

# 关于用定量 PCR 方法测定基因的 转录水平的几个问题

陆长德\*

(中国科学院生物化学与细胞生物学研究所, 上海 200031)

定量 PCR 的方法已经广泛应用于生命科学各学科的研究。作者发现有些用 SYBR Green I 的定量 PCR 的工作不甚规范, 影响了实验结果的准确性, 这里结合实验从原理说明 SYBR Green I 的定量 PCR 的应用方法。此外, 生物学研究中大量采用了相对测定, 然后对结果归一化的策略。这里我们分析相对测定策略存在的不足, 并介绍每拷贝基因的转录水平的概念和双外参跟踪标定法。

## 一 用 SYBR Green I 的定量 PCR 方法的探讨

### 1 SYBR Green I 定量 PCR 的工作原理

染料 SYBR Green I 与双链 DNA 有很高的亲和性, 它在比 EtBr 低得多的浓度就可以与 DNA 结合而发出绿色荧光, 因而比 EtBr 灵敏得多。由于它的本底噪音很低、灵敏度很高的特点, 非常适合于定量 PCR 仪器的应用。在反应开始, 模板 DNA 量很少, 与这些 DNA 结合的染料发出的荧光低于本底噪音, 随着 PCR 产物的指数扩增, 与 PCR 产物结合的染料发出的荧光也指数式地增加, 直到饱和。与 PCR 产物结合的染料发出的荧光超过本底噪音时的 PCR 圈数与模板 DNA 的起始量相关, 这就是 SYBR Green I 定量 PCR 的工作原理。由于各个管子的本底噪音略有不同, 实验中取同一个阈值 T 来确定各个 PCR 反应的圈数, 为 Ct。有的仪器已设置好阈值, 直接给出 Ct。从定量 PCR 的模板曲线可见: 模板多, Ct 小; 模板少, Ct 大。随着圈数增加, 酶会逐步失活, 扩增效率会逐步降低, 最后达到饱和的荧光值也不相同(图 1)。

需要提到一个问题, 即 PCR 扩增产物的长度。从仪器设计考虑, 定量 PCR 扩增产物设计得比较短, 每一周期中延伸一步的时间就比较短, 使总的测定时间不长。更重要的是短的 PCR 扩增产物可以确保每个周期中每一条链都得到完全延伸, 一般建议设计 PCR 扩增产物为 100~300 bp。然而, 有人设计的

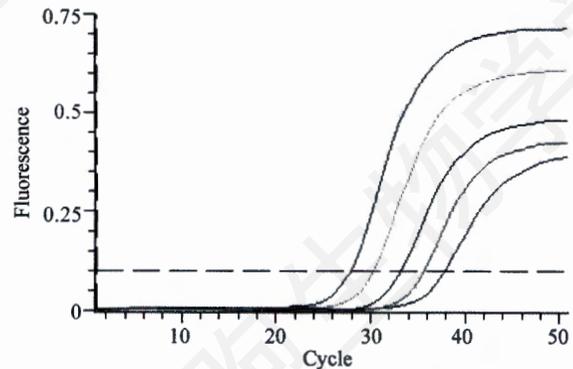


Fig.1 定量 PCR 的模板曲线

5 个反应的模板量由上而下依次为 60 ng/ $\mu$ l, 12 ng/ $\mu$ l, 2.4 ng/ $\mu$ l, 0.48 ng/ $\mu$ l 和 0.096 ng/ $\mu$ l。

PCR 扩增产物超过这个长度, 达 600~700 bp。PCR 扩增产物的长度超过仪器一般要求, 延伸时间应当加长, 否则有的片段可能没有得到完全延伸, 总的测定时间就应当变长, 酶失活会变得严重, dNTP 也会消耗过多。此外, 通常一个实验会同时测定多个基因, 人们有时设计不同 PCR 扩增产物的长度会差得很多。对于不同长度的 PCR 扩增产物除了很难采用不同的延伸时间外, 我们还应当想到, 一条较长的

\* 通讯作者。Tel: 021-54921234, E-mail: cdlu@sibs.ac.cn

PCR 扩增产物可以结合的染料比较多, 所产生的荧光信号就较大, 显然会在比较少的扩增周期就超过本底噪音, 因而使 Ct 值变小。这种情况产生的误差只有每个基因都与标准曲线比才会消除。因此, 设计定量 PCR 引物要注意它不同于一般 PCR 引物, 要设计在 100~300 bp 之间, 同时做的不同基因的 PCR 扩增产物的长度之差应当尽量小, 建议不超过 20 bp。

