



## Support de cours

# Épigénétique

Destiné aux étudiants de Licence 3

Spécialité Génétique

Docteur *REZGOUN Mohamed Larbi*  
Maître de conférences (A)

Année universitaire  
2020 - 2021

# Table des matières

Liste des abréviations

Liste des figures

Liste des tableaux

Préambule

<b>Chapitre I : Définition de l'épigénétique</b> .....	<b>Page 01</b>
1- Naissance du concept d'épigénétique .....	01
2- Définition de l'épigénétique .....	01
3- Information épigénétique .....	02
4- Définition du concept de l'épigénome .....	03
<b>Chapitre II : Organisation structurale et fonctionnelle de la chromatine</b> .....	<b>05</b>
1- Découverte de la chromatine .....	06
2- Composition de la chromatine .....	06
2-1- Protéines histones .....	06
2-2- Variants histones .....	08
2-3- Protéines non histones .....	09
3- Structure de la chromatine .....	10
3-1- Nucléosome et la fibre F10 .....	10
3-2- Fibre F30 .....	12
3-3- Boucles chromatiniennes .....	13
3-4- Quatrième niveau de condensation .....	14
4- Différentes étapes de condensation de la chromatine .....	15
4-1- Assemblage du nucléosome et formation de la F10 .....	15
4-2- Formation des niveaux d'organisation supérieurs de la chromatine .....	16
5- Territoires chromatiniens .....	19
<b>Chapitre III : Structures chromosomiques spécialisées</b> .....	<b>21</b>
1- Télomères des eucaryotes supérieurs .....	21
1-1- Historique et définition .....	21
1-2- Structure des télomères .....	22
1-3- Réplication des extrémités télomériques et télomérase .....	23
1-4- Fonctions des télomères .....	25
2- Centromères .....	27
2-1- Historique et définition .....	27
2-2- Structure des régions centromériques .....	28
2-3- Paradoxe du centromère .....	31
<b>Références bibliographiques</b> .....	<b>32</b>

# Liste des abréviations

**ADN** : Acide Désoxyribo-Nucléique

**ALT** : Alternative Lengthening of Telomeres

**ARN** : Acide Ribo-Nucléique

**ARNr** : ARN ribosomiques

**ARNt** : ARN de transfert

**ASF1** : Anti-Silencing Function 1

**ATP** : Adénosine Tri-Phosphate

**ATPase** : Adenosine Tri-Phosphatase

**CAF1** : Chromatin Assembly Factor 1

**kDa** : kilo-Daltons

**Mb** : Méga bases

**miRNA** : microRNA

**NAP1** : Nucleosome Assembly Protein 1

**pb** : paires de bases

**PCNA** : Proliferating Cell Nuclear Antigen

**PM** : Poids Moléculaire

**POT1** : Protection Of Telomere 1

**RAP1** : Repressor Activator Protein 1

**RC** : Région Centromérique

**SAR** : Scaffold Attachment Region

**SINES** : Short Interspersed Nuclear Element

**SMC** : Structural Maintenance of Chromosome

**TIN2** : TRF1 Interacting factor 2

**TPP1** : TriPeptidyl Peptidase 1

**TRF** : Telomeric Repeat Binding Factor 1

# Liste des figures

<b>Figure 01 :</b>	Épigénome, un second niveau d'information .....	<b>04</b>
<b>02 :</b>	Représentation schématique de la structure des histones majeurs .....	<b>07</b>
<b>03 :</b>	Cliché en microscopie électronique du nucléo-filament .....	<b>10</b>
<b>04 :</b>	Structure de l'histone cœur du nucléosome .....	<b>11</b>
<b>05 :</b>	Structure du nucléosome et de la fibre F10 .....	<b>12</b>
<b>06 :</b>	Passage de la F10 à la F30 .....	<b>13</b>
<b>07 :</b>	Modèles de structure de la F30 .....	<b>13</b>
<b>08 :</b>	Représentation schématique de la structure des boucles chromosomiques .....	<b>14</b>
<b>09 :</b>	Cliché en microscopie électronique de chromosomes métaphasiques .....	<b>15</b>
<b>10 :</b>	Cliché en microscopie électronique d'un chromosome sans histones .....	<b>17</b>
<b>11 :</b>	Représentation schématique de structure des condensines .....	<b>18</b>
<b>12 :</b>	Représentation schématique d'un noyau interphasique coloré au Feulgen .....	<b>20</b>
<b>13 :</b>	Représentation schématique de la structure des télomères .....	<b>22</b>
<b>14 :</b>	Modèle actuel du mécanisme de réplication des télomères .....	<b>24</b>
<b>15 :</b>	Classification des chromosomes en fonction de l'indice centromérique .....	<b>27</b>
<b>16 :</b>	Organisation des RCs entre les espèces .....	<b>29</b>
<b>17 :</b>	Modèle de structure secondaire de la chromatine centromérique .....	<b>30</b>

# Liste des tableaux

<b>Tableau I :</b>	Caractéristiques biochimiques des histones .....	<b>07</b>
--------------------	--	-----------

## **Préambule**

Le présent polycopié est destiné aux étudiants de troisième année Licence de la formation académique Génétique. Il peut constituer également un point de départ pour tout biologiste désireux de s'initier à l'épigénétique. Seuls sont présentés les principes et les concepts de base de cette nouvelle discipline très large, en s'appuyant le plus possible sur des données actualisées.

L'importance du principe épigénétique est mise en évidence du fait que toutes les cellules d'un organisme vivant partagent un génome identique. Cependant, elles n'ont pas toutes la même morphologie et ne contribuent pas aux mêmes fonctions biologiques. Cette hétérogénéité est le fruit d'une expression différente du génome en fonction du type cellulaire. La machinerie qui se met en place pour réguler cette expression fait appel à des mécanismes moléculaires complexes appartenant au domaine de l'épigénétique. Le terme épigénétique définit aujourd'hui les modifications transmissibles et réversibles de l'expression des gènes et ne s'accompagnant pas de changements des séquences nucléotidiques. Ce type de régulation peut cibler l'ADN ou l'ARN et agir au niveau du noyau ou du cytoplasme. La transmission d'un état épigénétique peut être réalisée par la méthylation de l'ADN, par des modifications post-traductionnelles des histones, et par la déposition de variants d'histones. De plus, il devient de plus en plus évident que les mécanismes médiés par les petits ARN non codants jouent un rôle central dans la mise en place et le maintien de l'état d'activité des gènes.

Afin de simplifier la lecture de ce polycopié et de susciter chez le lecteur des questions et des réflexions, son attention est attirée par une organisation en plusieurs chapitres disposés selon un enchaînement cohérent. Dans le premier chapitre, nous avons essayé de donner une définition très simplifiée du concept d'épigénétique et d'intégrer la terminologie particulière utilisée dans cette discipline au bagage linguistique de l'étudiant.

Pour qu'un gène puisse conduire à la synthèse d'une protéine ou d'un ARN fonctionnel, il doit être lisible, c'est-à-dire accessible à différents complexes protéiques qui interviennent dans ce processus. Les marques apposées sur les histones modifient l'état de compactage de la molécule d'ADN, favorisant ou au contraire limitant l'accessibilité aux gènes. Pour comprendre cet aspect, nous avons consacré le chapitre II à la description de la structure de la chromatine. En complément, nous décrirons dans le chapitre III, les structures chromosomiques spécialisées (centromères et télomères) indispensables au fonctionnement du chromosome comme support de l'information héréditaire.

# Chapitre I

## Définition de l'épigénétique

## 1- Naissance du concept d'épigénétique

Chacune de nos cellules renferme l'ensemble de notre patrimoine génétique. Mais, si toutes nos cellules contiennent la même information génétique, elles n'en font manifestement pas toutes le même usage : une cellule de la peau ne ressemble en rien à un neurone, une cellule du foie n'a pas les mêmes fonctions qu'une cellule du cœur. De même, deux jumeaux qui partagent le même génome ne sont jamais parfaitement identiques ! À travers ces exemples et dans bien d'autres, la clef du mystère se nomme « épigénétique ». L'organisme humain est composé de  $10^{13}$  cellules organisées en 106 types cellulaires différenciées ou spécialisés (épithéliales, musculaires, nerveuses ...etc.). Toutes ces cellules sont issues d'une seule cellule ; l'œuf fécondé, qui, par des mitoses successives, va donner tous les types cellulaires. Autrement dit : toutes les cellules qui composent un organisme sont génétiquement identiques car issues de mitoses ; processus conservateur de l'information génétique qui génère à partir d'une cellule mère deux cellules filles identiques entre elles et identiques par rapport à la cellule mère préexistante. Comment expliquer alors les variations de forme, de physiologie et de fonction qui existent entre les cellules différenciées alors qu'elles possèdent le même génome ?

La réponse réside dans le contrôle de l'expression des gènes, domaine de la biologie qu'on nomme épigénétique. Aujourd'hui il est admis que les cellules diffèrent n'ont pas du fait qu'elles contiennent des gènes différents, mais du fait qu'elles activent ou répriment l'expression de gènes différents. Alors que la génétique correspond à l'étude des gènes, l'épigénétique s'intéresse à une « couche » d'informations complémentaires qui définit comment ces gènes vont être utilisés par une cellule... ou ne pas l'être.

## 2- Définition de l'épigénétique

Venant du mot épigénèse, théorie selon laquelle l'embryon est indifférencié lors des phases précoces du développement, le terme « épigénétique » qui signifie littéralement « sur », « au-dessus de » ou « ce qui couvre » la génétique, a été introduit pour la première fois par *Conrad Waddington* en 1942, à une époque où on n'avait pas encore établi la nature Acide Désoxyribonucléique (ADN) du support de l'information génétique, pour décrire la discipline de la biologie qui étudie « les interactions des gènes avec leur environnement pour définir le phénotype des êtres ».



