How many proteins are there in humans

人体内有多少种蛋白质

王美林, 张润东, 苏丹*

四川大学华西医院生物治疗国家重点实验室, 成都 610041

* 联系人, E-mail: sudan@scu.edu.cn

2016-09-12 收稿, 2017-04-10 修回, 2017-04-11 接受, 2017-08-22 网络版发表

摘要 2005年*Science*提出了“人体内有多少个蛋白质?”这一科学问题。随着人类基因组计划的不断完善, 以及“后基因组时代”在蛋白质组学方面研究的不断深入, 科学家们都尝试来回答这一有趣的问题。经典分子生物学认为, 一个基因只能表达一种蛋白, 人类基因组含有约24000个基因, 但蛋白质种类有50000多种甚至更多, 显然可能有些基因可以表达的蛋白至少超过两种以上, 才能够使机体复杂的功能需求得以满足。至今为止, 科学家们对人体内蛋白质具体数量还没有达成共识。本文结合真核模式生物的进化、基因编码、蛋白质修饰以及当前蛋白质组学技术发展, 综合分析人体内可能存在多少种蛋白质, 希望能够为回答这一科学问题提供一定的帮助。

关键词 蛋白质, 基因编码, 蛋白剪接, 翻译后修饰, 质谱技术

蛋白质是一切生命活动的直接执行者, 在整个过程中都有参与, 如物种遗传、细胞发育、繁殖、物质能量代谢、应激等^[1]。生命体中如此纷繁复杂的生物过程, 自然需要大量不同种类的蛋白质来完成。奇妙的是, 生命体中的蛋白质仅由20种氨基酸组成^[2]。然而在理论上有限的氨基酸种类经过不同长度及排列组合方式可以形成数量庞大的蛋白质库。以一个由100个氨基酸组成的蛋白质为例, 其理论上可以拥有的种类为 $20^{100}=1.268 \times 10^{130}$, 如果以每个蛋白分子的大小为 15.6 nm^3 计算(假设是边长为 2.5 nm 的立方体。这是一个合理的估计, 因为肌红蛋白有大约150个氨基酸(amino acid, aa), 体积是 19.2 nm^3 , (PDB code: 1MBC)), 将所有的这些蛋白一个一个地排列, 其总体积达到 $1.98 \times 10^{104} \text{ m}^3$, 这是个无法想像的数量, 因为太阳的体积只有 $1.41 \times 10^{27} \text{ m}^3$ 。如果再将蛋白质不同长度考虑进去的话, 那理论上种类更加无法令人想像。显然在地球上的生命体中不可能有这么多种的蛋白质存在。生命体的进化过程中, 并不是理论上所有可能存在的蛋白都会出现, 即使能

够出现, 如果对生命体的生存没有益处的话, 其存在也可能只是短暂的。很容易想象到, 只有极其有限种类的蛋白质最终能够整合进精密的生命体中, 并在漫长的生存斗争中保留下来。

生命体中的蛋白都是从基因组DNA编码而来的, 因此考察基因组DNA中编码框的数目, 可以作为推算蛋白质种类的基础。虽然人们早期有“一个基因, 一个蛋白”的初浅认识, 但随着研究的深入, 尤其现代基因组学及蛋白质组学的研究发展, 人们认识到DNA中基因与蛋白的对应关系要复杂得多。从编码蛋白的基因到成熟蛋白整个过程, 需要经过转录、翻译、翻译后加工修饰等多个环节^[3]。每个环节都有产生多种蛋白的可能性, 因此最终得到的蛋白质种类要远多于基因数量。2003年人类基因组计划完成, 其中99%的人类基因组序列获得了精确测序, 人类只有近24000个基因, 但其中90%的基因具有一个以上的mRNA转录本, 这样就直接导致了蛋白质的数量远超过人类基因的数量^[4]。因此, 人体具有多少种蛋白质, 问题看似简单, 却很难给出一个精确的答案。

引用格式: 王美林, 张润东, 苏丹. 人体内有多少种蛋白质. 科学通报, 2017, 62: 3256–3261

Wang M L, Zhang R D, Su D. How many proteins are there in humans (in Chinese). Chin Sci Bull, 2017, 62: 3256–3261, doi: 10.1360/N972016-01016

同时, 蛋白氨基酸序列的突变, 为蛋白质数量计数带来了新的困难, 一个氨基酸的突变往往导致蛋白结构及其生物学功能的改变, 突变后的蛋白显然又变成了一种“新”的蛋白, 如何定义一个新的蛋白成为计数蛋白种类的另一个难题。