## 2 熔解曲线、熔点和荧光读取温度

仪器会给出一个 PCR 产物的熔解曲线, 它可以给我们很多信息(图 2)。

从图中可以看到, PCR 反应产物在其熔点处被解开, 荧光读数迅速降低, 在导数曲线中则显示出一个峰, 而杂峰往往与主峰熔点不同。从 PCR 产物的熔解曲线图也可以看到对于每一个定量 PCR 反应应当设置不同的荧光读取温度。从熔解曲线看, 一个峰的半高宽度一般在 5~6°C, 将荧光读取温度设置在熔点前 3~4°C, 既可以避免将杂峰的荧光读进来, 又可以使读到的数值最能反映主要 PCR 反应产物的荧光值的增加。

在正式测定前先要将所有的引物做一次预实验, 看各个 PCR 产物的熔点是, 专一性是否好, 并确定每个 PCR 反应的荧光读取温度。也可以同时看一下模板大约稀释多少比较好。DNA 双链解开

的温度取决于它的 G+C 含量(参考值: PCR 产物的 G+C 含量为 45% 时熔点约在 80°C 左右, 50% 时熔点约在 83°C 左右)。设计引物时, 要注意 PCR 反应产物的 G+C 含量不要太高或太低。

每次实验结束, 首先要检查 PCR 反应产物的熔解曲线, 熔点与标准不对或者有杂峰且对结果有影响的都不能要。

## 3 影响定量 PCR 反应的因素

### (1) 引物与退火温度的影响

PCR 反应的第一步是引物与模板的配对, 形成正确的模板-引物是合成的开始。影响这一步的主要是退火温度, 至于延伸反应都是 72°C, 变性都是 95°C, 没有什么差别。有一个实验很能说明问题。在同一个质粒上同时有 EGFP 和 Luciferase 两个报告基因, 设计了一对测 EGFP 的引物, 两对测 Luciferase 的引物(实际上只有其中一条引物的 3' 短 2 个碱基, 扩增产物长度相同)。引物及产物信息见表 1。

测定的退火温度设为 54°C, 上述 5 条引物中 4 条的 T<sub>m</sub> 比较接近, 在 57.30°C 至 57.80°C, 只有 Luc-F2 相差较大, T<sub>m</sub> 值为 55.02°C, 在 54°C 时 T<sub>m</sub> 在 57.30°C 至 57.80°C 的 4 条引物与模板配对比较稳定, 而 Luc-F2 形成的模板-引物稳定性差一些, 扩增效率受影响, 在同时进行的测定中, 相同模板起始量用 Luc-F2/Luc-R 测得的 Ct 就比用 Luc-F1/Luc-R 测得的 Ct 大。

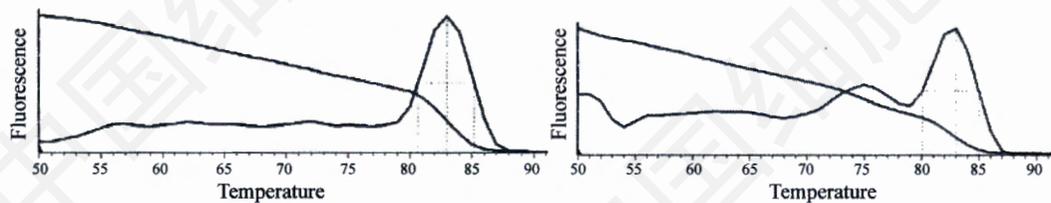


Fig.2 定量 PCR 产物的熔解曲线

左图显示引物专一性好的 PCR 反应产物的熔解曲线; 右图是引物专一性差的 PCR 反应产物的熔解曲线。逐渐降低的线是荧光读数随着温度升高而降低的曲线; 有峰的线是荧光读数对温度的导数曲线(有的仪器这两条线分别作在两个图中)。

Table 1 定量 PCR 的引物及产物

引物	序列	引物 T <sub>m</sub>	产物(bp) G+C	产物 T <sub>m</sub>
EGFP-F	5'-GAC GGC AAC TAC AAG ACC-3'	57.30°C	159 bp 57.8%	87°C
EGFP-R	5'-GTC GGC CAT GAT ATA GAC G-3'	57.56°C		
Luc-F1	5'-AAA CGC TGG GCG TTA ATC AG-3'	57.80°C	145bp 46.9%	82°C
Luc-F2	5'-AAA CGC TGG GCG TTA ATC-3'	55.02°C	145bp 46.9%	82°C
Luc-R	5'-TCG TCC CAG TAA GCT ATG TC-3'	57.80°C		

