Statistical Evaluation and Characterization of *Carica papaya* Metabolites

Vidhya Meleveetil

ABSTRACT

Background: Natural resources are often sought for potential molecules which are pharmacologically active. *Carica papaya* is a well-known plant, studied for its constituent secondary metabolites and their benefits. **Methods:** This study focuses on identifying low molecular weight secondary metabolites form *Carica papaya* as potential pharmacophores, based on their structure. We determined their similarity with pharmaceutically active anti-clotting molecules using clustering method. Cheminformatic and Statistical analysis was done with data on molecular characteristics. **Results:** We have identified *Carica papaya* compounds to be pharmacologically active. In this study, we have differentiated *Carica papaya* compounds, through statistical analysis using clustering method, on the basis of significant molecular properties. **Conclusion:** Molecular similarity of secondary metabolites revealed through this study identifies, *Carica papaya* metabolites as a possible substitute drug and a preventive medicine against thrombosis.

Key words: Clustering, Druglikeness, Ligands, Metabolites, Phytochemistry, Thrombogenesis.

INTRODUCTION

There are innumerable ways of interaction and dependency among plants and animals. Selective use of the resources and changes in the environment modifies plant growth and adaptation. These plant responses have enabled the evolution of various metabolites. They form flavours, colours or develop defence mechanisms to cope with competition or survival threats.[1] Some adaptations also provide protection to neighbouring plants.[2] Few such compounds produce a bitter smell, toxic contents or cyanogens.[3]

Production of secondary metabolites involves pathways which function for the plant survival, reproduction, etc. The metabolites produced have wide commercial applications in colouration, cosmetic, fragrances, or defense mechanisms.[4-8] These low molecular weight compounds have unique and complex structures. Methods as study of genetic maps of biosynthetic pathways, targeted metabolite analysis or cDNA amplified fragment length polymorphism (AFLP), quantitatively expresses profiles in metabolite analysis. In targeted metabolite analyses where transcript profiling was carried out with cDNA-amplified fragment length polymorphism, jasmonate elicited tobacco Bright yellow 2 cells. An extensive jasmonate mediated metabolic reprogramming was found focused on genes. This had reflected in metabolite biosynthesis.[9]

This study focusses on the herbaceous plant *Carica papaya* from family Caricaceae. It grows in tropical and many subtropical regions of the world. It has a single unbranched trunk, palmately lobed leaves on long hollow petioles with leaf blades divided into five to nine main segments. Flowers in inflorescence appear in the axils of the leaves. The fruit is melon-like, round or long containing many seeds inside. It's smooth and green skin turns yellow or orange when ripe [Figure 1].[10,11]

The plant is cultivated as a food and cash crop. The crop is developed to increase yield, fruit quality along with better nutritional content. The plant is also found to have high applications in the form of traditional medicines. Various studies based on the constituents indicate presence of phytochemicals, which helps to impart pharmacologic activity as an antioxidant, anti-inflammatory, immunomodulatory or antimicrobial. They are used alone or as combination drugs in treating diseases.[12] Papaya is also identified as a rich source of numerous bioactive secondary metabolites as alkaloids, flavonoids, tannins, saponins and steroids.

Carica papaya is rich in antioxidants. Papaya leaves extracted in water exhibit antioxidant activity which can inhibit *B. subtilis*, *P. aeruginosa* and *S. aureus* gram positive bacterias and *E. coli*, *S. typhi* and *K. pneumonia*, which are gram negative bacteria.[13]

In general, small molecules display unique properties by virtue of their molecular characteristics. Natural resources are used as a feasible resource of compounds, with pharmaceutical applications as drugs or nutraceuticals.

Vidhya Meleveetil

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka, INDIA.

Correspondence

Dr. M Vidhya

Dr. D. S. Kothari Postdoctoral Fellow, Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, Karnataka, INDIA.

Phone no: +91-09972545251

Email id: vidhyarmenon@gmail.com

Figure 1: *Carica papaya* plant (Family: Caricaceae).

