



Orbitrap Exploris 480质谱在定量蛋白质组学应用中的优化和评测*

周岳¹⁾ 杨湘云²⁾ 黄敏²⁾ 李想³⁾ 关锋^{1,3)**}

⁽¹⁾ 江南大学生物工程学院, 糖化学与生物技术教育部重点实验室, 无锡 214122;

⁽²⁾ 赛默飞世尔科技(中国)有限公司, 上海 200120; ⁽³⁾ 西北大学, 生命科学学院, 西安 710069)

摘要 基于数据依赖的扫描模式 (data-dependent acquisition, DDA) 和数据非依赖的扫描模式 (data-independent acquisition, DIA) 的非标记定量 (label-free quantitative, LFQ) 和同位素标记 TMT (tandem mass tag) 定量是蛋白质组学定量中较常见的技术. 本文利用最新的 Orbitrap Exploris 480 质谱, 优化了 DDA、FAIMS DDA、FAIMS DIA 的非标记定量方法以及 TMT 定量策略的关键质谱参数, 并将其应用在人细胞蛋白质组、单细胞蛋白质组、血浆蛋白质组和酵母蛋白质组分析. 结果表明, 在 DDA 实验中, 设置碰撞能量为 27、二级谱图的分辨率为 15 K、最大离子注入时间为 22 ms 是最佳的参数组合. 针对极微量样品 200 pg~5 ng, 可以根据样品量相应设置最佳的质谱参数. 使用 200 pg 和 500 pg 的 HeLa 细胞样品, 分别鉴定到 1 259 和 1 725 个蛋白质, 从而实现了单细胞蛋白质组学的深度覆盖. 在 FAIMS DDA 实验中, 60 min 或 90 min 梯度时选择 CV-45V 的补偿电压, 120 min 或 150 min 梯度时选择 CV-45V-65V 补偿电压组合, 可以获得最佳的蛋白质鉴定结果. 从 293T 蛋白质组中分别在 60、120 和 150 min 中鉴定 6 300、6 994 和 7 500 个蛋白质. 在 FAIMS DIA 实验中, 使用 CV-45V-65V 双补偿电压切换, 60 个隔离窗口来获得最佳的蛋白质鉴定数目和定量重现性. 在 60 min 内分别在 293T 细胞和去除高峰度的血浆中定量到 7 019 个细胞蛋白和 1 077 个血浆蛋白. 在 TMT 定量实验中设置 APD 开启、“Precursor Fit” 阈值 70 和 Turbo TMT 功能的组合同时增加了 TMT 定量蛋白质的数量和 TMT 定量的准确性. 在 TMT11-plex 标记的酵母蛋白质组中 60 min 即可定量 10 989 个多肽和 2 162 个蛋白质. 综上所述, 本研究中优化的多种非标记定量和 TMT 定量方法为 Orbitrap Exploris 480 在蛋白质组学更广泛的应用提供了参考.

关键词 定量蛋白质组学, 非标记定量, TMT, FAIMS, Orbitrap Exploris 480

中图分类号 O658, Q51

DOI: 10.16476/j.pibb.2020.0245

定量蛋白质组学已广泛应用于生物学研究、精准医学和药物发现等领域, 是蛋白质组学研究的热门方向^[1]. 近几年, 越来越多的新技术如新型质谱技术、新的软件算法等用来提高定量蛋白质组学的深度和准确性^[2]. Orbitrap 质谱是蛋白质组学分析领域应用最多的质谱仪, 已经经历多次的技术更迭^[3]. 最新的 Orbitrap Exploris 480 质谱仪仍沿用方形开口的高容量离子传输管, 后端配合使用电动离子漏斗, 从而提高分析灵敏度; 主动离子传输组件采用最新的双弯曲四极杆, 提高抗污染能力; 全新的分段四极杆进一步提升了灵敏度和选择性, 最小的隔离窗口可达 0.4 Th; 改良后的 Orbitrap 质量分析器, 最高分辨率为 480 000, 具有 40 Hz 的采集

速度, 并且嵌入了 Turbo-TMT (tandem mass tag) 算法^[4]; Orbitrap Exploris 480 可以兼容高场非对称波形离子迁移谱 (high-field asymmetric waveform ion mobility spectrometry, FAIMS) 离子淌度^[5]. FAIMS 技术的引入可以将肽段在进入质谱之前通过碰撞横截面积的差异实现快速气相分离, 提高蛋白质组学分析的灵敏度和仪器的稳定性^[6]. Bekker-Jensen 等^[5] 评测了 Orbitrap Exploris 480 在短梯度时加载 FAIMS、使用单补偿电压

* 国家科技重大专项(2018ZX10302205-003)资助项目.

** 通讯联系人.

Tel: 029-88302411, E-mail: guanfeng@nwu.edu.cn

收稿日期: 2020-07-20, 接受日期: 2020-09-09

(compensation voltage, *CV*) 时数据依赖性的扫描模式 (data-dependent acquisition, DDA) 和数据非依赖性的扫描模式 (data-independent acquisition, DIA) 的性能, 但是未见详细质谱参数的优化、长时间梯度分析、FAIMS DDA 和 FAIMS DIA 多 *CV* 切换扫描以及分析皮克级别微量样品的性能. 本文优化了最常见的非标记定量 (label-free quantification, LFQ) 方法包括 DDA、FAIMS DDA、FAIMS DIA^[7] 以及同位素标记相对和绝对定量 TMT^[8] 的关键质谱参数, 在此最优的质谱参数基础上评测了 Orbitrap Exploris 480 在细胞蛋白质组、微量样品蛋白质组、血浆蛋白质组和酵母蛋白质组学的表现.

1 实验部分

1.1 实验试剂和仪器

EasyPep™ 蛋白质组学样品前处理试剂盒、Pierce™ HeLa 酶解肽段标准品、Pierce™ TMT11 标酵母酶解肽段标准品、Acclaim™ PepMap™ C18 分析柱 (规格 2 μm、75 μm、25 cm)、Top14 去高丰度蛋白试剂盒、Pierce 肽段定量试剂盒和 SOLA HRP 固相萃取小柱均采购于 Thermo Scientific.

1.2 样品制备

293T 全细胞肽段的制备过程: 人肾上皮细胞 293T 按照 EasyPep™ 试剂盒操作方法对 100 μg 的 293T 细胞全蛋白样品进行还原、烷基化、酶切、纯化, 得到酶切肽段, 溶解于 0.1% FA / H₂O 中. 然后用肽段定量试剂盒对酶切后的肽段定量, 浓度分别为 100 mg/L (上样 1 μl 或 2 μl 用于质谱分析)、500 mg/L (上样 1 μl、2 μl 或 4 μl 用于质谱分析).

去高丰度血浆肽段制备过程: 用 Top14 去高丰度蛋白试剂盒去除 10 μl 人血浆样品 (约 60 μg) 高丰度蛋白, 将去除高丰度蛋白的蛋白质样品用 FASP (filter-aided sample preparation)^[9] 方法还原、烷基化、酶切, 酶切后肽段用 SOLA HRP 固相萃取小柱脱盐. 酶切肽段使用 0.1% FA / H₂O 复溶, 浓度为 500 mg/L, 上样 1 μl 用于质谱分析.

1.3 NanoHPLC-MS/MS 分析

酶切肽段用 Easy nLC1200 (Thermo Scientific) 和 Exploris 480 (Thermo Scientific) 进行 LC-MS 分析. 肽段以 800 bar 的压力直接上样到分析柱上 (250 mm×75 μm), 再用 300 nl/min 流速进行洗脱, 柱温箱的温度为 55℃. 色谱条件如下: 流动相 A 为 0.1% 甲酸 (formic acid, FA) 水溶液, 流动相 B 为

0.1% FA 80% 乙腈溶液. 色谱梯度设置为 60、90、120 和 150 min.

DDA、FAIMS DDA 和 TMT 质谱方法: 采用 Orbitrap Exploris 480. 一级质谱采用 Orbitrap 检测, 分辨率 120 000@*m/z* 200, 最大注入时间为 20 ms, 目标 AGC 设为 300%; 二级谱图分辨率设为 15 000@*m/z* 200, 扫描范围为 *m/z* 120~2 000, 隔离窗口设为 1.6 Th, 目标 AGC 设为 75%, 强度阈值为 8e3; TMT 二级谱图设置为 MS2 分辨率为 45 000@*m/z* 200, Turbo TMT 开启时 MS2 分辨率分别设置为 15 000 和 30 000, Precursor Fit 分别设置为 0.6、0.7 和 0.8, 扫描范围起始为 110, 隔离窗口设置为 0.7 Th, 归一化碎裂能量为 38%, 最大注入时间为 100 ms, 目标 AGC 设为 100%, 强度阈值为 1 e5; 设置全扫描谱图和多个二级谱图的扫描周期为 2 s; 动态排除时间为 40 s; FAIMS 补偿电压设置为 -45V 或者 -65V 或者 -45V-65V, 其余参数按照默认设置.

