

标签蛋白融合技术概述：从分子生物学基础和生化基础到商用系统

K. Terpe

摘要：随着蛋白质组学的迅猛发展，重组蛋白质的使用在近年来大大增加。重组杂合体含有一个多肽融合担体也就是亲和标签可用于辅助目标蛋白的纯化，这已经被广泛使用。许多不同的蛋白质，结构域或者肽类能与目标蛋白融合。利用融合蛋白的有助于重组蛋白纯化和检测这个优点被广泛赞同。然而，很难选择正确的纯化系统用于特定的目标蛋白。本综述概述了使用最频繁的和有趣的系统：精氨酸标签（Arg-tag），钙调蛋白结合肽（calmodulin-binding peptide），纤维素结合结构域（cellulose-binding domain），DsbA，c-myc-tag，谷胱甘肽S-巯基转移酶（glutathione S-transferase），FLAG-tag, HAT-tag, His-tag, 麦芽糖结合蛋白（maltose-binding protein），NusA, Stag, SBP-tag, Strep-tag, 和硫氧还蛋白等。

引言

生产高度纯化和性状确切的重组蛋白质已成为在医药工业工作的蛋白质化学家的主要任务，近年来，一些抗原表位的肽类和蛋白质已被用于大量生产重组蛋白质。这些亲和标签系统具有以下特征：（a）一步的吸附纯化，（b）对三级结构和生物活性影响小，（c）可方便且专一的去除以产生天然蛋白质，（d）在纯化过程中重组蛋白的分析简便准确，（e）适用于大量的不同蛋白质。然而，每一个亲和标签都在其特定的缓冲液条件下纯化得到（表1）。因此有几种不同的策略用于大规模生产重组蛋白质。其中一种办法是使用很小的肽标签，这些标签不会与融合的蛋白质发生干扰。使用最为广泛的有多聚精氨酸，FLAG-，多聚组氨酸-，c-myc-，S-，and Strep II-tag. 对于某些应用，小标签无需去除。这些标签不像大标签具有免疫原性，经常可以直接作为抗原用于产生抗体。小标签对于融合蛋白的三级结构和生物活性的影响取决于标签的位置和氨基酸组成（Bucher et al. 2002）。另一种方法是使用大的肽类或蛋白质作为融合担体。大担体的使用可以增加目标蛋白的溶解性。缺点是对一些应用如结晶或抗体产生等，标签必须加以去除。一般来说，对于特定的目标蛋白很难决定最佳的融合系统。这取决于目标蛋白本身（如稳定性，疏水性），表达系统，纯化后蛋白的用途。本综述概述了使用最频繁的和有趣的标签-蛋白融合系统（表2）。

多聚精氨酸-标签(Arg-tag)

精氨酸-标签最早在1984年首次描述(Sassenfeld和 Brewer 1984)，通常由5或6个精氨酸组成。它已被成功用作细菌C末端标签，产生的重组蛋白质纯度高达95%，得率达到44%。精氨酸是碱性最强的氨基酸，带5个精氨酸标签的蛋白质可以结合到阳离子交换树脂SP-Sephadex上，而大部分杂蛋白不结合。结合后，带标签的蛋白质在碱性pH下运行线性NaCl梯度洗脱得到。当C末端为疏水性区域时，多聚精氨酸可能影响蛋白质的三级结构(Sassenfeld和Brewer 1984)。来自 *Pyrococcus furiosus* 的带精氨酸标签的麦芽葡聚糖结合蛋白已获得结晶(Bucher et al. 2002)。天然蛋白质的晶体在视觉上是分辨不出差别的，然而晶体在镶嵌图案和衍射图谱上有差别。精氨酸残基的C末端序列可用羧肽酶B处理去除。这一酶促处理已被成功用于一些例子，但常常由于低的切割得率或者在期望的蛋白质序列间发生不必要的切割而受到限制(Nagai和 Thogerson 1987)。精氨酸标签可用于在扁平表面固定化功能蛋白质；这对于研究其与配体的相互作用是重要的。GFP在其末端加上6个精氨酸标签可以通过这一序列可逆、专一结合到云母的表面，这已经作为电子扫描电镜应用的标准底物(Nock et al. 1997)。然而精氨酸标签并不常用，与第二标签联用是很有趣的蛋白质纯化工具。

多聚组氨酸-标签(His-tag)

已广泛采用的方法是利用固定化金属螯合层析纯化带有由多聚组氨酸残基组成的一个短的亲和标签的重组蛋白质。固定化金属螯合层析（IMAC；由Porath et al. 1975提出）的基础是固定在基质上的过渡态金属离子(Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+})与特定的氨基酸侧链之间的相互作用。组氨酸是与固定化金属离子作用最强的氨基酸，组氨酸的咪唑环作为电子供体容易与固定的金属离子形成配位键。含有连续组氨酸残基序列的肽类可在IMAC中有效保留。基质材料洗涤之后，可通过调节柱缓冲液的pH或者添加游离咪唑洗脱含多聚氨基酸序列的肽类(表 1)。纯化带组氨酸残

