

A Computational Approach to Drug Discovery: Search for Chalcone Analogues as the Potential Candidates for Anti Colorectal Cancer (HT29)

Neni FRIMAYANTI^{1,*}, Ihsan IKHTIARUDIN¹, Rahma DONA¹,
Tiara Tri AGUSTINI¹, Fri MURDIYA² and Adel ZAMRI³

¹*Sekolah Tinggi Ilmu Farmasi Riau, Jalan Kamboja, Pekanbaru, Indonesia*

²*Department of Electrical Engineering, Faculty of Engineering Science, Universitas Riau, Kampus Bina Widya, Indonesia*

³*Department of Chemistry, Faculty of Mathematic and Natural Science, Universitas Riau, Kampus Bina Widya, Indonesia*

(* Corresponding author's e-mail: nenifrimayanti@gmail.com)

Received: 31 July 2018, Revised: 20 June 2019, Accepted: 29 July 2019

Abstract

A series of 46 chalcone derivative compounds with their inhibitory activity against colorectal cancer were used as data set for developing the quantitative structure activity relationship (QSAR). 2D QSAR and 3D QSAR models have been developed with high predictive ability with r^2 and $r^2(CV)$ of 0.81 and 0.78, respectively. Results from the 2D and 3D quantitative structure activity relationship models indicate that electrostatic parameter enhanced bioactivity of the chalcone derivatives. Further, docking and molecular dynamic simulation was performed using 2wft PDB ID as the molecular target of colon cancer. Based on the docking, molecular dynamic, and biological assay, it is confirmed that compound **2**, cpd **4**, cpd **21**, cpd **23**, cpd **27**, cpd **32**, cpd **38**, and cpd **39** show better activity (active) against colorectal cancer cells.

Keywords: 2D QSAR, 3D QSAR, docking, molecular dynamic, pharmacophore, HT29

Introduction

Chalcones, considered to be the precursor of flavonoids and isoflavonoids, are abundant in edible plants. They consist of open-chain flavonoids in which the 2 aromatic rings are joined by a 3-carbon α , β -unsaturated carbonyl system. Studies have revealed that compounds with a chalcone-based structure have anti-inflammatory [1], anti-bacterial [2], antifungal [3] and anti-tumor activities [4].

Cancer is a multifactorial disease characterized by uncontrolled and abnormal cell growth. Among the various types of cancers diagnosed, it is worth mentioning that colorectal cancer is the third most fatal malignancy, affecting both men and women [5]. Most patients die with a year after diagnosis. The primary treatments used for colorectal cancer are chemotherapy and surgery. However, synthetic drugs are costly, and have several side effects. Searching for potential anticancer agents with high efficacy and low side effects are hot trends in research nowadays.

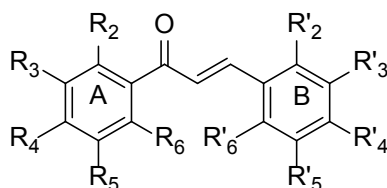
Drug design is an area where a computational approach has created a major impact, since it is very fast, cost efficient, and are almost universally applicable across a wide range of targets. The computational approach can aid in various ways and has played a critical role in the discovery of new inhibitors. This study was conducted to investigate the correlation between the structure of chalcone-based compounds and their activities; in addition, docking was used to model the interaction between a small molecule and a protein at the atomic level, which allows us to characterize the behavior of small molecules in the binding site of target proteins [6].

Materials and methods

2D, 3D QSAR modeling and pharmacophore

A data set of 45 chalcone-based compounds (presented in **Table 1**) were taken from the literature [7]; then, the data set was divided into a training set for QSAR model development and a test set for model validation. The 2D molecular structures of the ligands were sketched, converted into 3D structures, and followed with descriptor calculation. Some descriptors were generated from this calculation; these descriptors were then reduced to a smaller set of descriptors.

