

·学科进展与展望·

中国蛋白质组计划

贺福初*

(军事医学科学院放射医学研究所,北京 100850)

[摘要] “蛋白质组”指一种细胞、组织或完整的生物体所拥有的全套蛋白质。基因组是生命体遗传信息的载体,蛋白质组是生命活动的执行体;蛋白质组学的特点是采用高分辨率的蛋白质分离手段,结合高效率的蛋白质鉴定技术,全景式地研究在各种特定情况下的蛋白质谱;每种疾病平均与10个左右基因相关,而每种基因可能又与3—10种蛋白质相关,如果以人类主要的100—150种疾病进行计算,则应该有3 000—15 000种蛋白质具有成为药靶的可能;我国的“疾病基因组学”研究已取得明显成就,在重大疾病的功能蛋白质组学研究方面也起步良好,国家层面上对蛋白质组学的部署不容迟缓。

[关键词] 中国,蛋白质组,计划

1 国家层面上对蛋白质组学的部署不容迟缓

“蛋白质组”指一种细胞、组织或完整的生物体所拥有的全套蛋白质。

人类基因组计划被誉为20世纪的3大科技工程之一。其划时代的研究成果——人类基因组序列草图的完成宣告了一个新的纪元——“后基因组时代”的到来。其中,功能基因组学成为研究的重心,蛋白质组学则是其“中流砥柱”之一。正因如此,《自然》和《科学》杂志在公布人类基因组序列草图的同时,分别发表了述评与展望,将蛋白质组学的地位提到前所未有的高度,认为是功能基因组学这一前沿研究的战略制高点,蛋白质组学将成为新世纪最大战略资源——人类基因争夺战的重要“战场”。

新世纪来临,我国面临巨大的机遇也遭遇极大的挑战。

挑战之一关系到我国人民的健康保障:能否不断地对影响我国人民健康的重大疾病提出有效的防治措施?能否为社会提供来源充分、价格合理、疗效显著的防治药品?

目前的形势十分严峻。由于历史及经济的原

因,我国制药业长期重生产而轻创新,一度曾普遍以仿制国外新药为主要“研发”战略,致使我国的制药业极为缺乏独立知识产权。当我国的经济与国际接轨以后,拥有强大研发力量及大量知识产权的跨国制药企业涌入我国,如果我们仍然忽视创新和知识产权,不能在生物科技的发展上处于有竞争能力的地位,我国的民族制药工业以至整个医疗保健事业将面临严重危机。这对于一个拥有13亿人口而又处于发展中的国家来说决不能掉以轻心。

功能基因组学,特别是作为其重要组成部分的蛋白质组学的发展给我们提供了一个难得的机遇和机会。蛋白质组学的特点是采用高分辨率的蛋白质分离手段,结合高效率的蛋白质鉴定技术,全景式地研究在各种特定情况下的蛋白质谱。由于蛋白质是生物细胞赖以生存的各种代谢和调控途径的主要执行者,因此蛋白质不仅是多种致病因子对机体作用最重要的靶分子,并且也成为大多数药物的靶标乃至直接的药物。药靶,来源于对生命活动的生理病理过程的研究;药靶,又形成制药业的发展源头。蛋白质组学正是近年来新发展起来的强有力的发现药靶的技术平台,作为一个新的学科发展领域,它对所有及时进入的国家都将提供巨大的机会。机不可

* 中国科学院院士。

本文为作者在第七届海内外生命科学论坛学术研讨会的大会报告。

本文于2002年7月8日收到。

失,时不我待。

一项科学统计表明,在20世纪90年代中期,全世界制药业用于找寻新药的药靶共约483个,它们主要是蛋白质(受体占45%,酶占28%等);而当时全世界正在使用的药物总数约是2000种,其中的85%都是针对上述483种药靶。这483种药靶分子构成了全世界药厂的最重要的发展源泉。从功能基因组学的角度,人们认为每种疾病平均与10个左右基因相关,而每种基因又可能与3—10种蛋白质相关,如果以人类主要的100—150种疾病进行计算,则应该有3000—15000种的蛋白质具有成为药靶的可能。也就是说还可能几千到上万种的新药靶将被发现,这将是功能基因组学尤其是蛋白质组学研究有可能带来的一笔巨大的科学、经济财富。这也是为什么蛋白质组学作为发现药靶的主要技术平台,在20世纪90年代末期以来越来越受国际巨型跨国制药集团垂青的重要原因所在。

