

## 定位诱变

吴汝平

(中国科学院上海药物研究所, 上海)

自从重组 DNA 技术问世以来, 基因已从生物体的完整基因组中取出, 克隆到载体上。随着 DNA 顺序分析、寡核苷酸合成及其他 DNA 操作技术的出现和迅速发展, 人们已能按照自己的意愿在体外使基因的特定区域有效地产生突变。这种被称为定位诱变的技术已成为分子遗传学的一种重要工具, 它可以使产生突变的频率在百分之几的范围中, 这样就有可能在缺乏表型选择时进行遗传分析, 对研究基因表达调控、探索蛋白质结构与功能的关系, 推动蛋白质工程的发展, 阐明细胞中各种蛋白的作用和蛋白在细胞中的装配极为有用。

定位诱变 (Site-directed mutagenesis) 一般包括两个方面, 区域随机诱变 (Localized random mutagenesis, sitespecific mutagenesis) 和寡核苷酸定位诱变 (Oligonucleotide-directed mutagenesis), 前者是在带靶位点的一段 DNA 顺序上引入随机碱基取代; 后者可按预先设计的方案, 改变 DNA 顺序上某个或某些特定碱基。这两种方法各有所长, 选择那一种决定于试验要解决的问题。

### (一) 区域随机诱变

区域随机诱变是在体外随机诱变的基础上发展而来的。这是由于体外随机诱变虽然方法简单, 但突变效率不高, 限于少数“热点”碱基。提高诱变率常引起 DNA

分子上几处同时突变, 使分析结果复杂化。区域随机诱变能克服上述缺点, 只使突变集中在所要研究的 DNA 小片段上, 片段的大小一般在几十个到几百个碱基之间。最简单的方法是把靶 DNA 片段用合适的限制酶切出, 在体外与诱变剂作用, 再把已被诱变的片段与原来 DNA 分子的大片段重组。亚硝酸、羟胺及甲氧基胺等已成功地用于区域随机诱变<sup>[1-3]</sup>。采用此法的关键问题是靶 DNA 区域两端要有合适的限制酶切口, 不仅需考虑易于分离靶片段, 且还能有效地与原始 DNA 大片段重组。在实验设计时, 常因靶位置两端没有合适的限制酶切口而限制了此法的应用。

第二种方法是使 DNA 分子的靶区域形成局部单链区, 然后用只与单链 DNA 上碱基起反应的化学诱变剂处理, 这样诱变的区域只在 DNA 分子的单链区。常用的单链 DNA 诱变剂是亚硫酸钠<sup>[4]</sup>, 它可使 DNA 单链区的 C 脱氨基生成 U, 当 DNA 复制时, U 像 T 一样与 A 配对, 产生 GC→AT 转换。控制亚硫酸钠的浓度与反应时间, 可使诱变效率增高。但由于只能诱变 C→T, 故多样性不够。为增加突变的多样性, 可在 DNA 多聚酶 I 修复局部单链区的互补链时加入核苷类似物, 如用三磷

本文于1986年7月17日收到,

酸 N-4-羟基脱氧胞苷代替dTTP修复单链缺口。由于 N-4-羟基脱氧胞苷酮式与烯醇式的量相等,故与 A或C 配对的机率相同,经DNA复制,可得 AT→GC 转换<sup>[6]</sup>;也可在修复单链缺口时加 4 种三磷酸脱氧核苷 (dNTP) 中的三种,当 DNA 多聚酶 I 修补单链缺口到缺乏配对的 dNTP 处,因没有底物可中止链的伸长,或选用其他三种 dNTP 之一误作为底物继续修补,导致碱基错配。但由于 DNA 多聚酶 I 也有 3'—5' 外切酶活性,能校正错配的碱基对而减少了突变。为提高诱变效率,可用三磷酸 $\alpha$ -硫代磷酸核苷代替相对应的脱氧核苷, DNA 多聚酶不能校正  $\alpha$ -硫代磷酸核苷的错误参入,从而增加了诱变效率<sup>[8]</sup>。用核苷类似物诱变及三磷酸核苷的错误参入虽然诱变率和多样性均高,但也有严重的缺点,常会产生多重、成簇突变。