In this study, the secondary metabolites identified in the *Carica papaya* plant were chosen for analysis. We determined the structural characteristics specific to each compound. Based on the benefits identified in the plant parts, functional assignment was done. A molecular search was done for reference molecules, focussing on drugs and potential ligands that imparts the specific function. The structural parameters of compounds were compared and statistically characterized through clustering techniques to determine the structural proximity of *Carica papaya* compounds to the reference molecules.

MATERIALS AND METHODS

Ligand Dataset preparation

Identification of Compounds from Carica papaya

Plant secondary metabolite composition of *Carica papaya*, was identified from previous studies. Phytochemical screening results on various extracts of *Carica papaya* taken from the seed, leaf and pulp from the plant were taken. All three extracts, especially the leaf extract are identified to be rich in alkaloids. The extracts also contain glycosides, high amount of polyphenol. Tannins are available in seeds and leaves, but absent in pulp. Flavonoids are absent in seeds and pulp, but leaves are found to be rich in flavonoids.[14]

The leaf extract of *Carica papaya* contained flavonoids as Quercetin, 5,7-Dimethoxy Coumarin, Chlorogenic acid, Caffeic acid, Kaempferol, p-Coumaric acid, and Protocatecheuic acid which imparted anti-viral assembly properties against Dengue2 virus.[15]

Papaya plant composition identified with freeze dried *Carica papaya* leaf juice indicated the presence of twenty four compounds. Compounds including Quinic acid, Malic acid, Protocatechuic acid, Chlorogenic acid, p-Coumaric acid, Caffeic acid, Ferulic acid, Rutin, Nicotiflorin, Myricetin, Fisetin, Morin, Quercetin, Kaempferol, Citropten and Isorhamnetin were determined using mass spectral pattern and the retention times.[16]

From the above studies, many compounds like Quercetin, Fisetin were found to be commonly available in extracts of different parts of the plant. A ligand dataset with commonly identified compounds from *Carica*

papaya and the low molecular weight pharmacologically active drugs, taken as reference molecules, were prepared.

Determination of biological significance of *Carica papaya*

Functional assignment for the *Carica papaya* secondary metabolites were done based on previous studies that are carried out with the plant extracts. Effective biological activity was observed with *Carica papaya* extracts against diseases involving platelets and clot formation.

Papaya plant extracts were found to be effective against viral diseases as Dengue fever.[17] Dengue is caused by a virus which causes self-limiting acute febrile stage of disease which is followed by a critical phase of defervescence. This can also lead to severe illness. Thrombocytopenia along with dysfunction of platelet are found to occur during the disease.[18]

Therapeutic intervention involves understanding disease mechanisms and effective targeting.

Activation of platelet is important in regulating thrombocytopenia in dengue infection.[19] Platelet count is a very important accessory test, reliably used for determination of infections including dengue viral infection.[20]

The viral non-structural protein 1 induced thrombin and plasminogen cross reactive antibodies, are found to participate in the development of haemorrhage in patients with Dengue haemorrhagic fever. This is found to involve the processes of coagulation and fibrinolysis. Thrombogenesis happens through platelet aggregation formed by activating platelets, by attaching to strands of fibrinogen. During coagulation, a cascade of events involving enzymes leads to the formation of fibrin strands. Clotting can be therapeutically targeted using antiplatelet drugs or anticoagulants. Antiplatelet drugs inhibit platelet aggregation. Anticoagulants inhibit the formation of fibrin strands. Fibrinolytic drugs are involved in dissolving existing clots.[21,22]

These studies indicate, effective therapeutic intervention can be done with papaya plant extract in thrombocytopenia or dysfunction of platelets.

Identification of reference drug molecules

Reference molecules were chosen based on their drug like properties. Antiplatelet drugs and anticoagulants were selected as reference molecules. In spite of the benefits provided, these molecules also specifically pose other risks such as postoperative bleeding or other unfavourable drug interactions.[22] Molecular information of the compounds was collected from the PubChem database.[23]

Cheminformatic analysis using clustering and visualization

Biological activity of molecules was determined based on Lipinski's Rule of 5 (RO5).[24] Statistical analysis was carried out to determine the compound structural similarities among various molecules. Through, hierarchical cluster analysis, classification based on data similarity was done with the tool.[25] Numerical data with molecular parameter values were used for clustering data with, Wolfram Mathematica tool.[26] The system is based on usage of classic and modern machine learning-based clustering analysis.