FAIMS DIA 质谱方法: 用于样品分析的 DIA 方法, 一级质谱采用 Orbitrap 检测, 分辨率 120 000@*m/z* 200, 最大注入时间为 20 ms, 目标 AGC 设为 300%, 扫描范围 350~1 200; DIA 扫描分辨率设为 30 000@*m/z* 200, 扫描范围为 *m/z* 200~2 000, 最大离子注入时间为 Auto, 目标 AGC 设为 1 000%, 归一化碎裂能量为 32%, FAIMS 补偿电压设置为 -45V 或者 -65V 或者 -45V-65V, 其余参数按照默认设置. 用于色谱谱图库建立的 GPF-DIA (gas-phase fractionation)^[10-11] 方法, 293T 细胞酶切肽段和去除高丰度蛋白的血浆肽段样品进行 7 次 GPF-DIA 数据采集, 每次采集的 DIA 隔离窗口为 3 Th, 隔离窗口数目为 40, 每次 GPF-DIA 覆盖 120 Th 的 *m/z* 范围, 分别进行 *CV*-45V 和 *CV*-65V 的 GPF-DIA 数据采集.

1.4 数据库检索及功能分析

DDA 和 FAIMS DDA 非标记定量: 使用 Proteome Discoverer 2.4 (Thermo Scientific) 软件的 MSFragger^[12] 搜索引擎, 检索 Swissprot_Human (版本 2016 年 04 月, 包含 20 148 条蛋白质序列) 数据库, 一级谱中母离子质量误差为 ±10 ppm, 二级谱中碎片离子质量误差为 ±0.02 Da, 酶为胰蛋白酶 (trypsin), 完全酶切, 漏切数目为 2 个, 固定修饰为半胱氨酸的烷基化 (C, +57.021 50 Da), 可变修饰为甲硫氨酸的氧化 (M, +15.994 92 Da). Percolator^[13] 进行 PSM (peptide spectrum match)

FDR 的计算, PSM、肽段、蛋白质假阳性率均控制在 1%. 非标记定量用 Minora Feature Detector 进行色谱峰提取, Feature Mapper 进行色谱峰和肽段鉴定结果的匹配, 设置保留时间对齐, 最大的保留时间偏差为 10 min, Precursor Ions Quantifier 进行定量结果输出, 蛋白质强度为 top3 肽段强度的平均值.

微量样品分析: 使用 Proteome Discoverer 2.4 (Thermo Scientific) 软件的 MSepSearch^[14] 谱图匹配算法, 谱图库从 Proteome Discover 中下载, 为 NIST_Human_Orbitrap_HCD_20160923 和 NIST_ProteomeTools_Human_synthetic_HCD_20170530, 一级谱中母离子质量误差为 ± 10 ppm, 二级谱中碎片离子质量误差为 ± 0.02 Da, Precursor detector 用于从原始谱图中提取混合的二级谱图. Percolator 进行 PSM FDR 的计算, PSM、肽段、蛋白质假阳性率均控制在 1%. 非标记定量方法同 DDA 非标记定量方法.

TMT 定量: 使用 Proteome Discoverer 2.4 (Thermo Scientific) 软件的 SequestHT^[15] (TMT 实验) 搜索引擎, 一级谱中母离子质量误差为 ± 10 ppm, 二级谱中碎片离子质量误差为 ± 0.02 Da, 酶为胰蛋白酶, 完全酶切, 漏切数目为 2 个, 固定修饰为半胱氨酸的烷基化 (C, +57.021 50 Da), TMT6plex (K, N-term, +229.162 932 Da), 可变修饰为甲硫氨酸的氧化 (M, +15.994 92 Da). Percolator 进行 PSM FDR 的计算, PSM、肽段、蛋白质假阳性率均控制在 1%. Reporter Ions Quantifier 用于报告离子、肽段、蛋白质强度的计算, 共隔离干扰阈值设置为 50, S/N 阈值设置为 10.

FAIMS DIA 定量: GPF-DIA 数据用 DIA-NN^[16] 进行目标峰的抽提, 谱图库为 DDA 谱图库, 并将 GPF-DIA 结果生成色谱谱图库 (chromatography library). FAIMS DIA 数据用 DIA-NN 软件进行分析, 谱图库为 DDA 谱图库或者色谱谱图库, 蛋白质数据库为 Swissprot_Human (版本 2016 年 04 月, 包含 20 148 条蛋白质序列), 其余设置按照默认设置, 控制多肽和蛋白质水平的 FDR 为 1%.

2 结果与讨论

2.1 Orbitrap Exploris 480 在 DDA 定量分析中的参数优化和应用

基于数据依赖的质谱扫描模式 (data-

dependent acquisition, DDA) 的非标记定量是蛋白质组实验中最常用的方法, 通过 DDA 方法可以获得蛋白质鉴定和定量信息^[17]. 本文首先优化 Orbitrap Exploris 480 的 DDA 方法, 以获得单针最优的蛋白质鉴定数目. 合适的质谱参数是获得最佳结果的前提, 质谱采集的关键参数主要包括采集时长、二级谱图的分辨率、碎裂能量、二级碎片离子扫描最大离子注入时间等. 首先利用 293T 全细胞肽段样品, 对上述关键参数进行了优化. 如图 1a 所示, 在归一化碰撞能量 (normalized collision energy, NCE) 27~33 范围内的不同能量对鉴定结果影响有限, 推荐用 NCE 27 进行数据采集. 针对较为常用的 60 min 和 120 min 采集时长, 组合不同 MS2 分辨率和不同离子注入时间, 每个条件采集两次重复的质谱数据, 数据库检索得到多肽和蛋白质鉴定数目 (图 1b, c). 根据测试结果, 推荐采用二级谱图 15K 分辨率和 22 ms 离子注入时间的组合, 60 min 可以在 293T 全细胞酶解样品中鉴定到 5 636 个蛋白质和 35 339 条肽段, 在 120 min 则可以鉴定到 6 715 个蛋白质和 50 611 条肽段.

使用上述优化的质谱方法, 分析 100~1 000 ng 293T 全细胞酶解产物, 在 60 min 的梯度内, 本研究获得了较高的蛋白质组覆盖度, 在 100 ng 样品中鉴定到 4 480 个蛋白质, 在 200 ng 样品中鉴定出 4 830 个蛋白质, 而在 500~1 000 ng 样品中实现了超过 5 000 个蛋白质的鉴定 (图 1d, e), 蛋白质的平均序列覆盖度为 19%, 含有 2 个及以上肽段的蛋白质占总蛋白质数的 79%.

为评估 DDA 一级峰面积非标记定量数据重现性, 本研究对 500 ng 的 293T 全细胞酶解样品进行 3 次重复检测, 分别有 83.8% 的肽段和 90.29% 的蛋白质检测到超过 2 次, 其中 3 次检测共有的肽段和蛋白质占比分别为 70.28% 和 83.34%, 蛋白质强度变异系数的中位数值为 6.77% (图 1f, g).

