Identification of Significant Proteins in Hypertension Using The Clustering Molecular Complex Detection (MCODE) Method

Lusi Agus Setiani1,*, Wisnu Ananta Kusuma2, Silvia Alviani Zulkarnaen3

1,3Department of Pharmacy, Faculty of Mathematics and Natural Science, Pakuan University, Bogor, West Java, 16143, Indonesia
2Tropical Biopharmaca Research Center, Bogor Agricultural University, Jl. Taman Kencana No. 3, Bogor 16128, Indonesia

Abstract

Hypertension is a condition where the systolic blood pressure value is more than 140 mmHg and the diastolic blood pressure value is more than 90 mmHg. A significant protein is a protein that has the greatest effect or is the center of protein regulation in all biochemical processes. The purpose of this study was to determine the significant protein that has the greatest effect on hypertension by using the clustering Molecular Complex Detection (MCODE) method which will identify areas in the network with the highest density value locally and to determine the mechanism of action of the significant proteins obtained in the setting blood pressure using Gene Ontology and Kyoto Encyclopedia and Genome Analysis (KEGG) by looking at protein signaling pathways for hypertension. The results showed that the STAT3, MAPK3, AKT1, and EDN1 proteins were significant proteins involved in the mechanism of the response to leptin, the ERK1 and ERK2 cascades, the process of nitric oxide biosynthesis, and the cellular response to ROS.

Keywords: Hypertension, MCODE, PPI, Node, Edges

1. Introduction

Hypertension is a condition where the systolic blood pressure value is more than 140 mmHg and the diastolic blood pressure value is more than 90 mmHg [1]. A significant protein is a protein that has the greatest influence or is at the center of protein regulation in all biochemical processes. Therefore, protein-protein interaction (PPI) is very important in regulating blood pressure variations, especially the protein encoded by the causative gene [2]. The incidence of hypertension has increased in adults aged 30-79 years, from 650 million to 1.28 billion in the last 30 years [3].

The high incidence of hypertension is one of the reasons for the need to discover or develop new drugs that can treat this condition. One of the causes of hypertension is heredity or genetic factors. Therefore, the first step that can be taken is to identify significant proteins that are the target of therapy [4]. A significant protein is a protein that has the greatest influence or is at
the center of protein regulation in all biochemical processes. Therefore, protein-protein interaction (PPI) is very important in regulating blood pressure variations, especially the protein encoded by the causative gene [2].

Proteins have different levels of importance based on their association with other proteins. To determine the level of importance of these proteins, the Protein-Protein Interaction (PPI) network can be used, which is a tool used to determine the function of proteins, genes and their relationship to disease [5] and can be used to obtain a potential biomarker which can be a therapeutic target [6].

A computational approach is used to determine the PPI network because it has several advantages including economical costs, more efficient time and easier processing [7]. The PPI network is represented in graph form because it can provide a clear picture in understanding the function and complex structure of proteins. A graph is a combination of nodes (vertices) and edges (sides) where nodes describe proteins and edges describe the relationships between proteins. Research on protein interactions has been carried out in Indonesia in several diseases such as protein interactions in Alzheimer’s disease [8], Parkinson’s [9] and Type 2 Diabetes Mellitus [10]. While research on essential hypertension has been carried out using network topology analysis which produces the most significant protein, namely NOS3 protein [2]. In addition, several studies also reported several proteins involved in blood pressure regulation such as MAPK3 [11] and AKT1 [12].

The purpose of this study was to determine the significant protein that has the greatest influence on hypertension by using the clustering Molecular Complex Detection (MCODE) method which is one of the methods with an algorithm that will identify areas in the network that have a high density locally [13]. In addition, this study also aims to determine the mechanism of action of the significant protein obtained in the regulation of blood pressure using the Gene Ontology Method and the Kyoto Encyclopedia Gene Analysis and Genome (KEGG).

2. Methods

The tool used is an Acer-PC laptop with Windows 8 single-language 64-bit operating system, processor Intel® CoreTM i3-2365M CPU @ 1.40 GHz, 4.00 GB RAM, Cytoscape software version 3.8.2, STRING database, UniProt, OMIM, Malacard, and DAVID. The materials used are proteins obtained from the UniProt, OMIM, and Malacard databases. The steps taken in this research are:

2.1. Protein Candidate Data Collection

In the target or protein collection stage, the UniProt, OMIM, and Malacards databases are used, which are databases that contain a collection of protein data from a disease. This protein search step is carried out by entering the keyword "hypertension" in the search column, and then the protein data obtained is downloaded in Excel format.

