

重组人粒细胞刺激因子的 N 端氨基酸序列异质性分析*

韩春梅, 刘兰, 杨靖清, 陶磊, 范文红, 史新昌**, 饶春明**

(中国食品药品检定研究院, 卫生部生物技术产品检定方法及其标准化重点实验室, 北京 100050)

摘要 目的: 分析重组人粒细胞刺激因子(rhG-CSF)的 N 端异质性。方法: 采用反相超高效液相色谱法分离 rhG-CSF 中 N 端不一致的各组分; 色谱条件: 采用 Waters BEH 300 C₄(2.1 mm×100 mm, 1.7 μm) 色谱柱, 以 0.1% 三氟乙酸水溶液(A)-0.1% 三氟乙酸乙腈溶液(B)为流动相, 梯度洗脱, 流速 0.2 mL·min⁻¹。分别收集各组分并冷冻干燥, 采用 N 端测序仪测定各组分的 N 端氨基酸序列; 液质联用测定 rhG-CSF 的相对分子质量, 对 N 端测序结果进行验证质谱仪采用 Waters 公司 Xevo G2-S 质谱仪, 正离子电喷雾离子源, 毛细管电压 3.0 kV, 锥孔电压 40 V, 去溶剂气体温度 350 °C, 去溶剂气体流速 800 L·h⁻¹, 扫描范围(*m/z*) 500~3 000。结果: 色谱分离得到 3 个组分; N 端氨基酸序列测定结果显示, 3 个组分分别为 rhG-CSF 主成分及 2 个 N 端异质性相关蛋白; 质谱相对分子质量测定结果与 N 端测序结果一致。结论: 应用色谱法及质谱法可对 rhG-CSF 中的 N 端异质性相关蛋白进行分析检测, 检测结果可为国产 rhG-CSF 制品的质量控制和标准提升提供参考和数据支持。

关键词: 重组人粒细胞刺激因子; N 端异质性; 相关蛋白; N 端氨基酸序列; 反相超高效液相色谱; 质谱

中图分类号: R 917

文献标识码: A

文章编号: 0254-1793(2018)11-1887-06

doi: 10.16155/j.0254-1793.2018.11.06

Analysis of N-terminal amino acid sequence heterogeneity of recombinant human granulocyte colony-stimulating factor*

HAN Chun-mei, LIU Lan, YANG Jing-qing, TAO Lei,
FAN Wen-hong, SHI Xin-chang**, RAO Chun-ming**

(National Institutes for Food and Drug Control, Key Laboratory of the Ministry of Health for Research on Quality and Standardization of Biotech Products, Beijing 100050, China)

Abstract Objective: To analyze the N-terminal heterogeneity of recombinant human granulocyte colony-stimulating factor (rhG-CSF). **Methods:** The N-terminal inconsistent components in rhG-CSF were separated by reversed phase ultra high performance liquid chromatography. The chromatographic separation was carried out on

* 药品医疗器械审评审批制度改革专项课题(ZG2017-3-02); 十三五科技重大专项课题“生物类似药质量相似性评价体系建设研究”(2015ZX09501008)

** 通信作者 饶春明 Tel:(010)67095380; E-mail: raocm@nifdc.org.cn

史新昌 Tel:(010)67095604; E-mail: shixc@nifdc.org.cn

第一作者 韩春梅 Tel:(010)67095604; E-mail: hanem@nifdc.org.cn

刘兰 Tel:(010)67095604; E-mail: liulan@nifdc.org.cn

a Waters BEH 300 C₄ column (2.1 mm × 100 mm, 1.7 μm) with the mobile phase consisting of 0.1% trifluoroacetic acid aqueous solution (A)–0.1% trifluoroacetic acid acetonitrile solution (B) in a gradient mode at the flow rate of 0.2 mL · min⁻¹. Each component was collected and freeze-dried, and the N-terminal amino acid sequences of them were determined by N-terminal sequencer. The relative molecular mass of rhG-CSF was measured by liquid chromatography mass spectrometry so as to verify the results of N-terminal sequencing. Waters Xevo G2-S mass spectrometer with electrospray ionization source in positive mode was used, the capillary voltage was 3 kV, Cone voltage was 40 V, gas temperature was 350 °C, gas flow was 800 L · h⁻¹, mass scan range was *m/z* 500–3 000. **Results:** Three components were separated by chromatography, and the results of N-terminal amino acid sequence analysis showed that the three components were rhG-CSF principal components and two N-terminal heterogeneity related proteins, respectively. The results of relative molecular weight determination were consistent with N-terminal sequencing. **Conclusion:** N terminal heterogeneity of rhG-CSF can be analyzed and detected by chromatography and mass spectrometry, which could provide reference and data support for quality control and standard improvement of domestic rhG-CSF products.

