

Chapitre I: Biologie moléculaire et Génomique (partie2)

Dekkiche. samia

II-Génomique et annotation d'un génome

1- ADN et gènes

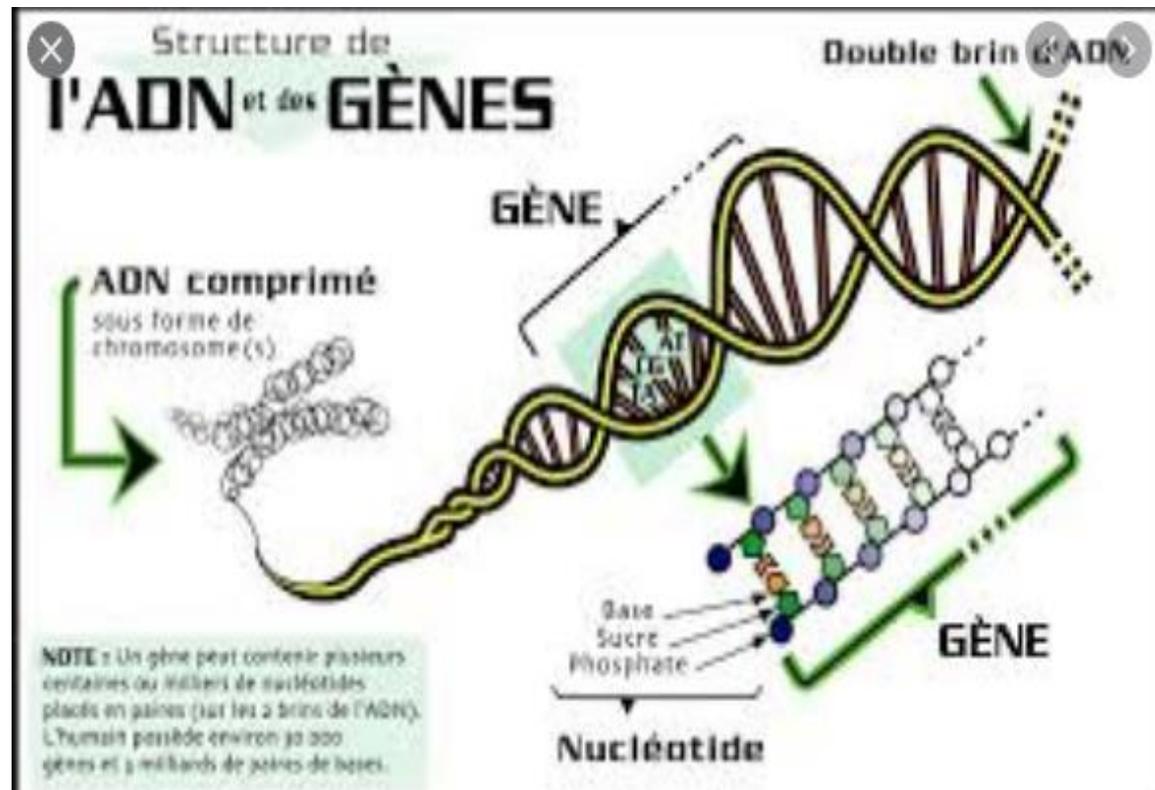
- Un gène est **une séquence d'ADN**, qui spécifie la synthèse d'une chaîne polypeptidique ou d'un acide ribonucléique fonctionnelle.
- Le gène représente l'unité d'information génétique
- Le gène est polymorphe (sous forme de plusieurs versions d'allèles)

NB:

-L'ensemble des gènes constitue le **génom**

-L'ensemble des ARNm constitue le **transcriptome**

-L'ensemble des protéines constitue le **proteome** ou **proteome**



II-Génomique et annotation d'un génome

2-La génomique

C'est la science qui étudie l'ensemble des gènes (étude des génomés).

La génomique structurale (structure des gènes)

- séquençage des gènes,
- localisation des gènes,
- leur disposition sur les chromosomes,
- nombre d'exons d'introns
- polymorphismes, ...)

La génomique fonctionnelle

- fonction des gènes.
- mécanismes de l'expression des gènes

(La génomique expressionnelle).

3-L'annotation d'un génome

Consiste à analyser les séquences nucléotidiques en se basant sur deux points essentiels:



-localiser les gènes et les régions non codantes
(**génomique structurale**)



-identifier ou prédire leurs fonctions biologiques
(**génomique fonctionnelle**).

NB: Prédire la fonction des gènes c'est leur attacher une étiquette, portant

- leurs nom probables,
- leurs fonctions probables,
- leurs interactions probables,

