Detection of high levels of recombination generated during PCR amplification of RNA templates

Wayne Yu, Karl J. Rusterholtz, Amber T. Krummel, and Niles Lehman

BioTechniques 40:499-507 (April 2006)
doi 10.2144/000112124

*Recombination during the PCR amplification of DNA templates can be a serious problem for those seeking to genotype heterogeneous populations, yet a boon to those seeking to enhance variation during in vitro evolution. Here, the extent to which PCR generates chimeric full-length products was estimated using a powerful restriction fragment-length polymorphism (RFLP) assay involving the use of fluorescently labeled PCR primers. Three different RNA-encoding DNA templates were assayed: (*i*) one for a group I ribozyme, (*ii*) one for a 16S ribosomal RNA (rRNA), and (*iii*) one for a messenger RNA (mRNA). In all cases, the observed frequency of chimeric PCR products exceeded 20%, and longer templates appear to produce more chimeric products. Although two of these templates have the potential to form secondary structures during the PCR, this tendency does not seem to heighten recombination frequency. These results corroborate previous studies that show that the production of chimeras can be best attenuated to a certain extent by varying the extension times in PCR.*

INTRODUCTION

Recombination can occur during PCR as a consequence of the tendency of *Taq* and other DNA polymerases to jump from one template to another. Thus, whenever a heterogeneous pool of similar sequences is being amplified, the potential exists to observe recombination, even under PCR conditions that would be considered standard for the error-free amplification of homogenous templates. Reports of the fraction of chimeric PCR products, which can be quantified as the recombination frequency (*rf*), range from <1% to as much as 7% (1–4). In fact, amplification of genes from polyploid sources has recently been shown to give *rf* values as high as 31% (5).

Recombination during PCR results from the existence of incompletely extended DNA oligomers after the extension (typically 72°C) step. Upon reannealing, these partial products can hybridize with a fully extended molecule, and if the two elements of these duplexes are not identical in sequence as a consequence of having been derived from different template molecules used to seed the PCR, a recombination event takes place. Subsequent rounds of PCR can amplify these recombinants and restore full sequence complementarity. Recombinatory PCR conditions exacerbate this phenomenon by lessening the probability of full-length extension; for example, the staggered extension process (StEP) utilizes 50–80 rounds containing no extension step, coupled with brief 0- to10-s annealing steps, to create an abundance of abortive extension products (6). However, in principle, any phenomenon that causes the polymerase to pause or dissociate from a single-stranded DNA template will create partial extension products and promote recombination. Judo et al. (4) noted this in their demonstration that standard PCR amplification of an equimolar mixture of two homologous plasmid-derived sequences results in 1% or 7% recombinants between markers separated by 282 bp when *Taq* or *Vent* DNA polymerases, respectively, are used. The cycling conditions used in that study were 25 rounds of 45-s steps at 94°, 50°, and 72°C. These and other authors noted that an increase of extension time to 3 min reduced recombination rates 2- to 4-fold by both polymerases (1,4).

Researchers who are using PCR to create DNA templates for either the in vivo or in vitro transcription of RNA should be acutely aware of these findings. On the one hand, reverse transcription PCR (RT-PCR) amplification from potentially heterogeneous genomic RNA sources, such as retroviral samples, can lead to erroneous conclusions about populational diversity if artifactual recombination is not taken into account. On the other hand, the use of recombinatory PCR can be deliberately performed during in vitro evolution to exploit the tremendous advantages that recombination offers in exploring sequence space, for both catalytic RNAs and proteins (6–15).

We have been investigating the efficacy of recombination among

Portland State University, Portland, OR, USA

catalytic RNA molecules in enhancing the evolutionary process in vitro. Unlike the templates assayed previously for their propensities to recombine during standard PCR protocols, templates for catalytic RNAs have the capacity to form strong secondary structures. These structures, though not identical to those of the RNAs that they encode, include thermodynamically stable elements such as paired regions and tetraloops that could stall DNA polymerases even at 72°C. Here we employ an improved version of the restriction fragment-length polymorphism (RFLP)-based assay of Judo et al. (4) to investigate the recombination of RNA templates during both typical and recombinatory PCR. We find that recombination of the 421-bp template for the *Tetrahymena* group I intron is remarkably common during standard PCRs, and we test various parameters designed to minimize the recombination. Moreover, we show that other RNA templates, such as those for *Escherichia coli* 16S ribosomal RNA (rRNA) and eukaryotic (seal) messenger RNA (mRNA), that should contain different types of secondary structure are also quite prone to PCR-generated recombination.

