

分子生物学在蛋白质结晶中的应用*

袁彩 刘钊 黄明东**

(中国科学院福建物质结构研究所, 福州 350002)

摘要 获得具有高分辨率的蛋白质晶体是目前蛋白质结构测定的主要瓶颈. 蛋白质结晶受很多因素影响, 蛋白质自身是结晶时最重要的变量, 可以说, 蛋白质的内在特性在某种程度上决定了其能否结晶以及所得晶体分辨率的高低. 近年来分子生物学尤其是蛋白质工程的应用有效地提高蛋白质的溶解度、均一性及可结晶性等内在特性, 促进蛋白质的结晶, 成为提高蛋白质结晶能力和蛋白质晶体分辨率的有效途径.

关键词 蛋白质工程, 蛋白质结晶, 溶解度, 均一性, 可结晶性

学科分类号 Q71

蛋白质 X 射线结晶学中, 以往蛋白质晶体生长和结构解析 (即所谓的相角问题) 都是困难的步骤. 近年来, 由于计算技术的快速进步、同步辐射光源的广泛应用以及 X 射线结晶学研究方法 (如反常散射法) 的发展和进步, 相角问题的解决和结构模型的构建以及结构的修正都变得相对容易. 相比之下, 蛋白质结晶则成为结构测定、乃至结构基因组学发展的主要瓶颈, 而获得足够量高纯度的和适于结晶的蛋白质则是蛋白质晶体结构测定的前提. 上世纪 80 年代以来, DNA 克隆、重组蛋白表达和蛋白质工程为生产大量的目的蛋白和对蛋白质进行改造提供了强有力的工具, 近年来这些技术在改进蛋白质的可结晶性方面得到了越来越广泛的应用, 是蛋白质结晶学的一个重要发展方向, 本文结合作者的工作对此做一介绍.

1 基因工程与用于晶体生长的蛋白质表达

外源蛋白表达系统有原核表达系统、真核表达系统和体外无细胞系统. 每种表达系统都有其特有的优势和不足. 原核表达系统之代表——大肠杆菌系统, 由于其可有效又经济地大量表达蛋白质, 经常成为许多科研人员表达异源蛋白的首选目标, 但其不适合表达那些需翻译后修饰才有活性或有多个二硫键的蛋白质, 同时大肠杆菌表达还存在真核蛋白表达密码子稀有性的问题. 目前一些商业化菌株如增强胞浆内二硫键形成的 Origami 菌株、增强带大肠杆菌稀有密码子真核蛋白表达的 Rosetta 宿主菌, 使这些问题有所改善. 而可进行表达蛋白甲硫氨酸标记的菌株, 尤适用于结晶学研究, 如 B834 为甲硫氨酸营养缺陷型, 可用 ^{35}S -甲硫氨酸和碘代

甲硫氨酸对目标蛋白进行高特异活性标记. 真核表达系统在表达需分子伴侣辅助折叠或需翻译后修饰的蛋白质时具有很强的优势. 其中酵母和昆虫细胞表达系统蛋白质表达水平高、生产成本低, 但它们的加工修饰体系与哺乳动物细胞不完全相同, 哺乳动物细胞产生的蛋白质更接近于天然蛋白质, 但其表达量低、操作繁琐. 体外无细胞系统合适表达那些结合碘代甲硫氨酸、同位素标记、有毒性的蛋白质或膜蛋白^[1]. 载体上的一些改造, 也使大量蛋白质的获得更为容易. 亲和标记 (如 His6) 的应用使蛋白质纯化更为容易. 融合标签 (如 GST) 的加入使外源蛋白更稳定, 超级感受态细胞得到高转化效率, Creator、Gateway 等通用克隆系统的应用可将目的基因同时克隆到不同的表达载体上. 这些都使对不同蛋白质基因进行克隆、获得大量的蛋白质更为容易.

