

- 15 Perrett S. Misshapes and misfits: protein misfolding and disease. *Chemistry & Industry*, 1998, **18**: 389~ 393
- 16 Jackson G S, Hosszu L L P, Power A, *et al.* Reversible conversion of monomeric human prion protein between native and fibrilogenic conformation. *Science*, 1999, **283** (5409): 1935~ 1937
- 17 Prusiner S B. Prion diseases and the BSE crisis. *Science*, 1998, **278** (5336): 245~ 251
- 18 Balter M. Prions: a lone killer or a vital accomplice? *Science*, 1999, **286** (5440): 660~ 662
- 19 Hardy J. The alzheimer family of diseases: many etiologies, one pathogenesis? *Proc Natl Acad Sci USA*, 1997, **94**: 2095~ 2097
- 20 Selkoe D J. Alzheimer's disease: genotypes, phenotype, and treatments. *Science*, 1997, **275**: 630~ 631
- 21 Vassar R. β -Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science*, 1999, **286**: 735~ 741
- 22 Haass C, Strooper B D. The presenilins in Alzheimer's disease: proteolysis holds the key. *Science*, 1999, **286** (5441): 916~ 919
- 23 Spillantini M G. α -Synuclein in Lewy bodies. *Nature*, 1997, **388**: 839~ 840
- 24 Polymeropoulos M H, Lavedent C, Leroy, E, *et al.* Mutation in the α -Synuclein gene identified in families with Parkinson's disease. *Science*, 1997, **276**: 2045~ 2047
- 25 Bullock A N, Henckel J, DeDecker B S, *et al.* Thermodynamic stability of wild-type and mutant p53 core domains. *Proc Natl Acad Sci USA*, 1997, **94**: 14338~ 14342
- 26 Janciauskiene S, Carlemalm E, Eriksson S. *In vitro* amyloid fibril formation from alpha 1-antitrypsin. *Biol Chem*, 1995, **376**: 103~ 109
- 27 Tagliavini F, McArthur R A, Canciani B, *et al.* Effectiveness of anthracycline against experimental prion disease in syrian hamsters. *Science*, 1997, **276** (5315): 1119~ 1121
- 28 Caughey B, Race R E. Potent inhibition of scrapie-associated PrP accumulation by congo red. *J Neurochem*, 1992, **59**: 768~ 771
- 29 Ingrosso L, Ladogana A, Pocchiari M. Congo red prolongs the incubation period in scrapie-infected hamsters. *J Virol*, 1995, **69**: 506~ 508
- 30 Foster B A, Coffey H A, Morin M J, *et al.* Pharmacological rescue of mutant p53 conformation and function. *Science*, 1999, **286** (5449): 2507~ 2510
- 31 Pennisi E. Bracing p53 for the war on cancer. *Science*, 1999, **286** (5449): 2431

Protein Misfolding and Disease. ZHOU Jun-Mei (*National Laboratory of Biomacromolecules, Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China*).

Abstract Proteins take center stage in directing the working of living cells. Every conceivable role within human bodies is played by proteins, from catalysis chemical reactions to defence against alien attack. A variety of quality control mechanisms that operate in the endoplasmic reticulum in downstream compartments of the secretory pathway to ensure the fidelity and regulation of protein expression during cell life and differentiation were introduced. The posttranslational quality control and its relation with protein misfolding were discussed in details. A number of diseases related with protein misfolding were introduced and the principles of therapy were discussed.

Key words protein misfolding, quality control, prion, disease, therapeutic agents

蛋白质组研究的技术体系及其进展*

成海平¹⁾ 钱小红²⁾

(军事医学科学院国家生物医学分析中心, 北京 100850)

摘要 随着后基因组时代的到来, 蛋白质组研究越来越受到国内外科学工作者的密切关注, 我国国家自然科学基金委员会已把蛋白质组研究列为重大科研项目。概述了蛋白质组研究中的基本技术, 包括双向凝胶电泳的样品制备和分离、蛋白质的检测、凝胶图像分析、蛋白质的鉴定以及蛋白质数据库构建等, 并就蛋白质鉴定的常用方法如氨基酸组成分析方法、蛋白质末端序列分析、肽质量指纹谱作了详细阐述。直观地列出了蛋白质组研究的技术体系流程图, 着重介绍了蛋白质组研究的最新技术及其进展。

