

基于质谱的 SRM 相关策略和方法的研究进展*

常 乘¹⁾ 吴松锋¹⁾ 马 洁¹⁾ 张 伟²⁾ 朱云平^{1) **}

(¹⁾ 军事医学科学院放射与辐射医学研究所, 北京蛋白质组研究中心, 蛋白质组学国家重点实验室, 北京 102206)

2) 国防科学技术大学, 机电工程与自动化学院自动控制系, 湖南长沙 410073)

摘要 随着蛋白质组研究的不断深入, 基于质谱的选择反应监测技术 (SRM) 已经成为以发现生物标志物为代表的定向蛋白质组研究的重要手段. SRM 技术根据假设信息, 特异性地获取符合假设条件的质谱信号, 去除不符合条件的离子信号的干扰, 从而得到特定蛋白质的定量信息. SRM 技术具有更高的灵敏度和精确性、更大的动态范围等优势. 该技术可分为实验设计、数据获取和数据分析三个步骤. 在这几个步骤中, 最重要的是利用生物信息学手段总结当前实验数据的结果, 并用机器学习方法和总结的经验规则进行 SRM 实验的母离子和子离子的预测. 针对数据质控和定量的生物信息学方法研究在提高 SRM 数据可靠性方面具有重要作用. 此外, 为方便 SRM 的研究, 本文还收集、汇总了 SRM 技术相关的软件、工具和数据库资源. 随着质谱仪器的不断发展, 新的 SRM 实验策略以及分析方法、计算工具也应运而生. 结合更优化的实验策略、方法, 采用更精准的生物信息学算法和工具, SRM 在未来蛋白质组学的发展中将发挥更加重要的作用.

关键词 定向蛋白质组学, 生物质谱, 选择反应监测技术, 蛋白质特异性肽段, 母离子-子离子对

学科分类号 Q51, Q811.4

* 资助项目: 国家重点基础研究发展计划 2011CB910601, 2010CB912700; 国家高技术研究发展计划 2012AA020409, 2012AA020201; 国家自然科学基金 21105121.

**通讯联系人.

Tel: 010-80705225, E-mail: zhuyunping@gmail.com

收稿日期: 2012-01-06, 接受日期: 2012-03-15

质谱技术虽然早已出现,但在蛋白质组学研究领域^[1],直到上世纪九十年代生物质谱技术的快速发展带来了高灵敏度和高检测范围^[2],质谱技术才逐步成为蛋白质组研究中的主流技术.目前,高性能液相色谱技术联合串联质谱技术(HPLC-MS/MS)已成为蛋白质组领域最有效的研究手段,并且绝大多数蛋白质组研究都基于质谱技术^[3].

当前,基于质谱的蛋白质组研究方法中的主流技术是鸟枪法.然而,鸟枪法对低丰度肽段和蛋白质的检测能力不强,蛋白质组样品的高复杂性更使得鸟枪法在低丰度样品中的应用受到限制.定向蛋白质组学(targeted proteomics)采用假设驱动(hypothesis-driven)的研究策略,能很好地弥补鸟枪法的不足^[4].该策略可以用来测定样品蛋白质水平的相对量或者绝对量变化,尤其是检测样品中低丰度肽段和蛋白质的差异变化,多用于疾病标志物发现、差异蛋白质表达验证等方面.

基于质谱的选择反应监测技术^[5](selected reaction monitoring, SRM),或称单反应监测技术(single reaction monitoring),是定向蛋白质组研究中常用的方法^[6].SRM根据实验前的假设信息,对质谱数据进行过滤,即在母离子中选择符合假设条件的母离子,进行碎裂后形成子离子,在多个子离子中选出符合假设条件的子离子,通过两次选择,有效去除了噪声和干扰,从而提高了灵敏度.质谱多反应监测技术(multiple reaction monitoring, MRM)^[7]被认为是并行SRM技术,一般统称为SRM.

近年来已经有一些综述概述了SRM技术^[7-9],但是在SRM相关的方法和策略方面,尤其是生物信息学在SRM上的应用仍缺乏系统和详细的总结.本文在介绍了SRM的实验原理和特点以及SRM的实验流程后,重点详述了SRM实验各步骤中所使用的策略和方法,突出了生物信息学在其中的作用,列举分析了SRM相关的常用软件及数据库资源,最后总结了SRM的最新进展及目前的不足,并展望了未来可能的发展方向.

1 SRM 的实验原理与流程

SRM技术根据待分析的目标蛋白质的氨基酸序列选择对应的蛋白质特异性肽段(ProteoTypic Peptide, PTP)^[10]及蛋白质特异性肽段对应的高响应子离子,最后用高响应子离子的信号强度来反映该蛋白质的表达丰度.这些信号强度大且能够确定唯一目标蛋白质的肽段被称为蛋白质特异性肽段,并与对应的高响应子离子组成一个母离子-子离子对(transition)^[7].SRM实验通常基于三重四级杆(triple quadrupole, QQQ)串联质谱仪,也可以在线性离子阱(Linear Ion Trap)质谱仪或者四级杆-飞行时间(Q-TOF)质谱仪上进行.基于三重四级杆质谱仪的SRM实验原理如图1所示.

