Ligation: Theory and Practice

Karthikeyan Kandavelou, *Johns Hopkins University, Maryland, USA*

Mala Mani, *Johns Hopkins University, Maryland, USA*

Sekhar PM Reddy, *Johns Hopkins University, Baltimore, Maryland, USA*

Srinivasan Chandrasegaran, *Johns Hopkins University, Baltimore, Maryland, USA*

Ligation is the process by which NAD DNA ligases catalyse the formation of phosphodiester bonds between juxtaposed 3′-hydroxyl group and 5′-phosphate termini in duplex deoxyribonucleic acid (DNA). Ligases use adenosine triphosphate or nicotinamide–adenine dinucleotide (NAD) as cofactors for this covalent joining of DNA.

Introduction

One of the key steps in recombinant deoxyribonucleic acid (DNA) technology is the joining of two separate DNA fragments covalently to form a single DNA molecule that is capable of autonomous replication in cells. The construction of recombinant DNA molecules involves the joining of insert sequences into plasmids or bacteriophage cloning vectors. This critical step is performed by DNA ligases. Like type II restriction enzymes, DNA ligases are essential tools in recombinant DNA technology for analysing and manipulating DNA. DNA ligases are essential enzymes in the synthesis of new DNA during cell division and for maintaining the integrity of the genome.

DNA Ligase Reaction Properties and Intermediates

DNA ligases catalyse the formation of phosphodiester bonds between juxtaposed 5′-phosphate groups and a 3′-hydroxyl termini in duplex DNA. They can repair single-stranded nicks in duplex DNA as well as covalently join restriction fragments with compatible cohesive ends or blunt ends (**Figure 1**). The commonly used DNA ligases in nucleic acid research are T4 DNA ligase and *Escherichia coli* DNA ligase. T4 DNA ligase was originally purified from T4 phage-infected cells of *E. coli*; it is the product of gene 30 of phage T4. *E. coli* DNA ligase is the product of the *lig* gene. Both genes have been cloned and the enzymes are obtained from overproducing strains. T4 DNA ligase and *E. coli* ligase differ from each other in two important properties: (1) T4 DNA ligase uses adenosine triphosphate (ATP) as a cofactor, whereas *E. coli* ligase uses NAD as a cofactor; (2) under normal reaction conditions, only T4 DNA ligase joins blunt ends efficiently. While *E. coli* DNA ligase repairs single-stranded nicks in duplex DNA and joins restriction fragments with compatible cohesive or sticky ends, it does not ligate restriction fragments with blunt ends under normal reaction conditions. **See also**: Bacteriophage T4; DNA ligases; DNA repair by reversal of damage; Nucleic acids: general properties

DNA ligases catalyse the DNA joining reaction in three distinct steps: (1) the formation of a covalent enzyme–adenosine monophosphate (AMP) complex; (2) the transfer of the AMP moiety to the 5′-phosphate at a nick in the duplex DNA; and (3) the formation of the phosphodiester bond with concomitant release of AMP. **See also**: DNA-binding enzymes: structural themes

The first step of the reaction is the best characterized. Studies with ATP-dependent T4 DNA ligase and NAD-dependent *E. coli* DNA ligase have shown the formation of a covalent enzyme–nucleoside monophosphate reaction intermediate. It appears that the AMP moiety is linked via a phosphoramidite bond to a lysine residue in an active site motif, KXDGXR, that is diagnostic for DNA ligases (**Figure 2**). A further five conserved motifs have been defined by protein sequence alignments of DNA ligases. **See also**: Protein motifs for DNA binding

The crystal structure of bacteriophage T7 DNA ligase has revealed that this enzyme comprises two distinct domains linked together to form a cleft. The active site motif, KXDGXR, is at the base of the cleft, and the other conserved motifs are located on the surface of the two domains close to the site of adenylation. It appears that the nicked duplex DNA binds the cleft between the two protein domains.

