

# Network Pharmacology Study of Curcuma Longa: Potential Target Proteins and their Functional Enrichment Analysis.

SANGEETA KUMARI (✉ [kumari13@live.com](mailto:kumari13@live.com))

---

## Research note

**Keywords:** Markov clustering, functional module, Centrality measure, Synergistic mechanism

**Posted Date:** June 5th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-31985/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on October 7th, 2020. See the published version at <https://doi.org/10.1186/s13104-020-05301-0>.

# Abstract

**Objective:** This study's primary goal is unraveling the mechanism of action of bioactives of *Curcuma longa* L. at the molecular level using protein-protein interaction network.

**Results:** We used target proteins to create protein-protein interaction network (PPI) and identified significant node and edge attributes of PPI. To find the function module, we identified the cluster of proteins which were further queried to GO enrichment analysis. Closeness centrality and jaccard score identified as most important node and edge attribute of the protein-protein interaction network respectively. The mapped pathways of various function module of the network were overlapped and showed synergistic mechanism of action. Three most important identified pathways were Gonadotropin-releasing hormone receptor pathway, Endothelin signaling pathway, and Inflammation mediated by chemokine and cytokine signaling pathway.

## Introduction

The *Curcuma longa* L. has been studied for antiinflammatory, anticancer effects [1]. The exact mechanism remains largely unexplored. Bioactives have shown the multi-components, multi-targets effect which can be studied using protein-protein interaction network (PPI). A target protein usually carries out a typical function by regulating other molecules, thus, study of PPI helps to understand relationship between target proteins and other interacting proteins in a systematic way [2]. Earlier study has shown that target proteins indeed have some special topological features that are significantly different than the normal proteins [3]. Thus, comparison study of a PPI network and a false interaction network was needed to identify them which further could be used to rank the nodes and edges in a PPI network.

## Methods

The four bioactive compounds namely curcumin, Desmethoxycurcumin, Bisdemethoxycurcumin and Turmerone of *Curcuma longa* were studied. The smile structure of all these four compounds were obtained from pubchem database. We used similarity ensemble approach (<http://sea.bkslab.org/>) to identify the potential target proteins of all these four bioactive compounds. We queried the target proteins to StringDB to identify all the possible interactions involving the target proteins. We used NetworkX library in python to build and study the protein-protein interaction network. We used Markov-clustering library in python to identify functional module in the network. We used GO enrichment analysis to identify important pathways associated with various module of the network [4].

## Calculation

We created network using StringDB interaction as edges and interacting proteins as nodes. We calculated jaccard\_score, preferential\_attachment, common\_neighbors, and resource\_allocation\_index values for all

the edges of network using NetworkX library in python. We calculated node topological property of network such as degree, eigenvector centrality, betweenness centrality, closeness centrality, local clustering coefficient, eccentricity values for all the nodes in the network. We used Markov clustering algorithm and optimized the hyperparameter inflation value by optimizing the modularity score of a cluster. We found optimized value of inflation parameter at 2.4 that gives modularity value of cluster 0.77. The codes available at [https://github.com/sangeeta97/Network\\_Pharmacology](https://github.com/sangeeta97/Network_Pharmacology) .

## Results And Discussion

### Similarity Ensemble Approach

The chemical-centric method can exploit the pharmacological relationships among protein targets in addition to biological [5]. Curcumin found to have 193 human target proteins. Desmethoxycurcumin was mapped to 166 human target proteins, Bisdemethoxycurcumin had 71 human target proteins and Turmerone was associated with 2 target proteins. After removing overlapped target proteins of all the studied bioactives, we had 219 unique target proteins for further study.

### Network formation and Property

We downloaded Human protein interaction data (scored links between proteins) from String DB and identified all the interaction in which any of the 219 target proteins were involved and this has led to 208125 interactions for which the interaction score was varied from 150 to 999. Further to reduce the complexity of the network and to increase the confidence of the interaction, we included only edges having interaction score above 300. This has led to total 58482 interactions (edge) involving 11979 proteins (nodes).