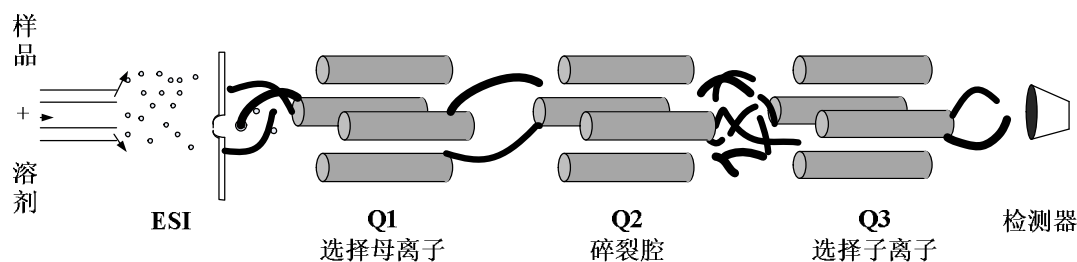


图1 质谱选择反应监测技术的原理图

Fig. 1 The theory of a typical SRM experiment

在三个串联的四级杆中,第一个四级杆(Q1)用来过滤母离子,只有预先被选中的母离子才能通过Q1,进入第二个四级杆(Q2).Q2作为碎裂腔,对过滤后的母离子进行碎裂,通常是碰撞诱导碎裂模式(Collision Induced Dissociation, CID).但de Graaf等人^[11]在刚刚发表的研究中使用高能碰撞诱导碎裂模式(High Energy CID, HCD)进行SRM实验,在灵敏度上比传统的CID碎裂模式更好.第三个四级杆(Q3)用来过滤子离子,最后只有符合预先设定条件的子离子才能通过Q3,进入到检测器中.

通常来说,一个典型的SRM实验包括三步:实验设计、数据获取、数据分析.其中,实验设计是指目

标蛋白质对应的蛋白质特异性肽段的选取, 以及相应的母离子-子离子对的选取. 数据获取是指设置好质谱仪的相应参数, 获取 SRM 质谱数据. 最后, 数据分析包括对 SRM 质谱数据的验证, 以及定性、定量分析.

相对于传统质谱仪的全扫描方式, SRM 只对符合预先设定条件的蛋白质特异性肽段和子离子进行扫描, 因此 SRM 具有更高的检测范围和灵敏度, 更适用于对复杂样品混合物的检测, 尤其是对低丰度肽段和蛋白质的检测. 而在重复性和准确性方面, SRM 由于采用了两级过滤策略^[12], 有效降低了背景噪声以及干扰, 因此比传统的质谱技术更具优势. 总的来说, SRM 技术具有更高的灵敏度和更好的准确性、重复性等特点. 目前, 结合相应的质谱技术, SRM 在蛋白质组学领域已经得到了广泛应用. Addona 等人^[13]的研究就证明了基于稳定同位素稀释 (stable isotope dilution, SID) 及质谱多反应监测技术结合串联质谱 (SID-MRM-MS) 的策略在不同实验仪器、不同平台、不同实验室之间具有良好的重复性和灵敏性, 表明 SRM 技术可以应用于临床大规模样品分析中.

2 SRM 实验步骤中的分析策略及方法

2.1 实验设计

实验设计是 SRM 实验的关键步骤. 在 SRM 实验中, 实验设计的好坏决定了实验质量的高低. 实验设计具体是指三个方面: (1) 待研究蛋白质对应的蛋白质特异性肽段预测; (2) 蛋白质特异性肽段的母离子-子离子对预测; (3) 蛋白质特异性肽段和母离子-子离子对的确定.

2.1.1 蛋白质特异性肽段的预测

研究者根据实验要求的精度和动态范围, 根据已有实验数据、相关文献的归纳总结, 或者根据相关数据库资源首先选择出研究感兴趣的蛋白质集合. 在确定好目标蛋白质集合后, 对于每个目标蛋白质, 研究人员需要选择相应的特异性肽段. 蛋白质特异性肽段是指那些能确定唯一目标蛋白质, 同时具有较强信号强度的肽段. 一个目标蛋白质被酶切后可以得到数十至数百个肽段, 其中可能有一个或者多个特异性肽段可用来进行 SRM 实验^[3]. 因此, 选择合适的蛋白质特异性肽段对于 SRM 实验至关重要.

目前, 蛋白质特异性肽段的预测有两大策略^[9]: 一是根据已有实验数据进行预测, 二是根据计算工具进行预测. 第一类根据已有实验数据预测蛋白质特异性肽段的策略又可以细分为两类: 一种是直接从已有实验数据中总结出蛋白质特异性肽段, 这也是最简单、直接的方法^[14-15]; 另一种是根据相关数据库进行蛋白质特异性肽段的预测. 比如以 Ruedi Aebersold 为主的研究团队整合了不同实验平台、不同搜库软件的数据, 选取高可信的肽段和蛋白质鉴定结果, 构建了一系列数据库: PeptideAtlas^[16-18]、SRMAtlas^[19]、MRMAtlas^[19]. 这些数据库存储了以往实验产生的高可信的蛋白质特异性肽段以及其相应的母离子-子离子对, 可供研究人员免费下载. 类似的工作还有 GPM^[20]、MRMaid-DB^[21]、PRIDE^[22]等数据库. 然而, 根据实验数据预测蛋白质特异性肽段的策略存在明显的局限性, 即只能针对之前实验检测到的蛋白质和肽段. 如果待检测的蛋白质和肽段没有在数据库中出现过, 那么这种利用已有数据预测蛋白质特异性肽段的方法就无能为力了.

于是研究人员又发展出第二类预测蛋白质特异性肽段的策略: 根据生物信息学的计算工具预测蛋白质特异性肽段. 该类方法弥补了第一类策略的缺点. 一般来说, 不管实验中待测的肽段和蛋白质之前是否被检测过, 研究者都可以用计算工具来预测每个蛋白质的特异性肽段. 该策略是利用肽段的理化性质, 在已有的大规模数据集中挑选高可信的结果进行分析, 根据肽段序列、样品预处理流程、肽段的丰度等多方面的综合信息, 在蛋白质的众多理论酶切肽段中区分出蛋白质特异性肽段.

