Biotechnological
Communication

Biosc.Biotech.Res.Comm Vol 12 (4) 1047-1062 (2019)

BBRC
Bioscience Biotechnology Research Communications

A Fuzzy Model for Minining Amino Acid Assosiations In Peptide Sequences of Flavivirus Sub Families

Priyanka Solanki[1*] Sowmya Dhawan[2], and Bhavna Kabra[3]
[1,3]Shri Vaishnav Institute of Management, Scheme No. 71, Gumasta Nagar, Indore,
[2]National Health Mission, M.P Bhopal,

ABSTRACT

The Genus Flavivirus cause significant human disease in the form of encephalitis or hemorrhagic fever. This genus of the family Flavi viridae comprises of 70 viruses, but vaccines are available for only yellow fever, Japanese and Tick Borne Encephalitis. Disease diagnosis can be difficult as all the members of Flaviviridae are antigenically and genetically closely related. Thus it is important to reveal relationships between amino acids and other parameters in molecular sequences of Flavivirus as it may assist in controlling of the diseases caused by these viruses. In this paper an attempt has been made to develop and explore a model for mining fuzzy amino acid association patterns in peptide sequences of Flavivirus and their relationships with secondary structures and physicochemical properties. The uncertainty arising due to variation in length of sequences and this is handled by employing fuzzy sets. A tool based on fuzzy approach was developed to find fuzzy amino acid association patterns by calculating support and confidence. It also calculates secondary structure and physicochemical properties of amino acid association patterns. Total 9160 sequences were taken from National Centre for Biotechnology Information. After that around 4004 non-redundant peptide sequences of Flavivirus subfamilies filtered to form the dataset. This dataset is transformed to fuzzy transaction dataset and their fuzzy support and confidence have been computed. The association patterns generated from this model can be useful in understanding the structure, function and interaction of the protein in the disease. This patterns generated may also be useful in gaining better insight about the structure and function of the genus leading to development of new vaccines.

KEY WORDS: *CONFIDENCE, FUZZY ASSOCIATION MINING, DATASET, SUPPORT, THRESHOLDETHANOL.*

Article Information:*Corresponding Author: solanki.priyanka28@gmail.com
Received 05/11/2019 Accepted after revision 29/12/2019
Published: 30th Dec 2019 Pp- 1047-1062

Published by Society for Science & Nature, Bhopal India.
Available at: https://bbrc.in/
Article DOI: http://dx.doi.org/10.21786/bbrc/12.4/28

INTRODUCTION

Flaviviruses of the family Flavi viridae are important arthropod-borne viruses in both human and veterinary medicine. The Flavivirus family contains many viral agents which produces encephalitis. Flavivirus encephalitis's are either mosquito- borne, tick-borne, or an unknown vector (Oya and Kurane 2007). Major symptoms include mild acute febrile syndromes, severe neurological, hepatic and hemorrhagic disease. The geographical diversity of Flavivirus has shown the occurrence of Japanese Encephalitis Virus (JEV) in Asia, causing menigo encephalitis in children and West Nile Virus (WNV) in West Africa, Middle East, and from 1999 in North America (Blitvich 2008). The overview of the host genes and variants on modify susceptibility or resistance to major mosquito-borne flaviviruses infections in mice and humans (Manet and Roth 2018). Mosquito-borne flaviviruses and their interactions with the innate immune response have been well-studied and reviewed extensively, thus this review will discuss tick-borne flaviviruses and their interactions with the host innate immune response (Lindqvist and Upadhyay 2018).

Flaviviruses exploit the ER function during infection to gain optimal replication. Multiple independent genome-wide screen studies have identified several ER-associated complexes and individual proteins that are important for flavivirus replication. Thus, these ER- complexes represent promising host targets for developing broad-spectrum anti-flavivirus drugs,(Rothan and Kumar 2019). The area of bioinformatics is known for association analysis, which is one of the most popular analysis paradigms in data mining (Gupta et al. 2009). The association rule mining has become one of the core task, and motivated tremendous interest among the data mining researchers and practitioners (Agrawal et al. 1995). The association rule mining research mainly focuses on discovery of patterns and algorithms. The first reported algorithm for finding frequent item sets is the Apriori algorithm (Agrawal and Srikant 1994).

Since then a good number of algorithms are reported in the literature for association rule mining. The traditional association rule mining algorithms lack in capability of handling inherent uncertainties present in the biological data. Thus there is high possibility of generation of over predicted or under predicted patterns in the data. The fuzzy set approach can be employed for mining association pattern in molecular sequences to overcome this challenge to some extent (Zadeh 1965). Association rules read the nature of different amino acids that are present in the protein. This very basic analysis provides understandings into the Co-occurrence of certain amino acids in a protein (Gupta Mangal et al. 2006). Attempts are also reported in the literature for mining associations in molecular sequences. In this paper an attempt has been made to explore fuzzy amino association patterns in peptide sequences of Flavivirus. To develop a model for mining amino acid association patterns in peptide sequences of MTBC has been discussed. The variation in the length of these sequences leads to variation in degree of relationship among amino acids present in each sequence. The fuzzy set is employed to model this uncertainty of degree of relationships among the amino acids of the peptide sequences of MTBC (Seth and Pardasani 2014).

An approach for mining fuzzy association patterns in peptide sequences of dengue virus employed to incorporate the degree of relationships among amino acids due to variation in length of the sequences. This approach is employed to incorporate the relationship of parameters with amino acid association patterns (Gour and Pardasani 2018). Analytical Study of Data Mining Applications in Malaria Prediction and Diagnosis. This study shows the large number of deaths occur annually as a result of many factors which include shortages of medical personnel, laboratory equipment, hospitals and wrong interpretation of laboratory results. It also established the fact that remote areas are majorly affected. The fusion of Medical Science and Computer Science (Information Technology) in managing deadly diseases as a result of the earlier mentioned challenges was also established. This collaboration has led to development of computer based predictive models in medical diagnosis and treatment (Boruah and Kakoty 2019). Protecting the Privacy of Cancer Patients Using Fuzzy Association Rule Hiding, a novel method was presented to hide the sensitive rule in quantitative data by decreasing the support of the RHS of the rule. Experimental results demonstrate that the proposed approach is more efficient as it facilitates better rule hiding and minimizes the number of lost rules and ghost rules. Also, this approach makes minimum modifications to the dataset (Krishnamoorthy and Murugesan 2018, Hussain and Kumar 2019).

In this study a Java EE platform based tool was developed of for studying of molecular sequences. The main feature of tool is its accuracy and intelligence in generating the results. The main aim is to analyze the fuzzy associations between various frequent patterns occur to handle upcoming challenges of uncertainty. The available bioinformatics tools provide information only about the secondary structure and physicochemical properties of entire peptide sequences without using any parameter like length of the sequences, length range of the sequences, creating difficulty in critical analysis due to under prediction and over prediction of the rules. The divergence and convergence of association patterns within the Flavivirus subfamilies is analyzed to generate the association rules. The results generated are also correlated with structural and physicochemical properties.

