CHAPTER I

INTRODUCTION

RNA interference (RNAi) in plants

Several years ago plant molecular biology advances were rapidly developed which lead to the discovery of RNA interference (RNAi). In plant, the RNAi phenomenon was first unfolded serendipitously during a search for transgenic petunia flowers that were expected to overexpress the *CHS* gene which encode a key enzyme (chalcone synthase) in anthocyanin biosysthesis. Surprisingly, it resulted in down-regulation of the antocyanin level (Napoli, 1990). Subsequently, it was known that plant transformation with a transcribed sense transgene could shut down the expression of a homologous endogenous gene. Not only in plant but it was also discovered in animals and fungi. The natural functions of RNAi and its related processes seem to be protection of the genome against invasion by mobile genetic elements such as viruses and transposons as well as orchestrated functioning of the developmental programs of eukaryotic organisms.

Mechanism of RNA interference

RNA interference (RNAi), also known as post-transcriptional gene silencing (PTGS) or co-suppression is a homology-based phenomenon that is initiated by the formation of double-stranded RNA (dsRNA) and leads to inhibition of gene expression at post-transcriptional level (Waterhouse, 1998). The dsRNA is recognized and cleaved by Dicer into small 21–25 nucleotide fragments, called "small interfering RNA (siRNA)". The siRNA is then incorporated into the RNA-induced silencing complex (RISC) which is subsequently targeted to degrade messenger RNA (mRNA)

that has a region complementary to the siRNA sequence (Hamilton and Baulcombe, 1999; Hutvagner and Zamore, 2002; Matzke et al., 2004). (see in Figure 1.1)

Different classes of RNAs

The most abundant types of RNA inside a cell include ribosomal RNA (rRNA), transfer RNA (tRNA) and messenger RNA (mRNA). In addition, there are double-stranded RNA (dsRNA), hairpin RNA (hpRNA), small interfering RNA (siRNA) and micro RNA (miRNA). Only the last four classes of RNA take active roles in RNAi.

The first of the classes of RNA taking an active role in RNAi is dsRNA, which is formed by complementary base pairing of two single-stranded fragments of RNA (Agrawal *et al.*, 2003). Found naturally in the cell, long dsRNAs generally derive from such events as transposition (Schramke and Allshire 2004) or virus induction (Rovere *et al.* 2002; Marathe *et al.* 2000).

Second, hpRNA is simply another form of dsRNA derived from a long piece of single-stranded RNA containing inverted repeats and a hairpin loop connecting them (Wesley *et al.* 2001).

Third, siRNA occurs when the long dsRNA is cut into smaller fragments. It is coined due to their use as a targeting sequence by RISC which aims at mRNA for degradation. It is composed of 21-25 bp with a 3′overhangs of two nucleotides. In addition to 3′ overhangs of two nucleotides, it also has 3′-hydroxyl and 5′-phosphate termini (Lipardi *et al.*, 2001).

Fourth, miRNA which closely resembles siRNA (Bartel 2004; Nelson *et al.* 2003). Both are short (generally in the range of 25 bp) and are used as targeting sequences aimed at degrading a specific mRNA. The miRNAs are endogenous and

