

## Sidik Jari Interaksi Protein-Ligan Ansambel pada Konstruksi dan Validasi Protokol Penapisan Virtual Bertargetkan Reseptor Kemokine C-X-C Tipe 4

### (Ensemble Protein-Ligand Interaction Fingerprints in Construction and Validation of Virtual Screening Protocol Targeting C-X-C Chemokine Receptor Type 4)

ENADE PERDANA ISTYASTONO<sup>1,2,\*</sup>, MUHAMMAD RADIFAR<sup>2</sup>

<sup>1</sup>Divisi Rancangan dan Penemuan Obat, Fakultas Farmasi, Universitas Sanata Dharma, Kampus 3 Paingan, Depok, Sleman, Yogyakarta 55282 - Indonesia

<sup>2</sup>Laboratorium Kimia Komputasi MOLMOD.ORG, Pogung Kidul, Sinduadi, Mlati, Sleman, Yogyakarta 55284 - Indonesia

Diterima 7 Desember 2016, Disetujui 24 Februari 2017

**Abstrak:** Protokol Penapisan Virtual Berbasis Struktur (PVBS) bertargetkan reseptor kemokine C-X-C tipe 4 (CXCR4) telah dikonstruksi dengan menggunakan PLANTS 1.2 untuk penambatan molekuler dan PyPLIF 0.1.1 untuk identifikasi Sidik jari Interaksi Protein-Ligan (SIPL). Penggunaan skor ChemPLP bawaan PLANTS 1.2 dan nilai Tc-PLIF bawaan dari PyPLIF 0.1.1 untuk memilih pose terbaik dalam PVBS retrospektif memberikan nilai Faktor Pengayaan (FP) di bawah nilai FP referensi (17,5). Meskipun demikian, dalam PVBS retrospektif juga dihasilkan SIPL untuk masing-masing pose hasil penambatan molekuler. Dalam penelitian yang dipaparkan dalam artikel ini, diujicobakan analisis Hubungan Kuantitatif Struktur-Aktivitas (HKSA) biner dengan prediktor baru yaitu SIPL ansambel hasil PVBS untuk menggantikan SIPL dari pose terbaik menurut skor ChemPLP maupun menurut nilai Tc-PLIF. SIPL ansambel dihitung dengan memperhitungkan seluruh pose yang memiliki skor ChemPLP di bawah skor ChemPLP tertentu dan untuk masing-masing *bitstring* SIPL dihitung persentase interaksi "on". Penentuan skor ChemPLP tertentu ini dilakukan secara sistematis untuk mendapatkan skor ChemPLP yang memberikan nilai *F-score* dan *Matthews correlation coefficient* (MCC) paling tinggi. Hasil penelitian menunjukkan nilai *F-score* dan MCC dari analisis HKSA biner dengan SIPL ansambel hasil PVBS sebagai prediktor mampu mencapai nilai 0,58 dan 0,61 dengan nilai FP mencapai 323,57, jauh lebih baik dari nilai FP protokol referensi.

**Kata kunci:** penapisan virtual berbasis struktur (PVBS), hubungan kuantitatif struktur-aktivitas (HKSA) biner, sidik jari interaksi protein-ligan (SIPL) ansambel, reseptor kemokine c-x-c tipe 4 (CXCR4).

**Abstract:** Structure-Based Virtual Screening (SBVS) protocols targeting C-X-C chemokine receptor type 4 (CXCR4) have been constructed by employing PLANTS 1.2 to perform molecular docking simulations and PyPLIF 0.1.1 to identify Protein-Ligand Interaction Fingerprint (PLIF). By using ChemPLP score from PLANTS 1.2 and Tc-PLIF from PyPLIF 0.1.1 to select best pose in the retrospective SBVS showed Enrichment Factor (EF) values of less than the EF value of the reference SBVS (17.5). Nevertheless, the retrospective SBVS campaigns have also resulted in PLIF bitstrings for all poses resulted from the molecular docking simulations. In this article, binary Quantitative Structure-Activity Relationship (QSAR) analysis employing new predictors ensemble PLIF resulted from the retrospective SBVS campaigns, instead of using PLIF bitstrings from the best pose only, are presented. The ensemble PLIF as predictors were calculated by taking into account all poses with ChemPLP score lower than a certain ChemPLP score as the predefined cutoff, and subsequently for every compound the percentage of "on" interactions was calculated for every PLIF bitstring. The predefined cutoff was selected by

\* Penulis korespondensi, Hp. 081510704594  
e-mail: enade@usd.ac.id

performing systematic trials to obtain a ChemPLP score as the cutoff with the highest F-score and Matthews correlation coefficient (MCC) value. The results showed that the F-score and MCC values could reach 0.58 and 0.61, respectively with EF value of 323.47, which was much better than the EF value of the reference SBVS protocol.

**Keywords:** structure-based virtual screening (SBVS), binary quantitative structure-activity relationship, ensemble protein-ligand interaction fingerprint (PLIF), C-X-C chemokine receptor type 4 (CXCR4).

