

# Aplikasi Teori Graf dalam Analisa Struktur Protein

Harum Lokawati, 13515109<sup>1</sup>  
Program Studi Teknik Informatika  
Sekolah Teknik Elektro dan Informatika  
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia  
<sup>1</sup>13515109@std.stei.itb.ac.id

**Abstract**—With proteins as one of cellular components which have crucial role in biological function, there is a growing interest in studies to explain the correlation of the structure, its biological function, and to develop new. Since the lab-intensive one have strong biases, and low accuracy, it is desirable to explore mathematical method to express the information. This little review deals with some of graph theory and its relevancy with protein interaction. The principle of graph theory has been applied mostly in electrical engineering and computer science. The basic concepts of graph in analysing the protein interaction are discussed.

**Keywords**—Protein; graph; Laplacian matrix, protein-protein interaction.

## I. INTRODUCTION

Teori graf merupakan salah satu subjek dalam matematika diskrit yang digunakan untuk merepresentasikan objek-objek dan hubungannya dengan pendekatan geometri. Objek diskrit dalam graf dapat direpresentasikan dalam bentuk noktah, bulatan, atau titik, serta hubungannya dalam bentuk garis atau sisi. Dengan direpresentasikan dalam graf, berbagai permasalahan seperti *networking*, *electrical circuits*, dsb tentu akan lebih mudah untuk dianalisis. Penggunaan graf sebagai representasi dari suatu masalah telah dikembangkan dalam berbagai disiplin ilmu. Sekarang ini banyak digunakan sistem dengan noktah dan sisi dalam jumlah besar yang tentu akan diproses dalam komputer dengan efektif. Salah satu topik yang juga akan dibahas dalam makalah ini adalah representasi graf dalam interaksi antara protein.

Protein merupakan komponen sel yang memiliki peran penting dalam proses biologis – control enzim, proses regulasi, aktifitas sel, dan transfer sinyal. Mempelajari jaringan dari interaksi antar protein dapat memberikan gambaran fungsi kerja sel dan penyakit-penyakit yang lebih kompleks. Interaksi antar protein sering disebut *protein-protein interaction* (PPI) yang jaringannya dapat direpresentasikan dalam graf, yaitu noktah menunjukkan protein dan sisi sebagai PPI. Representasinya dalam graf dapat digunakan untuk analisa skala besar dan karakter struktural dari protein.

## II. TEORI GRAF

### 2.1. Definisi

Graf  $G(V,E)$  didefinisikan sebagai pasangan himpunan  $V$  (*vertices* atau *nodes*) dan  $E$  (*edges*), dengan  $V$  adalah himpunan tidak kosong dan  $E$  mungkin kosong. Jadi, sebuah graf dimungkinkan hanya memiliki satu simpul (*vertex*) dan tanpa sisi, atau yang disebut dengan graf trivial. Dua simpul  $v_1$  dan  $v_2$  pada sebuah graf  $G$  dikatakan bertetangga jika ada sisi  $e_{12}$  yang menghubungkan keduanya. Dua sisi pada graf  $G$  dikatakan bertetangga jika kedua sisi tersebut menghubungkan paling tidak satu simpul yang sama. Derajat sebuah simpul ( $deg_i$ ) diartikan sebagai banyaknya simpul yang merupakan tetangga dari simpul tersebut.

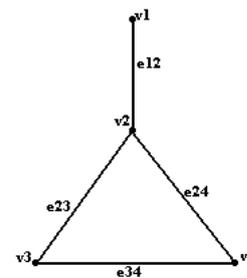


Fig. 1.

Graf diatas memiliki 4 simpul ( $v_1, v_2, v_3, v_4$ ) dan 4 sisi ( $e_{12}, e_{23}, e_{34}, e_{24}$ ) dan derajat dari tiap simpul adalah  $(1,3,2,2)$ .

### 2.2. Jenis-Jenis Graf

Berdasarkan orientasi arahnya, secara umum graf dibagi menjadi dua jenis yaitu graf tak-berarah (*undirected graph*) dan graf berarah (*directed graph*). Pada graf tak-berarah, suatu sisi  $e_{12}$  sama dengan sisi  $e_{21}$ . Contoh dari graf tak-berarah adalah representasi dari ikatan kimia, topologi molekular, dll.