MATERIAL AND METHODS

Description of the algorithm employed is as follows: In this paper we have taken molecular data of Flavivirus subfamilies like: mosquito borne, Tick borne, Known vector from NCBI. To calculate the Fuzzy frequent patterns in redundant and non-redundant dataset of Flavivirus subfamilies, the fuzzy membership of amino acid in respective sequence is calculated as

$$\mu_i(A)=\sum_i^n f_i(A)/L_i \qquad (1)$$

Where $\sum_i^n f$ (A) is the frequency of amino acid A and μ_i (A) is the membership of amino acid A in the i^{th} sequence. It is assumed that there are 20 amino acids and each amino acid will have equal likely chance of appearing in a sequence. Thus the threshold value can be calculated as:

$$T=0.05*N \qquad (2)$$

Here N is the Number of sequences.
The apriori algorithm is employed to find frequent patterns in all the sequences. These patterns are used to generate association rule .The Fuzzy Support from amino acid can be calculated as: The frequency Support for n amino acid can be calculated as:

$$\sum \mu_i(A1 \cap A2 \cap A3 \cap \ldots\ldots\ldots\ldots\ldots A_{n-1} \cap A_n) \qquad (3)$$

Confidence for n amino acid can be calculated by:

$$\frac{\sum_i^L \mu i(A1 \cap A2 \cap A3 \cap \ldots\ldots\ldots\ldots An-1 \cap An)}{\sum_i^L \mu i(A1 \cap A2 \cap A3 \cap \ldots\ldots\ldots\ldots An-1)} \qquad (4)$$

RESULTS AND DISCUSSION

After applying the fuzzy approach for finding

Step 1:

Sequence number (Transactions)	Amino Acids(Item Set) a_1 a_2 .... a_{19} a_{20}	Length of Sequence (Parameter)
s_1	$a_{1,1}$ $a_{1,2}$ .... $a_{1,19}$ $a_{1,20}$	l_1
s_2	$a_{2,1}$ $a_{2,2}$ .... $a_{2,19}$ $a_{2,20}$	l_2
.	.	.
s_n	$a_{n,1}$ $a_{n,2}$ .... $a_{n,19}$ $a_{n,20}$	l_n

Step 2:

Sequence number (Transactions)	Membership of Amino Acids(Item Set) a_1 a_2 .... a_{19} a_{20}	Sequence number (Transactions)
s_1	$\mu_1(a_1)$ $\mu_1(a_2)$ .... $\mu_1(a_{19})$ $\mu_1(a_{20})$	s_1
s_2	$\mu_2(a_1)$ $\mu_2(a_2)$ .... $\mu_2(a_{19})$ $\mu_2(a_{20})$	s_2
.	.	.
s_n	$\mu_n(a_1)$ $\mu_n(a_2)$ .... $\mu_n(a_{19})$ $\mu_n(a_{20})$	s_n

Step 3:

Sequence number	Amino Acids(Item Set) a_1 a_2 .... a_{19} a_{20}	Length of Sequence (Parameter)
s_1	$a_{1,1}$ $a_{1,2}$ .... $a_{1,19}$ $a_{1,20}$	l_1
s_2	$a_{2,1}$ $a_{2,2}$ .... $a_{2,19}$ $a_{2,20}$	l_2
.	.	.
s_n	$a_{n,1}$ $a_{n,2}$ .... $a_{n,19}$ $a_{n,20}$	l_n

the maximum and minimum frequency of each amino acid for all subfamilies of Flavivirus, it has been found that there are variations in frequent amino acid for redundant and non-redundant data set of all three subfamilies of Flavivirus :-mosquito borne, tick borne and known vector subfamilies. G(glycine) is most frequent amino acid for Japanese encephalitis, St. Louis encephalitis , West Nile, Louping ill and Summar encephalities while L((luecine) is most frequent for Marry Valley, Ilheus, ,Central European, Russion Spring-Rodant, Ricio and Rio Bravo whereas V(valine) amino acid is most frequent for Powassion. Amino acid C (cystein) is least frequent in Marry Vally, Ilheus , West Nile, Louping Ill, Russin Spring-Rodant, Summar encephalitis, Ricio and Rio Bravo whereas amino acid H(Histidine) is least frequent in Japanese encephalitis, St. Louis encephalitis, and Powassion ; and W(Tryptopher) is least frequent for Central European. Table 1 shows the maximum support for frequent amino acid among the subfamilies of mosquito borne family of Flavivirus. A, G, L, T and S (Alanine, Glycine, Leucine, Threonine

Figure 1. Algorithm of the method for finding frequent pattern

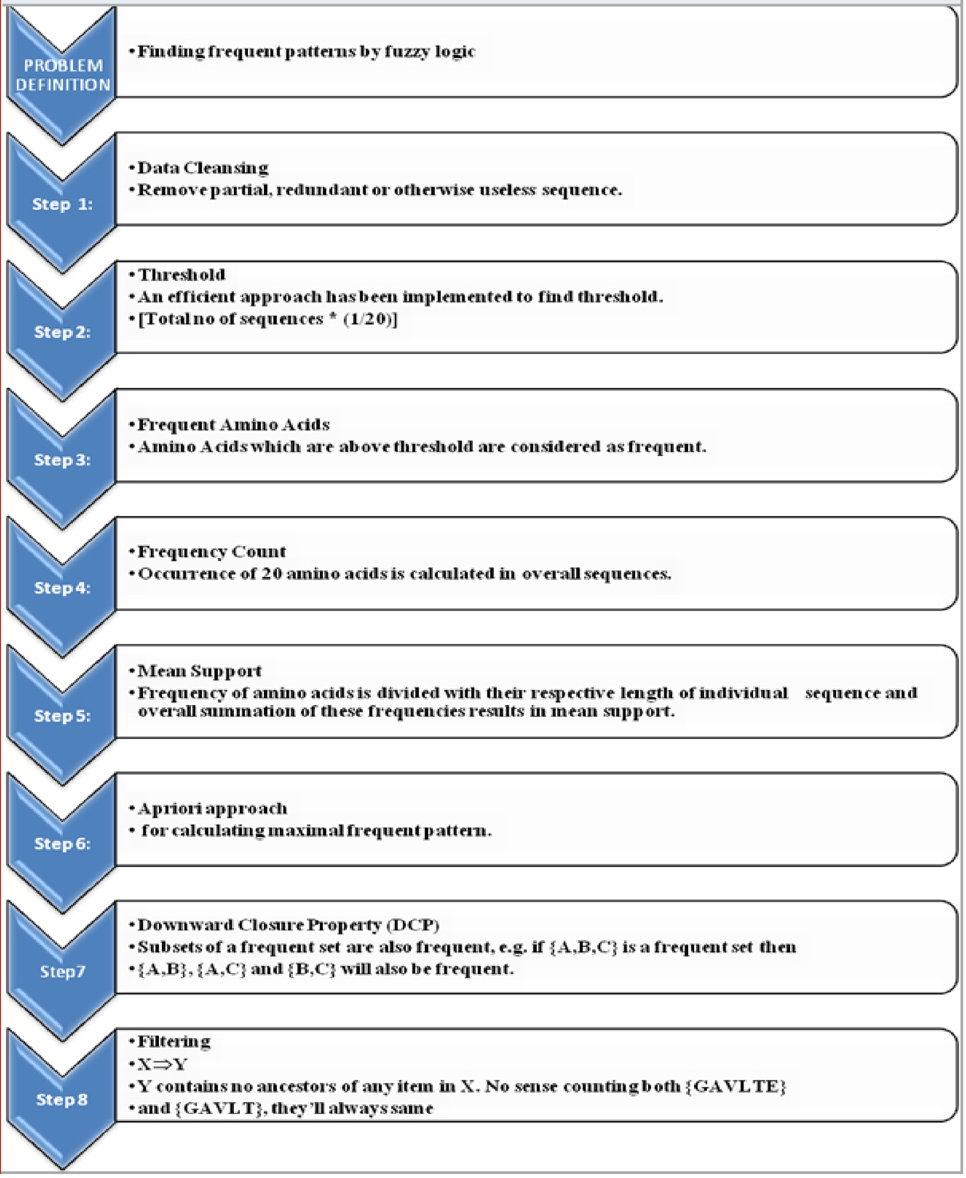


Table 1. Maximum support in case of Mosquito Borne									
Japanese encephalitis		Marry Vally encephalitis		St. Louis encephalitis			Ilheus		West Nile
R	Non-R	R	Non-R	R	Non-R	R	Non-R	R	Non-R
S OF SIX FP [G, A, L, T, V, S]	S OF SIX FP [G, A, L, T, V, S]	S OF EIGHT FP[E, G, A, L, K, T, V, S]	S OF SEVEN FP[E, G, A, L, K, T, V, S]	S OF SEVEN FP[G, A, L, K, T, V, S]	S OF SEVEN FP[E, G, A, L, K, T, V, S]	S OF FIVE FP[F, G, A, L, I, T, S]	S OF FIVE FP[G, A, L, I, T, V, S	S OF SIX FP[G, A, L, T, V, S]	S OF FIVE FP[G, A, L, T, V, S]
GALTVS	GALT	EGALKT	GALK TVS	GALKTVS	GALKTVS	GALIS	GALIS	GALTVS	GALTV
S= 127.35	VS S = 55.66	VS S= 5.29	S = 3.01	S = 13.93	S = 9.30	S= 8.22	S=4.28	S= 203.74	S= 115.42
TOTAL NO OF SIX	TOTAL NO OF SIX	TOTAL NO OF EIGHT	TOTAL NO OF SEVEN	TOTAL NO OF SEVEN	TOTAL NO OF SEVEN	GALTS S = 8.16	TOTAL NO OF FIVE	TOTAL NO OF SIX	TOTAL NO OF FIVE
FP: 1 -	FP: 1 -	F P: 1 -	FP: 1 -	FP: 1 -	FP: 1 -	GLITS	FP : 1	FP : 1	FP: 1
-	-	-				S= 8.25			
- -	- -	- -	-	-	-	ALITS S= 8.09			
-	-	-	-	-	-	TOTAL NO OF FIVE FP: 4	-	-	-
*S= Support, FP= Frequent Patterns									