关键词 蛋白质组, 双向凝胶电泳, 质谱, 凝胶图像分析, 数据库

学科分类号 Q51

* 国家自然科学基金重大项目资助 (39990600), 军事医学科学院科研创新启动基金资助 (1998~ 1999)。

¹⁾ 现工作单位: 总后卫生部药品检验所, 北京 100071。 ²⁾ 通讯联系人。

Tel: (010) 66930306, E-mail: qianxh@nic.bmi.ac.cn 收稿日期: 1999-12-31, 修回日期: 2000-06-22

人类自开展大规模基因组研究以来，迄今已取得了辉煌成就。十多种结构比较简单的生物基因组 DNA 全序列已经被先后测定，如大肠杆菌、酿酒酵母等等。人类基因组计划预计 2003 年前完成。在获得了生物基因的全部序列信息后，人类迫切需要了解这些基因的功能是什么，这些基因在生物体中如何发挥其作用等，由此标志着后基因组时代的到来。基因是遗传信息的携带者，生命功能的真正执行者是蛋白质，仅仅从基因的角度来研究是远远不够的，所以必须研究由基因编码和翻译的蛋白质，才能真正揭示生命活动的规律。由此产生了研究细胞内蛋白质组成及其活动规律的新兴学科——蛋白质组学 (proteomics)。

1994 年, 澳大利亚科学家 Wilkins 等首先提出了蛋白质组 (proteome) 的概念^[1], 并见诸

“Electrophoresis”：蛋白质组是指一个基因、一个细胞或组织所表达的全部蛋白质成分。蛋白质组研究是对不同时间和空间发挥功能的特定蛋白质群体的研究，从蛋白质水平上探索蛋白质作用模式、功能机理、调节控制以及蛋白质群体内相互作用，为临床诊断、病理研究、药物筛选、新药开发、新陈代谢途径研究等提供理论依据和基础。

蛋白质组研究技术体系的总流程图见图 1。蛋白质组研究的三大关键核心技术是：双向凝胶电泳技术、质谱鉴定技术、计算机图像数据处理与蛋白质组数据库。开展蛋白质组研究，首先是建立高通量、自动化的大规模蛋白质组研究技术体系^[2]，并且该体系具有灵敏度高、准确度好、重现性好的特点。下面根据蛋白质组研究的技术路线流程，分别加以阐述。

