

Dynamic Programming

Jianlin Cheng, PhD

Computer Science Department

University of Missouri, Columbia

Fall, 2014

Project

- Align two protein sequences using BLOSUM matrix
- Maximize alignment score and print out the optimal alignments

BLOSUM62 Matrix

Ala	4																				
Arg	-1	5																			
Asn	-2	0	6																		
Asp	-2	-2	1	6																	
Cys	0	-3	-3	-3	9																
Gln	-1	1	0	0	-3	5															
Glu	-1	0	0	2	-4	2	5														
Gly	0	-2	0	-1	-3	-2	-2	6													
His	-2	0	1	-1	-3	0	0	-2	8												
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4											
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4										
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5									
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5								
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6							
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7						
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4					
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5				
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11			
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7		
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	
Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val		

Gap penalty: 5

An Example

VVFNLK

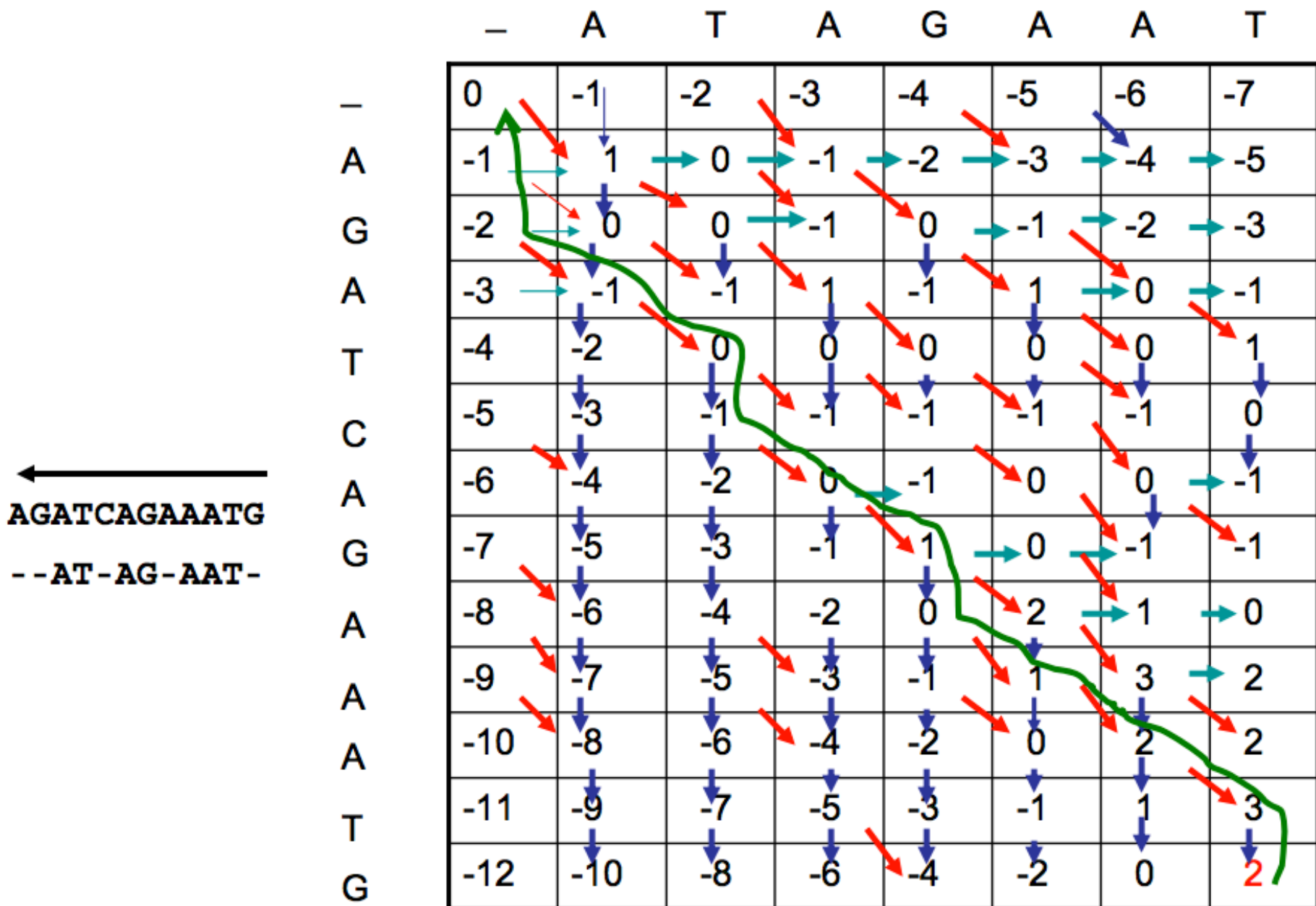
YV-QFK

$$1 + 4 - 5 + 0 + 0 + 5 = 5$$

Part I – Global Alignment

$$M[i,j] = \max [\begin{array}{l} M[i-1,j-1] + S(P[i],Q[j]), \\ M[i,j-1] + S(-, Q[j]) \\ M[i-1,j] + S(P[i], -) \end{array}]$$

3. Trace Back



Two Homologous Sequences

>4J8QA

DGYSLGDIAVDWATVRVVGDDTYSLNADRWGTLWPAATAIPF
YKPIDGQQRVITYFNPLYDNYEGYDHAVKVEHNYNVLTQVEDL
TAENESEFGNDPVWVNKDWIGGGYLVNIFRQNLKHLVSLVRD
RATAAEGEDDGYIHLELRYKTYDDQANGAVSFNLNSLDLTGKK
GIKVKLNSVKDGETEVVFNLK

>3K0Y

DIWIAVATVVPEGNVYYLRLDDGDKLWPAATNYPNYQPKN
QRALVNFTILADSHYIKVNAIHNILTKSIAKNEGAANDSIYGTDP
VSIYNNNWIGDGYLNIYFETLWGGKTAHFILIQPDAENDPYTL
EFRHNAYDDPQYTIGAGRVAFNLSSLPDTKGETVDLVVNYWTS
EGKQAYKLYN

Second Part of Project

- The first part is not challenging enough.
- So, develop an algorithm to do local sequence alignment

Local vs. Global Alignment

- Global Alignment

```
--T--CC-C-AGT--TATGT-CAGGGGACACG-A-GCATGCAGA-GAC
|  || |  ||  |  |  |  |||  |  |  |  |  |  ||||  |
AATTGCCGCC-GTCGT-T-TTCAG----CA-GTTATG-T-CAGAT--C
```

- Local Alignment—better alignment to find conserved segment

Transcription binding site

```
          ↗
tccCAGTTATGTCAGggggacacgagcatgcagagac
| | | | | | | | | |
```

```
aattgccgccgtcgttttcagCAGTTATGTCAGatc
```

Second Part of Project

- Where to change in the previous global alignment algorithm?

For j in 1..n (column)

 for i in 1..m (row)

$M[i,j] = \max(0, (M[i-1,j-1]+S(i,j), M[i,j-1]+S(-,j), M[i-1, j] + S(i,-))$

Discussion

- Input: two protein sequences; output: score and optimal alignment (global, local)
- Updating rule
- Algorithm
- Time and space complexity
- Implementation
- Difference between global and local
- Evaluation / Comparison (other alignment tools)
- Visualization of optimization process
- Challenges and solutions
- Task assignments