

正交设计法在文昌鱼同源框基因 PCR 引物设计中的应用*

APPLICATION OF ORTHOGONAL DESIGN IN THE DESIGN OF PCR PRIMER FOR AMPHIOXUS HOMEBOX GENE

王 勇¹ 郎刚华² 徐永立¹ 张培军¹

(¹中国科学院海洋研究所 青岛 266071)

(²青岛海洋大学海洋生命学院 266003)

在 PCR 实验中, PCR 实验结果的好坏主要是靠引物来保证的。因此,在合成引物之前除了作人工检查外,往往需要对所设计的引物进行计算机检查。一是检查在引物内部是否有发夹结构,二是检查引物对之间是否有较多的互补碱基,三是计算引物的 T_m 值。这已经成了引物设计中必不可少的步骤。现在的计算机软件只能对序列确定的引物(即不含有简并碱基)进行分析,如果所要分析的是简并的引物,那么需要

将每一种可能的序列写出来,逐一分析。这样一来,对于简并程度较高的引物,分析起来就相当困难。本文在文昌鱼同源框基因的 PCR 扩增实验所需引物的设计中,采用了正交设计法,获得了令人满意的结果。

1 方法

这对引物中,一条引物的核苷酸序列为:



* 这就有 128 种分子。如果逐一写出序列,输入计算机,那么所耗费的时间与精力实在太大了。而用正交设计法,只需写出 8 种序列即可达到目的。在这里,把简并碱基位点当作正交设计中的因子,把每一个简并

位点上的简并碱基当作同一个因子的不同水平。因为

* 中国科学院海洋研究所调查研究报告第 3198 号;
实验海洋生物学开放研究实验室研究报告第 164 号。
收稿日期: 1997-09-03

该引物序列中共有 7 个简并的碱基位点, 每种简并的碱基均有两种情况, 因此在这种情况下, 因子数为 7, 每个因子均为 2 水平因子(罗马数字表示因子序号, 阿拉伯数字表示水平序号), 这就是一个 2 水平 7 因子实验, 所以选用 $L_8(2^7)$ 正交表(见表 1)。根据 $L_8(2^7)$ 正交表就可列出该引物的正交表(见表 2)。由表

表 1 $L_8(2^7)$ 正交表

试验号	列号						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

2 可知, 只需写出 8 种排列组合的引物序列, 就可以代

	I	II	III	IV	V	VIV II
序列 1 5'	- ATG	GAC	ATC	GCA	TCT	GAC TGA TAC- 3'
序列 2 5'	- ATG	GAC	ATC	GCA	TCA	CAC GCA TAC- 3'
序列 3 5'	- ATG	GAT	ATT	GCA	TCT	GAC GCA TAC- 3'
序列 4 5'	- ATG	GAT	ATT	GCA	TCA	CAC TGA TAC- 3'
序列 5 5'	- ATC	GAC	ATT	GCA	TCT	CAC TCA TAC- 3'
序列 6 5'	- ATC	GAC	ATT	GCA	TCA	GAC GGA TAC- 3'
序列 7 5'	- ATC	GAT	ATC	GCA	TCT	CAC GGA TAC- 3'
序列 8 5'	- ATC	GAT	ATC	GCA	TCA	GAC TCA TAC- 3'

3 讨论

由于该引物是一套简并的引物, 因此在评估该套引物的设计质量时, 应考虑这 128 个序列的集合所表现出的总体特征, 而不应看重个别的引物序列(因 PCR 反应中, 引物参与反应的方式是随机的)。如果其总体特征符合要求, 那么该套引物也就符合要求。而这个集合的总体特征是根据正交表中的排列, 通过抽取了 8 个序列进行检查后得出的, 这一做法的可靠性又是由正交设计法本身的理论基础所保证的。因此, 将正交设计法用于 PCR 引物设计在理论上是可行的。

根据正交设计法的原理, 根据表 2 所列出的 8 种排列组合的引物序列, 就可以代表所有排列组合的引物序列。对这 8 个序列进行计算机分析的结果, 基本上代表了所有的简并序列。例如, 在所检查的这 8 个序列中, T_m 值的最高值是 72 °C, 最低值是 66 °C; 而在全部

表全部排列组合的特点。

表 2 引物序列正交表*

组号	碱基位点						
	I	II	III	IV	V	VI	VII
1	G	C	C	T	G	T	G
2	G	C	C	A	C	G	C
3	G	T	T	T	G	G	C
4	G	T	T	A	C	T	G
5	C	C	T	T	C	T	C
6	C	C	T	A	G	G	G
7	C	T	C	T	C	G	G
8	C	T	C	A	G	T	C

* 在这里, 组号就是指试验号, 碱基位点就是指列号。

2 结果

在表 2 中, 每一组就代表一种经排列组合得到的引物序列。在写出引物序列时, 凡是位于简并位点的碱基, 都根据表 2 的排列写出。结果如下:

128 个序列中, T_m 值的最高值也是 72 °C, 最低值也是 66 °C, 这表明这 8 个序列的 T_m 值范围可完全代表 128 个序列的 T_m 值范围。又比如, 假设在所检查的 8 个序列中, 有两个序列含有发夹结构, 即有 25 % 的序列含有发夹结构, 这表明在 128 个序列中也有 25 % 左右的序列含有发夹结构。从以上分析也可看出, 待分析的序列数从 128 减少到 8, 在不影响分析正确性的前提下, 工作量减少到原来的 1/16。用同样方法对另一条引物也进行了分析, 并据此对这一对引物序列进行了多次改进。用改进的引物进行 PCR 扩增后, 获得了令人满意的实验结果(结果待发表)。

当然, 以上所讨论的是一种比较简单情况。我们经常会遇到水平数不等的多因子情况, 这时应根据实际情况选用最合适的正交表, 例如, 假定某引物序列中有 3 个位点为 A/G/C/T 4 个碱基简并, 其余 6 个位点均为 2 个碱基简并。在这里, 有 3 个因子为 4 水平

因子, 6 个因子为 2 水平因子, 则应选用 $L_{16}(4^3 \times 2^6)$ 正交表。正交表的选择可参考有关正交设计法的文献^[1,2]。

参考文献

- 1 北京大学数学力学系概率统计组。正交设计法。北京: 石油化学工业出版社, 1976。161~179
- 2 中国科学院数学研究所数理统计组。正交试验法。北京: 人民教育出版社, 1975。105~150