基的蛋白质的方法首次于1987年提出 (Hochuli et al. 1987)。Hochuli开发了亚硝基三乙酸 (NTA) 吸附剂用于金属螯合亲和层析。NTA 树脂形成四齿的螯合物, 非常适合与金属离子形成6价配位键, 2价用于可逆保留生物高分子。带有多聚组氨酸标签的二氢叶酸还原酶在1988年通过 Ni^{2+} -NTA基质成功分离 (Hochuli et al. 1988)。这一系统的纯化效率取决于多聚组氨酸的长度和溶解系统(表 3)。虽然在变性条件下系统对于带6个组氨酸标签的蛋白质纯化有效, 带3个组氨酸标签的蛋白质在生理条件下可有效纯化。然而带6个组氨酸标签的蛋白质可以在天然条件下、在低或高浓度盐的缓冲液中结合到 Ni^{2+} -NTA 基质上。结合后, 目标蛋白可用0.8到250 mM咪唑梯度洗脱。低浓度的咪唑 (如0.8 mM) 可减少带有组氨酸的宿主蛋白质的非专一吸附。6个组氨酸标签的蛋白质可用20到250mM咪唑浓度范围洗脱 (Hefti et al. 2001; Janknecht et al. 1991)。使用咪唑的缺点是咪唑会影响NMR试验, 竞争性试验研究和晶体学试验, 咪唑的存在经常导致蛋白质的聚集(Hefti et al. 2001)。另一种用于纯化多聚组氨酸标签蛋白质的材料是TALON。这由 Co^{2+} -羧甲基天冬氨酸(Co^{2+} -CMA)组成, 并偶联到固相的树脂上。TALON使得标签蛋白在温和条件下洗脱, 已有报导与 Ni^{2+} -NTA树脂相比, 其非特异性蛋白结合更少, 从而达到更高的洗脱产物纯度(Chaga et al. 1999a, b)。一种酶制剂通过SDS-PAGE鉴定的纯度高于95%。利用 Co^{2+} -CMA纯化可以使用的天然多聚组氨酸亲和标签长达19个氨基酸残基(HAT-tag; 序列见表 2)。氯霉素乙酰基转移酶, 二氢叶酸还原酶和绿荧光蛋白在N末端加入HAT标签可在温和条件下一步纯化纯度大于95%。平衡缓冲液和上样缓冲液中添加5mM 咪唑去除结合较弱的非专一蛋白质吸附, 而用150mM 咪唑洗脱HAT标签的蛋白质。也可降低pH到5.0洗脱标签的蛋白质。与盐酸胍相比, 尿素对于HAT标签蛋白质结合力的降低作用更强。然而HAT标签蛋白的过量表达试验仅在细菌中获得证实。多聚组氨酸亲和标签通常置于重组蛋白质的N端或C端。最佳的标签放置位置是针对特定的蛋白质而言的。纯化多聚组氨酸标签蛋白已成功用于大量的表达系统包括细菌 (Chen 和 Hai 1994; Rank et al. 2001), 酵母(Borsing et al. 1997; Kaslow和 Shiloach 1994), 哺乳细胞 (Janknecht et al. 1991; Janknecht 和 Nordheim 1992), 以及杆状病毒感染的昆虫细胞(Kuusinen et al. 1995; Schmidt et al. 1998)。超过100种结构的组氨酸标签蛋白质登记在蛋白质数据库中 (Protein Data Bank)。具有组氨酸标签的蛋白质在镶嵌图案和衍射图谱与天然蛋白相比变化不大(Hakansson et al. 2000)。虽然多聚组氨酸亲和标签因为相对较小且带电荷而几乎不影响蛋白质的活性, 但是通常并不能排除亲和标签影响其蛋白质活性的可能(Wu 和 Filutowicz 1999)。将亲和标签移到另一端(Halliwel et al. 2001)或者在变性条件下进行纯化可以解决这个问题。由于金属离子可吸附到NTA树脂, 带金属离子中心的蛋白质不推荐用该法纯化。由于 Ni^{2+} -NTA可被还原, 在厌氧条件下也不推荐使用该办法。然而利用组氨酸标签纯化蛋白质是通常使用的办法。