Keywords: recombinant human granulocyte colony-stimulating factor; N-terminal heterogeneity; related protein; N terminal amino acid sequence; reverse phase ultra-high performance liquid phase; mass spectrometry

重组人粒细胞刺激因子 (recombinant human granulocyte colony-stimulating factor, rhG-CSF) 是人体内粒细胞刺激因子模拟分子, 与天然 hG-CSF 结构相似, 生物学效应基本一致^[1-2]。

rhG-CSF 在大肠杆菌中表达, 由 174/175 个氨基酸残基组成 (N 末端含有蛋氨酸残基) 则有 175 个氨基酸残基, 相对分子质量为 18.7 k ~ 18.8 k, 等电点 5.5 ~ 6.1, 虽无糖基修饰, 但其生物学活性与天然 hG-CSF 无明显区别^[3]。

rhG-CSF 主要生物学功能是促进中性粒细胞系造血细胞的增殖、分化和活化, 临床上用于治疗放疗、化疗以及骨髓移植后中性粒细胞减少症。rhG-CSF 于 1991 年由美国 FDA 首次批准上市, 我国已批准 15 家制药企业生产 rhG-CSF。

目前, 各家企业生产的 rhG-CSF 的 N 端序列并不完全一致, 多数企业的 rhG-CSF 序列是以蛋氨酸 (M) 起始, N 端 15 个氨基酸残基为 MTPLGPASSLPQSFL; 部分企业是以苏氨酸 (T) 起始, N 端 15 个氨基酸残基为 TPLGPASSLPQSFL。hG-CSF 的理论序列以 T 起始, M 是大肠杆菌表达系统引入的, 带有 M 的 rhG-CSF 实际上是 hG-CSF 的衍生物。这 2 个 N 端序列从 2010 年版《中华人民共和国药典》开始, 都被确认为主药成分。为确定某一 rhG-CSF 制品的 N 端氨基酸序列的异质性, 本文开展了方法学研究。

1 仪器、试剂及样品

N 末端序列分析仪 (PROTEIN SEQUENCER PPSQ-53, 岛津公司); 超高效液相色谱仪 (ACQUITY H-Class, Waters 公司); 高效液相色谱仪 (Alliance 2695, Waters 公司); 紫外检测器 (2489, Waters 公司); 质谱仪 (Xevo G2-S QToF, Waters 公司); BEH300 C₄ 色谱柱 ((2.1 mm × 100 mm, 1.7 μm), Waters 公司); 冻干机 (ModulyoD, Thermo 公司); Symmetry300 C₁₈ 色谱柱 (4.6 mm × 250 mm, 5 μm, Waters 公司); BEH 300 C₄ 色谱柱 (2.1 mm × 50 mm, 3.5 μm, Waters 公司)。

超纯水为密理博超纯水机制备; 乙腈为色谱级, 购自 Fisher Chemical 公司; 三氟乙酸为分析纯, 购自梯希爱 (上海) 化成工业发展有限公司; 甲酸为分析纯, 购自阿拉丁试剂 (上海) 有限公司; 氨基酸测序用标准品及所用实际均购自岛津公司。

