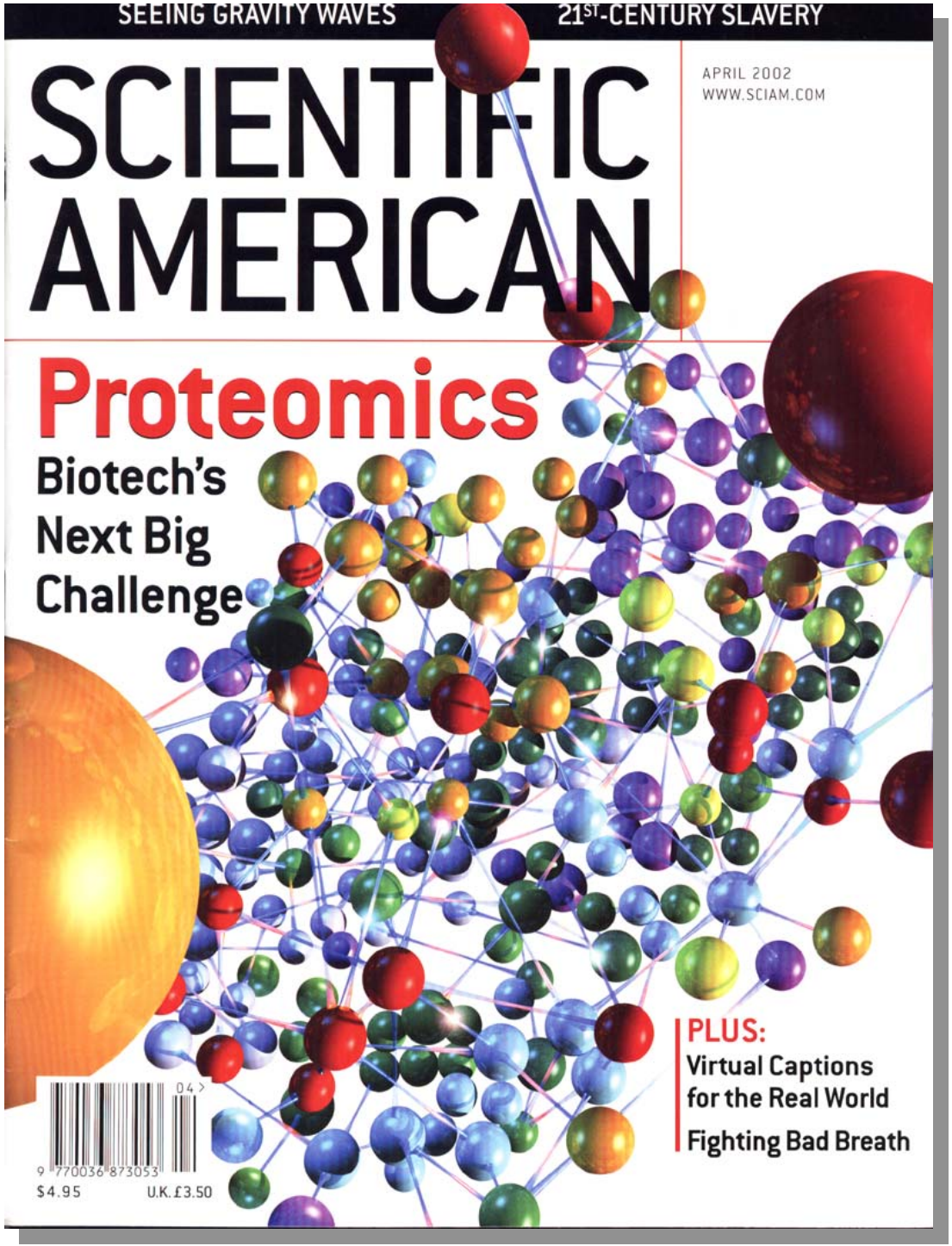
代謝路徑立體圖



Alberts et al (2002) Molecular Biology of the Cell (4e) p.107

Systems Biology

整體性的生物學觀念與工具



Scientific American (2002) cover page

燃燒的天堂 Medicine Man

一部老電影。一個特立獨行的科學家，在巴西原始叢林發現一種抗癌藥物，但是面對跨國公司的開發專案，那片叢林即將被毀 …

只有蛋白質體工具能夠及時挽救…

引申與想像的一些情節及問題：

(1) 你在巴西叢林中發現一種抗血癌植物

(2) 有效成份可能是一群揮發性的小分子

(3) 藥物微溶入血中與白蛋白結合後運送

(4) 藥物進入癌細胞後引發細胞程式死亡

(5) 這種植物要被特定螞蟻咬過才有效用

(6) 跨國公司在三個月後就要燒掉這森林

(7) **你如何用蛋白質體學工具進行探索？**

課後請與同學組成小組討論 (必考)