



评述

中国蛋白质组学研究进展

——以人类肝脏蛋白质组计划和蛋白质组学技术发展为主题

李衍常^①, 李宁^①, 徐忠伟^①, 翟琳辉^①, 樊锋旭^{①②}, 常乘^①, 张成普^①, 郑俊杰^①,
徐平^{①*}, 贺福初^{①*}

① 中国人民解放军军事医学科学院放射与辐射医学研究所, 蛋白质组学国家重点实验室, 国家蛋白质科学中心(北京),
北京蛋白质组研究中心, 蛋白质药物国家工程研究中心, 北京 102206;

② 安徽医科大学研究生学院, 合肥 230032

* 联系人, E-mail: hefc@nic.bmi.ac.cn; xupingghy@gmail.com

收稿日期: 2014-04-14; 接受日期: 2014-06-27

国家重点基础研究发展计划(批准号: 2011CB910600, 2013CB911201)和国家自然科学基金(批准号: 31070673, 31170780)资助项目

摘要 蛋白质组学是对细胞或生物体全部蛋白质的系统鉴定、定量并阐释其生物学功能的学科。自 21 世纪初期开始, 随着高精度、高灵敏度和快速扫描质谱仪的出现和快速发展以及微量蛋白质组样品高效分离技术的进步, 蛋白质组学获得了快速发展, 并在生理过程与病理机制研究等几乎所有生命科学研究领域得到了广泛的应用。过去 10 年, 中国蛋白质组学研究在政府的支持和广大蛋白质组学研究人员的努力下呈现出腾飞式的发展态势。本文综述了人类肝脏蛋白质组计划和 2010~2013 年中国蛋白质组学技术的发展。

关键词

蛋白质组学
人类肝脏蛋白质组计划
蛋白质相互作用
蛋白质翻译后修饰
生物信息学

自 2003 年“中国人类蛋白质组组织”(China human proteome organization, CNHUPO)成立至今, 中国的蛋白质组学研究经历了十年多的发展, 呈现出百家争鸣、百花齐放的局面。继中国科学家领衔“人类肝脏蛋白质组计划”(human liver proteome project, HLPP)之后, 2014 年 6 月, “中国人类蛋白质组计划”(china human proteome project, CNHPP)在京启动, 标志着中国科学家开始向全面、精确地阐释人体全器官蛋白质组这座高峰冲刺。本文在已有综述^[1-4]的基础上, 以人类肝脏蛋白质组计划和 2010~2013 年中国蛋白质组学技术的发展为主题进行综述。

1 人类肝脏蛋白质组计划的发展与成就

2003 年, 由贺福初及其科研团队^[5]提出的“人类肝脏蛋白质组国际计划”开始实施, 这是中国科学家首次领衔国际重大科研合作项目。近年来, 中国蛋白质组学研究团队密切合作、联合攻关, 在以下 3 个方面取得了阶段性的新进展(图 1): 系统性地注释肝脏蛋白质表达谱和蛋白质修饰谱(两谱); 最大纬度地绘制肝脏蛋白质的亚细胞定位与相互作用网络图(两图); 建设完成了大规模的肝脏蛋白质组学研究材料和数据库(肝脏蛋白质组组织样本库、抗体库和开源质谱数据库, 三库)。

引用格式: 李衍常, 李宁, 徐忠伟, 等. 中国蛋白质组学研究进展——以人类肝脏蛋白质组计划和蛋白质组学技术发展为主题. 中国科学: 生命科学, 2014, 44: 1099–1112

英文版见: Li N, Xu Z W, Zhai L H, et al. Rapid development of proteomics in China: from the perspective of the human liver proteome project and technology development. Sci China Life Sci, 2014, 57, in press, doi: 10.1007/s11427-014-4714-2

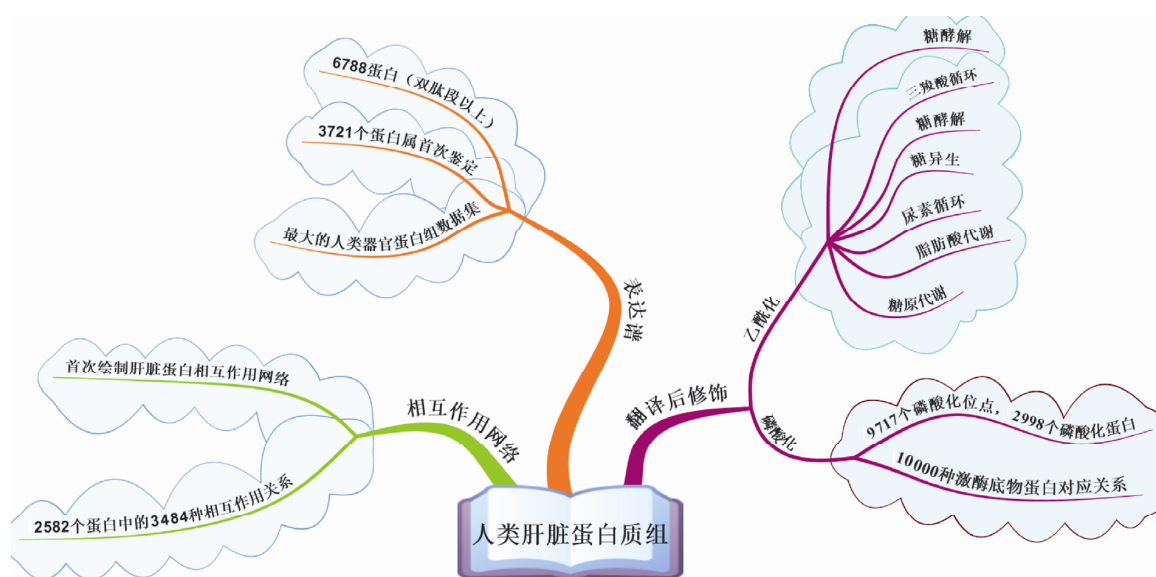


图1 中国人类肝脏蛋白质组计划(CNLPP)取得的进展

1.1 肝脏蛋白质组表达谱

人类个体间的遗传背景、生存环境乃至心理和精神状态等诸多方面存在显著差异,这决定了人类肝脏组织样本具有一定的异质性.为了绘制具有代表性的人类肝脏表达谱,中国蛋白质组学团队与国际同行合作,系统评价了肝脏组织样本个体差异对蛋白质组学研究结果的影响,建立了国际首份完整的人体组织器官蛋白质组学的样品制备标准化工作流程(standard operating procedures, SOPs),为人类肝脏蛋白质组国际计划的实施奠定了基础^[6].在此基础上,中国蛋白质组学研究团队对各种生理和病理状态下肝脏组织样本进行了系统的蛋白质组学研究,共鉴定到双肽段以上高可信的肝脏蛋白质 6788 个.其中 3721 个蛋白质在肝脏组织中被首次鉴定.这是迄今为止人类蛋白质组学研究计划中最大的单一组织脏器的蛋白质组数据集,引领并促进了国际蛋白质组合作计划的深入开展.

中国蛋白质组学研究团队进而对这些鉴定蛋白质的丰度信息进行了系统研究,发现这些蛋白质横跨 6 个数量级的丰度范围,其中 78% 的蛋白质(5294 个)位于中等或偏下信号强度区间,而首次鉴定的 3721 个蛋白中的 3069 个蛋白均属于低丰度蛋白.如在肝脏中低丰度表达的细胞色素 P450 家族的 4 个分子和 3 个离子通道相关蛋白在肝脏组织中被有效鉴定^[7,8].

1.2 肝脏蛋白质组修饰谱

肝脏蛋白质的翻译后修饰,如磷酸化、乙酰化等研究也被广泛开展.复旦大学管坤良和熊跃团队^[9]在肝脏蛋白质的乙酰化修饰谱方向开展了卓有成效的研究,拓展了人体生理和病理条件下代谢及其调控的研究领域.利用实验室自己研制的特异高效的乙酰化肽段富集抗体,实现了肝脏组织中大量乙酰化修饰肽段的富集和大规模鉴定.对这些鉴定的乙酰化肽段进行系统的生物信息学研究,发现几乎所有参与中间代谢的酶蛋白,如糖酵解、糖异生、三羧酸循环、尿素循环、脂肪酸和糖原合成等途径的蛋白质被乙酰化修饰.这些代谢酶蛋白质分子的乙酰化修饰程度与细胞内能量物质,如葡萄糖、氨基酸和脂肪酸的浓度关系密切.这些结果显示,各种酶蛋白的乙酰化修饰对细胞内的能量代谢起着重要的调节作用.不仅如此,赵国屏和管坤良、熊跃团队^[10]合作,以沙门氏菌(*Salmonella*)为研究对象,发现在不同碳源培养条件下,核心代谢酶分子的乙酰化修饰水平发生剧烈的波动,以适应细胞生长和能量代谢的需要.部分限速酶的乙酰化修饰还参与调控糖酵解/糖异生、柠檬酸循环/乙醛酸循环的代谢转化过程.这些研究不仅证实了基础和能量代谢酶类分子的乙酰化修饰在原核和真核生物中高度保守,同时也发现酶蛋白的乙酰化修饰参与了机体代谢过程的调控,奠定了

蛋白质乙酰化作为代谢调控者的基础. 这些研究结果被分别发表在 *Science* 同一期上. 为表彰管坤良与熊跃在蛋白质翻译后修饰蛋白质组学研究中的杰出贡献, 中国蛋白质组组织(CNHUPO)在第八届中国蛋白质组学大会(2013, 重庆)为其颁发了学术贡献奖.

