

Research Article

Identification of putative and potential cross-reactive chickpea (*Cicer arietinum*) allergens through an *in silico* approach



Anuja Kulkarni^a, Laxmi Ananthanarayan^{a,*}, Karthik Raman^b

^a Food Engineering & Technology Department, Institute of Chemical Technology, Mumbai, India[†]

^b Indian Institute of Technology (IIT) Madras, Chennai 600036, India

ARTICLE INFO

Article history:

Received 27 April 2013

Received in revised form 11 July 2013

Accepted 12 August 2013

Keywords:

Chickpea

Allergens

Allergen prediction

Bioinformatics

Proteomics

ABSTRACT

Background: Allergy has become a key cause of morbidity worldwide. Although many legumes (plants in the *Fabaceae* family) are healthy foods, they may have a number of allergenic proteins. A number of allergens have been identified and characterized in *Fabaceae* family, such as soybean and peanut, on the basis of biochemical and molecular biological approaches. However, our understanding of the allergens from chickpea (*Cicer arietinum* L.), belonging to this family, is very limited.

Objective: In this study, we aimed to identify putative and cross-reactive allergens from Chickpea (*C. arietinum*) by means of *in silico* analysis of the chickpea protein sequences and allergens sequences from *Fabaceae* family.

Methods: We retrieved known allergen sequences in *Fabaceae* family from the IUIS Allergen Nomenclature Database. We performed a protein BLAST (BLASTp) on these sequences to retrieve the similar sequences from chickpea. We further analyzed the retrieved chickpea sequences using a combination of *in silico* tools, to assess them for their allergenicity potential. Following this, we built structure models using FUGUE: Sequence-structure homology; these models generated by the recognition tool were viewed in Swiss-PDB viewer.

Results: Through this *in silico* approach, we identified seven novel putative allergens from chickpea proteome sequences on the basis of similarity of sequence, structure and physicochemical properties with the known reported legume allergens. Four out of seven putative allergens may also show cross reactivity with reported allergens since potential allergens had common sequence and structural features with the reported allergens.

Conclusion: The *in silico* proteomic identification of the allergen proteins in chickpea provides a basis for future research on developing hypoallergenic foods containing chickpea. Such bioinformatics approaches, combined with experimental methodology, will help delineate an efficient and comprehensive approach to assess allergenicity and pave the way for a better understanding of the biological and medical basis of the same.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

A generally accepted definition of allergy is that it constitutes a hypersensitivity reaction where symptoms appear rapidly and are caused by exposure to exogenous macromolecules, also known as antigens or allergens. This reaction is most commonly mediated by

a specific class of antibody, known as immunoglobulin E (IgE). It has been estimated that around 1–2% of the population, and up to 8% of children suffer from some type of food allergy (Mills et al., 2003). Majority of food allergies are caused by only around eight types of foods (Mills et al., 2003), including a number of foods of plant origin such as peanuts, tree nuts, wheat and soya. Although our understanding of plant food allergens has considerably advanced in the last decade (Kumar et al., 2012), there are still major gaps in our understanding of certain very important questions. For example, it is critical to understand as to what makes certain plant foods more allergenic than others, if we are to develop knowledge-based means of assessing the allergenic potential of novel foods as well as processing strategies to reduce allergen loads (Mills et al., 2003).

Bioinformatics tools available for database searching and biological sequence analysis enable the quick identification of sequences of interest and provide bibliographic, taxonomic and

Abbreviations: BLAST, Basic Local Alignment Search Tool; BLASTp, protein–protein BLAST; IgE, immunoglobulin E; WHO, World Health Organization; IUIS, International Union of Immunological Societies; FAO, Food and Agriculture Organization; Pfam, protein families' database.

* Corresponding author. Tel.: +91 22 33612506; fax: +91 22 33611020.

E-mail addresses: l.ananthanarayan@ictmumbai.edu.in, laxmi.ananth.iyer@gmail.com (L. Ananthanarayan).

[†] University Under Section 3 of UGC Act – 1956, Elite Status & Centre of Excellence – Government of Maharashtra, India.

feature information. Tools for sequence comparison, motif searching or sequence profiling assist researchers to identify biologically relevant sequence similarities and features (Petrovsky and Brusica, 2002).

Various tests to predict potential allergenicity like bioinformatics, *in vitro* digestibility of the protein, and binding to antisera of allergic patients make allergen prediction an intricate issue. The Codex alimentarius, which has established guidelines to assess potential allergenicity of proteins using bioinformatics in a step-by-step procedure, recommends that potential allergenicity should be considered when there is either:

- (a) More than 35% similarity over a window of 80 amino acids in the amino acid sequence of the query protein (without the leader sequence, if any) with an entry known as allergen, or
- (b) A stretch of identity of 6–8 contiguous amino acids (FAO and WHO, 2001, 2003).