2.2 Orbitrap Exploris 480 在微量样品分析中的优化和应用

随着质谱灵敏度的提高, 对于微量样品蛋白质组分析的需求越来越多, 如分析单细胞蛋白质组^[18] 以及临床上珍贵而且样品量有限的样品, 如外泌体、细针穿刺和活检组织切片等. 本研究用系列稀释的 HeLa 细胞蛋白质的酶切肽段, 对 Exploris 480 分析微量样品的质谱方法进行了优化, 同时评估了其在微量 HeLa 细胞蛋白质样品中鉴定和定量蛋白质组的能力. 首先, 对质谱上样量

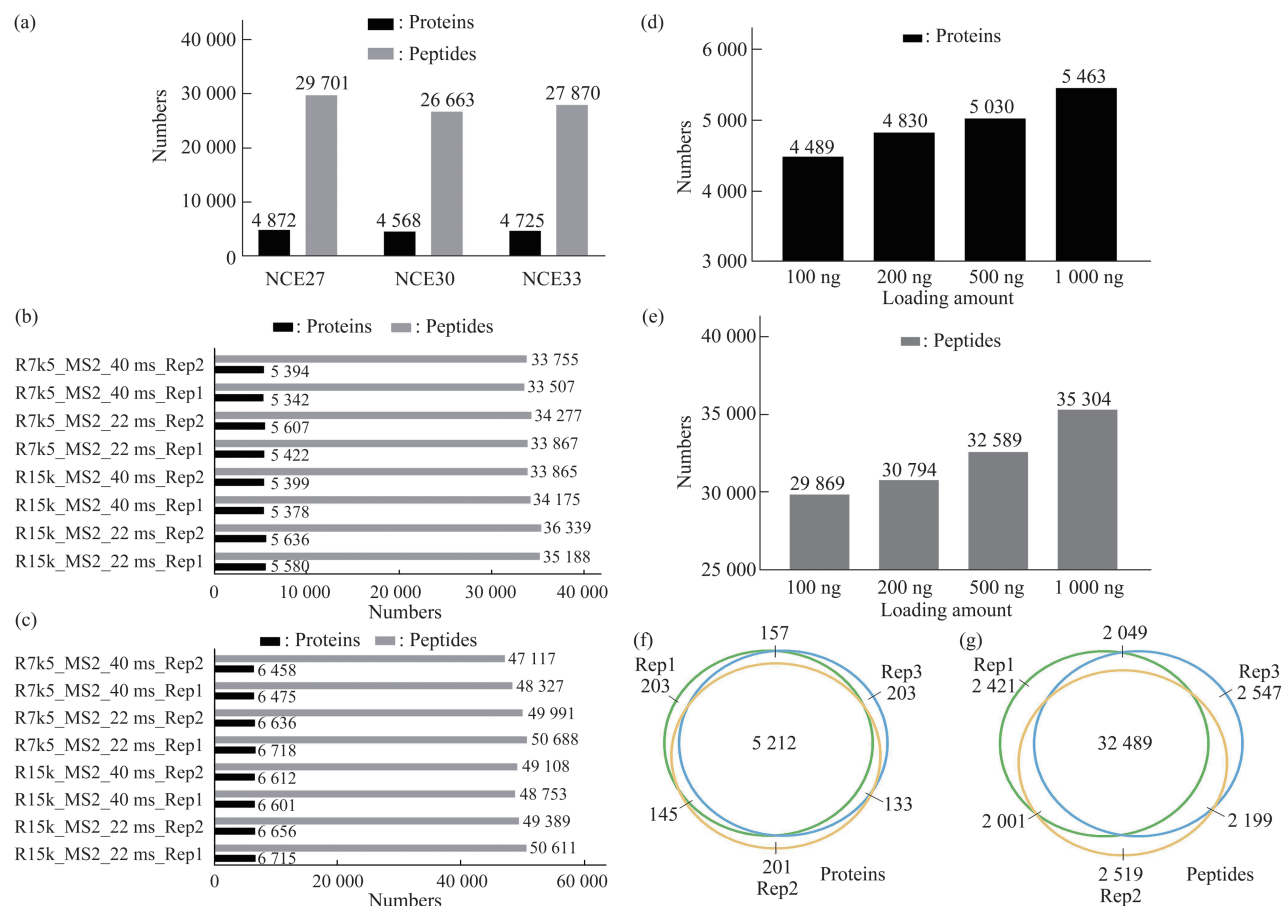


Fig. 1 Parameter optimization and application of Orbitrap Exploris 480 in label-free DDA quantification

The sample was 293T cell lysate. (a) The number of identified peptides and proteins under different fragmentation energies (200 ng sample, 60 min gradient, MS2 resolution 15 K, maximum MS2 ion injection time 40 ms). (b) The number of identified peptides and proteins at different MS2 resolution and ion injection time with 1 µg sample and 60 min gradient. (c) Identification results at different MS2 resolution and ion injection time with 1 µg sample and 120 min gradient. (d, e) Identification at 100–1 000 ng sample, 60 min gradient with optimized mass spectrometry parameters. (f, g) Venn diagrams of peptides and proteins in three technical replicates.

200 pg、1 ng 和 5 ng 的 HeLa 样品（分别对应 1 个细胞、5 个细胞和 25 个细胞的样品量）进行了质谱参数的优化，其中对微量蛋白质鉴定影响最大的参数为二级谱图的分辨率和二级碎片离子扫描最大离子注入时间。对于 200 pg 的样品量，设置二级谱图分辨率为 45 K，二级谱图最大离子注入时间为 118 ms；对于 1 ng 的样品量，设置二级谱图分辨率为 30 K，二级谱图最大离子注入时间为 86 ms；对于 5 ng 的样品量，设置二级谱图分辨率为 30 K，二级谱图最大离子注入时间为 54 ms，分别为该样品量下最佳的质谱参数，获得最多的多肽和蛋白质鉴定数目（图 2a, b）。但在实际的样品分析中，往往未知样品的蛋白质浓度，尤其某些微量样品的蛋白质浓度难以测定，通过本研究参数优化的结果

可以看出使用折中的质谱参数，设置二级谱图分辨率为 30 K 和二级谱图最大离子注入时间为 118 ms 可以获得 90% 左右的最优蛋白质鉴定结果（图 2a, b）。

根据每个样品浓度下最优的质谱参数，本研究分析了 200 pg、500 pg、1 ng、2 ng 和 5 ng 的 HeLa 蛋白酶切肽段样品，每个样品技术重复 2 次，单独检索每个样品量数据和合并检索所有样品量数据。单独检索非标记定量每个样品量数据时，蛋白质鉴定数据只依赖于二级谱图的匹配信息；合并检索非标记定量所有样品量数据时，蛋白质鉴定结果除依赖于二级谱图的匹配信息，软件会对根据母离子的精确保留时间和质量数在鉴定结果间进行传递，即 MBR（match between run）功能^[19]，高样品量的

蛋白质鉴定结果可以传递给低样品量,从而增加低样品量的蛋白质鉴定数目.通过MBR,200 pg样品的蛋白质鉴定数目从734增加到1 259,提高了71%,500 pg样品的蛋白质鉴定数目从1 234增加到1 725,提高了39%(图2c,d).高分辨质谱对低量样品的定量具有很高的重现性,每个样品量蛋白质强度的变异系数中值在15%以内(图2e),两针200 pg上样量的定量蛋白质强度的相关性为0.96(图2f).高分辨质谱对低量样品的定量具有良好的

线性范围,样品量和蛋白质总强度的 R^2 为0.99(图2g).

对于微量样品,本研究优化了Exploris 480不同浓度最优的质谱参数以及未知样品浓度时的折中质谱参数,在数据检索中MBR显著提高了微量样品的鉴定数目,200 pg的HeLa细胞酶切肽段样品可以鉴定到1 259个蛋白质,实现单细胞蛋白质组的深度覆盖.结果表明高分辨质谱对于微量样品蛋白质组具有精确和准确的定量性能.