De nos jours, le terme épigénétique peut être défini comme l'étude des modifications transmissibles et réversibles du profil d'expression des gènes, autrement dit la manière dont un gène est exprimé, qui ne s'accompagnent pas de changements dans la séquence d'ADN. La génétique, telle qu'elle a été pratiquée pendant de nombreuses années, se focalisait sur les modifications transmissibles et irréversibles de la séquence de l'ADN, comme les mutations, n'a effectivement pas suffi à élucider certains des mystères du vivant. L'épigénétique est donc une nouvelle science capable de combler les insuffisances de la génétique mendélienne classique. Elle représente une vision complémentaire du contrôle de l'information génétique et désigne la grande diversité des mécanismes modulant l'expression des gènes. L'information épigénétique est donc une « extension » de l'information génétique.

### **3- Information épigénétique**

Les modifications épigénétiques sont induites par l'environnement au sens large, à l'échelle cellulaire et celui de l'organisme : la cellule reçoit en permanence toutes sortes de signaux l'informant sur son environnement, de manière à ce qu'elle se spécialise au cours du développement, ou ajuste son activité à une situation particulière. Ces signaux, peuvent conduire à des modifications dans l'expression de nos gènes, sans pour autant affecter leur séquence. Le phénomène peut être transitoire, mais il existe des modifications épigénétiques durables, qui persistent même lorsque le signal qui les a induites disparaît. Concrètement, ces modifications sont matérialisées par des marques biochimiques, apposées par des enzymes spécialisées sur l'ADN ou sur des protéines qui le structurent (histones) et parfois même en faisant appel à des molécules d'Acide Ribonucléique (ARN) particulières qu'on appelle ARN non codant (ARNnc). Alors que l'information génétique s'exprime sous une seule forme ; une séquence d'ADN, l'information épigénétique s'exprime sous quatre formes possibles :

- modification de la structure de l'ADN par méthylation des cytosines,
- modifications post-traductionnelles histones par fixation covalente de différents groupements chimiques (acétylation, méthylation, phosphorylation, ubiquitinylation, ribosylation, etc.),
- incorporation de variants d'histones (H2A-X, H2A.Z, H3.3, etc.).
- régulations médiées par les petits ARN non-codant.

Un phénomène épigénétique peut être régulé par une ou plusieurs formes de cette information.

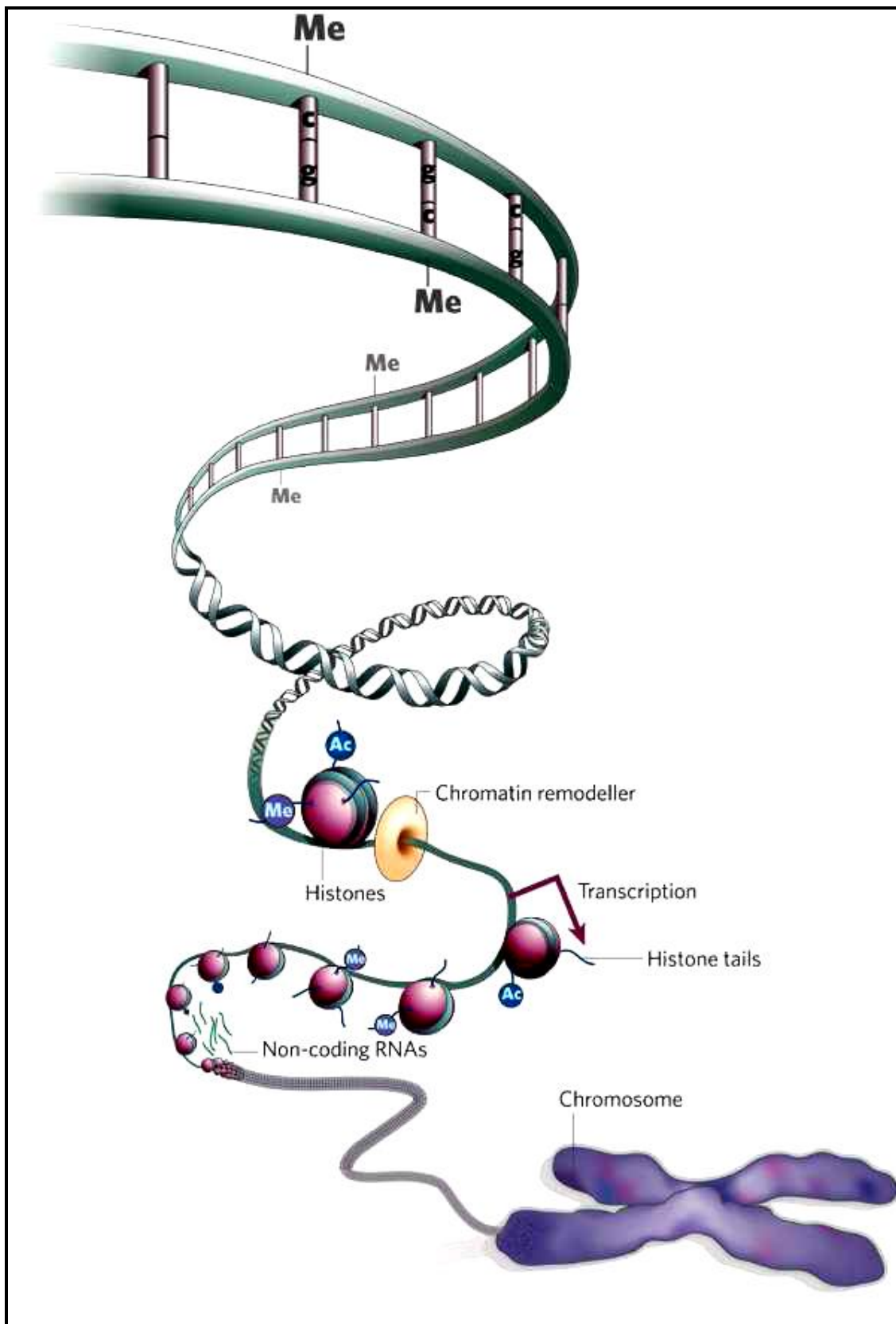
#### 4- Définition du concept de l'épigénome

De nos jours, l'épigénétique fait référence à la fois aux changements héréditaires de l'activité et de l'expression des gènes de la descendance d'une cellule ou d'un individu, et aux changements stables sur le long terme du programme transcriptionnel d'une cellule, sans pour autant être nécessairement transmis aux cellules filles. Ainsi, l'épigénétique représente l'ensemble des mécanismes permettant à une cellule de faire face aux différents événements de sa vie (différenciation, stress, etc...) en modulant son programme d'expression génétique. Les modifications de l'information épigénétique sont à la fois stables et modulables par un grand nombre de facteurs cellulaires, reflétant un état physiologique ou pathologique de la cellule ainsi qu'une adaptabilité à son environnement. En comparaison du génome qui représente l'ensemble des gènes d'une cellule, l'épigénome représente l'ensemble des marques et modifications épigénétiques d'une cellule. Cet épigénome est propre à chaque cellule à un instant  $t$ , et influe directement sur le génome et donc sur le programme génétique utilisé par la cellule.

En conclusion, un épigénome est défini comme étant : **l'ensemble des modifications épigénétiques propre à chaque type cellulaire différencié à un instant  $t$ .**

Il est important de noter qu'il existe une grande connectivité entre ces différentes informations épigénétiques : de nombreux travaux mettent en évidence une synergie et un enchaînement de modifications épigénétiques pour arriver *in fine* à la régulation spécifique de l'expression du génome. Ces différents mécanismes font appel à toute une batterie de facteurs et de complexes :

- « **écrivains** » et « **effaceurs** » : capables de modifier les marques épigénétiques,
- « **lecteurs et recruteurs** » : capables de reconnaître une ou plusieurs marques épigénétiques et de cibler ainsi des régions spécifiques de la chromatine afin d'y recruter différents complexes protéiques,
- « **effecteurs** » : capables de remanier la chromatine et/ou d'entraîner différents processus biologiques.



**Figure 01 : Épigénome, un second niveau d'information (Jones *et al.*, 2008).**  
 (La troisième forme de l'information épigénétique ou variants histones, n'est pas mentionné dans cette figure)

# Chapitre II

Organisation structurelle et  
fonctionnelle de la chromatine

La chromatine est la forme sous laquelle se présente l'ADN dans le noyau. C'est la substance de base des chromosomes eucaryotes. Elle correspond à l'association de l'ADN et de protéines structurales et fonctionnelles spécifiques.

À titre d'exemple, l'ADN contenu dans un noyau d'une cellule humaine est composé d'environ 50 000 gènes, repartis au sein de 23 paires de chromosomes pour un total de 6,4 milliards de paires de bases (pb). L'ADN est une macromolécule de taille conséquente, d'environ 2 mètres de long mis bout à bout, pour 2,50 nm de large, et tout le génie d'une cellule réside dans sa capacité à maintenir une si grande quantité d'ADN au sein de l'espace si restreint d'un noyau cellulaire. Une image couramment utilisée pour se représenter une échelle de taille est que si l'on compare le noyau d'une cellule à une orange, l'ADN replié à l'intérieur peut être assimilé à une pelote d'un long fil de pêche d'environ 20 kilomètres. La cellule doit donc structurer son ADN sans s'emmêler, tout en restant capable d'accéder à n'importe quel moment à l'information codée par un gène précis. Ceci réclame une étroite collaboration entre l'ADN et les protéines structurantes de la chromatine, afin de permettre l'accès en temps voulu à la machinerie cellulaire pour diverses activités telles que la réplication, la transcription, la réparation, etc... On peut donc attribuer deux fonctions à la chromatine :

- **Une fonction de compaction de l'ADN :** la chromatine permet en effet la compaction d'un polymère d'ADN d'environ deux mètres, dans un espace nucléaire dont les dimensions ne dépassent pas quelques micromètres dans le cas de l'homme. La chromatine est très organisée dans le noyau, et cette organisation résulte de différents niveaux hiérarchiques de compaction. Deux questions auxquelles on va essayer de répondre dans ce chapitre : comment arriver à ce niveau de compaction ? et comment garder quand même une grande activité en ce qui concerne la réplication, la transcription et la réparation ?
- **Une fonction dynamique qui permet la régulation de l'expression des gènes :** des marques apposées sur les histones modifient l'état de compactage de la molécule d'ADN, favorisant ou au contraire limitant l'accessibilité aux gènes.