提高人口健康水平的源头是阐明重要生理病理过程的发生机制及认清疾病的诱发启动和发展转归的过程。我国对肝炎、肝癌、心血管疾病等重大疾病及神经内分泌、造血、神经损伤等重要生理病理过程均曾进行过较系统深入的病理学或基因组学研究。而基因组学虽然可在基因水平对疾病的遗传因素提供了解,但仅此是往往不够的。基因组是生命体遗传信息的载体,而蛋白质组是生命活动的执行体。生命体的统一性在于基因组,生命体的复杂性与功能性却在于蛋白质组。人体中各组织、细胞间基因组完全相同,而其形态上、功能上的重大差别关键在于蛋白质组的不同。此外,细胞中的基因和蛋白质并不存在绝对的线性关系。一个开放阅读框架(ORF)并不一定预示存在一个相对应的功能性蛋白;mRNA水平并不一定与蛋白质的表达水平完全对应;翻译后修饰及同工蛋白质(isoforms)等现象在基因水平无从表现;蛋白质与蛋白质的相互作用更是难以在基因水平得以预知。所有这些都需要通过蛋白质组学的研究来加以阐明。

因此,我们建议通过对肝脏、造血、神经内分泌等重要生理病理过程的“组成性蛋白质组学”分析及肝炎、心脑血管病、肿瘤等重大疾病的“比较蛋白质组学”研究,在蛋白质组水平上为严重影响我国人群健康的重大疾病的预警、诊断、治疗与预防提供新的思路和新的策略。其中,心脑血管疾病是我国居第一位的健康杀手;原发性肝癌是世界上最常见的原发性肿瘤之一,在因此而死亡的病例中,我国占有一

半;在我国即将到来的老龄化社会中,老年痴呆症、帕金森症等神经系统疾患的危害将日益突出。因此开展对上述疾病的蛋白质组学研究势在必行。

到目前为止,生物技术药(指用重组技术、杂交瘤和转基因方法等生产的蛋白质、多肽、RNA、DNA等的治疗剂)的56%是针对所谓的“稀有疾病”。这主要是因为这些“稀有疾病”往往由单基因的异常所致,而此类异常可通过传统的单一基因或蛋白的分析予以揭示。基因组学研究使得人们可以从全基因组的水平,以相互关联、网络化、多基因的全局性观点来研究疾病。随着基因组学、蛋白质组学技术的发展,涉及到多基因、多蛋白甚至蛋白质作用网络的常见病、多发病,其发生发展机制的认识必将取得突破。我国蛋白质组学技术平台、研究体系的建立,其意义与作用将决不局限于上述所建议的少数几类重大疾病的研究,而会是一个通用性的研究平台,将适用于对更多种严重影响我国人民健康的重大、常见疾病的研究。

我国的基因组学研究已取得重大的成就,通过人类基因组项目及多种资源微生物的全基因组测序,我们不仅建立了国际先进水平的规模化工作能力,同时在人体重要生理病理体系的规模化表达谱的建立、中国人群优势致病菌株的全基因组序列分析等方面都取得了引人瞩目的成绩。其中,我国作为6个参与国之一,而且作为唯一的发展中国家的代表完成的1%人类基因组序列测定,赢得了国内外的广泛称赞;在国际上首次完成的下丘脑-垂体轴-肾上腺轴和人胎肝、CD34+造血干/祖细胞和海马大规模cDNA测序,使我国发现了一批国际上从未涉足的人类新基因,其新基因及基因表达谱已相继在国际核心学术刊物上发表并获国际学术界承认。上述研究使我国在国际基因组学领域占有了重要的一席之地,同时为开展上述体系的蛋白质组表达谱、功能连锁群的研究,并实现人类组织或细胞中从未有过的转录组与蛋白质组的对接与集成奠定了重要基础。

基于此,我们建议以我国基因组学研究的雄厚基础为出发点,从蛋白质组水平上继承并发扬其已有的成果,使其由单纯的核酸序列信息或理论推测转化成我国生物技术能进一步开发的系列实物性蛋白质;并通过其蛋白质组功能连锁群的建立,进一步提升研究水平,扩大我国在此类系统研究上的特色与优势;同时,基于上述体系的独特性及国际上研究的空白,可望通过一批人类新型蛋白质的确认与发