2.2. Protein Network Construction – Protein Interaction (PPI)

To determine protein-protein interactions, the PPI network was built using the STRING database so that interactions between proteins can be identified which are represented in the form of networks or graphs. The step taken is to enter protein data that has been obtained previously in the multiple protein column by selecting homo sapiens organisms and setting the Required score in the Advanced setting to high confidence (0.900) then searching. The network results obtained are exported in .tsv format for further analysis.

2.3. Clustering Molecular Complex Detection (MCODE)

Data in .tsv format obtained from the STRING database are then entered into the Cytoscape software for clustering on the MCODE plug-in with a cutoff degree value of 3 [1]. Clusters with a score > 5 [2] were taken for network topology analysis on the total protein obtained in the clusters as well as GO and KEGG analysis in each cluster.
2.4. Network Topology Analysis

After clustering using MCODE, the protein members that are included in the top cluster are analyzed for network topology by measuring the centrality of the network using the centrality parameter. The centrality parameters used are the degree and betweenness centrality parameters [2]. These parameter values can be obtained in the Cytoscape software by analyzing the network on the tools menu, which is then exported to the analysis data in Excel format. Then, the proteins were sorted based on the degree centrality value from the largest to the smallest using the short large to small menu in Microsoft Excel and also sorted based on the largest to the smallest betweenness centrality value using the same method. Furthermore, 5% of the total protein, which has the highest degree of centrality, is taken, which is a significant protein.

2.5. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Analysis of Each Cluster

GO analysis for each cluster was used to determine gene function based on biological processes, molecular functions, and cellular components in each cluster, while KEGG analysis was used to identify relevant pathways for a protein in each cluster. GO and KEGG analysis was carried out on the Database for Annotation, Visualization, and Integrated Discovery (DAVID), where the proteins of each cluster are entered in this database by entering several settings, including select identifiers filled in OFFICIAL_GENE_SYMBOL, species used filled with homo sapiens, and type list filled with gene list. After the GO and KEGG analysis results appeared, the settings used in this study were the default settings (count = 2, EASE = 0.1). Then the GO and KEGG results were downloaded in.txt format, which was then copied and saved in Excel format, and an analysis was carried out on the GO and KEGG results, which had a link to hypertension.

2.6. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Analysis of Significant Proteins

A GO and KEGG analysis of significant proteins was performed to validate the molecular mechanism of the resulting significant protein on blood pressure regulation. The significant proteins obtained were analyzed by GO and KEGG on the DAVID database by entering the names of the significant proteins obtained. Then the same settings are made as for the GO and KEGG analyses for each previous cluster, namely select identifiers filled in OFFICIAL_GENE_SYMBOL, the species used filled in homo sapiens, and the type list filled in the gene list. GO and KEGG results on this significant protein are also downloaded in.txt format, which is then copied and stored in Excel format for analysis on GO and KEGG results related to hypertension. The results of GO and KEGG analysis on significant proteins are displayed in the form of a network visualization that links significant proteins with their respective biological processes and KEGG. The step taken is to create data in Excel format that contains protein data, biological processes, and KEGG, which is then imported into Cytoscape.

2.7. KEGG Signaling Pathway Mapping

KEGG signaling pathway mapping was carried out to describe the pathways involved in various cellular processes and organelle systems involving significant proteins. This description of the KEGG signaling pathway was obtained in the DAVID database by suppressing each signaling pathway associated with hypertension.

3. Result and Discussion

3.1. Protein Candidate Data Collection

The protein candidate data was obtained from three databases, namely OMIM, Uniprot, and Malacards. These three databases have different updates, so the proteins contained in them are not all the same. In addition, the use of these three databases also aims to maximize the search
results for protein candidates so that broader protein data can be obtained. From the search results in the third database, it was found that there were 793 proteins related to hypertension, of which 276 were obtained from the Uniprot database, 69 from the OMIM database, and 448 from the Malacards database. Then remove duplicates to remove the same protein so that 732 proteins are obtained, which are entered in the STRING database to determine the interaction network between proteins.

### 3.2. Protein Network Construction – Protein Interaction (PPI)

The PPI network built using the STRING database obtained 423 nodes with 1618 edges. The PPI network is then imported into the cytoscape for network reduction, and nodes that have no interaction with large networks are not included in this study. This is because proteins that are not directly related to large networks have little or no interaction. After the reduction was carried out, 396 nodes and 1601 edges were obtained. The network visualization can be seen in Figure 1.