Quelle est la structure de la chromatine et comment est-elle modulable afin d'apporter une information supplémentaire à celle de l'ADN ?

## 1- Découverte de la chromatine

La substance nucléaire ou chromatine, a été décrite pour la première fois par *Miescher* en 1869 est nommée ainsi par *Fleming* en 1882. L'étymologie du mot chromatine est « corps colorés ». Elle est disposée dans le noyau de la cellule sous forme de réseaux de filaments enchevêtrés sans organisation apparente. Il semble que ce soit *Sutton* en 1903 et *Boveri* en 1904 qui proposèrent pour la première fois d'associer les gènes aux chromosomes qui deviendraient ainsi supports de l'hérédité. Ces deux chercheurs étaient intrigués par le comportement des chromosomes lors de la mitose décrit par *Flemming* en 1879. Au début de 20<sup>ème</sup> siècle se développèrent les observations cytologiques par l'utilisation de techniques de colorations spécifiques de la chromatine, dont la première était le Feulgen : immersion dans du HCl suivit d'une coloration par la fuschine, la technique de coloration la plus répandue de la chromatine utilise le Giemsa qui est un mélange de deux colorants : le bleu de méthylène à caractère basique et l'éosine à caractère acide. La chromatine présente également une très grande affinité pour l'hématoxyline qui la colore en noir, le bleu de méthylène en bleu, le bleu de toluidine et pyronine qui lui donne une teinte bleu-violacée.

## 2- Composition de la chromatine

La chromatine est composée d'acides nucléiques (ADN et ARN) et de protéines. Les protéines sont de deux types : essentiellement les histones à caractère basique, et les protéines non histones dites « acides ». L'ARN est représenté essentiellement par des chaînes d'ARNm naissantes qui sont en transit vers le cytoplasme où ils seront traduits.

### 2-1- Protéines histones

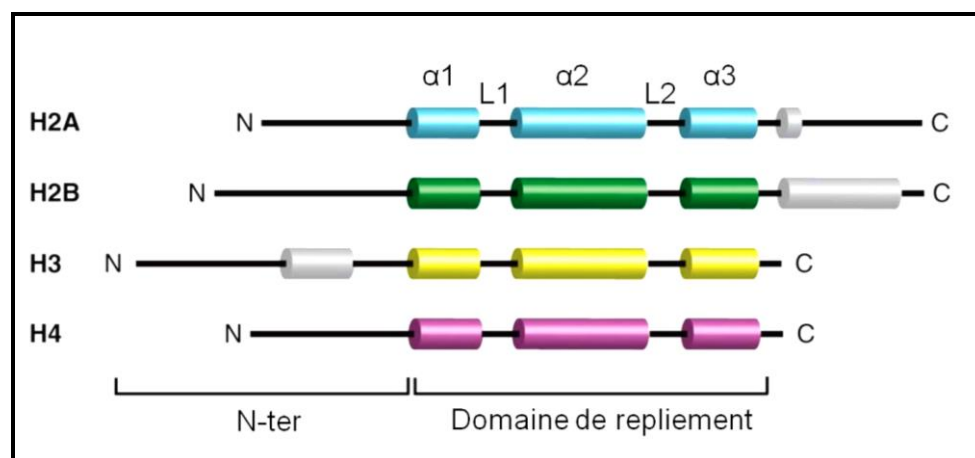
Petites protéines nucléaires d'environ 120 résidus, à caractère basique riches en acides aminés arginine (R) et en lysine (K) qui représentent plus de 25% de leurs compositions, ils possèdent ainsi de nombreuses chaînes latérales chargées positivement qui se lient aux groupements phosphates de l'ADN chargés négativement ; l'ADN est un polyanion possédant une charge négative globale due à la fonction acide libre de l'acide phosphorique. Leurs poids moléculaires (PM) varient de 11 à 23 kilodaltons (kDa). Cinq classes d'histones ont été caractérisées en fonction de leurs proportions relatives en acides aminés basiques : H1, H2A, H2B, H3 et H4 : H3 et H4 riches en Arginine, H2A et H2B légèrement riches en Lysine et la H1 très riche en Lysine. Récemment, on a pu isoler H5 des érythrocytes aviaires, H6 dans les cellules germinales de poissons, ainsi que de nombreux variantes histones.

Les histones sont les protéines liées à l'ADN les plus abondantes dans les cellules eucaryotes et parmi les protéines connues les plus conservées au cours de l'évolution. Les histones H2A, H2B, H3 et H4 présentes sont appelées histones majeures, histones cœur, ou réplication dépendante. Cependant, l'histone H1, appelée histone inter-nucléosomale ou histone de liaison, présente des caractéristiques fondamentalement différentes des autres classes d'histones (**tableau I**).

**Tableau I** : Caractéristiques biochimiques des histones (**Ridgway, 2006**).

Histones	Acides aminés basiques		Acides aminés Acides	Rapport basiques / acides	PM (Da)
	Lysine	Arginine			
<b>H1</b>	29 %	1 %	5 %	<b>5,4</b>	23000
<b>H2A</b>	11 %	9 %	15 %	<b>1,4</b>	13960
<b>H2B</b>	16 %	6 %	13 %	<b>1,7</b>	13774
<b>H3</b>	10 %	13 %	13 %	<b>1,8</b>	15342
<b>H4</b>	11 %	14 %	10 %	<b>2,5</b>	11282

Les histones majeures possèdent un domaine central hydrophobe de repliement caractéristique des histones appelé « motif histone-fold » formé de 70 acides aminés environ. Ce domaine est composé de trois hélices  $\alpha$  reliées entre elles par deux boucles L1 et L2, ainsi que de deux queues flexibles aux extrémités N-terminales (N-ter) et C-terminales (C-ter) linéaires dépourvues de structure secondaire. Ce domaine commun à toutes les histones assure leur dimérisation selon une configuration dite en poignée de main (**figure 02**).



**Figure 02** : Représentation schématique de la structure tridimensionnelle des histones majeures (**Young, 2014**).

Les histones H3 et H4 sont les protéines les plus conservées au cours de l'évolution chez toutes les espèces eucaryotes. La région la plus conservée de ces histones est leur domaine central. Par contre, les extrémités N-ter et C-ter, sont plus variables et sont dépourvues de structure secondaire. Ces extrémités sont particulièrement riches en résidus lysine et arginine et sujettes à de nombreuses modifications post-traductionnelles. Les H2A et H2B, de structures similaires à H3 et H3, présentent néanmoins des variations interspécifiques non négligeable.

L'histone de liaison H1 présente quant à elle une structure radicalement différente des autres classes d'histones. Elle possède le PM le plus élevé, très riche en lysine ; c'est la plus basique de toute. Elle possède un domaine globulaire hydrophobe GH1 très conservé, aussi appelé « wing hélice », composé de trois hélices  $\alpha$  et d'un feuillet  $\beta$  responsable de la liaison à l'ADN. Les extrémités N-ter et C-ter, très peu conservées, sont impliquées dans de nombreuses interactions protéine-protéine. Cette histone, au niveau de ces extrémités, présente des variations importantes entre tissu et espèces. La H1 ne présentent pas d'homologie structurale avec les autres histones et doit son nom du fait qu'elle est associée *in vivo* avec les nucléosomes dans un rapport de 1/1.

Les gènes des histones sont transcrits en phase S du cycle cellulaire de manière synchrone avec la réplication de l'ADN, elle est toujours étroitement couplée avec celle de l'ADN, cette synthèse est rapide puisqu'on ne trouve jamais de segments d'ADN nu au niveau des fourches de réplication : l'inhibition de la synthèse des histones diminue le taux de production de l'ADN et réciproquement.

Les gènes des histones se situent au niveau des chromosomes : 1, 6 et 12, très répétés, de 10 à 1000 copies suivant les espèces, ce qui reflète leurs importances pour la survie de l'organisme. Ces gènes sont regroupés en tandems mais transcrits séparément. La plupart des gènes des histones sont sans introns et leurs ARNm ne portent pas de queue de poly A.

## 2-2- Variants histones

Les histones font partie des protéines les mieux conservées au cours de l'évolution. Étant codées par de multiples allèles, les histones possèdent de nombreux variants définis comme étant des isoformes non-alléliques d'une histone canonique (dite aussi typique) qui diffèrent par leurs séquences. Il existe plusieurs variants différents pour chaque sous-type d'histone, excepté pour l'histone H4. Actuellement, ce ne sont pas moins de 55 variants uniques d'histones qui ont été décrits chez l'Homme.



Les histones sont ainsi classées en deux groupes majeurs : les histones canoniques et les variants d'histones, encore appelés histones de remplacement. Au-delà de leur séquence en acide aminés, la distinction entre ces deux groupes se fait selon que leur incorporation au sein de la chromatine est dépendante ou non de la réplication de l'ADN. Les histones canoniques sont exclusivement exprimées au cours de la phase S du cycle cellulaire et incorporées dans la chromatine pour répondre à un besoin important en histones lors de la réplication de l'ADN. Les variants d'histones peuvent, quant à eux, être synthétisés tout au long du cycle cellulaire et sont, la plupart du temps, incorporés dans la chromatine indépendamment de la synthèse d'ADN. Par ailleurs, les pré-ARNm des variants d'histones sont polyadénylés et contiennent des introns et des exons. De plus, il est important de noter qu'il existe des isoformes canoniques d'histones H2A, H2B, H3 et H1 qui ne diffèrent parfois que d'un seul acide aminé.

