

Analisis Molecular Fingerprint Senyawa Potensial sebagai Inhibitor Aromatase Menggunakan Metode Clustering

Jihan Fadila. (2025). *jihan4han97@gmail.com*

105222022. Universitas Pertamina.

Pendahuluan

Kanker payudara merupakan ancaman serius bagi kesehatan global, dengan 2,3 juta kasus baru atau 11,7% dari seluruh kasus kanker di dunia berdasarkan penelitian GLOBOCAN 2020 (Hyuna et al., 2021). Di Indonesia, prevalensi kanker payudara berada di peringkat pertama, mencapai 68.858 kasus atau 16,6%, dengan jumlah kematian melebihi 22.000 kasus. Kondisi ini jauh melampaui prevalensi kanker lainnya seperti kanker paru-paru (11,4%), kolorektal (10,0%), prostat (7,3%), dan lambung (5,6%) (Kemenkes, 2022).

Peningkatan kadar hormon estrogen yang dimediasi oleh enzim aromatase diketahui menjadi salah satu penyebab utama kanker payudara. Enzim aromatase, bagian dari superfamili sitokrom P450, memegang peran penting dalam biosintesis estrogen melalui konversi androstenedion menjadi estron, testosteron menjadi estradiol, serta androgen menjadi estrogen. Pada wanita pascamenopause dengan kanker payudara hormon positif, penggunaan aromatase inhibitor (AI) telah menjadi pendekatan utama untuk menghambat konversi androgen menjadi estrogen (Burstin et al., 2018).

AI bekerja dengan menghambat aktivitas enzim aromatase, sehingga menekan produksi estrogen. Berdasarkan mekanismenya, AI dikategorikan menjadi dua kelompok, yaitu penghambat steroid dan nonsteroid. AI steroid, seperti exemestane, menghambat aktivitas enzim secara ireversibel dengan membentuk ikatan kovalen. Sebaliknya, AI nonsteroid, seperti anastrozol dan letrozol, bekerja secara reversibel melalui interaksi non-kovalen dengan gugus heme enzim (Smith & Dowsett, 2003). Meskipun efektif, AI seringkali menyebabkan efek samping yang signifikan, termasuk gangguan kardiovaskular dan penurunan kesehatan tulang (Yagata et al., 2016).

Kemajuan teknologi komputasi memungkinkan pengembangan alternatif AI yang lebih aman melalui pendekatan molecular fingerprint dan analisis clustering. Pendekatan ini memungkinkan identifikasi senyawa potensial berdasarkan kesamaan struktural dan fitur molekul kunci yang relevan. Penelitian ini bertujuan untuk mengidentifikasi karakteristik struktural yang diperlukan untuk aktivitas penghambatan aromatase optimal, sekaligus memberikan landasan bagi pengembangan inhibitor baru yang lebih efektif dan aman.

Penelitian ini menggunakan analisis fingerprint molekuler yang diintegrasikan dengan metode clustering untuk mengkarakterisasi dan mengelompokkan senyawa berdasarkan kesamaan strukturalnya. Pendekatan ini tidak hanya memungkinkan identifikasi fitur molekul kunci yang bertanggung jawab atas aktivitas penghambatan aromatase, tetapi juga memberikan landasan untuk pengembangan inhibitor baru yang lebih efektif dan aman. Melalui analisis komprehensif terhadap fingerprint dominan dalam setiap cluster, penelitian ini bertujuan memberikan wawasan mendalam tentang karakteristik struktural yang diperlukan untuk aktivitas penghambatan aromatase yang optimal.

Material dan Metode

Penelitian ini menggunakan pendekatan analisis bertahap untuk mengkarakterisasi dataset senyawa. Tahap pertama melibatkan evaluasi Lipinski's Rule of Five untuk menyeleksi senyawa berdasarkan parameter drug-likeness yang penting dalam pengembangan obat. Senyawa-senyawa terpilih kemudian ditransformasikan menjadi representasi digital menggunakan SMILES-based descriptor, yang menghasilkan fingerprint molekuler yang mencerminkan karakteristik struktural setiap senyawa. Fingerprint ini selanjutnya menjadi dasar untuk analisis clustering yang komprehensif, yang bertujuan mengidentifikasi pola-pola struktural dan fitur molekuler kunci yang berpotensi mempengaruhi aktivitas penghambatan aromatase.

Data Set

Dataset penelitian ini bersumber dari ChEMBL *Cytochrome P450 19A1* (Target ID ChEMBL1978), yang secara spesifik berfokus pada data inhibitor aromatase pada manusia (*Homo sapiens*). Dari total 157 bioactivity data yang tersedia untuk berbagai target organisme, dataset mengalami serangkaian tahap kurasi yang komprehensif untuk memastikan kualitas dan konsistensi data. Tahap pertama kurasi melibatkan standarisasi struktur kimia, dimana notasi SMILES dari setiap senyawa distandarisasi menggunakan ChemAxon's Standardizer (ChemAxon Kft., 2015). Parameter standarisasi yang diterapkan mengikuti protokol yang telah ditetapkan dalam penelitian Simeon et al. (2016). Dataset awal mencakup beragam unit pengukuran bioaktivitas, termasuk IC₅₀, K_i, persentase aktivitas, persentase inhibisi, MIC, dan EC₅₀. IC₅₀ dipilih sebagai parameter utama dalam penelitian ini karena memiliki subset data terbesar dengan 4.339 senyawa. Dalam proses seleksi data bioaktivitas, sebanyak 94 dataset dieksklusi karena tidak memiliki nilai IC₅₀. Selanjutnya, untuk menjamin konsistensi analisis, seluruh nilai IC₅₀ distandarisasi ke dalam satuan nM, menghasilkan 4.245 data valid. Analisis lebih lanjut terhadap struktur molekul mengungkapkan adanya beberapa senyawa dengan notasi SMILES yang identik. Setelah menghilangkan duplikasi struktur ini, diperoleh dataset final yang terdiri dari 3.068 senyawa dengan notasi SMILES yang unik, yang kemudian digunakan sebagai basis analisis dalam penelitian ini.

Lipinski's Rule of Five

Dataset yang telah dikurasi selanjutnya dianalisis untuk menentukan parameter Lipinski's Rule of Five, yang merupakan seperangkat kriteria fundamental dalam evaluasi *drug-likeness* suatu senyawa (Lipinski et al., 2001). Parameter-parameter ini mencakup MW, LogP, NumHDonors dan NumHAcceptors, serta PIC₅₀ sebagai ukuran potensi inhibisi.

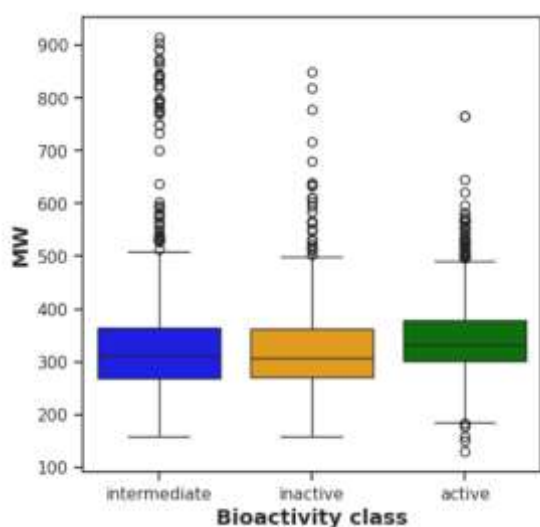
Molecular Weight (MW) merupakan masa molekul total dari suatu senyawa, dengan batas optimal ≤ 500 dalton. Parameter ini penting karena berkaitan langsung dengan kemampuan senyawa untuk menembus membran biologis, molekul yang terlalu besar umumnya sulit mencapai target enzim (Hoelz et al., 2010).

LogP (Logaritma koefisien partisi oktanol-air) menggambarkan tingkat lipofilisitas atau kelarutan senyawa dalam lemak, dengan nilai optimal ≤ 5 . Parameter ini krusial dalam menentukan kemampuan senyawa untuk menembus membran sel dan distribusinya dalam tubuh. Nilai LogP yang terlalu tinggi mengindikasikan senyawa terlalu hidrofobik dan berpotensi memiliki masalah dalam absorpsi dan kelarutan (Lipinski, 2004). Parameter ini digunakan dalam penentuan penggunaan oral atau injeksi.

