

文章编号: 1672 - 058X(2009)01 - 0061 - 04

结构基因组学研究中的亲和接头

姜怀春¹, 李 宏²

(1. 重庆工商大学 学报编辑部, 重庆 400067; 2 重庆教育学院 生命科学与化学系, 重庆 400067)

摘 要:介绍了亲和接头在高通量生产目标蛋白质中的运用,这些亲和接头包括 6 组氨酸、麦芽糖结合蛋白、谷胱甘肽 - S - 转移酶和硫氧还蛋白,并讨论了高通量生产的蛋白质的溶解和复性。

关键词:结构基因组学;亲和接头;融合蛋白

中图分类号: Q 34

文献标识码: A

由于普通蛋白质含量低,难于纯化,更难于得到结晶水平的蛋白质纯度和蛋白质量,所以一些科学家使用蛋白质亲和接头进行大规模地高通量生产目标蛋白,然后,进行结构基因组学研究。

1 用大亲和接头 (large - affinity tags)超表达目标蛋白质

把目标蛋白质融合到亲和接头上,能增加蛋白质的表达,增加蛋白质的溶解性,保护目标蛋白质免受蛋白酶水解,增进蛋白质折叠,然后,用亲和层析很容易对带有接头的目标蛋白质进行纯化。但是,这种融合蛋白接头诱导的构象异质性 (conformational heterogeneity)阻碍晶体生长,需要把接头去掉才能用于晶体生长^[1]。

Smyth 等^[1]认为融合蛋白质 (fusion protein 或 chimeric protein)被广泛用于蛋白质研究的许多方面,如生物化学中的纯化、免疫检测、蛋白质治疗、疫苗开发、功能基因组学、蛋白质运输分析 (analysis of protein trafficking)、蛋白质 - 蛋白质或蛋白质 - 核酸相互作用^[2]。

在结构生物学研究中,常常需要毫克水平的均一性蛋白质样品 (homogeneous protein sample),目前,最常用的方法是用亲和层析从细胞裂解物中分离融合蛋白质。最常用的亲和接头包括 6 组氨酸接头 (hexa - histiding tag, His - tag)^[3],大肠杆菌麦芽糖结合蛋白 (maltose - binding protein, MBP)^[4,5],日本血吸虫 (*Schistosoma japonicum*)谷胱甘肽 - S - 转移酶 (glutathione - S - transferase, GST)^[6],大肠杆菌硫氧还蛋白 (thioredoxin, TRX)^[7],抗生素蛋白质接头 (avidin/streptavidin Strep tags)^[8],其他一些接头也被开发出来^[9]。

要把目标蛋白质生长成晶体,用于 X - 射线衍射研究,那么,象麦芽糖结合蛋白和谷胱甘肽 - S - 转移酶这样的大亲和接头通常要被去掉,去掉这些大亲和接头的方法是在工程化的连接子区域 (engineered linker region)进行位点专一性水解 (site - specific proteolysis),然后,用柱层析把目标蛋白质从亲和接头及蛋白酶中分离和纯化出来。但是,目前特别困难的是:对目标蛋白质和接头融合体的水解步骤产量低,目标蛋白质要沉淀、沉闷的水解最佳条件的摸索、蛋白酶昂贵的成本和难以收集到有活力的和结构完整的目标蛋白质^[10]。如果不水解接头与目标蛋白质的蛋白融合体,对接头与目标蛋白质的融合体进行结晶,由于融合体是多个结构域蛋白质 (multidomain protein),柔性连接子区域 (flexible linker region)的构象异质性 (conformational heterogeneity)使融合体难以形成次序好的衍射晶体 (well - ordered diffracting crystals)。并且,由于融合体蛋白质太大,构基因组学 (high throughput structural genomics)研究中,因为用这样的小接头,就常常不需要在晶体生长前把接头去掉^[11]。

收稿日期: 2008 - 11 - 07;修回日期: 2008 - 12 - 04。

作者简介:姜怀春 (1961 -),男,四川剑阁人,副研究员,主要从事生物化学研究。

1.1 用小亲和接头和大亲和接头表达融合蛋白 (fusion protein)

在 2000 年, Edwards 和 Stevens^[9,12]估计:有三分之一到一半的原核蛋白不能使用 6 组氨酸接头在细菌中以可溶解蛋白质的形式 (soluble form) 进行超表达 (overexpression), Braun, Hammarstrom, Shih 的研究报道表明,真核蛋白在细菌中表达的这个数据还要高^[13-15]。如果遇到 6 组氨酸接头在大肠杆菌中超表达的蛋白质不溶解,就需要改变大肠杆菌的培养条件,或者同时表达伴侣蛋白 (chaperone),或者改变寄主菌株,或者另外使用像 MBP、GST、TRX 或 NusA 这样的接头^[9]。精确了解结构域 (domain) 的边界的基因图谱并且用有限的蛋白酶进行水解,可以得到有单一结构域的蛋白质片段进行结构研究^[16]。也可以使用真核表达系统来研究单一结构域的蛋白质片段。

