Chapter 2

A Brief Comparison Between *In Vivo* DNA Replication and *In Vitro* PCR Amplification

In principle, PCR generates large quantities of DNA from a minute amount of nucleic acid starting material using a methodology similar to (but much simpler than) that seen in living cells. For living cells, *in vivo* DNA synthesis is dependent upon a well defined but complex set of enzymes and co-factors, which have evolved to act in a concerted fashion during the synthetic phase (S-phase) of the cell cycle. In comparison, PCR facilitates in vitro DNA synthesis in a much simpler fashion, making use of a smaller set of defined ingredients and reaction conditions involving relatively high temperatures. The range of factors contributing to successful PCR amplification is reviewed below.

2.1 Nucleic Acid Targets

2.1.1 DNA

In vivo DNA duplication, which is essentially a form of limited DNA amplification, is performed under the direction of a select and diverse set of structural proteins, enzymes and additional co-factors (a detailed description of which is beyond the scope of this book and interested readers are referred to the international literature for a more in-depth look into this particular topic, e.g. Shcherbakova et al., 2003; Goren and Cedar, 2003; Kelman, 2000; Nasheuer et al., 2002; Nasmyth, 2002). In eukaryotes, the DNA molecule is intimately associated with positively charged proteins, which are strongly electrostatically bound to the phosphate moieties on the DNA chain. These “histone” proteins associate into octameric complexes, bind to approximately 400 base pairs (bp) of genomic DNA, and constitute approximately half of the mass of the eukaryotic chromosome. These histones do not completely disassociate from the DNA during replication. An additional level of complexity also exists in eukaryotic DNA, in that further folding processes induce secondary (300 nm) and tertiary (700 nm) order folding on the template DNA, thereby generating tightly coiled DNA with a high molecular density [Stewart, 1997] (Fig. 2.1). For prokaryotes, the situation regarding the *in vivo* state of DNA

© Springer Science + Business Media B.V. 2008

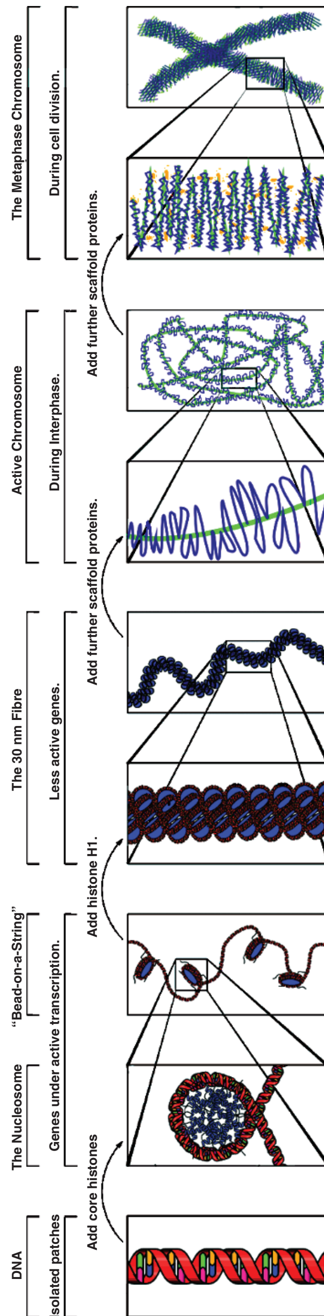

Fig. 2.1 *In vivo*, the double stranded DNA molecule is organized into a highly condensed structure with an impressive molecular density. The primary double helix structure is embedded into spherical nucleosomes (11 nm in diameter) via an interaction with chaperone or histone proteins. These nucleosomes are in turn ordered into 30 nm chromatin fibers which are looped into macrostructures called chromatin. This chromatin may be further condensed into heterochromatin, providing the framework that generates the final eukaryotic chromosomal structure. What is shown is a beads-on-a-string model of chromatin with a linker length of 20 nucleotide pairs in three dimensions (From http://en.wikipedia.org/wiki/Chromatin)

is generally considered to be somewhat simpler, with microorganisms lacking a visible nucleus and containing only one (circular) chromosomal copy per cell. However, it is known that prokaryotic DNA is also associated with a variety of histone-like proteins, which intimately interact with the cellular DNA, protecting it and conferring complex structure (though this structure is generally less ordered or regular than that associated with eukaryotic chromosomes). Therefore, one of the major requirements for DNA replication in *in vivo* systems is the systematic uncoiling of specific regions of genomic DNA in order to expose single strands of DNA ready for replication. In both eukaryotic and prokaryotic kingdoms this process is achieved via specific DNA topoisomerases, helicases and gyrases, which act together to help uncoil the double stranded DNA.