We carried out data analysis to produce a clustering tree, with all the compounds from the small molecule dataset. Molecular data, including various parameter values, were analysed after a symbolic tree based on hierarchical clustering of data was created. A binary weighted tree was created with weight of each vertex indicating the distance between sub trees sharing the vertex as root. The information was plotted for graphical

Table 1: Molecular properties of compounds from *Carica papaya*.

Compound Name	Molecular weight(Da)	Hydrogen bond (No of:)		No of rotatable bond	XLogP3	TPSA (Å²)	Heavy Atom COUNT
		Donor	Acceptor				
Caffeic acid	180.16	3	4	2	1.2	77.8	13
Coumaric acid	164.16	2	3	2	1.5	57.5	12
5-7-Dimethoxy Coumarin (Citropten)	206.19	0	4	2	1.9	44.8	15
Quercetin	302.235	5	7	1	1.5	127	22
Fisetin	286.24	4	6	1	2	107	21
Kaempferol	286.24	4	6	1	1.9	107	21
3,4-Dihydroxybenzoic acid (Protocatechuic acid)	154.12	3	4	1	1.1	77.8	11
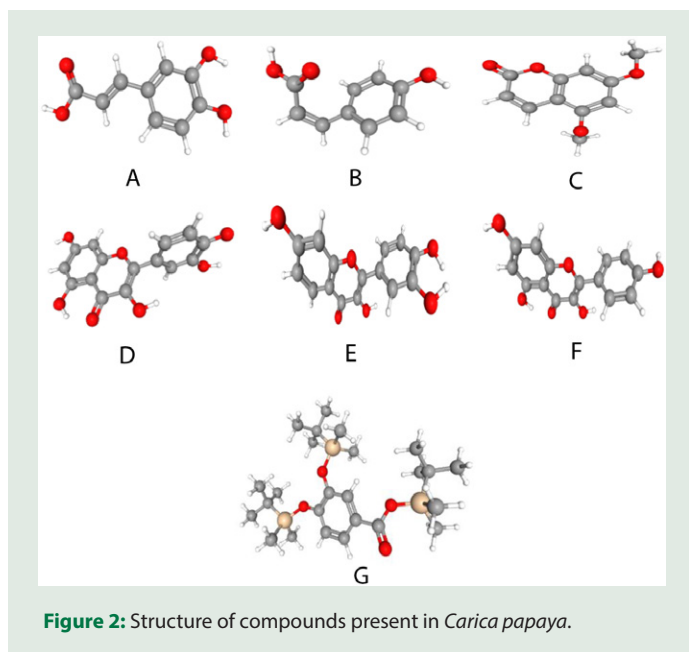
Figure 2: Structure of compounds present in *Carica papaya*.

visualisation and analysis, using charts and information visualization based on histograms.[26-28]

RESULTS AND DISCUSSION

Carica papaya identified as a rich source of useful compounds. Few compounds commonly identified in the *Carica papaya* plant through various studies were taken. [Table 1] [Figure 2]

This included Caffeic acid, Coumaric acid, 5-7-DimethoxyCoumarin (Citropten), Quercetin, Fisetin, Kaempferol, 3,4-Dihydroxybenzoic acid (Protocatechuic acid). [Figure 2: a-g] The molecular weight of compound ranges between: 154.12 to 302.235 Da. Citropten has no hydrogen donor atoms, whereas Coumaric acid has 2 donor atoms and 3 acceptors for hydrogen bonds. Caffeic acid and Protocatechuic acid have 3 donors and 4 acceptors. Fisetin and Kaempferol has 4 hydrogen donors and 6 acceptors. Quercetin had the maximum number of donors as 5 and acceptors as 7 atoms, for the hydrogen bond. Caffeic acid, Coumaric acid and citropten are more flexible with 3 hydrogen bonds. Fisetin, Kaempferol, Quercetin and Protocatechuic acid have one rotatable bond. Maximum logP3 value is found in Fisetin molecule. Quercetin exposes maximum polar surface area and has the maximum number of heavy

atoms. Further analysis focussed to determine similarities and differences in comparison to other molecules. RO5 based analysis shows, the molecules are druggable based on oral activity.

a. Caffeic acid b. Cis-p-Coumaric acid c. Citropten d. Quercetin e. Fisetin f. Kaempferol g. Protocatechuic acid.

