ASTHMA
ALLERGY
IMMUNOLOGY
ASTIM
ALLERJİ
İMMÜNOLOJİ

doi: 10.21911/aai.065
Asthma Allergy Immunol 2023;21:25-34

RESEARCH ARTICLE

Received: 27.07.2022 • Accepted: 15.11.2022
Online Published: 12.01.2023

Identification of Novel Endochitinase Class I Based Allergens

Levent CAVAS[1,2], Yesim YILMAZ-ABESKA[1]

[1]Department of Chemistry, Dokuz Eylül University, Graduate School of Natural and Applied Sciences, İzmir, Türkiye
[2]Department of Chemistry, Dokuz Eylül University, Faculty of Science, İzmir, Türkiye

Corresponding Author: Levent Cavas ✉ levent.cavas@deu.edu.tr

The study was presented as a poster presentation at the HIBIT'2020 Conference, Sabanci University, İstanbul, Turkey, 22-23 October 2020.

ABSTRACT

Objective: Bioinformatics analyses have been becoming important research tools in the field of life science since 2000. New tools have been developed for the detection of allergens in food sources. This study investigated potential allergen proteins similar to the endochitinase class 1(Pers a 1) protein found in avocado using bioinformatics analysis tools.

Materials and Methods: The potential allergens were examined with four different *in silico* tools: Clustal W for multiple sequence comparison, Swiss-MODEL for homology modeling, ProtParam for comparing protein parameters, and AlgPred for allergen prediction.

Results: *In silico* analyses revealed high rates of similarities in protein sequences among the studied species. The highest percent identity was found in *Cinnamomum micranthum f. kanehirae* as 97.55 while the lowest percent identity was found in *Diospyros kaki* as 70.23. It is very important to note that 9 of the 10 potential allergen species investigated are plats consumed in daily life and these species were identified as allergens. According to the algorithm of BLAST search on allergen representative peptides (ARPs), ARPs were found in eight potential allergen proteins except for *Prunus avium and Citrus unshiu*. The ARP sequence related to *Persea americana* is LQLTWNYNYGAAGNSIGFNGLSNP.

Conclusion: People who are allergic to avocado *(Persea americana)* may be advised to be careful while consuming foods such as *Phoenix dactylifera, Allium sativum, Citrus unshiu, Diospyros kaki,* and *Spinacia oleracea*. The present paper can be considered as a model study for understanding allergens in foods by using *in silico* tools.

Keywords: Allergenicity, avocado, bioinformatics, endochitinase, *in silico*

INTRODUCTION

Bioinformatics has developed after the genome project to evaluate the extensive data such as DNA and protein fast, efficiently and reliably through multidisciplinary work with the contributions of fields such as biochemistry, biology, computer science, statistics, and mathematics (1). With the developments in the bioinformatics data banks, these provide important information on the compounds that one wants to investigate, i.e. about the structure, function, nomenclature, taxonomy and physicochemical properties. Bioinformatics tools can be used to identify allergens from different sources and they can also be used to find the compounds that are associated with food allergens (2). Allergome and Allergen Nomenclature approved by the WHO and IUIS Allergen Nomenclature Commission are the databases where information about allergen proteins, including food allergens can be obtained (3-4). Bioinformatics analysis associated with the allergenicity of food components requires access to reliable and updated information about the structure and sequence of known allergens.

Latex fruit syndrome has been known for years and is described as someone with a latex allergy having hypersensitivity to some plant-foods (5). Allergen cross-reactivity has been studied with banana, avocado, kiwi, potato, chestnut, and several other plants (6-11). Avocado is a

ORCID Levent Cavas / 0000-0003-2136-6928, Yesim Yilmaz Abeska / 0000-0002-8105-921X

fruit that is highly cultivated and consumed in many countries, especially in the Dominican Republic, Mexico, and the USA (12, 13). It is rich in potassium; the B, E, and K vitamins; and fiber. Avocado also contains monounsaturated fats compared with other fruits (14). For this reason, the consumption of avocado is critical in countries where meat and fish consumption is limited. Previous studies have shown that consumption of avocado can regulate the level of cholesterol and potentially prevent cancer (15,16). At the same time, avocado is known to show an allergic reaction in some people. These allergic reactions can occur in two ways. The first one is local allergy syndrome and the second one is known as latex fruit syndrome and it is associated with latex allergy, and it also triggers an IgE-mediated immune response (17). It usually causes symptoms like urticaria, muscle pain, vomiting, and in some cases death. It is known that endochitinase class I (Pers a 1), a protein in avocado, causes allergic reactions (6).

In this study, an *in silico* methodology to find novel endochitinase-like proteins was studied using *in silico* tools. For this purpose, four *in silico* comparison tools were used and the potential allergens were examined with various parameters. Clustal W, Swiss-MODEL, ProtParam, and AlgPred were used for multiple sequence comparison, homology modeling, protein parameters, and allergenicity prediction, respectively.

MATERIALS and METHODS

The amino acid sequence of target allergen protein endochitinase class I found in *Persea americana* was obtained from the Allergen.org (ALLERGEN NOMENCLATURE WHO/IUIS Allergen Nomenclature Sub-Committee) database (4,18). The amino acid sequence of the allergen protein in allergen.org is accessed via uniprot.org. After retrieving the FASTA form and other functional information obtained from uniProt.org, BLAST (Basic Local Alignment Search Tool) found in the NCBI (National Center for Biotechnology Information) database was used for determining similar proteins (19,20). The maximum number of results was selected as 5000. Potential allergen proteins were chosen based on similarity index and E-value, and then further *in silico* analysis was performed. Various model fruits (Table I) were selected to see the relationship between percent identity and allergenicity.

Evaluation of Potential Allergen Proteins

Multiple sequence analysis and phylogenetic analysis of potential allergen proteins were performed via Clustal omega (21,22). The FASTA form of allergen protein (endochitinase class I found in *Persea americana*) and potential allergen proteins were compared, and preserved regions and similarity of sequences were studied (uniprot.org).

ProtParam tool in Expasy.ch was used for obtaining physical and chemical characteristics of protein sequence (23). Molecular weight, theoretical pI value, amino acid composition and percentages, atomical composition, and instability index can be found via this tool.

