NETWORK PHARMACOLOGY ANALYSIS TO IDENTIFY BLACK SEA CUCUMBER BIOACTIVE COMPOUNDS POTENTIAL

Muhammad Naufal Abdillah

Department of Computer Sciences, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, 16680, Indonesia
mnaufalabdillah@apps.ipb.ac.id

Wisnu Ananta Kusuma

Department of Computer Sciences, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, 16680, Indonesia
ananta@apps.ipb.ac.id

Mushthofa

Department of Computer Sciences, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, 16680, Indonesia
mush@apps.ipb.ac.id

Mala Nurilmala

Department of Aquatic Product Technology, Faculty of Fisheries and Marine Sciences, IPB University, Bogor, 16680, Indonesia
mnurilmala@apps.ipb.ac.id

Abstract

As a maritime country, Indonesia possesses abundant marine resources, including the black sea cucumber (Holothuria atra). The black sea cucumber has not been fully utilized due to its slightly bitter taste and relatively low price. However, several studies have shown that the black sea cucumber has various health benefits. Based on this, the present research aims to identify the compounds in the black sea cucumber that can be used as medicine, and the diseases that can be addressed with these compounds using a network pharmacology analysis approach. The network pharmacology approach includes topology analysis of protein-protein interactions and fuzzy clustering to group protein targets, as well as biological information such as gene ontology (GO) and pathways. This study identifies 43 compounds found in black sea cucumber can be used as medicine to address three diseases: Malignant Neoplasm of Breast, Leukemia Myelocytic Acute, and Colorectal Carcinoma by targeting seven protein targets associated with each disease such as AKT1, AR, ESR1, TP53, JAK2, CTNNB1, GNAS.

***Keywords*: Bioinformatics; black sea cucumber; fuzzy clustering; network pharmacology; topological analysis.**

1. Introduction

In Asia, sea cucumbers are commonly consumed as health food and considered a delicacy. When used as a health food, sea cucumbers have numerous health benefits, serving as tonics and possessing the ability to heal various diseases and disorders [Bordbar et al. 2011]. In recent years, the health benefits of sea cucumbers have been scientifically validated, showcasing their effectiveness as wound-healing agents, antibacterials, antioxidants, blood clot preventers, and inhibitors of cancer cell growth [Pangestuti and Arifin 2018]. One sea cucumber species that has seen increased demand in the international market is the black sea cucumber (*Holothuria atra*) [Hartati et al. 2021]. However, black sea cucumber is one of the sea cucumber species with relatively lower commercial value and is seldom consumed due to its slightly bitter taste [Salim Hanafi et al. 2019].

Black sea cucumber has been found to contain protein compounds that can act as immunomodulators, substances that can stimulate or suppress the immune system and potentially serve as medicinal agents [Fadhlia

2017]. However, to date, there is a lack of specific research elucidating the diseases that can be addressed using the protein compounds found in black sea cucumber.

The search for diseases that can be treated using specific compounds can be conducted through laboratory research involving in vivo or in vitro studies, as well as in silico approaches. Experimental research conducted in the laboratory, such as in vitro or in vivo studies, requires significant time and resources [Acencio and Lemke 2009]. Hence, in silico research can be conducted as an alternative. In silico research refers to studies conducted using computational approaches [Ekins et al. 2007].

The identification of diseases that can be addressed with specific drug compounds involves searching for the protein targets associated with those diseases that interact with the black sea cucumber compounds. This challenge can be overcome by using the network pharmacology approach [Umar et al. 2022], which systematically describes a disease and its interactions with drugs in the body through biological networks [Muhammad et al. 2018].

By exploring compound-target networks, protein targets associated with black sea cucumber compounds can be identified. Protein targets are proteins that are associated with at least one disease [Ghadermarzi et al. 2019]. However, there is many protein targets associated with black sea cucumber compounds. Therefore, the identification of significant proteins among the multitude of protein interactions is necessary.

Topological analysis of protein-protein interaction (PPI) networks can aid in identifying proteins with crucial roles in network connectivity [Soofi et al. 2020]. Topological analysis helps in pinpointing significant proteins by measuring the most important nodes using various centrality measurements [Pan et al. 2016]. There are several types of centrality measurements. Among the centrality measures commonly used for identifying important proteins, degree centrality is frequently employed [Lei et al. 2019; Wang et al. 2020; Umar et al. 2022; Zuhri et al. 2022].

Searching for significant proteins solely based on protein-protein interaction data can lead to a high rate of false positives [Mahdavi and Lin 2007]. This can be mitigated by incorporating biological information in the process of identifying significant proteins [Zhong et al. 2021] . Gene Ontology (GO) and pathways are biological data that can be utilized in the search for significant proteins [Reimand et al. 2019; Zhu et al. 2019]. With the integration of biological information such as GO and pathways, the scale of PPI data expands. Clustering methods can be used to cluster PPI data. A study by Fernando (2017) [Fernando 2017] employed MCODE clustering to find significant proteins; however, MCODE clustering lacks the ability to detect overlaps.

Another soft clustering method is Fuzzy K-Partite clustering, which clusters large-scale PPI data into small-scale, functionally related clusters [Ramadhani 2021]. Fuzzy K-Partite clustering was developed by Hartsperger et al. (2010) [Hartsperger et al. 2010]. This method enables tripartite clustering of disease-gene-protein data. The fuzzy approach is employed due to its capability for overlap clustering, where data can belong to more than one cluster. This is crucial as each network is multifunctional and can belong to multiple clusters [Hartsperger et al. 2010].

For example, Umar et al in 2022 [Umar et al. 2022] conducted the identification of drug candidates and protein targets from *Curculigo* spp. compounds for the treatment of diabetes mellitus using network pharmacology, molecular docking, and molecular dynamics simulation techniques. By using the same methods, Zuhri et al in 2022 [Zuhri et al. 2022] conducted the identification of drug candidates and protein targets from *Tinospora crispa* as insulin sensitizer. In this study, the identification of diseases, along with their significant protein targets that can be addressed by bioactive compounds in black sea cucumber, is performed using network pharmacology. The approach includes the construction of a compound-target network, topology analysis, and clustering of protein-protein interactions (PPI), as well as enrichment analysis using Gene Ontology (GO) and pathway analysis (Fig 1).