Functional assignment of papaya plant compounds was done based on studies of crude plant extract that lead to effective action against the functioning of platelets or thrombocytopenia.

Database and literature search provided numerous drug molecules. We prepared the dataset with compounds which are found to be effective against coagulation or thrombosis and secondary metabolites from *Carica papaya* plant. [Table 1-3] The molecular properties of the molecules were determined from the Pubmed database. Various molecular parameters that affect interactions were identified as hydrogen donors and acceptors, polar surface area exposed for other interactions and rotatable bonds which gives the flexibility to the molecular interactions. Among all the compounds, Eptifibatide had the highest molecular weight, maximum number of donor and rotatable bonds as 10 and the maximum number of acceptors as 12. Clopidogrel had 0 donors, 4 acceptors and rotatable bonds. Warfarin had 1 additional hbond donor than Clopidogrel. After Eptifibatide the maximum polar surface area was available for Ticagrelor.

The compounds from papaya plant are found to possess more similarity based on molecular properties to few drugs with respect to the analysed parameters.

Mean and Standard Deviation (SD) of values for parameters of molecular weight, partition coefficient and total polar surface area (TPSA), showed the SD values are lower than the mean values. This indicates the parameter values of the compounds from papaya plant are closer, to the mean values.

Molecular data based on various parameter values represented in the form of weighted tree were studied based on the distribution of parameter values in different clusters. Clustering of molecular weight of compound entries in the dataset indicated, Caffeic acid and Aspirin belonged to the same cluster. Coumarinic acid and Protocatechuic acid were found to be the closest cluster to Caffeic acid and Aspirin. 5-7-DimethoxyCoumarin was also closely related to these two clusters. Quercetin was clustered with oral anticoagulant Warfarin. The cluster was closely related to the antithrombotic drug Clopidogrel. Fisetin and Kaempferol which formed a cluster were found to be closely related to Clopidogrel. Beyond Clopidogrel, these compounds were found to be closer to antithrombotic drug Prasugrel and Cilostazol [Figure 3a].

Based on partition coefficients, Caffeic acid and Protocatechuic acid were clustered together and were found closer to Aspirin. 5-7-Dime-

Table 2: Antithrombotic and anticoagulant drug molecules.