2 蛋白质工程与蛋白质结晶

获得高分辨率的晶体要求重组蛋白不仅具有所需的活性, 还要稳定、均一、纯度高. 不同蛋白质的结晶特性存在很大差异, 有些蛋白质很容易长出晶体, 而对另外一些蛋白质来说, 尽管费尽周折, 还是无法得到适于进行 X 射线衍射分析的晶体, 改变蛋白质的内在特性则成为解决这一问题的有效

*中国科学院百人计划项目 (IB021061), 福建省青年创新基金项目 (2002J013) 资助项目和中国科学院福建物质结构研究所前沿交叉学科项目 (IP031049).

** 通讯联系人.

Tel: 0591-3704996, E-mail: mhuang@fjirsm.ac.cn

收稿日期: 2004-09-01, 接受日期: 2004-09-30

途径. 一般来说, 删除或突变暴露位于蛋白质表面或强柔顺性环区的不保守和亲水性的氨基酸, 或在这些部位加入融合蛋白, 并不会改变蛋白质总体结构和活性, 这些增加、缺失和突变成为人们提高蛋白质可结晶性的首选^[2]. 对一些难以结晶和晶体分辨率差的蛋白质表面进行修饰以增加其可结晶性, 可称为蛋白质晶体工程. 以下从提高蛋白质溶解度、均一性和可结晶性三方面加以阐述.

2.1 蛋白质工程提高蛋白质溶解度

X 射线结晶学要求用于蛋白质晶体生长的蛋白质浓度在 2~10 g/L 以上, 才能形成足够大的晶体用于 X 射线衍射分析^[3], 不少蛋白质在浓缩时发生沉淀, 因此, 提高蛋白质的溶解度成为蛋白质结晶的一个关键. 将蛋白质疏水性表面残基突变为亲水性残基可增加蛋白质的可溶性. HIV 整合酶 (integrase) 催化结构域 F185K 突变体的溶解度由野生型的 1 g/L 上升至 40 g/L^[4]; 二氢叶酸脱氢酶表面非极性氨基酸突变为带负电荷氨基酸的一系列突变体中, 突变体 N48E、N130D 溶解度都有所提高, 而双突变体 N48E/N130D 的溶解度则出现大幅度上升^[5]; 人源瘦素蛋白 (leptin) W100E 突变体有较好的溶解度, 获得了分辨率为 0.24 nm 的晶体^[6]; 溶解度差的铜绿假单胞菌 AHL 合成酶与同属另一种高溶解度的 AHL 合成酶同源比较发现, 前者在 $\beta 2$ 和 $\beta 3$ 片层间的 β 转角处多了 4 个氨基酸, 这 4 个氨基酸残基用甘氨酸取代后则得到分辨率为 0.23 nm 的晶体^[7]. 去除跨膜部分也可以提高蛋白质的可溶性, 去除跨膜区和 C 端的可溶性叶绿体细胞色素 F 蛋白能结晶为 0.23 nm 的晶体^[8].

由于突变位点较难预测, 定点突变提高蛋白质溶解性经常是一个大量的试错工作^[9], 因此, 随机或定向进化成为提高蛋白质溶解度的有效途径之一^[10,11]. 对分支杆菌属肺结核 Rv2002 基因以易错 PCR 为手段、GFP 蛋白为筛选标记进行溶解性筛选, 其中三突变体 Rv2002-M3 (I6T/V47M/T69K) 溶解性远大于野生型蛋白, 结果, 它与 NAD 的二元复合物形成分辨率为 0.18 nm 的晶体, 与雄甾酮、NADH 的三元复合物形成分辨率为 0.24 nm 的晶体^[12].

蛋白质工程提高膜蛋白或低溶解性蛋白结晶的一个新策略是将蛋白质与一个载体连接, 这种载体蛋白是一种可溶性、已知结构的、稳定的蛋白质, 并具有便于选择的光谱或酶学特性, 使融合蛋白比天然蛋白更容易操作. 很多膜蛋白的亲水性环区都

很短, 可以把载体蛋白引入目的蛋白亲水性环区, 增加目的蛋白的亲水性表面积, 有利于这种融合蛋白形成晶体^[13]. 大肠杆菌乳糖透性酶 (lac permease) 与细胞色素 b562 蛋白形成的融合蛋白具有和天然蛋白一样的运输活性, 又具细胞色素可见光吸收特性, 这种红色透性酶在表达、纯化、浓缩和结晶过程都能进行方便的检测, 并且, 与无细胞色素载体的蛋白质相比, 它在溶液中更为稳定, 且可以达到 5 g/L 的溶解度, 用这种融合蛋白培养也获得了晶体^[14,15].