2. Material and Methods

2.1. *Constructing database of known black sea cucumber compounds*

The compound data of black sea cucumber was obtained from Fadhlia's research in 2017 [Fadhlia 2017], and its chemical structures were searched for in the Simplified Molecular Input Line Entry System (SMILES) format using the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) [Kim et al. 2023]. Subsequently, the compound data was inputted into the SwissADME database (http://www.swissadme.ch/) [Daina et al. 2017] to assess the drug-likeness and druggability of the compounds. This assessment was based on the Lipinski Rule of Five and Abbott Bioavailability Score with a threshold of 0.5 [Lipinski et al. 2001; Lipinski 2004]. Only the compounds that met both criteria were considered in the analysis.

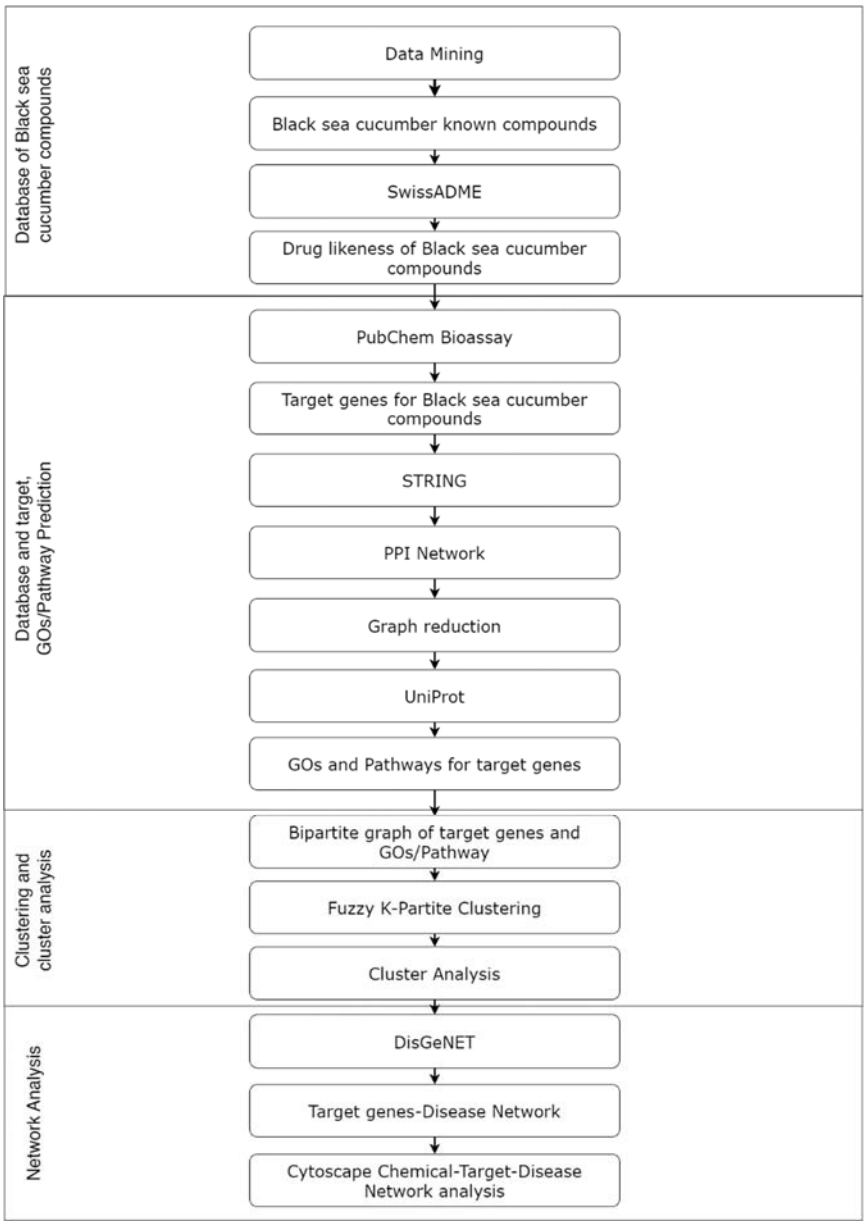

Fig. 1 General workflow of present work using network pharmacology analysis

2.2. *Constructing database of protein target involved to black sea cucumber compounds*

The compounds that have passed the data filtration stage are used as input in the PubChem BioAssays database (https://pubchem.ncbi.nlm.nih.gov/docs/bioassays) [Wang et al. 2009] to obtain protein target data associated with black sea cucumber compounds. This data is then used as input in the STRING database (https://string-db.org/) [Szklarczyk et al. 2023] with the parameter set to search for the full STRING network type and a required score set to medium confidence (0.400) to obtain protein-protein interaction (PPI) data in the form of a network. The resulting network is then reduced based on the degree centrality values of each protein. Top 50 protein targets with the highest degree centrality values are selected, this is done in order to limit the number and ensure that the proteins used for the next stage are indeed significant proteins. Degree centrality for each protein targets is calculated using following equation.

$$DC(u) = \Sigma_v a_{u,v} \qquad (1)$$

Where *DC(u)* represents the degree centrality value of node u. $a_{u,v}$ is the adjacency matrix entry for node u and node v, with a value of one if they are connected and zero otherwise. Next, for each of the 50 proteins, their respective gene ontology (GO) and pathway data are retrieved. The GO data includes three categories: molecular function, cellular component, and biological process. The GO data is obtained from the UniProt database

(https://www.uniprot.org/) [Bateman et al. 2015], while the pathway data is retrieved using Metascape (https://metascape.org/) [Zhou et al. 2019].