重组人粒细胞刺激因子样品为本实验室留样。

2 方法与结果

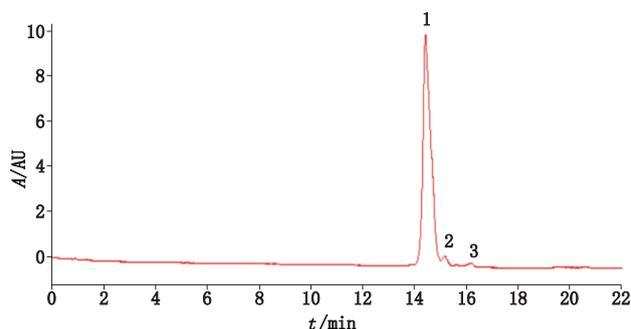
2.1 N 端序列异质体的分离、收集及冷冻干燥

色谱条件: 采用 Waters BEH 300 C₄ (2.1 mm × 100 mm, 1.7 μm) 色谱柱, 柱温 30 °C, 流动相 A 为 0.1% 三氟乙酸水溶液, 流动相 B 为 0.1% 三氟乙酸乙腈溶液, 梯度洗脱 (0~2 min 流动相 B 比例维持 50%, 2~22 min 流动相 B 比例由 50% 变为 60%), 流速 0.2 mL · min⁻¹, 样品保存温度 15 °C, 进样 50 μL, 检测波长 280 nm。根据色谱图检测结果分别收集各色谱峰

的流出液。

冷冻干燥: 将各收集液放入冻干机的离心浓缩组件, 35 °C离心浓缩 1 h; 取出置于 -70 °C冰箱放置 3 h; 取出并迅速放置于冻干机的冷冻干燥组件, 真空干燥 15 h。

液相色谱图见图 1, 峰 1 为主峰, 峰 2 和峰 1 的保留时间相距较近, 未有效地基线分离, 所以在收集峰 2 流出液时, 从峰 2 的后半部分开始收集。



1. 主蛋白 (target protein) 2、3. N 端异质性导致的制品相关蛋白 (product-related proteins caused by N-terminal sequence heterogeneity)

图 1 rhG-CSF 的超高效液相色谱图

Fig. 1 UPLC chromatogram of rhG-CSF

2.2 N 端氨基酸序列测定

应用 N 端氨基酸测序仪内置的 Edman 降解法分析 N 端氨基酸序列^[4-5]。取出收集并冻干的组分, 加入 10 μL 超纯水溶解, 加于玻璃膜上, 氮气吹扫干燥 10 min; 将干燥的玻璃膜组装入反应池, 运行 N 端氨基酸测定程序。仪器自动加载异硫氰酸苯酯与 N 端氨基结合, 在三氟乙酸蒸汽下裂解 N 端第 1 个肽键, 将裂解下来的氨基酸残基在线转化成稳定的 PTH 氨基酸, 过特异的氨基酸分离柱并记录结果。

N 末端测序仪分析结果如图 2~4 (用氨基酸单字符表示), 分别为峰 1、峰 2、峰 3 收集液的 N 端氨基酸序列谱图 (共 3 个循环)。其中峰 1 的纯度较好, 测序结果为 TPL; 峰 2 的总量较少不足 5%, 并受到峰 1 的干扰, 测序结果为 TPL 和 PLG, 扣除峰 1 干扰序列, 峰 2 的序列应为 PLG, 即 N 末端比峰 1 序列少 1 个氨基酸残基 T; 峰 3 的测序结果更为复杂, 可能受到多个方面的影响, 主要的序列为 TPL 和 GYG, TPL 应为峰 1 的干扰, GYG 为部分分泌信肽的序列, 参见质谱检测结果。

