QSAR, Molecular Docking, and Molecular Dynamic of Novel Coumarin Derivatives as α-Glucosidase Inhibitor

Mutista Hafshah 1*, Irvan Maulana Firdaus 2, Ika Nur Fitriani 3, Lutfiah Rahmawati 1

1 Department of Chemistry, Faculty of Science and Technology, Universitas Islam Negeri Walisongo, Semarang, 50181, Indonesia
2 Al Irsyad Al Islamiyyah Boarding School Purwokerto, Banyumas 53113, Indonesia
3 School of Chemistry, Faculty of Science, The University of Sydney, Camperdown NSW 2050, Australia

* Corresponding author: mutista.hafshah@walisongo.ac.id

https://doi.org/10.14710/jksa.27.7.316

Abstract

Diabetes mellitus (DM) is a chronic metabolic disorder posing a significant health risk. Effective treatments are continually sought. Coumarin derivatives with oxime ester groups have shown potential as antidiabetic agents by inhibiting the α-glucosidase enzyme, a key player in glycoprotein metabolism and postprandial hyperglycemia control. This makes lysosomal α-glucosidase a promising therapeutic target. A study used 28 coumarin derivatives with known α-glucosidase inhibitory IC₅₀ values for computer-assisted drug design (CADD) through quantitative structure-activity relationship (QSAR) analysis, yielding a statistically robust equation for guiding new compound development. Subsequently, eleven new coumarin derivatives with oxime ester groups were synthesized, showing enhanced α-glucosidase inhibitory activity compared to the initial set. Molecular docking assays indicated that compounds 32, 37, 38, and 39 had lower free energy values than the native ligand, suggesting higher stability in target protein interactions. Notably, compound 38 had the lowest free energy (~8.351) and demonstrated lower root mean square deviation (RMSD) and root mean square fluctuation (RMSF) than the original ligand, indicating greater stability. The QSAR equation derived is:

\[
\log IC_{50} = 2.886 - 0.054 (LUMO) + 0.073 (\mu) - 0.148 (\alpha) - 0.046 (RD) + 0.046 (BM) + 0.001 (V_{cav}) - 0.421 (qC2) + 1.138 (qC8) - 0.092 (qC9) + 2.61 (qC10) + 1.354 (qN1)
\]

(Eq 1) n=28; R=0.918; R²=0.843; SD=0.196; F calc/F tab=3.169; Sig <0.01; PRESS =1.376.

Compound 38's SMILES notation is:

\[
C(C=N/OC(=O))C=CCl=C(CBr)C=CCl=CCl=C(CCC(=O)=O)C-C(=O)C\]

1. Introduction

Diabetes mellitus (DM) causes hyperglycemia due to hormonal abnormalities influencing insulin production [1, 2, 3]. The pancreas regulates blood glucose using insulin [4]. DM has two types: Type I, which kills pancreatic beta cells and requires insulin injections, and Type II, which reduces insulin release and raises blood glucose [2, 4]. High hyperglycemia can induce kidney failure, heart attack, and eyesight loss. The International Diabetes Federation (IDF) estimated 9.3% of 483 million persons aged 20–79 had DM in 2019. This number will climb to 578 million by 2030 and 700 million by 2045. Indonesia has 10.7 million diabetics, placing it among the top 10 countries with the greatest frequency. Better prevention and treatment strategies are needed.

Treatment varies by diabetes type. Insulin treats Type I diabetes, while blood glucose-lowering medicines treat Type II. α-Glucosidase inhibitors that convert carbs into glucose treat Type II diabetes. Coumarin derivatives are promising natural and synthetic α-glucosidase inhibitors [5, 6, 7, 8]. Coumarins, found in Umbelliferae/Apiceae and Rutaceae [9]. The molecular structure of a chemical has an impact on its biological action [10, 11].
Molecular dynamics simulations can monitor protein and ligand binding stability. RMSD (≤ 3 Å), RMSF values are essential for molecular docking validation and protein–ligand binding stability. Molecular dynamics simulations can monitor protein and ligand binding conformational changes alongside molecular docking [29, 30, 31]. This unique approach uses molecular docking and molecular dynamics analysis to improve in silico analysis data and identify α-glucosidase inhibitors from coumarin derivative compounds with oxime ester groups.

2. Experimental

2.1. Tools and Materials

The research utilized an Intel Core i5 computer processor, 8 GB of Random Access Memory (RAM), and the Windows 10 operating system. The software tools employed were Hyperchem version 8.0, SPSS version 25, JASP, PLANTS (Protein–Ligand ANTSystem), Autodock Vina, and SketchChem. The study included 28 coumarin compounds (Figure 1), and the IC₅₀ values were derived from Zhang et al. [14]. The compounds used are categorized as 21 fitting compounds and 7 test compounds (Table 1). The selection of 7 test compounds is based on the steric representation of 28 coumarin derivative compounds with ester oxime groups.

