Disruption of phase during PCR amplification and cloning of heterozygous target sequences

Ruud Jansen⁺ and Fred D.Ledley

Howard Hughes Medical Institute, Department of Cell Biology and Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA

Received May 21, 1990; Revised and Accepted August 2, 1990

ABSTRACT

PCR amplification of genomic DNA or cDNA has become a standard tool for identification of mutations underlying genetic disease. There are inherent limitations in the application of this method in compound heterozygotes. One problem which is encountered is the disruption of phase (linkage) between heterozygous polymorphisms represented on heterologous alleles. A test system was used to demonstrate and quantitate the disruption of phase between two polymorphic restriction sites. Phase is disrupted in approximately 1% of the PCR amplified material, possibly due to incomplete chain elongations and subsequent priming on the heterologous allele. Phase is disrupted in approximately 1/4 of cloned PCR fragments, possibly due to excision repair of heteroduplexes during cloning. The implications of these disruptions for the use of PCR in identifying mutations are discussed.

INTRODUCTION

The polymerase chain reaction (PCR) is widely used for cloning cDNA (1, 2) or genomic segments (3) in order to identify pathogenic mutations. Many elegant schemes have been developed for localizing mutations directly from amplified PCR products or from PCR products cloned into plasmid vectors. Heterozygosity of the target DNA presents several analytical problems for methods which utilize PCR amplification.

First, some methods of mutation detection, such as direct sequencing of PCR products, will demonstrate heterozygosity by the coexistence of two distinct species; one representing the normal sequence and one representing a polymorphic variant. Methods based on hybridization kinetics or cleavage at mismatched residues, however, demonstrate only the presence of novel sequences, but cannot reliably distinguish the absence or presence of the normal sequences. Thus, PCR products are commonly cloned, and multiple clones are analyzed to distinguish homozygous, heterozygous, or artifactual sequence changes (e.g. 4, 5, 6).

Second, compound heterozygosity is commonly assumed when two distinct heterozygous mutations are observed in a gene. The

phase of multiple heterozygous mutations (i.e. whether they segregate on the same or heterologous alleles) cannot be determined by direct analysis of PCR products unless these sequences are incorporated in a single amplification unit and fortuitously alter restriction endonuclease sequences. 'Full length' clones are often sequenced to ascertain the structure of a single allele. Analysis of multiple 'full length' clones, however, can reveal an apparently random association of heterozygous polymorphisms which is incompatible with simple genetic inheritance (4, 6), suggesting that the phase of these polymorphisms between the alleles is disrupted during amplification or cloning.

In order to assess the frequency of these phase disruptions, we constructed a test system containing two polymorphic restriction endonuclease sites. We observed disruption of phase in a small fraction of molecules (1%) following PCR amplification and in a predictably larger percentage (25%) in cloned PCR products. The mechanisms and implications for these phase disruptions are discussed.

MATERIALS AND METHODS

Clones with Polymorphic Restriction Sites

Two clones containing 670 bp from the 5' end of the human methylmalonyl CoA mutase (MCM) cDNA (7) in the *KpnI*-*Clal* sites of the vector pGEM7zf(+) were used (fig. 1A). The first clone derived from cell line MAS contains a polymorphic *MspI* site at position 392 (6) and is designated Mh (fig. 1A). The second clone derived from cell line DS79 lacks the polymorphic *MspI* site but contains a polymorphic *HindIII*(+) site at position 713 (8) and is designated mH (fig. 1A). A segment spanning the two polymorphic sites was amplified with oligonucleotides to the T7 promoter and internal MCM sequences (oligo #52, MCM bases 329-348(sense): TTCCGGGAGTGAAGCCATTC) (fig. 1B).

