Screening and Prediction of Potential Compounds from Virgin Olive Oil Acting on Proteins Associated with Cancer Disease

Achmad Rodiansyah1,2*
1 Department of Biology, Universitas Negeri Malang, Indonesia
2 Department of Biotechnology, School of Life Sciences and Technology, Institut Teknologi Bandung, Indonesia

ABSTRACT

Virgin olive oil contains phenolic compounds that were potential for anti-inflammatory and cancer treatment. Computational biology is a beneficial method to understand how this compound can affect the biological process in humans. This research is conducted by the potential screening of VOO compounds, constructing the pharmacological network and enrichment, and docking simulation. The enrichment result showed that the EGFR, BRAF, MAPK1, CCND1, and MDM2 protein have multiple cancer contributions and related pathways. The docking simulation result showed that the interaction of EGFR-luteolin, BRAF-luteolin, MAPK1-luteolin, CCND1-apigenin, and MDM2-1-hydroxyxypinoresinol has the highest binding affinity. Further research with the in-vitro methods are required to check the reliable mechanisms of each compound to their protein target.

Keywords: Bioinformatics; Cancer; Molecular docking; Network pharmacology; Olive oils; Phenolics

INTRODUCTION

The virgin olive oil (VOO) is the main product of olive trees (Olea europaea subs. Eureupaea). This tree is a symbolic species in the Mediterranean region. Genomic research with this tree is important for facilitating study in metabolisms, developmental and physiological process, which is the research could contribute to improving the economic values of this tree (Cruz et al., 2016). Extra VOO contains dominant phenolic compounds (Visioli and Bernardini, 2011), which those compounds have different classes, such as phenyl ethyl alcohols, steroids, phenolic acids, hydroxy-isochromans, lignans and, flavonoids. The phenolics (also known as polyphenols) from a plant have some benefits in the human body to prevent various diseases. They work to bind several proteins, which may lead to a specific disease, and they have also antioxidants activity against free-radical (Preedy and Watson, 2010). In the in-vitro studies, they can modulate intracellular signaling pathways; so, VOO compounds, like hydroxytyrosol, tyrosol, and other minor compounds, have been the focus of research to see its effect in the biological process (Serrelli and Deiana, 2018).

Recent studies in-vitro and in-vivo with VOO compounds have been done. The recent studies, hydroxytyrosol from VOO has been reported could induce cell cycle arrest and apoptosis on various cells, such as colon cancer cells, bladder cancer and cholangiocarcinoma (Coccia et al., 2016; Li et al., 2014; Lópe d de las Hazas et al., 2017). Pinoresinol may also probably have antitumor activity in breast cancer cells (López-Biedma et al., 2016), and p-HPE-EDA can inhibit colon cancer cells with inhibition of AMP-activation protein kinase (AMPK) and cyclooxygenase-2 (COX-2) (Khanal et al., 2011). In silico and in vitro study showed that oleuropein confirmed could inhibit mTOR, which is responsible for tumors’ properties on breast cancer (Corominas-Faja et al., 2018).

World health organization (WHO) mentioned that cancer which causes the death of about 9.6 million death in 2018, is a group of diseases with abnormal cell growth that can invade other organs. The mechanisms in this disease are very complicated. Researchers mention that this disease is mainly caused by a gene mutation affecting cell functions, carcinogenic chemical compounds, and an unhealthy lifestyle (Hassanpour and Dehghani, 2017). Genetic materials and proteins play an essential role in this group’s diseases; with bioinformatics, it will more easily explore potential clinical applications and improve diagnosis, therapies, and cancer diseases prognosis (Wu et al., 2012). Computational network biology is a new research field that involves theory and applications to describe a molecule's interaction on living cells; this field also contributes to accelerating molecular biology, pharmacology, and genetics studies (Ni et al., 2018). This method is acceptable for integrating
and analyzing big data; with this network model, the extensive information of the biological system could be more easily to be understanding (Ideker and Nussinov, 2017; Ma’ayan, 2011).

