PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2014, 41(7): 693~703
www.pibb.ac.cn

四核苷酸重复序列单链扩增特性及机理探究 *

陶现明　王鹏飞　王　阳　梁兴国 **
(中国海洋大学食品科学与工程学院，青岛 266003)

摘要　简单重复序列广泛存在于多种生物基因组中，其生物学意义越来越受到人们的重视．许多简单重复序列易于扩增变长，某些重复序列的异常延伸是造成一些遗传疾病的直接原因．本研究以 20 nt 的 60 种四重复和 6 种二重复序列单链为模板，系统研究了它们在嗜热 DNA 聚合酶作用下等温扩增的特点．电泳结果显示，多数单链模板能扩增变长，即使链内没有互补碱基的序列也可被扩增，如$(AGGA)_5$．定量分析结果显示：回文序列扩增最快；二重复序列比相同碱基组成的四重复序列有更宽的适于扩增的温度范围；G 和 C 含量多的 DNA 较 G 和 C 含量少的序列更易扩增，而且 G 和 C 含量越多越适于在较高的温度下扩增；重复单位含两相同嘧啶的链多数比其互补链更易扩增；产物浓度与时间基本呈线性关系．限制性酶切产物结果显示，扩增产物与模板具有相同的重复单位，是重复序列的简单延伸．最后，根据实验结果和相关文献，提出了包括链内滑动扩增和发卡 DNA 介导扩增两阶段的重复序列单链扩增模型，以对重复序列非特异扩增和相关疾病发生机制的研究提供参考．

关键词　四核苷酸重复序列，等温扩增，非特异性扩增，分子进化，发卡结构
学科分类号　Q71，Q52　　**DOI**: 10.3724/SP.J.1206.2014.00002

串联重复序列由特定序列(即重复单位)首尾相连组成，也称为简单重复序列或微卫星 DNA，如$(A)_n$、$(AT)_n$、$(ATG)_n$、$(ATGC)_n$ 等．它广泛存在于多种生物基因组中，表现出种属、长度和序列的多态性[1-3]，并与某些遗传性疾病的发生、基因调控和分子进化等密切相关．利用其在不同种属及个体中的差异，串联重复序列分析还可用于遗传标记[4-5]．研究发现，某些三核苷酸重复序列的异常扩展会引起Ⅰ型强直性肌营养不良(DM1)、亨廷顿舞蹈症(HD)、脊髓和延髓的肌肉萎缩症(MD)及脊髓小脑共济失调(SCA)等至少 17 种遗传疾病[6-8]；$(CCTG/CAGG)_n$ 的异常变化与Ⅱ型强直性肌萎缩症(DM2)密切相关[9]．

DNA 从头合成能够生成一些简单重复序列[10-11]，支持重复序列的合成先于基因，基因组源于重复序列扩增的有关分子进化的说法．CRISPR/Cas 系统赋予了细菌和古细菌对入侵病毒和质粒的适应性免疫能力[12-14]，其中短回文重复序列是 CRISPR 的必要组成部分，显示了重复序列对原始生命的重要性．$(GATA)_n$ 在人和果蝇的细胞中具有增强子阻断活性[15]，并在人的 Y 染色体中含量丰富，参与了某些生物的性别调控[16]．Stevens 等[17]发现人细胞提取物能促进$(CTG)_{22}$ 等三核苷酸重复序列的延伸．

此外，DNA 等温扩增检测特定 DNA 序列时，重复序列导致的异常扩增易使实验产生假阳性结果[18]，使得重复序列扩增机制的研究成为必要．Ogata 等[19-21]分析了短链回文重复序列的扩增特点，认为形成发卡和分子间滑动分别在扩增初期和后期发挥重要作用．梁兴国等[22-23]发现了内切酶可大大促进 DNA 从头合成的现象，并提出了 Cut-Grow 模型．王阳等[24-25]研究了三重复序列的扩增现象，得出含 GC 多的序列更易扩增且高温下扩增较快等

* 山东省万人计划，山东省自然科学杰出青年基金(JQ201204)，国家青年千人计划，长江学者和创新团队发展计划(IRT1188)资助项目．
** 通讯联系人．
Tel: 0532-82031086, E-mail: liangxg@ouc.edu.cn
收稿日期：2014-01-03，接受日期：2014-05-16

结论，进一步丰富了双链滑动扩增的机制.

虽有文献报道了某些四重复双链的扩增特点[26]，探讨了序列对重复序列扩增的影响，但对四重复单链扩增特点的系统研究和扩增机制的研究很少. 本文对 60 种四重复和 6 种二重复的单链 DNA 进行等温扩增，分析了四重复序列产物分子质量和浓度随碱基序列、温度和时间变化的规律，并初步确定了产物的序列，提出了新的单链重复序列的扩增机理.

1 材料与方法

1.1 材料

60 种 20 nt 四重复 DNA 和 6 种 20 nt 二重复 DNA 用于系统研究四重复单链的扩增(购自 Integrated DNA Technologies 公司)；Vent (exo-) DNA 聚合酶、DNA Maker、限制性内切酶等购自 New England Biolabs；SYBR Green Ⅰ 购自 Invitrogen 公司；鲑鱼精 DNA 购自 Sigma 公司. 四重复单链模板的命名方法如下：以重复单位代指整个序列，如 TAAA 指代$(TAAA)_5$(为统一命名，以 ATAT 指代$(AT)_{10}$).

1.2 单链模板的恒温扩增与产物酶切

1.2.1 单链模板的恒温扩增

将反应溶液加入 200 μl EP 管中，置于 PCR 仪中恒温反应一定时间后定点取样. 反应体系：总体积 20 μl, 100 nmol/L 单链模板，20 U/ml 聚合酶，0.5 mmol/L dNTPs，1×Thermopol Buffer(20 mmol/L Tris-HCl, 10 mmol/L $(NH_4)_2SO_4$，10 mmol/L KCl，2.0 mmol/L $MgSO_4$，0.1% Triton X-100，pH 8.8@25℃). 反应温度各为 50℃、60℃、70℃、80℃.

1.2.2 产物的酶切

将纯化后产物加入相应的酶切体系：总体积 50 μl，纯化产物约 30 mg/L，适量内切酶(50～100 U/ml)，相应的 1×缓冲液. 反应一定时间后分次取样，每次取 10 μl，然后加入 0.2 倍体积 6×洗脱缓冲液终止反应.

