CHAPTER I

INTRODUCTION

1.1 Amino acids

Amino acids are biomolecule found in all organisms. They can be classified into two groups by ability of the polarized light plane rotation: L-formed and D-formed. L-amino acids play important role in all life by serving as building block of enzyme, hormone, antibody and protein. Furthermore, they can also balance buffering capacity in blood and often function as chemical messengers in cell communication (Holum, 1982). In contrast, D-formed amino acids are rarely found. The optically active amino acids have been extensively studied. They are commonly occurring moiety in the rational design of chiral drugs such as anticancer compounds and viral inhibitors (Taylor *et al.,* 1998).

L-amino acids have stimulated the research on various methods for their synthesis such as chemical reaction, extraction from protein hydrolysates and fermentation (Hummel *et al.,* 1987). However, the products contain both D-formed and L-formed amino acids. Therefore, some researchers have attempted to produce L-amino acid by enzymatic method e.g. L-amino acid transaminase, amino acid racemase (Berberich *et al.,* 1968), L-aminopeptidase (Kamphuis *et al.,* 1992), and L-amino acid-β-decarboxylase (Yamamoto *et al.,* 1980). These methods lead to enantiomerically pure compounds. A recent alternative method is the enzymatic reductive amination of α-keto acid by amino acid dehydrogenase, a route which has

the advantage of complete enantioselectivity and up to 100% towards selectivity of desired product.

1.2 L-alanine

Alanine is a non-essential amino acid that used for protein systhesis. L-Alanine was one of the first amino acids to be synthesized before its isolation from natural sources, in 1850. L-Alanine occurs in high levels in its free state in plasma. It is produced from pyruvate by transamination. It involves in sugar and acid metabolism, increases immunity, and provides energy for muscle tissue, brain, and the central nervous system. L-Alanine can be used as food additive and health care cosmetics. L-alanine is an important raw material of vitamin B6 production (Eisenberg and Star, 1968).

In addition, there are many reports described the using of L-alanine in the synthesis of DOPA (dihydroxy-phenylalanine) which is the precursor of many chemicals such as melanin which found in hair and skin, dopamine; the one of chemicals messengers in nervous system, norepinephrine and epinephrine (Reinhold *et al.,* 1987). The defect of tyrosine hydroxlase causes the decreasing in DOPA production, subsequently, the defect of melanin will occur. The defect of DOPA also causes Pakinson's disease, however, the using of DOPA drug can release the effect of this disease by repairing the activity of nervous system. The derivative of L-alanine like (*N*-(*p*-chlorobenzoyl)-2-(2-cyanoethyl)) alanine is also used for control immune system and can protect the development of tumor (Nagano *et al.,* 1985).

In prokaryote, alanine plays many important roles in the growth and physiology of entericbacteria. Both the L- and D-stereoisomers of alanine are major constituents of peptidoglycan layer (Matsuhashi, 1994).

1.3 L-alanine production

A variety methods has been used for the production of L-amino acid such as chemical synthesis, extraction from protein hydrolysates, fermentative or enzyme method. L-alanine can also be prepared by chemical synthesis used of the Stecker reaction in which acetaldehyde, prussic acid and ammonium are used as starting material, but this process is not necessarily desirable since L-alanine is used as food additive and, in the Stecker reaction, toxic cyanogens is used. L-alanine can be produced by culturing a microorganism, *Zymomonas mobilis* that has been genetically modified in such a manner that express it an alanine dehydrogenase gene. The alanine dehydrogenase converted glucose in the medium, via conversion into pyruvate, into a mixture of alanine and ethanol, with a maximum alanine yield of 16%, calculated on the amount of glucose converted (Holes *et al.*, 2003). Several enzymatic process of L-alanine synthesis have been reported previously. L-alanine also can be produced from decarboxylation of L-aspatic acid catalyzed by L-aspatate-β-decarboxylase of immobilized cells or cell suspension of *Pseudomonas dacunhae* (Yamamoto *et al.*, 1980 and Takamatsu *et al.*, 1981). However, this method has high substrate (aspatate) cost. Recently, the reductive amination of pyruvate catalyzed by L-alanine dehydrogenase can be alternative way to produce L-alanine (Suye *et al.*, 1992). This method was used to produce an L-amino acid from the corresponding keto-acid also applicable to produce various amino acid such as L-leucine (Ohshima

et al., 1985), L-valine, L-isoleucine (Gu and Chang, 1990) and L-phenylalanine (Matsunaga *et al.*, 1987).

