Introduction

Advances in biological engineering technologies have allowed for improved genetic manipulation of plants. The advent of plant genetic engineering and novel gene expression vectors in the 1970s saw the creation of novel crop lines (Chilton, 1979). No longer limited to selective breeding, geneticists were able to introduce single genes directly into plants. The result has been economic and intellectual gain in the form of enhanced agricultural productivity, increased understanding of plant biology, and novel biotechnologies.

Genetically modified plants

Increased crop productivity has been achieved through resistance to pests (e.g., cotton plants expressing *Bacillus thuringiensis cry* genes, Perlak et al., 1990), herbicides (e.g., *Roundup Ready* soybeans, Monsanto Corp.), and environmental stresses such as soil deficiencies or osmotic stress (e.g., transgenic tobacco, Nuccio et al., 1999). Twenty million acres of transgenic crops were planted worldwide in 1998 (ISAAA, 1999). Plants functioning as bio-reactors have been developed for large scale pharmaceutical and chemical production. Unlike bacterial systems, plants are amenable to genetic manipulation to ensure proper eukaryotic protein processing (Cramer et al., 1999). Antigen production for vaccine development (Yusibov et al., 1997), alternative fuel production (Wilke, 1999), and bio-remediation (He et al., 2001) are other examples of transgenic plant applications.

Transformation technologies have also led to an increased understanding of fundamental plant processes through improved experimental design. Genes may be introduced, over-expressed, or silenced in order to characterize, for example, plant

development (Frugis et al., 2001), metabolism (McNeil et al., 2001), or defense response (Shah et al., 2001).

Producing genetically modified plants

Non-targeted methods

DNA may be introduced into plant genomes using a variety of methods. Biolistics involves bombardment of tissue with DNA-coated particles (Seki et al., 1999). DNA may be injected directly into plant nuclei (Holm et al., 2000) or introduced into protoplasts via electroporation (Arencibia et al., 1998). These strategies depend on tissue survival and successful regeneration and selection of transgenic plants.

A more efficient method, *Agrobacterium* mediated transformation (Willmitzer et al., 1983), does not involve time- and labor-intensive plant regeneration. This simple process takes advantage of the naturally infectious *Agrobacterium tumefaciens.* This soil bacterium harbors a Ti-plasmid that causes crown gall tumors in dicotyledonous plants (Zupan et al., 1995). The Ti-plasmid contains genes associated with uptake of plasmid DNA into the plant (*vir* family, Zambryski et al., 1992), plant growth regulation, and amino acid production and catabolism (Van Montagu et al., 1980). Infection results in a rapid growth of undifferentiated, conjugated-amino acid producing, plant tissue.

Multiple copies of a region of the Ti-plasmid (T-DNA) are integrated randomly into the plant genome (Ten Hoopen et al., 1999). The T-DNA encodes auxin, cytokinin, and opines, and is flanked by two border sequences that function as transfer signals. Excess amounts of auxin and cytokinin result in plant tumor growth. Opines are a class of conjugated-amino acid metabolized by the bacterium. Five *vir* genes orchestrate this

integration event. The *vir* genes of laboratory strains of *Agrobacterium* are usually engineered into the chromosome or a separate plasmid (binary vector system).

Investigators have taken advantage of this natural DNA transfer to introduce novel genes into plants (Willmitzer et al., 1983). The T-DNA genes are often replaced with a gene(s) of interest, a selectable marker gene, and a reporter gene. The *vir* genes mediate integration of the T-DNA cassette into the plant genome.

Each of these non-targeted transformation methods typically results in some number of plants stably harboring and expressing the gene(s) of interest. However, the random nature of the integration event yields transgenes that are especially susceptible to silencing as a result of chromatin condensation or hypermethylation (Meyer et al., 1994). Described as position effects , resulting transgene expression levels are highly variable (Gelvin, 1998). This calls for laborious screening times in order to identify transgene-expressing transformants. Insertional mutagenesis is also possible, with potentially lethal or deleterious results for the plant.

Transgene silencing in plants is particularly problematic. This broad-range defense response to viral pathogens and transposition events leads to gene inactivation. Aberrant viral mRNA is converted to a double stranded form, degraded, and used to signal methylation of the gene (Bender, 2001; Vance and Vaucheret, 2001). While details surrounding this mechanism continue to unfold, conditions inducing transgene silencing are well characterized (Matzke et al.; Paszkowski (ed), 1994). Homology (to endogenous genes or other trans-sequences), passage through generations, and multiple copies each contribute to gene inactivation. The multiple, complex insertion patterns yielded by non-targeted integration methods contribute to transgene silencing.

