Computational analysis of SRK protein from *Brassica oleracea L.* for allergenic and antigenic characters

Rokade S.S., Tidke J.A. and Chikhale N.J.*
Department of Botany, Sant Gadge Baba Amravati University, Amravati – 444 602 (M.S.) India.
*Shri. Shivaji Agricultural College, Amravati– 444 602 (M.S.) India.
rokadesim@rediffmail.com

Abstract- The pollen antigens are variable and the study of the antigenic/allergenic fractions of the allergens is of paramount importance. The recent findings and techniques from the fields of molecular biology, nanotechnology, bioinformatics and information technology (IT) has given further stimulation for the identification, characterization and comprehensive classification of the allergenic modules and their sources. Therefore, during the present investigations SRK protein from Brassica oleracea was screened in-silico for its allergic and antigenic characters. SRK protein contains 658 amino acids and has 25 antigenic determinants. Out of which 4-SYTFSFLLVFLVLILFHPALSIYVNT-29 and 567-QHNNLVRLLGCCVYEGEKILIYEYLD NLSLDSHLFD-602 fragment shows highest antigenicity. It shows 1.0273 average antigenic propensity containing 50.7599% hydrophilic amino acids and 49.0881% hydrophobic amino acids. The SRK protein contains 72 basic amino acids and 69 acidic amino acids. The total linear charge density of the protein is 0.217325. The solubility of protein is 1.48797. The structural characters of the SRK protein show alpha-helix 29.9%, beta-sheet 28.3% and coil 41.8%. Motif map of SRK protein shows 509 motifs out of which 220 are MHC class I related, 200 motifs have functional relationship with B-cell membrane and others have functional relationship with T-cell membrane. Predicted results are sufficient to elicit the allergenicity of SRK protein.
Keywords: Pollen protein, Antigenicity, Pollen allergy, Brassica oleracea.

INTRODUCTION

Common incidence of the allergic diseases such as sneezing, a runny nose, itchy eyes and ears, sore threat, bronchial asthma, allergic rhinitis and atopic dermatitis may be due to pollen grains present in the surrounding environment. Some of the pollen types carry proteins that are more allergenic. In recent years considerable progress has been achieved in the elucidation of the molecular basis of pollen allergy with the characterization and cloning of allergic components involved in the allergic reaction [1-5].

Bioinformatics tests for allergenicity have become increasingly visible in the literature [6-15]. Largely this direction has emerged in response to a need for a relatively expedient method to screen for potential protein allergens [16]. The purpose of this screening is to determine if the introduced protein shares any sequence similarities with known allergen that would indicate the protein which could elicit a clinical reaction in an allergic population.

MATERIALS AND METHODOLOGY

Prediction of hydrophobicity

The SRK protein was scanned for Hydrophobicity and Parker hydrophilicity index [17]. Hydrophobicity (or hydrophilicity) plots are designed to display the distribution of polar and apolar residues along a protein sequence. Most commonly, this analysis has the goal of predicting membrane-spanning segments (highly hydrophobic) or regions that are likely exposed on the surface of proteins (hydrophilic domains) and therefore found to useful to identify potentially antigenic segments. Scale of hydrophobicity have been developed, which were derived from experimental studies on partitioning of peptides in apolar and polar solvents.

Prediction of antigenic sites

The SRK protein sequence of Brassica oleracea L. was analyzed and characterized to study the antigenicity and MHC class peptide binding, which allows potential drug targets to identify active sites against allergenic reactions. Antigenic epitopes are determined by using the method of Kolaskar and Tongaonkar [18]. Predictions are based on a table that reflects the occurrence of amino acid residues in experimentally known segmental epitopes. Prediction of antigenicity

program predicts those segments from within the SRK protein sequence that are likely to be antigenic by eliciting an antibody response. B-cell epitopes are predicted according to the method given by Saha and Raghava [19]. Flexibility recorded by using Karplus-Schulz index [20].

Secondary alignment

Secondary structure prediction was determined by using Chou and Fasman [21] and Garnier [22] method. These methods are based on information theory. The outputs of these programs are alpha-helix, beta-sheet and coil and gives probable value for each secondary structure. The predicted structure is one of the highest probably compatible with experimental structure.

Protein statistics

In-silico protein statistics was carried out with respect to several parameters such as atomic weight, average molecular weight, hydrophobic and hydrophilic amino acids percentage, number of basic and acidic amino acids, linear charge density and solubility of protein, amino acid frequency and hydrophobicity.

