



## Original article

Molecular cloning and in-depth bioinformatics analysis of type II ribosome-inactivating protein isolated from *Sambucus ebulus*Masoumeh Rezaei-Moshaei<sup>a</sup>, Ali Bandehagh<sup>a,\*</sup>, Ali Dehestani<sup>b</sup>, Ali Pakdin-Parizi<sup>b,\*</sup>, Majid Golkar<sup>c</sup><sup>a</sup> Department of Plant Breeding and Biotechnology, University of Tabriz, Tabriz, Iran<sup>b</sup> Genetics and Agricultural Biotechnology Institute of Tabarestan, Sari Agricultural Sciences and Natural Resources University, Sari, Iran<sup>c</sup> Pasteur Institute of Iran, Tehran, Iran

## ARTICLE INFO

## Article history:

Received 10 October 2019

Revised 29 January 2020

Accepted 9 February 2020

Available online 17 February 2020

## Keywords:

Physicochemical properties

Protein modeling

Recombinant RIP

## ABSTRACT

Plant ribosome-inactivating proteins (RIPs) are N-glycosidases which inhibit protein synthesis through depurination of the ribosomal RNA sequence. Type II RIPs are heterodimer proteins which can bind to cell surfaces. The cytotoxicity of these RIPs is different. *Sambucus* spp. are a rich source of RIP proteins with different properties. In the present study, a type II RIP was isolated from *S. ebulus* plant that grows widely in the north of Iran, and different bioinformatics tools were used for the evaluation of physicochemical, functional and 3D protein characteristics. The results showed significant differences among isolated RIP and other *Sambucus* RIP proteins. The study of these differences can not only expand our insight into the functioning mechanisms of plant RIPs but also provide information about a novel RIP protein with potential biological applications.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Ribosome-inactivating proteins (RIPs) are toxins that are extensively distributed in nature but are mainly found in plants, bacteria, fungi, and some algae. These proteins specifically and irreversibly inhibit protein translation through binding to the large ribosomal subunit and cleaving the specific adenine base of 28S rRNA by N-glycosidase activity (Barbieri et al., 1993; Girbes et al., 2004; Stirpe, 2004). However, the biological role of RIPs in plants is unclear (Lee-Huang et al., 1990). Existing evidence suggests that they might be involved in protecting plants against insects (Arias et al., 1992), fungi (Munoz et al., 1991) and viruses (Merino et al., 1990).

RIP proteins have been classified into three structural classes. Type I RIPs, such as saporins, dianthins, and beetins, are a single

polypeptide chain (approximately 30 kDa) that strongly displays the N-glycosidase activity. Type II RIPs such as ricin and abrin are heterodimeric proteins (60–65 kDa) with an enzymatic A chain similar to type I RIPs linked by a disulfide bond to a slightly larger B chain with lectin properties (Lord et al., 1994). Transfer of functional A chain to host cells is more convenient than other RIP types due to B chain binding ability to sugar units of cell surface receptors (Mathews et al., 2007). Type III RIPs are composed of a polypeptide chain containing a type I RIP in the N-terminal and a C-terminal domain with unknown function (Jimenez et al., 2014). Type III RIPs are also synthesized as inactive precursors that are converted into an active form by proteolytic processing events (Peumans et al., 2001).

*Sambucus ebulus* (dwarf elder) is an annual plant of the Adoxaceae family. This family of flowering plants includes about 190 species and four genera and is often dispersed in southern and central Europe, Northwest Africa, and Southwest Asia (Westwood, 1985). The *S. ebulus* extracts show high antioxidant, anti-inflammatory, anti-arthritis, anti-nociceptive, antimicrobial, and anticancer activities (Schwaiger et al., 2011; Tasinov et al., 2012; Tasinov et al., 2013). It has been suggested that part of the biological activities of plant extracts, e.g., anticancer properties, are possibly connected to RIP proteins (Benitez et al., 2005; Shokrzadeh et al., 2009). *Sambucus* species have a complex mixture of different types of RIPs and related lectins. The heterodimeric type II RIPs are present in the leaves (Girbes et al., 1993), rhizome (Citores et al., 1997) and

\* Corresponding authors at: Genetics and Agricultural Biotechnology Institute of Tabarestan, Sari Agricultural Sciences and Natural Resources University, Sari, Iran (A. Pakdin-Parizi); Department of Plant Breeding and Biotechnology, University of Tabriz, Tabriz, Iran (A. Bandehagh).

E-mail addresses: [bandehhagh@tabrizu.ac.ir](mailto:bandehhagh@tabrizu.ac.ir) (A. Bandehagh), [a.pakdin@sanru.ac.ir](mailto:a.pakdin@sanru.ac.ir) (A. Pakdin-Parizi).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

fruits (Nikolov, 2007) of *S. ebulus*. Ebulin I, the first RIP isolated from the dwarf elder, is non-toxic as compared with the highly toxic ricin (Stripe et al., 1992; Girbes et al., 1993). The structure of ebulin I has been characterized by X-ray diffraction analysis and it is very similar to type I RIPs and other type II RIPs such as ricin (Pascal et al., 2001).

Due to the diverse biological activities of RIPs, extensive research has been done to investigate their use as antiviral and antitumor agents (Zhu et al., 2016; Sipahioglu et al., 2017). The use of RIPs as part of the conjugates are the most promising applications of RIPs in medicine, especially in cancer therapy. In these conjugates, the enzymatic RIPs are attached to tumor-targeting ligands or antibodies that cause their binding and entry into malignant cells. Since the drugs of plant source are more efficient and more cost-effective than the synthetic drugs, new drug substitutes isolated from plants should be identified and designed to produce drugs with insignificant side effects (Kumar et al., 2010). In this regard, a RIP II gene was isolated from native *S. ebulus* plant in the north of Iran, and its primary and secondary structures, physico-chemical and functional properties were evaluated by various bioinformatics tools. The obtained results can be useful for a comprehensive understanding of the nature and mechanism of action of RIP proteins as well as the accurate design of new drug substitutes from plant sources.

## 2. Material and methods

### 2.1. Plant material and gene cloning

*Sambucus ebulus* plants were collected from the north of Iran (36°39.5'N 53°4.3'E) during July 2017. Fresh leaves were finely grinded and used for DNA extraction according to a CTAB method (Doyle and Doyle, 1987). Forward (5'-ATGATAGACTATCCCTCCGTC-3') and degenerate reverse primers (5'-CTAAMTTKGARG GACTTGTGT-3') were designed according to the deposited RIP gene sequences in the GenBank, NCBI (AJ400822.1, AF249280.1, U41299.1 and AF409135.1), using the Primer premier software. PCR amplifications were carried out with the following cycling parameters: 35 cycles of 94 °C for 1 min, 68 °C for 1 min, and 72 °C for 2 min with a final extension at 72 °C for 10 min. PCR products were subjected to electrophoresis in a 1% (w/v) agarose gel, DNA fragments of the expected length were extracted and purified with High Pure PCR Purification Kit (Cat. No. 11732668, Roche). The purified DNA fragments were cloned into pTZ57R/T vector (Thermo Scientific, InsTAclone PCR Cloning Kit) and were sequenced by universal M13 primers.

### 2.2. Nucleotide sequence accession number

The RIP II gene sequence obtained in this study is available in the GenBank under accession number MH053462. The isolated RIP gene was named as pebulin, in which p at the beginning of the phrase stands for Persian.

### 2.3. Gene ontology

For gene ontology and gene annotation of EBULIN protein, the keyword “ribosome-inactivating protein” is searched as gene product in QuickGO ([ebi.ac.uk/QuickGO/](http://ebi.ac.uk/QuickGO/)). The ancestor chart of related GO term was also obtained.

### 2.4. Conserved domains search

Using the NCBI conserved domain database (CDD, National Institutes of Health, Maryland, USA), identification of conserved domains within pebulin was done.

### 2.5. Primary protein structure and physicochemical properties analysis

Nucleotide sequence of isolated gene was translated to a protein sequence based on standard genetic code using ExPASy translate tool. The sequence similarity search was done by the blastp algorithm, BLAST tool ([blast.ncbi.nlm.nih.gov](http://blast.ncbi.nlm.nih.gov)). Sequences of *Sambucus* spp. RIP II proteins were retrieved from UniProt database (Q9AVR2, Q41358, P33183 and O22415). The primary structure analysis and physicochemical characterization of protein were conducted by ExPASy ProtParam tool (<http://expasy.org/tools/prot-param.html>) (Gasteiger et al., 2005).

### 2.6. Protein solubility upon overexpression in *E. coli*

Probability of protein solubility upon overexpression in *E. coli* was predicted by SOLpro (<http://scratch.proteomics.ics.uci.edu/>) tool (Magnan et al., 2009).

### 2.7. Subcellular localization and transmembrane protein prediction

The presence of the N-terminal pre-sequences in the protein sequence was investigated by TargetP 1.1 (<http://www.cbs.dtu.dk/services/TargetP/>) to predict the subcellular location of the protein (Emanuelsson et al., 2007). Furthermore, TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) is used for the prediction of transmembrane helices in the protein (Krogh et al., 2001).

### 2.8. Disulfide bond prediction

DIpro (<http://scratch.proteomics.ics.uci.edu/>) a cysteine disulfide bond predictor was used for prediction of the sequence disulfide bonds, estimation the number of disulfide bonds, and prediction the bonding state of each cysteine and the bonded pairs (Cheng et al., 2006).

### 2.9. Post-translational modification prediction

Different post-biosynthesis protein modifications, including ubiquitination, phosphorylation, methylation, and glycosylation were predicted in the pebulin. For prediction of potential ubiquitination sites in the pebulin protein, UbPred tool is used (Radivojac et al., 2010). NetPhos 3.1 server is used to predict phosphorylation sites at threonine, serine and tyrosine residues of the protein (Blom et al., 1999). Prediction of potential arginine methylation sites, and glycosylation is done using PRmePRed tool ([bioinfo.icgeb.res.in/PRmePRed](http://bioinfo.icgeb.res.in/PRmePRed)) (Kumar et al., 2017) and NetNGlyc 1.0 server ([cbs.dtu.dk/services/NetNGlyc/](http://cbs.dtu.dk/services/NetNGlyc/)) (Gupta et al., 2004), respectively.

### 2.10. Secondary protein structure analysis

The secondary structure of RIPII protein chains is predicted using GOR IV tool (Garnier et al., 1998), and the percentage of alpha helices, extended strand, beta-turn, and random coils were determined.

### 2.11. Protein modeling

Based on the availability of *S. ebulus* protein structure as a template, SWISS-MODEL server (<https://swissmodel.expasy.org/>) is used for built pebulin protein models. Validation of predicted models is checked by generated Ramachandran plots.

### 2.12. Protein-ligand interactions

The Protein-Ligand Interaction Profiler (PLIP) is used for the analysis of non-covalent interactions (hydrogen bonds, water bridges,

salt bridges, halogen bonds, hydrophobic interactions,  $\pi$ -stacking,  $\pi$ -cation interactions, metal complexes) in protein-ligand complexes from custom pebulin PDB file (Salentin et al., 2015).

### 2.13. Protein chains interaction prediction

For analysis of protein chains interactions, COCOMAPS tool is applied ([molnac.unisa.it/BioTools/cocomaps/](http://molnac.unisa.it/BioTools/cocomaps/)) (Vangone et al., 2011). Threshold distance to select interacting residues was set as the default value (8 Å).

### 2.14. Prediction of geometric and topological properties of protein structures

Computed Atlas of Surface Topography of proteins (CASTp) tool is used for locating, delineating, and measuring the geometric and topological properties of the pebulin protein structure (Wei et al., 2018).

### 2.15. Phylogenetic analysis

Similarity comparison of each protein chain sequence separately was done by DELTA-BLAST algorithm in protein databank (PDB) database. Sequences were producing significant alignments with E-value better than the threshold (0.05) after several iterations were selected and MSA was done by Clustal  $\Omega$  program. Phylogenetic analyses and tree construction were conducted with MEGA 6.

