

三核苷酸双链重复序列扩展合成特性及其机理*

王 阳 贾蕾敏 董 平 梁兴国**

(中国海洋大学食品科学与工程学院, 青岛 266003)

摘要 简单重复序列在各种生物基因组中广泛存在, 同分子进化、遗传多样性、分子标记和某些遗传性疾病等密切相关. 本文以全部 32 种 18 bp 三核苷酸双链重复序列作为研究对象, 对它们在聚合酶作用下的扩展合成进行了系统研究. 探讨了反应温度、序列本身等对扩展效率和产物长度的影响. 结果显示, 几乎所有的序列都能扩展变长, 但序列对其扩展效率有很大影响: GC 含量较多的短链, 尤其是一条链中同时有 G 和 C 的短链扩展效率较高; 全由 AC 和 GT 组成的双链也较易扩展. 短链的最适扩展温度与短链 GC 组成呈一定的正相关关系. 琼脂糖凝胶电泳和聚丙烯酰胺凝胶电泳分析发现, 大部分产物单一性良好, 且产物分子质量的大小随反应时间线性增加; 随着反应温度升高, 产物差异性增大. 最后分析了双链重复序列的“复制滑移”扩展机理, 有望为进一步研究重复序列的分子进化和基因检测中的非特异性扩增等奠定基础.

关键词 三核苷酸重复序列, 重复序列, 非特异性扩增, 分子进化

学科分类号 Q71, Q52

DOI: 10.3724/SP.J.1206.2012.00308

串联重复序列(tandem repeat sequences), 如 (AT)_n、(CAG)_n 等又称为卫星 DNA (satellite DNA), 由较短的核心序列以头尾相串联的方式组成, 广泛存在于真核生物和一些原核生物的基因组中, 所占比重较大, 并表现出长度、碱基组成等多态性^[1-3]. 串联重复序列同分子进化、遗传多样性、分子标记、某些遗传性疾病等密切相关. 如三碱基串联重复序列的长度变化会引起亨廷顿舞蹈症(Huntington's disease)、强直性肌萎缩症(myotonic dystrophy)、弗里德赖希共济失调(Friedreich's ataxia)等十几种严重的遗传性疾病^[4-7]. 另一方面, 正是由于这些简单重复序列容易变异, 其对生物进化和物种多样性等有很大贡献, 并可作为区分不同物种的特异性基因标记^[8-10]. 但由于 10 年前人类基因组计划才刚刚完成, 有关重复序列的全面研究还处于发展中, 以上现象的相关细节与原理还不十分清楚, 有待进一步深入研究.

关于重复序列的扩展机制, 存在着多种假说, 如分子内分子间重组(intra or intermolecular recombination)、滑动链错配(slipped strand mispairing)、非正常延伸模型(illegitimate elongation

model)等^[11-13]. 如 Ogata 等^[14-16]研究了短链回文重复序列 DNA 的扩展产物特性, 认为发夹型结构在扩展初始阶段起了重要作用, 而在扩展后期分子间的滑动更为重要. 但这些机制都有其欠缺之处, 如不能解释扩展的高效, 不适于解释较低浓度下的扩展反应, 或很难在某些反应条件下产生. Liang 等^[17-20]曾提出了末端形成发卡结构不断重复扩展的机理, 但还有待进一步实验证明.

另据报道, 在无模板和引物存在的体外反应体系中, 一些耐热性聚合酶能在体外直接由 dNTPs 从头合成一些串联重复的 DNA 序列, 这些序列容易生成并可很快扩展至 50 kb 以上^[21-23]. 这支持了核酸可能起源于简单序列的说法, 为生命起源前 DNA 的形成提供了一些线索. 更进一步的研究发现, 在聚合酶和内切酶共存的条件下, 从头合成

* 国家青年千人计划, 山东省万人计划, 长江学者和创新团队发展计划和山东省自然科学杰出青年基金(JQ201204)资助项目.

** 通讯联系人.

Tel: 0532-82031086, E-mail: liangxg@ouc.edu.cn

收稿日期: 2012-06-21, 接受日期: 2012-09-25

DNA 的效率可得到极大的提高^[14-17]. 这是一个非常有趣的现象, 将合成 DNA 的酶和破坏 DNA 的酶混合在一起, 却可以“无中生有”, 仅以 dNTP 为原料生成大量含有重复序列的 DNA. 这可用于解释如何获得用于进化成简单生命的大量核酸材料.

有零散的研究发现一些短链的简单重复序列如 (TA)_n、(GAA)_n/(TTC)_n、(CCTG)_n、(dG)_n/(dC)_n 和回文重复序列等可以在聚合酶的作用下进行扩展, 甚至生成各种长达 10 kb 的重复序列(表 1). 但大

多只是作为一种较为异常的(unusual)现象或是针对一些遗传性疾病而进行的独立研究, 对于机理的探讨也较为肤浅. 目前还没有发现对相关的简单重复序列的扩展进行系统研究的报道. 本研究通过对全部 32 种 18 bp 三核苷酸双链重复序列的扩展进行系统研究, 全面探索短链重复序列的扩展特性以及扩展机理, 希望能对研究核酸的分子进化、解释串联重复序列扩展而产生的疾病以及基因检测与诊断中非特异性扩增的抑制等提供线索和支持.

Table 1 Summary of repetitive DNA elongation with various DNA polymerases

Short DNA sequence (5'→3')	Reaction temperature/℃	DNA length/bp		DNA polymerase	Reference
		Before elongation	After elongation		
(dG) ₁₀ /(dC) ₁₀	37	10/10	10 000	Klenow(exo ⁻)	24
(AT) ₈	65	16	20 000	Taq pol	25
(TA) ₉	65	18	20 000	Taq pol	26
(CA) ₄₋₁₂ (TG) ₄₋₁₂	PCR	8~24	>20 000	Taq pol	27
(TC) ₁₀ /(GA) ₁₀	PCR	20/20	>20 000	Taq pol	27
(TTG) ₅ /(CAA) ₅	37~72	15/15	>20 000	Taq pol	27
(AAT) ₃ /(ATT) ₅	47	9/15	20 000	Klenow	28
(CAG) ₃ /(CTG) ₅	47	9/15	2 000	Taq pol	28
(GCC) ₅	37	15	250	Taq pol	29
(GCC) ₅	37	15	80	Klenow	29
(GCC) ₅	37	15	45	HIV/RT	29
(GAA) ₁₀₋₁₇	37	30, 51	>1 000	Klenow	30
(TTC) ₁₀₋₁₇	37	30, 51	>1 000	Klenow	30
(TAGG) ₁₀ /(CCTA) ₁₀	70	40/40	10 000	Taq pol	31
(TTAAGGG) ₅ /(CCCCTAA) ₅	75	40/40	10 000	Taq pol	31
(TACATGTA) ₆	74	48	60 000	Tth pol	32
(TACATGTT) ₆	74	48	60 000	ΔTth pol	32