The obtained PPI network had average shortest path length of 3.5, diameter of 6, transitivity of 0.03, avg\_clustering of 0.3, and density value of 0.0008, Average degree value of 9. Network was connected and scale-free with few nodes having very high degree and many nodes with a small degree value. The PPI network was small world network that had diameter of 6 and average shortest path length of 3.5.

### Real Biological interactions Network vs. False interaction Network

To understand the property of real biological PPI network interactions, we calculated the different edge attribute that computes the probability of forming the interaction between two nodes based only on their topological features. We calculated preferential attachment score (nodes with high degree having higher chances to connect), common neighbors score (nodes with higher number of common neighbor having higher chance to connect), jaccard score (nodes having higher number of common neighbor and having lesser degree having higher chance to connect), and resource allocation score (nodes having common

neighbor which have less degree value having higher chance to connect) by passing true edge list as parameter to the algorithms implemented in NetworkX library in python. Also, we calculated the edge properties for all the non-existent edges of the network and built the network with non-existent edges. We called this network as false interaction PPI network.

We calculated the correlation coefficient between calculated edge attributes for PPI interaction network and interaction score obtained from StringDB. We found a poor correlation between interaction score and each calculated edge attribute value. Thus, none of the calculated edge attributes resembles the biological interactions between two protein nodes. Further, we studied the difference in the edge attributes between PPI and false interaction PPI network. We found that jaccard score value was differentially high for false interaction PPI network (Fig. 1). We used the machine learning algorithms such as logistic regression and random forest to select best classifier (edge attribute) between the PPI and false interaction PPI network which identified the jaccard score as a significant discriminator between two network types.

## Difference in Centrality Measures

Further, we studied the node attributes of these two networks, and calculated different types centrality measures. We calculated the degree (the number of nodes it connects), closeness centrality (measure relative closeness of a node with respect to all other nodes), Eigenvector Centrality (eigenvector centrality cares if you are a hub, but it also cares how many hubs you are connected to), betweenness Centrality (betweenness centrality looks at all the shortest paths that pass through a particular node),

Local Clustering Coefficient (fraction of pairs of the node's neighbors that are adjacent to each other), Eccentricity (largest possible shortest path distance between node and all other nodes).

We calculated the correlation coefficient of all the centrality measures for the PPI and false interaction PPI network. For PPI network, we found the very strong correlation between degree and betweenness centrality (0.95) which shows that nodes with high degree control the information flow in the network by being present in shortest paths in PPI networks and may contribute to multiple pathways. Thus, it shows the nodes with high degree in PPI network are not good candidate of potential target proteins as they can impact multiple pathways simultaneously in a biological system.

For false interaction PPI network, we found the very strong correlation between degree and eigenvector centrality (0.93) but shows poor correlation between degree and betweenness centrality (0.56) which shows that unlike PPI network, high degree nodes do not control the information flow in the network. Further, we used the machine learning algorithm logistic regression and random forest to select most significant node attributes to differentiate between the PPI network and false interaction PPI network which identified the closeness centrality as a best classifier. For PPI network, nodes have relatively higher values for closeness centrality.

Using our study finding, we removed the insignificant edges and nodes from PPI network and made it sparse. We removed edges having jaccard score value above 75 percentile of PPI network and nodes that had closeness centrality value less than the 25 percentiles. The resulting network had 1900 nodes and 4637 edges.

## Module identification

We used Markov cluster (MCL) algorithm for module identification. MCL algorithm is particularly noise-tolerant as well as effective in identifying high-quality functional modules [6]. MCL is unsupervised cluster algorithm for graphs (also known as *networks*) based on manipulation of transition probabilities to identify functional module. Functional modules are generally highly overlapped but MCL is hard clustering algorithm and proteins are non-overlapping. The fundamental concept of identifying functional modules is that a pair of proteins interacting with each other has higher probability of sharing the same function (pathway) than two proteins not interacting with each other. The algorithm identified 6 clusters within network (Fig. 2).