蛋白质磷酸化是重要的功能信号传导分子, 参与并调控了众多生命过程. 中国科学院大连化学物理研究所邹汉法团队和华中科技大学薛宇团队^[11]合作, 开展了大规模的基于固定化金属亲和和层析法(immobilized metal affinity chromatography, IMAC)富集的肝脏蛋白质组磷酸化研究, 并发展了磷酸化的生物信息学技术, 鉴定了肝脏组织中 2998 蛋白质上的 9719 个磷酸化位点. 利用这个大规模的磷酸化蛋白质鉴定数据集, 发现人类肝脏可能包含 10000 多个节点的磷酸化激酶与特异性底物分子的磷酸化蛋白质分子网络, 为磷酸化信号途径及其网络的分子机制研究奠定了基础.

1.3 蛋白质相互作用和亚细胞定位网络构建

蛋白质相互作用(protein-protein interaction, PPI)信息的揭示不仅有助于了解蛋白质分子所处的细胞内的分子环境, 而且可进一步探索这些分子可能参与的代谢途径或信号通路, 从而为这些蛋白质的功能及其分子机制研究创造条件. 因此 PPI 网络的构建也是 HLPP 计划的重要研究内容和主要目标之一. 通过大规模蛋白质组学研究至今已经形成多种模式生物, 如多种病原细菌、酵母(*Saccharomyces cerevisiae*)、线虫(*Caenorhabditis elegans*)和果蝇(*Drosophila melanogaster*)等的蛋白质相互作用网络图. 这些相互作用的发现为整合生物学、疾病分子机制研究与药物筛选提供了有力的实验和数据支撑. 但目前 PPI 的建立主要面临的挑战是由于污染蛋白的存在带来的假阳性率偏高, 多数组学实验研究结果仅仅有少部分可被实验证实.

北京蛋白质组研究中心贺福初、杨晓明和王建团队^[12]选取了肝脏组织样本中的 5026 个蛋白分子进行了系统深入的蛋白质-蛋白质相互作用关系研究. 利用成熟的酵母双杂交技术平台和严格的假阳性排除技术, 该团队成功地鉴定了 2582 个蛋白质的 3484 种相互作用. 通过生物化学与分子生物学、细胞高内涵筛选系统验证发现相互作用的阳性率高达 72%. 深

入地分析这些相互作用数据, 该团队还发现了系列决定肝脏特征表型以及疾病状态的独有的蛋白质相互作用. 这是人类肝脏蛋白质相互作用网络(human liver protein interaction network, HLPN)国际合作项目中率先完成的首个脏器、器官蛋白相互作用大型数据集. 这个网络的系统构建对于理解人类肝脏蛋白质相互作用网络功能具有重要的价值.

人类肝脏蛋白质组计划(HLPP)的研究进展得益于蛋白质组学研究技术的进步和蛋白质组学的发展. HLPP 以构建“两谱、两图、三库”为总目标, 为蛋白质组学技术的研发提出了明确的科学问题和发展目标.

2 蛋白质组学技术的发展

蛋白质组学更高层次的研究离不开技术的发展, 而技术的发展又为蛋白质组学提供新的视角和思维方式. 在过去的 3 年中, 中国蛋白质组学研究团队在蛋白质组学样品制备、微量复杂样品的高效色谱分离、翻译后修饰蛋白质的富集、蛋白质组鉴定和定量分析以及生物信息学工具的发展等蛋白质组学研究的几乎所有方面都取得了显著发展.

2.1 蛋白质组样品的制备

蛋白质组学样品来源于细胞或者组织内部的全部蛋白质, 具有蛋白质种类复杂多样、丰度范围宽泛等特点, 这给蛋白质组的高覆盖鉴定和高精度定量造成了巨大的困难. 国内一些实验室对样品的制备环节进行了大量的探索, 开发了多种化学或生物介质材料, 在一定程度上消除了高丰度蛋白对高覆盖蛋白质组学鉴定的影响, 开拓了蛋白质组学研究的新领域.

(1) 低丰度蛋白质化学富集介质的开发. 以磁性微球为载体偶联多种亲和介质, 形成了多种高效的蛋白质分离和富集复合材料, 并已经成为化学介质方法中最有效的方法.

新型的磁性材料多以磁性无机粒子与有机高分子结合形成具有特殊结构的磁性微球, 在此过程中通过共聚及表面改性等方法赋予其表面不同的功能基团. 复旦大学杨芾原团队和张祥民团队^[13]根据金属离子亲和色谱原理, 将 Cu^{2+} 离子固定在高比表面积介孔二氧化硅微球载体上, 利用该微球的多孔

道特性和高密度的 Cu^{2+} 与肽段的结合力, 可以较高效地从微量的混合样品中富集肽段样品^[13]。邹汉法等人^[14,15]合成了具有 Yolk-Shell 结构的磁性介孔碳微球颗粒。利用该材料的中孔结构的强磁性响应, 可以从人血清中选择性地提取低丰度的内源性肽段。该团队利用这种微球从 20 μL 的人血清样品中高效地提取出了 3402 种不同的内源性多肽。这些内源性多肽通常丰度较低, 但生理活性显著。其高效鉴定为血清中的生物标志物的筛选创造了条件。中国科学院大连化学物理研究所张玉奎和张丽华团队与南开大学陈朗星团队^[16,17]合作, 通过点击化学的方法将 Fe_3O_4 纳米颗粒与亚氨基二乙酸共价连接, 制备成了强磁性和超高吸附容量新型材料。该材料可特异性吸附血红蛋白等血浆中高丰度蛋白。这些新技术的发展和新材料的开发有效地去除或者降低了血液样品中高丰度蛋白的干扰, 大大提高了低丰度蛋白的鉴定能力和序列覆盖度, 提高了从血液样品中发现生物标志物的能力, 因此具有良好的理论和实用价值。

(2) 富集低丰度蛋白质的生物介质的研发。针对特定的生物学问题, 设计开发具有针对性的生物介质元件, 可实现相关生物因子的高效特异的富集。北京蛋白质组研究中心秦钧团队^[18]发明了一种转录因子 DNA 结合序列串联阵列(catTFRE), 能够从微量细胞样品中高效率的富集转录因子。利用该亲和介质从单个细胞样品中鉴定到了 400 多个转录因子, 而从 11 个不同类型的细胞中共鉴定到 878 个转录因子, 涵盖了细胞内近 1/2 的基因组编码的转录因子产物, 实现了转录因子的高覆盖鉴定。同时, 该课题组的刘琼明等人^[19]利用荷尔蒙反应元件(HREs)DNA 序列作为亲和介质, 成功地从小鼠(*Mus musculus*)肝脏组织样品中富集到了低丰度的内源性核受体因子。生物介质方法的优势在于可针对不同的生物学问题, 设计不同的反应原件, 针对性地高效富集和鉴定低丰度的蛋白质因子, 进而揭示基因调控的复杂分子机制。有效鉴定目的蛋白是研究这些蛋白质功能的前提。这些高效的化学介质和生物介质的开发, 可实现目的蛋白质亚组的富集, 有效地增加了蛋白质组学的测序深度, 增强了利用蛋白质组学技术解析生命奥秘的能力。

(3) 蛋白质样品高效酶切处理技术的研究进展。目前, 蛋白质组学的鉴定和定量主要采用从肽段到蛋白, 即自下而上(bottom up)的策略, 因此蛋白质水

解成为肽段的过程是高端蛋白质组学技术研究中的重要一环。如何实现蛋白质组样品酶消化的高效性与特异性是蛋白质组学样品制备技术研究的热点。

复旦大学刘宝红等人^[20]发明了一种微芯片反应器, 通过将蛋白酶固定在具有较大比表面积的纳米粒子上, 最大程度地提高了蛋白酶和蛋白质样品的接触面积, 提高了底物蛋白质的水解效率。利用该反应器, 胰蛋白酶水解蛋白质底物的速度可达到 $400 \text{ mmol L}^{-1} \text{ min}^{-1} \mu\text{g}^{-1}$, 可实现蛋白质组样品的快速水解。此外, 该反应器还能够与后续的液相色谱和质谱联用蛋白质样品检测平台具有很好的兼容性, 可实现蛋白质水解和肽段的在线检测。