RESULTS AND DISCUSSION

The SRK protein is made up of 658 amino acids and its total atomic weight is 74991.719 daltons. From the In-silico studies, it was observed that the SRK protein comprising average molecular weight of the amino acids i.e. 113.969. It contains 50.7599% hydrophilic amino acids and 49.0881% hydrophobic amino acids. The protein contains 72 basic amino acids and 69 acidic amino acids. The total linear charge density of the protein is 0.217325. The solubility of protein is 1.48797. Structural characters of the protein observed in-silico are also listed “Table (1)”. The information from hydrophobicity was useful in identifying coil regions, exposed loops, interior domains, B-cell antigenic determinants and membrane spanning regions within the sequence. The maximum value of hydrophobicity for SRK protein is 4.00 “Fig. (1)”. The Parker [17] hydrophilicity index shows average value 0.945, minimum value -8.243 (in negative) and maximum value 7.029 “Fig. (2)”. From the antigenic peptide prediction program, it was observed that the antigenic determinants are in the area of greatest local hydrophilicity. The Kolaskar-Tongaonkar [18] antigenicity scale shows that SRK protein is highly antigenic nature and has 25 antigenic determinants “Fig. (3)”. Segment from amino acid number 4 to 29 shows antigenic propensity more than 1.26. The average antigenic propensity of this protein is 1.0273 “Table (2)”. The predicted B cell epitopes are ranked according to their score obtained by trained recurrent neural network. Higher score of the peptide means the higher probability to be as epitope “Table (3)”. All the peptides shown here are above the 0.51 threshold value. Karplus-Schulz flexibility index [20] displayed in graph of the main chain mobility within a protein based on sequence information alone “Fig. (4)”. Motif map of SRK protein shows antigenic motifs which are MHC class 1 related and also have functional relationship with B-cell membrane and T-cell membrane “Fig. (5)”. The SRK protein structure is predicted by using Chau and Fasman [21] and Garnier [22]. Each residue is assigned values for α-helix, □-sheet and extended coil using window size seven residues. Using this information parameter shows the structural characters of the SRK protein as α-helix 29.9%, □-sheet 28.3% and extended coil 41.8% “Fig. (6)”. The beta staircase shows how the amino acids are positioned in relation to one another. The hydrophobic amino acids are seen on outside of the beta staircase “Fig. (7)”. The helical wheel assumes a periodicity of 3.6 residues per helical turn. Individual residues represented as a color circles are placed successively at each node of the helix. Multiple turns of the helix spiral outwards from helix center. The inter connectivity bars indicate the residue arrangement along the helix “Fig. (8)”. Serine, tyrosine, histidine, threonine, leucine and phenylalanine are placed on outer side of the helix. Amino acid frequency and percentage amino acid weights are also determined in-silico by using peptool “Fig. (9 and 10)”. It was observed that Leucine was 12%, each in total residue of amino acid in SRK protein, and then it was followed by serine and glycine. Histidine represents 2% amino acid frequency. By weight leucine and arginine the dominating amino acids in SRK protein. Cystein and histidine residues are representing 2% by weight in SRK protein “Fig. (10)”.

CONCLUSION

Allergenic reactions to the pollen protein of Brassica oleracea are hazardous to human health causing some allergic dermatitis, hay-fever and respiratory problems. Small peptide nonamers or fragments from Brassica oleracea pollen involve multiple antigenic components to direct and empower the immune system to protect the host from allergic infections. Knowledge of epitopes may be used in the design of vaccines and diagnostics tests. It is therefore of interest to develop improved methods for predicting epitops “Table (2)”. The nonamers shows high antigenic response because of presence of beta sheets regions. The antigenic determinants site 4-SYTFSFLLVFLVLILFHPALSIYVNT-29 shows high antigenic nature and form beta sheets in secondary structure “Fig. (6)” and also showing hydrophobic characteristics. These small peptide fragments of antigen can induce immune response against whole antigen. This theme may be implemented in designing subunit and synthetic peptide vaccines. The antigenicity analysis method allows potential drug targets to identify active sites, which forms antibodies against Brassica oleracea L. pollen allergy. The results are further confirmed by studying the protein statistics antigenic motifs of SRK protein, which are found to contain antigenic sites. Antigenic epitopes of major SRK protein are important antigenic determinants against the allergic reactions.

REFERENCES

[1] Chakraborty, P., I. Roy, S. Chatterjee, S. Chanda and S. Gupta-Bhattacharya (2006). Phoenix sylvestris Roxb. pollen allergy: a 2-year randomized controlled trial and follow-up study of immunotherapy in patients with seasonal allergy in an agricultural area of West Bengal, India. J. Investig Allergol Clin Immunol; 16(6): 377 – 384.