局部单链区可用以下几种方法获得:

(1) 用 D-环(displacement loop) 法制造单链缺口<sup>[7]</sup>。把一段含有靶碱基的同源单链DNA片段与ccc DNA混合,在 Rec A蛋白质和ATP存在时,单链DNA片段可以在负超螺旋的压力下和ccc DNA中互补链配对,置换出与它顺序相同的单链,形成D-环,然后用 S1 酶在置换出的单链上制造一个缺口,该缺口使 DNA 分子的负超螺旋压力消失,单链 DNA 再被原有的 DNA 链置换出来,这样在 DNA 上就产生了一个离靶碱基极近的短单链区(图 1)。

(2) 在靶区域内或附近找一个合适的平端限制酶酶切后,再用DNA多聚酶 I 沿着 3'→5' 的方向在 DNA 链上切去一段单链,而形成局部单链区。(3) 在溴化乙锭存在时,限制酶只在ccc DNA上产生单链切口,然后再用外切酶切出短的单链区。

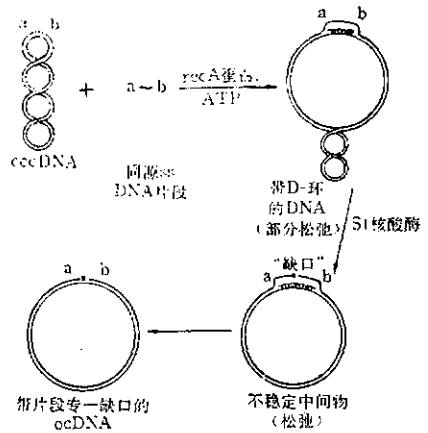


图 1 D-环法制造单链缺口

以上两种方法只限于带突变的分子有遗传标记选择,但选得的突变株也难以用表型区分突变位点是否相同。

第三种方法是 Myer 等发展的饱和诱变法 (Saturation mutagenesis)<sup>[8]</sup>,能在靶区域获得多样高效及不同位点的单一碱基突变。把靶DNA片段插入带有M13复制起始点及GC堆(GC-clamp,阻止DNA两条链完全解链)的两个方位相反的pBR 322 衍生质粒上,用化学诱变剂亚硝酸,甲酸及胍等诱变单链 DNA,用诱变过的单链 DNA 作模板合成互补链。显然在模板链中受损的碱基会使互补链的相对位置形成不同于野生型的碱基。诱变过的靶 DNA 片段可用物理方法在高温下进行尿素梯度胶电泳与野生型片段分离。再经次级克隆,分得单个菌落,比较每个菌落中含靶DNA及 GC堆的DNA在变性梯度胶上的电泳位置,就可区分突变位置是否相同,并可直接测定变种的 DNA 顺序。此法虽然步骤繁多,但能满足多样、高效及单一碱基突变的要求,在缺乏表型选择时进行遗传分析及研究蛋白质结构与功能的关系很有吸引力,适用于 30—600bp 的诱变。

用缺失或插入方法引入区域随机诱变

可参阅有关文献<sup>[9-11]</sup>, 本文不一一介绍。

## (二) 人工合成寡核苷酸的定位诱变

区域随机诱变引起的突变多数只能在某限制酶切点附近区域发生, 虽突变范围广、效率高, 但突变的特异性不高, 若希望插入、缺失或改变靶DNA片段的某个或某些特定碱基, 需采用寡核苷酸定位诱变。通常先合成一段含靶碱基及其两边序列的寡核苷酸, 若靶是单一碱基取代, 缺失或插入, 一般合成15个碱基左右的核苷酸, 靶的碱基数越多, 合成的片段越长, 靶碱基位于中间。除了靶碱基换成所需碱基外, 其他碱基顺序均需和原DNA分子中相应顺序完全互补。然后将合成的寡核苷酸和含靶位置的单链DNA混合、杂交, 并以此段寡核苷酸作为引物, 在DNA多聚酶I, 连接酶和4种NTP作用下合成完整的互补链, 转化或转染大肠杆菌, 在体内复制得野生型及突变型混合物, 经合适的筛选方法, 获得变种等位基因。