Table 1 Molecular structures of chalcone



No Cpd	Ring A					Ring B				
	R2	R3	R4	R5	R6	R'2	R'3	R'4	R'5	R'6
1	H	H	H	H	H	H	H	H	H	H
2	H	H	H	H	H	OH	H	H	H	H
3	H	H	H	H	H	H	OH	H	H	H
4	H	H	H	H	H	H	H	OH	H	H
5	OH	H	H	H	H	H	H	H	H	H
6	OH	H	H	H	H	OH	H	H	H	H
7	OH	H	H	H	H	H	OH	H	H	H
8	OH	H	H	H	H	H	H	OH	H	H
9	OH	H	H	H	H	H	H	OCH ₃	H	H
10	OH	H	H	H	H	H	OCH ₃	OCH ₃	H	H
11	OH	H	H	H	H	H	OCH ₃	OCH ₃	OCH ₃	H
12	OH	H	H	H	H	H	1,3 dioxolone	H	H	H
13	OH	H	H	H	H	H	H	CH ₃	H	H
14	OH	H	H	H	H	H	H	Cl	H	H
15	OH	H	H	H	H	H	Phenyl	H	H	H
16	OH	H	OCH ₃	H	OCH ₃	H	H	N(CH ₃) ₂	H	H
17	OH	H	OCH ₃	H	OCH ₃	H	H	OCH ₃	H	H
18	OH	OCH ₃	OCH ₃	H	H	H	H	Br	H	H
19	OH	H	OCH ₃	H	OCH ₃	H	H	H	H	H
20	OH	H	OCH ₃	H	OCH ₃	NO ₂	H	H	H	H
21	OH	H	OCH ₃	H	OCH ₃	H	H	NO ₂	H	H
22	OCH ₃	H	OCH ₃	H	H	H	H	NO ₂	H	H
23	OCH ₃	H	OCH ₃	H	H	H	H	Cl	H	H
24	OCH ₃	H	OCH ₃	H	H	H	H	Br	H	H
25	OCH ₃	H	OCH ₃	H	H	H	1,3-dioxolone	H	H	H
26	OCH ₃	H	H	OCH ₃	H	H	1,3-dioxolone	H	H	H
27	OCH ₃	H	H	OCH ₃	H	H	OCH ₃	OCH ₃	OCH ₃	H
28	H	OCH ₃	OCH ₃	H	H	H	H	Cl	H	H
29	H	OCH ₃	OCH ₃	H	H	H	H	SCH ₃	H	H
30	OCH ₃	H	H	H	H	H	H	F	H	H
31	H	H	OCH ₃	H	H	OCH ₃	H	H	H	H
32	H	H	OCH ₃	H	H	H	H	Cl	H	H
33	H	H	OCH ₃	H	H	H	Phenyl	H	H	H
34	H	H	CH ₃	H	H	H	H	N(CH ₃) ₂	H	H

No Cpd	Ring A					Ring B				
	R2	R3	R4	R5	R6	R'2	R'3	R'4	R'5	R'6
35	H	H	CH ₃	H	H	H	H	F	H	H
36	H	H	CH ₃	H	H	H	H	Cl	H	H
37	H	H	CH ₃	H	H	H	H	SMe	H	H
38	NH ₂	H	H	H	H	H	H	H	H	H
39	NH ₂	H	H	H	H	H	H	NO ₂	H	H
40	NH ₂	H	H	H	H	H	H	Cl	H	H
41	NH ₂	H	H	H	H	H	H	OCH ₃	H	H
42	NH ₂	H	H	H	H	H	Phenyl	H	H	H
43	H	H	NH ₂	H	H	H	H	H	H	H
44	H	H	NH ₂	H	H	H	H	H	H	H
45	H	H	NH ₂	H	H	H	H	Cl	H	H
46	H	H	NH ₂	H	H	Cl	H	Cl	H	H

Correlation matrix can be used to choose the best subset of descriptors by eliminating the high correlation between each descriptor. The next step is scaling descriptors; scaling descriptors are required to prevent the manipulation of each descriptor, since there may be underlying relationships between these descriptors. Scaling can be calculated as follows;

$$y_i = \frac{x_j - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

y_i is the scaled value, x_i is the original value, x_{min} is the minimum of the collection of x object and x_{max} is the maximum of the collection of x object. QSAR models were developed using multiple linear regression analysis (MLRA) technique. In this technique, leave-one-out (LOO) method was used to perform the cross-validation analysis.

The 3D QSAR model was carried out using partial least squares (PLS) technique. The 3D QSAR model consisted of 2 variables- independent and dependent variable. Biological activity was used as the dependent variable, and descriptors were used as the independent variables. Cross validation was performed to validate the QSAR model by predicting the biological activity of unknown compounds in the test set. The pharmacophore modeling was generated using some features, such as hydrogen bond donor, hydrogen bond acceptor, and hydrophobic atom. The pharmacophore features were then used to ensure that the 3D QSAR model had a predictive power to predict the biological activities of unknown compounds. The applicability of the 3D QSAR model to the molecules using the same properties as pharmacophore alignment was also shown.