[2] Ariano, R., G. Mistrello, G. Mincigrucci, E. Bricchi, O. Lannotti, G. Frenguelli, G. Passalaequa and R. C. Panzani (2006). In-vitro and in-vivo biological activities of old and fresh Cupressus arizonica pollen. J. Investig Allergol Clin Immunol; 16(3): 177 – 182.

[3] Shahali, Y., A. Majd, Z. Pourpark, G. Tajadod, M. Halftlang and M. Moin (2007). Comparative study of the pollen protein contents in two major varieties of Cupressus arizonica planted in Tehran. Iran J Allergy Asthma Immunol; 6(3): 123 – 127.

[4] Moverare, R., H. Everberg, R. Carlsson, A. Holtz, R. Thunberg, P. Olsson, P. Brostedt and E. Hogbom (2008). Purification and Characterization of the major oak pollen allergen Que a 1 for component-resolved diagnostics using immunoCAP® . Int Archs Allergy Immunol; 146(3): 203 – 211.

[5] Duffort, O., J. Quintana, H. Ipsen, D. Barber and F. Polo (2008). Antigenic similarity among group 1 allergens from grasses and qunantitation of ELISA using monoclonal antibodies to Phl p 1. Int Archs Allergy Immunol; 145(4): 283 – 190.

[6] Bjorklund, A. K., D. Soeria-Atmadja, A. Zorzet, U. Hammerling and M. G. Gustafsson (2005). Supervised identification of allergen-representative peptides for in-silico detection of potentiality allergenic proteins. Bioinformatics; 21: 39 – 50.

[7] Jiang, S. Y., P. X. H. Jasmin, Y. Y. Ting and S. Ramachandran (2005). Genome-wide identification of Ole_e_I, Allergen_1 and Allergen_2 Domain-containing pollen allergen-like genes in Oryza sativa. DNA Research; 12: 167 – 179.

[8] Domon, B. and R. Aebersold (2006). Challenges and oppournities in proteomics data analysis. Mol Cell Proteomics; 5: 1921 – 1926.

[9] Mari, A., P. Palazo and E. Scala (2006). Allergome: Bioinformatic tools to manage the increasing knowledge of allergens. New Horizons Allergy; 1; 1 – 8.

[10] Jiang, B., H. Qu, Y. Hu, T. Ni and Z. Lin (2007). Computational analysis of the relationship between allergenicity and digestibility of allergenic proteins in simulated gastric fluid. BMC Bioinformatics; 8: 375.

[11] Matthiesen, R. (2007). Methods, algorithms and tools in computational proteomics: a practical point of view. Proteomics; 7: 2815 – 2832.

[12] Zhang, Z. H., J. L. Y. Kho, G. L. Zhang, K. H. Choo, M. T. Tammi and J. C. Tong (2007). AllerTool: a web server for predicting allergenicity and allergic cross-reactivity in proteins. Bioinformatics; 23(4): 504 – 506.

[13] Gomase, V. S., J. A. Tidke and K. V. Kale (2008). Prediction of antigenic MHC binders of calreticulin protein from Parthenium argentatum. International Journal of Bioinformatics; 1 (1): 37 – 44.

[14] Hennig, J., K. D. M. Hennig and M. Sunnerhagen (2008). MTMDAT: Automated analysis and visualization of mass spectrometry data for tertiary and quaternary structure probing of proteins. Bioinformatics; 24(10): 1310 – 1312.
[15] Kumar, K. K. and P. S. Shelokar (2008). An SVM method using evolutionary information for the identification of allergic proteins. Bioinformation; 2(6): 253 – 256.
[16] Goodman, R. E., S. L. Hefle, S. L. Taylor and R. van Ree (2005). Asseing genetically modified crops to minimize the risk of increase food allergy: a review. Int Archs Allergy Immunol; 137: 153 – 166.
[17] Parker, J. M., D. Guo and R. S. Hodges (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry; 23:25.
[18] Gomase VS and Vankhede GN. Antigenicity prediction in melittin: approach for fragment based peptide drug development from Apis dorsata. World Allergy Organization Journal, Volume WAC 2007, November 2007, S235.
[19] Vankhede GN, Gomase VS, Chikhale NJ. Study of MHC nonamers and TAP activity of tertiapin from Apis mellifera. World Allergy Organization Journal, Volume WAC 2007, November 2007, S235.
[20] Gomase VS, Chikhale NJ and Vankhede GN. Prediction of antigenic MHC binders of neurotoxin M14 protein from Buthus eupeus. World Allergy Organization Journal, Volume WAC 2007, November 2007, S237.
[21] Chou, P. Y. and G. D. Fasman (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol.; 47:45-148.
[22] Garnier, J. and D.J. Osguthorpe and B. Robson (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol.; 120:97-120.