寡核苷酸定位诱变需要单链DNA靶位点, 以供诱变引物杂交之用, 因此必须把基因克隆到能提供分离单链DNA的载体上。一般用两种系统, 单链DNA噬菌体及单链环状质粒。单链噬菌体,  $\phi$ X174是第一个人工合成寡核苷酸定位诱变采用的载体, 在严格控制各步操作时, 其诱变效率最高可达30%以上<sup>[12, 13]</sup>。目前, 更多地采用噬菌体M13系列, 因为M13衍生物有多种限制酶切口, 不仅有利于DNA片段插入, 且插入后易于在X-gae选择性平板上筛选, 且又可直接用于DNA顺序分析。再者, M13噬菌体基因组的复制是不对称的, 在噬菌体DNA转染宿主后, 先以2:1倾向有利于负链作为模板的复制, 也即有利于带寡核苷酸定位诱变链作模板的复制<sup>[14, 15]</sup>。另一种是采用单链环状质

粒, 把需诱变的DNA片段克隆到带噬菌体复制起始点及形成表型所需基因(M13, T4, T7之一)的质粒载体上, 在辅助噬菌体存在时细胞释放单链质粒<sup>[16, 17]</sup>。这类质粒可插入较大的DNA片段, 稳定性比M13系列强。在无辅助噬菌体时与双链质粒完全相同, 进行定位诱变或DNA顺序分析时可以单链质粒存在, 是一类很有应用价值的质粒。此外也可用溴化乙锭和限制酶共存时在双链环状DNA的一条链上引入缺口, 再用外切酶切去一条链<sup>[18]</sup>。

采用上述两种系统诱变, 有两种不利因素会使诱变率减低。(1)宿主细胞对DNA的错配有校正作用。在开始复制时, 修复系统识别有甲基化DNA链(正链, 作外模板链)的原始信息, 而校正没有甲基化链(负链, 带突变的链)的错误结构<sup>[14-15]</sup>。

(2)完全单链的环状模板在用DNA多聚酶合成互补链时常会出现错误突变(False mutations), 且DNA多聚酶不易通过有次级结构的模板区, 因此在许多DNA分子的反应混合物中具有一种不完全的互补链<sup>[19]</sup>。

为克服这些障碍, 采用无碱基错配修复功能的大肠杆菌菌株作为转染杂合双链DNA的受体菌<sup>[14, 20]</sup>, 或使诱变区成单链缺口的双链DNA分子g<sub>ep</sub> duplex DNA缩写成(gd DNA)与寡核苷酸杂交, 再用DNA多聚酶I和4个dNTP把DNA缺口处补全封口。Kramer等<sup>[20]</sup>用M13的正链DNA先与大部份的负链(不包括待诱变区)杂交成为gd DNA, 再与合成的寡核苷酸杂交, 在体外补缺封口后转染错配修复功能缺陷的大肠杆菌, 得到的变种高达70%。Efimov等<sup>[19]</sup>把切成平端的线性质粒双链DNA分子用外切酶III处理, 由3'端切入, 伸出的单链5'端包括待诱变区, 然后与合成的寡核苷酸杂交, 诱变率

为10—20%。也可用抗修复的核苷类似物,如Dixon等<sup>[21]</sup>合成的寡核苷酸在靶位上用一种抗修复的脱氧鸟嘌呤核苷类似物2'-脱氧-7-去氮杂次黄嘌呤核苷(dDI)代替G的错配,在30个蚀斑中得到8个预期的突变,增加了诱变的效率。