A method for enzymatic synthesis of L-amino acid from α-keto acids with NAD(P)-dependent amino acid dehydrogenase has been purposed (Ohshima *et al.*, 1985, Gu and Chang, 1990 and Matsunaga *et al.*, 1987). However, the application of this method to industrial production of L-amino acid has been hampered by cost of coenzyme regeneration. A multienzyme reaction system for simultaneous coenzyme regeneration has been proposed to overcome this problem. In previously, L-leucine has been produced from α-ketoisocaproate (2-oxo-4methylpentanoic acid) with leucine dehydrogenase from *Bacillus sphaericus*, formate dehydrogenase from *Candida boidinii* and NADH, covalently bound to water soluble polyethyleneglycol, in a reactor with an ultrafiltration membrane. The process has been successfully scaled up for industrial production (Hummel and Kula, 1989). From previous application, formate dehydrogenase is a great catalyst for NADH regeneration in the enzymatic synthesis of amino acid. In 2004, a multienzyme reaction system approach for L-alanine production was recently attempted in *Escherichia coli* BL21(DE3) by the heterologous gene expression between formate dehydrogenase gene (*fdh*) and alanine dehydrogenase gene (*aladh*). The two methods were performed 1) cloning of heterologous gene of *aladh* and *fdh* in high expression vector pET-17b and 2) co-transformation of plasmid containing *aladh* and *fdh* under T7 promoter. The coupling reaction of these two enzymes were shown in Figure 1.1. However, the production of alanine by various recombinatnt clones were not significantly different with ratio of

หอสมุดกลาง ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

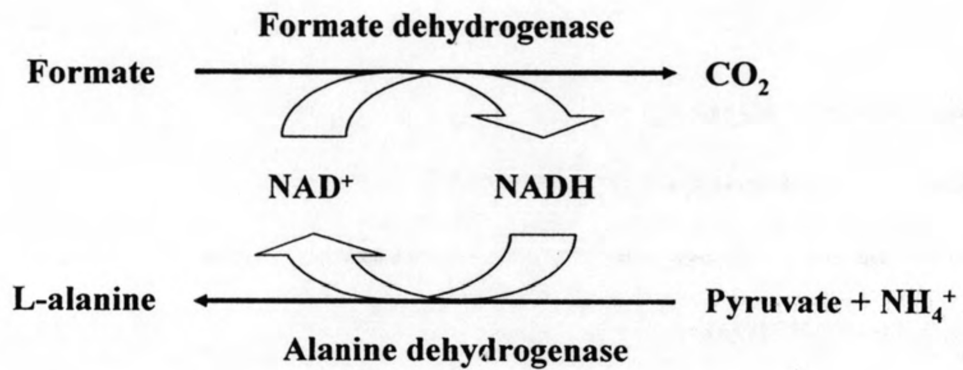

Figure 1.1 Conjugated enzyme system of alanine dehydrogenase and formate dehydrogenase for production of L-alanine

D:L form about 1.6:1 (Hatrongjitt, 2004) because of the presence of alanine racemase activity.

1.4 Alanine racemase

Alanine racemase (EC 5.1.1.1) is catalyst for the interconversion of the D and L forms of alanine (Figure 1.2), and therefore is a typical prokaryotic enzyme. The D enantiomer is used either as an energy source or as a component of bacterial cell wall, where it is directly involved in the cross-linking of adjacent peptidoglycan chain. Alanine racemase required a pyridoxal 5'-phosphate (PLP) as a coenzyme (Matsuhashi, 1994). While some organisms contain only one alanine racemase, other have two isozymes, encoded by the *alr* and the *dadX* genes. The latter group includes *E. coli* (Lobocka *et al.,* 1994) and *Pseudomonas aeruginosa* (Strych *et al.*, 2000). The *alr* gene encodes the constitutively expressed biosynthetic enzyme, sufficient to provide enough D-alanine for cell wall biosynthesis. The catabolic *dadX* gene encodes a second alanine racemase isozyme whose expression is subject to induction by L-alanine, thus is most active when L-alanine is used as a carbon and energy source, which through the subsequent action of D-amino acid oxidase and lactate dehydrogenase, is converted to pyruvate and is thus metabolized in *Lactococcus lactis* (Lobocka *et al.,* 1994).