Table 1- Protein statistics of SRK protein

Sr. No.	Parameter	Value
1	Molecular weight (Daltans)	74991.719
2	Number of amino acids	658
3	Mean amino acid weight (Daltans)	113.969
4	Average hydrophobicity	-0.209574
5	Ratio of hydrophilicity to hydrophobicity	1.166
6	Percentage of hydrophilic amino acids	50.7599
7	Percentage of hydrophobic amino acids	49.0881
8	Ratio of percentage hydrophilic to Percentage hydrophobic	1.03406
9	Mean beta hydrophobic moment	0.222673
10	Mean helix hydrophobic moment	0.164583
11	Number of basic amino acids	72
12	Number of acidic amino acids	69
13	Estimated pH for protein	8.7
14	Total linear charge density	0.217325
15	Polar area of extended chain ($Angs^2$)	44000.2
16	Non-polar area of extended chain ($Angs^2$)	74503.5
17	Total area of extended chain ($Angs^2$)	118504.0
18	Polar ASA of folded protein ($Angs^2$)	9301.98
19	Non-polar ASA of folded protein ($Angs^2$)	13191.4
20	ASA of folded protein ($Angs^2$)	22493.4
21	Ratio of folded to extended area	0.202676
22	Buried polar area of folded protein ($Angs^2$)	30862.1
23	Buried non-polar area of folded protein ($Angs^2$)	53788.2
24	Buried charge area of folded protein ($Angs^2$)	3527.1
25	Total buried surface ($Angs^2$)	88177.4
26	Number of buried amino acids	300
27	Packing volume (est) ($Angs^3$)	91346.9
28	Packing volume (act) ($Angs^3$)	89500.5
29	Interior volume of protein ($Angs^3$)	66130.8
30	Exterior volume of protein ($Angs^3$)	23510.1
31	Partial specific volume (ml/g)	0.724781
32	Fisher volume ratio (act)	0.355509
33	Fisher volume ratio (idealized)	0.460831
34	Protein solubility	1.48797
35	Estimated radius of folded protein (Angs)	33.7039
36	RMS end to end distance of extended chain (Angs)	269.035
37	Radius of gyration of extended chain (Angs)	109.833
38	Solv. Free energy folding (Kcal/mol)	-635.4