![PPI Network Visualization](image)

**Figure 1.** PPI Network Visualization  
\(a\) Before network reduction ; \(b\) After network reduction

### 3.3. Clustering Molecular Complex Detection (MCODE)

The PPI network that has been entered into the cytoscape application is then analyzed with Molecular Complex Detection (MCODE) to find the function of the unknown protein. In addition, MCODE is also used because clusters in protein-protein networks generally correspond to protein complexes involved in the same biological process [14]. According to the principle of "guilt by association," interacting proteins are proteins that have the same biological process and are located in the same cellular compartment. Therefore, in general, proteins of the same molecular complex have the same function [15]. In this study, 17 clusters with different scores were obtained, where the data collected was clusters with scores > 5 [16], which showed the most significant clusters among other clusters [17]. Figure 2 shows six clusters with a score > 5.

### 3.4. Network Topology Analysis

The parameters used in this network topology analysis are centrality parameters, namely degree and betweenness centrality. The calculation of the degree and betweenness centrality values was carried out for all protein members belonging to the six clusters. Table 1 represents 5% of the total 83 proteins in 6 clusters with the highest degree and betweenness centrality values.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Degree</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT3</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>MAPK3</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>AKT1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>EDN1</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Based on the degree and betweenness centrality values in Table 3, it can be seen that the significant proteins obtained were STAT3, MAPK3, AKT1, and EDN1 proteins. The degree value
3.4 Network Topology Analysis

Figure 2. MCODE clustering with a score $> 5$

<table>
<thead>
<tr>
<th>Protein</th>
<th>Degree Centrality</th>
<th>Betweenness centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT3</td>
<td>43</td>
<td>0.065376586</td>
</tr>
<tr>
<td>MAPK3</td>
<td>42</td>
<td>0.075388881</td>
</tr>
<tr>
<td>AKT1</td>
<td>42</td>
<td>0.072065683</td>
</tr>
<tr>
<td>EDN1</td>
<td>29</td>
<td>0.053732184</td>
</tr>
</tbody>
</table>

Table 1. 5% protein based on the best degree centrality and betweenness centrality values

indicates the number of proteins that interact with other proteins. Proteins with high degree values indicate that these proteins are central proteins or proteins that have the greatest role in protein regulation. The next parameter is the betweenness centrality parameter, which works by measuring the number of shortest paths that pass through the node. The shortest path is the shortest distance between two nodes. The more shortest paths that pass through a node, the more important that node is. Proteins with high betweenness values can functionally maintain communication between proteins [18]. The molecular mechanisms taken from several publications for the four significant proteins can be seen in Table 2.

Table 2. Significant Proteins and Their Mechanisms

<table>
<thead>
<tr>
<th>Significant Protein</th>
<th>Description</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT3 Signal Transducer And Activator of Transcription 3</td>
<td>STAT3 activation causes increased angiogenesis [19]</td>
<td></td>
</tr>
<tr>
<td>MAPK3 Mitogen-Activated Protein Kinase 3</td>
<td>Activate ERK 1 which is involved in vasoconstriction and vascular smooth muscle cell growth [11]</td>
<td></td>
</tr>
<tr>
<td>AKT1 AKT serine/threonine kinase 1</td>
<td>AKT1 hyperactivation triggers an increase in NO production which results in enlarged blood vessel diameter and increased blood flow [20]</td>
<td></td>
</tr>
<tr>
<td>EDN1 Endothelin-1</td>
<td>Increased Endothelin-1 can cause a vasoconstrictive effect [21]</td>
<td></td>
</tr>
</tbody>
</table>
3.5 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Analysis of Each Cluster

The proteins included in the six clusters were subjected to GO and KEGG analysis to determine biological processes, cellular components, molecular functions, and signaling pathways related to hypertension in each cluster. GO and KEGG analyses were performed on the DAVID database by selecting biological processes and signaling pathways that correlate with hypertension. In this study, the results of the GO and KEGG analyses were obtained, which can be seen in Figure 3.

![GO and KEGG analysis of 6 clusters](image)

**Figure 3.** GO and KEGG analysis results for each cluster; The red color indicates a biological process; the yellow color indicates cellular components; the blue color indicates molecular function; and the green color indicates the KEGG signaling pathway

Biological processes that occur in cluster 1 are involved in the process of angiogenesis. The biological process of angiogenesis is the process of forming new blood vessels with a vasodilating effect. Disruption of angiogenesis can cause an increase in peripheral resistance and an increase in blood pressure. One of the pro-angiogenic factors is Vascular Endothelial Growth Factor (VEGF), which helps in the process of angiogenesis. The presence of VEGFA inhibition in the body can trigger hypertension [22].