Drugname	Type	MW g/mol	Hdonor	Hacceptor	rot bond	XLogP3	TPSA(Å²)	Heavy Atom COUNT
Aspirin		180.16	1	4	3	1.2	63.6	13
Clopidogrel		321.8	0	4	4	3.8	57.8	21
Prasugrel		373.4	0	6	6	3.6	74.8	26
Ticagrelor		522.6	4	12	10	2	164	36
Dipyridamole	Antiplatelet	504.6	4	12	12	0.7	145	36
Cilostazol		369.5	1	5	7	3.1	81.9	27
Tirofiban		440.6	4	7	14	1.4	113	30
Eptifibatide		832	10	12	10	-2.4	377	57
Warfarin		308.3	1	4	4	2.7	63.6	23
Dabigatran		471.5	4	7	9	1.7	150	35
Rivaroxaban	Oral Anticoagul-ants	435.9	1	6	5	2.5	116	29
Apixaban		459.5	1	5	5	2.2	111	34
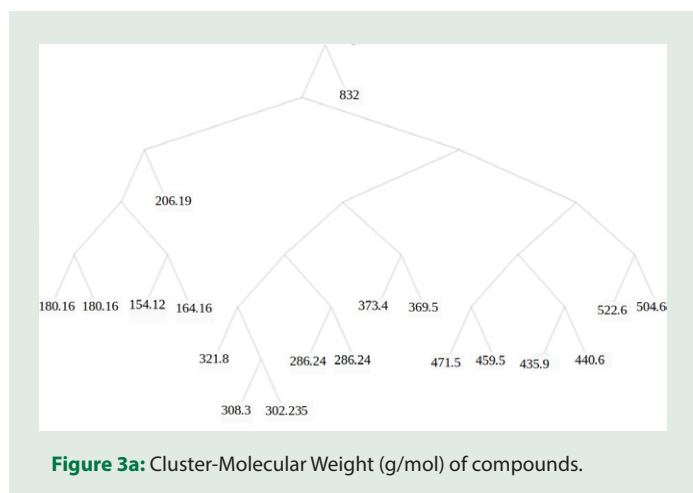
Figure 3a: Cluster-Molecular Weight (g/mol) of compounds.
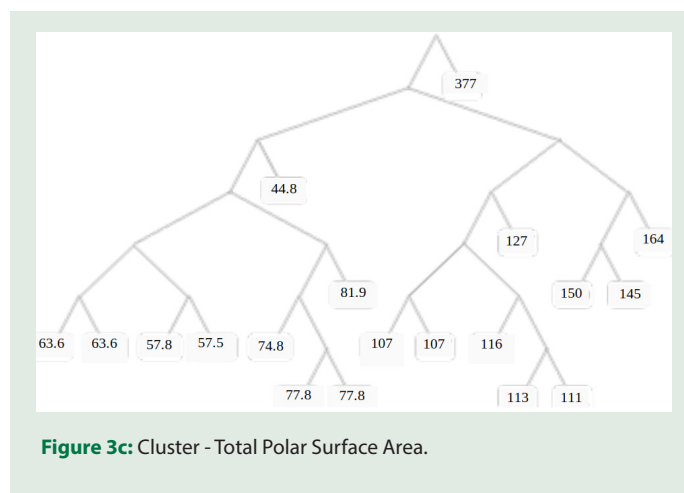
Figure 3c: Cluster - Total Polar Surface Area.
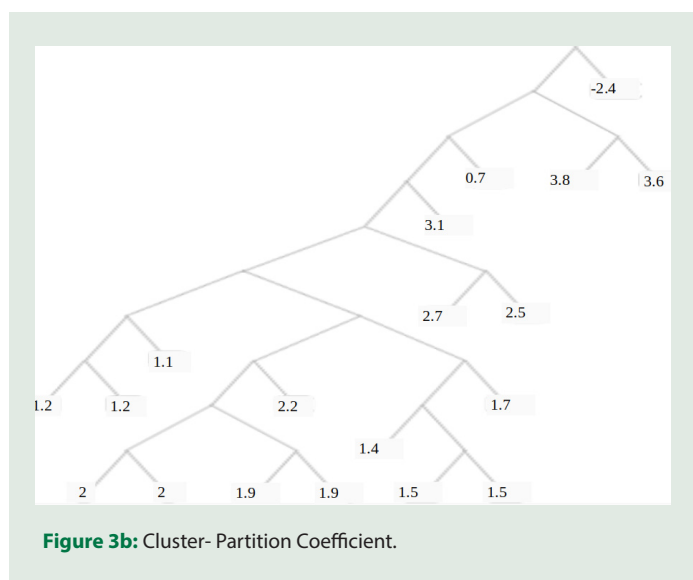
Figure 3b: Cluster- Partition Coefficient.

Table 3: Mean and SD among anti-thrombotic and anti-coagulant molecules and *Carica papaya* compounds.

Molecular weight		XLogP3		TPSA(Å²)	
Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
357.853	165.384	1.76842	1.30981	111.4	72.834

thoxy Coumarin, Fisetin, Kaempferol and Ticagrelor were found to closely clustered [Figure 3b].

Clustering based on TPSA indicated Caffeic acid, Protocatechuic acid belonged to the same cluster, which and more related to antiplatelet compounds, Dipyridamole and Prasugrel. 5-7-DimethoxyCoumarin closely relate to this cluster. Coumaric acid was clustered with Antithrombotic compound Clopidogrel. Fisetin, Kaempferol, Quercetin were found to closely relate to Oral anticoagulants Rivaroxaban, Apixaban and antithrombotic molecule Tirofiban [Figure 3c].

