

专论与综述

蛋白质工程

孟广震

(中国科学院微生物研究所,北京)

蛋白质工程 (Protein Engineering) 也称蛋白质的定点突变 (site-directed mutagenesis), 指的是根据蛋白质结构研究结果, 设计一个新蛋白质的氨基酸序列, 通过修饰编码原蛋白质的 DNA 序列, 最后创造出新的蛋白质的技术。

蛋白质工程是基因工程与蛋白质结构研究互相融合的产物。这一技术开辟了一条改变蛋白质结构的崭新途径, 使蛋白质和酶学研究与发展进入了一个新时期^[1]。

改变蛋白质的结构是研究蛋白质结构与功能的传统方法, 如天冬酸氨基甲酰转移酶活性中心一个残基的改变 (Gly → Asp), 不但导致活性丢失, 而且显著影响调节亚基与活性亚基之间的结合^[2]。在应用方面则经常要求通过改变蛋白质的结构来改变产品的特性。

用化学合成的办法人工全合成一种新的蛋白质目前在技术上和成本上还不现实。通用的方法仍然是随机突变及筛选技术^[3], 但盲目性极大。特别是当谋求改善某一蛋白质(或酶)的多方面性质时, 往往顾此失彼。如获得一个高活力的酶, 其稳定性可能极差; 或相反, 稳定性好, 活力却很不理想。而新近问世的蛋白质工程却为人们提供了一个改造蛋白质结构的有力手段。

一、重组 DNA 技术与蛋白质工程

重组 DNA 技术的建立使人们在很大程度上摆脱了对天然蛋白质的依赖, 特别当从天然来源获得蛋白质非常困难的时候, 重组 DNA 技

术更显示其无可争辩的优越性。

首先, 用于遗传物质的切割、聚合及拼接等工具酶的出现, 使重组 DNA 技术成为可能。重组 DNA 技术关键的一步是构建含某蛋白质结构基因的运载体(通常用质粒)。质粒的复制频率比染色体复制频率或细胞的分裂频率高, 这有利于高效表达。重组 DNA 技术能使蛋白质的合成脱离原寄主细胞调节基因的控制, 而置于新寄主细胞的调节基因的控制之下, 这也有利于高效表达。而且以微生物为重组 DNA 的受体细胞, 可以用发酵法大规模生产。

在重组 DNA 技术的基础上发展了蛋白质工程技术, 它通过对基因核苷酸序列的修饰来改变蛋白质的结构。但构成蛋白质的氨基酸数目毕竟太多了, 对其中任何一个氨基酸实行定向取代都须在已知的蛋白质结构信息的指导下完成。

二、蛋白质的结构测定与蛋白质工程

获得蛋白质立体结构的基本工具是 X 射线晶体衍射技术。根据电子密度图可以构建能显示键长和键角的结构模型, 这种分子模型过去用金属线、木棒及塑料等组装, 而现在逐渐为电子计算机所建立的模型所取代。在屏幕上可以清楚地显示蛋白质结构的骨架及在特定环境下的表面结构, 允许从分子的内部或外部观察蛋白质分子的某一断面的结构, 还可以用彩色图象、透视方法。在真时 (real time) 情况下描绘酶及底物分子的平移和转动。在屏幕上组建的

模型易于组装、操作、贮存及修正，这样的分子模型是指导蛋白质工程的有力工具^[1]。

从图 1 可以看到蛋白质结构测定在蛋白质工程的工作程序中的地位和作用。为获得丰富的野生型蛋白质，首先通过基因工程技术使编码该蛋白的基因克隆并表达。分析该蛋白的性质，估计其应用价值，在许多情况下可以直接开发利用野生型蛋白质并转化为有用的产物。为开辟新的应用需设计第二代蛋白质分子，这需借助于结构测定和分子模型。从蛋白质的三维结构出发，应用结构与作用机制的知识选择适当的修饰位点。首先在屏幕上设计出新的蛋白质模型，然后用蛋白质工程技术按设计蓝图生产出新一代的蛋白质。研究新一代蛋白质的三维结构又可为第三代蛋白质的设计提供信息反馈，上一次设计中未能预见到的细微结构改变可以在这次设计中考虑进去，如此反复无穷。实际上，一旦野生型的亲本系统被充分认识之后，便能很快地进行一群突变蛋白质的生产和分析。