2.2 蛋白质工程提高蛋白质均一性

进行结晶的蛋白质要求有很高的纯度, 这不仅包括化学上的还包括构象上的均一性. 引起蛋白质构象不均一性 (heterogeneity) 的原因包括: 蛋白质表面糖基化、分子内存在柔顺环区 (如多结构域蛋白)、蛋白质不稳定性 (如易被蛋白酶水解)、蛋白质聚集等等. 许多来自真核生物的蛋白质经常带有糖基, 这些蛋白质表面的碳水化合物会导致蛋白质构象和电荷的多样性, 进而影响蛋白质晶体的生长或晶体的 X 射线衍射能力^[16]. 有很多报道用糖苷酶进行脱糖基化, 并取得了较好的效果. 但是脱糖基酶脱糖基化也存在一些问题, 如糖苷酶相对昂贵, 脱糖基化反应不完全, 脱糖基化后糖苷酶不易去除等等. 因此很多研究人员也采用蛋白质工程手段来获得脱糖基化蛋白质, 例如, 通过蛋白质工程去掉糖基化位点的组织蛋白酶 (cathepsin) K 给出了较高分辨率的晶体^[17]. 含游离半胱氨酸残基的蛋白质通常会在溶液中聚集而引起不均一性, 在溶液中加入还原剂是结晶这类蛋白质的一种方法, 但是有时还原剂并不能完全解决这一问题, 定点突变在这方面卓有成效. 天然分支酸裂解酶在溶液中易聚集而无法获得高分辨率的晶体, 通过定点突变将其两个引起聚集的半胱氨酸突变为丝氨酸则解决了这一问题, 得到分辨率为 0.11 nm 的晶体^[18]. 细胞色素 P450cam 在溶液中易形成二聚体, 将暴露在表面的 Cys334 突变为丙氨酸 (Ala), 也获得了分辨率为 0.12 nm 的高分辨率晶体^[19]. 同样, 分裂素激活蛋白 (MAP) 激酶 p38 α C162S 突变体能够减少蛋白质间的相互聚集、增加稳定性和均一性, 从而提高了晶体衍射分辨率^[20]. 作者在 GTP 酶 Sar1 蛋白的晶体生长中发现, Sar1 易在溶液中聚集成可溶性聚集体, 动态光散射测定发现其在溶液中分子质量为 1 446.8 ku, 与序列相近的 Arf1 进行同源比较发现, 其 N 端多了 9 个疏水残基片段, 删除前 3 个、

6个氨基酸的突变体由于 N 端疏水性的增强, 聚集程度更为严重, 而切除 N 端 9 个残基的 Sar1 突变体 N 端为亲水性氨基酸, 溶液中分子质量为 71.9 ku, 这种以二聚体形式存在的 Sar1 突变体则得到 0.17 nm 分辨率的晶体(表 1)^[21].

很多蛋白质由几个结构域组成, 结构域间通过柔顺性强的环区连接, 这经常引起蛋白质构象的不均一, 去除结构域间柔顺环区, 或取单独结构域进行结晶能够取得不错的效果. 完整的尿激酶无法得到可衍射的晶体, 而只具催化结构域、无糖基化的低分子质量尿激酶则得到分辨率为 0.25 nm 的晶体^[22], 在此基础上, 删除 C 端松散多肽并突变游离半胱氨酸和糖基化位点的低分子质量尿激酶, 则给出了 0.20 nm 高分辨率的晶体^[23].