除了亲和纯化以外,前面提到过的大亲和接头有很多优势。比如,与 6 组氨酸接头相比,TRX 和 MBP 增加了 32 个小的蛋白质在大肠杆菌中的溶解性和表达^[17]。

1.2 结构基因组学研究中大亲和接头的使用

MBP、TRX 和 GST 的结构已经测定,这些结构可以用作搜索模型来解决分子置换法中的结晶学相位问题^[1]。另一个优势是成功测定这些大亲和接头的方法可以用来测定融合蛋白。然而,这些优势仅仅显示在用 GST 测定 5 - 42 个残基的蛋白质片段,成功研究融合蛋白质与大亲和接头的最大困难是在亲和接头和融合蛋白质之间的柔性连接物 (flexible linker) 诱导融合蛋白和大蛋白质接头产生构象异质性 (conformational heterogeneity)。大亲和接头 MBP 与融合蛋白质的晶体结构已经报道^[1], GST、TRX 和 MBP 与融合蛋白的 X 射线衍射晶体结构也已经报道^[18]。

1.3 融合到大亲和接头上的蛋白质的结晶

1.3.1 麦芽糖结合蛋白 (MBP)

人类 T 细胞 型白血病病毒外壳蛋白 gp21 的外功能区的两个片段与麦芽糖结合蛋白的融合蛋白已经被结晶^[19],由于 gp21 片段本身的溶解性很差,所以,把 gp21 融合到麦芽糖结合蛋白上,如果不修饰在 MBP 与 gp21 之间连接的氨基酸残基,MBP/gp21 融合蛋白 (335 - 445 个氨基酸残基) 不能生长成晶体。充分修饰以后,得到了晶体。另一个例子是将细胞表面抗原 CD38 与神经节苷脂 GT1b 的复合物与 MBP 接头融合,得到了晶体^[20]。Staphylococcus 附属调节子 R (SaR) 与 MBP 接头的蛋白质的晶体结构也已经报道^[21],通过序列分析显示在 SaR 与 MBP 之间的修饰与 MBP 与 gp21 之间连接的修饰相似,在 MBP C - 末端的带电荷的残基被突变成丙氨酸,连接的长度被缩短到 5 个残基,在 2.3 Å 水平上用分子置换法测定了这个融合蛋白的晶体衍射结构。

1.3.2 谷胱甘肽 - S - 转移酶 (GST)

已经报道了果蝇 DNA 复制相关元件结合因子与 GST 接头融合 (GST/DREF) 的晶体结构和小鼠雌激素受体结合区域与 GST 接头融合 (GST/ERHBD) 的结晶^[22,23],这两个结晶中,只有 GST/DREF 产生了能进行 X - 射线衍射的晶体,虽然较薄的 GST/ERHBD 结晶不能进行 X - 射线衍射分析,但是,电泳结构和电子显微镜图谱确认了该蛋白与 GST 接头的融合是完整的。

1.3.3 硫氧还蛋白 (TRX)

抗万古霉素蛋白 (VanH) 与硫氧还蛋白的融合体成功进行了克隆、表达、纯化、动力学分析和晶体生长,并在 3 Å 水平上测定了它的晶体结构,抗万古霉素蛋白是从屎肠球菌 (Enterococcus faecium) 中分离出来的一种 D 型乳酸脱氢酶^[24]。

Douglas 等发展和检测了一种简单而有效的蛋白质纯化方法,纯化的蛋白质或蛋白质片段可用于核磁共振和原二色散研究^[1],这种方法是在先前报道的基因表达和蛋白质溶解性筛选的方法的基础上发展而来的。使用这种方法可以在 1 d 以内将几个目标蛋白质与谷胱甘肽 - S - 转移酶融合,经过表达、纯化、用 NMR 研究来判断这些蛋白是否适用于进一步研究。只要这些蛋白质与接头蛋白形成可溶解的融合蛋白就可以测定。对于细胞培养、裂解和 NMR 样品制备仅需要将不同融合蛋白和蛋白酶的组合优化。

