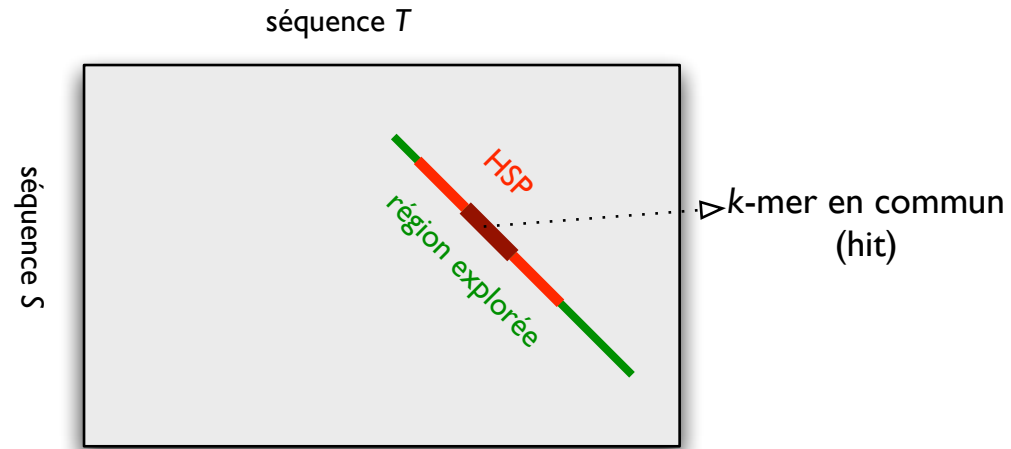


# EXTENSION D'UN HIT

Techniques :

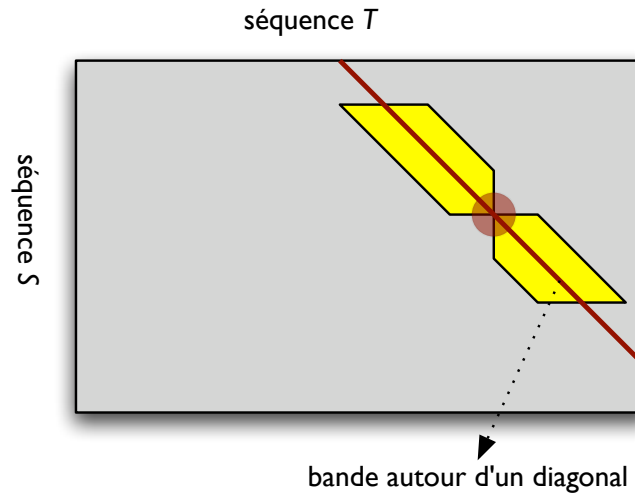
- extension rapide sur une diagonale
- X-drop
- programmation dynamique dans une bande

# EXTENSION RAPIDE



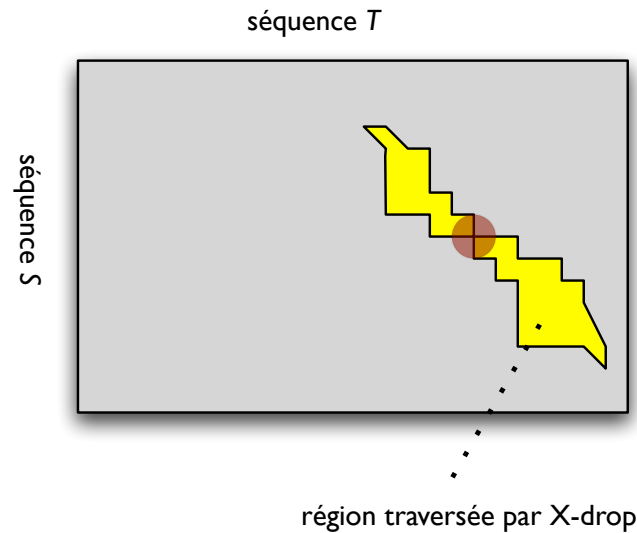
Rester sur la même diagonale ; explorer jusqu'à ce que le score devienne 0, prendre le meilleur segment (*high-scoring segment pair*, HSP)

# PD DANS UNE BANDE



Bande de  $\pm d$  sommets proches de la diagonale  $D : \{v_{i,j} : |(i - j) - D| \leq d\}$

# X-DROP



à partir d'une case initiale, explorer vers  $v_{0,0}$  et  $v_{|S|,|T|}$  ; arrêter si le score tombe par  $X$

(exploration de toute une région ou quelques [même 1] diagonales)

Altschul et al, *Nucleic Acids Res.* 25 : 3389.

## DÉTAILS : EXTENSION RAPIDE [VERS SUD-EST]

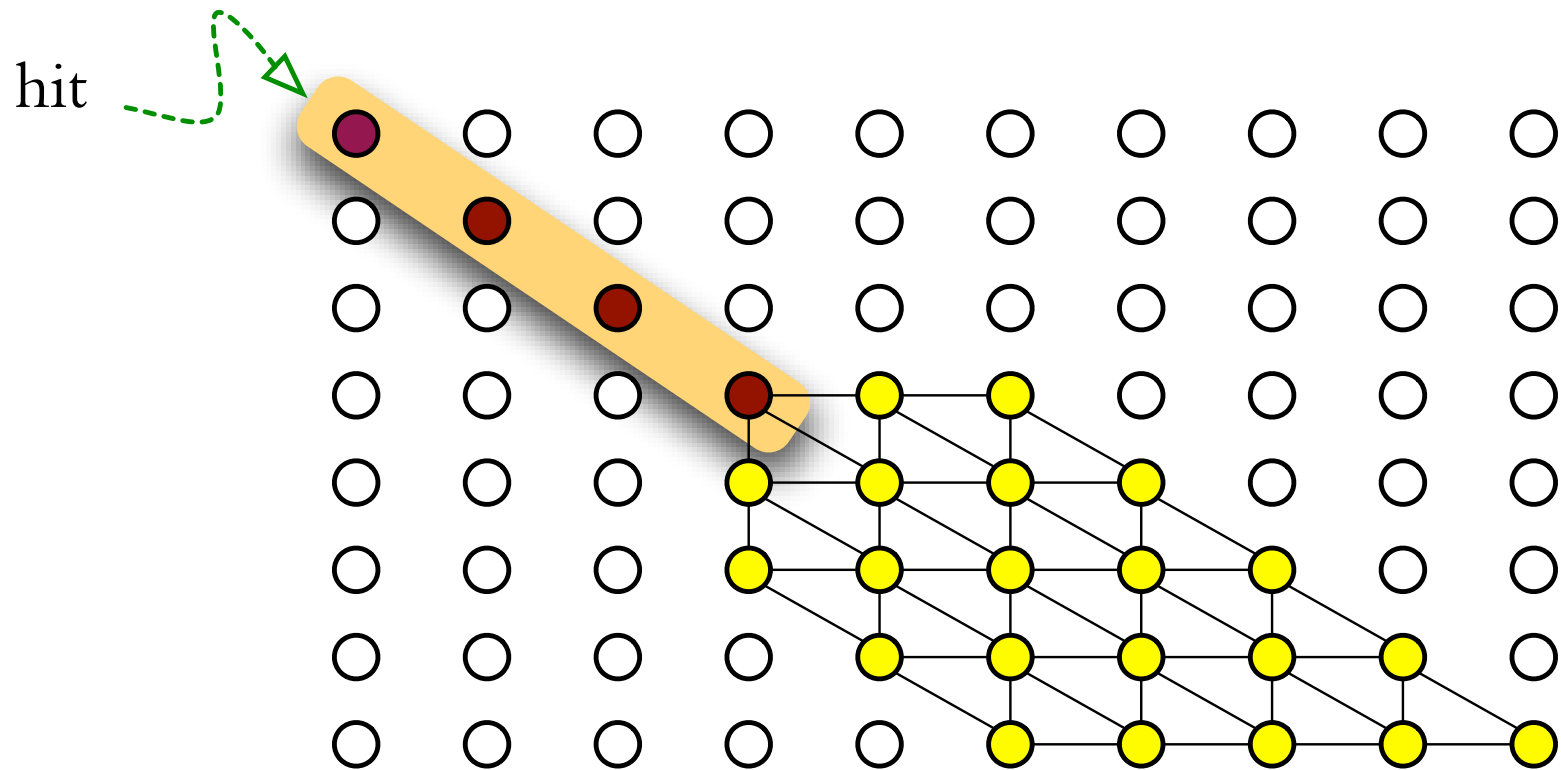
ER1 **Entrée**  $i_0, j_0$  départ de l'extension,  $s_0$  score initial  
ER2 meilleur  $\leftarrow s_0$ ; extension  $\leftarrow 0$   
ER3  $i \leftarrow i_0 + 1$ ;  $j \leftarrow j_0 + 1$ ; score  $\leftarrow s_0$   
ER4 **tant que**  $i \leq |S|, j \leq |T|, \text{score} \geq 0$   
ER5       score  $\leftarrow \text{score} + \mathbf{C} \begin{bmatrix} S[i] \\ T[j] \end{bmatrix}$   
ER6       **si** score  $\geq$  meilleur **alors** meilleur  $\leftarrow$  score, extension  $\leftarrow j - j_0$   
ER7        $i \leftarrow i + 1, j \leftarrow j + 1$   
ER8 **retourner** meilleur, extension