Cluster 2 predominates with BBSome proteins, which are important regulators of various cellular and physiological processes. BBSome is also a large protein complex consisting of eight proteins, namely BBS1, BBS2, BBS4, BBS5, BBS7, BBS8, BBS9, and BBS18. If one of these proteins is not present, it can stop the function of BBSome itself. An example is the highly prevalent obesity BBS mutation that impairs BBSome function [23]. Obesity is one of the factors that causes hypertension by stimulating the activity of the Renin-Angiotensin-Aldosterone System (RAAS) by mediators such as cytokines and hormones. The hormone aldosterone is related to water and sodium retention, which can cause increased blood pressure.
Furthermore, the biological process in cluster 3 is natriuresis pressure. Pressure natriuresis is a process of sodium excretion in the urine. Disturbances in natriuresis pressure can cause blood pressure to increase by interfering with sodium excretion in the kidneys, resulting in sodium accumulation, which triggers increased blood pressure [24].

The biological processes of proteins in cluster 4 are the ERK 1 and ERK 2 cascades. ERK (extracellular signal-regulated kinase) is a member of the mitogen-activated protein kinase family involved in vasoconstriction and the growth of vascular smooth muscle cells. ERK activity is increased in animal models of increased blood pressure. Therefore, inhibition of ERK can reduce the growth of smooth muscle cells and vasoconstriction, which can reduce blood pressure [11].

Cluster 5 proteins have biological functions related to the synthesis of ATP (adenosine triphosphate) in mitochondria. This is because mitochondria are organelles that can be found in all eukaryotic cells and can produce ATP through an oxidative phosphorylation process in which the respiratory chain is an energy source that is used by transferring electrons from NADH to oxygen molecules and then reducing oxygen to H2O and producing ROS. Excess ROS production can cause oxidative stress, which triggers a decrease in NO bioavailability through the chemical reaction of superoxide with NO [25].

The last cluster, namely cluster 6, shows biological processes related to NO production and cellular response to ROS with the PI3K/AKT signaling pathway, MAPK signaling pathway, and RAS (Renin Angiotensin System). ROS and NO play an important role in blood pressure regulation through modulation of the autonomic nervous system, especially in the central nervous system [26]. ROS can react with NO directly, thereby reducing NO bioavailability and triggering an increase in blood pressure [27].

3.6. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Analysis of Significant Proteins

Based on network topology analysis, four significant proteins were identified, including STAT3, MAPK3, AKT1, and EDN1. The significant protein was subjected to GO and KEGG analysis on DAVID to determine its molecular mechanism. To make it easier to read the GO and KEGG pathways of each protein, a network visualization was created. Figure 4 is a visualization of the GO and KEGG pathways of significant proteins related to blood pressure regulation with a p-value 0.05.

In Figure 4, it can be seen that the significant proteins STAT3 and EDN1 are involved in the biological process of response to leptin. Leptin is a hormone that regulates food intake, metabolism, and fat accumulation and can affect blood pressure and contribute to hypertension [28]. Leptin is a member of the adipocytokines produced by adipocytes that act on the endothelium by providing a vasodilatory effect and stimulating NO synthesis in endothelial cells and blood vessels [29]. EDN1 regulates leptin expression in adipocytes and stimulates leptin production via the endothelin-A receptor (ETA) [29]. Leptin inhibition induced by SOCS3 (Suppressor of Cytokine Signaling 3) upregulation is mediated by STAT3 activation. This can lead to a decrease in NO, which triggers hypertension [30].

Significant proteins, MAPK3 and EDN1, are involved in the biological processes of the ERK1 and ERK2 cascades. Extracellular signal-regulated kinase (ERK), commonly referred to as MAPK3, is a member of the mitogen-activated protein kinase family and is involved in the process of vasoconstriction and the growth of vascular smooth muscle cells. In differentiated contractile vascular smooth muscle, ERK1 and ERK2 are involved in the regulation of vascular smooth muscle contraction. ERK activity increases in the vascular smooth muscle cells of rats who experience hypertension, so inhibition of ERK activation can reduce the growth of vascular muscle cells and vasoconstriction.