Manipulating and Recombining Restriction Fragments with DNA Ligase

Figure 3a shows the cloning of a *Hin*dIII restriction fragment into a vector or plasmid containing a single *Hin*dIII
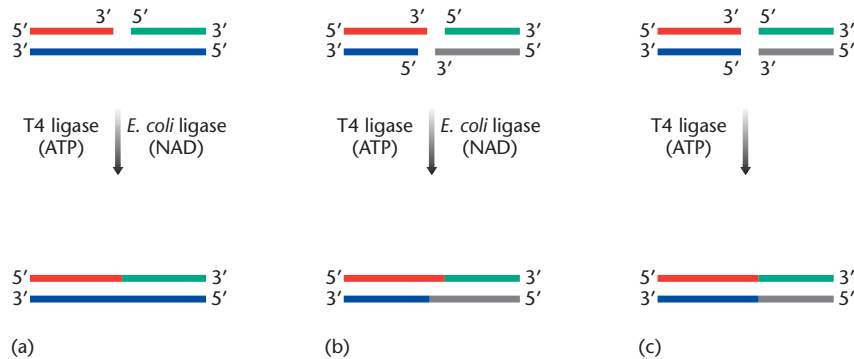
Figure 1 T4 and *E. coli* DNA ligase activity at single-stranded breaks or nicks (a) and at double-stranded breaks with cohesive or sticky ends. (b) T4 DNA ligase activity at blunt ends. (c) ATP, adenosine triphosphate; NAD, nicotinamide–adenine dinucleotide. They catalyse the formation of phosphodiester bonds between juxtaposed 5′ phosphate groups and 3′ hydroxyl termini in duplex DNA. While T4 DNA ligase uses ATP as a cofactor, *E.coli* ligase uses NAD.

site. The first step is to cleave the vector at the unique *Hin*dIII site by using the *Hin*dIII restriction endonuclease to produce the vector with homologous cohesive ends. The cleaved vector is then treated with calf intestinal phosphatase to remove the 5′-phosphates from the termini. This process inhibits self-ligation of the vector. Gel electrophoresis is used to purify the cleaved vector away from the uncleaved vector. Uncleaved and self-ligated vector DNA are the major source of background in the cloning of the restriction fragments. Because all the *Hin*dIII cohesive ends are equivalent, the restriction fragment will be cloned in either of the two possible orientations with respect to the vector sequences. These two possible recombinant products can be distinguished by restriction mapping by using an enzyme that cleaves asymmetrically within the insert DNA. **See also**: Restriction enzymes

Alternatively, directional cloning of the restriction fragments with heterologous ends may be used to force the insert in a particular orientation with respect to the vector sequences (**Figure 3b**). The restriction fragments that are to be subcloned have heterologous ends. These are generated by cutting the DNA with two different restriction enzymes, for example *Bam*HI and *Hin*dIII, respectively. The vector or plasmid DNA containing unique *Bam*HI and *Hin*dIII sites is also digested with *Bam*HI and *Hin*dIII. The cleaved vector DNA with heterologous ends is gel-purified away from the uncleaved vector DNA to reduce background. As the heterologous ends of the vector DNA are not compatible, there is no self-ligation. Directional cloning is possible because the heterologous ends of the insert are compatible with those of the vector DNA in only one orientation. The other orientation results in noncompatible ends for ligation and, hence, this recombinant product is eliminated from the reaction mixture. **See also**: Gel electrophoresis: one-dimensional; Plasmids

T4 DNA ligase can join DNA fragments with blunt ends, but the ligation efficiency is much lower than that achieved by restriction fragments with cohesive ends. The ligation efficiency may be increased by increasing the

concentration of DNA and by using more DNA ligase. The blunt ends are equally ligatible, irrespective of the restriction enzyme used to generate the ends. The ligation of cohesive ends requires compatible termini. The ligation of blunt ends produced by different restriction enzymes results in recombinants that cannot be cleaved by either enzyme.

