Advances in Weed Science
formerly Planta Daninha

RESEARCH ARTICLE

Structural characterization of phosphinothricin *N*-acetyltransferase (*RePAT*) and the glufosinate-resistant analysis for site-directed mutagenesis of *RePAT* in *Arabidopsis thaliana*

Shixiao Xu[1], Lingxiao He[1], Jutao Sun[1], Zhiqiang Zhang[1], Tiezhao Yang[1], Gang Xue[1]*, Yongle Ding[1]

[1] Henan Agricultural University, Coll Tobacco Sci, Natl Tobacco Cultivat & Physiol & Biochem Res Ctr, Scientific Observation and Experiment Station of Henan, Ministry of Agriculture, Zhengzhou 450002, Henan, Peoples R China.

Abstract: Background: Transferring herbicide resistance genes to plant cultivars is one of the most effective methods for managing weed growth in agricultural land. The *RePAT* gene, previously isolated from the marine bacterium *Rhodococcus sp*. strain YM12, was found to provide glufosinate resistance to plants. **Objective:** In this study, we further investigated the protein structure and function of RePAT isolated from the marine. **Methods:** The physicochemical properties, two-dimensional structure, three-dimensional structure, and functional domains of the RePAT protein were analyzed and predicted using bioinformatics tools, and *RePAT* was optimized according to codon bias present in *Arabidopsis thaliana*. Site-directed mutagenesis of *RePAT* was performed, and the wild-type (*RePAT*) and mutant ($RePAT_{V120I}$) genes were successfully transferred into *A. thaliana*. **Results:** Our results showed that RePAT was an unstable hydrophilic protein, and six phosphorylation sites and two N-glycosylation sites were predicted. In addition, conserved domains containing the NAT_SF super family and coenzyme A-binding pocket were predicted in RePAT. Transgenic experiments and glufosinate resistance assays showed that the glufosinate resistance of *A. thaliana* containing the mutant gene ($RePAT_{V120I}$) was lower than that of plants containing the wild-type gene, indicating that the missense mutation in $RePAT_{V120I}$ had a significant effect on its glufosinate resistance. **Conclusions:** Our study provided improvement result for knowing the transferring herbicide resistance gene *RePAT*.

Keywords: *RePAT*; Site-directed mutagenesis; Glufosinate resistance; *Arabidopsis thalia*

Journal Information:
ISSN - 2675-9462
Website: http://awsjournal.org
Journal of the Brazilian Weed Science Society

How to cite: Xu S, He L, Sun J, Zhang Z, Yang T, Xue G, Ding Y. Structural characterization of phosphinothricin N-acetyltransferase (*RePAT*) and the glufosinate-resistant analysis for site-directed mutagenesis of *RePAT* in *Arabidopsis thaliana*. Adv Weed Sci. 2022;40:e020220108.
https://doi.org/10.51694/AdvWeedSci/2022;40:00010

Approved by:
Editor in Chief: Carlos Eduardo Schaedler
Associate Editor: José Barbosa dos Santos

Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this manuscript.

Received: October 27, 2021
Approved: June 21, 2022

* Corresponding author:
<xueganghnauedu@163.com>

1. Introduction

Weeds are one of the major pests that affect agricultural production (Moss, 2019, Rao et al., 2007). Chemical weeding is the most economical and effective weed management method at present, and can effectively reduce grain yield loss and labor cost (Antralina et al., 2015, Yadav et al., 2010). However, the majority of chemical herbicides are selective herbicides that control the growth of only certain weeds, and even selective herbicides may damage crops and affect crop yields. With the development of biotechnology, the emergence of genetically modified crops brings new hope for solving these issues. Research shows that transgenic herbicide-resistant crops can not only improve weed control efficiency, but also aid water and soil conservation and reduce greenhouse gas emissions by reducing soil tillage requirements (Green, 2012). Therefore, cultivating herbicide-resistant transgenic crops can solve many problems facing weeding of agricultural land.

Tolerance to the herbicide glufosinate is one of the most widely used herbicide resistance traits. Currently, eight transgenic crops resistant to glufosinate have been approved (Duke, 2015). All glufosinate-resistant crops are developed by expression of a gene encoding the enzyme phosphinothricin N-acetyltransferase (PAT), which can detoxify L-phosphinothricin (L-PPT) by acetylation of the amino group (Cui et al., 2016). Two commercially used glufosinate-resistant genes include *bar* and *pat*, which were isolated from *Streptomyces hygroscopicus* and *Streptomyces viridochromogenes* in 1987 and 1988, respectively (Cui et al., 2016, Thompson et al., 1987, Wohlleben et al., 1988). However, studies show that different rice varieties expressing *bar* show different levels of glufosinate resistance (Oard et al., 1996), which is also reported in transgenic barley expressing *bar* (Bregitzer et al., 2007). Therefore, the *pat* and *bar* genes have different kinetic parameters and activities against glufosinate in different plant cells. Wu et al. (2014) isolated a novel PAT-encoding gene (*RePAT*) from the marine bacterium *Rhodococcus sp*. strain YM12 in 2014 (Wu et al., 2014). This study additionally found that the RePAT protein shows a 37% identity and different kinetic constants compared to the proteins encoded by *bar*

and *pat*, and it has a high catalytic activity in converting L-PPT in vitro (Wu et al., 2014). In addition, another study shows that *RePAT* could be stably integrated and expressed in transgenic rice, and the resistance and agronomic traits of the transgenic plants were evaluated and proved that *RePAT* is an efficient gene (Cui et al., 2016).