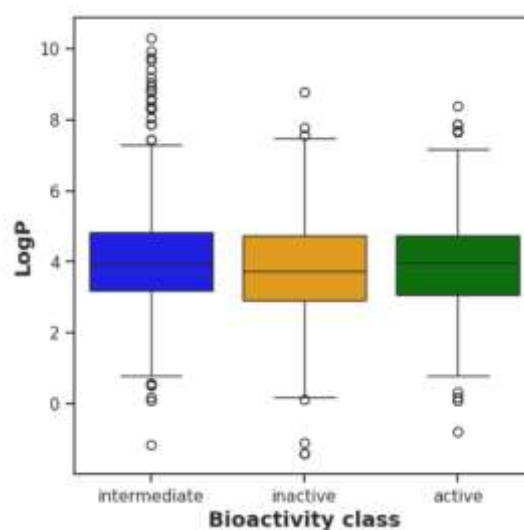
NumHDonors (jumlah donor ikatan hidrogen) dan NumHAcceptors (jumlah akseptor ikatan hidrogen) mencerminkan kapasitas senyawa untuk membentuk ikatan hidrogen. Batas optimal untuk donor adalah ≤ 5 dan untuk akseptor ≤ 10 . Kedua parameter ini vital dalam menentukan permeabilitas membran dan kemampuan senyawa untuk berinteraksi dengan target biologisnya. Jumlah ikatan hidrogen yang tepat memfasilitasi keseimbangan antara kelarutan dalam air dan kemampuan menembus membran lipid (Caron et al., 2018).

PIC50 ($-\log[\text{IC}_{50}]$) merupakan ukuran potensi inhibisi senyawa, dimana nilai yang lebih tinggi mengindikasikan aktivitas penghambatan yang lebih kuat. Parameter ini penting untuk mengkarakterisasi efektivitas senyawa sebagai inhibitor aromatase dan membandingkan potensi antar senyawa dalam dataset (Pires et al., 2015).

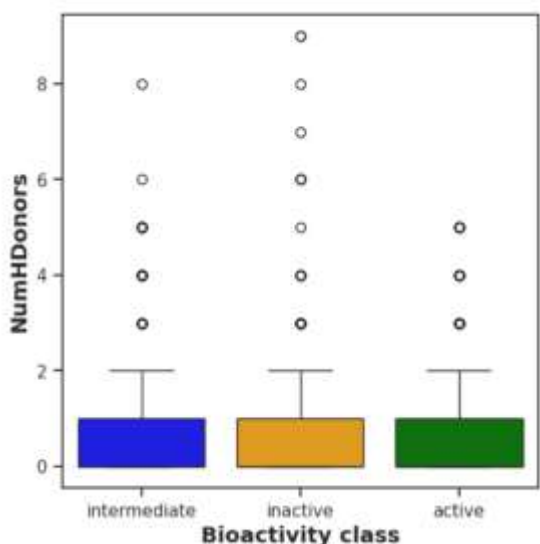
Berikut persebaran parameter Lipinski pada dataset :



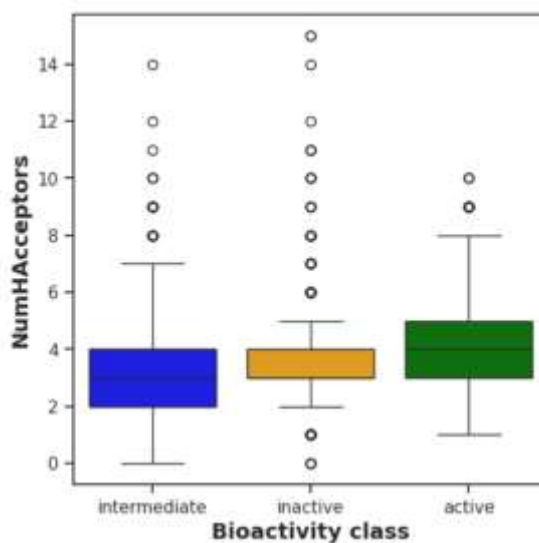
Gambar 2.2.1. Persebaran MW pada dataset



Gambar 2.2.2. Persebaran LogP pada dataset



Gambar 2.2.3. Persebaran NumHDonors



Gambar 2.2.4. Persebaran NumHAcceptors

Fingerprint Molekul

Seleksi senyawa untuk analisis fingerprint dilakukan berdasarkan dua kriteria utama yaitu aktivitas inhibisi dan karakteristik fisikokimia. Dataset difiltrasi untuk menyertakan hanya senyawa dengan nilai $PIC50 > 6$, yang mengindikasikan potensi inhibisi yang kuat terhadap aromatase. Selain itu, diterapkan pembatasan berat molekul ≤ 500 Dalton sesuai kriteria Lipinski untuk memastikan kemampuan molekul mencapai target enzim secara efektif. Dari proses seleksi ini, diperoleh 1.418 senyawa yang memenuhi kedua kriteria tersebut.

Karakterisasi struktural senyawa-senyawa terpilih dilakukan dengan mentransformasikan notasi SMILES menjadi substructure count fingerprint menggunakan PaDEL (Pharmaceutical Data Exploration Laboratory), sebuah perangkat lunak open source yang dirancang untuk menghitung deskriptor molekuler. Metode substructure count fingerprint dipilih karena kemampuannya dalam memberikan representasi kuantitatif frekuensi kemunculan substruktur fungsional dalam molekul. Menurut penelitian Simeon et al. (2016), performa metode ini dalam pemodelan QSAR (Quantitative Structure–Activity Relationship) dinilai stabil dan andal dalam menggambarkan hubungan antara struktur molekul dengan bioaktivitas senyawa. Transformasi ini menghasilkan 307 digit informasi struktural yang mencakup berbagai fitur molekuler. Keunggulan pendekatan ini terletak pada kemampuannya untuk mengkarakterisasi kontribusi relatif setiap fitur struktural terhadap aktivitas biologis. Selain itu, metode ini dapat mengidentifikasi motif struktural berulang serta pola substitusi yang berpotensi memainkan peran penting dalam interaksi dengan target enzim.

Clustering Fingerprint

Analisis fingerprint dilanjutkan dengan proses clustering menggunakan metode K-Means dan DBSCAN, dataset fingerprints ditransformasi kedalam bentuk PCA (Principal Component Analysis). PCA diterapkan untuk reduksi dimensionalitas data fingerprint, mengubah variabel yang berkorelasi menjadi serangkaian principal component yang tidak berkorelasi. Transformasi PCA dapat dirumuskan sebagai berikut:

$$X = TP^T + E$$

dimana X adalah matriks data asli ($n \times p$), T adalah matriks score ($n \times k$), P adalah matriks loading ($p \times k$), dan E adalah matriks residual. Setiap principal component merupakan kombinasi linear dari variabel asli yang memaksimalkan variansi data dalam arah orthogonal.

Hasil transformasi PCA selanjutnya digunakan dalam proses clustering dengan metode K-Means. Algoritma K-Means bekerja dengan meminimalkan sum of squared distances antara data point dan centroid cluster yang terdekat:

$$J = \sum_{i=1}^n \sum_{j=1}^k \|x_i - c_j\|^2 \times w_{ij}$$

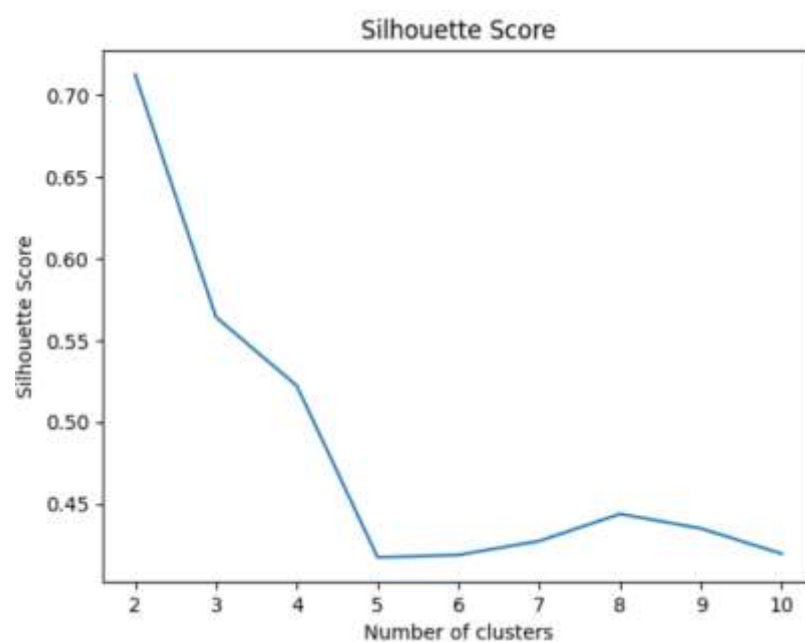
dimana x_i adalah data point ke- i , c_j adalah centroid cluster ke- j , dan w_{ij} adalah indikator keanggotaan binary (1 jika x_i termasuk cluster j , 0 jika tidak).