For the estimation of allergen proteins and screening IgE epitopes on allergen proteins, AlgPred (Prediction of

Table I: List of endochitinase like proteins (BLASTp)

Species	Name of protein (UniProt)	NCBI Reference Sequence	E-value	% Identity
Persea americana	endochitinase	CAB01591.1	0.00	100.00
Cinnamomum micranthum f. kanehirae	basic endochitinase-like protein	RWR79276.1	0.00	97.55
Phoenix dactylifera	chitinase 3-like	XP_026656006.2	2e-168	81.73
Spinacia oleracea	basic endochitinase	XP_021860442.1	9e-158	78.42
Prunus avium	Endochitinase 2-like	XP_021807867.1	2e-146	74.16
Citrus unshiu	hypothetical protein CUMW_194000	GAY59367.1	9e-152	75.49
Vigna unguiculata	endochitinase	XP_027904895.1	1e-146	73.83
Pisum sativum	RecName: Full=Endochitinase A2; Flags: Precursor	P21226.2	4e-140	75.68
Allium sativum	chitinase	AAA32640.1	2e-144	73.13
Raphanus sativus	PREDICTED: endochitinase CH25	XP_018490597.1	4e-147	70.86
Diospyros kaki	basic chitinase 1	ASL69226.1	8e-140	70.23

Allergen Proteins and Mapping of IgE Epitopes) was used (24). Mapping of IgE epitopes and PID, SVM module based on amino acid composition, SVM module based on the dipeptide composition, Blast search on allergen representative peptides (ARPs), and Prediction by Hybrid Approach tools were used for estimation of allergenicity. Swiss Model in Expasy.ch was used for homology modeling of potential proteins (25,26).

RESULTS

Endochitinase found in *Persea americana* with known allergenic effects were also investigated for other species via BLAST analysis. In this study, we investigated 10 potential allergen proteins by using the NCBI database (Table I). The highest percent identity was found in *Cinnamomum micranthum f. kanehirae* as 97.55 while the lowest percent identity was found in *Diospyros kaki* as 70.23. Percent identity was maximum when the E value was minimum. The E (Expect) value is a parameter that describes the expected number of hits when searching in a database with limits.

Clustal omega

The Clustal omega tool was used for analyzing similarity between potential allergen proteins. The similarity of 10 potential allergen proteins is shown in Figure 1A. According to the phylogenetic tree, there are three clusters. Endochitinase of *Persea americana,* basic endochitinase-like protein of *Cinnamomum micranthum f. kanehirae,* chitinase 3-like of *Phoenix dactylifera,* chitinase of *Allium sativum,* hypothetical protein CUMW_194000 of *Citrus unshiu*, basic chitinase 1 of *Diospyros kaki*, and basic endochitinase of *Spinacia oleracea* are located in the same cluster. Chitinase 3-like of *Phoenix dactylifera* is not described as an allergen but Profilin (Pho d 2) of *Phoenix dactylifera* has been mentioned as an allergen in allergen.org.

The first amino acid found in endochitinase class I protein is methionine (M). Amino acids at the position of 27–29, 31, 34, 37, 40, 42–44, 47, 49, 50, 55, 56, 60–64, 84, 88, 90–92, 96, 101–103, 106, 107, 110, 111, 114, 117, 121, 122, 127, 129, 131-133, 136-144, 148-153, 155-158, 162 (the numbers are according to *Persea americana*) were the same for all potential proteins and they were demonstrated with "*" in Figure 1B. Since these amino acids are the same in all selected proteins, it could be said that these regions are conserved.

The amino acids are A, M, D, I, R, W, F at the position of 35, 87, 93, 108, 128, 145, 159 in endochitinase of *Persea americana.* All proteins have the same amino acids at the 35th position except for *Phoenix dactylifera,* which has T at the same position. *Citrus unshiu* has L at position 87 but others have the same amino acid with endochitinase of *Persea americana*. *Spinacia oleracea, Pisum sativum, Raphanus sativus, Allium sativum* and *Prunus avium* have E, L, K, Y, and Y at the position of 93, 108, 128, 145, and 159. This similarity was shown as ":".