Physico-chemical properties of the protein such as lipid binding, ionic charge and post translational modifications may also be important in eliciting the allergenicity. Hence, we could assume that potent allergens would possess both such physicochemical properties, as well as relevant structural epitopes; while non-allergens would lack either (Singh et al., 2009).

In this study, by means of *in silico* analyses of the proteome sequences, we have performed a proteome-wide analysis of allergens from *Fabaceae* families, with the aim of identifying putative and cross-reactive allergens from Chickpea (*Cicer arietinum*). Our *in silico* approaches can act as groundwork to perform a preliminary screening of allergens and also provide a foundation for further functional analysis of the allergens.

2. Methods

2.1. Sequence retrieval

All databases and software used in these studies are publicly available on the world-wide web. We retrieved the primary sequences of known allergens of *Fabaceae* family from IUIS Allergen Nomenclature Sub-Committee (IUIS Allergen Nomenclature Home Page), a publicly available database. Further, we performed a BLASTp search on the retrieved sequences with default gap penalties, an expectation value threshold of 1.0 and the default score matrix, so that depending upon the length of the sequence the appropriate matrix can be selected. Further, by applying the filter of *C. arietinum* species, seven similar sequences from *C. arietinum* were selected and the sequences were acquired from UniProt (UniProt Consortium, 2012). This provided data required to predict the primary structure (sequence) and to perform the allergenicity assessment study.

2.2. Sequence analysis for potential allergenicity of proteins

For prediction of the potential allergenicity of proteins according to the current recommendations of the FAO/WHO Expert Consultation, as outlined in the Codex alimentarius, we used two web-based tools, Allermatch™ (Allermatch: Allermatch, 2012) and AllergenOnline (Kim et al., 2009). Following a BLASTp search, we retrieved the resulting sequence alignment for each of the seven similar *C. arietinum* sequences. We used another tool, AlgPred (Saha and Raghava, 2006), for the prediction of allergenicity on the basis of parameters like mapping of IgE epitopes, mapping of motifs (sequence pattern that occurs repeatedly in a group of related protein sequences) by MEME (Multiple Em for Motif Elicitation)/MAST (Motif Alignment and Search Tool) search tools, SVM (support

vector machine) module based on amino acid composition, SVM module based on dipeptide composition, and BLAST search on allergen representative peptides (ARPs).

2.3. Physicochemical analysis

We used the PROTPARAM tool (Gasteiger et al., 2005) for physicochemical analysis of the sequences, namely, parameters such as number of amino acids, molecular weight, theoretical pI, total number of negatively charged residues (Asp + Glu), total number of positively charged residues (Arg + Lys), net charge, instability index.

2.4. Structural analysis

We used Fugue: Sequence-Structure Homology recognition tool (Shi et al., 2001) for structural analysis for allergenicity prediction; the model structure retrieved by this tool was visualized in the Swiss-PDB viewer, available freely (Guex and Peitsch, 1997).

3. Results and discussion

Availability of the protein sequences of chickpea has made it possible for the first time to identify the putative allergen in this plant species. Although it has been reported clinically that chickpea proteins do have allergens (Patil et al., 2001), no chickpea allergens have been listed by IUIS Allergen Nomenclature Sub-Committee as shown in Table 1. Therefore, an attempt was made on the identification of possible allergen in chickpea in this *in silico* approach. In order to identify the allergens, known allergen sequences of *Fabaceae* family (Table 1) were analyzed by BLASTP against protein sequences of chickpea. This analysis resulted in seven putative allergens from chickpea as described in Table 2. The seven putative allergens from chickpea were UniProtKB/TrEMBL ID: Q9SMK8, UniProtKB/TrEMBL ID: Q39450, UniProt/Swiss-Prot ID: Q9SMJ4, UniProtKB/TrEMBL ID: Q304D4, UniProtKB/TrEMBL ID: G1K3R9, UniProtKB/TrEMBL ID: G1K3S0, UniProt/Swiss-Prot ID: Q23758. Each protein is described in more detail below.

3.1. Q9SMK8 and Q39450

On the BLASTP analysis (Table 2) of the Ara h 8, Gly m 4 and Vig r 1 allergens, which are pathogenesis related proteins (PR proteins) from peanut, soybean and mung bean respectively, we retrieved Q9SMK8 and Q39450 as the similar sequences from chickpea. As shown in Table 2, both these proteins too belong to same PR group of proteins. Many plant-derived food allergens can be found within this group of proteins. Characteristics like small size, stability at low pH, and resistance to proteolysis, make PR proteins good candidates for provoking an immune response in predisposed individuals (Ebner et al., 2001; Hoffmann-Sommergruber, 2002). The Pfam ID's in Table 2 shows that these chickpea proteins consist of the cupin domain which have been identified in major plant food allergens like 7S globulins of soybean (P-conglycinin), peanut (conarachin; Ara h 1) etc. (Breiteneder and Mills, 2005; Hauser et al., 2008; Mills et al., 2002).