**METHODOLOGY**

**Preparation of ligand and protein target**

VOO bioactive compounds were collected from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) based on relevant literature (Alavi and Golmakani, 2017; Basiricò et al., 2019; Rupani, 2014; Waterman, 2007; Zalejska-Fiolka et al., 2015). Finally, 16 bioactive compounds used in this study were listed in Table I. The 3D structure and canonical SMILES of all compounds were downloaded from the PubChem database. Canonical SMILES compounds were used for identifying the potential target proteins on the Phrammapper web server (http://lilab-ecust.cn/pharmmapper/) and Swiss Target Prediction web server (http://www.swisstargetprediction.ch/). Phrammapper web server is a platform for identifying the potential protein target from bioactive compounds with statistical method calculation include more than 7000 target pharmacophores (X. Liu et al, 2010; X. Wang et al., 2017). The specificiation for searching target proteins was set as Druggable Pharmacophore Models (v2017.16159) (Liang et al., 2019). Like a Phrammapper webserver, SwissTargetPrediction is an accruable webserver to predict the target proteins of bioactive compounds (Gfeller et al., 2014); this web server has been updated for efficient prediction of protein targets (Daina et al., 2019). The result from Phrammapper and SwissTargetPrediction webserver was saved as .csv format and used to construct network compound-protein interactions.

**Network construction and analysis**

The network compound-protein interaction was constructed by Cytoscape software and its plugins (https://cytoscape.org/). Excel files from Pharmmapper and SwissTargetPrediction that contains compound-protein interaction data were imported to Cytoscape software with the menu “import network form file system”. The compound table was selected as a “source node”, and the protein target was selected as a “target node”. The following analysis was used ClueGO Cytoscape’s plugins to show the interpretation of list genes on the metabolical process (Bindea et al., 2009). The ClueGO plugin’s setting performs as pathways with p-values ≤ 0.001, and the kappa score was set as 0.4. The Gene Ontology databases were used from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Wikipathways. The selected target proteins from the ClueGO enrichment were used for docking simulation for determining ligand-protein interaction.

**Protein Preparation and Molecular docking simulation**

The ClueGO results relating to cancer diseases were selected as a target protein for molecular docking. The target protein used for molecular docking were retrieved from Protein Data Bank (https://www.rcsb.org/). The 3D structure of proteins used in this study was listed as follows: EGFR-PDB.id 6LUD (Kashima et al., 2020), BRAF-PDB.id 6NSQ (Assadieskandar et al.,

---

### Table I. Bioactive compounds in virgin olive oil

<table>
<thead>
<tr>
<th>No</th>
<th>PubChem ID</th>
<th>Compound Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5280934</td>
<td>alpha-Linolenic acid</td>
</tr>
<tr>
<td>2</td>
<td>5280450</td>
<td>Linoleic acid</td>
</tr>
<tr>
<td>3</td>
<td>445639</td>
<td>Oleic acid</td>
</tr>
<tr>
<td>4</td>
<td>985</td>
<td>Palmitic acid</td>
</tr>
<tr>
<td>5</td>
<td>5281</td>
<td>Stearic acid</td>
</tr>
<tr>
<td>6</td>
<td>10393</td>
<td>Tyrosol (p-HPEA)</td>
</tr>
<tr>
<td>7</td>
<td>82755</td>
<td>Hydroxytyrosol (3,4-DHPEA)</td>
</tr>
<tr>
<td>8</td>
<td>131750845</td>
<td>1-(3-Methoxy-4-hydroxy)-phenyl-6,7-dihydroxy-isochroman</td>
</tr>
<tr>
<td>9</td>
<td>442831</td>
<td>1-Acetoxyphinaresinol</td>
</tr>
<tr>
<td>10</td>
<td>13824420</td>
<td>1-Hydroxyphinaresinol</td>
</tr>
<tr>
<td>11</td>
<td>131750844</td>
<td>1-Phenyl-6,7-dihydroxy-isochroman</td>
</tr>
<tr>
<td>12</td>
<td>124202093</td>
<td>3,4-DHPEA-EA</td>
</tr>
<tr>
<td>13</td>
<td>5280443</td>
<td>Apigenin</td>
</tr>
<tr>
<td>14</td>
<td>5280445</td>
<td>Luteolin</td>
</tr>
<tr>
<td>15</td>
<td>16681728</td>
<td>p-Hpea-edu</td>
</tr>
<tr>
<td>16</td>
<td>234817</td>
<td>Pinoresinol</td>
</tr>
</tbody>
</table>
Screening and Prediction of Potential Compounds from Virgin Olive Oil

