

Avant Propos

Objectifs Généraux du module :

Les objectifs généraux de ce module sont :

1. La bonne connaissance de la structure et des propriétés des acides nucléiques.
2. L'assimilation des grands principes de base impliqués dans la transmission et la réparation du matériel génétique, ainsi que l'expression des gènes et sa régulation.
3. L'assimilation des principes de la manipulation des acides nucléiques
- 4.

Objectifs des Travaux Pratiques :

Les séances de travaux pratiques de Biologie Moléculaire visent à initier l'étudiant de la Filière des Sciences de la Vie aux outils techniques de base de biologie moléculaire. Cette partie du TP a pour but de vous faire découvrir quelques outils utilisés pour comprendre et décrire des molécules d'ADN ou de protéines. En effet, à l'issue du séquençage d'un fragment de chromosome, ou plus largement du génome d'un organisme, les données obtenues ressemblent à :

```
GGCTCTACTACCTGTTACCCTGGGGTAGTGTCTAGTGACAGCAAAAATAGAATAGTGTCAGTCAT
AGTCAGGTGGCTTTGGAAACTGAAGCTGCCTTCCCAGGACCTGGCCAGTCTCTTGTGGCAGCCAG
GGCTGACAGGCAACTCTTCCTCACAGTCACTGGTCCAGTCACGGAAAGCAGGGATTACTCCGCT
ATGGCTACACGTACCTCTCTCAAGGATGAAGATCTGGTAAGTGAAGTACTCTTGGACATTT
```

Les questions que les biologistes se posent sont alors les suivantes :

- * cette séquence contient-elle des gènes ?
- * si oui, ces gènes codent-ils pour des protéines, pour des ARNs ?
- * les protéines putatives codées par ces gènes sont-elles déjà connues ?

Dans cette partie du TP, nous vous proposons d'endosser le rôle de chercheur pour analyser un fragment de chromosome, séquencé à partir d'ADN donné, par un logiciel Serial Cloner.

L'étudiant doit être capable d'interpréter des résultats expérimentaux obtenus par les techniques les plus courantes de biologie moléculaire ou à l'exploration de matériel génétique, en général, par les laboratoires de biologie.

A la fin des travaux pratiques, les étudiants doivent rédiger un compte rendu concernant les séances de TP.

A. Le séquençage du DNA par la méthode de Sanger

Autrefois, toutes les séquences de protéines étaient déduites de la séquence d'une série de peptides produits par diverses protéases de spécificité bien déterminée. Bien que ces techniques soient encore d'un grand intérêt aujourd'hui, il est généralement plus facile d'établir la structure primaire d'une protéine d'une manière indirecte, par séquençage du DNA complémentaire (cDNA). Les cDNA sont des copies d'ARN messagers produites in vitro par une transcriptase réverse.

Pour des raisons de sécurité et de temps, il n'est pas possible de vous faire séquencer un fragment de DNA ; nous nous contenterons donc de vous expliquer la méthode la plus communément utilisée, celle de Sanger, encore appelée "méthode de séquençage par terminaison des chaînes avec des didéoxynucléotides". Vous pourrez aussi vous familiariser avec la lecture d'une autoradiographie d'un gel de séquençage.

Pour séquencer un fragment de DNA, il faut d'abord l'insérer dans un vecteur de séquençage. Il en existe de plusieurs types. Ils dérivent de phages et de plasmides (= phagemides) et portent des noms commerciaux bizarres : M13mp18, M13mp19, pTZ18R, pTZ19R, etc...

Ces phagemides comportent (voir Figure 1) :

- pBR322ori : origine de répllication du plasmide pBR322, permettant la production en masse de DNA recombinant double brin par des bactéries.
- f1-ori : origine de répllication du phage f1, nécessaire à la production de DNA simple brin encapsidé.
- AMP^r : zone conférant la résistance à l'ampicilline et empêchant la croissance des bactéries sans phagemides en présence de cet antibiotique.