Optimizing the Formation of Recombinant Molecules

During the ligation reaction, several factors appear to influence the formation of recombinant molecules: (1) the purity of the DNA, (2) the relative concentrations of the insert and vector DNA and (3) increased amount of T4 DNA ligase for blunt-end ligations. The probability of obtaining the desired recombinant DNA molecules is greatly improved with the purity of the individual DNA components of the ligation reaction. Highly purified restriction fragments and vector DNA free of minor contaminants are required. The minor contaminants arise from incomplete digestion by restriction endonucleases. For example, even if the cleavage efficiency of the vector DNA by a restriction enzyme is 99%, the uncleaved 1% of vector DNA readily transfects the competent cells during transformation. This results in an unacceptably high background in these experiments. Gel purification of the cleaved vector DNA away from the uncleaved vector greatly reduces this background. The other major source of background is from intramolecular self-ligation of vector DNA. This can be eliminated by using vector DNA with heterologous ends. In the case of vector or plasmid DNA with homologous ends, intramolecular joining can be prevented by removing the 5′-phosphate from each end with calf intestinal phosphate. This 5′ dephosphorylation cannot be applied to both the vector DNA and the DNA fragment to be ligated. Thus, the untreated partner remains capable of self-ligation. Recently, Ukai and colleagues
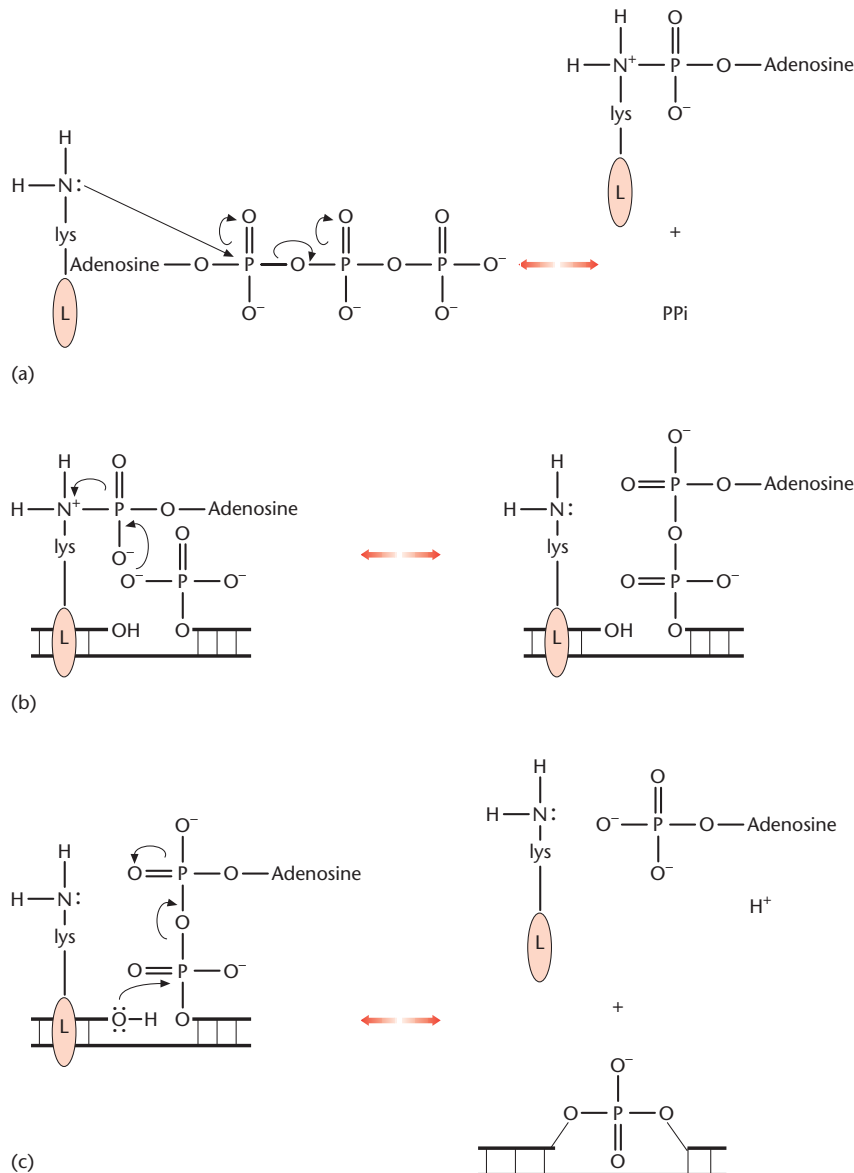
Figure 2 Reaction mechanisms for T4 DNA ligases. (a) T4 DNA ligase (L) reacts with ATP to form a phosphoramide-linked AMP with the amino group of the active lysine site. Pyrophosphate (PPi) is released. (b) The 5′-phosphate at the nick attacks the activated phosphoryl group of the AMP to form an adenylated DNA. (c) The enzyme catalyses joining of the 3′-OH of the DNA at the nick to the activated 5′-phosphate to form the phosphodiester bond and concomitant release AMP.