Sequence analysis by Allermatch™, as shown in Fig. 1a and b, indicated that both the sequences showed high similarity against pollen allergen Bet v 1 from *Betula pendula* (European white birch) (*Betula verrucosa*) with 100% hits identical and thus satisfying the criteria of Codex alimentarius (FAO and WHO, 2001, 2003) for being potential allergens. Also by AllergenOnline database, Q9SMK8 showed similarity with 153 allergen sequences while Q39450 showed similarity with 151 allergen sequences. Bet v 1 is again another PR protein. By considering Bet v 1 as the template and structural analysis by FUGUE: Sequence-structure homology

Table 1Lists known allergen sequences of *Fabaceae* family from IUIS Allergen Nomenclature Sub-Committee with their ID's followed by biochemical details [www.allergen.org/].

Species	Allergen	Biochemical name	MW (kDa)	UniProt/Swiss-Prot ID or UniProtKB/TrEMBL ID	Pfam ID
<i>Arachis hypogaea</i> (Peanut)	Ara h 1	Cupin (Vicillin-type, 7S globulin)	64	P43238 ^a	PF00190
<i>Arachis hypogaea</i> (Peanut)	Ara h 2	Conglutin (2S albumin)	17	Q6PSU2 ^a , Q6PSU2-2 ^a , Q6PSU2-3 ^a , Q6PSU2-4 ^a	PF00234
<i>Arachis hypogaea</i> (Peanut)	Ara h 3	Cupin (Legumin-type, 11S globulin, Glycinin)	60, 37 (fragment)	O82580	PF00190
<i>Arachis hypogaea</i> (Peanut)	Ara h 4	Nutrient reservoir activity	60, 37	Q9SQH7	PF00190
<i>Arachis hypogaea</i> (Peanut)	Ara h 5	Profilin	15	Q9SQI9 ^a	PF00235
<i>Arachis hypogaea</i> (Peanut)	Ara h 6	Conglutin (2S albumin)	15	Q647G9 ^a	PF00234
<i>Arachis hypogaea</i> (Peanut)	Ara h 7	Conglutin (2S albumin)	15	Q9SQH1, B4XID4	PF00234
<i>Arachis hypogaea</i> (Peanut)	Ara h 8	Pathogenesis-related protein, PR-10	17	Q6VT83, B0YIU5	PF00407
<i>Arachis hypogaea</i> (Peanut)	Ara h 9	Nonspecific lipid-transfer protein 1	9.8	B6CEX8, B6CG41	PF00234
<i>Arachis hypogaea</i> (Peanut)	Ara h 10	16 kDa oleosin	16	Q647G5, Q647G4	PF01277
<i>Glycine max</i> (Soybean)	Gly m 1	Hydrophobic protein from soybean	7	Q9S8F2, Q9S8F3	–
<i>Glycine max</i> (Soybean)	Gly m 3	Profilin	14	O65809 ^a , O65810 ^a	PF00235
<i>Glycine max</i> (Soybean)	Gly m 4	PR-10 protein	17	P26987 ^a	PF00407
<i>Glycine max</i> (Soybean)	Gly m 5	Beta-conglycinin (vicilin, 7S globulin)	–	O22120, Q9FZP9, P25974 ^a	PF00190
<i>Glycine max</i> (Soybean)	Gly m 6	Glycinin (legumin, 11S globulin)	–	P04776 ^a , P04405 ^a , Q7GC77, Q9SB11 Q702P1 ^a , Q702P0	PF00190
<i>Pisum sativum</i> (Pea)	Pis s 1	Vicilin	44	–	PF00190
<i>Pisum sativum</i> (Pea)	Pis s 2	Convicillin	63	–	–
<i>Lens culinaris</i> (Lentil)	Len c 1	Gamma-vicilin subunit	47	Q84UI1, Q84UI0	PF00190
<i>Lens culinaris</i> (Lentil)	Len c 2	Seed-specific biotinylated protein	66	–	–
<i>Lens culinaris</i> (Lentil)	Len c 3	Nonspecific lipid transfer protein type 1	9	A0AT29 ^a	PF00234
<i>Lupinus angustifolius</i> (Narrow-leaved blue lupin)	Lup an 1	Conglutin beta (7S seed storage globulin, vicilin)	55–61	B8Q5G0	PF00190
<i>Phaseolus vulgaris</i> (Green bean, French bean)	Pha v 3	Non-specific lipid transfer protein type 1	8.8–9.0	–	–
<i>Vigna radiata</i> (Mung bean)	Vig r 1	Pathogenesis related (PR-10) protein	16	Q2VU97	PF00407
<i>Vigna radiata</i> (Mung bean)	Vig r 2	8S Globulin (Vicilin)	52	Q198W3, B1NPN8	PF00190
<i>Vigna radiata</i> (Mung bean)	Vig r 4	Seed albumin	30	Q43680	PF00045
<i>Vigna radiata</i> (Mung bean)	Vig r 6	Cytokinin-specific binding protein (Bet v 1 family member)	18	Q9ZWP8	PF00407

