

重组蛋白的HCP定性和定量分析方法思考——从前处理到数据分析

龙珍 齐英姿 祝翔

赛默飞世尔科技（中国）有限公司

关键词：重组蛋白，HCP，定性，定量，软件，前处理，超高分辨质谱，DIA，DDA

1. 引言

重组蛋白类药物由遗传修饰的原核或真核宿主细胞培养/发酵产生。在目标蛋白的发酵过程中，由于宿主细胞的凋亡/死亡/裂解，宿主细胞产生的其它蛋白质也可能释放到细胞培养基/发酵液中，即为宿主细胞蛋白（HCP）。HCP可能会引起机体免疫反应，从而产生类佐剂作用——增强机体对药物的免疫应答。HCP也可能是一些酶类，具有分解目标蛋白或人体蛋白的作用，导致药物不稳定或进入人体后引发一系列不良反应。因此，生物制药公司不断优化纯化工艺，从而控制终产品中HCP的种类和含量。

目前常用的HCP检测方法为酶联免疫法（ELISA）。该方法具有较高的灵敏度，但也存在一些局限性。LC-MS方法是ELISA的有利补充。与ELISA相比，LC-MS方法具有以下优点：第一，可以鉴定HCP的种类；第二，可以获得单个HCP的含量信息；第三，可用于不同工艺所得产品的横向对比。由于LC-MS的以上优点，LC-MS作为HCP的检测受到广泛关注和广泛使用。例如，Huang等[1]以NIST mab为例，发展了基于LC-MS检测HCP的前处理方法。LC-MS在HCP检测中的广泛使用，推动了相关分离、检测和数据分析技术的发展。但多样的技术选择也为HCP的分析带来一些疑问：

第一、 大部分关于HCP的研究均以抗体为分析对象。在抗体的分析中，native酶解可有效抑制主蛋白的酶切，从而提升HCP的鉴定。在生物制药中，除抗体以外，还有大量的重组蛋白产品，这些产品的酶解应该采用什么前处理方法？

第二、 在HCP的检测中有nanoLC-MS和UHPLC-MS两种分离检测方法，这两种方法各自的优缺点是什么，应该选择哪一种方法？

第三、 常听人说，DIA采集数据有利于提升鉴定量，是否用DIA采集HCP可显著提升HCP鉴定量？

本文以毕赤酵母发酵的某重组蛋白产品为例，展示重组蛋白产品中毕赤酵母残留蛋白的定性和定量分析方法。

2. 实验部分

2.1 样品及前处理

纯化工艺1得到的蛋白样品（1.5 mg/mL）和纯化工艺2得到的蛋白样品（1.54 mg/mL）。

2.1.1 native酶切

- 分别取200 μ L样品和1.5 μ L（USP2标品，200 μ L/支），再加入200 μ L Rapigest混匀，60 $^{\circ}$ C，变性15 min；
- 加入10 μ L 500 mM DTT，60 $^{\circ}$ C还原1 h；
- 冷却到室温，加入10 μ L 1M IAM，避光，反应30 min；
- 将样品转移到10 KDa超滤膜，21,000 g离心10 min，去掉滤过液，加入200 μ L 50 mM碳酸氢铵置换溶液1次，每次置换，去掉滤过液；
- 在超滤管中加入200 μ L 50 mM碳酸氢铵，20 μ L Trypsin（20 μ g/200 μ L，50 mM碳酸氢铵溶解）后，移液枪吹打，混合均匀，37 $^{\circ}$ C胰蛋白酶，酶切过夜；
- 将上述样品加入4 μ L甲酸后，21,000 g离心15 min，取上清液再次用于LCMS分析。

2.1.2 变性酶切

- 分别取200 μ L样品和1.5 μ L（USP2标品，200 μ L/支）混匀，10 KDa超滤膜，21,000 g离心10 min，去掉滤过液，加入200 μ L 50 mM碳酸氢铵置换溶液1次，每次置换，去掉滤过液；
- 在超滤管中加入200 μ L 50 mM碳酸氢铵，20 μ L Trypsin（20 μ g/200 μ L，50 mM碳酸氢铵溶解）后，移液枪吹打，混合均匀，37 $^{\circ}$ C胰蛋白酶，酶切过夜；
- 加入4 μ L甲酸后，混匀，21,000 g离心15 min，取上清液用于LCMS分析。

