International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2018; 6(2): 3283-3291

Received: 08-01-2018
Accepted: 10-02-2018

Neha Gupta
Department of Chemistry and Biochemistry, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

Gagan Rani
Department of Chemistry and Biochemistry, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

Neeru Singh Redhu
Department of Chemistry and Biochemistry, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

Sudhir Kumar
Department of Chemistry and Biochemistry, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

Correspondence
Neha Gupta
Department of Chemistry and Biochemistry, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

In silico characterization and modelling of *Drosophila melanogaster* krotzkopf verkehrt protein

Neha Gupta, Gagan Rani, Neeru Singh Redhu and Sudhir Kumar

Abstract
Chitin synthase is known to be an important enzyme that catalyzes last step of chitin biosynthetic pathway. It causes linear polymerization of chitin from activated UDP-N-acetylglucosamine monomers in manner joined together by β-(1, 4)-linked. Chitin synthases are thus critical enzymes for synthesis of chitin and as well as for growth and development of insects. In this study, physiochemical properties and modeling of isoforms of chitin synthase (krotzkopf verkehrt protein) which are encoded by kkv genes were analyzed using *in silico* approach. Physiochemical properties such as molecular weight, theoretical isoelectric point, extinction coefficient, aliphatic index, instability index, total number of negatively and positively charged residues and grand average of hydrophathicity were computed. Along with these physiochemical properties cellular localilization, no. of transmembrane helices, other proteins with which these isoforms interact were also depicted using various tools.

Keywords: modelling, characterization, krotzkopf verkehrt, protein

Introduction
After cellulose, chitin is earth's second most abundant organic compound and is synthesized by a broad variety of organisms of different taxonomic groups. Chitin is found not only in arthropods including insects, arachnids and crustaceans but also in lower invertebrates such as sponges, coelenterates, nematodes and molluscs (Merzendorfer, 2011) [2]. Chitin is a linear polymer of N-acetyl- β-D-glucosamine and is a major component of insect cuticle and peritrophic matrices. Chitin is found in the exo- and endocuticle or in the newly secreted, unsclerotized procuticle but not in the epicuticle, the outermost part of the integument. Insects cuticles form an exoskeleton that exhibits only a limited capacity to keep pace with body growth because it is a more or less rigid structure due to the presence of chitin and sclerotized proteins. To allow growth and development, insects are therefore periodically forced to replace their old cutical with a new one during molting (ecdysis). Chitin functions as light but mechanically strong scaffold material and is always associated with cuticle proteins that mainly determine the mechanical properties of the cuticle (Merzendorfer and Zimoch, 2003) [3]. Chitin is also an integral part of insect peritrophic matrices which functions as a permeability barrier between the food bolus and the midgut epithelium, enhance digestive processes and protect the brush border from mechanical disruption as well as from attack by toxins and pathogens (Tellam, 1996). Thus, insect growth and development is strictly dependent on the capability to remode chitinous structures. Therefore, insects consistently synthesize and degrade chitin in a highly controlled manner to allow ecdysis nad regeneration of the peritrophic matrices. Formation of the different forms of chitin is catalyzed by chitin synthase (CHS) (EC 2.4.1.16), a highly conserved enzyme found in every chitin synthesizing organism. Krotzkopf verkehrt (kkv) gene encodes the Drosophila chitin synthase enzyme. Inhibiting the chtin synthase encoded by kkv gene leads to disruption of chtin bio synthetic path way (Ostrowski, 2002) [5]. Thus chtin synthases represents an attractive target site for combating insects pests as insect growth and development are strictly dependent on precisely tuned chitin biosynthesis and this pathway is absent in humans and other vertebrates.

Methodology
Sequence retrieval
The amino acid sequences of various isoforms of krotzkopf verkehrt protein (Isoform A, C and D) were retrieved from NCBI having the accession number NP_52433.1, NP_9961521.1,

and NP_730928.2 respectively. The characterization of various isoforms was done by using bioinformatics tools *in silico.*

Characterization of target sequence

The physiochemical property of the protein was determined by using Protparam tool such as molecular weight, theoretical pI, total number of negatively and positively charged amino acids, amino acid composition of the protein, extinction coefficient, aliphatic index and GRAVY index. The subcellular localization of the protein was found out using Cello v2.5. The transmembrane helices were predicted using TMHMM tool of Expasy. Signal P 4.1 server was used to check if the protein is a signal peptide or not. Motifs and Domain in the protein was identified using the conserverd domain search of NCBI database and the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database was used to identify proteins interacting with the sodium channel protein. Multiple sequence alignment of all the isoforms of both the proteins was done using CLUSTAL OMEGA. Secondary structure prediction was done using SOPMA. *Ab initio* modelling was carried out using phyre 2 server.