Although the *RePAT* gene has been comprehensively described and successfully expressed in rice plants, and its resistance to glufosinate has been demonstrated, the structure of the *RePAT* protein and the loci of resistance in this protein are still poorly understood. In this study, we performed systematic analysis of the *RePAT* protein structure and predicted the functional sites using bioinformatics tools to study the characteristics of *RePAT*. In addition, to determine whether a specific amino acid residue can affect the properties of glufosinate-resistance conferred by *RePAT*, suspected amino acid sites of *RePAT* were subjected to mutation. Moreover, we constructed plant expression vectors containing wild-type (*RePAT*) and mutated ($RePAT_{V120I}$) genes and transferred by using agrobacterium-mediated transformation to successfully transfer the vector into *Arabidopsis thaliana* plants. Our results provide a theoretical basis for further study of the applications of *RePAT* in plant resistance to glufosinate.

2. Material and Methods

2.1 Plant materials

A. thaliana (Arabidopsis ecotype Col-0 plants) seeds were collected from Huazhong Agricultural University (Wuhan, China). Seeds were sown in plastic pots containing a 1:1 (v:v) mixture of peat and loam. The plastic pots were placed at 22°C in a greenhouse and exposed to a light intensity of 100-150 μmol/m2/s for 18 h per day. Daily watering was performed to maintain good plant growth. Plants were dipped after 5-6 weeks when fruits grew out from the biggest inflorescences.

2.2 Bioinformatics analysis of RePAT

The basic physical and chemical properties of *RePAT* were analyzed using the ProtParam tool (http://web.expasy.org/protparam/); the hydrophilic/hydrophobic properties using the ProtScale tool (http://web.expasy.org/protscale/); potential phosphorylation sites using the NetPhos tool (http://www.cbs.dtu.dk/services/NetPhos/); potential N-glycosylation sites using the NetNGlyc tool (http://www.cbs.dtu.dk/services/NetNGlyc/); and potential O-glycosylation sites using the NetOGlyc tool (http://www.cbs.dtu.dk/services/NetOGlyc/).

2.3 Codon optimization and RePAT gene synthesis

The wild-type *RePAT* gene was isolated from the marine bacterium *Rhodococcus sp.* strain YM12 (Wu et al., 2014). The nucleotide and amino acid sequences of *RePAT* were obtained from the GeneBank database (accession number: JQ398613). Owing to the codon bias present in *Arabidopsis*, the sequence of *RePAT* was optimized using DNAWorks 2.4 (NIH, Bethesda, Maryland, USA) (Hoover, Lubkowski, 2002). The protein sequence was input into DNAWorks, and then the optimization algorithm was executed. The initial and final synthetic gene sequences and a list of optimized oligonucleotide sequences, along with the scores for each section from both the initial and final sequence, were generated (Hoover, Lubkowski, 2002).

The assembly of the synthetic gene from oligonucleotides was performed according to a previously described protocol (Stemmer et al., 1995). First, equal volumes of the oligonucleotide solutions were mixed together and diluted with water, and then the oligonucleotide mixture was diluted 5-fold with the polymerase chain reaction (PCR) solution. For gene amplification, the mixture resulting from the gene assembly reaction was used as the template, with the outermost oligonucleotide sequences used as primers (Hoover, 2012).

2.4 Construction of the plant expression vector and site-directed mutagenesis

The synthesized *RePAT* gene was inserted into a pCAMBIA2301 vector containing restriction sites for Ncol and Pmll, and the recombinant plasmid was named pCAMBIA2301-RePAT (35S:RePAT:Nos polyA). Then, PCR site-directed mutagenesis was performed. The primers of the mutant gene were designed using the template of the recombinant plasmid pGEX-RePAT. The pairs of the complementary oligonucleotides for the desired mutant primers were F-RePAT-V120I: 5'-GTTATAGTGGCATCTATAGAGTCTACTAACGCTAC-3' and R-RePAT-V120I: 5'-TAGATGCCACTATAACGTGGATTC CTC-3'. The site-directed mutagenesis was performed according to a previously described protocol: the Easy Mutagenesis system (TransGen Biotech, Beijing, China) was used, and the PCR reaction was carried out with the following parameters: an initial denaturation for 2 min at 97°C, 20 cycles of 20 s at 95°C, 20 s at 56°C, and 3 min 30 s at 72°C, and a final incubation for 5 min at 72°C. The plasmids bearing the desired mutations were transformed into Escherichia. Coli BL21-CodonPlus (DE3)-RIL cells for protein expression and purification (Wu et al., 2014).

2.5 Agrobacterium-mediated transformation

The final two plant expression vectors, pCAMBIA2301-RePAT (35S:RePAT:Nos polyA) and pCAMBIA2301-RePATV120I (35S:RePATV120I:Nos polyA), were introduced into Agrobacterium tumefaciens 1301 by using electroporation. The supernatant containing the bacteria was collected after centrifugation at 4,000 RPM, and then was re-suspended in half-strength Murashige and Skoog

(MS) liquid medium (containing 5% sucrose). The OD600 of the medium was adjusted to 0.6, to which 2-(N-morpholino) ethanesulfonic acid (0.05%) and a surfactant, Silwet L-77 (0.02%), were added and mixed thoroughly. Following this, the inflorescence of A. thaliana was infected and transformed using the floral dip method, and the seeds of T1 generation plant were harvested.

