
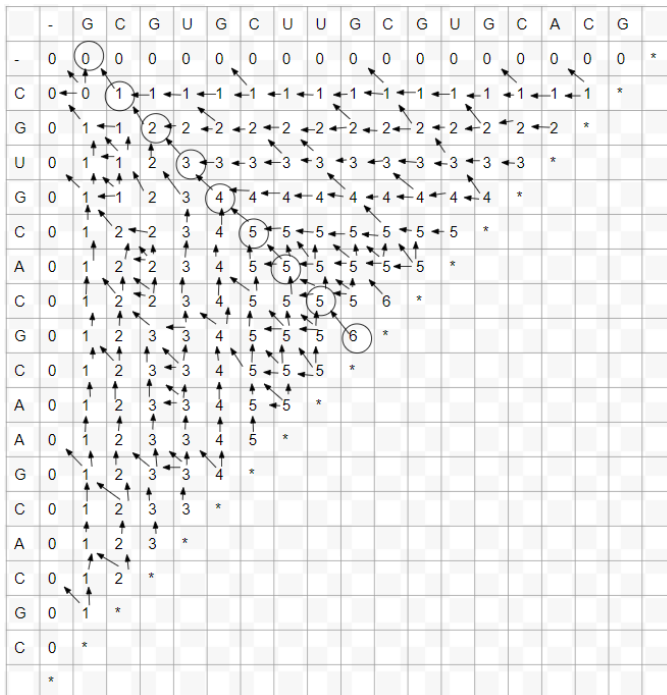


## IFT3295 - TP2

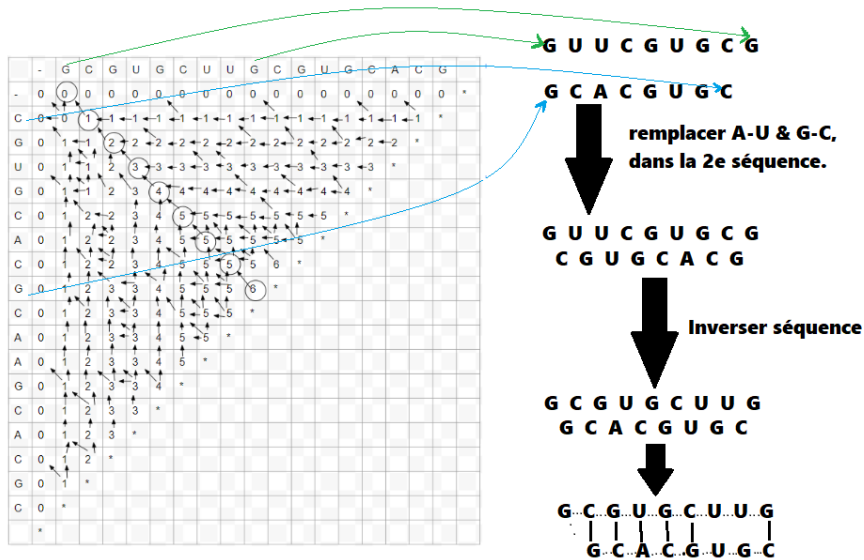
### Repliement d'ARNs

1. Les cellules de M sur l'anti-diagonale ( $M[i, |S| - i]$ ) correspondent aux différents scores des alignements. Il faut seulement remplir la moitié de la table.
2.
 
$$v(i,0) = v(j,0) = 0$$

$$v(i,j) = \max\{ v(i-1,j), v(i,j-1), v(i-1,j-1) + 1 \text{ si match ou } + 0 \text{ sinon} \}$$
3. Il suffit de remplir la table en utilisant les règles définies précédemment. Ensuite, sur l'anti-diagonale, on trouve le score maximale. Ensuite, suivre les flèches jusqu'à la case 0,0.
4. GCGUGCUUGCGUGCACG -> CGUGCACGCAAGCACGC  tp2 bio



Score : 6



5.

6. Afin de permettre les appariements "G-U", on doit modifier les règles de récurrences. On peut croire qu'il suffit simplement d'ajouter un autre match possible entre "G-U", mais ce n'est pas aussi simple. En effet, il faut se rappeler que nous avons remplacé les "G" par des "C" et les "U" par des "A" et inversement (Sr). Donc, nous allons accepter les appariements "G-A" et "C-U".

$$v(i,0) = v(j,0) = 0$$

$$v(i,j) = \max \{ v(i-1,j), \\ v(i,j-1), \\ v(i-1,j-1) + 1 \text{ si match} \\ v(i-1,j-1) + 1 \text{ si "G-A" ou "C-U"} \\ v(i-1,j-1) + 0 \text{ sinon} \}$$

7. a)

b)

