75079

A STUDY ON BIOMOLECULAR SEQUENCE ALIGNMENT USING MACHINE LEARNING TECHNIQUES

MUHAMAD RAZIB OTHMAN
NAOMIE SALIM
ROZITA ABDUL JALIL
SAFAAI DERIS
SAFIE MAT YATIM
ROSLI MD ILLIAS

FAKULTI SAINS KOMPUTER DAN SISTEM MAKLUMAT
UNIVERSITI TEKNOLOGI MALAYSIA

2004

ABSTRAK

Penjajaran jujukan berpasangan digunakan untuk membandingkan jujukan *nucleotides* atau protein dengan tujuan untuk mengetahui struktur, fungsi dan hubungan evolusi yang wujud bagi jujukan yang dikaji. Matlamat utama bagi penjajaran jujukan adalah untuk mencari jajaran yang optimal. Kaedah yang sering digunakan dalam penyelidikan dan diakui dapat menghasilkan jajaran yang optimal ialah kaedah Pengaturcaraan Dinamik *Smith-Waterman* bagi jajaran setempat. Berdasarkan penyelidikan terdahulu, skema permarkahan yang terdapat dalam pengaturcaraan dinamik boleh diperbaiki dengan menggunakan matriks penggantian dan memperkenalkan jurang dengan fungsi jurang penalti. Ianya bertujuan untuk mengoptimumkan hasil jajaran disamping mengekalkan konsep biologi iaitu wujudnya perubahan evolusi dalam biomolekul disebabkan mutasi. Tetapi sehingga kini, tiada teori umum yang memberikan panduan bagi pemilihan matriks penggantian dan jurang penalti bagi jajaran jujukan setempat. Oleh kerana itu projek ini akan mengimplementasi algorithma pengaturcaraan dinamik *Smith-Waterman* dengan menggunakan parameter matriks penggantian dan fungsi jurang penalti yang berbeza dalam skema permarkahan. Matriks penggantian yang akan digunakan ialah BLOSUM45, BLOSUM62 dan BLOSUM80. Manakala fungsi jurang penalti linear dengan julat nilai parameter dari (–d=1 hingga –d=10) dan jurang penalti *affine* dengan julat nilai parameter bukaan jurang dari (–d=1 hingga -d=12) dan tambahan jurang dari (–e=1 hingga –e=5). Perbandingan secara intensif akan dilakukan bagi menguji keberkesanan dan menentukan parameter matriks penggantian dan jurang penalti yang efektif bagi jajaran jujukan. Jajaran jujukan dilakukan terhadap 27 set data jujukan protein yang dikategorikan mengikut ukuran panjang dan peratusan kesamaan identiti. Hasilnya adalah panduan pemilihan parameter matriks penggantian dan jurang penalti yang efektif bagi penjajaran jujukan.

ABSTRACT

Pairwise sequence alignment is used to compare the sequence of *nucleotides* or protein with the aims of inferring structural, functional and evolutionary relationships. The main reason of sequence alignment is to find an optimal alignment. The most used method in research and have been certify to produce an optimal sequence alignment are dynamic programming methods *Smith-Waterman* for local alignment. Based from the previous research, scoring schemes in dynamic programming can be improved by using substitutions matrices and introduction of gap in alignment with gap penalty function. The reasons are to optimize result of alignments with perpetuate biology concept like evolution changes in molecular structures caused by mutation. Today, no general theory guides the selection of substitution matrices and gap penalties for local sequence alignment. Because of that, this project will implement dynamic programming method *Smith-Waterman* with different parameter of substitution matrices and gap penalty function in scoring schemes. Substitution matrices that will be used are BLOSUM45, BLOSUM62 and BLOSUM80. While linear gap penalty with range values parameter from (–d=1 to –d=10) or affine gap penalty with range values parameter for opening gap from (–d=1 to –d=12) and extension gap from (–e=1 to–e=5). Intensive comparison will be done to test the efficiency and determine the effective substitution matrices and gap penalty parameter for sequence alignment. 27 sets of data protein sequences categorized by length and percentage similarity identity will be used for sequence alignment. The results will give the guideline for the selection of effective substitution matrices and gap penalty parameter for sequence alignment.

KANDUNGAN

SENARAI RAJAH

SENARAI SIMBOL

SIMBOL		PENERANGAN
α	-	Matriks penggantian BLOSUM
β_{-d}	-	Fungsi jurang penalti linear
$\delta_{-d,-e}$	-	Fungsi jurang penalti affine
q	-	Frekuensi
p	-	Kebarangkalian (*probability)*
$-d$	-	Nilai penalti pembukaan jurang
$-e$	-	Nilai penalti penambahan jurang
Subs	-	Matriks penggantian
J_x	-	Jadual hasil
S	-	Jujukan pendek
M	-	Jujukan sederhana
L	-	Jujukan panjang
a	-	Peratusan kesamaan <25%
b	-	Peratusan kesamaan 20%-40%
c	-	Peratusan kesamaan >35%
μ	-	Mean
σ	-	Sisihan piawai
σ^2	-	Varian

SENARAI SINGKATAN

SINGKATAN		PENERANGAN
3D	-	3 dimensi
AL	-	*alignment length*
BAliBASE	-	*benchmark alignment database*
BLOSUM	-	*blocks subsitution matrix*
BP	-	*base pairs*
CA	-	*correct alignment*
DNA	-	*deoxyribonucleic acid*
MAX	-	maksimum
OM	-	*optimal mark*
PAM	-	*point accepted mutation*
RZ-score	-	*reform of Z-score*
SW	-	Smith-Waterman
UTM	-	Universiti Teknologi Malaysia

SENARAI DAFTAR ISTILAH

ISTILAH		TRANSLASI
asid amino	-	*amino acid*
banyak pasangan	-	*multiple sequence*
denda	-	*penalized*
jarak hubungan	-	*distant relationship*
jujukan	-	*sequence*
jujukan yang berkait rapat	-	*closely related sequences*
jujukan berjarak rapat	-	*distantly related sequences*
jurang	-	*gap*
jurang penalti	-	*gap penalty*
jurang penalti linear	-	*linear gap penalty or cost*
jurang penalti affine	-	*affine gap penalty or cost*
kadar evolusi	-	*evolution rate*
kadar kemunculan	-	*ratio of appearance*
kebarangkalian bagi kejadian	-	*probability of occurrence*
kenyataan kesamaan	-	*equation*
kesamaan setempat	-	*local similarity*
kodon	-	*codon*
komputasi biologi	-	*computational biology*
markah kesamaan optimal	-	*optimal mark*
matriks penggantian	-	*substitution matrices*
matriks jarak mutasi minimum.	-	*minimum mutation distance matrix*
padanan jajaran	-	*correct alignment*

panjang jajaran	-	*alignment length*
pasangan jujukan	-	*pairwise sequence*
pelupusan	-	*deletion*
penalti tambahan jurang	-	*gap extension penalty*
penalti pembukaan jurang	-	*gap opening penalty*
penambahan	-	*insertion*
pencapahan	-	*divergence*
pendaraban tukaran matriks	-	*matrix-chain multiplication*
penentuan berjujukan	-	*sequential decision*
pengaturcaraan dinamik	-	*dynamic programming*
pengecaman pertuturan	-	*speech recognition*
pengisian matriks	-	*matrix fill / tabular computation*
penilaiawalan	-	*initialization or recurrent relation*
penjadualan himpunan laluan	-	*assembly-line scheduling*
penjajaran jujukan	-	*sequence alignment*
penjajaran setempat	-	*local alignment*
penjajaran global	-	*global alignment*
penjejakan balik	-	*traceback*
penuding	-	*pointer*
peratusan kesamaan identiti	-	*precentage similarity identity*
permarkahan kesamaan kimia	-	*chemical similarity scoring*
permarkahan kod genetik	-	*genetic code scoring*
pertuturan berdigit	-	*digitized speech*
piawai	-	*standard*
serpihan	-	*fragment*
substruktur optimal	-	*optimal substructure*
tindanan submasalah	-	*overlapping subproblem*

SENARAI LAMPIRAN

BAB 1

PENGENALAN PROJEK

1.1 Pendahuluan

Penemuan struktur DNA *(deoxyribonucleic acid)* pada tahun 1953 telah membawa impak besar terhadap perkembangan dunia biologi. Ianya telah membuka satu lembaran baru kepada penerokaan dunia sains yang unik dan menarik. Kini, ahli biologi giat mentafsir struktur DNA bagi setiap bentuk hidupan yang ditemui di muka bumi ini. Hasilnya adalah jumlah data yang luar biasa yang perlu dianalisis. Maka, tidak menjadi kesangsian lagi kenapa pada dekad kini ramai sainstis dan pakar komputer tertarik untuk membangunkan penyimpanan dan capaian maklumat *(information storage and retrieval*) serta kaedah-kaedah analisis bagi mentafsir data-data biologi.

Percantuman di antara bidang biologi dan sains komputer mencipta satu peraturan di antara lapangan yang dikenali sebagai komputasi biologi *(computational biology)*[13] yang menerokai bagaimana kapasiti komputer menerima atau mengekstrak pengetahuan dari data biologi. Penyelidik boleh mempelajari berkaitan jujukan biomolekul dengan membandingkannya terhadap jujukan yang sudah dikaji. Oleh kerana itu perbandingan jujukan merupakan satu masalah asas atau utama bagi

komputasi biologi, di mana ianya selalu diselesaikan dengan kaedah yang dikenali sebagai penjajaran jujukan *(sequence alignment)* [27].

Penjajaran jujukan merupakan perbandingan dan penyusunan dua atau lebih input bagi jujukan, sama ada untuk mengira kesamaan di antara jujukan tersebut atau untuk mencari jujukan induk yang mana setiap input bagi jujukan berkongsi kriterianya. Penjajaran jujukan merupakan peralatan penting yang digunakan secara meluas dalam pelbagai aplikasi sainstifik [9]. Contohnya dalam bidang molekul biologi, jujukan dibandingkan di antara protein dan *nucleotides* manakala dalam bidang geologi [26], ianya mengambarkan struktur *stratigraphic* bagi persampelan utama dan dalam bidang pengecaman pertuturan (*speech recognition*) ianya merupakan sampel bagi pertuturan berdigit (*digitized speech*).

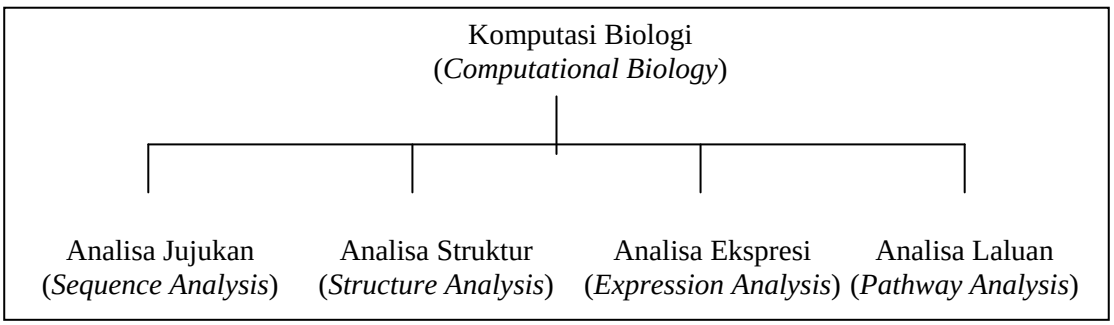
Penyelidikan ini berkaitan kaedah untuk melakukan perbandingan terhadap jujukan biomolekul, difokuskan kepada penjajaran setempat *(local alignment)* dan pasangan jujukan *(pairwise sequence).* Secara umumnya, bab ini akan memberikan gambaran ringkas tentang keseluruhan penyelidikan yang dilakukan. Bermula dengan latarbelakang bagi masalah, motivasi, matlamat, objektif dan skop bagi penyelidikan. Perincian lanjut boleh diperolehi dalam bab 2 dan 3.

1.2 Latarbelakang Masalah dan Cabaran Penjajaran Jujukan

Perkembangan pesat dalam bidang biologi dengan penemuan biomolekul baru menyebabkan pertambahan pangkalan data genome yang mendadak [4, 28]. Para saintis terdahulu dalam bidang biologi telah melakukan penyelidikan terhadap struktur biomolekul ini dengan menterjemahkannya ke bentuk jujukan biomolekul. Jujukan ini diwakilkan dengan rentetan aksara yang mana setiap aksara telah dipiawaikan yang membawa maksud tertentu dan dikenali sebagai kod genetik

[6,13]. Ianya bertujuan bagi memudahkan proses analisa dan kajian terperinci dilakukan terhadap biomolekul tanpa melibatkan bahan atau jisim tersebut. Jujukan biomolekul mewakili set lengkap bagi organisma hidup yang mana boleh terdiri dari dua iaitu jujukan DNA atau *nucleotides* dan jujukan protein atau asid amino [13]. Oleh itu analisa jujukan boleh dilakukan sama ada terhadap jujukan *nucleotides* atau *asid amino.* Penjelasan terperinci berkaitan jujukan biomolekul boleh diperoleh dalam Bab 2 dan Lampiran A.

Analisa jujukan merupakan salah satu cabang utama dalam 4 cabangan komputasi biologi di antaranya ialah analisa jujukan, analisa struktur, analisa ekspresi dan analisa laluan [37]. Sila rujuk Rajah 1.1. Penjajaran jujukan yang merupakan sub topik atau masalah dalam cabangan analisa jujukan, merupakan peralatan komputasi yang penting bagi melakukan analisa terhadap jujukan DNA dan protein dalam era biomolekul moden ini. Penjajaran jujukan digunakan bagi membandingkan jujukan dengan tujuan untuk mendapatkan struktur, fungsi dan hubungan evolusi yang wujud bagi jujukan yang dikaji. Sebagai contoh jika satu jujukan baru ditemui, saintis akan melakukan proses penjajaran terhadap jujukan tersebut dengan jujukan yang telah diketahui kefungsiannya. Ia bertujuan untuk meramalkan fungsi bagi jujukan berdasarkan hubungan yang wujud dari hasil jajaran. Penjajaran jujukan juga merupakan asas sebelum sesuatu jujukan itu dianalisa untuk membentuk pepohon *phylogenetic* atau penentuan homolog. Lampiran B menunjukkan langkah pembentukkan pepohon *phylogenetic* dari jujukan DNA atau protein [20], di mana penjajaran jujukan merupakan langkah kedua yang perlu dilaksana untuk membina pepohon *phylogenetic.*



Rajah 1.1 : Cabang utama bagi komputasi biologi.

Penjajaran jujukan juga boleh dilakukan secara berpasangan *(pairwise sequence alignment)* atau secara lebih dari satu pasangan atau banyak pasangan iaitu *(multiple sequence alignment)* [9]. Sila rujuk Lampiran C. Ianya juga boleh terdiri dari dua jenis iaitu iaitu jajaran global *(global alignment)* dan jajaran setempat *(local alignment),* yang mana ini perlu ditentukan sebelum menjajarkan sesuatu jujukan [9]. Sila rujuk Lampiran D dan penjelasan lanjut dalam bab 2. Matlamat penjajaran jujukan adalah untuk memadankan jujukan dengan memaksimakan padanan yang sama dan meminimakan padanan yang tidak sama atau dalam erti kata lain untuk mendapatkan jajaran yang optimal [37, 40].

Oleh kerana penjajaran jujukan ini merupakan proses yang rumit jika melibatkan jujukan yang panjang dan mengambilkira kewujudan mutasi iaitu boleh berlakunya penambahan, pembuangan dan penggantian dalam jujukan, maka keperluan kepada pengautomasian dengan kaedah yang efektif amat diperlukan bagi menyelesaikan masalah ini [39]. Hasil dari penyelidikan oleh para penyelidik terdahulu, pelbagai kaedah pembelajaran mesin dibangunkan bagi tujuan penjajaran jujukan seperti *Brute-Force* [15], *Rabin-Karp* [32], *Dot Matrik* [16] dan *Dynamic Programming* yang terdiri dari *NeedleMan-Wunsch* [27] dan *Smith-Waterman* [36].

Berdasarkan kajian dari hasil penyelidikan para penyelidik yang terdahulu terhadap penjajaran jujukan biomolekul, didapati kaedah Pengaturcaraan Dinamik merupakan kaedah yang sering digunakan dan diakui dapat menghasilkan jajaran yang optimal iaitu kaedah *NeedleMan-Wunsch* bagi penjajaran global [27] dan *Smith-Waterman* bagi penjajaran setempat [38]. Oleh kerana penjajaran setempat amat diperlukan bagi pencarian pangkalan data dan banyak digunakan oleh saintis biologi [3], maka projek ini akan memfokuskan kepada jajaran berpasangan setempat bagi jujukan protein. Algorithma *Smith-Waterman* akan diimplementasikan kerana kesesuaiannya bagi penjajaran setempat [3, 38]. Berdasarkan journal-journal yang dikaji, sehingga ke hari ini para penyelidik masih terus menyelidik kaedah penjajaran jujukan bagi mendapatkan jajaran yang optima dan kebanyakannya menggunakan algorithma *Smith-Waterman* khususnya bagi kes jajaran setempat [25, 31, 36].

Beberapa penemuan dan cadangan dari penyelidik terdahulu yang tujuan memperbaiki jajaran jujukan ialah dengan penggunaan matriks penggantian PAM (*Point Accepted Mutation*) [10] dan BLOSUM (*Blocks Subsitution Matrix*) [19] dalam skema permarkahan bagi pengaturcaraan dinamik, serta memperkenalkan jurang dalam jajaran dan cadangan pengiraan jurang penalti [2, 5, 17]. Secara umumnya kesemua penyelidikan dan cadangan tersebut bertujuan memperbaiki algoritma pengaturcaraan dinamik yang asal khususnya terhadap skema permarkahan bagi jajaran.

1.3 Motivasi

Berdasarkan kajian yang telah dijalankan terbukti DNA boleh menentukan wujudnya hubungan di antara suatu organisma dengan organisma yang lain. Penjajaran jujukan diperlukan bagi meramalkan fungsi, struktur dan hubungan evolusi yang wujud bagi jujukan yang dikaji. Huraian lanjut adalah berkaitan permasalahan yang wujud dalam penjajaran jujukan berdasarkan penyelidikan yang terdahulu [1, 2, 16].

1.3.1 Menyusun Jajaran Optimal

Permasalahannya adalah bagaimana untuk mendapatkan jajaran yang optimal iaitu memaksimakan padanan jajaran yang sama dan meminimum padanan yang tidak sama, iaitu menjajarkan satu jujukan x terhadap satu jujukan y bagi mendapatkan susunan dan mewujudkan hubungan yang sama pada aksara. Oleh kerana wujudnya perbezaan panjang di antara dua jujukan, kewujudan jurang, penambahan jujukan, pelupusan jujukan dan pengosongan akan menyebabkan

penjajaran jujukan menjadi lebih rumit. Selain itu terdapat lebih dari satu padanan jajaran akan terhasil dari satu jajaran jujukan, maka timbul masalah tentang bagaimana untuk mendapatkan jajaran yang paling optimum bilangan kesamaannya. Kaedah pengaturcaraan dinamik telah terbukti berkesan bagi membantu masalah ini [35, 36, 39].

1.3.2 Memperbaiki Skema Permarkahan

Setiap kaedah penjajaran jujukan memerlukan skema permarkahan bagi mengira nilai padanan dan tidak padanan, begitu juga dalam kaedah pengaturcaraan dinamik. Sebagai contoh, markah akan diumpukkan bagi setiap posisi dalam jujukan bergantung kepada padanan bagi posisi tersebut. Markah bagi semua posisi dalam jajaran kemudiannya akan ditambah untuk mendapatkan jumlah markah. Ini digunakan bagi menentukan jajaran yang optimal di antara jajaran alternatif. Skema permarkahan mudah adalah dengan mengumpukkan satu nilai bagi padanan dan satu nilai bagi tidak padanan. Matriks permarkahan sebegini dikenali sebagai matriks unitari.

Bagi jajaran nucleotide, matriks permarkahan unitari sudah memadai. Secara umumnya, perubahan atau peristiwa mutasi dalam jujukan asid amino lebih bermaklumat berbanding perubahan dalam jujukan nucleotide. Ini kerana kefungsian protein dan kemungkinan wujud hubungan secara terus kepada warisan keturunan. Oleh itu, terdapat dua kriteria yang perlu diambil kira bagi memperbaiki skema permarkahan iaitu mengukur perubahan evolusi dan mengawal jurang. Seterusnya adalah merupakan perincian berkaitan dua kriteria tersebut.

(i) Mengukur Perubahan Evolusi

Informasi genetik yang berubah mengikut masa dinamakan mutasi [35]. Terdapat tiga cara bagaimana mutasi boleh berlaku iaitu:-

a) Penambahan asid amino atau *nucleotides*
b) Pelupusan asid amino atau *nucleotides*
c) Penggantian bagi satu *nucleotides* dengan yang lain.

Maka, matriks penggantian akan digunakan dalam permarkahan jajaran kerana ianya dapat mengukur yang mengambil kira perubahan evolusi tersebut.

(ii) Mengawal Jurang

Untuk mendapatkan jajaran yang optimal atau padanan yang baik, penambahan atau perlupusan aksara jujukan dalam jajaran dilakukan. Kebiasannya dalam keadaan sebenar, penambahan dan pembuangan bagi sub jujukan dinamakan sebagai peristiwa mutasi. Satu mutasi yang berlaku boleh menyebabkan wujudnya jurang yang mempunyai saiz yang berlainan. Jurang merupakan ruang kosong yang terdapat dalam jujukan bagi membolehkan jajaran. Jumlah keseluruhan jurang semasa menjajarkan dapat dikaitkan dengan kos mutasi. Oleh itu fungsi jurang penalti akan digunakan bagi pengiraan jurang dalam jajaran. Penggunaan matriks penggantian dan jurang penalti dalam skema permarkahan pengaturcaaan dinamik dapat menghasilkan jajaran yang optima.

1.4 Penyataan Masalah

Penyelidik Reese dan Pearson [31] menyatakan, sehingga kini tiada teori umum yang memberikan panduan bagi pemilihan matriks penggantian dan jurang penalti bagi jajaran jujukan setempat. Oleh itu projek ini akan mengimplementasi algorithma pengaturcaraan dinamik *Smith-Waterman* dengan menggunakan jurang penalti dan matriks penggantian yang berbeza dalam skema permarkahan jajaran. Seterusnya, perbandingan secara intensif akan dilakukan bagi menguji keberkesanannya dan menentukan parameter matriks penggantian dan jurang penalti yang efektif bagi jajaran.

1.5 Matlamat

Menentukan kombinasi parameter matriks penggantian dan jurang penalti (linear dan affine) yang efektif bagi pengaturcaraan dinamik Smith-Waterman untuk penjajaran jujukan protein.

1.6 Objektif

Objektif yang dikenalpasti untuk penyelidikan ini ialah :-

(i) Merekabentuk dan memformulasikan skema permarkahan dalam pengaturcaraan dinamik Smith-Waterman yang asal dengan menggunakan matriks penggantian dan jurang penalti yang berbeza.

(ii) Membangunkan dan melaksanakan model pengaturcaraan dinamik Smith-Waterman yang diubahsuai untuk penjajaran jujukan.

(iii) Menganalisa keberkesanan dan menentukan parameter matriks penggantian dan jurang penalti linear yang efektif dalam pengaturcaraan dinamik Smith-Waterman.

(iv) Menganalisa keberkesanan dan menentukan parameter matriks penggantian dan jurang penalti affine yang efektif dalam pengaturcaraan dinamik Smith-Waterman.

1.7 Skop

Skop bagi penyelidikan ini merangkumi perkara-perkara berikut :-

(i) Kaedah penjajaran jujukan hanya difokuskan kepada pasangan setempat sahaja.

(ii) Menggunakan algorithma pengaturcaraan dinamik iaitu *Smith-Waterman*.

(iii) Menggunakan matriks penggantian BLOSUM (*Blocks Subsitution Matrix)* kerana ianya sesuai untuk penjajaran setempat [19, 20]. Tiga jenis matriks penggantian BLOSUM yang digunakan ialah BLOSUM45, BLOSUM62 dan BLOSUM80.

(iv) Menggunakan fungsi jurang penalti linear (*linear gap penalty)* dengan julat nilai –d =1 hingga –d=10 dan jurang penalti affine (*affine gap penalty)* dengan julat nilai jurang pembukaan –d =1 hingga –d=12 dan jurang tambahan –e=1 hingga –e=5.

(v) Penjajaran jujukan dilakukan hanya pada set data jujukan protein sahaja.

(vi) Penganalisaan keputusan dan perbandingan keberkesanan dilakukan dari segi permarkahan kesamaan optimal, panjang jajaran dan padanan yang terhasil mengikut kategori data jujukan iaitu ukuran panjang dan peratusan kesamaan.

1.8 Susunan Laporan

Susunan laporan ini dimulai dengan :

(i) Bab 1 merupakan pendahuluan berkaitan projek merangkumi latarbelakang masalah, motivasi, objektif, matlamat dan skop.

(ii) Bab 2 merupakan kajian literatur bagi projek yang memperincikan jujukan biomolekul, komputasi biologi, kaedah pembelajaran mesin bagi penjajaran jujukan dan pengaturcaraan dinamik Smith-Waterman berserta skema permarkahan bagi mendapatkan penjajaran optima.

(iii) Bab 3 menghuraikan lapan langkah utama bagi metodologi projek.

(iv) Bab 4 menerangkan tentang formulasi bagi model pengaturcaraan dinamik yang terdiri daripada model pengaturcaraan dinamik Smith-Waterman yang asal dan model pengaturcaraan dinamik Smith-Waterman yang diubahsuai.

(v) Bab 5 merupakan pembangunan dan perlaksanaan pengaturcaraan dinamik Smith-Waterman yang telah diubahsuai untuk penjajaran jujukan.

(vi) Bab 6 menghuraikan analisa keputusan dan perbincangan terhadap parameter matriks penggantian BLOSUM dan jurang penalti linear dalam pengaturcaraan dinamik.

(vii) Bab 7 menghuraikan analisa keputusan dan perbincangan terhadap parameter matriks penggantian BLOSUM dan jurang penalti affine dalam pengaturcaraan dinamik.

(viii) Bab 8 merupakan huraian kesimpulan kajian dan cadangan masa hadapan.

BAB 2

KAJIAN LITERATUR

2.1 Pendahuluan

Bab ini membincangkan mengenai kajian latarbelakang yang akan memberikan huraian tentang bidang yang sedia ada yang berhubungkait dengan projek ini iaitu kajian terhadap jujukan biomolekul, komputasi biologi, pembelajaran mesin, jajaran jujukan dan pengaturcaraan dinamik. Ia turut membincangkan masalah dan kemungkinan-kemungkinan penyelesaiannya.

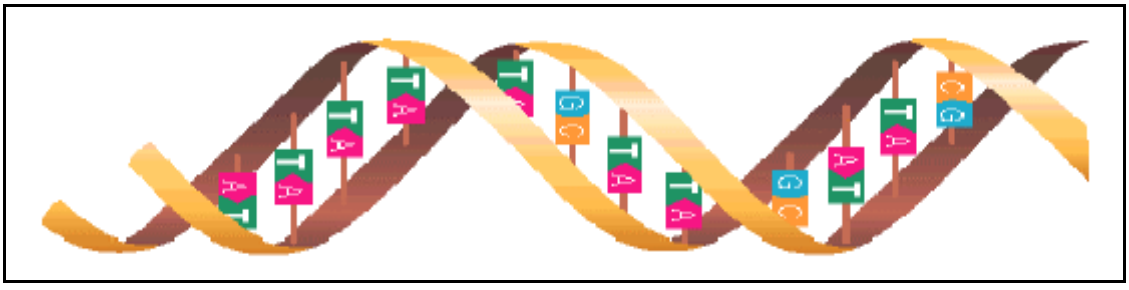
2.2 Jujukan Biomolekul

Bahagian ini akan memperincikan struktur jujukan biomolekul terutama jujukan protein yang akan digunakan dalam penyelidikan ini. Pangkalan data di mana set data jujukan protein diperolehi juga akan diterangkan.

2.2.1 Jujukan DNA dan Jujukan Protein

Genome merupakan set lengkap bagi molekul DNA dalam mana-mana organisma hidup yang akan diwarisi dari satu generasi kepada generasi yang lain. DNA boleh dianggap sebagai *"blue print of life"* kerana ianya mengkodkan segala informasi yang berkaitan untuk membentuk keperluan protein bagi semua proses bersel [13]. Selain itu ianya merupakan agen pengenalpastian untuk menguji sama ada dua hidupan itu serupa atau berbeza dari segi biologi.

DNA pada asasnya adalah rantaian berganda *(double chain)* bagi molekul mudah yang dipanggil *nucleotides*, di mana molekul ini diikat atau dihubungkan bersama dalam struktur berlingkar yang lebih dikenali sebagai *double helix*, seperti ditunjukkan pada Rajah 2.1. *Nucleotides* dibezakan oleh 4 jenis asas nitrogen yang iaitu *adenosine*, *cytosine*, *guanine* dan *thymine* [37]. Asas ini dihubungkan untuk membentuk rantaian yang mengikat *double helix* bersama. Manakala *base pairs (bp)* merupakan unit untuk mengukur panjang bagi DNA. DNA boleh ditentukan secara unik dengan menyenaraikan jujukan bagi *nucleotides*. Oleh kerana itu, untuk tujuan praktikal DNA diabstrak sebagai teks panjang yang terdiri dari 4 huruf abjad yang mewakili *nucleotides* A, C, G dan T iaitu diambil dari awalan nama bagi setiap *nucleotides*. Jujukan yang terdiri dari 4 kombinasi aksara ini dikenali sebagai jujukan DNA [6].



Rajah 2.1 : Gambaran *double helix*

Protein merupakan molekul yang menyempurnakan kebanyakkan fungsi bagi sel hidup [37], menentukan bentuknya dan struktur. Protein adalah jujukan bagi molekul mudah yang dikenali sebagai asid amino. Terdapat 20 asid amino yang berbeza yang boleh dijumpai dalam protein. Ianya dikenalpasti dengan huruf abjad atau 3 kod huruf. Sila rujuk Rajah 2.2. Sebagai contoh asid amino *alanine* diwakili dengan huruf A atau 3 kod huruf iaitu ALA.

Satu-huruf (One-letter)	**Tiga-huruf (Three-letter)**	**Nama (Name)**	**Satu-huruf (One-letter)**	**Tiga-huruf (Three-letter)**	**Nama (Name)**
A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic Asid	P	Pro	Proline
E	Glu	Glutamic Asid	Q	Gla	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	U	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

Rajah 2.2: Kod asid amino

Seperti DNA, protein boleh diwakilkan dengan rentetan huruf yang mengambarkan jujukan bagi asid amino. Ianya membentuk hubungan yang rapat di antara jujukan DNA dan jujukan protein. Untuk menghasilkan protein, sel akan membaca jumlah bagi 3 *nucleotides* dari jujukan DNA yang dinamakan kodon *(codon)* bagi menjana setiap asid amino [37].

Sebagai contoh :

Rangkai AAG yang dijumpai pada jujukan DNA yang panjang mengarahkan sel untuk membentuk asid amino *lysine*. Keserupaan di antara kodon dan asid amino ini dikenali sebagai kod genetik. Sila rujuk Rajah 2.3

Posisi Pertama	**Posisi Kedua**				**Posisi Ketiga**
	G	**A**	**C**	**T**	
G	Gly	Glu	Ala	Val	G
	Gly	Glu	Ala	Val	A
	Gly	Asp	Ala	Val	C
	Gly	Asp	Ala	Val	T
A	Arg	Lys	Thr	Met	G
	Arg	Lys	Thr	Ile	A
	Ser	Asn	Thr	Ile	C
	Ser	Asn	Thr	Ile	T
C	Arg	Gln	Pro	Leu	G
	Arg	Gln	Pro	Leu	A
	Arg	His	Pro	Leu	C
	Arg	His	Pro	Leu	T
T	Trp	Stop	Ser	Leu	G
	Stop	Stop	Ser	Leu	A
	Cys	Tyr	Ser	Phe	C
	Cys	Tyr	Ser	Phe	T

Rajah 2.3: Kod genetik yang memetakan DNA kepada asid amino

Penyelidikan ini akan menggunakan set data protein iaitu jujukan protein. Sila rujuk Lampiran A untuk mengetahui perincian kod asid amino dan *nucleotides*.

2.2.2 Pangkalan Data Jujukan

Pangkalan data jujukan menyusun dan menyimpan maklumat jujukan dalam kapasiti yang banyak, segala maklumat ini dikumpulkan dari makmal seluruh dunia

dan dilonggokkan sehingga mencapai kadar eksponen. Setiap pangkalan data mempunyai format yang spesifik. Pangkalan data ini boleh dicapai secara melayari laman webnya di internet. Terdapat tiga organisasi utama yang dipertanggungjawabkan untuk menyelenggara kebanyakan data biologi [15].

(i) *National Center for Biotechnology Information (NCBI)* di United States, divisyen bagi *National Library of Medicine (NLM)* di *National Institute of Health (NIH),* menyokong dan mengagihkan pangkalan data GENBank nucleic asid dan *GenPept CDS (Coding Sequence*) ke *National Biomedical Research Fondation* untuk diterjemah disamping menyelenggara pangkalan data *Protein Identification Resource (PIR).*

(ii) *European Molecular Biology Laboratory* menyelenggara pangkalan data nucleic asid *EMBL* dan pangkalan data jujukan protein *Swiss-Prot* yang mana turut dibantu oleh *Swiss Institute of Bioinformatics (SIB).* Selain itu pangkalan data *TrEMBL* yang diterjemah dari *EMBL,* pangkalan data jujukan protein di Cambridge, UK, Heidelberg dan Geneva. Terdapat juga pangkalan data yang kurang diketahui umum iaitu, *DNA Data Bank of Japan (DDBJ).*

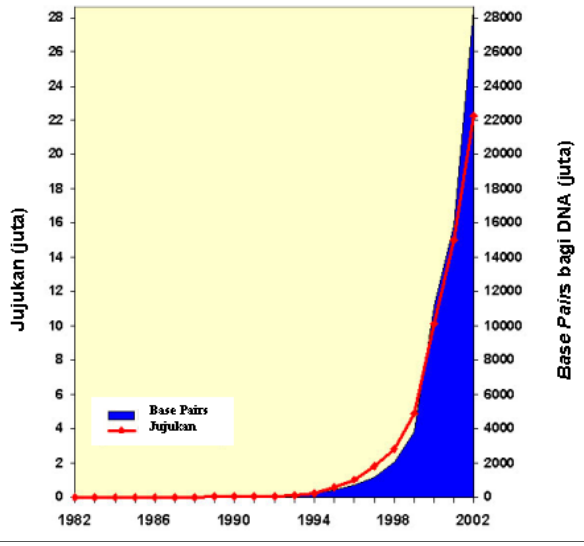
(iii) *NRL_3D* merupakan pangkalan data bagi jujukan protein berstruktur 3 dimensi dari *Protein Data Bank (PDB)* yang mana menyediakan maklumat dari *primary* (protein dalam bentuk 2 dimensi) kepada *tertiary* (protein dalam bentuk 3 dimensi).

Kebanyakan pangkalan data jujukan mengandungi data ASCII atau binari serta fail teks yang panjang dengan berbagai maklumat tentang jujukan tersebut. Fail binari memudahkan proses pemegangan bersama fail lain dengan menyediakan fungsi mengindeks. Pangkalan data *nucleic* asid dan *TrEMBL* dibahagikan kepada subdivisyen berdasarkan kepada atau sejarah pewarisannya (*taxanomy).*

Perkembangan pesat komputasi biologi memberikan kesan terhadap pertumbuhan pangkalan data jujukan genome dengan pembangunan *Human Genome*

Project dan beberapa projek genome yang menghasilkan pelbagai data. Pada Disember 2002 (GenBank version 121.0) 40 genome yang lengkap boleh diperolehi secara terbuka untuk dianalisis, tidak termasuk genome virus dan viroid yang juga boleh dimuat turun. Berdasarkan statistik pertumbuhan GenBank [46], pangkalan data GenBank meningkat 2 kali ganda setiap tahun, seperti ditunjukkan pada Rajah 2.4.

Tahun	BasePairs	Jujukan
1982	680338	606
1983	2274029	2427
1984	3368765	4175
1985	5204420	5700
1986	9615371	9978
1987	15514776	14584
1988	23800000	20579
1989	34762585	28791
1990	49179285	39533
1991	71947426	55627
1992	101008486	78608
1993	157152442	143492
1994	217102462	215273
1995	384939485	555694
1996	651972984	1021211
1997	1160300687	1765847
1998	2008761784	2837897
1999	3841163011	4864570
2000	11101066288	0106023
2001	15849921438	4976310
2002	**28507990166**	**22318883**



Rajah 2.4 : Statistik pertumbuhan GenBank 1982-2002

2.3 Komputasi Biologi

Kajian seterusnya adalah berkaitan dengan penjajaran jujukan. Ianya merupakan satu masalah dalam analisa jujukan yang merupakan salah satu cabang utama dalam komputasi biologi seperti yang telah dinyatakan sebelum ini. Huraian selanjutnya ialah berkaitan perbezaan penjajaran global dan setempat serta penjajaran berpasangan dan banyak pasangan.

2.3.1 Penjajaran Jujukan dan Kesamaan

Perbandingan jujukan boleh didefinasikan sebagai masalah untuk menentukan bahagian mana dalam suatu jujukan adalah sama dan bahagian mana yang berbeza. Ianya dianggap sebagai blok pembinaan kepada masalah yang lebih kompleks seperti penjajaran sekumpulan jujukan dan pembinaan pepohon *phylogenetic* yang mana menerangkan hubungan evolusi di antara spesis [44]. Perbandingan jujukan merupakan masalah yang diketahui umum dalam sains komputer. Bagi sainstis komputer, jujukan biomolekul merupakan salah satu sumber bagi data. Oleh kerana perkembangan pesat saiz pangkalan data biologi, algorithma yang lebih baik diperlukan [30]. Pendekatan untuk menyelesaikan masalah bagi menentukan kesamaan dan perbezaan di antara dua jujukan ialah dengan menggunakan kaedah penjajaran jujukan [27, 36]. Berdasarkan skema permarkahan yang tersusun, kesamaan boleh dikira.

Secara umumnya idea bagi menjajarkan dua jujukan yang mungkin terdiri dari saiz yang berbeza ialah dengan membariskan satu jujukan ke atas yang lain. Seterusnya memecahkannya kepada bahagian yang kecil dengan memasukkan ruang kosong pada salah satu jujukan supaya subjujukan dijajarkan dengan hubungan satu kepada satu. Kebiasaannya ruang kosong tidak dimasukkan kepada kedua-dua jujukan di lokasi yang sama. Akhirnya adalah hasil jajaran jujukan yang mempunyai saiz yang sama. Sebagai contoh dua jujukan dari protein Ferredoxin digunakan iaitu fer1_equar dan fer1_anasp. Sebahagian dari dua jujukan tersebut diambil dan diwakili dengan J1 = AYKTVLKTPS dan J2 = ATFKVTLI seperti ditunjukkan dalam rajah di bawah. Simbol “-” mewakili ruang kosong atau jurang manakala simbol “|” mewakili padanan yang sama.

```
J1 =  A-YK- TVLKTPS
      |    |   | |
J2 =  AT FKVT-L I - - -
```

Rajah 2.5 : Jajaran bagi dua jujukan

Objektifnya adalah untuk memadankan subjujukan yang sama sebanyak mungkin, dalam contoh di atas terdapat 4 padanan *(match)* bagi jajaran tersebut. Sekiranya jujukan tidak sama, wujudnya jajaran tidak padan (*mismatch)* di mana aksara berbeza dijajarkan bersama. Dua jajaran tidak padan boleh dikenalpasti dalam contoh di atas iaitu aksara “Y” pada jujukan J1 dijajarkan dengan aksara “F” pada jujukan J2, dan aksara “K” pada jujukan J1 dijajarkan dengan aksara “I” pada jujukan J2. Penambahan *(insertion)* bagi ruang kosong menghasilkan jurang *(gaps)* dalam jujukan. Ianya penting untuk mewujudkan jajaran yang baik di antara tiga aksara terakhir pada kedua-dua jujukan ini, kerana jika jurang tidak diwujudkan atau penambahan ruang tidak berlaku pada jujukan J1 maka hasil padanan jajaran semakin berkurang iaitu hanya 1 padanan sahaja.

Jajaran boleh dilihat dari cara perubahan satu jujukan terhadap jujukan lain. Ketidakpadanan boleh dianggap sebagai penggantian *(substitution)* bagi aksara. Jurang pada jujukan pertama dianggap sebagai penambahan *(insertion)* bagi aksara dari jujukan kedua kepada yang pertama. Manakala jurang yang terbentuk pada jujukan kedua dianggap sebagai pelupusan *(deletion)* bagi aksara dari jujukan pertama. Berdasarkan contoh sebelum ini, terdapat enam cara di mana J1 boleh ditukar kepada J2.

(i) Penambahan aksara “T”
(ii) Pengantian aksara “Y” kepada “F”
(iii) Penambahan aksara “V”
(iv) Pelupusan aksara “V”
(v) Pengantian aksara “K” kepada “I”
(vi) Pelupusan aksara “T”

Apabila jajaran sudah dihasilkan, markah boleh diumpukkan kepada setiap pasang aksara yang dijajarkan yang dinamakan pasangan jajaran *(aligned pair).* Permarkahan ini berdasarkan kepada skema permarkahan yang dipilih. Kebiasaanya padanan akan diberi ganjaran manakala ketidakpadanan dan jurang akan didenda *(penalized).* Secara keseluruhan markah bagi jajaran boleh dikira dengan menambah markah bagi setiap pasangan jajaran. Misalnya menggunakan skema permarkahan

mudah yang memberi nilai +2 kepada padan, –2 kepada tidak padan dan –1 kepada jurang. Sebagai contoh markah bagi jajaran yang terhasil menggunakan contoh pada Rajah 2.5 adalah $[\ 4\times(2)+2\times(-2)+3\times(-1)\ ]=1$.

Kesamaan *(similarity)* bagi dua jujukan boleh didefinisikan sebagai markah yang terbaik di antara semua jajaran yang mungkin terhasil. Ianya bergantung kepada pilihan skema permarkahan. Bahagian seterusnya memberi kupasan lanjut berkaitan skema permarkahan.

2.3.2 Perbezaan Penjajaran Global dan Penjajaran Setempat

Secara umumnya terdapat dua jenis jajaran iaitu global dan setempat. Jajaran global ialah padanan yang merangkumi keseluruhan jujukan, iaitu penjajaran jujukan dari aksara pertama bagi satu jujukan hingga aksara terakhir bagi jujukan tersebut. Ahli biologi lebih berminat dalam penjajaran pendek bagi kesamaan setempat *(local similarity)*. Dalam erti kata lain, penjajaran setempat merupakan kaedah di mana seseorang mencari jajaran terbaik di antara cebisan atau subrentetan dalam jujukan. Sebagai contoh sebahagian dari dua jujukan protein Ferredoxin digunakan iaitu fer1_equar dan fer1_anasp. Jujukan tersebut diwakili dengan J1 = AYKTVLKTPS dan J2 = ATFKVTLI yang mempunyai panjang n =10 dan m =8.

(i) Bagi jajaran global: jajaran kesemua jujukan J1 dengan kesemua jujukan J2

```
J1 =  A-YK- TVLKTPS
      |    |   | |
J2 =  ATFKVT-L I - - -
```

(ii) Bagi jajaran setempat: mencari markah kesamaan yang tertinggi bagi subjujukan dalam jujukan J1 dan J2

```
J1 =  A-YK- TVL
      |    |   | |
J2 =  ATFKVT-L
```

2.3.3 Perbezaan Penjajaran Berpasangan dan Banyak Pasangan

Jajaran boleh dilakukan sama ada secara berpasangan atau banyak pasangan. Jajaran berpasangan melibatkan hanya dua input jujukan sahaja untuk dipadankan bersama. Manakala jajaran banyak pasang melibatkan satu set input jujukan yang terdiri dari lebih dari dua jujukan untuk dipadankan bersama. Sila rujuk Lampiran C.