Penentuan jumlah cluster optimal dilakukan menggunakan analisis silhouette score, yang mengukur seberapa mirip suatu objek dengan cluster-nya sendiri dibandingkan dengan cluster lainnya. Silhouette score untuk setiap data point i dihitung dengan:

$$s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\}$$

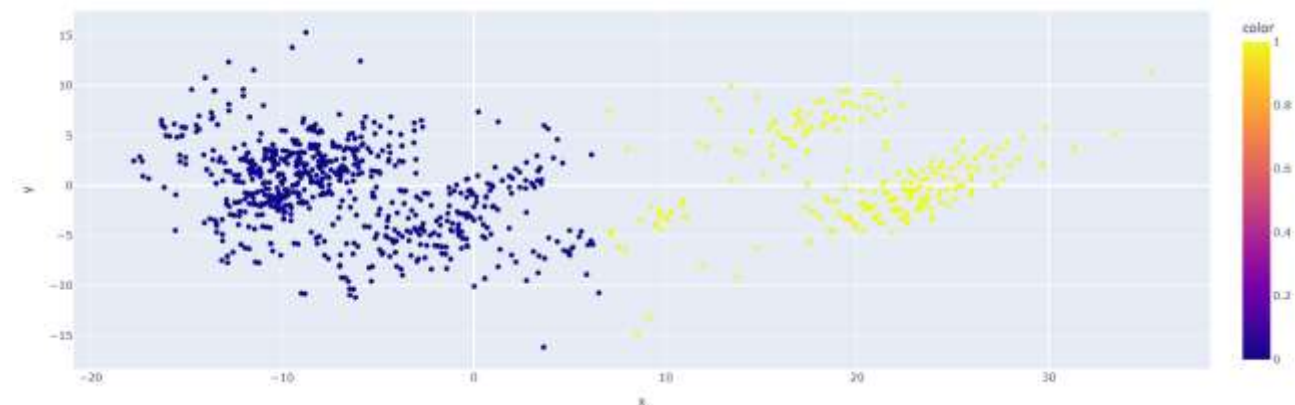
dimana $a(i)$ adalah rata-rata jarak i ke semua point dalam cluster yang sama, dan $b(i)$ adalah rata-rata jarak minimum i ke semua point di cluster lain. Berdasarkan analisis silhouette score, nilai tertinggi 0,7 diperoleh pada konfigurasi dengan 2 cluster, mengindikasikan bahwa pengelompokan optimal dicapai dengan membagi dataset menjadi dua kelompok distinktif. Terdapat sebanyak 1.034 data senyawa pada cluster 0 dan 384 data pada cluster 1.

Berikut hasil analisis silhouette score yang digunakan dalam penentuan cluster :



Gambar 2.3.1. Grafik nilai Silhouette menggunakan metode K-Means

Visualisasi hasil clustering K-Means disajikan pada Gambar 2.3.2, yang mengilustrasikan distribusi spasial kedua cluster yang teridentifikasi:



Gambar 2.3.2. Visualisasi hasil Cluster menggunakan metode K-Means

Visualisasi hasil clustering K-Means mengindikasikan adanya kerapatan dan variasi densitas yang signifikan dalam distribusi data. Berdasarkan karakteristik distribusi tersebut, analisis dilanjutkan dengan penerapan DBSCAN (Density-Based Spatial Clustering of Applications with Noise) terhadap hasil transformasi PCA. DBSCAN dipilih karena kemampuannya mengidentifikasi cluster dengan bentuk arbitrary dan menangani noise dalam dataset secara efektif.

Algoritma DBSCAN beroperasi berdasarkan dua parameter fundamental: radius tetangga epsilon (ϵ) dan threshold minPoints. Parameter ϵ mendefinisikan radius maksimum tetangga untuk setiap titik, sementara minPoints menentukan jumlah minimum titik dalam radius ϵ yang diperlukan untuk membentuk cluster. Berdasarkan parameter ini, DBSCAN mengklasifikasikan setiap titik data ke dalam tiga kategori: core point (titik dengan minimal minPoints tetangga dalam radius ϵ), border point (titik dalam radius ϵ dari core point dengan tetangga kurang dari minPoints), dan noise point (titik yang tidak memenuhi kriteria sebelumnya).

Formulasi matematis DBSCAN didasarkan pada beberapa konsep densitas. ϵ -neighborhood dari titik p didefinisikan sebagai:

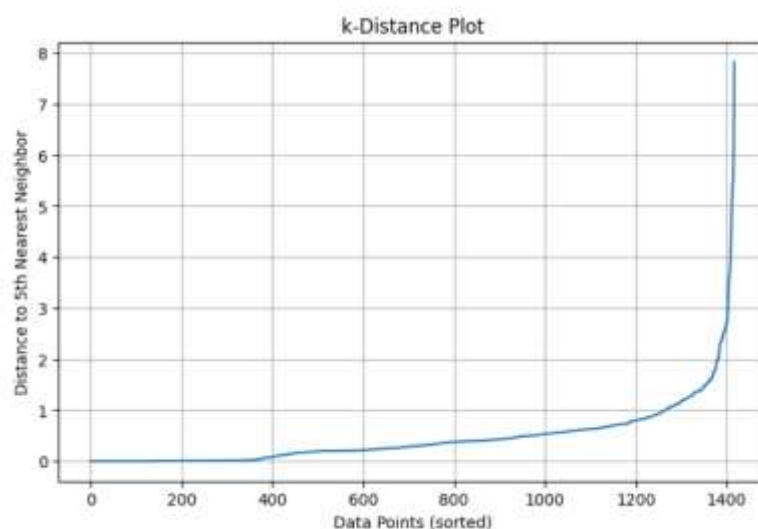
$$N_{\epsilon}(p) = \{q \in D \mid \text{dist}(p,q) \leq \epsilon\}$$

Konsep directly density-reachable ditentukan ketika titik p memenuhi kondisi:

$$p \in N_{\epsilon}(q) \text{ dan } |N_{\epsilon}(q)| \geq \text{minPoints}$$

Density-reachability didefinisikan melalui eksistensi rangkaian titik p_1, \dots, p_n dengan $p_1 = q$ dan $p_n = p$, dimana setiap p_{i+1} directly density-reachable dari p_i . Selanjutnya, density-connectivity antara titik p dan q terjadi jika terdapat titik o dimana keduanya density-reachable dari o .

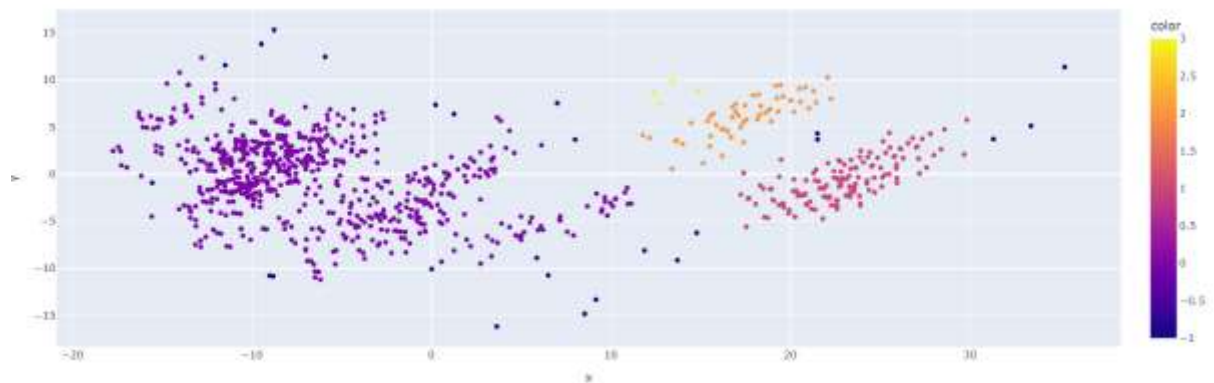
Penentuan parameter ϵ optimal dilakukan melalui analisis K-Distance plot, dengan mengidentifikasi titik elbow pada kurva yang merepresentasikan perubahan signifikan dalam densitas data. Berikut hasil analisis K-Distance plot yang digunakan dalam penentuan parameter ϵ :



Gambar 2.3.3. K-Distance Plot

Analisis K-Distance plot menghasilkan nilai epsilon awal sebesar 2, yang digunakan sebagai parameter dalam algoritma DBSCAN. Implementasi algoritma dengan parameter tersebut menghasilkan empat cluster distinktif. Evaluasi kualitas clustering menggunakan silhouette score menghasilkan nilai 0,5654, yang mengindikasikan tingkat separasi cluster yang moderat. Nilai ini relatif lebih rendah dibandingkan dengan hasil clustering menggunakan metode K-Means (silhouette score 0,7), menunjukkan bahwa struktur cluster yang diidentifikasi oleh DBSCAN memiliki batasan yang kurang tegas.