2.2 仪器

质谱仪器: Orbitrap Exploris 240 (赛默飞世尔科技, 美国);

色谱仪器: Vaquish Flex 液相色谱系统和UltiMate™ 3000 RSLCnano液相系统 (赛默飞世尔科技, 美国);

2.3 仪器方法

2.3.1 nanoLC-MS DDA分析

色谱仪: UltiMate™ 3000 RSLCnano; Trap柱: Acclaim PepMap™ 100Å 75 µm x 20 mm, 2 µm, 货号164535; 分析柱: Acclaim PepMap™ 100Å 75 µm x 250 mm, 2 µm, 货号164941; 柱温: 50 °C; 流动相: A: 0.1% FA-水; B: 0.1% FA: 乙腈 (20:80, v/v); 上样流速: 8 µl/min; 分析流速: 0.3 µl/min; 上样量: 2 µl; 上样泵: 100% A;

分析泵梯度:

时间 (min)	B/%
-5	2
0	2
3	8
98	28
113	35
116	99
125	99

质谱条件:

检测模式	+
喷雾电压	2.4 kV
毛细管温度	320 °C
分辨率设置	一级60,000@m/z 200; 二级15,000@m/z 200
母离子扫描范围	m/z 400-1800
隔离窗口	1.6
MicroScans	1
HCD CE (%)	30
RF-Lens	70

2.3.2 nanoLC-MS DIA采集

仪器硬件、色谱柱、流动相和梯度同2.3.1。

检测模式	+
喷雾电压	2.4 kV
毛细管温度	320 °C
分辨率设置	一级60,000@m/z 200; 二级15,000@m/z 200
母离子扫描范围	m/z 400-1200
隔离窗口	1.6
MicroScans	1
HCD CE (%)	30
RF-Lens	70
窗口	平均分为100个窗口

2.3.3 UHPLC-MS DDA分析

色谱仪: Vanquish Flex; 分析柱: C18 130Å 2.1 μm x 150 mm, 1.7 μm ; 柱温: 60 °C

流动相: A: 0.1% FA-水; B: 0.1% FA-乙腈; 分析流速: 0.2 ml/min; 上样量: 10 μl

梯度:

时间 (min)	B/%
0	3
3	3
90	32
95	90
100	90
100.1	3
110	3

质谱条件:

检测模式	+
喷雾电压	3.8 kV
毛细管温度	320 °C
雾化温度	350 °C
分辨率设置	一级120,000@m/z 200; 二级15,000@m/z 200
母离子扫描范围	m/z 350-2000
AGC	一级300%; 二级100%
(n)CE	30%

2.4 数据采集及分析

Thermo Xcalibur 4.4采集

2.4.1 BioPharma Finder 4.1处理数据

The screenshot displays the BioPharma Finder 4.1 software interface with several configuration panels:

- Basic Parameters:**
 - Enable HCP:
 - Protein Database: uniprot-bichi_jaomu_3A4922_and_USP2
 - Acquisition Type: High-High (high res MS1 and high res MS2)
 - Precursor Mass Tolerance: 20 ppm
 - Ions to Search: a ion, b ion, c ion, x ion, y ion, z ion, NL ion
 - Mass Range (MH+peptide mass range to analyze): 600.0000 to 5,000.0000
 - Identification Filters: E-Value Cutoff: 0.010000000 (all identification less than this value will pass filter)
- Protease Parameters:**
 - Select Protease: Trypsin
 - Protease Termini: Fully-digested, Semi-digested
 - Max Num Internal Miscalveages: 0 0 1 2
 - Enabled Decoy Search: Yes, No
- Advanced Search:**
 - Search by Full MS Only: Yes, No
 - Use MS/MS: Use All MS/MS
 - Maximum Peptide Mass: 11,000
 - Mass Accuracy (ppm): 10
 - Minimum Confidence: 0.80
 - Maximum Number of Modifications for a Decade: 1
- Disulfide Search:**
 - Perform Disulfide Bond Search: Yes, No
 - Allow Free Cys:
 - Maximum Number of Hits: 2,048
 - Maximum Number of Disulfide Bonds: 1
 - Maximum Number of Identical Chains in the Molecule: 2
- Variable Modification:**
 - Max # of variable modification per peptide: 1
 - 1. Variable Side Chain Modification: Deamidation (N) / +0.984 (N), N
 - 2. Variable Side Chain Modification: Deamidation (Q) / +0.984 (Q), Q
 - 3. Variable Side Chain Modification: Oxidation (MW) / +15.995 (M,W), M
 - 4. Variable Side Chain Modification: None