2.6 Glufosinate resistance assay in MS medium

The seeds of two transgenic *Arabidopsis* plants (35S:RePAT:Nos polyA and 35S:RePATV120I:Nos polyA) were dulled, sterilized, and transferred into the half-strength MS medium containing different concentrations of glufosinate (1, 5, 10, 20 mg/L). They were cultured at 25°C in 16 h light/8 h dark cycles, and the growth was observed.

3. Results

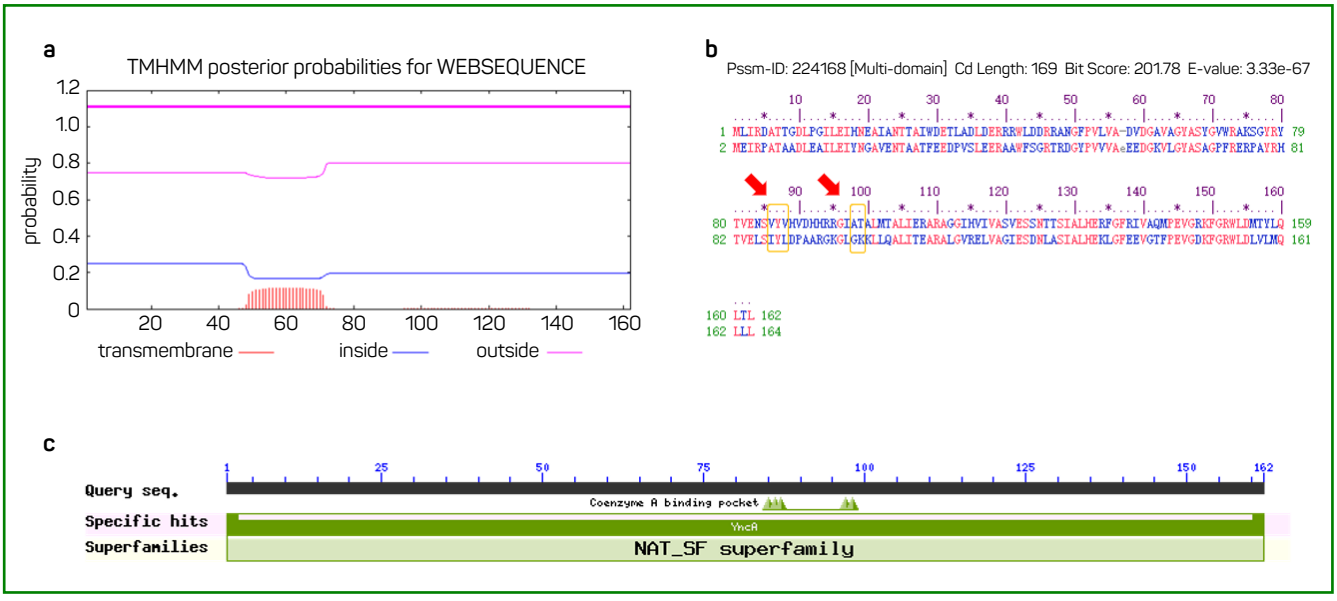
3.1 Characterization of RePAT protein structure and functional sites

The bioinformatics analysis showed that *RePAT* encodes 162 amino acids, and the estimated half-life is 30 hours in mammalian reticulocytes in vitro (>20 hours in yeast, in vivo; >10 hours in *E.coli*, in vivo). The aliphatic index of RePAT was found to be 96.98, and the instability index (II) was calculated as 40.23 (>40 indicates as unstable protein). The total average of hydropathicity (GRAVY) was -0.089, and RePAT was found to be a hydrophilic protein. Six phosphorylation sites containing two specific protein kinase-binding (four unsp and two CK1) sites were found in the region. Two potential N-glycosylation sites were found at 24 and 124 amino acid loci. There were no potential O-glycosylation sites found in the protein. Moreover, the transmembrane region could not be located by using the TMHMM software (Figure 1a). According to the result of conserved domain analysis by NCBI, *RePAT* contained a conserved domain similar to that of NAT_SF super family proteins (*YncA*: accession number, COG1247) (Figure 1b). In addition, coenzyme A-binding pockets were found at 86-88 and 98-99 amino acid loci of *RePAT* (Figure 1c).

3.2 Structural models of the RePAT protein

The two-dimensional (2D) structure of RePAT is shown in Figure 2. α-Helixes, α-folds, and irregular curls were identified in RePAT. The 2D structure formed by 162 amino acid residues comprised five α-helixes (amino acid loci: 11-24, 36-45, 96-110, 125-133, 159-160; accounting for 30.86% of all the amino acid resides), eight b-folds (amino acid loci: 3-4, 52-58, 62-70, 80-87, 114-121, 137-141, 146-149, 151-158; accounting for 31.48% of all the amino acid resides). The remaining 62 amino acid residues formed an irregular curl, and these amino acid residues account for 38.28% of the total amino acid residue content. The three-dimensional (3D) structure of RePAT was predicted using the SWISS-MODEL online software (Figure 3). Figure 3 shows that the 3D structure of the protein mainly comprised α-helixes, b -folds, and irregular curls. The predicted results of the 3D structure and secondary structure were similar.