## PENDAHULUAN

RESEPTOR kemokine C-X-C tipe 4 (CXCR4) merupakan target penemuan obat yang bertanggung jawab dalam inflamasi, *cancer metastasis*, regulasi migrasi sel sistem imun, dan infeksi HIV-1<sup>(1,2)</sup>. Di bidang pemodelan molekul dan penemuan obat berbantuan komputer, struktur CXCR4 sebagai target penemuan obat merupakan target yang menantang<sup>(2-4)</sup>. Pada usaha global melakukan prediksi interaksi ligan pada kristal-kristal baru *G-protein coupled receptor* (GPCR) di tahun 2010 yang melibatkan 35 grup riset didapati total 25 grup riset mengusulkan 103 model CXCR4 dengan ligan IT1t<sup>(2)</sup>. Hanya satu model dari 103 model tersebut yang memiliki nilai *z-score* di atas 2,0 dengan akurasi kontak ligan dengan protein sebesar 36%<sup>(1,2,4)</sup>. Hal ini menggambarkan tantangan yang dihadapi dengan pemilihan CXCR4 sebagai target penemuan obat<sup>(4)</sup>. Meskipun demikian, pada Penapisan Virtual Berbasis Struktur (PVBS) retrospektif menggunakan *Database of Useful Decoys-Enhanced* (DUD-E) dengan *software* DOCK 3.6 berhasil diperoleh nilai *Enrichment Factor* (Faktor Pengayaan (FP)) sebesar 17,5<sup>(5)</sup>.

Model yang paling akurat dalam usaha global prediksi interaksi IT1t dengan CXCR4 pada tahun 2010 menggunakan fungsi obyektif alternatif sebagai substitusi maupun komplementer dari skor penambatan molekuler berupa sidik jari interaksi yang dikembangkan oleh Marcou dan Rognan pada tahun 2007<sup>(2,4,6)</sup>. Penggunaan sidik jari interaksi sebagai fungsi obyektif alternatif terbukti mampu meningkatkan kualitas PVBS retrospektif untuk identifikasi ligan pada reseptor histamin H1 (HRH1) dan H4 (HRH4) dengan nilai FP mencapai 57,6 untuk HRH1 dan 40,3 untuk HRH4<sup>(7,8)</sup>. Pada PVBS prospektif untuk HRH1 dan HRH4 secara berturut diusulkan 26 dan 37 senyawa baru untuk diverifikasi secara *in vitro* setelah menapis sekitar 13 juta senyawa, dengan hasil verifikasi diperoleh 19 senyawa aktif dengan *Ki* antara 6 - 10000 nM untuk HRH1 dan 9 senyawa aktif dengan nilai *Ki* antara 140 - 6900 nM untuk HRH4<sup>(7,8)</sup>. Hasil ini menginspirasi untuk mengembangkan *software* identifikasi sidik jari interaksi yang ditulis Marcou dan Rognan menggunakan bahasa pemrograman C++ dengan pustaka berbayar untuk ditulis ulang

dengan bahasa pemrograman Python dengan pustaka yang disediakan gratis dan berlisensi *open source* yang diberi nama PyPLIF<sup>(6,9-11)</sup>. PyPLIF yang saat ini memiliki versi 0.1.1 (PyPLIF 0.1.1) secara khusus dikembangkan untuk membaca luaran hasil simulasi penambatan molekuler menggunakan PLANTS<sup>(10,11)</sup>.

Pada awalnya PyPLIF dikembangkan untuk mendukung penelitian untuk rancangan dan penemuan obat baru bertargetkan reseptor estrogen alfa (ER $\alpha$ )<sup>(10,12)</sup>. Pemanfaatan PyPLIF mampu meningkatkan nilai FP dari PVBS reguler, yaitu dari 21,2 menjadi 53,8<sup>(10,12)</sup>. Secara prospektif, protokol PVBS yang dikembangkan berhasil mengidentifikasi analog eugenol yang optimal sebagai ligan pada ER $\alpha$  dan mengidentifikasi inhibitor COX-2 celecoxib sebagai ligan aktif juga pada ER $\alpha$ <sup>(10,12-14)</sup>. Pada perkembangan selanjutnya, dengan memanfaatkan data di DUD-E dilakukan pemutakhiran PVBS retrospektif bertargetkan ER $\alpha$  dan diperoleh FP sebesar 18,5 dengan PVBS reguler menggunakan skor ChemPLP bawaan PLANTS 1.2<sup>(15)</sup>. PyPLIF 0.1.1 selanjutnya digunakan untuk melakukan identifikasi Sidik jari Interaksi Protein-Ligan (SIPL) pada pose terbaik menurut skor ChemPLP dan hasilnya berupa *bitstring* SIPL digunakan sebagai prediktor pada Hubungan Kuantitatif Struktur-Aktivitas (HKSA) biner dengan metode Recursive Partition And Regression Tree (RPART)<sup>(16,17)</sup>. Pendekatan tersebut mampu meningkatkan nilai FP dari 18,5 menjadi 247,9<sup>(15,16)</sup>. PyPLIF juga digunakan untuk target lain yaitu enzim siklooksigenase-1 (COX-2)<sup>(11)</sup>, sitokrom P450 3A4 (CYP3A4)<sup>(18)</sup>, reseptor adrenergik  $\beta_2$  (ADRB2)<sup>(19)</sup> dan enzim histon deasetilase 2 (HDAC2)<sup>(20)</sup>.

