



UNIVERSITÀ DEGLI STUDI DI UDINE

Dottorato di Ricerca in Scienze e Biotecnologie Agrarie
Ciclo XXIV
Coordinatore: prof. Mauro Spanghero

TESI DI DOTTORATO DI RICERCA

**Miglioramento dei livelli di espressione di enzimi
lisosomiali umani prodotti in endosperma di
Oryza sativa L.**

DOTTORANDA
Dott.ssa Serena Versolatto

SUPERVISORE
Prof. Stefano Marchetti

ANNO ACCADEMICO 2010/2011

SUMMARY

The present study describes the successful attempt to increase the expression levels of two recombinant proteins, the acid- α -glucosidase (GAA) and the acid- β -glucosidase (GCase), in the rice endosperm.

GAA and GCase are human lysosomal enzymes, whose deficiency or insufficient catalytic activity causes the rare autosomal recessive storage disorders known as Gaucher and Pompe disease, respectively.

Pompe disease is characterized by glycogen accumulation in lysosomes, leading to lysosomal swelling, cellular damage and organ malfunction; in early-onset patients (infantile and juvenile forms), glycogen accumulation may cause death.

The enzymatic deficiency of GCase causes a process of sphingoglycolipid engulfment especially into cells of the monocyte-macrophage system endowed with an intense phagocytic activity. In these cells, the stored material mainly grows out of the membranes of phagocytised peripheral hemopoietic cells.

A clinical approach for the treatment of lysosomal storage diseases is based on enzyme replacement therapy (ERT) which consists of regular infusions of the missing enzyme by intravenous administration of the recombinantly-produced enzyme. Currently, recombinant forms of human GAA e GCase have been produced and purified from CHO cells (Chinese hamster ovary cells) and approved for the enzyme replacement therapy. However, the CHO technology is rather inefficient and shows a series of disadvantages such as low production and dramatically high production costs which reflect in a limited availability and in a more expensive drug. To overcome these problems and to guarantee the drug access to all patients, transgenic plants have been proposed as an alternative safe and cost-effective production system. In recent years plant systems have been found ideal bioreactors for producing recombinant proteins because of their high yields, low production cost, large storage ability and low risk of contamination from human- and animal-derived pathogens. To date, many recombinant proteins have been successfully produced with the rice seed system used as bioreactor.

In this respect, endosperm-specific expression was chosen because endosperm is an ideal platform for the production of recombinant proteins. Compared with other expression systems, the use of endosperm has many advantages: 1. it

allows an accumulation of the recombinant proteins for long time without compromising enzymatic activity in a small volume and in a stable biochemical environment; 2. in the seed there is a relatively low content of lipids and phenols which could hinder protein extraction and purification processing; 3. proteins expressed in endosperm have no interference with plant metabolism and growth. However, how to further increase the level of recombinant enzymes is still a major problem for the practical application of plant based production system.

The main scope of the present research was to improve the production of human α - and β -glucosidase exploiting the factors affecting the expression levels. In order to achieve the aforementioned objective, experiments were designed to: a. create synthetic versions of human GCase, applying the criteria of Codon Context (CC) and Codon Usage (CU); b. evaluate different 3'-UTR elements as terminators; c. construct a suitable GCase and GAA expression vector for *Agrobacterium*-mediated rice transformation; d. develop a selection system and progeny analysis based on protein extraction and subsequent immunological assay (DAS-ELISA).

Gene design technology was applied to convert the native GCase coding sequence into synonymous versions optimized for a high expression in rice seed. GCase seed content was determined in a population of 50 individuals harboring the GCase CC and the GCase CU compared with the native version of the sequence. Results showed that the GCase CC version determines a remarkable increase in the seed content of the recombinant protein; it was therefore demonstrated experimentally the improvement of the expression levels using the Codon Context method.

In the attempt to achieve high levels of accumulation of recombinant GAA, the effect of two different 3'-UTR sequences, NOS and GluB4 terminator, combined with the GluB4 promoter was studied. Compared to NOS, the 3'-UTR of GluB4 significantly enhanced the accumulation of recombinant GAA leading to a 2-fold increase of GAA production in rice endosperm.

In order to identify the lines with the higher content of GCase and GAA into the population of primary transformants, a selection system based on protein extraction and subsequent DAS_ELISA was optimized. The average GAA and GCase amount in the crude, seed protein extracts from best primary transformants was about 120 mg/Kg of rice seed; however, major improvements

in productivity can be achieved through the use of refined extraction techniques, further selection of transformants and transgene fixation at an homozygous state. For this reason, in the present work the lines which presented the highest amount of each protein were chosen to produce the next plant generation in order to fix the transgene at an homozygous state.

In conclusion, both recombinant enzymes were successfully produced in rice endosperm demonstrating the possibility to use rice as a bioreactor for recombinant protein production. Moreover, the yield of recombinant proteins in transgenic rice can be increased by optimizing parameters that affect endosperm-specific expression. In particular, both codon optimization and selection of appropriate 3'-UTR sequences can certainly improve accumulation of recombinant enzymes in rice seed. The endosperm-specific expression system developed in rice has all the prerequisites to develop into a technological platform for a more cost-effective, industrial-scale production of GAA and GCase.

1.	INTRODUZIONE	8
1.1	I disordini da accumulo lisosomiale (LSD)	8
1.2	L' α -glucosidasi acida umana (GAA)	9
1.2.1	Generalità sull' α -glucosidasi acida umana	9
1.2.2	Caratteristiche strutturali, funzionali e sintesi della proteina	10
1.2.3	Relazioni filogenetiche	13
1.2.4	La malattia di Pompe	14
1.2.5	La GAA come farmaco nella terapia enzimatica sostitutiva	16
1.3	La β -glucosidasi acida umana (GCasi)	18
1.3.1	Generalità sulla β -glucosidasi acida umana	18
1.3.2	Caratteristiche strutturali, funzionali e sintesi della proteina	19
1.3.3	Organizzazione genomica	21
1.3.4	Relazioni filogenetiche	22
1.3.5	La malattia di Gaucher	23
1.3.6	Principali sedi di accumulo della malattia di Gaucher	25
1.3.7	Mutazioni responsabili della malattia di Gaucher	26
1.3.8	La GCasi come farmaco nella terapia enzimatica sostitutiva	28
1.4	Terapia enzimatica sostitutiva nel trattamento dei pazienti affetti da malattia di Gaucher o da glicogenosi di tipo II: problematiche generali e costi della terapia	29
1.5	Le piante come bioreattori verdi	31
1.5.1	Il sistema di espressione in pianta: le modificazioni post-traduzionali	32
1.5.2	Strategie di "umanizzazione" delle proteine ricombinanti prodotte in pianta	36
1.5.3	Strategie di miglioramento dell'espressione di proteine ricombinanti in pianta	40
1.6	GAA e GCasi in pianta	43
1.7	Il riso come specie modello	44
1.8	Il metodo di trasformazione con <i>Agrobacterium tumefaciens</i>	46
1.8.1	Il plasmide pTi	49
1.9	La normativa sugli OGM	52
1.10	Scopo della tesi	55
2.	MATERIALI E METODI	56
2.1	Scrittura in silico del gene GAA ottimizzato per l'espressione in riso	56
2.2	Scrittura in silico del gene GCasi ottimizzato per l'espressione in riso	57
2.3	Analisi della sequenza	58
2.4	Metodiche comuni di biologia molecolare	59
2.5	Vettori utilizzati	62
2.5.1	pUC18	62
2.5.2	pGEM-T	63
2.5.3	pCAMBIA	64

2.6	Realizzazione dei vettori di espressione per GAA e GCasi	65
2.6.1	Realizzazione dei vettori di espressione per GAA	66
2.6.1.1	pCAMBIA1300_GAA CC-NOS ter_PMI e pCAMBIA1300_GAA CC-NOS ter_PMI_FxAntiFx	67
2.6.1.2	pCAMBIA1300_GAA CC-GluB4 ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx	69
2.7	Realizzazione dei vettori di espressione per GCasi	70
2.8	Trasformazione di <i>Oryza sativa</i> mediante <i>Agrobacterium tumefaciens</i>	71
2.8.1	Preparazione e sviluppo di calli embrionici da scutello di riso	71
2.8.2	Co-coltura dei calli con <i>A. tumefaciens</i>	72
2.8.3	Selezione dei calli basata su PMI	73
2.8.4	Rigenerazione di piantine di riso da calli trasformati	73
2.9	Estrazione di DNA da foglie di riso con i vettori pCAMBIA1300_GAA CC-GluB4 ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx	75
2.10	Verifica della presenza del gene GAA in piante putativamente trasformate	76
2.11	PCR di controllo per saggiare l'amplificabilità del DNA estratto	76
2.12	PCR per l'amplificazione del gene GAA	76
2.13	PCR per l'amplificazione del <i>backbone</i> del vettore	76
2.14	Analisi sulle progenie di piante GCasi e GAA	77
2.14.1	Estrazione di proteine totali da semi GAA e GCasi	77
2.14.2	Estrazione di proteine totali da singoli semi GCasi	78
2.14.3	Metodiche comuni di analisi proteica	79
2.14.4	Saggio immunoenzimatico (DAS-ELISA)	79
2.14.5	Validazione del saggio DAS-ELISA	80
2.15	Precipitazione frazionata delle proteine dall'estratto crudo con solfato di ammonio	81
2.16	Purificazione di GAA	81
2.16.1	Cromatografia di affinità con concanavalina A	81
2.16.2	Cromatografia a interazioni idrofobiche (HIC)	82
2.16.3	Gel-filtrazione (GF)	82
SEZ. A - RISULTATI PER GAA		83
3.1	Riscrittura della sequenza codificante la GAA umana e costruzione del vettore pUC18_GAA CC	83
3.1.1	Analisi della sequenza	94
3.1.2	Assemblaggio dei frammenti in pUC18	94
3.2	Realizzazione dei vettori di espressione per GAA	96
3.2.1	Isolamento di PMI	96
3.2.2	Realizzazione della cassetta per il silenziamento FxAntiFX	98
3.2.3	pCAMBIA1300_GAA CC-NOS ter_PMI e pCAMBIA1300_GAA CC-NOS ter_PMI_FxAntiFx	103
3.2.4	pCAMBIA1300_GAA CC-GluB4 ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx	107

3.3	Trasformazione di riso con <i>Agrobacterium tumefaciens</i>	116
3.4	Analisi molecolari sulle piante trasformate con i vettori pCAMBIA1300_GAA CC-GluB4 ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx	117
3.5	Analisi sulla progenie	119
3.5.1	Prove di estrazione di proteine totali da seme	119
3.5.2	Validazione del saggio DAS-ELISA	120
3.5.3	Valutazione dell'effetto del terminatore sui livelli di espressione della proteina ricombinante	121
3.5.4	Selezione del migliore espressore GAA CC-GluB4 ter_PMI_FxAntiFX	122
3.5.5	Selezione del migliore espressore GAA CC-GluB4 ter_PMI	123
3.6	Purificazione di GAA	124
	 SEZ. B - RISULTATI PER GCasi	 128
3.7	Resintesi genica di GCasi	128
3.7.1	Riscrittura della sequenza codificante la GCasi umana secondo il metodo del <i>codon context</i>	128
3.7.2	Riscrittura della sequenza codificante la GCasi umana secondo il metodo del <i>codon usage</i>	133
3.7.3	Analisi della sequenza	136
3.8	Realizzazione dei vettori di espressione per GCasi	137
3.8.1	pCAMBIA1300_GCasi nativa_PMI e pCAMBIA1300_GCasi nativa_PMI_FxAntiFx	137
3.8.2	pCAMBIA1300_GCasi CC-NOS ter_PMI-FxAntiFx e pCAMBIA1300_Gcasi CU-NOS ter_PMI_FxAntiFx	143
3.9	Analisi sulla progenie	146
3.9.1	Prove di estrazione di proteine totali da semi Gcasi	146
3.9.2	Validazione del saggio DAS-ELISA	147
3.9.3	Valutazione dell'effetto della riscrittura genica sui livelli di espressione della proteina ricombinante	148
3.9.4	Selezione del migliore espressore GCasi CC	152
3.9.5	Identificazione delle piante T ₂ altamente produttive	154
4.	DISCUSSIONE	156
5.	CONCLUSIONI	169
5.1	Prospettive future	170
6.	BIBLIOGRAFIA	171

1. INTRODUZIONE

1.1 I disordini da accumulo lisosomiale (LSD)

I disordini da accumulo lisosomiale (LSD) rappresentano un vasto gruppo di malattie genetiche umane caratterizzate da un difetto genetico a carico di uno o più enzimi lisosomiali specifici coinvolti nel metabolismo cellulare (Richards, 2002).

I lisosomi sono organelli citoplasmatici di forma quasi sferica, circondati da un unico strato membranoso e presenti nelle cellule animali. Sono deputati alla degradazione di polimeri biologici (proteine, acidi nucleici, carboidrati e lipidi) e al successivo riciclo di componenti cellulari sotto forma di sostanze monomeriche che vengono rilasciate nel citosol (Neufeld, 1991).

Gli enzimi degradativi sarebbero dannosi se liberi di agire su tutti i componenti cellulari. Pertanto, essi vengono convenientemente racchiusi all'interno del lisosoma. Una seconda modalità di difesa contro la distruzione incontrollata di macromolecole citosoliche da parte degli enzimi lisosomiali è rappresentata dalla differenza di pH che si instaura tra l'interno dei lisosomi e il citosol a opera di una pompa protonica attivata dall'ATP e presente nella membrana lisosomiale. Il compartimento lisosomiale è, infatti, più acido ($\text{pH} \leq 5$) rispetto al citosol (pH intorno a 7). Gli enzimi degradativi delle proteine, dei lipidi, dei mono- e polisaccaridi, le nucleasi e le fosfatasi sono attivi a pH acido mentre la loro attività diminuisce notevolmente al pH più elevato del citosol (Nelson e Cox, 2002).

L'insufficiente attività degli enzimi lisosomiali richiesti per il catabolismo cellulare è causa delle malattie da accumulo lisosomiale. In relazione alla particolare mutazione, i pazienti colpiti possono presentare un'assenza totale o parziale di attività enzimatica. Conseguentemente si osserva un accumulo progressivo del substrato che non può essere degradato a livelli sufficienti all'interno del lisosoma (Tab. 1.1). La mancata degradazione degli intermedi biochimici porta a un malfunzionamento delle cellule e dei tessuti che si manifestano in una varietà di sintomi clinici sia cronici, sia progressivi (Hopwood e Brooks, 1997).

Malattia	Disordine metabolico	Enzima difettivo	Substrato accumulato
Malattia di Pompe	Idrolisi difettiva di glicogeno	α -glucosidasi acida	Glicogeno
Malattia di Gaucher	Idrolisi difettiva di glucosfingolipidi	β -glucosidasi acida	Glucosilceramide
Malattia di Fabry	Idrolisi difettiva di glicosfingolipidi	α -galattosidasi A	Globotriaosilceramide
Mucopolisaccaridosi I	Catabolismo difettivo di glicosaminoglicani	α -L-iduronidasi	Eparano solfati
Malattia di Niemann-Pick B	Idrolisi difettiva di sfingomieline	Sfingomielinasi	Sfingomieline

Tabella 1.1 Principali malattie da accumulo lisosomiale.

Due importanti disordini da accumulo lisosomiale sono rappresentati dalla malattia di Gaucher e dalla glicogenosi di tipo II (anche nota come malattia di Pompe), determinate da una deficienza nell'attività enzimatica rispettivamente della β -glucosidasi acida e dell' α -glucosidasi acida. Entrambe sono proteine umane che svolgono specifiche funzioni e presentano caratteristiche biochimiche tali da farle assumere anche importanti funzioni terapeutiche. Sono accumulate da una disponibilità limitata e da un processo di sintesi e produzione complesso. Nei prossimi paragrafi seguirà una descrizione particolareggiata per ognuna di esse.

1.2 L' α -glucosidasi acida umana (GAA)

1.2.1 Generalità sull' α -glucosidasi acida umana

L'alfa glucosidasi lisosomiale (alfa glucosidasi acida, maltasi acida, in breve GAA; EC 3.2.1.3) è una proteina lisosomiale solubile appartenente alla famiglia delle γ -amilasi con specificità per il glicogeno e diversi oligoglucosidi naturali e sintetici.

1.2.2 Caratteristiche strutturali, funzionali e sintesi della proteina

L'enzima è una proteina lisosomiale solubile con attività ottimale a valori di pH compresi tra 4.0 e 5.0 (Jeffrey et al., 1970; Palmer, 1971; Rosenfeld, 1975; Kostner e Slee, 1977).

Nei lisosomi, GAA catalizza la reazione di idrolisi dei legami glucosidici α -1,4 e α -1,6 con degradazione del glicogeno a glucosio.

Il gene codificante l' α -glucosidasi umana è localizzato sul braccio lungo (q) del cromosoma 17 tra le posizioni 25.2 e 25.3. E' lungo approssimativamente 28 Kb e contiene 19 introni e 20 esoni (Martiniuk et al., 1991). Il primo introne è situato all'interno della regione non tradotta 5' e il primo esone è, perciò, non codificante. Il codone di inizio ATG è localizzato a 33 bp dall'inizio del secondo esone. Nei vertebrati la porzione non tradotta del primo esone codificante dei geni è generalmente corta e raramente supera i 40 nucleotidi.

L'RNA messaggero del gene GAA comprende una sequenza codificante di 2856 bp fiancheggiata da una regione non tradotta in 5' di 218 bp e una regione non tradotta in 3' di 555 bp. La CDS codifica per una proteina precursore di 952 a.a. con una massa molecolare predetta di 110 KDa. Il sito catalitico è situato in corrispondenza del residuo di acido aspartico 518.

Il precursore di GAA porta un peptide segnale aminotermine per il trasporto co-traduzionale nel lume del reticolo endoplasmico dove viene N-glicosilato a livello di sette siti di glicosilazione. Da qui il precursore sintetizzato è trasportato al trans Golgi. Dopo la traslocazione, la maggior parte della molecola è indirizzata ai lisosomi mentre una piccola quantità viene portata al plasmalemma e secreta (Fig. 1.1). L'invio ai lisosomi può essere mediato dai recettori del mannosio-6-fosfato (MPRs) o seguire un cammino indipendente dal mannosio-6-fosfato (M6P) attraverso un recettore ancora sconosciuto (Tsuji et al., 1988). L'aggiunta dei residui di M6P è il risultato di modificazioni post-traduzionali a opera di N-acetilglucosaminidasi e N-acetilglucosaminil fosfotransferasi localizzate nel compartimento del Golgi. L'enzima modificato contenente M6P è poi riconosciuto da MPRs catione-dipendenti e catione-indipendenti che alla fine trasferiscono l'enzima ai lisosomi (Zhu et al., 2005). Come detto, sebbene molti enzimi siano traslocati ai lisosomi attraverso i recettori MPRs, una piccola parte sfrutta una via indipendente da tali recettori (Aerts et al., 1988). Ciò è stato dimostrato da studi condotti sulla malattia a cellule I causata da una mancanza di fosfotransferasi. Da tali ricerche è emerso che GAA è presente a livelli normali

nel lisosoma, indicando che un enzima non fosforilato potrebbe essere trasportato ai lisosomi indipendentemente dagli MPRs (Tsuji et al., 1988). Questi risultati sono confermati da lavori su fibroblasti MPR-difettivi (Ludwig et al., 1994).

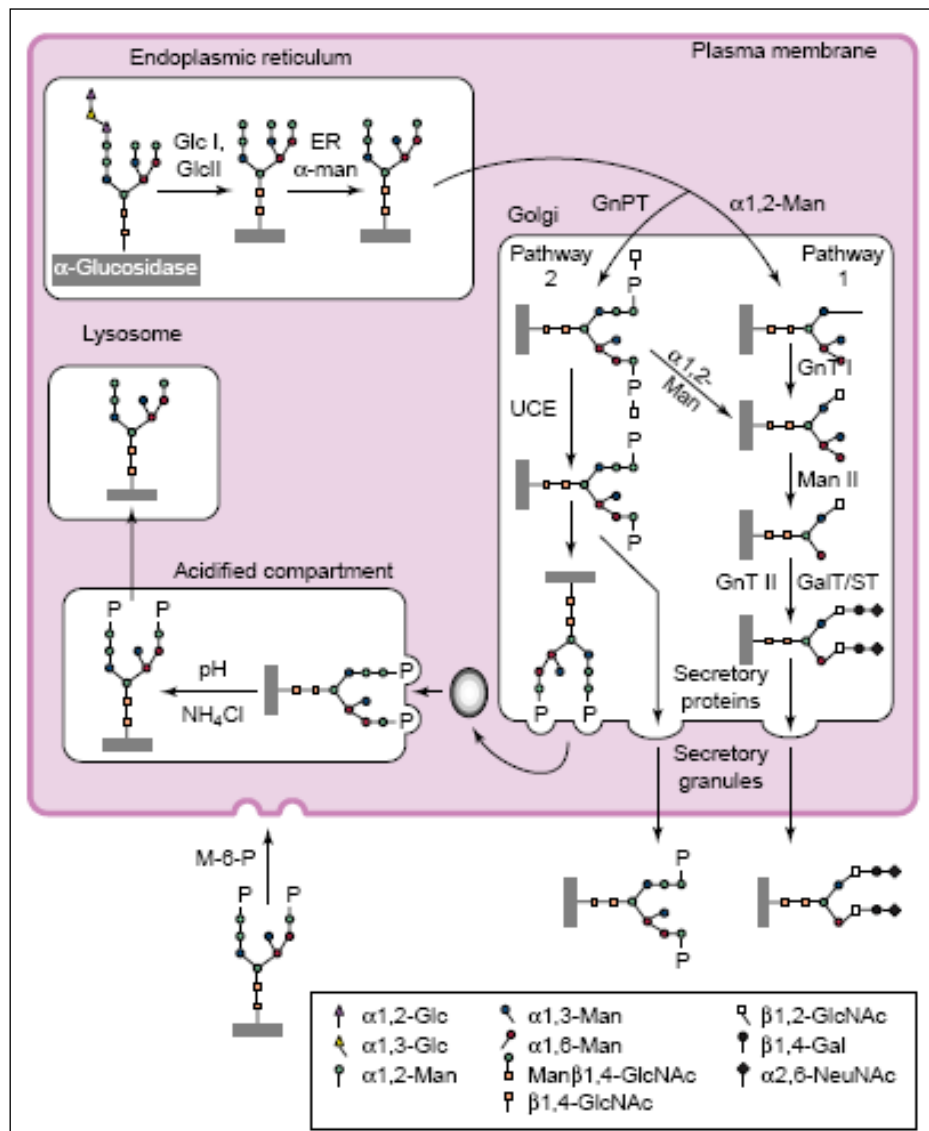


Figura 1.1 Traffico degli enzimi lisosomiali.

Dopo la rimozione della sequenza segnale dal precursore di 110 KDa, un processamento proteolitico che ha luogo nell'endosoma/lisosoma conduce alla formazione di un intermedio di 95 KDa e da questo alle forme lisosomiali pienamente attive (76 KDa e 70 KDa) (Geel et al., 2007) (Fig. 1.2).

In particolare, la forma di 76 KDa si forma a seguito del taglio proteolitico dell'intermedio di 95 KDa tra gli aminoacidi in posizione 816 e 881. Un'ulteriore

proteasi che agisce al terminale amminico della forma di 76 kDa porta alla produzione della molecola di 70 kDa.

Entrambe le forme della GAA sono responsabili dell'idrolisi del glicogeno a glucosio (Raben et al., 2002; Chen et al., 2000; Hirschhorn et al., 2001).

Sebbene l'identità delle proteasi coinvolte nella maturazione dell' α -glucosidasi non sia stabilita, il processamento proteolitico sembra essere richiesto per l'attività ottimale verso il substrato naturale. Infatti, le specie 76-70 kDa mostrano un'affinità per il glicogeno 7-10 volte superiore rispetto al precursore di 110 kDa.

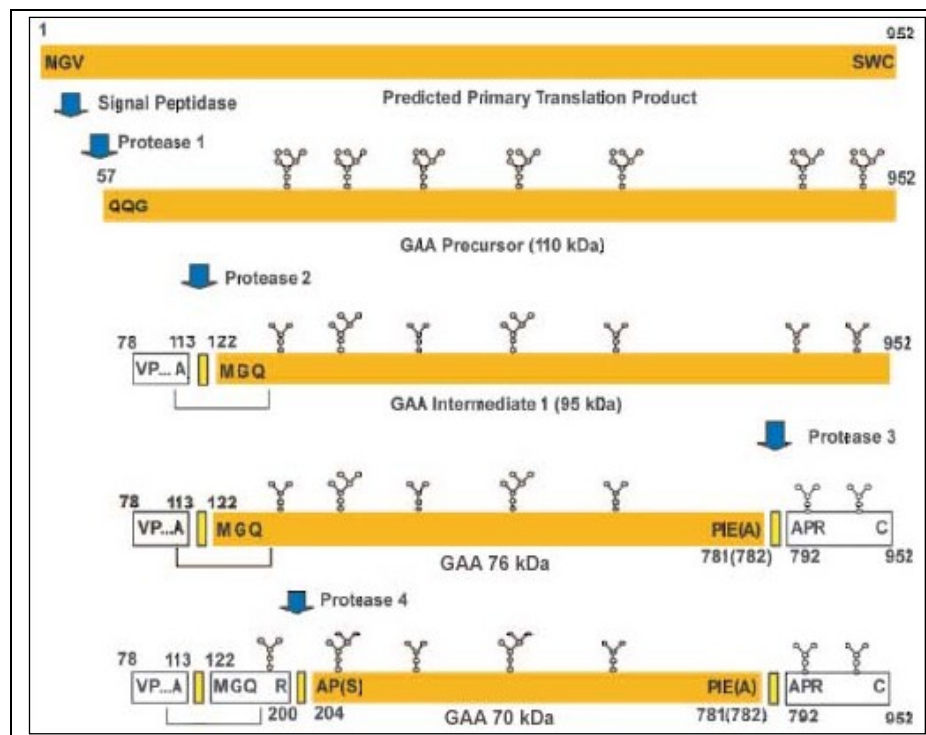


Figura 1.2 Modello della maturazione di GAA.

Nel citoplasma l'idrolisi del glicogeno ha luogo quando il glucosio è richiesto per il mantenimento dei livelli di zucchero nel sangue o come fonte di energia.

Mutazioni a carico del gene codificante GAA portano a una perdita o riduzione di attività dell'enzima con limitata degradazione del glicogeno e conseguente accumulo dello stesso.

Tali mutazioni sono causa di effetti specifici che interferiscono sulla sintesi dell'enzima, fosforilazione, processamento proteolitico, attività catalitica e localizzazione intracellulare dell' α -glucosidasi acida (Reuser et al., 1985, 1987).

La deficienza enzimatica dovuta a queste mutazioni comporta l'insorgenza della sindrome di Pompe o glicogenosi di tipo II.

1.2.3 Relazioni filogenetiche

La sequenza amminoacidica dell' α -glucosidasi acida umana mostra un'evidente omologia con le sequenze amminoacidiche della saccarasi-isomaltasi di coniglio e dell'isomaltasi umana (Hoefsloot et al., 1988) mentre non presenta alcuna omologia significativa con la sequenza di altre proteine inserite nella libreria del National Biochemical Research Foundation (NBRF). Il grado di omologia tra GAA, saccarasi e isomaltasi è sorprendentemente alto considerando le differenti caratteristiche di queste proteine. Infatti, diversamente dagli altri due enzimi che sono ancorati alla membrana apicale delle cellule epiteliali intestinali e manifestano la massima attività a pH neutro, l' α -glucosidasi acida umana è, come anzidetto, una proteina lisosomiale solubile con optimum di pH tra 4 e 5. Inoltre, il glicogeno rappresenta il substrato naturale della GAA mentre saccarasi e isomaltasi non agiscono su questo polisaccaride.

Come suggerito da Hunziker et al. (1986), il precursore a catena singola della saccarasi e isomaltasi deriva dalla duplicazione di un gene ancestrale per l'isomaltasi. Poichè l' α -glucosidasi presenta un ugual livello di omologia con questi due enzimi e non è il prodotto di un gene duplicato, si ipotizza che essa si sia evoluta dal gene ancestrale prima o contemporaneamente alla duplicazione genica.

L'evidenza di una omologia di sequenza permette di trarre alcune significative conclusioni relativamente ai domini funzionalmente importanti. In particolare, attorno al sito attivo putativo della saccarasi e della isomaltasi, è stata individuata una sequenza di sei amminoacidi che include il residuo di acido aspartico (Quaroni e Semenza, 1976; Hunziker et al., 1986) e che risulta identica ai corrispondenti amminoacidi della α -glucosidasi lisosomiale. Ciò indica che il residuo di acido aspartico è essenziale anche per l'attività catalitica di GAA.

Regioni non omologhe tra i tre enzimi considerati sono, invece, presenti sia al terminale amminico, sia a quello carbossilico e risultano svolgere ruoli importanti nella specificità di substrato e nelle modificazioni post-traduzionali.

1.2.4 La malattia di Pompe

La malattia di Pompe è un disordine autosomale recessivo del metabolismo del glicogeno, caratterizzato da una deficienza dell' α -glucosidasi acida lisosomiale. Come conseguenza di tale difetto genetico, gli individui affetti dalla malattia di Pompe sono incapaci di degradare il glicogeno che viene così accumulato all'interno dei lisosomi in molti tessuti del corpo; in particolare il muscolo cardiaco e scheletrico (Fig. 1.3, 1.4) risultano essere quelli più seriamente colpiti (Chen e Amalfitano, 2000).

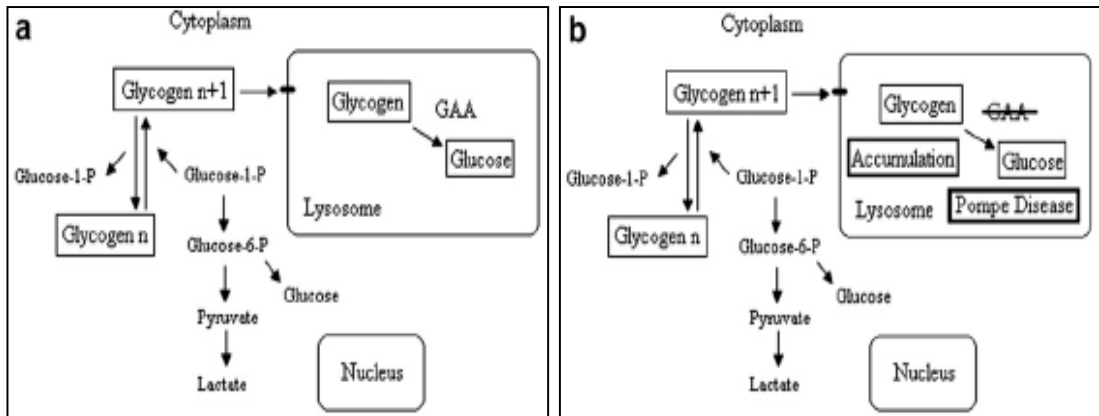


Figura 1.3 Patogenesi della malattia di Pompe. Normale degradazione del glicogeno a opera della GAA (a). Accumulo del glicogeno in condizioni di deficienza della GAA (b).

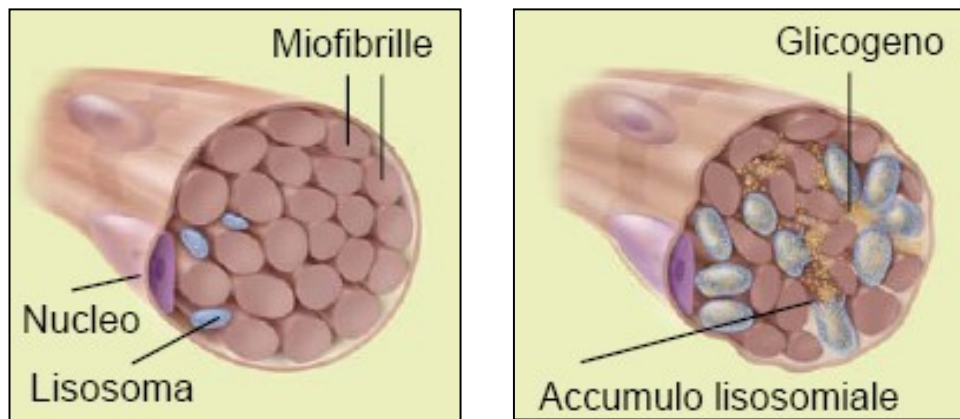


Figura 1.4 Rappresentazione schematizzata di una fibra muscolare sana (a sinistra) e una fibra muscolare affetta da glicogenosi di tipo II (a destra).

La glicogenosi di tipo II comprende una serie di fenotipi clinici che differiscono per l'età dei pazienti, ma anche per l'organo coinvolto e la severità clinica; vengono indicati come forma infantile, giovanile e adulta (Pompe, 1932; Hers, 1963; Courtecuissf et al., 1965; Engel et al., 1973; Mehler e DiMauro, 1977). Il fenotipo più severo è quello infantile, originariamente descritto da J.C. Pompe. È

caratterizzata da un ingente deposito di glicogeno in sede cardiaca e nel muscolo scheletrico, che porta invariabilmente a una rapida progressione della cardiopatia, debolezza muscolare generalizzata e ipotonia. Spesso lo sviluppo motorio si interrompe del tutto o, alternativamente, dopo il raggiungimento dei principali stadi di sviluppo motorio, si osserva una retrocessione. Normalmente, il decesso sopraggiunge per insufficienza cardiaca e/o respiratoria entro il primo anno d'età.

In uno studio retrospettivo della storia naturale in pazienti con forma infantile della malattia di Pompe (n=168), l'età mediana di esordio dei sintomi era 2 mesi e quella del decesso era 9 mesi. I tassi di sopravvivenza di Kaplan-Meier a 12, 24 e 36 mesi erano rispettivamente pari a 26%, 9% e 7%.

La forma giovanile si presenta tipicamente come difficoltà a camminare nei bambini durante l'infanzia. Questi segni fisici sono solitamente seguiti da debolezza della porzione prossimale dei muscoli e coinvolgimento del muscolo respiratorio. La morte è causata da insufficienza respiratoria e può colpire i pazienti prima della seconda decade d'età.

La forma tardiva della malattia di Pompe si manifesta nella prima infanzia, in età pediatrica, nell'adolescenza o persino nell'età adulta e ha una progressione molto meno rapida rispetto alla forma infantile. È normalmente caratterizzata da una sufficiente attività residua di GAA tale da precludere l'esordio di cardiomiopatia, ma si è riferito un coinvolgimento cardiaco fino al 4% circa dei pazienti con forma tardiva della patologia.

I pazienti con malattia di Pompe in forma tardiva presentano tipicamente miopatia progressiva, principalmente a carico dei muscoli prossimali della pelvi e del cingolo della spalla, oltre a coinvolgimento respiratorio di vario grado. In ultima analisi, questi sintomi sfociano in profonda disabilità e/o nell'esigenza di ricorrere a un supporto per la ventilazione. I tempi di progressione della patologia sono molto variabili e imprevedibili, con rapido deterioramento della funzione muscolare respiratoria e scheletrica in alcuni pazienti, seguita da perdita della deambulazione e insufficienza cardiaca, o con progressione meno celere in altri pazienti. Infine, taluni pazienti possono presentare dissociazione nella progressione del coinvolgimento muscolare scheletrico e respiratorio.

L'eterogeneità delle manifestazioni della malattia di Pompe deriva almeno in parte dal verificarsi di mutazioni diverse a carico del gene GAA lisosomiale che

influenzano in modo variabile la capacità funzionale dell'enzima mutante (Hermans et al., 2004); sono note più di 200 mutazioni del gene. Per la malattia di Pompe è disponibile un database delle mutazioni al sito www.pompecenter.nl mantenuto dal Pompe Center Erasmus MC Rotterdam e contenente la descrizione delle mutazioni più comuni per ciascun fenotipo; la natura delle mutazioni può servire per la predizione del fenotipo.

In linea generale, ciascun fenotipo clinico della malattia di Pompe è correlato con l'attività residuale dell'enzima GAA che risulta essere, rispetto i normali livelli di attività, inferiore all'1% nella forma più grave (infantile), 1-10% in quella giovanile fino a una riduzione di attività pari anche al 10-40%.

L'incidenza reale di tutte e tre le forme non è stata ancora chiarita ma le stime riportano un caso di GSD-II ogni 40000-100000 nuovi nati. In particolare, in Australia (Meikle et al., 1999) si registra un'incidenza di 1 su 201000 mentre frequenze più alte (circa 1 malato su 40000 nati) vengono rilevate negli Stati Uniti (Martiniuk et al., 1998) e in Olanda (Ausems et al., 1999).

La diagnosi clinica della glicogenosi di tipo II è confermata dall'assenza virtuale (nella forma infantile) o dalla riduzione dell'attività α -glucosidasi nelle biopsie muscolari e nelle colture di fibroblasti. La diagnosi prenatale può essere condotta determinando l'attività dell' α -glucosidasi acida nelle cellule amniotiche coltivate (Park et al., 1992) e attraverso l'analisi delle mutazioni (Kleijer et al., 1995). Gli strumenti diagnostici attuali non sono tuttavia vantaggiosi e convenienti per uno screening su larga scala.

1.2.5 La GAA come farmaco nella terapia enzimatica sostitutiva

Attualmente non sono disponibili trattamenti definitivi per la cura della malattia di Pompe. Tuttavia, sono state sviluppate due strategie principali: la terapia enzimatica sostitutiva (ERT) e la terapia genica.

La terapia genica mira a compensare le mutazioni responsabili della GSD-II negli individui malati attraverso la veicolazione di DNA contenente la versione funzionale del gene; questo tipo di terapia è tutt'ora irrealizzabile nella pratica clinica e appare di difficile conseguimento. Per questo motivo si sta cercando di seguire vie alternative di compensazione genica. In tale contesto si inseriscono studi condotti con oligonucleotidi chimerici RNA-DNA che possono riparare

alterazioni di una o due basi nei genomi delle cellule di mammifero. Un'altra possibilità è rappresentata dall'aggiunta della sequenza codificante della GAA normale nelle cellule ammalate mediante vettori virali per il trasferimento dell'informazione necessaria (Nicolino et al., 1998; Pauly et al., 1998; Zaretsky et al., 1997; Amalfitano et al., 1999). I ricercatori hanno ottenuto qualche successo utilizzando diversi ceppi di un comune virus del raffreddore per trasferire il gene dell' α -glucosidasi lisosomiale nei topi e nelle quaglie. In entrambi i modelli animali, il virus ha raggiunto il fegato, dove ha cominciato a produrre l'enzima e a inviarlo alle cellule muscolari attraverso il flusso sanguigno. Nel giro di qualche giorno l'enzima ha cominciato a eliminare il glicogeno, migliorando la funzionalità muscolare. Sono necessari, tuttavia, ulteriori studi per valutare la sicurezza e l'efficacia di questo approccio prima che possa essere tentato negli esseri umani. Secondo molti autori la terapia enzimatica sostitutiva risulta essere più vantaggiosa e costituisce, infatti, la strada principale per il trattamento dei pazienti affetti da glicogenosi di tipo II. Essa consiste nella somministrazione dell'enzima mancante per via venosa. Purtroppo, le fonti naturali di GAA sono scarse e non disponibili a causa della richiesta di sicurezza, specificità ed efficacia. L'enzima somministrato deve, infatti, rispondere a una serie di requisiti:

- a) essere stabile;
- b) raggiungere i lisosomi dei tessuti bersaglio;
- c) mantenere l'attività catalitica dopo la somministrazione;
- d) non indurre risposta immunitaria.

Per motivi di immunogenicità, i lavori più recenti si sono focalizzati sulla produzione di GAA ricombinante umana (rhGAA). La rhGAA è stata prodotta in latte di coniglio transgenico e in cellule di ovaio di criceto (CHO) (Bijvoet et al., 1999; Van Hove et al., 1996).

Attualmente è disponibile in commercio il farmaco Myozyme[®] della ditta Genzyme Corp., contenente una forma ricombinante dell' α -glucosidasi acida umana, prodotta mediante tecnologia del DNA ricombinante utilizzando colture di cellule di ovaio di criceto (CHO). Il trattamento con Myozyme prevede la somministrazione del farmaco alla dose consigliata di 20 mg/kg di peso corporeo, una volta ogni 2 settimane come infusione endovenosa sotto controllo

di un medico esperto nella gestione dei pazienti affetti da malattia di Pompe o da altre patologie metaboliche o neuromuscolari ereditarie.

Le infusioni vanno somministrate in modo incrementale. La velocità iniziale di infusione consigliata è 1 mg/kg/ora, con graduale aumento di 2 mg/kg/ora ogni 30 minuti in assenza di reazioni connesse all'infusione, fino al raggiungimento del tasso massimo di 7 mg/kg/ora.

La sicurezza e l'efficacia di Myozyme sono state prevalentemente valutate nei bambini, in una fascia d'età compresa tra l'infanzia e l'adolescenza.

Gran parte dei pazienti con malattia di Pompe infantile trattati con Myozyme mostra una migliorata funzione cardiaca, oltre che una stabilizzazione o un miglioramento dei parametri della crescita. Nel contempo, le risposte motorie e respiratorie alla terapia sono state più variabili. Va rilevato che i pazienti con GSD-II infantile rispondenti sul piano motorio sono stati quelli che alla visita basale mostravano funzione motoria più conservata e un ridotto tenore di glicogeno nel muscolo quadricipite.

1.3 La β -glucosidasi acida umana (GCasi)

1.3.1 Generalità sulla β -glucosidasi acida umana

La β -glucosidasi acida umana (D-glucosil-acilsfingosina glucoidrolasi, EC 3.2.1.45; in breve GCasi) (Grabowski et al., 1990) è una glicoproteina lisosomiale associata alla membrana appartenente alla famiglia delle idrolasi (Weinreb et al., 1968) e responsabile dell'idrolisi del legame β -glucosidico dell'*N*-acilsfingosil-1-O- β -D-glucoside, suo substrato naturale, e di composti β -glucosidici sintetici.

L'*N*-acilsfingosil-1-O- β -D-glucoside è anche chiamato glucosilceramide, ceramide β -glucoside o glucocerebroside. Pertanto ulteriori denominazioni dell'enzima sono glucosilceramidasi, ceramide β -glucosidasi, glucocerebrosidasi o β -glucosidasi acida.

1.3.2 Caratteristiche strutturali, funzionali e sintesi della proteina

La β -glucosidasi acida umana catalizza la reazione di idrolisi del glucosilceramide in ceramide e glucosio (Brady et al., 1965) attraverso la scissione del legame β -glucosidico (Fig. 1.5). Da un punto di vista strutturale il glucosilceramide mostra affinità con il galattosilceramide presente come prodotto anabolico nel sistema nervoso, principalmente nella guaina mielinica, ma praticamente assente negli altri tessuti. Diversamente, il glucosilceramide è ampiamente distribuito in molti tessuti mammiferi come intermedio metabolico nella sintesi o degradazione di glicosfingolipidi più complessi, come i gangliosidi o globosidi.

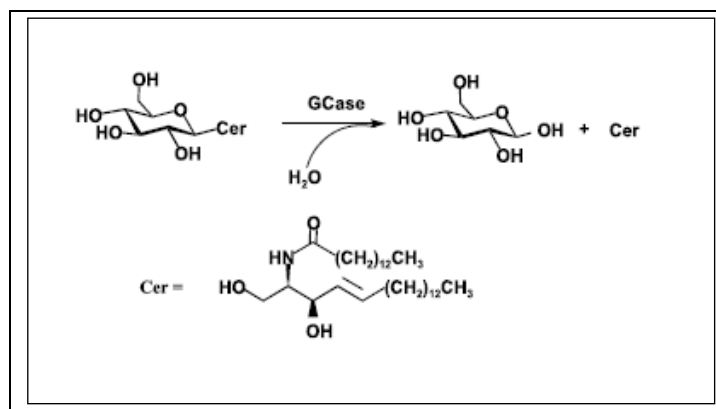


Figura 1.5 Reazione di idrolisi catalizzata dalla GCasi.

La forma di GCasi estraibile mediante detergenti è presente in una varietà di tessuti e costituisce quasi il 70-90% della GCasi totale (Grabowski et al., 1990; Imai, 1985). L'enzima non presenta segmenti transmembrana putativi, nè code citoplasmiche rilevabili (Erickson et al., 1985). *In vitro* la proteina purificata, completamente delipidizzata manifesta un'elevata affinità per le membrane fosfolipidiche cariche negativamente; l'attacco a tali membrane è richiesto per ricostituire l'attività (Qi e Grabowski, 1998; Basu et al., 1985). Nei lisosomi l'attacco a simili interfacce è probabilmente essenziale per l'attività *in vivo*.

La sequenza amminoacidica parziale (Osiecki et al., 1986) e completa (Tsuji et al., 1986) è stata determinata chimicamente dall'enzima placentale e dedotta dalle sequenze nucleotidiche isolate da librerie a cDNA ottenute da fibroblasti, tessuto epatico e linfoidi (Ginns et al., 1984; Sorge et al., 1985; Reiner et al., 1987; Graves et al., 1988). Eccetto per la presenza di artefatti di clonazione e mutazioni casuali ascrivibili alla malattia di Gaucher, le sequenze

amminoacidiche sono risultate identiche. Tali risultati, unitamente ad analisi di immunoelettroblotting condotti in diversi tessuti sani, suggeriscono che un singolo polipeptide venga tradotto a partire dall'mRNA specifico per la GCasi in tutti i tessuti.

La proteina umana matura (Fig. 1.6) si compone di 497 amminoacidi (Furbish et al., 1977) con una massa molecolare compresa tra 59 e 67 KDa a causa di un'irregolare glicosilazione (Beutler e Grabowski, 2001). Sono, infatti, presenti cinque presunti siti di N-glicosilazione, ma solo quattro sono effettivamente occupati (Erickson et al., 1985). La glicosilazione al primo sito svolge un ruolo essenziale nello sviluppo di forme enzimatiche cataliticamente attive, come dimostrato da studi di mutagenesi sito-diretta (Berg-Fussman et al., 1993).

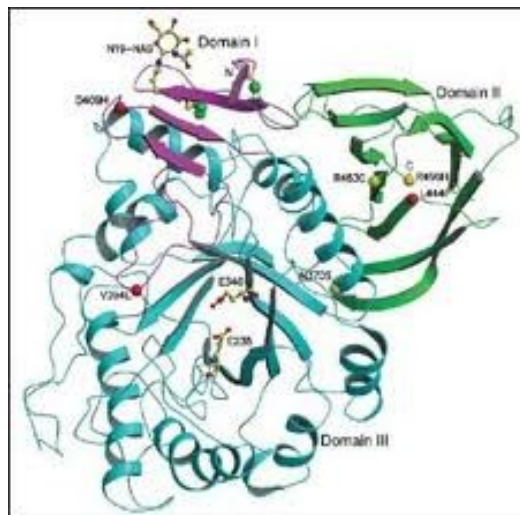


Figura 1.6 Struttura tridimensionale di GCasi.

La composizione amminoacidica del polipeptide si configura costituita per il 13% da residui basici (lisina, arginina, istidina), per l'11% da residui di leucina mentre il 45% è rappresentato da amminoacidi non polari. Il pI calcolato è pari a 7.2. Nonostante l'abbondanza dei residui non polari e di leucine, simulazioni al computer basate sull'uso di diversi algoritmi e scale di idrofobicità non hanno individuato sequenze di amminoacidi idrofobici coerenti con la presenza di domini transmembrana nella proteina matura. Inoltre, studi di digestione proteolitica ottenuti da traduzioni *in vitro* in presenza di membrane microsomiali non hanno rilevato un largo dominio citoplasmatico (Erickson et al., 1985). Tali evidenze indicano che la β -glucosidasi acida non sia una proteina

transmembrana bensì una proteina associata alla membrana, sebbene non sia possibile riconoscere alcuna prova dell'associazione stretta alla membrana sulla base della sequenza primaria della proteina.

Come tipicamente avviene per le idrolasi lisosomiali (Hasilik e Neufeld, 1980), l'enzima normale è sintetizzato a partire da una proteina precursore portante una sequenza segnale che subisce un taglio proteolitico durante il transito attraverso la membrana del reticolo endoplasmico (Erickson et al., 1985).

Nessun processamento proteolitico addizionale avviene nel corso del trasporto dall'apparato di Golgi al lisosoma (Bergmann e Grabowski, 1989; Erickson et al., 1985). Successivamente alla sintesi e al *folding* iniziale a pH neutro nel reticolo endoplasmico, la proteina è indirizzata ai lisosomi dove il pH acido fornisce l'ambiente ottimale per la funzione e l'avvolgimento stabile dell'enzima. Il polipeptide maturo, non glicosilato presenta un peso molecolare calcolato di circa 56 KDa. La glicosilazione co-traduzionale ha luogo a livello di quattro dei cinque siti predetti, Asn-Xaa-(Ser/Thr). La caratterizzazione completa della glucocerebrosidasi placentale dimostra la presenza di tipiche strutture complesse bi- e triantennarie.

Diversamente dagli enzimi lisosomiali solubili come GAA, GCasi non contiene residui di mannosio-6-fosfato (M-6-P). Esperimenti condotti sia su cellule di pazienti affetti da mucopolipidosi II (in cui il traffico lisosomiale attraverso la via del M-6-P viene interrotto), sia su fibroblasti di individui colpiti da malattia a cellule I (causata da una mancanza di fosfotransferasi) hanno evidenziato livelli normali dell'enzima GCasi (Owada e Neufeld, 1982). Ciò indica che la proteina non utilizza il sistema di targeting mediato da recettori M-6-P per la localizzazione all'interno del lisosoma.

1.3.3 Organizzazione genomica

Le sequenze complete del gene strutturale codificante la β -glucosidasi acida umana (7604 bp) e del suo pseudogene sono state studiate da Horowitz et al. (1989); alcune correzioni alla sequenza genica originale sono state riportate successivamente da Beutler et al. (1992).

Il gene GCasi è localizzato sul braccio lungo del cromosoma 1 in posizione 21 (1q21) (Barneveld et al., 1983; Ginns et al., 1985) e contiene 11 esoni e 10 esoni. Il pseudogene si trova 16 Kb a valle rispetto al gene funzionale (Zimran et

al., 1990) con cui mantiene un alto grado di omologia (96%). Esso risulta di dimensioni inferiori (circa 5 Kb) rispetto al gene funzionale per l'effetto di delezioni nelle regioni introniche 2 (313 bp), 4 (626 bp), 6 (320 bp) e 7 (277 bp), e di due delezioni esoniche (esone 9, 55 bp; esone 4, 5 bp). La presenza del pseudogene altamente omologo aumenta la frequenza di riarrangiamenti cromosomiali e di allineamenti non corretti, processi che rendono ragione dell'elevato numero di alleli ricombinanti individuati nel morbo di Gaucher. Tali differenze e omologie rivestono, quindi, un ruolo importante nell'identificazione e caratterizzazione delle mutazioni responsabili della comparsa della malattia nonché nell'accuratezza diagnostica dei test genetici per l'individuazione delle sue varianti.

L'analisi delle sequenze genomiche a monte del gene strutturale e dello pseudogene indicano la presenza di elementi promotori in queste regioni (Reiner et al., 1988).

Il promotore del gene strutturale contiene due presunte TATA-box e CAAT-box identificate a circa 260 bp a monte del codone di inizio ATG; più precisamente la prima TATA-box tra la posizione -23 e -27 dal sito di inizio trascrizione, la seconda tra -33 e -38 e le due CAAT-box tra -90 e -94 e -96 e -99 (Grabowski et al., 1990).

Il cDNA di GCasi è lungo circa 2.5 Kb. Studi di traduzione *in vitro* della proteina hanno evidenziato la presenza di due ATG entrambi utilizzati come segnali di inizio traduzione. Nonostante l'importanza/funzionalità relativa di questi due codoni *start* sia ancora sconosciuta, studi compiuti *in vivo* dimostrano che entrambi i siti possono produrre forme funzionali dell'enzima.

Un'elevata espressione di GCasi nelle cellule animali non è possibile poiché è presente la proteina TCP80, che inibisce la traduzione della proteina di interesse mediante legame all'RNA messaggero (Xu e Grabowski, 1998).

1.3.4 Relazioni filogenetiche

La sequenza amminoacidica dell'enzima umano non presenta omologia con quella di lievito e pianta (Beutler e Grabowski, 2001) mentre mostra un'elevata omologia (88%) con quella murina. Il 100%, il 91% e l'89% dei residui di cisteina (7), prolina (34) e glicina (34) risultano rispettivamente conservati.

Presenza e localizzazione delle 5 sequenze consenso per l'N-glicosilazione sono altamente conservate ma solo quelle tra gli amminoacidi 146-148 e 462-464 appaiono identiche nei due organismi (Beutler e Grabowski, 2001). Questo risultato, suffragato da ulteriori esperimenti di espressione (Grace et al., 1989), suggerisce un potenziale ruolo funzionale della glicosilazione nell'attività catalitica.

1.3.5 La malattia di Gaucher

La malattia di Gaucher (GD), la più comune patologia da accumulo lisosomiale, è determinata dalla deficienza nell'attività della β -glucosidasi acida umana, causata da mutazioni a carico del gene codificante l'enzima.

E' una malattia ereditaria autosomica recessiva con un'incidenza di 1:57000-200000 nella popolazione caucasica e 1:640-10000 nella popolazione ebraica di origine Ashkenazy (Cox e Schofield, 1997; Meikle et al., 1999; Beutler e Grabowski, 2001).

La deficienza enzimatica si riflette in un accumulo di sfingoglicolipidi, principalmente glucocerebrosidi, nei macrofagi presenti in fegato, milza, midollo osseo, ossa e sistema nervoso centrale (Beutler e Grabowski 2001). A causa dell'accumulo di sostanze lipidiche, si assiste a un ingrossamento dei macrofagi stessi (Fig. 1.7) e alla comparsa di una sintomatologia particolare ascrivibile alla malattia di Gaucher, che include epatosplenomegalia, anemia, trombocitopenia e lesioni scheletriche (Grabowski e Hopkin, 2003; Jmoudiak e Futerman, 2005).

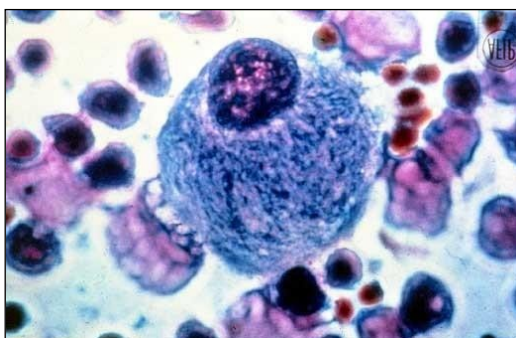


Figura 1.7 Immagine di un macrofago ingrossato (cellula di Gaucher) in un preparato istologico ottenuto da un paziente affetto dalla malattia.

Da un punto di vista clinico, la GD, sebbene altamente variabile, può essere suddivisa in tre classi principali sulla base della sintomatologia, del decorso

clinico e delle manifestazioni neurologiche (Tab. 1.2). La classe di tipo 1, anche nota come GD non neuropatica, colpisce bambini e adulti ed è il fenotipo più comune, caratterizzato dall'assenza di compromissione del sistema nervoso centrale (SNC). I pazienti di tipo 1 mostrano un ampio spettro di sintomi clinici, che comprendono organomegalia, citopenia, interessamento scheletrico, e un'età variabile di comparsa della malattia.

In presenza di coinvolgimento del SNC, a seconda dell'età di insorgenza e della sintomatologia neurologica, si distinguono due classi: la forma infantile (tipo 2) e quella giovanile (tipo 3), rispettivamente indicate come forma acuta e subacuta. La classe di tipo 2 rappresenta il fenotipo più raro e severo per l'accumulo consistente di glucosilceramide in una varietà di tessuti e cellule, tra cui il SNC; tale condizione conduce a una rapida progressione neurodegenerativa con esito fatale entro i 2-3 anni di vita.

Il fenotipo di classe 3, per definizione, include tutti i pazienti con manifestazioni neurologiche che siano sopravvissuti ai primi anni di vita.

E' stato, inoltre, descritto un fenotipo intermedio tra le forme neuropatiche 2 e 3 (Goker-Alpan, 2003)

Caratteristiche cliniche	Tipo 1	Tipo 2	Tipo 3
Insorgenza	Infanzia / maturità	Infanzia	Infanzia
Epatosplenismo	+	-	-
Ipersplenismo	+	-	-
Fragilità ossea	+	-	-
Decorso neurodegenerativo	-	---	--
Sopravvivenza	6-80 + anni	<2 anni	da 2° a 4° decade
Predilezione etnica	ebraica	panetnico	nord Svezia

Tabella 1.2 Tipi clinici della malattia di Gaucher.

1.3.6 Principali sedi di accumulo delle cellule di Gaucher

I principali organi colpiti, sedi di accumulo delle cellule di Gaucher, sono fegato, milza, midollo osseo, ossa e il sistema nervoso centrale. Di seguito viene brevemente trattata la sintomatologia clinica associata a ciascun tessuto e organo colpito.

1. FEGATO

Quando la malattia è grave, il fegato può ingrossarsi così tanto da occupare buona parte della cavità addominale; solitamente si presenta consistente e liscio, con i lobi destro e sinistro aventi circa le stesse dimensioni. Fegati notevolmente ingrossati sono in genere duri e hanno una superficie irregolare. L'anormale volume di quest'organo può causare dolore al paziente. Minori anomalie del fegato, presenti in pazienti meno gravemente ammalati, si traducono in un incremento dell'attività transaminasica e γ -glutamyl transferasica nel sangue. L'iperbilirubimia è solitamente causa di infezione (Beutler e Grabowski 2001).

2. MILZA

La milza è la sede classica di accumulo delle cellule di Gaucher. In condizioni normali la milza distrugge le cellule del sangue invecchiate che vengono sostituite da quelle neoprodotte dal midollo osseo. L'accumulo di cellule di Gaucher nella milza induce un aumento del suo volume e della sua attività. Ciò causa una distruzione più rapida delle cellule del sangue, per cui:

- la carenza di globuli rossi, deputati al trasporto dell'ossigeno, determina un'insufficiente ossigenazione dei tessuti: di conseguenza, le cellule muscolari, ad esempio, non producono l'energia necessaria al mantenimento e il soggetto si sente molto stanco;
- la carenza di globuli bianchi riduce la capacità dell'organismo a combattere le infezioni batteriche;
- la carenza di piastrine determina un difetto di coagulazione del sangue. Sono pertanto frequenti ematomi e sanguinamenti (es. ecchimosi nasali).

Nei pazienti dove si è operata una rimozione della milza, il fegato può ingrossarsi drammaticamente in quanto viene a mancare la sede di accumulo splenico.

3. MIDOLLO OSSEO

Il midollo osseo, deputato alla produzione delle cellule del sangue, è l'altra sede tipica delle cellule di Gaucher. L'accumulo di tali cellule riduce la sua funzionalità in particolare nelle zone in cui è maggiore la loro concentrazione, con le seguenti conseguenze:

- riduzione della produzione delle cellule del sangue; tale evento aggrava la situazione determinata dall'iperfunzionamento della milza;
- riduzione del flusso sanguigno: la mancanza di un sufficiente apporto di ossigeno provoca "crisi ossee" estremamente dolorose, o addirittura distruzione del tessuto osseo (necrosi asettica) con conseguenti problemi di mobilità.

4. OSSA

I pazienti affetti da malattia di Gaucher spesso lamentano dolori alle articolazioni e alle ossa.

Le ossa dei pazienti colpiti dalla malattia di Gaucher sono più sottili e deboli del normale e possono deformarsi o essere soggette a fratture spontanee: se quest'ultime avvengono a livello di colonna spinale (fratture da compressione) i pazienti possono subire notevoli danni neurologici.

5. SISTEMA NERVOSO CENTRALE

Nelle forme neurologiche, il sistema nervoso centrale viene colpito da un notevole accumulo di cellule di Gaucher nel cervello e da una compressione della spina dorsale in seguito al collasso delle vertebre (Beutler e Grabowski 2001).

1.3.7 Mutazioni responsabili della malattia di Gaucher

Sono state identificate più di 250 mutazioni a carico del gene GCasi, che causano l'assenza totale o parziale di attività enzimatica della proteina (Horowitz et al., 1989; Beutler, 1992; Firon et al., 1990; Latham et al., 1990; Grace et al., 1994; Beutler and Gelbart, 1996; Tayebi et al., 2003; Beutler e Grabowski, 2001). Tali mutazioni includono inserzioni, delezioni, transizioni ed eventi di ricombinazione con il pseudogene. Nonostante vi sia una ricca letteratura riguardante lo screening di specifici gruppi di pazienti in funzione della presenza

o assenza di alcune mutazioni comuni, la comprensione delle correlazioni genotipo-fenotipo appare ancora non del tutto chiara. Esiste, infatti, una variazione significativa non solo all'interno del medesimo gruppo fenotipico ma anche tra pazienti con identico genotipo. Numerosi studi hanno dimostrato che tipo clinico e severità della malattia possono variare tra fratelli, anche tra gemelli omozigoti, e tra pazienti portanti le medesime mutazioni (Sidransky et al., 1994; Grabowski, 1997; Lachmann et al., 2004). La validità della predizione fenotipica sulla base del genotipo risulta, dunque, limitata.

Tuttavia, una variabilità nella frequenza di specifici alleli mutati è stata osservata in funzione del tipo di popolazione preso in esame (Diaz et al., 1998). In particolare, in gruppi di pazienti di diversa provenienza sono state analizzate alcune mutazioni, accertate come responsabili del morbo di Gaucher:

- N370S: transizione A → G: produce una proteina matura con una sostituzione Asn → Ser in posizione 370;
- L444P: transizione T → C al nucleotide 1448 che causa la sostituzione Leu₄₄₄ → Pro₄₄₄;
- 84insG: risulta dall'aggiunta di una guanina in posizione 84;
- IVS2+1: transizione G → A al sito di taglio in 5' del secondo introne del gene GCasi.

I risultati ottenuti indicano che i difetti genetici N370S, L444P, 84insG, IVS2+1 costituiscono il 90-96% degli alleli mutati in pazienti Ashkenazy (Beutler et al., 1992), mentre nelle popolazioni non ebraiche le mutazioni N370S e L444P si registrano rispettivamente nel 38% e 33% dei casi (Beutler e Gelbart, 1993; Horowitz et al., 1993). In particolare, la mutazione L444P è presente con un'elevata frequenza nella popolazione che vive nel nord della Svezia, ma viene riconosciuta anche in altre popolazioni benché con un'incidenza minore (Beutler e Grabowski 2001). In effetti, nei pazienti Gaucher giapponesi L444P appare relativamente comune mentre le mutazioni N370S e 84insG non compaiono.

Studi condotti su individui Gaucher di origine portoghese hanno evidenziato la presenza del difetto N370S con una frequenza pari al 63% e la comparsa di due mutazioni rare (G377S e N396T) (Amaral et al., 1993; 1999).

1.3.8 La GCasi come farmaco nella terapia enzimatica sostitutiva

La malattia di Gaucher è divenuta un prototipo per la diagnosi e per la cura delle malattie genetiche rare attraverso la terapia enzimatica sostitutiva. Attualmente è presente in commercio l'imiglucerasi (Cerezyme®, Genzyme Corp.), un enzima prodotto in cellule trasformate di CHO (cellule di ovario di criceto) per la cura di pazienti Gaucher di tipo 1. La sequenza amminoacidica dell'imiglucerasi differisce da quella inizialmente rilevata per GCasi, purificata dalla placenta umana, per la presenza di un residuo di arginina in luogo di una istidina in posizione 495. Va rilevato che His₄₉₅ o Arg₄₉₅ è un polimorfismo naturale che non dà luogo a fenomeni patologici; entrambe le sequenze si ritrovano, infatti, in soggetti sani.

L'enzima prodotto in cellule CHO è successivamente sottoposto a un processo di deglicosilazione parziale (Grabowski et al., 1998) allo scopo di esporre residui di mannosio, necessari per consentirne l'internalizzazione e l'uptake da parte dei macrofagi attraverso la via dei recettori del mannosio (Furbish et al., 1981) (Fig. 1.8). Pertanto, l'imiglucerasi differisce dalla GCasi di placenta anche per la presenza di catene glicaniche raccorciate.

I risultati clinici ottenuti dall'applicazione della terapia su pazienti Gaucher hanno chiaramente dimostrato la sua efficacia nella cura di forme non neurologiche della malattia.

Altre forme di GCasi ricombinanti approvate dall'FDA includono la velaglucerasi alfa (Shire HGT, Cambridge, MA, USA), forma gene-attivata della GCasi nativa, e la taliglucerasi alfa (Shaaltiel et al., 2007) prodotta in colture cellulari di carota dalla ditta Protalix. Entrambi le forme di GCasi hanno concluso la fase di sperimentazione III.