A previous study showed that RePAT shared 37% identity with the amino acids of PAT and BAR and shared

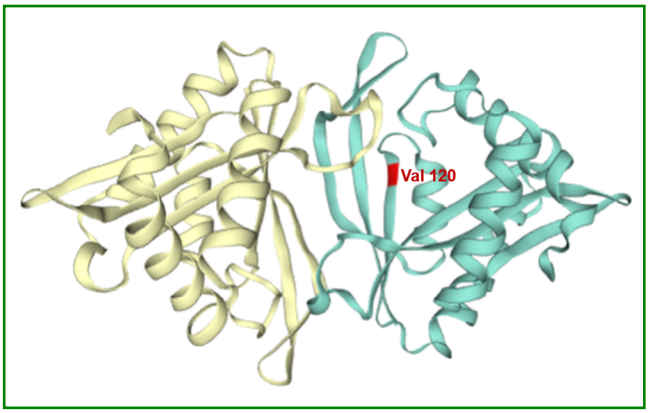


a. The transmembrane structure of the RePAT protein. b. Prediction of the functional domain. The red arrow indicates the amino acid site of the coenzyme A-binding pocket. c. Prediction of the conserved domain

Figure 1 - Structural characterization of the RePAT protein

Figure 2 - Secondary structure prediction of RePAT using PSIPRED analysis

33%, 59%, and 58% identity with ScPAT, MAT, and Pita, respectively (Wu et al., 2014). Based on the results of BLASTP amino acid alignments, we found that the amino acid present at site 120 in RePAT is valine; however, isoleucine was found at the same site in PAT, Mat, ScPAT, BAR, and Pita. We then speculated that the amino acid present at site 120 is important to the function of RePAT (Figure 3).



The Val120 is the mutant site

Figure 3 - The 3D-structure model of RePAT protein

3.3 Codon optimization and synthesis of RePAT gene

The codons of microorganisms and plants are not completely identical; the codons of the microbe-derived *RePAT* gene were optimized in this study to improve the expression efficiency in plant cells. In comparison, the optimized *RePAT* gene sequence changed 69 bases, which accounts for 14.11% of the total bases (Figure 4). After optimizing the *RePAT* sequence according to the codon bias present in *Arabidopsis*, we obtained 24 oligonucleotides using DNAWorks, and then synthesized these oligonucleotides (Table 1). Finally, the sequence of *RePAT* was synthesized using mixed amplification with the outermost oligonucleotides used as primers.

```
origin          ATGCTGATCCGCGACGCCACCACAGGTGACCTGCCCGGGATCCTCGAGATCCACAACGAG
optimized       ATGCTGATTAGAGATGCTGTTCCTGGAGACCTCCCTGGTATTTTAGAAATTCACAACGAA
                ******** .*.** ** .  .*:**:***** ** ** **  *.**.** ********.

origin          GCGATCGCCAACACGACGGCGATCTGGGACGAGACACTCGCCGATCTGGACGAGCGGCGC
optimized       GCTATTGCTAATTCAACTGCTATTTGGGATGAAACTCCTGCAGATCTTGATGAAAGACGA
                ** ** ** ** :*.** ** ** ***** **.**:*  **.***** ** **..*.**.

origin          CGTTGGCTCGACGATCGTCGGGCCAACGGCTTCCCGGTGCTCGTCGCCGACGTCGATGGA
optimized       AGATGGTTTGATGACAGACGTGCTAATGGGTTTCCTGTTCTTGTAGCTGATGTTGATGGA
                .*:*** * ** ** .*:** ** ** ** ** ** ** ** **.** ** ** ******

origin          GCAGTCGCGGGGTACGCCTCGTACGGGGTGTGGCGGGCCAAGAGCGGCTACCGATACACC
optimized       GTTGTGGCTGGATATGCATCTTATGGGGTATGGAGAGCTAAGTCTTCTTATCGACACACT
                * :** ** **.** **.** ** *****.***.*.** ***:     ** *** ****

origin          GTCGAGAACTCGGTGTACGTCCACGTCGATCACCATCGGCGGGGCATCGCGACCGCGCTC
optimized       GTCGAGAACTCGGTGTACGTTCATGTAGACCATCATAGAAGAGGAATTGCCACTGCTCTT
                ********************* ** **.** ** ***.*..*.**.** ** ** ** **

origin          ATGACCGCGTTGATCGAGCGGGCCCGCGCCGGCGGGATCCACGTGATCGTCGCCAGCGTC
optimized       ATGACTGAACTTATTGAGCGAGCTAGAGCTGGAGGAATCCACGTTATAGTGGCATCTGTA
                ***** *..  * ** *****.** .*.** **.**.******** **.** **.:  **.

origin          GAATCGTCCAACACGACGTCCATCGCCCTGCACGAACGGTTCGGTTTCCGCATCGTGGCG
optimized       GAGTCTACTAACGCTACTTCAGTAGCTCTTCACGAAAGGTTTGGATTTCGTATTGTTGCT
                **.** :* ***.* ** **.*.** ** ******.**** **:** ** ** ** **

origin          CAGATGCCGGAGGTGGGACGCAAGTTCGGCCGCTGGCTGGACATGACGTACCTGCAGCTG
optimized       CACATGCCTGAAGTTGGAAGAAAGTTCGGACGTTGGCTTGATATGACATACTTACAACTG
                ** ***** **.** ***.*.********.** ***** ** *****.*** *.**.***

origin          ACCCTCTAG
optimized       ACATTG---
                **. *
```

Figure 4 - Alignment of the codon sequences before and after optimization for *RePAT*

3.4 Transformation and glufosinate resistance assay results

Recombinant expression vectors containing the genes *RePAT* and *RePAT*$_{V120I}$ were constructed using the plasmid pCAMBIA2301 (Figure 5a). The pCAMBIA2301-RePAT and pCAMBIA2301-RePATV120I plasmids were then transferred into *A. thaliana* using *Agrobacterium*-mediated transformation. Two T0 transgenic *Arabidopsis* plants were obtained.