1.5 Intron

In bacteria, a polypeptide chain is generally encoded by a DNA sequence that is colinear with the amino acid sequence, continuing along the DNA template without interruption until the information needed to specify the polypeptide is complete.

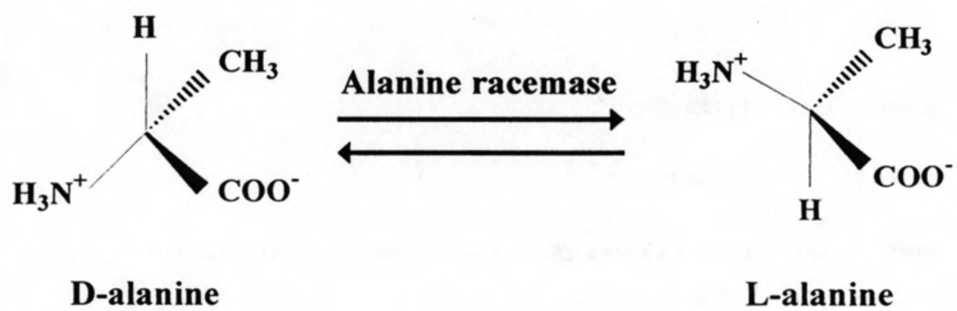

Figure 1.2 Reaction catalyzed by alanine racemase

However, the notion that all genes are continuous was disproved in 1977 when Phillip Sharp and Richard Roberts independently discovered that many genes for polypeptides in eukaryotes are interrupted by noncoding sequences (introns). The occurrence of introns in eukaryotes varies. Many genes in the yeast *Saccharomyces cerevisiae* lack introns, although in some other yeast species introns are more common. Introns are also found in a few eubacterial and archaebacterial genes. Introns in DNA are transcribed along with the rest of the gene by RNA polymerases. The introns in the primary RNA transcript are then spliced, and the exons are joined to form a mature, functional RNA.

There are four classes of introns. The first two, the group I and group II introns, differ in the details of their splicing mechanisms but share one surprising characteristic: they are self-splicing, no protein enzymes are involved. Group I introns are found in some nuclear, mitochondrial, and chloroplast genes coding for rRNAs, mRNAs, and tRNAs. Group I introns carry out transesterification reactions in order to excise themselves from a precursor transcript (Cech, 1993). Group II introns are generally found in the primary transcripts of mitochondrial or chloroplast mRNAs in fungi, algae, and plants. Neither class requires a high energy cofactor (such as ATP) for splicing. The splicing mechanisms in both groups involve two transesterification reaction steps. These reactions are very similar to the DNA breaking and rejoining reactions promoted by topoisomerases and site-specific recombinases.

Most introns are *not* self-splicing, and these types are not designated with a group number. The third and largest class of introns includes those found in nuclear mRNA primary transcripts. These are called spliceosomal introns, because their removal occurs within and is catalyzed by a large protein complex called a

spliceosome. The fourth class of introns, found in certain tRNAs, is distinguished from the group I and II introns in that the splicing reaction requires ATP and an endonuclease. Although spliceosomal introns appear to be limited to eukaryotes, the other intron classes are not (Nelson *et al.*, 2004).

1.5.1 Group II intron

Group II introns, as mobile genetic element, were discovered in the mitochondrial (mt) and chloroplast (cp) genome of lower eukaryotes and higher plants, where they interrupt conserved gene. The organellar group II introns encode an open reading frame (ORF) with homology to reverse transcriptase (RT) (Michel and Lang, 1985 and Toor and Zimmerly, 2001). Among them first identified and studied were the yeast mtDNA introns *coxI*-I1 and -I2 (also known as aI1 and aI2, respectively), which are found at two sites in the cytochrome oxidase subunit I gene (Bonitz *et al.*, 1980).

