

DNA 单链二级结构预测与单链构象多态性分析的一致性研究*

何云刚 谭德勇** 钱伟 赖建华 谢咏芳

(云南大学生物技术系, 昆明 650091)

摘要 单链构象多态性 (SSCP) 分析是一种简便, 快速检测 DNA 突变的方法, 它在基因突变检测、遗传分析、进化研究等领域有着广泛的应用价值。但是这种方法的突变检出率随 DNA 序列不同而变化, 一般只能达到 70% ~ 80%。这主要是有的碱基突变对单链 DNA 的构象影响较小, 不能通过 SSCP 检测出来。将计算机对 DNA 二级结构的预测结果和实验结果作了对比, 发现二者有很高的一致性。这一结果表明计算机的 DNA 单链二级结构预测分析可用于 PCR-SSCP 分析的辅助设计, 提高 SSCP 的突变检出率。

关键词 突变检测, DNA 单链二级结构预测, 单链构象多态性分析

学科分类号 Q319.3

单链构象多态性 (SSCP) 分析技术最开始由 Orita^[1] 于 1989 年建立, 后来与 PCR 技术结合建立 PCR-SSCP 技术^[2], 是进行 DNA 的 SNP 检查和突变检测所用的筛查方法中较好的一种方法。虽然现在也有一些其他的技术可进行突变检测, 如变性梯度凝胶电泳 (DDGC) 方法, 测序, 芯片技术等, 但由于 PCR-SSCP 方法简便易行, 耗时短, 检出率较高, 仍然被广泛地应用^[3~8]。

SSCP 方法的建立, 几乎全靠经验来进行各种条件的控制和优化。由于有的碱基突变对 DNA 分子构象影响较小, 因而不能用这种方法检测到突变。虽然有人已经注意到突变位点周围的碱基构成对突变的检出率产生影响, 但并不知道哪种情况的突变不能检测^[8]。这使 PCR-SSCP 方法的检出率无法提高, 影响这一方法的应用。

在 SSCP 分析中, 变性时 DNA 双链解开成单链, 在复性时, 单链自身配对形成二级结构, 再在单链自身的二级结构基础上进一步形成高级结构。因此, DNA 碱基突变对单链的二级结构影响可以反映出它对空间结构的影响, 通过了解单个碱基对 DNA 单链二级结构的影响就可以预测这种变化能否通过 SSCP 分析检测到。近年来, 随着结构分子生物学的发展, 大量的 DNA/RNA 二级结构预测软件出现, 使人们通过计算机预测 DNA/RNA 的二级结构成为可能。本文采用核酸二级结构预测软件, 对 DNA 单碱基突变被 SSCP 分析检出的可能性与实际的 SSCP 分析结果进行了比较, 以期用对 DNA 结构的理论预测结果指导 SSCP 分析的实验

设计^[9~12]。

1 材料与方法

1.1 材料

野生型 p53 基因质粒 pc53-SN3 和突变型 p53 基因质粒 pc53-SCX3 由美国 Hopkins 大学 Vogelstein 教授赠送。突变型 p53 基因所编码蛋白的第 143 位由 Val 变为 Ala, 即密码子由 GTG 变为 GCG。

Factor IX 基因的 SSCP 分析由 Sarkar^[13] 在 1992 年发表, 基因序列由 NCBI 的核酸序列数据库中获。RNAstructure3.2 由中国科学院昆明动物研究所刘次全研究员实验室提供。

1.2 试剂

PCR 试剂盒购于基因公司; 引物由上海生物工程公司合成。其他试剂购于云大生化公司。

1.3 方法

1.3.1 PCR 扩增

引物 p53E5A: 5'-ACTCCCCTGCCCTCAACAAG-3'; p53E5B: 5'-CATCGCTATCTGAGCAGCG-3'; p53E7B: 5'-TGGACTCTTCCACTGACAAG-3'。p53E5A, p53E5B 扩增第五外显子, 片段长 184 bp; p53E5A 和 p53E7B 扩增第五, 第六和第七外显子, 片段长 400 bp。

* 国家自然科学基金资助项目 (39760034 和 39960030) 云南省基金资助项目 (1999C002Z)。

** 通讯联系人。

Tel: 0871-5032061, E-mail: dytan3@public.km.cn

收稿日期: 2001-03-16, 接受日期: 2001-05-28

1.3.2 SSCP 分析

取 5 μ l 扩增产物, 加入 5 μ l 变性剂 (95% 甲酰胺, 10 mmol/L EDTA, 0.02% 溴酚蓝). 沸水浴 5 min, 于冰浴中急剧冷却. 用 10% 的聚丙烯酰胺凝胶, 电泳 (8 V/cm) 4 h. EB 染色后, 紫外光下观察并在 Bio-Rad 凝胶扫描仪上扫描处理.