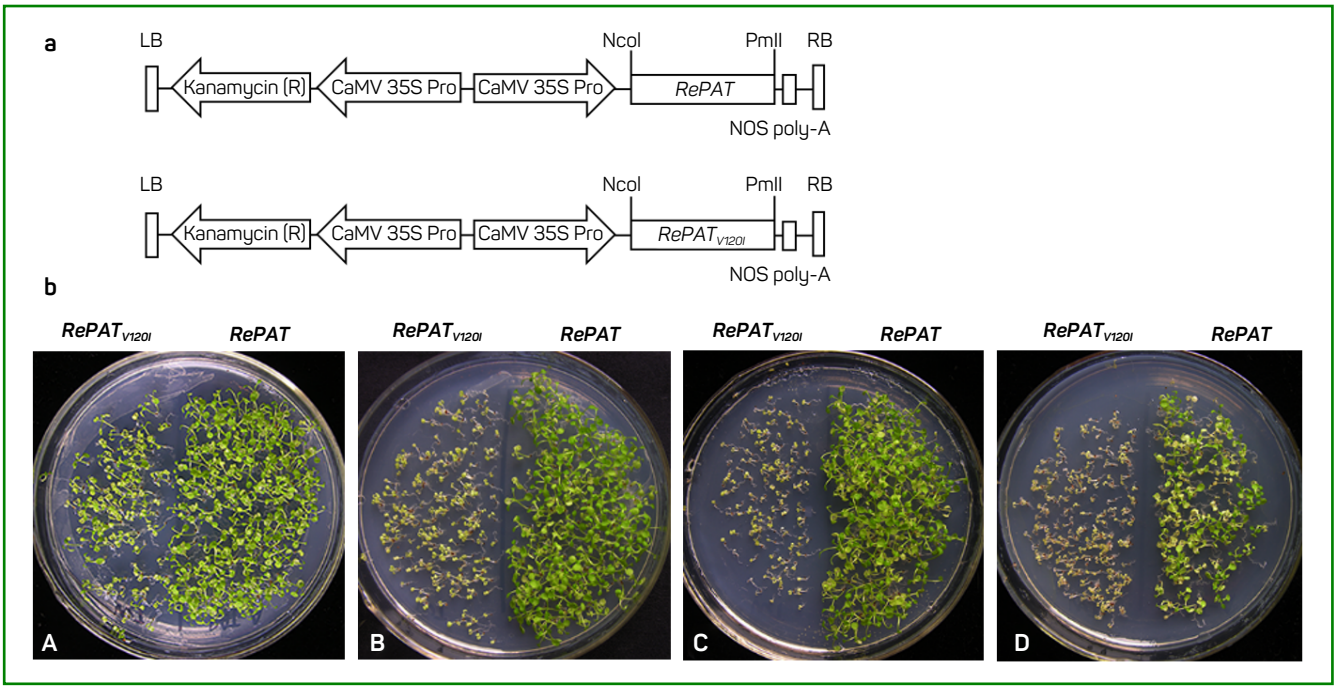
The T1 *Arabidopsis* transgenic seeds were sterilized and planted on the half-strength MS medium containing 50 mg/L kanamycin to screen for the transformed plant. After obtaining the T1-positive plants, the T2 *Arabidopsis* transgenic seeds were harvested using T1 *Arabidopsis* selfing. The T2 *Arabidopsis* transgenic seeds were then planted on the half-strength MS medium containing 50 mg/L kanamycin to select homozygous transgenic plants. The T3 *Arabidopsis* transgenic plant was obtained by T2 homozygous plant selfing. The T3 *Arabidopsis* transgenic seeds were planted on the MS media containing 1, 5, 10, and 20 mg/L glufosinate-ammonium, respectively. After 7 days, the transgenic *Arabidopsis* plant (*RePAT*) showed normal growth on the medium containing 1, 5, and 10 mg/L glufosinate-ammonium, but was inhibited on the medium containing 20 mg/L glufosinate-ammonium (Figure 5b). In addition, the resistance of the transgenic *Arabidopsis* plant (*RePAT*$_{V120I}$) was decreased as the concentration of glufosinate-ammonium increased. Compared with the *Arabidopsis* plant containing *RePAT*$_{V120I}$, the *Arabidopsis* plant containing *RePAT* showed a higher resistance to glufosinate (Figure 5b).