1.3 反应产物的电泳分析

将 5.0 μl 反应产物用 0.8%琼脂糖凝胶(酶切产物用 2%琼脂糖凝胶)电泳 50 min 左右. EB 染色约 15 min 后，置于伯乐凝胶成像仪(Gel Doc XR+)上成像分析.

1.4 SYBR Green I 荧光定量扩增产物

配制鲑鱼精 DNA 梯度浓度溶液，用 1×SYBR Green Ⅰ 避光染色 15 min. 用 Thermo Scientific Varioskan Flash 全波长多功能酶标仪测定荧光值，激发波长 497 nm，发射波长 520 nm. 实验表明，双链 DNA 浓度在 0.8～431.5 μg/L 浓度范围内与荧光值呈良好的线性关系，r^2=0.9993. 本实验将产物用 1×TE 缓冲液稀释 20～1000 倍后，使荧光值保持在线性范围之内，据标准曲线计算产物浓度.

1.5 短链熔点测定及相关参数计算

模板对应双链 T_m 测定：2.0 μmol/L双链模板(互补单链各 2.0 μmol/L 或 4.0 μmol/L回文单链)，10 mmol/L 磷酸盐(Na_2HPO_4-NaH_2PO_4)，100 mmol/L NaCl，pH 7.1@25℃，使用岛津 UV1800 测定样品在 30℃～90℃的吸光度值，计算出 T_m. 用 DINAMelt Web Server 软件计算双链模板、80 bp 和 96 bp 串联重复双链的 T_m 及单链模板自身折叠的相关参数.

2 结 果

文献报道，具有相同重复单元的 3 种三重复单链具有相近的扩增特性[24]，如$(ATG)_6$、$(TGA)_6$ 和$(GAT)_6$ 除末端碱基不同外，内部序列完全相同，其扩增效率以及对温度的依存性都很接近. 20 nt 四重复单链共 240 种，考虑重复单元相同的序列，如$(ACGA)_5$、$(CGAA)_5$、$(GAAC)_5$ 和$(AACG)_5$ 可能有相似的扩增特性，我们选 4 种序列中的一种，$(ACGA)_5$(简称 ACGA)进行实验，这样共有 60 种序列. 另外有 6 种 20 nt 二重复序列，包括$(GA)_{10}$(简称 GAGA)，$(CT)_{10}$(简称 CTCT)，$(CA)_{10}$(简称 CACA)，$(TG)_{10}$(简称 TGTG)，$(AT)_{10}$(简称 ATAT)，$(GC)_{10}$(简称 GCGC)也用于扩增实验. 主要分析了不同序列产物长度的分布，碱基组成及顺序、T_m、温度和时间与扩增特点的联系等.

2.1 扩增产物分子质量随温度及时间的变化

我们对所有序列的扩增产物都进行了电泳分析. 图 1 为几种有代表性的序列的产物分子质量随温度和时间变化的电泳分析结果. ATAT 和 GCGC 为二重复回文序列，CACA 为二重复非回文序列，GCTA 为四重复回文序列，CAGC 为四重复非回文序列. 如图 1 所示，扩增产物多呈现弥散条带，温度对扩增效率产生很大影响. 50℃时 ATAT 扩增效率很高，CACA 在 16 h 有很长产物生成；60℃时除 ATAT 外其余序列扩增明显增加，尤其是 GCTA 有大量 DNA 生成；70℃和 80℃时 ATAT的扩增进一步减弱，其他所选序列都有明显扩增. 结合其他序列的扩增产物分析可以得出如下结果：非回文序列扩增较慢，一般只在 16 h 时扩增明显，

且产物长度多大于 10 kb；回文序列比非回文序列扩增更快，1 h 反应后即有大量产物生成，分子质量随时间呈逐渐变大，且在产物量较少时呈现明显较窄的条带；ATAT 低于 70℃时扩增较多，GCGC 高于 70℃时扩增较多．特别地，在 60℃和 80℃下，CACA 扩增产物的长度有随时间(3 h 内)呈线性增长的趋势．

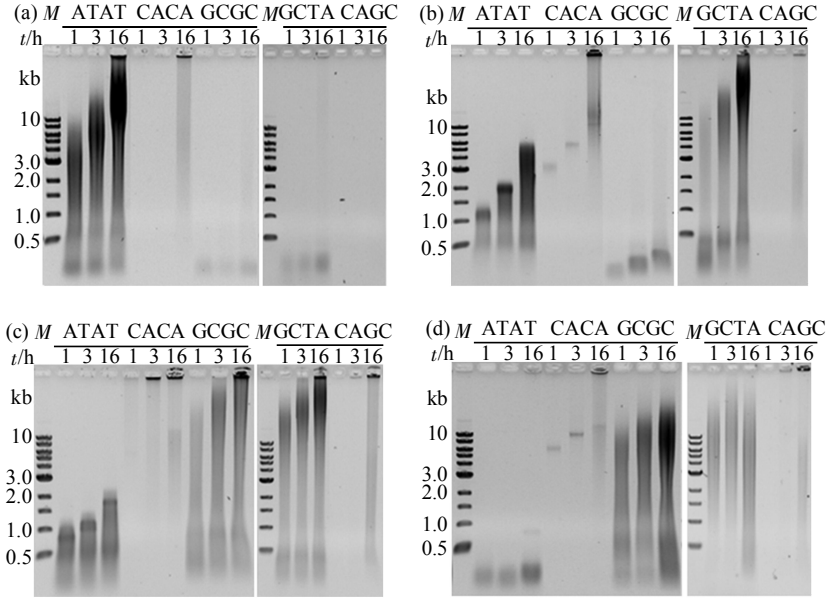

Fig. 1　Agarose gel electrophoresis analysis of expansion products from five short sequences

Reaction was performed at 50℃ (a), 60℃ (b), 70℃ (c), and 80℃ (d) for 1.0, 3.0 and 16 h, respectively. Reaction conditions: 100 nmol/L short repeats, 20 U/ml Vent, 0.5 mmol/L dNTPs, in 20 μl of 1×Thermopol Buffer (20 mmol/L Tris-HCl, 2.0 mmol/L $MgSO_4$, 10 mmol/L KCl, 10 mmol/L $(NH_4)_2SO_4$, 0.1% Triton X-100, pH 8.8 (25℃)). Analyzed by 0.8% agrose gel.