Targeted methods

Problems associated with position effects and multiple insertions may be alleviated by targeting of the transgene to a specific region in the plant genome. Two recombination-based strategies have been used to accomplish this, homologous and site-specific recombination.

An established tool in mammalian and yeast systems (Shapira et al., 1983, Folger et al., 1982), homologous recombination involves the introduction of extrachromosomal DNA with homology to endogenous sequence. Insertion occurs as a result of crossing over during mitotic division. Transgenic production via homologous recombination has not seen the same success in plants as in animal or fungal models, as targeted integration occurs at very low frequencies (e.g., 0.1%, Kempin et al., 1997; <0.1%, Halfter et al., 1992; 0.08%, Miao et al., 1995). Surprisingly, extrachromosomal DNA integrates almost exclusively at random, non-homologous sites (Mengiste et al., 1999). Cellular factors that promote homologous recombination are currently under investigation as facilitators of targeted transgene insertion (Gherbe et al., 2001).

Four site-specific recombination systems have been used in plants, Cre/*lox*, FLP/FRT, R/RS, and Gin/*gix* (Odell and Russell; Paszkowski (ed), 1994). Derived from yeast and bacteriophage, these are well characterized, two-component systems consisting of a recombination site and a recombinase. Genes have been both inserted and excised using each of these mechanisms (Dale and Ow, 1991). However, it seems excision is often favored over integration, and the use of transiently expressed recombinases and mutated integration sites are being investigated for routine production of transformants (Dale and Ow, 1991).

Goals of this project

To date, no reliable means of effectively targeting foreign DNA to the plant genome has been demonstrated. We hope to demonstrate site-specific transgene integration into plants using a tri-partite, human parvovirus-based system. A target integration sequence has been introduced into plants using *Agrobacterium* (Zabaronick, personal communication). The work presented here involves transformation of plants with a novel recombinase, and represents the second chapter of this system.

Biology of adeno-associated virus

Viral vectors for gene therapy

Gene therapy involves the introduction of functional, therapeutic genes to correct genetic disorders. Given their natural ability to deliver genetic material to human cells, viruses are attractive gene therapy vectors. Common viral vectors currently used in clinical trials include retroviruses (Iyama et al., 2001), adenovirus (Chung-faye et al., 2001), and adeno-associated virus type 2 (AAV-2, Keir et al., 2001). Vectors designed to target and treat cancers have also been developed (Samani et al., 2001). Retroviruses and AAV-2 integrate into the host genome, offering the potential for long term, constitutive expression of the therapeutic gene. AAV-2 is unique in that its genome integrates into a specific region on chromosome 19 (Kotin et al., 1990). Combined with a lack of associated disease symptoms, this targeted insertion event makes AAV-2 a particularly attractive viral gene therapy vector.

AAV-2 structure and genome organization

Classified as a dependent parvovirus, AAV-2 is a small icosahedral virus with a single copy of a single-stranded, 4.7 kb DNA genome. The capsid consists of repeating

units of three proteins, VP1 (90 kDa), VP2 (72 kDa), and VP3 (60 kDa). It has a diameter of approximately 20 nm. VP1 is translated from a 2.3 kb mRNA containing an intron from nt 1907 to nt 2200 (Fig. 1). VP2 and VP3 use two different initiation codons within a 2.3 kb mRNA containing an intron from nt 1907 to nt 2227 (Becerra et al., 1988). The p40 promoter controls both transcripts.

The p5 and p19 promoters control transcription of four nonstructural proteins, Rep78 (78 kDa), Rep68, Rep52, and Rep40 (Srivastava et al., 1983; Trempe et al., 1987). Rep78 and 68 are translated from a 4.2 kb mRNA containing a splice site from nt 1907 to nt 2227. Rep52 and 40 are translated from a 3.6 kb mRNA containing the same intron. Maintaining complementary termini, most parvoviruses have a high degree of secondary structure at the genome ends. Two T-shaped, inverted terminal repeats (ITRs) flank the AAV-2 genome (Spear et al., 1977).

