

RESEARCH ARTICLE

# Virtual Screening of Inhibitor Compounds for HL-60 Cell Line Using QSAR and Molecular Docking Approaches

(Virtual Screening Senyawa Inhibitor untuk HL-60 Cell Line Menggunakan Pendekatan QSAR dan Docking Molekular)

Maulina Surindri Putri Handayani<sup>1</sup>, Sudarko<sup>1\*</sup>, Ari Satia Nugraha<sup>2</sup>

<sup>1</sup>Jurusan Kimia, FMIPA, Universitas Jember, Jl. Kalimantan 37, 68121, Jember, Indonesia

<sup>2</sup>Prodi S-1 Farmasi, Universitas Jember, Jl. Kalimantan 37, 68121, Jember, Indonesia

## ABSTRACT

Leukaemia is a type of heterogeneous malignancy disease in which the number of immature white blood cells in the bone marrow is increased or poorly differentiated. Inhibitors of HL-60 cell line are important drugs in the treatment of leukaemia. This study aims to identify potential inhibitors for HL-60 leukaemia cells using virtual screening with QSAR and molecular docking approaches. The best QSAR model obtained was the model with  $R^2$  testing value of 0.72. The best model was applied to the Enamine database, obtained 3,841,278 compounds with  $pIC_{50}$  above 9 were clustered to end up with 50 inhibitor candidate compounds. The compounds were further studied through molecular docking against CDK-2 and BCL-2 target proteins with binding free energy analysis. Screened compounds have high inhibitory potential against CDK-2 and BCL-2 target proteins in HL-60 cells which can be proposed as potential inhibitors of HL-60 cells. These compounds can be further tested for the development of more effective antileukaemia drugs.

Leukemia adalah jenis penyakit keganasan heterogen di mana jumlah sel darah putih yang belum matang di sumsum tulang meningkat atau berdiferensiasi buruk. Inhibitor garis sel HL-60 adalah obat penting dalam pengobatan leukemia. Penelitian ini bertujuan untuk mengidentifikasi potensi penghambat sel leukemia HL-60 menggunakan virtual screening dengan QSAR dan pendekatan molecular docking. Model QSAR terbaik yang diperoleh adalah model dengan nilai pengujian  $R^2$  sebesar 0,72. Model terbaik diterapkan pada basis data Enamine, diperoleh 3.841.278 senyawa dengan  $pIC_{50}$  di atas 9 diklasifikasi hingga menghasilkan 50 senyawa kandidat inhibitor. Senyawa tersebut dipelajari lebih lanjut melalui penambatan molekuler terhadap protein target CDK-2 dan BCL-2 dengan analisis energi bebas pengikatan. Senyawa yang disaring memiliki potensi penghambatan yang tinggi terhadap protein target CDK-2 dan BCL-2 dalam sel HL-60 yang dapat diusulkan sebagai penghambat potensial sel HL-60. Senyawa ini dapat diuji lebih lanjut untuk pengembangan obat antileukemia yang lebih efektif.

**Keywords:** HL60, Inhibitor, QSAR, Virtual Screening, Docking Molecular.

\*Corresponding author:  
Sudarko  
E-mail: darko@unej.ac.id

## PENDAHULUAN

Leukaemia yaitu jenis keganasan heterogen di mana jumlah sel darah putih tidak matang dalam sumsum tulang meningkat atau berdiferensiasi buruk [1]. Yayasan Onkologi Anak Indonesia menyatakan bahwa terdapat sekitar 11.000 kasus kanker baru pada anak yang ditemukan setiap tahunnya di seluruh Indonesia dan 30 - 40% diantaranya merupakan leukaemia [2].

Metode Komputasi dapat digunakan untuk proses desain obat. Desain obat dapat digunakan untuk penemuan obat baru dalam berbagai penyakit dengan menggunakan metode *virtual screening*. Metode *virtual*

*screening* dapat diklasifikasikan menjadi metode desain obat berbasis struktur dan berbasis ligan berdasarkan ketersediaan informasi strukturalnya. Metode desain obat berbasis struktur menggambarkan pengikatan molekul sementara metode berbasis ligan membahas hubungan aktivitas struktur kuantitatif dan pemodelan farmakologis [3].