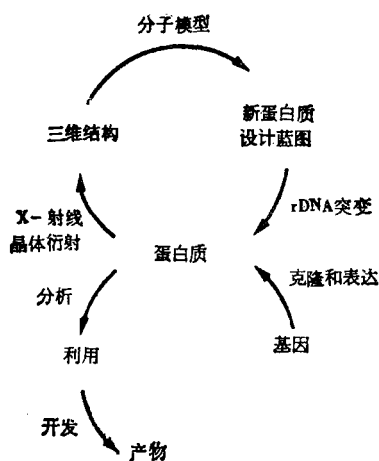


图 1 蛋白质工程工作程序图

三、基因修饰的方法学

如前所述，蛋白质结构的改变是通过基因修饰完成的。改变基因的核苷酸序列大体可以归纳为四种方法。现举例说明如后：

1. 基因的全化学合成 现以 Edge 等^[4]的人白细胞干扰素基因的全合成为例说明之。人

干扰素 α_1 由 166 个氨基酸组成。首先根据已知的核苷酸序列设计一个由 514 个碱基对组成的双链核酸序列，其中包括干扰素 α_1 的结构基因。起始和终止信号以及适当的限制性内切酶的位点。随后用快速固相法化学合成 66 个核苷酸片段，其中有 71 个核苷酸与天然序列有区别，但改变后的密码子依旧为相同的氨基酸编码，这种改变有助于在细菌寄主中获得高水平的表达。最后一步是按一定程序将各核苷酸片断用连接酶逐次连接成完整的基因。

合成的干扰素 α_1 基因两侧有单链粘性末端，这便于随后与质粒运载体的连接。该工作选用从质粒 pAT 153 派生出的 pPM 50 为运载体克隆这一合成的基因（图 2）。质粒 pAT 153 的抗四环素区域在 EcoR1 和 Ram H1 的切点之间。质粒 pAT 153 经以上两种内切酶消化之后，分离水解物中的大片段，用 DNA 聚合酶对切口作适当修饰，将大肠杆菌乳糖操纵子中 95bp 的启动子插入切口，构成质粒 pPM 50。用限制酶 Bam H1 和 Sal 1 消化 pPM 50，提纯水解物中的大片段，再将合成的干扰素基因插入，构成重组的质粒 pIFS。用此质粒转化大肠杆菌，干扰素 α_1 基因可获得高水平的表达。

上述全合成基因技术有许多优点。可以按需要在一个或多个位置上任意改变核苷酸序列，最后获得多点突变的蛋白质产物。可以按需要在核苷酸序列中安排适当的限制酶切点，便于随后进一步的修饰及基因操作。但该技术的最大困难是消耗大量人工合成的寡核苷酸，这是一般实验室所承受不了的。除上述干扰素外，用全合成法还成功地完成了生长激素释放的抑制因子，胰岛素及脱乙酰基胸腺素基因的全合成。预期这一技术将会得到进一步发展。

2. 基因直接修饰法 从野生型基因中将靶子密码子用酶法除去，然后按需要引进化学合成核苷酸。现以 β -内酰胺酶为例作以下说明：

β -内酰胺酶是破坏青霉素内酰胺环的水解酶，它的存在是微生物对青霉素耐药性的根源。该酶编码于质粒 pBR 322 的抗氨苄青霉素的基因 ApR 上，因此抗药性可以作为筛选该酶的