The first bacterial group II introns were discovered about ten years ago by PCR screens (Ferat and Michel, 1993). Only recently, The recent genome sequencing revealed that group II introns are surprisingly common in both gram-negative and gram-positive bacteria (Dai and Zimmerly, 2003; Toro, 2003). Group II introns are rare in archaea, being found in only 2 closely related species out of 17 whose genomes have been sequenced (Rest, 2003). In other example, group II intron was discovered in a relaxase gene (*ltrB*) in a conjugative element from *Lactococcus lactis* Ll.LtrB, where its splicing is required for conjugation (Mills *et al.*, 1997 and Shearman *et al.*, 1996).

1.5.2 Splicing mechanism and structure of group II introns

The group II introns splice via two sequential transesterification reactions that yield ligated exons and an excised intron lariat with a 2'-5' phosphodiester bond. At first, nucleophilic attack at the 5'-splice site by the 2'-OH of bulged A-residue in DVI results in cleavage of the 5'-splice site coupled to formation of lariat intermediate (Figure 1.3 B). Finally, nucleophilic attack at the 3'-splice site by the 3'-OH of the cleaved 5' exon results in exon ligation and release of the intron lariat (Figure 1.3 A) (Peebles *et al.*, 1986, Schmelzer and Schweyen, 1986 and Vanderveen *et al.*, 1986). In the case of group II introns, however, the splicing reactions are catalyzed by the intron RNA itself.

To accomplish this, the RNA folds into conserved secondary and tertiary structures, which form an active site containing catalytically essential Mg^{2+} ions (Michel and Ferat, 1995 and Qin and Pyle, 1998). The conserved secondary structure consists of six double-helical domains (DI-DVI) radiating from a central wheel (Figure 1.3 B). DV interacts with DI to form the minimal catalytic core; DVI contains the branch-point nucleotide residue, usually a bulged A and DII and DIII contribute to RNA folding and catalytic efficiency (Fedorova *et al.*, 2003). DIV, which encodes the intron ORF, is dispensable for ribozyme activity. NMR and X-ray crystal structures have been determined for segments DV, DV and DVI (Sigel *et al.*, 2004 and Zhang and Doudna, 2002), enabling structural modeling of the active site (Costa *et al.*, 2000, Noah and Lambowitz, 2003 and Swisher *et al.*, 2001).