表1 亲和标签和基质洗脱条件

亲和标签	基质	洗脱条件
Poly-Arg	阳离子交换树脂	在碱性 pH>8.0 NaCl 线性梯度 0 - 400mM
Poly-His	Ni^{2+} -NTA, Co^{2+} -CMA (Talon)	150 mM 咪唑或低 pH
FLAG	Anti-FLAG 单抗	pH 3.0 或 2 - 5 mM EDTA
Strep-tag II	Strep-Tactin (修饰的链球菌抗生物素蛋白)	2.5 mM 脱硫生物素
c-myc	单抗	低pH
S	RNaseA酶 的S片断	3M硫酸胍, 0.2M 柠檬酸pH 2, 3M 氯化镁
HAT (天然组氨酸亲和标签)	Co^{2+} -CMA (Talon)	150 mM 咪唑或低 pH
钙调蛋白结合肽	钙调蛋白	EGTA 或1 M NaCl中加 EGTA
纤维素结合结构域	纤维素	1型: 盐酸胍或者尿素大于4M

		2/3型：乙二醇
SBP	Streptavidin (链球菌抗生物素蛋白)	2 mM 生物素
壳聚糖结合结构域	壳聚糖	内含子融合：30–50mM 二硫苏糖醇, β 巯基乙醇, 或半胱氨酸
谷胱甘肽S-巯基转移酶	谷胱甘肽	5–10mM 还原型谷胱甘肽
麦芽糖结合蛋白	交联淀粉	10mM 麦芽糖

表2 亲和和标签的序列和大小

标签	残基	序列	大小 (KDa)
Poly-Arg	5–6 (通常5个)	RRRRR	0.80
Poly-His	2–10 (通常6g)	HHHHHH	0.84
FLAG	8	DYKDDDDK	1.01
Strep-tag II	8	WSHPQFEK	1.06
c-myc	11	EQKLISEEDL	1.20
S	15	KETAAAKFERQHMS	1.75
HAT	19	KDHLIHNVHKEFHAAHANK	2.31
3x FLAG	22	DYKDHDGDYKDHDIDYKDDDDK	2.73
钙调蛋白结合肽	26	KRRWKNFIAVSAANRFKKISSSGAL	2.96
纤维素结合结构域	27 – 189	结构域	3.00 – 20.00
SBP	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQQGQREP	4.03
壳聚糖结合结构域	51	TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWE PSNVPALWQLQ	5.59
谷胱甘肽S-巯基转移酶	211	蛋白质	26.00
麦芽糖结合蛋白	396	蛋白质	40.00

表3 在6M盐酸胍和0.05M磷酸盐缓冲液中对于Ni²⁺-NTA吸附剂的亲和力 (Hochuli et al. 1988)

	磷酸盐		盐酸胍	
	截留 (%)	洗脱 (%)	截留 (%)	洗脱 (%)
多聚组氨酸二氢叶酸还原酶				
(His) ₂ -DHFR	30	10	<5	-
(His) ₃ -DHFR	90	75	<10	-
(His) ₄ -DHFR	>90	30	10	10
(His) ₅ -DHFR	>90	20	50	50
(His) ₆ -DHFR	>90	10	>90	90
DHFR-(His) ₂	>90	90	<5	-
DHFR-(His) ₃	>90	80	<10	-
DHFR-(His) ₄	>90	50	10	10
DHFR-(His) ₅	>90	40	50	50
DHFR-(His) ₆	>90	30	>90	90

FLAG-标签

FLAG-标签系统利用一短的亲水性8氨基酸肽(表1)并融合到目标蛋白(Hopp et al. 1988)。FLAG肽可与抗体M1结合。结合是否是钙离子依赖性的(Hope et al. 1996)或者无关的(Einhauer和Jungbauer 2000)仍有争议。FLAG-GFP的动力学结合试验通过BIACORE分析,结果表明在有 Ca^{2+} 离子存在下是一致的。其他目标是单抗M2和M5,每一个都具有不同的识别和结合特性。FLAG-标签可位于蛋白质的C端或N端。这一系统已用于各种各样的细胞类型,包括细菌(Blanar和Rutter 1992, Su et al. 1992),酵母(Einhauer et al. 2002; Schuster et al. 2000)和哺乳细胞(Kunz et al. 1992; Zhang et al. 1991)。该系统的纯化条件是非变性的,因此可以纯化有活性的融合蛋白,复合物的解离可通过添加螯合剂如EDTA或瞬间降低pH得到(表1)。这一系统的缺点是单抗纯化基质不如其他基质稳定如 Ni^{2+} -NTA或者Strep-Tactin。分离得到的蛋白质纯度在90%左右(Schuster et al. 2000)。通常小标签可用专门的单抗加以检测。为了改善FLAG标签的检测,已开发出3x FLAG系统。这一三联体FLAG抗原表位是亲水的,由22个氨基酸组成(表2)且可以检测表达低到10fmol水平的融合蛋白。来自*Pyrococcus furiosus*的带FLAG标签的麦芽葡聚糖结合蛋白已获得晶体(Bucher et al. 2002),晶体质量与无标签的蛋白质晶体非常相似。最后,FLAG-标签可加肠激酶处理去除,肠激酶专一识别该肽序列C末端的5个氨基酸残基(Maroux et al. 1971)。