掘,为人类基因组草图完成时未曾确认的大批推测性基因从编码产物的层次提供确凿的规模化识别、鉴定与确认,从而为 HGP 战略目标的实现作出我国学者的历史性贡献。

我国在蛋白质化学与蛋白质科学领域曾取得举世瞩目的成就,近年在蛋白质组学这一新兴领域已有了良好的基础,一批研究结果已在国际蛋白质组学核心期刊发表,技术平台的部分指标达到国际先进水平。但是,现有的规模与层次难以提供我国医药卫生事业与生物技术产业迅猛发展所急需的、强有力的蛋白质组学学术与技术支撑,难以适应我国基因组学等现代生命科学前沿领域对蛋白质组学的广泛需求,难以应对国际在蛋白质组学这一战略“高地”的激烈竞争。因此,国家层面上对蛋白质组学的部署及大力支持不容迟缓。

2 国内外研究现状和发展趋势

2.1 国外研究现状

基因,尤其是基因组研究形成了 20 世纪生命科学一道亮丽的风景线。截至 2001 年底,已公开发表至少 84 个生物体的基因组全序列(含染色体),其中包括古细菌 12 个、细菌 57 个、真核生物 15 个,此外有 200 余个原核生物和 150 余个真核生物或其染色体的基因组正在测序。基因组研究所产生的“海量”数据和大量新型技术以前所未有的深度与广度极大地推动了生物医学多学科的飞速发展。

但是,随着大量生物体全基因组序列的获得,特别是人类基因组序列草图的完成,基因组研究的战略重点不可避免地结构基因组学转向功能基因组学,而蛋白质组学正是作为功能基因组学的重要支柱,在 20 世纪 90 年代中期应运而生。近年,《自然》和《科学》以及其他一些重要杂志,接连刊登有关蛋白质组学的评论文章或原始论著,明确表明了蛋白质组学已经成为新世纪生命科学研究的前沿。

蛋白质组学的第一篇原始论著 1995 年发表于国际上并不著名的《电泳》杂志(1995,16:1 090)。如果说蛋白质组学刚诞生时没有立即得到国际生命科学界主流的高度重视,那么近 3 年已发生巨大的变化:美国国立卫生研究院(NIH)所属的国立癌症研究所(NCI)投入了大量经费支持蛋白质组研究,其中 1 000 万美元用于在密执安医学院建立一个有关肺、直肠、乳腺和卵巢等肿瘤的蛋白质组数据库;此外,国立癌症研究所和美国食品与药物管理局(FDA)联合,投入数百万美元,资助建立一个有关癌症不同发

病阶段和治疗阶段的蛋白质组数据库。美国能源部不久前也启动了蛋白质组项目,旨在研究涉及环境和能源的微生物和低等生物的蛋白质组。

欧共体目前正资助酵母蛋白质组研究。英国生物技术和生物科学研究委员会最近也资助了 3 个研究中心,对一些已完成或即将完成全基因组测序的生物进行蛋白质组研究。在法国,新成立了 5 个区域性遗传基地,它们将得到为期 3 年的资助,每年约为 500 万美元,这些经费将平均用于基因组、转录组和蛋白质组研究。德国的联邦研究部提供 700 多万美元,在德国东部罗斯托克建立了一个蛋白质组学中心。1997 年澳大利亚政府即着手建立第一个全国性的蛋白质组研究网 APAF。APAF 为该国的有关实验室提供一流的仪器设备,并把它们整合在一起,进行大规模的蛋白质组研究。日本的科学与技术委员会也已先期由政府出资 300 万美元开展蛋白质组研究,近期更是以 1 000 万美元启动水稻的蛋白质组研究。韩国即将以上亿美元的资金启动肝脏的蛋白质组研究。台湾已举资 2 亿美元开展衰老的蛋白质组研究。

由于蛋白质组学研究比基因组学研究更可能接近实用,具有巨大的市场前景,企业与制药公司纷纷斥巨资开展蛋白质组研究。如独立完成人类基因组测序的赛莱拉公司已宣布投资上亿美元于此领域;又如日内瓦蛋白质组公司与布鲁克质谱仪制造公司联合成立了国际上最大的蛋白质组研究中心。由此可见,蛋白质组学虽然问世不到 10 年,但鉴于其战略的重要性的技术的先进性,已成为西方各主要发达国家、各跨国制药集团竞相投入的“热点”与“焦点”。

蛋白质组学的前沿大致分为 3 大方向:(1)针对已完成基因组或转录组研究的生物体或组织/细胞,建立其蛋白质组或亚蛋白质组(或蛋白质表达谱)及其蛋白质组连锁群,即组成性蛋白质组学研究;(2)以重要生命过程或人类重大疾病为对象,进行重要生理/病理体系或过程的比较蛋白质组学研究;(3)蛋白质组学支撑技术平台和生物信息学的研究。