2.3.4 Motivasi Penjajaran Berpasangan Setempat

Penyelidikan akan memfokuskan kepada jajaran setempat sahaja dengan menggunakan input berpasangan. Terdapat banyak cara untuk menjajarkan 2 jujukan. Menggunakan contoh yang sama sebelum ini pertimbangkan jujukan pertama
J1 = AYKTVLKTPS dan jujukan kedua J2 = ATFKVTLI. Jajaran yang mungkin terhasil adalah seperti berikut :

```
AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS
ATFKVTLI--  ATFKVTL-I-  ATFKVTL--I  ATFKVT-L-I  ATFKVT-LI-

AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS
ATFKVT--LI  ATFKV-T-LI  ATFKV-TLI-  ATFKV-TL-I  ATFKV--TLI

AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS  AYKYVLKTPS
ATFK-V-TLI  ATFK-VT-LI  ATFK-VTL-I  ATFK-VTLI-  ATFK--VTLI

dan ratusan jajaran yang lain.
```

Bilangan jajaran berpasangan mungkin meningkat dengan bertambahnya panjang bagi sesuatu jujukan. Dua jujukan protein dengan panjang 100 asid amino boleh dijajarkan dalam lingkungan 10^{60} cara yang berbeza [29]. Sekiranya jajaran

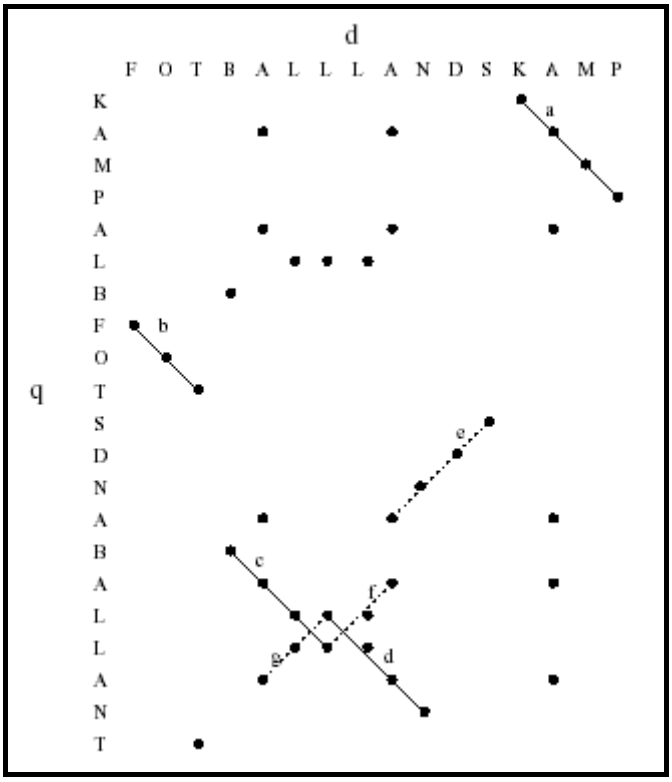
ini dijana secara manual sudah tentulah kemungkinan kesilapan akan berlaku kerana bilangan jajarannya terlalu banyak. Oleh itu persoalannya di sini ialah bagaimana proses penjajaran dapat dipermudahkan bagi mendapatkan jajaran yang paling banyak padanannya, iaitu jajaran optima.

2.4 Pembelajaran Mesin Untuk Penjajaran Jujukan

Kajian terhadap pembelajaran mesin difokuskan kepada penyelesaian penjajaran jujukan.Terdapat dua kaedah yang popular yang digunakan untuk menyelesaikan penjajaran jujukan iaitu dot matriks [14] dan pengaturcaraan dinamik *(dynamic programming)* [27, 38].

2.4.1 Dot Matriks

Kaedah pertama yang digunakan untuk menjajarkan jujukan adalah dengan menggunakan dot matriks yang juga dikenali sebagai dot plot. Matriks *m* x *n* dibina dengan asid amino *q* berada pada garisan menegak dan *d* berada di garisan melintang. Padanan bagi matriks ialah dot atau kosong. Dot merupakan sel *(i,j)* iaitu *q(i) = d(j)* seperti pada Rajah 2.6.



Rajah 2.6 : Dot matriks

Berikut merupakan kaedah bagi dot matriks:

(i) Satu jujukan dibariskan secara melintang dan satu jujukan dibariskan secara menegak di bahagian kiri.

(ii) Bergerak dari satu baris ke baris dan letakkkan dot jika menemui ruang di mana memiliki aksara yang sama.

(iii) Berterusan kepada setiap baris sehingga semua kemungkinan padanan aksara di antara jujukan diwakili dengan dot.

(iv) Baris pepenjuru bagi dot menunjukkan kesamaan jujukan.

(v) Manakala dot yang bertaburan menunjukkan kesamaan rawak iaitu tidak berkaitan bagi jajaran.

Ringkasan kaedah dot matriks:

(i) Kaedah ini mudah difahami kerana memberi gambaran visual.

(ii) Mudah untuk mencari subrentetan, sebagai dot yang ditemukan atau baris secara pepenjuru iaitu a, b, c dan d. Sila rujuk Rajah 2.6.

(iii) Mudah untuk mencari subrentetan yang terbalik, sebagai dot yang ditemukan atau baris secara pepenjuru iaitu e, g dan t. Sila rujuk Rajah 2.6.

(iv) Mudah menjumpai perubahan dalaman bagi subjujukan seperti dalam Rajah 2.6 contohnya aksara a dan b bertukar.

(v) Mudah menjumpai penggantian aksara.

(vi) Walau bagaimanapun dot matriks boleh mengandungi *noise* iaitu kebanyakkan dot yang bertaburan tidak menunjukkan subjujukan.

(vii) Selain itu matriks boleh menjadi besar sekiranya jujukan yang lebih panjang yang ingin dijajarkan. Oleh itu ia tidak boleh dilihat secara visual.

2.4.2 Pengaturcaraan Dinamik

Pengaturcaraan dinamik merupakan satu lagi kaedah yang selalu digunakan untuk melaksanakan jajaran jujukan [9]. Terdapat dua kaedah pengaturcaraan dinamik iaitu Needleman-Wunsh [27] untuk jajaran global dan Smith-Waterman [38] untuk jajaran setempat. Dot matriks hanya menunjukkan bahagian kesamaan tetapi bukan laluan yang berhubung kepada bahagian yang tidak sama. Matlamat utama bagi penjajaran jujukan adalah untuk mencari jajaran yang optimal, oleh itu pengaturcaraan dinamik ini adalah suatu kaedah bagi memastikan hasil jajaran

adalah yang terbaik [15]. Pengaturcaraan dinamik merupakan kelas bagi penyelesaian optimum yang mana mencari penyelesaian terbaik dengan memecahkan masalah yang besar kepada bahagian yang kecil kemudian diselesaikan. Jawapan bagi masalah besar bergantung kepada pergantungan turutan *(sequential dependency)* yang mana jawapan submasalah i^{th} boleh didapati dari jawapan submasalah $i-1^{th}$. Setiap submasalah diselesaikan dan penyelesaiannya disimpan sebagai markah di dalam jadual. Jujukan atau laluan bagi markah submasalah yang paling tinggi dipilih sebagai penyelesaian optimal bagi keseluruhan masalah.

Seperti yang telah dijelaskan sebelum ini, matlamat utama jajaran jujukan adalah untuk mendapatkan markah yang maksimum (optima) bagi dua jujukan yang dijajarkan iaitu:

(i) Memaksimakan markah bagi pasangan aksara yang padan
(ii) Meminimakan markah bagi pasangan aksara yang tidak padan
(iii) Meminimakan jurang

Keseluruhan masalah dibahagikan kepada submasalah iaitu penjajaran aksara jujukan dengan setiap aksara jujukan yang lain. Penyelesaian terbaik dipilih menggunakan tiga pilihan iaitu:

(i) Menjajarkan aksara
(ii) Memperkenalkan jurang dalam jujukan 1 atau
(iii) Memperkenalkan jurang dalam jujukan 2

Secara ringkasnya algorithma pengaturcaraan dinamik ini menggunakan matriks seperti pada dot matriks dengan jujukan disusun pada baris pertama melintang dari kiri ke kanan dan lajur pertama menegak dari atas ke bawah. Pada setiap posisi dalam matriks, algoritma mengira markah terbaik dan menyimpan penuding dari posisi sebelumnya dari mana markah tertinggi itu dihasilkan atau diperolehi. Akhir sekali langkah penjejakan balik *(trace back)* dibuat untuk mencari markah tertinggi berdasarkan pemetaan penuding yang disimpan sebelum ini.

Pengaturcaraan dinamik sama ada Needleman-Wunch atau Smith-Waterman mempunyai tiga komponen utama iaitu:

(i) Penilaiawalan *(Initialization) / recurrent relation*

(ii) Pengisian Matrik *(Matrix fill)/ tabular computation*

(iii) Penjejakan balik *(Traceback*)

Selain itu, ada juga yang mengatakan komponen pertama penilaiawalan bagi pengaturcaraan dinamik ialah sebagai hubungan pengulangan *(recurrent relation)* dan pengisian matriks juga dikenali sebagai komputasi jadual *(tabular computation)* [9]. Skema permarkahan diperlukan dalam pengaturcaraan dinamik semasa langkah pengisian matriks. Berdasarkan penyelidikan terdahulu skema permarkahan yang terdapat dalam pengaturcaraan dinamik perlu diperbaiki dengan menggunanakan matriks penggantian dan memperkenalkan jurang dalam jajaran serta cadangan pengiraan jurang penalti [1,17,19,20]. Ianya bertujuan untuk mengoptimumkan hasil jajaran disamping mengekalkan konsep biologi. Huraian berkaitan formulasi teknik pengaturcaraan dinamik ini boleh diperolehi dalam bab 4.

2.5 Skema Permarkahan Bagi Membentuk Permarkahan Optima

Bagi menentukan jujukan mana yang optimal kita memerlukan skema permarkahan. Pertimbangkan contoh di bawah.

Berikut merupakan hasil yang mungkin bagi penjajaran jujukan J1 = AYKTVLKTPS dan jujukan J2 = ATFKVTLI. Jajaran mana yang optima?

(i)
```
AYKTVLKTPS
|    |  |
ATFKV - - TLI
```

(ii)
```
A-YKTVLKTPS
|  |  |
ATFK-V TLI - -
```

(iii)
```
A-YKTV- LKTPS
|  |  |  |
ATFK- VTLI- - -
```

(iv)
```
A-YK -TVLKTPS
|  |  |  |
ATFKVT-L I- - -
```

Berdasarkan contoh sebelum ini, cuma empat jajaran sahaja yang ditunjukkan tetapi sebenarnya terdapat ratusan jajaran yang boleh terhasil dari dua input jujukan tersebut. Ianya bergantung kepada panjang jujukan. Jajaran yang munasabah bagi panjang n adalah $\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \approx \frac{2^{2n}}{\sqrt{2\pi n}}$ [39]. Melalui pendekatan pengaturcaraan dinamik jajaran optimal dapat diperolehi dengan berkesan, tetapi ianya bergantung kepada skema permarkahan yang digunakan. Terdapat dua faktor yang perlu diambil kira semasa membentuk skema permarkahan iaitu :

(i) Matriks permarkahan bagi padan dan tidak padan serta mengambilkira perubahan yang mungkin berlaku akibat mutasi.

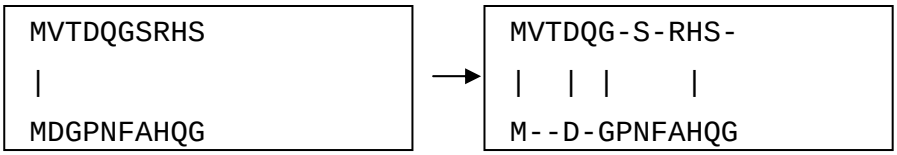
(ii) Kehadiran jurang dalam jajaran.

Jadi bagi menangani masalah mengukur perubahan evolusi yang telah dinyatakan sebelum ini, matriks permarkahan atau lebih dikenali sebagai matriks penggantian diperlukan [19]. Manakala penyelesaian bagi menangani masalah mengawal jurang memerlukan fungsi jurang penalti [17, 15].

2.5.1 Mengawal Jurang Menggunakan Jurang Penalti

Jurang merupakan maksima ruang kosong yang berturutan pada jujukan. Jurang diperkenalkan semasa penjajaran bagi mendapatkan hubungan kesamaan atau kemungkinan jajaran yang lebih baik di antara dua jujukan. Sebagai contoh 10 aksara pertama dari jujukan protein yeast digunakan iaitu put3_yeast dan yhx8_yeast. Jujukan ini diwakili oleh jujukan A = MVTDQGSRHS dan jujukan
B = MDGPNFAHQG.

Berikut adalah contoh jajaran jujukan yang diperbaiki dengan memperkenalkan jurang. Berdasarkan Contoh 2 pada rajah di bawah, didapati dengan kehadiran jurang jumlah padanan yang terhasil adalah lebih banyak iaitu 4 berbanding hanya 1 apabila tanpa jurang.



Contoh 1 Contoh 2

Rajah 2.7: Perbezaan penjajaran jujukan dengan kehadiran jurang

Jurang boleh terjadi:

(i) Sebelum aksara pertama bagi jujukan.

(ii) Dalam jujukan.

```
MVTDQG-S-RHS-
|   | |     |
M--D-GPNFAHQG
```

(iii) Selepas aksara terakhir bagi jujukan.

```
MVTDQG-S-RHS-
|   | |     |
M--D-GPNFAHQG
```

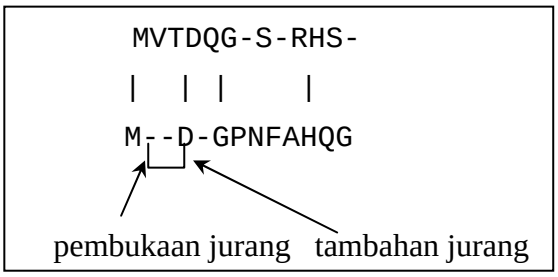
(iv) Ianya juga boleh terjadi dalam jujukan pertama, jujukan kedua atau kedua-duanya.

Sekirannya kewujudan jurang dibenarkan semasa penjajaran maka perlunya satu fungsi jurang penalti bagi mengawal dan mengambilkira kewujudan jurang tersebut. Terdapat dua fungsi jurang penalti yang biasa digunakan apabila jurang k dibenarkan dalam jajaran iaitu:

(i) *Linear gap cost*: $\beta(k) = -dk$, d = jurang

(ii) *Affine gap cost* : $\delta(k) = -d - e*(k-1)$, d = bukaan jurang, e= tambahan jurang.

Jurang penalti affine terdiri dari dua bahagian iaitu penalti pembukaan jurang *(gap opening penalty) –d* dan penalti tambahan jurang *(gap extension penalty) –e*. Rujuk Rajah 2.8. Kebiasaannya penalti tambahan jurang lebih kecil (e<d) untuk mengambarkan 10 penambahan bagi 1 neuleotides lebih susah dari 1 penambahan bagi 10 neucleotides [15].



Rajah 2.8: Pembukaan dan tambahan jurang

Jurang penalti affine merupakan model untuk menilai penambahan dan pembuangan menggunakan fungsi linear di mana satu syaratnya adalah panjang yang bebas dan satu lagi panjang yang saling bergantungan. Model ini menggalakkan penambahan jurang berbanding pengenalan jurang baru. Berbeza pula dengan jurang penalti linear, yang mana mengumpukkan nilai yang tetap per jurang. Berikut merupakan contoh perbezaan kiraan di antara dua fungsi jurang penalti ini. Diberi dua jujukan, berserta markah +2 bagi padan, -2 bagi jurang, –1 bagi tidak padan, -2 bagi jurang pembukaan dan -1 bagi jurang tambahan.

Jurang Penalti Linear	Jurang Penalti Affine
MVTDQG-S-RHS- | | | | M--D-GPNFAHQG + - -+-+--- - +-- 2 2 2 2 2 2 2 1 2 1 2 1 2 = -7 markah	MVTDQG-S-RHS- | | | | M--D-GPNFAHQG + -v+-+--- - +-- 2 2 **1** 2 2 2 2 1 2 1 2 1 2 = -6 markah
Bagi jajaran kedua, aksara S dianjakkan ke kanan satu posisi maka markahnya adalah:	
MVTDQG--SRHS- | | | | M--D-GPNFAHQG + - -+-+--- - +-- 2 2 2 2 2 2 2 2 1 1 2 1 2 = -7 markah	MVTDQG--SRHS- | | | | M-vD-GPvNFAHQG + - -+-+--- - +-- 2 2 **1** 2 2 2 2 **1** 1 1 2 1 2 = -5 markah

Rajah 2.9: Contoh perbezaan kiraan jurang linear dan affine

Berdasarkan Rajah 2.9, apabila dikira semula jajaran kedua menggunakan jurang penalti affine, markah jajaran meningkat berbanding jajaran pertama. Ini menunjukkan jurang penalti affine menyediakan insentif bagi algorithma jajaran untuk memastikan jujukan sentiasa bersama sebanyak mungkin berbanding memasukkan ratusan jurang kecil. Walaubagaimana pun parameter julat nilai jurang yang digunakan harus ditentukan dan perlu bersesuaian bagi memastikan jajaran yang optima diperolehi, tidak kira sama ada menggunakan fungsi linear atau affine. Satu panduan bagi penentuan parameter julat nilai bagi fungsi jurang penalti linear dan affine adalah perlu sebelum jajaran dilakukan. Maka, projek ini akan mengimplementasikan kedua fungsi jurang penalti ini bagi menghasilkan satu garis panduan pemilihan parameter julat nilai yang efektif secara empirikal.

2.5.2 Matriks Permarkahan

Matriks permarkahan yang digunakan bagi perbandingan jujukan protein adalah lebih komplek berbanding hanya menggunakan matrik unitari. Pelbagai alternatif kepada matriks unitari dicadangkan. Satu cadangan awalan adalah matriks permarkahan berdasarkan nombor minimum bagi *bases* yang perlu diubah untuk penukaran kodon bagi satu asid amino ke dalam kodon bagi asid amino kedua. Matriks ini dikenali sebagai matriks jarak mutasi minimum *(minimum mutation distance matrix)*. Ianya berjaya mengenalpasti lebih banyak jarak hubungan di antara jujukan protein berbanding kaedah matriks unitari. Matriks ini berkesan kerana ianya memasukkan maklumat tentang proses bagi pembentukkan mutasi dari satu asid amino kepada yang lain. Walau bagaimanapun ianya masih meminggirkan proses bagi pemilihan yang menentukan mutasi mana yang boleh hidup dalam sesuatu populasi. Oleh itu berberapa skema permarkahan yang dibangunkan berdasarkan kepada ciri fizikal, kimia atau struktur bahan seperti permarkahan kod genetik *(genetic code scoring*), permarkahan kesamaan kimia *(chemical similarity scoring)* dan matrik penggantian *(substitution matrices)* juga dikenali sebagai *log odds matrix* [18, 19].

Kepentingan matriks permarkahan adalah:

(i) Matriks permarkahan wujud dalam semua analisis termasuk perbandingan jujukan.

(ii) Pemilihan matriks boleh membawa kesan besar terhadap hasil analisis.

(iii) Matriks permarkahan yang tersirat mewakili teori evolusi yang khusus.

(iv) Memahami teori disebalik matriks permarkahan boleh membantu membuat pilihan yang tepat.

Oleh kerana matlamat utama penjajaran jujukan adalah untuk mendapatkan jajaran yang optima dan berasaskan kepada perubahan evolusi, maka matriks penggantian sesuai digunakan dalam skema permarkahan pengaturcaraan dinamik tradisional bagi mengukur perubahan evolusi supaya jajaran yang terhasil masih menerapkan unsur-unsur biologi [19]. Matriks penggantian yang digunakan secara meluas masa kini adalah PAM *(Point Accepted Mutation)* [10] dan BLOSUM *(Blocks Subsitution Matrix)* [19].

2.5.2.1 Matriks Penggantian PAM

Salah satu pembaikan yang penting kepada matriks unitari adalah berdasarkan perubahan evolusi *(evolutionary distances)*. Margeret Dayhoff merupakan pengasas bagi pendekatan ini. Pada tahun 1970, beliau melakukan kajian intensif terhadap frekuensi bagaimana asid amino saling bergantian sesama sendiri semasa evolusi. Kajiannya merangkumi penjajaran terhadap semua protein dalam beberapa keluarga protein dan kemudiannya membina pepohon phylogenetik bagi setiap keluarga tersebut. Ianya menjadi panduan bagi pembinaan jadual frekuensi

relatif terhadap kejadian asid amino dalam kajian protein yang dicantumkan semasa membuat pengiraan skema permarkahan bagi keluarga PAM [10].

Siri PAM adalah berdasarkan kadar peratusan anggaran mutasi dari protein yang mempunyai hubungan rapat dan kerana itu penguasaan mutasi asid amino disebabkan dari perubahan satu base. Matriks ini juga dikenali sebagai matriks log-odds dan formula bagi pengiraan kadar *log-odds* matrik adalah $S_{i,j} = \log \frac{q_{ij}}{p_i p_j}$. S adalah kadar *log odds* bagi dua kebarangkalian dua aksara, i dan j, dijajarkan oleh evolusi keturunan *(aligned by evolutionary descent)* dan kebarangkalian bahawa ia dijajarkan secara kebetulan *(aligned by chance).* Manakala q_{ij} adalah frekuensi bagi i dan j dikira untuk dijajar dalam jujukan yang diketahui wujud hubungan. Ianya diperolehi dari *(transition probability matrix).* p_i dan p_j merupakan frekuensi bagi kejadian aksara i dan j dalam set jujukan. Ianya dihasilkan dari jajaran global bagi jujukan yang berkait rapat *(closely related sequences)*. Nombor bagi matriks (PAM40, PAM100) merujuk kepada jarak evolusi, semakin besar nombornya semakin besar jaraknya. Contoh model matrik penggantian PAM20 ditunjukkan pada Rajah 2.10.

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	6	-3	-5	-4	-3	-5	-3	-3	-3	-6	-7	-3	-6	-9	-2	-1	-1	-16	-9	-3
R	-3	9	-7	-12	-9	-2	-11	-11	-3	-6	-10	-1	-5	-10	-5	-4	-3	-3	-11	-9
N	-5	-7	8	1	-13	-5	-3	-4	-1	-6	-8	-2	-11	-10	-7	-1	-3	-9	-5	-9
D	-4	-12	1	8	-16	-4	2	-4	-5	-9	-15	-6	-13	-17	-9	-5	-6	-17	-13	-9
C	-3	-9	-13	-16	10	-16	-16	-11	-3	-7	-17	-16	-16	-15	-9	-4	-9	-18	-5	-7
Q	-5	-2	-5	-4	-16	9	0	-3	0	-9	-6	-4	-5	-15	-4	-6	-7	-15	-14	-3
E	-3	-11	-3	2	-16	0	8	-5	-6	-6	-10	-5	-8	-16	-7	-5	-7	-19	-9	-3
G	-3	-11	-4	-4	-11	-3	-5	7	-10	-13	-12	-3	-10	-10	-7	-3	-7	-17	-16	-7
H	-3	-3	-1	-5	-3	0	-6	-10	9	-11	-7	-3	-13	-7	-5	-7	-3	-3	-4	-7
I	-6	-6	-6	-9	-7	-9	-6	-13	-11	9	-2	-7	-2	-3	-10	-3	-3	-16	-7	1
L	-7	-10	-8	-15	-17	-6	-10	-12	-7	-2	7	-9	0	-4	-3	-9	-3	-7	-3	-3
K	-3	-1	-2	-6	-16	-4	-5	-3	-3	-7	-9	7	-3	-16	-3	-5	-4	-14	-10	-10
M	-6	-5	-11	-13	-16	-5	-3	-10	-13	-2	0	-3	11	-5	-9	-6	-5	-15	-13	-2
F	-9	-10	-10	-17	-15	-15	-16	-10	-7	-3	-4	-16	-5	9	-11	-7	-10	-6	1	-9
P	-2	-5	-7	-9	-9	-4	-7	-7	-5	-10	-3	-3	-9	-11	8	-3	-5	-16	-16	-7
S	-1	-4	-1	-5	-4	-6	-5	-3	-7	-3	-9	-5	-6	-7	-3	7	0	-6	-3	-3
T	-1	-3	-3	-6	-9	-7	-7	-7	-3	-3	-3	-4	-5	-10	-5	0	7	-15	-7	-4
W	-16	-3	-9	-17	-18	-15	-19	-17	-3	-16	-7	-14	-15	-6	-16	-6	-15	13	-6	-18
Y	-9	-11	-5	-13	-5	-14	-9	-16	-4	-7	-3	-10	-13	1	-16	-3	-7	-6	10	-3
V	-3	-9	-9	-9	-7	-3	-3	-7	-7	1	-3	-10	-2	-9	-7	-3	-4	-18	-3	7

Rajah 2.10: PAM 20

2.5.2.2 Matriks Penggantian BLOSUM

Satu andaian dari model Dayhoff adalah kadar evolusi adalah seragam bagi keseluruhan jujukan protein [10]. Andaian ini tidak semestinya betul, kerana kadar evolusi adalah lebih rendah dalam *conserved region* dan tinggi dalam *non-conserved region* [18]. Oleh itu siri matriks BLOSUM yang dibangunkan oleh Steve Henikoff adalah satu cara menjelaskan kepentingan jarak hubungan *(distant relationship)* [19]. Ianya dibina dengan menggunakan blok bagi serpihan *(fragment)* jujukan dari keluarga protein yang berbeza yang boleh dijajarkan tanpa kehadiran jurang. Oleh kerana matriks PAM adalah berdasarkan jujukan protein sekurang-kurangnya 85% kesamaan, penulis bagi matriks BLOSUM ingin membina matriks yang boleh memodelkan jujukan protein yang mempunyai hanya sedikit darjah pencapahan *(divergence)* [20]. Selain itu, data jujukan protein pada masa matriks PAM dibina adalah terhad.

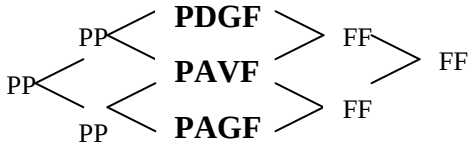
Berikut merupakan ringkasan langkah pembinaan matriks BLOSUMx dari BLOK:
Sebagai contoh, andaikan berikut merupakan BLOK dari X% pengkelasan:

PDGF
PAVF
PAGF

Langkah 1 :

Membina jadual yang mengandungi nombor pasangan asid amino bagi setiap ruang dalam BLOK. Jadual ini akan digunakan untuk membina matriks yang mengandungi kadar kemunculan *(ratio of appearance)* bagi pasangan dalam BLOK *(pairs in BLOCK)* berbanding kemunculan pasangan secara kebetulan *(pairs by chance)*.

Contoh :



3 pasangan PP	2 pasangan GV
2 pasangan DA	1 pasangan GG
1 pasangan AA	3 pasangan FF

Langkah 2 :

Kira frekuensi kebarangkalian bagi kejadian *(probability of occurrence)* setiap pasangan. Caranya, membahagikan pasangan kejadian dengan jumlah bagi pasangan

Dengan formula. $q_{ij} = \frac{f_{ij}}{\sum_{i=1}^{20}\sum_{j=1}^{i} f_{ij}}$

Contoh:

Terdapat 12 pasangan dalam BLOK. Kebarangkalian bagi kejadian bagi pasangan PP iaitu (qPP) adalah 3/12 =0.25. Bagi qDA = 2/12 = 0.166, qAA = 1/12 = 0.0833, qGV = 2/12 = 0.166, qGG = 1/12 = 0.0833 dan qFF = 3/12 = 0.25.

Langkah 3:

Kira kebarangkalian bagi kejadian asid amino i dalam pasangan i, j .

Formulanya, $p_i = q_{ii} + \sum_{j \neq i} \frac{q_{ij}}{2}$

Contoh:

Berdasarkan formula di atas, kebarangkalian bagi kejadian asid amino P iaitu (pP) dalam pasangan adalah [3+0]/12 =0.25. Bagi pA = [1 + (2/2)]/12 = 2/12 = 0.166,
pD = [0 + (2/2)]/12 = 1/12 = 0.0833, pG = [1 + (2/2)]/12 = 2/12 = 0.166,
pV = [0 + (2/2)]/12 = 1/12 = 0.0833 dan pF = [3 + 0]/12 = 0.25.

Langkah 4 :

Kira jangkaan kebarangkalian kejadian untuk semua pasangan i, j .

Apabila $i = j$, maka $e_{ij} = p_i p_j$, sekiranya $i \neq j$ maka $e_{ij} = 2 p_i p_j$

Contoh :

Jangkaan kebarangkalian kejadian untuk pasangan PP iaitu (ePP) adalah
0.25 x 0.25 = 0.0625. Bagi eDA = 2 x 0.0833 x 0.166 = 0.0276,
eAA = 0.166 x 0.166 = 0.0276, eGV = 2 x 0.166 x 0.0833 = 0.0276.
eGG = 0.166 x 0.166 = 0.0276 dan eFF = 0.25 x 0.25 = 0.0625.

Langkah 5:

Kira *odds-matrix* menggunakan formula di bawah, kemudian ditukar ke *log-odds matrix* dengan mengumpukkan logarithma asas 2 pada setiap masukan. Nilai bagi matriks BLOSUM adalah hasil *log-odds matrix* di mana setiap nilai didarab dengan nilai 2 dan dibundarkan kepada integer terhampir.
Formulanya:

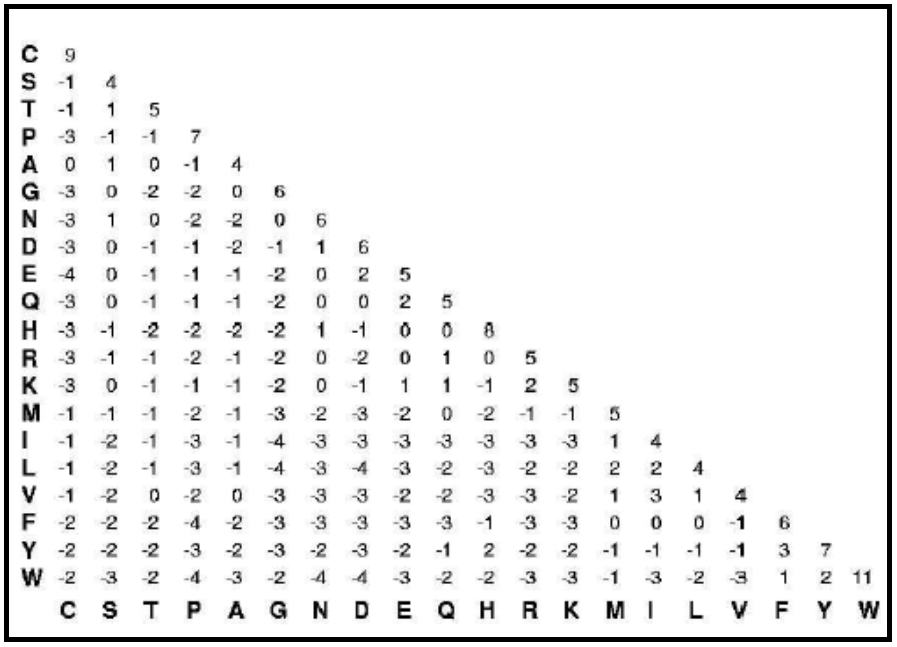
Odds matrix, Om= $\frac{q_{ij}}{e_{ij}}$

Log-odds matrix, s_{ij} = log2 (Om)

Matriks BLOSUM dihasilkan dari penjajaran setempat bagi jujukan berjarak rapat *(distantly related sequences*) dengan tujuan memperbaiki matriks PAM. Terdapat banyak siri BLOSUM contohnya : BLOSUM 90, BLOSUM 80, BLOSUM 62, BLOSUM 50, BLOSUM 45.dan BLOSOM 30. Nombor bagi matriks BLOSUM62 merujuk kepada peratusan identiti minimum bagi blok yang digunakan untuk membina matriks. Semakin besar nombor semakin kurang jaraknya *(lesser distances).* Ianya dihasilkan berdasarkan nilai *threshold* sebagai contoh *threshold* 80% identiti menghasilkan BLOSUM 80, *threshold* 45% identiti menghasilkan BLOSUM 45 dan seterusnya[20]. Contoh model matrik penggantian BLOSUM62 yang dibina Henikoff pada tahun 1993, ditunjukkan pada Rajah 2.11.

Oleh kerana pembinaan matriks penggantian BLOSUM mengekalkan konsep biologi (berlakunya peristiwa mutasi dalam struktur protein), iaitu menerapkan hubungan evolusi serta mengambil kira jujukan berjarak rapat. Maka matriks penggantian BLOSUM sesuai digunakan untuk projek ini yang mengkaji jajaran setempat [19]. Kaedah pengaturcaraan dinamik Smith-Waterman diimplement bagi mendapatkan jajaran yang optimal dengan mengubah skema permarkahan asal dalam pengaturcaraan dinamik dengan menggunakan matriks penggantian BLOSUM berserta fungsi jurang penalti linear dan affine. Perbandingan keberkesanan akan dibuat terhadap tiga jenis matriks penggantian iaitu BLOSUM45, BLOSUM62 dan BLOSUM80 dan parameter fungsi jurang penalti linear dan affine yang berbeza.

Rujuk Lampiran E, F dan G iaitu matriks penggantian BLOSUM 45, BLOSUM62, dan BLOSUM80 yang akan digunakan dalam projek ini.



	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	9																			
S	-1	4																		
T	-1	1	5																	
P	-3	-1	-1	7																
A	0	1	0	-1	4															
G	-3	0	-2	-2	0	6														
N	-3	1	0	-2	-2	0	6													
D	-3	0	-1	-1	-2	-1	1	6												
E	-4	0	-1	-1	-1	-2	0	2	5											
Q	-3	0	-1	-1	-1	-2	0	0	2	5										
H	-3	-1	-2	-2	-2	-2	1	-1	0	0	8									
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5								
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5							
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5						
I	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4					
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4				
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4			
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6		
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7	
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11

Rajah 2.11: BLOSUM62

2.6 Ringkasan

Kajian literatur dilakukan bagi menganalisa kajian-kajian terdahulu berkaitan projek yang bakal dilaksanakan, fokusnya kepada permasalahan penjajaran jujukan. Ianya bertujuan bagi mengenalpasti permasalahan masa kini dan cadangan penyelesaiannya. Hasil kajian didapati kaedah pengaturcaraan dinamik dapat mempermudahkan tugasan penjajaran. Skema permarkahan dalam pengaturcaran dinamik perlu diperbaiki bagi menghasilkan jajaran yang optima iaitu mempunyai bilangan padanan maksima. Berdasarkan hasil penyelidikan dari penyelidik terdahulu matriks penggantian dan fungsi jurang penalti dapat mengoptimumkan hasil jajaran disamping mengekalkan konsep biologi iaitu wujudnya perubahan evolusi disebabkan mutasi [13, 15]. Oleh itu kedua fungsi ini akan diimplementasikan dalam skema permarkahan dalam pengaturcaraan dinamik dan menganalisa keberkesanan skema permarkahan tersebut bagi menentukan parameter yang efektif.

BAB 3

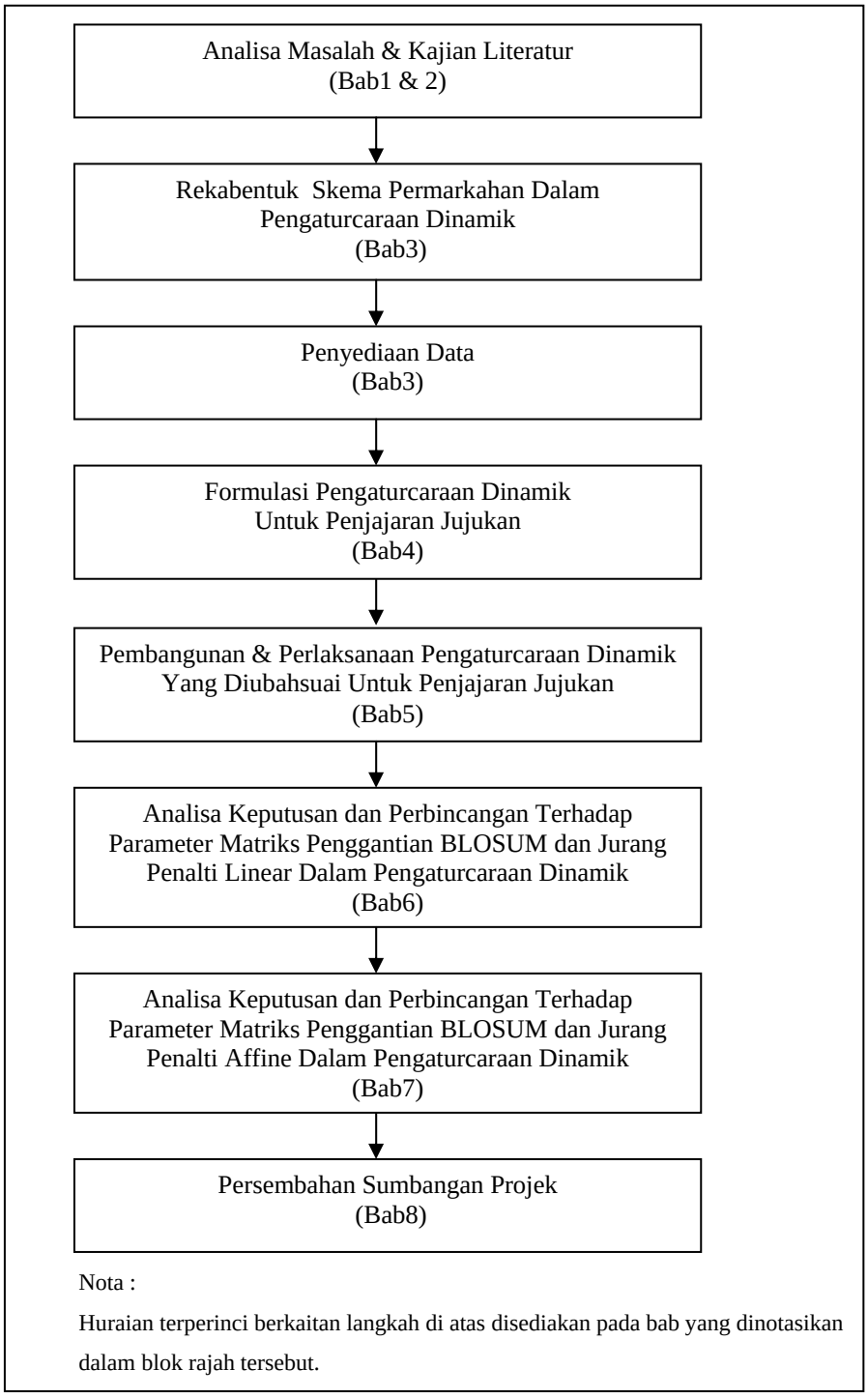
METODOLOGI

3.1 Pendahuluan

Bagi membangunkan suatu projek penyelidikan, terlebih dahulu pembangun perlu merancang fasa-fasa yang harus dilalui dalam kitar hayat pembangunan projek tersebut. Keperluan kepada penentuan metodologi dan kaedah dalam pembangunan projek merupakan perkara yang penting untuk menghasilkannya secara efektif.

3.2 Organisasi Bagi Projek

Metodologi merupakan suatu garis panduan untuk diikuti dalam membangunkan suatu projek. Ia juga merujuk kepada keseluruhan proses pembangunan. Setiap metodologi mengandungi beberapa fasa tertakrif dengan matlamatnya tersendiri. Terdapat lapan langkah utama bagi menghasilkan projek ini, seperti Rajah 3.1. Setiap langkah tersebut akan diterangkan dalam bahagian seterusnya.



Rajah 3.1 : Metodologi projek

3.3 Analisa Masalah dan Kajian Literatur

Analisa dan kenalpasti masalah merupakan langkah permulaan bagi membangunkan sesuatu projek. Ianya bagi memastikan penyelidikan yang dilakukan benar-benar diperlukan. Setelah masalah dikenalpasti kajian literatur dilakukan bagi mengkaji penyelesaian yang dihasilkan dari penyelidikan terdahulu, seterusnya mengenalpasti penyelesaian yang berpotensi untuk diformulakan sebagai cabaran projek. Objektif dan skop projek dibangunkan berdasarkan analisa masalah dan kajian yang dilakukan. Selain itu, projek yang dibangunkan harus mempunyai matlamat yang jelas dengan jadual perancangan kerja disusun mengikut jangka masa yang ditetapkan.. Secara keseluruhannya fasa ini merupakan rangka projek yang bakal dilaksanakan.

3.4 Rekabentuk Skema Permarkahan dalam Pengaturcaraan Dinamik

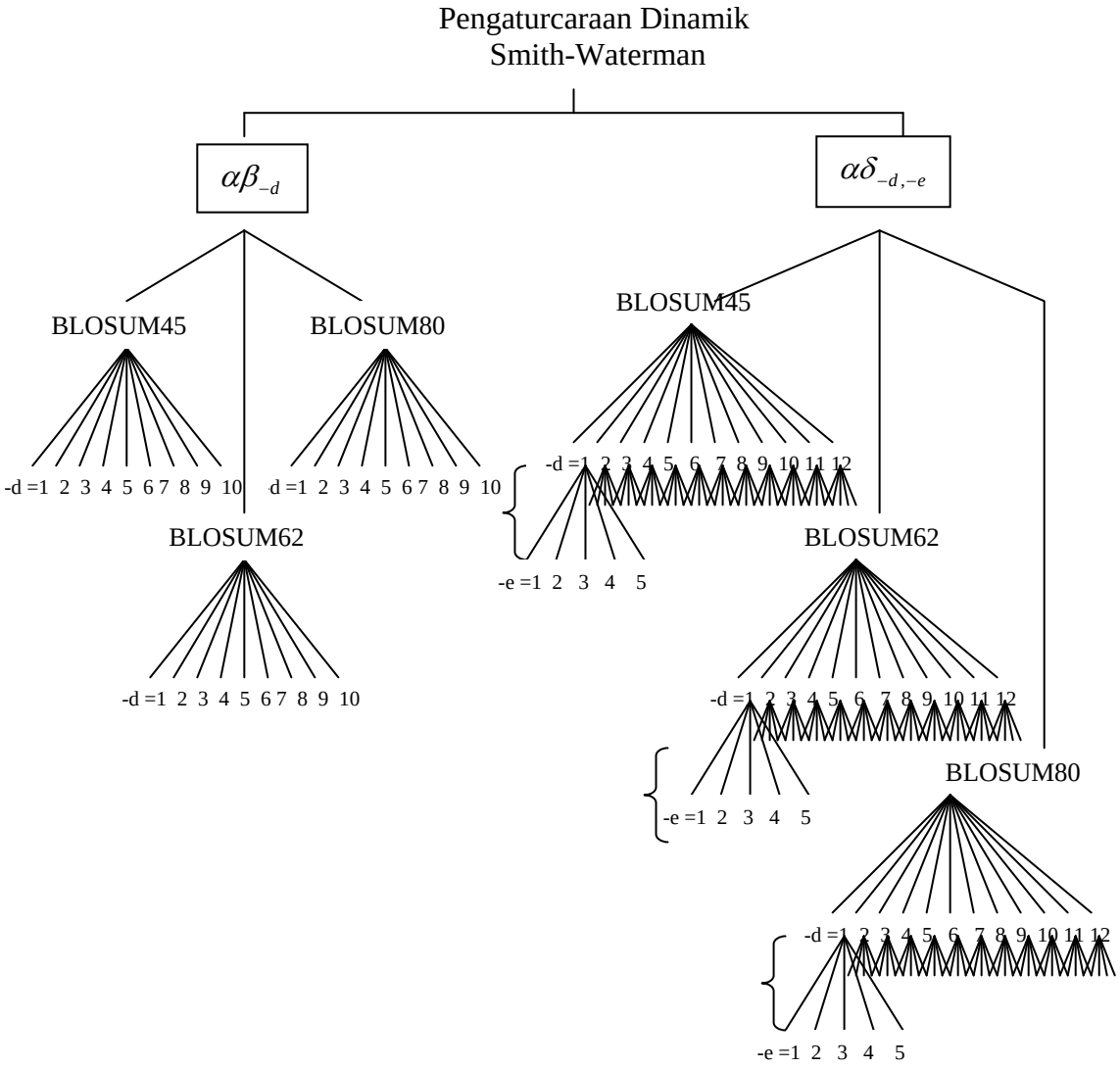
Seterusnya merupakan rekabentuk bagi kajian yang akan dilakukan bagi projek ini iaitu menentukan parameter skema permarkahan yang efektif dalam pengaturcaraan dinamik Smith-Waterman. Skema permarkahan yang akan digunakan terdiri dari kombinasi matriks penggantian BLOSUM dan fungsi jurang penalti (linear dan affine). Permilihan skema permarkahan ini adalah berdasarkan kepada penyelidikan terdahulu [16, 17, 24] dan masalahnya adalah tiada sebarang panduan bagi pemilihan parameter matriks penggantian dan fungsi jurang penalti bagi penjajaran jujukan [31].

Berikut merupakan skema permarkahan yang akan digunakan dalam projek ini ialah :

(i) Matriks penggantian BLOSUM (α) dengan jurang penalti linear (β_{-d})

(ii) Matriks penggantian BLOSUM (α) dengan jurang penalti affine ($\delta_{-d,-e}$)

Parameter siri matriks penggantian BLOSUM yang akan digunakan adalah BLOSUM45, BLOSUM62 dan BLOSUM80. Parameter julat nilai –d bagi jurang penalti linear (β_{-d})yang akan digunakan adalah 1 hingga 10. Manakala parameter jurang penalti affine ($\delta_{-d,-e}$), julat nilai –d adalah 1 hingga 12 dan julat nilai –e adalah 1 hingga 5. Rujuk Rajah 3.2 bagi rekabentuk skema permarkahan yang akan digunakan.



Rajah 3.2 : Rekabentuk skema pemarkahan dalam pengaturcaraan dinamik

3.5 Penyediaan Data

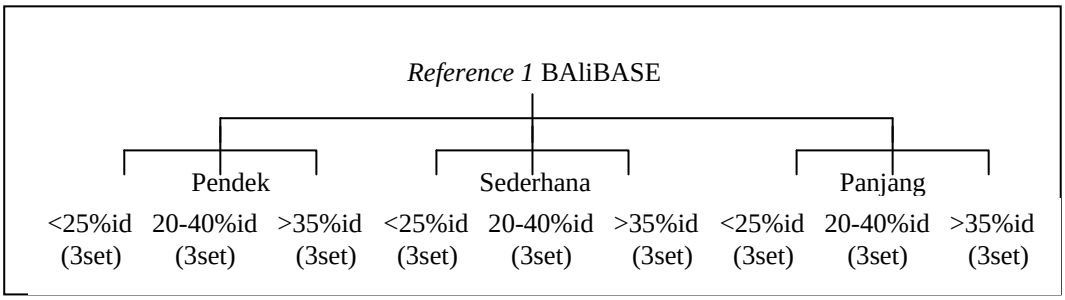
3.5.1 Perolehan dan Pra-pemprosesan Data Protein

Proses seterusnya adalah penyediaan data jujukan protein yang bakal digunakan bagi tujuan penjajaran. Hanya saintis biologi sahaja yang mengetahui data jujukan protein yang digunakan itu merupakan data setempat dan sesuai untuk dijajarkan. Oleh itu bagi memastikan kesahihan data, rujukan bagi nombor id jujukan protein dari pangkalan data BAliBASE *(Benchmark Alignment Database)* digunakan [41, 42]. Rujukan dari pangkalan data BAliBASE dipilih kerana penyelidikan terdahulu yang mengkaji penjajaran jujukan setempat turut menggunakan pangkalan data ini [8, 42].