proposed a new technique to overcome this problem. This technique replaces the 2′ deoxyribose at the 3′-end of the DNA fragment with 2′,3′dideoxyribose to prevent self-ligation. The 5′-phosphate in this fragment remains capable of ligation to the 3′-OH of the vector DNA. By combining this 3′ replacement technique with 5′ de-phosphorylation of vector DNA, self-ligation of both the partners in the reaction can be simultaneously prevented. The relative concentrations of the insert DNA and vector DNA in the ligation mixture may also influence the frequency of specific ligation products. Cloning experiments using vector DNA require the joining of two or more separate DNA molecules followed by the circularization of the product. The optimal cloning efficiency results in DNA concentrations that are high enough to permit sufficient intermolecular ligation, but not so high as to reduce intramolecular ligation. In situations where the background is low, the DNA concentration has been determined experimentally to be between 1 and $50\,\mu g\,mL^{-1}$. It must be pointed out that genetic methods are available to identify

```
                                    AAGCTT ━━━━ AAGCTT
                                    TTCGAA ━━━━ TTCGAA
              HindIII
5' pAGCTT ━━━━ A 3'
        3' A ━━━━ TTCGAp 5'
   HindIII

         +          T4 ligase          +
                    ─────────▶
                      (ATP)

   A 3'        5' pAGCTT        AAGCTT ━━━━ AAGCTT
TTCGAp 5'           3' A        TTCGAA ━━━━ TTCGAA

(a)
```

```
              HindIII
5' pGATCC ━━━━ A 3'
        3' G ━━━━ TTCGAp 5'
   BamHI

                               GGATCC ━━━━ AAGCTT
                               CCTAGG ━━━━ TTCGAA
         +          T4 ligase
                    ─────────▶
                      (ATP)

   G 3'        5' pAGCTT
CCTAGp 5'           3' A

(b)
```
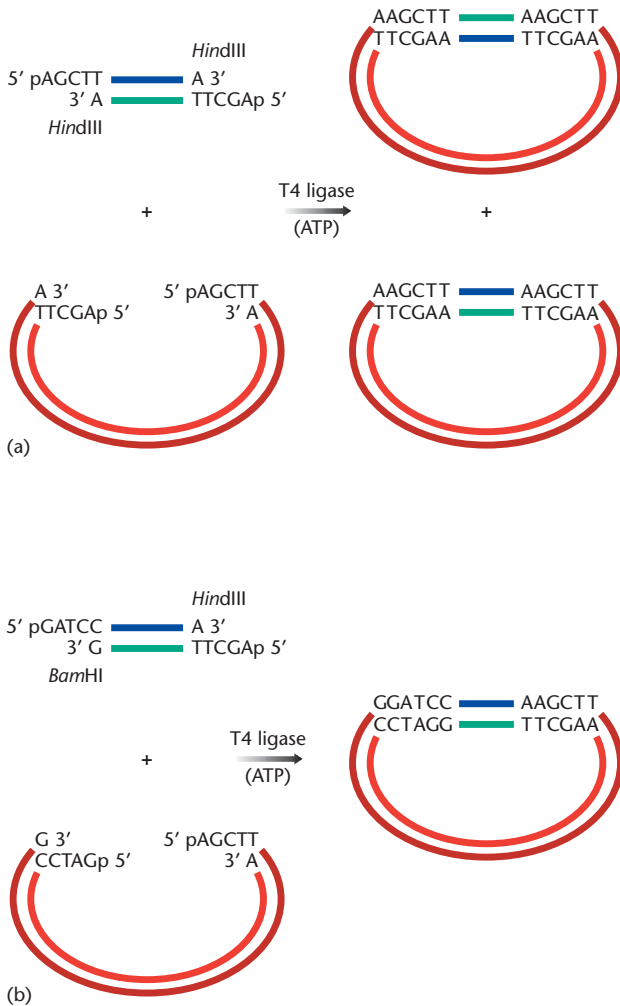
Figure 3 Ligation of a restriction fragment into a plasmid or vector DNA. (a) Cloning of a restriction fragment into a vector with cohesive ends. The insert can be cloned into the vector in two possible orientations with respect to the vector sequences. Both recombinant molecules are shown and they can be distinguished by restriction mapping. (b) Directional cloning of a restriction fragment into a vector with heterologous ends. The restriction fragment is digested with two different restriction enzymes to generate heterologous ends. The plasmid or vector DNA is also cleaved with the same enzymes to generate heterologous ends. This results in the cloning of the insert DNA in only one particular orientation with respect to the vector sequences. The other orientation results in noncompatible ends for ligation, and hence this recombinant product is eliminated from the ligation reaction mixture. ATP, adenosine triphosphate.