Compounds as 5-7-Dimethoxy Coumarin, Fisetin and Kaempferol shared the same number of hydrogen donor atoms mostly with anti-

thrombotic molecules as Clopidogrel, Prasugrel, Ticagrelor. The number of acceptors as 4, 7 or 6 were available for compounds in anticoagulant and antithrombotic molecules. The number of rotatable bonds was the least for compounds from the papaya plant which were either 2 or 1 indicating lowest rotational freedom during the interaction of the molecules in comparison to the antithrombotic or anticoagulant drugs. The majority

Figure 4(a): Molecular weight(g/mol) distribution of compounds.

Figure 4b: Partition coefficient distribution of compounds.

Figure 4c: Total Polar Surface Area distribution of compounds.

of the compounds are found to be of low molecular weight, with a partition coefficient ranging from 0 and 4 having a topological surface area ranging from 51 to 150.

The molecular weight distribution of compounds was maximum between 200 to 400 g/mol and one above 800g/mol. Partition coefficient of few compounds was found to be below 0 and the maximum number of compounds possessed a coefficient between 0-2. The polar surface area in the range 0-50 and 350-400 were found to be present in few compounds. A majority of compounds possessed TPSA value in the range 50 – 100. The maximum number of rotatable bonds was found to be between 20 and 30 numbers [Figure 4a-c].

CONCLUSION

The *Carica papaya* plant is found to be a rich source of secondary metabolites. Compounds from *Carica papaya* are found to be pharmacologically active with respect to Lipinski's RO5 specifications. Statistical analysis through clustering, segregated the compounds based on molecular properties where *Carica papaya* compounds were grouped in most of the cases with anti-thrombotic drugs than with anti-coagulant drugs. Standard deviations of parameter values for compounds from the papaya plant, indicate the compound relatedness with referenced antithrombotic or anticoagulatory molecules.

This study, based on Pharmacognosy of *Carica papaya* compounds, shows similarity through clustering of molecular properties with the referenced anti-thrombotic or anticoagulant molecules. Clustering analysis recognised them to be part of a specific group of pharmaceutically active molecules. Grouping with Aspirin or Clopidogrel was significant in classification with respect to other drug properties. The secondary metabolites from *Carica papaya*, indicate a high potential structurally as pharmacologically active molecule which can be studied aginst clot related aetiology.

ACKNOWLEDGEMENT

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ABBREVIATIONS

cDNA: Circular Deoxyribo Nucleic Acid; **AFLP:** Amplified Fragment Length Polymorphism; ***B. subtilis:*** *Bacillus subtilis*; ***P. aeruginosa:*** *Pseudomonas aeruginosa*; ***S. aureus:*** *Staphylococcus aureus*; **E. coli:** *Escherichia coli*; ***S. typhi:*** *Salmonella typhi*; ***K. pneumonia:*** *Klebsiella pneumoniae*; **RO5:** Lipinski's Rule of 5; **Da:** Dalton; **TPSA:** Total Polar Surface Area; **SD:** Standard Deviation; **XlogP3:** Partition Coefficient; **Hacceptor:** Hydrogen bond acceptor; **Hdonor:** Hydrogen bond donor; **MW:** Molecular Weight.

