Chapter two
Literature Review

2.1 *Cucumber mosaic virus*

Cucumber mosaic virus (CMV) is a virulent pathogen which severely infects a very diverse range of host encompassing more than 1000 plant species (Roossinck, 1999). The many different strains of this virus (Carrère *et al*, 1998) shows differences in pathology and thus is a serious menace to global agriculture.

In the family of *Bromoviridae*, *Cucumber Mosaic Virus* is the type species of the *Cucumovirus* genus and of the superfamily of alphavirus-like-viruses. The genus also includes *Peanut Stunt Virus* (PSV) and *Tomato Aspermy Virus* (TAV). CMV has long been divided into two Subgroups (I and II), according to its biological and serological properties and genomic homology (Gallitelli, 2000). More recently, Subgroup I has been classified into Subgroups IA and IB on the basis of phylogenetic analysis of the CP ORF and nucleotide sequence of the 5' nontranslated region of RNA 3 (Palukaitis and Zaitlin, 1997 ; Roossinck *et al*, 1999).

The genome of CMV comprises of tripartite plus sense single stranded RNAs with 5' cap structure and 3' conserved regions which can form into tRNA-like structures, separately encapsidated in isometric particles. These three genomic RNAs are designated RNA 1 (3.3 kb in length), RNA 2 (3.0 kb) and RNA 3 (2.2 kb) in the order of decreasing composition. The only monocistronic mRNA, RNA 1 encodes 1a protein, known as subunit of CMV replicase complex (Hayes and Buck, 1990), which is vital both for viral genome replication (Roossinck, 2002) and regulation of systemic infection (Kim *et al*, 2006). The RNA 2 codes two overlapping open reading frames (ORFs) which are translated into 2a and 2b proteins. The 2a protein is a viral RNA-dependent RNA polymerase subunit of replicase complex. The 2b protein is translated

from the low abundance subgenomic RNA 4a (Ding *et al*, 1994) which is transcribed from the small 3' ORF of RNA 2, only found in Subgroup II strains (Roossinck, 2001). This protein has been implicated not only in suppression of host post-translational gene silencing mechanism (Bonnet *et al*, 2005), but also in symptom severity (Salánki *et al*,2006). The RNA 3 is a bicistronic RNA encoding the movement protein (MP), and coat protein (CP) (Palukaitis and García-Arenal, 2003) expressed from subgenomic RNA 4, that is encoded by 3' ORF of RNA 3. Both proteins are responsible for the efficient of cell-to-cell movement of CMV infection (Canto *et al*, 1997) and CMV encapsidation, respectively (Roossinck, 2001)(Kim *et al*, 2006).

2.2 Application of scFv Antibodies in Detection of Plant Pathogen

2.2.1 Methods for detection of plant pathogens

As part of a primary disease management strategy against any plant pathogen, preventive measures relate closely to methods of detection. Detection relates to identifying the presence of pathogenic viruses and bacteria within a sample. Precise disease detection requires techniques of high specificity, sensitivity and efficiency (Lopez *et al*, 2003). In other words, specificity is defined as the ability to detect the pathogen without false positives and negatives while sensitivity relates to the minimum amount of pathogens detected per sample (Lopez *et al*, 2003).

Methods developed for virus detection vary and has dynamically evolved. Each technique has its advantages or disadvantages or limitations. Currently, serological and molecular techniques are the most commonly used and include techniques such as PCR, FISH (Fluorescence In Situ Hybridization), ELISA and TIBA (Tissue Blot Immunoassay) (Lopez *et al*, 2003; Webster *et al*, 2004). The molecular methods are known for its highly rapidity and specificity. However these methods not only require expert-labour but are also costly for large scale testing. Immunoassays such as ELISA

has proved to be a suitable plant pathology detection method for field purposes (Lopez *et al*, 2003).

To date, many companies still adopt the use of monoclonal antibodies for the detection of pathogens (Lopez *et al*, 2003). For example, Agdia (USA), Adgen (Scotland), Ingenasa (Spain), Durviv (Spain) and others. In contrast, the use of recombinant antibodies for detection of plant pathogens is still at the research level (Lopez *et al*, 2003). Several papers in recent years have shown that the scFv antibodies displayed similar detection characteristic as commercially-available monoclonal antibodies. Thus, the scFv is potentially a valuable reagent for detection systems. With vast technologies available now, the generation of scFv against any antigen could be possibly cheap, fast and easy.

2.2.2 Methods for detection of *Cucumber mosaic virus*

For the past three decades, advance in molecular biology, biochemistry and immunology has intensified the development of de novo, precise, rapid and less laborious methods for CMV detection. This major breakthrough can be categorized into two; protein based techniques and viral nucleic acid based techniques (Akinjogunla *et al*, 2008). Both serological and molecular techniques are applied for certain purposes : (i) characterization of CMV isolates to its subgroups, (ii) diagnosis of CMV as a cause of specific plant diseases, (iii) CMV detection for quarantine and certification purposes and (iv) virus population studies (Gallitelli, 2000). Most commonly used protein based techniques include ELISA, Dot Blot assay and Western Blot assay. While the viral nucleic acid based techniques include DNA/RNA hybridization, RT-PCR, RT-PCR RFLP and real time RT-PCR. Both groups mentioned possess its own advantages and disadvantages.