复旦大学张祥民团队和杨芃原团队^[21,22]发现激光辅助水解可提高蛋白质水解的效率, 缩短水解反应的时间。该方法不仅价廉, 还可在几秒钟内实现低至 2 μL 样品的快速、高效的酶解和快速检测, 因此适于高通量蛋白质组学的研究^[22]。为了提高蛋白质组的覆盖度, 邹汉法等人^[23]采用多种酶组合串联水解的策略, 取得了良好的效果。张玉奎和张丽华课题组^[24]还合作开发出了在线的自动化蛋白质快速水解技术和装置, 促进了高通量蛋白质组学研究。

有趣的是, 邹汉法等人^[25]发现胰蛋白酶不仅具有水解酶的作用, 还有连接酶的活性。这是世界上首次报告该蛋白酶具有连接酶的活性。该团队发现蛋白酶在水相条件下具有水解酶的活性, 但在含有较高有机相的环境中则起到连接酶的作用。该酶促标记反应条件温和, 标记区域的特异性高, 可有效减少常规化学标记方法中肽段降解和副反应的发生。据此, 该团队发展了特异的 N 端稳定同位素标记肽段的新方法, 并成功应用于定量蛋白质组学分析。N 端特异性标记的肽段可以形成丰富的 b 和 y 离子, 这为肽段从头测序技术的开发奠定了良好的基础。

2.2 蛋白质组鉴定技术的发展

受益于色谱系统的高效分离技术的发展以及快速、高精度高分辨质谱仪器的开发, 基于色谱-质谱联用技术平台的蛋白质组学深度覆盖成为现实。多维液相色谱耦联串联质谱技术是大规模、高覆盖蛋白质组学研究的重要手段。北京蛋白质组研究中心丁琛等人^[26]开发了一套快速、高效的鉴定和定量蛋白质组的研究策略, 使得鉴定以及定量的深度和效率得到了双重提高。利用这种方法, 在 12 h 的质谱分析时间

内达到 8000 个人类细胞基因产物的鉴定量. 这种“Fast-seq”策略省略了“脱盐”这个步骤, 并且通用于常规的无标定量和基于标记技术的相对和绝对定量蛋白质组定量技术. 同样利用传统的 2D-MS 技术平台, 对肝脏蛋白质组学研究取得了不错的成绩^[27,28]. 杨芑原团队^[27]结合密度梯度离心和 2D-MS 技术系统分析了 C57BL/6J 小鼠肝脏核蛋白质组, 有效鉴定了 748 个低丰度蛋白, 为肝脏核蛋白质组研究提供了借鉴. 多维液相色谱分离技术也同样适用于磷酸化蛋白质组学研究. 邹汉法团队^[29]建立了一套新的反相-反相(reverse phase-reverse phase, RP-RP)二维色谱联用技术, 其中第一维选择离线高 pH 反相液相色谱分离, 第二维是在线的低 pH 反相液相色谱, 并直接串联质谱仪. 这两维分离技术交叉极大地降低了蛋白质样品的复杂性, 成功地将磷酸化肽段和位点鉴定数量提升到 8000 多个. 除此之外, 蛋白质组学在生物标志物的鉴定与发现以及药物疗效评估等方面发挥着独特的作用^[30-32]. 血浆分泌蛋白质组学研究有力的支撑了生物标志物发现, 利用质谱技术鉴定比较了正常细胞与乙型肝炎病毒(hepatitis B virus, HBV)相关肝癌细胞分泌蛋白质组, 共鉴定了 1365 个蛋白, 第一次报道了 MMP1(Metalloproteinase 1)比现在常用的肿瘤标志物甲胎蛋白(α -fetoprotein, AFP)具有更好的灵敏度与特异性^[30].

2.3 定量蛋白质组学及其应用

由定性到定量是蛋白质组学发展的必然趋势. 对蛋白质组进行准确、高效的定量研究将极大地促进人类对以蛋白质为基础的生理和病理分子机制的理解. 近年来, 多种新型定量标记技术被开发, 并在生命科学研究领域得到了广泛的应用. 这些技术根据样品制备时标记与否, 分为无标定量、化学标记定量和代谢标记定量等 3 种主要的技术策略. 三者各有优缺点, 但后两者的定量精度较高, 在定量蛋白质组学研究中得到了有效的推广.

(1) 化学标记法定量策略的发展. 化学标记法是在体外对制备的蛋白质或多肽样品进行标签标记, 可对任何细胞、组织和器官等几乎所有的蛋白质组样品进行定量研究, 操作简便, 应用范围广. 但传统的技术存在一定的技术难题, 有待深入研究和改进. 针对同位素相对标记与绝对定量技术(isobaric tags for relative and absolute quantitation, iTRAQ)在复杂生物

样本存在标记偏性, 中国科学院北京基因组研究所刘斯奇团队^[33]开发了新的同位素标记策略——氘同位素氨基反应标签标记策略(deuterium isobaric amine reactive tag, DiART), 并对这两种标记方法在碎裂机制、可定量蛋白数以及定量精度方面做了详实的对比. 结果表明, DiART 比 iTRAQ 策略中的报告离子信号强度高, 定量精度更为准确, 并且显现出较 iTRAQ 更低的报告离子定量比例的压制, 更有利于在肽段层面的精确定量研究.

翻译后修饰定量蛋白质组学因兼具富集修饰和准确定量双重挑战, 难度更大. 邹汉法团队^[34]发展了一种新的磷酸化蛋白质组学定量策略, 能够显著提高蛋白质定量精度和通量. 将对照组、实验组和一个重复对照组分别用“轻”、“中”和“重”稳定性同位素标记的二甲基正常肝组织、肝癌细胞的磷酸化肽进行三重标记, 并利用反相色谱/强离子交换色谱/反相色谱的在线多维液相色谱对标记后混合的样品进行深度分离简化, 经过 42 h 的分离鉴定耗时, 实现了 1800 个磷酸化肽段的近 2000 个磷酸化位点的鉴定和定量^[35].

在定量标签方面, 邹汉法团队^[36]也取得了较好的成果. 该小组开发的新颖的肽段 N-端同位素标记策略, 该策略使用含有不同稳定同位素的甲醛作为标记试剂, 可以在肽段的 N-端发生二甲基化的修饰, 从而进行标记定量. 该策略在复杂生物样本中可以达到 98% 的 N-端位点选择特异性和 99% 的标记效率.

(2) 代谢标记法定量策略的发展. 稳定同位素代谢标记技术(stable isotope labeling by amino acids in cell culture, SILAC)是定量蛋白质组学中的金标准. 杨芑原等人^[35]交替使用 $^{13}\text{C}_6$ -Arg 和 $^{13}\text{C}_6$ -Lys 对两组样品进行稳定性同位素代谢标记. 将样品等比例混合后, 使用特异性的蛋白内切酶 Lys-N 和 Arg-C 进行消化, 产生以 Lys 为 N-端和 Arg 为 C-端的肽段. 对非起始或者结尾的样品蛋白肽段产物, 两组样品在一级质谱中分子离子的 m/z 值一致, 这些母离子在碎裂后生成的二级质谱谱图中将产生相同质量迁移单位的 b 和 y 系列离子对. 依据 b 和 y 系列离子对的信号强度比, 可实现更为精确的定量. 但对于 N 端或 C 端的肽段, 能可在一级质谱谱图中形成一定质量单位迁移的离子对, 不仅有利于区分这些末端肽段, 而且可实现这些末端离子对的直接定量.

中国科学院上海生命科学研究院的曾嵘团队^[37]则将基于 SILAC 的定量蛋白质组学方法向转化医学应用推进了一步. 该课题组将 SILAC 小鼠应用到 IgA 肾病患者血清和尿液的蛋白质组学研究中, 发现了一系列可提供临床诊断的候选关键蛋白分子, 如: 补体 C3, 白蛋白, VDBP, ApoA1, IGFBP7 等. 这是 SILAC 标记小鼠在人类疾病小鼠模型的体液蛋白质组学的直接应用, 这为新的生物标志物筛选提供了新的系统鉴定方法.

除了定量蛋白质组样品外, 中国科学院水生生物研究所葛峰团队^[38,39]将 SILAC 技术成功应用到 microRNA(miR)生物功能的定量研究, 选择 siRNA 方法敲低 U266 骨髓瘤细胞内源性 miR-21 的细胞系作为研究材料, 然后利用 SILAC 方法对 miR-21 潜在的靶标分子进行系统的定量研究, 最终使用 Western blot 和报告基因策略证实了 STAT 3 蛋白抑制分子 (protein inhibitor of activated STAT 3, PIAS3) 能够抑制信号转导子和转录激活子 3 (signal transducer and activator of transcription 3, STAT3) 的激活的关键分子是 miR-21 的直接底物.

2.4 蛋白质组翻译后修饰研究的技术发展

蛋白质翻译后修饰 (post translational modification, PTM) 在各种生物进程中都发挥着重要的作用. 比如目前研究比较深入的磷酸化、糖基化、泛素化等修饰, 在蛋白质功能和定位等方面都发挥着重要的作用. 种类繁多的蛋白质翻译后修饰极大地增加了蛋白质组的复杂度和动态范围. 同时由于高丰度蛋白的抑制效应, 使得对蛋白质翻译后修饰的鉴定和定量成为挑战. 因此, 发展有效的 PTM 富集策略对质谱的鉴定分析显得尤为重要.