大的融合标签如 MBP、TRX、GST 对增加蛋白质溶解性、防止目的蛋白被蛋白酶水解、增加蛋白质折叠及方便蛋白质纯化都有很大的好处. 但是, 这些多结构域蛋白不利于形成高序、高分辨率的晶体. 减少融合标签与目的蛋白间接头氨基酸数目或改变接头氨基酸特性, 使融合蛋白两个结构域间形成刚性接触是解决这个问题有效方法^[24]. 在 HIV-1 包膜蛋白 gp21 片段的结晶中, 将融合标签与目的蛋白间 25 个氨基酸组成的接头缩短为 3 个丙氨酸, 并将融合标签 C 端的带电氨基酸也变为丙氨酸后, 就得到了该片段的高分辨率(0.25 nm)晶体^[25], 用同样的方法也得到了酵母 MATa1 蛋白分辨率为 0.21 nm 的晶体^[26].

Table 1 Effect of N terminal deletion removes the aggregation of hamster Sar1 GTPase, leading to its crystallization

表 1 N 端缺失消除 Sar1 GTP 酶在溶液中的聚集而得到晶体

蛋白质	部分序列 (从 N 端开始)	分子质量 /ku	蛋白质晶体
野生型 Sar1 GTPase	MSFIFDWIYSGFSSVLQFLGLY-KKTGKLVF	1 446.8	无
三肽缺失体 Sar1	IFDWIYSGFSSVLQFLGLY-KKTGKLVF	7 770.8	无
六肽缺失体 Sar1	WIYSGFSSVLQFLGLY-KKTGKLVF	2 384.6	无
九肽缺失体 Sar1	SGFSSVLQFLGLY-KKTGKLVF	71.9	有
Arf1 GTPase*	GNIFANLFKGLFGKKEMRILM		有

*Sar1 同源蛋白.

2.3 蛋白质工程提高蛋白质可结晶性

蛋白质的可结晶性与其表面残基类型有着密切的关系, 蛋白质工程可以从晶格接触、熵值、空间位阻等多个方面改良蛋白质结晶性质从而得到高质量、高衍射分辨率的晶体. 在结晶过程中自由能的变化为 $\Delta G = \Delta H - T(\Delta S_{\text{protein}} + \Delta S_{\text{solvent}})$, 必须满足自由能的变化小于零, 才能保证结晶反应自发进行. 游离蛋白质分子形成有序结晶是一个熵减过程($\Delta S_{\text{protein}} < 0$), 但在蛋白质形成晶体结构时会释放出原来结合在蛋白质接触面上的水分子, 而使溶剂的熵增加($\Delta S_{\text{solvent}} > 0$), 部分补偿蛋白质有序化引起的熵减少. 所以, 能够释放足够量结合水的蛋白质将有利于形成晶体^[27], 反之, 含大亲水性残基的蛋白质可能有一个熵保护不易形成晶体. 研究也发现构象熵较大的赖氨酸和谷氨酸在晶格接触中是最少出现的, 但这 2 种氨基酸残基却常常在蛋白质表面出

现^[28]. 为了进一步具体了解这 2 种氨基酸对蛋白质结晶性质的影响, 将人源 RhoGDI 蛋白表面赖氨酸突变成丙氨酸, 谷氨酸突变成丙氨酸或天冬氨酸^[29], 发现除一个例外, 所有突变体蛋白都得到了晶体, 而野生型蛋白质却无法被结晶. 比较所获得的这些晶体解出的 4 个结构, 发现其中 3 个突变体出现新晶格间的晶格接触, 并且这些新晶格接触都是出现在突变位点上. 对嗜热水生菌天冬氨酸 tRNA 合成酶 AspRS-1 进行研究, 分析正交晶系空间群中的晶格接触, 并且在接触位置上进行不同点突变, 获得了 7 种不同的蛋白质突变体^[30]. 这些突变体在蛋白质表面的电荷分布、疏水性及晶格接触中的氢键分布都与野生型蛋白质不同. 当引入的突变破坏原来的晶格接触, 突变阻碍蛋白质结晶; 引入的突变增加新的晶格接触时, 突变蛋白则更易于结晶. 将 Vsp9p 蛋白 CUE 结构域中的 2 个赖氨酸

突变成丙氨酸, 该突变体与泛素 (ubiquitin) 形成的复合物产生了衍射分辨率高达 0.17 nm 的高质量晶体^[31].

