

Bioinformatics II: Alignment Dynamic Programming

Dr Manaf A Guma
University of Anbar- college of applied sciences-Heet
Department of chemistry

1

Before we start, lets remember!

- How many types of changes which can present between two sequence alignment?
 1. A mutation that replace one character with other.
 2. An insertion that adds one or more positions.
 3. A deletion that deletes one or more positions.

2

2

What is the Pair-wise sequence alignment?

Idea:

1. Display one sequence above another with spaces inserted in both to reveal similarity.
2. It does show how programs work dynamic.

```
A: C A T - T C A - C
    |   |       | |   |
B: C - T C G C A G C
```

3

3

Do you remember BLAST?

- What BLAST look for?
- It looks for the score of the alignment.
- What does it indicate?
- The BLAST Score indicates the quality of the best alignment between the query sequence and the found sequence (**hit**).
- What does high/ low score represent?
- The higher the score, the better the alignment.
- When the score of the alignment is reduced?
- Scores **are** reduced by mismatches and gaps in the best alignment

4

4

How many types of alignments? There are 2 types of alignment

S = CTGTCGCTGCACG
T = TGCCGTG

Global alignment
dealing with big
sequence of DNA

Local alignment:
dealing with small
segment of DNA

CTGTCG-CTGCACG
-TGC-CG-TG----

CTGTCGCTGCACG--
-----TGC-CGTG

5

5

How do you do Global alignment: Scoring?

CTGTCG-CTGCACG
-TGC-CG-TG----

Let's assume that
Reward for matches: α
Mismatch penalty: β
Space penalty: γ

$$\text{score}(A) = \alpha w - \beta x - \gamma y$$

$w = \# \text{matches}$ $x = \# \text{mismatches}$ $y = \# \text{spaces}$

6

6

Global alignment: Calculate Scoring?

Reward for matches: 10
 Mismatch penalty: 2
 Space penalty: 5

C	T	G	T	C	G	-	C	T	G	C
-	T	G	C	-	C	G	-	T	G	-
-5	10	10	-2	-5	-2	-5	-5	10	10	-5

Total = 11

7

7

Optimum Alignment

- What does the score measure?
- The score of an alignment is a measure of its quality.
- **How does it solve the problem of the alignment?**
- Given a pair of sequences X and Y , find an alignment (global or local) with maximum score.
- What does maximum score of an alignment represent?
- The **similarity** between X and Y , denoted $sim(X,Y)$, is the maximum score of an alignment of X and Y

8

8

Who designed the Alignment algorithms?

- Global: **Needleman-Wunsch**
- Local: **Smith-Waterman**
- NW and SW use *dynamic programming*.
- **How do the** algorithms express the variations of alignment?
 1. Gap penalty functions
 2. Scoring matrices

9

9

Follow the role to find the Global alignment in two sequencing?

Two seq. ATCG and TCG?
Scoring ?

- Match - 2
- Mis-match -2
- Gap -2

- Rules:
 1. Put the gap in the first.
 2. The higher is score the best because it shows the maximum alignment.
 3. The first box always zero.
 4. Add the initial column and row by adding gap value -2.
 5. To calculate the score, it will be the highest value in the diagonal.

10

- Once we finish, we need to do trace back to find the score.
- Trace back: we go back to find the highest value in the diagonal only.(the small score is, the best is)
- Then to express that:
- We start from the back to starting point which is the zero. It has to be a straight line to the zero.
- We need to do the alignment to find why the score is +1.
- ATCG ←
-TCG

11

11

Solve out this alignment?

G	-8				1. box beside +gap (-2)
C	-6				2. box bottom + gap (-2)
T	-4				3. Diagonal box (match /mismatch).
A	-2	-1	-3	-5	Match + 1
-	0	-2	-4	-6	Mis-match -1
	-	T	C	G	Gap -2

The direction mean the number it can come form this direction or from the other direction.

12

It is not complicated

<https://www.youtube.com/watch?v=vqxc2EfpWdk&t=1113s>

G	-8	10 7 5	4 4 2	4 +1 4
G	-6	5 5 3	5 5 0	2 2 2
T	-4	6 3 -1	2 2 -2	4 4 -4
A	-2	2 4 -1	1 2 -2	3 3 -3
O	O	-2	-4	-6
	O	T	C	G

13

13

Local Alignment

- We follow the same rules as same of the global alignment.
- Example:
- Find the local alignment of two seq:
- ATCG and TCG.

14

14

Solve out this alignment?

C	-6				
C	-4				
T	-2				
-	0	-2	-4	-6	-8
-	A	T	C	G	

Match + 1
 Mis-match -1
 Gap -2

. box beside +gap (-2)
 . box bottom + gap (-2)
 . Diagonal box (match /mismatch).

15

Find the score using local alignment matrices. For local alignment only: place any value (-) with zero, whenever!

C	0				
C	0				
T	0				
-	0	0	0	0	0
-	A	T	C	G	

Match + 1
 Mis-match -1
 Gap -2

. box beside +gap (-2)
 . box bottom + gap (-2)
 . Diagonal box (match /mismatch).

16

Go on for this alignment?

C	0	0	0	1	0	Score= 1
C	0	0	0	1	0	
T	0	0	1	0	0	Match + 1 Mis-match -1 Gap -2
-	0	0	0	0	0	
	-	A	T	C	G	box beside +gap (-2) box bottom + gap (-2) Diagonal box (match /mismatch).

17

Local alignment: Example

S = g g t c t g a g
T = a a a c g a

If you assume that

Match: +2

Mismatch and space: -1

Best local alignment:

g g t c t g a g
a a a c - g a -

Score = 5

18

18

Another method for alignment is called **End-gap free alignment**

- Gaps at the start or end of alignment are not penalized

$S = c a c t g t a c$
 $T = g a c a c t t g$

Match: +2

Mismatch and space: -1

Best global

Best end-gap free

c a c - - t - g t a c
g a c a c t t g - - -

- - c a c - t g t a c
g a c a c t t g - - -

Score = 1

Score = 9

19