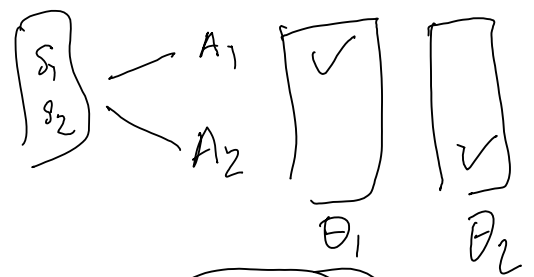


Scoring scheme: Gap model

Wednesday, February 3, 2021 11:58 AM



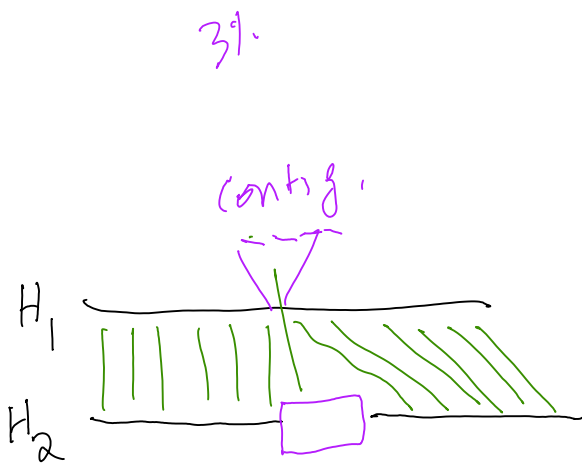
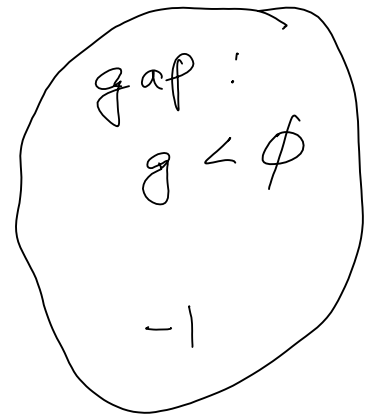
Parameters:

match:
 $m_a > \phi$

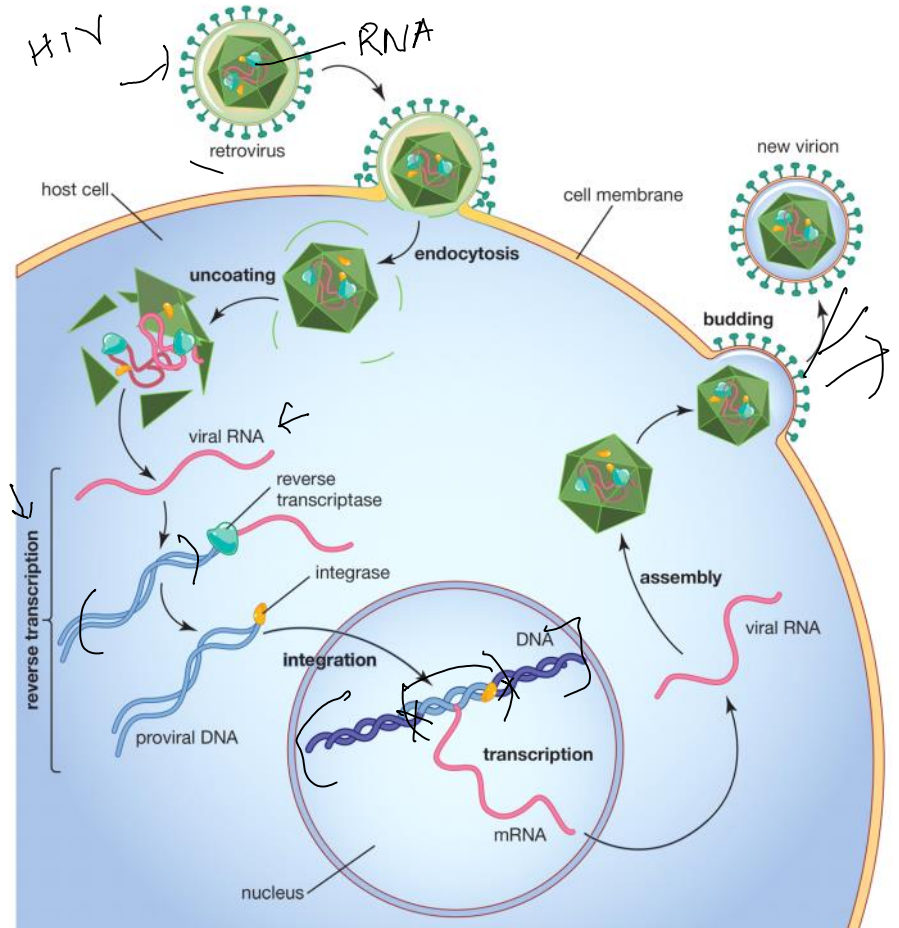
$\theta_1: +1$

mismatch:
 $m_i < \phi$

-2



Retrovirus infection and reverse transcription



© 2012 Encyclopædia Britannica, Inc.