3 结 语

蛋白质结晶受很多因素影响, 如培养条件、蛋白质浓度、纯度及蛋白质内在性质. 对有些蛋白质虽然进行了很多条件的筛选, 还是无法获得适于结构解析的晶体, 无疑蛋白质本身是最重要的变量. 近年分子生物学的迅速发展对蛋白质晶体生长领域产生了很大的影响, 蛋白质工程的应用, 改变了蛋白质的内在化学性质, 很多难以长成晶体的蛋白质形成了高分辨率的结晶, 相信随着对蛋白质结晶机理的深入了解, 蛋白质工程在蛋白质结晶上的应用会越来越引起人们的关注. 可以预计, 蛋白质工程和晶体生长方法学的结合最终将能够解决蛋白质结晶这一重要问题.

参 考 文 献

- Mittl P R, Grutter M G. Structural genomics: opportunities and challenges. *Curr Opin Chem Biol*, 2001, **5** (4): 402~408
- Kan C C. Impact of recombinant DNA technology and protein engineering on structure-based drug design: case studies of HIV-1 and HCMV proteases. *Curr Top Med Chem*, 2002, **2** (3): 247~269
- Stevens R C. Design of high-throughput methods of protein production for structural biology. *Structure Fold Des*, 2000, **8** (9): 177~185
- Dyda F, Hickman A B, Jenkins T M, *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, 1994, **266** (5193): 1981~1986
- Dale G E, Broger C, Langen H, *et al.* Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase. *Protein Eng*, 1994, **7** (7): 933~939
- Zhang F, Basinski M B, Beals J M, *et al.* Crystal structure of the obese protein leptin-E100. *Nature*, 1997, **387** (6629): 206~209
- Gould T A, Watson W T, Choi K H, *et al.* Crystallization of *Pseudomonas aeruginosa* AHL synthase LasI using beta-turn crystal engineering. *Acta Crystallogr D Biol Crystallogr*, 2004, **60** (3): 518~520
- Martinez S E, Huang D, Szczepaniak A, *et al.* Crystal structure of chloroplast cytochrome f reveals a novel cytochrome fold and unexpected heme ligation. *Structure*, 1994, **2** (2): 95~105
- Dale G E, Oefner C, D'Arcy A. The protein as a variable in protein crystallization. *J Struct Biol*, 2003, **142** (1): 88~97
- Pedelacq J D, Piltch E, Liong E C, *et al.* Engineering soluble proteins for structural genomics. *Nat Biotechnol*, 2002, **20** (9): 927~932
- Waldo G S. Genetic screens and directed evolution for protein solubility. *Curr Opin Chem Biol*, 2003, **7** (1): 33~38
- Yang J K, Park M S, Waldo G S, *et al.* Directed evolution approach to a structural genomics project: Rv2002 from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA*, 2003, **100** (2): 455~460
- Prive G G. Fusion proteins as tools for crystallization: the lactose permease from *Escherichia coli*. *Acta Crystallogr D Biol Crystallogr*, 1994, **50** (4): 375~379
- Prive G G, Kaback H R. Engineering the lac permease for purification and crystallization. *J Bioenerg Biomembr*, 1996, **28** (1): 29~34
- Zhuang J, Prive G G, Werner G E, *et al.* Two-dimensional crystallization of *Escherichia coli* lactose permease. *J Struct Biol*, 1999, **125** (1): 63~75
- Baker H M, Day C L, Norris G E, *et al.* Enzymatic deglycosylation as a tool for crystallization of mammalian binding proteins. *Acta Crystallogr D Biol Crystallogr*, 1994, **50** (4): 380~384
- Linnevers C J, McGrath M E, Armstrong R, *et al.* Expression of human cathepsin K in *Pichia pastoris* and preliminary crystallographic studies of an inhibitor complex. *Protein Sci*, 1997, **6** (4): 919~921
- Stover C, Mayhew M P, Holden M J, *et al.* Crystallization and 1.1-Å diffraction of chorismate lyase from *Escherichia coli*. *J Struct Biol*, 2000, **129** (1): 96~99
- Nickerson D, Wong L L, Rao Z. An improved procedure for the preparation of X-ray diffraction-quality crystals of cytochrome p450cam. *Acta Crystallogr D Biol Crystallogr*, 1998, **54** (3): 470~472
- Patel S B, Cameron P M, Frantz-Wattley B, *et al.* Lattice stabilization and enhanced diffraction in human p38 alpha crystals by protein engineering. *Biochim Biophys Acta*, 2004, **1696** (1): 67~73
- Huang M, Weissman J T, Wang C, *et al.* Protein engineering for crystallization of the GTPase Sar1 that regulates ER vesicle budding. *Acta Crystallogr D Biol Crystallogr*, 2002, **58** (4): 700~703
- Spraggon G, Phillips C, Nowak U K, *et al.* The crystal structure of the catalytic domain of human urokinase-type plasminogen activator. *Structure*, 1995, **3** (7): 681~691
- Nienaber V, Wang J, Davidson D, *et al.* Re-engineering of human urokinase provides a system for structure-based drug design at high resolution and reveals a novel structural subsite. *J Biol Chem*, 2000, **275** (10): 7239~7248
- Smyth D R, Mrozkiewicz M K, McGrath W J, *et al.* Crystal structures of fusion proteins with large-affinity tags. *Protein Sci*, 2003, **12** (7): 1313~1322
- Kobe B, Center R J, Kemp B E, *et al.* Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proc Natl Acad Sci USA*, 1999, **96** (8): 4319~4324
- Ke A, Wolberger C. Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. *Protein Sci*, 2003,