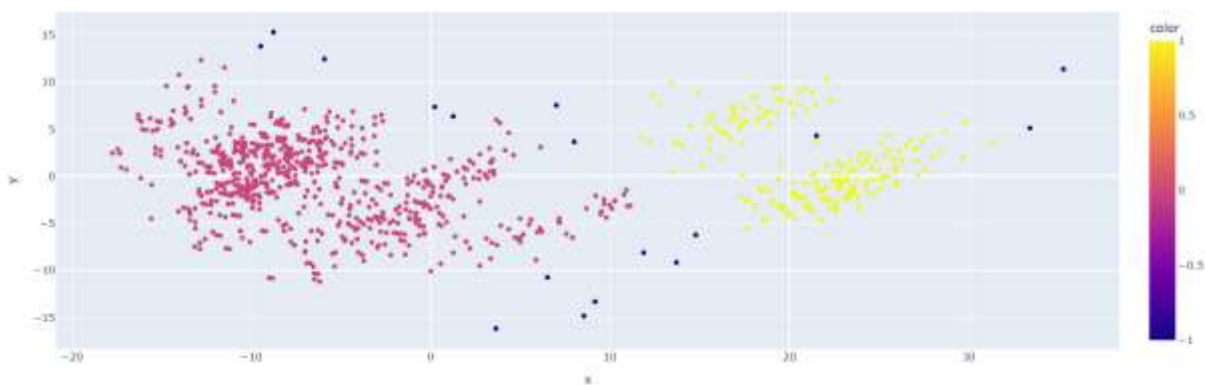
Visualisasi hasil clustering DBSCAN disajikan pada Gambar 2.3.4, yang mengilustrasikan distribusi spasial keempat cluster yang teridentifikasi:



Gambar 2.3.4. Hasil cluster DBSCAN $\epsilon = 2$ samples = 5

Selanjutnya, melakukan optimasi parameter DBSCAN dengan melakukan pencarian sistematis nilai epsilon dan minimum samples. Range pencarian epsilon ditentukan berdasarkan analisis K-Distance plot, dengan rentang 0,8 hingga 2,7, menggunakan 50 titik evaluasi. Hasil optimasi mengidentifikasi parameter optimal dengan nilai epsilon 2,5836734693877554 dan minimum samples sebanyak 8 data point. Implementasi DBSCAN dengan parameter optimal tersebut menghasilkan dua cluster distinktif dengan silhouette score 0,617, mengindikasikan peningkatan kualitas clustering dibandingkan hasil sebelumnya. Terdapat sebanyak 1.055 data senyawa pada cluster 0, 345 data senyawa pada cluster 1, serta 18 senyawa teridentifikasi sebagai noise points.

Representasi visual hasil clustering optimal disajikan pada Gambar 2.3.5 :

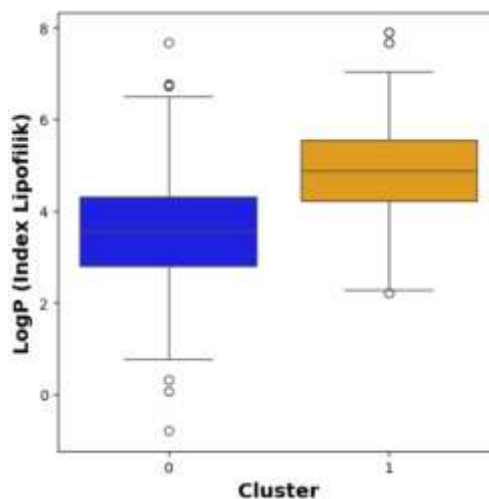
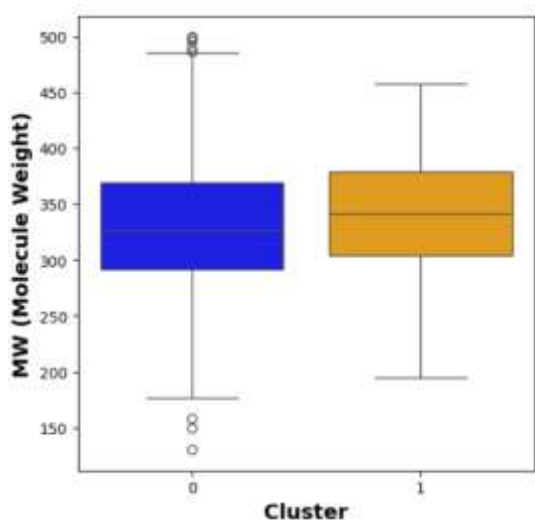


Gambar 2.3.5. Hasil cluster DBSCAN $\epsilon = 2,5836734693877554$ samples = 8

Hasil dan Pembahasan

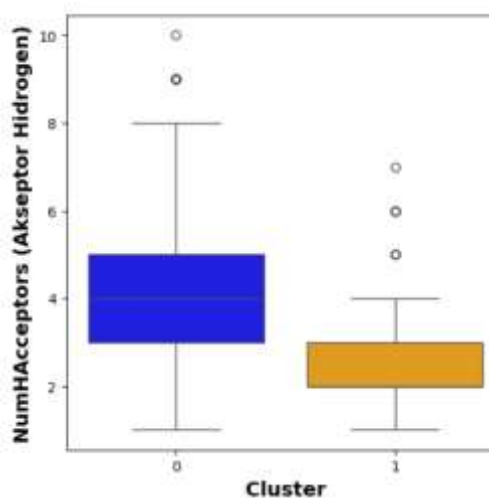
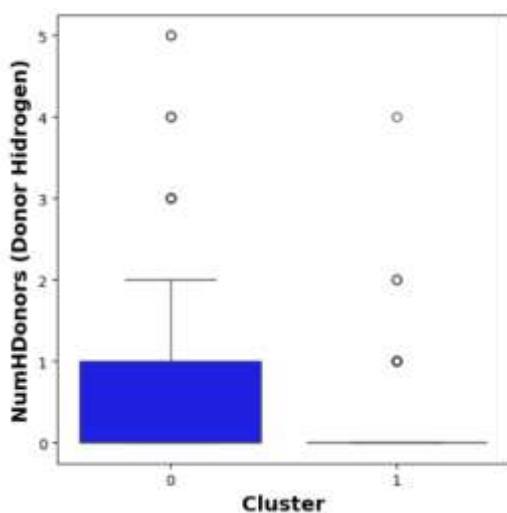
Parameter Lipinski's Rule of Five

Karakterisasi cluster dilakukan melalui analisis distribusi parameter Lipinski's Rule of Five. Analisis ini mengungkapkan perbedaan karakteristik fisikokimia yang signifikan antar cluster. Distribusi berat molekul (MW) menunjukkan pola yang relatif serupa antara kedua cluster, dengan Cluster 0 terkonsentrasi pada rentang 290-370 Dalton dan Cluster 1 pada rentang 300-380 Dalton. Perbedaan yang lebih substansial teramati pada distribusi LogP, dimana Cluster 0 menunjukkan rentang 2,8-4,2, sementara Cluster 1 memiliki karakteristik yang lebih lipofil dengan rentang 4,2-5,5. Perbedaan paling distinktif terobservasi pada parameter ikatan hidrogen. Jumlah donor ikatan hidrogen (NumHDonors) pada Cluster 0 bervariasi antara 0-1, sedangkan Cluster 1 secara konsisten menunjukkan nilai 0. Demikian pula, jumlah akseptor ikatan hidrogen (NumHAceptors) menunjukkan perbedaan yang jelas, dengan Cluster 0 memiliki rentang 3-5 akseptor dan Cluster 1 menunjukkan distribusi yang lebih terbatas pada rentang 2-3 akseptor.



