

Penerapan Algoritma String Matching pada RNA Splicing dan Translasinya ke Protein

Rici Trisna Putra - 13522026
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail (13522026@std.stei.itb.ac.id):

Abstract— Bioinformatika memainkan peran penting dalam bidang kesehatan dengan membantu mengorganisasi, memproses, dan memahami data biologi yang kompleks dan berukuran besar. Penemuan-penemuan penting dalam kesehatan tidak lepas dari kontribusi bioinformatika, terutama dalam memahami struktur molekuler seperti DNA, RNA, dan protein serta peran mereka dalam pewarisan sifat. Kemajuan teknologi sekuensing telah menghasilkan data molekuler yang melimpah, yang sulit dipahami tanpa bantuan komputer. Salah satu proses kunci adalah pembuatan protein melalui RNA (mRNA), yang dikenal sebagai Dogma Sentral biologi molekuler. Memahami proses ini memungkinkan peneliti menemukan solusi untuk penyakit genetik dan melakukan modifikasi genetik untuk tujuan tertentu.

Keywords—component; DNA; RNA; Splicing; String;Tranlasi;Transkripsi;

I. PENDAHULUAN

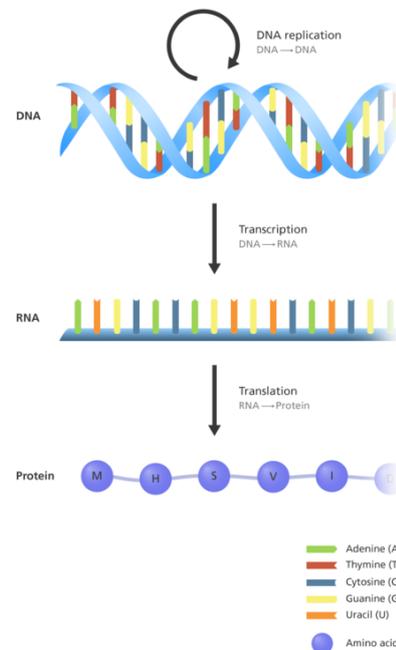
Tidak dapat dipungkiri bahwa kemajuan pada bidang kesehatan saat ini merupakan salah satu faktor penting bagi kesejahteraan hidup manusia. Keterkaitan bioinformatika dengan biologi molekuler yang merupakan bidang yang termasuk garda terdepan sains dalam bidang kesehatan saat ini membuat bioinformatika sangat menarik. Berbagai penemuan penting dalam dunia Kesehatan yang sedang ditemukan saat ini dan mengubah sejarah manusia tidak lepas dari kontribusi bioinformatika.

Salah satu peran penting dari bidang bioinformatika adalah memberikan kemudahan untuk mengorganisasi, memproses dan mencari tahu makna pada suatu data biologi yang berukuran besar dan kompleks. Hal tersebut sangat wajar tentunya karena semakin kompleksnya pemahaman kita mengenai struktur molekuler makhluk hidup seperti DNA, RNA dan Protein serta peran mereka pada pewarisan sifat. Selain itu, kemajuan teknologi sekuensing telah memberikan kita data struktur molekuler yang sangat melimpah dan sulit untuk dipahami secara manual tanpa bantuan komputer.

Salah satu fitur penting pada struktur molekuler adalah keterkaitan mereka satu sama lain dan bagaimana tiap struktur molekuler memiliki “bahasa” mereka sendiri untuk mengkodekan informasi. DNA yang merupakan bentuk struktur molekuler yang digunakan untuk pewarisan dari orang tua ke anak, tidak dapat bekerja sendiri untuk membentuk

protein (yang selanjutnya akan membentuk sifat) melainkan membutuhkan bantuan RNA sebagai perantara.

