

液质联用研究人重组干细胞因子一级结构

郑建华¹, 张平², 饶晓璐², 王世媛², 颜江华^{1*}

(1. 厦门大学医学院, 福建 厦门 361005; 2. 厦门特宝生物工程股份有限公司, 福建 厦门 361028)

摘要: 通过高效液相色谱(HPLC)和质谱技术,对人重组干细胞因子(rhSCF)分子进行分析,以确证其一级结构.将rhSCF样品胰酶酶切,HPLC-RPC分离,基质辅助激光解吸电离-飞行时间串联质谱仪(MALDI-TOF-MS)对rhSCF的酶切肽段进行解析,并对rhSCF的O-糖基化和N-糖基化位点进行了分析.结果:(1)rhSCF分子质量得到测定;(2)rhSCF的一级结构得到确认,测定的肽质量谱的覆盖率达到94%,分子中有两对二硫键;(3)鉴定了样品中的N-糖基化位点,并对rhSCF产品的O及N-连接糖型上的糖结构进行初步探讨.结果显示:rhSCF一级氨基酸序列和二硫键形成位置与理论上一致,糖基化特性符合毕赤酵母表达蛋白的糖基化特点.

关键词: MALDI-TOF-MS; 人重组干细胞因子; 二硫键; 糖基化

中图分类号: Q 516

文献标志码: A

文章编号: 0438-0479(2011)03-0612-05

干细胞因子(stem cell factor, SCF)^[1,2]是构成造血中心的一种细胞分子,它在骨髓基质细胞、血管内皮细胞等的细胞里最初是以膜结合的形式合成的,通过蛋白酶水解产生可溶型.它的主要生物学功能是作用于极早期的具有多向分化潜能的造血干细胞,促使这类细胞由静止、休眠状态进入细胞周期或加快细胞周期的进程等^[3].SCF是一种糖蛋白,分子质量为20 000~36 000 u,等电点为5.0.它的可溶型由膜外1~165位氨基酸组成^[4],含有SCF主要生物学活性区域.人源SCF具有两对分子内二硫键,分别为Cys⁴-Cys⁸⁹和Cys⁴³-Cys¹³⁸,两对二硫键都是维持SCF生物活性所必需的^[4].SCF分子的糖基约占相对分子质量的40%,包括N-和O-两种糖基化类型.毕赤酵母表达的SCF分子的糖基化略有不同,但分析认为糖基化与否对SCF的活性没有影响^[5].本文通过对人重组干细胞因子(rhSCF)的一级结构进行解析,确证了rhSCF的一级氨基酸序列、二硫键结合位置和糖基化的一些特点,同时为重组蛋白的一级结构确认、理化性质的了解和质量控制提供一个可行的方法.

1 材料与方法

1.1 材料

Autoflex 质谱仪、相应基质及 Flexanalysis 数据库检索软件(Bruker 公司); Agilent 1200 高效液相系

统; 东富隆 LYO-0.2 冻干机; 37 °C 水浴锅; 反相柱 Jupiter C18(Phenomenex 公司); 乙腈(HPLC 纯, Fisher 公司); 三氟乙酸(HPLC 纯, TEDIA 公司); 二硫苏糖醇(DTT)(Invitrogen 公司); 3K 超滤膜(Amicon Ultra 0.5 lot R9DN68652); 测序级蛋白内切酶胰酶(Promega 公司); 碳酸氢铵(AR, 上海试四赫维化工有限公司); rhSCF 原液由本单位自制.

1.2 方法

1.2.1 样品制备

1) rhSCF 原液的制备. rhSCF 酵母工程菌发酵后, 发酵液离心除去菌体, 上清经过阳离子交换、分子筛层析, 无菌过滤等技术制备得到 rhSCF 原液.