## DÉTAILS : BANDE [VERS SUD-EST]

B1 **Entrée**  $i_0, j_0$  départ de l'extension,  $s_0$  score initial,  $d$  épaisseur  
B2  $A^* \leftarrow s_0, i \leftarrow i_0, D \leftarrow i_0 - j_0$   
B3 **tant que**  $i \leq |S|$   
B4      $j \leftarrow \max\{j_0, i - D - d\}$   
B5     **tant que**  $j \leq \min\{|T|, i - D + d\}$   
B6         calculer  $A(i, j)$ ; **si**  $A^* < A(i, j)$  **alors**  $A^* \leftarrow A(i, j)$   
B7          $j \leftarrow j + 1$   
B8     **si**  $\forall j: A(i, j) \leq 0$  **alors** sauter à Ligne B10.  
B9      $i \leftarrow i + 1$   
B10 reporter  $A^*$

# DÉTAILS BANDE 2 — GRAPHE

Calcul en ligne B6 ( $d = 2$ )



## DÉTAILS BANDE 2 — CODE

Calcul en ligne B6 : pondération par  $\mathbf{C}$

- si  $i = i_0, j = j_0$ , alors  $A(i, j) = s_0$
- si  $i = i_0$ , et  $j > j_0$ , alors  $A(i, j) = A(i, j - 1) + \mathbf{C} \begin{bmatrix} \bar{T}[j] \\ - \end{bmatrix}$
- si  $i > i_0$  et  $j = j_0$ , alors  $A(i, j) = A(i - 1, j) + \mathbf{C} \begin{bmatrix} S[i] \\ - \end{bmatrix}$
- si  $i > i_0 + d$  et  $j = i - D - d$ , alors
$$A(i, j) = \max \left\{ A(i - 1, j - 1) + \mathbf{C} \begin{bmatrix} S[i] \\ T[j] \end{bmatrix}, A(i - 1, j) + \mathbf{C} \begin{bmatrix} S[i] \\ - \end{bmatrix} \right\}$$