Penelitian yang dipaparkan dalam artikel ini bertujuan untuk menguji kemampuan PyPLIF dalam meningkatkan kualitas prediksi PVBS bertargetkan CXCR4. Protokol referensi memberikan nilai FP sebesar 17,5<sup>(5)</sup>. Diawali dengan PVBS reguler yang dilanjutkan dengan analisis HKSA biner metode RPART pada *bitstring* SIPL pose-pose terbaik<sup>(16,17)</sup>. Pada penelitian ini selanjutnya diujicobakan strategi baru yang mempertimbangkan *bitstring* SIPL seluruh pose hasil penambatan molekuler yang memiliki skor ChemPLP lebih baik dari suatu nilai tertentu. Persentase *bitstring* SIPL yang teridentifikasi berikatan

dari seluruh pose luaran simulasi penambatan molekuler dihitung sebagai SIPL ansambel, yang kemudian digunakan sebagai prediktor pada HKSA biner metode RPART. Pemilihan skor ChemPLP sebagai *cutoff* dipilih secara sistematis dan dipilih skor ChemPLP yang memberikan nilai *F-score* dan MCC terbaik<sup>(21,22)</sup>.

## BAHAN DAN METODE

**ALAT DAN BAHAN.** Struktur kristal yang diunduh dari *Protein Data Bank* (PDB) dengan kode PDB 3ODU<sup>(1)</sup> digunakan sebagai referensi struktur target. Ligan-ligan CXCR4 sejumlah 40 dan pengecohnya sejumlah 3406 yang disediakan di basis data DUD-E<sup>(5)</sup> digunakan untuk melakukan validasi retrospektif protokol PVBS yang dikembangkan. Aplikasi-aplikasi kimia medisinal komputasi yang digunakan dalam penelitian ini adalah SPORES<sup>(23)</sup>, PLANTS1.2<sup>(24,25)</sup>, *Open Babel* 2.2.3<sup>(26)</sup> dan PyPLIF 0.1.1<sup>(10)</sup>. Analisis HKSA biner dilakukan menggunakan paket “*rpart*” pada *software* statistika komputasi R versi 3.2.3 (R-3.2.3)<sup>(17,27)</sup>. Seluruh komputasi dilakukan di server beralamat *Internet Protocol* (IP) 172.23.1.42 di jaringan internet Universitas Sanata Dharma dengan sistem operasi Linux distribusi Ubuntu 12.04.3 LTS “*Precise Pangolin*”, prosesor *Quad Core* Intel(R) Xeon(R) CPU E31220 @ 3.10GHz dan *random access memory* (RAM) 8 GB.

**METODE.** Mengacu pada prosedur yang dipublikasikan Istyastono dan Setyaningsih dilakukan PVBS retrospektif untuk mengenali ligan-ligan CXCR4 di antara pengecoh-pengecohnya<sup>(5,19)</sup>. Struktur kristal CXCR4 dengan kode PDB 3ODU diunduh dari situs PDB<sup>(1)</sup>. Hanya rantai A dari struktur kristal yang digunakan lebih lanjut. Modul *splitpdb* di SPORES digunakan untuk memisah struktur reseptor, ligan kokristal, dan molekul-molekul air yang ada pada berkas *3odu.pdb* dan kemudian mengubah menjadi berkas-berkas *mol2* yang siap digunakan dalam simulasi penambatan molekuler menggunakan *software* PLANTS1.2 Langkah ini menghasilkan target virtual *protein.mol2* dan ligan kokristal *ligand\_ITD1500\_0.mol2*.

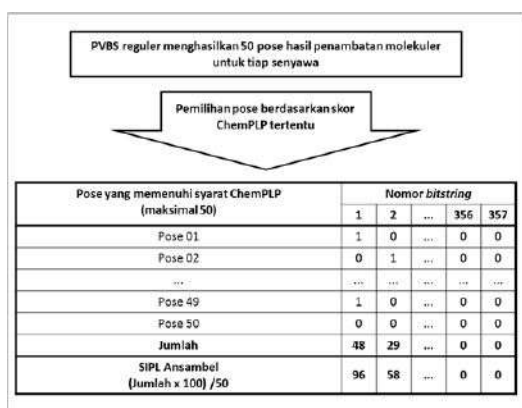
Ligan-ligan aktif pada CXCR4 dengan nilai  $IC_{50}$  kurang dari 1  $\mu$ M sejumlah 40 dan pengecoh-pengecohnya diunduh dalam format SMILES format dari situs DUD-E<sup>(5)</sup> yang disimpan di server lokal dengan nama *actives\_final.ism* dan *decoys\_final.ism*. Berkas-berkas tersebut kemudian disatukan dalam sebuah berkas dengan nama *all.smi*. Setiap senyawa dalam berkas tersebut dikonversi menjadi format tiga dimensi (3D) pada pH 7,4 sebagai berkas *mol2* menggunakan *software* Open Babel 2.2.3, Modul

*set types* pada SPORES kemudian digunakan untuk memeriksa dan memperbaiki berkas-berkas *mol2* supaya layak ditambatkan menggunakan PLANTS1.2. Sejumlah 50 pose dihasilkan untuk setiap senyawa yang ditambatkan menggunakan fungsi skor ChemPLP pada aturan kecepatan 2. Kantung ikatan CXCR4 didefinisikan sebagai residu-residu asam amino berjarak 5 Å dari ligan referensi (ligan kokristal). Pusat kantung ikatan adalah titik tengah dari ligan referensi. Opsi-opsi lain digunakan dalam pengaturan bawaan PLANTS1.2<sup>(11)</sup>. Setiap senyawa ditambatkan secara independen 3 kali. Dari tiga kali simulasi independen ini diambil simulasi yang menghasilkan pose dengan nilai ChemPLP terbaik untuk diproses lebih lanjut.

