

填充型纳流液相色谱长柱的蛋白质组学分析性能考察

王晓飞¹ 秦飞² 郭睿³ 张博*¹

¹(厦门大学化学化工学院, 厦门 361005) ²(厦门医学院, 厦门 361023)

³(厦门宜柱科技有限公司 厦门 361102)

摘要 纳流液相色谱长柱用于蛋白质组学分析中,可以显著提高多肽的分离度和检出率,进而提高蛋白质的鉴定数目和可信度。本研究基于纳流液相色谱-电喷雾质谱联用平台,考察了柱床长度为 1.5 m 的超长纳流液相色谱柱,将其用于蛋白质组学分析中。与常规纳流液相色谱柱相比,使用长柱时,复杂多肽样品得到了优良的分离效果。蛋白质组学数据检索结果表明,使用 1.5 m 长柱进行前端的色谱分离,采用质谱识别肽段数提升了 83.6%,肽段数目提升了 74.4%,唯一肽段数目提升了 75.0%,蛋白质鉴定数目提升了 38.4%。研究结果表明,这类填充型纳流液相色谱长柱为高分辨蛋白质组学提供了一种新的分离工具。

关键词 纳流液相色谱;毛细管液相色谱;填充柱;蛋白质组学;质谱

1 引言

近年来,蛋白质组学技术获得了蓬勃的发展,并在众多领域得到了广泛应用,这主要得益于面向高度复杂和多样化生物样品分离检测技术水平的提高。在蛋白质组学分析过程中,需对高度复杂蛋白质酶解物进行分离,因此对分离技术的分辨能力提出了巨大挑战。纳流液相色谱具有柱效优良、灵敏度高、洗脱流速低、样品需求量少、易与质谱偶联等优势,已成为复杂多肽样品分离分析中最重要的工具^[1]。

开发高效的纳流液相色谱柱,能够极大提升多肽的分离度和鉴定数目,进而可以增加低丰度蛋白质鉴定的可能性,为蛋白质组学分析提供深度全面的数据信息。依据色谱理论,减小填料尺寸和柱管内径,以及增加柱床长度是提高液相色谱柱峰容量和分离性能的有效方法,商业化单分散 1~3 μm 色谱填料应用广泛,同时,超细内径 (ID=15 μm) 纳流液相色谱柱也得到了应用^[2,3]。与继续减小填料粒径和毛细管内径相比,增加柱床长度是更简单、直接且有效的方法^[4,5]。Shen 等^[6]采用匀浆法制备了 30 μm 内径、柱长为 87 cm 的纳流液相色谱长柱,并用于酵母菌酶解肽段的分离,获得了 1100 的高峰容量。Smith 课题组制备了一系列内径和长度的纳流液相色谱长柱,并对纳流液相色谱长柱在复杂多肽样品分离中的性能和方法进行了探究^[3,7,8]。Wang 等^[9]采用组装的方式,将 2 根 7.5 cm 与 3 根 15.0 cm 的壳核型固定相填充的毛细管柱联接 (60 cm),以牛血清白蛋白的酶解肽段作为测试样品,获得了大于 500 的峰容量。Thomas 等^[10]在不同洗脱条件下对比了内径 75 μm 、柱床长度分别为 25 和 50 cm 的商品化纳流液相色谱柱,通过对 HeLa 细胞蛋白酶解物进行分离鉴定,结果表明,峰容量越高,获得的蛋白质鉴定数越多。Thakur 等^[11]选用 1.8 μm 的 C_{18} 色谱填料,制备了柱长 50 cm,内径 75 μm 纳流液相色谱长柱,分离鉴定了复杂的酵母蛋白组样品。3 次运行总共鉴定 2990 种酵母蛋白质,约覆盖 68% 酵母蛋白组,并指出:使用纳流液相色谱柱越长,单次运行获得的蛋白质信息越多。de Vos 等^[4]通过串联 6 根柱长 15 cm、内径 2.1 mm 的核壳型色谱柱,最终获得 ~1360 的超高峰容量。

