**LOCAL ALIGNMENT** : Pairwise SEQ. Align.

**GLOBAL**

\[
\begin{align*}
X & \quad Y \\
\cdots & \\
\end{align*}
\]

**LOCAL**

\[
\begin{align*}
X & \quad Y \\
\end{align*}
\]

---

Consider all substrings of the X seq \( \alpha \) and all the substrings of the Y seq \( \beta \). Then align globally \( \alpha \) and \( \beta \).

The local optimal alignment is one of max score.

Substrings = contiguous subsequences

\[ACTAAASGT\]
Suerose

Subsequence

\[
\text{ACTAAA} \quad \text{GT} \quad \text{distinct subseq.}
\]

In local align we consider global aligns between contiguous subseq of \(X\) with contiguous subseq of \(Y\).

\[
\begin{array}{cccc}
X & & \cdots & N \\
Y & & \cdots & N \\
\end{array}
\]

\(O(m^2), \Theta(n^2)\)

all against all

\(O(m^2n^2)\) global aligns

Amazing: Finding optimal (max)
Local alignment: $O(MN)$

same time complexity as for global alignment.

"D" zero

**GLOBAL ALIGNMENT**

**INPUT:** Two sequences $X, Y$

and a scoring scheme $S$

**OUTPUT:**

MAX

GLOBAL OPTIMAL

GLOBAL ALIGNMENT

OF $X$ and $Y$ under

scoring scheme $S$

The Global Alignment Algorithm works with every scoring scheme.

```
ACGT
scoring A/C 11
```


\[ S_1 \]

\[ S_2 = S_1 + \text{add 27 to all entries} \]

For local alignment the Smith-Waterman Scheme is very special

**Biology**

The Smith-Waterman Arg

\( X, Y \) input seqs \& \( \delta \) the scoring scheme

\( X = \text{AACCTG} \quad Y = \text{AAGTGG} \)

prefix, suffix

Prefix of \( X = \{ \lambda, A, AA, A-AC, A-ACT, A-ACTG \} \)

\( \lambda = \text{empty} \)
$\text{Suffixes of } X = \{ TCG, ACTG, AACTG \}$

$\text{Substrings of } X = \{ \text{Suffixes U, Prefixes U} \}$

$\text{Suffices (Prefixes (X))} = \text{ACTG}$

$\text{Suffices (Suffices (X))} = \text{ACTG}$

$\text{The Local Alignment PE}$

Given $X, Y, \delta$
Find $X, Y$ such that $\gamma$
\[ \begin{align*}
\text{substring of } x \\
\beta & \text{ substring of } y \\
\text{s.t. } \text{global align } x \text{ and } \beta \\
& \text{ is max among all choices of } x \text{ and } \beta.
\end{align*} \]

\[ v^* = \text{optimal local alignment score (i.e. global align of)} \]
\[ \max \text{ cost of } x \text{ and } \beta \]

\[ x \]

\[ Y \]

\[ Y_i, Y_j \text{ prefixes of } x \text{ and } y \]
Suffixes (Prefixes) = All suffixes
v*(i, j) = value of the optimal global alignment between suffixes of x_i and y_j

\[ V^* = \max \left\{ V(i, j), 0 \leq i \leq N, 0 \leq j \leq M \right\} \]

Proof:
1) \( V^* \geq \max \left\{ V(i, j), \ldots \right\} \)
   by def of \( V^* \)

2) \( V^* \leq \max \left\{ V(i, j), \ldots \right\} \)
   Indeed
   \[ a^* \]
\[
V^* = V(i^*, j^*) \leq \text{MAX} \left\{ V(i, j) \right\}
\]

The Smith-Waterman Alg

The Local Align Alg

\[
V(i, 0) = V(0, j) = 0, \quad \forall i, j \\
\text{FOR } i > 0, j > 0 \text{ DO} \\
\quad V(i, j) = \text{max} \left[ 0, \\
\quad \quad V(i-1, j-1) + \delta(x_i, x_j), \\
\quad \quad V(i-1, j) + \delta(x_i, -), \\
\quad \quad V(i, j-1) + \delta(-, x_j) \right]
\]
\( \sum_{(i,j) \neq (-1,1)} + \delta(-1,1) \)

PAM

BLOSUM

3 Pillars:

- Computer Science
- Statistics
- Biology = Math Models of Evolution

Margaret Dayhoff

BLOSUM (Henikoff & Henikoff 1992)
20 letter alphabet for protein sequences

20 amino acids (capital letters)

WWYtR
WFYtR
WYYtR
WYYtR
WFYKR

Example:

BABA
AAAC
AACc
AABA
AACc
AABC

Think about for a column think of a pair of rows as a substitution

"block"

multiple gap-less align
of one amino acid (letter) into another

24 amino acids observed
4 × 6
14 are A
4 are B
6 are C

<table>
<thead>
<tr>
<th>aa</th>
<th>freq obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>14</td>
</tr>
<tr>
<td>B</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
</tr>
</tbody>
</table>

4. \( \binom{6}{2} = 60 \) aligned pairs
<table>
<thead>
<tr>
<th>A to A</th>
<th>( \frac{26}{60} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A to B</td>
<td>( \frac{8}{60} )</td>
</tr>
<tr>
<td>A to C</td>
<td>( \frac{10}{60} )</td>
</tr>
<tr>
<td>B to B</td>
<td>( \frac{3}{60} )</td>
</tr>
<tr>
<td>B to C</td>
<td>( \frac{6}{60} )</td>
</tr>
<tr>
<td>C to C</td>
<td>( \frac{7}{60} )</td>
</tr>
</tbody>
</table>

**Result:**

\[ P_{AB} = \frac{3}{20} \]

\[ P_{AC} = \frac{14}{24} \]

\[ P_{BC} = \frac{14}{24} \]

\[ P_{\text{total}} = 0.6 \]
SCORING SCHEME
Hypothesis Testing
Likelihood Ratio

\[ 2 \log_2 \left( \frac{\text{observed}}{\text{expected}} \right) \]

BLOSUM

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>B</td>
<td>-1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>-2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

A to A: 0.70
A to B: -1.09
A to C: -1