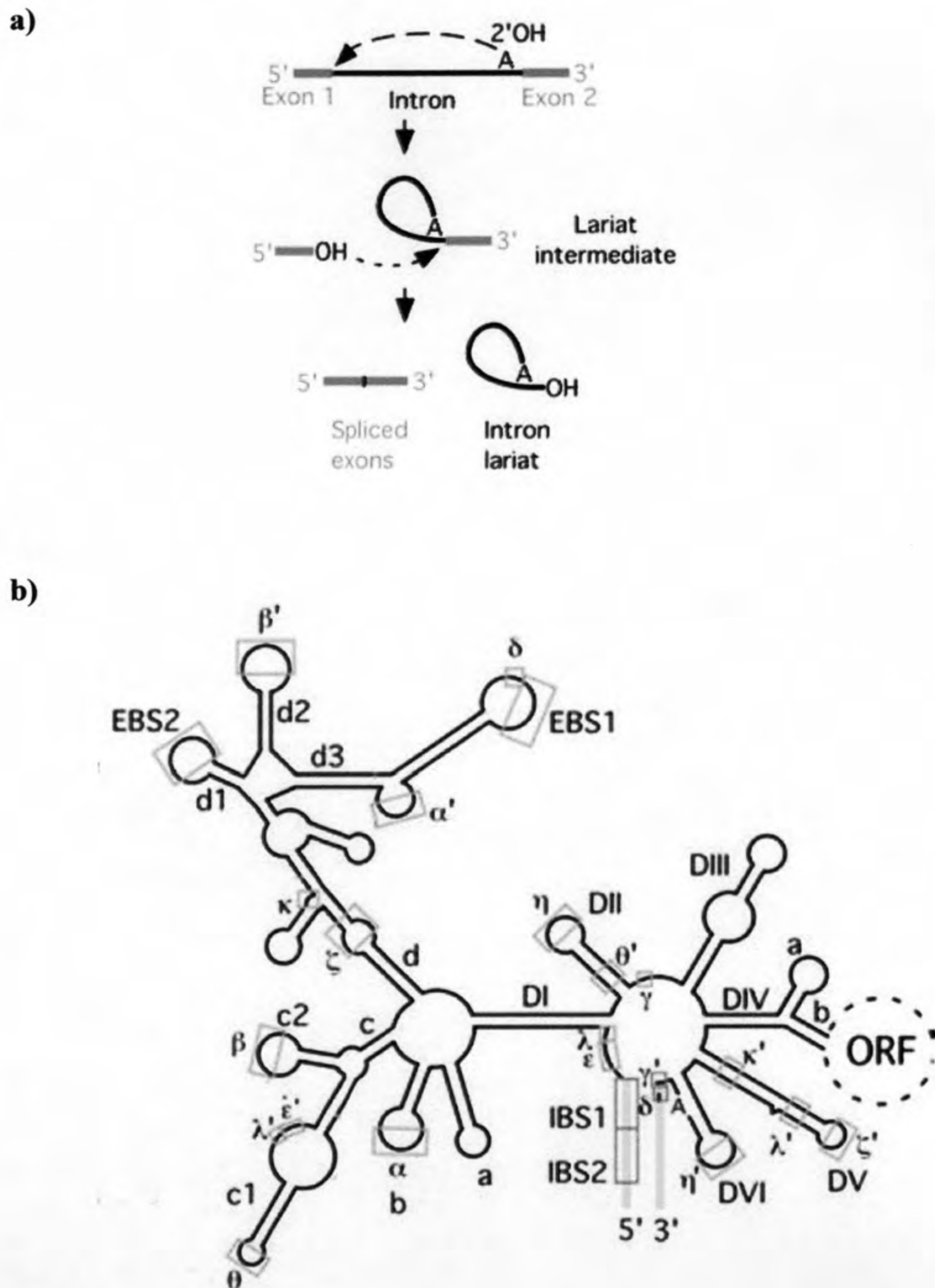

Figure 1.3 Group II introns spicing mechanism and secondary structure

a) Spicing mechanism of group II introns

b) Secondary structure of group II introns

1.5.3 Retrohoming reaction

Mobile group II introns are retroelements that encode reverse transcriptase (RTs) and insert into specific DNA target sites at high frequency by a process termed retrohoming (Lambowitz *et al.*, 1999 and Belfort *et al.*, 2002). The retrohoming reactions are carried out by a ribonucleoprotein (RNP) complex that is formed during RNA splicing and consists of the intron-encoded protein (IEP) and the excised intron lariat RNA (Zimmerly *et al.*, 1995). This RNP initiates mobility by recognizing a specific double-stranded DNA target site, with both the IEP and base pairing of the intron RNA contributing to DNA target site recognition (Figure 1.4) (Guo *et al.*, 1997). The IEP first recognizes a small number of `fixed' positions in duplex DNA, triggering local DNA unwinding, which enables the intron RNA to base pair to the adjacent 14-16 nt DNA sequence (Singh and Lambowitz, 2001). These base pairing interactions involve three short sequence elements in the DNA target site [intron-binding sites 1 and 2 (IBS2 and IBS1) and δ'], which are recognized by complementary sequence elements [exon-binding sites 1 and 2 (EBS1 and EBS2) and δ' in the intron RNA (Figure 1.4). The intron RNA then inserts directly into one strand of the DNA by reverse splicing, while the IEP cleaves the opposite strand and uses the cleaved 3' end as a primer for reverse transcription of the inserted intron RNA. The resulting intron cDNA is integrated into DNA target site. (Cousineau *et al.*, 1998).

Because most of the DNA target site is recognized by base pairing, the intron-insertion site can be controlled simply by modifying the intron RNA. From this strategy, genomic DNA can be modified. The *Lactococcus lactis* L1.LtrB intron, used as a model for integration into desired DNA target site (Kerberg *et al.*, 2001),

has a wide host range. Group II intron being efficiently mobile both in its gram-positive host and in gram-negative (Cousineau *et al.*, 1998). In *Lactococcus lactis* L1.LtrB intron system, the IEP, denoted LtrA, has RT activities. The mechanism of retrohoming of L1.LtrB intron is shown in Figure 1.4.