MATERIALS AND METHODS

Template Preparation

For the *Tetrahymena* ribozyme, the sources of the two parental molecules were plasmids, each of which encoded a *Tetrahymena* group I intron L-21 mutant. These sequences were both obtained from in vitro selection experiments for calcium-dependent activity (16). The sequence in plasmid NL137 (P1) has a total of four mutations, one of which (C47A) destroys the wild-type *Scr*FI restriction site at position 55. The ribozyme sequence in NL151

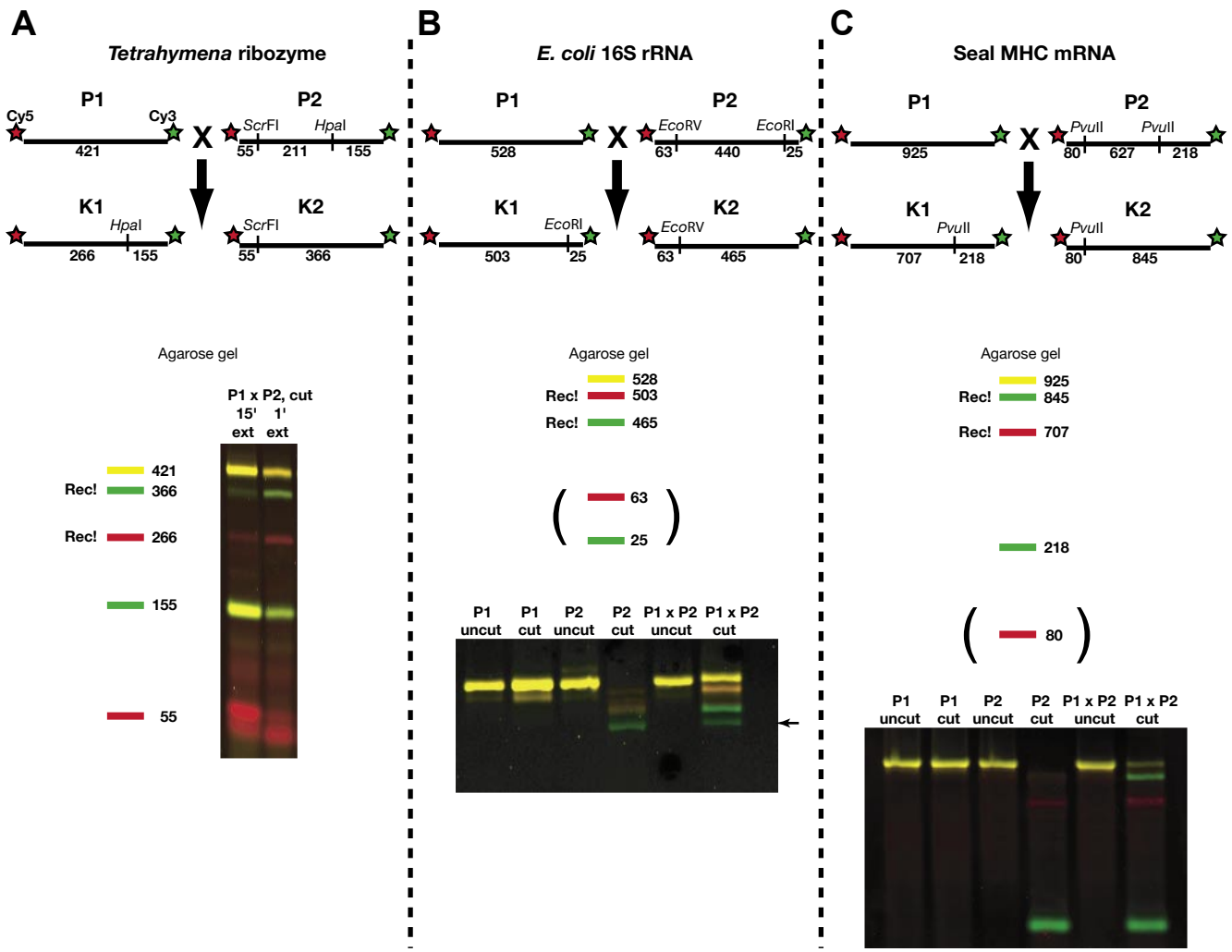

Figure 1. Restriction fragment-length polymorphism (RFLP) strategies to quantify recombination during the PCR amplification of a hetergeneous mix of DNA templates for three different RNA-encoding systems. The faithful amplification of parental templates (P1 and P2) can be distinguished from the artifactual production of recombinant products (K1 and K2) by a restriction digest when fluorescently labeled PCR primers are used. The bands marked Rec! can only result from either recombination during the PCR, copying errors by *Taq* DNA polymerase, or partial digest by the restriction enzymes. The band indicated with an arrow in the *Escherichia coli* 16S ribosomal RNA (rRNA) system is unexplained. It may be a consequence of cross-amplification of another *E. coli* locus and was not figured into the recombination frequency (*rf*) analyses. The bands in parentheses in the schematics are not visible on the example gels shown. MHC, major histocompatibility; mRNA, messenger RNA.