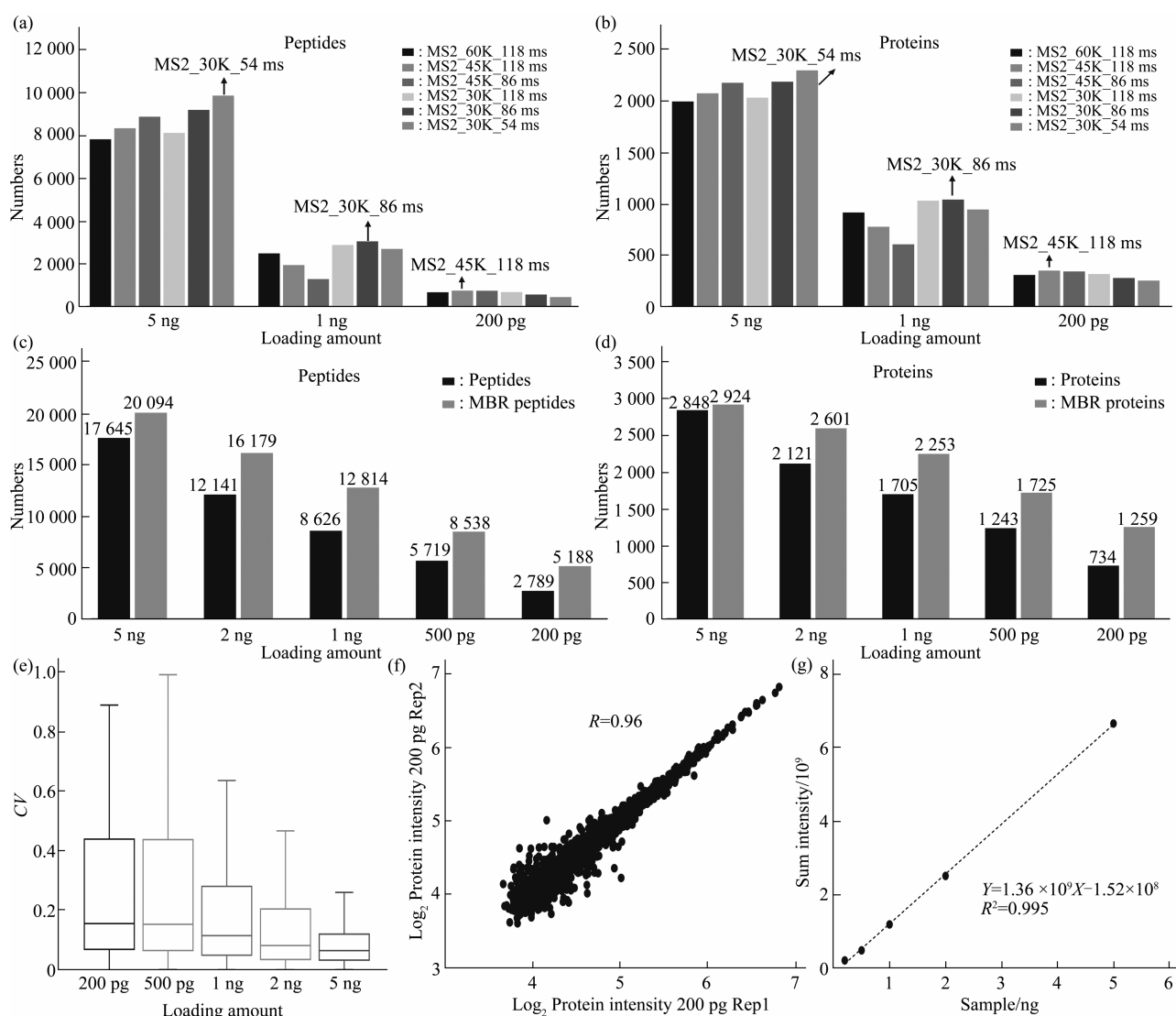


Fig. 2 Optimization and application of Orbitrap Exploris 480 in the analysis of ultra-low amount proteomic samples

We used Pierce HeLa Standard Peptides Digests for method development. (a, b) The number of peptides and proteins identified with different mass spectrometry parameters in 200 pg, 1 ng, and 5 ng samples. (c, d) The number of peptide and proteins identified in 200 pg, 500 pg, 1 ng, 2 ng, and 5 ng samples with optimized mass spectrometry parameters. Two methods were used for database searching: individual search of each sample loading (peptides, proteins) and combined search of all samples (MBR peptides, MBR proteins). (e) Distribution of coefficient of variation of protein intensities in different sample amounts. (f) Correlation of protein intensities of two technical replicates of 200 pg samples. (g) Correlation of sample amount and total protein intensity.

2.3 Orbitrap Exploris 480结合FAIMS离子淌度在DDA定量分析中的优化和应用

FAIMS是一种在大气压条件下分离离子的技术,通过调节FAIMS的补偿电压(compensation voltage, *CV*)值,可以分离带电性质不同的肽段,从而作为一种在线的气态样品分级手段来提高复杂样品的定性和定量效率. Schweppe等^[20]和Pfammater等^[6]评测了FAIMS DDA在三合一质谱上采用一级高分辨质谱全扫描、二级低分辨质谱碎片离子扫描模式不同分析梯度的最佳*CV*组合,但是Exploris 480采用一级高分辨质谱全扫描、二级高分辨质谱碎片离子扫描模式的不同分析梯度*CV*组合尚未见报道.本文用293T细胞的酶切肽段进行*CV*优化,结果显示60 min或90 min梯度采用*CV*-45V的单电压可以获得更多的蛋白质鉴定数目,但是采用*CV*-45V-65V可以获得更多的多肽鉴定数目;而对于120 min、150 min等长梯度则选用-45V和-65V的双电压组合获得最多的多肽和蛋白质鉴定数目(图3a, b).

结合DDA扫描最优的MS2质谱参数及FAIMS最优的*CV*值,本研究测试不同分析时长(60、120和150 min),不同样品量(200、500、1 000和2 000 ng)的293T全细胞酶解肽段多肽和蛋白质鉴定数目(图3c, d).相较于图1中不加载FAIMS的结果,FAIMS DDA在相同60 min梯度内比DDA鉴定的蛋白质数目提升约15%~25%,延长质谱采集时间至120 min,FAIMS DDA在2 μ g细胞样品中的蛋白质鉴定数目接近7 000个,150 min鉴定接近7 500个蛋白质,体现了FAIMS在蛋白质组学应用中的巨大潜力.为了评估Exploris 480搭载FAIMS的重现性,本研究对293T全细胞酶解样品进行3次90 min梯度的重复检测.相比于不加载FAIMS的结果,此次试验的数据质量和数据重现性具有非常明显的提升,检测到超过2次的肽段和蛋白质占比分别高达96.05%和98.26%,而其中3次共有肽段占比为91.79%和96.44%,3次技术重复蛋白质强度变异系数的中位数值小于5%(图3e, f).可见FAIMS在提高蛋白质鉴定深度的同时也提高了蛋白质定量的重现性.

为探究上述基于细胞样品优化的条件是否对于其他样品同样适用,本研究对最常见的也是解析难度较大的血浆样品进行测试.发现,*CV*-45V和60 min色谱梯度,*CV*-45V-65V双电压切换和90 min两个方法,分别可以从1 μ g去除14个高丰

蛋白的血浆样品中鉴定到549和665个蛋白质,以及3 864和5 165个肽段(图3g, h).

2.4 Orbitrap Exploris 480结合FAIMS离子淌度在DIA定量分析中的优化和应用

DIA定量方法由于其重现性好、缺失值少、变异系数小等优点被广泛应用于临床蛋白质组学研究^[21],目前FAIMS DIA采集方法文献报道较少,只有Bekker-Jensen等^[5]评测了单*CV*条件下21 min的梯度进行DIA定量的实验结果.鉴于FAIMS在DDA非标记定量中的提升,本研究尝试优化双*CV*切换和单*CV*的FAIMS DIA方法,并应用于细胞蛋白质组和血浆蛋白质组分析. DIA数据分析基于样品特异的谱图库,通过第一维高pH反向色谱分离收集馏份,第二维低pH反相色谱-质谱联用的DDA方法建立了人细胞的蛋白质组谱图库,包含有248 769条肽段和10 965个蛋白质;同样本研究建立了血浆蛋白质组谱图库,包含有50 151条肽段和5 190个蛋白质.

首先,本研究用293T细胞酶切肽段来优化FAIMS DIA的方法.根据非标记定量的优化结果,选取*CV*-45V和-65V两个电压,每个电压下分别进行独立的DIA扫描,组成一个循环(图4a)或者设定单个电压进行一级的全扫描和DIA扫描. DIA方法中,选取合适的DIA隔离窗口数目至关重要,其影响多肽和蛋白质定量数目以及定量的重现性^[21].在60 min和25 cm色谱柱的色谱条件下,设置30、60和80个隔离窗口,分别对应平均每个色谱峰7个、4个和3个点.比较单*CV*值下30个隔离窗口、每个*CV*值下各30和40个隔离窗口的蛋白质和多肽定量数目(图4b)以及重现性(DIA *CV*-45V_30w、DIA *CV*-65V_30w、DIA *CV*-45V-65V_60w、DIA *CV*-45V-65V_80w)(图4c). DIA *CV*-45V-65V_60w的定量蛋白质数目比DIA *CV*-45V-65V_80w高6%,同时展现出更好的定量重现性, DIA *CV*-45V-65V_60w 3次技术重复蛋白质强度的变异系数中位数为5.7%,而DIA *CV*-45V-65V_80w为12%.本研究同时比较了双*CV* DIA方法与单*CV* DIA方法定量多肽和蛋白质数目和重现性.单*CV*-45V_30w和*CV*-65V_30w方法的定量多肽数目是双*CV*-45V-65V_60w方法的50%和70%,蛋白质鉴定数目是*CV*-45V-65V_60w方法的88%和90%,同时3种DIA方法对于定量的重现性没有差别.所以*CV*-45V-65V_60w在提高多肽和蛋白质定量数目的同时也能获得较好的定量重现性,本研究选取该方法应

