

# Kajian Pendekatan Penempatan Ligan pada Protein Menggunakan Algoritma Genetika

Hartanto Setiawan dan Mohammad Isa Irawan

Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)  
*e-mail: mii@its.ac.id*

**Abstrak**—Penempatan ligan pada protein atau *molecule docking* merupakan bidang komputasi yang sedang berkembang. Metode *molecular docking* adalah metode yang bermanfaat untuk mencari kombinasi interaksi protein dan ligan serta menjadi dasar penemuan obat secara simulasi. *Molecular docking* yang digunakan adalah *flexible docking* dan jenis protein-ligand *docking*. Pendekatan algoritma genetika merupakan metode alternatif yang bisa digunakan untuk simulasi *molecular docking*. Hasil dari pendekatan algoritma genetika yaitu berupa penempatan posisi *docking* yang optimum. Penerapan algoritma genetika dalam *docking* tidak berlaku untuk semua protein dan ligan. Dalam penerapannya tingkat homologi mempengaruhi keberhasilan dari *docking*.

**Kata Kunci**—Algoritma Genetika, *Molecular docking*, *Protein-ligand docking*, *Flexible docking*.

## I. PENDAHULUAN

SEIRING berkembangnya teknologi informasi, komputasi pengolahan data menjadi lebih cepat untuk dilakukan. Salah satu bidang teknologi yang berkembang sekarang ini adalah komputasi biologi. Komputasi biologi adalah bidang ilmu yang berfokus pada penyusunan sebuah model matematika dalam menyelesaikan dan menganalisis masalah sekuen biologi. Selain itu dalam perkembangan komputasi biologi juga digunakan untuk menemukan prinsip-prinsip baru yang mendasar dalam ilmu biologi seperti pencarian obat.

Komputasi pada bidang biologi atau dikenal sebagai bioinformatika, pada umumnya merupakan kombinasi biologi dan komputasi dimana menggunakan aplikasi dari alat komputasi dan analisis untuk menangkap dan menginterpretasikan data-data biologi. Salah satu nya seperti perkembangan teknologi DNA rekombinan merupakan suatu pengetahuan baru dalam rekayasa genetika organisme yang dikenal sebagai bioteknologi.

Bidang kajian bioinformatika yang sedang berkembang sekarang ini adalah *molecular docking* (penempatan molekul). *Molecular docking* merupakan metode berbasis genetika yang dapat digunakan untuk mencari pola interaksi yang paling tepat dan melibatkan antara dua molekul, yaitu reseptor dan ligan. Ligan sendiri merupakan molekul sinyal kecil yang terlibat dalam kedua proses anorganik dan biokimia.

*Molecular docking* bertujuan meniru peristiwa interaksi suatu molekul ligan dengan protein yang menjadi targetnya pada uji *in-vitro* [1]. *Molecular docking* dapat diklasifikasikan menjadi 3 berdasarkan fleksibilitas molekul yaitu *Rigid Docking* (bersifat rigid/kaku), *semi-fleksible docking* (bersifat semi fleksibel) dan *flexible docking* (bersifat fleksibel).

Tujuan dari *docking* adalah untuk mencapai konformasi protein dan ligan yang optimal. *Docking* membantu dalam mempelajari obat / ligan atau interaksi reseptor / protein dengan mengidentifikasi situs aktif yang cocok pada protein, mendapatkan geometri terbaik dari kompleks ligan – reseptor. *Docking* menjadi dasar untuk penemuan obat secara simulasi komputasi. Langkah pertama dari desain obat dibantu komputer adalah menemukan situs pengikatan ligan protein, yang merupakan kantong atau celah pada permukaan protein yang digunakan untuk mengikat ligan (obat terlarang) [2].

Dengan peningkatan jumlah struktur biologi molekuler yang tersedia, pendekatan *docking* telah menjadi alat yang sangat penting dan berguna dalam penemuan obat rasional berbasis struktur dan desain [2]. Untuk reseptor protein dengan struktur tiga dimensi yang dikenal, masalahnya *docking* ligan-protein pada dasarnya terdiri dalam memprediksi konformasi terikat ligan molekul dalam situs aktif protein.