## 3. Results and discussion

### 3.1. Gene cloning and sequence analysis of the pebulin gene

The pebulin gene (1641 bp) was isolated from *S. ebulus* leaves using PCR with specific primers. The purified PCR product was

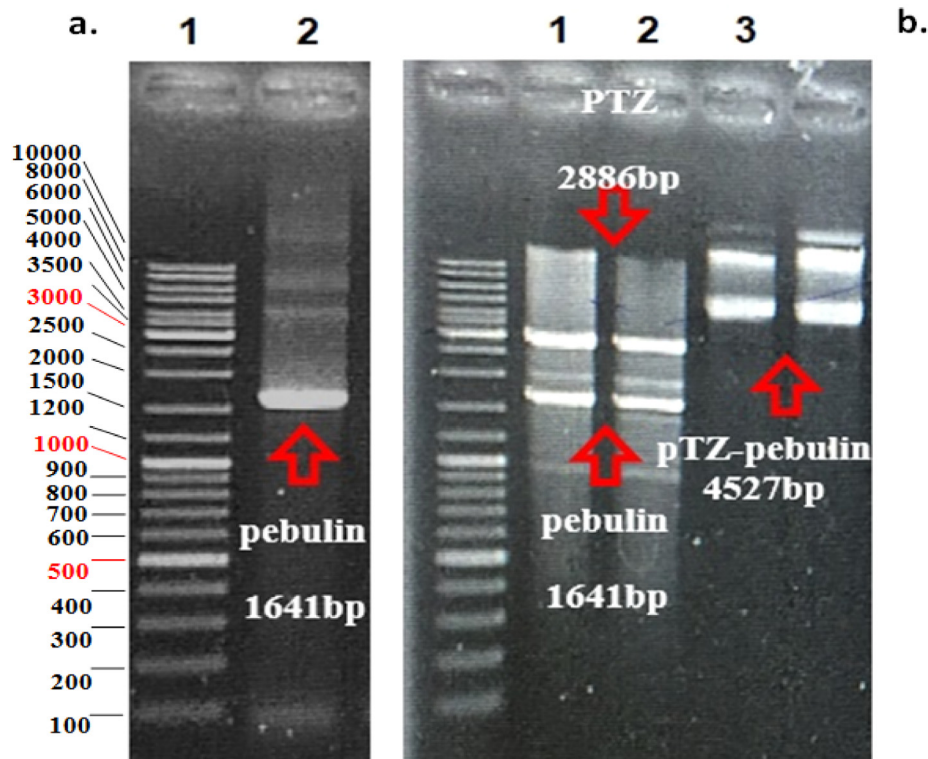
cloned into pTZ57R/T vector and transformed into *E. coli* DH5 $\alpha$  competent cells. The accuracy of cloning was confirmed by restriction enzyme digestion and colony PCR. The results of PCR and digestion are shown in Fig. 1. After sequencing the desired gene with the M13 universal primers, the validated full-length sequence of the pebulin gene was submitted to the GenBank database (MH053462). The obtained sequence was used for subsequent bioinformatics analyses.

### 3.2. Gene ontology

Two GO terms were obtained by searching the “ribosome-inactivating protein” as the gene product in QuickGO. Generally, negative regulation of translation is the biological process of RIP (GO:0017148) that prevents and reduces protein synthesis rate. Another GO term (GO: 0030598) indicated the molecular function of the RIP, which catalyzes the hydrolysis of an N-glycosidic bond at adenine base in 28S rRNA. As shown in Fig. 2, GO:0017148 is related to the negative regulation of biological processes such as cellular protein metabolic process (GO:0032269), cellular macromolecule biosynthetic process (GO:2000113) and gene expression (GO:0010629).

### 3.3. Protein conserved domain and chain determination

The pebulin A and B chains amino acid sequences were obtained by comparing the protein sequence with the previously determined *S. ebulus* type II RIP chains (UniProtKB: Q9AVR2) based on the X-Ray crystal structure determination method (Fig. 3). The length of A and B chains was 254 and 267 aa, respectively, and there is a 24 aa intermediate sequence between the pebulin chains (Table 1). Analyses of the conserved region using CDD suite showed that all the three type II RIPs of *Sambucus* spp. shared similar conserved domains with identical locations including a domain



**Fig. 1.** (a) PCR product of pebulin gene. Lane 1: GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific); Lane 2: pebulin gene amplicon with predicted size, 1641 bp. (b) Gel electrophoresis analysis of recombinant plasmids by BamHI/XhoI restriction enzyme digestion on a 1% agarose gel. Lane 1: GeneRuler 1 kb DNA Ladder; Lanes 2, 3: digested pTZ-pebulin recombinant plasmid; Lanes 4, 5: undigested pTZ-pebulin recombinant plasmid.

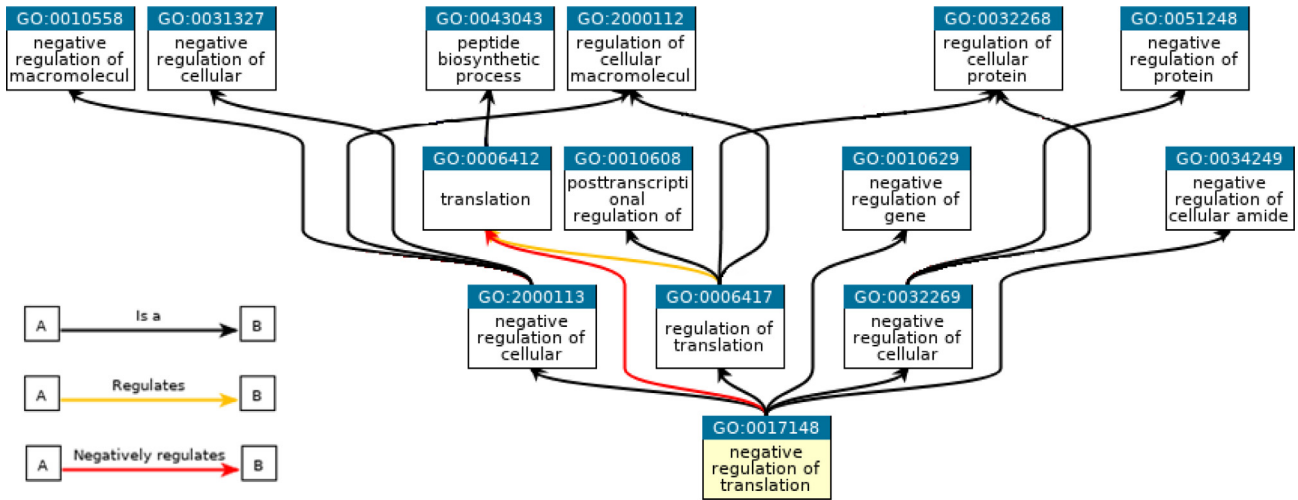


Fig. 2. Simplified ancestor chart for pebulin (ebulin isolated from Persian *S. ebulus*), GO: 0017148.

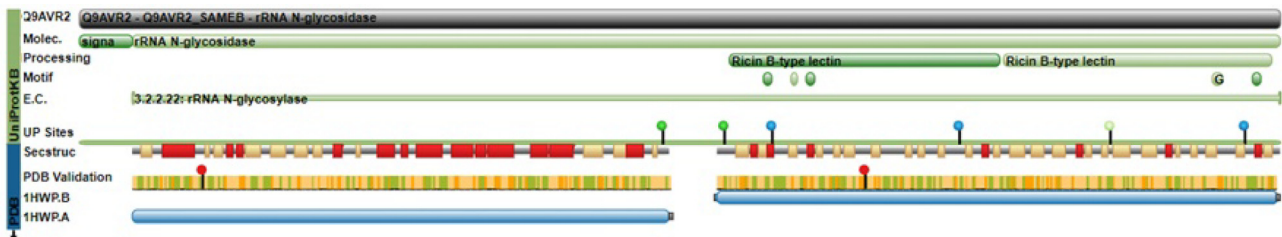


Fig. 3. Structure of ebulin (*S. ebulus*) RIP II protein chains (UniProtKB: Q9AVR2).

**Table 1**  
The sequence and length of pebulin protein chains.

Chain	length	sequence
A	254	IDYPSVSNLTGAKWTTYRDFIKDLRQIVANGTYEVNGLPVLRRNEVEQEKNRFLVLL TNYNGDTVTLAVDVTNLYVAFMANGTSYFFNDITPLERNNLFRETTQHILPYTGNYE HLERAARSTRESTNLGPDPLDEAITLWYNGSIARSLLVVIQMVSEAAARFGYIEQIRRSIR KQVCFTPSALMLSMENNWSSMSLEVQQSGDN VSPFSGTVQLQNYNHTLRLVDNFEELYQITGIAILLFRCVSPRS B
B	B1	DGETCPVAASFTRKISGGRDGLCVDRVNGYDTDGTPIQLFPCGSEKNQWTFYKDGTT RSMGKCMTANGLNSGSIMTFNCDTAVENATKWALPIDGSIINPSS GRVITAPSAASRTLLLDNNIHAASQGWTVSN
	B2	DVQPIVTSIVGYNETCLQANGENNRVWMECEITSLQQVWVLFGRDTRVNSDRGLC VTSNGYSSKDLIILKCCQGLASQSWLFNSDGTIVNLTATLVMDVKQSDVSLRQIIVPP TGNPNQQWRTQVPQI
Intermediate sequence	24	SSSYCNDKALRMPLVLAGEDNKYN
Signal peptide	25	MRVVKAAMLYLHIVVLAIYSVGIQG

on A chain belonging to RIP superfamily (cl08249) and two domains on B chain from Ricin-B-lectin superfamily (cl26069). The reported results predict that these proteins have been highly conserved and have the same activities. The schematic representation of pebulin conserved domains is shown in Fig. 4.

### 3.4. Gene and protein sequence identity

The identities of gene and protein sequences from *Sambucus* spp. Type II RIPs are shown in Table 2. The protein sequence identity of pebulin with *S. ebulus* (CAC33178.1) and *S. nigra* (AAC15886.1) protein sequences was 78.1% and 79.8%, respectively. The identity of pebulin with *S. nigra* B chain sequences was 81.2%, and the highest similarity was observed between B1 sub-chains (83.8%) compared to B2 sub-chains (78.6%). These sequence differences determine the specificity of each RIP and

are the main reason for the contrasting results reported for RIPs in previous research. The ribosome-binding specificity of each RIP and the differences between ribosomes among various species determine the efficiency of RIP in blocking the ribosome activity (Domashevskiy and Goss, 2015).

### 3.5. Primary protein structure and physicochemical properties analysis

The primary structure analysis and physicochemical properties of pebulin and other RIP II genes are shown in Table 3. The estimated half-life of all proteins was similar within 20 h (mammalian reticulocytes, in vitro), 30 min (yeast, in vivo) and >10 h (*Escherichia coli*, in vivo).

The relative volume of a protein occupied by its aliphatic side chains (alanine, valine, isoleucine, and leucine) is defined as aliphatic index, which plays an important role in determining protein

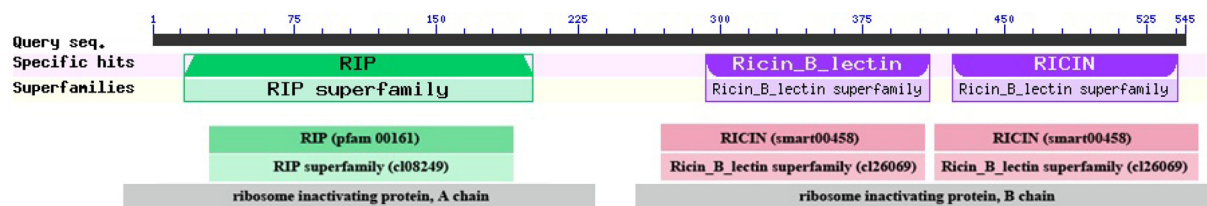


Fig. 4. Conserved domains of pebulin protein.

Table 2

The nucleotide and amino acid sequence identity of *Sambucus* spp. Type II RIPs and sequence identity comparison of protein chains.