^a UniProtKB/Swiss-Prot ID.**Table 2**

Seven putative allergens from chickpea after the BLASTp analysis with their percentage similarities and ID's.

<i>Cicer arietinum</i> – BLAST retrieved entry	Allergen	Molecular function	E-value for BLAST	% Identity	Positives	UniProt/Swiss-Prot ID or UniProtKB/TrEMBL ID	Pfam ID
G1K3R9.CICAR	Vig r 4	Metal ion binding	5.0×10^{-79}	57.00%	Q43680-73	G1K3R9	PF00045
G1K3S0.CICAR	Vig r 4	Metal ion binding	1.0×10^{-74}	56.00%	Q43680-72	G1K3S0	PF00045
O23758.NLTP.CICAR	Ara h 9	Non-specific lipid-transfer protein	8.0×10^{-41}	62.00%	B6CEX8-75, B6CG41-80	O23758 ^a	PF00234
Q304D4.CICAR	Ara h 1	Nutrient reservoir activity	4.0×10^{-82}	48.00%	P43238-68	Q304D4	PF00190
	Gly m 5		4.0×10^{-82}	48.00%	O22120-70, Q9FZP9-70, P25974-71	Q304D4	PF00190
	Pis s 1		4.0×10^{-82}	48.00%	Q702P1-87, Q702P0-87	Q304D4	PF00190
	Len c 1		4.0×10^{-82}	48.00%	Q84UI1-87, Q84UI0-87	Q304D4	PF00190
	Lup an 1		4.0×10^{-82}	48.00%	B8Q5G0-76	Q304D4	PF00190
	Vig r 2		4.0×10^{-82}	48.00%	Q198W3-68, B1NPN8-69	Q304D4	PF00190
Q39450.CICAR	Ara h 8	Pathogenesis-related protein	1.0×10^{-74}	70.00%	Q6VT83-83, B0YIU5-73	Q39450	PF00407
	Gly m 4		1.0×10^{-74}	70.00%	P26987-85	Q39450	PF00407
	Vig r 1		1.0×10^{-74}	70.00%	Q2VU97-78	Q39450	PF00407
Q9SMK8.CICAR	Ara h 8	Pathogenesis-related protein	4.0×10^{-65}	62.00%	Q6VT83-81, B0YIU5-76	Q9SMK8	PF00407
	Gly m 4		4.0×10^{-65}	62.00%	P26987-78	Q9SMK8	PF00407
	Vig r 1		4.0×10^{-65}	62.00%	Q2VU97-74	Q9SMK8	PF00407
Q9SMJ4.LEG.CICAR	Ara h 3	Alpha-amylase inhibitor, Seed storage protein, Storage protein	1.0×10^{-145}	50.00%	O82580-64	Q9SMJ4 ^a	PF00190
	Ara h 4		1.0×10^{-150}	49.00%	Q9SQH7-63	Q9SMJ4 ^a	PF00190
	Gly m 6		1.0×10^{-150}	49.00%	P04776-70, P04405-70, Q7GC77-58, Q9SB11-56	Q9SMJ4 ^a	PF00190

^a UniProtKB/Swiss-Prot ID.