In contrast, the uncoiling of DNA during PCR amplification is not enzymatically controlled but is in fact achieved using a far less complicated procedure, i.e. the DNA to be amplified is usually chemically extracted from its host chaperone proteins, and the residual tertiary or secondary structure removed by heating the naked DNA. This heating step also provides the mechanism by which the two DNA strands are separated (melted) into single strands ready for PCR amplification. As a further point, whereas *in vivo* DNA chromosomal replication involves the replication of many millions of nucleotides, PCR amplification products are generally designed to be shorter than 1,000 bp in length (a parameter largely imposed on the PCR amplification process by the type of heat-stable DNA dependent DNA polymerases used). However in extreme cases, the successful PCR amplification of regions of DNA over 35 kilobases in length have been reported using special heat-stable DNA dependent DNA polymerases combinations [Hu et al., 2002; Davies and Gray, 2002]. In principle, DNA from all types of viruses, cells (plant, animal and bacterial), or tissues (lung, brain, etc.) may be amplified using nucleic acid extraction techniques coupled to PCR amplification.

2.1.2 RNA

As defined in the "central dogma of molecular biology", the flow of genetic information *in vivo* proceeds from DNA to RNA (via transcription) and then finally to protein (via translation). However, since the central dogma was formulated, scientific discoveries have shown that the flow of genetic information may occur in the reverse direction, i.e. from RNA to DNA (reverse transcription). This process of reverse transcription occurs during the replication cycle of certain virus families (*Retroviridae*, *Hepadnaviridae* and *Caulimoviridae*), and the relevant enzymes (reverse transcriptases) have been isolated and utilized to generate DNA from RNA molecules in specially adapted PCR protocols. This special form of PCR is known as reverse transcriptase- or reverse transcription-PCR (often shortened to RT-PCR), and allows the detection and quantification of several types of RNA molecules, including messenger RNA (mRNA) from both prokaryotic and eukaryotic cells, as well as RNA-based genomes, e.g. Coronaviruses and Picornaviruses. Several non-PCR-based amplification methods have also been developed which do not

utilise reverse transcriptase enzymes and hence do not make use of an intermediate DNA template. These techniques amplify RNA (and not double stranded DNA as is the case with PCR) using an initial RNA template.

2.2 Target DNA Strand Separation and Primer Annealing

During *in vivo* DNA replication, cell associated proteins (including DNAa, DNAb, DNAc, single stranded binding proteins, helicases and gyrases, ligase and a variety of polymerase subunit proteins), all act in concert to uncoil the stable α-helical DNA structure, break the hydrogen bonds between the purine and pyrimidine bases, and expose a DNA replication origin [Nasheuer et al., 2002]. In contrast, DNA required for PCR amplification is separated from its chaperone proteins using chemical and/or enzymatic extraction methods, with the DNA then being separated into single strands via thermal disassociation, i.e. via incubation at approximately 94°C for 30 seconds to 5 minutes, which causes breaking of the hydrogen bonds between the complementary base pairs present on opposite strands. This method is not available to living cells as proteins are rapidly denatured at 94°C and cellular integrity irretrievably breaks down. In fact, the melting temperature (T_m) of a DNA molecule is defined as the temperature at which half of it's constitutive bases are no longer paired to their complementary partner on the opposite strand. For most natural species of DNA, the melting temperature lies somewhere between 70°C and 100°C, with the actual T_m being dependent on both the length of the DNA molecule to be melted and on the base composition of the strands. Moreover, there exists a linear relationship between the percentage of guanosine and cytosine bases present in the DNA strands (referred to as the "GC content") and the melting temperature of that particular DNA helix, with GC-rich DNA melting at a higher temperature than adenosine and thymidine rich (AT-rich) DNA. Because of the fact that DNA melting curves are relatively steep, temperatures of T_m+5°C and T_m−5°C above or below the calculated T_m value will result in the double helix being completely denatured or completely intact, respectively.

Several other factors may also facilitate DNA helix destabilization, all of them operating by destabilizing the interaction between the various complementary DNA base moieties. Extreme pH values, denaturants such as formamide or urea and the overall salt concentration are important parameters in this respect (Rauch et al., 2000). Essentially, the addition of these reagents shifts the DNA melting curve to the left, with the result that the absolute T_m for the double stranded DNA value decreases. This decrease in T_m may be tens of degrees centigrade dependent on the factor included in the reaction mix. Conversely, several compounds actually stabilize the DNA double helix, including magnesium ions and elevated salt concentrations, which help neutralize the triple-negative phosphate charges on the opposing DNA strands in the duplex (therefore limiting electrostatic repulsion), as well as (partially) neutralizing the effect that the water dipole has on DNA hydrogen bonding.