最终, 根据真核模式生物蛋白质数据库统计规律研究蛋白质, 在人体内存在50000种不同类别蛋白质。然而, 通过体内进一步的酶切、修饰加工等处理, 人体内可以产生超过200万种的蛋白质分子^[5]。本文将从单个基因如何翻译生成一个成熟蛋白的全过程, 来揭示蛋白多样性的原因, 希望对回答蛋白数目这一问题提供参考。

1 基因序列多态性

基因是细胞合成蛋白的蓝图。基因序列在遗传复制上的准确性及稳定性对于生命物种的传承非常重要。然而, 基因突变导致的个体基因多态性, 使得生命个体之间在同一个基因序列上往往存在着一些细小的差异。这些基因序列微小的变化, 可能使得转录翻译后最终形成的蛋白在结构和功能上具有明显的差异。例如, 人体肝脏中的乙醛脱氢酶(aldehyde dehydrogenase, ALDH), 负责将乙醛转化为乙酸, 是酒精代谢中的关键酶^[6]。虽然体内有ALDH1和ALDH2两种同工酶, 但因为后者的活性远高于前者, 因此ALDH2是主要负责乙醛转化的酶。乙醛脱氢酶ALDH2基因位于人的12号染色体(12q24.2), 正常人群中部分人体内基因的1510位存在单个碱基G→A的突变, 从而造成最终生成的酶在487位残基由谷氨酸(Glu)变为赖氨酸(Lys), 这种点突变造成生成的酶催化活性丧失, 记为ALDH2*2, 与其对应的具有正常催化活性的酶为ALDH2*1。ALDH2*2基因在人类族群中的分布不同, 但基本全部出现在亚洲人中。因为ALDH2*2的表型又是显性的, 携带ALDH2*2基因的人, 即使是只携带1/2 ALDH2*2基因的杂合子, 也很难正常代谢乙醇的氧化产物乙醛, 所以饮酒后血液乙醛浓度增高会出现一些不良反应, 如脸红、头晕、心跳加快等^[7]。正常人群中这种等位基因中单个碱基突变的情况很多, 称为遗传多态性(genetic polymorphism)或基因多态性(gene polymorphism)^[8]。遗传多态性常发生在非编码区, 在编码区也常常不影响蛋白的编码。编码区突变造成蛋白结构功能发生变化的例子还有很多, 另一个著名的例子是引起

镰刀形细胞贫血症的血红蛋白中的珠蛋白^[9]。其 β 基因发生单碱基突变, 正常 β 基因的第6位密码子GAG突变为GTG, 造成谷氨酸(Glu)变为缬氨酸(Val), 使之成为镰型红细胞症血红蛋白HbS。在纯合子状态, 当形成HbS后, HbS在脱氧状态下聚集成多聚体, 达到一定数量时, 细胞膜就会由正常的双凹形盘状变成异常的镰刀形。该细胞变形性差, 僵硬, 容易破裂而溶血, 从而造成贫血症。由此可见, 基因的多态性会影响蛋白结构功能发生变化, 直接影响蛋白质的数量。

2 基因转录过程中的选择性剪接

选择性剪接, 也称可变剪接(alternative splicing), 指通过不同的剪接方式(选择不同的剪接位点组合), 从一个mRNA前体(pre-mRNA)中, 将不同外显子片段融合在一起, 从而产生多种成熟mRNA剪接异构体的过程。不同的剪切、拼接方式能够表达形成多种不同功能和作用的产物, 如降钙素基因是由6个外显子片段构成, 由于对其mRNA前体的剪切方式的不同, 细胞可以将外显子片段1, 2, 3, 4, 5拼接成降钙素来参与调节体内血钙水平; 外显子片段1, 2, 3, 4, 6却拼接成与降钙素基因相关的肽(calcitonin gene-related peptide, CGRP)负责调控机体血压及血流量^[10]。通过对mRNA前体不同的剪切方式, 形成功能不同的蛋白分子, 这种类似的情况还有很多。转录过程中的选择性剪接使得基因最终的蛋白表达产物表现出不同的功能和结构特性, 成为不同的蛋白亚型, 这也是真核生物增加mRNA和蛋白多样性的重要手段。