这些研究表明,基于纳流液相色谱长柱的超高分辨率,显著提高了多肽的质谱检出数目,进而有利于蛋白组学样品信息的深度挖掘,并显著提高蛋白序列鉴定的覆盖率,因此毛细管液相色谱长柱的开发与应用获得了广泛关注^[12]。但是,以上均为各研究组自行制备的毛细管液相色谱长柱,目前还未见商品化的 1 m 级长柱,这显然限制了更高色谱分辨率的获得,以及高分辨蛋白质组学的发展。工业化蛋白质组学分析是未来精准医疗的重要支撑技术,高分辨纳流液相色谱长柱是重要的分离工具。

2018-11-24 收稿;2018-12-11 接受

本文系国家自然科学基金项目 (Nos. 21475110, 21535007, J1310024)、中国博士后基金项目 (No. 2018M642572)、福建省中青年骨干教师教育科研项目 (No. JT180645)、厦门市科技计划项目 (No. 3502Z0173019) 和中央高校基本科研业务费 (No. 20720160051) 资助

* E-mail: bozhang@xmu.edu.cn

基于单颗粒柱塞技术,研究者已发展了高效毛细管填充柱的制备工艺,并已广泛用于纳流液相色谱与毛细管电色谱^[13-16]。除了经典的全填充柱型,还有短柱床^[17]、超长柱长^[18]、多级柱床^[19]等多种柱型,为微纳尺度液相分离提供了丰富工具,且部分柱型已实现了商品化。

在本课题组前期的工作^[18]中,针对一米长纳流液相色谱柱的制备和色谱性能评价做了系统研究,并获得了高峰容量(800),预示了其在复杂生物体系(如蛋白质组学)分析中的应用前景。本研究考察了长1.5 m商品化纳流液相色谱长柱的蛋白组学分析性能。以HeLa细胞蛋白酶解物为样品,对此长柱在蛋白质组学中高效分离能力进行评价。结果表明,与常规纳流液相色谱柱相比,此长柱具有分辨率高、多肽和蛋白鉴定数目多、结果更可靠的优点,有利于高分辨蛋白质组学研究中的信息深度挖掘。

2 实验部分

2.1 仪器与试剂

ACQUITY Nano-UPLC(美国 Waters 公司)/Q-Exactive 质谱仪(美国 Thermo Scientific 公司)联用系统。

HeLa 细胞标准蛋白酶解物(美国 Thermo Fisher 公司);乙腈、甲醇(色谱级,德国 Merck Darmstadt 公司);实验用水为 Milli-Q 系统(美国 Millipore 公司)制备的超纯水;其它试剂至少为分析纯。

纳流液相色谱长柱、Smartube C₁₈柱(100 μm×150 cm, 5 μm)购自厦门宜柱科技有限公司;Acquity Peptide BEH C₁₈常规纳流液相色谱柱(100 μm×10 cm, 1.7 μm, 13 nm, 美国 Waters 公司);Nanoport 接头备件(Upchurch Scientific 公司)。

2.2 实验方法

样品为 HeLa 细胞标准蛋白酶解物,使用 Acquity Nano-UPLC 2G-VIM Trap C₁₈柱(20 mm × 180 μm, 5 μm)作为样品捕集柱。流动相 A 为含有0.1%甲酸的乙腈-水(98:2, V/V)溶液,流动相 B 为含有0.1%甲酸的水-乙腈(98:2, V/V)溶液。梯度洗脱,常规纳流液相色谱柱的流动相洗脱条件是:0~5 min, 2% A; 5~105 min, 10% A; 105~125 min, 35% A; 125~145 min, 65% A; 145~结束 min, 95% A。进样量为 1 μg,在 300 nL/min 体积流量下进行洗脱。1.5 m 长纳流液相色谱柱的流动相洗脱条件为:0~20 min, 2% A; 20~420 min, 10% A; 420~500 min, 45% A; 500~540 min, 65% A; 540~580 min, 95% A; 580 min~结束, 95% A; 进样量为 1 μg,在 300 nL/min 体积流量下进行洗脱。