Proses pembuatan protein melalui perantara RNA (mRNA) merupakan salah satu proses yang sangat penting dalam bidang biologi molekuler karena merupakan proses penting bagi ekspresi sifat pada suatu makhluk hidup. Bahkan proses ini sampai disebut sebagai Dogma sentral (*Central Dogma*) dari bidang biologi molekuler. Dengan menguasai proses ini dan beberapa proses yang mengikutinya (seperti RNA Splicing), maka peneliti dapat memahami bagaimana suatu DNA berubah dari suatu informasi hingga dapat diekspresikan sebagai sifat. Dari pemahaman tersebut, kita dapat mencari berbagai solusi masalah seperti penyakit dan cacat yang disebabkan oleh kelainan genetik atau bahkan modifikasi genetik untuk sifat yang sesuai pada suatu tumbuhan misalnya.



Gambar 1.1 Central Dogma Biologi Molekuler

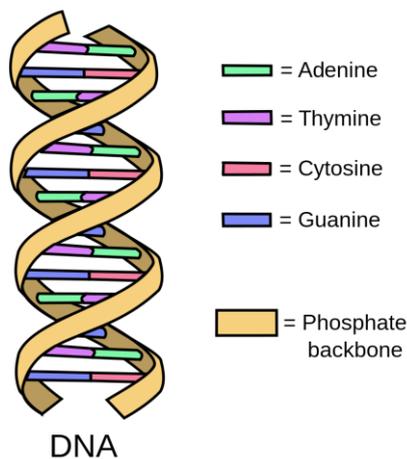
Sumber : <https://www.yourgenome.org/theme/how-is-dna-turned-into-protein-the-central-dogma-of-molecular-biology/>

Oleh karena itu, pada makalah ini penulis akan menerapkan proses pembuatan RNA dari DNA (Transkripsi) yang kemudian RNA yang disebut mRNA tersebut akan diubah ke bentuk Protein (Translasi) melalui mekanisme RNA Splicing. Semua proses di atas akan diimplementasikan dalam bahasa Python sehingga dapat memberikan pembaca wawasan bagaimana bidang Bioinformatika dapat memberikan kemudahan bagi peneliti di bidang biomolekuler untuk dapat mempercepat dan mengotomasi proses ini dan memiliki yang besar untuk kemajuan dunia kesehatan.

II. DASAR TEORI

A. DNA dan RNA

Deoxyribonucleic acid atau yang biasa disingkat DNA merupakan suatu asam nukleat yang terdiri dari susunan rantai nukleotida. DNA berbentuk helix ganda melingkar seperti terlihat pada gambar di bawah. Bentuk tersebut mengakibatkan DNA memiliki dua rantai polinukleotida yang saling terhubung oleh ikatan hidrogen dari basa nitrogen yang ada di tiap nukleotidanya yang melekat pada suatu kerangka fosfat. Setiap nukleotida yang ada di DNA terdiri dari salah satu jenis basa nitrogen yakni Adenin(A), Timina(T), Sitosin(C), atau Guanin(G). Rantai Polinukleotida yang ada pada DNA saling terhubung karena basa nitrogen Adenine berikatan hidrogen dengan Thymine sedangkan basa nitrogen Cytosine berikatan hidrogen dengan Guanine. Oleh sebab itu dua rantai polinukleotida pada DNA dikatakan sebagai saling komplemen (*Complementary*). Tentunya informasi dari salah satu rantai DNA cukup untuk membuat rantai lainnya sehingga dua rantai ini bisa dibilang membawa informasi yang sama (dalam bentuk komplemen).