1.3.3 DNA 的单链扩增

在 25 μ l 的 PCR 扩增反应系统中, 加入第一次 PCR 扩增产物 5 μ l (约 0.5 μ g), PCR 引物 p53E5A 或 p53E5B 25 pmol (2.5 μ l), 水 11 μ l, 其他成分不变. 分别扩增 p53 基因 5 号外显子的 184 bp 的单链. 其余操作同于双链 PCR 扩增.

1.3.4 计算机预测分析

将 p53E5 和 factor IX 的标准 DNA 序列片段和相应的含有突变位点的 DNA 序列输入计算机, 用 RNAstructure3.2 程序对两条单链分别进行二级结构预测分析^[9-12].

2 结 果

2.1 p53 基因 SSCP 分析与二级结构预测结果比较

2.1.1 p53 第五外显子的突变检测预测

图 1 显示的是 p53E5 片段的计算机预测结果. 从图 1 中可以看出, 突变型和野生型的正链二级结构没有任何差异, 第 53 位碱基由 T 突变为 C 后, 在此处碱基发生了错配, 但整个二级结构中的其他任何位置的碱基配对都未发生变化. 而负链的二级结构在野生型与突变型间发生了很大的变化, 表明单个碱基的突变会引起空间结构的很大变化.

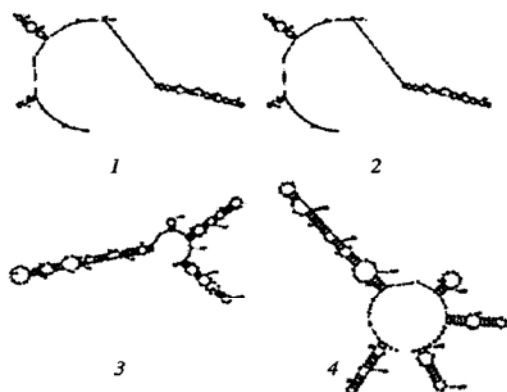


Fig. 1 Predicted second structures of exon 5

1: *cis*-strand of wild-type p53; 2: *cis*-strand of mutate-type p53; 3: *trans*-strand of wild-type p53; 4: *trans*-strand of mutate-type p53.

在图 2a 中, 下方的是双链 DNA 带, 上方两条距离较近的是单链 DNA 带. 在单链带中, 迁移率

较大的单链在野生型和突变型间没有差别, 而迁移率较小的单链带在野生型与较突变型间差异很大. 为确定单链碱基突变导致迁移率改变的单链是正链还是负链, 将 PCR 产物进行单链扩增, 结果见图 2b, 突变型与野生型的单链带有差异的正好是负链, 二级结构预测中, 正链在野生型与突变型间无差异, 而负链在野生型与突变型间差异较大, 预测结构与实验结果很好地吻合.

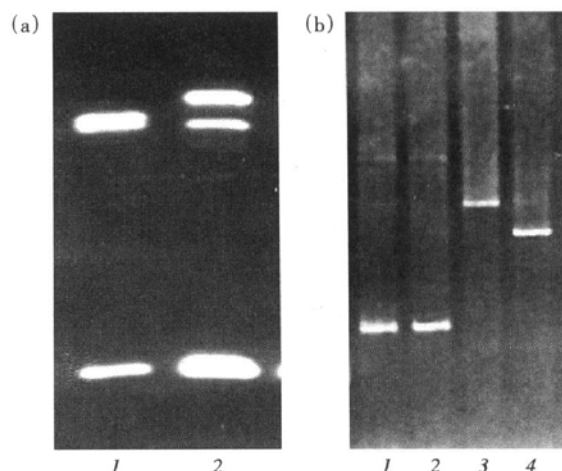


Fig. 2 SSCP analysis of exon 5

(a) double strands SSCP analysis (1: mutated-type; 2: wild-type); (b) single strand SSCP analysis (1: *cis*-strand of wild-type p53; 2: *cis*-strand of mutated-type p53; 3: *trans*-strand of wild-type p53; 4: *trans*-strand of mutated-type p53).