Docking and molecular dynamic simulation

Docking was performed using Autodock. 45 chalcone compounds and macromolecules (downloaded from pdb database www.pdb.org: PDB ID 2wft) were prepared before the docking process. The protein had been previously minimized, the water molecule removed, non-polar hydrogen merged, Kollman charges assigned, and the solvation parameters added. The ligands were sketched using the ChemDraw software package, followed with the addition of the Gasteiger charges to the minimized ligand structures, and all bonds were made rotatable and flexible by allowing the detection of root torsion. Grid maps with grid spacing of 0.39 Å in the x, y, and z-dimensions of 126×126×126 points were set to cover the entire protein. The Lamarckian Genetic Algorithm (LGA) was used to search for the lowest binding energy by implementing local minimization of the genetic algorithm, to enable modification of the gene population. Docking was then performed with the clustered output based on the root-mean-square deviation (RMSD) tolerance of 2.0 Å.

The relationship between structure and protein function is directly connected to atomistic aspects of protein dynamics. As a major methodology in structural biology, molecular dynamics (MD) simulations permit the exploration of the physical mechanism underlying the function of proteins by examining their

dynamical behavior. In this study, MD was performed using CHARMM (Chemistry at HARvard Macromolecular Mechanics) as a force field. NAMD (Nanoscale Molecular Dynamics program; v 2.9) and visual MD (VMD), 2 widely-used software programs, were used to run the MD simulation.

Results and discussion

2D, 3D QSAR modeling and pharmacophore

Multiple regression calculates an equation describing the relationship between a single dependent Y variable (biological activity) and several explanatory X variable (descriptors). The statistical output of this model is shown in **Table 2**, with the equation presented as follows;

$$Y = -0.92513633 * \text{vamp octupole XXX} + 0.89910382 * \text{vamp octupole YYX} \\ + 2.3147724 * \text{vamp octupole ZZY} + 1.212097 * \text{vamp octupole ZZZ} \\ - 1.4879624 \quad (2)$$

The $r^2(CV)$ value of 0.78 indicates the very good internal predictive capability of the developed model. The model also exhibited a non-cross-validated correlation coefficient r^2 of 0.81. The high value of this parameter adds to its usefulness as a predictive tool. The statistical output of this QSAR model is presented in **Table 2**.

Table 2 Statistical output of QSAR model

Statistical output	Value
r^2	0.81
$r^2(CV)$	0.78
Residual sum of square (RSS)	2.01
Predictive sum of square (PRESS)	4.68

Based on the QSAR model developed above, the electrostatic parameter is observed to be positively correlated with the biological activity. The electrostatic parameters are properties of a molecule which are related to the electron affinity, and characterize the susceptibility of the molecules toward attack nucleophiles. Biological activity improves with an increase in the electrostatic parameter.

A plot of experimental *vs.* predicted IC₅₀ is shown in **Figure 1**. This plot is important to graphically demonstrate the predictive capability of QSAR, hence, the developed QSAR is considered to be stable. The inhibition concentration of the compounds in the test set was predicted using the QSAR model (Eq. (2)); this is listed in **Table 3**.

The 3D QSAR model was then validated by predicting the IC₅₀ of compounds in the test set, and the descriptors in the 3D QSAR model were validated using pharmacophore alignment. A pharmacophore can be defined as the effect of the steric or electronic features that are used to confirm the interaction between a protein with the biological target. The pharmacophore was developed with the most active ligand, and also generated the importance of features; they are features of hydrophobic (green sphere), hydrogen bond donor (purple sphere), and aromatic ring features (orange sphere). These features were located in the pharmacophore alignment, as presented in **Figure 2**. It can be concluded and presumed that the biological activity of the most active compounds, such as compound **38**, will be influenced by the existence of these 3 features.