A selection of representatives with varying degrees of steric hindrance, including mild, medium, and substantial hindrances, were chosen as test sets [32]. Fitting compounds are used to determine the equation, whereas test compounds are used to validate the resulting equation. The equation is validated by calculating the predicted residual error sum of squares (PRESS) value from the experimental log IC₅₀ data and the predicted log IC₅₀ data of 7 test compounds.

2.2. Geometric Optimization

HyperChem 8.0.8 was used to create a two-dimensional model of the molecular structure of molecules that are derivatives of coumarins. Then, hydrogen atoms were added to each of the structure's constituent atoms. Geometric principles were applied to construct each structure to minimize molecular energy and produce the most stable conformation possible. The convergence limit was established at 0.001 kcal/mol using the Polak-Ribiere algorithm, and the optimization calculations were performed using the PM3 semiempirical technique. The atomic charge (qC₁, qC₂, qO₁, qO₂, and qN₁), dipole moment (µ), log P, polarizability (α), molecular weight, van der Waals area (AᵥdW), van der Waals volume (VᵥdW), refractive index (RD), and the HOMO and LUMO energies were among the descriptors used. Descriptors were included in Table 2, along with instructions on how to retrieve them [24].

2.3. Determining Descriptors and Correlation Analysis

The analysis involved a bivariate correlation utilizing the two-tailed technique and Pearson’s coefficient. At this point, each descriptor was carefully examined to evaluate its degree of association with log IC₅₀ activity. Next, the atomic charge (qC₁–qC₂, qO₁–qO₂, and qN₁), dipole moment (µ), log P, polarizability (α), molecular weight, van der Waals Area (AᵥdW), van der Waals volume (VᵥdW), refractive index (RD), and the HOMO and LUMO energies were selected. A strong relationship indicates a significant impact on log IC₅₀ activity. Afterward, the chosen descriptors were utilized as independent variables in Multiple Linear Regression (MLR) studies [33].
Table 1. Coumarin compound derivatives

<table>
<thead>
<tr>
<th>Compound number</th>
<th>R¹</th>
<th>R²</th>
<th>R³</th>
<th>IC₅₀</th>
<th>Log IC₅₀</th>
<th>Compound number</th>
<th>R¹</th>
<th>R²</th>
<th>R³</th>
<th>IC₅₀</th>
<th>Log IC₅₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H</td>
<td>H</td>
<td></td>
<td>93.230</td>
<td>1.969</td>
<td>16</td>
<td>H</td>
<td>H</td>
<td></td>
<td>8.710</td>
<td>0.940</td>
</tr>
<tr>
<td>2</td>
<td>H</td>
<td>H</td>
<td></td>
<td>15.910</td>
<td>1.201</td>
<td>17</td>
<td>H</td>
<td>H</td>
<td></td>
<td>19.66</td>
<td>1.293584</td>
</tr>
<tr>
<td>3</td>
<td>H</td>
<td>H</td>
<td></td>
<td>10.950</td>
<td>1.039</td>
<td>18</td>
<td>H</td>
<td>H</td>
<td></td>
<td>21.540</td>
<td>1.333</td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td>H</td>
<td></td>
<td>6.460</td>
<td>0.810</td>
<td>19</td>
<td>H</td>
<td>Cl</td>
<td></td>
<td>3.420</td>
<td>0.534</td>
</tr>
<tr>
<td>5</td>
<td>H</td>
<td>H</td>
<td></td>
<td>29.210</td>
<td>1.466</td>
<td>20</td>
<td>H</td>
<td>Br</td>
<td></td>
<td>2.540</td>
<td>0.405</td>
</tr>
<tr>
<td>6</td>
<td>H</td>
<td>H</td>
<td></td>
<td>19.340</td>
<td>1.286</td>
<td>21</td>
<td>H</td>
<td>F</td>
<td></td>
<td>3.920</td>
<td>0.593</td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>H</td>
<td></td>
<td>26.120</td>
<td>1.417</td>
<td>22</td>
<td>H</td>
<td>CH₃</td>
<td></td>
<td>5.240</td>
<td>0.719</td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>H</td>
<td></td>
<td>26.160</td>
<td>1.418</td>
<td>23</td>
<td>H</td>
<td>OCH₃</td>
<td></td>
<td>6.430</td>
<td>0.808</td>
</tr>
<tr>
<td>9</td>
<td>H</td>
<td>H</td>
<td></td>
<td>26.140</td>
<td>1.417</td>
<td>24</td>
<td>Cl</td>
<td>H</td>
<td></td>
<td>4.250</td>
<td>0.628</td>
</tr>
<tr>
<td>10</td>
<td>H</td>
<td>H</td>
<td></td>
<td>34.350</td>
<td>1.536</td>
<td>25</td>
<td>Br</td>
<td>H</td>
<td></td>
<td>3.810</td>
<td>0.581</td>
</tr>
<tr>
<td>11</td>
<td>H</td>
<td>H</td>
<td></td>
<td>13.390</td>
<td>1.127</td>
<td>26</td>
<td>F</td>
<td>H</td>
<td></td>
<td>3.730</td>
<td>0.572</td>
</tr>
<tr>
<td>12</td>
<td>H</td>
<td>H</td>
<td></td>
<td>6.340</td>
<td>0.802</td>
<td>27</td>
<td>CH₃</td>
<td>H</td>
<td></td>
<td>6.170</td>
<td>0.790</td>
</tr>
<tr>
<td>13</td>
<td>H</td>
<td>H</td>
<td></td>
<td>5.060</td>
<td>0.704</td>
<td>28</td>
<td>OCH₃</td>
<td>H</td>
<td></td>
<td>7.840</td>
<td>0.894</td>
</tr>
<tr>
<td>14</td>
<td>H</td>
<td>H</td>
<td></td>
<td>11.550</td>
<td>1.063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>H</td>
<td>H</td>
<td></td>
<td>15.090</td>
<td>1.179</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.4. Determining the QSAR Equation