Génomique structurale

1-Etudes portantes sur un gène

Structure d'un gène

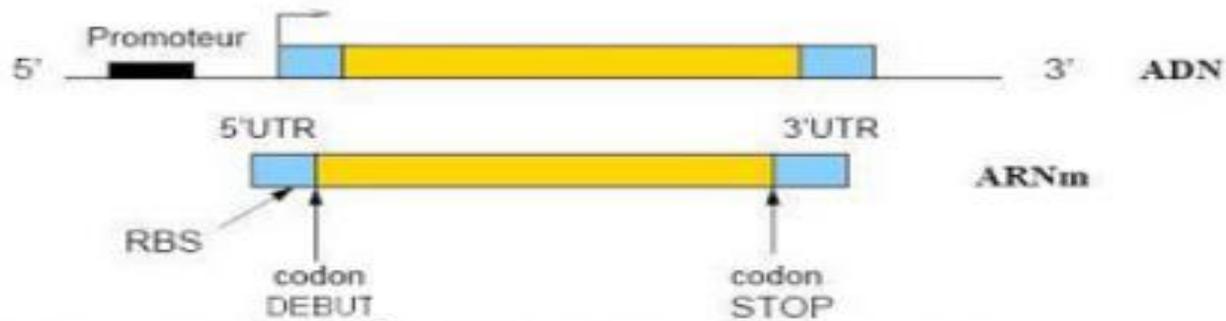
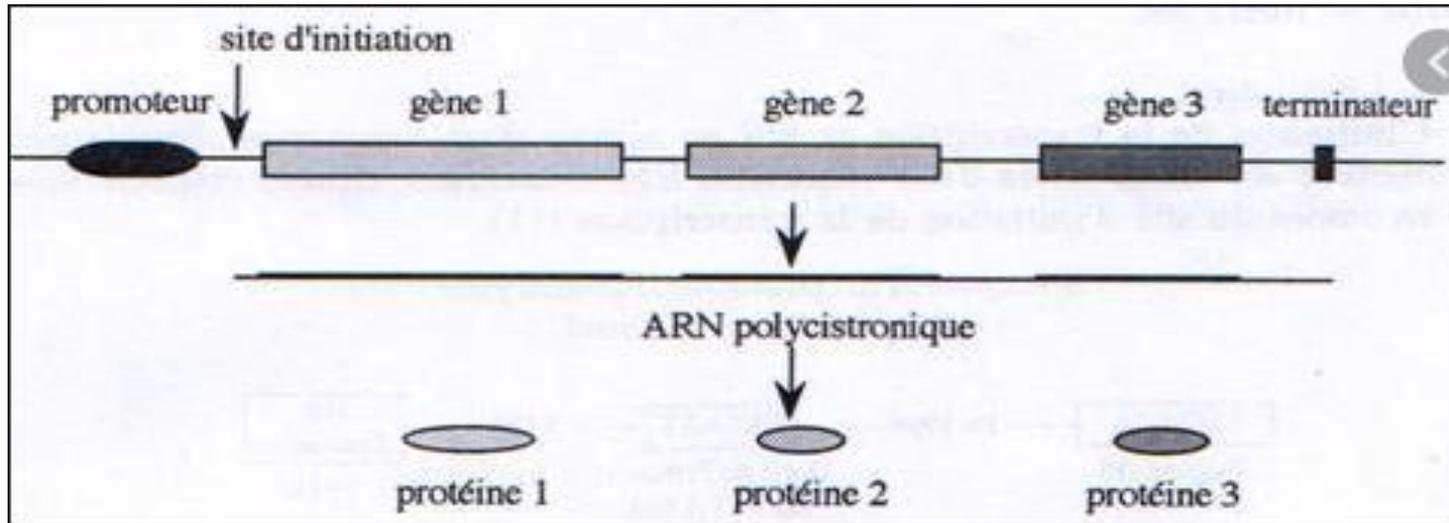
Quatre régions importantes

- **Une région régulatrice** (le promoteur et l'opérateur ainsi que d'autres séquences participant dans la régulation de la transcription)
- **Une région contenant le gène** ou les gènes codants (gènes de structure)
- Des séquences pour **initiation de la transcription**
- Des séquences pour **terminaison de la transcription**

NB: des différences dans les séquences de ces quatre régions sont remarquées entre les procaryotes et les eucaryotes :

Génomique structurale

Structure d'un gène

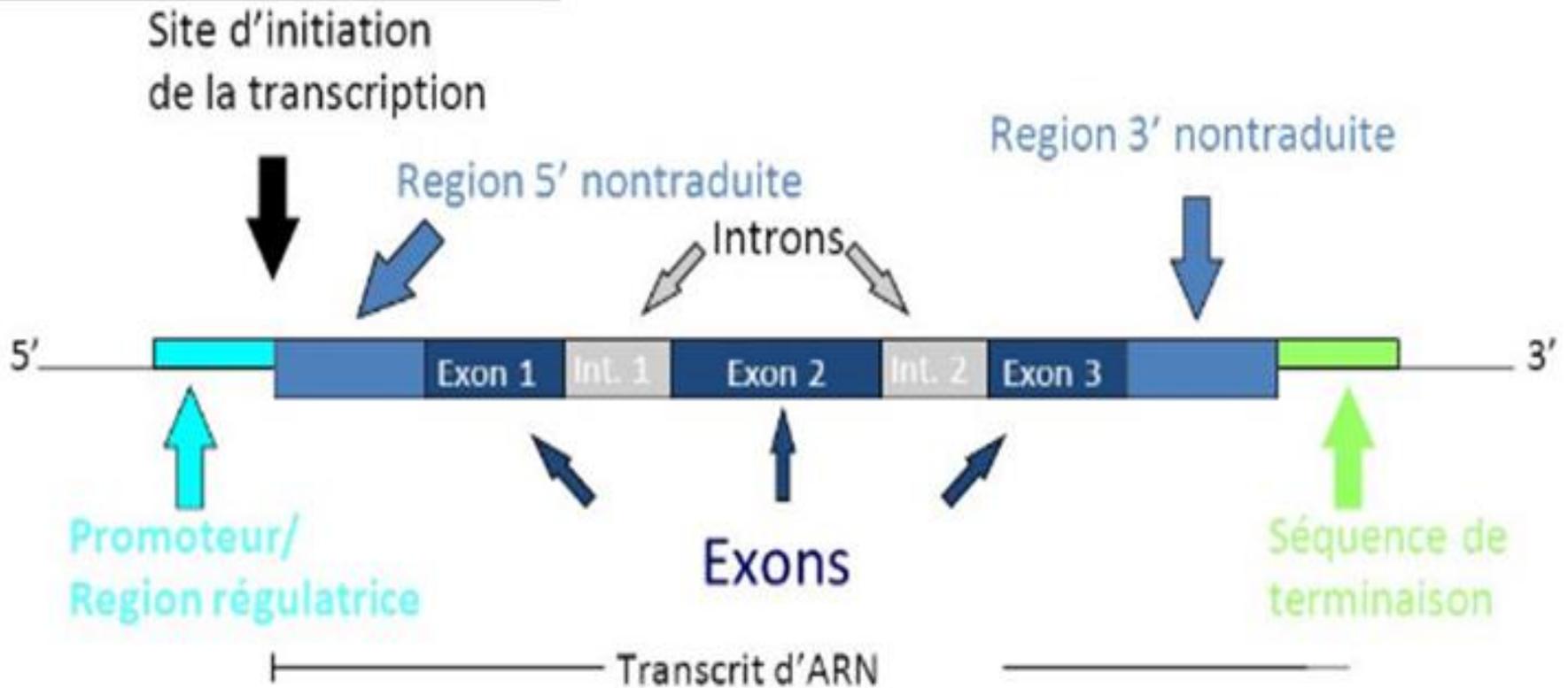


UTR : Région non traduite (UnTranslated Region)
RBS : Site de fixation ribosome (Ribosome binding site)

chez les procaryotes

Génomique structurale

Structure d'un gène



chez les eucaryotes

Génomique structurale

Identification de la séquence d'un gène

Le gène peut être identifié soit par la détermination de sa séquence complète soit en se basant sur ses séquences spécifiques par ses marqueurs spécifiques (empreintes génétiques)

Techniques d'identification



• Hybridation moléculaire,
(Southern blot, Northern blot, puce d'ADN, FISH et autres techniques),



RFLP
(polymorphisme des longueurs de fragments de restriction).