Results and Discussion

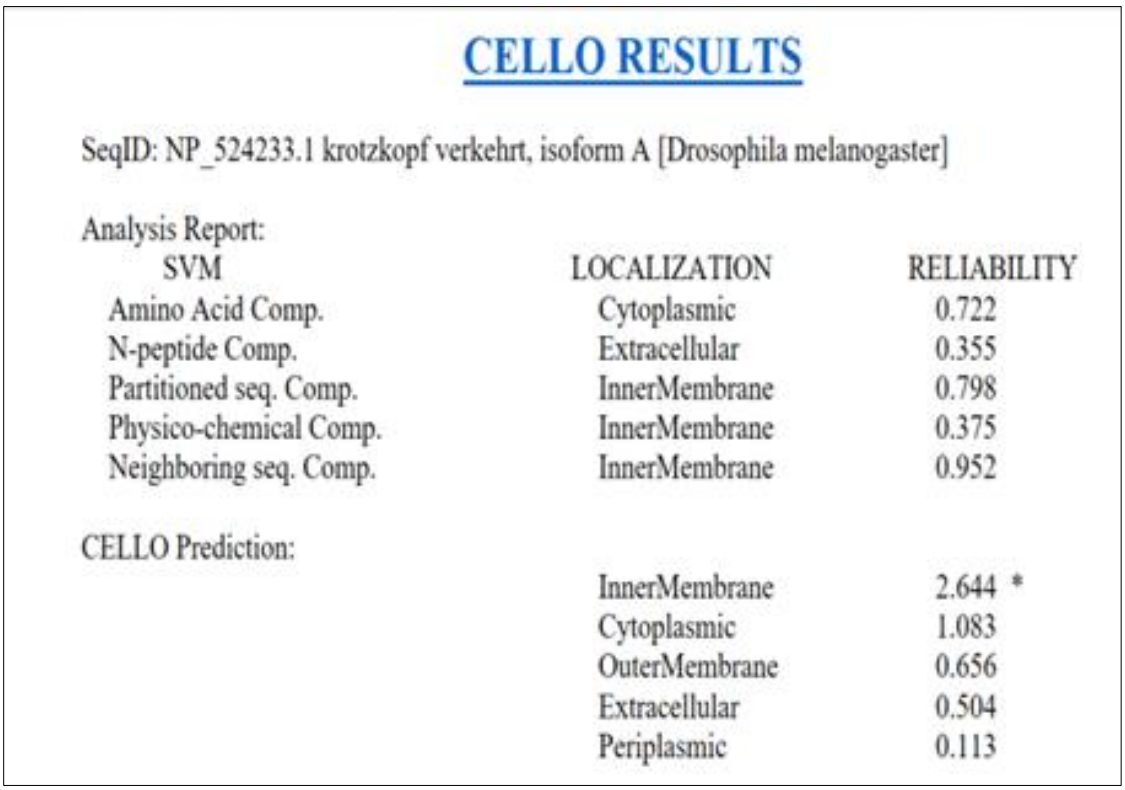
The results of Protparam tool showed that the molecular weight of all the three isoforms of krotzkopf verkehrt protein is same. Theoretical pI of the three is forms of krotzkopf verkehrt protein have a the ortical pI of 6.7. Total no. of negatively charged (Asp+Glu) residues in the isoform A & C of krotzkopf verkehrt protein is 187 and total number of positively charged (Lys+Arg) residues is 174 leaving them with total charge of -13 whereas isoform D has total negative charge of 186 and total positive charge is 173, with total charge of -13.Extinction coefficient (in M^{-1} cm^{-1} & at 280 nm measured in water) for the isoforms of krotzkopf verkehrt protein assuming all pairs of Cys residues form cystines is 241835and under same conditions extinction coefficient assuming all Cys residues are reduced for the for all three isoforms was found to be 239960. Extinction coefficient can be used to separate the protein from the solution. The instability index (II) determines the stability of the protein in a test tube and is computed to be 40.22 for the isoforms of kkv protein. Since instability index for all the isoforms is more than 40, it indicates that all the isoforms may be unstable. The value 91.55 of aliphatic index for the isoforms indicates the relative volume of a protein that is occupied by aliphatic side chains which in turn contributes to the increased thermo stability of isoforms of the protein. Negative Grand average of hydropathicity (GRAVY) index value for isoforms of kkv protein points towards their hydrophilic nature.

Table 1: Various physiochemical properties by PROTPARAM

Protein Name		Molecular Weight (Dalton)	Theortical pi	Total Charge	Instability Index	Aliphatic Index	Gravy Index
Krotzkopf Verkehrt Protein	ISOFORM A	182829.51	6.37	-13	40.22	91.55	-0.108
	ISOFORM C	182829.51	6.37	-13	40.22	91.55	-0.108
	ISOFORM D	182829.51	6.37	-13	40.22	91.55	-0.108

Cello v2.5

The sub-cellular localization of the protein was predicted by using the tool Cello v2.5 which shows that all three isoforms of krotzkopf verkehrt protein are located in inner membrane with isoform D having the highest reliability 2.675.