Sequence	Pebulin		<i>S. ebulus</i>		<i>S. nigra</i>	
	Nucleotide	Protein	Nucleotide	Protein	Nucleotide	Protein
Pebulin QAU19548.1	ID	ID	0.882	0.781	0.895	0.798
<i>S. ebulus</i> CAC33178.1	0.882	0.781	ID	ID	0.904	0.840
<i>S. nigra</i> AAC15886.1	0.895	0.798	0.904	0.840	ID	ID
Chain	Pebulin protein		<i>S. ebulus</i> protein		<i>S. nigra</i> protein	
	A	B (B1, B2)	A	B (B1, B2)	A	B (B1, B2)
pebulin	ID	ID	0.771	0.790 (0.794, 0.786)	0.783	0.812 (0.838, 0.786)
<i>S. ebulus</i>	0.771	0.790 (0.794, 0.786)	ID	ID	0.889	0.793 (0.851, 0.732)
<i>S. nigra</i>	0.783	0.812 (0.838, 0.786)	0.889	0.793 (0.851, 0.732)	ID	ID

Table 3

Comparison of physicochemical properties of *Sambucus* spp. type II RIPs.

Protein	No. aa	MW (KDa)	The. pl	Asp + Glu	Arg + Lys	Ext. coefficient* M <sup>-1</sup> cm <sup>-1</sup> Abs 0.1%	Ext. coefficient** M <sup>-1</sup> cm <sup>-1</sup> Abs 0.1%	Instability index	Aliphatic index	GRAVY
Pebulin	521	58.020	5.19	51	42	79465, 1.37	78840, 1.359	34.99	85.83	-0.245
<i>S. ebulus</i>	520	57.792	5.12	53	44	79465, 1.375	78840, 1.364	32.36	84.69	-0.242
<i>S. nigra</i>	520	57.649	5.69	47	42	73965, 1.283	73340, 1.272	37.76	84.69	-0.193

\* Assuming all pairs of Cys residues form cysteines.

\*\* Assuming all Cys residues are reduced.

Table 4

Predicted disulfide bonds in *Sambucus* spp. RIP II proteins.

Protein	Number of disulfide bonds	Positions
Pebulin	5	249–259, 277–296, 318–335, 406–421, 447–464
<i>S. ebulus</i> / <i>S. nigra</i> RIPs	4	249–259, 276–295, 405–420, 446–463

thermostability (Panda and Chandra, 2012). Based on an aliphatic index, it is predicted that pebulin is more thermostable than the other known RIPs. The high aliphatic index of the studied RIPs (84.69 and 85.83) indicated that these proteins have high amounts of aliphatic side chains; hence they are thermally stable and could be used at high temperatures and maintain their activities. Thermostability of pebulin can also be attributed to the greater number and different position of disulfide bonds in pebulin compared to other *Sambucus* spp. RIP proteins (shown in Table 4) (Khoo and Norton, 2011).

In a study, the thermal stability of abrin II, a heterodimeric RIP isolated from *Abrus pulchellus*, and its subunits were evaluated under different pH and ligand binding conditions, using differential scanning calorimetry (DSC). Two peaks were observed in DSC scan

of abrin II, which may be attributed to the presence of 2 different entities which are unfolded at various temperatures (Krupakar et al., 1999).

All *Sambucus* spp. RIP proteins included in this study showed negative GRAVY values; hence they are hydrophilic and can easily interact with water. However, the GRAVY value of *S. nigra* is lower than other proteins. RIP of *S. nigra* had the highest pl (5.69) and the pl values of pebulin (5.19) and *S. ebulus* RIP (5.12) were almost similar. Knowing the calculated isoelectric point (pl) will be essential for the development of buffering systems in the purification process. The values of instability index indicated that almost all RIP proteins are stable in nature (instability index < 40).

The total number of negatively charged residues (Asp + Glu) was different among *Sambucus* spp. RIPs, as the highest number (53) and the lowest number (47) were observed in *S. ebulus*, and *S. nigra*, respectively. The highest number of positively charged residues (Arg + Lys) was in *S. ebulus* RIP and the number of these residues was similar in pebulin and *S. nigra*.

### 3.6. Protein solubility upon overexpression in *Escherichia coli*

There is a high demand for pure, soluble, and functional proteins in industry, food production, agriculture, and bioengineering. So, high-level production of soluble recombinant proteins in bacte-

rial expression hosts such as *E. coli* has great importance (Rosano and Ceccarelli, 2014). Prediction of protein solubility upon overexpression in *E. coli* helps to employ appropriate strategies for correct protein folding and handle protein misfolding (Baneyx and Mujacic, 2004).

In the present study, the SOLpro tool was used to predict the solubility of recombinant RIP proteins from *Sambucus* spp., overexpressed in *E. coli*. The results indicated that pebulin, ebulin and *S. nigra* RIP proteins are insoluble with the probability of 0.72, 0.6 and 0.53, respectively.

In some cases, the growth of the host bacterial cell was remarkably affected by expression of a recombinant toxin in *E. coli*, which might have resulted from depurination of ribosomes (Chaddock et al., 1994). Using *E. coli* strains specifically selected for toxic protein expression and modifying the expression conditions can be implemented to overcome this issue (Zuppone et al., 2019).

The mature  $\alpha$ -luffin (a type I RIP from *Luffa cylindrica* seeds) gene linked with SUMO fusion tag was cloned, and the recombinant  $\alpha$ -luffin was successfully expressed in *E. coli*. To prevent the toxic effect of  $\alpha$ -luffin on the host expression, a rhamnose promoter was used to tightly control the protein expression. The high level of soluble expression of  $\alpha$ -luffin was achieved by reducing the induction temperature, fed-batch cultivation system, and SUMO fusion tag (Namvar et al., 2018). The curcin (a type I RIP from *Jatropha curcas*) was expressed in *E. coli* strain M15, and the target protein was produced in the form of inclusion bodies (Luo et al., 2007). Likewise, the unglycosylated A-chain sequence of MLI (a type II RIP) was expressed in BL21 strain of *E. coli*. The resulting recombinant protein was insoluble, and further purification and refolding of recombinant MLI led to a pure and homogeneous protein species with an apparent molecular mass of 27 kDa and a pI value of 6.4 (Eck et al., 1999).

### 3.7. Subcellular prediction

Based on the TMHMM results, a transmembrane helix with the 94.65% probability was predicted in the N-terminal of the pebulin sequence. The expected number of amino acids in transmembrane helix is 21.08 amino acids. The results were re-evaluated by TargetP server using PLANT networks that showed the sequence contains a signal peptide with the 0.989 score that was used by the secretory pathway and membrane translocation of the protein. The sequences with a score higher than 0.5 are considered as putative signal peptides. Thus, sequences that have a score above 0.7 are more probable to do the same activity (Petersen et al., 2011).

### 3.8. Disulfide bonds prediction

Disulfide bonds play an important role in protein folding and structural stability. Predicting accurately disulfide bonds from protein sequences is important for modeling the structural and functional characteristics of proteins. Five disulfide bonds were predicted in pebulin protein by Dipro algorithm. In all *Sambucus* spp. Type II RIPs, the inter-chain disulfide bond (between 249 and 259 Cys residues) that links A chain to the B chain was identical. However, the number and position of intra-chain disulfide bonds in pebulin B chain were more than those in *S. ebulus* and *S. nigra* RIP proteins (Table 4).

Considering the fact that the formation of disulfide bonds is often the main part of the folding pathway of a protein and their position in the protein structure may also influence its role in stabilizing or folding of the protein. The differences among *Sambucus* RIPs may lead to an alternative folding of pebulin protein. Moreover, disulfide bonds maintain protein integrity and protect proteins from damage by stabilizing the tertiary and/or quaternary structures of proteins, increase the thermodynamic stability, making them less susceptible to denaturation and degradation and more resistant to high temperature and pH (Khoo and Norton, 2011).

Proteins with high sequence identity do not always exhibit the same disulfide bond pattern. In a study on families of proteins, the conservation of disulfide bonds in homologous protein domains was investigated, and only 54% conservation was found, showing a poor relationship between sequence identity and disulfide bond conservation (Thangudu et al., 2008). In general, disulfide bonds appear to play important structural or functional roles and are well conserved in proteins with similar folding (Khoo and Norton, 2011). However, it has been shown that some disulfide bonds have no direct role in protein's function, and removing them may have minor effects on the structure and function of the protein. In many cases, the removal of a disulfide bond has been accompanied by local changes in the protein structure, but the folding has remained unchanged (Khoo and Norton, 2011).

### 3.9. Post-translational modifications prediction

#### 3.9.1. Ubiquitination site prediction

Attachment of ubiquitin to lysine residues in proteins (ubiquitination) is one of the most important reversible post-translational modifications that controls the activity of proteins (Peng et al., 2003). Furthermore, protein ubiquitination plays significant roles

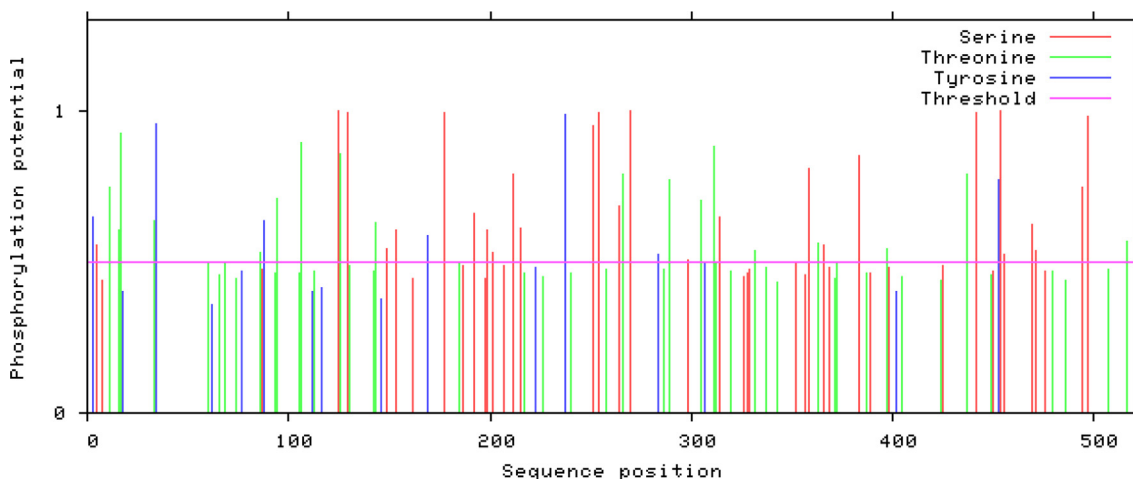


Fig. 5. Predicted phosphorylation sites in pebulin protein sequence isolated from Iranian *S. ebulus*. (NetPhos 3.1a).

**Table 5**  
Arginine methylation sites prediction of pebulin.

SeqId		R site	Peptides	Prediction Score
A Chain		121	YTGNYEHLERAARSTREST	0.561527
		124	NYEHLERAARSTRESTNLG	0.776309
		127	HLERAARSTRESTNLGPDP	0.866791
		248	ITGIAILLFRCVSPRSXXX	0.540343
B Chain	B1	14	CPVAASFTRKRISGGRDGLC	0.879204
		19	SFTKRISGGRDGLCVDVRN	0.869622
		27	GRDGLCVDVRNGYDTDGTP	0.74138
		59	WTFYKDGTRSMGKCMTAN	0.569186
		106	GSIIINPSSGRVITAPSAAS	0.911514
	B2	116	VITAPSAASRTLLLDNNI	0.667303
		54	DRTIRVNSDRGLCVTSNGY	0.792351

in various biological functions, immune response, signal transduction, DNA repair, receptor modulation, and transcription regulation (Ebner et al., 2017). Potential ubiquitination sites in *Sambucus* spp. RIP II proteins were evaluated by UbPred tool. Based on the results, a ubiquitination site at position 300 with score 0.67 was predicted, which falls in the range of low confidence probabilities ( $0.62 < s < 0.69$ ). No possible ubiquitination sites were detected in the *S. ebulus* and *S. nigra* RIP proteins.

