

基因工程产品的N末端均一性及其中 残留的痕量杂质蛋白的分析

金冬雁 张智清 周 圆 侯云德

(中国预防医学科学院病毒学研究所, 病毒基因工程国家重点实验室, 北京)

利用 Edman 降解法, 辅之以邻苯二甲醛阻断等技术, 在蛋白/多肽测序仪-PTH 氨基酸分析仪上微量自动测定了在大肠杆菌系统获得高效表达的重组人 γ 干扰素、 α 2 干扰素和白细胞介素-2 等多种基因工程产品的N末端氨基酸序列。根据氨基酸色谱分析的原始资料, 对上述蛋白质样品的N末端均一性进行了细致分析。在此基础上, 借助序列库检索比较等计算机方法检定了这些样品中可能残留的痕量杂质蛋白, 并对分析痕量残留蛋白的一般方法进行了讨论。分析结果表明重组人 γ 干扰素N末端均一; 而 α 2 干扰素和白细胞介素-2 的N末端不均一, 其中主要含大致等量的N末端带有蛋氨酸与不带蛋氨酸的两种分子, 仅在个别送检样品中发现痕量N末端丢失一个氨基酸的分子。通过在序列库中检索所测出的一段八肽疑似序列, 推断在一批送检样品中含有痕量的卵清溶菌酶, 经进一步纯化样品并测序后得以证实。本文结果为进一步优化和确定有关产品的生产流程并建立产品质量控制的标准提供了重要依据。

关键词 蛋白质序列测定与分析; 蛋白质N末端均一性; 痕量杂质蛋白

本实验室先后在大肠杆菌系统获得重组人 α 1、 α 2 和 γ 干扰素、重组人白细胞介素-2 的高效表达^[1-4], 并且已经或正在进行中试生产, 从而标志着我国基因工程高技术产业的兴起。在建立和完善基因工程产品纯化及生产流程的过程中, 必须对表达产物的均质度, 尤其是N末端均一性作出准确的鉴定, 并对样品中可能残留的痕量杂质蛋白进行必要的定性分析。N末端氨基酸序列测定, 就是分析蛋白质N末端均一性及痕量残留蛋白的重要指标。

气相及脉冲液相蛋白/多肽微量自动 Edman 降解技术以及与之联用的微内径高效液相色谱 (Microbore HPLC) PTH 氨基酸分析技术, 集中了快速、微量、灵敏与自动化等特点, 配用先进的色谱分析软件, 可以用100pmol样品顺利分析20个以上的N末端氨基酸序列, PTH氨基酸检

出下限达0.5—1.0pmol^[5]。如果Edman降解中结合运用邻苯二甲醛阻断的方法封闭样品中所有暴露在N末端的一级氨基, 还可测定混合肽中特定的一种N末端为Pro的氨基酸序列。目前这一技术已经成为进行蛋白和多肽微量测序的常规技术。

利用 FASTA 快速序列检索比较程序在蛋白质序列库中搜寻某一肽段的同源序列, 是辅助分析测序结果的有力手段。统计资料指出^[6], 在现行蛋白质序列库中随机出现两段五肽相同序列的概率约为10%, 而随机出现两段六肽相同序列的概率小于1%。因此可以期望通过寡肽的序列库检索对所检出的蛋白质进行定性, 乃至判定测序结果的可靠性。

本文报道在利用应用生物系统公司477A蛋白多肽测序仪-120A PTH氨

本文于1990年7月1日收到。

基酸分析仪测定重组人 γ 干扰素、 $\alpha 2$ 干扰素和白细胞介素-2等样品N末端氨基酸序列的过程中,借助邻苯二甲醛阻断和蛋白质序列库的计算机检索比较等方法,分析测序样品均质度及其中痕量蛋白的结果,并在此基础上讨论了进行上述分析的一般策略。

材料与方法

(一) 基因工程产品的生产与纯化

重组人 γ 干扰素、重组人 $\alpha 2a$ 干扰素、重组人白细胞介素-2的生产与纯化分别根据文献^[2-4]提供的流程进行,每批样品均用SDS-聚丙烯酰胺凝胶电泳检查纯度,确认达到电泳纯。

(二) $\alpha 2a$ 干扰素的过甲酸氧化^[7]

0.9ml甲酸与0.1ml 30%过氧化氢在室温下反应1h后制成过甲酸,冰浴冷却备用,以60 μ l甲酸溶解约5 nmol蛋白质,加入20 μ l过甲酸在-20 $^{\circ}$ C保温2h,再加入2ml超纯水,真空干燥备用。所用化学试剂均为国产分析纯或优级纯试剂,超纯水用美国Millipore公司Milli-Q SP纯水仪制备。

(三) Edman降解与PTH氨基酸分析^[5]