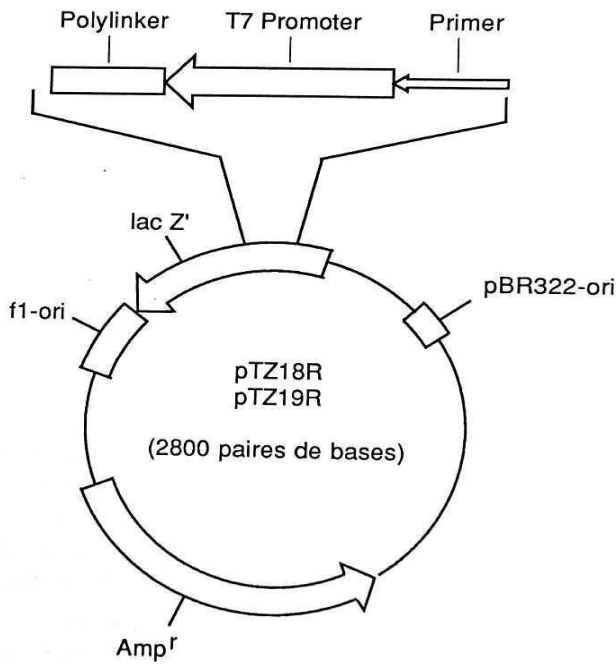


Figure 1. Structure d'un vecteur de séquençage.

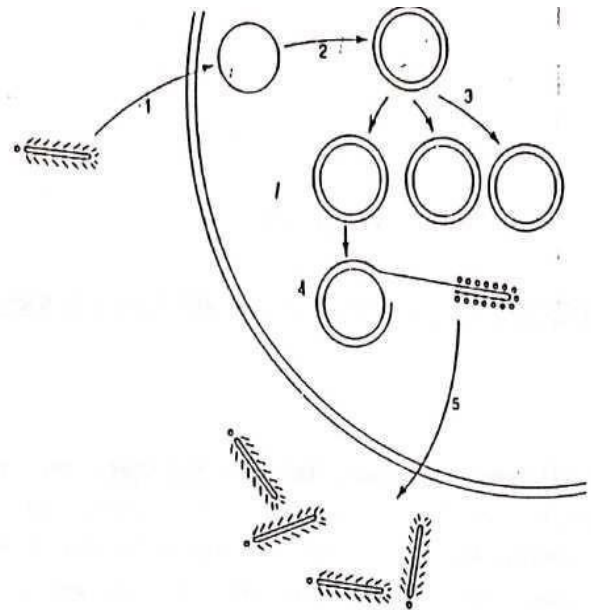


Figure 2. Amplification du DNA recombinant et production du DNA simple brin.

