Sequence Alignment: II

Dr Safdar Ali
Department of Biological Sciences
Aliah University, Kolkata
Date: 04/04/2020

Topics to be covered

- Sequence formats
- Dot Plot
- Global and Local alignment
- Scoring Matrices
- Needleman Wunsch Algorithm
- Smith Waterman Algorithm
- BLAST and FASTA
- HMM
- Multiple sequence alignment

Alignment of Sequences

1. Dot matrix analysis

2. The dynamic programming (or DP) algorithm

3. Word or *k-tuple methods, such as used by the programs FASTA and BLAST*

Dynamic Programming Algorithm

- Needleman and Wunsch (1970)
  - Global Alignment

- Smith and Waterman (1981)
  - Local Alignment

Needleman–Wunsch Algorithm

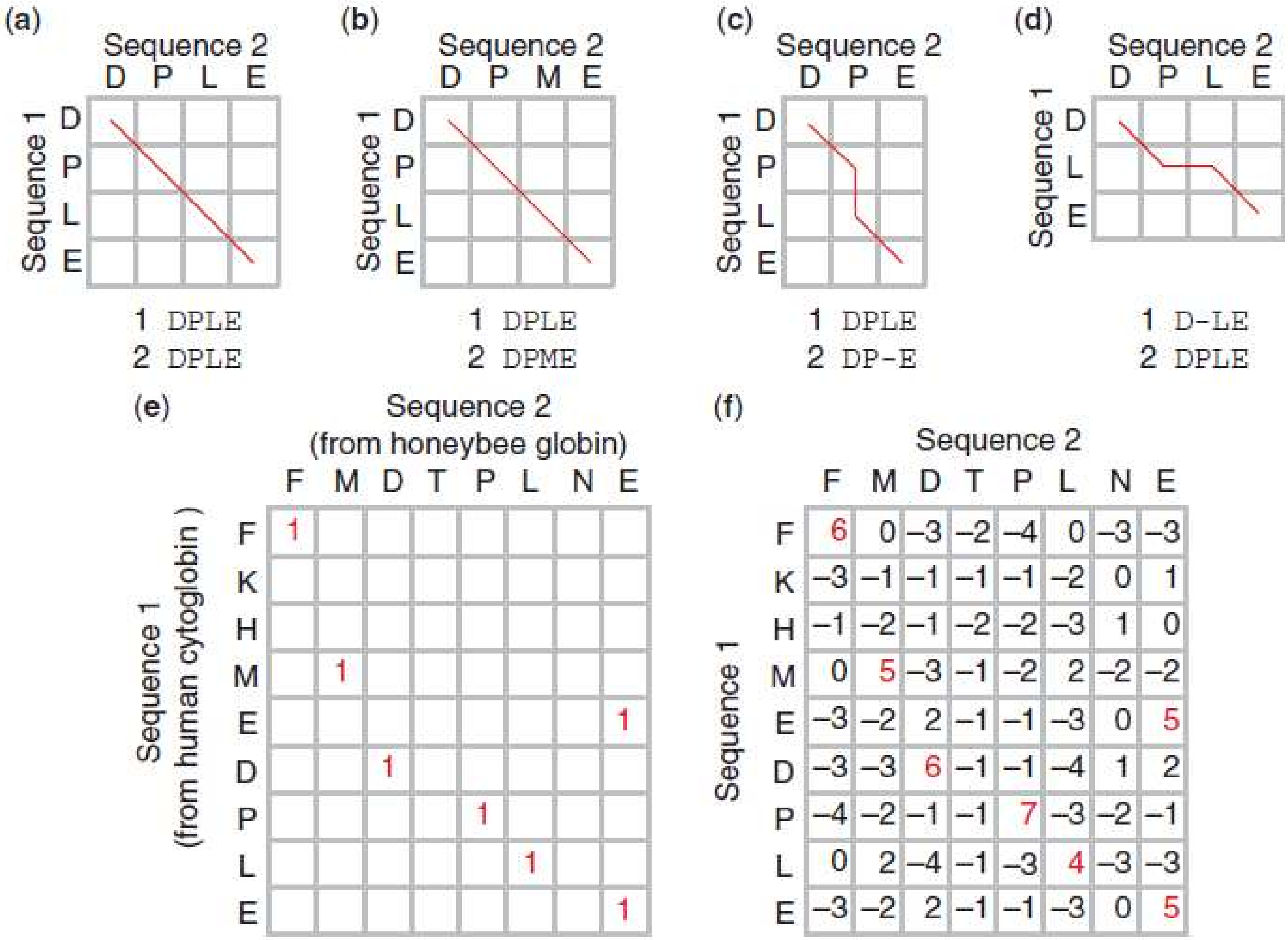
(1) Setting up a matrix

(2) Scoring the matrix

(3) Identifying the optimal alignment

Bioinformatics & Functional Genomics 3rd Edition, 2015,
Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