### 3.9.2. Phosphorylation site prediction

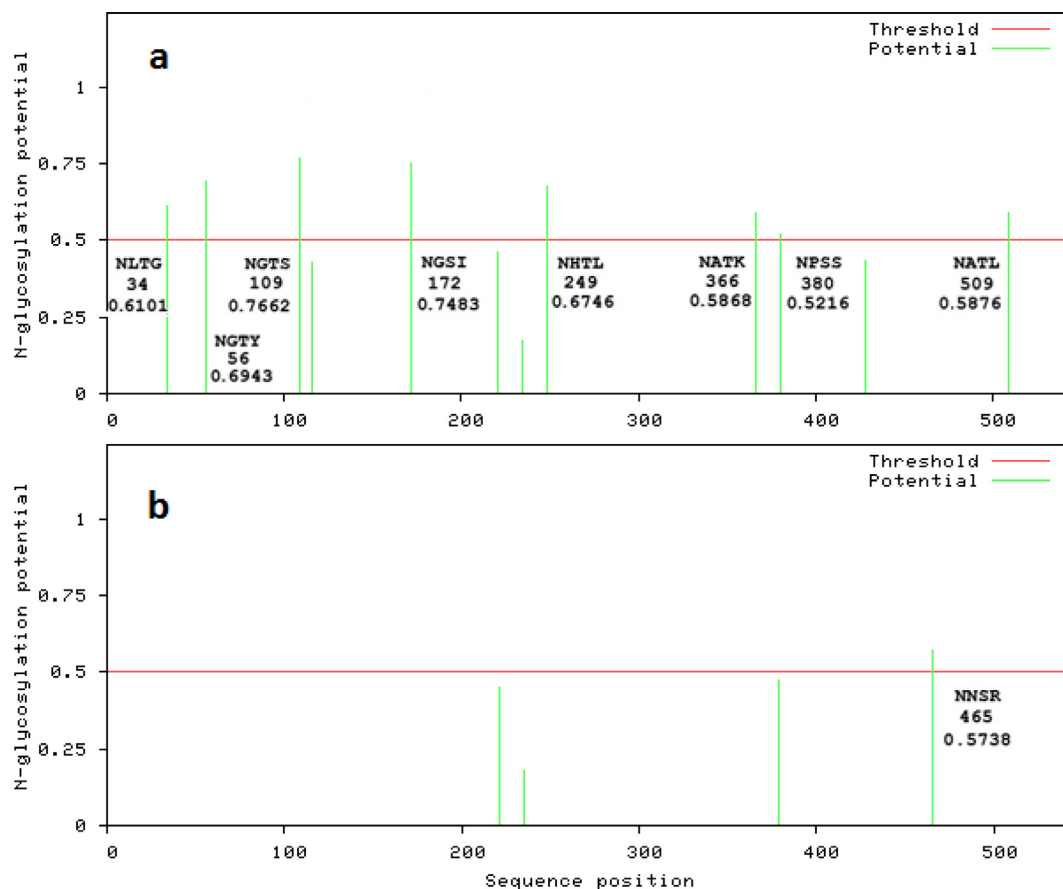
Protein phosphorylation plays important roles in biological processes, e.g., metabolism regulation, proliferation, differentiation, motility, survival and death, apoptosis, and subcellular trafficking (Wang et al., 2014). This type of post-translational modification mainly takes place on serine and threonine residues of eukaryotic proteins, whereas tyrosine phosphorylation is less abundant compared to the other PTMs (van Bentem and Hirt, 2009). As expected,

the number of potential Ser and Ther phosphorylation sites (26 and 18 sites, respectively) in pebulin were greatly higher than Tyr phosphorylation sites (7 sites). The results of phosphorylation site prediction are shown in Fig. 5. It is indicated that the functional effects of phosphorylation in a protein are site-dependent and they are functional if phosphorylation takes place on a specific site, not random (Olsen et al., 2006).

### 3.9.3. Arginine methylation sites prediction

Based on the PRmePred tool results, 11 arginine methylation sites were predicted in pebulin, four sites of which were in chain A and seven sites in chain B. The position and score of predicted arginine methylation sites are shown in table 5. The predicted arginine methylation sites in *S. ebulus* and *S. nigra* RIP proteins were 8 and 10, respectively.

Post-translational methylation at arginine residues significantly influences the structure and functions of the protein by changing



**Fig. 6.** N-Glycosylation sites prediction in pebulin (a) and *S. ebulus* RIP protein (b). The position and score of predicted sites are shown adjacent of each potential glycosylation site.

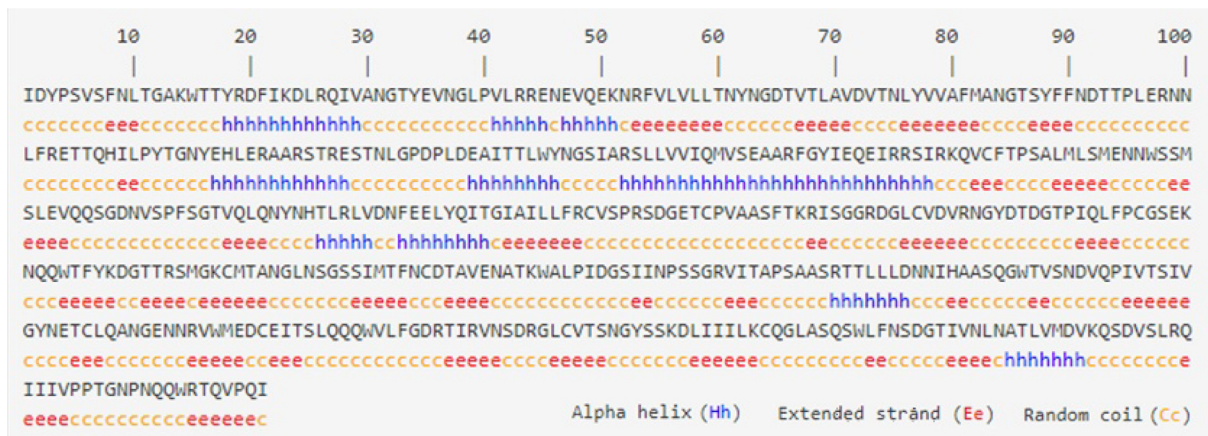


Fig.7. Secondary structure prediction of pebulin protein.

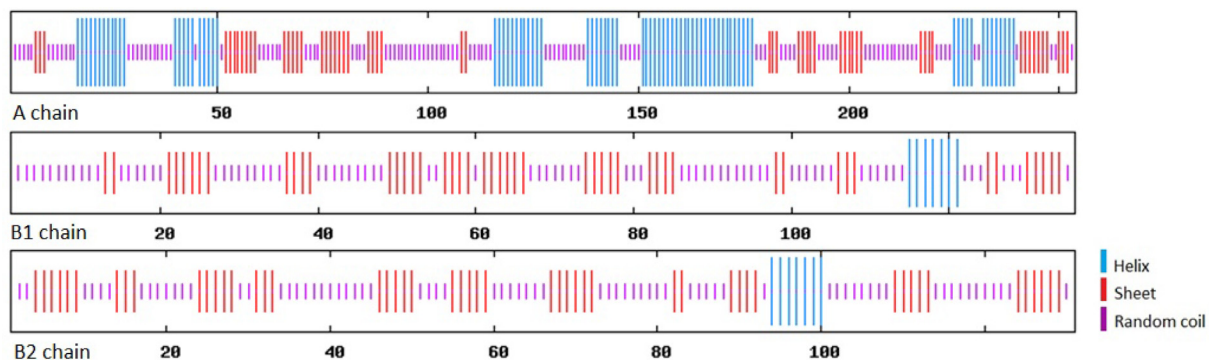


Fig. 8. Detailed secondary structure of pebulin protein chains.

Table 6  
Secondary structure prediction of *Sambucus* sp. RIP II proteins by GOR IV method.

Structure	Pebulin		<i>S. ebulus</i>		<i>S. nigra</i>	
	aa No.	%	aa No.	%	aa No.	%
Alpha helix	96	18.43%	98	18.18%	105	19.52%
Extended strand	149	28.60%	177	32.84%	157	29.18%
Random coil	276	52.98%	264	48.98%	276	51.30%

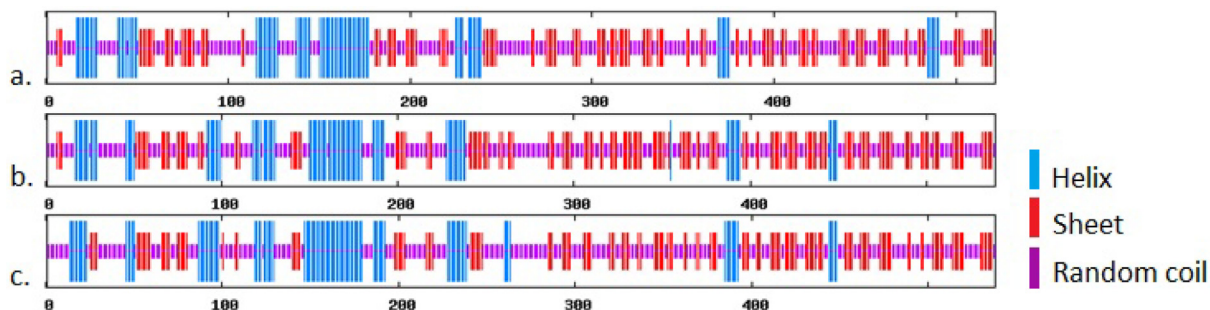
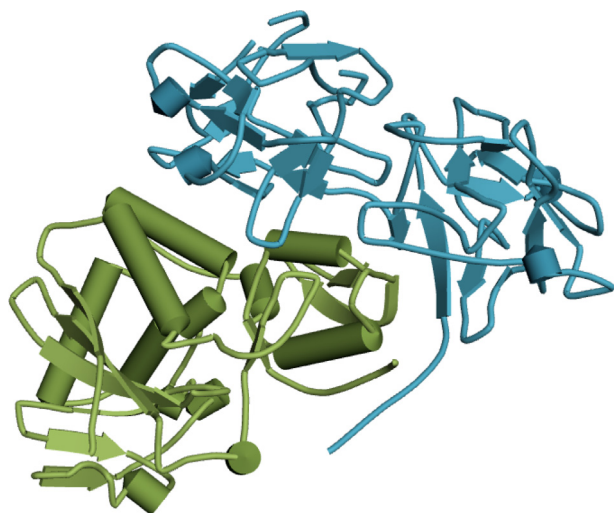


Fig. 9. Comparison of Secondary structure of *Sambucus* sp. RIP II proteins. (a) pebulin, (b) *S. ebulus*, and (c) *S. nigra*.

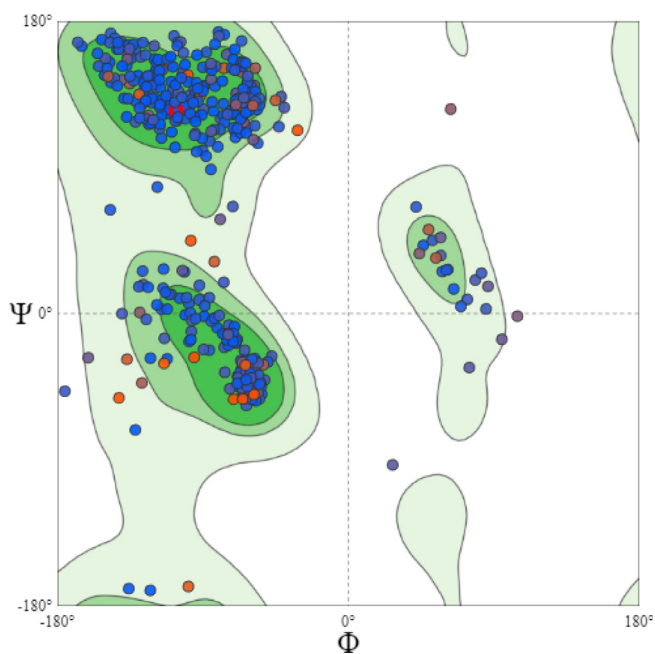
Table 7  
The structural template used for protein modeling.

Template	Seq Identity	Oligo-state	QSQE	Found by	Method	Resolution	Seq Similarity	Coverage	Description
1hwm.1	78.27	Hetero-1-1-mer	0.78	BLAST	X-ray	2.80 Å	0.54	1.00	EBULIN; EBULIN





**Fig. 10.** Schematic representation of pebulin protein. The chains are shown in different colors, Chain A (green) and Chain B (blue).