Table 1. PCR Primers Used in this Study

Template	Primers
Tetrahymena ribozyme	TAS 2.1 = 5′-Cy5-CTGCAGAATTCTAATACGACTCACTATAG-GAGGGAAAAGTTATCAGGC-3′ T12 = 5′-Cy3-CGAGTACTCCAAAACTAATC-3′
Escherichia coli 16S rRNA	R519 = 5′-Cy5-ATTACCGCGGCTGCTGG-3′ F8 = 5′-Cy3-AGAGTTTGATCCTGGCTCAG-3′
Seal DQA MHC mRNA	MDQA-2 = 5′-Cy5-CCGGATCCCCAGTGCTCCACCTTGCAGTC-3′ MDQA-1 = 5′-Cy3-CCGGATCCCAGTACACCCATGAATTTGATGG-3′

rRNA, ribosomal RNA; MHC, major histocompatibility; mRNA, messenger RNA.

(P2) contained a C260A mutation that creates a new restriction site *Hpa*I at position 266 while retaining the wild-type *Scr*FI restriction site. Each plasmid was amplified using PCR (see below for protocol), and the products were digested with *Bme*1390I and *Ksp*AI restriction endonucleases (isoschizomers of *Scr*FI and *Hpa*I, respectively) to ensure the plasmids were homogeneous. For the 16S *E. coli* rRNA, cloned bacterial genomic DNA was amplified using R519 and F8 primers that span 528 bp of the 5′ portion of the rDNA gene. The naturally occurring sequence of this gene contains neither *Eco*RI nor *Eco*RV restriction sites and could thus be used as the P1 sequence. To engineer P2, primers were designed with mismatches that create double mutations engendering *Eco*RV and *Eco*RI restriction sites at positions 63 and 503, respectively, in the resulting PCR products (Figure 1B). To insure homogeneity of the P2 sample, the original PCR product was diluted 1000-fold, reamplified with R519 and F8 primers, and digested with *Eco*RV and *Eco*RI, both separately and together. The complete absence of any undigested material was indicative of P2 DNA uncontaminated with P1 DNA. For the seal major histocompatibility (MHC) mRNA, genomic DNA from a Weddell seal with a homozygous genotype W6/W6 at the DQA locus was amplified with primers that generate a 925-bp product spanning approximately 475 bp of exon sequence and 450 bp of intron sequence (17). This gene contains no *Pvu*II restriction sites and could thus be used as the P1 sequence. To engineer P2, a leopard seal with homozygous genotype L1/L1 at this locus was used as a starting point because it contained a single intrinsic *Pvu*II restriction site 80 bp from the 5′ end of the gene. The QuickChange® Mutagenesis Kit (Stratagene, La Jolla, CA, USA) was used to create a second *Pvu*II restriction site 707 bp from the 5′ end of the gene (Figure 1C). PCR amplification of a single bacterial colony was performed with this mutation to produce a pure population of the P2 genotype.

PCR Amplification

The PCR DNAs corresponding to P1 and P2 templates for each system were diluted to approximately 10 fg/μL and equilibrated in concentration. PCR primers were made with 5′ fluorescent dyes for all three systems (Table 1). Amplifications were carried out in 50-μL volumes with 1.25 U *Taq* DNA polymerase (USB Biochemical, Cleveland, OH, USA) and either 2 μL P1 alone, 2 μL P2 alone, or 1 μL P1 and 1 μL P2 mixed as a cross. PCR cycling conditions are given in Tables 2 and 3. For dimethyl sulfoxide (DMSO)-containing reactions, a final concentration of 10% DMSO was used in the PCR; otherwise all other parameters remained the same. Except to check the effects of ramp speed, in all cases, reactions were performed in 200-μL thin-walled microcentrifuge tubes in an MJ Research PTC-100® thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA) with a heated lid.

RFLP Analysis

The resulting PCR products were precipitated in ethanol and resuspended in 40 μL water. Typically, between 5–10 μL were used in either single or double restriction digests. For the *Tetrahymena* ribozyme system, double digestions utilized 8.5 U *Bme*1390I and 1.5 U *Ksp*AI (Fermentas, Hanover, MD, USA). For the *E. coli* 16S rRNA system, double digestions utilized 2.5 U *Eco*RV and 7.5 U *Eco*RI (New England Biolabs, Ipswich, MA, USA). For the seal MHC system, digestions utilized 5.0 U *Pvu*II (New England Biolabs). All restrictions were performed for 1–3 h at 37°C using the buffers supplied by the manufacturers. The resulting fragments were separated in 1× TAE running buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 7.0) on 2.5% agarose gels and visualized on a Typhoon™ PhosphorImager™ (Amersham Biosciences, Piscataway, NJ, USA). All DNA and gel manipulations were optimized such as to minimize both partial digestions and star enzyme activity and were kept in the dark as much as possible to preclude degradation of the Cy™3 and Cy5 labels.