Although highly specific, sensitive and effective, RT-PCR and other nucleic acid based methods can be laborious for routine testing of large number of samples which need to be processed quickly. Moreover, these methods require bulky and expensive hardware such as microarray reader for fluorescent probe detection or thermocycler for polymerase chain reaction. The advantages of ELISA and other immunoblot assays are its sensitivity, simplicity, relatively inexpensiveness which make them suitable even in field conditions. The disadvantage of ELISA is that it may sometimes give false positive reactions while immunoassays in general require a certain minimum amount of sample for sensitivity.

2.3 Single chain Fv Fragment (scFv) antibodies

The scFv comprises two variable domains (V_H and light V_L) linked into a single polypeptide by a peptide linker ($Glycine_4Serine_3$). This hydrophilic linker is flexible, allowing the assembly of both domains to form an intact antigen binding site. The antigen-binding site of an antibody is usually conserved, even when only the V_H and V_L domains are used (Wörn & Plückthun, 2001).

Like most Ig domains, the variable regions contain two conserved cysteine residues forming a disulfide bridge (Glockshuber *et al*, 1992), as a consequence of spatial proximity in an oxidizing environment (De Jaeger *et al*, 1999). Therefore, a typical scFv contains two disulfide bridges, one in each variable region, needed to maintain its folding and antigen-binding properties (Brinkmann, 2010). In addition, disulfide bridges are required in some folding pathways and for the stability of certain 8folded polypeptides (Wörn and Plückthun, 2001 ; Brinkmann, 2010). For some specific scFv fragments, disulphide bond formation has been shown to be necessary for functional accumulation in *E. coli* (Glockshuber *et al*, 1992), and it has been suggested

that this is the case for most of the scFv fragments or any other antibody fragment (Plückthun, 1992).

Unlike whole antibodies, the scFv does not contain additional regions whose unfolding under force may give rise to structural changes (Rief *et al*, 1997a; Rief *et al*, 1997b) that might influence the unbinding event. The scFv fragments are particularly interesting models, because they can be generated against all conceivable antigenic targets and mutants with various binding properties (Ros *et al*, 1998).

2.4 Generation of scFv antibodies

The casettes for an antibody fragment are available from a variety of sources, which include the genetic material from non-immunized or immunized animal or humans and genes from a repertoire of naïve, semi-synthetic or synthetic genes (Daly *et al*, 2001).

With the development of libraries from these sources which includes recombinant phage display antibody libraries and ribosomal display libraries , it is now possible to generate the scFv *de novo* bypassing the traditional hybridoma technology and the need for immunizing animals (Blazek and Celer,2003).

In this study, the genes of V_H and V_L of scFv antibodies were isolated from spleen tissue of a mouse immunized with *Cucumber Mosaic Virus* coat protein (Chua and Othman, 2007). The DNA fragments of both chains were obtained via PCR using oligonucleotide primer mix, producing a pool of libraries with diverse of V_H and V_L genes. Subsequently, the DNA fragments were combined via synthetic peptide linker and inserted into a plasmid. The phage display technique enables the scFv to be expressed on the surface of recombinant phage. The biopanning of the scFv increases the selection of antibodies with high affinity and specificity to the CMV coat protein or

any given target, thus facilitates the *in vitro* isolation of best scFv construct from the libraries of V-genes (Chua *et al*, 2003; Padoa and Crowther, 2006).

2.5 Expression systems for the production of scFv antibodies

Various expression systems have been discovered in the production of scFv antibodies, such as bacterial, mammalian, insect, yeast, plant and *in vitro* translation systems. Each system has its own advantages and disadvantages. In this study, the scFv can be produced as functional antibody fragments without requiring post-translational modifications, unlike complex whole antibodies (Daly *et al*, 2001; Blazek and Celer, 2003).

2.5.1 Prokaryotic Expression System

One of the common methods of expressing scFv antibodies is the prokaryotic expression system. The advantages of using this system are the ease of genetic manipulation of bacteria, the nature of rapid growth of bacteria cells which lead to high yield, and cost-effective production (Verma *et al*, 1998; Daly *et al*, 2001). Thus, this method eases the purification, analysis and availability of the expressed scFv in a shorter time.