1 材料与方法

1.1 材料

64 种 18 bp 串联重复序列由 Integrated DNA Technologies 公司合成, 如(AGC)₆、(GCT)₆等; 双链形式可表示为(AGC)₆/(GCT)₆和(ATA)₆/(TAT)₆等; Vent 和 Vent (exo⁻) DNA 聚合酶、DNA 分子质量标记(Marker)等购自 New England Biolabs; SYBR Green I 购自 Invitrogen 公司; 鲑鱼精 DNA 购自 Sigma 公司; 核酸共沉淀试剂盒购自 Takara 公司, 其他试剂均为国产分析纯试剂.

1.2 短链重复序列的恒温扩展

32 种互补配对的 18 bp 三核苷酸重复序列反应液加入 96 孔板. 为使形成的互补 DNA 达到最佳复性状态, 将短链进行预处理: 置于 10 μl 1×耐热聚合酶缓冲液(Thermopol Buffer)体系中 (1×Thermopol Buffer: 20 mmol/L Tris-HCl, pH 8.8 (25℃), 10 mmol/L (NH₄)₂SO₄, 10 mmol/L KCl, 2.0 mmol/L MgSO₄, 0.1% Triton X-100), 加热至 90℃ 热变性 3.0 min, 缓慢冷却至室温. 加入 10 μl 含 40 U/ml Vent 或 Vent(exo⁻)DNA 聚合酶, 1.0 mmol/L dNTPs, 1×Thermopol Buffer 至上述预处理短链溶

液中, 轻微混匀后盖上 PCR 板硅胶盖, 于恒温培养箱或 PCR 仪中反应. 最终反应条件为: 总体积 20 μ l, 20 U 聚合酶, 0.5 mmol/L dNTPs. 实验设置反应温度为 37 $^{\circ}$ C、50 $^{\circ}$ C、60 $^{\circ}$ C、70 $^{\circ}$ C、80 $^{\circ}$ C. 反应一定时间后定点取样.

1.3 SYBR Green I 荧光定量法分析反应产物

配置标准鲑鱼精 DNA 梯度浓度溶液, SYBR Green I 染色, 染料终浓度为 1 $\times$, 避光染色 10~30 min. 设定荧光酶标仪(美国分子 M5)的激发波长为 497 nm, 发射波长为 520 nm, 测定荧光值. 实验表明, dsDNA 浓度在 12.5~500 mg/L 浓度范围内呈良好的线性关系, $r^2 = 0.9992$, 检测限可达 1.56 μ g/L. 本研究中将产物用 TE 缓冲液稀释 25~400 倍后, 使 dsDNA 的浓度保持在 12.5~500 mg/L 范围内进行反应产物的定量测定.

1.4 扩增产物的电泳分析

取 5.0 μ l 扩增产物, 0.8% 琼脂糖凝胶电泳 1.0 h 或 6% 聚丙烯酰胺凝胶电泳 1.5 h, EB 染色 30 min 后置于伯乐凝胶成像仪(Gel Doc XR⁺)上成像分析.

1.5 短链熔点测定

将短链稀释成终浓度为 1.0 μ mol/L 溶解于 1 $\times$ 不含变性剂的反应缓冲液(20 mmol/L Tris-HCl, 10 mmol/L (NH₄)₂SO₄, 10 mmol/L KCl, 2.0 mmol/L MgSO₄, pH 8.8 25 $^{\circ}$ C). 使用岛津 UV1800 测定样品吸光值随温度 20~90 $^{\circ}$ C 的变化情况, 同步测定吸光值. 同时用 DINAMelt Web Server 软件对短链及其扩展后的长链产物的熔点进行计算.

2 结 果

2.1 短链三核苷酸重复序列扩展反应的序列依存性

本研究使用了 18 个碱基长的三核苷酸重复序列, 这样的长度足够形成发卡结构或作为引物进行 DNA 合成(一般 PCR 引物大于 18 个碱基). 如果把末端碱基不同而重复单位相同的三核苷酸重复序列看作不同的序列, 理论上共有 64(4³)种三核苷酸重复序列. 如 (ACC)₆、(ACT)₆、(ACA)₆、(ACG)₆ 等. 这些序列可以形成 32 种完全互补的双链, 如 (AAG)₆/(CTT)₆、(AGC)₆/(GCT)₆ 等. Vent DNA 聚合酶是存在于深海古生菌 *Thermococcus litoralis*(生存于深海的热水口处, 可在 98 $^{\circ}$ C 的高温下生存)中的一种具有高保真性能的耐热性聚合酶, 被用于 PCR 和引物延伸等反应, 它具有很强的 3'→5' 外切

酶活性. Ogata 等^[23, 33]曾报道 Vent DNA 聚合酶可以进行 *de novo* DNA 合成, 而且混入内切酶可大大提高其合成效率^[17-18]. 因 Vent DNA 聚合酶的以上特点和耐热性古生菌在进化过程中的重要性, 在此我们选用它来进行简单重复序列的扩展反应.

首先我们研究了 Vent DNA 聚合酶对各个序列的扩展. 在 70 $^{\circ}$ C 下反应 4.0 h 的琼脂糖电泳分析的结果如图 1 所示. 由于完全互补的双链不能形成用于进行 DNA 生物合成的“模板/引物”结构, 一般双链 DNA 不能扩展变长. 但反应后的结果显示, 除 (GGG)₆/(CCC)₆ 外, 其他 31 种双链都或多或少有产物生成. 由于所用核苷酸的浓度较低 (50 nmol/L), 无产物生成时用 SYBR Green I 染色基本不能观测到条带. 从图 1 可以看出, 序列不同得到产物的长度各异. 主要可以分为两种, 一是较单一的片段, 产物长度分布很窄, 一是产物在很大范围呈弥散带状, 甚至有些滞留在胶孔里, 产物分