引物 T<sub>m</sub> 值是由合成引物的上海生工给出的数据。

Table 2 定量 PCR 的数据

引物	拷贝数 $N_0$	$10^2$	$10^3$	$10^4$	$10^5$	$10^6$	$10^7$
EGFP-F/ EGFP-R	Ct EGFP-1	26.91	23.76	21.15	17.68	14.30	10.52
EGFP-F/ EGFP-R	Ct EGFP-2	28.91	26.01	22.71	19.41	15.91	12.11
Luc-F1/ Luc-R	Ct Luc-1	29.34	25.19	22.38	18.65	15.07	10.70
Luc-F2/ Luc-R	Ct Luc-2	36.19	31.50	27.66	22.85	18.74	13.13

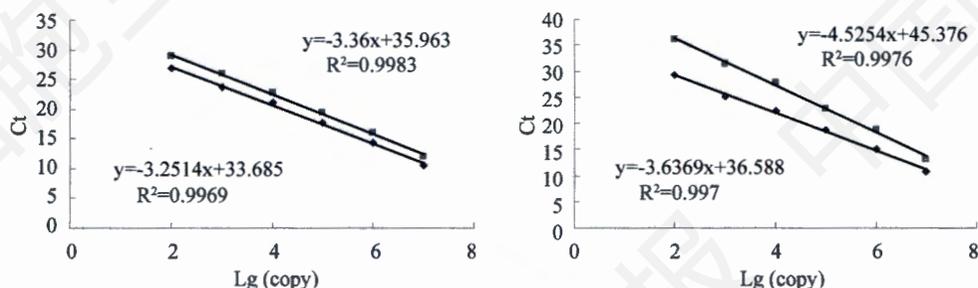


Fig.3 引物对定量 PCR 结果的影响

左图: 同样引物对 EGFP 基因的二次测定。右图: 同时用不同引物对 Luc 基因的测定。

测定标准曲线结果见表 2 和图 3。

从图得到标准曲线公式:

$$\text{EGFP-1: } y = -3.2514x + 33.685, R^2 = 0.9969$$

$$\text{EGFP-2: } y = -3.36x + 35.963, R^2 = 0.9983$$

$$\text{Luc-1: } y = -3.6369x + 36.588, R^2 = 0.9970$$

$$\text{Luc-2: } y = -4.5254x + 45.376, R^2 = 0.9976$$

图 3 的左图二条线之差是重复同一实验的实验误差, 可以通过多次实验来消除; 右图的二条线是因为引物的不同而产生的误差, 不能通过多次实验来消除。拿同时测定的 EGFP-1 与 Luc-1 比, 它们的引物  $T_m$  还是相当接近的, 被测样品是同样的 10 倍稀释系

列样品, 只是测同一个质粒上的 2 个不同基因, 扩增产物长度差 14 bp, 从直线方程的斜率  $a$  看, 它们之间的差别还是大于 EGFP 的二次测定间的误差。

虽然测定不同基因的引物不一样, 或测定同一基因用不同引物得到的标准曲线不同, 用各自的标准曲线计算得到的未知样品中起始拷贝数大致还是差不多的。因为在相同的条件下, 由引物不同引起的影响是相同的, 在测定和数据处理中被消除了。

从原理分析和实验结果说明, 用相同引物测定同一个基因在数据处理时, 可以消除由 PCR 扩增而产生的误差, 而用不同引物测定时不能消除由 PCR 扩增效率不同而产生的误差。

由于引物对于 PCR 扩增的影响很大, 在设计引物时使引物的  $T_m$  值尽量相同非常重要。最佳的退火温度需要靠实验确定, 一般比引物  $T_m$  低  $4^\circ\text{C}$  左右引物能够与模板形成好的配对。如果退火温度太低, 容易发生非专一的配对而影响扩增的专一性, 出现杂峰。

## (2) 其他因素

在定量 PCR 反应中, 试剂商已配好了主要反应液, 操作者另外加入的就是引物和样品(含模板 DNA), 引物对每个管子都相同, 一般将合成的引物加  $\text{ddH}_2\text{O}$  溶解, 不会有什么问题, 因此, 最需要分析的是样品中可能的影响因素。