Furthermore, the significant proteins MAPK3 and AKT1 are involved in the biological processes of cellular response to reactive oxygen species. Increased production of ROS (reactive oxygen species) plays a role in various chronic diseases, including hypertension. ROS are involved in inflammation, hypertrophy, migration, fibrosis, and angiogenesis [31]. ROS control endothelial
Figure 4. Visualization of biological processes and the KEGG pathway protein is significantly related to blood pressure regulation with a p-value < 0.05. Yellow color indicates significant proteins; blue color indicates biological processes; and green color indicates the KEGG signaling pathway function and vascular tone in the vascular system, which means that increased ROS production contributes to endothelial and vascular smooth muscle cell dysfunction. In addition, ROS can also cause increased contractility, vascular smooth muscle cell growth, and apoptosis [32].

The significant proteins AKT1 and EDN1 are involved in the positive regulation of biosynthetic nitric oxide. According to [33], AKT1 preferentially phosphorylates endothelial nitric oxide synthase (eNOS) and promotes the release of nitric oxide (NO). NO is a vasodilator and relaxation factor derived from the endothelium [34]. Decreased NO bioavailability causes endothelial dysfunction and is a risk factor for hypertension. Previous studies reported that mice with impaired NO-producing genes (eNOS) had higher blood pressure levels compared to control animals [35]. In addition, an imbalance between NO and endothelin-1 can contribute to changes in vascular tone associated with increased blood pressure [36].

3.7 KEGG Signaling Pathway Mapping

KEGG signaling pathways involving significant proteins obtained from this study are the JAK-STAT signaling pathway, Relaxin signaling pathway, Adrenergic signaling in cardiomyocytes, Vascular smooth muscle contraction, RAS signaling pathway, Lipid and atherosclerosis, and RAP-1 signaling pathway.

Of the seven signaling pathways, the relaxin signaling pathway and lipid and atherosclerosis are signaling pathways that mostly involve the four significant proteins. The relaxin signaling pathway is related to blood pressure regulation because relaxin is a peptide hormone that has a mechanism for relaxing smooth muscles. Previous research stated that rats with spontaneous hypertension induced by relaxin experienced a decrease in blood pressure. In lowering blood pressure, relaxin works by increasing NOS (Nitric Oxide Synthase) so that NO levels increase [37]. In addition to its effect on increasing NO production, relaxin also has effects on endothelin inhibition, angiotensin II inhibition, VEGF production, and matrix metalloproteinase production. These effects lead to
systemic and renal vasodilatation, increased arterial compliance, and other vascular changes [38].

Elevated blood pressure is associated with the development of atherosclerosis in that the volume of atherosclerotic lesions may increase and the vessel walls exhibit collagen deposition and plaque accumulation [39]. Several studies have reported that STAT3 activation has an important role in the development of atherosclerosis. STAT3 activation contributes to endothelial dysfunction and inflammation and thus becomes an important modulator during atherosclerosis [40]. MAPK3 also has an important role in the development of atherosclerosis, where MAPK3 plays a role in modulating atherosclerotic lesions through regulation of macrophage foam cell formation [41]. In addition, loss of AKT1 signaling can also lead to extensive coronary atherosclerosis due to endothelial dysfunction and increased apoptosis in vascular cells [42]. Figure 5 is a signaling pathway map of the relaxin signaling pathway, lipids, and atherosclerosis.

Figure 5. Relaxin signaling pathway, lipid and atherosclerosis
4. Conclusion

Based on the Clustering Molecular Complex Detection (MCODE) method, four significant proteins associated with hypertension were identified, namely STAT3, MAPK3, AKT1, and EDN1. The STAT3 and EDN1 proteins are involved in the biological process of response to leptin, which is a member of the adipocytokines produced by adipocytes and acts on the endothelium by providing a vasodilatory effect and stimulating NO synthesis in endothelial cells and blood vessels, so that leptin inhibition can cause hypertension. In addition, the significant proteins MAPK3 and AKT1 are involved in the biological processes of cellular response to reactive oxygen species. Increased production of ROS (Reactive Oxygen Species) plays a role in various chronic diseases, including hypertension. ROS are involved in inflammation, hypertrophy, migration, fibrosis, and angiogenesis. The AKT1 and EDN1 proteins are involved in the positive regulation of nitric oxide biosynthetic biological processes in which reduced NO bioavailability causes endothelial dysfunction and is a risk factor for hypertension.

5. Acknowledgement

Thank you to the Department of Mathematics for providing support and input in this research.

References