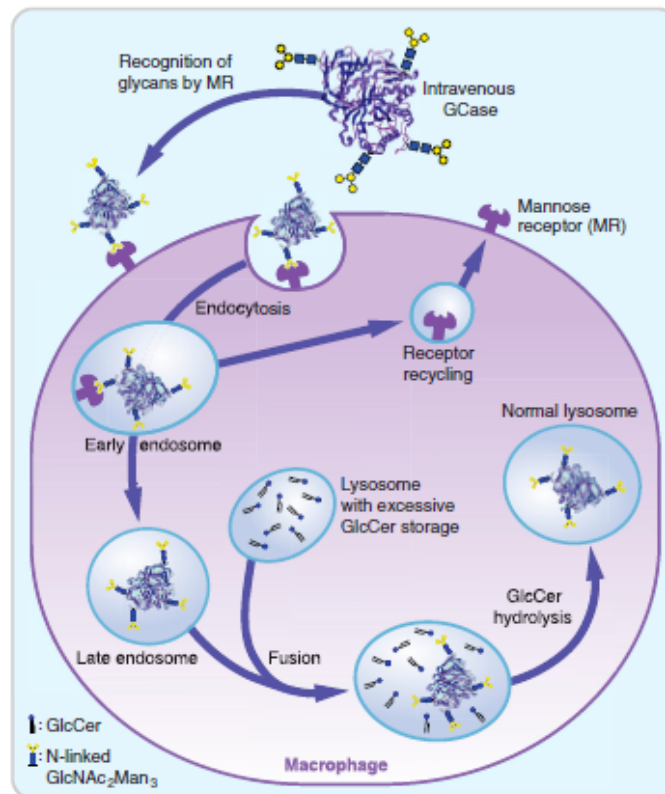


Figura 1.8 Rappresentazione dell'uptake di GCasi ricombinante somministrata per via intravenosa. L'enzima ricombinante prodotto in cellule CHO è modificato per esporre gli N-glicani con residui di mannosio terminali che conferiscono alta affinità per i recettori del mannosio (MR) residenti nei macrofagi. L'internalizzazione della GCasi avviene tramite un processo di endocitosi del complesso recettore MR/GCasi.

1.4 Terapia enzimatica sostitutiva nel trattamento dei pazienti affetti da malattia di Gaucher o da glicogenosi di tipo II: problematiche generali e costi della terapia

La malattia di Gaucher e la glicogenosi di tipo II, come tutti i disordini da accumulo lisosomiale, rientrano nella categoria delle malattie rare (MR). Nell'UE è considerata "rara" una patologia che colpisce non più di 5 persone su 10000 abitanti, che equivale a circa 253000 persone negli Stati membri della UE. La bassa incidenza nella popolazione non significa però che i pazienti affetti da MR siano pochi dal momento che il numero di MR conosciute e diagnosticate oscilla tra i 7000 e gli 8000, arrivando a colpire circa il 6-8% della popolazione totale; in altre parole, tra i 27 e i 36 milioni di persone nell'Unione Europea e circa 2 milioni in Italia. La maggior parte di tali malattie è dovuta a difetti genetici e si manifesta

con sintomi che possono comparire non solo alla nascita o durante l'infanzia ma anche in età adulta, comportando spesso disabilità e debilitazione cronica fino a un esito infausto nei casi più severi.

La rarità della malattia si traduce in un minore interesse della ricerca eziologica e patogenetica; una maggiore difficoltà nel descrivere la storia naturale con le sue possibili varianti e nel progettare ricerche cliniche; un mercato minore, incapace di ammortizzare i costi relativi alla ricerca farmacologica specifica. Quando si parla di MR, infatti, si deve parlare, e non si può non farlo, di farmaci orfani, che proprio a causa della frammentazione delle singole patologie, faticano a incontrare l'interesse economico delle case farmaceutiche. Le MR pongono, dunque, problematiche impegnative a vari livelli: dalla diagnosi alla presa in carico del paziente; dall'identificazione di appropriate soluzioni terapeutiche alla ricerca. Sebbene solitamente non esistano cure definitive per le MR, diagnosi precoci seguite da adeguati trattamenti possono contribuire a migliorare la qualità di vita dei pazienti e aumentarne la speranza di vita. Tali malattie costituiscono, quindi, una sfida interessante per le applicazioni mediche che sfruttano le tecnologie del DNA ricombinante.

Nel contesto delle MR e delle problematiche ad esse associate, si inseriscono gli studi legati alla terapia enzimatica sostitutiva nel trattamento della malattia di Gaucher e della glicogenosi di tipo II, volti a sviluppare piattaforme industriali tali da aumentare l'accessibilità dei farmaci, agendo a livello di costi, ma anche di quantitativi effettivamente disponibili degli stessi. È importante sottolineare, inoltre, che la presenza di farmaci registrati è limitata ai Paesi economicamente ricchi dove i sistemi sanitari nazionali possono coprirne i relativi costi. L'espansione del mercato non sarebbe comunque possibile a causa di fattori tecnologici limitanti che determinano una sottoproduzione del farmaco rispetto alle reali esigenze. Ne sono dimostrazione le recenti crisi del farmaco registrate per la malattia di Pompe, Gaucher e Fabry. Il costo per lo sviluppo di farmaci in questa area dipende, infatti, dall'efficacia del sistema di produzione utilizzato nella sintesi degli enzimi; tale costo risulta particolarmente elevato nel caso della produzione di GCasi e GAA in cellule CHO.

L'imiglucerasi è ritenuta uno dei farmaci più costosi al mondo a causa soprattutto delle basse rese del sistema produttivo per la presenza negli animali della proteina inibitrice TCP80 e delle modificazioni strutturali che deve subire per

essere riconosciuta e internalizzata dalle cellule bersaglio umane. In molte cliniche in cui viene effettuata la ERT, il costo della terapia si è attestato sui 270.000 € annui per uomo adulto. Le vendite annue mondiali di imiglucerasi sono state stimate essere pari a un miliardo di euro (Beutler, 2006).

Anche per quanto concerne il trattamento della glicogenosi di tipo II con il farmaco commerciale Myozyme[®] sono emerse criticità rilevanti, che ne limitano la disponibilità:

1. la metodologia attualmente impiegata pone seri limiti alla produzione di grandi quantità di farmaco. Questo è un problema rilevante ai fini di una completa efficacia della terapia, considerato che la risposta terapeutica è dose-dipendente;
2. la produzione su CHO è molto costosa e determina un prezzo di commercializzazione del farmaco estremamente elevato (circa 500 euro per flacone da 50 mg). Poiché il protocollo terapeutico prevede la somministrazione di 20 mg/kg ogni 15 giorni, il farmaco necessario al trattamento di un paziente adulto del peso medio di 70 kg costa al sistema sanitario nazionale circa 340.000 €/anno.

Inoltre, la necessità di infondere elevate dosi di enzima GCasi e GAA a cadenza bisettimanale per tutta la vita dei pazienti e gli alti costi del farmaco impongono lo sviluppo di metodi più economici per la produzione dell'enzima sostitutivo.

Per questi motivi, è importante esplorare il potenziale di bioproduzione di piante appositamente trasformate per assumere la funzione di fabbriche cellulari di composti terapeutici. Produrre in pianta la proteina biologicamente attiva a costi accessibili potrebbe consentire ai pazienti colpiti da malattia di Pompe e Gaucher il trattamento con l'enzima sostitutivo e offrire loro una speranza di vita e una buona qualità di vita.

1.5 Le piante come bioreattori verdi

Le piante si sono rivelate nei secoli organismi estremamente duttili, capaci di adattarsi alle necessità dell'uomo, sia per scopi agricoli che ornamentali. Circa trent'anni fa, con i primi studi sulla tecnica del DNA ricombinante, è anche stata scoperta la possibilità di introdurre e far esprimere in un tessuto vegetale geni derivanti da altre specie vegetali, batteriche o animali.

Questa scoperta ha aperto nuovi orizzonti per l'enorme numero di applicazioni che si potevano proporre. I primi lavori sono stati rivolti al miglioramento delle piante stesse, per la loro difesa contro organismi patogeni e infestanti. A questo scopo sono stati introdotti geni di resistenza di origine batterica, fungina o vegetale. Da qui ad immaginare che fosse possibile far produrre alla pianta altri generi di proteine, non necessariamente a suo vantaggio, il passo è stato breve. In particolare, grande interesse ha suscitato la possibilità di produrre molecole a uso sanitario, come antigeni per la produzione di vaccini, anticorpi e loro frammenti, o proteine con funzione farmacologica (molecular farming).

L'impiego di bioreattori vegetali può rappresentare la giusta soluzione per lo sviluppo di proteine farmaceutiche a bassi costi di produzione. Le piante, infatti, possono sintetizzare elevate quantità di proteina ad alto profilo qualitativo grazie alle modificazioni post-traduzionali simili a quelle animali, con parziale eccezione per i processi di glicosilazione (Fischer e Emans, 2000). Alcuni esempi di proteine farmaceutiche prodotte in pianta sono elencati in tabella 1.3.

Proteina	Pianta ospite	Applicazioni/commenti
Ormone della crescita	tabacco, girasole	prima proteina umana in pianta
Siero albumina	tabacco, patata, riso	cirrosi epatica
Fattore di crescita epidermide	tabacco	riparazione ferite
Interferone α	riso, rapa	epatite B e C
Eritropoietina	tabacco	anemia
GM-CSF	tabacco	neutropenia
IL-2 e IL-4	sosp. cell tabacco	ematopoiesi
Antitripsina α -1	riso	fibrosi cistica
Fosfatasi alcalina di placenta	tabacco	prodotta per secrezione da foglia
Aprotinina	mais	inibitore della tripsina
Collagene	tabacco	primo polimero strutturale umano

Tabella 1.3 Esempi di sintesi di proteine farmaceutiche in pianta (a esclusione di vaccini, anticorpi, nutraceutici).

1.5.1 Il sistema di espressione in pianta: le modificazioni post-traduzionali

Le piante vengono modificate geneticamente allo scopo di produrre proteine eterologhe di alto valore qualitativo destinate all'uso farmacologico e terapeutico. Al fine di ottenere una specifica proteina in pianta, il DNA codificante viene inserito nel genoma vegetale attraverso diversi metodi di trasformazione. Il DNA

inserito è poi tradotto in proteina che viene indirizzata e accumulata in uno specifico tessuto. Attualmente la produzione si è concentrata sull'uso di foglie in sistemi ospite come il tabacco e di tessuti di riserva come i semi nei cereali. Sono state sviluppate tecniche per l'ottimizzazione dei livelli di espressione e il targeting delle proteine in particolari compartimenti cellulari (Stoger et al., 2004). La struttura e la funzionalità di una proteina è data dalla sua sequenza amminoacidica che, a sua volta, ne determina la conformazione tridimensionale o struttura. I ponti disolfuro e i legami a idrogeno che si instaurano tra gli amminoacidi conferiscono alla molecola la sua forma nativa. Le proteine subiscono, inoltre, ulteriori processi post-traduzionali quali la fosforilazione (aggiunta di gruppi fosfato) o la glicosilazione (aggiunta di carboidrati), che modificano le funzioni della molecola.

La qualità delle proteine prodotte in pianta, valutata in termini di folding, modificazioni post-traduzionali e attività biologica, è un parametro cruciale che richiede una particolare attenzione.

La sintesi proteica è altamente conservata tra piante e animali. Questo consente di produrre in pianta una proteina con sequenza amminoacidica identica alla controparte nativa. Tuttavia, esistono alcune importanti differenze relative alle modificazioni post-traduzionali, in particolare al processo di glicosilazione, consistente nell'aggiunta di catene glicaniche alle proteine.

Negli organismi eucarioti, più del 50% delle proteine sono glicosilate (Apweiler et al., 1999) e la loro funzione spesso dipende da particolari glicoforme che ne determinano emivita, targeting tissutale e attività biologica. Allo stesso modo, più di un terzo dei biofarmaci approvati sono glicoproteine (Walsh e Jeffens, 2006), la cui efficienza e attività è influenzata dalla presenza e composizione glicanica.

I glicani sono normalmente classificati in base alla natura del legame che li unisce alla proteina in due classi principali: O-glicani e N-glicani.

Gli O-glicani presentano un legame O-glicosidico tra N-acetilgalattosammina e il gruppo ossidrilico di un residuo di serina o di treonina (O-glicani). Generalmente costituiscono strutture biantennali non molto complesse; in presenza di cluster di O-glicani si generano molecole di tipo mucinico che sono frequentemente secrete o compongono la superficie cellulare.

Gli N-glicani sono caratterizzati da un legame N-glicosidico tra l'N-acetilglucosammina e l'azoto del gruppo ammidico di residui asparagina (N-

glicani) presenti nella proteina neosintetizzata. Nelle cellule animali e vegetali, la sequenza consenso per la N-glicosilazione è Asn-X-Ser/Thr, dove X rappresenta un qualsiasi amminoacido eccetto Pro e Asp.

La principale differenza tra proteine prodotte in pianta e in animali riguarda la sintesi delle catene glicaniche (Fig. 1.9). Come per gli N-glicani dei mammiferi, anche quelli vegetali possono essere distinti in due categorie principali: i glicani ad alto contenuto di mannosio, che presentano la formula generale $\text{Man}_{5-9}\text{GlcNAc}_2$ e una struttura relativamente semplice, e i glicani complessi che derivano dal rimodellamento dei glicani ad alto contenuto di mannosio, con taglio e aggiunta di diversi residui monosaccaridici (Rayon et al., 1998).

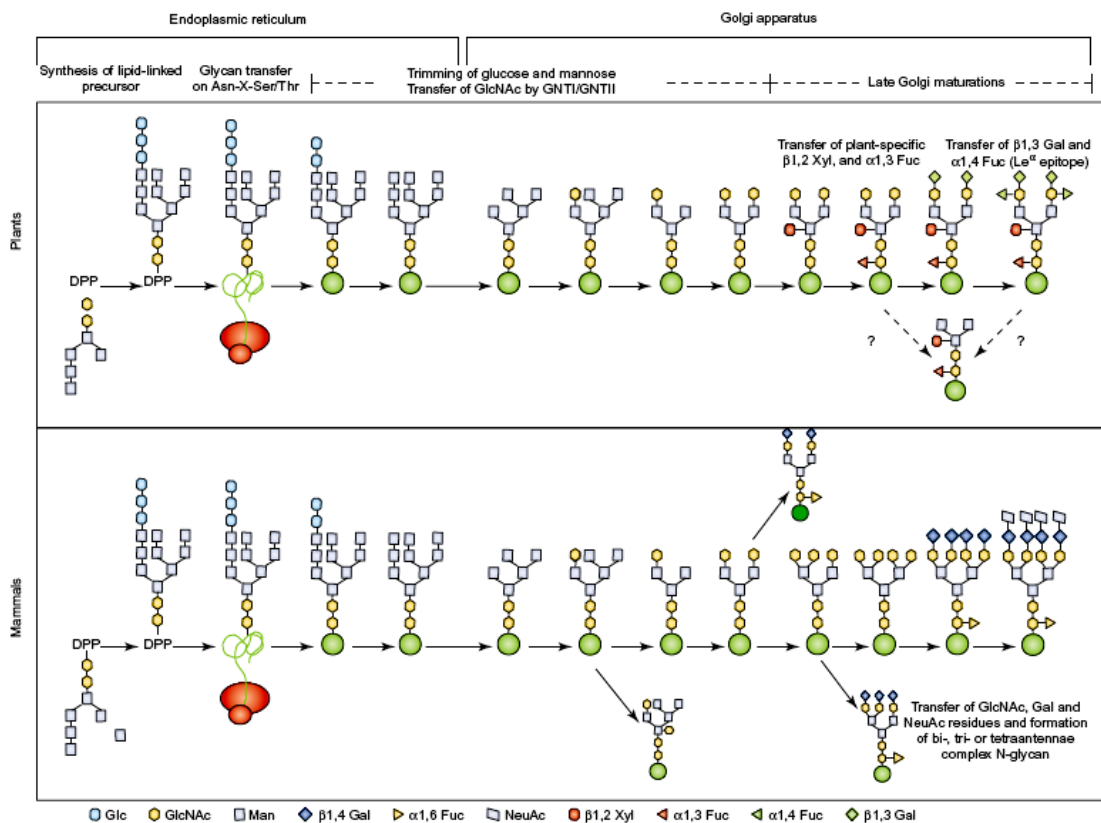


Figura 1.9 Processo di N-glicosilazione in piante e animali. Un precursore oligosaccaridico assemblato in carrier lipidico è trasferito su uno specifico residuo Asn presente nel polipeptide nascente. Successivamente gli N-glicani sono processati tramite rimozione dei glicosil e mannosil residui. Differenze nello sviluppo degli N-glicani complessi tra piante e animali hanno luogo durante gli eventi di maturazione nell'apparato di Golgi.

Le piante producono glicani di tipo complesso diversi da quelli animali (Fig. 1.10).

In particolare, i glicani complessi vegetali (Ma et al., 2003):

1. generalmente mancano di residui terminali di galattosio e acido sialico;
2. possono contenere residui di α -1,3-fucosio;
3. possono contenere residui di β -1,2-xilosio;
4. possiedono generalmente una struttura biantennale semplice.

Le modificazioni che gli N-glicani possono subire nell'apparato di Golgi differiscono, quindi, in piante e animali. Per questo, nella produzione di proteine eterologhe in pianta è necessario indagare gli effetti esercitati dalle modificazioni post-traduzionali.

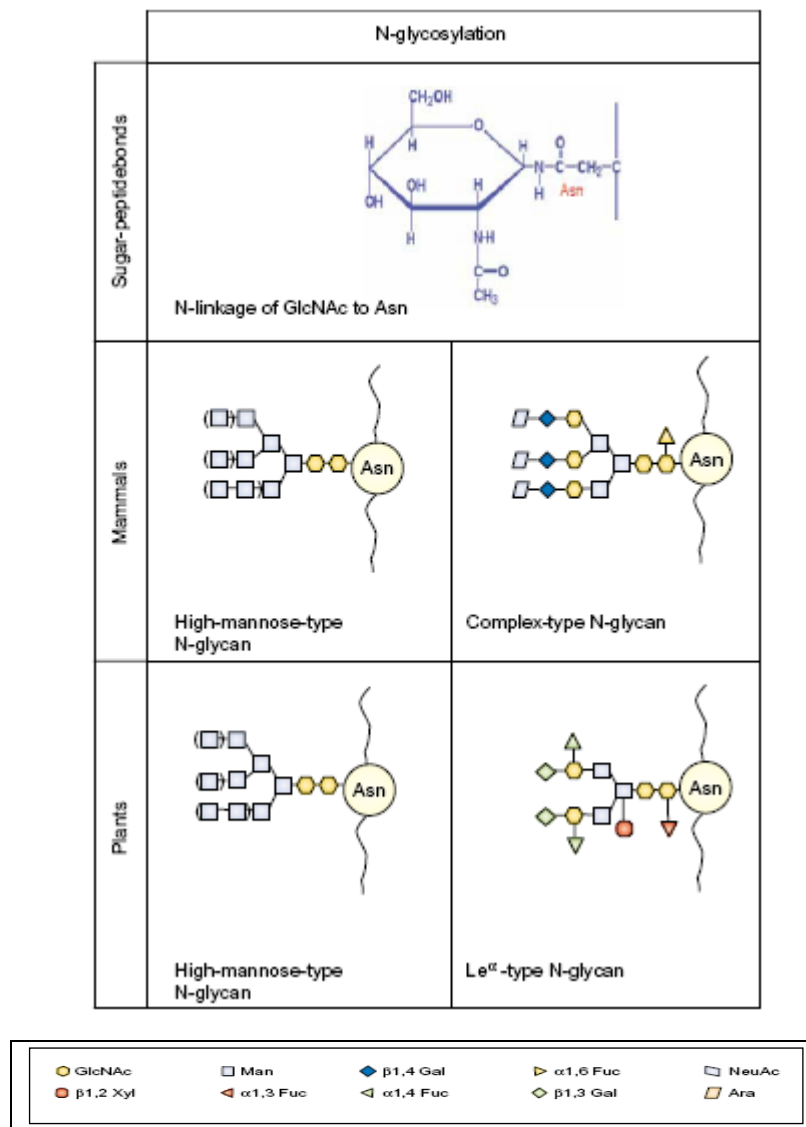


Figura 1.10 Tipiche strutture degli N-glicani vegetali e animali.

1.5.2 Strategie di “umanizzazione” delle proteine ricombinanti prodotte in pianta

Il processo di N-glicosilazione influenza significativamente proprietà quali il ripiegamento, la stabilità, la solubilità, la suscettibilità a proteasi, il livello di purezza e l'antigenicità delle proteine stesse. Conseguentemente, l'attività biologica di molte proteine terapeutiche dipende dal loro stato di glicosilazione. Ciò contribuisce a spiegare le ragioni per le quali i biofarmaci vengono prodotti in sistemi eterologhi di espressione potenzialmente in grado di attuare i processi di glicosilazione. Tale capacità costituisce, dunque, un requisito fondamentale per assicurare un'effettiva bioattività e farmacocinetica delle proteine ricombinanti in un contesto terapeutico e porta a sostenere l'uso di piante transgeniche a sfavore di sistemi di produzione fondati su batteri (es. *Escherichia coli*) e lieviti. Tuttavia, le piante sono incapaci di attuare allo stesso modo i processi di costruzione e rimodellamento degli N-glicani che avvengono negli organismi mammiferi. Per questo motivo, un'analisi dettagliata dei glicani è importante per le proteine farmaceutiche prodotte in piante transgeniche, soprattutto se si prevede una somministrazione per via parenterale o intravenosa. Poiché $\beta(1,2)$ -xilosio e $\alpha(1,3)$ -fucosio sono generalmente assenti nelle glicoproteine animali, essi possono rappresentare epitopi per anticorpi reattivi contro i carboidrati (Lerouge et al., 2000). E' stato dimostrato che residui di xilosio e fucosio fungono da epitopi nell'indurre una reazione allergica glicano-specifica e proteina-indipendente nell'uomo (Garcia-Casado et al., 1996) e che tali residui sono bersaglio di IgE nelle persone allergiche a pollini o alimenti vegetali (Fotisch e Vieths, 2001). Sebbene la presenza di anticorpi nel siero non implichi sempre un riscontro dannoso, la percezione di un effetto negativo derivante da strutture glicaniche “estrane” costituisce un aspetto importante nell'accettazione di biofarmaci prodotti in pianta destinati alla terapia umana. Al fine di superare tali problematiche, che rischiano di limitare l'impiego delle piante come fabbriche cellulari per la produzione di proteine ricombinanti, sono state sviluppate diverse strategie di “umanizzazione” delle glicoproteine prodotte in sistemi di espressione vegetali, che includono: 1. il targeting della proteina ricombinante nel reticolo endoplasmico (ER); 2. l'inattivazione genica degli enzimi endogeni con attività glicosiltransferasica attraverso tecniche di knock-out e silenziamento genico ; 3. l'espressione *de novo* di glicosiltransferasi eterologhe per l'aggiunta di residui

oligosaccaridici tipici dei glicani complessi animali (Saint-Jore-Dupas et al., 2007).

Una prima strategia per prevenire l'aggiunta di glicani immunogenici consiste nello stoccaggio delle proteine terapeutiche all'interno del RE attraverso segnali di ritenzione (Gomord et al., 1997), posti al terminale carbossilico della proteina (KDEL, HDEL, SEQDEL). Il RE è un compartimento localizzato a monte delle cisterne dell'apparato di Golgi dove ha luogo l'addizione dei glico-epitopi immunogenici agli N-glicani vegetali. Le proteine residenti nel lume del RE, pertanto, dovrebbero contenere esclusivamente strutture glicaniche ad alto mannosio strutturalmente simili tra piante e animali (Navazio et al., 1996; Pagny et al., 2000). La ritenzione nel RE, tuttavia, non è assoluta, in quanto esiste un passaggio continuo di proteine verso il Golgi ma anche dal Golgi verso il RE (Vitale e Denecke, 1999).

Una seconda strategia di ingegnerizzazione del processo di glicosilazione in pianta è basata sull'inibizione di glicosiltransferasi presenti nell'apparato di Golgi e responsabili della maturazione degli N-glicani. In particolare, gli enzimi $\alpha(1,3)$ -fucosiltransferasi e $\beta(1,2)$ -xilosiltransferasi, coinvolti, rispettivamente, nell'attacco di fucosio $\alpha(1,3)$ e xilosio $\beta(1,2)$ (Koprivova et al., 2004; Lienard et al., 2007) alla catena glicanica, rappresentano i target principali del silenziamento genico.

Recenti progressi nel campo della biologia molecolare hanno dimostrato che il silenziamento genico può essere indotto in maniera altamente specifica attraverso il meccanismo dell'RNA interferente (RNAi). Questo sistema si basa su un processo di inattivazione genica post-trascrizionale che prevede la generazione di piccoli frammenti di RNA di 21-22 ribonucleotidi, denominati *small interfering* RNA (siRNA), a partire da un dsRNA più lungo omologo al gene da sopprimere (Hamilton e Baulcombe, 1999; Zamore et al., 2000; Elbashir et al., 2001). Descritto inizialmente in pianta e in *Caenorhabditis elegans* (Fire et al., 1998), successivi studi hanno dimostrato che tale fenomeno è conservato in molti organismi eucarioti (Cogoni e Macino, 2000) dove riveste un ruolo importante in diversi processi biologici, quali la regolazione dello sviluppo, la formazione e il mantenimento dell'eterocromatina durante la mitosi e la meiosi, la difesa antivirale (Waterhouse et al., 2001). È stato, infatti, osservato che in natura la cellula necessita di questo tipo di soppressione post-trascrizionale per

contrastare eventuali infezioni causate da virus a dsRNA e da microorganismi patogeni in grado di produrre dsRNA durante il loro ciclo replicativo.

Nelle piante, un metodo di silenziamento genico particolarmente efficiente (Chuang e Meyerowitz, 2000) consiste nel trasformare la pianta con un plasmide contenente una sequenza che codifica per un shRNA (*small hairpin* RNA) (Fig. 1.11). In tale vettore il gene target è clonato come sequenza ripetuta invertita, separata da uno spacer e posta sotto il controllo di un promotore forte, quale il promotore 35SCaMV nelle specie dicotiledoni e quello dell'ubiquitina 1 di mais nelle monocotiledoni. Tali elementi così disposti formano, a trascrizione avvenuta, una struttura a forcina per l'appaiamento della sequenza senso e antisense. Altri vettori di espressione per shRNA codificano le sequenze senso e antisense sotto il controllo di promotori separati. Una volta transfettata, la cellula produrrà i due filamenti dell'RNA che si uniranno a formare il dsRNA.

In entrambi i casi, a seguito del taglio ad opera dell'enzima Dicer, si generano i frammenti siRNA che vengono reclutati nel complesso ribonucleoproteico (RISC) dove mediano la degradazione specifica dell'mRNA bersaglio.

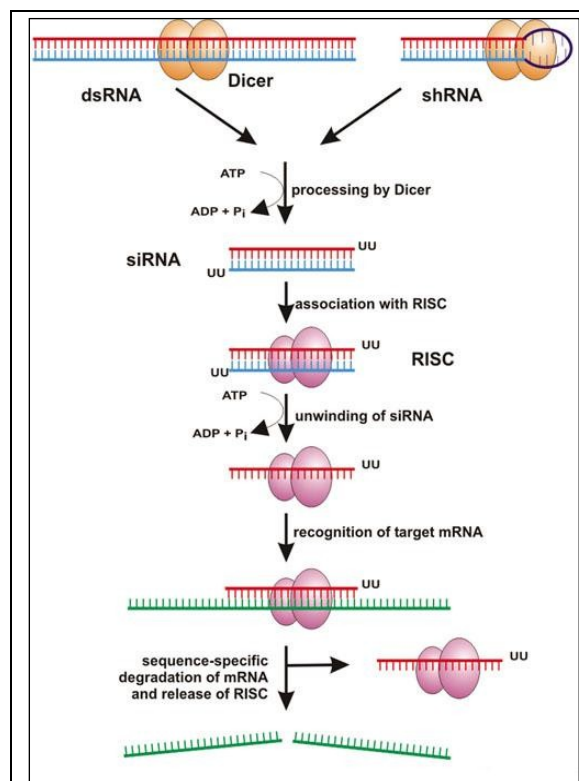


Figura 1.11 Rappresentazione schematica del meccanismo di silenziamento genico mediato da shRNA.

Nell'ambito del *molecular farming* in pianta, il silenziamento genico mediante RNA interferente si è dimostrato un efficace strumento per generare fenotipi quasi-null per $\alpha(1,3)$ -fucosiltransferasi e $\beta(1,2)$ -xilosiltransferasi in due diversi sistemi di espressione: nell'alga *Lemna minor* e in *Medicago sativa*. In particolare, questo sistema si è rivelato altamente efficiente nella produzione di un anticorpo monoclonale in *Lemna minor* (Cox et al., 2006). Tale molecola ha presentato un tipo di glicosilazione omogeneo e privo di catene glicaniche immunogeniche (Lienard et al., 2007).

Esperimenti di *knock out* sono stati condotti sulla specie di muschio *Physcomitrella patens*; tale organismo presenta un'elevata frequenza di ricombinazione omologa che consente di eseguire il *knock out* secondo modalità relativamente semplici. In *P. patens* il *knock out* dei geni codificanti l' $\alpha(1,3)$ -fucosiltransferasi e la $\beta(1,2)$ -xilosiltransferasi ha permesso di ottenere N-glicani privi dei glicoeptopi pianta-specifici senza interferire con la secrezione della proteina (Koprivova et al., 2004). Dal momento che in *P. patens* il processo di N-glicosilazione è simile a quello delle piante superiori, questi risultati hanno aperto la strada all'uso di tale tecnica applicata alla produzione di proteine ricombinanti di interesse farmaceutico.

Una terza strategia di modificazione del processo di glicosilazione in pianta volta all'ottenimento di pattern di glicosilazione compatibili con quello delle glicoproteine umane consiste nell'utilizzo degli enzimi esogeni $\beta(1,4)$ -galattosiltransferasi e sialiltransferasi per la modificazione in vitro delle proteine derivate da pianta (Blixt, 2002). I glicani vegetali mancano generalmente di residui terminali di galattosio e acido sialico presenti, invece, nelle glicoproteine vegetali. E' stato dimostrato che l'assenza di tali residui riduce l'efficienza clinica dei biofarmaci prodotti in pianta a causa di una rapida *clearance* attraverso la via dei recettori di asialoglicoproteine epatici galattosio-specifici (Stockert, 1995). In piante trasgeniche di tabacco ingegnerizzate per la produzione di un anticorpo (una proteina umana modello) e per l'espressione della $\beta(1,4)$ -galattosiltransferasi umana, il 30% degli N-glicani portati dall'anticorpo ha presentato sequenze con residui di N-acetilgalattosamina terminale tipici dei glicani animali.

1.5.3 Strategie di miglioramento dell'espressione di proteine ricombinanti in pianta

L'esito di una corretta trasformazione e l'ottenimento di elevati livelli di espressione della proteina ricombinante prodotta in pianta sono influenzati da una molteplicità di parametri; alcuni dei quali dipendono dalla sequenza codificante inserita, mentre altri ne sono indipendenti.

Di seguito vengono presi in considerazione alcuni fattori particolari quali l'adattamento codonico della sequenza codificante in funzione del *codon usage* della specie vegetale ospite e gli elementi di regolazione a livello trascrizionale (promotore e terminatore).

Codice genetico e ottimizzazione codonica

Gli organismi viventi hanno sviluppato un codice formato da 64 codoni che costituisce il sistema più efficiente possibile per codificare almeno 20 elementi diversi (in questo caso amminoacidi) utilizzando parole di 3 lettere (i codoni) formate usando un alfabeto di 4 diversi caratteri (le basi azotate).

Come conseguenza del soprannumero di triplette minime necessarie all'interno del codice, 18 di questi amminoacidi sono codificati da più codoni sinonimi e solamente gli amminoacidi metionina (Met) e triptofano (Trp) sono codificati da un unico codone. Inoltre, 3 diversi codoni sono utilizzati per codificare il messaggio di fine traduzione proteica (codoni stop: UAA, UAG, UGA).

Le triplette degli amminoacidi codificati da più di un codone presentano una radice codonica fissa formata dalle prime due basi, indicate con la dicitura N_1N_2 , e da una terza posizione variabile (*wobble*) indicata con N_3 . Solo per gli amminoacidi arginina (Arg), serina (Ser) e leucina (Leu), rappresentati da sei possibili codoni, esistono due diverse radici codoniche. La corrispondenza degli amminoacidi con i 64 codoni nucleotidici costituisce il codice genetico e poiché quasi tutti gli organismi viventi codificano con gli stessi codoni gli stessi amminoacidi, tale codice è definito universale (Fig. 12).

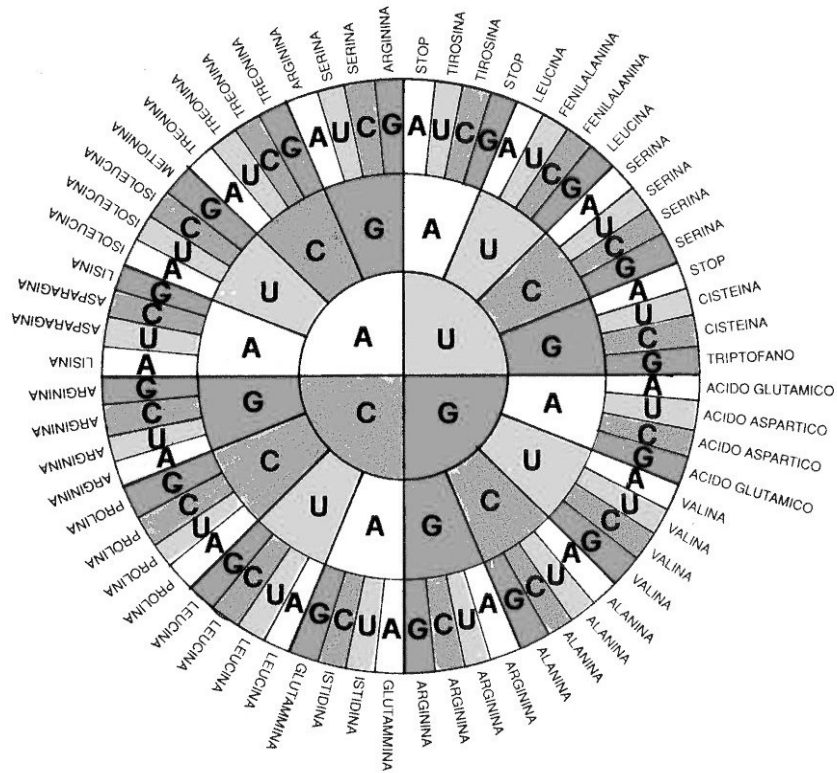


Figura 1.12 Codice genetico universale.

Data la natura degenerata del codice genetico universale, è teoricamente possibile (e sperimentalmente dimostrato) che la stessa proteina possa essere codificata utilizzando un diverso assortimento di codoni sinonimi.

A tale proposito va rilevato che le analisi di sequenza del DNA hanno dimostrato che gli organismi viventi non utilizzano indifferentemente i codoni sinonimi ma che, al contrario, esiste una scelta selettiva, non casuale dei codoni stessi. Tale scelta ha una forte valenza biologica: evidenze sperimentali hanno dimostrato che le frequenze dei codoni sinonimi utilizzati variano in modo netto e sorprendente nei vari organismi contribuendo assieme ad altre proprietà alla cosiddetta firma genomica di una specie. La scelta codonica specie-specifica è stata chiamata "dialetto codonico" (Ikemura, 1985).

Sono state osservate variazioni del *codon usage* anche internamente alla stessa specie e al medesimo individuo; in effetti le frequenze di utilizzazione dei codoni sinonimi risultano variare in relazione al gene considerato (Ikemura, 1985).

In alcuni organismi è stata dimostrata una correlazione positiva tra il livello di espressione genica e la scelta di codoni sinonimi (Grantham et al., 1981; Powell e Moriyama, 1997; Duret e Mouchiroud, 1999; Kliman et al., 2003) e ciò è stato spiegato ipotizzando un meccanismo di selezione traduzionale.

Secondo questa ipotesi, ai codoni più frequenti nel genoma corrisponderebbero di norma gli isoaccettori più abbondanti a livello cellulare e di conseguenza ciò implicherebbe una maggiore efficienza traduzionale.

Un ulteriore elemento da tenere in considerazione riguarda la composizione generale del DNA nei diversi genomi. Infatti, non tutte le specie utilizzano le stesse percentuali delle coppie di basi azotate puriniche (G, A) e pirimidiniche (C, T), dando luogo a “contesti” generali per le sequenze codificanti molto diversi. Inoltre, dalle analisi effettuate sulle sequenze di DNA genomiche sono emerse significative differenze tra le diverse specie per quanto riguarda la frequenza dei vari elementi dinucleotidici. Similmente a quanto osservato per l'utilizzo dei codoni, anche per gli elementi dinucleotidici è possibile identificare un insieme di frequenze che risultano specie-specifiche, non casuali e altamente stabili (Russel et al., 1976; Viotti et al., 1978).

Sulla base di queste considerazioni, disponendo di dati corretti sull'effettivo utilizzo dei codoni sinonimi nei diversi organismi è teoricamente possibile adattare il *codon usage* di una sequenza eterologa al genoma dell'organismo ospite. Per migliorare i livelli di espressione e massimizzare la produzione di una proteina ricombinante nel sistema di espressione vegetale scelto, può essere, quindi, conveniente riadattare la sua sequenza codificante in funzione del dialetto codonico della pianta ospite, attuando una resintesi del gene che si vuole esprimere sulla base delle frequenze codoniche e intercodoniche ed eliminando tutte le eventuali sequenze di disturbo per l'espressione.

Elementi di regolazione a livello tradizionale: il promotore e il terminatore

Al fine di ottenere elevati livelli di trascrizione, i due elementi più importanti da considerare sono il promotore e il sito di poliadenilazione (Ma et al., 2003).

I segnali di poliadenilazione più diffusi sono quelli relativi ai geni 35S di CaMV (cauliflower mosaic virus), *nos* di Agrobacterio e *ssu* di pisello (Ma et al., 2003). A livello di promotore, la scelta è senz'altro più ampia e diversificata, e per un'espressione costitutiva a livelli medio-alti si ricade su promotori virali (es. 35S

di CaMV) o di proteine ubiquitarie (es. l'ubiquitina-1 di mais) (Sanger et al., 1990).

Livelli di espressione troppo elevati possono, tuttavia, causare un ridotto vigore della pianta (nel caso in cui buona parte delle risorse energetiche siano impiegate nella produzione di proteine eterologhe non essenziali) o anche mettere in pericolo la sua stessa sopravvivenza. Per questo motivo, i promotori tessuto-specifici, che regolano l'espressione del transgene esclusivamente in un determinato distretto, sono particolarmente utili poiché consentono la produzione di proteine eterologhe altrimenti fitotossiche e facilitano le successive fasi di purificazione. In particolare, l'espressione di una proteina ricombinante nel seme può efficacemente prevenire effetti tossici alla pianta ospite (Chinn e Comai, 1996). Tra i promotori seme-specifici più utilizzati troviamo quelli delle proteine di riserva del seme di riso, tra cui i promotori delle gluteline 1 e 4, i promotori delle prolamine da 10 e 16 KDa e il promotore della globulina-1 (Qu e Takaiwa, 2004).

1.6 GAA e GCasi in pianta

I bioreattori vegetali sono particolarmente interessanti per la produzione degli enzimi lisosomiali oggetto di studio: una serie di aspetti positivi, infatti, giocano a favore di tale sistema. Anzitutto, diversamente dai metodi basati sui batteri, le piante sono in grado di attuare le modificazioni post-traduzionali della proteina senza indurre, però, l'iperglicosilazione che si osserva con l'impiego di funghi. Inoltre, viene superato il rischio di contaminazione da patogeni umani o animali, concreto, invece, in cellule ingegnerizzate di mammifero, poiché le piante non rappresentano un loro ospite.

Da un punto di vista strettamente economico, i sistemi vegetali permettono poi una riduzione dei costi di produzione su larga scala (Fig. 1.13), con possibile abbassamento dei prezzi della molecola farmaceutica ottenuta.

Tra i sistemi ospite vegetali, l'endosperma di riso offre alcuni importanti vantaggi per l'espressione di GAA; questo tessuto consente l'accumulo stabile della suddetta proteina per lunghi periodi di tempo senza compromissione della sua attività biologica (Stoger et al., 2000). L'enzima, infatti, in tale ambiente è protetto dalla degradazione proteolitica a motivo della sintesi di quantità significative di inibitori di proteasi. A ciò si aggiunge il contenuto relativamente basso di

composti fenolici e lipidi che permette di ottenere una migliore efficienza nei successivi processi di estrazione e purificazione della proteina dal tessuto vegetale. Inoltre, le piccole dimensioni del seme consentono il raggiungimento di un'alta concentrazione della molecola ricombinante, aspetto di particolare interesse per la produzione industriale. Infine, la proteina espressa in seme non manifesta caratteristiche di fitopatogenicità e non interferisce, quindi, con lo sviluppo e la crescita della pianta.

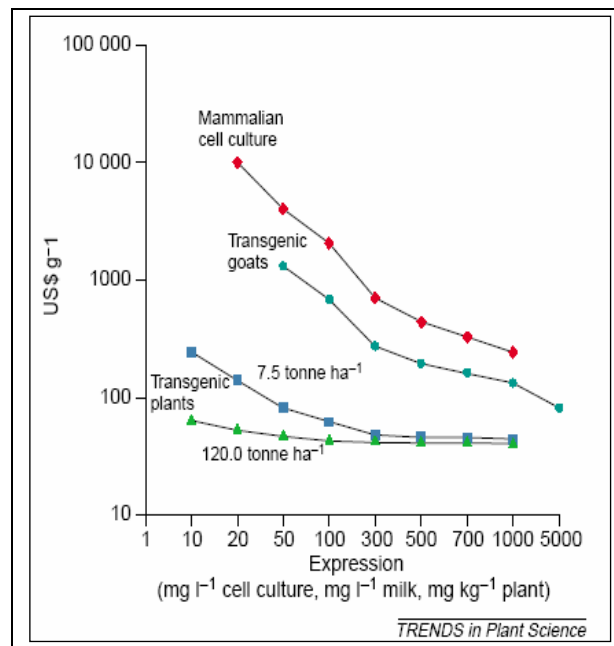


Figura 1.13 Costi per grammo dell'immunoglobulina A purificata prodotta in diversi sistemi di espressione. I costi delle colture cellulari mammifere sono derivati dai costi industriali delle stesse e delle strutture di purificazione. I costi in capre transgeniche sono stime disponibili pubblicamente fornite da Genzyme Transgenics (Farmingham, MA, USA). I costi in pianta confrontano la produzione in biomasse verdi (120.00 tonnellate ha⁻¹) e in seme (7.5 tonnellate ha⁻¹). Le differenze nei diversi sistemi di produzione sono basati principalmente sui costi di produzione, assumendo che i costi di purificazione e le perdite durante il processo di purificazione siano gli stessi per tutti i sistemi.

1.7 Il riso come specie modello

Il riso (*Oryza sativa* L.) è uno dei cereali più coltivati al mondo (secondo solo al frumento) e il principale alimento per circa il 60% della popolazione umana, soprattutto nei paesi asiatici dove fornisce i tre quarti dell'apporto calorico

giornaliero a una popolazione in crescita a un tasso annuale dell'1.8% (Khush, 1997).

Negli ultimi anni si è assistito a un maggiore impiego delle proteine di riso come potenziali ingredienti alimentari dato il loro alto valore nutrizionale, le proprietà ipoallergeniche e il sapore delicato (Chrastil, 1992).

Il riso è considerato specie modello per le piante appartenenti alla classe delle monocotiledoni grazie a una serie di caratteristiche che possono essere così esemplificate:

- genoma completamente sequenziato, di dimensioni minori (431 Mb) rispetto alle altre graminacee;
- facilità di trasformazione e rigenerazione;
- tecniche colturali semplici e bassa richiesta di manodopera dato l'elevato grado di meccanizzazione;
- brevità del ciclo biologico (da 140 a 170 gg in funzione del tipo di cultivar);
- elevate rese per ettaro di superficie coltivata;
- sistema di unioni fondato sull'autogamia e sulla cleistogamia, caratteristica importante ai fini della biosicurezza di una PGM;
- possibilità di utilizzo di varietà inadatte al consumo (no-food) e destinate a scopi diversi rispetto all'alimentazione umana (es. estrazione di amido);
- presenza di marcatori genetici legati alle caratteristiche del seme (colore, forma, indice di collosità e consistenza) (Yuan et al., 2001; Nature, 2005) al fine di evitare l'ingresso nella filiera alimentare di sementi prodotte da piante transgeniche.

Nel presente lavoro, per la manipolazione genetica, è stata utilizzata una varietà *waxy* di riso (CR W3), con endosperma glutinoso, costituita dall'Ente Nazionale Risi per scopi industriali.

I semi di riso, come tutti quelli dei cereali, presentano un tessuto endospermico preminente, che può occupare fino al 90% del seme, e un embrione di dimensioni ridotte (Yang et al., 2002). Tutti i tessuti del seme svolgono una funzione di riserva (contengono minerali, oli, proteine e amido). I semi di cereali contengono fino al 16% in peso secco di proteine. Le proteine di riserva del seme vengono accumulate ad alti livelli in tessuti specifici come l'endosperma e i cotiledoni.

L'endosperma di riso contiene quattro classi principali di proteine distinte in base al loro grado di solubilità ed estraibilità in acqua (albumine), soluzioni saline diluite (globuline), miscele alcol/acqua (prolamine) e acidi o basi diluite (gluteline) (Kawagoe et al., 2005).

Sia le prolamine, sia le gluteline vengono inizialmente sintetizzate sulle membrane del RE, ma sono poi immagazzinate sotto forma di corpi proteici in diversi compartimenti intracellulari (Muench et al., 1998). Le prolamine vengono accumulate e assemblate in granuli d'inclusione intracisternale di forma sferica, di 1-2 μm di diametro, all'interno del lume del RE (Hamada et al., 2003). Nel riso, i corpi proteici di prolamine vengono estrusi dal RE e rimangono separati dal compartimento vacuolare. È stato dimostrato che il chaperone BiP (binding protein), appartenente alla famiglia delle heat shock proteins 70 (hsp70), è strettamente associato alle prolamine di riso, suggerendo che BiP promuova un folding corretto delle prolamine e il loro assemblaggio sulla superficie di corpi proteici crescenti (Li et al., 1993).

Le gluteline vengono inizialmente indirizzate al lume del RE attraverso un processo di immissione co-traduzionale; vengono quindi trasportate al Golgi e successivamente a piccoli vacuoli di immagazzinamento a formare corpi proteici elettro-densi, di forma irregolare e diametro di 2-3 μm (Hamada et al., 2003).

Le proteine di riserva sono così rappresentate nell'endosperma di riso (Van der Borgh et al., 2006): 3.8-8.8% di albumine; 9.6-10.8% di globuline; 2.6-3.3% di prolamine; 66-78% di gluteline.

1.8 Il metodo di trasformazione con *Agrobacterium tumefaciens*

Fino a 30 anni fa, la possibilità di utilizzare il comune batterio del suolo, *Agrobacterium tumefaciens*, come vettore per la trasformazione genetica controllata delle piante era solo una prospettiva lontana. Oggi, invece, molte specie agronomicamente importanti sono trasformate routinariamente con questo batterio e la lista delle piante suscettibili alla trasformazione, mediata da agrobatterio, è destinata ad aumentare. Tuttavia, la ricerca deve ancora risolvere alcuni interrogativi relativi alla trasformazione genotipo-indipendente di specie agrarie economicamente importanti e alla stabilità di espressione del transgene.

Agrobacterium tumefaciens è un batterio Gram-negativo del suolo di forma bastoncellare fitopatogeno ma non invasivo, noto per indurre tumori al colletto che interferiscono con la normale crescita delle piante infette (Fig. 1.14). In natura, *Agrobacterium tumefaciens* infetta un'ampia gamma di specie vegetali, soprattutto dicotiledoni. A seconda della specie vegetale considerata, si identificano ceppi a bassa oncogenicità e ceppi altamente oncogenici.



Figura 1.14 Tipico callo generato da Agrobatterio.

Al genere *Agrobacterium* sono ascrivibili cinque specie distinte in base al carattere fitopatogeno: *Agrobacterium tumefaciens*, agente del tumore al colletto; *Agrobacterium rhizogenes*, causa di piccoli tumori a livello radicale con iperramificazioni; *Agrobacterium rubi* che colpisce le parti aeree, *Agrobacterium radiobacter* considerata una specie avirulenta (Otten et al., 1984) e *Agrobacterium vitis*. In realtà, è dimostrato che i sintomi per la maggior parte seguono il tipo di plasmide tumorigenico presente all'interno di un particolare ceppo. Infatti, il tipo di malattia scatenata è determinata non tanto da fattori ereditabili del batterio, quanto piuttosto da fattori portati sul plasmide. Il tumore al colletto causato da *Agrobacterium tumefaciens* si sviluppa proprio in seguito al trasferimento e alla successiva integrazione ed espressione di un particolare segmento di DNA plasmidico, detto T-DNA (DNA trasferito), all'interno di uno o più cromosomi della pianta (Glick e Pasternack, 1999). Successivamente all'introduzione del gene di interesse nel T-DNA, è possibile rigenerare da una singola cellula piante transgeniche caratterizzate da un'espressione stabile del gene introdotto e da un background genetico identico a quello della pianta

madre. Alla trasformazione segue la selezione delle piante che presentano il gene d'interesse mediante un gene di resistenza selezionabile introdotto nel vettore di espressione o in un secondo vettore nel caso si operi una co-trasformazione. La possibilità di ottenere piante transgeniche trasformate in modo stabile dipende dalla varietà e il tempo richiesto per saggiare le proteine espresse in modo eterologo può variare da 3 a 9 mesi.

Agrobacterium tumefaciens manifesta la capacità di trasformare geneticamente le piante che presentano ferite rilascianti sostanze come acetosiringone e idrossiacetosiringone, in grado di richiamare il batterio e di indurre l'attività dei geni *vir*, rendendo le cellule ferite suscettibili all'attacco di *A. tumefaciens*.

La trasformazione genetica, che si manifesterà come formazione del tumore al colletto, è il risultato dell'integrazione del segmento T del DNA plasmidico dell'agrobatterio all'interno del genoma della pianta e la conseguente espressione dei geni interni a tale segmento (Stachel et al., 1986) (Fig. 1.15). L'integrazione può aver luogo in uno o più punti di uno stesso cromosoma o di cromosomi diversi.

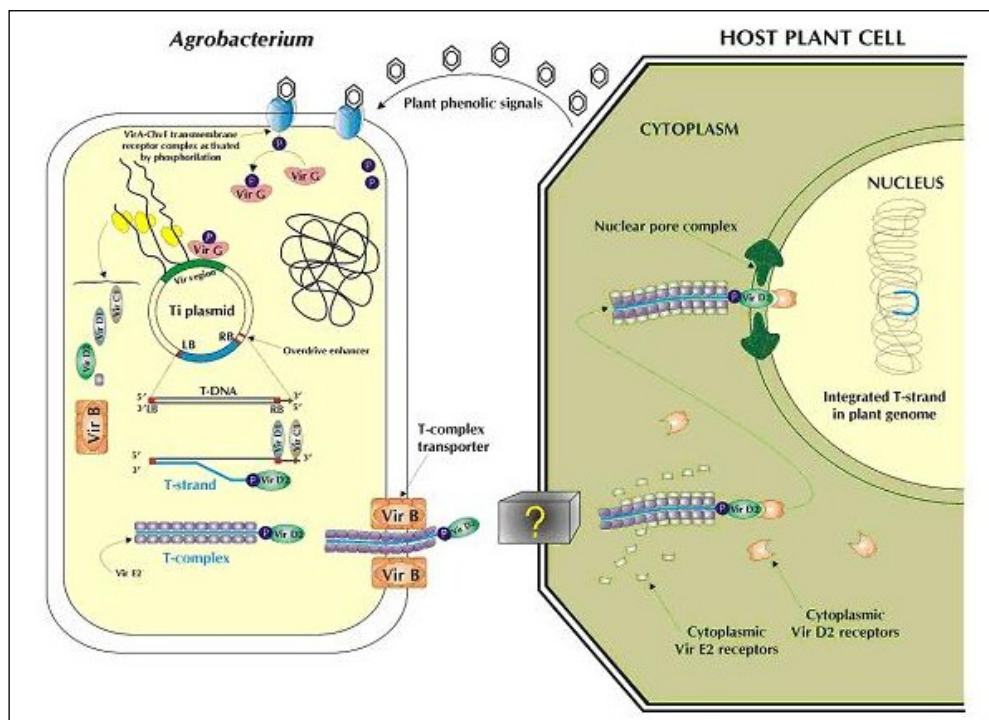


Figura 1.15 Meccanismo di infezione di *Agrobacterium tumefaciens*.

1.8.1 Il plasmide pTi

Le basi molecolari della trasformazione genetica mediata da *Agrobacterium tumefaciens* consistono nel trasferimento dal batterio e successiva integrazione nel genoma nucleare della pianta di una porzione del plasmide pTi (tumor-inducing, nome derivato dai sintomi che induce nelle piante infettate) residente nel batterio.

pTi è un plasmide a basso numero di copie di dimensioni comprese tra 200 e 800 kb (Goodner et al., 2001). In pTi si individuano due regioni principali: la regione del T-DNA (DNA di trasferimento) e la regione Vir (Fig. 1.16).

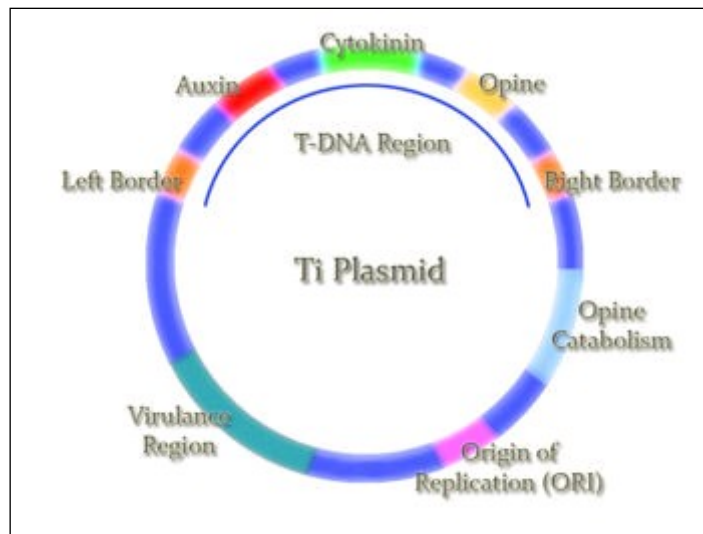


Figura 1.16 Rappresentazione schematica del plasmide pTi.

Il T-DNA è lungo approssimativamente da 10 a 30 kb, rappresentando così meno del 10% del plasmide. E' fiancheggiato da sequenze ripetute di 25 bp altamente omologhe (left e right border repeat), essenziali per l'integrazione nel genoma delle piante (Hansen et al., 1999). La regione del T-DNA porta oncogeni, responsabili della iperproduzione di auxine e citochine, e geni per la sintesi di opine. Auxine e citochinine sono fitoormoni che regolano la crescita delle cellule vegetali e lo sviluppo della pianta. Gli oncogeni possono sintetizzare direttamente tali ormoni, oppure aumentare la sensibilità delle cellule vegetali agli stessi, determinando una crescita cellulare rapida e disorganizzata e l'insorgenza di tumori, quali quelli al colletto (Glick e Pasternack, 1999). Le opine rappresentano una classe nuova di amminoacidi modificati estranei alle piante e derivati dalla condensazione tra un amminoacido e un chetoacido oppure tra un

amminoacido e uno zucchero. Il tipo di opina dipende dal ceppo di *Agrobacterium tumefaciens* utilizzato per la trasformazione. Si distinguono, così, ceppi che producono mannopina (ceppi mannopinici), nopalina (ceppi nopalinici), opalina (ceppi opalinici) e octopina (ceppi octopinici). Le opine sintetizzate costituiscono un'importante fonte di azoto organico utilizzabile da *Agrobacterium tumefaciens*.

Nella regione Vir (virulenza) di 35 Kb, posta al di fuori del T-DNA, mappano circa 10 geni Vir, codificanti proteine che esplicano un ruolo essenziale nel trasferimento e integrazione della porzione T-DNA nel genoma della pianta.

L'applicabilità della trasformazione mediata da Agrobatterio, dapprima limitata al tabacco e a poche altre specie, bersagli naturali dell'infezione, è ora ampliata alla maggior parte delle specie vegetali di interesse agronomico, comprese graminacee e leguminose (Lee et al., 2001).

La selezione delle cellule geneticamente trasformate da cui rigenerare le piante transgeniche prevede spesso l'impiego di substrati selettivi contenenti un antibiotico che uccide le cellule non trasformate in quanto prive del gene marker per la resistenza all'antibiotico stesso; tale gene è contenuto nel T-DNA ed è generalmente una componente associata al gene di interesse.

La presenza di geni marcatori nelle linee di piante transgeniche può costituire una fonte di potenziali rischi per la salute umana e per l'ambiente; i prodotti di alcuni geni marker, infatti, possono risultare tossici o allergenici per l'uomo o per gli animali, mentre i geni per la resistenza agli antibiotici potrebbero essere trasferiti orizzontalmente ai batteri presenti nell'ambiente, sebbene non esista alcuna evidenza scientifica che avvalorino tali ipotesi. Ciò nonostante la normativa europea vigente raccomanda, ove tecnicamente possibile, di eliminare le resistenze agli antibiotici entro il 31 dicembre 2008 (cfr. § 1.9). Per questo motivo sono stati proposti diversi metodi basati sull'eliminazione del marcatore selezionabile dal trasformato primario e finalizzati alla produzione di varietà transgeniche marker-free. Uno di questi consiste nella co-trasformazione della pianta con marcatori selezionabili e geni di interesse posti ciascuno in una regione T-DNA presente in un differente plasmide di *A. tumefaciens*. Si ottengono così linee transgeniche in cui il gene marker è integrato in un locus cromosomico diverso da quello in cui si trova il gene di cui si desidera

l'espressione e, sottoponendo queste linee a cicli di autofecondazione, si possono selezionare le piante transgeniche prive del marcatore.

Vi sono diversi approcci a questo tipo di metodica: possono essere impiegati due ceppi di *A. tumefaciens* ospitanti ciascuno un diverso plasmide (Depicker et al., 1985; McKnight et al., 1987), oppure entrambi i plasmidi possono essere portati dallo stesso ceppo contenente, quindi, due vettori binari separati (Daley et al., 1998).

Un miglioramento significativo nelle procedure di co-trasformazione è stato introdotto con l'uso di plasmidi binari contenenti due T-DNA (Hiei et al., 1994; Komari et al., 1996), detti anche "vettori superbinari", nei quali un T-DNA contiene il gene di interesse, mentre il secondo contiene un gene marcatore selezionabile. È stato dimostrato che questo approccio porta a frequenze di co-trasformazione più alte rispetto ai metodi precedentemente descritti (Komari et al., 1996).

Più di recente si sta considerando la possibilità di usare elementi trasponibili per costruire piante transgeniche prive di marcatori senza ricorrere necessariamente alla separazione del gene marker in anafase meiotica.

Un approccio alternativo all'impiego di geni per la resistenza ad antibiotici è rappresentato da sistemi di selezione positiva basati sull'aggiunta di zuccheri che non possono essere metabolizzati dalle cellule non trasformate. Tra i metodi sviluppati una particolare attenzione è stata rivolta all'utilizzo del mannosio come agente selettivo e del gene PMI di *Escherichia coli* codificante l'enzima fosfomannosio isomerasi, come marker selezionabile (Miles e Guest, 1984). Tale strategia si basa sull'osservazione che il mannosio, convertito in mannosio-6-fosfato dall'azione di esochinasi endogene, non può essere utilizzato dalle piante come fonte di nutrimento per la mancanza dell'enzima fosfomannosio isomerasi necessario per la sua conversione a fruttosio-6-fosfato. L'accumulo del mannosio fosforilato nelle cellule inibisce l'attività della 6-fosfo-glucosio isomerasi con conseguente blocco della glicolisi (Goldsworthy e Street, 1965). Inoltre, si osserva il concomitante consumo dei pools intracellulari di fosfato e ATP che si traduce in una inibizione severa della crescita (Ferguson et al., 1958; Malca et al., 1967). Quando le cellule sono trasformate con il gene PMI, acquisiscono la capacità di convertire il mannosio-6-fosfato a fruttosio-6-fosfato che può essere facilmente metabolizzato mediante gluconeogenesi e utilizzato come fonte di

carbonio, ottenendo un vantaggio metabolico rispetto alle cellule non trasformate.

1.9 La normativa sugli OGM

Il contesto normativo che a livello europeo disciplina il rilascio nell'ambiente a scopo di ricerca o finalizzata alla commercializzazione di un OGM si basa sui principi di precauzione e prevenzione. È costituito dalla Direttiva 2001/18/CE che, abrogando la precedente direttiva 90/220/CE, regola l'emissione deliberata nell'ambiente di un OGM, definito dalla normativa stessa come *“un organismo, diverso da un essere umano, il cui materiale genetico è stato modificato in modo diverso da quanto si verifica in natura mediante accoppiamento o incrocio con la ricombinazione genetica naturale”*.

Tale direttiva è stata recepita dagli Stati membri dell'Unione e integrata nella legislazione italiana con il Decreto 8 luglio 2003, n. 224. La legge riscrive le regole base per l'approvazione di un OGM a fini di ricerca o di commercializzazione e all'articolo 4 definisce gli obblighi per gli stati membri:

1. *“Gli Stati membri, nel rispetto del principio precauzionale, provvedono affinché siano adottate tutte le misure atte a evitare effetti negativi sulla salute umana e sull'ambiente che potrebbero derivare dall'emissione deliberata o dall'immissione in commercio di OGM (...)”*.
2. *“prima di presentare una notifica (...), i notificanti effettuano una valutazione del rischio ambientale (...). Gli Stati membri e la Commissione si assicurano che gli OGM che contengono geni che esprimono resistenza agli antibiotici utilizzati per trattamenti medici o veterinari siano presi in particolare considerazione, al momento della valutazione del rischio ambientale, per individuare ed eliminare gradualmente negli OGM i marcatori di resistenza agli antibiotici che possono avere effetti negativi sulla salute umana e sull'ambiente. Questa eliminazione graduale avverrà entro il 31 dicembre 2004 per gli OGM immessi in commercio ed entro il 31 dicembre 2008 per gli OGM autorizzati a norma della parte B”*.
3. *“Gli Stati membri designano la o le autorità competenti responsabili dell'attuazione delle prescrizioni della presente Direttiva”*.

La Direttiva stabilisce una serie di valutazioni per ottenere il via libera alla sperimentazione, tra cui:

- ✓ descrizione del/i tratto/i e delle caratteristiche introdotte o modificate;
- ✓ eventuali modifiche della capacità della pianta geneticamente modificata (PGM) di trasferire materiale genetico ad altri organismi;
- ✓ informazioni su eventuali effetti tossici, allergenici o altri effetti nocivi per la salute umana riconducibili alla modificazione genetica;
- ✓ informazioni sulla sicurezza della PGM per la salute animale, con particolare riguardo a eventuali effetti tossici, allergenici o altri effetti nocivi riconducibili alla modificazione genetica se si intende impiegare la PGM negli alimenti per animali;
- ✓ meccanismi di interazione tra la PGM e gli organismi bersaglio (se del caso);
- ✓ potenziali cambiamenti nelle interazioni della PGM con organismi non bersaglio risultanti dalla modificazione genetica;
- ✓ potenziali interazioni con l'ambiente abiotico.

In Italia, il Decreto di recepimento dell'8 luglio 2003, in aggiunta alle norme già previste negli allegati alla Direttiva europea, prevede la valutazione preventiva di:

- abbandono o sostituzione di colture divenute, in seguito all'impatto dell'OGM, non più adatte o economicamente non più convenienti, con particolare riguardo alle varietà locali;
- danni all'immagine dei prodotti locali e/o della zona di emissione e costi da sostenere per difenderla;
- cambiamento dei percorsi commerciali per i prodotti provenienti dalla zona di emissione dovuti a impossibilità di accesso alla vendita OGM FREE (grande distribuzione, alimenti per lattanti e bambini) o ad altre tipologie commerciali;
- modificazioni del paesaggio con impatto negativo sull'attività agrituristica;
- abbandono e/o marginalizzazione della zona di emissione in seguito alla compromissione di forme di agricoltura praticate nella zona divenute meno redditizie in seguito all'impatto dell'OGM.