样品中的 DNA: 反应中总 DNA 量太高会竞争与酶结合而影响 PCR 反应, 见图 4, 用相同量的质粒外

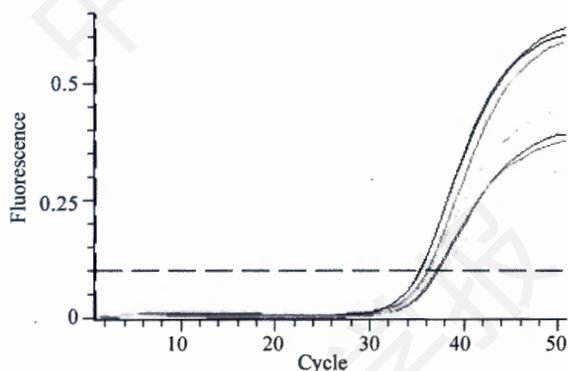


Fig.4 总 DNA 对 PCR 扩增反应的影响

各反应中质粒质量相同, 外加组织 DNA 在反应体系中的浓度由上而下分别为  $0 \text{ ng}/\mu\text{l}$ ,  $0.0192 \text{ ng}/\mu\text{l}$ ,  $0.096 \text{ ng}/\mu\text{l}$ ,  $0.48 \text{ ng}/\mu\text{l}$ ,  $2.4 \text{ ng}/\mu\text{l}$ ,  $12 \text{ ng}/\mu\text{l}$  和  $60 \text{ ng}/\mu\text{l}$ 。

加不同量的组织DNA, 总DNA量在一定范围内(按仪器说明书建议)影响不大, 太高则对被测基因的Ct稍有影响。测定的结果见图4。

样品中的其他成分: PCR扩增反应是一个酶催化的反应, 反应体系的 $Mg^{2+}$ 浓度、pH值、盐浓度对酶反应是重要的, 它们可以受样品的因素影响, 例如, 如果样品中有EDTA可以络合 $Mg^{2+}$ 离子, 而 $Mg^{2+}$ 离子是PCR扩增反应所必需的; 因此溶解样品不要使用含高盐、高EDTA、pH值太高或太低的浓的缓冲液, 在做定量PCR测定时模板不宜太浓, 一般需要适当稀释, 可以用 $ddH_2O$ 稀释模板, 这样经稀释后的模板样品一般就不会对反应有多少影响了。

#### 4 数据处理

对定量PCR测定结果有多种数据处理方法。对于生物学研究来讲, 直观而简明的方法更有利于说明问题和进行交流。

PCR扩增反应中 $N_n = N_0 \times (1+E)^n$ , ( $E \leq 1$ ),  $N_0$ 是被扩增基因的起始拷贝数,  $N_n$ 是扩增到第n圈体系中扩增产物的拷贝数, E是扩增效率, 当 $E=1$ , 扩增倍数为 $2^n$ 。

取以10为底的对数得到 $\lg N_n = \lg(1+E) \times n + \lg N_0$ , 为直线方程,  $\lg N$ 即反应体系中拷贝数的对数, 即 $\lg(\text{copy})$ , n即扩增圈数, 荧光超过本底噪音时的PCR圈数为Ct。

以 $\lg N_0$ 对Ct作图, 可得直线。也可写成 $Ct = a \times \lg N_0 + b$ ,  $N_0$ : 反应中初始拷贝数。

用作图法是个直观而简明的方法。将质粒浓度测准确, 10倍稀释成系列, 做标准曲线,  $\lg N_0$ 与Ct成一条直线, 处理起来非常方便。这种表示法给出的结果是拷贝数, 也非常明了(表2, 图3)。

把用被测样品测得的Ct代入标准曲线公式即可得到样品拷贝数。我们把根据标准曲线计算得到的未知样品的拷贝数的测定方法称为绝对测定, 一般做三个测定取平均以提高测定结果的准确性, 同时可以计算测量误差。

通常定量PCR测定的目标(不论是DNA或RNA)有两种: 1) 测定样品中某基因的拷贝数, 例如血液或细胞培养液中某种病毒的量。2) 研究某个生物过程中某个基因量的变化, 例如某基因在某些条件下表达的变化。第一种即绝对测定, 与标准样品一起测, 从标准曲线计算即可得到。第二种测定是不同情况下的比较, 由于考虑到样品处理可能带来的误差, 往往采用一个或几个参考基因进行归一化, 也可称为相对