Calculation of Recombination Frequency

Using the ImageQuant® software program (Amersham Biosciences), the band intensities of the restriction fragments were determined separately for Cy3 fluorescence (using excitation at 532 nm) and for Cy5 fluorescence (using excitation at 633 nm). After these values were equilibrated, background fluorescence of each band was subtracted using a band-free portion of the gel, and partial digests were accommodated by subtracting the equivalent positions from lanes containing digestions of only one parental molecule (see Figure 1). In each case, the *rf* could be calculated using the general formula: where K1, K2, P1, and P2 represent the fractions of the total PCR product that are derived from the corresponding molecular type. However, because not all bands in the RFLP assays are uniquely ascribable to a single source, and because not all bands are easily visualized on agarose gels (the smallest bands, denoted in parentheses in Figure 1, were not used), this formula was modified slightly for each system. For the *Tetrahymena* ribozyme system,

the following equation was used: where the bracketed numerals such as [266] refer to the corrected intensity of the band at that size on the gel. In this case, the 155-bp band includes molar representatives from both P2 and K1 molecules, and thus its intensity could be used as their sum. Also, the 421-bp band intensity was divided by two to reflect the amount of P1, because this molecule is doubly labeled. For the *E. coli* rRNA system, the following equation was used: In this case, the two smallest bands (63 and 25 bp) could not be reliably observed on the gel, and thus the P2 value had to be estimated indirectly. This was done by assuming that the P1 and P2 inputs would co-amplify equally—a reasonable assumption if they are input in equal amounts—such that the 528-bp band, being doubly labeled, would be a surrogate for the sum of both parental molecule concentrations. For the seal mRNA system, the following equation was used: Here, the 218-bp band represents the sum of P2 and K1, such that K1 need not be added separately in the denominator. Again, the intensity of the 925-bp band needs to be divided by two, because this molecule is doubly labeled. An example step-by-step calculation of the *rf* in the seal MHC system from a gel is provided in the supplementary materials available online at www.BioTechniques.com.

In all cases except for the DMSO trials, the *rf* values reported are the averages of three independent PCR/RFLP experiments, and in each experiment, the quantification using ImageQuant was performed three times on different days and averaged.

Cloning

$$rf = \frac{K1 + K2}{K1 + K2 + P1 + P2} \times 100\%$$

To corroborate the RFLP-based *rf* values in the seal MHC system, PCR products of the P1 × P2 reaction were ligated into the pCR2.1 vector using the TA Cloning® kit (Invitrogen, Carlsbad, CA, USA), and then the vector was transformed into competent *E. coli* cells. Plasmid DNA was isolated from each colony using the boiling-lysis method (18). A total of 120 of these were chosen for genotyping by digestion with *Eco*RI, which reveals an approximately 1000-bp product on an agarose gel if an insert were present. Second, DNA samples from plasmids containing inserts were then digested with *Pvu*II, which could differentiate among the P1, P2, K1, and K2 genotypes. The *rf* was then estimated using the general formula above.

$$rf = \frac{K1 + K2}{P1 + K2 + (P2 + K1)} \times 100\% =$$

$$\frac{[266] + [366]}{[421]/2 + [366] + [155]} \times 100\%$$

RESULTS

General Strategy

$$rf = \frac{K1 + K2}{K1 + K2 + (P1 + P2)} \times 100\% =$$

$$\frac{[503] + [465]}{[503] + [465] + [528]} \times 100\%$$

Our assay of recombination during the PCR is based on the creation of unique restriction fragments that can only arise via a recombination event between two RFLP marker sites (Figure 1). An RFLP-based assay obviates the need for the cloning and sequence determination of a large number of PCR products (4) and can thus screen a large fraction of the PCR pool (i.e., tens of billions of molecules). We first used as input molecules two DNA templates for the *Tetrahymena* L-21 ribozyme that differ only at two of their 421 bp. These differences create or destroy unique restriction sites for *Scr*FI and *Hpa*I, which have recognition sequences 5′-CC/NGG-3′ and 5′-GTT/AAC-3′, respectively, spaced 211 bp apart. The two input (parental) molecules contained neither restriction site (P1) or both sites (P2) restriction sites, while the two possible recombinant (kid) molecules have either the *Scr*FI site alone (K1) or the *Hpa*I site alone (K2). The recombinant molecules thus produce unique fragments of 266 or 366 bp upon double digestion with *Scr*FI and *Hpa*I (Figure 1A) and allow a determination of the percentage of all full-length PCR products that are recombinants, defined here as *rf*. We also used PCR primers that each contained a single noninterfering 5′ fluorescent label (Cy3 or Cy5), so that the molar amounts of each PCR product molecule could be directly calculated from an agarose gel subject to fluorescent imaging.