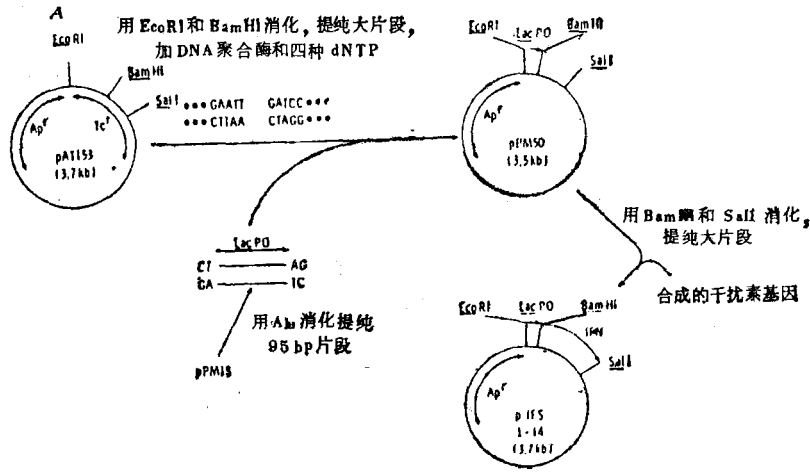


图2 重组 pIFS 质粒的构建

标记。从 DNA 及蛋白质的序列分析得知该酶组成均一。从对该酶各种底物和自杀性钝化剂的研究得知 Ser-70 是该酶活性中心的亲核基团, β -内酰胺的水解作用显然通过涉及 Ser-70 的羰基中间物进行。为进一步研究 Ser-70 对催化作用的重要性, Sigel 等^[5]用蛋白质工程将编码 Cys 的密码子 TGC 取代了 Ser-70 的密码子 AGC。取代程序如图 3 所示。

作者发现限制酶 HgiA 1 的切点 AGCA (410—413) 恰好位于 Ser-70 的密码子上。首先用该酶消化质粒 pBR 322, 然后用 3'→5' 核酸外切酶从 3' 到 5' 消除以上四种核苷酸, 形成一个有缺口的质粒。由于质粒序列太短, 用连接酶不容易将四体寡核苷酸直接插入, 因此先插入一个带有 Kpn1 切点的八体寡核苷酸 T-G-G-T-A-C-C-A, 构成质粒 pIS 1, 然后用限制酶 Kpn 1 消化, 用 3'→5' 核酸外切酶消除多余的 G-T-A-C 四聚核苷酸, 最后环化为质粒 pIS2。如此便完成了以 TGC 取代 AGC 程序。

用上述不同质粒转化大肠杆菌, 对氨苄青霉素表现不同抗性。带有 pBR 322 质粒的大肠杆菌能在 40mg/L 的氨苄青霉素的浓度下生长 (耐药性强); 带有 pIS 2 质粒的大肠杆菌在 10 mg/L 浓度下生长下降 90% (耐药性下降, β -

内酰胺酶活力下降); 而带有 pIS 1 质粒的大肠杆菌在 5mg/L 浓度下完全不形成菌落 (耐药性消失), 这是因为八聚核苷酸的引入使质粒 pIS1 密码的阅读框架偏移, 不再能合成 β -内酰胺酶的缘故。测定带有不同质粒的大肠杆菌提取物的 β -内酰胺酶活力获得相似结果: 带有 pBR 322 的与 pIS 2 的活力比为 30:1。被 Cys 取代的酶为巯基 β -内酰胺酶, 容易被巯基试剂 pCMB 所钝化, 而野生型 β -内酰胺酶对 pCMB 不敏感。

基因直接修饰法避免了化学合成方法需消耗大量合成寡核苷酸的缺点, 但要求找到一种限制性内切酶, 其切口应恰好位于被修饰的核苷酸上。

3. 寡核苷酸引物存在下新 DNA 分子的体外复制 现以酪氨酰 tRNA 合成酶 (TyrRS) 为例说明如下:

嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*) 的 Tyr RS 催化酪氨酸和 tRNA 之间的氨酰化反应。反应的第一步是用 ATP 活化酪氨酸形成酪氨酰腺苷酸; 第二步是将酪氨酰基转移到 tRNA 分子上形成酪氨酰 tRNA^[6]。从克隆在质粒 pBR 322 的该酶基因测定了它的核苷酸序列。该酶在大肠杆菌中的表达可达 10% 可溶性蛋白, 在 3 Å 分辨率水平上完成了该酶

