

THESES, SIS/LIBRARY
R.G. MENZIES BUILDING NO.2
Australian National University
Canberra ACT 0200 Australia

Telephone: +61 2 6125 4631
Facsimile: +61 2 6125 4063
Email: library.theses@anu.edu.au

USE OF THESES

**This copy is supplied for purposes
of private study and research only.
Passages from the thesis may not be
copied or closely paraphrased without the
written consent of the author.**

GREEN FLUORESCENT PROTEIN-BASED ENHANCER TRAPS FOR RICE FUNCTIONAL GENOMICS

Xiqin Fu

Research School of Biological Sciences (RSBS)

The Australian National University

Canberra, Australia

And

The Center for the Application of Molecular Biology to

International Agriculture (CAMBIA)

Canberra, Australia

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

February 2004



Corrections page

1. P15, L15 Shimomura et al. 1962

Missing reference added: Shimomura, O., Johnson, F. H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. J. Cell. Comp. Physiol. 59, 223-239.

2. P19, L13 Replace 'transfected' with 'introduced'
3. P19, L16 Upadhyaya et al., 1998; 2000 (1998 added)
4. P28, L20 Hoekema et al, 1983 (missing year)
5. P45, L8 Sanger et al., 1977

Missing reference added: Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.

6. P66, L7 Approximate number of plants obtained: 500
7. The following citations can be found in the references: Chilton et al., 1974 (P50); Chu et al., 1975 (P53); Omirulleh et al., 1993 (P74); Wilson et al., 1995 (P70)
8. FU01 297 had GFP pattern AEOPRS. In Chapter 7 (Genetic test and confirmation of the transactivator's functionality), only AEO (root, stomata, lodicule) with relatively strong signal were investigated in the F1 populations.

STATEMENT

I hereby declare that the results presented in this thesis are original work, except where acknowledged, conducted by myself in CAMBIA during the course of my enrolment as a PhD student at the Australian National University. No materials in this thesis have been previously submitted for an academic record at any other university.



Handwritten signature of Xiqin Fu, consisting of a stylized 'Xiq' followed by a long horizontal stroke and a small 'Fu' to the right.

Xiqin Fu

11/2/09

ACKNOWLEDGEMENTS

First and foremost, my deepest appreciation goes to my supervisor Dr. Richard A. Jefferson for giving me the opportunity to pursue my Ph.D. research in CAMBIA and for his financial support. Richard is not only an excellent scientific resource to me but also treats me very friendly. I would have not accomplished my Ph.D. program without his advice, guidance and encouragement throughout my graduate study.

My deepest appreciation also goes to Professor Longping Yuan, the Father of Hybrid Rice, for offering me a position in his institute and allowing me away from the institute for such a long time, and for his invaluable advice, encouragement and support all the way.

I would like to extend my deep gratitude to other members of my supervisory panel, Professor Murray Badger, Dr. Andrzej Kilian and Dr. Dean Price, for their scientific guidance, warm-hearted advice and patience. It would be impossible to make the present achievement without their help. Professor Murray Badger, as chair of the panel, always shows his gracious compassion and willingness to help. A special thanks goes to Dr. Andrzej Kilian for critically reading my thesis.

During my graduate research in CAMBIA, I have been fortunate to work with people from many different countries who are all generous in giving help. I thank those colleagues and friends, including Wei Yang, Liz Deaves, Chris Pratt, Carol Nottenburg, Jorge Mayer, Eric Huttner, Sujin Patarapuwadol, Damian Jaccoud, Shiyang Yang, Lijun Tian, Kaiman Peng, Thach Tran, Oanh Kim Hoang, Selvameena Rajagopal, Leon Smith, Peter Wenzl, Sri Koerniati, Tuan Nguyen, and Satya Nugroho, for their friendship and for their advice on both scientific and interpersonal matters.

My special thanks goes to the Rockefeller Foundation for generous financial support for my graduate study.

Last, but not the least, I would like to thank my parents, my wife Xiaoping Liu, and my daughter Baihua Fu for their love, support and endurance, which has been my invaluable source of strength throughout my graduate study.

ABSTRACT

The enhanced green fluorescent protein (EGFP) reporter was cloned into 26 binary vectors including enhancer traps and positive/negative controls, and successfully used for tracing dynamic gene expression in the transformation process of rice, and for screening enhancer trap pattern lines. GUS::EGFP and GUSPlus::EGFP fusion reporters were also created to make good use of the advantage of GFP as a vital marker for dynamic monitoring of gene expression and the advantage of GUS for its high sensitivity in histochemical staining.

An *Agrobacterium*-mediated transformation system for the japonica rice variety Millin was well established. In a case study with enhancer trap constructs harboring GUSPlus::EGFP fusion reporter, a total of 1,021 transgenic rice lines were obtained in a single transformation experiment starting with 1,000 scutellum-derived calli. Further improvement of the transformation efficiency was achieved by decreasing the temperature to 22°C during co-cultivation of rice calli with agrobacteria. The efficiency can fulfill the prerequisite for the TransGenomics project in which large populations of transgenic lines are needed.

Several sets of enhancer traps were constructed and tested through transformation of rice calli. GFP expression of the enhancer traps, along with corresponding Gal4-deletion constructs, was studied extensively during the callus stage and at plant level. Constructs carrying the CaMV 35S promoter exhibited serious within T-DNA cis-activities, imposing significant background problem for screening genomic enhancers being trapped. The problem was in a great extent solved by replacing the CaMV 35S promoter with Ubi-1 promoter in the selection cassette of enhancer trap constructs. Out of 393 enhancer trap lines obtained with the improved constructs, 129 (32.8%) lines were found GFP-positive with diverse expression patterns, which were valuable genetic resources for functional studies of gene of interest in rice.

Sexual crosses, between enhancer trap lines with EGFP reporter and target gene lines harboring 6xUAS-MP-GUS, were made to verify functionality of the

transactivator in the system. Clear co-expression of GFP and GUS in F₁ progenies were observed, implying that the transactivator performed its functions as theoretically expected and the system established could be used for creating gain-of-function mutagenesis in rice. This is the first genetic testing of the Gal4-UAS system in rice, providing solid evidences for a successful establishment of the Transcriptional Activator-Facilitated Enhancer Trap (TAFET) system in rice. Potential usage of the system for innovative rice breeding was also discussed.

TABLE OF CONTENTS

Statement	ii
Acknowledgements	iii
Abstract	v
Table of contents	vii
List of figures	xiv
List of tables	xvi
Chapter 1 Introduction and literature review	1
1.1 Rice as a model system for cereal genomics.....	1
1.1.1 Rice as an important crop plant.....	1
1.1.2 Challenges for rice breeding in the future.....	2
1.1.3 Rice as a model system for cereal genomics.....	3
1.2 Functional genomics - approaches and limitations	5
1.2.1 Gene expression by defined promoter.....	6
1.2.2 Enhancer trap mutagenesis.....	7
1.2.2.1 The concept of enhancer traps.....	7
1.2.2.2 Development of enhancer trap systems.....	8
1.2.3 The Gal4 system.....	10
1.2.4 Chemical-inducible transcription factors	11
1.2.4.1 Glucocorticoid receptor-based, dexamethasone inducible system (GVG).....	12
1.2.4.2 Tet and glucocorticoid receptors-based, dual control system (TGV)	14
1.3 Use of the green fluorescent protein as a reporter.....	15
1.3.1 Discovery of the green fluorescent protein	15
1.3.2 Molecular structure of GFP.....	16
1.3.3 Structure of the GFP fluorophore.....	17

1.3.4	Basic characteristics of GFP	18
1.3.5	GFP as a reporter	18
1.3.6	Important variants of GFP	19
1.4	<i>Agrobacterium</i> -mediated T-DNA transformation.....	23
1.4.1	Characteristics of <i>Agrobacterium tumefaciens</i>	23
1.4.2	<i>Agrobacterium</i> -mediated T-DNA transfer process	24
1.4.3	T-DNA binary vectors.....	27
1.4.4	Methods of transformation	29
1.4.5	<i>Agrobacterium</i> -mediated transformation in monocots	30
1.4.5.1	Target cells with high division activity and strong regeneration potential	31
1.4.5.2	Increasing vir gene expression	32
1.4.5.3	Bacterium strains and T-DNA vectors	33
1.4.5.4	Other influencing factors.....	34
1.5	Aims of the project.....	34
Chapter 2	Materials and methods	36
2.1	GFP detection	36
2.2	GUS histochemical assay	37
2.3	DNA manipulation and cloning	38
2.3.1	Plasmid DNA isolation (miniprep protocol)	38
2.3.2	Quantitation of DNA.....	40
2.3.3	DNA purification.....	40
2.3.3.1	Purification from the agarose gel	40
2.3.3.2	Purification from other sources	40
2.3.4	Fill-in reaction	41
2.3.5	SAP treatment of vector DNA	41
2.3.6	Rapid DNA ligation	42

2.3.7	Electroporation	43
2.3.7.1	Transformation of <i>E. coli</i> through electroporation	43
2.3.7.2	Transformation of <i>Agrobacterium</i> through electroporation.....	43
2.3.8	Polymerase chain reaction (PCR)	44
2.3.9	Oligo synthesis and DNA sequencing.....	45
2.4	Preparation of electrocompetent cells	45
2.4.1	Preparation of electrocompetent <i>Agrobacterium</i> cells	45
2.4.2	Preparation of electrocompetent <i>E. coli</i> cells.....	46
2.5	<i>Agrobacterium</i> -mediated transformation of rice callus	47
2.5.1	Callus Induction from rice seeds.....	47
2.5.2	Callus subculture	48
2.5.3	Growing agrobacteria for transformation.....	48
2.5.4	Transformation	48
2.5.5	Media used for transformation	50
2.6	Rice genomic DNA isolation	54
2.7	Southern hybridization	56
2.7.1	Enzyme digestion of genomic DNA	57
2.7.2	DNA gel electrophoresis and blotting.....	57
2.7.3	Probe preparation	58
2.7.4	Prehybridization and hybridization	58
2.7.5	Immunological detection.....	59
2.7.6	Chemicals for Southern hybridization:	60
Chapter 3	Construction and expression of GFP and its fusion reporter systems	62
3.1	Introduction	62
3.2	Construction of the enhanced green fluorescent protein (EGFP) reporters.....	64
3.2.1	Comparison between EGFP and mgfp5-ER	64