Moda ikatan ligan kokristal pada struktur kristal CXCR4 digunakan untuk menghasilkan SIPL referensi menggunakan PyPLIF. Tujuh tipe ikatan (ikatan hidrofobik, ikatan aromatis *face-to-face*, ikatan aromatis *edge-to-face*, ikatan hidrogen dengan residu asam amino sebagai donor, ikatan hidrogen dengan residu asam amino sebagai akseptor, ikatan ionik dengan residu asam amino sebagai kation, dan ikatan ionik dengan residu asam amino sebagai anion) digunakan untuk mendefinisikan SIPL<sup>(9,10)</sup>. Residu-residu yang terlibat dalam identifikasi SIPL mengacu pada kantung ikatan yang digunakan pada penambatan molekuler<sup>(11,19)</sup>.

Pose hasil penambatan dengan skor ChemPLP terbaik dan dengan koefisien Tanimoto kemiripan dengan ligan referensi terbaik (Tc-SIPL) dipilih untuk setiap senyawa yang ditapis. Set data berisi daftar pose terbaik menurut skor ChemPLP dan nilai Tc-SIPL disandikan dalam berkas *def.chemplp.lst* dan *def.ifp.lst*. Data-data tersebut kemudian diurutkan berdasarkan skor ChemPLP maupun Tc-SIPL. Ligan disandikan sebagai Positif (P) sementara pengecoh disandikan sebagai Negatif (N). Dari data yang sudah diurutkan hanya senyawa berlokasi di atas senyawa N urutan 35 (1% dari semua pengecoh) yang dipilih untuk dianalisis lebih lanjut<sup>(7)</sup>. Berdasarkan data terpilih dibuat *confusion matrix* (2 x 2) yang terdiri dari *True Positives* (TPs), *False Negatives* (FNs), *True Negatives* (TNs), dan *False Positives* (FPs)<sup>(21)</sup>. Nilai Faktor Pengayaan (FP) kemudian dihitung  $(FP = (TPs/P)/(FPs/N))^{(7)}$  dan dibandingkan dengan nilai FP protokol referensi ( $FP = 17,5$ )<sup>(5)</sup>.

Setiap simulasi penambatan molekuler bertargetkan CXCR4 menghasilkan 50 pose untuk setiap senyawa yang sukses ditambatkan. Untuk setiap senyawa dipilih pose-pose yang memiliki nilai ChemPLP di bawah nilai tertentu (nilai *cutoff*), dan dari pose-pose yang terpilih dilakukan perhitungan SIPL ansambel seperti yang dicontohkan secara skematis pada Gambar 1.



**Gambar 1.** Skema perhitungan SIPL ansambel, yaitu persentase jumlah interaksi “on” dibagi jumlah pose hasil penambatan molekuler. Dalam penelitian jumlah pose hasil penambatan molekuler adalah 50.

Nilai *cutoff* dianalisis secara sistematis dimulai dari skor ChemPLP -25 hingga -125 dengan resolusi maksimal 1 unit skor ChemPLP dan dipilih nilai *cutoff* yang memberikan pohon keputusan (*decision tree*) hasil analisis HKSA biner metode RPART dengan nilai *F-score* dan MCC terbaik<sup>(17,21,22)</sup>. *Shell script* untuk perhitungan SIPL ansambel disajikan pada Gambar 2.

```

1  for k in $(seq 1 357);
2  do
3  echo "awk -F, '{print \$\$k}' pyplif.csv > $k.csv" > $k.sh
4  chmod u+x $k.sh
5  ./$k.sh
6  rm $k.sh
7  positif=$(grep "1" $k.csv | wc -l)
8  echo "python -c 'print \$positif/50.0'" > calc.sh
9  chmod u+x calc.sh
10 ./calc.sh > $k.calc
11 rm $k.csv calc.sh
12 done

```

**Gambar 2.** Contoh *Shell Script* untuk perhitungan SIPL ansambel.

## HASIL DAN PEMBAHASAN

Penelitian ini bertujuan menguji efektivitas SIPL ansambel dalam meningkatkan kualitas prediksi PVBS dengan target yang relatif sulit CXCR4<sup>(2-5)</sup>. Validasi retrospektif PVBS reguler menggunakan PLANTS1.2 yang dilanjutkan dengan PyPLIF<sup>(15,19)</sup> bertargetkan CXCR4<sup>(1)</sup> memanfaatkan data ligan-ligan CXCR4 dan pengecohnya yang disediakan DUD-E<sup>(5)</sup> menghasilkan 516600 pose hasil penambatan. Diketahui ada 300 pose hasil penambatan 300 yang hilang karena 2 dari 3406 pengecoh tidak lolos

proses preparasi menggunakan SPORES. Hal ini mengindikasikan bahwa sejak proses preparasi sudah mampu membantu mengurangi jumlah pengecoh dalam proses penapisan virtual. Dengan menggunakan skor ChemPLP untuk memilih dan mengurutkan pose terbaik hasil penambatan<sup>(24)</sup> didapati bahwa tidak satu pun ligan ditemukan pada set data yang mengandung 35 pengecoh dengan skor ChemPLP terbaik. Oleh karena itu nilai FP dari set data *def.chemplp.lst* adalah 0, hasil dari 0 TPs dan 35 FPs. Dengan demikian, protokol ini tidak layak untuk digunakan lebih lanjut.