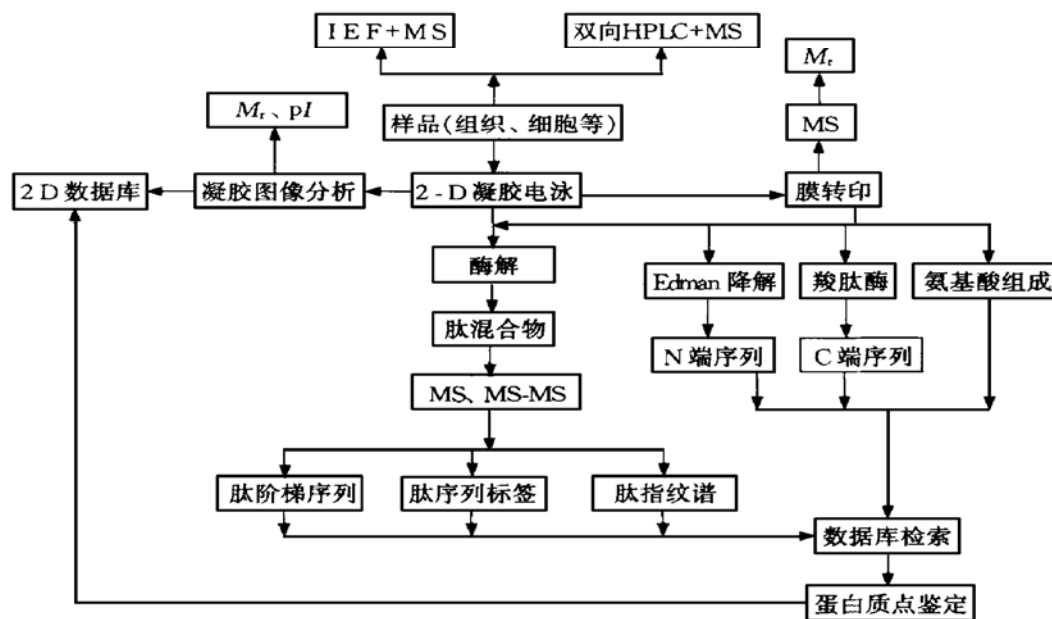


图 1 蛋白质组研究技术路线流程图

1 样品制备

样品制备是双向凝胶电泳成功的关键,直接影响到蛋白质组研究结果.其一般过程是:对细胞、组织等样品进行破碎、溶解、失活和还原,断开蛋白质之间的连接键,提取全部蛋白质,除去非蛋白质部分.但溶解度差的蛋白质,如膜蛋白、与膜相连的蛋白质以及来自具有高抗性组织(头发、皮肤等等)的蛋白质,需要添加破膜剂、兼性离子表面活性剂等以增加蛋白质的溶解度,提高蛋白质的提取率.有报道^[3]采用有机溶剂提取亲脂性蛋白;或采用不同溶出缓冲液多步提取各种膜蛋白.

2 双向聚丙烯酰胺凝胶电泳

1975年, O' Farrell、Klose 和 Scheele 分别建立了双向凝胶电泳 (two-dimensional polyacrylamide gel electrophoresis, 2-D PAGE)。目前在一张凝胶上能同时分离几千个甚至上万个蛋白质点。

双向凝胶电泳的第一向是等电聚焦 (isoelectric focusing, IEF), 根据蛋白质等电点 (pI) 不同而将之分离。最早采用载体两性电解质 pH 梯度等电聚焦。其中包括平衡 pH 梯度电泳 (ISO-DALT); 非平衡 pH 梯度电泳 (nonequilibrium pH gradients electrophoresis, NEPHGE)。后者常用于分离碱性

蛋白质, pI 7~11.5, 电泳展开时间相对较短。

1982年, Bjellqvist等发明了固相化 pH 梯度 (immobilized pH gradients, IPG) 等电聚焦。它的优点是: a. 不产生电极漂移; b. pH 梯度稳定, pH 相差 1 pH 单位的商品胶条已经上市; c. 上样量大, 可达几十毫克, 因此常用于蛋白质微量制备; d. 重复性好, 使不同人、不同实验室之间的实验结果能够相互比较, 这是它最突出的优点。

双向凝胶电泳的第二向是 SDS-聚丙烯酰胺凝胶电泳 (SDS-PAGE), 根据蛋白质分子质量不同而将之分离。一般采用垂直电泳或水平电泳, 两种方法的实验结果没有明显的差别, 均适用于分子质量在 10~150 ku 蛋白质。

判断双向凝胶电泳的好坏通过分析其图谱而获得。一张好的图谱必须低背景、少纹理、分辨率高、灵敏度高, 而且图谱之间必须存在良好的重现性。为了达到上述目的, 已经对双向凝胶电泳进行了改进^[2]。市售的微型 2-D PAGE 系统, 从 IPG 胶条的重泡胀到考马斯亮蓝染色, 整个操作 9 h 内完成。另外 Large Scale Biology 公司 (Rockville, MD, 美国) 开发研制了每周展开 100 块凝胶的仪器。还有 Proteome 公司 (Beverly, MA, 美国) 已经开发了一种完全自动化的仪器——Proteomatron, 6 张大胶 (40 cm × 40 cm) 从灌制、展开、染色全部实现自动化。从商业购买预制 IPG 胶条、SDS-聚丙烯酰胺垂直预制胶或水平预制胶。对样品采用同位素标记, 以此提高检测灵敏度。对样品采用两种不同的同位素进行标记, 如 ¹²⁵I、¹³¹I, 能在两种不同实验条件下, 在同一块凝胶上进行双向凝胶电泳, 根据标记的同位素不同, 可得到二张凝胶电泳图谱, 并能观察到不同的蛋白质表达。

为了获得一张包含细胞内表达的全部蛋白质图谱, 有报道采用“蛋白质组重叠群” (proteome contigs)^[4]方法。该法与高灵敏度的质谱联用能检测出低丰度蛋白质。双向凝胶电泳蛋白质点的矢量图 (vector map of the 2-D gel)^[5]是专门用来研究蛋白质翻译后是否修饰的一种方法, 通过此法能了解蛋白质翻译后修饰程度。