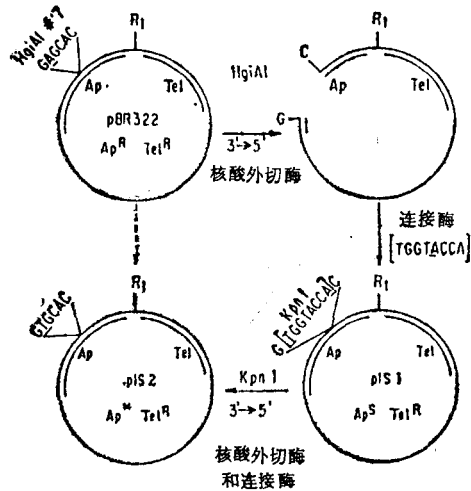


图3 将质粒 pBR 322 转为 pIS2 的步骤

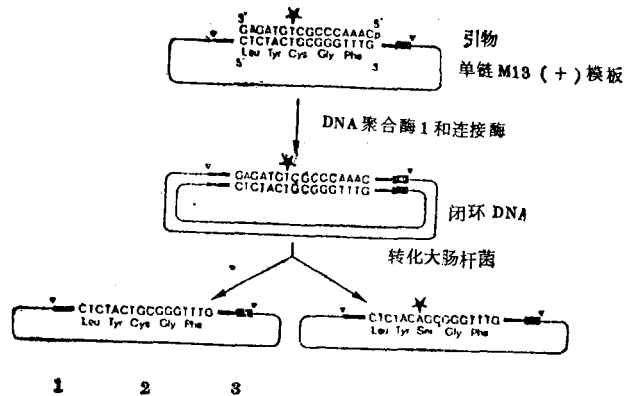


图4 酪氨酸 tRNA 合成酶的基因在引物引导下的体外突变

结晶的衍射图谱，并鉴定了酪氨酸与 ATP 在该酶的结合部位^[7]。在该酶分子的模型中可以看到 Cys-35 与酪氨酰腺苷酸的核糖 3' 羟基相结合，故 Winter 等^[8]选定 Cys-35 为蛋白质工程靶子位点。如果用与 Cys 结构十分类似的 Ser 取代 Cys-35，估计会对底物的结合产生影响。

用 Ser 取代 Cys-35 的主要工作程序如图 4 所示。1) 先把克隆在 pBR 322 的该酶基因及其启动子一起再克隆到噬菌体 M13mp 93 的 DNA 上。2) 人工合成引物寡聚核苷酸 5' CAAACCCGCT*GTAGAG3'，该引物除一个核苷酸 (T) 失配外，其它均与噬菌体单链核酸模板互补。3) 在上述模板核酸存在下，在 DNA 聚合酶的催化下引物寡核苷酸延伸，最后新合成的链用连接酶封闭成环。4) 用克隆的双链环状核酸分子转染大肠杆菌细胞。5) 从大肠杆菌培养物中分离各空斑噬菌体，点在硝酸纤维素滤纸上，用 ³²P 标记的引物与突变体核酸杂交，然后用放射自显影鉴定突变的核酸。被突变噬菌体侵染的大肠杆菌对 Tyr-RS 表达效率甚高，占细胞可溶性蛋白的 50%。

从突变体细胞培养物所获得的提取物具有相当高的酶活力，这说明，在 35 位上对 SH 基的要求不太严格，但突变酶的比活力较野生型

低得多。还发现突变的酶对酪氨酸的 K_m 不变 (15 μM)，但对 ATP 的 K_m 则由 0.9 增加到 4.1mM， V_{max} 从 1.4 降到 0.9 s^{-1} 。从 Tyr RS 的分子模型得知 Cys-35 位于 ATP 的结合部位，Cys-35 被 Ser 取代后影响了酶对 ATP 的 K_m 值，两个结果相符，说明 Cys-35 为 ATP 的结合提供了最合适的结构。

引物存在下的 DNA 体外复制方法被认为是蛋白质工程的通用方法^[9]，耗用化学合成寡核苷酸的量比基因全合成法少得多，适用范围又比基因直接修饰法广泛得多。

4. 盒式突变技术 上述三种方法都能成功地进行基因修饰。但氨基酸取代后的结果，预见性是有限的。在此情况下，只好以其他 19 种氨基酸一一取代，然后逐个分析各被修饰蛋白质的性质，这样做显然是很费时费力的。为提高工作效率，最近 Wells 等^[10]报告了一种盒式诱变 (Cassette mutagenesis) 技术，用此技术蛋白质中的某氨基酸残基，可以一气呵成地分别为其他 19 种氨基酸取代。

枯草杆菌蛋白酶 (subtilisin) 的活性中心涉及 Ser-221 残基，与其邻近的是 Met-222。由于 Met-222 的存在，使该酶易被氧化失活。因为事先无法确定哪种氨基酸取代 Met-222 后会既改善对氧化的稳定性，又不致剧烈降低酶的