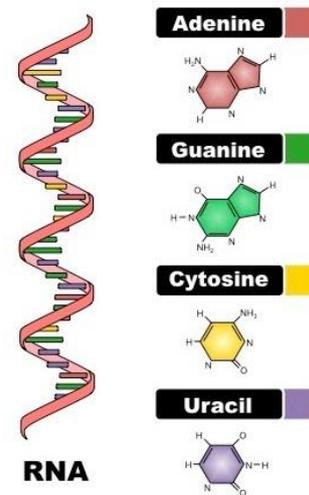
Gambar 2.1 Struktur DNA

Sumber :

https://id.wikipedia.org/wiki/Berkas:DNA_simple2.svg

Ribonucleic Acid atau biasa disingkat RNA merupakan suatu asam nukleat seperti DNA, namun RNA hanya memiliki satu buah rantai dan dihasilkan oleh proses transkripsi dari suatu rantai DNA. Selain itu, RNA juga

terdiri dari susunan rantai nukleotida seperti DNA, namun pada RNA basa nitrogen Timina (T) diubah menjadi Urasil (U). Terdapat beberapa jenis RNA yakni RNA Duta, RNA Ribosomal, dan RNA Transfer. RNA jenis RNA Duta (mRNA) merupakan tipe RNA yang dihasilkan dari proses transkripsi untuk selanjutnya digunakan oleh Ribosom untuk pembuatan protein (proses translasi)



Gambar 2.2 Struktur RNA

Sumber :

<https://www.kompas.com/skola/read/2020/10/31/235706569/rna-pengertian-struktur-dan-fungsi?page=all>

B. Transkripsi dan RNA Splicing

Transkripsi merupakan proses pembuatan RNA (Terutama mRNA) dari suatu bagian tertentu pada suatu rantai DNA. Proses transkripsi ini dilakukan oleh suatu enzim yang disebut RNA polimerase. Transkripsi dibagi menjadi tiga tahapan yakni :

- Inisiasi

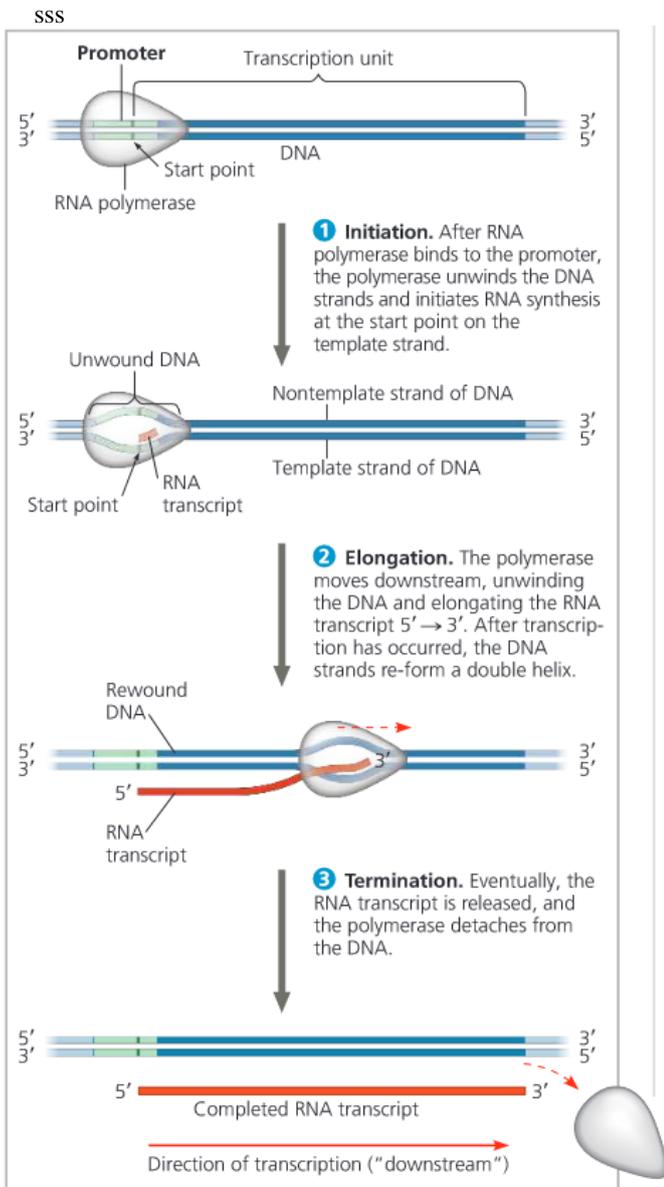
Merupakan tahap pencarian titik awal transkripsi pada suatu rantai DNA oleh RNA polimerase.