BAliBASE telah dibangunkan pada tahun 1999 oleh Julie D. Thompson, Frederic Plewniek dan Oliver Poch, dengan tujuan sebagai set pengujian atau rujukan bagi menilai program penjajaran banyak pasangan [41]. BAliBASE merupakan pangkalan data yang menapis secara manual sekumpulan jajaran jujukan. Secara spesifiknya ianya direka bagi penilaian dan perbandingan bagi sekumpulan jajaran jujukan. Jajaran dikategorikan dari segi panjang jujukan, kesamaan dan kehadiran pernambahan dan N/C-tambahan terminal. BAliBASE (Versi 1.0) mengandungi 142 set rujukan jajaran yang dibahagikan kepada 4 hieraki set rujukan, setiap satu mengandungi sekurang-kurangnya 12 perwakilan jajaran seperti Lampiran H. Hanya *Reference1* dalam BAliBASE sahaja yang akan digunakan dalam projek ini [51].

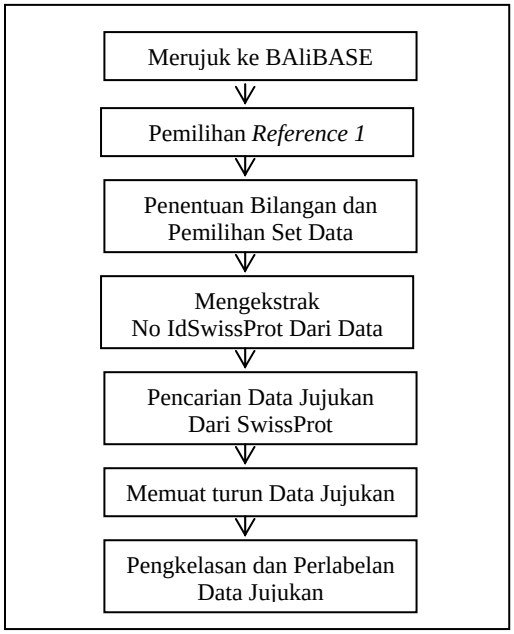
Reference1 mengandungi 3 kategori data yang dikelaskan mengikut panjang jujukan iaitu pendek (60-130 bp), sederhana (200-480 bp) dan panjang (480-1000 lebih bp). Bagi setiap kelas ini dikumpulankan mengikut peratusan kesamaan hubungan yang wujud *(similarity identity)* iaitu <25% id, 20-40% id dan >35% id. Oleh kerana kekangan masa, hanya 3 pasangan data akan diambil dari setiap kumpulan yang mana mewakili 9 pasangan data mengikut pengkelasan panjang

jujukan seperti Rajah 3.3. Secara keseluruhannya 27 set pasangan data yang dijadikan sebagai data seperti Rajah 3.5.



Rajah 3.3 : Bilangan set yang diambil dari *Reference1* BAliBASE

Setiap set pasangan yang dipilih dari *Reference 1* mempunyai nombor id SwissProt iaitu nombor id jujukan dalam pangkalan data protein universal SwissProt. Berdasarkan nombor id tersebut, data jujukan protein akan dicari dan dimuat turun dari laman web pangkalan data SwissProt. Rujuk [50, 52] bagi alamat laman tersebut. Selepas data dimuat turun, proses seterusnya adalah mengkelaskan dan melabelkan data tersebut mengikut kumpulan dari *Reference 1* dan disimpan dalam pangkalan data bagi memudahkan proses perlaksanaan kelak. Berikut merupakan rajah bagi proses perolehan dan prapemprosesan data.



Rajah 3.4 : Proses perolehan dan pra pemprosesan data kajian

Kategori Jujukan		Nama Kumpulan BAliBASE	Nama Jujukan	Id Swiss-Prot
Ukuran	Identiti			
Pendek (60-130 *bp*)	<25%	Repressor	1r69	P16117
			1neq	P06020
		Pertussis Toxin	1tvxA	P02775
			1prtF	P04981
		Ubiquitin	1ubi	P02248
			1awd	P56408
	20%-40%	High-Potential Iron-Sulfur Protein	1hpi	P38524
			1hip	P00260
		Cytochrome C	cy2_rhoge	P00097
			c550_bacsu	P24469
		Cytochrome E	1ycc	P00044
			1c2r	P00094
	>35%	Toxin	scxa_buteu	P01490
			scx1_titse	P01496
		Ferredoxin [2fe-2s]	1frr	P00235
			1fxa	P06543
			1fxa	P06543
		Hemerythrin	1hrb	P02244
			hem1_phago	P27686
Sederhana (200-320 *bp*)	<25%	Uridylate Kinase	1uky	P15700
			1dvrA	P07170
		3 Alpha, 20 Beta Hydroxysteroid Dehydrogenase	2hsdA	P19992
			1fds	P14061
		Phtalate Reductase	1fdr	P28861
			1ndh	P07514
	20%-40%	Ribosomal Protein L1	rk1_cyapa	P48125
			rl1_metja	P54050
		Tonin	1ton	P00759
			1try	P35049
		Anhydrase	cah1_human	P00915
			cah4_rat	P48284
	>35%	Triose Phosphate Isomerase	1amk	P48499
			1timA	P00940
		Lectin	1led	P24146
			1lte	P16404
		Thymidylate Sythase	tysy_bpt4	P00471
			tysy_ecoli	P00470
Panjang (415-1000 *bp*).	<25%	Cyclodextrin	1pamA	P05618
			1smd	P04745
		Myrosinase	2myr	P29092
			1gowA	P22498
		Gal4	put3_yeast	P25502
			yhx8_yeast	P38699
	20%-40%	Carboxypeptidase	1ac5	P09620
			1ivy	P10619
		Eftu	erf2_picpi	P23637
			selb_ecoli	P14081
		Acetylcholinesterase	bal_human	P19835
			est1_culpi	P16854
	>35%	Glycogen Phosphorylase B	gpb	P00489
			ahpA	P00490
		Lactoferrin	1lcf	P02788
			trfe_rabit	P19134
		Phosphoglucomutase	pgmu_rabit	P00949
			pgmu_agrtu	P39671

Rajah 3.5: Set data jujukan protein mengikut kategori

3.5.2 Perolehan Matriks Penggantian BLOSUM

Elemen terpenting untuk menilai kualiti bagi sesuatu jajaran jujukan berpasangan adalah matriks penggantian, yang mana ianya mengumpukkan markah untuk menjajarkan pasangan yang munasabah bagi setiap aksara. Teori bagi matriks penggantian asid amino telah dinyatakan dalam [1], dan diaplikasikan kepada perbandingan jujukan DNA dalam [45]. Secara umumnya, matriks penggantian yang berbeza adalah khusus untuk mengesan kesamaan di antara jujukan yang mempunyai darjah pencapahan yang pelbagai. Satu matriks tidak mungkin dapat merangkumi kesemua nilai bagi perubahan evolusi. Bagi kajian ini matriks penggantian BLOSUM akan digunakan. Matriks BLOSUM diperolehi dari laman web dijana oleh matblas. Rujuk [49] bagi alamat laman tersebut. Siri matriks BLOSUM ini akan dimuat turut dan dilabelkan sebelum disimpan.

3.6 Formulasi Pengaturcaraan Dinamik Untuk Penjajaran Jujukan

Kajian ini akan mengimplementasikan skema permarkahan yang terdiri dari matriks penggantian dan jurang penalti dalam model pengaturcaraan dinamik Smith-Waterman yang asal. Oleh itu, formulasi bagi pengubahsuaian itu harus dirangka terlebih dahulu bagi mempermudahkan proses pembangunan aturcara kelak. Formulasi ini merangkumi formulasi pengaturcaraan dinamik secara umum, formulasi pengaturcaraan dinamik bagi penjajaran jujukan iaitu Smith-Waterman yang asal dan formulasi pengaturcaraan dinamik Smith-Waterman yang diubahsuai untuk digunakan dalam kajian. Hasilnya merupakan model pengaturcaraan dinamik Smith-Waterman yang diubahsuai. Perincian formulasi bagi model ini boleh diperolehi dalam Bab 4.

3.7 Pembangunan dan Perlaksanaan Pengaturcaraan Dinamik Yang Diubahsuai Untuk Penjajaran Jujukan

Fasa ini akan menjelaskan proses-proses yang perlu dilakukan bagi membangunkan model pengaturcaraan dinamik Smith-Waterman yang diubahsuai pada fasa sebelumya. Pengubahsuaian yang dilakukan adalah terhadap skema permarkahan yang terdiri dari kombinasi jurang penalti dan matriks penggantian yang berbeza. Rekabentuk dan formulasi bagi model yang dibina akan diterjemahkan kepada kod aturcara. Setelah algorithma siap dikod, proses seterusnya adalah perlaksanaan penjajaran jujukan bagi set data jujukan protein yang telah ditentukan sebelum ini.

3.8 Analisa Keputusan dan Perbincangan Terhadap Parameter Matriks Penggantian BLOSUM dan Jurang Penalti Linear Dalam Pengaturcaraan Dinamik

Proses seterusnya adalah analisa hasil keputusan dan perbincangan terhadap keberkesanan parameter skema permarkahan yang digunakan dalam pengaturcaraan dinamik Smith-Waterman iaitu kombinasi siri matriks penggantian BLOSUM ($\alpha 45, a62$ atau $\alpha 80$) dengan julat nilai fungsi jurang penalti linear ($\beta_{1...10}$). Hasil keputusan bagi setiap larian dengan kombinasi parameter skema permarkahan yang berbeza ($\alpha\beta_{-d}$) direkodkan dalam jadual. Selain itu, langkah perolehan hasil dan bagaimana proses olahan dilakukan bagi memperolehi hasil akhiran turut diperincikan.

Seterusnya kajian perbandingan dan penilaian terhadap jadual dan graf yang dijana dari hasilan akhir, berdasarkan kategori set data yang digunakan. Ianya

bertujuan untuk menentukan parameter yang efektif bagi matriks penggantian BLOSUM dan nilai –d bagi β_{-d} . Hasilnya adalah satu panduan pemilihan parameter skema permarkahan $\alpha\beta_{-d}$ yang efektif.

3.9 Analisa Keputusan dan Perbincangan Terhadap Parameter Matriks Penggantian BLOSUM dan Jurang Penalti Affine Dalam Pengaturcaraan Dinamik

Proses seterusnya adalah analisa hasil keputusan dan perbincangan terhadap keberkesanan parameter skema permarkahan yang digunakan dalam pengaturcaraan dinamik Smith-Waterman iaitu kombinasi siri matriks penggantian BLOSUM ($\alpha 45, a62$ atau $\alpha 80$) dengan julat nilai fungsi jurang penalti affine $\delta_{1..12,1..5}$. Hasil keputusan bagi setiap larian dengan kombinasi skema permarkahan yang berbeza ($\alpha\delta_{-d,-e}$) direkodkan dalam jadual. Selain itu, langkah perolehan hasil dan bagaimana proses olahan dilakukan bagi memperolehi hasil akhiran turut diperincikan.

Seterusnya kajian perbandingan dan penilaian terhadap jadual dan graf yang dijana dari hasilan akhir, berdasarkan kategori set data yang digunakan. Ianya bertujuan untuk menentukan parameter yang efektif bagi matriks penggantian BLOSUM dan nilai –d serta –e bagi $\delta_{-d,-e}$. Hasilnya adalah satu panduan pemilihan parameter skema permarkahan $\alpha\delta_{-d,-e}$ yang efektif.

3.10 Persembahan Sumbangan Projek

Penyediaan persembahan sumbangan projek merupakan langkah terakhir bagi metodologi projek di mana segala hasil bagi projek ini akan didokumentasikan dan format penulisannya mengikut spesifikasi format penulisan tesis UTM. Ini termasuklah analisa masalah, kajian literatur, metodologi projek, model pengaturcaraan dinamik, pembangunan dan perlaksanaan penjajaran jujukan, analisa keputusan dan perbincangan, diakhiri dengan kesimpulan. Melalui hasil projek ini dapat memudahkan rujukan bagi penyelidikan selanjutnya khususnya terhadap penjajaran jujukan. Selain itu ianya dapat memberikan panduan terhadap pemilihan parameter skema permarkahan yang efektif bagi penjajaran jujukan menggunakan pengaturcaraan dinamik. Hasil projek ini boleh digunakan bagi kesimbungan penyelidikan akan datang.

3.11 Ringkasan

Perancangan metodologi merupakan kriteria penting sebelum pembinaan sesuatu projek bagi memastikan projek berjalan lancar dan sistematik serta dapat disiapkan mengikut masa yang ditetapkan. Secara keseluruhannya bab ini menerangkan metodologi bagi projek yang terdiri dari lapan langkah utama iaitu analisa masalah dan kajian literatur, rekabentuk, penyediaan data, formulasi pengaturcaraan dinamik, analisa keputusan dan perbincangan, serta diakhiri dengan sumbangan projek. Sebahagian langkah tersebut diterangkan secara ringkas sahaja dalam bab ini kerana ianya akan diperincikan dalam bab selanjutnya.

BAB 4

MODEL PENGATURCARAAN DINAMIK UNTUK PENJAJARAN JUJUKAN

4.1 Pendahuluan

Permulaan bab ini akan membincangkan secara ringkas teori pengaturcaraan dinamik secara umum merangkumi elemen yang perlu ada bagi membolehkan pengaturcaraan dinamik diaplikasikan. Manakala perbincangan seterusnya adalah terhadap formulasi bagi model pengaturcaraan dinamik untuk penjajaran jujukan iaitu Smith-Waterman yang bakal digunakan bagi kajian ini. Penerangan dimulai dengan model pengaturcaraan dinamik Smith-Waterman yang asal dan diikuti dengan model pengaturcaraan dinamik Smith-Waterman yang diubahsuai bagi tujuan kajian. Pengubahsuaian yang dilakukan adalah terhadap skema permarkahan dalam pengaturcaraan dinamik Smith-Waterman yang asal dengan menggunakan matriks penggantian BLOSUM dan fungsi jurang penalti iaitu linear dan affine.

4.2 Pengaturcaraan Dinamik Secara Umum

Model pengaturcaraan dinamik diwakili dengan cara yang berbeza berbanding model matematik yang lain. Berbeza dengan kefungsian objektif dan kekangan, model pengaturcaraan dinamik mengambarkan tempoh proses bagi keadaan (states), penentuan *(decision)*, peralihan *(transition)* dan pemulangan *(return)*. Ianya mengandungi koleksi bagi kenyataan kesamaan *(equation)* yang mengambarkan proses penentuan berjujukan *(sequential decision)*. Pengaturcaraan dinamik kebiasaannya diaplikasikan untuk menangani masalah optima, di mana permasalahan tersebut mempunyai banyak kemungkinan penyelesaiannya [48]. Setiap penyelesaian mempunyai nilai dan pengaturcaraan dinamik akan mencari penyelesaian optima iaitu yang mempunyai nilai maksima dari kemungkinan penyelesaian tersebut.

Pembangunan algorithma pengaturcaraan dinamik boleh dipecahkan kepada empat langkah iaitu seperti di bawah [48].

Langkah 1 : Kenalpasti Struktur
Menggambarkan sifat struktur bagi penyelesaian optima dengan menunjukkan ia boleh diuraikan kepada sub masalah optima.

Langkah 2 : Penyelesaian Rekursif
Mentakrifkan nilai bagi penyelesaian optima.

Langkah 3 : Pengiraan bawah-atas
Mengira nilai bagi penyelesaian optima dalam corak bawah-atas dengan menggunakan struktur jadual.

Langkah 4 : Membina penyelesaian optima
Bina penyelesaian optima dari informasi yang dikira.

Langkah 1 hingga 3 membentuk asas penyelesaian pengaturcaraan dinamik kepada masalah. Langkah 4 boleh diabaikan sekiranya hanya nilai bagi penyelesaian optimal sahaja yang diperlukan.

Terdapat banyak permasalahan yang menggunakan pengaturcaraan dinamik, di antara contoh masalah adalah pengiraan nombor fibonanci, penjadualan himpunan laluan *(assembly-line scheduling),* pendaraban tukaran matrik *(matrix-chain multiplication),* carian semua pasangan laluan pendek *(all-pairs shortest paths),* subjujukan terpanjang *(longest common subsequence),* polygon bersegi optimal *(optimal polygon triangulation)* dan banyak lagi. Algorithma pengaturcaraan dinamik dibangunkan menggunakan empat langkah di atas bagi menyelesaikan setiap masalah tersebut, setiap masalah berbeza formulasi pengaturcaraan dinamiknya.

Seterusnya adalah bagaimana untuk mengenalpasti bahawa permasalahan yang sesuai diselesaikan oleh pengaturcaraan dinamik. Oleh itu terdapat dua elemen utama yang diperlukan bagi pengaturcaraan dinamik bagi membolehkan teknik ini digunakan. Berikut adalah perincian ringkas element tersebut [48].

Elemen 1 : Substruktur optimal *(optimal substructure)*

Langkah pertama menyelesaikan masalah optima menggunakan pengaturcaraan dinamik adalah mengambarkan struktur bagi penyelesaian optimal. Oleh itu perlu dipastikan bahawa masalah mempunyai substruktur optimal iaitu penyelesaian optimal bagi masalah mengandungi penyelesaian optimal bagi submasalah.

Elemen 2 : Tindanan submasalah *(overlapping subproblem)*

Merupakan ramuan kedua yang mesti wujud dalam masalah optima yang ingin diaplikasi menggunakan pengaturcaraan dinamik iaitu tindanan submasalah. Cirinya adalah nombor atau ruang bagi sub masalah adalah kecil dan algorithma rekursif menyelesaikan submasalah yang sama berulang-ulang kali.

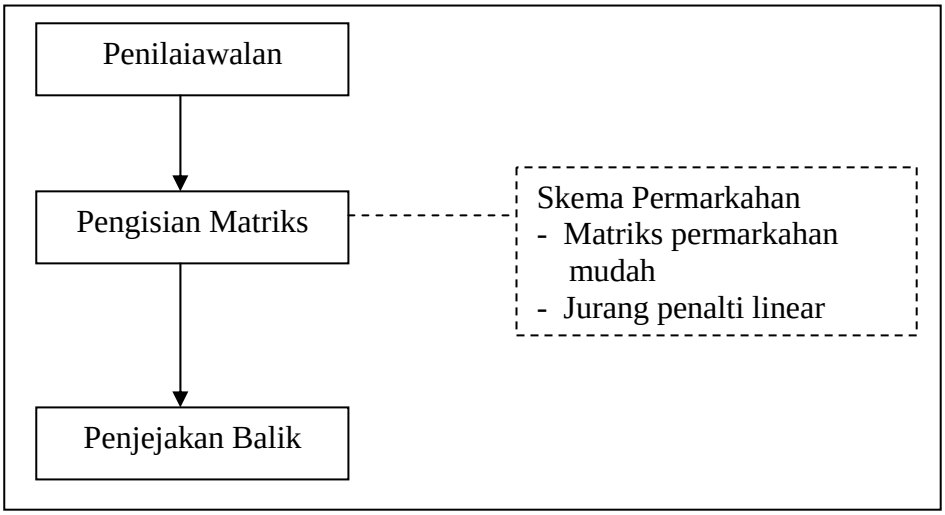
Elemen 3 : *Memoization*

Merupakan pembolehubah mudah bagi pengaturcaraan dinamik yang mengambil manfaat dari elemen tindanan submasalah. Ideanya diguna pada algorithma rekursif dengan menyimpan keputusan terdahulu kepada fungsi rekursif dalam jadual.

Bagi permasalahan penjajaran jujukan dua kaedah pengaturcaraan dinamik yang dinamakan NeedleMan-Wunsh [27] bagi jajaran global dan Smith-Waterman [38] bagi jajaran setempat telah dibangunkan. Seperti yang telah dinyatakan dalam bab 2. Perincian seterusnya adalah terhadap formulasi pengaturcaraan dinamik Smith-Waterman akan digunakan bagi kajian ini.

4.3 Model Pengaturcaraan Dinamik Smith-Waterman Asal

Projek ini hanya mengkaji berkaitan penjajaran setempat dengan menggunakan kaedah Smith-Waterman. Rajah 4.1 di bawah menunjukkan langkah-langkah dalam model pengaturcaraan dinamik Smith-Waterman.



Rajah 4.1 : Turutan proses pengaturcaraan dinamik Smith-Waterman

Berikut merupakan huraian bagi setiap langkah dalam pengaturcaraan dinamik.

Langkah 1 : Penilaiawalan

1. Diberi *m* aksara bagi jujukan *x* dan *n* aksara bagi jujukan *y*.

2. Langkah pertama ialah membentuk matriks *F* dengan *m* baris (*m* +1) dan *n* lajur (*n* + 1) yang merupakan saiz bagi jujukan yang ingin dijajarkan. Mengatur baris pertama dari kiri hingga ke kanan dan pada lajur pertama dari atas ke bawah

3. Matriks $F_{(i,j)}$ = markah bagi jajaran terbaik bagi jujukan $x_{(1\ldots i)}$ dan $y_{(1\ldots j)}$

4. Menilaiawal bagi $F_{(i,0)} = F_{(0,j)} = 0$. Seperti pada Rajah 4.2

Sebagai contoh penjajajaran dua jujukan protein Ferredoxin digunakan, sama seperti contoh dalam bab 2. Jujukan tersebut diwakili dengan x = AYKTVLKTPS dan y = ATFKVTLI

		A	**Y**	**K**	**T**	**V**	**L**	**K**	**T**	**P**	**S**
	0	0	0	0	0	0	0	0	0	0	0
A	0										
T	0										
F	0										
K	0										
V	0										
T	0										
L	0										
I	0										

Rajah 4.2 : Penilaiawalan

Langkah 2 : Pengisian Matriks

1. Pengisian matriks memerlukan skema permarkahan iaitu terdiri dari matriks permarkahan mudah dan fungsi jurang penalti yang perlu ditentukan sebelum pengisian matriks dilakukan.

2. Simbol $S_{(i,j)}$ merupakan permarkahan bagi matriks dan -d merupakan jurang penalti. Sebagai contoh menggunakan matriks permarkahan iaitu:

$$S_{(i,j)} = \begin{cases} 2 & x_i = y_j \\ -2 & x_i \neq y_j \end{cases}$$
$$-d = 1$$

   $S_{(i,j)} = 2$ sekiranya aksara dilokasi *i* pada jujukan *x* dan aksara dilokasi *j* pada jujukan *y* adalah sama atau padan.

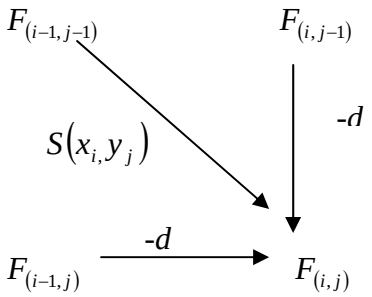
   $S_{(i,j)} = -2$ sekiranya aksara dilokasi *i* pada jujukan *x* dan aksara dilokasi *j* pada jujukan *y* adalah tidak serupa.

   $-d = 1$ merupakan jurang penalti (sekiranya nilai $-d = 0$ tidak mengambilkira jurang penalti).

3. Matriks diisi dari atas kiri dan mencari markah maksimum bagi $F_{(i,j)}$ bagi setiap posisi *i, j* dalam matriks.

4. Bagi mendapatkan markah kesamaan optimal (maksimum) formula di bawah digunakan. Simbol max bermaksud nilai maksimum. Rujuk Rajah 4.3 bagi illustrasi pengiraan formula.

$$F_{(i,j)} = \max \begin{cases} F_{(i-1,j-1)} + S_{(i,j)} & \text{Padan /tidak padanan} \\ F_{(i,j-1)} - d & \text{Jurang pada x (jujukan pertama)} \\ F_{(i-1,j)} - d & \text{Jurang pada y (jujukan kedua)} \\ 0 \end{cases}$$



Rajah 4.3 : Ilustrasi pengiraan markah Smith-Waterman

5. Setelah mendapatkan markah yang maksimum bagi sel, hasil tersebut catatkan di dalam matriks mengikut lokasi bagi sel berserta penuding yang menunjukkan dari mana hasil markah itu diperolehi.

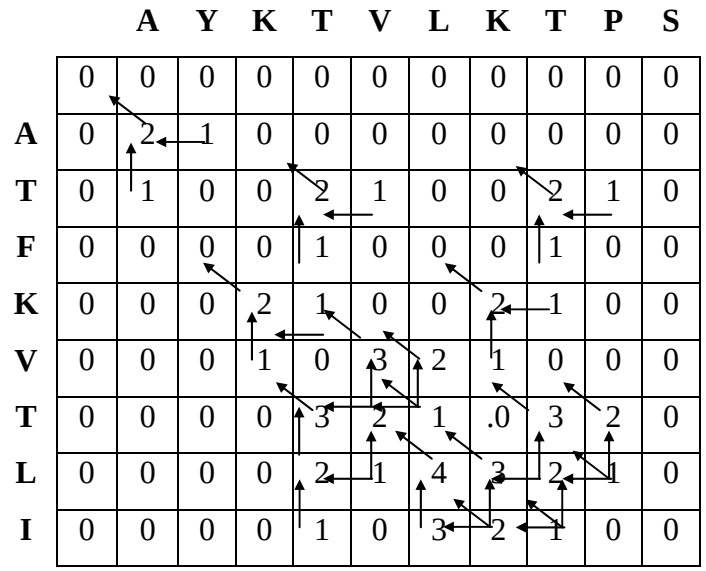
   Berikut merupakan contoh pengisian matriks dengan pada lokasi $F_{(2,4)}$ menggunakan matriks permarkahan pada langkah 2, ke atas contoh jajaran sebelum ini.

   Pada lokasi $F_{(2,4)} = \max\begin{cases} 0+2 \\ 0-1 \\ 0-1 \\ 0 \end{cases} = 2$

6. Ulang langkah 3 sehingga kesemua permarkahan matriks siap diisi. Sila rujuk Rajah 4.4 dan Rajah 4.5.

		A	**Y**	**K**	**T**	**V**	**L**	**K**	**T**	**P**	**S**
	0	0	0	0	0	0	0	0	0	0	0
A	0	2	1	0	0	0	0	0	0	0	0
T	0	1	0	0	2	1	0	0	2	1	0
F	0	0	0	0	1	0	0	0	1	0	0
K	0	0	0	2	1	0	0	2	1	0	0
V	0	0	0	1	0	3	2	1	0	0	0
T	0	0	0	0	3	2	1	.0	3	2	0
L	0	0	0	0	2	1	4	3	2	1	0
I	0	0	0	0	1	0	3	2	1	0	0

Rajah 4.4 : Pengisian matriks pada lokasi $F_{(2,4)}$



Rajah 4.5 : Pengisian penuh matriks

Langkah 3 : Penjejakan balik

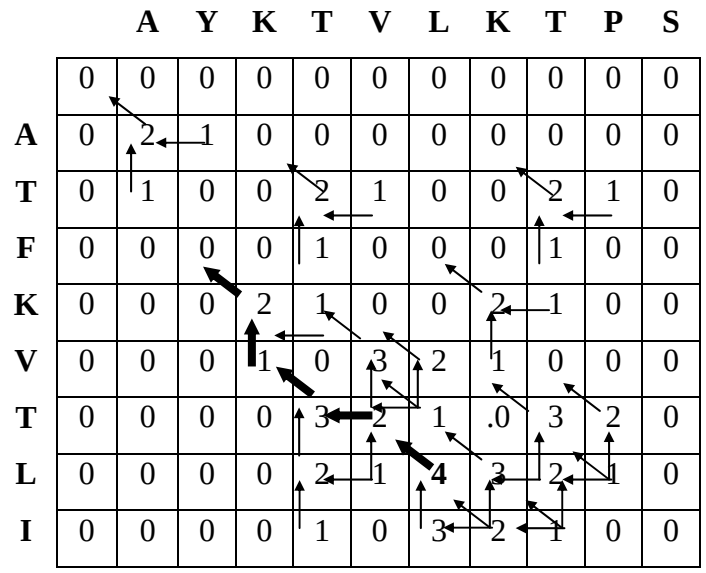
1. Selepas langkah pengisian matriks, dapatkan permarkahan maksimum jajaran setempat bagi dua jujukan tersebut. Iaitu nilai maksimum bagi $F_{(i,j)}$. Berdasarkan Rajah 4.5 permarkahan optimal adalah 4.

2. Langkah penjejakan balik akan menentukan jajaran yang menghasilkan markah yang maksimum. Bermula pada lokasi makimum $F_{(i,j)}$.

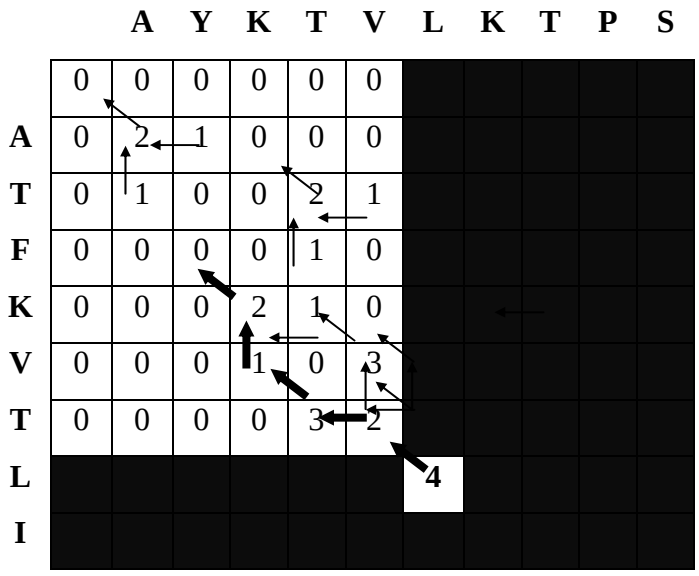
3. Bagi setiap sel, lihat kepada penuding berdasarkan di mana hasil maksimum bagi sel tersebut diperolehi.

4. Ikut laluan yang ditunjukkan oleh penuding, berhenti menjejak sekiranya berjumpa dengan sel bernilai 0.

5. Sila rujuk Rajah 4.6 hingga 4.8 bagi proses menjejak semula.



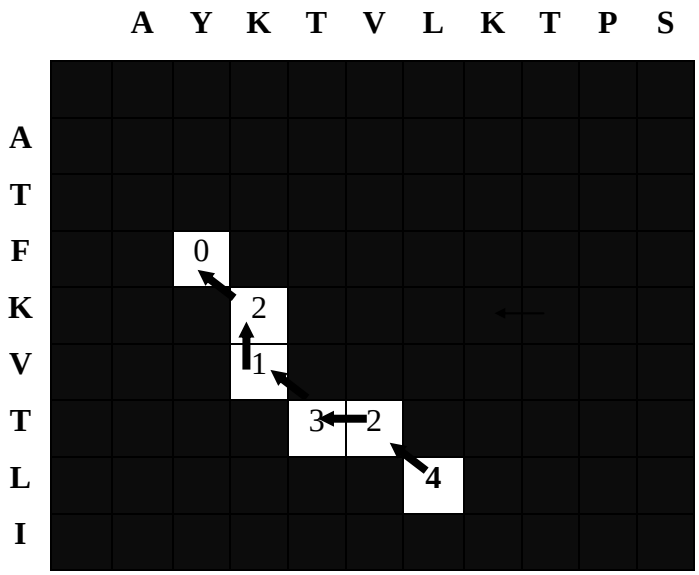
Rajah 4.6 : Langkah pertama proses menjejak semula



Ianya menghasilkan jajaran berikut:

L
|
L

Rajah 4.7 : Langkah kedua proses menjejak semula



Ianya menghasilkan jajaran berikut:

Y K _ T V L
| | |
F K V T _ L

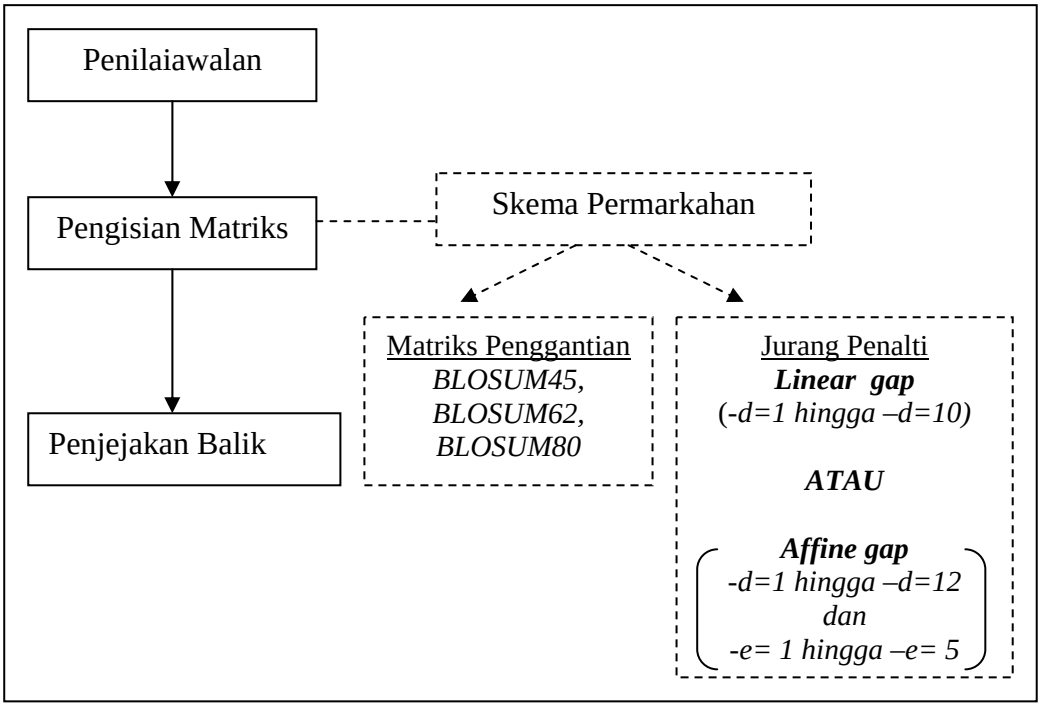
Rajah 4.8 : Langkah terakhir proses menjejak semula

4.4 Model Pengaturcaraan Dinamik Smith-Waterman Yang Diubahsuai

Seterusnya adalah penerangan formulasi bagi model pengaturcaraan dinamik Smith-Waterman yang diubahsuai berdasarkan penyelidikan yang terdahulu untuk tujuan kajian ini. Pengubahsuaian dilakukan terhadap skema permarkahan pengaturcaraan dinamik pada langkah 2, di mana matriks penggantian dan fungsi jurang penalti linear dan affine digunakan. Permasalahannya adalah tiada sebarang panduan bagi pemilihan parameter matriks penggantian dan fungsi jurang penalti bagi penjajaran jujukan. Berdasarkan model ini satu kod aturcara akan dibangunkan pada fasa seterusnya.

Rajah 4.9 menunjukkan skema permarkahan yang akan digunakan dalam pengaturcaran dinamik baru dengan parameter nilai julat yang berbeza. Berdasarkan rajah tersebut terdapat tiga langkah utama dalam pengaturcaraan dinamik. Bagi langkah penilaiawalan dan penjejakan balik caranya sama seperti yang telah

dijelaskan dalam pengaturcaraan dinamik asal. Huraian seterusnya adalah berkaitan langkah pengisian matriks khususnya terhadap skema permarkahan yang bakal digunakan.



Rajah 4.9 : Pengaturcaraan dinamik dengan skema permarkahan berbeza

Pengisian matriks yang merupakan langkah kedua bagi pengaturcaraan dinamik memerlukan skema permarkahan untuk menghasilkan nilai bagi setiap baris dan lajur matriks. Bagi memperbaiki skema permarkahan sedia ada dalam pengaturcaraan dinamik asal, matriks penggantian dan fungsi jurang penalti akan diperkenalkan dalam skema permarkahan.

Skema permarkahan yang akan digunakan dalam projek ini ialah :

(i) Matriks penggantian BLOSUM (α) dengan jurang penalti linear (β_{-d}).

(ii) Matriks penggantian BLOSUM (α) dengan jurang penalti affine ($\delta_{-d,-e}$).

Siri matriks penggantian BLOSUM yang akan digunakan adalah BLOSUM45, BLOSUM62 dan BLOSUM80. Julat nilai bagi –d bagi β_{-d} adalah 1 hingga 10 manakala julat nilai bagi $\delta_{-d,-e}$ adalah 1 hingga 12 dan julat nilai –e adalah 1 hingga 5.

(i) Skema Permarkahan Menggunakan Matriks Penggantian BLOSUM dan Jurang Penalti Linear ($\alpha\beta_{-d}$)

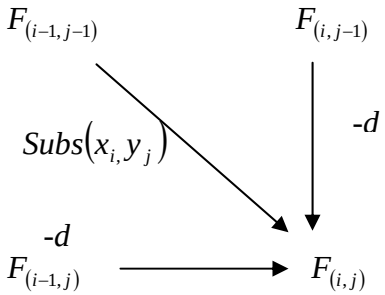
1. Matriks diisi dari atas kiri dan mencari markah maksimum bagi $F_{(i,j)}$ bagi setiap posisi *i, j* dalam matriks. Memerlukan penentuan satu α iaitu $\alpha 45$, $\alpha 62$ atau $\alpha 80$ dan satu nilai bagi jurang penalti *-d* yang akan digunakan iaitu di antara 1 hingga 10.

2. Pengisian markah matriks $Subs_{(i,j)}$ pada setiap posisi menggunakan α .Contoh pengisian matriks menggunakan $\alpha 45$ adalah seperti Rajah 4.11. Sila rujuk di Lampiran E bagi memperolehi $\alpha 45$.

3. Mengira markah kesamaan optimal pada setiap posisi menggunakan formula:

$$F_{(i,j)} = \max\begin{cases} F_{(i-1,j-1)} + Subs_{(i,j)} & \text{Padan /tidak padanan} \\ F_{(i,j-1)} - d & \text{Jurang pada x (jujukan pertama)} \\ F_{(i-1,j)} - d & \text{Jurang pada y (jujukan kedua)} \\ 0 \end{cases}$$

Simbol $Subs_{(i,j)}$ merupakan markah bagi pasangan asid amino dari matriks penggantian dan -d merupakan jurang penalti dan max bermaksud nilai maksimum.



Rajah 4.10 : Ilustrasi pengiraan skema permarkahan $\alpha\beta_{-d}$

Sebagai contoh dengan menggunakan Rajah 4.12 dan nilai –d =2

$$F_{(1,1)} = \max\begin{cases} 0+5 \\ 0-2 \\ 0-2 \\ 0 \end{cases} = 5 \qquad F_{(2,2)} = \max\begin{cases} 5-1 \\ 0-2 \\ 0-2 \\ 0 \end{cases} = 4$$

4. Setelah mendapatkan markah yang maksimum bagi setiap posisi i, j, hasil tersebut catatkan di dalam matriks mengikut lokasi i, j berserta penuding yang menunjukkan dari mana hasil markah itu diperolehi.

5. Ulangi langkah 3 hingga kesemua matriks siap diisi. Sila rujuk rajah di bawah.

6. Seterusnya langkah penjejakan balik adalah sama seperti yang dijelaskan sebelum ini dan hasil jajarannya seperti Rajah 4.12.

		A	**Y**	**K**	**T**	**V**	**L**	**K**	**T**	**P**	**S**
	0	0	0	0	0	0	0	0	0	0	0
A	0	-5	-2	1	0	0	-1	-1	0	-1	1
T	0	0	-1	-1	5	0	-1	-1	5	-1	2
F	0	-2	3	-3	-1	0	1	-3	-1	-3	-2
K	0	-1	-1	5	-1	-2	-3	5	-1	-1	-1
V	0	0	-1	-2	0	5	1	-2	0	-3	-1
T	0	0	-1	-1	5	0	-1	-1	5	-1	2
L	0	-1	0	-3	-1	1	5	-3	-1	-3	-3
I	0	-1	0	-3	-1	3	2	-3	-1	-2	-2

Rajah 4.11 : Pengisian matriks penggantian BLOSUM 45

		A	**Y**	**K**	**T**	**V**	**L**	**K**	**T**	**P**	**S**
	0	0	0	0	0	0	0	0	0	0	0
A	0	5	3	1	0	0	0	0	0	0	1
T	0	3	4	2	6	4	2	0	5	3	2
F	0	1	6	4	4	6	5	3	3	2	1
K	0	0	4	11	9	7	5	10	8	6	4
V	0	0	2	9	11	14	12	10	10	8	6
T	0	0	0	7	14	12	13	11	15	13	11
L	0	0	0	5	12	15	**17**	15	13	12	10
I	0	0	0	3	10	15	17	15	14	12	10

Ianya menghasilkan jajaran berikut:

A - Y K - T V L
| | | | |
A T F K V T - L

Rajah 4.12 : Pengisian matriks dan penjejakan balik

(ii) Skema Permarkahan Menggunakan Matriks Penggantian BLOSUM dan Jurang Penalti Affine $(\alpha\delta_{-d,-e})$

1. Matriks diisi dari atas kiri dan mencari markah maksimum bagi $F_{(i,j)}$ bagi setiap posisi *i, j* dalam matriks. Memerlukan penentuan satu α iaitu $\alpha 45$, $\alpha 62$ atau $\alpha 80$. Manakala bagi $\delta_{-d,-e}$ memerlukan penentuan satu nilai bagi pembukaan jurang *-d* iaitu 1 hingga 12 dan satu nilai tambahan jurang *–e* iaitu di antara 1 hingga 5.

2. Pengisian markah matriks $Subs_{(i,j)}$ pada setiap posisi menggunakan α yang dipilih, sama seperti yang telah dibincangkan sebelum ini.

3. Mengira markah kesamaan optimal pada setiap posisi *i, j* . Bagi mengesan jurang pembukaan tiga jenis jajaran yang diwakilkan dengan matriks harus didefinasikan iaitu :

   $F_{(i,j)}$ = Markah optimal bagi penjajaran terhadap $x_1...x_i$ dan $y_1...y_i$ yang diakhiri tanpa jurang.

   $I_{x(i,j)}$ = Markah optimal bagi penjajaran terhadap $x_1...x_i$ dan $y_1...y_i$ yang diakhiri dengan jurang maksima yang dijajarkan dengan x.

   $I_{y(i,j)}$ = Markah optimal bagi penjajaran terhadap $x_1...x_i$ dan $y_1...y_i$ yang diakhiri dengan jurang maksima yang dijajarkan dengan y.

   Oleh itu penjajaran setempat dengan jurang penalti *affine* boleh dilakukan dalam pengaturcaraan dinamik menggunakan tiga matriks $F_{(i,j)}, I_x, I_y$ [10]. Formula bagi mengira markah maksimum adalah seperti berikut:

$$F_{(i,j)} = \max\begin{cases} F_{(i-1,j-1)} + Subs_{(i,j)} \\ I_{x(i-1,j-1)} + Subs_{(i,j)} \\ I_{y(i-1,j-1)} + Subs_{(i,j)} \\ 0 \end{cases},$$

$$I_{x(i,j)} = \max\begin{cases} F_{(i-1,j)} - d \\ I_{x(i-1,j)} - e \\ I_{y(i-1,j)} - d \end{cases} \quad \text{dan} \quad I_{y(i,j)} = \max\begin{cases} F_{(i,j-1)} - d \\ I_{x(i,j-1)} - d \\ I_{y(i,j-1)} - e \end{cases} \text{ dengan penilaiawalan}$$

$F_{(0,0)} = 0$, $F_{(i,0)} = 0$ bagi $i = 1...m$, $F_{(0,j)} = 0$ bagi $j = 1...n$ selain itu -∞

4. Setelah mendapatkan markah yang maksimum bagi setiap posisi i, j hasil tersebut catatkan di dalam matriks mengikut lokasi i, j berserta penuding yang menunjukkan dari mana hasil markah itu diperolehi. Penuding tersebut perlu disimpan untuk digunakan dalam langkah penjejakan balik.

5. Ulangi langkah 3 hingga kesemua matriks siap diisi.

6. Langkah penjejakan balik sama seperti pengaturcaraan dinamik Smith-Waterman yang asal, perbezaannya penjejakan akan bermula mencari dari maksimum nilai $(F_{(i,j)}, I_{x(i,j),} I_{y(i.,j)})$.

4.5 Ringkasan

Secara umumnya bab ini menerangkan formulasi bagi model pengaturcaraan dinamik untuk penjajaran jujukan khususnya kaedah Smith-Waterman. Perbincangan terhadap formulasi bagi model pengaturcaraan dinamik Smith-Waterman yang diubahsuai dari model pengaturcaraan dinamik Smith-Waterman yang asal. Pengubahsuaian dilakukan terhadap skema permarkahan dalam pengaturcaraan dinamik. Skema permarkahan ini terdiri dari kombinasi parameter matriks penggantian BLOSUM dan jurang penalti linear dan affine. Model pengaturcaraan dinamik yang telah diubahsuai ini bertujuan bagi mempermudahkan pembangunan kod aturcara pada fasa perlaksanaan disamping membantu kajian.