利用 Waters ACQUITY Nano-UPLC/Thermo Q-Exactive Mass Spectrometer 液质联用系统进行蛋白质组数据采集:使用电喷雾离子源(ESI),在正离子扫描模式下采集数据,扫描范围 m/z 350~1800, Full MS 分辨率 70000,自动增益控制目标设置为 1×10^6 ,允许最大注入时间为 50 ms; 单次选择 15 个丰度最高的肽离子用于二级检测;二级质谱采集分辨率为 17500,自动增益控制目标设置为 1×10^5 ,允许最大注入时间为 60 ms; 前驱体选择质量窗口为 m/z 2.0; 二级碰撞能量为 28%,动态排除窗口时间为 30 s。

从 UniProt 下载人类蛋白质数据库(Reviewed(Swiss-Prot) UP000005640(20397)),使用 Proteome Discoverer 1.4 进行蛋白质搜库鉴定,参数设置如下:Min Peptide length: 4; Max Peptide length: 30; Precursor Mass Tolerance: 10 ppm; Fragment Mass Tolerance: 0.8 Da。

3 结果与讨论

3.1 色谱分离结果

在蛋白质组学研究中,“Single shot proteomics”研究策略逐渐被广泛应用,该策略使用单根色谱柱进行较“缓”的梯度洗脱,并结合快速的质谱扫描检测,从而获得高的分离度和峰容量^[20,21]。Kocher 等^[22]的研究表明,蛋白质鉴定数目与峰容量具有线性相关性,且相同洗脱条件下,柱床较长时液相色谱柱具有更高的峰容量。

本研究中采用了常规的 10 cm 纳流液相色谱柱和 1.5 m 纳流液相色谱长柱进行对比。由于 1.5 m 纳流液相色谱柱长度显著长于常规柱,因而可以施加更长且更缓的梯度洗脱程序。在本研究中,HeLa 细胞蛋白酶解样品在长柱上获得了近 10 h(580 min)的洗脱,而短柱梯度时间仅为 145 min。图 1 是两次分离分析对应

的总离子流色谱图 需要指出 两图是在相同的时间尺度(横轴)下进行对比。可以看出,在长柱上,复杂的多肽组分能够在 580 min 宽的时间窗口进行分布,尤其是 100~550 min 宽的窗口空间均被多肽覆盖占据。显然,450 min 的时间窗口已被充分利用,并且对于复杂的多肽样品而言,这个宽度仍然是不够用的。对于常规纳流柱,同样的复杂多肽组分只能拥挤在 145 min 的时间窗口内,其中 50~145 min 的窗口空间是被满覆盖的。

对于如此复杂的多肽混合物,本研究采用的长柱和短柱都未实现所有组分的基线分离。然而,在长柱上,整体的分辨率显著更高,且组分分布的时间窗口更宽,这对于后续的质谱检测有重要意义。因为现代高分辨生物质谱(如这里使用的 Thermo Q-Exactive 系统)其分辨能力很大程度上来源于质谱仪的扫描速度。在质谱仪的扫描速度固定的情况下,样品越缓慢地进入(即进样的窗口越宽),越有利于质谱的充分扫描。这显著降低了质谱对多肽分子的漏检率和离子化抑制现象,从而提高多肽离子的检测数量,尤其是显著提高了低丰度多肽离子的检出率和质谱信号值^[23~26],可挖掘出更多的多肽分子信息,用于多肽/蛋白的最终鉴定。