3.2.2	Construction of the binary vector pFX-B114.1 with EGFP reporter	65
3.2.3	Performance of the binary vector pFX-B114.1	65
3.3	Construction of the GUS::EGFP and GUSPlus::EGFP fusion reporters	68
3.4	Co-expression of GUS/GUSPlus and GFP in KW1	70
3.5	Co-expression of GUS/GUSPlus and GFP in rice calli	71
3.6	Co-expression of GUS/GUSPlus and GFP in rice plants.....	74
3.7	Discussion	75
3.8	Conclusion.....	78
Chapter 4	Optimization of a high throughput <i>Agrobacterium</i> -mediated transformation system	79
4.1	Transformation of rice callus: a case study	79
4.1.1	Rice cultivar, bacterium strain and binary vectors used in the experiment	79
4.1.2	Induction and pre-culture of calli derived from the scutellum of rice seeds	81
4.1.3	Transformation and selection of transgenic calli	81
4.1.4	Regeneration of transgenic plantlets	82
4.1.5	Summary of the single transformation experiment.....	82
4.2	Improvements in transformation efficiency	83
4.3	Discussion	87
4.4	Conclusion.....	89
Chapter 5	Testing elements of enhancer trap constructs	91
5.1	The TAFET system and its elements	91
5.2	Design and construction of the testing vectors	93
5.2.1	Design and construction of enhancer trap vectors	93
5.2.1.1	The intermediate plasmid pVALE15.5	93
5.2.1.2	The intermediate binary vector pFX-B61-1	94

5.2.1.3	The intermediate plasmid pFX-C70-6	94
5.2.1.4	Binary enhancer trap vectors pFX-C90.1-12R and pFX-C97.1-2.....	95
5.2.1.5	Binary enhancer trap vector pFX-C90.2-17.....	95
5.2.2	Design and construction of Gal4-deletion vectors	97
5.3	Reporter gene expression in rice calli transformed with enhancer trap constructs.....	98
5.3.1	Transformation and GFP expression of rice calli in enhancer trap vectors	98
5.3.2	Transformation and GFP expression of rice calli in Gal4-deletion vectors	99
5.3.3	GFP expression of rice calli with more negative control vectors ..	102
5.4	GFP expression in T ₀ rice plants transformed by the enhancer trap and Gal4-deletion constructs	103
5.4.1	Frequency and strength of GFP expression in transgenic lines	103
5.4.2	Frequency of GFP expression in different tissues and organs	104
5.4.3	GFP expression in Gal4-deletion lines.....	106
5.5	Discussion	107
5.6	Conclusion.....	109
Chapter 6	Improvement of the TAFET system	110
6.1	Requirement for the new TAFET constructs' design	110
6.2	Characteristics of the Ubi-1 promoter.....	110
6.3	Construction of enhancer trap vectors and Gal4-deletions	112
6.3.1	Construction of pFX-G74.1 and pFX-G85.2	112
6.3.2	Construction of Gal4-deletion vectors	114
6.4	Construction of the positive control vector	114
6.5	Reporter gene expression in rice calli transformed with enhancer trap and Gal4-deletion constructs.....	115
6.5.1	GFP expression of rice calli in the positive control	116
6.5.2	GFP expression of rice calli in the enhancer traps.....	116

6.5.3	GFP expression of rice calli in the Gal4-deletion constructs	118
6.6	GFP expression in T ₀ rice plants transformed by the enhancer trap and Gal4-deletion constructs	119
6.6.1	Frequency of GFP expression in transgenic lines	119
6.6.2	Patterns of GFP expression in enhancer trap lines	120
6.6.3	Patterns of GFP expression in Gal4-deletion lines	125
6.7	T-DNA copy numbers in T ₀ lines	127
6.8	GFP expression in T ₁ generation	127
6.9	Discussion	128
6.10	Conclusion	131
Chapter 7	Genetic test and confirmation of the transactivator's functionality	133
7.1	Introduction and experimental design	133
7.2	Characteristics of parental lines used for genetic test and combinations of sexual crosses	137
7.2.1	Enhancer trap lines	137
7.2.2	Target gene lines	137
7.2.3	Parental combinations of the genetic crosses	140
7.3	Making crosses using hot water emasculation method	140
7.3.1	Selection and early growth of T-DNA harboring plants	140
7.3.2	Panicle selection and trimming	141
7.3.3	Emasculation	141
7.3.4	Pollination	141
7.3.5	Harvesting hybrid seeds	142
7.4	Growing F ₁ population	142
7.5	Reporter gene expression in T ₂ parental lines	142
7.6	Validation of transactivator functionality through analysis of the F ₁ populations	143
7.6.1	Co-expression of GFP and GUS in F ₁ plants	143

7.6.2	Results of genetic crosses involving the DTA target gene	145
7.6.3	Cytoplasmic effects of the transactivator activities	146
7.7	Discussion	147
7.8	Conclusion.....	150
References		152
Appendix A.	List of abbreviations	175
Appendix B.	List of different expression patterns of GFP and the number of transgenic rice lines representing these patterns.	177

LIST OF FIGURES

Figure 1.1 The overall shape of GFP..	16
Figure 1.2 Mechanism proposed by Cubitt et al (1995) for the intramolecular biosynthesis of the GFP chromophore, with rate constants estimated for the Ser65Thr mutant by Reid & Flynn (1997) and Heim et al (1995).	17
Figure 1.3 Two-way exchange of chemical signals between <i>Agrobacterium tumefaciens</i> and host plants.	25
Figure 2.1 Wavelengths of the filters used for detecting EGFP expression.	36
Figure 3.1 The pEGFP plasmid introduced from CLONTECH	63
Figure 3.2 The binary vector pFX-B114.1 harboring <i>egfp</i> gene driven by CaMV 35S promoter	65
Figure 3.3 Expression of GFP in transgenic rice tissues/organs obtained with the binary vector pFX-B114.1	67
Figure 3.4 The binary vector pFX-D102-9 harboring GUSPlus::EGFP fusion driven by CaMV 35S promoter.	70
Figure 3.5 Co-expression of GFP (left) and GUS (right) in the same rice callus transformed with pFX-D102-9.	72
Figure 3.6 Dynamics of GFP expression with the positive control (pFX-B114.1 harboring EGFP) and GUS/GUSPlus::EGFP fusion reporters.	73
Figure 3.7 T-DNA components of the enhancer trap pFX-E24.2-15R.	77
Figure 4.1 Enhancer trap constructs used for the case study of <i>Agrobacterium</i> -mediated transformation.	80
Figure 4.2 T-DNAs of the four binary vectors used to study the temperature effect on transformation efficiency.	84
Figure 4.3 Dynamic GFP expression in calli of two constructs (pFX-B114.1 and pFX-J55.7R), transformed with 2 different co-cultivation temperatures (26°C and 22°C).	86
Figure 5.1 Components of the TAFET system.	92
Figure 5.2 Arrangement of elements on T-DNAs of enhancer trap vectors.	96
Figure 5.3 T-DNA structures of the Gal4-deletion vectors pFX-C103.1-2 (based on pFX-C90.1-12R) and pFX-C103.2-3 (based on pFX-C90.2-17).	97
Figure 5.4 Dynamics of GFP expression in rice calli transformed with enhancer	

trap vectors..	99
Figure 5.5 Dynamics of GFP expression in rice calli transformed with enhancer traps and their corresponding Gal4-deletion vectors.	100
Figure 5.6 Constructs designed for testing cis-activities of enhancer elements in the CaMV 35S promoter.	101
Figure 5.7 Comparison of GFP expression in Gal4-deletion constructs with different orientation of the CaMV 35S promoter-HPT cassette.	102
Figure 6.1 T-DNA components of the enhancer trap pFX-G74.1.	112
Figure 6.2 T-DNA components of the enhancer trap pFX-G85.2.	113
Figure 6.3 T-DNAs of the Gal4-deletion vectors pFX-G99.3 and pFX-H13.3.	113
Figure 6.4 T-DNA components of the positive control binary vector pFX-J99.2.	115
Figure 6.5 Dynamics of GFP expression in rice calli transformed with enhancer trap constructs and their corresponding Gal4-deletion vectors.....	117
Figure 7.1 Directed gene expression in <i>Drosophila</i>	133
Figure 7.2 Experiment design to verify functions of the transactivator.....	135
Figure 7.3 T-DNA components of the enhancer trap pFX-G74.1.	136
Figure 7.4 T-DNA components of the target gene constructs.	138
Figure 7.5 Seeds treated in water containing hygromycin (50 mg/ml).....	140
Figure 7.6 Co-expression of GFP and GUS in F1 progenies of enhancer trap lines crossed with target lines.....	144

LIST OF TABLES

Table 2.1 Components of the GUS staining solution.....	38
Table 2.2 Components of N6 salts and vitamins.....	53
Table 2.3 Components of MS salts and vitamins.....	54
Table 3.1 Comparison of GFP expression among different reporters, showing percentage of calli with GFP expression in replication I and II.....	73
Table 3.2 Levels of GFP expression in transgenic rice plants with different reporters.....	75
Table 4.1 Results of a single transformation experiment.....	83
Table 4.2 Effects of co-cultivation temperature on transformation efficiency. ...	85
Table 5.1 Number and percentage of transgenic lines showing different strength of GFP expression.	103
Table 5.2 Expression frequencies of GFP in different tissues/organs of enhancer trap lines.....	105
Table 5.3 GFP expression in various tissues and organs of rice enhancer trap lines and Gal4-deletion control lines.....	106
Table 6.1 Number and percentage of transgenic lines showing GFP expression.....	119
Table 6.2 Complexity of reporter gene expression patterns among the lines obtained with enhancer trap construct pFX-G74.1 and pFX-G85.2.	121
Table 6.3 The frequencies of GFP expression in various tissues and organs of rice plants transformed with enhancer trap constructs.....	123
Table 6.4 Frequencies of GFP expression in various tissues/organs of rice plants transformed with Ubiquitin-HPT and 35S-HPT series enhancer trap constructs.....	124
Table 6.5 The frequencies of GFP expression in various tissues/organs of transgenic plants obtained with enhancer trap and Gal4-deletion constructs.....	126
Table 6.6 T-DNA copy numbers in T ₀ sample lines.	127
Table 7.1 Characters of enhancer trap lines used for crossing.....	136
Table 7.2 Code numbers of the target gene lines used for crossing.....	139

Table 7.3 Combinations and code numbers of genetic crosses.....	139
Table 7.4 Co-expression of GFP and GUS in F1 populations.	145
Table 7.5 Results of GFP and GUS co-expression in F1 populations arranged in groups of crossing.	146
Table 7.6 Co-expression of GFP and GUS in F1 populations generated through reciprocal crosses.....	147

Chapter 1 Introduction and literature review

1.1 Rice as a model system for cereal genomics

1.1.1 Rice as an important crop plant

The word “rice” generally indicates a plant and a crop of the species *Oryza sativa* L. The genus *Oryza* L. is classified under the tribe Oryzeae, subfamily Oryzoideae, of the grass family Poaceae (Gramineae) (Lu, 1999). This genus has two cultivated species (*O. sativa* L. and *O. glaberrima* Steud.) and more than 20 wild species distributed throughout the tropics and subtropics. The so-called Asian cultivated rice (*O. sativa*) is actually cultivated worldwide, while *O. glaberrima* is only cultivated in a few countries in West and Central Africa.