测定, 常用相对于某个基因的量来表示。关于实验误差的来源、归一化的策略下面专门讨论。就数据处理和结果的表示而言, 有人简单地比较两个不同基因测得的Ct, 用 $\Delta Ct$ 数据来表示生物事件的变化, 而不与各个基因的标准曲线比较, 这样做往往是由于没有每一个基因的标准质粒。然而, 从上面的影响因素分析可以看到, 对不同基因, 由于引物不同, PCR扩增产物长度不同, 扩增标准曲线会有很大差别, 因此, 用 $\Delta Ct$ 等数据来表示生物事件的变化不是一种精确的方法, 当然变化趋势还是可以肯定的。再说直接用 $\Delta Ct$ 表示结果也不如用拷贝数表示那么明了。

两个样品之间, 经过数学转换可以得到被测基因的两个Ct之差 $\Delta Ct$ , 以及参考基因的两个Ct之差 $\Delta Ct$ , 用两个 $\Delta Ct$ 之差 $\Delta \Delta Ct$ 表示, 引物不同造成的误差是消除了, 但这个结果看起来也只显示了一个趋势。这里被测基因都与参考基因比, 问题还在于参考基因是否不变, 这将在下面再讨论。

数据处理中也有用 $2^{-\Delta \Delta Ct}$ 表示被测基因都与参考基因比的。虽然 $N_n = N_0 \times (1+E)^n$ , 当 $E=1$ 时扩增倍数才等于 $2^n$ 。实际上E总是小于1的, 提出这个处理方法的作者在文中指出只有当各个条件很好地优化后E才接近1, 因此, 用 $2^{-\Delta \Delta Ct}$ 表示也是近似的。

## 二 有关基因的转录水平定量测定策略的探讨

### 1 前言

1970年沃森等确立了基因表达的中心法则: DNA  $\rightarrow$  mRNA  $\rightarrow$  Protein 以来, 一个基因的转录水平一直是科学家研究的焦点问题之一。人们根据当时的技术条件不断地提出新的方法来测定基因的转录水平。1977年出现的Northern blot方法是基于核酸的转移、杂交技术和同位素标记技术。80年代逆转录酶的发现和PCR技术的发明结合产生了RT-PCR技术, 在1986年报道了它在基因转录水平的测定中的应用。高温DNA Polymerase的发现促使PCR技术走向自动化, PCR技术成为几乎所有分子生物学实验室都使用的方法, 定量PCR的仪器很快被创造出来也是必然的了, 1996年开始有Real-Time qPCR的应用报道。90年代初基因组计划的实施还促使cDNA array和chip开始用于基因转录水平的高通量的测定。Northern blot方法和RT-PCR方法得到的只能是半定量的结果, 而Real-Time qPCR则可以得到定量的结果而且灵敏度极高, 可以说对基因转录水平的

定量测定是个飞跃。有了定量测定的方法又如何来确定一个基因的转录水平呢？还是一个需要进行探讨的问题。

## 2 实验过程的误差分析和测定结果的归一化策略

正如前面提到的,在生物学研究中,从 Northern blot到定量PCR,乃至测定酶活性的报告基因方法,考虑到样品处理的复杂性和处理过程可能带来的误差,往往采用相对测定的策略对实验数据进行归一化处理。在研究基因的转录水平时,整个实验中样品需要经过组织或细胞破碎、匀浆, RNA 抽提, 逆转录, PCR 测定等步骤,这中间有离心、样品转移、酒精沉淀与洗涤等实验过程,也有酶反应。各个步骤的各个过程(或反应)的回收率和效率可能不同,都会带入一定的误差。尤其是,由于我们处理的对象是 RNA,生物体内丰富的核酸酶使得它不稳定,因此更容易产生误差。在这些过程中如果有内参或外参分子一起经历,误差就可以通过归一化处理消除。关于定量 PCR 测定中的影响因素前面已经讨论了。应用内参或外参加上标准曲线也可以消除 PCR 测定中的误差。

对于半定量结果或者重点在看看趋势的实验,应用相对测定的策略可以达到实验的要求。然而对于定量测定来说,目前的策略的本身还存在令人不尽满意之处。归纳起来,目前的归一化策略可以分为两种:一种是以 RNA 或蛋白质的总量为基数进行比较;例如用相同量的 RNA 或蛋白质走电泳,用相同量 RNA 做逆转录等等。另一种以某一个或几个内源的分子做参考,因为参考分子与研究对象经历相同的实验处理,虽然有损失它们之间仍然维持相同的比例关系,有时这两种考虑同时用到。这些策略确实很实用,也给研究带来很多方便,为大家广泛使用。然而,仔细推敲起来还是存在令人不尽满意之处的。