The most efficient QSAR equation was determined using multilinear regression statistical analysis. The JASP application was utilized to analyze 21 compounds. These compounds were determined to be appropriate for the dependent variable log IC₅₀, while the independent variable was produced via the correlation test [34].

2.5. Development of Novel Coumarin Derivatives

The satisfactory completion of the molecular design was achieved by modifying the substituents at positions R₁, R₂, and R₃. Afterward, the molecules that have been generated undergo geometry optimization and descriptor computation calculations. We employed the most precise QSAR equation to calculate the theoretical activity of the chemical, represented by the notation log IC₅₀. The compounds exhibiting the lowest log IC₅₀ values had the highest levels of anti-diabetic efficacy [35].
2.6. Molecular Docking

Molecular docking involves several steps: protein and ligand preparation, addition of hydrogen atoms and charges to proteins and ligands, determination of grid box, and actual molecular docking [36]. The protein and ligand preparation process commences with The protein code utilized in this investigation, 2OLE, obtained from the website https://www.rcsb.org/structure/2OLE and stored in PDB format. Dehydration was performed to obtain a protein molecule, and the initial ligand was extracted, resulting in the preservation of a protein molecule. This molecule was then recorded under the file name 2OLE_protein.pdb. Subsequently, a preparatory procedure was conducted to isolate the native ligand molecule exclusively. Next, the file (2OLE_ligand.pdb) in the current director was stored. Hydrogen atoms and charges were introduced to the proteins and ligands. Specifically, the protein molecule (2OLE_protein.pdb) and the ligand molecule (2OLE_ligand.pdb) were modified by adding hydrogen atoms.

The protein and ligand were assigned Gasteiger charges and saved in pdbqt format, leading to the generation of a protein file named 2OLE_protein.pdbqt and a ligand file named 2OLE_ligand.pdbqt. Grid box determination The AutodockTools software was used to determine the dimensions of the grid box by changing the spacing to 1 Å. The measurements of the grid box were stored under the file name “Grid.txt”. The docking procedure was executed in the current directory, including the 2OLE_protein.pdbqt, 2OLE_ligand.pdbqt, and Grid.txt files. Generate a Conf.txt file that includes the specified details: receptor = 2OLE_protein.pdbqt, ligand = 2OLE_ligand.pdbqt, center x,y,z based on the grid box dimensions, size_x, y, z based on the grid box dimensions, num_modes = 10, and exhaustiveness = 32. Launch Windows Power Shell in the current directory and input the command vina.exe --config conf.txt --cpu 4. Save the docking results in a text format and then separate the ligand from the docking results using the command vina_split.exe –2OLE_ligand_out.pdbqt. Choose a ligand with the highest stability level in terms of affinity value.