2.4.3 Spectronaut处理数据

版本号: 15.6

Protein Inference Workflow	From protein-db matching	Interference Correction	<input checked="" type="checkbox"/>
Digest Type	Specific	Only Identified Peptides	<input checked="" type="checkbox"/>
Enzymes/Cleavage Rules	<input checked="" type="checkbox"/> Trypsin/P <input type="checkbox"/> Trypsin <input type="checkbox"/> LysC <input type="checkbox"/> LysC/P <input type="checkbox"/> GluC <input type="checkbox"/> chymotrypsin(FLWY)	Exclude All Multi-Channel Interferences	<input checked="" type="checkbox"/>
Inference Algorithm	IDPicker	MS1 Min	2
Toggle N-Terminal M	<input checked="" type="checkbox"/>	MS2 Min	3
Exclude Duplicate Assays	True	Protein LFQ Method	Automatic
Generate Decoys	True	Proteotypicity Filter	None
Decoy Method	Mutated	Major (Protein) Grouping	by Protein Group Id
Preferred Fragment Source	NN Predicted Fragments	Minor (Peptide) Grouping	by Stripped Sequence
Decoy Limit Strategy	Dynamic	Major Group Quantity	Mean peptide quantity
Library Size Fraction	0.1	Major Group Top N	<input checked="" type="checkbox"/>
Machine Learning	Per Run	Max	3
Precursor PEP Cutoff	0.2	Min	1
Precursor Qvalue Cutoff	0.01	Minor Group Quantity	Mean precursor quantity
Protein Qvalue Cutoff (Experiment)	0.01	Minor Group Top N	<input checked="" type="checkbox"/>
Protein Qvalue Cutoff (Run)	0.05	Max	3
Single Hit Definition	By Stripped Sequence	Min	1
Exclude Single Hit Proteins	<input type="checkbox"/>	Quantity MS-Level	MS1
Pvalue Estimator	Kernel Density Estimator	Quantity Type	Area
		Data Filtering	Qvalue sparse
		Imputing Strategy	Global Imputing
		Cross Run Normalization	<input checked="" type="checkbox"/>
		Normalization Filter Type	None

3. 结果与讨论

3.1 Native酶解和变性酶解

对比native酶解和变性酶解的BPC图，如图1所示。在变性酶解中，主蛋白酶解更为充分（强度约为native酶解的3倍）。在变性酶解所得样品的BPC图中，由于主蛋白肽段过载，导致主蛋白肽段呈现出较大的拖尾因子（ $As > 2.0$ ）和较宽的峰宽（约5 min）。主峰拖尾和展宽导致部分小峰信号被掩盖，如保留时间在50-60 min的信号。

BPF检索并对比native酶解和变性酶解所得样品中HCP的检出情况，如图2所示。变性酶切和native酶切各13和29种HCP被鉴定到（包括添加的标准蛋白）。其中添加的8种标准蛋白在native酶切中共6种被检出，在变性酶切中仅2种被检出。在变性酶切所得样品检出的13种HCP中，有10种在native酶切样品中检出。