突变种的检测常用<sup>32</sup>P末端标记同样的人工合成寡核苷酸作为探针、进行菌落杂交,筛选含等位基因的变种。最近随着人工合成寡核苷酸诱变方法的不断改进,已建立了用正筛选方法检测变种。一种是Kunkel<sup>[22]</sup>建立的带尿嘧啶的DNA模板法,此方法的依据是dut<sup>-</sup>ung<sup>-</sup>的大肠杆菌没有dut基因产的dUTPase,导致在细胞内增加dUTP,它与dTTP竞争参入DNA,由于细胞缺乏ung基因产物尿嘧啶糖苷酶,参入的尿嘧啶不会从DNA链上除去,带U的DNA在ung<sup>-</sup>宿主中有几乎正常的模板编码能力,由这样的宿主制得的单链噬菌体DNA中含U,以它作为模板用于定位诱变,得到的杂合双链转染ung<sup>+</sup>宿主,此时含U的正链模板易受尿嘧啶糖苷酶的作用,移去U,破坏了正链模板,因此得到的基本上是带突变的寡核苷酸负链为模板的子代。另一种是Carter设计的偶联引导法(Coupled priming)<sup>[14]</sup>,根据EcoK

(5' AAC N6 GTGC 3')及EcoB (5' TGA N8 TGCT 3')系统的识别顺序,合成一段DNA小片段,带EcoK或EcoB限制位点,并能通过一个碱基取代使切口对接。把靶DNA片段克隆到带EcoK或EcoB限制位点的M13载体上,用两种合成的寡核苷酸作为引物,一种用于构建靶DNA突变,另一种用来消除负链上可选择的限制标记。若正链模板上带有EcoB限制位点,则选择引物中的碱基顺序是用EcoK代替EcoB,在体外合成负链后,转染错配修复功能缺失的*E. coli* B菌株,正链的

子代有EcoB切口,不能生存,而负链模板的子代已消除EcoB切口,故生长的是偶联引物合成的负链子代,由此法得到的变种可达70%。若得到的突变噬菌体还需在其他部位诱变,则选择引物就可用消除EcoK带EcoB的寡核苷酸,转染*E. coli* K菌株,同样能得到只带负链子代的突变噬菌体。此法可用于多次诱变循环(图2)。相似的方法是构建噬菌体必须基因的琥珀(am)突变,这种载体不能在无校正抑制(sup<sup>o</sup>)的菌株中复制,以此作为模板,在杂交的互补链上带有消除am的片段及诱变靶位点的片段,转染sup<sup>o</sup>菌株,只得到负链模板的子代,从而提高了诱变效率<sup>[14, 23]</sup>。

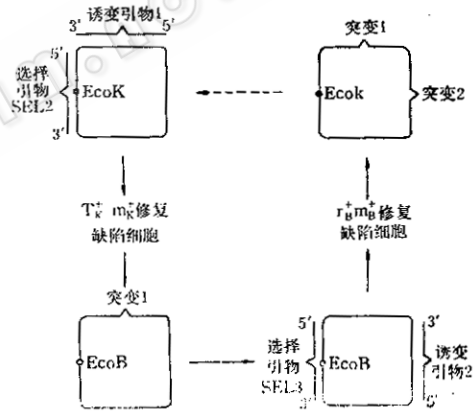


图2 “偶联引导法”寡核苷酸定位诱变和循环选择

寡核苷酸定位诱变与区域随机诱变各有其应用范围。对于已知DNA顺序的基因,其靶顺序较长或已知靶区域还未确定靶位点的DNA片段,采用区域随机诱变较为合适,可获得各种不同类型不同位点的突变;对于已确定靶位点的基因则用寡核苷酸定位诱变更为有利。