Rice is the world’s single most important food crop that is the staple food for more than one-half of the world’s population. It is a nutritious grain crop which contains carbohydrates, proteins, lipids, minerals, etc. Rice is used for food in various forms. Grains are heated in water to become cooked rice. Rice flour is usually kneaded with water, boiled and used for various rice products. The bran is an important source of oil for food and manufacturing. Husks are used for fertilizers and animal feed, and rice straw is used as an important animal feed and for making various wrapping materials and mats.

Rice is planted on about 150 million hectares annually. According to FAO data, 603 million tons of rice was harvested from plantings of 154 million hectares worldwide (FAO, 2003). More than 91% of world rice production comes from Asia, 5% from the Americas, 3% from Africa, and another 1% from Europe and Oceania. Rice is the predominate staple food for fifteen countries in Asia and the Pacific, ten countries in Latin America and the Caribbean, one country in North Africa and seven countries in Sub-Saharan Africa (FAO, 1999). In developing countries, rice accounts for 715 kcal/capita/day which corresponds to 27% of dietary energy supply, 20% of dietary protein and 3% of dietary fat. Countries in

Southeast Asia are heavily reliant upon rice; in Bangladesh, Laos, Viet Nam, Myanmar and Cambodia, rice supplies more than fifty percent of per capita dietary energy and protein supply and 17-27 % of dietary fat. Rice is also an important staple food for several countries in Africa. In Guinea, Guinea-Bissau, Gambia, Liberia, Senegal and Cote d'Ivoire, rice supplies between 22-40 % of dietary energy and 23-39 % of dietary protein.

1.1.2 Challenges for rice breeding in the future

The population of rice consumers is increasing at the rate of 1.8% annually. The present annual rice production of 560 million tons must be increased to 850 million tons by 2025 (FAO, 2002b). There are no additional lands available for rice cultivation. In fact, the area planted to rice is going down in several countries due to pressures of urbanization. Thus, we need the rice varieties with higher yield potential and yield stability for meeting the challenges of increased rice production.

Major advances have been made in increasing rice production worldwide as a result of large-scale adoption of modern high-yielding rice varieties and improved cultural practices. The rapid rice production increase was mainly achieved through the application of principles of classical Mendelian genetics and conventional plant breeding methods, including the hybrid rice technology.

Hybrid rice is the most significant technology since the identification of dwarf plant-types. Hybrids have consistently shown a 15-20 % increase in yield compared to conventional varieties. Scientists in China initiated hybrid rice breeding efforts in the late 1960s and the first commercial hybrid was introduced in 1974. In year 2000, it was estimated that hybrid rice was cultivated on approximately 16 million hectares in China; 300,000 hectares in Vietnam; 150,000 hectares in India and 30,000 hectares in Bangladesh (IRRI, 2000).

Because of consistent efforts by plant breeders, rice production has doubled between 1966 and 1990, but it must increase further by 60% by 2025 in order to

feed the additional rice consumers (Khush, 1997). However, the contribution of traditional breeding methods to increased rice yield has reached a plateau, for the average rice yield per unit area has been stagnated for several years according to the FAO data (FAO, 2003). Yield-stagnation also appears to be a problem even with hybrid rice (Yuan, 1998). The growth in rice yields has decelerated to slightly more than 1% per year or approximately equal to population growth (FAO, 2002a).

Rice demand in year 2030 is projected to be approximately 533 million tons of milled rice (FAO, 2002b). To meet this goal, rice varieties with higher yield potential, durable resistance to diseases and insects, and tolerance to abiotic stresses are needed. Sustainability in rice production can only be achieved by addressing the major yield constraints imposed by drought, floods, salinity, soil degradation and pests and diseases. Addressing these problems requires a coordinated approach using the best available research tools at molecular level. These include use of transgenics, isolation and transfer of novel genes that confer tolerance or resistance to particular abiotic/biotic stresses as well as the use of molecular markers in breeding programs to achieve faster genetic improvement.

1.1.3 Rice as a model system for cereal genomics

Apart from its economic significance, rice has been also an important model plant for genetic and genomic studies. Rice has relatively small genome (430 Mb) compare to other cereal crops, such as maize (2,400 Mb), barley (4,900 Mb) and wheat (16,000 Mb) (Bennetzen, 2002). Rice researchers have developed important tools for genetic analysis, including high-density molecular genetic maps and efficient genetic transformation techniques. Comparative genetic maps within the grass family indicate the existence of conserved gene content and gene order among grasses genomes (Ahn et al., 1993; Devos and Gale, 1997, 2000).

In rice, the first publication of a molecular genetic map using RFLP was in 1988 by S. McCouch et al. (McCouch et al., 1988), and subsequently, higher resolution maps were published by Kurata et al. (Kurata et al., 1994), Causse et

al. (Causse et al., 1994), and Harushima et al. (Harushima et al., 1998). More recently, a sophisticated linkage map has been developed by Yano et al. (Yano, 2000) in which 3267 DNA markers were located and many of them were converted from RFLP to CAPS (Cleaved Amplified Polymorphic Sequence) markers (Glazebrook et al., 1998). CAPS facilitates the identification of polymorphisms within a PCR product using a restriction enzyme as an alternative to the laborious Southern hybridization technique. The quality and number of markers on this map are undoubtedly the most outstanding among all organisms genetically analyzed by DNA markers. The total number of co-dominant markers such as RFLP, SSR or CAPS published to date is about 5000. This means that the average marker density within the rice genome is one marker in every 80 kb.

In April 2002, a rough draft sequence of indica rice variety 93-11 was published by an academic institute, the Beijing Genomics Institute (Yu et al., 2002). In the same month, a draft sequence of japonica variety 'Nipponbare' was published by a private company, Syngenta (Goff et al., 2002). Both draft sequences were obtained by a whole-genome shotgun approach. In November 2002, a few month later, two members of the International Rice Genome Sequencing Project (IRGSP), namely the Rice Genome Research Program (RGP, Japan) and the National Center for Gene Research (NCGR, China) published high-quality phase-3 sequences of chromosome 1 and chromosome 4, respectively (Feng et al., 2002; Sasaki et al., 2002). Both sequences were obtained using complete sequencing of BAC/PAC contigs. On December 18th 2002, the IRGSP announced the completion of a high-quality draft with at least phase-2 sequences of the 12 rice chromosomes (Sasaki and Sederoff, 2003).

The number of cereal genes that are estimated based on partial genomic sequencing and EST clusters ranges from 44,700 for rice (Wu et al., 2002) to 55,000 for maize (Brendel et al., 2002). Two independent analyses that were based on the shotgun genomic sequencing of the rice genome estimate the number of predicted rice genes to be 32,000–50,000 (Goff et al., 2002) and 46,022–55,615 (Yu et al., 2002). More recent estimates, obtained by

extrapolating the annotation of the finished rice chromosomes 1 and 4, predict a slightly higher gene number of 57,000–62,500 (Feng et al., 2002; Sasaki et al., 2002). Together, these estimates suggest that rice has a transcriptome that contains nearly twice the number of genes in *Arabidopsis* (Arabidopsis Genome Initiative, 2000) and humans (Lander et al., 2001), and more than three times the number in *Caenorhabditis elegans* (The C. elegans Sequencing Consortium, 1998) and *Drosophila* (Adams et al., 2000).

Whole-genome analyses from *Arabidopsis* and rice suggest that higher estimate of gene number in cereals can be explained by a combination of factors that includes gene amplification and inaccurate gene predictions (Ware and Stein, 2003). These findings are supported by reports from the International Rice Genome Sequencing Project (IRGSP) on rice chromosomes 1 and 4 (Feng et al., 2002; Sasaki et al., 2002).

A finished rice genome will provide a complete index of potential rice genes but will not tell us which of these genes are important in providing desired traits in cereal crops. Genome sequences are just a beginning: they provide a necessary resource for powerful methods for proteome analysis that require sequence knowledge. In the future, techniques such as gene, protein and metabolic profiling will provide insights into the function and expression patterns of genes and into how these genes ultimately contribute to a crop's ability to react to an environment and reproduce (Ware and Stein, 2003).

1.2 Functional genomics - approaches and limitations

The raw data produced by genome sequencing projects currently provides little insight into the precise workings of an organism at the molecular level. Therefore, the goal of functional genomics is to complement the genomic sequence by assigning useful biological information to every gene. Through this, we can aim to improve our understanding of how the different biological molecules contained within the cell (i.e., DNA, RNA, proteins, and metabolites) interact to

make the organism viable. Clearly, the main challenge is the elucidation of all molecular, cellular, and physiological functions of each gene product.

1.2.1 Gene expression by defined promoter

Ectopic misexpression of a single gene in *Drosophila* is often sufficient to disrupt normal development, generating dominant phenotypes that may provide important clues regarding gene function. The simplest method for controlled gene expression is to use transcriptional regulatory elements from a well-characterized promoter. With the regulatory elements, the gene of interest will express in a defined tissue or stage-specific manner.

A good example of this technique is the use of the *sevenless* promoter to drive gene expression in a subset of cells in the eye imaginal disc of *Drosophila* (Basler et al., 1991). Gain-of-function mutants were obtained by overexpressing the *sevenless* gene with duplicated transcriptional regulatory elements in the promoter. This approach has proven invaluable in analyzing signal transduction and cell fate determination in the *Drosophila* eye.