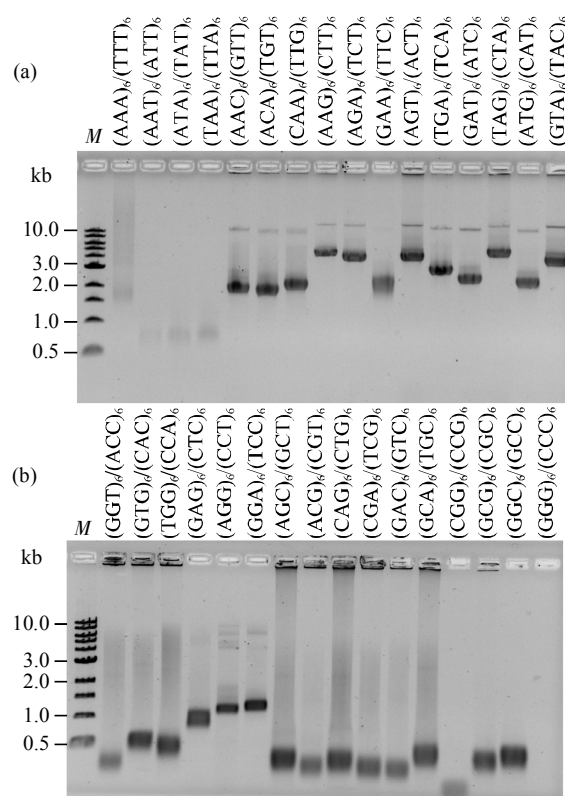


Fig. 1 Agarose gel electrophoresis analysis of expansion products of all 18 bp short trinucleotide repeats

(a) Short repeats with 0 or 1/3 GC content. (b) Short repeats with 2/3 or 100% GC content. Reaction conditions: 50 nmol/L short repeats, 10 U/ml Vent, 0.5 mmol/L dNTPs, in 20 μ l of 1 $\times$ Thermopool Buffer (20 mmol/L Tris-HCl, 10 mmol/L (NH₄)₂SO₄, 10 mmol/L KCl, 2.0 mmol/L MgSO₄, 0.1% Triton X-100, pH 8.8 25 $^{\circ}$ C), 70 $^{\circ}$ C, 4.0 h. Analyzed by 1% agarose gel.

子较大. 参照相关文献, 单一长度的产物可判断为由双链 DNA 的 2 条链间滑移而逐步变长的 DNA 产物, 而滞留在胶孔里的产物多为按照一种未知的高效扩增机制而生成的长达数 10 kb 的 DNA^[17-18].

如图 1 所示, 对于只有 A/T 碱基对的序列, 产物量较少; 对于只有 G/C 碱基对的序列, 产物一般很短. 还可以发现, 对于既有 A/T 又有 G/C 碱基对的序列, 当一条链上全为嘌呤碱基(如 (AAG)₆、(AGG)₆), 而互补链全为嘧啶碱基时, 几乎只得到单一的产物, 当 GC 含量为 1/3 时, 基本上以单一产物为主, 当 GC 含量为 2/3, 尤其是当一条链上既有 G 又有 C 时, 产物以在电泳图上呈弥散状的长产物为主.

在 4.0 h 内各个序列生成单一条带产物的平均速率列于表 2. 可见序列不同其生成单一条带产物的速率也各不相同. GC 含量为 1/3 的序列, 具有较高的生成速率, 基本上能在 4.0 h 的扩展反应后得到 2.0 kb 以上的产物, 扩展速率达到 8.0 bp/min 以上. 可以看出每 3 种序列的内部重复序列相同, 只是末端序列不同, 如(AAC)₆、(ACA)₆和(CAA)₆的内部都可以看作是 AAC 的重复. 这 3 种序列都具有相似的扩展速率. 有些序列也有较大的不同, 如(AAG)₆/(CTT)₆较(GAA)₆/(TTC)₆扩展速率快很多. 当一条链上全是嘌呤碱基时, 如(AAG)₆、(AGG)₆等, 扩展速率相对较高.

Table 2 Length of final product and synthesis rate for different 18 bp simple sequence repeats

Sequence	Length after 4.0 h/bp	Synthesis rate/(bp·min ⁻¹) ¹⁾
(AAA) ₆ /(TTT) ₆	1 500	6.3
(AAT) ₆ /(ATT) ₆ , (ATA) ₆ /(TAT) ₆ , (TAA) ₆ /(TTA) ₆	800	3.3
(AAC) ₆ /(GTT) ₆ , (ACA) ₆ /(TGT) ₆ , (CAA) ₆ /(TTG) ₆	2 000	8.3
(AAG) ₆ /(CTT) ₆ , (AGA) ₆ /(TCT) ₆	4 000	16.7
(GAA) ₆ /(TTC) ₆	2 000	8.3
(AGT) ₆ /(ACT) ₆ , (GTA) ₆ /(TAC) ₆ , (TAG) ₆ /(CTA) ₆	4 000	16.7
(TGA) ₆ /(TCA) ₆	3 000	12.5
(GAT) ₆ /(ATC) ₆ , (ATG) ₆ /(CAT) ₆	2 000	8.3
(GAG) ₆ /(CTC) ₆ , (AGG) ₆ /(CCT) ₆ , (GGA) ₆ /(TCC) ₆	1 000	4.2
(GTG) ₆ /(CAC) ₆ , (TGG) ₆ /(CCA) ₆	500	2.1
(GGT) ₆ /(ACC) ₆	< 500	< 2.1
(AGC) ₆ /(GCT) ₆ , (GCA) ₆ /(TGC) ₆ , (CAG) ₆ /(CTG) ₆	< 500	< 2.1
(CGA) ₆ /(TCG) ₆ , (GAC) ₆ /(GTC) ₆ , (ACG) ₆ /(CGT) ₆	< 500	< 2.1
(CCG) ₆ /(CGG) ₆ , (CGC) ₆ /(GCG) ₆ , (GCC) ₆ /(GGC) ₆	< 500	< 2.1
(GGG) ₆ /(CCC) ₆	< 500	< 2.1

¹⁾ The rates were determined under standard reaction conditions.

2.2 短链三核苷酸重复序列扩展反应的温度依存性

所用双链 DNA 的熔点(T_m)无疑会对其扩展的温度依存性产生很大影响. 我们测定了一些有代表性的序列的熔点(表 3). 可以看出, 除 GC 含量在 100%以上的序列外, 其他双链的 T_m 都低于 70℃, 即扩展反应的温度. 如(AAG)₆/(CTT)₆的 T_m 只有

50.3℃, 比扩展反应温度低 20℃. 即(AAG)₆/(CTT)₆在扩展反应的初始阶段, 绝大部分呈单链状态. 为了考察温度对扩展反应的影响, 我们选用了 (GAT)₆/(ATC)₆和(AGC)₆/(GCT)₆, 研究了在 37℃, 50℃、60℃、70℃和 80℃各个温度下的扩展反应, 产物的电泳分析结果如图 2 所示.