3 蛋白质点的染色

凝胶上的蛋白质点染色常用方法有银染法、考马斯亮蓝染色法, 另外还采用以下染色试剂^[6]: 丽春红 S、氨基黑、印度墨、³⁵S 硫脲银、胶态金、

咪唑锌等。银染法广泛用于双向凝胶电泳分离后蛋白质点的染色, 该法灵敏度为 4 ng, 但对质谱测定有干扰, 必须先脱银^[7]。考马斯亮蓝染色法可用于胶上或膜上蛋白质点染色, 该法操作方便、重现性好, 同银染法一样, 质谱测定时必须先脱色。

4 凝胶图像分析

凝胶图像分析首先是建立双向凝胶电泳图谱, 可用图像扫描仪、激光光密度仪、磷光或荧光测定仪等, 把双向凝胶电泳后凝胶上的蛋白质点数字化。

其次是对不同组织、不同细胞表达的蛋白质图谱进行比较。常用的软件有 Amersham Pharmacia 公司 ImageMaster2-D、Bio-Rad 公司 PDquest, 另外还有^[8] Tycho、Gellab、Kepler、GR42、Lips、Hermes、Elsie、Gemini 和 Melanie。运用软件对凝胶图谱去除纹理及背景, 找出表达上差异的蛋白质点, 并对其进行分析, 发现有生物学意义的蛋白质。

5 蛋白质鉴定

5.1 氨基酸组成分析

氨基酸组成分析作为蛋白质组研究中的鉴定方法之一, 该法经济、快速。但灵敏度低, 约十几皮摩尔。原理是通过测定蛋白质中各氨基酸所占摩尔百分数 (%) 或各氨基酸的摩尔比率, 然后与数据库中已知蛋白质的理论值进行比较, 给出匹配分数值大小, 分数值越小, 表示其与真正蛋白质越接近。

测定氨基酸组成常用的方法是酸水解, 酸水解时某些氨基酸被破坏^[8], 必须用已知蛋白质作校正, 常用人血清白蛋白。另外还有放射性标记法、蛋白酶水解法。

蛋白质经过 Edman 降解, 切除 N 端 3~4 个氨基酸后, 仍可测定氨基酸组成, 检索软件为改良 AACompIdent。用该法测定了大肠杆菌^[9]双向凝胶电泳后蛋白质点。

当测得氨基酸组成后, 最后一步是进行数据库检索。常用的软件是 AACompIdent (从 ExPASy 获得), 另外还有软件^[8] ASA、FINDER、AAC-pI、PROP-SEARCH。

5.2 蛋白质和多肽的 N 端、C 端氨基酸序列分析

蛋白质的末端氨基酸序列具有惊人的专一性。43%~83% 蛋白质可用 N 端 4 个氨基酸残基来确

定, 74%~97% 蛋白质可用 C 端 4 个氨基酸残基确定, 因此 C 端比 N 端更具专一性. 若测定出蛋白质 N 端或 C 端 5 个氨基酸残基, 鉴定蛋白质的专一性则更高^[10].

N 端氨基酸序列分析仪是依据 Edman 降解原理测定蛋白质 N 端氨基酸序列, 其灵敏度达几百飞摩尔. 对于 N 端封闭的蛋白质, 可用溴化氰、内肽酶或外肽酶使蛋白质断裂, 产生新的 N 端, 再测定其序列. 但是该方法测定时间长、费用高.

测定 C 端氨基酸序列时, 需样量约 100 pmol, 只要测定出 C 端 3~5 个氨基酸残基, 即可鉴定蛋白质. 常用的测定方法有羧肽酶法、化学降解法、C 端片段分离法以及物理法等. 目前蛋白质 C 端氨基酸序列分析仪已逐步实现自动化.

最后依据 N 端、C 端氨基酸序列检索蛋白质. 常用的软件是 TagIdent, 一次只能输入 6 个氨基酸残基.

5.3 质谱技术

质谱鉴定蛋白质, 其优点是灵敏度高、准确度高, 而且容易实现自动化. 因此, 质谱技术是蛋白质组研究中最重要鉴定技术.

伴随电喷雾电离 (ESI) 和基质辅助激光解吸电离 (MALDI) 两项具有划时代的“软电离”技术的出现, 各种新质谱技术不断涌现.

ESI 是以连续离子化方式使样品电离, 此技术已经运用到四极杆质谱^[11]、飞行时间质谱 (TOF)、离子阱质谱 (ITMS) 和傅里叶变换离子回旋加速器质谱 (FTMS), 但还不能用于磁质谱中.