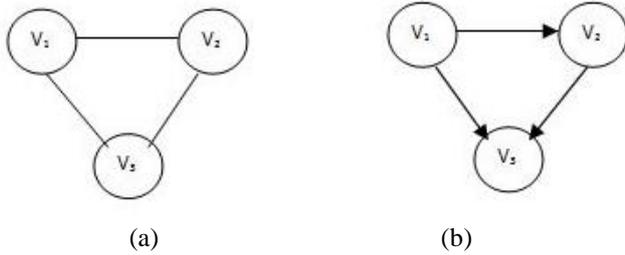


Fig. 2. (a) Graf tak-berarah, (b) Graf berarah.

Berdasarkan ada atau tidaknya gelang (*loop*) pada suatu graf, maka dapat dibedakan menjadi 2, yaitu graf sederhana (*simple graph*) dan graf tidak sederhana (*unsimple graph*). Terdapat dua macam graf tidak sederhana, yaitu graf ganda (*multigraph*) dan graf semu (*pseudograph*). Pada graf ganda, terdapat dua simpul yang dihubungkan oleh lebih dari satu sisi. Sedangkan disebut graf semu jika pada graf tersebut terdapat gelang (*loop*). Karena pada bahasan kali ini, lebih megarah ke bentuk dari interaksi protein, maka akan lebih terfokus ke graf sederhana.

Jenis graf lain yang umum digunakan adalah graf lengkap dan graf berbobot. Suatu graf disebut graf lengkap ( $K_n$ ) jika setiap simpul memiliki  $n-1$  simpul tetangga. Sedangkan graf berbobot adalah suatu graf dimana setiap simpul dan sisinya dibedakan dengan pemberian nilai tertentu sebagai bobot dari simpul. Graf berbobot biasanya digunakan untuk merepresentasikan struktur kimia yang mengandung berbagai atom dan ikatan-ikatan yang beragam.

### 2.3. Upagraf

Suatu graf  $G_1(V_1, E_1)$  dapat didefinisikan sebagai upagraf dari graf  $G(V, E)$  jika  $V_1 \subseteq V$  dan  $E_1 \subseteq E$ . Jika upagraf  $G_1$  mengandung semua simpul dari  $G$ , maka  $G_1$  disebut upagraf merentang. Jadi, secara umum upagraf suatu graf yang mengandung simpul-simpul dari graf tsb. Pada bahasan kali ini upagraf digunakan untuk menganalisa ketersamaan antara suatu struktur protein.

### 2.4. Graf Isomorfik

Dua graf  $G_1$  dan  $G_2$  dikatakan isomorfik jika ada korespondensi satu-satu antara simpul-simpul keduanya dan antara sisi-sisi keduanya sehingga hubungan antara sisi-sisi dan simpul-simpulnya sama. Dari teori tersebut, dapat disimpulkan bahwa syarat dari dua graf agar isomorfik antara lain:

1. Jumlah simpul sama,
2. Jumlah sisi sama,
3. Jumlah derajat tiap simpul sama.



Fig. 3.

Pada fig. 3, graf a dan b memenuhi syarat isomorfik, tetapi jika dilihat pada graf b, simpul 4 tidak bertetangga dengan simpul 5, berbeda dengan graf a. Sehingga, untuk mengidentifikasi isomorfisme dua graf, masih harus dilakukan pemeriksaan secara visual untuk menentukan apakah dua graf tertentu isomorfik.

### 2.5. Representasi Graf

Representasi suatu data dengan graf sangat berguna dalam hal visualisasi. Tetapi untuk mengolah datanya dalam algoritma tertentu, dibutuhkan representasi lain yaitu dengan matriks. Jenis matriks yang digunakan untuk merepresentasikan graf bergantung dari sifat atau algoritma mana yang akan digunakan. Jenis yang paling umum digunakan adalah matriks ketetanggaan, matriks bersisian, matriks Laplacian.