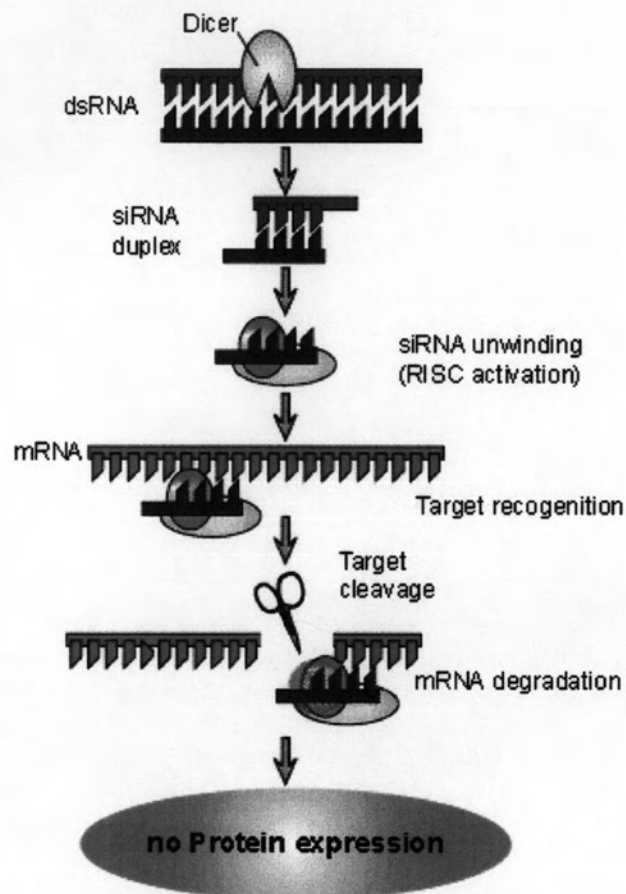

Figure 1.1 The mechanism of RNA interference (Michael *et al.*, 2006)

appear to represent a novel way of regulating gene expression during development.

Hairpin RNA (hpRNA)

Hairpin RNA (hpRNA) is formed as a result of the folding back of two closely positioned complementary sequences from a single-stranded RNA molecule (Wesley *et al.*, 2001; Smith *et al.*, 2000). This folding brings the two complementary sequences together where they will hybridize. hpRNA can be accomplished through sequences of inverted repeats where a normal sequence read in the 5′-3′ direction followed by the same sequence read in the 3′-5′ direction (Figure1.2). However, hpRNA may have a slight problem which is its lack of stability.

Stabilization of hpRNA can be accomplished through the introduction of a spacer sequence located between the two inverted repeat sequences (Wesley *et al.*, 2001; Smith *et al.*, 2000). The spacer sequence can be composed of any sequence, not complementary to the inverted repeats, and creates the loop structure of the hairpin. Recent studies illustrate that a much higher rate of silencing may be achieved, if the constructed spacer contains an established active intron sequence (Wesley *et al.* 2001). The use of an intron sequence as a spacer has been termed intron-hairpin RNA (ihpRNA). In the previous study, Smith *et al.* (2000) and Stoujesdijk *et al.* (2002) reported that the use of ihpRNA further increased the silencing efficiency to nearly 100% of transformants. The reason for this added stability and silencing may be due to the alignment of the two arms (inverted sequences) of the hpRNA during its docking at the splicing machinery. In other words, the splicing machinery creates a clean and perfectly aligned dsRNA sequence that exits the nucleus and initiates RNAi in the cytosol.

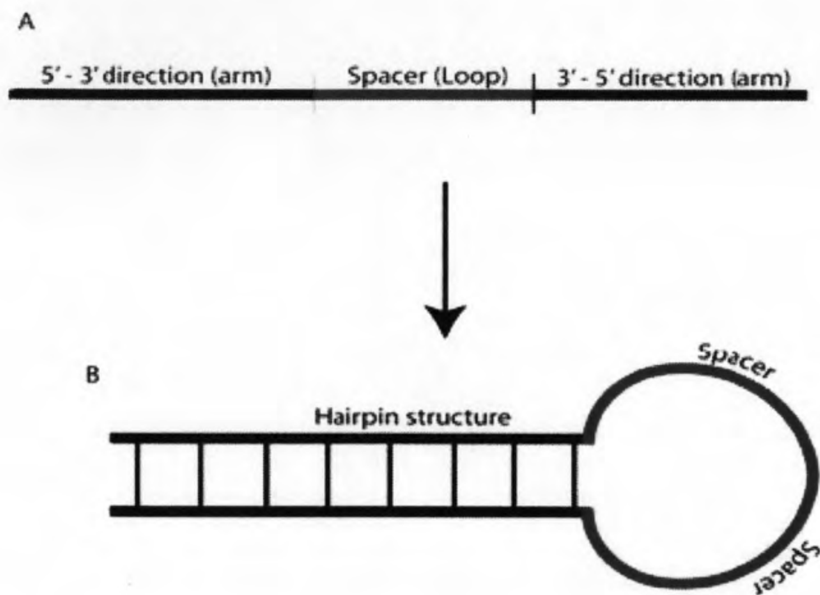

Figure 1.2 Diagram processing of ssRNA into hpRNA (Williams *et al.*, 2004) (A) A diagram of hpRNA prior to folding into the characteristic hairpin structure. It has two sequences in inverted repeat orientation with a spacer in between. The two sequences compose the arms and the spacer composes the loop. (B) The unfolded sequence in (A) folded into a hairpin structure.