Strep-标签 (Strep-tag)

Strep-标签是一氨基酸肽开发来用于在链球菌抗生物素蛋白(streptavidin)柱上亲和纯化相应的融合蛋白(Schmidt和Skerra 1993)。链球菌抗生物素蛋白(streptavidin)在特定位点第44, 45, 和47位的突变体与天然形式相比,对八肽Strep-tag II的亲合力更强(序列见表2, Schmidt et al. 1996; Voss和Skerra 1997; Korndrfer和Skerra 2001)。这一链球菌抗生物素蛋白突变体称为Strep-Tactin。Strep标签的蛋白质在生理缓冲液条件下结合到生物素结合区域中,并可用生物素衍生物温和洗脱。洗脱时推荐使用2.5 mM脱硫生物素。基质可以用4-羟基-偶氮苯-2-羧酸再生,这种物质在溶液中呈黄色,当结合到基质上则呈红色。结合条件是非常专一的。生物素化的蛋白质如大肠杆菌的羧基载体蛋白可以结合到Strep-Tactin上,但生物素或者生物素化的蛋白质则用抗生物素封闭。纯化条件是多种多样的。螯合剂,温和性去污剂,还原型去污剂,和高达1M的盐可以加到缓冲液中。变性的纯化条件如6M尿素可破坏Strep-tag/Strep-Tactin相互作用,但不破坏Strep-Tactin。标签与Strep-Tactin的作用接近 $1\mu\text{M}$ (Voss和Skerra 1997)。融合蛋白可以通过Strep-Tactin嵌合物或抗体专一地检测。标签可以加在蛋白质的C端或N端。重组Strep-标签-杂合体在细菌(Fontaine et al. 2002),酵母(Murphy和Lagarias 1997),哺乳系统(S_rdy et al. 2002; Smyth et al. 2000),植物(Drucker et al. 2002)和杆状病毒感染的昆虫细胞中都有产生。这一方法被推荐用于在厌氧条件下纯化带小标签的有活性的融合蛋白(Hans和Buckel 2000; Juda et al. 2001)以及含金属离子的酶。标签蛋白也有可能整合到膜中(Groß et al. 2002)。那些不带标签的膜蛋白亚基则不会被同时纯化。这一标签的特殊用途是它可用于真核表面展示(Ernst et al. 2000)。Strep-Tactin结合生物素和Strep-标签的相容性被用于观测EF0EF1-F-ATPase酶c-亚基寡聚体的旋转(Pänke et al. 2000)。近年来Strep-标签的使用大大增加。具有该标签的重组蛋白质可用于NMR和结晶实验研究(Ostermeier et al. 1997)。Strep-标签系统适合用于研究蛋白质之间的相互作用以及当那些大的标签或者带电荷的标签无效时的特殊用途。

c-myc-标签

在1985年开发出鼠抗c-myc抗体9E10(Evan et al. 1985)并被作为免疫化学试剂用于细胞生物学和蛋白质工程领域中。抗体的11个氨基酸抗原表位(表2)表达在不同的蛋白质框架中仍可识别9E10免疫球蛋白(Munro和Pelham 1986)。c-myc-标签已成功应用在Western-blot杂交技术,免疫沉淀和流式细胞计量术中(Kipriyanov 1996)。因此可用于监测重组蛋白质在细菌(Dreher et al. 1991; Vaughan et al. 1996),酵母(Sequi-Real et al. 1995; Weiss et al. 1998),昆虫细胞(Schioth et al. 1996),和哺乳细胞(McKern 1997; Moorby和Gherardi 1999)中的表达情况。表达在经土壤细菌转化的Arabidopsis细胞中的作用蛋白也有成功进行共免疫纯

化的报导(Ferrando et al. 2001)。c-myc-标签的蛋白质可通过偶联9E10单抗到二乙烯砷活化的琼脂糖上而进行亲和纯化。洗涤条件是生理条件的,接着用低pH洗脱。而低pH往往降低蛋白质的活力。纯化的c-myc标签蛋白质已获得结晶(McKern et al. 1997)。c-myc-标签可放在C端或N端(Manstein et al. 1995)。这一系统广泛用于检测但很少用于纯化。

S-标签

S-标签是个融合肽标签可以通过快速灵敏的均匀分析或者在Western blot中用比色法加以检测。该系统的基础是15个氨基酸残基的S-tag(表2)与103氨基酸残基的S-蛋白质之间的强烈相互作用,这两个都来自RNaseA(Karpeisky et al. 1994; Kim 和 Raines 1994)。S-蛋白质/S-标签复合物的Kd接近0.1 μ M,这取决于pH,温度和离子强度(Connelly et al. 1990)。标签由4个带正电荷残基,3个负电荷残基,3个不带电的极性残基和5个非极性残基组成。这使S-标签保持可溶。S-标签的快速分析是基于核酸降解活力的恢复。标签蛋白可以结合到S蛋白质基质上。洗脱条件较苛刻,如pH2.0的缓冲液(表1)。然而为了获得有功能的蛋白质推荐用蛋白酶切割标签。该系统可用于纯化的重组蛋白质来源包括细菌(Lellouch和Geremia 1999),哺乳细胞和杆状病毒感染的昆虫细胞抽提物。该系统经常与第二标签联用。RNase酶A高度灵敏的荧光底物的发现使得该系统可用于与高通量筛选联用的检测(Kelemen et al. 1999)。

