Introduction

The development of protocols for the direct DNA sequence analysis of PCR products and the use of PCR to facilitate conventional sequencing strategies have been the focus of investigators since 1986. Direct PCR sequencing enables rapid and precise determination of sequence identity and variation, which is useful in most aspects of molecular biology and for diagnostic genetic applications. In a more general way, PCR improves the ease and capacity of all DNA sequencing activities, by simplifying the screening, preparation, and manipulation of DNA templates.

Many different variations of the integration of sequencing and PCR amplification have therefore been described. This chapter provides a general overview of issues and methods, and the subsequent chapters describe some particular approaches in detail, to illustrate both procedural aspects and the situations in which individual protocols may be favored. Interested readers may also refer to previous reviews on this subject (Gibbs et al., 1991; Gyllensten, 1989; Bevan et al., 1992).

Direct DNA Sequencing versus Cloning and Sequencing

"Direct DNA sequencing" refers to the analysis of unfractiomed PCR amplified DNA molecules, which contrasts to the process of molecularly cloning fragments prior to the sequence determination (Fig. 17.1). The two procedures are similar to carry out, and offer similar results; however there are several important distinctions. The foremost of these is that direct sequencing simultaneously views an entire mixture of PCR amplified DNA molecules in a single assay. If the sequences to be analyzed are homogeneous, then a single unambiguous sequence results. However, if a mixture of similar sequences is present then a direct sequence analysis will reveal ambiguous signals at the mixed positions. For example, a PCR-amplified human DNA sample that is heterozygous for a single DNA base substitution will yield a mixture of two kinds of DNA fragments. The data will contain unambiguous sequences at the homozygous positions, but an equal representation of the two bases at the polymorphic position.
Cloning PCR products allows DNA sequence variants to be separated before the sequencing (see Chapter 2 for details of PCR cloning procedures). In the example of a human heterozygous mutation, an equal number of clones is expected to be recovered with each of the base substitutions. When each separate clone is sequenced, the results are unambiguous, but when different clones are compared about one-half will contain one of the polymorphic bases, and the other half the different base. In theory, the identification of these mixed bases is a simple statistical problem, and provided a sufficient number of clones is examined it is highly likely that polymorphic base positions will be correctly identified.

When more complex mixtures of bases are to be analyzed a larger number of individual clones can be sequenced. In these experiments the cloning of PCR products has an additional advantage compared to direct sequencing, as the relative linkage of multiple base differences can be determined. For example, if DNA base differences were to result from the presence of two bases at two positions in an amplified fragment, direct sequencing would not show which of the alternate bases at each position were to be found on the same molecule. By cloning the DNA prior to sequencing the individual molecules are separated and so the linkage is identified.

The Effect of Polymerase Errors on Sequencing Strategy

Polymerase errors introduced during PCR amplification are of concern primarily when cloning is used prior to the DNA sequencing, but can also influence direct sequencing (Keohavong and Thilly, 1989; Eckert and Kunkel, 1991). In either case, when the abundance of the initial template DNA is very low, such as when single molecules or cells are amplified, then a polymerase error that occurs early in the amplification will result in a significant fraction of the molecules containing a misincorporated base. In an extreme case, an error in the first cycle of a PCR can result in 25% of all amplified molecules being in error. This would be identified by either cloning or direct PCR sequencing strategies.

These are unlikely scenarios, however, and the vast majority of amplifications begin with larger numbers of template molecules. For example, amplification of a single copy gene from a nanogram of total human DNA involves approximately 500 starting template molecules. Assuming that at least 10% of the templates are copied in early PCR cycles, even if there was a “first cycle error,” the representation of PCR errors in the final product would still be masked in a direct sequencing assay by