Affine Gap Model (vs. linear gap model)

Friday, February 5, 2021 10:14 AM

Biological Expectation: Gaps tend to occur in "blocks" (contiguously) rather than scattered.

An example biological context: sets of transposons / transposable elements

Current scoring scheme: (linear gap model)
 match ($m_a > \phi$), mismatch ($m_i < \phi$), gap ($g < \phi$)

An example:

Two alignments between

The same two sequences:

Alignment A:

S_1 : g c a a a a a a a g t
 S_2 : g c a - a - a - a - g t

Score: $8 \times m_a + 4 \times g$

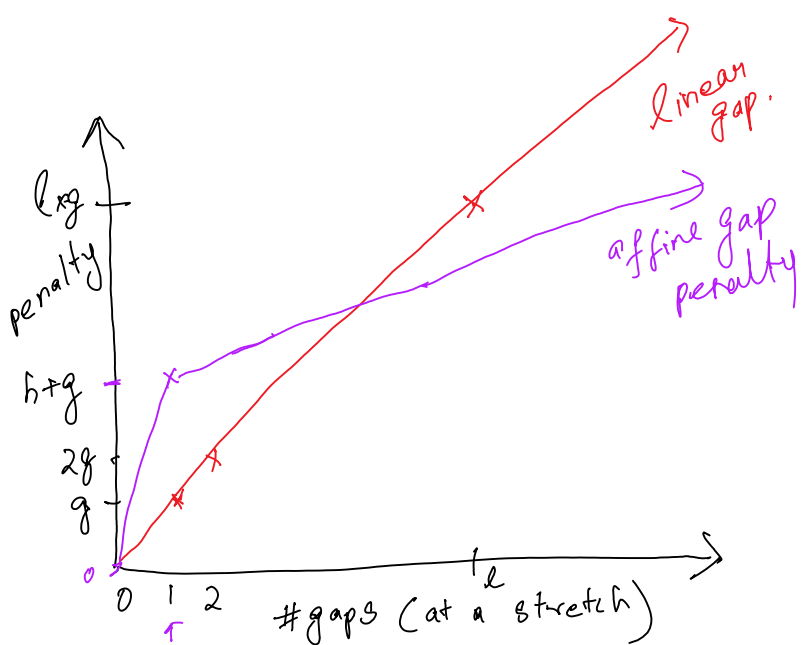
(Affine) score = $8 \times m_a + 4(h + g) <$

Alignment B:

S_1 : g c a a a a a a a g t
 S_2 : g c - - - - a a a a g t

Score: $8 \times m_a + 4 \times g$

(Affine) score = $8 \times m_a + h + 4g$



h : opening gap penalty
 g : continuing gap penalty

$h+g$ g g g g g
 a c c a t c

gap opening
 gap continuation

$h = -5$
 $g = -1$

$h + 6g$

Alignments with Affine Gaps

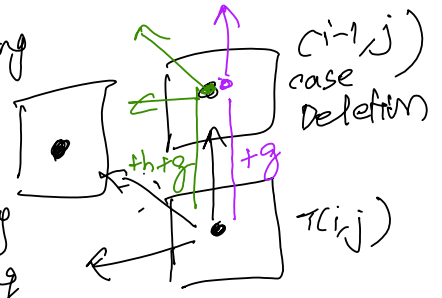
Friday, February 5, 2021 10:14 AM

$$\left\{ \begin{array}{l} \text{Global} \\ \text{Local} \end{array} \right\} \times \left\{ \begin{array}{l} \text{Linear} \\ \text{Affine} \end{array} \right\}$$

Main recurrence:

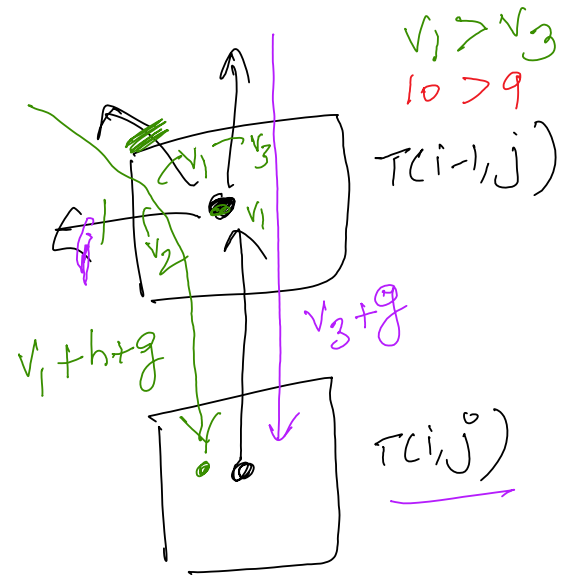
$$TC(i, j) = \max \left\{ \begin{array}{l} TC(i-1, j-1) + \beta(a_i, b_j) \checkmark \\ TC(i-1, j) + \begin{cases} h+g, & \text{if opening} \\ g, & \text{continuing} \end{cases} \\ TC(i, j-1) + \begin{cases} h+g, & \text{if opening} \\ g, & \text{continuing} \end{cases} \end{array} \right.$$

h : opening
 g : continuation



Quick fix
will it work?
⇒ will not guarantee optimality

S_1



⇒ We need a different algorithmic strategy!

$v_1 > v_3$
but $v_1 + h + g < v_3 + g$

Optimal alignments under the Affine Gap model (Gotoh'83)

Friday, February 5, 2021 10:14 AM

store all three values at every cell

alignment ending

let $S(i, j) \leftarrow \text{opt. score} \begin{pmatrix} s_1 [1 \dots i] \\ s_2 [1 \dots j] \end{pmatrix}$ s.t. $\begin{pmatrix} a_i \\ b_j \end{pmatrix}$

$D(i, j) \leftarrow$ " s.t. $\begin{pmatrix} a_i \\ - \end{pmatrix}$

$I(i, j) \leftarrow$ " s.t. $\begin{pmatrix} - \\ b_j \end{pmatrix}$

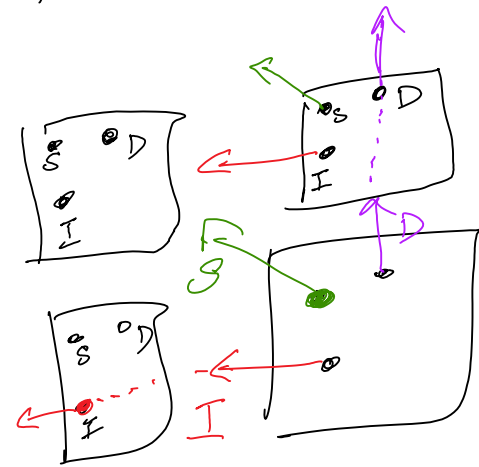
Recurrences:

$$S(i, j) = \max \begin{cases} S(i-1, j-1) \\ D(i-1, j-1) \\ I(i-1, j-1) \end{cases} + b(a_i, b_j)$$

$$D(i, j) = \max \begin{cases} D(i-1, j) + g \\ S(i-1, j) + h + g \\ I(i-1, j) + h + g \end{cases}$$

$$I(i, j) = \max \begin{cases} I(i, j-1) + g \\ S(i, j-1) + h + g \\ D(i, j-1) + h + g \end{cases}$$

store all three values at (i, j)



Initialization:

$$\left. \begin{array}{l} S(0, 0) = 0 \\ S(i, 0) = -\infty \\ S(0, j) = -\infty \end{array} \right\} \begin{array}{l} D(0, 0) = 0 \\ D(i, 0) = h + i \times g \\ D(0, j) = -\infty \end{array} \left| \begin{array}{l} I(0, 0) = 0 \\ I(i, 0) = -\infty \\ I(0, j) = h + j \times g \end{array} \right.$$

Retrace procedure (for affine gaps)

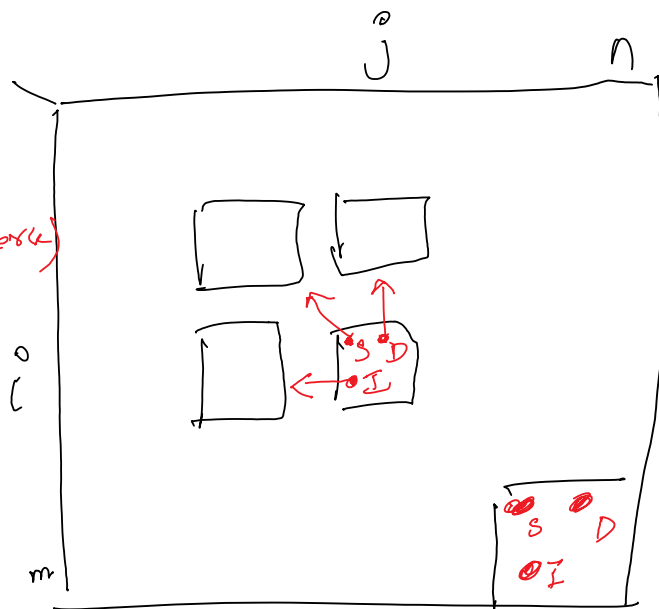
Monday, February 8, 2021 9:07 AM

Illustration for the Global Alignment case:
(Local align. case is similar)