2) 胰酶酶切样品制备. rhSCF 原液采用 50 mmol/L 碳酸氢铵 pH 7.8 缓冲液通过 3K 超滤膜超滤替换缓冲液, 并浓缩至约 1 mg/mL. 取替换过缓冲液的样品 250 μL, 按蛋白与酶质量比为 50:1 加入测序级胰酶, 混合均匀, 在 37 °C 条件下水浴保温 24 h. 酶切结束后加入 10% 体积的 50% 醋酸混合均匀, 取 200 μL 样品加入 1 mol/L DTT 1 μL, 50 μL 样品加入超纯水 1 μL, 分别置于 37 °C 水浴中保温 30 min.

3) 酶切肽段的 HPLC-RPC 分析. 将胰酶酶切后的 rhSCF 酶切产物用终浓度为 5 mmol/L 的 DTT 混合 37 °C 保温 30 min, 然后取样进行 HPLC-RPC 肽图分析. 肽图采用 HPLC-RPC-C18 柱分离, 柱温为 30 °C, 以含 0.1% (体积分数, 下同) 三氟乙酸的水溶液为流动相 A 液, 以 0.1% 三氟乙酸的乙腈溶液为流动相 B 液, 流速为 1 mL/min, 梯度洗脱 70 min(A 液从

收稿日期: 2010-07-14

* 通信作者: jhyan@jingxian.xmu.edu.cn

100%~30%, B液从0~70%), 检测波长为214、280 nm, 分别收集各肽段峰。

4) 冻干各峰样品, 然后用10 μ L超纯水溶解。

5) 分别对原液、酶切后的样品、酶切后还原的样品和液相分离的各峰样品进行质谱分析。除原液外各样品分别取样品1 μ L与含0.1%三氟乙酸的70%乙腈溶液饱和的 α -氰基-4-羟基肉桂酸溶液1 μ L充分混合, 点样, 待自然干燥后质谱分析。rhSCF原液样品取样品1 μ L与含0.1%三氟乙酸的50%乙腈溶液饱和的DHB溶液1 μ L充分混合, 点样, 待自然干燥后质谱分析。

1.2.2 质谱条件

Daltonics Autoflex III MALDI-TOF-MS, 采用RP模式测定肽段分子质量, 采用Lift模式测定二级序列, 采用LP模式测定整体蛋白分子质量。

2 结果与讨论

2.1 rhSCF分子质量的质谱分析

本文纯化得到的rhSCF是由一系列的不同糖基化程度的样品组成的混合物, 由于不均一的糖链的存在, 质谱分析时形成一个从约18 000~22 000 u的连续分布的分子质量峰。同时可以检测到相应的2质子峰和3质子峰, 分子质量在6 905和10 360 u附近。

2.2 rhSCF分子的一级序列分析

本文对胰酶酶切后的样品进行高效液相反相分离, 收集各肽峰样并得到rhSCF的胰酶肽图, 随后对每个液相峰进行单独处理, 检测该峰所含的肽段, 并进行二次解析分析, 得到rhSCF的液相峰肽对应图。

rhSCF蛋白通过胰酶酶切将产生20个独立肽段, 本文中除T2、T12肽段由于分子质量太小而不适合质谱解析, T20肽段由于肽键断裂使得部分肽段没有得到解析, T10肽段由于完全糖基化而未打出未糖基化的肽峰外, 其余样品均能很好地检测分析。最终经过解析确定的氨基酸序列占总氨基酸残基数量的94%; 如表1所示。

2.3 rhSCF分子中两对二硫键分析

本文在rhSCF氨基酸序列已经清楚的情况下, 通过比较由二硫键还原前后的肽段差异, 确定二硫键的形成位点。非还原状态下, 形成二硫键的两个肽段通过二硫键彼此连接, 分子质量为两个肽段之和减2 u, 还原状态下二硫键被打开, 形成两个分子质量较小的肽段。通过比较还原前后肽段的差异, 可确定形成二硫键