CELLO RESULTS

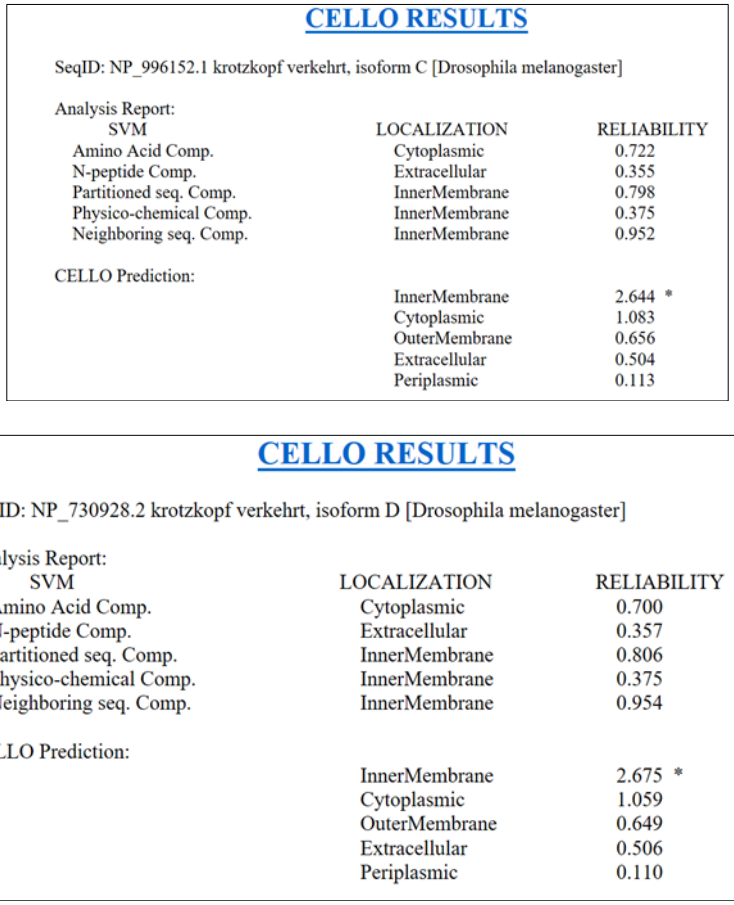
SeqID: NP_524233.1 krotzkopf verkehrt, isoform A [Drosophila melanogaster]

Analysis Report:

SVM	LOCALIZATION	RELIABILITY
Amino Acid Comp.	Cytoplasmic	0.722
N-peptide Comp.	Extracellular	0.355
Partitioned seq. Comp.	InnerMembrane	0.798
Physico-chemical Comp.	InnerMembrane	0.375
Neighboring seq. Comp.	InnerMembrane	0.952

CELLO Prediction:

InnerMembrane	2.644 *
Cytoplasmic	1.083
OuterMembrane	0.656
Extracellular	0.504
Periplasmic	0.113



CELLO RESULTS

SeqID: NP_996152.1 krotzkopf verkehrt, isoform C [Drosophila melanogaster]

Analysis Report:

SVM	LOCALIZATION	RELIABILITY
Amino Acid Comp.	Cytoplasmic	0.722
N-peptide Comp.	Extracellular	0.355
Partitioned seq. Comp.	InnerMembrane	0.798
Physico-chemical Comp.	InnerMembrane	0.375
Neighboring seq. Comp.	InnerMembrane	0.952

CELLO Prediction:

InnerMembrane	2.644 *
Cytoplasmic	1.083
OuterMembrane	0.656
Extracellular	0.504
Periplasmic	0.113

CELLO RESULTS

SeqID: NP_730928.2 krotzkopf verkehrt, isoform D [Drosophila melanogaster]

Analysis Report:

SVM	LOCALIZATION	RELIABILITY
Amino Acid Comp.	Cytoplasmic	0.700
N-peptide Comp.	Extracellular	0.357
Partitioned seq. Comp.	InnerMembrane	0.806
Physico-chemical Comp.	InnerMembrane	0.375
Neighboring seq. Comp.	InnerMembrane	0.954

CELLO Prediction:

InnerMembrane	2.675 *
Cytoplasmic	1.059
OuterMembrane	0.649
Extracellular	0.506
Periplasmic	0.110

Fig 1: Cello v2.5 results showing the cellular localization of the respective isoform (field marked as * indicates the cellular localization with the reliability)

TMHMM

The TMHMM tool predicted the total number of transmembrane helices (TMHs) to be 15 in all cases. Expected no. of amino acids in TMHs were found to be same in isoform A & C but differ in case of isoform D and expected no. of TMHs in first sixty amino acids present in the various isoforms of the protein were found to be 0. The total probability that the N-terminal is on the cytoplasmic side of the membrane for the three isoforms of krotzkopf verkehrt protein was found to be 0.992.