当前对这些肽段进行分析预测的数学模型及其应用包括如下几类:

- 1) Sanders 等人^[23]运用了神经网络的方法推出了预测蛋白质特异性肽段的工具 PepFly, 在两个大数据集上进行十倍交叉验证, 精确度可达 80%左右.
- 2) Mallick 等人^[10]使用了分级爬山搜寻法 (hierarchical hill-climbing search approach), 对于每个特征在训练集中建立了高斯混合判别函数, 利用四个大规模数据集产生的高可信肽段, 生成了分类器来预测蛋白质特异性肽段, 精确度达到了 85%以上.
- 3) Webb-Robertson 等人^[24]首次利用支持向量机 (support vector machine, SVM) 的方法, 基于 35 种肽段理化性质, 生成了预测工具 STEPP, 用来在蛋白的理论酶切肽段集中预测蛋白质特异性肽段.
- 4) Fusaro 等人^[25]考虑了 550 种肽段理化性质^[26], 使用了随机森林分类器进行聚类分析, 并推出了预测蛋白质特异性肽段的工具 ESP Predictor, 已整合在了 GenePattern^[27]中.

2. 1. 2 母离子-子离子对中的子离子预测

母离子-子离子对中的子离子的选择与 SRM 质谱数据的碎裂模式有关. 三重四级杆中的 CID 碎裂模式通常只产生较少的 b 离子和二价离子, 这些离子的信号比较弱, 而 CID 碎裂更倾向于产生 y 离子^[28], 生成的 y 离子信号也比较强. 因此在一般的 CID 碎裂模式中, 选择母离子-子离子对时可以多考虑 y 离子. 根据子离子的质核比一般都在整个质核比范围的中上部等经验规则^[9], 研究人员通常不考虑在肽段 N 端和 C 端的 y 离子, 以及和母离子质核比太接近的 y 离子^[29]. Sherwood 等人^[30]的研究表明肽段序列对产生的 y 离子信号强度有显著影响, 而这种影响与仪器类型无关. 他们的研究还表明任何离子阱产生的母离子-子离子对都可以直接用来进行基于三重四级杆的 MRM 实验.

结合上述的子离子选择策略, 研究者通常使用生物信息学计算工具来预测母离子-子离子对中的子离子. 首先, 各个质谱仪生产商都推出了各自基于 SRM 实验的商业软件, 比如 Applied Biosystems 公司的 MIDAS、MRMPilot、MultiQuant, Thermo Scientific 公司的 PinPoint 等. 其次, 文献报道的预测方法与工具也有不少. 有些工具是根据 SRM 相关数据库资源进行预测的, 比如 Lange 等人^[31]推出的 TIQAM, TIQAM 调用了 PeptideAtlas 的数据进行蛋白质特异性肽段和母离子-子离子对预测. Mead 等人^[12]则是利用 GAPP^[32]数据库的相关数据资源推出了一个基于网页的预测母离子-子离子对的工具 MRMAid, 并能结合保留时间来校正预测的母离子-子离子对. 之后 Cham 等人^[21]在此基础上又进一步推出 MRMAid-DB, 一个专门提供蛋白的母离子-子离子对的数据库. 还有一些预测工具不依赖于数据库资源, 仅仅根据蛋白质和肽段的理化特点进行预测, 比如 Bertsch 等人^[33]的研究使用从头测序法 (de novo), 仅利用蛋白序列信息来预测母离子-子离子对, 适用于小规模 SRM 实验设计.

2. 1. 3 蛋白质特异性肽段及母离子-子离子对的确定

经过实验前对蛋白质特异性肽段及母离子-子离子对的预测, 研究人员往往还要通过实验数据来最后确定目标蛋白质的蛋白质特异性肽段及相应的母离子-子离子对, 这样尽可能的避免了预测带来的假阳性, 提高了 SRM 实验的准确性, 同时也节约了实验成本, 提高了实验效率. 研究者既可以根据原有数据及鉴定结果, 在预测的蛋白特异性肽段及母离子-子离子对中选择信号最强的那些结果, 也可以针对待研究的目标蛋白质, 用标准样品做预实验, 然后根据预实验的结果来确定最终的蛋白特异性肽段和母离子-子离子对.

2. 2 数据获取

当目标蛋白集和对应的蛋白质特异性肽段与母离子-子离子对都确定好后, 下一步就是 SRM 实验. 每个蛋白质特异性肽段一般选择 2-4 个母离子-子离子对. 实验中能够有效检测的母离子-子离子对数目是由每个母离子-子离子对的滞留时间 (dwell time, 通常为 5-40ms) 与色谱峰的典型宽度之间的关系决定的^[34]. 由于 SRM 实验中质谱仪的循环时间 (cycle time) 是实验前设置好的, 因此单次实验分析的母离子-子离子对数目有限. 单次实验分析的母离子-子离子对数目越多, 每个母离子-子离子对的滞留时间就越少. 仪器的灵敏度、检测范围和检测通量都因此受到了限制. 所以研究者往往会折中考虑各个参数的设置, 期望实验效

果达到最佳. Stahl-Zeng 等人^[35]在研究中提出了 scheduled SRM 策略, 增加了单次检测的母离子-子离子对数目, 在保持原有灵敏度的情况下, 提高了检测范围和检测通量. Picotti 等人^[36]进一步提出了 time-scheduled SRM 策略, 并用于酵母数据的分析中, 能够很好地鉴定到低丰度肽段和蛋白质.