$$rf = \frac{K1 + K2}{K1 + K2 + P1 + P2} \times 100\% =$$

$$\frac{K1 + K2}{K2 + P1 + (P2 + K1)} \times 100\% =$$

$$\frac{[707] + [845]}{[845] + [925]/2 + [218]} \times 100\%$$

Comparison with Other Methods to Estimate Recombination Frequency

Judo et al. (4) used this basic strategy to deduce the 1% *rf* produced by *Taq* DNA polymerase in a standard PCR on a pUC-derived fragment of 590 bp. Their approach used RFLP markers a comparable distance apart (282 bp) on a template that was essentially devoid of secondary structure complications. In that study however, there were four methodological shortcomings that we have mainly solved in our strategy. First, their study relied on ethidium bromide staining of the digested fragments. Thus, the amounts of ethidium binding per length of double-stranded DNA molecule had to be considered. Our strategy employs 5′ fluorescent dyes on the PCR primers, allowing for a correspondence closer to 1:1 between band intensity and molar amounts of the restriction fragments. Second, each of their parental molecules contained one of the two restriction sites (*Sma*I and *Pvu*II), while the recombinants contained both restriction sites (K1) or neither one (K2). As a consequence, if a P1/P2 heteroduplex DNA were to form during PCR between the two parental templates prior to any recombination, then it would appear in an RFLP assay as a molecule with no restriction sites (K2), resulting actually in an overestimate of *rf*. Our strategy, in contrast, recombines a zero and a double mutant to produce two single mutants. In this case, if any P1/P2

heteroduplexes exist during the PCR, they would also appear as a molecule with no restriction sites, mimicking one of the two parental molecules (P1), resulting in an underestimate of *rf*. Heteroduplexes that arise after recombination in either system can appear equally as parental or recombinant molecules, but these heteroduplexes would be the least common even at an *rf* of 35% (see the online supplementary materials). Thus, our strategy ironically provides a conservative estimate of *rf* should any heteroduplex formation occur. Third, the formulation of *rf* in the Judo et al. (4) strategy assumes that the frequencies of the two types of recombinants, K1 and K2, are equal, but ours does not. And finally, as a consequence of the means by which they obtained their parental molecules (mutagenic PCR), there was a low-level of contamination—about 5%—of the uncleavable (K2) sequence in the parental pools. Our strategy used two cloned and separately amplified mutants that were free of detectable cross-contamination. The Judo et al. (4) study was able to correct for all of these complications, but our approach requires far less correction and thus should give more accurate and precise estimations of *rf*. We did need to correct for small amounts of partial digest during the RFLP assays and equalize the Cy3 and Cy5 fluorescent outputs, but these were easily done (see the supplementary materials).

Recombination During Amplification of the *Tetrahymena* Ribozyme Template

When we apply a recombinatory PCR protocol (no extension step) to an equimolar mixture of P1 and P2 sequences for the *Tetrahymena* ribozyme template, we observe 30% ± 1% *rf*. This can be clearly observed visually by the presence of the 366-bp (green) and 266-bp (red) bands on an agarose gel after a double digest with *Hpa*I and *Scr*FI (e.g., Figure 1A). Barring incomplete digestion or *Taq* DNA polymerase errors, these bands can only result from the recombinant PCR products K2 and K1, respectively. The uncleaved P1 DNA appears yellow because it contains both the Cy3 and Cy5 labels. To our surprise however, a standard PCR protocol (30 cycles of 1 min at 92°C, 1 min at 50°C, and 1 min at 72°C) did not eliminate recombination (Figure 1A). Under these conditions, the *rf* is lowered only to 22.4% (Table 2). The only significant correction we needed to make to calculate *rf* was a subtraction of the percentage of parental molecules that did not fully digest with both restriction enzymes in a single-buffer double-digest, a value that ranged between 1% and 30% of the intensity of the recombinant band produced in the P1 × P2 PCR. The approximate 22% *rf* value was reproducible over more than 30 independent trials and is minimally affected by the model of the thermal cycler used (i.e., ramp speed), the absolute input template concentration, the magnesium ion concentration, and the extension temperature (data not shown). None of these parameters can be varied in a way that lowers the *rf* below 20%.