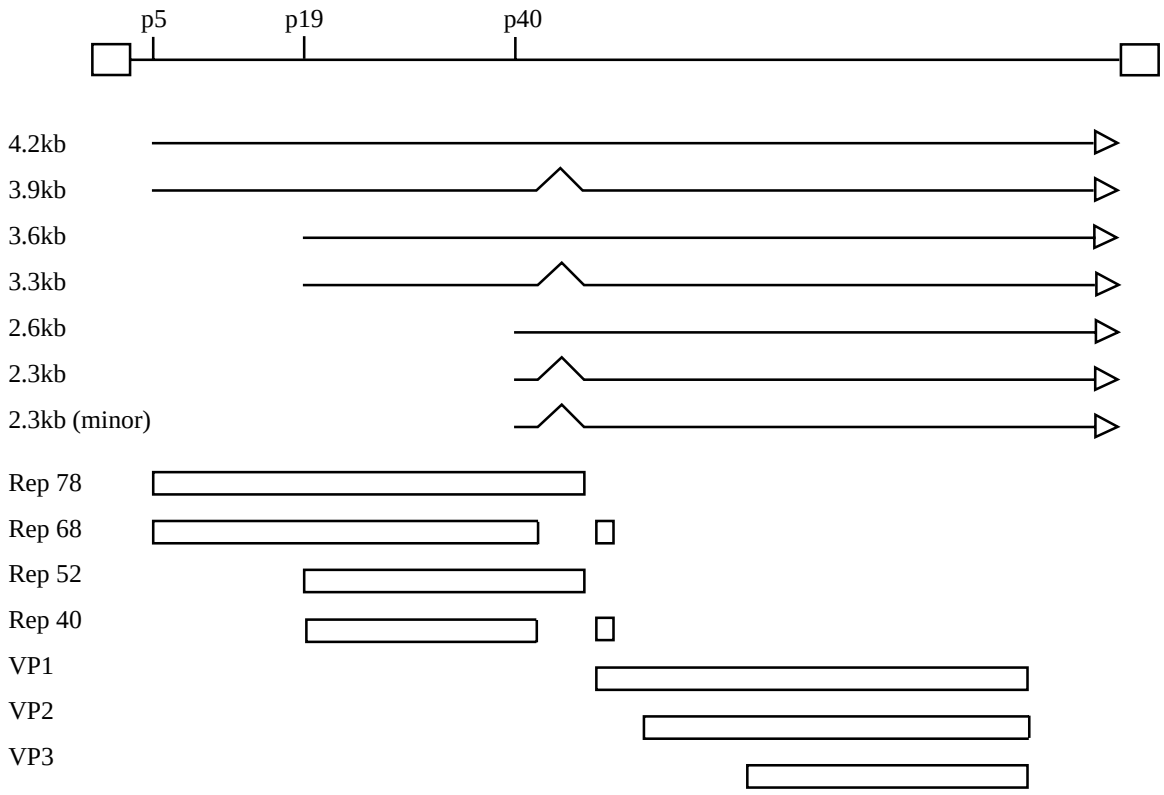

Figure 1. AAV-2 transcription map (adapted from Redmann et al., 1989). Viral mRNAs (lines), introns (carets), and corresponding proteins (open boxes) are shown. Including inverted terminal repeats (ITRs, open boxes), genome length is 4680 bp. Rep78/68 is involved in genome replication. Rep52/40 is involved in genome packaging. VP 1/2/3 are coat proteins.

AAV-2 life cycle

Heparan sulfate proteoglycans have been identified as primary receptors for AAV-2 (Summerford et al., 1998), with alpha5beta5 integrin (Summerford et al., 1999) and fibroblast growth factor receptor 1 (Qing et al., 1999) functioning as coreceptors. These are believed to be the primary determinants of infectivity. The virus enters the cell via a clathrin coated pit and proceeds through the endocytic pathway (Bartlett et al., 2000). Acidification of the late endosome is required for disruption of the vesicle and subsequent escape of virus. A Rac1/P13 kinase cascade is thought to mediate movement to the nucleus via microfilaments and microtubules (Sanlioglu et al., 2000).

As a dependent parvovirus, adeno-associated virus requires co-infection by a helper virus to replicate effectively. Once the ssDNA genome is in the nucleus, cellular factors are able to convert a small number of strands to a double stranded, replicative form (Yakobson et al., 1987). However, co-infection by adenovirus (Ad) greatly facilitates this conversion, thus ensuring a productive AAV-2 infection (Rose et al., 1972). Ad provides a DNA-binding protein that increases the processivity of replicating AAV DNA (Ward et al., 1998). This duplex replicative form is then converted to progeny monomers, dimers, and concatemers (Berns et al., 1979). Rep78/68 binds to, nicks, and unwinds the inverted terminal repeats, thereby resolving the hairpin structures and enabling replication (Im and Muzyczka, 1990). Single strands of both polarities are then shuttled to pre-formed, empty capsids in the nucleus by Rep52/40 (King et al., 2001). Both Ad and AAV particles accumulate until the cell lyses.