Table 3 T_m values of several short repeats

Sequence	Measured T_m of short repeats / $^{\circ}\text{C}$ ¹⁾	Calculated T_m of short repeats / $^{\circ}\text{C}$ ^{1,2)}	Calculated T_m of long products / $^{\circ}\text{C}$ ²⁾
AAA/TTT	44.7	41.8	66.1
AAT/ATT	35.8	34.9	59.4
AAC/GTT	55.9	57.2	78.5
AAG/CTT	50.3	49.7	75.1
ACT/AGT	51.1	47.5	71.3
ATC/GAT	53.6	50.1	75.9
ACC/GGT	66.7	66.3	92.0
AGG/CCT	63.5	62.8	89.0
ACG/CGT	70.1	66.4	89.5
AGC/GCT	67.5	68.4	93.1
CGG/CCG	81.4	82.1	107.8
GGG/CCC	88.9	79.5	109.6

¹⁾ The length of short repeats was 18 bp. ²⁾ The T_m was obtained by DINAMelt Web Server which is the software to predict melting profiles for nucleic acids. The length of long products was set to be 102 bp, and T_m will not increase for longer duplex.

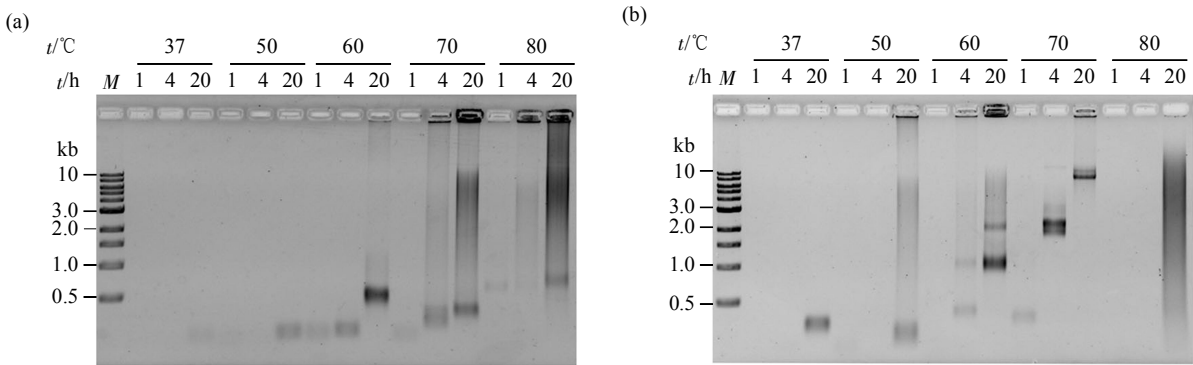


Fig. 2 Effect of temperature on expansion of short trinucleotide repeats

(a) $(\text{AGC})_6/(\text{GCT})_6$. (b) $(\text{GAT})_6/(\text{ATC})_6$. Reaction was carried out at 37°C , 50°C , 60°C , 70°C and 80°C for 1.0, 4.0, and 20 h, respectively. Other reaction conditions are the same as shown in Figure 1.

对于 $(\text{AGC})_6/(\text{GCT})_6$ ，在低于其 $T_m(67.5^{\circ}\text{C})$ 的 37°C 、 50°C 和 60°C 下反应，基本呈单一条带(图 2a); 而在高于其 T_m 的 70°C 和 80°C 下反应产物则以弥散带为主。在 37°C 和 50°C 下的反应速度很慢，只有在 20 h 后才能观测到少量产物，在 60°C 以上时，1.0 h 就有产物出现。在 60°C 下，单一条带产物的生成速度最快，可达到约 0.50 bp/min。对于 $(\text{GAT})_6/(\text{ATC})_6(T_m = 53.6^{\circ}\text{C})$ 只有在 37°C 和 70°C 下反应时产物以单一条带为主，而在 50°C 和 80°C 下反应时产物以弥散带为主，在 60°C 时的反应产物是单一条带和弥散带共存(图 2b)。在 60°C 时

单一条带产物生成速度约为 4.2 bp/min，在 70°C 下，单一条带产物的生成速度最快，可达到约 8.5 bp/min。在各个温度下， $(\text{GAT})_6/(\text{ATC})_6$ 的单一条带生成速率明显比 $(\text{AGC})_6/(\text{GCT})_6$ 快。虽然所用耐热性 Vent 聚合酶的最佳反应温度范围为 $70\sim 80^{\circ}\text{C}$ ，两者在 37°C 都有产物生成。

为进一步研究 Vent 聚合酶在 37°C 下生成扩展产物的长度与反应时间的关系，我们详细研究了 $(\text{AAA})_6/(\text{TTT})_6$ 的产物长度随时间的变化，结果如图 3 所示。反应 0.5 h 后产物即达到 100 bp 以上，产物的长度主要分布在 $100\sim 130$ bp 之间。反应产

物分子大小随着反应时间的延长而增大, 呈较单一的明亮条带, 说明长度的分布较窄(图 3a). 如图 3b 所示, 在 12 h 以内, 反应产物的长度同反应时间有很好的线性相关关系, 平均反应速率达到 1.6 bp/min. 24 h 后合成速率有所降低(图 3a). 在

反应条件下, 当双链足够长(> 50 bp)时 Poly-A/Poly-T 的 T_m 可达到 66.1°C (表 3), 因此, 在 37°C 反应温度下, 反应进行 0.5 h 以后, 产物都以双链的形式存在.