The heating procedure used to melt DNA during the PCR process is a simple and reliable method for ensuring that DNA strand separation occurs and that DNA binding sites for specifically designed PCR primers are exposed. However, at DNA melting temperatures, PCR primers also remain disassociated from the target DNA, and effective binding of the PCR primers to the target DNA can only take place at a reduced temperature when the thermal energy is low enough to allow complementary base pairing. In most applications (random amplification of polymorphic DNA or RAPD excepted), PCR primers are designed to specifically bind to known sequences of target DNA and to anneal to the melted target DNA at a temperature of between 45°C and 65°C (the "annealing" temperature). Of course, at this temperature the target DNA also re-anneals to its complementary strand. However, the excess concentration of primers added to PCR mixes ensures that binding of at least some PCR primers will occur during each thermocycle (dependent of course on the presence of primer-complementary sequences in the target DNA).

2.3 DNA Dependent DNA Polymerase and Oligonucleotide Primers

DNA dependent DNA polymerase is an essential component of both *in vivo* DNA replication and the PCR process (Fig. 2.2), though the DNA dependent DNA polymerases found in the vast majority of organisms are heat sensitive (thermolabile), one of the major stumbling blocks hindering the initial success of PCR. However, the discovery of thermophilic organisms and the subsequent isolation of thermostable DNA dependent DNA polymerases from these organisms heralded a new chapter in PCR amplification technology [Saiki et al., 1988], such that successive cycles of heating and cooling (necessary for target DNA melting and primer annealing during PCR thermocycling) no longer resulted in the concomitant denaturation of the thermolabile DNA dependent DNA polymerase enzyme. The discovery of these thermostable enzymes allowed the whole PCR process to become far more convenient, less time consuming and more user-friendly for laboratory personnel. The first thermostable DNA dependent DNA polymerase enzyme to be widely used in PCR (Taq polymerase) was derived from the hot spring-dwelling bacterium *Thermus aquaticus*, though many other commercially available non-Taq thermostable DNA dependent DNA polymerase enzymes are now available on the market, e.g. Pfu from *Pyrococcus furiosus*, Vent from *Thermococcus litoralis*, etc. The advantages and disadvantages of some of these different thermostable polymerases with respect to PCR thermocycling are described more fully in Chapter 7.

DNA dependent DNA polymerases *per se* are actually incapable of performing DNA synthesis from a purely single stranded piece of DNA and require an additional shorter "priming" oligonucleotide to initiate DNA replication. In *in vivo* DNA replication, an RNA oligonucleotide (generated by a DNA dependent RNA polymerase called primase) acts as the primer for DNA replication. This primase enzyme synthesizes a short RNA molecule (of approximately ten nucleotides),

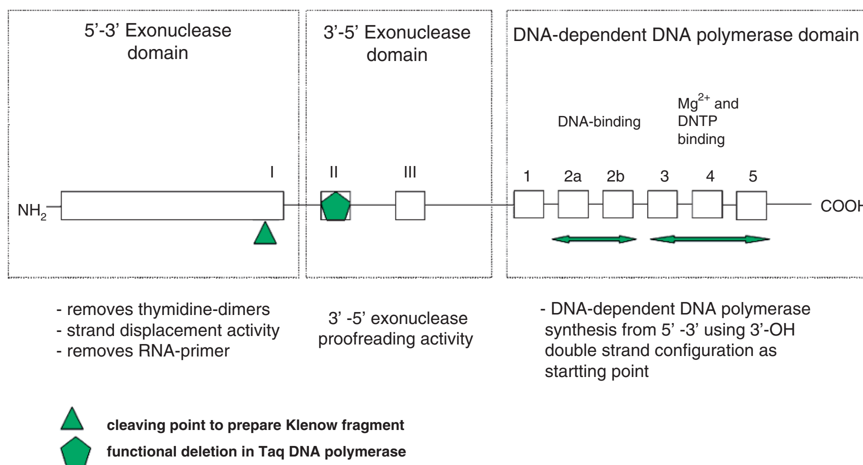

Fig. 2.2 Schematic outline of the structure of the DNA dependent DNA polymerase I holoenzyme from *Escherichia coli*. The amino terminal domain has exonuclease activity, which is able to remove RNA primers and thymidine dimers and to induce strand displacement. The deletion of this domain generates an enzyme that is deficient in 5′- to 3′- exonuclease activity (Klenow fragment). In contrast, the central domain exhibits 3′- to 5′- exonuclease activity, and the carboxy terminal domain is involved in the binding of DNA and co-factors such as magnesium ions and dNTPs. *Taq* polymerase essentially has the same structural organisation