在美国应用生物系统公司生产的ABI 477A蛋白多肽序列仪和120A PTH氨基酸分析仪上进行。上样量在100—600pmol之间,采用该公司提供的1.61版序列测定与分析软件设定降解、转化和色谱分析条件,并作序列分析。所用各种化学试剂均为该公司提供的蛋白质测序级试剂。

(四) 邻苯二甲醛阻断Edman降解的技术^[8]

将20mg邻苯二甲醛和50 μ l β -巯基乙醇溶于10ml乙腈中,制成邻苯二甲醛试剂,装入477A测序仪,待测序进行到预

期Pro残基暴露于样品中一种组分蛋白的N末端时,采用应用生物系统公司提供的特定程序进行邻苯二甲醛阻断。邻苯二甲醛是Fluka公司分析纯产品, β -巯基乙醇是Merck公司合成级试剂。

(五) 在蛋白质序列库中检索同源序列^[9,10]

快速蛋白检索软件FASTA由美国弗吉利亚大学Pearson博士赠送,美国生物数学研究中心蛋白质序列库(NBRF-PIR)19.0版购自Pharmacia公司,由本文作者将序列资料转换FASTA程序可识别的数据形式并补入部分最近发表的序列。同源序列的检索按文献^[9]提供的方法进行。

结果

(一) γ 干扰素样品N末端均一性的分析

测定 γ 干扰素N末端序列的结果归纳在表1中。从表1可见,第2至第19个氨基酸序列为Q-D-P-Y-V-K-E-A-E-N-L-K-K-Y-F-N-A-G。测序结果不排除第1个氨基酸为C,而C在所采用的系统中无法检出。表中“pmol数”一项是在Edman降解的某一周期实际测得的PTH氨基酸量,而“pmol比值”一项则是“pmol”项经过校正之后与标准差之比,可作为评判测出该氨基酸的可能性的统计学标准。根据表中“pmol比值”以及每一周期的色谱分析数据判定,N末端连续19个Edman降解周期均测出单一氨基酸残基,说明样品N末端均一,均质度很好,达到氨基酸序列纯。为了检查样品中是否含有痕量杂质,又将测序的样品量从20pmol增加到150pmol,结果仍未检出任何其它蛋白的序列。

(二) 白细胞介素-2和 $\alpha 2a$ 干扰素样品N末端均一性的分析

表 1 γ 干扰素N末端测序结果
Table 1 N terminal sequence of IFN- γ

序号 No.	实际保留时间 Retention time(s)	标准保留时间 Calibration time(s)	pmol数 pmol	pmol比值 pmol ratio	氨基酸 Amino acid
1					?
2	7.47	7.52	14.99	29.75	Gln
3	5.48	5.53	14.58	61.18	Asp
4	19.67	19.58	15.27	37.81	Pro
5	16.22	16.15	11.78	108.38	Tyr
6	21.00	20.90	9.67	31.62	Val
7	26.40	26.25	11.53	70.87	Lys
8	9.37	9.38	7.95	14.89	Glu
9	12.43	12.43	11.75	27.83	Ala
10	9.20	9.38	6.74	22.18	Glu
11	6.02	6.08	8.86	48.92	Asn
12	26.92	26.78	8.41	19.19	Leu
13	26.40	26.25	7.95	24.28	Lys
14	26.40	26.25	9.02	24.34	Lys
15	16.20	16.15	5.67	17.07	Tyr
16	25.15	24.98	5.52	20.28	Phe
17	6.07	6.08	7.05	33.40	Asn
18	12.50	12.43	6.55	12.82	Ala
19	8.57	8.57	8.36	17.13	Gly

初步测定白细胞介素-2和 $\alpha 2a$ 干扰素序列时,发现每一降解周期的色谱分析图均出现两种PTH氨基酸的色谱峰;同样从每种氨基酸在各个周期的分布图(如图1)上往往可见连续两个高度相仿的高峰。据此我们推断白细胞介素-2和 $\alpha 2a$ 干扰素样品的N末端并不均一,其中主要含有两种N末端相差一个蛋氨酸的序列。根据初步测序的结果,预期在白细胞介素-2样品的第3个和 $\alpha 2a$ 干扰素样品的第4个降解周期,Pro残基将暴露在一种主要蛋白组分的N末端。为了准确读出N末端不

均一样品之中的一种序列,以便对样品进行必要鉴定,我们采用邻苯二甲醛阻断法,分别在第3、第4周期对白细胞介素-2和 $\alpha 2a$ 干扰素样品的Edman降解进行了阻断。图2比较了采用和未采用邻苯二甲醛阻断时,Thr在降解白细胞介素-2的各个反应周期的分布情况。图中可见阻断后本底大为降低,第4、第8、第11位的Thr信号显露突出。阻断后测得的单一序列证实了前面对样品组分进行的分析。又将 $\alpha 2a$ 干扰素的测序样品量从40pmol增大到200pmol,同时在降解的第3周期施