In addition, scientists try to enhance silencing strength by placing an inverted repeat downstream of a strong, constitutive promoter within a binary vector (Figure 1.3). This approach has been found to increase the strength of RNA silencing when compared with traditional PTGS methods (Wesley *et al.*, 2001).

Dicer

Dicer first discovered by Bernstein *et al.* (2001) in *Drosophila*, is a complex enzyme belonging to the ribonuclease III (RNase III) family that show specificity for dsRNAs and cleave them with 3′ overhangs of 2-3 nucleotides and 5′-phosphate and 3′-hydroxyl termini (Elbashir, 2001) as shown in Figure 1.4.

RNA-induced silencing complex (RISC)

RISC first discovered by Hammond *et al.* (2000) in *Drosophila*, is the multi-protein siRNA complex (Figure 1.5) that seperates dsRNAs and binds short antisense RNA strands which are then able to bind complementary strands of target mRNA. When it finds a complementary strand, it activates ribonuclease activity and cleaves the target RNA. (Gregory, 2005; Preall *et al.*, 2006; Sen *et al.*, 2005).

RNAi as tool for gene function analysis in plants

Recently, plant genome and EST sequencing efforts have yielded an abundance of genetic information, including the complete genomic sequences of *Arabidopsis thaliana* and rice. Genome sequencing projects are also currently under way for maize, *Nicotiana tabacum, Medicago truncatula* and *Lotus japonicus*, among others. In addition, a large number of EST sequencing projects are being pursued in a variety of plants (Louisa, 2004). This wealth of information has created a need for

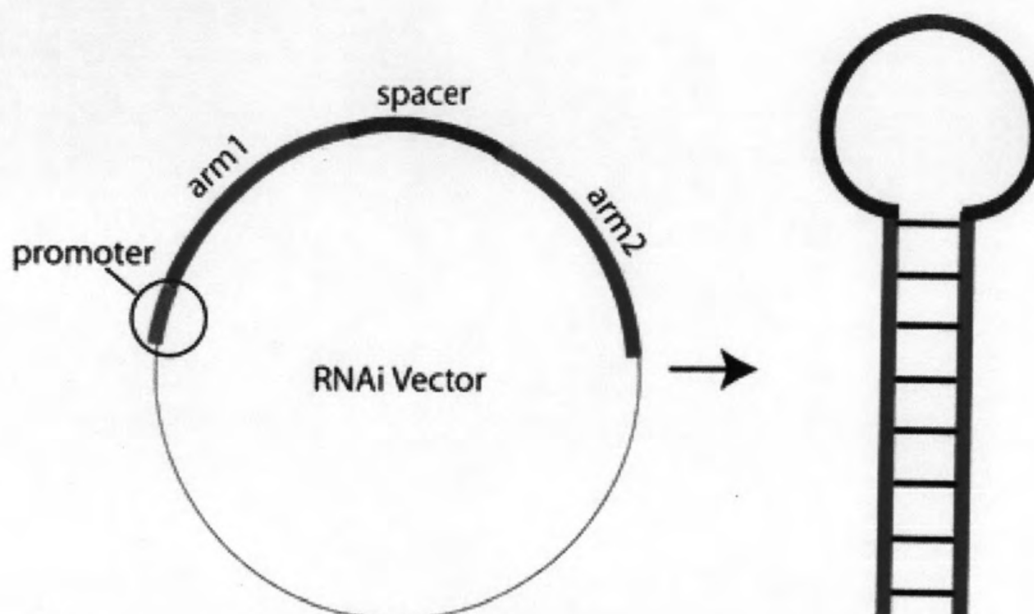

Figure 1.3 Diagram representing DNA-directed RNAi. An RNAi vector contains a promoter region (green), two arms (red), and a spacer (blue). As RNA polymerase transcribes the RNAi vector, it will produce hairpin structures as shown on the right (Williams *et al*., 2004).