The *Drosophila* Rh1 promoter contains a pentamer of binding sites for the GLASS transcriptional activator. Hay et al. (Hay et al., 1994) subcloned the regulatory sequence from Rh1 into the GMR (glass multimer reporter) vector and studied ectopic gene expression in the developing *Drosophila* eye. GLASS expression is largely restricted to the developing eye and the larval photoreceptor organs, showing a tight spatial regulation on expression of the P35 gene concerned.

This method is the easiest way for studying ectopic gene expression since only a single or a few transgenic lines are needed. A great limitation of the method is the availability of well-characterized promoters. Moreover, constitutive promoters cannot be used to investigate genes whose constant over- or under-expression has deleterious effects on the plant. Furthermore, when the gene

product to be expressed is toxic to the organism, it becomes impossible to establish stable transgenic lines carrying the gene.

1.2.2 Enhancer trap mutagenesis

Enhancer traps are reporter gene constructs that can respond to cis-acting transcriptional signals when inserted in the genome of certain organism. The concept of enhancer trap has been successfully used in *Drosophila* (O'Kane and Gehring, 1987; Bellen et al., 1989), mouse (Allen et al., 1988) and *Arabidopsis* (Klimyuk et al., 1995; Sundaresan et al., 1995; Campisi et al., 1999), leading to the isolation of genes important in development whose phenotypes would have been missed in conventional mutagenesis screens.

1.2.2.1 The concept of enhancer traps

Transcriptional enhancers are DNA sequences that control expression of nearby genes and direct tissue-specific, positive regulation of those genes (Karp, 2002). The general mode of action of these sequences is that they associate with specific activator proteins that are found in only certain tissues and this association of enhancer and activator facilitates transcription initiation at the target gene's promoter. Enhancers might be located upstream, downstream or even within the transcription unit that they control. Under the influence of one or more enhancers, a gene can be abundantly expressed in certain tissues of the organism where the activator protein is found and expressed weakly or not in other tissues. By identifying and studying enhancers, we can study how genes are regulated in time and space within multicellular organisms and we are able to use enhancers to tag tissues or even individual cells (Karp, 2002).

Conventional mutagenesis approaches are dependent on the presence of a recognizable phenotype when a particular gene is mutated. These approaches have at least two limitations. First, many genes play multiple roles and phenotypic screens may fail to detect such genes that exhibit discrete functions at different stages in development. Second, if a gene is functionally redundant,

disruption of the gene may result in only subtle phenotype or no phenotype at all unless all genes specifying that function are mutated simultaneously.

In recent years, a better method called enhancer trap mutagenesis has been developed for detecting and studying enhancers. This method utilizes a recombinant molecule construct that can be inserted at essentially random sites within the genome and that, when inserted near an enhancer, will express a reporter gene product within the tissue or cells containing activator protein for that enhancer. The enhancer trap transposon contains a minimal promoter element linked to a reporter gene that is inactive, except when influenced by a nearby enhancer. Because most enhancers direct tissue-specific gene expression (Karp, 2002), insertion of the enhancer trap molecule near an enhancer usually results in reporter gene expression within the particular tissue or group of tissues in which that enhancer normally functions. By mobilizing the transposon to diverse, essentially random chromosomal sites, one can detect diverse enhancers at those sites and study the tissue-specific gene control that they direct. Furthermore, enhancer trap constructs have been engineered to permit cloning of genomic DNA adjacent to the site of any particular insertion, enabling the isolation of the enhancer and its target gene from genomic DNA. This has proved to be a powerful system for studying enhancers and promoters that direct highly specific gene expression patterns.

1.2.2.2 Development of enhancer trap systems

The first cis-acting enhancer trap was established in *Drosophila* from transposable P-element (O'Kane and Gehring, 1987) primarily due to high efficiency of the transformation system. In these constructs, the β -galactosidase gene (*lacZ*) has been placed under the control of a weak promoter so that there should be no detectable *lacZ* expression without enhancer activity. In integrated copies of the P-*lacZ*, the *lacZ* gene acts as a reporter for genomic transcriptional enhancer elements that are located sufficiently close to the detector to activate the P-element promoter.

Bellen et al. (Bellen et al., 1989) generated and characterized more than 500 *Drosophila* strains that carry single copies of a P-element enhancer trap. A remarkable diversity of spatially and temporally regulated staining patterns was observed in embryos carrying different insertions. The expression patterns evident in those early enhancer studies showed that 5-10% of the lines with single P-element insertions expressed lacZ in very specific tissues or cells, providing many excellent new tissue and cell markers (Bellen, 1999). These markers are very useful in studying the embryonic development of these tissues and cells.

The enhancer piracy approach developed by Noll et al. (Noll et al., 1994) is a modification of enhancer detection technique. In their enhancer piracy vectors, the *Drosophila* rhomboid (*rho*) gene was substituted for the *lacZ* reporter gene present in enhancer trap vectors. Upon introducing the modified construct into the *Drosophila* genome, *rho* expression came under the influence of neighboring genomic enhancer elements. A variety of dominant *rho* alleles were generated through the enhancer piracy approach, providing important insight into the developmental function of the *rho* gene. Enhancer piracy may prove to be a general strategy for obtaining dominant alleles of a gene of interest in diverse insects, worms, plants and potentially in vertebrates (Noll et al., 1994).

Enhancer trap mutagenesis approach has also been employed in plants initially in *Arabidopsis* (Klimyuk et al., 1995; Sundaresan et al., 1995; Campisi et al., 1999) and recently in rice (Wu et al., 2003). A well studied enhancer trap system has been established by Sundaresan et al. (Sundaresan et al., 1995) based on the two component Ac/Ds transposable elements (Bancroft et al., 1992). The system uses a modified Ds element carrying a β -glucuronidase gene fused to a 35S minimal promoter as enhancer trap. Various starter lines, each containing a single stable DsE insertion, are generated by *Agrobacterium*-mediated transformation. Transposition of the DsE is achieved by crossing the DsE starter lines with lines expressing the Ac transposase. Since parental Ac transposase gene is linked to a gene coding for the indole acetic acid hydrolase (IAAH), which confers sensitivity to naphthalene acetamide (NAM), progenies that are free of

transposase activity can be selected so that each selected DsE insertion is stable. The IAAH gene is also linked to the DsE in the T-DNA donor site to act as a counter-selectable marker for eliminating the progenies retaining the T-DNA donor site in the presence of NAM. At the same time, plants with a transposition event are selected by the kanamycin resistance gene *nptII* contained in the DsE element. This selection procedure ensures that only the plants with transposition to more distal sites can survive. In this way, the genome can be saturated with random DsE insertions.

1.2.3 The Gal4 system

The yeast transcriptional activator Gal4 provides the most powerful method currently available for directing tissue- or cell-specific ectopic expression (Brand and Perrimon, 1993). The Gal4 protein can activate transcription of any target genes bearing Gal4 binding site UAS (upstream activation sequence) and a minimal promoter near the UAS. The Gal4 system is designed to generate lines that express a transcriptional activator, rather than a target gene, in numerous specific expression patterns. Two approaches have been used to generate different patterns of Gal4 expression. The first is to drive Gal4 expression using characterized promoters (eg. a tissue-specific or heat shock promoter). The second is based on enhancer trapping, in which the Gal4 expression pattern is directed by flanking genomic enhancer elements proximal to insertion site of the transactivator construct.

The UAS-target gene lines are generated separately. A cassette of UAS-minimal promoter-target gene is constructed and used for generating target transgene lines. Usually several independent insertion lines should be tested for expression experiments due to the fact that expression of UAS transgenes is subject to position effects (differences in expression potential of various chromatin domains).

Rorth (Rorth, 1996) developed a method of screening for phenotypes caused by misexpression of endogenous genes. In this method, a minimal promoter under

the control of UAS elements is introduced into multiple sites in the *Drosophila* genome to drive endogenous genes flanking downstream the inserted UAS-regulated promoter. When combined with a specific Gal4 line, the flanking endogenous gene will express ectopically in a pattern directed by the Gal4 activator. Rorth et al. (Rorth et al., 1998) generated 2300 independent UAS lines that were screened for dominant phenotypes in combination with various Gal4 pattern lines. Dominant phenotypic abnormalities were detected in 2-7% of the UAS lines depending on which Gal4 line was used. These gain-of-function screens seems to be an important approach to study functions of genes, for mis- or overexpression phenotypes are much less likely to arise by traditional loss-of-function mutagenesis.

The key feature of the Gal4 system is that the Gal4 gene and UAS-target gene are initially separated into two distinct transgenic lines. This ensures that the generated parental lines are viable since in Gal4 pattern lines, expression of Gal4 only has no detectable effects on development of organisms; in UAS lines, the target gene is silent in the absence of Gal4 activator. Only on crossing these two lines will the target gene be activated in the progeny of the cross, and can the phenotypically defective or lethal consequences of misexpression be conveniently studied.

Another feature of the system is the potential to create numerous distinct expression patterns of target gene by crossing target lines with a range of pattern lines.

1.2.4 Chemical-inducible transcription factors

The ability to tightly regulate gene expression in plant cells is an effective tool for the elucidation of gene function and manipulating complex developmental pathways. Chemical-inducible gene expression systems have advantages over other systems as heterologous transcription activators are used and transgene induction can be regulated in a spatial and temporal fashion using tissue or developmental specific promoters to express the activator, and by applying

inducer at a desired time and location. Gene expression can also be restricted to a particular window by constructing transactivators that can be switched on with one chemical and switched off with another chemical. In addition, transgenes can be expressed at physiologically relevant levels by varying the inducer concentration.

The chemical-inducible systems developed for plants so far include: Tet repressor-based, tetracycline de-repressible (Gatz et al., 1992; Love et al., 2000); tTA-based, tetracycline inactivatable (Weinmann et al., 1994); glucocorticoid receptor-based, dexamethasone inducible (Aoyama and Chau, 1997); AlcR-based, ethanol inducible (Caddick et al., 1998; Roslan et al., 2001); Ecdysone receptor (EcR)-based, EcR agonist inducible (Martinez et al., 1999); estrogen receptor-based, β -estradiol inducible (Bruce et al., 2000; Zuo and Chua, 2000); and Tet and glucocorticoid receptors-based, dual control (Bohner et al., 1999).