## DÉTAILS BANDE 2 — CODE (CONT.)

- si  $i > i_0$  et  $\max\{j_0, i - D - d\} < j < \min\{|T|, i - D + d\}$ , alors  $A(i, j) = \max\left\{A(i-1, j) + \mathbf{C}\left[\begin{smallmatrix} S[i] \\ - \end{smallmatrix}\right], A(i-1, j-1) + \mathbf{C}\left[\begin{smallmatrix} S[i] \\ T[j] \end{smallmatrix}\right], A(i, j-1) + \mathbf{C}\left[\begin{smallmatrix} - \\ T[j] \end{smallmatrix}\right]\right\}$
- si  $i > i_0$  et  $j = \min\{|T|, i - D + d\}$ , alors  $A(i, j) = \max\left\{A(i-1, j-1) + \mathbf{C}\left[\begin{smallmatrix} S[i] \\ T[j] \end{smallmatrix}\right], A(i, j-1) + \mathbf{C}\left[\begin{smallmatrix} - \\ T[j] \end{smallmatrix}\right]\right\}$

# DÉTAILS X-DROP

Idée : maintenir  $A^*$  score du meilleur alignement et ne pas continuer l'extension si  $A(i, j) < A^* - X$

Stocker  $col_g, col_d$  : colonnes de la dernière rangée que l'on a explorée.

(Ou code plus simple si l'exploration est dans une bande seulement : on n'a pas besoin de  $col_g, col_d$ )

[Code pour extensions vers sud-est seulement]

## DÉTAILS X-DROP 2

```
XD1 Entrée  $i_0, j_0$  départ de l'extension,  $s_0$  score initial,  $X$   
XD2  $A^* \leftarrow s_0, \text{col}_g \leftarrow j_0, \text{col}_d \leftarrow |T|$   
XD3  $i \leftarrow i_0$   
XD4 tant que  $i \leq |S|, \text{col}_g \leq \text{col}_d$   
XD5      $j \leftarrow \text{col}_g$   
XD6     tant que  $j \leq \min\{\text{col}_d + 1, |T|\}$   
XD7         calculer  $A(i, j)$   
XD8         si  $A(i, j) > A^*$  alors  $A^* \leftarrow A(i, j)$   
XD9         si  $A(i, j) < A^* - X$  alors  $A(i, j) \leftarrow -\infty$   
XD10         $j \leftarrow j + 1$   
XD11        tant que  $\text{col}_g \leq \text{col}_d$  et  $A(i, \text{col}_g) = -\infty$ ,  $\text{col}_g \leftarrow \text{col}_g + 1$   
XD12         $\text{col}_d \leftarrow \text{col}_d + 1$ ; tant que  $\text{col}_d \geq \text{col}_g$  et  $A(i, \text{col}_d) = -\infty$ ,  $\text{col}_d \leftarrow \text{col}_d - 1$   
XD13         $i \leftarrow i + 1$   
XD14 retourner  $A^*$ 
```

## DÉTAILS X-DROP 3

Calcul en ligne XD7 : pondération par  $\mathbf{C}$

- si  $i = i_0$  et  $j > j_0$ , alors  $A(i, j) \leftarrow A(i, j - 1) + \mathbf{C} \left[ \begin{array}{c} \bar{-} \\ T[j] \end{array} \right]$
- si  $i > i_0$  et  $j = \text{col}_g$  alors  $A(i, j) \leftarrow A(i - 1, j) + \mathbf{C} \left[ \begin{array}{c} S[i] \\ \bar{-} \end{array} \right]$
- si  $i > i_0$  et  $j = \text{col}_d + 1$  alors  $A(i, j) \leftarrow \max \left\{ A(i - 1, j - 1) + \mathbf{C} \left[ \begin{array}{c} S[i] \\ T[j] \end{array} \right], A(i, j - 1) + \mathbf{C} \left[ \begin{array}{c} \bar{-} \\ T[j] \end{array} \right] \right\}$
- sinon  $A(i, j) \leftarrow \max \left\{ A(i - 1, j) + \mathbf{C} \left[ \begin{array}{c} S[i] \\ \bar{-} \end{array} \right], A(i - 1, j - 1) + \mathbf{C} \left[ \begin{array}{c} S[i] \\ T[j] \end{array} \right], A(i, j - 1) + \mathbf{C} \left[ \begin{array}{c} \bar{-} \\ T[j] \end{array} \right] \right\}$