钙调蛋白结合肽

1992年首次提出了含钙调蛋白结合肽的融合蛋白质的纯化(Stofko-Hahnet al. 1992)。该肽来自骨骼肌的肌浆球蛋白轻链激酶的C端26个残基(序列见表2)。它在0.2mM CaCl₂存在下与钙调蛋白的亲合力可达纳摩尔水平(Blumenthal et al. 1985)。紧密的结合可以进行使洗涤更严谨,确保仅少量的杂蛋白与融合蛋白共纯化。如果在第一步蛋白质没有完全洗脱下来,可进一步用EGTA和1M NaCl洗脱。由于没有可与钙调蛋白相互作用的内源蛋白质,该系统在纯化大肠杆菌的重组蛋白时具有很高的专一性。融合蛋白的回收率在80-90%。该系统可耐受还原剂和高达0.1%的去污剂(Vaillancourt et al. 2000)。真核细胞的纯化则不是很理想,因为有许多内源蛋白质可与钙调蛋白作用,也是钙离子依赖型的(Head 1992)。对于使用蛋白激酶A进行 γ [³²P]ATP同位素标记来说,钙调蛋白结合肽凝血酶融合标签是个极好的对象(Vaillancourt et al. 2000)。组氨酸标签的激酶可通过Ni²⁺-NTA层析去除。这可以用于研究蛋白质之间的相互作用或者噬菌体表达库筛选。钙调蛋白结合肽可放在N端或C端。钙调蛋白结合肽放在N端可能降低翻译的效率,而在C端则可高表达(Zheng et al. 1997)。

纤维素结合结构域

具有纤维素结合结构域(CBDs)的蛋白质种类已超过13种。CBD大小从4到20kDa,出现在多肽内的不同位置:N端,C端,或者中间。一些CBD与纤维素不可逆结合,因此可用于固定有活力的酶类(Xu et al. 2002);其它则可以结合,因此可用于分离和纯化。I型的CBD与透明的纤维素可逆结合,可作为亲和层析的有用标签。氢键的形成和范德华作用是结合的主要的推动力(Tomme et al. 1998)。纤维素的优点是它是惰性的,非专一的亲合力低,它具有多种形式,也已经被批准用于许多药物和人类的使用。CBD在相当宽的pH范围内,从3.5到9.5,均可与纤维素结合。标签可放在目标蛋白的N端或C端。标签的亲合力如此强,以致结合的融合蛋白需要用含有尿素或盐酸胍的缓冲液才能解离。这种变性的洗脱条件需要复性融合的目标蛋白。与II和 III型的CBD融合的蛋白质用乙二醇即可从纤维素上温和洗脱(McCormick和Berg 1997)。这种低极性的溶剂被推测破坏了结合位点的疏水作用。乙二醇可通过透析去除。重组CBD杂合体已在细菌,酵母,哺乳细胞和杆状病毒感染的昆虫细胞中产生(Tomme et al. 1998)。

SBP-标签

SBP-标签是一种新的链球菌抗生物素蛋白(streptavidin)结合肽,由38个氨基酸组成(序列见表2; Wilson et al. 2001)。标签与streptavidin的解离常数为2.5nM。SBP-标签蛋白质可用固定化的streptavidin进行纯化。洗脱条件非常温和,用2mM生物素。C端SBP标签的蛋白质已在细菌中表达并成功纯化(Keefe et al. 2001)。对于进一步的应用还不多,但这一标签似乎成为在包被有streptavidin的芯片上固定化蛋白质的有趣工具。

壳聚糖结合结构域

来自环状杆菌的壳聚糖结合结构域由51个氨基酸组成(Watanabe et al. 1994)。亲和标签通常与自切内含子联用。通常使用的内含子来自酿酒酵母的VMA1基因, 编码454个氨基酸(Chong et al. 1996, 1997)。其他短的内含子也有使用(Xu et al. 2000)。硫酯键的自切可用硫醇试剂如1, 4-二硫苏糖醇或者 β -巯基乙醇诱发(表2)。目标蛋白的C端或N端残基可影响体内和体外的切割(Xu et al. 2000)。高浓度盐或者非离子型去污剂可用于降低非专一结合, 从而增加纯度。当目标蛋白洗脱时, 未切割的融合前体和内含子蛋白质标签仍结合在壳聚糖树脂上, 再用1% SDS或者6M盐酸胍清除。C端或者N端融合壳聚糖结合结构域的蛋白质已在细菌中表达(Cantor 和 Chong 2001; Sweda et al. 2001; Wiese et al. 2001)。