2.1.2 p53E5-E7 野生型和突变型的比较

第五、第六和第七外显子共 408 bp, 单链二级结构预测结果见图 3. 正链有突变的碱基在第 53 位产生错配, 所以, 与单独的第五外显子的二级结构预测结果一样, 在突变型与野生型间无差异. 负链的 5' 端也没差异, 但在 3' 端有碱基的重新配对

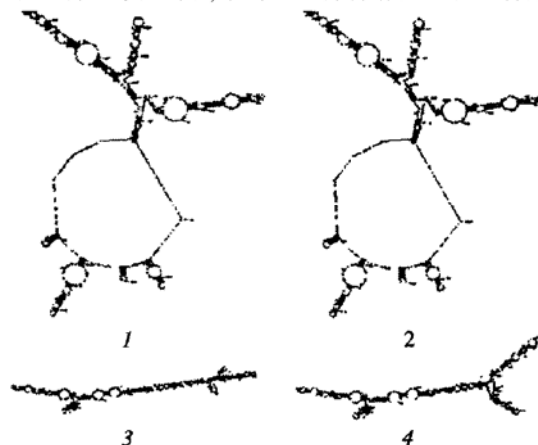


Fig. 3 Predicted second structures of exon 5~7

1: *cis*-strand of wild-type p53; 2: *cis*-strand of mutated-type p53; 3: *trans*-strand of wild-type p53; 4: *trans*-strand of mutated-type p53.

产生二级结构变化. 实验结果 (图 4) 与上述结果也相对应, 迁移率大的链在野生型与突变型间无差异, 而迁移率小的链在野生型与突变型间有差异, 预测结果与实验结果非常吻合. 由于该序列较长, 理论预测还是实验结果都较单一的第五外显子的检测效果差, 表明 SSCP 分析片段不能太长.

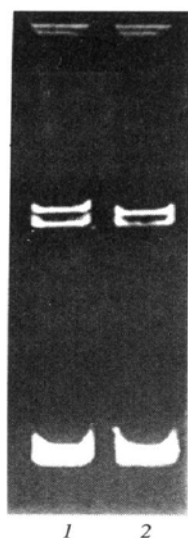


Fig. 4 SSCP analysis of exon 5~7

1: wild-type p53; 2: mutated-type p53.

2.2 Factor IX E8 DNA 二级结构预测与实验结果的比较

Sarkar 等^[13]用凝血因子 IX 的第 8 外显子序列对 DNA SSCP 和 RNA SSCP 的突变检出率进行了比较分析. 为了进一步明确 DNA 二级结构预测与 SSCP 分析的相关性, 我们利用 Sarkar 等的实验结果进一步比较二级结构预测与 SSCP 实验结果的一致性. 分析结果见表 1 和表 2.

Table 1 Comparing result of short segment (183 bp)

No. of mutated sample	Predicted result of <i>trans</i> -strand	Predicted result of <i>cis</i> -strand	Result of SSCP	Result of comparison
1	+	-	-	-
2	+	+	+	+
3	-	+	+	+
4	-	+	-	-
5	+	+	+	+
6	+	+	+	+
7	+	+	+	+
8	+	+	+	+
9	+	+	+	+
10	+	+	+	+
11	+	-	+	+
12	-	+	+	+

Table 2 Comparing result of long segment (307 bp)

No. of mutated sample	Predicted result of <i>trans</i> -strand	Predicted result of <i>cis</i> -strand	Result of SSCP	Result of comparison
1	-	-	-	+
2	-	-	-	+
3	-	+	-	-
4	-	+	+	+
5	+	+	+	+
6	+	-	+	+
7	+	-	+	+
8	+	+	+	+
9	+	-	+	+
10	+	+	+	+
11	+	+	+	+
12	+	-	+	+
13	-	+	+	+
14	+	+	-	-
15	+	+	+	+
16	+	+	+	+
17	-	+	+	+
18	-	+	+	+
19	+	+	+	+
20	+	+	+	+
21	+	-	-	-
22	-	-	+	-

在表 1 中, 被检测的序列长 183 bp, 表中“+”表示突变片段的预测结果和标准对照之间有显著差异,“-”表示二者之间无显著差异. 突变片段的二级结构和标准比较, 碱基未发生重新配对、茎环结构未发生改变视为无差异. 否则, 视为有差异. 在实验结果和预测结果比较时“+”表示二者结果符合,“-”表示不符. 预测结果中正链和负链只要有一条链与野生型序列有差异, 即可判为预测有突变.

从表 1 可以看出, 183 bp 的片段预测结果和实验结果非常吻合, 12 个位点的突变有 10 个位点 SSCP 分析结果与二级结构预测结果一致. 在不吻合的两个位点, 计算机预测指出了有一条链二级结构无差异的 DNA 链, 而另一条链有差异, 但实验结果定为无差异. 从原文给出的实验图谱看, 实验结果可能因差异太小而结果判错^[13].