Pendekatan penelitian menggunakan analisis hubungan kuantitatif *structure-activity* (QSAR) dan *docking* molekuler secara teoritis dapat digunakan untuk memprediksi aktivitas biologis yang paling optimal dari sebuah senyawa. Pendekatan teoritis dengan bantuan kimia Komputasi sangat penting

dalam proses perancangan obat-obatan baru. Metode QSAR menggambarkan hubungan antara deskriptor molekul dan aktivitas biologisnya [4]. Interaksi senyawa dengan protein target dapat divisualisasikan dengan metode *docking* dan dapat digunakan untuk mengetahui *pharmacophore* dari suatu senyawa [5].

Dalam penelitian ini, kami melaporkan desain senyawa-senyawa potensial yang mampu menghambat sel HL-60 menggunakan analisis QSAR untuk merekomendasikan beberapa senyawa dengan aktivitas penghambatan yang lebih baik. Selanjutnya, senyawa-senyawa potensial di *docking* untuk memprediksi interaksi senyawa dengan protein target CDK-2 (PDB ID: 7KJS) dan BCL-2 (PDB ID: 6O0K).

## METODE PENELITIAN

### Alat dan Bahan

Alat perangkat keras yang di gunakan adalah Komputer/PC Laptop Lenovo Idea Center AIO 520-5MID sebanyak 5 buah dengan sistem operasi *Linux* Ubuntu 20.04 LTS, RAM 4GB dan prosesor Intel Core i5. Perangkat lunak menggunakan Smina, OpenBabel, PyMOL, dan Biovia Discovery Studio.

Bahan-bahan yang digunakan adalah website protein data bank (<https://www.ebi.ac.uk/chembl/>) dan Enamine database ([www.enamine.net](http://www.enamine.net)).

### Konstruksi Model

Dataset sebanyak 2.903 dari HL-60 *cell line* untuk masa inkubasi 72 jam diunduh dari database ChEMBL 31 (<https://www.ebi.ac.uk/chembl/>) dengan kode ChEMBL383. Dataset berisi data senyawa dan nilai aktivitas dibagi menjadi tiga bagian yaitu data *training* (85%), data *validation* (5%) dan data *testing* (10%).

Komputer dengan sistem operasi *Linux* Ubuntu 20.04 LTS yang telah dilengkapi dengan program Anaconda dan Python disiapkan. Dataset dari HL-60 dilakukan optimalisasi model QSAR sebanyak 375 kombinasi *hyperparameter neural network*. Hasil yang diperoleh berupa model terbaik yang kemudian digunakan untuk proses prediksi nilai aktivitas.

### Virtual Screening

Skrining virtual menggunakan molekul pada *database* Enamine ([www.enamine.net](http://www.enamine.net)) sebanyak 6 milyar molekul untuk dilakukan pemeringkatan. Proses pemeringkatan dilakukan berdasarkan prediksi nilai  $pIC_{50}$  menggunakan model terbaik yang telah

dihasilkan pada prosedur sebelumnya. Skrining tersebut akan menghasilkan molekul dengan nilai  $pIC_{50}$  terbaik. Molekul hasil pemeringkatan selanjutnya disaring kembali menggunakan metode *clustering* K-Means. *Clustering* bertujuan untuk membantu mengidentifikasi kelompok-kelompok senyawa yang memiliki karakteristik serupa.

### Docking Molecular

Smina digunakan untuk proses *docking* molekuler karena tersedia secara gratis. Smina dikembangkan berdasarkan AutoDock Vina dengan peningkatan performa dan efisiensi [6]. *Docking* dengan Smina dijalankan pada Jupyter Notebooks berdasarkan Talktorial T015 [7].