• Séquençage (soit par la méthode Sanger automatisé soit par l'une des méthodes de séquençage de nouvelle génération.



Génomique structurale

2-Etudes portantes sur le génome

Taille et diversité du génome

A-Unité de mesure de la taille

-**nucléotides**

-**bases** : kilobases (kb), Megabases (Mb), sachant que $1\text{Kb}=10^3\text{b}$ et $1\text{Mb}=10^6\text{b}$

-**paire de bases** (pb)

-**picogramme** (pg) (sachant que $1\text{pg}=10^3\text{Mpb}$)

B- variabilité de la taille

La taille du génome varie de quelques Kilo bases chez les virus a plusieurs centaines de milliers de Mb chez les eucaryotes (environ 3 milliards pour le génome humain).

virus	Eucaryotes
Quelque Kilobases	- <i>Paris japonicum</i> (plante), son génome est de 150milliard pb=prés de 50 fois la taille du génome humain

Génomique structurale

2-Etudes portantes sur le génome

Nombre de gène dans un génome

organisme	Nombre de gènes	Taille du génome (Mpb)
<i>E.coli</i>	4300	4,6
Levure de bière	6000	12,1
Drosophile (insecte)	Environ 14500	150
Nématodes	Environ 21000	110
Souris	Environ 30000	27000
homme	Environ 30000	3400

NB: Le nombre de gènes dans un génome est variable beaucoup moins que la taille du génome.

Génomique structurale

2-Etudes portantes sur le génome

Localisation des gènes codants



Outils de statistiques = bioinformatique



A- recherche et identification des codons

Reconnaissance des régions codantes sur la base de l'analyse de la fréquence des codons



identification des régions comportant probablement les gènes

B-recherche et identification des motifs spécifiques des gènes



Recherche des signatures caractéristiques du démarrage et de la fin, des jonctions entre les introns et les exons, ...

- séquences promotrices
- terminatrices,
- Sites de fixation du ribosome (RBS).

Génomique et annotation d'un génome

Génomique structurale

Séquençage d'un génome

Le séquençage d'un génome  détermination de la séquence nucléotidique de l'ADN présent dans chaque cellule d'un organisme donné

Difficultés du séquençage 

- La grandeur du génome
- Sa richesse en séquences répétées

Les génomes séquencés

Virus	Procaryotes	Eucaryotes
Plus de 10 ³ (VIH, virus d'immuno Déficience Humaine)	- Archea et bactéries (plus de 17 souches et espèces)	21 espèces
	Eubactéries (plus de 125 souches et espèces)	

NB: Le site GOLD (Génome Online Database), contient tous les génomes séquencés ainsi que les projets en cours concernant le séquençage

Génomique structurale

Séquençage d'un génome

Techniques de séquençage

le principe de base consiste à fragmenter de façon aléatoire le génome pour obtenir des fragments d'ADN de quelques milliers de pb (facile à manipuler), puis de rassembler les séquences de ces fragments après les avoir séquencés



Séquençage aléatoire

globale (Whole génome shotgun)

Ou encore séquençage du génome entier en pistolet

- rapide et économique
- utilisée depuis 1982 (Bactériophage λ
- retenue pour tous les génomes bactériens

Stratégies de séquençage « clone par clone » (« Shotgun hiérarchique »)

exige la construction préalable ou concomitante d'une carte physique (utilisée pour le **séquençage** du génome humain.. (4etapes)

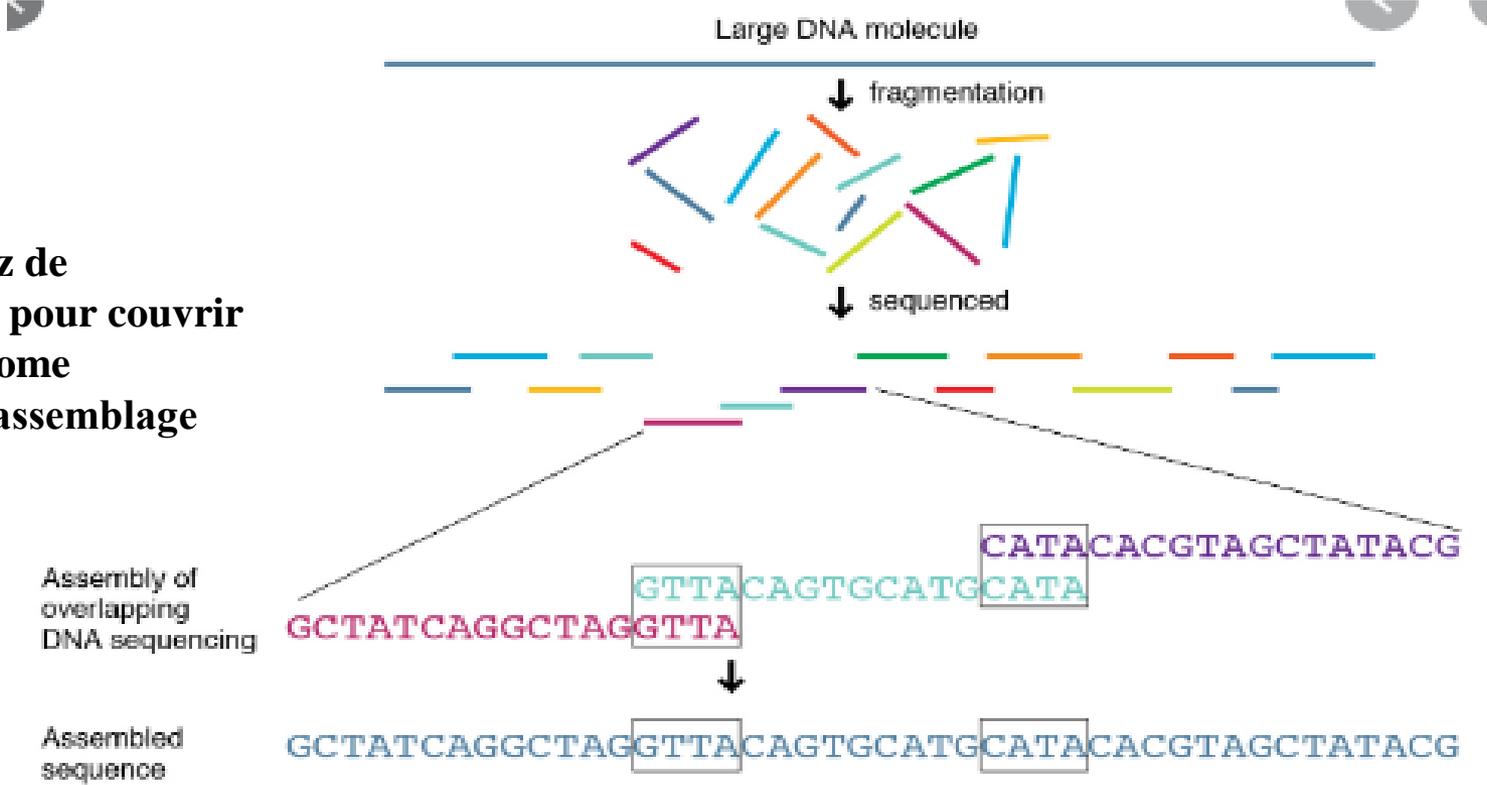
- Etablissement d'une carte physique
- Obtention de clones de taille grande (choix du vecteur)
- Séquençage de chaque clone (type Shotgun)
- Assemblage des clones