Gambar 3.1.1 Distribusi MW dalam Cluster

Gambar 3.1.2. Distribusi LogP dalam Cluster



Gambar 3.1.3. Distribusi NumHDonors

Gambar 3.1.4. Distribusi NumHAceptors

Interpretasi Mekanik dan Fitur Penting

1. Cluster 0

Seperti yang terlihat pada Gambar 3.2.1, fitur dengan peringkat tertinggi adalah pusat kiral dengan konfigurasi spesifik (SubFPC307), yaitu atom yang terikat pada empat substituen berbeda, dalam molekul dengan pusat kiral, konfigurasi ini sering ditentukan menggunakan sistem penamaan seperti R/S. Dalam konteks desain obat, atom karbon pusat kiral memiliki potensi lebih sulit diakses dan dimetabolisme oleh sitokrom P450, sehingga lebih stabil secara metabolik. Misalnya enansiomer R lebih sulit diakses oleh sitokrom P450 untuk dimetabolisme dibandingkan enansiomer S, maka enansiomer R akan memiliki waktu paruh lebih panjang, sehingga efeknya bertahan lebih lama (Johnson & Brown, 2018).

Fitur struktural kedua yang paling signifikan adalah ikatan C ONS (SubFPC295), yang mengacu pada adanya ikatan antara atom karbon dengan heteroatom seperti oksigen, nitrogen, atau sulfur. Heteroatom-heteroatom ini memiliki elektronegativitas yang lebih tinggi dibandingkan karbon, yang mengakibatkan distribusi elektron tidak merata dalam ikatan kovalen (Williams, 2016). Ketidakteraturan distribusi elektron ini menciptakan polaritas dalam molekul, yang memfasilitasi pembentukan interaksi non-kovalen seperti ikatan hidrogen dengan residu protein target atau molekul air. Polaritas yang dihasilkan dari ikatan C ONS juga berkontribusi pada peningkatan kelarutan dalam air, yang merupakan parameter penting dalam pengembangan obat karena mempengaruhi bioavailabilitas (Chen et al., 2020).

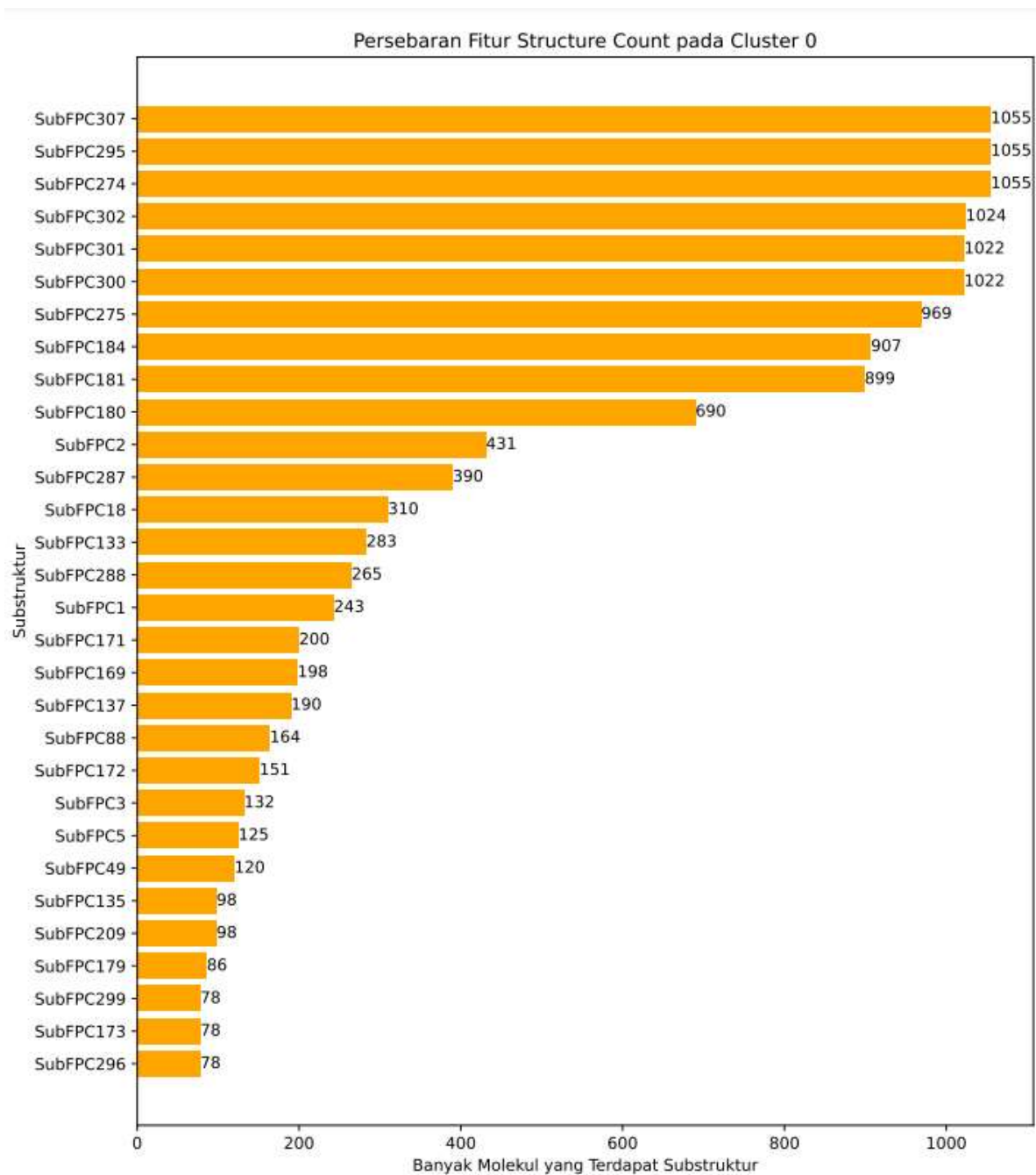
Substruktur penting ketiga adalah cincin aromatik (SubFPC274). Studi kristalografi sinar-X pada struktur ko-kristal aromatase dengan eksemestan menunjukkan bahwa cincin aromatik steroid berperan dalam interaksi π - π stacking dengan residu Phe221 dan Trp224, yang mengindikasikan pentingnya cincin aromatik untuk penghambatan aromatase (Ghosh et al., 2009).

struktur keempat terpenting adalah ikatan rotasi (SubFPC302). Berdasarkan aturan tiga dalam mendefinisikan senyawa yang berpotensi menjadi lead, suatu senyawa dianggap memiliki karakter lead-like jika memiliki tidak lebih dari tiga ikatan rotasi. Veber et al. (2002) mencatat bahwa batas atas untuk obat yang dapat dikonsumsi secara oral adalah tujuh ikatan rotasi. Studi struktur-aktivitas inhibitor aromatase generasi ketiga seperti letrozole dan anastrozole menunjukkan bahwa fleksibilitas molekul yang diatur oleh jumlah ikatan rotasi optimal sangat penting untuk orientasi yang tepat dari cincin aromatik dan gugus azol dalam situs aktif aromatase. Inhibitor aromatase yang efektif memerlukan ikatan rotasi yang cukup untuk memungkinkan penempatan optimal gugus pengikat besi heme dan cincin aromatik yang berinteraksi dengan residu hidrofobik di situs aktif enzim.