2.3. *Clustering and clusters analysis*

The significant protein targets data, GO molecular function GO cellular component, GO biological process, and pathway data obtained are then used to construct a bipartite graph. The data is transformed into an adjacency matrix, which serves as input for the clustering step. Clustering is performed between proteins and GO molecular function, GO cellular component, GO biological process, or pathway data using Fuzzy K-Partite clustering.
The initialization of the number of clusters to be formed should be done in advance, as the Fuzzy K-Partite clustering method is a non-hierarchical clustering method. The equations to determine the maximum number of clusters for each GO or pathway data and the maximum number of clusters for protein data can be seen in Equation 2 and Equation 3 [Hartsperger et al. 2010; Ramadhani 2021; Ode et al. 2022].

$$C_{go/pathway} = \frac{N_{go/pathway}}{10} \qquad (2)$$

As in Eq. 2, $C_{go/pathway}$ represents the maximum number of clusters for each GO or pathway data that can be created using Fuzzy K-Partite clustering, and $N_{go/pathway}$ is the number of nodes in each GO or pathway data. The maximum number of clusters for each GO or pathway obtained from the calculation process will be used to determine the maximum number of protein cluster, which can be seen in this following equation.

$$C_{protein} = C_{go/pathway}\sqrt{\frac{N_{protein}}{N_{go/pathway}}} \qquad (3)$$

In Eq. 3, $C_{protein}$ represents the maximum number of protein clusters. $N_{protein}$ is the number of nodes in the protein targets data. This algorithm outputs the membership values of proteins and GO/pathways in each cluster, as well as the inter-cluster connectivity values between protein clusters and GO/pathway clusters. The inter-cluster connectivity value will be high if the percentage of cluster members is low, and vice versa. And the equation to calculate the cost function value can be seen in Equation 4.

$$f(H,C) = \Sigma_{i<j}\left\|A^{(ij)} - C^{(i)}B^{(ij)}C^{(j)}\right\|_F^2 \qquad (4)$$

$\|.\|_F^2$ is a value that represents the squared Frobenius norm, which is the sum of squares of the matrix elements. A is an adjacency matrix between protein targets and GOs/pathways, each element in matrix A is valued one if there is an edge connecting between protein target and GO/pathway, conversely, it is valued zero. B represents the connectivity matrix between each protein clusters and GO/pathway clusters. Membership value between each protein and its protein clusters or GOs/pathways and its GO/pathway clusters represented by C in form of membership value matrix. These three matrices are used as input for Fuzzy K-Partite clustering, which the algorithm can be seen in Table 1 [Hartsperger et al. 2010].

In this study, there are four sets of data paired with protein target data, namely molecular function GO, cellular component GO, biological process GO, and pathway. Based on this, the clustering process is performed four times, which includes clustering of proteins with molecular function GO, cellular component GO, biological process GO, and pathway. This is done due to the absence of information regarding the relationships between each type of GO and pathway. The formed cluster by using fuzzy k-partite clustering, whether they are protein target clusters or GO/pathway clusters, have interconnectivity values between the two independent sets. By comparing the interconnectivity values among clusters, those with highest interconnectivity are selected. Subsequently, the members of each best cluster are the ones that consist of protein targets and GO term molecular function, GO term cellular component, GO term biological process, or pathway associated with the compounds from the black sea cucumber.

2.4. *Prediction of diseases treatable by black sea cucumber compounds*

All the significant protein targets associated with black sea cucumber compounds obtained from the previous clustering step are then used as input in the DisGeNET database (https://www.disgenet.org/search) [Piñero et al. 2015] to obtain associations between the protein targets and diseases. The process of searching for diseases related to significant protein targets is carried out by using the API of the DisGeNET database, which takes significant protein targets obtained from the subsequent stage as input. Subsequently, information regarding diseases associated with the input significant protein targets is retrieved. These protein-target-disease associations are then linked back to the bioactive compounds in black sea cucumber that can be used as potential drugs. The final

visualization is formed as an interaction network between the bioactive compounds in black sea cucumber, protein targets, and diseases that have the potential to be treated with these compounds.

Fuzzy K-Partite clustering algorithm

Input: *k-partite* protein target graph P with possibly non-negatively weighted edge matrices $A^{(ij)}$, $\boldsymbol{i} < \boldsymbol{j}$, number of clusters $\boldsymbol{m_1}, \ldots, \boldsymbol{m_k}$

Output: fuzzy clustering membership value $C^{(i)}$ and *k-partite* cluster interconnection value graph H given by matrices $B^{(ij)}$

```
Initialize C^(i), B^(ij) to random non-negative matrices
Normalize C_rs^(i) ← C_rs^(i) / (Σ_t C_rt^(i)) for all i, r, s
repeat
    update fuzzy cluster
    for i ← 1, ..., k do
        C^(i) ← C^(i) ⊗ (Σ_{j≠i} A^(ij) C^(i) B^(ij)T) ⊘ (C^(i) B^(ij) C^(j)T C^(j) B^(ij)T)
        Normalize C_rs^(i) ← C_rs^(i) / (Σ_t C_rt^(i)) for all r, s
    end
    Update k-partite cluster graph H
    for i ← 1, ..., k − 1 do
        for j ← i + 1, ..., k do
            B^(ij) ← B^(ij) ⊗ (C^(i)T A^(ij) C^(j)) ⊘ (C^(i)T C^(i) B^(ij)T C^(j)T C^(j))
        end
    end
Until convergences;
```

Note: ⊗ and ⊘ symbolize element-wise multiplication and division, respectively

Table 1 Fuzzy K-Partite clustering algorithm

3. Results and Discussion

3.1. *Database of known black sea cucumber compounds*

A total of 102 compounds were obtained from Fadhlia's research in 2017 [Fadhlia 2017]. After inputting these compounds into the PubChem and SwissADME databases [Daina et al. 2017; Kim et al. 2023], their chemical structures were obtained in the SMILES format, and the Abbott Bioavailability Score was calculated. Out of the 102 compounds, 13 compounds are duplicated, and 28 compounds did not meet the criteria as drugs, as their Abbott Bioavailability Score was less than 0.5. The molecular weight ranges from 109.15 to 613.79, with Hypotaurine as the lightest molecule and Indinavir as the heaviest molecule. Predicted gastrointestinal absorption showed 47 of 61 compounds classified as highly absorbed in GI tract. The drug likeness of compounds was obtained through the Lipinski Rule of Five, where 40 compounds satisfy all five criteria according to Lipinski's rules, and the remaining 21 compounds have 1-2 violations against the Lipinski Rules. The selected compounds can be seen in Table 2.