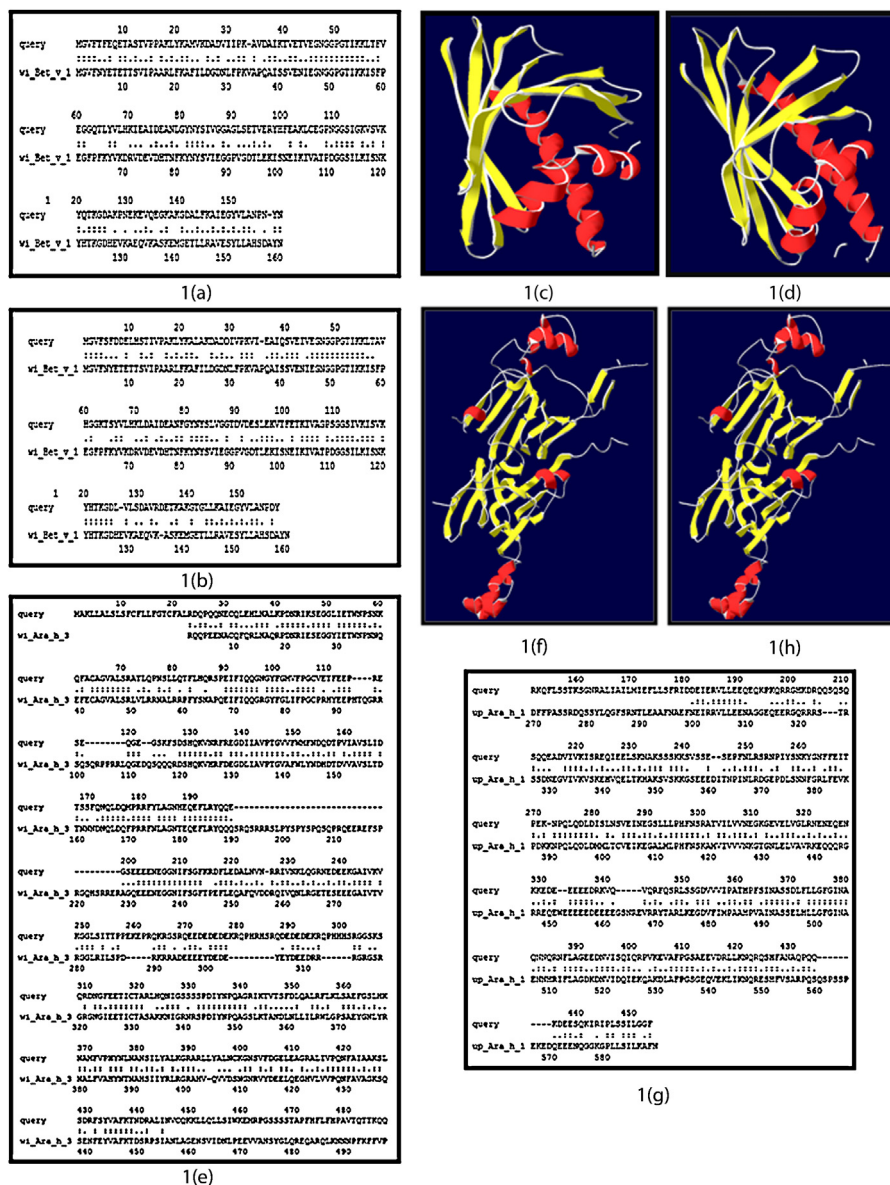


Fig. 1. AllergyMatch™ and Fugue: Sequence-Structure Homology online tool recognition analysis results showing the sequence similarities and the structural model respectively. (a) and (c) Sequence alignment and structural model for Q9SMK8, respectively. (b) and (d) Sequence alignment and structural model for Q39450, respectively. (e) and (f) Sequence alignment and structural model for Q9SMJ4, respectively. (g) and (h) Sequence alignment and structural model for Q304D4, respectively.

recognition online tool, it showed structural similarity of both proteins with Bet v 1. The rough structural models generated by the FUGUE: Sequence-structure homology analysis for Q9SMK8-Bet v 1 and Q39450-Bet v 1 are shown in Fig. 1c and d having a z-score of 36.40 and 33.97 respectively. The above analysis shows that the Q9SMK8 and Q39450 are very similar to Bet v 1, at the sequence as well as structural level and may thus show a similarity at the functional level. This conclusion is supported by the AlgPred™ analysis, as shown in Table 3 which predicts the above two proteins as allergens on the basis of amino acid composition, dipeptide composition and BLAST search on allergen representative peptides (ARPs). In addition to this, the physiochemical analysis of these two proteins by ProtParam (Table 4) suggest that Q9SMK8 and Q39450 have net negative charge (−6 and −3, respectively) and acidic pI (5.17 and 5.48, respectively), which mostly allergens tend to have (Singh et al., 2009). On the basis of the above results and outcomes it can be thus concluded that Q9SMK8 and Q39450 show similarity with the known allergens Ara h 8, Gly m 4, Vig r 1 and Bet v 1 and thus can be predicted to be the putative allergens from chickpea. It

can also be inferred that this sequence and/or structural similarity between the putative allergens and known allergens may also elicit allergic cross-reactivity within them (Schein et al., 2007).

3.2. Q9SMJ4

Ara h 3 and Gly m 6 are known allergens which belong to the 11S globulin family of seed storage proteins from peanut and soybean respectively as shown in Table 1. We retrieved Q9SMJ4 as the similar sequence from chickpea on the BLASTP analysis (Table 2) of these allergens, which again belongs to the same 11S globulin family. Major seed storage globulins are a part of the cupin superfamily which has a general characteristic of resistance to denaturation and proteolysis. This special stability may play a role in determining the allergenic activity of globulin storage proteins (Breiteneder and Mills, 2005; Hauser et al., 2008; Mills et al., 2002).