- Elongasi

yakni pembuatan untai RNA yang mengacu pada salah satu rantai DNA saja. Pemilihan rantai DNA bergantung pada RNA polimerase dan bagaimana tahap inisiasi dilakukan. Untai yang digunakan untuk melakukan tahap elongasi disebut sebagai untai template. Sedangkan untai yang tidak dipakai untuk template disebut untai koding

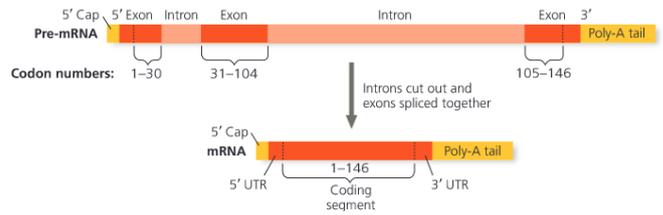
- Terminasi

Setelah Elongasi dilakukan dan ditemukan sinyal stop (ditemukan poliadenilasi) maka akan terjadi tahap terminasi dan RNA polimerase akan terlepas pada untai template. mRNA yang dihasilkan oleh proses transkripsi disebut sebagai mRNA primer karena akan dilakukan proses splicing untuk membuat RNA benar-benar siap dipakai pada proses translasi.



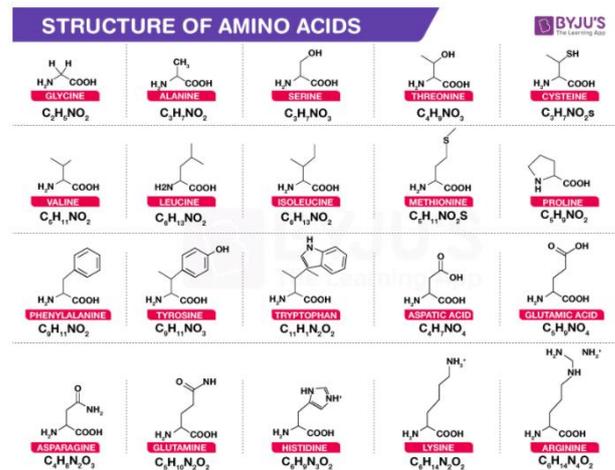
Gambar 2.3 Proses Transkripsi
Sumber : Campbell Biology

RNA Splicing merupakan suatu proses perubahan mRNA primer yang dihasilkan oleh proses transkripsi menjadi mRNA matang yang siap dikirim ke Ribosom untuk di translasi. RNA Splicing melibatkan penghapusan *semua* bagian Intron yang ada pada segmen RNA. *Intron* sendiri merupakan rangkaian nukleotida yang tidak akan ikut berperan pada pembentukan protein yang diinginkan sehingga harus dipotong dan dibuang dari mRNA. Adapun bagian yang ikut berperan pada pembentukan protein disebut sebagai *Exon*.



C. Protein

Protein merupakan suatu biomolekul yang disusun oleh satu rantai panjang asam amino yang memiliki bentuk dan fungsi tertentu. Protein merupakan molekul yang sangat penting karena proteinlah yang mengatur banyak fungsi dalam tubuh suatu makhluk hidup seperti metabolisme, melakukan aktivitas seluler, dll. Bentuk dan fungsi suatu protein bergantung pada gen yang mengkodekan protein tersebut. Protein dibentuk pada proses translasi RNA oleh Ribosom. Pada makhluk hidup dikenal 20 asam amino standar yang digunakan untuk membangun semua protein yang dibutuhkan makhluk hidup tersebut.