L'incorporation des différents variants d'histones a une répercussion directe sur les interactions entre histones au sein des nucléosomes. Ce phénomène peut influencer la stabilité des nucléosomes et avoir des conséquences sur la conformation de la chromatine, et sur sa perméabilité à la transcription. Certains variants jouent même directement un rôle dans divers processus cellulaires (mitose, réparation de l'ADN, etc.). Par conséquent, le remplacement des histones canoniques par des variants d'histones représente à lui seul un niveau d'information épigénétique. Des variants d'histones ont été découverts chez plusieurs espèces eucaryotes, différant de façon plus ou moins importantes des protéines histones conventionnelles au niveau de leurs séquences mais également dans leur assemblage au sein des nucléosomes. Ces différences ont une répercussion au niveau de la fonction de ces variants, et présentent des rôles spécifiques dans de nombreux processus biologiques, tels que l'organisation centromérique (H3 CenpA), la réparation de l'ADN, l'inactivation du chromosome X (macro-H2A) ou encore la transcription.

### **2-3- Protéines non histones**

Sous l'appellation générique de protéines non histones, dite aussi protéines « acides » on regroupe toutes les protéines nucléaires à l'exception des histones. Il en existe plusieurs et jouent de nombreux rôles :

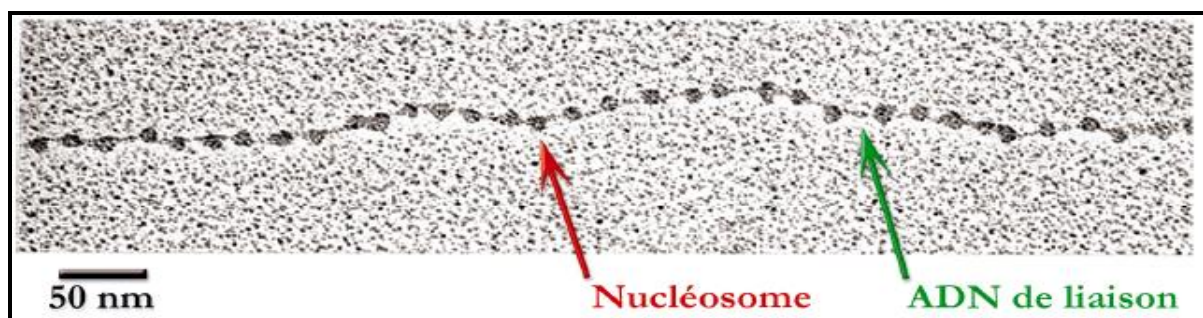
- protéines de structure contribuant à la formation du squelette chromosomique.
- protéines intervenant dans la réplication de l'ADN, sa transcription et sa réparation.
- protéines jouant le rôle de facteurs d'assemblage et de remodelage de la chromatine comme les protéines chaperonnes.

### 3- Structure de la chromatine

L'ADN contenu dans le noyau est fortement compacté. Cette compaction de l'ADN résulte de plusieurs niveaux d'organisation. Le premier niveau de repliement est fourni par la double hélice d'ADN elle-même. Puis, l'enroulement de l'ADN autour d'un cœur protéique va constituer ce que l'on appelle un nucléosome. La répétition de cette structure tout au long de l'ADN donnera à la chromatine l'aspect d'un « collier de perles » d'une épaisseur de 11 nm. Dans un second temps, cette fibre de 11 nm se replie pour adopter une structure en hélice de 30 nm de diamètre. Enfin, elle est à son tour repliée en une fibre de 300 nm pour constituer des domaines en boucles de 15 000 à 100 000 pb. Au cours de la mitose et de la méiose, il existe transitoirement un niveau extrême de compaction, les chromosomes. Ils sont tellement condensés qu'on peut étudier leur morphologie au microscope optique.

#### 3-1- Nucléosome et la fibre F10

La chromatine est constituée de particules régulièrement espacées appelées nucléosomes. L'existence de cette unité fondamentale a été décelée dans les années 1970, suite à la digestion de l'ADN par des nucléases et l'obtention de fragments résiduels d'environ 200 pb et leurs multiples et ne sont pas obtenus lorsque la même expérience est effectuée sur de l'ADN purifié. Ceci a permis de proposer qu'il existe, au sein du noyau, une unité structurale de base. Cette interprétation a été largement confortée par des observations de microscopie électronique. Une digestion plus fine permet d'isoler un complexe de 200 kDa, composé d'ADN et de protéines. Peu de temps après, l'enchaînement des nucléosomes en dont l'aspect rappelle celui d'un chapelet a pu être visualisé grâce au progrès de la microscopie électronique en 1974. La même année, *Kornberg* proposera l'existence d'une unité de base au sein du noyau, de deux cents paires de base d'ADN complexées avec 4 paires d'histones, ce que *Chambon* appellera en 1975 le nucléosome (**figure 03**).

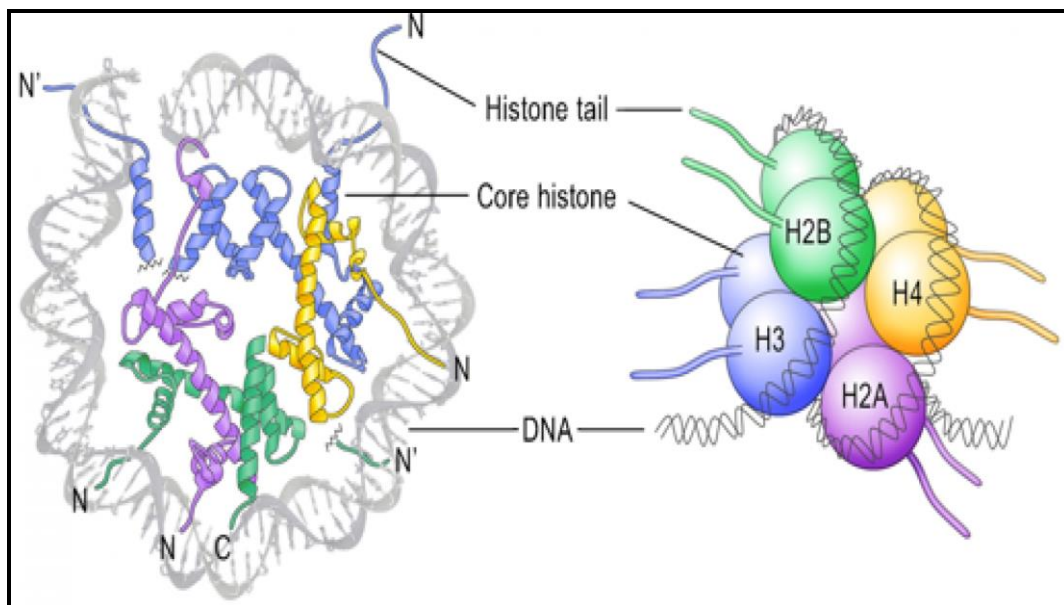


**Figure 03** : Cliché en microscopie électronique du nucléofilament (**Adkins, 2004**).

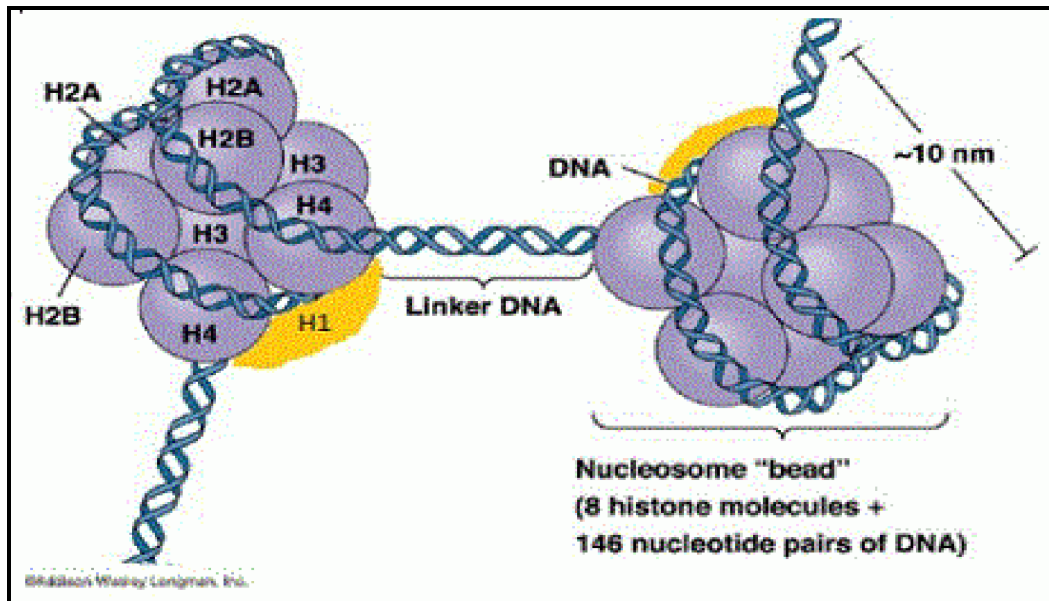
Un nucléosome est composé d'une particule cœur et une région de liaison appelée région inter-nucléosomale ou ADN linker qui relie les particules cœurs adjacentes. La particule cœur dont la structure est très conservée parmi les espèces, est composée de 146 pb d'ADN enroulés selon environ 2 tours (1,7 plus exactement) autour d'un octamère protéique formant deux disques arrangés en parallèles comprenant chacun des histones H3, H4, H2A et H2B ; la molécule d'histone H1 est placée à l'extérieur et joue le rôle de socle. La longueur de région inter-nucléosomale varie de 8 à 116 pb selon les espèces ; elle est de 60 pour l'humain.