因此,通过选择使用更长柱床的纳流液相色谱柱和较缓梯度洗脱条件,可以获得高峰容量、高分辨率的前端分离,从而有效地降低多肽混合物样品的复杂程度,显著降低后端质谱检测的压力,使质谱数据更能真实地反映出多肽信息。

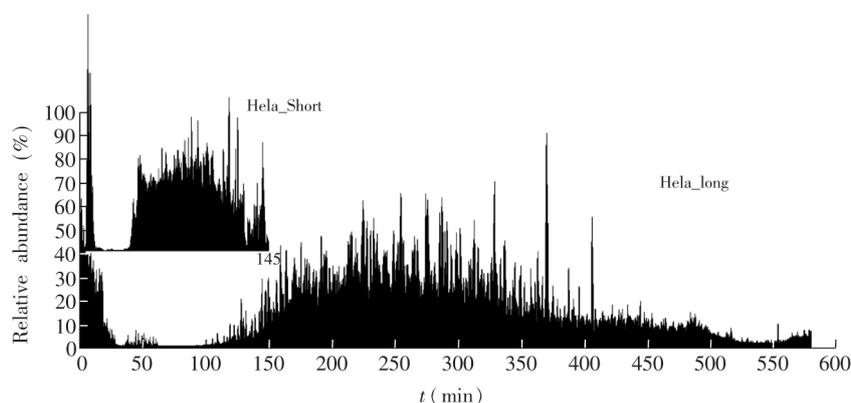


图 1 HeLa 细胞蛋白质酶解物在 1.5 m 和 0.1 m 纳流液相色谱柱(插图)上分离的总离子流色谱图

Fig.1 Total ion chromatograms of HeLa protein hydrolysate with 1.5 m long and 0.1 m (inset) columns

3.2 蛋白质组学鉴定结果

采用人工注释且非冗余的蛋白序列数据库(Reviewed(Swiss-Prot)UP000005640(20397)),使用软件 Proteome Discoverer 1.4 对所得多肽质谱数据进行蛋白质鉴定,综合评价和对比了商品化 10 cm 柱和 1.5 m 长柱的多肽和蛋白质鉴定能力。

如图 2 所示,与常规纳流色谱柱相比,长柱质谱数据中的识别肽段数(Peptide spectrum matches, PSMs,显示蛋白质的已识别肽段序列数)为 33283,短柱为 18130,提升了 83.6%;长柱质谱数据中的肽段数目(Peptides,显示蛋白质组的不同肽段序列数)为 18665,短柱为 10705,提升了 74.4%;长柱质谱数据中的唯一肽段数目(Unique Peptides,显示对蛋白质组唯一的肽段序列数)为 17384,短柱为 9935,提升了 75.0%;长柱质谱数据中的蛋白质组数目(Protein groups,肽段所属的蛋白质组的数量)为 3681,短柱为 2660,提升了 38.4%。以上结果表明,1.5 m 长柱在蛋白质组学分析平台中的应用,显著提高

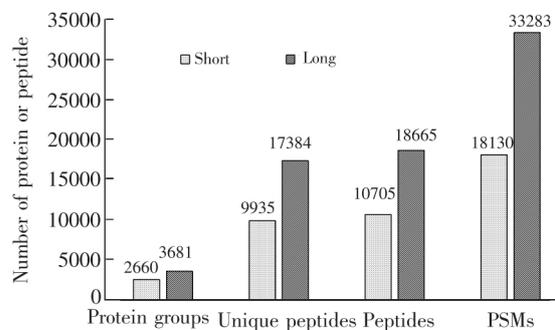


图 2 长柱与常规短柱的蛋白质组学搜库结果对比

Fig.2 Comparison of proteomics search results between 1.5 m long and 0.1 m short nanoLC columns