An ideal inducible gene expression system should have the following desirable properties (Zuo and Chua, 2000): low basal expression levels, high inducibility, specificity to inducer, high dynamic range to inducer concentrations, fast response, switch-off after removal of inducer, and low toxicity. Among the systems reported, the following two systems make use of the function of Gal4.

1.2.4.1 Glucocorticoid receptor-based, dexamethasone inducible system (GVG)

GVG is a chimeric transcription factor consisting of the DNA-binding domain of the yeast transcription factor Gal4, the transactivation domain of the herpes virus VP16, and the receptor domain of the rat glucocorticoid receptor (GR). The GVG system is quite similar to the Gal4 system in its components with the only difference that the hormone-binding domain (HBD) of the glucocorticoid receptor is fused to the Gal4/VP16 transactivator, rendering the system controllable by HBD. The system is based on the rationale that HBD of vertebrate steroid hormone receptors has repressive effects on covalently linked, neighboring domains in the absence of their cognate ligands and de-repression occurs when appropriate ligands bind to the HBD (Picard, 1993).

Aoyama and Chua (Aoyama and Chau, 1997) reported the first GVG system for transcription induction in tobacco. In their system, the GVG chimeric transactivator is driven by the constitutive promoter CaMV 35S. The *gvg* gene is introduced into tobacco together with the Luc receptor gene which is under the control of a minimal promoter with 6 copies of Gal4-binding sequence (UAS). In the absence of the chemical inducer dexamethasone (DEX), activity of the transactivator is repressed due to the effect of GR. When transgenic tobacco plants are treated with DEX in the concentration range from 0.1 to 10 μ M, *gvg* is activated and expression of the reporter gene is stimulated even over 100-fold depending on the concentrations of DEX applied. The transcription induction by DEX is rapid (60 min after DEX addition) and can be maintained for several days after removal of the inducer (Aoyama and Chau, 1997).

Several studies have demonstrated the utility of the GVG-based glucocorticoid-inducible gene expression system in transgenic tobacco and *Arabidopsis* plants (McNellis et al., 1998). Their results indicate that the GVG system should be suitable for studying the effects of expression of most genes of interest, including those that are extremely toxic or deleterious to the plants.

Using the GVG system, Kunkel et al. (Kunkel et al., 1999) developed a method for marker-free transformation. The *ipt* gene of *Agrobacterium* Ti plasmid is known to cause cytokinin production in transformed cells leading to shoot regeneration, though the uncontrolled production of cytokinins causes developmental abnormalities, and the transgenic shoots were unable to produce roots and their flowers were infertile. Kunkel et al. (Kunkel et al., 1999) placed the *ipt* gene under the control of the GVG system (Aoyama and Chau, 1997). *Agrobacteria* carrying the GVG-*ipt* construct were used to inoculate tobacco leaf disc in a medium without auxin and cytokinin. No shoot regeneration was observed in the absence of dex. In the presence of the inducer many shoots regenerated though they were developmentally abnormal. These shoots were then transferred to a medium without the inducer, and after several weeks normal marker-free transgenic plants developed.

However, some shortcomings of GVG system were also reported (Kang et al., 1999). In transgenic *Arabidopsis*, some severe developmental and growth defects were related with the strong *gvg* expression, and the inducer chemical DEX caused non-specific induction of defense-related genes in control plants.

1.2.4.2 Tet and glucocorticoid receptors-based, dual control system (TGV)

The TGV system is in a way a combination of TetR de-repression system (Gatz et al., 1992; Gatz, 1997; Gatz and Lenk, 1998) and DEX-mediated GVG system (Aoyama and Chau, 1997). Gal4, the DNA-binding element of the GVG transactivator, is replaced by TetR and correspondingly, the Gal4-binding site (UAS) is replaced by tet operator sequence for the binding of TetR. The feasibility for TetR's binding to the operators is controllable by tetracycline, while the activity of the transactivator is conditioned by whether the hormone inducer DEX is added or not. Therefore, the TGV system is a dual-control inducible system, where target gene expression can be switched on with DEX and switched off with tetracycline, allowing precise control of induced expression of a transgene.

In this system, the chimeric transcription activator TGV consists of TetR, the regulatory region of the rat GR and the VP16 transactivating sequence. The system has been established and characterized with tobacco plants (Bohner et al., 1999). In their experiments, the TGV transactivator is constitutively driven by the CaMV 35S promoter. The target promoter, consisting of 7 tet operators upstream of a TATA-box, is used to control expression of the GUS reporter gene.

When the plants were treated by DEX at a concentration of 30 μ M for 14 days, GUS activity reached to 2000U, which was the same inducing activity as in the tTA system and about 20% of the GUS activity driven by CaMV 35S promoter (10000U). After 14 days of DEX feeding, GUS activity could maintain at maximal levels for a further 6 weeks under DEX-free conditions, leading to quasi-constitutive levels of gene expression (Bohner et al., 1999).

On the other hand, application of tetracycline caused efficient shut down of the induced promoter. When GUS expression reached its maximal level and DEX was removed, application of tetracycline to the plants led to a decline of GUS mRNA levels after 6 h, and no gus expression being detectable after 24 h (Bohner et al., 1999).

However, there are also some potential problems in this system. First, the induction level is not desirably high and no induction is detected in flowers. Second, tetracycline causes browning of roots and reduction of root growth, and affects photosystem II efficiency by 5-15%. Third, the target promoter construction is still questionable. The one used in the tTA system causes silence of the target gene over time, while the modified target promoter in the TGV system leads a higher background expression.

1.3 Use of the green fluorescent protein as a reporter

1.3.1 Discovery of the green fluorescent protein

The green fluorescent protein (GFP) was discovered by Shimomura et al (1962) as a companion protein to aequorin, the famous bioluminescent protein from jellyfish *Aequorea victoria*. GFP was first isolated in 1974 and has two excitation peaks, one at 395 nm and a second, smaller peak at 475 nm. A single emission peak is found at 508 nm (Morise et al., 1974). The fluorescence in jellyfish is generated by sequential excitation of the two photoproteins (aequorin and GFP). The binding of calcium to aequorin causes the emission of a blue chemiluminescence. The blue light produced by aequorin is absorbed by GFP, which then emits green fluorescence efficiently without the assistance of a cofactor (Chalfie et al., 1994). The crucial breakthroughs in GFP studies came with the cloning of the gene by Prasher et al (Prasher et al., 1992) and the demonstrations by Chalfie et al (Chalfie et al., 1994) and Inouye and Tsuji (Inouye and Tsuji, 1994) that expression of the gene in other organisms creates fluorescence. In just a few years, GFP has become one of the most widely studied and exploited reporter proteins in biochemistry, cell and molecular

biology. The number of published papers involving GFP reporter or marker has since risen to several thousands (5728 based on a PubMed search).

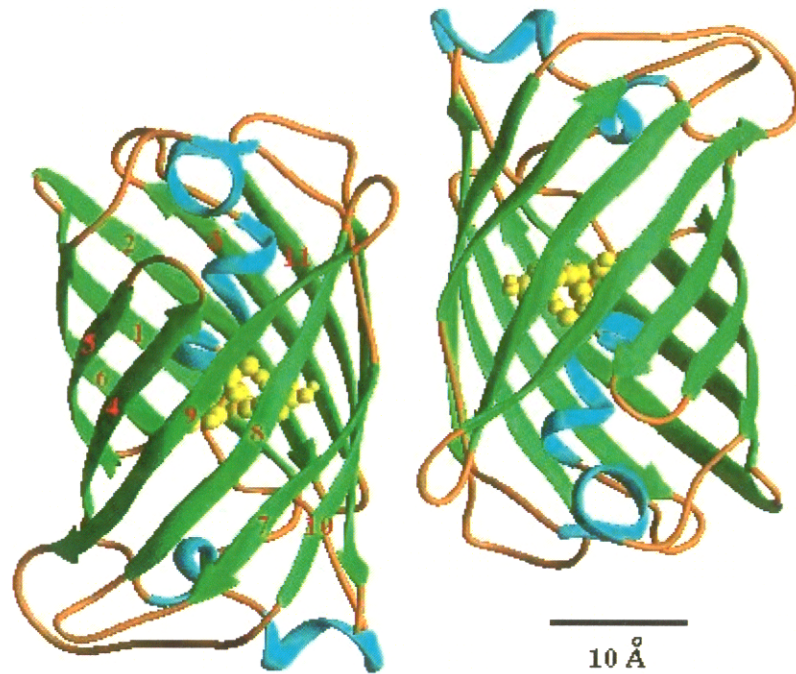


Figure 1.1 The overall shape of GFP. Eleven strands of β -sheet (green) form the walls of a cylinder. Short segments of α -helices (blue) cap the top and bottom of the “ β -can” and also provide a scaffold for the fluorophore which is near geometric center of the can. Two monomers are associated into a dimer in the crystal and in solution at low ionic strengths (Yang et al., 1996a).

1.3.2 Molecular structure of GFP

GFP is a small protein containing 238 amino acids (Prasher et al., 1992). From crystallographic data (Figure 1.1), GFP has been shown to exist as a barrel-like cylinder, composed of 11 closely-packed β -sheets (the slats of the barrel) with an α -helix inside (Yang et al., 1996a). The ends of the cylinder are capped by short helical segments of the polypeptide, forming an internal environment comprised of only the protein and some structured interior water molecules (Boxer, 1996). The structural motif of GFP is unique and has been termed the β -can (Yang et al.,

1996a; Phillips, 1997). It is inside this 'can,' in the center of the internal α -helix, where the GFP chromophore is located. The chromophore is a p-hydroxybenzylideneimidazolinone (Prasher et al., 1992; Cody et al., 1993) formed from residues 65–67.