1.5.4 Group II intron as gene-targeting vectors

In bacteria, retargeted group II intron are generally expressed from a donor plasmid (in this study, pACD4K-C and pACD4K-C-loxP). Intron donor plasmid contains a 0.9-kb Ll.LtrB-ΔORF derivative of the *L. lactis* L1.LtrB intron with flanking exon, cloned behind a T7 promoter. The ORF encoding the IEP (denoted LtrA protein) is deleted from its original location in intron DIV and expressed from a position just downstream of 3' exon. The protein expressed from this position promotes efficient RNA splicing and intron mobility, leading to insertion of the ΔORF intron at a new DNA location. In the absence of the IEP, the ΔORF intron is unable to splice after insertion at new location (Zhong *et al.*, 2003).

The Ll.LtrB intron is targeted to desired sites with the aid of a computer algorithm, which scans the target sequence for the best matches for positions recognized by the IEP and then designs PCR primers for optimal base-pairing interactions between the intron RNA's EBS1, EBS2, and sequences in the DNA target site, IBS1, IBS2, (Perutka *et al.*, 2004). In *E. coli*, targetrons commonly insert at frequencies of 1%–80%, and insertions are detected either by colony PCR or by incorporating a genetic marker within the intron (Zhong *et al.*, 2003 and Perutka *et al.*, 2004). Thus, targetrons can potentially be used to disrupt any gene for which a

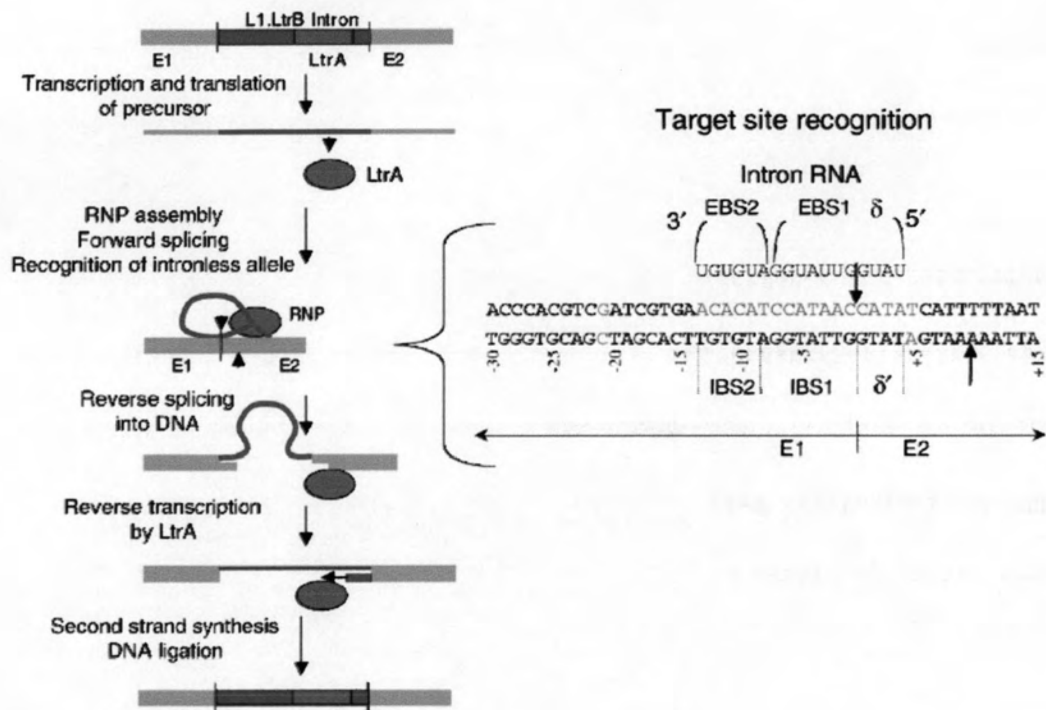

Figure 1.4 Retrohoming of the *Lactococcus lactis* L1.LtrB intron

sequence is available. A targetron derived from the *L. lactis* Ll.LtrB intron has been used for efficient targeted gene disruption in both gram-negative and gram-positive bacteria (Frazier *et al.*, 2003, Karberg *et al.*, 2001, Perutka *et al.*, 2004 and Zhong *et al.*, 2003) and minimally dependent on host factor.