PSM: Peptide spectrum matches

了质谱对多肽分子的检出率和准确度,最终使得更多蛋白质被更精确地鉴定。

如图3所示,将长柱与短柱所鉴定出多肽和蛋白质组的种类进行比较,结果表明,共同鉴定出的多肽数占短柱鉴定多肽总数的73.65%,蛋白质组的共同鉴定数占短柱鉴定总数的82.82%,以上结果表明,长柱的应用可以实现对短柱所鉴定多肽和蛋白质组的大比例覆盖。

蛋白质组鉴定结果的得分(Score)和蛋白质的肽段覆盖率(Coverage)直接决定了质谱数据的质量。如图4所示,利用1.5 m长柱进行质谱前端分离,不仅增加了蛋白质组的鉴定数目,同时显著增加了得

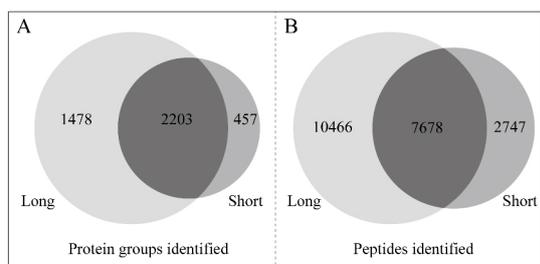


图3 长柱与常规短柱所鉴定多肽和蛋白质组种类覆盖差异比较

Fig.3 Overlap of identified protein groups and peptides between 1.5 m long and short nanoLC columns

A. Protein groups identified; B. Peptides identified.

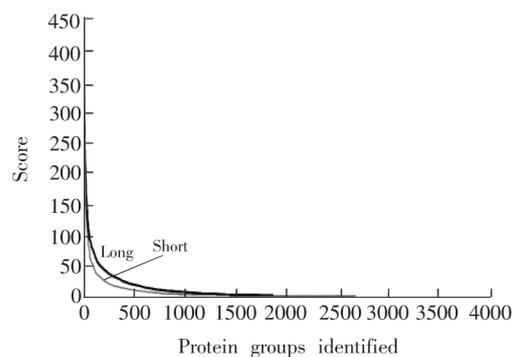


图4 长柱与常规短柱的蛋白质组得分对比

Fig.4 Comparison of protein groups score between 1.5 m long and short nanoLC columns

分高的蛋白质数目。由图5A可见,基于长柱的蛋白质组学鉴定结果中,蛋白质组的覆盖率总体比常规短柱高;考察了每隔10%覆盖率区间中蛋白质组数目在所有蛋白质组中所占比例的分布情况(图5B),结果表明,在长柱鉴定的结果中,高覆盖率的蛋白质组所占比例比常规短柱鉴定结果中更大。得分越高和多肽覆盖率越大,表明质谱检测数据质量越高,所得蛋白质数据的可信度越高。尽管目前蛋白质数据库中包含了上万个已经发现的蛋白质信息,但是生物体内仍有大量未未知的蛋白质组成,在此1.5 m长柱的分析结果中包含了更多(1.5 m长柱212705种,常规短柱76649种,增加了177.5%)未知肽段信息,这可以为未知蛋白和低丰度蛋白的发现和鉴定提供有价值的情报信息。