生物体中每一个细胞的 RNA 总量(主要是 rRNA)或 mRNA 总量与一个细胞的蛋白质合成的水平有关,不同的细胞可以相差很大。如果以细胞内源分子做参考,常用的有组成细胞骨架的 actin、细胞能量代谢中的甘油醛-3-磷酸脱氢酶(GAPDH)、蛋白质合成机器的 rRNA 和其他管家基因等等,实际上在不同组织的细胞中,它们也可以是相差很大的。对于相同细胞、相同组织这些内参可以认为差别不是很大;但是,很多结果讨论的对象是相差甚大的不同细胞,例如基因表达的组织专一性等问题,或癌细胞与普通细胞比,这显然就不够精确了。至于用于相同细胞的

不同状态,严格地讲也是不够精确的。例如研究肿瘤细胞被某个药物抑制这样一个问题,药物处理后细胞生长变慢(这是常常看到的现象),这说明 DNA 复制、蛋白质合成都变慢了,其细胞骨架的合成,能量代谢也不会保持不变。因此,用一个本身在变化的东西做参考,势必影响结论的准确性。于是人们有时同时用几个内参,试图找到稳定不变的内参,但是很难找到这样的内参。人们还用外参来归一化实验数据,称为 spike-in technique,中文是跟踪标定的意思,在样品抽提时加入一定量的人工制备的 RNA,与样品经历相同的处理,就可以归一化实验数据了。外参不受细胞变化而变化,困难在于如何确定对各个样品应该加多少量?不少做法是对相同总 RNA 量加入相同量的外参 RNA(有些基因芯片的分析中也这样做),但仍然没有解决不同细胞总 RNA 量不同的问题。

## 3 每拷贝基因的转录水平的概念和双外参跟踪标定法

为了解决在基因转录水平定量分析中遇到的问题,我们提出了“每拷贝基因的转录水平”这么一个概念,数学上等于一个基因的 mRNA 的数量与 DNA 拷贝数之比。这实际上是取被测基因的 DNA 来作为参照物对实验数据进行归一化。从它的名称和数学式可以看到其优点在于它反映了这个基因的转录水平。对于多拷贝基因或病毒等基因拷贝数与细胞变化不一致的病原体,细胞中它们的 RNA 量的多少应当从两方面来讨论,一部分是看这个基因本身转录强不强,即每拷贝基因的转录水平;另一部分是基因拷贝数增加对表达的贡献。这正如生物体内表达特别高的基因不仅靠启动子活性强,更要靠基因拷贝数高这个事实,也是利用病毒载体可以得到基因的高表达的原因。每拷贝基因的转录水平对于讨论基因转录的组织专一性和启动子的活性特别合适。

提出一个概念必需有相应的方法来测定。细胞内的 RNA 和 DNA 并不经历相同的处理,如何进行归一化?参考已有的方法,我们提出“双外参跟踪标定法(dual-spike-in technique)”的策略。

策略:在样品抽提时,加入按比例混合的 RNA 外参和 DNA 外参。用 RNA 外参跟踪标定样品的 RNA,用 DNA 外参跟踪标定样品的 DNA,一起进行各种处理(酚抽提、酒精沉淀、RNA 逆转录、PCR 扩增或芯片实验中 DNA 的引物延伸等),以消除实验步骤带来的误差;最后利用加入的混合液中 RNA 外参和

DNA 外参的比例来建立样品中 mRNA 和 DNA 之间的联系。任何在被测样品中不存在的序列都可以选做外参。

计算公式推导:

$$Ct = a \times \lg N_0 + b \dots\dots\dots (1)$$

Ct: 阈值圈数,  $N_0$ : 反应中初始拷贝数, 对每一个扩增片段有一个斜率  $a$  和截距  $b$ 。不同样品:  $Ct_1 = a_1 \times \lg N_{01} + b_1$ ,  $Ct_2 = a_2 \times \lg N_{02} + b_2$