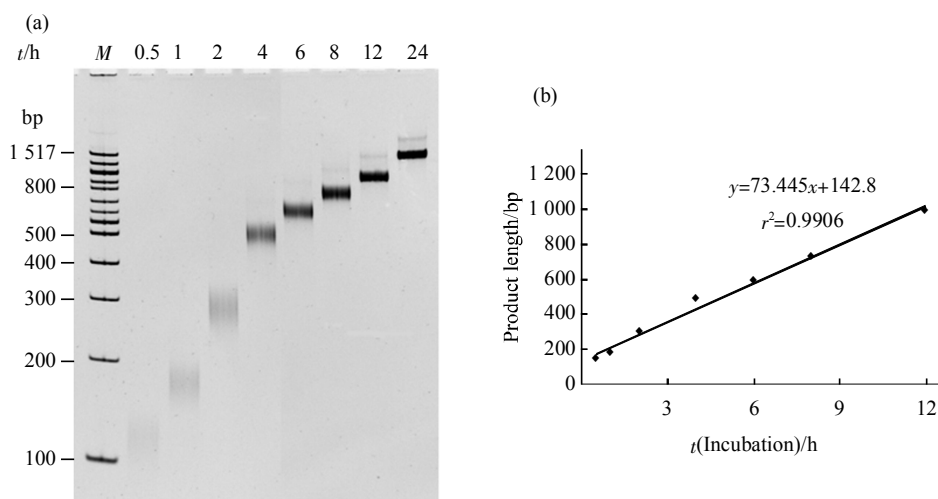


Fig. 3 Time course for the expansion of (AAA)₆/(TTT)₆ at 37°C

Aliquots were withdrawn after time intervals as shown on the top of the PAGE figure. (a) The reaction products were resolved on 6% PAGE gel and stained with ethidium bromide. (b) Dependence of the product length estimated from (a) on reaction time.

以上线性扩展也可在高温条件下进行. 如 (GCC)₆/(GGC)₆ 和 (CCG)₆/(CGG)₆ 在 80°C 下的扩展反应在前 8.0 h 内都有很好的线性关系, 扩展速率可达到 2.0 bp/min 以上. 8.0 h 后扩展反应难以继续进行(data not shown).

2.3 短链三核苷酸重复序列扩展反应的系统研究

最后我们将所有序列的双链三碱基重复序列分别在 37°C 、 50°C 、 60°C 、 70°C 和 80°C 进行了反应, 以系统研究它们的扩展特性. 由于样品数量太多, 我们用 96 孔样品板进行反应, 反应产物直接用 SYBR Green I 进行染色测定荧光, 对生成的 DNA 产物进行了定量测定(图 4). 可以看出, 几乎所有短链都有 DNA 生成. 但是大多数序列的扩展效率随温度变化差异很大, 每一种序列都有自己适宜的扩增温度. 只有 AT 碱基的如 (AAT)₆/(ATT)₆、(TAA)₆/(TTA)₆ 和 (ATA)₆/(TAT)₆ 等的适宜扩展温

度为 60°C . GC 含量为 1/3 的 (TCA)₆/(TGA)₆、(AGT)₆/(ACT)₆、(ACA)₆/(TGT)₆ 和 (ATG)₆/(CAT)₆ 等的适宜扩展温度为 70°C . GC 含量为 2/3 的 (ACG)₆/(CGT)₆、(CGA)₆/(TCG)₆ 和 (CAC)₆/(GTG)₆ 等的适宜扩展温度为 80°C , 但在 70°C 都有较快速的扩展. 而 GC 含量为 100% 的 (CCG)₆/(CGG)₆ 等只在 80°C 有高效的扩展. 即一般而言, GC 含量越高, 越易在较高的温度进行扩展.

尽管各自的最适宜扩展温度不同, 但在短链各自的最适扩展温度下, GC 含量高而且同一条链中同时有 G 和 C 碱基的序列扩展效率最高, 其次为 GC 含量为 1/3 的序列. 但由于有些序列的扩展产物中混有一些呈弥散带的产物, 荧光值不能直接反映扩展长度, 即荧光值高时产物的长度不一定长. 如 (ACC)₆/(GGT)₆ 的荧光值很高, 但单一片段的长度很短(图 1).

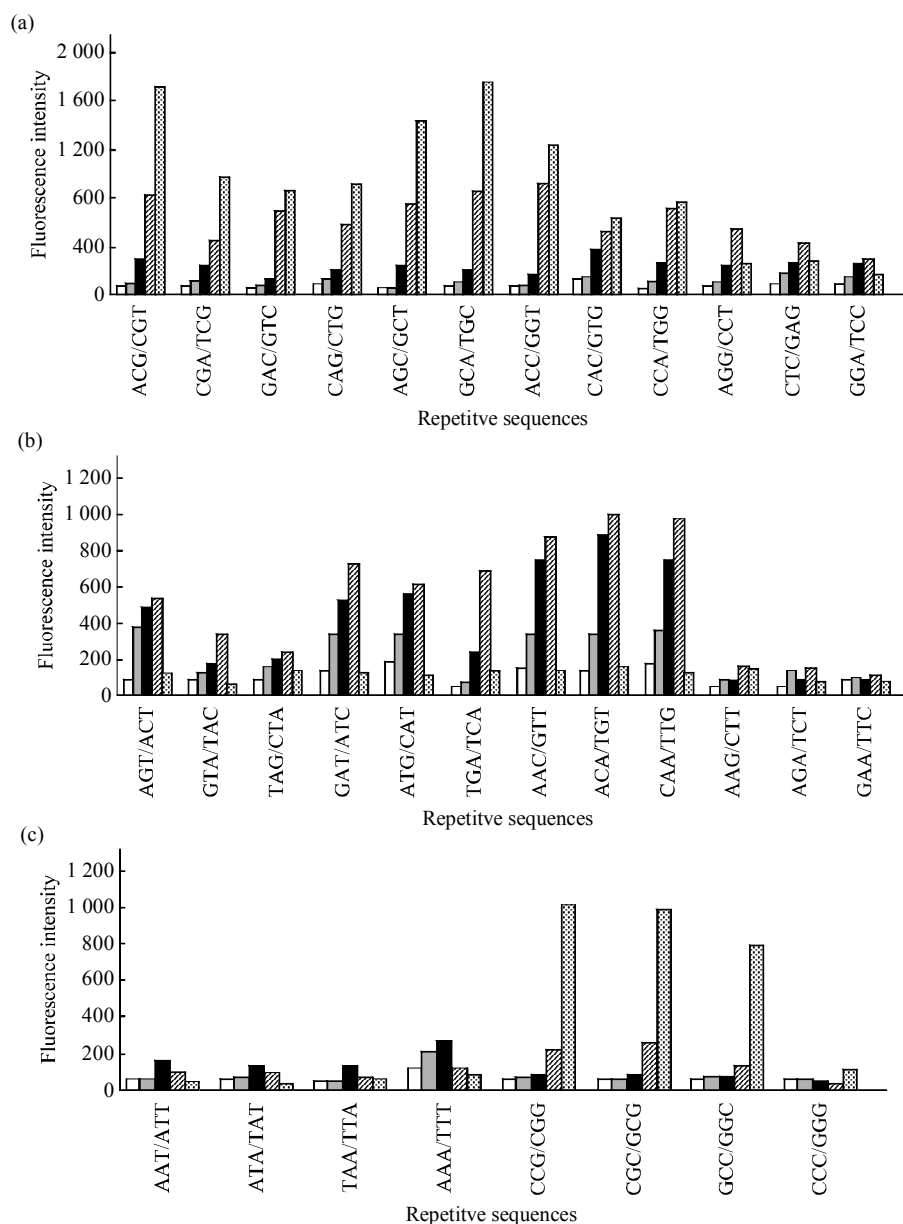


Fig. 4 Systematic study of expansion of all short trinucleotide repeats at various temperatures