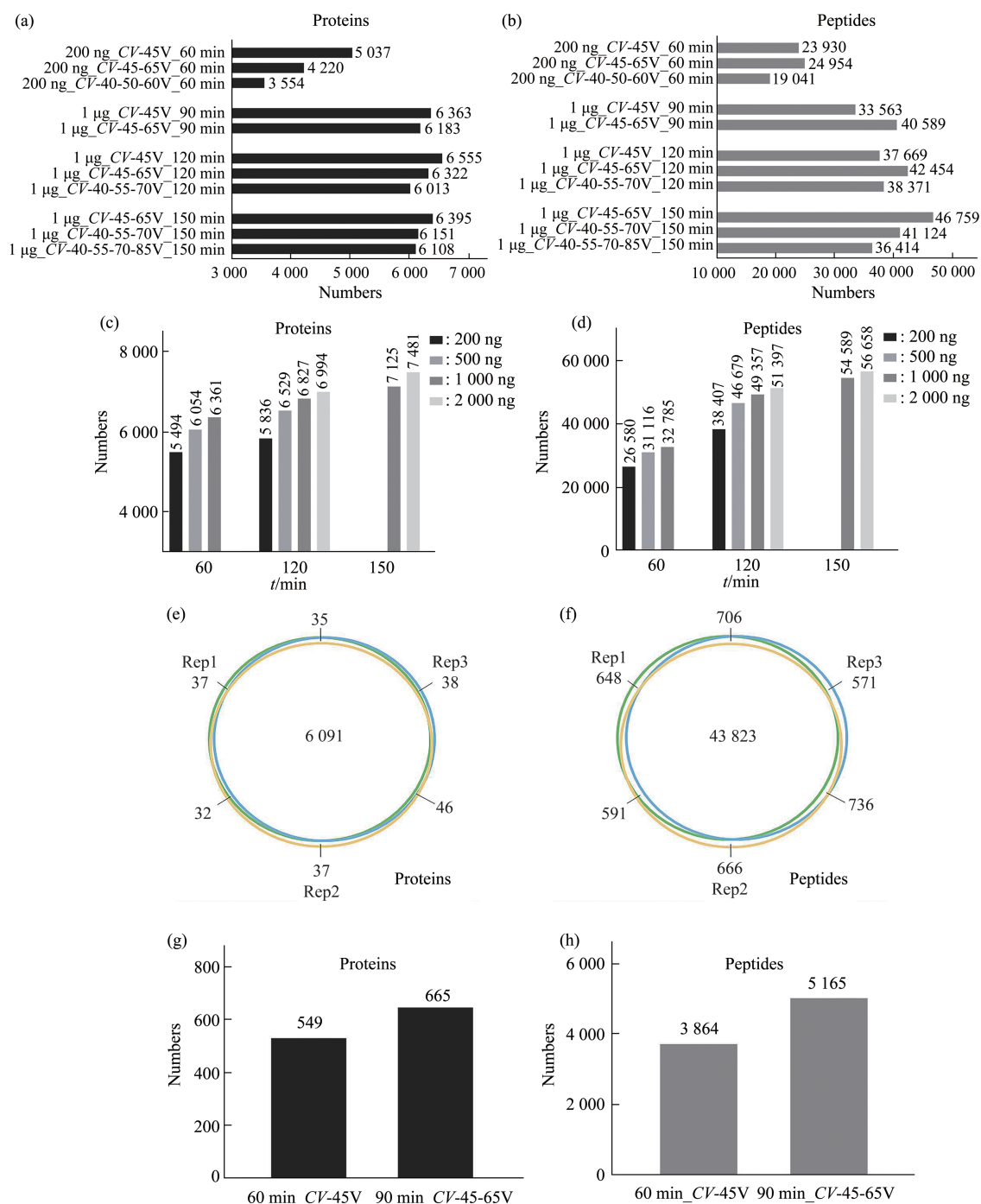


Fig. 3 Optimization and application of FAIMS DDA in label free quantification

(a, b) Identification of 293T cell lysate in 60, 90, 120 and 150 min gradients with different FAIMS CVs. (c, d) Identification of protein and peptide from 200–2 000 ng of 293T cell lysate using optimized FAIMS CV and mass spectrometry parameters in 60, 90, 120 and 150 min gradients. (e, f) Venn diagrams of peptides and proteins in three technical replicates of 293T cell lysate. (g, h) Protein and peptide identifications with depleted plasma in 60 and 90 min gradients.

用于293T细胞蛋白质组和血浆蛋白质组分析。

Searle等^[11]在DDA谱图库的基础上提出了色谱谱图库的概念,进一步提高了DIA定量多肽和

蛋白质的数目.本研究首次将色谱谱图库应用于FAIMS DIA的数据分析.按照文献报道的方法,首先建立了293T蛋白质组60 min梯度的色谱谱图库。

在建立色谱谱图库时, 采用单 *CV*-45V 或者 *CV*-65V, DIA 的隔离窗口为 3 Th, 40 个隔离窗口数目为 1 个循环, 每次 DIA 扫描可以覆盖 120 Th 的质量范围, 覆盖 *m/z* 350~1190, 需要 7 次重复进样扫描, 每次扫描覆盖的 *m/z* 范围分别为 350~470、470~590、590~710、710~830、830~950、950~1 070、1 070~1 190; DIA-NN 分析软件对生成的窄窗口 DIA 数据进行目标峰抽提, 采用的谱图库为已建立的 DDA 细胞蛋白质组谱图库, 最后软件将鉴定到的多肽和蛋白质生成色谱谱图库. 293T 细胞蛋白质组的色谱谱图库中包含 8 742 个蛋白质, 99 285 条多肽. 采用色谱谱图库的 293T FAIMS DIA 比 DDA 谱图库定量的多肽数目提高 7%, 蛋白质数目提高 5.5%, 比不加载 FAIMS 的常规 DIA 定量蛋白质数目提高 14% (图 4d), 同时几乎涵盖了所有定量到的蛋白质 (图 4f). FAIMS DIA 采用色谱谱图库, 依然具有很好的定量重现性, 3 次技术重复蛋白质强度变异性系数的中位数为 8% (图 4e). 同样, 本研究建立了去除 14 个高丰度蛋白的血浆蛋白质组色谱谱图库, 包含 1 175 个蛋白质和 8 372 条肽段. 结果发现, 对于去除高丰度蛋白的血浆蛋白质组, FAIMS DIA 采用色谱谱图库和 DDA 谱图库的定量多肽和蛋白质数目相当, 但是比不加载 FAIMS 的常规 DIA 定量多肽数目提高 5%, 蛋白质数目提高 42% (图 4g, i). FAIMS DIA 分析血浆蛋白质组时依然具有很好的重现性, 3 次技术重复蛋白质强度变异性系数的中位数为 9% (图 4h).

经过数据采集方法和谱图库的优化, FAIMS DIA 在 60 min 的梯度内能达到 7 019 个蛋白质的细胞蛋白质组定量深度, 在 60 min 的梯度内能达到 1 077 个蛋白血浆蛋白质组的定量深度, 实现了细胞和血浆蛋白质组的快速和深度覆盖.

2.5 Orbitrap Exploris 480 在 TMT 定量分析中的优化和应用

基于报告离子定量是另一种常见的蛋白质组定量方法, 一次实验可以同时定量 11 个甚至 16 个样品, 大大提高了定量的通量, 但是 TMT 定量受到比值压缩的影响降低了定量的准确度. 为了提升 TMT 二级定量的准确性和定量通量, Orbitrap Exploris 480 中新增了 “Precursor Fit”^[22] 和 TurboTMT 功能. “Precursor Fit” 通过 *m/z* 和价态计算母离子的同位素峰簇在隔离窗口内的比例. 若隔离窗口内目标母离子同位素峰簇比例高, 表明肽段定量特异性高, 从而触发二级; 若隔离窗口内目

标母离子同位素峰簇比例低, 表明肽段定量特异性差, 便不会触发二级. 因此, 启用 “Precursor Fit” 选项可获得特异性更高的 MS/MS 谱图, 提高定量准确性.