Because extension time has been reported as an effective means to minimize recombination during the PCR (1,4,6), we next examined this parameter on the *Tetrahymena* template. Lengthening the extension time while the extension temperature is held constant at 72°C does indeed reduce the *rf* for this template (Figure 1A and Table 2). However, inordinately long extension times, up to 15 min or more, are required to reduce the apparent *rf* to <10%. We could not reduce the *rf* below approximately 8%–9% even when 25- to 30-min extension times were used, at which point the productivity of the PCR begins to fail. Increasing the extension time at elevated extension temperatures not only was less effective in lowering the *rf* as noted above, but also served to reduce the yield in the PCR at temperatures above 83°C. Clearly, the optimal strategy to minimize recombination during standard PCRs with this ribozyme template is to perform as few amplification cycles as possible while utilizing 15-min extension times at 72°C. We obtained similar results with *Vent* DNA polymerase (data not shown).

Table 2. Recombination Frequencies from Amplification of Two Equimolar *Tetrahymena* Ribozyme Templates via PCR with Various Extension Times at Three Different Temperatures

Extension Time (s)	*rf* (% ± standard error)		
	72°C	83°C	88°C
0	29.8 ± 1.1	30.0 ± 0.9	32.5 ± 1.4
10	26.0 ± 1.1	27.0 ± 0.2	27.6 ± 2.7
20	24.8 ± 0.7	26.8 ± 0.7	23.7 ± 0.1
60	22.4 ± 2.0	25.8 ± 0.8	24.0 ± 0.7
120 (2 min)	21.4 ± 0.4	24.0 ± 1.1	25.9 ± 0.3
300 (5 min)	15.4 ± 0.6	—	25.0 ± 1.8
600 (10 min)	12.1 ± 0.8	—	—
900 (15 min)	9.4 ± 0.7	—	—

The recombination frequency (*rf*) values are determined by a restriction fragment-length polymorphism (RFLP)-based assay. Standard errors were derived from three identical trials. A dash (—) indicates no PCR product produced.

Recombination During Amplification of 16S rRNA and Seal mRNA Templates

To test the generality of these results, we assayed the *rf* resulting from standard PCRs using two other RNA-encoding templates. These two templates are expected to contain differing secondary structural characteristics. We examined a 521-bp fragment of the *E. coli* 16S rRNA and a 925-bp fragment of seal mRNA (MHC locus DQA). The former encodes an RNA that has a similar number of stem-loop elements as the *Tetrahymena* ribozyme, but that differs in the degree of compaction and requirement to form a catalytic pocket. The latter encodes an mRNA that spans two exons and one intron and should have far less secondary or tertiary structure. The predicted melting temperature (T_m)

Table 3. Recombination Frequency of PCR Products Resulting from the Amplification of Equimolar Quantities of Various RNA-Encoding DNA Fragments under Standard PCR Cycling Conditions

Template	Size (PCR Product/Marker Distance)	PCR Cycling Conditions	Average *rf* by RFLP Analysis (No DMSO/with DMSO)	*rf* by Cloning Analysis
Tetrahymena ribozyme	421 bp/211 bp	Thirty cycles of 1 min at 92°C, 1 min at 50°C, 1 min at 72°C	22.4% ± 1.2%/nd	nd
Escherichia coli 16S rRNA	528 bp/440 bp	Thirty cycles of 30 s at 94°C, 30 s at 50°C, 90 s at 72°C	32.4% ± 3.2%/ 30.3%	nd
Seal DQA MHC mRNA	925 bp/627 bp	35 cycles of 45 s at 92°C, 45 s at 53°C, 45 s at 72°C	35.7% ± 5.0%/ 51.3%	26.4% (91 clones)

RFLP, restriction fragment-length polymorphism; DMSO, dimethyl sulfoxide; rf, recombination frequency; rRNA, ribosomal RNA; MHC, major histocompatibility; mRNA, messenger RNA; nd, not determined.

values under PCR conditions (1.5 mM $MgCl_2$, 50 mM NaCl) are 73.8°, 75.8°, and 65.9°C for the ribozyme, rRNA, and mRNA templates, respectively. These values suggest a dichotomy between the stable RNA-coding and the protein-coding templates, but comparison of exact secondary structure predictions at 72°C would be difficult and problematic because of the high concentrations of antisense strands in solution. Analogous RFLP schemes to that used in the *Tetrahymena* ribozyme template were used (Figure 1, B and C); in fact the MHC locus RFLP strategy required only one restriction enzyme. Remarkably, both of these other loci demonstrated substantial levels of recombination during the PCR (Table 3). In the case of the *E. coli* 16S fragment, we used PCR cycling parameters based on those employed by environmental microbiologists (19). This cross under these PCR conditions generated 32.4% ± 3% *rf*. In the case of the seal MHC mRNA, we used cycling parameters commonly used to amplify this gene from dilute genomic DNA samples (17). This cross under these PCR conditions generated 35.7% ± 5% *rf*. Thus, a variety of RNA-encoding templates, even those whose source DNA should have minimal secondary structures, are highly prone to PCR chimera generation.