### (三) 定位诱变的应用

基因的定位诱变已应用于分子遗传学的许多领域,而且应用范围正在扩大,成功的例子已是举不胜举。应用最广泛的可

算是研究蛋白质结构与功能的关系，尤其是对已知活性中心的蛋白质。定向改变编码蛋白质活性中心氨基酸的密码子，常会引起底物特异性的改变，有时可得到增加卓越生物活性的蛋白质，为开发蛋白质工程，创造性地设计和构建人造蛋白提供了强有力的手段。现已开始用定位诱变制造人类蛋白的变种。例如，一种血浆蛋白 $\alpha_1$ -抗胰蛋白酶是弹性硬蛋白酶的抑制剂，弹性硬蛋白酶使肺丧失弹性引起肺气肿， $\alpha_1$ -抗胰蛋白酶可用于治疗肺气肿，但需要大剂量，且易氧化而降低活性。已知其活性中心是Met<sup>358</sup>，用15个碱基的寡核苷酸定位诱变使编码Met的密码子ATG变为GTG，即Met<sup>358</sup>→Val<sup>358</sup>，得到的变异蛋白抗氧化失活，而对弹性硬蛋白酶的抑制活性不变。用同样方法又使Met<sup>358</sup>→Arg<sup>358</sup>得到的Arg<sup>358</sup>变种改变了底物专一性其底物已从原来的弹性硬蛋白酶改为凝血酶<sup>[24-26]</sup>。这些新的抑制剂有治疗肺气肿和血栓的潜在力，可能成为有效的治疗剂。定位诱变也应用于改变工业用酶的底物专一性，如加酶洗涤剂用的枯草杆菌蛋白酶，活性中心是Gly<sup>166</sup>，它的靶位是底物在P1位的芳香或疏水氨基酸，用定位诱变先在活性位两侧取代个别碱基，引入两个限制酶切口，然后合成多种37碱基的寡核苷酸，改变Gly<sup>166</sup>的密码子，用不同氨基酸取代Gly，结果发现当Gly<sup>166</sup>→Asp<sup>166</sup>或Gly<sup>166</sup>→Glu<sup>166</sup>时，由于活性中心的负电荷，使它的底物专一性由丝氨酸蛋白酶改为靶位带正电荷氨基酸的蛋白酶<sup>[27]</sup>。此方法旨在改变酶的功能，扩大酶的用途，创造更有工业价值的酶。

定位诱变对分子遗传学中的应用也示有独到之处。它可改变和确定功能尚未完全阐明的DNA顺序区域的作用，已用于研究DNA复制、转录和翻译。如用于

确定腺病毒DNA复制起始点<sup>[28]</sup>和酵母染色体复制子的结构需要<sup>[29]</sup>；证明真核基因转录起始必须有TATA顺序，在此顺序中的碱基颠换(T→G)会明显减低转录效率<sup>[30]</sup>。寡核苷酸定位诱变对研究难以用表型区分的转录调控也是很有用的。如在大肠杆菌半乳糖操纵子上，有两个重叠的启动子（他们的转录起始点只相距5bp）和两个操纵基因，后者能以相等的亲和力与大肠杆菌半乳糖阻遏物结合，为了解两个操纵基因如何调控启动子转录，构建了一系列单碱基取代的突变种、失活两个操纵基因中的任何一个或启动子中的任何一个，由此得出上游操纵基因单独控制启动子P1，而启动子P2需有两个操纵基因同时控制。又如用随机定位诱变在*C.elegans*的tRNA<sup>P10</sup>启动子的第二个组分周围构建一系列单、双碱基取代的变种，分析转录起始中单个碱基的重要性，指出在编码密码子发夹结构(Stem-loop)上的突变不改变转录子转录和加工，但在编码T<sub>ψ</sub>CG发夹结构区的大部份变种都减弱表达<sup>[32,33]</sup>。最近Kozak<sup>[34]</sup>研究前胰岛素原基因翻译起始密码子AUG两侧顺序对翻译的调节作用时，采用寡核苷酸定位诱变。在AUG附近创造各种点突变，结果指出真核基因翻译起始最适顺序是ACCAUGG，尤其是AUG上游第三个核苷酸(-3位)起重要作用。Morle'等<sup>[35]</sup>分析地中海贫血是由于 $\alpha$ -球蛋白起始位置上的顺序由正常的CACCAUG突变为CCCCAUG造成的，说明-3位上A的重要性。