Amplification and Subcloning

PCR was performed with 5 ng of plasmid as template and 30 cycles of 92°C×1 min; 55°C×1 min; 74°C×1 min; *TaqI* polymerase (Perkin Elmer Cetus), and a Perkin Elmer Cetus Thermal Cycler (9, 10). Control reactions were performed without template. PCR products were digested with *AccI* and

⁺ Present address: Laboratorium voor Moleculaire Biologie, Centraal Diergeneeskundig Instituut, Edelhertweg 15, The Netherlands

KpnI and subcloned into the *AccI-KpnI* sites of a vector containing the first 392 bases of MCM up to the *AccI* site (fig. 1C). The ligation products were introduced into JM109 (Stratagene) by electroporation using a BioRad electroporator with conditions of 2.5 kV, 25 uF, and 200 Ohms (11). Inserts from 48 individual clones were amplified from overnight bacterial cultures using the SP6 and #52 oligonucleotide primers with the addition of [32 P]-dCTP (5 uCi, 200 uM total dCTP) (fig. 1D). These products were pooled and analyzed on 5% polyacrylamide gels after digestion with *SmaI* to remove polylinker sequences. Bands were quantitated by densitometry using an LKB Ultrascan densitometer and corrected for the dCTP content of the fragment.

For methylation experiments, the PCR products were ligated with vector, methylated with *dam* methylase (BioLabs) for one hour according to manufacturers specifications, subjected to digestion with *MboI*, then introduced into bacteria by electroporation.

RESULTS AND DISCUSSION

Disruption of Phase During PCR Amplification

A 1:1 mixture of plasmids with the Mh and mh inserts were added to PCR reactions. As a control, the Mh and mh plasmids were amplified independently, then aliquots of the PCR products were combined and subject to 30 cycles of PCR in the absence of polymerase. The products of these reactions are illustrated in fig. 2A and quantitated in fig. 2B. The products amplified from

mH alone digest with *HindIII*, but not with *MspI* (lane 1). The products amplified from Mh alone digest with *MspI*, but not *HindIII* (lane 2). Products resulting from the amplification of the Mh and mH mixture yield four bands corresponding to the uncut fragment, the *MspI* cut fragment, the *HindIII* cut fragment, and a fragment cut with both *MspI* and *HindIII* (lane 3). The products resulting from cycling the mixture of Mh or mH produce three bands, representing the uncut, the *MspI* cut, and the *HindIII* cut molecules (lane 4). In both lane 3 and lane 4, the uncut fragments represent approximately 1/2 of the PCR product and are presumably comprised of heteroduplexes between Mh and mh strands or molecules which have lost both the *HindIII* and *MspI* sites (designated mh) due to a disruption in the phase between these polymorphisms. Approximately 1% of the product produced by amplification of the heterologous templates is digested with both *HindIII* and *MspI* (designated MH) and represents molecules in which there has been disruption of phase between the two polymorphic sites. This loss of phase requires polymerization as evidenced by the absence of the MH material when the Mh and mH products are run through a mock PCR reaction in the absence of polymerase (fig. 2A, lane 4). The MH molecules

A. TEMPLATES FOR PCR REACTION

B. PCR PRODUCTS

C. CLONES IN pGEM-MCM7.1 (*AccI/KpnI* site)

D. CLONED PRODUCTS

Figure 1. Schematic of experiments to quantitate disruption of phase during amplification and cloning of heterozygous target sequences. A. PCR products resulting from amplification and cloning of templates containing either a polymorphic *HindIII* (H) site or *MspI* (M) site were analyzed. The presence (upper case letters: M or H) or absence (lower case letters: m or h) of restriction recognition sites within each type are indicated. A. The template molecules were derived from cDNA clones for human methylmalonyl CoA mutase in pGEM7zf(+) (7). One template (Mh) contains a polymorphic *MspI* site at position 392. The second template (mh) contains a polymorphic *HindIII* site at position 713. A region spanning these sites was amplified with oligonucleotides to the T7 promoter primer and an internal oligonucleotide. The position of polymorphic restriction sites and constant sites used for analysis are illustrated. B. Five major duplex forms could be present in the PCR products as illustrated. These include authentic copies of the templates (mH and Mh), heteroduplex molecules (mH/Mh), and products with disruption of phase having both polymorphic sites (MH) or neither polymorphic site (mh). C. The PCR products were digested with *AccI* and *KpnI* and cloned into the cDNA vector. The composition of the polylinker adjacent to the *KpnI* site is shown. D. The potential products of cloned PCR products include authentic clones of the templates (mH and Mh) and clones with disrupted phase (MH or mh) as illustrated.