综合BPC图信息和HCP检出情况信息可知，native酶切可限制主蛋白的酶解，减少主蛋白信号对HCP信号的抑制，实现更多HCP的检出。

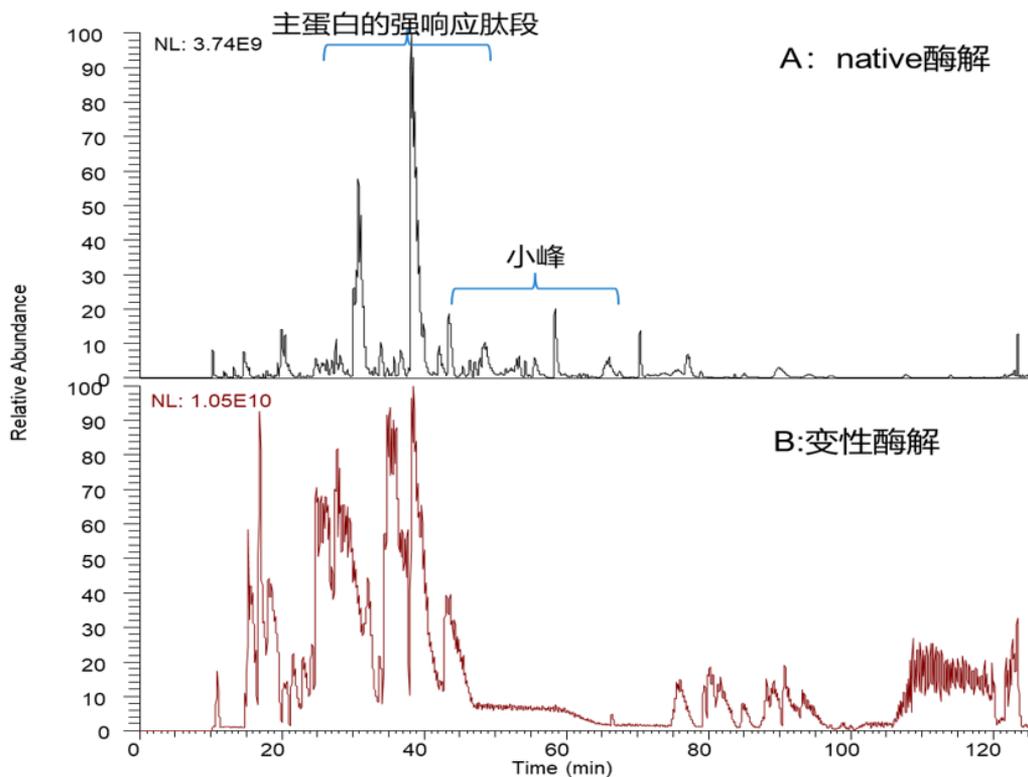


图1 native酶解和变性酶解所得样品BPC图

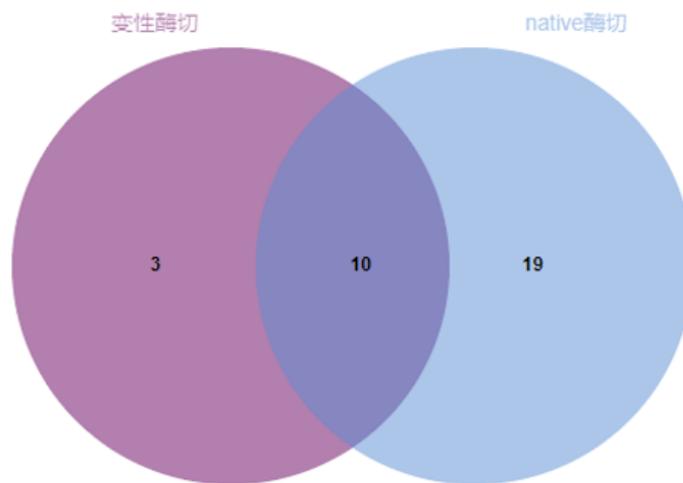


图2 变性酶解和native酶解鉴定HCP比较

3.2 nanoLC-MS和UHPLC-MS分析

一方面，Nano源与常规源相比具有更高的雾化效率，理论上nano源的雾化效率是常规源的6400倍。另一方面，相比UHPLC (1.7 μm)，nanoLC由于使用更大粒径的色谱填料 (5 μm)，更小的流速等，色谱峰峰宽普遍比UHPLC-MS大，从而降低了nanoLC-MS的响应。那么，nanoLC-MS和UHPLC-MS相比，两者在HCP检测中有什么异同呢。本节继续以毕赤酵母生产的某重组蛋白中HCP的检测为例，对比nanoLC-MS和UHPLC-MS在HCP检测中的使用。

如图3所示，尽管nanoLC-MS (2 μL) 的进样量仅UHPLC-MS (10 μL) 的五分之一，但BPC图响应是UHPLC-MS BPC图的约4倍。以BPF为软件检索两种条件所得数据并对比HCP检出情况，如图4所示。NanoLC-MS由于具有更高的灵敏度，检出的HCP个数比UHPLC高出近一倍。以蛋白A0A1B2JBG9为例对比两种分离检测方法，UHPLC-MS的检出强度为1.32E+05，nanoLC-MS的检出强度为7.26E+06，两者相差55倍。因此，虽然nanoLC-MS由于柱效等原因导致灵敏度降低，但nanoLC-MS的灵敏度仍显著高于UHPLC-MS，更有利于HCP的检测。