(1) 磷酸化蛋白质组学的技术发展. 蛋白质磷酸化修饰是目前研究最为成熟的翻译后修饰类型. 中国蛋白质组学家在磷酸化肽段富集方法的开发、质谱分析、生物信息学支撑和功能研究等诸多方面均取得了突出的成绩和进展.

新型富集材料的开发为高效特异的捕捉磷酸化肽段提供了保障. 蛋白质组学家从亲和离子、载体材料、结构等诸多方面进行尝试, 发展了多种新型材料^[40-43], 很好地改善了磷酸化肽段的富集效率和特异性, 为后续大规模研究磷酸化蛋白质组提供了多种候选材料.

磷酸化蛋白的消化和洗脱条件与后续 MALDI-TOF 质谱分析不兼容, 进行分析之前需要进行调节 pH 值、脱盐等繁琐的步骤. 邹汉法等人^[44]将强阴离子交换毛细管液相色谱与 MALDI-TOF 质谱联用, 简化了进行质谱分析前样品处理步骤, 同时降低了样品损失. 该方法适合于单一磷酸化蛋白的磷酸化位点的鉴定, 能够检测到低至 10~50 amol 的磷酸化肽段.

在磷酸化肽段富集过程中, 多磷酸化修饰的肽段的富集和鉴定是其难点. 为了理解富集过程中磷酸化肽段的竞争关系, 曾嵘等人^[45]详细评估了结合到 Ti 珠子的磷酸化肽段的比例, 提高了富集效率. 同时发现当 TiO₂ 珠子量不足时, 更加倾向于富集多磷酸化肽段. 此外, 北京大学纪建国等人^[46]发现柠檬酸的浓度对于单、多磷酸化肽段与 TiO₂ 珠子的结合有密切关系, 该团队发展了级联富集的方法, 通过改变柠檬酸的浓度来区分多磷酸化和单磷酸化富集.

为了提高磷酸化蛋白质组的覆盖度, 邹汉法等人^[23]通过利用组合不同酶对蛋白进行酶切来制备样品. 组合 Glu-C 和 Trypsin 对蛋白质组样品进行消化和磷酸化肽段富集, 该小组从 HeLa 细胞系中共鉴定了 8062 个特异磷酸化肽段的 8507 个磷酸化位点, 相比于单一 Trypsin 酶切, 鉴定量提高了近 1 倍.

复杂的蛋白质组样品的分离简化也有利于磷酸化肽段的富集和鉴定. 邹汉法团队^[29]发展了利用二维液相色谱的策略进行磷酸化肽段的分离和质谱鉴定的高效技术体系. 在这个体系中, 首先利用离线的高 pH 反相色谱进行肽段分离, 然后利用低 pH 液相色谱在线分离并进行质谱分析, 由此可提高 30% 的磷酸化肽段的鉴定量. 通过不同磷酸化肽段富集策略的组合同样能够提高富集效率, 且解决单一方法存在的富集偏性的问题. 曾嵘等人^[47]发展了 1 套利用强阴离子交换结合 TiO₂ 富集的策略, 第一步采用强阴离子交换柱富集磷酸化肽段, 流穿峰部分采用 TiO₂ 富集磷酸化肽段. 该组合策略实现了对样品磷酸化蛋白质组的高效且无偏性的富集分析.

除了定性研究, 翻译后修饰的定量研究对于揭示其生物学意义更加重要. 邹汉法等人^[48]利用 Ti(4+)-EPO 纳米颗粒为吸附剂, 富集同位素标记了的内源性磷酸化肽段. 还发展了 pseudo-triplex 稳定性同位素双甲基化标记, 实现了高准确性、高通量的定量磷酸化蛋白质组分析. 这些方法的开发为疾病发生发展过程中磷酸化修饰相关生物标志物的发现

和分子机制研究创造了条件^[34].

(2) 糖蛋白质组学的技术发展. 蛋白质的糖基化修饰及失调在正常生理状态和多种疾病发生发展过程中扮演着重要的角色. 天冬酰胺(N)-连接的糖蛋白质组研究已经形成了较为成熟的技术体系, 并得到了比较深入的研究. 为了富集 N 修饰的糖肽, 邹汉法团队和复旦大学陆豪杰团队^[49-52]开发了基于磁性纳米填料的毛细管柱子, 实现了 N-修饰糖肽的高效富集. 为了更加高效地分析 N-糖苷修饰的蛋白质组, 很多实验室系统研究并发展了相应的蛋白质组学平台, 实现了对血清以及组织样品的糖蛋白质组的分析^[53-55]. 核心岩藻糖化是蛋白质糖基化中的另一种重要的修饰形式, 但至今研究尚少. 北京蛋白质组研究中心的钱小红团队^[56]在大规模鉴定核心岩藻糖化蛋白质的基础上发展了基于多级反应监测的定量蛋白质组学技术方法, 实现了对核心岩藻糖化蛋白特异性位点的定向检测和精确定量.

除了糖蛋白和其特异性修饰位点外, 来自糖蛋白底物上的多糖同样可被详细表征, 并可作为生物标志物. 钱小红等人^[57]发展了一套基于 PCGO(1-pyrenebutyryl chloride functionalized free graphene oxide)实现快速、高效的多聚糖苷富集策略, 可比较简便地分析这些多聚糖型. 邹汉法团队^[58]同样发展了基于 OMC 的 N-糖苷的富集策略, 该方法可在有效去除蛋白质的同时又可特异性的富集糖链, 大大提高了糖链的信号强度. 杨芃原等人^[59]发展了基于酶切过程中在 N-糖末端引入 ¹⁸O 标记糖苷的方法, 实现了对修饰糖链的定量比较. 这些研究极大地推动了糖蛋白质组学和生物标志物的研究.

(3) 泛素化蛋白质组学研究技术的发展. 泛素化修饰是常见的蛋白质翻译后修饰之一. 泛素链修饰的特异性也是蛋白质底物发生特异性泛素化修饰的一种形式. 不同的泛素链修饰携带不同的化学结构信息, 传递底物蛋白特异的生物学功能. 泛素化修饰的特异性主要由泛素连接酶(E3s)来决定. 然而, 绝大多数被泛素化修饰的底物蛋白的特异性泛素连接酶尚不为我们所知. 中国医学科学院高友鹤等人^[60]建立了一套基于 E3 与蛋白质结合结构域相互作用的蛋白质组学技术特异性鉴定 E3 的底物蛋白的策略, 并且通过体外反应实验验证了该技术策略的有效性. 这是一种高通量的特定 E3 酶特异性修饰底物的筛选方法. 通过基于 E3 与蛋白质结合结构域相互作用的

蛋白质组筛选技术还可鉴定特定泛素链修饰的底物, 发现并证实了泛素连接酶(E3 ubiquitin-protein ligase, LNX1)介导的 PDZ 连接激酶(PDZ-binding kinase, PBK)经泛素化降解途径, 从而抑制细胞的增殖, 增加细胞对阿霉素的敏感性.

中国科学院上海生命科学研究院李林团队^[61]利用基于质谱的蛋白质翻译后修饰鉴定策略, 阐明了 Smurf1 能介导轴蛋白(axin)K29 位的多聚泛素化修饰. 进一步研究发现这种 K29 的多聚泛素链并不介导轴蛋白经泛素-蛋白酶体系的降解, 而是引导被修饰的轴蛋白通过与 Wnt 共受体 LRP5/6 发生相互作用, 引发 LRP5/6 的磷酸化修饰, 进而抑制 Wnt/ β -catenin 信号通路.

(4) 乙酰化蛋白质组学研究进展. 蛋白质乙酰化修饰是另一种同样具有重要生物功能的蛋白质翻译后修饰形式. 乙酰化修饰在染色体重塑、转录因子激活以及调节代谢过程中酶的活性等诸多方面具有重要作用. 研究蛋白质乙酰化修饰后的功能和机制的前提是鉴定蛋白质乙酰化修饰位点的. 然而, 细胞内存在大量的高丰度的乙酰化修饰的组蛋白, 这些乙酰化组蛋白的存在将极大的干扰动能多样的乙酰化非组蛋白的富集和质谱鉴定. 因此, 对于乙酰化修饰的研究主要受限于针对细胞内低丰度的乙酰化肽段的亲和纯化技术的发展.

复旦大学赵世民团队^[62]设计了可以运用于细胞内低丰度乙酰化肽段富集的新方法. 该团队首先获得了泛乙酰化修饰肽段的特异性抗体和针对精氨酸琥珀酸裂解酶第 288 位赖氨酸发生乙酰化修饰的特异性抗体; 然后通过亚细胞组分的分离, 降低乙酰化组蛋白的高丰度; 最后, 利用针对乙酰化修饰的抗体进行亲和富集, 结合 LC-MS/MS 技术, 鉴定乙酰化修饰的肽段和位点.