Berbagai metode komputasi telah dikembangkan untuk menentukan struktur dan fungsi protein dan mempelajari lebih lanjut tentang protein lipat mekanisme [3]. Efisien komputasi teknik dapat membantu untuk memecahkan masalah struktur protein (yaitu Homologi pemodelan, *threading* dan *ab initio*), yang memungkinkan aplikasi yang beragam dalam banyak bidang penelitian ilmiah [3]. Pengetahuan tentang struktur 3D protein ini juga sangat penting untuk studi mitra makromolekul lain dengan siapa mereka dapat berinteraksi. Selain itu dengan penerapan komputasi dalam hal ini juga dapat mengurangi biaya yang sangat besar, dan menghemat waktu dalam rekombinasinya, serta mempermudah dalam perhitungan kompleks suatu rekombinasi. Dalam hal ini masalah utama *docking* adalah sulit optimasi yang melibatkan banyak derajat kebebasan, dan pengembangan efisien *docking* algoritma dan metodologi akan menjadi manfaat besar dalam desain obat baru. Dalam hal ini pendekatan algoritma optimasi sangat membantu optimasi untuk mendapatkan desain obat secara simulasi. Kombinasi dari algoritma genetika (GA) [4][5], rotamer Perpustakaan dan dinamika molekuler (MD) atau normal mode (NM) [6][7], dapat merupakan pendekatan yang baik untuk melakukan *docking* studi. Algoritma genetika juga dapat digunakan untuk menyelesaikan permasalahan CVRP, Penempatan pegawai, penjadwalan, optimasi BTS, dll [8], [9], [10], dan [11].

Berdasarkan latar belakang diatas, bagaimana mengenai pendekatan penempatan ligan pada protein. Jenis *molecular docking* yang digunakan adalah *flexible docking*. Pendekatan yang digunakan adalah algoritma genetika. Algoritma genetika

digunakan untuk rekombinasi dan untuk mencari kombinasi terbaik dari struktur konformasi protein.

## II. PERANCANGAN PROGRAM

### A. Data Protein dan Ligan

Data yang digunakan adalah protein dari virus *plasmodium malariae* (Malaria) dan virus *dengu* (DBD). Ligan yang digunakan adalah ligan dari tumbuhan kina dan tumbuhan jambu merah. Tumbuhan kina diasumsikan sebagai obat dari penyakit Malaria sedangkan tumbuhan jambu merah diasumsikan sebagai obat dari penyakit DBD. Protein menggunakan jenis docking *protein-ligand* docking serta *flexibel* docking. Pemodelan sekuen diperlukan untuk mendapatkan nilai parameter untuk menghitung energi dari docking molecule. Pemodelan sekuen menggunakan web : <http://swissmodel.expasy.org/>.

### B. Algoritma Genetika

Algoritma genetika digunakan sebagai solusi pemecahan permasalahan *docking molecule*. Algoritma genetika menghasilkan kromosom acak bernilai 0 atau 1 dimana kromosom merepresentasikan solusi penempatan ligan pada protein sesuai dengan *binding site* protein. Algoritma genetika akan berhenti saat melebihi maksimum generasi. Tahap awal adalah dibangkitkan kromosom awal kemudian kromosom diterjemahkan untuk dilakukan perhitungan nilai fitness. Tahap selanjutnya adalah operator algoritma genetika meliputi seleksi, crossover, dan mutasi. Proses seleksi menggunakan metode roulette sedangkan proses crossover menggunakan one-point crossover. Sebelum dilakukan operator genetika dilakukan perhitungan nilai fitness. Setelah tahap mutasi di bangkitkan individu baru dan dilakukan proses seperti tahap awal hingga generasi melebihi maksimum generasi.

#### 1. Teknik Penyandian

Teknik penyandian disini meliputi penyandian gen dari kromosom. Gen merupakan bagian dari kromosom. Selain itu kromosom diasumsikan sebagai solusi dari permasalahan penempatan docking. Satu gen biasanya akan mewakili satu variabel. Gen dapat dipresentasikan dalam bentuk: string bit, pohon, array bilangan real, daftar aturan, elemen permutasi, elemen program, atau representasi lainnya yang dapat diimplementasikan untuk operator genetika.

#### 2. Prosedur Inisialisasi

Ukuran populasi tergantung pada masalah yang akan dipecahkan dan jenis operator genetika yang akan diimplementasikan. Setelah ukuran populasi ditentukan, kemudian harus inisialisai terhadap kromosom yang terdapat pada populasi tersebut. Inisialisasi kromosom dilakukan secara acak, namun demikian harus tetap memperhatikan domain solusi dan kendala permasalahan yang ada.