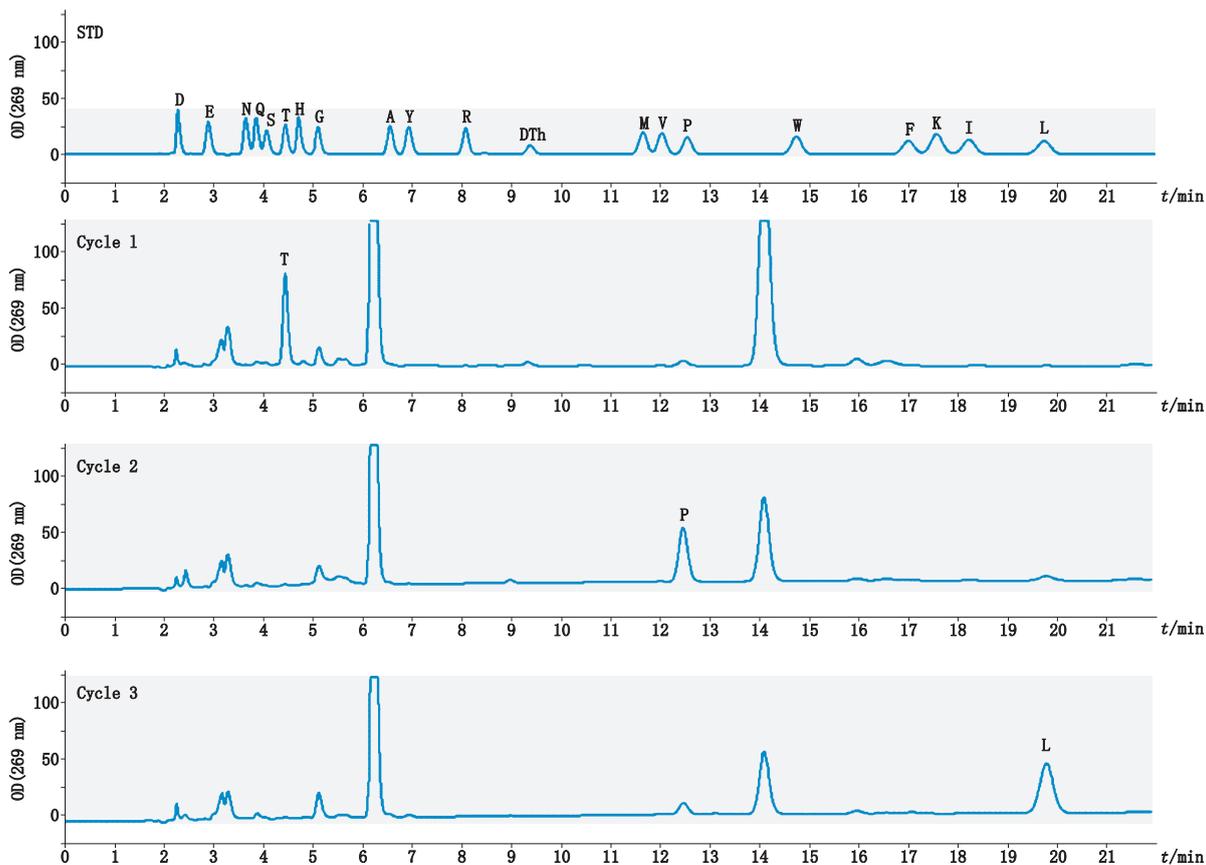


图 2 色谱峰 1 收集液的 N 端氨基酸序列检测结果 (3 个循环)

Fig. 2 The detection results of N terminal amino acid sequence of liquid collected by chromatographic peak 1 (3 cycles)