Génomique structurale

Séquençage aléatoire globale

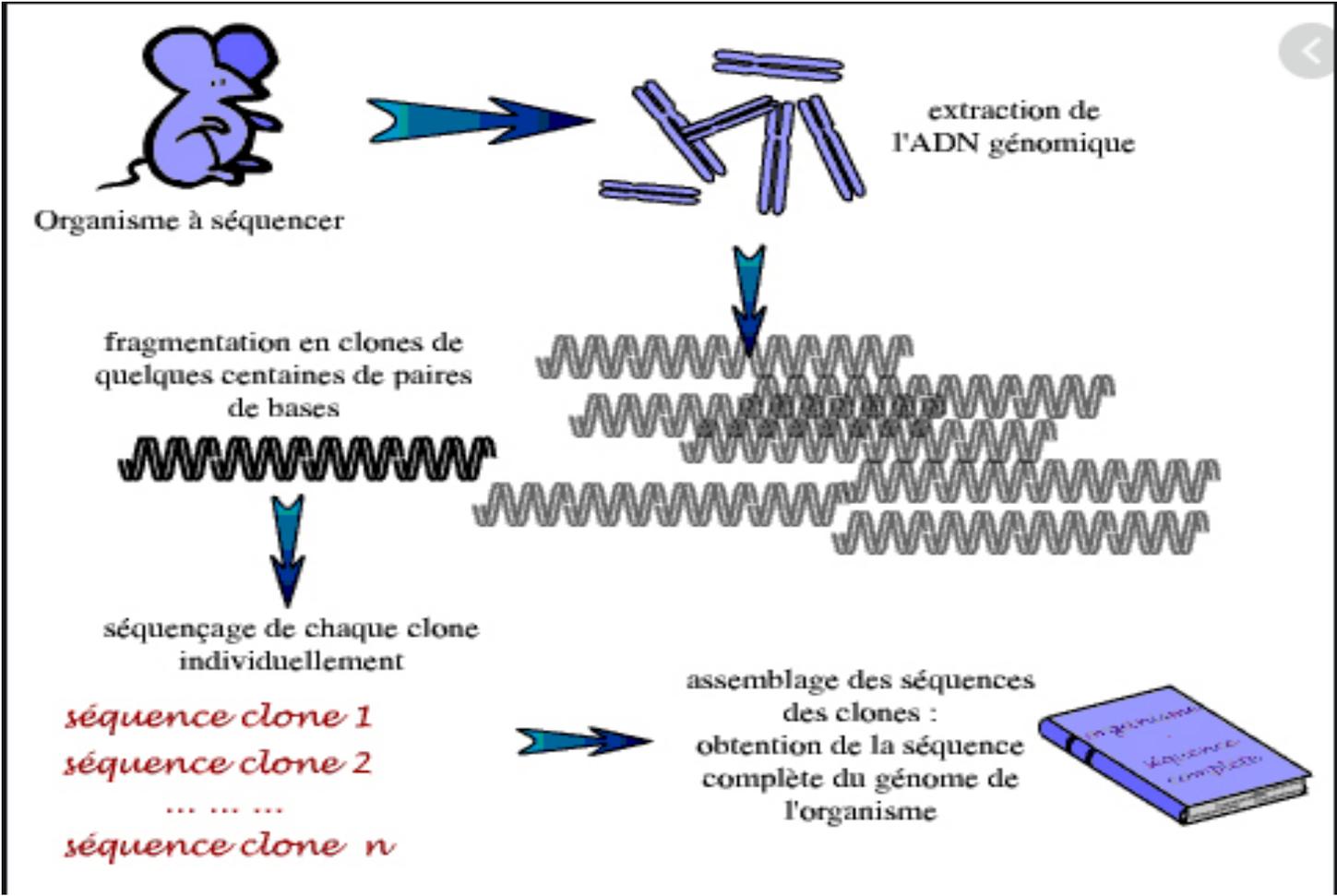
- Découper d'une façon aléatoire le génome en un grand nombre de fragments ayant une petite taille, (amplification par PCR ou par clonage).
- Séquençage des extrémités de ces fragments (Sanger ou par séquençage de nouvelle génération.)
- Assemblage des fragments sur la base de leur chevauchement (programmes d'informatique)

NB/
-avoir assez de
fragments pour couvrir
tout le génome
-Réussir l'assemblage



Génomique structurale

Stratégies de séquençage « clone par clone »



Génomique et annotation d'un génome

Génomique structurale

Stratégies de séquençage « clone par clone »

Construction d'une carte physique



nécessite la construction d'une carte de liaison,



points de repères ordonnés (marqueurs), sur le long du chromosome et sur la mesure de la distance existante entre chaque deux entre eux.