Skor penambatan molekuler alternatif dalam penelitian ini adalah nilai Tc-SIPL yang bergantung pada kantung ikatan dan ligan referensi<sup>(10,11)</sup>. Kantung ikatan pada penelitian ini terdiri dari 51 residu asam amino, yaitu: ARG30, GLU32, ASN37, LYS38, LEU41, PRO42, TYR45, PHE87, THR90, LEU91, PHE93, TRP94, ALA95, VAL96, ASP97, ALA98, VAL99, ALA100, ASN101, TRP102, PHE104, LEU108, CYS109, LYS110, VAL112, HIS113, TYR116, THR117, ASP171, ALA175, ASN176, VAL177, SER178, ARG183, TYR184, ILE185, CYS186, ASP187, ARG188, PHE189, TYR190, TYR255, ILE259, HIS281, LYS282, ILE284, SER285, ILE286, GLU288, ALA289, dan PHE292<sup>(1,2,28)</sup>. Secara keseluruhan dihasilkan *bitstring* SIPL sejumlah 184426200 untuk 516600 pose hasil penambatan karena setiap residu asam amino menghasilkan 7 *bitstring* SIPL. *Bitstring-bitstring* ini kemudian digunakan sebagai prediktor dalam analisis HKSA biner. Protokol reguler menggunakan Tc-SIPL sebagai fungsi obyektif memilih dan mengurutkan pose hasil penambatan mampu mengidentifikasi hanya satu ligan pada set data dengan 35 pengecoh (*def.ifp.lst*). Set data ini memiliki nilai FP 2,5 hasil dari 1 TP and 35 FP. Nilai ini masih terlalu rendah dibandingkan dengan nilai FP referensi 17,5<sup>(5)</sup>.

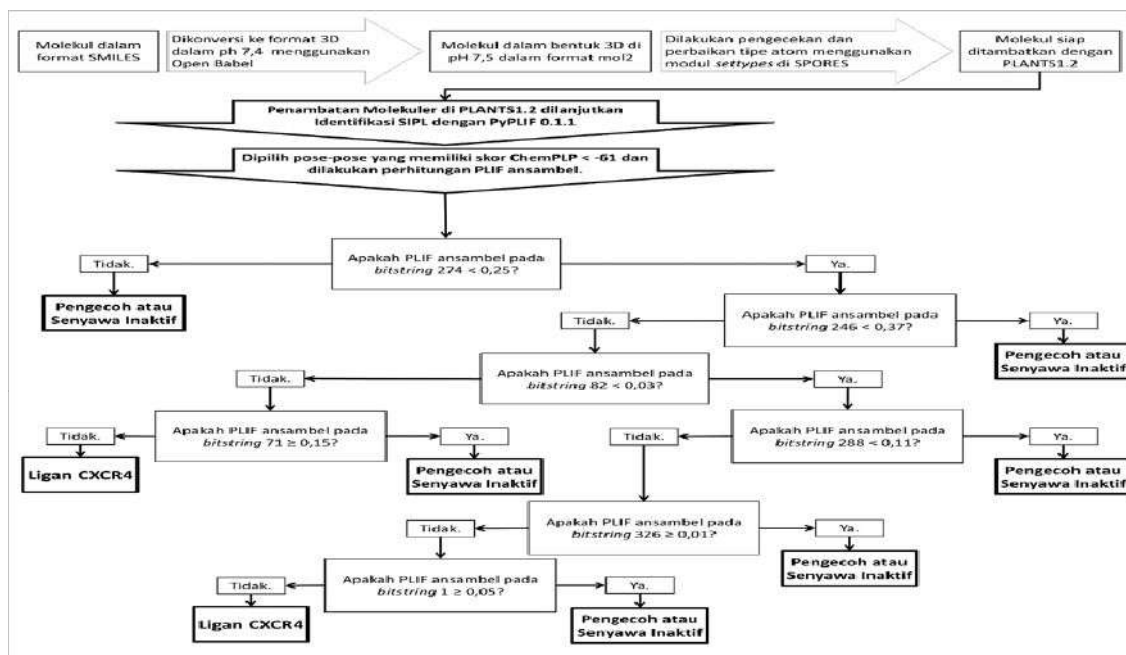
*Bitstring* SIPL hasil dari PVBS ini dapat digunakan untuk meningkatkan kualitas PVBS. Patut dicatat, meskipun nilai FP PVBS menggunakan Tc-SIPL masih tidak dapat digunakan, nilai FP PVBS ini masih lebih baik dari nilai FP PVBS menggunakan skor ChemPLP. Hasil ini sesuai dengan beberapa PVBS pada penelitian-penelitian sebelumnya yang menggunakan sidik jari interaksi sebagai fungsi obyektif<sup>(8,9)</sup>. Namun demikian, berbeda dengan analisis HKSA biner metode RPART pada *bitstring* SIPL pose-pose terbaik sebagai prediktor bertargetkan ER $\alpha$ <sup>(16,17)</sup>, dengan CXCR4 sebagai target tidak dapat dikonstruksi pohon keputusan sehingga berturut-turut nilai TPs, FNs, TNs, dan FPs adalah 0, 40, 3406, dan 0. Oleh karena itu dibutuhkan alternatif prediktor lain yang mampu meningkatkan kualitas PVBS bertargetkan CXCR4. Dalam penelitian ini