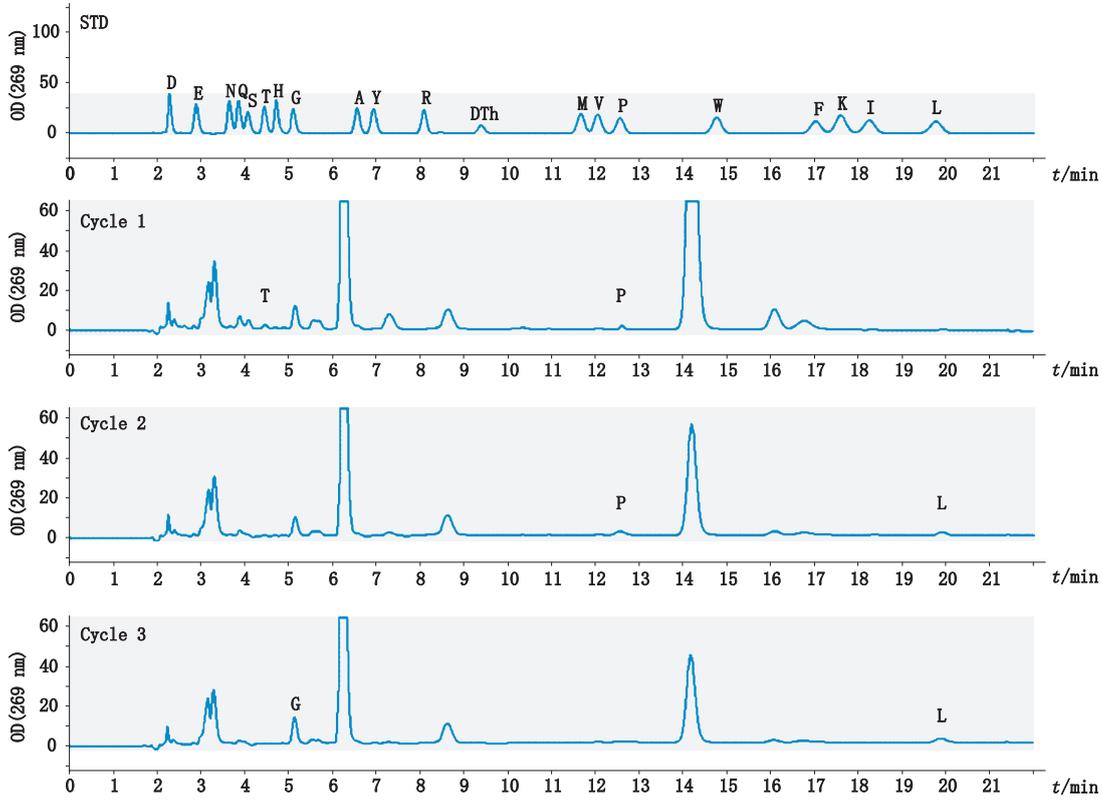


图 3 色谱峰 2 收集液的 N 端氨基酸序列检测结果 (3 个循环)

Fig. 3 The detection results of N terminal amino acid sequence of liquid collected by chromatographic peak 2(3 cycles)

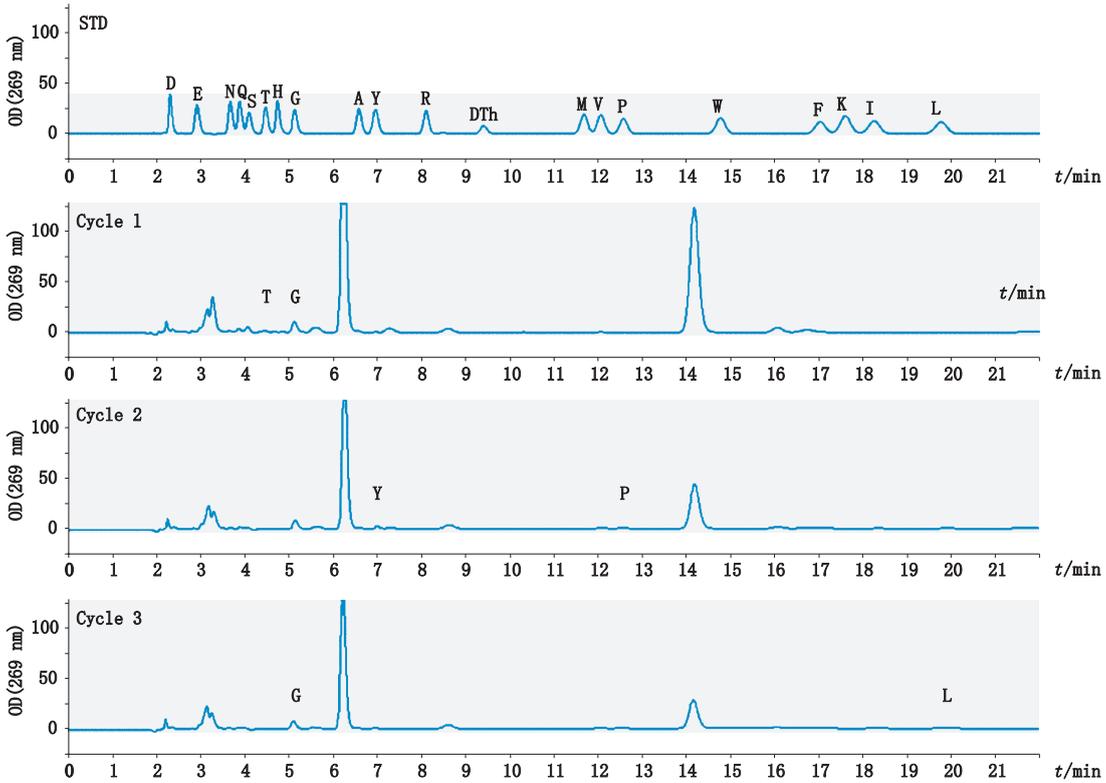


图 4 色谱峰 3 收集液的 N 端氨基酸序列检测结果 (3 个循环)