Table 2a: TMHMM result showing transmembrane helices and their position

Protein Name		No. Of Predicted TMHs	Expected No. of Aas TMHs	Expected No. of TMHs In First 60 Aas
Krotzko pf verkehrt protein	ISOFORM A	15	353.06638	0
	ISOFORM C	15	353.06638	0
	ISOFORM D	15	352.32925	0

TMHMM

1. # NP_524233.1 Length: 1615				
2. # NP_524233.1 Number of predicted TMHs: 15				
3. # NP_524233.1 Exp number of AAs in TMHs: 353.06638				
4. # NP_524233.1 Exp number, first 60 AAs: 0				
5. # NP_524233.1 Total prob of N-in: 0.99247				
6. NP_524233.1	7. TMHMM2.0	8. inside	9. 1	10. 88
11. NP_524233.1	12. TMHMM2.0	13. TMhelix	14. 89	15. 111
16. NP_524233.1	17. TMHMM2.0	18. outside	19. 112	20. 144
21. NP_524233.1	22. TMHMM2.0	23. TMhelix	24. 145	25. 164
26. NP_524233.1	27. TMHMM2.0	28. inside	29. 165	30. 182
31. NP_524233.1	32. TMHMM2.0	33. TMhelix	34. 183	35. 205
36. NP_524233.1	37. TMHMM2.0	38. outside	39. 206	40. 209
41. NP_524233.1	42. TMHMM2.0	43. TMhelix	44. 210	45. 232
46. NP_524233.1	47. TMHMM2.0	48. inside	49. 233	50. 315
51. NP_524233.1	52. TMHMM2.0	53. TMhelix	54. 316	55. 338
56. NP_524233.1	57. TMHMM2.0	58. outside	59. 339	60. 385
61. NP_524233.1	62. TMHMM2.0	63. TMhelix	64. 386	65. 408
66. NP_524233.1	67. TMHMM2.0	68. inside	69. 409	70. 420

71. NP_524233.1	72. TMHMM2.0	73. TMhelix	74. 421	75. 443
76. NP_524233.1	77. TMHMM2.0	78. outside	79. 444	80. 476
81. NP_524233.1	82. TMHMM2.0	83. TMhelix	84. 477	85. 496
86. NP_524233.1	87. TMHMM2.0	88. inside	89. 497	90. 923
91. NP_524233.1	92. TMHMM2.0	93. TMhelix	94. 924	95. 946
96. NP_524233.1	97. TMHMM2.0	98. outside	99. 947	100. 960
101. NP_524233.1	102. TMHMM2.0	103. TMhelix	104. 961	105. 980
106. NP_524233.1	107. TMHMM2.0	108. inside	109. 981	110. 986
111. NP_524233.1	112. TMHMM2.0	113. TMhelix	114. 987	115. 1006
116. NP_524233.1	117. TMHMM2.0	118. outside	119. 1007	120. 1015
121. NP_524233.1	122. TMHMM2.0	123. TMhelix	124. 1016	125. 1038
126. NP_524233.1	127. TMHMM2.0	128. inside	129. 1039	130. 1042
131. NP_524233.1	132. TMHMM2.0	133. TMhelix	134. 1043	135. 1065
136. NP_524233.1	137. TMHMM2.0	138. outside	139. 1066	140. 1287
141. NP_524233.1	142. TMHMM2.0	143. TMhelix	144. 1288	145. 1305
146. NP_524233.1	147. TMHMM2.0	148. inside	149. 1306	150. 1344
151. NP_524233.1	152. TMHMM2.0	153. TMhelix	154. 1345	155. 1367
156. NP_524233.1	157. TMHMM2.0	158. outside	159. 1368	160. 1615

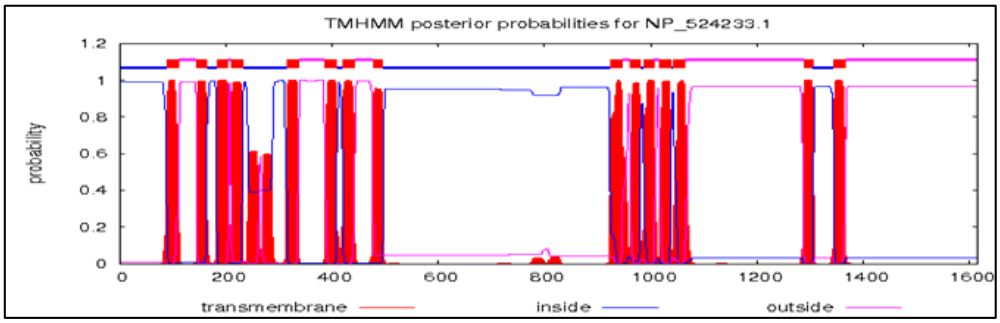


Fig 2a: Graphical view of transmembrane helices of isoform A

Table 2b: TMHMM result showing transmembrane helices and their position

TMHMM

NP_996152.1 Length: 1615
NP_996152.1 Number of predicted TMHs: 15
NP_996152.1 Exp number of AAs in TMHs: 353.06638
NP_996152.1 Exp number, first 60 AAs: 0
NP_996152.1 Total prob of N-in: 0.99247