2.7. Molecular Dynamics

A molecular dynamics simulation was performed on the compound, which was predicted to be the best in each activity. The process involves preparing the ions.mdp, em.mdp, nvt.mdp, npt.mdp, and md.mdp files. Then, the preparation and merging of receptor and ligand topologies, addition of ions, energy minimization, temperature and pressure equilibration, production simulation, analysis, and visualization of results (RMSD, RMSF) are performed [36].

3. Results and Discussion

This study examines the quantitative correlation between coumarin derivative architectures and their antidiabetic effectiveness. Using Hyperchem 8.0.8, 28 coumarin oxime ester compounds were geometrically optimized to find the most stable molecular structure by minimizing potential energy. The semi–empirical PM3 method was chosen for this optimization, as it uses experimental data parameterization and simplified integral computations for faster and more accurate predictions compared to the DFT method for complex organic compounds.

The DFT approach requires a large database since optimizing geometry and performing computations is time-consuming. Therefore, the semi–empirical method is a very efficient strategy for developing new organic compounds to improve their capabilities [37]. Multiple studies have been carried out to analyze the QSAR of organic molecules. This research utilized the semi–empirical PM3 approach as a geometric optimization technique to calculate the descriptor values of the molecules, among other factors. The following studies were conducted by [29, 30, 38, 39, 40, 41, 42, 43, 44, 45]. The semi–empirical PM3 approach optimizes geometry using the Polak-Ribiere algorithm at the ground state with an RMS gradient of 0.001 kcal/Amol. The optimized compounds have larger structures due to the inclusion of oppositely charged groups, resulting in increased spatial separation from the charge distribution around atoms and groups.

**Table 2. Descriptors and methods of determination**

<table>
<thead>
<tr>
<th>No</th>
<th>Symbol</th>
<th>Descriptors</th>
<th>Unit</th>
<th>Methods of determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>q</td>
<td>Net atomic charge of atoms C1–C12 and atoms O1–O4 and N1</td>
<td>Coulomb</td>
<td>Semiempirical PM3, Hyperchem, optimization of coumarin compounds with oxime ester groups</td>
</tr>
<tr>
<td>2</td>
<td>μ</td>
<td>Dipole moment</td>
<td>Debye</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Log P</td>
<td>n-octanol/water partition coefficient</td>
<td>Å³</td>
<td>QSAR properties, Hyperchem</td>
</tr>
<tr>
<td>4</td>
<td>α</td>
<td>Molecular polarizability</td>
<td>Å³</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>BM</td>
<td>Molecular weight</td>
<td>s.m.a</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>A_dw</td>
<td>Van der Waals surface area</td>
<td>Å³</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>V_dw</td>
<td>Van der Waals volume</td>
<td>Å³</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>RD</td>
<td>Refractive index</td>
<td>Å</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>E_HOMO</td>
<td>HOMO energy</td>
<td>eV</td>
<td>Orbitals, Hyperchem</td>
</tr>
</tbody>
</table>
Table 3. Results of correlation test analysis with JASP

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Correlation value</th>
<th>Significance value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM (amu)</td>
<td>-0.851</td>
<td>0.001</td>
</tr>
<tr>
<td>(\alpha) (Å)</td>
<td>-0.825</td>
<td>0.001</td>
</tr>
<tr>
<td>RD (Å)</td>
<td>-0.809</td>
<td>0.001</td>
</tr>
<tr>
<td>qN1</td>
<td>-0.67</td>
<td>0.001</td>
</tr>
<tr>
<td>qO3</td>
<td>0.607</td>
<td>0.001</td>
</tr>
<tr>
<td>HOMO</td>
<td>-0.6</td>
<td>0.001</td>
</tr>
<tr>
<td>(V_{\text{vdw}}) (Å)</td>
<td>-0.593</td>
<td>0.001</td>
</tr>
<tr>
<td>qC8</td>
<td>0.588</td>
<td>0.001</td>
</tr>
<tr>
<td>LUMO</td>
<td>0.577</td>
<td>0.001</td>
</tr>
<tr>
<td>(\mu) (Debye)</td>
<td>0.527</td>
<td>0.001</td>
</tr>
<tr>
<td>(A_{\text{vdw}}) grid</td>
<td>-0.484</td>
<td>0.009</td>
</tr>
<tr>
<td>qC10</td>
<td>0.48</td>
<td>0.01</td>
</tr>
<tr>
<td>qC1</td>
<td>0.46</td>
<td>0.014</td>
</tr>
<tr>
<td>qC2</td>
<td>-0.39</td>
<td>0.04</td>
</tr>
<tr>
<td>qC9</td>
<td>0.383</td>
<td>0.044</td>
</tr>
<tr>
<td>qO1</td>
<td>0.37</td>
<td>0.053</td>
</tr>
<tr>
<td>qC6</td>
<td>-0.319</td>
<td>0.098</td>
</tr>
<tr>
<td>(\log P)</td>
<td>-0.304</td>
<td>0.116</td>
</tr>
<tr>
<td>qO4</td>
<td>-0.218</td>
<td>0.265</td>
</tr>
<tr>
<td>qC5</td>
<td>-0.2</td>
<td>0.308</td>
</tr>
<tr>
<td>qC11</td>
<td>-0.172</td>
<td>0.383</td>
</tr>
<tr>
<td>qC7</td>
<td>-0.165</td>
<td>0.401</td>
</tr>
<tr>
<td>qC4</td>
<td>0.159</td>
<td>0.418</td>
</tr>
<tr>
<td>qC3</td>
<td>0.097</td>
<td>0.623</td>
</tr>
<tr>
<td>qO2</td>
<td>-0.065</td>
<td>0.744</td>
</tr>
<tr>
<td>(A_{\text{vdw}}) approx.</td>
<td>0.41</td>
<td>0.837</td>
</tr>
<tr>
<td>qC12</td>
<td>0.01</td>
<td>0.959</td>
</tr>
</tbody>
</table>