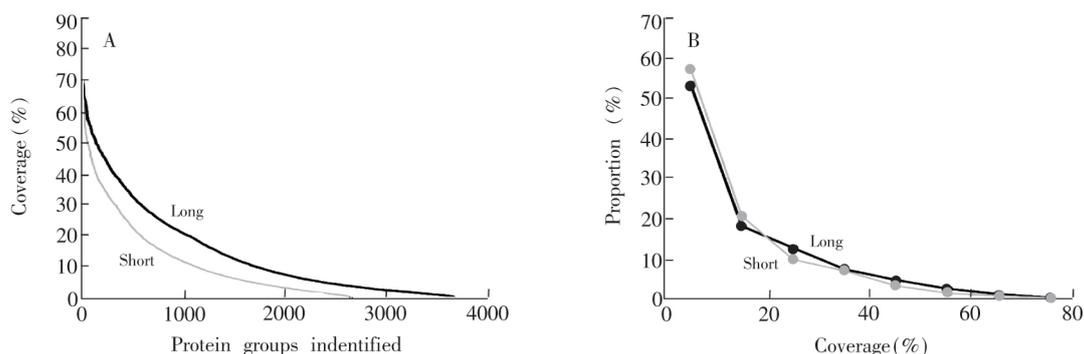


图5 长柱与常规短柱的蛋白质组肽段覆盖率对比:(A)蛋白质组中覆盖率分布;(B)蛋白质组覆盖率比例

Fig.5 Comparison of protein groups coverage between 1.5 m long and short nanoLC columns

(A) Coverage distribution in the protein groups; (B) Proportion of protein groups' coverage.

4 结论

考察了1.5 m长纳流液相色谱长柱的蛋白质组分析性能。结果表明,1.5 m长柱具有蛋白质鉴定数目更多和鉴定结果更可靠的优势,在蛋白质信息深度挖掘和未知蛋白质发现中具有很大应用潜力,这也充分体现了纳流液相色谱长柱在蛋白质组学分析方面的优越性以及发展长柱的必要性。此类单维液相

色谱长柱/质谱联用技术能有效提高复杂多肽组样品的分离性能与检测结果的准确性 随着一米级长柱制备技术的不断发展,单维液相长柱色谱/质谱联用技术有望成为蛋白组学信息深度挖掘最为有效的分析工具。

References

- 1 Xie F, Smith R D, Shen Y. *J. Chromatogr. A*, **2012**, 1261(10): 78–90
- 2 Shen Y, Zhao R, Berger S J, Anderson G A, Nestor Rodriguez A, Smith R D. *Anal. Chem.*, **2002**, 74(16): 4235–4249
- 3 Shen Y, Tolić N, Masselon C, Pasa-Tolić L, Camp D G, Hixson K K, Zhao R, Anderson G A, Smith R D. *Anal. Chem.*, **2004**, 76(1): 144–154
- 4 de Vos J, Stassen C, Vaast A, Desmet G, Eeltink S. *J. Chromatogr. A*, **2012**, 1264(22): 57–62
- 5 Wu C, Liang Y, Liang Z, Zhang L, Zhang Y. *Anal. Chim. Acta*, **2018**, 1019(8): 128–134
- 6 Shen Y, Zhao R, Berger S J, Anderson G A, Rodriguez N, Smith R D. *Anal. Chem.*, **2002**, 74(16): 4235–4249
- 7 Shen Y, Moore R J, Zhao R, Blonder J, Auberry D L, Masselon C. *Anal. Chem.*, **2003**, 75(14): 3596–3605
- 8 Shen Y, Zhang R, Moore R J, Kim J, Metz T O, Hixson K K, Zhao R, Livesay E A, Udseth H R, Smith R D. *Anal. Chem.*, **2005**, 77(10): 3090–3100
- 9 Wang X, Barber W E, Carr P W. *J. Chromatogr. A*, **2006**, 1107(1): 139–151
- 10 Thomas K, Remco S, Karl M. *Anal. Chem.*, **2011**, 83(7): 2699–2704
- 11 Thakur S S, Geiger T, Chatterjee B, Bandilla P, Fröhlich F, Cox J, Mann M. *Mol. Cell. Proteomics*, **2011**, 10(8): M110.003699
- 12 Wang H, Yang Y, Li Y, Bai B, Wang X, Tan H, Liu T, Beach T G, Peng J, Wu Z. *J. Proteome Res.*, **2015**, 14(2): 829–838
- 13 Zhang B, Liu Q, Yang L J, Wang Q Q. *J. Chromatogr. A*, **2013**, 1272(1): 136–140
- 14 Xiao Z L, Wang L, Liu Y, Wang Q Q, Zhang B. *J. Chromatogr. A*, **2014**, 1325(1): 109–114
- 15 Liu Q, Wang L, Zhou Z H, Wang Q Q, Yan L J, Zhang B. *Electrophoresis*, **2014**, 35(6): 836–839
- 16 Yang L J, Xu L J, Guo R, Gao T Y, Guo H X, Yu Y, Lv J J, Wang Q Q, Zhang B. *Anal. Chim. Acta*, **2018**, 1033(11): 205–212
- 17 Liu L N, Zhang B, Zhang Q, Shi Y H, Guo L P, Yang L. *J. Chromatogr. A*, **2014**, 1352(7): 80–86
- 18 Han J, Ye L Q, Xu L J, Zhou Z H, Gao F, Xiao Z L, Wang Q Q, Zhang B. *Anal. Chim. Acta*, **2014**, 852(12): 267–273
- 19 Liu Q, Yang L J, Wang Q Q, Zhang B. *J. Chromatogr. A*, **2014**, 1349(7): 90–95
- 20 Taylor P, Nielsen P A, Trelle M B, Hørring O B, Andersen M B, Vorm O, Moran M F, Kislinger T. *J. Proteome Res.*, **2009**, 8(8): 1610–1616
- 21 Zhou F, Cardoza J D, Ficarro S B, Adelmant G O, Lazaro J B, Marto J A. *J. Proteome Res.*, **2010**, 9(12): 6242–6255
- 22 Kocher T, Swart R, Mechtler K. *Anal. Chem.*, **2011**, 83(7): 2699–2704
- 23 Yamana R, Iwasaki M, Wakabayashi M, Nakagawa M, Yamanaka S, Ishihama Y. *J. Proteome Res.*, **2013**, 12(1): 214–221
- 24 Reinders J, Sickmann A. *Proteomics Methods and Protocols*. Humana Press, **2009**: 123–141
- 25 Hu S, Michels D A, Fazal M A, Ratisoontorn C, Cunningham M L, Dovichi N J. *Anal. Chem.*, **2004**, 76(14): 4044–4049
- 26 Negm R S, Mukesh V, Sudhir S. *Trends Mol. Med.*, **2002**, 8(6): 288–293