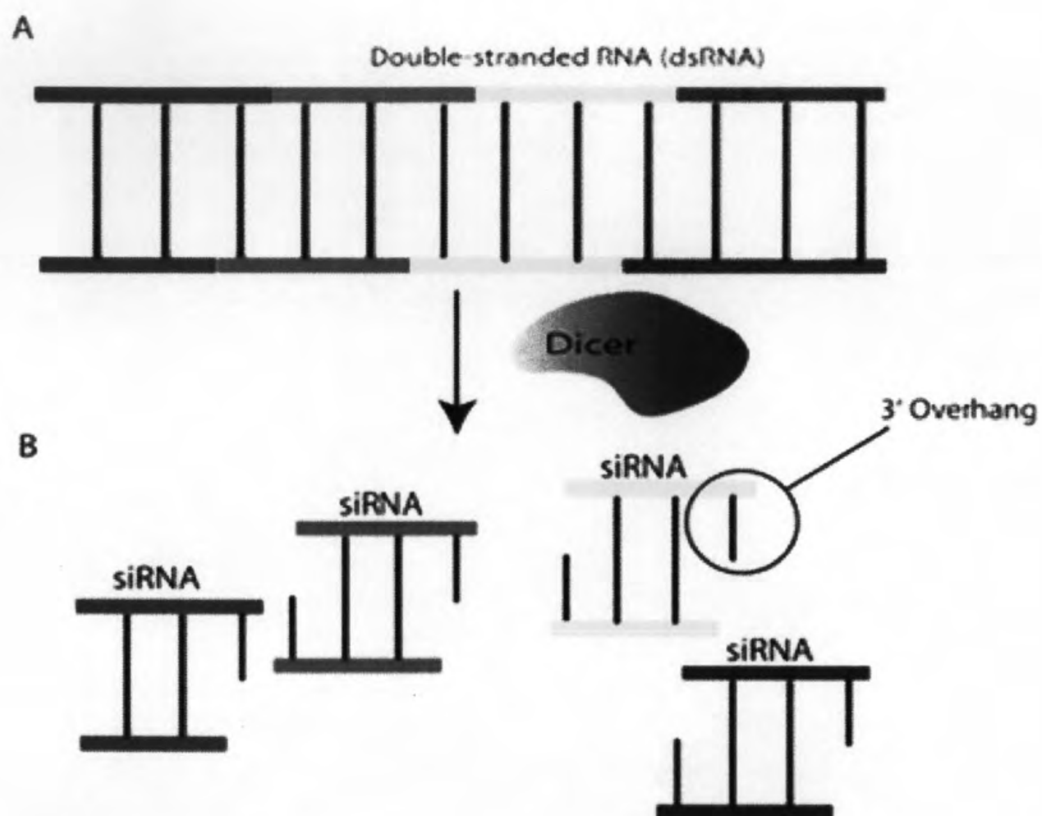

Figure 1.4 Diagram processing of dsRNA into siRNA by Dicer (Williams *et al.*, 2004)

(A) A visual representation of a fragment of long dsRNA. Colored fragments represent siRNAs prior to cutting by Dicer.

(B) A visual representation of siRNAs after Dicer has excised them from dsRNA. Colored fragments represent their position in the dsRNA.

Figure 1.5 Structure of RNA-induced silencing complex (RISC)

efficient, high-throughput methods to utilize such data for functional genomic analyses.

Current techniques for functional genomic analysis include chemical mutagenesis using ethylmethane sulphonate (EMS), radiation treatment mutagenesis, insertional mutagenesis using transposons or T-DNA, antisense RNA suppression and RNAi (Bevan, 2002). Chemical, radiation treatment and insertional mutagenesis approaches have been applied extensively to the functional characterization of plant genes, and many successful examples of their uses have been published. However, these approaches have disadvantages stemming from the randomness of mutagenesis. It cannot be targeted to specific plant DNA sequences, so the mutated loci must be sequenced or mapped to confirm their correct positions. Another disadvantage, mutation can not be used to study the functional genomic analysis, if the interesting gene is necessary for the survival of organisms. Antisense RNA suppression was an early form of RNA silencing employed mainly by plant scientists. This process involves the introduction into the cell of the antisense strand of RNA that corresponds to the target mRNA, the transcript intended to silence (Brant, 2002; Knee and Murphy, 1997). After entry into the cell, the introduced antisense RNA and the native target mRNA would bind via complementary base pairing preventing translation of the mRNA. This is due to the inability of ribosomes to bind to dsRNA (Arenz and Schepers, 2003; Brant, 2002). This process, however, did not always result in a loss of function of the targeted gene and the early research found that it has lower efficiency than RNAi. Therefore, RNAi by production hpRNA was widely used as a tool for gene function analysis in plants.