Compound	Formula	MW	GI absorption	Lipinski	Bioavailability Score
Polygodial	C15H22O2	234.33	High	0	0.55
Astaxanthin	C40H52O4	596.84	High	0	0.55
Carvacrol	C10H14O	150.22	High	0	0.55
Phenelzine	C8H12N2	136.19	High	0	0.55
Meclizine	C25H27ClN2	390.95	High	1	0.55
Arachidonate	C20H32O2	304.47	High	1	0.55
Riboflavin	C17H20N4O6	376.36	Low	0	0.55
Hypotaurine	C2H7NO2S	109.15	High	0	0.55
Xylitol	C5H12O5	152.15	Low	0	0.55
Hydrocortisone cypionate	C29H42O6	486.64	High	0	0.55

Table 2 Prediction of black sea cucumber compounds and its druggability

Compound	Formula	MW	GI absorption	Lipinski	Bioavailability Score
20-Hydroxyecdysone	C27H44O7	480.63	High	1	0.55
Buprenorphine	C29H41NO4	467.64	High	0	0.55
Phylloquinone	C31H46O2	450.7	Low	1	0.55
Calcidol	C2CaMgO6	184.4	Low	0	0.55
Solanidine	C27H43NO	397.64	High	1	0.55
Chenodeoxycholate	C24H40O4	392.57	High	0	0.56
Anandamide	C22H37NO2	347.53	High	0	0.56
Quercetin	C15H10O7	302.24	High	0	0.55
Aphidicolin	C20H34O4	338.48	High	0	0.55
Cetraxate	C17H23NO4	305.37	High	0	0.55
Taxifolin	C15H12O7	304.25	High	0	0.55
Tamoxifen	C26H29NO	371.51	Low	1	0.55
(9Z)-Octadecenoic acid	C18H34O2	282.46	High	1	0.85
Inosine	C10H12N4O5	268.23	Low	0	0.55
Albendazole	C12H15N3O2S	265.33	High	0	0.55
N-Acetylserotonin	C12H14N2O2	218.25	High	0	0.55
Dehydroemetine	C29H38N2O4	478.62	High	0	0.55
Nicotinamide	C6H6N2O	122.12	High	0	0.55
Naproxen	C14H14O3	230.26	High	0	0.85
Sterol	C17H28O	248.4	High	1	0.55
Carpaine	C28H50N2O4	478.71	High	0	0.55
Terfenadine	C32H41NO2	471.67	High	1	0.55
Puromycin	C22H29N7O5	471.51	Low	1	0.55
Glycyrrhetinate	C30H45O4-	469.68	High	1	0.56
Hydroxytamoxifen	C26H29NO2	387.51	High	1	0.55
Morphine	C17H19NO3	285.34	High	0	0.55
Taurine	C2H7NO3S	125.15	High	0	0.55
Procaine	C13H20N2O2	236.31	High	0	0.55
Telmisartan	C33H30N4O2	514.62	Low	2	0.85
Selegiline	C13H17N	187.28	High	0	0.55
Mannitol	C6H14O6	182.17	Low	1	0.55
Flavonol	C15H10O3	238.24	High	0	0.55
Khellin	C14H12O5	260.24	High	0	0.55
Phencyclidine	C17H25N	243.39	High	0	0.55

Table 2 Prediction of black sea cucumber compounds and its druggability (Continued)

Compound	Formula	MW	GI absorption	Lipinski	Bioavailability Score
Pyridoxal phosphate	C8H10NO6P	247.14	High	0	0.56
Indinavir	C36H47N5O4	613.79	High	1	0.55
Fosinopril	C30H46NO7P	563.66	Low	1	0.56
Ambenonium	C28H42Cl2N4O2++	537.56	High	1	0.55
Fucidic acid	C31H48O6	516.71	Low	1	0.56
Acadesine	C9H14N4O5	258.23	Low	0	0.55
Tetracaine	C15H24N2O2	264.36	High	0	0.55
Orphenadrine	C18H23NO	269.38	High	0	0.55
Cycloheximide	C15H23NO4	281.35	High	0	0.55
Pregnenolone	C21H32O2	316.48	High	0	0.55
Danazol	C22H27NO2	337.46	High	0	0.55
Betulinic acid	C30H48O3	456.7	Low	1	0.85
Fluocinolone	C21H26F2O6	412.42	High	0	0.55
Butoconazole	C19H17Cl3N2S	411.78	High	1	0.55
Oxacillin	C19H19N3O5S	401.44	High	0	0.56
Amsacrine	C21H19N3O3S	393.46	High	0	0.55

Table 2 Prediction of black sea cucumber compounds and its druggability (Continued)

As can be seen in Table 1, there are 61 black sea cucumber compounds that meet Lipinski Rule of Five and Abbott Bioavaibility Score criteria and can be used as drugs. By using these compounds, a search for protein targets associated with these compounds can be conducted.

3.2. *Protein-protein interaction network related to black sea cucumber compounds*

By using the selected black sea cucumber compounds, 1226 protein targets that interact with black sea cucumber compounds were obtained from the PubChem BioAssays database [Wang et al. 2009]. All of these protein targets were used as input in the STRING database [Szklarczyk et al. 2023] to obtain a protein-protein interaction network. The resulting network consists of 4,598 proteins and 27,695 interactions. (Fig 2).

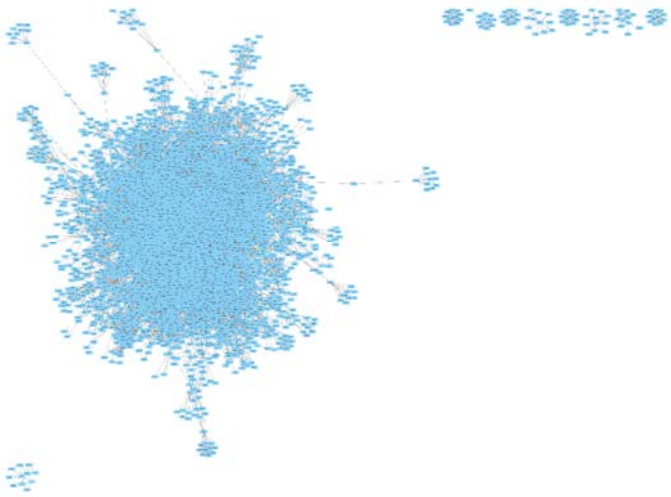

Fig. 2 Protein-protein interaction network

Obtained protein-protein interaction network was reduced by removing subgraphs that are not connected to the main network, as shown in the red box in Fig 3. This process must be done because the subgraph that are not connected to the main network can affect the degree centrality calculation for each protein. After this process, resulting in a network comprising 4,499 protein targets and 27,330 interactions.