1.6 Cre recombinase

Cre recombinase of the P1 bacteriophage efficiently catalyzes recombination between two of its consensus 34 base pair DNA recognition sites (loxP sites) in any cellular environment and on any kind of DNA. By adding the Cre recombinase system, we can excise or invert loxP-flanked DNA segments or create intermolecular recombination between different DNA molecules. By combining all of these tools, we are capable of creating any desired modification of the any genome, from introducing specific point mutations to large site-specific chromosomal aberrations, can be performed.

The Cre recombinase of the P1 bacteriophage belongs to the integrase family of site-specific recombinases. It is a 38 kDa protein that catalyzes the recombination between two of its recognition sites, called loxP (Hamilton and Abremski, 1984). This is a 34 bp consensus sequence, consisting of a core spacer sequence of 8 bp and two 13 bp palindromic flanking sequences. One of the great advantages of the Cre/loxP recombination system is that there is no need for additional co-factors or sequence elements for efficient recombination regardless of the cellular environment. Concerning the molecular mechanism of recombination, a single recombinase molecule binds to each palindromic half of a loxP site, then the recombinase

molecules form a tetramer, thus bringing two loxP sites together (Voziyanov *et al.*, 1999). The recombination occurs within the spacer area of the loxP sites (Figure1.5).

The bacteriophage P1 derived Cre-loxP system is a powerful and versatile tool for *in vivo* DNA recombination. It is widely used in gene targeting research. This system is also made it possible to delete gene flanking by loxP site.

In this study, to improved L-alanine production in *E. coli* Bl21(DE3) from heterologous gene expression and co-transformation under T7 system (Hatrongjitt, 2004) between alanine dehydrogenase gene and formate dehydrogenase gene, alanine racemase genes (*alr* and *dadX* gene) were inactivated. Group II intron was choosen for gene disruption.

The outline of this research

1. To inactivate *alr* gene in *E. coli* BL21(DE3)
2. To inactivate *dadX* gene in *E. coli* BL21(DE3)
3. To inactivate *dadX* and *alr* genes in *E. coli* BL21(DE3)

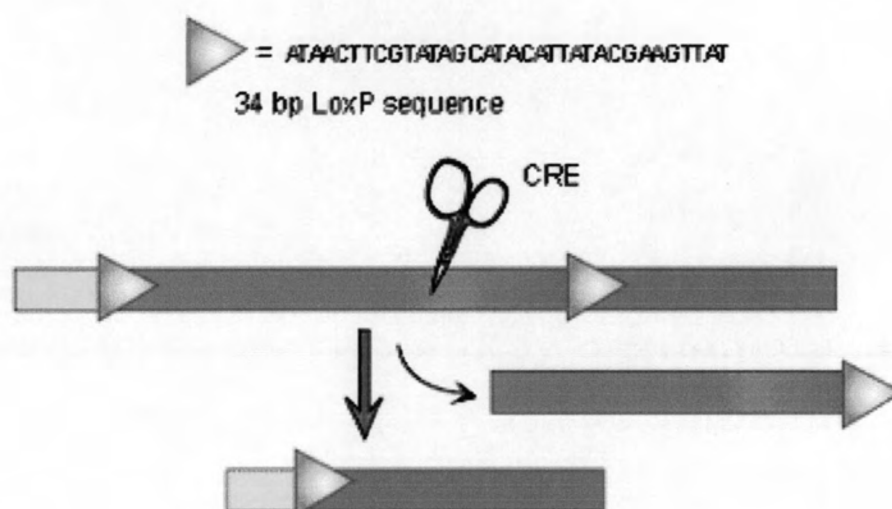

Figure 1.5 Simple mechanism of Cre/loxP system

When Cre-recombinase is expressed in the cell, the loxP site will be cut and joined together, removing the piece of DNA between the two sites.