### C. Fungsi Evaluasi

Ada 2 hal yang harus dilakukan dalam melakukan evaluasi kromosom, yaitu: evaluasi fungsi objektive (fungsi tujuan) dan konversi fungsi objektive kedalam fungsi fitness. Secara umum, fungsi fitness diturunkan dari fungsi objektive dengan nilai tidak negatif. Apabila ternyata fungsi objektive memiliki

nilai negatif, maka perlu ditambahkan suatu konstanta C agar nilai fitness yang terbentuk menjadi tidak negatif.

$$\sum_{\text{Protein}} \sum_{\text{Ligan}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right) + \sum_{\text{Ligan}} \sum_{\text{Ligan}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right) + \sum_{\text{Dihedral Angle}} \frac{V_k}{2} (1 + \cos(n\tau - \phi)) \quad (1)$$

### D. Seleksi

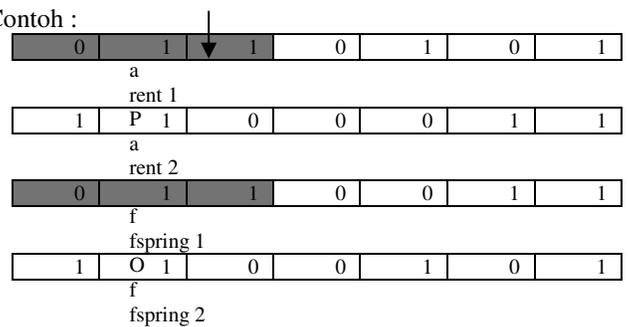
Seleksi ini bertujuan untuk memberikan kesempatan reproduksi yang lebih besar bagi anggota populasi yang paling fit.

### E. Operator Genetika

Ada 2 operator genetika, yaitu:

1. Operator untuk melakukan rekombinasi, yang terdiri dari: rekombinasi bernilai biner (*crossover*).

Contoh :



Gambar 1. Contoh single point *crossover*.

2. Mutasi menciptakan individu baru dengan melakukan modifikasi satu atau lebih gen dalam individu yang sama. Mutasi berfungsi untuk menggantikan gen yang hilang dari populasi selama proses seleksi serta menyediakan gen yang tidak ada dalam populasi awal. Sehingga mutasi akan meningkatkan variasi populasi.

<b>Parent</b>	1	2	3	4	5	6
<b>Offspring</b>	1	2	6	4	5	3

Gambar 2. Contoh single point *crossover*.

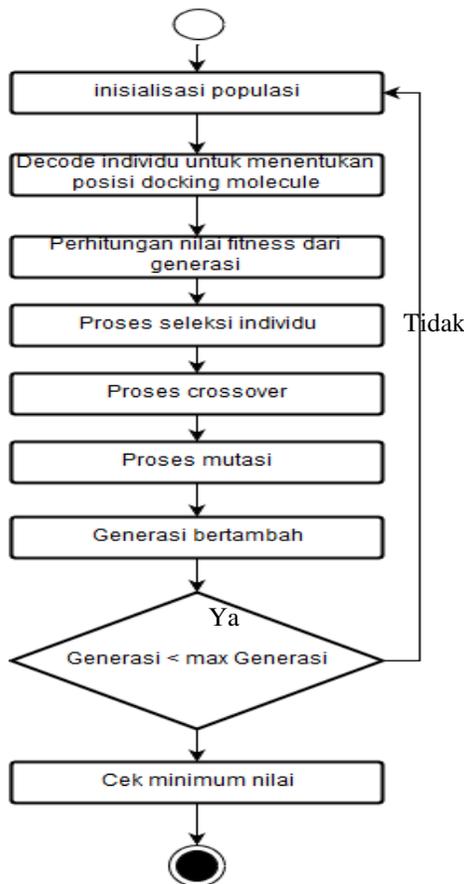
### F. Penentuan parameter

Yang disebut parameter disini adalah parameter kontrol algoritma genetika, yaitu: ukuran populasi (*popsi*), peluang *crossover* ( $P_c$ ), dan peluang mutasi ( $P_m$ ). Nilai parameter ini ditentukan juga berdasarkan permasalahan yang akan dipecahkan.

## III. HASIL DAN PEMBAHASAN

### A. Algoritma Genetika dalam Docking Molecule

Algoritma genetika pada *docking molecule* untuk penempatan ligan pada protein. Diagram Alur untuk Algoritma Genetika pada docking molecule dapat dilihat pada Gambar 3.