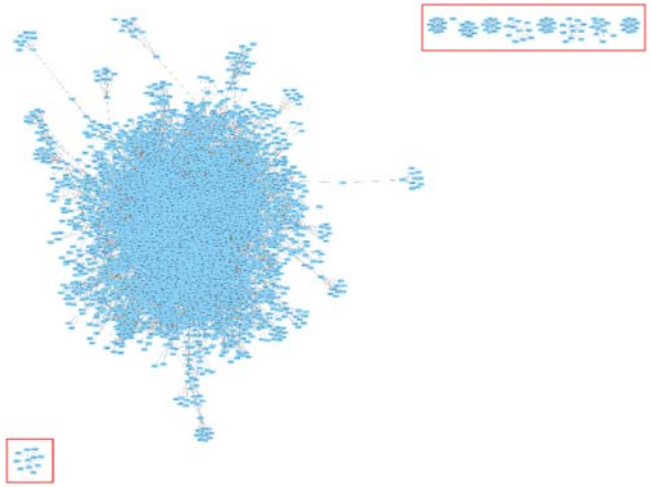

Fig. 3 Graph reduction process by eliminating subgraph that are not connected to the main network

The degree centrality values for each protein target in the reduced network are calculated by using Eq. 1. After calculation process, protein targets are sorted from the highest degree centrality and the top 50 protein targets with the highest degree centrality values are selected. Table 3 below shows the 50 protein targets and its degree centrality:

Protein Target	Degree Centrality
TP53	290
EP300	181
HSP90AA1	174
SRC	164
CTNNB1	159
GNAQ	144
AKT1	121
MAPK3	120
EGFR	114
HDAC1	110
CYP3A4	103
GNAS	102
JUN	99
STAT3	99
MAPK1	98
MYC	96
HRAS	95
PTPN11	93
GRB2	91
CDC42	89
ESR1	86
GNB1	86
UBC	86

Table 3 Top 50 protein targets with highest degree centrality

Protein Target	Degree Centrality
GNAI1	85
CREBBP	83
PPARGC1A	82
CYP1A1	79
RXRA	78
SHC1	77
GNA11	77
HIF1A	75
PCNA	73
DLG4	72
JAK2	71
UBE2I	71
CBL	70
ALB	70
CDK1	70
NCOA1	70
PPARG	69
ARRB1	69
GNAI2	68
PPARA	68
RELA	68
RPS27A	68
TRAF6	67
PLCG1	67
CASP3	66
APP	65

Table 3 Top 50 protein targets with highest degree centrality (Continued)

By using the 50 proteins obtained as input in the UniProt and Metascape databases [Bateman et al. 2015; Zhou et al. 2019], a total of 1,458 biological process GO terms with the top three most frequent are positive regulation of transcription by RNA polymerase II, positive regulation of DNA-templated transcription, and signal transduction, 229 cellular component GO terms with the top three most frequent are cytosol, cytoplasm, and nucleoplasm, 326 molecular function GO terms with the top three most frequent are enzyme binding, identical protein binding, and chromatin binding, 148 pathways were obtained with the top three most frequent are thyroid cancer, bladder cancer, and endometrial cancer. These were then used to construct a bipartite graph connecting the protein targets using Cytoscape [Shannon et al. 2003]. The visualization of the formed bipartite graph can be seen in Fig 4.

Each bipartite graph is then transformed into an adjacency matrix, which will be used as input in the clustering process. Adjacency matrix is a matrix that contains values of zero or one. If an edge has an interaction between a protein target and molecular functions GO term, cellular components GO term, biological processes GO term, or pathways, it will have a value of one, and vice versa.

3.3. *Clustering and clusters analysis*

Using the adjacency matrix formed as input in the Fuzzy K-Partite clustering, four results were obtained from the clustering process. The clusters that have been formed, whether they are protein clusters or GO/pathway clusters, have inter-cluster connectivity values with each other. By comparing these inter-cluster connectivity values, the cluster with the highest connectivity is selected. The members of each selected cluster represent a cluster that includes significant protein targets and their associated molecular function GO terms, cellular component GO terms, biological process GO terms, or pathways related to the compounds from black sea cucumber. In other words, these clusters consist of proteins and functional annotations or pathways that are

relevant to the compounds from black sea cucumber. The best results of clustering process can be seen in Table 4.