3 翻译后的蛋白剪接

蛋白剪接是一个翻译完成后自催化及加工过程, 它不需要相关酶或者其他的辅助因子的参与^[11~13]。蛋白剪接与RNA剪接有所不同, RNA的剪接发生在RNA水平而蛋白剪接则发生在蛋白水平; 同样, 蛋白剪接与蛋白翻译后加工也不相同, 蛋白翻译后加工过程通常只是切除前体蛋白的C端或者N端的一些片段, 往往是没有多肽链的形成; 当然也不同于蛋白酶原活化, 蛋白剪接是有新的肽键形成。在剪接过程中, 前体蛋白的内含肽(intein)被切离, 其两侧的外显肽(extein)连接在一起, 多数intein一般由134~650个氨基酸残基组成^[14]。内含肽的特殊性在于其蛋白的剪接功能, 即能以1个正常肽键连接两个多肽。它的

蛋白剪接功能又可以分解成N端和C端的切割功能^[15], 这些特殊的内含肽(或经修饰后的内含肽)自主催化完成的生化反应为蛋白的合成以及加工提供了新的途径. 目前, 已经发现的intein超过了400个, 在生命的真细菌、古细菌和真核生物中都有发现, 但是在人体内尚未发现具有相同功能的蛋白^[16], 因此这种蛋白翻译后的自身剪切机制在人体内还未见报道.

4 活性肽分子在体内的酶解

活性肽分子在人体内的酶解程度不同, 可以形成许多种不同的中间产物. 活性肽是由具有低活性或者无活性的大分子形式的前体肽原(preproptide)生成, 在体内则酶解成多个肽段, 不同的肽段又具有相对独立的生物学活性, 不同的多肽段间也可以通过彼此的相互作用来共同参与完成各项生理调节. 例如, 由230 aa组成的 β -内啡肽的前体蛋白(proopiomelanocortin, POMC), 它可分解成 β -内啡肽、g-趋脂素、 α -促黑色素细胞激素(α -melanocyte-stimulating hormone, α -MSH)、促肾上腺皮质激素释放激素、g-MSH等, β -内啡肽又可以分解为g-内啡肽(1~17)、 α -内啡肽、内啡肽(1~26)和内啡肽(1~27)等十几种不同产物^[17]. 另外, 血管紧张素是一种结构非常简单但对心血管活性却十分重要的多肽, 血管紧张素的前体分子是由447个氨基酸残基组成, 去掉信号肽后形成的血管紧张素原会在肾素的作用下形成血管紧张素I, 而血管紧张素I又能够在血管紧张素转换酶(angiotensin converting enzyme, ACE)的作用下进一步形成血管紧张素II. 近年来的一些研究发现, 血管紧张素I和II也能够通过一种内肽酶的作用从而形成血管紧张素1~7(Asp-Arg-Val-Tyr-Ile-His-Pro)^[18]. 由此看来, 从前体蛋白到成熟的功能蛋白之间的酶解过程具有十分重要的意义.

5 蛋白翻译后在生理状态下的修饰

蛋白翻译后修饰是指在mRNA被翻译成蛋白后, 对蛋白上的个别氨基酸残基进行共价修饰的过程. 蛋白翻译后修饰作用在生命体中具有极其重要的作用, 它使蛋白的结构更为复杂, 功能更加完善, 调节作用更为精细. 常见的蛋白翻译后修饰有泛素化、糖基化、磷酸化、脂基化、甲基化以及乙酰化等. 哺乳动物体内, 通常DNA密码仅能指导编码20种基本氨基酸的生物合成, 但是经过修饰后, 蛋白内的氨基酸

或氨基酸的衍生物至少能达到140种^[19]. 人类生命过程的复杂性不单是基因直接表达的结果, 而且还是因为蛋白翻译后修饰, 使得一个基因不只对应一个蛋白, 从而赋予了人类整个生命过程更多的复杂性.

6 自身免疫与蛋白翻译后修饰

抗体是介导体液免疫的重要效应分子, 它是B细胞在接受抗原刺激后增殖分化为浆细胞后所产生的糖蛋白, 在人体免疫过程中会产生多种抗体, 每一种抗体的产生代表着一种新的蛋白. B淋巴细胞是人体内一类重要的免疫细胞, 它与抗体的产生直接相关. B细胞从原始细胞到抗体分泌细胞要经过下面几个阶段: 干细胞、前B细胞、不成熟B细胞、成熟B细胞、浆细胞^[20]. B细胞从激活到增殖分化整个过程, 因刺激因素以及抗原的不同而产生不同的抗体, 如IgG, IgA, IgM, IgE, IgD等, 每一种又分为不同的亚型. 因免疫因素的复杂以及免疫过程的多边性, 可以产生数量庞大的抗体分子, 理论上在一个人体内可能存在 10^{11} 种不同的抗体.