L'art. 8 stabilisce che chiunque intenda effettuare una emissione deliberata nell'ambiente di un OGM sia tenuto a presentare una notifica al Ministero

dell'Ambiente che si avvale della Commissione Interministeriale di Coordinamento per le Biotecnologie. La notifica deve contenere una serie di documenti allegati, comprendente un fascicolo tecnico con tutte le informazioni relative alla valutazione dei rischi, immediati e futuri, per la salute umana; la valutazione del rischio ambientale; la valutazione del rischio per l'agrodiversità, i sistemi agrari e la filiera agro-alimentare. Concluso l'iter amministrativo previsto per la fase di ricerca e sviluppo, chi abbia intenzione di effettuare una immissione sul mercato di OGM, come tali o contenuti in prodotti, deve presentare nuova notifica preventiva all'autorità nazionale competente che svolgerà l'istruttoria preliminare. La notifica deve contenere una serie di indicazioni relative alla valutazione dei rischi ambientali, alle condizioni di immissione sul mercato, nonché alle opportune misure di sicurezza e di intervento in caso di emergenza. La procedura autorizzativa segue poi una seconda fase che coinvolge UE e gli Stati membri.

Particolare attenzione è posta nei riguardi delle resistenze agli antibiotici che dovranno essere eliminate entro il 31 dicembre 2008. Questo obbligo impone, quindi, al ricercatore la messa a punto di nuovi protocolli per la trasformazione e la selezione degli OGM. Infatti, negli attuali sistemi di trasformazione, la selezione delle cellule trasformate portanti il gene d'interesse si ottiene coltivando le cellule su un substrato selettivo spesso caratterizzato dalla presenza dell'antibiotico che uccide le cellule sensibili (Endo et al., 2002). Proprio questa procedura ha richiamato l'attenzione del Legislatore per la preoccupazione che il gene di resistenza possa trasferirsi verticalmente dalla pianta OGM a batteri potenzialmente patogeni per l'uomo.

Si ritiene importante sottolineare come il divieto imposto dalla normativa relativo all'uso di geni di resistenza come marcatori selezionabili sia fondato sul principio di precauzione e non su quello di prevenzione: ciò implica la possibilità di intervento da parte del legislatore in ambiti in cui i rischi siano anche solo ipotizzati. Il principio di prevenzione, invece, si applica solamente a situazioni in cui siano stati individuati e verificati rischi certi per l'ambiente e la salute dell'uomo. Nel caso specifico non è mai stata dimostrata non solo la reale pericolosità ma nemmeno il potenziale rischio dell'impiego degli antibiotici come geni marker. È dunque opportuno evidenziare come ad oggi non esista alcuna certezza scientifica relativa alla loro dannosità e che la normativa, seppur certo

da rispettare, sia una manifestazione chiara di un atteggiamento aprioristico di diffidenza e chiusura verso tutto ciò che concerne gli OGM e la loro produzione, non sostenuto dall'evidenza scientifica.

1.10 Scopo della tesi

Nei ultimi 15 anni si è assistito alla produzione, sperimentazione e successiva commercializzazione di una serie di proteine enzimatiche curative per un gruppo di patologie lisosomiali (malattia di Gaucher, mucopolissacaridosi I, II, VI, Fabry, e malattia di Pompe). La tecnologia utilizzata in tutti questi casi è basata sull'utilizzo del DNA ricombinante in cellule eucariotiche umane (linee fibroblastoidi) o murine (CHO) (tecnologia utilizzata anche per produrre la GAA ricombinante umana) che necessitano di bioreattori meccanici ad elevato volume. Il sistema utilizzato presenta i difetti di una produzione limitata e di elevati costi di investimento. Negli ultimi anni, è stato dimostrato che le piante transgeniche rappresentano sistemi competitivi per la produzione di proteine farmaceutiche. Riuscire ad esprimere una proteina ricombinante con potenziale terapeutico in pianta potrebbe riflettersi nello sviluppo di un farmaco più accessibile ai pazienti colpiti da tali malattie. Ciò risulta particolarmente importante nel caso del trattamento di patologie rare come la malattia di Gaucher e la glicogenosi di tipo II. Alla luce di queste considerazioni lo scopo principale del presente dottorato è stato il miglioramento dell'espressione di α - e β -glucosidasi acida umana ricombinante, esplorando il potenziale di bioproduzione delle piante. Il riso (*Oryza sativa* L. ssp. *japonica*, var. CR W3) è stato scelto come specie ospite e il suo seme come organo d'accumulo delle proteine lisosomiali di interesse. Per la realizzazione di questo scopo, il lavoro è consistito di diverse fasi: 1) design genico in silico finalizzato all'espressione endosperma-specifica e valutazione dell'effetto della riscrittura genica sui livelli di espressione della β -glucosidasi acida; 2) determinazione del contenuto dell' α -glucosidasi acida in funzione dell'impiego di diverse sequenze 3'-UTR; 3) realizzazione dei vettori di espressione finali per la trasformazione di riso via *Agrobacterium tumefaciens*; 4) sviluppo di un protocollo di selezione e analisi sulla progenie.

2. MATERIALI E METODI

2.1 Scrittura in silico del gene GAA ottimizzato per l'espressione in riso

La sequenza codificante (CDS, coding sequence) la GAA umana (GenBank NM_000152) presenta numerosi polimorfismi naturali; punto di partenza in questa tesi è stata una CDS portante 7 nucleotidi diversi rispetto alla sequenza depositata in GenBank collegabili ad altrettante mutazioni puntiformi sinonime. Tale sequenza è stata gentilmente fornita dal laboratorio di malattie metaboliche dell'Istituto "Burlo Garofalo" di Trieste.

Allo scopo di ottimizzare la sequenza di interesse per l'espressione endosperma-specifica in riso, si è proceduto alla realizzazione di un gene sintetico che presentasse caratteristiche favorevoli alla produzione della proteina umana in questo compartimento del seme di riso. La CDS della GAA così modificata è stata riscritta in silico secondo il metodo del *codon context* stabilito da Venturini (2006), che prevede l'impiego per ogni amminoacido del codone sinonimo preferito in funzione del contesto intercodonico, ovvero del primo nucleotide del codone successivo (indicato con N₄). A tal fine, dopo aver determinato la sequenza amminoacidica corrispondente, il gene è stato totalmente riscritto sostituendo a ciascun amminoacido un codone scelto sulla base del *codon context* a partire dal codone di stop per proseguire a ritroso lungo il gene. Eventuali elementi di disturbo (omotetrameri) e sequenze di instabilità (TATA box spurie, AATAAA, ATTTA) sono stati successivamente individuati ed eliminati nel gene sintetico attraverso l'inserimento del secondo codone più frequente in relazione al contesto. Per favorire l'espressione in riso si è deciso di sostituire la sequenza codificante il peptide segnale naturale con quello della glutelina 4 (Glub4), anch'essa riscritta secondo il medesimo criterio.

La sequenza riscritta e modificata è stata sottoposta ad analisi con uno specifico software (Webcutter, www.firstmarket.com/cgi-bin/cutter/cut2.html), capace di creare siti di restrizione attraverso mutazioni puntiformi sinonime tali da non alterare la sequenza amminoacidica della proteina. In questo modo, è stato possibile inserire i seguenti siti di restrizione:

- *Bam* HI (G/GATCC) (904-909 bp);
- *Kpn* I (GGTAC/C) (1872-1877 bp).

Tali siti sono stati realizzati per poter provvedere alla sintesi della CDS in tre tratti; poiché infatti la lunghezza della CDS è pari a 2862 bp, la sintesi dell'intera CDS nonché le successive operazioni di controllo apparivano oltremodo difficoltose. Per ridurre le difficoltà operative e diminuire i costi di sintesi, la CDS è stata così suddivisa:

- a) frammento A di 915 bp;
- b) frammento B di 974 bp;
- c) frammento C di 985 bp.

All'inizio dei frammenti B e C sono stati inseriti rispettivamente i siti di restrizione *Bam* HI e *Kpn* I: questa operazione è finalizzata all'assemblaggio dei tre frammenti nel vettore pUC18 in modo tale da ricostituire la CDS completa senza alcuna alterazione amminoacidica rispetto alla proteina nativa. Va infine rilevato che all'inizio del frammento A e al termine di quello C sono stati rispettivamente inseriti i siti *Xba* I (T/CTAGA) e *Sac* I (GAGCT/C), e ciò per consentire la sub-clonazione dell'intera CDS da pUC18 al vettore di espressione finale.

La sintesi della nuova sequenza così ottenuta, comprendente il peptide segnale Glub4 e divisa nei tre frammenti, è stata poi affidata a una ditta specializzata (GenScript) che ha provveduto a clonare ciascun frammento in un vettore pUC18 utilizzando i siti di restrizione inseriti e a fornire anche i necessari controlli qualità.

2.2 Scrittura in silico del gene GCasi ottimizzato per l'espressione in riso

Per verificare i miglioramenti dei livelli di espressione apportabili tramite design genico in silico, dopo aver determinato la sequenza amminoacidica corrispondente, la CDS della GCasi (GenBank D16328) è stata totalmente riscritta sostituendo a ciascun amminoacido un codone scelto secondo due diversi criteri:

- a. metodo del *codon usage* basato sull'utilizzo dei codoni in assoluto più frequenti modificato per discriminare tra geni espressi a diversi livelli di espressione e in relazione al tessuto di provenienza;
- b. metodo del Codon Context, già descritto dettagliatamente nel paragrafo precedente riguardante la riscrittura dell' α -glucosidasi, che prevede l'impiego del codone sinonimo con la maggiore frequenza in un determinato contesto intercodonico (Venturini, 2006).

In quest'ultimo caso, le regioni codificanti sono state sintetizzate in direzione 3'→5', ovvero partendo dal codone stop e procedendo a ritroso lungo il gene. Nell'eventualità che si formassero omotetrameri congiuntamente all'utilizzo di un codone preferito si è optato per l'eliminazione dell'elemento inserendo il secondo codone preferito.

Per la clonazione delle sequenze sintetiche nel vettore finale di espressione, anche in questo caso, sono stati inseriti i siti *Xba* I e *Sac* I rispettivamente al terminale 5' e 3' della sequenza codificante.

La sintesi delle nuove sequenze così ottenute, comprendenti il peptide segnale Glub4 ottimizzato, è stata poi affidata alla ditta GenArt.

2.3 Analisi della sequenza

E' stata eseguita l'analisi della sequenza codificante sia la GAA nativa, sia quella ottenuta applicando le regole del *codon context*. In particolare, si è proceduto all'analisi dei seguenti elementi:

- determinazione del contenuto totale in GC e TA espresso in termini percentuali;
- conteggio dei dinucleotidi CG e TA;
- conteggio degli intercodoni C_3pG_4 e T_3pA_4 (dove i numeri in pedice definiscono la posizione codonica; 4 indica la prima base della tripletta successiva);
- conteggio della frequenza di codoni terminanti in G o C ottenuta dividendo il numero di C_3 o G_3 calcolati per il numero dei codoni totali;
- determinazione della percentuale di C e G presenti in terza posizione dividendo il numero di C_3 o G_3 calcolati per il numero di nucleotidi totali;

In entrambe le sequenze è stata, inoltre, individuata la presenza di eventuali elementi di instabilità, quali TATA box spurie, AATAAA, ATTTA e omotetrameri.

Con l'ausilio di tre diversi software disponibili in rete (NetGene, GenScan e GeneSplicer), sono state identificate le sequenze che potevano rappresentare siti criptici di splicing, cioè tratti erroneamente interpretati come introni. Tali introni criptici sono stati eliminati attraverso il design genico.

La medesima procedura è stata adottata per l'analisi della sequenza codificante la GCasi nativa, la GCasi riscritta secondo le regole del *codon context* e la GCasi ottenuta applicando il criterio del *codon usage*. Infatti, i geni sintetici ottenuti con i due diversi metodi sono stati comparati *in silico* tra loro e con la sequenza nativa al fine di identificare il grado di similarità complessiva, la presenza di eventuali elementi strutturali come TATA box spurie, omotetrameri, dinucleotidi CpG.

2.4 Metodiche comuni di biologia molecolare

Tutte le reazioni di amplificazione sono state effettuate utilizzando il termociclatore MyCycler™ Thermal cycler (Bio-Rad) o il termociclatore PCR Sprint® (Hybrid). Gli oligonucleotidi sintetici utilizzati, di lunghezza variabile da 15 a 30 basi, sono stati sintetizzati dalla ditta Sigma Genosys.

Le separazioni elettroforetiche di DNA su gel di agarosio (GellyPhor® LE, EuroClone) sono state effettuate con l'apparato BioRad alle condizioni di corsa 4 V/cm per circa 1 h. Il recupero da gel di amplificati PCR e di prodotti ottenuti da digestione enzimatica è stato realizzato mediante il kit Wizard® SV Gel & PCR Clean-up System (Promega). Nel caso di recuperi da gel di plasmidi di grandi dimensioni (>10 kb) è stato utilizzato l'enzima agarasi (NEB) in abbinamento a gel di agarosio low melting SeaPlaque GTG agarose (FMC BioProducts).

Nota: i marcatori utilizzati nella visualizzazione dei frammenti di DNA in gel di agarosio sono 1 kb DNA ladder (NEB) (Fig. 2.1 A), 1 kb DNA ladder Plus (Invitrogen) (Fig. 2.1.B) e 100 bp DNA ladder (NEB) (Fig. 2.1C)

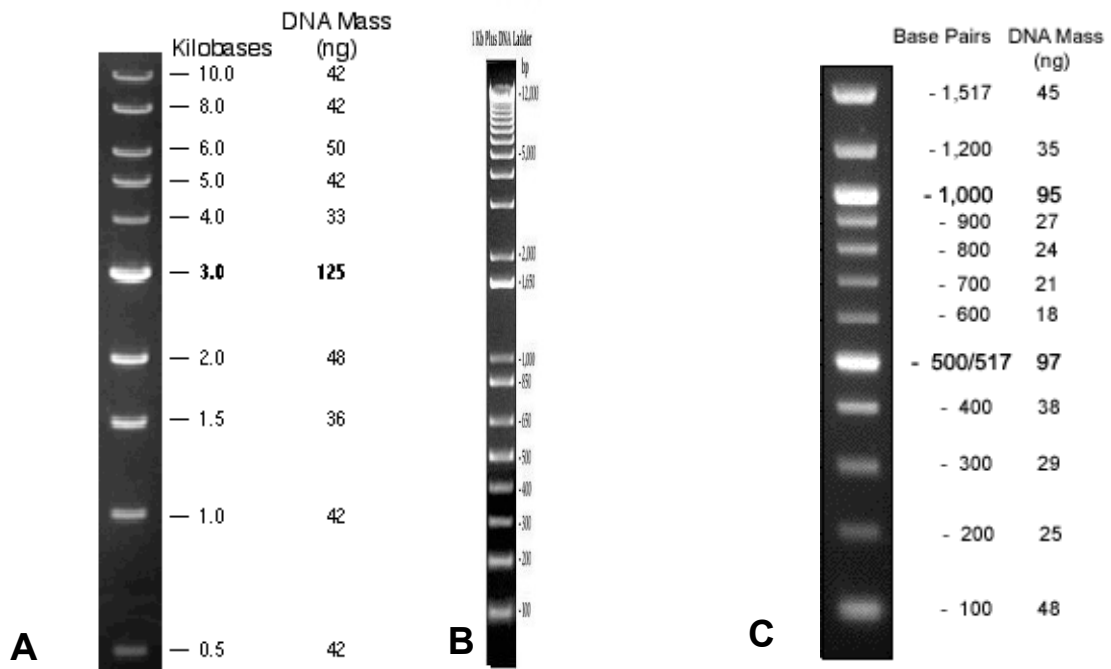


Figura 2.1 A, B, C. Marcatori molecolari utilizzati.

Le digestioni enzimatiche sono state effettuate utilizzando enzimi di restrizione e tamponi di digestione forniti da New England Biolabs (NEB).

Ciascuna reazione di digestione è stata allestita utilizzando i seguenti componenti:

Componenti	Quantità in μL
Buffer opportuno 10X	1.6 μL
BSA 100X	0.2 μL
Enzimi di restrizione (uno o più)	5-10 U
DNA da digerire	secondo necessità
H ₂ O	a volume
Volume finale	16.0 μL

Le reazioni di digestione sono state singole, doppie o triple a seconda dell'uso rispettivamente di 1, 2 o 3 enzimi di restrizione. Le digestioni, condotte con più di un enzima, hanno richiesto necessariamente una compatibilità di buffer (quanto maggiore possibile) tra i diversi enzimi di restrizione impiegati.

Le reazioni sono state lasciate in incubazione a 37°C per tempi variabili, da 2 a 14 ore (O/N, over-night) in funzione di diversi fattori, quali la quantità di DNA da digerire, l'efficienza dell'enzima e il possibile rischio di *star activity*. Per digestioni prolungate e in presenza di enzimi caratterizzati da *star activity*, alla reazione è stata aggiunta come stabilizzante albumina di siero bovino (BSA) (NEB).

Le reazioni di saldatura tra inserti e vettori sono state realizzate mediante l'enzima T4 ligasi (Promega) in presenza del tampone specifico, con incubazioni a 4°C overnight (ligation lenta) oppure a temperatura ambiente per almeno un'ora (ligation rapida). I quantitativi di inserto e vettore necessari per ciascuna reazione sono stati calcolati secondo la seguente formula:

$$\frac{\text{ng vettore} \times \text{bp inserto}}{\text{bp vettore}} \times \frac{3}{1} = \text{ng inserto}$$

considerando per ogni reazione una quantità di 25 ng di vettore, e un rapporto inserto/vettore di 3/1.

Ligation "rapida":

Vettore	25.0 ng
Inserto	ng calcolati
Ligasi buffer 2x (Promega)	5.0 µL
T4 ligasi	1.0 µL
Volume finale	10.0 µL

Incubazione a RT, minimo 1h

Ligation "lenta":

Vettore	25.0 ng	
Inserito	ng calcolati	
Ligasi buffer 2x (Promega)	1.0 μ L	
T4 ligasi	1.0 μ L	
Volume finale	10.0 μ L	Incubazione a 4°C O/N

Per la trasformazione di *Escherichia coli* è stato utilizzato il ceppo JM101 reso competente con CaCl_2 (Sambrook et al., 1989); la trasformazione di cellule competenti di *A. tumefaciens* (ceppo EHA105), preparate secondo il protocollo di Lin et al. (1995), è stata effettuata mediante elettroporazione con l'elettroporatore Micro Pulser™ (Bio-Rad).

Il DNA plasmidico dei ceppi trasformati è stato estratto mediante Wizard® plus minipreps (Promega) o QIAprep® spin miniprep Kit (Qiagen). Il sequenziamento dei tratti di DNA sintetizzati e/o subclonati è stato attuato da Primm srl.

2.5 Vettori utilizzati

Nell'ambito di questa tesi, per le operazioni di realizzazione e assemblaggio delle cassette geniche da trasferire nei vettori finali di espressione sono stati utilizzati i vettore commerciali pUC18, pGEM-T e pCAMBIA1300.

2.5.1 pUC18

pUC18 è un plasmide di *E. coli* ad alto numero di copie. Porta il gene per la resistenza all'ampicillina e presenta un MCS all'interno della regione codificante per il frammento α della beta-galattosidasi, rendendo possibile lo screening secondo colore delle colonie trasformate (Fig. 2.2). Infatti, l'inattivazione inserzionale permette ai cloni ricombinanti di essere identificati direttamente tramite α -complementazione: grazie alla presenza di un substrato cromogeno per la β -galattosidasi, le colonie che non presentano l'inserito appaiono blu, mentre quelle correttamente trasformate presentano il caratteristico colore bianco crema di *E. coli*.

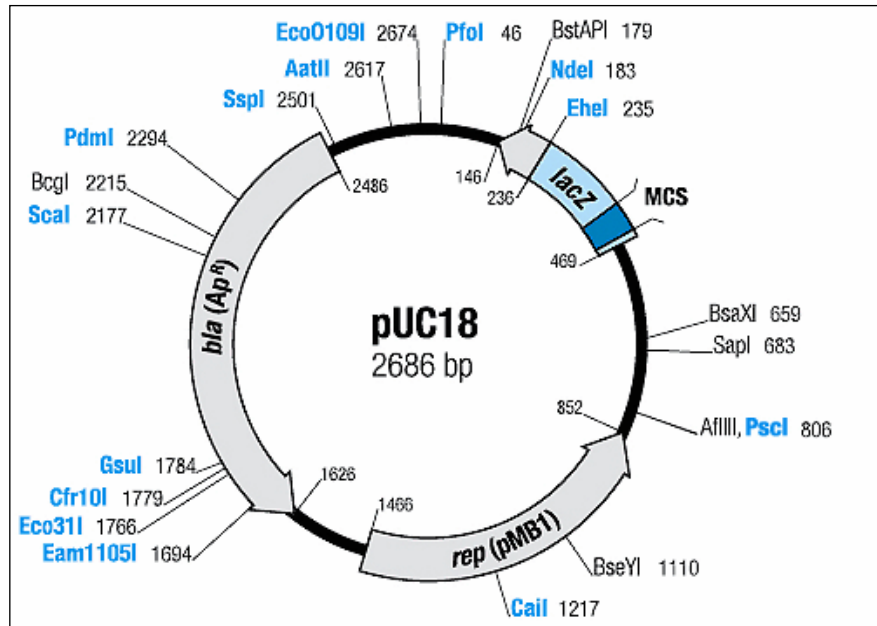


Figura 2.2. Rappresentazione del vettore pUC18.

2.5.2 pGEM-T

Il vettore pGEM[®]-T presenta ai lati del MCS i siti di legame per diversi primer, fra cui M13 forward e reverse, utilizzabili per l'amplificazione o il sequenziamento di inserti (www.promega.com).

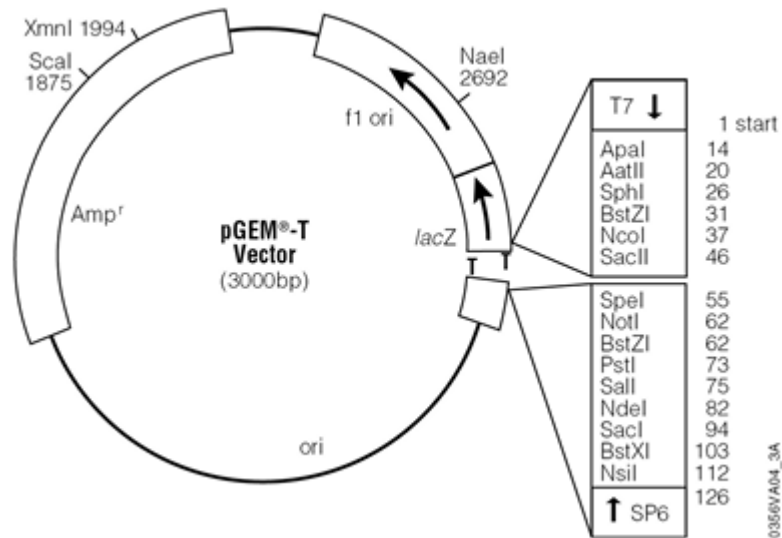


Figura 2.3. Rappresentazione del vettore pGEM-T vector.

Il vettore pGEM[®]-T è un plasmide high copy number derivato da pUC18 contenente i promotori T7 e SP6 nella regione fiancheggiante il sito di clonazione multiplo (MCS, multiple cloning site); quest'ultimo mappa all'interno della regione codificante per il frammento α dell'enzima β -galattosidasi (Fig. 2.3). L'inattivazione inserzionale permette ai cloni ricombinanti di essere identificati direttamente tramite α -complementazione.

2.5.3 pCAMBIA

Come base per il vettore finale di espressione in riso è stato, invece, utilizzato un vettore pCAMBIA opportunamente modificato, in cui sono state inserite le cassette geniche di interesse.

I vettori appartenenti alla famiglia pCAMBIA sono stati sviluppati appositamente per la trasformazione di piante mono- e dicotiledoni e, sebbene vadano ancora perfezionati, presentano alcune caratteristiche che li rendono adatti a questo tipo di lavoro:

- vengono replicati in alto numero di copie in *E. coli*;
- presentano un'elevata stabilità in agrobatterio grazie al replicone di pVS1;
- sono di dimensioni relativamente ridotte (7-12 Kb);
- presentano geni per la resistenza antibiotica utilizzabili per la selezione di cellule trasformate batteriche (kanamicina) e vegetali (igromicina).

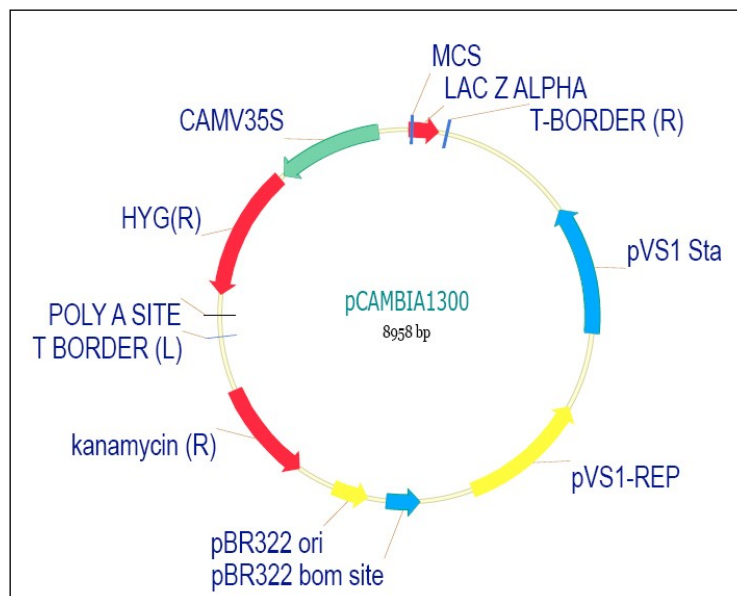


Figura 2.4. Struttura del vettore pCAMBIA1300.

Il vettore pCAMBIA 1300 (Fig. 2.4) è un vettore binario contenente un polylinker derivato da pUC18 e i geni *Kan^r* (nptIII) e *hptII* per la resistenza agli antibiotici kanamicina e igromicina, rispettivamente. Il gene *Kan^r* codifica per una neomicina fosfotransferasi in grado di inattivare gli antibiotici appartenenti alla famiglia della neomicina, in particolare kanamicina e amikomicina, conferendo un elevato livello di resistenza alle cellule batteriche che possiedono il plasmide portante tale gene e il promotore batterico. Il gene *hptII* codifica per una igromicina fosfotransferasi e permette di selezionare le piante trasformate utilizzando igromicina come agente di selezione sotto il controllo di un promotore attivo in pianta.

Il vettore utilizzato in questo lavoro è un pCAMBIA1300 modificato, indicato pCAMBIA13xx. Esso diverge dal vettore commerciale nella zona del MCS, che risulta quasi totalmente rimossa (delezione di 50 bp) al fine di impedire interferenze con i siti di restrizione unici presenti nella cassetta di interesse.

2.6 Realizzazione dei vettore di espressione per GAA e GCasi

Modificando il vettore commerciale pCAMBIA1300 (n. acc. GenBank AF234296) sono stati realizzati i vettori finali di espressione pCAMBIA_CDS di interesse-NOS ter_PMI e pCAMBIA_CDS di interesse-NOS ter_PMI_FXantiFX portanti rispettivamente due e tre cassette geniche. Entrambi i vettori sono caratterizzati da un unico T-DNA dotato di cassette per l'espressione del gene di interesse (GAA_CC; sequenza codificante l' α -glucosidasi riscritta secondo il criterio del Codon Context) e dell'agente selettivo PMI (fosfomannosio isomerasi) mentre differiscono tra loro per la presenza o meno della cassetta per l'espressione dell'RNAi diretto al silenziamento dei geni $\alpha(1,3)$ -fucosiltransferasi e $\beta(1,2)$ -xilosiltransferasi (FXantiFX). Mediante RNA interference si è voluto, infatti, silenziare due geni presenti nel genoma di riso codificanti per gli enzimi $\alpha(1,3)$ -fucosiltransferasi (GenBank acc. n° NM_001068555) e $\beta(1,2)$ -xilosiltransferasi (GenBank acc. n° NM_001068712). Tali enzimi sono rispettivamente responsabili dell'aggiunta dei residui glucosidici $\alpha(1,3)$ fucosio e $\beta(1,2)$ xilosio alle catene glicaniche delle proteine.

Per la sostituzione e l'introduzione delle cassette geniche, sono stati utilizzati specifici siti di restrizione presenti all'interno del T-DNA riportati in Fig. 2.5.

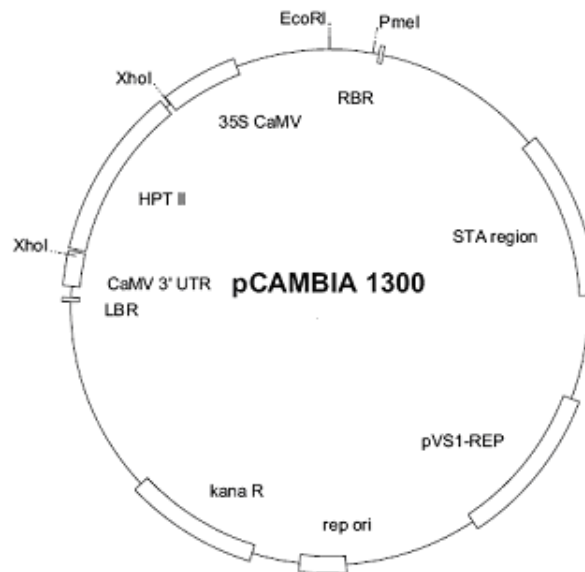


Figura 2.5 Rappresentazione schematica del vettore commerciale pCAMBIA1300 con evidenziati gli elementi strutturali e i siti di restrizione sfruttati per la preparazione dei vettori finali di espressione.

2.6.1 Realizzazione dei vettori di espressione per GAA

Prima di procedere alla descrizione dei vettori finali di espressione, si riportano le operazioni eseguite per l'ottenimento della sequenza completa della GAA CC.

I vettori pUC18 portanti ciascuno un frammento della GAA sono stati, infatti, sottoposti a una serie di digestioni enzimatiche al fine di ottenere la rimozione e il recupero dei frammenti d'interesse dai vettori per l'assemblaggio e la realizzazione del costrutto pUC18_GAA CC.

Come strategia da seguire per creare la sequenza intera della GAA si è deciso di tagliare il vettore pUC18_frammento C in modo tale da aprire il vettore e consentire dapprima l'inserimento del frammento B, e successivamente di quello A. A tale fine, sia per pUC18_frammento C, sia per pUC18_frammento B è stata impostata una digestione con *Aat* II e *Kpn* I.

Si è scelto di utilizzare l'enzima *Aat* II, anziché l'enzima *Bam* HI, per meglio distinguere i prodotti di digestione che altrimenti sarebbero più difficili da visualizzare in gel date le dimensioni simili. Le bande corrispondenti al vettore aperto pUC18_C e all'inserto B sono state recuperate e saldate in modo da ottenere pUC18_B+C. È stata, quindi, impostata la digestione sequenziale impiegando *Bam* HI e *Aat* II per aprire il nuovo vettore portante sia il frammento

C sia il frammento B e separare il frammento A dal vettore pUC18. La saldatura tra il vettore pUC18_B+C e l'inserto A ha consentito la produzione del vettore pUC18 portante la sequenza completa della GAA CC (pUC18_GAA CC).

2.6.1.1 pCAMBIA1300_GAA CC-NOS ter_PMI e pCAMBIA1300_GAA CC-NOS ter_PMI_FXAntiFX

In linea generale, la strategia seguita per la realizzazione del vettore a tre cassette ha previsto i seguenti adattamenti molecolari nel vettore pCAMBIA1300:

- 1) Inserimento della cassetta di espressione dell'enzima GAA;
- 2) Sostituzione del gene codificante igromicina fosfotransferasi (*hptII*), responsabile della resistenza all'antibiotico igromicina, con il gene *manA* codificante per l'enzima fosfomannosio isomerasi;
- 3) Inserimento della cassetta di espressione per il silenziamento combinato degli enzimi fucosiltransferasi e xilosiltransferasi in corrispondenza di *Pme I* (GTTT/AAAC).

Nel passaggio di clonazione finale di ciascuna cassetta si è operato in modo da orientare i costrutti secondo una direzione unica (più precisamente, promotore-gene-terminatore sono sempre disposti in senso antiorario).

Di seguito vengono descritte nel dettaglio le operazioni di assemblaggio e creazione delle cassette geniche effettuate nei diversi vettori preparativi.

Isolamento di PMI: il gene *manA* è stato isolato dal batterio *Escherichia coli* mediante PCR, utilizzando primer specifici disegnati sulla base della sequenza di *manA* depositata in banca dati (GenBank acc. n° M15380):

PMI for: 5'- CTC GAG ATG CAA AAA CTC ATT AAC T-3'

PMI rev: 5'- CTC GAG TTA CAG CTT GTT GTA AAC A -3'

Per favorire le operazioni di clonazione nel vettore finale di espressione, i primer PMI for e PMI rev portano il sito restrizione *Xho I* (C/TCGAG) ai terminali 5' e 3', rispettivamente.

Il gene così isolato è stato clonato all'interno del vettore pGEM-T

Realizzazione della cassetta per il silenziamento FxAntiFx: fatte salve le sequenze nucleotidiche relative ai geni da silenziare, tutte le altre derivano da un vettore ideato dal Commonwealth Scientific and Industrial Research Organisation

(CSIRO, Australia) per indurre silenziamento genico mediante *hairpin RNAi* (RNA interferente basato su strutture a forcina di RNA) in piante monocotiledoni. Per quanto riguarda le sequenze senso e antisenso dei geni $\alpha(1,3)$ -fucosiltransferasi e $\beta(1,2)$ -xilosiltransferasi, esse sono state isolate mediante PCR da cDNA di *O. sativa* mediante utilizzo dei seguenti primer, disegnati su un tratto di sequenza altamente conservata dei geni NM_001068555 e NM_001068712.

FucS for: GGATCCCGTCCTCGTCGGGGGCGCC

FucS rev: TCTAGAGCGCCTTCTCCAGGAGAGA

FucA for: GGTACCGGATCCCGT CCTCGTCGGG

FucA rev: TCTAGAGCGCCTTCTCCAGGAGAGA

XiIS for: GCTAGCAGGACGGGGCCGTTGGTGG

XiIS rev: TTAATTAAGTCCTCTCGCCGAACAA

XiIA for: GCTAGCAGGACGGGGCCGTTGGTGG

XiIA rev: ACT AGTGCCTCTCGCCGAACAAAG

La parte sottolineata della sequenza dei primer indica siti di restrizione introdotti per favorire le operazioni di assemblaggio della cassetta genica nel vettore pSTARLING.

Creazione del vettore pCAMBIA1300_GAA CC: la cassetta genica completa per l'espressione dell' α -glucosidasi acida, costituita dal promotore GluB4 LLTCK, dalla CDS della GAA e dal terminatore NOS, è stata escissa dal vettore pSV_GAA, realizzato in un precedente lavoro, tramite digestione con *EcoR* I (G/AATTC) e successivamente inserita nel vettore pCAMBIA1300 opportunamente aperto con il medesimo enzima.

Creazione del vettore pCAMBIA1300_GAA CC_PMI: per l'introduzione del gene *manA* il vettore pCAMBIA1300 è stato digerito con *Xho* I al fine di eliminare il gene *hptII* e consentirne la sostituzione in corrispondenza dello stesso sito di restrizione, realizzando il vettore a due cassette.

Creazione del vettore pCAMBIA1300_GAA CC_PMI_ FxAntiFx: l'inserto rappresentato dalle sequenze senso e antisenso è stato recuperato dal vettore pSTARLING che lo portava mediante digestione con *Not* I (GC|GGCCGC),

sottoposto a trattamento con T4 DNA polimerasi per eliminare le estremità *sticky* generate da *Not I* e favorire la *ligation blunt* con il vettore pCAMBIA1300_GAA CC_PMI in corrispondenza del sito *Pme I*. A saldatura avvenuta, è stato attuato il controllo del corretto orientamento della cassetta di silenziamento mediante digestione enzimatica con *Spe I* (A/CTAGT).

2.6.1.2 pCAMBIA1300_GAA CC-GluB4 ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FXAntiFX

Nel tentativo di incrementare i livelli di espressione in endosperma di riso, si è deciso di sostituire il terminatore NOS con quello della Glutelina 4 di riso (GluB4 ter). Sono stati, quindi, realizzati i vettori pCAMBIA1300_GAA CC-GluB4ter_PMI_ e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FXantiFX.

Data l'impossibilità di eseguire la sostituzione del terminatore NOS con quello della glutelina4 di riso direttamente nei vettori a due e tre cassette precedentemente realizzati (pCAMBIA1300_GAA CC-NOS ter_PMI e pCAMBIA1300_GAA CC-NOS ter_PMI_FxAntiFx) a causa della presenza di più siti *EcoR I* e *Sac I*, utili per il taglio del terminatore NOS, si è proceduto inserendo consecutivamente terminatore GluB4, promotore GluB4 e gene d'interesse GAA in vettori preparativi. Il suddetto promotore è stato recuperato dal vettore pGEM-T_GluB4 LLTCK già ottenuto in precedenti lavori di clonazione e caratterizzato dalla sostituzione del leader nativo presente nel promotore GluB4 con il leader sintetico LLTCK dotato di elementi TC e CAA ripetuti e contraddistinto dalla presenza della sequenza consenso Kozak (De Amicis et al., 2007). Il terminatore GluB4 munito dei siti *Sac I* e *EcoR I* rispettivamente ai terminali 5' e 3' è stato, invece, sintetizzato artificialmente dalla ditta GeneArt che ha provveduto alla sua clonazione all'interno di un opportuno vettore (pMK) mediante i siti di restrizione *Sac I* e *EcoR I*.

Come strategia da seguire per la sostituzione del terminatore si è deciso di ottenere dapprima i vettori intermedi pUC18_GluB4 ter e pUC18_GluB4 LLTCK_GluB4 ter e successivamente procedere al recupero della sequenza codificante l' α -glucosidasi acida riscritta secondo il metodo del codon context da pUC18_GAA CC e al suo inserimento in pUC18_GluB4 LLTCK_GluB4 ter. Si è, quindi, proceduto sostituendo il terminatore NOS, presente nel vettore pUC18_NOS ter, con quello della glutelina 4 di riso clonato nel vettore pMK

fornito dalla ditta incaricata della sua sintesi. A tal fine sia pUC18_NOS ter sia pMK_GluB4 ter sono stati digeriti con *EcoR* I e *Sac* I in modo da ottenere rispettivamente il vettore e l'inserito per produrre, tramite saldatura, il vettore intermedio pUC18_GluB4 ter. Lo step successivo ha previsto l'inserimento del promotore GluB4 LLTCK nel vettore intermedio. pUC18_GluB4 ter è stato, quindi, sottoposto a digestione con *Xba* I e *Sph* I (GCATG/C) così da ottenere il vettore per la saldatura del promotore GluB4 LLTCK escisso dal vettore che lo portava (pGEM-T_GluB4 LLTCK) mediante digestione con i medesimi enzimi. Ha fatto seguito l'introduzione della sequenza GAA CC nel vettore pUC18_GluB4 LLTCK_GluB4 ter prodotto. Tale vettore è stato aperto con *Xba* I e *Sac* I e saldato con l'inserito GAA CC recuperato da pUC18_GAA tramite digestione con *Xba* I, *Sac* I e *Aat* II, generando il vettore pUC18_GluB4 LLTCK_GAA_GluB4 ter portante il nuovo terminatore.

Per la realizzazione dei vettori finali a due e tre cassette, l'intera cassetta di espressione del gene di interesse presente in pUC18_GluB4 LLTCK_GAA CC_GluB4 ter è stata recuperata attraverso taglio con *EcoR* I.

Al fine di produrre il vettore pCAMBIA1300_GAA CC-GluB4 ter_PMI, il vettore pCAMBIA1300_GAA CC-NOS ter_PMI è stato digerito con *EcoR* I in modo da aprire il vettore e consentire la successiva saldatura della cassetta di espressione precedentemente recuperata, ottenendo il vettore finale a due cassette. Per il controllo del corretto orientamento dell'inserito si è proceduto alla digestione con *Sac* I.

Per la realizzazione del vettore a tre cassette, l'inserimento della sequenza FxAntiFX è avvenuto secondo la medesima procedura operativa già descritta nel paragrafo precedente (si veda *Creazione del vettore pCAMBIA1300_GAA CC_PMI_FxAntiFx*)

2.7 Realizzazione dei vettore di espressione per GCasi

Analogamente a quanto descritto per la GAA (cfr. § 2.6.1.1), sono stati realizzati i seguenti vettori di espressione finale della GCasi a tre cassette, pCAMBIA1300_GCasi CC-NOS ter_PMI_FXantiFX, pCAMBIA1300_GCasi CU-Nos ter_PMI_FXantiFX, pCAMBIA1300_GCasi nativa-NOS ter_PMI_FXantiFX, portanti rispettivamente la versione sintetica ottenuta applicando le regole del

codon context, la CDS riscritta secondo il metodo del *codon usage* e la sequenza nativa.

È stato, inoltre, costruito il vettore di espressione pCAMBIA1300_GCasi nativa-NOS ter_PMI privo della cassetta genica per il silenziamento e portante la sequenza GCasi nativa.

2.8 Trasformazione di *Oryza sativa* mediante *Agrobacterium tumefaciens*

Per la trasformazione di riso, varietà CR W3, è stato utilizzato il protocollo di Hiei et al. (1994), modificato da Hoge (Rice Research Group, Institute of Plant Science, Leiden University) e Guiderdoni (programma Biotrop, Cirad, Montpellier, France) fino all'ottenimento dei calli trasformati. Per la successiva fase di selezione è stato applicato, invece, il protocollo di Datta e Datta (2006). Qui di seguito sono brevemente descritte le principali fasi dell'intera procedura.

2.8.1 Preparazione e sviluppo di calli embrionici da scutello di riso

La trasformazione di riso è avvenuta su calli embrionici derivati da scutello. Per indurre la proliferazione di calli da tessuto scutellare è stato utilizzato il seguente protocollo operativo:

- è stata eseguita la sbramatura (eliminazione delle glume) di 200 semi di riso;
- per eliminare potenziali patogeni e saprofiti contaminanti che potrebbero interferire con la coltura di calli, si è proceduto quindi alla disinfezione delle cariossidi private delle glume:
 - a. il primo trattamento di disinfezione ha previsto la permanenza per 2 min in una soluzione di etanolo al 70%;
 - b. dopo il passaggio in etanolo, i semi sono stati trasferiti in una soluzione di ipoclorito di sodio al 5% con 2 gocce di detergente Tween20 e ivi mantenuti in agitazione lenta per 30 min;
 - c. per eliminare ogni traccia di ipoclorito di sodio che avrebbe potuto inibire l'induzione a callo degli scutelli, sono stati eseguiti una serie di lavaggi in H₂O sterile della durata di 15 min ciascuno;
- dopo aver effettuato l'ultimo lavaggio, i semi sono stati asciugati su carta bibula sterile;

- sulla superficie del substrato per l'induzione a callo (CIM, callus-induction medium), dispensato nel volume di 25 mL all'interno di piastre Petri (Ø 90 mm), sono stati posizionati 12 semi per piastra;
- le piastre così ottenute sono state incubate al buio, a una temperatura di 28°C per 21 giorni; dopo 1 settimana di incubazione si è proceduto all'eliminazione dell'endosperma e della radichetta per favorire lo sviluppo del callo proveniente dallo scutello (lo scutello si riconosce per la sua massa compatta, parzialmente inclusa nell'endosperma di colorazione gialla);
- terminate le 3 settimane di induzione, si è operato il trasferimento del callo su substrato CIM rinnovato, a cui ha fatto seguito la frammentazione delle masse callose senza l'utilizzo di bisturi, seguendo le linee di fratturazione naturalmente presenti sul callo;
- la sub-coltura è stata fatta proseguire per altri 10 giorni in modo da sviluppare il callo embriogenico e renderlo idoneo alla trasformazione.

2.8.2 Co-coltura dei calli con *A. tumefaciens*

1. Per ottenere quantità sufficienti di *A. tumefaciens* per la trasformazione, i ceppi portanti il plasmide contenente la cassetta di interesse e il gene per la fosfomannosio isomerasi sono stati incubati per 3 giorni a 30°C in LB-agar;
2. ottenute le colture di agrobatterio, le relative patine di crescita batterica sono state prelevate e sospese nel mezzo liquido di co-coltivazione (CCML, co-cultivation medium liquid), fino a ottenere una O.D. 600 di circa 1.0, corrispondente a $3\text{-}5\cdot 10^9$ cellule/mL;
3. i calli migliori, cioè quelli con un diametro di circa 2 mm, compatti e dal colore tendente al bianco, sono stati trasferiti in una piastra Petri contenente 35 mL di sospensione batterica e lasciati in immersione per 15 min in agitazione;
4. si è proseguito quindi all'asciugatura del callo utilizzando carta bibula sterile;
5. è stato disposto un numero massimo di 20 calli per piastra Petri high-edge (Sarstedt) contenente il substrato solido per la co-coltura (CCMS, co-cultivation medium solidified);

6. i calli sono stati quindi incubati in ambiente buio, a una temperatura di 25°C per 3 giorni.

2.8.3 Selezione dei calli basata su PMI

Dopo aver effettuato la co-coltura dei calli embriogenici di riso assieme all'agrobatterio, si è proceduto alla selezione dei tessuti trasformati sfruttando il sistema di selezione positiva basato su PMI (fosfomannosio isomerasi) come marker selezionabile e mannosio come agente selettivo. Tale metodo prevede l'impiego di substrati di coltura contenenti concentrazioni crescenti di mannosio e concentrazioni decrescenti di saccarosio.

La procedura impiegata è stata la seguente:

- trasferimento dei calli provenienti dalla co-coltura con *A. tumefaciens* su substrato PSM (pre-selection medium) privo di mannosio e contenente il 3% di saccarosio; incubazione per 1 settimana al buio a una temperatura di 28 °C;
- trasferimento dei calli su substrato di selezione SMI (selection medium I) contenente 2% saccarosio e 1.5% mannosio e incubazione al buio per 2 settimane a una temperatura di 28 °C;
- trasferimento dei calli su substrato di selezione SMII (selection medium II) contenente 1% saccarosio e 2% mannosio e incubazione al buio per 2 settimane a una temperatura di 28 °C;
- segue rigenerazione.

2.8.4 Rigenerazione di piantine di riso da calli trasformati

Dopo che il T-DNA portato dall'agrobatterio ha avuto modo di inserirsi in maniera stabile nel genoma di riso, si è proceduto alla rigenerazione delle piantine putativamente trasformate: questo processo è avvenuto grazie a un'opportuna stimolazione ormonale del callo trasformato seguendo la procedura qui riportata:

1. i calli embriogenici di riso selezionati sono stati trasferiti su piastre Petri high-edge contenenti il substrato per la pre-rigenerazione PRM (pre-regeneration medium) contenente 0.5% saccarosio e 2.5% mannosio e incubazione al buio per 2 settimane a una temperatura di 28 °C;
2. dopo il passaggio sul substrato PRM i calli sono stati trasferiti sul substrato per la rigenerazione RM (regeneration medium), privo di mannosio, nel numero

massimo di 8-10 unità per piastra Petri high-edge. La crescita delle piantine è avvenuta in presenza di luce, a 28°C per 3-4 settimane;

3. quando le piantine sono risultate sufficientemente grandi da poter essere separate dal callo (≥ 3 cm di altezza), si è proceduto al loro trasferimento in tubi di coltura contenenti 25 mL del substrato per la radicazione rm (rooting medium);

4. la sub-coltura all'interno dei tubi è proseguita per circa 3 settimane sempre a 28°C alla luce;

5. a conclusione del processo rigenerativo, le piante sono state trasferite in torba e allevate in serra.

Di seguito si riporta la composizione di ciascun substrato impiegato nel processo di trasformazione genetica mediato da *Agrobacterium tumefaciens* e le dosi per litro dei vari componenti.

Composizione	CIM	CCML	CCMS	PSM	SMI	SMII	PRM	RM	rm
N6 Macroelementi I (mL)	50				50	50	50	50	
N6 Macroelementi II (mL)	50				50	50	50	50	
MS FeNaEDTA (mL)	10				10	10	10	10	
B5 Vitamine (mL)	10				10	10	10	10	
B5 Microelementi (mL)	1				1	1	1	1	
Prolina (mg)	500				500	500	500	500	
Glutamina (mg)	500				500	500	500	500	
CEH (mg)	300				300	300	300	300	
MES (mg)	500				500	500	500	500	500
RA (mL)		25	25	25					
RB (mL)		25	25	25					
RC (mL)		25	25	25					
Tiamina (mg)		1	1	1					0.1
2,4-D (mg)		2.5	2.5	2.5	2.5	2.5			
Glucosio (g)		10	10						
MES (mg)		500	500	500					
Acetosiringone (mM)		0.1	0.1						
Agarosio SPI (g)			7	7	7	7			
Phytigel (g)							4.5	4.5	2.5
Saccarosio (g)	30			30	20	10	5	30	50
Mannosio (g)					15	20	25		
Cefotaxime (mg)				400	400	400	250	50	
ABA (mg)							5		
BAP (mg)							2	3	
NAA (mg)							1	0.5	
MS Sali (mL)									100
Glicina (mg)									2
Acido nicotinico (mg)									0.5
Piridossina (mg)									0.5
Inositolo (mg)									100
H ₂ O	q.b. a 1 L	q.b. a 1 L	q.b. a 1 L	q.b. a 1 L	q.b. a 1 L	q.b. a 1 L	q.b. a 1 L	q.b. a 1 L	q.b. a 1 L

N6 macroelementi II	Dose per 1 L
CaCl ₂ 2H ₂ O	3.32 g

N6 macroelementi I	Dose per 1 L
KNO ₃	56.60 g
(NH ₄)SO ₄	9.26 g
KH ₂ PO ₄	8.00 g
MgSO ₄ 7H ₂ O	3.70 g

B5 microelementi	Dose per 1 L
MnSO ₄ H ₂ O	10 g
KI	0.75 g
H ₃ BO ₃	3 g
Zn ₄ 7H ₂ O	2 g
CuSO ₄	0,025 g
Na ₂ MoO ₄ 2H ₂ O	0,25 g
CoCl ₂ 6H ₂ O	0,025 g

B5 Vitamine	Dose per 1 L
Inositolo	10 g
Tiamina-HCl	1 g
Acido nicotinico	0.1 g
Piridossina-HCl	0.1 g

Composizione R_A	Dose per 1 L
KNO ₃	162 g

Composizione R_C	Dose per 1 L
CaCl ₂ 2H ₂ O	6 g
H ₃ BO ₃	114.4 mg
Na ₂ MoO ₄ 2H ₂ O	5.2 mg
FeSO ₄ 7H ₂ O	496 mg
Na ₂ EDTA2H ₂ O	668 mg

MS FeNaEDTA	Dose per 1 L
FeSO ₄ 7H ₂ O	2.784 g
Na ₂ EDTA2H ₂ O	3.724 g

Composizione MS sali	Dose per 1 L
MnSO ₄ H ₂ O	16.9 mg
CuSO ₄ 5H ₂ O	0.025 mg
ZnSO ₄ 7H ₂ O	8.63 mg
CaCl ₂ 2H ₂ O	440 mg
KH ₂ PO ₄	170 mg
KI	0.83 mg
NH ₄ NO ₃	1.650 g
KNO ₃	1.9 g
NaMoO ₄ 2H ₂ O	0.25 mg
EDTAFENa	40 mg
H ₃ BO ₃	6.2 mg
MgSO ₄ 7H ₂ O	370 mg
CaCl ₂ 6H ₂ O	0,025 mg

Composizione R_B	Dose per 1 L
MgSO ₄ 7H ₂ O	10 g
(NH ₄) ₂ SO ₄	13.2 g
NaH ₂ PO ₄ H ₂ O	11 g
ZnSO ₄ 7H ₂ O	88 mg
MnSO ₄ H ₂ O	80 mg
CuSO ₄ 5H ₂ O	8 mg

2.9 Estrazione di DNA da foglie di riso trasformato con i vettori pTRS_GAA CC-GluB4 ter _PMI_ e pTRS_GAA CC-GluB4 ter _PMI_FXantiFX_

Per estrarre il DNA dalle foglie di riso trasformato con i vettori a due o a tre cassette portanti la sequenza GAA_CC e il terminatore GluB4 è stato utilizzato il protocollo di Doyle e Doyle (1997) modificato opportunamente per tenere conto delle minori quantità di materiale fresco di partenza (< 0.5 g).

2.10 Verifica della presenza del gene GAA in piante putativamente trasformate

Per verificare la presenza del gene GAA nelle piante di riso putativamente trasformate con i suddetti vettori, il DNA genomico estratto da foglia è stato sottoposto ad analisi PCR.

2.11 PCR di controllo per saggiare l'amplificabilità del DNA estratto

Per controllare che il DNA estratto fosse effettivamente amplificabile e, dunque, prevenire errate o dubbie interpretazioni dei risultati ottenuti con PCR, è stata programmata l'amplificazione di un tratto di DNA presente a livello di cpDNA di riso.

I primer utilizzati, denominati CPC e CPD, hanno la seguente sequenza nucleotidica:

CPC: 5'-CGAAATCGGTAGACGCTACG-3'

CPD: 5'-GGGGATAGAGGGACTTGAAC-3'

Ciclo: 95°C per 5' 40x (95°C per 30", 62°C per 30", 72°C per 2'), 72°C per 7'.

2.12 PCR per l'amplificazione del gene GAA

Per verificare la presenza del gene codificante la GAA, le piante sono state saggiate tramite PCR impiegando i seguenti primer:

PRIMER GAA A_for: 5'-GTCCATCTACTTCTGCGTGC-3'

PRIMER GAA A_rev: 5'-AGGTAGAAGGGATGGGAGC-3'

PRIMER GAA B_for: 5'-GTTCTGCTCAACTCCAATG-3'

PRIMER GAA B_rev: 5'-GTTGTGCAGGTTGTAATGGG-3'

PRIMER GAA C_for: 5'-AGATCCTGCAGTTCAACCTG-3'

PRIMER GAA C_rev: 5'-ATGTCCAGCACTTTGGTGTC-3'

Ciclo: 95°C per 5' 40x (95°C per 30", 62°C per 30", 72°C per 2'), 72°C per 7'.

2.13 PCR per l'amplificazione del backbone del vettore

Per verificare il corretto inserimento nelle piante trasformate delle sole cassette geniche di interesse comprese tra RBR e LBR, sono state allestite reazioni di amplificazione di un tratto del vettore esterno agli stessi. A tale fine è stata utilizzata la seguente coppia di primer:

LB_5: 5'-GTTTTATGTTTCAGATAATGCCCGATGAC-3'

KANRF2: 5'-CGGCCAGATCGTTATTCAGTAAGTAATC-3'

2.14 Analisi sulle progenie di piante GCasi e GAA

Il seme di riso prodotto dai trasformati primari è stato analizzato al fine di identificare le linee aventi il maggior contenuto di enzima ricombinante nell'endosperma. A questo scopo è stato sviluppato un protocollo di selezione basato sull'estrazione di proteina totali da seme e sulla quantificazione della proteina di interesse mediante saggio immunologico DAS-ELISA.

2.14.1 Estrazione di proteine totali da semi GAA e GCasi

I semi di riso trasformati sono stati sottoposti a sbramatura e, quando desiderato, a sbiancatura meccanica con apposita sbiancatrice Satake TO-92 (Satake Corporation), fornita dall'Ente Nazionale Risi.

Per la quantificazione del contenuto di enzima ricombinante nell'endosperma tramite analisi DAS-ELISA, le proteine totali sono state estratte da semi non sbiancati, mentre ai fini della purificazione cromatografica della proteina di interesse, gli estratti sono stati ottenuti da semi sbiancati.

Per ottenere estratti di proteine totali da saggiare mediante DAS-ELISA è stato messo a punto un protocollo di estrazione che ha previsto le seguenti fasi:

- Raccolta di 1-2 pannocchie mature da ciascun individuo
- Essiccazione delle pannocchie in locale asciutto e areato per circa 3 giorni fino al raggiungimento di una umidità relativa del seme pari al 14%
- Campionatura casuale di 40 semi per ciascuna linea
- Sbramatura dei semi con sbramino manuale da tavolo per riso
- Macinazione del campione con macina caffè da tavolo o mulino a sfere da laboratorio Retsch MM2 a velocità di 30 oscillazioni/min per 2 minuti e prelievo di 70 mg della farina ottenuta
- Omogeneizzazione in mortaio della farina con 1 mL di opportuno tampone di estrazione scelto in funzione della proteina di interesse da saggiare
- Successiva diluizione con ulteriori 7 mL del medesimo tampone
- Incubazione in agitazione continua secondo le seguenti condizioni:
 - 4°C per 1 h nel caso di estratti GCasi
 - 15 minuti a RT per i campioni GAA
- Prelievo di 1 mL e centrifugazione a 20000xg per 40 min a 4°C

- Recupero della fase liquida contenente le proteine e conservazione a -20°C.

I tamponi di estrazione sono stati i seguenti:

- Tris-HCl 50 mM, NaCl 0.5 M pH=7.0 per l'estrazione di proteine totali da semi GCasi
- Idrogenofosfato disodico 250 mM + NaCl 250 mM pH=7.4 per l'estrazione di proteine totali da campioni GAA.

Per la produzione di estratti GAA da sottoporre a purificazione, i semi sbiancati sono stati sfarinati mediante mulino analitico; 100 g di farina sono stati prelevati ed è seguita omogeneizzazione con tampone di estrazione (PBS 3X + 2 mM DTT + 10 mM EDTA), utilizzando un rapporto tra volume di tampone (mL) e peso dei semi (g) pari a 5:1. Dopo incubazione a 4°C per 40' in agitazione e a RT per 20', si è eseguita una centrifugazione a 2490 RPM per 20 minuti. Dopo il recupero del surnatante, al fine di estrarre tutta la proteina di interesse, il pellet residuo è stato sottoposto a due ulteriori estrazioni utilizzando un rapporto tra volume di tampone (mL) e peso dei semi (g) pari a 2:1, ed effettuando un'incubazione per 10' a RT, seguita da centrifugazione a 2490 RPM per 10'.

2.14.2 Estrazione di proteine totali da singoli semi GCasi

Per una valutazione del contributo offerto dai singoli semi all'ammontare complessivo di enzima ricombinante GCasi, i due migliori trasformati primari GCasi CC-NOS ter_PMI_FxAntiFx sono stati sottoposti all'analisi su seme singolo. A tal fine, sono stati isolati 22 semi i quali sono stati sbramati manualmente e singolarmente omogeneizzati in mortaio utilizzando un volume di tampone di estrazione pari a 2 mL. L'omogenato è stato recuperato, lasciato in incubazione per 1 h a 4°C in agitazione e successivamente centrifugato 20000xg per 40 min a 4°C. Il surnatante recuperato è stato conservato a -20° C.

Per valutare il contenuto di GCasi, gli estratti proteici ottenuti dai singoli semi sono stati analizzati in DAS-ELISA, applicando le condizioni descritte nei paragrafi successivi.

2.14.3 Metodiche comuni di analisi proteica

Una prima quantificazione del contenuto proteico degli estratti è stata ottenuta con un saggio biochimico di tipo Bradford, tramite lettura dell'estratto a Abs₅₉₅ utilizzando il lettore di piastra Tecan.

Per la determinazione dell'enzima ricombinante di interesse gli estratti proteici crudi sono stati sottoposti a saggio immunologico DAS-ELISA.

2.14.4 Saggio immunoenzimatico (DAS-ELISA)

Il saggio DAS-ELISA si basa su doppio riconoscimento immunologico. Un anticorpo specifico per la proteina che si vuole rilevare, chiamato *capturing*, viene ancorato al fondo dei pozzetti di una piastra di polistirene (reazione di *coating*). Terminata la reazione di *coating*, ciascun pozzetto viene messo a contatto con una soluzione di BSA per saturare la residua capacità legante del polistirene (*blocking*). Dopo una serie di lavaggi, i campioni che si suppongono contenere la proteina di interesse vengono seminati nei pozzetti. Dopo successivi lavaggi, un secondo anticorpo, specifico per la proteina di interesse, coniugato a una enzima in grado di catalizzare una reazione che permette rilevare la presenza della proteina, viene dispensato in tutti pozzetti. Dopo i lavaggi, il complesso *capturing*-proteina-coniugato viene rilevato tramite contatto con una soluzione contenente un adeguato substrato cromogenico, come ad es. il 3,3',5,5'-tetramethylbenzidine (TMB).

Il *capturing* è stato prodotto dalla purificazione su proteina A dell'antisiero di coniglio anti-Myozyme[®] o anti-Cerezyme[®] a seconda che la proteina di interesse da rilevare fosse rispettivamente la GAA o la GCasi.

L'anticorpo policlonale anti-GCasi o anti-GAA coniugato alla perossidasi di rafano (HRP) è stato ottenuto con il kit EZ-link Plus Activated Peroxidase, prodotto da Pierce.

a) Soluzione di *coating*:

PBS diluito 1:5, sodio azide 0.01% + anti- Myozyme[®] o anti- Cerezyme purificato a una concentrazione finale rispettivamente di circa 4 ng/μL e 2 ng/ μL. Dopo aver dispensato 100 μL di questa soluzione in ciascun pozzetto, la piastra è stata incubata O/N a 4°C.

b) Soluzione di *blocking*:

PBS + BSA 2.5% + sodio azide 0.01%; 300 μ L di questa soluzione sono stati immessi in ciascun pozzetto e ivi mantenuti per almeno 1 h a RT prima di procedere al loro lavaggio con tampone PBS + Tween20 0.1%.

c) I campioni (estratti proteici grezzi contenenti GAA, GCasi o frazioni cromatografiche) sono stati opportunamente diluiti in tampone di diluizione (PBS + Tween20 0.1% + BSA 1%). Una volta diluiti, i campioni sono stati seminati nel volume di 50 μ L nei pozzetti e incubati per 30 minuti a 37°C in agitazione.

d) Dopo successivi lavaggi, l'anticorpo anti-Myozyme[®] o anti-Cerezyme[®] coniugato a HRP è stato opportunamente diluito in tampone di diluizione (stesso tampone utilizzato per i campioni) e seminato (50 μ L) nei pozzetti. Dopo 30 minuti di incubazione a 37°C in agitazione, si è proceduto con i lavaggi e con la rilevazione del segnale.

e) Per lo sviluppo del segnale, è stato utilizzato il substrato TMB, tetrametilbenzidina (Ultra-TMB, Pierce) (100 μ L per pozzetto) e la reazione è stata bloccata con 100 μ L acido cloridrico 1 M, dopo circa 5 minuti di incubazione. L'assorbanza è stata misurata a 450 nm utilizzando il lettore di piastra Tecan.

Per calcolare la concentrazione della proteina di interesse nei campioni, in ciascun saggio è stata costruita una curva di taratura con concentrazioni note di Myozyme[®] o Cerezyme[®].

Per l'elaborazione dei dati si è proceduto all'immissione dei valori di assorbanza nel software Curve Fitting Data Analysis (Promega) e all'assegnazione dei valori di concentrazione nota degli standard. I valori di concentrazione dei campioni ottenuti come output del software utilizzando una curva lineare a quattro parametri sono stati successivamente elaborati considerando il fattore di diluizione adottato al fine di ottenere le concentrazioni reali degli estratti.

2.14.5 Validazione del saggio DAS-ELISA

Allo scopo di verificare la validità del metodo DAS-ELISA utilizzato nella determinazione del contenuto dell'enzima ricombinante negli estratti proteici

ottenuti da endosperma di riso in termini di specificità e selettività, sia per GAA sia per GCasi, sono stati allestiti tre distinti test ELISA, ciascuno con il medesimo estratto proteico saggiato a tre diverse diluizioni. A ciascuna diluizione è stato successivamente aggiunto lo standard (Myozyme o Cerezyme a seconda che la molecola da quantificare fosse α - o β -glucosidasi) alla concentrazione nota di 0.8 ng/ μ L. Il metodo DAS-ELISA è stato validato quando i risultati relativi ai campioni con aggiunta dello standard sono ricaduti all'interno dell'intervallo di validità stabilito, corrispondente a \pm 15% del valore atteso.

2.15 Precipitazione frazionata delle proteine dall'estratto crudo con solfato d'ammonio

Prima di procedere al primo passaggio cromatografico, sull'estratto grezzo è stata effettuata una precipitazione frazionata con solfato di ammonio. Aggiungendo una soluzione satura $(\text{NH}_4)_2\text{SO}_4$ goccia a goccia, l'estratto grezzo è stato portato al 15% di saturazione di solfato di ammonio. La soluzione così ottenuta è stata incubata a 0°C per 45 minuti, sotto agitazione. È seguita una centrifugazione a 2490 RPM per 20 minuti. Al surnatante recuperato è stato aggiunto $(\text{NH}_4)_2\text{SO}_4$ per arrivare a un livello di saturazione del 50%. In seguito, esso è stato incubato a 0°C per 1 h e 30 min in agitazione e poi centrifugato a 2490 RPM per 20 min. Il surnatante è stato scartato e il pellet contenente la proteina di interesse è stato recuperato, asciugato con carta assorbente Whatman e ridisciolti nel tampone di carico per il primo passaggio cromatografico.

2.16 Purificazione di GAA

Per la purificazione di GAA è stato messo a punto un protocollo basato su tre step di purificazione: una prima fase di capturing eseguita mediante cromatografia di affinità con concanavalina A, una fase intermedia attuata con una cromatografia a interazioni idrofobiche (HIC) e una fase finale di *polishing* realizzata mediante gel-filtrazione (GF). Tutte le cromatografie sono state effettuate con il cromatografo AKTA Prime (GE Healthcare).

2.16.1 Cromatografia con concanavalina A

La colonna derivatizzata con concanavalina A (*hitrap fast flow* 5 mL) è stata equilibrata con *binding buffer* composto da Mes-HCl pH 5.4, NaCl 0.5 M, CaCl₂, 1 mM, MnCl₂ 1mM. Il pellet ottenuto dalla precipitazione frazionata (cfr. § 2.15) è stato opportunamente risospeso in *binding buffer* e caricato in colonna a una velocità di flusso pari a 1mL/min.

Dopo opportuni lavaggi con *binding buffer*, è stata impostata un'eluizione con tampone di eluizione (*binding buffer* + α -D-glucopiranoside 0.25M)

Il *flow-through* e l'eluato sono stati raccolti in frazioni da sottoporre a successive analisi immunologiche (ELISA).

2.16.2 Cromatografia a interazioni idrofobiche (HIC)

E' stata utilizzata una colonna preimpaccata Hitrap butyl FF da 1 mL (GE Healthcare). Inizialmente la colonna è stata equilibrata con un volume di tampone di carico (25 mM sodio fosfato, 0.8 M ammonio solfato, pH 6.5). E' seguito il carico delle frazioni positive derivate da cromatografia IEC portate a 0.8 M di solfato di ammonio. Il carico in colonna è stato eseguito a una velocità di 1 mL/minuto. La colonna è stata quindi lavata con tampone di carico fino a ottenere un cromatogramma piatto. L'eluizione è stata eseguita con un gradiente da 10% a 60% di tampone B (sodio fosfato 15 mM, pH 6.5). Alla fine della procedura, la colonna è stata lavata con tampone B e rigenerata con 20% etanolo.

2.16.3 Gel-filtrazione (GF)

E' stata utilizzata una colonna preimpaccata Sephacryl S-100 16/60 (GE Healthcare) e un tampone di corsa costituito da 15 mM sodio fosfato, pH 6.5. La colonna è stata inizialmente equilibrata con il tampone a cui è seguito il carico in colonna dell'eluato concentrato prodotto mediante cromatografia HIC. L'intera procedura è stata eseguita a una velocità di flusso pari a 0.5 mL/minuto. Le frazioni raccolte sono state saggiate in ELISA e SDS-PAGE.

Sez A – RISULTATI PER GAA

3.1 Riscrittura della sequenza codificante la GAA umana e costruzione del vettore di espressione

L'analisi della sequenza nativa codificante la GAA umana ha evidenziato una serie di caratteristiche non favorevoli all'espressione in endosperma di riso. In particolare, è stato osservato un numero elevato di elementi di disturbo quali omotetrameri con base C/G, C- e G-runs nonché la presenza di una TATA box spuria.

CDS della GAA umana (lunghezza di 2859 bp)

```
ATGGGAGTGAGGCACCCGCCCTGCTCCCACCGGCTCCTGGCCGTCTGCGCCCTCGTGTC
CTTGGAACCGCTGCACTCCTGGGGCACATCCTACTCCATGATTTCTGCTGGTTCCCC
GAGAGCTGAGTGGCTCCTCCCCAGTCCCTGGAGGAGACTCACCCAGCTCACCAGCAGGGA
GCCAGCAGACCAGGGCCCCGGGATGCCAGGCACACCCCGGCCGTCCCAGAGCAGTGCC
CACACAGTGCAGCGTCCCCCCCAACAGCCGCTTCGATTGCGCCCCTGACAAGGCCATCA
CCCAGGAACAGTGCGAGGCCCGCGGCTGCTGCTACATCCCTGCAAAGCAGGGGCTGCAG
GGAGCCCAGATGGGGCAGCCCTGGTGCTTCTTCCCACCCAGCTACCCCAGCTACAAGCT
GGAGAACCTGAGCTCCTCTGAAATGGGCTACACGGCCACCCTGACCCGTACCACCCCCA
CCTTCTTCCCCAAGGACATCCTGACCCTGCGGCTGGACGTGATGATGGAGACTGAGAAC
CGCCTCCACTTACAGATCAAAGATCCAGCTAACAGGGCGCTACGAGGTGCCCTTGGAGAC
CCCGCGTGTCCACAGCCGGGCACCGTCCCCACTCTACAGCGTGGAGTTCTCTGAGGAGC
CCTTCGGGGTGATCGTGCGCCGGCAGCTGGACGGCCGCGTGCTGCTGAACACGACGGTG
GCGCCCCGTGTTCTTTGCGGACCAGTTCTTTCAGCTGTCCACCTCGCTGCCCTCGCAGTA
TATCACAGGCCTCGCCGAGCACCTCAGTCCCCTGATGCTCAGCACCAGCTGGACCAGGA
TCACCCTGTGGAACCGGGACCTTGCGCCACGCCCGGTGCGAACCTCTACGGGTCTCAC
CCTTTCTACCTGGCGCTGGAGGACGGCGGGTTCGGCACACGGGGTGTTCTGCTAAACAG
CAATGCCATGGATGTGGTCTGCAGCCGAGCCCTGCCCTTAGCTGGAGGTGCACAGGTG
GGATCCTGGATGTCTACATCTTCTGGGCCAGAGCCCAAGAGCGTGGTGCAGCAGTAC
CTGGACGTTGTGGGATAACCGTTTCATGCCGCCATACTGGGGCTGGGCTTCCACCTGTG
CCGCTGGGGCTACTCCTCCACCGCTATCACCCGCCAGGTGGTGGAGAACATGACCAGGG
CCCACCTTCCCCCTGGACGTCCAATGGAACGACCTGGACTACATGGACTCCCGGAGGGAC
TTCACGTTCAACAAGGATGGCTTCCGGGACTTCCCGGCCATGGTGCAGGAGCTGCACCA
GGGCGGCCGGCGCTACATGATGATCGTGGATCCTGCCATCAGCAGCTCGGGCCCTGCCG
GGAGCTACAGGCCCTACGACGAGGGTCTGCGGAAGGGGGTTTTTCATACCAACGAGACC
GGCCAGCCGCTGATTGGGAAGGTATGGCCCGGGTCCACTGCCTTCCCCGACTTCACCAA
CCCCACAGCCCTGGCCTGGTGGGAGGACATGGTGGCTGAGTTCCATGACCAGGTGCCCT
TCGACGGCATGTGGATTGACATGAACGAGCCTTCCAACCTCATCAGGGGCTCTGAGGAC
GGCTGCCCAACATGAGCTGGAGAACCACCTACGTGCCTGGGGTGTTGGGGGAC
CCTCCAGGCGGCCACCATCTGTGCCTCCAGCCACCAGTTTCTCTCCACACACTACAACC
TGCACAACCTCTACGGCCTGACCGAAGCCATCGCCTCCCACAGGGCGCTGGTGAAGGCT
CGGGGACACGCCCATTTGTGATCTCCCGCTCGACCTTTGCTGGCCACGGCCGATACGC
CGGCCACTGGACGGGGGACGTGTGGAGCTCCTGGGAGCAGCTCGCCTCCTCCGTGCCAG
AAATCCTGCAGTTTAACTGCTGGGGGTGCCTCTGGTCCGGGCGGACGTCTGCGGCTTC
CTGGGCAACACCTCAGAGGAGCTGTGTGTGCGCTGGACCCAGCTGGGGCCTTCTA CCC
CTTCATGCGGAACCACAACAGCCTGCTCAGTCTGCCAGGAGCCGTACAGCTTCAGCG
AGCCGGCCAGCAGGCCATGAGGAAGGCCCTCACCTGCGCTACGCACTCCTCCCCCAC
```

CTCTACACGCTGTTCCACCAGGCCACGTCGC GGGG GAGACCGTGGCCCGG CCCC TCTT
CCTGGAGTT CCCC AAGGACTCTAGCACCTGGACTGTGGACCACCAGCTCCTGT GGGGGG
AGGCCCTGCTCATCA CCCC AGTGCTCCAGGCCGGAAGGCCGAAGTGACTGGCTACTT C
CCC TTGGGCACATGGTACGACCTGCAGACGGTGCCAATAGAGGCCCTTGGCAGCCTCCC
A CCCC CACCTGCAGCT CCCC GTGAGCCAGCCATCCACAGCGA GGGG CAGTGGGTGACGC
TGCCGG CCCCCC TGGACACCATCAACGTCCACCTCCGGGCTGGGTACATCAT CCCC CTG
CAGGGCCCTGGCCTCACAACCACAGAGTCCC GCCAGCAGCCATGGCCCTGGCTGTGGC
CCTGACCAAGGGTGGAGAGGCCCGA GGGG AGCTGTTCTGGGACGATGGAGAGAGCCTGG
AAGTGCTGGAGCGA GGGG CCTACACACAGGTCATCTTCTGGCCAGGAATAACACGATC
GTGAATGAGCTGGTACGTGTGACCAGTGAGGGAGCTGGCCTGCAGCTGCAGAAGGTGAC
TGTCTGGGCGTGGCCACGGCG CCCC AGCAGGTCCTCTCCAACGGTGTCCCTGTCTCCA
ACTTCACCTACAG CCCC GACACCAAGGTCCTGGACATCTGTGTCTCGCTGTTGATGGGA
GAGCAGTTTCTCGTCAGCTGGTGT TAG

Traduzione:

MGVRHPPCSHRL LAVCALVSLATAALLGHILLHDFLLVPRELSGSSPVLEETHPAHQOQ
ASRPGPRDAQAHPRPRAVPTQCDVPPNSRFD CAPDKAITQEQCEARGCCYIPAKQGLQ
GAQMGPWCFFPPSYPSYKLENLSSSEMGYTATLTRTPTFFPKDILTLRLDVMETEN
RLHFTIKDPANRRYEVPLET PRVHSRAPS PLYSVEFSEEPFGVIVRRQLDGRVLLNTTV
APLFFADQFLQLSTSLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTGANLYGSH
PFYLALEDGSSAHGVFLNNSNAMDVVLQPSALSWSRSTGGILDVYIFLGPEPKSVVQQY
LDVVGYPFMPYWG LGFHL CRWGYSSTAITRQVENMTRAHFPLDVQWNDLDYMDSRD
FTFNKDGFRDFPAMVQELHQGRRYMMIVDPAISSSGPAGSYRPHYDEGLRRGVFITNET
GQPLIGKVWPGSTAFPDTNPTALAWWEDMVAEFHDQVPFDGMWIDMNEPSNFIRGSED
GCPNNELENPPYVPGVVGGTLQAATICASSHQFLSTHYNLHNLGLTEAIAASHRALVKA
RGTRPFVISRSTFAGHGRYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGADVCGF
LGNTSEELCVRWTQLGAFYPFMRNHNSLLSLPQEPYSFSEPAQQAMRKALTLRYALLPH
LYTLFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVLQAGKAEVTGYF
PLGTWYDLQTVPIEALGSLPPPPAAPREPAIHSEGQWVTLPAPLDTINVHLRAGYI I PL
QGPGLTTTESRQQPMALAVALT KGGEARGELFWDDGESLEVLERGAYTQVIFLARNNTI
VNELVRVTSEGAGLQLQKVTVLGVATAPQQVLSNGVPVSNFTYSPDTKVLDICVSLLMG
EQFLVSWC

La CDS nativa della GAA umana è stata inoltre analizzata con l'ausilio di tre diversi software disponibili in rete (NetGene, GenScan e GeneSplicer) per l'identificazione di sequenze che potevano rappresentare siti criptici di splicing intronico. Qui di seguito vengono riportati i risultati ottenuti con i tre programmi.

1. Analisi con NetGene, selezionando come organismo modello *Arabidopsis thaliana*: il programma non ha individuato alcun introne criptico:

The sequence: GAA has the following composition:
Length: 2859 nucleotides.

```
17.8% A, 35.0% C, 29.6% G, 17.6% T, 0.0% X, 64.6%
G+C

Donor splice sites, direct strand
-----
No donor site predictions above threshold.

Donor splice sites, complement strand
-----
No donor site predictions above threshold.

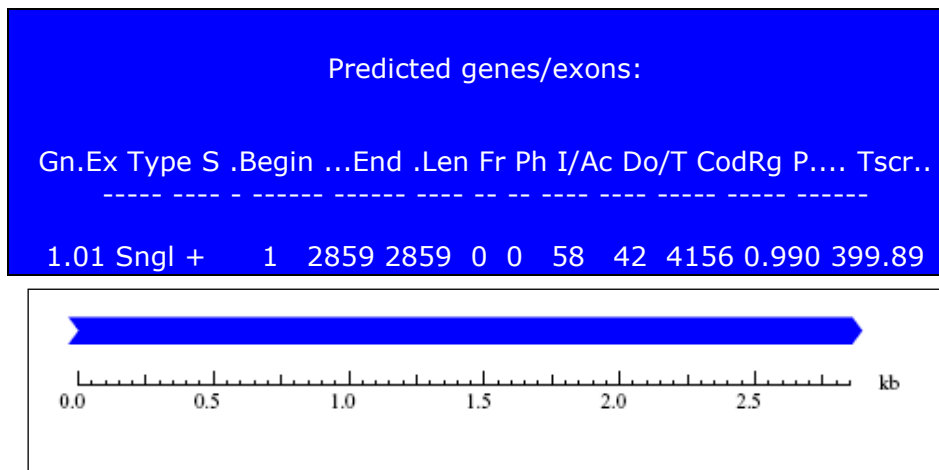
Acceptor splice sites, direct strand
-----
No acceptor site predictions above threshold.

Acceptor splice sites, complement strand
-----
No acceptor site predictions above threshold.

Branch points, direct strand
-----
No branch point predictions above threshold.

Branch points, complement strand
-----
No branch point predictions above threshold.
```

2. Analisi con GenScan, selezionando come organismo modello *A. thaliana*:
il programma non ha identificato siti critici di *splicing*:



3. Analisi con GeneSplicer selezionando come organismo modello *O. sativa*:
il programma ha previsto un possibile introne criptico:

Your sequence has 2859 bp.

Organism : O. sativa (Rice)					
acc_Sensitivity (%)		98	acc_threshold		-0.221536
don_Sensitivity (%)		98	don_threshold		0.811606
Da=60, Dd=60					
End5	End3	Score	Confidence	Splice_site_type	
-----	-----	-----	-----	-----	-----
1438	1439	1.641898	Medium	donor	
1579	1580	2.225420	Medium	acceptor	

GeneSplicer è l'unico software che permette di selezionare *O. sativa* come organismo modello; poiché il meccanismo di *splicing* intronico presenta una variabilità significativa fra le specie (Wang e Brendel, 2006), l'analisi con tale programma deve essere ritenuta maggiormente affidabile.

Considerati i risultati ottenuti si è deciso di procedere alla riscrittura della CDS applicando le regole del *codon context* messe a punto dal nostro gruppo di ricerca al fine di ovviare ai problemi appena descritti e migliorare l'espressione endosperma-specifica.

Prima bozza della CDS di GAA (lunghezza di 2859 bp)

ATGGGCGTGCGGCACCCTCCCTGCTCCCACCGGCTGCTGGCCGTGTGCGCGCTGGTGTCC
 CCTCGCCACCGCCGCGCTGCTGGGCCACATCCTGCTGCATGACTTCTGCTGGTGCCTC
 GGGAGCTGTCCGGCTCCTCCCCGCTGCTGGAGGAGACCACCCGGCGCACCAGCAGGGC
 GCGTCCCGGCCGGGCCCTCGGGATGCGCAGGCGCACCCGGGCCGGCCTCGGGCCGTGCC
 CACCCAGTGCATGTGCTCCCAACTCCCGGTTGACTGCGCGCCGGACAAAGCCATCA
 CCCAGGAGCAGTGCAGGGCGCGGGCTGCTGCTACATCCCGGCCAAGCAGGGCCTGCAG
 GCGCGCAGATGGGCCAGCCCTGGTGTCTTCTCCCTCCCTCCTACCCCTCCTACAAGCT
 GGAGAACCTGTCTCCTCCGAGATGGGCTACACCGCCACCCTCACCCGGACCACCCCA
 CCTTCTTCCCCAAAGACATCCTCACCCCTGCGGCTGGATGTGATGATGGAGACCGAGAAC
 CGGCTGCATTTACCATCAAAGACCCGGCCAACCGGCGGTATGAGGTGCCTCTGGAGAC
 CCCTCGGGTGCATTTCCCGGGCGCCCTCCCCCTCTGTACTCCGTTGAGTTCTCCGAGGAGC
 CCTTCGGCGTGATTGTGCGGCGGCAGCTGGATGGCCGGGTGCTGCTCAACACCACCGTT
 GCGCCTCTGTTCTTCGCCGACCAGTTCTGCAGCTGTCCACCTCCCTGCCCTCCAGTA
 CATCACCGCCCTGGCCGAGCACCTGTCCCCCTCATGCTGTCCACCTCCTGGACCCGGA
 TCACCCTGTGGAACCGGGACCTGGCGCCCACCCCGGGCGCAACCTGTATGGCTCCCA
 CCCCTTCTACCTGGCGCTGGAGGATGGCGGCTCCGCGCATGGCGTGTTCCTGCTCAACTC
 CAATGCCATGGATGTTGTGCTGCAGCCCTCCCCGGCGTGTCTGGCGGTCCACCGGCG
 GCATCCTGGATGTGTACATCTTCCCTGGGCCCGGAGCCGAAGTCCGTTGTGCAGCAGTAC
 CTGGATGTTGTTGGCTACCCCTTCATGCCTCCCTACTGGGCTGGGCTTCCACCTGTG
 CCGGTGGGCTACTCCTCCACCGCCATCACCCGGCAGGTTGTTGAGAACATGACCCGGG
 CGCATTTCCCTCTGGATGTGCAGTGAATGACCTGGACTACATGGACTCCCGGCGGGAC
 TTCACCTTCAACAAAGATGGCTTCCGGGACTTCCCGGCCATGGTGCAGGAGCTGCACCA
 GGGCGGCCGGCGGTACATGATGATTGTTGACCCGGCCATCTCCTCCTCCGGCCCGGCCG

GCTCCTACCGGCCCTATGATGAGGGCCTGCGGC GGGG CGTGTTTCATCACCAATGAGACC
GGCCAG CCCC TCATTGGCAAAGTGTGGCCGGGCTCCACCGCGTTCCCGGACTTCACCAA
CCCC ACCGCGCTGGCGTGGTGGGAGGACATGGTTGCCGAGTTCATGACCAGGTGCCCT
TCGATGGCATGTGGATTGACATGAATGAGCCCTCCAAC TTCATCC GGGG CTCCGAGGAT
GGCTG CCCC AACAATGAGCTGGAGAACCCTCCCTATGTGCCGGGCGTTGTTGGCGGCAC
CCTGCAGGCCGCCACCATCTGCGCGTCTCCACCAGTTCCTGTCCACCCATTACAACC
TGCACAACCTGTATGGCCTCACCGAGGCCATTGCGTCCCACCGGGCGCTGGTGAAAGCG
C GGGG CACCCGGCCCTTCGTGATCTCCCGGTCCACCTTCGCCGGCCATGGCCGGTATGC
CGGCCATTGGACCGGCGATGTGTGGTCTCCTGGGAGCAGCTGGCGTCTCCGTGCCGG
AGATCCTGCAGTTCAACCTGCTGGGCGTGCCTCTGGTTGGCGCCGATGTGTGCCGGCTTC
CTGGGCAACACCTCCGAGGAGCTGTGCGTGGGTGGACCCAGCTGGGCGCGTTCTA CCC
CTTCATGCGGAACCACAACCTCCCTGCTGTCCCTGCCTCAGGAGCCCTACTCCTTCTCCG
AGCCGGCGCAGCAGGCCATGCGGAAGGCGCTCACCTGCGGTACGCGTGTGCCTCAC
CTGTACACCCTGTTCCACCAGGCGCATGTTGCCGGCGAGACCGTTGCGCGGCCCTCTGTT
CCTGGAGTT CCCC AAAGACTCCTCCACCTGGACCGTTGACCACCAGCTGCTGT GGGG CG
AGGCGCTGCTCATCA CCCC GGTGCTGCAGGCCGGCAAAGCCGAGGTGACCGGCTACTTC
CCTCTGGGCACCTGGTATGACCTGCAGACCGTGCCATTGAGGCGCTGGGCTCCCTGCC
TCCTCCTCCGGCCGCGCTCGGGAGCCGGCCATCCATTCCGAGGGCCAGTGGGTGACCC
TGCCGGCGCCTCTGGACACCATCAATGTGCACCTGCCGGCCGGCTACATCATCCCTCTG
CAGGGCCCGGCCCTCACCACCAGGAGTCCCGGCAGCAGCCCATGGCGCTGGCCGTTGC
GCTCACCAAAGGCGGCGAGGCGC GGGG CGAGCTGTTCTGGGATGATGGCGAGTCCCTGG
AGGTGCTGGAGC GGGG CGCGTACACCCAGGTGATCTTCTGGCGCGGAACAACACCAT
GTGAATGAGCTGGTCCGGGTGACCTCCGAGGGCGCCGGCCTGCAGCTGCAGAAAGTGAC
CGTGCTGGGCGTTGCCACCGCGCCTCAGCAGGTGCTGTCCAATGGCGTGCCGGTGTCCA
ACTTCACCTACT CCCC GGACACCAAAGTGCTGGACATCTGCGTGTCCCTGCTCATGGGC
GAGCAGTTCCTGGTGTCTGGTGCTAG

Successivamente, la sequenza è stata sottoposta a un'ulteriore modifica con eliminazione degli omotetrameri attraverso l'impiego di codoni scelti ancora in base al contesto codonico risultante.

Seconda bozza della CDS di GAA:

ATGGGCGTGCGGCACCCTCCCTGCTCCCACCGGCTGCTGGCCGTGTGCGCGCTGGTGTG
CCTCGCCACCGCCGCGCTGCTGGGCCACATCCTGCTGCATGACTTCCTGCTGGTGCCTC
GGGAGCTGTCCGGCTCCAGCCCGGTGCTGGAGGAGACCCACCGGCGCACCAGCAGGGC
GCGTCCCGGCCGGGCCCTCGGGATGCGCAGGCGCACCCGGGCCGGCCTCGGGCCGTGCC
CACCCAGTGCATGTGCTCCCAACTCCCGGTTGACTGCGCGCCGGACAAAGCCATCA
CCCAGGAGCAGTGCAGGGCGCGTGGCTGCTGCTACATCCCGGCCAAGCAGGGCCTGCAG
GGCGCGCAGATGGGCCAGCCCTGGTGTCTTCCCTCCCTCCTATCCCTCCTACAAGCT
GGAGAACCTGTCTCCTCCGAGATGGGCTACACCGCCACCCTCACCCGGACCACGCCCA
CCTTCTTTCCCAAAGACATCTCACCCCTGCGGCTGGATGTGATGATGGAGACCGAGAAC
CGGCTGCATTTACCATCAAAGACCCGGCCAACCGGCGGTATGAGGTGCCTCTGGAGAC
GCCTCGGGTGCATTCGCGGGCGCCAGCCCTCTGTACTCCGTTGAGTTCTCCGAGGAGC
CCTTCGGCGTGATTGTGCGGCGGCAGCTGGATGGCCGGGTGCTGCTCAACACCACCGTT
GCGCCTCTGTTCTTCGCCGACCAGTTCCTGCAGCTGTCCACCTCCCTGCCCTCCAGTA
CATCACCGGCCTGGCCGAGCACCTGAGCCCTCTCATGCTGTCCACCTCCTGGACCCGGA
TCACCCTGTGGAACCGGGACCTGGCGCCACGCCGGGCGCCAACCTGTATGGCTCCCAT
CCCTTCTACCTGGCGCTGGAGGATGGCGGCTCCGCGCATGGCGTGTTCCTGCTCAACTC

CAATGCCATGGATGTTGTGCTGCAGCCCAGCCCGGCGCTGTCCTGGCGGTCCACCGGCG
GCATCCTGGATGTGTACATCTTCCTGGGCCCCGAGCCGAAGTCCGTTGTGCAGCAGTAC
CTGGATGTTGTTGGCTATCCCTTCATGCCTCCCTACTGGGGCCTGGGCTTCCACCTGTG
CCGGTGGGCTACTCCTCCACCGCCATCACCCGGCAGGTTGTTGAGAACATGACCCGGG
CGCATTTCCCTCTGGATGTGCAGTGGAATGACCTGGACTACATGGACTCCCGGCGGGAC
TTCACCTTCAACAAAGATGGCTTCCGGGACTTCCCGGCCATGGTGCAGGAGCTGCACCA
GGGCGGCCCGGCGGTACATGATGATTGTTGACCCGGCCATCTCCTCCTCCGGCCCCGGCCG
GCTCCTACCGGCCCTATGATGAGGGCCTGCGGCGTGGCGTGTTCATCACCAATGAGACC
GGCCAGCCTCTCATTTGGCAAAGTGTGGCCGGGCTCCACCGCGTTCCCGGACTTACCAA
TCCACCGCGCTGGCGTGGTGGGAGGACATGGTTGCCGAGTTCATGACCAGGTGCCCT
TCGATGGCATGTGGATTGACATGAATGAGCCCTCCAACCTTCATCCGTGGCTCCGAGGAT
GGCTGTCCCAACAATGAGCTGGAGAACCCTCCCTATGTGCCGGGCGTTGTTGGCGGCAC
CCTGCAGGCCGCCACCATCTGCGCGTCTCCACCAGTTCCTGTCCACCCATTACAACC
TGCACAACCTGTATGGCCTCACCGAGGCCATTGCGTCCACCGGGCGCTGGTGAAAGCG
CGTGGCACCCGGCCCTTCGTGATCTCCCGGTCCACCTTCGCCGGCCATGGCCGGTATGC
CGGCCATTGGACCGGCGATGTGTGGTCTCCTGGGAGCAGCTGGCGTCTCCGTGCCGG
AGATCCTGCAGTTCAACCTGCTGGGCGTGCCTCTGGTTGGCGCCGATGTGTGCCGCTTC
CTGGGCAACACCTCCGAGGAGCTGTGCGTGGGTGGACCCAGCTGGGCGCGTTCTATCC
CTTCATGCGGAACCACAACCTCCCTGCTGTCCCTGCCTCAGGAGCCCTACTCCTTCTCCG
AGCCGGCGCAGCAGGCCATGCGGAAGGCGCTCACCTGCGGTACGCGCTGCTGCCTCAC
CTGTACACCCTGTTCACCAGGCGCATGTTGCCGGCGAGACCGTTGCGCGGCCCTGTGT
CCTGGAGTTTCCCAAAGACTCCTCCACCTGGACCGTTGACCACCAGCTGCTGTGGGCG
AGGCGCTGCTCATCACGCCGGTGCTGCAGGCCGGCAAAGCCGAGGTGACCGGCTACTTC
CCTCTGGGCACCTGGTATGACCTGCAGACCGTGCCCATGAGGGCGCTGGGCTCCCTGCC
TCCTCCTCCGGCCGCGCCTCGGGAGCCGGCCATCCATTCCGAGGGCCAGTGGGTGACCC
TGCCGGCGCCTCTGGACACCATCAATGTGCACCTGCGGGCCGGCTACATCATCCCTCTG
CAGGGCCCCGGCCCTCACACCACCGAGTCCCGGCAGCAGCCCATGGCGCTGGCCGTTGC
GCTCACCAAAGGCGGCGAGGCGCGTGGCGAGCTGTTCTGGGATGATGGCGAGTCCCTGG
AGGTGCTGGAGCGTGGCGCGTACACCCAGGTGATCTTCCCTGGCGCGGAACAACACCAT
GTGAATGAGCTGGTCCGGGTGACCTCCGAGGGCGCCGGCCTGCAGCTGCAGAAAGTGAC
CGTGCTGGGCGTTGCCACCGCGCCTCAGCAGGTGCTGTCCAATGGCGTGCCGGTGTCCA
ACTTACCTACAGCCCGGACACCAAAGTGCTGGACATCTGCGTGTCCCTGCTCATGGGC
GAGCAGTTCCTGGTGTCTGGTGCTAG

Va rilevato che i tre omotetrameri GGGG evidenziati in giallo sono ineliminabili poiché si formano dall'unione del codone codificante l'amminoacido triptofano (TGG) con la radice codonica corrispondente all'amminoacido glicina (GGC).

Per favorire l'espressione in riso, la sequenza codificante il peptide segnale (PS) umano nativo di GAA è stata sostituita con quella relativa al peptide navetta della glutelina 4 di riso (PS Glub4), anch'essa riscritta secondo il metodo del codone contesto.

PS nativo (81 bp)

ATGGGAGTGAGGCACCCGCCCTGCTCCACCGGCTCCTGGCCGTCTGCGCCCTCGTGTC
CTTGGCAACCGCTGCACTCCTG

PS GluB4 ottimizzato per riso (72 bp)

ATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCTGTGCCA
CGGCTCCATGGCC

Terza bozza della CDS di GAA ottimizzata per l'espressione in endosperma di
riso (in blu la sequenza relativa a PS Glub4) (lunghezza 2850 bp):

ATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCTGTGCCA
CGGCTCCATGGCCGGCCACATCCTGCTGCATGACTTCCTGCTGGTGCCTCGGGAGCTGT
CCGGCTCCAGCCCAGTGTGCTGGAGGAGACCCACCCGGCGCACCAGCAGGGCGCGTCCCGG
CCGGGCCCTCGGGATGCGCAGGCGCACCCGGGCCGGCTCGGGCCGTGCCACCCAGTG
CGATGTGCCTCCCAACTCCCGGTTGACTGCGCGCCGGACAAAGCCATCACCCAGGAGC
AGTGCAGGGCGCGTGGCTGCTGCTACATCCCGGCCAAGCAGGGCCTGCAGGGCGCGCAG
ATGGGCCAGCCCTGGTGTCTTCCCTCCCTCCTATCCCTCCTACAAGCTGGAGAACCT
GTCTCCTCCGAGATGGGCTACACCGCCACCCTCACCCGGACCACGCCACCTTCTTTC
CCAAAGACATCCTCACCTGCGGCTGGATGTGATGATGGAGACCAGAAACCGGCTGCAT
TTCACCATCAAAGACCCGGCCAACCGGCGGTATGAGGTGCCTCTGGAGACGCCTCGGGT
GCATTCCCGGGCGCCAGCCCTCTGTACTCCGTTGAGTTCTCCGAGGAGCCCTTCGGCG
TGATTGTGCGGCGGCAGCTGGATGGCCGGGTGCTGCTCAACACCACCGTTGCGCCTCTG
TTCTTCGCGGACCAGTTCCCTGCAGCTGTCCACCTCCCTGCCCTCCCAGTACATCACCGG
CCTGGCCGAGCACCTGAGCCCTCTCATGCTGTCCACCTCCTGGACCCGGATCACCTGT
GGAACCGGGACCTGGCGCCACGCCGGGCGCCAACCTGTATGGCTCCCATCCCTTCTAC
CTGGCGCTGGAGGATGGCGGCTCCGCGCATGGCGTGTTCCTGCTCAACTCCAATGCCAT
GGATGTTGTGCTGCAGCCCAGCCGGCGCTGTCCCTGGCGGTCCACCGGCGGCATCCTGG
ATGTGTACATCTTCCCTGGGCCCCGAGCCGAAGTCCGTTGTGCAGCAGTACCTGGATGTT
GTTGGCTATCCCTTCATGCCTCCCTACTGGGGCCTGGGCTTCCACCTGTGCCGGTGGGG
CTACTCCTCCACCGCCATCACCCGGCAGGTTGTTGAGAACATGACCCGGGCGCATTTCC
CTCTGGATGTGCAGTGGAAATGACCTGGACTACATGGACTCCCGGCGGGACTTCACCTTC
AACAAAGATGGCTTCCGGGACTTCCCGGCCATGGTGCAGGAGCTGCACCAGGGCGGCCG
GCGGTACATGATGATTGTTGACCCGGCCATCTCCTCCTCCGGCCCCGGCGGCTCCTACC
GGCCCTATGATGAGGGCCTGCGGCGTGGCGTGTTCATCACCAATGAGACCGGCCAGCCT
CTCATTGGCAAAGTGTGGCCGGGCTCCACCGCGTTCCCGGACTTCACCAATCCCACCGC
GCTGGCGTGGTGGGAGGACATGGTTGCCGAGTTCCATGACCAGGTGCCCTTCGATGGCA
TGTGGATTGACATGAATGAGCCCTCCAATTCATCCGTGGCTCCGAGGATGGCTGTCCC
AACAAATGAGCTGGAGAACCCTCCCTATGTGCCGGGCGTTGTTGGCGGCACCCTGCAGGC
CGCCACCATCTGCGCGTCCCTCCCACCAGTTCCCTGTCCACCATTACAACCTGCACAACC
TGTATGGCCTCACCGAGGCCATTGCGTCCCACCGGGCGCTGGTGAAGCGCGTGGCACC
CGGCCCTTCGTGATCTCCCGGTCCACCTTCGCCGGCCATGGCCGGTATGCCGGCCATTG
GACCGGCGATGTGTGGTCCCTCCTGGGAGCAGCTGGCGTCCCTCCGTGCCGGAGATCCTGC
AGTTCAACCTGCTGGGCGTGCCTCTGGTTGGCGCCGATGTGTGCGGCTTCCTGGGCAAC
ACCTCCGAGGAGCTGTGCGTGCAGTGGACCCAGCTGGGCGCGTTCTATCCCTTCATGCG
GAACCACAACCTCCCTGCTGTCCCTGCCTCAGGAGCCCTACTCCTTCTCCGAGCCGGCGC
AGCAGGCCATGCGGAAGGCGCTCACCCCTGCGGTACGCGCTGCTGCCTCACCTGTACACC
CTGTTCCACCAGGCGCATGTTGCCGGCGAGACCGTTGCGCGGCCTCTGTTCCCTGGAGTT
TCCCAAAGACTCCTCCACCTGGACCGTTGACCACCAGCTGCTGTGGGGCGAGGCGCTGC
TCATCACGCCGGTGTGTCAGGCCGGCAAAGCCGAGGTGACCGGCTACTTCCCTCTGGGC
ACCTGGTATGACCTGCAGACCGTGCCATTGAGGCGCTGGGCTCCCTGCCTCCTCCTCC
GGCCGCGCCTCGGGAGCCGGCCATCCATTCCGAGGGCCAGTGGGTGACCCTGCCGGCGC
CTCTGGACACCATCAATGTGCACCTGCGGGCCGGCTACATCATCCCTCTGCAGGGCCCCG
GGCCTCACACCACCGAGTCCCGGCAGCAGCCCATGGCGCTGGCCGTTGCGCTCACCAA

AGGCGGCGAGGCGCGTGGCGAGCTGTTCTGGGATGATGGCGAGTCCCTGGAGGTGCTGG
AGCGTGGCGCGTACACCCAGGTGATCTTCCCTGGCGCGGAACAACACCATTGTGAATGAG
CTGGTCCGGGTGACCTCCGAGGGCGCCGGCCTGCAGCTGCAGAAAGTGACCGTGCTGGG
CGTTGCCACCGCGCCTCAGCAGGTGCTGTCCAATGGCGTGCCGGTGTCCAACCTCACCT
ACAGCCCCGACACCAAAGTGCTGGACATCTGCGTGTCCCTGCTCATGGGCGAGCAGTTC
CTGGTGTCCCTGGTGCTAG

Traduzione:

MATIAFSRLSIYFCVLLLLCHGSMAGHILLHDFLLVPRELSGSSPVLEETHPAHQGASR
PGPRDAQAHPRPRAVPTQCDVPPNSRFDCAFDKAITQEQCEARGCCYIPAKQGLQGAQ
MGQFWCFFPPSYPSYKLENLSSSEMGYTATLTRTPTFFPKDILTLRLDVMETENRLH
FTIKDPANRRYEVPLETPRVHSRAPSPLYSVEFSEEPFGVIVRRQLDGRVLLNNTVAPL
FFADQFLQLSTSLPSQYITGLAEHLSPLMLSTSWTRITLWNRDLAPTGANLYGSHPFY
LALEDGGSAGHVFLNSNAMDVVLQPSPALSWRSTGGILDVYIFLGPEPKSVVQQYLDV
VGYPFMPYWGGLGFHLCRWGYSSAITRQVVENMTRAHFPLDVQWNLDYMDSRRDFTF
NKDGRFRDFPAMVQELHQGRRYMMIVDPAISSSGPAGSYRKYDEGLRRGVFITNETGQP
LIGKVPWGSTAFPDTNPTALAWWEDMVAEFHDQVPPFDGMWIDMNEPSNFIRGSEDGCP
NNELENPPYVPGVVGGLQAATICASSHQFLSTHYNLHNLGLTEAIAASHRALVKARGT
RPFVISRSTFAGHGRYAGHWTGDVWSSWEQLASSVPEILQFNLLGVPLVGADVCGFLGN
TSEELCVRWTQLGAFYPFMRNHNLSLSPQEPYSFSEPAQQAMRKALTRYALLPHLYT
LFHQAHVAGETVARPLFLEFPKDSSTWTVDHQLLWGEALLITPVLQAGKAEVTGYFPLG
TWYDLQTVPIEALGSLPPPPAAPREPAIHSEGWVTLPLAPLDTINVHLRAGYIIPLQGP
GLTTTESRQQPMALAVALTGGGEARGELFWDGSESLEVLERGAYTQVIFLARNNTIVNE
LVRVTSEGAGLQLQKVTVLGVATAPQQVLSNGVPVSNFTYSPDTKVLDICVSLLMGEQF
LVSWC

La sequenza riscritta e modificata come descritto è stata sottoposta ad analisi con uno specifico software (Webcutter), capace di creare siti di restrizione attraverso mutazioni puntiformi sinonime. È stato così possibile inserire due siti di restrizione in corrispondenza del termine del primo e secondo terzo della CDS:

1. *Bam*H I (G/GATCC) (904-909 bp);
2. *Kpn* I (GGTAC/C) (1872-1877 bp).

Sono stati inoltre introdotti due ulteriori siti di restrizione utili alla sub-clonazione del gene nel vettore di espressione binario. Tali siti sono rappresentati da *Xba* I (T/CTAGA) al terminale 5' della CDS e *Sac* I (GAGCT/C) al terminale 3' della stessa.

Versione finale della CDS di GAA:

TCTAGAATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCT
GTGCCACGGCTCCATGGCCGGCCACATCCTGCTGCATGACTTCCCTGCTGGTGCCTCGGG
AGCTGTCCGGCTCCAGCCCGGTGCTGGAGGAGACCCACCCGGCGCACCAGCAGGGCGCG
TCCCGGCCGGGCCCTCGGGATGCGCAGGCGCACCCGGGCCGGCCTCGGGCCGTGCCAC
CCAGTGCATGTGCCTCCCAACTCCCGGTTGACTGCGCGCCGGACAAAGCCATCACCC

AGGAGCAGTGCGAGGCGCGTGGCTGCTGCTACATCCCGGCCAAGCAGGGCCTGCAGGGC
GCGCAGATGGGCCAGCCCTGGTGCTTCTTCCCTCCCTCCTATCCCTCCTACAAGCTGGA
GAACCTGTCTCTCCGAGATGGGCTACACCGCCACCCTCACCCGGACCACGCCACCT
TCTTTCCCAAAGACATCCTCACCCCTGCGGCTGGATGTGATGATGGAGACCAGAAACCGG
CTGCATTTACCCATCAAAGACCCGGCCAACCGGGCGGTATGAGGTGCCTCTGGAGACGCC
TCGGGTGCATTCCTGGGCGCCAGCCCTCTGTACTCCGTTGAGTTCTCCGAGGAGCCCT
TCGGCGTGATTGTGCGGCGGCAGCTGGATGGCCGGGTGCTGCTCAACACCACCGTTGCG
CCTCTGTTCTTCGCCGACCAGTTCCTGCAGCTGTCCACCTCCCTGCCCTCCAGTACAT
CACCGCCCTGGCCGAGCACCTGAGCCCTCTCATGCTGTCCACCTCCTGGACCCGGATCA
CCCTGTGGAACCGGGACCTGGCGCCACGCCGGGCGCCAACCTGTATGGCTCCCATCCC
TTCTACCTGGCGCTGGAGGATGGC**GGATCC**GCGCATGGCGTGTTCCTGCTCAACTCCAA
TGCCATGGATGTTGTGCTGCAGCCCAGCCCGGCGCTGTCTGGCGGTCCACCGGCGGCA
TCCTGGATGTGTACATCTTCCCTGGGCCCGGAGCCGAAGTCCGTTGTGCAGCAGTACCTG
GATGTTGTTGGCTATCCCTTCATGCCTCCCTACTGGGGCCTGGGCTTCCACCTGTGCCG
GTGGGGCTACTCCTCCACCGCCATCACCCGGCAGGTTGTTGAGAACATGACCCGGGCGC
ATTTCCCTCTGGATGTGCAGTGAATGACCTGGACTACATGGACTCCCGGCGGGACTTC
ACCTTCAACAAGATGGCTTCCGGGACTTCCCGGCCATGGTGCAGGAGCTGCACCAGGG
CGGCCGGCGGTACATGATGATTGTTGACCCGGCCATCTCCTCCTCCGGCCCGGCGGCT
CCTACCGGCCCTATGATGAGGGCCTGCGGCGTGGCGTGTTCATACCAATGAGACCGGC
CAGCCTCTCATTGGCAAAGTGTGGCCGGGCTCCACCGCGTTCCCGGACTTCACCAATCC
CACCGCGCTGGCGTGGTGGGAGGACATGGTTGCCGAGTTCATGACCAGGTGCCCTTCG
ATGGCATGTGGATTGACATGAATGAGCCCTCCAACCTTCATCCGTGGCTCCGAGGATGGC
TGTCCCAACAATGAGCTGGAGAACCCTCCCTATGTGCCGGGCGTTGTTGGCGGCACCCT
GCAGGCCCGCCACCATCTGCGCGTCTCCACCAGTTCCTGTCCACCCATTACAACCTGC
ACAACCTGTATGGCCTCACCGAGGCCATGCGTCCCACCGGGCGCTGGTGAAAGCGCGT
GGCACCCGGCCCTTCGTGATCTCCCGGTCCACCTTCGCCGGCCATGGCCGGTATGCCGG
CCATTGGACCGGCGATGTGTGGTCTCCTGGGAGCAGCTGGCGTCTCTC**GGTACC**GGAGA
TCCTGCAGTTCAACCTGCTGGGCGTGCCCTGGTTGGCGCCGATGTGTGCGGCTTCTTG
GGCAACACCTCCGAGGAGCTGTGCGTGCAGGTGGACCCAGCTGGGCGCGTTCTATCCCTT
CATGCGGAACCACAACCTCCCTGCTGTCCCTGCCTCAGGAGCCCTACTCCTTCTCCGAGC
CGGCGCAGCAGGCCATGCGGAAGGCGCTCACCTGCGGTACGCGCTGCTGCCTCACCTG
TACACCCTGTTCCACCAGGCGCATGTTGCCGGCGAGACCCTTGCAGCGCCTCTGTTCT
GGAGTTTCCCAAAGACTCCTCCACCTGGACCGTTGACCACCAGCTGCTGTGGGGCGAGG
CGCTGCTCATCACGCCGGTCTGCAGGCCGGCAAAGCCGAGGTGACCGGCTACTTCCCT
CTGGGCACCTGGTATGACCTGCAGACCGTGCCATTGAGGCGCTGGGCTCCCTGCCTCC
TCCTCCGGCCGCGCCTCGGGAGCCGGCCATCCATTCCGAGGGCCAGTGGGTGACCCTGC
CGGCGCCTCTGGACACCATCAATGTGCACCTGCGGGCCGGCTACATCATCCCTCTGCAG
GGCCCGGGCCTCACACCACCGAGTCCCGGCAGCAGCCATGGCGCTGGCCGTTGCGCT
CACCAAAGGCGGCGAGGCGGTGGCGAGCTGTTCTGGGATGATGGCGAGTCCCTGGAGG
TGCTGGAGCGTGGCGCGTACACCAGGTGATCTTCCCTGGCGCGGAACAACACCATTTGTG
AATGAGCTGGTCCGGGTGACCTCCGAGGGCGCCGGCCTGCAGCTGCAGAAAGTGACCGT
GCTGGGCGTTGCCACCGCGCCTCAGCAGGTGCTGTCCAATGGCGTGCCGGTGTCCAAC
TCACCTACAGCCCGGACACCAAAGTGTGGACATCTGCGTGTCCCTGCTCATGGGCGAG
CAGTTCCTGGTGTCTGGTGCTAG**GAGCTC**

La sequenza è stata nuovamente analizzata con GenScan, NetGene e GeneSplicer per verificare l'assenza di siti criptici di *splicing*:

1. Analisi con NetGene, selezionando come organismo modello *A. thaliana*: il programma non ha individuato alcun possibile introne criptico:

```
The sequence: GA has the following composition:

      Length:      2859 nucleotides.
14.6% A, 35.9% C, 30.3% G, 19.2% T, 0.0% X, 66.2%
                G+C

      Donor splice sites, direct strand
      -----
No donor site predictions above threshold.

      Donor splice sites, complement strand
      -----
No donor site predictions above threshold.

      Acceptor splice sites, direct strand
      -----
No acceptor site predictions above threshold.

      Acceptor splice sites, complement strand
      -----
No acceptor site predictions above threshold.

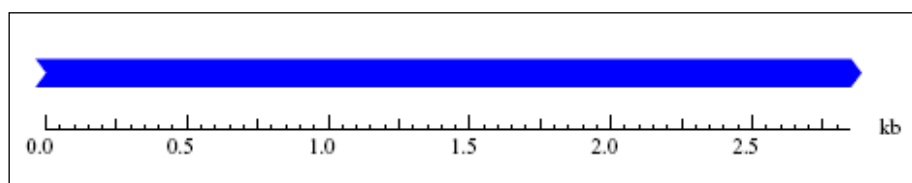
      Branch points, direct strand
      -----
No branch point predictions above threshold.

      Branch points, complement strand
      -----
No branch point predictions above threshold.
```

2. Analisi con GenScan, selezionando come organismo modello *A. thaliana*: il programma non ha individuato possibili siti critici di *splicing*:

```
Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
-----
1.01 Sngl +   1  2850 2850  0  0  77  42  4541 0.987 442.32
```



4. Analisi con GeneSplicer selezionando come organismo modello *O. sativa*:
il programma non ha previsto introni criptici:

```
Your sequence has 2859 bp.

Organism : O. sativa (Rice)

acc_Sensitivity (%):98      acc_threshold :-0.221536
don_Sensitivity (%):98      don_threshold :0.811606
Da=60, Dd=60

      End5|      End3|      Score| Confidence|
          Splice_site_type|
-----|-----|-----|-----|-----
                    -----|
```

Poiché la CDS di GAA è lunga 2862 bp, la sequenza è stata suddivisa in tre frammenti (A, B, C). All'inizio dei frammenti B e C sono stati inseriti rispettivamente i siti di restrizione *Bam*H I e *Kpn* I; va rilevato che questa operazione è stata programmata per consentire l'assemblaggio dei tre frammenti nel vettore pUC18 in modo tale da ricostituire la CDS completa senza alterazioni della cornice di lettura o aggiunta di amminoacidi estranei.

Le sequenze corrispondenti ai tre frammenti ottenuti *in silico* sono state successivamente sintetizzate e clonate nel vettore pUC18 dalla ditta GeneArt.

3.1.1 Analisi della sequenza

E' stata eseguita l'analisi dell'intera sequenza codificante sia la GAA nativa, sia quella ottenuta applicando le regole del *codon context*. I risultati ottenuti sono riportati in Tab. 3.1.

ANALISI	GAA NATIVA	GAA RISCRISSA
% GC totale	64.60	66.17
di cui:		
% GC in terza posizione	27.91	29.10
% codoni terminanti in G o C	84.36	87.30
% TA totale	35.40	33.83
CpG totale	149	214
TpA totale	45	32
TATA	1	0
AATAAA	0	0
ATTTA	0	0
AAA	5	11
AAAA	0	0
TTT	9	4
TTTT	1	0
CCC	93	86
CCCC	30	0
GGG	62	48
GGGG	21	3
Intercodoni CG	72	93

Tabella 3.1. Risultati ottenuti attraverso l'analisi della sequenza GAA nativa e sintetica.

L'omologia complessiva tra le due sequenze è risultata pari all'83%.

3.1.2 Assemblaggio dei frammenti in pUC18

Come descritto al paragrafo 2.6.1, la sequenza intera della GAA è stata ottenuta mediante una serie di digestioni enzimatiche e reazioni di saldatura dei vettori pUC18 portanti i tre diversi frammenti sintetizzati. La strategia seguita ha previsto il taglio del plasmide pUC18_frammento C per aprire il vettore e consentire dapprima l'inserimento del frammento B, successivamente, di quello A.

Dalla digestione di pUC18_frammento C e di pUC18_frammento B con *Aat* II e *Kpn* I, si sono ottenute le bande corrispondenti all'inserito B (1472 bp) e al vettore aperto pUC18_C (3152 bp). Successivamente alla ligation e alla trasformazione di cellule JM101 di *E. coli*, è stata eseguita la miniprep di due colonie e il saggio di positività mediante le seguenti digestioni enzimatiche:

- digestione singola con *Kpn* I per linearizzare il plasmide;
- digestione doppia con *Aat* II e *Kpn* I di pUC18_B+C per verificare la presenza dell'inserito portante il frammento B.

La digestione singola ha prodotto una banda di 4618 bp (Fig. 3.1); l'analisi elettroforetica dei prodotti di digestione doppia ha evidenziato una banda solo parzialmente digerita corrispondente alla lunghezza del plasmide linearizzato, una banda di 3152 bp corrispondente al vettore aperto portante il frammento B e una terza banda di 1472 bp corrispondente all'inserto.

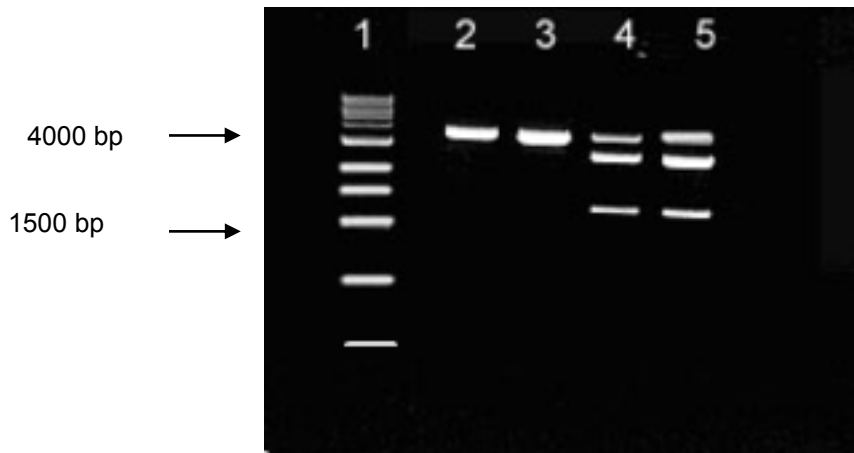
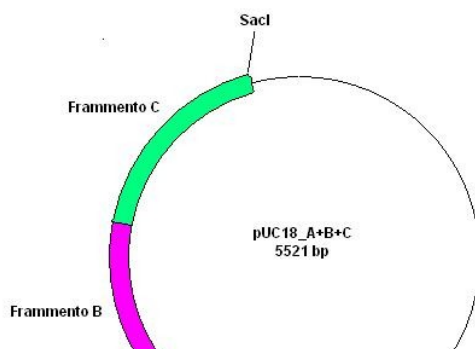


Figura 3.1 Corsia 1: Ladder 1 kb (NEB); 2, 3: colonie 1, 2 linearizzate con *Kpn* I; 4, 5: colonie 1, 2 digerite con *Aat* II e *Kpn* I.

Per l'inserimento del frammento A all'interno del vettore pUC18_B+C, il plasmide pUC18_A di 3589 bp e il vettore pUC18_B+C di 4618 bp sono stati entrambi digeriti con gli enzimi di restrizione *Bam*H I e *Aat* II, ottenendo le bande corrispondenti all'inserto A (1407 bp) e al vettore aperto pUC18_B+C (4124 bp) che sono state sottoposte a saldatura.

Per alcune delle colonie JM101 trasformate con il prodotto di *ligation* si è proceduto all'estrazione miniprep. Queste sono state analizzate mediante digestione di controllo con *Xba* I per linearizzare il plasmide allo scopo di verificare se l'inserto si fosse realmente legato al vettore pUC18_B+C (Fig. 3.2 B). Nella foto riportata si osserva un'unica banda di 5521 bp corrispondente alla lunghezza del vettore pUC18_A+B+C (Fig. 3.2 A).



3.2 Realizzazione dei vettori di espressione per GAA

3.2.1 Isolamento del gene PMI

Il gene *manA* codificante la fosfomannosio isomerasi è stato isolato da *E. coli* mediante PCR con primer (PMI for e PMI rev) disegnati sulla sequenza depositata in banca dati (GenBank acc. n° M15380). La PCR ha fornito l'amplificato atteso di 1176 pb (Fig. 3.3A) che è stato quindi clonato in pGEM-T (Fig.3.3B). Per favorire le successive operazioni di clonazione i primer PMI for e PMI rev sono stati programmati per portare il sito restrizione *Xho* I ai terminali 5' e 3', rispettivamente.

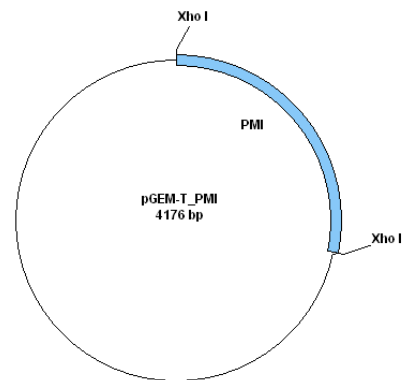
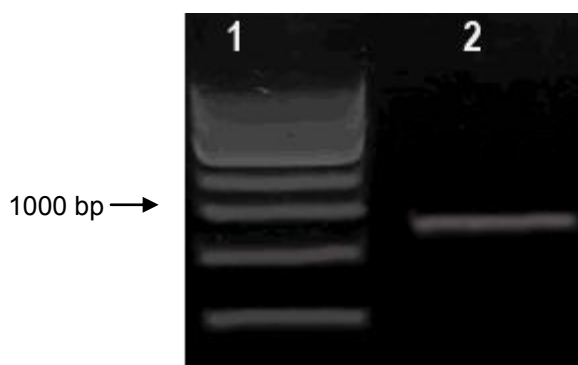


Figura 3.3A Corsia 1: Ladder 1 kb (NEB);

Il vet 2: amplificato PMI.

Figura 3.3B Rappresentazione

schematica del plasmide pGEM-

T_PMI.

1101 e

sottoposto a PCR di controllo su colonia con .

3.4 mostra gli amplificati ottenuti utilizzando i primer M13 che presentano la lunghezza attesa di 1424 bp.

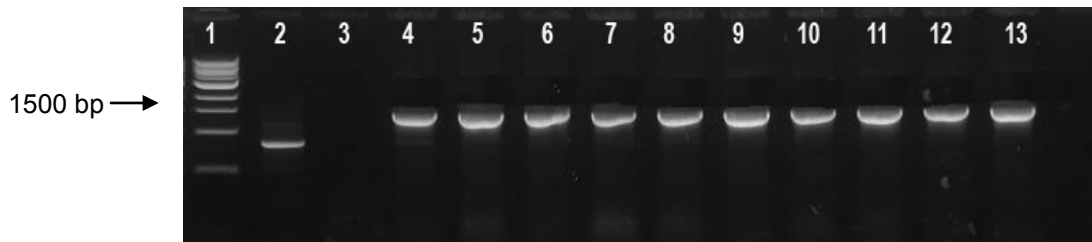


Figura 3.4 Corsia 1: Ladder 1 kb (NEB); 2: CP; 3: acqua; 4-13 colonie saggate in PCR.

Su tre colonie risultate positive all'analisi PCR è stata impostata anche una digestione enzimatica per verificare la correttezza del tratto sintetizzato. In un solo caso la digestione con *Xho* I, che consente la rimozione dell'insero PMI, ha dato i frammenti attesi corrispondenti al tratto sintetizzato e al vettore (3000 bp) (Fig. 3.5).

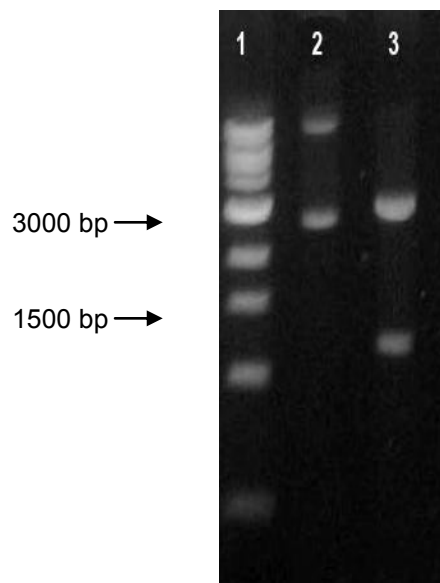


Figura 3.5 Corsia 1: Ladder 1 kb (NEB); 2: campione non digerito; 3: digestione *Xho* I.

La colonia identificata è stata completamente sequenziata ed è risultata perfettamente coincidente a quella depositata in banca dati.

3.2.2 Realizzazione della cassetta per il silenziamento FxAntiFx

Per la realizzazione della cassetta è stato impiegato il vettore pSTARLING (Fig. 3.6) creato da Commonwealth Scientific and Industrial Research Organisation (CSIRO, Australia) per indurre silenziamento genico mediante *hairpin RNAi*.

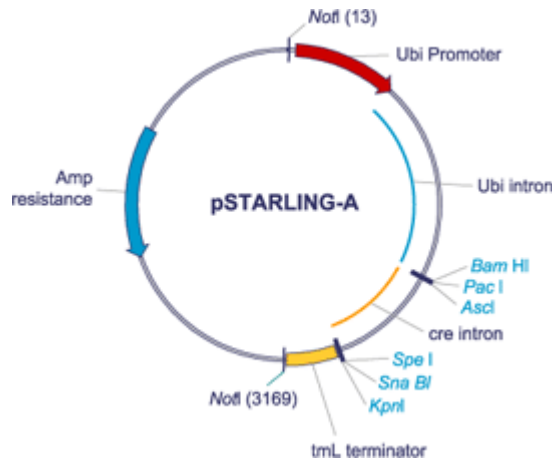


Figura 3.6 Rappresentazione del vettore pSTARLING (CSIRO) con evidenziati gli enzimi di restrizione presenti e gli elementi regolatori dell'espressione: promotore Ubi1 (in rosso) e terminatore tml (in giallo).

La cassetta FxAntiFx è costituita dai seguenti elementi: promotore dell'ubiquitina 1 di mais (Ubi1), sequenze senso dei geni $\alpha(1,3)$ -fucosiltransferasi e $\beta(1,2)$ -xilosiltransferasi di *Oryza sativa*, introne cre, sequenze antisense dei geni fucosiltransferasi e xilosiltransferasi di *O. sativa* e terminatore Tml (Tml ter). Tale cassetta è stata prodotta dall'assemblaggio delle sequenze nucleotidiche relative ai geni da silenziare nel plasmide pSTARLING. Le sequenze senso e antisense dei geni $\alpha(1,3)$ -fucosiltransferasi e $\beta(1,2)$ -xilosiltransferasi sono state isolate singolarmente mediante PCR da cDNA di *O. sativa* impiegando coppie di primer specifici disegnati su un tratto di sequenza altamente conservata dei geni NM_001068555 e NM_001068712 e studiati in modo da introdurre siti di restrizione utili alle operazioni di assemblaggio in pSTARLING. La Fig. 3.7 mostra gli amplificati ottenuti per ciascuna sequenza aventi le lunghezze attese di 606 bp, 614 bp, rispettivamente per le sequenze senso (FucS) e antisense (FucA) del gene fucosiltransferasi, e di 612 bp per ambedue le sequenze relative al gene xilosiltransferasi (XilS, XilA).

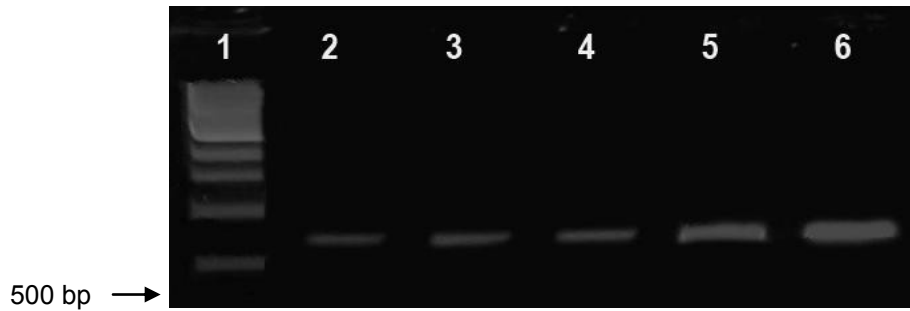


Figura 3.7 Corsia 1: Ladder 1 kb (NEB); 2: amplificato frammento XilS; 3-4: amplificato XilA; 5-6: amplificato FucS.

Gli amplificati sono stati sottoposti a reazione di *A-tailing* e clonati separatamente in pGEM-T; i vettori ingegnerizzati sono stati, quindi, utilizzati nella trasformazione di *E. coli* JM101. Anche in questo caso, il prodotto clonato è stato verificato mediante PCR su colonia con primer M13 for e primer rev specifico per ciascuna sequenza al fine di controllare l'orientamento (Fig. 3.8).

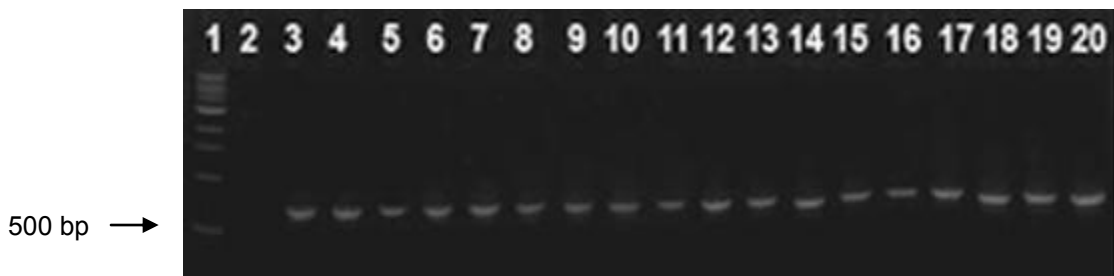


Figura 3.8 Corsia 1: Ladder 1 kb (NEB); 2: acqua; 3-20: colonie analizzate in PCR.

Visti i risultati positivi ottenuti, per ogni costrutto una colonia è stata inviata al sequenziamento. Il tratto sequenziato non ha presentato variazioni rispetto alla sequenza genica programmata.

Lo step successivo ha previsto l'assemblaggio separato delle sequenze senso (FucS-XilS) e antisenso (FucA-XilA) in pGEM-T generando due costrutti (pGEM-T_FucS-XilS e pGEM-T_FucA-XilA). Tale operazione è stata realizzata mediante digestione di pGEM-T_FucS e pGEM-T_FucA con *Xba* I e *Sac* I in modo da aprire i vettori e permettere la saldatura rispettivamente degli inserti XilS e XilA recuperati tramite digestione con *Nhe* I e *Sac* I. In Fig. 3.9 sono mostrati i profili di digestione ottenuti. Entrambe le digestioni hanno dato i frammenti attesi: le

bande a 3600 bp e 3612 bp corrispondono rispettivamente al vettore pGEM-T_FucS e pGEM-T_FucA mentre i frammenti di 653 bp e 651 bp rappresentano gli inserti XilS e XilA.

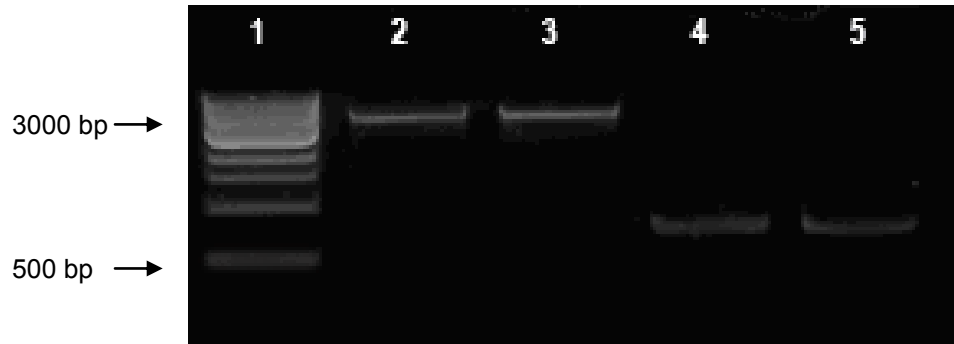


Figura 3.9 Corsia 1: Ladder 1 kb (NEB); 2-3: digestione *Xba* I e *Sac* I; 4-5: digestione *Nhe* I e *Sac* I.

Successivamente alla ligation e alla trasformazione di cellule JM101 di *E. coli*, è stata eseguita la PCR su tre colonie trasformate con pGEM-T_FucS-XilS che sono risultate positive all'analisi (Fig. 3.10).

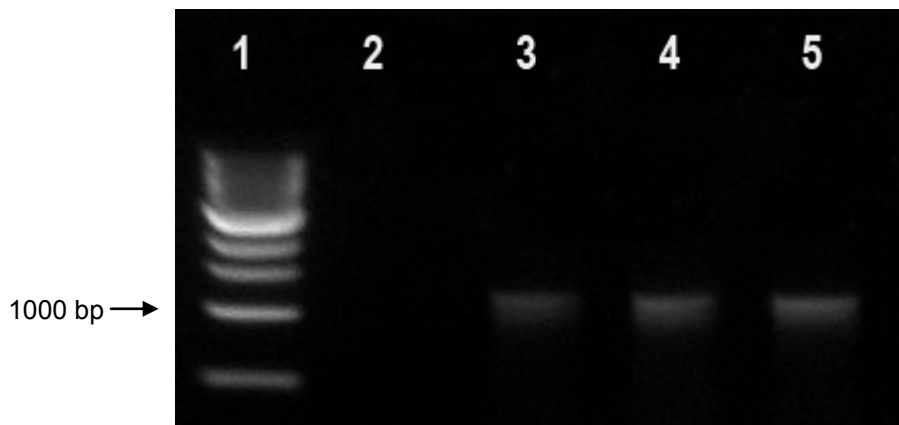


Figura 3.10 Corsia 1: Ladder 1 kb (NEB); 2: acqua; 3-4-5: colonie saggate mediante PCR.

Parallelamente è stata eseguito il controllo PCR con i primer FucA for-XilA rev in numerose colonie putativamente trasformate con pGEM-T_FucA-XilA; sette di esse hanno mostrato l'amplificato atteso (Fig. 3.11). Il sequenziamento di una

colonia per ciascun costrutto ha ulteriormente dimostrato che le sequenze inserite in pGEM-T sono perfettamente uguali a quella programmate.

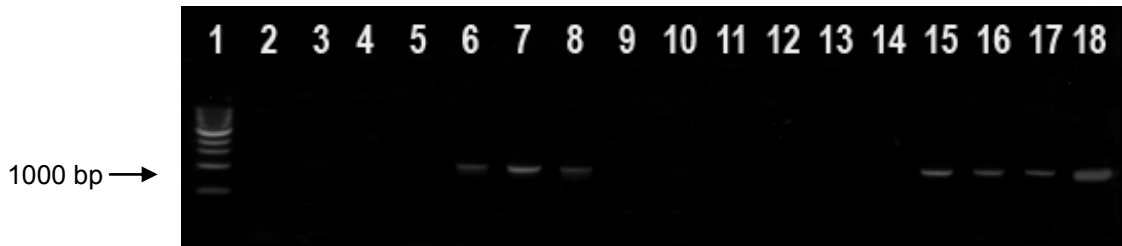


Figura 3.11 Corsia 1: Ladder 1 kb (NEB); 2: acqua; 3-18: colonie analizzate in PCR.

Ha fatto seguito l'introduzione dei due bracci della forcina (senso e antisense) nel vettore pSTARLING per la realizzazione della cassetta di espressione FxAntiFx completa. A questo scopo il vettore è stato recuperato sia dopo digestione con *Pac I/BamH I*, sia dopo digestione con *Kpn I/Spe I* per consentire l'inserimento rispettivamente della sequenza FucS-XiIS di 1200 bp recuperata mediante digestione *Pac I/BamH I* e di quella FucA-XiIA ottenuta dalla digestione con *Kpn I/Spe I*. La Fig. 3.12 evidenzia i frammenti attesi prodotti.

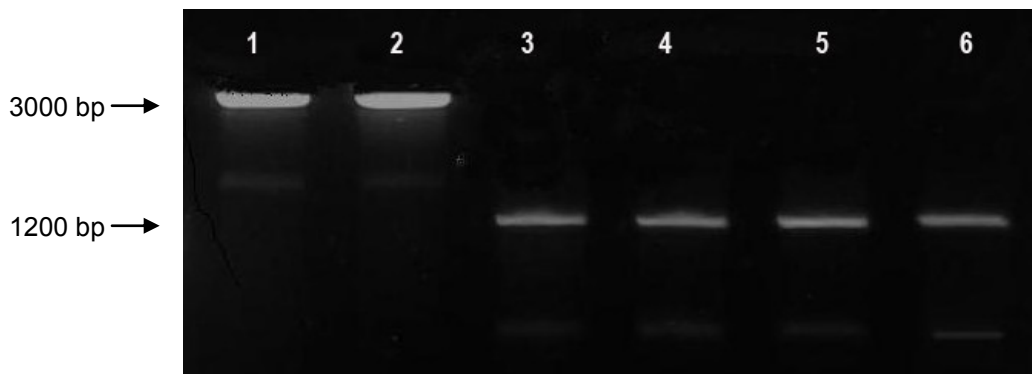


Figura 3.12 Corsia 1: digestione pSTARLING con *Pac I/BamH I*; 2: digestione pSTARLING con *Kpn I/Spe I*; 3: digestione inserto FucS-XiIS con *Pac I/BamH I*; digestione inserto FucA-XiIA con *Kpn I/Spe I*.

A saldatura avvenuta sono stati prodotti i seguenti vettori pSTARLING_FucS-XiIS e pSTARLING_FucA-XiIA impiegati per la trasformazione di cellule JM101 di *E. coli*. Un numero variabile di colonie trasformate con i due costrutti sono state

sottoposte a controllo attraverso PCR con primer specifici (FucS for-XiIS rev e FucA for-XiIA rev). Tutte le colonie saggiate, ad eccezione di una, hanno evidenziato l'amplificazione desiderata (Fig. 3.13).

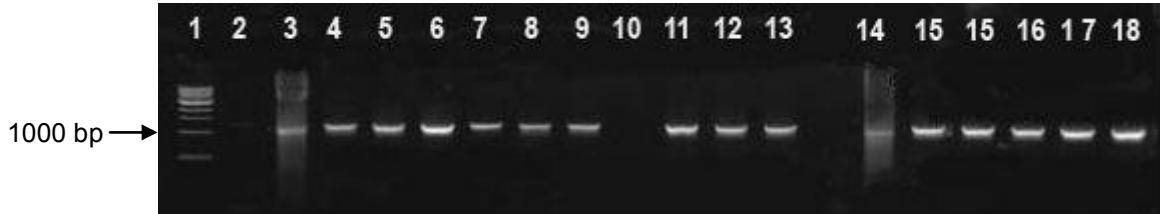


Figura 3.13 Corsia 1: Ladder 1 kb (NEB); 2: acqua; 3-13: colonie pSTARLING_FucS-XiIS saggiate in PCR; 14-18: colonie pSTARLING_FucA-XiIA saggiate in PCR.

Successivamente sia pSTARLING_FucS-XiIS, sia pSTARLING_FucA-XiIA sono stati digeriti con *Spe* I e *Kpn* I in modo da ottenere rispettivamente il vettore e l'inserito per produrre, tramite saldatura, il vettore portante la cassetta completa FxAntiFX. Come primo controllo, si è deciso di impostare analisi PCR su 10 colonie per amplificare separatamente i due bracci (senso e antiseno) utilizzando sia i primer FucS for-XiIS rev, sia FucA for-XiIA rev. Tutte le colonie hanno dato esito positivo all'amplificazione con ambedue le coppie di primer (Fig. 3.14).

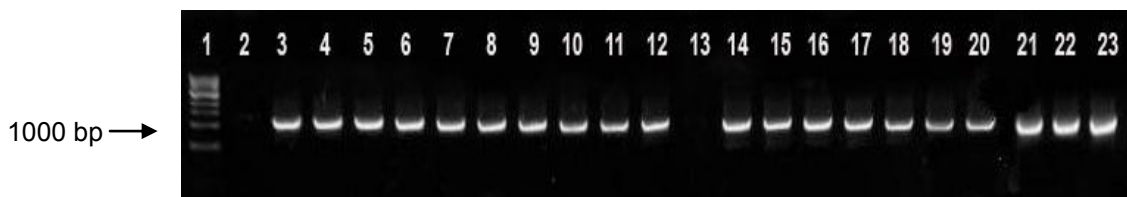


Figura 3.14 Corsia 1: Ladder 1 kb (NEB); 2-12: colonie saggiate in PCR con FucS for-XiIS rev; 13: vuoto; 14-23: colonie saggiate in PCR con FucA for-XiIA rev.

Da due colonie si è proceduto all'estrazione del DNA plasmidico su cui sono state condotte le seguenti digestioni di controllo:

1. digestione doppia con *Kpn* I/*Spe* I per verificare il corretto inserimento del braccio costituito dalle sequenze antiseno di 1200 bp;
2. digestione doppia con *Bam*H I/*Pac* I atta a rimuovere il braccio formato dalle sequenze senso di 1200 bp;

3. digestione con *Not I* per la rimozione dell'intera cassetta di espressione di 5556 bp.

La Fig. 3.15 mostra le bande prodotte dalle digestioni di cui al punto 1 e 2, aventi lunghezza corrispondente a quella attesa; in Fig. 3.16 sono invece riportati i risultati ottenuti dalla digestione 3.

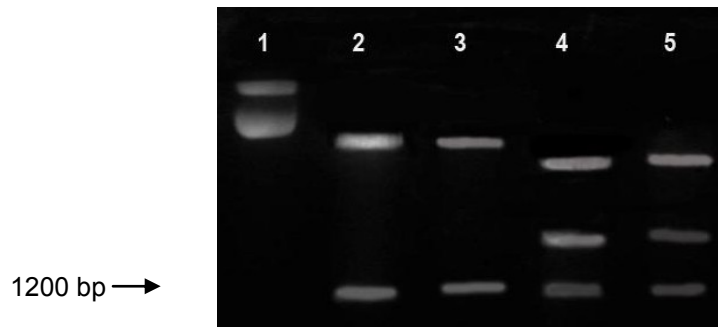


Figura 3.15 Corsia 1: campione non digerito; 2-3: digestione *Kpn I/Spe I*; 4-5: digestione *BamH I/Pac I*.

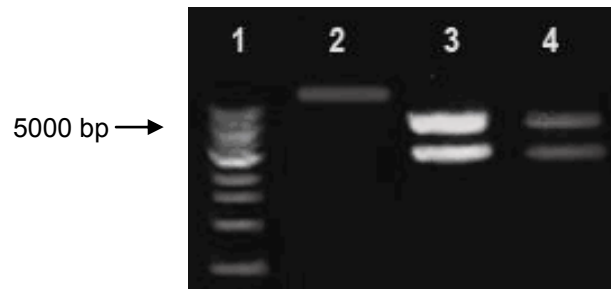


Figura 3.16 Corsia 1: Ladder 1 kb; 2: campione non digerito; 3-4: digestione *Not I*.

Per la realizzazione dei vettori a tre cassette sia per GAA sia per GCasi, pSTARLING_FucS-XiIS + FucA-XiIA è stato successivamente digerito con *Not I* per consentire il recupero della cassetta completa per il silenziamento.

3.2.3 pCAMBIA1300_GAA CC-NOS ter_PMI e pCAMBIA1300_GAA CC-NOS ter_PMI_FxAntiFx

Per la realizzazione del vettore a due cassette (pCAMBIA1300_GAA CC-NOS ter_PMI) si è proceduto per sostituzione della cassetta di espressione GCasi nel vettore pCAMBIA1300_GCasi nativa-NOS ter_PMI (cfr. § 3.8.1) con quella della GAA presente nel vettore pSV_GAA ottenuto in un precedente lavoro. A tal fine

ambidue i vettori sono stati sottoposti a digestione enzimatica con *EcoR* I (Fig. 3.17).

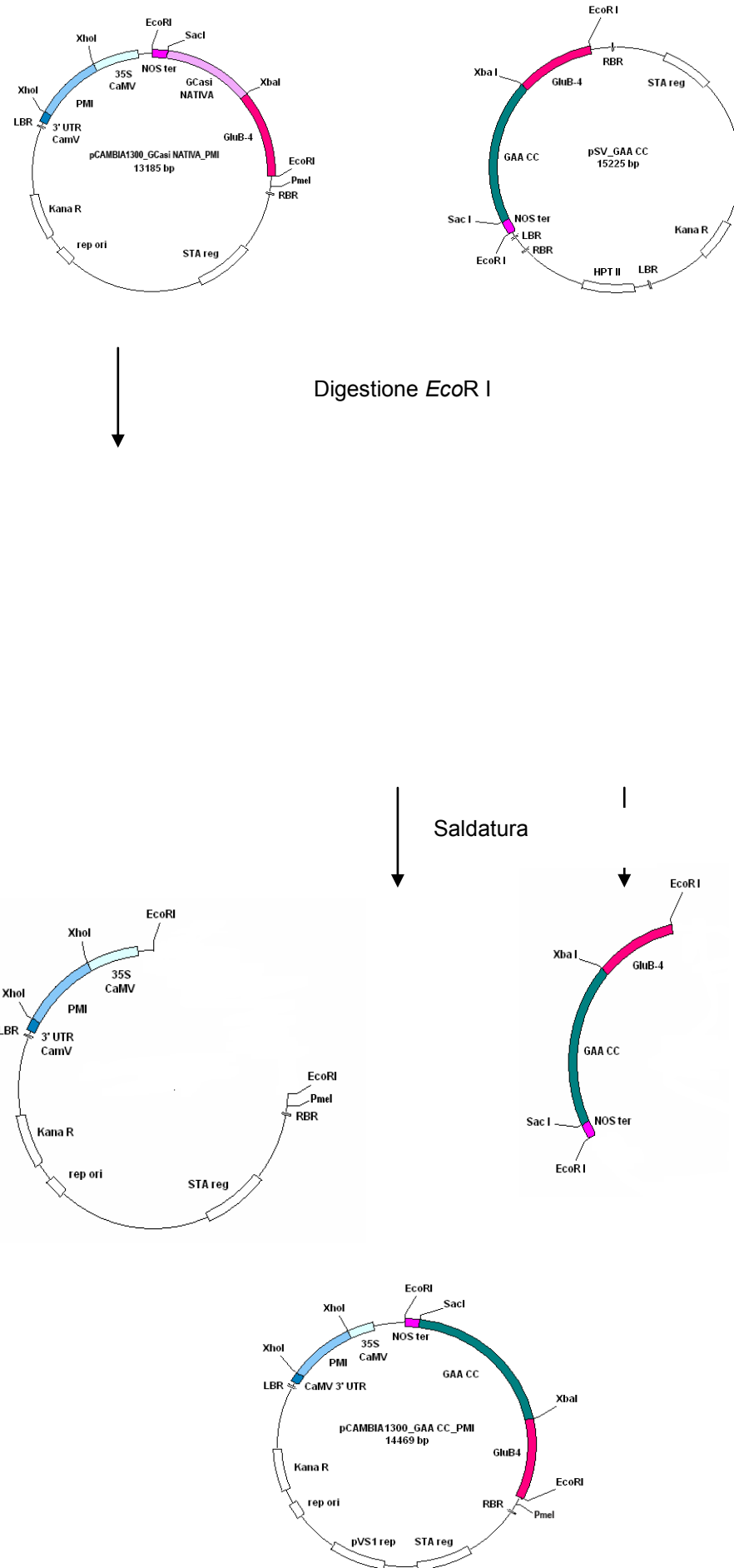


Figura 3.17 Rappresentazione schematica dei passaggi che hanno portato alla formazione del costrutto pCAMBIA1300_GAA CC_PMI.

Successivamente alla saldatura dell'inserto GAA nel vettore recuperato dalla digestione e alla trasformazione di cellule JM101 di *E. coli*, 3 colonie putativamente trasformate sono state verificate mediante digestione con *Xho* I, *EcoR* V, *Aat* II, *Sac* I e *Pme* I. Tutte e tre le colonie saggate hanno presentato i *pattern* di restrizione attesi (Fig. 3.18).



Figura 3.18 Corsia 1: Ladder 1 kb (NEB); 2: campione non digerito; 3: colonia 1 digestione *Xho* I; 4: colonia 1 digestione *EcoR* V; 5: colonia 1 digestione *Aat* II; 6: colonia 1 digestione *EcoR* I; 7: colonia 1 digestione *Sac* I; 8: colonia 1 digestione *Pme* I; 9: colonia 2 digestione *Xho* I; 10: colonia 2 digestione *EcoR* V; 11: colonia 2 digestione *Aat* II; 12: colonia 2 digestione *EcoR* I; 13: colonia 2 digestione *Sac* I; 14: colonia 2 digestione *Pme* I; 15: colonia 3 digestione *Xho* I; 16: colonia 3 digestione *EcoR* V; 17: colonia 3 digestione *Aat* II; 18: colonia 3 digestione *EcoR* I; 19: colonia 3 digestione *Sac* I; 20: colonia 2 digestione *Pme* I.

Il costrutto pCAMBIA1300_GAA CC-NOS ter_PMI prodotto è stato impiegato per la trasformazione di *Agrobacterium tumefaciens* ceppo EHA 105, effettuata mediante elettroporazione di cellule rese competenti.

Nel costrutto a due cassette è stata inserita la cassetta FxAntiFx tramite digestione del vettore pSTARLING_FucS-XilS + FucA-XilA con *Not* I. L'inserto recuperato è stato trattato con T4 DNA polimerasi e saldato nel vettore aperto con *Pme* I, producendo il plasmide pCAMBIA1300_GAA CC-NOS ter_PMI_FxAntiFx (Fig. 3.19).

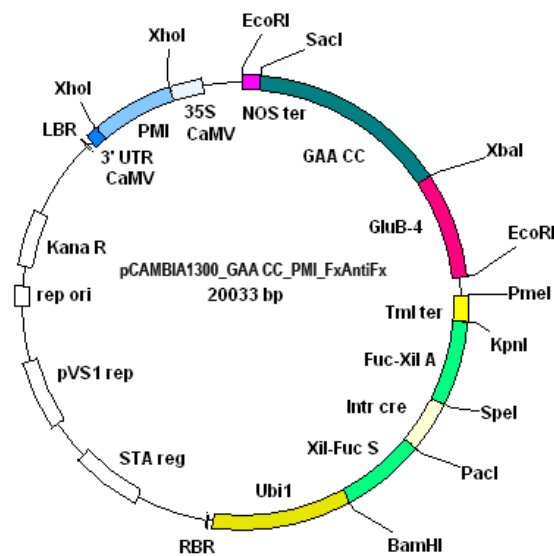


Figura 3.19 Rappresentazione schematica del vettore pCAMBIA1300_GAA CC-NOS ter_PMI_FxAntiFx.

Quattro colonie ottenute dalla trasformazione con pCAMBIA1300_GAA CC-NOS ter_PMI FxAntiFx sono state sottoposte a digestione di controllo con *Spe* I per verificarne l'orientamento. La Figura 3.20 mostra il pattern di restrizione ottenuto. La colonia risultata positiva all'analisi è stata ulteriormente analizzata mediante PCR per confermare la presenza dei geni PMI (primer PMI for/PMi rev), GAA (primer GAA_B for/GAA_B rev) e delle sequenze senso e antisense (primer FucS for/ XilS rev; FucA for/XilA rev) (Fig. 3.21). La colonia identificata è stata, quindi, impiegata per l'elettroporazione del ceppo di agrobatterio. Il sequenziamento della colonia di agrobatterio trasformata ha prodotto una sequenza identica a quella programmata.

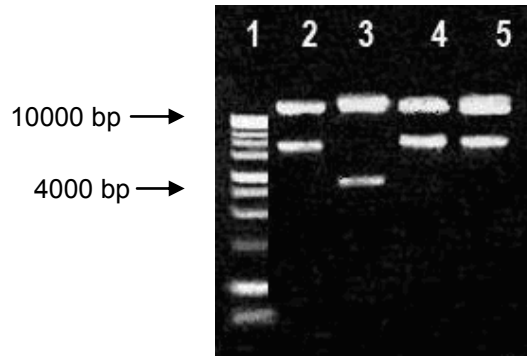


Figura 3.20 Corsia 1: Ladder 1 kb (NEB); 2: colonia 1 digestione *Spe I*; 3: colonia 2 digestione *Spe I*; 4: colonia 3 digestione *Spe I*; 5 colonia 4: digestione *Spe I*.

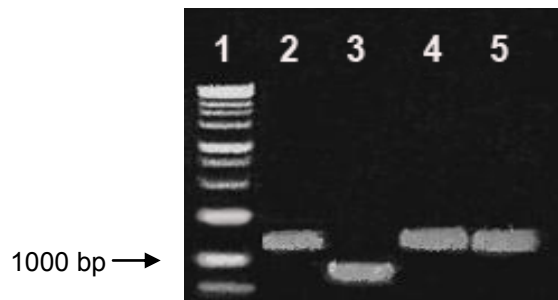


Figura 3.21 Corsia 1: Ladder 1 kb (NEB); 2: amplificato PMI for/PMI rev; 3: amplificato GAA_B for/GAA_B rev; 4: amplificato FucS for/XilS rev; 5: amplificato FucA for/XilA rev.

3.2.4 pCAMBIA1300_GAA CC-GluB4 ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx

Per la sostituzione del terminatore NOS con quello della glutelina 4 di riso si è proceduto inserendo consecutivamente terminatore GluB4, promotore GluB4 e gene d'interesse GAA in pUC18 attraverso una serie di digestioni enzimatiche e reazioni di saldatura.

Dalla digestione di pUC18_NOS ter e di pMK_GluB4 ter con *EcoR I* e *Sac I*, si sono ottenute le bande corrispondenti al vettore e all'inserto GluB4 ter (Fig. 3.23).

Successivamente alla *ligation* e alla trasformazione di cellule JM101 di *E. coli*, le colonie trasformate sono state selezionate mediante PCR. La Fig. 3.22 mostra gli amplificati ottenuti utilizzando i primer specifici per il terminatore GluB4 (GluB4ter for e GluB4ter rev).

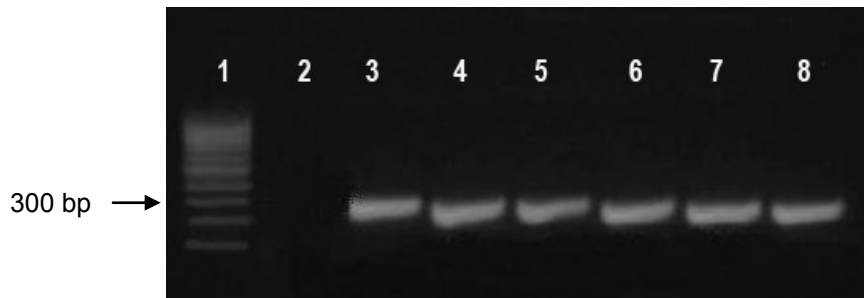
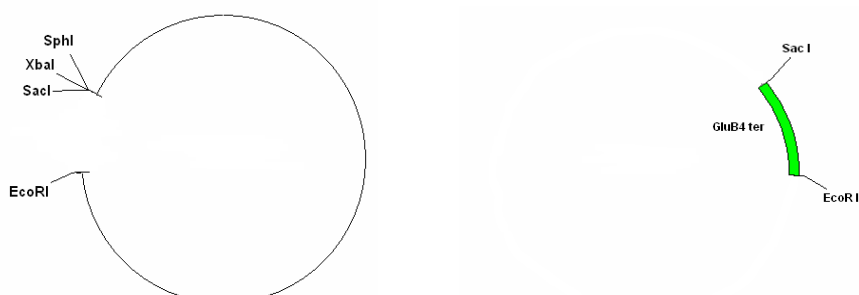


Figura 3.22 Corsia 1: Ladder 100 bp (NEB); 2: CN; 3: CP; 4-8: colonie saggiate con PCR.



Digestione *Sac I*/*EcoR I*

Saldatura



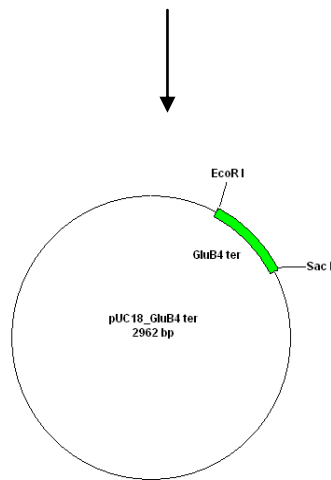


Figura 3.23 Rappresentazione schematica del passaggio che ha portato alla realizzazione del vettore intermedio pUC18_GluB4 ter.

pUC18_GluB4 LLTCK_GluB4 ter è stato ottenuto grazie all'assemblaggio del promotore GluB4 LLTCK nel vettore pUC18_GluB4 ter mediante digestione con *Xba* I e *Sph* I (Fig. 3.24).

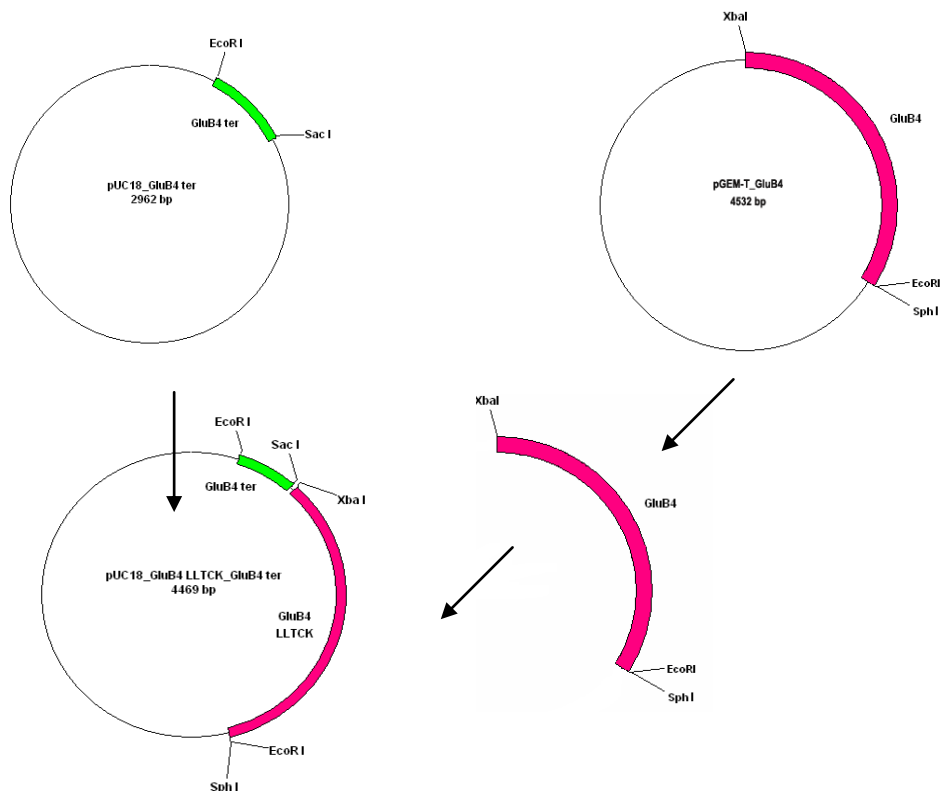


Figura 3.24 Realizzazione del vettore intermedio pUC18_GluB4 LLTCK_GluB4 ter.

La presenza e la correttezza del promotore GluB4 LLTCK sono state verificate mediante digestione congiunta con i medesimi enzimi *Xba* I e *Sph* I, che rimuovono dal vettore la zona del promotore stesso (Fig. 3.25).

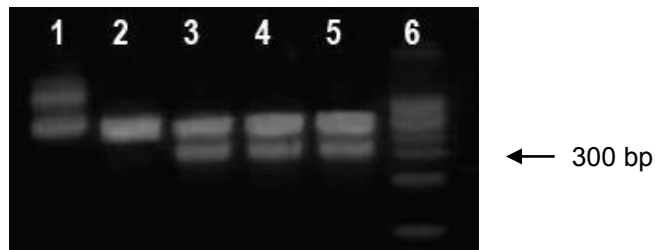
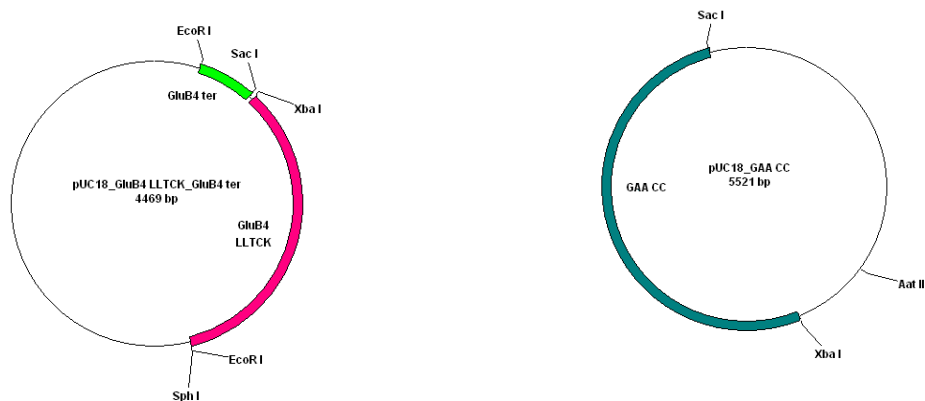


Figura 3.25 Corsia 1: campione non digerito; 2-5: colonie 2, 6, 7, 8 sottoposte a digestione *Xba* I/*Sph* I; 6: Ladder 1 kb. Le colonie 6, 7, 8 risultano positive per la presenza del promotore.

Il vettore pUC18_GluB4 LLTCK_GluB4 ter prodotto è stato successivamente aperto con *Xba* I e *Sac* I in modo da consentire l'inserimento della sequenza codificante la GAA riscritta secondo il codon context recuperata dal vettore pUC18_GAA CC tramite digestione con *Xba* I, *Sac* I e *Aat* II (Fig. 3.26).



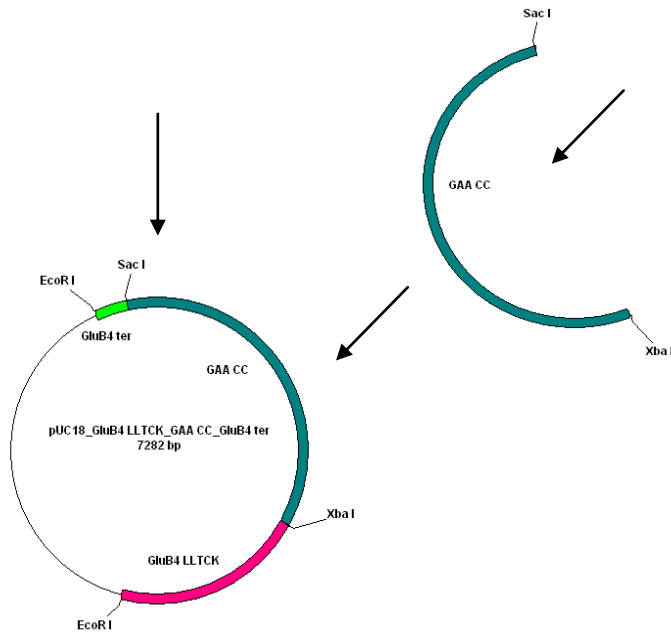


Figura 3.26 Realizzazione del vettore intermedio pUC18_GluB4 LLTCK_GAA CC_GluB4 ter.

È stata eseguita la miniprep di due colonie (A e B) che sono state saggiate per verificare i cloni positivi mediante digestione con *Xba* I e *Sac* I atta a rimuovere completamente la CDS GAA dal vettore. La Fig. 3.27 mostra i frammenti attesi di 2850 bp e di 4709 bp corrispondenti rispettivamente alla CDS GAA e al vettore residuo.

Il sequenziamento ha ulteriormente dimostrato che la sequenza GluB4 LLTCK_GAA CC_GluB4 ter è perfettamente uguale a quella programmata.

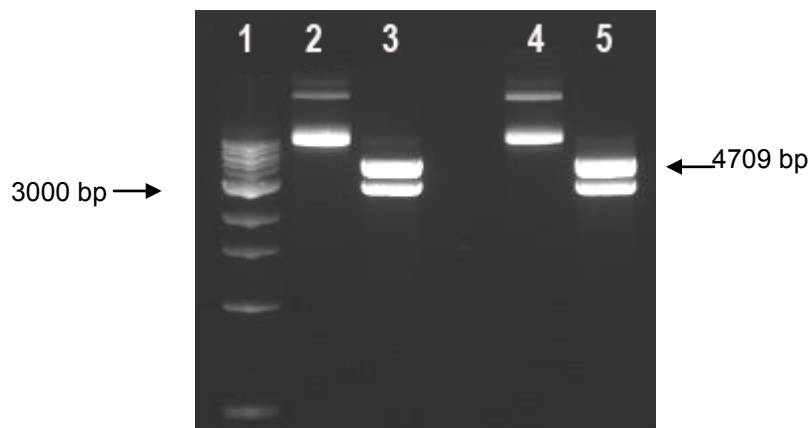


Figura 3.27 Corsia 1: Ladder 1 kb (NEB); 2: campione A non digerito; 3: campione A verificato mediante digestione *Xba* I/*Sac* I; 4: campione B non digerito; 5: campione B verificato mediante digestione *Xba* I/*Sac* I.

Per la realizzazione del vettore finale a due cassette, nel vettore pCAMBIA 1300_GluB4 LLTCK_GAA CC-NOS ter_PMI già prodotto in precedenza (cfr. § 3.2.2) si è proceduto alla sostituzione della cassetta di espressione portante il terminatore NOS con quella presente in pUC18_GluB4 LLTCK_GAA CC_GluB4 ter mediante digestione con *Eco*R I (Fig. 3.29)

Per il controllo del corretto orientamento dell'inserito si è proceduto alle digestioni con *Eco*R I e *Sac* I. La Fig. 3.28 mostra i profili di restrizione ottenuti con i due enzimi.

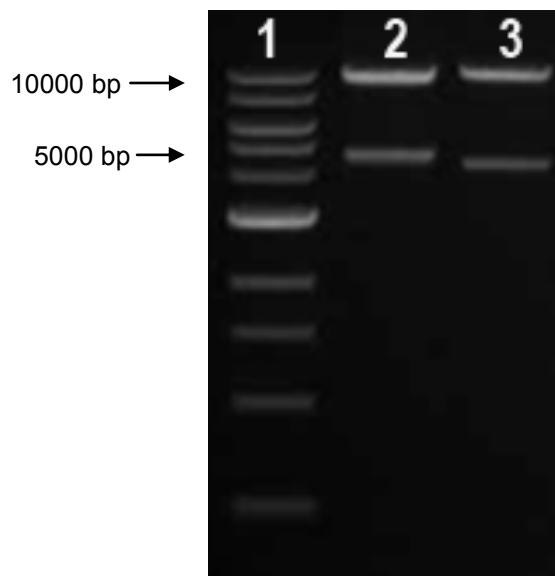


Figura 3.28 Corsia 1: Ladder 1 kb (NEB); 2: digestione con *Eco*R I; 3: digestione con *Sac* I.

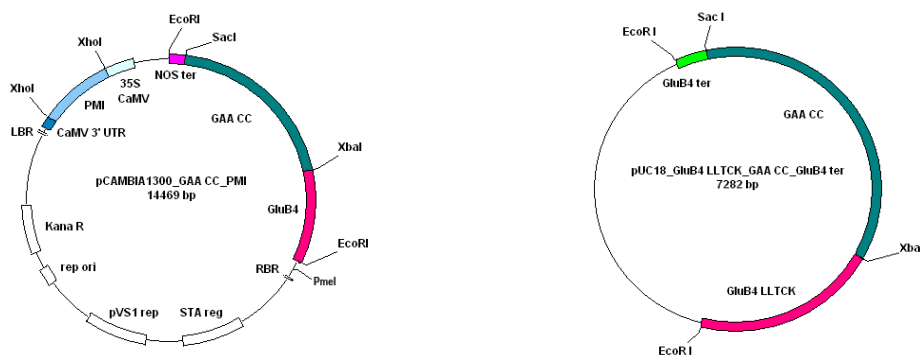




Figura 3.29 Rappresentazione schematica della realizzazione del vettore pCambia1300_GAA-CC-GluB4 ter_PMI tramite sostituzione della cassetta di espressione.

Il costrutto corretto a due cassette è stato impiegato per la trasformazione di *A. tumefaciens* e nuovamente verificato attraverso le seguenti digestioni.

- digestione tripla con *Xba* I, *Sac* I, *Eco*R I volta a rimuovere in tratti separati CDS, promotore e terminatore;
- digestione con *Xho* I per verificare la presenza del gene PMI.

In Fig. 3.30 sono mostrati i frammenti attesi di 9796 bp, 2852 bp, 1519 bp, 286 bp corrispondenti a vettore residuo, CDS GAA, promotore e terminatore generati dalla digestione tripla e il frammento di 1182 bp del gene PMI ottenuto dalla digestione con *Xho* I.

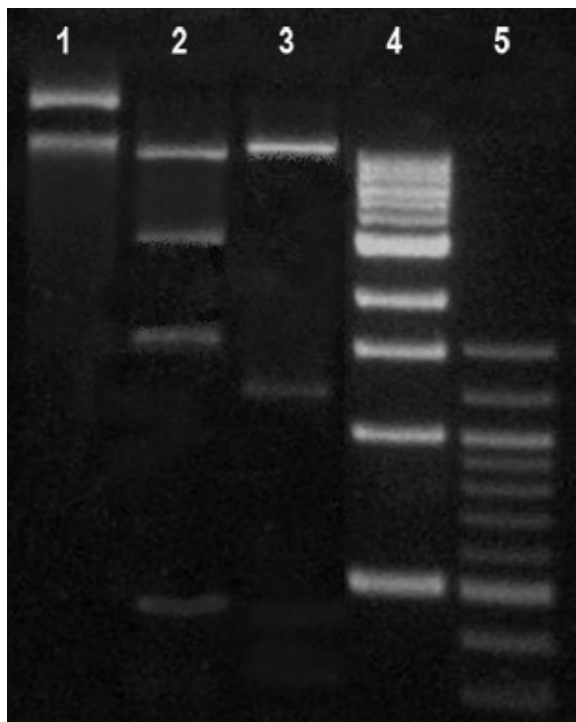


Figura 3.30 Corsia 1: costrutto non digerito; 2: digestione tripla; 3: digestione *Xho* I; 4: Ladder 1 Kb; 5: Ladder 100 bp (NEB).

Per ottenere il vettore finale a tre cassette (Fig. 3.31), pCAMBIA1300_GluB4 LLTCKGAA CC-GluB4 ter_PMI è stato tagliato con *Pme* I in modo da consentire l'introduzione della cassetta FxAntiFx rimossa dal vettore che la portava (pSTARLING_FxAntiFx) tramite digestione con *Not* I e sottoposta a trattamento con T4 DNA polimerasi per favorire la *ligation blunt* con il vettore.

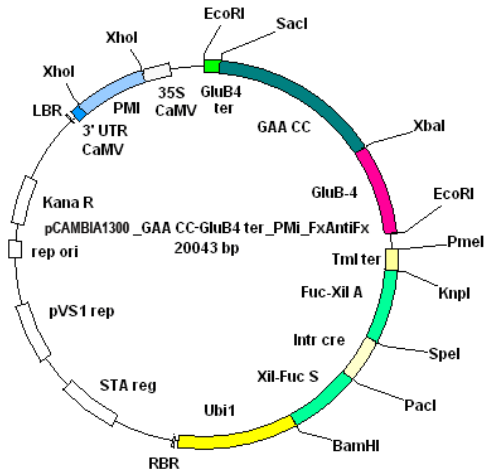


Figura 3.31 Rappresentazione schematica del vettore finale a tre cassette.

Tre diverse colonie (16-17-18) sono state analizzate per verificare il corretto orientamento della cassetta inserita attraverso digestione con *Spe I*, ma solo in un caso si è evidenziato il prodotto della lunghezza attesa (Fig. 3.32). In caso di orientamento corretto, le digestioni originano frammenti lunghi rispettivamente 17288 bp (vettore restante) e 2755 bp (cassetta FxAnti Fx). Se viceversa l'orientamento non è quello desiderato, si generano frammenti di 14893 bp e 5150 bp.

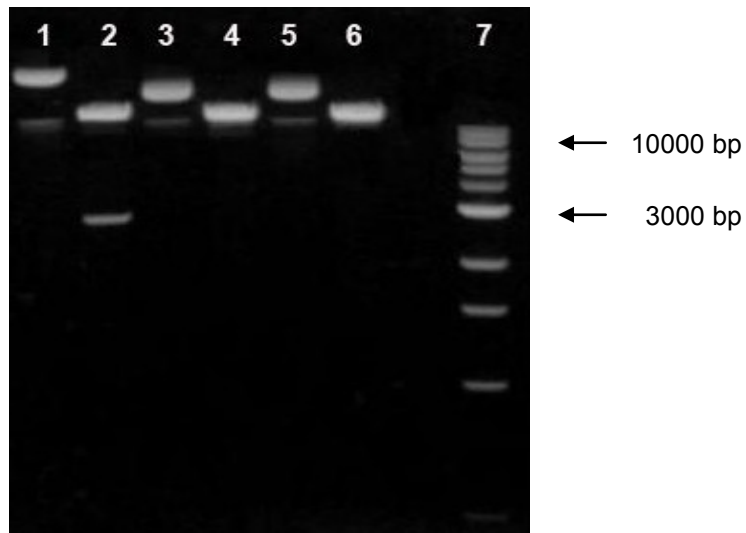


Figura 3.32 Corsia 1: campione 16 non digerito; 2: campione 16 verificato mediante digestione *Spe I*; 3: campione 17 non digerito; 4: campione 17 verificato mediante digestione *Spe I*; 5: campione 18 non digerito; 6: campione 18 verificato mediante digestione *Spe I*; 7: Ladder 1 kb (NEB).

Il costrutto finale è stato, quindi, controllato mediante sequenziamento e utilizzato per la trasformazione di *A. tumefaciens* EHA 105 effettuata mediante elettroporazione. La verifica del trasferimento del costrutto in agrobatterio è avvenuta tramite digestione di controllo con gli enzimi *EcoR* I, *Sac* I, *Spe* I. Tutte e tre le digestioni hanno prodotto il *pattern* di restrizione atteso (Fig. 3.33).

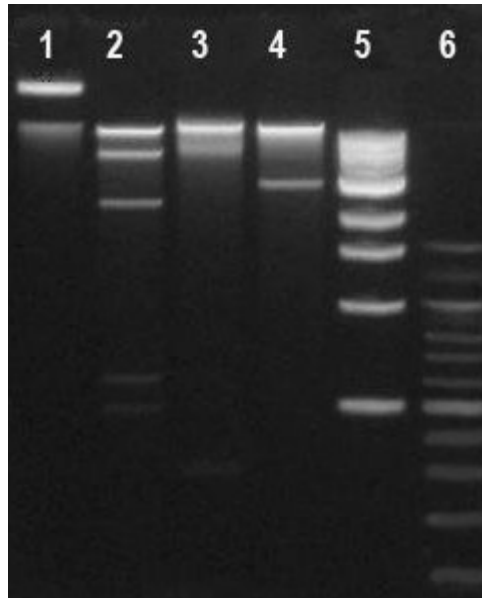


Figura 3.33 Corsia 1: campione non digerito; 2: digestione *EcoR* I; 3: digestione *Sac* I; 4. digestione *Spe* I; 5: Ladder 1 kb; 6: Ladder 100 bp (NEB).

3.3 Trasformazione del riso con *Agrobacterium tumefaciens*

La procedura di infezione del riso con cellule di *A. tumefaciens* portanti il costrutti di interesse si è svolta senza inconvenienti e secondo i tempi e modi previsti dal protocollo di riferimento. Di seguito vengono riportate alcune immagini fotografiche rappresentative dei punti fondamentali del processo di trasformazione genetica (Fig. 3.34).

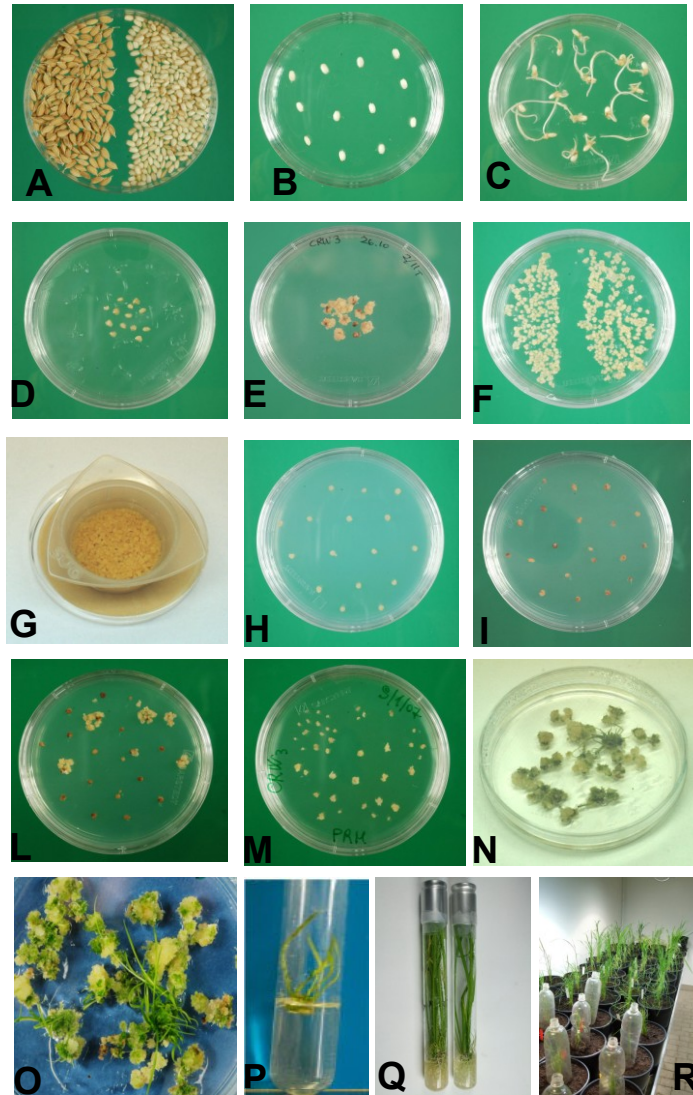


Figura 3.34 Tavola raffigurante le principali fasi di trasformazione del riso. A: sbramatura; B: disinfezione; C: germinazione; D: isolamento scutelli; E: callogenesi; F: selezione embrioidi; G: infezione con coltura liquida di agrobatterio; H: embrioidi su substrato CCM dopo infezione; I: embrioidi trasformati e non trasformati su substrato SMI; L: calli da embrioidi trasformati su substrato SMII; M: calli trasformati su substrato PRM; N: rigenerazione; O: differenziazione; P-Q: radicazione; R: trasferimento in torba.

3.4 Analisi molecolari sulle piante trasformate con i vettori pCAMBIA1300_GAA CC-GluB4 ter _PMI e pCAMBIA1300_GAA CC-GluB4 ter _PMI_FXantiFX

I vettori pCAMBIA1300_GAA CC-GluB4 ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx sono stati utilizzati per la trasformazione mediata da agrobatterio di scutelli derivati da semi di riso. Sono state rigenerate rispettivamente 147 piante putativamente trasformate con il costrutto a tre cassette e 136 con il vettore a due cassette. In entrambe le popolazioni di piante, l'effettiva presenza del gene GAA è stata verificata mediante PCR. I saggi PCR sono stati condotti con coppie di primer specifiche per il gene GAA, GAA_B for/GAA_B rev (Fig. 3.35) e GAA_C for/GAA_C rev (Fig. 3.36), in presenza di un controllo negativo (CN), un controllo positivo (CP) e un controllo di amplificabilità del DNA genomico (Fig. 3.37).



Figura 3.35 Corsie 1-19: piante di riso putativamente trasformate GAA-GluB4 ter_PMI_FxAntiFx saggiate mediante PCR con i primer GAA_B for/ GAA_B rev; 20: CP; 21: acqua; 22: CN; 23: Ladder 1 Kb (NEB).

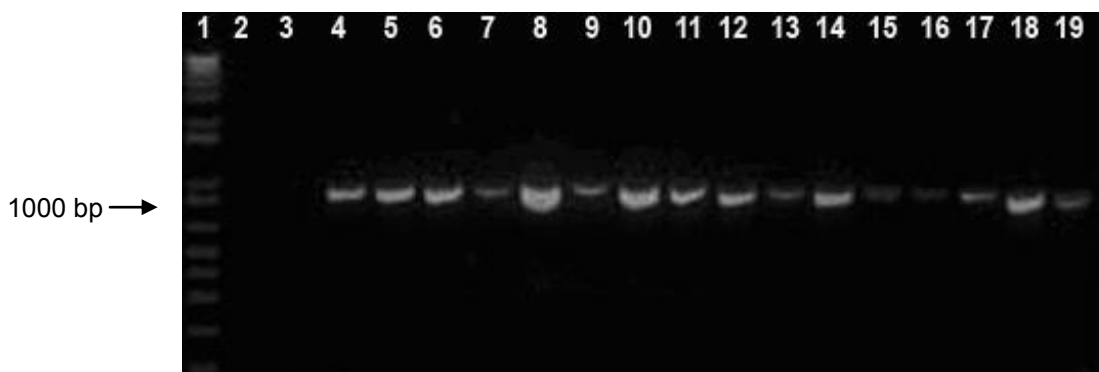


Figura 3.36 Corsia 1: Marker 1 kb (INVITROGEN); 2: acqua; 3: CN; 4: CP; 5-19: piante riso putativamente trasformate saggiate con i primer GAA_C for e GAA_C rev.

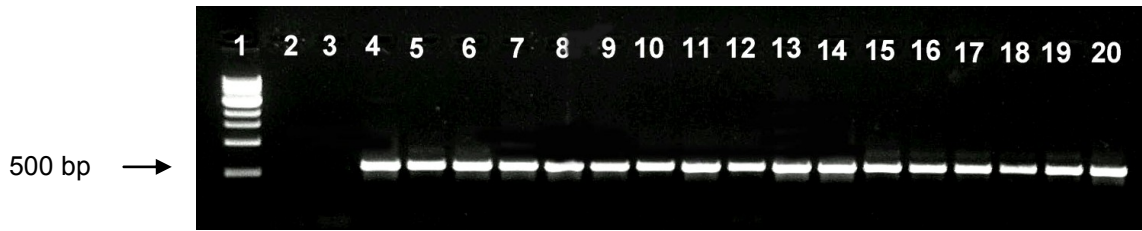


Figura 3.37 Corsia 1: Ladder 1 kb (NEB); 2: acqua; 3: CN; 4: CP; 5-20: piante saggate con primer specifici per DNA di riso (CPC e CPD).

I primer CPC e CPD impiegati sono stati costruiti su una regione conservata del DNA cloroplastico di riso (cfr. § 2.11); la lunghezza dell'amplificato atteso è stata calcolata in 614 bp. La presenza dell'amplificato è stata osservata in tutti i campioni saggiati, dimostrando la correttezza della procedura di estrazione del DNA da foglia. Al contrario, la percentuale di piante rigenerate risultate PCR positive al gene di interesse è stata pari al 78.91% e 98.52% rispettivamente nella popolazione portante il costrutto a tre cassette e in quella generata dalla trasformazione con pCAMBIA1300_GAA CC-GluB4ter_PMI.

Le piante positive per il gene di interesse sono state successivamente saggate tramite PCR per verificare la presenza di un tratto esterno ai due border repeats del T-DNA come indicatore dell'inserimento di parte del *backbone* del vettore nelle piante trasformate (Fig. 3.38).

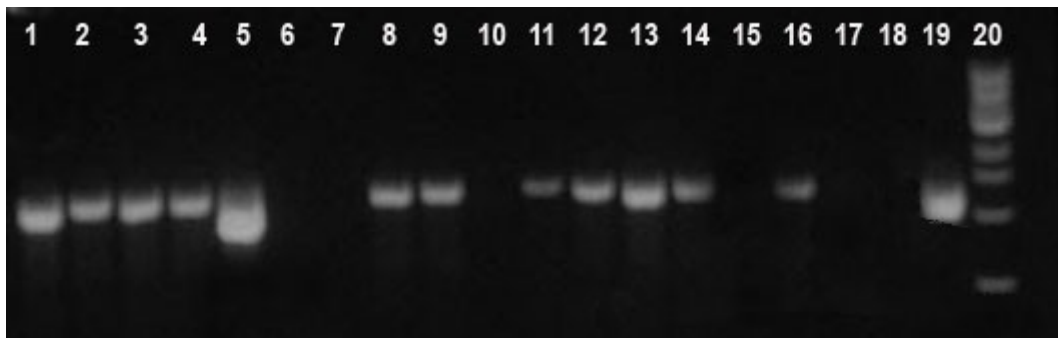


Figura 3.38 Corsie 1-16: piante saggate; 17: acqua; 18: CN; 19: CP; 20: Ladder 1 kb (NEB).

La percentuale di individui positivi all'amplificazione è stata particolarmente elevata nella popolazione di trasformati primari portanti le tre cassette (64.25%) mentre è risultata pari al 29.10% negli individui trasformati con il vettore a due cassette (Graf. 3.1).

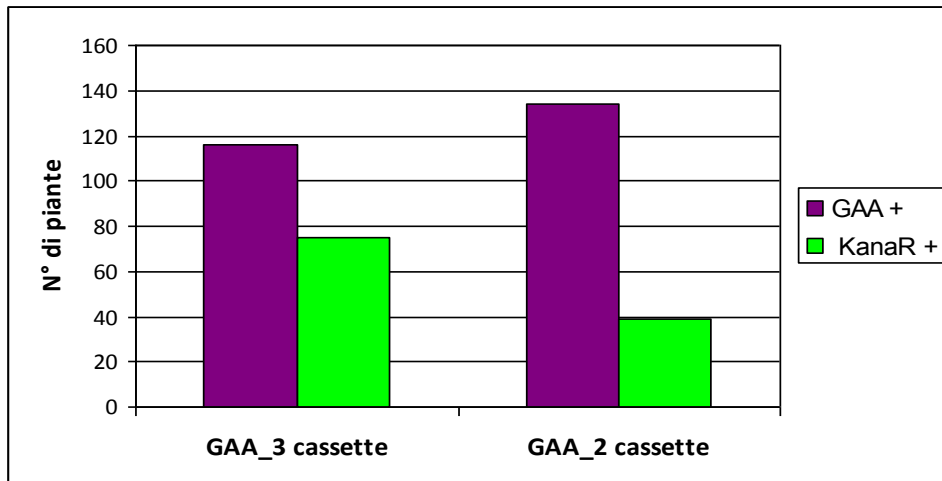


Grafico 3.1 Confronto tra il numero di individui positivi al gene d'interesse e al backbone del vettore nei trasformati primari a due e tre cassette.

3.5 Analisi sulla progenie

L'analisi DAS-ELISA sul seme prodotto dai trasformati primari, ottenuti dalla trasformazione con i vettori di espressione pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx e pCAMBIA1300_GAA CC-NOS ter_PMI_FxAntiFx, ha permesso di valutare l'effetto del terminatore sui livelli di espressione della proteina ricombinante.

A questo scopo è stato sviluppato un protocollo basato sull'estrazione di proteine totali da seme e sulla quantificazione della proteina di interesse mediante saggio immunologico DAS-ELISA.

3.5.1 Prove di estrazione di proteine totali da seme

Per la determinazione del contenuto di proteina ricombinante nell'endosperma di riso è stato messo a punto un protocollo che ha previsto l'estrazione di proteine totali a partire da 40 semi di ciascun individuo e il successivo saggio immunoenzimatico per la quantificazione dell'enzima. La prima fase nella procedura di estrazione è stata la sbramatura e macinazione dei 40 semi campionati casualmente. Nel tentativo di incrementare l'efficienza del processo di estrazione proteica sono state condotte prove di macinazione con due diversi sistemi, mulino a sfere e macina caffè da tavolo. I dati ottenuti, riportati nel

Grafico 3.2, hanno evidenziato una più elevata quantità di proteina GCasi e GAA nel caso di campioni macinati con mulino a sfere.

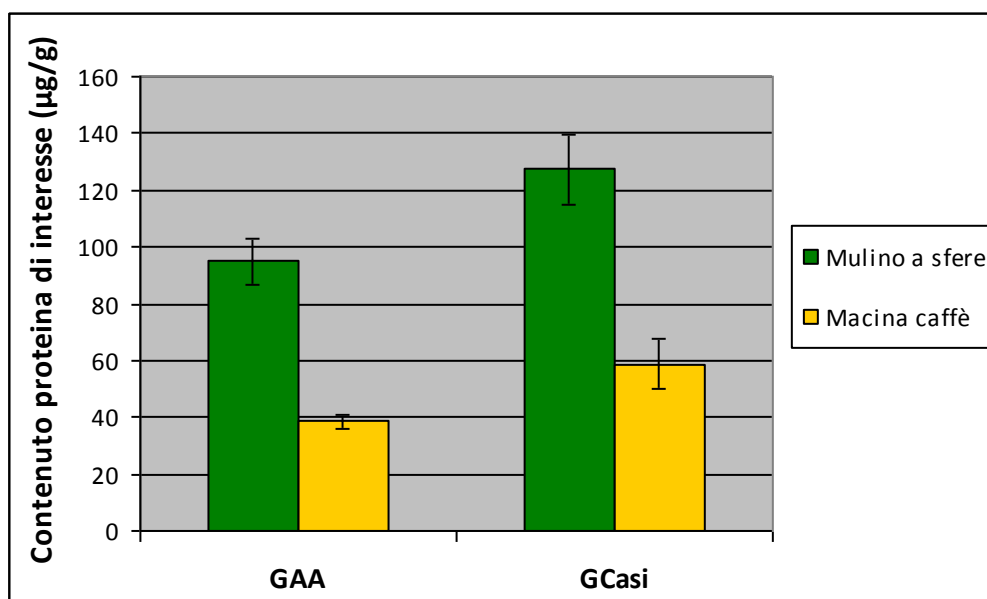


Grafico 3.2 Contenuto di proteina di interesse in campioni sottoposti a macinazione con mulino a sfere e macina caffè.

3.5.2 Validazione del saggio DAS-ELISA

Allo scopo di valutare la validità in termini di specificità e selettività del saggio immunoenzimatico sviluppato per determinare il contenuto di enzima ricombinante negli estratti proteici da seme, diluizioni 1:5, 1:10, 1:15 dell'estratto tal quale sono state confrontate con le medesime diluizioni del campione cui è stato aggiunto lo standard (Myozyme) a concentrazione nota (Tab. 3.2).

Diluizioni	[] campione tal quale senza aggiunta standard	[] campione con aggiunta standard	Valore atteso
1:5	1.93	1.02	1.82 +/- 0.273
1:10	2.02	1.36	2.16 +/- 0.324
1:15	2.24	1.57	2.37 +/- 0.355

Tabella 3.2. Concentrazioni dell'estratto GCasi diluito 1:5, 1:10, 1:15 rilevate in assenza/presenza dello standard a concentrazione nota e valori attesi nel range di validità (+/- 15%).

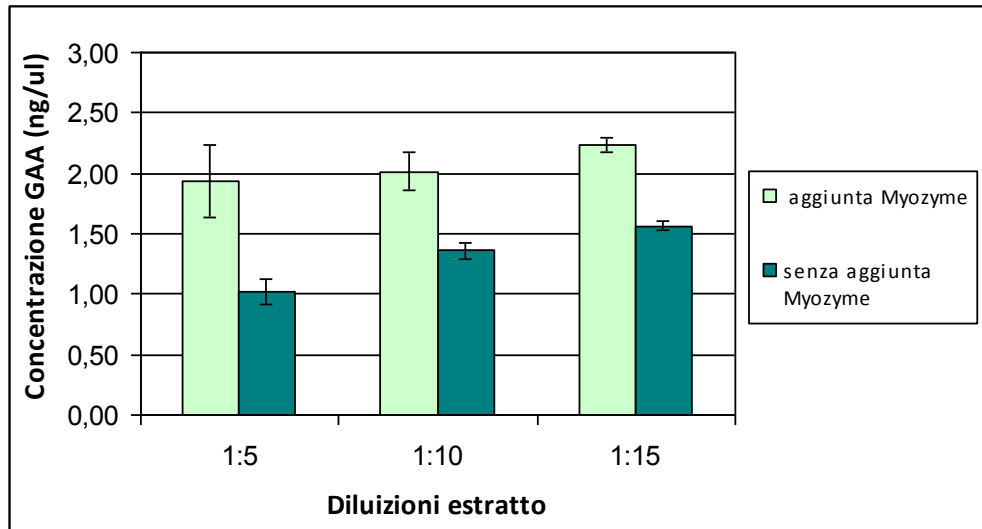


Grafico 3.3. Confronto tra le concentrazioni di GAA ottenute a tre diverse diluizioni dell'estratto nel campione tal quale e dopo aggiunta dello standard in quantità nota (Myozyme).

Come si osserva nel Grafico 3.3, le concentrazioni di GAA ottenute alle diverse diluizioni sono state pressocchè le stesse; inoltre, i dati relativi ai campioni trattati con lo standard sono ricaduti all'interno dell'intervallo di validità (+/- 15%) stabilito come indice di non interferenza (Tab. 3.2).

3.5.3 Valutazione dell'effetto del terminatore sui livelli di espressione della proteina ricombinante

Per valutare l'effetto dell'elemento terminatore sulla produzione di enzima ricombinante nell'endosperma, è stato analizzato il seme di 40 trasformati primari scelti casualmente ottenuti con i costrutti pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx. Confrontando il contenuto di proteina di interesse nelle due popolazioni considerate, si è osservato un netto incremento nei livelli di espressione di GAA a seguito dell'inserimento del terminatore GluB4 (Graf. 3.4). In particolare, nel seme della linea GAA CC-GluB4 ter avente il maggior contenuto di GAA si sono raggiunti livelli di produzione pari a circa 150 µg/g, più di due volte superiori a quelli rilevati nel migliore espressore GAA CC_NOS ter.

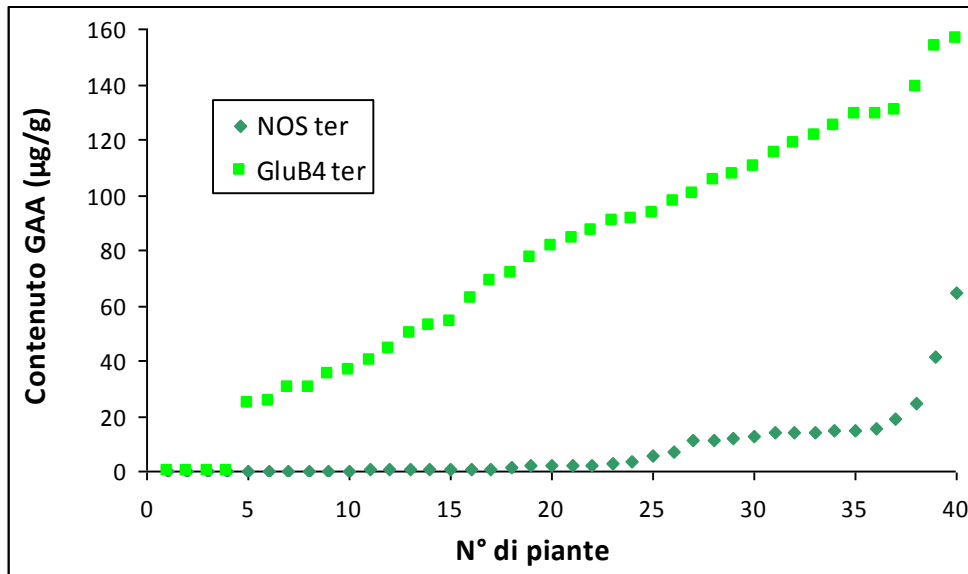


Grafico 3.4. Confronto dei livelli di espressione nelle popolazioni di 40 trasformati primari scelti a caso ottenuti con i vettori portanti rispettivamente il terminatore NOS e GluB4.

3.5.4 Selezione del migliore espressore GAA CC-GluB4 ter_PMI_FxAntiFX

Alla luce dei risultati sperimentali ricavati dal confronto tra i terminatori, l'analisi DAS-ELISA è stata estesa all'intera popolazione di trasformati primari GAA CC-GluB4 ter risultati PCR-positivi per la presenza del gene GAA (confr. §. 3.4) al fine di identificare, da un lato, l'individuo a maggior contenuto di enzima ricombinante nel seme e, dall'altro, verificare l'ipotesi di una possibile correlazione tra produttività e inserimento di parti del vettore esterne al T-DNA; quest'ultimo evento è stato valutato tramite l'amplificazione del tratto esterno ai *border repeats*.

Il seme di 101 trasformati primari è stato, quindi, sottoposto a estrazione di proteine totali e al successivo saggio DAS-ELISA. La progenie di 11 piante si è dimostrata pressoché priva di enzima ricombinante (nell'estratto saggiato in ELISA erano, infatti, presenti concentrazioni inferiori al valore minimo utilizzato nella costruzione della curva di taratura, ovvero 2 pg/µL) mentre il contenuto di proteina di interesse è risultato variare notevolmente tra le restanti linee come si osserva in Graf. 3.5.

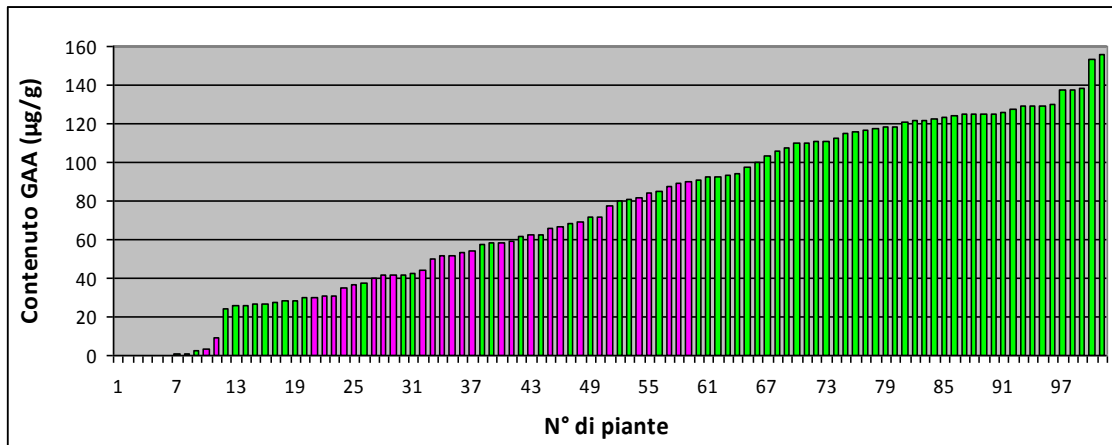


Grafico 3.5. Contenuto GAA rilevato nel seme prodotto dai trasformati primari GAA CC-GluB4 ter a tre cassette.

Dall'identificazione degli individui risultati positivi all'amplificazione del gene KanaR (in verde), si è osservata una corrispondenza significativa tra il maggior contenuto di proteina GAA e presenza del gene KanaR. Infatti, le linee prive del gene KanaR, che presumibilmente hanno integrato il solo T-DNA portante le cassette di interesse (in rosa), hanno presentato livelli di GAA nel seme relativamente bassi (non superiori a 90 µg/g) rispetto a quelli rilevati negli individui PCR-positivi al tratto esterno che si sono dimostrati i migliori espressor con un contenuto di enzima ricombinante nell'endosperma che ha raggiunto i 150 µg/g.

3.5.5 Selezione del migliore espressore GAA CC-GluB4 ter_PMI

Il medesimo lavoro di selezione è stato condotto sulla progenie dei trasformati primari ottenuti con il costrutto a due cassette GAA CC-GluB4 ter_PMI. Di seguito (Graf. 3.6) sono riportati i risultati ottenuti all'analisi DAS-ELISA.

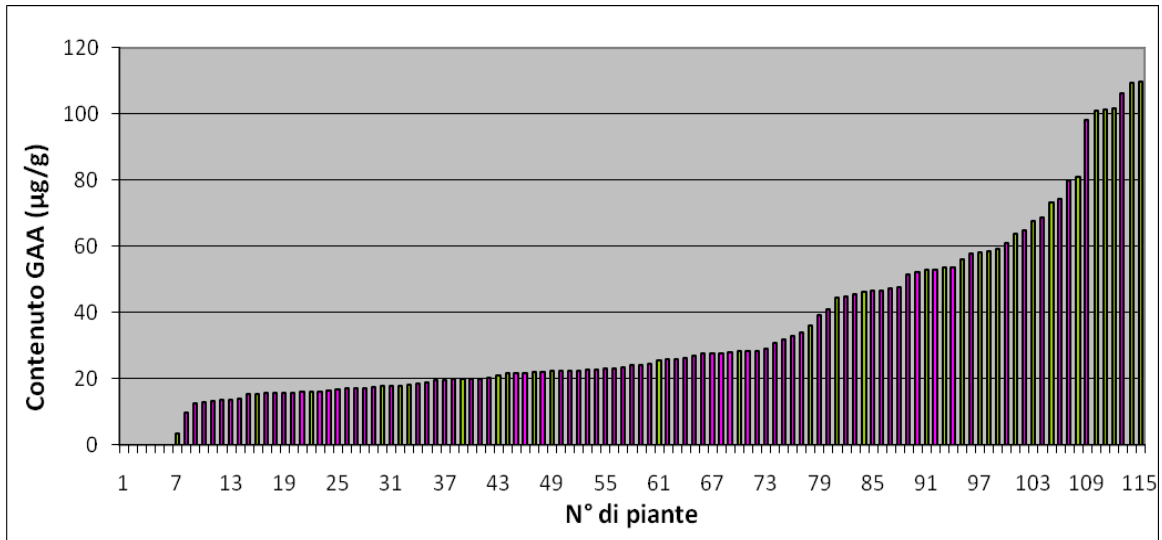


Grafico 3.6 Contenuto di GAA rilevato nel seme prodotto dai trasformati primari GAA CC-GluB4 ter a due cassette.

Il contenuto di proteina di interesse determinato nel seme di 115 trasformati primari è risultato variabile tra gli individui e in 5 linee ha raggiunto livelli di espressione superiori a 100 µg/g.

Diversamente dai dati ottenuti saggiando le progenie dei trasformati primari a tre cassette, il contenuto di α -glucosidasi acida in funzione della presenza o meno del *backbone* del vettore ha mostrato una distribuzione maggiormente omogenea tra le linee sebbene i migliori espressori siano risultati positivi all'amplicazione del tratto esterno.

3.6 Purificazione di GAA

Per la purificazione di GAA da seme di riso, è stato messo a punto un protocollo che prevede una fase iniziale di preparazione del campione seguita da tre fasi cromatografiche distinte. Nella Fig. 3.39 viene schematicamente illustrata l'intera procedura.

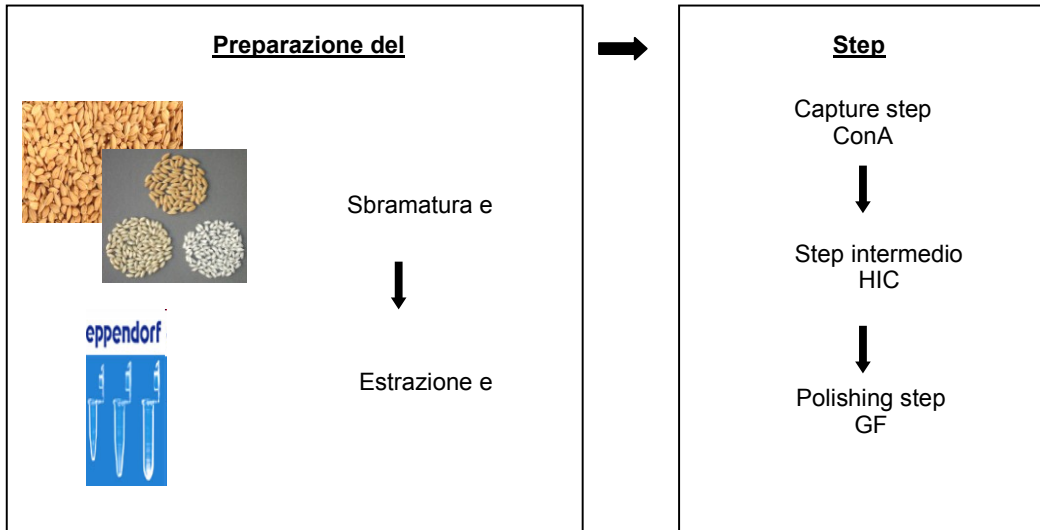


Figura 3.39 Schema riportante le attività svolte per la purificazione di GAA.

Nel protocollo di purificazione è stata determinata la combinazione tra una cromatografia di affinità con concanavalina A e una successiva cromatografia a interazione idrofobica (HIC). Di seguito (Fig. 3.40, 3.41) si riportano i cromatogrammi ottenuti nei due casi.

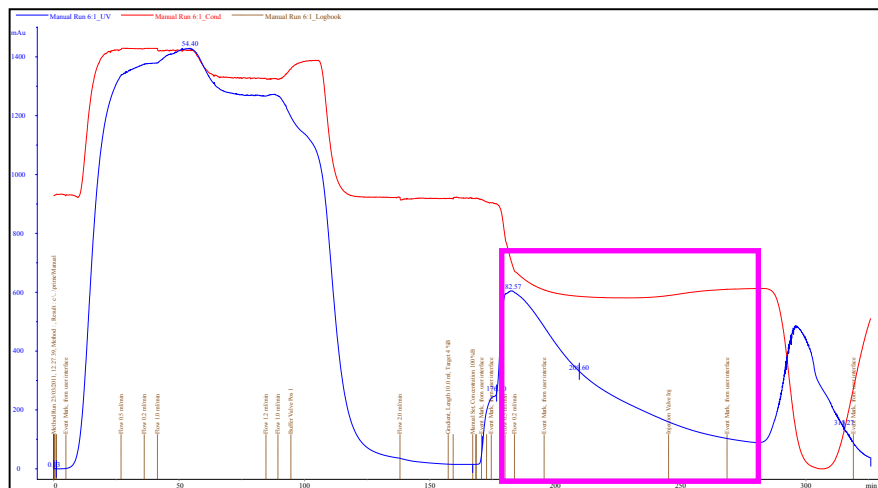


Figura 3.40 Cromatogramma ConA con evidenziazione delle frazioni contenenti la proteina di interesse.

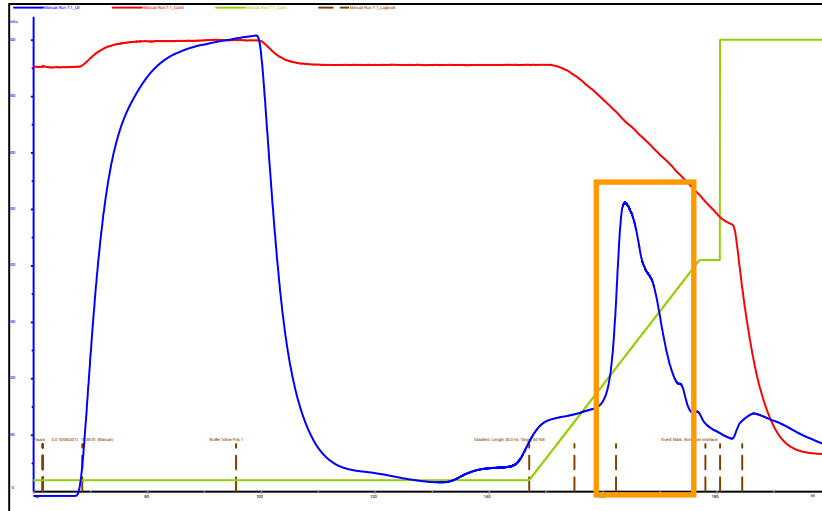


Figura 3.41 Cromatogramma HIC con evidenziazione delle frazioni contenenti la proteina di interesse.

I saggi ELISA assieme alle analisi in SDS-PAGE hanno permesso di individuare le frazioni arricchite con GAA aventi la minore quantità di proteine contaminanti. Le frazioni HIC di interesse sono state riunite e caricate in colonna di gel-filtrazione per la fase finale di purificazione (*polishing*). Di seguito (Fig. 3.42) è riportato il cromatogramma ottenuto.

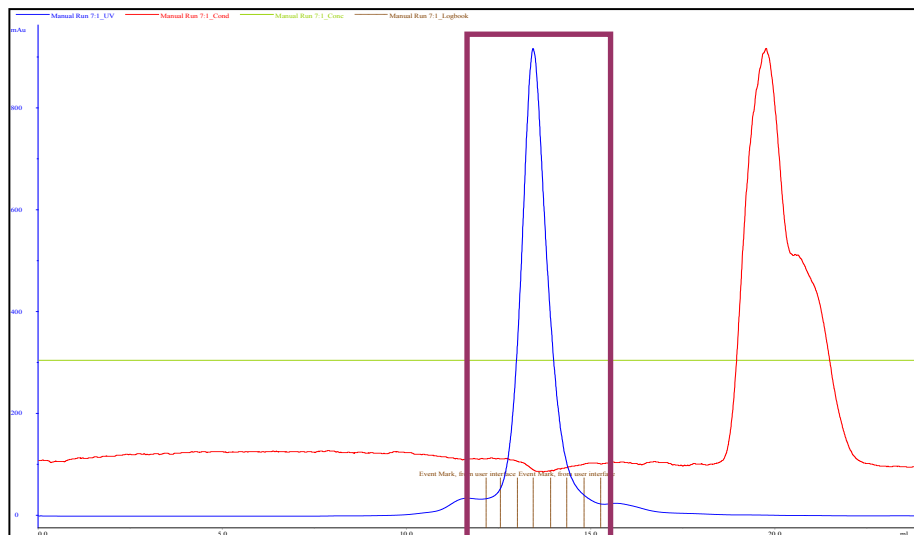


Figura 3.42 Cromatogramma GF con evidenziazione delle frazioni contenenti la proteina di interesse.

Il passaggio in colonna di gel-filtrazione ha permesso di eliminare la quasi totalità delle proteine contaminanti presenti nell'eluato HIC (Fig. 3.43).

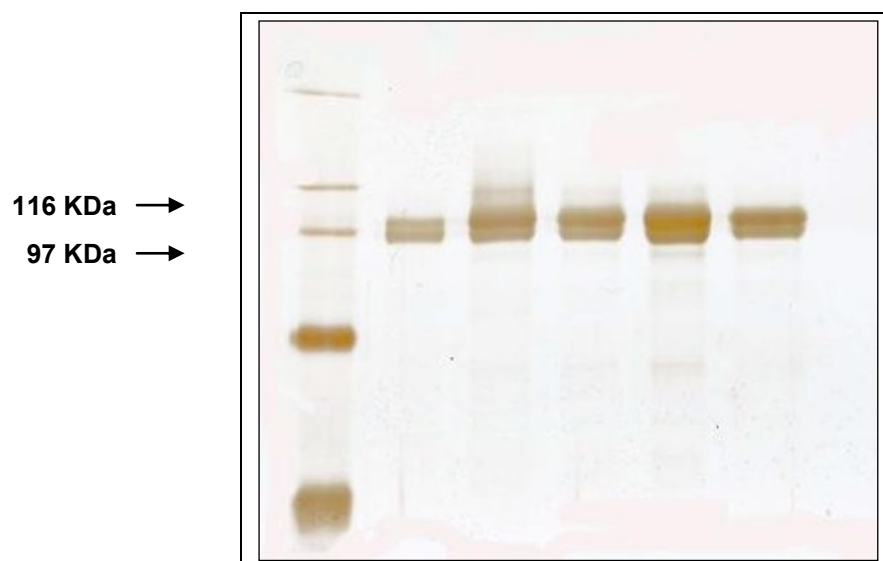


Figura 3.43 Colorazione argentea di SDS-PAGE sul campione ottenuto da gel filtrazione (Marker SDS 6D2).

SEZ. B – RISULTATI PER GCasi

3.7 Resintesi genica della GCasi

Il gene codificante la β -glucosidasi acida umana è stato tradotto, *in silico*, in proteina; successivamente, a partire dalla sequenza amminoacidica si è operata una riscrittura genica applicando sia il metodo standard di design fondato sul *codon usage* (modificato in relazione ai livelli di espressione genica e al tessuto di provenienza), sia quello fondato sul *codon context*, ovvero sulle regole di vicinanza tra codoni.

3.7.1 Riscrittura della sequenza codificante la GCasi umana secondo il metodo del *codon context*

La CDS codificante la GCasi nativa e portante il PS Glub4 già ottimizzato secondo il criterio del *codon context* è stata riscritta applicando le regole del metodo; per ciascun amminoacido sono stati, quindi, scelti i codoni sinonimi più frequenti in funzione del contesto intercodonico per geni alto espressi in seme di riso.

GCasi riscritta secondo il *codon context* con PS Glub4 sottolineato (1566 bp):

ATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCTGTGCCA
TGGCTCCATGGCCGCGCGGCCCTGCATCCCCAAGTCCTTCGGCTACTCCTCCGTTGTGT
GCGTGTGCAATGCCACCTACTGCGACTCCTTCGACCCTCCCACCTTCCCGGCGCTGGGC
ACCTTCTCCCGGTATGAGTCCACCCGGTCCGGCCGGCGGATGGAGCTGTCCATGGGCCC
CATCCAGGCCAACCACACCGGCACCGGCCTGCTGCTCACCCCTGCAGCCGGAGCAGAAGT
TCCAGAAAGTGAAAGGCTTCGGCGGCGCCATGACCGATGCCGCCGCGCTCAACATCCTG
GCGCTGTCCCCTCCGGCGCAGAACCTGCTGCTCAAGTCCTACTTCTCCGAGGAGGGCAT
TGGCTACAACATCATCCGGGTGCCCATGGCGTCCTGCGACTTCTCCATCCGGACCTACA
CCTATGCCGACACCCCGGATGACTTCCAGCTGCACAACTTCTCCCTGCCGGAGGAGGAC
ACCAAGCTCAAGATCCCTCTCATCCACCGGGCGCTGCAGCTGGCGCAGCGGCCGGTGTG
CCTGCTGGCGTCCCCCTGGACCTCCCCACCTGGCTCAAGACCAATGGCGCCGTGAATG
GCAAAGGCTCCCTCAAAGGCCAGCCGGGCGACATCTACCACCAGACCTGGGCGCGGTAC
TTCGTGAAGTTCCTGGATGCGTATGCCGAGCACAAGCTGCAGTTCTGGGCCGTGACCGC
CGAGAATGAGCCCTCCGCCGGCCTGCTGTCCGGCTACCCCTTCCAGTGCCTGGGCTTCA
CCCCGGAGCACCAGCGGGACTTCATTGCGCGGGACCTGGGCCCCACCTGGCCAACCTCC
ACCCACCACAATGTGCGGCTGCTCATGCTGGATGACCAGCGGCTGCTGCTGCCTCATTG
GGCCAAAGTTGTGCTCACCGACCCGGAGGCCGCCAAGTATGTGCATGGCATTGCCGTGC
ATTGGTACCTGGACTTCTGGCGCCGGCCAAAGCCACCCTGGGCGAGACCCACCGGCTG
TCCCCAACACCATGCTGTTTCGCGTCCGAGGCGTGCCTGGCTCCAAGTTCTGGGAGCA

GTCCGTGCGGCTGGGCTCCTGGGACCGGGGCATGCAGTACTCCCATTCCATCATCACCA
ACCTGCTGTACCATGTTGTTGGCTGGACCGACTGGAACCTGGCGCTCAACCCGGAGGGC
GGCCCCAACTGGGTGCGGAACTTCGTTGACTCCCCCATCATTGTTGACATCACCAAAGA
CACCTTCTACAAGCAGCCCATGTTCTACCACCTGGGCCATTTCTCCAAGTTCATCCCGG
AGGGCTCCCAGCGGGTTGGCCTGGTTGCGTCCCAGAAGAATGACCTGGATGCCGTTGCG
CTCATGCACCCGGATGGCTCCGCGGTTGTTGTTGTGCTCAACCGGTCTCCAAAGATGT
GCCTTCTACCATCAAAGACCCGGCCGTTGGCTTCCTGGAGACCATCTCCCCGGGCTACT
CCATCCACACCTACCTGTGGCACCGGCAGTGA

La sequenza nucleotidica così ottenuta è stata tradotta in sequenza
amminoacidica:

M A T I A F S R L S I Y F C V L L L C H G S M A A R P C I P
K S F G Y S S V V C V C N A T Y C D S F D P P T F P A L G T
F S R Y E S T R S G R R M E L S M G P I Q A N H T G T G L L
L T L Q P E Q K F Q K V K G F G G A M T D A A A L N I L A L
S P P A Q N L L L K S Y F S E E G I G Y N I I R V P M A S C
D F S I R T Y T Y A D T P D D F Q L H N F S L P E E D T K L
K I P L I H R A L Q L A Q R P V S L L A S P W T S P T W L K
T N G A V N G K G S L K G Q P G D I Y H Q T W A R Y F V K F
L D A Y A E H K L Q F W A V T A E N E P S A G L L S G Y P F
Q C L G F T P E H Q R D F I A R D L G P T L A N S T H H N V
R L L M L D D Q R L L L P H W A K V V L T D P E A A K Y V H
G I A V H W Y L D F L A P A K A T L G E T H R L F P N T M L
F A S E A C V G S K F W E Q S V R L G S W D R G M Q Y S H S
I I T N L L Y H V V G W T D W N L A L N P E G G P N W V R N
F V D S P I I V D I T K D T F Y K Q P M F Y H L G H F S K F
I P E G S Q R V G L V A S Q K N D L D A V A L M H P D G S A
V V V V L N R S S K D V P L T I K D P A V G F L E T I S P G
Y S I H T Y L W H R Q *

Nella sequenza sono stati poi individuati elementi omotetrameri CCCC e GGGG:

ATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCTGTGCCA
TGGCTCCATGGCCGCGCGGCCCTGCAT CCCC AAGTCTTCGGCTACTCCTCCGTTGTGT
GCGTGTGCAATGCCACCTACTGCGACTCCTTCGACCCTCCCACCTTCCCGGCGCTGGGC
ACCTTCTCCCGGTATGAGTCCACCCGGTCCGGCCGGCGGATGGAGCTGTCCATGGG CCC
CATCCAGGCCAACCACACCGGCACCGGCCTGCTGCTCACCTGCAGCCGGAGCAGAAGT
TCCAGAAAGTGAAAGGCTTCGGCGGCGCCATGACCGATGCCGCCGCGCTCAACATCCTG
GCGCTGT CCCC TCCGGCGCAGAACCTGCTGCTCAAGTCTACTTCTCCGAGGAGGGCAT
TGGCTACAACATCATCCGGGTGCCATGGCGTCTGCGACTTCTCCATCCGGACCTACA
CCTATGCCGACA CCCC GGATGACTTCCAGCTGCACAACTTCTCCCTGCCGGAGGAGGAC
ACCAAGCTCAAGATCCCTCTCATCCACCGGGCGCTGCAGCTGGCGCAGCGGCCGGTGT
CCTGCTGGCGT CCCC CTGGACCT CCCC CACCTGGCTCAAGACCAATGGCGCCGTGAATG
GCAAAGGCTCCCTCAAAGGCCAGCCGGGCGACATCTACCACCAGACCTGGGCGCGGTAC
TTCGTGAAGTTCCTGGATGCGTATGCCGAGCACAAGCTGCAGTTCTGGGCCGTGACCGC
CGAGAATGAGCCCTCCGCCGGCCTGCTGTCCGGCTA CCCC TTCCAGTGCCTGGGCTTCA

CCCCGGAGCACCAGCGGGACTTCATTGCGCGGGACCTGGGCCCCACCCTGGCCAACTCC
ACCCACCACAATGTGCGGCTGCTCATGCTGGATGACCAGCGGCTGCTGCTGCCTCATTG
GGCCAAAGTTGTGCTCACCGACCCGGAGGCCGCAAGTATGTGCATGGCATTGCCGTGC
ATTGGTACCTGGACTTCCTGGCGCCGGCCAAAGCCACCCTGGGCGAGACCCACCGGCTG
TTCCCCAACACCATGCTGTTTCGCGTCCGAGGCGTGCGTTGGCTCCAAGTTCTGGGAGCA
GTCCGTGCGGCTGGGCTCCTGGGACC GGGGCATGCAGTACTCCCATTCATCATCACCA
ACCTGCTGTACCATGTTGTTGGCTGGACCGACTGGAACCTGGCGCTCAACCCGGAGGGC
GGCCCCAACTGGGTGCGGAACTTCGTTGACTCCCCCATCATTGTTGACATCACCAAAGA
CACCTTCTACAAGCAGCCATGTTCTACCACCTGGGCCATTTCTCCAAGTTCATCCCGG
AGGGCTCCCAGCGGGTTGGCCTGGTTGCGTCCCAGAAGAATGACCTGGATGCCGTTGCG
CTCATGCACCCGGATGGCTCCGCCGTTGTTGTTGTGCTCAACCGGTCTCCAAAGATGT
GCCTCTCACCATCAAAGACCCGGCCGTTGGCTTCCTGGAGACCATCTCCCCGGGCTACT
CCATCCACACCTACCTGTGGCACCGGCAGTGA

Gli omotetrameri identificati sono stati eliminati attraverso l'impiego del secondo o terzo codone sinonimo scelti in base al contesto codonico risultante. Di seguito viene riportata la sequenza ottenuta sopprimendo tali elementi di disturbo:

ATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCTGTGCCA
TGGCTCCATGGCCGCGCGGCCCTGCATTCCCAAGTCTTCGGCTACTCCTCCGTTGTGT
GCGTGTGCAATGCCACCTACTGCGACTCCTTCGACCCTCCCACCTTCCCGGCGCTGGGC
ACCTTCTCCCGGTATGAGTCCACCCGGTCCGGCCGGCGGATGGAGCTGTCCATGGGTCC
CATCCAGGCCAACCACACCGGCACCGCCTGCTGCTCACCTGCAGCCGGAGCAGAAGT
TCCAGAAAGTGAAAGGCTTCGGCGGCGCCATGACCGATGCCGCCGCGCTCAACATCCTG
GCGTGAGCCCTCCGGCGCAGAACCTGCTGCTCAAGTCTACTTCTCCGAGGAGGGCAT
TGGCTACAACATCATCCGGGTGCCCATGGCGTCTGCGACTTCTCCATCCGGACCTACA
CCTATGCCGACACGCCGGATGACTTCCAGCTGCACAACTTCTCCCTGCCGGAGGAGGAC
ACCAAGCTCAAGATCCCTCTCATCCACCGGGCGCTGCAGCTGGCGCAGCGGCCGGTGTG
CCTGCTGGCGTCGCCCTGGACCTCGCCACCTGGCTCAAGACCAATGGCGCCGTGAATG
GCAAAGGCTCCCTCAAAGGCCAGCCGGGCGACATCTACCACCAGACCTGGGCGCGGTAC
TTCGTGAAGTTCTGGATGCGTATGCCGAGCACAAGCTGCAGTTCTGGGCCGTGACCGC
CGAGAATGAGCCCTCCGCCGGCCTGCTGTCCGGCTATCCCTTCCAGTGCTGGGCTTCA
CGCCGGAGCACCAGCGGGACTTCATTGCGCGGGACCTGGGTCCCACCTGGCCAACTCC
ACCCACCACAATGTGCGGCTGCTCATGCTGGATGACCAGCGGCTGCTGCTGCCTCATTG
GGCCAAAGTTGTGCTCACCGACCCGGAGGCCGCAAGTATGTGCATGGCATTGCCGTGC
ATTGGTACCTGGACTTCCTGGCGCCGGCCAAAGCCACCCTGGGCGAGACCCACCGGCTG
TTTCCCAACACCATGCTGTTTCGCGTCCGAGGCGTGCGTTGGCTCCAAGTTCTGGGAGCA
GTCCGTGCGGCTGGGCTCCTGGGACCGTGGCATGCAGTACTCCCATTCATCATCACCA
ACCTGCTGTACCATGTTGTTGGCTGGACCGACTGGAACCTGGCGCTCAACCCGGAGGGC
GGGCCAACTGGGTGCGGAACTTCGTTGACTCGCCATCATTGTTGACATCACCAAAGA
CACCTTCTACAAGCAGCCATGTTCTACCACCTGGGCCATTTCTCCAAGTTCATCCCGG
AGGGCTCCCAGCGGGTTGGCCTGGTTGCGTCCCAGAAGAATGACCTGGATGCCGTTGCG
CTCATGCACCCGGATGGCTCCGCCGTTGTTGTTGTGCTCAACCGGTCTCCAAAGATGT
GCCTCTCACCATCAAAGACCCGGCCGTTGGCTTCCTGGAGACCATCAGCCCGGGCTACT
CCATCCACACCTACCTGTGGCACCGGCAGTGA

La sequenza è stata poi analizzata con GenScan, NetGene e GeneSplicer per verificare l'assenza di siti criptici di *splicing*.

GeneSplicer (software specifico per *O. sativa*) ha individuato un possibile introne criptico:

```
Your sequence has 1566 bp.

Organism : O. sativa (Rice)

acc_Sensitivity (%) :98          acc_threshold :-0.221536
don_Sensitivity (%) :98          don_threshold :0.811606
                                Da=60, Dd=60

      End5|      End3|      Score| Confidence| Splice_site_type|
-----|-----|-----|-----|-----|
              1252|      1253|      1.148815|      Medium|      donor
```

Al fine di risolvere questo aspetto, la sequenza è stata ulteriormente modificata sempre considerando per ciascun amminoacido le scelte codoniche preferenziali per il riso (la variazione nucleotidica sinonima ottenuta impiegando il criterio del *condon context* è evidenziata in azzurro):

```
ATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCTGTGCCA
TGGCTCCATGGCCGCGCGGCCCTGCATTCCCAAGTCCTTCGGCTACTCCTCCGTTGTGT
GCGTGTGCAATGCCACCTACTGCGACTCCTTCGACCCTCCCACCTTCCCGGCGCTGGGC
ACCTTCTCCCGGTATGAGTCCACCCGGTCCGGCCGGCGGATGGAGCTGTCCATGGGTCC
CATCCAGGCCAACCACACCGGCACCGGCCTGCTGCTCACCCCTGCAGCCGGAGCAGAAGT
TCCAGAAAGTGAAAGGCTTCGGCGGCGCCATGACCGATGCCGCCGCGCTCAACATCCTG
GCGCTGAGCCCTCCGGCGCAGAACCTGCTGCTCAAGTCTACTTCTCCGAGGAGGGCAT
TGGCTACAACATCATCCGGGTGCCCATGGCGTCCCTGCGACTTCTCCATCCGGACCTACA
CCTATGCCGACACGCCGGATGACTTCCAGCTGCACAACTTCTCCCTGCCGGAGGAGGAC
ACCAAGCTCAAGATCCCTCTCATCCACCGGGCGCTGCAGCTGGCGCAGCGGCCGGTGTG
CCTGCTGGCGTCGCCCTGGACCTCGCCCACCTGGCTCAAGACCAATGGCGCCGTGAATG
GCAAAGGCTCCCTCAAAGGCCAGCCGGGCGACATCTACCACCAGACCTGGGCGCGGTAC
TTCGTGAAGTTCTGGATGCGTATGCCGAGCACAAGCTGCAGTTCTGGGCCGTGACCGC
CGAGAATGAGCCCTCCGCCGGCTGCTGTCCGGCTATCCCTTCCAGTGCCTGGGCTTCA
CGCCGGAGCACCAGCGGGACTTCATTGCGCGGGACCTGGGTCCCACCTGGCCAACTCC
ACCCACCACAATGTGCGGCTGCTCATGCTGGATGACCAGCGGCTGCTGCTGCCTCATTG
GGCAAAGTTGTGCTCACCGACCCGGAGGCCGCCAAGTATGTGCATGGCATTGCCGTGC
ATTGGTACCTGGACTTCTGGCGCCGGCCAAAGCCACCCTGGGCGAGACCCACCGGCTG
TTTCCCAACACCATGCTGTTTCGCTCCGAGGCGTGCCTGGCTCCAAGTTCTGGGAGCA
GTCCGTGCGGCTGGGCTCCTGGGACCGTGGCATGCAGTACTCCATTCCATCATCACCA
ACCTGCTGTACCATGTTGTTGGCTGGACCGACTGGAACCTGGCGCTCAACCCGGAGGGC
```

GGGCCCAACTGGGTCCGGAACCTTCGTTGACTCGCCCATCATTGTTGACATCACCAAAGA
CACCTTCTACAAGCAGCCCATGTTCTACCACCTGGGCCATTTCTCCAAGTTCATCCCGG
AGGGCTCCAGCGGGTTGGCCTGGTTGCGTCCCAGAAGAATGACCTGGATGCCGTTGCG
CTCATGCACCCGGATGGCTCCGCCGTTGTTGTTGTGCTCAACCGGTCCTCCAAAGATGT
GCCTCTACCATCAAAGACCCGGCCGTTGGCTTCCTGGAGACCATCAGCCCGGGCTACT
CCATCCACACCTACCTGTGGCACCGGCAGTGA

Ai terminali 5' e 3' della nuova sequenza sono stati rispettivamente aggiunti i siti di restrizione *Xba* I e *Sac* I. Di seguito viene riportata la sequenza sintetica definitiva della GCasi riscritta secondo il metodo del *codon context* per geni alto espressi in seme di riso con i siti di restrizione ai terminali:

TCTAGAATGGCCACCATTGCGTTCTCCCGGCTGTCCATCTACTTCTGCGTGCTGCTGCT
GTGCCACGGCTCCATGGCCGCGCGGCCCTGCATTCCCAAGTCCCTTCGGCTACTCCTCCG
TTGTGTGCGTGTGCAATGCCACCTACTGCGACTCCTTCGACCCTCCCACCTTCCCGGCG
CTGGGCACCTTCTCCCGGTATGAGTCCACCCGGTCCGGCCGGCGGATGGAGCTGTCCAT
GGGTCCCATCCAGGCCAACACACCGGCACCGGCCTGCTGCTCACCTGCAGCCGGAGC
AGAAGTTCAGAAAGTGAAAGGCTTCGGCGGCGCCATGACCGATGCCGCCGCGCTCAAC
ATCCTGGCGCTGAGCCCTCCGGCGCAGAACCTGCTGCTCAAGTCCTACTTCTCCGAGGA
GGGCATTGGCTACAACATCATCCGGGTGCCATGGCGTCCTGCGACTTCTCCATCCGGA
CCTACACCTATGCCGACACGCCGGATGACTTCCAGCTGCACAACTTCTCCCTGCCGGAG
GAGGACACCAAGCTCAAGATCCCTTCTCATCCACCGGGCGCTGCAGCTGGCGCAGCGGCC
GGTGTCCCTGCTGGCGTCGCCCTGGACCTCGCCACCTGGCTCAAGACCAATGGCGCCG
TGAATGGCAAAGGCTCCCTCAAAGGCCAGCCGGGCGACATCTACCACCAGACCTGGGCG
CGGTACTTCGTGAAGTTCCTGGATGCGTATGCCGAGCACAAGCTGCAGTTCGGGCCGT
GACCGCCGAGAATGAGCCCTCCGCCGGCCTGCTGTCCGGCTATCCCTTCCAGTGCCTGG
GCTTCACGCCGGAGACCAGCGGGACTTCATTGCGCGGGACCTGGGTCCCACCCTGGCC
AACTCCACCCACCACAATGTGCGGCTGCTCATGCTGGATGACCAGCGGCTGCTGCTGCC
TCATTGGGCCAAAGTTGTGCTCACCGACCCGGAGGCCGCAAGTATGTGCATGGCATTG
CCGTGCATTGGTACCTGGACTTCCTGGCGCCGGCCAAAGCCACCCTGGGCGAGACCCAC
CGGCTGTTTCCCAACACCATGCTGTTGCGTCCGAGGCGTGCCTGGCTCCAAGTTCG
GGAGCAGTCCGTGCGGCTGGGCTCCTGGGACCGTGGCATGCAGTACTCCCATTCATCA
TCACCAACCTGCTGTACCATGTTGTTGGCTGGACCGACTGGAACCTGGCGCTCAACCCG
GAGGGCGGGCCCAACTGGGTCCGGAACCTTCGTTGACTCGCCCATCATTGTTGACATCAC
CAAAGACACCTTCTACAAGCAGCCCATGTTCTACCACCTGGGCCATTTCTCCAAGTTC
TCCCGGAGGGCTCCCAGCGGGTTGGCCTGGTTGCGTCCCAGAAGAATGACCTGGATGCC
GTTGCGCTCATGCACCCGGATGGCTCCGCCGTTGTTGTTGTGCTCAACCGGTCCTCCAA
AGATGTGCCTCTCACCATCAAAGACCCGGCCGTTGGCTTCCTGGAGACCATCAGCCCGG
GCTACTCCATCCACACCTACCTGTGGCACCGGCAGTGA**GAGCTC**

Traduzione:

M A T I A F S R L S I Y F C V L L L C H G S M A A R P C I P
K S F G Y S S V V C V C N A T Y C D S F D P P T F P A L G T
F S R Y E S T R S G R R M E L S M G P I Q A N H T G T G L L
L T L Q P E Q K F Q K V K G F G G A M T D A A A L N I L A L
S P P A Q N L L L K S Y F S E E G I G Y N I I R V P M A S C

D F S I R T Y T Y A D T P D D F Q L H N F S L P E E D T K L
K I P L I H R A L Q L A Q R P V S L L A S P W T S P T W L K
T N G A V N G K G S L K G Q P G D I Y H Q T W A R Y F V K F
L D A Y A E H K L Q F W A V T A E N E P S A G L L S G Y P F
Q C L G F T P E H Q R D F I A R D L G P T L A N S T H H N V
R L L M L D D Q R L L L P H W A K V V L T D P E A A K Y V H
G I A V H W Y L D F L A P A K A T L G E T H R L F P N T M L
F A S E A C V G S K F W E Q S V R L G S W D R G M Q Y S H S
I I T N L L Y H V V G W T D W N L A L N P E G G P N W V R N
F V D S P I I V D I T K D T F Y K Q P M F Y H L G H F S K F
I P E G S Q R V G L V A S Q K N D L D A V A L M H P D G S A
V V V V L N R S S K D V P L T I K D P A V G F L E T I S P G
Y S I H T Y L W H R Q *

3.7.2 Riscrittura della sequenza codificante la GCasi umana secondo il metodo del *codon usage*

La sequenza codificante la β -glucosidasi acida umana nativa è stata rivista sulla base del *codon usage* di *Oryza sativa* e dei dati ottenuti dal lavoro di analisi di sequenze nucleotidiche codificanti per geni altamente espressi in seme. Per migliorare il costrutto di espressione di *GCasi* in riso alla sequenza della proteina matura si è anteposto il peptide della glutelina 4 e si è proceduto alla sintesi artificiale.

GCasi riscritta con sottolineato il PS Glub4 ottimizzato secondo il *codon usage* (1566 bp)

ATGGCCACCATCGCCTTCAGCAGGCTCAGCATCTACTTCTGCGTGCTCCTCCTTGCCA
CGGCAGCATGGCCGCCAGGCCTTGCATCCCTAAGAGCTTCGGCTACAGCAGCGTGGTGT
GCGTGTGCAACGCCACCTACTGCGACAGCTTCGACCCTCCTACCTTCCCTGCCCTCGGC
ACCTTCAGCAGGTACGAGAGCACCAGGAGCGGCAGGAGGATGGAGCTCAGCATGGGCC
TATCCAGGCCAACCACACCGGCACCGGCCTCCTCCTCACCCTCCAGCCTGAGCAGAAGT
TCCAGAAGGTGAAGGGCTTCGGCGGCGCCATGACCGACGCCGCCCTCAACATCCTC
GCCCTCAGCCCTCCTGCCAGAACCTCCTCCTCAAGAGCTACTTCAGCGAGGAGGGCAT
CGGCTACAACATCATCAGGGTGCCTATGGCCAGCTGCGACTTCAGCATCAGGACCTACA
CCTACGCCGACACCCCTGACGACTTCCAGCTCCACAACTTCAGCCTCCCTGAGGAGGAC
ACCAAGCTCAAGATCCCTCTCATCCACAGGGCCCTCCAGCTCGCCCAGAGGCCTGTGAG
CCTCCTCGCCAGCCCTTGACCAGCCCTACCTGGCTCAAGACCAACGGCGCCGTGAACG
GCAAGGGCAGCCTCAAGGGCCAGCCTGGCGACATCTACCACCAGACCTGGGCCAGGTAC
TTCGTGAAGTTCTCGACGCCTACGCCGAGCACAAGCTCCAGTTCTGGGCCGTGACCGC
CGAGAACGAGCCTAGCGCCGGCCTCCTCAGCGGCTACCCTTTCCAGTGCCTCGGCTTCA
CCCCTGAGCACCAGAGGGACTTCATCGCCAGGGACCTCGGCCCTACCCTCGCCAACAGC
ACCCACCACAACGTGAGGCTCCTCATGCTCGACGACCAGAGGCTCCTCCTCCCTCACTG
GGCCAAGGTGGTGCCTACCGACCCTGAGGCCGCCAAGTACGTGCACGGCATCGCCGTGC
ACTGGTACCTCGACTTCTCGCCCTGCCAAGGCCACCCTCGGCGAGACCCACAGGCTC

TTCCCTAACACCATGCTCTTCGCCAGCGAGGCCTGCGTGGGCAGCAAGTTCTGGGAGCA
GAGCGTGAGGCTCGGCAGCTGGGACAGGGGCATGCAGTACAGCCACAGCATCATCACCA
ACCTCCTCTACCACGTGGTGGGCTGGACCGACTGGAACCTCGCCCTCAACCCTGAGGGC
GGCCCTAACTGGGTGAGGAACTTCGTGGACAGCCCTATCATCGTGGACATCACCAAGGA
CACCTTCTACAAGCAGCCTATGTTCTACCACCTCGGCCACTTCAGCAAGTTCATCCCTG
AGGGCAGCCAGAGGGTGGGCCTCGTGGCCAGCCAGAAGAACGACCTCGACGCCGTGGCC
CTCATGCACCCTGACGGCAGCGCCGTGGTGGTGGTCAACAGGAGCAGCAAGGACGT
GCCTCTCACCATCAAGGACCCTGCCGTGGGCTTCCTCGAGACCATCAGCCCTGGCTACA
GCATCCACACCTACCTCTGGCACAGGCAGTGA

Traduzione:

M A T I A F S R L S I Y F C V L L L C H G S M A A R P C I P
K S F G Y S S V V C V C N A T Y C D S F D P P T F P A L G T
F S R Y E S T R S G R R M E L S M G P I Q A N H T G T G L L
L T L Q P E Q K F Q K V K G F G G A M T D A A A L N I L A L
S P P A Q N L L L K S Y F S E E G I G Y N I I R V P M A S C
D F S I R T Y T Y A D T P D D F Q L H N F S L P E E D T K L
K I P L I H R A L Q L A Q R P V S L L A S P W T S P T W L K
T N G A V N G K G S L K G Q P G D I Y H Q T W A R Y F V K F
L D A Y A E H K L Q F W A V T A E N E P S A G L L S G Y P F
Q C L G F T P E H Q R D F I A R D L G P T L A N S T H H N V
R L L M L D D Q R L L L P H W A K V V L T D P E A A K Y V H
G I A V H W Y L D F L A P A K A T L G E T H R L F P N T M L
F A S E A C V G S K F W E Q S V R L G S W D R G M Q Y S H S
I I T N L L Y H V V G W T D W N L A L N P E G G P N W V R N
F V D S P I I V D I T K D T F Y K Q P M F Y H L G H F S K F
I P E G S Q R V G L V A S Q K N D L D A V A L M H P D G S A
V V V V L N R S S K D V P L T I K D P A V G F L E T I S P G
Y S I H T Y L W H R Q *

La sequenza è stata analizzata per la ricerca di introni criptici; GeneSplicer ha individuato possibili siti criptici di *splicing*, come riportato nella figura seguente:

```
Your sequence has 1566 bp.  
Organism : O. sativa (Rice)  
acc_Sensitivity (%) :98          acc_threshold :-0.221536  
don_Sensitivity (%) :98          don_threshold :0.811606  
Da=60, Dd=60
```

End5	End3	Score	Confidence	Splice_site_type
-----	-----	-----	-----	-----
1275	1274	2.388000	Medium	acceptor
1342	1343	2.822243	Medium	acceptor

Per ovviare a questo problema si è deciso di modificare la sequenza riscritta tramite l'uso del secondo codone più frequente (la variazione nucleotidica sinonima ottenuta è evidenziata in azzurro):

```
ATGGCCACCATCGCCTTCAGCAGGCTCAGCATCTACTTCTGCGTGCTCCTCCTCTGCCA
CGGCAGCATGGCCGCCAGGCCTTGCATCCCTAAGAGCTTCGGCTACAGCAGCGTGGTGT
GCGTGTGCAACGCCACCTACTGCGACAGCTTCGACCCTCCTACCTTCCCTGCCCTCGGC
ACCTTCAGCAGGTACGAGAGCACCAGGAGCGGCAGGAGGATGGAGCTCAGCATGGGCCC
TATCCAGGCCAACCACACCGGCACCGGCCTCCTCCTCACCTCCAGCCTGAGCAGAAGT
TCCAGAAGGTGAAGGGCTTCGGCGGCATGACCGACGCCGCCCTCAACATCCTC
GCCCTCAGCCCTCCTGCCAGAACCTCCTCCTCAAGAGCTACTTCAGCGAGGAGGGCAT
CGGCTACAACATCATCAGGGTGCCTATGGCCAGCTGCGACTTCAGCATCAGGACCTACA
CCTACGCCGACACCCCTGACGACTTCCAGCTCCACAACCTCAGCCTCCCTGAGGAGGAC
ACCAAGCTCAAGATCCCTCTCATCCACAGGGCCCTCCAGCTCGCCCAGAGGCCTGTGAG
CCTCCTCGCCAGCCCTTGGACCAGCCCTACCTGGCTCAAGACCAACGGCGCCGTGAACG
GCAAGGGCAGCCTCAAGGGCCAGCCTGGCGACATCTACCACCAGACCTGGGCCAGGTAC
TTCGTGAAGTTCTCGACGCCTACGCCGAGCACAAGCTCCAGTTCTGGGCCGTGACCGC
CGAGAACGAGCCTAGCGCCGGCCTCCTCAGCGGCTACCCTTTCAGTGCCTCGGCTTCA
CCCCTGAGCACCAGAGGGACTTCATCGCCAGGGACCTCGGCCCTACCCTCGCCAACAGC
ACCCACCACAACGTGAGGCTCCTCATGCTCGACGACCAGAGGCTCCTCCTCCCTCACTG
GGCCAAGGTGGTGCCTACCGACCCTGAGGCCGCCAAGTACGTGCACGGCATCGCCGTGC
ACTGGTACCTCGACTTCTCGCCCCTGCCAAGGCCACCCTCGGCGAGACCCACAGGCTC
TTCCTAACACCATGCTCTTCGCCAGCGAGGCTGCGTGGGCAGCAAGTTCTGGGAGCA
GAGCGTGAGGCTCGGCAGCTGGGACAGGGGCATGCAGTACAGCCACAGCATCATCACCA
ACCTCCTTACCACGTGGTGGGCTGGACCGACTGGAACCTCGCCCTCAACCCTGAGGGC
GGCCCTAACTGGGTGAGGAACCTTCGTGGACAGCCCGATCATCGTGGACATCACCAAGGA
CACCTTCTACAAGCAGCCTATGTTCTACCACCTCGGCCACTTTCCAAGTTCATCCCTG
AGGGCAGCCAGAGGGTGGGCCCTCGTGGCCAGCCAGAAGAACGACCTCGACGCCGTGGCC
CTCATGCACCCTGACGGCAGCGCCGTGGTGGTGGTGCCTCAACAGGAGCAGCAAGGACGT
GCCTCTCACCATCAAGGACCCTGCCGTGGGCTTCTCCTCGAGACCATCAGCCCTGGCTACA
GCATCCACACCTACCTCTGGCACAGGCAGTGA
```

La sequenza così ottenuta è stata nuovamente analizzata con GeneSplicer per verificare l'eliminazione dei siti critici di splicing identificati:

```
Your sequence has 1566 bp.

Organism : O. sativa (Rice)

acc_Sensitivity (%) :98                acc_threshold :-0.221536
don_Sensitivity (%) :98                don_threshold :0.811606
                                Da=60, Dd=60

End5|      End3|      Score| Confidence| Splice_site_type|
-----|-----|-----|-----|-----|
```


Per la sub-clonazione della CDS riscritta nel vettore di espressione, sono stati aggiunti i siti di restrizione *Xba* I e *Sac* I, rispettivamente al terminale 5' e 3' della nuova sequenza. Dal momento che l'analisi di restrizione ha evidenziato la presenza di un ulteriore sito *Sac* I all'interno della versione sintetica, si è deciso di modificare la sequenza di riconoscimento dell'enzima di restrizione impiegando il secondo codone sinonimo (evidenziato in rosso nella sequenza in basso).

Di seguito si riporta la sequenza artificiale definitiva e completa:

```
TCTAGAATGGCCACCATCGCCTTCAGCAGGCTCAGCATCTACTTCTGCGTGCTCCTCCT
CTGCCACGGCAGCATGGCCGCCAGGCCTTGCATCCCTAAGAGCTTCGGCTACAGCAGCG
TGGTGTGCGTGTGCAACGCCACCTACTGCGACAGCTTCGACCCTCCTACCTCCCTGCC
CTCGGCACCTTCAGCAGGTACGAGAGCACCAGGAGCGGCAGGAGGATGSAACTCAGCAT
GGGCCCTATCCAGGCCAACACACCCGGCACCCGGCCTCCTCCTCACCTCCAGCCTGAGC
AGAAGTTCCAGAAGGTGAAGGGCTTCGGCGGCGCCATGACCGACGCCGCCGCCCTCAAC
ATCCTCGCCCTCAGCCCTCCTGCCCAGAACCTCCTCCTCAAGAGCTACTTCAGCGAGGA
GGGCATCGGCTACAACATCATCAGGGTGCCTATGGCCAGCTGCGACTTCAGCATCAGGA
CCTACACCTACGCCGACACCCCTGACGACTTCCAGCTCCACAACCTTCAGCCTCCCTGAG
GAGGACACCAAGCTCAAGATCCCTCTCATCCACAGGGCCCTCCAGCTCGCCCAGAGGCC
TGTGAGCCTCCTCGCCAGCCCTTGGACCAGCCCTACCTGGCTCAAGACCAACGGCGCCG
TGAACGGCAAGGGCAGCCTCAAGGGCCAGCCTGGCGACATCTACCACCAGACCTGGGCC
AGGTACTTCGTGAAGTTCTCGACGCCTACGCCGAGCACAAGCTCCAGTTCTGGGCCGT
GACCGCCGAGAACGAGCCTAGCGCCGGCCTCCTCAGCGGCTACCCTTTCCAGTGCCTCG
GCTTCACCCCTGAGCACCAGAGGGACTTCATCGCCAGGGACCTCGGCCCTACCCTCGCC
AACAGCACCCACCACAACGTGAGGCTCCTCATGCTCGACGACCAGAGGCTCCTCCTCCC
TCACTGGGCCAAGGTGGTGCTCACCGACCCTGAGGCCGCCAAGTACGTGCACGGCATCG
CCGTGCACTGGTACCTCGACTTCCTCGCCCCTGCCAAGGCCACCCTCGGCGAGACCCAC
AGGCTCTTCCCTAACACCATGCTCTTCGCCAGCGAGGCCTGCGTGGGCAGCAAGTTCTG
GGAGCAGAGCGTGAGGCTCGGCAGCTGGGACAGGGGCATGCAGTACAGCCACAGCATCA
TCACCAACCTCCTCTACCACGTGGTGGGCTGGACCGACTGGAACCTCGCCCTCAACCT
GAGGGCGGCCCTAACTGGGTGAGGAACCTTCGTGGACAGCCCGATCATCGTGGACATCAC
CAAGGACACCTTCTACAAGCAGCCTATGTTCTACCACCTCGGCCACTTCTCCAAGTTCA
TCCCTGAGGGCAGCCAGAGGGTGGGCCTCGTGGCCAGCCAGAAGAACGACCTCGACGCC
GTGGCCCTCATGCACCCTGACGGCAGCGCCGTGGTGGTGGTGGTCAACAGGAGCAGCAA
GGACGTGCCTCTCACCATCAAGGACCCTGCCGTGGGCTTCCTCGAGACCATCAGCCCTG
GCTACAGCATCCACACCTACCTCTGGCACAGGCAGTGAGAGCTC
```

3.7.3 Analisi della sequenza

Le due versioni sintetiche (GCasi CC e GCasi CU) sono state analizzate per valutare nel dettaglio la composizione strutturale di ciascuna sequenza. Confrontando le CDS artificiali con quella nativa si ottengono i seguenti risultati:

ANALISI	GCasi NATIVA	GCasi CC	GCasi CU
% identità		82	80
% GC totale	55.81	63.28	63.66
di cui			
% GC in terza posizione		29.24	31.09
% codoni terminanti in G o C	67.04	87.73	93.29
% TA totale	44.19	36.72	36.34
CpG totale	44	107	88
TpA totale	47	20	31
TATA	1	0	0
AATAAA	0	0	0
ATTTA	0	0	0
AAA	10	9	0
AAAA	3	0	0
TTT	11	2	1
TTTT	1	0	0
CCC	37	38	37
CCCC	11	0	3
GGG	23	24	23
GGGG	5	0	1
Intercodoni CG	16	45	88
Intercodoni TA	14	0	11

Tabella 3.3 Risultati ottenuti attraverso l'analisi della sequenza GAA nativa, riscritta secondo il metodo del *codon context* (CC) e del *codon usage* (CU).

3.8 Realizzazione dei vettori di espressione per GCasi

3.8.1 pCAMBIA1300_GCasi nativa_PMI e pCAMBIA1300_GCasi nativa_PMI_FxAntiFx

Il gene PMI (*manA*) precedentemente isolato è stato (cfr. § 3.2.1) inserito nel vettore intermedio ottenuto dalla digestione di pCAMBIA1300 e del plasmide portante la cassetta completa per l'espressione della β -glucosidasi nativa (promotore GluB4-GCasi nativa- terminatore NOS ter) con *EcoR* I (Fig. 3.45). Successivamente alla *ligation* e alla trasformazione di cellule JM101 di *E. coli*, le colonie trasformate sono state selezionate mediante PCR. La Fig. 3.44 mostra gli amplificati ottenuti utilizzando primer specifici per GCasi.

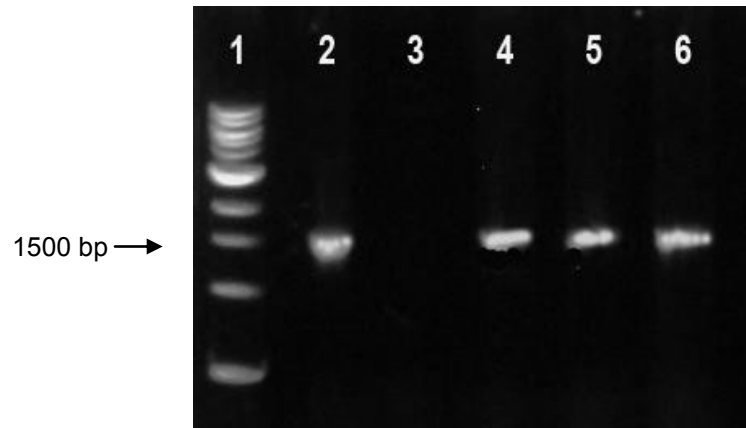


Figura 3.44 Corsia 1: Ladder 1 kb (NEB); 2-6: colonie saggiate mediante PCR.

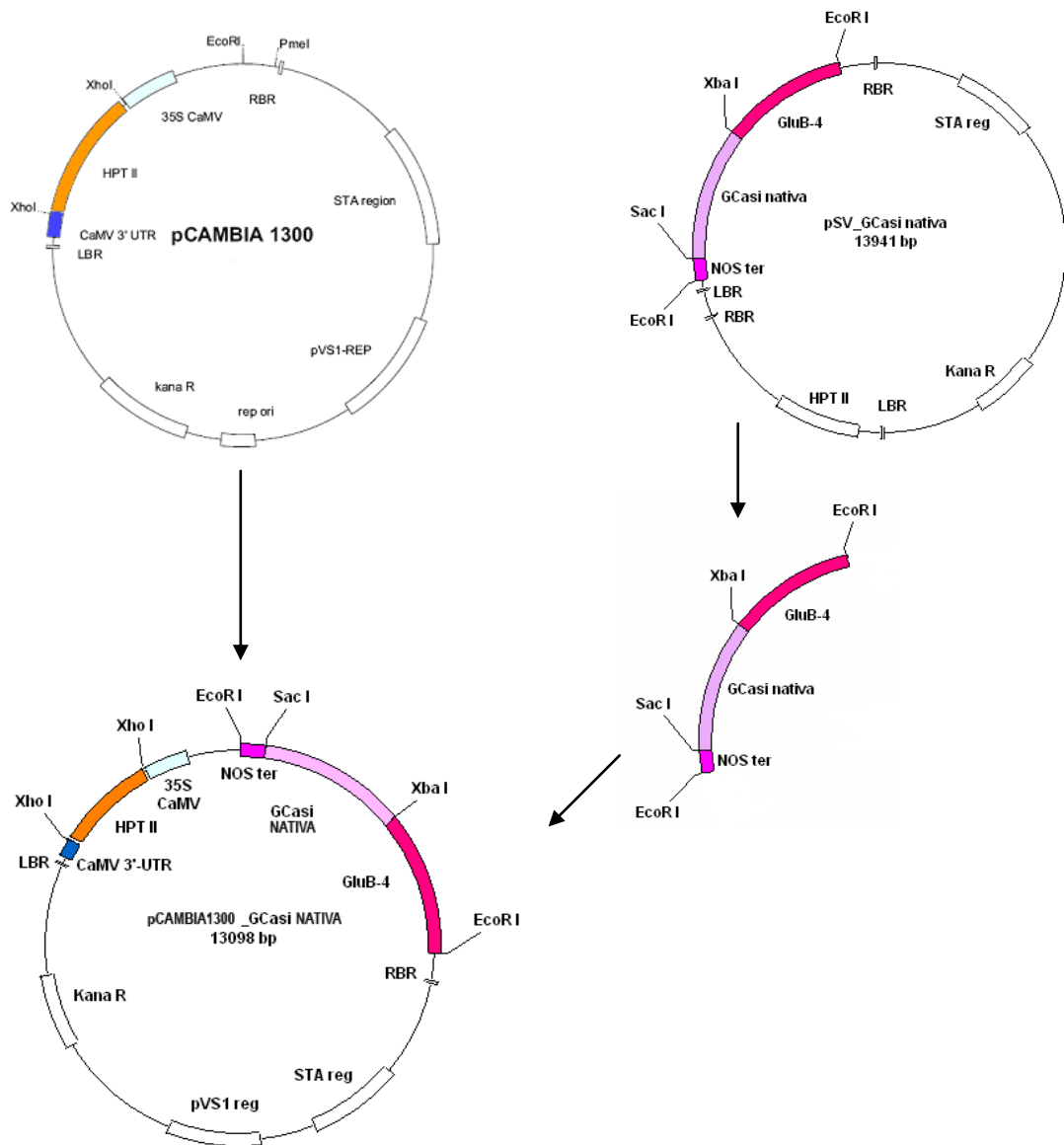


Figura 3.45 Realizzazione del vettore pCambia1300_GCasi nativa.

Come ulteriore controllo, due colonie sono state verificate mediante digestioni enzimatiche con *Xba* I ed *EcoR* I (Fig. 3.46).

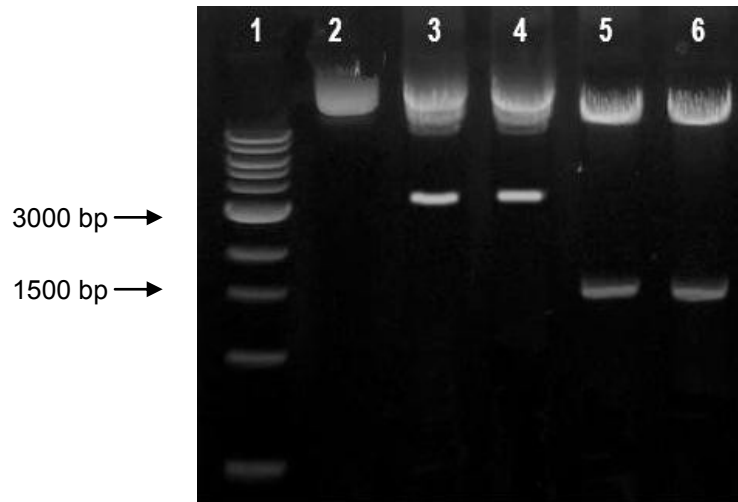


Figura 3.46 Corsia 1: Ladder 1 kb (NEB); 2: campione non digerito; 3-4: digestione con *EcoR* I; 5-6: digestione con *Xba* I.

Un ulteriore controllo mediante digestione con *Sac* I ha consentito di verificare l'orientamento della cassetta inserita. In caso di orientamento corretto i due siti *Sac* I sono distanti tra loro e generano un frammento di ~3104 bp; se viceversa l'orientamento non è quello desiderato, i due siti sono vicini tra loro rendendo un frammento di 276 bp. La Fig. 3.47 mostra i profili di restrizione ottenuti che sono risultati corrispondere a quelli attesi in caso di corretto orientamento della cassetta.