Carte de liaison

Utilisation de marqueurs génétiques polymorphes (cartes génétiques)

Ordonnés sur le chromosome grâce à l'étude des fréquences de recombinaison génétique (liaison génétique).

-1987, Obtention de la première carte génétique humaine (marqueur=RFLP)

--depuis 1990, les microsatellites (marqueurs préférés)

-1996, carte génétique du génome humain, ordonnant 5264 microsatellites

Utilisation des marqueurs moléculaires non nécessairement polymorphes (cartes obtenues par hybrides d'irradiation.)

séquences d'ADN présentes de manière unique dans le génome (*STS : Sequence Tagged Sites*),

-ils sont ordonnés en mesurant la fréquence avec laquelle deux d'entre eux sont séparés par une cassure induite par rayons X.

Génomique et annotation d'un génome

Génomique structurale

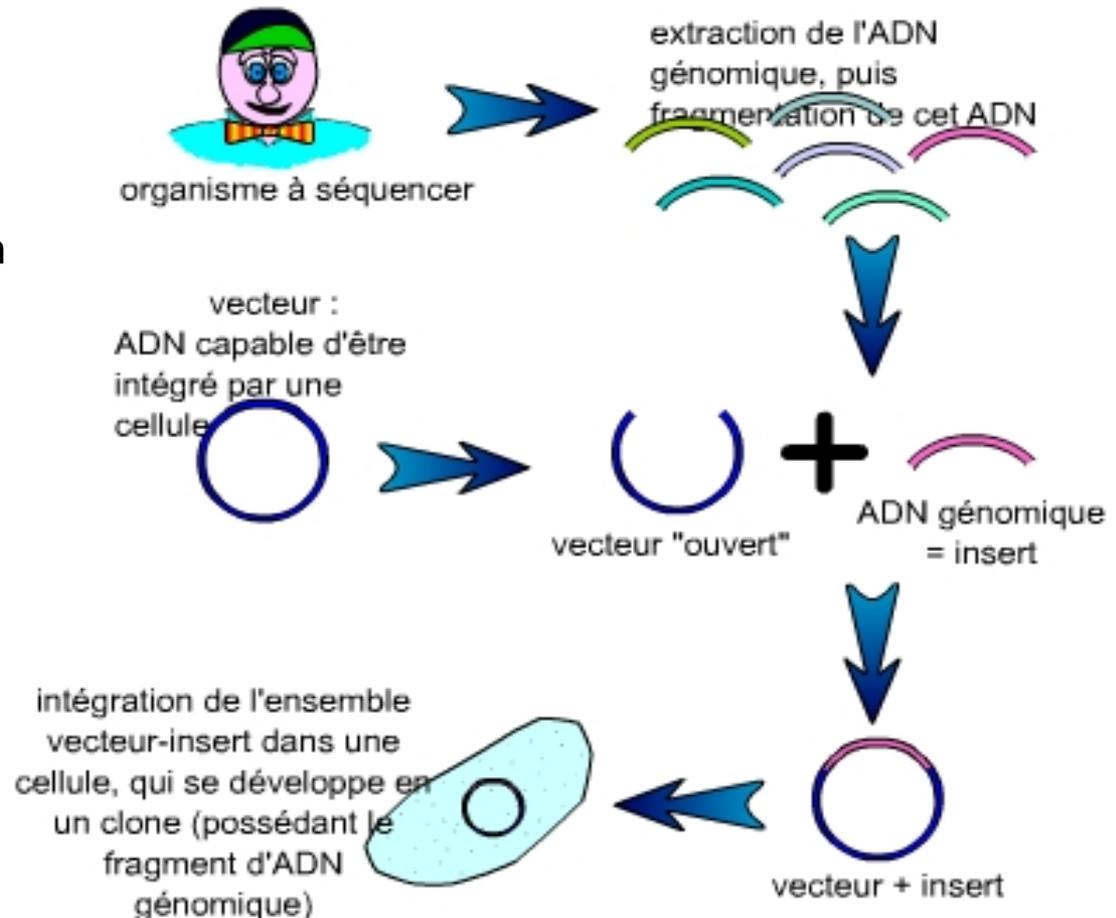
Stratégies de séquençage « clone par clone »

Obtention de clones de grandes tailles

Etape 1: le génome est cassé en fragments de grande taille (Les fragments nécessaires à la réalisation d'une carte physique mesurent en moyenne plus de 100 000 paires de bases (100 kilobases))

Etape 2: clonage des fragments séparément dans des cellules hôtes

NB: le premier problème = trouver des vecteurs supportant des inserts d'une telle taille (plasmides et cosmides → moins de 45kb)



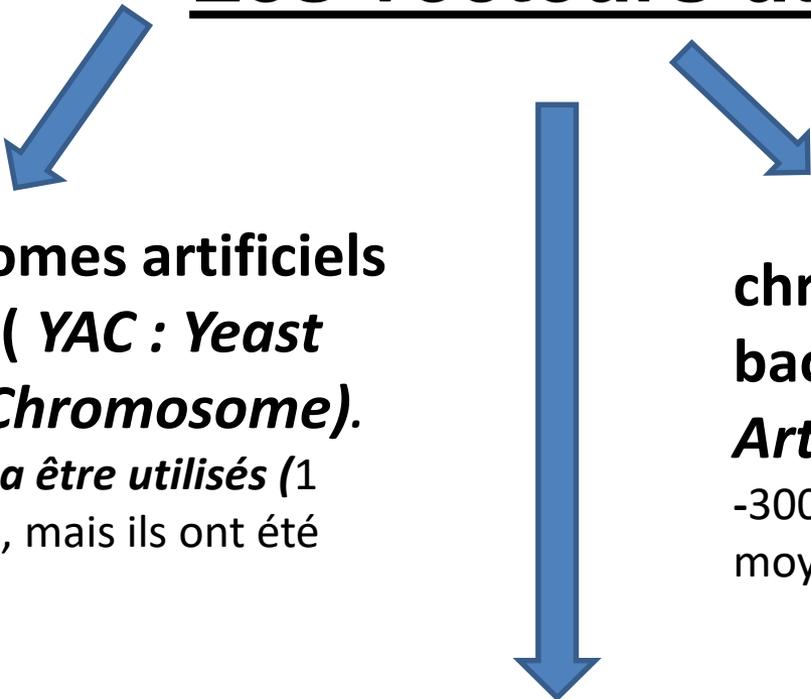
Fabrication d'une banque d'ADN génomique

Génomique et annotation d'un génome

Génomique structurale

Stratégies de séquençage « clone par clone »

Les vecteurs utilisés



chromosomes artificiels
de levure (*YAC : Yeast
Artificial Chromosome*).

