Dynamic Programming

Jianlin Cheng, PhD
Computer Science Department
University of Missouri, Columbia
Fall, 2014
• Align two protein sequences using BLOSUM matrix
• Maximize alignment score and print out the optimal alignments
BLOSUM62 Matrix

Gap penalty: 5
An Example

VVFNKLK
YV - QFK

1 + 4 - 5 + 0 + 0 + 5 = 5
$M[i,j] = \max \left[ \begin{array}{l}
M[i-1,j-1] + S(P[i],Q[j]),
M[i,j-1] + S(-, Q[j]),
M[i-1,j] + S(P[i], -)
\end{array} \right]$
3. Trace Back

<table>
<thead>
<tr>
<th></th>
<th>-</th>
<th>A</th>
<th>T</th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>A</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-4</td>
<td>-5</td>
<td>-6</td>
<td>-7</td>
</tr>
<tr>
<td>A</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
<td>-4</td>
<td>-5</td>
</tr>
<tr>
<td>G</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>-3</td>
</tr>
<tr>
<td>A</td>
<td>-3</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>C</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>-5</td>
<td>-3</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>-6</td>
<td>-4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>G</td>
<td>-7</td>
<td>-5</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>A</td>
<td>-8</td>
<td>-6</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>T</td>
<td>-9</td>
<td>-7</td>
<td>-5</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>-10</td>
<td>-8</td>
<td>-6</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>-11</td>
<td>-9</td>
<td>-7</td>
<td>-5</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>T</td>
<td>-12</td>
<td>-10</td>
<td>-8</td>
<td>-6</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

AGATCAGAAATG
--AT-AG-AAT--
Two Homologous Sequences

>4J8QA
DGYSGLGDIAVDWATVRVVGGGDTYSNLNADDRWGTWLWPAATAIPFYKPIDGQRVITYFNPLYDNYEGYDHAVKVEHNYNVLTQKVLEDLTRANESEFGNDPWWVNKDWIGGGYLNVIIFRQNLKHLVSLVRSRAATAEGDDGYIHLELRYKTYDDQANGAVSFNLNSLDLTDGKKGIKVKLNSVKGGETERVVFNLK

>3K0Y
DIWIAVATTVPEGNNVYYLRLDDGDKLWPAATNYPNYPKPNQRALVNFTILADSHYIKVNAIHNILTKSIAKNEGAANDSIVTDPVSIYNNNWIGDGYNIIYFETLWGGKTAFHFINLIQPDAENDPYTLEFRHNAYDDPQYTIGAGRVAFFNLSSLPDFTKGTVDLDVVNYWTSEGKQAYKLKYN
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
```

```
AATTGCGC GCC-GTCGT-T-TTCAG----CA-GTTATG-T-CAGAT---C
```

- **Local Alignment**—better alignment to find conserved segment

```
tccCAGTTATGTCA Ggggacacgagcatgcagagac
```

```
attgccgccggctcgcgttttcagCAGTTATGTCAGatc
```

Transcription binding site
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