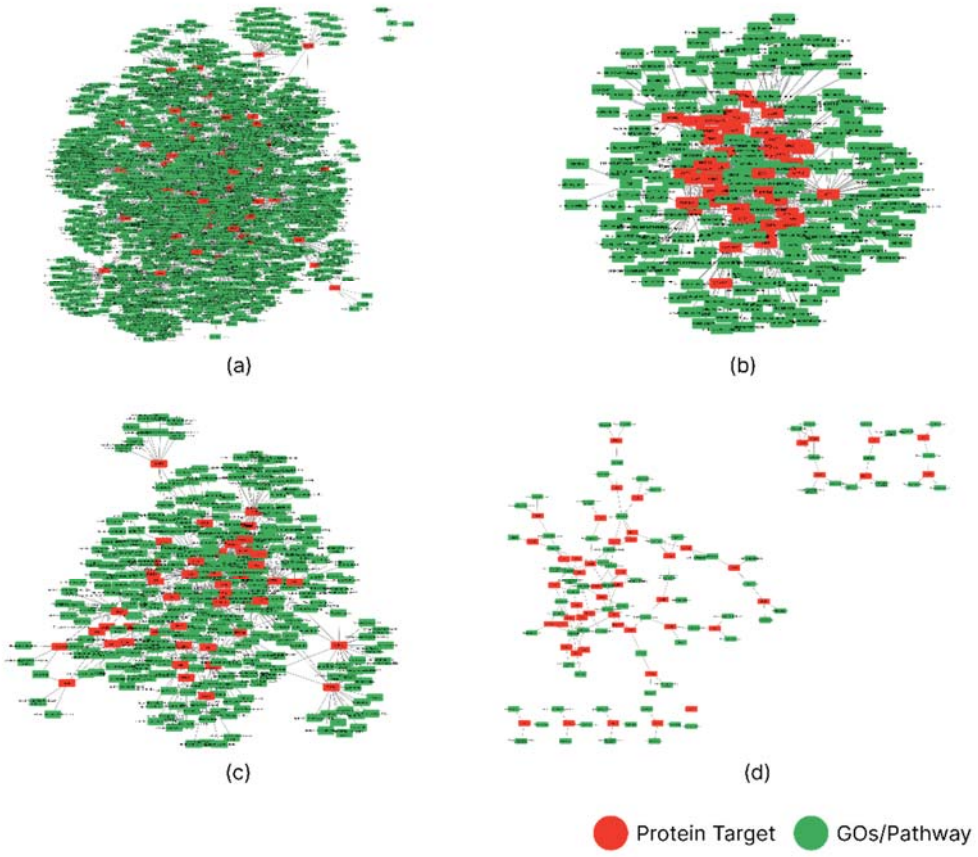

Fig. 4 (a) Bipartite graph of protein targets and biological process GO terms (b) Bipartite graph protein targets and cellular component GO terms (c) Bipartite graph protein targets and molecular function GO terms (d) Bipartite graph protein targets and pathway

It can be observed that the process of clustering protein targets and pathways is the quickest to converge, reaching convergence in 12 iterations with the smallest cost function value of 138.95. This could be attributed to the relatively smaller amount of data used and the presence of many proteins linked to the same pathway. The largest cost function value is found in the clustering process of protein targets and biological process GO terms, reaching 2081.25. The largest formed cluster comprises 42 proteins in the process of clustering protein targets and biological process GO terms, and the smallest one contains 23 proteins in the process of clustering protein targets and molecular function GO terms.

No.	Clustering process	Iteration	Cost function	Protein target members	GOs/ Pathway members	Inter-cluster connectivity value
1	Protein targets-Biological Process GO term	19	2081.25	42 protein targets	1401 Biological Process GO term	0.041
2	Protein targets-Cellular Component GO term	25	573.51	24 protein targets	153 Cellular Component GO term	0.122
3	Protein targets-Molecular Function GO term	26	687.78	23 protein targets	185 Molecular Fucntion GO term	0.094
4	Protein targets-Pathway	12	138.95	40 protein targets	51 Pathway	0.008

Table 4 Clustering results using Fuzzy K-Partite clustering and the best cluster members values

The highest inter-cluster connectivity value is observed in the process of clustering protein targets and cellular component GO terms, with a value of 0.122, and the lowest in the process of clustering protein targets with pathways, with a value of 0.008. All the protein targets that are members of the best clusters from each clustering process are selected and combined, resulting in a total of 50 proteins. No significant protein targets are eliminated during the clustering process. This means that all the identified protein targets are considered relevant and are retained in the final set of 50 proteins.

3.4. *Prediction of diseases treatable by black sea cucumber compounds and network analysis*

The significant protein targets obtained from the clustering process are used as input in the DisGeNET database to obtain protein-disease associations. The visualization of the protein-disease network can be seen in Fig 5.

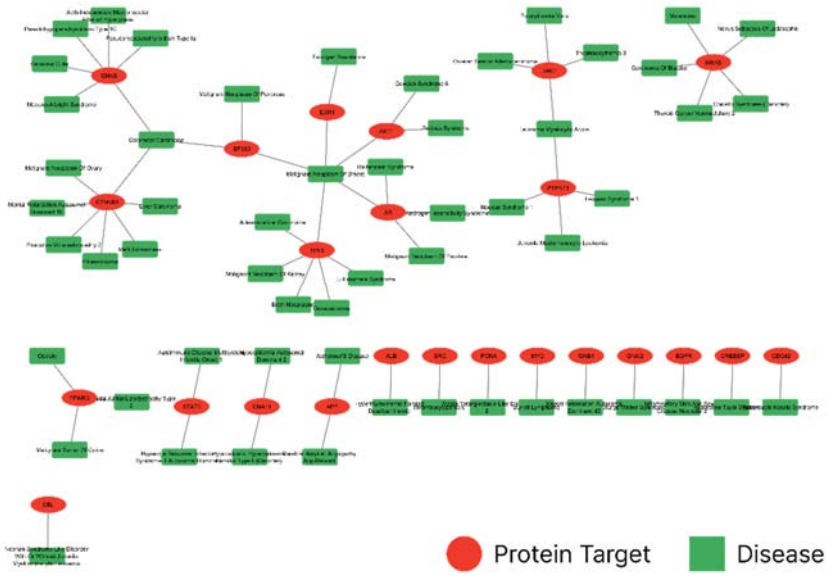

Fig. 5 Visualization of protein targets and diseases network

The protein-disease network resulted in a total of 24 protein targets associated with 56 diseases. The diseases with the highest occurrence in the network are Malignant Neoplasm of Breast, Colorectal Carcinoma, and Leukemia Myelocytic Acute. These diseases, along with their associated protein targets, are linked back to the bioactive compounds in black sea cucumber that can be used as potential drugs in the form of a network. The visualization of the network, which includes the compounds, protein targets, and potential diseases that can be treated, can be seen in Fig 6.