Struktur protein target yang digunakan dapat diunduh dengan format \*.pdb pada Protein Data Bank. Protein dipreparasi menggunakan modul Pybel yang merupakan bagian dari OpenBabel untuk mengubah parameter-parameter seperti muatan parsial pada atom dan penambahan atom hidrogen. Struktur protein target tersebut kemudian diubah menjadi format \*.pdbqt menggunakan OpenBabel.

Preparasi ligan dilakukan dengan mengambil ligan dalam format SMILES yang telah diperoleh dari *virtual screening*. Ligan dengan format SMILES diubah menjadi format \*.pdbqt menggunakan OpenBabel. Selanjutnya, penentuan *binding site* berupa *gridbox* yang dilakukan dengan memasukkan Ligan\_ID dari ligan yang terikat pada protein target. Struktur Protein dan ligan yang sudah dipreparasi kemudian dilakukan proses *docking* menggunakan Smina. Hasil dari *docking* berupa nilai *binding affinity* dan RMSD. Visualisasi hasil *docking* dilakukan menggunakan PyMOL dan Biovia Discovery Studio.

## HASIL DAN PEMBAHASAN

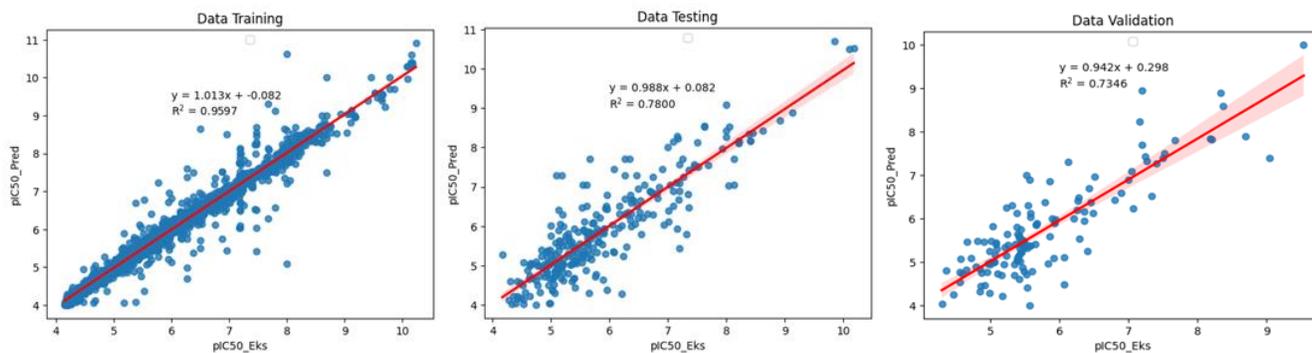
### Model QSAR

Dataset dengan kode ChEMBL383 untuk masa inkubasi 72 jam yang diperoleh dari database ChEMBL (<https://www.ebi.ac.uk/chembl/>) digunakan untuk mendapatkan model. Dataset dengan format \*.csv terdiri dari 2.903 data struktur molekul yang ditulis dalam bentuk SMILES dan nilai aktivitas molekul yang dinyatakan dengan nilai  $pIC_{50}$ . Pemodelan QSAR dilakukan dengan menjalankan 375 x 1000 *jobs* berdasarkan hasil variasi pemisahan data dan kombinasi *hyperparameter neural network* yang

digunakan pada proses pelatihan menggunakan *machine learning*. Kombinasi *hyperparameter* ini bertujuan untuk menemukan model dengan nilai *validation loss* terkecil. Model yang memiliki nilai *validation loss* terkecil dari 375 model tersebut akan dipilih untuk dilakukan uji dengan data *testing* agar akurasi model dapat diketahui berdasarkan parameter  $R^2$  dan MSE.