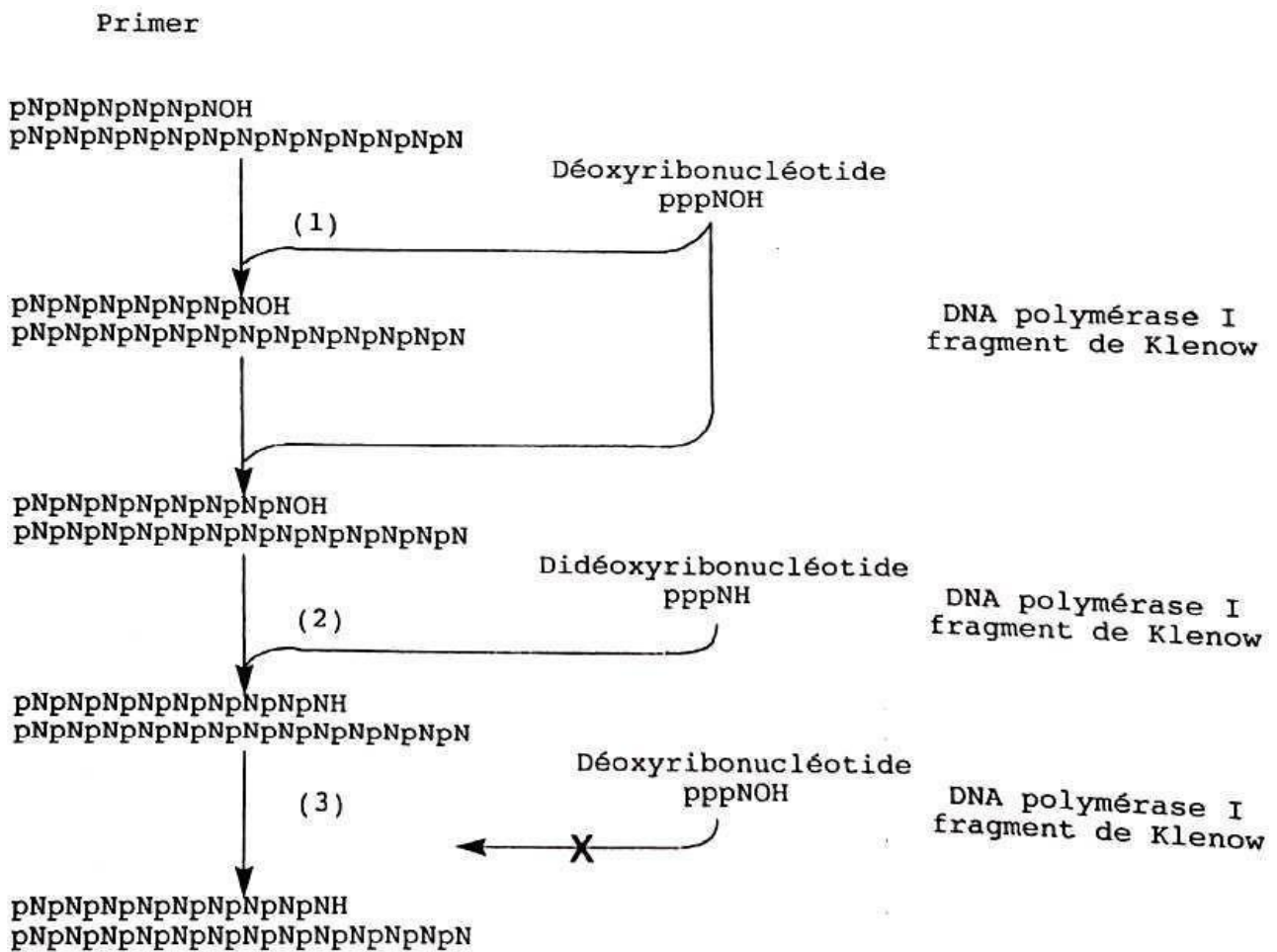


Figure 3. Terminaison des chaînes par les didéoxyribonucléotides.

- un polylinker : inséré ici dans une portion de l'opéron lactose de E. coli (lac Z'). Le polylinker comprend une série de sites de restriction ; c'est au niveau d'un de ces sites que le fragment de DNA à séquencer sera intégré.
- une séquence complémentaire au "primer", oligonucléotide synthétique pouvant servir d'amorce à la DNA polymérase.
- le promoteur de la T7 RNA polymérase, sans intérêt pour nous.

L'insertion du fragment de DNA à séquencer dans le vecteur de séquençage ne produit qu'une très petite quantité de DNA recombinant ; il faut donc l'amplifier (Figure 2). On fait pénétrer le DNA recombinant dans une bactérie par transfection (1) ; il se réplique dans la bactérie comme un plasmide (2), produisant rapidement un grand nombre de copies (3) ; puis l'un des deux brins du plasmide est répliqué préférentiellement (4) ; les bactéries découpent ce DNA simple brin en fragments de longueur unitaire et, après encapsidation, les expulsent dans le milieu (5). C'est ce DNA simple brin purifié à partir du milieu de culture qui est utilisé pour déterminer la séquence du DNA étudié.

Principe de la méthode de séquençage du DNA par la méthode des terminaisons de chaînes par des didéoxyribonucléotides

La méthode tire profit des propriétés de certaines DNA polymérases (Figure 3) :

- celles-ci peuvent, en présence des 4 désoxyribonucléosides triphosphates (dNTP), synthétiser un brin de DNA complémentaire au DNA simple brin utilisé comme matrice, par prolongement d'un primer (= amorce) dont l'extrémité 3'-OH est libre (1).

- elles acceptent aussi les 2',3'-didéoxyribonucléosides triphosphates (2), mais l'incorporation de ces analogues de substrats constitue un obstacle au prolongement ultérieur de la chaîne ; en effet, celle-ci ne se termine plus alors par un groupe 3'-OH libre, indispensable à la liaison du nucléotide suivant (3) et Figure 5, page 8.

En pratique on procède comme indiqué à la Figure

4, page 5 :

- on hybride le DNA simple brin purifié avec l'amorce ; cette hybridation ne se produit que dans des conditions bien déterminées de force ionique et de température.
- on ajoute le fragment de Klenow de la DNA polymérase I.
- le mélange est réparti dans 4 petits tubes renfermant les 4 dNTP, dont un radioactif, et une très petite quantité de ddCTP, ddGTP, ddTTP ou ddATP ; ces petits tubes sont marqués C°, G°, T° et A°, respectivement.