Steps: // init: Path P (in reverse)

1) Start at the "last" cell (m, n)

Set direction: $next \leftarrow$ whichever case is the max for cell (m, n)



At each cell (i, j) of trace back:

if $next == D$

- Add (a_i) to Path P .
- $i \leftarrow i-1$
- $next \leftarrow$ whichever case gave max for $D(i, j)$

if $next == S$

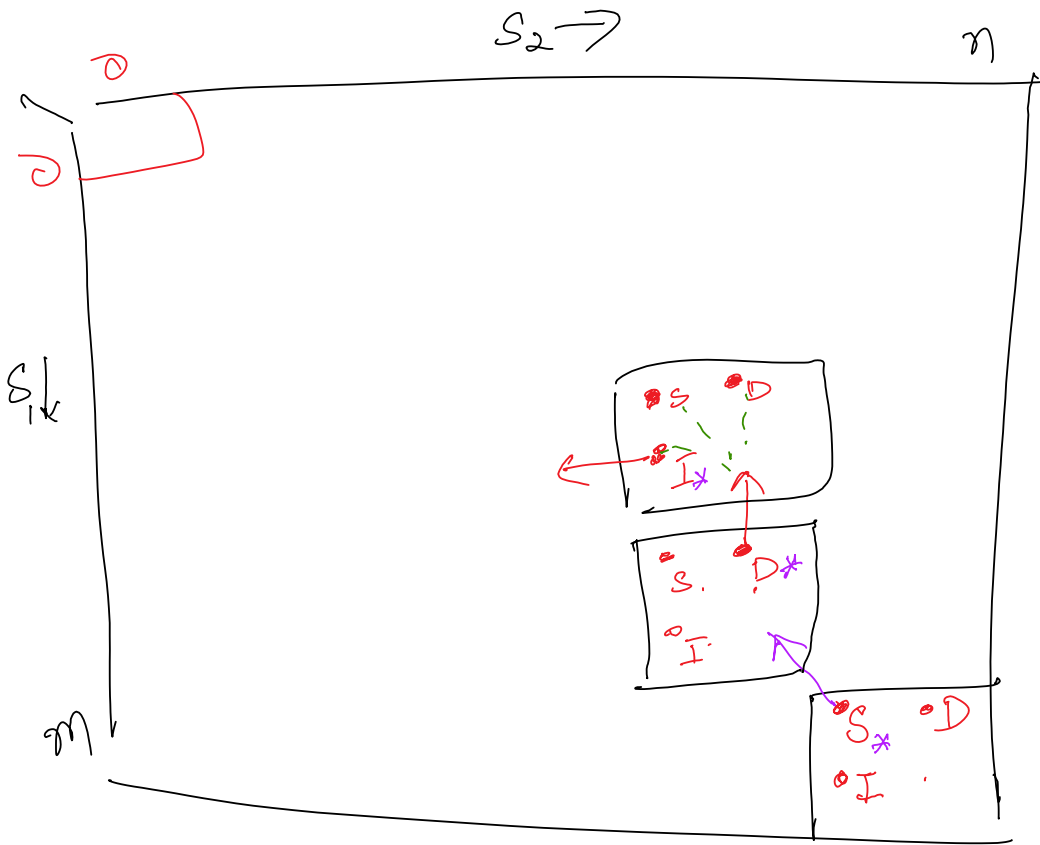
- Add (a_i, b_j) to Path P .
- $i \leftarrow i-1 ; j \leftarrow j-1$
- $next \leftarrow$ whichever case gave max for $S(i, j)$

if $next == I$

- Add (b_j) to Path P .
- $j \leftarrow j-1$
- $next \leftarrow$ whichever case gave max for $I(i, j)$

Retrace procedure (for affine gaps)

Monday, February 8, 2021 9:07 AM



(For global alignment)

1. Start at cell (m, n)
Find best case $(S, D \text{ or } I)$
 2. At each step: (i, j)
 - a. Find the case that yielded the current cell's case (along opt. path)
 - b. Retrace to that predecessor
- (i.e., Return the occurrence)

Opt Path \mathcal{P} :