的半胱氨酸所在的肽段, 由此确定二硫键的位置。

rhSCF分子中含4个半胱氨酸, 理论上其第4位的半胱氨酸和第89位的半胱氨酸形成二硫键, 第43位的半胱氨酸和第138位的半胱氨酸形成二硫键。rhSCF经过胰酶酶切后, T1和T9之间以二硫键连接在一起, 分子质量为2 034 u, 经过DTT还原后, 二硫键连接的2 034 u肽段将被还原打开, 形成分子质量为577 u(T1)和1 458 u(T9)的两个肽段, 分子质量为2 034 u的肽段峰将消失。T7和T18之间以二硫键连接在一起, 分子质量为5 690 u, 经过加入DTT还原后, 二硫键连接的5 690 u肽段将被还原打开, 形成分子质量为3 492 u(T1)和2 200 u(T9)的两个肽段, 分子质量为5 690 u的肽段峰将消失。质谱比较分析图谱如图1和2所示:

通过对还原前后样品的质谱解析, 可确定rhSCF的两对二硫键的位置为Cys⁴-Cys⁸⁹和Cys⁴³-Cys¹³⁸。

2.4 毕赤酵母表达蛋白的O和N糖基化在质谱解析中的特征差异

N-糖基化和O-糖基化中修饰蛋白的糖链与蛋白具有不同的连接方式。N-糖基化中糖链通过N原子与天冬酰胺相连接, 糖链中包含一个由3个甘露糖和2个GlcNAc组成的5糖核心, 靠近蛋白连接端的为2个GlcNAc。在进行二次质谱解析时, 由于分子内的化学键在质谱激光的作用下产生不同的裂解方式, 与肽段连接的部分通常会出现一种固定的断裂模式如图3(a), 这个断裂方式在质谱上表现为一系列的特征质谱峰。O-糖基化中修饰蛋白的糖链则通过O原子与苏氨酸或者丝氨酸相连接, 糖链的长度通常不超过3个单糖残基。在进行二次质谱解析时连接部位呈现具有如图3(b)所示的裂解模式^[69]。

2.4.1 rhSCF分子中的O-糖修饰分析

糖基或糖链的还原端与蛋白质肽链中的丝氨酸、苏氨酸或羟赖氨酸羟基中的氧原子相连称为O-连接糖链。在已有的文献中通常O-糖基化的糖链由于所占比例较小而不被重视, 而毕赤酵母所表达的通常为不高于3个甘露糖糖残基组成的糖链, 所形成Man1-3-Ser/Thr结构, 在已知的酵母表达糖蛋白中通常为Man α 1, 2Man α 1, 2ManSer/Thr结构。

rhSCF有多个O-糖基化位点, 在本文中共确证了4个O-糖基化肽段, 分别为T6、T10、T18、T20肽段。酵母的O-糖基化程度是不均一的、不完全的, 因此同一个位点有未被糖基化的, 也有糖基化的, 而糖基化

表 1 MALDI-TOF MS 测定的高效液相反相分离后的胰酶酶切 rhSCF 肽段的解析结果

Tab. 1 HPLC RPC separated rhSCF digested peptide analyzed by MALDI-TOF MS

序列起止	肽段序号	MH ⁺ /u		出峰时间/min	理论序列	检测到含有的修饰
		实测值	理论值			
1~ 5	T 1	577.3	576.7	16.668	EGICR	
6~ 7	T 2		288.3		NR	
8~ 13	T 3	674.4	673.8	13.404	VTNNVK	
14~ 17	T 4	462.1	461.5	9.084	DVTK	
18~ 24	T 5	754.7	753.9	23.134	LVANLPK	
25~ 31	T 6	883.4	883.1	28.939	D YMITLK	O-糖基化
32~ 62	T 7	3491.2	3492.0	59.265	YVPG DVLPSHCWI SEMVVQLSDS LTDLLDK	
63~ 78	T 8	1787.0	1787.0	31.923	FSN ISE GLSNYSI IDK	
79~ 91	T 9	1458.8	1458.7	42.305	LVNIVDD LVECVK	
92~ 96	T 10		563.6	23.638	ENSS K	N/O-糖基化
97~ 99	T 11	375.1	374.4	10.687	DLK	
100~ 100	T 12		146.2		K	
101~ 103	T 13	380.9	380.4	13.404	SFK	
104~ 108	T 14	585.2	584.6	12.087	SP EPR	
109~ 117	T 15	1185.6	1185.3	37.996	LFTPEEF FR	
118~ 121	T 16	549.4	548.6	17.372	IFNR	N-糖基化
122~ 127	T 17	680.5	679.8	23.638	SIDA FK	
128~ 148	T 18	2200.2	2200.4	31.480	DFVVASET SDCVVSSTLS PEK	O-糖基化
149~ 151	T 19	376.9	376.4	2.735	DSR	
152~ 165	T 20		1456.8	25.972*	VSVT KPFMLPPVAA	O-糖基化