Le nucléosome ainsi formé est maintenu par un jeu d'interactions électrostatiques et de liaisons hydrogènes entre les groupements phosphate de l'ADN et les histones, et par des contacts non polaires avec le sucre désoxyribose de l'ADN et les histones. Comme aucun contact direct n'existe entre les bases azotées et les histones, le nucléosome ne présente aucune spécificité à la séquence d'ADN.

Cette structure a un diamètre de l'ordre de 11 nm, c'est le premier niveau d'organisation de la chromatine à l'intérieur du noyau et qu'on appelle : Fibre de 10 nm ou F10. Cette organisation qui provoque une compaction de l'ADN selon un facteur 7 : une structure condensée ainsi va se retrouver à 7 fois moins sa longueur initiale (**figures 04 et 05**).



**Figure 04** : Structure de l'histone cœur du nucléosome (Ridgway, 2006).



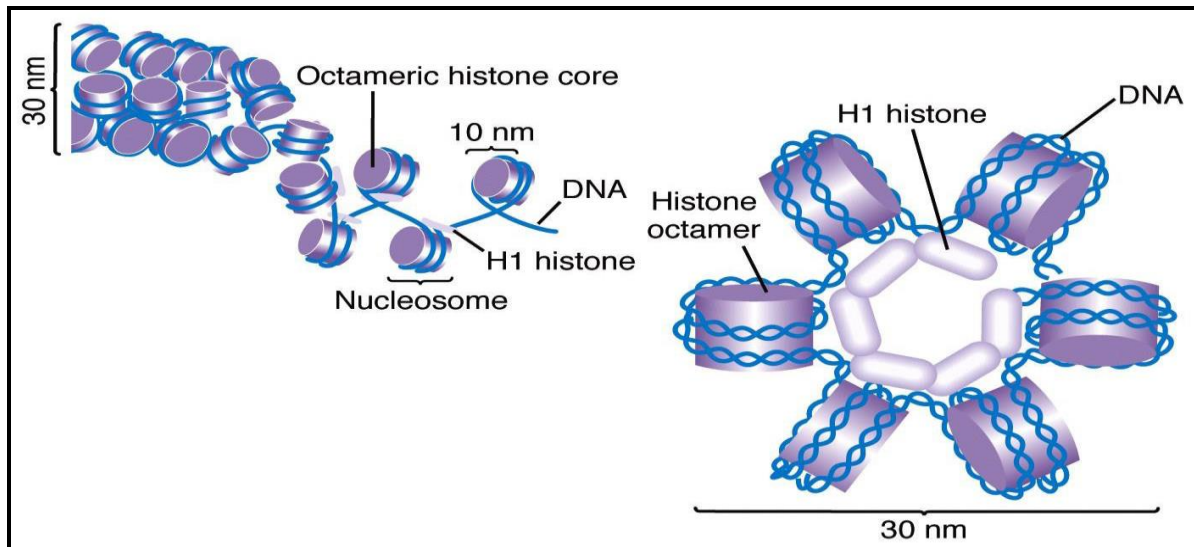
**Figure 05 :** Structure du nucléosome et de la fibre F10 (Young, 2014).

Il est important de signaler que l'histone de liaison H1 n'intervient pas dans la formation de la sous-unité fondamentale de la chromatine. Cependant, avec des protéines non histones, elles peuvent interagir avec les nucléosomes pour former des structures de plus haut degré d'enroulement.

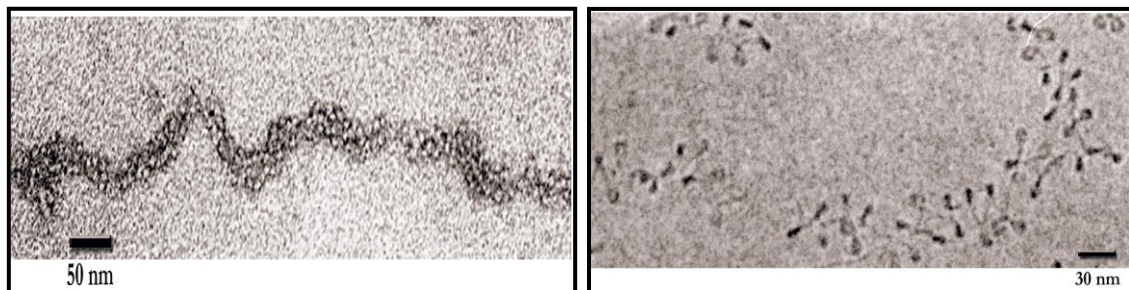
### 3-2- Fibre F30

Les niveaux supérieurs d'organisation des fibres de chromatine ont été visualisés dans différentes conditions salines. À une force saline basse, la chromatine se structure en collier de perles. À une force saline plus élevée (100 mM NaCl), elle forme un cylindre irrégulier d'environ 30 nm de large, communément appelé fibre de 30 nm ou F30.

La F10 présente une structure secondaire enroulée. Chaque tour comprend environ 6 nucléosomes ce qui correspond à un facteur de condensation d'environ 40 : chaque  $\mu\text{m}$  le long de l'axe de la fibre contient 40  $\mu\text{m}$  d'ADN. Dans cette fibre, les nucléosomes s'enroulent de façon hélicoïdale en une structure appelée solénoïde ; c'est le deuxième niveau de compaction de la chromatine. Pour cette fibre, deux modèles sont actuellement discutés : le premier dit en solénoïde, décrit précédemment, et qui le plus largement accepté. Le deuxième modèle dit en Zig-Zag (**figure 06 et 07**).



**Figure 06 :** Passage de la F10 à la F30 (Ridgway, 2006).



**Figure 07 :** Modèles de structure de la F30 (Adkins, 2004).

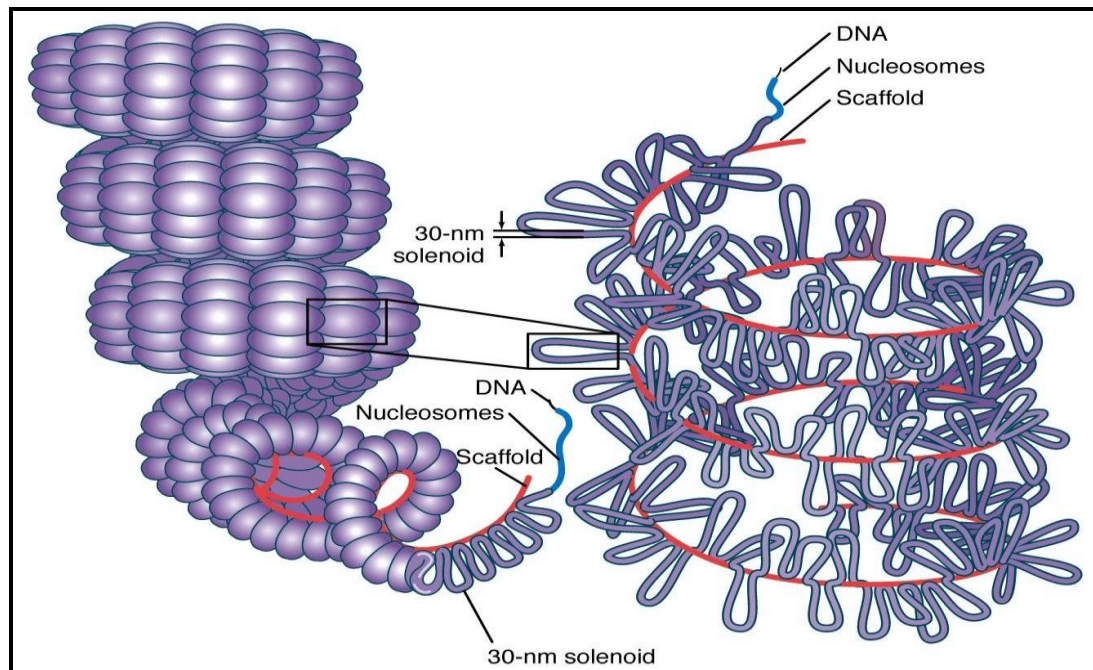
À gauche, un cliché en microscopie électronique de la fibre de 30 nm à salinité physiologique (solénoïde). À droite, un cliché en cryo-microscopie électronique à 5 mM d'ion monovalent.

Dans leur modèle dit de solénoïde, les nucléosomes consécutifs se localisent les uns à proximité des autres dans la fibre, formant une hélice simple. Depuis, cette fibre de 30 nm a été observée par microscopie électronique par de nombreux groupes, mais sa structure intrinsèque est encore débattue. Dans ce second modèle, les nucléosomes sont arrangés en Zig-Zag entre eux, de sorte à former une hélice double brin dont chaque nucléosome se retrouve à proximité du nucléosome.

### 3-3- Boucles chromatiniennes

La F30 peut également se replier et former des structures dont la topologie est encore moins bien décrite. Ces repliements d'ordre supérieur forment ce que l'on appelle l'organisation à grande échelle de la chromatine, qui repose sur la formation de boucles géantes de chromatine.