NP_996152.1	TMHMM2.0	inside	1	88
NP_996152.1	TMHMM2.0	TMhelix	89	111
NP_996152.1	TMHMM2.0	outside	112	144
NP_996152.1	TMHMM2.0	TMhelix	145	164
NP_996152.1	TMHMM2.0	inside	165	182
NP_996152.1	TMHMM2.0	TMhelix	183	205
NP_996152.1	TMHMM2.0	outside	206	209
NP_996152.1	TMHMM2.0	TMhelix	210	232
NP_996152.1	TMHMM2.0	inside	233	315
NP_996152.1	TMHMM2.0	TMhelix	316	338
NP_996152.1	TMHMM2.0	outside	339	385
NP_996152.1	TMHMM2.0	TMhelix	386	408
NP_996152.1	TMHMM2.0	inside	409	420
NP_996152.1	TMHMM2.0	TMhelix	421	443
NP_996152.1	TMHMM2.0	outside	444	476
NP_996152.1	TMHMM2.0	TMhelix	477	496
NP_996152.1	TMHMM2.0	inside	497	923
NP_996152.1	TMHMM2.0	TMhelix	924	946
NP_996152.1	TMHMM2.0	outside	947	960
NP_996152.1	TMHMM2.0	TMhelix	961	980
NP_996152.1	TMHMM2.0	inside	981	986
NP_996152.1	TMHMM2.0	TMhelix	987	1006
NP_996152.1	TMHMM2.0	outside	1007	1015
NP_996152.1	TMHMM2.0	TMhelix	1016	1038
NP_996152.1	TMHMM2.0	inside	1039	1042
NP_996152.1	TMHMM2.0	TMhelix	1043	1065
NP_996152.1	TMHMM2.0	outside	1066	1287

NP_996152.1	TMHMM2.0	TMhelix	1288	1305
NP_996152.1	TMHMM2.0	inside	1306	1344
NP_996152.1	TMHMM2.0	TMhelix	1345	1367
NP_996152.1	TMHMM2.0	outside	1368	1615

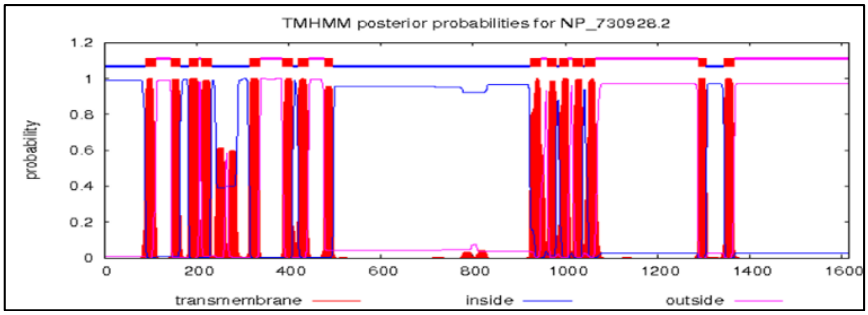
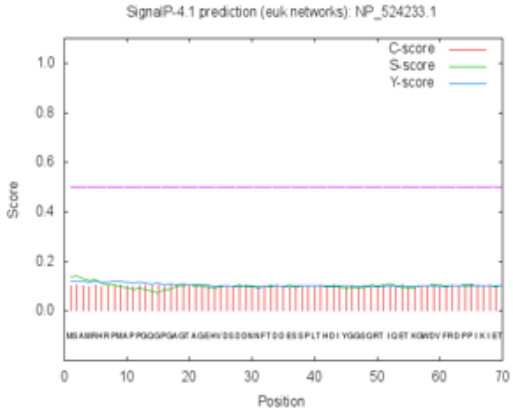