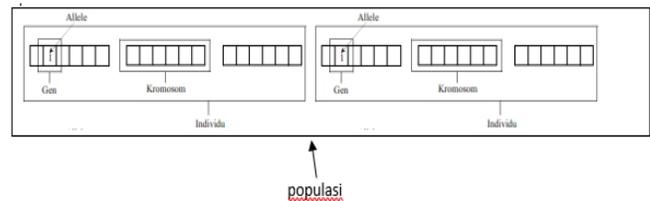


Gambar 3. Alur proses algoritma genetika.

Algoritma genetika digunakan sebagai sebuah solusi docking. Setiap individu merupakan solusi permasalahan dari protein-ligan docking, dimana individu merupakan posisi ligan sebagai reseptor dari protein [5]. Oleh karena itu, konformasi ligan diwakili oleh kromosom yang dibentuk oleh nyata saat gen yang mewakili ligan translasi, orientational dan konformasi derajat kebebasan. Penelitian molekul docking memfokuskan pada produk simulasi Proses pengenalan molekuler. Hal ini bertujuan untuk mencapai konformasi dioptimalkan untuk kedua protein dan ligan dan relatif orientasi antara protein dan ligan sedemikian rupa sehingga energi bebas dari sistem keseluruhan diminimalkan.

Algoritma genetika digunakan sebagai pemecahan solusi pada docking ligan protein. Algoritma genetika digunakan untuk membangun atau mengenerate titik penempatan. Titik penempatana dilakukan pada *binding site* atau sisi aktif sebagai reseptor untuk proses docking. Dalam algoritma genetika terdapat istilah generasi, populasi, individu, kromosom serta alel.

1. Alel adalah nilai dari gen.
2. Kromosom adalah gabungan gen-gen yang membentuk nilai tertentu.
3. Individu adalah satu nilai atau keadaan yang menyatakan salah satu solusi yang mungkin dari permasalahan yang diangkat.
4. Populasi adalah sekumpulan individu yang akan diproses bersama dalam satu siklus proses evaluasi
5. Generasi adalah satu siklus proses evolusi atau iterasi dalam algoritma genetika.



Gambar 4. Ilustrasi representasi Algoritma genetika.

Binding site diperoleh dari nilai random yang unik dengan batas maksimumnya adalah panjang protein. Banyaknya binding site tergantung pada nilai inputan yang diberikan. Algoritma genetika menghasilkan nilai 0 sampai banyaknya *binding site* secara acak disetiap individu dalam populasi.

- a. Tahap awal dalam pembentukan individu adalah menghitung panjang bit kromosom tiap individu. Individu sendiri terdiri dari kumpulan alel atau kromosom.

$$P = \lceil \log_2 \left( (b - a) * 10^d + 1 \right) \rceil$$

Dengan

P = panjang kromsoms

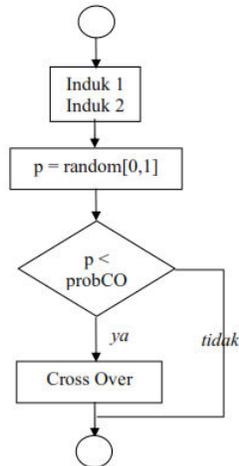
b = banyaknya *binding site*,

a = batas awal,

d = ketepatan angka dibelakang koma.

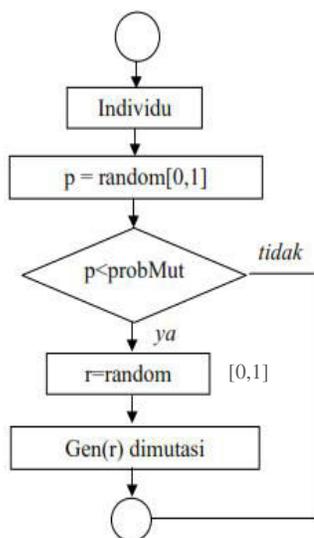
- b. Setelah diperoleh panjang bit kromosm kemudian dilakukan inisialisasi acak dengan cara memberikan nilai biner 1 atau 0 secara acak untuk setiap alel dari gen.
- c. Kemudian bilangan biner pada setiap individu yang telah diperoleh dari tahap inisialisasi acak diubah ke dalam bentuk decimal.
 