谷胱甘肽S-转移酶-标签

1988年首次描述了融合有谷胱甘肽S-转移酶(GST)的多肽的一步纯化(Smith 和 Johnson 1988)。来自日本血吸虫(*Schistosoma japonicum*)、26-kDa的GST(Taylor et al. 1994)被克隆在大肠杆菌表达载体中。融合蛋白可从未经处理的裂解液中用固定化的谷胱甘肽亲和层析加以纯化。结合的融合蛋白在非变性条件下用10mM 还原型谷胱甘肽洗脱。在大多数情况下, 融合蛋白在水溶液中是可溶的, 并形成二体。GST标签可用酶学分析或免疫分析很方便的检测。标签有助于保护重组蛋白免受胞外蛋白酶的降解并提高其稳定性。在大多数情况下GST融合蛋白是完全或部分可溶的。仍不清楚那些导致不溶解的因素, 但一些例子是GST融合蛋白不溶解与疏水区域的存在有关。其它不溶的融合蛋白或是富含带电荷残基或者分子量大于100kDa。有时候如果不溶的融合蛋白在1% Triton X-100, 1% Tween, 10mM DTT, 0.03% SDS或1.5% sarcosyl缓冲液中是可溶的话, 也可进行亲和层析纯化(Frangioni和Neel 1993)。Sarcosyl抑制蛋白质与细菌外膜组分的共聚集。其它不溶蛋白的纯化必须采用常规的办法。推荐采用位点专一的蛋白酶如凝血酶或Xa因子从融合蛋白切除GST标签。PreScission 蛋白酶含有带GST-标签的人鼻病毒3C蛋白酶; 蛋白酶解后, GST载体和蛋白酶通过谷胱甘肽琼脂糖亲和层析去除。GST标签可位于N端或C端, 可用于细菌(Smith 和Johnson 1988), 酵母(Lu et al. 1997), 哺乳细胞(Rudert et al. 1996), 和杆状病毒感染的昆虫细胞(Beekman et al. 1994)。GST融合蛋白已成为分子生物学家的基本工具。它广泛用于研究蛋白质-DNA相互作用(Beekman et al. 1994; Lassar et al. 1989), 蛋白质-蛋白质相互作用(Mayer et al. 1991; Ron 和 Dressler 1992) 也作为抗原用于免疫学或疫苗研究(McTigue et al. 1995)。

麦芽糖结合结构域

40kDa的麦芽糖结合蛋白(MBP)由大肠杆菌K12的malE基因编码(Duplay et al. 1988)。1988年首次报导了通过与MBP融合的载体来促进外源蛋白在大肠杆菌中的表达和纯化(Di Guan et al. 1988)。融合蛋白可通过交联淀粉亲和层析一步纯化。结合的融合蛋白可用10mM麦芽糖在生理缓冲液中进行洗脱。结合亲和力在微摩尔范围。一些融合蛋白在0.2% Triton X-100或0.25% Tween 20存在下不能有效结合, 而其他融合蛋白则不受影响。缓冲条件为pH7.0到8.5, 盐浓度可高达1M。不能使用变性剂。MBP可增加在细菌中过量表达的融合蛋白的溶解性, 尤其是真核蛋白(Sachdev和Chirgwin 1999)。编码10个天冬酰胺残基的手臂序列插入到MBP和目标蛋白之间可增加特殊融合的机会, 而这种特殊的融合(蛋白)将与淀粉树脂紧密结合。MBP标签可通过免疫分析很方便的检测。有必要用位点专一的蛋白酶切割标签。如果蛋白在细菌中表达, MBP可以融合在蛋白的N端或C端(Sachdev 和 Chirgwin 2000)。放在N端可能降低翻译的效率。MBP系统与小的亲和标签联用而广泛使用(Hamilton et al. 2002; Podmore和 Reynolds 2002)。

NusA, TrxA 和 DsbA

当外源蛋白质在大肠杆菌中生产时的一个缺点是蛋白质经常聚集成不溶的折叠中间体, 也就是包涵体。为了获得有活力的蛋白质, 必须使用变性剂如8M尿素或6M盐酸胍溶解。避免包涵体的一种可能是使用大的亲和标签如GST或者MBP。亲水标签, 如转录终止抗终止因子(NusA), 大肠杆菌硫氧还蛋白(TrxA), 或者蛋白质二硫键异构酶I(DsbA)可增加溶解性。缺点是带有这些标签的蛋白质不能用专门的亲和基质纯化。融合蛋白构建时必须与可用于纯化的小亲和标签联用。尤其是当NusA蛋白增加了融合蛋白的溶解性时(Davis et al. 1999)。一般来说, 大肠杆菌NusA蛋白促进发夹状折叠和终止(Gusarov和 Nudler 2001)。一些在大肠杆菌中表达时不溶的蛋白质