Fig 2c: Graphical view of transmembrane helices of isoform D

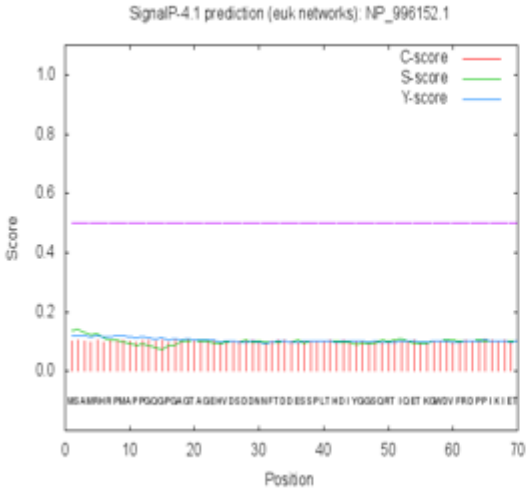
SignalP 4.0

The SignalP tool predicts whether the protein is a signal peptide or not. Fig. 3a to 3e shows the results of signal for the isoforms of chitin synthase as well as for the isoforms of krotzkopf verkehrt protein. Results clearly shows that none of the isoform of both the proteins is a signal peptide since the D-score (discrimination score) is lesser than the cutoff in each case.



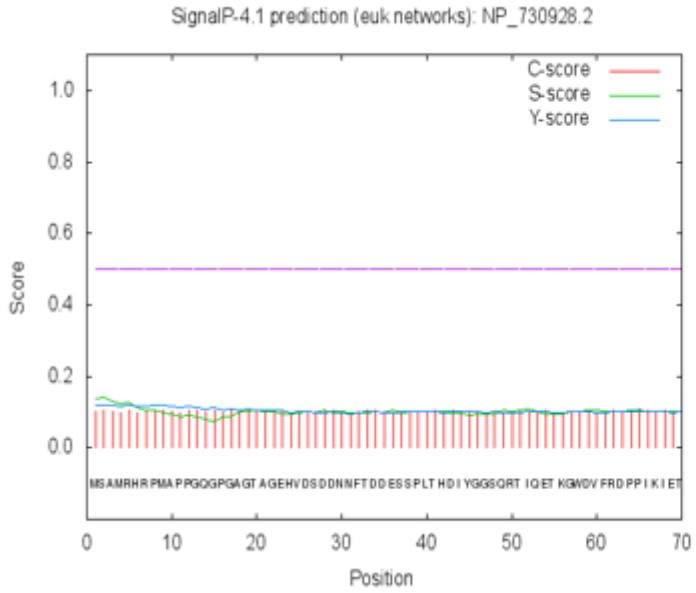
Measure	Position	Value	Cut off	signal peptide?
max. C	65	0.109		
max. Y	12	0.115		
max. S	2	0.142		
mean S	1-11	0.115		
D	1-11	0.115	0.450	NO

Fig 3a: SignalP graphical output showing C-score (raw cleavage site score), S-score (signal peptide score), Y-score (combined cleavage site score) for the isoform A



Measure	Position	Value	Cut off	signal peptide?
max. C	65	0.109		
max. Y	12	0.115		
max. S	2	0.142		
mean S	1-11	0.115		
D	1-11	0.115	0.450	NO

Fig 3b: SignalP graphical output showing C-score (raw cleavage site score), S-score (signal peptide score), Y-score (combined cleavage site score) for the isoform C

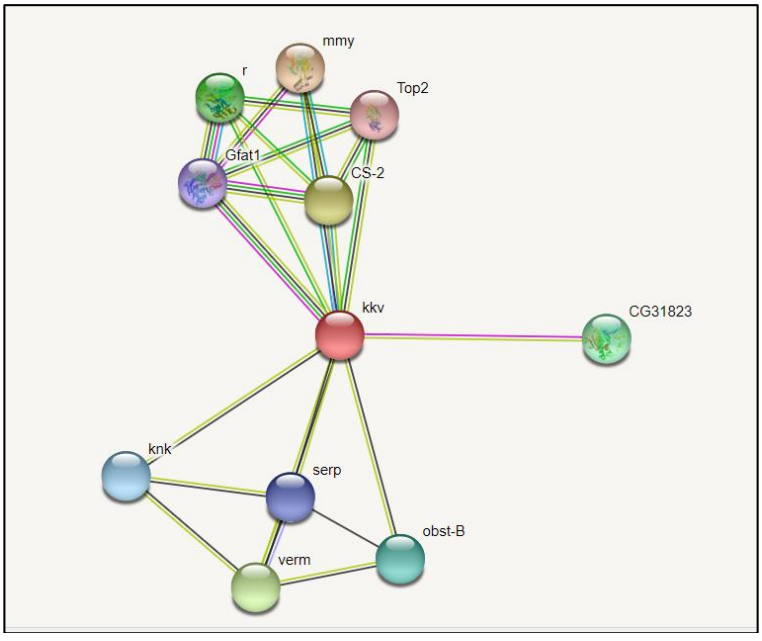


Measure	Position	Value	Cut off	signal peptide?
max. C	65	0.109		
max. Y	12	0.115		
max. S	2	0.142		
mean S	1-11	0.115		
D	1-11	0.115	0.450	NO

Fig 3c: SignalP graphical output showing C-score (raw cleavage site score), S-score (signal pepti de score), Y-score (combined cleavage site score) for the isoform D protein

STRING

STRING predicted that all three isoforms of kertzkopf verkehrt protein interacts with the same proteins such as Mummy having 520 amino acids, chitin synthase 2 having 1393 amino acids, vermiform, rudimentary and many other proteins as listed in fig. 4.