Setting up a matrix



Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

1. Two residues may be perfectly matched (i.e., identical)

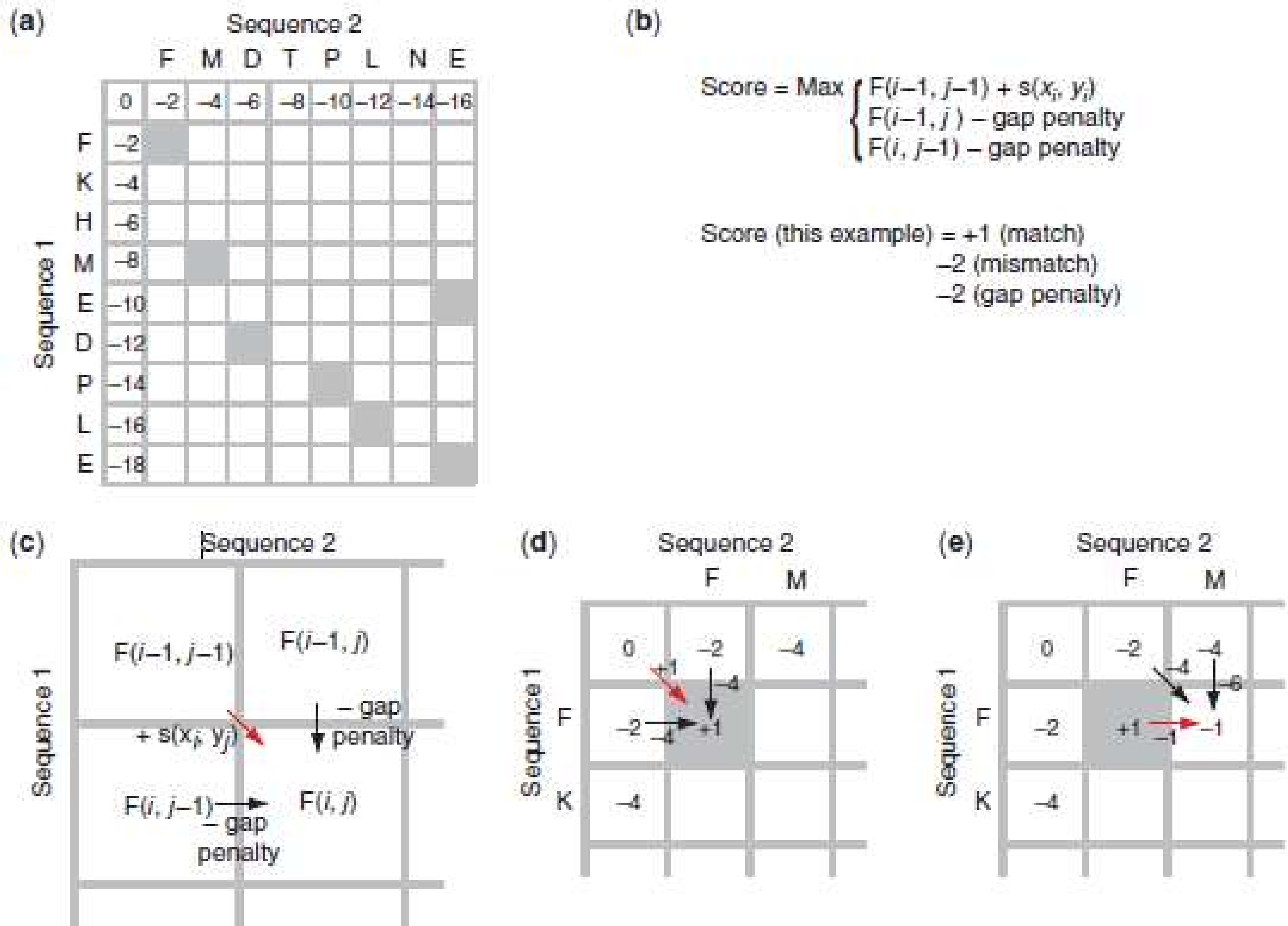