Performance of Long Packed Nanoflow Liquid Chromatographic Columns in Proteomics Analysis

WANG Xiao-Fei¹, QIN Fei², GUO Rui³, ZHANG Bo^{*1}

¹(College of Chemistry & Chemical Engineering, Xiamen University, Xiamen 361005, China)

²(Xiamen Medical College, Xiamen 361023, China)

³(Column Scientific Ltd, Xiamen 361102, China)

Abstract Long nanoLC columns have been used in proteomics to improve the resolution and detectability of peptides to enhance protein identification number and analysis confidence. In this study, based on a nanoflow liquid chromatography-electrospray ionization mass spectrometry platform, the performance of a 1.5-m long packed nanoLC column and its applicability in proteomics were investigated. Compared with the commonly used short length nanoLC column, the long packed column achieved excellent resolution for the complex peptide mixture. The proteomics data showed that, using the 1.5-m long nanoLC column, the numbers of peptide spectrum matches, the identified peptides, the unique peptides and the protein groups were increased by 83.6%, 74.4%, 75.0% and 38.4%, respectively. The results demonstrated that the packed long nanoLC columns could be used as a new separation tool for high-resolution proteomics.

Keywords Nanoflow liquid chromatography; Capillary liquid chromatography; Packed columns; Proteomics; Mass spectrometry

(Received 24 November 2018; accepted 11 December 2018)

This work was supported by the National Natural Science Foundation of China (Nos. 21475110, 21535007, J1310024), the China PostDoc Foundation (No. 2018M642572), the Fujian Provincial Education and Research Fund for Young Teachers (No. JT180645), the Xiamen Municipal Science and Technology Plan Project (No. 3502Z20173019) and the Fundamental Research Funds for Central Universities of China (No. 20720160051).