2.2　16 h 产物的定量分析

我们采用扩增产物染色后测定荧光的方法进行了定量分析．由于大多数序列在 1 h、3 h 的产物生成量较少，本文只分析了 4 个温度下 16 h 的产物浓度(表 1)．为了找出扩增效率与温度的关系，对短链 DNA 的 T_m 进行了测定；同时假定扩增产物是与模板具有相同重复单位的串联重复双链 DNA，利用相关软件(The DINAMelt Web Server)计算长链产物的 T_m(表 1)．首先考察序列对单链模板扩增特点的影响．

2.2.1　碱基组成及顺序对扩增能力的影响

由表 1 可见，斜体部分 18 种序列在 4 个温度下扩增都很弱，其余序列至少在一个温度下扩增明显(10 mg/L 以上)，从重复单位碱基成分和顺序具体分析如下：

a．A～F 组含 3 个相同碱基，绝大多数情形下不易扩增，其中包括在灵长类和啮齿类基因组外显子中含量丰富的 AAAB(A 为腺嘌呤，B 为除 A 外的碱基)类重复序列[3]．特例是 F01 (CCGC)，可在 80℃下大量扩增．

b．G、H 组为只含两种不能形成 Watson-Crick 碱基对的碱基(如 AGGA)组成的非回文序列，一般都可大量扩增．二重复单链比四重复单链最适扩增温度略低，前者为 60℃，后者为 70℃．同碱基组成的二重复单链比四重复单链能在更宽的温度范围内扩增，如 H1 较 H2 在各个温度扩增更多．70℃时四重复序列产物浓度排序为：AGGA>CCAA>CCTT>GTTG.

c．I、J 组(A、T 占 3/4)与 M、N 组(G、C 占 3/4)相比扩增较弱，特别是重复单位含两个 A 的序列很难扩增，只有 J31(CAAT)在 60℃时有明显扩增．含两个 T，尤其两个 T 不相邻时扩增较强．M、N 组整体扩增很强，只有 M12(GCTC)稍弱．

d．K、L 组为 GC 含量为 50%且含一对配对碱基(AT 或 GC)的非回文序列，除 K31(ATGG)外都有较强扩增．对含 A 和 T 的 K 组，间隔 G 或连续

Table 1 T_m Values of short and long repeats and concentrations of products amplified at four temperatures

T_m/(℃)			No.[3]	Seq	ρ(Products)/(mg•L^{-1})[4]				No.	Seq	ρ(Products)/(mg•L^{-1})			
20 bp Mea	20 bp Cal[1]	Prod Cal[2]			50℃	60℃	70℃	80℃			50℃	60℃	70℃	80℃
41.0	41.9	60.5	*A01*[5]	*TAAA*	0.2	0.2	0.2	0.1	*A02*	*TATT*	0.2	0.2	0.2	0.2
48.3	52.9	72.0	*B01*	*GAAA*	0.1	0.1	<0.1	0.1	*B02*	*CTTT*	<0.1	0.3	0.1	0.1
55.7	55.6	74.2	*C01*	*ACAA*	<0.1	<0.1	<0.1	<0.1	*C02*	*GTTT*	<0.1	0.5	3.3	0.2
66.9	69.9	90.6	*D01*	*AGGG*	<0.1	<0.1	0.1	0.1	*D02*	*CTCC*	<0.1	<0.1	4.6	0.1
72.4	72.0	92.4	*E01*	*CCAC*	<0.1	<0.1	0.7	0.1	*E02*	*GGGT*	<0.1	<0.1	0.2	1.9
82.6	83.1	102.05	*F01*	*GGCG*	<0.1	<0.1	0.1	1.4	F02	CCGC	0.1	0.1	37.4	**118**[6]
59.2	60.4	78.5	G11	GAGA	2.6	**24.7**	21.5	5.5	G12	CTCT	24.6	**71.7**	41.1	1.0
60.0	62.9	81.2	G21	AGGA	2.1	0.4	**113**	1.2	G22	CCTT	0.1	8.5	**62.2**	1.1
67.5	64.9	82.4	H11	CACA	15.8	**44.1**	22.6	12.1	H12	TGTG	2.6	**35.7**	30.1	4.4
64.4	63.7	83.2	H21	CCAA	<0.1	0.1	**79.9**	2.1	H22	GTTG	0.2	<0.1	**20.4**	0.6
48.8	48.6	67.3	*I11*	*TAGA*	6.4	7.2	0.3	0.6	I12	CTAT	**11.1**	7.1	0.2	0.1
52.5	51.8	69.4	*I21*	*AGTA*	1.9	6.7	2.2	0.6	*I22*	*TACT*	0.3	0.1	0.1	0.2
53.7	54.1	72.4	*I31*	*ATGA*	0.3	1.1	0.1	0.2	I32	CATT	0.1	**14.4**	0.1	0.1
51.2	50.5	69.1	*J11*	*CATA*	0.1	1.1	4.3	0.2	J12	GTAT	**56.4**	19.1	0.2	0.2
50.5	50.1	69.3	*J21*	*CTAA*	<0.1	0.6	0.5	<0.1	J22	AGTT	**23.7**	15.1	0.3	0.2
54.4	53.7	72.2	J31	CAAT	0.1	**16.2**	0.2	0.1	J32	TTGA	1.9	**25.5**	4.5	0.3
61.4	61.0	80.2	K11	AGTG	0.3	**34.8**	16.2	4.9	K12	CTCA	<0.1	15.2	**50.6**	0.1
60.0	59.6	78.5	K21	AGGT	1.3	**11.8**	3.2	0.4	K22	CCTA	0.4	5.0	**91.3**	0.4
63.8	62.5	81.6	*K31*	*ATGG*	0.2	0.2	8.8	2.1	K32	CCAT	2.6	42.5	**124**	60.6
61.3	61.5	80.2	L11	CAGA	0.1	**12.5**	2.0	0.1	L12	CTGT	10.3	15.6	**50.9**	0.2
67.4	65.3	81.8	L21	ACGA	2.7	31.2	**58.9**	12.1	L22	TCGT	0.2	16.9	**288**	3.1
67.2	66.0	84.0	L31	AGCA	0.1	0.3	13.8	**23.1**	L32	GCTT	4.9	35.6	43.8	**49.5**
77.7	72.7	91.0	M11	CGAG	0.1	0.2	13.3	**41.3**	M12	GCTC	3.5	**12.5**	3.4	1.2
73.9	72.4	90.5	M21	GACG	0.4	23.3	**156**	44.2	M22	CGTC	0.1	0.1	**40.8**	33.6
73.4	73.0	92.9	M31	CAGG	1.0	12.1	29.1	**33.9**	M32	CTGC	4.7	3.6	**219**	42.7
75.3	72.1	90.3	N11	CGAC	0.2	0.5	**83.4**	21.1	N12	TCGG	7.1	29.4	**178**	33.8
80.2	74.7	92.4	N21	ACGC	4.6	23.1	112	**312**	N22	TGCG	23.9	22.6	245	**309**
75.8	73.8	92.9	N31	CAGC	0.4	5.2	**128**	15.1	N32	GCTG	0.2	2.5	20.5	**67.9**
63.8	61.9	80.2	O11	CAGT	1.7	**40.9**	30.8	8.7	O12	GACT	0.7	**47.8**	35.6	12.9
			76.5	66.6	82.8	O20	ACGT	33.9	90.5	**253**	56.2			
			72.4	62.4	82.9	O30	GCTA	2.0	198	**263**	170			
			70.3	65.3	83.1	O40	CGAT	**4.4**	21.6	**262**	223			
			78.2	71.0	88.1	O50	TGCA	3.2	52.0	**229**	130			
			48.1	43.1	56.0	P10	ATAT	224	**242**	52.0	10.6			
			43.9	47.3	61.6	P20	TAAT	64.6	72.4	**104**	5.1			
			81.0	85.8	104	Q10	GCGC	2.8	14.7	**336**	270			
			86.8	84.6	105	Q20	GGCC	0.8	1.2	5.4	**159**			