2. They may be mismatched

3. A gap may be introduced from the first sequence

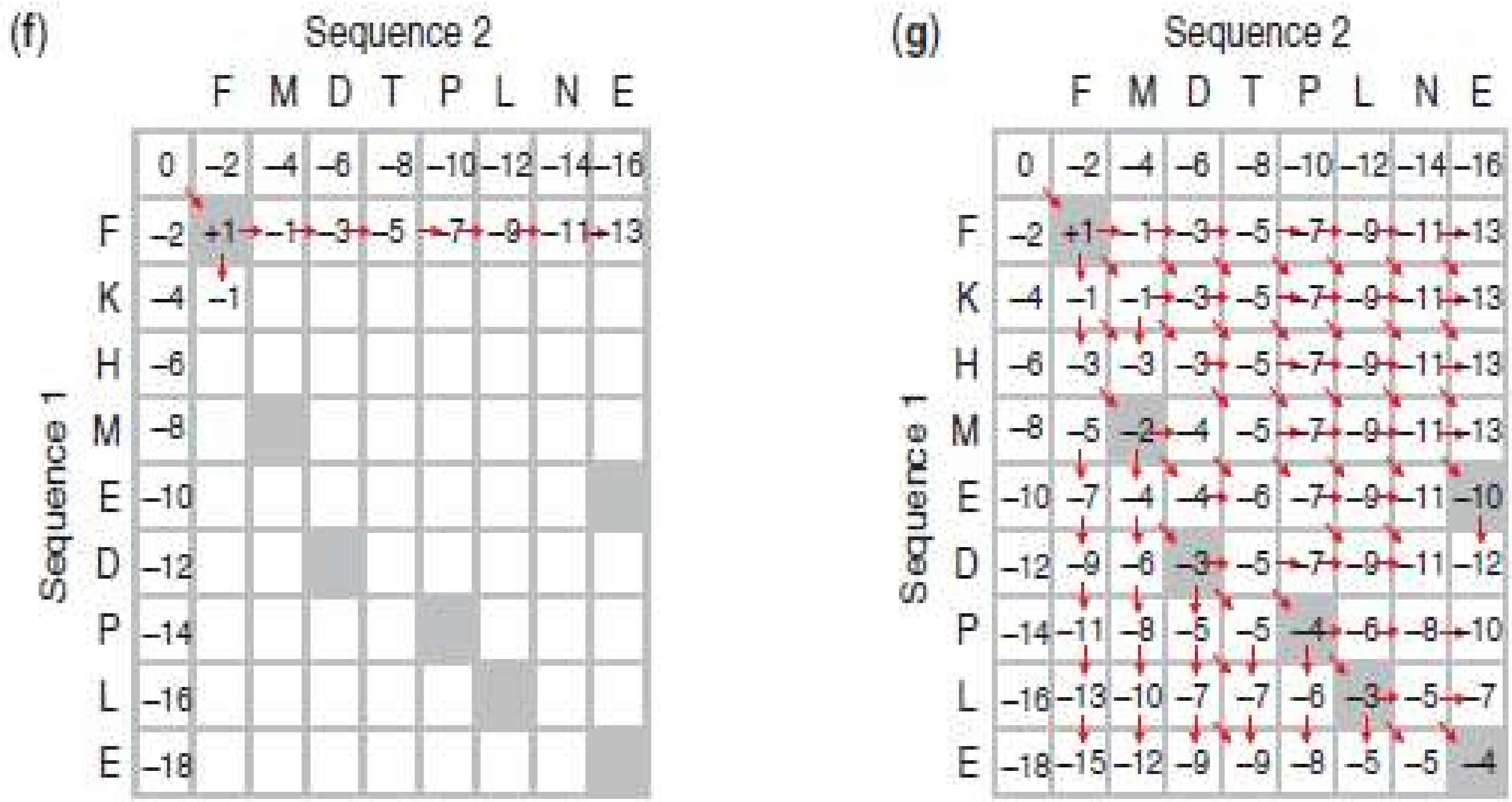
4. A gap may be introduced from the second sequence