REFERENCES

1. Simms EL. Plant–Animal Interactions. 2013;39-55.
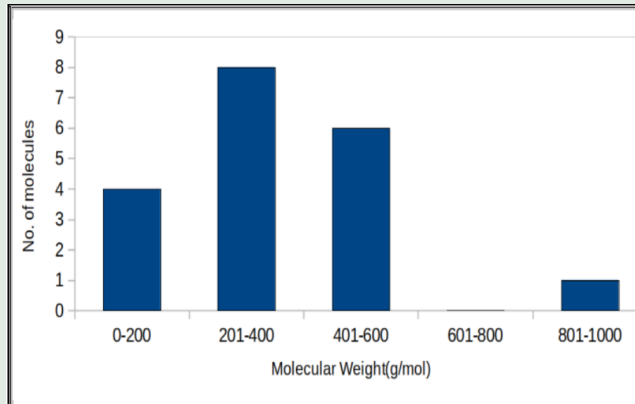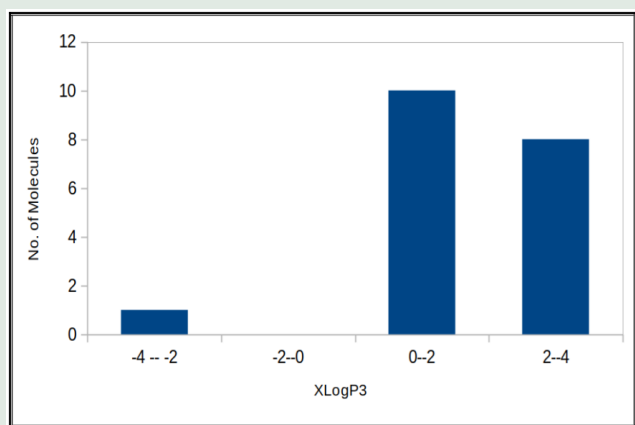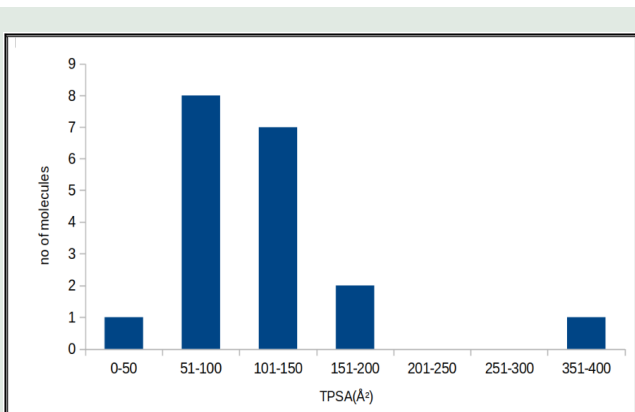2. Harbone JB. Introduction to Ecological Biochemistry. Academic Press, London.

1993;71-103.

3. Reed DR, Knaapila A. Genetics of taste and smell: poisons and pleasures. Progress in Molecular Biology and Translational Science. 2010;94:213-40.

4. Hussein RA, El-Anssary AA. Plants secondary metabolites: The key drivers of the pharmacological actions of medicinal plants. Herbal Medicine. 2019;1:13.

5. Isah T. Stress and defense responses in plant secondary metabolites production. Biological Research. 2019;52.

6. Pagare S, Bhatia M, Tripathi N, Pagare S, Bansal YK. Secondary metabolites of plants and their role: Overview. Current Trends in Biotechnology and Pharmacy. 2015;9(3):293-304.

7. Wink M. Plant secondary metabolites modulate insect behavior-steps toward addiction?. Frontiers in Physiology. 2018;9:364.

8. Böttger A, Vothknecht U, Bolle C, Wolf A. Plant secondary metabolites and their general function in plants. InLessons on Caffeine, Cannabis and Co. Springer, Cham. 2018;3-17.

9. Goossens A, Häkkinen ST, Laakso I, Seppänen-Laakso T, Biondi S, Sutter DV, *et al*. A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proceedings of the National Academy of Sciences. 2003;100(14):8595-600.

10. Abdulazeez MA, Sani I. Use of Fermented Papaya (*Carica papaya*) Seeds as a Food Condiment, and Effects on Pre-and Post-implantation Embryo Development. InNuts and Seeds in Health and Disease Prevention, Academic Press. 2011;855-63.

11. *Carica papaya* plant image: "FreeImages.com/Vasant Dave". 2021. [cited 2021 January 5].

12. Available from: https://www.freeimages.com/photo/papaya-tree-1-1323924

13. Nafiu AB, Alli-Oluwafuyi AM, Haleemat A, Olalekan IS, Rahman MT. Papaya (*Carica papaya* L., Pawpaw). In Nonvitamin and Nonmineral Nutritional Supplements. Academic Press. 2019;335-59.

14. Sarjono PR, Putri LD, Budiarti CE, Mulyani NS, Kusrini D, Prasetya NB. Antioxidant and antibacterial activities of secondary metabolite endophytic bacteria from papaya leaf (*Carica papaya* L.). InIOP Conference Series: Materials Science and Engineering, IOP Publishing. 2019;509(1):012112.