Pierce™ TMT11 标酵母酶解肽段标准品为基因敲除和野生型酵母菌株 (*met6Δ*、*his4Δ*、*ura2Δ*、BY4741) 酶解肽段的 TMT11 标混合样品^[23]. 本研究评估了 “Precursor Fit” 值对定量准确性的影响, 其中准确性以无干扰指数 *IFI* (interference free index) 值进行评估^[23], 根据 *met6Δ*、*his4Δ*、*ura2Δ* 酵母菌株样品的隔离窗口内干扰来计算, *IFI* 值越大表示准确性越高. 蛋白质组学实验中, 高级峰识别算法 APD (advanced peak determination) 功能可大大增加用于二级碎裂的母离子数目^[24], 并获得更多的 MS2 谱图、PSM 和肽段鉴定数目. 但在 TMT 实验中需要将此功能关闭, 以降低共流出母离子的二级谱图采集, 提高定量准确性^[25]. 因此本研究比较了 APD 关闭、APD 开启以及 APD 开启结合不同的 “Precursor Fit” 值时的蛋白质肽段鉴定及定量准确性, 结果如图 5 所示 (上样量 500 ng, 60 min 梯度). APD 开启后, 可获得更多的肽段鉴定 (图 5a, b), 但 *IFI* 值降低 (图 5c), 说明定量准确性降低. 叠加使用 Precursor Fit 功能后, *met6*、*his4* 和 *ura2* 的 *IFI* 值有不同程度的提升, 当 “Precursor Fit” 设置为 80 时, *IFI* 值最高, *his4* 和 *ura2* 的 *IFI* 值基本达到 APD off 时的数值. 说明 Precursor Fit 可有效提高 TMT 标记定量的准确性. APD 开启后, 蛋白质和肽段的定量比例均降低, 但在叠加使用 Precursor Fit 功能后, 除在保持 APD 的优势 (肽段鉴定数目提高) 之外, 蛋白质和肽段定量比例提升, 接近 99.9%. 对比不同的 “Precursor Fit” 阈值, 当设置为 70 和 80 时, 蛋白质肽段鉴定、定量数目、及定量比例较高, 且定量准确性高, 可用于 TMT 定量实验.

TurboTMT 通过 Φ SDM 谱图后处理算法提高报告离子区域的分辨率, 可在较低分辨下完成原来需要高分辨才可实现的 TMT10 标以上报告离子的分辨, 提高扫描速度, 从而提升定量通量. 本研究采用相同的 TMT11 标酵母酶解肽段标准品, 评估了 TurboTMT 功能对鉴定性能的影响 (上样量 500 ng, 60 min 梯度). 在 30 K 和 15 K 的二级谱图分辨率下分别开启 TurboTMT 功能, 所得蛋白质和肽段鉴定数目均比 45 K 的分辨率时提高, 其中 15 K 分辨率时蛋白质和肽段鉴定超过 45 K 以及 APD 开启条件,

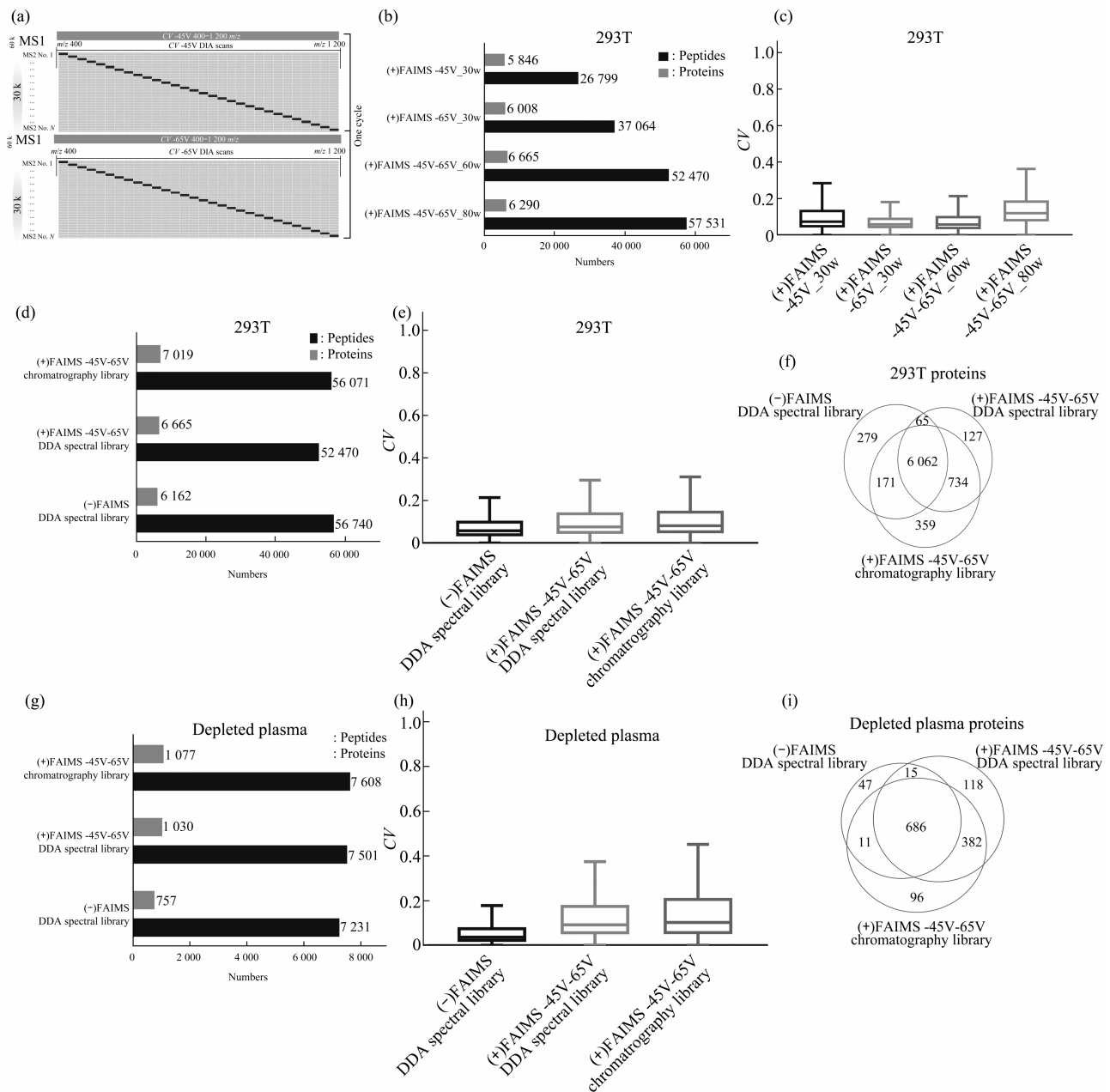


Fig. 4 Optimization and application of Orbitrap Exploris 480 in FAIMS DIA quantification

(a) FAIMS DIA scanning mode switching CV -45V and CV -65V. (b) Number of quantified peptides and proteins of (+)FAIMS-45V-65V_80w, (+)FAIMS-45V-65V_60w, (+)FAIMS-45V_30w and (+)FAIMS-65V_30w. (c) Protein intensities coefficient of variation distributions from (+)FAIMS-45V-65V_60w, (+)FAIMS-45V_30w and (+)FAIMS-65V_30w for triplicate analysis of 293T lysates. (d-f) FAIMS DIA analysis of 293T cell proteome in triplicates. Number of quantified peptides and proteins (d), coefficient of variation distributions of protein intensities (e) and venn diagram of quantitative protein with (+)FAIMS-45V-65V chromatography library (FAIMS DIA data targeted extraction from chromatography library), (+)FAIMS-45V-65V DDA spectral library (FAIMS DIA data targeted extraction from DDA spectral library) and (-)FAIMS DDA spectral library (DIA data without FAIMS targeted extraction from DDA spectral library) (f). (g-i) FAIMS DIA analysis of depleted plasma proteome in triplicates. Number of quantified peptides and proteins (g), coefficient of variation distributions of protein intensities (h) and venn diagram of quantitative protein with (+)FAIMS-45V-65V chromatography, (+)FAIMS-45V-65V DDA spectral library and (-)FAIMS DDA spectral library (i).