diujicobakan SIPL ansambel sebagai prediktor. Permasalahan yang kemudian muncul jika hendak menggunakan SIPL ansambel adalah pemilihan pose luaran dari PLANTS1.2 karena algoritma stokastik yang digunakan berakibat pada pose-pose awal dari seluruh 50 pose yang dihasilkan merupakan coba-coba untuk pemilihan pose berikutnya yang memberikan skor ChemPLP lebih baik<sup>(25)</sup>. Untuk mengatasi hal tersebut maka perlu dipilih satu nilai skor ChemPLP sebagai *cutoff*<sup>(7,20)</sup>.

Hasil pemilihan nilai skor ChemPLP sebagai *cutoff* untuk perhitungan SIPL ansambel sebagai prediktor pada konstruksi pohon keputusan dilakukan diawali dari skor ChemPLP -25, -50, -75, -100, dan -125 (per 25 unit) dengan hasil bahwa nilai *F-score* dan MCC tertinggi yaitu berturut-turut 0,56 dan 0,58 diperoleh di skor ChemPLP -25 dan -50. Oleh karena itu berikutnya dilakukan pada skor ChemPLP -55, -60, -65, dan -70 (per 5 unit), dan didapati bahwa nilai *F-score* dan MCC tertinggi yaitu berturut-turut 0,58 dan 0,58. Selanjutnya untuk memperkecil selisih nilai skor ChemPLP sebagai *cutoff* menjadi 1 unit, analisis dilakukan lagi pada -56, -57, -58, -59, -61, -62, -63, dan -64. Pada analisis dengan nilai-nilai tersebut sebagai *cutoff* diperoleh bahwa nilai *F-score* dan MCC tertinggi yaitu berturut-turut 0,59 dan 0,61 adalah saat menggunakan skor ChemPLP -61 maupun -62 sebagai *cutoff*. Skor ChemPLP yang lebih tinggi sebagai *cutoff*

Strategi optimasi dengan selisih unit berjenjang dilakukan karena setiap perhitungan SIPL ansambel untuk 184426200 *bitstring* SIPL yang dihasilkan dalam penelitian ini dibutuhkan waktu komputasi sekitar 9 jam. Total dibutuhkan sekitar 150 jam. Jika hendak dilakukan per 1 unit dari -25 hingga -125 akan dibutuhkan waktu komputasi sekitar 900 jam. Strategi unit berjenjang ini mampu menurunkan waktu komputasi hingga sekitar 17%. Secara utuh, *confusion matrix* dari analisis HKSA biner dengan SIPL ansambel sebagai prediktor pada berbagai nilai *cutoff* yang dilakukan pada penelitian ini disajikan di Tabel 1.

Protokol terbaik untuk penemuan ligan-ligan baru bertargetkan CXCR4 dalam penelitian ini memiliki nilai FP 323,57 yang jauh lebih baik dari protokol referensi (17,5)<sup>(6)</sup>. Nilai FP ini juga jauh lebih baik dari protokol-protokol PVBS yang terbukti berguna secara prospektif mengidentifikasi senyawa aktif baru<sup>(3,7,8,29)</sup>. Nilai-nilai *F-score* dan MCC di atas 5 menunjukkan bahwa prediksi secara retrospektif memberikan korelasi positif yang signifikan pada jumlah data yang mencapai 3446 (nilai  $p < 0,05$ ). Berdasarkan Gambar 3, suatu senyawa dikatakan aktif jika setelah ditambahkan pada target virtual CXCR4 menggunakan PLANTS1.2 dan identifikasi SIPL menggunakan PyPLIF dan dipilih pose-pose dengan skor ChemPLP  $m < -61$ , senyawa tersebut memiliki nilai SIPL ansambel pada



Gambar 3. Protokol PVBS terbaik untuk identifikasi ligan-ligan pada CXCR4.

memberi peluang lebih banyak pose yang terpilih<sup>(7)</sup>, sehingga skor ChemPLP -61 dipilih sebagai *cutoff*. Skema protokol SBVS dengan kemampuan prediksi terbaik hasil penelitian ini disajikan pada Gambar 3.