最近, Kiyonami 等人^[37]提出了 iSRM (intelligent SRM) 策略, 将一个蛋白质特异性肽段的所有母离子-子离子对分为定性和定量两类母离子-子离子对, 挑选信号强度最高的几个母离子-子离子对进行定量分析, 剩下的母离子-子离子对进行定性验证, 进一步提高了 SRM 实验的效率和通量.

2.3 数据分析

2.3.1 数据验证

在一些复杂的样品中, 经过 SRM 实验的两级过滤后可能仍会在结果中混有错误的母离子-子离子对. 此外, 样品在洗脱时或在质谱仪中可能会出现丢失, 这些情况都会造成一定程度上假阳性结果的产生. 因此对于 SRM 实验数据有必要进行验证, 从而确保分析的质量.

在验证母离子-子离子对时, 又可按是否使用二级谱图分为两种方法. 在使用二级谱图的验证方法中, 首先获得一个蛋白特异性肽段所对应的全部母离子-子离子对. 然后获取二级谱图, 结合搜库鉴定结果对母离子-子离子对进行验证. Unwin 等人^[38]提出了一种改进方法, 每次只选择超过设定阈值的母离子触发质谱仪获取二级谱, 验证肽段是否选择正确. 这种触发方式比数据依赖分析模式^[31] (data-dependent acquisition mode) 更有效率, 并且灵敏度和特异性也得到增加. 还有一种不需要获取二级谱图的验证方法, 即同位素标记法^[39]. 在样品进质谱仪前加入同位素重标的肽段后, 研究者可以在重构的离子流色谱峰 (extracted ion chromatogram, XIC) 中查看同一个肽段对应的母离子-子离子对是否共洗脱. 这种方法对低丰度的肽段检测也很灵敏. 比如 Picotti 等人^[40]采用了 SPOT-synthesis^[41]技术合成定量肽段, 并用此肽段进行 SRM 检测的验证和参数的优化, 能有效地提高 SRM 对于复杂样本和低丰度样本的检测能力, 达到高通量的目的.

Reiter 等人^[42]推出了 MRM 数据分析工具 mProphet, 在 SRM 数据分析中引入了质量控制的概念, 提出了对母离子-子离子对进行质量控制, 使用新的诱饵母离子-子离子对 (decoy transition) 构建方法, 联合多种特性组成一个综合打分, 用来对母离子-子离子对进行验证. 而 Brusniak 等人^[43]开发了多用户的 SRM 数据优化和验证工具 ATAQS, 使用并改进了 SIMPLE (Semi-Implicit Method for Pressure Linked Equations) 算法, 用于生成诱饵母离子-子离子对, 能够更方便的对母离子-子离子对进行验证.

除了对质谱数据中实际得到的母离子-子离子对进行验证外, 还有一些研究者关注于母离子-子离子对的预测准确性问题. Abbatiello 等人^[44]就此推出工具 AuDIT, 验证预测得到的母离子-子离子对, 将认为预测不准的母离子-子离子对去除. AuDIT 使用了两种互不相关的方法来客观估计 MRM-MS 数据. 第一种是比较蛋白特异性肽段碎片离子与内标肽段碎片离子的相对强度, 使用 t-检验来确定它们是否有显著性差异. 第二种是计算蛋白特异性肽段碎片离子与内标肽段碎片离子的相对强度比值的变异系数 (coefficient of variation, CV).

2.3.2 定量分析

SRM 的定量分析从方法上讲, 一般分为相对定量分析和绝对定量分析. SRM 实验的相对定量一般是指同一个肽段在实验组与对照组的信号强度比作为其相对定量值, 可以用目前常用的无标记 (label-free) 定量算法^[45-47]或者标记定量算法^[48-49]来进行计算. 在无标记定量算法方面, Eissler 等人^[50]发展了一种基于无标记算法的 SRM 定量方法, 并应用于磷酸化的研究. Tang 等人^[51]的研究表明肽段的检测性 (detectability) 大小——一个肽段成为蛋白质特异性肽段的概率大小——还可以用来对蛋白质丰度进行定量, 并由此提出了一种新的无标定量方法. 另一方面, 若在实验中引入了同位素标记的肽段, 那么可以直接计算重标和轻标的肽段信号强度比作为结果. 目前常用的标记方法有 ¹⁸O、SILAC^[52]、iCAT^[53]、iTRAQ^[54]等方法. Wolf-Yadlin 等人^[55]在 MRM 实验中结合了 iTRAQ 标记肽段, 成功的对一个细胞信号网络进行了定量分析, 约 88% 的网络节点在四次重复实验中均能被定量分析, 而传统的信息依赖分析模式^[56] (IDA,

information-dependent analysis) 只能定量到 34%左右的节点.

结合同位素标记的肽段, SRM 实验也可以进行绝对定量. 基于稳定同位素稀释(stable isotope dilution, SID)或者标记参考肽段(labeled reference peptide)的 SRM-MS 策略能做到对目标蛋白质及肽段的绝对定量^[57]. 绝对定量要求同位素标记的肽段上样量已知. 同时, 参考肽段和实验环境的选择都是在绝对定量时需要考虑的因素^[6]. 和相对定量相比, 绝对定量对实验条件以及质量控制的要求比较高. 近来, 研究人员开发了一些根据内标肽段进行定量分析的 SRM 数据分析工具. 如 Martin 等人^[34]推出了 SRM 数据处理软件 MRMer, 根据 SILAC 标记、AQUA^[58-59]策略进行肽段的精确的绝对定量和相对定量. Hammad 等人^[60]使用内部肽段, 发展了一种新的 SRM 绝对定量策略, 用于果蝇的研究.

3 SRM 实验相关软件和资源

随着 SRM 相关研究的不断增多和 SRM 技术的逐步发展, 有越来越多的 SRM 软件和数据库等网络资源和数据资源被研究者所用.