#### 2.5.1. Matriks Ketetanggaan (*adjacency matrix*)

Matriks ketetanggaan  $A=A(G)$ , dari graf  $G$  dengan  $n$  simpul, mempunyai ukuran  $n \times n$  dan simetris.

$$[A]_{ij} = \begin{cases} W_{ij}, & \text{jika } i \text{ bukan } j, \text{ simpul } i \text{ dan } j \text{ bertetangga} \\ 0, & \text{jika simpul } i \text{ dan } j \text{ tidak bertetangga} \\ 1, & \text{jika simpul } i \text{ dan } j \text{ bertetangga untuk graf tidak berbobot.} \end{cases}$$

Matriks ketetanggaan dari fig. 3a dapat digambarkan sebagai berikut:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

#### 2.5.2. Matriks Bersisian (*incidency matrix*)

Matriks bersisian menyatakan kebersisian simpul dengan sisi. Sehingga, dimensi matriks yang digunakan adalah kolom sebagai sisi yang ada pada graf dan baris sebagai simpul-simpulnya.

$$[A]_{ij} = \begin{cases} 0, & \text{jika simpul } i \text{ tidak bersisian dengan sisi } j, \\ 1, & \text{jika simpul } i \text{ bersisian dengan sisi } j. \end{cases}$$

#### 2.5.3. Matriks Laplacian

Matriks Laplacian dari sebuah graf  $G$ ,  $L(G)$ , didefinisikan sebagai

$$L(G) = \text{Deg}(G) - A(G)$$

dengan  $\text{Deg}(G)$  adalah matriks derajat dan  $A(G)$  adalah matriks ketetanggaan dari graf  $G$ . Matriks derajat adalah diagonal matriks yang berisi derajat dari setiap simpul.

$$\text{Deg}_{i,j} = \text{deg}_i, \quad i=j$$

$$0, \quad i \neq j.$$

Matriks Laplacian dari fig. 3a.

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 \end{pmatrix}$$

### III. INTERAKSI ANTAR PROTEIN DAN TOPOLOGI

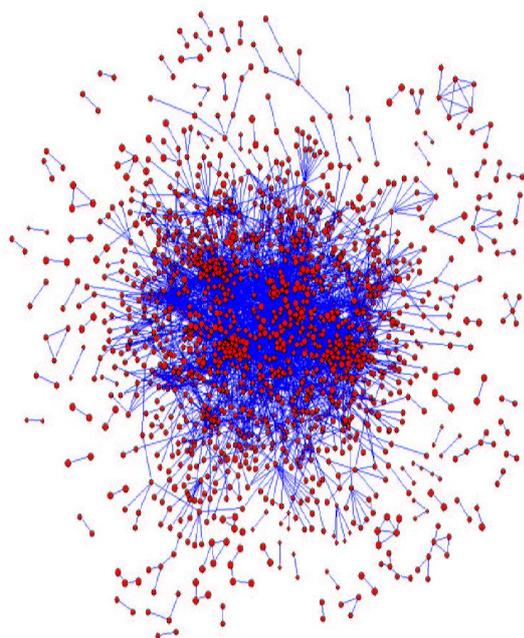


Fig.4.

Protein merupakan sebuah polimer yang terdiri dari asam amino terurut dan dihubungkan oleh ikatan peptida. Sehingga susunan asam amino yang unik menentukan struktur dari suatu protein tertentu. Sedangkan struktur sekunder dari protein sendiri terdiri dari  $\alpha$ -helix (pilinan rantai asam amino seperti spiral),  $\beta$ -sheet (lembaran rantai asam amino),  $\beta$  turn (lekukan beta), dan  $\gamma$ -turn (lekukan gamma). Protein dalam jumlah besar yang bergabung untuk melakukan fungsi tertentu disebut dengan *protein complexes*. Beberapa bukti menunjukkan bahwa protein kompleks merupakan salah satu upagraf dari jaringan PPI.