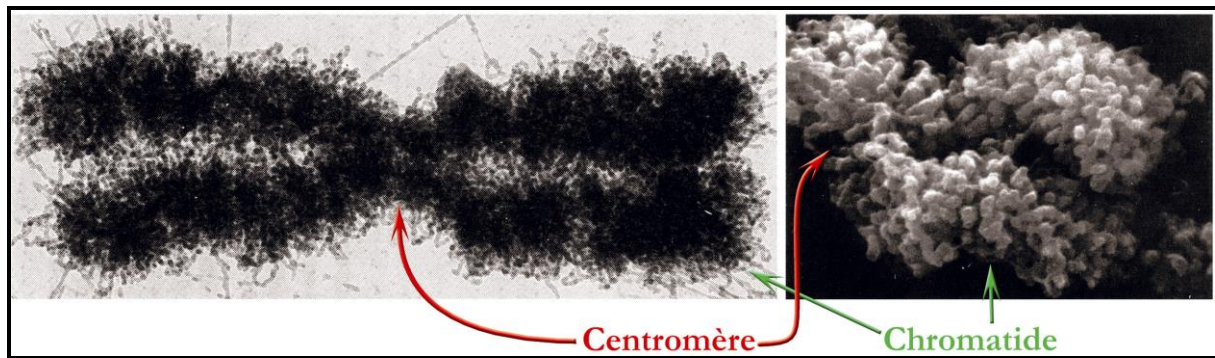
Le troisième niveau de compaction est assuré par le repliement de la fibre de 30 nm en boucles, ou domaines topologiques fermés, d'un diamètre de 250 à 300 nm environ, et contenant de l'ordre de 40 à 100 kb d'ADN, ces supers-tours apparaissent sous microscope électronique ancrés autour d'une partie centrale appelée armature ou matrice nucléaire, formée de protéines non histones. De cette armature protéique partent des boucles d'ADN, les fibres formants ces boucles sont les solénoïdes. La compaction atteint un facteur 1 000. Chaque boucle d'ADN commence et finit au niveau de l'armature, cette dernière est constituée en grande partie de protéines non-histones. Les boucles sont fixées à l'armature au niveau de régions spéciales le long de l'ADN appelées : région de liaison à l'armature ; SAR (Scaffold Attachment Région). Ces régions sont riches en AT et correspondent fréquemment à des fragments d'ADN courbés sujets à des ouvertures locales de la double hélice (comme les origines de réplication) (**figure 08**).



**Figure 08** : Représentation schématique de la structure en boucles (**Ridgway, 2006**).

### 3-4- Quatrième niveau de condensation

Enfin, un quatrième niveau de compaction correspond à l'enroulement des boucles en une hélice de l'ordre de 700 à 850 nm de diamètre qui constitue un chromatide. Le degré de compaction total de l'ADN atteint ainsi plus de 10 000 dans la chromatine condensée. Le chromosome métaphasique observé en mitose représente le niveau de compaction maximal et est capable d'être à la fois résistant, maniable et compact pour assurer une bonne ségrégation du matériel génétique (**figure 09**).



**Figure 09 :** Cliché microscopie électronique d'un chromosome métaphasique (Adkins, 2004).

#### 4- Différentes étapes de condensation de la chromatine

Le facteur de condensation général du matériel génétique suggère d'emblée que l'ADN ne peut être condensé directement dans la structure finale de la chromatine. Une hiérarchie au sein de l'organisation doit exister. L'assemblage de l'ADN en chromatine comprend plusieurs étapes successives initiées dès la réplication et qui commencent par la formation de la sous-unité fondamentale, le nucléosome, pour aboutir à des niveaux de compaction supérieurs organisés en domaines spécifiques dans le noyau.

La formation de la chromatine peut se découper en deux étapes principales : l'assemblage du nucléosome et la formation de la F10 et la F30, processus qui ne nécessitent pas l'hydrolyse de l'Adénosine Tri-Phosphate (ATP) (ATP-indépendant), puis l'organisation en structures de plus en plus compacte (supertours et boucles) par des processus qui mettent en jeu l'hydrolyse de l'ATP (ATP-dépendant), il s'agit du remodelage de la chromatine.

##### 4-1- Assemblage du nucléosome et formation de la F10

L'assemblage du nucléosome s'effectue grâce à des facteurs d'assemblage de la chromatine qui fonctionnent en coordination avec des protéines chaperonnes qui accompagnent les histones jusqu'à leur dépôt sur l'ADN. Selon le modèle généralement admis, l'assemblage du nucléosome se déroule en deux étapes : un tétramère d'histones H3 et H4 est d'abord déposé sur l'ADN nouvellement répliqué, puis c'est le tour des deux dimères de type H2A-H2B. Ces deux étapes sont sous le contrôle du même facteur d'assemblage de la chromatine CAF1 (Chromatin Assembly Factor 1). Le couplage avec la réplication se ferait par le recrutement de CAF1 au niveau de la fourche de réplication par le facteur PCNA (Proliferating Cell Nuclear Antigen) : complexe protéique joue le rôle d'une pince glissant le long de l'ADN et de facteur de progressivité de l'ADN polymérase.

La première étape est assistée par une chaperonne d'histone ASF1 (Anti-Silencing Function 1) quant au dépôt des histones H2A et H2B est assisté par le facteur NAP1 (Nucleosome Assembly Protein 1). Enfin, l'assemblage du nucléosome se termine par l'addition d'une histone H1 aux brins entrants et sortants de l'ADN entourant l'octamère. Le nucléosome est alors formé. La F30 est formé par l'enroulement spontanée de la F10 sous l'effet du fort caractère basique de la H1. Cette F30 peut subir des compactions supplémentaires pour atteindre le repliement des chromosomes mitotiques. La structure précise de ces formes compactes de la chromatine n'est pas connue. Néanmoins, des facteurs de remodelage ATP-dépendant sont impliqués.

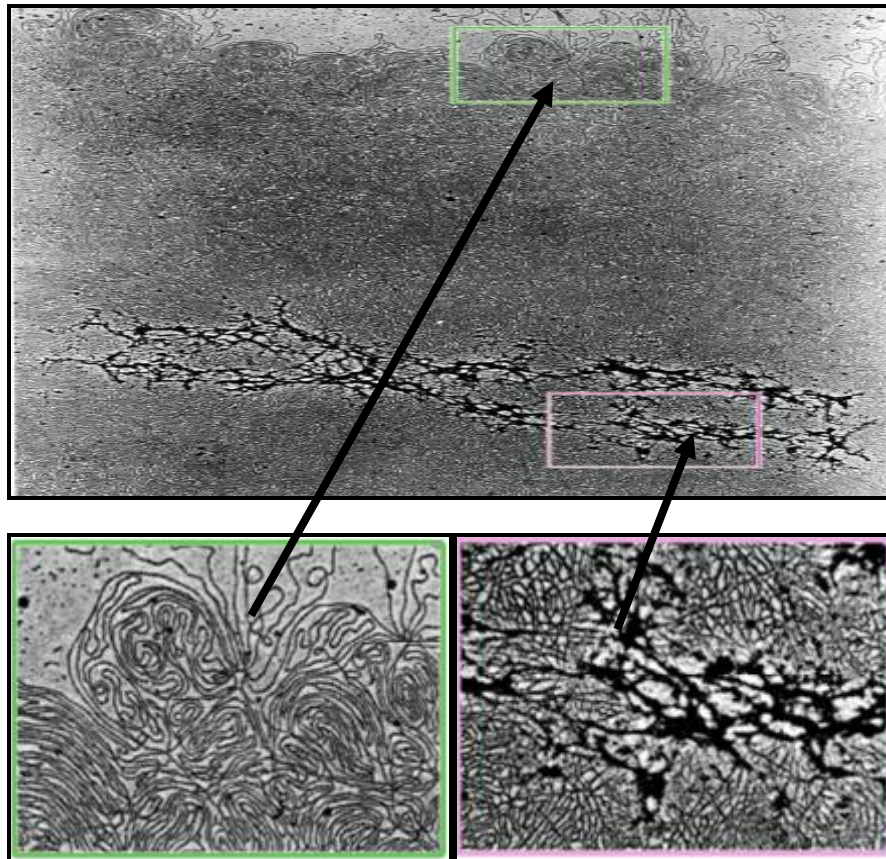
#### 4-2- Formation des niveaux d'organisation supérieurs de la chromatine

L'utilisation de la microscopie électronique a permis d'obtenir les premières informations sur l'architecture interne du chromosome métaphasique. Après extraction des histones par une solution de forte force ionique très riche en polyanions, on observe un halo de boucles d'ADN organisées en F30, super-enroulées en périphérie d'un résidu protéique appelé « matrice nucléaire » ou « squelette chromosomique » central ayant grossièrement la forme du chromosome métaphasique. Ces boucles représentent environ 100 kb d'ADN et correspondent aux rosettes de la fibre de 30 nm, insérées par leur base dans la structure protéique centrale au niveau des SAR. En faisant varier les conditions ioniques du milieu, on observe des variations morphologiques du squelette protéique, qui s'étire ou se contracte de façon réversible (**figure 10**).

Deux principales protéines constituent le squelette interne du chromosome : la topoisomérase II et les condensines I et II.

La Topoisomérase est une protéine qui présente une dualité fonctionnelle : de structure, elle fait partie des composants du squelette chromosomique, enzymatique car elle permet de supprimer les supertours positifs créés par l'ouverture localisée de la double hélice. Cette fonctionnalité est essentielle pour permettre une séparation correcte des deux chromatides sœurs pendant l'étape de condensation des chromatides en prophase. L'invalidation du gène de la Topoisomérase II se traduit alors par un défaut de condensation de la chromatine, qui survient de façon imparfaite et incomplète. Cette suppression a pour conséquence une séparation incomplète des deux lots chromosomiques en anaphase et la persistance de masse de chromatine au niveau du plan de division cellulaire. Cette protéine n'est cependant pas un composant fixe de l'axe protéique, elle en constitue au contraire un élément dynamique en renouvellement permanent.