Fig. 4 The detection results of N terminal amino acid sequence of liquid collected by chromatographic peak 3(3 cycles)

2.3 液质联用测定 rhG-CSF 的相对分子质量

色谱条件: 采用 BEH300 C₄ 色谱柱 (2.1 mm × 50 mm, 3.5 μm), 柱温为 80 °C, 流动相 A 为超纯水, 流动相 B 为乙腈, 流动相 C 为 1% 的甲酸水溶液, 洗脱梯度见表 1; MS 条件: 正离子、分辨率模式采集数据, 毛细管电压 3 kV, 锥孔电压 40 V, 去溶剂气体温度 350 °C, 源温 120 °C, 去溶剂气体流速 800 L · h⁻¹, 扫描范围 (*m/z*) 500~3 000; 样品保存温度 10 °C, 用超纯水将样品稀释至约 1 mg · mL⁻¹, 进样量 10 μL^[6]。

表 1 液质联用分析的洗脱梯度

时间 (time) / min	流速 (flow rate) / (mL · min ⁻¹)	流动相比例 (ratio of mobile phase) /%			曲线 (curve)
		A	B	C	
0.00	0.4	85	5	10	线性 (linear)
1.00	0.4	85	5	10	线性 (linear)
1.01	0.2	85	5	10	线性 (linear)
8.00	0.2	40	50	10	线性 (linear)
8.01	0.4	40	50	10	线性 (linear)

质谱检测结果见图 5, 共有 3 个峰, 其中主峰相对分子质量为 18 667.20, 与以 TPL 为起始序列的 rhG-CSF 的相对分子质量一致; 18 565.80 与主峰的相对分子质量差值为 101.40, 与苏氨酸 (T) 残基的相对分子质量一致, 故 18 565.80 为以 PLG 为起始序列的 rhG-CSF 的相对分子质量; 18 944.40 与主峰相对分子质量的差值为 277.20, 与 N 端氨基酸序列检测的 GYG 三肽残基 (见图 4) 的相对分子质量 277.28 一致, 故应为以 GYGTPL 为起始序列的 rhG-CSF 的相对分子质量。图 5 中丰度较低的、靠近主峰的小峰相对分子质量为 18 724.95, 与主峰相对分子质量的差值为 57.29, 即 1 个甘氨酸残基相对分子质量 57.05, 所以 18 724.95 应为以 GTPL 为起始序列的 rhG-CSF 的相对分子质量。从总体上看, 这是表达系统信号肽切除时留下的痕迹, 即在主切割位点附近, 仍然留存非特异性切割位点, 产生的蛋白起始序列呈现了梯次形态, 及 GYGTPL...、YGTPL... (未检出)、GTPL... (占比较低)、TPL... 和 PL...。

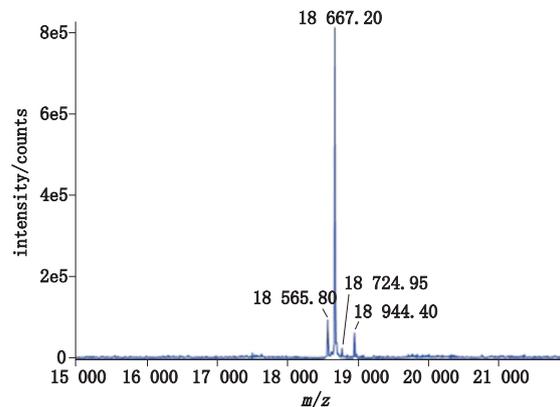


图 5 质谱检测 rhG-CSF 的相对分子质量

Fig. 5 Relative molecular mass of rhG-CSF detected by mass spectrometry

2.4 反相高效液相色谱检测 rhG-CSF 的纯度

色谱条件: 采用 Symmetry 300 C₁₈ (4.6 mm × 250 mm, 5 μm) 色谱柱, 柱温 30 °C, 流动相 A 为 0.1% 三氟乙酸水溶液, 流动相 B 为 0.1% 三氟乙酸乙腈溶液, 梯度洗脱 (0~2 min, 流动相 B 的比例维持 0%; 2~10 min, 流动相 B 的比例由 0% 到 40%; 10~50 min, 流动相 B 的比例由 30% 到 70%), 流速 1.0 mL · min⁻¹, 检测波长 214 nm。样品保存温度 15 °C, 直接进样 20 μL 进行检测。