Traditional Medicine Journal, 26(1), 2021

2019), MDM2-PDB.id 3LNJ (M. Liu et al., 2010), MAPK/ERK-PDB.id 5NHH (Ward et al., 2017), and CCND1 6P6G (Guiley et al., 2019). Those target proteins were sterilized from waters and ligands using PyMOL software (available in https://pymol.org/2/). The ligands used for molecular docking simulation were converted to .pdbqt format using open babel on PyRx software (available in https://pyrx.sourceforge.io/), and the reverse docking simulation was performed using Vina Wizard in PyRx (Dallakyan and Olson, 2015; Trott and Olson, 2010). The docking results were visualized and evaluated using PyMOL software. The interaction between amino acid residues and ligands in 2D view were visualized using BIOVIA DiscoveryStudio 2019 software (available in https://www.3dsbiovia.com/).

RESULT AND DISCUSSION
Preparation of ligand and network data construction
Sixteen potential ligands from VOO were collected from PubChem, and the list of protein targets from Phrammapper and SwissTargetPrediction were downloaded. Each compound interacted with various proteins that may contribute to many functions on biological pathways to build the compound–proteins network. Network construction with Cytoscape software showed 1639 nodes and 4016 edges. A node represents the target proteins and VOO compounds, while an edge represents the interaction of compounds–target proteins on the biological pathways (Figure 1). The VOO compounds labeled with red nodes in Figure 1 interact with blue nodes labeled as target proteins. Luteolin, 1-Hydroxypinorisenol, apigenin, linoleic acid, pinoresinol, oleic acid, stearic acid, 1-Acetoxypinorisenol, palmitic acid, p-Hpea-edu, and alpha-linolenic acid are located on central of nodes target protein, and other compound’s nodes are located on outside of central interaction. Table panel protein target interactions from the Cytoscape network were saved as an Excel file. All protein was filtering and grouping to determine their biological pathways using ClueGO Cytoscape’s plugin.

Network construction and analysis
ClueGO analysis with KEGG and gene ontologies database showed 26 nodes and 38 edges with three pathways (Figure 2a), and WikiPathways database 35 nodes and 59 edges with four pathways (Figure 2b). ClueGO’s plugin pathways were listed in Table 2; this result showed that VOO compounds are associated with various pathways, especially in tumor and cancer. The VOO compounds interact with various proteins that work on several biological pathways and are interconnected with its pathway or other pathways. From the KEGG database and WikiPathways database enrichment, it has known that VOO compounds’ target protein was probably related to cancer disease and tumor with a p-value ≤ of 0.001. A significant test of the p-value is vital for decision-making. A very small p-value indicated that the hypothesis is probably correct (Panagiotakos, 2008). These enrichment pathways indicated that they have a high confidence value.