La DNA polymérase synthétise du DNA complémentaire au DNA matrice dans les 4 tubes : dans le tube marqué C°, l'enzyme aura le choix entre dCTP et ddCTP, et produira une série de chaînes incomplètes, mais toutes terminées par C ; dans le tube G°, toutes les chaînes seront terminées par G, etc...

Les produits de réaction sont analysés dans 4 pistes adjacentes d'un gel de polyacrylamide : les produits sont chauffés à 100 °C avant le chargement sur le gel, pour séparer les brins radioactifs du DNA matrice. La séparation se fait en présence d'urée et à haute tension (1.500 volts durant plusieurs heures). L'urée et la chaleur produite par la haute tension empêchent le réappariement des brins complémentaires en cours de migration. La vitesse de migration est une fonction inverse de la longueur des fragments. Les petits fragments migrent donc le plus vite. Leur position est repérée par autoradiographie du gel séché : le pouvoir de résolution est si élevé qu'on peut distinguer aisément des fragments dont la longueur ne diffère que par

un seul nucléotide. On arrive à lire de 250 à 400 nucléotides. Il faut donc sous-cloner une série de fragments chevauchants pour élucider la structure de la protéine entière. Le primer est le même dans tous les cas puisqu'il est complémentaire d'une petite zone du vecteur située juste en amont du site d'insertion du DNA cloné.

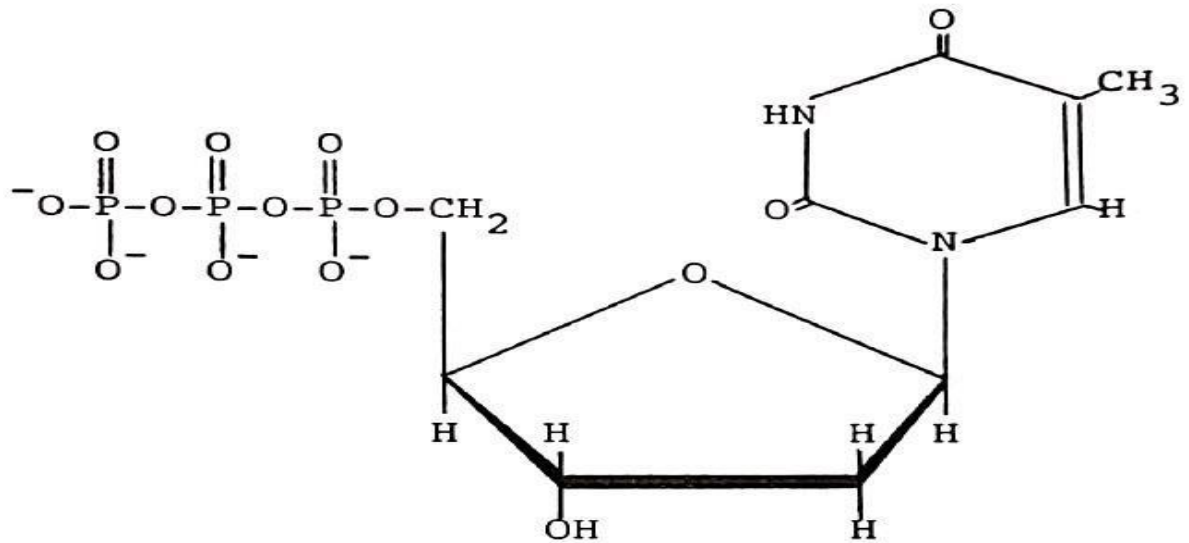
La séquence d'une protéine est déterminée par la séquence du DNA du gène correspondant. Ce DNA est d'abord transcrit en ARN, dit messenger, parce qu'il porte l'information inscrite dans le gène jusqu'aux ribosomes où se produit la synthèse de la protéine.

La traduction de la séquence des nucléotides de l'ARN messenger en séquence d'acides aminés se fait selon le code génétique (Voir Fig 6, p.13). Chaque acide aminé est spécifié par une succession bien particulière de 3 nucléotides appelée "codon", ou "triplet". Il y a 64 codons différents : 3 servent à la terminaison de la traduction ; ici ils sont marqués "Term*", mais sont aussi appelés "Ocre", "Ambre" et "Opale" pour UAA, UAG et UGA, respectivement ; les 61 codons restants déterminent les 20 acides aminés qui entrent dans la constitution des protéines. Remarquez qu'il est fréquent que plusieurs codons différents codent pour le même acide aminé : on dit pour cela que le code génétique est dégénéré. Le codon AUG code pour Met, mais sert aussi à l'initiation de la traduction (Init.).