2.5 生物信息学发展

随着高通量蛋白质组学测序技术的快速发展, 海量质谱数据特别是高精度质谱数据的“爆炸式”涌现, 对数据的管理、共享和挖掘提出了挑战. 为此, 国内生物信息学家在数据质量控制、肽段和蛋白质的鉴定和定量、从头测序、基因组重新注释、翻译后修饰蛋白质的鉴定和蛋白质交联鉴定的算法和软件工具发展等方面, 都开展了系统深入的研究工作并取得了一系列进展. 质谱数据的深度挖掘使蛋白质组学

成为新蛋白和新修饰发现的有力工具。此外, 蛋白质相互作用预测、蛋白质 motif 数据库和基于蛋白质的生物标志物数据库等开发也取得了较好的研究进展。

(1) 蛋白质鉴定、定量相关的生物信息工具开发。基于质谱数据的蛋白质鉴定流程包括实验和计算两部分。计算部分一般包括图谱预处理、数据库搜索和质量控制 3 个部分。

在图谱预处理部分, 精确的母离子质量挑选能够降低鉴定的假阳性, 提高图谱解析率和定量准确性。中国科学院计算技术研究所贺思敏团队^[63]开发了 pParse, 利用同位素峰簇中第一同位素峰和最高峰的位置关系挑选候选峰簇, 并根据相似性和色谱峰强度确定单一同位素峰, 识别、区分共洗脱肽段, 有效地提高了图谱解析率。

在数据的质量控制方面, 北京蛋白质组研究中心朱云平团队^[64]开发了肽段鉴定质量控制软件 PepDistiller, 用于 MASCOT 搜库结果的数据质量控制。相比于业内广泛使用的质控软件 MASCOT Percolator^[65], 其对半酶切搜库结果质控的灵敏度有显著提升。PepDistiller 内置的多线程技术也能大大加快高通量质谱数据的处理速度。曾嵘团队^[66]开发了质控工具 BuildSummary, 通过将图谱鉴定结果按照特定属性分类后单独过滤, 并以蛋白质水平错误发现率(false discovery rate, FDR)为最终过滤条件, 实现了多搜索引擎鉴定结果的数据整合, 也适合来自不同样品或仪器类型的不同数据集的整合。

搜索引擎是质谱数据解析的基本工具。贺思敏团队^[67,68]经过多年努力, 研制了我国具有完全自主知识产权的搜索引擎 pFind, 并在业内得到了广泛推广。中国科学院上海生命科学研究院李亦学团队^[69]也开发了基于支持向量机的多特征整合引擎, 用于一级图谱鉴定, 其灵敏度和准确性都要优于已有搜索引擎的 MASCOT, ProFound 等。

从头测序技术不依赖数据库而直接利用图谱信息解析肽段, 常被用于新蛋白和新修饰的鉴定。但从头测序算法的图谱质量较高, 限制了该算法在低精度质谱数据的实际应用。研究表明, 能被数据库搜索鉴定的图谱中只有不超过一半的图谱可被从头测序算法正确鉴定^[70]。高能碰撞诱导碎裂(higher-energy collisional dissociation, HCD)和电子转移碎裂(electron transfer dissociation, ETD)技术所生成的高精度二级质谱谱图的二级碎裂分子离子的连续性好、

离子类型多、质量精度高, 因此使用从头测序算法对其进行解析, 可取得较好的效果。而 HCD 和 ETD 图谱的碎裂离子的类型还具有互补性, 可进一步提升高精度质谱谱图的从头测序的解析率。中国科学院计算技术研究所贺思敏团队和北京生命科学研究董梦秋团队^[71]合作开发了从头测序软件 pNovo, 充分利用了 HCD 图谱二级离子高精度以及具有丰富的 internal 离子和 immonium 离子等特点, 使得 pNovo 正确解析图谱数覆盖常规数据库搜索鉴定数的 80% 以上, 并能同时有效鉴定出脱酰氨基修饰和氨基酸突变。在该软件升级版 pNovo+ 中, 除 HCD 图谱外, 来自同一肽段的 ETD 图谱的二级离子信息(包括 c, z 离子及氢重排离子)被整合入图谱节点构建和动态规划算法中, 使得常规搜库约 95% 的鉴定结果可被 pNovo+ 覆盖^[72]。pNovo 在新蛋白鉴定方面得到了很好的应用, 例如中国科学院生物物理研究所苗龙课题组^[73]利用 pNovo 鉴定到了 1 种丝氨酸蛋白酶 As_TRY-5 及其抑制因子 As_SRP-1。这 2 种蛋白质是线虫纲物种中调节精子活化或精子竞争优势的关键蛋白。

此外, 贺思敏团队和董梦秋团队^[74]合作开发了用于肽段交联质谱鉴定的软件 pLink, 实现了规模化的化学交联蛋白质的结构鉴定。pLink 通过图谱预过滤、候选交联肽段初筛及 KSDP 图谱匹配打分的优化^[75], 实现了交联肽段的快速鉴定, 并设计了适用于交联肽段鉴定的 target-decoy 策略, 有效地控制了交联鉴定的 FDR 水平。pLink 的有效性进一步在纯化的蛋白质样品、蛋白质复合物、免疫共沉淀样品、模式生物的全细胞裂解液上得到了验证。

定量蛋白质组学根据是否需要稳定同位素标记可以分为无标定量和有标定量两大类。在蛋白质无标定量方面, 朱云平团队组和国防科学技术大学谢红卫团队^[76]合作开发了无标定量工具 LFQuant, 实现了无标实验策略下的蛋白质精确定量。LFQuant 采用了新的交叉搜索算法, 大大降低了计算复杂度, 提升了软件的效率, 实现了对多种格式质谱数据的定量分析。其核心算法在多组标准数据集上进行了测试和比较, 优于目前的定量软件 MaxQuant^[77]以及 IDEAL-Q^[78], 目前已经成功应用于中国人类染色体计划等大规模数据的定量分析中^[79]。针对基于稳定同位素标记数据的定量分析, 朱云平、谢红卫团队^[80]还合作开发了有标定量工具 SILVER。和其他定量软

件相比, SILVER 中额外包含了新的定量可靠性评估指标和过滤打分, 在不降低定量灵敏性的前提下, 提高了定量准确性. 与目前常用的定量软件 MaxQuant 和 Proteome Discoverer 相比, SILVER 不仅能够对 SILAC、 ^{18}O 等标记方式进行精确定量, 还能够额外支持 ^{15}N 标记的定量数据分析, 扩大了质谱数据定量软件的使用范围. 此外, LFQuant 和 SILVER 都具有用户友好的界面, 能够进行定量结果的系统展示和统计学分析, 方便用户使用. 贺思敏团队和董梦秋团队^[81]也联合推出了有标定量软件 pQuant, 在肽段定量时额外考虑了不完全标记的情况, 同时在蛋白质定量值计算时采用了非参数模型进行评估, 有效提高其定量准确性.

(2) 蛋白质基因组学在基因组重注释上的应用. 后基因组时代, 蛋白质组学不仅仅依靠基因组数据库进行蛋白质组的鉴定和定量, 也可以反过来对已有的基因结构注释进行验证和修正. 原核生物基因组小, 基因结构简单, 可直接通过六框搜索进行基因组重注释和修正. 真核生物基因组庞大, 基因结构复杂, 因而对于注释修正算法以及结果的假阳性控制均有更高的要求, 以保证结果的准确性和灵敏度.

目前, 国外在利用蛋白质组数据进行基因组注释的修正方面已发展了一系列算法或工具, 并在小鼠或人类等高等生物样品数据集上得到了较好的应用. 国内学者在这个方向也进行了一定的尝试. 上海生物信息技术研究中心谢鹭和中国科学院上海生命科学研究院李亦学团队^[82,83]合作在使用外显子组合数据库对小鼠的高精度质谱数据进行了尝试并开发了 iGepros, 成功应用到基因和蛋白质的整合注释的研究中. 值得一提的是中国科学家积极参与和领导国际 HUPO 组织的国际染色体蛋白质组计划 (chromosome-centric human proteome project, C-HPP), 旨在通过国际合作对各条染色体上基因编码的蛋白质产物进行鉴定, 并进一步完善基因组的注释. 在该计划中, 中国团队由中国人民解放军军事医学科学院、北京蛋白质组研究中心、复旦大学、北京基因组研究所和暨南大学构成, 并由贺福初、杨芑原、刘斯奇、何庆瑜等人分别领衔、负责人类 1、8 和 20 号染色体上编码基因的蛋白质产物的鉴定工作, 经过两年的努力, 取得了阶段性进展, 已鉴定 62% 左右的人类基因组编码基因产物, 并有望发现系列新的漏注释基因^[79,84-86].