colonies containing the desired DNA molecule from background colonies. These include the use of a combination of drug markers, as well as the use of designed cloning vectors to select or screen for recombinants on appropriate indicator plates. Furthermore, polymerase chain reaction (PCR) methods may be employed directly to screen colonies for recombinants by using oligonucleotide primers complementary to the vector sequences flanking the inser-

tion site. Finally, it is important to perform appropriate control experiments simultaneously with the ligation reaction of interest. The controls reflect on the relative success of the cloning experiment. This is particularly helpful because the preparation and analysis of the recombinant DNA molecules is a laborious and time-consuming part of the cloning process. Blunt-end ligation is facilitated by using an even higher concentration of DNA components and about 10-fold more ligase. All ligations are carried out at $16°C$ overnight in a relatively low reaction volume, usually $10–20\,\mu L$. **See also**: Bacterial restriction–modification systems; DNA: methods for preparation; Polymerase chain reaction (PCR); Polymerase chain reaction (PCR): specialized reactions

In practice, the critical parameters for the design of a ligation reaction include the number of inserts, the type of DNA ends, the concentration of the DNA fragment, the preparation and purity of the restriction fragments, and the availability of a selection or screening system that can be employed to identify the desired recombinant DNA molecule. Sometimes, it is important to obtain maximum number of recombinants especially during the construction of genomic or complementary DNA (cDNA) libraries. This ensures a complete coverage or representation of the genome or cDNA. In other instances, it is better to obtain a few colonies containing the desired recombinant product than a large number of colonies in which only relatively few contain the desired product. In each case, the experimental conditions of ligation are adjusted to favour the desired recombinant product. **See also**: Microorganisms: applications in molecular biology

Formation of the desired recombinant product during the ligation reaction can be increased by using an excess of the dephosphorylated vector or plasmid DNA to that of the insert that is to be cloned. An alternative method is to perform the ligation of the vector to the insert with compatible ends in the presence of the restriction enzyme that was used to generate the inserts. The enzyme will inhibit ligation of the inserts to themselves, as these will be cleaved continuously by the enzyme to regenerate the monomer inserts. On the other hand, the ligation of the inserts to the vector DNA is favoured because it leads to the generation of a fusion site that is not recognized by the restriction endonuclease, and hence not cleaved. This results in the accumulation of the desired recombinant molecules during the ligation reactions.