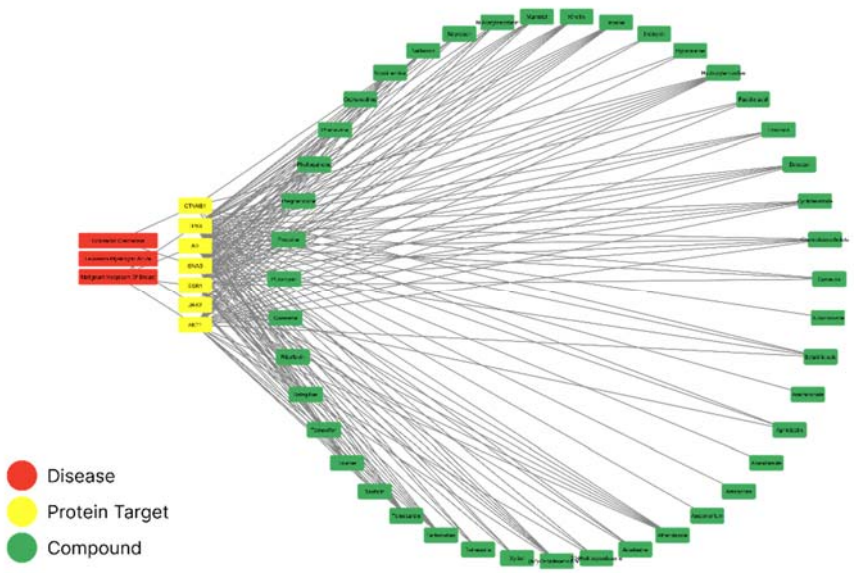

Fig. 6 Visualization of compounds, protein targets and diseases network

This network consists of 43 black sea cucumber compounds, namely Taurine, Naproxen, Procaine, Arachidonate, Telmisartan, Hypotaurine, Selegiline, Mannitol, N-Acetylserotonin, Flavonol, Khellin, Indinavir, Nelfinavir, Ambenonium, Fucidic acid, Acadesine, Tetracaine, Orphenadrine, Cycloheximide, Pregnenolone, Danazol, Hydroxytamoxifen, Terfenadine, Puromycin, Betulinic acid, Butoconazole, Amsacrine, Chenodeoxycholate, Xylitol, Carvacrol, Phenelzine, Riboflavin, 20-Hydroxyecdysone, Phylloquinone,

Anandamide, Quercetin, Aphidicolin, Taxifolin, Tamoxifen, (9Z)-Octadecenoic acid, Inosine, Albendazole, and Nicotinamide. These compounds have the potential to treat three diseases: Malignant Neoplasm of Breast, with protein targets AKT1, AR, ESR1, and JAK2. Leukemia Myelocytic Acute, with the protein target JAK2. And Colorectal Carcinoma, with protein targets CTNNB1 and GNAS.

4. Conclusion and Future Scope

This research aimed to identify potential diseases that can be treated with black sea cucumber compounds using a network pharmacology approach. By employing the compound-target network approach and topological analysis with centrality measurements from protein-protein interaction data, significant protein targets that interact with black sea cucumber compounds and can be used as drugs were discovered. Utilizing Fuzzy K-Partite clustering and conducting enrichment analysis with Gene Ontology (GO) and pathways, the significant protein targets found were validated, confirming their significance. These protein targets were then used to identify diseases associated with them, which were subsequently linked back to black sea cucumber compounds that can be used as potential drugs. A total of 43 compounds found in black sea cucumber can be used to address three diseases: Malignant Neoplasm of Breast, Leukemia Myelocytic Acute, and Colorectal Carcinoma by targeting specific protein targets associated with each disease. For further research, protein-protein interaction data can be utilized by cross-checking it with gene expression data, resulting in the construction of a weighted graph. Additionally, comparisons can be made with other soft clustering methods.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1]. Acencio ML, Lemke N. 2009. Towards the prediction of essential genes by integration of network topology, cellular localization and biological process information. BMC Bioinformatics. 10:290. doi:10.1186/1471-2105-10-290.
[2]. Bateman A, Martin MJ, O'Donovan C, Magrane M, Apweiler R, Alpi E, Antunes R, Arganiska J, Bely B, Bingley M, et al. 2015. UniProt: A hub for protein information. Nucleic Acids Res. 43(D1):D204–D212. doi:10.1093/nar/gku989.
[3]. Bordbar S, Anwar F, Saari N. 2011. High-value components and bioactives from sea cucumbers for functional foods - A review. Mar Drugs. 9(10):1761–1805. doi:10.3390/md9101761.
[4]. Daina A, Michielin O, Zoete V. 2017. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 7(October 2016):1–13. doi:10.1038/srep42717. http://dx.doi.org/10.1038/srep42717.
[5]. Ekins S, Mestres J, Testa B. 2007. In silico pharmacology for drug discovery: Methods for virtual ligand screening and profiling. Br J Pharmacol. 152(1):9–20. doi:10.1038/sj.bjp.0707305.
[6]. Fadhlia IR. 2017. Karakterisasi Konsentrat dan Hidrolisat Protein Teripang Keling (Holothuria atra) dan Potensinya sebagai Imunomodulator. Institut Pertanian Bogor.
[7]. Fernando D. 2017. Clustering Interaksi Antara Protein Dan Proses Biologis Yang Terkait Diabetes Mellitus Tipe 2 Menggunakan Fuzzy K-Partite. Institut Pertanian Bogor.
[8]. Ghadermarzi S, Li X, Li M, Kurgan L. 2019. Sequence-Derived Markers of Drug Targets and Potentially Druggable Human Proteins. Front Genet. 10(November):1–18. doi:10.3389/fgene.2019.01075.
[9]. Hartati R, Zainuri M, Ambariyanto A, Widianingsih W, Mustagpirin, Panji F. 2021. Ranching of sea cucumber in Indonesia : A study case of Holothuria atra. IOP Conf Ser Earth Environ Sci. 777(1). doi:10.1088/1755-1315/777/1/012019.
[10]. Hartsperger ML, Blöchl F, Stümpflen V, Theis FJ. 2010. Structuring heterogeneous biological information using fuzzy clustering of k-partite graphs. BMC Bioinformatics. 11(1):522. doi:10.1186/1471-2105-11-522. http://www.biomedcentral.com/1471-2105/11/522.
[11]. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, et al. 2023. PubChem 2023 update. Nucleic Acids Res. 51(D1):D1373–D1380. doi:10.1093/nar/gkac956.
[12]. Lei X, Wang S, Wu FX. 2019. Identification of essential proteins based on improved HITS algorithm. Genes (Basel). 10(2):1070–1080. doi:10.3390/genes10020177.
[13]. Lipinski CA. 2004. Lead- and drug-like compounds: The rule-of-five revolution. Drug Discov Today Technol. 1(4):337–341. doi:10.1016/j.ddtec.2004.11.007.
[14]. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 64(SUPPL.):4–17. doi:10.1016/j.addr.2012.09.019.
[15]. Mahdavi MA, Lin YH. 2007. False positive reduction in protein-protein interaction predictions using gene ontology annotations. BMC Bioinformatics. 8:1–10. doi:10.1186/1471-2105-8-262.
[16]. Muhammad J, Khan A, Ali A, Fang L, Yanjing W, Xu Q, Wei D-Q. 2018. Network Pharmacology: Exploring the Resources and Methodologies. Curr Top Med Chem. 18(12):949–964. doi:10.2174/1568026618666180330141351.
[17]. Ode W, Agus R, Manarfa U, Kusuma WA, Sitanggang IS. 2022. Correlation of Diabetes Mellitus and Cellular Components using Fuzzy K-Partite. 2022(2011):56–64. doi:10.11594/nstp.2022.2509.
[18]. Pan A, Lahiri C, Rajendiran A, Shanmugham B. 2016. Computational analysis of protein interaction networks for infectious diseases. Brief Bioinform. 17(3):517–526. doi:10.1093/bib/bbv059.
[19]. Pangestuti R, Arifin Z. 2018. Medicinal and health benefit effects of functional sea cucumbers. J Tradit Complement Med. 8(3):341–351. doi:10.1016/j.jtcme.2017.06.007.
[20]. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. 2015. DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes. Database. 2015:1–17. doi:10.1093/database/bav028.
[21]. Ramadhani HF. 2021. Analisis Jejaring Interaksi Protein-Protein untuk Mengidentifikasi Protein Penting Terkait Hiperinflamasi pada COVID-19. IPB University.
[22]. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, et al. 2019. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 14(2):482–517. doi:10.1038/s41596-018-0103-9. http://dx.doi.org/10.1038/s41596-018-0103-9.