4. Discussion

To improve the glufosinate-resistance in plants, the *RePAT* gene obtained from the marine bacterium *Rhodococcus sp*. strain YM12 is introduced in plant cells and is used in rice crops (Cui et al., 2016). In a previously study, Cui et al. (2016) found that more than 70% of the independent T0 transgenic plants containing *RePAT* grew normally without chlorosis when they were treated with a high concentration of glufosinate (Cui et al., 2016). They also verified that transgenic rice plants with a lower *RePAT* expression would die or show severe chlorosis when treated with glufosinate (Cui et al., 2016). In addition, agronomic

Table 1 - The 24 oligonucleotides for synthesized *RePAT*

Oligonucleotide No.	Sequences (5′-3′)
RePAT-1	5′-ATGCTGATTAGAGATGCTGTTCCTGGAG-3′
RePAT-2	5′-ATTTCTAAAATACCAGGGAGGTCTCCAGGAACAGCATCTC-3′
RePAT-3	5′-CCTCCCTGGTATTTTAGAAATTCACAACGAAGCTATTGCT-3′
RePAT-4	5′-ATCCCAAATAGCAGTTGAATTAGCAATAGCTTCGTTGTGA-3′
RePAT-5	5′- AATTCAACTGCTATTTGGGATGAAACTCCTGCAGATCTTG-3′
RePAT-6	5′-CAAACCATCTTCGTCTTTCATCAAGATCTGCAGGAGTTTC-3′
RePAT-7	5′-ATGAAAGACGAAGATGGTTTGATGACAGACGTGCTAATGG-3′
RePAT-8	5′-ATCAGCTACAAGAACAGGAAACCCATTAGCACGTCTGTCA-3′
RePAT-9	5′-TTCCTGTTCTTGTAGCTGATGTTGATGGAGTTGTGGCTGG-3′
RePAT-10	5′-TCCATACCCCATAAGATGCATATCCAGCCACAACTCCATC-3′
RePAT-11	5′-GCATCTTATGGGGTATGGAGAGCTAAGTCTTCTTATCGAC-3′
RePAT-12	5′-ACCGAGTTCTCGACAGTGTGTCGATAAGAAGACTTAGCTC-3′
RePAT-13	5′-ACTGTCGAGAACTCGGTGTACGTTCATGTAGACCATCATA-3′
RePAT-14	5′-GCAGTGGCAATTCCTCTTCTATGATGGTCTACATGAACGT-3′
RePAT-15	5′-AGAGGAATTGCCACTGCTCTTATGACTGAACTTATTGAGC-3′
RePAT-16	5′-TTCCTCCAGCTCTAGCTCGCTCAATAAGTTCAGTCATAAG-3′
RePAT-17	5′-GAGCTAGAGCTGGAGGAATCCACGTTATAGTGGCATCTGT-3′
RePAT-18	5′-GAAGTAGCGTTAGTAGACTCTACAGATGCCACTATAACGT-3′
RePAT-19	5′-AGAGTCTACTAACGCTACTTCAGTAGCTCTTCACGAAAGG-3′
RePAT-20	5′-AGCAACAATACGAAATCCAAACCTTTCGTGAAGAGCTACT-3′
RePAT-21	5′-TTTGGATTTCGTATTGTTGCTCACATGCCTGAAGTTGGAA-3′
RePAT-22	5′-TCAAGCCAACGTCCGAACTTTCTTCCAACTTCAGGCATGT-3′
RePAT-23	5′-TCGGACGTTGGCTTGATATGACATACTTACAACTGACATT-3′
RePAT-24	5′-CTACAATGTCAGTTGTAAGTATGTCA-3′



a. The plant expression vector. *RePAT* and *RePAT*$_{V120I}$ transcriptions were initiated by the CaMV 35S Pro promoter and terminated by NOS polyA. Kanamycin was used for screening plasmid-positive T1 plants. NcoI and PmlI were the two restriction enzyme digestion sites used to construct the vector. LB, left border; RB, right border; b. Glufosinate resistance assays of transgenic *Arabidopsis* plants. A. Germination of the *Arabidopsis* group containing *RePAT*$_{V120I}$ was slightly inhibited on half-strength MS medium containing 1 mg/L glufosinate, whereas that of the *Arabidopsis* group containing *RePAT* was normal. B. Germination of the *Arabidopsis* group containing *RePAT*$_{V120I}$ was moderately inhibited on half-strength MS medium containing 5 mg/L glufosinate, whereas that of the *Arabidopsis* group containing *RePAT* was normal. C. Germination of the *Arabidopsis* group containing *RePAT*$_{V120I}$ was completely inhibited on half-strength MS medium containing 10 mg/L glufosinate, whereas that of the *Arabidopsis* group containing *RePAT* was normal. D. Germination of the *Arabidopsis* group containing *RePAT*$_{V120I}$ was completely inhibited on half-strength MS medium containing 20 mg/L glufosinate, whereas that of the *Arabidopsis* group containing *RePAT* was certainly inhibited. NOS, nopaline synthase; MS, Murashige and Skoog

Figure 5 - Plant expression vector and the results of glufosinate resistance assay

analyses showed that although *RePAT* provides high resistance to glufosinate, it does not affect the agronomic characteristics of transgenic rice plants (Cui et al., 2016). These results suggest that the novel glufosinate-resistance gene *RePAT* is an ideal substitute for the traditional herbicide-resistance genes bar and pat. In this study, based on known information on *RePAT*, the 2D and 3D structures of the RePAT protein were analyzed using bioinformatics tools to further understand its structural and functional characteristics. Additionally, we performed a site-specific mutation in *RePAT* at the site of amino acid 120, which was combined with the alignment result of a previous study to explore the amino acid sites affecting glufosinate resistance in RePAT. We additionally obtained meaningful results on glufosinate resistance of the mutated *RePAT* gene. It is worth mentioning that we synthesized the RePAT gene using primers designed privately, established a plant expression vector using the optimized codons, and successfully transferred the vectors into *Arabidopsis* using Agrobacterium. Our study provides a theoretical basis for further studies of the biological functions of the *RePAT* gene at the structural and functional levels of RePAT protein and gene mutagenesis.