Selection or screening methods are currently available to identify the desired recombinant molecule from a ligation mixture. Cloning vectors have been designed to screen or select for the recombinants. pUC or M13mp derivatives are the commonly used vectors. They contain the α-complementation region of the *E. coli lacZ* gene. They produce blue colonies or plaques on appropriate indicator plates. A white colony or plaque is seen when there is a successful insertion of a restriction fragment in the *lacZ* region of the vector. This white colony or plaque is easily

detected or identified among a background of blue transformants. An alternative approach is to use plasmids or vectors that contain two selectable drug markers, such as ampicillin or tetracycline resistance, as in the case of pBR322. Insertion of a restriction fragment into one of these marker genes renders the colony sensitive to that drug. These recombinant colonies can easily be screened by means of replica plating. **See also**: Genetic engineering: reporter genes

Biochemical Properties of DNA Ligases

The ATP-dependent DNA ligases are widespread in nature from bacteriophage to mammals. The NAD-dependent DNA ligases are unique to bacteria. This uniqueness of NAD-dependent ligases to eubacteria has led to the suggestion that it could be a potential target for novel antibiotics. In addition to ATP or NAD as cofactors, these enzymes require Mg^{2+} for activity. The lower organisms appear to encode only a single essential DNA ligase, whereas mammalian cells carry multiple genes. DNA ligases vary widely in their molecular mass, ranging from 41 kDa (T7 ligase) to 100 kDa (mammalian DNA ligase I). Protein regions with additional regulatory functions account for much of the variation in the size. For instance, mammalian DNA ligase I comprises two distinct and separable domains: a carboxy (*C*)-terminal catalytic domain and an amino (*N*)-terminal region with regulatory functions. The *N*-terminal domain is not required for the catalytic activity. However, it is required both for nuclear localization and for recruitment of the enzyme at the replication sites during the S phase. While significant sequence homology has been observed between ATP-dependent ligases, none is found between NAD-dependent DNA ligases of eubacteria. **See also**: Adenosine triphosphate; Coenzymes and cofactors; NAD^+ and $NADP^+$ as *prosthetic* groups for enzymes; Protein–DNA complexes: specific

Barany and co-workers have compared the biochemical properties of seven NAD-dependent DNA ligases of thermophilic bacteria, collected from *Thermus* species worldwide. The enzymes are highly homologous, with amino acid sequence identities ranging from 85 to 98%. The enzymes have different levels of tolerance for mismatch ligation when Mn^{2+} is substituted for Mg^{2+}. The sequence divergence and subtle structural variation among these DNA ligases appear to underlie the enzyme's recognition preferences toward different mismatched base pairs.

In vivo Function of DNA Ligases

Study of mammalian DNA ligases has shown that they play crucial roles in cellular functions such as DNA rep-

lication, DNA repair and genetic recombination. Four distinct DNA ligase activities have been shown to arise from three mammalian genes (I, III, IV) encoding DNA ligases. DNA ligase I joins the Okazaki fragments generated by lagging strand DNA synthesis during DNA replication. Ligase I is also involved in base excision repair (BER). It directly interacts with DNA polymerase β, the enzyme responsible for BER. DNA ligase I-deficient cell lines are sensitive to DNA damage by alkylating agents, ionizing radiation and ultraviolet exposure. This suggests that DNA ligase I is involved in DNA repair pathways as well. **See also**: Developmentally programmed DNA rearrangements; DNA repair; Eukaryotic replication fork