表 1 列举了目前一些常用的 SRM 实验设计(包括预测蛋白质特异性肽段以及母离子-子离子对)和数据分析软件、软件对应的网络资源以及参考文献. 其中不少软件都是提供免费下载的, 有一些还公开了源代码.

表 2 列举了 SRM 实验设计相关的数据库资源及对应的介绍和参考文献.

表 1 SRM 实验预测、分析常用软件

Table1. The common softwares for SRM experiments

软件名称	网址	文献	软件功能	备注
ESP Predictor	http://www.genepattern.org	[25]	a	B/S 模式, 整合在 GenePattern 中
STEPP	http://omics.pnl.gov/software/STEPP.php	[24]	a	C/S 模式
PepFly	http://www.agbase.msstate.edu/cgi-bin/tool/s/index.cgi	[23]	a	命令行模式, 开源, 基于 Sequest 结果
MaRiMba	http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP-MaRiMba	[61]	b	B/S 模式, 开源, 整合在 TPP ^[62] 中
MRMaid	http://www.mrmaid.info/	[12, 21]	b	B/S 模式, 整合在 GAPP 中
AuDIT	http://www.genepattern.org	[44]	b	B/S 模式, 整合在 GenePattern 中
mProphet	http://www.mprophet.org	[42]	c	命令行模式
MIDAS	http://www.appliedbiosystems.com	-	a,b,c,d	C/S 模式, AB 公司商业软件
PinPoint	http://www.thermo.com/pinpoint	-	a,b,c,d	C/S 模式, Thermo 公司商业软件
MRMPilot	http://www.absciex.com/Products/Software/MRMPilot-Software	-	a,b,c,d	C/S 模式, Waters 公司商业软件
TIQAM	-	[31]	a,b,c,d	-
Skyline	http://proteome.gs.washington.edu/software/skyline	[63]	a,b,c,d	C/S 模式
ATAQS	http://tools.proteomecenter.org/ATAQS/ATAQS.html	[43]	a,b,c,d	B/S 模式
MRMer	-	[34]	a,b,c,d	-
jTraML	http://iomics.ugent.be/jtraml	[64]	SRM 数据格式转换	B/S 模式

备注:

1. B/S 模式是指浏览器/服务器模式 (Browser/Server), 用户不需安装客户端, 直接通过本地浏览器与服务器连接进行操作; C/S 模式是指客户端/服务器模式 (Client/Server), 用户需要先下载、安装客户端到本地, 然后在本地使用客户端连接服务器.

2. TIQAM 和 MRMer 在文章中并未直接给出下载方法和地址, 需要与作者联系.
3. 软件功能一列中: a 是指蛋白质特异性肽段预测, b 是指母离子-子离子对预测, c 是指数据验证, d 是指数据定量分析.

表 2 SRM 实验设计常用数据库

Table2. The common databases for SRM experiment design

数据库	网址	文献	备注
PeptideAtlas	http://www.peptideatlas.org/	[16-18]	蛋白质组综合数据库
MRMAtlas	http://www.srmatlas.org/original.php	[19]	现已被 SRMAtlas 替代
SRMAtlas	http://www.srmatlas.org/	[19]	SRM 数据库
GPM	http://www.thegpm.org/	[20]	蛋白质组综合数据库
MRMaid-DB	http://mo-box.ccc.cranfield.ac.uk:5555/mrmdb_home.html	[21]	SRM 数据库
PRIDE	http://www.ebi.ac.uk/pride	[22]	蛋白质组综合数据库
ProteomeCommons	https://www.proteomecommons.org/	[65-66]	蛋白质组综合数据库
GAPP	http://www.gapp.info/	[32]	蛋白质组数据分析平台
SRM ORS	https://www-s.nist.gov/srmors/	-	NIST 的 SRM 数据库

在表 1 列举的众多 SRM 相关软件中, ESP Predictor、PepFly、STEPP 是蛋白质特异性肽段预测的工具. STEPP 还加入了蛋白理论酶切功能, 用户可以直接载入蛋白序列库, 使用方便. mProphet 和 AuDIT 用来验证蛋白质特异性肽段和母离子-子离子对. MRMaid、MaRiMba 专门用来预测母离子-子离子对, 并且 MaRiMba 已经整合在了 TPP 中, 使用方便. 而 TIQAM 和 ATAQS、Skyline、MRMer 都提供了 SRM 实验的完整流程, 不仅包括了实验设计, 还有后面的数据验证、分析. Skyline 通过计算 SRM 实验数据与谱图库的子离子信号相对强度之间的点积来评价实验数据的质量, 该方法需要先提供谱图库的信息, 缺点是没有明确给出数据的错误发现率 (False Discovery Rate, FDR). MIDAS、MRMPilot、PinPoint 是质谱仪生产商的商业软件, 提供了 SRM 实验的完整分析流程, 但只提供给购买质谱仪的客户.

与此同时, 如表 2 所示, SRM 相关的数据库资源也越来越多, 越来越完善. 这些数据库功能各异, 有些是专门的 SRM 数据库如 SRMAtlas、MRMAtlas (现在已经合并到 SRMAtlas 中)、MRMaid-DB. 这些数据库不仅提供了之前实验用的蛋白质特异性肽段和母离子-子离子对, 还可以对蛋白质的功能进行进一步分析.