国外大部分蛋白质组表达谱的研究论文发表于 2000 年下半年以后,且大多建立在已完成基因表达谱的基础上,表明目前在基因组或转录组基础上开展蛋白质组表达谱的研究是一个新的方向。人类重大疾病的蛋白质组研究通常采用比较蛋白质组分析方法。近年来,蛋白质组学技术在研究细胞的增殖、分化、异常转化、肿瘤形成等方面进行了有力的探

索,涉及到白血病、乳癌、结肠癌、膀胱癌、前列腺癌、肺癌、肾癌和成神经细胞瘤等,鉴定了一批肿瘤相关蛋白,为肿瘤的早期诊断、药靶的发现、疗效判断和预后提供了重要依据。高通量、高灵敏度和规模化的双向凝胶电泳-质谱,是目前最流行和较可靠的技术平台;酵母双杂交技术已被用于研究蛋白质连锁群和蛋白质功能网络系统,且分别发表于《自然》与《科学》。生物信息学方法在蛋白质组学研究领域亦得到有效的利用,其中突出的代表是 Eisenberg 等联合采用系统发育分布图法、Rosetta stone 法和基因邻居法,成功地建立了酵母 SIR(沉默信息调节子)作用网络和酵母朊病毒(prion)功能连锁网络。

2.2 存在的问题和发展趋势

方法学上,二维凝胶电泳-质谱虽然仍是目前最流行和较可靠的技术平台,但其通量、灵敏度和规模化均有待进一步加强。二维凝胶电泳有分离容量的先天限制,染色转移等环节操作困难费时,低丰度蛋白难以辨别,和质谱技术的联用已成为瓶颈。因此国际上开始重视研究以色谱/电泳-质谱为主的技术平台。另一方面,酵母双杂交技术虽已被用于研究蛋白质连锁群和蛋白质功能网络系统,但仍缺乏快速、高效的手段获取复杂蛋白质相互作用的多维信息。蛋白质组的生物信息学研究,虽然已有成功的先例,但其应用范围与准确率仍需提高,所面临的更大挑战是如何进行信息综合,准确分析蛋白质的相互作用,界定相互作用连锁群。

学术上,在基因组、转录组基础上的蛋白质组全谱研究微生物已有成功的报道,但是在高等生物尤其是哺乳动物中未见报道,人类组织或细胞的蛋白质组全谱研究则基本未涉及。而由于物种演化中进化上的差别,人类基因组、转录组、蛋白质组的全景式比较,对于不同层次人类基因表达调控规律的认识必不可少;此外,人类基因组序列草图虽已公布,但是所估计的3.5万左右基因中一半左右纯属理论推测,需要从蛋白质组水平予以检验与确认,因此开展人类组织或细胞的蛋白质组表达谱的分析势在必行。

受基因调控的细胞内各种信号转导途径之间是相互交错和彼此关联的。虽然近年来人们对转导途径以及相互关系的认识取得了不少进展,但是针对任一生物体或组织/细胞开展全方位的蛋白质组相互作用网络的分析鲜有报道。而此类相互作用网络的揭示对于深刻认识重要生理、病理过程的机制不可缺少。

2.3 国内研究现状

在1995年国际上发表第一篇蛋白质组学的研究论文后不久,国家自然科学基金委员会即酝酿并于1997年设立了重大项目“蛋白质组学技术体系的建立”。在此前后,中国科学院生物化学研究所、军事医学科学院与湖南师范大学迅速启动了蛋白质组研究,建立并分别组合了二维电泳蛋白质组分离技术、2D-PAGE 图像分析技术和蛋白质鉴定的质谱技术;先后举办了多次全国性的蛋白质组学术研讨会,并在国际上较早提出了功能蛋白质组学的研究战略。

经过几年努力,中国科学院上海生命科学研究院、军事医学科学院与复旦大学相继成立了专门的蛋白质组学研究中心。整体上,我国蛋白质组学技术平台的建设有了飞跃的发展,若干研究单位重点建立了技术平台,并在方法学的跟踪与创新上作了不少工作,部分技术平台基本达到国际水平。我国科学家已经在蛋白质组分析技术与方法、在重大疾病如肝癌、维甲酸诱导白血病细胞凋亡启动模型及维甲酸定向诱导胚胎干细胞向神经系统分化的模型等的比较蛋白质组研究以及一些重要生理和病理体系的蛋白质组成分研究方面获得了不少成绩,并在国际蛋白质组学核心刊物上发表了一系列高水平的论文。对不断涌现的新技术,如衍生出的多种分离和鉴定模式,我国已经进行了很好的跟踪和发展,在某些方面孕育着新的突破。