ProtParam

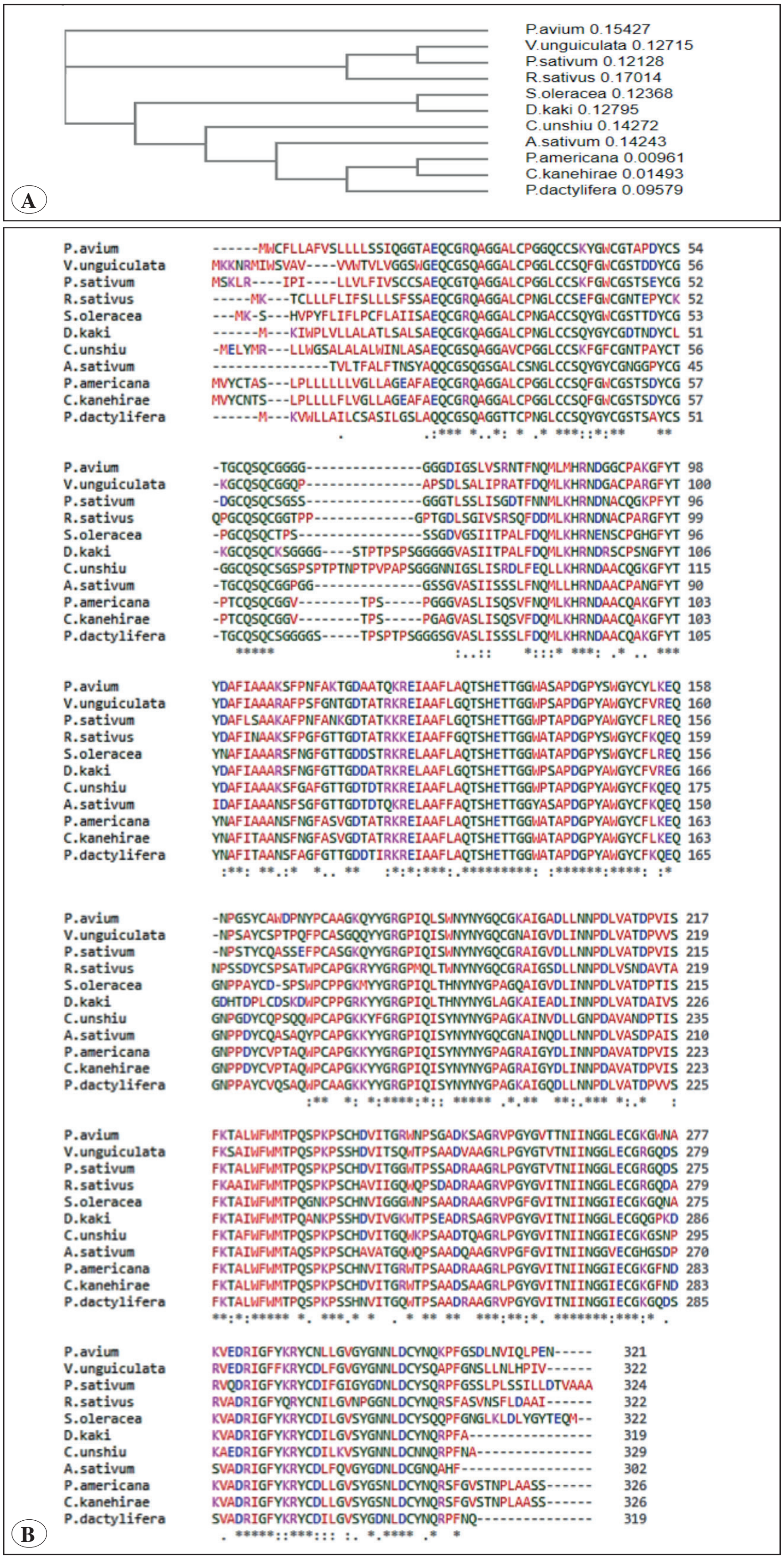
ProtParam was used for determining the physicochemical characteristics of proteins. There are 326 amino acids in endochitinase of *Persea americana* and basic endochitinase-like protein of *Cinnamomum micranthum f. kanehirae* (Table II). Maximum and minimum amino acid numbers are 329 and 302 in *Citrus unshiu* and *Allium sativum*, respectively. The pI value of endochitinase of *Persea americana* is 8.25 and this is the maximum pI value while the minimum pI value is the 5.18 for *Allium sativum*. Net charge of endochitinase of *Persea americana* was found as +4. The lowest net charge was found as -6, for *Allium sativum*. The instability index is an indicator of the stability of the protein in the test tube. The protein is stable if the stability index is below 40 and unstable if above it (27). The instability index of *Persea americana* was found as 37.44. The minimum value was found as 28.86, for *Allium sativum,* and the maximum value was found as 42.99, for *Spinacia oleracea*.

According to Table III, when sequences of potential allergen proteins were examined, it can be said that the most abundant amino acid found in endochitinase of *Persea americana* is glycine and this is valid for other selected allergen proteins. Other abundant amino acids found in endochitinase of *Persea americana* are alanine and serine. Potential allergen proteins have similar percentages of amino acids.

AlgPred

Six different prediction algorithms found in AlgPred were used for estimation of allergenicity. These are:

- Mapping of IgE epitopes and PID
- Support vector machine module based on amino acid composition
- Support vector machine based on dipeptide composition



```
P.avium         ------MWCFLLAFVSLLLLSSIQGGTAEQCGRQAGGALCPGGQCCSKYGWCGTAPDYCS 54
V.unguiculata   MKKNRMIWSVAV----VVWTVLVGGSWGEQCGSQAGGALCPGGLCCSQFGWCGSTDDYCG 56
P.sativum       MSKLR----IPI----LLVLFIVSCCSAEQCGTQAGGALCPGGLCCSKFGWCGSTSEYCG 52
R.sativus       -----MK---TCLLLFLIFSLLLSFSSAEQCGRQAGGALCPNGLCCSEFGWCGNTEPYCK 52
S.oleracea      ---MK-S---HVPYFLIFLPCFLAIISAEQCGRQAGGALCPNGACCSQYGWCGSTTDYCG 53
D.kaki          ------M---KIWPLVLLALATLSALSAEQCGKQAGGALCPNGLCCSQYGYCGDTNDYCL 51
C.unshiu        -MELYMR---LLWGSALALALWINLASAEQCGSQAGGAVCPGGLCCSKFGFCGNTPAYCT 56
A.sativum       ---------------TVLTFALFTNSYAQQCGSQGSGALCSNGLCCSQYGYCGNGGPYCG 45
P.americana     MVYCTAS---LPLLLLLLVGLLAGEAFAEQCGRQAGGALCPGGLCCSQFGWCGSTSDYCG 57
C.kanehirae     MVYCNTS---LPLLLLFLVGLLAGEAFAEQCGRQAGGALCPGGLCCSQFGWCGSTSDYCG 57
P.dactylifera   ------M---KVWLLAILCSASILGSLAQQCGSQAGGTTCPNGLCCSQYGYCGSTSAYCS 51
                .          .:*** *..*: * .* ***::*:**    **

P.avium         -TGCQSQCGGGG---------------GGGDIGSLVSRNTFNQMLMHRNDGGCPAKGFYT 98
V.unguiculata   -KGCQSQCGGQP---------------APSDLSALIPRATFDQMLKHRNDGACPARGFYT 100
P.sativum       -DGCQSQCSGSS---------------GGGTLSSLISGDTFNNMLKHRNDNACQGKPFYT 96
R.sativus       QPGCQSQCGGTPP-------------GPTGDLSGIVSRSQFDDMLKHRNDNACPARGFYT 99
S.oleracea      -PGCQSQCTPS----------------SSGDVGSIITPALFDQMLKHRNENSCPGHGFYT 96
D.kaki          -KGCQSQCKSGGGG----STPTPSPSGGGGGVASIITPALFDQMLKHRNDRSCPSNGFYT 106
C.unshiu        -GGCQSQCSGSPSPTPTNPTPVPAPSGGGNNIGSLISRDLFEQLLKHRNDAACQGKGFYT 115
A.sativum       -TGCQSQCGGPGG--------------GSSGVASIISSSLFNQMLLHRNDAACPANGFYT 90
P.americana     -PTCQSQCGGV--------TPS-----PGGGVASLISQSVFNQMLKHRNDAACQAKGFYT 103
C.kanehirae     -PTCQSQCGGV--------TPS-----PGAGVASLISQSVFDQMLKHRNDAACQAKGFYT 103
P.dactylifera   -TGCQSQCSGGGGS-----TPSPTPSGGGSGVASLISSSLFDQMLKHRNDAACQAKGFYT 105
                   *****                       :..::    *:::* ***: .* .. ***

P.avium         YDAFIAAAKSFPNFAKTGDAATQKREIAAFLAQTSHETTGGWASAPDGPYSWGYCYLKEQ 158
V.unguiculata   YDAFIAAARAFPSFGNTGDTATRKREIAAFLGQTSHETTGGWPSAPDGPYAWGYCFVREQ 160
P.sativum       YDAFLSAAKAFPNFANKGDTATKKREIAAFLGQTSHETTGGWPTAPDGPYAWGYCFLREQ 156
R.sativus       YDAFINAAKSFPGFGTTGDTATRKKEIAAFFGQTSHETTGGWATAPDGPYSWGYCFKQEQ 159
S.oleracea      YNAFIAAARSFNGFGTTGDDSTRKRELAAFLAQTSHETTGGWATAPDGPYSWGYCFLREQ 156
D.kaki          YDAFIAAARSFNGFGTTGDDATRKRELAAFLGQTSHETTGGWPSAPDGPYAWGYCFVREG 166
C.unshiu        YDAFIAAAKSFGAFGTTGDTDTRKREIAAFLAQTSHETTGGWPTAPDGPYAWGYCFKQEQ 175
A.sativum       IDAFIAAANSFSGFGTTGDTDTQKRELAAFFAQTSHETTGGYASAPDGPYAWGYCFKQEQ 150
P.americana     YNAFIAAANSFNGFASVGDTATRKREIAAFLAQTSHETTGGWATAPDGPYAWGYCFLKEQ 163
C.kanehirae     YNAFITAANSFNGFASVGDTATRKREIAAFLAQTSHETTGGWATAPDGPYAWGYCFLKEQ 163
P.dactylifera   YNAFITAANSFAGFGTTGDDTIRKREIAAFLAQTSHETTGGWATAPDGPYAWGYCFKQEQ 165
                 :**: **.:*  *.. **   :*:*:***:.*********: :******:****: :*

P.avium         -NPGSYCAWDPNYPCAAGKQYYGRGPIQLSWNYNYGQCGKAIGADLLNNPDLVATDPVIS 217
V.unguiculata   -NPSAYCSPTPQFPCASGQQYYGRGPIQISWNYNYGQCGNAIGVDLINNPDLVATDPVVS 219
P.sativum       -NPSTYCQASSEFPCASGKQYYGRGPIQISWNYNYGQCGRAIGVDLLNNPDLVATDPVIS 215
R.sativus       NPSSDYCSPSATWPCAPGKRYYGRGPMQLTWNYNYGQCGRAIGSDLLNNPDLVSNDAVTA 219
S.oleracea      GNPPAYCD-SPSWPCPPGKMYYGRGPIQLTHNYNYGPAGQAIGVDLINNPDLVATDPTIS 215
D.kaki          GDHTDPLCDSKDWPCPPGRKYYGRGPIQLTHNYNYGLAGKAIEADLINNPDLVATDAIVS 226
C.unshiu        GNPGDYCQPSQQWPCAPGKKYFGRGPIQISYNYNYGPAGKAINVDLLGNPDAVANDPTIS 235
A.sativum       GNPPDYCQASAQYPCAPGKKYYGRGPIQISYNYNYGQCGNAINQDLLNNPDLVASDPAIS 210
P.americana     GNPPDYCVPTAQWPCAPGKKYYGRGPIQISYNYNYGPAGRAIGYDLINNPDAVATDPVIS 223
C.kanehirae     GNPPDYCVPTAQWPCAPGKKYYGRGPIQISYNYNYGPAGRAIGYDLINNPDAVATDPVIS 223
P.dactylifera   GNPPAYCVQSAQWPCAAGKKYYGRGPIQISYNYNYGPAGKAIGQDLLNNPDLVATDPVVS 225
                            :**  *: *:****:*:: ***** .*.**  **:.*** *:.*   :

P.avium         FKTALWFWMTPQSPKPSCHDVITGRWNPSGADKSAGRVPGYGVTTNIINGGLECGKGWNA 277
V.unguiculata   FKSAIWFWMTPQSPKPSSHDVITSQWTPSAADVAAGRLPGYGTVTNIINGGLECGRGQDS 279
P.sativum       FKTALWFWMTPQSPKPSCHDVITGGWTPSSADRAAGRLPGYGTVTNIINGGLECGRGQDS 275
R.sativus       FKAAIWFWMTPQSPKPSCHAVIIGQWQPSDADRAAGRVPGYGVITNIINGGLECGRGQDA 279
S.oleracea      FKTAIWFWMTPQGNKPSCHNVIGGGWNPSAADRAAGRVPGFGVITNIINGGIECGKGQNA 275
D.kaki          FKTAIWFWMTPQANKPSSHDVIVGKWTPSEADRSAGRVPGYGVITNIINGGLECGQGPKD 286
C.unshiu        FKTAFWFWMTPQSPKPSCHDVITGQWKPSAADTQAGRLPGYGVITNIINGGIECGKGSNP 295
A.sativum       FKTAIWFWMTAQSPKPSCHAVATGQWQPSAADQAAGRVPGFGVITNIINGGVECGHGSDP 270
P.americana     FKTALWFWMTPQSPKPSCHNVITGRWTPSAADRAAGRLPGYGVITNIINGGIECGKGFND 283
C.kanehirae     FKTALWFWMTPQSPKPSCHDVITGRWTPSAADSAAGRLPGYGVITNIINGGIECGKGFND 283
P.dactylifera   FKTALWFWMTPQSPKPSSHNVITGQWTPSAADRAAGRVPGYGVITNIINGGIECGKGQDS 285
                **:*:***** *. ***.* *  . * ** **  ***:**:*. *******:***:* .

P.avium         KVEDRIGFYKRYCNLLGVGYGNNLDCYNQKPFGSDLNVIQLPEN-----     321
V.unguiculata   RVEDRIGFFKRYCDLFGVGYGNNLDCYSQAPFGNSLLNLHPIV------     322
P.sativum       RVQDRIGFYKRYCDIFGIGYGDNLDCYSQRPFGSSLPLSSILLDTVAAA     324
R.sativus       RVADRIGFYQRYCNILGVNPGGNLDCYNQRSFASVNSFLDAAI------     322
S.oleracea      KVADRIGFYKRYCDILGVSYGNNLDCYSQQPFGNGLKLDLYGYTEQM--     322
D.kaki          KVADRIGFYKRYCDILGVSYGNNLDCYNQRPFA----------------     319
C.unshiu        KAEDRIGFYKRYCDILKVSYGNNLDCNNQRPFNA---------------     329
A.sativum       SVADRIGFYKRYCDLFQVGYGDNLDCGNQAHF-----------------     302
P.americana     KVADRIGFYKRYCDLLGVSYGSNLDCYNQRSFGVSTNPLAASS------     326
C.kanehirae     KVADRIGFYKRYCDLLGVSYGSNLDCYNQRSFGVSTNPLAASS------     326
P.dactylifera   SVADRIGFYKRYCDILGVSYGDNLDCYNQRPFNQ---------------     319
                 . *****::***::: :. *.**** .*  *
```