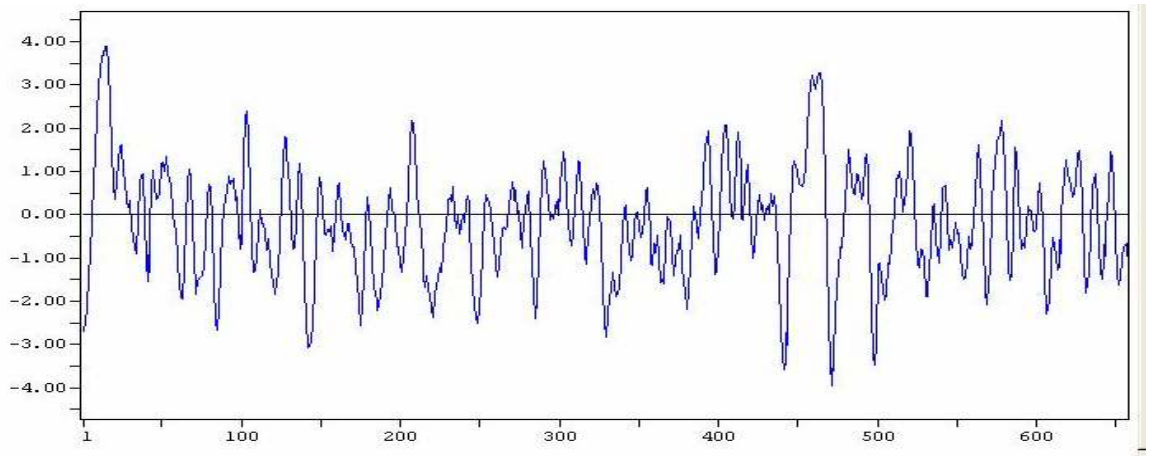

Fig. 1- Hydrophobicity scale of SRK protein

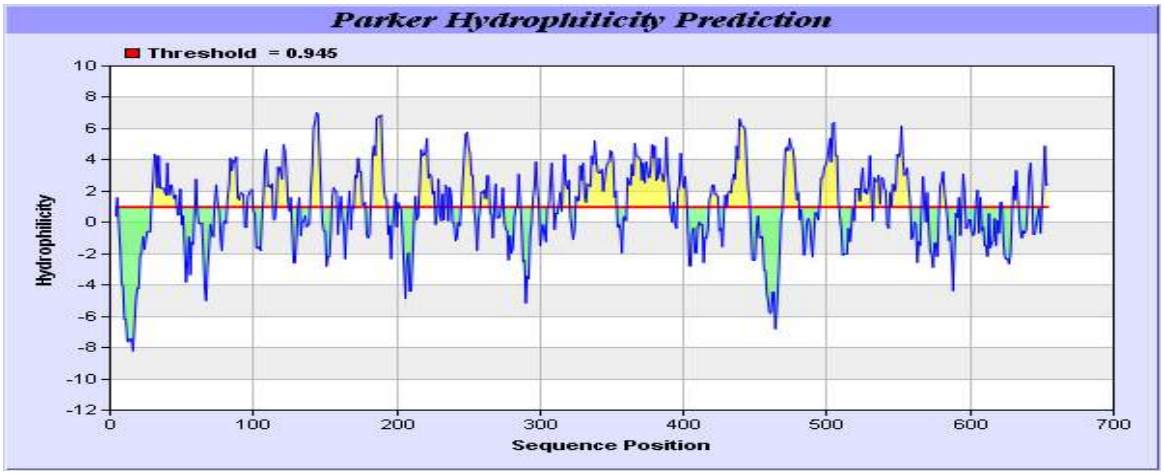

Fig. 2- Parker hydrophilicity scale of SRK protein

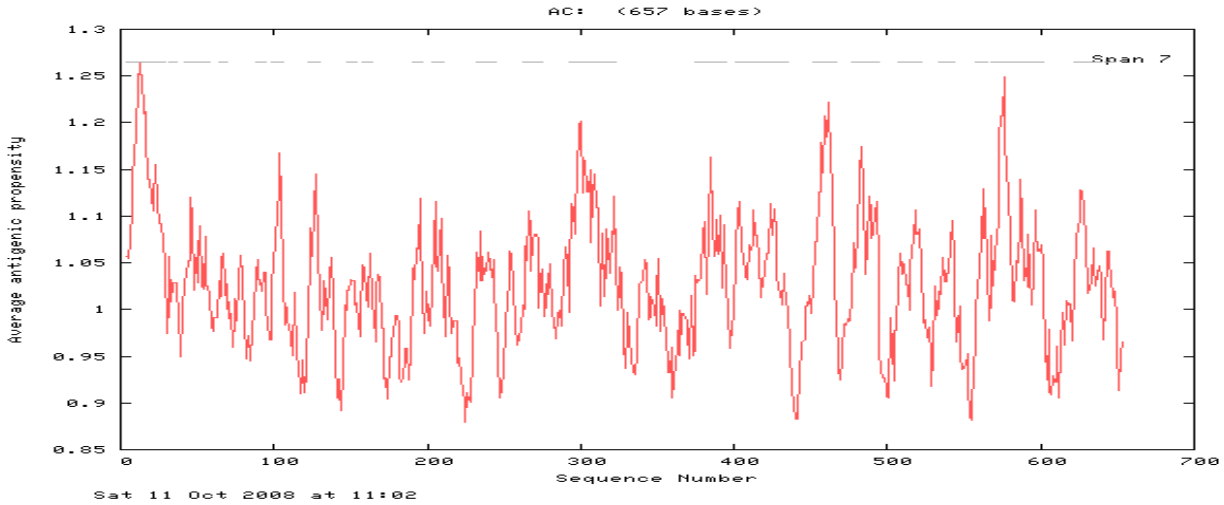

Fig. 3- Antigenic determinant plot for SRK protein sequence

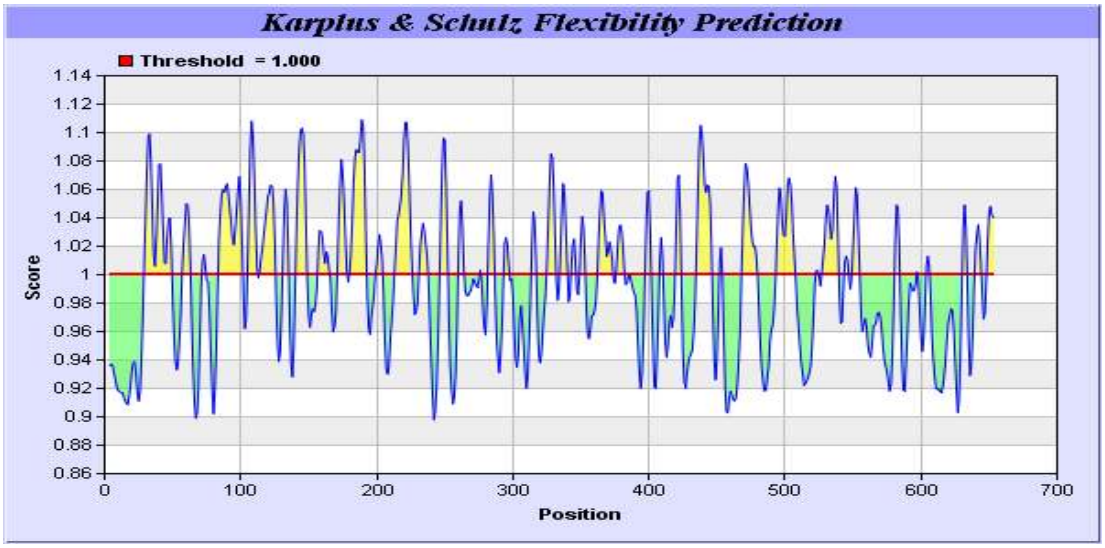

Fig. 4- Karplus-Schulz flexibility of SRK protein sequence

Table 2- Antigenic determinants in SRK protein

Sr. No.	Start Position	Sequence	End Position
1	4	SYTFSFLLVFLVLILFHPALSIYVNT	29
2	31	SSSESLT	37
3	41	NRTLVSPGGVFELGFFKP	58
4	64	WYLGIWY	70
5	88	LSSSIGTL	95
6	99	GNNLVLLS	106
7	122	ARSPVIAEL	130
8	148	GFLWQSF	154
9	158	TDTLLPE	164
10	190	SGNFVYKL	197
11	203	LPEFILINQ	211
12	232	GIPEVQGLNYMVYN	245
13	262	QSIYSRLTVSELTL	275
14	293	LPTDVCDPLYLCGSYSYCDLITSPNCNCIRG	323
15	375	DVKKCEERCLSDCNCTSFAIA	395
16	401	GLGCVFWTGELVAIRKFAVGGQDLYVRLNAADLDI	435