总之,我国的“疾病基因组学”研究已取得明显的成就,在神经系统遗传病致病基因,肝癌、心血管疾病、白血病等相关基因,造血干/祖细胞、下丘脑-垂体-肾上腺轴系统、海马体、胎肝、心血管和神经系统等组织或细胞的基因表达谱(即转录组)方面均做出了与国际前沿水平相当的工作,且有我国的特色与优势。所取得的丰富数据可直接成为我国开展对应蛋白质组学研究的基础与出发点。我国在重大疾病的蛋白质组学研究方面也取得了良好的起步,已进行肝癌细胞系及正常肝细胞蛋白质组的比较分析研究,发现了两者间不同的蛋白表达群;此外,还进行了我国自行建立的肝癌高/低转移细胞系、肺癌高/低转移细胞系、原位食管癌/转移食管癌间的比较蛋白质组研究,初步发现了一批与肿瘤转移相关的蛋白质群。同时,心肌细胞与应激损伤的心肌细胞的比较蛋白质组研究也已有一定的基础。

上述研究,一方面证明我国已有的蛋白质组学技术平台已能支撑一定规模的研究,一方面亦为我

国在国际蛋白质组学研究领域争得了一席之地,同时为未来的发展奠定了良好的基础。国家与企业如能加大投入,前景会一片光明。

THE CHINESE PROJECT OF HUMAN PROTEOMICS

He Fuchu

(Beijing Institute of Radiation Medicine, Beijing 100850)

Abstract The goals of Chinese Project of Human Proteomics are as follows: To build a powerful platform in proteomic research with high throughput, high accuracy and high stability; based on their gene expression profiles, to find 1-2 interaction networks of the proteins correlated with some important physiological and pathological processes; to identify more than 500 unknown human proteins and to probe their interaction and regulation characteristics. Discover several protein groups in correlation with the disease in high risk, such as hepatitis, liver cancer and heart disease, and screen some proteins as the bio-markers for their early stage diagnosis and the new targets for new drug development; to develop a systematic bioinformatics technique with the protein database, modules and algorithms for functional analysis of protein-protein interaction.

The final objective of this program is to give powerful academic and technical supports for proteomics research in our country, to make the progress of biomedical research in high speed and to reach the high level in the world.

Key words proteomics, human, Chinese

·资料·信息·

国家基础科学人才培养基金骨干教师培训计划正式启动

国家自然科学基金委员会组织的国家基础科学人才培养基金骨干教师培训计划动物生物学教师培训班于2002年7月29日在陕西师范大学结束,国家自然科学基金委员会副主任王乃彦院士和全国高等学校生物学教学指导委员会负责人王喜中教授、乔守怡教授以及来自40余所院校的60余位青年教师参加了培训结业典礼。至此,国家基础科学人才培养基金骨干教师培训计划正式启动。

“九五”期间,国家基础科学人才培养基金拨出专款2000万元用于“国家基础科学人才培养基地名牌课程”的建设。在基地广大教师的努力下,“名牌课程”建设取得了一大批教学改革研究成果,为了发挥“名牌课程”的示范和辐射作用,提高基地青年教师以及西部院校和地方院校基础学科青年教师的教学水平,国家自然科学基金委员会决定组织实施“国家基础科学人才培养基金骨干教师培训计划”。骨干教师培训计划以“名牌课程”为依托,以其

创建人为主讲教师,适当吸收国内外名师参加,组成示范教学师资队伍;以名师授课、集中研讨和交流以及必要的现场考察相结合的方式进行;培训工作向西部院校和地方院校倾斜,除基地青年骨干教师参加外,吸收西部院校和地方院校非基地青年骨干教师参加。经与教育部有关部门和相关学科全国高等学校教学指导委员会负责人商议,骨干教师培训计划2002年实施工作共安排4个学科7门课程的培训:动物生物学、植物生物学、量子力学、数学物理方法、物理化学、结构化学以及地学多媒体教学。2003年的培训工作将在认真总结今年培训工作经验的基础上在各个学科全面展开。

为配合国家西部开发战略,国家自然科学基金委员会组织的骨干教师培训计划还将开展多种形式的培训工作,如资助一批西部院校的青年骨干教师到东部名校进修等。

(计划局 谢焕瑛 供稿)