Model terbaik dipilih berdasarkan nilai  $R^2$  dan MSE pada setiap *state id*. Parameter  $R^2$  memiliki rentang nilai antara 0 hingga 1. Model terbaik yang didapatkan yaitu pada *state id* 82 dengan kombinasi l1; l2; l3; *drop out rate*;

dan *batch size* berturut - turut yaitu 256; 256; 64; 0,4; dan 64. Nilai parameter  $R^2$  yang dihasilkan untuk data *training*, *validation*, dan *testing* berturut - turut yaitu 0,96; 0,73; dan 0,78. Nilai-nilai tersebut sudah memenuhi syarat minimal nilai  $R^2$ . Menurut Hermansyah *et al.*, [8], menyatakan bahwa nilai minimum yang direkomendasikan untuk model QSAR regresi agar dapat menghasilkan prediksi yang akurat adalah  $R \geq 0,8$  dan  $R^2 \geq 0,6$ . Plot regresi antara nilai  $pIC_{50}$  prediksi *vs* eksperimen dari model terbaik dapat dilihat pada Gambar 1.



Gambar 1. Plot Regresi Nilai  $pIC_{50}$  Prediksi vs Eksperimen pada Model Terbaik

### Virtual Screening pada Database Enamine

Model terbaik dengan nilai  $R^2$  *testing* sebesar 0,72 digunakan untuk proses *virtual screening* dari 6 milyar senyawa pada database Enamine ([www.enamine.net](http://www.enamine.net)). Senyawa dari hasil *virtual screening* tersebut di peringkat dari yang memiliki nilai  $pIC_{50}$  tertinggi dan dipilih senyawa dengan nilai  $pIC_{50}$  diatas 9,0 yaitu sebanyak 3.841.278 senyawa. Pemilihan senyawa dengan nilai  $pIC_{50}$  sebesar 9 didasarkan pada obat leukaemia seperti venetoclax yang sudah beredar memiliki nilai  $pIC_{50}$  sebesar 9, yang bertujuan untuk mendapatkan senyawa dengan aktivitas penghambatan yang lebih baik jika dibandingkan [9].

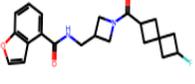
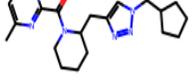
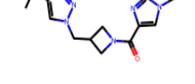
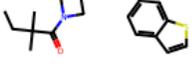
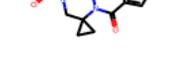
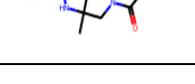
Senyawa hasil *screening* tersebut disaring kembali berdasarkan kesamaan struktur menggunakan pengelompokan (*clustering*) dengan metode K-Means. *Clustering* K-Means pada 3.841.278 senyawa menghasilkan 50 *cluster* yang terdapat senyawa dengan nilai  $pIC_{50}$  tertinggi pada setiap *cluster*-nya. Hasil *clustering* beberapa senyawa ditunjukkan pada Tabel 1.

Senyawa-senyawa yang dihasilkan dengan prediksi aktivitas nilai  $pIC_{50}$  memiliki kemungkinan untuk menjadi calon obat baru leukaemia yang efektif karena tersebut telah memenuhi aturan 5 Lipinski. Senyawa-senyawa tersebut kemudian di docking untuk mengetahui interaksi senyawa dengan protein target.

### Docking Molecular

Hasil dari docking berupa nilai binding affinity dan RMSD yang kemudian dilakukan pemeringkatan dari nilai binding affinity yang terkecil. Nilai binding affinity diukur dengan satuan energi bebas Gibbs ( $\Delta G$ ). Nilai  $\Delta G$  yang negatif memengaruhi kemampuan sebuah ligan untuk berinteraksi dengan molekul target [10]. Interaksi antara ligan dan protein target menghasilkan residu asam amino yang dapat menentukan kekuatan interaksi yang terjadi. Hasil docking dari beberapa senyawa terhadap protein target CDK-2 dapat dilihat pada Tabel 2.