Alternatively, without co-infection by adenovirus, adeno-associated virus latently infects the host cell by integrating into the genome (Cheung et al., 1980). AAV-2 is

unique in its ability to integrate into a specific region of human chromosome 19 (Kotin et al., 1990). Termed AAVS1, this ~500 bp segment on the q arm of chromosome 19 contains a Rep78/68 binding site and cleavage site, nearly identical to those located on the AAV-2 ITRs (Weitzman et al., 1994, Berns and Linden, 1995). This replication-based, non-homologous recombination event results in multiple, head-to-tail copies of viral genome being inserted into AAVS1. A working model for integration has been proposed based on observed, Rep78/68-dependent AAV-AAVS1 junctions (Fig. 2; Dyall et al., 1999; Linden et al., 1996). A productive infection is established upon adenovirus infection, as Rep nicks the integrated genome and AAV is efficiently replicated.

Rep78/68

Mutational and complementation analyses involving the first 1900 nucleotides of AAV-2 internal to the ITRs identified an ORF required for replication of AAV-2 (Hermonat et al., 1984; Tratschin et al., 1984). Antibodies raised to synthetic oligopeptides identified the four Rep proteins from infected cell extracts (Mendelson et al., 1986). Rep78 and 68 are required for DNA replication (Hermonat et al., 1984; Tratschin et al., 1984), site-specific integration (Linden et al., 1996; Surosky et al., 1997), integrated genome excision (rescue), and promoter regulation (Kyostio et al., 1994) as shown by site-specific mutagenesis. *In vitro*, Rep78 and 68 exhibit biochemical activities consistent with these functions, including DNA binding (McCarty et al., 1994), helicase activity (Im and Muzyczka, 1990), and site-specific endonuclease activity (Im and Muzyczka, 1992). These and other studies have allowed for identification of functional domains within Rep78/68. (Fig. 3; McCarty et al., 1992; Walker et al., 1997; Yang et al., 1992). The N terminus is involved in site-specific binding and nicking (Urabe et al.,

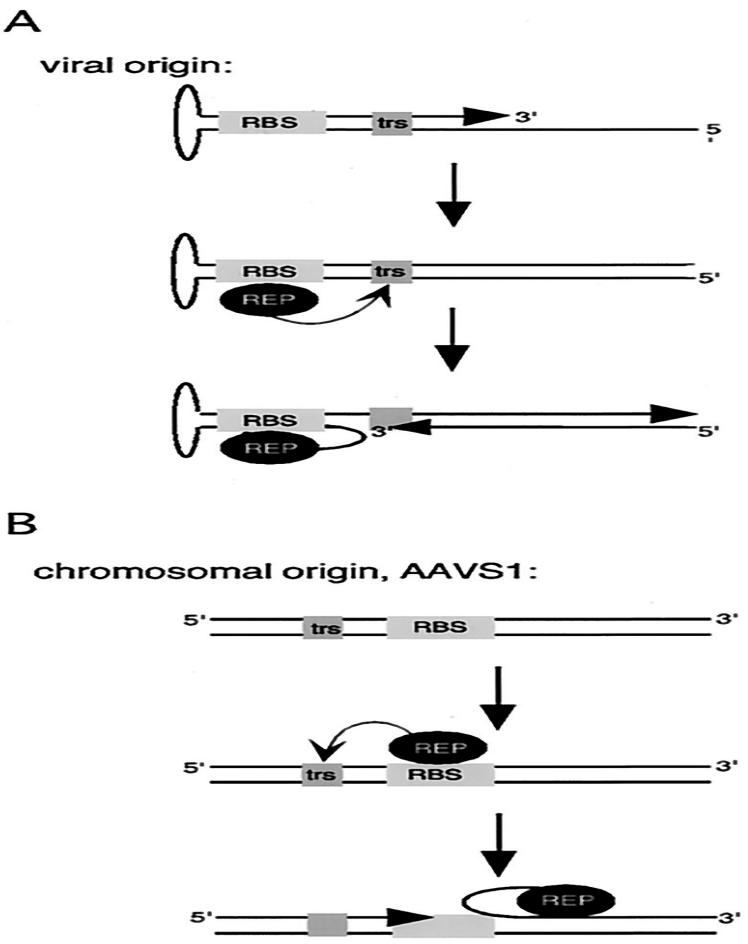

Figure 2. Integration of AAV-2 into AAVS1 (taken from Yoon et al., 2001). Rep binding (RBS) and cleavage sites (trs) within both AAV-2 (A) and AAVS1 (B, chromosome 19) are shown. Rep78/68 forms a complex with both origins simultaneously. Cellular replication machinery alternates between the proximal strands to recombine AAV-2 to AAVS1.