The 18 bp short repeats is the abbreviation for trinucleotide, such as (ACG)₆/(CGT)₆ is written as ACG/CGT. The reaction conditions are the same as shown in Figure 1. Fluorescence was measured after the reaction solution was diluted by 1× TE buffer containing 1× SYBR Green I. (a) Duplex with more G/C base pairs. (b) Duplex with more A/T base pairs. (c) Duplex with only A/T or G/C base pair. □: 37°C; ■: 50°C; ■: 60°C; ▨: 70°C; ▩: 80°C.

3 讨 论

由以上结果可知, 三核苷酸串联重复序列的扩展具有普遍性。几乎所有三核苷酸重复序列都有非正常扩展的特性, 但不同的序列扩展能力差异较大, 荧光定量显示 GC 含量较高的短链较易扩展, 电泳显示 AT 含量较高的短链扩展产物单一性较好。短链在 37°C ~ 80°C 范围内的扩展具有一定的

温度依存性, 随着短链 GC 含量增大, 最适扩展温度普遍升高, 在低于熔点但与熔点接近的温度下最易进行扩展反应。此研究中采用的是耐热性聚合酶, 考虑到聚合酶的适宜反应温度, 扩展反应速率并不完全和温度呈规律性的正相关。常温 37°C 反应条件下虽然扩展反应较慢, 但只要时间足够长, 也可发生扩展。

在互补短链组成的扩展反应体系中, 除双链自

身的扩展外, 还存在着其他副反应. 副反应主要为单链的扩展反应, 其次是 DNA 的从头合成, 主反应和副反应之间存在一定的竞争关系. 当温度较高, 且所用 DNA 链中互补碱基较多时, 容易发生单链自身的伸长反应, 即在 3' 端容易形成分子内的错配, 并在聚合酶的作用下发生类似于引物 / 模板的引物伸长反应. 由于 Vent 聚合酶在只有 dNTP 的存在下容易从头合成 DNA^[15, 19], 当单链和双链都很难扩展伸长时, 也可能发生从头合成. 单链的自身伸长的产物多在电泳加样孔中移动度很小, 而从头合成的产物为范围较宽的弥散带^[15, 19]. 因此, 本文中在电泳图中表现为从胶孔至明亮单一带之间的

弥散带的产物可判断为相应的副产物. 随着反应温度的升高或时间的延长, 副产物有所增多. 但在多数条件下对双链重复序列扩展反应的影响不太大. 已有研究表明单链短重复序列扩增多以发夹型结构开始. 单链自身在恒温条件下形成末端发夹型特殊结构, 聚合酶结合在 3' 端开始延伸至 5' 端. 自我互补的单链 DNA 产物在不稳定体系中一经形成末端发夹型结构, 便以此种方式继续延伸扩展^[17]. 从头合成反应需要一定的滞后时间, 一般在 70℃ 以上较高温度下效率较高, 并且一旦开始反应便迅速生成数十 kb 以上的产物. 本研究中的一些现象也同以上解释相一致.

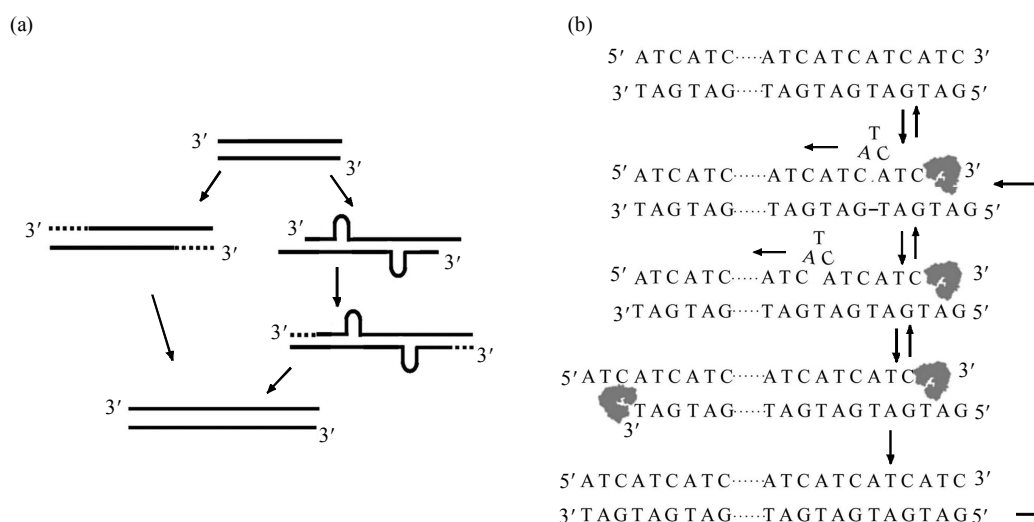


Fig. 5 Proposed mechanism for elongation of tandem repetitive sequence

(a) Mechanism proposed previously by Christian^[34]. (b) Mechanism we proposed. The (ATC)₆/(GAT)₆ was set as an example to explain this slippage synthesis model. A three base loop is formed and the polymerase binds quickly to the template/primer-like structure for primer extension. The unstable loop structure will slip and wave quickly in the duplex till it is released at the 5' end. The polymerase binds the new-born template/primer-like structure to carry out primer extension. Accordingly, the length of DNA is elongated by 6 nucleotides for every cycle.