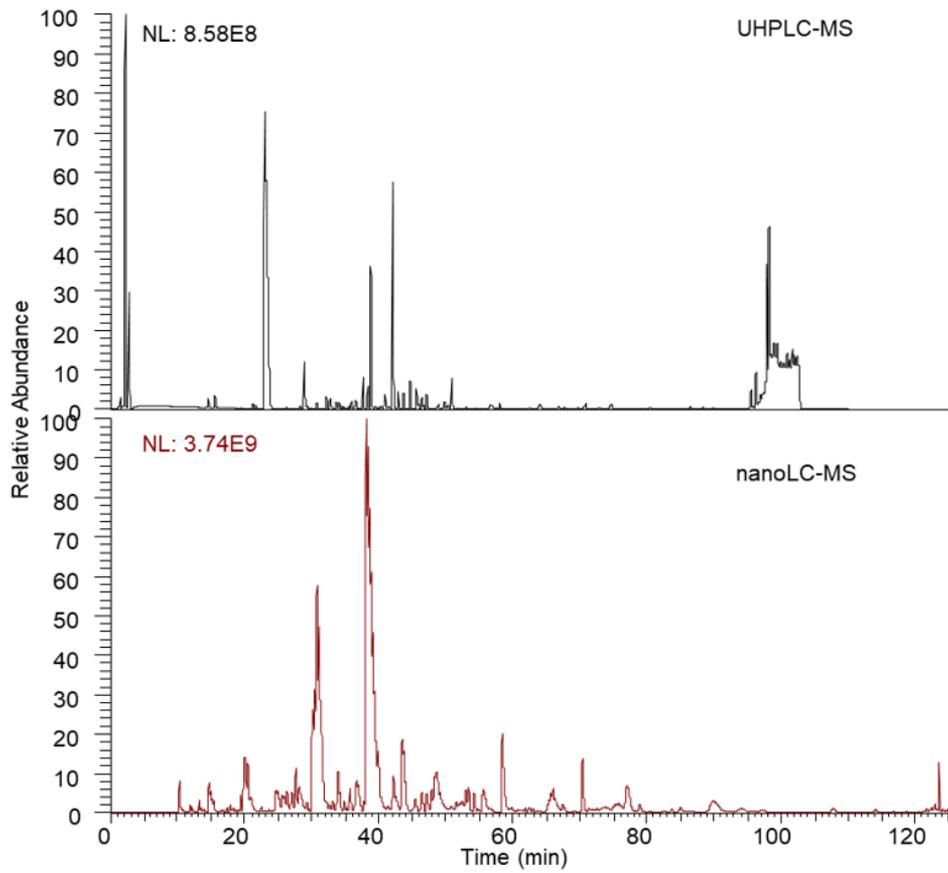


图3 对比UHPLC-MS和nanoLC-MS分析native酶解的某重组蛋白

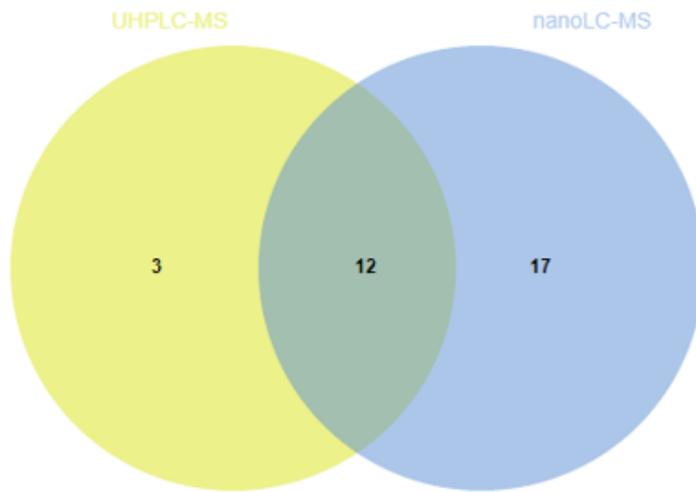


图4 对比UHPLC-MS和nanoLC-MS分析native酶解样品中HCP

3.3 DIA和DDA分析HCP

以Spectronaut软件为检索平台，对比DIA和DDA采集的数据，两者在HCP的检出个数上无显著差异，且结果高度一致（86%重叠）。因此，对组成较为简单的HCP样品体系，DIA与DDA在HCP的鉴定上没有显著差异。