(3) 蛋白质翻译后修饰的鉴定和质控. 蛋白质翻译后修饰 (PTM) 的鉴定存在两个难点: (i) 实验检测难. PTM 的检测属于亚化学计量水平, 且存在时间短, 一般是瞬时出现或动态变化, 必须依靠高灵敏度的分析化学检测技术才可能捕获; (ii) 计算难. 蛋白质修饰类型丰富, 而每种蛋白质的多个氨基酸残基都可能发生修饰, 使得 PTM 的鉴定碰到“组合爆炸”问题, 难于处理. 常规的数据库搜索策略需预先指定修饰类型, 单次搜索允许的单肽段修饰种类和修饰个数都受到限制, 只能鉴定数据集中的部分质谱数据.

然而即使采用双高精的质谱仪 (例如 HCD), 图谱的解析率也只有 50% 左右, 因此, 人们认为未解析的图谱中仍蕴含着丰富的新蛋白、可变剪切、氨基酸突变和翻译后修饰信息, 亟待挖掘. 目前人们普遍认为, 未修饰肽段和修饰肽段在样品中是同时存在的, 因此通过挖掘高频出现的肽段母离子质量差和色谱保留时间差, 则有望快速发现潜在的修饰类型. 利用这一原理, 中国科学院计算技术研究所付岩^[87]开发了修饰挖掘工具 DeltAMT. 该软件采用二维高斯混合模型进行建模, 并设计了 D-score 进行修饰类型过滤, 可以有效地对单一修饰或混合修饰进行鉴定. 在实际运用中, DeltAMT 对高丰度修饰 (例如一些体外修饰) 有较好的鉴定效果, 但对低丰度修饰的挖掘仍存在一定困难.

在对富集的特定修饰类型的实验数据进行处理时, 翻译后修饰鉴定的数据质量控制对鉴定的准确性也尤为重要. 不同磷酸化肽段具有不同的碎裂行为. 邹汉法团队^[88]利用这个特征, 在磷酸化三级质谱策略中运用了分类过滤策略, 即根据二级谱图是否有中性丢失峰将磷酸化肽段进行分类, 每类单独过滤, 提高了磷酸化肽段鉴定的分辨率和磷酸化蛋白质测序的覆盖度, 提高了磷酸化鉴定的灵敏度, 实现了高效的鉴定. 该团队还发展了人类血清已鉴定蛋白的正反数据库, 并利用二次搜库质控的策略提高了磷酸化肽段的鉴定灵敏度, 也降低了搜库时间, 促进了血液磷酸化蛋白质组学的研究^[89]. 利用磷酸化蛋白质组的大数据集, 李亦学等人^[90]发现磷酸化位点在脊椎动物内特定功能分子的保守性高于其他基本分子和其侧翼区域. 基此, 认为磷酸化位点在脊椎动物的进化过程中发挥着重要的作用. 姚雪彪和薛宇等人^[91]分别用特异激酶的磷酸化位点信息预测工具 (GPS2.0) 分析了单核苷酸多态性 (single nucleotide

polymorphisms, SNP)数据,发现约 70%报道的 SNP 位点是潜在的磷酸化 SNP. 大约有 74.6%的潜在的磷酸化 SNP 包含了激酶类型相关的磷酸化位点的改变,而不是直接产生或者去除磷酸化位点. 这些结果有助于深入理解人类疾病过程中的基因差异,并更好地为个性化医疗服务.

(4) 蛋白质相互作用预测. 蛋白质自相互作用的研究有助于蛋白质相互作用网络进化的探索,揭示蛋白质的生物学功能. 但目前尚无蛋白质自相互作用预测工具研究的报道. 在开发蛋白质相互作用预测工具 PRINCESS 之后,北京蛋白质组研究中心李栋团队^[92]又相继开发了蛋白质自相互作用预测工具 SLIPPER. SLIPPER 通过整合功能注释和网络拓扑结构特征,采用特征筛选和 Logistic 回归对自相互作用进行预测发现,自相互作用蛋白质倾向于含有更多的结构域,进化上也更保守,生物学功能上更倾向于酶类、看家基因和药物靶点等蛋白质. 由于这些蛋白质在相互作用网络中多占据关键节点,因此蛋白质自相互作用预测研究还为蛋白质相互作用网络构建和验证创造条件.

(5) 蛋白质组学研究相关数据库构建的研究进展. 海量高通量质谱数据的处理催生了质谱数据分析软件的蓬勃发展. 在基本解决了鉴定和定量问题后,这些大数据蕴含的丰富生物功能信息的挖掘同样需要生物信息学的支撑. 谢鹭团队^[93]通过收集覆盖 20 种人类癌症的 331 套实验数据,提供了人癌症差异蛋白质组数据分析可供参考的数据库. 高友鹤团队^[94]通过文献挖掘和人工确认,建立了人和动物的尿液蛋白质生物标志物数据库 UPB,并且发现不同方法鉴定的生物标志物的低重叠率可能与不同实

验室的实验流程差异有关.

最近薛宇团队^[95]在蛋白质的泛素化和乙酰化预测方面开发了一些列数据库,促进了翻译后修饰的注释研究. 例如,通过文献挖掘和人工确认,收集了 26 种 E1, 105 种 E2, 1003 种 E3 和 148 种去泛素化酶或去类泛素化酶的信息,结合 E3 的分类研究,最终构建了泛素结合和类泛素结合数据库(ubiquitin and ubiquitin-like conjugation database, UUCD). 该数据库收录了横跨 70 个真核生物物种的约 6 万种酶的信息. 与此类似,还建立了 CPLA 等赖氨酸乙酰化位点数据库^[96,97],以及收录了包括人在内的 7 种模式生物的上千种在细胞分化和有丝分裂过程中与中间体、中心体和着丝粒相结合的蛋白质数据库 MiCroKit^[98],为染色体相关的蛋白质组学的研究创造了条件.

3 展望

在中国蛋白质组学家不懈努力和各种科学基金的鼎力支持下,中国的蛋白质组学研究在过去的 3 年中继续保持了欣欣向荣、蓬勃发展的态势. 中国蛋白质组学研究团队承担的国际肝脏蛋白质组计划取得了阶段性成果,引领了国际器官蛋白质组学的发展,并影响着后续的国际人类蛋白质组学研究. 在国际肝脏蛋白质组计划实施过程中形成的思路、策略、技术体系和人才队伍为刚刚启动的中国人类蛋白质组计划的开展奠定了坚实的基础. 可以预见,通过未来几年的努力,人类蛋白质组学的研究成果在阐释生理与病理过程的分子机制的研究中将会发挥更加积极的作用,并为人类的健康事业和生命科学的发展奠定基础.

致谢 衷心感谢中央政府和地方各级科技部门,特别是国家大科学设施计划、国家自然科学基金以及北京市政府对蛋白质组学发展的支持和资助. 同时感谢 CNHUPO 组织以及中国蛋白质组学家的支持与帮助,特别是引文中的作者在蛋白质组学发展中做出的努力与贡献. 限于篇幅限制,其他 CNHUPO 成员卓有成效的工作没有被提及,在此一并表示对他们的敬意与感谢.

参考文献

- 1 Gao X, Zhang X, Zheng J, et al. Proteomics in china: ready for prime time. *Sci China Life Sci*, 2010, 53: 22–33
- 2 He F. At a glance: proteomics in china. *Sci China Life Sci*, 2011, 54: 1–2
- 3 贺福初. 大发现时代的“生命组学”(代序). *中国科学: 生命科学*, 2013, 43: 1–15