ESI 是液相进样, 它可以与高效液相色谱 (HPLC)、毛细管电泳 (CE) 等技术相连. 双向凝胶电泳分离的蛋白质点, 经酶解后, 生成的肽混合物不需分离, 可直接导入下列带有 ESI 的串联质谱 (MS/MS) 分析仪中: HPLC-MS/MS 适合含 10 pmol 以上的肽样品; 毛细管 HPLC-MS/MS 适合含几个皮摩尔到几十个飞摩尔肽样品; CE-MS/MS 适合含飞摩尔或更低肽样品, 此仪器灵敏度最高. LC-MS/MS 已经实现了自动化; 固相提取毛细管区带电泳与串联质谱联用初步实现自动化. 有研究小组已经实现 ESI-MS 分析的高通量, 把标准的 96 孔微装置与独立的 ESI 头相连, 采用计算机控制, 在 10 s 内即可完成每个样品的分析^[12, 13].

MALDI 是以脉冲离子化方式使样品电离. 它常与飞行时间质谱 (TOF-MS) 联用, 称为基质辅

助激光解吸电离飞行时间质谱 (MALDI-TOF-MS). 为了提高检测精确度, 研制出了反射飞行时间质谱和延迟飞行时间质谱. 目前蛋白质的肽指纹谱 (peptide fingerprinting, PMF) 测定常用 MALDI-TOF-MS, 也有用 ESI-MS. 这两种方法精确度高, 均可达 0.1 个质量单位; 灵敏度高, 分解亚皮摩尔量的蛋白质, 所产生的小肽段, 也可以做质谱分析; 分析时间短, 测定仅需几分钟. MALDI-TOF-MS 测定蛋白质肽指纹谱已经实现了高度自动化集成系统^[14], 在不到 10 个工作日内鉴定了大肠杆菌双向凝胶上 95 个蛋白质点.

有报道^[2]将蛋白质转印到偶合了蛋白酶的聚偏二氟乙烯 (PVDF) 膜上, 直接进行 MALDI-MS 肽指纹谱测定; 还可采用红外激光质谱, 直接对 PVDF 膜上的未修饰蛋白质质量进行测定.

MS/MS 可直接测定肽混合物中某一特定肽段的肽阶梯序列 (peptide ladder sequencing) 或肽序列标签 (peptide sequence tag). 因此对于翻译后修饰的蛋白质用串联质谱来鉴定具有独特的优点, 如蛋白质磷酸化^[15]测定. 纳喷串联质谱 (nano-electrospray tandem mass spectrometry) 灵敏度达到了飞摩尔. 用该法对 Burkitt 淋巴细胞株 2-D 凝胶电泳后 36 个蛋白质点进行鉴定^[16], 其中 33 个蛋白质点能从数据库找出, 3 个蛋白质点未能从数据库中找出.

目前, 有两种分离、鉴定蛋白质的方法类似于双向凝胶电泳: 一种是采用等电聚焦-固相 pH 梯度 (IEF-IPG) 作为第一向分离^[17], 质谱测定分子质量作为第二向. 另一种是采用双向高效液相与质谱联用来鉴定蛋白质^[18].

质谱数据的数据库检索: 三条肽段的精确质量数便可用来鉴定蛋白质, 但是, 最好有 4~6 个肽段质量. 根据所测数据, 可采用不同的软件检索^[19]: PeptideSearch、TagIdent、Mass Search、MsFit、ProFound、MsEdman、MOWSE、MsTag、PepFrag、Sequest、MutIdent.

6 蛋白质组数据库

蛋白质组数据库贮存了有机体、组织或细胞所表达的全部蛋白质信息, 通过用鼠标点击双向凝胶图谱上的蛋白质点获得这些信息, 如蛋白质鉴定结果; 蛋白质的亚细胞定位; 在不同条件下有机体、组织或细胞蛋白质的表达水平等. 目前可从 ExPASy 上下载 Make2ddb 软件来构建蛋白质组的

双向凝胶数据库. 查找蛋白质组数据库网址是 2Dhunt (<http://www.expasy.ch/ch2d/2DHunt>), 一个有机体、组织或细胞的蛋白质组数据库可能由不同研究小组构建, 它们各有特点, 检索时数据库之间互相连接, 可用索引检索: WORLD-2DPAGE (<http://www.expasy.ch/ch2d/2d-index.html>).

同基因组静态过程相比较, 蛋白质组是一种动态过程. 因此, 目前仅有酵母蛋白质组数据库 (YPD, <http://www.ypd.com>) 是完整的蛋白质组数据库^[20], 其他数据库需要不断完善.