检测色谱图见图 6, 按峰面积归一化法积分, 主峰面积大于 98.3%。

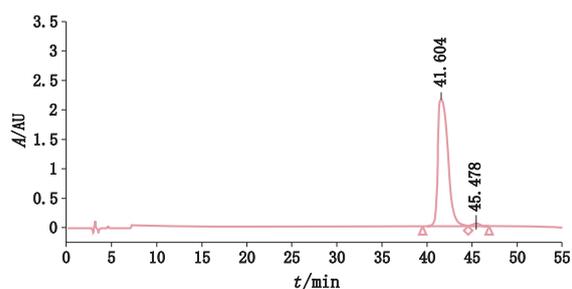


图 6 rhG-CSF 的反相高效液相色谱图

Fig. 6 RP-HPLC chromatogram of rhG-CSF

3 讨论

部分单链重组蛋白类制品的蛋白质主链 N- 末端具有异质性, 即有 2 个以上 N- 末端起始肽链, 长的肽链序列包含短的肽链序列, 或短的序列缺失了 1 个或几个氨基酸残基。产生这种现象的原因, 可能与生产工艺相关或与肽链本身性质相关^[7-9]。较常见的情况是: 如果重组蛋白质是通过分泌表达, 在过细胞膜的过程中需要膜肽酶的切割, 而发酵环境的变化可使切割位点偏移^[10], 从而产生 N 端异质体;

另外,脯氨酸残基在 N 端第 2 位的多肽,其 N 端第 1 位与第 2 位脯氨酸之间的肽键较其他肽键更易断裂^[10],这也会导致 N 端异质体的出现。

具有 N-末端异质性的各种蛋白是否有不同的生物学活性尚无定论^[11-13]。如果缺失的氨基酸残基影响了蛋白的活性中心或改变了蛋白质的空间结构,那么蛋白的比活性可能损失较多;如果缺失的氨基酸残基使其与配体的结合力下降,也可能引起比活性下降。2015 年版《中华人民共和国药典》三部中将 rhG-CSF 以蛋氨酸残基起始的序列和以苏氨酸残基起始的序列均视为主药成分,这是不同生产企业制品标准合并的结果,部分生产企业制品属前者,部分属后者。2 个 N 端序列不同的主药成分在比活性上未做明显区分,说明相差不显著。

很多学者倾向于将这种 N 端异质性蛋白看作相关蛋白。从逻辑上讲,蛋白分子上有任何基团的丢失或加入都带来分子结构的变化,这一过程在制品的生产和保藏过程中每时每刻都在发生,所以任何一个人们认为的纯蛋白类制品,也许只是“理论上的纯品”,实际上是各种相似物质的混合物,只是主药蛋白质占主体。只有占比较高的相关蛋白才容易引起人们的注意。

在制品申报过程中,人们是按照既定目标进行申报,即认定组分(主药蛋白)与杂质(如宿主蛋白等),而游离在两者之间的蛋白种类繁多。从检测手段上看,一些常规的质量控制检测方法并不能有效分离所有“主药蛋白”之外的其他蛋白,如本文中提及的 RP-HPLC 纯度检测法就不能完全区分部分 N 端异质性蛋白,纯度大于 98.3%,即将部分相似物质也列入其中。

从分子结构上看,有些蛋白与主药蛋白非常相近且具有较高的同源性,性质上也较为接近,但如果其具有毒副作用,就应该格外关注并且加以控制。从理论上推测,N 端异质性蛋白应具有与“主药蛋白”相似的生物学活性,但由于目前实验条件所限,本文未进行验证。后续将进一步对 rhG-CSF 样品进行分析和测定,如纯化收集足量的相关蛋白,测定蛋白含量,除菌过滤后测定生物学活性并统计分析其比活性,还需评价乙腈等有机溶剂对生物学活性测定的影响等。