In comparison with chemical mutagenesis using, radiation treatment, T-DNA or transposon insertion and antisense suppression approach, there are a number of

advantages for using RNAi to produce silencing (loss-of-function) mutants, especially in plant species with a large-size genome. First, RNAi allows targeted and effective silencing of specific genes at a high frequency without random and laborious screening of loss-of-function mutants from large mutant populations. Second, simultaneous suppression of redundant or homologous genes, for example multiple members of a gene family can be achieved with RNAi (Lawrence and Pikaard, 2003). Third, inducible RNAi may provide an effective way for functional analysis of genes whose mutation will lead to embryonic or early developmental lethality (Guo *et al.*, 2003). Furthermore, a large population of gene silencing lines can be generated through high throughput RNAi (Wesley *et al.*, 2001; Brummell, 2003), which will complement other mutagenesis approaches for both forward and reverse genetics-based functional genomic studies.

Use of transgenic plants for characterizing gene functions

Plant transformation mediated by *Agrobacterium tumefaciens*, a soil plant pathogenic bacterium which naturally infects only dicotyledoneous plants, is widely used to create transgenic plants because it has several merits compared with direct gene transfer methods such as particle bombardment, and electroporation. The advantages are (1) stable gene expression because of the insertion of a foreign gene into the host plant chromosome (2) low copy number of the transgene and (3) large size DNA segments can be transferred. In addition, it is a single cell transformation system not forming mosaic plants, which are more frequent when direct transformation is used (Ensiquez-Obregon *et a*l., 1998).

Agrobacterium carries three genetic components required for plant cell transformation. The first two components which are located on the large Ti (Tumor-

inducing) plasmid are the T-DNA (Transferred DNA) and the virulence (*vir*) region (Figure 1.6) (Binns and Thomashow, 1988). The T-DNA is the DNA fragment that can be transferred from *Agrobacterium* to the plant cell. The *vir* region is organized into six complementation groups that are either absolutely essential for (*virA, vir B, vir D and virG*) or that enhance the efficiency of (*virC and virE*) plant transformation. The third bacterial component of the T-DNA transfer process resides in the *Agrobacterium* chromosome. Three chromosomal virulence loci, chvA and chvB and pscA, encode products involved in the binding of *Agrobacterium* to plant cells during the infection process (Zambryski, 1988).

During infection by *Agrobacterium*, T-DNA is transferred from the bacterium to the plant cell (Figure 1.6). The T-DNA is delimited by 25-bp direct repeats that flank the T-DNA. Any DNA between these borders will be transferred to a plant cell (Walden, 1993). Wild-type T-DNA encodes enzymes for the synthesis of the plant growth regulators; auxin and cytokinin, and the production of these compounds in transformed plant cells results in the tumorous phenotype. In addition, wild-type T-DNA also encodes enzymes for the synthesis of novel amino acid derivatives called opines. The Ti-plasmid encodes enzymes for their catabolism; hence, *Agrobacterium* has evolved to genetically commandeer plant cells and use them to produce compounds that they can uniquely utilize as a carbon/nitrogen source (Kahl and Weising, 1993).