It appears that two forms of DNA ligase III, IIIα and IIIβ, are produced by alternative splicing in a tissue- and cell type-specific manner. DNA ligase IIIα interacts with the DNA strand-break repair protein XRCC1 to form a complex during the repair of DNA single-stranded breaks in all tissues and cells. This interaction is mediated by dimerization of their carboxy terminal BRCT (BRCA *C*-terminal) modules. BRCT modules are present in various proteins involved in cell cycle and DNA replication and it has been proposed that these processes are controlled in part by interaction of these BRCT domains. Ligase IIIα is also involved in the repair and replication of mitochondrial DNA. DNA ligase IIIβ does not interact with XRCC1 and appears to be involved in the meiotic recombination. Ligase IV takes part in the nonhomologous end joining (NHEJ) repair of DNA double-strand breaks caused by radiation and chemical agents. This enzyme interacts with XRCC4, which stabilizes and directs it to DNA double-strand breaks. It has been shown to be essential for embryonic development and V(D)J recombination in mice. DNA ligases also appear to be involved in the repair of DNA double-stranded breaks by homologous or non-homologous recombination pathways. **See also**: Alternative splicing: cell-type-specific and developmental control; Meiotic recombination pathways; Recombinational DNA repair in eukaryotes

The clones carrying chimaeric nucleases can be made more viable by increasing the levels of the DNA ligase within the cells. As there are no counterpart methylases available for the hybrid endonucleases that are formed by the fusion of the isolated nuclease domain of *Fok*I to other DNA-binding motifs, production of these engineered nucleases *in vivo* is often lethal to cells.

Ligase Chain Reaction

The availability of cloned thermostable ligases makes it possible for the application of ligases in DNA diagnostics. This method is called the ligase chain reaction (LCR). It complements the PCR that has revolutionized DNA diagnostics. PCR is a simple and powerful *in vitro* method that

allows enzymatic synthesis of specific DNA sequences, using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA. A repetitive series of automated cycles, which include template denaturation, primer annealing and the extension of the annealed primers by a thermostable DNA polymerase, results in the exponential amplification of a specific fragment whose termini are defined by the 5′-ends of the primers. LCR, on the other hand, utilizes a thermostable ligase both to amplify DNA and to discriminate a single base substitution in target DNA (**Figure 4**). The thermostable ligase specifically links two juxtaposed oligonucleotides when hybridized at 65°C to a complementary target at the junction. Oligonucleotide products are amplified exponentially by thermal cycling of the ligation reaction in the presence of a second set of adjacent oligonucleotides, complementary to the first set and the target. A single base mismatch at the junction prevents ligation, and hence amplification. Barany has used this method to discriminate between normal β^A and sickle β^S globin genotypes of human population by using 10-μL blood samples. More recently, Barany and co-workers have developed a universal DNA microarray method for multiplex detection of low abundance point mutations in cancer by combining PCR with LCR. Alternatively, point mutations and SNPs can be sensitively detected by using padlock probes. These are oligonucleotides, which include one target-complementary sequence at each end and are designed such that the two ends are placed immediately next to each other on hybridization wit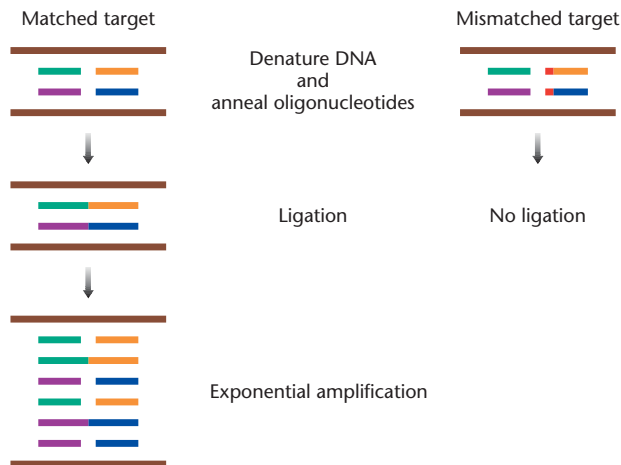h a perfectly matched target sequence. These perfectly matched probes can be linked covalently by ligase. Hence, they can specifically discriminate point mutations on the target sequence. The circularized probes can be amplified by rolling circle amplification (RCA) using single primer and ϕ 29 DNA polymerase. This reaction, unlike PCR, proceeds at 37°C for 3 h and yields a tandem repeat of the target sequence. The RCA product can be directly stained with Sybr-Gold and visualized under UV light.

