

**Why is DP (Dynamic programming) table so useful and efficient?**

1. Simply by looking at the table, we can easily deduce what is the best alignment for each sequence.

Example

Q. We want to know the best alignment between ACCG and AC

		A	C	C	G
	0	-10	-20	-30	-40
A	-10	2	-8	-18	-28
C	-20	-8	4	-6	-16
G	-30	-18	-6	-3	-4

A. Simply by looking at the table, we can deduce the best alignment that this sequence may have.

2. The table excludes any hopeless score.

Example

Q. The user is confused whether ACCG\_\_\_ can be a good alignment or not.

\_\_\_ACG

A. This alignment has 3 gaps and one mismatch between G and A. According to the scoring matrix, this alignment would give score of  $-10 -10 -5 -10 = -35$

However, if we look at the table,

		A	C	C	G
	0	-10	-20	-30	-40
A	-10	2	-8	-18	-28
C	-20	-8	4	-6	-16
G	-30	-18	-6	-3	-4

We cannot see a score of -35. Then what does this indicate?

- It indicates that the alignment above can never be the best alignment. So the user can simply ignore that alignment above when he or she is looking for the best alignment.

3. We can simply guess the different form of decomposition.

-By looking at the direction, we understand that into which type of sequence that the certain sequence has been or can be decomposed into.

1) If the direction is toward horizontal, it indicates that the sequence was decomposed into smaller

sequence such that one base has been decomposed from the first sequence. → gives out  $F(\text{base}, \_)$

2) If the direction is toward vertical, it indicates that the sequence was decomposed into smaller sequence such that one base has been decomposed from the second sequence. → gives out  $F(\_, \text{base})$

3) If the direction is toward diagonal, it indicates that the sequence was decomposed into smaller sequence such that one base has been decomposed from each sequence. → gives out  $F(\text{base}, \text{base})$

### Local alignment vs Global alignment

	Local alignment	Global alignment
Definition	Alignment that only finds similar components, locally similar regions (only align regions between gaps)	Alignment that finds from beginning to the end. So regardless of similarity, alignment is still produced. So in a context of biology, it uses all the gaps including insertions or deletions
Optimized score from the DP table	4 → If we only treat the similar components and only care about the 'local' region, 4 can be the optimized score from that specific local area in the DP table	-4 → since we need to care about the whole alignment, the optimized score would be -4 which is located at the very bottom and right side of the DP table.

### Some useful facts

Q. What would affect the alignment of two sequences using DP?

A. the scoring matrix

Q. How do we define the sequence similarity in this course?

A. The alignment score of the best alignment.

### Gene expression

Why do we need to study Gene expression?

- Even though there is not much variation within the human genome, we are very different in terms of 'phenotype'.
- The gene expression difference may account for the phenotype difference, therefore it is very important to study gene expression.

### Gene expression matrix

Q. How do we interpret this?

A. By comparing color. (More red indicates that gene expression value is high in that certain circumstance whereas green indicates lower gene expression)

Q. How do we know the gene expression level

A.

- It is very difficult to do the protein sequencing. So what we do is we play with RNA.
- More specifically, for our DNA to be translated to protein, it definitely needs go under transcription first. So we can count how many copies of corresponding RNA of that specific gene we have.
- High number of RNA indicates we have high gene expression level in that circumstance.

Q. How do we make gene expression matrix?

A.

- We first isolate RNA and using a high technology such as 'Illumina sequencing', then we gather information we need.
- Then we create gene count matrix based on the information we have, such as number of RNA reads.

Example

	Sample 1	Sample 2
Gene A	500	100
Gene B	250	300

- Then based on the gene count matrix we have, and using a lot of mathematical calculations, we get a 'Gene expression matrix'.

Example of gene expression matrix:

