The Journal of Animal & Plant Sciences, 25(1): 2015, Page: 254-260
ISSN: 1018-7081

CODON OPTIMIZATION OF *SERRATIA MARCESCENS CHI*A AND ITS EXPRESSION IN TOBACCO

M. A. Pasha*, A. Belgaumwala, R. Kumar, P. U. Krishnaraj and M. S. Kuruvinashetti

*Department of Biotechnology, University of Agricultural Sciences, Dharwad-580005, Karnataka, India
Corresponding Author e-mail:malikiabt@gmail.com

ABSTRACT

Most plant diseases are caused by fungal pathogens, which contain chitin as their major cell wall component, except Oomycetes. Chitinases cleave between C1 and C4 of two consecutive N-acetylglucosamine of chitin. Among bacterial chitinases, *Serratia marcescens* chitinases have been studied well. In order to develop intrinsic resistance, genes from various sources have been cloned and expressed in plants but in most cases, the transgene failed to produce desired quantity of foreign protein in the heterologous system. Codon optimization is a general approach for improving expression of gene in any heterologous system. Full length *chi*A (1692 bp) from *S. marcescens* 141 (*sm141chi*A) was considered for removal of regulatory signals (partial optimized- *sm141chi*A*pm*) and replacement of rare codons with preferred codons by tobacco (fully optimized- *sm141chi*A*fm*). Of 1692 bp, 22 and 432 nucleotides were altered to get partial and fully optimized genes, respectively. The synthesized gene was sub-cloned in plant expression vectors and transferred to tobacco. The plants containing *sm141chi*A*pm* produced 1.82 fold more reducing equivalents whereas plants containing *sm141chi*A*fm* produced 4.39 fold more reducing equivalents compared to the plants containing *sm141chi*A. Optimization of bacterial codons to suit the plant codon dictionary improved the expression level of chitinase in tobacco by 2.57 fold. For high level expression of transegene in heterologous system, the codon frequency of transgene should be similar to the codons preferred by high level expressing genes in host.

Key words: Chitinase, *Serratiamarcescens*, codon optimization, transgenic tobacco, regulatory sequences.

INTRODUCTION

Diseases cause significant loss in crop production. Majority of plant diseases are caused by fungi and most of the fungi contain chitin as major cell wall component. Endochitinases (pathogenesis related type 3 protein) catalyze the hydrolysis of chitin which leads to inhibition of hyphal elongation and lysis of cell (Gokul *et al.,* 2000). Almost all organisms including bacteria produce chitinases. Among bacteria, chitinases from *Serratia marcescens* have been studied well. It produces *chi*A, *chi*B, *chiC,* chitobiase and chitin binding protein, of which *chi*A has been found to be most effective (Brurberg *et al.,* 2000). In order to develop intrinsic resistance, genes from various sources have been cloned and expressed in plants. In most cases, the transgene failed to produce desired quantity of foreign protein in the heterologous system (Haseloff *et al.,* 1997). Protein expression, especially in a heterologous system, is governed by many factors including those that affect transcription, mRNA processing, its stability and initiation of translation (Conway and Wickens, 1988). Degeneracy of genetic code is yet another factor. Codon optimization is one of the techniques developed to increase the protein expression in plants (Streatfield, 2007). Several authors reported the increased expression of heterologous gene in plants by codon optimization (Xiu-ying *et al.,* 2013; Weng *et al.,* 2011 and Jabeen *et al.,* 2010). Codon optimization by codon randomization strategy increases the accumulation of calf prochymosin in *E. coli* by 70 % (Menzella, 2011). The DNA sequence used to encode a protein in one organism is often quite different from the sequences that would be used to encode the same protein in another organism. To enhance the expression level of a foreign gene in a particular expression system, it is important to adjust the codon frequency of the transgene to match that of the host expression system. This study aims at modification of codons and other regulatory elements of *chiA* from *S. marcescens* and check its transient expression in tobacco plant.

MATERIALS AND METHODS

Full length coding sequence of *S. marcescens* 141 *chiA*(*sm*141*chiA)* cloned earlier at the Department of Biotechnology, University of Agricultural Sciences, Dharwad, India and submitted to NCBI (No.DQ990373) was considered for modification (Ningaraju, 2006). A codon usage table for tobacco was downloaded from www.kazusa.or.jp/codon/. A modified codon usage table for tobacco was constructed by taking 10% threshold level; all frequencies below 10 per cent were set to zero, so that rare codons were completely eliminated. By using Evolving Code's online tool, Synthetic Gene Designer, the nucleotide sequences were back translated to their

amino acid sequence, and then the amino acids were converted back to nucleotide sequences by referring to the modified codon preference table for tobacco. Regulatory sequences viz, premature mRNA termination signal, mRNA degradation signals, intron splice sites and mRNA stable secondary signals were also removed by degenerate codons manually. PolyAH and splice sites were predicted with the help of online tool Softberry (www.softberry.com), which predicts the potential splice sites present in the genomic DNA. The designed sequence were named as *sm*141*chiAfm* (full or complete optimized) and *sm*141*chiApm* (partial optimized) where only signals were removed.

Designed sequences were custom synthesized at Geneart, Germany, subcloned in plant transformation vector, pHS100 at *Xba*I and *Bam*HI sites. pHS100 containing *sm*141*chiA*, *sm*141*chiApm* and *sm*141*chiAfm* were named as pNKK0205, pMAPKK0302 and pMAPKK0302, respectively and were mobilized in *Agrobacterium tumefaciens* LBA4404 with the help of *tra* plasmid, pRK2013. Finally, tobacco transformation was done by protocol of Hooykaas and Schilperoopt (1992).

The fully regenerated plants on MS media containing 50 mg/ml Kanamycin were screened with gene specific primers. Total protein from PCR positive plants was isolated according to Carmenza and Roxanne (2002). The level of chitinase expression in plants was quantified by enzymatic activity assay as described by Omumasaba *et al.* (2001) with slight modifications. To 0.5 ml of 0.2 % colloidal chitin (prepared in 0.1 M of sodium acetate buffer, pH 5.0) 200 µg of total protein was added and incubated at 45ºC for 30 minutes. The reaction was stopped by addition of 0.5 ml of dinitro salicylic acid and boiling for 10 minutes. The reaction volume was made up to 3 ml by adding water and reducing sugars released were estimated by DNSA method with N Acetyl D glucosamine as standard at 540 nm.

RESULTS AND DISCUSSION

In this study, efforts were made to modify *chiA* from *S. marcescens*141 to make it “plant like” by modifying its codon usage pattern to that of tobacco. The native *chiA* cloned previously from *S. marcescens*141 (Ningaraju, 2006) submitted to NCBI (Accession DQ990373) was modified with the aim of achieving high expression levels in transgenic tobacco plants. The full length nucleotide sequence of native *chiA* was considered for modification to overcome codon bias and other undesirable regulatory sequences for its enhanced expression in tobacco.

G+C composition is the major factor affecting codon usage variation. *In silico* analysis of 1357 tobacco CDSs showed that for amino acids *viz.*, alanine, arginine, threonine, proline and serine, codons ending with CG, GC and GG are preferred less than 10 per cent. The codons ending with CC are also preferred less but above 10 per cent. In case of leucine and valine, among A/T dinucleotide, the codons ending with TA are least preferred. The differences found in coding sequences became much larger when third codon positions were compared. At the wobble position, among T and C, T is preferred to C and among A and G, in most of the cases, A is preferred to G except for leucine, lysine and valine. In contrary to codon usage pattern of tobacco, *in silico* analysis of coding region of *sm*141*chiA* revealed that in almost all the cases, codons ending with CG, GC, GG and CC were preferred. At the wobble position, among T and C, C is preferred to T and among A and G, G is preferred to A.

When the nucleotide sequences of *sm*141*chiA* were analyzed, we found 58.5 per cent overall G+C content and 26.7 per cent at wobble position which resembles the overall G+C content of prokaryotes which vary from 25 per cent to 75 per cent (Sueoka, 1962) and more than 20 per cent to less than 90 per cent at third position (Muto and Osawa, 1987). But when the same analysis was done for modified codon usage table for tobacco, we found 43.62 per cent overall G+C content and 39.84 per cent at wobble position which is similar to the overall G+C content of plants, especially dicots which is around 45 per cent and 38.8 per cent at wobble position (Kawabe and Miyashita, 2003). The details of codon usage pattern of tobacco, modified codon usage table for tobacco, *sm*141*chiA*, *sm*141*chiApm* and *sm*141*chiAfm* is presented in table 1.

For the amino acids like alanine, arginine, threonine, proline and serine, codons ending with CG, GC and GG are least preferred in tobacco. The codons ending with CC are also preferred less but above cut off value (10 per cent). In case of leucine and valine, among A/T dinucleotide, the codons ending with TA are least preferred. These results are in accordance with the results obtained by Elizabeth *et al.* (1989) wherein the codons ending with G for threonine, proline, alanine and serine and the codons ending with CG dinucleotide are strongly avoided and the doublet TA is also avoided in codon position II and III in most eukaryotes. This is to avoid regulation due to methylation in plants and other eukaryotes (Elizabeth *et al.*, 1989).

The nucleotide sequence adapted for Gram negative *S. marcescens* may not be the appropriate sequence for efficient plant expression. Examination of coding sequence of *sm*141*chiA* indicated the presence of two ATTTA, three-AATAAA, one-AATAAT and one {CAN(7-9)AGTNNA}signals that have a regulatory significance in eukaryotes. When the sequence was fed to Softberry for the analysis of potential intron splice sites, we found three splice acceptor (AG) and six splice donor (GT) sites. These signals are similar to the signals found

in wild type *cryIA(b)* which are found to be mRNA degradation, polyadenylation, RNA polymerase II termination signals and potential intron splice sites (Perlak *et al.*, 1991). Similar to the studies of Perlak *et al.* (1991), to differentiate whether the enhanced expression of transgene in plant is due to improved translation efficiency (by codon optimization) or by the combination effect of mRNA stability and improved translation efficiency, two sets of nucleotides were designed, one by removing only the regulatory sequences without considering codon usage pattern of tobacco, which was termed as partial optimization (*sm*141*chiApm*). Another set with codon optimization and removal of regulatory sequences, was termed as complete optimization (*smchiA141fm*). To enhance the expression of *cry1A(b)* in tobacco, Perlak *et al.* (1991) modified 3 per cent and 21 per cent of nucleotide sequences to get partially and fully modified *cry1A(b)*. In the present study, 1.3 per cent and 25.24 per cent of nucleotides were changed to get *sm*141*chiApm* and *smchiA141fm* respectively. After codon optimization, *sm*141*chiA* had its GC per cent reduced from 58.5 per cent to 44.0 per cent, similar to GC content of tobacco. The differences between CDS's of tobacco and *sm141chiA* and the modifications done to design *sm*141*chiApm* and *sm*141*chiAfm* are presented in table 2.

The secondary structure of mRNA can severely affect gene expression by influencing translation initiation efficiency (Gualerzi and Pon, 1990). A very stable structure within the leader (ΔG > -50 kcal/mol) completely block ribosome scanning, and a moderately strong hairpin (-30 kcal/ mol) located near the 5' end repressed translation by influencing the binding of the pre-initiation complex to the mRNA (Kozak, 1989). When the nucleotide sequence of *Sm*141*chiA* was analyzed for the stable secondary structure, no structure with >-30kcal/mol energy has been found when predicted using Genebee tool. The highest stable secondary structure found *in Sm*141*chiA* is ΔG = -22.9kcal/mol. (Supplementary figures S1, S2 and S3).

To validate the expression of codon optimized *chiA* gene in tobacco, *sm*141*chiA*, *sm*141*chiApm* and *sm*141*chiApfm* were cloned in plant transformation vector pHS100 and introduced by *Agrobacterium* mediated transformation into tobacco. All the transgenic events in tobacco were confirmed by PCR using gene specific primers (Fig1). Transgenic plants derived were assayed for the synthesis of chitinase A in leaf by reducing sugar estimation. The amount of monomers of glycol chitin released and the enzyme unit in each treatment is given in table 3. The non-transgenic plants and the plants containing *sm*141*chiA*, *sm*141*chiApm* and *sm*141*chiApfm* produced reducing equivalent of 0.1387, 0.1997, 0.2502 and 0.4069 µmoles/min/mg of protein, respectively. Considering chitinase activity of the non-transgenic plants, the plants containing *sm*141*chiA*, *sm*141*chiApm* and *sm*141*chiApfm* produced 1.44, 1.80 and 2.93 fold more reducing equivalents, respectively. The plants containing *sm*141*chiApm* produced 1.82 fold more reducing equivalents whereas the plants containing *sm*141*chiAfm* produced 4.39 fold more reducing equivalents compared to plants containing *sm*141*chiA*. This difference in expression between native and modified genes is due to codon optimization and removal of potential regulatory sequences from bacterial *chiA*. But there was 2.57 fold more production of reducing equivalent in *sm*141*chiAfm* compared to *sm*141*chiApm*, which may be due to improved translation efficiency by substituting the rare codons with more preferred codons by plants. Perlak *et al.* (1991) modified the coding sequences of *cryIA(b)* to improve its expression in plants by removing only the regulatory sequences (partially optimized) and by substituting rare codons with more preferred codons along with removal of regulatory sequence (fully optimized). They got 10 fold more production of CryIA(b) in plants containing partially optimized gene and 100 fold more production in plants containing fully optimized gene compared to plants containing unmodified *cryIA(b)*.

This study revealed that the bias in codon preference could be one of the factors in getting reduced levels of expression in transgenics. In our transient expression analysis, we observed only four fold increased expression. Our analysis is based on few positive plants; the results could be more promising with screening more number of positive plants.

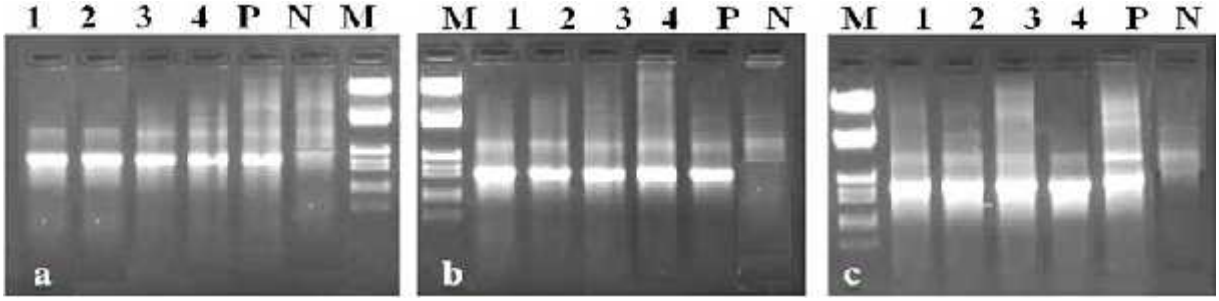

Figure1. PCR confirmation of *sm*141*chiA* (a), *sm*141*chiApm* (b) and *sm*141*chiAfm* (c) in tobacco plants. 1-4 transgenic plants, P – *sm*141 *chiA* in *Agrobacterium* (Positive control), N – Non-transgenic tobacco plant (Negative control) and M - ^{a}*Eco* RI/*Hind* III double digest marker

Table 1. Original and modified codon dialect of tobacco and codon usage pattern in *sm*141*chiA*, *sm*141*chiApm* and *sm*141*chiAfm*

Amino Acid	codon	Tobacco codon usage table form Kazusa (%)	Modified codon usage table for tobacco (%)	Codon usage pattern in *sm*141*chiA* (%)	Codon usage pattern in *sm*141*chiApm* (%)	Codon usage pattern in *sm*141*chiAfm* (%)	Amino Acid	codon	Tobacco codon usage table form Kazusa (%)	Modified codon usage table for tobacco (%)	Codon usage pattern in *sm*141*chiA* (%)	Codon usage pattern in *sm*141*chiApm* (%)	Codon usage pattern in *sm*141*chiAfm* (%)
Ala	GCG	8	0	43	43	0	Leu	TTG	24	26	18	20	27
	GCA	31	34	4	4	33		TTA	14	16	7	4	16
	GCT	43	47	6	9	48		CTG	11	12	73	73	11
	GCC	17	19	48	44	19		CTA	10	0	0	0	0
Arg	AGG	25	30	0	0	33		CTT	27	30	0	0	31
	AGA	32	38	0	0	42		CTC	14	16	2	2	16
	CGG	8	0	8	8	0	Lys	AAG	51	51	48	48	50
	CGA	11	13	0	0	8		AAA	49	49	52	52	50
	CGT	16	19	8	8	17	Met	ATG	100	100	100	100	100
	CGC	8	0	83	83	0	Phe	TTT	58	58	12	8	58
Asn	AAT	60	60	47	44	59		TTC	42	42	88	92	42
	AAC	40	40	53	56	41	Pro	CCG	10	0	75	75	0
Asp	GAT	69	69	35	35	70		CCA	39	43	0	0	46
	GAC	31	31	65	65	30		CCT	37	41	21	21	42
Cys	TGT	57	57	20	20	60		CCC	13	15	4	4	13
	TGC	43	43	80	80	40	Ser	AGT	17	18	6	6	19
Stop	TGA	39	0	0	0	0		AGC	13	14	34	31	13
	TAG	20	0	0	0	0		TCG	7	0	19	19	0
	TAA	42	100	100	100	100		TCA	23	25	3	3	25
Gln	CAG	43	43	77	65	42		TCT	26	28	6	9	28
	CAA	57	57	23	35	58		TCC	14	15	31	31	16
Glu	GAG	46	46	63	63	47		ACG	9	0	23	23	0
	GAA	54	54	37	37	53	Thr	ACA	33	36	0	0	43
Gly	GGG	15	15	8	8	16		ACT	39	43	0	0	37
	GGA	34	34	2	2	34		ACC	19	21	77	77	20
	GGT	33	33	10	7	33	Trp	TGG	100	100	100	100	100
	GGC	17	17	80	84	16	Tyr	TAT	57	57	45	45	55
His	CAT	61	61	50	50	67		TAC	43	43	55	55	45
	CAC	39	39	50	50	33		GTG	25	25	51	51	27
Ile	ATA	25	25	0	0	26	Val	GTA	17	17	3	0	16
	ATT	50	50	22	22	48		GTT	41	41	22	22	41
	ATC	25	25	78	78	26		GTC	17	17	24	27	16

Table 2. Comparison between *sm*141*chiA*, *sm*141*chiApm* and *sm*141*chiAfm*

S No	Signal	*sm*141*chiA*	*sm*141*chiApm*	*sm*141*chiAfm*
1	GC content (%)	58.5	58.5	44.0
2	mRNA degradation signal (ATTTA)	2	0	0
	Polyadenylation signal			
	AATAAA	3	0	0
3	AATAAT	1	0	0
	AATTAA	0	0	0
	AACCAA	0	0	0
4	PolyAH	1	0	0

	Intron Splice site			
5	Acceptor site (AG)	3	0	0
	Donor site (GT)	6	0	0
6	RNA polymerase II termination signal {CAN(7-9)AGTNNA}	1	0	0
7	mRNA secondary structure (Gibbs energy G>-30 kcal/mol)	0	0	0
7	Sites for common restriction endonucleases	*Eco*RV, *Hae*III, *Sma*I, *Xma*I	*Eco*RV, *Hae*III, *Sma*I, *Xma*I	*Eco*RV, *Hae*III, *Hind*III, *Kpn*I
8	Number of nucleotides to be changed	0	22	432
9	% similarity with *SmchiA141*	100	98.70	74.46

Table 3. Amount of chitinaseA produced in transgenic plants

Plant	Non transgenic-1	Non transgenic-2	*sm*141*chiA*-20 (C_2-20)	*sm*141*chiA* (C_2-21)	*sm*141*chiApm* (P_1-81)	*sm*141*chiApm* (P_1-83)	*sm*141*chiAfm* (P_2-99)	*sm*141*chiAfm* (P_2-120)
Total protein used (µg). A	200	200	200	200	200	200	200	200
Values on X-axis. B	168	200	280	250	330	320	560	520
Equiv. of NAG released B/221.2 /30min (µmoles/min)	0.025	0.030	0.042	0.038	0.052	0.048	0.084	0.078
Specific activity (µmoles/min/mg of protein)	0.126	0.151	0.211	0.188	0.259	0.241	0.422	0.392
Average	0.1387		0.1997		0.2502		0.4069	
Corrected specific activity for non-transgenics			0.061		0.1115 (1.82 fold)		0.2682 (4.39 fold)	

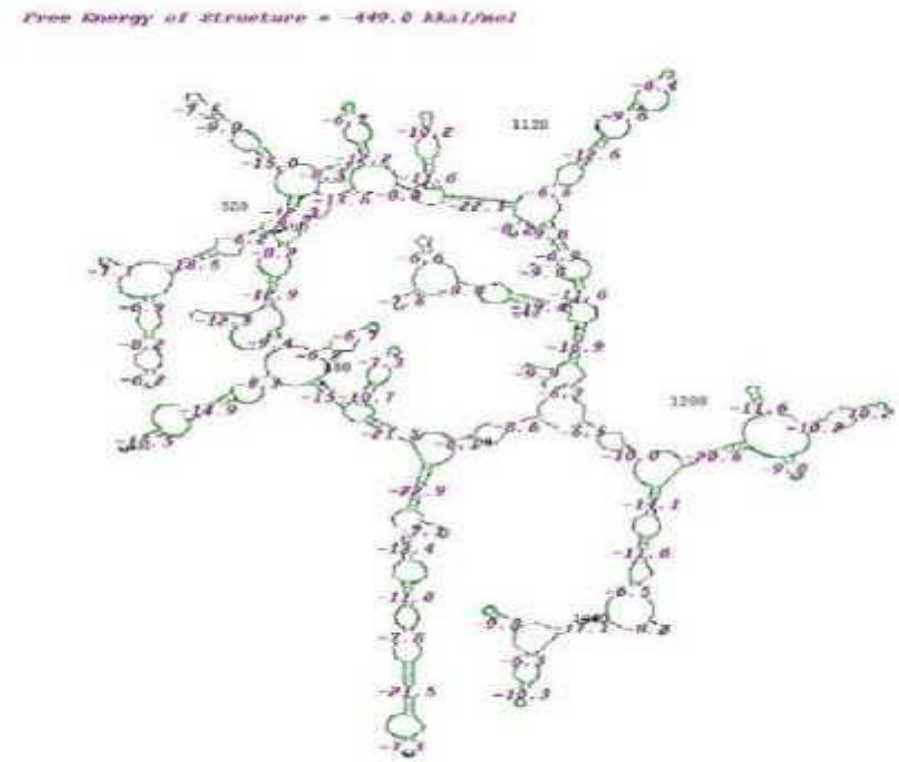

Figure S1. Predicted mRNA secondary structure of *Sm*141*chiA* using Genebee tool.

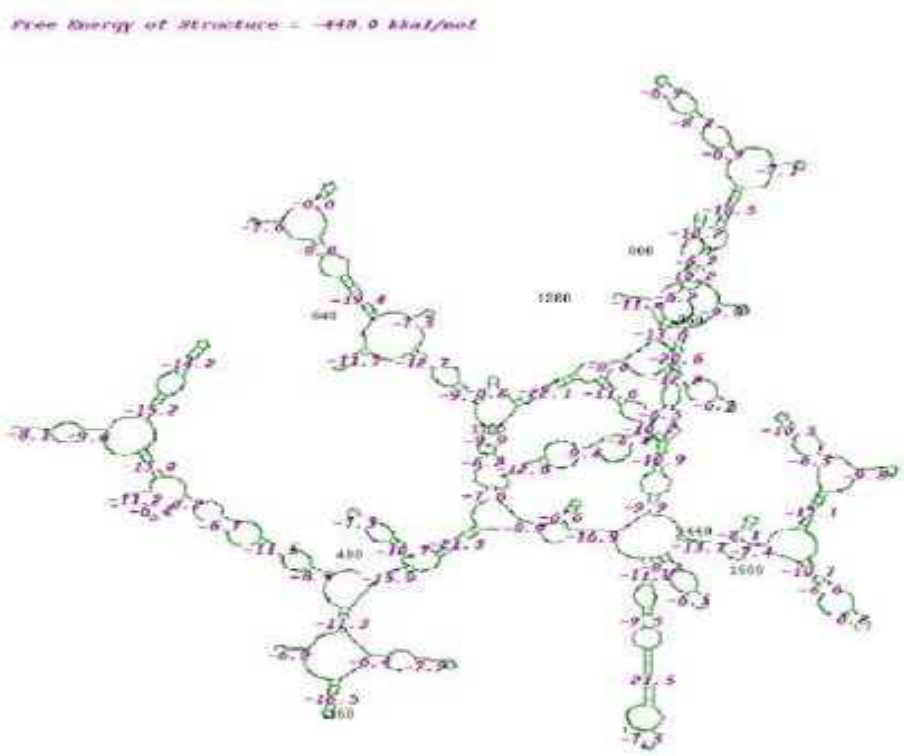

Figure S2. Predicted mRNA secondary structure of *Sm*141*chiApm* using Genebee tool.

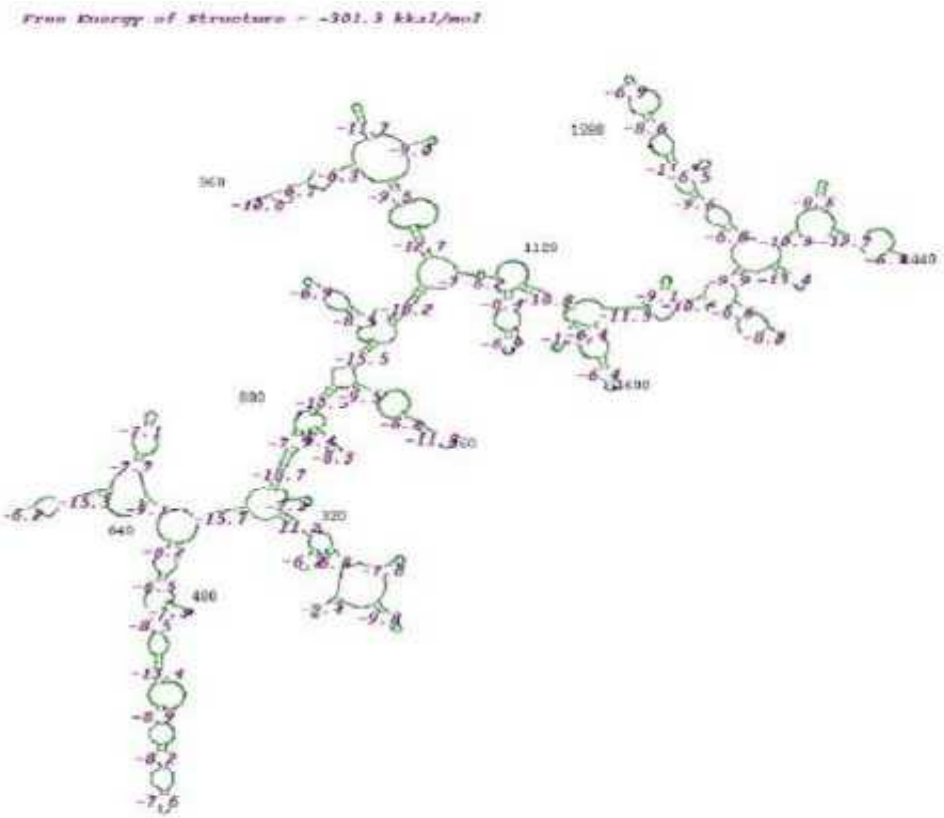

Figure S3. Predicted mRNA secondary structure of *Sm*141*chiAfm* using Genebee tool.

Acknowledgements: We would like to thank Department of Biotechnology, Government of India for funding this project.

REFERENCES

Brurberg, M. B., I. F. Nes, and V. G. H. Eijsink (2000). Comparative studies of chitinase A and B fom *Serratia marcescens*. Microbiol., 142**:**1581-1589

Carmenza, E. G. and M. B. Roxanne (2002). Plant growth and development influenced by transgenic insertion of bacterial chitinolytic enzymes. Mol. Breed., 9**:**123–135.

Conway, L. and M. Wickens (1988). RNA Processing, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York pp. 40

Elizabeth, E. M., L. Jeff and E. Mery (1989). Codon usage in plant genes. Nucleic Acids Res., 17(2):125-129.

Gokul, B., J. H. Lee, K. B. Song, S. K. Rhee, C. H. Kim and T. Panda (2000). Characterization and application of chitinases from *Trichodermaharzianum* - a review. Bioprocess Engg., 23:691-694

Gualerzi, C. O. and C. L. Pon (1990). Initiation of mRNA translation in prokaryotes. Biochem., 29: 5881-5889.

Haseloff, J., K. R. Siemering, D. Prasher and S. Hodge (1997). Removal of cryptic intron and subcelluar localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. Proc. Natl. Acad. Sci. USA., 94:2122-2127

Hooykaas, P. J. J. and Schilperoort (1992). *Agrobacterium* and plant genetic engineering. Plant Mol. Biol., 19:15-38.

Jabeen, R., M. S. Khan, Y. Zafar and T. Anjum (2010). Codon optimization of cry1Ab gene for hyper expression in plant organelles. Mol. Bio. Rep., 37:1011-1017.

Kawabe, A and Miyashita N. T (2003). Patterns of codon usage bias in three dicot and four monocot plant species. Genes Genet Syst., 78(5):343-52.

Kozak, M (1989). Circumstances and mechanisms of inhibition of translation by secondary structure in eukaryotic mRNAs. Mol. Cell. Biol., 9:9134–9142.

Muto, A and S. Osawa (1987). The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA, 84:166–169.

Ningaraju, T. M. (2006). Cloning and characterization of chitinase gene/s from native isolates of *Serratiamarcescens*, M.Sc.(Agri.) Thesis, Univ. Agril. Sci. Dharwad, India.

Omumasaba, C. A., N. Yoshida and K. Ogawa (2001). Purification and characterization of a chitinase from *Trichodermaviride* .J. Gen. Appl. Microbiol., 47: 53–61

Perlak, F. J., R. L. Fuchs, D. A. Dean, S. L. Mcpherson and D. A. Fischhoff (1991). Modification of the coding sequence enhances plant expression of insect control protein genes. Proc. Natl. Acad. Sci. USA, 88:3324-3328

Sueoka, N (1962). On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA, 48:582–591.

Weng, L. X., H. H. Deng, J. L. Xu, Q. Li, Y. Q. Zhang, Z. D. Jiang, Q. W. Li, J. W. Chen and Zhang L H (2011). Transgenic sugarcane plants expressing high levels of modified *cry1Ac* provide effective control against stem borers in field trials. Transgenic Res., 20:759-772.

Xiu-ying, L., L. Zhi-hong, Z. Jie, H. Kang-lai, Z. Li and H. Da-fang (2013). Acquisition of Insect-Resistant Transgenic Maize Harboring a Truncated cry1Ah Gene via Agrobacterium-Mediated Transformation. J. Integr. Agri., 13(5):937-944.

Streatfield, S. J. (2007). Approaches to achieve high-level heterologous protein production in plants. Plant Biotech. J., 5(1): 2-15.

Menzella, H. G. (2011). Comparison of two codon optimization strategies to enhance recombinant protein production *in Escherichia coli*. Microb. Cell Fact., 10:15