- 12 (2): 306~312
- 27 Vekilov P G. Solvent entropy effects in the formation of protein solid phases. *Methods Enzymol*, 2003, **368**: 84~105
- 28 Dasgupta S, Iyer G H, Bryant S H, *et al.* Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins*, 1997, **28** (4): 494~514
- 29 Mateja A, Devedjiev Y, Krowarsch D, *et al.* The impact of Glu-->Ala and Glu-->Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 Å resolution. *Acta Crystallogr D Biol Crystallogr*, 2002, **58** (12): 1983~1991
- 30 Charron C, Kern D, Giege R. Crystal contacts engineering of aspartyl-tRNA synthetase from *Thermus thermophilus*: effects on crystallizability. *Acta Crystallogr D Biol Crystallogr*, 2002, **58** (10): 1729~1733
- 31 Prag G, Misra S, Jones E A, *et al.* Mechanism of ubiquitin recognition by the CUE domain of Vps9p. *Cell*, 2003, **113** (5): 609~620

Application of Molecular Biology in Protein Crystallization*

YUAN Cai, LIU Shan, HUANG Ming-Dong**

(Fujian Institute of Research on The Structure of Matter, The Chinese Academy of Sciences, Fuzhou 350002, China)

Abstract Protein crystallization is the primary bottleneck step in X-ray protein crystallography. Among many variables that affect protein crystallization, such as growth condition, temperature, protein purity and concentration, the intrinsic characteristics of the protein is the most critical factor for protein crystallization. Common strategies and examples of protein engineering in improving protein solubility, homogeneity, and crystallizability are presented.

Key words protein engineering, crystallization, solubility, homogeneity, crystallizability

*This work was supported by grants from Hundred Talents Project from The Chinese Academy of Sciences (IB021061), Fujian Youth Innovation Grant (2002J013), and FJIRSM Frontier.

**Corresponding author . Tel: 86-591-3705814, E-mail: mhuang@fjirsm.ac.cn

Received: September 1, 2004 Accepted: September 30, 2004