表 2 中, 被分析的序列长 307 bp. 二级结构预测结果与实验结果之间也吻合得很好, 22 个位点的 SSCP 分析结果有 18 例与二级结构预测结果一致. 突变 3、14、21 和 22 的预测结果和实验不吻合, 但是计算机预测显示突变 3 和 21 中一条链突变型与野生型间只有很小的差别. 所以预测结果与

实验结果完全不吻合的只有突变 14 和 22.

3 讨 论

在本文所考察的序列中, DNA 单链二级结构预测结果有大于 90% 的例子与 SSCP 实验结果吻合, 所以用计算机进行 DNA 二级结构预测, 可以较好地判断所考察的 DNA 序列能否用 SSCP 分析检测出某一特定点的突变. 一些实验结果表明 DNA 的 SSCP 分析对突变的检出率随序列的不同而有很大差异^[8]. 我们对不同片段的 DNA 单链二级结构预测结果也表明单个碱基的突变对二级结构的影响与序列相关. 对 p21 基因的一个片段进行随机点突变预测, 35 个位点只有 3 个位点的突变对二级结构没有影响 (结果未显示). 结果表明对有些序列中的突变检测, SSCP 分析是不很适合的.

我们在计算机模拟 DNA 单链二级结构时还发现, 碱基突变对折叠的影响与突变在序列的中间或两边 (但不是末端) 并无显著关系, 而与折叠后的位置有关, 位于环上碱基的突变比位于茎上的突变对折叠的影响要小一些.

DNA 单链二级结构预测结果与实验结果的很好吻合, 表明结构预测对提高 SSCP 在遗传突变分析中的检测效果很有应用价值. 至少有以下两方面的应用:

a. 在已知突变分析中的应用. 遗传病的基因突变常常都有突变热点, 比如 p53 基因, ras 基因, 引起镰刀型贫血病的 β -珠蛋白基因等. 随机选择一段包含点突变的序列进行 SSCP 分析只有大约 80% 的可能性检测到突变. 计算机 DNA 二级结构预测技术可望用于指导 SSCP 分析片段的选取和 PCR 引物设计, 选择既有利 PCR 扩增, 又有利于 SSCP 分析的序列用于 SSCP 分析, 将对特定点突变的检出大大提高. 对于一个片段上有多个突变热点的序列, 可通过 DNA 二级结构预测, 选择一段能同时检测到几个突变点的序列进行 SSCP 分析.

b. 在对未知基因突变的检测中, 这种 DNA 二级结构预测可以准确估计实际突变率. 在结构模拟中, 一些 DNA 片段中由于被选取片段的一个或几个突变点位于对分子折叠影响不大的, 显示出较低的突变检出率, 这意味着对这些片段的 SSCP 分析结果不可避免地会漏掉一些突变. 这种遗漏可能通过二级结构预测对实验测定的突变率进行矫正. 首

先对被测片段进行随机突变, 然后通过二级结构预测可检测的突变位点数, 计算出突变的理论检测率. 实际突变率可用下式计算:

$$\text{实际突变率} = \frac{\text{实验突变率}}{\text{理论检出率}}$$

参 考 文 献

- Orita M, Iwahana H, Kanazawa H. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA*, 1989, **86** (8): 2766 ~ 2770
- Wawer C, Ruggeberg H, Meyer G, *et al.* Simple and rapid electrophoresis method to detect sequence variation in PCR-amplified DNA fragments. *Nucleic Acids Res*, 1995, **23** (23): 4928~ 4929
- 赵永忠, 徐湘民, 徐 晴. PCR-LIS-SSCP 快速分析非缺失型 α -地中海贫血病点突变. *中华医学遗传学杂志*, 1999, **16** (4): 113~ 115
Zhao Y Z, Xu X M, Xu Q. *Chin J Med Genet*, 1999, **16** (4): 113~ 115
- 唐冬生, 禹宽平, 汤熙翔, 等. 生物信息学技术克隆人类神经鞘蛋白零家族基因. *生物化学与生物物理学学报*, 2000, **32** (4): 364~ 368
Tang D S, Yu K P, Tang X X, *et al.* *ACTA Bioch et Bioph*, 2000, **32** (4): 364~ 368
- 邢德印, 谭 文, 宋 南, 等. DNA 修复基因 hOGG1 多态性与食管癌遗传易感性. *中华医学遗传学杂志*, 2000, **17** (4): 241~ 243
Xing D Y, Tan W, Song N, *et al.* *Chin J Med Genet*, 2000, **17** (6): 377~ 380
- Gasser R B, Monti J R. Identification of parasitic nematodes by PCR-SSCP of ITS-2 rDNA. *Mol Cell Probes*, 1997, **11** (3): 201~ 209
- Becker K F, Reich U, Schott C, *et al.* Identification of eleven novel tumor-associated E-cadherin. *Hum Mutat*, 1999, **13** (2): 171
- Nataraj A J, Olivos-Glander I, Kusakawa N, *et al.* Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. *Electrophoresis*, 1999, **20** (6): 1177~ 1185
- Walter A E, Turner D H, Kim J, *et al.* Stacking of helices enhances binding of oligoribonucleotides and improves predictions of RNA folding. *Proc Natl Acad Sci USA*, 1994, **91** (20): 9218~ 9222
- SantaLucia Jr J. Unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci USA*, 1998, **95** (25): 1460~ 1465
- Sugimoto N, Nakano S, Katoh A, *et al.* Thermodynamic parameters to predict the stability of RNA/DNA hybrid duplexes. *Biochemistry*, 1995, **34** (35): 11211~ 11216
- SantaLucia Jr J, Allawi H T, Seneviratne P A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 1996, **35** (11): 3555~ 3562
- Sarkar G, Yoon H S, Sommer S S. Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP nucleic acids research. 1992, **20** (4): 871~ 878