[1] The conditions of calculated T_m of 20 bp dsDNA were the same as that of measurement: 116.7 mmol/L NaCl, 2.0 μmol/L DNA. [2] The T_m of 96 bp dsDNA was calculated using the conditions of DNA amplification: 10 mmol/L NaCl, 2 mmol/L magnesium ion, 0.1 μmol/L DNA. [3] The number of each sequence was designated according to the kinds of nucleotides, ascending T_m of the 96 bp repetitive sequences. The left sequences were complimentary with the right ones. Take group I for example, TAGA, AGTG and ATGA, composed of the same bases, were classified as group I-1 and numbered as I11, I12, and I13 respectively, in consideration of their 96 bp T_m. [4] Products concentration of sequences were the mean of two repeated experiments, the relative error was less than 15%. [5] The data of products concentration of italic No. and corresponding sequences are less than 10. [6] The bold data, more than 10, are maximum of products concentration of corresponding sequences under four temperatures, reflecting the most suitable temperature for amplification.

C 有利扩增；对含 G 和 C 的 L 组，连续 A 或连续 T 有利扩增，即 G 和 C 相邻时更易扩增．如 L22 (TCGT)的扩增能力强于 L12(CTGT)，这与文献报道的双链扩增结果一致[26]．可以认为，两个互补的碱基(A 和 T 或 G 和 C)相邻时一般更易扩增．

e．O、P、Q 组中，回文序列(包括近似回文序列 O30、O40、P20)扩增很强，非回文序列(O11 和 O12)扩增相对弱一些．除 ATAT 适于 50℃或 60℃，GGCC 适于 80℃外，其他序列最适扩增温度为 70℃．二重复的 ATAT 较四重复的 TAAT 更易扩增．TAAT 的最适扩增温度高于 ATAT，可能由于 5′-AA/TT-5′比 5′-TA/AT-5′更稳定[27]，表 1 也显示 TAAT 的长链产物 T_m 更高．

f．从 I 至 O1 组数据考察逆向序列(如 AGTT 和 TTGA，只是 A 和 G 颠倒)，大多具有相同的扩增特征，最适扩增温度也相近(10℃以内)．

此外，与疾病相关的序列，如含 CTG、CAG 的四核苷酸重复序列[9]，除 L11(CAGA)外，大多具有较强的扩增能力．具有增强子阻断活性的 GATA[15]扩增很弱．

2.2.2 互补序列扩增特点的比较

表 1 中左右两排序列的重复单元互补，如 TAGA (I11) 与 CTAT (I12)，可形成……TAGATAGA……/……TCTATCTA……这样的双链．对于可以顺利扩增的序列，可以发现具有以下特点：a．互补序列最适扩增温度相近(差值在 10℃之内)；b．除个别序列外(如 H1 和 N2)，大部分互补序列最大产物浓度相差 1 倍以上；c．当单链模板重复单位中含两 C 或两 T 时，包括 G1、H 组(含两 C 的更强)和 I3 等 17 对，扩增能力往往强于其互补链(含两 G 或两 A)，即最大扩增量更多或适于扩增的温度范围更宽．当有连续的相同碱基时，这种现象尤其明显．相反的情况有 G2、M1、M2、N1 和 N2 等 5 对，如 M1、M2 中含两 G 序列扩增能力远强于含两 C 的序列，这可能因为错配 AG 比错配 TC 更稳定[27]．

综合 **2.2.1** 至 **2.2.2** 可知：回文序列扩增最强；二重复非回文序列扩增普遍较强，比同碱基组成的四重复序列有更宽的适宜扩增的温度范围；对于四重复非回文序列而言，从重复单位上看，含有较多 AT(3/4)或 3 个相同碱基的序列扩增最弱，含有较多 GC(3/4)的序列扩增较强，GC 与 AT 各半且含一对配对碱基的序列总体扩增能力居中；互补序列最适扩增温度相近，但产物浓度有一定差别，重复单位中含两 C 或两 T 的序列比其互补序列扩增更强．

接下来我们进一步考察了序列相关的 T_m、温度和时间与单链模板扩增特点的联系．

2.2.3 T_m 与单链扩增量的关系

以 DNA 为模板在 DNA 聚合酶作用下合成的 DNA 都会形成双螺旋结构，因此等温扩增过程中双螺旋打开的难易程度对扩增过程有较大影响．虽然起始的寡核苷酸是单链，但在扩增过程中会形成双螺旋结构，因此相应重复序列双链的 T_m 应该和扩增的难易有很大关系．根据扩增结果总结出以下规律：

a．四个温度下都很难扩增的序列(斜体部分，产物浓度小于 3 mg/L 的 12 种：A0、B0、C01、D01、E0、F01、I22、I31 和 J21；在 3～9 mg/L 之间的有 6 种：C02、D02、I11、I21、J11 和 K31)的 T_m 大多太低或太高，前者如 A0、B0、I11 和 I2-等：短链 T_m 在 41~56℃，长链 T_m 在 75℃以下．G、C 含量较高的 D0、E0 和 F01 等，长链 T_m 都在 90℃以上．

b．至少在一个温度下扩增明显的序列中，大多数短链 T_m 与最适扩增温度(加粗数据对应的温度)相差 10℃以内．例外情况是，M12(GCTC)的最适扩增温度较低，而 P20(TAAT)的 T_m 较低，但在 70℃更易扩增．有意思的是最适扩增温度大多低于长链 T_m，有 21 种序列长链 T_m 高于其最适扩增温度约 10℃，如 G2、H2 和 I32 等．21 种序列长链 T_m 高于最适扩增温度 20℃左右，如 G1、H1、F02、J12、Q10 和 M12 等．总体而言，短链 T_m 10℃上下或低于长链 T_m 10℃～20℃的温度有利于序列的扩增．对于 I～N 组的序列，随着 G、C 含量的增多，T_m 逐渐升高，最适扩增温度也逐渐升高；而且适于扩增的温度范围逐渐变宽，最大扩增量逐渐增多．这与三核苷酸单链重复序列扩增特点相似[24]．