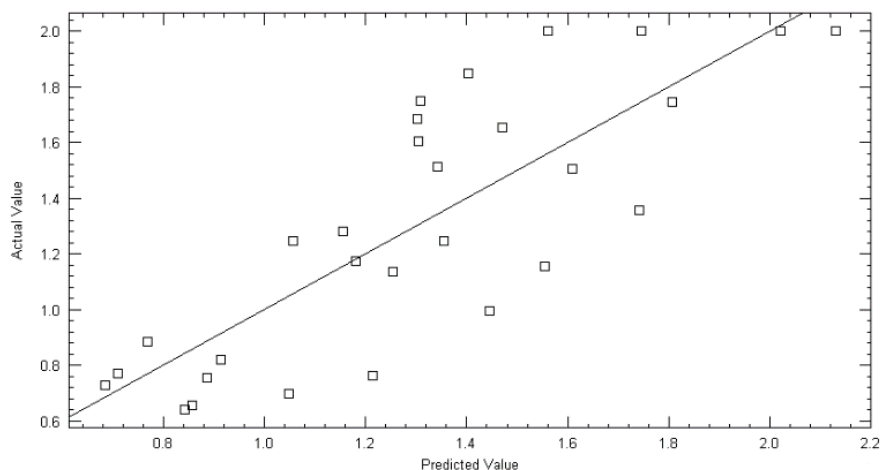


Figure 1 Plot of actual value vs predicted value

Table 3 Predicted value of IC₅₀ compounds in the test set based on 2D QSAR model

No of compound	Log 1/IC ₅₀	Predicted log 1/IC ₅₀
16	-2	-2.01
17	-2	-2.13
20	-2	-1.97
22	-2	-1.96
25	-2	-1.98
29	-2	-2.04
30	-2	-2.03
31	-2	-2.1
33	-2	-1.95
34	-2	-1.95
35	-2	-1.98
36	-2	-1.97
37	-2	-1.98
41	-2	-2.11
42	-2	-2.03
43	-2	-1.9
46	-2	-2.01

Docking and molecular dynamic simulation

Molecular docking is a good tool to predict and match the desired binding site, understand the possible conformation of the compounds, and further clarify the binding interactions in the binding pocket method [8]. Knowledge of the preferred orientation, in turn, may be used to predict the strength of association or binding affinity between 2 molecules. The best poses from the docking results were selected based on the lowest docked energy values. Eight compounds (the most active ligands) of 46 compounds were observed in a similar manner orientation of the protein active site. Superimposition of these 8 ligands is shown in **Figure 3**.

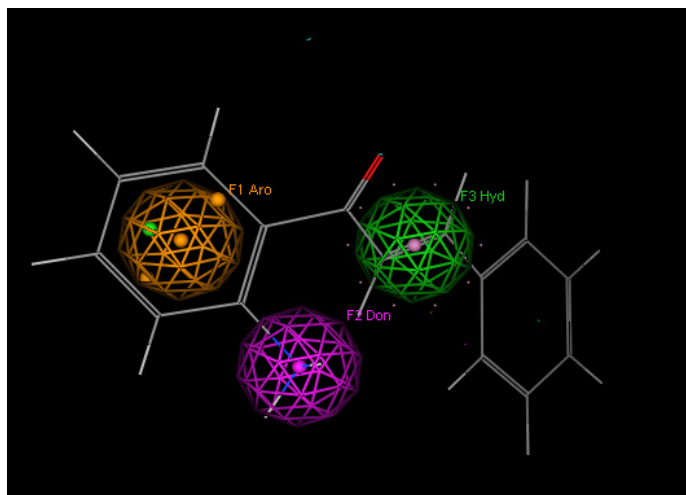


Figure 2 Best pharmacophore hypothesis for the most active ligand (compound **38**). Orange sphere is featured for aromatic ring, green sphere for hydrophobic, and purple for hydrogen bond donor

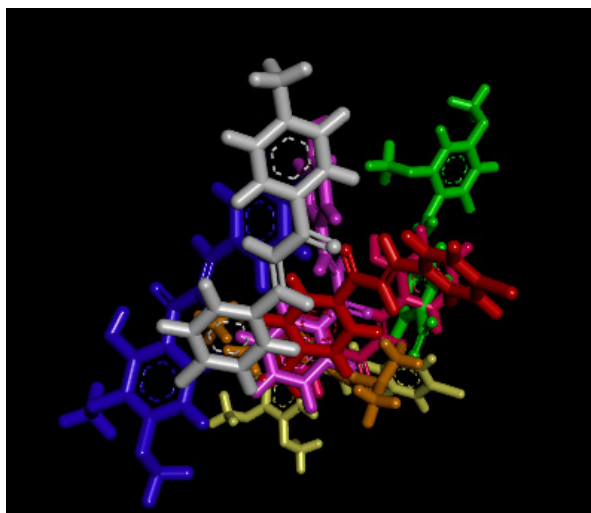


Figure 3 Superimposition of compound **2** (pink), compound **4** (purple), compound **21** (orange), compound **23** (blue), compound **27** (green), compound **32** (yellow), compound **38** (red), and compound **39** (white)

Compound **2** was shown to have one interaction via hydrophobic interaction with Ser578, and one van der Waals interaction with the residue Glu257. The best docking poses for compound **4** exhibited the hydrogen bond between the hydroxyl group of the ligand with Ser224. In addition, van der Waals interaction was also observed involving Glu221 amino acid. The binding interaction for compound **2** and compound **4** are depicted in **Figure 4**.