注: T20 肽段质谱解析到肽段 VSVT KPF 及其 O-糖基化产物,完整的 T20 肽段未检测到。

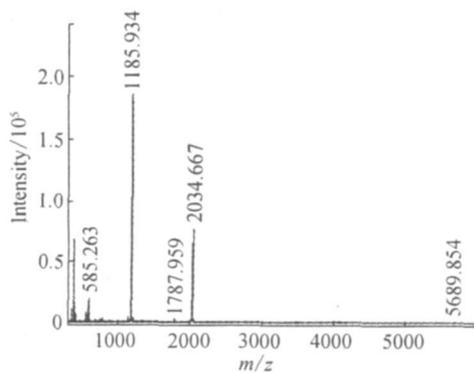


图 1 还原前胰酶酶切后的 rhSCF 肽段样品质谱图

Fig. 1 The mass spectrum of the rhSCF peptides digested by trypsin before deoxidized

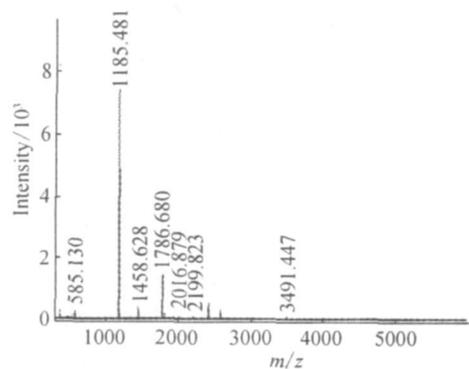
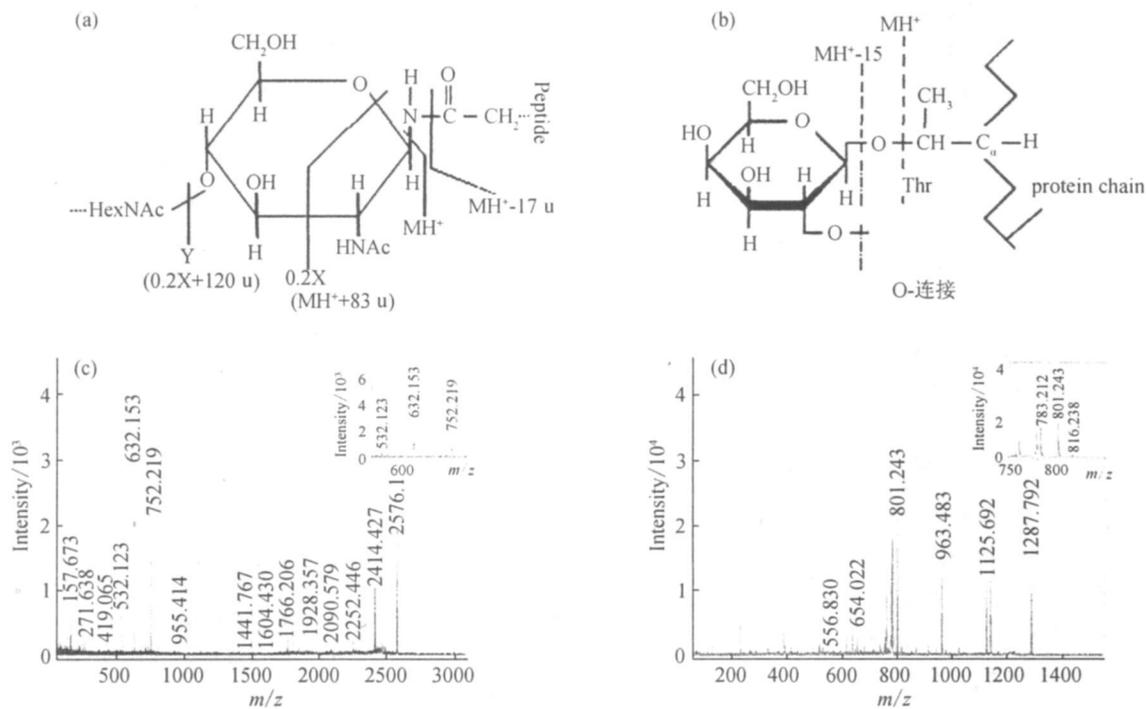