L'ordinateur traduit les lectures des gels (voir exemples page 10) en séquence de DNA double brin et indique la position des sites de restriction (voir pages 9 et 11). Le programme traduit aussi ces séquences de nucléotides en séquences d'acides aminés. Puisque les codons sont des triplets et que l'information est différente dans les 2 brins d'ADN, il y a 6 phases de lecture possibles (voir pages 10 et 12). On voit que certaines phases de lecture sont interrompues par des codons stop (marqués ici "OCH", "AMB" et "OPA", pour "Ochre", "Ambre" et "Opale") et produisent des fragments d'ADN trop courts pour contenir l'information déterminant la structure de la protéine étudiée ; ces phases-là sont donc éliminées d'emblée. Les

autres phases de lecture possibles se dégagent de l'enchaînement d'une série de fragments ininterrompus. L'identification absolue de la vraie phase de lecture se fait sur base de toutes les propriétés connues de la protéine : P.M., pI, présence de signal peptide, hydrophobicité, etc...

Déoxy TTP



Didéoxy TTP

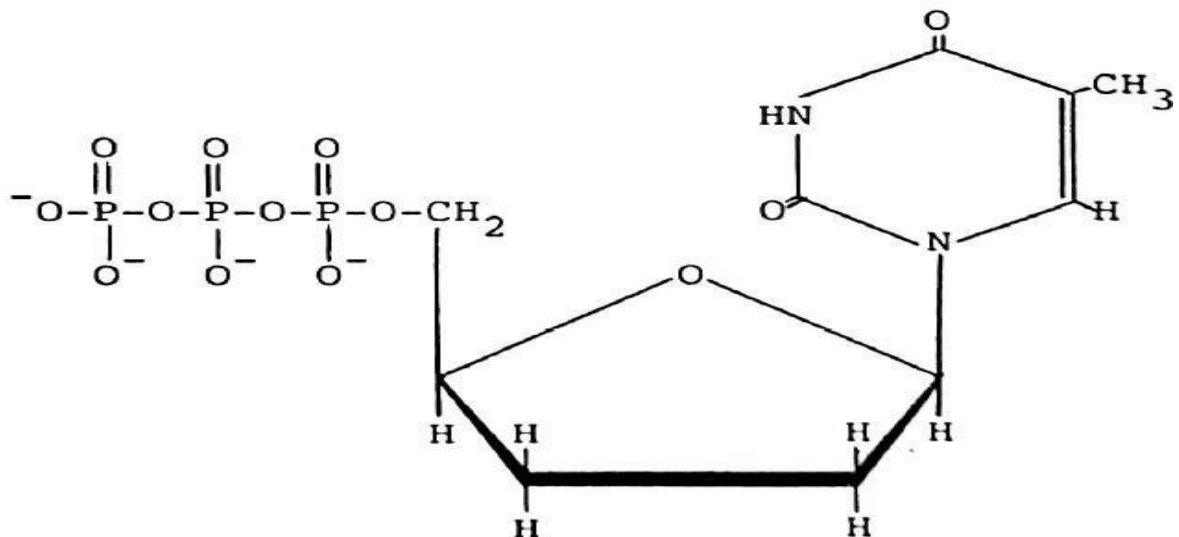


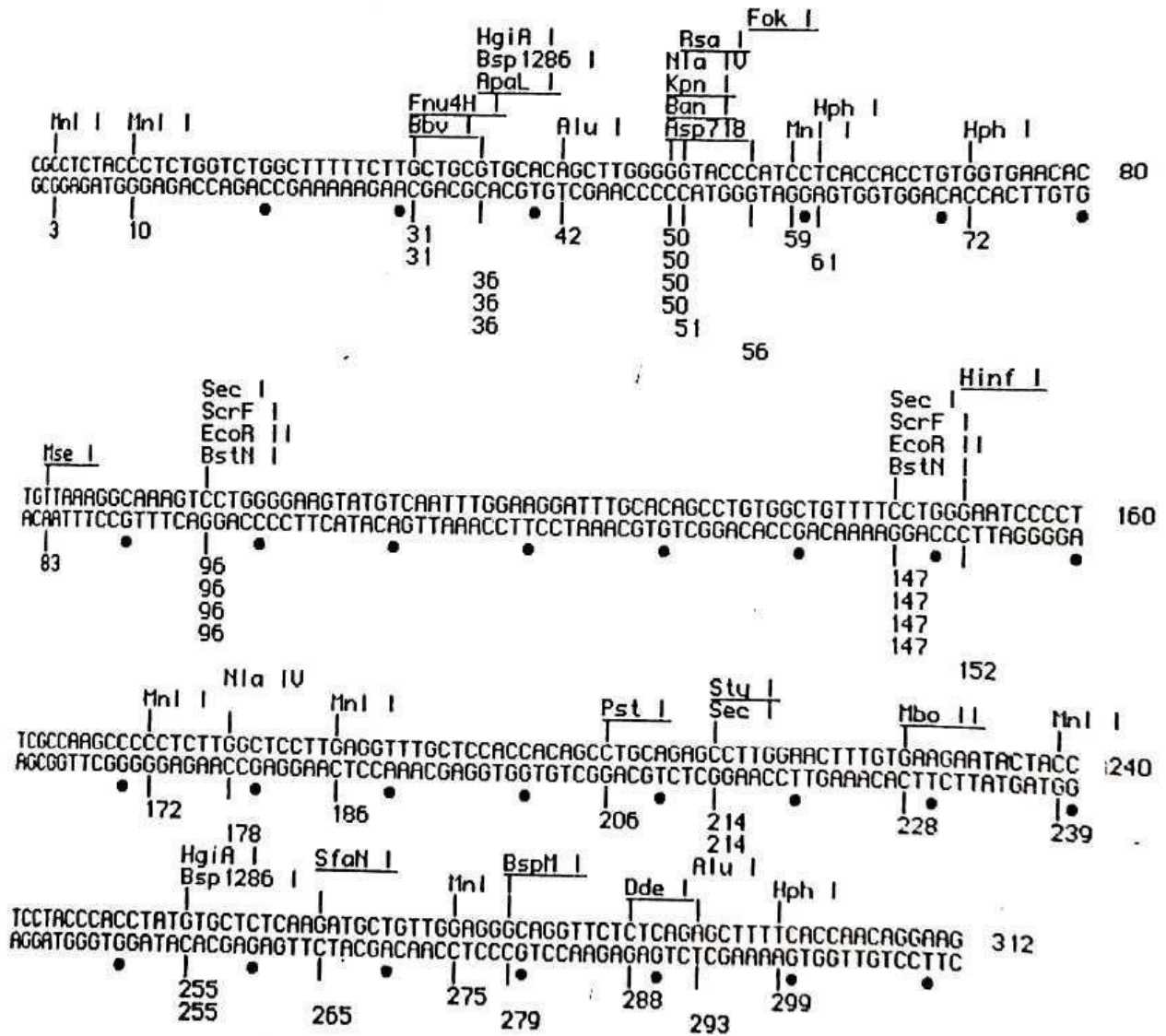
Figure 5. Différences entre les Déoxy TTP et les Didéoxy TTP.

Exemples de séquences lues

Pour le fragment de DNA N° 1

	10	20	30	40	50	60	
1	CGCCTCTACC	CTCTGGTCTG	GCTTTTTCTT	GCTGCGTGCA	CAGCTTGGGG	GTACCCATCC	60
61	TCACCACCTG	TGGTGAACAC	TGTTAAGGC	AAGTCCTGG	GGAGTATGT	CAATTTGGAA	120
121	GGATTTGCAC	AGCCTGTGGC	TGTTTTCTG	GGATCCCT	TCGCCAGCC	CCCTCTTGGC	180
181	TCCTTGAGGT	TTGCTCCACC	ACAGCCTGCA	GAGCCTTGG	ACTTTGTGAA	GAATACTACC	240
241	TCCTACCCAC	CTATGTGCTC	TCAAGATGCT	GTTGGAGGGC	AGGTTCTCTC	AGAGCTTTTC	300
301	ACCAACAGGA	AG					312
	10	20	30	40	50	60	