Figure 1. A) Phylogenetic tree of the studied sequences. **B)** Multiple sequence alignment analysis of studied sequences (The name of the proteins in these species were given in Table 1).

Table II: Protein parameters for endochitinase like proteins in various plants (The name of the proteins in these species were given in Table 1).

Species	# aa	Mw (Kda)	Theoretical pI	Negatively Charged Residues (Asp+Glu)	Positively Charged Residues (Arg+Lys)	Net Charge	Instability Index
Persea americana	326	34586.04	8.25	19	23	4	37.44
Cinnamomum micranthum f. kanehirae	326	34639.99	7.36	21	22	1	35.87
Phoenix dactylifera	319	33641.52	8.09	18	21	3	39.42
Spinacia oleracea	322	34652.86	6.56	22	21	-1	42.99
Prunus avium	321	34381.58	7.77	22	24	2	35.67
Citrus unshiu	329	35135.39	8.08	23	26	3	41.41
Vigna unguiculata	322	34595.74	6.40	23	22	-1	35.30
Pisum sativum	324	34677.94	7.34	24	24	0	42.50
Allium sativum	302	31727.93	5.18	19	13	-6	28.86
Raphanus sativus	322	34762.94	7.35	24	25	1	39.16
Diospyros kaki	319	34066.19	6.90	28	28	0	36.01

Table III: Amino acid number and percentages in endochitinase like proteins from various plants (The name of the proteins in these species were given in Table 1)

	Persea americana		*Cinnamomum micranthum f. kanehirae*		*Phoenix dactylifera*		*Spinacia oleracea*		*Prunus avium*		*Citrus unshiu*	
	#	%	#	%	#	%	#	%	#	%	#	%
Ala (A)	36	11	35	10.7	32	10	26	8.1	28	8.7	32	9.7
Arg (R)	12	3.7	11	3.4	9	2.8	11	3.4	9	2.8	10	3
Asn (N)	17	5.2	16	4.9	15	4.7	19	5.9	20	6.2	20	6.1
Asp (D)	13	4	15	4.6	14	4.4	15	4.7	15	4.7	15	4.6
Cys (C)	17	5.2	17	5.2	16	5	17	5.3	18	5.6	16	4.9
Gln (Q)	14	4.3	14	4.3	20	6.3	15	4.7	16	5.0	17	5.2
Glu (E)	6	1.8	6	1.8	4	1.3	7	2.2	7	2.2	8	2.4
Gly (G)	41	13	40	12.3	42	13	43	13.4	47	14.6	42	12.8
His (H)	3	0.9	3	0.9	3	0.9	6	1.9	3	0.9	3	0.9
Ile (I)	14	4.3	14	4.3	16	5	18	5.6	12	3.7	16	4.9
Leu (L)	22	6.7	21	6.4	16	5	17	5.3	22	6.9	18	5.5
Lys (K)	11	3.4	11	3.4	12	3.8	10	3.1	15	4.7	16	4.9
Met(M)	3	0.9	3	0.9	3	0.9	5	1.6	4	1.2	3	0.9
Phe (F)	14	4.3	15	4.6	11	3.4	15	4.7	12	3.7	15	4.6
Pro (P)	22	6.7	22	6.7	19	6	24	7.5	19	5.9	27	8.2
Ser (S)	23	7.1	24	7.4	30	9.4	21	6.5	20	6.2	21	6.4
Thr (T)	19	5.8	20	6.1	21	6.6	19	5.9	16	5	20	6.1
Trp (W)	7	2.1	7	2.1	7	2.2	7	2.2	10	3.1	8	2.4
Tyr (Y)	18	5.5	18	5.5	18	5.6	18	5.6	18	5.6	15	4.6
Val (V)	14	4.3	14	4.3	11	3.4	9	2.8	10	3.1	7	2.1
Pyl (O)	0	0	0	0	0	0	0	0	0	0	0	0
Sec (U)	0	0	0	0	0	0	0	0	0	0	0	0

Table III continue

	Pisum sativum		*Vigna unguiculata*		*Allium sativum*		*Raphanus sativus*		*Diospyros kaki*	
	#	%	#	%	#	%	#	%	#	%
Ala (A)	25	7.7	28	8.7	33	10.9	29	9	28	8.8
Arg (R)	13	4	14	4.3	6	2	15	4.7	13	4.1
Asn (N)	14	4.3	15	4.7	18	6	18	5.6	15	4.7
Asp (D)	17	5.2	17	5.3	15	5	17	5.3	21	6.6
Cys (C)	19	5.9	16	5	17	5.6	18	5.6	15	4.7
Gln (Q)	15	4.6	18	5.6	22	7.3	17	5.3	11	3.4
Glu (E)	7	2.2	6	1.9	4	1.3	7	2.2	7	2.2
Gly (G)	41	13	41	13	42	13.9	39	12.1	43	13.5
His (H)	3	0.9	4	1.2	5	1.7	3	0.9	5	1.6
Ile (I)	16	4.9	14	4.3	13	4.3	14	4.3	15	4.7
Leu (L)	23	7.1	16	5	13	4.3	19	5.9	21	6.6
Lys (K)	12	3.7	8	2.5	7	2.3	10	3.1	15	4.7
Met(M)	3	0.9	4	1.2	2	0.7	4	1.2	3	0.9
Phe (F)	15	4.6	15	4.7	16	5.3	17	5.3	11	3.4
Pro (P)	19	5.9	23	7.1	17	5.6	22	6.8	22	6.9
Ser (S)	31	9.6	25	7.8	25	8.3	24	7.5	21	6.6
Thr (T)	20	6.2	17	5.3	16	5.3	18	5.6	19	6
Trp (W)	7	2.2	10	3.1	4	1.3	8	2.5	7	2.2
Tyr (Y)	15	4.6	14	4.3	18	6	14	4.3	16	5
Val (V)	9	2.8	17	5.3	9	3	9	2.8	11	3.4
Pyl (O)	0	0	0	0	0	0	0	0	0	0
Sec (U)	0	0	0	0	0	0	0	0	0	0