Jaringan PPI secara umum direpresentasikan dalam sebuah graf, dengan simpul sebagai protein dan sisi sebagai PPI itu sendiri. Contoh dari PPI yang dikonstruksi dalam graf ditunjukkan pada fig.4. Graf tersebut dibentuk dari 2.401 protein. Secara umum, jaringan PPI menggunakan graf tak-berarah dan direpresentasikan hanya dalam *binary interactions*. PPI dalam jumlah besar

Protein memiliki fungsi yang kompleks, serta mengalami perubahan-perubahan yang memungkinkan adanya struktur protein kompleks baru. Untuk memahami hal-hal tersebut, diperlukan adanya gambaran yang jelas pada struktur tiga dimensi dari protein. Beberapa representasi graf sederhana pada protein ditunjukkan pada fig.4. Pada gambar dibawah struktur sekunder protein,  $\beta$  direpresentasikan dalam bentuk segitiga dan  $\alpha$  dalam bentuk segiempat sebagai simpul dari graf dan ikatan diantaranya sebagai sisi.



(a)  $\beta\alpha\beta$



(b)  $\beta$ -hairpin

Fig. 5.

Dengan memanfaatkan analisis dari matriks laplacian dari susunannya, dapat diprediksi susunan asam amino yang berguna untuk merancang protein untuk kebutuhan medis ataupun studi. Topik yang sering dijadikan sebagai masalah adalah bagaimana mengidentifikasi interaksi antara protein dan molekul lain, sebagai salah satu usaha untuk memahami berbagai penyakit dan keadaan medis lain, serta pada pembuatan obat.

Berbagai cara yang dilakukan untuk mengidentifikasi interaksi antar protein lebih terfokus dalam lab dan menghasilkan akurasi yang rendah. Sedangkan, secara matematis, jaringan PPI dapat direpresentasikan dengan menggunakan graf dengan hasil yang lebih efektif. Disamping itu, dibutuhkan pula identifikasi dengan metode lain seperti *yeast 2-hybrid assay*, *orthology-based interaction maps*, dan *literature-based interaction maps* yang tidak akan dibahas kali ini.

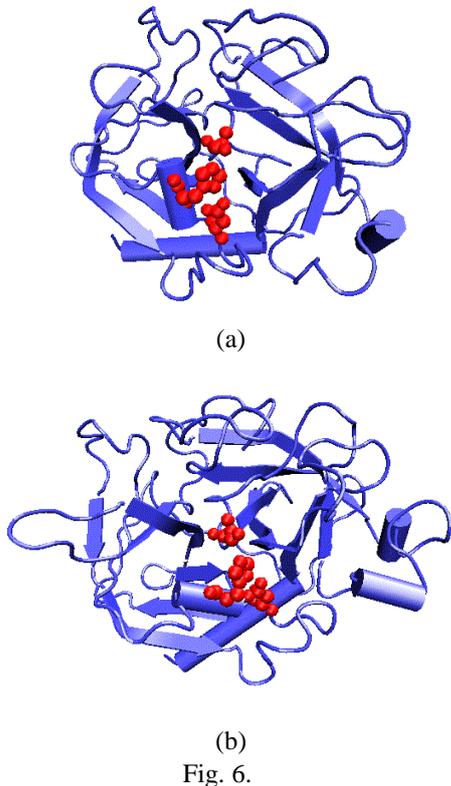
Struktur protein memiliki bentuk geometri tertentu yang dibentuk dari *protein backbone* dan *side chains*. Berbagai struktur dibedakan berdasarkan bentuk geometri tetapi memiliki topologi yang sama. Salah satu formulasi matematikal dari topologi protein dilakukan oleh Koch dkk, yang mengenalkan konsep matematika dari graf yang merepresentasikan struktur *beta*. Pada graf *beta* simpul merepresentasikan *beta-strand* dan dua sisinya menunjukkan urutan dan ikatan hydrogen yang menghubungkannya. Teori graf sangat berperan dalam merepresentasikan topologi protein yang sangat rumit.

### IV. APLIKASI TEORI GRAF DALAM ANALISA STRUKTUR PROTEIN

Jika telah terbentuk suatu graf yang menunjukkan struktur sebuah protein, maka dapat dilakukan analisa-analisa berkaitan dengan struktur dari protein itu sendiri, pola dari topologi protein, prediksi dari struktur suatu

protein, dll.

Struktur dari protein di alam tentu jumlahnya tak-terhitung, tetapi protein-protein tersebut mengadopsi struktur yang terbatas yang dikonstruksi menjadi *protein folds* atau struktur yang lebih kompleks lainnya. Sehingga dapat ditentukan kesamaan struktur dari protein yang berbeda. Hal tersebut berguna untuk mengidentifikasi korelasi antara struktur dan fungsinya.