BAB 5

PEMBANGUNAN DAN PERLAKSANAAN PENGATURCARAAN DINAMIK YANG DIUBAHSUAI UNTUK PENJAJARAN JUJUKAN

5.1 Pendahuluan

Bab ini akan membincangan mengenai pembangunan dan perlaksanaan model pengaturcaraan dinamik Smith-Waterman yang telah diubahsuai dengan skema permarkahan berbeza. Pembangunan aturcara ini adalah berdasarkan rekabentuk dan formulasi bagi model pengaturcaraan dinamik yang telah dibina sebelum ini. Sila rujuk Bab 3 bagi rekabentuk dan formulasi modelnya dalam Bab 4. Aturcara ini dibangunkan bertujuan sebagai platform bagi mempermudahkan perlaksanaan kajian iaitu penjajaran jujukan setempat. Bagi melaksanakan penjajaran jujukan data kajian yang dikategorikan mengikut panjang jujukan dan peratusan kesamaan digunakan dan hasil keputusannya direkod. Perlaksanaan dilakukan satu persatu mengikut kategori data dengan menggunakan kombinasi parameter skema permarkahan yang berbeza. Jangkamasa yang panjang diperlukan bagi melaksanakan kesemua kombinasi parameter skema permarkahan dan kompleksiti perlaksanaannya turut dijelaskan dalam bab ini.

5.2 Pembangunan Aturcara Pengaturcaraan Dinamik dengan Skema Permarkahan Berbeza

Berdasarkan rekabentuk dan formulasi bagi model pengaturcaraan dinamik Smith-Waterman yang telah diubahsuai. Sebuah aturcara yang mengkodkan model berkenaan dibangunkan bagi mempermudahkan proses kajian. Aturcara ini dinamakan SWAlign. Objektif pembangunannya, keperluan dan rekabentuk aturcara diperincikan pada bahagian selanjutnya.

5.2.1 Objektif Pembangunan Aturcara

Objektif utama membangunkan SWAlign adalah untuk mengimplementasikan algorithma pengaturcaraan dinamik Smith-Waterman dengan menggunakan skema permarkahan yang berbeza iaitu:

(i) Matriks penggantian BLOSUM (α) dengan jurang penalti linear (β_{-d}).

(ii) Matriks penggantian BLOSUM (α) dengan jurang penalti affine $(\delta_{-d,-e})$.

Formulasi bagi skema permarkahan ini telah diterangkan dalam bab 4. Terdapat 2 kebaikan yang diperolehi dari penghasilan SWAlign. Pertama, ianya merupakan algorithma yang piawai *(standard)* bagi penjajaran jujukan berpasangan setempat yang menghasilkan asas bagi perbandingan keberkesanan. Kedua, ianya digunakan bagi menguji perbezaan kombinasi parameter skema permarkahan yang digunakan.

SWAlign direka adalah untuk pembelajaran dan penilaian algoritma jujukan dalam bidang komputasi biologi. Ianya dikod menggunakan bahasa Java, memandangkan Java merupakan bahasa pengaturcaraan berorientasikan objek yang moden dan popular selain kod aturcaranya mudah dibaca dan boleh diguna semula. Walau apa pun objektifnya adalah mengkod pendek dan semudah mungkin.

5.2.2 Keperluan Aturcara

Keperluan aturcara SWAlign memerlukan tindakbalas pengguna untuk perkara berikut:

(i) Menginput dua jujukan protein.

(ii) Memilih skema permarkahan $(\alpha\beta_{-d})$ atau $(\alpha\delta_{-d,-e})$.

(iii) Sekiranya memilih $(\alpha\beta_{-d})$,
setkan nilai bagi (α) dan setkan nilai penalti –d bagi (β_{-d}).

(iv) Sekiranya memilih $(\alpha\delta_{-d,-e})$,
setkan nilai bagi (α) dan setkan nilai penalti -d dan –e bagi $(\delta_{-d,-e})$.

(v) Penjajaran jujukan yang menghasilkan jajaran optimal berserta markah optimal, panjang jajaran terhasil dan peratusan padanan.

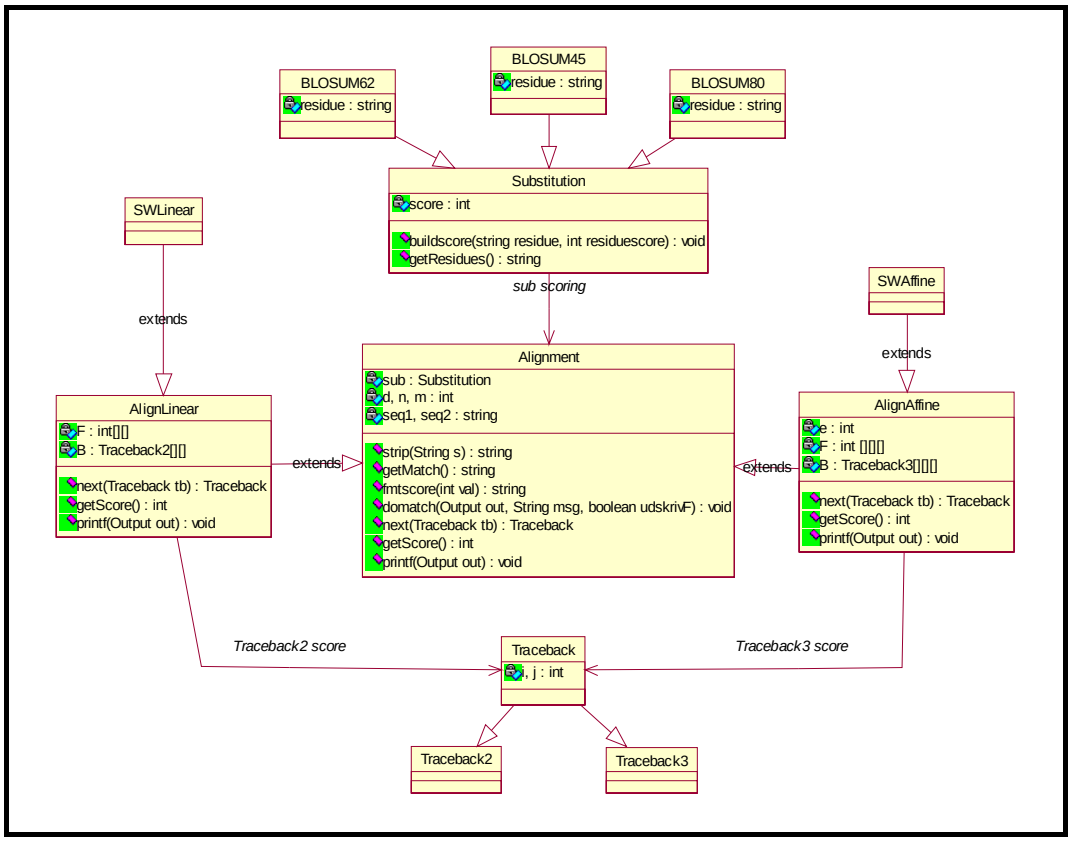
5.2.3 Rekabentuk Aturcara

Terdapat beberapa kelas dalam SWAlign, dua kelas utama bagi mengimplementasi algorithma pengaturcaraan dinamik dengan skema permarkahan berbeza.

(i) SWLinear : mengimplementasikan penjajaran setempat Smith-Waterman dengan skema permarkahan $(\alpha\beta_{-d})$.

(ii) SWAffine : mengimplementasikan penjajaran setempat Smith-Waterman dengan skema permarkahan $(\alpha\delta_{-d,-e})$.

Rajah di bawah menggambarkan perhubungan di antara kelas. Berdasarkan rajah tersebut, kelas BLOSUM45, BLOSUM62, BLOSUM80 mewarisi atribut dan operasi bagi kelas *Substitution*. Begitu juga dengan kelas *Alignment* yang akan diwarisi oleh *AlignLinear* atau *AlignAffine*. Sekiranya penjajaran jujukan dilakukan, kelas *Alignment* akan dipanggil dan seterusnya bergantung kepada kombinasi skema permarkahan yang ingin digunakan. Ini kerana melalui kelas *Alignment*, markah matriks penggantian berdasarkan padanan aksara jujukan pertama dan kedua diperolehi dengan operasi *getresidue()* dalam kelas *Substitution*. Manakala bagi mengkombinasikan markah matriks penggantian dengan fungsi jurang penalti, maka kelas *AlignLinear* dan *AlignAffine* dibina.



Rajah 5.1: Kelas dalam SWAlign

Kelas *Traceback* pula menyimpan penuding dan markah bagi langkah penjejakan balik, di mana *Traceback2* untuk penjejakan balik untuk kelas *SWLinear* manakala *Traceback3* untuk kelas *SWAffine.* Sekiranya penjajaran SW menggunakan skema permarkahan $\alpha\beta_{-d}$ dikehendaki, kelas *SWLinear* akan didefinasikan seperti dalam Rajah 5.2.

```
Procedure SWLinear(sub,d,seq1,seq2)
   super(sub, d, seq1, seq2)
   int n ← this.n, m ← this.m
   int[][] score ← sub.score
   int maxi ← n, maxj ← m
   int maxval ← NegInf
   int i, j
   //Pengiraan kos permarkahan berdasarkan formula linear dalam bab4
   for i ← 1 to n do
     for j ← 1 to m do
       int s ← score[seq1.charAt(i-1)][seq2.charAt(j-1)]
       int val ← max(0, F[i-1][j-1]+s, F[i-1][j]-d, F[i][j-1]-d)
       F[i][j] ← val
       if val = 0 then
          B[i][j] ← null
       else if val = F[i-1][j-1]+s then
          B[i][j] ← new Traceback2(i-1, j-1)
       else if val =F[i-1][j]-d then
          B[i][j] ← new Traceback2(i-1, j)
       else if val = F[i][j-1]-d then
          B[i][j] ← new Traceback2(i, j-1)
       else  throw new Error("SW 1")
        if val > maxval then
          maxval ← val
          maxi ← i; maxj ← j
       endif
     endif
   B0 ← new Traceback2(maxi, maxj)
  repeat
repeat
end SWLinear
```

Rajah 5.2 : Prosedur bagi algoritma SW $\alpha\beta_{-d}$

Manakala penjajaran Smith-Waterman menggunakan skema permarkahan $\alpha\delta_{-d,-e}$, prosedur algoritmanya didefinasikan dalam Rajah 5.3.

```
Procedure SWAffine(sub, d, e, seq1, seq2)
begin
  super(sub, d, e, seq1, seq2)
  int n ← this.n, m← this.m
  int maxi← n, maxj ← m
  int maxval← NegInf
  int[][] score ← sub.score
  int[][] M ←F[0], Ix ← F[1], Iy ← F[2]
  int i,j, val
 //Penilaiawalan matriks
  for i ← 1 to n do
    M[i][0] ← 0
    B[0][i][0] ← new Traceback3(0, i-1,0)
  repeat
  for i ← 1 to n do
     Ix[i][0] ← Iy[i][0] ← NegInf
  repeat
  for j ← 1 to m do
    M[0][j] ← 0
    B[0][0][j] ← new Traceback3(0,0,j-1)
  repeat
  for j ← 1 to m do
     Ix[0][j] ← Iy[0][j] ← NegInf
  repeat
//Pengiraan kos permarkahan berdasarkan formula
affine dalam bab4
  for i ← 1 to n do
  for j ← 1 to m do
  int s ← score[seq1.charAt(i-1)][seq2.charAt(j-1)]
    val ← M[i][j] ← max(0,M[i-1][j-1]+s,
                        Ix[i-1][j-1]+s, Iy[i-1][j-1]+s)
    if val = 0 then
      B[0][i][j] ← null
    else if val = M[i-1][j-1]+s then
      B[0][i][j] ← new Traceback3(0, i-1, j-1)
    else if val = Ix[i-1][j-1]+s then
      B[0][i][j] ← new Traceback3(1, i-1, j-1)
    else if (val = Iy[i-1][j-1]+s then
      B[0][i][j] ← new Traceback3(2, i-1, j-1)
    else throw new Error("SWAffine 1")
      if val > maxval then
        maxval ← val
        maxi ← i, maxj ← j
      endif
    endif
    val ← Ix[i][j] ← max(M[i-1][j]-d, Ix[i-1][j]-e,
                         Iy[i-1][j]-d)
    if val = M[i-1][j]-d then
       B[1][i][j] ← new Traceback3(0, i-1, j)
    else if val = Ix[i-1][j]-e then
       B[1][i][j] ← new Traceback3(1, i-1, j)
    else if val = Iy[i-1][j]-d then
       B[1][i][j] ← new Traceback3(2, i-1, j)
    else throw new Error("SWAffine 2")
       if val > maxval then
         maxval ← val
         maxi ← i; maxj ← j
       endif
    endif
    val ← Iy[i][j] ← max(M[i][j-1]-d, Iy[i][j-1]-e,
                         Ix[i][j-1]-d)
    if val = M[i][j-1]-d then
       B[2][i][j] ← new Traceback3(0, i, j-1)
    else if val = Iy[i][j-1]-e then
       B[2][i][j] ← new Traceback3(2, i, j-1)
    else if val = Ix[i][j-1]-d then
       B[2][i][j] ← new Traceback3(1, i, j-1)
    else throw new Error("SWAffine 3")
       if val > maxval then
         maxval ← val
         maxi ← i, maxj ← j
       endif
    endif
   repeat
  repeat

  // Menentukan markah optimal
   int maxk ← 0
   maxval ← F[0][maxi][maxj]
   for k ← 1 to 3 do
      if maxval < F[k][maxi][maxj] then
        maxval ← F[k][maxi][maxj]
        maxk ← k
      endif
     B0 ← new Traceback3(maxk,maxi,maxj)
   repeat
end SWAffine
```

Rajah 5.3 : Prosedur bagi algoritma SW $\alpha\delta_{-d,-e}$

Berikut adalah prosedur untuk pembinaan matriks penggantian dalam SWAlign.

```
Procedure Substitution
begin
int[][] score
int i, j
 call void buildscore(String residues, int[][] residuescores)
 begin
  // Benarkan aksara lowercase dan uppercase (ASCII code <= 127)
  score ← new int[127][127]
  for i ←0  to residues.length do
   char res1  ← residues.charAt(i)
   for j ←0 to i do
    char res2  ← residues.charAt(j);
    score[res1][res2] ← score[res2][res1]
          ← score[res1][res2+32] ←score[res2+32][res1]
          ←score[res1+32][res2] ← score[res2][res1+32]
          ← score[res1+32][res2+32] ←score[res2+32][res1+32]
         ←residuescores[i][j];
   repeat
  repeat
 end buildscore
 call String getResidues()
end Substitution
```

Rajah 5.4 : Prosedur bagi membina markah matriks penggantian

```
Procedure BLOSUMx extends Substitution
begin
    String residues ← "ARNDCQEGHILKMFPSTWYVBZX"
    call String getResidues()
      return residues
     int[][] residuescores  ←  {//masukkan matriks BLOSUMX // }
     call BLOSUMx()
       buildscore(residues, residuescores)
end BLOSUMx
```

Rajah 5.5 : Prosedur bagi matriks BLOSUM

5.3 Perlaksanaan Penjajaran Jujukan

Perlaksanaan atau larian data dilakukan menggunakan komputer yang mempunyai kelajuan pemproses Pentium 3.1GHz dengan kapasiti ingatan 512RAM dan kapasiti storan 40GB. Bagi melaksanakan penjajaran jujukan 27 set data protein yang telah dikategorikan mengikut panjang jujukan dan peratusan kesamaan digunakan dan hasil keputusannya dicatatkan. Perlaksanaan dibahagikan kepada dua kategori mengikut rekabentuk skema permarkahan seperti yang ditunjukkan dalam Bab 3 iaitu Rajah 3.2. Bagi setiap kategori 3 nilai α yang berbeza digunakan iaitu BLOSUM45, BLOSUM 62 dan BLOSUM 80 berserta dengan nilai bagi jurang penalti linear (β_{-d}) atau affine $(\delta_{-d,-e})$ yang berlainan.

Perlaksanaan dilakukan satu persatu dengan menggunakan kombinasi parameter skema permarkahan yang berbeza iaitu $\alpha\beta_{-d}$ dan $\alpha\delta_{-d,-e}$ pada setiap set data. Berdasarkan Rajah 3.2 dalam Bab 3, sebanyak 30 kali laluan perlu dilalui bagi melaksanakan parameter skema permarkahan $\alpha\beta_{-d}$ dan 180 kali laluan bagi melaksanakan parameter skema permarkahan $\alpha\delta_{-d,-e}$. Rajah berikut merupakan pengiraan kompleksiti masa perlaksanaan kerja per jam bagi kajian ini.

Matriks Penggantian & Jurang Penalti Linear	Matriks Penggantian & Jurang Penalti Affine
Laluan bagi 1pasang data = 3×10 = 30 kali	Laluan bagi 1pasang data = $3\times12\times5$ = 180 kali
Data Kajian: Pendek = 9 set pasangan; Sederhana = 9 set pasangan; Panjang = 9 set pasangan	Data Kajian: Pendek = 9 set pasangan; Sederhana = 9 set pasangan; Panjang = 9 set pasangan
Jumlah = 27 set pasangan	Jumlah = 27 set pasangan
Jumlah ms diperlukan = $30\times27\times5$ (minit) = 4050 minit = 67jam 30 minit	Jumlah ms diperlukan = $180\times27\times5$ (minit) = 24300 minit = 405jam

Rajah 5.6 : Pengiraan kompleksiti masa perlaksanaan

Oleh kerana masalah kekangan masa hanya 27 set data protein akan dilaksanakan bagi kedua-dua skema permarkahan. Ini kerana berdasarkan Rajah 5.6 laluan yang diperlukan adalah 180 kali bagi sepasang data dengan masa 5 minit menggunakan skema $\alpha\delta_{-d,-e}$. Jika 27 set data dilaksanakan jumlah masa yang diperlukan adalah 405 jam atau lebih kurang 51 hari jika 8 jam bekerja per hari. Maka, keseluruhan masa perlaksanaan bagi menghasilkan setiap keputusan menggunakan nilai parameter skema permarkahan yang berbeza adalah 472 jam 30 minit atau bersamaan lebih kurang 60 hari jika 8 jam bekerja per hari.

Jangkamasa hampir 2 bulan merupakan jangkamasa yang agak lama bagi melaksanakan penjajaran, ini tidak termasuk proses-proses yang perlu dilaksanakan selepas hasil diperoleh. Oleh itu hanya 27 set data sahaja yang dilaksanakan. Penyelidik terdahulu yang mengunakan jajaran jujukan berpasangan ada menyatakan 5 hingga 9 set pasangan sudah memadai untuk dijadikan data kajian [28, 29, 8]. Hasil keputusan bagi setiap larian direkodkan bagi tujuan penilaian. Penjelasan lanjut tentang proses-proses terlibat bagi mengukur keberkesanan dari hasil keputusan larian akan diperincikan dalam bab seterusnya.

5.4 Ringkasan

SWAlign merupakan aturcara java yang dikod berdasarkan rekabentuk dan formulasi bagi model pengaturcaraan dinamik Smith-Waterman yang diubahsuai. Ianya dibangunkan bertujuan menjadi platform bagi perlaksanaan data kajian. Seterusnya adalah fasa perlaksanaan atau larian data yang mana merupakan fasa yang agak lama dan segala turutan aktivitinya perlu dilakukan dengan sabar dan teliti. Ini kerana hasil keputusan yang diperolehi akan diproses untuk dinilai pada fasa seterusnya.

BAB 6

ANALISA KEPUTUSAN DAN PERBINCANGAN TERHADAP PARAMETER MATRIKS PENGGANTIAN BLOSUM DAN JURANG PENALTI LINEAR DALAM PENGATURCARAAN DINAMIK

6.1 Pendahuluan

Secara umumnya bab ini akan membincangkan hasil larian penjajaran jujukan menggunakan parameter skema permarkahan yang terdiri dari matriks pengantian BLOSUM dan jurang penalti linear dalam pengaturcaraan dinamik Smith-Waterman yang diperoleh dari fasa perlaksanaan. Hasil larian tersebut akan diolah bagi mendapatkan keputusan akhir yang lebih baik. Seterusnya adalah proses analisa keputusan dan perbincangan terhadap hasil keputusan akhir yang diperoleh dengan menggunakan kombinasi parameter skema permarkahan yang berbeza mengikut kategori data kajian.

6.2 Proses Olahan Hasil Larian

Bagi memperolehi hasil, 27 set data jujukan protein akan dijajarkan menggunakan kombinasi parameter matriks penggantian dan nilai jurang penalti linear yang berbeza. Seperti yang telah dijelaskan dalam bab 5, 30 pengulangan perlu dilakukan bagi melaksanakan satu set pasangan. Hasil jajaran yang diperoleh akan direkodkan dalam jadual. Satu jadual dibina bagi setiap pasangan jajaran seperti contoh rajah di bawah.

Kategori Data : Pendek <25%ID (P02775 128bp & P04981 133bp)

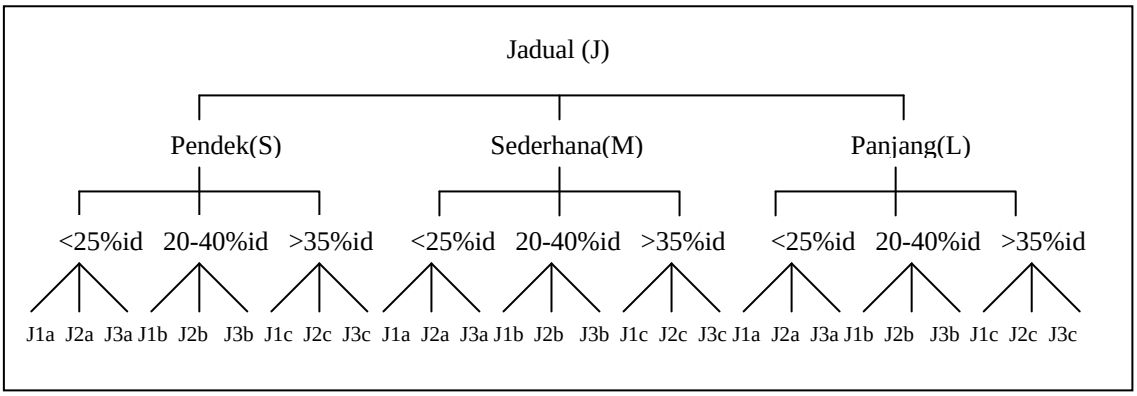
Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
	BLOSUM45	Markah	229	171	126	96	69	49	45	42	40	40
		Panjang Jajaran	168	152	127	126	113	60	53	53	21	21
		Padanan	26.79%	26.97%	28.35%	27.78%	28.32%	21.67%	20.75%	20.75%	33.33%	33.33%
	BLOSUM62	Markah	181	126	90	65	45	39	38	38	38	38
Pendek		Panjang Jajaran	162	148	114	104	103	60	15	15	15	15
<25%		Padanan	25.93%	27.03%	28.07%	26.92%	27.18%	25.00%	40.00%	40.00%	40.00%	40.00%
	BLOSUM80	Markah	180	112	75	46	34	33	32	31	30	29
		Panjang Jajaran	170	132	115	75	16	16	16	16	16	12
		Padanan	26.47%	28.03%	28.70%	32.00%	43.75%	43.75%	43.75%	43.75%	43.75%	41.67%

Rajah 6.1: Contoh jadual hasil SWLinear bagi jajaran sepasang jujukan

Hasil jajaran yang diambilkira bagi tujuan penganalisaan adalah markah kesamaan optimal, panjang jajaran yang terhasil dan peratusan aksara yang padan dari jajaran yang terhasil. Jajaran yang baik boleh dikenalpasti dengan mencari markah tertinggi atau maksimum bagi setiap ciri hasil yang diperoleh. Oleh itu, proses olahan hasil jajaran perlu dilakukan bagi mempermudahkan proses analisa. Huraian seterusnya merupakan proses-proses yang dilakukan untuk menghasilkan keputusan akhir.

6.2.1 Penjumlahan Jadual Hasil Larian

Setiap pasangan jujukan mempunyai satu jadual hasil seperti di atas. Oleh itu, terdapat 9 jadual hasil bagi setiap kategori jujukan iaitu pendek, sederhana dan panjang. Bagi setiap kategori dipecahkan kepada 3 kategori peratusan kesamaan identiti yang diwakilkan dengan 3 jadual. Rujuk Rajah 6.2.



Rajah 6.2: Ilustrasi jadual hasil

Bagi membezakan hasil berdasarkan kategori panjang jajaran satu jadual hasil perlu dihasilkan iaitu dengan mencampurkan kesemua jadual mengikut kategori data yang sama dan mengambil markah purata bagi markah optimal, jajaran dan peratusan padanan.

Menggunakan formula berikut :

$$J_x = \frac{(J1a + J2a + J3a + J1b + J2b + J3b + J1c + J2c + J3c)_x}{9} \quad ,$$

x mewakili kategori panjang jujukan. Hasilnya adalah jadual J_S, J_M dan J_L. Bagi membezakan hasil berdasarkan peratusan kesamaan identiti formula berikut digunakan.

$$Ja_x = \frac{J1a_x + J2a_x + J3a_x}{3}, \quad Jb_x = \frac{J1b_x + J2b_x + J3b_x}{3}, \quad Jc_x = \frac{J1c_x + J2c_x + J3c_x}{3}$$

Hasilnya adalah jadual $Ja_S, Jb_S, Jc_S, Ja_M, Jb_M, Jc_M, Ja_L, Jb_L$ dan Jc_L.

Contoh jadual kategori pendek yang dihasilkan seperti Rajah 6.3 atau Rujuk Lampiran I bagi keseluruhan hasil penjumlahan mengikut kategori data.

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
	BLOSUM45	Markah	257.11	226.78	206.00	191.33	178.22	167.11	160.11	154.33	149.67	145.44
		Panjang Jajaran	111.56	101.44	92.56	92.00	89.22	79.22	72.22	65.44	61.89	60.00
		Padanan	37.71%	38.38%	38.57%	38.33%	38.05%	37.57%	38.11%	38.85%	40.25%	41.16%
	BLOSUM62	Markah	207.22	177.11	158.56	144.78	133.67	126.89	122.11	117.78	114.56	111.78
Pendek		Panjang Jajaran	111.00	97.11	92.00	88.00	79.00	67.78	59.67	55.67	52.11	52.11
		Padanan	37.83%	39.49%	38.75%	38.27%	39.32%	39.47%	41.80%	41.84%	42.41%	42.41%
	BLOSUM80	Markah	213.89	177.33	155.89	141.11	131.00	124.78	120.33	116.22	112.67	110.00
		Panjang Jajaran	111.78	96.22	90.22	84.00	66.67	58.67	57.11	53.11	51.78	44.00
		Padanan	38.24%	40.19%	40.10%	39.61%	42.03%	43.25%	47.26%	47.30%	48.05%	50.30%

Rajah 6.3 : Jadual hasil SWLinear bagi data kategori pendek (J_S)

6.2.2 Pernormalan Hasil Menggunakan *Z-score*

Setelah jadual baru diperoleh seperti contoh jadual Rajah 6.3, proses penilaian terhadap kombinasi parameter skema permarkahan jajaran dilakukan iaitu dengan mencari nilai maksimum bagi markah optimal, panjang jajaran dan peratusan padanan. Permasalahannya adalah, format bagi hasil tidak sama dan julat perbezaannya agak jauh. Oleh itu, proses pernormalan perlu dilakukan bagi menseragamkan format untuk memudahkan proses penilaian. *Z-score* merupakan satu cara statistik untuk menpiawaikan data menjadi satu skala supaya analisa keputusan boleh boleh dilakukan menggunakan markah tersebut. *Z-score* boleh digunakan untuk semua jenis data, setiap nilai *Z-score* sejajar dengan titik pengagihan normal *(normal distribution).*

Z-score boleh dikira menggunakan formula berikut:

$$Z = \frac{(x-\mu)}{\sigma} = \frac{titik\_data - mean}{sisihan\_piawai},$$

$$\text{mean} = \frac{\sum X}{N}, \text{ varian } = \sigma^2 = \frac{\sum (x-\mu)^2}{N} \text{ dan } \sigma = \sqrt{\sigma}$$

Manakala formula bagi pengiraan *Z-score* berdasarkan jadual ujikaji ialah seperti di bawah. Simbol $\alpha\beta_{-d}$ merupakan kombinasi skema permarkahan matriks BLOSUM dengan jurang penalti linear. α = B45, B62 dan B80 manakala M mewakili nilai bagi markah optimal, panjang jajaran dan padanan.

Mean bagi $M\alpha\beta_{-d} = \mu M\alpha\beta_{-d} = \dfrac{M\alpha\beta_{-1} + M\alpha\beta_{-2} + M\alpha\beta_{-3}...M\alpha\beta_{-10}}{10}$,

Sisihan piawai bagi $M\alpha\beta_{-d} = \sqrt{\dfrac{\sum_{i=1}^{10}\left(M\alpha\beta_{-i} - \mu M\alpha\beta_{-d}\right)^2}{10}}$

$$= \sqrt{\frac{\left(M\alpha\beta_{-1} - \mu M\alpha\beta_{-d}\right)^2 + ... + \left(M\alpha\beta_{-10} - \mu M\alpha\beta_{-d}\right)^2}{10}}$$

Z-score bagi $M\alpha\beta_{-i} = \dfrac{M\alpha\beta_{-i} - \mu M\alpha\beta_{-d}}{\sigma M\alpha\beta_{-d}}, i = 1...10$

Rajah 6.4 merupakan contoh hasil pernormalan *Z-score* ke atas setiap nilai jadual bagi kategori data pendek J_S pada Rajah 6.3. Rujuk Lampiran J.

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
	BLOSUM45	z-score Markah	2.00	1.18	0.61	0.21	-0.15	-0.45	-0.64	-0.80	-0.92	-1.04
		z-score Panjang	1.65	1.08	0.57	0.54	0.38	-0.19	-0.59	-0.97	-1.18	-1.28
		z-score Padanan	-0.87	-0.28	-0.11	-0.32	-0.57	-0.98	-0.52	0.14	1.36	2.15
Pendek	BLOSUM62	z-score Markah	2.11	1.15	0.11	0.11	-0.25	-0.47	-0.62	-0.76	-0.86	-0.95
		z-score Panjang	1.69	1.03	0.79	0.60	0.17	-0.37	-0.75	-0.94	-1.11	-1.11
		z-score Padanan	-1.32	-0.38	-0.79	-0.47	-0.47	-0.39	0.93	0.95	1.27	1.27
	BLOSUM80	z-score Markah	2.21	1.11	0.47	0.02	-0.28	-0.47	-0.60	-0.72	-0.83	-0.91
		z-score Panjang	1.78	1.10	0.83	0.56	-0.21	-0.56	-0.63	-0.81	-0.86	-1.21
		z-score Padanan	-1.27	-0.81	-0.83	-0.95	-0.38	-0.09	0.85	0.86	1.04	1.57

Rajah 6.4: Contoh jadual SWLinear dengan *Z-score*

6.2.3 Pengabungan Hasil *Z-score* Menggunakan *RZ-score*

Hasil yang diperoleh dari proses pernormalan ialah *Z-score* bagi markah optimal, panjang jajaran dan padanan. Jajaran yang baik harus mempunyai nilai maksima bagi markah optimal, panjang jajaran dan padanan. Persoalannya disini, bagaimana memperoleh hanya satu nilai maksima yang memenuhi kesemua ciri ini. Satu teknik ukuran yang dinamakan *Goodness of Hit* atau lebih dikenali sebagai *G-H score* telah diperkenalkan oleh Gunry dan Henry [12, 14]. Ukuran ini digunakan bagi mengira peratusan kompaun aktif dan tidak aktif dalam pangkalan data kimia. Bertujuan mengukur dan menganalisa kualiti bagi hasil pengkelasan data.

Berikut adalah formula *G-H score:*

$$G\text{-}H\ score = \frac{\alpha P + \beta R}{2},$$

Simbol α dan β adalah pemberat yang mengambarkan kepentingan relatif bagi *recall* dan *precision*. Jika pemberat disetkan sebagai uniti, maka markah adalah mean bagi *recall* dan *precision* iaitu $G\text{-}H\ score = \frac{P+R}{2}$. Berdasarkan idea dan pengubahsuaian formula dari *G-H score*, satu ukuran baru yang dinamakan *Reform of Z-score (RZ-score)* diperkenalkan bagi mengatasi permasalahan yang wujud.

Berikut merupakan formula bagi *RZ-score*:

$RZ\text{-}score = \frac{Z_{OM} + Z_{AL} + Z_{CA}}{3}$, di mana Z = *z-score* ,

OM = Markah Kesamaan Optimal *(Optimal Mark),*

AL = Panjang Hasil Jajaran *(Alignment Length)* dan

CA = Padanan Jajaran *(Correct Alignment)*

Oleh itu, proses seterusnya selepas pernormalan hasil adalah mengira *RZ-score* yang akan menghasilkan satu nilai maksima memenuhi tiga ciri tersebut. Hasil pengiraan *RZ-score* ini direkodkan dalam jadual, seperti contoh Rajah 6.5 iaitu

jadual *RZ-score* bagi kategori data pendek. Rujuk Lampiran K dan L bagi kesemua jadual SWLinear.

Data	Matriks\ -d	1	2	3	4	5	6	7	8	9	10
	BLOSUM45	0.93	0.66	0.36	0.14	-0.11	-0.54	-0.58	-0.54	-0.25	-0.06
Pendek	BLOSUM62	0.83	0.60	0.03	0.08	-0.18	-0.41	-0.15	-0.25	-0.24	-0.27
	BLOSUM80	0.91	0.47	0.16	-0.12	-0.29	-0.37	-0.13	-0.22	-0.22	-0.18

Rajah 6.5: Contoh jadual SWLinear dengan *RZ-score*

6.2.4 Menjana Graf

Bagi memudahkan proses penilaian terhadap hasil keputusan *RZ-score*, proses menjana graf berdasarkan jadual keputusan merupakan langkah terakhir yang perlu dilakukan. Ianya bertujuan untuk membuat analisa perbandingan keberkesanan parameter skema permarkahan sekaligus dapat mencari kombinasi parameter skema permarkahan yang terbaik

6.3 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM dan Jurang Penalti Linear

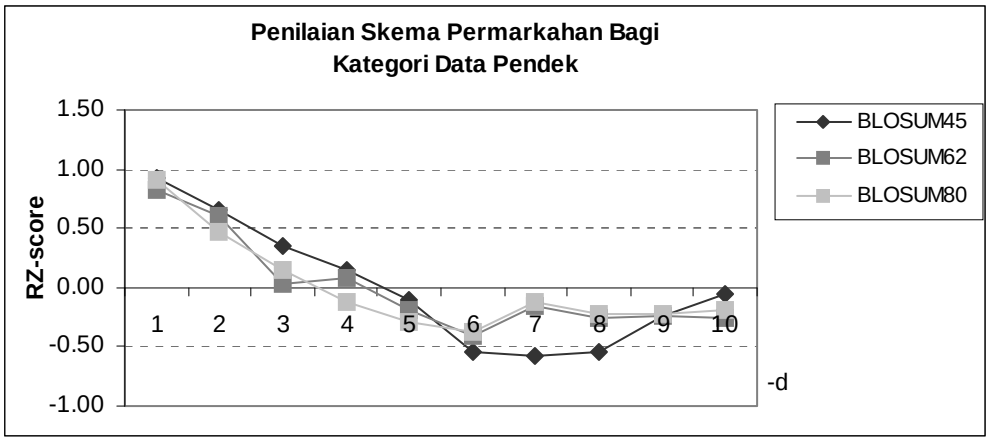
Hasil ujikaji penjajaran jujukan yang dilakukan bertujuan untuk menganalisa keberkesanan kombinasi matriks penggantian BLOSUM dan fungsi jurang penalti linear yang digunakan dalam skema permarkahan pengaturcaraan dinamik Smith-Waterman. Berikut merupakan perincian hasil ujikaji penjajaran mengikut kategori data jujukan yang berbeza.

6.3.1 Hasil Ujikaji Penjajaran Jujukan Bagi Kategori Data Jujukan Pendek

Hasil keputusan penjajaran jujukan pendek yang digunakan terdiri dari koleksi set jujukan protein pendek yang mempunyai peratusan kesamaan identiti yng berbeza iaitu <25%, 20-40% dan >35%. Oleh itu analisa hasil keputusan *RZ-score* terhadap parameter skema permarkahan yang terdiri dari matriks penggantian BLOSUM dan jurang penalti linear, dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti.

6.3.1.1 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM

Berdasarkan graf pada Rajah 6.6 dan Lampiran K, nilai *RZ-score* lebih tinggi dengan menggunakan BLOSUM45 pada julat nilai jurang penalti –d 1 hingga 5 berbanding BLOSUM62. Manakala jika dilihat dari kategori data pendek dengan peratusan kesamaan identiti yang berbeza seperti dalam Lampiran L.



Rajah 6.6: Hasil ujikaji SWLinear bagi kategori data pendek

Kesimpulan yang boleh dibuat untuk julat nilai –d 1 hingga 2 adalah seperti berikut:

(i) Kategori pendek <25% identiti, didapati BLOSUM80 lebih baik.

(ii) Kategori pendek 20-40% identiti, didapati BLOSUM62 lebih baik.

(iii) Kategori pendek >35% identiti, didapati BLOSUM62 lebih baik.

Ini menunjukkan peratusan kesamaan mendatangkan kesan terhadap hasil jajaran dan ianya berperanan menentukan pemilihan parameter siri BLOSUM bagi menghasilkan jajaran yang lebih berkesan. Secara umumnya bagi data berjujukan pendek, tiada matriks penggantian yang dapat disarankan. Ini kerana secara purata nilai *RZ-score* tertinggi bagi data pendek ialah BLOSUM45 (seperti Rajah 6.6) tetapi hasilnya bercanggah sekiranya menggunakan data pendek dengan peratusan identiti berbeza..

6.3.1.2 Analisa Keputusan Terhadap Parameter Jurang Penalti Linear

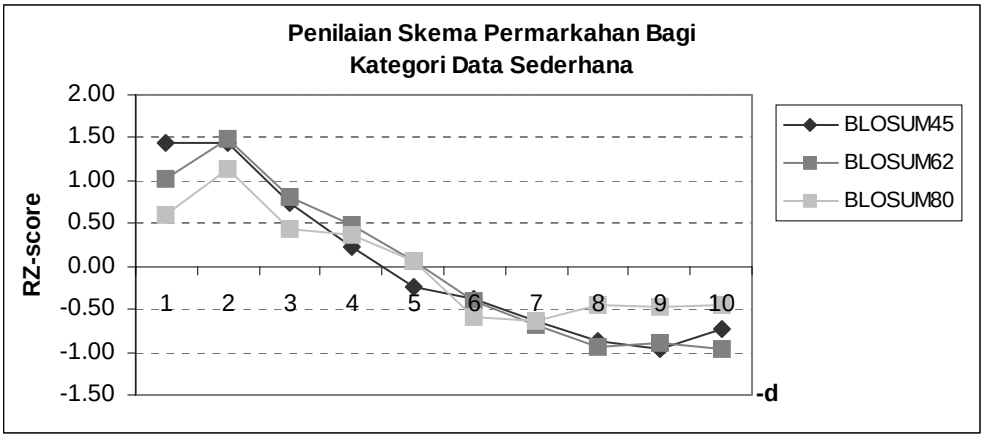
Graf pada Rajah 6.6 dan jadual di lampiran K dan L dirujuk dalam menganalisa parameter nilai jurang penalti linear (-d) yang digunakan dengan kombinasi matriks penggantian BLOSUM. Berdasarkan sumber yang dinyatakan, di dapati nilai jurang penalti linear yang paling efektif bagi data pendek adalah 1 hingga 2. Nilai yang sama juga diperolehi walaupun menggunakan data pendek dengan perbezaan peratusan identiti. Semakin besar nilai –d semakin kurang markah *RZ-score*, sekiranya nilai –d terlalu besar maka markah *RZ-score* menjadi tidak menentu atau boleh dikatakan penjajaran padanan menjadi semakin kurang. Keadaan ini berlaku apabila nilai –d >6 seperti yang ditunjukkan dalam graf.

6.3.2 Hasil Ujikaji Penjajaran Jujukan Bagi Kategori Data Sederhana

Hasil keputusan penjajaran jujukan sederhana yang digunakan terdiri dari koleksi set jujukan protein sederhana yang mempunyai peratusan kesamaan identiti yng berbeza iaitu <25%, 20-40% dan >35%. Oleh itu analisa hasil keputusan *RZ-score* terhadap parameter skema permarkahan yang terdiri dari matriks penggantian BLOSUM dan jurang penalti linear, dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti.

6.3.2.1 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM

Berdasarkan graf pada Rajah 6.7 dan Lampiran K, nilai *RZ-score* lebih tinggi dengan menggunakan BLOSUM62 pada julat nilai jurang penalti –d 1 hingga 7 berbanding parameter siri matriks yang lain. Manakala jika dilihat dari kategori data sederhana dengan peratusan kesamaan identiti yang berbeza seperti dalam Lampiran L.



Rajah 6.7: Hasil ujikaji SWLinear bagi kategori data sederhana

Kesimpulan yang boleh dibuat untuk julat nilai –d 1 hingga 2 adalah seperti berikut:

(i) Kategori sederhana <25% identiti, didapati BLOSUM62 lebih baik.

(ii) Kategori sederhana 20-40% identiti, didapati BLOSUM62 lebih baik.

(iii) Kategori sederhana >35% identiti, didapati BLOSUM62 lebih baik.

Ini menunjukkan peratusan kesamaan mendatangkan kesan terhadap hasil jajaran dan ianya berperanan menentukan pemilihan parameter siri BLOSUM bagi menghasilkan jajaran yang lebih berkesan. Secara umumnya jika data jujukan jajaran adalah sederhana maka BLOSUM62 disarankan.

6.3.2.2 Analisa Keputusan Terhadap Parameter Jurang Penalti Linear

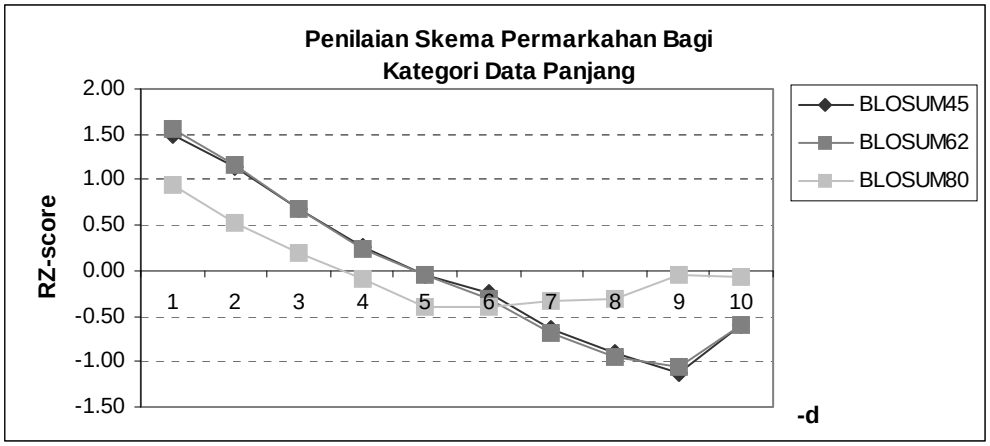
Graf pada Rajah 6.7 dan jadual di Lampiran K dan L dirujuk dalam menganalisa parameter nilai jurang penalti linear (-d) yang digunakan dengan kombinasi matriks penggantian BLOSUM. Berdasarkan sumber yang dinyatakan, didapati nilai jurang penalti linear yang paling efektif bagi data sederhana adalah di antara julat 1 dan 2. Nilai yang sama juga diperolehi walaupun menggunakan data sederhana dengan perbezaan peratusan identiti. Semakin besar nilai –d semakin kurang markah *RZ-score*, sekiranya nilai –d terlalu besar maka markah *RZ-score* menjadi tidak menentu atau boleh dikatakan penjajaran padanan menjadi semakin kurang. Keadaan ini berlaku apabila nilai –d >7 seperti yang ditunjukkan dalam graf.

6.3.3 Hasil Ujikaji Penjajaran Bagi Kategori Data Panjang

Hasil keputusan penjajaran jujukan panjang yang digunakan terdiri dari koleksi set jujukan protein panjang yang mempunyai peratusan kesamaan identiti yng berbeza iaitu <25%, 20-40% dan >35%. Oleh itu analisa hasil keputusan *RZ-score* terhadap parameter skema permarkahan yang terdiri dari matriks penggantian BLOSUM dan jurang penalti, dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti.

6.3.3.1 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM

Berdasarkan Rajah 6.8 dan Lampiran K, nilai *RZ-score* lebih tinggi dengan menggunakan parameter siri BLOSUM62 dan BLOSUM45 pada julat nilai jurang penalti –d 1 hingga 6, nilai *RZ-score* bagi kedua siri matriks ini hampir serupa bagi setiap perubahan nilai –d . Manakala jika dilihat dari kategori data panjang dengan peratusan kesamaan identiti yang berbeza seperti dalam Lampiran L.



Rajah 6.8: Hasil ujikaji SWLinear bagi kategori data panjang

Kesimpulan yang boleh dibuat untuk julat nilai –d 1 hingga 2 adalah seperti berikut:

(i) Kategori panjang <25% identiti, didapati BLOSUM45 lebih baik.

(ii) Kategori panjang 20-40% identiti, didapati BLOSUM45 lebih baik.

(iii) Kategori panjang >35% identiti, didapati BLOSUM62 lebih baik.