对于双链三核苷酸重复序列的恒温扩展的机理, 研究者们多认可图 5a 所示的滑动模式. 但并没有实验证据和合理的理论解释予以支持. 也有人持怀疑态度, 认为当反应温度大大低于双链 DNA 的熔点时, 完全的链间滑动需要很高的活化能. 本研究发现, 几乎所有的三核苷酸重复序列都可随时间成线性的扩展关系, 即在一定链长范围内, 链长同扩展反应的速度无关, 而现有机理很难解释这一现象. 因此我们提出了如图 5b 所示的三核苷酸小环(loop)在双链中波动传递而产生滑动的模型, 探

索了其解释三核苷酸重复序列扩展的可能性. 有文献报道, 在大大低于 DNA 双链熔点的情况下, 少数碱基对仍然可以进行瞬间氢键断裂而部分解链的呼吸运动, 即在 DNA 双螺旋中不停地有碱基对打开^[35]. 在 DNA 双链的末端, 会有更多的碱基对容易打开^[36]. 由于重复序列的特点, 在以上的碱基对打开的过程中就可能发生滑移错位. 如果末端的 6 个碱基对打开并形成三核苷酸小环结构, 就会发生三碱基的滑移(图 5b). 特别是聚合酶会更快地同已经错位的呈模板 / 引物结构的 3' 端结合以防止形

成的三核苷酸小环重新回到常规双链状态。同时, 聚合酶会进行 DNA 合成以补齐新生成的缺口。这样生成的不稳定的三核苷酸小环会迅速滑动到另外一端也形成模板 / 引物结构, 并在聚合酶的作用下发生 DNA 合成反应。至此短链完成第一轮扩展, 增长了 6 bp。此后, 扩展后的短链会再次形成三核苷酸小环, 由聚合酶补齐缺口, 如此循环使得双链逐渐扩展。由于三核苷酸小环处于高能的不稳定状态, 在张力作用下它会很快传播到末端, 而且一旦形成缺口, 聚合酶也会迅速做出反应, 所以在每个扩展循环过程中, 三核苷酸小环的形成阶段是整个扩展反应的限速阶段。由于双链末端形成三核苷酸小环的速度不发生变化, 短链双链重复序列的 DNA 扩展呈线性, 这可以很好地解释我们的实验结果。当双链长度很长时, 如 1.0 kb 以上, 三核苷酸小环在链中左右滑移的几率增大, 达到末端的时间变长, 从而使链扩展反应的速度降低, 扩展反应偏离线性关系。当反应温度较高时, 也可能在末端有更多的碱基对打开, 产生更大的小环, 如六核苷酸小环。

生物体内含有很多简单重复序列, 如人的基因组中可达 3%, 在 PCR 等方法检测 DNA 时会发生非特异性的扩增从而产生假阳性, 找到抑制简单重复序列扩展的方法将有助于提高基因检测过程中的准确性。另外简单重复序列在基因组中的功能还没有弄清楚。越来越多的实验证明, 简单重复序列易于扩展并发生变异, 很可能是进化过程中的远古序列的残留。一些科学家提出了组成生物遗传物质的核酸起源于简单重复序列这一假说。在核酸分子的进化过程中, 正是由于重复序列的不稳定性, 容易发生变异和扩展等变化, 使得有意义的基因片段得以形成。尤其是当各种简单重复序列混在一起时, 可以高效地衍生出各种序列。无疑, 系统探寻它们扩展和变异的规律将有助于建立简单重复序列和基因组序列之间的关系。本研究对所有三核苷酸重复序列的扩展反应进行了初步研究, 拓展了核酸分子进化研究的思路。

本研究提出的重复序列扩展模型也有望帮助解释体内三核苷酸重复序列的非正常扩展。在 DNA 复制扩增过程中, 会发生 DNA 聚合酶脱落现象, 或在冈崎片段生成的过程中产生开放的双链末端, 这就有可能发生类似的滑动现象, 使三核苷酸重复序列扩展发生。有待进一步进行试验设计以检验这一模型发生的效率。

参 考 文 献

- [1] Hastings P J, Lupski J R, Rosenberg S M, *et al.* Mechanisms of change in gene copy number. *Nat Rev Genet*, 2009, **10**(8): 551-564
- [2] Levinson G, Gutman G A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*, 1987, **4**(3): 203-221
- [3] Takanori M. Origin of genomic DNA: Discussion from reverse-transcription and expansion of repetitive oligonucleotides. *Viva Origino*, 2003, **31**(2003): 46-61
- [4] Heidenfelder B L, Topal M D. Effects of sequence on repeat expansion during DNA replication. *Nucleic Acids Res*, 2003, **31**(24): 7159-7164
- [5] Margolis R L, Ross C A. Expansion explosion: new clues to the pathogenesis of repeat expansion neurodegenerative diseases. *Trends Mol Med*, 2001, **7**(11): 479-482
- [6] Thibodeau S N, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. *Science*, 1993, **260**(5190): 816-819
- [7] 高 焕, 孔 杰. 串联重复序列的物种差异及其生物功能. *动物学研究*, 2005, **26**(5): 555-564
Gao H, Kong J. *Zoological Research*, 2005, **26**(5): 555-564
- [8] 王化坤, 乔玉山, 姜晓鸣, 等. 拟南芥叶绿体 DNA 全序列微卫星分布规律的分析. *中国生物化学与分子生物学报*, 2006, **22**(10): 845-850
Wang H K, Qiao Y S, Lou X M, *et al.* *Chin J Biochem Mol Biol*, 2006, **22**(10): 845-850
- [9] 刘丽华, 王立新, 赵昌平, 等. 光温敏二系杂交小麦恢复系遗传多样性和群体结构分析. *中国生物化学与分子生物学报*, 2009, **25**(9): 867-875
Liu L H, Wang L X, Zhao C P, *et al.* *Chin J Biochem Mol Biol*, 2009, **25**(9): 867-875
- [10] 李新海, 袁力行, 李晓辉. 利用 SSR 标记划分 70 份我国玉米自交系的杂种优势群. *中国农业科学*, 2003, **36**(6): 622-627
Li X H, Yuan L X, Li X H, *et al.* *Scientia Agricultura sinica*, 2003, **36**(6): 622-627
- [11] Heale S M, Petes T D. The stabilization of repetitive tracts of DNA by variant repeats requires a functional DNA mismatch repair system. *Cell*, 1995, **83**(4): 539-545
- [12] Aquilina G, Hess P, Branch P, *et al.* A mismatch recognition defect in colon carcinoma confers DNA microsatellite instability and a mutator phenotype. *Proc Natl Acad Sci USA*, 1994, **91**(19): 8905-8909
- [13] da Silva E F, Reha-Krantz L J. Dinucleotide repeat expansion catalyzed by bacteriophage T4 DNA polymerase *in vitro*. *J Biol Chem*, 2000, **275**(40): 31528-31535
- [14] Ogata N, Miura T. Genetic information 'created' by archaeobacterial DNA polymerase. *Biochem J*, 1997, **324**(Pt 2): 667-671
- [15] Ogata N, Miura T. Creation of genetic information by DNA polymerase of the *Thermus thermophilus*. *Nucleic Acids Res*, 1998, **26**(20): 4657-4661
- [16] Ogata N, Miura T. Elongation of tandem repetitive DNA by the DNA polymerase of the hyperthermophilic archaeon *Thermococcus*