Bioinformatics is an emerging discipline combining molecular biology and informatics (Akalin, 2006, Bartlett et al., 2017). It provides new insights for future experiments to understand the structural and functional characteristics of biological macromolecules using bioinformatics data (Bartlett et al., 2017, Kim, 2014). In this study, the structural and functional sites of RePAT were predicted using bioinformatics tools for the first time, providing some basic theoretical materials for further studies of the function of *RePAT* isolated from the marine bacterium *Rhodococcus sp*. Regarding post-translational modification of proteins, phosphorylation is one of the most important covalent modifications in the process of translation, which is closely related to gene expression, signal transduction, cell division, the cell growth cycle, and cell growth and development (Uhrig et al., 2019, Vu et al., 2018). Moreover, protein glycosylation also plays an important role in plant growth and development, hormonal network regulation, signal transduction, and plant virus infection (Strasser, 2016, Veit et al., 2015). Our study showed that the RePAT protein had six phosphorylation sites containing two specific protein kinase-binding sites (four unsp and two CK1) and two potential *N*-glycosylation sites. The expression level

of *RePAT* may be affected when these sites are modified. These observations indicate that future studies on *RePAT* should focus on the glycosylation and phosphorylation sites predicted in this study. In addition, we predicted that the RePAT protein has a conserved domain similar to those of NAT_SF superfamily proteins. Additionally, a coenzyme A-binding pocket was identified in the protein.

There are 61 codons encoding 20 basic amino acids of natural proteins; each amino acid can be encoded by 1 to 6 different codons (Sanhong et al., 2003). Codon bias or preference is important for finding the optimal heterogenous expression system or host of a gene, and therefore improving the expression level of the gene (Paul et al., 2018). Our study analyzed the codon preference of *RePAT* and optimized the codons such that it can be normally expressed in *A. thaliana*. These results provide a reference for future studies that involve the transformation of *Arabidopsis* with *RePAT*.

Site-directed mutagenesis is a PCR-based technique used to mutate specified nucleotides in a sequence within a plasmid vector. This technique allows the study of the relative importance of a particular amino acid in protein structure and function (Bachman, 2013). To investigate the amino acid that affects protein function in glufosinate resistance, the Val120 amino acid of RePAT was mutated, and the glufosinate resistance assay showed that the resistance of *Arabidopsis* containing mutated *RePAT* ($RePAT_{V120I}$) was lower than that of the *Arabidopsis* containing wild-type *RePAT*. This result reveals that the Val120 amino acid is important for the glufosinate resistance provided by *RePAT*.

5. Conclusions

This research further investigated the protein structure and function of RePAT isolated from the marine bacterium *Rhodococcus sp.* strain YM12. The *RePAT* gene encoded 162 amino acids and contained six phosphorylation sites and two potential *N*-glycosylation sites. The 2D structure was mainly composed of α-helixes and irregular curls. Moreover, the effect of specific amino acid residues on glufosinate resistance provided by *RePAT* was evaluated using site-directed mutagenesis. We demonstrated that the amino acid present at site 120 (Val) is important for the glufosinate resistance provided by *RePAT*. These mutants should be helpful in further applications of *RePAT* introduced in transgenic plants to improve glufosinate resistance. This study further analyzed the characteristics of the RePAT protein at the levels of protein structure and function, codon preference, and gene mutation to provide a theoretical basis for further research on the biological functions of the *RePAT* gene.

Author's contributions

All authors read and agreed to the published version of the manuscript. SX, GX, and YD: conceptualization of the manuscript and development of the methodology. LH, JS, and ZZ: data collection and curation. JS, ZZ, and TY: data analysis, data interpretation. GX: funding acquisition and resources, project administration, supervision. SX: writing the original draft of the manuscript. SX, LH, JS, ZZ, TY, GX, and YD: writing, review and editing.

Funding

This research was funded by the Open Project Program of Key Laboratory of Tobacco Biology & Processing (201803), China National Tobacco Corporation Henan company (2018410000270035), Guizhou Tobacco Corporation Guiyang company (2016-07), China National Tobacco Corporation (110201002008, 110201402004).

Acknowledgements

We thank the four fund projects (the Open Project Program of Key Laboratory of Tobacco Biology & Processing (201803), China National Tobacco Corporation Henan company (2018410000270035), Guizhou Tobacco Corporation Guiyang company (2016-07), China National Tobacco Corporation (110201002008, 110201402004) for supporting this research. We also appreciate the anonymous reviewers for their constructive comments.