## Alignement multiple de séquences

1. Voir main.py, résultat :

```
[0, 54, 20, 34, 80]
[54, 0, 94, 143, 68]
[20, 94, 0, 86, 15]
[34, 143, 86, 0, 31]
[80, 68, 15, 31, 0]
```

2. voir main.py. Ici, il suffit simplement de faire le résultat de chaque ligne et retenir le max:

La séquence centrale S\* est la séquence 2 avec la plus grande somme de 359

```
MEKVPGEMEIERRERSEELSEAERKAVQATWARLYANCEDVGVAIVRFFVNFPSAKQYFSQFKHMEEPLE
MERSPQLRKHACRVMGALNTVVENLHDPEKVSSVLSLVGKAHALKHKVEPVYFKILSGVILEVIAEEFANDFP
PETQRAWAKLRGLIYSHVTAAYKEVGWVQVQPNATTPPATLPSSGP
```

- 3.

S' 0 :

```
M*****G*EIGF***TEKQ*EAL***VKESWEILKQDIPKYSLHFFSQILEIAPAAKGLFSFLRDSDE**VPHNNPKLKAHAVKVKMTCETAIQLREEGKVVVADTTQLYLSIHL
KSGVIDP*HFEVVKEALLRTLKEGLGEKYNEEVEGAWS**Q**AYDHLALA*****IK*****TEMKQES
```

S\* 1 :

```
MEKVPGEMEIERRERSEELSEAERKAVQATWARLYANCEDVGVAIVRFFVNFPSAKQYFSQFKHMEEPLEMERSPQLRKHACRVMG*ALNTVVENLHDPEK*V*SSVLSLV
GKAHALKHKVEPVYFKILSGVILEVIAEEFANDFPETQRAWAKLRGLIYSHVTAAYKEVGWVQVQPNATTPPATLPSSGP
```

## Outils bioinformatique

- 1.

La séquence S1 est le gène qui code pour l'hémoglobine 2 de l'espèce *Arabidopsis thaliana*, une espèce de plante considérée comme mauvaise herbe. Sa fonction est de transporter l'oxygène.

La séquence S2 est le gène qui code pour le cytoglobine qui se retrouve dans beaucoup d'espèces dont notamment le boeuf domestique (*Bos taurus*).

La séquence S3 est le gène qui code pour une sous-unité alpha d'hémoglobine qui se retrouve dans beaucoup d'espèces dont notamment le coq (*Gallus gallus*).

La séquence S4 est le gène qui code pour la myoglobine, une protéine qu'on retrouve dans les vertébrés et qui a pour fonction de stocker l'oxygène.

S3	-----MVL SAADKNNVKGIFTKIAGHAAEEYGAETLERMFTTYPPTKYTF	44
S2	MEKVPGMEIERRERSEELSEAERKAVQATWARLYANCEDVGVALVRFVFNFP SAKQYF	60
S4	-----MGLSDGEWQLVLNVWGKVEADIPGHGQEV LIRLFKGHPETLEKF	44
S1	-----MGEIGFTEKQEALVKESWEILKQDIPKYSLHFFSQILEIAPA AKGLF	47
S5	-----MERLESELIRQSWRAVSRSPLEHGTVLFSRLFALEPSLLPLF	42
	: : : : : . : :: * *	
S3	PHFDLS-H-----GSAQIKGHGKKVVAALIE-----AANHIDDIAGT LSKLSDLHAHK	91
S2	SQFKHM-EEPLEMERSPQLRKHACRMV GALNTVV---ENLHDPEKVSSVSLV GKAHALK	116
S4	DKFKHL-KSEDEMKASEDLKKHGATVLTALGGIL---KKKGHH---EAEIKPLAQSHATK	97
S1	SFLRDSDEVPHN---NPKLKAHAVKVKMT CETAIQLREEGKVVVADTT LQYLGSIHLK-	103
S5	QYNGRQFSSPEDCLSSPEFLDHIRKVMLVID-AA--VTNVEDLSSLEEY LATLGRKHRA-	98
	. : : * *, . : : *	
S3	LRVDPVNFKLLGQCFLVWVAIHHPAALTPEVHASLDKFLCAVGTVLTAKYR-----	142
S2	HKVEPVYFKILSGVILEVIAEEFANDFP PETQRAWAKLRGLIYSHVTAAYKEVGWVQQVP	176
S4	HKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG---	154
S1	SGVIDPHFEVVKEALLRTLKEGLGEKYNEEVEGAWSQAYDHLALAIKTE-----MKQEE S	158
S5	VGVR LSSFSTVGESLLYMLEKCLGPDFTPATRTAWSQLYGAVVQAMS RG-----WDGE--	151
	: : . : : : : . : :	
S3	-----	142
S2	NATTPPATLPSSGP	190
S4	-----	154
S1	-----	158
S5	-----	151