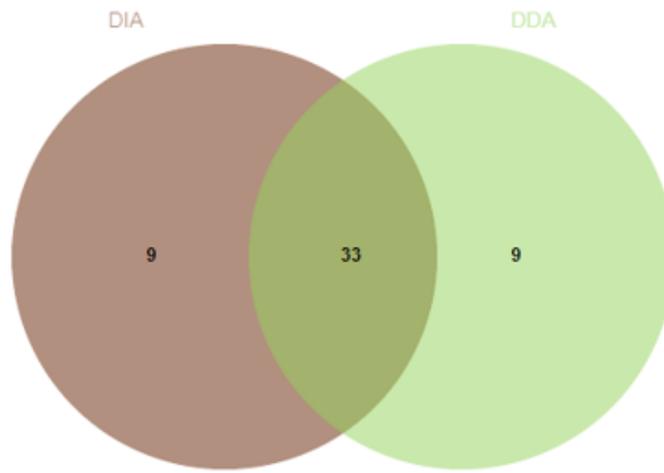


图6 DDA和DIA分析HCP数据结果对比

3.4 HCP定量

如3.3节所示，当unique peptide卡值为1时，检索结果的随机性较大，导致不同软件检索结果差异较大。因此，在HCP定量中，建议卡值unique peptide为大约等于2。用PD软件所得abundance或者BPF软件中HCP的MS area数据作为蛋白的Intensity计算HCP含量，公式如下：

$$\frac{\text{标准蛋白 n 的物质的量}}{\text{HCP(n)的物质的量}} = \frac{\text{标准蛋白 n 的 Intensity}}{\text{HCP(n)的 Intensity}} \quad \text{公式1}$$

公式1的出处是Jeffrey C. Silva的文献[2-4]：用已知蛋白的响应除以物质的量，得到一个常数（universal signal response factor）。未知蛋白的物质的量等于这个蛋白的强度除以这个常数，翻译出来就是公式1。蛋白的响应用这个肽段被检测到的丰度最高的肽段计算。注意，参与计算的peptide必须为unique，unique peptide数目在HCP定量中，建议不少于2)。

不能只用1个标准蛋白，需要用多个。出处是Huang的文献[1]。在这篇文献中，用了3个蛋白。在本文档的HCP计算中，我们使用了8个标品。有6个标准品被检测到。其中有3个HCP标品的unique peptide不小于2。每个符合要求的HCP标品都参与计算，然后得到一个HCP(n)的含量范围，如1-5。最后，HCP(n)的含量为这个范围的平均值（2.5）。

文档中呈现的HCP通常以ppm为单位，它的计算按公式2进行。其中，HCP(n)的物质的量用公式1计算，单位为pmol。Mn是n蛋白的摩尔质量，单位为g/mol，主药的含量为前处理所用蛋白质量，单位为μg。

$$\text{HCP(n)含量 (ppm)} = \frac{\text{HCP(n)含量(公式1计算结果)} * \text{Mn}}{\text{主药含量 (如本操作流程中前处理中写的取 300 } \mu\text{g 前处理)}} \quad \text{公式2}$$

用上述公式处理nanoLC-MS采集、BPF分析的HCP数据，如表1和图7所示。NanoLC-MS具有极高的灵敏，进一步优化nanoLC色谱柱，改善峰形、提高柱效，可检测到含量更低的HCP。