The Comparison of Predicted Second Structure of DNA Single Strand with Single Strand Conformation Polymorphism Analysis Results*

HE Yun-Gang, TAN De-Yong**, QIAN Wei, LAI Jian-Hua, XIE Yong-Fang

(Department of Biotechnology, Yunnan University, Kunming 650091, China)

Abstract Single strand conformation polymorphism (SSCP) has been widely used for mutation detecting. But the theory about it is not strong enough to guide application. Recently, many computer programs have been designed for predicting the folding of single strand DNA or RNA. Second structures of three DNA segments with RNAstructure 3.2 have been predicted and have been compared with their results of SSCP. The comparison shows there is closely relation between them. The kinds of prediction can guide operator to select primers for SSCP analysis and accurately estimate mutated ratio.

Key words single strand conformation polymorphism (SSCP), mutation detect, structure prediction of DNA single strand

* This work was supported by grants the National Natural Science Foundation of China (39760034, 39960030) and the National Natural Science Foundation of Yunnan Province (1999C002Z).

** Corresponding author. Tel: 86-871-5032061, E-mail: dytan3@public.km.cn

Received: March 16, 2001 Accepted: May 28, 2001

基因分析新技术及其应用学术交流会通知

为了促进分子遗传学和基因分析技术在生命科学各领域的推广和应用,总结经验、交流成果,定于2002年6月上旬在浙江温州举办“基因分析新技术及其应用学术交流会”。

一、主办单位:中国遗传学会《遗传》杂志编委会

二、承办单位:浙江温州医学院检验医学与公共卫生学院

三、协办单位:温州市生物化学与分子生物学学会

四、大会主席:《遗传》杂志副主编曾溢滔院士、夏家辉院士

五、大会秘书长:《遗传》杂志编辑部主管李绍武编审;温州医学院检验医学与公共卫生学院院长、细胞与分子医学研究所所长吕建新教授

六、交流内容:1. 基因序列分析与基因组研究;2. 新基因的克隆、鉴定、定位与表达;3. 基因诊断与基因治疗;4. DNA技术的法医学应用;5. 基因突变的检测;6. 基因芯片技术的应用;7. 国内外基因分析新技术、新方法、新进展;8. 生物信息学;9. 基因分析新设备、新仪器及新书刊展销。

七、收费标准:会期4~5天,每位代表收会务及餐饮费650元,研究生550元。发放会议论文集和2002年全年《遗传》杂志。2002年5月1日前交费者会议费600元,研究生500元,请汇到温州市生物化学与分子生物学学会。

开户银行:工商银行温州市分行营业部 帐号:120320209064008035

八、征文与报名:会议征文2000~4000字,参见《遗传》杂志的格式,内容有:题目、作者姓名、工作单位及邮政编码、正文、参考文献、基金项目、第一作者简介、电话与电子信箱。文内可有少量图、表和照片。请注明“会议征文”字样。征文截止日期:2002年4月1日,书面投稿的同时请发来电子邮件(Word文件),E-mail:swli@genetics.ac.cn

经审查选用的论文编入《基因分析新技术原理及其应用》一书由中国科技出版社正式出版,部分论文由《遗传》杂志选用。大会将评选优秀论文,颁发证书和奖金。

有无征文均可报名参加会议。欢迎公司企业参展。国内公司1500元/展位,外商3000元/展位,并列入大会通讯录,发放会议资料。

联系地址:北京市安定门外大屯路917大楼 中国科学院遗传研究所编辑室

邮编:100101 联系人:李绍武 电话:010-64889354 传真:010-64889348