Tabel 1. Hasil *Clustering* Beberapa Senyawa

ID Enamine	Struktur Molekul	pIC <sub>50</sub>
PV-008392343646		9,01
PV-005945500657		9,92
PV-005944179637		9,65
PV-005944793561		9,46
PV-005943606370		9,16
PV-008392560869		10,00
PV-004072924727		9,95

Tabel 2. Hasil Docking Molecular CDK-2

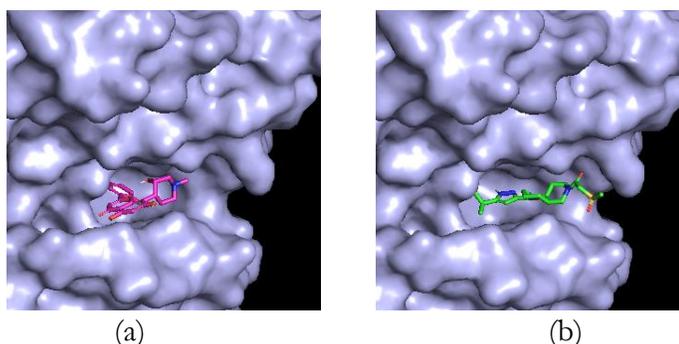
Ligan	Binding Affinity (kcal/mol)	RMSD (Å)	Residu Asam Amino	
			Ikatan Hidrogen	Interaksi Hidrofobik
PV-008392343646	-9,3	1,4	Asp86; Lys89;	Ala31; Val64; Leu134; Ile10;
PV-005945500657	-9,2	1,4	Asp145;	Ile10; Ala31; Val64; Ala144; Val18; Leu134;
PV-005944179637	-9,0	1,9	Phe82; Glu81;	Leu134; Ala31; Ile10; Val18; Ala144;
PV-005944793561	-8,9	0,9	Asp86;	Leu134; Phe80; Ala31; Val18; Lys33;
PV-005943606370	-8,8	1,8	Asp145; Gln131; Ile10;	Phe80; Ile10;
Alvocidib	-10,2	1,2	Lys33; Asp86;	Val18; Ala31; Val64; Ala144;

Hasil docking yang tertera pada Tabel 2 menunjukkan ligan dengan kode PV-008392343646 memiliki nilai binding affinity paling negatif yaitu -9,3 kcal/mol. Ligan uji dan ligan pembanding yang di docking dengan protein CDK-2 memiliki nilai binding affinity kurang dari 0 yang menunjukkan semua ligan uji dan ligan pembanding mudah berinteraksi dengan protein target. Nilai RMSD antara ligan dan residu

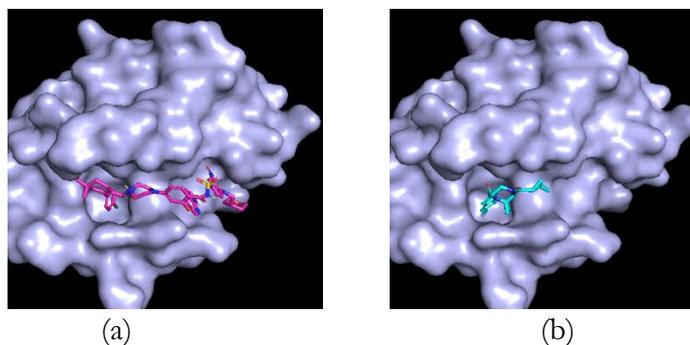
WG1 digunakan untuk validasi protokol docking. Hasil docking dapat dikatakan valid jika kalkulasi nilai RMSD pada ligan kurang dari 2 Å. Ligan uji juga di docking dengan protein BCL-2 menggunakan parameter binding affinity. Hasil docking molecular ligan terhadap protein BCL-2 dapat dilihat pada Tabel 3.

Tabel 3. Hasil Docking Molecular BCL-2

Ligan	Binding Affinity (kcal/mol)	RMSD (Å)	Residu Asam Amino	
			Ikatan Hidrogen	Interaksi Hidrofobik
PV-008392560869	-9,0	1,9		Ala34; Val41;
PV-004072924727	-8,9	1,0		Val33;
PV-005945500657	-8,7	0,8	Gly30;	Tyr7; Ala34; Arg31; Val33;
PV-005944793561	-8,2	1,1	Arg31;	Ala34;
PV-005944179637	-7,9	1,8		Ala34; Arg31; Val18; Leu22;
Venetoclax	-11,4	1,7	Gly30;	Tyr7; Leu22;



Gambar 2. Visualisasi Hasil Docking dari Ligan (a) Alvocidib; (b) PV-008392343646 Terhadap CDK-2



Gambar 3. Visualisasi Hasil Docking dari Ligan (a) Venetoclax; (b) PV-008392560869 Terhadap BCL-2

Hasil visualisasi ligan menggunakan PyMOL yang dapat dilihat pada Gambar 2 dan Gambar 3 menunjukkan kesamaan posisi untuk ligan berinteraksi dengan protein target. Ligan-ligan yang di *docking* memiliki *binding site* yang sama dengan ligan pembanding.