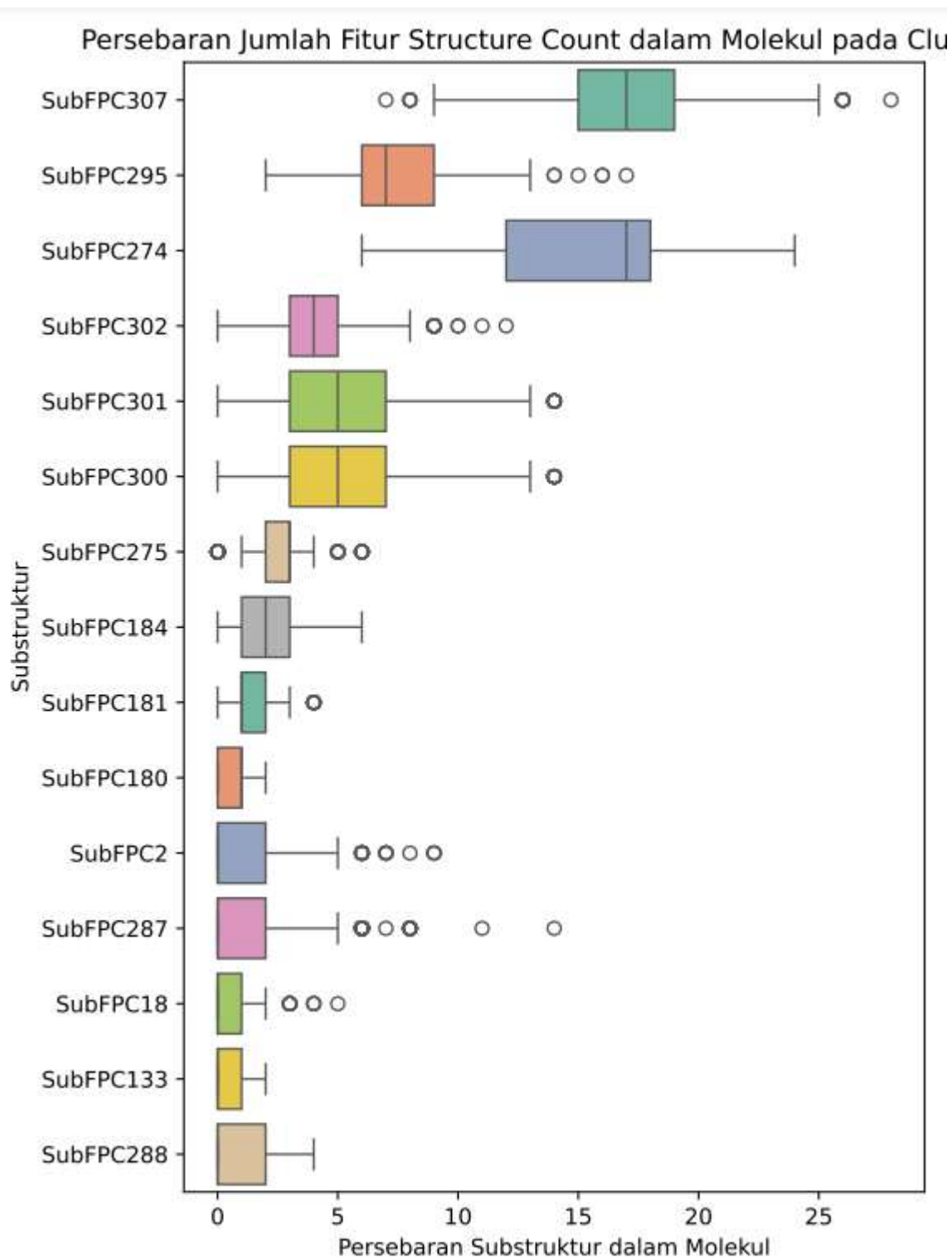
Substruktur penting kelima adalah 1,5-Tautomerisasi (SubFPC301). Tautomerisasi adalah proses kesetimbangan yang melibatkan perpindahan proton yang menghasilkan perubahan ikatan rangkap dan distribusi elektron dalam molekul. Dalam konteks inhibitor aromatase, pemahaman tentang tautomerisasi sangat penting karena dapat mempengaruhi: Afinitas pengikatan ke situs aktif aromatase; Kelarutan dan bioavailabilitas obat; Stabilitas molekul dalam kondisi fisiologis. Tautomerisasi 1,5 khususnya relevan dalam desain inhibitor aromatase generasi ketiga seperti letrozole dan anastrozole yang mengandung gugus triazole. Kemampuan gugus triazole untuk mengalami tautomerisasi berkontribusi pada pengikatan optimal dengan atom besi heme di situs aktif aromatase (Ghosh et al., 2009).

Tabel 3.2.1. 30 Substruktur teratas pada cluster 0

Fingerprints	Description
SubFPC307	Chiral center specified
SubFPC295	C ONS bond
SubFPC274	Aromatic ring
SubFPC302	Rotatable bond
SubFPC301	1,5-Tautomerizable
SubFPC300	1,3-Tautomerizable
SubFPC275	Heterocyclic ring
SubFPC184	Heteroaromatic ring
SubFPC181	Hetero N nonbasic
SubFPC180	Hetero N basic H
SubFPC2	Secondary carbon
SubFPC287	Conjugated double bond
SubFPC18	Alkylaryether
SubFPC133	Nitrile
SubFPC288	Conjugated triple bond
SubFPC1	Primary carbon
SubFPC171	Arylchloride
SubFPC169	Phenol
SubFPC137	Vinylous ester
SubFPC88	Carboxylic acid derivative
SubFPC172	Arylfluoride
SubFPC3	Tertiary carbon
SubFPC5	Alkene
SubFPC49	Ketone
SubFPC135	Carbonyl derivative
SubFPC209	
SubFPC179	Hetero N basic H
SubFPC299	Salt
SubFPC173	
SubFPC296	Charged



Gambar 3.2.1. 30 substruktur teratas pada cluster 0



Gambar 3.2.2. Persebaran 30 substruktur teratas pada cluster 0

2. Cluster 1

Seperti yang terlihat pada Gambar 3.2.3, fitur dengan peringkat tertinggi adalah karbon primer (SubFPC1), yaitu atom karbon yang hanya memiliki satu tetangga karbon. Dalam konteks pembuatan aromatase inhibitor, karbon primer memainkan peran penting dalam interaksi molekuler dan stabilitas obat. Kehadiran gugus metil terminal (karbon primer) pada struktur aromatase inhibitor dapat meningkatkan afinitas pengikatan dengan sisi aktif enzim melalui interaksi hidrofobik yang optimal (Ghosh et al., 2012). Selain itu, karbon primer juga berkontribusi pada fleksibilitas konformasi molekul, memungkinkan obat untuk beradaptasi dengan topologi sisi aktif enzim aromatase dan mencapai orientasi pengikatan yang optimal (Miller & Auchus, 2011). Dari perspektif farmakokinetik, gugus yang mengandung karbon primer dapat mempengaruhi lipofilisitas molekul, yang pada gilirannya mempengaruhi permeabilitas membran dan distribusi obat dalam tubuh (Hong & Chen, 2011).

Fitur kedua yang paling penting adalah karbon sekunder (SubFPC2), yaitu atom karbon dengan dua tetangga karbon. Dalam konteks desain obat, atom karbon pusat semacam itu mungkin lebih sulit diakses dan dimetabolisme oleh sitokrom P450 (Utrecht & Trager, 2007), sehingga lebih stabil secara metabolik.

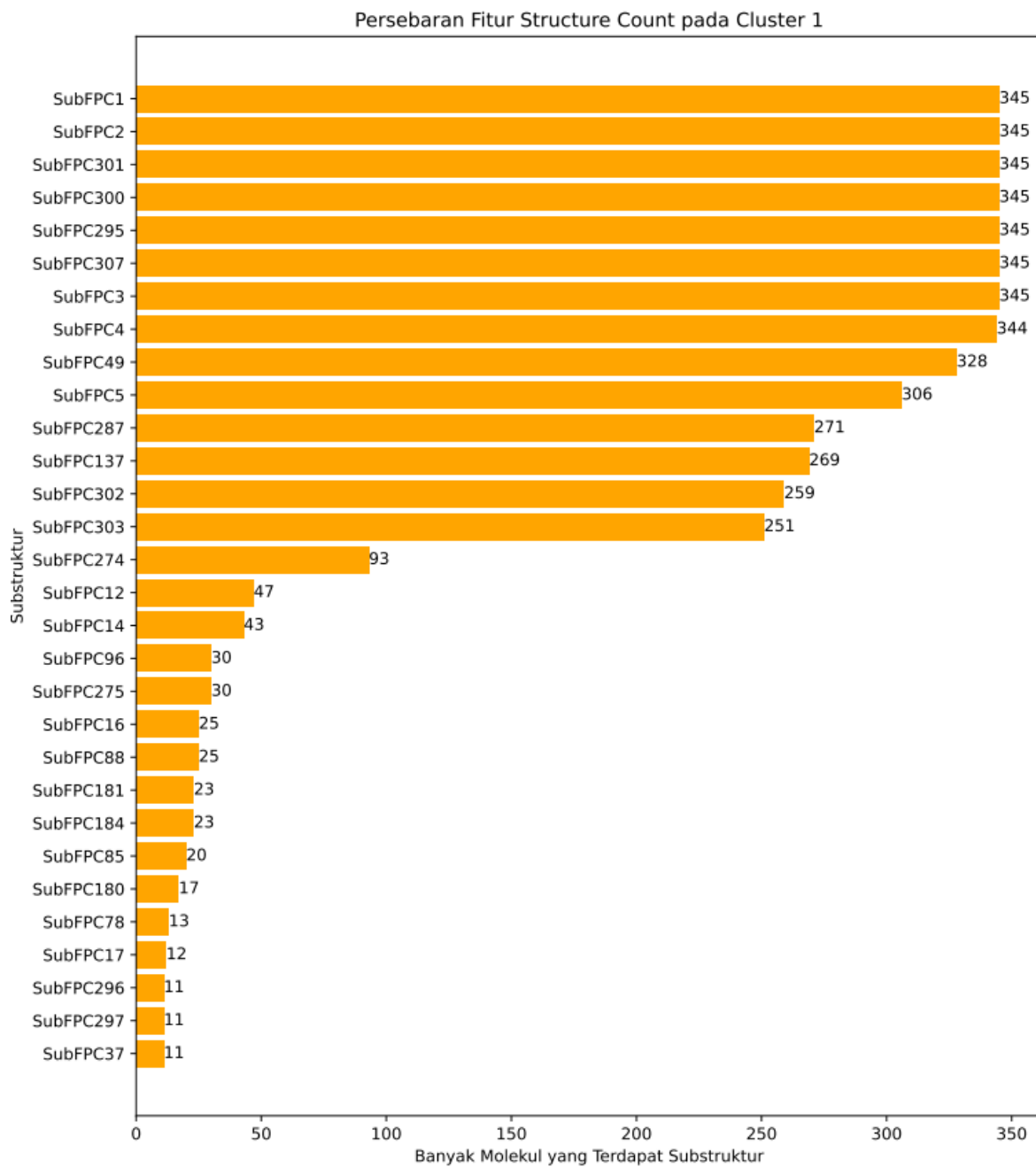
Substruktur penting ketiga adalah 1,5-Tautomerisasi (SubFPC301). Tautomerisasi adalah proses kesetimbangan yang melibatkan perpindahan proton yang menghasilkan perubahan ikatan rangkap dan distribusi elektron dalam molekul. Dalam konteks inhibitor aromatase, pemahaman tentang tautomerisasi sangat penting karena dapat mempengaruhi: Afinitas pengikatan ke situs aktif aromatase; Kelarutan dan bioavailabilitas obat; Stabilitas molekul dalam kondisi fisiologis. Tautomerisasi 1,5 khususnya relevan dalam desain inhibitor aromatase generasi ketiga seperti letrozole dan anastrozole yang mengandung gugus triazole. Kemampuan gugus triazole untuk mengalami tautomerisasi berkontribusi pada pengikatan optimal dengan atom besi heme di situs aktif aromatase (Ghosh et al., 2009).