The processing and transfer of T-DNA are mediated by products encoded by the *vir* region, which is also resident on the Ti-plasmid (Stachel and Nesyer, 1986). The *vir* genes, whose products are directly involved in T-DNA processing and transfer, are tightly regulated so that their expression occurs only in the presence of wounded plant cells, the target of infection. Control of gene expression is mediated

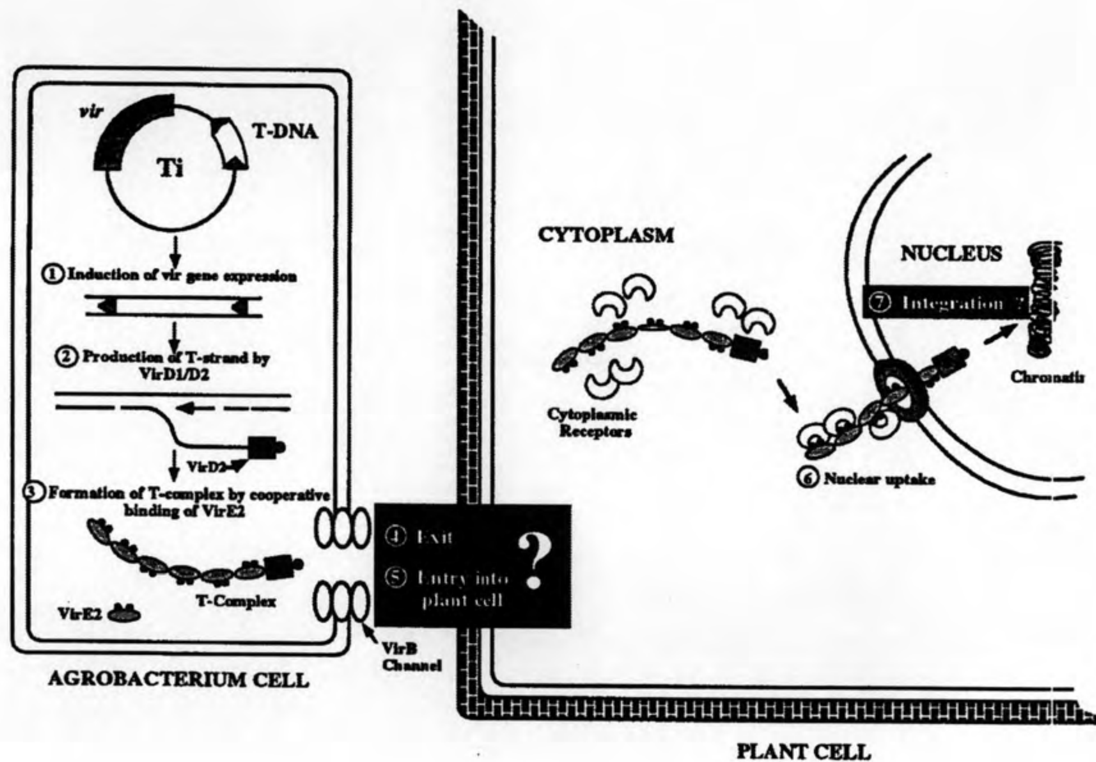

Figure 1.6 Basic step in the transformation of plants cells by *A. tumefaciens* (Zupan and Zambryski, 1995).

by the VirA and VirG proteins, a two compounds regulatory system. VirA detects the small phenolic compounds released by wounded plants resulting in VirG being activated and leading to transcriptional activation of other *vir* genes (Winans, 1992) (Figure 1.6, step 1).

Following *vir* gene induction, the production of a transfer intermediate begins with the generation of the T-strand, a single strand copy of the T-DNA (Stachel *et al.*, 1986). VirD1 and VirD2 are essential for this process (Filichkin and Gelvin, 1993). Together, VirD1/VirD2 recognizes the 25-bp border sequences and produces a single strand endonucleolytic cleavage in the bottom strand of each border (Figure 1.6, step 2). These nicks are used as the initiation and termination sites for T-strand production. T-strand production is thought to result from the displacement of the bottom strand of the T-DNA between the nicks (Zupan and Zambryski, 1995). After nicking, VirD2 remains tightly associated with the 5′ end of the T-strand. The lone VirD2 at the 5′ end gives the T-complex a polar character that may ensure that, in subsequent steps, the 5′ end is the leading end.