Table 4. Six equations from backward analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.885 + 0.289 (HOMO) - 0.925 (LUMO) + 0.04 ((\mu)) - 0.167 ((\alpha)) - 0.081 (RD) - 0.003 (BM) + 0.012 ((V_{\text{vdw}})) + 1.261 \times 10^{-5} (A_{\text{vdw}}) + 5.982 (qC1) - 1.657 (qC2) + 13.803 (qC8) + 3.289 (qC9) + 6.471 (qC10) + 6.487 (qN1) + 9.331 (qO3)</td>
</tr>
<tr>
<td>2</td>
<td>8.885 + 0.289 (HOMO) - 0.925 (LUMO) + 0.04 ((\mu)) - 0.167 ((\alpha)) - 0.081 (RD) - 0.003 (BM) + 0.012 ((V_{\text{vdw}})) + 5.98 (qC1) - 1.657 (qC2) + 13.803 (qC8) + 3.288 (qC9) + 6.471 (qC10) + 6.485 (qN1) + 9.329 (qO3)</td>
</tr>
<tr>
<td>3</td>
<td>4.659 - 0.655 (LUMO) + 0.045 ((\mu)) - 0.138 ((\alpha)) - 0.057 (RD) - 0.003 (BM) + 0.009 ((V_{\text{vdw}})) + 3.434 (qC1) - 1.572 (qC2) + 9.932 (qC8) + 2.285 (qC9) + 4.944 (qC10) + 5.463 (qN1) + 6.862 (qO3)</td>
</tr>
<tr>
<td>4</td>
<td>3.892 - 0.469 (LUMO) + 0.067 ((\mu)) - 0.123 ((\alpha)) - 0.052 (RD) - 0.002 (BM) + 0.008 ((V_{\text{vdw}})) - 2.228 (qC2) + 7.142 (qC8) + 1.667 (qC9) + 4.223 (qC10) + 3.874 (qN1) + 3.834 (qO3)</td>
</tr>
<tr>
<td>5</td>
<td>3.741 - 0.298 (LUMO) + 0.062 ((\mu)) - 0.139 ((\alpha)) - 0.036 (RD) - 0.002 (BM) + 0.006 ((V_{\text{vdw}})) - 2.038 (qC2) + 5.067 (qC8) + 1.339 (qC9) + 3.799 (qC10) + 3.174 (qN1)</td>
</tr>
<tr>
<td>6</td>
<td>3.946 - 0.288 (LUMO) + 0.052 ((\mu)) - 0.132 ((\alpha)) - 0.048 (RD) + 0.006 ((V_{\text{vdw}})) - 2.066 (qC2) + 5.261 (qC8) + 1.523 (qC9) + 3.615 (qC10) + 3.038 (qN1)</td>
</tr>
</tbody>
</table>
The optimized structure was assessed for its physicochemical properties using the descriptors (Table 2). A statistical correlation approach is employed to evaluate the 27 descriptors. Correlation studies investigated the relationship between descriptors and their anti-diabetic activity. The test allows for identifying the characteristics that exert the greatest influence on the IC₅₀ value of coumarin derivative compounds containing oxime ester groups, which are used as anti-diabetic agents. Correlation analysis was conducted using the two-tailed approach and Pearson correlation coefficient. Correlation analysis results are utilized to select descriptors as independent variables in MLR statistical calculations. The results of the correlation test between variables and log IC₅₀ are presented in Table 3.