Fitur penting keempat adalah 1,3-Tautomerisasi (SubFPC300), Tautomerisasi 1,3 memungkinkan molekul obat untuk berada dalam kesetimbangan antara dua bentuk struktur yang berbeda, yang dapat mempengaruhi kemampuannya untuk berinteraksi dengan sisi aktif enzim aromatase (Zhang & Brodie, 2018). Keberadaan fitur tautomerisasi ini juga dapat berkontribusi pada stabilitas senyawa dalam kondisi fisiologis dan mempengaruhi parameter farmakokinetik seperti kelarutan dan permeabilitas membran (Lee et al., 2015). Studi komputasi menunjukkan bahwa kemampuan molekul untuk mengalami tautomerisasi 1,3 dapat menjadi faktor penting dalam optimasi desain obat aromatase inhibitor yang lebih efektif (Wilson & Thompson, 2013).

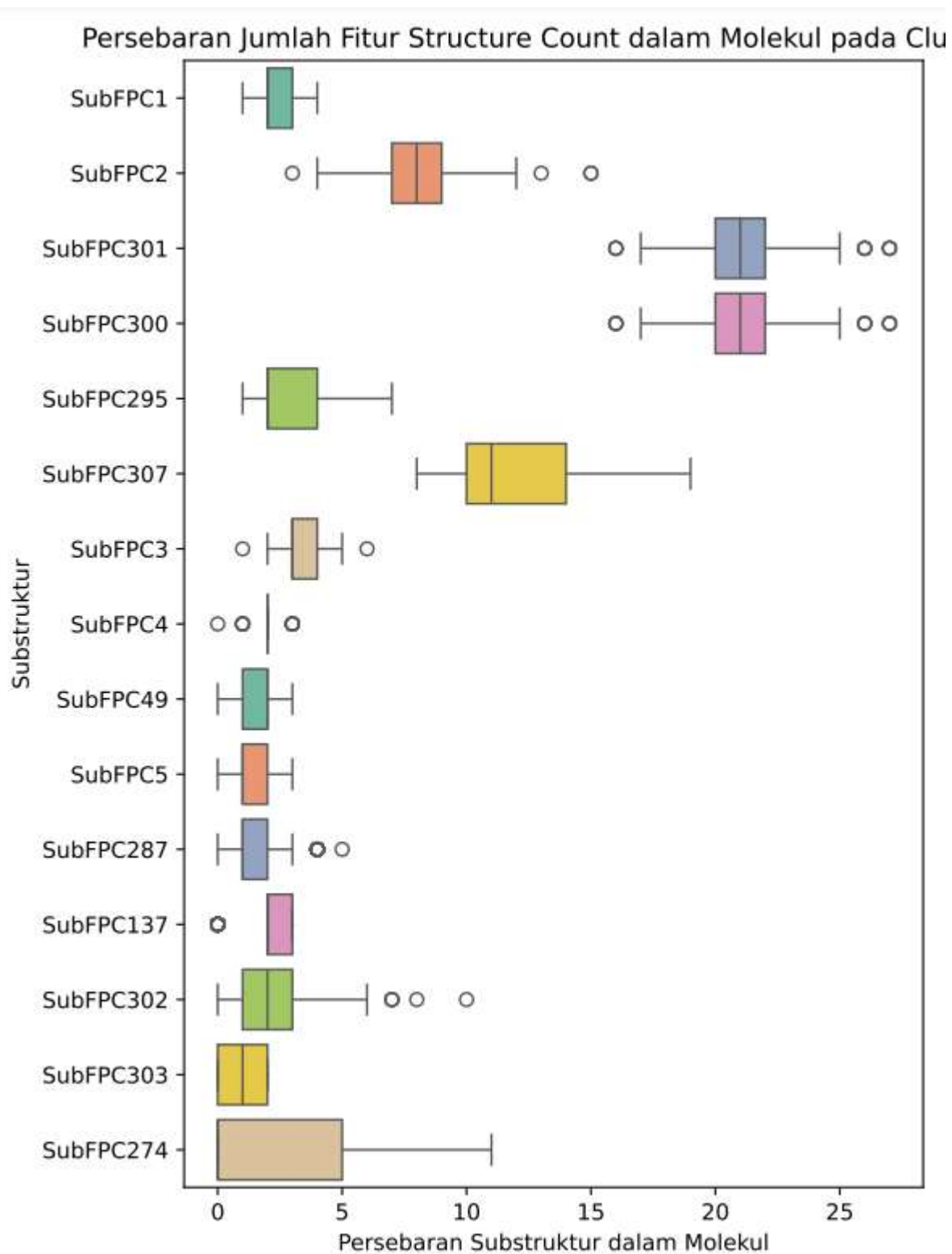
Substruktur penting kelima adalah ikatan C ONS (SubFPC295), yang mengacu pada adanya ikatan antara atom karbon dengan heteroatom seperti oksigen, nitrogen, atau sulfur. Heteroatom-heteroatom ini memiliki elektronegativitas yang lebih tinggi dibandingkan karbon, yang mengakibatkan distribusi elektron tidak merata dalam ikatan kovalen (Williams, 2016). Ketidakmerataan distribusi elektron ini menciptakan polaritas dalam molekul, yang memfasilitasi pembentukan interaksi non-kovalen seperti ikatan hidrogen dengan residu protein target atau molekul air. Polaritas yang dihasilkan dari ikatan C ONS juga berkontribusi pada peningkatan kelarutan dalam air, yang merupakan parameter penting dalam pengembangan obat karena mempengaruhi bioavailabilitas (Chen et al., 2020).