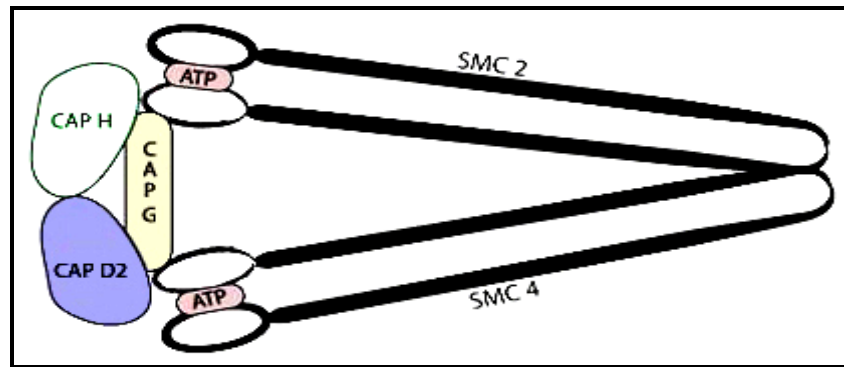




**Figure 10 :** Modèle d'organisation du chromosome (Adkins, 2004).

Cliché en microscopie électronique d'un chromosome mitotique dont les histones ont été enlevées. Les zones, encadrées en bas de la figure, correspondent à des agrandissements des zones de la même figure. Des boucles d'ADN d'environ 50 à 100 kb (à gauche) sont reliées à une structure d'apparence squelettique (à droite) formée de protéines non-histones.

Le deuxième constituant essentiel de l'axe chromosomique est la condensine, dont il existe deux sous types : I et II. Ces protéines sont en fait des complexes multiprotéiques associant 2 sous unités SMC (SMC2 et SMC 4 pour Structural Maintenance of Chromosome) et 3 sous unités non SMC (CAP H/H2, CAP D2/D3, CAP G/G2). Les deux sous unités SMC s'associent pour former un V porteur d'un site de fixation pour l'ATP à l'extrémité de chacune des deux branches. Les sous-unités non SMC permettent de contrôler l'ouverture de la pince constituée par les deux protéines SMC. Grâce à l'énergie libérée par l'hydrolyse de l'ATP, les condensines ont la capacité de se lier à l'ADN et de générer des supertours et des boucles, ce qui entraîne une condensation de la chromatine (**figure 11**).



**Figure 11** : Représentation schématique de structure des condensines (Becker, 2002).

Le rôle exact des deux sous-types condensine I et II n'est pas clairement identifié, mais il semble que la condensine II intervienne en premier dans les étapes précoces de la condensation, alors que la condensine I interviendrait plus tard pour stabiliser la chromatine dans cet état compacté. De manière surprenante au regard du rôle présumé des condensines, leur inactivation n'entraîne pas une abolition de la compaction, qui survient malgré tout mais de façon retardée et incomplète. Cette compaction imparfaite entraîne cependant une morphologie anormale des chromosomes en raison d'une désorganisation de l'axe protéique des chromatides, ainsi qu'une ségrégation anormale en raison de masses chromatiniennes persistantes au niveau de la plaque équatoriale en fin d'anaphase.

Un dernier type de protéine ne faisant pas partie du squelette protéique, joue néanmoins un rôle fondamental dans la structure et la physiologie du chromosome métaphasique : il s'agit des cohésines. Ces protéines appartiennent à la même famille que les condensines et ont une structure semblable, constituées de 4 sous-unités spécifiques (2 sous unités SMC1 et 3, et deux sous unités non SMC), RAD21 (REC8 pour les Cohésines méiotiques) et SA1. Les cohésines sont associées à la chromatine pendant la phase S, au moment de la réplication de l'ADN et servent à maintenir ensemble les deux chromatides sœurs jusqu'à leur séparation.

La topoisomérase II et condensine II sont présentes pendant l'interphase dans le noyau. En revanche, la condensine I a une localisation cytoplasmique et ne peut s'associer aux chromosomes qu'après la disparition de l'enveloppe nucléaire. En début de prophase, on observe une phosphorylation de la topoisomérase II et de la condensine II entraînant leur association avec la périphérie de la chromatine en cours de compaction. Cette association permet un premier niveau de compaction de la fibre de 30 nm pour aboutir à une fibre de 200 à 250 nm de diamètre. Puis, un deuxième niveau de compaction apparaît en fin de prophase suite à la réorganisation des protéines de structures qui deviennent axiales et assurent la stabilisation de l'enroulement de la fibre de 250 nm en une fibre de 700 nm.

La condensine I semble jouer un rôle prépondérant dans la stabilisation finale de la structure du chromosome. Un point essentiel de cette architecture chromosomique est son caractère dynamique, caractérisée par un flux permanent de protéines qui s'associent et se dissocient.

## 5- Territoires chromatiniens

Exception faite de la mitose où elle existe sous forme de chromosomes individuels compactés, la chromatine est diffuse dans le noyau. Elle apparaît sous forme de régions plus ou moins condensées et cet arrangement en compartiments distincts influence les activités fonctionnelles du noyau. Ainsi, la chromatine n'est pas organisée de manière chaotique au sein du noyau, bien au contraire, cette structure est hautement régulée.

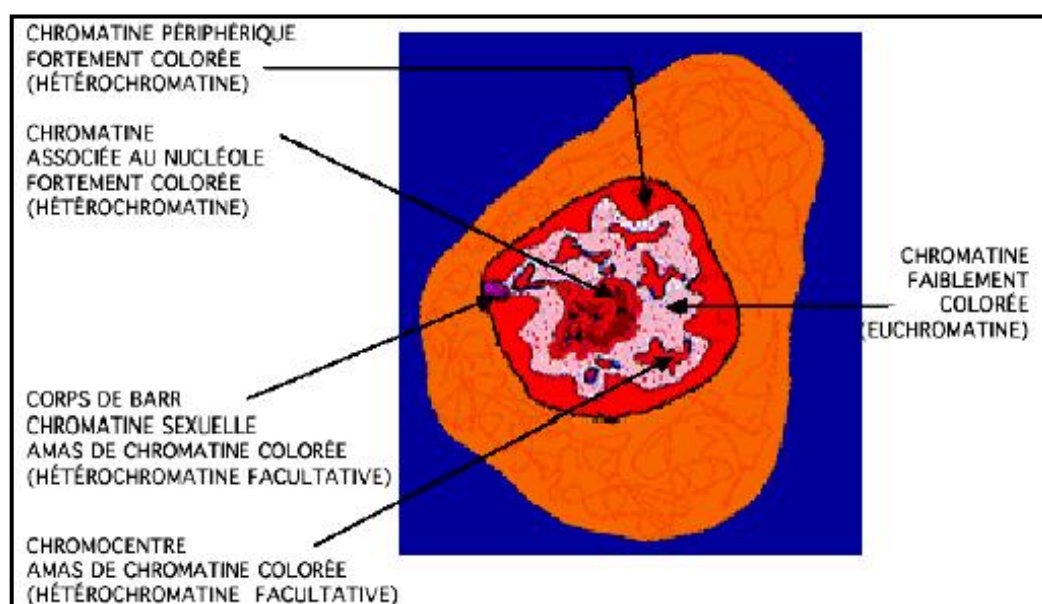
En 1928, basé sur des observations exclusivement histologiques, *Heitz* remarqua « l'existence de deux régions contiguës dans le noyau interphasique, mais non délimitées, présentant des affinités tinctoriales distinctes ». En effet, cette différence de coloration est due au fait que certaines zones du noyau cellulaire qui sont plus condensées que d'autres. Il a introduit le terme d'euchromatine (dite active), représentant la chromatine lâche, et d'hétérochromatine (dite inerte), représentant les zones plus denses et compactes : ces deux types de chromatine dans les cellules eucaryotes présentant des critères structuraux et fonctionnels différents. Par la suite, l'euchromatine a été décrite comme associée à une transcription active des gènes, alors que la transcription est majoritairement réprimée au sein de l'hétérochromatine. Il a été communément admis depuis un certain nombre d'années que l'euchromatine, contiendrait des gènes particulièrement actifs, contrairement à l'hétérochromatine qui serait de la chromatine transcriptionnellement inactive. En fait, cette idée est une simplification de la réalité.

Le degré de condensation de la chromatine varie au cours du cycle cellulaire. Pendant l'interphase, l'euchromatine est assez bien décondensée et dispersée dans tout le volume du noyau. Durant cette phase du cycle cellulaire, les gènes sont transcrits et l'ADN se réplique. Une partie de la chromatine interphasique conserve un état très condensé, c'est l'hétérochromatine. Elle échappe à la transcription et contient des séquences d'ADN non codantes, hautement répétitives comme celles que l'on trouve au niveau des centromères et des télomères.

Dans les cellules somatiques, on peut distinguer également, au cours de l'interphase, les zones de réplication précoce correspondant à l'euchromatine, et celles de la réplication tardive qui concernent l'hétérochromatine : c'est ce qu'on appelle le profil répliatif.

On distingue deux sous-catégories d'hétérochromatine, en fonction de leur stabilité : l'hétérochromatine constitutive et l'hétérochromatine facultative.

- Toutes les cellules d'une même espèce vont regrouper des régions similaires de l'ADN sous la forme d'une hétérochromatine constitutive. Par exemple, pour l'Homme, les chromosomes 1, 9, 16 et Y contiennent de larges régions d'hétérochromatine constitutive, majoritairement réprimées. Cette hétérochromatine constitutive est généralement constituée de séquences répétées, et semble jouer un rôle préférentiel de structuration de régions telles que les centromères et télomères. Elle peut également avoir un impact direct sur l'expression des gènes à proximité, en les réprimant plus ou moins en fonction de la propagation de l'hétérochromatine.
- L'hétérochromatine facultative quant à elle n'est pas répétée, et peut structurer des régions d'ADN différentes au sein de cellules d'une même espèce. Bien qu'elle partage la même structure compactée que l'hétérochromatine constitutive, elle peut se décondenser sous certaines conditions et permettre ainsi la transcription de gènes normalement réprimés. La formation de l'hétérochromatine est un moyen de réprimer spécifiquement certaines régions du génome, et participe ainsi à l'élaboration d'un programme génétique propre à chaque cellule. Cette organisation de la chromatine est donc extrêmement régulée et fait appel à de nombreux acteurs épigénétiques. Le meilleur exemple est le chromosome X inactif chez la femelle des mammifères, on peut l'observer sous forme d'un point dense dans le noyau interphasique appelé : corps de Barr (**figure 12**).