Les premiers à être utilisés (1
000 kilobases, mais ils ont été
abandonnés)

chromosomes artificiels
bactériens (*BAC : Bacterial
Artificial Chromosome*).

-300 kilobases au maximum, 150 kilobases en
moyenne)

dérivés du phage P1 (les *PAC*),

-70 à 100 kb

Remarque: Après clonage, chaque clone sera séquencé selon la méthode précédente de séquençage (Whole génome shotgun)

Génomique et annotation d'un génome

Génomique structurale

Stratégies de séquençage « clone par clone »

Ordonnement des clones

Le positionnement des clones les uns par rapport aux autres repose sur l'ordre donné par la carte physique et par la recherche des parties communes (plusieurs méthodes)

Utilisation des profils de restriction

on digère les clones grâce à des enzymes de restriction, puis on recherche si les différents clones présentent des fragments de même taille.

Hybridation des clones entre eux

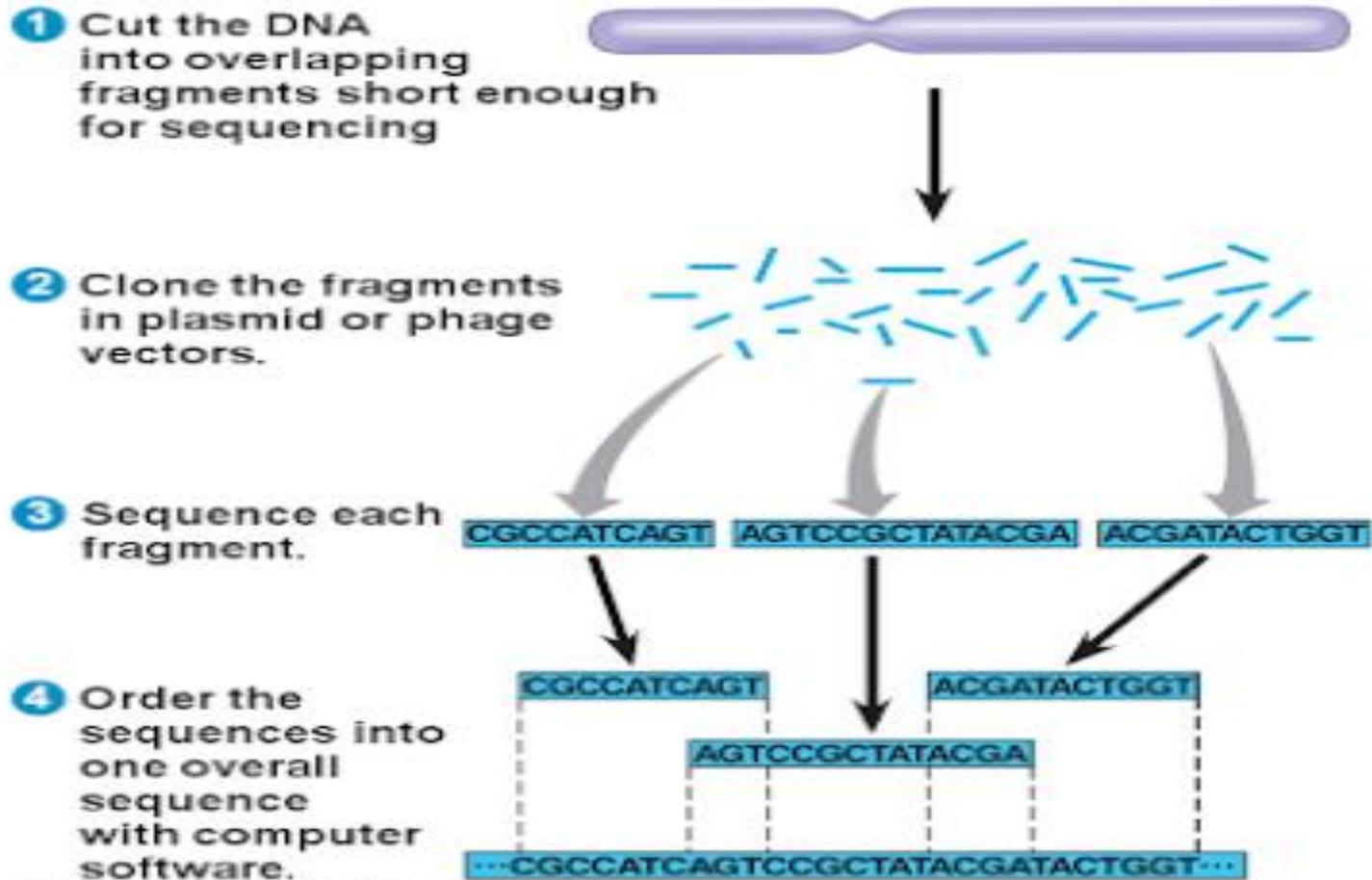
Si Hybridation= séquences communes

Utilisation des *STS* (*Séquence Tagged Sites*)

si deux clones peuvent s'hybrider avec l'une de ces séquences uniques du génome humain, c'est qu'ils possèdent une région commune.

Stratégies de séquençage « clone par clone »

Ordonnement des clones



Copyright © 2009 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

NB: Ordonnement des clones est actuellement facilité via les outils automatisés et bio informatisés

Génomique fonctionnelle



Détermination de la
fonction des gènes



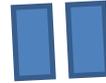
Disruption systématique des
gènes (modification de la
fonction du gène, Knock-outs,
RNAi)



Analyse globale de l'expression génétique

Génomique fonctionnelle

1-Détermination de la fonction des gènes



analyser les protéines données par ces gènes

(plusieurs méthodes se basent sur des programmes de bioinformatiques)



Recherche d'homologie de séquence

Homologie de structure= homologie probable de fonction et donc attribuer une fonction probable au gène correspondant.