Summary

DNA ligases play essential roles in many cellular functions, including DNA replication, DNA repair and DNA recombination. They catalyse the formation of phosphodiester bonds between adjacent 3′-hydroxyl and 5′-phosphate termini at single- or double-stranded breaks. DNA ligases fall into two major classes based on their cofactor requirements: (1) enzymes that use ATP and (2) enzymes that are NAD dependent. Four distinct DNA ligase activities from three mammalian genes encoding DNA ligases have been identified. They have been shown to play important roles in the maintenance of genomic stability and integrity.

DNA ligases have also proven to be essential tools in the recombinant DNA technology. These enzymes are used to join two or more separate DNA segments to generate a single DNA molecule that is capable of autonomous replication in cells and bacteria.

LCR is a powerful *in vitro* method that utilizes a thermostable ligase both to amplify DNA and to discriminate a single base substitution in target DNA. LCR thus complements PCR, which has revolutionized DNA diagnostics.

Acknowledgements

Figure 4 Diagram depicting DNA amplification and detection by means of the LCR. The target DNA is heat denatured and four complementary oligonucleotides are then hybridized to the target at 65°C. A thermostable ligase is used to link covalently adjacent oligonucleotides that are perfectly matched to the target. Products from one cycle of ligation become targets for the next cycle, and thus the number of products increases exponentially. Oligonucleotides that contain a single base mismatch at the junction do not ligate efficiently, and therefore do not amplify the ligated product. The single base mismatch at the junction is shown in red.

Further Reading

Barany F (1991) Genetics disease detection and DNA amplification using cloned thermostable ligase. *Proceedings of the National Academy of Sciences of the USA* **88**: 189–193.

Ciarrocchi G, MacPhee DG, Deady LW and Tilley L (1999) Specific inhibition of eubacterial DNA ligase by arylamino compounds. *Antimicrobial Agents and Chemotherapy* **43**: 2766–2772.

Doherty AJ and Wigley DB (1999) Functional domains of an ATP-dependent DNA ligase. *Journal of Molecular Biology* **285**: 63–71.

Frank KM, Sekiguchi JM, Seidi KJ *et al*. (1998) Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**: 173–177.

Grossman L (1997) DNA repair. *Encyclopedia of Human Biology* **3**: 447–454.

Gumport RI and Lehman IR (1971) Structure of the DNA-ligase adenylate intermediate: lysine-linked adenosine monophosphoramidite. *Proceedings of the National Academy of Sciences of the USA* **68**: 2559–2563.

Kubota Y, Nash RA, Klungland A *et al*. (1996) Reconstitution of DNA base extension-repair with purified human proteins: interaction between DNA polymerase β and the XRCC1 protein. *EMBO Journal* **15**: 6662–6670.

Struhl K and Tabor S (2000) Enzymatic manipulation of DNA and RNA: DNA ligases. In: Ausubel FM, Brent R, Kingston RE *et al*. (eds) *Current Protocols in Molecular Biology*, vol. I, chap. 3, pp. 3.14–3.14.3. New York: John Wiley.

Timson DJ and Wigley DB (1999) Functional domains of an NAD$^+$-dependent DNA ligase. *Journal of Molecular Biology* **285**: 73–83.

Timson DJ, Singleton MR and Wigley DB (2000) DNA ligases in the repair and replication of DNA. *Mutation Research* **460**: 301–318.

Tomkinson AE and Levin DS (1997) Mammalian DNA ligases. *BioEssays* **19**: 893–901.

Ukai H, Ukai-Tadenuma M, Ogiu T and Tsuji H (2002) A new technique to prevent self-ligation of DNA. *Journal of Biotechnology* **97**: 233–242.

Qi X, Bakht S, Devos KM, Gale MD and Osbourn A (2001) L-RCA (liagation-rolling circle amplification): a general method for genotyping of single nucleotide polymorphisms (SNPs). *Nucleic Acids Research* **29**: E116.