The T-strand must travel through numerous membranes and cellular spaces before arrival in the plant nucleus. Thus, to preserve its integrity, it was hypothesized that the T-DNA likely travels as a single stranded DNA-protein complex. VirE2 is an inducible single strand nucleic acid-binding protein encoded by the virE locus which binds without sequence specificity. VirE2 binds tightly and cooperatively, which means that a T-strand would be completely coated (Figure 1.6, step 3). Consequently, degradation by nucleases would be prevented and, indeed, *in vitro* binding of VirE2 renders single stranded DNA resistant to nucleolytic degradation. Finally, binding of VirE2 unfolds and extends single strand DNA to a narrow diameter of 2 nm, which may facilitate transfer through membrane channels. The T-strand along with VirD2

and VirE2 are termed the T-complex (Zupan and Zambryski, 1995).

Subsequencely, the T-complex must exit the bacterium cell (Figure 1.6, step 4) passing through the inner and outer membranes as well as the bacterial cell wall. It must then cross the plant cell wall and membrane (Figure 1.6, step 5). Once inside the plant cell, the T-complex targets to the plant cell nucleus and crosses the nuclear membrane (Figure 1.6 step 7). In the context of experimental chronology and relevant results, step 1 through 3 (Figure 1.6) have been studied. Current and recent research has related to step 4 and 6. Entering the plants cell (step 5) and the mechanics of integration (step 7) are almost completely uncharacterized (Zupan and Zambryski, 1995).

Reporter gene

Recently, new techniques of the introduction of DNA into plant cells have been developed. It is necessary to use a reliable gene-expression system to monitor DNA uptake, transcription and translation. The reporter genes are coding sequences that express in transgenic plants to provide a clear indication that genetic transformation did take place. In general, reporter genes should have the following characteristics (Crazzolara *et al.*, 1995; Alicja, 2001):

1) The gene products should not interfere with cell growth or metabolism.
2) The gene products should not be present in the organism or tissue under study.
3) The product of the reaction catalyzed by the reporter gene should be stable, easily detectable and quantifiable.
4) The gene products should be detectable in minimal quantities.

The four reporter gene systems that have been most widely used in experiments to transform plants and to perform assays with plant material are chloramphenicol acetyl transferase (CAT), luciferase, green fluorescent protein (GFP) and β-glucuronidase (GUS).

β-glucuronidase gene (GUS)

β-glucuronidase (GUS) has been the predominant reporter used to study gene expression in plants , especially in higher plants due to the absence of endogenous GUS activity in most plant species. The system which used the gene encoding β-glucuronidase, *gusA* (formerly *uidA*) from *Escherichia coli* was developed in 1987 by Richard Jefferson *et al.* The β-glucuronidase enzyme catalyses the hydrolysis of a wide variety of glucuronides. Its substrates consist of glucuronic acid conjugated through a β-O-glycosidic linkage to a range of aglycones (Figure 1.7). GUS has a monomeric molecular weight of approximately 68,000 Da, but exists *in vivo* as a tetramer. Higher plants transformed with a *gus* gene are healthy, develop normally and are fertile.

GUS is widely used as a reporter gene in plant because (1) absence of endogenous GUS activity (2) β-glucuronidase catalyses the hydrolysis of a wide variety of glucuronides and many substrates are commercially available. (3) The β-glucuronidase enzyme is very stable and will tolerate many detergents.

Because GUS is able to catalytically hydrolyze β-O-glycosidic linkages of a wide varity of glucuronides, and it is this substrate versatility that has been employed for either qualitative or quantitative assessments of patterns of gene expression. The GUS substrates that are commercially available find different applications due to the distinct chromogenic (histochemical), spectrophotometric and fluorimetric properties

of their aglycone forms, therefore, GUS activity can be measured by three basic methods. First, the most widely used of these is the colorless chromogenic substrate, 5-bromo-4-chloro-3-indolyl-β-D-glucuronide (X-Gluc). Cleavage of its glycosidic linkage by the action of GUS and subsequent oxidative dimerization/dehydrogenation of the resulting aglycone (Figure 1.7, reaction A) produce the characteristic indigo blue chromogenic precipitate that is extensively utilized to delineate patterns of gene expression in intact biological tissues. Both the intensity of color of the final dichloro-dibromo-indigo (ClBr-indigo) blue dye precipitate, and the precision of its spatial resolution in identifying the location of gene expression, are enhanced when ferricyanide/ferrocyanide salts are added. Second, albeit less commonly employed, GUS substrate is *p*-nitrophenyl-β-D-glucuronide (PNPG), which is typically used in standard spectrophotometric assays. In this case, GUS activities *in vitro* can be quantified by measuring the absorbance of the aglycone, *p*-nitrophenol (λ_{max} = 415 nm), which is released following GUS hydrolysis (Figure 1.7, reaction B). Third, substrate used for *in vitro* assays is 4-methylumbelliferyl-β-D-glucuronide (4-MUG). In this case, GUS hydrolysis generates the aglycone, 4-methylumbelliferone (4-MU) (Figure 1.7, reaction C), which gives a detectable fluorescent emission at 455 nm following excitation at 365 nm.