## GO pathway enrichment Analysis

Target identification and synergistic interaction among multiple target is important unravel the pharmacological mechanism of action of bioactives. Target proteins belonging to each cluster were searched into Gene Ontology database (<http://pantherdb.org/webservices/go/overrep.jsp>). We uploaded the protein list of each cluster, we selected the option of statistical overrepresentation test which tested the set of queried proteins against human genes and identified the statistically significant pathways. Out of 6 cluster, one cluster did not identify any significant pathway. The cluster number 1 had 16 proteins. Cluster 1 had RARA target protein along with its interacting proteins and mapped to Transcription regulation by bZIP transcription factor (P00055) pathway after enrichment analysis. The cluster number 2 has 608 proteins. Cluster 2 had two target proteins; TF65 and JUN where JUN had higher degree value. Cluster 2 was enriched in 26 pathways such as Toll receptor signaling pathway (P00054), Ras Pathway (P04393), FAS signaling pathway (P00020) and Inflammation mediated by chemokine and cytokine signaling pathway (P00031). The cluster number 3 had 330 proteins. Cluster 3 had two target proteins i.e. EP300 and HDAC1. Cluster 3 proteins were enriched in DNA replication (P00017) pathway. The cluster 4 had 647 proteins with 3 target proteins namely EGFR, MMP9, and PTN11. The proteins in cluster 4 were significant in 35 pathways which included EGF receptor signaling pathway (P00018), Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032). The cluster 5 proteins were not significant in any pathway. The cluster 6 had 9 proteins and it had one target protein GSK3B. The proteins in cluster 6 were significant in 37 pathways such as Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027), PI3 kinase pathway (P00048), Wnt signaling pathway (P00057), and Gonadotropin-releasing hormone receptor pathway (P06664). The target proteins present in different clusters act synergistically to play a pivotal role in bioactive-mediated

response. We identified 3 common pathway namely Gonadotropin-releasing hormone receptor pathway, Endothelin signaling pathway, and Inflammation mediated by chemokine and cytokine signaling pathway present in cluster 2, 4, and 6 as important pathways.

## **Limitations**

- Experimental study is not included in the which was ideal to assess pathways enrichment.
- Lack of complete information about target proteins and their interaction which might lead to false positive or false negative interactions in the network.

## **Declarations**

## **Ethics approval and consent to participate**

Not applicable.

## **Consent for publication**

Not applicable.

## **Availability of data and material**

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request

## **Competing interests**

Not applicable.

## **Funding**

There was no funding source for this study

## **Author's contributions**

Dr. Sangeeta Kumari contributed to conception and design, acquisition of data, analysis and interpretation of data, drafted the manuscript and accountable for all aspects of the work.