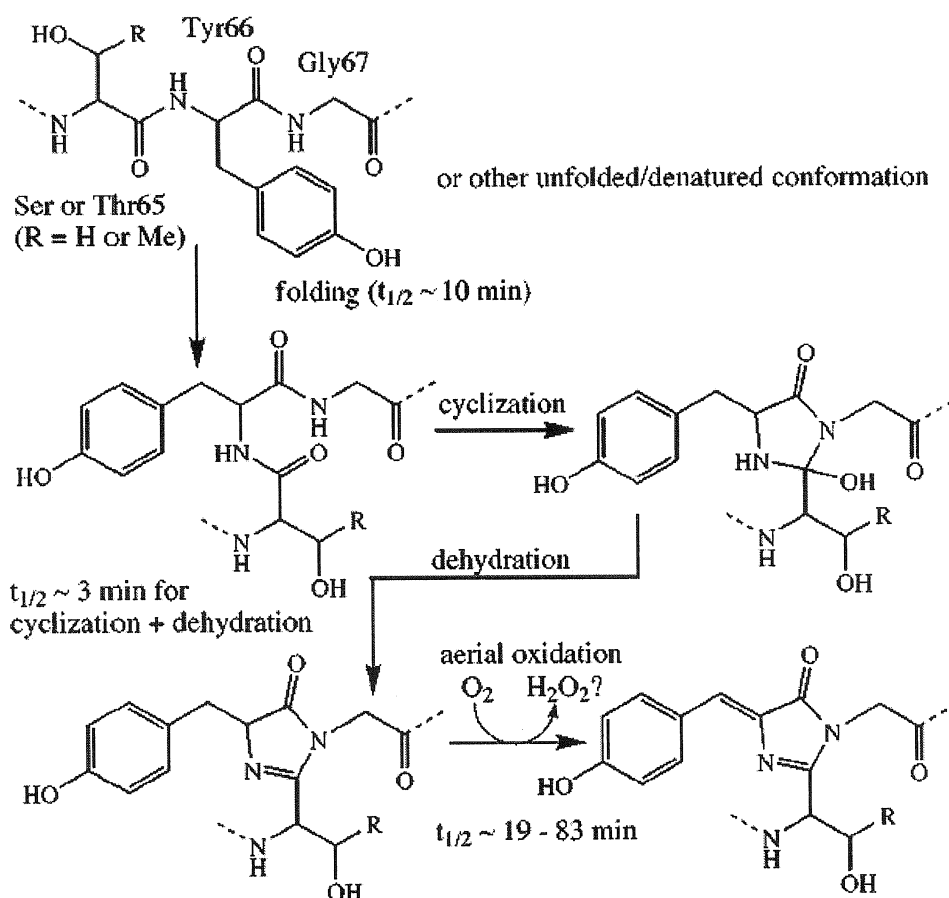


Figure 1.2 Mechanism proposed by Cubitt et al (Cubitt et al., 1995) for the intramolecular biosynthesis of the GFP chromophore, with rate constants estimated for the Ser65Thr mutant by Reid & Flynn (Reid and Flynn, 1997) and Heim et al (Heim et al., 1995).

1.3.3 Structure of the GFP fluorophore

Figure 1.2 shows the currently accepted mechanism (Heim et al., 1994; Cubitt et al., 1995; Reid and Flynn, 1997) for chromophore formation of GFP. The fluorophore is derived from three consecutive amino acids, which are Ser65,

Tyr66 and Gly76 in the native protein. Following synthesis of GFP, these three amino acids undergo autocatalytic cyclization and oxidation to form the functional fluorophore. The carbonyl group (C=O) of Ser65 forms a covalent bond to the amino group (NH₂) of Gly67 and finally the α - β bond of Tyr66 is oxidized to a double bond to put its aromatic group into conjugation with the imidazolinone. Only at this stage does the chromophore acquire visible absorbance and fluorescence.

1.3.4 Basic characteristics of GFP

GFP is an extremely stable protein (Prasher et al., 1992). The photostability of GFP is thought to be due to the tight packing of the fluorochrome within the protein core, which shields it from solvent and from reactive oxygen species responsible for photobleaching. The fluorescent properties of the protein are unaffected by prolonged treatment with 6M guanidine HCl, 8M urea or 1% SDS, and two day treatment with various proteases such as trypsin, chymotrypsin, papain, subtilisin, thermolysin and pancreatin at concentrations up to 1 mg/ml fail to alter the intensity of GFP fluorescence (Bokman and Ward, 1981). GFP is stable in neutral buffers up to 65°C, and displays a broad range of pH stability from 5.5 to 12. The protein is intensely fluorescent, with a quantum efficiency of approximately 80% and molar extinction coefficient of $2.2 \times 10^4 \text{ cm}^{-1} \text{ M}^{-1}$ (Morise et al., 1974).

1.3.5 GFP as a reporter

The use of GFP as a reporter molecule has several advantages over current reporter systems such as β -galactosidase, luciferase, alkaline phosphatase, chloramphenicol acetyltransferase, and GUS. These reporters all require exogenous substrates, co-factors, or antibodies for detection, while GFP only requires oxygen and illumination by the proper wavelength of light (Prasher, 1995; Sheen et al., 1995), thereby facilitating non-invasive, in vivo detection. GFP's relatively small size (27kDa), monomeric structure, and unusual stability allow fusion proteins to be made without disturbing GFP's native structure or

function (Boxer, 1996; Mitra et al., 1996; Plautz et al., 1996). In fact, Roger Tsien's group has produced constructs where a protein of interest is genetically inserted into one of the exterior β -sheets, such that the GFP maintains proper folding and is fully fluorescent (Tsien, 2000).

GFP retains fluorescence when fused to another protein at either the N- or C-terminus (Leffell et al., 1997). This property makes it an ideal fluorescent tag to monitor subcellular localization of proteins, organelles, and biochemical processes. GFP was fused to organelle targeting tags. This resulted in GFP localized to subcellular structures including the ER (Haseloff and Amos, 1995), nucleus (Grebenok et al., 1997), mitochondria (Kohler, 1998), cytoskeleton (Katz et al., 1998), Golgi apparatus (Boevink et al., 1998), chloroplasts (Kohler et al., 1997), and peroxisomes (Jedd and Chua, 2002).

GFP has been successfully transfected and expressed in a number of heterologous plant systems including rice, corn, wheat, oat, sugarcane, tobacco, soybean, and *Arabidopsis* (Hu and Cheng, 1995; Casper and Holt, 1996; Zhang et al., 1996; Elliott et al., 1999; Jordan, 2000; Kaeppler et al., 2000; Upadhyaya et al., 2000). Microprojectile bombardment of GFP fusion constructs into onion epidermal peels as well as roots from intact *Arabidopsis* seedlings has proved effective in transient transfection experiments (Sheen et al., 1995; Scott et al., 1999), while *Agrobacterium*-mediated transformation of *Arabidopsis* is now routinely used to establish stable transgenic plants expressing GFP (Pang et al., 1996; Cutler et al., 2000; Hasezawa et al., 2000). Several uses of GFP include measurement of gene expression in vivo, monitoring dynamic cellular processes in real time, easy visual identification of transformants, and monitoring protein subcellular location and protein-protein interactions (Prasher, 1995).

1.3.6 Important variants of GFP

Cubitt et al. (Cubitt et al., 1995) created the first important mutation of wild-type GFP, which shifted the primary excitation peak from 395 nm to 488 nm, combining the area under the two peaks into a single larger peak. This change

was the result of a single mutation in the chromophore, mutating serine 65 to a threonine (Cubitt et al., 1995). The mutation had several other beneficial effects, including six-fold brighter fluorescence, faster oxidation, and little or no photobleaching of the chromophore (Cubitt et al., 1995). The S65T derivatives of GFP are ideal for confocal microscopy because they are very efficiently excited by the 488 nm argon laser illumination.

The synthetic S65T gene with the cryptic intron removed and humanized codon-usage was called sGFP (S65T) (Haas et al., 1996). Codon engineering conferred a significant increase in expression efficiency and the combination of the S65T mutation and codon optimization resulted in a DNA segment encoding a highly visible mammalian marker protein. Haas et al. (Haas et al., 1996) reported that there was a net improvement in fluorescence per cell of between 40–120 folds, depending on detection conditions. Chiu et al. (Chiu et al., 1996) demonstrated that the Ser65 to Thr (S65T) mutation provided a fluorescence gain of up to 100-fold in plant cells after human codon optimization was performed. The sGFP-S65T mutation variant, which has a single blue excitation peak (489 nm optimum) and a redshifted excitation peak (511 nm optimum), has been used very often in monocots (Elliott et al., 1999; Jordan, 2000; Kaeppler et al., 2000; Upadhyaya et al., 2000).

Another GFP variant popularly used for expression in plants (frequently dicots) is the mutant mGFP5 created by Jim Haseloff (Cambridge, U.K.). Haseloff et al. (Haseloff et al., 1997) optimized wild-type GFP for expression in plants by demonstrating that an 84-nt cryptic intron was removed from GFP cDNAs between nucleotides 400 and 483 in *Arabidopsis* (between residues 380-463 of the GFP coding sequence). The resultant protein was non-fluorescent (Haseloff and Amos, 1995). A modified version of the gene was created that mutated the splice sites and altered codon usage to reduce the AU content of the mRNA. Transgenic plants stably expressing the modified GFP (mGFP series) were created, demonstrating proper expression of the protein in plants (Siemering et al., 1996; Haseloff et al., 1997). mGFP5, however, is similar in spectral properties to wild-type GFP, save a larger excitation shoulder at 475 nm

(excitation = 395/475 nm; emission = 509 nm). Functionally, this means that mGFP5 can be excited equally well by UV and blue spectral components (DAPI and FITC filter settings), while the native GFP from *Aequorea victoria* is preferentially excited by UV radiation (395 nm). Excitation of mGFP5 in living plant tissues by blue light is preferable to UV due to the greater incidence of photodamage induced by the latter.

The enhanced green fluorescent protein (EGFP) is a popular humanized (Yang et al., 1996b) and bright commercial variant from Clontech (Clontech Laboratories, Palo Alto, CA). The characteristics of EGFP are as follows:

- The coding sequence of the EGFP gene contains more than 190 silent base changes which correspond to human codon-usage preferences (Haas et al., 1996)
- A mutation of Ser65 to Thr to promote chromophore ionization and a mutation of Phe64 to Leu to improve folding at 37°C, contributing to a greatly increased (100-fold) fluorescence intensity compared to the wild-type (Cormack et al., 1996)
- A mutation of His231 to Leu, which was probably inadvertent and neutral
- Upstream sequences flanking EGFP have been converted to a Kozak consensus translation initiation site (CGCCACCATGG) (Kozak, 1987) to further increase the translation efficiency in eukaryotic cells, including insertion of a new codon GTG (Val) right after the start codon ATG. The inserted amino acid Val is often numbered 1a to maintain correspondence with the wild-type numbering.
- Excitation maximum = 488 nm; emission maximum = 507 nm

Many other groups optimized and reshaped GFP, creating spectral variants altered in both excitation and emission maxima. The most distinguishable mutant form of GFP features a tyrosine to histidine change at residue 66, resulting in blue fluorescence rather than green (Palm et al., 1997). The existence of two spectrally distinguishable fluorescent proteins made dual-labeling experiments a

possibility and spurred the development of additional variants (Davis and Vierstra, 1998).