2.2.4 16 h 产物浓度统计分析

为对所有模板总体扩增情况进行分析，将 16 h 产物浓度由低到高不等分为 5 组：几乎不扩增组 n (0～1.0 mg/L，almost no detection)，低量组 l (1.0～10 mg/L, low yield)，中量组 m(10～50 mg/L, middle yield)，高量组 h(50～150 mg/L，high yield)，超高量组 sh(> 150 mg/L，super high yield)．n、l 组合称泛低量组 L，h、sh 组合称泛高量组 H．考虑到回文与否和重复单位长短的区别，将序列类别细分为四重复非回文(nonpalindromic TSR，54 种)、四

重复回文(palindromic TSR，6 种)、二重复非回文(nonpalindromic DSR，4 种)及二重复回文(ATAT 和 GCGC). 据表 1 中数据绘制成统计图如图 2；为直观地表现产物浓度随温度的分布，作 L、m、H 组序列数 - 时间曲线图，如图 3.

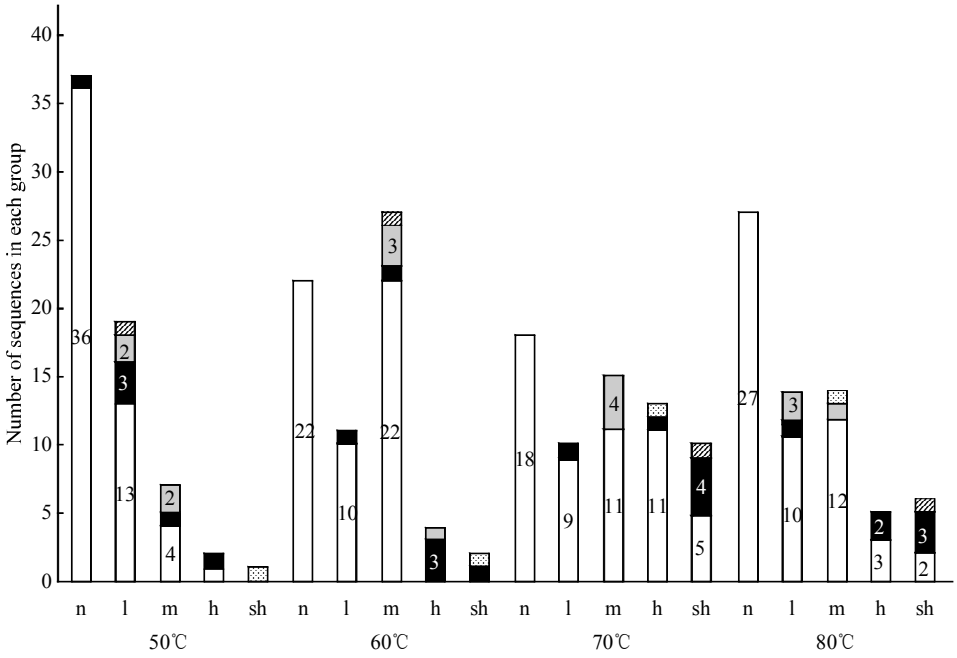

Fig. 2　Statistical histogram of the amplification of repetitive sequences

□: Nonpalindromic TSR; ■: Palindromic TSR; ▨: Nonpalindromic DSR; ⊠: ATAT; ▨: GCGC.

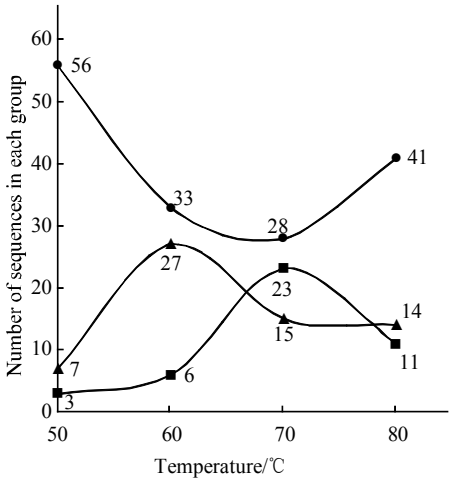

Fig. 3　Variation of the number of sequences of L, m and H amplification efficiency along with temperature

●—●: L; ▲—▲: m; ■—■: H.

a．五浓度组序列数随温度的分布

由图 2 可以看出，比较其他温度，70℃ 时 n 和 l 组的序列数目最小(分别 18 和 10)，而 h 和 sh 组的序列数目最多(分别为 13 和 10)，这表明大多序列 70℃ 时更易扩增．一般耐热性 DNA 聚合酶的最适温度在 70℃～80℃ 之间，可以看出 80℃ 时，相对于 50℃ 和 60℃ 也更容易扩增一些.

b．三浓度组序列数随温度的变化

由图 3 可见，随温度升高，三个浓度组序列数基本呈现抛物线状态：L 组序列数先减后增，70℃ 时最少；m、H 组先增后减，分别在 60℃ 和 70℃ 组达最高点．L 与 H 组序列数随温度升高此消彼长，也直观地显示 70℃ 适于大多数序列的扩增，温度过高或过低都不适宜.

c．不同序列类型扩增能力不同

由图 2 看出，回文序列扩增普遍很强．对于四重复回文序列，多数不适于在 50℃ 下扩增，但多半可在其他温度下大量扩增．对于二重复回文序列(ATAT、GCGC)，除 50℃ 时 GCGC 扩增较少外，在各温度下扩增能力普遍较强．而非回文序列扩增能力不一，二重复较四重复非回文序列整体扩增能力稍强，这跟 **2.2.1** 中的分析结果相似．对于四重复非回文序列而言，50℃ 时，除 J12 (GTAT) 外产