2 带有亲和接头的超表达蛋白质的复性

用大肠杆菌来生产用于结构基因组研究的蛋白质的主要制约瓶颈是超产的蛋白质形成不溶解的蛋白质

聚集体(包含体),由于超表达的 promoter使目标蛋白质基因产生大量的目标蛋白质,在大肠杆菌体内形成包含体,从包含体恢复蛋白质性质成为结构基因组学研究的瓶颈。Berkeley结构基因组研究中心(Berkeley Structural Genomics Center)建立了一种从层析柱上恢复包含体中蛋白质性质的方法,这种方法是利用“人造伴侣蛋白质帮助的复性”与亲和层析相结合的方法,该方法利用了亲和层析步骤,较少的时间消耗,无需过滤和浓缩,可以容易地用于蛋白质生产自动化和高通量生产程序。

到 2005年,许多国家组织了 25个结构基因组学研究计划^[25,26],在每一个结构基因组学研究流程中,报道了大量的从克隆、表达、纯化到结构测定的研究报告。收集的数据揭示了蛋白质结构、功能和调控的研究工作中,可溶解蛋白质的生产在结构基因组学高通量研究中起关键作用。由于大肠杆菌生长快、容易控制、成本低,大肠杆菌已经成为结构基因组学研究的主要表达系统。然而,在大肠杆菌中超产的重组蛋白质常常以不溶解的聚集体形式聚集在一起形成了包含体。通过克隆和表达同源基因,选择不同的表达载体,寄主菌株和生产条件,就可以生产可溶解的蛋白质。把包含体变成可溶解的蛋白质有很多优势,这些优势包括:包含体产量高,即使包含体对细菌有毒,产量也高,仅有的问题是如何把包含体变成折叠正确并且具有活力的蛋白质。通常使用的方法是缓慢地用透析,或者用盐酸脲或盐酸胍的缓冲液稀释,或者用柱层析将目标蛋白质恢复活力。层析方法包括离子交换层析,就是把变性的蛋白质固定到基质上,然后,用变性剂稀释来使蛋白质恢复活力。通过加入低相对分子质量的添加剂来降低蛋白质聚集,可以使包含体恢复生物活力。Roze-ma和 Gellman^[25]发展了一种新的以稀释为基础的让蛋白质恢复折叠的方法,就是变性的蛋白质首先暴露到含有洗涤剂的溶液中,来阻止蛋白质聚集,然后,用环糊精(cyclodextrin)剥去洗涤剂来促进蛋白质折叠。折中方法已经被用来模拟 GroEL - GroES chaperonin在细胞内的作用,并且被命名为人造伴侣蛋白(chaperone)帮助下的蛋白质折叠恢复方法(artificial chaperone - assisted refolding),用折中方法,有几个蛋白质被成功地恢复了折叠^[25]。但耗费时间,并且在大体积的透析和浓缩过程中,蛋白质容易损失。

Oganesyan等^[25]报道了一种在层析柱上让不溶解的变性蛋白质恢复活力的方法,并且用这种方法让在大肠杆菌中超表达的含有组氨酸接头(His - tagged)的蛋白质。步骤是:把溶解在脲里的包含体(inclusion body)首先挂到亲和层析柱上,让洗涤剂洗涤不让蛋白质错误折叠,然后,用 B - 环糊精(B - cyclodextrin)洗涤,去除洗涤剂,并促进正确的折叠,然后,用咪唑(imidazole)洗提目标蛋白质,再进一步用离子交换层析或者分子筛层析恢复折叠。Oganesyan等研究了 10个蛋白质,有 7个蛋白质恢复了 30% ~ 100%的折叠。

3 展 望

随着结构基因组学研究方法的突破,高通量制备目标蛋白已成为可能,再加上让高通量制备的蛋白质溶解方法的发展,必将使结构基因组学研究更上一层楼。

参考文献:

- [1] SMYTH D, MROZKIEW ICZM, MCGRATH W, et al Crystal structures of fusion proteins with large - affinity tags[J]. Protein Science, 2003, 12: 1 313 - 1 322
- [2] BECKW ITH J. The all purpose gene fusion[J]. Methods in Enzymology, 2000, 326: 3 - 7
- [3] BORNHORST J A, FALKE J J. Purification of proteins using polyhistidine affinity tags[J]. Methods in Enzymology, 2000, 326: 245 - 254
- [4] SACHDEV D, CHIRGW N J M. Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose - binding protein or thioredoxin[J]. Protein Expr Purif, 1998, 12: 122 - 132
- [5] SACHDEV D, CHIRGW N J M. Fusion to maltose - binding protein: Control of folding and solubility in protein purification[C]. Methods in Enzymology, 2000, 326: 312 - 321
- [6] SM ITH D B. Generating fusions to glutathione S - transferase for protein studies[C]. Methods in Enzymology, 2000, 326: 254 - 270
- [7] LAVALL IE E, LU Z, D BLASD - Smith E, et al Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli[C]. Methods in Enzymology, 2000, 326: 322 - 340
- [8] SKERRA A, SCHM DT TGM. Use of the Strep - tag and streptavidin for detection and purification of recombinant proteins[C]. Methods in Enzymology, 2000, 326: 271 - 304