In a comprehensive review paper, Tsien (Tsien, 1998) summarized the factors affecting the detectability of GFP as the following four categories, which are practically helpful in utilization of GFP reporter.

Total amount of GFP

- Number of copies of gene, duration of expression
- Strength of transcriptional promoters and enhancers
- Efficiency of translation including Kozak sequence and codon usage
- Absence of mRNA splicing, protein degradation and export

Efficiency of posttranslational fluorophore formation

- Solubility vs. formation of inclusion bodies
- Availability of chaperones
- Hindrance to folding because of unfortunate fusions to host proteins
- Time, temperature, availability of O₂, and intrinsic rate of cyclization/oxidation

Molecular properties of mature GFP

- Wavelengths of excitation and emission
- Extinction coefficient and fluorescence quantum yield
- Susceptibility to photoisomerization/bleaching
- Dimerization

Competition with noise and background signals

- Autofluorescence of cells or culture media at preferred wavelengths
- Location of GFP, diffuse vs. confined to small subregions of cells or tissues
- Quality of excitation and emission filters and dichroic mirrors
- Sensitivity, noise, and dark current of photodetector

1.4 *Agrobacterium*-mediated T-DNA transformation

Agrobacterium tumefaciens is a Gram-negative soil pathogenic bacterium that causes crown galls or tumours on a wide range of dicotyledonous plant species (Zhu et al., 2000). It has played a major role in the development of plant genetic engineering and the basic researches in molecular biology. Initially, it was believed that only dicots, gymnosperms and a few monocot species could be transformed by this bacterium, but recent achievements totally changed this view by showing that many “recalcitrant” species not included in its natural host-range such as monocots and fungi can now be transformed (Chan et al., 1993; Hiei et al., 1994; Bundock et al., 1995; Dong et al., 1996; Ishida et al., 1996; Rashid et al., 1996; Cheng et al., 1997; Tingay et al., 1997; Toki, 1997). In addition, the transformed cells usually carry single or low copy number T-DNA integrated in their genome with less rearrangement, and very large DNA segments can be transformed into the plants (Hamilton et al., 1996; Liu et al., 1999). Twenty-five years ago, the concept of using *Agrobacterium tumefaciens* as a vector to create transgenic plants was viewed as a prospect and a “wish.” Today, many agronomically and horticulturally important species are routinely transformed using this bacterium. Plant transformation mediated by *Agrobacterium tumefaciens* has become the most used method for the introduction of foreign genes into plants (Gelvin, 2003).

1.4.1 Characteristics of *Agrobacterium tumefaciens*

Virulent strains of *Agrobacterium tumefaciens*, when interacting with susceptible plant cells, induce diseases known as crown gall. These strains contain a large megaplasmid which is more than 200 kbp in size (Goodner et al., 1999; Goodner et al., 2001; Wood et al., 2001) and plays a key role in tumor induction and for this reason it was named (tumor-inducing) Ti plasmid. The Ti plasmid is a large, circular double stranded DNA molecule, existing as an autonomous replicating unit within the bacteria.

The transferred DNA (T-DNA) is referred to as the T-region when located on the Ti plasmid. T-regions on native Ti plasmids are approximately 10 to 30 kbp in size (Zambryski et al., 1980; Suzuki et al., 2000).

T-regions are defined by T-DNA border sequences. These borders (right border and left border) are 25 bp in length and highly homologous in sequence (direct repeats). They flank the T-region in a directly repeated orientation (Wang et al., 1984). The right border is generally considered to be more important than the left border (Wang et al., 1984; Jen and Chilton, 1986; Sen et al., 1989).

The Ti plasmid also contains other functional parts for virulence (*vir*), conjugation (*con*) and the origin of its own replication (*ori*). Proteins responsible for T-DNA processing and transfer are encoded by the *vir* region of the Ti plasmid. Approximately twenty genes in this region are essential for wild-type levels of pathogenesis on most host plants and are expressed in six operons, *virA*, *virB*, *virC*, *virD*, *virE*, and *virG* (Sheng and Citovsky, 1996; Zhu et al., 2000).

1.4.2 *Agrobacterium*-mediated T-DNA transfer process

The mechanism of T-DNA processing and transfer during *Agrobacterium* infection has been subjected to a number of excellent reviews (Zambryski, 1992; Zupan and Zambryski, 1995; Zhu et al., 2000; Gelvin, 2003). The process of gene transfer from *Agrobacterium tumefaciens* to plant cells implies several essential steps: (1) bacterial colonization (2) induction of bacterial virulence system, (3) generation of T-DNA transfer complex (4) T-DNA transfer and (5) integration of T-DNA into plant genome. A brief description of the process is given below (Figure 1.3).

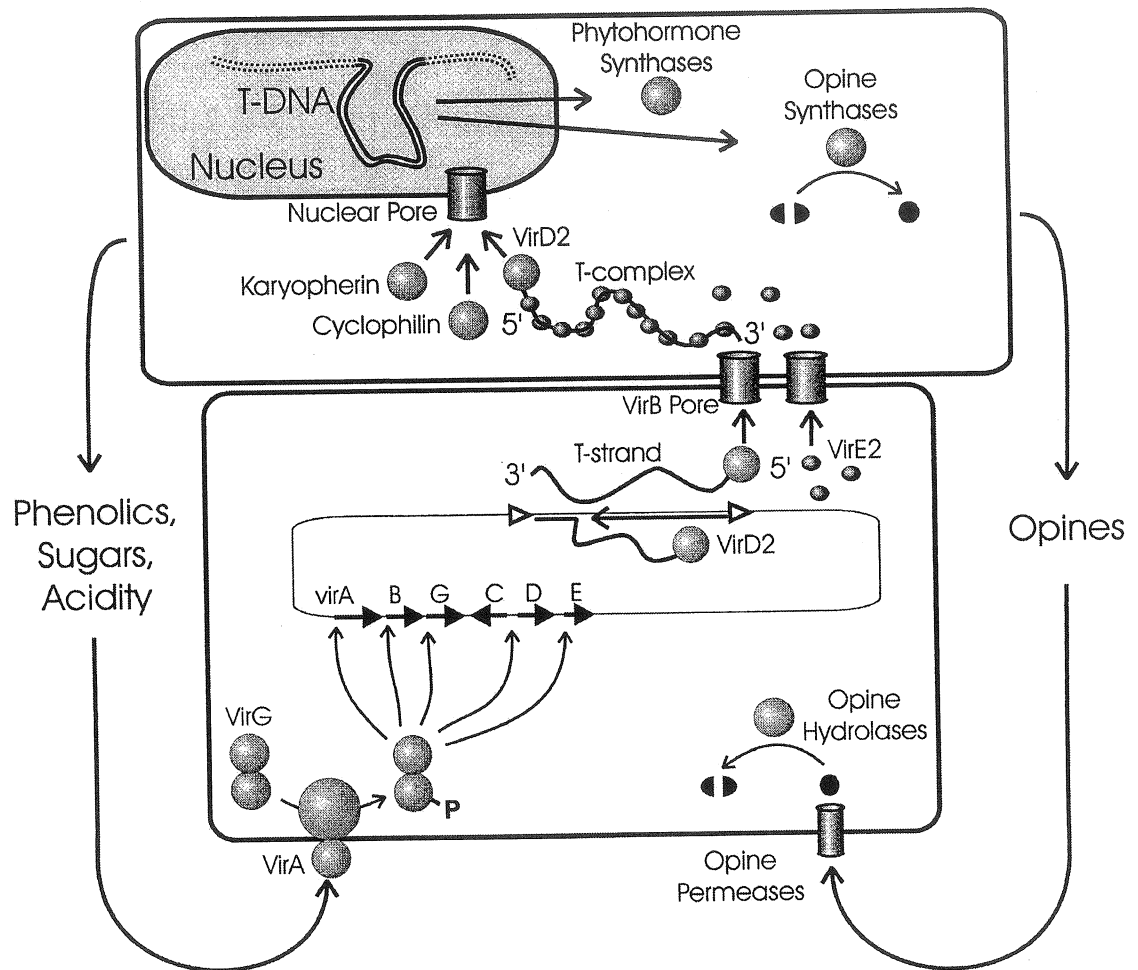


Figure 1.3 Two-way exchange of chemical signals between *Agrobacterium tumefaciens* and host plants. Wound-released chemical stimuli are perceived by the VirA to VirG proteins, which leads to transcription of vir promoters. T-DNA is processed by the VirD2 protein, and single-stranded linear T strands are formed by strand displacement. T strands and VirE2 are translocated from the bacteria via a pore encoded by the virB operon and form a T complex within the plant cytoplasm. T complexes are transported into the nucleoplasm via the host protein karyopherin alpha, and the T-DNA is integrated into genomic DNA. Transferred genes encode phytohormone synthesases that lead to plant cell proliferation and opine synthesases that provide nutrients to the colonizing bacteria. Opines are released from the plant cell, enter the bacteria via dedicated opine permeases, and are catabolized via opine-specific catabolic proteins. Opine permeases and catabolic enzymes are encoded by the Ti plasmid (Zhu et al., 2000).

- Bacterial colonization is an essential and the earliest step in tumor induction and it takes place when *Agrobacterium tumefaciens* is attached

to the plant cell surface. This attachment is achieved by a group of genes located within the bacterial chromosome. The precise mechanism for attachment is still unknown.