其中还有不少是综合的蛋白质组数据库, 列举如下:

1. PeptideAtlas 是一个公开的蛋白质组学数据库, 存储多个物种的串联质谱数据. 该数据库收集人类、小鼠、酵母和一些其他物种的质谱数据, 使用最新的搜索引擎和蛋白质序列数据库鉴定实验产生的数据, 内容丰富而全面.
2. PRIDE (PRoteomics IDentifications) 数据库是一个公开的拥有自身标准的蛋白质组学数据库. PRIDE 能提供蛋白质鉴定和肽段鉴定的相关证据, 同时提供与肽段翻译后修饰鉴定相关的信息.
3. ProteomeCommons 是一个公开的质谱数据综合平台, 使用 Tranche^[66]进行数据的快速上传与下载, 同时还提供了一些开源的数据分析工具. 其最大的特点是免费提供了很多原始质谱文件.
4. GPM (Global Proteome Machine) 是一个综合蛋白质数据库, 包含了人类、小鼠等哺乳动物以及其他一些生物的蛋白质组数据.
5. GAPP (The Genome Annotating Proteomic Pipeline) 是一个基于网络的蛋白质组质谱数据分析平台, 存储了质谱相关的蛋白质组数据, 并且提供了相应的分析工具, 其中包括 MRMaid, 可以用来产生预测的母离子-子离子对.
6. 美国标准技术研究所 (National Institute of Standards and Technology, NIST) 也提供了一些 SRM 的

相关数据资源, 如 SRM Order Request System (SRM ORS) .

然而, 由于各个质谱仪生产商定义的 SRM 数据格式不同, 目前不少数据库中存储的 SRM 数据格式各不相同, 为数据分析带来了不便. 为此, 质谱的蛋白质组标准计划工作组(the Mass Spectrometry working group of the Proteomics Standards Initiative , PSI-MS) 为 SRM 数据设定了一个标准格式 TraML (Transition Markup Language) ^[67], 使得 SRM 数据的共享、分析更加方便. Helsens 等人^[64]刚刚推出了基于 Java 的格式转换工具 jTraML, 可以用于各种数据格式与标准格式 TraML 之间的互相转换.

4 总结与展望

本文介绍了基于质谱的选择反应监测技术的背景以及实验原理和特点, 总结了该技术的三大实验步骤(实验设计、数据获取、数据分析), 详细分析了每个步骤中的分析策略和方法, 最后汇总了 SRM 实验中常用的软件、数据库以及其他相关网络资源.

目前, SRM 为代表的定向蛋白质组研究越来越多, SRM 因为其天生的高准确性、高灵敏性、大动态范围以及在低丰度肽段和蛋白质鉴定、定量方面的优势而被越来越多的研究者所采用. 但是目前 SRM 技术的通量并不高, 检测效率低而且灵敏度和动态范围还有待进一步提高, 以更好地适应复杂蛋白质组样品的研究要求.

结合不断发展的样品制备、预处理方法, 同位素内标的合成肽段, 以及更精准的生物信息学计算方法及工具, SRM 实验将在生物标志物的发现和验证方面发挥更大的作用^[68-69]. 同时, SRM 也将在翻译后修饰的发现、蛋白相互作用网络的验证、新蛋白的发现等众多方面都将发挥重要作用. 目前, 结合 SRM 技术的研究新方法、新策略也层出不穷^[70-71]. 可以预见, 今后 SRM 将会在蛋白质组的研究中占据重要的地位.

致谢 感谢北京蛋白质组研究中心的应万涛博士及江静同学的讨论和建议, 感谢北京蛋白质组研究中心的生物信息组全体成员的支持和帮助.

参 考 文 献

1. 胡志远, 贺福初. 蛋白质组研究进展. 生物化学与生物物理进展, 1999, 26(3):202-205
Hu Z Y, He F C. Progress in Proteome Research. Progress in Biochemistry and Biophysics, 1999, 26(3): 202-205
2. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature, 2003, 422(6928):198-207
3. Kuster B, Schirle M, Mallick P, *et al.* Scoring proteomes with proteotypic peptide probes. Nat Rev Mol Cell Biol, 2005, 6(7):577-583
4. Domon B, Aebersold R. Options and considerations when selecting a quantitative proteomics strategy. Nat Biotechnol, 2010, 28(7):710-721
5. Baty JD, Robinson PR. Single and multiple ion recording techniques for the analysis of diphenylhydantoin and its major metabolite in plasma. Biomed Mass Spectrom, 1977, 4(1):36-41
6. Bronstrup M. Absolute quantification strategies in proteomics based on mass spectrometry. Expert Rev Proteomics, 2004, 1(4):503-512
7. Lange V, Picotti P, Domon B, *et al.* Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol, 2008, 4:222
8. 赵焱, 应万涛, 钱小红. 质谱 MRM 技术在蛋白质组学研究中的应用. 生命的化学, 2008, 28(2):210-213
Zhao Y, Ying W T, Qian X H. Application of multiple reaction monitoring in proteome research. Chemistry of Life, 2008, 28(2):210-213
9. Calvo E, Camafeita E, Fernandez-Gutierrez B, *et al.* Applying selected reaction monitoring to targeted