*bitstring* 274 kurang dari 0,25 dan SIPL ansambel pada *bitstring* 246 tidak kurang dari 0,37 dan: (i) SIPL ansambel pada *bitstring* 82 tidak kurang dari 0,03 dan SIPL ansambel pada *bitstring* 71 kurang dari

Tabel 1. Kualitas prediksi PVBS bertargetkan CXCR4 berupa confusion matrix, nilai F-score dan MCC hasil analisis HKSA biner dengan SIPL ansambel pada beberapa *cutoff* skor ChemPLP

No.	Skor ChemPLP	Confusion Matrix				F-score (21,22)	MCC (21,22)
		TPs <sup>a)</sup>	FNs <sup>a)</sup>	TNs <sup>a)</sup>	FPs <sup>a)</sup>		
1.	-25	17	23	3402	4	0,56	0,58
2.	-50	17	23	3402	4	0,56	0,58
3.	-55	19	21	3396	10	0,55	0,58
4.	-56	19	21	3396	10	0,55	0,58
5.	-57	19	21	3396	10	0,55	0,58
6.	-58	17	23	3403	3	0,57	0,60
7.	-59	17	23	3403	3	0,57	0,60
8.	-60	19	21	3399	7	0,58	0,58
9.	-61	19	21	3401	5	0,59	0,61
10.	-62	19	21	3401	5	0,59	0,61
11.	-63	9	31	3404	2	0,35	0,43
12.	-64	8	32	3406	0	0,33	0,45
13.	-65	9	31	3406	0	0,37	0,47
14.	-70	0	40	3406	0	0	-
15.	-75	14	26	3399	7	0,46	0,48
16.	-100	0	40	3406	0	0	-
17.	-125	0	40	3406	0	0	-

<sup>a)</sup> True Positives (TPs), False Negatives (FNs), True Negatives (TNs), dan False Positives (FPs)<sup>(21)</sup>.

0,15; atau (ii) SIPL ansambel pada *bitstring* 82 kurang dari 0,03 dan SIPL ansambel pada *bitstring* 288 tidak kurang dari 0,11 dan SIPL ansambel pada *bitstring* 326 kurang dari 0,01 dan SIPL ansambel pada *bitstring* 1 kurang dari 0,05. Hal ini menunjukkan bahwa interaksi pada *bitstring* 274 sebagai prediktor pada cabang utama merupakan interaksi yang kurang disukai, sementara interaksi pada cabang penting berikutnya yaitu *bitstring* 246 merupakan interaksi yang disukai. Interaksi pada cabang-cabang berikutnya tidak dapat ditentukan sebagai interaksi yang disukai maupun tidak disukai karena *bitstring* 82 sebagai pokok cabang di sini memiliki dua cabang yang di keduanya didapati kemungkinan untuk mengidentifikasi ligan CXCR4.

*Bitstring-bitstring* yang teridentifikasi penting dalam prediksi ligan untuk CXCR4 menurut hasil analisis HKSA biner (Gambar 3), yaitu *bitsring* 274, 246, 82, 288, 71, 326, dan 1 mewakili interaksi-interaksi berikut secara berurutan: ikatan hidrofobik dengan PHE189, ikatan hidrofobik dengan ILE185, ikatan hidrogen dengan TRP94 dengan ligan sebagai donor, ikatan hidrofobik dengan TYR255, ikatan hidrofobik dengan PHE93, ikatan hidrogen dengan SER285 dengan ligan sebagai akseptor, dan ikatan hidrofobik dengan ARG30<sup>(10)</sup>. Pengecekan dengan studi *site-directed* mutagenesis (SDM) secara retrospektif dengan mengacu data-data mutasi yang dikelola oleh GPCRdb (<http://gpcrdb.org/>) menunjukkan bahwa PHE189, ILE185, TYR255, dan ARG30 merupakan determinan molekuler dalam ikatan antara ligan dengan CXCR4<sup>(28)</sup>. Selain itu, analisis visual pada interaksi IT1t dengan CXCR4 menunjukkan bahwa IT1t berinteraksi dengan ILE185, TRP94 dan TYR255<sup>(1,2)</sup>. Hasil yang serupa ini menunjukkan potensi analisis HKSA biner menggunakan SIPL ansambel tidak hanya mampu

meningkatkan kualitas PVBS namun juga mampu mengenali determinan molekuler dalam ikatan protein-ligan.

## SIMPULAN

Analisis HKSA biner menggunakan SIPL ansambel olahan dari SIPL luaran PyPLIF pasca simulasi penambatan molekuler mampu meningkatkan kemampuan PVBS bertargetkan CXCR4. Protokol referensi memberikan nilai FP = 17,5<sup>(5)</sup> dan protokol PVBS reguler menggunakan 1% FPs memberikan nilai FP = 0 untuk penggunaan skor ChemPLP sebagai fungsi obyektif dan FP = 2,5 dengan Tc-SIPL sebagai fungsi obyektif. Penggunaan SIPL dari pose dengan skor ChemPLP terbaik sebagai prediktor seperti yang berhasil diujicobakan pada target ER $\alpha$ <sup>(16)</sup>, tidak mampu meningkatkan kualitas PVBS pada target CXCR4 ini. Protokol PVBS terbaik hasil penelitian ini menggunakan SIPL ansambel dari pose-pose yang memiliki skor ChemPLP < -61 mampu memberikan nilai FP = 323,57. Protokol PVBS dengan nilai *F-score* = 0,59 dan MCC = 0,61 ini juga berhasil mengenali PHE189, ILE185, TRP94, TYR255, dan ARG30 sebagai determinan molekuler interaksi ligan pada CXCR4.

## DAFTAR PUSTAKA

1. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*. 2010. 330(6007):1066–71.
2. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure*. 2011. 19(8):1108–26.