Your Input:

		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
kkv	*Krotzkopf verkehrt (1615 aa)*									

Predicted Functional Partners:

mmy	*Mummy (520 aa)*						●			0.926
CS-2	*Chitin synthase 2 (1393 aa)*						●	●	●	0.811
verm	*Vermiform (555 aa)*				●			●		0.694
r	*Rudimentary; This protein is a "fusion" protein encoding four enzymatic activities of the pyrimidine pathway (GATase, CPSase,...*	●						●		0.689
CG31823	*Serine-type carboxypeptidase activity (427 aa)*					●				0.684
obst-B	*obstructor-B (337 aa)*				●			●		0.593
knk	*Knickkopf (689 aa)*				●			●		0.582
serp	*Serpentine (541 aa)*				●			●		0.569
Gfat1	*Glutamine-fructose-6-phosphate aminotransferase 1 (694 aa)*	●			●	●		●		0.555
Top2	*Topoisomerase 2; Control of topological states of DNA by transient breakage and subsequent rejoining of DNA strands (Pub...*	●						●		0.519

Fig 4: STRING results showing the proteins with which isoforms of krotzkopf verkehrt protein interact

Secondary structure prediction

Secondary structure prediction was done using SOPMA. The results of secondary structure prediction is shown in fig. 5 to 7. Different secondary structures are colour coded with different colours in the sequence, alpha helix(h) with blue, extended strand(e) with red, beta turn (b) with green and random coil(c) with yellow colour in SOPMA. Predicted secondary structure in SOPMA showsthat isoform A & C of kertzkopf verkehrt protein consists of exactly same percentage of alpha helix, extended strand, beta turn and random coil, whereas the isoform D differs from the other two isoforms in percentage of alpha helix, extended strand and random coil. Higher number of helices makes the protein more flexible for folding that might increase interactions. There are no 3_{10} helix, Pi helix, Beta Bridge, bend region or ambiguous states in either of the isoform.

Fig 5a: SOPMA result showing secondary structure for isoform A of kertzkopf verkehrt protein. (Helix represented by 'h' in blue, random coil by 'c' in yellow, extended strand by 'e' in red and beta turn by 't' in green color.)

Fig 5b: Pie chart showing the percentage of alpha helix, extended strand, beta turn and random coil for isoform A of kertkzkopf verkehrt protein

Fig 6a: SOPMA result showing secondary structure for isoform C of kertzkopf verkehrt protein. (Helix represented by ‘h’ in blue, random coil by ‘c’ in yellow,
extended strand by ‘e’ in red and beta turn by ‘t’ in green color.)

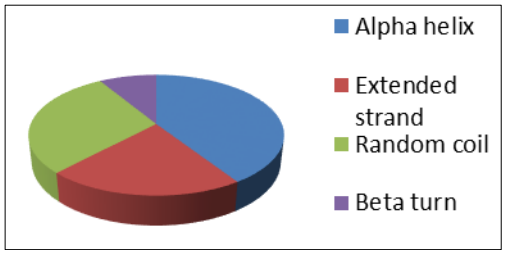


Fig 6b: Pie chart showing the percentage of alpha helix, extended strand, beta turn and random coil for isoform C of kertkzkopf verkehrt protein

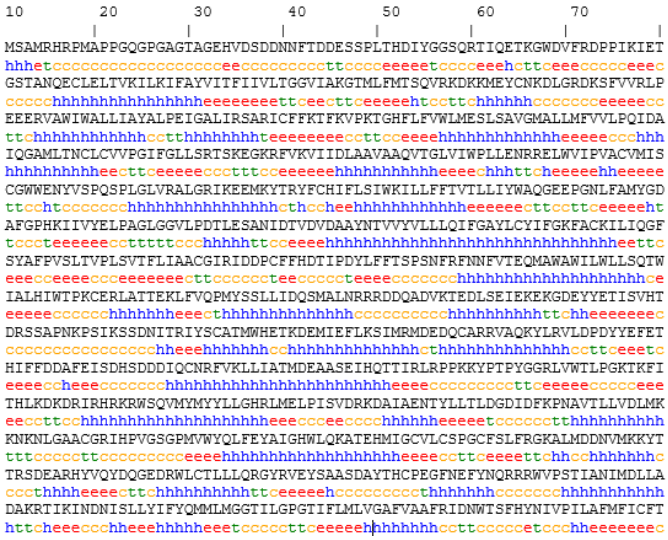


Fig. 7a: SOPMA result showing secondary structure for isoform D of kertzkopf verkehrt protein. (Helix represented by 'h' in blue, random coil by 'c' in yellow, extended strand by 'e' in red and beta turn by 't' in green color.)