Gambar 2.4 Asam Amino Dasar

Sumber : <https://byjus.com/chemistry/amino-acid-structure/>

D. Translasi dan Kodon

Seperti yang sudah dijelaskan sebelumnya, translasi merupakan suatu proses pembentukan protein oleh Ribosom menggunakan mRNA sebagai templatnya. Hasil dari translasi ini adalah rantai asam amino yang sesuai dengan sekuens nukleotida yang ada pada RNA. Rantai asam amino tersebut kemudian akan melipat dan berubah menjadi protein. Pada proses translasi, satu jenis protein dikodekan oleh tiga nukleotida bersebelahan yang biasa disebut kodon sehingga ribosom akan membaca tiga nukleotida berurutan dan membuat satu asam amino untuk tiap kodon tersebut. Karena terdapat 64 kombinasi yang mungkin untuk tiga kodon maka terdapat sebuah tabel kodon yang dapat memetakan suatu kodon ke sebuah protein. Namun karena hanya terdapat 20 jenis protein

dasar maka terdapat beberapa kodon berbeda yang mengkodekan suatu protein yang sama.

UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp
CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }
AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }
GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }

Gambar 2.5 Tabel Kodon

Sumber : <https://byjus.com/chemistry/amino-acid-structure/>

Proses translasi juga dibagi menjadi tiga bagian seperti proses transkripsi yakni:

- Inisiasi
Ribosom akan mencari start kodon yakni kodon AUG (Methylamine) dan memulai translasi untuk tiap kodon.
- Elongasi
Ribosom akan menambahkan asam amino untuk tiap kodon yang dibaca sesuai dengan yang ada di tabel kodon.
- Terminasi
Ketika stop kodon ditemukan, maka ribosom akan berhenti menambahkan asam amino dan protein yang sudah jadi akan dilepas.

E. String Matching

String matching merupakan pencarian semua kemunculan suatu pola pada sebuah string (teks). Pola merupakan suatu string yang ukurannya lebih kecil atau sama dengan string yang akan dicocokkan dengan string matching. Adapun persoalan string matching dapat disederhanakan sebagai berikut:

1. T : teks, yang merupakan string dengan panjang n
2. P : pattern, yakni suatu string dengan panjang m dengan asumsi $m \ll n$ yang akan dicari di T

maka masalah di string matching adalah menemukan lokasi pertama atau semua lokasi kemunculan pattern di dalam teks.

Terdapat dua klasifikasi dari algoritma string matching yakni:

- Left to Right
Merupakan algoritma yang mencari dari bagian kiri yang menelusur ke kanan. Contoh algoritma string matching left to right adalah algoritma brute force dan algoritma Knuth-Morris-Pratt (KMP).
- Right to Left
Merupakan algoritma yang mencari dari bagian kanan terlebih dahulu kemudian menelusur ke kiri. Contoh algoritma string matching right to left adalah algoritma Boyer – Moore (BM).

F. Algoritma Brute Force

Merupakan algoritma naif left to right yang mengecek satu persatu huruf pada pattern dan text kemudian apabila terjadi suatu mismatch akan bergeser satu posisi ke kanan.

```

NOBODY NOTICED HIM
1 NOT
2 NOT
3 NOT
4 NOT
5 NOT
6 NOT
7 NOT
8 NOT

```

Gambar 2.6 Pencarian Algoritma Brute Force

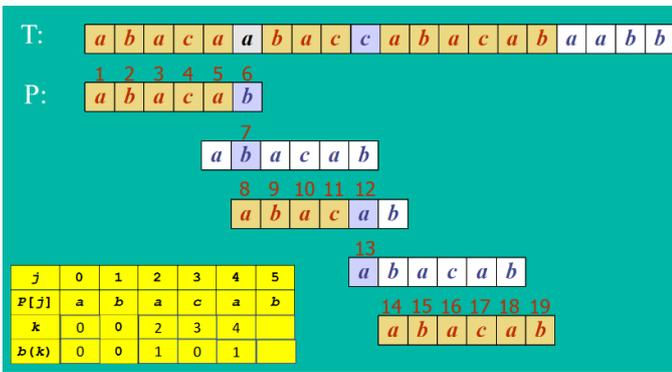
Sumber :

<https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2020-2021/Pencocokan-string-2021.pdf>

G. Algoritma KMP (Knuth-Morris-Pratt)

Algoritma KMP merupakan contoh algoritma left to right yang memiliki prinsip sama dengan algoritma brute force akan tetapi algoritma KMP memiliki mekanisme untuk menggeser semaksimal mungkin pola apabila terjadi suatu mismatch dengan teks.