7 当前蛋白质鉴定相关技术应用

近年来, 蛋白质组学方法的迅速发展主要得益于蛋白质高通量分离的高分辨二维电泳技术, 以及可对蛋白质成分进行灵敏鉴定的质谱学方法, 同时还依赖于基因组DNA测序推测出的可能存在的所有蛋白质氨基酸序列数据库^[21,22]. 自20世纪90年代开始, 蛋白质组学方面的研究取得了飞速的发展, 伴随着高效肽电离技术和质谱技术(mass spectrometry, MS)的进一步发展, 已经实现了蛋白质的高效高灵敏度鉴定^[23]. 同时, 生物信息技术也在2D电泳和MS分析中得到了大量的应用, 包括蛋白质鉴定(将MS数据与数据库进行比对)、运用反向遗传学分析相关基因的特征及分析蛋白翻译后修饰过程^[24]等方面.

当前, 蛋白质组学研究的核心内容之一是对翻译后修饰蛋白进行系统分析, 包括利用相关的分析技术对修饰蛋白及其位点进行鉴定、对相应的修饰结构进行确认以及对不同生理状态下翻译后修饰的变化情况进行定量分析等. 但是, 由于目前蛋白质数据库中的蛋白信息大多是来源于已知功能特点的蛋白, 而大多数不确定功能蛋白的具体序列数据较少, 这也正是导致目前关于蛋白计数研究相对困难的一个原因^[25].

人类基因组计划已经开始由结构基因组学转向功能基因组学。关于基因组的表达及其调控、基因组的多样性、模式生物体基因组研究等将是未来研究领域的核心问题。随着模式生物体中越来越多的蛋白构建编码单位逐渐被识别,基因和蛋白同源关系的搜寻和家族的分类变得更加简便。同时,生物信息学的算法和程序也在不断改善使得研究者们不仅能够从一级结构,也能从二级结构上来发现同源关系。但是,利用计算机模拟所获得的大量的蛋白数目关系的理论数据,还需要经过进一步的实验来验证和修正。

8 总结

随着蛋白质组学高效分离方法以及质谱鉴定技术的发展,人体中蛋白数量的问题已经不再局限于一个基础生命科学问题,而是可以通过分析蛋白种类及数量的变化,从而分析出机体在功能层面的改变。各种新型富集材料、高效分离方法、新型质谱仪和生物信息学计算方法的日新月异,使得大规模、高通量蛋白分析成为可能,尤其是在蛋白翻译后修饰的规模化分析、位点的准确定位和低灵敏度检测等方面。相信在不久的将来,有可能越来越深刻地认识到蛋白质多样性的生物学意义。