According to the docking results, compound **21** interacts with one amino acid residue Glu258 via van der Waals interaction, and this compound also builds hydrophobic interaction with amino acid residue Arg544. For compound **23** and compound **27**, both of these compounds have the same interaction

as each other. As other active compounds, these 2 compounds have 2 kinds of interactions with the amino acid residue, van der Waals interaction with Glu227 and π interaction with residue Val598. The spatial arrangement of compound **21**, compound **23**, and compound **27** are depicted in **Figure 5**.

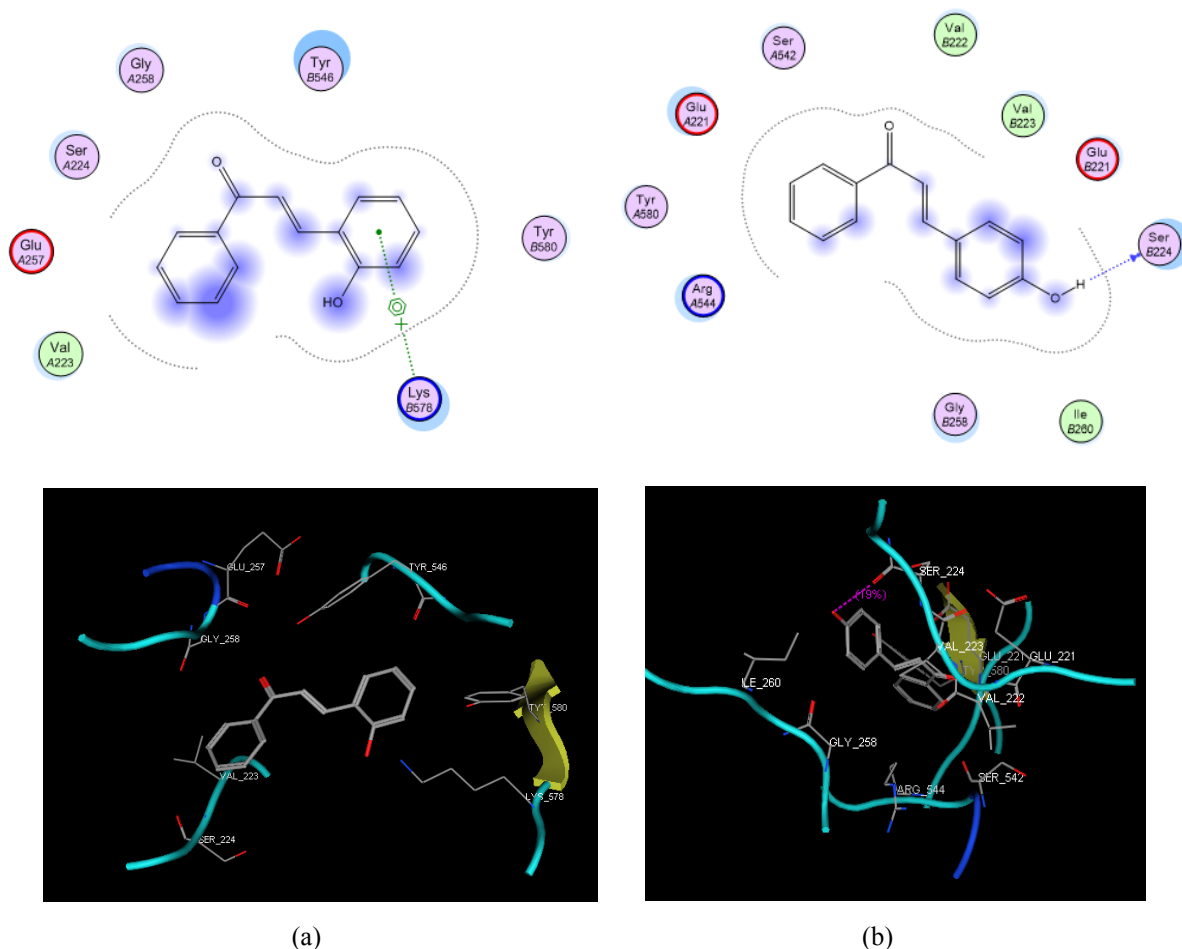
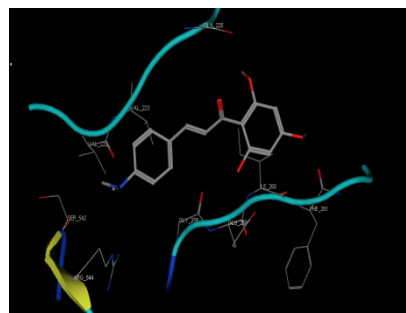
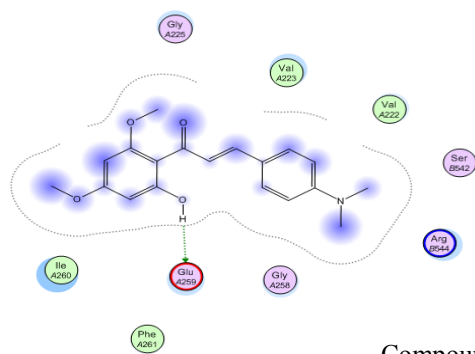
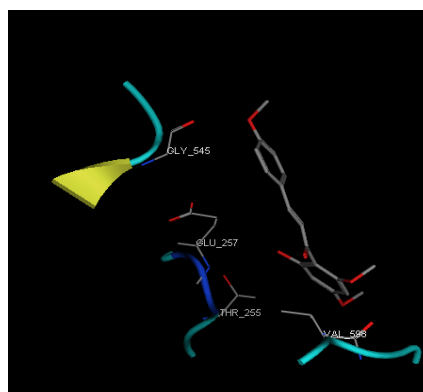
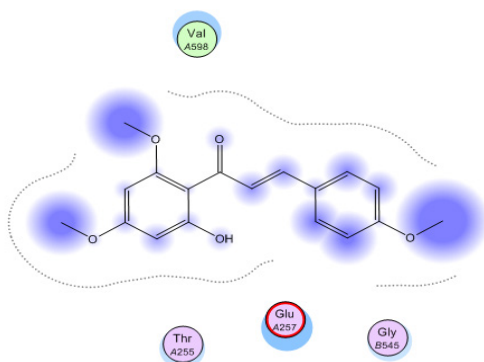