All pathways from ClueGO are related to each other indicated that the several nodes of the gene connect with all pathways in Figure 2 (marked with a red circle). Glioma is the one type of tumours disease that occurs in the Central
Nervous System (CNS) and spinal cord; this disease is a common type of primary malignant brain tumor derived from glial cells (Ernest and Sontheimer, 2009). Glioblastoma or glioblastoma multiforme (GBM) is a subclass of glioma disease (JOVČEVSKA et al., 2013). Oncostatin M (OSM) is an important pathway associated with the biological process and cellular responses. This pathway is vital for clinical and biomedical therapeutic on the human disease (Dey et al., 2013). OSM pathway has been reported associating with cancer cell’s plasticity (Junk et al., 2017); also contribute to breast tumor specifically mediated by OSMRβ (Underhill-Day and Heath, 2006) and prostate cancer (Godoy-Tundidor et al., 2005). Prostate cancer and bladder cancer are double cancer primary cancer with high frequency reported; this data suggest that the patients diagnosed with bladder or prostate cancer should be followed second malignant urologic diagnosis.
(Kinoshita et al., 2004). Disturbing the OSM pathway may be potential for cancer cell treatment (Caffarel and Coleman, 2014; Stroeder et al., 2018).

The other pathway, Matrix metalloproteinases (MMPs), is a group of enzymes responsible for the degradation of extracellular matrix protein during organogenesis and normal tissue replacement. This group of enzymes is also associated with oral cancer (Sorsa et al., 2004). Based on ClueGO enrichment, it is shown that this pathway has a connection with bladder cancer and OSM. The interruption of this protein activity could lead to various diseases (Laronha and Caldeira, 2020).

From KEGG and WikiPathways database enrichment, five genes contribute to OCM, glioma, bladder, and prostate cancer pathways. Those genes are CCND1 encodes G1/S-specific cyclin-D1 protein, MAPK1 encodes dual specificity mitogen-activated protein kinase 1, MDM2 encodes E3 ubiquitin-protein ligase Mdm2, EGFR encodes epidermal growth factor receptor protein, BRAF encodes serine/threonine-protein kinase B-RAF. These proteins were chosen for molecular docking simulation docked with the compound from VOO as the ligand to know their interaction. They could be evaluated to be used as therapeutic compounds for treating cancer disease.

The CCND1 (Cyclin D1) gene has a function in the regulation of CDK kinase in the cell cycle; when this gene had the mutation and overexpressed, it can promote a various type of cancers disease in humans, such as breast cancer, endometrial cancer, colon cancer, and prostate cancer (Fu et al., 2004; Ikeda et al., 2013; Moreno-Bueno et al., 2003; Xu and Lin, 2018). This gene is also used as a biomarker in breast cancer (Lundberg et al., 2019). Cyclin D1 protein functions as cyclin-dependent-kinase (CDK) in a subunit of CDK4 or CDK6 to regulate the cell cycle from G1 to S phase transition. They act as apoptosis regulators interacting with tumor suppressor protein retinoblastoma (Rb) to interrupting the cell cycle. CCND1 also acts on chromatin recruitment, mitochondrial biogenesis, and DNA Damage Response (DDR) (Fu et al., 2004; Massagué, 2004; Pestell, 2013). When this gene act as an oncogene caused by point mutation, local DNA rearrangements, or chromosomal translocation, then they could overexpress in a cell, and that cell will have rich of complex CDK-cyclin; so, it could stimulate the progression in cell cycles, stimulate tumorigenesis, and metastases even in the absence of growth factor (Fu et al., 2004; Hardin et al., 2012; Kim and Diehl, 2009).