Table 2. Maximum support in case of Tick Borne									
Central European (4033)			Louping Ill (45)		Powassan (126)		Russian Spring-Rodents (7)		Summer encephalitis 2)
R	Non-R	R	Non-R	R	Non-R	R	Non-R	R	Non-R
S OF FOUR FP [G, A, L, I, T, V, S]	S OF FOUR FP [G, A, L, I, T, V, S]	S OF FIVE FP [E, G, A, L,K, T, V, S, R]	S OF FIVE FP [E, G, A, L, K, T, V, S, R]	S OF SIX FP [D, E, G, A, L, K, T, V, S]	S OF SIX FP [E, G, A, L, K, T, V, S, R]	S OF SIX FP [G, A, L, I, T, V]	S OF FIVE FP [E, A, L, I, K, T, V, R]	S OF EIGHT FP [D, E, G, A, L, V, S, R]	S OF EIGHT FP [D, E, G, A, L, V, S, R]
GALI S = 203.74	GALV S = 127.43	EGALS S = 2.30	EGALS S = 1.76	EGATVS S = 6.32	EGALVS S= 3.81	GALITV S = 0.36	EALKR S = 0.29	DEGAL VSR S = 0.10	DEGAL VSR S =0.10
GALV S = 218.84	GALS S=123.58	GALTV S= 2.86	GALTV S= 2.22	EAKTVS S= 6.53	TOTAL NO OF SIX FP: 1	TOTAL NO OF SIX FP: 1	ALITV S = 0.26	TOTAL NO OF EIGHT FP: 1	TOTAL NO OFEIGHT FP: 1
GALS SUPP ORT= 209.47	ALVS SUPPORT = 121.30	TOTAL NO: OF FIVE FP:2	TOTAL NO OF FIVE FP: 2	GAKTVS S = 6.40	-	-	TOTAL NO OF FIVE FP: 2	-	-
TOTAL NO OF FOUR FP: 3	TOTAL NO OF FOUR FP:3	-	-	ALKTVS S = 6.30	-	-	-	-	-
-	-			TOTAL NO OF SIX FP: 4	-	-	-	-	-

and Serine) are frequent for all subfamilies. V (Valine) is also frequent for all subfamilies except of Ilheus virus for non-redundant dataset. E (Glutamic) is frequent in Marry Valley for both redundant and non-redundant dataset and in St. Louis encephalitis (non-redundant dataset). K (Lysine) is frequent in marry valley and St. Louis encephalities. I (Isoleucine) is frequent for Ilheus. F (Phenylalanine) is frequent for redundant dataset of Ilheus virus. Table 2 represents maximus support found for tick borne subfamily. In tick borne subfamily maximum support for frequent amino acid are A, L, V (Alanine, Leucine, Valine) for all subfamilies. G(Glycine) is frequent for all subfamilies except the Russion Spring-Rodant non-redundant dataset. T (Threonine) is also frequent for all subfamilies except Summer encephalitis. S (Serine) is frequent for all subfamilies except Russion Spring-Rodent dataset. E (Glutamic) is frequent for Louping Ill, Powassan , Summer encephalitis and Russion Spring-Rodent (non-redundant datase). K(Lysine) is frequent for Louping Ill, Powasson and Pussion Spring-Rodent (non-redundant dataset),R is frequent for Louping Ill, Powassan (non-redundant dataset) and Russion Spring-Rodent(non-redundant dataset), I is frequent for Central european and Russion Spring-Rodent, D (Aspartic) is frequent for Summar encephalitis and Powassan (redundant dataset).

Table 3 depicts that in known vector subfamilies of Flavivirus; L,I,T,V,S (Leucine, Isoleucine, Threonine, Valine, Serine) are frequent for all subfamilies, G(Glycine) and A(Alanine) are frequent for Rico and Rio-bravo (redundant dataset), S(Serine) is also frequent for Racio and K(Lysine) is frequent for Rocio non-redundant dataset. Kumari and Pardasani (2013,14), have applied the same method in their research but with different dataset i.e., GPCRs. Table 4 shows the probable structure (helix, beta, and coil) and physicochemical properties of Favivirus subfamilies: Mosquito Borne, Tick Borne and

Table 3. Maximum support in case of Known Vector

Rocio R	Non-R	Rio Bravo R	Non-R
S OF THREE FP [G, A, P, S] I, T, V, P, S] GLI S = 18.40 GLT S = 17.54	S OF THREE FP [G, A, L, I, K, T, V, P, S] GLI S= 16.64 GLT S = 15.79	SOF FOUR FP [G,A, L, I,T, V,S] LIVS S = 2.43 TOTAL NO OF FOUR FP: 1	S OF FOUR FP [L, I, T, V, S] LIVS S = 1.75 TOTAL NO OF FOUR FP: 1
GLV S = 18.43	GLV S = 16.62	-	-
GLP S = 17.86	GLP S = 16.20	-	-
GLS S = 17.45	GLS S = 15.68	-	-
TOTAL NO OF THREE FP:5	TOTAL NO OF THREE FP:5	-	-

Known Vector. The observation reveals that the most of the amino acids G, A, L (Glycine, Alanine, Leucine) are common in all subfamilies and Helix is the Probable Structure in maximum subfamilies. The physicochemical properties like hydrophobicity, CBetaBranched, polar aliphatic and uncharged, non-polar aliphatic groups are common in all subfamilies. Table 5 shows the probable helix structure of protein for Flavivirus subfamilies based on amino acid associations. Amino acids like A, R, E, Q, L, K, M, H are responsible for helix structure formation. It has been revealed that the maximum frequent patterns for helix formation are 4-frequent patterns. Table 6 shows the probable sheet structure of protein for Flavivirus subfamilies based on amino acid associations. Amino acids like V, I, T, C, W, F, Y are responsible for secondary sheet structure formation.

It has been observed that the maximum frequent patterns are 3 for secondary structure of sheet formation i.e. ITV. Table 7 presents the secondary structure of coil formation in which almost all subfamilies have 2-frequent patterns GS which are responsible for formation of secondary structure. It has been observed in Table 5, 6 and 7 that in all the 12 subfamilies of Flavivirus association patterns of amino acid exposed high tendency to form secondary structure Helix rather than Sheet and Coil. Table 8 and Figure 1 and 2 depict the percentage wise calculations of physicochemical Properties of mosquito borne subfamilies. It reveals that west Nile virus (non-redundant dataset) have high percentage of Molecular weight and Extension Coefficient among all the subfamilies of mosquito borne.