- [9] STEVENS R. Design of high - throughput methods of protein production for structural biology[J]. *Struct Fold Des*, 2000 (8): 177 - 185
- [10] BANEYX F. Recombinant protein expression in *Escherichia coli*[J]. *Curr Opin Biotechnol*, 1999 (10): 411 - 421
- [11] BUCHER M, EVDOKMOV A, Waugh D. Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin - binding protein[J]. *Acta Crystallography*, 2002, 58: 392 - 397
- [12] EDWARDS A, ARROW SMITH C, CHRISTENDAT D, et al Protein production: Feeding the crystallographers and NMR spectroscopists[J]. *Nature Structural Biology*, 2000 (7): 970 - 972
- [13] BRAUN P, HU Y, SHEN B, et al Proteome - scale purification of human proteins from bacteria[J]. *Proc Natl Acad Sci*, 2002, 99: 2 654 - 2 659
- [14] HAMMARSTROM M, HELLGREN N, VAN D B, et al Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*[J]. *Protein Science*, 2002(11): 313 - 321
- [15] SHIH Y, KUNG W, CHEN J, et al High - throughput screening of soluble recombinant proteins[J]. *Protein Science*, 2002 (11): 1 714 - 1 719
- [16] MARSDEN R, MCGUFFIN L, Jones D. Rapid protein domain assignment from amino acid sequence using predicted secondary structure[J]. *Protein Science*, 2002(11): 2 814 - 2 824
- [17] HAMMARSTROM M, HELLGREN N, VAN D B, et al Rapid screening for improved solubility of small human proteins produced as fusion proteins in *Escherichia coli*[J]. *Protein Science*, 2002(11): 313 - 321
- [18] KE A, WOLBERGER C. Insights into binding cooperativity of MATa1/MATa2 from the crystal structure of a MATa1 homeodomain - maltose binding protein chimera[J]. *Protein Science*, 2003(12): 306 - 312
- [19] CENTER R, KOBEB, WILSON K, et al Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose - binding protein[J]. *Protein Science*, 1998(7): 1 612 - 1 619
- [20] KUKUMOTO M, NUREKIO, SHIROUZUM, et al Crystallization and preliminary X - ray diffraction analysis of the extracellular domain of the cell surface antigen CD38 complexed with ganglioside [J]. *J Biochem*, 2000, 127: 181 - 184
- [21] LU Y, MANNA A, LIR, et al Crystal structure of the Sar R protein from *Staphylococcus aureus*[J]. *Proc Natl Acad Sci*, 2001, 98: 6 877 - 6 882
- [22] KUGEM, FUJII Y, SHIMIZU T, et al Use of a fusion protein to obtain crystals suitable for X - ray analysis: Crystallization of a GST - fused protein containing the DNA - binding domain of DNA replication - related element - binding factor DREF [J]. *Protein Science* 1997 (6): 1 783 - 1 786
- [23] LALLY J, NEWMAN R, KNOWLES P, et al Crystallization of an intact GST - estrogen receptor hormone binding domain fusion protein [J]. *Acta Crystallogr*, 1998, 54: 423 - 426
- [24] STOLL V, MANOHAR A, GILLON W, et al A thioredoxin fusion protein of VanH, a D - lactate dehydrogenase from *Enterococcus faecium*: cloning, expression, purification, kinetic analysis and crystallization[J]. *Protein Science*, 1998(7): 1 147 - 1 155
- [25] OGANESYAN N, KMS - H, KMR. On - column protein refolding for crystallization [J]. *Journal of structural and Functional Genomics*, 2005(6): 177 - 182
- [26] HOU J, SMS G, ZHANG C, et al A global representation of the protein fold space [J]. *Proc Natl Acad Sci U S A*, 2003, 100(5): 2 386 - 2 390

Affinity tags in structural genomics research

JIANG Hua¹ - chun¹, LI Hong²

(1. Editorial Office, Chongqing Technology and Business University, Chongqing 400067; 2. Department of Life Science and Chemistry, Chongqing College of Education, Chongqing 400067, China)

Abstract: This paper introduces application of affinity tags to high - throughput production of targeted proteins, these tags include hexa - histidine, maltose - binding protein, glutathione - S - transferase and thioredoxin and discusses protein solubility and refolding for the proteins produced by high - throughput methods

Keywords: structural genomics; affinity tag; fusion protein

责任编辑:田 静