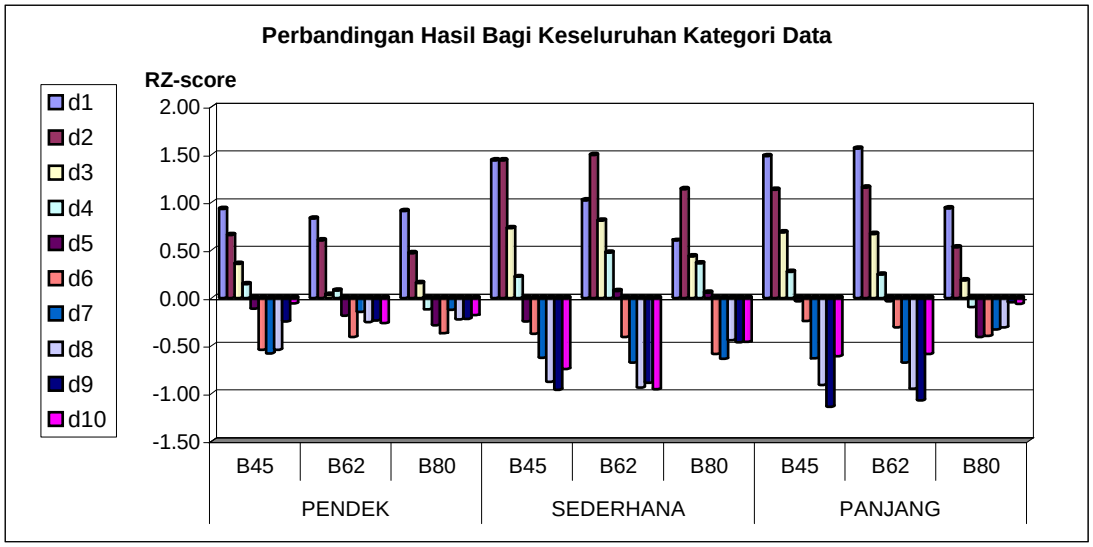
Ini menunjukkan peratusan kesamaan mendatangkan kesan terhadap hasil jajaran dan ianya berperanan menentukan pemilihan parameter siri BLOSUM bagi menghasilkan jajaran yang lebih berkesan. Secara umumnya bagi data berjujukan panjang, disarankan matriks penggantian BLOSUM62 dan BLOSUM45. Ini kerana hasil yang diperolehi secara purata mengikut kategori panjang jujukan hampir sama dengan hasil yang diperolehi mengikut sub kategori peratusan identiti.

6.3.3.2 Analisa Hasil Keputusan Terhadap Parameter Jurang Penalti Linear

Graf pada Rajah 6.8 dan jadual di Lampiran K dirujuk dalam menganalisa parameter nilai jurang penalti linear (-d) yang digunakan dengan kombinasi matriks penggantian BLOSUM. Berdasarkan sumber yang dinyatakan, di dapati nilai jurang penalti linear yang paling efektif bagi data panjang adalah di antara julat 1. Nilai yang sama juga diperolehi walaupun menggunakan data panjang dengan perbezaan peratusan identiti. Semakin besar nilai –d semakin kurang markah *RZ-score*, sekiranya nilai –d terlalu besar maka markah *RZ-score* menjadi tidak menentu atau boleh dikatakan penjajaran padanan menjadi semakin kurang. Keadaan ini berlaku apabila nilai –d >6 seperti yang ditunjukkan dalam graf.

6.4 Perbincangan

Berikut merupakan graf keseluruhan keputusan mengikut kategori panjang jujukan menggunakan skema permarkahan yang berbeza.



Rajah 6.9 : Graf perbandingan hasil SWLinear

Hasil yang diperoleh dari kajian ini dapat membuktikan secara empirikal keberkesanan parameter skema permarkahan yang digunakan bagi menghasilkan jajaran yang optimal. Berdasarkan jadual analisa hasil pada rajah 6.10 didapati julat nilai jurang penalti linear yang kecil adalah lebih efektif bagi menghasilkan markah *RZ-score* yang lebih tinggi. Jika diperhatikan sekiranya julat nilai –d yang besar diguna didapati perubahan markah *RZ-score* semakin berkurang dan menghampiri sifar seterusnya markah akan menjadi tidak menentu. Keadaan ini berlaku disebabkan kemungkinan jajaran yang terhasil terlalu pendek yang menyebabkan padanan yang terhasil terlalu sedikit dan tidak memberi makna. Hasil ini berjaya membuktikan teori bagi jurang penalti bahawa hasil jajaran boleh diperbaiki dengan memperkenalkan jurang [6, 33]. Oleh itu parameter julat nilai jurang penalti linear yang dicadangkan adalah 1 dan 2.

Ciri Jujukan		**Nilai Parameter Cadangan**	
Kategori Data	**Peratusan Identiti**	**Nilai –d**	**Siri BLOSUM**
Pendek	Secara Purata	1-5	BLOSUM45
	<25%	1-5	BLOSUM80
	20%-40%	1-2	BLOSUM62
	>35%	1-2	BLOSUM62
Nilai Parameter Saranan Bagi Kategori Data Pendek		Tidak dapat ditentukan.	
Sederhana	Secara Purata	1-7	BLOSUM62
	<25%	1-2	BLOSUM62
	20%-40%	1-2	BLOSUM62
	>35%	1-2	BLOSUM62
Nilai Parameter Saranan Bagi Kategori Data Sederhana		1-2	BLOSUM62
Panjang	Secara Purata	1-6	BLOSUM62 BLOSUM45
	<25%	1-2	BLOSUM62
	20%-40%	1-2	BLOSUM45
	>35%	1-2	BLOSUM45
Nilai Parameter Saranan Bagi Kategori Data Panjang		1-2	BLOSUM62 & BLOSUM45

Rajah 6.10 : Analisa hasil keputusan bagi SW$\alpha\beta_{-d}$

Bagi keputusan parameter siri matriks penggantian BLOSUM pula, faktor panjang jujukan dan peratusan kesamaan memainkan peranan dalam menentukan pemilihan matriks penggantian BLOSUM. Analisa dilakukan mengikut kategori data jujukan satu persatu, bagi data pendek cadangan siri BLOSUM tidak dapat ditentukan. Ini kerana ketidakseragaman pilihan matriks penggantian BLOSUM yang efektif, kemungkinan disebabkan saiz jujukan yang terlalu pendek atau ketidakserasian data menggunakan skema permarkahan tersebut. Manakala bagi

kategori data berjujukan sederhana terbukti penggunaan BLOSUM62 lebih efektif berbanding siri BLOSUM lain bagi julat nilai 1 hingga 2. Ini kerana walaupun diuji menggunakan data yang mempunyai peratusan identiti yang berbeza namun BLOSUM62 tetap yang terbaik. Bagi data kategori panjang pula, disyorkan menggunakan BLOSUM45 atau BLOSUM62 dengan julat nilai 1-2. Ini kerana jika dilihat secara purata markah *RZ-score* bagi kedua siri ini hampir serupa.

Jika dilihat dari sudut peratusan kesamaan yang digunakan, matriks yang mempunyai jarak evolusi yang tinggi (BLOSUM45) adalah lebih efektif bagi menjajarkan jujukan yang bersaiz panjang dan mempunyai peratusaan kesamaan yang tinggi iaitu >35%. Kesimpulan berdasarkan hasil ini adalah, data yang mempunyai saiz jujukan panjang memerlukan darjah pencapahan yang besar atau jarak evolusi yang besar. Ianya menyokong teori Henikoff [19, 20] yang mementingkan jarak hubungan sekaligus menyangkal andaian dari model Dayhoff yang menyatakan kadar evolusi adalah seragam bagi keseluruhan jujukan protein [10].

6.5 Ringkasan

Secara keseluruhannya bab ini menerangkan proses olahan hasil, analisa keputusan dan perbincangan terhadap parameter skema permarkahan yang digunakan dalam pengaturcaraan dinamik Smith-Waterman iaitu kombinasi parameter matriks penggantian BLOSUM dan jurang penalti linear. Hasil yang diperolehi dari fasa perlaksanaan akan diolah untuk mendapatkan keputusan akhir, ianya terdiri dari empat langkah utama iaitu penjumlahan jadual hasil larian mengikut kategori data, pernormalan hasil menggunakan *Z-score*, pengabungan hasil *Z-score* menggunakan *RZ-score* dan pembinaan graf. Berdasarkan kepada keputusan akhir iaitu jadual dan graf *RZ-score* mengikut kategori data yang diperoleh, proses seterusnya adalah analisa dan perbincangan keputusan. Proses analisa keputusan dan perbincangan

terhadap parameter matriks penggantian BLOSUM dan jurang penalti linear diperincikan mengikut kategori data jujukan kajian yang digunakan, iaitu dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti. Berdasarkan hasil analisa ini, suatu panduan pemilihan kombinasi parameter matriks penggantian dan jurang penalti linear yang efektif bagi pengaturcaraan dinamik Smith-Waterman berjaya dihasilkan.

BAB 7

ANALISA KEPUTUSAN DAN PERBINCANGAN TERHADAP PARAMETER MATRIKS PENGGANTIAN BLOSUM DAN JURANG PENALTI AFFINE DALAM PENGATURCARAAN DINAMIK

7.1 Pendahuluan

Secara umumnya bab ini akan membincangkan mengenai proses-proses yang terlibat bagi mengolah hasil larian skema permarkahan matriks penggantian BLOSUM dan jurang penalti affine yang diperolehi dari fasa perlaksanaan untuk mendapatkan keputusan akhir yang lebih baik. Seterusnya adalah analisa keputusan dan perbincangan terhadap hasil keputusan yang diperoleh mengikut kategori data kajian dengan kombinasi parameter skema permarkahan yang digunakan.

7.2 Proses Olahan Hasil Larian

Bagi memperolehi hasil, 27 set data jujukan protein akan dijajarkan menggunakan kombinasi parameter matriks penggantian dan nilai jurang penalti

affine yang berbeza. Seperti yang telah dijelaskan dalam bab 5, 180 pengulangan perlu dilakukan bagi melaksanakan satu set pasangan. Hasil jajaran direkodkan dalam jadual, satu jadual dibina bagi setiap pasangan jajaran seperti contoh rajah di bawah.

Kategori Data : Pendek <25%ID (P02775 128bp & P04981 133bp)

Data	Matriks	Output	-e k-	1	2	3	4	5	6	7	8	9	10	11	12
		Markah		229	185	151	120	98	84	73	64	58	54	50	47
		Pjg Jajaran	1	168	156	155	148	114	113	113	112	111	111	100	100
		Padanan		26.79%	26.92%	27.10%	26.35%	26.32%	23.89%	23.89%	23.21%	18.92%	18.92%	19.00%	19.00%
		Markah		218	171	136	108	87	72	60	49	46	43	40	40
		Pjg Jajaran	2	159	152	150	126	113	112	112	100	100	100	21	21
		Padanan		26.42%	26.97%	27.33%	27.78%	28.32%	25.00%	25.00%	19.00%	19.00%	19.00%	33.33%	33.33%
		Markah		213	165	126	102	81	62	50	43	41	40	40	40
Pendek		Pjg Jajaran	3	158	152	127	126	113	112	102	53	53	21	21	21
<25%ID	BLOSUM45	Padanan		25.95%	26.97%	28.35%	27.78%	28.32%	25.89%	24.51%	20.75%	20.75%	33.33%	33.33%	33.33%
		Markah		209	160	122	96	75	56	46	42	40	40	40	40
		Pjg Jajaran	4	158	151	127	126	113	102	61	53	21	21	21	21
		Padanan		24.68%	25.83%	28.35%	27.78%	28.32%	26.47%	22.95%	20.75%	33.33%	33.33%	33.33%	33.33%
		Markah		208	158	119	92	69	52	45	42	40	40	40	40
		Pjg Jajaran	5	158	151	127	126	113	102	53	53	21	21	21	21
		Padanan		24.68%	25.17%	28.35%	27.78%	28.32%	26.47%	20.75%	20.75%	33.33%	33.33%	33.33%	33.33%
		Markah		181	140	105	82	66	57	53	50	47	44	41	38
		Pjg Jajaran	1	162	154	151	124	113	102	100	100	100	100	100	15
		Padanan		25.93%	26.62%	20.53%	25.00%	25.66%	21.57%	19.00%	19.00%	19.00%	19.00%	19.00%	40.00%
		Markah		172	126	95	73	57	47	43	40	38	38	38	38
		Pjg Jajaran	2	158	148	114	104	103	100	100	100	15	15	15	15
		Padanan		25.95%	27.03%	28.07%	26.92%	26.21%	20.00%	20.00%	19.00%	40.00%	40.00%	40.00%	40.00%
		Markah		165	121	90	69	53	39	38	38	38	38	38	38
		Pjg Jajaran	3	157	129	114	104	103	60	15	15	15	15	15	15
	BLOSUM62	Padanan		25.48%	27.91%	28.07%	26.92%	27.18%	25.00%	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%
		Markah		162	117	86	65	49	39	38	38	38	38	38	38
		Pjg Jajaran	4	157	129	114	104	103	60	15	15	15	15	15	15
		Padanan		24.84%	27.91%	28.07%	26.92%	27.18%	25.00%	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%
		Markah		161	114	82	62	45	39	38	38	38	38	38	38
		Pjg Jajaran	5	157	129	114	104	103	60	15	15	15	15	15	15
		Padanan		24.84%	27.13%	28.07%	26.92%	25.93%	25.00%	40.00%	40.00%	40.00%	40.00%	40.00%	40.00%
		Markah		180	134	94	71	57	44	33	31	30	29	29	29
		Pjg Jajaran	1	170	167	158	106	104	103	49	16	16	12	12	12
		Padanan		26.47%	26.95%	27.22%	27.36%	27.88%	27.18%	30.61%	43.75%	43.75%	41.67%	41.67%	41.67%
		Markah		163	112	81	58	40	33	32	31	30	29	29	29
		Pjg Jajaran	2	160	132	115	114	74	16	16	16	16	12	12	12
		Padanan		26.25%	28.03%	28.70%	28.95%	31.08%	43.75%	43.75%	43.75%	43.75%	41.67%	41.67%	41.67%
		Markah		157	108	75	52	36	33	32	31	30	29	29	29
		Pjg Jajaran	3	160	131	115	114	74	16	16	16	16	12	12	12
	BLOSUM80	Padanan		68.33%	27.48%	28.70%	28.95%	31.08%	43.75%	43.75%	43.75%	43.75%	41.67%	41.67%	41.67%
		Markah		154	105	71	46	36	33	32	31	30	29	29	29
		Pjg Jajaran	4	159	131	115	75	74	16	16	16	16	12	12	12
		Padanan		25.16%	27.48%	28.70%	32.00%	31.08%	43.75%	43.75%	43.75%	43.75%	41.67%	41.67%	41.67%
		Markah		152	102	68	43	34	33	32	31	30	29	29	29
		Pjg Jajaran	5	159	130	115	75	16	16	16	16	16	12	12	12
		Padanan		24.53%	26.92%	28.70%	32.00%	43.75%	43.75%	43.75%	43.75%	43.75%	41.67%	41.67%	41.67%

Rajah 7.1: Contoh jadual hasil SWAffine bagi jajaran sepasang jujukan

Hasil jajaran yang diambilkira bagi tujuan analisa dan penilaian adalah markah kesamaan optimal, panjang jajaran yang terhasil dan padanan dari jajaran yang terhasil. Jajaran yang baik boleh dikenalpasti dengan mencari markah tertinggi atau maksimum bagi setiap ciri hasil yang diperoleh. Oleh itu, proses olahan hasil jajaran perlu dilakukan bagi memudahkan proses penilaian. Huraian seterusnya merupakan proses-proses yang dilakukan bagi menghasilkan keputusan akhir.

7.2.1 Penjumlahan Jadual Hasil Larian

Proses penjumlahan hasil ini dan formula pengiraannya sama seperti yang dilaksanakan pada bab 6, iaitu dengan mencampurkan kesemua jadual dalam kategori data yang sama dan mengambil markah purata bagi markah optimal, jajaran dan padanan. Hasilnya adalah 3 jadual mengikut kategori panjang jujukan J_S, J_M dan J_L dan 9 jadual mengikut kategori panjang jujukan dan peratusan kesamaan identiti. Rajah 7.2 merupakan contoh penjumlahan hasil jadual bagi kategori data pendek J_S. Rujuk Lampiran M.

Data	Matriks	Output	-a %	1	2	3	4	5	6	7	8	9	10	11	12
Pendek	BLOSUM45	Markah	1	249	227	210	196	184	177	170	165	161	157	154	152
		PjgJajar		106	101	96	95	85	89	82	81	81	81	72	72
		Padanan		35.40%	35.18%	35.91%	36.91%	39.83%	37.24%	36.89%	36.93%	36.21%	35.78%	40.17%	57.30%
		Markah	2	242	219	202	190	179	170	164	159	155	152	149	147
		PjgJajar		120	110	102	97	91	88	85	83	81	80	77	77
		Padanan		35.42%	35.74%	35.90%	36.97%	36.86%	38.35%	37.44%	36.08%	36.08%	47.33%	40.41%	40.41%
		PjgJajar	3	238	216	198	186	175	166	159	154	151	148	146	144
		PjgJajar		99	97	86	86	83	81	80	71	66	57	55	51
		Padanan		38.57%	38.86%	39.69%	39.67%	39.41%	38.94%	37.96%	36.89%	37.00%	40.46%	41.59%	42.42%
		Markah	4	236	213	197	184	173	163	156	152	149	147	145	143
		PjgJajar		99	97	86	86	83	80	69	66	60	53	51	51
		Padanan		38.23%	38.71%	39.69%	39.55%	39.41%	39.03%	38.35%	37.27%	39.36%	41.72%	42.85%	42.85%
		Markah	5	236	212	195	182	170	161	155	151	147	145	143	141
		PjgJajar		99	97	86	86	83	80	66	66	56	53	51	51
		Padanan		38.23%	38.42%	39.57%	39.55%	39.41%	39.03%	37.55%	37.70%	40.36%	41.72%	42.85%	42.85%
	BLOSUM62	Markah	1	199	178	162	151	142	135	130	127	124	121	118	116
		PjgJajar		104	98	95	87	85	82	78	75	74	74	72	58
		Padanan		35.37%	35.49%	37.33%	37.36%	37.85%	36.89%	41.86%	37.38%	37.04%	39.04%	43.97%	41.67%
		Markah	2	193	171	156	144	135	129	125	122	119	117	115	114
		PjgJajar		100	92	85	83	82	81	75	74	60	52	50	50
		Padanan		37.16%	37.88%	39.86%	39.40%	38.69%	36.85%	37.54%	37.53%	40.59%	43.40%	43.53%	43.53%
		PjgJajar	3	190	168	152	141	132	125	122	119	117	115	114	112
		PjgJajar		100	88	85	82	81	64	57	52	52	52	50	50
		Padanan		38.19%	39.79%	39.86%	39.36%	38.45%	39.19%	41.69%	42.84%	42.84%	42.40%	43.53%	43.53%
		Markah	4	188	166	151	139	129	123	120	118	116	114	112	110
		PjgJajar		99	88	84	81	80	64	52	52	52	52	50	50
		Padanan		37.95%	40.00%	39.85%	39.65%	38.95%	39.19%	42.68%	42.84%	42.84%	42.84%	43.53%	43.53%
		Markah	5	188	165	149	138	127	122	119	117	114	112	110	109
		PjgJajar		99	88	83	81	76	64	52	52	52	52	50	50
		Padanan		38.06%	39.87%	39.85%	39.90%	39.16%	39.62%	42.68%	42.84%	42.84%	42.84%	43.96%	43.53%
	BLOSUM80	Markah	1	205	181	162	150	140	132	125	122	119	117	115	113
		PjgJajar		106	100	97	87	83	81	63	57	55	48	45	45
		Padanan		37.00%	39.25%	40.15%	40.21%	40.39%	39.71%	47.35%	49.69%	50.05%	51.98%	51.98%	51.67%
		Markah	2	196	171	155	143	132	125	122	119	117	115	113	112
		PjgJajar		102	89	84	84	77	61	57	48	46	45	45	45
		Padanan		38.16%	40.70%	41.19%	40.46%	40.91%	42.98%	49.32%	51.28%	53.27%	51.92%	51.97%	51.67%
		PjgJajar	3	192	169	152	139	130	123	120	118	115	113	112	110
		PjgJajar		101	88	84	83	76	51	49	48	46	45	45	45
		Padanan		39.64%	40.59%	41.19%	40.38%	40.86%	45.55%	51.35%	51.14%	53.27%	51.92%	51.97%	51.67%
		Markah	4	191	166	150	137	128	122	119	117	114	112	110	109
		PjgJajar		100	88	83	76	76	51	48	48	46	45	45	45
		Padanan		38.34%	40.59%	40.61%	40.88%	40.86%	45.74%	51.44%	51.37%	52.27%	51.92%	51.92%	51.67%
		Markah	5	190	165	149	136	127	122	119	116	113	111	109	107
		PjgJajar		100	87	83	76	62	51	48	48	46	45	45	45
		Padanan		38.17%	39.83%	40.61%	40.88%	43.39%	45.42%	51.44%	51.59%	52.50%	51.92%	51.92%	51.62%

Rajah 7.2 : Contoh jadual hasil SWAffine bagi data kategori pendek (J_S)

7.2.2 Pernormalan Hasil Menggunakan *Z-score*

Proses pernormalan hasil dilakukan bagi menseragamkan format untuk memudahkan analisa keputusan. Prosesnya sama seperti yang diterangkan dalam bab 6 iaitu menggunakan *Z-score*. Cuma formula bagi pengiraanya yang perlu diubahsuai mengikut jadual yang dihasilkan.

Berikut adalah formula pengiraan *Z-score*:
Simbol $\alpha\beta_{-d,-e}$ merupakan kombinasi skema permarkahan matriks BLOSUM dengan jurang penalti affine. α = B45, B62 dan B80 manakala M mewakili nilai bagi markah optimal, panjang jajaran dan padanan.

$$\mu M\alpha\beta_{-d,-e} = \frac{(M\alpha\beta_{-1,-1} + ...M\alpha\beta_{-1,-5}) + (M\alpha\beta_{-2,-1} + ...M\alpha\beta_{-2,-5}) + (M\alpha\beta_{-12,-1}...M\alpha\beta_{-12,-5})}{12 * 5}$$

$$\sigma M\alpha\beta_{-d.-e} = \sqrt{\frac{\left(M\alpha\beta_{-1,-1} - \mu M\alpha\beta_{-d,-e}\right)^2 + ... + \left(M\alpha\beta_{-12,5} - \mu M\alpha\beta_{-d,-e}\right)^2}{12 * 5}}$$

$$\text{z-score } M\alpha\beta_{-i,-j} = \frac{M\alpha\beta_{-i,-j} - \mu M\alpha\beta_{-d,-e}}{\sigma M\alpha\beta_{-d,-e}}, \begin{matrix} i = 1..12 \\ j = 1..5 \end{matrix}$$

Rajah 7.3 merupakan contoh hasil pernormalan *Z-score* ke atas setiap nilai jadual bagi kategori data pendek J_S pada Rajah 7.2. Rujuk Lampiran N untuk jadual pernormalan hasil.

Data	Matriks	Output	o k-d	1	2	3	4	5	6	7	8	9	10	11	12
		Markah		1.6019	1.5962	1.6217	1.5412	1.4553	1.4813	1.4684	1.5208	1.5071	1.5285	1.5592	1.5446
		Pjg Jajar	1	0.1480	0.1305	0.6477	0.8918	0.0549	1.1623	0.6661	0.9503	1.0269	1.1132	0.8753	0.9238
		Padanan		-1.0963	-1.2415	-1.0920	-1.1184	0.7019	-1.6628	-1.3573	-0.0732	-0.8180	-1.3655	-1.0953	1.7693
		Markah		0.3031	0.2756	0.2427	0.3692	0.4671	0.4347	0.5306	0.4545	0.4646	0.4219	0.3568	0.3882
Pendek		Pjg Jajar	2	1.6891	1.6943	1.4525	1.2784	1.7175	1.0025	1.0303	1.1526	1.0640	1.0583	1.2690	1.2522
		Padanan		-1.0863	-0.9272	-1.0982	-1.0709	-1.7686	-0.2141	-0.3525	-1.4970	-0.8856	1.4389	-0.9052	-0.6936
	BLOSUM45	Pjg Jajar		-0.3463	-0.2699	-0.3365	-0.2461	-0.1617	-0.1825	-0.3001	-0.3311	-0.3098	-0.3043	-0.3025	-0.3139
		Pjg Jajar	3	-0.5999	-0.5694	-0.6926	-0.7033	-0.5908	-0.4876	0.4212	-0.2579	-0.2660	-0.5091	-0.4688	-0.7254
		Padanan		0.8695	0.8329	0.7499	0.7856	0.3556	0.5425	0.5814	-0.1434	-0.4146	-0.2293	0.0101	-0.4007
		Markah		-0.7174	-0.7005	-0.6398	-0.6856	-0.6707	-0.6655	-0.7289	-0.7239	-0.6969	-0.6847	-0.6516	-0.6443
		Pjg Jajar	4	-0.6186	-0.6277	-0.6926	-0.7334	-0.5908	-0.8386	-0.8649	-0.9225	-0.7157	-0.8312	-0.8378	-0.7254
		Padanan		0.6565	0.7488	0.7499	0.7018	0.3556	0.6677	1.2918	0.4930	0.8027	0.0779	0.9952	-0.3375
		Markah		-0.8411	-0.9015	-0.8881	-0.9786	-1.0899	-1.0680	-0.9700	-0.9203	-0.9650	-0.9614	-0.9619	-0.9747
		Pjg Jajar	5	-0.6186	-0.6277	-0.7150	-0.7334	-0.5908	-0.8386	-1.2527	-0.9225	-1.1092	-0.8312	-0.8378	-0.7254
		Padanan		0.6565	0.5870	0.6905	0.7018	0.3556	0.6677	-0.1635	1.2206	1.2154	0.0779	0.9952	-0.3375
		Markah		1.5893	1.6320	1.5704	1.5667	1.5215	1.5723	1.5476	1.5737	1.5896	1.5490	1.4526	1.3344
		Pjg Jajar	1	1.7692	1.6530	1.7576	1.6830	1.3235	1.1656	1.2104	1.1088	1.6655	1.7889	1.7889	1.7889
		Padanan		-1.6806	-1.5996	-1.7888	-1.7475	-1.5183	-1.0723	0.2656	-1.1209	-1.6524	-1.7516	1.1046	-1.7889
		Markah		0.3268	0.2114	0.3475	0.3348	0.3966	0.3726	0.3959	0.3378	0.2573	0.3078	0.4304	0.5600
		Pjg Jajar	2	-0.2054	0.2488	-0.2549	0.1012	0.2814	1.0225	0.9455	1.0820	0.2158	-0.4472	-0.4472	-0.4472
		Padanan		-0.1599	-0.3726	0.4509	0.2631	0.1364	-1.0989	-1.7475	-1.0696	-0.2503	0.7430	-0.7303	0.4472
	BLOSUM62	Pjg Jajar		-0.3416	-0.3171	-0.3282	-0.2653	-0.2250	-0.3852	-0.3239	-0.3213	-0.2481	-0.1887	-0.1076	-0.0357
		Pjg Jajar	3	-0.4423	-0.6339	-0.2549	-0.3417	0.1251	-0.7294	-0.4714	-0.7303	-0.6271	-0.4472	-0.4472	-0.4472
		Padanan		0.7194	0.6071	0.4509	0.2209	-0.3380	0.6180	0.1845	0.7302	0.6242	0.1712	-0.7303	0.4472
		Markah		-0.7130	-0.6475	-0.6500	-0.7075	-0.6986	-0.6693	-0.6838	-0.6509	-0.6156	-0.6355	-0.6456	-0.6314
		Pjg Jajar	4	-0.5213	-0.6339	-0.5904	-0.7213	-0.2918	-0.7294	-0.8422	-0.7303	-0.6271	-0.4472	-0.4472	-0.4472
		Padanan		0.5162	0.7157	0.4471	0.5093	0.6563	0.6180	0.6487	0.7302	0.6242	0.4187	-0.7303	0.4472
		Markah		-0.8615	-0.8787	-0.9396	-0.9286	-0.9946	-0.8903	-0.9358	-0.9393	-0.9831	-1.0327	-1.1298	-1.2271
		Pjg Jajar	5	-0.6003	-0.6339	-0.6574	-0.7213	-1.4382	-0.7294	-0.8422	-0.7303	-0.6271	-0.4472	-0.4472	-0.4472
		Padanan		0.6050	0.6494	0.4399	0.7541	1.0632	0.9363	0.6487	0.7302	0.6242	0.4187	1.0863	0.4472
		Markah		1.6562	1.6809	1.6325	1.5957	1.6464	1.6752	1.5867	1.5140	1.4290	1.3361	1.2790	1.2921
		Pjg Jajar	1	1.6746	1.7799	1.7824	1.2116	1.0938	1.6843	1.4909	1.7886	1.7889	1.7889	0.0000	0.0000
		Padanan		-1.3464	-1.4898	-1.3478	-1.1389	-0.7424	-1.6112	-1.5512	-1.7463	-1.6814	1.7874	0.9706	0.4365
		Markah		0.1793	0.0838	0.2305	0.2863	0.1787	0.1226	0.2851	0.3820	0.4714	0.5938	0.6323	0.6048
		Pjg Jajar	2	0.1647	-0.3303	-0.3610	0.5757	0.2712	0.1623	0.5843	-0.4159	-0.4472	-0.4472	0.0000	0.0000
		Padanan		-0.1084	0.8065	0.9945	-0.3507	-0.3124	-0.3460	-0.4741	0.3444	0.7564	-0.5174	0.5973	0.4365
	BLOSUM80	Pjg Jajar		-0.3745	-0.3090	-0.3614	-0.2923	-0.3318	-0.3677	-0.2727	-0.1839	-0.1179	-0.0742	-0.0144	-0.0137
		Pjg Jajar	3	-0.4530	-0.4220	-0.3610	0.2744	0.1378	-0.6113	-0.6749	-0.4575	-0.4472	-0.4472	0.0000	0.0000
		Padanan		1.4692	0.6208	0.9945	-0.6111	-0.3526	0.6448	0.6400	0.1699	0.7564	-0.4233	0.5973	0.4578
		Markah		-0.6382	-0.6493	-0.6418	-0.6882	-0.6190	-0.6129	-0.6446	-0.6792	-0.6335	-0.6680	-0.6610	-0.6323
		Pjg Jajar	4	-0.6589	-0.4220	-0.5302	-1.0309	0.1378	-0.6113	-0.7001	-0.4575	-0.4472	-0.4472	0.0000	0.0000
		Padanan		0.0842	0.6208	-0.3205	1.0503	-0.3526	0.7163	0.6927	0.4673	-0.0010	-0.4233	-1.0826	0.4578
		Markah		-0.8228	-0.8064	-0.8599	-0.9014	-0.8742	-0.8172	-0.9545	-1.0329	-1.1491	-1.1876	-1.2358	-1.2509
		Pjg Jajar	5	-0.7275	-0.6055	-0.5302	-1.0309	-1.6407	-0.6240	-0.7001	-0.4575	-0.4472	-0.4472	0.0000	0.0000
		Padanan		-0.0986	-0.5782	-0.3205	1.0503	1.7610	0.5961	0.6927	0.7647	0.1696	-0.4233	-1.0826	-1.7888

Rajah 7.3: Contoh jadual SWAffine dengan *Z-score*

7.2.3 Pengabungan *Z-score* Menggunakan *RZ-score*

Proses ini diperlukan bagi menghasilkan satu ukuran yang mewakili nilai maksima bagi markah optimal, panjang jajaran dan padanan yang diperolehi selepas penormalan *Z-score*. Proses ini dilakukan dengan menggunakan formula *RZ-score* yang dinyatakan dalam Bab 6.2.3 bagi mengira jadual mengikut kategori panjang jujukan dan peratusan identiti. Hasil pengiraan *RZ-score* ini direkodkan dalam jadual, seperti contoh Rajah 7.4 iaitu jadual *RZ-score* bagi kategori data pendek. Rujuk Lampiran O dan P.

Data	Matriks	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Pendek		1	0.22	0.16	0.39	0.44	0.74	0.33	0.26	0.80	0.57	0.43	0.45	1.41
		2	0.30	0.35	0.20	0.19	0.14	0.41	0.40	0.04	0.21	0.97	0.24	0.32
	BLOSUM45	3	-0.03	0.00	-0.09	-0.05	-0.13	-0.04	0.23	-0.24	-0.33	-0.35	-0.25	-0.48
		4	-0.23	-0.19	-0.19	-0.24	-0.30	-0.28	-0.10	-0.38	-0.20	-0.48	-0.16	-0.57
		5	-0.27	-0.31	-0.30	-0.34	-0.44	-0.41	-0.80	-0.21	-0.25	-0.57	-0.27	-0.68
		1	0.56	0.56	0.51	0.50	0.44	0.55	1.01	0.52	0.53	0.53	1.45	0.44
		2	-0.01	0.03	0.18	0.23	0.27	0.10	-0.14	0.12	0.07	0.20	-0.25	0.19
	BLOSUM62	3	-0.02	-0.11	-0.04	-0.13	-0.15	-0.17	-0.20	-0.11	-0.08	-0.15	-0.43	-0.01
		4	-0.24	-0.19	-0.26	-0.31	-0.11	-0.26	-0.29	-0.22	-0.20	-0.22	-0.61	-0.21
		5	-0.29	-0.29	-0.39	-0.30	-0.46	-0.23	-0.38	-0.31	-0.33	-0.35	-0.16	-0.41
		1	0.66	0.66	0.69	0.56	0.67	0.58	0.51	0.52	0.51	1.64	0.75	0.58
		2	0.08	0.19	0.29	0.17	0.05	-0.02	0.13	0.10	0.26	-0.12	0.41	0.35
	BLOSUM80	3	0.21	-0.03	0.09	-0.21	-0.18	-0.11	-0.10	-0.16	0.06	-0.31	0.19	0.15
		4	-0.40	-0.15	-0.50	-0.22	-0.28	-0.17	-0.22	-0.22	-0.36	-0.51	-0.58	-0.06
		5	-0.55	-0.66	-0.57	-0.29	-0.25	-0.28	-0.32	-0.24	-0.48	-0.69	-0.77	-1.01

Rajah 7.4: Contoh jadual SWAffine dengan *RZ-score*

7.2.4 Menjana Graf

Bagi memudahkan proses penilaian terhadap hasil keputusan *RZ-score*, proses menjana graf berdasarkan jadual keputusan merupakan langkah terakhir yang perlu dilakukan. Ianya bertujuan untuk membuat analisa perbandingan keberkesanan parameter skema permarkahan sekaligus dapat mencari kombinasi parameter skema permarkahan yang terbaik.

7.3 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM dan Jurang Penalti Affine

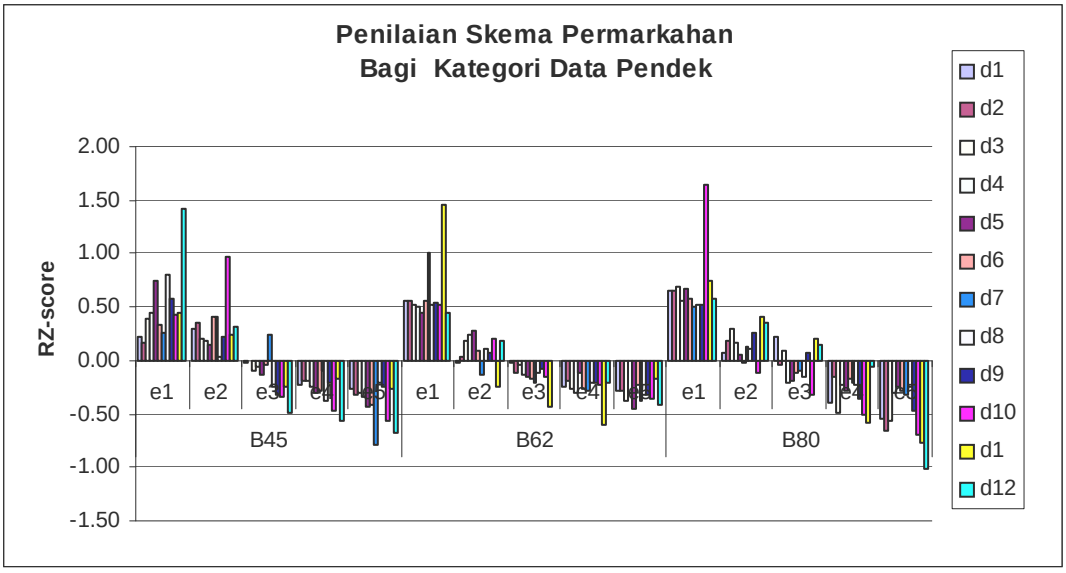
Hasil ujikaji penjajaran jujukan yang dilakukan bertujuan untuk menganalisa keberkesanan kombinasi matriks penggantian BLOSUM dan fungsi jurang penalti affine yang digunakan dalam skema permarkahan pengaturcaraan dinamik Smith-Waterman. Berikut merupakan perincian hasil ujikaji penjajaran mengikut kategori data jujukan yang berbeza.

7.3.1 Hasil Ujikaji Penjajaran Jujukan Bagi Kategori Data Jujukan Pendek

Hasil keputusan penjajaran jujukan pendek yang digunakan terdiri dari koleksi set jujukan protein pendek yang mempunyai peratusan kesamaan identiti yng berbeza iaitu <25%, 20-40% dan >35%. Oleh itu analisa hasil keputusan *RZ-score* terhadap parameter skema permarkahan yang terdiri dari matriks penggantian BLOSUM dan jurang penalti affine, dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti.

7.3.1.1 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM

Berdasarkan graf pada Rajah 7.5 dan Lampiran O, nilai *RZ-score* yang tertinggi atau berkedudukan pertama adalah dengan menggunakan siri BLOSUM80 dengan nilai –d =10 dan –e =1 bagi jurang penalti affine. Manakala jika dilihat dari kategori data pendek dengan peratusan kesamaan identiti yang berbeza seperti dalam Lampiran P.



Rajah 7.5: Hasil ujikaji SWAffine bagi kategori data pendek

Kesimpulan yang boleh dibuat adalah seperti berikut:

(i) Kategori pendek <25% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM80 dengan nilai –d =7 dan –e =1.

(ii) Kategori pendek 20%-40% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM80 dengan nilai –d =7 dan –e =1.

(iii) Kategori pendek >35% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM62 dengan nilai –d =5 dan –e =1.

Ini menunjukkan peratusan kesamaan mendatangkan kesan terhadap hasil jajaran dan ianya berperanan menentukan pemilihan parameter siri BLOSUM bagi menghasilkan jajaran yang lebih berkesan. Secara umumnya bagi data berjujukan pendek, disarankan matriks penggantian BLOSUM80. Ini kerana hasil yang diperolehi secara purata mengikut kategori panjang jujukan hampir sama dengan hasil yang diperolehi mengikut sub kategori peratusan identiti. Ini kerana secara purata nilai *RZ-score* tertinggi bagi data pendek ialah dengan menggunakan BLOSUM80 dan jurang penalti affine (-d,-e) adalah (10, 1) seperti Rajah 7.5. Selain itu BLOSUM80 turut memperolehi markah tertinggi iaitu berkedudukan pertama dengan menggunakan sub kategori peratusan identiti <25% dan 20%-40%.

7.3.1.2 Analisa Keputusan Terhadap Parameter Jurang Penalti Affine

Bagi jurang penalti affine terdapat dua nilai parameter yang mesti ditentukan iaitu –d nilai pembukaan jurang dan –e nilai penambahan jurang. Graf pada Rajah 7.5 dan jadual di lampiran O dan P dirujuk dalam menganalisa jurang penalti affine iaitu -d dan –e bagi kategori data pendek. Nilai ini digunakan bersama kombinasi siri matriks penggantian BLOSUM. Berdasarkan Lampiran O, markah *RZ-score* didapati semakin meningkat apabila nilai (-d >-e) bagi kategori data pendek. Bagi semua nilai

–d dalam julat 1 hingga 12, nilai –e yang paling efektif adalah antara julat 1 hingga 2. Hasil yang sama juga diperolehi walaupun menggunakan kategori data pendek dengan peratusan identiti yang berbeza.

Berikut merupakan kombinasi parameter $\alpha\beta_{-d,-e}$ yang efektif bagi kategori data berjujukan pendek:

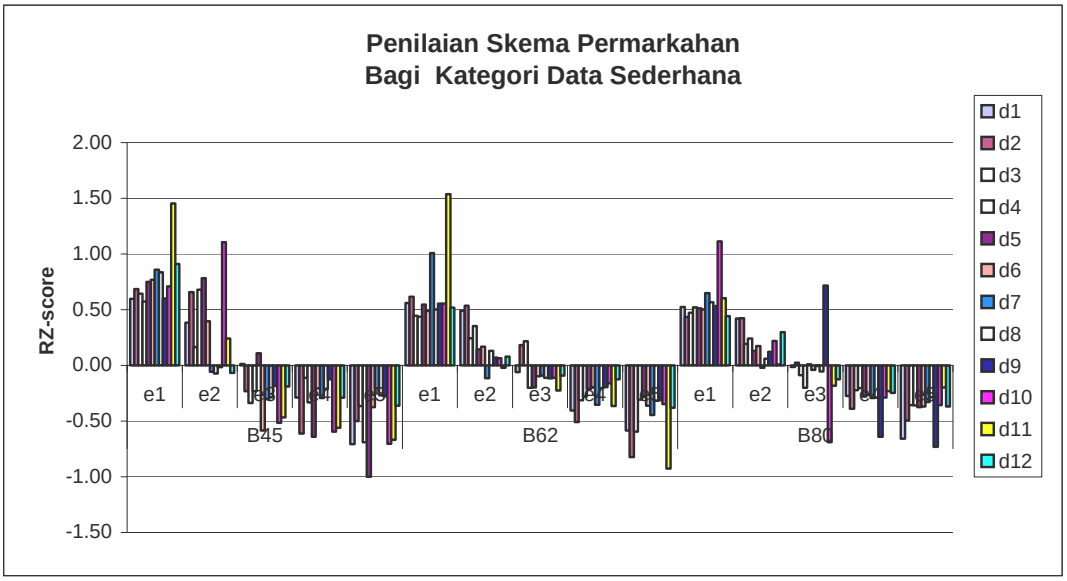
Matriks	Kedudukan Pertama		Kedudukan Kedua	
	(-d,-e)	*RZ-score*	(-d,-e)	*RZ-score*
BLOSUM45	(12,1)	1.41	(10,2)	0.97
BLOSUM62	(11,1)	1.45	(7,1)	1.01
BLOSUM80	(10,1)	1.64	(11,1)	0.75

7.3.2 Hasil Ujikaji Penjajaran Bagi Kategori Data Sederhana

Hasil keputusan penjajaran jujukan sederhana yang digunakan terdiri dari koleksi set jujukan protein sederhana yang mempunyai peratusan kesamaan identiti yng berbeza iaitu <25%, 20-40% dan >35%. Oleh itu analisa hasil keputusan *RZ-score* terhadap parameter skema permarkahan yang terdiri dari matriks penggantian BLOSUM dan jurang penalti affine, dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti.

7.3.2.1 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM

Berdasarkan graf pada Rajah 7.6 dan Lampiran O, nilai *RZ-score* yang tertinggi dan berkedudukan pertama adalah dengan menggunakan parameter siri BLOSUM62 dengan nilai –d =11 dan –e =1 bagi jurang penalti affine. Manakala jika dilihat dari kategori data sederhana dengan peratusan kesamaan identiti yang berbeza seperti dalam Lampiran P.



Rajah 7.6: Hasil ujikaji SWAffine bagi kategori data sederhana

Kesimpulan yang boleh dibuat adalah seperti berikut:

(i) Kategori sederhana <25% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM62 dengan nilai –d =11 dan –e =1.

(ii) Kategori sederhana 20%-40% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM62 dengan nilai –d =7 dan –e =1.

(iii) Kategori sederhana >35% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM62 dengan nilai –d =11 dan –e =1.

Ini menunjukkan peratusan kesamaan mendatangkan kesan terhadap hasil jajaran dan ianya berperanan menentukan pemilihan parameter siri BLOSUM bagi menghasilkan jajaran yang lebih berkesan. Secara umumnya bagi data berjujukan sederhana, disarankan matriks penggantian BLOSUM62. Ini kerana hasil yang diperolehi secara purata mengikut kategori panjang jujukan hampir sama dengan hasil yang diperolehi mengikut sub kategori peratusan identiti. Ini kerana secara purata nilai *RZ-score* tertinggi bagi data sederhana ialah dengan menggunakan BLOSUM62 dan jurang penalti affine (-d,-e) adalah (11, 1) seperti Rajah 7.6. Selain itu BLOSUM62 turut memperolehi markah tertinggi atau berkedudukan pertama dengan menggunakan sub kategori peratusan identiti <25%, 20%-40% dan >35%.

7.3.2.2 Analisa Keputusan Terhadap Parameter Jurang Penalti Affine

Bagi jurang penalti affine terdapat dua nilai parameter yang mesti ditentukan iaitu –d nilai pembukaan jurang dan –e nilai penambahan jurang. Graf pada Rajah 7.6 dan jadual di lampiran O dan P dirujuk dalam menilai jurang penalti affine iaitu -d dan –e bagi kategori data sederhana. Nilai ini digunakan bersama kombinasi siri matriks penggantian BLOSUM. Berdasarkan Lampiran O, markah *RZ-score* didapati semakin meningkat apabila nilai (-d >-e) bagi kategori data sederhana. Bagi semua nilai –d dalam julat 1 hingga 12, nilai –e yang paling efektif adalah antara julat 1 hingga 2. Hasil yang sama juga diperolehi walaupun menggunakan kategori data sederhana dengan peratusan identiti yang berbeza.