本文对 rhG-CSF 蛋白质 N-末端异质性进行了分析,希望为国产 rhG-CSF 制品的质量标准提升提供参考和数据支持。

参考文献

[1] ASANO S. Function, molecular structure and gene expression of granulocyte colony-stimulating factor[J]. *Nihon Rinsho*, 1992, 50

(8): 1854

- [2] MORSTYN G, FOOTE M, PERKINS D, *et al.* The clinical utility of granulocyte colony-stimulating factor: early achievements and future promise[J]. *Stem Cells*, 1994, 12(Suppl 1): 213
- [3] BABAEIPOUR V, KHANCHEZAR S, MOFID MR, *et al.* Efficient process development of recombinant human granulocyte colony-stimulating factor (rh-GCSF) production in *Escherichia coli*[J]. *Iran Biomed J*, 2015, 19(2): 102
- [4] 梁逊,郭小丽. 多肽、蛋白质固相分析[J]. *生物化学与生物物理学进展*, 1988, 15(2): 105
- LIANG X, GUO XL. Solid phase analysis of polypeptide and protein[J]. *Prog Biochem Biophys*, 1988, 15(2): 105
- [5] 史新昌,杨靖清,韩春梅,等. N-末端氨基酸测序数据估计 N-末端不均一的单链重组蛋白制品主肽链比例的方法[J]. *中国生物制品学杂志*, 2016, 29(10): 1073
- SHI XC, YANG JQ, HAN CM, *et al.* Estimate most-peptide-chain-component-ratio of heterogeneous N-terminal single subunit recombinant protein biological products using N-terminal sequencer data[J]. *Chin J Biol*, 2016, 29(10): 1073
- [6] 毕华,陶磊,韩春梅,等. 重组人血管内皮生长因子抑制剂理化对照品质控方法及质量标准的建立[J]. *中国生物制品学杂志*, 2017, 30(8): 833
- BI H, TAO L, HAN CM, *et al.* Establishment of methods and standard for quality control of physicochemical reference substance for recombinant human vascular endothelial growth factor inhibitor[J]. *Chin J Biol*, 2017, 30(8): 833
- [7] BRORSON K, JIA AY. Therapeutic monoclonal antibodies and consistent ends: terminal heterogeneity, detection, and impact on quality[J]. *Curr Opin Biotechnol*, 2014, 30C: 140
- [8] RENZ A, SCHIKORA S, SCHMID R, *et al.* cDNA sequence and heterologous expression of monomeric spinach pullulanase: multiple isomeric forms arise from the same polypeptide[J]. 1998, 331(Pt 3): 937
- [9] VIGUERA AR, SERRANO L. Stable proline box motif at the N-terminal end of alpha-helices[J]. *Protein Sci*, 1999, 8(9): 1733
- [10] 何冰芳,米兰,陈文华. 大肠杆菌蛋白质分泌机理及其重组蛋白分泌表达新进展[J]. *食品与生物技术学报*, 2012, 31(6): 561
- HE BF, MI L, CHEN WH. Research advances in mechanism of protein secretion and secretory expression of recombinant prote in *E. coli*[J]. *J Food Sci Biotechnol*, 2012, 31(6): 561
- [11] RAUCCI R, COSTANTINI S, CASTELLO G, *et al.* An overview of the sequence features of N- and C-terminal segments of the human chemokine receptors[J]. *Cytokine*, 2014, 70(2): 141
- [12] HESS AD, THOBURN C, CHEN W, *et al.* The N-terminal flanking region of the invariant chain peptide augments the immunogenicity of a cryptic “self” epitope from a tumor-associated antigen[J]. *Clin Immunol*, 2001, 101(1): 67
- [13] EL-SHAMY A, SHOJI I, KIM SR, *et al.* Sequence heterogeneity in NS5A of hepatitis C virus genotypes 2a and 2b and clinical outcome of pegylated-interferon/ribavirin therapy[J]. *PLoS One*, 2012, 7(2): e30513

(本文于 2018 年 8 月 3 日收到)