17	451	IGSSVMLILSVILFCFW	467
18	477	DATPIVGYQVLMNEVVLP	494
19	507	ENLELPL	513
20	515	EFEAVVTA	522
21	533	RLVDGQEIAVKR	544
22	558	NEVRLIAK	565
23	567	QHNNLVRLLGCCVYEGEKILIYEYLDNLSLDSHLFD	602
24	622	ARGILYLHHDSSIRIIHRD	640
25	642	KASNILL	648

Table 3- Predicted B-cell epitop in SRK protein

Rank	Sequence	Start position	Score
1	NRTLVSPGGV	41	0.85
2	LPEFILINQF	203	0.80
3	DDPSSGNFVY	186	0.78
4	VFELGFFKPL	50	0.77
5	TRGCVRTTQM	338	0.76
5	NGNFVIRHSN	133	0.76
6	EERCLSDCNC	380	0.75
6	NLPDTKTATV	361	0.75
7	TSPNCNCIRG	314	0.74
7	LDIRRGLPEF	197	0.74
8	WTLPTDVCDP	291	0.73
8	PWNGMEFSGI	224	0.73
9	SGNFVYKLDI	190	0.72
9	SFDFPTDTLL	153	0.72
10	PWKTYAWVAN	74	0.71
10	QFLNQRVETQ	211	0.71
10	FLWQSFDFPT	149	0.71
11	QQWDLRDGTR	330	0.70
11	RNRFLTSWKG	175	0.70
11	NKDSSGFLWQ	143	0.70
12	SIGTLKISGN	91	0.69
12	VRTTQMSCSG	342	0.69
12	SGIPEVQGLN	231	0.69