3. Kufareva I, Katritch V, Stevens RC, Abagyan R. Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure*. 2014. 22(8):1120–39.
4. Roumen L, Scholten DJ, De Kruijf P, De Esch IJP, Leurs R, De Graaf C. C(X)CR in silico: Computer-aided prediction of chemokine receptor-ligand interactions. *Drug Discov Today Technol*. 2012. 9(4):e281–91.
5. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J Med Chem*. 2012. 55(14):6582–94.
6. Marcou G, Rognan D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J Chem Inf Model*. 2007. 47(1):195–207.
7. de Graaf C, Kooistra AJ, Vischer HF, Katritch V, Kuijter M, Shiroishi M, et al. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H1 receptor. *J Med Chem*. 2011. 54(23):8195–206.
8. Istyastono EP, Kooistra AJ, Vischer H, Kuijter M, Roumen L, Nijmeijer S, et al. Structure-based virtual screening for fragment-like ligands of the G protein-coupled histamine H4 receptor. *Med Chem Commun*. 2015. 6:1003–17.
9. Salentin S, Haupt VJ, Daminelli S, Schroeder M. Polypharmacology rescored: Protein-ligand interaction profiles for remote binding site similarity assessment. *Prog Biophys Mol Biol*. 2014. 116(2-3):174–86.
10. Radifar M, Yuniarti N, Istyastono EP. PyPLIF: Python-based protein-ligand interaction fingerprinting. *Bioinformatics*. 2013. 9(6):325–8.
11. Radifar M, Yuniarti N, Istyastono EP. PyPLIF-assisted redocking indomethacin-(R)-alpha-ethyl-ethanolamide into cyclooxygenase-1. *Indones J Chem*. 2013. 13(3):283–6.
12. Anita Y, Radifar M, Kardono L, Hanafi M, Istyastono EP. Structure-based design of eugenol analogs as potential estrogen receptor antagonists. *Bioinformatics*. 2012. 8(19):901–6.
13. Anita Y, Sundowo A, Dewi NLP, Filaila E, Mulyani H, Risdian C, et al. Biotransformation of eugenol to dehydroeugenol catalyzed by *Brassica juncea* Peroxidase and its cytotoxicity activities. *Procedia Chem*. 2015. 16(2015):265–71.
14. Istyastono EP, Riswanto FDO, Yuliani SH. Computer-aided drug repurposing: a cyclooxygenase-2 inhibitor celecoxib as a ligand for estrogen receptor alpha. *Indones J Chem*. 2015. 15(3):274–80.
15. Setiawati A, Riswanto FDO, Yuliani SH, Istyastono EP. Retrospective validation of a structure-based virtual screening protocol to identify ligands for estrogen receptor alpha and its application to identify the alpha-mangostin binding pose. *Indo J Chem*. 2014. 14(2):103–8.
16. Istyastono EP. Employing recursive partition and regression tree method to increase the quality of structure-based virtual screening in the estrogen receptor alpha ligands identification. *Asian J Pharm Clin Res*. 2015. 8(6):21–4.
17. Therneau T, Atkinson B, Ripley B. rpart: Recursive partitioning and regression Trees. R package version 4.1-9. 2015. diambil dari: <http://CRAN.R-project.org/package=rpart>.
18. Setyaningsih D, Radifar M, Murti YB, Istyastono EP. Construction of in silico structure-based screening tools to study the oxidative metabolites formation of curcumin by human cytochrome 450 3A4. *Indones J Pharm*. 2013. 24(2):75–85.
19. Istyastono EP, Setyaningsih D. Construction and retrospective validation of structure-based virtual screening protocols to identify potent ligands for human adrenergic beta-2 receptor. *Indones J Pharm*. 2015. 26(1):20–8.
20. Istyastono EP, Nurrochmad A, Yuniarti N. Structure-based virtual screening campaigns on curcuminoids as potent ligands for histone deacetylase-2. *Orient J Chem*. 2016. 32(1):275–82.
21. Cannon EO, Amini A, Bender A, Sternberg MJE, Muggleton SH, Glen RC, et al. Support vector inductive logic programming outperforms the naive Bayes classifier and inductive logic programming for the classification of bioactive chemical compounds. *J Comput Aided Mol Des*. 2007. 21:269–80.
22. Desaphy J, Raimbaud E, Ducrot P, Rognan D. Encoding protein-ligand interaction patterns in fingerprints and graphs. *J Chem Inf Model*. 2013. 53(3):623–37.
23. ten Brink T, Exner TE. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. *J Chem Inf Model*. 2009. 49(6):1535–46.
24. Korb O, Stütze T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model*. 2009. 49(1):84–96.
25. Korb O, Stütze T, Exner TE. An ant colony optimization approach to flexible protein-ligand docking. *Proc IEEE Swarm Intell Symp*. 2007. 1(2):115–34.
26. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform*. 2011. 3(1):33–47.
27. R Core Team. R: A Language and Environment for Statistical Computing. Vienna. 2015. diambil dari: <http://www.r-project.org/>
28. Vroling B, Sanders M, Baakman C, Borrmann A, Verhoeven S, Klomp J, et al. GPCRdb: information system for G protein-coupled receptors. *Nucleic Acids Res*. 2011. 39(Suppl. 1):D309–19.
29. Andrews SP, Brown GA, Christopher JA. Structure-based and fragment-based GPCR drug discovery. *ChemMedChem*. 2014(9(2)):256–75.