图 2 还原后胰酶酶切后的 rhSCF 肽段样品质谱图

Fig. 2 The mass spectrum of the deoxidized rhSCF peptides digested by trypsin



(a) 为 N-糖基化的结合处的结构示意图和 N-糖基化的肽段在质谱二次解析时分子碎片的断裂情况;
 (b) 为 O-糖基化的结构示意图和 O-糖基化的肽段在质谱二次解析时分子碎片的裂解情况;
 (c) 为 N-糖基化肽链的二级质谱解析图谱, 图中小图为 N-糖基化位点质谱解析的特征图谱;
 (d) 为 O-糖基化肽链的二级质谱解析图谱, 图中小图为 O-糖基化位点质谱解析的特征图谱.

图 3 O-糖基化和 N-糖基化的结构示意图及 rhSCF 糖基化肽段的二级质谱解析图谱

Fig. 3 Structure of N and O glycosylation and the second order mass spectrum of rhSCF glycosylated peptide

表 2 已质谱解析的糖基化肽段的汇总

Tab. 2 Summary of mass analysed glycosylated peptide

肽段序号	氨基酸序列	糖基化类型	分子量/ u	糖链结构
T6	D Y MIT LK	O	1045	肽 Hex
T10	ENSS K	N / O	2473	肽 HexNAc HexNAc 6Hex Hex _(a+b) *
T16	IFNR	N	2576	肽 HexNAc HexNAc 8Hex Hex Hex
		N	2414	肽 HexNAc HexNAc 8Hex Hex
T18	DFVVASETSD	O	2524	肽 Hex Hex
	CVVSSTLSPE K	O	2362	肽 Hex
T20 片段	VSVT KPF	O	1287	肽 Hex Hex Hex
		O	1125	肽 Hex Hex

注: * a + b = 3, a 为 O-糖链的己糖数, b 为 N-糖链的尾端的己糖数.

的糖链的长度从 1~ 3 个糖残基都有发现, 见表 2.

2.4.2 rhSCF 分子中的 N-糖修饰分析

在糖蛋白中仅有一种糖残基与天冬酰胺相连, 即 N-乙酰-β-D-葡糖胺, 生成的是 4N-(2 乙酰氨基-2 脱氧-β-D-吡喃葡糖基)-L-天冬酰胺, 这种连接方式有时称为 N(或 Asn) 连接型糖链或 N-聚糖. 酵母 N-糖基化

表达的糖链主要是高甘露糖型, 不含唾液酸. 巴斯德毕氏酵母合成的糖链一般为 8~ 14 个糖基组成. 酵母表达的重组糖蛋白存在糖基化不均一现象, 主要表现为糖基化程度的差异和糖链长度的一致, 其主要结构由一个 5 糖核心和若干甘露糖组成. 本实验分析的样品的 N-糖基化样品的糖基化位点发生在 N₁₂₀ 位(肽段