Berikut merupakan kombinasi parameter $\alpha\beta_{-d,-e}$ yang efektif bagi kategori data berjujukan sederhana:

Matriks	Kedudukan Pertama		Kedudukan Kedua	
	(-d,-e)	*RZ-score*	(-d,-e)	*RZ-score*
BLOSUM45	(11,1)	1.45	(10,2)	1.11
BLOSUM62	(11,1)	1.54	(7,1)	1.01
BLOSUM80	(10,1)	1.12	(7,1)	0.65

7.3.3 Hasil Ujikaji Penjajaran Bagi Kategori Data Panjang

Hasil keputusan penjajaran jujukan panjang yang digunakan terdiri dari koleksi set jujukan protein sederhana yang mempunyai peratusan kesamaan identiti yng berbeza iaitu <25%, 20-40% dan >35%. Oleh itu analisa hasil keputusan *RZ-score* terhadap parameter skema permarkahan yang terdiri dari matriks penggantian BLOSUM dan jurang penalti affine, dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti.

7.3.3.1 Analisa Keputusan Terhadap Parameter Matriks Penggantian BLOSUM

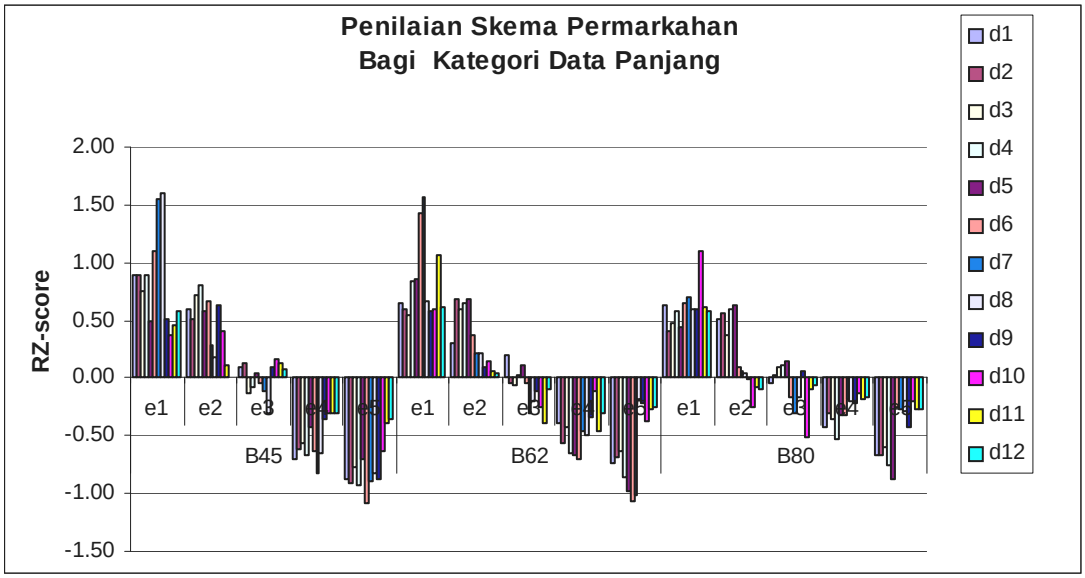
Berdasarkan graf pada Rajah 7.7 dan Lampiran O, nilai *RZ-score* yang tertinggi dan berkedudukan pertama adalah dengan menggunakan parameter siri BLOSUM45 dengan nilai –d =8 dan –e =1 bagi jurang penalti affine. Manakala kedudukan kedua tertinggi adalah dengan menggunakan siri BLOSUM62 dengan nilai –d =11 dan –e =1 bagi jurang penalti affine. Jika dilihat dari kategori data panjang dengan peratusan kesamaan identiti yang berbeza seperti dalam Lampiran P.

Kesimpulan yang boleh dibuat adalah seperti berikut:

(i) Kategori panjang <25% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM45 dengan nilai –d =8 dan –e =1.

(ii) Kategori panjang 20%-40% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM62 dengan nilai –d =7 dan –e =1.

(iii) Kategori panjang >35% identiti, nilai *RZ-score* yang tertinggi adalah dengan menggunakan siri BLOSUM62 dengan nilai –d =7 dan –e =1.



Rajah 7.7: Hasil ujikaji SWAffine bagi kategori data panjang

Ini menunjukkan peratusan kesamaan mendatangkan kesan terhadap hasil jajaran dan ianya berperanan menentukan pemilihan parameter siri BLOSUM bagi menghasilkan jajaran yang lebih berkesan. Secara umumnya bagi data berjujukan panjang, disarankan dua matriks penggantian iaitu BLOSUM45 dan BLOSUM62. Ini kerana secara purata nilai *RZ-score* tertinggi bagi data panjang ialah dengan menggunakan BLOSUM45 dan jurang penalti affine (-d,-e) adalah (8, 1) diikuti BLOSUM62 yang mempunyai markah kedua tertinggi dengan jurang penalti affine (11,1) seperti Rajah 7.7. BLOSUM62 turut memperolehi markah tertinggi atau berkedudukan pertama dengan menggunakan sub kategori peratusan identiti 20%-40% dan >35%.

7.3.3.2 Analisa Keputusan Terhadap Parameter Jurang Penalti Affine

Bagi jurang penalti affine terdapat dua nilai parameter yang mesti ditentukan iaitu –d nilai pembukaan jurang dan –e nilai penambahan jurang. Graf pada Rajah 7.7 dan jadual di lampiran O dan P dirujuk dalam menilai jurang penalti affine iaitu

-d dan –e bagi kategori data panjang. Nilai ini digunakan bersama kombinasi siri matriks penggantian BLOSUM. Berdasarkan Lampiran O, markah *RZ-score* didapati semakin meningkat apabila nilai (-d >-e) bagi kategori data panjang. Bagi semua nilai –d dalam julat 1 hingga 12, nilai –e yang paling efektif di adalah antara julat 1 hingga 2. Hasil yang sama juga diperolehi walaupun menggunakan kategori data panjang dengan peratusan identiti yang berbeza.

Berikut merupakan kombinasi parameter $\alpha\beta_{-d,-e}$ yang efektif bagi kategori data berjujukan panjang:

Matriks	Kedudukan Pertama		Kedudukan Kedua	
	(-d,-e)	*RZ-score*	(-d,-e)	*RZ-score*
BLOSUM45	(8,1)	1.60	(7,1)	1.54
BLOSUM62	(7,1)	1.57	(11,1)	1.44
BLOSUM80	(10,1)	1.10	(7,1)	0.71

7.4 Perbincangan

Hasil yang diperoleh dari kajian ini dapat membuktikan secara empirikal keberkesanan parameter skema permarkahan yang digunakan bagi menghasilkan jajaran yang optimal. Berdasarkan jadual analisa hasil pada Rajah 7.8 dan Lampiran O dan P didapati nilai penalti penambahan iaitu (–e) lebih kecil dari nilai penalti pembukaan iaitu (-d), adalah lebih efektif bagi menghasilkan markah *RZ-score* yang tinggi. Jika diperhatikan sekiranya julat nilai –e yang besar diguna iaitu (-e>-d) didapati perubahan markah *RZ-score* semakin berkurang dan bernilai negatif. Keadaan ini berjaya membuktikan teori bahawa fungsi jurang penalti affine yang menggalakkan penambahan jurang berbanding pengenalan jurang baru [3, 17]. Berdasarkan hasil keputusan yang diperoleh mengikut kategori data pendek, sederhana, panjang dan peratusan kesamaan, didapati nilai jurang penalti affine agak

sukar ditentukan dengan tepat. Rumusan yang boleh dibuat adalah nilai –e yang efektif adalah dalam julat 1 hingga 2 bagi sebarang padanan nilai -d. Manakala nilai –d ditentukan mengikut kategori data dan matriks BLOSUM yang digunakan.

Ciri Jujukan		**Nilai Parameter Cadangan**			
Kategori Data	**Peratusan Identiti**	**Kedudukan Pertama**		**Kedudukan Kedua**	
		(-d, -e)	**BLOSUM**	**(-d, -e)**	**BLOSUM**
Pendek	Secara Purata	(10,1)	BLOSUM80	(11,1)	BLOSUM62
	<25%	(7,1)	BLOSUM80	-	-
	20%-40%	(7,1)	BLOSUM80	(10,1)	BLOSUM80
	>35%	(5,1)	BLOSUM62	-	-
Nilai parameter saranan bagi kategori data pendek		BLOSUM80 dengan jurang penalti (7,1) dan (10,1)			
Sederhana	Secara purata	(11,1)	BLOSUM62	(7,1)	BLOSUM62
	<25%	(11,1)	BLOSUM62	(12,1)	BLOSUM62
	20%-40%	(7,1)	BLOSUM62	(11,1)	BLOSUM62
	>35%	(11,1)	BLOSUM62	(7,1)	BLOSUM62
Nilai parameter saranan kategori data sederhana		BLOSUM62 dengan jurang penalti (7,1), (11,1) dan (12,1)			
Panjang	Secara purata	(8,1)	BLOSUM45	(7,1)	BLOSUM62
	<25%	(8,1)	BLOSUM45	(7,1)	BLOSUM62
	20%-40%	(7,1)	BLOSUM62	(7,1)	BLOSUM45
	>35%	(7,1)	BLOSUM62	(7,1)	BLOSUM45
Nilai parameter saranan kategori data panjang		BLOSUM 45 dengan jurang penalti (8,1) dan (7,1) dan BLOSUM62 dengan jurang penalti (7,1)			

Rajah 7.8 : Analisa hasil keputusan bagi SW $\alpha\delta_{-d,-e}$

Bagi keputusan parameter siri matriks penggantian BLOSUM pula, faktor panjang jujukan dan peratusan kesamaan memainkan peranan dalam menentukan pemilihan matriks penggantian BLOSUM. Analisa dilakukan mengikut kategori data jujukan satu persatu, bagi data pendek cadangan siri BLOSUM yang disyorkan adalah BLOSUM80. Ini kerana jika dilihat secara purata markah *RZ-score* bagi siri ini dengan nilai jurang penalti (10,1) adalah yang terbaik. Berpandukan hasil dari peratusan kesamaan, didapati BLOSUM80 berkedudukan pertama dan kedua dengan jurang penalti (7,1) dan (10,1) bagi peratusan kesamaan 20%-40%. Manakala bagi kategori data berjujukan sederhana terbukti penggunaan BLOSUM62 lebih efektif berbanding siri BLOSUM lain dengan nilai jurang penalti (7,1), (11,1) dan (12,1). Ini kerana walaupun diuji menggunakan data yang mempunyai peratusan identiti yang berbeza namun BLOSUM62 tetap yang terbaik.

Ciri Jujukan	**Nilai Parameter Cadangan**				
	Matriks	**Kedudukan Pertama**		**Kedudukan Kedua**	
		(-d,-e)	***RZ-score***	**(-d,-e)**	***RZ-score***
Pendek	BLOSUM45	(12,1)	1.41	(10,2)	0.97
	BLOSUM62	(11,1)	1.45	(7,1)	1.01
	BLOSUM80	(10,1)	1.64	(11,1)	0.75
Sederhana	BLOSUM45	(11,1)	1.45	(10,2)	1.11
	BLOSUM62	(11,1)	1.54	(7,1)	1.01
	BLOSUM80	(10,1)	1.12	(7,1)	0.65
Panjang	BLOSUM45	(8,1)	1.60	(7,1)	1.54
	BLOSUM62	(7,1)	1.57	(11,1)	1.44
	BLOSUM80	(10,1)	1.10	(7,1)	0.71
Keseluruhan	**BLOSUM45**	**(12,1), (10,2), (11,1), (8,1) dan (7,1)**			
	BLOSUM62	**(11,1) dan (7,1)**			
	BLOSUM80	**(10,1), (11,1) dan (7,1)**			

Rajah 7.9 : Analisa hasil parameter nilai jurang penalti terhadap matriks BLOSUM

Bagi data kategori panjang pula, disyorkan menggunakan BLOSUM45 dengan julat (7,1) dan (8,1) atau BLOSUM62 dengan julat (7,1). Ini kerana secara purata nilai *RZ-score* tertinggi bagi data panjang ialah dengan menggunakan BLOSUM62 yang merupakan markah kedua tertinggi dengan jurang penalti affine (11,1). BLOSUM62 turut memperolehi markah tertinggi atau berkedudukan pertama dengan menggunakan sub kategori peratusan identiti 20%-40% dan >35%. Kesimpulannya, faktor panjang jujukan memainkan peranan dalam menentukan pemilihan matriks penggantian BLOSUM. Terbukti bahawa matriks yang mempunyai jarak evolusi yang rendah lebih efektif digunakan untuk menjajarkan jujukan yang pendek [19].

Secara umumnya berdasarkan analisa terhadap matriks BLOSUM pada Rajah 7.9, julat nilai (-d,-e) yang disarankan bagi BLOSUM45 adalah (7,1), (8,1), (10,2), (11,1) dan (12,1). Bagi BLOSUM62 adalah (11,1) dan (7,1) manakala BLOSUM80 adalah (7,1), (10,1) dan (11,1). Jika dilihat dari sudut peratusan kesamaan yang digunakan, bagi jujukan panjang dengan peratusan kesamaan rendah markah *RZ-score* bagi BLOSUM45 adalah tinggi berbanding matriks lain berbeza pula bagi jujukan pendek dengan peratusan kesamaan tinggi markah *RZ-score* bagi BLOSUM80 adalah tinggi. Kesimpulannya, matriks yang mempunyai jarak evolusi yang rendah (BLOSUM80) adalah lebih efektif bagi menjajarkan jujukan yang pendek dan mempunyai peratusaan kesamaan yang rendah iaitu <25%. Manakala matriks yang mempunyai jarak evolusi yang tinggi (BLOSUM45) adalah lebih efektif untuk menjajarkan jujukan panjang dan mempunyai peratusan tinggi.

Hasil dari kajian empirikal ini, berjaya membuktikan teori kepentingan jarak hubungan *(distant relationship)* bagi menghasilkan jajaran optimal [1, 2, 19, 20]. Rumusannya berdasarkan hasil ini adalah, data yang mempunyai saiz jujukan panjang memerlukan darjah pencapahan yang besar atau jarak evolusi yang besar. Ianya menyokong teori Henikoff [19, 20] yang mementingkan jarak hubungan sekaligus menyangkal andaian dari model Dayhoff yang menyatakan kadar evolusi adalah seragam bagi keseluruhan jujukan protein [10].

7.5 Ringkasan

Secara keseluruhannya bab ini menerangkan proses olahan hasil, analisa keputusan dan perbincangan terhadap parameter skema permarkahan yang digunakan dalam pengaturcaraan dinamik Smith-Waterman iaitu kombinasi parameter matriks penggantian BLOSUM dan jurang penalti affine. Hasil yang diperolehi dari fasa perlaksanaan akan diolah untuk mendapatkan keputusan akhir, ianya terdiri dari empat langkah utama iaitu penjumlahan jadual hasil larian mengikut kategori data, pernormalan hasil menggunakan *Z-score*, pengabungan hasil *Z-score* menggunakan *RZ-score* dan pembinaan graf. Berdasarkan kepada keputusan akhir iaitu jadual dan graf *RZ-score* mengikut kategori data yang diperoleh, proses seterusnya adalah analisa dan perbincangan keputusan. Proses analisa keputusan dan perbincangan terhadap parameter matriks penggantian BLOSUM dan jurang penalti affine diperincikan mengikut kategori data jujukan kajian yang digunakan, iaitu dinilai dari sudut panjang jujukan dan peratusan kesamaan identiti. Berdasarkan hasil analisa ini, suatu panduan pemilihan kombinasi parameter matriks penggantian dan jurang penalti affine yang efektif bagi pengaturcaraan dinamik Smith-Waterman berjaya dihasilkan.

BAB 8

KESIMPULAN DAN KERJA MASA HADAPAN

8.1 Pendahuluan

Penjajaran jujukan merupakan perbandingan dan penyusunan dua atau lebih input bagi jujukan, sama ada untuk mengira kesamaan di antara jujukan tersebut atau untuk mencari jujukan induk yang mana setiap input bagi jujukan berkongsi kriterianya. Penjajaran jujukan digunakan untuk membandingkan jujukan dengan tujuan untuk mendapatkan struktur, fungsi dan hubungan evolusi yang wujud bagi jujukan yang dikaji. Matlamat penjajaran jujukan adalah untuk memadankan jujukan dengan memaksimakan padanan yang sama dan meminimakan padanan yang tidak sama atau dalam erti kata lain untuk mendapatkan jajaran yang optimal. Pengaturcaraan dinamik Smith-Waterman merupakan kaedah yang selalu digunakan untuk mendapatkan jajaran jujukan setempat yang optima menggunakan skema permarkahan [8]. Matriks penggantian dan jurang penalti dimasukkan dalam skema permarkahan pengaturcaraan dinamik yang asal bertujuan untuk melihat keberkesanannya terhadap penjajaran jujukan setempat. Algorithma pengaturcaraan dinamik dikodkan, dataset ditentukan dan hasil larian aturcara direkodkan bagi tujuan analisa dan perbandingan. Ianya bertujuan bagi menganalisa keberkesanan kombinasi parameter skema permarkahan yang digunakan bagi menentukan dan menghasilkan panduan pemilihan kombinasi parameter yang efektif.

8.2 Kesimpulan

Hasil yang diperoleh dari kajian ini dapat membuktikan secara empirikal keberkesanan skema permarkahan yang digunakan bagi menghasilkan jajaran yang optimal. Oleh kerana kajian ini melakukan ujikaji ke atas dua skema permarkahan yang berbeza maka, kesimpulan diperincikan mengikut hasil bagi setiap skema permarkahan.

8.2.1 Kesimpulan Terhadap Parameter Matriks Penggantian BLOSUM dan Jurang Penalti Linear Dalam Skema Permarkahan Pengaturcaraan Dinamik

Hasil yang diperoleh dari kajian yang menggunakan skema permarkahan matriks penggantian BLOSUM dan jurang penalti linear dapat membuktikan secara empirikal keberkesanan skema permarkahan yang digunakan bagi menghasilkan jajaran yang optimal. Elakkan mengguna jurang penalti linear –d yang terlalu besar kerana, sekiranya julat nilai –d yang besar diguna kemungkinan jajaran yang terhasil terlalu pendek yang menyebabkan padanan yang terhasil terlalu sedikit dan tidak memberi makna. Oleh itu julat nilai jurang penalti linear yang dicadangkan adalah 1 dan 2 bagi semua siri BLOSUM. Ini membuktikan teori bagi jurang penalti, bahawa hasil jajaran boleh diperbaiki dengan hanya memperkenalkan jurang [6, 33].

Bagi keputusan siri matriks penggantian BLOSUM pula, faktor panjang jujukan dan peratusan kesamaan memainkan peranan dalam menentukan pemilihan matriks penggantian BLOSUM. Analisa dilakukan mengikut kategori data jujukan satu persatu, dan disyorkan penggunaan BLOSUM62 bagi kategori data sederhana. Manakala BLOSUM45 dan BLOSUM62 sesuai digunakan untuk data berjujukan panjang. Bagi data pendek tiada cadangan siri BLOSUM dapat ditentukan. Ini kerana

ketidakseragaman pilihan matriks penggantian BLOSUM yang efektif, kemungkinan disebabkan saiz jujukan yang terlalu pendek atau ketidakserasian data menggunakan skema permarkahan tersebut.

Jika dilihat dari sudut peratusan kesamaan yang digunakan, matriks yang mempunyai matriks yang mempunyai jarak evolusi yang tinggi (BLOSUM45) adalah lebih efektif bagi menjajarkan jujukan yang panjang dan mempunyai peratusaan kesamaan yang tinggi iaitu >35%. Ini bermakna data yang mempunyai saiz jujukan panjang memerlukan darjah pencapahan yang besar atau jarak evolusi yang besar. Keputusan kajian ini didapati menyokong teori Henikoff [19,20] yang mementingkan jarak hubungan sekaligus menyangkal andaian dari model Dayhoff yang menyatakan kadar evolusi adalah seragam bagi keseluruhan jujukan protein [10].

8.2.2 Kesimpulan Terhadap Parameter Matriks Penggantian BLOSUM dan Jurang Penalti Affine Dalam Skema Permarkahan Pengaturcaraan Dinamik

Berdasarkan hasil yang diperoleh dari kajian yang menggunakan skema permarkahan matriks penggantian BLOSUM dan jurang penalti affine. Penentuan jurang penalti affine dan matriks penggantian yang efektif dapat dilakukan secara empirikal. Bagi jurang penalti affine didapati nilai penalti penambahan iaitu (–d) lebih besar dari nilai penalti pembukaan iaitu (-e), adalah lebih efektif bagi menghasilkan jajaran yang optimal berdasarkan markah *RZ-score* yang tinggi. Jika diperhatikan sekiranya julat nilai –e yang besar diguna iaitu (-e>-d) didapati perubahan markah *RZ-score* semakin berkurang dan bernilai negatif. Keadaan ini berjaya membuktikan teori bahawa fungsi jurang penalti affine yang menggalakkan menggalakkan penambahan jurang berbanding pengenalan jurang baru [3, 17]. Berdasarkan hasil kombinasi parameter yang diperoleh mengikut kategori data iaitu pendek, sederhana, panjang dan peratusan kesamaan, didapati nilai jurang penalti

affine agak sukar ditentukan dengan tepat. Rumusan yang boleh dibuat adalah nilai –e yang efektif adalah dalam julat 1-2 bagi sebarang padanan nilai -d. Manakala nilai –d ditentukan mengikut kategori data dan matriks BLOSUM yang digunakan.

Bagi keputusan siri matriks penggantian BLOSUM pula, faktor panjang jujukan dan peratusan kesamaan memainkan peranan dalam menentukan pemilihan matriks penggantian BLOSUM. Analisa dilakukan mengikut kategori data jujukan satu persatu, bagi data pendek cadangan siri BLOSUM yang disyorkan adalah BLOSUM80 dengan nilai jurang penalti (7,1) dan (10,1). Manakala bagi kategori data berjujukan sederhana terbukti penggunaan BLOSUM62 lebih efektif berbanding siri BLOSUM lain dengan nilai jurang penalti (7,1), (11,1) dan (12,1). Bagi data kategori panjang pula, disyorkan menggunakan BLOSUM45 dengan julat (7,1) dan (8,1) atau BLOSUM62 dengan julat (7,1). Kesimpulannya, faktor panjang jujukan memainkan peranan dalam menentukan pemilihan matriks penggantian BLOSUM. Terbukti bahawa matriks yang mempunyai jarak evolusi yang rendah lebih efektif digunakan untuk menjajarkan jujukan yang pendek [19].

Secara umumnya berdasarkan analisa terhadap matriks BLOSUM, julat nilai (-d,-e) yang disarankan bagi BLOSUM45 adalah (7,1), (8,1), (10,2), (11,1) dan (12,1). Bagi BLOSUM62 adalah (11,1) dan (7,1) manakala BLOSUM80 adalah (7,1), (10,1) dan (11,1). Bagi jarak hubungan yang jauh dan jujukan yang panjang, BLOSUM62 dengan nilai penalti (7,1) disyorkan. Secara kebetulan nilai (11,1) bagi BLOSUM62 dan (10,1) adalah sama dengan julat yang disarankan oleh [3] dan nilai (7,1) adalah nilai yang digunakan oleh [28] untuk mencari pangkalan data jujukan protein.

Jika dilihat dari sudut peratusan kesamaan yang digunakan, bagi jujukan panjang dengan peratusan kesamaan rendah markah *RZ-score* bagi BLOSUM45 adalah tinggi berbanding matriks lain berbeza pula bagi jujukan pendek dengan peratusan kesamaan tinggi markah *RZ-score* bagi BLOSUM80 adalah tinggi. Kesimpulannya, matriks yang mempunyai jarak evolusi yang rendah (BLOSUM80)

adalah lebih efektif bagi menjajarkan jujukan yang pendek dan mempunyai peratusaan kesamaan yang rendah iaitu <25%. Manakala matriks yang mempunyai jarak evolusi yang tinggi (BLOSUM45) adalah lebih efektif untuk menjajarkan jujukan panjang dan mempunyai peratusan tinggi. Hasil dari kajian empirikal ini, berjaya membuktikan teori kepentingan jarak hubungan *(distant relationship)* bagi menghasilkan jajaran optimal [1, 2, 19, 20]. Bagi data yang mempunyai saiz jujukan panjang memerlukan darjah pencapahan yang besar atau jarak evolusi yang besar. Ianya menyokong teori Henikoff [19, 20] yang mementingkan jarak hubungan sekaligus menyangkal andaian dari model Dayhoff yang menyatakan kadar evolusi adalah seragam bagi keseluruhan jujukan protein [10].

8.2.3 Kesimpulan Hasil Ujikaji di antara SW$\alpha\beta_{-d}$ dan SW$\alpha\delta_{-d,-e}$

Lampiran K dan O digunakan bagi tujuan pembandingan dua skema permarkahan SW$\alpha\beta_{-d}$ dan SW$\alpha\delta_{-d,-e}$. Ianya dilakukan dengan membandingkan *RZ-score* yang tertinggi (maksimum) mengikut kategori data dan penggunaan matriks BLOSUM. Nilai *RZ-score* mewakili hasil jajaran, semakin tinggi nilai ini semakin optimal lah hasil jajarannya. Jika dibandingkan didapati nilai maksimum *RZ-score* bagi skema permarkahan SW$\alpha\delta_{-d,-e}$ adalah lebih baik dari nilai maksimum *RZ-score* bagi skema permarkahan SW$\alpha\beta_{-d}$ seperti yang ditunjukkan dalam Rajah 8.1. Oleh itu secara kesimpulannya skema permarkahan SW$\alpha\delta_{-d,-e}$ lebih efektif berbanding skema permarkahan SW$\alpha\beta_{-d}$. Ini kerana secara teori fungsi jurang penalti affine menggalakkan penambahan jurang dalam jajaran berbanding wujudnya jurang baru [17]. Berbeza pula dengan jurang penalti linear yang mengumpukkan nilai yang tetap per jurang. Sekiranya wujud jurang kecil yang banyak maka sudah tentulah padanan jajaran yang terhasil akan berkurang.

Kategori Data	Matriks	SW $\alpha\beta_{-d}$	SW $\alpha\delta_{-d,-e}$.
Pendek	BLOSUM45	0.93	1.41
	BLOSUM62	0.83	1.45
	BLOSUM80	0.91	1.64
Sederhana	BLOSUM45	1.44	1.45
	BLOSUM62	1.49	1.54
	BLOSUM80	1.02	1.12
Panjang	BLOSUM45	1.48	1.60
	BLOSUM62	1.56	1.57
	BLOSUM80	0.94	1.10

Rajah 8.1 : Perbandingan di antara SW $\alpha\beta_{-d}$ dan SW $\alpha\delta_{-d,-e}$

8.3 Sumbangan

Berikut merupakan sumbangan dari projek ini:

(i) Panduan pemilihan parameter siri matriks penggantian BLOSUM yang efektif mengikut kategori data yang digunakan iaitu panjang jujukan dan peratusan kesamaan.

(ii) Panduan pemilihan parameter julat nilai yang efektif bagi $-d$ bagi jurang penalti linear dan julat nilai yang efektif bagi $-d$ dan $-e$ bagi jurang penalti affine.

(iii) Panduan pemilihan kombinasi parameter matriks penggantian dan jurang penalti yang efektif mengikut kategori panjang jujukan dan peratusan kesamaan identiti.

Panduan pemilihan skema permarkahan memudahkan proses pemilihan dilakukan, ianya boleh digunakan bagi penjajaran jujukan atau pencarian jujukan dalam pangkalan data.

8.4 Kerja Masa Hadapan

Berdasarkan hasil dari penyelidikan ini, sebahagian perkara memerlukan lanjutan kajian dan berikut merupakan cadangan kerja masa hadapan :

(i) Menggunakan panduan pemilihan skema permarkahan yang disarankan untuk mengkaji hasil jajaran menggunakan kategori rujukan data yang berlainan, contohnya kategori rujukan dalam BAliBASE.

(ii) Menggunakan panduan pemilihan skema permarkahan bagi menjajarkan jujukan banyak pasangan.

(iii) Mengubahsuai algorithma pengaturcaraan dinamik dalam projek ini untuk digunakan bagi menjajarkan struktur bagi data protein.

(iv) Perbandingan di antara jajaran jujukan dan jajaran struktur.

8.5 Penutup

Perancangan projek dan analisa keperluan bukan satu tugas yang mudah, sebaliknya memerlukan penelitian terhadap permasalahan yang dihadapi. Fasa 1 melibatkan perancangan dan analisa masalah hingga proses penyedian data kajian,

manakala proses pembangunan dan proses seterusnya dilaksanakan di dalam fasa dua. Oleh itu metodologi yang ditakrifkan akan menjadi rujukan bagi kesinambungan proses-proses seterusnya sehingga dapat menghasilkan suatu panduan pemilihan parameter skema permarkahan yang efektif dalam pengaturcaraan dinamik Smith-Waterman.

SENARAI RUJUKAN

1. Altschul, S.F. Amino acid substitution matrices from an information theoretic perspective. *J Mol Biol,* 219:555–65; 1991

2. Altschul, S.F. A protein alignment scoring system sensitive at all evolutionary distances. *J Mol Evol,* 36:290–300; 1993

3. Altschul, S.F. and Gish, W. Local alignment statistics. *Meth. Enzymol,* 266:460-480; 1996

4. Altschul S.F., Madden T.L.and Schaffer, A.A. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs.*Nucl. Acids Res,* 25:3389-3402; 1997

5. Altscul S.F.Generalize affine gap costs for protein sequence Alignment. *Protein.Strc.Func.Gen,* 32:88-96; 1998

6. Andreas, D.B. BIOINFORMATICS: A Pratical Guide to the Analysis of Genes and Protein. Wiley Interscience 187-212; 2001

7. Bairoch, A. and Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucl. Acids Res.* 26:38-42; 1998

8. Bellgard M. and Gamble T. Gap mapping: a paradigm for aligning two sequences. Applied Bioinformatics, 2(3)S31-S35; 2003

9. Bergeron, B. Bioinformatics Computing. Prentice Hall, 302-339; 2003

10. Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure*, 5:345-352; 1978

11. Durbin, R., Eddy, S., Krogh, A. and Mitchison, G. Biological Sequence Analysis, Probabilistic Models of Proteins and Nucleic Acids.Cambridge University Press, 1998

12. Edgar, S.J., Holiday, J.D. and Willett, P. Effectiveness of retrieval in similarity searches of chemical databases: A review of performance measures. *Journal of Molecular Graphics and Modelling*, 18:343-357; 2000

13. Feitelson, D.G. and Treinin, M. The Blueprint for Life, IEEE Computer, July 2002

14. Gunner, O. and Henry, F. Formula for Determination the Goodness of Hit Lists in 3D Database Searches, http://www.netsci.org/Science/Cheminform/feature09.html 5/6/04 1998

15. Gusfield, D. Algorithms on Strings, Trees and Sequences. Cambridge University Press, 1997

16. Gibbs, A. J. and McIntyre, G.A. The diagram method for comparing sequences its use with amino acid and nucleotide sequences. *Eur. J. Biochem*, 16:1-11; 1970

17. Gotoh, O. An improved algorithm for matching biological sequences. *J. Mol. Biol*, 162:705-708; 1982

18. Henikoff, S. and Henikoff, J.G. Automated assembly of protein blocks for database searching. *Nucl. Acids Res*, 19:6565-6572; 1991

19. Henikoff, S. and Henikoff, J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA*, 89:10915-10919; 1992

20. Henikoff, S. and Henikoff, J.G. Performance evaluation of amino acid substitution matrices.*Proteins. Struct. Funct. Gene,* 17:49-61;1993

21. Heath, L.S. and Ramakrishnan N. The Emerging Landscape of Bioinformatics Software Systems. IEEE Computer; 2002

22. Hillis, D.M., Allard, M.W. and Miyamoto, M.M. Analysis of DNA sequence data: phylogenetic inference. *Methods in Enzymology,* 224:456-487; 1993

23. Karlin, S. and Altschul, S.F. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. U.S.A,* 87: 2264-2268;1990

24. Leung, M.Y., Blaisdell, B.E., Burge, C. and Karlin, S. An efficient algorithm for identifying matches with errors in multiple long molecular sequences. *J. Mol. Biol*, 221:1367-1378; 1991

25. Mott, R. Local sequence alignments with monotonic gap penalties.*Bioinformatics,* 15(6):455-462; 1999

26. Michael S.W. and Temple, F. S. New Stratigraphic Correlation Techniques, *J.Geo,* 88:451-457; 1980

27. Needleman, S.B. and Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequences of two proteins *J. Mol.Biol*, 48:443-453; 1970

28. Pearson W.R. Comparison of methods for searching protein sequence databases. *Protein Sci,* 4:1145–60; 1995

29. Pearson W.R. Identifying distantly related protein sequences. *Computer Applications in the Biosciences*, 13(4):325-332; 1997

30. Pearson W.R. Training for bioinformatics and computational biology. *Bioinformatics*, 17(9):761-762; 2001

31. Reese, J.T and Pearson, W.R. Empirical determination of effective gap penalties for sequence comparison. *Bioinformatics*, 18:1500-1507; 2002

32. Karp, R. and Rabin, M. Efficient Randomized Pattern Matching algorithms.*IBM Journal of Research and Development*, 31:249-260; 1987

33. Sankoff, D. Matching sequences under deletion-insertion constraints. *Proc. Natl. Acad. Sci. U.S.A.* 69: 4-6; 1972

34. Sankoff, D. and Kruskal, J. Time Warps, String Edits, and Macromolecules:The Theory and Practice of Sequence Comparison. Addison-Wesley, 1987

35. Schwartz, R.M.and Dayhoff, M.O. Matrices for detecting distant relationships. In: *Atlas of Protein Sequence and Structure*. 5:353-358; 1976

36. Schlosshauer M and Ohlsson M. A novel approach to local reliability of sequence alignments. *Bioinformatics*, 18(6):847-54; 2002

37. Sellers, P.H. Pattern recognition in genetic sequences by mismatch density. *Bull. Math. Bio*, 46:501-514; 1984

38. Smith, T.F. and Waterman, M.S. Identification of common molecular subsequences.
J. Mol. Biol. 147: 195-197; 1981

39. Setubal, J.C. and Meidanis, J. Introduction to Computational Molecular Biology. PWS Publishing Company, 1997

40. Tatusov, R.L., Altschul, S.F.and Koonin, E.V. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc. Natl. Acad. Sci. U.S.A*, 91:12091-12095; 1994

41. Thompson, J., Plewniak, F. and Poch, O. BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs. *Bioinformatics*, 15: 87–88; 1999

42. Thompson, J., Plewniak, F. and Poch, O. A Comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Research,* 27:13-15; 1999

43. Waterman, M.S. Estimating statistical of sequence alignments. *Phil. Trans .R. Soc. Lond,* 344:383-390; 1994

44. Waterman, M.S. Introduction to Computational Biology. Chapman and Hall Press; 1995

45. Zhang, Z., Pearson, W.R and Miller, W. Aligning a DNA sequence with protein sequence. RECOMB 337-343; 1995

46. GENBANK.Growth of GenBank, http://www.ncbi.mlm.nih.gov/Genbank/genbankstats.html, 3/1/04 3.34pm

47. States, D.J., Gish, W. and Altschul, S.F. Improved sensitivity of nucleic acid database searches using application-specific scoring matrices. *Proc. Natl. Acad. Sci*, 3:66-70; 1991

48. Thomas, H.C., Charles, E., Leiserson, R.L. and Rivest, C.S. Introduction to Algorithms, McGraw-Hill, 1996

49. Alamat laman web matlab, http://eta.embl-heidelberg.de:8000/misc/mat/; 9/1/04 2.00pm

50. Alamat laman web BAliBASE, http://www. igbmc.ustrasbg.fr/BioInfo/BAliBASE/ ; 9/1/04 2.30pm

51. Reference 1 BALIBASE, http://www-igbmc.u- strasbg.fr/BioInfo/BAliBASE/align_index.html; 9/1/04 2.35pm

52. Alamat laman web Swiss-Prot, http://www.ebi.ac.uk/swissprot/ ; 9/1/04 4.21pm

LAMPIRAN A

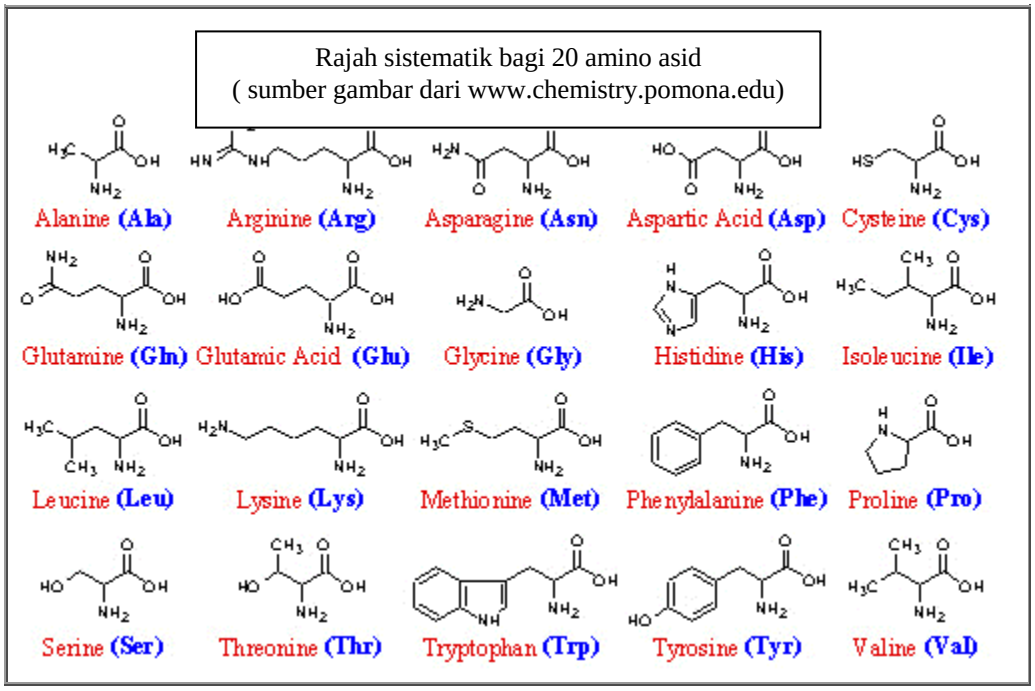
Kod Bagi Jujukan Protein *(asid amino)* dan Jujukan DNA(*nucleotides)*

Lampiran A : Kod Bagi Jujukan Protein *(asid amino)* dan Jujukan DNA(*nucleotides)*

Jujukan Protein

Mengandungi satu aksara atau tiga kod aksara yang mewakili asid amino.

Satu kod aksara	Tiga kod aksara	Nama asid amino
A	Ala	Alanine
R	Arg	Arginine
N	Asn	Asparagine
D	Asp	Aspartic asid
C	Cys	Cysteine
Q	Gln	Glutamine
E	Glu	Glutamic asid
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine
B	Asx	Aspartic asid or Asparagine
Z	Glx	Glutamic asid or Glutamine
X	Xaa	Sebarang asid amino



Jujukan Nucleotide

Nucleotide bases terdiri dari dua kategori bergantung kepada struktur cincin*(ring structure)* bagi base. Purines (Adenine and Guanine) merupakan dua *ring bases*, pyrimidines (Cytosine and Thymine) are satu *ring bases*. Mutasi di DNA mengubah satu *bases* dengan menggantikan yang lain.