In order to obtain a functional soluble scFv antibodies, the gene expression casette must be constructed to have the appropriate sequences for proper transcription and translation. An inducible promoter (lac promoter) can be inserted to prevent mutation of the gene, if the production appears to be toxic to the bacteria, which is regulated by the lac repressor (Tabor and Richardson, 1985; Studier and Moffatt, 1986). Inducer concentration (IPTG) and incubation conditions also influence the level of secreted scFv (Daly *et al*, 2001). Additionally, the T7 RNA promoter is included to control the high level of expression, and start and stop codons are placed between the

cassette to avoid transcription of unwanted gene, adding stability to the DNA fragment (Verma *et al*, 1998). In order to bypass the *in vitro* refolding process, vector with leader sequence or signal sequence mediates the export of scFv into the periplasmic space (space between the inner and outer membrane of bacteria) in *E. coli*. The oxidizing environment of periplasm will naturally form the correct disulphide bond, due to the presence of chaperone-like molecules and disulphide isomerases (Padoa and Crowther, 2006). The signal sequence will be removed during the transport through the cytoplasmic membrane (Ferenci and Silhary, 1987). Thus, a soluble and correctly folded scFv fragment can be easily released from the periplasm either by lysis of cell, freeze-thaw or osmotic shock.

2.5.2 Plant Expression System

Plants are potentially very good bioreactor for the production of plantibodies. Plants can be used as temporary storage of scFv in leaves or for longer period of storage in seeds (Conrad and Fiedler, 1998). The highlights of trangenic plants are its ease of genetic manipulation, product safety as plants are neither host nor carrier if so involve human pathogens, cost-effective, and lastly possess eukaryotic cell machinery for complex protein processing (Fischer, 2003).

Among all the proteins produced in plants, the commonly studied are the monoclonal antibodies (mAbs) and antibody fragments (Smith and Glick, 2000). The level of expression depends on the the type of tissue, plant species and intrinsic properties of the antibodies (De Wilde *et al*, 1999).

The tobacco plant has been the principal model for research in this field, with *Agrobacterium-mediated* leaf disk transformation being utilized as the regeneration method to produce transgenic tobacco lines (Daly *et al*, 2001). This transformation method targets the T-DNA complex to the nucleus of the plant cells. The T-DNA

complex includes the expression construct design to optimize all stages of gene expression, from transcription to protein stability.

In the expression construct, the transgene is flanked with various regulatory sequences which conform to the KOZAK consensus for plant (Fischer *et al*, 2003). A strong constitutive plant promoter (CAMV35S) can be included to drive the scFv expression level and to control the antibiotic resistance gene and GUS gene separately.

For a complex multicellular organism, protein targeting is the most important in the design of expression construct. As most of the recombinant antibodies are more stable in certain intra-cellular compartment, this brings benefits for high yield, protein isolation and purification. Comparison studies have shown that the secretory pathway is a better environment for folding, assembly and post-translation modification (Schillberg *et al*, 1999). Insertion of signal peptide at the N-terminal brings the scFv to the golgi apparatus (Fischer *et al*, 2003) while insertion of retention signal (KDEL) at C-terminal retrives and brings the scFv to the endoplasmic reticulum (ER) (Conrad *et al*, 1998). The ER is an oxidizing environment with a number of molecular chaperones for the formation of intra- and inter-chain disulphide bond, thus a place for correct folding and maturity of nascent proteins (Twyman *et al*, 2003).

2.6 Downstream Processing

The realities in downstream processing in any given expression system demand for scalability. For commercialization, the protein needs to be purified before use. However, each system has its own limitations. Therefore, several strategies have been developed and evaluated on an empirical basis to optimize the downstream processing.

With fermentation, the modest protein expression of total cellular protein is 2-5 % in the bacterial cells (Wingfield, 2002). For plants, the yield of 0.1-1.0 % of total soluble protein is the typical level observed for production, which is competitive to

other expression systems as its cost is 2-10 % of microbial fermentation systems (Giddings, 2001).

Most of the laboratory-scale procedures often are too elaborate and expensive for large-scale efforts. The usage of certain reagents such as protease inhibitors and ammonium sulfate are not preferred beyond the bench scale, due to either the expense or toxicity (Larrick *et al*, 2002). In addition, the types of equipment used are often considered as the personnel in charge should not need to be an expert in handling and maintainence. Thus, minimalist approaches to large-scale downstream processing would lead to not only low costs but also to less errors.

The main prerequisite factor is the protein yield and stability, which is highly influenced by the conditions under which proteins are isolated and purified (Hwang *et al*, 2004). The speed of the process is also important to maintain protein stability and structure.

The isolation of protein in plants normally involves grinding while periplasmic proteins in bacterial cells can be released either by osmotic shock or enzymatic lysis or repeated freezing and thawing of cells (Wingfield, 2002). Clarification of the extract is either by filtration or centrifugation. Simplicity in the approach is important to mimic the commercial-scale production as most of the equipments commonly used are available in the food and beverage industry (Larrick *et al*, 2002). The introduction of the commonly used polyhistidine tag either on the N- or C- terminal of a protein does not interfere with the structure or function of the protein (Hwang *et al*, 2004). In fact, it facilitates the purification and minimilizes the contamination of recombinant proteins.

In conclusion, the requirement for downstream processing will depend on the choice of expression system, cost of production, simplicity in the approach and lastly, the degree of purity required.