Redundant dataset of West Nile virus have shown higher tendency to form a secondary structure Coil (29.135%) among all the subfamilies of Mosquito Borne. Marry Vally has shown high tendency of Absorbance among all the subfamilies. All the subfamilies of Mosquito borne viruses show negative hydrophobicity. Illeus virus has high percentage of aliphatic index and Aromaticity among all the subfamilies of mosquito borne virus. C-Beta Sheets are higher in Japanese encephalitis redundant dataset. Protein stability is high for Marry Vally virus in redundant and non-redundant dataset among all the subfamilies of mosquito borne. In Louis encephalitis both (redundant and non-redundant) datasets are showing high Salt Bridged, Positive Charged, and

Table 4. Probable Structures and physicochemical Properties of Protein Sequences of Sub families

Subfamily	R F-Amino Acid	Non-R Probable Structure	Physicochemical Properties	F-Amino Acid	Probable Structure	Physicochemical Properties
Mosquito Borne Subfamilies						
Japanese encephalitis	G, A, L, T, V, S	Helix, Sheet and Coil	Polar aliphatic (G), polar uncharged(S, T), non-polar aliphatic(A,L,V) and hydrophobic (G,L,V), CBeta Branched(T,V)	G, A, L, T, V, S	Helix	Acidic Negative charged protein stable (E), Polar aliphatic (G),polar uncharged(S,T), non-polar aliphatic(A,L,V) and hydrophobic (G,L,V), CBeta Branched(T,V)
Marry Vally encephalitis	E, G, A, L, K, T, V, S	, Helix	Acidic Negative charged protein stable(E), Polar aliphatic (G), non-polar aliphatic(A,L,V), and polar uncharged(T,S), basic charged (K) and hydrophobic (G,L,V), CBeta Branched(T,V)	E, G, A, L, K, T, V, S	Helix	Acidic Negative charged protein stable (E), Polar aliphatic hydrophobic (G), non-polar aliphatic(A,L,V), polar uncharged (T,S),basic charged(K) and hydrophobic(G,L,V), CBetaBranched (T,V)
St. Louis encephalitis	F,G, A, L, I, T, V,S	Helix & Sheet	Aromatic, Aliphatic(F) Polar aliphatic (G),polar uncharged (T,S),non-polar aliphatic (A,L,V,I) and hydrophobic (G,L,V,F), CBetaBranched(T,V)	G, A, L, I, T, V,S	Helix	Polar aliphatic (G), polar uncharged(T,S), non-polar aliphatic(A,L,V,I) and hydrophobic(G,L,V), CBetaBranched(T,V)
Ilheus	G, A, L, T, V, S	Sheet	Polar aliphatic (G), polar uncharged(S, T), non-polar aliphatic(A,L,I,V) and hydrophobic(G,L,V) CBetaBranched(T,V)	G, A, L, T, V, S	Sheet	Polar aliphatic (G), polar uncharged(S, T), non-polar aliphatic(A,L,V) and hydrophobic(G,L,V), CBetaBranched(T,V)
West Nile	G, A, L, T, V, S	Helix	Polar aliphatic (G), non-polar aliphatic(A,L,V), and polar uncharged(T,S) and hydrophobic(G,L,V), CBetaBranched(T,V),	G, A, L, T, V, S,	Helix	Polar aliphatic (G), non-polar aliphatic (A,L,V), and polar uncharged(T,S) and hydrophobic(G,L,V), CBetaBranched (T,V),
Tick Borne Subfamilies						
Central European	G, A, L, T, V, S	Helix, Sheet and Coil	Polar aliphatic (G), polar uncharged(S, T), non-polar aliphatic(A,L,V) and hydrophobic(G,L,V), CbetaBranched(T,V)	G, A, L, T, V, S	Helix	Acidic Negative charged protein stable (E), Polar aliphatic(G), polar uncharged (S,T),non-polar aliphatic(A,L,V) and hydrophobic(G,L,V), CbetaBranched(T,V)
Louping Ill	E, G, A, L, K, T, V, S, R	Helix	Acidic Negative charged protein stable(E), Polar aliphatic (G), non-polar aliphatic(A,L,V), and polar uncharged(T,S), basic charged(K) and hydrophobic(G,L,V), CbetaBranched(T,V), basic positive charged protein stable ®	E, G, A, L, K, T, V, S, R	Helix	Acidic Negative charged protein stable (E), Polar aliphatic hydrophobic (G), non-polar aliphatic (A,L,V), polar uncharged (T,S),basic charged (K) and hydrophobic(G,L,V), CbetaBranched (T,V), basic positive charged protein stable ®
Powassan	D,F,G, A, L, K, T, V,S	Helix & Sheet	Acidic Negative charged protein stable(E), Polar aliphatic (G),polar uncharged(T,S), non-polar aliphatic(A,L,V) and hydrophobic(G,L,V), CbetaBranched(T,V), basic charged(K), Acidic Charged (D)	E , G, A, L, K, T, V,S , R	Helix	Acidic Negative charged protein stable(E),Polar aliphatic (G),polar uncharged(T,S),non-polar aliphatic(A,L,V) and hydrophobic(G,L,V), CbetaBranched(T,V),

Russian Spring-Rodents	G, A, L, I, T, V	Sheet	Polar aliphatic (G), polar uncharged(T), non-polar aliphatic (A,L,I,V) and hydrophobic (G,L,V) CbetaBranched(T,V	E, A, L, I, K, T, V, R	Sheet	Basic charged amino acid®, basic charged(K), Acidic Negative charged protein stable(E),polar uncharged(S,T), non-polar aliphatic(A,L,I,V) and basic charged(K), hydrophobic (G,L,V), CbetaBranched (T,V), Basic charged amino acid®
Summer encephalitis	D,E, G, A, L, T, V, S,R	Helix	Acidic Charged (D), Acidic Negative charged (E) ,Polar aliphatic (G), non-polar aliphatic (A,L,V), and polar uncharged(T,S) and hydrophobic(G,L,V), CbetaBranched(T,V), Basic charged amino acid®, protein stable(D,R,E)	D,E, G, A, L, T, V, S,R	Helix	Acidic Charged (D), Acidic Negative charged protein stable (E),Polar aliphatic (G), non-polar aliphatic (A,L,V), and polar uncharged(T,S) and hydrophobic(G,L,V), CbetaBranched(T,V), Basic charged amino acid®, protein stable(D,R,E)
Known Vector subfamilies						
Rocio	G, A, L, I, T, V, P, S	Coil and Helix	Polar aliphatic (G), polar uncharged(T,S), non-polar aliphatic (A,L,I,P,V) and hydrophobic (G,L,V), CBeta Branched(T,V)	G, A, L, I, K, T, V, P, S (A,I,L,V,P)	Helix and Coil	Polar aliphatic (G), polar uncharged(S,T), non-polar aliphatic and, hydrophobic (G,L,V), CBetaBranched (T,V),basic charged (K)
Rio Bravo	G, A, L I,T, V,S	Helix, Coil	Polar aliphatic (G), polar uncharged(S,T), non-polar aliphatic (A,L,I,V), and hydrophobic (G,L,V), CBetaBranched(T,V)	L, I, T, V,S	Coil	non-polar aliphatic(L,I,V), polar uncharged(T,S) and hydrophobic(L,V), CBetaBranched(T,V)

Negative Charged parameters. Polarity is high in Marry Vally among all the subfamilies of mosquito borne of Flavivirus. Marry Vally shows the high tendency of Helix formation with respect to other subfamilies of Mosquito borne. Beta Sheet formation tendency is high in Japanese encephalitis redundant dataset than other subfamilies. Table 9 and Figure 3 and 4 present the physicochemical properties of tick borne subfamilies ,it has been found that non redundant dataset of Russian Spring-Rodent have high percentage of molecular weight ,extension coefficient ,positive charge, negative charge, salt bridged, polarity, protein stability and helix formation among all the subfamilies of Tick borne subfamilies, Central European redundant dataset are having high percentage of hydrophaticity, aromalicity and beta sheet formation. Among all subfamilies of tick borne, coil formation is high in non redundant dataset of Central European. Absorbance is high in Louping ill redundant dataset, C-Beta branched is high in Powassion, aliphatic index is high in Russian Spring-Rodent among all subfamilies of Tick Borne virus. Table 10 and Figure 5 & 6 present some physicochemical properties of known vector subfamilies. It has been found that molecular weight, aliphatic index, aromaticity, extension coefficient absorbance , C-Beta branched and helix formation are high in Rio bravo(redundant dataset) and beta sheet formation tendancy is high in Rio bravo non-redundant dataset. Hydrophaticily, polarity, salt bridged and coil formation are high in Rocio (redundant dataset) and protein stability and positive-negative charge are high in non-redundant dataset of Rocio. Some of the researcher have applied the same method in their research but with different dataset like Shankar and Pardasani (2013) worked on the dataset Apphaproteo bacteria; Seth Pardasani (2015), worked on the dataset MTBC; Gour and Pardasani (2018) studied the dataset Dengue Virus. The mathematical expressions represent the degree of relationships among amino acids in peptide sequences of flavivirus subfamilies and association relationships among amino acids in peptide sequences of flavivirus subfamilies(mosqitu born, tike born and known vector). These relationships are characterized by fuzzy membership, fuzzy support and fuzzy confidence. The relationships are interpreted in terms of associations rules of amino acids in peptide sequences.