且可定量的蛋白质和肽段比例更高 (图 5a, b). 说明 TurboTMT 功能可提高 TMT 实验的鉴定通量,

且此时 11 标 TMT 的报告离子分辨率可分别达到 200 K 和 100 K 左右, 实现报告离子的分辨. 此外,

met6、his4 和 ura2 的 *IFI* 值在使用 TurboTMT 功能前后一致, 说明 TurboTMT 功能的开启不会影响 TMT 定量准确性 (图 5c)。

接着, 尝试联合测试 APD 开启/关闭、不同“Precursor Fit”阈值和 TurboTMT 功能开启/关闭, 评估它们对鉴定和定量性能的影响。联合使用 APD 开启, “Precursor Fit”阈值 70 和 TurboTMT 功能开

启后, 鉴定和定量蛋白质和肽段数目最多 (图 5a, b), 且不影响定量准确性 (图 5c)。在 60 min 梯度, 在 11 个酵母样本中, 实现 10 989 条肽段和 2 162 个蛋白质的定量结果。综上所述, 可知联合这三种功能有利于提高 TMT 标记实验的鉴定和定量通量, 以及定量准确性。

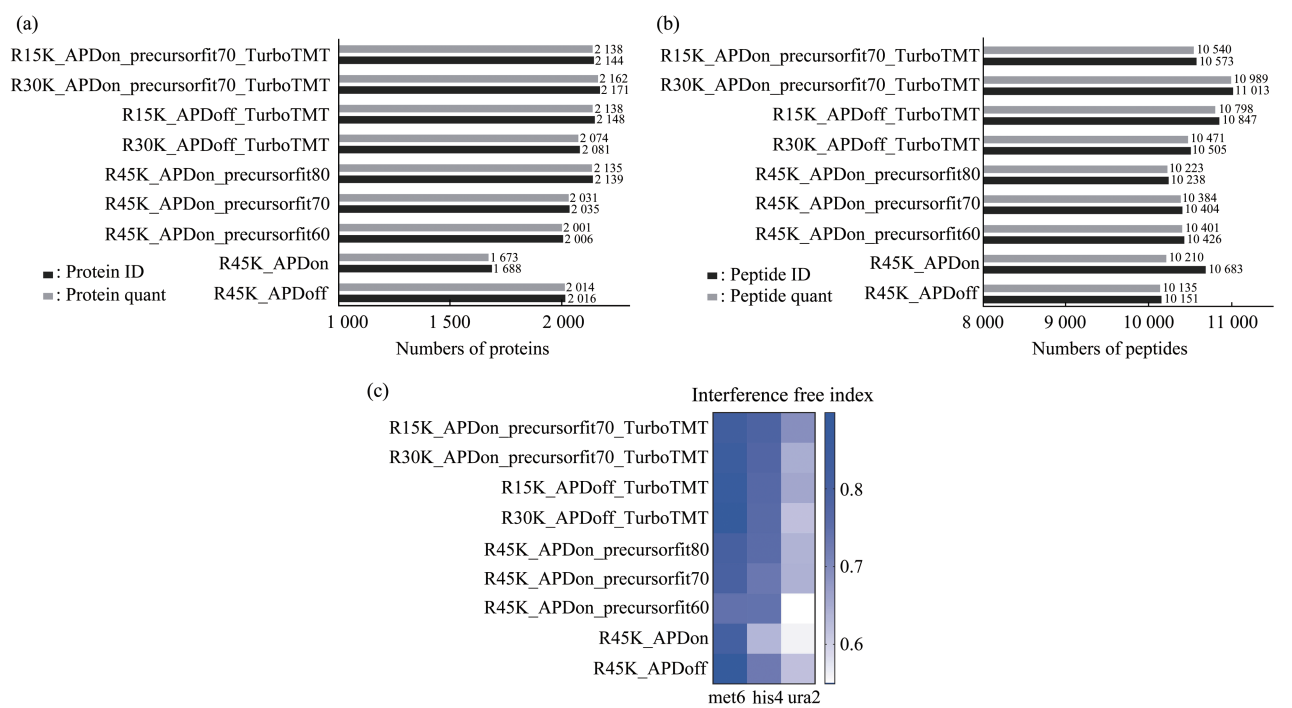


Fig. 5 Optimization and application of Orbitrap Exploris 480 in TMT quantification

The sample is Pierce TMT11 Yeast Digest Standard. Number of identified and quantified proteins (a), number of identified and quantified peptides (b) and met6, his4 and ura2 proteins *IFI* (c) in conditions of R45k_APDOff (Resolution=45 K and APD off), R45K_APDon (Resolution=45 K and APD on), R45K_APDon_precursorfit60 (Resolution=45 K, APD on and Precursor fit threshold 60%), R45K_APDon_precursorfit70 (Resolution=45 K, APD on and Precursor fit threshold 70%), R45K_APDon_precursorfit80 (Resolution=45 K, APD on and Precursor fit threshold 80%), R30K_APDOff_TurboTMT (Resolution=30 K, APD off and Turbo TMT on), R15K_APDOff_TurboTMT (Resolution=15 K, APD off and Turbo TMT on); R30K_APDon_Precursorfit70_TurboTMT (Resolution=30 K, APD on, Precursor fit threshold 70% and Turbo TMT on), R15K_APDon_Precursorfit70_TurboTMT (Resolution=15 K, APD on, Precursor fit threshold 70% and Turbo TMT on).

3 结 论

Orbitrap Exploris 480 是最新的用于蛋白质组学分析的质谱仪, 其在蛋白质组学中的应用还需进行优化和评估。本研究对 3 种非标记定量方法 DDA、FAIMS DDA 和 FAIMS DIA, 以及同位素标记定量 TMT 的关键参数进行了一系列优化和评估。在 DDA 非标记定量实验中, 设置碰撞能量 27, 二级谱图 15 000 分辨率和 22 ms 离子注入时间的组合可以获得最优的蛋白质鉴定数目, 同时 DDA 定量可

以获得较高的定量重现性, 3 次技术重复蛋白质强度变异系数的中位数值为 6.77%。FAIMS 接口可以选择性地排除干扰离子, 简化样品复杂度, 更高效、更稳定地完成蛋白质组学深度解析, 本研究首次在 Orbitrap Exploris 480 质谱上对 FAIMS DDA 和 FAIMD DIA 两种扫描模式进行了方法优化并评测了细胞和血浆蛋白质组的定量深度。通过优化 FAIMS DDA 不同分析时长的 *CV* 组合, 60 min 或者 90 min 的梯度选用 *CV*-45V, 120 min 或者 150 min 梯度选用 *CV*-45V-65V 的电压组合可以获得

最多的蛋白质鉴定数目。FAIMS DDA 相比于不加载FAIMs的DDA分析,在相同梯度时间内蛋白质鉴定数目提高15%~25%,在293T细胞蛋白质组中,60 min可以鉴定6 300个蛋白质,120 min可以鉴定6 994个蛋白质,150 min的长梯度分析可以达到7 500个蛋白质的鉴定深度,同时定量的重现性也进一步提高,3次技术重复的变异系数的中位数小于5%。对于FAIMS DIA, CV-45V-65V_60w的方法比 CV-45V_30w、CV-65V_30w 定量到更多的多肽和蛋白质数目,同时定量的重现性相当;使用色谱谱图库相比于DDA谱图库可以进一步提高FAIMS DIA定量到293T蛋白质组多肽和蛋白质数目,60 min的梯度,可以定量7 019个293T细胞蛋白质,同时3次技术重复的蛋白质变异系数中位数为8%;使用色谱谱图库和DDA谱图库对去除14个高丰度蛋白的血浆蛋白质组学定量多肽和蛋白质数目相当,60 min梯度,FAIMS DIA定量到1 077个血浆蛋白质,同时3次技术重复的蛋白质变异系数中位数为9%。比较DDA、FAIMS DDA、FAIMS DIA三种常见的非标记定量方法,FAIMS DIA具有最佳的蛋白质组覆盖度,同时具有良好的定量重现性,值得进一步深入优化。

对极低量样品甚至单细胞蛋白质组分析是蛋白质组学发展的新趋势。采用DDA非标记定量的方法,本研究优化了分析微量样品的关键质谱参数,二级谱图的分辨率和二级谱图的最大离子注入时间,质谱上样为200 pg、1 ng、5 ng时均有各自最佳质谱参数,但是本研究发现,二级谱图分辨率设置为30K和二级谱图的最大离子注入时间设置为118 ms是相对折中的方法,可以达到最佳鉴定结果的90%。使用优化的质谱方法分析200 pg、500 pg、1 ng、2 ng和5 ng的HeLa细胞样品,非标记定量数据分析MBR的功能显著提高低浓度的多肽和蛋白质鉴定数目,200 pg的样品量鉴定1 259个蛋白质,500 pg的样品量鉴定1 725个蛋白质,实现单细胞蛋白质组水平的深度覆盖。同时我们的数据表明即使极微量的样品,也可以获得较高的定量重现性,200 pg~5 ng样品量蛋白质强度的变异系数中位数在15%以内。