Bioinformatics & Functional Genomics 3rd Edition, 2015,
Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

Scoring the Matrix

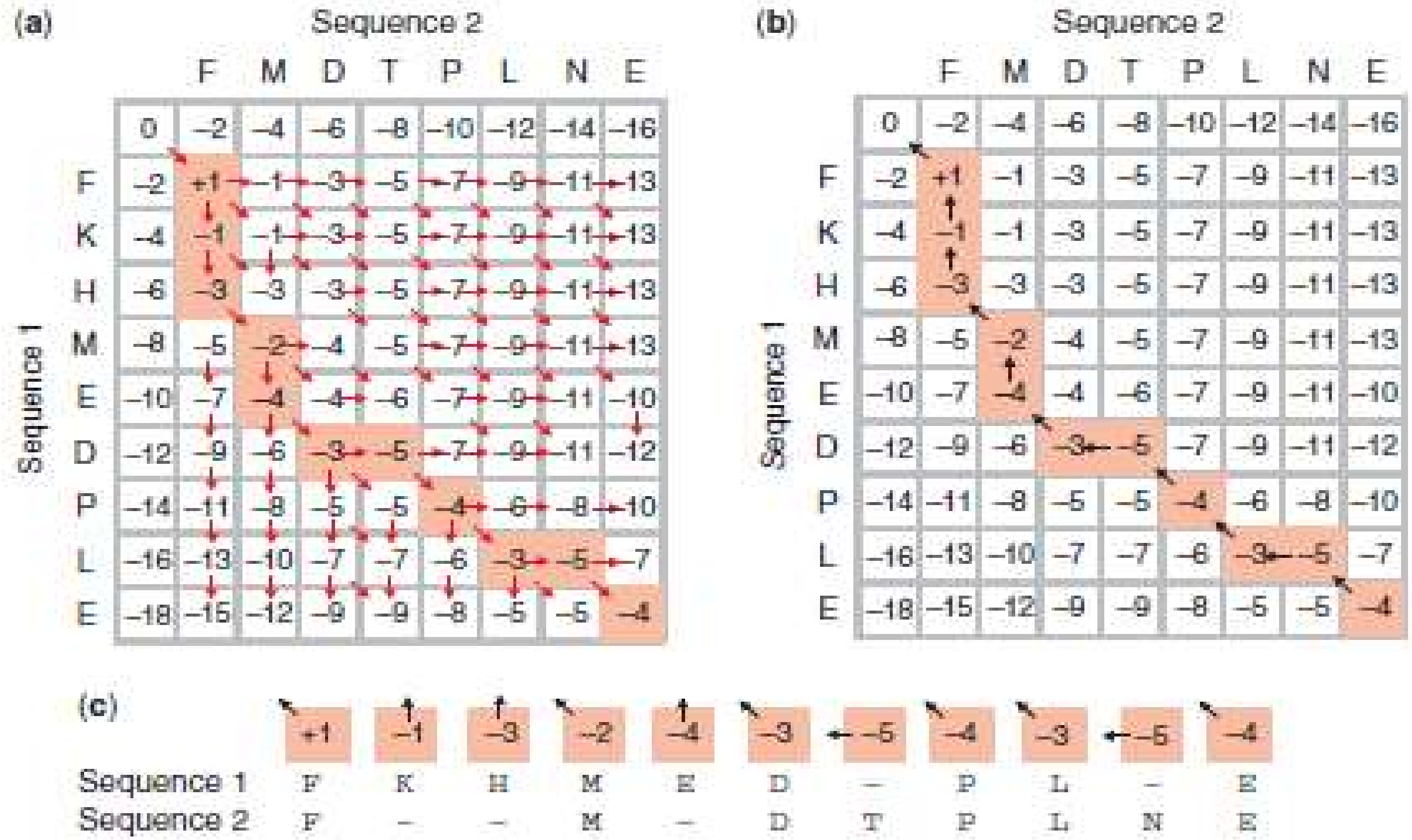


Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3



Bioinformatics & Functional Genomics 3rd Edition, 2015,
Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

Identifying the Optimal Alignment



Note that the final score of -4 equals sum of matches (6) and gaps (-10)

Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

Smith and Waterman Algorithm

Sequence 1

Sequence 2		C	A	G	C	C	U	C	G	C	U	U	A	G
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
A	0.0	0.0	1.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.7
U	0.0	0.0	0.0	0.7	0.3	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.7
G	0.0	0.0	0.0	1.0	0.3	0.0	0.0	0.7	1.0	0.0	0.0	0.7	0.7	1.0
C	0.0	1.0	0.0	0.0	2.0	1.3	0.3	1.0	0.3	2.0	0.7	0.3	0.3	0.3
C	0.0	1.0	0.7	0.0	1.0	3.0	1.7	1.3	1.0	1.3	1.7	0.3	0.0	0.0
A	0.0	0.0	2.0	0.7	0.3	1.7	2.7	1.3	1.0	0.7	1.0	1.3	1.3	0.0
U	0.0	0.0	0.7	1.7	0.3	1.3	2.7	2.3	1.0	0.7	1.7	2.0	1.0	1.0
U	0.0	0.0	0.3	0.3	1.3	1.0	2.3	2.3	2.0	0.7	1.7	2.7	1.7	1.0
G	0.0	0.0	0.0	1.3	0.0	1.0	1.0	2.0	3.3	2.0	1.7	1.3	2.3	2.7
A	0.0	0.0	1.0	0.0	1.0	0.3	0.7	0.7	2.0	3.0	1.7	1.3	2.3	2.0