Visualisasi ligan uji juga dilakukan menggunakan *Biovia Discovery Studio* untuk melihat interaksi ligan pada protein target. Interaksi yang terjadi diantara ligan uji dan protein target yaitu ikatan hidrogen dan interaksi hidrofobik. Visualisasi dari *Biovia Discovery Studio* untuk protein CDK-2 dapat dilihat pada Gambar 4 dan untuk protein BCL-2 dapat dilihat pada Gambar 5.

Hasil interaksi antara ligan uji dan ligan pembanding dengan protein target yang tertera pada Tabel 2 dan Tabel 3 menunjukkan adanya beberapa

kesamaan interaksi. Ligan uji PV-008392343646 memiliki interaksi paling mirip dengan alvocidib yaitu untuk ikatan hidrogen pada residu Asp86; serta interaksi hidrofobik pada residu Ala31 dan Val64. Interaksi pada ligan PV-005945500657 juga memiliki kesamaan dengan venetoclax yaitu untuk ikatan hidrogen pada residu Gly30 dan interaksi hidrofobik pada residu Tyr7. Hasil visualisasi hasil *docking* yang menunjukkan potensial yang paling baik untuk protein target CDK-2 dan BCL-2 yaitu senyawa dengan kode PV-005945500657. Senyawa PV-005945500657 menunjukkan nilai energi bebas *binding* yang rendah terhadap kedua protein target dan memiliki interaksi yang sama dengan ligan pembanding. Menurut Prasetiawati *et al.*, [11] ligan uji yang memiliki ikatan hidrogen pada residu yang sama dengan ligan pembanding dapat memiliki aktivitas yang serupa juga.



- [5] F. F. Setiawan and E. P. Istyastono, "Uji in silico senyawa 2,6-Dihidroksiantraquinon sebagai ligan pada reseptor estrogen alfa," *J. Farm. Sains dan Komunitas*, vol. 12, no. 2, pp. 77-80, 2015.
- [6] D. R. Koes, M. P. Baumgartner, and C. J. Camacho, "Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise," *J. Chem. Inf. Model.*, vol. 53, no. 8, pp. 1893-1904, 2013.
- [7] D. Sydow *et al.*, "TeachOpenCADD 2022: Open source and FAIR python pipelines to assist in structural bioinformatics and cheminformatics research," *Nucleic Acids Res.*, vol. 50, no. W1, pp. W753-W760, 2022.
- [8] O. Hermansyah, A. Bustamam, and A. Yanuar, "Virtual screening of DPP-4 inhibitors using QSAR-based artificial intelligence and molecular docking of hit compounds to DPP-8 and DPP-9 Enzymes," *Res. Sq. Prepr.*, 2020.
- [9] S. Krishna *et al.*, "Pharmacoinformatics based screening of combined synthetic and natural compounds to identify novel and in silico potential Bcl-2 inhibitors," *J. Mol. Liq.*, vol. 366, p. 120250, 2022.
- [10] T. A. Balogun *et al.*, "Computational evaluation of bioactive compounds from colocasia affinis schott as a novel EGFR inhibitor for cancer treatment," *Cancer Inform.*, vol. 20, no. 2021, pp. 1-12, 2021.
- [11] R. Prasetiawati, M. Suherman, B. Permana, and Rahmawati, "Molecular docking study of anthocyanidin compounds against Epidermal Growth Factor Receptor (EGFR) as anti-lung cancer," *Indones. J. Pharm. Sci. Technol.*, vol. 8, no. 1, pp. 8-20, 2021.