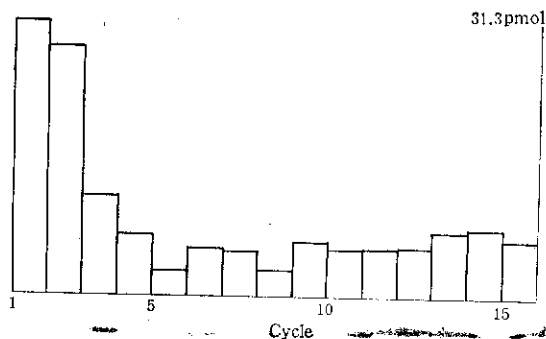


图1 Ala在白细胞介素-2各个降解周期的分布情况
Fig. 1 Ala histogram in each degradation cycle of interleukin-2

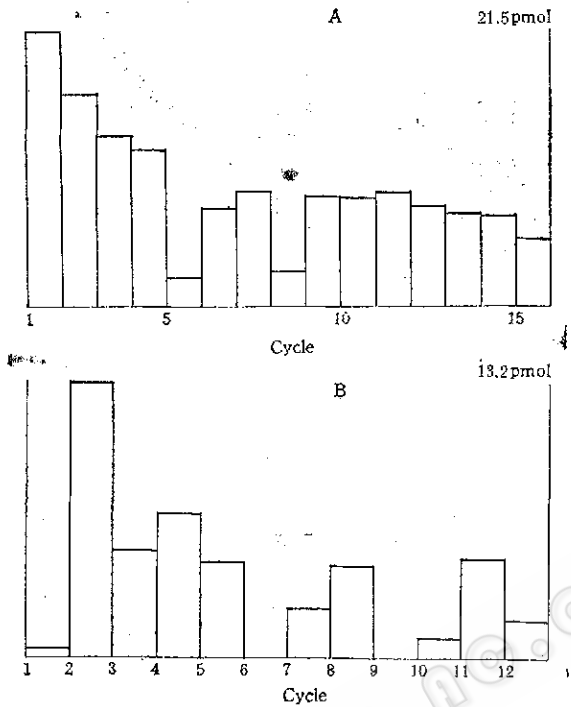


图 2 Thr在白细胞介素-2 各个降解周期的分布情况
Fig. 2 Thr histogram in each degradation cycle of interleukin-2
A. 未采用邻苯二甲醛阻断 No blockage
B. 在第三周期采用邻苯二甲醛阻断 OPA blockage at the third cycle

表 2 杂质蛋白N末端测序结果
Table 2 N terminal sequence of the contaminated protein

序号 No.	实际保留时间 Retention time(s)	标准保留时间 Calibration time(s)	pmol数 pmol	pmol比值 pmol ratio	氨基酸 Amino acid
1	25.82	26.08	33.49	35.50	Lys
2	20.93	20.73	21.23	28.04	Val
3	25.13	24.80	33.68	80.44	Phe
4	8.63	8.45	34.54	20.58	Gly
5	17.82	17.88	23.94	29.72	Arg
6					?
7	9.33	9.25	19.81	24.30	Glu
8	26.92	26.60	23.62	42.41	Leu
9	12.45	12.32	32.88	18.03	Ala
10	12.43	12.32	35.73	17.12	Ala
11	12.45	12.32	26.78	11.67	Ala
12	20.50	20.32	23.75	65.09	Met
13	26.30	26.08	21.56	12.35	Lys
14	18.03	17.88	17.99	23.86	Arg
15	13.68	13.58	5.48	21.43	His

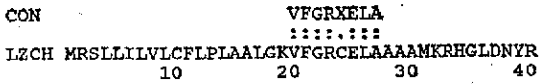


图 3 痕量残留蛋白质序列与卵清溶菌酶序列的同源排列
Fig. 3 Sequence alignment between the residual contaminant and lysozyme
CON: 痕量残留蛋白 Residual contaminant, LZCH: 卵清溶菌酶前体Chicken eggwhite lysozyme

行邻苯二甲醛阻断, 结果发现样品中可能有含量约为5pmol、N末端为D-L-P-Q的蛋白, 其N末端比天然干扰素分子缺少一个氨基酸, 推测是在发酵或后处理过程中由于某种细菌内源性或外源性因素造成的。