- proteomics. *Expert Rev Proteomics*,2011, 8(2):165-173
10. Mallick P, Schirle M, Chen SS, *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol*,2007, 25(1):125-131
11. de Graaf E, Altelaar M, van Breukelen B, *et al.* Improving SRM Assay Development: A Global Comparison between Triple Quadrupole, Ion Trap and Higher Energy CID Peptide Fragmentation Spectra. *J Proteome Res*,2011
12. Mead JA, Bianco L, Ottone V, *et al.* MRMAid, the web-based tool for designing multiple reaction monitoring (MRM) transitions. *Mol Cell Proteomics*,2009, 8(4):696-705
13. Addona TA, Abbatiello SE, Schilling B, *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol*,2009, 27(7):633-641
14. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics*,2006, 5(4):573-588
15. Keshishian H, Addona T, Burgess M, *et al.* Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics*,2007, 6(12):2212-2229
16. Desiere F, Deutsch EW, King NL, *et al.* The PeptideAtlas project. *Nucleic Acids Res*,2006, 34(Database issue):D655-658
17. Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep*,2008, 9(5):429-434
18. Deutsch EW, Eng JK, Zhang H, *et al.* Human Plasma PeptideAtlas. *Proteomics*,2005, 5(13):3497-3500
19. Picotti P, Lam H, Campbell D, *et al.* A database of mass spectrometric assays for the yeast proteome. *Nat Methods*,2008, 5(11):913-914
20. Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res*,2004, 3(6):1234-1242
21. Cham JA, Bianco L, Barton C, *et al.* MRMAid-DB: a repository of published SRM transitions. *J Proteome Res*,2010, 9(1):620-625
22. Jones P, Cote RG, Martens L, *et al.* PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res*,2006, 34(Database issue):D659-663
23. Sanders WS, Bridges SM, McCarthy FM, *et al.* Prediction of peptides observable by mass spectrometry applied at the experimental set level. *BMC Bioinformatics*,2007, 8 Suppl 7:S23
24. Webb-Robertson BJ, Cannon WR, Oehmen CS, *et al.* A support vector machine model for the prediction of proteotypic peptides for accurate mass and time proteomics. *Bioinformatics*,2008, 24(13):1503-1509
25. Fusaro VA, Mani DR, Mesirov JP, *et al.* Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat Biotechnol*,2009, 27(2):190-198
26. Kawashima S, Kanehisa M. AAindex: amino acid index database. *Nucleic Acids Res*,2000, 28(1):374
27. Reich M, Liefeld T, Gould J, *et al.* GenePattern 2.0. *Nat Genet*,2006, 38(5):500-501
28. Paizs B, Suhai S. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev*,2005, 24(4):508-548
29. Sandhu C, Hewel JA, Badis G, *et al.* Evaluation of data-dependent versus targeted shotgun proteomic approaches for monitoring transcription factor expression in breast cancer. *J Proteome Res*,2008, 7(4):1529-1541
30. Sherwood CA, Eastham A, Lee LW, *et al.* Correlation between y-type ions observed in ion trap and triple quadrupole mass spectrometers. *J Proteome Res*,2009, 8(9):4243-4251
31. Lange V, Malmstrom JA, Didion J, *et al.* Targeted quantitative analysis of *Streptococcus pyogenes* virulence factors by multiple reaction monitoring. *Mol Cell Proteomics*,2008, 7(8):1489-1500
32. Shadforth I, Xu W, Crowther D, *et al.* GAPP: a fully automated software for the confident identification of human peptides from tandem mass spectra. *J Proteome Res*,2006, 5(10):2849-2852
33. Bertsch A, Jung S, Zerck A, *et al.* Optimal de novo design of MRM experiments for rapid assay development in targeted proteomics. *J Proteome Res*,2010, 9(5):2696-2704
34. Martin DB, Holzman T, May D, *et al.* MRMer, an interactive open source and cross-platform system for data extraction and visualization of multiple reaction monitoring experiments. *Mol Cell Proteomics*,2008, 7(11):2270-2278
35. Stahl-Zeng J, Lange V, Ossola R, *et al.* High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics*,2007, 6(10):1809-1817
36. Picotti P, Bodenmiller B, Mueller LN, *et al.* Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell*,2009, 138(4):795-806

37. Kiyonami R, Schoen A, Prakash A, *et al.* Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol Cell Proteomics*, 2011, 10(2):M110 002931
38. Unwin RD, Griffiths JR, Leverentz MK, *et al.* Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics*, 2005, 4(8):1134-1144
39. Zhang H, Liu Q, Zimmerman LJ, *et al.* Methods for Peptide and protein quantitation by liquid chromatography-multiple reaction monitoring mass spectrometry. *Mol Cell Proteomics*, 2011, 10(6):M110 006593
40. Picotti P, Rinner O, Stallmach R, *et al.* High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods*, 2010, 7(1):43-46
41. Frank R, Overwin H. SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. *Methods Mol Biol*, 1996, 66:149-169
42. Reiter L, Rinner O, Picotti P, *et al.* mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods*, 2011, 8(5):430-435
43. Brusniak MY, Kwok ST, Christiansen M, *et al.* ATAQS: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. *BMC Bioinformatics*, 2011, 12:78
44. Abbatiello SE, Mani DR, Keshishian H, *et al.* Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin Chem*, 2010, 56(2):291-305
45. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol*, 2010, 2010:840518
46. Wong JW, Cagney G. An overview of label-free quantitation methods in proteomics by mass spectrometry. *Methods Mol Biol*, 2010, 604:273-283
47. 张伟, 张纪阳, 刘辉, *et al.* 蛋白质质谱分析的无标记定量算法研究进展. *生物化学与生物物理进展*, 2011, 38(6):506-518
Zhang W, Zhang J Y, Liu H, *et al.* Development of Algorithms for Mass Spectrometry-based Label-free Quantitative Proteomics. *Progress in Biochemistry and Biophysics*, 2011, 38(6):506-518
48. Colaert N, Vandekerckhove J, Martens L, *et al.* A case study on the comparison of different software tools for automated quantification of peptides. *Methods Mol Biol*, 2011, 753:373-398
49. Becker CH, Bern M. Recent developments in quantitative proteomics. *Mutat Res*, 2011, 722(2):171-182
50. Eissler CL, Bremmer SC, Martinez JS, *et al.* A general strategy for studying multisite protein phosphorylation using label-free selected reaction monitoring mass spectrometry. *Anal Biochem*, 2011
51. Tang H, Arnold RJ, Alves P, *et al.* A computational approach toward label-free protein quantification using predicted peptide detectability. *Bioinformatics*, 2006, 22(14):e481-488
52. Ong SE, Blagoev B, Kratchmarova I, *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*, 2002, 1(5):376-386
53. Gygi SP, Rist B, Gerber SA, *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*, 1999, 17(10):994-999
54. Ross PL, Huang YN, Marchese JN, *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 2004, 3(12):1154-1169
55. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, *et al.* Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci U S A*, 2007, 104(14):5860-5865
56. Huang J, Bathena SP, Alnouti Y. Metabolite profiling of praziquantel and its analogs during the analysis of in vitro metabolic stability using information-dependent acquisition on a hybrid triple quadrupole linear ion trap mass spectrometer. *Drug Metab Pharmacokinet*, 2010, 25(5):487-499
57. Zhao Y, Jia W, Sun W, *et al.* Combination of improved (18)O incorporation and multiple reaction monitoring: a universal strategy for absolute quantitative verification of serum candidate biomarkers of liver cancer. *J Proteome Res*, 2010, 9(6):3319-3327
58. Kirkpatrick DS, Gerber SA, Gygi SP. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods*, 2005, 35(3):265-273
59. Gerber SA, Rush J, Stemman O, *et al.* Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A*, 2003, 100(12):6940-6945
60. Hammad LA, Cooper BS, Fisher NP, *et al.* Profiling and quantification of *Drosophila melanogaster* lipids using liquid chromatography/mass spectrometry. *Rapid Commun Mass Spectrom*, 2011, 25(19):2959-2968
61. Sherwood CA, Eastham A, Lee LW, *et al.* MaRiMba: a software application for spectral library-based MRM transition list assembly. *J Proteome Res*, 2009, 8(10):4396-4405