- The infection begins at the wounded sites. The injured plant cells release signaling molecules. *Agrobacterium* can recognize plant compounds such as acetosyringone, sinapinic acid, coniferyl alcohol, caffeic acid and methylsyringic acid which induce the bacteria's virulence (Bolton et al., 1986).
- Signaling molecules are recognized by the dimeric transmembrane receptor complex VirA-ChvE, where ChvE is a sugar binding protein; VirA is activated by a way of autophosphorylation.
- Phosphorylation of cytoplasmic VirG by autophosphorylated VirA.
- Activated VirG binds to the regulatory regions of other *vir* operons, acting as a transcriptional activator; and inducing expression of *vir* genes. There are at least six *vir* operons (*virA*, *virB*, *virC*, *virD*, *virE* and *virG*) which mediate the actual transfer (Zambryski, 1992). The *virA* and *virG* are required for overall *vir* expression, indicating they are *vir* master control genes.
- The VirD1 and VirD2 are responsible for the T-strand (a single-stranded copy of the T-DNA) generation by recognizing the 25bp T-DNA borders and specifically cutting the bottom strand at the two borders (endonucleolytic cleavages between the third and fourth base of the 25bp border repeats). After cutting, VirD2 remains covalently attached to the 5'-end of the T-strand, forming a complex with polar character, i.e. the 5'-end is the leading end in subsequent steps of the transfer. Once the T-strand is generated, it remains as single strand linear DNA and must be protected from degradation by bacterial endo- and exonucleases. Protection from endo- and exonucleases is thought to be mediated by the VirE2 protein by binding the T-strand, preventing degradation both in the plant cytoplasm and perhaps in the nucleus (Rossi et al., 1993).
- The T-strand-VirD2 complex and another single strand binding protein VirE2 are believed to be transferred separately into plant cells through a

pilus-like structure made of a complex set of VirB proteins (Zupan and Zambryski, 1995; Baron and Zambryski, 1996; Fullner et al., 1996; Sundberg et al., 1996). There are 11 VirB proteins (Christie, 1997), each except VirB1 is essential for the transfer and tumorigenesis.

- Once into the plant cell, many VirE2 molecules bind cooperatively to the T-strand, forming T-complex (Sundberg et al., 1996; Gelvin, 1998), which is then targeted into the nucleus by the nuclear target signals (NLSs) of its associated VirD2 and VirE2, where the T-DNA randomly integrates into the plant genome as single or multiple copies.
- The integration is shown to preferentially occur in the transcription active and/or repetitive regions of the genome through a process of illegitimate recombination (Mayerhofer et al., 1991; Zambryski, 1992; De Buck et al., 1999), requiring short homology between the T-DNA and the target site of the genome, producing various forms of rearrangement such as small duplications, deletions and fillings. The mechanisms of integration into the host genome are still relatively unknown, However, it is known that the T-complex polarity is maintained upon integration. It is believed that the VirD2-bound 5'-end of the T-complex may help unwind the host DNA and form a gap. Next, the 5'-end of the T-complex is bound to the gap in host DNA and 3'-end is incorporated down-stream. The plant repair and recombination mechanisms may then replace the DNA using the T-strand as a template and ligate the ends together (Zupan and Zambryski, 1995).

1.4.3 T-DNA binary vectors

The native Ti plasmids are very large and T-DNA regions do not generally contain unique restriction endonuclease sites not found elsewhere on the Ti plasmid. Therefore, one cannot simply clone a gene of interest into the T-region. Scientists therefore developed a number of strategies to introduce foreign genes into the T-DNA.

The early method employed to produce a delivery vector from Ti plasmids used cointegration to make the necessary changes to the Ti-plasmid. An intermediate shuttle vector containing T-DNA (excluding the repeat borders), a multiple cloning site, an NPTII gene (for Kanamycin resistance selection in plants) and a selectable marker for *Agrobacterium* was prepared. The gene of interest was ligated into the multiple cloning site, transformed into *E. coli*. and Amp^R colonies were selected and mated with *Agrobacterium*, allowing the shuttle vectors to be transferred into *Agrobacterium*. The T-DNA of the shuttle vector recombined with the T-DNA of the Ti plasmid between the border repeats, incorporating the gene of interest, the NPTII gene and the selectable marker for *Agrobacterium*. Bacteria containing only the modified Ti plasmid and not the intermediate shuttle vector were selected for and propagated in culture. This recombination step was critical in that it removed the tumor inducing genes from the T-region and replaced them with the desired genes. Thus, the *Agrobacterium* has only the ability to integrate T-strands into the host genome and lacks oncogenic activity. The recombinant *Agrobacterium* was allowed to infect plant cells and integrate its T-strand into the host's nuclear genome.

As the technology progressed, the long and cumbersome process of cointegration was replaced by the binary system, which is now the standard for *Agrobacterium* vector production. This strategy was based on findings of Hoekema et al. that the T-region and the *vir* genes could be separated into two different replicons. When these replicons were within the same *Agrobacterium* cell, products of the *vir* genes could act in *trans* on the T-region to effect T-DNA processing and transfer to a plant cell. Hoekema et al. (Hoekema et al., 1983) called this a binary-vector system; the replicon harboring the T-region constituted the binary vector, whereas the replicon containing the *vir* genes became known as the *vir* helper. The *vir* helper plasmid generally contained a complete or partial deletion of the T-region, rendering strains containing this plasmid unable to incite tumors.

The binary vector is produced in vitro. The vector contains a bacterial selectable marker, a bacterial origin of replication and an integration region. The integration region is bordered by the T-region left and right border sequences and contains a

multiple cloning site and a resistant gene for selection in plants. The gene of interest is inserted into the multiple cloning site.

T-DNA binary vectors revolutionized the use of *Agrobacterium* to introduce genes into plants. Scientists without specialized training in microbial genetics could now easily manipulate *Agrobacterium* to create transgenic plants. These plasmids are small and easy to manipulate in both *E. coli* and *Agrobacterium* and generally contain multiple unique restriction endonuclease sites within the T-region into which genes of interest could be cloned. Moreover, the use of two separate plasmids in the binary system enhances the production of VirG and VirE, which increases the effectiveness of transformation. Furthermore, the binary system can accommodate foreign inserts up to 150kb, where the cointegration system can only contain up to 20kb of foreign DNA (Birch, 1997). An excellent review and guideline on binary vectors, *Agrobacterium* strains and their use was published recently by Hellens et al. (Hellens et al., 2000). The most comprehensive and updated review is the White Paper (*Agrobacterium*-Mediated Transformation of Plants) compiled by C. Roa-Rodríguez and C. Nottenburg in CAMBIA

(<http://www.cambiaip.org/Whitepapers/Transgenic/AMT/books/whole.html>).

1.4.4 Methods of transformation

According to Hansen and Wright (Hansen and Wright, 1999), successful transformation of plants demands that certain criteria be met. Among the requirements for transformation they listed are:

- Target tissues competent for propagation or regeneration.
- An efficient DNA delivery method.
- Agents to select for transgenic tissues.
- The ability to recover fertile transgenic plants at a reasonable frequency.
- A simple, efficient, reproducible, genotype-independent and cost-effective process.

- A tight timeframe in culture to avoid somaclonal variation and possible sterility.

The *Agrobacterium*-mediated transformation techniques appear to fulfill these criteria. Several methods have been successfully used for *Agrobacterium*-mediated transformation of plants, including co-culture of agrobacteria with various explants, such as protoplasts, leaf discs and the tissue in whole plants, and other methods integrating non-*Agrobacterium* factors such as the systematic virus infection (agroinfection) and the biolistic method for gene transfer (agrolistics). The choice of the methods depends on the plant species to be transformed and mostly relies on in vitro regeneration of the transformed tissues. Typically, *Agrobacterium*-mediated transformation of dicots is performed using sterile leaf pieces, cotyledons, stem segments, callus suspension cultures and germinating seeds. For *Arabidopsis* however, efficient in planta transformation protocols were developed (Bechtold et al., 1993; Clough and Bent, 1998; Richardson et al., 1998). These in planta simplified transformation protocols have allowed the large scale production of transgenic plants necessary for T-DNA tagging strategies.

1.4.5 *Agrobacterium*-mediated transformation in monocots

The first demonstrated monocot transgenic plant mediated by *Agrobacterium* was *Asparagus officinalis* (Bytebier et al., 1987). Before that, nearly all the early evidence showing the transformation of monocot cells by *Agrobacterium* failed to provide sufficient molecular data to confirm the stable T-DNA integration and inheritance (Smith and Hood, 1995). The major breakthrough occurred in rice, for which two independent groups reported unequivocal molecular and genetic evidence for its transformation (Chan et al., 1993; Hiei et al., 1994). Molecular analysis confirmed that the characteristic features of T-DNA integration were basically similar to those in dicots, and the transformation frequency was also equally high (30%) (Hiei et al., 1994). These greatly stimulated further studies on using *Agrobacterium* to transform other cereal crop species including maize

(Ishida et al., 1996), wheat (Cheng et al., 1997), barley (Tingay et al., 1997) and sugarcane (Enriquez-Obregon et al., 1998).

It was 10 years after the first dicot plants were transformed (Barton et al., 1983), that transgenic rice plants were obtained from immature embryos infected with *Agrobacterium* (Chan et al., 1993). One reason for the delay in monocot transformation by *Agrobacterium* was the weak, or lack of wounding response, from injured monocot tissues. This response is essential for successful infection of dicot plants as factors, such as acetosyringone (AS), released by damaged dicot cells induce expression of virulence genes located on the Ti-plasmid. The wounding response of monocots is very different. Monocot cells usually tend to be lignified or sclerified without active cell division. This is one of the major reasons accounting for the insusceptibility of most monocots to *Agrobacterium* infection (Hiei et al., 1997), and the widely used methods based on wounding response such as leaf-disc transformation are obviously unsuitable for monocots.

Despite the weak wounding induction response from monocot cells, many attempts were made to adapt the *Agrobacterium* system for monocot transformation. There are many factors affecting T-DNA transfer into monocot cells, and recent studies have identified several key factors for successful transformation of monocots by *Agrobacterium*, which are listed below.

1.4.5.1 Target cells with high division activity and strong regeneration potential

Hiei (Hiei et al., 1994) reported, for the first time in monocot plants, an efficient *Agrobacterium*-mediated transformation system for rice. They examined various tissues, namely, shoot apices and segments of roots from young seedlings, scutella, immature embryos, calli induced from young roots and scutella, and cells in suspension cultures induced from scutella, as donor for *Agrobacterium tumefaciens*. The greatest production of transgenic cells was obtained from scutellum callus. In addition, scutellum callus have high ability of propagation and regeneration, which showed scutellum callus was suitable for a donor material of *Agrobacterium*-mediated transformation.