综上所述, 蛋白质组研究的技术体系为人类从蛋白质水平上揭开生命活动规律提供了有力的武器.

参 考 文 献

- Wasinger V C, Cordwell S J, Anne C D, *et al.* Progress with gene-product mapping of the *Mollicutes*: *Mycoplasma genitalium*. Electrophoresis, 1995, **16** (17): 1090~1094
- Quardroni M, James P. Proteomics and automation. Electrophoresis, 1999, **20** (4~5): 664~677
- Herbert B. Advances in protein solubilisation for two-dimensional electrophoresis. Electrophoresis, 1999, **20** (4~5): 660~663
- Ian H S, Walter B. Proteome analysis: genomics via the output rather than the input code. J Protein Chem, 1997, **16** (5): 537~544
- Wilkins M R, Sanchez J C, Williams K L, *et al.* Current challenges and future application for protein maps and post-translation vector maps in proteome projects. Electrophoresis, 1996, **17** (5): 830~838
- Wilkins M R, Sanchez J C, Gooley A A, *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genetic Eng Rev, 1996, **13** (1): 19~50
- Gharahdaghi F, Weinberg C R, Meagher D A, *et al.* Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. Electrophoresis, 1999, **20** (3): 601~605
- Ian H S, Cordwell S J, Blackstock W P. Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis, 1997, **18** (8): 1217~1242
- Wilkins M R, Gasteiger E, Sanchez J C, *et al.* Protein identification with sequence tags. Curr Biol, 1996, **6** (12): 1543~1544
- Wilkins M R, Gasteiger E, Tonella L, *et al.* Protein identification with N and C-terminal sequence tags in proteome projects. J Mol Biol, 1998, **278** (3): 599~608
- Yates J R. Mass spectrometry and the age of the proteome. J Mass Spectrometry, 1998, **33** (1): 1~9
- Figeys D, Gygi S P, Zhang Y N, *et al.* Electrophoresis combined with novel mass spectrometry techniques: Powerful tools for the analysis of proteins and proteome. Electrophoresis, 1998, **19** (10): 1811~1818
- Haynes P A, Gygi S P, Figeys D, *et al.* Proteome analysis: Biological assay or data archive? Electrophoresis, 1998, **19** (11): 1862~1871
- Traini M, Gooley A A, Ou K, *et al.* Towards an automated approach for protein identification in proteome projects. Electrophoresis, 1998, **19** (11): 1941~1949
- Carr S A, Huddleston M J, Annan R S. Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. Anal Biochem, 1996, **239** (2): 180~192
- Muller E C, Schumann M, Rickers A, *et al.* Study of Burkitt lymphoma cell line proteins by high resolution two-dimensional gel electrophoresis and nanoelectrospray mass spectrometry. Electrophoresis, 1999, **20** (2): 320~330
- Joseph A L, Jeffrey B, Glenn C, *et al.* High sensitivity mass spectrometric methods for obtaining intact molecular weights from gel-separated proteins. Electrophoresis, 1999, **20** (4~5): 743~748
- Gregory J O, Suzanne M R, James W J. Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. Anal Biochem, 1998, **258** (2): 349~361
- 丁达夫, 盛泉虎, 解涛. 蛋白质组信息学. 生命的化学, 1998, **18** (6): 10~16
- Ding D F, Sheng X H, Xie T. Chem Life, 1998, **18** (6): 10~16
- Payne W E, Garrels J I. Yeast protein database (YPD): a database for the complete proteome of *Saccharomyces cerevisiae*. Nucleic Acids Res, 1997, **20** (4~5): 57~62

The Technological System of Proteome Research and Its Progress. CHENG Hai-Ping, QIAN Xiao-Hong (*National Center for Biomedical Analysis, Beijing 100850, China*).

Abstract With the coming of the post-genome era, proteome research, which was sponsored by National Natural Science Foundation of China in our country, has increasingly caught many biochemists' attention. The fundamental technologies of proteome research are briefly introduced, including sample preparation, protein separation, protein spot detection, gel image analysis, protein identification, database construction, *et al.* The commonly used identification methods are amino acid analysis, protein sequence tags and peptide mass fingerprinting. The flow chart of proteome analysis is shown, and in particular the latest proteome technologies are summarized.

Key words proteome, two-dimensional gel electrophoresis, mass spectrometry, gel image analysis, database