References

Akalin PK. Introduction to bioinformatics. Mol Nutr Food Res. 2006;50(7):610-9. Available from: https://doi.org/10.1002/mnfr.200500273

Antralina M, Istina IN, YuyunYuwariah, Simarmata T. Effect of difference weed control methods to yield of lowland rice in the SOBARI. Proc Food Sci. 2015;3:323-9. Available from: https://doi.org/10.1016/j.profoo.2015.01.035

Bachman J. Site-directed mutagenesis. Met Enzymol. 2013;529:241-8. Available from: https://doi.org/10.1016/b978-0-12-418687-3.00019-7

Bartlett A, Penders B, Lewis J. Bioinformatics: indispensable, yet hidden in plain sight? BMC Bioinf. 2017:18(1):1-7. Available from: https://doi.org/10.1186/s12859-017-1730-9

Bregitzer P, Cooper LD, Hayes PM, Lemaux PG, Singh J, Sturbaum AK. Viability and bar expression are negatively correlated in Oregon Wolfe Barley dominant hybrids. Plant Biotech J. 2007;5(3):381-8. Available from: https://doi.org/10.1111/j.1467-7652.2007.00247.x

Cui Y, Liu Z, Li Y, Zhou F, Chen H, Lin Y. Application of a novel phosphinothricin N-acetyltransferase (RePAT) gene in developing glufosinate-resistant rice. Scientific Rep. 2016;6(1):1-11. Available from: https://doi.org/10.1038/srep21259

Duke SO. Perspectives on transgenic, herbicide-resistant crops in the United States almost 20 years after introduction. Pest Manag Sci. 2015;71(5):652-7. Available from: https://doi.org/10.1002/ps.3863

Green JM. The benefits of herbicide-resistant crops. Pest Manag Sci. 2012;68(10):1323-31. Available from: https://doi.org/10.1002/ps.3374

Hoover D. Using DNAWorks in designing oligonucleotides for PCR-based gene synthesis. Gene Synth. 2012;852:215-23. Available from: https://doi.org/10.1007/978-1-61779-564-0_16

Hoover DM, Lubkowski J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucl Acids Res. 2002;30(10):1-7. Available from: https://doi.org/10.1093/nar/30.10.e43

Kim S. Recent development in bioinformatics for utilizing omics data. Methods. 2014;69(3):205-6. Available from: https://doi.org/10.1016/j.ymeth.2014.09.008

Moss S. Integrated weed management (IWM): why are farmers reluctant to adopt non-chemical alternatives to herbicides? Pest Manag Sci. 2019;75(5):1205-11. Available from: https://doi.org/10.1002/ps.5267

Oard JH, Linscombe SD, Braverman MP, Jodari F, Blouin DC, Leech M et al. Development, field evaluation, and agronomic performance of transgenic herbicide resistant rice. Mol Breed. 1996;2:359-68. Available from: https://doi.org/10.1007/BF00437914

Paul P, Malakar AK, Chakraborty S. Codon usage and amino acid usage influence genes expression level. Genetica. 2018;146(1):53-3. Available from: https://doi.org/10.1007/s10709-017-9996-4

Rao AN, Johnson DE, Sivaprasad B, Ladha JK, Mortimer AM. Weed management in direct-seeded rice. Adv Agron. 2007;93:153-255. Available from: https://doi.org/10.1016/s0065-2113(06)93004-1

Sanhong F, Aiguang G, Liwei S, Xiaoping H. Analysis of genetic code prederence in *Arabidopsis thaliana*. Prog Biochem Biophys. 2003;30(2):221-5. Chinese.

Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene. 1995;164(1):49-53. Available from: https://doi.org/10.1016/0378-1119(95)00511-4

Strasser R. Plant protein glycosylation. Glycobiology. 2016;26(9):926-39. Available from: https://doi.org/10.1093/glycob/cww023

Thompson CJ, Movva NR, Tizard R, Crameri R, Davies JE, Lauwereys M et al. Characterization of the herbicide-resistance gene bar from *Streptomyces hygroscopicus*. EMBO J. 1987;6(9):2519-23. Available from: https://doi.org/10.1002/j.1460-2075.1987.tb02538.x

Uhrig RG, Schläpfer P, Roschitzki B, Hirsch-Hoffmann M, Gruissem W. Diurnal changes in concerted plant protein phosphorylation and acetylation in Arabidopsis organs and seedlings. Plant J Cell Mol Biol. 2019;99(1):176-94. Available from: https://doi.org/10.1111/tpj.14315

Veit C, Vavra U, Strasser R. N-glycosylation and plant cell growth. Met Mol Biol. 2015;1242:183-94. Available from: https://doi.org/10.1007/978-1-4939-1902-4_16

Vu LD, Gevaert K, Smet I. Protein language: post-translational modifications talking to each other. Trends Plant Sci. 2018;23(12):1068-80. Available from: https://doi.org/10.1016/j.tplants.2018.09.004

Wohlleben W, Arnold W, Broer I, Hillemann D, Strauch E, Pühler A. Nucleotide sequence of the phosphinothricin N-acetyltransferase gene from *Streptomyces viridochromogenes* Tü494 and its expression in *Nicotiana tabacum*. Gene. 1988;70(1):25-37. Available from: https://doi.org/10.1016/0378-1119(88)90101-1

Wu G, Yuan M, Wei L, Zhang Y, Lin Y, Zhang L et al. Characterization of a novel cold-adapted phosphinothricin N-acetyltransferase from the marine bacterium *Rhodococcus* sp. strain YM12. J Mol Catal B Enz. 2014;104:23-8. Available from: https://doi.org/10.1016/j.molcatb.2014.03.001

Yadav V, Singh L, Singh R. Effect of crop establishment methods and weed management practices on nutrient uptake, yield and quality of rice (*Oryza sativa* L.). Pakistan J Weed Sci Res. 2010;16(4):379-85.