- 4 Leng F. Opportunity and challenge: ten years of proteomics in china. *Sci China Life Sci*, 2012, 55: 837–839
- 5 He F. Human liver proteome project: plan, progress, and perspectives. *Mol Cell Proteomics*, 2005, 4: 1841–1848
- 6 Zhang X, Guo Y, Song Y, et al. Proteomic analysis of individual variation in normal livers of human beings using difference gel electrophoresis. *Proteomics*, 2006, 6: 5260–5268
- 7 Sun A, Jiang Y, Wang X, et al. Liverbase: a comprehensive view of human liver biology. *J Proteome Res*, 2010, 9: 50–58
- 8 CHLPP C. First insight into the human liver proteome from proteome(sky)-liver(hu) 1.0, a publicly available database. *J Proteome Res*, 2010, 9: 79–94
- 9 Zhao S, Xu W, Jiang W, et al. Regulation of cellular metabolism by protein lysine acetylation. *Science*, 2010, 327: 1000–1004
- 10 Wang Q, Zhang Y, Yang C, et al. Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science*, 2010, 327: 1004–1007
- 11 Song C, Ye M, Liu Z, et al. Systematic analysis of protein phosphorylation networks from phosphoproteomic data. *Mol Cell Proteomics*, 2012, 11: 1070–1083
- 12 Wang J, Huo K, Ma L, et al. Toward an understanding of the protein interaction network of the human liver. *Mol Syst Biol*, 2011, 7: 536
- 13 Liu S, Chen H, Lu X, et al. Facile synthesis of copper(ii) immobilized on magnetic mesoporous silica microspheres for selective enrichment of peptides for mass spectrometry analysis. *Angew Chem Int Ed Engl*, 2010, 49: 7557–7561
- 14 Wan H, Qin H, Xiong Z, et al. Facile synthesis of yolk-shell magnetic mesoporous carbon microspheres for efficient enrichment of low abundance peptides. *Nanoscale*, 2013, 5: 10936–10944
- 15 Qin H, Gao P, Wang F, et al. Highly efficient extraction of serum peptides by ordered mesoporous carbon. *Angew Chem Int Ed Engl*, 2011, 50: 12218–12221
- 16 Jian G, Liu Y, He X, et al. Click chemistry: a new facile and efficient strategy for the preparation of Fe₃O₄ nanoparticles covalently functionalized with ida-cu and their application in the depletion of abundant protein in blood samples. *Nanoscale*, 2012, 4: 6336–6342
- 17 Zhang M, He X, Chen L, et al. Preparation and characterization of iminodiacetic acid-functionalized magnetic nanoparticles and its selective removal of bovine hemoglobin. *Nanotechnology*, 2011, 22: 065705
- 18 Ding C, Chan D W, Liu W, et al. Proteome-wide profiling of activated transcription factors with a concatenated tandem array of transcription factor response elements. *Proc Natl Acad Sci USA*, 2013, 110: 6771–6776
- 19 Liu Q, Ding C, Liu W, et al. In-depth proteomic characterization of endogenous nuclear receptors in mouse liver. *Mol Cell Proteomics*, 2013, 12: 473–484
- 20 Liu Y, Xue Y, Ji J, et al. Gold nanoparticle assembly microfluidic reactor for efficient on-line proteolysis. *Mol Cell Proteomics*, 2007, 6: 1428–1436
- 21 Zhang X, Zhu S, Xiong Y, et al. Development of a MALDI-TOF ms strategy for the high-throughput analysis of biomarkers: on-target aptamer immobilization and laser-accelerated proteolysis. *Angew Chem Int Ed Engl*, 2013, 52: 6055–6058
- 22 Yao G, Deng C, Zhang X, et al. Efficient tryptic proteolysis accelerated by laser radiation for peptide mapping in proteome analysis. *Angew Chem Int Ed Engl*, 2010, 49: 8185–8189
- 23 Bian Y, Ye M, Song C, et al. Improve the coverage for the analysis of phosphoproteome of hela cells by a tandem digestion approach. *J Proteome Res*, 2012, 11: 2828–2837
- 24 Sun L, Ma J, Qiao X, et al. Integrated device for online sample buffer exchange, protein enrichment, and digestion. *Anal Chem*, 2010, 82: 2574–2579
- 25 Pan Y, Ye M, Zhao L, et al. N-terminal labeling of peptides by trypsin-catalyzed ligation for quantitative proteomics. *Angew Chem Int Ed Engl*, 2013, 52: 9205–9209
- 26 Ding C, Jiang J, Wei J, et al. A fast workflow for identification and quantification of proteomes. *Mol Cell Proteomics*, 2013, 12: 2370–2380
- 27 Zhang Y, Fang C, Bao H, et al. Nuclear proteome profile of C57BL/6J mouse liver. *Sci China Life Sci*, 2013, 56: 513–523
- 28 赵倩倩, 冯艳玲, 贾小芳, 等. 免疫性肝纤维化大鼠肝非实质细胞蛋白质组分析. *中国科学: 生命科学*, 2013, 43: 928–938
- 29 Song C, Ye M, Han G, et al. Reversed-phase-reversed-phase liquid chromatography approach with high orthogonality for multidimensional separation of phosphopeptides. *Anal Chem*, 2010, 82: 53–56
- 30 杨磊, 荣维淇, 肖汀, 等. 应用分泌/释放蛋白质组学鉴定HBV相关的肝细胞癌血浆标志物. *中国科学: 生命科学*, 2013, 43: 663–671
- 31 刘玥, 殷惠军, 陈可冀. 血小板蛋白质组学及其在血瘀证与活血化瘀中药研究中的探索应用. *中国科学: 生命科学*, 2013, 43: 619–625
- 32 周圣涛, 刘锐, 赵霞, 等. 病毒蛋白质组学: 病毒学研究前沿. *中国科学: 生命科学*, 2010, 40: 767–777
- 33 Chen Z, Wang Q, Lin L, et al. Comparative evaluation of two isobaric labeling tags, diart and itraq. *Anal Chem*, 2012, 84: 2908–2915
- 34 Song C, Wang F, Ye M, et al. Improvement of the quantification accuracy and throughput for phosphoproteome analysis by a pseudo triplex stable isotope dimethyl labeling approach. *Anal Chem*, 2011, 83: 7755–7762

- 35 Nie A Y, Zhang L, Yan G Q, et al. *In vivo* termini amino acid labeling for quantitative proteomics. *Anal Chem*, 2011, 83: 6026–6033
- 36 Qin H, Wang F, Zhang Y, et al. Isobaric cross-sequence labeling of peptides by using site-selective N-terminus dimethylation. *Chem Commun (Camb)*, 2012, 48: 6265–6267
- 37 Zhao S, Li R, Cai X, et al. The application of SILAC mouse in human body fluid proteomics analysis reveals protein patterns associated with IgA nephropathy. *Evid Based Complement Alternat Med*, 2013, 2013: 275390
- 38 Xiong Q, Zhong Q, Zhang J, et al. Identification of novel miR-21 target proteins in multiple myeloma cells by quantitative proteomics. *J Proteome Res*, 2012, 11: 2078–2090
- 39 Li C, Xiong Q, Zhang J, et al. Quantitative proteomic strategies for the identification of microRNA targets. *Expert Rev Proteomics*, 2012, 9: 549–559
- 40 Zhang Y, Chen C, Qin H, et al. The synthesis of Ti-hexagonal mesoporous silica for selective capture of phosphopeptides. *Chem Commun (Camb)*, 2010, 46: 2271–2273
- 41 Yan Y, Zheng Z, Deng C, et al. Hydrophilic polydopamine-coated graphene for metal ion immobilization as a novel immobilized metal ion affinity chromatography platform for phosphoproteome analysis. *Anal Chem*, 2013, 85: 8483–8487
- 42 Xu B, Zhou L, Wang F, et al. Selective capture of phosphopeptides by hierarchical Ti-aluminophosphate-5 molecular sieves. *Chem Commun (Camb)*, 2012, 48: 1802–1804
- 43 Zhang L, Zhao Q, Liang Z, et al. Synthesis of adenosine functionalized metal immobilized magnetic nanoparticles for highly selective and sensitive enrichment of phosphopeptides. *Chem Commun (Camb)*, 2012, 48: 6274–6276
- 44 Dong M, Wu M, Wang F, et al. Coupling strong anion-exchange monolithic capillary with MALDI-TOF ms for sensitive detection of phosphopeptides in protein digest. *Anal Chem*, 2010, 82: 2907–2915
- 45 Li Q R, Ning Z B, Tang J S, et al. Effect of peptide-to-TiO₂ beads ratio on phosphopeptide enrichment selectivity. *J Proteome Res*, 2009, 8: 5375–5381
- 46 Zhao X, Wang Q, Wang S, et al. Citric acid-assisted two-step enrichment with TiO₂ enhances the separation of multi- and monophosphorylated peptides and increases phosphoprotein profiling. *J Proteome Res*, 2013, 12: 2467–2476
- 47 Nie S, Dai J, Ning Z B, et al. Comprehensive profiling of phosphopeptides based on anion exchange followed by flow-through enrichment with titanium dioxide (AFET). *J Proteome Res*, 2010, 9: 4585–4594
- 48 Qin H, Wang F, Wang P, et al. Phosphoric acid functionalized mesoporous organo-silica (EPO) as the adsorbent for in situ enrichment and isotope labeling of endogenous phosphopeptides. *Chem Commun (Camb)*, 2012, 48: 961–963
- 49 Xiong Z, Qin H, Wan H, et al. Layer-by-layer assembly of multilayer polysaccharide coated magnetic nanoparticles for the selective enrichment of glycopeptides. *Chem Commun (Camb)*, 2013, 49: 9284–9286
- 50 Zhang Y, Kuang M, Zhang L, et al. An accessible protocol for solid-phase extraction of N-linked glycopeptides through reductive amination by amine-functionalized magnetic nanoparticles. *Anal Chem*, 2013, 85: 5535–5541
- 51 Liu J, Wang F, Lin H, et al. Monolithic capillary column based glycoproteomic reactor for high-sensitive analysis of N-glycoproteome. *Anal Chem*, 2013, 85: 2847–2852
- 52 Xiong Z, Zhao L, Wang F, et al. Synthesis of branched peg brushes hybrid hydrophilic magnetic nanoparticles for the selective enrichment of N-linked glycopeptides. *Chem Commun (Camb)*, 2012, 48: 8138–8140
- 53 Ma C, Zhao X, Han H, et al. N-linked glycoproteome profiling of human serum using tandem enrichment and multiple fraction concatenation. *Electrophoresis*, 2013, 34: 2440–2450
- 54 Zhu J, Wang F, Chen R, et al. Centrifugation assisted microreactor enables facile integration of trypsin digestion, hydrophilic interaction chromatography enrichment, and on-column deglycosylation for rapid and sensitive N-glycoproteome analysis. *Anal Chem*, 2012, 84: 5146–5153
- 55 Qu Y, Xia S, Yuan H, et al. Integrated sample pretreatment system for N-linked glycosylation site profiling with combination of hydrophilic interaction chromatography and PNGase F immobilized enzymatic reactor via a strong cation exchange precolumn. *Anal Chem*, 2011, 83: 7457–7463
- 56 Zhao Y, Jia W, Wang J, et al. Fragmentation and site-specific quantification of core fucosylated glycoprotein by multiple reaction monitoring-mass spectrometry. *Anal Chem*, 2011, 83: 8802–8809
- 57 Zhang W, Han H, Bai H, et al. A highly efficient and visualized method for glycan enrichment by self-assembling pyrene derivative functionalized free graphene oxide. *Anal Chem*, 2013, 85: 2703–2709
- 58 Qin H, Zhao L, Li R, et al. Size-selective enrichment of N-linked glycans using highly ordered mesoporous carbon material and detection by MALDI-TOF MS. *Anal Chem*, 2011, 83: 7721–7728
- 59 Zhang W, Wang H, Tang H, et al. Endoglycosidase-mediated incorporation of ¹⁸O into glycans for relative glycan quantitation. *Anal Chem*, 2011, 83: 4975–4981