AllergyMatch™ showed high level of sequence similarity between Ara h 3 and Q9SMJ4, as shown in Fig. 1e, with 92.09% hits identical and thus satisfying the criteria of Codex alimentarius

Table 3
AlgPred analysis of the seven putative allergens from chickpea (X: – absence, ✓: – presence).

Cicer arietinum – BLAST retrieved entry	AlgPred Mapping of IgE epitopes and PID	SVM module based on amino acid composition				SVM module based on dipeptide composition				Blast search on allergen representative peptides (ARPs)	
		Allergen prediction	Score	Positive predictive value (%)	Negative predictive value (%)	Allergen prediction	Score	Positive predictive value (%)	Negative predictive value (%)		
G1K3R9.CICAR	X	✓	0.97968961	85.64	67.96	✓	0.29442982	85.88	72.01	X	
G1K3S0.CICAR	X	✓	1.0804723	85.64	67.96	✓	0.25609052	85.88	72.01	X	
O23758.NLTP.CICAR	X	✓	0.5094765	81.83	74.03	✓	0.1470044	74.14	79.04	✓	(LNLNNAASIPKCNVNVPTIS)
Q304D4.CICAR	✓(ELHLIGF- GIN)(KDLAFPGSGE)	✓	–0.10149276	64.55	86.61	X	–0.38678952	39.4	89.34	✓	(YSNQFGQFEACPEEHRLQEMD)
Q39450.CICAR	X	✓	1.6158648	85.64	67.96	✓	0.94717994	100	59.74	✓	(EGNGGPCTIKKISFEGPFFKYV)
Q9SMK8.CICAR	X	✓	1.0414618	85.64	67.96	✓	0.28953275	85.88	72.01	✓	(TFLEDGETKFLHKIESIDEANLG)
Q9SMJ4.LEG.CICAR	✓(IETWNPNNQFEFCAG) (GNFSGFTPEFLEQA)	✓	0.021663922	70.05	80.74	X	–0.5323609	27.66	92.4	✓	(QPEENACQFRLNAQRPNRIES)

(FAO and WHO, 2001, 2003) for being a potential allergen. By AllergenOnline, it has showed similarity with 34 allergen sequences. Rough model generated by Fugue: Sequence-Structure Homology online tool, considering Ara h 3 as a template, showed high level of structural similarity (Fig. 1f), between Ara h 3 and Q9SMJ4, with a z-score of 64.19, which may also predict the functional similarity. The AlgPred™ analysis (Table 3) predicts Q9SMJ4 as allergen on the basis of presence of IgE epitopes, amino acid composition and presence of allergen representative peptides (ARPs). ProtParam analysis (Table 4) shows that the protein has net negative charge (–6) and acidic pI (6.2) predicting higher potential of allergenicity (Singh et al., 2009). Hence we can conclude here Q9SMJ4 similarity with the known allergen Ara h 3 and can be predicted to be the putative allergen from chickpea. Here too it can be inferred that this sequence and/or structural similarity between the putative allergen and known allergen may also elicit allergic cross-reactivity within them (Schein et al., 2007).

3.3. Q304D4

Q304D4 is a protein involved in nutrient reservoir activity from chickpea. We retrieved it as a similar entry on the BLASTP analysis (Table 2) of the Ara h 1, Gly m 5, Len c 1, Lup an 1 and Vig r 2 allergens from peanut, soybean, lentil, narrow-leaved blue lupin and mung bean as shown in Tables 1 and 2, respectively. Pfam ID's showed that this chickpea protein consists of the cupin domain, like Q9SMK8, Q39450 and Q9SMJ4, which have been identified in major plant food allergens (Breiteneder and Mills, 2005; Hauser et al., 2008; Mills et al., 2002). As shown in Fig. 1g, sequence analysis by AllerMatch™ shows that the sequence has high similarity against allergen Ara h 1 from peanut with 56.15% hits identical and also by AllergenOnline, it showed similarity with the 26 allergen sequences thus satisfying the criteria of Codex alimentarius (FAO and WHO, 2001, 2003) for being a potential allergen. By considering Ara h 3 as a template a rough model with a z-score of 28.29 was generated by FUGUE: Sequence-Structure Homology online tool, which showed structural similarity of this protein with Ara h 3 (Fig. 1h). The above analysis shows that the protein Q304D4 may shows similarity at functional level with either Ara h 1 or Ara h 3. AlgPred™ analysis (Table 3) predicted this protein as allergen on the basis of presence of IgE epitopes and PID, amino acid composition, Blast search on allergen representative peptides (ARPs). ProtParam analysis (Table 4) showed the above protein has net negative charge (–5) and acidic pI (6.04) predicting potential allergenicity (Singh et al., 2009). Thus, Q304D4 shows similarity with the known allergens Ara h 1 and Ara h 3 can be thus predicted to be the putative allergen from chickpea and may also elicit allergic cross-reactivity within them (Schein et al., 2007).