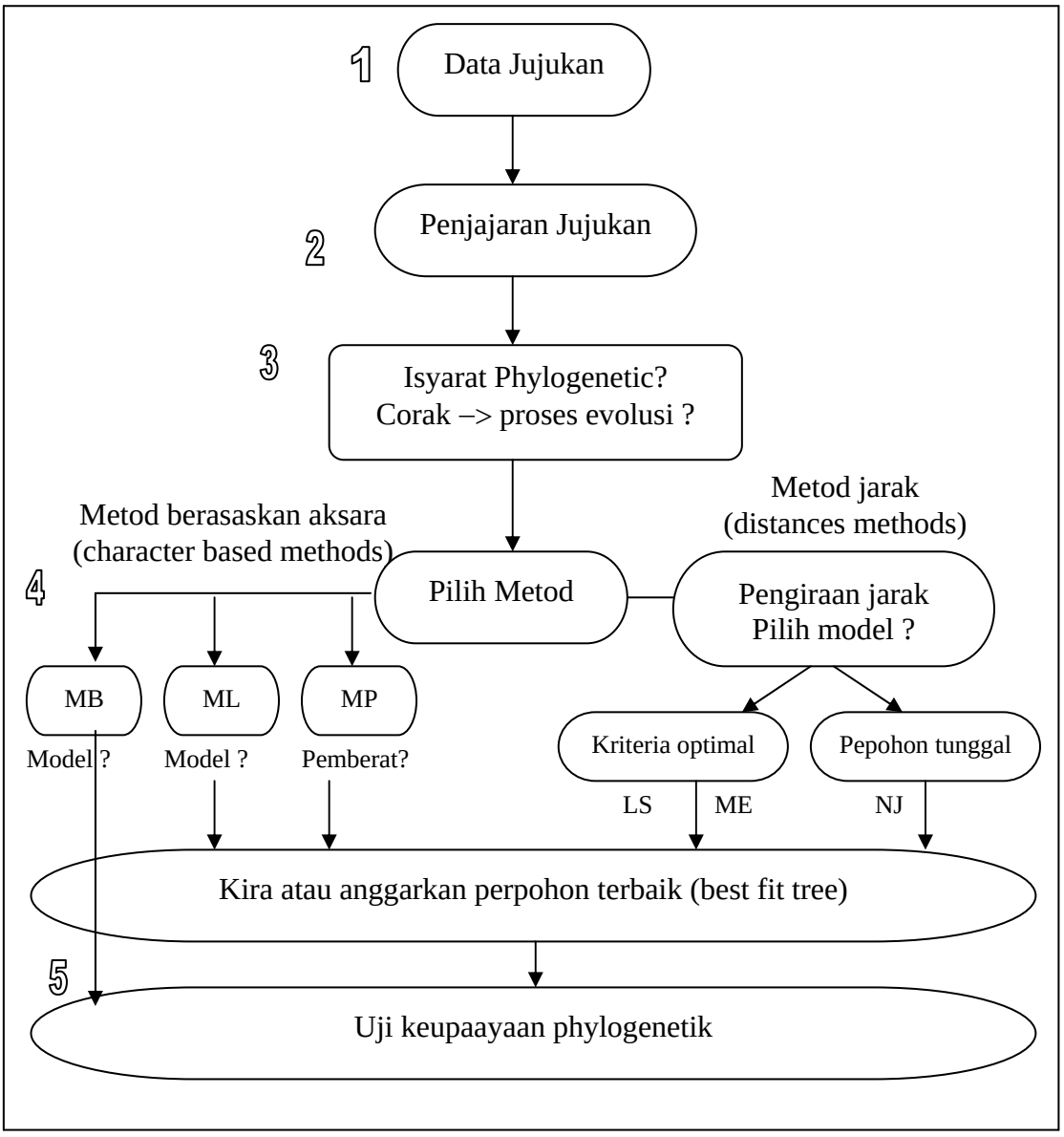
Satu kod aksara	Nama	Lokasi
A	Adenine	DNA/RNA
G	Guanine	DNA/RNA
C	Cytosine	DNA/RNA
T	Thymine	DNA
U	Uracil	RNA

LAMPIRAN B

Pembentukkan Pepohon *Phylogenetic* Dari Jujukan DNA/Protein

Lampiran B : Pembentukkan Pepohon *Phylogenetic* Dari Jujukan DNA/Protein

Rajah ini diambil dari Hillis et al., (1993). Methods in Enzymology 224, 456-487



LAMPIRAN C

Penjajaran Berpasangan *(pairwise alignment)* & Banyak Pasang *(multiple alignment)*

Lampiran C : Penjajaran Berpasangan *(pairwise alignment)* & Banyak Pasang *(multiple alignment)*

Contoh : Penjajaran Berpasangan *(pairwise alignment)*

```
HBB_HUMAN    FFESFGDLSTPDAVMGNPKVKAHGKKVL - - - - -GAFSDGLAHLDNLKGTF
HBB_HORSE    FFDSFGDLSNPGAVMGNPKVKAHGKKVL- - - - - HSFGEGVHHLDNLKGTF
```

Contoh : Penjajaran Banyak Pasang *(multiple alignment)*

```
HBB_HUMAN    FFESFGDLSTPDAVMGNPKVKAHGKKVL - - - - -GAFSDGLAHLDNLKGTF
HBB_HORSE    FFDSFGDLSNPGAVMGNPKVKAHGKKVL- - - - -HSFGEGVHHLDNLKGTF
HBA_HUMAN    YFPHF-DLS - - -  - - HGSAQVKGHGKKVA- - - - -DALTNAVAHVDDMPNAL
HBA_HORSE    YFPHF-DLS -  - - -  - HGSAQVKAHGKKVG - - - -- DALTLAVGHLDDLPGAL
MYG_PHYCA    KFDRFKHLKTEAEMKSSEDLKKHGVTVL - - - - - TALGAILKKKGHHEAEL
GLB5_PETMA   FFPKFKGLTTADQLKKSADVRWHAER I I - - - - -NAVNDAVASMDDTEKMS
```

LAMPIRAN D

Penjajaran Global dan Penjajaran Setempat

Lampiran D : Penjajaran Global dan Penjajaran Setempat

Berikut merupakan contoh penjajaran jujukan bagi dua jujukan protein iaitu Human alpha-1 hemoglobin dan plant Leghemoglobin

Penjajaran Global:

```
  1 MGAFSEKQESLVKSSWEAFKQNVPHHSAVFYTLILEKAPAAQNMFSFLSNGVDPNNPKLK   60
    |   |   :: ||::|     :   : |     :    |  :  |      :   : ::|
  1 M-VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFD--LSHGSAQVK   57

 61 AHAEKVFKMTVDSAVQLRAKGEVVLADPTLGSVHVQKGVLDP-HFLVVKEALLKTFKEAV   119
     | :||     ::   :    ::  |   |  :|  |  :|| :| ::   || |    :
 58 GHGKKVADALTNAVAHV---DDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL   114

120 GDKWNDELGNAWEVAYDELAAAIKKAMGS--A   149
          |   |   : |:  |::   : |
115 ----PAEFTPAVHASLDKFLASVSTVLTSKYR   142
```

Penjajaran Setempat:

```
  5 SEKQESLVKSSWEAFKQNVPHHSAVFYTLILEKAPAAQNMFSFLSNGVDPNNPKLKAHAE   64
    |   :: ||::|     :   : |     :    |  :  |      :   : ::| | :
  4 SPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFD--LSHGSAQVKGHGK   61

 65 KVFKMTVDSAVQLRAKGEVVLADPTLGSVHVQKGVLDP-HFLVVKEALLKT   114
    ||     ::   :    ::  |   |  :|  |  :|| :| ::   || |
 62 KVADALTNAVAHV---DDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVT   109
```

LAMPIRAN E

Matriks Penggantian BLOSUM 45

Lampiran E : Matriks Penggantian BLOSUM 45

Identiti	Nilai identiti
Positif	Nilai positif

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	5	-2	-1	-2	-1	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-2	-2	0	-1	-1	0	-5
R	-2	7	0	-1	-3	1	0	-2	0	-3	-2	3	-1	-2	-2	-1	-1	-2	-1	-2	-1	0	-1	-5
N	-1	0	6	2	-2	0	0	0	1	-2	-3	0	-2	-2	-2	1	0	-4	-2	-3	4	0	-1	-5
D	-2	-1	2	7	-3	0	2	-1	0	-4	-3	0	-3	-4	-1	0	-1	-4	-2	-3	5	1	-1	-5
C	-1	-3	-2	-3	12	-3	-3	-3	-3	-3	-2	-3	-2	-2	-4	-1	-1	-5	-3	-1	-2	-3	-2	-5
Q	-1	1	0	0	-3	6	2	-2	1	-2	-2	1	0	-4	-1	0	-1	-2	-1	-3	0	4	-1	-5
E	-1	0	0	2	-3	2	6	-2	0	-3	-2	1	-2	-3	0	0	-1	-3	-2	-3	1	4	-1	-5
G	0	-2	0	-1	-3	-2	-2	7	-2	-4	-3	-2	-2	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	-5
H	-2	0	1	0	-3	1	0	-2	10	-3	-2	-1	0	-2	-2	-1	-2	-3	2	-3	0	0	-1	-5
I	-1	-3	-2	-4	-3	-2	-3	-4	-3	5	2	-3	2	0	-2	-2	-1	-2	0	3	-3	-3	-1	-5
L	-1	-2	-3	-3	-2	-2	-2	-3	-2	2	5	-3	2	1	-3	-3	-1	-2	0	1	-3	-2	-1	-5
K	-1	3	0	0	-3	1	1	-2	-1	-3	-3	5	-1	-3	-1	-1	-1	-2	-1	-2	0	1	-1	-5
M	-1	-1	-2	-3	-2	0	-2	-2	0	2	2	-1	6	0	-2	-2	-1	-2	0	1	-2	-1	-1	-5
F	-2	-2	-2	-4	-2	-4	-3	-3	-2	0	1	-3	0	8	-3	-2	-1	1	3	0	-3	-3	-1	-5
P	-1	-2	-2	-1	-4	-1	0	-2	-2	-2	-3	-1	-2	-3	9	-1	-1	-3	-3	-3	-2	-1	-1	-5
S	1	-1	1	0	-1	0	0	0	-1	-2	-3	-1	-2	-2	-1	4	2	-4	-2	-1	0	0	0	-5
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-1	-1	2	5	-3	-1	0	0	-1	0	-5
W	-2	-2	-4	-4	-5	-2	-3	-2	-3	-2	-2	-2	-2	1	-3	-4	-3	15	3	-3	-4	-2	-2	-5
Y	-2	-1	-2	-2	-3	-1	-2	-3	2	0	0	-1	0	3	-3	-2	-1	3	8	-1	-2	-2	-1	-5
V	0	-2	-3	-3	-1	-3	-3	-3	-3	3	1	-2	1	0	-3	-1	0	-3	-1	5	-3	-3	-1	-5
B	-1	-1	4	5	-2	0	1	-1	0	-3	-3	0	-2	-3	-2	0	0	-4	-2	-3	4	2	-1	-5
Z	-1	0	0	1	-3	4	4	-2	0	-3	-2	1	-1	-3	-1	0	-1	-2	-2	-3	2	4	-1	-5
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-2	-1	-1	-1	-1	-1	-5
*****	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	1

LAMPIRAN F

Matriks Penggantian BLOSUM 62

Lampiran F : Matriks Penggantian BLOSUM 62

Identiti	Nilai identiti
Positif	Nilai positif

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	-4
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	-4
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	-4
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3	4	1	-1	-4
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	-3	-3	-2	-4
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2	0	3	-1	-4
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	-4
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3	0	0	-1	-4
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3	-3	-3	-1	-4
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	-4
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	-4
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	-4
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-3	-1	-4
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	-2	-4
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0	0	0	-4
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	-1	-1	0	-4
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3	-4	-3	-2	-4
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1	-3	-2	-1	-4
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	-3	-2	-1	-4
B	-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-3	-2	0	-1	-4	-3	-3	4	1	-1	-4
Z	-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-1	-4
*****	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	1

LAMPIRAN G

Matriks Penggantian BLOSUM 80

Lampiran G : Matriks Penggantian BLOSUM 80

Identiti	Nilai identiti
Positif	Nilai positif

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	7	-3	-3	-3	-1	-2	-2	0	-3	-3	-3	-1	-2	-4	-1	2	0	-5	-4	-1	-3	-2	-1	-8
R	-3	9	-1	-3	-6	1	-1	-4	0	-5	-4	3	-3	-5	-3	-2	-2	-5	-4	-4	-2	0	-2	-8
N	-3	-1	9	2	-5	0	-1	-1	1	-6	-6	0	-4	-6	-4	1	0	-7	-4	-5	5	-1	-2	-8
D	-3	-3	2	10	-7	-1	2	-3	-2	-7	-7	-2	-6	-6	-3	-1	-2	-8	-6	-6	6	1	-3	-8
C	-1	-6	-5	-7	13	-5	-7	-6	-7	-2	-3	-6	-3	-4	-6	-2	-2	-5	-5	-2	-6	-7	-4	-8
Q	-2	1	0	-1	-5	9	3	-4	1	-5	-4	2	-1	-5	-3	-1	-1	-4	-3	-4	-1	5	-2	-8
E	-2	-1	-1	2	-7	3	8	-4	0	-6	-6	1	-4	-6	-2	-1	-2	-6	-5	-4	1	6	-2	-8
G	0	-4	-1	-3	-6	-4	-4	9	-4	-7	-7	-3	-5	-6	-5	-1	-3	-6	-6	-6	-2	-4	-3	-8
H	-3	0	1	-2	-7	1	0	-4	12	-6	-5	-1	-4	-2	-4	-2	-3	-4	3	-5	-1	0	-2	-8
I	-3	-5	-6	-7	-2	-5	-6	-7	-6	7	2	-5	2	-1	-5	-4	-2	-5	-3	4	-6	-6	-2	-8
L	-3	-4	-6	-7	-3	-4	-6	-7	-5	2	6	-4	3	0	-5	-4	-3	-4	-2	1	-7	-5	-2	-8
K	-1	3	0	-2	-6	2	1	-3	-1	-5	-4	8	-3	-5	-2	-1	-1	-6	-4	-4	-1	1	-2	-8
M	-2	-3	-4	-6	-3	-1	-4	-5	-4	2	3	-3	9	0	-4	-3	-1	-3	-3	1	-5	-3	-2	-8
F	-4	-5	-6	-6	-4	-5	-6	-6	-2	-1	0	-5	0	10	-6	-4	-4	0	4	-2	-6	-6	-3	-8
P	-1	-3	-4	-3	-6	-3	-2	-5	-4	-5	-5	-2	-4	-6	12	-2	-3	-7	-6	-4	-4	-2	-3	-8
S	2	-2	1	-1	-2	-1	-1	-1	-2	-4	-4	-1	-3	-4	-2	7	2	-6	-3	-3	0	-1	-1	-8
T	0	-2	0	-2	-2	-1	-2	-3	-3	-2	-3	-1	-1	-4	-3	2	8	-5	-3	0	-1	-2	-1	-8
W	-5	-5	-7	-8	-5	-4	-6	-6	-4	-5	-4	-6	-3	0	-7	-6	-5	16	3	-5	-8	-5	-5	-8
Y	-4	-4	-4	-6	-5	-3	-5	-6	3	-3	-2	-4	-3	4	-6	-3	-3	3	11	-3	-5	-4	-3	-8
V	-1	-4	-5	-6	-2	-4	-4	-6	-5	4	1	-4	1	-2	-4	-3	0	-5	-3	7	-6	-4	-2	-8
B	-3	-2	5	6	-6	-1	1	-2	-1	-6	-7	-1	-5	-6	-4	0	-1	-8	-5	-6	6	0	-3	-8
Z	-2	0	-1	1	-7	5	6	-4	0	-6	-5	1	-3	-6	-2	-1	-2	-5	-4	-4	0	6	-1	-8
X	-1	-2	-2	-3	-4	-2	-2	-3	-2	-2	-2	-2	-2	-3	-3	-1	-1	-5	-3	-2	-3	-1	-2	-8
*****	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	1

LAMPIRAN H

Set Rujukan BALIBASE

Lampiran H : Set Rujukan BALIBASE

Rajah ini diambil dari Thompson. J, Plewniak.F and Poch O. “BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs.” *Bioinformatics*, 15: 87–88; 1999

A

Rujukan 1	**<100aksara**	**200<300aksara**	**>500aksara**
<25% identiti	**7**	**8**	**8**
20-40% identiti	**10**	**9**	**10**
>35% identiti	**10**	**10**	**8**

Rujukan 2	**9**	**8**	**7**

Rujukan 3	**5**	**3**	**5**

Rujukan 4	**Tambahan (Extension)**	**Masukkan (Insertion)**
	12	**12**

B

1csy - reference 1

```
Name                    SH2

Number of sequences                5
Alignment Length                   110
Longest Sequence                   104
Shortest Sequence                  100
Average Percent Identity           30
Maximum Percent Identity           38
Minimum Percent Identity           27

Sequence Name     SWISSPROT Accession
1csy                  P43405
1gri                  P29354
1aya                  P35235
2pna                  P23727
1bfi                  P27986

Family    1csy 1gri 1aya 2pna 1bfi


1csy          1  shekmpWFHGKISREESEQIVligskTNGKFLIRARD..nnGSYALCLLH
1gri          1  emkphpWFFGKIPRAKAEEML.skqrHDGAFLIRESEs.apGDFSLSVKF
1aya          1  ...mrrWFHPNITGVEAENLLltrg.VDGSFLARPSKs.npGDFTLSVRR
2pna          1  .lqdaeWYWGDISREEVNEKLrdt..ADGTFLVRDAStkmhGDYTLTLRK
1bfi          1  hhdektWNVGSSNRNKAENLLrgk..RDGTFLVRESS..kqGCYACSVVV

1csy         49  EGKVLHYRIdkdktgklsipegk.kFDTLWQLVEHYsyka......dgll
1gri         49  GNDVQHFKVlrdgagkyfl.wvv.kFNSLNELVDYHrsts.vsrnqqifl
1aya         46  NGAVTHIKIqn..tgdyydlyggekFATLAELVQYYmehhgqlkekngdv
2pna         48  GGNNKLIKIfh.rdgkygfsdpl.tFNSVVELINHYrnes.laqynpkld
1bfi         47  DGEVKHCVInktatg.ygfaepynlYSSLKELVLHYqhts.lvqhndsln

1csy         92  rvl.tVPcqk
1gri         96  rdieqVPqq.
1aya         94  iel.kYPln.
2pna         95  vkl.lYPvs.
1bfi         95  vtl.aYPvya
Key

alpha helix          RED
beta strand          GREEN
core blocks          UNDERSCORE
```

You can also look at the alignment in *RSF format*, or *MSF format* with a *Feature Table*

Back to Index

LAMPIRAN I

Jadual Hasil SWLinear Dari Proses Penjumlahan Mengikut Kategori Data

LAMPIRAN I : Jadual Hasil SWLinear Dari Proses Penjumlahan Mengikut Kategori Data

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
Pendek	BLOSUM45	Markah	257.11	226.78	206.00	191.33	178.22	167.11	160.11	154.33	149.67	145.44
		Panjang Jajaran	112	101	93	92	89	79	72	65	62	60
		Padanan	37.71%	38.38%	38.57%	38.33%	38.05%	37.57%	38.11%	38.85%	40.25%	41.16%
	BLOSUM62	Markah	207.22	177.11	158.56	144.78	133.67	126.89	122.11	117.78	114.56	111.78
		Panjang Jajaran	111	97	92	88	79	68	60	56	52	52
		Padanan	37.83%	39.49%	38.75%	38.27%	39.32%	39.47%	41.80%	41.84%	42.41%	42.41%
	BLOSUM80	Markah	213.89	177.33	155.89	141.11	131.00	124.78	120.33	116.22	112.67	110.00
		Panjang Jajaran	112	96	90	84	67	59	57	53	52	44
		Padanan	38.24%	40.19%	40.10%	39.61%	42.03%	43.25%	47.26%	47.30%	48.05%	50.30%

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
Sederhana	BLOSUM45	Markah	644.22	559.22	504.56	466.78	438.44	416.67	367.89	352.33	347.56	298.30
		Panjang Jajaran	307	278	265	255	246	239	228	215	213	212
		Padanan	35.89%	37.04%	36.23%	35.68%	35.12%	35.15%	35.18%	35.06%	34.98%	35.73%
	BLOSUM62	Markah	542.67	500.11	459.00	459.89	434.89	317.22	302.11	288.78	279.78	274.89
		Panjang Jajaran	303	274	260	239	228	223	210	208	188	184
		Padanan	35.87%	37.99%	36.92%	36.55%	36.00%	35.87%	35.61%	35.11%	35.73%	35.69%
	BLOSUM80	Markah	540.44	441.89	381.56	344.00	316.56	297.67	283.89	272.11	260.78	258.89
		Panjang Jajaran	312	280	261	235	222	206	192	184	175	172
		Padanan	35.75%	37.44%	36.91%	37.29%	37.11%	36.42%	36.57%	37.01%	37.13%	37.18%

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
Panjang	BLOSUM45	Markah	1374.44	1116.11	956.00	846.78	745.56	714.67	679.00	654.00	643.67	584.89
		Panjang Jajaran	836	746	700	630	628	615	555	455	348	287
		Padanan	33.44%	34.20%	33.53%	32.91%	32.14%	31.48%	30.43%	30.19%	30.17%	33.52%
	BLOSUM62	Markah	1587.78	1313.33	1135.44	1017.00	929.33	859.33	804.44	761.67	727.56	702.11
		Panjang Jajaran	830	743	692	635	625	615	587	527	499	421
		Padanan	33.44%	34.14%	33.52%	32.91%	32.24%	31.48%	30.43%	30.19%	30.17%	33.52%
	BLOSUM80	Markah	1316.33	1009.78	820.11	690.89	607.67	567.78	550.78	540.89	532.33	524.44
		Panjang Jajaran	847	752	654	625	476	319	245	240	227	227
		Padanan	33.02%	34.43%	34.85%	33.86%	33.87%	37.35%	39.80%	40.32%	43.91%	43.87%

LAMPIRAN J

Jadual Hasil SWLinear Dari Proses Pernormalan Mengikut Kategori Data

LAMPIRAN J : Jadual Hasil SWLinear Dari Proses Pernormalan Mengikut Kategori Data

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
Pendek	BLOSUM45	z-score Markah	2.00	1.18	0.61	0.21	-0.15	-0.45	-0.64	-0.80	-0.92	-1.04
		z-score Panjang	1.65	1.08	0.57	0.54	0.38	-0.19	-0.59	-0.97	-1.18	-1.28
		z-score Padanan	-0.87	-0.28	-0.11	-0.32	-0.57	-0.98	-0.52	0.14	1.36	2.15
	BLOSUM62	z-score Markah	2.11	1.15	0.11	0.11	-0.25	-0.47	-0.62	-0.76	-0.86	-0.95
		z-score Panjang	1.69	1.03	0.79	0.60	0.17	-0.37	-0.75	-0.94	-1.11	-1.11
		z-score Padanan	-1.32	-0.38	-0.79	-0.47	-0.47	-0.39	0.93	0.95	1.27	1.27
	BLOSUM80	z-score Markah	2.21	1.11	0.47	0.02	-0.28	-0.47	-0.60	-0.72	-0.83	-0.91
		z-score Panjang	1.78	1.10	0.83	0.56	-0.21	-0.56	-0.63	-0.81	-0.86	-1.21
		z-score Padanan	-1.27	-0.81	-0.83	-0.95	-0.38	-0.09	0.85	0.86	1.04	1.57

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
Sederhana	BLOSUM45	z-score Markah	1.91	1.12	0.61	0.25	-0.01	-0.21	-0.67	-0.82	-0.86	-1.32
		z-score Panjang	1.96	1.01	0.62	0.30	0.01	-0.22	-0.56	-0.98	-1.06	-1.09
		z-score Padanan	0.44	2.19	0.96	0.11	-0.74	-0.70	-0.64	-0.83	-0.96	0.19
	BLOSUM62	z-score Markah	1.52	1.11	0.71	0.72	0.48	-0.67	-0.81	-0.94	-1.03	-1.08
		z-score Panjang	1.87	1.11	0.76	0.19	-0.10	-0.23	-0.58	-0.62	-1.15	-1.25
		z-score Padanan	-0.32	2.26	0.96	0.51	-0.17	-0.32	-0.64	-1.25	-0.49	-0.54
	BLOSUM80	z-score Markah	2.19	1.12	0.46	0.05	-0.25	-0.46	-0.61	-0.74	-0.86	-0.88
		z-score Panjang	1.84	1.18	0.79	0.24	-0.03	-0.38	-0.67	-0.84	-1.03	-1.10
		z-score Padanan	-2.24	1.11	0.06	0.80	0.45	-0.92	-0.62	0.26	0.50	0.60

Data	Matriks	Jurang Penalti	1	2	3	4	5	6	7	8	9	10
Panjang	BLOSUM45	z-score Markah	2.17	1.14	0.50	0.06	-0.34	-0.47	-0.61	-0.71	-0.75	-0.99
		z-score Panjang	1.48	0.96	0.69	0.29	0.28	0.20	-0.14	-0.72	-1.34	-1.69
		z-score Padanan	0.80	1.30	0.86	0.46	-0.04	-0.47	-1.15	-1.30	-1.32	0.86
	BLOSUM62	z-score Markah	2.10	1.15	0.53	0.12	-0.19	-0.43	-0.62	-0.77	-0.89	-0.98
		z-score Panjang	1.78	1.05	0.62	0.15	0.06	-0.02	-0.25	-0.76	-0.99	-1.64
		z-score Padanan	0.81	1.26	0.86	0.46	0.02	-0.47	-1.16	-1.31	-1.33	0.86
	BLOSUM80	z-score Markah	2.29	1.12	0.40	-0.10	-0.41	-0.56	-0.63	-0.67	-0.70	-0.73
		z-score Panjang	1.60	1.21	0.80	0.68	0.06	-0.59	-0.90	-0.92	-0.97	-0.97
		z-score Padanan	-1.08	-0.74	-0.64	-0.88	-0.87	-0.04	0.54	0.67	1.52	1.51

LAMPIRAN K

Jadual *RZ-Score* Bagi Hasil SWLinear Mengikut Kategori Data

LAMPIRAN K : Jadual RZ-Score Bagi Hasil SWLinear Mengikut Kategori Data

Data	Matriks\ -d	1	2	3	4	5	6	7	8	9	10
Pendek	BLOSUM45	0.93	0.66	0.36	0.14	-0.11	-0.54	-0.58	-0.54	-0.25	-0.06
	BLOSUM62	0.83	0.60	0.03	0.08	-0.18	-0.41	-0.15	-0.25	-0.24	-0.27
	BLOSUM80	0.91	0.47	0.16	-0.12	-0.29	-0.37	-0.13	-0.22	-0.22	-0.18

Data	Matriks\ -d	1	2	3	4	5	6	7	8	9	10
Sederhana	BLOSUM45	1.44	1.44	0.73	0.22	-0.24	-0.38	-0.63	-0.88	-0.96	-0.74
	BLOSUM62	1.02	1.49	0.81	0.47	0.07	-0.41	-0.68	-0.94	-0.89	-0.95
	BLOSUM80	0.60	1.14	0.43	0.36	0.06	-0.59	-0.64	-0.44	-0.46	-0.46

Data	Matriks\ -d	1	2	3	4	5	6	7	8	9	10
Panjang	BLOSUM45	1.48	1.13	0.68	0.27	-0.04	-0.24	-0.63	-0.91	-1.14	-0.61
	BLOSUM62	1.56	1.15	0.67	0.24	-0.04	-0.31	-0.68	-0.95	-1.07	-0.59
	BLOSUM80	0.94	0.53	0.18	-0.10	-0.41	-0.40	-0.33	-0.31	-0.05	-0.06

LAMPIRAN L

Jadual *RZ-Score* Bagi Hasil SWLinear
Mengikut Kategori Panjang Jujukan Dengan
Peratusan Kesamaan Identiti

LAMPIRAN L : Jadual RZ-Score Bagi Hasil SWLinear Mengikut Kategori Panjang Jujukan Dengan Peratusan Kesamaan Identiti

Pendek <25%ID

Matriks\ -d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	0.97	0.40	0.04	-0.24	-0.39	-0.42	-0.08	-0.09	-0.10	-0.14
BLOSUM62	0.88	0.50	0.10	-0.19	-0.21	-0.41	-0.16	-0.16	-0.17	-0.18
BLOSUM80	1.04	0.57	0.32	0.13	-0.05	-0.57	-0.60	-0.62	-0.21	0.07

Pendek 20-40%ID

Matriks\-d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	1.30	0.93	0.24	-0.07	-0.29	-0.55	-0.29	-0.55	-0.34	-0.37
BLOSUM62	1.31	0.98	0.21	-0.01	-0.28	-0.45	-0.53	-0.36	-0.41	-0.46
BLOSUM80	1.09	0.77	0.38	0.01	-0.25	-0.34	-0.40	-0.63	-0.66	0.03

Pendek >35%ID

Matriks\ -d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	0.59	0.39	0.28	0.12	-0.01	-0.19	-0.27	-0.36	-0.22	-0.43
BLOSUM62	0.70	0.59	0.57	0.18	-0.12	-0.21	-0.30	-0.33	-0.42	-0.50
BLOSUM80	0.54	0.50	0.68	0.41	-0.19	-0.28	-0.36	-0.37	-0.46	-0.54

Sederhana <25%ID

Matriks\ -d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	0.99	0.92	-0.21	0.14	0.18	0.22	-0.51	-0.61	-0.31	-0.82
BLOSUM62	1.43	1.40	0.54	0.10	-0.34	-0.41	-0.04	-0.70	-1.03	-0.95
BLOSUM80	0.71	0.55	0.18	0.21	0.14	-0.42	-0.50	-0.49	-0.19	-0.19

Sederhana 20-40%ID

Matriks\ -d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	1.71	1.25	0.65	0.16	-0.24	-0.46	-0.53	-0.68	-0.91	-1.02
BLOSUM62	1.84	1.32	0.66	0.20	-0.24	-0.49	-0.57	-0.73	-0.91	-1.02
BLOSUM80	1.65	1.32	0.67	0.37	-0.11	-0.49	-0.82	-0.79	-0.90	-0.90

Sederhana >35%ID

Matriks\ -d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	0.60	1.03	0.53	0.14	-0.07	0.05	-0.10	-0.34	-0.81	-1.03
BLOSUM62	0.73	1.00	0.40	0.16	0.18	0.04	0.05	-0.82	-0.85	-0.90
BLOSUM80	0.53	0.75	0.35	0.02	0.05	-0.18	-0.12	-0.10	-0.56	-0.73

Panjang <25%ID

Matriks	1	2	3	4	5	6	7	8	9	10
BLOSUM45	1.49	1.05	0.66	0.17	-0.07	-0.32	-0.59	-0.93	-1.01	-0.45
BLOSUM62	1.10	0.69	0.31	-0.03	-0.35	-0.57	-0.46	-0.36	-0.16	-0.17
BLOSUM80	1.00	0.56	0.21	-0.09	-0.46	-0.44	-0.47	-0.40	0.04	0.04

Panjang 20-40%ID

Matriks\ -d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	1.56	1.16	0.46	0.36	-0.05	-0.23	-0.70	-0.82	-1.06	-0.68
BLOSUM62	1.09	0.65	0.27	0.08	-0.33	-0.39	-0.50	-0.53	-0.21	-0.14
BLOSUM80	0.91	0.48	0.12	-0.11	-0.30	-0.31	-0.18	-0.19	-0.20	-0.22

Panjang >35%ID

Matriks\ -d	1	2	3	4	5	6	7	8	9	10
BLOSUM45	1.07	0.79	0.46	0.09	-0.43	-0.54	-0.55	-0.55	-0.16	-0.18
BLOSUM62	1.53	1.18	0.82	0.31	0.06	-0.23	-0.67	-0.94	-1.03	-1.02
BLOSUM80	0.74	0.56	0.32	-0.03	-0.41	-0.43	-0.04	-0.20	-0.22	-0.28

LAMPIRAN M

Jadual Hasil SWAffine Dari Proses Penjumlahan Mengikut Kategori Data

LAMPIRAN M : Jadual Hasil SWAffine Dari Proses Penjumlahan Mengikut Kategori Data

Data	Matriks	Output	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Pendek	BLOSUM45	Markah		249	227	210	196	184	177	170	165	161	157	154	152
		Pjg Jajaran	1	106	101	96	95	85	89	82	81	81	81	72	72
		Padanan		35.40%	35.18%	35.91%	36.91%	39.83%	37.24%	36.89%	36.93%	36.21%	35.78%	40.17%	57.30%
		Markah		242	219	202	190	179	170	164	159	155	152	149	147
		Pjg Jajaran	2	120	110	102	97	91	88	85	83	81	80	77	77
		Padanan		35.42%	35.74%	35.90%	36.97%	36.86%	38.35%	37.44%	36.08%	36.08%	47.33%	40.41%	40.41%
		Pjg Jajaran		238	216	198	186	175	166	159	154	151	148	146	144
		Pjg Jajaran	3	99	97	86	86	83	81	80	71	66	57	55	51
		Padanan		38.57%	38.86%	39.69%	39.67%	39.41%	38.94%	37.96%	36.89%	37.00%	40.46%	41.59%	42.42%
		Markah		236	213	197	184	173	163	156	152	149	147	145	143
		Pjg Jajaran	4	99	97	86	86	83	80	69	66	60	53	51	51
		Padanan		38.23%	38.71%	39.69%	39.55%	39.41%	39.03%	38.35%	37.27%	39.36%	41.72%	42.85%	42.85%
		Markah		236	212	195	182	170	161	155	151	147	145	143	141
		Pjg Jajaran	5	99	97	86	86	83	80	66	66	56	53	51	51
		Padanan		38.23%	38.42%	39.57%	39.55%	39.41%	39.03%	37.55%	37.70%	40.36%	41.72%	42.85%	42.85%
	BLOSUM62	Markah		199	178	162	151	142	135	130	127	124	121	118	116
		Pjg Jajaran	1	104	98	95	87	85	82	78	75	74	74	72	58
		Padanan		35.37%	35.49%	37.33%	37.36%	37.85%	36.89%	41.86%	37.38%	37.04%	39.04%	43.97%	41.67%
		Markah		193	171	156	144	135	129	125	122	119	117	115	114
		Pjg Jajaran	2	100	92	85	83	82	81	75	74	60	52	50	50
		Padanan		37.16%	37.88%	39.86%	39.40%	38.69%	36.85%	37.54%	37.53%	40.59%	43.40%	43.53%	43.53%
		Pjg Jajaran		190	168	152	141	132	125	122	119	117	115	114	112
		Pjg Jajaran	3	100	88	85	82	81	64	57	52	52	52	50	50
		Padanan		38.19%	39.79%	39.86%	39.36%	38.45%	39.19%	41.69%	42.84%	42.84%	42.40%	43.53%	43.53%
		Markah		188	166	151	139	129	123	120	118	116	114	112	110
		Pjg Jajaran	4	99	88	84	81	80	64	52	52	52	52	50	50
		Padanan		37.95%	40.00%	39.85%	39.65%	38.95%	39.19%	42.68%	42.84%	42.84%	42.84%	43.53%	43.53%
		Markah		188	165	149	138	127	122	119	117	114	112	110	109
		Pjg Jajaran	5	99	88	83	81	76	64	52	52	52	52	50	50
		Padanan		38.06%	39.87%	39.85%	39.90%	39.16%	39.62%	42.68%	42.84%	42.84%	42.84%	43.96%	43.53%
	BLOSUM80	Markah		205	181	162	150	140	132	125	122	119	117	115	113
		Pjg Jajaran	1	106	100	97	87	83	81	63	57	55	48	45	45
		Padanan		37.00%	39.25%	40.15%	40.21%	40.39%	39.71%	47.35%	49.69%	50.05%	51.98%	51.98%	51.67%
		Markah		196	171	155	143	132	125	122	119	117	115	113	112
		Pjg Jajaran	2	102	89	84	84	77	61	57	48	46	45	45	45
		Padanan		38.16%	40.70%	41.19%	40.46%	40.91%	42.98%	49.32%	51.28%	53.27%	51.92%	51.97%	51.67%
		Pjg Jajaran		192	169	152	139	130	123	120	118	115	113	112	110
		Pjg Jajaran	3	101	88	84	83	76	51	49	48	46	45	45	45
		Padanan		39.64%	40.59%	41.19%	40.38%	40.86%	45.55%	51.35%	51.14%	53.27%	51.92%	51.97%	51.67%
		Markah		191	166	150	137	128	122	119	117	114	112	110	109
		Pjg Jajaran	4	100	88	83	76	76	51	48	48	46	45	45	45
		Padanan		38.34%	40.59%	40.61%	40.88%	40.86%	45.74%	51.44%	51.37%	52.27%	51.92%	51.92%	51.67%
		Markah		190	165	149	136	127	122	119	116	113	111	109	107
		Pjg Jajaran	5	100	87	83	76	62	51	48	48	46	45	45	45
		Padanan		38.17%	39.83%	40.61%	40.88%	43.39%	45.42%	51.44%	51.59%	52.50%	51.92%	51.92%	51.62%

Data	Matriks	Output	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Sederhana	BLOSUM45	Markah		612	555	516	487	464	444	428	412	400	391	384	376
		Pjg Jajaran	1	292	280	269	264	260	258	257	255	251	228	228	227
		Padanan		35.12%	36.20%	36.25%	35.27%	35.13%	34.70%	34.63%	34.40%	34.15%	34.89%	40.37%	34.36%
		Markah		588	531	492	466	444	426	411	398	387	378	371	364
		Pjg Jajaran	2	278	268	260	254	251	249	244	243	229	228	227	227
		Padanan		36.71%	37.05%	36.52%	36.09%	35.45%	34.85%	34.29%	33.86%	34.50%	38.84%	33.67%	33.49%
		Pjg Jajaran		576	518	479	452	431	413	399	387	377	369	362	355
		Pjg Jajaran	3	277	264	256	252	249	245	236	231	218	216	216	215
		Padanan		36.57%	36.55%	36.37%	35.53%	35.28%	34.58%	34.59%	34.42%	35.05%	34.84%	34.72%	34.50%
		Markah		569	510	471	444	423	407	393	382	373	364	357	350
		Pjg Jajaran	4	276	263	248	243	237	236	231	219	217	216	216	215
		Padanan		36.31%	36.37%	37.05%	35.87%	35.19%	34.96%	34.95%	35.23%	35.34%	35.09%	35.03%	34.54%
		Markah		566	507	468	441	419	404	390	379	369	360	353	346
		Pjg Jajaran	5	273	259	248	242	237	236	220	219	217	216	215	215
		Padanan		35.82%	36.69%	36.86%	36.62%	35.06%	34.92%	35.58%	35.29%	35.28%	35.09%	35.16%	34.59%
	BLOSUM62	Markah		500	445	408	382	360	342	329	317	309	301	294	288
		Pjg Jajaran	1	288	274	265	259	258	234	232	229	229	228	228	227
		Padanan		35.06%	35.76%	35.33%	35.60%	34.88%	35.19%	35.98%	34.03%	33.73%	33.55%	35.96%	33.41%
		Markah		479	424	388	364	344	330	316	305	296	288	282	276
		Pjg Jajaran	2	277	265	255	249	235	233	232	228	227	227	209	192
		Padanan		36.34%	37.04%	36.14%	35.94%	35.62%	35.26%	32.10%	34.24%	33.67%	33.41%	35.92%	35.78%
		Pjg Jajaran		468	413	377	352	334	320	309	298	290	283	277	271
		Pjg Jajaran	3	272	261	253	238	233	219	217	210	210	209	193	192
		Padanan		36.18%	37.13%	36.59%	35.92%	35.53%	35.92%	35.90%	35.68%	35.47%	34.96%	35.94%	35.78%
		Markah		462	406	370	346	329	316	305	295	286	279	273	267
		Pjg Jajaran	4	272	258	243	229	219	213	211	210	210	209	193	192
		Padanan		35.85%	36.34%	36.52%	36.09%	36.17%	36.23%	36.17%	35.77%	35.62%	35.30%	35.94%	36.21%
		Markah		459	403	367	344	326	313	301	291	283	275	269	263
		Pjg Jajaran	5	270	255	242	228	219	212	211	210	209	198	192	192
		Padanan		35.80%	36.20%	36.20%	36.12%	36.17%	36.13%	36.25%	35.77%	35.66%	35.99%	35.92%	35.62%
	BLOSUM80	Markah		518	454	410	377	352	330	313	299	288	277	269	262
		Pjg Jajaran	1	290	284	274	270	263	256	230	226	225	224	206	198
		Padanan		35.12%	35.89%	35.21%	35.20%	35.99%	35.79%	36.37%	36.86%	36.38%	37.92%	37.13%	35.89%
		Markah		491	424	383	352	331	313	297	285	275	267	261	255
		Pjg Jajaran	2	290	272	266	242	237	229	224	207	200	188	188	182
		Padanan		36.01%	37.45%	36.13%	36.93%	36.80%	36.74%	36.26%	37.14%	37.81%	37.97%	37.33%	37.75%
		Pjg Jajaran		477	407	367	339	319	302	290	279	276	262	256	251
		Pjg Jajaran	3	280	262	255	238	220	207	205	201	200	188	168	162
		Padanan		36.29%	37.62%	36.83%	36.48%	37.49%	37.44%	37.17%	37.31%	39.99%	37.33%	37.84%	38.29%
		Markah		471	399	359	332	313	297	285	274	265	259	254	249
		Pjg Jajaran	4	280	255	247	221	217	206	204	200	168	168	168	162
		Padanan		36.05%	37.42%	37.36%	37.58%	37.26%	37.26%	37.02%	37.23%	37.85%	38.07%	37.92%	38.29%
		Markah		467	397	356	328	309	294	281	271	263	257	252	246
		Pjg Jajaran	5	279	254	247	220	209	206	204	181	168	168	168	158
		Padanan		35.65%	37.32%	37.14%	37.36%	37.45%	37.22%	37.09%	37.66%	37.72%	38.07%	38.07%	38.55%

Data	Matriks	Output	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Panjang	BLOSUM45	Markah		1672	1526	1419	1336	1276	1232	1198	1168	1143	1123	1105	1089
		Pjg Jajaran	1	756	738	712	656	631	613	610	607	601	600	589	580
		Padanan		35.44%	35.86%	35.23%	36.12%	33.88%	35.65%	36.29%	37.93%	33.82%	33.32%	33.30%	33.14%
		Markah		1589	1440	1336	1262	1209	1170	1138	1111	1088	1069	1053	1038
		Pjg Jajaran	2	724	686	670	619	600	583	580	578	568	560	553	550
		Padanan		38.20%	36.87%	36.89%	37.01%	36.22%	36.11%	35.36%	34.88%	34.33%	33.90%	33.53%	33.18%
		Pjg Jajaran		1544	1392	1286	1216	1167	1130	1099	1073	1052	1034	1017	1003
		Pjg Jajaran	3	703	672	650	593	576	572	567	560	549	544	542	526
		Padanan		38.13%	36.83%	36.12%	36.47%	36.20%	35.70%	35.35%	34.78%	34.26%	33.99%	33.77%	33.63%
		Markah		1521	1368	1260	1188	1138	1101	1071	1046	1025	1005	991	978
		Pjg Jajaran	4	696	659	643	577	570	565	551	550	546	522	501	501
		Padanan		35.12%	35.94%	35.66%	36.07%	35.50%	35.33%	34.76%	34.60%	34.12%	33.91%	33.85%	33.66%
		Markah		1508	1355	1246	1173	1119	1081	1051	1025	1002	985	971	957
		Pjg Jajaran	5	694	656	641	576	564	558	551	547	526	501	500	500
		Padanan		34.62%	35.58%	35.42%	35.85%	35.16%	35.05%	34.84%	34.61%	34.02%	33.91%	33.88%	33.72%
	BLOSUM62	Markah		1347	1207	1110	1042	990	953	939	894	874	855	839	824
		Pjg Jajaran	1	754	722	647	635	616	605	603	584	575	570	550	550
		Padanan		34.10%	35.29%	35.17%	36.03%	35.71%	36.53%	41.99%	34.62%	34.11%	33.82%	38.84%	33.52%
		Markah		1271	1128	1037	977	932	898	869	845	826	810	797	786
		Pjg Jajaran	2	707	679	612	591	579	565	556	542	515	512	484	484
		Padanan		35.07%	36.82%	37.43%	37.03%	36.55%	36.11%	35.69%	34.99%	34.78%	34.71%	36.18%	33.84%
		Pjg Jajaran		1235	1087	998	939	895	861	833	813	796	782	771	762
		Pjg Jajaran	3	699	657	585	576	565	553	515	511	511	462	434	409
		Padanan		35.39%	36.23%	37.11%	36.70%	36.33%	36.05%	35.82%	34.91%	34.91%	34.79%	35.82%	37.38%
		Markah		1214	1068	977	912	868	836	809	789	775	763	754	747
		Pjg Jajaran	4	685	651	580	566	550	538	518	464	442	405	404	370
		Padanan		34.91%	35.59%	36.58%	36.09%	35.69%	35.51%	35.46%	35.29%	35.86%	37.85%	37.33%	38.16%
		Markah		1205	1056	967	899	852	817	790	772	759	749	744	739
		Pjg Jajaran	5	684	607	576	565	547	537	497	432	432	364	316	316
		Padanan		34.48%	36.09%	36.32%	35.91%	35.44%	35.10%	32.34%	36.99%	36.79%	37.97%	41.63%	41.36%
	BLOSUM80	Markah		1390	1225	1113	1032	968	918	880	851	827	818	792	779
		Pjg Jajaran	1	776	735	663	642	627	601	582	560	538	535	504	493
		Padanan		33.22%	35.24%	35.83%	36.43%	36.33%	36.80%	36.91%	35.49%	37.13%	45.65%	34.49%	34.23%
		Markah		1303	1132	1030	952	897	856	825	802	783	770	761	754
		Pjg Jajaran	2	727	692	631	607	572	556	526	502	443	413	365	323
		Padanan		35.58%	37.04%	37.07%	37.43%	37.44%	36.66%	36.31%	36.09%	38.60%	42.49%	39.66%	41.40%
		Pjg Jajaran		1259	1081	977	907	855	816	792	775	764	755	749	745
		Pjg Jajaran	3	719	676	589	574	558	528	446	408	358	317	298	298
		Padanan		34.92%	36.58%	37.81%	37.50%	37.19%	37.05%	37.50%	39.36%	43.51%	43.88%	46.98%	46.64%
		Markah		1233	1051	951	877	824	792	774	762	753	746	742	738
		Pjg Jajaran	4	711	615	583	567	554	429	362	358	318	292	292	292
		Padanan		34.47%	36.96%	37.22%	36.93%	36.84%	39.22%	41.51%	41.84%	43.19%	47.74%	47.72%	47.53%
		Markah		1217	1037	938	860	805	777	762	752	745	740	736	732
		Pjg Jajaran	5	709	607	576	561	469	365	362	318	311	292	292	292
		Padanan		34.15%	36.43%	36.96%	36.80%	36.89%	41.73%	41.49%	43.08%	42.33%	47.74%	47.74%	47.72%

LAMPIRAN N

Jadual Hasil SWAffine Dari Proses Pernormalan Mengikut Kategori Data

LAMPIRAN N : Jadual Hasil SWAffine Dari Proses Pernormalan Mengikut Kategori Data