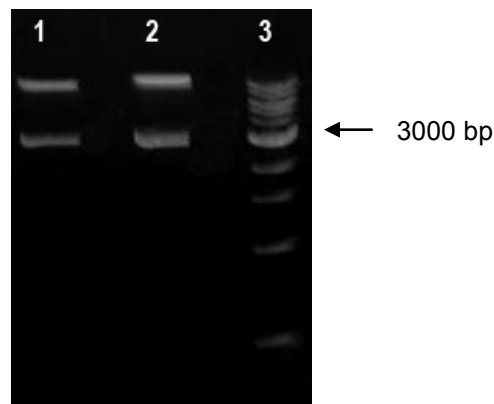
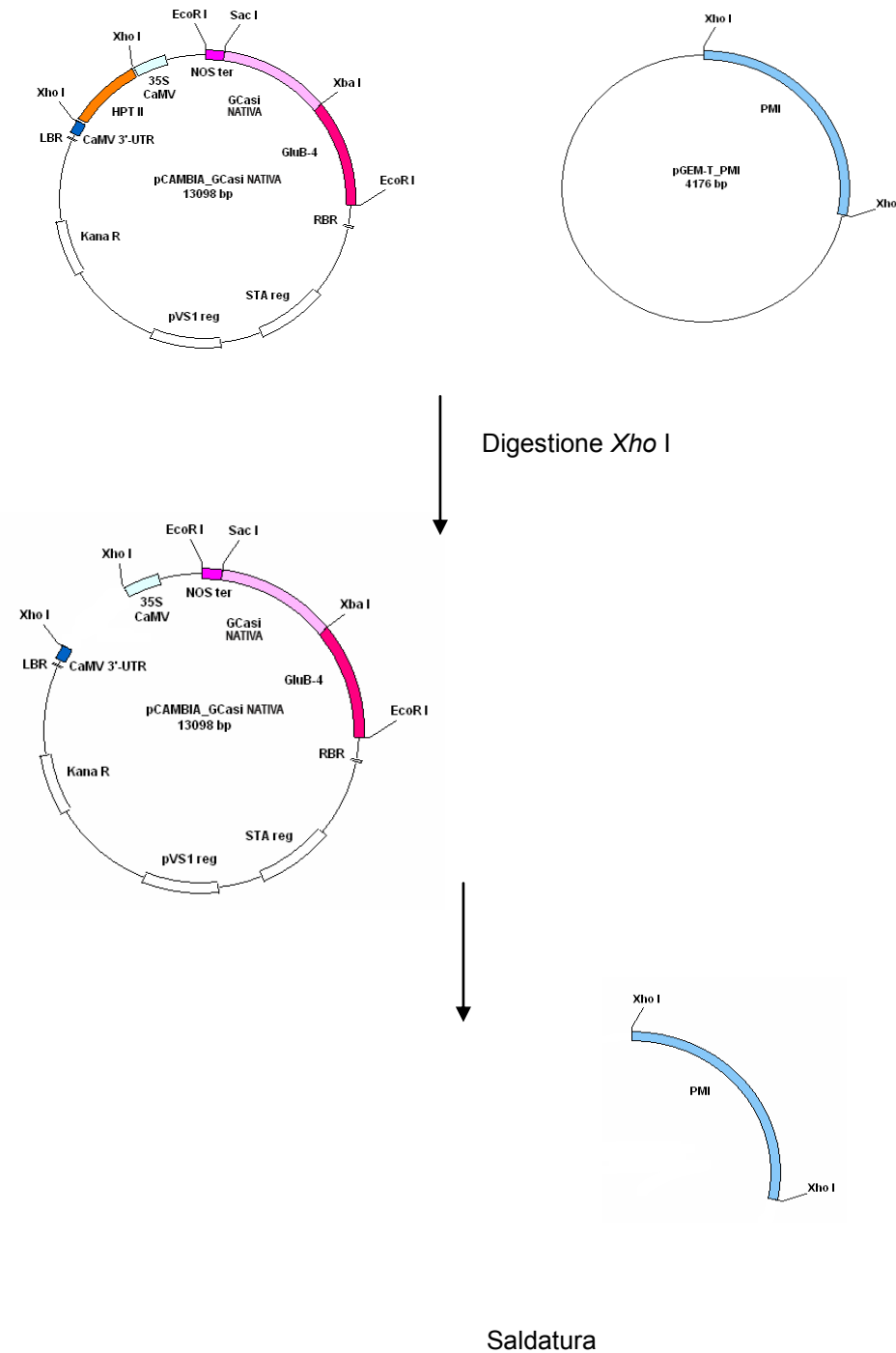


Figura 3.47 Corsie 1-2: digestione con *Sac* I; 3: Ladder 1 kb (NEB).

Nel vettore pCAMBIA1300_GCasi nativa così ottenuto si è proceduto all'introduzione del gene PMI precedentemente sintetizzato e clonato in pGEM-T mediante digestione con *Xho* I. Tale operazione ha consentito la rimozione del gene *hptII* da pCAMBIA1300_GCasi nativa_NOS ter e la sua sostituzione con il gene PMI a saldatura avvenuta (Fig. 3.48).



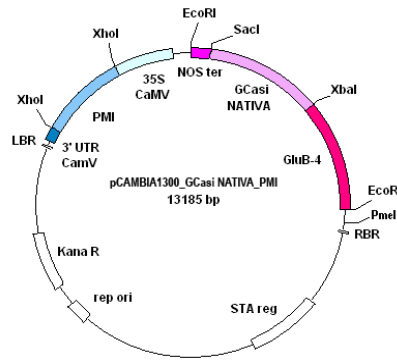


Figura 3.48 Realizzazione del vettore finale a due cassette pCAMBIA_GCasi nativa_PMI.

Il costrutto pCAMBIA1300_GCasi nativa-NOS ter_PMI così ottenuto è stato controllato attraverso digestioni enzimatiche e utilizzato per la trasformazione di *A. tumefaciens* EHA 105 effettuata mediante elettroporazione.

La verifica del costrutto trasferito in agrobatterio è avvenuta su due colonie trasformate (1 e 2) sia tramite digestioni di controllo con gli enzimi *EcoR* I, *EcoR* V, *Xho* I, *Pme* I (Fig. 3.49), sia mediante PCR.

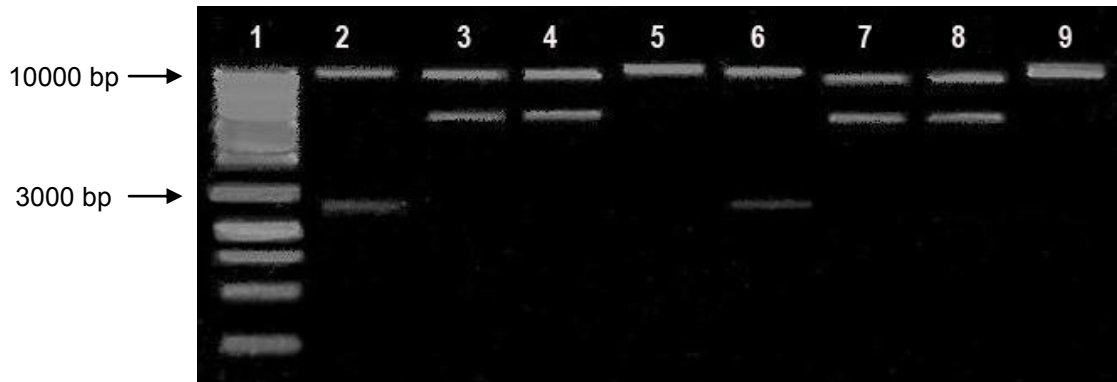


Figura 3.49 Corsia 1: Ladder 1 kb (NEB); 2: colonia 1 digestione *Xho* I; 3: colonia 1 digestione *EcoR* V; 4: colonia 1 digestione *EcoR* I; 5: colonia 1 digestione *Pme* I; 6: colonia 2 digestione *Xho* I; 7: colonia 2 digestione *EcoR* V; 8: colonia 2 digestione *EcoR* I; 9: colonia 1 digestione *Pme* I.

In Fig. 3.50 sono mostrati gli amplificati ottenuti impiegando primer specifici per il gene PMI (PMI for/PMI rev) e GCasi (riceTPO for/GLU rev). Solo la colonia 1 di agrobatterio è risultata positiva all'analisi PCR.

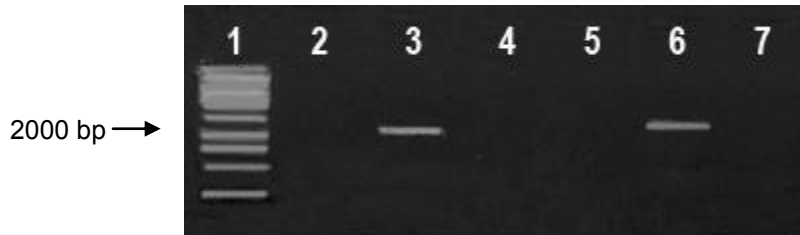


Figura 3.50 Corsia 1: Ladder 1 kb (NEB); 2: acqua; 3: colonia 1 saggiata con PMI for/PMI rev; 4: colonia 2 saggiata con PMI for/PMI rev; 5: acqua; 6: colonia 1 saggiata con riceTPO for/GLU rev; 7: colonia 2 saggiata con riceTPO for/GLU rev.

Nel vettore pCAMBIA_GCasi nativa-NOS ter_PMI è stata introdotta la terza cassetta (FxAntiFx), secondo la medesima procedura già descritta precedentemente nel caso della GAA (cfr. § 3.2.4), producendo il vettore finale a tre cassette (Fig. 3.51).

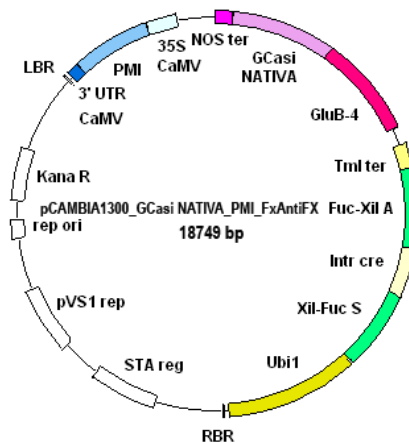


Figura 3.51 Rappresentazione schematica del vettore finale a tre cassette pCAMBIA1300_GCasi nativa_PMI_FxAntiFx.

Quattro colonie sono state analizzate per verificare il corretto orientamento della cassetta inserita attraverso digestione con *Spe I* (Fig. 3.52) e nuovamente controllate mediante PCR (Fig. 3.53) con primer specifici (PMI for/PMI rev; riceTPO for/GLU rev). Due colonie hanno dato risultati positivi a entrambi i controlli; una di queste è stata scelta per la trasformazione di agrobatterio.

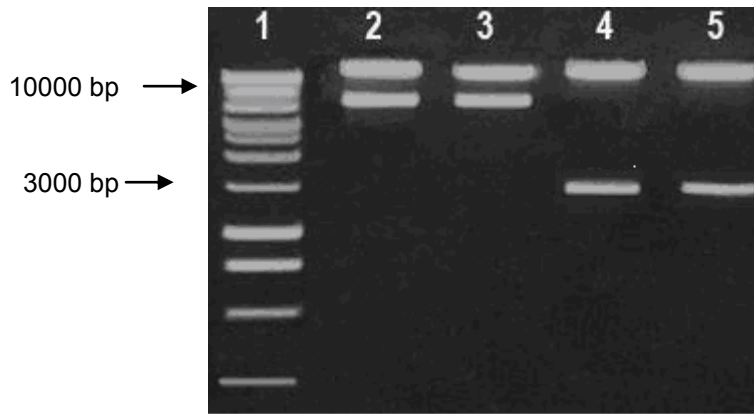


Figura 3.52 Corsia 1: Ladder 1 kb (NEB); 2: colonia 1 digestione *Spe I*; 3: colonia 2 digestione *Spe I*; 4: colonia 3 digestione *Spe I*; 5: colonia 4 digestione *Spe I*.

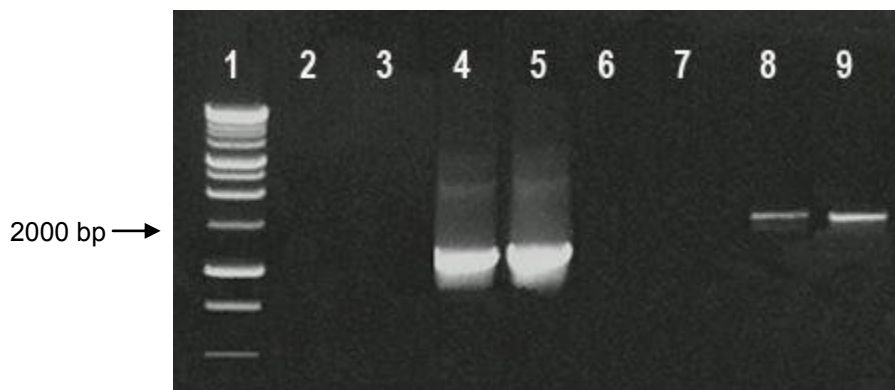


Figura 3.53 Corsia 1: Ladder 1 kb (NEB); 2: colonia 1 saggiata con PMI for/PMI rev; 3: colonia 2 saggiata con PMI for/PMI rev; 4: colonia 3 saggiata con PMI for/PMI rev; 5: colonia 4 saggiata con PMI for/PMI rev; 6: colonia 1 saggiata con riceTPO for/GLU rev; 7: colonia 2 saggiata con riceTPO for/GLU rev; 8: colonia 3 saggiata con riceTPO for/GLU rev; 9: colonia 4 saggiata con riceTPO for/GLU rev.

3.8.2 pCAMBIA1300_GCasi CC-NOS ter_PMI_FxantiFX e pCAMBIA1300_GCasi CU-Nos ter_PMI_FxantiFX

Per la realizzazione dei vettori finali di espressione portanti le versioni sintetiche del gene GCasi (GCasi CC e GCasi CU) si è proceduto per sostituzione della CDS codificante la β -glucosidasi nativa nel vettore preparativo pSV_GCasi mediante digestione con *Xba I* e *Sac I*. Gli inserti corrispondenti alle sequenze GCasi CC e GCasi CU sono state recuperate dai plasmidi, in cui sono stati clonati dalla ditta incaricata della loro sintesi, attraverso digestione congiunta con i medesimi enzimi. La presenza delle sequenze riscritte è stata verificata

mediante digestioni con *Xho* I e *Sma* I. La scelta di operare tali digestioni enzimatiche va ricercata nel fatto che la sequenza CC e CU portano rispettivamente *Sma* I e *Xho* I come sito aggiuntivo. La Figura 3.54 evidenzia i profili di restrizione ottenuti: due colonie putativamente trasformate con il plasmide pSV_GCasi CC hanno mostrato i risultati attesi mentre la digestione di tutte e tre le colonie pSV_GCasi CU ha dato esito positivo.



Figura 3.54 Corsia 1: Ladder 1 kb (Promega); 2-4: colonie pSV_GCasi CC saggate mediante digestione con *Xho* I; 5-7: colonie pSV_GCasi CU saggate mediante digestione con *Xho* I; 8-10: colonie pSV_GCasi CC saggate mediante digestione con *Sma* I; 11-13: colonie pSV_GCasi CU saggate mediante digestione con *Sma* I.

L'integrità dei siti terminali (*Xba* I e *Sac* I) è stata confermata tramite digestione doppia con gli stessi enzimi (Fig. 3.55).

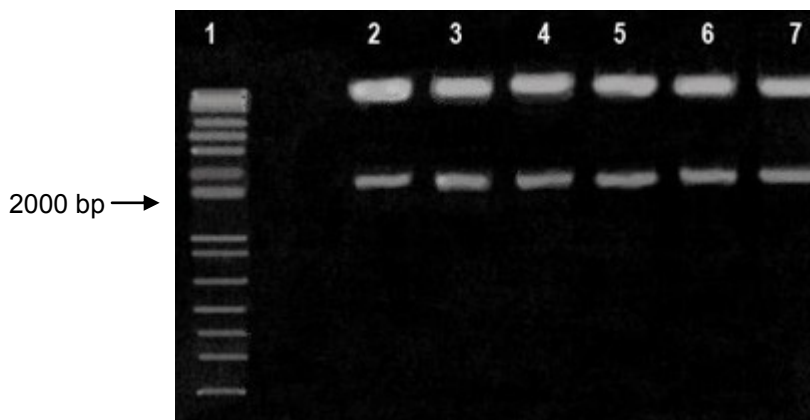


Figura 3.55 Corsia 1: Ladder 1 kb (Invitrogen); 2-4: colonie pSV_GCasi CC digerite con *Xba* I e *Sac* I; 5-7: colonie pSV_GCasi CU digerite con *Xba* I e *Sac* I.

I plasmidi portanti le CDS CC e CU sono stati sottoposti a digestione con *EcoR* I in modo da recuperare gli inserti e consentirne l'introduzione nel vettore pCAMBIA1300_GCasi nativa_PMI anch'esso tagliato con *EcoR* I. Successivamente alla saldatura e alla trasformazione di cellule JM101 di *E. coli*, sono state eseguite digestioni di controllo con i seguenti enzimi di restrizione: *Xho* I, *Sma* I, *EcoR* I (Fig. 3.56).

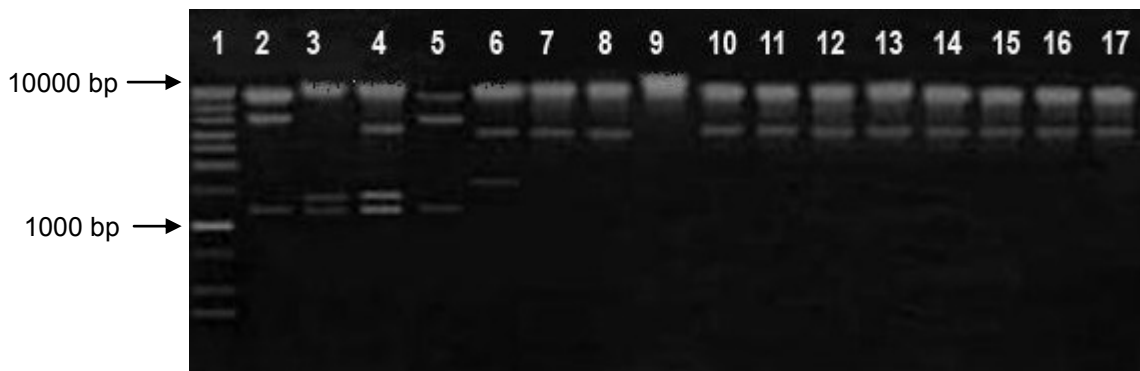


Figura 3.56 Corsia 1: Ladder 1 kb (Promega); 2-5: colonie pCAMBIA1300_GCasi CU_PMI digerite con *Xho* I; 6-9: colonie pCAMBIA1300_GCasi CC_PMI digerite con *Sma* I; 10-13: pCAMBIA1300_GCasi CC_PMI digerite con *EcoR* I; 13-17: pCAMBIA1300_GCasi CU_PMI digerite con *EcoR* I.

Successivamente sono stati prodotti i vettori finali a tre cassette tramite inserimento della cassetta per il silenziamento rispettivamente nel vettore pCAMBIA1300_GCasi CC_PMI e pCAMBIA1300_GCasi CU_PMI. I costrutti finali pCAMBIA1300_GCasi CC_PMI_FxAntiFx e pCAMBIA1300_GCasi CU_PMI_FxAntiFx (Fig. 3.57) sono stati, quindi, utilizzati per la trasformazione di *A. tumefaciens* EHA 105 effettuata mediante elettroporazione.

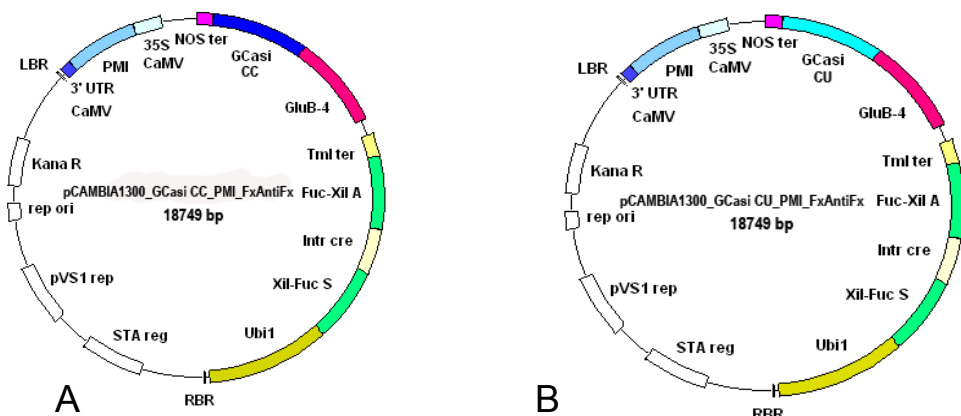


Figura 3.57 Rappresentazione schematica dei vettori finali pCAMBIA_GCasi CC_PMI_FxAntiFx (A) e pCAMBIA_GCasi CU_PMI_FxAntiFx (B).

3.9 Analisi sulla progenie

L'analisi DAS-ELISA sul seme prodotto dai trasformati primari, ottenuti dalla trasformazione con i diversi vettori di espressione, ha permesso di valutare l'effetto delle tre sequenze sinonime (nativa, riscritta CC e CU) sui livelli di espressione della proteina ricombinante e, successivamente, di individuare le linee aventi il maggior contenuto di enzima nell'endosperma di riso. In particolare, il migliore espressore GCasi è stato scelto per la semina al fine di ottenere la successiva generazione di piante T₂.

A questo scopo è stato sviluppato un protocollo di selezione basato sull'estrazione di proteine totali da seme e sulla quantificazione della proteina di interesse mediante saggio immunologico DAS-ELISA.

3.9.1 Prove di estrazione di proteine totali da semi GCasi

Alla luce delle evidenze ottenute negli esperimenti di estrazione proteica impiegando due diversi sistemi di macinazione (cfr. § 3.5.1), si è deciso di allestire ulteriori prove variando la velocità delle oscillazioni/minuto del mulino a sfere. In particolare, sono state impiegate velocità di 10, 20 e 30 oscillazioni/minuto, indicate rispettivamente V1, V2, V3. Come riportato in Grafico 3.7, è stato osservata una riduzione della capacità di estrazione proteica al diminuire della velocità impiegata per la macinazione. Infatti, il maggior contenuto di GCasi è stato rilevato in campioni macinati utilizzando lo strumento alla velocità più elevata testata (V3), che è risultata essere anche la velocità massima utilizzabile, oltre la quale la capacità di estrazione viene quasi totalmente abbattuta (dati non mostrati).

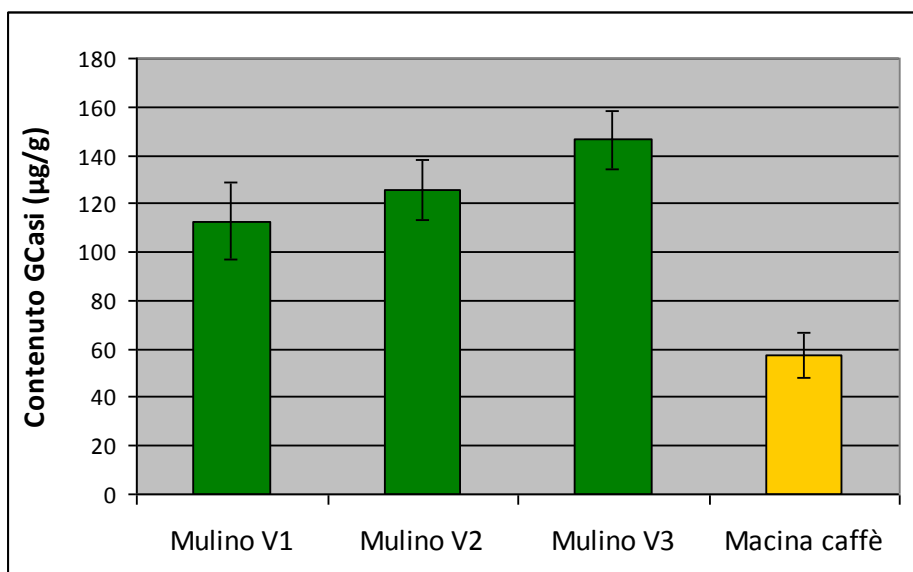


Grafico 3.7 Contenuto di GCasi in campioni sottoposti a macinazione con mulino a sfere a tre diverse velocità di oscillazioni/minuto e macina caffè.

3.9.2 Validazione del saggio DAS-ELISA

Analogamente a quanto descritto nel paragrafo 3.5.2, è stato allestito un esperimento per la validazione del saggio DAS-ELISA impiegato nella determinazione del contenuto di proteina GCasi nel seme. Diluizioni 1:20, 1:30, 1:40 dell'estratto tal quale sono state confrontate con le medesime diluizioni del campione cui è stato aggiunto lo standard (Cerezyme) a concentrazione nota (Tab. 3.4).

Diluizioni	[] campione tal quale senza aggiunta standard	[] campione con aggiunta standard	Valore atteso
1:20	0.97	1.88	1.77 +/- 0.265
1:30	1.07	2.00	1.87 +/- 0.280
1:40	1.08	1.85	1.88 +/- 0.282

Tabella 3.4 Concentrazioni dell'estratto GCasi diluito 1:20, 1:30, 1:40 rilevate in assenza/presenza dello standard a concentrazione nota e valori attesi nel range di validità (+/- 15%).

Come si osserva nel Grafico 3.8, le concentrazioni ottenute alle diverse diluizioni sono state pressoché le stesse; inoltre, i dati relativi ai campioni trattati con lo standard sono ricaduti all'interno dell'intervallo di validità (+/- 15%) stabilito come indice di non interferenza.

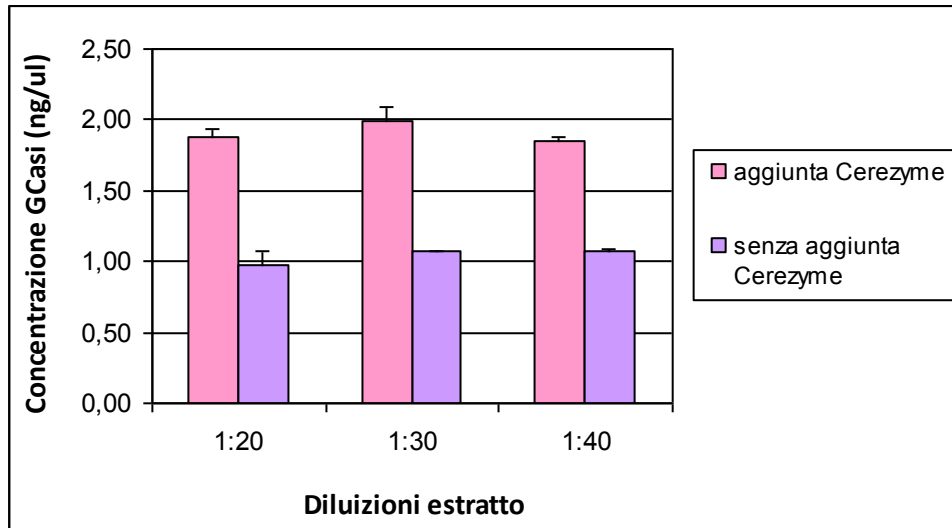
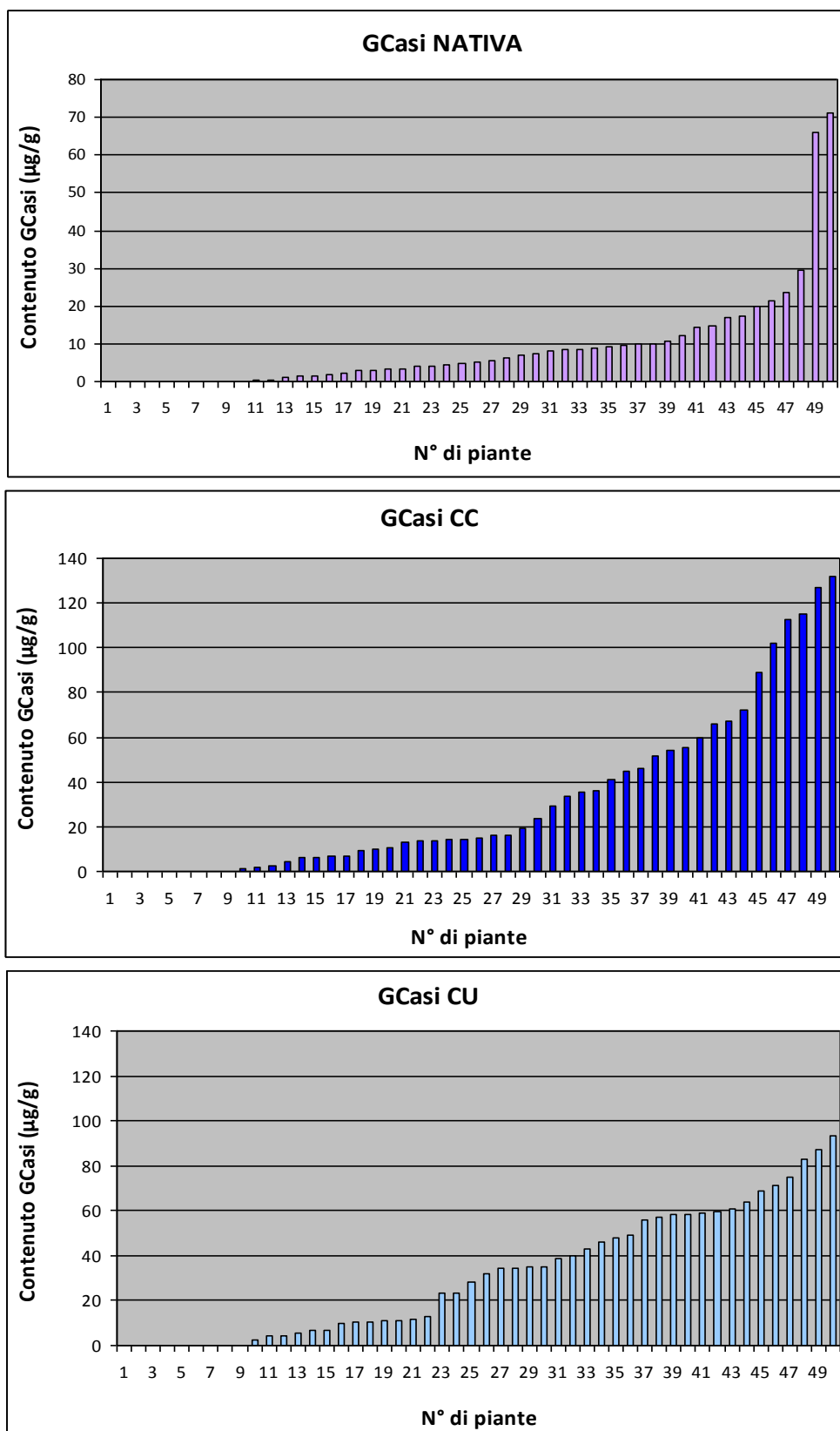


Grafico 3.8 Confronto tra le concentrazioni di GCasi ottenute a tre diverse diluizioni dell'estratto nel campione tal quale e dopo aggiunta dello standard in quantità nota (Cerezyme).

3.9.3 Valutazione dell'effetto della riscrittura genica sui livelli di espressione della proteina ricombinante

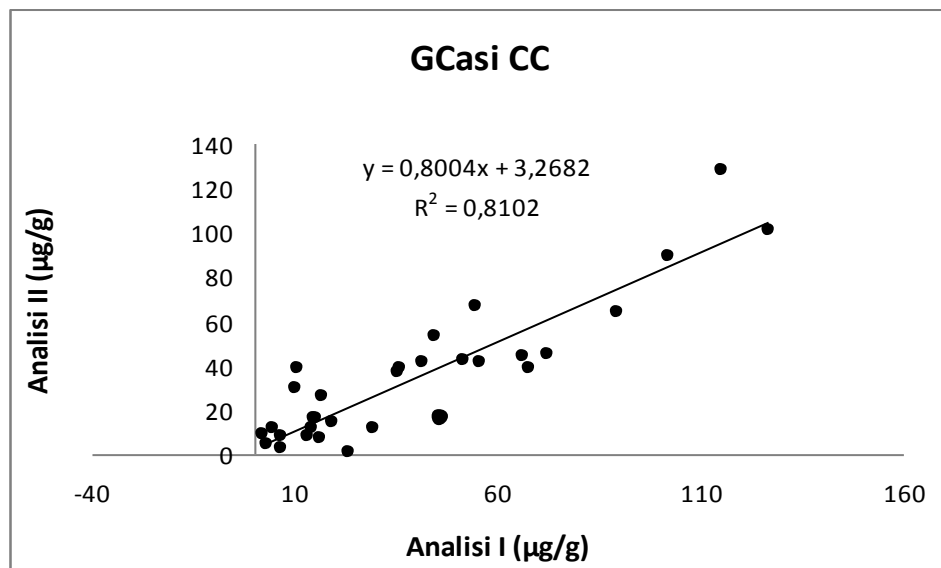
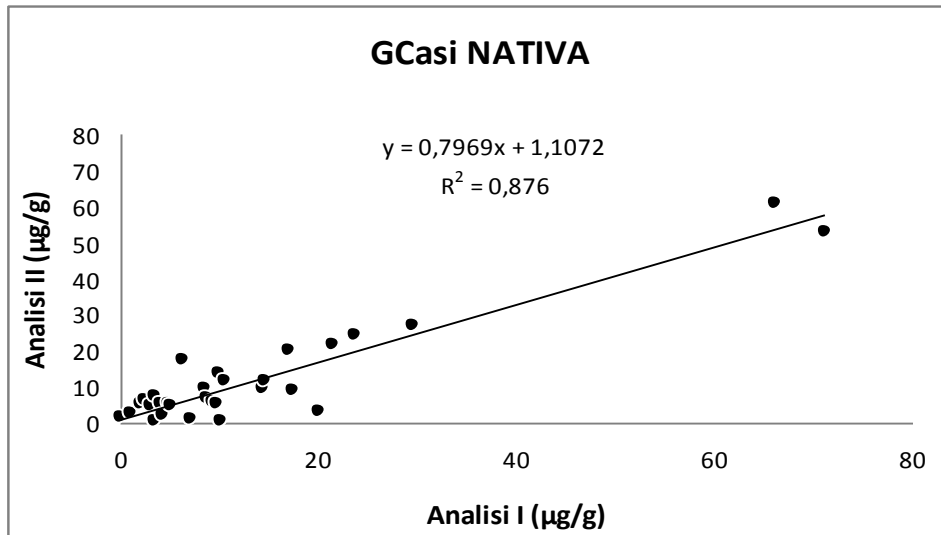
Per valutare sperimentalmente i miglioramenti dei livelli di espressione apportabili tramite design genico in silico, le progenie di 50 trasformati primari ottenuti con i costrutti pCAMBIA1300_GCasi nativa_PMI_FxAntiFx, pCAMBIA 1300_GCasi CC e pCAMBIA1300_GCasi CU_FxAntiFx sono state analizzate mediante estrazione di proteine totali e successivo saggio immunologico DAS-ELISA. Le linee considerate sono state scelte a caso (cioè indipendentemente da qualsiasi criterio specifico, quali ad es. periodo di raccolta, entità della produzione, posizione all'interno dell'ambiente di coltivazione).

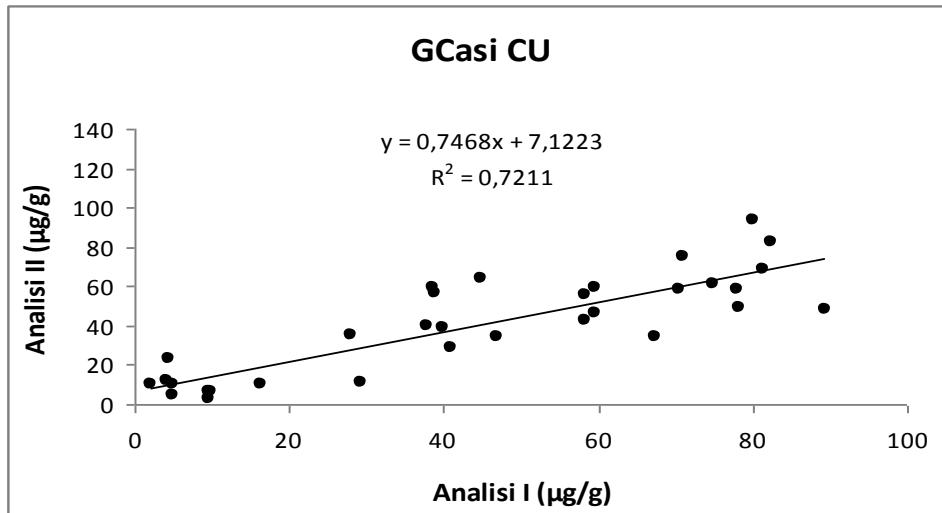
Di seguito si riportano i dati ottenuti, ordinati in senso crescente, per ciascuna popolazione di trasformati primari (Graf. 3.9). Il contenuto di enzima ricombinante nel seme è espresso come microgrammi di proteina di interesse per grammo di farina di riso ottenuta tramite macinazione.



Grafici 3.9 Contenuto in GCasi rilevato nel seme prodotto da 50 trasformati primari scelti a caso e ottenuti con i costrutti a tre cassette portanti rispettivamente le sequenze GCasi nativa, GCasi CC, GCasi CU.

Per riconfermare l'attendibilità del metodo, l'intera procedura di estrazione e valutazione in ELISA è stata ripetuta sui medesimi 50 trasformati primari, analizzando statisticamente i risultati ottenuti nei due saggi indipendenti. Dall'analisi statistica sono stati esclusi i campioni che in almeno uno dei due saggi avevano evidenziato un contenuto in GCasi inferiore a 2 pg/ μ L, in quanto valore estremo della curva di taratura. Come evidenziato in Graf. 3.10, 3.11, 3.12, è stata osservata una buona ripetibilità dei dati analitici. In particolare, circa l'88%, l'81% e il 72% della variazione osservata rispettivamente nei dati relativi alla progenie GCasi nativa, Gcasi CC e GCasi CU, è risultata essere spiegata in base a una correlazione dei valori ottenuti in due estrazioni separate; inoltre, la pendenza della retta di regressione (coefficiente angolare b) non è apparsa significativamente diversa da 1.





Grafici 3.10, 3.11, 3.12. Analisi della regressione condotta sulle serie di dati ottenuti per le medesime piante GCasi nativa, GCasi CC e GCasi CU in due estrazioni separate.

Al fine di determinare l'effetto delle tre sequenze codificanti sinonime sul livello massimo di espressione della proteina, è stato confrontato il contenuto medio di enzima ricombinante ottenuto dalle due analisi indipendenti nelle tre popolazioni transgeniche analizzate come riportato in Grafico 3.13. I dati ottenuti hanno evidenziato un aumento dei livelli di proteina nel seme dei trasformati primari GCasi CC, confermando sperimentalmente il miglioramento nell'espressione endosperma-specifica realizzabile attraverso la riscrittura *in silico* secondo il metodo del *codon context*.

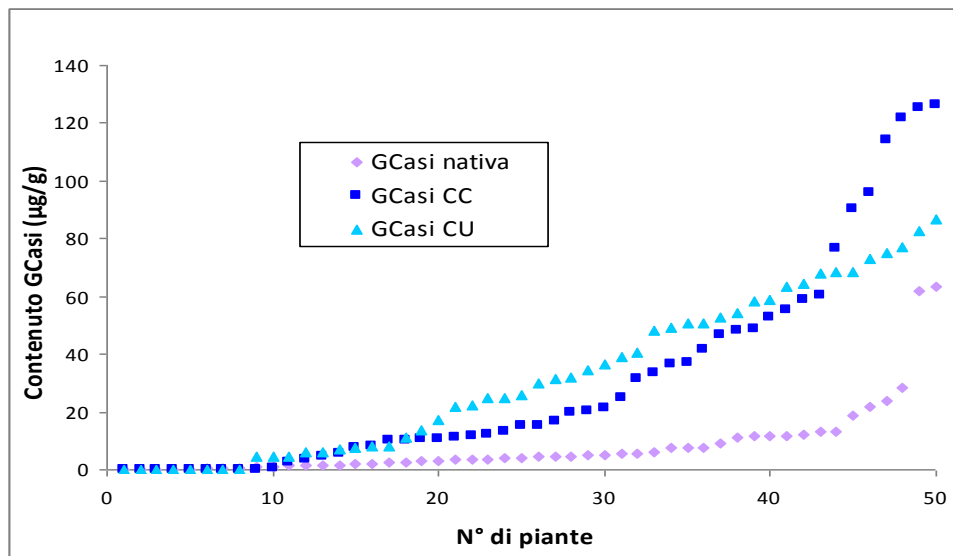


Grafico 3.13 Confronto dei livelli di espressione nelle popolazioni GCasi nativa, GCasi CC, GCasi CU.

3.9.4 Selezione del migliore espressore GCasi CC

Sulla base delle evidenze sperimentali ottenute dal confronto dei livelli di espressione delle tre sequenze sinonime, si è deciso di estendere l'analisi all'intero numero di trasformati primari GCasi CC allo scopo di identificare il miglior espressore la cui progenie potesse essere impiegata per produrre la successiva generazione T₂; 40 semi di ciascuna pianta sono stati, quindi, sottoposti a estrazione proteica e saggio DAS-ELISA. La selezione è stata condotta su un totale di 114 linee; il seme di 37 piante è risultato pressoché privo di GCasi (nell'estratto saggiato in ELISA erano, infatti, presenti concentrazioni inferiori al valore minimo utilizzato nella costruzione della curva di taratura, ovvero 2 pg/ μ L) mentre il contenuto di proteina di interesse è risultato variare notevolmente tra le restanti linee come si osserva nel seguente grafico.

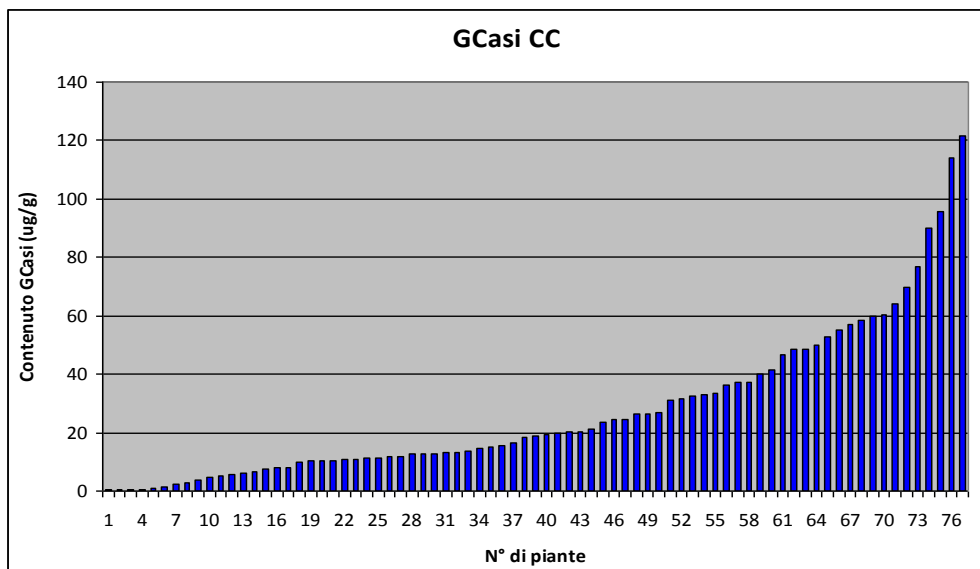
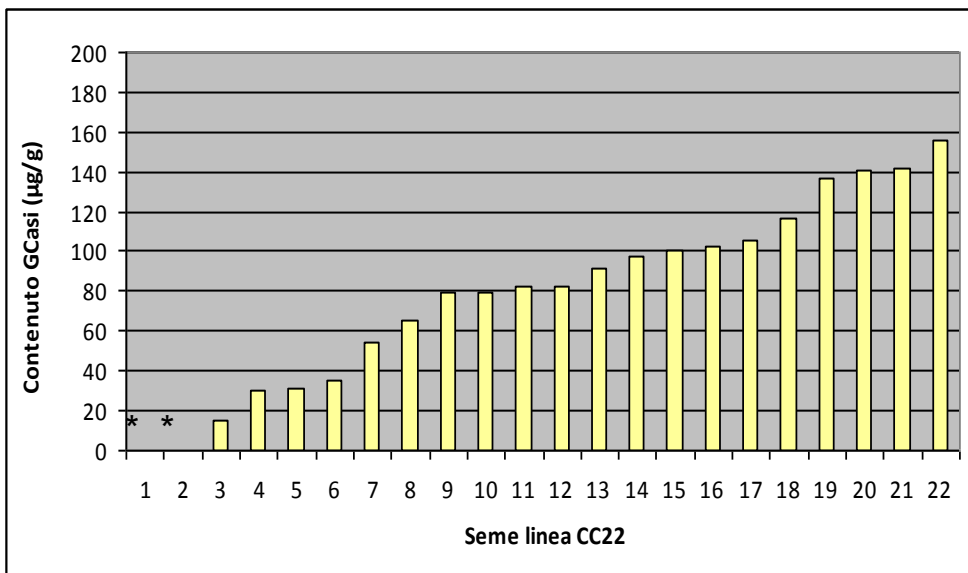
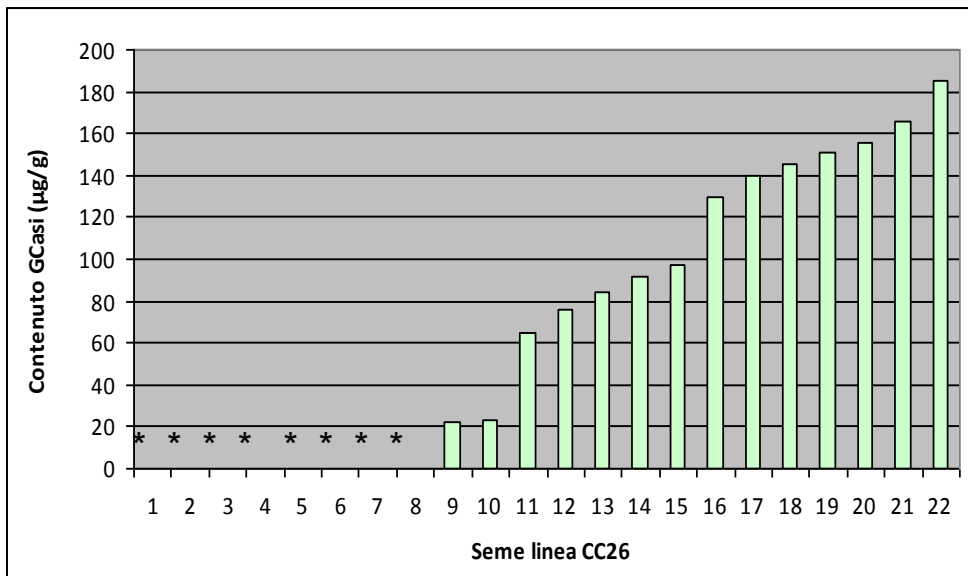


Grafico 3.14 Contenuto di GCasi rilevato nel seme prodotto dai trasformati primari GCasi CC.

Considerata la netta superiorità di due trasformati primari (contenuto GCasi maggiore di 100 μ g/g), essi sono stati sottoposti a esami più approfonditi per una valutazione del contributo offerto dai singoli semi all'ammontare complessivo di enzima ricombinante presente nei rispettivi lotti. A tal fine, dai trasformati primari CC22 e CC26 sono stati isolati 22 semi, analizzati singolarmente. Anche in questo caso, per valutare il contenuto di GCasi, gli estratti proteici ottenuti dai singoli semi sono stati analizzati in DAS-ELISA.

I risultati ottenuti nelle due linee sono riportati in ordine crescente nei Grafici 3.15 e 3.16. Appare evidente la presenza in ciascun gruppo di una variabilità di espressione imputabile alla segregazione del/i transgene/i. E' inoltre facilmente osservabile una distribuzione dei dati maggiormente discreta nel caso della linea CC26.



Grafici 3.15 e 3.16. Risultati ottenuti su estratti proteici da seme singolo delle linee CC22 e CC26. La presenza di concentrazioni di GCasi inferiori al valore minimo utilizzato nella costruzione della curva di taratura (2 pg/µL) è indicata con un asterisco.

Dati i risultati ottenuti, la linea identificata per la semina è stata il trasformato primario CC26. Tale linea, infatti, ha presentato una distribuzione dei dati maggiormente discreta a livello di analisi del singolo seme e il più alto livello di espressione di GCasi.

3.9.5 Identificazione delle piante T₂ altamente produttive

Dalla semina della linea GCasi CC26 è stata generata una popolazione di piante T₂ che sono state analizzate al fine identificare gli individui altamente produttivi. La selezione delle linee GCasi CC26 T₂ è stata condotta su un totale di 250 piante. Di queste 53 non hanno evidenziato espressione della proteina di interesse. Per una semplificazione grafica di seguito vengono rappresentati i livelli di enzima ricombinante nel seme relativo ai 50 migliori espressoristi (Graf. 3.17) sui quali successivamente è stato eseguita l'analisi su seme singolo. Le piante T₂ altamente produttive hanno mostrato un contenuto di β -glucosidasi acida variabile, compreso tra 150 e 270 $\mu\text{g/g}$, evidenziando un significativo incremento nella produzione di proteina ricombinante rispetto alla generazione precedente.

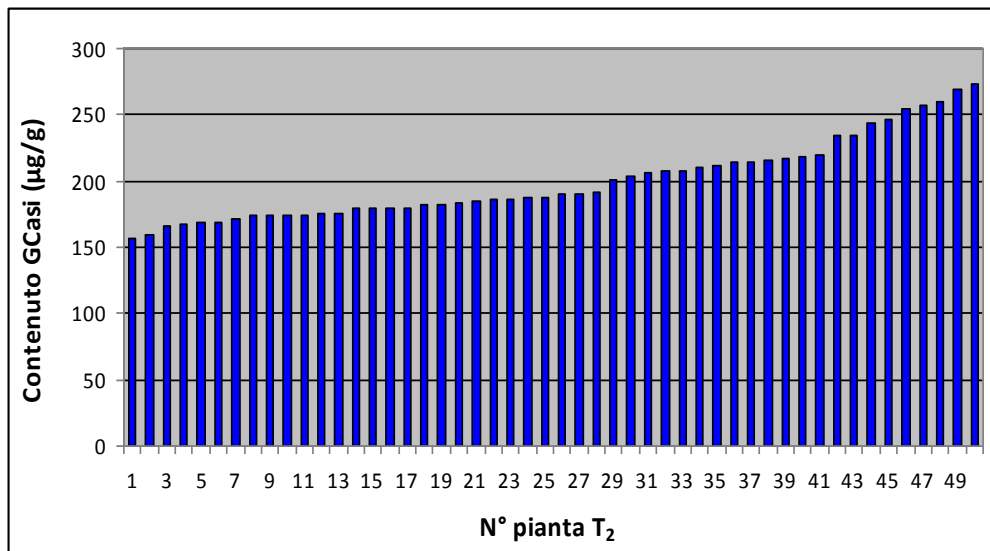


Grafico 3.17 Contenuto GCasi nel seme dei 50 migliori espressoristi T₂.

Dopo aver effettuato la selezione delle piante T₂ altamente produttive, sulle medesime è stato determinato il contenuto di beta-glucosidasi acida mediante analisi su seme singolo.

A titolo esemplificativo vengono riportati nei Grafici. 3.18 e 3.19 i risultati relativi all'analisi degli estratti proteici ottenuti da 12 singoli semi valutati in DAS-ELISA per 2 individui T₂ che hanno presentato semi privi di proteina di interesse (*null*), indicati con un asterisco, e 2 individui T₂ per i quali si è osservata una maggiore stabilità del transgene.

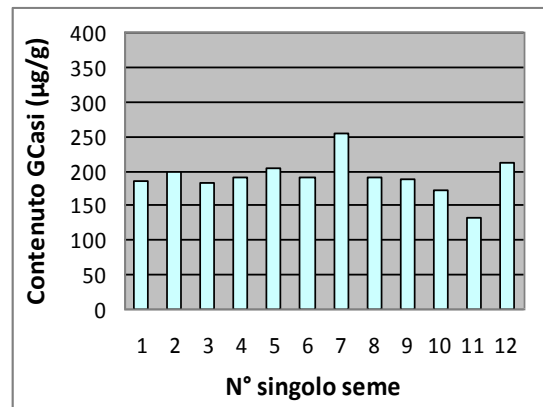
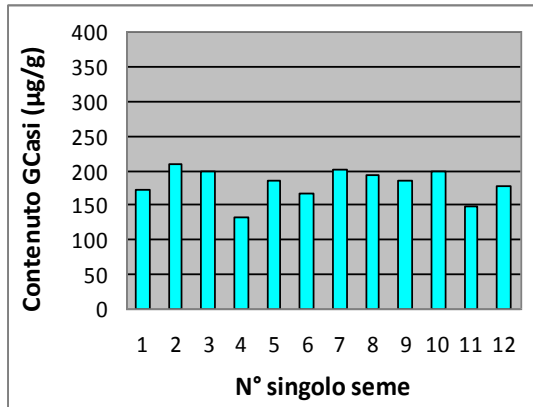


Grafico 3.18 Risultati relativi all'analisi degli estratti proteici ottenuti da 12 singoli semi valutati in DAS-ELISA per 2 individui T₂ privi di semi *null*.

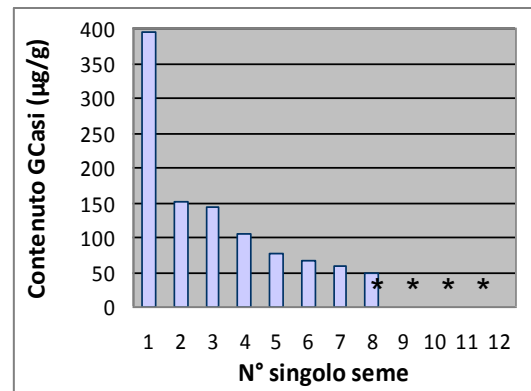
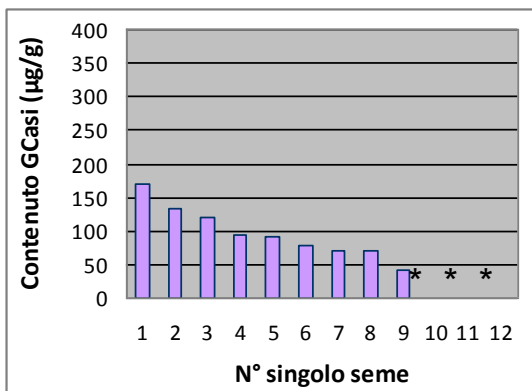


Grafico 3.19 Risultati relativi all'analisi degli estratti proteici ottenuti da 12 singoli semi valutati in DAS-ELISA per 2 individui T₂ con semi *null*.

4. DISCUSSIONE

La presente tesi si inserisce nell'ambito della produzione di proteine eterologhe di interesse farmacologico basata sull'impiego di piante come bioreattori. In particolare, la ricerca è stata rivolta al miglioramento dei livelli di espressione di α - e β -glucosidasi agendo a livello sia di ottimizzazione codonica, sia di elementi di regolazione (sequenze 3'-UTR). Il lavoro svolto si è, quindi, articolato nella realizzazione e sviluppo delle prime fasi del processo di produzione: dalla trascrizione alla traduzione della proteina ricombinante, all'accumulo in comportamenti subcellulari adeguati sino all'estrazione delle proteine dal tessuto target e alla valutazione del contenuto degli enzimi ricombinanti di interesse.

Nell'ambito del presente lavoro, la scelta di utilizzare il riso come specie ospite e, in particolare, la varietà *waxy* CR W3 è stata compiuta per soddisfare i criteri di biosicurezza nell'ottica di un eventuale rilascio in campo delle linee trasformate.

La varietà di riso CR W3, costituita dall'Ente Nazionale Risi per l'estrazione industriale di amido, presenta caratteri merceologici molto differenti rispetto alle varietà destinate al consumo alimentare. Allo stato attuale, questa varietà è coltivata unicamente dall'Ente Nazionale Risi presso il Centro Ricerche sul Riso, Castello d'Agogna (PV) su una superficie alquanto limitata e al solo scopo di mantenimento nel registro nazionale delle varietà di riso. E' interessante rilevare che l'impiego di CR W3 ai fini del presente lavoro è stato suggerito proprio dall'Ente costituente nella consapevolezza dell'esistenza in CR W3 di una serie di caratteri assolutamente peculiari. E' innanzitutto una varietà del tutto inadatta al consumo alimentare, poiché non tiene la cottura (Fig. 4.1). Tale caratteristica è determinata dalla presenza di un endosperma di tipo glutinoso, ossia con scarsissimo contenuto in amilosio; quest'ultimo aspetto associato a un profilo della cariosside arrotondato (rapporto lunghezza/larghezza inferiore a 1.75) la differenzia nettamente dalla stragrande maggioranza delle varietà iscritte al registro (Fig. 4.2).

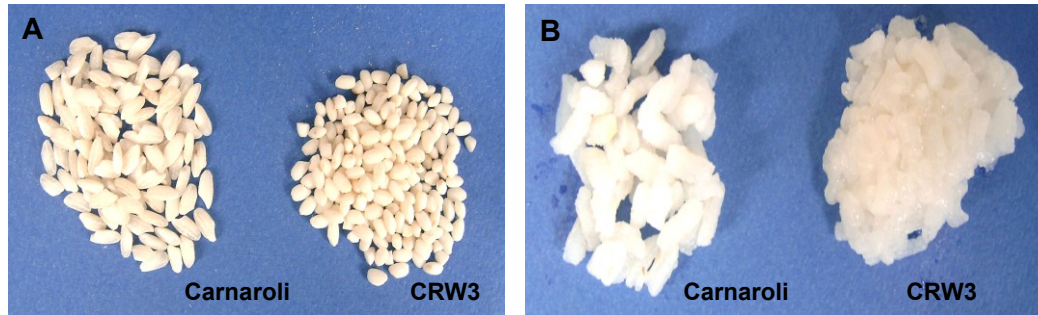


Figura 4.1 Differenze tra un tipo di riso alimentare adatto al consumo (Carnaroli) e la var. CR W3. A: forma della cariosside; B: tenuta alla cottura.

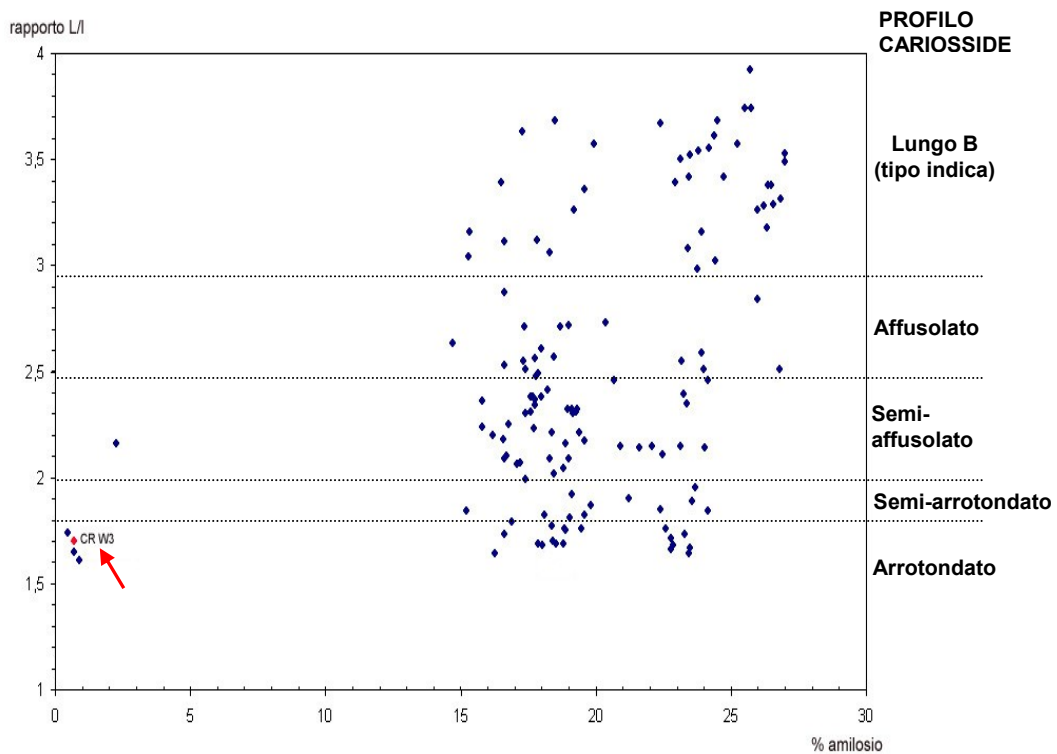


Figura 4.2 Distribuzione delle varietà di riso iscritte al registro nazionale in relazione al contenuto in amilosio e al rapporto lunghezza/larghezza della cariosside.

La precocità di fioritura e l'elevato livello di autogamia rende meno probabile l'esoincrocio di CR W3 con varietà coltivate e con il riso crodo. Diversamente da altre varietà a granello tondo come Selenio, Balilla e Centauro, CR W3 presenta apice della cariosside pigmentato, visibile già a maturazione latteo-cerosa. In CR W3 il grado di suberificazione del pedicello di attacco al rachide è inferiore rispetto a molte varietà; ciò determina una maggiore resistenza del panicolo alla sgranatura e in generale una maggiore difficoltà di distacco delle cariossidi,

anche a seguito di eventi meteorologici avversi (Fig. 40). Dal confronto della resistenza dei semi alla sgranatura di una linea CR W3 e CR W3 trasformata con 7 varietà di riso commerciali rappresentative di un diverso comportamento alla trebbiatura si è osservato che la pianta GM mantiene la caratteristica resistenza alla sgranatura tipica di CR W3 (Fig. 4.3). Inoltre, CR W3 non alletta ed è resistente al brusone.

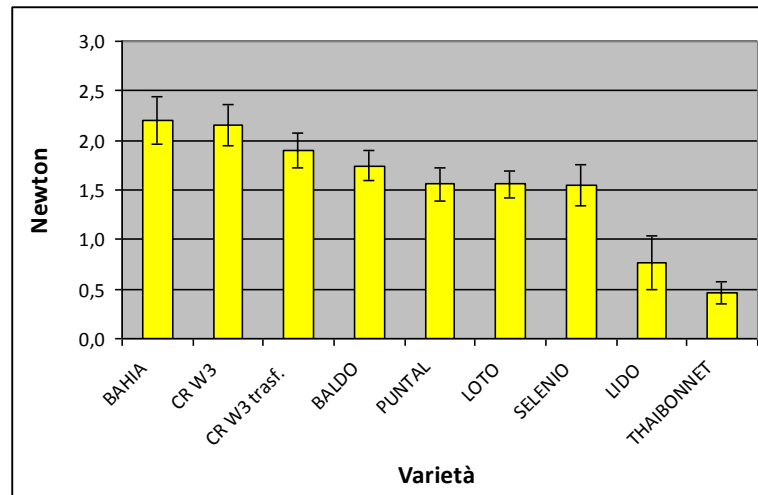


Figura 4.3 Resistenza dei semi alla sgranatura in CR W3, CR W3 trasformata e 7 varietà di riso commerciali. La forza di distacco della cariosside dal pedicello è stata misurata con un dinamometro ed espressa in Newton (N).

La scelta di sviluppare una strategia di sintesi nell'endosperma di riso è stata dettata da diverse caratteristiche favorevoli del sito di accumulo.

Nell'endosperma, infatti, le proteine ritrovano un ambiente naturalmente predisposto al loro stoccaggio in una biomassa altamente conservabile e ricca di compartimenti subcellulari.

Quando espresse in seme, le proteine ricombinanti rimangono stabili per anni, anche se conservate a temperatura ambiente; nel seme infatti esse trovano una matrice inerte con bassa umidità (14%), a basso contenuto lipidico (2.2%), ricca in inibitori di proteasi, isomerasi, chaperoni e povera di sostanze che possono interferire con i successivi processi di estrazione e purificazione. Inoltre, a differenza di altre specie vegetali come ad esempio il tabacco, il seme di riso può essere processato in modo da separare le componenti metabolicamente attive del seme, quali l'embrione, da quelle inerti, come l'endosperma. Tale aspetto

risulta fondamentale nella produzione di proteine eterologhe fitotossiche; le proteine di interesse possono, infatti, essere accumulate in maniera specifica nell'endosperma evitando possibili interferenze con il metabolismo della pianta e, in particolare, dell'embrione.

E' noto da precedenti studi che la GCasi ricombinante è di per sé dannosa all'organismo vegetale; la produzione costitutiva di GCasi in foglia risulta a tal punto problematica da escludere ogni reale utilità del sistema (Reggi et al., 2005). Il confinamento dell'espressione di GCasi nel seme di tabacco ha consentito una risoluzione seppur parziale dei problemi di fitotossicità. Tuttavia, l'accumulo di elevate quantità di β -glucosidasi acida ha indotto la perdita di germinabilità del seme di tabacco (Reggi et al., 2005).

Per affrontare queste problematiche nel presente lavoro si è cercato di sviluppare una piattaforma di sintesi delle proteine di interesse in endosperma di riso sfruttando gli aspetti favorevoli di questa matrice.

Inoltre, tale soluzione consente di raggiungere un altro importante risultato: ridurre il grado di contaminazione proteica dell'ambiente, ovvero la diffusione di proteine al di fuori dei normali confini biologici.

Per realizzare un'espressione endosperma-specifica degli enzimi lisosomiali oggetto di studio, le relative CDS sono state poste sotto il controllo del promotore della glutelina 4 di riso e di una sequenza leader artificiale LL-TCK; ciò allo scopo di ottenere elevati livelli di accumulo della proteina.

Il leader LL-TCK è una sequenza non tradotta in 5' (5'-UTR) adatta all'ottenimento di livelli di espressione elevati e rispondente a una serie di requisiti importanti per un'efficiente traduzione dell'mRNA:

1. sito di inizio trascrizione (*Inr*) del promotore CaMV 35S (Guilley et al., 1982) per un'efficiente *capping* dell'mRNA;
2. lunghezza superiore a 40 nt per favorire il riconoscimento del primo AUG come codone di inizio autentico (Kozak, 1989) e il reclutamento di più subunità ribosomiali 40S;
3. sequenza ricca in elementi CT, in analogia a molti leader vegetali (Bolle et al., 1996);
4. contenuto complessivo in GC inferiore al 40%.