**Fig. 11.** General Ramachandran Plot of pebulin chains structure assessment. The Darkest region is the most favored region consisting of more than 90% of residues.

**Table 8**

The result of the model quality evaluation of pebulin obtained from MolProbity 4.4 tool.

MolProbity Score	2.38	
Clash Score	3.37	
Ramachandran Favored	92.77%	
Ramachandran Outliers	0.98%	B265 PRO, B4 THR, A104 GLU, B87 ASN, B148 TYR
Rotamer Outliers	10.96%	B89 THR, B135 SER, B149 ASN, B37 ILE, A221 GLN, B77 MET, B70 LEU, B39 LEU, B112 SER, A215 SER, A163 GLU, A249 CYS, B63 LYS, B204 LEU, A128 GLU, A92 ASP, B261 ARG, B134 VAL, B116 ARG, B103 SER, A29 VAL, A175 ARG, A230 VAL, A220 LEU, B120 LEU, A173 GLU, B227 VAL, B248 ILE, B229 LEU, A152 ARG, B15 ILE, A137 LEU, B192 LEU, B266 GLN, A96 LEU, B61 MET, B71 ASN, B242 VAL, A154 LEU, A202 LEU, A23 LYS, A44 ARG, A47 GLU, A49 GLN, A219 GLN, A245 LEU, B38 GLN, B90 LYS, B100 ILE, B238 LYS
C-Beta Deviations	4	B232 THR, B19 ARG, B116 ARG, A251 SER
Bad Bonds	1/4136	B187 ASN_1 NAG
Bad Angles	19/5638	(B187 ASN_1 NAG), A92 ASP, (B266 GLN-B267 ILE), (A212 SER-A213 PRO), (B264 VAL-B265 PRO), B250 VAL, B57 THR, (A135 ASP-A136 PRO), (B40 PHE-B41 PRO), A93 THR, B77 MET, (B93 LEU-B94 PRO), (B251 PRO-B252 PRO), (B98 SER-B99 ILE), (A185 THR-A186 PRO), B108 ILE, B126 HIS, A118 HIS

**Table 9**

Prediction of Ligand- pebulin interactions.

Ligand	Added to Model	Description
NAG	✓	N-ACETYL-D-GLUCOSAMINE
BMA	× - Not in contact with the model.	BETA-D-MANNOSE
GAL	× - Binding site not conserved.	BETA-D-GALACTOSE
MAN	× - Not in contact with the model.	ALPHA-D-MANNOSE

the bulkiness and hydrophobicity of modified residues. Hence, it modulates a myriad of essential biological processes, including transcriptional regulation, RNA metabolism, DNA repair, signal transduction, protein sorting, and apoptosis, etc. (Ahmad et al., 2011). PRMTs (Protein arginine methyltransferases) methylate a large number of essential proteins and a wide variety of substrates, most of which are either RNA binding proteins or proteins involved in transcription such as histones, transcription factors and a number of proteins involved in RNA processing, transport, and translation. RNA methylation as an epigenetic mechanism occurs in different RNA species including tRNA, rRNA, mRNA, tmRNA, snRNA, snoRNA, miRNA, and viral RNA, affecting several biological processes such as RNA stability and mRNA translation (Wang et al., 2014, 2015; Ji and Chen, 2012; Dev et al., 2017), through a variety of RNA methyltransferases.

### 3.9.4. N-glycosylation sites prediction

Using NetNGlyc 1.0 server, eight potential N-glycosylation sites were predicted in pebulin protein, while only one glycosylation site was predicted in *S. ebulus* RIP protein (Fig. 6). More potential motifs were found in pebulin than in *S. ebulus* RIP protein, and the position of potential glycosylation sites was different between the two proteins. In eukaryotes, N-glycosylation takes place in the endoplasmic reticulum and/or the Golgi apparatus on secreted or membrane-bound proteins. N-linked glycosylation alters the structure and function of the eukaryotic proteins (Schwarz and Aebi, 2011). N-glycans can also affect protein stabilization and folding acceleration through an intrinsic chemical mechanism (Glozman et al., 2009). Many studies have revealed that glycoproteins are more stable than their corresponding unglycosylated counterparts, even when there are no major structural changes associated with glycosylation (Imperiali and O'Connor, 1999). It is shown that N-linked glycans may also facilitate the formation of a key segment of a secondary structure, serving to enhance the overall stability of the protein or to enable some specific function performed around the attached site (Qi et al., 2014). Accordingly, the higher potential N-glycosylation sites in pebulin in comparison with *S. ebulus* RIP may enhance the reversibility of the unfolding process and improve protein stabilization.

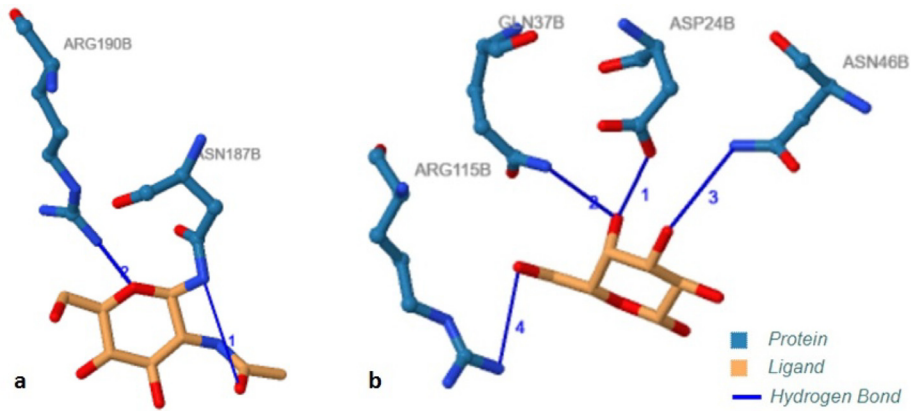


Fig. 12. Protein-Ligand Interaction of (a) pebulin and (b) *S. ebulus* RIP protein with N-ACETYL-D-GLUCOSAMINE. The number of amino acid residues in the PDB file of protein B chain are indicated.

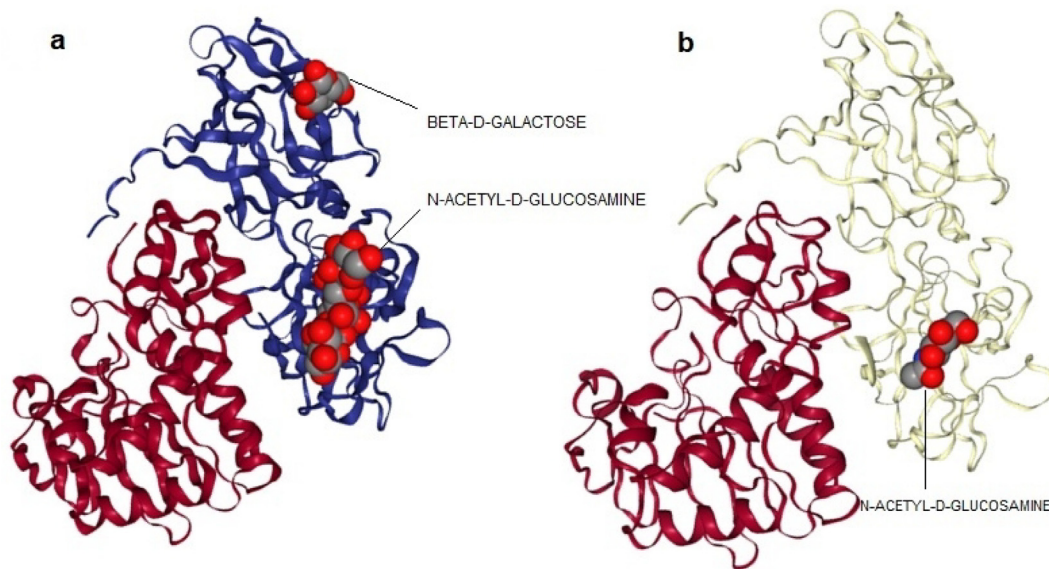


Fig. 13. Representation of ligand and protein interaction, (a) *S. ebulus* RIP and (b) pebulin.

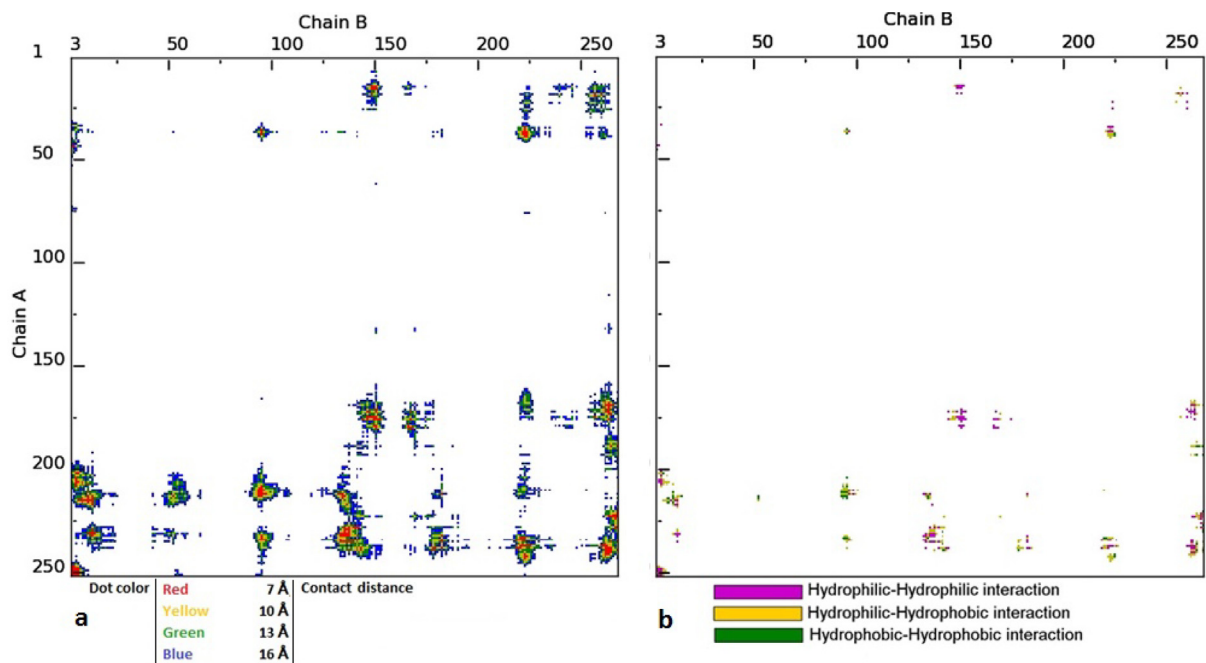


Fig. 14. (a) Contact map of distance ranges and (b) Contact map of property between pebulin protein chains.

**Table 10**  
Accessible surface area table for pebulin and *S. ebulus* RIP protein chains.

Title	Pebulin	<i>S. ebulus</i>
Buried area upon the complex formation (Å <sup>2</sup> )	3333.7	3251.6
Buried area upon the complex formation (%)	13.84	14.14
Interface area (Å <sup>2</sup> )	1666.85	1625.8
Interface area Chain A (%)	14.14	14.71
Interface area Chain B (%)	13.55	13.61
POLAR Buried area upon the complex formation (Å <sup>2</sup> )	1891.8	1898.9
POLAR Interface (%)	56.75	58.40
POLAR Interface area (Å <sup>2</sup> )	945.9	949.45
NON-POLAR Buried area upon the complex formation (Å <sup>2</sup> )	1441.9	1352.6
NON-POLAR Interface (%)	43.25	41.60
NON-POLAR Interface area (Å <sup>2</sup> )	720.95	676.3
Residues at the interface_TOT (n)	86	83
Residues at the interface_Chain A	43	41
Residues at the interface_Chain B	43	42

**Table 11**  
Interaction summary statistics of pebulin and *S. ebulus* RIP chains.

Title	Pebulin	<i>S. ebulus</i> RIP
Number of interacting residues Chain A	77	76
Number of interacting residues Chain B	79	80
Number of hydrophilic-hydrophobic interaction	208	205
Number of hydrophilic-hydrophilic interaction	176	170
Number of hydrophobic-hydrophobic interaction	59	58
Number of salt bridges	3	3
Number of hydrogen bonds	14	15
Number of disulfide bonds	1	1

### 3.10. Secondary structure prediction

GOR IV method was applied to predict the secondary structures (SS) of *Sambucus* spp. RIP II proteins. The result of SS prediction of pebulin is shown in Fig. 7 and Fig. 8. The pebulin protein is comprised of alpha helix, beta-sheet, and random coil with 18.43%, 28.6%, and 52.98%, respectively (Table 6). Furthermore, the distribution and percentage of different secondary structures in the *Sambucus* spp. RIP II proteins are compared and shown in Fig. 9 and Table 6, respectively. GOR IV analysis revealed that random coil has the maximum amount in all the proteins studied. The ran-

dom coil is not a true secondary structure but is the class of conformations that indicate an absence of regular secondary structure.