相同基因, 不论 cDNA 或 DNA 模板,  $a_1 = a_2 = a$ ,  $b_1 = b_2 = b$ , 二次测定的公式相减得到

$$\lg(N_{01}/N_{02}) = (Ct_1 - Ct_2) \div a, N_{01}/N_{02} = 10^{(Ct_1 - Ct_2) \div a} \dots\dots\dots (2)$$

这里可见, 计算中只需要直线公式的斜率“ $a$ ”即可。分别测出 mRNA 和 DNA 的  $Ct_1$  和  $Ct_2$ , 用这个公式可以计算出靶(mRNA/DNA)<sub>测定</sub> 和外参(RNA/DNA)<sub>测定</sub>。

$$\text{每拷贝基因的转录水平(mRNA/DNA)} = \frac{\text{靶(mRNA/DNA)}_{\text{测定}}}{\text{外参(RNA/DNA)}_{\text{测定}}} \times \text{外参(RNA/DNA)}_{\text{混合液}} \dots\dots\dots (3)$$

这个策略中实际有二次归一化处理。第一次是样品 mRNA, DNA 分别用外参归一化, 得到(靶 mRNA/外参 RNA)<sub>测定</sub> 和(靶 DNA/外参 DNA)<sub>测定</sub>。再经数学运算转化得到的是两个 mRNA 和 DNA 的比值相除的商: (靶 RNA/外参 RNA)<sub>测定</sub>  $\div$  (靶 DNA/外参 DNA)<sub>测定</sub> = 靶(mRNA/DNA)<sub>测定</sub>  $\div$  外参(RNA/DNA)<sub>测定</sub>。首先, 用外参去除了处理过程造成的样品间的差别。其次, 把不同基因的数据之比转换成相同基因的数据之比; 去除不同引物造成的 PCR 扩增效率的差别。可以从 Ct 和  $a$  值, 用前面的公式(2)进行计算, 使运算更方便。

第二次是把前面得到的商再乘以外参(RNA/

DNA)<sub>混合液</sub>, 得到靶 mRNA/DNA。

这里, 外参(RNA/DNA)<sub>混合液</sub> 是加入到样品的外参混合液中外参 RNA 与 DNA 拷贝数之比。这样用加入到样品中的外参 RNA 与 DNA 拷贝数之比重新建立样品中 RNA 与 DNA 的关系, 去除了 RNA 与 DNA 不同处理带来的差别。

#### 4 应用例子

家蚕各组织功能差别非常明显, 是研究的好材料。实验步骤和结果如下:

(1) 家蚕幼虫 5 龄第 3 天解剖得到 5 种组织: 中部丝腺, 后部丝腺, 脂肪体, 马氏管和中肠。

(2) 选转座酶 IFP2 基因做外参, IFP2 质粒 DNA 用电泳割带、回收。用 T7 RNA 聚合酶体外转录得到 IFP2 RNA, CsCl 超离心纯化 RNA。紫外吸收定量 IFP2 质粒 DNA 和 IFP2 RNA, 根据质粒大小和转录的 RNA 的长度计算摩尔浓度, 按比例(例如 RNA/DNA=50~100)混合得到混合液。外参样品可以一次制备多次使用, 加乙醇到 75%, -70~ -80°C 保存, 用时减压抽掉乙醇即可。

(3) RNA 和 DNA 抽提。在研钵中将组织研碎, 向约 100 mg 组织中加入几  $\mu$ l 的 RNA-DNA 混合液(ng 级,  $10^9$ - $10^8$  RNA 分子,  $10^7$ - $10^8$  DNA 分子), 用总 RNA 抽提试剂 TRNzol-A+ 抽提, 按试剂盒说明分别收集 DNA 部分和 RNA 部分。DNA 部分用酚/氯仿处理、酒精沉淀、重新溶解。RNA 部分用 RNase-free DNaseI 除去可能残留的 DNA, 加入(dN)<sub>6</sub> 和 oligo(dT)进行逆转录, 合成第一条 cDNA 链。

(4) 为了比较, 共设计了 7 对定量 PCR 引物: 外参 IFP2 基因, A3, GAPDH, 28S rRNA 三种常用的内参基因, 丝素重链 FibH 和轻链 FibL, 丝胶 Ser-1 三种被测

Table 3 各基因片段扩增直线方程的斜率

Gene	IFP2	A3	GAPDH	28S r	FibH	FibL	Ser-1
“a”	-3.6882	-3.120	-3.102	-3.620	-3.089	-3.584	-3.499