The association rules generated on the basis of above Flavivirus subfamilies results are given below:-

Mosquito Borne Subfamilies:-
For 2 frequent Patterns:-

1. {A(Frequent)$\cap$L(Frequent)=>Tendency for Helix Formation}
2. {A(Frequent)$\cap$E(Frequent)=>Tendency for Helix

Table 5. Probable Helix Structure of Protein of Flavivirus Subfamilies based on amino acid association

Subfamily	Helix Formation (A,R,E,Q,L,K,M,H)							
	R				Non-R			
	1-FP	2- FP	3- FP	4- FP	1- FP	2- FP	3- FP	4- FP
Japanese encephalitis	A,E,L	AL	None	None	A,E,L,K	AL	None	None
Marry Vally encephalitis	A,L, E,K,R	AL,A E,AK, LE,LK, EK,ER	EAL, EAK, ELK, ALK ER	EALK	A,L, E,K,R	AL,AE, AK,LE, LK,EK,	EAL, EAK, ELK, ALK,	EALK
St. Louis encephalitis	A,L, E,K	AL,AE, AK,LE, LK,EK	ALK	None	A,L, E,K	AL,AE, AK,LE, LK,EK	EAL, EAK, ELK, ALK	EALK
Ilheus	A,L	AL	None	None	A,L	AL	None	None
West Nile	A,L, K	AL, LK	None	None	A,L, E,K	AL, LK	-	-
Central European	A,L, E,K	AL	None	None	A,L, E,K	AL	None	None
Louping Ill	A,L, E,K,R	AL,AE ,AK LE,LK, EK,AR. LR	EAL, ALK, ALR	None	A,L, E, K,R	AL, AE, AK LE,LK, EK, AR.LR	EAL, ALK, ALR	None
Powassan	A,L, E,K,R	AL, AE, AK, AR, LE,LK, LR , EK,ER	EAL, EAK, ALK, ALR	None	A,L, E,K,R	AL,AE, AK, LE,LK, EK, AR,LR	EAL, EAK, ELK, ALK	None
Russian Spring-Rodents	A,L, E,K	AL,AE, EK,LE	EAL	None	A,L, E,K,R	AL,AE, AR,AK, LE,LK, LR,EK, ER,KR	EAL, EAK, ELR, EKR, ALK, ALR, AKR, LKR	EALK, EALR, EAKR, ALKR
Summer encephalitis	A,L, R,E,K	AR,AK, AE,,LE, LK,LR, ER,AL	EAL, EAR, ELR, ALK, ALR	EALR	A,L, R,E,K	AR, AK, AE,,LE, LK,LR, ER,AL	EAL, EAR, ELR, ALK, ALR	-
Rocio	A,L,R,K	AL	None	None	A,L, K	AL,LK	None	None
Rio Brivo	A,L,K	AL,LK	None	None	A,L,K	AL,LK	None	None

Formation}

3. {E(Frequent)∩K(Frequent)=>Tendency for Helix Formation and Protein Solubility}
4. {E(Frequent)∩R(Frequent)=>Tendency for Helix Formation and Protein Solubility }
5. {V(Frequent)∩T(Frequent)=>Tendency for Sheet Formation}
6. G(Frequent)∩S(Frequent)=>Tendency for Coil Formation}

For 3 frequent Patterns:-

1. { E(Frequent)∩A(Frequent)∩L(Frequent)=>Tendency for Helix Formation}
2. {A(Frequent)∩L(Frequent)∩K(Frequent)=> Tendency for Helix Formation}

Table 6. Probable Sheet Structure of Protein of Flavivirus Subfamilies based on amino acid association

Subfamily	Sheet Formation (V,I,T,C,W,F,Y)							
	R				Non-R			
	1- FP	2- FP	3- FP	4- FP	1- FP	2- FP	3- FP	4- FP
Japanese encephalitis	V,T,I	VT,VI	None	None	V,T,I	VT,VI	None	None
Marry Vally encephalitis	V,T	VT	None	None	V,T	VT	None	None
St. Louis encephalitis	V,T,I	VT	None	None	V,T	VT	None	None
Ilheus	V,F,I,T	VI,VT,IF,IT	None	None	V,I,T	VI,IT	None	None
West Nile	V,T	VT	None	None	V,T	VT	None	None
Central European	V,T,I	VT,VI	None	None	V,T,I	VT	None	None
Louping Ill	V,T	VT	None	None	V,T	VT	None	None
Powassan	V,T	VT	None	None	V,T	VT	None	None
Russian Spring-Rodents	V,I,T	VI,VT,IT	ITV	None	V,I,T	VI,VT,IT	ITV	None
Summer encephalitis	V,T	VT	None	None	V,T	VT	None	None
Rocio	V,T,I	VT,VI	None	None	V,T,I	VT	None	None
Rio Brivo	V,T,I	VI,IT	None	None	V,T,I	VI,IT	None	None

Table 7. Probable Coil Structure of Protein of Mosquito Borne Subfamilies based on amino acid association

Subfamily	Coil (N,D,P,S,G)							
	R	Non-R						
	1- FP	2- FP	3- FP	4- FP	1- FP	2- FP	3- FP	4- FP
Japanese encephalitis	G,S,D	GS	none	none	G,S	GS	None	None
Marry Vally encephalitis	G,S	GS	None	none	G,S	GS	None	None
St. Louis encephalitis	G,S	GS	None	None	G,S	GS	None	None
Ilheus	G,S	GS	None	None	G,S	GS	None	None
West Nile	G,S	GS	None	None	G,S	GS	None	None
Central European	G,S	GS	none	none	G,S,D	GS	None	None
Louping Ill	G,S	GS	None	none	G,S	GS	None	None
Powassan	G,S,D	GS	None	None	G,S,D	GS	None	None
Russian Spring Rodents	G,S	None	None	None	G,S	G	None	None
Summer encephalitis	G,S,D	GS,GD, SD	DGS	None	G,S,D	GS,GD ,SD	DGS	None
Rocio	G,S,P	GS,GP	none	none	G,S,P	GS,GP	None	None
Rio Brivo	G,S	GS	None	none	G,S	None	None	None