物量都在中等(m)或更低的水平；60℃时没有序列的产物量达到泛高量组(H)水平；70℃时有5种(GACG，TCGT，CTGC，TCGG，TGCG)，80℃时有2种(TGCG，ACGC)序列扩增到150 mg/L以上．再结合表1可以看出，含有TGC(GCA)，TCG(CGA)或CGT(ACG)的序列容易扩增．

2.2.5　五条序列70℃产物浓度随时间的变化

为考察单链扩增的速度，按照16 h产物浓度由高到低选取了五条序列(CGAT，CAGC，CCAT，CGAC，ACGA)，测定了它们在70℃下扩增1 h、3 h、8 h、16 h、24 h的产物浓度，结果如图4所示．

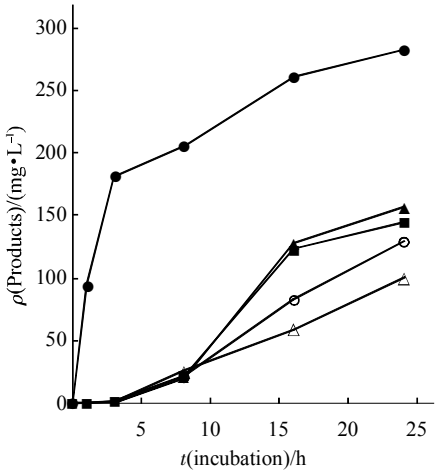

Fig. 4　Variation of products concentraton along with reaction time at 70℃

●—●: CGAT; ▲—▲: CAGC; ■—■: CCAT; ○—○: CGAC; △—△: ACGA.

回文序列CGAT扩增很快，在1 h已有大量扩增，3 h内基本完成50%以上的扩增，3 h内产物量也随时间基本呈线性增长．8 h后增长缓慢，可能由于dNTPs基本消耗较多以及生成的焦磷酸阻碍反应进一步进行．其他四条序列经过一个3 h或更长一些的无明显扩增阶段，然后反应产物呈线性扩增趋势．值得关注的是，对于相同碱基组成的序列CAGC和CGAC，前者扩增明显较快．对于更难扩增的序列也有类似特点，即先经过数小时或更长时间的潜伏期，然后出现较快增长，说明扩增的起始阶段是扩增反应的律速阶段．

2.3　部分序列扩增产物酶切分析

我们还采用限制性内切酶处理的方法分析了扩增产物的序列，考察扩增产物是否为主要含有起始模板寡核苷酸的重复序列．如果扩增产物都由限制性内切酶的识别序列组成，酶切后产物断片会很小，否则应该含有其他序列．GCGC、CGAT、CTCC扩增产物的酶切结果如图5所示．

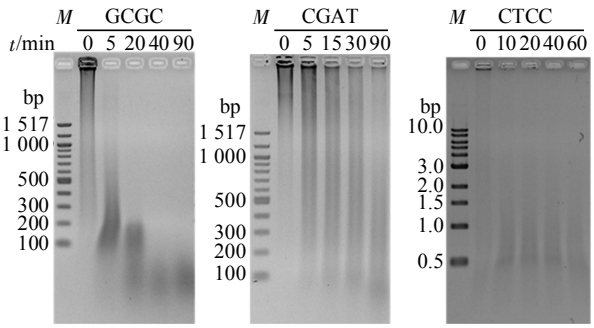

Fig. 5　Electrophoresis analysis of the digestion of amplification products by restriction enzyme

The products of GCGC were digested by Bsh1236I (CG↓ CG). The products of CGAT were digested by BsiEI (CGRY↓ CG). The products of CTCC were digested by Mn1I (CCTC(N)$_7$↓).

可以看出，扩增产物都可以被酶切，酶切产物的片段大小随时间的延长而逐渐变短．GCGC最明显，5 min时大部分产物已在500 bp以下；CGAT产物分子质量减小较慢，最终产物呈很大范围的弥散条带，可能因为分子内或分子间易形成复杂结构阻碍了酶切反应．CTCC由于产物较少，条带比较微弱，但也显示出了相似的酶切效果．此外AATT和GCTA等序列与GCGC产物酶切电泳图也很相似(未列出)．由此可推断，重复序列单链扩增产物基本上是相同重复单位串联而成的双链DNA．

3　讨　　论

实验显示，回文序列与非回文序列具有不同的扩增特点．回文序列比非回文序列扩增快且产物分子质量比较集中，应该由于回文单链容易形成分子内配对而很快启动扩增，同时每个分子都易扩增使得序列延伸具有同步性．非回文单链需形成含错配的3′发卡结构启动扩增进而形成大分子，错配不稳定使得初始扩增的分子具有随机性，错配积累慢、延伸快及扩增不同步等性质应该是非回文序列起始扩增慢和条带弥散的原因．CAGC等序列70℃下产物浓度变化曲线表明，扩增起始阶段制约着非回文序列的扩增速度，反映了初始错配积累的困难．相反，回文序列由于序列互补的特点，初始

扩增很快.

短链 T_m 10℃上下或低于长链 T_m 10℃～20℃的温度有利于短链模板的扩增，可作如下解释．在实验温度下，短链 T_m 太低则不易形成双链结构而难以引发扩增，太高又使双链部分不易打开进行下一步扩增．长链产物的 T_m 高于实验温度时，可以保持较为稳定的双链结构，同时又可以通过链内滑动稳定增长．互补序列的最适扩增温度相近应归因于产物序列的相似性，而最大产物浓度的差别可能由于初始发卡形成的难易不同．多数情况下，含 2 个相同嘧啶的序列比其互补序列扩增更强，可能因为嘌呤碱基太大形成空间位阻而使双链不稳定或聚合酶更难结合.

碱基组成和顺序都影响序列的扩增特性．多的 G、C 利于大部分非回文序列的扩增．这可能由于 G•C 的存在利于双链结构的形成，尤其 G、C相邻时；还可能因为含 G 的错配比较稳定[27]，利于初始发卡结构的形成，如 G•T、G•A．TAAT 的最适扩增温度高于 ATAT，归因于前者碱基排列方式提高了双链的稳定性.

16 h 产物浓度统计分析表明，四重复序列整体扩增能力不强，大多数更适合在 70℃ 下扩增．这一方面可能由于序列本身的特性，如错配引发扩增的难易、双链 T_m 的高低等，还可能由于 Vent(exo-) DNA 聚合酶在 75℃ 具有最高活性.