15. Ikpeme EV, Ekaluo UB, Kooffreh ME, Udensi O. Phytochemistry and heamatological potential of ethanol seed leaf and pulp extracts of *Carica papaya* (Linn.).

Pakistan Journal of Biological Sciences. 2011;14(6):408-11.

16. Senthilvel P, Lavanya P, Kumar KMRS, Anitha P, Bag S, Sarveswari S, *et al*. Flavonoid from *Carica papaya* inhibits NS2B-NS3 protease and prevents Dengue 2 viral assembly. Bioinformation. 2013;9(18):889.

17. Razak AMR, Misnan NM, Jelas NH, Norahmad NA, Muhammad A, Ho TC, *et al*. The effect of freeze-dried *Carica papaya* leaf juice treatment on NS1 and viremia levels in dengue fever mice model. BMC Complementary and Alternative Medicine. 2018;18(1):1-3.

18. Sarala N, Paknikar SS. Papaya extract to treat dengue: A novel therapeutic option?. Annals of Medical and Health Sciences Research. 2014;4(3):320-4.

19. Azeredo EL, Monteiro RQ, Pinto DOLM. Thrombocytopenia in dengue: Interrelationship between virus and the imbalance between coagulation and fibrinolysis and inflammatory mediators. Mediators of Inflammation. 2015.

20. Ojha A, Nandi D, Batra H, Singhal R, Annarapu GK, Bhattacharyya S, *et al*. Platelet activation determines the severity of thrombocytopenia in dengue infection. Scientific Reports. 2017;7(1):1-0.

21. Mohan S, Dutta R, Singh A. Association of serological parameters and platelet counts in dengue fever: A hospital-based study. International Journal of Health and Allied Sciences. 2019;8(3):164.

22. Chuang YC, Lin YS, Liu HS, Wang JR, Yeh TM. Antibodies against thrombin in dengue patients contain both anti-thrombotic and pro-fibrinolytic activities. Thrombosis and Haemostasis. 2013;110(08):358-65.

23. Becker DE. Antithrombotic drugs: Pharmacology and implications for dental practice. Anesthesia Progress. 2013;60(2):72-80.

24. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, *et al*. PubChem substance and compound databases. Nucleic Acids Research. 2016;44(D1):D1202-13.

25. Walters WP, Murcko AA, Murcko MA. Recognizing molecules with drug-like properties. Current Opinion in Chemical Biology. 1999;3(4):384-7.

26. Wolfram Research, Clustering Tree, Wolfram Language function. 2016. https://reference.wolfram.com/language/ref/ClusteringTree.html (updated 2017).

27. Wagon S. Mathematica in action. Springer Science and Business Media; 1999.

28. Mangano S. Mathematica Cookbook: Building Blocks for Science, Engineering, Finance, Music, and More. O'Reilly Media, Inc. 2010.

29. Szalay A, Gray J. Science in an exponential world. Nature. 2006;440(7083):413-4.

GRAPHICAL ABSTRACT

METABOLITES--LIGAND--DRUG(THROMBOSIS)
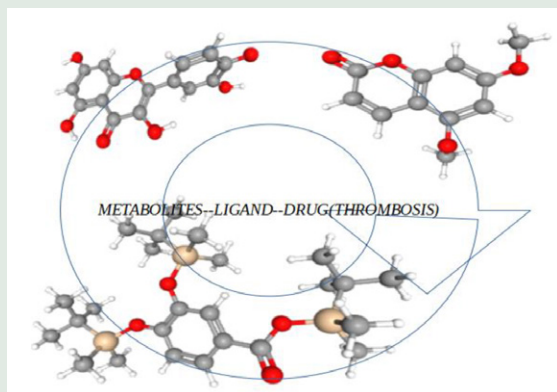
SUMMARY

The *Carica papaya* plant is identified for many of its medicinal uses. Various secondary metabolites are identified in the plant source. This work analyses the secondary metabolites from *Carica papaya* plant based on its structural and molecular properties to determine their pharmacognostic characteristics. The metabolites structural characteristics are compared and statistically clustered through hierarchical clustering methods with anti-thrombotic molecules. The results identify molecular similarities with specific anti-thrombotic drugs, indicating the structural relatedness which could contribute to their functional benefits.