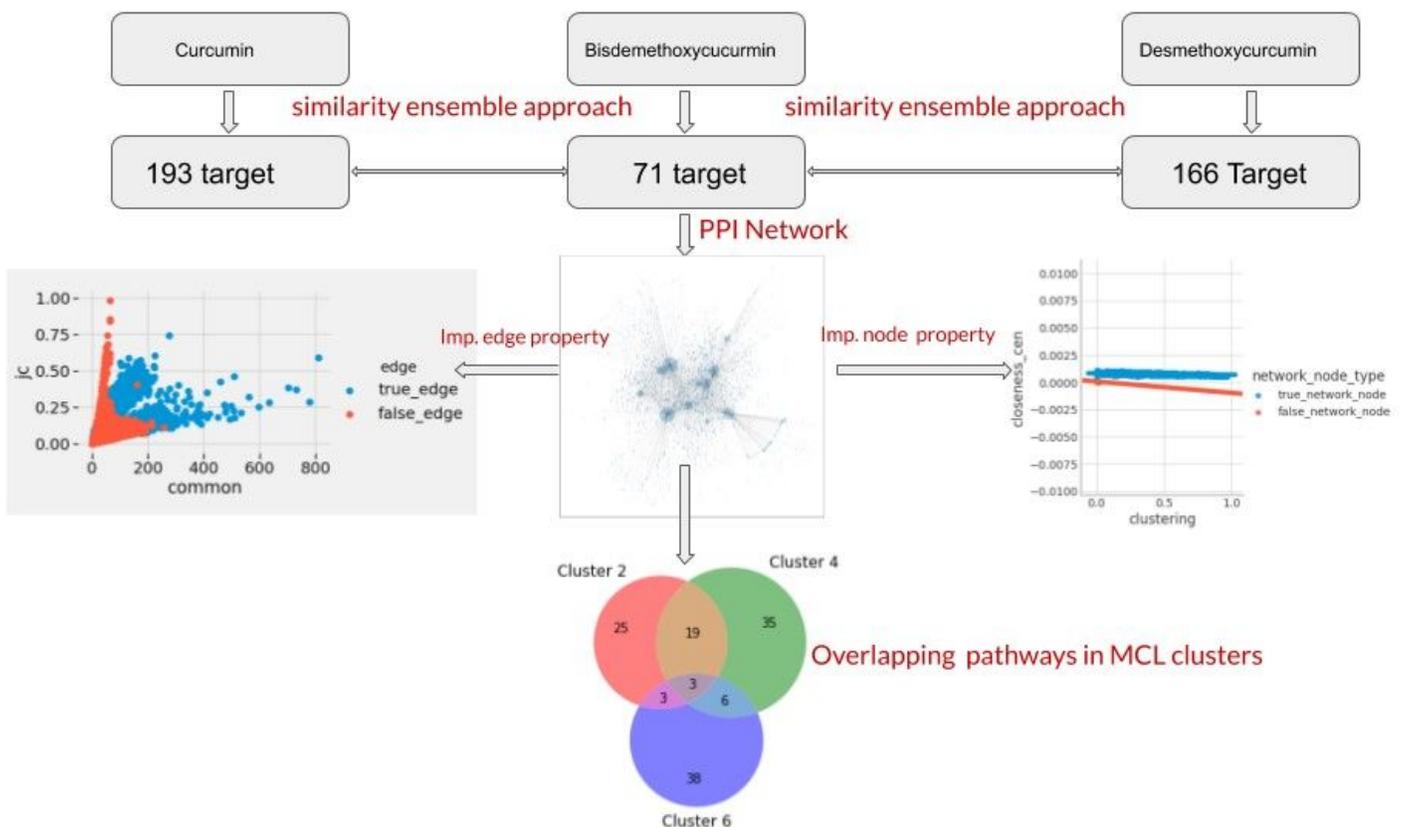
## **Acknowledgements**

Dr. Sangeeta Kumari is grateful to Prof. H S Subramanya and Dr. R. Srivatsan, from IBAB, for helpful discussions.