Carte de restriction du fragment de DNA N° 1

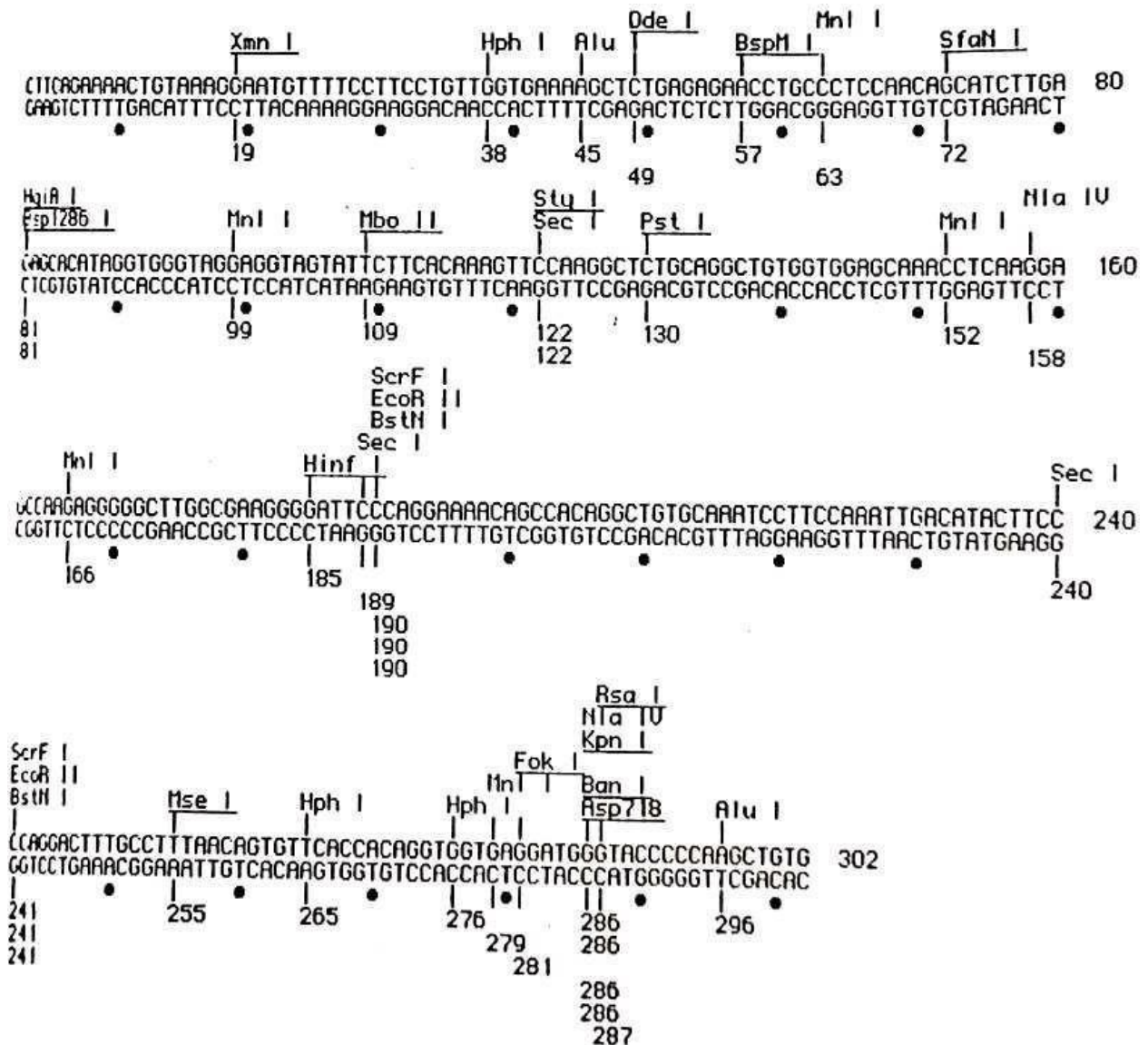


Pour le fragment de DNA N° 2

```

      | 10      | 20      | 30      | 40      | 50      | 60
1  CTTCAGAAAA CTGTAAGGA ATGTTTTCTT TCCTGTTGGT GAAAAGCTCT GAGAGAACCT 60
61 GCCCTCCAAC AGCATCTTGA GAGCACATAG GTGGGTAGGA GGTAGTATC TTCACAAAGT 120
121 TCCAAGGCTC TGCAGGCTGT GGTGGAGCAA ACCTCAAGGA GCCAAGAGGG GGCTTGGCGA 180
181 AGGGGATTCC CAGGAAACA GCCACAGGCT GTGCAATCC TTCCAATTG ACATACTTCC 240
241 CCAGGACTTT GCCTTTACA GTGTTACCA CAGGTGGTGA GGATGGGTAC CCCCAGCTG 300
301 TG
      | 10      | 20      | 30      | 40      | 50      | 60
    
```