Based on the correlation study, only 15 of the 27 descriptors had a significance value below 0.05. Therefore, a total of 15 descriptors that exhibited the highest correlation values and met the criteria for acceptable significance were selected as independent variables to conduct a multi-linear regression analysis to establish the QSAR equation. The variables used as descriptors for the MLR test, based on the JASP results, are BM, α, RD, qN1, qC3, E HOMO, Vvdw, qC8, E LUMO, μ, AVdw, qC1, qC2, and qC9. Correlation data is presented in Table 3. For the multilinear regression analysis of the log IC₅₀ value, 15 specific descriptors were used as independent variables. Utilizing the backward technique, multilinear regression analysis was conducted on 21 fitting compounds, deriving 6 equation models as displayed in Table 4. The statistical values employed for the test compound (7 compounds) included the partition coefficient, correlation coefficient, standard deviation, estimated F/F table, and PRESS value.

Upon analyzing the data shown in Table 5, it is evident that the estimated values of R, R², SD, and F and F table values exhibit little differences. However, the PRESS value stands out as significantly divergent. Consequently, this research selects equation model 5, which corresponds to the lowest PRESS value. The MLR calculation results obtained using the backward method (model Equation 5) will serve as a fundamental reference for the MLR calculation using the Enter method. This will be done to fit the compounds used for training and the test compounds (28 compounds). The resulting data will then be analyzed for statistical values, including correlation coefficient, partition coefficient, standard deviation, Fcount/Ftable, and PRESS for all the compounds. The ideal QSAR equation is derived from the findings of the Enter approach, and Figure 2 demonstrates the comparison between observed log IC₅₀ and predicted log IC₅₀. The correlation coefficient exceeds 0.8 according to the optimal QSAR equation and the correlation image (Figure 4) between experimental log IC₅₀ and anticipated log IC₅₀. This suggests the predicted log IC₅₀ values are widely scattered over the experimental log IC₅₀ correlation line. This suggests that the equation is highly effective and can be utilized in creating innovative molecules using Oxime Ester Coumarin derivatives.

---

Table 5. Value R, R², SD, F calc/F table, and PRESS from six equations

<table>
<thead>
<tr>
<th>Equation</th>
<th>R</th>
<th>R²</th>
<th>SD</th>
<th>F calc/F table</th>
<th>PRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.991</td>
<td>0.982</td>
<td>0.084</td>
<td>1.923</td>
<td>24,826</td>
</tr>
<tr>
<td>2</td>
<td>0.991</td>
<td>0.982</td>
<td>0.076</td>
<td>3.160</td>
<td>24,843</td>
</tr>
<tr>
<td>3</td>
<td>0.987</td>
<td>0.974</td>
<td>0.084</td>
<td>3.278</td>
<td>17,739</td>
</tr>
<tr>
<td>4</td>
<td>0.982</td>
<td>0.964</td>
<td>0.093</td>
<td>3.241</td>
<td>7,677</td>
</tr>
<tr>
<td>5</td>
<td>0.980</td>
<td>0.960</td>
<td>0.094</td>
<td>3.782</td>
<td>5,227</td>
</tr>
<tr>
<td>6</td>
<td>0.978</td>
<td>0.956</td>
<td>0.094</td>
<td>4.453</td>
<td>6,758</td>
</tr>
</tbody>
</table>

---

Figure 2. Correlation of experimental log IC₅₀ with predicted log IC₅₀

Figure 3. (a) Protein and (b) native ligand

Figure 4. (a) Grid box, (b) docking protein and native ligand
The novel product was designed by modifying substituents at the R1, R2, and R3 positions in the original coumarin oxime ester compound (Table 6). The resulting compound was further refined utilizing the PM3 method’s geometric optimization. Based on the calculation results of log IC₅₀ prediction using Equation 1, it is known that eleven newly modified compounds had lower predicted log IC₅₀ values compared to 28 synthesized coumarin derivative compounds. This demonstrates that modifying the structure produces compounds that are more potent in inhibiting the activity of the α-glucosidase enzyme, as predicted.