Cloning and DMSO Control Experiments

We performed two additional experiments to investigate possible explanations of these data. First, we chose the system that consistently generated the highest *rf* values, the seal MHC, and cloned the PCR products in a typical fashion to ensure that the high *rf* was not some unforeseen artifact of the RFLP strategy. We isolated plasmid DNA from 102 bacterial colonies derived from P1 × P2 PCR amplifications under standard conditions. Ninety-one of these colonies that contained inserts were subjected to a *Pvu*II digest that could discriminate among P1, P2, K1, and K2. We found that 36 clones were P1, 31 were P2, 12 were K1, and 12 were K2. Although 91 is a small number compared with the $>10^9$ assayed by RFLP, the cloning served to corroborate the magnitude of the *rf* observed via our RFLP strategy. We observed an *rf* of 26.4% and note that P1 and P2 were amplifying and generating recombinants in roughly equal amounts. This 26.4% value lies just within the 95% confidence interval of our estimate of 35.7% *rf* based on RFLP alone, assuming a binomial distribution centered at *Prob*(recombination) = 0.357. Second, to explore the notion that secondary structure in the template DNA was contributing to our results, we repeated both the *E. coli* rRNA and the seal mRNA using standard PCRs in the presence of DMSO and determined the *rf* values. DMSO is often used in PCR specifically to render amplifiable difficult templates, such as those with secondary structure (20,21). However, DMSO did not noticeably lower the *rf* in either system (Table 3).

DISCUSSION

Our results indicate that template length, or specifically the distance between the two restriction markers, is the best correlate with *rf*, not secondary structure. The system with perhaps the most severe secondary constraints was the *Tetrahymena* ribozyme, but these templates gave the lowest *rf*, significantly less than the next highest, the 16S rRNA value ($P = 0.038$; Student's *t*-test). Conversely, the seal mRNA likely possesses the fewest secondary constraints, but gave the highest *rf*, although not significantly different from the 16S value. Yet the distance between RFLP markers in the latter case was more than three times greater. On the other hand, the *Tetrahymena* system does display the highest rate of recombination per unit length (*rf*/bp) of the three systems, and thus a residual effect of secondary structure on the dissociation of *Taq* DNA polymerase from its template is still conceivable. The relatively minor variations in normal extension times across the three systems (45–90 s) should not significantly affect *rf* (Table 2). In addition, DMSO does not reduce the *rf* values greatly in our assays. DMSO in fact increased *rf* in the mRNA system, suggesting that secondary structure could actually serve to decrease the likelihood of recombination when intrastrand annealing can successfully compete against annealing of a second DNA molecule. Thus, while we cannot rule out the influence of the pausing of thermostable polymerases on base-paired elements during PCR, we speculate that the single greatest influence on *rf* is the fraction of incompletely extended templates that can

exist. This value in turn derives from the staggering of extension events and the consequent effect that template length will play. The most effective way to reduce recombination in PCR still appears to be lengthening of the extension time, regardless of the template (22).

We acknowledge that certain factors likely raise *rf* in our assay above the true value, but posit that these factors would not result in *rf* values more than 5% different from those in Table 2. For example, the in vitro error rate of *Taq* DNA polymerase is estimated at 8.0×10^{-6}/bp, meaning that 16% of 1-kb sequences should be mutant (23). This could account for a portion of the observed *rf* if mutations destroyed restriction sites, because this would convert P2 sequences into K1 or K2 sequences. We do correct for this in calculating *rf*, by subtracting the background in the P2 cut lane at the sites where recombinant bands occur from those positions in P1 × P2 lanes (see the supplementary materials). Furthermore, K1 and K2 sequences can be converted to P1 sequences by the same mechanism. In all cases, the same number of restriction sites (and thus the same number of nucleotides) is a target for mutation, and yet average *rf* values clearly differ by system.