Sebelum pengecekan, KMP memerlukan preprocessing pola dengan membuat suatu border function untuk pola tersebut. Border function ini dicari dengan mencari jumlah huruf tertinggi prefix dan suffix yang sama. Setelah mendapatkan border function untuk tiap posisi mismatch j maka selanjutnya pengecekan string dapat dimulai satu persatu layaknya algoritma Brute Force. Apabila ditemukan mismatch maka algoritma akan mengubah posisi j pola sesuai nilai yang ada pada border function dan posisi j pada pola saat mismatch terjadi. Karena memiliki mekanisme pergeseran yang lebih cerdas dari algoritma Brute Force, maka algoritma KMP memiliki kompleksitas waktu yang lebih efisien.



Gambar 2.7 Pencarian Algoritma KMP

Sumber :

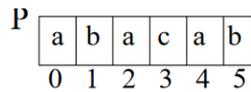
<https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2020-2021/Pencocokan-string-2021.pdf>

H. Algoritma BM (Boyer-Moore)

Algoritma Boyer-Moore merupakan salah satu algoritma string matching yang memiliki pendekatan pencarian dari kanan ke kiri (right to left) yakni dari akhir pola sampai ke awal pola. Terdapat 3 kasus yang mungkin terjadi pada pengecekan dengan algoritma ini, dan untuk tiap kasus terdapat strategi berbeda yang dapat dilakukan.

Sama seperti Algoritma KMP, terdapat preprosesing pola yang harus dilakukan sebelum pengecekan dimulai yakni mencari last occurrence functionnya terlebih dahulu. Konsep last occurrence function ini adalah untuk setiap huruf yang ada pada teks maka dicari posisi kemunculan terakhirnya dalam pola dan kemudian dicatat. Untuk tiap huruf yang ada di teks namun tidak muncul pada pola maka akan diberikan nilai -1.

- A = {a, b, c, d}
- P: "abacab"



x	a	b	c	d
L(x)	4	5	3	-1

Gambar 2.8 Last Occurrence Function Algoritma BM

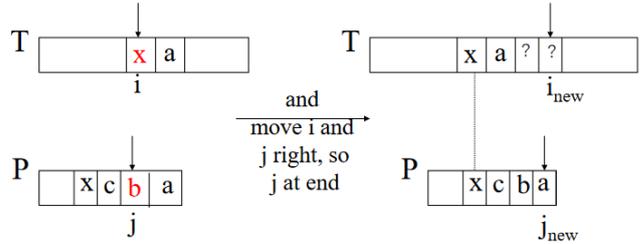
Sumber :

<https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2020-2021/Pencocokan-string-2021.pdf>

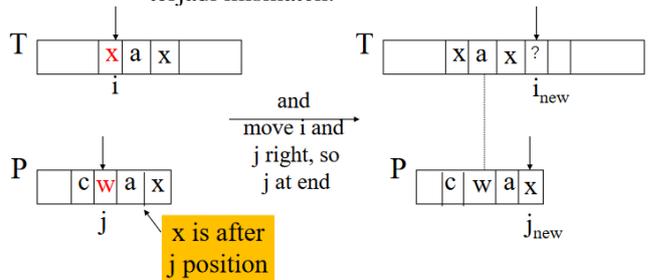
Selanjutnya seperti yang sudah dijelaskan pada bagian sebelumnya, terdapat 3 kasus berbeda yang dapat terjadi saat pencarian menggunakan algoritma BM yakni:

1. Jika terdapat mismatch pada teks dengan posisi i dan pola pada posisi j dan huruf yang mismatch pada teks adalah huruf x dan terdapat huruf x pada bagian kiri pola maka i akan digeser hingga match dengan huruf x pada pola yang berada di bagian kiri tadi. Adapun rumusnya adalah $i = i +$

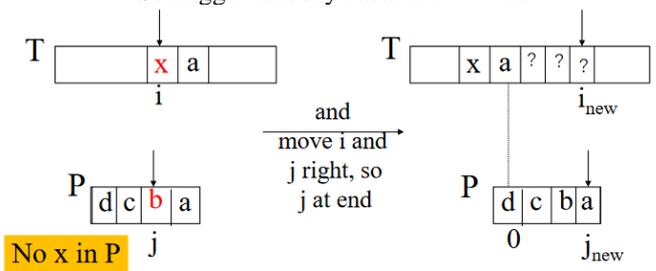
$m - (lo + 1)$ dengan m panjang pola dan lo adalah last occurrence dari huruf x.



2. Apabila terjadi mismatch seperti kasus 1 namun huruf x hanya terdapat di bagian kanan mismatch maka pergeseran i sesuai dengan rumus $i = i + m - j$ dengan m panjang pola dan j indeks pola saat terjadi mismatch.



3. Apabila terjadi mismatch seperti kasus 1 namun huruf x tidak terdapat pada pola maka i digeser sejauh panjang pola yakni m agar melewati pencocokan huruf yang tidak ada di pola. Sehingga rumusnya adalah $i = i + m$.



III. IMPLEMENTASI DAN PEMBAHASAN

Pada Bagian ini akan dijelaskan bagaimana proses pemanfaatan algoritma String Matching dalam proses Pembuatan RNA dari DNA template dan kemudian melakukan RNA Splicing dan hasil RNA Splicing akan diubah ke Protein.

Program yang diimplementasikan dalam bahasa Python. Program pertama-tama akan melakukan pengubahan sekuens DNA ke RNA dengan mengubah Timina pada untai koding dengan Urasil.

```
def transcription(sequence):
    return sequence.replace("T", "U")
```


V. UCAPAN TERIMA KASIH

Puji Syukur kepada Allah SWT yang telah memberikan Rahmat sehingga penulis dapat menyelesaikan makalah ini dengan tepat waktu yang tentunya keberhasilan ini tidak lepas dari kontribusi berbagai pihak.

Pertama-tama penulis mengucapkan terima kasih yang sebesar-besarnya kepada dosen pengampu Mata Kuliah IF2211 Strategi Algoritma K-02 ibu Nur Ulfa Maulidevi, S.T, M.Sc. Karena atas bimbingan beliau selama ini saya berhasil menyelesaikan mata kuliah ini dengan baik. Tentunya arahan dan ajaran beliau sangat memberikan dampak bagi diri saya untuk bisa lebih baik kedepannya.

Selain itu penulis juga berterima kasih kepada kedua orang tua penulis yang telah mendidik dan membesarkan penulis hingga sampai saat ini.

Terakhir penulis sadar bahwa terdapat berbagai kesalahan pada tulisan ini sehingga apabila terdapat kritik dan saran maka penulis sangat mengharapkan untuk dapat disampaikan.

LINK GITHUB

<https://github.com/RiciTrisnaP/makalahStima>

REFERENCES

- [1] <https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2020-2021/Pencocokan-string-2021.pdf> Diakses 12 Juni 2024
- [2] <https://www.genome.gov/genetics-glossary/Deoxyribonucleic-Acid> Diakses 12 Juni 2024
- [3] <https://www.britannica.com/science/RNA>. Diakses 12 Juni 2024
- [4] <https://www.ncbi.nlm.nih.gov/gene/3630>. Diakses 12 Juni 2024

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, 12 Juni 2024



Rici Trisna Putra
13522026