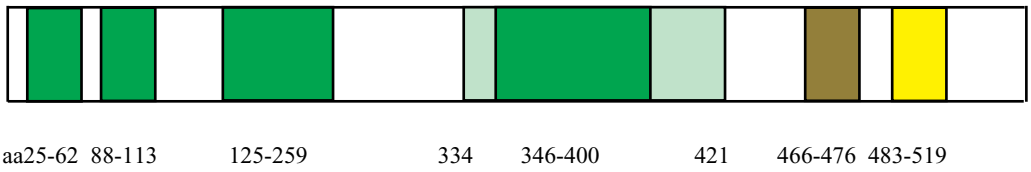

Figure 3. Rep78/68 functional domain map (adapted from Gavin et al., 1999). DNA binding (dark green), ATPase and helicase activity (light green, amino acids 334-421), oligomerization (brown), and nuclear localization (yellow) regions are shown.

1999; Davis et al., 2000), while the central portion of the Rep78/68 amino acid sequence is responsible for helicase activity (Walker et al., 1997) and site-specific endonuclease activity. The endonuclease domain has recently been crystallized and critical residues identified (Hickman et al., 2002). The C terminus consists of zinc finger motifs and a nuclear localization signal. Identification of residues critical to these biochemical properties has been difficult, as most Rep78/68 mutants compromise multiple Rep functions. Temperature sensitive mutants (Gavin et al., 1999) and domain swapping studies (Yoon et al., 2001) are currently being used to aid in biochemical characterization.

Rep78/68 has also been observed to regulate AAV-2 transcription. Without helper virus co-infection, Rep negatively regulates transcription from the p5 and p19 promoters (Beaton et al., 1989). In the presence of helper virus, Rep positively regulates transcription from all three AAV-2 promoters, p5, p19, and p40 (Pereira et al., 1997).

AAV-2 applications

Unlike autonomous parvoviruses, dependent parvoviruses are able to latently infect non-dividing cells, and AAV-2 does not exhibit a strictly defined tropism. While most people test seropositive for exposure to AAV-2, the virus is not associated with any symptoms of disease. With a well characterized, controlled ability to site-specifically insert DNA into the host genome, AAV-2 has been successfully used to introduce genes into patients. Clinical trials involving the transfer of the cystic fibrosis transmembrane conductance regulator (CFTR) gene to treat cystic fibrosis (Aitken et al., 2001) and the factor 9 gene to treat hemophilia (Larson et al., 2001) are two examples of promising results of AAV-2 gene therapy studies.

AAV-2 gene therapy technology to transform plants

Current plant transformation techniques, while effective, are sub-optimal. Transformation efficiencies of popular techniques can be as low as 0.48%, as with *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* (Clough and Bent, 1998). Combined with the highly variable transgene expression levels of these non-targeted techniques, this inefficient method involves a labor-intensive screening process. An ideal technique would involve a simple, efficient method of gene transfer, yielding a large percentage of transformed plants with predictable levels of transgene expression. We have designed an AAV-2 based plant transformation system to accomplish this.

The minimum requirements for AAV-2 integration into AAVS1 are 1) AAVS1, containing Rep binding and cleavage sites, 2) ITRs (inverted terminal repeats), also containing Rep binding and cleavage sites (Young and Samulski, 2001), and 3) Rep78/68 to orchestrate the insertion event. Plants harboring AAVSI and either Rep78 or Rep68 should, then, be able to target any ITR-flanked transgene to AAVS1.

The work presented here involves the introduction of Rep78 into *Arabidopsis thaliana*. Both wild type and AAVS1-containing plants were transformed with the full-length *rep78* using *Agrobacterium tumefaciens* in order to examine the feasibility of producing plants constitutively expressing Rep. This functional Rep will then be used to catalyze the integration of an ITR-flanked transgene into the AAVS1 target sequence.