Data	Matriks	Output	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Pendek		Markah	1	1.6019	1.5962	1.6217	1.5412	1.4553	1.4813	1.4684	1.5208	1.5071	1.5285	1.5592	1.5446
		Plg Jajaran		0.1480	0.1305	0.6477	0.8918	0.0549	1.1623	0.6661	0.9503	1.0269	1.1132	0.8753	0.9238
		Padanan		-1.0963	-1.2415	-1.0920	-1.1184	0.7019	-1.6638	-1.3573	-0.0732	-0.8180	-1.3655	-1.0953	1.7693
		Markah	2	0.3031	0.2756	0.2427	0.3692	0.4671	0.4347	0.5306	0.4545	0.4646	0.4219	0.3568	0.3882
		Plg Jajaran		1.6891	1.6943	1.4525	1.2784	1.7175	1.0025	1.0303	1.1526	1.0640	1.0583	1.2690	1.2522
		Padanan		-1.0863	-0.9272	-1.0982	-1.0709	-1.7686	-0.2141	-0.3525	-1.4970	-0.6856	1.4389	-0.9052	-0.6936
	BLOSUM45	Plg Jajaran	3	-0.3463	-0.2699	-0.3365	-0.2461	-0.1617	-0.1825	-0.3001	-0.3311	-0.3096	-0.3043	-0.3025	-0.3139
		Plg Jajaran		-0.5999	-0.5694	-0.6926	-0.7033	-0.5908	-0.4876	0.4212	-0.2579	-0.2660	-0.5091	-0.4688	-0.7254
		Padanan		0.8695	0.8329	0.7499	0.7856	0.3556	0.5425	0.5814	-0.1434	-0.4146	-0.2293	0.0101	-0.4007
		Markah	4	-0.7174	-0.7005	-0.6398	-0.6656	-0.6707	-0.6655	-0.7289	-0.7239	-0.6969	-0.6847	-0.6516	-0.6443
		Plg Jajaran		-0.6186	-0.6277	-0.6926	-0.7334	-0.5908	-0.6386	-0.8649	-0.9225	-0.7157	-0.8312	-0.8378	-0.7254
		Padanan		0.6565	0.7488	0.7499	0.7018	0.3556	0.6677	1.2918	0.4930	0.8027	0.0779	0.9952	-0.3375
		Markah	5	-0.8411	-0.9015	-0.8881	-0.9786	-1.0899	-1.0680	-0.9700	-0.9203	-0.9650	-0.9614	-0.9619	-0.9747
		Plg Jajaran		-0.6186	-0.6277	-0.7150	-0.7334	-0.5908	-0.6386	-1.2527	-0.9225	-1.1092	-0.8312	-0.8378	-0.7254
		Padanan		0.6565	0.5870	0.6905	0.7018	0.3556	0.6677	-0.1635	1.2206	1.3154	0.0779	0.9952	-0.3375
		Markah	1	1.5893	1.6320	1.5704	1.5667	1.5215	1.5723	1.5476	1.5737	1.5896	1.5490	1.4526	1.3344
		Plg Jajaran		1.7692	1.6530	1.7576	1.6830	1.3235	1.1656	1.2104	1.1088	1.6655	1.7889	1.7889	1.7889
		Padanan		-1.6806	-1.5996	-1.7888	-1.7475	-1.5183	-1.0733	0.2656	-1.1209	-1.6524	-1.7516	1.1046	-1.7889
		Markah	2	0.3268	0.2114	0.3475	0.3348	0.3966	0.3726	0.3959	0.3378	0.2573	0.3078	0.4304	0.5600
		Plg Jajaran		-0.2054	0.2488	-0.2549	0.1012	0.2814	1.0225	0.9455	1.0820	0.2158	-0.4472	-0.4472	-0.4472
		Padanan		-0.1599	-0.3726	0.4509	0.2631	0.1364	-1.0989	-1.7475	-1.0696	-0.2503	0.7430	-0.7303	0.4472
	BLOSUM62	Plg Jajaran	3	-0.3416	-0.3171	-0.3282	-0.2653	-0.2250	-0.3852	-0.3239	-0.3213	-0.2481	-0.1887	-0.1076	-0.0357
		Plg Jajaran		-0.4423	-0.6339	-0.2549	-0.3417	0.1251	-0.7294	-0.4714	-0.7303	-0.6271	-0.4472	-0.4472	-0.4472
		Padanan		0.7194	0.6071	0.4509	0.2209	-0.3380	0.6180	0.1845	0.7302	0.6342	0.1712	-0.7303	0.4472
		Markah	4	-0.7130	-0.6475	-0.6500	-0.7075	-0.6986	-0.6693	-0.6838	-0.6509	-0.6156	-0.6355	-0.6456	-0.6314
		Plg Jajaran		-0.5213	-0.6339	-0.5904	-0.7213	-0.2918	-0.7294	-0.8422	-0.7303	-0.6271	-0.4472	-0.4472	-0.4472
		Padanan		0.5162	0.7157	0.4471	0.5093	0.6568	0.6180	0.6487	0.7302	0.6342	0.4187	-0.7303	0.4472
		Markah	5	-0.8615	-0.8787	-0.9396	-0.9286	-0.9946	-0.8903	-0.9358	-0.9393	-0.9831	-1.0327	-1.1298	-1.2271
		Plg Jajaran		-0.6003	-0.6339	-0.6574	-0.7213	-1.4382	-0.7294	-0.8422	-0.7303	-0.6271	-0.4472	-0.4472	-0.4472
		Padanan		0.6050	0.6494	0.4399	0.7541	1.0632	0.9363	0.6487	0.7302	0.6342	0.4187	1.0863	0.4472
		Markah	1	1.6562	1.6809	1.6325	1.5957	1.6464	1.6752	1.5867	1.5140	1.4290	1.3361	1.2790	1.2921
		Plg Jajaran		1.6746	1.7799	1.7824	1.2116	1.0938	1.6843	1.4909	1.7886	1.7889	1.7889	0.0000	0.0000
		Padanan		-1.3464	-1.4898	-1.3478	-1.1389	-0.7434	-1.6112	-1.5512	-1.7463	-1.6814	1.7874	0.9706	0.4365
		Markah	2	0.1793	0.0838	0.2305	0.2863	0.1787	0.1226	0.2851	0.3820	0.4714	0.5938	0.6323	0.6048
		Plg Jajaran		0.1647	-0.3303	-0.3610	0.5757	0.2712	0.1623	0.5843	-0.4159	-0.4472	-0.4472	0.0000	0.0000
		Padanan		-0.1084	0.8065	0.9945	-0.3507	-0.3124	-0.3460	-0.4741	0.3444	0.7564	-0.5174	0.5973	0.4365
	BLOSUM80	Plg Jajaran	3	-0.3745	-0.3090	-0.3614	-0.2923	-0.3318	-0.3677	-0.2727	-0.1839	-0.1179	-0.0742	-0.0144	-0.0137
		Plg Jajaran		-0.4530	-0.4220	-0.3610	0.2744	0.1378	-0.6113	-0.6749	-0.4575	-0.4472	-0.4472	0.0000	0.0000
		Padanan		1.4692	0.6308	0.9945	-0.6111	-0.3526	0.6448	0.6400	0.1699	0.7564	-0.4233	0.5973	0.4578
		Markah	4	-0.6382	-0.6493	-0.6418	-0.6882	-0.6190	-0.6129	-0.6446	-0.6792	-0.6335	-0.6680	-0.6610	-0.6323
		Plg Jajaran		-0.6589	-0.4220	-0.5302	-1.0309	0.1378	-0.6113	-0.7001	-0.4575	-0.4472	-0.4472	0.0000	0.0000
		Padanan		0.0842	0.6308	-0.3205	1.0503	-0.3526	0.7163	0.6927	0.4673	-0.0010	-0.4233	-1.0826	0.4578
		Markah	5	-0.8228	-0.8064	-0.8599	-0.9014	-0.8742	-0.8172	-0.9545	-1.0329	-1.1491	-1.1876	-1.2358	-1.2509
		Plg Jajaran		-0.7275	-0.6055	-0.5302	-1.0309	-1.6407	-0.6240	-0.7001	-0.4575	-0.4472	-0.4472	0.0000	0.0000
		Padanan		-0.0986	-0.5782	-0.3205	1.0503	1.7610	0.5961	0.6927	0.7647	0.1696	-0.4233	-1.0826	-1.7888

Data	Matriks	Output	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
		Markah		1.5863	1.5839	1.5658	1.5435	1.5340	1.5333	1.5356	1.5159	1.5121	1.5028	1.5003	1.4952
		Plg Jajaran	1	1.7331	1.6246	1.4437	1.4607	1.3375	1.4348	1.4017	1.3881	1.6685	1.1203	1.1202	1.0954
		Padanan		-1.5245	-1.1544	-1.0791	-1.2868	-0.6190	-0.6597	-0.3650	-0.3961	-1.3773	-0.4948	1.7439	0.1397
		Markah		0.3222	0.3321	0.3646	0.4070	0.4185	0.4438	0.4310	0.4540	0.4471	0.4640	0.4596	0.4584
Sederhana		Plg Jajaran	2	-0.1057	0.1682	0.4125	0.3207	0.4110	0.4389	0.4460	0.6097	0.2071	1.0701	1.0699	1.0954
		Padanan		0.9294	1.4699	-0.2776	1.3116	1.5156	0.3086	-1.0508	-1.2810	-0.7011	1.7849	-0.8069	-1.7569
	BLOSUM45	Plg Jajaran		-0.3187	-0.3236	-0.3169	-0.3154	-0.3005	-0.3656	-0.3487	-0.3271	-0.3024	-0.3049	-0.2838	-0.2647
		Plg Jajaran	3	-0.3586	-0.3036	0.0115	0.0889	0.2426	-0.0037	-0.1085	-0.1686	-0.6099	-0.7134	-0.7133	-0.7469
		Padanan		0.7181	-0.0639	-0.7019	-0.4664	0.3857	-1.3855	-0.4495	-0.3571	0.3609	-0.5286	-0.4075	0.4445
		Markah		-0.7104	-0.7324	-0.7258	-0.7295	-0.7338	-0.7057	-0.6952	-0.6811	-0.6706	-0.6691	-0.6758	-0.6739
		Plg Jajaran	4	-0.4735	-0.4882	-0.9052	-0.8772	-0.9703	-0.9258	-0.4743	-0.9146	-0.6214	-0.7385	-0.7133	-0.7220
		Padanan		0.3186	-0.6111	1.3006	0.6138	-0.2235	1.0132	0.2920	0.9632	0.9149	-0.3808	-0.2904	0.5213
		Markah		-0.8795	-0.8601	-0.8877	-0.9057	-0.9162	-0.9056	-0.9226	-0.9618	-0.9862	-0.9928	-1.0002	-1.0150
		Plg Jajaran	5	-0.7953	-1.0010	-0.9624	-0.9931	-1.0208	-0.9442	-1.2648	-0.9146	-0.6444	-0.7385	-0.7635	-0.7220
		Padanan		-0.4416	0.3595	0.7580	-0.1721	-1.0588	0.7234	1.5733	1.0710	0.8026	-0.3808	-0.2392	0.6513
		Markah		1.5857	1.5705	1.5672	1.5584	1.5463	1.5110	1.5362	1.5625	1.5753	1.5658	1.5551	1.5304
		Plg Jajaran	1	1.6815	1.5473	1.4190	1.3667	1.5794	1.0943	1.0948	1.1505	1.1519	1.0656	1.5896	1.7888
		Padanan		-1.5904	-1.2641	-1.6476	-1.6165	-1.4840	-1.1380	0.3919	-1.2067	-1.0656	-0.9694	1.4660	-1.7660
		Markah		0.3338	0.3466	0.3632	0.3685	0.4143	0.4632	0.4044	0.3307	0.2980	0.3041	0.3064	0.3342
		Plg Jajaran	2	0.1322	0.3186	0.3907	0.6633	0.1158	1.0328	1.0326	1.0386	1.0370	0.9684	0.3969	-0.4367
		Padanan		1.0058	0.9387	-0.0237	0.0060	-0.1005	-0.9902	-1.7831	-0.9748	-1.1203	-1.0977	-0.7720	0.3385
	BLOSUM62	Plg Jajaran		-0.3418	-0.2996	-0.3178	-0.3413	-0.3565	-0.3226	-0.3048	-0.2691	-0.2587	-0.2297	-0.2098	-0.1759
		Plg Jajaran	3	-0.5058	-0.2503	0.0994	-0.2161	0.0316	-0.3197	-0.3203	-0.7191	-0.7187	-0.3887	-0.6587	-0.4367
		Padanan		0.6719	1.0935	0.8700	-0.0492	-0.2679	0.3490	0.3460	0.6573	0.6252	0.2841	0.1958	0.3385
		Markah		-0.7094	-0.7206	-0.7323	-0.7170	-0.6937	-0.6948	-0.6670	-0.6419	-0.6354	-0.6503	-0.6260	-0.6333
		Plg Jajaran	4	-0.5058	-0.5461	-0.9289	-0.8944	-0.8634	-0.8883	-0.8957	-0.7350	-0.7187	-0.4144	-0.6587	-0.4577
		Padanan		-0.0013	-0.2611	0.7223	0.7703	0.9262	0.9886	0.5025	0.7621	0.7656	0.5872	0.1958	0.7161
		Markah		-0.8684	-0.8969	-0.8803	-0.8887	-0.9105	-0.9568	-0.9688	-0.9822	-0.9792	-0.9900	-1.0256	-1.0555
		Plg Jajaran	5	-0.8020	-1.0694	-0.9803	-0.9195	-0.8634	-0.9191	-0.9113	-0.7350	-0.7515	-1.2509	-0.6692	-0.4577
		Padanan		-0.0860	-0.5070	0.0789	0.8893	0.9262	0.7905	0.5428	0.7621	0.7952	1.1957	-1.0855	0.3729
		Markah		1.6060	1.5898	1.5847	1.5906	1.5734	1.5568	1.5750	1.5913	1.4590	1.5908	1.5601	1.5482
		Plg Jajaran	1	1.5152	1.4795	1.3506	1.5714	1.5745	1.6023	1.3288	1.4310	1.3534	1.6124	1.5542	1.5024
		Padanan		-1.5444	-1.7671	-1.5172	-1.5999	-1.6072	-1.6487	-0.9570	-1.3160	-1.2129	0.1428	-1.3019	-1.7232
		Markah		0.3041	0.3407	0.3483	0.3211	0.3445	0.3812	0.3161	0.2776	0.1496	0.3231	0.3674	0.3511
		Plg Jajaran	2	0.5401	0.4932	0.6944	0.1751	0.3669	0.3680	0.8249	0.2521	0.3224	0.0336	0.4690	0.5500
		Padanan		0.4194	0.4395	-0.4602	0.2327	-0.3213	-0.2268	-1.2079	-0.3492	-0.1056	0.3027	-0.8080	-0.0019
	BLOSUM80	Plg Jajaran		-0.3833	-0.3761	-0.3679	-0.3430	-0.3135	-0.3264	-0.2732	-0.2747	0.2687	-0.3517	-0.3244	-0.2607
		Plg Jajaran	3	-0.6788	-0.2336	-0.2351	-0.0149	-0.4459	-0.6133	-0.6472	-0.1342	0.3086	0.0263	-0.6744	-0.6064
		Padanan		1.0231	0.6819	0.3399	-0.2447	0.7901	0.8185	0.9044	0.2490	1.5748	-1.7447	0.4507	0.4950
		Markah		-0.6744	-0.7168	-0.7222	-0.6877	-0.7006	-0.6916	-0.6750	-0.6478	-0.8536	-0.6789	-0.6584	-0.6331
		Plg Jajaran	4	-0.6413	-0.8436	-0.8913	-0.8493	-0.5620	-0.6670	-0.7533	-0.1951	-0.9922	-0.8362	-0.6744	-0.6064
		Padanan		0.4909	0.3961	0.9411	0.9230	0.4250	0.5580	0.5506	-0.0272	-0.0779	0.6496	0.6386	0.4950
		Markah		-0.8523	-0.8375	-0.8429	-0.8611	-0.9038	-0.9199	-0.9429	-0.9464	-1.0237	-0.8833	-0.9447	-1.0055
		Plg Jajaran	5	-0.7351	-0.8955	-0.9186	-0.8624	-0.9336	-0.6900	-0.7533	-1.3538	-0.9922	-0.8362	-0.6744	-0.8396
		Padanan		-0.3689	0.2496	0.6964	0.6889	0.7134	0.4990	0.7099	1.4434	-0.1781	0.6496	1.0205	0.7351

Data	Matrika	Output	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Panjang		Markah	1	1.5832	1.5815	1.5649	1.5326	1.5014	1.4831	1.4845	1.4700	1.4578	1.4589	1.4615	1.4561
		Plg Jajaran		1.5897	1.6717	1.6479	1.5313	1.5528	1.6164	1.5548	1.5577	1.5157	1.4394	1.3849	1.4260
		Padanan		-0.5004	-0.6026	-0.9584	-0.4036	-1.5733	0.2015	1.5909	1.7830	-1.4415	-1.7720	-1.4909	-1.1640
		Markah	2	0.3377	0.3408	0.3708	0.4116	0.4396	0.4562	0.4519	0.4688	0.4740	0.4751	0.4759	0.4860
		Plg Jajaran		0.3688	0.1110	0.2308	0.4403	0.4341	0.2118	0.3217	0.3864	0.3548	0.3886	0.4173	0.5506
		Padanan		1.1006	1.0964	1.5474	1.5539	0.8643	1.3532	0.0710	-0.3337	1.0862	0.3410	-0.5430	-1.0061
	BLOSUM45	Plg Jajaran	3	-0.3442	-0.3432	-0.3343	-0.2878	-0.2373	-0.2126	-0.2117	-0.2026	-0.1800	-0.1747	-0.1865	-0.1927
		Plg Jajaran		-0.4578	-0.3003	-0.4414	-0.3263	-0.4547	-0.2717	-0.1742	-0.3272	-0.3053	-0.0334	0.1409	-0.1682
		Padanan		1.0588	1.0361	0.3863	0.3741	0.8401	0.3273	0.0429	-0.4043	0.7463	0.6789	0.4287	0.5728
		Markah	4	-0.6940	-0.6935	-0.7022	-0.7120	-0.7019	-0.6955	-0.6893	-0.6796	-0.6719	-0.6958	-0.6887	-0.6794
		Plg Jajaran		-0.7249	-0.6966	-0.6766	-0.7981	-0.6663	-0.6094	-0.8511	-0.7378	-0.4231	-0.6093	-0.9604	-0.9017
		Padanan		-0.6863	-0.4556	-0.3115	-0.5168	0.1126	-0.5873	-0.9194	-0.5260	0.0379	0.3797	0.7307	0.6785
		Markah	5	-0.8627	-0.8855	-0.8993	-0.9443	-1.0018	-1.0313	-1.0354	-1.0566	-1.0799	-1.0636	-1.0622	-1.0700
		Plg Jajaran		-0.7758	-0.7858	-0.7606	-0.8472	-0.8659	-0.9471	-0.8511	-0.8792	-1.1421	-1.1852	-0.9827	-0.9066
		Padanan		-0.9727	-1.0742	-0.6638	-1.0077	-0.2437	-1.2947	-0.7854	-0.5190	-0.4289	0.3724	0.8745	0.9188
		Markah	1	1.6048	1.5992	1.5871	1.5311	1.4926	1.4790	1.5469	1.4761	1.4922	1.5015	1.5161	1.5313
		Plg Jajaran		1.6836	1.3938	1.5751	1.6690	1.5850	1.6266	1.5438	1.2843	1.3626	1.3018	1.2847	1.3422
		Padanan		-1.3682	-1.2067	-1.5465	-0.6706	-0.5040	1.1999	1.6313	-0.7854	-1.1281	-1.0354	0.3723	-1.0229
		Markah	2	0.2916	0.3108	0.3305	0.3997	0.4492	0.4630	0.3594	0.4645	0.4421	0.4318	0.4220	0.4051
		Plg Jajaran		0.0351	0.3897	0.4090	0.1533	0.2900	0.2015	0.4252	0.5791	0.3381	0.6061	0.5264	0.6295
		Padanan		0.5571	1.3729	1.0459	1.3888	1.2884	0.4414	-0.1626	-0.3954	-0.4881	-0.5783	-0.7513	-0.9235
	BLOSUM62	Plg Jajaran	3	-0.3403	-0.3622	-0.3437	-0.2559	-0.2243	-0.2143	-0.2556	-0.2064	-0.2225	-0.2298	-0.2493	-0.2717
		Plg Jajaran		-0.2280	-0.1500	-0.5205	-0.3711	-0.2351	-0.2457	-0.5387	0.0702	0.2698	-0.0045	-0.0418	-0.1821
		Padanan		1.1794	0.3808	0.6745	0.7193	0.8180	0.3483	-0.1241	-0.4778	-0.3648	-0.5340	-0.9053	0.1617
		Markah	4	-0.6969	-0.6823	-0.6994	-0.7200	-0.7084	-0.6915	-0.6666	-0.6950	-0.6778	-0.6854	-0.7114	-0.7067
		Plg Jajaran		-0.7190	-0.2929	-0.6614	-0.7112	-0.7782	-0.7704	-0.4673	-0.7013	-0.8999	-0.7002	-0.3820	-0.6050
		Padanan		0.2395	-0.7004	0.0623	-0.5306	-0.5435	-0.6307	-0.2291	-0.0716	0.5466	1.0423	-0.2668	0.4013
		Markah	5	-0.8592	-0.8655	-0.8744	-0.9550	-1.0091	-1.0363	-0.9840	-1.0391	-1.0340	-1.0181	-0.9773	-0.9580
		Plg Jajaran		-0.7716	-1.3406	-0.8022	-0.7400	-0.8618	-0.8121	-0.9631	-1.2323	-1.0706	-1.2032	-1.3873	-1.1845
		Padanan		-0.6078	0.1534	-0.2362	-0.9069	-1.0590	-1.3588	-1.1156	1.7301	1.4343	1.1054	1.5511	1.3834
		Markah	1	1.5784	1.5713	1.5568	1.5408	1.5104	1.5195	1.5472	1.5818	1.6109	1.6677	1.6286	1.5848
		Plg Jajaran		1.7283	1.3026	1.4615	1.5314	1.2537	1.0919	1.2895	1.3042	1.4984	1.5742	1.6839	1.7589
		Padanan		-1.4192	-1.6749	-1.5923	-1.3146	-1.4536	-0.6826	-0.7190	-1.0924	-1.3182	0.0643	-1.4730	-1.6007
		Markah	2	0.3315	0.3546	0.3927	0.3870	0.4161	0.4274	0.3943	0.3299	0.2676	0.1429	0.2124	0.2641
		Plg Jajaran		-0.0617	0.4958	0.6102	0.4844	0.2864	0.6202	0.7176	0.7224	0.5139	0.4113	0.1619	-0.1888
		Padanan		1.2616	0.8135	0.1315	0.9227	1.2066	-0.7450	-0.9546	-0.9134	-0.8124	-1.2911	-0.6106	-0.3634
	BLOSUM80	Plg Jajaran	3	-0.3123	-0.3214	-0.3436	-0.2709	-0.2235	-0.2751	-0.3156	-0.3316	-0.3305	-0.3636	-0.3149	-0.2368
		Plg Jajaran		-0.3278	0.2042	-0.5188	-0.4844	0.0365	0.3344	-0.1039	-0.2092	-0.3668	-0.5015	-0.5726	-0.4817
		Padanan		0.5131	0.1830	1.1560	1.0724	0.6000	-0.5710	-0.4872	0.0555	0.6837	-0.6947	0.6115	0.5411
		Markah	4	-0.6871	-0.7161	-0.7176	-0.7063	-0.7013	-0.7031	-0.6917	-0.6666	-0.6600	-0.6248	-0.6237	-0.6285
		Plg Jajaran		-0.6180	-0.9284	-0.6846	-0.6850	-0.0459	-0.6935	-0.9508	-0.7065	-0.7873	-0.7420	-0.6366	-0.5492
		Padanan		0.0066	0.7061	0.3320	-0.2009	-0.2424	0.4267	1.0860	0.7917	0.7720	0.9608	0.7342	0.6955
		Markah	5	-0.9105	-0.8884	-0.8883	-0.9506	-1.0018	-0.9687	-0.9342	-0.9136	-0.8881	-0.8221	-0.9024	-0.9837
		Plg Jajaran		-0.7208	-1.0742	-0.8683	-0.8464	-1.5306	-1.3529	-0.9525	-1.1110	-0.8582	-0.7420	-0.6366	-0.5492
		Padanan		-0.3620	-0.0278	-0.0271	-0.4795	-0.1105	1.5719	1.0747	1.1587	0.4749	0.9608	0.7379	0.7276

LAMPIRAN O

Jadual *RZ-Score* Bagi Hasil SWAffine Mengikut Kategori Data

LAMPIRAN O : Jadual RZ-Score Bagi Hasil SWAffine Mengikut Kategori Data

Data	Matriks	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Pendek		1	0.22	0.16	0.39	0.44	0.74	0.33	0.26	0.80	0.57	0.43	0.45	1.41
		2	0.30	0.35	0.20	0.19	0.14	0.41	0.40	0.04	0.21	0.97	0.24	0.32
	BLOSUM45	3	-0.03	0.00	-0.09	-0.05	-0.13	-0.04	0.23	-0.24	-0.33	-0.35	-0.25	-0.48
		4	-0.23	-0.19	-0.19	-0.24	-0.30	-0.28	-0.10	-0.38	-0.20	-0.48	-0.16	-0.57
		5	-0.27	-0.31	-0.30	-0.34	-0.44	-0.41	-0.80	-0.21	-0.25	-0.57	-0.27	-0.68
		1	0.56	0.56	0.51	0.50	0.44	0.55	1.01	0.52	0.53	0.53	1.45	0.44
		2	-0.01	0.03	0.18	0.23	0.27	0.10	-0.14	0.12	0.07	0.20	-0.25	0.19
	BLOSUM62	3	-0.02	-0.11	-0.04	-0.13	-0.15	-0.17	-0.20	-0.11	-0.08	-0.15	-0.43	-0.01
		4	-0.24	-0.19	-0.26	-0.31	-0.11	-0.26	-0.29	-0.22	-0.20	-0.22	-0.61	-0.21
		5	-0.29	-0.29	-0.39	-0.30	-0.46	-0.23	-0.38	-0.31	-0.33	-0.35	-0.16	-0.41
		1	0.66	0.66	0.69	0.56	0.67	0.58	0.51	0.52	0.51	1.64	0.75	0.58
		2	0.08	0.19	0.29	0.17	0.05	-0.02	0.13	0.10	0.26	-0.12	0.41	0.35
	BLOSUM80	3	0.21	-0.03	0.09	-0.21	-0.18	-0.11	-0.10	-0.16	0.06	-0.31	0.19	0.15
		4	-0.40	-0.15	-0.50	-0.22	-0.28	-0.17	-0.22	-0.22	-0.36	-0.51	-0.58	-0.06
		5	-0.55	-0.66	-0.57	-0.29	-0.25	-0.28	-0.32	-0.24	-0.48	-0.69	-0.77	-1.01

Data	Matriks	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Sederhana		1	0.60	0.68	0.64	0.57	0.75	0.77	0.86	0.84	0.60	0.71	1.45	0.91
		2	0.38	0.66	0.17	0.68	0.78	0.40	-0.06	-0.07	-0.02	1.11	0.24	-0.07
	BLOSUM45	3	0.01	-0.23	-0.34	-0.23	0.11	-0.59	-0.30	-0.28	-0.18	-0.52	-0.47	-0.19
		4	-0.29	-0.61	-0.11	-0.33	-0.64	-0.21	-0.29	-0.21	-0.13	-0.60	-0.56	-0.29
		5	-0.71	-0.50	-0.36	-0.69	-1.00	-0.38	-0.20	-0.27	-0.28	-0.70	-0.67	-0.36
		1	0.56	0.62	0.45	0.44	0.55	0.49	1.01	0.50	0.55	0.55	1.54	0.52
		2	0.49	0.53	0.24	0.35	0.14	0.17	-0.12	0.13	0.07	0.06	-0.02	0.08
	BLOSUM62	3	-0.06	0.18	0.22	-0.20	-0.20	-0.10	-0.09	-0.11	-0.12	-0.11	-0.22	-0.09
		4	-0.41	-0.51	-0.31	-0.28	-0.21	-0.20	-0.35	-0.20	-0.20	-0.16	-0.36	-0.12
		5	-0.59	-0.82	-0.59	-0.31	-0.28	-0.36	-0.45	-0.32	-0.31	-0.35	-0.93	-0.38
		1	0.53	0.43	0.47	0.52	0.51	0.50	0.65	0.57	0.53	1.12	0.60	0.44
		2	0.42	0.42	0.19	0.24	0.13	0.17	-0.02	0.06	0.12	0.22	0.01	0.30
	BLOSUM80	3	-0.01	0.02	-0.09	-0.20	0.01	-0.04	-0.01	-0.05	0.72	-0.69	-0.18	-0.12
		4	-0.27	-0.39	-0.22	-0.20	-0.28	-0.27	-0.29	-0.29	-0.64	-0.29	-0.23	-0.25
		5	-0.66	-0.49	-0.36	-0.36	-0.37	-0.37	-0.33	-0.29	-0.73	-0.36	-0.20	-0.37

Data	Matriks	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Panjang		1	0.89	0.88	0.75	0.89	0.49	1.10	1.54	1.60	0.51	0.38	0.45	0.57
		2	0.60	0.52	0.72	0.80	0.58	0.67	0.28	0.17	0.64	0.40	0.12	0.01
	BLOSUM45	3	0.09	0.13	-0.13	-0.08	0.05	-0.05	-0.11	-0.31	0.09	0.16	0.13	0.07
		4	-0.70	-0.62	-0.56	-0.68	-0.42	-0.63	-0.82	-0.65	-0.35	-0.31	-0.31	-0.30
		5	-0.88	-0.92	-0.77	-0.93	-0.70	-1.09	-0.89	-0.82	-0.88	-0.63	-0.39	-0.35
		1	0.64	0.60	0.54	0.84	0.86	1.44	1.57	0.66	0.58	0.59	1.06	0.62
		2	0.29	0.69	0.60	0.65	0.68	0.37	0.21	0.22	0.10	0.15	0.07	0.04
	BLOSUM62	3	0.20	-0.04	-0.06	0.03	0.12	-0.04	-0.31	-0.20	-0.11	-0.26	-0.40	-0.10
		4	-0.39	-0.56	-0.43	-0.65	-0.68	-0.70	-0.45	-0.49	-0.34	-0.11	-0.45	-0.30
		5	-0.75	-0.68	-0.64	-0.87	-0.98	-1.07	-1.02	-0.18	-0.22	-0.37	-0.27	-0.25
		1	0.63	0.40	0.48	0.59	0.44	0.64	0.71	0.60	0.60	1.10	0.61	0.58
		2	0.51	0.55	0.38	0.60	0.64	0.10	0.05	0.05	-0.01	-0.25	-0.08	-0.10
	BLOSUM80	3	-0.04	0.02	0.10	0.11	0.14	-0.17	-0.30	-0.16	0.06	-0.52	-0.09	-0.06
		4	-0.43	-0.31	-0.36	-0.53	-0.33	-0.32	-0.19	-0.19	-0.23	-0.14	-0.18	-0.16
		5	-0.66	-0.66	-0.59	-0.76	-0.88	-0.25	-0.27	-0.29	-0.42	-0.20	-0.27	-0.27

LAMPIRAN P

Jadual *RZ-Score* Bagi Hasil SWAffine
Mengikut Kategori Panjang Jujukan Dengan
Peratusan Kesamaan Identiti

LAMPIRAN P : Jadual RZ-Score Bagi Hasil SWAffine Mengikut Kategori Panjang Jujukan Dengan Peratusan Kesamaan Identiti

Data	Mat	-e\ -d	1	2	3	4	5	6	7	8	9	10	11	12
Pendek		1	0.40	0.50	0.52	0.52	0.55	0.50	0.60	0.70	0.69	0.62	0.60	0.60
<25%id		2	0.38	0.07	0.40	0.48	0.49	0.17	0.52	0.00	0.10	0.70	-0.15	-0.15
	B45	3	-0.07	-0.13	-0.58	-0.39	-0.46	0.00	0.30	-0.04	-0.48	-0.18	-0.15	-0.15
		4	-0.83	-0.64	-0.66	-0.53	-0.63	-0.26	-0.46	-0.67	-0.16	-0.18	-0.15	-0.15
		5	-0.87	-0.89	-0.72	-0.64	-0.79	-0.38	-1.29	-0.67	-0.16	-0.18	-0.15	-0.15
		1	0.32	0.33	0.41	0.44	0.52	0.50	0.59	0.59	0.60	0.60	0.60	
		2	0.30	0.24	0.19	0.23	0.40	0.13	0.01	-0.05	-0.15	-0.15	-0.15	
	B62	3	-0.14	-0.10	-0.06	-0.76	0.60	-0.21	-0.20	-0.18	-0.15	-0.15	-0.15	
		4	-0.78	-0.19	-0.28	-0.22	-0.29	-0.21	-0.20	-0.18	-0.15	-0.15	-0.15	
		5	-0.82	-0.46	-0.36	-0.29	-0.91	-0.21	-0.20	-0.18	-0.15	-0.15	-0.15	
		1	0.50	52.00	0.53	0.46	0.59	0.60	0.76					
		2	-0.07	0.21	0.33	0.24	-0.07	-0.21	-0.15					
	B80	3	0.32	-0.01	0.11	-0.15	-0.27	-0.14	-0.15					
		4	-0.63	-0.14	-0.56	-0.25	-0.27	-0.11	-0.15					
		5	-0.69	-0.85	-0.62	-0.31	-0.16	-0.14	-0.15					
20-40%id		1	0.38	0.36	0.49	0.51	0.52	0.60	0.41	0.52	0.53	0.46	0.53	0.69
		2	-0.10	-0.11	-0.05	0.04	0.52	-0.43	0.10	0.13	0.17	0.47	0.12	-0.06
	B45	3	-0.04	-0.19	-0.22	-0.17	-0.45	0.04	0.12	-0.05	-0.16	-0.12	-0.11	-0.22
		4	-0.18	-0.32	-0.35	-0.71	-0.61	-0.12	-0.27	-0.38	-0.20	-0.24	-0.23	-0.28
		5	-0.26	-0.94	-1.00	-0.80	-0.73	-0.26	-0.35	-0.23	-0.33	-0.32	-0.31	-0.31
		1	0.36	0.42	0.46	0.50	0.60	0.58	0.51	0.53	0.63	0.72	0.50	0.46
		2	-0.28	-0.18	0.24	-0.17	0.32	0.00	0.09	0.14	0.09	0.02	0.09	0.17
	B62	3	-0.32	-0.35	0.07	0.25	-0.43	-0.24	-0.11	-0.12	-0.10	-0.15	-0.08	-0.02
		4	-0.46	-0.52	-0.07	-0.38	-0.11	-0.36	-0.24	-0.22	-0.21	-0.14	-0.26	-0.21
		5	-0.54	-0.65	-1.07	-0.45	-0.39	-0.14	-0.32	-0.31	-0.32	-0.26	-0.25	-0.40
		1	0.45	0.50	0.52	0.55	0.60	0.61	0.74	0.55	0.50	0.73		
		2	0.57	0.41	-0.24	-0.21	0.08	0.02	0.00	0.06	0.12	0.19		
	B80	3	-0.55	-0.50	-0.38	-0.39	-0.11	-0.12	-0.14	-0.19	-0.08	-0.02		
		4	-0.66	-0.62	-0.50	-0.51	-0.27	-0.19	-0.21	-0.23	-0.23	-0.23		
		5	-0.77	-0.71	-0.61	-0.62	-0.31	-0.32	-0.27	-0.20	-0.31	-0.40		
>35%id		1	0.55	0.57	0.55	0.50	1.19	1.64						
		2	0.08	0.00	0.05	0.16	0.93	-0.11						
	B45	3	-0.21	-0.11	-0.12	-0.11	-0.54	-0.31						
		4	-0.21	-0.23	-0.20	-0.24	-0.73	-0.51						
		5	-0.21	-0.23	-0.29	-0.31	-0.85	-0.71						
		1	0.74	0.92	0.53	0.98	1.68							
		2	-0.11	0.09	0.11	-0.16	-0.11							
	B62	3	-0.25	-0.73	-0.13	-0.37	-0.31							
		4	-0.38	-0.14	-0.22	-0.57	-0.51							
		5	0.01	-0.14	-0.29	0.12	-0.71							
		1	0.84	0.63	0.56	0.52								
		2	0.34	-0.21	0.05	0.12								
	B80	3	0.02	-0.14	-0.20	-0.08								
		4	-0.07	-0.14	-0.20	-0.28								
		5	-1.13	-0.14	-0.20	-0.28								

Data	Mat	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Sederhana		1	0.20	0.22	0.25	0.33	0.35	0.50	0.58	0.58	0.54	0.60	0.55	0.56
<25%id		2	0.11	0.19	0.23	0.28	0.47	0.39	0.12	0.07	0.11	0.03	0.07	0.05
	B45	3	-0.01	-0.05	-0.44	-0.38	0.14	-0.33	-0.20	-0.20	-0.17	-0.14	-0.17	-0.18
		4	-0.29	-0.50	-0.20	-0.24	-0.47	-0.22	-0.30	-0.20	-0.19	-0.22	-0.22	-0.20
		5	-0.89	-0.30	-0.29	-0.34	-0.49	-0.33	-0.21	-0.25	-0.29	-0.27	-0.23	-0.23
		1	0.36	0.38	0.42	0.41	0.54	0.52	0.54	0.53	0.58	0.59	0.67	0.64
		2	0.22	0.28	0.32	0.39	0.19	0.13	0.09	0.07	-0.01	-0.06	-0.08	-0.10
	B62	3	-0.17	0.28	-0.39	-0.36	-0.23	-0.15	-0.16	-0.16	-0.14	-0.14	-0.15	-0.12
		4	-0.70	-0.41	-0.12	-0.20	-0.23	-0.22	-0.20	-0.20	-0.19	-0.14	-0.17	-0.21
		5	-0.77	-0.64	-0.41	-0.21	-0.27	-0.29	-0.28	-0.24	-0.25	-0.25	-0.19	-0.16
		1	0.34	0.34	0.47	0.49	0.55	0.55	0.56	0.61	0.60	0.59	0.59	0.58
		2	0.01	0.29	0.35	0.39	0.06	0.08	0.02	0.15	-0.07	-0.11	-0.08	-0.06
	B80	3	0.09	-0.30	-0.31	-0.35	-0.15	-0.17	-0.12	-0.18	-0.18	-0.13	-0.13	-0.12
		4	-0.34	-0.11	-0.13	-0.15	-0.20	-0.21	-0.21	-0.27	-0.17	-0.16	-0.17	-0.18
		5	-0.69	-0.29	-0.29	-0.26	-0.26	-0.24	-0.25	-0.31	-0.19	-0.19	-0.21	-0.23
20-40%id		1	0.58	0.52	0.57	0.42	1.53	1.14	1.40	0.54	0.16	0.48	1.27	0.68
		2	0.42	0.23	0.01	-0.02	0.39	0.75	-0.38	0.08	0.45	0.22	0.15	0.44
	B45	3	-0.01	-0.51	-0.31	-0.42	-0.26	-0.45	-0.25	-0.15	0.20	-0.15	-0.14	-0.54
		4	-0.47	-0.65	-0.15	-0.55	-0.67	-0.66	-0.14	-0.16	-0.40	-0.20	-0.56	-0.53
		5	-1.20	-0.81	-0.59	-1.02	-0.99	-0.77	-0.62	-0.31	-0.41	-0.36	-0.72	-0.05
		1	0.58	0.57	0.55	0.48	1.13	1.09	1.63	1.13	0.77	1.04	1.60	1.03
		2	0.32	0.25	0.33	0.09	0.16	0.85	-0.21	0.06	0.00	-0.35	-0.06	-0.08
	B62	3	-0.14	0.05	-0.30	-0.01	-0.34	-0.21	-0.14	-0.50	-0.40	-0.15	0.20	-0.30
		4	-0.85	-0.81	-0.66	-0.84	-0.67	-0.79	-0.32	-0.41	-0.24	-0.34	0.01	-0.24
		5	-0.93	-1.02	-0.85	-1.06	-0.78	-0.94	-0.47	-0.59	-0.14	-0.20	-0.91	-0.41
		1	0.57	0.62	0.69	0.68	1.38	1.26	1.56	0.63	0.42	0.50	0.43	0.44
		2	0.23	0.28	0.38	0.42	0.44	0.73	0.30	0.01	0.16	0.26	0.26	0.21
	B80	3	-0.15	0.27	-0.51	-0.52	0.05	-0.22	-0.36	0.01	0.01	-0.19	-0.16	-0.11
		4	-0.43	-0.81	-0.61	-0.41	-0.81	-0.84	-0.69	-0.27	-0.27	-0.24	-0.26	-0.22
		5	-1.15	-0.91	-0.79	-0.99	-1.06	-0.93	-0.81	-0.38	-0.34	-0.32	-0.27	-0.33
>35%id		1	0.32	0.38	0.22	0.19	0.17	-0.16	0.69	0.18	0.06	0.00	0.06	-0.12
		2	0.01	0.09	0.30	0.12	0.86	-0.53	-0.20	-0.34	-0.25	-0.09	-0.22	-0.01
	B45	3	0.05	0.17	0.40	0.33	0.24	0.69	-1.07	-0.61	-0.66	-0.74	-0.65	-0.63
		4	-0.38	0.02	-0.56	-0.34	-0.35	0.02	0.31	0.27	0.48	0.46	0.35	0.45
		5	-0.44	-1.24	-1.39	-0.97	-0.93	-0.02	0.27	0.50	0.37	0.37	0.46	0.32
		1	0.38	0.27	0.19	0.18	0.73	0.59	0.98	0.34	0.24	0.35	1.12	0.47
		2	0.00	0.30	0.01	0.12	0.29	0.42	0.58	0.24	0.23	0.25	0.18	0.16
	B62	3	-0.02	0.33	-0.14	0.08	0.28	0.27	0.30	0.15	0.19	0.03	-0.15	-0.02
		4	-0.14	-0.47	-0.83	-0.07	-0.06	-0.25	-0.22	-0.38	-0.12	-0.20	-0.33	-0.21
		5	-0.65	-1.13	-0.98	-0.86	-0.15	-0.38	-0.07	-0.35	-0.22	-0.35	-0.20	-0.39
		1	0.44	0.35	0.34	0.29	0.54	0.58	0.60	0.48	0.56	0.52	0.53	0.20
		2	-0.01	0.00	0.09	0.01	0.43	0.10	0.19	0.18	-0.03	0.05	-0.01	0.30
	B80	3	-0.24	0.01	-0.15	0.17	0.11	-0.12	0.07	-0.14	-0.09	-0.18	-0.23	0.09
		4	-0.48	-0.31	-0.34	-0.51	-0.03	-0.06	-0.47	-0.22	-0.18	-0.15	-0.08	-0.11
		5	-0.92	-0.41	-1.06	-0.65	-0.17	-0.51	-0.29	-0.30	-0.26	-0.24	-0.06	-0.48

Data	Mat	-e \-d	1	2	3	4	5	6	7	8	9	10	11	12
Panjang		1	0.43	0.46	0.65	0.58	0.49	1.26	1.28	1.34	1.01	0.52	0.36	0.60
<25%id		2	0.23	0.15	0.01	-0.02	0.46	0.49	0.48	0.55	0.70	0.70	0.32	0.00
	B45	3	0.15	0.23	-0.01	-0.08	0.04	0.03	0.15	0.07	0.08	0.24	0.14	0.12
		4	-0.65	-0.59	-0.64	-0.75	-0.43	-0.67	-0.89	-0.88	-0.45	-0.79	-0.51	-0.32
		5	-1.01	-1.01	-0.95	-0.98	-0.56	-1.10	-1.02	-1.04	-1.34	-0.66	-0.32	-0.40
		1	0.55	0.60	0.37	0.21	1.05	1.10	1.31	0.68	0.67	0.66	0.73	0.69
		2	0.32	0.22	0.18	0.20	0.61	0.60	0.58	0.25	0.13	0.11	-0.10	-0.11
	B62	3	-0.04	-0.05	-0.24	0.01	-0.02	0.23	-0.06	-0.24	-0.27	-0.38	-0.29	-0.15
		4	-0.65	-0.67	-0.72	-0.80	-0.86	-0.76	-0.27	-0.55	-0.37	-0.18	-0.16	-0.18
		5	-0.98	-0.93	-0.90	-0.92	-1.03	-1.17	-1.31	-0.14	-0.15	-0.21	-0.18	-0.23
		1	0.44	0.60	0.79	0.63	1.00	0.65	0.64	0.64	0.58	0.72	0.65	0.64
		2	0.44	0.60	0.79	0.63	0.59	0.08	0.06	-0.06	-0.02	-0.20	-0.21	-0.19
	B80	3	-0.10	0.05	-0.04	-0.19	-0.34	-0.21	-0.33	-0.11	-0.07	-0.32	-0.08	-0.04
		4	-0.56	-0.47	-0.78	-0.71	-0.62	-0.29	-0.15	-0.19	-0.14	-0.08	-0.15	-0.15
		5	-1.00	-0.90	-1.00	-1.07	-0.62	-0.22	-0.22	-0.28	-0.35	-0.13	-0.22	-0.26
20-40%id		1	0.42	0.52	0.63	0.77	0.46	0.66	1.15	0.63	0.50	0.50	0.65	0.56
		2	0.23	0.51	0.34	0.47	0.59	0.76	0.42	0.17	0.26	0.09	-0.17	-0.08
	B45	3	0.05	-0.08	-0.46	-0.38	0.07	-0.08	0.06	-0.47	-0.09	0.04	-0.06	0.12
		4	-0.66	-0.71	-0.70	-0.77	-0.46	-0.51	-0.51	-0.09	-0.31	-0.15	0.00	-0.31
		5	-0.88	-0.93	-0.84	-0.89	-1.17	-0.84	-0.42	-0.23	-0.35	-0.47	-0.41	-0.30
		1	0.32	0.45	0.48	0.66	1.00	0.87	1.26	0.56	0.51	0.50	0.53	0.54
		2	0.27	0.34	0.41	0.48	0.55	0.50	0.18	-0.02	-0.02	0.15	0.18	0.17
	B62	3	-0.21	-0.26	-0.06	-0.24	0.07	-0.01	0.14	-0.09	0.08	-0.06	-0.04	-0.04
		4	-0.70	-0.72	-0.63	-0.66	-0.61	-0.47	-0.41	-0.08	-0.25	-0.21	-0.33	-0.40
		5	-0.91	-0.63	-1.00	-0.85	-1.01	-0.89	-0.95	-0.37	-0.32	-0.37	-0.34	-0.27
		1	0.55	0.60	0.37	0.21	0.96	0.62	0.60	0.56	0.56	1.12	0.98	0.59
		2	0.38	0.27	0.19	0.18	0.51	0.13	0.11	0.13	0.07	0.05	-0.09	-0.07
	B80	3	-0.28	-0.45	-0.28	-0.16	0.07	-0.12	-0.15	-0.16	-0.11	-0.13	-0.18	-0.19
		4	-0.72	-0.16	-0.58	-0.78	-0.67	-0.31	-0.24	-0.21	-0.20	-0.23	-0.16	-0.15
		5	-0.85	-0.73	-1.04	-0.87	-1.06	-0.32	-0.31	-0.31	-0.32	-0.26	-0.20	-0.18
>35%id		1	0.56	0.15	0.24	0.44	0.64	0.80	1.16	0.97	0.41	0.40	0.59	0.60
		2	0.12	0.11	0.01	-0.02	0.16	0.55	0.16	0.40	0.41	0.36	-0.02	0.35
	B45	3	-0.07	0.22	0.13	0.06	0.30	-0.72	-0.42	-0.39	0.03	-0.09	-0.08	-0.50
		4	-0.82	-0.45	-0.34	-0.32	-0.48	-0.44	-0.81	-0.57	-0.16	-0.15	-0.13	-0.14
		5	-0.96	-0.55	-0.53	-0.75	-0.61	-0.74	-0.40	-0.96	-0.69	-0.52	-0.37	-0.30
		1	0.42	0.48	0.46	0.38	0.80	0.87	1.23	0.78	0.60	0.51	0.98	0.55
		2	0.26	0.44	0.26	0.30	0.23	0.08	0.23	0.11	0.14	0.11	0.21	0.10
	B62	3	-0.60	-0.14	0.06	-0.13	0.07	-0.06	-0.43	-0.02	-0.01	0.00	-0.13	-0.06
		4	-0.16	-0.39	-0.36	-0.16	-0.29	-0.19	-0.22	-0.22	-0.22	-0.25	-0.20	-0.26
		5	-0.75	-0.60	-0.62	-0.49	-0.47	-0.34	-0.17	-0.38	-0.39	-0.36	-0.32	-0.33
		1	0.48	0.52	0.52	0.51	0.56	0.63	0.95	0.57	0.46	0.44	0.53	0.50
		2	0.49	0.13	0.24	0.13	0.34	0.54	0.31	0.27	0.23	0.21	0.12	0.17
	B80	3	-0.10	0.06	0.11	0.00	-0.01	0.02	-0.77	-0.16	-0.09	-0.09	-0.14	-0.20
		4	-0.14	-0.25	-0.40	-0.27	-0.17	-0.27	-0.28	-0.33	-0.22	-0.20	-0.20	-0.21
		5	-0.73	-0.45	-0.47	-0.37	-0.73	-0.92	-0.42	-0.36	-0.37	-0.36	-0.31	-0.26