which is complementary to the DNA region to be replicated, and once hybridized acts as a "primer" for DNA replication by cellular DNA dependent DNA polymerase. The primase enzyme itself does not require a primer in order to generate this short RNA oligonucleotide. In contrast, a typical PCR uses two specifically designed and synthetically synthesized 15–25 base pair long oligonucleotides to act as primers for the DNA dependent DNA polymerase enzyme (see Fig. 1.1, Chapter 1). PCR primers are designed such that one primer is complementary to the sense (coding) template strand and one primer complementary to the antisense (non-coding) template strand, and as such are the major components, in determining PCR specificity. Further, the two PCR primers determine the length of the DNA region to be amplified and facilitate the exponential (two-fold or doubling) of amplification products during each PCR cycle.

2.4 Deoxyribonucleotides and Additional Factors

In vivo DNA replication and PCR not only require specific primers (RNA for *in vivo* replication, DNA for PCR amplification) and a (thermostable) DNA polymerase, but also several other factors. The most essential of these factors are the building

blocks or free deoxynucleotide triphosphate molecules that are incorporated into the growing DNA chain by the DNA dependent DNA polymerase. These compounds can be acquired *in vivo* as nutrients, but are generally synthesized via complex pathways involving the reduction of ribonucleotide diphosphates and addition of a further phosphate group by a kinase enzyme. The four major deoxynucleotides required for DNA synthesis (adenosine-, guanosine-, thymidine- and cytosine-triphosphate) contain an energy-rich triphosphate moiety, which is utilized by the DNA polymerase to catalyze a phosphodiester link between the 3′-hydroxy terminus of the primer (or previously added deoxynucleotide triphosphate) and the newly acquired deoxynucleotide triphosphate on the growing DNA strand. Hydrogen bonding between complementary nucleotides on adjacent strands then completes the double stranded primary structure of the DNA molecule. The mechanism by which deoxyribonucleotide triphosphates (dNTPs) are added to the growing DNA chain is identical for both *in vivo* replication and PCR amplification processes, though the concentration of dNTPs in *in vitro* amplification reactions may be easily manipulated to artificially high levels. As well as deoxyribonucleotides, successful *in vivo* DNA synthesis and amplification requires a large number of chemical components such as (1) the bivalent metal ions magnesium (a cofactor for the DNA polymerase enzyme) and manganese and (2) simple chemicals such as sodium chloride, etc. which help to maintain the correct pH of the reaction or allow for the synthesis of new DNA with a complex secondary structure (see Chapters 4–7 and Wilson et al., [2002] for a more detailed discussion).

References

Davies PA, Gray G. 2002. Long-range PCR. Methods Mol Biol 187:51–55.

Goren A, Cedar H. 2003. Replicating by the clock. Nat Rev Mol Cell Biol 4:25–32.

Hu M, Chilton NB, Gasser RB. 2002. Long PCR based amplification of the entire mitochondrial genome from single parasitic nematodes. Mol Cell Probes 16:261–267.

Kelman Z. 2000. DNA replication in the third domain of life. Curr Protein Pept Sci 1:139–154.

Nasheuer HP, Smith R, Bauerschmidt C, Grosse F, Weisshart K. 2002. Initiation of eukaryotic DNA replication: regulation and mechanisms. Prog Nucleic Acid Res Mol Biol 72:41–94.

Nasmyth K. 2002. Segregating sister genomes: the molecular biology of chromosome separation. Science 297:559–565.

Rauch J, Wolf D, Hausmann M, Cremer C. 2000. The influence of formamide on thermal denaturation profiles of DNA and metaphase chromosomes in suspension. Z Naturforsch 55:737–746.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491.

Shcherbakova PV, Bebenel K, Kunkel TA. 2003. Functions of eukaryotic DNA polymerases. Sci Aging Knowl Environ 26:RE3.

Stewart RD. 1997. A theoretical investigation of cell cycle effects and interspecies radiosensitivities. Ph.D. thesis.

Wilson T, Carson J, Bowman J. 2002. Optimisation of one-tube PCR-ELISA to detect femtogram amounts of genomic DNA. J Microbiol Methods 51:163–170.