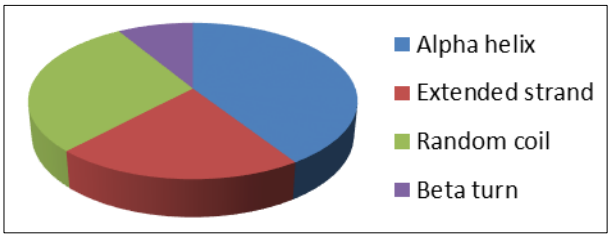


Fig 7b: Pie chart showing the percentage of alpha helix, extended strand, beta turn and random coil for isoform D of kertkzkopf verkehrt protein

Ab initio modeling

Ab initio modeling of isoform A, C & D was carried out using Phyre2 server. For the isoform C which contains 1392 amino acids 402 residues(29%) were modelled at >90% accuracy whereas for isoform D which consists of 1393 residues 403(29%) residues were modeled at >90% accuracy.

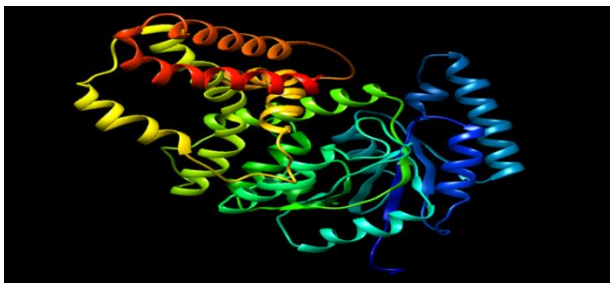


Fig 8a: *ab initio* model for isoform A

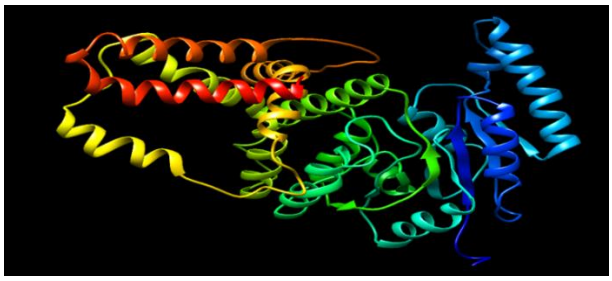


Fig 8b: *ab initio* model for isoform C

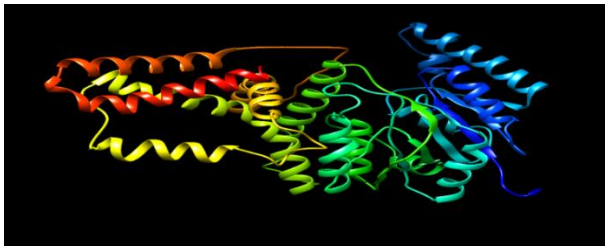


Fig 8c: *ab initio* model for isoform D

Conclusion

The current study is performing ab initio modeling of the isoforms of krotzkopf verkehrt protein from *Drosophila melanogaster* using *in silico* methods. The physiochemical properties were investigated using various in silico tools. All the isoforms were classified as inner membrane proteins with no signal peptide and their functional partners were also found to be same. Alpha helices and random coil were computed to be dominating in secondary structure of both the isoforms followed by extended strand.

Acknowledgement

Authors thank Department of Science and Technology (DST), Government of India, for supporting the research by providing INSPIRE fellowship.

References

1. Merzendorfer H, Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases, Journal of Experimental Biology. 2003; 206:4393-4412.
2. Xin Zhang, Jianzhen Zhang, Yoonscong Park, Kun Yan Zhu, Identification and characterization of two chtin synthase genes in African malaria mosquito, Anopheles gambiae, Insect Biochemistry and Molecular Biology. 2012; 42: 674-682
3. Merzendorfer H., The cellular basis of chitin synthesis in fungi and insects: Common principles and differences, European journal of Cell Biology. 2011; 90:759-769
4. Bhagath KP, Nagendra PS, Kasi VK, Sampath KR, Dinakara RA. UDP-N-acetylglucosamine pyrophosphorylase as novel target for controlling *Aedes aegypti* - molecular modeling, docking and simulation studies. International Journal of Mosquito Research. 2014; 1 (4):17-24.
5. Ostrowski S, Dierick HA, Bejsovec A. Genetic control of cuticle formation during embryonic development of Drosophila melanogaster. Genetics. 2002; 161:171-182
6. Peterson TN., Brunak S., Von Heijne G., Nielsen H., SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 2011; 8:785-786
7. Kelley LA, *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 2015; 10:845-858.