Gambar diatas merupakan bentuk tiga dimensi dari tripsin(a) dan elastin(b) yang mempunyai fungsi sebagai katalis dan mengandung rangkaian Ser-His-Asp. Dengan menggunakan teori graf, dapat diidentifikasi pola dan secara spesifik dengan simpul dan sisinya. Jika graf telah dibuat, maka polanya akan ditentukan dengan mengenali sifat isomorfis dari graf tersebut dengan *protein atoms* pada *Protein Data Bank*. Dengan melakukan berbagai analisis graf protein, maka pola-pola tertentu dapat ditentukan bersamaan dengan fungsinya.

Penentuan simpul dan sisi dapat dilakukan dengan menggunakan berbagai parameter seperti struktur sekunder dari protein (A dan B), ikatan hydrogen, serta sudut dari struktur yang membentuknya. Dua protein dapat dibandingkan, dilihat dari struktur graf yang membentuknya dengan *graph isomorphism detection method*. Salah satu yang banyak digunakan adalah dengan *Ullman's algorithm* yang secara umum adalah dengan membentuk matriks yang menunjukkan keterhubungan antara simpul dan sisi, lalu dilakukan indentifikasi isomorfis dari suatu graf dengan *matrix permutations*.

Metode selanjutnya adalah dengan menggunakan algoritma *maximal common subgraph (MCS)*. Permasalahan dalam penggunaan MCS adalah untuk

membandingkan dua graf sederhana dengan  $m$  simpul,  $n$  simpul, dan upagraf dengan  $k$  simpul, maka jumlah komputasi komparasinya akan sebanyak

$$\frac{m!n!}{(m-k)!(n-k)!k!}$$

sehingga komputasi ini sangat tidak efektif. Metode yang digunakan untuk mengurangi jumlah komparasi adalah dengan menggunakan *clique-detection algorithm* yang jauh lebih efisien. Pada algoritma *clique-detection* sebuah *clique* dalam graf  $G$  adalah upagraf lengkap dari  $G$ . *Maximal clique* adalah *clique* dimana tidak ada simpul lain yang bisa ditambahkan.

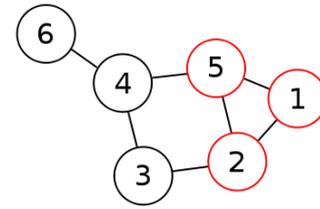


Fig.7.

Pada graf diatas, 1-2-5 adalah *maximal clique* dari graf tersebut.

Pola yang didapatkan akan mengarah ke struktur protein yang isomorfis terhadap struktur lain dengan metode *subgraph isomorphism*. Hal tersebut juga berkorelasi pada hubungan topologi protein atau struktur dengan fungsinya. Salah satu graf yang dihasilkan dari protein-protein *interaction* adalah schizoprenia PPI.

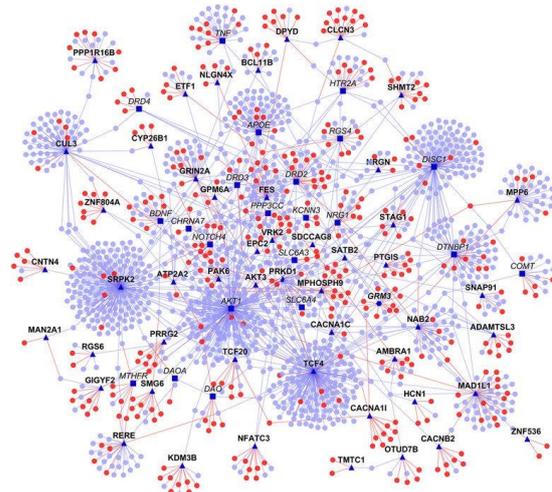


Fig.8.

Representasi lain dalam graf adalah dengan *signed network interaction* yang mendeskripsikan jenis interaksi yang ada pada protein-protein tersebut seperti pada fig.9.