BRAF encodes Raf protein and MAPK1/2 gene-encoded mitogen-activated protein kinase 1/2 or Extracellular Signal-Regulated Kinases (ERKs) protein which play on signal transduction in the MAPK signaling pathway. This protein is known as signal transduction which works on the Ras-Raf-MEK-ERK signaling pathway (Ursem et al., 2018). Activating this pathway starts with growth factor family protein like Epidermal Growth Factor (EGF) binds their receptor to phosphorylate Raf protein as downstream of Ras on the MAPK signaling pathway (Ursem et al., 2018). Altered of this signaling pathway or mutation in genes encoding the protein in this pathways have been reported correlated and detected in tumors and cancer disease, including in pituitary tumorigenesis, cervical cancer tissue, and significantly correlated on breast cancer with axillary lymph node metastasis (Suojun et al., 2012; Manousaridis et al., 2013 Jan 1; Li et al., 2015: 1; AACR Project GENIE 2017; Shao et al., 2018). The inhibitors with targeting on these kinases protein could treat malignant tumors; this inhibitor would be promising and challenging in future research (Liu et al., 2018; Suojun et al., 2012). National Comprehensive Cancer Network (NCCN) also recommending BRAF testing gene for diagnosis of colorectal cancer (CRC) (Ursem et al., 2018).

**Molecular Docking Simulation**

Computational methods recently used in the biotechnology and pharmaceutical industries were...
beneficial for drug discovery with high success and accuracy and reduced costs (Parenti and Rastelli, 2012; Suortti, 1997). Reverse docking is a promising drug prediction technique that acts on protein-related disease as inhibitors (Kharkar et al., 2014); this method has excellent drug design and drug discovery success. The docking simulation result showed that VOO compounds have various docking scores to target proteins (Figure 3).

The bar chart in Figure 3 shows the docking simulation score from VOO compounds with their target proteins. The deep-blue chart is an EGFR protein; an orange chart is a BRAF protein; a yellow chart is a MAPK protein; a grey chart is an MDM2 protein; and a blue chart is a CCND1 protein. Luteolin from VOO has the highest docking score, -7.9 kcal/mol on EGFR protein and -9.5 kcal/mol on BRAF protein. 1-Hydroxypinoresinol has the highest docking score with -7.0 kcal/mol on MDM2 protein. Apigenin has the highest docking score with -6.7 kcal/mol on CCND1 protein; also, on MAPK protein, luteolin has the highest docking score of -9.0 kcal/mol, respectively (Figure 3). Compared to the control compound, only the docking score from Luteolin-BRAF and luteolin-MAPK were more stable. The docking score from control compounds was listed in Table III.
Binding affinity is a critical aspect of drug design to produce potential ligands with high binding affinity to the target protein and low target binding affinity to non-targeted protein (Kairys et al., 2019). The Gibbs energy of binding ($\Delta G_b$) or binding affinity is value to define a strong interaction between two molecules, so this is a crucial quantity for study molecule interaction (Vangone et al., 2018). The type of bond on reverse docking is critical on binding affinity as a docking score, especially the hydrogen bond in ligand-protein interaction. The 2D visual interaction of the VOO ligands colored with green within their target protein can be seen in Figure 4, and the list of amino acid residues can be seen in Table IV.

The results from molecular docking showed that the most of the ligands occupied the vital region of the protein. This interaction could potentially disrupt the work of target protein, like preventing phosphorylation and inhibiting protein activation. A ligands’ effectiveness to interfere with proteins’ action can be predicted by the docking score from the inhibitor ligand to the protein. Many factors that contribute to binding affinity score as follow: the role of water, the existence of H-bonds, the different types of the bind of ligand-protein interaction include ionic interactions, Van der Waals and hydrogen bond interaction, hydrophobic interaction, Pi-alkyl bond, and Pi-sulfur bond; those factors must be considered to evaluate docking results (Arthur and Uzairu, 2019; Pantsar and Poso, 2018). However, not all parameters can be calculated on this docking simulation, so an in-vitro and in-vivo study must be carried to determine the reliable responses (Pintilie and Stefaniu, 2019).

The illustration of VOO compounds to inhibit the target protein can be seen in Figure 5. From this illustration, the VOO compounds probably can work on several fields on the cells, like on the extracellular-membrane layer, cytoplasm, and nucleus cells field. EGFR protein is located on membrane cell, MAPK and BRAF protein work on the cytoplasm, MDM2 and CCND1 protein work on the nucleus.