- Blast search on allergen representative peptides (ARPs)
- MEME/MAST motif
- Prediction by Hybrid Approach

The algorithm of Mapping of IgE epitopes and PID predicted 3 of 10 proteins as an allergen and these are *Cinnamomum micranthum f. kanehirae*, *Spinacia oleracea* and *Vigna unguiculata*. Three allergen proteins have the same IgE epitope as SQWGWC as *Persea americana*. PID values of three allergen proteins are the same as *Persea americana* (Table IVA). According to the algorithm of Blast search on allergen representative peptides (ARPs), ARPs were found in eight potential allergen proteins except for *Prunus avium* and *Citrus unshiu*. The ARP sequence related to *Persea americana* is LQLTWNYNYGAAGNSIGFNGLSNP and it is the same for *Cinnamomum micranthum f. kanehirae* and *Vigna unguiculata*. According to the results of the algorithm of the SVM module based on amino acid composition, SVM module based on dipeptide composition, and Prediction by Hybrid Approach, all proteins being studied are allergens (Table IV A,B). The algorithm of MEME/MAST algorithm predicted 10 proteins as non-allergens. When all algorithms' results are considered, all proteins being studied are predicted as allergens.

Swiss-MODEL

Homology modeling is performed using Swiss-MODEL. The 3D structure of proteins is important since it is linked with function. Also, 3D structure may be used for explanation of allergenicity of endochitinase class I like proteins. 3D structures of some of the proteins are seen in Figure 2. Swiss-MODEL uses proteins whose 3D structures are already known to create homology models. Four different templates were used by Swiss-MODEL tool for modeling potential allergen proteins. As a result, the 3D structure of potential proteins was found to be nearly the same (Supplementary material Table I).

Table IV: AlgPred analysis of different allergen endochitinase like proteins with a) Mapping of IgE epitopes and PID, Blast search on allergen representative peptides (ARPs), Prediction by Hybrid Approach algorithms b) SVM module based on amino acid composition, SVM module based on dipeptide composition and MEME/MAST motif (The name of the proteins in these species were given in Table 1)

A)

AlgPred Parameters	Mapping of IgE epitopes and PID					Blast search on allergen representative peptides (ARPs)	Prediction by Hybrid Approach
Species	Allergen prediction	IgE epitope	Sequence matched	Position	PID		
Persea americana	√	SQWGWC	SQFGWC	41	83.3	LQLTWNYNYGAAGNSIGFNGLSNP	√
Cinnamomum micranthum f. kanehirae	√	SQWGWC	SQFGWC	44	83.3	LQLTWNYNYGAAGNSIGFNGLSNP	√
Phoenix dactylifera	X					GVTPSP-----GGGVASLISQSVFNQMLK	√
Spinacia oleracea	√	SQWGWC	SQYGWC	40	83.3	SAAKAFPAFGTTGDVDTCKREIAA	√
Prunus avium	X						√
Citrus unshiu	X						√
Vigna unguiculata	√	SQWGWC	SQFGWC	43	83.3	LQLTWNYNYGAAGNSIGFNGLSNP	√
Pisum sativum	X					SAAKAFPAFGTTGDVDTCKREIAA	√
Allium sativum	X					GRQAGGKLCPNNLCCSQWGWCGS	√
Raphanus sativus	X					LISASLFDQMLKYRNDPRCKSNGF	√
Diospyros kaki	X					GGVASLISQSVFNQMLKHRNDAAC	√

B)

AlgPred Parameters	SVM module based on amino acid composition				SVM module based on dipeptide composition				MEME/ MAST motif
Species	Allergen prediction	Score	Positive predictive value (%)	Negative predictive value (%)	Allergen prediction	Score	Positive predictive value (%)	Negative predictive value (%)	
Persea americana	√	0.79200327	87.05	71.53	√	0.99798747	100	59.74	x
Cinnamomum micranthum f. kanehirae	√	0.88663897	85.64	67.96	√	1.1388986	100	59.74	x
Phoenix dactylifera	√	1.0673486	85.64	67.96	√	1.0628199	100	59.74	x
Spinacia oleracea	√	1.0022276	85.64	67.96	√	0.94676503	100	59.74	x
Prunus avium	√	1.1683693	85.64	67.96	√	1.2215008	100	59.74	x
Citrus unshiu	√	1.0116458	85.64	67.96	√	1.074682	100	59.74	x
Vigna unguiculata	√	0.45049336	81.83	74.03	√	0.52013055	86.55	66.47	x
Pisum sativum	√	0.43198172	81.83	74.03	√	0.72480486	82.97	62.4	x
Allium sativum	√	1.210904	85.64	67.96	√	1.4961617	100	59.74	x
Raphanus sativus	√	0.50937095	81.83	74.03	√	0.53714987	86.55	66.47	x
Diospyros kaki	√	1.3516591	85.64	67.96	√	0.92057654	100	59.74	x

DISCUSSION

Since the discovery of the latex fruit syndrome, which triggers the IgE-mediated immune response, many new plants have been associated with this condition. When these plants are consumed in daily life, they are not known to cause an allergic reaction but can be life-threatening. For this reason, studies on this subject are increasing day by day. With the developments in bioinformatics and data