[23]. Salim Hanafi P, Sutrisno A, Murniasih T, Harijono, Yunovilsa Putra M, Untari F. 2019. Acute Oral Toxicity Assessment of the Ethanol Extract of Holothuria Atra in Mice. Asian J Pharm Clin Res. 13(2):150–153. doi:10.22159/ajpcr.2020.v13i2.36316.
[24]. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. (Karp 2001):2498–2504. doi:10.1101/gr.1239303.metabolite.
[25]. Soofi A, Taghizadeh M, Tabatabaei SM, Tavirani MR, Shakib H, Namaki S, Alighiarloo NS. 2020. Centrality analysis of protein-protein interaction networks and molecular docking prioritize potential drug-targets in type 1 diabetes. Iran J Pharm Res. 19(4):121–134. doi:10.22037/ijpr.2020.113342.14242.
[26]. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, et al. 2023. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 51(1 D):D638–D646. doi:10.1093/nar/gkac1000.
[27]. Umar AH, Ratnadewi D, Rafi M, Sulistyaningsih YC, Hamim H, Kusuma WA. 2022. Drug candidates and potential targets of Curculigo spp. compounds for treating diabetes mellitus based on network pharmacology, molecular docking and molecular dynamics simulation. J Biomol Struct Dyn. 0(0):1–17. doi:10.1080/07391102.2022.2135597. https://doi.org/10.1080/07391102.2022.2135597.
[28]. Wang M, Wang Q, Du Y, Zhang X. 2020. Network Pharmacology-Based Strategy to Investigate Pharmacological Mechanisms of Qiaoshao Formula for Treatment of Premature Ejaculation. Evidence-based Complement Altern Med. 2020:1–12. doi:10.1155/2020/1418634.
[29]. Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TO, Wang J, Xiao J, Zhang J, Bryant SH. 2009. An overview of the PubChem BioAssay resource. Nucleic Acids Res. 38(SUPPL.1):255–266. doi:10.1093/nar/gkp965.
[30]. Zhong J, Tang C, Peng W, Xie M, Sun Y, Tang Q, Xiao Q, Yang J. 2021. A novel essential protein identification method based on PPI networks and gene expression data. BMC Bioinformatics. 22(1):1–21. doi:10.1186/s12859-021-04175-8. https://doi.org/10.1186/s12859-021-04175-8.
[31]. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 10(1). doi:10.1038/s41467-019-09234-6. http://dx.doi.org/10.1038/s41467-019-09234-6.
[32]. Zhu J, Zhao Q, Katsevich E, Sabatti C. 2019. Exploratory Gene Ontology Analysis with Interactive Visualization. Sci Rep. 9(1):1–9. doi:10.1038/s41598-019-42178-x. http://dx.doi.org/10.1038/s41598-019-42178-x.
[33]. Zuhri UM, Purwaningsih EH, Fadilah F, Yuliana ND. 2022. Network pharmacology integrated molecular dynamics reveals the bioactive compounds and potential targets of Tinospora crispa Linn. as insulin sensitizer. PLoS One. 17(6 Jun):1–15. doi:10.1371/journal.pone.0251837. http://dx.doi.org/10.1371/journal.pone.0251837.

Authors Profile

Muhammad Naufal Abdillah, is a Computer Science Master Student in the Department of Computer Science at IPB University, Indonesia. He has received his Bachelor degree in Computer Science from IPB University, Indonesia in 2022. This paper is a part of his master's research.

Wisnu Ananta Kusuma, is an Associate Professor at the Department of Computer Science, and a researcher at Tropical Biopharmaca Research Center, IPB University, Indonesia. He is also an executive committee of the Asia Pacific Bioinformatics Network and president of the Indonesian Society of Bioinformatics and Biodiversity. His research interests include bioinformatics, machine learning, network pharmacology, and high-performance computing.

Mushthofa, is a lecturer at the Department of Computer Science, IPB University, Indonesia. In 2009, he completed his Master's in Computational Logic at the Vienna University of Technology, Austria. In 2018, he received his Ph.D. in Computer Science and Bioinformatics at Ghent University, Belgium. His research interests include artificial intelligence, machine learning, data science, knowledge representation and reasoning, and bioinformatics.

Mala Nurilmala, is a Professor at the Department of Aquatic Product Technology, Faculty of Fisheries and Marine Science, IPB University Indonesia and a researcher at Halal Science Center, IPB University, Indonesia. She is also a General Secretary of Indonesian fishery product processing community association. The research interests include fisheries processing as well as marine natural products.