Objective of this research

With the completion of *Arabidopsis* and rice genome sequencing and the expanding crop sequence databases, the practical use of RNA silencing to reduce gene expression in a sequence-specific manner promises to be an essential and routine reverse genetics approach in plant functional genomics. Technologically reliable and high-throughput methods of RNA silencing are being developed by the recent progress

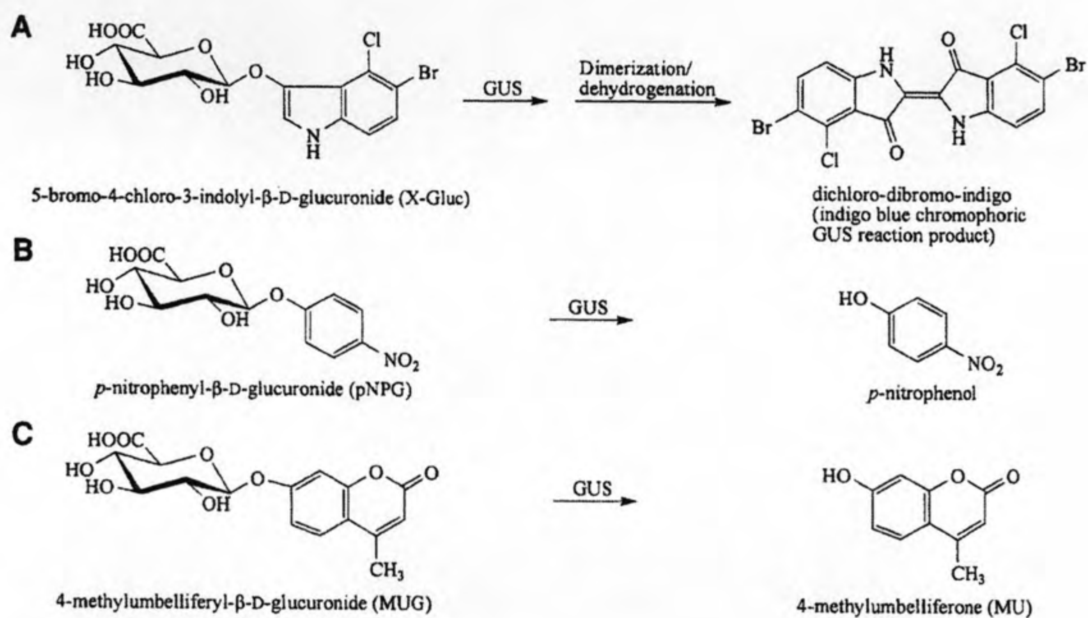

Figure 1.7 β-glucuronidase specifically hydrolyzes β-linked D-glucuronides to D-glucuronic acid and aglycone.

(A) 5-bromo-4-chloro-3-indilyl-β-D-glucoronide

(B) p-nitrophenyl-β-D-glucuronide

(C) 4-methylumbelliferyl-β-D-glucuronide

on the understanding of the core RNA silencing mechanism.

Therefore, the objective of this study is to silence a *gus* gene in transgenic tobacco by introducing a construct consisting of the hpGUS fragment driven by a 35SCaMV promoter into tobacco leaf discs (*Nicotiana tabacum* L. cv. Virginia Coker) by *Agrobacterium*-mediated transformation and to determine the expression level of the *gus* gene in transgenic plants compare with the control plants. The knowledge obtained from this study will provide us with a basic system for developing RNAi to be used as an efficient system for gene function analysis in the future.