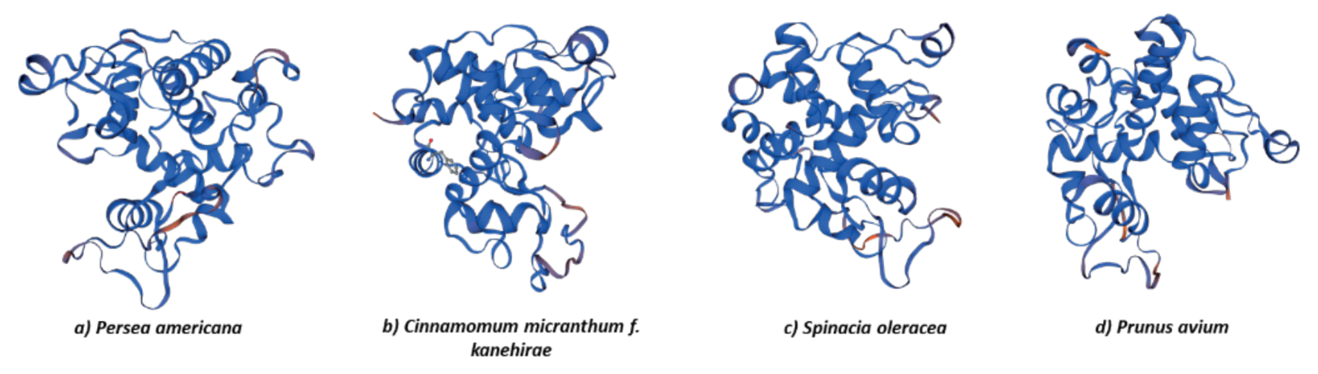


Figure 2. 3D visualization of each protein by Swiss-MODEL.

banks, the number of *in silico* studies has also increased. In our study, under the light of previous research and using *in silico* tools, we studied some plants that may be related to the latex fruit syndrome. We examined plants containing protein structures similar to endochitinase class I protein found in avocado. Sowka et. al. mentioned that avocado is an allergen source and they described this as latex fruit syndrome due to cross-sensitization between avocado and latex proteins (28). They identified Pers 1, a 32-kDa major avocado allergen recognized by 15 of 15 patients with avocado and/or latex allergies and purified the Pers a 1 protein. They detected two tryptic peptide sequences at the N-terminal of the Pers a 1 protein. They expressed Pers a 1 in 50 mg/liter culture medium in *Pichia pastoris* yeast. Recombinant Pers a 1 had endochitinase activity. They determined that *Fusarium oxysporum* has IgE binding capacity due to its inhibitory effect on hyphae. They concluded that latex proteins containing an allergen of 20 kDa, probably prohevein, provided explanation for cross-sensitivity with avocado proteins. Based on sequence comparison, Pers a 1 and pro-hevein had 70% similarity in the binding sites of the chitin. They demonstrated that mature Pers a 1 proteins have a conserved region of 43 amino acids at the N-terminal and this is an area capable of binding and recognizing the chitin. They showed that this domain exists in Endochitinase class 1 protein.

Some proteins found in the structure of plants and used to cope with environmental stress may have allergic effects when consumed by other living things (29). Considering the studies on plant-derived food allergens, some protein classes stand out. Class I chitinases, which are allergens homologous to PR-3 (pathogenesis-related) type proteins, are highly abundant proteins found in seed-producing plants (30). Most plant chitinases are endochitinases that can hydrolytically break down chitin, and chitin is found in the skeleton of insects that can damage the plant (31, 32). There are many studies in the literature on class I chitinase in avocado, chestnut, and banana, which are known to be allergens.

Blanco et al. researched class 1 chitinase causing the latex fruit syndrome (17). Class I chitinase found in chestnuts revealed positive SPT (skin prick test) responses in 13 of 18 (72%) patients with latex-fruit allergy, besides class I chitinase in avocado was positive in 18 (67%) of similar allergic patients. SPT (skin prick test) responses were positive in 13 of 18 (72%) patients with latex-fruit allergy, besides class I chitinase in avocado was positive in 18 (67%) of similar-allergic patients. However, class II enzymes have no SPT responses and class II enzymes do not contain hevein-like domains. As a result, they stated that class I chitinase found in chestnut and avocado with N-terminal hevein-like domains are cross-reacting major allergens. Therefore, according to the results of Blanco et al. hevein-like domains are probably pan-allergens responsible for the latex-fruit syndrome.

Posch et al. mentioned that endochitinase class 1 protein containing hevein-like domains causes avocado allergy. 4.7 kDa chitin-binding domain, hevein, is the protein that causes latex allergy and belongs to the pathogen-related protein class (6). The 43 amino acid conserved region found in hevein , which is capable of binding N-acetyl glucosamine, can be found in lectin and Endochitinase class 1 protein which induced with wound formation in plants. In 80% (n = 15) of avocado-sensitive individuals, IgE antibodies directed against 31-kDa allergen were detected by immunoblotting. As a result, Posch et al. stated that the main mechanism underlying avocado allergy is hevein-like domains in endochitinase class 1 protein (6). It is very interesting to note that Karisola et al. showed that purified hevein like domains causes IgE-mediated reactions in the latex-fruit syndrome. Moreover, they underline that 31-kd endochitinases do not show any reactivity in this process (33).

In this study, based on endochitinase class I protein found in *Persea americana*, which is known to be an allergen in the literature, plants that may have an allergen like Pers a 1 were investigated. Using the NCBI database, the similarity rates between the sequences of the endochitinase proteins were used in the selection of plants. Different plant species have been researched from previous studies and new results have been obtained that can contribute to the literature. According to the results of our study, 8 of 10 species we researched do not have any records on allergen.org regarding any profilin or chitinase associated with latex fruit allergy, and our results make up for this deficiency. The results we obtained from AlgPred also contribute to evaluating these plants as allergens.

CONCLUSION

In conclusion, 10 potential endochitinase class I like proteins were investigated using five different *in silico* tools. Physicochemical characteristics, allergenicity, and 3D homology of investigated allergen proteins were evaluated. Profilin, chitinase or endochitinase records were not found for *Cinnamomum micranthum f. kanehirae*, *Spinacia oleracea*, *Citrus unshiu*, *Vigna unguiculata*, *Pisum sativum*, *Allium sativum*, *Raphanus sativus*, and *Diospyros kaki*. On the other hand, profilin was found in allergen.org for *Phoenix dactylifera* and *Prunus avium*. According to the results of six prediction algorithms of AlgPred, *Cinnamomum micranthum f. kanehirae, Spinacia oleracea,* and *Vigna unguiculata* were concluded to be allergens based on five of six allergen algorithms.

Disclosure of Interest

None.

Acknowledgments

Financial support was provided by Dokuz Eylül University - Department of Scientific Research Projects with the project number 2019.KB.FEN.044.