62. Deutsch EW, Mendoza L, Shteynberg D, *et al.* A guided tour of the Trans-Proteomic Pipeline. *Proteomics*, 2010, 10(6):1150-1159
63. MacLean B, Tomazela DM, Shulman N, *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, 2010, 26(7):966-968
64. Helsen K, Brusniak MY, Deutsch E, *et al.* jTraML: An open source java API for TraML, the PSI standard for sharing SRM transitions. *J Proteome Res*, 2011
65. Hill JA, Smith BE, Papoulias PG, *et al.* ProteomeCommons.org collaborative annotation and project management resource integrated with the Tranche repository. *J Proteome Res*, 2010, 9(6):2809-2811
66. Smith BE, Hill JA, Gjukich MA, *et al.* Tranche distributed repository and ProteomeCommons.org. *Methods Mol Biol*, 2011, 696:123-145
67. Deutsch E, Chambers M, Bertsch A, *et al.* TraML - a standard format for exchange of selected reaction monitoring transition lists. 2010
68. 廖明, 李彦, 舒伟. 质谱 MRM 技术在生物标志物研究中的应用. *分子诊断与治疗杂志*, 2010, 2(4):275-279
Liao M, Li Y, Shu W. Application of multiple reaction monitoring in biomarkers research. *J Mol Diagn Ther*, July 2010, 2(4):275-279
69. 韩冰, 王立顺. 基于 MRM-MS 技术的蛋白质生物标志物研究的前景. *中华检验医学杂志*, 2011, 34(6):484-488
Han B, Wang L S. Strategies for novel protein biomarker identification through multiple reaction monitoring mass spectrometry. *Chin J Lab Med*, 2011, 34(6):484-488
70. Ortea I, Canas B, Gallardo JM. Selected tandem mass spectrometry ion monitoring for the fast identification of seafood species. *J Chromatogr A*, 2011, 1218(28):4445-4451
71. Ramanathan R, Jemal M, Ramagiri S, *et al.* It is time for a paradigm shift in drug discovery bioanalysis: from SRM to HRMS. *J Mass Spectrom*, 2011, 46(6):595-601

Methods and Progress of Mass Spectrometry-based Selected Reaction Monitoring^{*}

CHANG Cheng¹⁾, WU Song-Feng¹⁾, MA Jie¹⁾, ZHANG Wei¹⁾, ZHU Yun-Ping¹⁾^{**}

(¹⁾ State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing 102206, China;

²⁾ Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan 410073, China)

Abstract Selected reaction monitoring is one of the major approaches in targeted proteomics researches which mainly focus on the discovery of biomarkers. According to the research purpose, the specific proteins and peptides are selected before an SRM experiment. SRM increases the sensitivity, the accuracy and the dynamic range during the analysis especially in the detection capabilities of low abundance proteins and peptides. A typical SRM experiment includes three steps: experiment design; MS data acquisition; data validation and quantification. Among the three steps, it is vital to predict the special peptides and fragment ions using bioinformatics methods. The validation and quantification of the SRM data are also important to SRM data analysis. Meanwhile, the corresponding softwares, tools and databases are summarized. New experiments strategies and new bioinformatics tools are generated along with the developments of SRM instruments. Using improved experiment strategies, parameters and better quantitative algorithms as well as more sophisticated bioinformatics tools, SRM will play a more important role in the future of proteomics research.

Key words Targeted Proteomics, Biological Mass Spectrometry, Selected Reaction Monitoring, ProteoTypic Peptide, Transition

^{*} This work was supported by National Basic Research Program of China (No. 2011CB910601, 2010CB912700), National High Technology Research and Development Program of China (No. 2012AA020409, 2012AA020201) and National Natural Science Foundation of China (21105121)

^{**}Corresponding author. Tel: 010-80705225, E-mail: zhuyunping@gmail.com

Received: January 6, 2012 Accepted: March 15, 2012 Available online: March 19, 2012