

# **Dynamic Programming**

**Jianlin Cheng, PhD**

**Computer Science Department  
University of Missouri, Columbia**

**Fall, 2013**

# Project

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

# BLOSUM62 Matrix

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

**Gap penalty: -10**

# Two Homologous Sequences

- Glyma02g01040.1  
MVSQRQRLARKRYKAEHPELFPKPEPTPPKDPEKKKKKKNSAFKRKRPEP  
KPGSRKRHPLRVPGMKPGESCFICKAMDHIAKLCPEKAEWEKNKICLRCRR  
RGHRAKNCPLEVLDGAKDAMYCYNCENGHSLTQCPHPLQEGGTKFAECF  
VCNQRGHLSKNCQPQNTHGIYPKGCCCKICGGVTHLAKDADKSGVAA  
NRPADGWMRIERPMGQVTKFVSGDDIEDDFMTDDIHSGDKKPAKSTE  
DHVKPKKKEGPKVVNF
- Glyma16g23370.1  
QVKKHSKHCRTCNRCVEGFDHHCRWLNNCVGKRNYTAFFLLMIFLLIKGG  
TAIAIFIRCFVDRRGIEKELQRKLYVDFPRGVLATICVFLLLTA  
YSSAALGQLFFFHVVVLIRKTNTHAIKSIIISLRKTYDYILAMKEENEAMELESFDDSDFSSDES  
DFDSPEKPTLMSGFLCKGNQGKKALLAAEKARERIMREKPMGEHNSLKPL  
PLETKCGPLMNTYKNMDTEDFGSTSFIKGRLNESPGRFSSPRRRFSAGSPT  
VFSSSMMMASPHHKYRSSFDLKLTVSRELETHISRQVLCSVISKDDSEPSR

# Discussion

- Input: two protein sequences; output: score and optimal alignment
- Subproblem
- Updating rule
- Algorithm
- Time and space complexity
- Implementation
- Evaluation
- Visualization of optimization process
- Challenges and solutions
- Task assignments