Table 8. Physicochemical Properties of Mosquito Borne Subfamilies										
Physicochemical Properties/ Parameters	Japanese encephalitis	Marry Vally encephalitis		St. Louis encephalitis		Ilheus	West Nile			
	R	Non-R	R	Non-R	R	Non-R	R	Non-R	R	Non-R
MOLECULAR WEIGHT	55938.88	97201.14	50446.12	61154.543	83670.14	99688.87	32957.98	33770.816	89999.94	111316.75
EXTENSION COEFFICIENT	87659.75	157137.25	81971.6	100050.16	131802.67	158459.83	53105.535	48602.41	144807.45	183429.42
ABSORBANCE	1.52	1.50	1.56	1.57	1.54	1.52	1.54	1.35	1.55	1.56
HYDROPHATICITY [GRAVY]	-0.04	-0.10	-0.30	-0.22	-0.18	-0.20	0.23	0.11	-0.13	-0.17
ALIPHATIC INDEX	86.74%	83.97%	80.47%	84.11%	78.93%	78.49%	102.01%	99.14%	82.76%	83.21%
AROMATICITY	8.22%	8.21%	8.65%	8.54%	8.85%	8.82%	10.86%	9.66%	9.17%	9.18%
PROTEIN STABILITY	22.97%	22.92%	24.90%	24.26%	22.55%	23.24%	16.67%	19.57%	21.78%	22.72%
C-BETA BRANCHED	2190%	21.18%	20.06%	20.67%	21.70%	21.27%	20.75%	19.95%	19.82%	19.61%
POLARITY	49.279%	49.60%	51.33%	50.29%	49.55%	49.96%	43.96%	45.30%	48.87%	49.28%
SALT BRIDGED	19.20%	19.68%	23.66%	22.58%	20.28%	20.61%	15.53%	17.68%	19.12%	20.01%
HELIX FORMATION	38.10%	39.04%	44.12%	43.46%	38.80%	39.43%	40.54%	41.52%	39.92%	40.72%
BETA SHEET	33.08%	32.02%	30.45%	30.89%	32.68%	32.39%	32.48%	30.77%	30.95%	30.68%
COIL	28.83%	28.94%	25.43%	25.65%	28.53%	28.19%	26.98%	27.71%	29.13%	28.60%
POSITIVE CHARGED	9.48%	10.10%	12.23%	11.84%	10.49%	10.72%	7.33%	8.88%	10.47%	11.00%
NEGATIVE CHARGED	10.52%	10.19%	10.93%	10.74%	9.93%	10.21%	8.46%	9.53%	9.35%	9.83%

Figure 1. Secondary structure Formation of Mosquito Borne Subfamilies

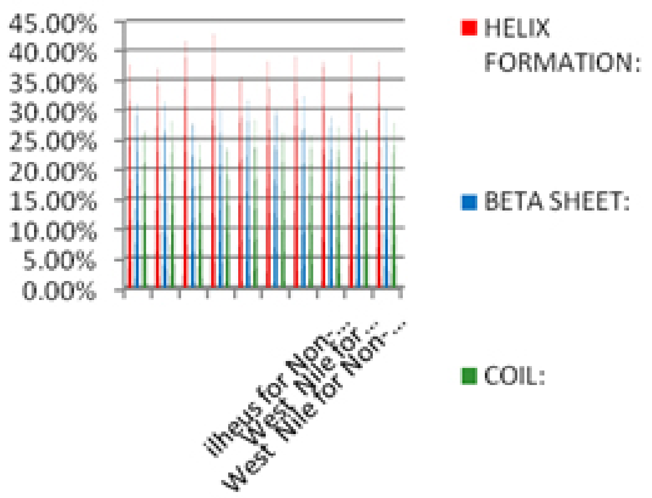


Figure 2. Protein stability of Mosquito of Borne Subfamilies

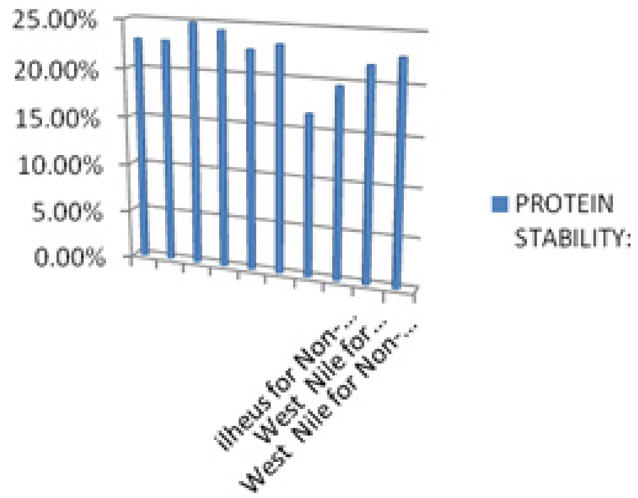


Table 9. Physicochemical Properties of Tick Borne Subfamilies

Physicochemical Properties (Parameters)	Central European		Louping l Il		Powassan		Russion Spring-Rodent	Summar encephalitis		
	R	Non-R	R	Non-R	R	Non-R	R	Non-R	R	Non-R
MOLECULAR WEIGHT	51949.90	64864.06	62048.28	48462.6	76665.56	104627.31	975770.7	1356700.0	46053.8	46053.8
EXTENSION COEFFICIENT	59115.016	73226.59	103324.78	79950.14	134810.8	188640.06	1096284.2	1520818.0	75842.5	75842.5
ABSORBANCE	1.20	1.17	1.75	1.74	1.62	1.75	1.02	0.83	1.71	1.71
HYDROPHATICITY GRAVY]	-0.04	-0.17	-0.09	-0.07	-0.31	-0.26	-0.04	-0.42	-0.21	-0.21
ALIPHATIC INDEX	91.54%	86.54%	87.55%	88.60%	77.86%	82.12%	98.51%	.15%	82.46%	82.46%
AROMATICITy	9.33%	9.00%	8.61%	8.41%	7.80%	7.95%	7.22%	5.72%	8.31%	8.31%
PROTEIN STABILITY	21.45%	23.28%	23.30%	23.36%	27.17%	26.40%	24.13%	.41%	24.73%	24.73%
C-BETA BRANCHED	18.76%	18.02%	18.75%	18.86%	19.80%	19.01%	18.86%	18.19%	18.54%	18.54%
POLARITY	47.21%	49.77%	46.279%	46.08%	51.11%	48.97%	45.98%	53.04%	48.90%	48.90%
SALT BRIDGED	19.08%	20.29%	20.65%	20.63%	22.53%	22.31%	22.26%	28.424%	19.957%	19.95%
HELIX FORMATION:	41.39%	42.01%	43.85%	44.28%	43.87%	44.86%	48.83%	52.47%	42.58%	42.58%
BETA SHEET:	29.73%	29.04%	29.25%	29.18%	20.01%	28.86%	26.83%	24.95%	29.08%	29.08
COIL:	28.85%	28.90%	26.900%	26.55%	26.12%	26.28%	24.35%	22.58%	28.35%	28.35%
POSITIVE CHARGED:	9.80%	10.27%	11.38%	11.50%	12.19%	12.50%	9.85%	3.06%	11.04%	11.04%
NEGATIVE CHARGED:	10.01%	11.00%	11.00%	9.96%	12.55%	11.991%	13.53%	17.30%	11.46%	11.46%

Figure 3 Secondary structure Formation Tendency of Tick Borne Subfamilies

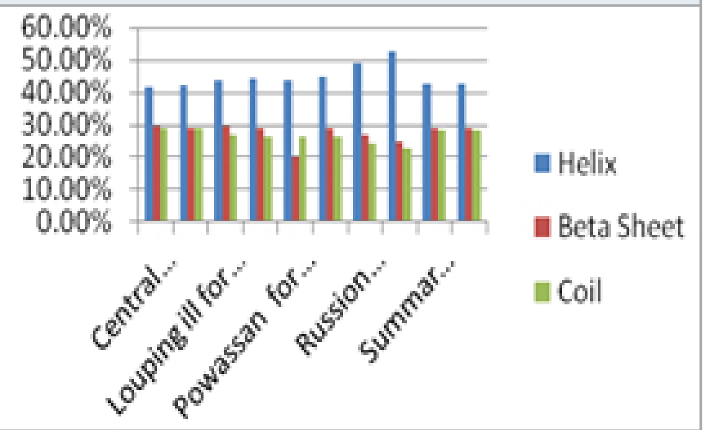
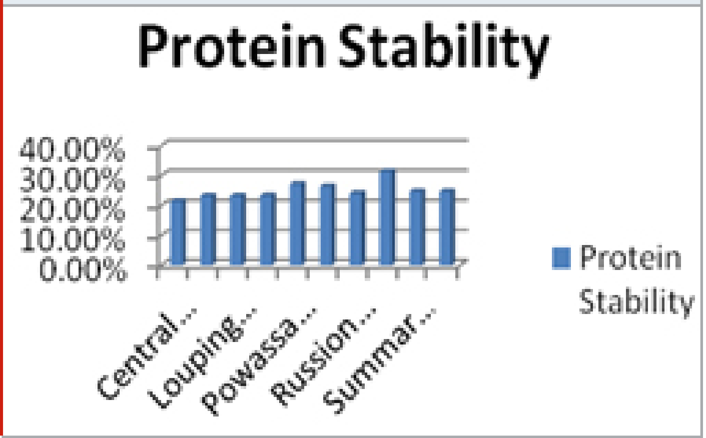