### 3.11. Protein modeling

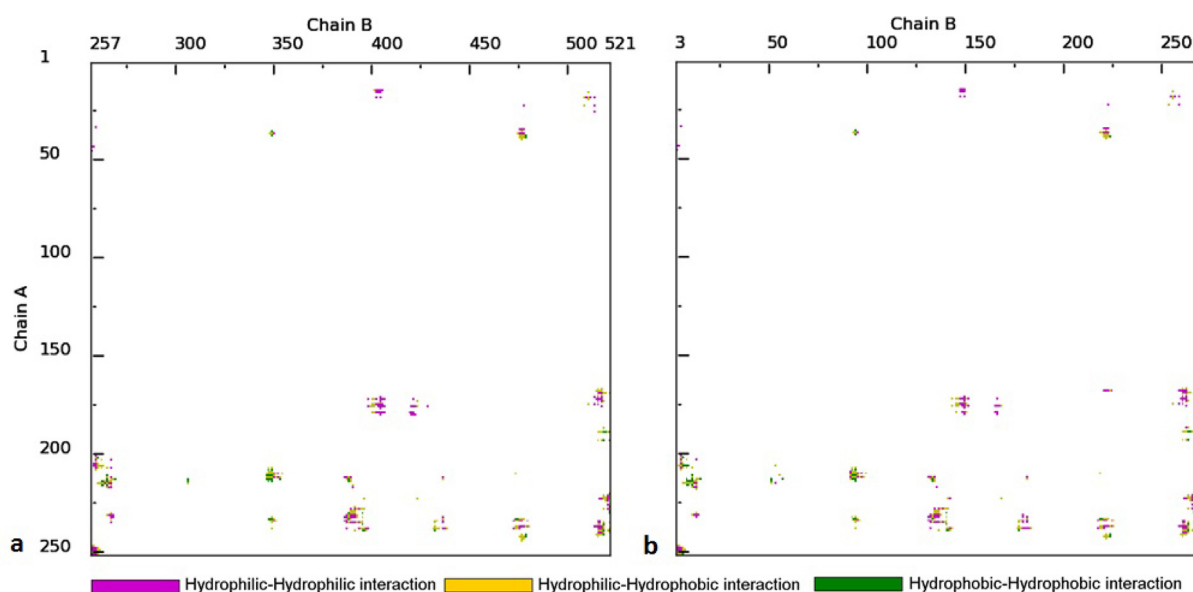
The SWISS-MODEL template library (SMTL version 2019-06-27, PDB release 2019-06-21) was searched with BLAST (Camacho et al., 2009) and HHblits (Remmert et al., 2012) for evolutionary related structures matching the templates to the target sequences. Among 144 templates found, 1hwm.1 was selected as the most appropriate template (Table 7) and used for protein model building. The built model was a hetero-1-1-mer, 1 × NAG: N-ACETYL-D-GLUCOSAMINE ligand with GMQE: 0.93 and QMEAN: −1.65. The Schematic representation of the model is shown in Fig. 10. The folding of the pebulin B chain is very similar to type II RIP B-chains and lectins and is composed of two beta domains (I and II) (Ferrerias et al., 2010).

#### 3.11.1. Model quality evaluation

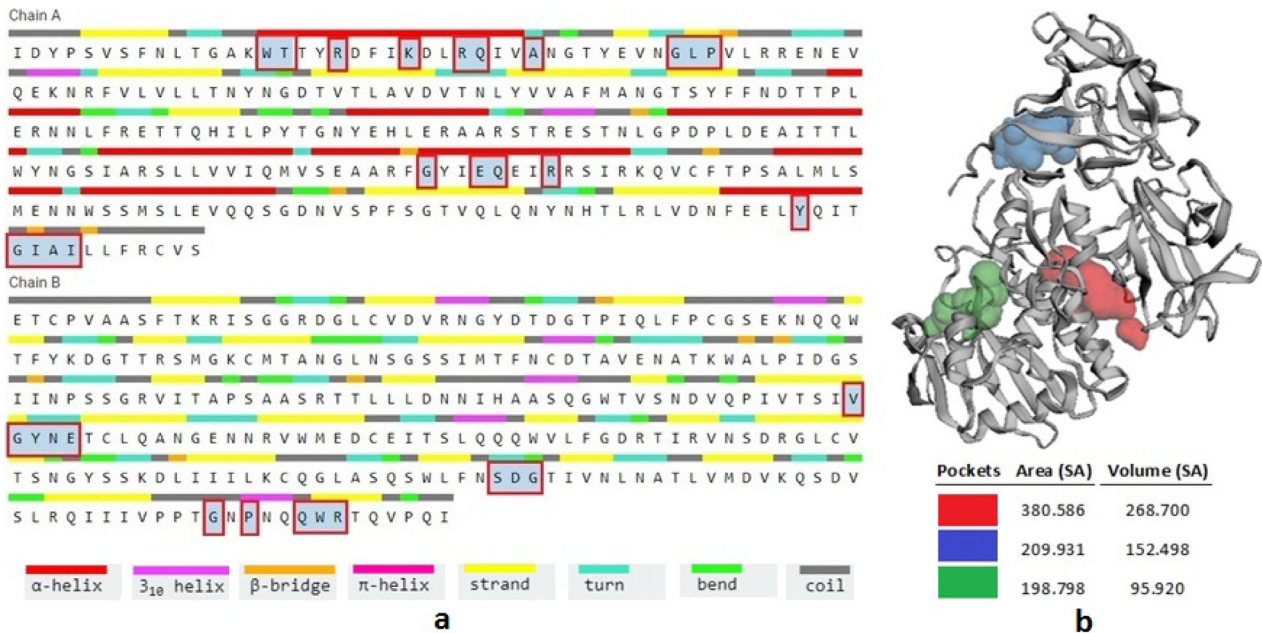
The results of structure assessment of the protein model were shown by Ramachandran Plot (Fig. 11) and MolProbity Results (Table 8). Ramachandran plot visualizes the backbone dihedral angles (Psi against Phi) of amino acid residues in the protein structure, and MolProbity gives information about problems in a 3D model of the protein. There is no Gly in Ramachandran outliers, but a Pro residue (265) in chain B is among outliers. Although Ramachandran outliers are the consequence of mistakes during the data processing, (Ramachandran and Sasisekharan, 1968) sometimes Ramachandran outliers might play a unique role in the function (Lenfant et al., 2013).

#### 3.11.2. Ligand modeling

The Protein-Ligand Interaction Profiler (PLIP) tool is used for the analysis of non-covalent interactions in protein-ligand complexes from the predicted pebulin protein model. The results of protein-ligand interaction prediction are shown in Table 9. Accordingly, among different ligands, only N-acetyl-D-glucosamine was added to the pebulin protein. The interaction of ligand with amino acid residues and spatial ligand position on the proteins are shown in Figs. 13 and 14, respectively. The N-acetyl-D-glucosamine molecule was attached to ASN187B and ARG190B by hydrogen bonds



**Fig. 15.** Contact map of (a) pebulin and (b) *S. ebulus* RIP protein chains.



**Fig. 16.** (a) Secondary structures and the functional active sites of pebulin largest surface concavity (red boxes). (b) Top three empty concavities of the pebulin protein (based on area and volume).

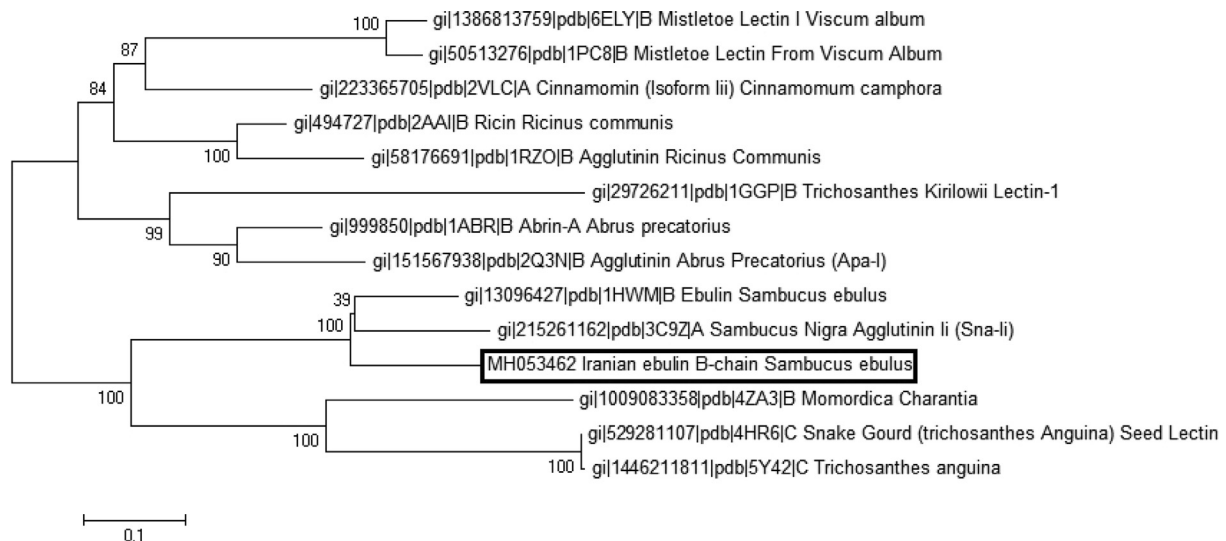
in pebulin, whereas this molecule was attached to four different amino acid residues by hydrogen bond (ASP24B, GLN37B, ASN46B and ARG115B) in *S. ebulus* RIP protein (Figs. 12 and 13).