Figure 4 Spatial arrangements of compound **2** (a) and compound **4** (b)

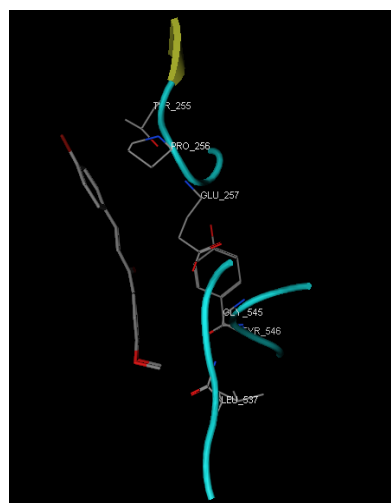
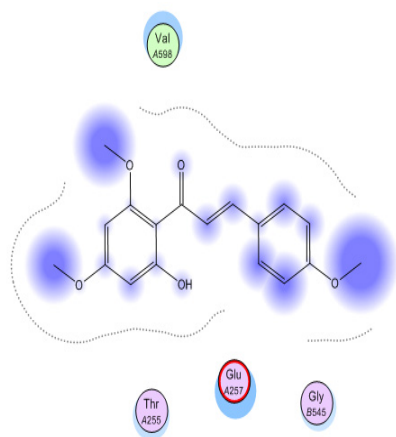
Based on the biological activity, and from the docking results, it is also confirmed that compounds **32**, compound **38**, and compound **39** are active compounds. Based on their spatial arrangement, compound **32** has 3 types of hydrogen bonding; they are between carbonyl of ligand with amino acid residue Lys262, and between methoxy of ligand with Asn330 and Lys329. Compound **38** only has one hydrogen bonding between carbonyl group of the ligand with Ala601; in addition, this compound also has a hydrophobic interaction with His332, probably caused by the compound becoming active. Compound **39** has one hydrogen bonding interaction with His332 and van der Waals interaction with Glu257. The spatial arrangement of these compounds is presented in **Figure 6**. Based on this, the biological activity and the docking results suggests that the presence of carbonyl may enhance the hydrogen bonding and make the compound became active. In addition, the presence of van der Waals interaction or hydrophobic interaction with the residues surrounding the active site may increase the biological activity.