与NusA在N端融合时则变成可溶。NusA经常与组氨酸标签联合使用(Harrisson 2000)。硫氧还蛋白可以融合在目标蛋白的C端或N端(Katti et al. 1990; LaVallie et al. 2000),但通常将trxA放在5'末端。DsbA增加目标蛋白在大肠杆菌胞质和周质中的溶解性。推荐采用位点专一的蛋白酶从融合蛋白切割NusA, TrxA或DsbA。切割位点可用作连接肽。

其他标签系统

也有其他在用的标签系统,在本综述中没有详细阐述:葡萄球菌蛋白A基因融合载体被开发来利用IgG亲和层析纯化融合蛋白(Uhl_n et al. 1983; Nilsson et al. 1985)。这一蛋白由于可与许多种属包括人类的免疫球蛋白的FC区域专一结合而非常适用于亲和纯化。与蛋白A相似,来自链球菌菌株G148的蛋白G也由于可与IgG的FC区结合而同样适用(Goward et al. 1990)。使用小肽标签的蛋白质的生物素化通常用于检测,固定化以及纯化(Cronan 1990)。不同的标签,如AviTag, PinPoint Xa蛋白纯化系统,和Bio-tag(Schatz 1993; Tucker和 Grisshammer 1996)也进行了描述。噬菌体The T7和V5 抗原表位是用于灵敏检测的有趣标签。其他用于检测的抗原表位有: ECS(肠激酶识别位点), HA(血粘素A), 和谷氨酸-谷氨酸。

标签的切割

亲和标签的存在可能影响待研究蛋白的重要特性或功能。可以通过位点专一的蛋白酶去除目标蛋白的标签,且切割不能降低蛋白质的活力。切割后蛋白酶的去除对于使用带亲和标签的重组蛋白酶或者生物素化的蛋白酶较为容易。生物素化的蛋白酶可通过Strep-tag/Strep-Tactin亲和层析直接加以纯化,或者通过第二步链球菌抗生物素蛋白亲和层析。标签的切割不加蛋白酶也是可行的,可加入自切位点(Xu et al. 2000)。最普遍使用的蛋白酶是:肠激酶,烟草蚀刻病毒(TEV),凝血酶,和Xa因子。目标蛋白的回收取决于切割效率。肠激酶由于专一识别5个氨基酸多肽(D-D-D-D-K-X1)并在赖氨酸的羧基上切割,是N末端融合通常选择的蛋白酶,在其他残基的不规则切割发生水平较低,这取决于蛋白质底物的构象(Choi et al. 2001)。肠激酶轻链的分子量为26.3kDa。一个单位定义为在23℃下切割50ug 融合蛋白在8小时内达到95%切割率所需的酶量。生化分析已经表明切割效率取决于D4K识别位点下游的X1 氨基酸残基(Table 4; Hosfield 和Lu 1999)。与其他标签不同, FLAG-tag(DYKDDDK)具有内在的肠激酶识别位点。TEV 蛋白酶是位点专一的蛋白酶,具有7个氨基酸残基的识别位点,序列为 E-XX-Y-X-Q-S,切割在两个保守的谷氨酰胺和丝氨酸之间进行(Dougherty et al. 1989)。X可以是各种氨基酸,但不是全部氨基酸都可行。最佳的切割序列为 E-N-L-Y-F-Q-S(Carrington 和 Dougherty 1988; Dougherty et al. 1988)。当TEV蛋白酶识别位点在两个结构域之间是结果最理想。当切割不理想时,插入增加结构灵活性的小连接肽可改善效率。高专一性,多种底物的活力以及低温下的有效切割使TEV蛋白酶成为从融合蛋白移除标签的理想工具(Parks et al. 1994)。凝血酶是一种广泛用于切割标签的蛋白酶。切割可在20到37℃之间切割0.3到16小时。与肠激酶相反,Xa因子和凝血酶切割后,在目标蛋白C末端增加了两个氨基酸。凝血酶最佳的切割位点是X4-X3-P-R[K]-X1'-X2',这里X4和X3是疏水氨基酸而X1'和X2'是非酸性氨基酸(Chang 1985; Chang et al. 1985; Haun 和 Moos 1992)。一些经常使用的识别位点是L-V-P-R-G-S, L-V-P-R-G-F, 和MY-P-R-G-N在X4-X3-P-R-G-X2'之间切割比在X4-X3-P-K-L-X2'更有效。其他识别位点是X2-R[K]-X1',这里X2或者X1'是甘氨酸。例如A-R-G和G-K-A,这里切割发生在第二个残基后。在凝血酶切割位点和N末端标签之间插入五个甘氨酸残基可改善切割(Guan 和 Dixon 1991)。通过这一“甘氨酸连接肽”只需较少酶量就可达到完全切割,而且也可以避免可能发生的错误切割。有效的消化缓冲液是纯的Tris缓冲液,pH8.0。在缓冲液中存在的NaCl具有抑制作用(Haun and Moos 1992)。凝血酶可从切割产物中用p-氨基琼脂糖亲和纯化,或者superose-12 FPLC柱凝胶过滤(Yu et al. 1995)或者苯甲脒琼脂糖移去。在标签和目标蛋白之间的Xa因子识别位点可作为有效的工具完全去除N末端的亲和标签。Xa因子在四氨基酸肽I-E[D]-G-RX1的羧基切割(Nagai和Thogerson 1984),这里X1可以是除了精氨酸和脯氨酸以外的其他氨基酸。切割可以在4到25℃之间进行。绝大多数Xa因子的分子量大约43 kDa,由两个二硫键连接的多肽组成,一条27 kDa,一条16 kDa。在SDS-PAGE,还原的肽链具有30 kDa和20 kDa的表观分子量。通过位点专一的蛋白酶如Xa因子切除标签有时候是无效的,有Xa因子切