In esperimenti precedenti (De Amicis et al., 2007) si è evidenziato che LL-TCK è in grado di determinare un innalzamento dei livelli sia di traduzione, sia di trascrizione del gene di interesse. L'effetto di LL-TCK è stato studiato in tabacco, utilizzando il promotore costitutivo CaMV 35S e il gene *uidA* codificante per l'enzima β -glucuronidasi (GUS) determinando un aumento della concentrazione di enzima rispetto alla sequenza leader naturale di circa 12 volte.

I promotori delle gluteline sono già stati impiegati con successo nell'espressione o nella sovraespressione di molte proteine ricombinanti in seme di riso, quali ad es.: la β -faseolina per l'incremento del contenuto in lisina (Zheng et al., 1995); la ferritina di soia per migliorare l'apporto di ferro (Goto et al., 1999); la lattoferrina per aumentare l'assorbimento del ferro e proteggere l'organismo dall'attacco di virus e batteri (Nandi et al., 2002); il lisozima per aumentare le difese contro microorganismi patogeni (Yang et al., 2003); gli enzimi coinvolti nella biosintesi del β -carotene ovvero della provitamina A affinché attraverso la dieta possa essere migliorato lo stato delle mucose e ridotta l'incidenza delle varie forme di cecità nella popolazione infantile (Ye et al., 2000).

Il promotore più frequentemente utilizzato per l'espressione di proteine eterologhe in endosperma di riso è quello della Glutelina 1 (GluB1) (Yu et al., 2005); in questo lavoro, è stato invece scelto il promotore della Glutelina 4 in quanto la sua sequenza viene attivata da fattori transagenti presenti unicamente nell'endosperma e la sua forza trascrizionale appare superiore a quella di altri promotori seme-specifici, incluso GluB1 (Qu e Takaiwa, 2004). In effetti, dai dati sperimentali raccolti sulla progenie dei trasformati migliori è anticipabile che il livello di espressione raggiunto per le proteine ricombinanti di interesse sia almeno 120 mg di proteina ricombinante per kg di seme. Questi valori indicano livelli di produzione ancora superiori ad altre proteine ricombinanti prodotte in seme di riso (Nandi et al., 2005) e confermano la competitività del sistema pianta rispetto ad altri sistemi di espressione rappresentati da fermentatori batterici, colture cellulari e sistemi animali. Questi risultati sono da considerarsi comunque preliminari; notevoli miglioramenti potranno essere apportati alle rese con la messa a punto di sistemi di estrazione e purificazione su scala industriale, l'ulteriore selezione delle linee produttive, la fissazione del transgene allo stato omozigote e condizioni di crescita naturali delle piante attraverso la coltivazione in campo. Dai dati ottenuti nel presente lavoro è emerso che la sola

sostituzione del macinacaffè con il mulino a sfere ha consentito di migliorare la capacità di estrazione delle proteine GAA e GCasi di più di due volte (cfr. Graf. 3.2) e che l'attività di selezione condotta sulle progenie GCasi CC può condurre a un aumento di produzione pari a 2 volte il quantitativo iniziale. Tali risultati assumono un'importanza tanto maggiore nel caso della produzione di α - e β -glucosidasi acida ricombinante considerando le elevate dosi di somministrazione degli enzimi richieste per sostenere un'efficace terapia enzimatica sostitutiva.

L'accumulo, l'attività e la stabilità di una proteina ricombinante sono inoltre vincolate al rispetto di condizioni chimico-fisiche e biochimiche. Non tutti i compartimenti cellulari si prestano in ugual modo a fornire un ambiente idoneo al processamento e allo stoccaggio di proteine diverse (Schillberg et al., 1999); in particolare, il *targeting* sub-cellulare influenza i processi di ripiegamento, assemblaggio e modificazione post-traduzionale di una proteina (Twyman et al., 2003). A tale proposito, va rilevato che sia GAA sia GCasi sono glicoproteine caratterizzate da una complessa struttura terziaria. Inoltre, GCasi acquisisce attività catalitica solo a fronte della presenza di un oligosaccaride nel primo sito di N-glicosilazione. Pertanto, è assolutamente necessario provvedere a un invio di entrambe le proteine nel lume del reticolo endoplasmatico (RE).

A tal fine, il peptide segnale nativo umano dell' α -glucosidasi acida è stato sostituito con quello della glutelina 4 (GluB 4) di riso, la cui sequenza nucleotidica è stata ottimizzata applicando le regole del *codon context* in modo da ottenere un innalzamento nell'efficienza di sintesi della proteina. La decisione di utilizzare PS Glub4 in luogo di quello nativo trova spiegazione nella possibilità che peptidi segnali umani espressi in pianta possano non essere riconosciuti e rimossi; la mancata rimozione del peptide segnale può determinare l'ancoraggio della proteina alla membrana del RE ostacolandone il corretto avvolgimento nella forma biologicamente attiva (Yan et al., 1997).

Va rilevato, inoltre, che in precedenti lavori condotti dalla sezione di Genetica del DiSA tale peptide segnale è risultato funzionale ed è stato correttamente rimosso dalla proteina precursore all'atto del versamento nel RE.

L'invio della proteina ricombinante al RE e il successivo indirizzamento ai vacuoli di immagazzinamento proteico (PSV) è stato effettivamente confermato da prove di immunolocalizzazione in semi trasformati GAA e GCasi svolte esternamente a questa tesi.

Il vettore scelto per l'espressione degli enzimi lisosomiali è stato pCAMBIA1300 opportunamente modificato al fine di ottemperare agli obblighi imposti dalla Direttiva 2001/18/CE e vigenti all'interno degli Stati membri dell'Unione Europea. Tale normativa ha stabilito una progressiva eliminazione dei marcatori di resistenza agli antibiotici presenti nelle piante geneticamente modificate (PGM), entro la scadenza posta al 31 dicembre 2008 per le PGM rilasciate a scopo di ricerca o di sviluppo. Questa linea guida impone, quindi, al ricercatore la messa a punto di nuovi protocolli per la trasformazione e la selezione delle PGM allo scopo di ottenere piante *marker-free*. In tale contesto normativo è stato messo a punto un sistema di selezione positiva, in cui le piante portanti il gene di selezione possiedono un vantaggio metabolico o di sviluppo rispetto a quelle non trasformate, basato sull'impiego del gene codificante la fosfomannosio isomerasi come marcatore selezionabile. In ragione di tale decisione si è proceduto alla sostituzione del gene *hptII* con il gene *manA* (PMI) isolato da *E. coli* nei vettori di espressione. Il metodo sviluppato ha previsto tre cicli di selezione caratterizzati dall'impiego di substrati di coltura contenenti concentrazioni crescenti di mannosio e concentrazioni decrescenti di saccarosio. La Figura 4.4 mostra una tipica piastra di selezione dei calli embriogenici di riso con PMI: i calli non trasformati appaiono di colore bruno; viceversa quelli bianchi portano il gene per la fosfomannosio isomerasi e acquisiscono la capacità di crescere nel mezzo selettivo.

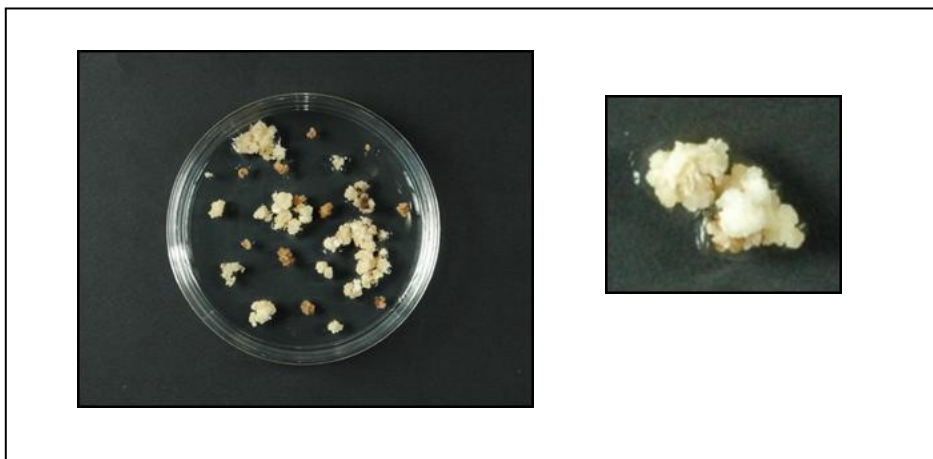


Figura 4.4 Selezione dei calli embriogenici di riso basata su PMI come *marker* selezionabile e mannosio come agente selettivo.

In vista di un eventuale rilascio delle PGM in campo e in ottemperanza alla vigente normativa europea, sono state condotte, inoltre, analisi molecolari allo scopo di rilevare la presenza di sequenze estranee al T-DNA, in generale definite come *backbone* del vettore. Nonostante per molto tempo si sia assunto che solo le porzioni comprese tra i due *border repeats* si trasferissero alla pianta, numerosi studi hanno dimostrato che il trasferimento di segmenti "non T-DNA" del vettore utilizzato per la trasformazione mediata da *A. tumefaciens* non è infrequente (De Buck et al., 2000; Martineu et al., 1994; Yin e Wang, 2000). In molti casi, tali sequenze risultano legate al T-DNA attraverso il *left border (LB)*, il *right border (RB)* o entrambi. Sono stati proposti due meccanismi per spiegare il fenomeno: da un lato, il LB non agirebbe da sito bersaglio nella reazione di taglio richiesta per rilasciare il T-DNA a singolo filamento (van der Graaff et al., 1996); dall'altro il LB potrebbe funzionare come RB (Huang et al., 2004).

Nel presente lavoro l'inserimento del backbone del vettore è stato verificato tramite PCR sui trasformati primari pCAMBIA1300_GAA CC-GluB4ter_PMI e pCAMBIA1300_GAA CC-GluB4 ter_PMI_FxAntiFx. I risultati hanno indicato una percentuale di individui positivi all'amplificazione pari al 29.10% nelle linee trasformate con il vettore a due cassette; tale valore è risultato aumentare al 64.25% nella popolazione di trasformati primari portanti le tre cassette. Dai dati disponibili in letteratura emerge che la percentuale di inserimento di sequenze backbone è tipicamente compresa tra il 20 e il 50% e può talvolta raggiungere valori intorno al 75% (Kononov et al., 1997; Kim et al., 2003).

In uno studio condotto da Yin e Wang (2000) in piante di riso trasformate è stato riportato, inoltre, il trasferimento non solo di un tratto esterno ai *border repeats* ma dell'intero vettore e di una seconda copia di T-DNA. Se questo tipo di trasferimento ha luogo, il numero di copie di T-DNA può aumentare. Alla luce di questi risultati, è stata verificata l'ipotesi di una possibile correlazione tra produttività e presenza del *backbone*.

In particolare, una corrispondenza significativa tra il maggior contenuto di proteina GAA e presenza di sequenze diverse dal T-DNA è stata osservata nella popolazione pCAMBIA_GAA CC-GluB4 ter_PMI_FxAntiFx; infatti, gli individui PCR-positivi al tratto esterno si sono dimostrati i migliori espressor. Ciò potrebbe essere spiegato con il trasferimento in pianta di più copie di T-DNA che

porterebbero a un aumento del contenuto di α -glucosidasi acida rilevato nel seme.

Diversamente dai dati ottenuti saggiando le progenie dei trasformati primari a tre cassette, il contenuto di GAA ricombinante in funzione della presenza o meno del *backbone* vettoriale ha mostrato una distribuzione maggiormente omogenea tra le linee pCAMBIA_GAA CC-GluB4 ter_PMI sebbene i migliori espressorini siano risultati positivi all'amplicazione del tratto esterno.

La possibile correlazione tra produttività e presenza del *backbone* è un aspetto particolarmente controverso per le proteine ricombinanti prodotte in pianta di cui si intenda effettuare il rilascio in campo. Se da un lato, infatti, l'incremento del contenuto di enzima ricombinante rappresenta un parametro fondamentale per l'applicabilità industriale di tale sistema eterologo di espressione ed è, quindi, auspicabile, dall'altro l'inserimento di sequenze del vettore estranee al T-DNA può costituire un ostacolo all'accettazione della notifica per la possibile presenza di geni che conferiscono resistenza ad antibiotici nel *backbone* del vettore. Infatti, la normativa europea, come già ricordato, ne ha disposto la graduale eliminazione, sebbene la FDA (Food and Drug Administration) ritenga che l'uso di geni di resistenza agli antibiotici debba essere valutato caso per caso.

Nonostante le problematiche di carattere regolatorio-normativo, le piante transgeniche costituiscono senza dubbio una piattaforma promettente per l'espressione di una varietà di proteine eterologhe. Tuttavia i livelli di produzione delle stesse rappresentano ancora un fattore critico determinante per le applicazioni dei sistemi basati sui bioreattori vegetali.

Come noto, espressione e accumulo delle molecole ricombinanti in pianta sono regolati a livello sia trascrizionale sia post-traduzionale e coinvolgono una molteplicità di fattori quali la forza del promotore, la stabilità dell'mRNA, l'efficienza di traduzione, il *targeting* subcellulare e l'impiego di appropriate sequenze non tradotte in 3' (3'-UTR) come terminatori (Streatfield, 2007; Knirsch et al., 2000).

Numerosi studi sono stati condotti per esplorare gli aspetti relativi al controllo dell'espressione genica concentrando l'attenzione sull'isolamento e caratterizzazione di promotori forti (Guilchley et al., 1982; Christensen et al., 1992; Lamacchia et al., 2001; Qu e Takaiwa, 2004; Qu et al., 2008).

Le regioni 3'-UTR rivestono un ruolo importante nella regolazione dell'espressione genica influenzando l'accumulo dell'mRNA, la stabilità e l'efficienza del processo di traduzione (Atwater et al., 1990; Bashirullah et al., 2001; Hollans et al., 2002; Merritt et al., 2008). Nelle piante non esiste una sequenza universalmente conservata, quale l'esamero AATAAA identificato nei mammiferi, responsabile del processamento all'estremità 3'; questo rende difficile l'elaborazione di un modello predittivo. Gli ipotetici segnali consistono di un elemento FUE (*Far Upstream Element*), una e o più regioni ricche in A note come NUE (*Near Upstream Element*) e una regione ricca in U localizzata a monte, a valle o fiancheggiante il CS (*Cleavage polyadenylation site*).

Il terminatore Nos (nopalina sintasi) combinato con promotori eterologhi è ampiamente utilizzato per esprimere transgeni in pianta (Depicker et al., 1982). Tuttavia, per un accumulo ottimale delle proteine ricombinanti e per una produzione su scala industriale si rende necessario indagare ulteriormente le funzioni delle sequenze 3'-UTR.

Terminatori di geni vegetali diversi si sono, infatti, dimostrati efficienti nel contribuire alla regolazione quantitativa dell'espressione genica. Il terminatore del gene *Me1* ha portato un aumento dei livelli di espressione del reporter GUS in foglia senza alterarne il pattern di espressione (Ali e Taylor, 2001). La regione 3'-UTR del gene manganese superossido dismutasi ha assunto una funzione di *enhancer* traduzionale *in vivo* (Knirsch e Clerch, 2000) mentre il 3'-UTR del gene *GS₁* di soia ha influenzato la stabilità dell'mRNA e l'accumulo della proteina in piante transgeniche di tabacco ed erba medica (Ortega et al., 2006). Più recentemente Takaiwa et al. (2008) hanno dimostrato che l'impiego combinato di promotore e terminatore derivante dallo stesso gene (*GluB1*) aumenta la stabilità dell'mRNA inducendo un incremento dei livelli di espressione che in alcuni casi ha raggiunto il 15% delle proteine totali. Il miglioramento apportabile con l'utilizzo di promotore-terminatore del medesimo gene è stato dimostrato anche in specie diverse dal riso.

Alla luce di queste evidenze e della necessità di aumentare i livelli di espressione degli enzimi ricombinanti di interesse per abbattere i costi di produzione e conseguentemente il prezzo dei farmaci per la ERT, si è deciso di combinare il terminatore *GluB4* con il promotore del medesimo gene; quest'ultimo già impiegato in precedenti esperimenti.

Di seguito è riportata la sequenza 3'-UTR di GluB4 sintetizzata evidenziando i segnali ipotetici che ne regolano il processamento:

```

                -98                -65
GAGCTCTG TACTAATGAAATAGTATA GGTGTA TCAAAAA AATAAT
                -70
GCCACAAGT ATGTGAAACTTTGTGGCGGTTCTGTTCTAAGTTTAT
                -26                -10                0
GAAGCACT AATAAAA ATGCTACAAC TGTTTT GTTC ATTTGTT GTGT
                +28                +46
CAATATAGGTGAGACCA TGTTTT GGTAGTTTACAT AAAAGA TTTG
CCCATAGTAGATAGATCTCAAAGCAGATTCGGTGCCAGGTATAAC
                +145
TTGTCAATGCTATATTTTAATCTCAGAATTATTATGAAGAGATAAA
GAAA TGCTTATGAATTC
    
```

Figura 4. Sequenza del terminatore GluB-4 impiegato con evidenziazione del sito CS (in rosso), degli elementi FUE (in azzurro) e NUE (in verde), dell'esamero AAGAAA (in rosa).

Per valutare l'effetto del terminatore, sono stati valutati i livelli di espressione della proteina GCasi in piante ingegnerizzate con il costrutto portante il terminatore NOS e in piante ottenute dalla trasformazione con il nuovo terminatore GluB4. In entrambi i casi la sequenza 3'-UTR è stata associata al promotore della glutelina 4. Confrontando il contenuto di proteina di interesse nelle due popolazioni considerate, si è osservato un netto incremento nei livelli di espressione di GAA a seguito dell'inserimento del terminatore GluB4 (cfr. Graf. 3.4). In particolare, nel seme della linea GAA CC-GluB4 ter avente il maggior contenuto di enzima ricombinante si sono raggiunti livelli di produzione pari a circa 150 µg/g, più di due volte superiore a quelli rilevati nel migliore espressore GAA CC_NOS ter (circa 65 µg/g).

Nel presente lavoro sono state saggiate due versioni artificiali sinonime codificanti la GCasi matura utilizzando come termine di confronto il cDNA umano, al fine di valutare l'effetto sul livello massimo di espressione della proteina delle 3 sequenze codificanti. In particolare, le due versioni sintetiche del gene, denominate GCasi CC e GCasi CU sono il risultato di un'attività di *design* genico. Ciascuna sequenza codificante è stata inserita nel vettore di espressione

in sostituzione della versione umana nativa, dando luogo a tre cassette tra loro differenti unicamente per questo aspetto.

La decisione di operare la riscrittura genica trova evidenza nella natura degenerata del codice genetico universale; una stessa proteina, infatti, può essere codificata da diversi codoni sinonimi. È stato dimostrato che i differenti organismi mettono in atto una scelta selettiva, non casuale e specie-specifica di tali codoni, nota come dialetto codonico o *codon usage*. L'esistenza dei dialetti codonici influisce sulla espressione del transgene sia in termini di riconoscimento della sequenza, sia sul livello di espressione della stessa. Alcuni studi hanno permesso di evidenziare differenze nel *codon usage* tra piante monocotiledoni e dicotiledoni (Kawabe e Miyshita, 2003).

Per tali motivi in questa tesi si è deciso di riscrivere il gene GCasi utilizzando sia il metodo del *codon context*, sia quello del *codon usage*, entrambi indirizzati a codificare proteine alto-esprese in endosperma di riso.

La CDS della β -glucosidasi acida è stata, quindi, tradotta *in silico* in proteina; successivamente, a partire dalla sequenza amminoacidica si è operata una doppia scrittura della GCasi applicando in un caso il metodo standard di *design* genico fondato sul *codon usage* modificato per ottenere un'elevata espressione in seme di riso e, in un altro, quello desunto sulla base del *codon context* basato sull'uso dei codoni più frequenti in relazione al contesto intercodonico.

Dalle analisi di sequenza eseguite è emerso che le versioni sintetiche mostrano una composizione strutturale molto simile tra loro sia relativamente al contenuto GC espresso in termini percentuali (63%), sia per quanto riguarda la soppressione dei segnali di instabilità e di elementi di disturbo quali omotetrameri costituiti da sole citosine e guanine, seppur con una diversa incidenza dei dinucleotidi CpG e degli intercodoni C₃pG₄. Il contributo di questi elementi è, infatti, risultato variare nei due casi: nella sequenza CC si è osservato una più alta frequenza di CpG (107 vs. 88) mentre in CU l'aumento nel contenuto in GC è per lo più imputabile al maggior numero di intercodoni C₃pG₄ (88 vs. 45). Nonostante la diversa distribuzione degli elementi CG intracodonici e intercodonici, le due sequenze non appaiono tra loro molto diverse. Questo dato non deve, però, sorprendere poiché già studi precedenti (Venturini et al., 2000) avevano evidenziato che la differenza tra le due metodologie di *design* genico risulta più rilevante nel caso di geni appartenenti alla classe dicotiledoni.

Sebbene i geni sintetici CC e CU abbiano mostrato un alto livello di identità, con la sequenza umana nativa (rispettivamente pari all'82 e 80%), non sono sovrapponibili rispetto alla CDS umana a causa di un differente contenuto in GC totale e una diversa incidenza di elementi CpG intracodonici e intercodonici (C_3pG_4), nonché di elementi omotrimeri e omotetrameri formati da sole citosine e guanine. Se comparate con la sequenza originale, le sequenze sintetiche mostrano una composizione strutturale diversa; la percentuale di GC totale calcolata è stata particolarmente alta, pari al 63%, come atteso per geni altamente espressi in cereali (Guo et al., 2007) che possono raggiungere valori in GC intorno al 60-70% (Bernardi, 1985). In particolare, si è osservata una più alta frequenza di dinucleotidi CpG, a dimostrazione del fatto che i genomi delle monocotiledoni non evitano l'utilizzo dei codoni NCG (De Amicis e Marchetti, 2000) e di elementi C_3pG_4 . Inoltre, sono stati aboliti gli omotetrameri costituiti da sole citosine e quelli formati da sole guanine, in linea con le osservazioni ottenute in un precedente lavoro (De Amicis e Marchetti, 2000) e con la dimostrazione degli effetti negativi dei *G-runs* sulla stabilità dell'mRNA (Williamson, 1994).

In aggiunta, gli omotrimeri sono risultati ridotti ed eventuali sequenze che potevano costituire siti critici di *splicing* eliminate.

L'analisi delle sequenze sintetiche ha, inoltre, evidenziato un'alta frequenza di codoni terminanti in G o C, pari all'87.73 e 93.29% nel caso rispettivamente di GCasi CC e CU; in studi precedenti (Venturini, 2006) era stato osservato un netto aumento nella incidenza dei codoni terminanti in G o in C passando da geni debolmente espressi a quelli altamente espressi, a conferma dell'esistenza di una relazione tra contenuto in GC in terza posizione (N_3) e livello di espressione. Al fine di determinare l'effetto delle tre sequenze codificanti sinonime sul livello massimo di espressione della proteina, è stato confrontato il contenuto medio di enzima ricombinante ottenuto dalle due analisi indipendenti nelle tre popolazioni transgeniche analizzate come riportato (cfr. Graf. 3.13). I dati ottenuti hanno evidenziato un aumento dei livelli di proteina nel seme dei trasformati primari GCasi CC rispetto sia alla sequenza nativa, sia a quella riscritta secondo il criterio del *codon usage*. I risultati sperimentali hanno, quindi, indicato una superiorità del metodo basato sull'impiego dei codoni contesto a fronte di una composizione strutturale delle versioni sintetiche non dissimile tra loro.

5. CONCLUSIONI

La produzione e l'impiego di piante geneticamente modificate rappresenta senza dubbio uno degli argomenti più discussi e controversi degli ultimi anni soprattutto per le perplessità suscitate e percepite da una parte dell'opinione pubblica sull'onda di una accentuata emotività e di un timore diffuso e atavico verso ciò che ancora non si è sperimentato e che scardina le convinzioni comuni. Se da un lato questi timori possono essere compresi come parte dell'animo umano, dall'altro essi appaiono il frutto di una disinformazione generalizzata e di un allarmismo che nulla ha a che vedere con la ricerca e il riscontro scientifico. È, infatti, importante sottolineare come non vi siano evidenze relative a un reale danno sull'ambiente e sulla salute umana provocato dall'utilizzo e immissione di piante geneticamente modificate. La stessa normativa europea estremamente restrittiva verso le PGM si fonda sul principio di precauzione. Tra le problematiche connesse alla trasformazione genetica delle piante l'accettazione sociale assume, quindi, una particolare rilevanza.

Al di là di questo scetticismo culturale, senza alcun dubbio l'impiego di piante geneticamente modificate apre la strada a una evoluzione dei sistemi agricoli attuali e al progresso sociale attraverso lo sviluppo di nuove varietà che renderanno disponibili alimenti arricchiti in vitamine, sali minerali o nutraceutici capaci di ridurre l'incidenza di determinate carenze alimentari oppure la realizzazione di bioreattori verdi per la sintesi di proteine di interesse farmaceutico nella lotta contro alcune malattie rare e forme tumorali. Alcuni esempi sono forniti dalla produzione del Golden Rice, varietà di riso arricchite con pro-vitamina A (Ye et al., 2000) o biofortificate per l'assunzione di una quantità maggiore di Fe e Zn (Vasconcelos et al., 2003) o piante geneticamente modificate per la produzione di vaccini anti-tumorali paziente specifici.

L'espressione di proteine eterologhe in pianta offre, in particolare, prospettive estremamente interessanti nell'abbattimento dei costi di produzione e nell'ampliamento dei volumi produttivi (Kusnadi et al., 1997). I bioreattori vegetali costituiscono una valida ed efficiente alternativa ai sistemi tradizionali basati su cellule batteriche (Fischer e Emans, 2000), di lievito o mammifere. Ai costi di investimento ridotti e alla rapidità nella produzione, si aggiungono la possibilità di ottenere elevate quantità di proteina ricombinante e la capacità di attuare

modificazioni post-traduzionali simili a quelle animali con solo alcune differenze relative alla glicosilazione (cfr. § 1.3.1). Inoltre, l'utilizzo di piante transgeniche elimina i rischi, esistenti invece per le colture cellulari animali, di contaminazione e trasmissione all'uomo di agenti infettivi (Fischer e Emans, 2000).

Per rendere ancora più competitivo il sistema pianta nell'area della produzione di proteine eterologhe a scopo terapeutico e farmacologico è, tuttavia, necessario soddisfare ulteriori requisiti, quali un efficace accumulo della molecola ricombinante nei tessuti vegetali e la possibilità di perfezionare le fasi di estrazione e purificazione della proteina con un duplice obiettivo: ridurre i costi e raggiungere un elevato grado di purificazione al fine di evitare la contaminazione da allergeni a basso peso molecolare. In riso sono state, infatti, individuate proteine in grado di agire come potenziali allergeni (Usui et al., 2001).

In conclusione, si può affermare che la produzione di proteine eterologhe in pianta costituisce un sistema efficiente, economicamente conveniente, con importanti margini di miglioramento e di grande interesse soprattutto per implementare gli strumenti terapeutici disponibili.

5.1 Prospettive future

Nel presente lavoro si è tentato di agire sulla piattaforma tecnologica di sintesi di proteine ricombinanti in endosperma di riso al fine di migliorarne i livelli di espressione. Dall'attività di ricerca condotta è emerso che sia l'ottimizzazione codonica, sia l'impiego combinato di promotore-terminatore della glutelina 4 hanno consentito un innalzamento dei livelli di espressione.

Per quanto riguarda le prospettive future delle due molecole studiate, l'attività sperimentale si concentrerà sullo sviluppo di metodi di purificazione scalabili industrialmente allo scopo di produrre una proteina pura a elevate rese per procedere successivamente alla caratterizzazione e valutazione pre-clinica degli enzimi lisosomiali ricombinanti. In particolare, sia per GCasi, sia per GAA la ricerca proseguirà a breve termine con i seguenti obiettivi:

- analisi della sequenza N-terminale;
- esatta determinazione della struttura degli N-glicani della molecola e valutazione dell'efficacia della cassetta per il silenziamento introdotta;
- saggi *in vitro* di *uptake* su fibroblasti umani da individui malati;
- avvio dei saggi pre-clinici su modello murino.

6. BIBLIOGRAFIA

Aerts JM, Schram AW, Strijland A, van Weely S, Jonsson LM, Tager JM, Sorrell SH, Ginns EI, Barranger JA, Murray GJ, 1988. Glucocerebrosidase, a lysosomal enzyme that does not undergo oligosaccharide phosphorylation. *Biochim Biophys Acta* 964: 303-308.

Ali S, Taylor W (2001) The 30 non-coding region of a C4 photosynthesis gene increases transgene expression when combined with heterologous promoters. *Plant Mol Biol* 46:325–333.

Amalfitano A, McVie-Wylie AJ, Hu H, Dawson TL, Raben N, Plotz P, 1999. Systemic correction of the muscle disorder glycogen storage disease type II after hepatic targeting of a modified adenovirus vector encoding human acid- α -glucosidase. *Proc Natl Acad Sci USA* 96: 8861-6.

Amaral O, Lacerda L, Marcao A, Pinto E, Tamagnini G, Miranda MCS, 1999. Homozygosity for two mild glucocerebrosidase mutations of probable Iberian origin. *Clin Genet* 56: 100-102.

Amaral O, Lacerda L, Santos R, Pinto A, Aerts H, Miranda MCS, 1993. Type 1 Gaucher disease: Molecular, biochemical, and clinical characterization of patients from northern Portugal. *Biochem Med Metab Biol* 49: 97-107.

Apweiler R, Hermjakob H, Sharon N, 1990. On the frequency of protein glycosylation, as deduced from analysis of SWISS-PROT database. *Biochim Biophys Acta* 1473: 4-8.

Atwater JA, Wisdom R, Verma IM, 1990. Regulated mRNA stability. *Annu Rev Genet* 24: 519-541.

Ausems MG, Verbiest J, Hermans MMP, Kroos MA, Beemer FA, Wokke JHJ, 1999. Frequency of glycogen storage disease type II in The Netherlands: implications for diagnosis and genetic counselling. *Eur J Hum Genet* 7: 713-6.

Barneveld RA, Keijzer W, Tegelaers FPW, [Ginns EI](#), [Geurts van Kessel A](#), [Brady RO](#), [Barranger JA](#), [Tager JM](#), [Galjaard H](#), [Westerveld A](#), [Reuser AJ](#), 1983. Assignment of the gene coding for human beta-glucocerebrosidase to the region q21-q31 of the chromosome 1 using monoclonal antibodies. *Hum Genet* 64: 227-231.

Bashirullah A, Cooperstock RL, Lipshitz HD, 2001. Spatial and temporal control of RNA stability. *Proc Natl Acad Sci USA* 98: 7025-7028.

Basu A, Prenc E, Garrett K, Glew RH, Ellingson JS, 1985. Comparison of N-acyl phosphatidylethanolamines with different N-acyl groups as activators of glucocerebrosidase in various forms of Gaucher's disease. *Arch Biochem Biophys* 243: 28-34.

Basu S, Kaufman B, Roseman S, 1968. Enzymatic synthesis of ceramide-glucose and ceramide-lactose by glycosyltransferases from embryonic chicken brain. *J Biol Chem* 243: 5802-5804.

Basu S, Kaufman B, Roseman S, 1973. Enzymatic synthesis of glucocerebroside by a glycosyltransferase from embryonic chicken brain. *J Biol Chem* 248: 1388-1394.

Bernardi G, Olofsson B, Filipinski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F, 1985. The mosaic genome of warm-blooded vertebrates. *Science* 228: 953-958.

Berg-Fussman A, Grace ME, Ioannou Y, Grabowski GA, 1993. Human beta-glucosidase: N-glycosylation site occupancy and the effect of glycosylation on enzymatic activity. *J Biol Chem* 268: 14861-14866.

Bergmann JE, Grabowski GA, 1989. Posttranslational processing of human lysosomal acid beta-glucosidase: a continuum of defects in Gaucher disease type 1 and type 2 fibroblasts. *Am J Hum Genet* 44: 741-750.

Beutler E, 2006. Lysosomal storage diseases: natural history and ethical and economic aspects. *Mol Genet Metab* 88: 539-547.

Beutler E, 1992. Gaucher disease: New molecular approaches to diagnosis and treatment. *Science* 256: 794-799.

Beutler E, Gelbart T, 1993. Gaucher disease mutations in non Jewish patients. *Br Haematol* 85: 401-405.

Beutler E, Gelbart T, Kuhl W, Zimran A, West C, 1992. Mutations in Jewish patients with Gaucher disease. *Blood* 79: 1662-1666.

Beutler E, Grabowsky GA, 2001. Gaucher disease In: Scriver CR, Beudet A, Sly WS, Valle D, editors Molecular and metabolic basis of Inherited diseases 8th edn New York: McGraw-Hill: p 3635-3656

Beutler E, Gelbart T, 1996. Glucocerebrosidase (Gaucher disease). *Hum Mutation* 8: 207-213.

Beutler E, West C, Gelbart T, 1992. Polymorphisms in the human glucocerebrosidase gene. *Genomics* 12: 795-800.

Bijvoet AG, Van Hirtum H, Kroos MA, Van de Kamp EH, Schoneveld O, Visser P, Brakenhoff JP, Weggeman M, van Corven EJ, Van der Ploeg AT, Reuser AJ, 1999. Human acid alpha-glucosidase from rabbit milk has therapeutic effect in mice with glycogen storage disease type II. *Hum Mol Genet* 8(12): 2145-2153.

Blixt O et al., 2002. Efficient chemoenzymatic synthesis of O-linked sialyl oligosaccharides. *J Am Chem Soc* 124: 5739-5746.

Bolle C, Herrmann RG, Oelmüller R, 1996. Different sequences for 5'-untranslated leaders of nuclear genes for plastid proteins affect the expression of the β -glucuronidase gene. *Plant Mol Biol* 32: 861-868.

Brady RO, Kanfer JN, Shapiro D, 1965. Metabolism of glucocerebrosidase. II. Evidence of an enzymatic deficiency in Gaucher's disease. *Biochem Biophys Res Commun* 18: 221-225.

Chen YT, Amalfitano A, 2000. Towards a molecular therapy for glycogen storage disease type II (Pompe disease). *Mol Med Today* 6: 245-251.

Chinn AM, Comai L, 1996. Gene Transcription. In: Owen MRL, Pen J (eds.) *Transgenic Plants: a production system for industrial and pharmaceutical proteins* John Wiley and Sons, London, UK: 27-48.

Chrastil J, 1992. Correlations between the physicochemical and functional properties of rice. *Journal of Agricultural and Food Chemistry* 40: 1683-1686.

Christensen AH, Sharrock RA, Quail PH, 1992. Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol Biol* 18: 675-689.

Chuang CF, Meyerowitz EM, 2000. Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 97: 4985-4990.

Cogoni C, Macino G, 2000. Post-transcriptional gene silencing across kingdoms. *Curr Opin Genet* 6: 638-643.

Conrad U, Feidler U, 1998. Compartment-specific accumulation of recombinant immunoglobulins in plant cells: an essential tool for antibody production and immunomodulation of physiological functions and pathogen activity. *Plant Mol Biol* 38: 101-109.

Cox TM, Schofield JP, 1997. Gaucher's disease: clinical features and natural history. In Gaucher's disease. *Bailliere Clin Haematol* 10: 657–689.

Cox KM, Sterling JD, Regan JT, Gasdaska JR, Frantz KK, Peele CG, Black A, Passmore D, Moldovan-Loomis C, Srinivasan M, Cuisson S, Cardarelli PM e Dickey LF, 2006. Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat Biotechnol* 24: 1591-1597.

Courtecuissf V, Royer F, Habib R, Monnifer C, Denvos J, 1965. Glycogérose musculaire par déficit d'alpha 1-4 glucosidase simulant un dystrophie musculaire progressive. *Arch Fr Pediat* 22: 1153-1164.

Datta K e Datta SK, 2006. *Agrobacterium* Protocols, 2/e, vol. 1. *Methods in Molecular Biology*, vol. 343: pagg. 201-212. Edited by: Kan Wang © Humana Press Inc., Totowa, NJ.

De Amicis F, Marchetti S, 2000. Intercodon dinucleotides affect codon choice in plant genes. *NAR* 28: 3339-3345.

De Amicis F, Patti T, Marchetti S, 2007. Improvement of the pBI121 plant expression vector by leader replacement with a sequence combining a poly(CAA) and a CT motif. *Transgenic Res* 16: 731-738.

De Buck S, De Wilde C, van Montagu M, Depicker A, 2000. T-DNA vector backbone sequences are frequently integrated into the genome of transgenic plants obtained by *Agrobacterium*-mediated transformation. *Mol Breed* 6: 459-468.

Depicker A, Stachel S, Dhaese P, Zambryski P, Goodman HM (1982) Nopaline synthase: transcript mapping and DNA sequence. *J Mol Appl Genet* 1:561–573.

Diaz A, Montfort M, Cormand B, Zeng BJ, Pastores GM, Chabas A, Vilageliu L, et al., 1999. Gaucher disease: The N370S mutation in Ashkenazy Jewish and

Spanish patients has a common origin and arose several thousand years ago. *Am J Hum Genet* 64: 1233-1238.

Doyle JJ, Doyle JL, 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Duret L, Mouchiroud D, 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci USA*, 96: 4482-4487.

Elbashir SM, Lendeckel W, Tuschl T, 2001. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* 15: 188-200.

Engel AG, Gomez MR, Seybold ME, Lambert EH, 1973. The spectrum and diagnosis of acid maltase deficiency. *Neurology* 23: 95-106.

Erickson AH, Ginns EI, Barranger JA, 1985. Biosynthesis of the lysosomal enzyme glucocerebrosidase. *J Biol Chem* 260: 14319-14324.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC, 1998. Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. *Nature*, 391: 806-811.

Firon N, Eyal N, Kolodny EH, Horowitz M, 1990. Genotype assignment in Gaucher disease by selective amplification of the active glucocerebrosidase gene. *Am J Hum Genet* 46: 527-532.

Furbish FS, Steer CJ, Krett NL, Barranger JA, 1981. Uptake and distribution of placental glucocerebrosidase in rat hepatic cells and effects of sequential deglycosylation. *Biochim Biophys Acta* 673: 425-434.

Ferguson JD, Street BE, David SB, 1958. The carbohydrate nutrition of tomato roots. IV. The inhibition of excised root growth by galactose and mannose and its reversal by dextrose and xylose. *Ann Bot* 22: 523-538.

Fischer R, Emans N. 2000. Molecular farming of pharmaceutical proteins. *Transgenic Res*, 9: 279-299.

Furbish FS, Blair HE, Shiloach J, Pentchev PG, Brady RO, 1977. Enzyme replacement therapy in Gaucher's disease: large-scale purification of glucocerebrosidase suitable for human administration. *Proc Natl Acad Sci USA* 74: 3560-3563.

Geel TM, McLaughlin PMJ, de Leij LFMH, Ruiters MHJ, Niezen-Koning KE, 2007. Pompe disease: current state of treatment modalities and animal models. *Mol Genet Metab* 92: 299-307.

Ginns EI, Choudary PV, Martin BM, Winfield S, Stubblefield B, Mayor J, Merkle-Lehan D, Murray GJ, Bowers LA, Barranger JA, 1984. Isolation of cDNA clones for human β -glucocerebrosidase using the λ -gt II expression system. *Biochem Biophys Res Commun* 123: 574-80.

Ginns EI, Choudary PV, Tsuji S et al., 1985. Gene mapping and leader polypeptide sequence of human glucocerebrosidase: implication for Gaucher disease. *Proc Natl Acad Sci USA* 82: 7101-7105.

Glick BR, Pasternak JJ. 1999. *Biotecnologia molecolare*, Zanichelli.

Goker-Alpan O, Schiffmann R, Park JK, Stubblefield BK, Tayebi N, Sidransky E, 2003. Phenotype continuum in neuronopathic Gaucher disease: an intermediate phenotype between type 2 and type 3. *J Pediatr* 143: 273-276.

Goldsworthy A, Street HE, 1965. The carbohydrate nutrition of tomato roots VIII. The mechanism of the inhibition by D-mannose of the respiration of excised roots. *Ann Bot* 29: 45-58.

Gomord V, Denmat LA, Fichette-Laine AC, Satiat-Jeunemaitre B, Hawes C, Faye L, 1997. The C-terminal HDEL sequence is sufficient for retention of

secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. *Plant J* 11: 313-325.

Goto F, Yoshihara T, Shigemoto N, Toki S, Takaiwa F, 1999. Iron fortification of rice seed by the soybean ferritin gene. *Nat Biotechnol* 17: 282-286.

Grabowski GA, 1997. Gaucher disease: gene frequencies and genotype/phenotype correlations. *Genet Test* 1: 5-12.

Grabowski GA, Gatt S, Horowitz M, 1990. Acid β -glucosidase: enzymology and molecular biology of Gaucher disease. *Biochemistry and Molecular Biology* 25: 385-414.

Grabowski GA, Hopkin RJ, 2003. Enzyme therapy for lysosomal storage disease: principles, practice, and prospects. *Annu Rev Genom Hum Genet* 4: 403-436.

Grabowski GA, Leslie N, Wenstrup R, 1998. Enzyme therapy for Gaucher disease: the first 5 years. *Blood Rev* 12: 115-133.

Grace ME, Goldberg L, Berg A, Grabowski GA, 1989. Gaucher disease: dissection of the enzymatic pathology by site-directed mutagenesis and expression of acid β -glucosidase cDNAs. *Am J Hum Genet* 45: A190.

Grace ME, Newman KM, Scheinkert V et al., 1994. Analysis of human acid β -glucosidase by site directed mutagenesis and heterologous expression. *J Biol Chem* 269 (3): 2283-2291.

Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R, 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res* 9: 43-74.

Graves PN, Grabowski GA, Elmer R, Palese P, Smith FI, 1988. Gaucher disease type 1: cloning and characterization of a cDNA encoding acid β -glucosidase from an Ashkenazi Jewish patient. *DNA* 7: 521-528.

Guilley H, Dudley RK, Jonard G, Balazs, Richards KE, 1982. Transcription of Cauliflower Mosaic Virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* 30: 763-773.

Guo X, Bao J, Fan L, 2007. Evidence of selectively driven codon usage in rice: implications for GC content evolution of Gramineae genes. *FEBS Lett* 581: 1015-1021.

Hamada S, Ishijama K, Choi S-B, Wang C, Singh S, Kawai N, Franceschi VR, Okita TW, 2003. The transport of prolamine RNAs to prolamine protein bodies in living rice endosperm cells. *The Plant Cell* 15 : 2253-2264.

Hamilton AJ, Baulcombe DC, 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286: 950-952.

Hansen G, Chilton MD, 1999. Lessons in gene transfer to plants by a gifted microbe. In: Hammond J, McGarvey P, Yusibov V, editors. *Plant Biotechnology: New products and Applications* 240: 21-58.

Hasilik A, Neufeld EF, 1980. Biosynthesis of lysosomal enzymes in fibroblasts. Synthesis as precursors of higher molecular weight. *J Biol Chem* 255: 4937-4945.

Hermans MM, van Leenen D, Kroos MA, 2004. Twenty-two novel mutations in the lysosomal alpha-glucosidase gene (GAA) underscore the genotype-phenotype correlation in glycogen storage disease type II. *Hum Mutat* 23: 47-56.

Hers HG, 1963. Alpha-glucosidase deficiency in generalized glycogen storage Disease (Pompe's Disease). *Biochem J.* 86: 11-16.

Hiei Y, Ohta S, Komari T, Kumashiro T, 1994. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *The Plant Journal* 6: 271-282.

Hirschhorn R, Reuser AJ, 2001 The metabolic and molecular bases of inherited disease. *McGraw Hill*, New York, pp: 3389-3420.

Hollams EM, Giles KM, Thomson AM, Leedman PJ, 2002. mRNA stability and the control of gene expression: implications for human disease. *Neurochem Res* 27: 957-980.

Hopwood JJ, Brooks DA, 1997. An introduction to the basic science and biology of the lysosome and storage disease. *Organelle diseases*. London: Chapman Hall: p. 7-35.

Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E, 1989. The human glucocerebrosidase gene and pseudogene: structure and evolution. *Genomics* 4: 87-96.

Huang S, Gilbertson LA, Adams TH, Malloy KP, Reisenbigler EK, Birr DH, Snyder MW, Zhang Q, Luethy MH, 2004. Generation of marker-free transgenic maize by regular two-border *Agrobacterium* transformation vectors. *Transgenic Res* 13: 451-461.

Hunziker W, Spiess M, Semenza G, Lodish HF, 1986. The sucrase-isomaltase complex: Primary structure, membrane-orientation and evolution of a stalked, intrinsic brush border protein. *Cell* 46: 227-234.

Ikemura T, 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2(1): 13-34.

Imai K, 1985. Characterization of beta-glucosidase as a peripheral enzyme of lysosomal membranes from mouse liver and purification. *J Biochem (Tokyo)* 98:1405-1416.

Jeffrey PL, Brown DH, Brown B, 1970. Studies of lysosomal α -glucosidase. I. Purification and properties of the rat liver enzyme. *Biochemistry* 9: 1403-1415.

Jeffrey PL, Brown DH, Brown B, 1970. Studies of lysosomal α -glucosidase. Kinetics of action of the rat liver enzyme. *Biochemistry* 9: 1416-1423.

Jmoudiak M, Futerman AH, 2005. Gaucher disease: pathological mechanisms and modern management. *Br J Haematol* 129: 178-188.

Kawabe A, Miyashita NT, 2003. Patterns of codon usage bias in three dicot and four monocot plant species. *Genes Genet Syst* 78: 343-52.

Khush GS, 1997. Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 35: 25-34.

Kim SR, Lee J, Jun SH, Park S, Kang HG, Kwon S, An G, 2003. Transgenic structures in T-DNA-inserted rice plants. *Plant Mol Biol* 52: 761-773.

Kleijer WJ, van der Kraan M, Kroos MA, Groener JE, van Diggelen OP, Reuser AJJ, 1995. Prenatal diagnosis of glycogen storage disease type II: enzyme assay or mutation analysis? *Pediatr Res* 38: 103-6.

Kliman RM, Irving N, Santiago M, 2003. Selection conflicts, gene expression, and codon usage trends in yeast. *J Mol Evol* 57: 98-109.

Knirsch L, Clerch LB, 2000. A region in the 3' UTR of MnSOD RNA enhances translation of a heterologous RNA. *Biochem Biophys Res Commun* 272:164-168.

Komari T, Hiei Y, Saito Y, Murai N, Kumashiro T, 1996. Vectors carrying two separated T-DNAs for co-transformation of higher plants mediated by *Agrobacterium tumefaciens* and segregation of transformants free from selection markers. *Plant J* 10: 165-174.

Kononov ME, Bassuner B, Gelvin SB, 1997. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. *Plant J* 11:945-957.

Koprivova A, Stemmer C, Hoffmann A, et al., 2004. Targeted knockouts of *Physcomitrella* lacking plant-specific immunogenic N-glycans. *Plant Biotechnology Journal* 2: 517-523.

Koster JF, Slee RG, 1977. Some properties of human liver acid alpha-glucosidase. *Biochim Biophys Acta* 482: 89-97.

Kozak M, 1989. The scanning model for translation: an update. *J Cell Biol* 108: 229-241.

Kusnadi A, Nikolov ZL, Howard JA, 1997. Production of recombinant proteins in transgenic plants: practical consideration. *Biotechnol Bioeng* 56: 473-484.

Lachmann RH, Grant IR, Halsall D, Cox TM, 2004. Twin pairs showing discordance of phenotype in adult Gaucher's disease. *QJM* 97: 199-204.

Lamacchia C, Shewry PR, Di Fonzo N, Forsyth JL, Harris N, Lazzeri PA, Napier JA, Halford NG, Barcelo P, 2001. Endosperm specific activity of a storage protein gene promoter in transgenic wheat seed. *J Exp Bot* 52: 243-250.

Latham T, Grabowski GA, Theophilus BDM, Smith FI, 1990. Complex alleles of the acid beta-glucosidase gene in Gaucher disease. *Am J Hum Genet* 47: 79-86.

Li X, Wu Y, Zhang DZ, Gillikin JW, Boston RS, Franceschi VR, Okita TW, 1993. Rice prolamine protein body biosynthesis: a BiP-mediated process. *Science* 262: 1054-1056.

Lienard D, Sourrouille C, Gomord V, Faye L, 2007. Pharming and transgenic plants. *Biotechnol Ann Rev* 13: 115-147.

Lin JJ, 1995. Electrotransformation of *Agrobacterium*. In *Methods in Molecular Biology*, Nickoloff, JA, ed., Humana Press, Totowa, NJ.

Ludwig T, Munier-Lehmann H, Bauer U, Hollinshead M, Ovitt C, Lobel P, Hoflack B, 1994. Differential sorting of lysosomal enzymes in mannose 6-phosphate receptor-deficient fibroblasts. *EMBO J* 13: 3430-3437.

Ma JK-C, Drake PMW, Christou P, 2003. The production of recombinant pharmaceutical proteins in plants. *Nature Rev Genetics* 4: 794-804.

Malca I, Endo RM, Long MR, 1967. Mechanism of glucose counteraction of inhibition of root elongation by galactose, mannose and glucosamine. *Phytopathology* 57: 272-278.

Martineau B, Voelker TA, Sanders RA, 1994. On defining T-DNA. *Plant Cell* 6: 1032-1033.

Martiniuk F, Bodkin M, Tzall S, Hirschhorn R, 1991. Isolation and partial characterization of the structural gene for human acid alpha glucosidase. *DNA Cell Biol.* 10: 283-292.

Martiniuk F, Chen A, Mack A, Arvanitopoulos E, Chen Y, Rom WN, 1998. Carrier frequency for glycogen storage disease type II in New York and estimates of affected individuals born with the disease. *Am J Med Genet* 79: 69-72.

Mehler M, Di Mauro S, 1977. Residual acid maltase activity in late onset acid maltase deficiency. *Neurology* 27: 178-184.

Meikle PJ, Hopwood JJ, Clague AE, Carey WF, 1999. Prevalence of lysosomal storage disorders. *JAMA* 282: 249-254.

Merritt C, Rasoloson D, Ko D, Seydoux G, 2008. 3'-UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Curr Biol* 18: 1476-1482.

Miles JS e Guest JR, 1984. Nucleotide sequence and transcriptional start point of the phosphomannose isomerase gene (*manA*) of *Escherichia coli*. *Gene* 32: 41-48.

Muench DG, Wu Y, Coughlan SJ, Okita TW, 1998. Evidence for a cytoskeleton-associated binding site involved in prolamine mRNA localization to the protein bodies in rice endosperm tissue. *Plant Physiol* 116 : 559-569.

Navazio L, Baldan B, Mariani P, Gerwig GJ, Vliegenthart JF, 1996. Primary structure of the N-linked carbohydrate chains of calreticulin from spinach leaves. *Glycoconj J* 13: 977-983.

Nelson DL, Cox MM, 2002. *I principi di biochimica* di Lehninger: p.32-33.

Neufeld EF, 1991. Lysosomal storage disease. *Annu Rev Biochem* 60: 257-282.

Nicolino MP, Puech JP, Kremer EJ, Reuser AJJ, Mbebi C, Verdieresahuque M, 1999. Adenovirus-mediated transfer of the acid α -glucosidase gene into fibroblasts, myoblasts and myotubes from patients with glycogen storage disease type II leads to high level expression of enzyme and corrects glycogen accumulation. *Hum Mol Genet* 7: 1695-702.

O'Neill R, Tosbihou T, Kozak CA, Brady RO, 1989. Comparison of the chromosomal localization of murine and human glucocerebrosidase genes and of deduced amino acid sequences. *Proc Natl Acad Sci USA* 86: 5049-5053.

Ortega JL, Moguel-Esponda S, Potenza C, Conklin CF, Quintana A, Sengupta-Gopalan C (2006) The 3' untranslated region of a soybean cytosolic glutamine synthetase (GS1) affects transcript stability and protein accumulation in transgenic alfalfa. *Plant J* 45:832-846.

Osiecki-Newman IM, Fabbro D, Mnur T, Boas S, Gatt S, Legler G, Desnick RJ, Grabowski GA, 1986. Human acid β -glucosidase: affinity purification of the

normal placental and Gaucher disease splenic enzymes on N-alkyl-deoxynojirimycin-sepharose. *Enzyme* 3: 147-153.

Otten L, DeGreve H, Leemans J, Hain R, Hooykaas P, Schell J, 1984. Restoration of virulence of vir region mutants of *Agrobacterium tumefaciens* strain B6S3 by coinfection with normal and mutant *Agrobacterium* strain. *Mol Gen Genet* 195: 159-163.

Owada M, Neufeld EF, 1982. Is there a mechanism for introducing acid hydrolases into liver lysosomes that is independent of mannose 6-phosphate? Evidence from I-cell disease. *Biochem Biophys Res Commun* 105: 814-820.

Pagny S, Cabanes-Macheteau M, Gillikin JW, Leborgne-Castel N, Lerouge P, Boston RS, Faye L, Gomord V, 2000. Protein recycling from the Golgi apparatus to the endoplasmic reticulum in plants and its minor contribution to calreticulin retention. *Plant Cell* 12: 739-756.

Palmer TN, 1971. The substrate specificity of acid α -glucosidase from rabbit muscle. *Biochem J.* 124: 701-711.

Park HK, Kay HH, McConkie-Rosell A, Lanman J, Chen YT, 1992. Prenatal diagnosis of Pompe's disease (type II glycogenosis) in chorionic villus biopsy using maltose as a substrate. *Prenat Diagn* 12: 169-73.

Pauly DF, Johns DC, Matelis LA, Lawrence JH, Byrne BJ, Kessler PD, 1998. Complete correction of acid α -glucosidase deficiency in Pompe disease fibroblasts *in vitro*, and lysosomally targeted expression in neonatal rat cardiac and skeletal muscle. *Gene Ther* 5: 473-80.

Pompe JC, 1932. Over idiopatische hypertrophic van het hart. *Ned Tijdschr Geneeskde* 76: 304-311.

Powell JR e Moriyama E, 1997. Evolution of codon usage bias in *Drosophila*. *Proc Natl Acad Sci USA* 94: 7784-7790.

Qi X, Grabowski GA, 1998. Acid beta-glucosidase: Fluorescence and conformational alterations accompanying membrane interaction and insertion. *Biochemistry* 37:1154-11554.

Qu LQ, Yoshihara T, Ooyama A, Goto Y, Takaiwa F, 2005. Iron accumulation does not parallel the high expression level of ferritin in transgenic rice seeds. *Planta* 222: 225-233.

Qu LQ, Takaiwa F, 2004. Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. *Plant Biotechnol J* 2: 113-125.

Quaroni A, Semenza G, 1976. Partial amino acid sequences around the essential carboxylate in the active sites of the intestinal sucrase-isomaltase complex. *J Biol Chem* 251: 3250-3253.

Raben N, Plotz P, Byrne BJ, 2002. Acid alpha-glucosidase deficiency (glycogenosis type II, Pompe disease). *Curr Mol Med* 2: 145-166.

Rayon C, Lerouge P, Faye L, 1998. The protein N-glycosylation in plants. *J Exp Bot* 49: 1463-1472.

Reggi S, Marchetti S, Patti T et al., 2005. Recombinant human acid β -glucosidase stored in tobacco seeds is stable, active and taken up by human fibroblasts. *Plant Molecular Biology* 57: 101-113.

Reiner O, Widler S, Glvol D, Horowitz M, 1987. Efficient *in vitro* and *in vivo* expression of human glucocerebrosidase cDNA. *DNA* 6: 101-108.

Reiner O, Wigderson M, Horowitz M, 1988. Structural analyses of the human glucocerebrosidase genes. *DNA* 7: 107-116.

Reuser AJJ, Kroos M, Oude Elferink RPJ, Tager JM, 1985. Defects in synthesis, phosphorylation and maturation of acid alpha-glucosidase in glycogenosis type II. *J Biol Chem* 260: 8336-8341.

Reuser AJJ, Kroos M, Willemsen R, Swallow D, Tager JM, Galjaard H, 1987. Clinical diversity in glycogenosis type II. Biosynthesis and *in situ* localization of acid alpha-glucosidase in mutant fibroblasts. *J Clin Invest* 79: 1689-1699.

Richards SM, 2002. Immunologic considerations for enzyme replacement therapy in the treatment of lysosomal storage disorders. *Clinical and Applied Immunology Reviews* 2: 241-253.

Rijnboutt S, Aerts HM, Gauze HJ, Tager JM, Strous GJ, 1991. Mannose 6-phosphate-independent membrane association of cathepsin D, glucocerebrosidase, and sphingolipid-activating protein in HepG2 cells. *J Biol Chem* 266: 4862-4868.

Russell GJ, Walker PMB, Elton RA, Subak-Sharpe JH, 1976. Doublet frequency analysis of fractionated vertebrate nuclear DNA. *J Mol Biol* 108: 1-23.

Saint-Jore-Dupas C, Faye L, Gomord V, 2007. From planta to pharma with glycosylation in the toolbox. *Trends Biotechnol* 25: 317-323.

Sambrook J, Fritsch EF, Maniatis T, 1989. Molecular cloning: a laboratory manual 2nd edition. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY.

Sanger M, Daubert S, Goodman RM, 1990. Characteristics of strong promoter from figwort mosaic virus: comparison with the analogous 35S promoter from cauliflower mosaic virus and the regulated mannopine synthase promoter. *Plant Mol Biol* 14: 433-443.

Schillberg S, Zimmermann S, Voss A, Fischer R, 1999. Apoplastic and cytosolic expression of full-size antibodies and antibody fragments in *Nicotiana tabacum*. *Transgenic Res* 8: 255-263.

Shaaltiel Y, Barfeld D, Hashmueli S et al., 2007. Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol J* 5: 579-590.

Sidransky E, Bottler A, Stubblefield B, Ginns EI, 1994. DNA mutational analysis of type 1 and type 3 Gaucher patients: How well do mutations predict phenotype? *Hum Mutat* 3: 25-28.

Sorge J, West C, Westwood B, Beutler E, 1985. Molecular cloning and nucleotide sequence of human glucocerebrosidase cDNA. *Proc Natl Acad Sci USA* 82: 7289-7293.

Stachel SE, Nester EW, Zambryski PC, 1986. A plant cell factor induces *Agrobacterium tumefaciens vir* gene expression. *Proc Natl Acad Sci USA* 83: 379-383.

Streatfield SJ, 2007. Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnol J* 5: 2-15.

Stockert RJ, 1995. The asialoglycoprotein receptor: relationships between structure, function, and expression. *Physiol Rev* 75: 591-609.

Stoger E, Marcel S, Twyman RM, 2004. Crop plants for molecular farming. In: *Handbook of Plant Biotechnology*. Eds.: Paul Christou and Harry Klee. John Wiley and Sons Ltd., pp. 747-758.

Takaiwa F, Yang L, Yasuda H (2008) Health-promoting transgenic rice: application of rice seeds as a direct delivery system for bioactive peptides in

human health. In: Hirano HY, Hirai A, Sano Y, Sasaki T (eds) *Biotechnology in agriculture and forestry*, vol 62. Springer-Verlag, Berlin, pp 357–373.

Tayebi N, Stubblefield BK, Park JK, Orvisky E et al., 2003. Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in Gaucher disease. *Am J Hum Genet* 72 (3): 519-534.

Tsuji S, Choudazy PV, Martin BM, Winfield S, Barranger JA, Ginns EI, 1986. Nucleotide sequence of cDNA containing the complete coding sequence for human lysosomal glucocerebrosidase. *J Biol Chem* 261: 50-53.

Tsuji A, Omura K, Suzuki Y, 1988. Intracellular transport of acid alpha-glucosidase in human fibroblasts: evidence for involvement of phosphomannosyl receptor-independent system. *J Biochem (Tokyo)* 104: 276-278.

Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R, 2003. Molecular pharming in plants: host system and expression technology. *Trends Biotechnol* 21: 570-578.

Usui Y, Nakase M, Hotta H, Urisu A, Aoki N, Kitajima K, Matsuda T, 2001. A 33 kDa allergen from rice (*Oryza sativa* L. *japonica*): DNA cloning, expression, and identification as a novel glyoxalase I. *J Biol Chem* 276: 11376-11381.

Van Der Borght A, Vandeputte GE, Derycke V, Brijs K, Daenen G, Delcour JA, 2006. Extractability and chromatographic separation of rice endosperm proteins. *Journal of Cereal Science* 44: 68-74.

Van Der Graaff E, den Dulk-Ras A, Hooykaas PJJ, 1996. Deviating T-DNA transfer from *Agrobacterium tumefaciens* to plants. *Plant Mol Biol* 13: 677-681.

Van Hove JL, Yang HW, Wu JY, Brady RO, Chen YT, 1996. High-level production of recombinant human lysosomal acid alpha-glucosidase in chinese hamster ovary cells which targets to heart muscle and corrects glycogen

accumulation in fibroblasts from patients with Pompe disease. *Proc Natl Acad Sci USA* 93(1): 65-70.

Van Larebeke N, Engler G, Holsters M, Van den Elsacker S, Zaenen I, Schilperoort RA, Schell J, 1974. Large plasmid in *Agrobacterium tumefaciens* essential for crown gall inducing ability. *Nature* 252: 169-170.

Vasconcelos M, Datta K, Oliva N, Khalekuzzaman M, Torrizo L, Krishnan S, Oliveira M, Goto F, Datta SK, 2003. Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. *Plant Science* 164: 371-378.

Venturini E, 2006. Nuovo metodo di design genico fondato sulle regole di vicinanza codonica: applicazioni per la sintesi di inteine utilizzabili in processi di purificazione proteica. Università degli studi di Udine.

Viotti A, Balducci C, Weil JH, 1978. Adaptation of the tRNA population of maize endosperm for zein synthesis. *Biochim Biophys Acta* 517: 125-132.

Vitale A, Denecke J, 1999. The endoplasmic reticulum-gateway of the secretory pathway. *Plant Cell* 11 : 615-628.

Walsh G, Jefferis R, 2006. Post-translational modifications in the context of therapeutic proteins. *Nat Biotechnol* 24: 1241-1252.

Waterhouse PM, Wang MB, Lough T, 2001. Gene silencing as an adaptive defence against viruses. *Nature* 441:834-842.

Weinreb N, Brady RO, Tappel AL, 1968. The lysosomal localization of sphingolipid hydrolases. *Biochem Biophys Acta* 159: 141-146.

Williamson JR, 1994. G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct* 23: 703-730.

Yan X, Gonzales RA, Wagner GJ, 1997. Gene fusions of signal sequences with a modified β -glucuronidase gene results in retention of the β -glucuronidase protein in the secretory pathway/plasma membrane. *Plant Physiol* 115: 915-924.

Yang D, Guo F, Liu B, Huang N, Watkins SC, 2003. Expression and localization of human lysozyme in the endosperm of transgenic rice. *Planta* 216: 597-603.

Yang J, Zhang J, Huang Z, Wang Z, Zhu Q, Liu L, 2002. Correlation of cytokinin levels in the endosperms and roots with cell number and cell division activity during endosperm development in rice. *Annals of Botany* 90: 369-377.

Ye X, Al-Babili S, Kloti A, Zhang J, Lucca P, Beyer P, Potrykus I, 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287: 303-305.

Yin Z e Wang GL, 2000. Evidence of multiple complex patterns of T-DNA integration into the rice genome. *Theor Appl Genet* 100: 461-470.

Yu et al., 2005. Rice glutelin gene promoters. *United States Patent Application* 20050066387.

Yuan Q, Quackenbush J, Sultana R, Pertea M, Salzberg SL, Buell CR, 2001. Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. *Plant Physiol* 125: 1166-1174.

Zamore PD, Tuschl T, Sharp PA, Bartel DP, 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101: 25-33.

Zaretsky JZ, Candotti F, Boerkoel C, Adams EM, Yewdell JW, Blaese RM, 1997. Retroviral transfer of acid α -glucosidase cDNA to enzyme-deficient myoblasts results in phenotypic spread of the genotypic correction by both secretion and fusion. *Hum Gene Ther* 8: 1555-63.

Zhu Y, Li X, McVie-Wylie A, Jiang C, Thurberg BL, N. Raben, Mattaliano RJ, Cheng SH, 2005. Carbohydrate-remodelled acid alphaglucoosidase with higher affinity for the cation-independent mannose 6-phosphate receptor demonstrates improved delivery to muscles of Pompe mice. *Biochem J* 389: 619-628.

Zimran A, Sorge J, Gross E, Kubitz M, West C, Beutler E, 1990. A glucocerebrosidase fusion gene in Gaucher disease. Implication for the molecular anatomy, pathogenesis and diagnosis of this disorder. *J Clin Invest* 85: 219-222.

Xu YH, Grabowski GA, 1998. Translational inefficiency of acid β -glucosidase mRNA in transgenic mammalian cells. *Mol Genet Metab* 64: 87-98.