Recherche des motifs d'acides aminés caractéristiques

Présence de motifs d'acides aminés caractéristiques de certaines classes de protéines (kinases, protéases...) = attribuer une fonction probable au gène correspondant.

Génomique fonctionnelle

2-Disruption systématique des gènes



comprendre et vérifier le rôle suspecté des gènes

A-Modification de la fonction du gène par mutagénèse

But: avoir ou obtenir un gène qui fonctionne mal



réaliser des mutagénèses soit ciblées soit aléatoires.

Remarque: beaucoup d'inconvénients:

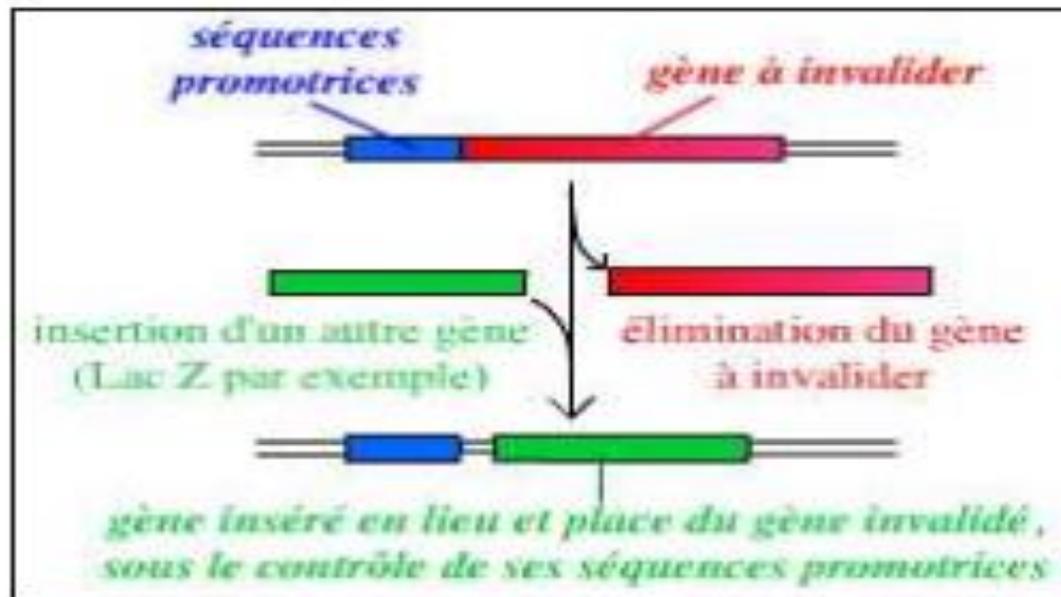
- aucun contrôle sur les mutations qui vont avoir lieu → fastidieux travail de criblage pour rechercher les mutants
- diminution de l'expression du gène ou une diminution de l'efficacité de la protéine codée

Génomique fonctionnelle

2-Disruption systématique des gènes

B- Invalidation d'un gène (ou le KO = knock-out)

But : bloquer l'activité d'une gène par son remplacement dans le génome de l'organisme, par une version modifiée, inactive de ce gène.



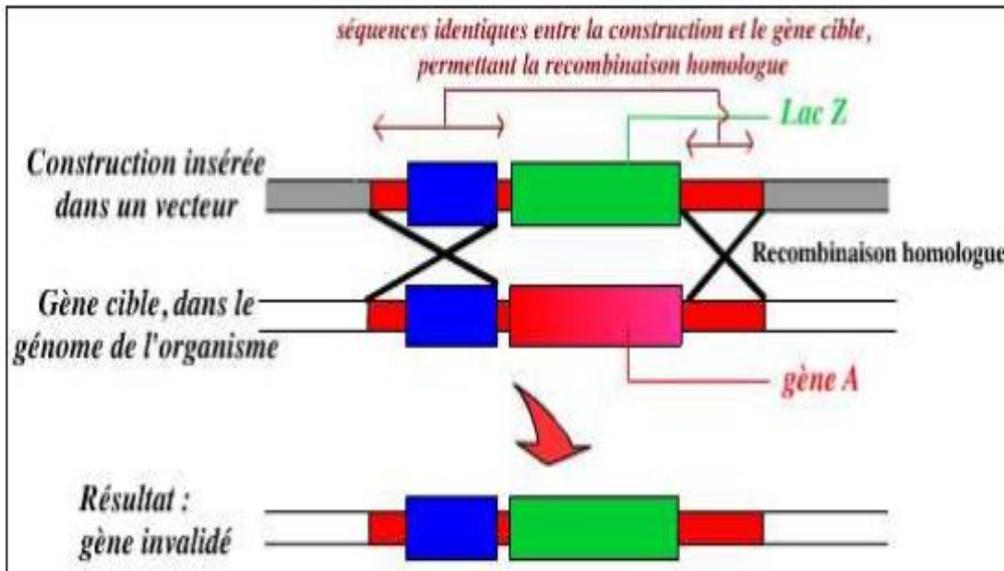
Exemple de méthode pour invalider un gène

Génomique fonctionnelle

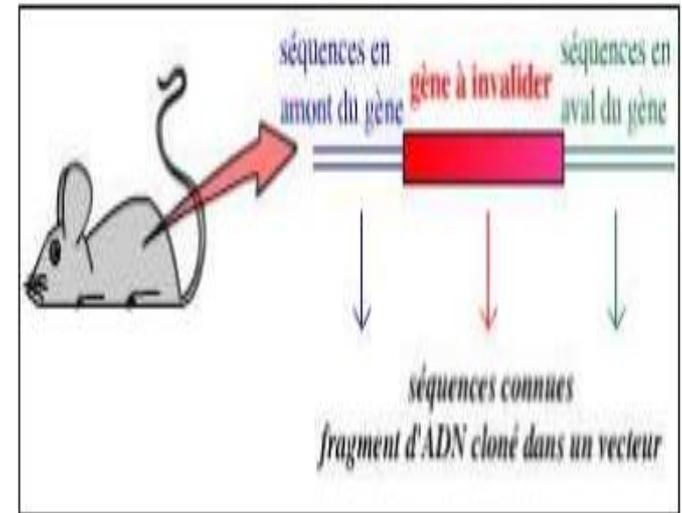
2-Disruption systématique des gènes

B- Invalidation d'un gène (ou le KO = knock-out)

Principe: technique est basée sur la recombinaison homologue



Principe de KNOOUT-OUT



Remarques: Pour invalider un gène,

- il est nécessaire de l' identifier au préalable (connaitre sa séquence)
- Il est souvent aussi nécessaire de disposer de la séquence nucléotidique des fragments d'ADN bordant ce gène dans le génome de l'organisme étudié.