Figure 4 Protein stability of Tick Borne Subfamilies



For 4 frequent Patterns:-

1. {E(Frequent)∩A(Frequent)∩L(Frequent)∩K(Frequent)=>Tendency for Helix Formation}

Tick Borne Subfamilies:-

For 2 Frequent Patterns

1. {A(Frequent)∩L(Frequent)=>Tendency for Helix Formation}
2. {A(Frequent)∩E(Frequent)=>Tendency for Helix Formation}
3. {E(Frequent)∩K(Frequent)=>Tendency for Helix Formation and Protein Solubility}
4. {L(Frequent)∩R(Frequent)=>maintain charge of protein and help in protein stability }
5. {V(Frequent)∩I(Frequent)=>Tendency for Sheet Formation}
6. {G(Frequent)∩S(Frequent)=>Tendency for Coil Formation}

For 3 frequent Patterns:-

1. { E(Frequent)∩A(Frequent)∩L(Frequent)=>Tendency for Helix Formation}

Table 10. Physicochemical Properties of Known Vector Subfamilies

Physicochemical Properties (Parameters)	Rocio R	Non-R	Rio Bravo R	Non-R
MOLECULAR: WEIGHT	47890.18	48100.777	50921.996	7222.977
EXTENSION COEFFICIENT [assuming all residues of tyr,trp,cys]:	57442.98	58009.14	95937.62	88235.15
ABSORBANCE :	1.406	1.415	1.556	1.551
HYDROPHATICITY [GRAVY]:	-0.175	-0.210	0.321	0.275
ALIPHATIC INDEX:	88.693%	87.751%	110.873%	08.928%
AROMATICITY:	9.205%	9.201%	9.558%	9.106%
PROTEIN STABILITY:	21.997%	22.564%	17.356%	18.189%
C-BETA BRANCHED:	18.433%	18.334%	20.401%	20.638%
POLARITY:	48.404%	49.059%	43.085%	43.980%
SALT BRIDGED:	19.403%	19.775%	16.250%	16.853%
HELIX FORMATION:	41.123%	41.160%	41.653%	40.991%
BETA SHEET:	29.418%	29.392%	31.452%	31.468%
COIL:	29.460%	29.448%	26.896%	27.541%
POSITIVE CHARGED:	10.590%	10.8585%	9.099%	9.249%
NEGATIVE CHARGED:	9.626%	9.850%	6.764%	7.216%

Figure 5. Secondary structure Formation Tendency of Known Vector Subfamilies

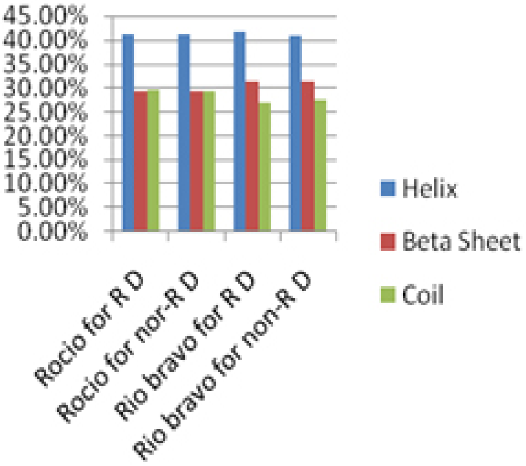
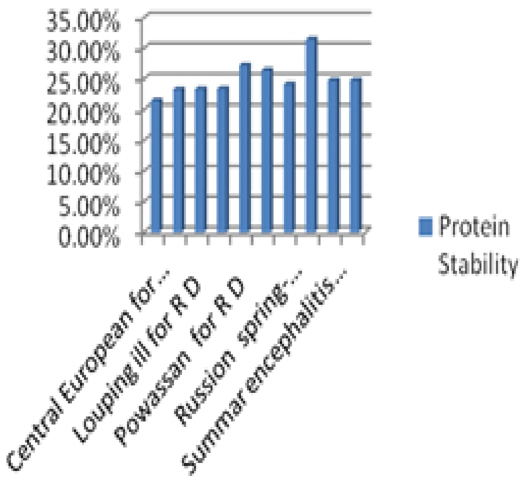


Figure 6. Protein stability of Known Vector Tendency of Known Vector Subfamilies



2. {A(Frequent)∩L(Frequent)∩R(Frequent)=> Tendency for Helix Formation}
3. {E(Frequent)∩K(Frequent)∩R(Frequent)=> Tendency for Helix Formation and Protein Solubility }
4. {I(Frequent)∩T(Frequent)∩V(Frequent)=> Tendency for Sheet Formation}
5. {D(Frequent)∩G(Frequent)∩S(Frequent)=>Tendency for Coil Formation}

For 4 frequent Patterns:-

1. {E(Frequent)∩A(Frequent)∩L(Frequent)∩K(Frequent)=>Tendency for Helix Formation}
2. {E(Frequent)∩A(Frequent)∩L(Frequent)∩R(Frequent)=>Tendency for Helix Formation}

For 5 frequent Patterns:-

1. {E(Frequent)∩A(Frequent)∩L(Frequent)∩K(Frequent)∩R(Frequent)=>Tendency for Helix Formation}

Known Vector Subfamilies:-

For 2 Frequent Patterns:-

1. {A(Frequent)∩L(Frequent)=>Tendency for Helix Formation}
2. {L(Frequent)∩K(Frequent)=>Tendency for Helix Formation}
3. {V(Frequent)∩T(Frequent)=>Tendency for Sheet Formation}
4. {V(Frequent)∩I(Frequent)=>Tendency for Sheet Formation}
5. {G(Frequent)∩S(Frequent)=>Tendency for Coil Formation}
6. {G(Frequent)∩P(Frequent)=>Tendency for Coil Formation}

According to rule 1 for 2 frequent patterns of all subfamilies, it has been observed that amino acids A and L favour helix formation. According to rule 1 for 3 frequent patterns of mosquito and tick borne subfamilies, it has been observed that amino acids E, A and L favour helix formation. According to rule 1 for 4 frequent patterns of mosquito and tick borne subfamilies, it has been observed that amino acids E, A, L and K favour helix formation. According to rule 1 for 5 frequent patterns of tick borne subfamilies, it has been observed that amino acids E, A, L, K and R favour helix formation.Similar interpretation can be inferred by rest of the rules for frequent patterns of all subfamilies in frequent patterns. The above result shows frequent pattern for amino acid in helix formation is maximum than sheet and coil formation.

CONCLUSION

The fuzzy set approach is proposed and employed for prediction of amino acid association patterns in peptide sequences of flavivirus subfamilies. The association rules generated have been used to predict the physiochemical properties and secondary structures as an illustration. The association patterns generated gives the insights of various relationships among amino acids, physiochemical properties and secondary structures. Such models can be developed to generate the information on molecular relationships and mechanisms involved in the disease which could be useful to bio medical scientists for development of methodology for diagnosis and treatment of diseases.

REFERENCES

Agrawal R., Mannila H., Srikant R., Toivonen H., Verkamo A. I. (1995), Fast discovery of association rules, Advances in Knowledge Discovery and Data Mining , Pages 307-328,AAA I Press, mento Park, CA.

Agrawal R., Imielinski T., Swami A.N. (1993), Mining association rules between sets of items in large database In : Proc. SIGMOD, pp.207-216.

Akey D.L., Brown W.C., Dutta S., Konwerski J., Jose J., et al. (2014), Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. Science, 343, 881–885.