12	EMKLGYDLKT	164	0.69
12	TLLPEMKLGY	160	0.69
12	LLSQSTNTVW	104	0.69
13	RGFVPKNPQQ	322	0.68
13	SYSYCDLITS	306	0.68
13	DPLYLCGSYS	299	0.68
13	LTWIPPSRDW	278	0.68
13	HMTNQSIYSR	258	0.68
13	QGLNYMVYNY	237	0.68
14	KPLGRSRWYL	57	0.67
14	CDLITSPNCN	310	0.67
15	SGNNLVLLSQ	98	0.66
15	MDVKKCEERC	374	0.66
15	IYSRLTVSEL	264	0.66
15	VIRHSNNKDS	137	0.66
16	IWYKKAPWKT	68	0.64
16	IAELLPNGNF	127	0.64
16	RSPVIAELLP	123	0.64
17	ADVRNGGLGC	395	0.63
17	YTENSEEIAY	246	0.63
17	LVFLVLILFH	11	0.63
18	SDCNCTSFAI	385	0.62
19	GDGFLRLNNM	351	0.61
20	NRDNPLSSSI	83	0.60
21	MSCSGDGFLR	347	0.59
21	YHHSYTFSFL	1	0.59
22	AWVANRDNPL	79	0.58
22	ELTLDRLTWI	272	0.58
22	NLTRGNARSP	116	0.58
23	SSSESLTISS	31	0.57
23	NTVWSTNLTR	110	0.57
24	RLNNMNLPDT	356	0.56
24	AYSFHMTNQS	254	0.56
25	GGLGCVFWTG	400	0.55
26	SIYVNTLSSS	24	0.54
27	CVFWTGELVA	404	0.53
27	RVETQRSGPW	216	0.53
28	LTVSELTLDR	268	0.51