为了提高TMT定量的准确度和深度开发了“Precursor fit”阈值和Turbo TMT功能,通过对TMT11标酵母酶解肽段标准品的优化,本研究发现联合使用APD开启、“Precursor Fit”设置为70和Turbo TMT开启时可兼顾定量的深度和准确度;

在60 min梯度的TMT-11标记酵母蛋白质组中,定量了10 989条肽段和2 162个蛋白质。总之,本研究通过多个蛋白质组实验的优化和测试,为后续Orbitrap Exploris 480在定量蛋白质组学中的应用提供方法参考,通过本研究结果可以预见Orbitrap Exploris 480在翻译后修饰蛋白质组如磷酸化和糖基化修饰中也能有出色的表现,FAIMS可以排除背景离子的干扰,同时对相同肽段序列不同修饰位点的肽段会有一定的分离作用,将会是未来研究的重要方向。

参 考 文 献

- [1] Schubert O T, Rost H L, Collins B C, *et al.* Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc*, 2017, **12**(7): 1289-1294
- [2] Ankney J A, Muneer A, Chen X. Relative and absolute quantitation in mass spectrometry-based proteomics. *Annu Rev Anal Chem*, 2018, **11**(1): 49-77
- [3] Eliuk S, Makarov A. Evolution of orbitrap mass spectrometry instrumentation. *Annu Rev Anal Chem*, 2015, **8**: 61-80
- [4] Kelstrup C D, Aizikov K, Batth T S, *et al.* Limits for resolving isobaric tandem mass tag reporter ions using phase-constrained spectrum deconvolution. *J Proteome Res*, 2018, **17**(11): 4008-4016
- [5] Bekker-Jensen D B, Martinez-Val A, Steigerwald S, *et al.* A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. *Mol Cell Proteomics*, 2020, **19**(4): 716-729
- [6] Pfammatter S, Bonnell E, Mcmanus F P, *et al.* A novel differential ion mobility device expands the depth of proteome coverage and the sensitivity of multiplex proteomic measurements. *Mol Cell Proteomics*, 2018, **17**(10): 2051-2067
- [7] Chelius D, Bondarenko P V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res*, 2002, **1**(4): 317-323
- [8] Thompson A, Schafer J, Kuhn K, *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*, 2003, **75**(8): 1895-1904
- [9] Wisniewski J R, Zougman A, Nagaraj N, *et al.* Universal sample preparation method for proteome analysis. *Nat Methods*, 2009, **6**(5): 359-362
- [10] Searle B C, Swearingen K E, Barnes C A, *et al.* Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat Commun*, 2020, **11**(1): 1548
- [11] Searle B C, Pino L K, Egerton J D, *et al.* Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat Commun*, 2018, **9**(1): 5128
- [12] Kong A T, Leprevost F V, Avtonomov D M, *et al.* MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods*, 2017, **14**(5):

- 513-520
- [13] Kall L, Canterbury J D, Weston J, *et al.* Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*, 2007, **4**(11): 923-925
- [14] Zhang Z, Burke M, Mirokhin Y A, *et al.* Reverse and random decoy methods for false discovery rate estimation in high mass accuracy peptide spectral library searches. *J Proteome Res*, 2018, **17**(2): 846-857
- [15] Diamant B J, Noble W S. Faster SEQUEST searching for peptide identification from tandem mass spectra. *J Proteome Res*, 2011, **10**(9): 3871-3879
- [16] Demichev V, Messner C B, Vernardis S I, *et al.* DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods*, 2020, **17**(1): 41-44
- [17] Clough T, Key M, Ott I, *et al.* Protein quantification in label-free LC-MS experiments. *J Proteome Res*, 2009, **8**(11): 5275-5284
- [18] Slavov N. Single-cell protein analysis by mass spectrometry. *Curr Opin Chem Biol*, 2020, **60**: 1-9
- [19] Cox J, Hein M Y, Lubner C A, *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*, 2014, **13**(9): 2513-2526
- [20] Schweppe D K, Prasad S, Belford M W, *et al.* Characterization and optimization of multiplexed quantitative analyses using high-field asymmetric-waveform ion mobility mass spectrometry. *Anal Chem*, 2019, **91**(6): 4010-4016
- [21] Ludwig C, Gillet L, Rosenberger G, *et al.* Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol*, 2018, **14**(8): e8126
- [22] Yu Q, Paulo J A, Naverrete-Perea J, *et al.* Benchmarking the Orbitrap tribrid Eclipse for next generation multiplexed proteomics. *Anal Chem*, 2020, **92**(9): 6478-6485
- [23] Paulo J A, O'connell J D, Gygi S P. A triple knockout (TKO) proteomics standard for diagnosing ion interference in isobaric labeling experiments. *J Am Soc Mass Spectrom*, 2016, **27**(10): 1620-1625
- [24] Hebert A S, Thoing C, Riley N M, *et al.* Improved precursor characterization for data-dependent mass spectrometry. *Anal Chem*, 2018, **90**(3): 2333-2340
- [25] Myers S A, Klaeger S, Satpathy S, *et al.* Evaluation of advanced precursor determination for tandem mass tag (TMT) -based quantitative proteomics across instrument platforms. *J Proteome Res*, 2019, **18**(1): 542-547

Optimization and Evaluation of Orbitrap Exploris 480 Mass Spectrometry for Quantitative Proteomics*

ZHOU Yue¹⁾, YANG Xiang-Yun²⁾, HUANG Min²⁾, LI Xiang³⁾, GUAN Feng^{1,3)}**

¹⁾Key Laboratory of Carbohydrate Chemistry & Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi 214122, China;

²⁾Themofisher Scientific, Shanghai 200120, China;

³⁾The College of Life Sciences, Northwest University, Xi'an 710069, China)

Abstract Data-dependent acquisition (DDA) and data-independent acquisition (DIA) based label-free quantification (LFQ) and tandem mass tag (TMT) based isotope labeling quantification are two key techniques for quantitative proteomics. Here we optimized the latest Orbitrap Exploris 480 mass spectrometer parameters in DDA, FAIMS DDA, FAIMS DIA based LFQ and TMT, and show its performance in cell line proteomics, single-cell proteomics, plasma proteomics and yeast proteomics. The results showed that the collision energy of 27, the resolution of the fragment spectrum of 15K, and maximum ion injection time of 22 ms are the best parameter combination for DDA experiment. For the proteomic analysis of ultra-low samples ranging from 200 pg–5 ng, the individual mass spectrometer parameters should be considered. We identified 1 259 and 1 725 proteins in 200 pg and 500 pg of HeLa cell lysate, respectively, achieving deep coverage of single-cell proteomics. In FAIMS DDA experiment, we chose *CV*-45V for 60 min or 90 min gradient, *CV*-45V-65V combinations for 120 min or 150 min gradient to obtain the optimal protein identifications, and identified 6 300, 6 994 and 7 500 proteins in 60 min, 120 min and 150 min from 293T proteome, respectively. In FAIMS DIA experiment, we used *CV*-45V-65V voltage sweeping, 60 isolation windows, and obtained the optimal proteins identifications and quantitative reproducibility. We quantified 7 019 proteins in 293T cell lysate and 1 077 proteins in depleted plasma in 60 min gradient. The combination of APD on, "Precursor Fit" threshold of 70 and Turbo TMT could simultaneously increase the number of protein identifications and the accuracy of TMT quantification. 10 989 peptides and 2 162 protein were quantified in TMT11-plex labeled yeast proteome. Taken together, the various LFQ methods and TMT quantification method optimized in this study showed high performance and various applications in quantitative proteomics.

Key words quantitative proteomics, label-free quantitative, TMT, FAIMS, Orbitrap Exploris 480

DOI: 10.16476/j.pibb.2020.0245

* This work was supported by a grant from National Science and Technology Major Projects(2018ZX10302205-003).

** Corresponding author.

Tel: 86-29-88302411, E-mail: guanfeng@nwu.edu.cn

Received: July 20, 2020 Accepted: September 9, 2020