3.4. G1K3R9, G1K3S0 and O23758

G1K3R9 and G1K3S0 are the metal ion binding proteins from chickpea retrieved as similar sequences by BLASTP analysis (Table 2) of Vig r 4 from mung bean. These proteins neither showed any identity with any reported allergens by AllergenOnline, AllerMatch™ or Fugue: Sequence-Structure Homology online tool nor had the physicochemical properties, as per ProtParam analysis, as allergens are reported to have (Singh et al., 2009). But they have been predicted as allergens by AlgPred™ analysis (Table 3) on the basis of amino acid composition and dipeptide composition. Thus, on the basis of BLASTP and AlgPred™ analysis, G1K3R9 and G1K3S0 can be predicted to be putative allergens from chickpea. But cross-reactivity of these proteins with known allergens cannot be predicted here since no structural similarities were found in between them (Schein et al., 2007).

Table 4
Physiochemical properties of the seven putative chickpea allergens by PROTPARAM.

Cicer arietinum – BLAST retrieved entry	PROTPARAM						
	Number of amino acids	MW (KDa)	Theoretical pI	Total number of negatively charged residues (Asp + Glu)	Total number of positively charged residues (Arg + Lys)	Net charge	Instability index
G1K3R9.CICAR	227	25,698	8.24	27	29	+2	16.96
G1K3S0.CICAR	227	25,686	8.24	27	29	+2	15.41
O23758 NLTP.CICAR	116	11,687.7	8.92	3	9	+6	26.26
Q304D4.CICAR	453	51,421.4	6.04	65	60	–5	54.56
Q39450.CICAR	158	16,887.1	5.48	21	18	–3	14.64
Q9SMK8.CICAR	157	16,669.9	5.17	23	17	–6	13.06
Q9SMJ4.LEG.CICAR	496	56,251	6.2	65	59	–6	54.34

O23758 is a lipid transfer protein from chickpea retrieved as similar sequence by BLASTP analysis (Table 2) of Ara h 9 and Len c 3 allergens from peanut and lentil respectively (Table 1). Plant lipid transfer proteins are considered as pan-allergens because of their thermo stability as well as resistance to proteolysis and harsh pH conditions also rendering them potent class I food allergens (Breiteneder and Mills, 2005; Hauser et al., 2008; Mills et al., 2002). This protein did not show any similarity with allergens by AllerMatch™, Fugue: Sequence-Structure Homology online Tool or ProtParam analysis but was predicted to be allergen by AlgPred analysis (Table 3) on the basis of amino acid composition and dipeptide composition and also showed similarity with 48 allergen sequences from AllergenOnline. On the basis of BLASTP analysis and the above discussions, it can be said that this may show allergenicity and thus can be putative allergen from chickpea. Although predicting cross-reactivity here would be again difficult since no structural similarity was found between putative and reported allergens (Schein et al., 2007).

3.5. Limitations

The main limitation of bioinformatics approach is that it makes predictions and can only be used to suggest experiments in the lab for testing those predictions. Further, modifications made to nucleic acids and proteins and the cell environment are often an important consideration for predicting significance of functional role. These things are hard to model every time using computational approaches. Also, choosing a database and algorithm can influence the error rates towards either a higher rate of false positives or false negatives results. A too high detection rate of false positives would generate an unnecessary and undesirable burden of additional testing whereas a too high detection rate of false negatives would generate undesirable potential health risks for consumers. Hence larger amount of databases and consideration of more parameters to decrease the amount of false negatives and positives results, respectively, should be utilized for greater accuracy and efficiency of results (FAO and WHO, 2001, 2003). Also, as per the current database entry, complete chickpea proteome sequence has not been sequenced and only 498 proteins are available from chickpea (UniProt Consortium, 2012). Thus, this analysis could have missed important allergens owing to gaps in the data.

4. Conclusions

The availability of the proteome sequences of chickpea (*C. arietinum*) permitted the identification of seven putative allergens Q9SMK8, Q39450, Q9SMJ4, Q304D4, G1K3R9, G1K3S0 and O23758. The sequence, structural and physiochemical similarities of four of the seven putative allergens namely Q9SMK8, Q39450, Q9SMJ4 and Q304D4 with known allergens may also predict the cross-reactivity within them.