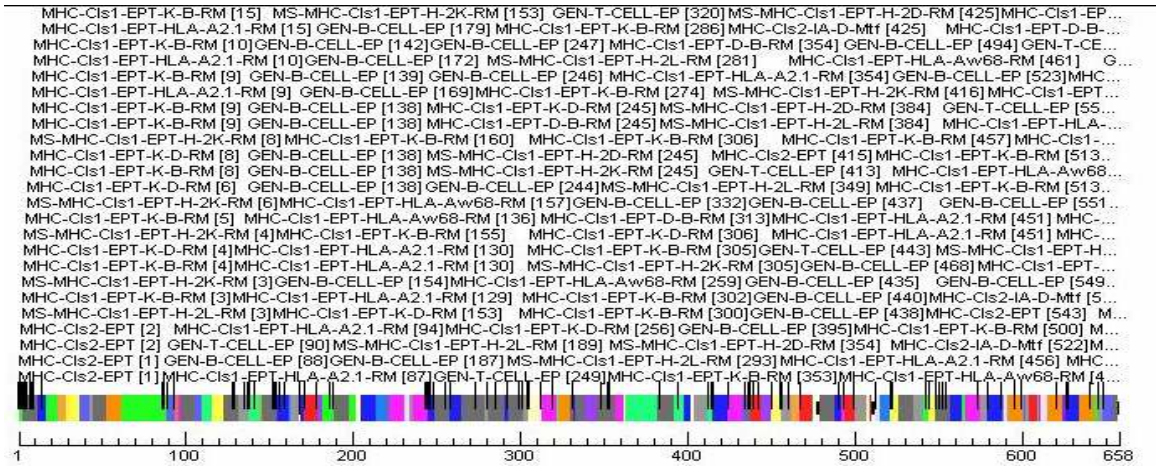

Fig. 5- Motif map of SRK protein

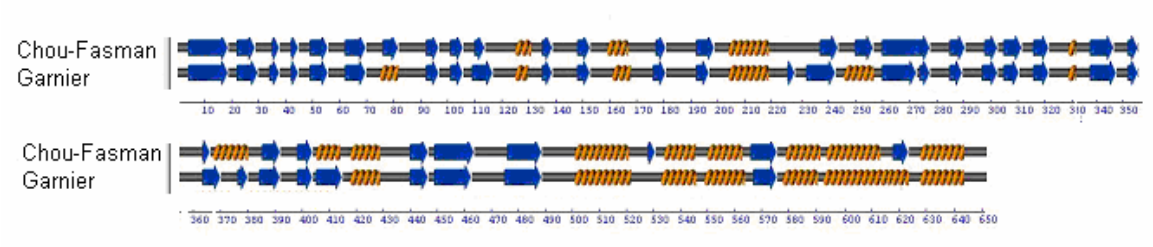

Fig. 6- Structure prediction of SRK protein

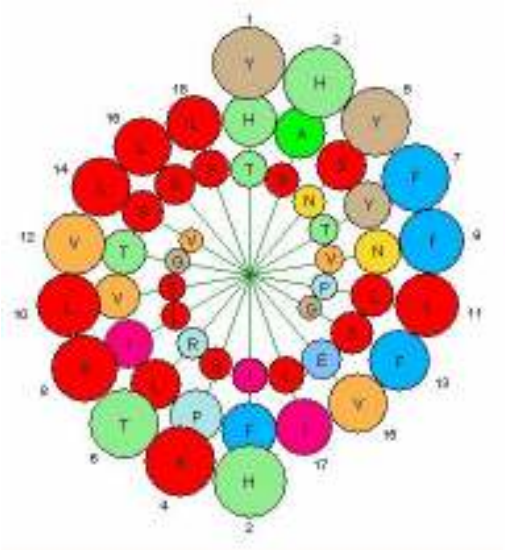

Fig. 7- Beta staircase arrangement of amino acids in SRK protein

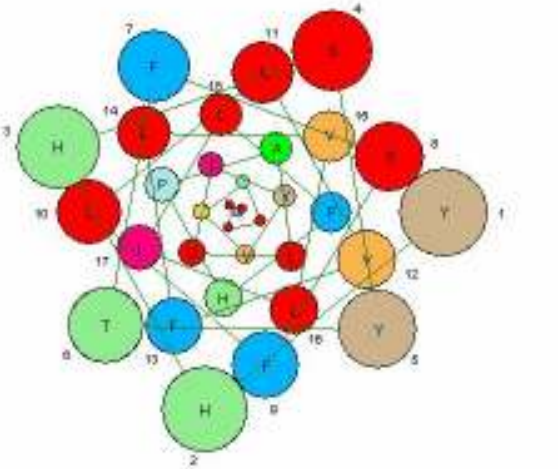

Fig. 8-Helical wheel arrangement of amino acids SRK protein

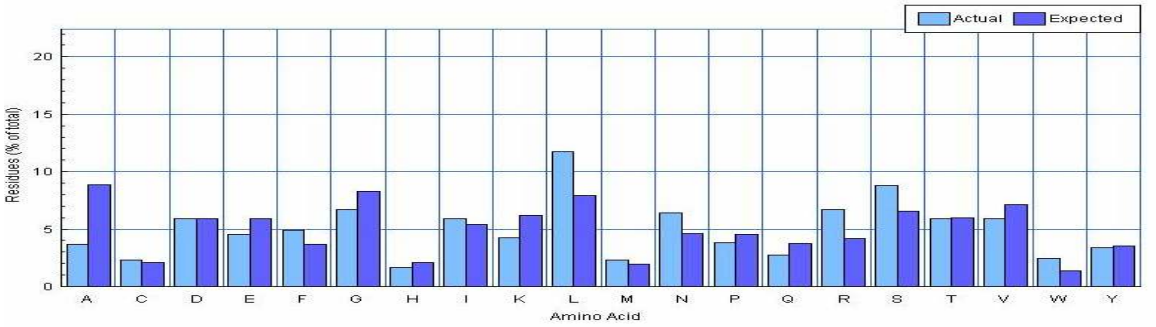

Fig. 9- Actual and expected amino acid frequency in SRK protein

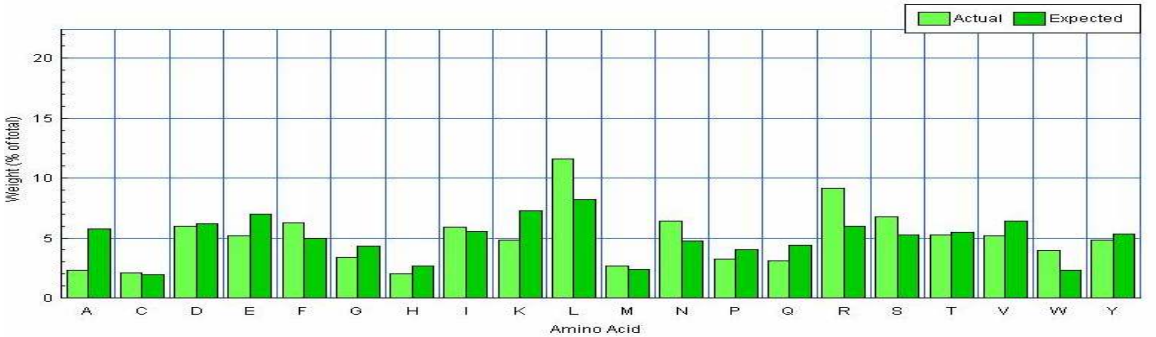

Fig. 10- Actual and expected percent amino acid weight in SRK protein