### 3.12. Interactions between protein chains

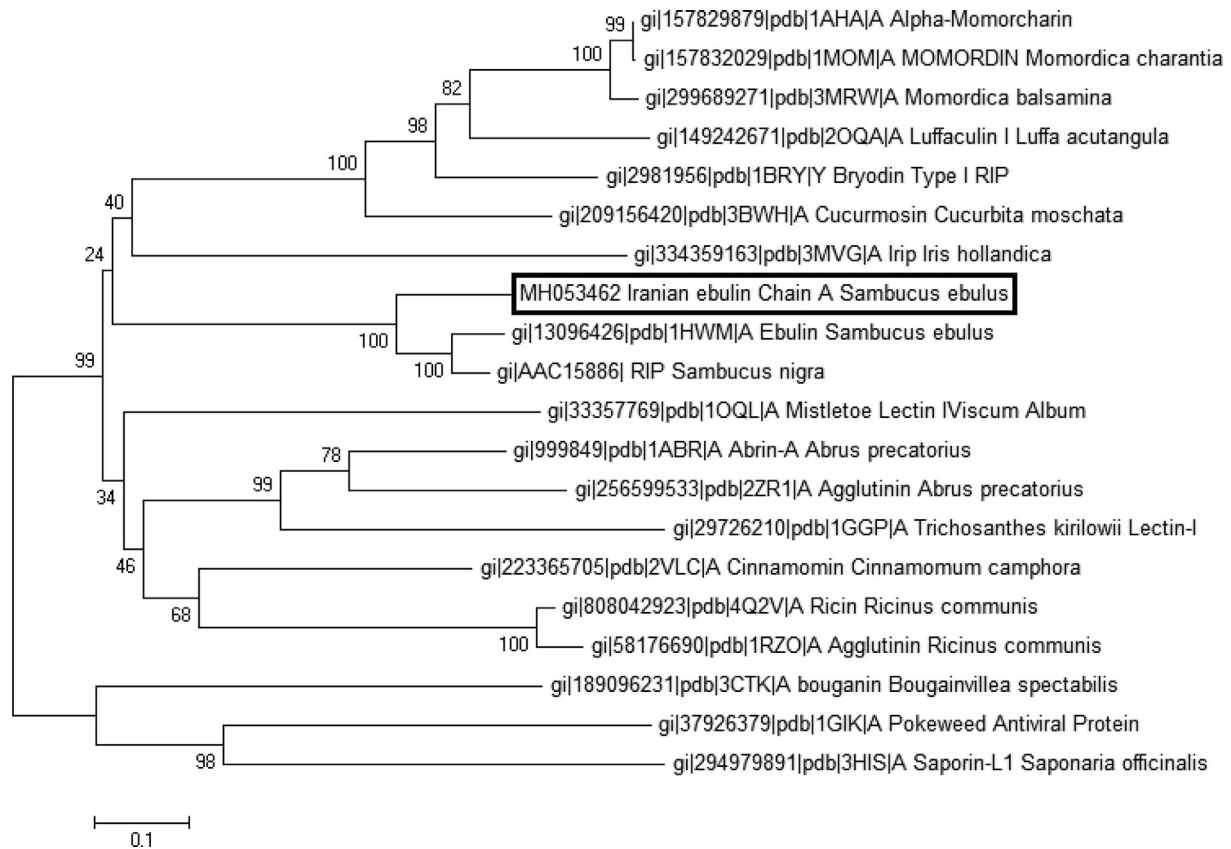
Accessible surface area and the interface statistics between the protein chains are shown in Tables 10 and 11, respectively. Based on the results, the amino acids that contributed in salt bridge formation are constant in both pebulin and *S. ebulus* RIP proteins and Cys249A and Cys5B that contributed to inter-chain sulfide bond are conserved. There are three substitutions in amino acids that contributed in H bond formation, Gly instead of Arg in position 168, Gln instead of Lys in position 238 in A chain, and Lys instead of Arg in position 13 in B chain.

In Fig. 14a., the residue distances between A and B chains are shown with color codes. The region of 150–250 aa of chain A is closer than 16 Å to the region of 100–250 aa of Chain B. The highest amount of interactions between chain A and chain B of pebulin are hydrophilic- hydrophobic interactions (Fig. 14b and Table 11).

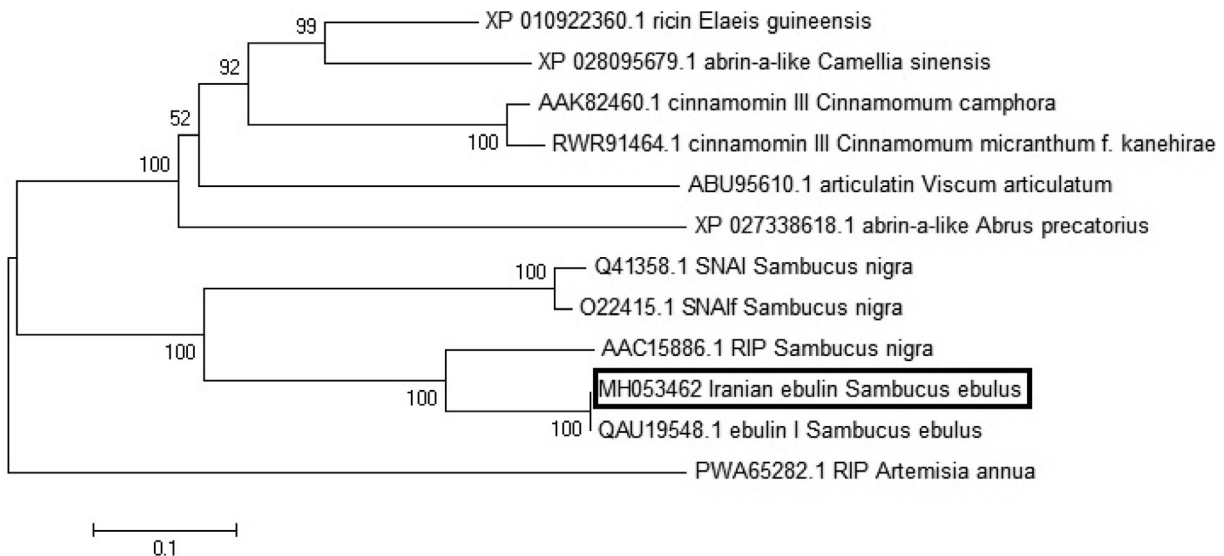
In Fig. 15, the contact maps of pebulin and *S. ebulus* RIP are compared. There is no significant difference between the contact position and the type of residue interaction in pebulin and another RIP. However, around position 175 of *S. ebulus* chain, some more hydrophilic-hydrophilic interactions are observed in comparison with pebulin. The C-terminal (200–250 residues) of both chain A RIP proteins have the highest hydrophobic-hydrophobic interactions with the two regions of B chain N-terminal, consisting of residues around position 10 and 100. On the whole, the results show that residue interactions in both proteins are reasonably similar; however, there are differences between hydrophilic-hydrophobic



**Fig. 17.** Phylogenetic tree of the Iranian ebulin (pebulin) B chain similar sequences constructed by Neighbor-joining method. Bootstrap values (1000 replicates) are reported as percentages. The scale bar represents the number of differences between sequences.



**Fig. 18.** Phylogenetic tree of the Iranian ebulin (pebulin) chain A similar sequences constructed by Neighbor-joining method. Bootstrap values (1000 replicates) are reported as percentages. The scale bar represents the number of differences between sequences.



**Fig. 19.** Phylogenetic tree of the Iranian ebulin (pebulin) similar sequences constructed by Neighbor-joining method. Bootstrap values (1000 replicates) are reported as percentages. The scale bar represents the number of differences between sequences.

and hydrophobic-hydrophobic interaction numbers which may lead to distinctive protein functions and are a good starting point to define the protein interaction network.

### 3.13. Geometric and topological properties of the protein structure

The function of proteins is fundamentally related to the geometric and topological properties of protein structures, including

surface empty concavities, internal cavities, and channels (Edelsbrunner et al., 1998). The detailed secondary structure of pebulin and the probable functional sites are shown in Fig. 16. Pockets are empty concavities on the protein surface into which the solvent can gain access, i.e., these concavities have mouth openings connecting their interior with the outside bulk solution. Three empty concavities based on area and volume are shown on pebulin protein in Fig. 16.

### 3.14. Phylogenetic analysis results

After three iterations in PSI blast, no new sequences were found above the 0.005 threshold in PDB protein database. The phylogenetic relationship of A and B chains of RIP proteins are shown in Figs. 17 and 18, separately. The phylogenetic tree revealed that as expected, A and B chains of pebulin with other *Sambucus* spp. RIP II protein chains were grouped together. This demonstrates that these proteins are highly conserved during evolution and carry out the same functions. The close phylogenetic relationship of isolated RIP from native Iranian *S. ebulus* with *Sambucus* spp. RIPs is shown in Fig. 19.

## 4. Conclusion

Research and study of RIPs have increased in recent years because of their use as part of immunotoxins, conjugates or recombinant chimeras with the aim of the treatment of several important diseases such as cancer (Allahyari et al., 2017; Naran et al., 2018), AIDS (Hogan et al., 2018) and autoimmune diseases. In addition, some RIPs have antiviral activity against animal and plant viruses (Zhu et al., 2018; Parikh and Tumer, 2004). Pebulin is a type II RIP isolated from native Iranian *S. ebulus*. In this study, the sequence of pebulin was evaluated using different bioinformatics tools and compared with two type II RIPs obtained from *Sambucus* species. Generally, the main biological function of *Sambucus* spp. RIP is the negative regulation of translation which prevents protein synthesis by the hydrolysis of a N-glycosidic bond at adenine base in 28S rRNA. Based on the vast bioinformatics analysis, this protein has distinct properties that differ from other *Sambucus* spp. type II RIPs. Hence, it can be a novel and promising protein in the researches related to RIP proteins.

## References

Ahmad, A., Dong, Y., Cao, X., 2011. Characterization of the PRMT gene family in rice reveals conservation of arginine methylation. *PLoS ONE* 6 (8), e22664.

Allahyari, H., Heidari, S., Ghamsosha, M., Saffarian, P., Amani, J., 2017. Immunotoxin: A new tool for cancer therapy. *Tumor Biol.* 39 (2), 1010428317692226.

Arias, F.J., Rojo, M.A., Ferreras, J.M., Iglesias, R., Munoz, R., Rocher, A., Mendez, E., Barbieri, L., Girbes, T., 1992. Isolation and partial characterization of a new ribosome-inactivating protein from *Petrocoptis glaucifolia* (Lag.) Boiss. *Planta* 186 (4), 532–540.

Baneyx, F., Mujacic, M., 2004. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* 22 (11), 1399.

Barbieri, L., Battelli, M.G., Stirpe, F., 1993. Ribosome-inactivating proteins from plants. *Biochim. Biophys. Acta* 1154 (3–4), 237–282.

Benitez, J., Ferreras, J.M., Munoz, R., Arias, Y., Iglesias, R., Cordoba-Diaz, M., et al., 2005. Cytotoxicity of an ebulin I-anti-human CD105 immunotoxin on mouse fibroblasts (L929) and rat myoblasts (L6E9) cells expressing human CD105. *Med. Chem.* 1 (1), 65–70.

Blom, N., Gammeltoft, S., Brunak, S., 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 294 (5), 1351–1362.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinf.* 10 (1), 421–430.

Chaddock, J.A., Lord, J., Hartley, M.R., Roberts, L.M., 1994. Pokeweed antiviral protein (PAP) mutations which permit *E. coli* growth do not eliminate catalytic activity towards prokaryotic ribosomes. *Nucl. Acids Res.* 22 (9), 1536–1540.

Cheng, J., Saigo, H., Baldi, P., 2006. Large-scale prediction of disulphide bridges using kernel methods, two-dimensional recursive neural networks, and weighted graph matching. *Proteins: Struct. Funct. Bioinf.* 62 (3), 617–629.

Citores, L., de Benito, F.M., Iglesias, R., 1997. Characterization of a new non-toxic two-chain ribosome-inactivating protein a structurally-related lectin from rhizomes of dwarf elder (*Sambucus ebulus* L.). *cell. Mol. Biol.* 43 (4), 485–499.

Dev, R.R., Ganji, R., Singh, S.P., Mahalingam, S., Banerjee, S., Khosla, S., 2017. Cytosine methylation by DNMT2 facilitates stability and survival of HIV-1 RNA in the host cell during infection. *Biochem. J.* 474 (12), 2009–2026.

Domashevskiy, A., Goss, D., 2015. Pokeweed antiviral protein, a ribosome inactivating protein: activity, inhibition and prospects. *Toxins* 7 (2), 274–298.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19 (1), 11–15.

Ebner, P., Versteeg, G.A., Ikeda, F., 2017. Ubiquitin enzymes in the regulation of immune responses. *Crit. Rev. Biochem. Mol. Biol.* 52 (4), 425–460.

Eck, J., Langer, M., Mockel, B., Baur, A., Rothe, M., Zinke, H., Lentzen, H., 1999. Cloning of the mistletoe lectin gene and characterization of the recombinant A-chain. *Eur. J. Biochem.* 264 (3), 775–784.

Edelsbrunner, H., Facello, M., Liang, J., 1998. On the definition and the construction of pockets in macromolecules. *Disc. Appl. Math.* 88 (1–3), 83–102.

Emanuelsson, O., Brunak, S., Von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2 (4), 953.

Ferreras, J.M., Citores, L., Iglesias, R., Jimenez, P., Girbes, T., 2010. *Sambucus* ribosome inactivating proteins and lectins. In: Lord, J.M., Hartley, M.R. (Eds.), *Toxic Plant Proteins-series Plant Cell Monographs*. Springer, Berlin, pp. 107–131.