上面只是许多定位诱变方法应用中的几个例子。随着诱变方法的不断改进和寡核苷酸合成方法的简化，定位诱变应用范围会越来越广，将成为分子生物学基础研究和生物工程特别是蛋白质工程研究必不

可少的手段。

### 参 考 文 献

- [1] Solnick, D., *Nature* (London), 291:508, 1981.
- [2] Busby, S. et al., *J. Mol. Biol.*, 154:197, 1982.
- [3] Kadonaga, J.T. and Knowles, L.R., *Nucleic Acids Res.*, 13:1733, 1985.
- [4] Shortle, D. and Botstein, D., *Methods Enzymol.*, Vol.100, editor by Ray Wu and Lawrence Grossman, p.457, 1983.
- [5] Muller, W. et al., *J. Mol. Biol.*, 124:343, 1978.
- [6] Shortle, D. et al., *Proc. Natl. Acad. Sci. USA*, 79:1588, 1982.
- [7] Shortle, D. et al., *Proc. Natl. Acad. Sci. USA*, 77:5375, 1980.
- [8] Myers, R.M. et al., *Science*, 229:242, 1985.
- [9] Haltiner, M. et al., *Nucleic Acids Res.*, 13:1015, 1985.
- [10] Shortle, D. et al., *Ann. Rev. Genet.*, 15:265, 1981.
- [11] Botstein, D. and Shortle, D., *Science*, 229:1193, 1985.
- [12] Hutchison, C.A. et al., *J. Biol. Chem.*, 253:6551, 1978.
- [13] Gillam, S. and Smith, M., *Gene*, 8:81, 1979.
- [14] Carter, P. et al., *Nucleic Acids Res.*, 13:4431, 1985.
- [15] Fritz, H.-J., *DNA Cloning*, vol. 1., editor by D.M. Glover, p.151, 1985.
- [16] Dente, L. et al., *DNA Cloning*, vol. 1., editor by D.M. Glover, p.101, 1985.
- [17] Schmidt, B.J. et al., *Gene*, 41:331, 1986.
- [18] Wallace, R.B. et al., *Science*, 209:1396, 1980.
- [19] Efimov, V.A. et al., *FEBS Letters*, 181:407, 1985.
- [20] Kramer, W. et al., *Nucleic Acids Res.*, 12:944, 1984.
- [21] Dixon, L. et al., *Gene*, 41:225, 1986.
- [22] Kunkel, T.A., *Proc. Natl. Acad. Sci. USA*, 82:488, 1985.
- [23] Bauer, C.E. et al., *Gene*, 37:73, 1985.
- [24] Carrell, R., *Nature*, 312:12, 1984.
- [25] Rosenberg, S., *Nature*, 312:77, 1984.
- [26] Courtney, M. et al., *Nature*, 313:149, 1985.
- [27] Estell, D.A. et al., *Biotechnology Report '84 USA*, 181, 1984.
- [28] Tamanoi, F., *Proc. Natl. Acad. Sci. USA*, 80:6446, 1983.
- [29] Kearsey, S., *Cell*, 37:299, 1984.
- [30] Wasylyk, B. et al., *Proc. Natl. Acad. Sci. USA*, 77:7024, 1980.
- [31] Kuhnke, G. et al., *EMBO J.*, 5:167, 1986.
- [32] Traboni, C. et al., *EMBO J.*, 1:415, 1982.
- [33] Traboni, C. et al., *Cell*, 36:179, 1984.
- [34] Kozak, M., *Cell*, 44:283, 1986.
- [35] Morl'e et al., *EMBO J.*, 4:1245, 1985.