$$110 = (2^2 * 1) + (2^1 * 1) + (2^0 * 0) = 4 + 2 + 0 = 6$$
- d. Nilai decimal yang diperoleh kemudian dicocokkan dengan urutan random binding site yang diperoleh dari tahap pembentukan individu. Kemudian dihitung nilai evaluasi atau nilai fitness dengan index posisi sesuai dengan binding site. Jika nilai hasil perubahan biner ke decimal melebihi atau tidak cocok dengan urutan random dair iniding site maka tidak dapat dilakukan docking sehingga nilai evaluasi bernilia 0 dengan kata lain sekuen tidak memenuhi.
- e. Tahap selanjutnya adalah operator algoritma genetika. Dalam algoritma genetika, operator yang digunakan ada 2 yaitu
  1. Operasi Evaluasi yang melibatkan proses seleksi (*selection*) didalamnya.
  2. Operasi Genetika yang melibatkan operator pindah silang (*crossover*) dan mutasi (*mutation*)
- f. Tahap seleksi digunakan untuk memilih individu-individu mana saja yang akan dipilih untuk proses kawin silang dan mutasi. Langkah pertama yang dilakukan dalam seleksi adalah pencarian nilai fitness. Nilai fitnesss akan digunakan pada tahap-tahap seleksi berikutnya. Masing-masing individu dalam selskis akan menerima probabilitas reproduksi yang tergantung pada nilai obyektif terhadap nilai obyektif dari semua individ dalam seleksi tersebut.
- g. Tahap Selanjutnya adalah kawin silang atau (*crossover*). Kawin silang adalah operator dari algoritma genetika tidak melibatkan dua induk untuk membentuk kromosom

baru. Pindah silang menghasilkan titik baru dalam ruang pencarian yang diuji. Operasi ini tidak selalu dilakukan pada semua individu yang ada. Individu dipilih secara acak untuk dilakukan crossing dengan probabilitas kawin silang antara 0 sampai dengan 1. Jika dipindah silang tidak dilakukan, maka nilai induk akan diturunkan kepada keturunan.



Gambar 5. Diagram alir proses *crossover*.

h. Tahap berikutnya pada algoritma genetika adalah mutasi gen. Operator ini berperan untuk menggantikan gen yang hilang dari populasi akibat proses seleksi yang memungkinkan munculnya kembali gen yang tidak muncul pada inisialisasi populasi. Kromosom anak dimutasi dengan menambahkan nilai acak yang sangat kecil dengan probabilitas yang rendah. Jika nilai probabilitas tiap gen lebih kecil dari probabilitas mutase maka gen mengalami mutasi dan terjadi perubahan biner dari 1 ke 0 dan sebaliknya. Sedangkan jika tidak memenuhi maka tidak terjadi perubahan nilai.



Gambar 6. Diagram alir proses mutase.

i. Kemudian kembali ke tahap generate kromosom sampai ke tahap mutasi. Algoritma genetika berhenti ketika generasi melebihi maksimum generasi.

j. Tahap selanjutnya mencari nilai minimum dari semua generasi. Hasil digunakan untuk cek homology kecocokan hasil docking dengan menggunakan software MATLAB

A. Hasil Implementasi dan penyejajaran sekuen

Hasil dari implementasi menggunakan algoritma genetika dapat dilihat pada tabel 1 dimana menghasilkan nilai evaluasi dan posisi dari penempatan ligan pada protein.

Tabel 1.  
Hasil implementasi

No	Protein	Ligan	Nilai evaluasi	Posisi
1	Malaria	Kina	-18234.47	134
2	Malaria	Jambu merah	-13851.972	41
3	DBD	Kina	-3291.149	28
4	DBD	Jambu merah	-3833.07	52

Implementasi juga menggunakan *software* PLANTS (*Protein-Ligand ANT System*). PLANTS menggunakan algoritma *Ant Colony* (koloni semut) untuk *docking molecule*. Tabel 2 merupakan hasil implementasi dengan *software* PLANTS

Tabel 2.  
Hasil docking dengan PLANTS

No	Protein	Ligan	Nilai evaluasi
1	Malaria	Kina	9898.66
2	Malaria	Jambu merah	9933.13
3	DBD	Kina	27641.1
4	DBD	Jambu merah	27559.9

Hasil implementasi menggunakan java parameter yang digunakan adalah parameter CHARMM dengan nilai hasil docking evaluasi sekitar -13254.93. Implementasi menggunakan *software* PLANTS parameter yang digunakan adalah parameter GOLD didapatkan nilai 9898.66 Terlihat bahwa nilai yang dihasilkan memiliki selisih yang besar di karena pamater yang digunakan berbeda. Tahap selanjutnya adalah penyejajaran sekuen hasil docking dengan sebelum docking.