割非专一的报导 (Ko et al. 1994)。原因可能是融合蛋白的不溶性或者变性剂的存在。可通过引入5个氨基酸的多聚甘氨酸区域来增加切割(Rodriguez 和 Carrasco 1995)。丹磺酰氯-谷氨酸-甘氨酸-精氨酸-氯甲基酮在室温下1分钟内可使Xa因子活性不可逆失活95%。虽然由于切割需要更长的孵化时间而且效率较低而使Xa因子越来越少使用, 但仍有使用Xa因子的成功报导 (Pryor 和 Leiting 1997)。

表4 不同X1氨基酸残基的肠激酶切割情况 (通过分光光度法测定, Hosfield 和 Lu 1999)。具有序列...-GSDYKDDDDK-X1-ADQLTEEQIA-... 的GST-钙调蛋白融合蛋白用于检测; 5mg蛋白质加0.2U肠激酶在37°C下切割16小时。

X1位置的氨基酸	切割效率 (%)
丙氨酸	88
甲硫氨酸	86
赖氨酸	85
亮氨酸	85
天冬酰胺	85
苯丙氨酸	85
异亮氨酸	84
天冬氨酸	84
谷氨酸	80
谷氨酰胺	79
缬氨酸	79
精氨酸	78
苏氨酸	78
酪氨酸	78
组氨酸	76
丝氨酸	76
半胱氨酸	74
甘氨酸	74
色氨酸	67
脯氨酸	61

结论

亲和标签在蛋白质纯化过程中是很重要的, 可以有助于稳定蛋白质并提高其溶解性。亲和层析通常可以达到90—99%的纯度。纯化系统的选择取决于蛋白质本身及其进一步的应用。有时候融合蛋白由于其标签没有暴露到表面而不能加以纯化。在变性条件下或者将标签置于蛋白质的另一个末端可以解决这个问题。在许多情况下, 第二亲和标签用于第二步亲和层析以增加纯度。(Pryor 和 Leiting 1997; Schioth et al. 1996)。作为选择, 一个标签用于纯化而另一个用于检测 (Vaughan et al. 1996; Lu et al. 1997)。如果两个不同的标签放在不同的末端, 则全长产物可通过两步亲和层析得到(Ostermeier et al. 1995; Sun 和 Budde 1995)。多个标签也是可能的, 每一个标签用于特定的用途。多个标签还可以连续多步纯化达到高纯度。这些高度纯化的蛋白质使得检测蛋白质之间的相互作用成为可能。结合的蛋白质可以通过质谱加以确认 (Honey et al. 2001)。一个特殊的多个标签用于串联亲和层析(TAP; Rigaut et al. 1999; Puig et al. 2001)。它由目标蛋白质, 钙调蛋白结合肽, TEV蛋白酶切割位点, 及用于固定化的蛋白A组成。TAP标签可以快速纯化特定的化合物。这一过程的用途类似于酵母双杂交筛选 (Fromont-Racine et al. 1997)。TAP标签可以作为蛋白质组开发的工具 (Gavin et al. 2002)。这一方法已经在酵母中获得证实但也可以用于其他细胞或微生物。许多对于其结合担体具有高亲和力的标签可以作为有用的工具在表面固定化肽类或蛋白质。有生物活性蛋白质的固定化对

于研究和工业应用来说都是重要的。此外，亲和标签技术的重要性将增加其在肽类/蛋白质芯片设计，高通量纯化，肽类/蛋白质库，大规模生产系统以及药物释放策略中的应用。

致谢 作者感谢A. Steinb_chel教授给予的大力支持，以及 F. Mayer教授在准备该文稿时的建议。

参考文献（略）