B

type	length	PCR AMPLIFIED area(% total)	NO POLYMERASE area(% total)
MH	321	1702 (1)	0 (0)
Mh	384	52515 (36)	45049 (33)
Mh	408	24770 (16)	26543 (18)
mh	471	84932 (47)	81740 (49)

Figure 2. Disruption of phase during PCR amplification of heterologous templates. A. *HindIII/MspI* double digests of PCR products produced from various templates. Lane 1. PCR products amplified from a template containing a *HindIII* site (mH). Lane 2. PCR products amplified from a template containing a *MspI* site (Mh). Lane 3. PCR products amplified from an equal mixture of the Mh and mH templates. Lane 4. Equal amounts of the products amplified from the template with the *HindIII* site (mH, lane 1) and the template with the *MspI* site (Mh, lane 2) were mixed and run through 30 cycles of annealing, extension, and denaturation in the absence of polymerase. The products of these reactions are illustrated schematically in fig. 1. B. Quantitation of the relative amount of each product by scanning densitometry. The integrated areas of each peak (arbitrary units) and percent of total [32 P]-dCTP incorporation corrected for the length of each segment is shown for the PCR amplified material (equivalent to lane 3) and amplified material run through 30 cycles of PCR without polymerase (equivalent to lane 4). The quantitative difference between mH and Mh species reflects unequal amounts of initial template.

presumably arise from incomplete extension during one PCR cycle and subsequent use of this incomplete fragment as a primer annealed to a template with the opposite polymorphism during a subsequent PCR cycle. This is the same process as the recently described 'jumping PCR' (12) which was found to be dependent in part upon the integrity of the template.

Loss of Phase During Cloning of PCR Products

The co-amplified Mh and mH PCR products (material shown in fig. 2A, lane 3) were digested with *KpnI* and *AccI*, subcloned, and analyzed (fig. 3A). The inserts were completely digestible with a non-polymorphic enzyme *XhoI* (lane 2). The majority of the cloned molecules could be cut with either *HindIII* (lane 4) or *MspI* (lane 5). Double digestion with *HindIII* and *MspI* revealed four bands (lane 6) which differed in relative intensity from the composition of the material before cloning (compare fig. 2B and fig. 3B). There is a decrease in the uncut fraction which now represents only mh molecules lacking both the *HindIII* and *MspI* sites. There is an increase in the amount of material in the double digested fraction, representing MH molecules with both the *HindIII* and *MspI* sites.

This result demonstrates that a significant proportion (approximately 1/4) of the cloned PCR products exhibit disruption in the phase between the *HindIII* and *MspI* sites. The observed frequency is consistent with a model in which heteroduplexes

between Mh and mH strands in the PCR product are subjected to excision repair following introduction into *E. coli*. Such repair would randomly 'correct' the heteroduplex mismatch resulting in creation or elimination of the restriction recognition sequence. Approximately 1/2 of the molecules would be heteroduplexes following PCR amplification of a heterologous template. Random repair would restore the normal phase in 1/2 of these molecules yielding Mh or mH forms and would alter the phase in the other 1/2 resulting in the mh or MH types. This would produce 1/8 mh; 3/8 Mh; 3/8 mH; and 1/8 MH which is tolerably close to the proportions observed.

It is theoretically possible for cross-over to occur between plasmids with mH and Mh types, giving rise to a similar loss of phase. Because these experiments were performed in a RecA deficient host, we consider this an unlikely explanation for the observed frequency of phase loss.

Is Phase Disruption During Cloning Sensitive to *dam* Methylase?

We considered the possibility that methylation of PCR products with *dam* methylase would prevent mismatch repair and prevent disruption of phase (13). The PCR products were ligated, treated with *dam* methylase, digested with *MboI* to eliminate non-methylated molecules, then electroporated into *E. coli* (data not shown). The frequency of phase disruption was not affected by this treatment, reflecting either incomplete *dam* methylation (asymmetrical methylation might prevent digestion, but not mismatch repair) or the activity of other excision repair mechanisms in *E. coli*.