Authorship Contributions

Concept: **Levent Cavas,** Design: **Levent Cavas,** Data Collection or processing: **Levent Cavas, Yesim Yilmaz Abeska,** Analysis or Interpretation: **Levent Cavas, Yesim Yilmaz Abeska,** Literature search: **Levent Cavas, Yesim Yilmaz Abeska,** Writing: **Levent Cavas, Yesim Yilmaz Abeska,** Approval: **Levent Cavas, Yesim Yilmaz Abeska.**

REFERENCES

1. Tu M, Cheng S, Lu W, Du M. Advancement and prospects of bioinformatics analysis for studying bioactive peptides from food-derived protein: Sequence, structure, and functions. TrAC Trends in Analytical Chemistry 2018;105:7-17.
2. Hayes M, Rougé P, Barre A, Herouet-Guicheney C, Roggen EL. In silico tools for exploring potential human allergy to proteins. Drug Discov Today Dis Models 2015;17:3-11.
3. Mari A, Rasi C, Palazzo P, Scala E. Allergen databases: Current status and perspectives. Current Allergy and Asthma Reports 2009;9(5):376-83.
4. Pomés A, Davies JM, Gadermaier G, Hilger C, Holzhauser T, Lidholm J, et al. WHO/IUIS allergen nomenclature: Providing a common language. Mol Immunol 2018;100:3-13.
5. Ricci G, Piccinno V, Calamelli E, Giannetti A, Pession A. Latex-fruit syndrome in Italian children and adolescents with natural rubber latex allergy. Int J Immunopathol Pharmacol 2013;26(1):263-8.
6. Posch A, Wheeler CH, Chen Z, Flagge A, Dunn MJ, Papenfuss F, et al. Class I endochitinase containing a hevein domain is the causative allergen in latex-associated avocado allergy. Clin Exp Allergy 1999;29(5):667-72.
7. Berthelot K, Peruch F, Lecomte S. Highlights on *Hevea brasiliensis* (pro) hevein proteins. Biochimie 2016;127:258-70.
8. Igetei JE, El-Faham M, Liddell S, Doenhoff MJ. Antigenic cross-reactivity between *Schistosoma mansoni* and peanut: A role for cross-reactive carbohydrate determinants (CCDs) and implications for the hygiene hypothesis. Immunology 2017;150(4):506-17.
9. Parisi CAS, Kelly KJ, Ansotegui IJ, Gonzalez-Díaz SN, Bilò MB, Cardona V, et al. Update on latex allergy: New insights into an old problem. World Allergy Organ J 2021;14(8):100569.
10. Ukleja-Sokołowska N, Zacniewski R, Lis K, Żbikowska-Gotz M, Kuźmiński A, Bartuzi Z. Exercise induced anaphylaxis in kiwi allergic patient: case report. Allergy Asthma Clin Immunol 2021;17(1):91.
11. Vojdani A, Lerner A, Vojdani E. Cross-reactivity and sequence homology between al-pha-synuclein and food products: A step further for parkinson's disease synucleinopathy. Cells 2021;10(5):1111.
12. Tan CX. Virgin avocado oil: An emerging source of functional fruit oil. Journal of Functional Foods 2019;54(26):381-92.
13. Martín-Torres S, Jiménez-Carvelo AM, González-Casado A, Cuadros-Rodríguez L. Authentication of the geographical origin and the botanical variety of avocados using liquid chromatography fingerprinting and deep learning methods. Chemometrics and Intelligent Laboratory Systems 2020;199: 103960.

14. Fulgoni VL 3rd, Dreher M, Davenport AJ. Avocado consumption is associated with better diet quality and nutrient intake, and lower metabolic syndrome risk in US adults: results from the National Health and Nutrition Examination Survey (NHANES) 2001-2008. Nutr J 2013;12:1.
15. Dreher ML, Davenport AJ. Hass avocado composition and potential health effects. Crit Rev Food Sci Nutr 2013;53(7):738-50.
16. Bhuyan DJ, Alsherbiny MA, Perera S, Low M, Basu A, Devi OA, et al. The odyssey of bioactive compounds in avocado (*Persea americana*) and their health benefits. Antioxidants 2019; 8(10): 1-53.
17. Blanco C, Carrillo T, Castillo R, Quiralte J, Cuevas M. Avocado hypersensitivity. Allergy 1994;49(6):454-9.
18. Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J, et al. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. Allergy 2014;69:413-9.
19. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: A more efficient report with usability improvements. Nucleic Acids Res 2013;41(W1): W29-33.
20. Madden TL, Busby B, Ye J. Reply to the paper: Misunderstood parameters of NCBI BLAST impacts the correctness of bioinformatics workflows. Bioinformatics 2018;35(15):2699-700.
21. Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Bio. 2011;7:539.
22. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Science 2018; 27:135-45.
23. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server, in Walker, J.M. (Ed), The Proteomics Protocols Handbook, Humana Press, 2005; 571-607.
24. Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 2006;34(Web Server issue):W202-9.
25. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep 2017;7(1):10480.
26. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46(W1):W296-303.
27. Guruprasad K, Reddy BVB, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 1990;4(2):155-61.
28. Sowka S, Hsieh LS, Krebitz M, Akasawa A, Martin BM, Starrett D, et al. Identification and cloning of Prs a 1, a 32-kDa endochitinase and major allergen of avocado, and its expression in the yeast *Pichia pastoris*. J Biol Chem 1998;273(43):28091-7.
29. Breiteneder H, Ebner C. Molecular and biochemical classification of plant-derived food allergens. J Allergy Clin Immunol 2000;106(1 Pt 1):27-36.
30. Graham LS, Sticklen MB. Plant chitinases. Canadian Journal of Botany 1994;72(8):1057-83.
31. Bishop JG, Dean AM, Mitchell-Olds T. Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. Proc Natl Acad Sci U S A 2000;97(10):5322-7.
32. Kasprzewska A. Plant chitinases--regulation and function. Cell Mol Biol Lett 2003;8(3):809-24.
33. Karisola P, Kotovuori A, Poikonen S, Niskanen E, Kalkkinen N, Turjanmaa K, et al. Isolated hevein-like domains, but not 31-kd endochitinases, are responsible for IgE-mediated in vitro and in vivo reactions in latex-fruit syndrome. J Allergy Clin Immunol 2005;115(3):598-605.