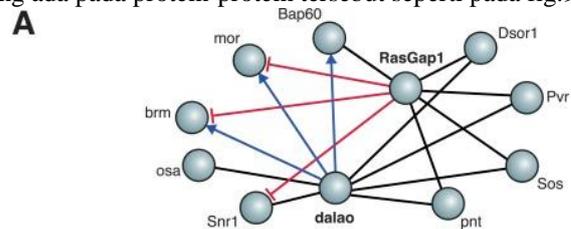


Fig.9.

## V. CONCLUSION

Dasar teori graf dapat diaplikasikan dalam berbagai disiplin ilmu lain seperti biochemical atau biasa disebut bioinformatics. Salah satunya adalah untuk mengidentifikasi hubungan antara protein yang kemudian dimanfaatkan dalam penelitian untuk membentuk protein baru, kebutuhan medis, dan studi lain. Representasi protein dalam graf yaitu simpul graf menunjukkan protein ataupun struktur sekundernya dan sisi menunjukkan hubungan kedekatan spasial atau interaksi didalamnya. Representasi dalam graf akan dianalisis untuk menemukan pola susunan asam amino atau pola struktur interaksinya sehingga pola yang didapatkan dapat dimanfaatkan untuk mengidentifikasi fungsi dan peta untuk membentuk protein baru atau rekayasa protein. Berbagai algoritma lain juga digunakan dalam proses pengidentifikasian yaitu *Ullman's Algorithm*, *Clique-detection algorithm*, *Subgraph Isomorphism*, dll.

## VII. UCAPAN TERIMA KASIH

Pertama-tama ucapan terima kasih penulis sampaikan kepada Allah SWT, karena berkat rahmat-Nya penulis dapat menyelesaikan makalah ini dengan baik. Selain itu, ucapan terima kasih juga penulis sampaikan kepada orang tua dan teman-teman seperjuangan yang telah mendukung penulis selama ini. Ucapan terimakasih, penulis juga ingin menyampaikannya kepada para pembimbing yaitu Ir. Rinaldi Munir, MT dan Harlili M.Sc atas ilmu dan bimbingan yang telah diberikan selama ini.

## REFERENCES

- [1] Saraswathi Vishveshwara, "Protein Structure: Insights from Graph Theory" 2002.
- [2] C. Chothia, Nature 357(6379), 543 (1992).
- [3] Kenneth H. Rosen, "Discrete Mathematics and Its Applications" 7ed. New York: McGraw Hill, 2011.
- [4] J.R. Ullman, "An Algorithm for Subgraph Isomorphism", 1975.
- [5] Natasa Przulj, "Knowledge Discovery in Proteomics: Graph Theory Analysis of Protein-Protein Interactions", University of Toronto, Canada, 2005.
- [6] Futschik, E.M "Graph-Theoretical comparison reveals structural divergence of human protein interaction networks", Germany.-
- [7] Munir, Rinaldi . "Diktat Kuliah IF2120 Matematika Diskrit," edisi keempat, Program Studi Teknik Informatika ITB, 2006.
- [8] <http://www.differencebetween.com/difference-between-directed-and-vs-undirected-graph/> diakses pada 9 Desember 2016.
- [9] [https://en.wikipedia.org/wiki/Clique\\_problem](https://en.wikipedia.org/wiki/Clique_problem) ,Bomze et al. (1999); Gutin(2004) diakses pada 9 Desember 2016.
- [10] [https://en.wikipedia.org/wiki/Protein%E2%80%93protein\\_interaction](https://en.wikipedia.org/wiki/Protein%E2%80%93protein_interaction) diakses pada 9 Desember 2016.
- [11] Ganapathiraju MK, Thahir M, Handen A, Sarkar SN, Sweet RA, Nimgaonkar VL, Loscher CE, Bauer EM, Chaparala S (April 2016). "Schizophrenia interactome with 504 novel protein-protein interactions". *NPJ Schizophrenia*. 2: 16012
- [12] Fischer B, Sandmann T, Horn T, Billmann M, Chaudhary V, Huber W, Boutros M (2015-04-02). "A map of directional genetic interactions in a metazoan cell". *eLife*. 4.

## PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 8 Desember 2016



Harum Lokawati  
13515109