CONCLUSION

This work demonstrates that disruption of phase of heterozygous polymorphisms occurs in up to 1/4 of cloned PCR products. This phenomenon complicates the generally accepted scheme and logic for identifying mutations in heterozygotes using the PCR. When a full length cDNA is cloned, sequenced, and found to contain only one deviation from normal, this variation is commonly presumed to be a pathogenic mutation. If the clone is obtained by PCR, this assumption is not valid, since there is a significant probability that additional heterozygous sequence variations may have been eliminated from this clone and missed by the analysis. Thus, it is technically impossible to prove that any single allele has been completely sequenced unless the target sequences are homozygous. Furthermore, in individuals who are compound heterozygotes, the identification of two heterozygous mutations does not ensure that these mutations are on opposite alleles. It is difficult to rule out the possibility that the two variations are on the same allele and that additional pathogenic sequence changes have been missed. This provides another reason why extreme caution must be exercised in demonstrating the pathogenic character of putative mutations.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Gail Herman and Paul Modrich for their advise in this project and Tammy Reid for preparation of the manuscript. FDL is an Assistant Investigator of the Howard Hughes Medical Institute. This work was supported by NIH grants HD-24186 (FDL) and HD-24064 (Baylor MRRC).

Figure 3. Disruption of phase during cloning of PCR products from heterologous templates. A. PCR products resulting from the amplification of equal amounts of template containing the *MspI* site (Mh) or the *HindIII* site (mH) (material shown in figure 2A, lane 3) were cloned into pGEM7zf(+). The cloned inserts were independently amplified by PCR, pooled, and digested with various enzymes. Lane 1. length markers as indicated; Lane 2. *XhoI* digest. Lane 2. Undigested. Lane 3. *HindIII* digest. Lane 4. *MspI* digest. Lane 5. *HindIII/MspI* double digest. The products of these reactions are illustrated schematically in fig. 1. B. Quantitation of the relative amount of each product by scanning densitometry. The integrated areas of each peak (arbitrary units) and percent of total [32 P]-dCTP incorporation corrected for the length of each segment is shown for the cloned material after digestion with *HindIII* (equivalent to lane 3), *MspI* (equivalent to lane 4), or a *HindIII/MspI* double digest (equivalent to lane 5).

REFERENCES

1. Gibbs,R.A., Chamberlain,J.S., and Caskey,C.T. (1989) In Erlich,H.A. (ed.), PCR Technology. M Stockton Press, New York, pp. 171-191.
2. Kawasaki,E.S. (1990) In Innis,M.A., Gelfand,D.H., Sninsky,J.J., and White,T.J. (eds.), PCR Protocols. Academic Press Inc., New York, pp. 21-27.
3. Saiki,R.K. (1990) In Innis,M.A., Gelfand,D.H., Sninsky,J.J., and White T.J. (eds) PCR Protocols. Academic Press Inc., New York, pp. 13-20.
4. Kobayashi,K., Jackson,M.J., Tick,D.B., O'Brien,W.E., and Beaudet,A.L. (1990) *J. Biol. Chem.* (in press).
5. Ledley,F.D., Jansen,R., Nham,S-U., Fenton,W.E., and Rosenberg,L.E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3147-3150.
6. Jansen,R., and Ledley,F.D. (1990) *Am. J. Hum. Genet.* (in press).
7. Jansen,R., Kalousek,F., Fenton,W., Rosenberg,L.E., and Ledley,F.D. (1989) *Genomics* **4**, 198-205.
8. Ledley,F.D., Crane,A.M., and Lumetta,M. (1990) *Am. J. Hum. Genet.* **46**, 539-547.
9. Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B., and Erlich,H.A. (1988) *Science* **239**, 487-491.
10. Scharf,S.J., Horn,G.T., and Erlich,H.A. (1986) *Science* **233**, 1076-1078.
11. Dower,W.J., Miller,J.F., and Ragsdale,C.W. (1988) *Nucleic Acids Res.* **16**, 6127-6145.
12. Paabo,S., Irwin,D.M., and Wilson,A.C. (1990) *J. Biol. Chem.* **265**, 4718-4721.
13. Pukkila,P.J., Peterson,J., Herman,G., Modrich,P., and Meselson,M. (1983) *Genetics* **104**, 571-582.