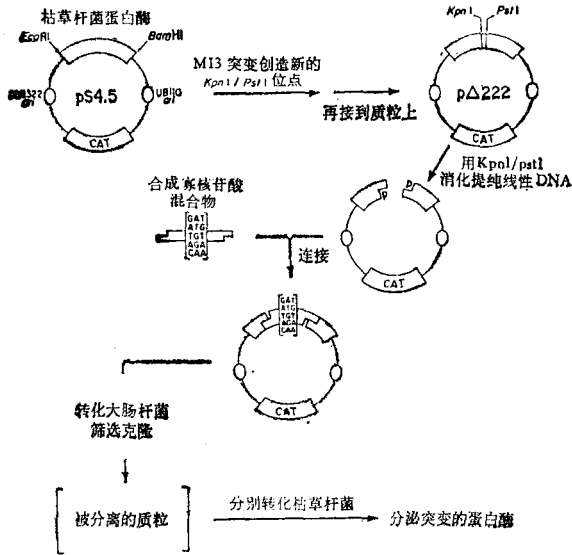


图5 盒式突变法进行枯草杆菌蛋白酶的氨基酸取代示意图
活力,故采用这种盒式突变技术对该酶进行修
饰。

完整的枯草杆菌蛋白酶的基因在质粒 PS 4.5 的一个 1.5Kb 的 EcoRI-BamHI 片段中(图 5)。将此片段连接到噬菌体 M13mp9 上,构成一个单链重组噬菌体 DNA (M13mp9 SUBT)。另外,合成一个 38 体寡聚脱氧核苷酸引物,该引物缺失包括 Met-222 在内的 10 个核苷酸,并在其两侧分别创造了限制酶 Kpn I 和 pst I 的新切点(图 6)以此寡核苷酸为引物,以 M13 mp9 SUBT 为模板,在 DNA 聚合酶 1(Klenow)和 T₄ DNA 连接酶催化下体外复制成新的突变 DNA 分子。用 EcoRI-BamHI 消化后所获得的 DNA 片段,再克隆到穿梭质粒 pBS 42 上,获得质粒 pΔ 222。用限制酶 Kpn I 和 Pst I 消化这个质粒,形成切口。将 5 种双链 25 体寡聚核苷酸混合物(库)在连接酶催化下插入切口,形成五种不同的重组质粒。用此混合质粒转化大肠杆菌,培养后提取质粒,用限制酶 KpnI 再次处理以消除未突变的质粒 pΔ 222 所造成的

	220	222	230
野生型氨基酸序列	Ala	Tyr	Asn Gly Thr Ser Met Ala Ser Pro His Val Ala Gly Ala Ala
野生型 DNA 序列	5'-GCG TAC AAC GGT ACG TCA ATG GCA TCT CCG CAC GTT GCC GGA GCG GCT-3'		
	3'-CGC ATG TTG CCA TGC AGT TAC CGT AGA GGC GTG CAA CGG CCT CGC CGA-5'		
Δp222 DNA 序列	5'-GCG TAC AAC <u>GGT ACC</u> TCA	CG CAC	<u>GCT GCA</u> GGA GCG GCT-3'
	3'-CGC ATG TTG CCA TGG AGT	GC GTG CGA CGT CCT CGC CGA-5'	
	<u>Kpn I</u>		<u>Pst I</u>
被 Kpn I/pst I 消化后的 Δp222 序列	5'-GCG TAC AAC GGT AC		<u>pGGA GCG GCT-3'</u>
	3'-CGC ATG TTG Cp		A CGT CCT CGC CGA-5'
突变后的 DNA 序列	5'-GCG TAC AAC GGT ACG TCA NNN GCA TCT CCG CAC GTT GCA GGA GCG GCT-3'		
	3'-CGC ATG TTG CCA TGC AGT NNN CGT AGA GGC GTG CAA CGT CCT CGC CGA-5'		

图6 盒式突变技术的程序

任何污染。用净化过的质粒再转化大肠杆菌,分离单菌落的质粒 DNA,测定 DNA 序列,筛选突变体质粒。将 5 种突变体质粒分别转化蛋白酶缺陷型枯草杆菌突变株 BG 2036,从这些转化体细胞中分别分泌不同突变的枯草杆菌蛋白酶。上述一个实验组一举可得 5 种不同的蛋白酶(以 5 种不同氨基酸取代 Met-222,其他序列与野生型相同)。分 4 个实验组进行则可以用全部 19 种氨基酸分别取代 Met-222,从而大