Garnier, J., Gibrat, J. F., Robson, B., 1998. GOR secondary structure prediction method version IV, *Meth Enzym.*, RF Doolittle Ed. 266, p. 540–553.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExpASY server. In: *The Proteomics Protocols Handbook*. Humana Press, pp. 571–607.

Girbes, T., Citores, L., Ferreras, J.M., Rojo, M.A., Iglesias, R., Munoz, R., Arias, F.J., Calonge, M., Garcia, J.R., Mendez, E., 1993. Isolation and partial characterization of nigrin b, a non-toxic novel type 2 ribosome-inactivating protein from the bark of *Sambucus nigra* L. *Plant Mol. Biol.* 22 (6), 1181–1186.

Girbes, T., Citores, L., Iglesias, R., Ferreras, J.M., Munoz, R., Rojo, M.A., Arias, F.J., Garcia, J.R., Mendez, E., Calonge, M., 1993. Ebulin 1, a nontoxic novel type 2 ribosome-inactivating protein from *Sambucus ebulus* L. leaves. *J. Biol. Chem.* 268 (24), 18195–18199.

Girbes, T., Ferreras, J.M., Arias, F.J., Stirpe, F., 2004. Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria. *Mini Rev. Med. Chem.* 4 (5), 461–476.

Glozman, R., Okiyoneda, T., Mulvihill, C.M., Rini, J.M., Barriere, H., Lukacs, G.L., 2009. N-glycans are direct determinants of CFTR folding and stability in secretory and endocytic membrane traffic. *J. Cell. Biol.* 184 (6), 847–862.

Gupta, R., Jung, E., Brunak, S., 2004. Prediction of N-glycosylation sites in human proteins. *NetNGlyc 1.0 Server*.

Hogan, L.E., Vasquez, J., Hobbs, K.S., Hanhauser, E., Aguilar-Rodriguez, B., Hussien, R., et al., 2018. Increased HIV-1 transcriptional activity and infectious burden in peripheral blood and gut-associated CD4+ T cells expressing CD30. *PLoS Pathog.* 14 (2), e1006856.

Imperiali, B., O'Connor, S.E., 1999. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* 3 (6), 643–649.

Ji, L., Chen, X., 2012. Regulation of small RNA stability: methylation and beyond. *Cell Res.* 22 (4), 624.

Jimenez, P., Cabrero, P., Basterrechea, J.E., Tejero, J., Cordoba-Diaz, D., Cordoba-Diaz, M., 2014. Girbes T., Effects of short-term heating on total polyphenols, anthocyanins, antioxidant activity and lectins of different parts of dwarf elder (*Sambucus ebulus* L.). *Plant Food Hum. Nutr.* 69 (2), 168–174.

Khoo, K., Norton, R.S., 2011. Role of disulfide bonds in peptide and protein conformation. In: *Amino Acids, Peptides and Proteins in Organic Chemistry: Analysis and Function of Amino Acids and Peptides*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 395–417.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305 (3), 567–580.

Krupakar, J., Swaminathan, C.P., Surolia, A., Podder, S.K., 1999. Calorimetric studies on the stability of the ribosome-inactivating protein abrin II: effects of pH and ligand binding. *Biochem. J.* 338 (2), 273–279.

Kumar, P., Joy, J., Pandey, A., Gupta, D., 2017. PRmePred: a protein arginine methylation prediction tool. *PLoS ONE* 12 (8), e0183318.

Kumar, D.S., Sharathnath, K.V., Yogeswaran, P., Hara-ni, A., Sudhakar, K., Sudha, P., David, B., 2010. A Medicinal Potency of *Momordica charantia*. *Int. J. Pharm. Sci. Rev. Res.* 1 (2), 95–100.

Lee-Huang, S., Huang, P.L., Nara, P.L., Chen, H.C., Kung, H.F., Huang, P., Huang, H.I., Huang, P.L., 1990. MAP 30: a new inhibitor of HIV-1 infection and replication. *FEBS Lett.* 272 (1–2), 12–18.

Lenfant, N., Hotelier, T., Bourne, Y., Marchot, P., Chatonnet, A., 2013. Proteins with an alpha/beta hydrolase fold: relationships between subfamilies in an ever-growing superfamily. *Chem. Biol. Interact.* 203 (1), 266–268.

Lord, J.M., Roberts, L.M., Robertus, J.D., 1994. Ricin: Structure, mode of action, and some current applications. *FASEB J.* 8 (2), 201–208.

Luo, M.J., Liu, W.X., Yang, X.Y., Xu, Y., Yan, F., Huang, P., Chen, F., 2007. Cloning, expression, and antitumor activity of recombinant protein of curcin. *Russ. J. Plant Physiol.* 54 (2), 202–206.

Magnan, C.N., Randall, A., Baldi, P., 2009. SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* 25 (17), 2200–2207.

Mathews, M.B., Sonenberg, N., Hershey, J.W.B., 2007. *Translational Control in Biology and Medicine*. Cold Spring Harbor Press, NY.

Merino, M.J., Ferreras, J.M., Munoz, R., Iglesias, R., Girbes, T., 1990. Plant species containing inhibitors of eukaryotic polypeptide synthesis. *J. Exp. Bot.* 41 (1), 67–70.

Munoz, R., Ferreras, J.M., Iglesias, R., Merino, M.J., Girbes, T., 1991. Changes in sensitivity of in vitro rat brain protein synthesis to the acute action of ethanol and isopropanol as a consequence of the long-term ingestion of isopropanol. *Brain Res.* 65 (6), 500–504.

Namvar, S.H., Barkhordari, F., Raigani, M., Jahandar, H., Nematollahi, L., Davami, F., 2018. Cloning and soluble expression of mature  $\alpha$ -luffin from *Luffa cylindrica* in *E. coli* using SUMO fusion protein. *Turk. J. Biol.* 42 (1), 23–32.

Naran, K., Nundalall, T., Chetty, S., Barth, S., 2018. Principles of immunotherapy: implications for treatment strategies in cancer and infectious diseases. *Front. Microbiol.*, 9

- Nikolov, S., 2007. Specialized Encyclopedia of Medicinal Plants in Bulgaria, Sofia, Trud. (in Bulgarian).
- Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., Mann, M., 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127 (3), 635–648.
- Panda, S., Chandra, G., 2012. Physicochemical characterization and functional analysis of some snake venom toxin proteins and related non-toxin proteins of other chordates. *Bioinformatics* 8 (18), 891.
- Parikh, B.A., Tumer, N.E., 2004. Antiviral activity of ribosome inactivating proteins in medicine. *Mini Rev. Med. Chem.* 4 (5), 523–543.
- Pascal, J.M., Day, P.J., Monzingo, A.F., Ernst, S.R., Robertus, J.D., Iglesias, R., Perez, Y., Ferreras, J.M., Citores, L., Girbes, T., 2001. 2.8-Å crystal structure of a nontoxic type-II ribosome-inactivating protein, ebulin I. *Proteins* 43 (3), 319–326.
- Peng, J., Schwartz, D., Elias, J.E., Thoreen, C.C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., Gygi, S.P., 2003. A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* 21 (8), 921–926.
- Petersen, T.N., Brunak, S., von Heijne, G., von Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8 (10), 785.
- Peumans, W.J., Hao, Q., Van Damme, E.J., 2001. Ribosome-inactivating proteins from plants: More than N-glycosidases? *FASEB J.* 15 (9), 1493–1506.
- Qi, F., Zhang, W., Zhang, F., Chen, G., Liu, W., 2014. Deciphering the effect of the different N-glycosylation sites on the secretion, activity, and stability of cellobiohydrolase I from *Trichoderma reesei*. *Appl. Environ. Microbiol.* 80 (13), 3962–3971.
- Radivojac, P., Vacic, V., Haynes, C., Cocklin, R.R., Mohan, A., Heyen, J.W., Goebel, M.G., Iakoucheva, L.M., 2010. Identification, analysis and prediction of protein ubiquitination sites. *Proteins* 78 (2), 365–380.
- Ramachandran, G.T., Sasisekharan, V., 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* 23, 283–437.
- Remmert, M., Biegert, A., Hauser, A., Soding, J., 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* 9 (2), 173–175.
- Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5, 172.
- Salentin, S., Schreiber, S., Haupt, V.J., Adasme, F., Schroeder, M., 2015. PLIP: fully automated protein-ligand interaction profiler. *Nucl. Acids Res.* 43 (1), 443–447.
- Schwaiger, S., Zeller, I., Polzelbauer, P., Frotschnig, S., Laufer, G., Messner, B., et al., 2011. Identification and pharmacological characterization of the anti-inflammatory principal of the leaves of dwarf elder (*Sambucus ebulus* L.). *J. Ethnopharmacol.* 133 (2), 704–709.
- Schwarz, F., Aebi, M., 2011. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* 21 (5), 576–582.
- Shokrzadeh, M., Saeedi Saravi, S.S., Mirzayi, M., 2009. Cytotoxic effects of ethyl acetate extract of *Sambucus ebulus* compared with etoposide on normal and cancer cell lines. *Pharmacogn. Mag.* 5 (20), 316–319.
- Sipahioglu, H.M., Kaya, I., Usta, M., Unal, M., Ozcan, D., Ozer, M., et al., 2017. Pokeweed (*Phytolacca americana* L.) antiviral protein inhibits Zucchini yellow mosaic virus infection in a dose-dependent manner in squash plants. *Turk. J. Agric. For.* 41 (4), 256–262.
- Stirpe, F., 2004. Ribosome-inactivating proteins. *Toxicol.* 44 (4), 371–383.
- Stripe, F., Barbieri, L., Battelli, M.G., Soria, M., Lappi, D.A., 1992. Ribosome-inactivating proteins from plants: present status and future prospects. *Biotechnol. J.* 10 (4), 405–412.
- Tasinov, O., Kiselova-Kaneva, Y., Ivanova, D., 2012. Antioxidant activity, total polyphenol content and anthocyanins content of *Sambucus ebulus* L. aqueous and aqueous-ethanolic extracts depend on the type and concentration of extractant. *Sci. Technol.* 1 (2), 37–41.
- Tasinov, O., Kiselova-Kaneva, Y., Ivanova, D., 2013. *Sambucus ebulus* L. fruit aqueous infusion modulates GCL and GPx4 gene expression. *Bulg. J. Agric. Sci.* 19 (2), 143–146.
- Thangudu, R.R., Manoharan, M., Srinivasan, N., Cadet, F., Sowdhamini, R., Offmann, B., 2008. Analysis on conservation of disulphide bonds and their structural features in homologous protein domain families. *BMC Struct. Biol.* 8 (1), 55.
- van Bentem, S.D.L.F., Hirt, H., 2009. Protein tyrosine phosphorylation in plants: more abundant than expected? *Trends Plant Sci.* 14 (2), 71–76.
- Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., Oliva, R., 2011. COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics* 27 (20), 2915–2916.
- Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., et al., 2014. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505 (7481), 117–120.
- Wang, Y.C., Peterson, S.E., Loring, J.F., 2014. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* 24 (2), 143.
- Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., et al., 2015. N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161 (6), 1388–1399.
- Wei, T., Chang, Ch., Xue, L., Jieling, Zh., Jie, L., 2018. CASTp 3.0: computed atlas of surface topography of proteins. *Nucl. Acids Res.* 46 (1), 363–367.
- Westwood, J.A., 1985. A guide to legendary Britain. Grafton Books, London, p. 103.
- Zhu, F., Yuan, S., Zhang, Z.W., Qian, K., Feng, J.G., Yang, Y.Z., 2016. Pokeweed antiviral protein (PAP) increases plant systemic resistance to Tobacco mosaic virus infection in *Nicotiana benthamiana*. *Eur. J. Plant Pathol.* 146 (3), 541–549.
- Zhu, F., Zhou, Y.K., Ji, Z.L., Chen, X.R., 2018. The plant ribosome-inactivating proteins play important roles in defense against pathogens and insect pest attacks. *Front. Plant Sci.* 9, 146.
- Zuppone, S., Fabbri, M.S., Vago, R., 2019. Hosts for hostile protein production: the challenge of recombinant immunotoxin expression. *Biomedicines* 7 (2), 38.