IFNR) 和 N⁹³位(肽段 ENSSK), 糖链长度主要在 8~11 个单糖残基, 其中与糖链相连的 2 个为 N-乙酰己糖, 其余为己糖。质谱解析结果如图 3 所示, 汇总见表 2。

3 结 论

本研究结合 HPLC 和质谱法对 rhSCF 进行一级结构解析, 建立了多种 HPLC 与质谱法联用研究糖基化蛋白的方法, 如胰酶肽图、平均分子量、N-糖基化位点的确定、O-糖基化位点的分析等。同时结合酶学、高效液相色谱和质谱的技术方法测定了 rhSCF 的峰肽对应图谱, 使得对 rhSCF 的氨基酸序列的解析程度达到 94%。在此基础上, 对 rhSCF 结构中的两对二硫键进行了研究, 确定了二硫键位置的结构信息; 并对 N-糖及 O-糖的结构进行了初步的探讨。本方法灵敏度高、重现性好, 为生物工程产品的质量控制和研究提供了一种可行的方法。

参考文献:

[1] Martin F H, Suggs S V, Langley K E, et al. Primary structure and functional expression of rat and human

stem cell factor DNA[J]. Cell, 1990, 63: 203-211.

[2] Anderson D M, Lyman S D, Daird A, et al. Molecular cloning of mast cell growth factor, a hematopoietin that is active in both membrane bound and soluble forms[J]. Cell, 1990, 63: 235-243.

[3] Andrews R G, Knitter G H, Bart elmer S H, et al. Recombination human stem cell factor, a kit ligand, stimulates hematopoiesis in primates[J]. Blood, 1991, 78(8): 1975-1980.

[4] Galli S J, Zsebo K M, Geissler E N. The kit ligand, stem cell factor[J]. Advances in Immunology, 1994, 55: 1-96.

[5] 王志军. 生物技术药物研究开发和质量控制[M]. 北京: 科学出版社, 2007.

[6] Herscovics A, Orlean P. Glycoprotein biosynthesis in yeast[J]. FASEB J, 1993, 7: 540-550.

[7] Trent R Gemmill, Robert B Trimble. Overview of N- and O-linked oligosaccharide structures found in various yeast species[J]. Biochimica et Biophysica Acta, 1999, 1426(2): 227-237.

[8] 詹洁, 吴军. 酵母 N-糖基化工程研究进展[J]. 生物技术通讯, 2004, 15(3): 272-274.

[9] 王黎, 刘洋, 陈敏. 酵母表达人源化糖蛋白研究进展[J]. 中国生物工程杂志. 2007, 27(4): 120-125.

Primary Structure Determination of Recombinant Human Stem Cell Factor by HPLC and MALDI-TOF

ZHENG Jiarhua¹, ZHANG Ping², RAO Xiaolu², WANG Shiyuan²,
YAN Jianghua^{1*}

(1. Medical College of Xiamen University, Xiamen 361005, China;

2. Xiamen Amoytop Biotech Corporation Limited, Xiamen 361028, China)

Abstract: High performance liquid chromatography (HPLC) and matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS) were used to determine the primary structure of the recombinant human stem cell factor (rhSCF) samples. The rhSCF samples were digested by TPCK-trypsin, and separated by HPLC. The molecular weight of rhSCF and the amino sequences of its the peptides were determined by MALDI-TOF, and the glycosylated peptides were analyzed. The results are as follows: (1) the mass of rhSCF was detected, (2) the rhSCF primary structure was determined, and approximately 94% amino sequence of rhSCF was confirmed, including two couples of disulfide bonds, (3) the N-linked and O-linked glycosylated positions were analyzed and the oligosaccharide structure was proposed. The the primary amino sequence of rhSCF corresponded to the theory sequence, and the two couples of disulfide bonds have properly formed. And the character of the glycosylation conformed to the pichia yeast expressed proteins.

Key words: MALDI-TOF-MS; recombinant human stem cell factor; disulfide bond; oligosaccharide