大简化了操作程序,与其他几种方法比较有其独到优越性。

四、蛋白质工程的应用及发展前景

迄今通过蛋白质工程已完成了 20 多种蛋白质分子的结构改造。全面总结它的成就可能还为时尚早,但已显示了这一新技术广阔的发展前景。

在蛋白质的结构与功能方面已获得不少有

价值的信息。如将 β -内酰胺酶的 Ser-70 和 Thr-71 进行了倒位双突变,氨基酸序列变成 Thr-70 和 Ser-71。双突变的 β -内酰胺酶活力丢失,说明 Ser-70 对该酶催化功能的重要性^[9]。从枯草杆菌蛋白酶的空间结构得知 Gly-166 残基伸向活性中心裂缝内部,用带有负电荷的 Asp 和 Glu 分别取代这个残基后,对底物的亲合力和催化效率发生了戏剧性的改变。而用中性的 Asn 和 Gln 取代后未发现生化性质的明显改变^{[11][12]}。成熟的胰岛素分子由 A 链和 B 链组成,而仅存在于胰岛素原子中由 35 个残基组成的 C 链在分泌过程中被切除。Wetzel 等^[13]构建了仅为 6 个氨基酸残基编码的微型 C 肽(mini-C)胰岛素原基因,虽然 C 链大大缩短,胰岛素仍能获得正确的折叠。此外蛋白质的三维结构归根结底决定于它的一级结构和其所处的外部环境。理论上,从一级结构预测空间结构是完全可能的,但蛋白质的结构必竟太复杂了,解决这个问题太困难了。蛋白质工程无疑将会提供大量关于一级结构与高级结构之间相互关系的有用信息,并为最终解决这个问题做出贡献。

蛋白质工程的研究在日后对商业上有重要价值的蛋白质的开发显然会带来利益。某些蛋白质已引起科学家们的注意,如核糖-1.5-二磷酸羧化酶在光合生物中司 CO_2 的固定,但也能利用分子氧进行光呼吸,使约 50% 的被固定的碳白白损失。如能消除或降低该酶的光呼吸作用,必将大大提高光合作用的效率,在这方面,或许蛋白质工程也能发挥重要作用。蛋白质工

程在医学上的应用也受到重视,如 α -抗胰蛋白酶有抑制弹性蛋白酶的作用,后者能水解肺组织的胶原蛋白。吸烟者 α -抗胰蛋白酶中的甲硫氨酸易被氧化,失去对弹性蛋白酶的抑制作用,最后导致肺气肿形成。如果用缬氨酸取代甲硫氨酸可望对肺气肿有防治作用。

蛋白质本身的多样性为其多种应用创造了条件,而随着蛋白质工程的发展,在不久的将来又会有大批人工创造的新蛋白质家族出现。总之,蛋白质工程对于探索者确是一块沃土,在付出必要的劳动之后,一定会结出丰硕的果实。

参 考 文 献

- [1] Rastetter, W. H.: *Applied Biochem. Biotech.*, 8, 423—436, 1983.
- [2] Kim, R. et al.: *J. Biol. Chem.*, 256, 4691, 1981.
- [3] Ulmer, K. M.: *Science*, 219, 666—671, 1983.
- [4] Edge, M. D. et al.: *Nature*, 292, 756—762, 1981.
- [5] Siga, I. S. et al.: *Proc. Natl. Acad. Sci. USA*, 79 (23), 7157—7160, 1982.
- [6] Fersht, A. R. and Jakes, R.: *Biochemistry*, 14, 3350—3356, 1975.
- [7] Rubin, J. and Blow, D. M.: *J. Mol. Biol.*, 145, 489—500, 1981.
- [8] Winter, G. et al.: *Nature*, 299, 756, 1982.
- [9] Dalbadie-McFarland, G. et al.: *Proc. Natl. Acad. Sci. USA*, 79, 6409—6413, 1982.
- [10] Wells, J. A. et al.: *Gene*, 34, 315—323, 1985.
- [11] Estell, D. A. et al.: *World Biotech. Report*, 2, 181—187, Online, 1984.
- [12] 孟广震:《生物工程学报》, 1(2), 1—8, 1985.
- [13] Wetzel, R. et al.: *Gene*, 16, 63—71, 1981.
- [14] Perutz, M.: *New Scientist*, 13, June, 12, 1985.

[本文于 1985 年 10 月 14 日收到]