一般认为单链 DNA 可以扩增变长是由于 3′端形成发卡结构作为引物而引发延伸反应，这是简单重复序列扩增的起始阶段．我们的结果也显示，重复序列的起始阶段是扩增反应的律速阶段．据此，同时参考 CAGG 单链结构特点[9]、双链滑动扩增原理[25]，并利用 DINAMelt Web Server 计算相关结构的参数，以 CAGG 为例提出非回文序列等温扩增过程(图 6).

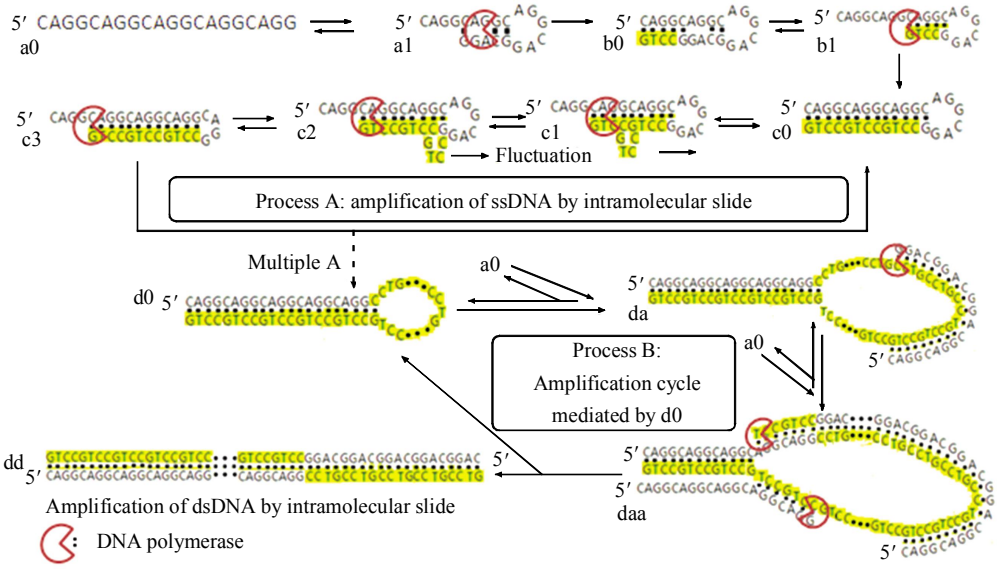

Fig. 6　Mechanism of the amplification of single-stranded tetranucleotide repeats

Process from a0 to d0 was amplifiction by intrachain slip. Process from d0 to dd and d0 was amplification mediated by DNA containing hairpin structure.

如图 6 所示，单链模板 a0 的 3′端形成含错配的小发卡分子 a1，结合聚合酶而扩增为 b0，这一步由于存在错配而较慢．b0 双链部分不够稳定易解链再结合形成 b1，进而延伸为 c0．c0 类分子双链部分较长不易解链时，会以末端波动传递造成链内滑动的形式进行延伸，每一过程至少伸长一个重复单位，这一过程记为 A．最后生成含有大发卡结构的 d0 类分子．由于 d0 的环部分不易形成双链[9]且与单链模板互补，它会与模板 a0 杂交形成分子 da；da 结合聚合酶而延伸，后期逐渐置换下 d0 的 3′端，此末端可与模板 a0 进一步杂交形成中间分子 daa；daa 分子的延伸形成双链分子 dd，同时置

换下 d0 类分子．dd 类双链分子可通过链内滑动扩增逐渐变长．由 d0 类分子介导的生成 dd 类和 d0 类分子的循环过程记为 B．实际反应时，过程 B 中 a0 与 d0 的杂交位置应该很多，图中只阐述了较为简单的情况．

b0 的形成由于存在错配而较慢，并且具有随机性．由 c 类分子经过程 A 生成 d0 类分子需要较长时间的温育．一旦 d0 类分子形成后，会进行以 d0 为模板，以 a0 为引物的 B 循环，生成数量可观的大分子．这可以解释非回文序列初始扩增很慢，而温育较长时间后生成大量大分子的现象．此外，过程 B 中 a0 与 d0 分子结合位置的多样性可能导致弥散条带的形成．

此模型中存在一定的问题，即过程 B 并未使 d0 类的分子增长很多，大分子的形成最终依赖于 dd 类双链分子的链间滑动．受限于链间滑动本身的低速性，不能很好地解释某些模板 1 h 内大分子的形成．重复序列的扩增变长机理还有待通过设计相应实验，作进一步深入研究，如对产物进行测序，获得 dd 类分子的具体序列．

4 结　论

本实验对 60 种四重复和 6 种二重复单链的恒温扩增性质进行了研究，发现大部分序列在实验温度下能明显扩增，特别是含两对相同碱基的四重复非回文单链仍能有效扩增(如 AGGA)．所得主要结论有：a．除回文序列在某些温度下的产物分子质量较为集中外，其他情况下产物大多呈现弥散条带．b．影响产物浓度的因素有碱基组成、GC 含量、反应温度等．回文序列扩增最强，二重复非回文序列扩增较强，含有三个相同碱基的序列(除 CCGC 外)扩增都较弱；大多序列的最适扩增温度低于长链 T_m 10℃～20℃，GC 含量越高，最适扩增温度越高，扩增能力趋于增强；对互补的两序列来讲，最适扩增温度相近但扩增能力大多有显著差别，如含有两个相同嘧啶的序列扩增能力往往强于其互补序列．c．大多数四重复序列适合在 70℃ 下扩增，某些 G、C 含量高的序列可以在 80℃ 下大量扩增．d．回文序列初始扩增很快，数小时后扩增逐渐变慢；非回文序列初始扩增很慢，数小时后基本呈线性扩增．e．扩增产物与原重复序列具有相同的重复单位．

该实验进一步揭示了重复单链易扩增的特点，可部分解释体内串联重复序列延伸的原因．也应看到，该实验条件(单链模板、较高温度、嗜热聚合酶等)与人体内实际情况有一定差异．有研究显示，人体温度(37℃)下，某些三重复[25]、少数四重复和多数二重复双链(后两项为初步试验结果，未列出)在嗜热聚合酶作用下也能很快扩增，相信这些结果对了解体内重复序列的变化机制也有所帮助．

本研究丰富了重复序列体外等温扩增特性的研究，提出了重复单链扩增的两阶段模型．此模型可对 PCR 等方法的引物设计提供参考，避免某些特定序列导致的非特异性扩增，也可为扩增技术开辟新思路．由于本机制并不能很好地解释短时间内大的 DNA 分子的形成，这一机制有待进一步设计实验以验证与改进．