Tabel 3.  
Hasil Pensejajaran

No	Protein	Ligan	Persentasi kemiripan	
			Needleman-wunsch	Smith-Waterman
1	Malaria	Kina	75%	82%
2	Malaria	Jambu merah	76%	82%
3	DBD	Kina	75%	75%
4	DBD	Jambu merah	64%	64%

Pada tabel 3 menunjukan hasil pensejajaran untuk protein virus Malaria dan DBD, untuk ligan adalah kina dan jambu merah. Dari tabel terlihat untuk protein virus malaria dan ligan kina memiliki persentasi kemiripan nilai lebih kecil sama

dengan dibandingkan dengan protein virus malaria dan ligan jambu merah. Kemudian dari tabel juga terlihat untuk protein virus DBD dan ligan jambu merah memiliki persentasi kemiripan nilai lebih kecil dibandingkan dengan nilai dari protein virus DBD dan ligan kina. Persentasi kemiripan mempengaruhi suatu docking, untuk nilai persentasi kemiripan (homologi) kecil maka protein tersebut sudah memiliki sifat yang berbeda dari sekuen awal protein.

#### IV. KESIMPULAN

Berdasarkan analisis terhadap hasil pengujian program, maka dapat diambil kesimpulan bahwa Algoritma Genetika dapat diterapkan untuk *molecule docking* ligan dan protein dengan protein virus malaria dan DBD, ligan tumbuhan kina dan jambu merah. Penerapan algoritma genetika dalam docking tidak berlaku untuk semua protein dan ligan. Dalam penerapannya tingkat homologi mempengaruhi keberhasilan. Semakin tinggi homologi sekuen setelah docking dan sekuen awal virus, maka docking tidak berhasil. maka docking tidak berhasil.

#### DAFTAR PUSTAKA

- [1] D. Motiejunas and R. Wade, "Structural, Energetics, and Dynamic Aspects of Ligand-Receptor Interactions. In J.B. Taylor & D. J. Triggle (Eds.)," *Compr. Med. Chem. II Elsevier*, vol. 4, pp. 193–214, 2006.
- [2] P. J. Gane and P. M. Dean, "Recent Advances in Structure-Based Rational Drug Design," *Curr. Opin. Struct. Biol.*, vol. 10, pp. 401–404, 2000.
- [3] I. Belda and et al, "ENPDA: An Evolutionary Structure-Based De Novo Peptide Design Algorithm," *J. Comput. Aided. Mol. Des.*, vol. 19, pp. 585–601, 2005.
- [4] F. Osterberg, G. Morris, M. Sanner, A. Olson, and D. Goodsell, "Automated Docking to Multiple Target Structure: Incorporation of Protein Mobility and Structural Water Heterogeneity in Auto Dock," *Protein*, vol. 46, pp. 34–40, 2002.
- [5] C. d. Magalhães, "A Genetic Algorithm for The Ligand-Protein Docking Problem," *Genet. Mol. Biol.*, vol. 27, pp. 605–610, 2004.
- [6] N. Floquet, J. Marechal, M. Badet-Denisot, C. Robert, M. Dauchez, and D. Perahia, "Using Normal Modes Analysis as A Prerequisite for Drug Design : Application to Matix Metalloproteinase Inhibitors," *FEBS Lett.*, vol. 580, pp. 5130–5136, 2006.
- [7] L. Mouawad, D. Perahia, Q. Cui, and I. Bahar, *Normal Mode Analysis Theory and Applications to Biological and Chemical Systems*. Chapman & Hall, 2006.
- [8] M. L. Shahab and M. I. Irawan, "Algoritma Genetika Ganda untuk Capacitated Vehicle Routing Problem," *J. Sains dan Seni ITS*, 2016.
- [9] M. Muhtaromi and M. I. Irawan, "Perancangan Prototipe Perangkat Lunak Untuk Penempatan Pegawai Dengan Model Pilihan Dari Perspektif Dua Arah Berbasis Algoritma Genetika," *J. Sains dan Seni ITS*, 2016.
- [10] A. Firmansyah, D. B. Utomo, and M. I. Irawan, "Algoritma Genetika Ddengan Modifikasi Kromosom Untuk Penyelesaian Masalah Penjadwalan Flowshop," *J. Sain dan Seni*, vol. 1, no. 1, 2016.
- [11] Y. Pramsistya, D. B. Utomo, and M. I. Irawan, "Optimasi Penempatan BTS dengan Menggunakan Algoritma Genetika," *J. Sain dan Seni*, vol. 1, no. 1, 2010.