(三) 样品中残留的痕量卵清溶菌酶的检出与分析

通过仔细分析一批送检样品测序的色谱分析原始资料, 发现样品中可能残留有一种含量约为4pmol的杂质蛋白, 其N端第2至第9个残基的序列疑似V-F-G-R-X-E-L-A。利用FASTA程序在NBRF-PIR序列库中寻找这一八肽的同源序列, 结果在16336个序列中找到10个与之最相似的序列, 均属溶菌酶超家族。序列的同源排列表明(图3), 所测序列中的不明残基很可能是半胱氨酸, 而这一八肽序列最可能来自裂解细菌时所引入的卵清溶菌酶。利用反相FPLC技术对样品进行进一步纯化, 收集杂蛋白峰测序, 终于检出卵清溶菌酶N端15个氨基酸序列(表2), 证实了前面的分析。

讨 论

从几种样品测序结果看, γ 干扰素均质性最好, 其N末端均一, 终产品中几乎完全不含Met残基; 而 $\alpha 2a$ 干扰素和白细

胞介素-2则N末端不均一, 大肠杆菌去除Met残基的比例大致在40%左右。大肠杆菌加工起始氨基酸的情况以及终产品中带有Met起始氨基酸的蛋白所占的比例, 随菌株、表达产物、表达方式(包括载体和其他条件)以及生产流程而异, 利用不同宿主-载体系统表达同一产物, 加工情况可能大不一样。带有Met残基的产物在免疫原性和生物活性方面与天然人体蛋白质的细微差异, 至今尚无定论^[11]。因此通过各种细菌内源性或外源性措施去除Met的技术方法及其必要性, 仍有待进一步研究。

分析蛋白质样品中的痕量残留物, 在建立蛋白质纯化流程, 检定蛋白质混合物的组成成分以及鉴定基因工程产品等方面, 都具有十分重要的意义。我们在检出测序样品中的溶菌酶以后, 根据卵清溶菌酶在分子量及其他理化特性与待分离产物具有一定相似性的特点, 改用超声裂解细菌的方案彻底排除了杂质源, 最终获得良好效果。我们认为, 为分析蛋白质样品中残留的痕量蛋白, 可以分别或者同时采用以下几种方法: (1) 增大测序样品量, 提高痕量蛋白的绝对量; (2) 采用邻苯二甲醛阻断法突出痕量蛋白的检出信号; (3) 对测出的疑似序列进行序列库的检索比较; (4) 利用高分辨率的纯化技术分离痕量杂质蛋白单独测序。

参 考 文 献

- [1] 周建华等: 中华微生物学免疫学杂志, 5:69, 1985.
- [2] 智 刚等: 病毒学报, 7(2):142—147, 1991.
- [3] 张德震等: 病毒学报, 5(1):37, 1989.
- [4] 张智清等: 病毒学报, 4(3):97, 1988.
- [5] Simpson, R. J. et al.: *Anal. Biochem.*, 177:221, 1989.
- [6] George, D. G. et al.: *Protein: Structure and Function*, Ed. L'Italien, J. J., Plenum Press, New York, pp. 445—458, 1987.
- [7] Schlesinger, D. H. et al.: *Biochem. Biophys. Res. Commun.*, 81:282, 1974.
- [8] Margolies, M. N. et al.: *Protein: Structure and Function*, Ed. L'Italien, J. J., Plenum Press, New York, pp. 333—341, 1987.

[9] Pearson, W. R. et al.: *Proc. Natl. Acad. Sci. USA*, 85:2444, 1988.

[10] 金冬雁等: 病毒学报, 7(1):81, 1991。

[11] Sharma, S. K.: *Advanced Drug Delivery Reviews*, 4:87, 1990.

Evaluation of Recombinant Proteins—N-terminal Micro-sequencing and Detection of Residual Contaminants

Jin Dongyan Zhang Zhiqing Zhou Yuan Hou Yunde

(National Laboratory of Molecular Virology and Genetic Engineering, Institute of Virology, Chinese Academy of Preventive Medicine, Beijing)

The N terminal amino acid sequence of purified recombinant human interferon- γ , interferon- α 2a and interleukin-2 expressed in *E. coli* was determined on Applied Biosystems 477A Protein/Peptide Sequencer and 120A PTH Amino Acid Analyzer. From the raw chromatographic data of these samples, the identity and heterogeneity, the amount of methionine-plus species remaining in the final product, and the probable process contaminants were evaluated with the help of computer methods including database searching. General methods to characterize trace contaminants in protein sample were also discussed. Among the sequenced samples, only interferon- γ was shown to be N-terminal homogenous. Methionine-containing species were found in interleukin-2 and interferon- α 2a. Chicken eggwhite lysozyme was detected at very small amount in one batch of sample. The results provide valuable information for the development and improvement of preparation methods as well as regulatory responses to recombinant products.

Key words

Protein sequencing and sequence analysis; N-terminal heterogeneity; protein contaminants in recombinant products