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] Takanori M. Origin of genomic DNA: disscussion from reverse-transcription and expansion of repetitive oligonucleotides. Viva Origino, 2003, **31**(2003): 46−61

[3] 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

[4] 徐建欣, 王云月, 姚　春, 等. 利用 SSR 分子标记分析云南陆稻品种遗传多样性. 中国水稻科学, 2012, **26**(2): 155−164
Xu J X, Wang Y Y, Yao C, *et al*. Chin Rice Sci, 2012, **26**(2): 155−164

[5] 杨琳琳, 欧阳鸿, 徐湘民. 应用短串联重复序列快速诊断 21 三体. 中华医学遗传学杂志, 2004, **21**(5): 466−469
Yang L L, Ou J J, Xu X M. Chin J Med Genet, 2004, **21**(5): 466−469

[6] Thibodeau S N, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science, 1993, **260**(5190): 816−819

[7] Mangiarini L, Sathasivam K, Seller M, *et al*. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell, 1996, **87**(3): 493−506

[8] Mirkin S M. Expandable DNA repeats and human disease. Nature, 2007, **447**(7147): 932−940

[9] Dere R, Napierala M, Ranum L P W, *et al*. Hairpin structure-forming propensity of the (CCTG•CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. JBC, 2004, **279**(40): 41715−41726

[10] Ohno S. Original domain for the serum albumin family arose from repeated sequences. Proc Natl Acad Sci USA, 1981, **78**(12): 7657−7661

[11] 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

[12] Wiedenheft B, Sternberg S H, Doudna J A. RNA-guided genetic silencing systems in bacteria and archaea. Nature, 2012, **482**(7385): 331-338

[13] Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet, 2011, **2011**(45): 273-297

[14] Terns M P, Terns R M. CRISPR-based adaptive immune systems. Curr Opin Microbiol, 2011, **14**(3): 321-327

[15] Kumar R P, Krishnan J, Singh N P, *et al*. GATA simple sequence repeats function as enhancer blocker boundaries. Nat Commun, 2013, **2013**(4): 1844-1860

[16] Subramanian S, Mishra R K, Singh L. Genome-wide analysis of Bkm sequences (GATA repeats): Predominant association with sex chromosomes and potential role in higher order chromatin organization and function. Bioinformatics(Oxford, England), 2003, **19**(6): 681-685

[17] Stevens J R, Lahue E E, Li Guo Min, *et al*. Trinucleotide repeat expansions catalyzed by human cell-free extracts. Cell Res, 2013, **23**(4): 565-572

[18] Kato T, Liang X G. Hiroyuki A. Model of elongation of short DNA sequence by thermophilic DNA polymerase under isothermal conditions. Biochemistry, 2012, **51**(40): 7846-7853

[19] Ogata N, Miura T. Genetic information 'created' by archaebacterial DNA polymerase. Biochem J, 1997, **324**(2): 667-671

[20] Ogata N, Miura T. Creation of genetic information by DNA polymerase of the Thermus thermophilus. Nucleic Acids Res, 1998, **26**(20): 4657-4661

[21] 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

[22] Liang X G, Jensen K, Frank-Kamenetskii M D, *et al*. Very efficient template/primer-independent DNA synthesis by thermophilic DNA polymerase in the presence of a thermophilic restriction endonuclease. Biochemistry, 2004, **43**(42): 13459-13466

[23] Liang X G, Kato T, Asanuma H. Mechanism of DNA elongation during de novo DNA synthesis. Nucleic Acids Symp Ser, 2008, **52**(1): 411-412

[24] Wang Y, Dong P, Liang X G. Elongation of trinucleotide repeats by DNA polymerase. Lecture Notes in Electrical Engineering, 2014 (251): 1383-1392

[25] 王 阳, 贾蕾敏, 董 平, 等. 三核苷酸双链重复序列扩展合成特性及其机理. 生物化学与生物物理进展, 2013, **40**(4): 345-355
Wang Y, Jia L M, Dong P, *et al*. Prog Biochem Biophys, 2013, **40**(4): 345-355

[26] Heidenfelder B L, Topal M D. Effects of sequence on repeat expansion during DNA replication. Nucleic Acids Res, 2003, **31**(24): 7159-7164

[27] Aboul-ela F, Koh D, Tinoco J I, *et al*. Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G(X, Y=A, C, G, T). Nucleic Acids Res, 1985, **13**(13): 4811-4824

Amplification Characteristics of Single-strand Tetranucleotide Repetitive Sequences and Its Mechanism*

TAO Xian-Ming, WANG Peng-Fei, WANG Yang, LIANG Xing-Guo**

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

Abstract Simple sequence repeats(SSR), whose biological significance causes people's increasing attention, are widely distributed in genomes of many organisms. Many of them can be elongated easily and abnormal extension can directly result in certain hereditary diseases in some cases. In this research, sixty kinds of tetranucleotide repetitive sequences (TRS) and six kinds of dinucleotide repetitive sequences (DRS) of 20 nt single strands were used for isothermal amplification by thermophilic DNA polymerase. The electrophoresis results demonstrated that most of single-strand repeats, even the sequences with no complementary bases inside like AGGA, can be elongated. The results of quantitative analysis demonstrated: palindromic sequences were amplified most easily; DRS could be amplified at a broader range of temperature than TRS; DNA with more G and C, were more suitable for amplification under higher temperature; Most strands whose repetitive unit contains two same pyrimidines were amplified more easily than their complimentary ones; the concentration of products exhibited linear relationship with time. The results of restriction endonuclease digestion indicated that the products had the same repetitive unit with their original repetitive sequences. Finally, an two-stage amplification model, including amplification by intra-chain slide and mediated by hairpin-contained structure, was proposed to provide information for the study of nonspecific amplification of repetitive sequences and pathogenetic mechanisms of relevant diseases.

Key words tetranucleotide repetitive sequences, isothermal amplification, non-specific amplification, molecular evolution, hairpin structure

DOI: 10.3724/SP.J.1206.2014.00002

*This work was supported by grants from Recruitment Programs of "Wanren Plan", "Fund for Distinguished Young Scholars" of Shandong Province (JQ201204), "National Youth Qianren Plan" and Program for Changjiang Scholars and Innovative Research Team in University (IRT1188).

**Corresponding author.

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

Received: January 3, 2014 Accepted: May 16, 2014