C	0.0	1.0	0.0	0.7	1.0	2.0	0.7	1.7	1.7	3.0	2.7	1.3	1.0	2.0
G	0.0	0.0	0.7	1.0	0.3	0.7	1.7	0.3	2.7	1.7	2.7	2.3	1.0	2.0
G	0.0	0.0	0.0	1.7	0.7	0.3	0.3	1.3	1.3	2.3	1.3	2.3	2.0	2.0

Bioinformatics & Functional Genomics 3rd Edition, 2015,
Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

1. The score from the cell at position $i-1, j-1$; that is, the score diagonally up to the left. To this score, add the new score at position $s[i, j]$, which consists of either a match or a mismatch.
2. $s(i, j-1)$ (i.e., the score one cell to the left) minus a gap penalty.
3. $s(i-1, j)$ (i.e., the score immediately above the new cell) minus a gap penalty.
4. The number zero. This condition assures that there are no negative values in the matrix. In contrast negative numbers commonly occur in global alignments because of gap or mismatch penalties (note the log-odds matrices in this chapter).

Bioinformatics & Functional Genomics 3rd Edition, 2015, Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

Global vs Local

(b)

```
sequence 1   GCC-UCG
sequence 2   GCCAUUG
```

(c)

```
sequence 1   CAGCC-UCGCUUAG
sequence 2   AAUGCCAUUGACGG
```

Bioinformatics & Functional Genomics 3rd Edition, 2015,
Jonathon Pevsner, Wily and Blackwell Publishing, Chapter 3

Thank You