Table 6. Design of new compounds and their predicted log IC₅₀

<table>
<thead>
<tr>
<th>Compound number</th>
<th>Structure</th>
<th>Log IC₅₀ Pred.</th>
<th>Compound number</th>
<th>Structure</th>
<th>Log IC₅₀ Pred.</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>![Structure 29]</td>
<td>-0.00018</td>
<td>35</td>
<td>![Structure 35]</td>
<td>0.314485</td>
</tr>
<tr>
<td>SMILES:C(C(=N\OC(=O))C(C=C/C1=CC=C(Br)C=C1)C1=CC2=C(O1)=C(Br)C=C1)C(C1=CC2=CC(((_N))(Br)=C2OC1=O)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>![Structure 30]</td>
<td>0.377253</td>
<td>36</td>
<td>![Structure 36]</td>
<td>-0.06001</td>
</tr>
<tr>
<td>SMILES:C(C(=N\OC(=O))C(C=C/C1=CC=C(Br)C=C1)C1=CC2=C(O1)=C(Br)C=C1)C(CC(C(=O)=O)C=C2OC1=O)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>![Structure 31]</td>
<td>0.140997</td>
<td>37</td>
<td>![Structure 37]</td>
<td>0.250982</td>
</tr>
<tr>
<td>SMILES:C(C(=N\OC(=O))C(C=C/C1=CC=C(Br)C=C1)C1=CC2=C(O1)=C(Br)C=C1)C(CC(C(=O)=O)C=C2OC1=O)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>![Structure 32]</td>
<td>0.285733</td>
<td>38</td>
<td>![Structure 38]</td>
<td>0.320</td>
</tr>
<tr>
<td>SMILES:C(C(=N\OC(=O))C(C=C/C1=CC=C(Br)C=C1)C1=CC2=C(O1)=C(Br)C=C1)C(CC(C(=O)=O)C=C2OC1=O)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>![Structure 33]</td>
<td>0.334779</td>
<td>39</td>
<td>![Structure 39]</td>
<td>0.355</td>
</tr>
<tr>
<td>SMILES:C(C(=N\OC(=O))C(C=C/C1=CC=C(Br)C=C1)C1=CC2=C(O1)=C(Br)C=C1)C(CC(C(=O)=O)C=C2OC1=O)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>![Structure 34]</td>
<td>0.344635</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMILES:CC1=CC2=C(O1)=C2CC(((_N))(OC(=O))C=C=C(Br)C=C2)C=C1Br</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7. Docking results of Novel Coumarin derivatives

<table>
<thead>
<tr>
<th>No.</th>
<th>Compound</th>
<th>Binding energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ligand native</td>
<td>-7.283</td>
</tr>
<tr>
<td>2</td>
<td>Compound 29</td>
<td>-6.667</td>
</tr>
<tr>
<td>3</td>
<td>Compound 30</td>
<td>-6.882</td>
</tr>
<tr>
<td>4</td>
<td>Compound 31</td>
<td>-6.307</td>
</tr>
<tr>
<td>5</td>
<td>Compound 32</td>
<td>-7.374</td>
</tr>
<tr>
<td>6</td>
<td>Compound 33</td>
<td>-7.05</td>
</tr>
<tr>
<td>7</td>
<td>Compound 34</td>
<td>-5.678</td>
</tr>
<tr>
<td>8</td>
<td>Compound 35</td>
<td>-6.554</td>
</tr>
<tr>
<td>9</td>
<td>Compound 36</td>
<td>-6.844</td>
</tr>
<tr>
<td>10</td>
<td>Compound 37</td>
<td>-7.505</td>
</tr>
<tr>
<td>11</td>
<td>Compound 38</td>
<td>-8.351</td>
</tr>
<tr>
<td>12</td>
<td>Compound 39</td>
<td>-7.45</td>
</tr>
</tbody>
</table>

In addition to predicting IC_{50} values, the determination of the potential of a new compound can also be studied by in silico analysis. In silico analysis can be conducted using the molecular docking approach, utilizing proteins targeted based on their biological activity. Molecular docking analysis is used to observe and compare the interactions between the target protein, the predicted compound, and its native ligand.

Molecular docking analysis was performed on the protein 2OLE using eleven new compounds derived from coumarin with ester oxime groups. The protein 2OLE, downloaded from https://www.rcsb.org, was prepared by removing water molecules and separating them from their native ligands. The separated protein and ligand, which have lost their water molecules, are then supplemented with hydrogen atoms and charges to adjust the docking environment to approximate the pH conditions inside the body and to restore hydrogen atoms in the macromolecule, allowing the formed hydrogen bonds to be observed [46]. Figure 4 presents the native protein and ligand resulting from the preparation. Before conducting docking, the molecular docking method must be validated. The validity of the method is determined by redocking the native ligand onto the target protein using the AutoDockTools 1.5.6 tool.

In the validation of molecular docking methods, the configuration of a grid box is also performed to serve as the space for the native ligand to form conformations when docked with the target protein, which is done after the protein. The grid box (Figure 5a) serves as the location for the ligand to interact with the amino acid residues at the binding site of the target protein. The grid box is determined to ascertain the coordinates of the binding site of a protein. The grid box configuration involves setting the grid center coordinates and size. RMSD (Root Mean Square Deviation) validates molecular docking by comparing the native ligand conformation from docking with that from crystallographic measurements [31].

The acceptable range for the RMSD value is < 3 Å [28]. The RMSD value of protein 2OLE for the native ligand is 1.64 Å, and the affinity energy is -7.28 kcal/mol. Based on these results, the method used is valid, and the docking process for the new compound derivative of coumarin with an ester oxime group can be performed. The visualization of the grid box and the interaction between the target protein and the native ligand is presented in Figures 5a and b. Figure 6 also illustrates the interaction of native ligands with the amino acids of the target protein. The amino acids bound to the native ligand include TYRB:666, TYRB:547, SERB:630, HISB:740, TYRB:662, ARGB:125, ASNB:710, GLUB:205, and PHEB:357.

Eleven novel compounds, coumarin derivatives containing ester oxime groups, were utilized for molecular docking on a specific protein. The docking procedure employed was identical to the validation approach. By employing consistent grid box size and coordinates throughout validation, the test compounds are accurately docked into the active region of the target protein, as confirmed. The chemicals exhibit flexibility while the target protein remains rigid, enabling it to adopt its ideal shape for binding to its active site.
The molecular dynamics analysis aims to determine the stability of ligand–receptor interactions, while molecular docking does not provide information regarding the stability of these interactions over space and time. The parameters used to evaluate the stability of ligand–enzyme interactions in molecular dynamics simulations are RMSD and RMSF.

During the simulation, the stability of the genuine protein ligand and the predicted protein ligand (compound 38) were carefully studied. Both interactions have an average RMSD value below 0.4 nm. The interaction between native and predicted protein–ligand has shown a stable RMSD with an average of 0.385 nm and 0.311 nm. Based on these values, it is evident that the RMSD of the predicted protein–ligand complex is lower compared to the RMSD of the native protein–ligand complex. The graph illustrating the RMSD fluctuations is presented in Figure 8. This indicates that the prediction of protein–ligand interactions is more stable than native protein–ligand interactions.

The parameter RMSF (Root Mean Square Fluctuation) is evaluated to determine the fluctuation of ligand–amino acid interactions in the enzyme throughout the simulation. Unlike RMSD, RMSF is calculated for each constituent residue of the protein, measuring the extent of fluctuation in the movement of each residue during the simulation. Generally, the RMSF value describes the flexibility of ligand interactions with each amino acid residue. The residual amino acid protein is measured after each fluctuation with RMSF. The protein–ligand fluctuation prediction shows a lower RMSF value than the native protein–ligand. The RMSF residue projection is below 0.45 nm (Figure 8), indicating that the protein is highly stable, which supports the RMSD results. This indicates that the predicted ligand or compound 38 is more stable than its native ligand in terms of bonding and interactions with the amino acids in protein 2OLE.

The interaction between coumarin derivatives and the oxime ester group with the amino acid acting as the catalytic site of the hydrolyze 2OLE enzyme is mediated by residues TYRB:666, TYRB:547, PHEB:357, GLUB:205, SERB:630, HISB:740, and GLUB:206. The lower the value of RMSF, the more stable the interaction between the ligand and the amino acid [47]. The residual amino acid protein is measured after each fluctuation with RMSF. The protein–ligand fluctuation prediction shows a lower RMSF value than the native protein–ligand. The RMSF residue projection is below 0.45 nm, indicating that the protein is highly stable, which supports the RMSD results. This indicates that the predicted ligand or compound 38 is more stable than its native ligand in terms of bonding and interactions with the amino acids in protein 2OLE.

4. Conclusion

The best QSAR equation (Equation 1) has been obtained for the derivative compound of coumarin with an ester oxime group. Based on the equation, eleven new, more potent compounds have been designed. An in silico study was conducted on eleven new compounds with the hydrolyze protein from Homo sapiens (PDB:2OLE) using molecular docking and molecular dynamics. This study identified compound 38 (SMILES: C\{C(\=N/OC(\=O))C\=C/
C1=CC=C(Br)C=C1)C1=CC2=CC(O)=C(CC(O)=O)C=C2OC1) as the most potent compound, with the lowest binding energy value, in addition to having smaller RMSD and RMSF values than the native ligand.

Acknowledgment

The author thanks the BOPTN UIN Walisongo Semarang (980/Un.10.0/R/HK.01.14/4/2023) from the Ministry of Religion of Indonesia for funding this work through the Applied Research Grant program.

References