Further clinical analysis of above specified allergens would be required for confirmation of their allergenicity and cross-reactivity which would also help us to design the appropriate methods for the production of hypoallergenic chickpea containing food. Combined with experimental methodology validated on animal models that has been extensively used for allergenicity testing, a bioinformatics approach will help define a more efficient and comprehensive approach to the assessment of chickpea and other food proteins. Also, numbers of genetically modified chickpea are being developed and produced using biotechnology approaches for the improvement of agronomic characteristics (such as resistance to herbicides) or quality (such as improved shelf life) wherein proteomic *in silico* analysis could be used to determine the safety of such products (GMO Database – Authorisation and approval status in the EU, 2013).

Acknowledgements

We gratefully acknowledge support and funding given by the University Grants Commission, India.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.compbiolchem.2013.08.003>.

References

- Allermatch, 25 August 2012, <http://www.allermatch.org/>
- Breiteneder, H., Mills, E.N.C., 2005. Plant food allergens—structural and functional aspects of allergenicity. *Biotechnology Advances* 23, 395–399, PMID: 15985358.
- Ebner, C., Hoffmann-Sommergruber, K., Breiteneder, H., 2001. Plant food allergens homologous to pathogenesis-related proteins. *Allergy* 56, 43–44, PMID: 11298007.
- FAO, WHO, 2001. Evaluation of allergenicity of genetically modified foods: report of a Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology, 22–25 January 2001. FAO, Rome. <http://www.who.int/foodsafety/publications/biotech/en/ec.jan2001.pdf>
- FAO, WHO, 2003. Codex principles and guidelines on foods derived from biotechnology. FAO, Rome. <ftp://ftp.fao.org/codex/standard/en/CodexTextsBiotechFoods.pdf>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, J.M. (Ed.), *The Proteomics Protocols Handbook*. Humana press, pp. 571–601, PMID: 10027275.
- GMO Database – Authorisation and approval status in the EU, 2013. GMO Compass, <http://www.gmo-compass.org/eng/gmo/db/>
- Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis* 18, 2714–2723, PMID: 9504803.
- Hauser, M., Egger, M., Wallner, M., Wopfner, N., Schmidt, G., Ferreira, F., 2008. Molecular properties of plant food allergens: a current classification into protein families. *The Open Immunology Journal* 1, 1–12.
- Hoffmann-Sommergruber, K., 2002. Pathogenesis-related (PR)-proteins identified as allergens. *Biochemical Society Transactions* 30 (Part 6), 930–935, PMID: 12440949.
- IUIS Allergen Nomenclature Home Page, <http://www.allergen.org/index.php>

- Kim, C.K., Kwon, S.J., Lee, G.S., Lee, H.K., Choi, J.W., Kim, Y.H., Hahn, J.H., 2009. A database for allergenic proteins and tools for allergenicity prediction. *Bioinformatics* 3 (8), 344–345, PMID: 19707297.
- Kumar, S., Verma, A.K., Das, M., Dwivedi, P.D., 2012. Molecular mechanisms of IgE mediated food allergy. *International Immunopharmacology* 13, 432–439, PMID: 22668720.
- Mills, E.N.C., Jenkins, J., Marigheto, N., Beltont, P.S., Gunning, A.P., Morris, V.J., 2002. Allergens of the cupin superfamily. *Biochemical Society Transactions* 30 (Part 6), 925–929, PMID: 12440948.
- Mills, E.N.C., Madsen, C., Shewryx, P.R., Wichers, H.J., 2003. Food allergens of plant origin—their molecular and evolutionary relationships. *Trends in Food Science & Technology* 14, 145–156.
- Patil, S.P., Niphadkar, P.V., Bapat, M.M., Chickpea, 2001. a major food allergen in the Indian subcontinent and its clinical and immunochemical correlation. *Annals of Allergy Asthma Immunology* 87 (2), 140–145, PMID: 11527247.
- Petrovsky, N., Brusic, V., 2002. Computational immunology: the coming of age. *Immunology and Cell Biology* 80, 248–254, PMID: 12067412.
- Saha, S., Raghava, G.P.S., 2006. AllPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research* 34, W202–W209, PMID: 16844994.
- Schein, C.H., Ivanciuc, O., Braun, W., 2007. Bioinformatics approaches to classifying allergens and predicting cross-reactivity. *Immunology and Allergy Clinics of North America*, 1–27, PMID: 17276876.
- Shi, J., Blundell, T.L., Mizuguchi, K., FUGUE, 2001. sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of Molecular Biology* 310 (1), 243–257, PMID: 11419950.
- Singh, S., Taneja, B., Salvi, S.S., Agrawal, A., 2009. Physical properties of intact proteins may predict allergenicity or lack thereof. *PLoS ONE* 4 (7), e6273, PMID: 19609445.
- The UniProt Consortium, 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Research* 40, D71–D75.