Carte de restriction du fragment de DNA N° 2



1ère base du codon		2ème base du codon				3ème base du codon				
		U	C	A	G					
U	UUU]Phe	UCU]Ser	UAU]Tyr	UGU]Cys	U	
	UUC		UCC		UAC		UGC		C	
	UUA		UCA		UAA		UGA		Term.	A
	UUG		UCG		UAG		UGG		Trp	G
C	CUU]Leu	CCU]Pro	CAU]His	CGU]Arg	U	
	CUC		CCC		CAC		CGC		C	
	CUA		CCA		CAA		CGA		A	
	CUG		CCG		CAG		CAG		G	
A	AUU]Ile	ACU]Thr	AAU]Asn	AGU]Ser	U	
	AUC		ACC		AAC		AGC		C	
	AUA		ACA		AAA		AGA		A	
	*AUG		ACG		AAG		AGG		Arg	G
G	GUU]Val	GCU]Ala	GAU]Asp	GGU]Gly	U	
	GUC		GCC		GAC		GGC		C	
	GUA		GCA		GAA		GGA		A	
	**GUG		GCG		GAG		GGG		G	

*Codon d'initiation ; ** Codon d'initiation alternatif.

	1 lettre	3 lettres
Alanine	A	Ala
Cystéine	C	Cys
Acide aspartique	D	Asp
Acide glutamique	E	Glu
Phénylalanine	F	Phe
Glycocolle	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Lysine	K	Lys
Leucine	L	Leu
Méthionine	M	Met
Asparagine	N	Asn
Proline	P	Pro
Glutamine	Q	Gln
Arginine	R	Arg
Sérine	S	Ser
Thréonine	T	Thr
Valine	V	Val
Tryptophane	W	Trp
Tyrosine	Y	Tyr

Figure 6. Le code génétique et abréviations standards des acides aminés naturels.

Rapport sur les travaux Pratiques de Biologie Moléculaire

Séquençage des protéines

I- Lisez un fragment de séquence sur l'autoradiographie d'un gel (Figure 7). Inscrivez vos lectures dans le cadre prévu ci-dessous. Faites vérifier la séquence notée par les membres du groupe de TP et analysez les par le logiciel Serial Cloner.

5

10

15

1															
---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

16															
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

31															
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

46															
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

61															
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

76															
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

91															
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

106															
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

121															
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

136															
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

151															
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

166															
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

181															
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



Figure 7. Autoradiographie d'un gel de séquençage.

II- Examinez les cartes de restriction des pages 9 et 11 :

Pensez-vous que :

- a- les 2 fragments de l'ADN sont tout à fait différents ?
- b- les 2 fragments de l'ADN sont insérés dans le même sens dans le vecteur de séquençage, mais décalés l'un par rapport à l'autre ?
- c- les 2 fragments de l'ADN se recouvrent partiellement, mais sont insérés en sens opposé dans le vecteur ?

III- Examinez les phases de lecture des pages 10 et 12 :

- a- Repérez les codons "stop" dans les différentes phases ouvertes de lecture possibles.
- b- Soulignez la phase de lecture ouverte la plus longue dans ces deux séquences.
- c- Comparez ces 2 phases de lecture soulignées, compte tenu de votre réponse au point II.

IV- Examinez la séquence suivante par une approche *in silico* et ceci en utilisant le programme de recherche BLAST (Basic Local Alignement Search Tool) tout en recherchant toutes les informations relatives à cette séquence (Voir cours magistral).

1	A	T	A	C	A	C	C	G	G	A	T	T	T	G	C
16	C	A	A	G	A	C	A	G	A	G	T	G	T	G	G
31	A	T	T	C	T	G	C	C	A	T	T	G	A	A	A
46	A	A	G	C	T	T	T	G	A	A	G	G	T	C	T
61	G	G	G	A	G	G	A	G	G	T	G	A	C	C	C
76	C	A	C	T	C	A	C	T	T	T	C	T	C	C	A
91	G	G	A	T	C	T	C	T	G	A	A	G	G	A	G
106	A	G	G	C	T	G	A	C	A	T	A	A	T	G	A
121	T	C	T	C	C	T	T	T	G	C	A	G	T	T	G
135	G	A	G	A	A	C	A	T	G	G	A	G	A	C	T
151	T	T	G	T	C	C	C	T	T	T	T	G	A	T	G
166	G	G	C	C	T	G	G	A	A	C	A	G	T	C	T