- litoralis* at a hairpin-coil transitional state: a model of amplification of a primordial simple DNA sequence. *Biochemistry*, 2000, **39**(45): 13993–14001
- [17] Liang X G, Jensen K, Frank-Kamenetskii M D, *et al.* Very efficient template/primer-independent DNA synthesis by 2004, thermophilic DNA polymerase in the presence of a thermophilic restriction endonuclease. *Biochemistry*, **43**(42): 13459–13466
- [18] Liang X G, Li B C, Jensen K, *et al.* Ab initio DNA synthesis accelerated by endonuclease. *Nucleic Acids Symp Ser(Oxf)*, 2006, **50**(1): 95–96
- [19] Liang X G, Tomohiro Kato, Hiroyuki A. Unexpected efficient ab initio DNA synthesis at low temperature by using thermophilic DNA polymerase. *Nucleic Acids Symp Ser*, 2007, **51**(1): 351–352
- [20] Liang X G, Tomohiro Kato, Hiroyuki A. Mechanism of DNA elongation during de novo DNA synthesis. *Nucleic Acids Symp Ser*, 2008, **52**(1): 411–412
- [21] Ohno S. Original domain for the serum albumin family arose from repeated sequences. *Proc Natl Acad Sci USA*, 1981, **78**(12): 7657–7661
- [22] Ohno S. Evolution from primordial oligomeric repeats to modern coding sequences. *J Mol Evol*, 1987, **25**(4): 325–329
- [23] Ogata N, Miura T. Creation of genetic information by DNA polymerase of the archaeon *Thermococcus litoralis*: influences of temperature and ionic strength. *Nucleic Acids Res*, 1998, **26**(20): 4652–4656
- [24] Alexander B, Kotlyar, Natalia B, *et al.* *In vitro* synthesis of uniform poly (dG)-poly (dC) by Klenow exo- fragment of polymerase I. *Nucleic Acids Res*, 2005, **33**(2): 525–535
- [25] G K, Kandala V R, Basuthkar J R. Fold-back structures at the distal end influence DNA slippage at the proximal end during mononucleotide repeat expansions. *Nucleic Acids Res*, 1999, **27**(19): 3851–3858
- [26] Hanaki K, Odawara T, Nakajima N, *et al.* Two different reactions involved in the primer/template-independent polymerization of dATP and dTTP by Taq DNA polymerase. *Biochem Biophys Res Commun*, 1998, **244**(1): 210–219
- [27] Nad'a R, Jaroslav K. Expansion during PCR of short single-stranded DNA fragments carrying nonselfcomplementary dinucleotide or trinucleotide repeats. *Mol Biol Rep*, 2003, **30**(3): 155–163
- [28] Lyons-Darden T, Topal M D. Effects of temperature, Mg^{2+} concentration and mismatches on triplet-repeat expansion during DNA replication *in vitro*. *Nucl Acids Res*, 1999, **27** (11): 2235–2240
- [29] Ji J, Clegg N J, Peterson K R, *et al.* *In vitro* expansion of GGC:GCC repeats: identification of the preferred strand of expansion. *Nucleic Acids Res*, 1996, **24**(14): 2835–2840
- [30] Wu M J, Chow L W, Hsieh M L. Amplification of GAA/TTC triplet repeat *in vitro*: preferential expansion of (TTC)_n strand. *Biochim Biophys Acta*, 1998, **1407**(2): 155–162
- [31] Tuntiwechapikul W, Salazar M. Mechanism of *in vitro* expansion of long DNA repeats: Effect of temperature, repeat length, repeat sequence, and DNA polymerases. *Biochemistry*, 2002, **41** (3): 854–860
- [32] Norio O, Hirofumi M. Elongation of repetitive DNA by DNA polymerase from a hyperthermophilic bacterium *Thermus Thermophilus*. *Nucleic Acids Res*, 2000, **28**(20): 3999–4004
- [33] Ogata N, Miura T. Genetic information 'created' by archaeobacterial DNA polymerase. *Biochem J*, 1997, **324**(Pt 2): 667–671
- [34] Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res*, 1992, **20**(2): 211–215
- [35] Krueger A, Protozanova E, Frank-Kamenetskii M D. Sequence-dependent base pair opening in DNA double helix. *Biophys J*, 2006, **90**(9): 3091–3099
- [36] Smolina I V, Demidov V V, Soldatenkov V A, *et al.* End invasion of peptide nucleic acids (PNAs) with mixed base composition into linear DNA duplexes. *Nucleic Acids Res*, 2005, **33**(17): e146

Expansion of Trinucleotide Repetitive Sequences and Its Mechanism*

WANG Yang, JIA Lei-Min, DONG Ping, LIANG Xing-Guo**

(College of Food Science and Engineering, Ocean University of China, Qingdao 266003, China)

Abstract Simple repetitive sequences are widely present in genomes of most organisms. They are closely related to molecular evolution, genetic diversity, molecular marker and some hereditary diseases. In this article, all of the 18 bp trinucleotide repetitive double strand sequences were used, and their expansion by DNA polymerase was researched systematically. We discussed the factors including reaction temperature and sequence which influence the amplification efficiency and the length of product. The result showed that almost all of the repeats can be elongated, and the variance in sequence had great effect on the amplification efficiency. The repeats with more GC, especially those with G and C in one of the strands were much easier to grow. Duplexes with one strand containing only A and C were amplified a lot as well. The optimum reaction temperature and the GC content of the duplexes had a positive correlation. Agarose and polyacrylamide gel electrophoresis showed that the product has narrow size distribution of molecules, and the expansion has a linear dependence of the length of products on the reaction time. Differentiation of the products occurred more at higher reaction temperatures. At last, a slippage model demonstrating the mechanism of duplex expansion was discussed, and it is promising to be used for explanation of molecular evolution and unusual expansion of repetitive sequences during gene amplification and detection.

Key words trinucleotide repetitive sequences, repetitive sequences, abnormal amplification, molecular evolution

DOI: 10.3724/SP.J.1206.2012.00308

*This work was supported by grants from National Youth Qianren Plan, Recruitment Programs of "Wanren Plan", Program for Changjiang Scholars and Innovative Research Team in University, and Fund for Distinguished Young Scholars of Shandong Province (JQ201204).

**Corresponding author.

Tel: 86-532-82031086, E-mail: liangxg@ouc.edu.cn

Received: June 21, 2012 Accepted: September 25, 2012