Artamonova I.I., Frishman G., Gelfend M.S., Frishman D. (2005), Mining sequence annotation databanks for association patterns, Bioinformatics,21:iii49-iii57.

Atluri G., Gupta R., Fang G., Pandey G., Steinbach M., Kumar V. (2009), Association analysis techniques for Bioinformatics Problems, Department of Computer Science and Engineering, University of Minnesota, Springer-Verlag Berlin Heidelberg, S. Rajasekaran (Ed.): BICoB, LNBI 5462, pp. 1–13.

Avirutnan P., Fuchs A., Hauhart R. E. (2010), Antagonism of the complement component C4 by flavivirus nonstructural protein NS1, J Exp Med 207:793-806.

Blitvich B. J. (2008), Transmission dynamics

and changing epidemiology of West Nile virus, Animal Health Research Reviews, vol. 9, no. 1, pp. 71–86.
Borgelt C., Kruse R. (2002), Induction of Association Rules: Apriori Implementation, In 15th conference on Computational Statistics Physical Verlag pp:395-400.
Cai C.H., Fu W.C., Cheng W.W. (1998), Mining association rules with weighted items, In Proceedings of the International Database Engineering and Applications Symposium, Cardiff, Wales, Uk,pp.68-77.
Gaunt M. W., Sall A. A., Lamballerie X., Falconar A. K. I., Dzhivanian T. I., Goul E. A. (2001), Phylogenetic relationships of flaviviruses correlate with their epidemiology, Disease association and biogeography, Journal of General Virology, vol. 82, part 8, pp. 1867–1876.
Gould E., Solomon T. (2008), Pathogenic flaviviruses, The Lancet, vol. 371, no. 9611, pp. 500–509.
Gour, A., Pardasani, K.R.(2018), Soft Fuzzy Set Approach for Mining Frequent Amino Acid Associations in Peptide Sequences of Dengue Virus,. Proceedings of the National Academy of Sciences., India, Sect. A Physical. Sciences, Vol. 88: 529-538.
Gupta N., Mangal N., Tiwari K., Mitra P. (2006), Mining Quantitative Association Rules in Protein Sequences, Lecture Notes in Computer Science Volume 3755, pp 273-281.
Halstead S. B. (2007), Dengue, The Lancet, vol. 370, no. 9599, pp. 1644–1652.
Han J., Yongjian (1995), Discovery of multiple level association rule from large database, VLDB Conference, pp. 420-431.
Hilda G., Maria A. C., Amelia P. A. (2018), Characterization of Three New Insect-Specific Flaviviruses: Their Relationship to the Mosquito-Borne Flavivirus Pathogens, Am. J. Trop. Med. Hyg., The American Society of Tropical Medicine and Hygiene, 98(2), pp. 410–419.
Hussain A. R., Kumar M. (2019), Role of Endoplasmic Reticulum-Associated Proteins in Flavivirus Replication and Assembly Complexes, Department of Biology, College of Arts and Sciences, Georgia State University, Atlanta, GA 30303, USA.
Hong T.P., Lin K.Y., Wang S.L. (2003), Fuzzy data mining for interesting generalized association rules, Fuzzy Sets and Systems, 138(2):255-269.
Hung-Chang K. et al (2011), Discovering amino acid patterns on binding sites in protein complexes, Bio information , Bio information 6(1):10-14.
Kaya M., Alhajj R. (2003), A Clustering algorithm with genetically optimized membership functions for fuzzy association rule mining, In Proceedings of IEEE International Conference on Fuzzy Systems, pp.881-886.
Khare N., Adlakha N., Pardasani K. R. (2010), An Algorithm for Mining Multidimensional Association Rules using Boolean Matix, Recent Trends in Information,Telecommunication and Computing (ITC), International Conference on IEEE.
Khare N., Adlakha N., Pardasani K. R. (2009), An Algorithm for Mining Multidimensional Fuzzy Association Rules, arXiv preprint arXiv:0909.5166.
Khare N., Adlakha N., Pardasani K. R. (2010), An algorithm for mining conditional hybrid dimensional association rule using Boolean Matix, Computer and automation engineering (ICCAE), The 2nd International Conference on,vol.2, pp.644,648,26-28.
Khare N., Adlakha N., Pardasani K. R. (2009), Karnaugh MAP Model for Mining Association Rules in Large Databases, (IJCNS) International Journal of Computer and Network Security 1.1.
Kuok C., Fu A., Wong M. (1998), Mining Fuzzy association rules in database, SIGMOD Record, 27(1):41-46.
Kumari T., Pardasani K. R. (2013), Mining Fuzzy amino acid Association Patterns in Class C GPCRs, Computational Life sciences, Springer.
Kumari T., Pardasani K. R. (2012), Mining Fuzzy associations among amino acids of class A GPCRs, OnlineJ Bioinform.,13(2):202-213.
Lindqvist R., Upadhyay A., Överby A. K. (2018), Tick-Borne Flaviviruses and the Type I Interferon Response, Department of Clinical Microbiology, Virology, Umeå University.
Manet C., Roth C., Tawfik A., Cantaert T. (2018), Host genetic control of mosquito-borne Flavivirus infections, Springer Science and Business Media, LLC, part of Springer Nature.

Mannila H. et al.(1997), Methods and Problems in Data Mining, In Proceedings of the 6th International Conference on Database Theory, LCNS, Springer-Verleg, 1186:41-55.

Neufeldt C. J., Cortese M., Acosta E. G., Bartenschlager R. (2018), Rewiring cellular networks by members of the Flaviviridae family. Nat. Rev. Genet., 16, pp: 125–142.

Oya A., Kurane I. (2007), Japanese encephalitis for a reference to international travelers, Journal of Travel Medicine, vol. 14, no. 4, pp. 259–268.

Panday A., Pardasani K. R. (2009), PPCI algorithm for mining temporal association rules in large database, Journal of Information and Knowledge Management 8.04:345-352.

Pandey A., Pardasani K. R. (2009), Rough set Model for Discovering Multidimensional Association Rules. International Journal of Computer Science and Network Security9.6:159-164.

Pandey A., Bhargava N., Pardasani K. R. (2007), Counting inference appoach to discover calendar based temporal association rules, SPIT-IEEE Colloquium. Vol.5.

Piatetsky-shapiro G. (1991), Discovery, analysis and presentation of strong rules, Knowledge Discovery in Database, Amino Acidai, Mit Press, pp. 229-248.

R. Agrawal, R. Srikant (1994), Fast algorithms for mining association rules, In Proc of the 20th Int,l Conference on Very Large Database, pp. 497-419.

Seth A., Pardasani K. R. (2015), Soft Set Model for Mining Amino Acid Associations in Peptide Sequences of Mycobacterium Tuberculosis Complex (MTBC), The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 3, No. 7.

Seth A., Pardasani K. R. (2016), Soft fuzzy model for mining amino acid associations in peptide sequences of Mycobacterium tuberculosis complex, Current Science, Vol. 110, No. 4.

Shankar A., Pardasani K. R. (2013), Mining Fuzzy amino acid association patterns in various orders of class Apphaproteo bacteria, Journal of Medical Imaging and Health Informatics.

Shankar A., Pardasani K. R. (2013), Amino acid composition based model for prediction and identification of Alpa and Epsilon-proteobacteria, Online Journal of Bioinformatics, 14(1).

Shi P-Y, (editor) (2012), Molecular Virology and Control of Flaviviruses, Caister Academic Press. ISBN 978-1-904455-92-9.

Wei Q., Chen G. (1999), Mining generalized association rules with fuzzy taxonomic structures, In Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society(NAFIPS), NY, USA, pp. 477-481.

Zadeh L.A. (1965), Fuzzy sets, Information and Control, 8(3):338-353.

Zhang R., Miner J.J., Gorman M.J., Rausch K., et al. (2016), A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. Nature, 535, 164–168.

Zou J., Xie X., Lee L.T., Chandrasekaran R., Reynaud A., et al. (2014), Dimerization of Flavivirus NS4B Protein. J. Virol., 88, 3379–3391.