参考文献

- 1 Shi M G. The study of the method for predicting protein-protein interactions (in Chinese). Doctor Dissertation. Hefei: University of Science and Technology of China, 2009 [史明光. 蛋白质相互作用预测方法的研究. 博士学位论文. 合肥: 中国科学技术大学, 2009]
- 2 Luo L. Study on models of protein structure prediction (in Chinese). Doctor Dissertation. Wuhan: Huazhong University of Science and Technology, 2010 [罗亮. 蛋白质结构预测模型研究. 博士学位论文. 武汉: 华中科技大学, 2010]
- 3 Hu J, Guo Y T, Li Y M. Advances in post-translational modification of protein (in Chinese). Chin Sci Bull, 2005, 50: 1061–1072 [胡箭, 郭燕婷, 李艳梅. 蛋白质翻译后修饰研究进展. 科学通报, 2005, 50: 1061–1072]
- 4 Chen Z, Huang W, Fu G, et al. Present situation and prospect of Human Genome Project (in Chinese). Chin J Nat, 2000, 22: 125–133 [陈竺, 黄薇, 傅刚, 等. 人类基因组计划现状与展望. 自然杂志, 2000, 22: 125–133]
- 5 Feng L Q, Li H. Study on the statistical law of eukaryotic biological protein database (in Chinese). Chin J Med Phys, 2014, (6): 5318–5321 [冯立芹, 李宏. 真核模式生物蛋白质数据库统计规律研究. 中国医学物理学杂志, 2014, (6): 5318–5321]
- 6 Tan Z. Association of aldehyde dehydrogenase 2 activity levels and genotype with coronary heart disease (in Chinese). Doctor Dissertation. Shijiazhuang: Hebei Medical University, 2013 [谭峥. 乙醛脱氢酶 2 活性水平及基因型与冠心病的相关性研究. 博士学位论文. 石家庄: 河北医科大学, 2013]
- 7 Wang R, He Z Y, Chen L M. Advances in alkaline dehydrogenase 2 genotyping methods for key elements of alcohol metabolism (in Chinese). J Guangdong Coll Pharm, 2013, 29: 682–685 [王琰, 何震宇, 陈亮明. 酒精代谢关键酶乙醛脱氢酶 2 基因分型方法研究进展. 广东药学院学报, 2013, 29: 682–685]
- 8 Deng W G. Development of genetic polymorphism of species (in Chinese). Biol Bull, 1994, (1): 7–9 [邓务国. 物种遗传多态性研究方法的进展. 生物学通报, 1994, (1): 7–9]
- 9 Chen Y Q. Molecular biology basis of two hemoglobinopathies (in Chinese). Biol Bull, 2003, 38: 21–23 [陈玉芹. 两种血红蛋白病的分子生物学基础的比较. 生物学通报, 2003, 38: 21–23]
- 10 Goncalves V, Matos P, Jordan P. Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. Hum Mol Genet, 2009, 18: 3696–3707
- 11 Anraku Y, Mizutani R, Satow Y. Protein splicing: Its discovery and structural insight into novel chemical mechanisms. Iubmb Life, 2005, 57: 563–574
- 12 Noren C J, Wang J M, Perler F B. Dissecting the chemistry of protein splicing and its applications. Angew Chem Int Edit, 2000, 39: 450–466
- 13 Choi J J, Nam K H, Min B, et al. Protein *trans*-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*. J Mol Biol, 2006, 356: 1093–1106
- 14 Perler F B. Protein splicing mechanisms and applications. Iubmb Life, 2005, 57: 469–476
- 15 Liu X Q. Protein-splicing intein: Genetic mobility, origin, and evolution. Annu Rev Genet, 2000, 34: 61–76
- 16 Perler F B. InBase: The intein database. Nucleic Acids Res, 2002, 30: 383–384
- 17 Tang J. Functional diversity of bioactive proteins and intramolecular regulation (in Chinese). Chin J Med Sci, 2000, 80: 727–728 [汤健. 生物活性蛋白的功能多样性与分子内调控. 中华医学杂志, 2000, 80: 727–728]

- 18 Tang J, Tang C S, Chen G H, et al. Intramolecular regulation of cardiovascular active polypeptides (in Chinese). *Sci Technol Herald*, 1996, 14: 3–7 [汤健, 唐朝枢, 陈光慧, 等. 心血管活性多肽的分子内调控. *科技导报*, 1996, 14: 3–7]
- 19 Zhao X W. The research of protein post-translational modifications and protein-protein interactions prediction methods (in Chinese). Doctor Dissertation. Shenyang: Northeast Normal University, 2013 [赵晓威. 蛋白质翻译后修饰及其相互作用预测方法研究. 博士学位论文. 沈阳: 东北师范大学, 2013]
- 20 Feng J S, Chi H G, Zheng X B. Study on human regulatory B cell phenotype and related functional molecules (in Chinese). *J Cell Mol Immunol*, 2013, 29: 320–323 [冯锦山, 迟宏罡, 郑学宝. 人体调节性 B 细胞表型及相关功能分子研究. *细胞与分子免疫学杂志*, 2013, 29: 320–323]
- 21 Abbott A. And now for the proteome ... *Nature*, 2001, 409: 747–747
- 22 Pierce J D, Fakhari M, Works K V, et al. Understanding proteomics. *Nurs Health Sci*, 2007, 9: 52–58
- 23 Yates J R, Speicher S, Griffin P R, et al. Peptide mass maps: A highly informative approach to protein identification. *Anal Biochem*, 1993, 214: 397–408
- 24 Anderson L B, Maderia M, Ouellette A J A, et al. Posttranslational modifications in the CP43 subunit of photosystem II. *Proc Natl Acad Sci USA*, 2002, 99: 14676–14681
- 25 Liu H H, Tian Y, Lu X Y, et al. Macro proteomics: A new strategy for studying microbial communities (in Chinese). *J Microbiol*, 2010, 30: 88–92 [刘虎虎, 田云, 卢向阳, 等. 宏蛋白质组学: 研究微生物群落的一种新策略. *微生物学杂志*, 2010, 30: 88–92]