## References

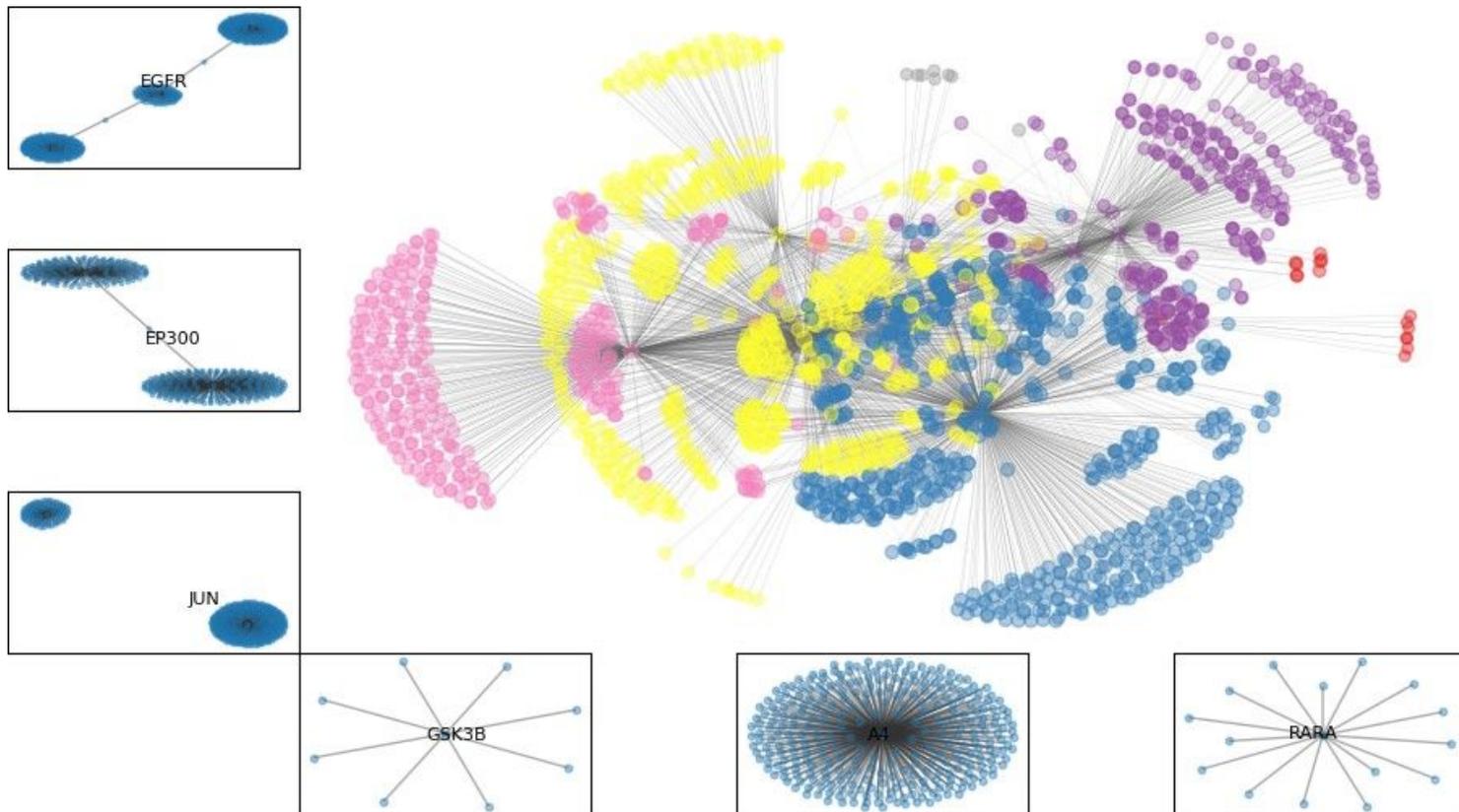
1. Gan Y, Zheng S, Baak JP, et al. Prediction of the anti-inflammatory mechanisms of curcumin by module-based protein interaction network analysis. *Acta Pharm Sin B*. 2015;5(6):590-595. doi:10.1016/j.apsb.2015.09.005
2. Feng Y, Wang Q, Wang T. Drug Target Protein-Protein Interaction Networks: A Systematic Perspective. *Biomed Res Int*. 2017;2017:1289259. doi:10.1155/2017/1289259
3. Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, Raval A. Identifying hubs in protein interaction networks. *PLoS One*. 2009;4(4):e5344. doi:10.1371/journal.pone.0005344
4. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013;8(8):1551-1566. doi:10.1038/nprot.2013.092
5. Wang Z, Liang L, Yin Z, Lin J. Improving chemical similarity ensemble approach in target prediction. *J Cheminform*. 2016;8:20. Published 2016 Apr 23. doi:10.1186/s13321-016-0130-x
6. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 2002;30(7):1575-1584. doi:10.1093/nar/30.7.1575

## Figures



**Figure 1**

Schematic representation of workflow used in this study: identification of important network attributes and GO enrichment pathway analysis of clusters.



**Figure 2**

Module identification of PPI network using MCL clustering and target protein of each cluster MCL clustering in the network: Each of the 6 clusters were shown along with their target protein (highest degree). The set of interacting proteins (forming edges with a particular target protein) are involved in a particular pathway.