Compound 21



Compound 23



Compound 27

Figure 5 Binding interaction of compound 21, compound 23, and compound 27

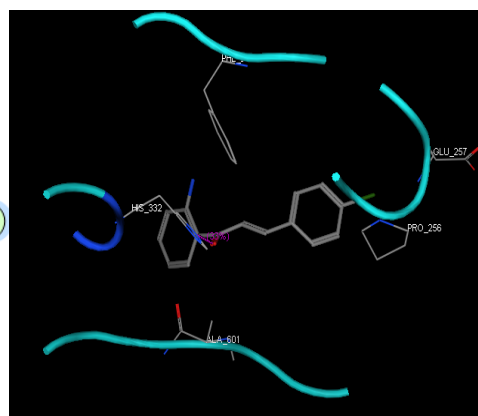
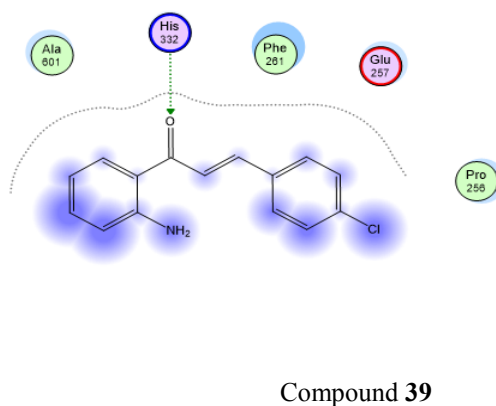
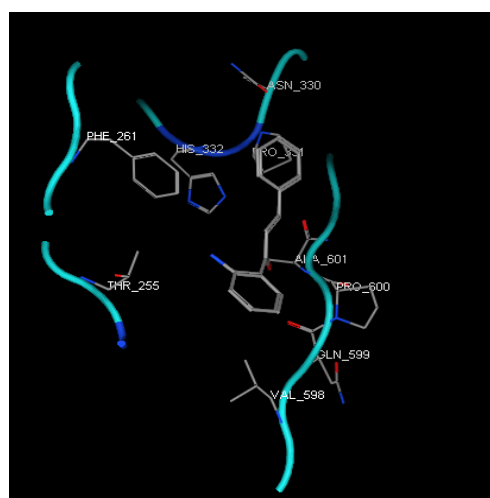
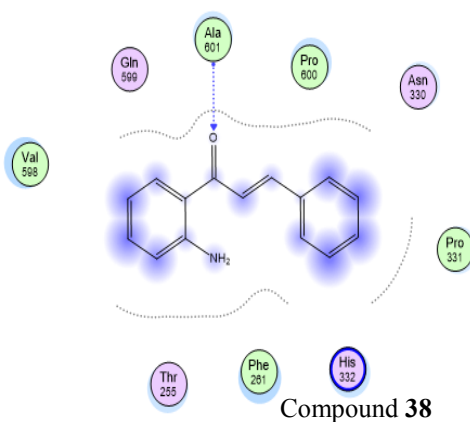
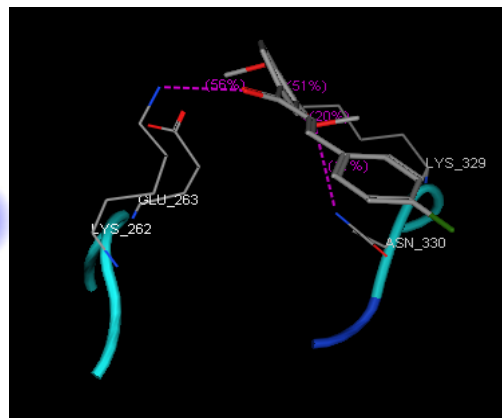
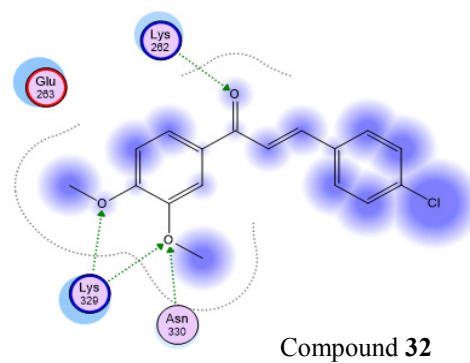


Figure 6 Binding site of compound 32, compound 38, and compound 39

Another computational technique involved in this study is Molecular Dynamic simulation (MD). MD simulations have enabled researchers to rationalize experimentally-measured properties, to analyze ligand-receptor interactions, and to refine models of biomolecules determined by X-ray or NMR methods [9]. A properly-constructed MD simulation, in which a solvated protein and its unbound ligand are subjected to physical laws, can yield a stable protein-ligand complex [10]. Using MD, it is possible to theoretically characterize how protein structure and stability are affected by an explicit solvent.

In this study, MD simulation is carried out on these 8 active compounds to further explore the interaction between ligand and receptor [11]. The properties of each complex were inspected to check the stability of MD. In this study, the MD simulations were performed at a temperature of 300⁰K to check the affinity of the ligand to the binding site. Generally, 8 of these ligands are maintained to binding with the essential residues, such as Gly258 and Lys578. The MD simulation results are depicted in **Table 4**. After the MD simulation, 8 of these active ligands were maintained to interact with the same residues, and those residues (Gly258 and Lys578), presumably, can enhance biological activity.