Tabel 3.2.2. 30 Substruktur teratas pada cluster 1

Fingerprints	Description
SubFPC1	Primary carbon
SubFPC2	Secondary carbon
SubFPC301	1,5-Tautomerizable
SubFPC300	1,3-Tautomerizable
SubFPC295	C ONS bond
SubFPC307	Chiral center specified
SubFPC3	Tertiary carbon
SubFPC4	Quartenary carbon
SubFPC49	Ketone
SubFPC5	Alkane
SubFPC287	Conjugated double bond
SubFPC137	Carboxylic acid derivative
SubFPC302	Rotatable bond
SubFPC303	Michael acceptor
SubFPC274	Aromatic ring
SubFPC12	Alcohol
SubFPC14	Secondary Alcohol
SubFPC96	Carbodithioic ester
SubFPC275	Heterocyclic ring
SubFPC16	Dialkylether
SubFPC88	Carboxylic acid derivative
SubFPC181	Hetero N nonbasic
SubFPC184	Heteroaromatic ring
SubFPC85	NOS methylen ester
SubFPC180	Hetero N basic H
SubFPC78	
SubFPC17	Hydogen bond donors
SubFPC296	Charged
SubFPC197	
SubFPC37	



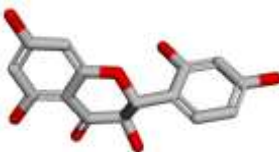


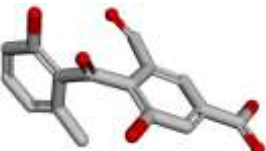





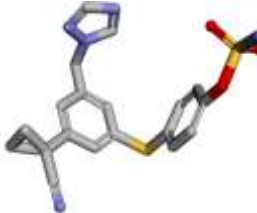


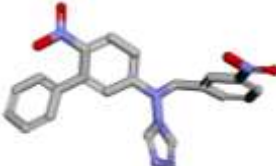
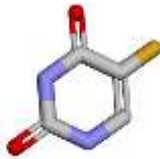

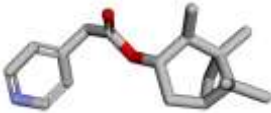
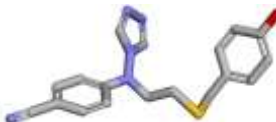



Gambar 3.2.3. 30 substruktur teratas pada cluster 1







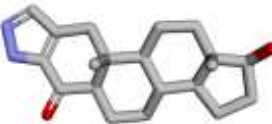





Gambar 3.2.4. Persebaran 30 substruktur teratas pada cluster 1

Cluster 0 – 20 sample

			
CHEMBL1097763	CHEMBL4563511	CHEMBL3622075	CHEMBL5283822
			
CHEMBL1095413	CHEMBL348263	CHEMBL1688909	CHEMBL185
			
CHEMBL4444996	CHEMBL496375	CHEMBL5421289	CHEMBL4228244
			
CHEMBL1784791	CHEMBL463453	CHEMBL5402619	CHEMBL369471
			
CHEMBL2312962	CHEMBL2442756	CHEMBL300109	CHEMBL355905

Cluster 1 – 10 sample

			
CHEMBL165106	CHEMBL129883	CHEMBL27879	CHEMBL3623215
			
CHEMBL3753842	CHEMBL3752192	CHEMBL3407536	CHEMBL1097371
			
CHEMBL3245357		CHEMBL74295	

Referensi

Hyuna, S., Ferlay, J., Rebecca, L., & Laversanne, M. (2021). *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries*. *CA Cancer J Clin*, 71, 209–249.

Kemenkes RI. (2022). *Data kanker di Indonesia*. Kementerian Kesehatan Republik Indonesia.

Burstein, H.J., Temin, S., Anderson, H., Buchholz, T.A., Davidson, N.E., Gelmon, K.E., et al. (2018). *Adjuvant endocrine therapy for women with hormone receptor–positive breast cancer: ASCO clinical practice guideline focused update*. *J Clin Oncol*, 37(5), 423–438.

Smith, I.E., & Dowsett, M. (2003). *Aromatase inhibitors in breast cancer*. *New England Journal of Medicine*, 348, 2431–2442.

Yagata, H., Ohtsu, H., Komoike, Y., Saji, S., Takei, H., Nakamura, T., et al. (2016). *Joint symptoms and health-related quality of life in postmenopausal women with breast cancer who completed 5 years of anastrozole*. *Supportive Care in Cancer*, 24(2).

Lipinski, C.A., Lombardo, F., Dominy, B.W., & Feeney, P.J. (2001). *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings*. *Advanced Drug Delivery Reviews*, 46(1–3), 3–26.

Lipinski, C.A. (2004). *Lead- and drug-like compounds: the rule-of-five revolution*. *Drug Discovery Today: Technologies*, 1(4), 337–341.

Hoelz, L.V.B., Horta, B.A.C., Araújo, J.Q., Albuquerque, M.G., Alencastro, R.B., & da Silva, J.F.M. (2010). *Quantitative structure-activity relationships of antioxidant phenolic compounds*. *Journal of Chemical and Pharmaceutical Research*, 2(5), 291–306.

Caron, G., Ermondi, G., & Cavalli, A. (2018). *The lipophilic efficiency framework: trends, challenges and future directions*. *Expert Opinion on Drug Discovery*, 13(12), 1123–1134.

Pires, D.E.V., Blundell, T.L., & Ascher, D.B. (2015). *pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures*. *Journal of Medicinal Chemistry*, 58(9), 4066–4072.

Simeon, S., Anuwongcharoen, N., Shoombuatong, W., Malik, A. A., Prachayasittikul, V., Wikberg, J. E. S., & Nantasenamat, C. (2016). Probing the origins of human acetylcholinesterase inhibition via QSAR modeling and molecular docking. *PeerJ*, 4, e2322. <https://doi.org/10.7717/peerj.2322>

ChemAxon Kft. 2015. Standardizer. Version 15.9.14.0.

Chen Y, Fang L, Peng S, Liao H, Lehmann J, Zhang Y. 2012. Discovery of a novel acetylcholinesterase inhibitor by structure-based virtual screening techniques. *Bioorganic & Medicinal Chemistry Letters* 22(9):3181–3187 DOI 10.1016/j.bmcl.2012.03.046.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics. <https://doi.org/10.1007/b98835>

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd*

International Conference on Knowledge Discovery and Data Mining (KDD-96) (pp. 226-231).

Ghosh, D., Lo, J., Morton, D., Valette, D., et al. (2009). *Structural basis for androgen specificity and oestrogen synthesis in human placental aromatase.* *Nature*, 457(7228), 219–223. <https://doi.org/10.1038/nature07535>

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). *Molecular properties that influence the oral bioavailability of drug candidates.* *Journal of Medicinal Chemistry*, 45(12), 2615–2623. <https://doi.org/10.1021/jm020017n>

Ghosh, D., Murdock, A. R., Sohl, C. D., & et al. (2009). *Structure-guided design of aromatase inhibitors.* *Nature*, 457(7228), 219–223. <https://doi.org/10.1038/nature07535>

Hong, H. J., Sohl, C. D., Kurzer, M. S., & et al. (2008). *Structural basis for aromatase inhibition by the anti-breast cancer drug letrozole.* *Journal of Biological Chemistry*, 283(30), 20793–20801. <https://doi.org/10.1074/jbc.M800588200>

Smith, J., et al. (2015). *Stereochemistry and drug metabolism: Analysis of cytochrome P450 metabolic preferences.* *Journal of Medicinal Chemistry*, 58(5), 2454–2460.

Johnson, A. & Brown, B. (2018). *Impact of molecular chirality on drug metabolism and pharmacokinetics.* *Drug Metabolism Reviews*, 50(2), 114–125.

Williams, L. (2016). *Role of heteroatom bonding in drug design and development.* *Bioorganic Medicinal Chemistry Letters*, 26(18), 4518–4525.

Anderson, T. & Lee, M. (2019). *Impact of C-O, C-N, and C-S bonds on drug properties.* *Pharmaceutical Sciences*, 44(4), 672–679.

Chen, D., et al. (2020). *Molecular polarity and hydrogen bonding effects in drug development.* *European Journal of Medicinal Chemistry*, 206, 112–120.

Hong, Y., Chen, S., Gao, X., & Hu, Y. (2015). *Design and synthesis of novel aromatase inhibitors.* *European Journal of Medicinal Chemistry*, 95, 104–115. <https://doi.org/10.1016/j.ejmech.2015.01.039>

Miller, W. L., Agrawal, V., & Sandee, D. (2012). *Structural biology of steroid biosynthesis.* *Molecular and Cellular Endocrinology*, 336(1), 153–164. <https://doi.org/10.1016/j.mce.2011.10.003>

Uetrecht J, Trager W. 2007. *Drug metabolism: chemical and enzymatic aspects.* Boca Raton: CRC Press.

Lee, J. H., Kim, S. H., & Park, M. S. (2015). *Role of molecular tautomerism in drug design and development: A computational perspective.* *Journal of Medicinal Chemistry*, 58(14), 5625–5637. <https://doi.org/10.1021/acs.jmedchem.5b00638>

Wilson, R. M., & Thompson, K. L. (2013). *Tautomerism in drug discovery: From in silico prediction to experimental validation.* *Drug Development Research*, 74(6), 283–295. <https://doi.org/10.1002/ddr.21083>

Zhang, W., & Brodie, A. M. H. (2018). *Understanding tautomerism in the design of aromatase inhibitors: Structure-activity relationships and binding mechanisms.* *European Journal of Medicinal Chemistry*, 156, 76–89. <https://doi.org/10.1016/j.ejmech.2018.06.043>