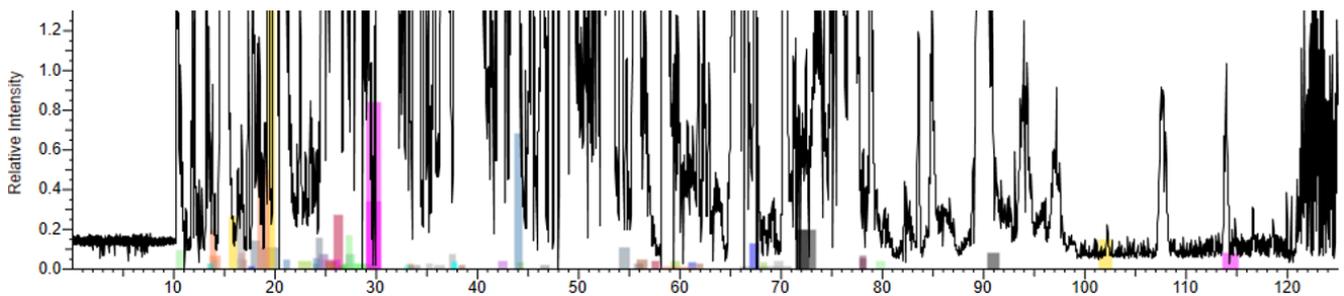


图7 总离子流图中HCP的检测（彩色标出的区域）

表1 某重组蛋白中残留HCP（毕赤酵母蛋白）定量结果

Protein ID	Protein Mass	Unique peptide	Average MS Area	ppm
AOA1B2JG30	45732.9	3	1.07E+06	59.0
AOA1B2JBG9	36962.4	3	7.26E+06	323.8
P00915ups	28738.8	3	5.70E+05	-
AOA1B2JHD1	74126	3	1.45E+06	129.7
B3VDJ1	12124.8	3	3.48E+06	50.9
P00918ups	29114.6	3	6.45E+05	-
P01127ups	12294.4	3	5.68E+06	-
AOA1B2JHJ8	57381.9	3	6.13E+05	42.4
AOA1B2JDU3	36069.4	2	4.65E+06	202.4
V9H1K4	31515	2	3.77E+05	14.3
AOA1B2J755	35915.6	2	1.71E+06	74.1
Q56D08	74213.1	2	5.83E+05	52.2

-: HCP标品

3.5 纯化工艺对HCP种类的影响

建立上述HCP检测方法后，将其用于纯化过程的监控。如图8所示，卡值unique peptide为2，纯化前共35种HCP检出。纯化过程中有29种HCP被去除，有6种蛋白未被去除。纯化过程引入3种HCP，分别是AOA1B2JDU3、B3VDJ1和V9H1K4。需要对工艺工程进行讨论，以便确定这三种蛋白的来源。



图8 纯化前和纯化后样品中HCP的检出情况

结论

本文系统对比了变性酶解和native酶解、nanoLC-MS和UHPLC-MS、DIA采集和DDA在HCP分析中的异同。相对于变性酶解，native酶解更有利于HCP的检出。相对于UHPLC-MS检测，在本案例中，nanoLC-MS的灵敏度更高（对丰度最高的HCP，两者相差55倍）。相对于DDA采集，DIA采集在本案例中没有显著优势。此外，本文还详细解释了HCP计算的流程、公式，并对公式出处和公式中的每一项做了详细解释，以便读者参考和查证。最后，我们将优化所得方法用于纯化工艺的监控，为重组蛋白中HCP的检测提供思路。

参考文献

- [1] L. Huang, N. Wang, C.E. Mitchell, T. Brownlee, S.R. Maple, M.R. De Felippis, A Novel Sample Preparation for Shotgun Proteomics Characterization of HCP in Antibodies, *Analytical Chemistry*, 89 (2017) 5436-5444.
- [2] J.C. Silva, R. Denny, C. Dorschel, M.V. Gorenstein, G.-Z. Li, K. Richardson, D. Wall, S.J. Geromanos, Simultaneous Qualitative and Quantitative Analysis of the Escherichia coli Proteome: A Sweet Tale *, *Molecular & Cellular Proteomics*, 5 (2006) 589-607.
- [3] J.C. Silva, M.V. Gorenstein, G.-Z. Li, J.P.C. Vissers, S.J. Geromanos, Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition *S, *Molecular & Cellular Proteomics*, 5 (2006) 144-156.
- [4] J.C. Silva, M.V. Gorenstein, G.-Z. Li, J.P.C. Vissers, S.J. Geromanos, Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition * S, *Molecular & Cellular Proteomics*, 5 (2006) 144-156.



赛默飞
官方微信

热线 800 810 5118
电话 400 650 5118
www.thermofisher.com

thermo scientific