Génomique fonctionnelle

2-Disruption systématique des gènes

C- Le mécanisme RNAi (interférence ARN)

But : réduction ou blocage de l'expression d'un gène

principe

ARNdb  Silençage génique (dégradation de l'ARNm)  voie de l'ARNi

Invivo dans les systèmes **non mammifères**, l'introduction de l'ARN db

La nucléase cytoplasmique (Dicer), clive le long ARNdb en petits ARN interférents (**siRNA**) de **21 à 23 pb**

Les (siRNA), se déroulent et s'assemblent avec les molécules (RISC)

Formation d'un complexe de silencing (siRNA+RISC)

Le brin d'ARNsi antisens guide alors le RISC vers des molécules d'ARN complémentaires

le RISC clive l'ARNm

Silençage génique

Génomique fonctionnelle

2-Disruption systématique des gènes

C- Le mécanisme RNAi (interférence ARN)

Remarques

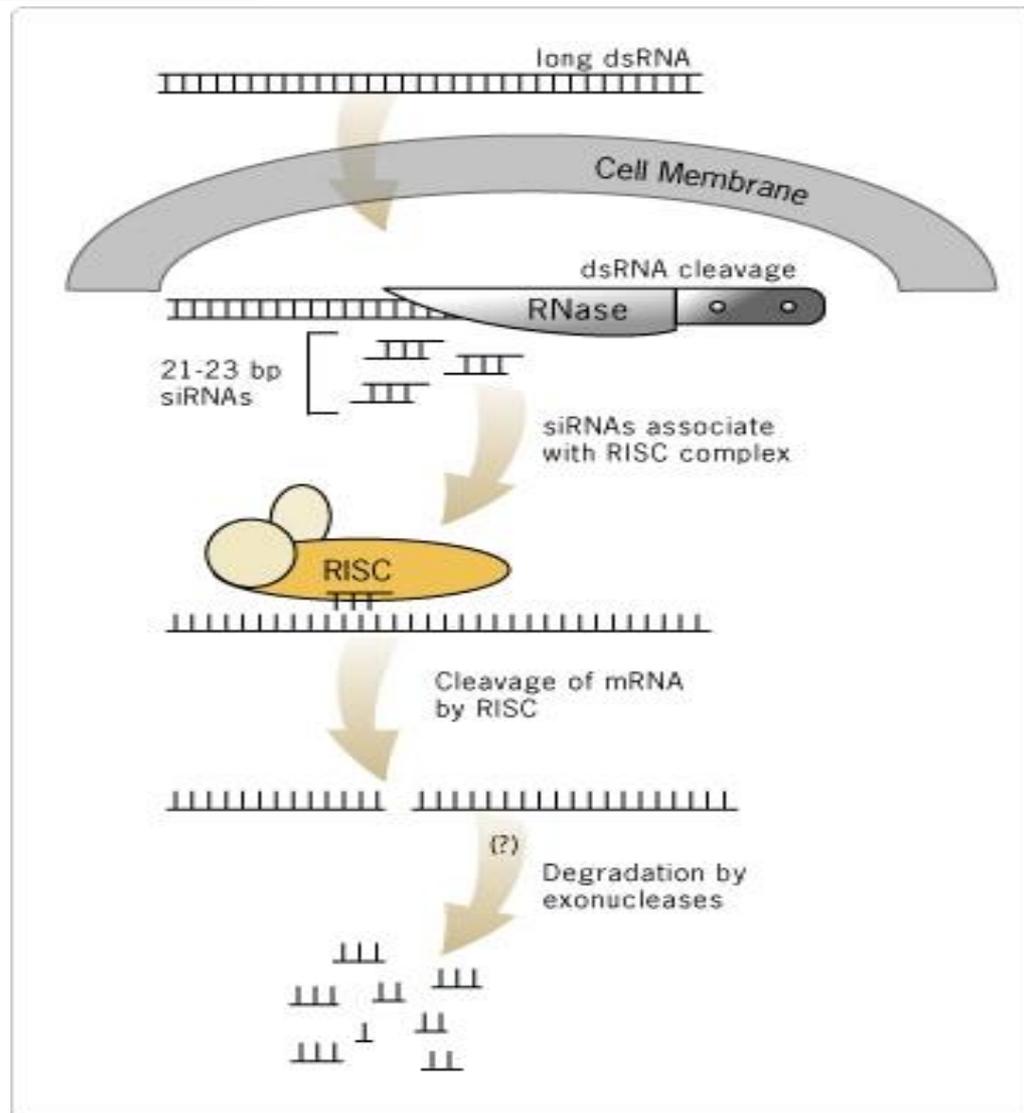
1- Pour l'application du mécanisme RNAi chez les cellules de mammifères il faut introduire l'ARNdb de moins de 30 pb (,transfection des cellules avec des siARN de 21 à 23 pb pour induire l'ARNi

2- les siRNA sont préparés in vitro soit

- Synthèse chimique
- Transcription in vitro
- Digestion de long dsRNA in vitro par RNase III ou Dicer

3- Les siARN sont Introduits dans des vecteurs ou de cassettes à base d'ADN

- Expression à partir d'un plasmide
- Expression d'un vecteur viral



Génomique fonctionnelle

3-Analyse globale de l'expression génétique



(Génomique expression elle)

L'expression génique consiste à:

caractériser et quantifier les produits d'expression de l'ADN (ARNm, protéines) de manière à identifier, dans un tissu, dans un état et à un moment donné du développement, les séquences actives et donc à révéler ainsi le niveau d'expression des gènes dont elles sont issues.



Cette partie sera plus détaillée dans le chapitre suivant:

Analyse de l'expression d'un gène