Table 4 家蚕 5 种组织中 A3, GAPDH, 28S rRNA 的每拷贝基因的转录水平

组织	A3	GAPDH	28S rRNA
中部丝腺	112.35	10.620	5 463.7
后部丝腺	1 207.6	166.64	8 849.4
脂肪体	522.12	11 780	95 425
马氏管	400.70	142.41	2 872.5
中肠	9.9050	6.3244	2 121.2

Table 5 丝素、丝胶基因的每拷贝基因转录水平和相对测定的比较

Expression	FibH in PSG	FibL in PSG	Ser1 in MSG	Ser1/FibL
Expression level per gene copy	98 509	105 117	166 875	1.588
Relative expression to A3	95.92	102.4	1 485	14.51
Relative expression to GAPDH	591.2	630.8	16 015	25.39
Relative expression to 28S rRNA	0.03711	0.03959	0.1018	2.572

基因。它们的PCR扩增产物长度在144~152 bp,  $T_m$  在55.02~60.07°C之间。扩增产物熔点在81~85°C之间。并测定各基因在扩增条件下的直线方程的斜率“a”, (由于计算中不需要直线方程的截距“b”, 在没有标准质粒的情况, 测定样品的10倍稀释曲线也可以得到“a”值)。

(5) 进行定量PCR测定, 计算每拷贝基因的转录水平。A3, GAPDH, 28S rRNA的测定结果显示不同组织细胞中三种常用的内参基因的转录水平相差很大。在不同组织间比较转录水平时它们中的如何一个都不适合做内参。后部丝腺A3表达最高与后部丝腺在5龄长大最快的事实一致, GAPDH和rRNA在脂肪体表达最高, 与脂肪体担负的能量代谢和大量蛋白质合成功能相符。

测定了家蚕丝素重链和轻链、丝胶-1基因的每拷贝基因转录水平, 如果把它们和相对测定进行比较, 结果相差甚远。从中部丝腺表达的丝胶蛋白与在后部丝腺表达的丝素蛋白轻链的比值Ser1/FibL来看, 用每拷贝基因转录水平比较合理, 二者比值为1.588, 两个基因的转录水平很接近, 考虑到中部丝腺与后部丝腺的细胞数, 这个转录水平之比与茧壳中两种蛋白的含量相一致。如果以常用的三种内参来比, 则结果各不相同, 例如以A3为内参, Ser1/FibL比值达14.51, Ser1转录活性比FibL高出十几倍, 这是由于在5龄中部丝腺不如后部丝腺长大快(这个现象从解剖可见), 因而A3表达较低之故。

## 5 讨论

(1) 每拷贝基因转录水平给出了细胞内某基因RNA与DNA拷贝数之比, 是个直接而明了的结果。

我们看到高表达的基因这个数最高可达 $2 \times 10^5$ , 例如丝腺中的丝蛋白和rRNA。中等表达的基因这个数一般在几百到几千, 如大部分组织中的Actin。因此, 我们建议可以大致将每拷贝基因转录水平 $> 5000$ 的称为高表达的基因;  $50 \sim 5000$ 称为中等表达的基因;  $1 \sim 50$ 的称为低表达的基因; 这个数 $< 1$ 为不表达。外参RNA与DNA的比值取 $50 \sim 100$ 相当于中等表达的基因, 用于参考还是合适的。

(2) 正如前面所说, 相同细胞不同状态的定量PCR测定用内参或外参RNA加上标准曲线基本上能够得到比较准确的结果。而对于不同组织的定量PCR测定要用RNA、DNA双外参跟踪标定的策略测定每拷贝基因转录水平才可靠。

(3) 每拷贝基因转录水平与每细胞基因转录水平两个概念的主要不一样在于有些基因有多拷贝, 或在某些时候一个细胞的全部或部分基因会倍增, 因而细胞的DNA也不是不变的。每拷贝基因转录水平更重视基因的转录活性或者说启动子的上调或抑制。

(4) 在实验操作上“双外参跟踪标定法”工作量增加不太多。在抽提RNA样品的同时抽提DNA样品, 在测定cDNA的同时用同样的引物测定DNA样品, 这一点样品量的增加是可以接受的, 数据处理也是很容易掌握的。

(关于“每拷贝基因的转录水平的概念和双外参跟踪标定法”请参考文献: Zhang Y, Wei ZG, Li YY, Chen YH, Shen WD, Lu CD. Transcription level of messenger RNA per gene copy determined with dual-spike-in strategy. Analytical Biochemistry 2009; 394: 202-8)