苏丹

教授, 四川大学博士生导师, 中国生物物理学会理事, 国家“千人计划”青年科学家, 四川省“千人计划”领军人才, 四川省侨联特聘专家, 成都市人才计划专家. 长期从事结构生物学领域研究, 综合利用 X 射线蛋白晶体衍射、核磁共振及电子显微镜技术, 以表观遗传学相关蛋白及蛋白复合体的空间结构为研究对象, 结合细胞生物学功能实验, 深入开展了表观遗传学分子调控机制的基础研究. 至今已在包括 *Nature*, *Cell*, *J Biol Chem*, *J Mol Biol*, *FEBS J*, *J Virol* 等国际期刊上发表研究论文多篇, 担任 *Acta Crystallogr D Struct Biol*, *BMC Struct Biol*, *Protein Cell* 等多个杂志的审稿人.

How many proteins are there in humans

WANG MeiLin, ZHANG RunDong & SU Dan*

State Key Laboratory of Biotherapy, West Hospital, Sichuan University, Chengdu 610041, China

* Corresponding author, E-mail: sudan@scu.edu.cn

On this stage, there is no consensus on the *Science* question: “How many proteins are there in humans”. But with the improved efficiency on proteomics separation and the development of mass spectrometry technique, the question about protein type and quantity in the human body can be answered in part. Via analyzing the changes in the type and quantity of protein, the alternations in functional level of body can be disclosed. For better comprehensive understanding, the study of number of proteins in the human body must be solved from the synthesis, processing, post-translational modification, and the main source of protein in the human body. This article summarized the relationship between the number of proteins and the number of genes in the human body based on current researches, combining the evolution of eukaryotic model organisms, gene coding, protein modification and current proteomics techniques development, it comprehensively analyzed the problem about how many protein species may exist in human body.

Proteins are the direct performer of all kinds of life activities and only consist of 20 kinds of amino acids. However, by different length and arrangement and combination, the theoretically limited amino acids varieties indeed can form vast number of protein bank. Since proteins are derived from genomic DNA, researching the numbers of coding frame could be treated as the basis to calculate protein species. Gene is the blueprint of synthesizing protein. Maintaining accuracy and stability of gene translation is vital for the inheritance of life species, thus even tiny variation of gene sequence may induce proteins to be different structurally and functionally formed by transcription and translation. In the meantime, by comparing and analyzing the protein splicing and gene's alternative splicing, we are aware that in protein splicing, new peptide bonds come into being: during splicing, precursor protein's intron are excised and the exon on both sides bond together. Almost Introns consist of 134–650 aa, and gene's alternative splicing means starting from a mRNA precursor (pre-mRNA), fusing different exons and produce multiple types of mature mRNA splicing isotypes by different splicing patterns (select different splicing sites and combine). Understanding this process can lead to a grasp of difficulty about counting proteins. In addition, from proteins to the maturation of proteins, various modifications are involved, and other processes like the bioactive peptide *in vivo* enzymolysis. The complexity of human life process is not the result of the direct expression of gene, but also the modifications of proteins' post-translation, which make one gene correspond to more than one protein, so endow more complexity to human's whole life process. As for antibody-one big source in human body, multiple types of antibodies are generated during immunogenesis. Every kind of antibody is a novel protein, and in the process of antibody-secreting immune cells activation, proliferation and differentiation, different antibodies are produced because of different stimulus factors and antigens, and in theory one person may have more than 1011 types of antibodies. Apparently, it's quite challenging to analyze the questions about how many proteins are inside human body from either genomics or proteomics directions, which lead us failed to figure out how many proteins are in human body. With the advance of new enrichment materials, high-performance separation method, novel mass spectroscopy and biointelligence algorithm, large-scale and high throughput protein analysis must be applicable, especially in aspects like scale analysis of post-translation modifications of proteins, precise siting of target and low-sensitive inspections. In the near future, we are probably able to have a deeper understanding on the biological meaning of protein's diversity, which will progress the problem about how many proteins are in human body.

protein, gene coding, protein splicing, post-translational modification, mass-spectrometric technique (MS)

doi: 10.1360/N972016-01016