Table 4 Interaction of ligands through molecular dynamic simulation and docking

No	Cpd	Docking	MD
1	Cpd 2	Val223, Ser224, Glu257, Gly258, Tyr546, Lys578	Val223, Ser224, Glu257, Tyr580, Tyr546, Lys578
2	Cpd 4	Glu221, Ser224, Gly258, Tyr580, Ser542, Lys578	Glu221, Ile260, Lys262, Asn576, Gly549, Lys578
3	Cpd 21	Gly256, Gly258, Ser542, Arg544	Lys247, Leu245, Glu246, Gly258, Lys578
4	Cpd 23	Thr255, Glu257, Gly545	Gly258, Gly545, Lys578
5	Cpd 27	Thr255, Glu257, Gly545	Leu245, Glu257, Gly258, Lys578
6	Cpd 32	Lys262, Glu263, Asn330, Lys329	Asn330, Glu257, Gly258
7	Cpd 38	Phe251, Thr255, Pro331, His322, Asn330, Ala601	Phe251, Glu257, Lys578
8	Cpd 39	Glu257, Asn330, Gly258, Lys578	Asn330, Gly258, Gly545, Lys578

Conclusions

The computational approach can be used in the design of new drugs in the pharmaceutical industries. From the 2D QSAR model, it is shown that the electrostatic properties can increase inhibitory activity and, from the 3D QSAR model, it is shown that biological activity will be enhanced with influence from the hydrogen bond donor and hydrophobic and aromatic ring properties. Docking and molecular dynamic simulation showed the maintain hydrogen bond and van der Waals interaction in the binding pocket. It is indicated that these chalcone can be used as potential candidates for second generation drug discovery.

References

- [1] HK Sieh, LT Tsao, JP Wang and CN Lin. Synthesis and anti-inflammatory effect of chalcones. *J. Pharm. Pharmacol.* 2000; **52**, 163-71.
- [2] D Dragana, BM Marina, I Branka and C Ivana. Antibacterial activity of three newly-synthesized chalcones & synergism with antibiotics against clinical isolates of methicillin-resistant *Staphylococcus aureus*. *Indian J. Med. Res.* 2014; **140**, 130-7.
- [3] G Deepa and DK Jain. Chalcone derivatives as potential antifungal agents: Synthesis, and antifungal activity. *J. Adv. Pharm. Tech. Res.* 2015; **6**, 114-7.
- [4] DK Mahapatra, SK Bharti and V Asati. Anti-cancer chalcones: Structural and molecular target perspectives. *Eur. J. Med. Chem.* 2015; **15**, 69-114.

- [5] AM Zeinab, ME Mastoura, AM Rasha, FMG Sobhi, SA Seham and NM Yahia. Synthesis, antitumor evaluation and molecular docking of new morpholine based heterocycles. *Molecules* 2017; **22**, 1211-25
- [6] C Yu-Chien. Beware of docking. *Trend Pharm. Sci.* 2014; **36**, 617.
- [7] CW Mai, M Yaeghoobi, NA Rahman, YB Kang and MR Pichika. Chalcones with electron-withdrawing and electron-donating substituents: Anticancer activity against TRAIL resistant cancer cells, structure-activity relationship analysis and regulation of apoptotic proteins. *Eur. J. Med. Chem.* 2014; **77**, 378-87.
- [8] J Jasril, I Ihsan, Z Adel, YT Hilwan and F Neni. New fluorinated chalcone and pyrazoline analogs: synthesis, docking and molecular dynamic studies as anticancer agents. *Thai J. Pharmaceut. Sci.* 2017; **41**, 93-8.
- [9] R Amit, R Vinit, KS Ashok, KK Amit, K Umesh, K Dinesh and S Sudipta. Design and synthesis of 1,4-benzothiazine derivatives with promising effects against coloerectal cells. *Cogent Chem.* 2017; **3**, 1-24.
- [10] Z Adel, F Neni and YT Hilwan. Docking and molecular dynamic simulations: Study of 1, 3, 4-oxadiazolechalcone hybrid derivatives to search new active anticancer agents. *Thai J. Pharmaceut. Sci.* 2016; **40**, 179-84.
- [11] H Adam, RG Josep, O Modesto and LG Josep. Molecular dynamic simulations: Advances and application. *Adv. Appl. Bioinform. Chem.* 2015; **8**, 37-47.