In conclusion, we have found strikingly high percentages of recombinant PCR products for RNA-encoding templates from three very disparate systems. These percentages range from 20%–35%, even under PCR conditions not specifically designed to produce recombination. Template length is the best candidate for a parameter that influences *rf*. Those seeking to minimize *rf* in the PCR should lengthen extension times and perform a minimum number of cycles, while those seeking to enhance *rf* should reduce extension times and perform additional cycles. Note that the absolute values of the *rf* values seen in this study may not be realized in other systems in which the ratio of input parental molecules is greatly skewed from 1:1. Amplifications from populations with one dominant and a few rare genotypes should be less prone to recombination. But the phenomenon of frequent chimeric PCR products should be manifest in many situations relevant to RNA biology. It should particularly be an issue in the amplification of alleles from heterozygous diploid or polyploidy sources, or from heterogenous pools with square or broad distributions of genotype frequencies. These results should be viewed as an encouragement to those employing recombination during the in vitro selection of RNAs. On the other hand, researchers anticipating faithful amplification when amplifying RNA-encoding DNA templates should realize that PCR recombination is a significant issue and that this problem is not easily resolved by changing amplification parameters.

ACKNOWLEDGMENTS

We thank A. Burton, D. Decker, E. Hayden, and S. Soll for useful advice and the National Science Foundation (NSF; DEB-0315286 to N.L.) for funding.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Meyerhans, A., J.P. Vartanian, and S. Wain-Hobson.** DNA recombination during PCR. Nucleic Acids Res. *18*:1687-1691.
2. **Yang, Y.L., G. Wang, K. Dorman, and A.H. Kaplan.** 1996. Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. AIDS Res. Hum. Retroviruses *12*:303-306.
3. **Bradley, R.D. and D.M. Hillis.** 1997. Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. *14*:592-593.
4. **Judo, M.S., A.B. Wedel, and C. Wilson.** 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. *26*:1819-1825.
5. **Cronn, R.C., M. Cedroni, T. Haselkorn, C. Grover, and J.F. Wendel.** 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. Theor. Appl. Genet. *104*:482-489.
6. **Zhao, H., L. Giver, Z. Shao, J.A. Affholter, and F.H. Arnold.** 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. *16*:258-261.
7. **Lehman, N.** 2004. Assessing the likelihood of recurrence during RNA evolution in vitro. Artif. Life *10*:1-22.
8. **Burke, D.H. and J.H. Willis.** 1998. Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. RNA *4*:1165-1175.
9. **Kauffman, S.A. and W.G. Macready.** 1995. Search strategies for applied molecular evolution. J. Theor. Biol. *173*:427-440.
10. **Shao, Z., H. Zhao, L. Giver, and F.H. Arnold.** 1998. Random-priming in vitro recombination: an effective tool for directed evolution. Nucleic Acids Res. *26*:681-683.
11. **Lehman, N. and P.J. Unrau.** 2005. Recombination during in vitro evolution. J. Mol. Evol. *61*:245-252.
12. **Jaschke, A. and M. Helm.** 2003. RNA sex. Chem. Biol. *10*:1148-1150.
13. **Joern, J.M., P. Meinhold, and F.H. Arnold.** 2002. Analysis of shuffled gene libraries. J. Mol. Biol. *316*:643-656.
14. **Moore, G.L. and C.D. Maranas.** 2000. Modeling DNA mutation and recombination for directed evolution experiments. J. Theor. Biol. *205*:483-503.
15. **Stemmer, W.P.** 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*:389-391.
16. **Lehman, N. and G.F. Joyce.** 1993. Evolution in vitro of an RNA enzyme with altered metal dependence. Nature *361*:182-185.
17. **Lehman, N., D.J. Decker, and B.S. Stewart.** 2004. Divergent patterns of variation in major histocompatibility complex class II alleles among Antarctic phocid pinnipeds. J. Mammal. *85*:1215-1224.
18. **Sambrook, J., E. Fritsch, and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual. CSH Laboratory Press, Cold Spring Harbor, NY.
19. **Reysenbach, A.L., G.S. Wickham, and N.R. Pace.** 1994. Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. Appl. Environ. Microbiol. *60*:2113-2119.
20. **Varadaraj, K. and D.M. Skinner.** 1994. Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. Gene *140*:1-5.
21. **Shammas, F.V., R. Heikkilä, and A. Osland.** 2001. Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. Clin. Chim. Acta *304*:19-28.
22. **Rozak, D.A. and P.N. Bryan.** 2005. Offset recombinant PCR: a simple but effective method for shuffling compact heterologous elements. Nucleic Acids Res. *33*:e82.
23. **Cline, J., J.C. Braman, and H.H. Hogrefe.** 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. *24*:3546-3551.

Received 21 September 2005; accepted 21 December 2005.

Address correspondence to:

Niles Lehman
Department of Chemistry
Portland State University
PO Box 751
Portland, OR 97207, USA
e-mail: niles@pdx.edu