**CONCLUSION**

The VOO compounds have multiple protein targets on various pathways, especially in cancers and tumors. Five proteins that act on multiple
Figure 5. The illustration of VOO compounds work on their target proteins in the cell (The illustration was created with BioRender. (Available at: https://biorender.com/)

Table IV. Amino acid residues from molecular docking simulation with the highest score and the information domain from UniProt

<table>
<thead>
<tr>
<th>No</th>
<th>Protein</th>
<th>VOO ligand</th>
<th>Amino acid residues</th>
<th>Domain UniProt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EGFR</td>
<td>luteolin</td>
<td>LEU718, GLY719, VAL726, ALA743, LYS 745 GLU762, LEU792, MET793, PRO794, GLY796, MET798, LEU844, THR854, ASP855</td>
<td>Uniprot Id: P00533 Region: 712-972 (Protein kinase domain)</td>
</tr>
<tr>
<td>2</td>
<td>BRAF</td>
<td>luteolin</td>
<td>ILE463, VAL471, ALA481, VAL482, LYS483, GLU501, VAL504, LEU505, THR508, ILE513, LEU514, ILE527, THR529, GLN530, TRP531, CYS532, LEU567, PHE583, GLY593, ASP594, PHE595</td>
<td>Uniprot Id: P15056 Region: 457-717 (Protein kinase domain)</td>
</tr>
<tr>
<td>3</td>
<td>MDM2</td>
<td>1-Hydroxypinoresinol</td>
<td>LEU54, LEU57, GLY58, ILE61, MET62, TYR67, GLN72, HIS73, VAL93, VAL75, H96, ILE99, TYR100</td>
<td>Uniprot Id: Q00987 Region: 1-110 (USP interaction domain)</td>
</tr>
<tr>
<td>4</td>
<td>MAPK1</td>
<td>Luteolin</td>
<td>ILE31, GLY32, GLU33, GLY34, VAL39, ALA52, LYS54, ILE84, GLN105, ASP106, LEU107, MET108, GLU109, THR110, LYS114, LEU156, CYS166, ASP167</td>
<td>Uniprot Id: P28482 Region: 25-313 (Protein kinase domain) Region: 105-108 (Inhibitor binding domain)</td>
</tr>
<tr>
<td>5</td>
<td>CCND1</td>
<td>Apigenin</td>
<td>ASN174, ILE178, ILE177, HIS181, VAL212, GLY214, LEU217, ARG218, PRO220, ASN222</td>
<td>Uniprot Id: P24385 Region: 2-208 (Interaction region) Region: 2-19 (Region for RPLP0 &amp; TCF3) Region: 150-360 (Region for TCF3)</td>
</tr>
</tbody>
</table>
central pathways from KEGG and WikiPathways database were MAPK1, BRAF, EGFR, MDM2, and CCND1. The docking simulation showed that the luteolin compound was stable with EGFR protein and BRAF protein, with the docking score reaching about -7.9 kcal/mol and -9.5 kcal/mol, respectively; 1-hydroxypinoresinol was stable to interact with MDM2 protein, reaching a score of -7.0 kcal/mol; apigenin and stearic acid were stable to interact with CCND1 protein, which has a docking score of -9.0 kcal/mol. Those compounds, especially luteolin, probably have the potential for therapeutic on various cancers and tumors. These docking simulation results also report that luteolin has a stronger binding affinity than the control compound for interfering with the BRAF and MAPK protein. The in-vitro and in-vivo study must be carried out to validate the specific response from ligands on proteins involved in cancer pathways.

ACKNOWLEDGEMENT
The author is grateful to his previous affiliation, "Universitas Negeri Malang, Indonesia", to facilitate internet connection and give a permit to access articles and software used in this study.

REFERENCES
Achmad Rodiansyah


Screening and Prediction of Potential Compounds from Virgin Olive Oil