- 60 Guo Z, Song E, Ma S, et al. Proteomics strategy to identify substrates of LNX, a PDZ domain-containing E3 ubiquitin ligase. *J Proteome Res*, 2012, 11: 4847–4862
- 61 Fei C, Li Z, Li C, et al. Smurf1-mediated Lys29-linked nonproteolytic polyubiquitination of axin negatively regulates Wnt/-catenin signaling. *Mol Cell Biol*, 2013, 33: 4095–4105
- 62 Guan K L, Yu W, Lin Y, et al. Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides. *Nat Protoc*, 2010, 5: 1583–1595
- 63 Yuan Z F, Liu C, Wang H P, et al. Pparse: a method for accurate determination of monoisotopic peaks in high-resolution mass spectra. *Proteomics*, 2012, 12: 226–235
- 64 Li N, Wu S, Zhang C, et al. Pepdistiller: a quality control tool to improve the sensitivity and accuracy of peptide identifications in shotgun proteomics. *Proteomics*, 2012, 12: 1720–1725
- 65 Brosch M, Yu L, Hubbard T, et al. Accurate and sensitive peptide identification with mascot percolator. *J Proteome Res*, 2009, 8: 3176–3181
- 66 Sheng Q, Dai J, Wu Y, et al. Buildsummary: using a group-based approach to improve the sensitivity of peptide/protein identification in shotgun proteomics. *J Proteome Res*, 2012, 11: 1494–1502
- 67 Li D, Fu Y, Sun R, et al. Pfind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics*, 2005, 21: 3049–3050
- 68 Wang L H, Li D Q, Fu Y, et al. Pfind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. *Rapid Commun Mass Spectrom*, 2007, 21: 2985–2991
- 69 Li Y, Hao P, Zhang S, et al. Feature-matching pattern-based support vector machines for robust peptide mass fingerprinting. *Mol Cell Proteomics*, 2011, 10: M110 005785
- 70 Pevtsov S, Fedulova I, Mirzaei H, et al. Performance evaluation of existing de novo sequencing algorithms. *J Proteome Res*, 2006, 5: 3018–3028
- 71 Chi H, Sun R X, Yang B, et al. Pnovo: de novo peptide sequencing and identification using hcd spectra. *J Proteome Res*, 2010, 9: 2713–2724
- 72 Chi H, Chen H, He K, et al. pNovo+: de novo peptide sequencing using complementary HCD and ETD tandem mass spectra. *J Proteome Res*, 2013, 12: 615–625
- 73 Zhao Y, Sun W, Zhang P, et al. Nematode sperm maturation triggered by protease involves sperm-secreted serine protease inhibitor (Serp). *Proc Natl Acad Sci USA*, 2012, 109: 1542–1547
- 74 Yang B, Wu Y J, Zhu M, et al. Identification of cross-linked peptides from complex samples. *Nat Methods*, 2012, 9: 904–906
- 75 Fu Y, Yang Q, Sun R, et al. Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics*, 2004, 20: 1948–1954
- 76 Zhang W, Zhang J, Xu C, et al. LFQuant: a label-free fast quantitative analysis tool for high-resolution LC-MS/MS proteomics data. *Proteomics*, 2012, 12: 3475–3484
- 77 Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, 2008, 26: 1367–1372
- 78 Tsou C C, Tsai C F, Tsui Y H, et al. IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. *Mol Cell Proteomics*, 2010, 9: 131–144
- 79 Wu S, Li N, Ma J, et al. First proteomic exploration of protein-encoding genes on chromosome 1 in human liver, stomach, and colon. *J Proteome Res*, 2013, 12: 67–80
- 80 Chang C, Zhang J, Han M, et al. SILVER: an efficient tool for stable isotope labeling LC-MS data quantitative analysis with quality control methods. *Bioinformatics*, 2014, 30: 586–587
- 81 Liu C, Song C Q, Yuan Z F, et al. Pquant improves quantitation by keeping out interfering signals and evaluating the accuracy of calculated ratios. *Anal Chem*, 2014, 86: 5286–5294
- 82 Xing X, Li Q, Sun H, et al. The discovery of novel protein-coding features in mouse genome based on mass spectrometry data. *Genomics*, 2011, 98: 343–351
- 83 Zheng G, Wang H, Wei C, et al. iGepros: an integrated gene and protein annotation server for biological nature exploration. *BMC Bioinformatics*, 2011, 12 Suppl 14: S6
- 84 Zhang Y, Yan G, Zhai L, et al. Proteome atlas of human chromosome 8 and its multiple 8p deficiencies in tumorigenesis of the stomach, colon, and liver. *J Proteome Res*, 2013, 12: 81–88
- 85 Wang Q, Wen B, Yan G, et al. Qualitative and quantitative expression status of the human chromosome 20 genes in cancer tissues and the representative cell lines. *J Proteome Res*, 2013, 12: 151–161

- 86 Guo F, Wang D, Liu Z, et al. Caper: a chromosome-assembled human proteome browser. *J Proteome Res*, 2013, 12: 179–186
- 87 Fu Y, Xiu L Y, Jia W, et al. Deltamt: a statistical algorithm for fast detection of protein modifications from LC-MS/MS data. *Mol Cell Proteomics*, 2011, 10: M110 000455
- 88 Jiang X, Ye M, Han G, et al. Classification filtering strategy to improve the coverage and sensitivity of phosphoproteome analysis. *Anal chem*, 2010, 82: 6168–6175
- 89 Zhu J, Wang F, Cheng K, et al. Analysis of human serum phosphopeptidome by a focused database searching strategy. *J Proteomics*, 2013, 78: 389–397
- 90 Wang Z, Ding G, Geistlinger L, et al. Evolution of protein phosphorylation for distinct functional modules in vertebrate genomes. *Mol Biol Evol*, 2011, 28: 1131–1140
- 91 Ren J, Jiang C, Gao X, et al. Phosnp for systematic analysis of genetic polymorphisms that influence protein phosphorylation. *Mol Cell Proteomics*, 2010, 9: 623–634
- 92 Liu Z, Guo F, Zhang J, et al. Proteome-wide prediction of self-interacting proteins based on multiple properties. *Mol Cell Proteomics*, 2013, 12: 1689–1700
- 93 Shao C, Li M, Li X, et al. dbDEPC 2.0: updated database of differentially expressed proteins in human cancers. *Nucleic Acids Res*, 2012, 40: D964–971
- 94 Shao C, Li M, Li X, et al. A tool for biomarker discovery in the urinary proteome: a manually curated human and animal urine protein biomarker database. *Mol Cell Proteomics*, 2011, 10: M111 010975
- 95 Gao T, Liu Z, Wang Y, et al. UUUCD: a family-based database of ubiquitin and ubiquitin-like conjugation. *Nucleic Acids Res*, 2013, 41: D445–D451
- 96 Liu Z, Cao J, Gao X, et al. Cpla 1.0: an integrated database of protein lysine acetylation. *Nucleic Acids Res*, 2011, 39: D1029–1034
- 97 Liu Z, Wang Y, Gao T, et al. Cplm: a database of protein lysine modifications. *Nucleic Acids Res*, 2014, 42: D531–536
- 98 Ren J, Liu Z, Gao X, et al. Microkit 3.0: an integrated database of midbody, centrosome and kinetochore. *Nucleic Acids Res*, 2010, 38: D155–160