**Figure 12** : Représentation schématique du noyau coloré au Feulgen (Baxter, 2002).

# Chapitre III

## Structures chromosomiques spécialisées

Le noyau des cellules est une structure très organisée, dont l'agencement joue un rôle important dans la régulation de l'expression des gènes. La compréhension des mécanismes à l'origine de cette organisation est donc essentielle à la compréhension du fonctionnement des génomes. Malgré les progrès réalisés par les techniques d'imagerie et d'analyse biochimique, la structure du chromosome métaphasique reste dans une large mesure inconnue. Une partie de la difficulté qu'il y a à comprendre son architecture est liée au fait qu'il s'agit d'une structure dynamique, transitoire, qui nécessite donc d'être bloquée à un moment précis pour être analysée. Mais dans le même temps où l'on fige un chromosome par une technique d'analyse, on prend le risque d'en modifier la structure. De ce point de vue, le chromosome au sens microscopique n'est donc pas un élément stable de la cellule mais une structure dynamique et transitoire.

## 1- Télomères des eucaryotes supérieurs

### 1-1- Historique et définition

Le terme de télomère, du grec *telos* (fin) et *meros* (partie), a été introduit par *Muller* en 1938. Il désigne actuellement des structures nucléoprotéiques spécialisées localisées aux extrémités des chromosomes linéaires eucaryotes, nécessaires à leur stabilité et à leur réplication. La notion de télomère est apparue à la fin des années 1930s à partir de travaux menés sur la drosophile par *Muller* et le maïs par *McClintock*, 1941 qui ont remarqué que les extrémités des chromosomes ayant subi des cassures, induites par les UV, fusionnaient entre elles, aboutissant ainsi à la fusion de deux chromosomes ou la re-circularisation du chromosome, alors que les extrémités chromosomiques naturelles ne fusionnaient pas.

Ensuite, il a été observé par *Watson* en 1972 que, du fait du mode semi-conservatif de la réplication de l'ADN, une molécule d'ADN linéaire est incomplètement répliquée et se raccourcit à chaque cycle de réplication. Ce problème de réplication des extrémités a été mis en avant pour expliquer le nombre limité de divisions des cellules en culture par *Hayflick* en 1965. Les liens entre raccourcissement télomérique et sénescence (vieillesse cellulaire) ont été étudiés (et prouvés) plus tard.

La séquence des télomères chez l'homme a été identifiée en 1988. Les différents composants fonctionnels du télomère ont ensuite été caractérisés comme la télomérase et les protéines associées aux télomères TRF1 (1995) et TRF2 (1997). Les publications sur le sujet n'ont cessé d'augmenter et ces recherches ont été couronnées en 2009 par l'attribution du Prix Nobel de Médecine et Physiologie à *Blackburn*, *Greider* et *Szostak*.



À l'extrémité proche du centromère existe des séquences dites subtélomères et qui sont définies comme les régions de transition (dont la fonction est inconnue à ce jour) entre les séquences chromosomiques spécifiques et les répétitions télomériques terminales (TTAGGG)<sub>n</sub>. Donner une définition plus précise est difficile car ces régions sont extrêmement dynamiques et variables. Elles sont composées de séquences s'étendant sur une distance de 8 kb à 300 kb environ et sont particulièrement riches en duplications (segments d'ADN génomique répétés d'au moins 1 kb de long et de plus de 90 % de similarité de séquence).

### **b- Complexe télosome (Shelterin)**

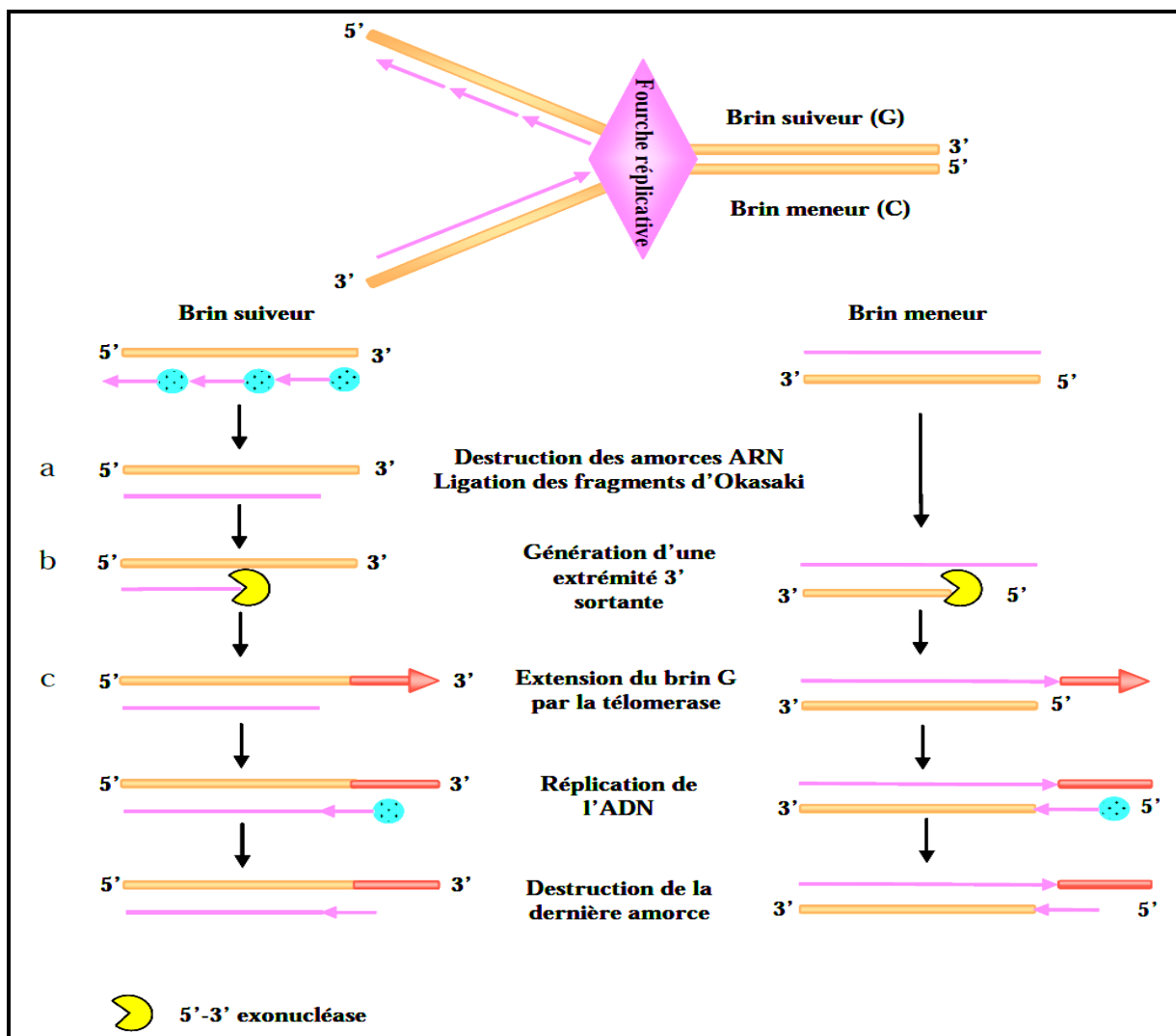
Le complexe protéique spécialisé associé au télomère chez les mammifères s'appelle télosome. Il est formé de six protéines: TRF1 (Telomeric Repeat Binding Factor 1), TRF2 (Telomeric Repeat Binding Factor 2), POT1 (Protection Of Telomere 1), TIN2 (TRF1 Interacting factor 2), TPP1 (TriPeptidyl Peptidase 1) et RAP1 (Repressor Activator Protein 1). Ces protéines peuvent se lier à l'ADN télomérique double brin ou à l'extrémité 3' sortante simple brin. La région double brin est reconnue spécifiquement par TRF1 et TRF2, deux protéines de la même famille qui se distinguent par leur domaine amino-terminal (acide pour TRF1, basique pour TRF2). L'ADN simple brin est reconnu par TPP1 et POT1, TPP1. Les éléments du télosome peuvent former des complexes en absence d'ADN télomérique et des sous-complexes ont également été identifiés dans des extraits cellulaires (**figure 13**). Le télosome intervient dans la régulation de la longueur des télomères par la télomérase et dans la protection des extrémités télomériques contre les systèmes de signalisation et de réparation des dommages à l'ADN.

### **1-3- Réplication des extrémités télomériques et télomérase**

Du fait de l'incapacité de l'ADN polymérase classique à répliquer complètement l'ADN linéaire, l'extrémité des chromosomes eucaryotes est confrontée à une perte de matériel génétique à chaque cycle de réplication. En effet, lors de la réplication semi-conservative de l'ADN, le brin retardé (brin suiveur) est synthétisé de façon discontinue par l'intermédiaire de fragments d'Okazaki. Chaque fragment d'Okazaki est initié par une amorce ARN qui est ensuite détruite, remplacée par de l'ADN et liguée. En revanche, l'amorce la plus distale ne peut pas être remplacée sur un ADN linéaire, ce qui a pour conséquence une délétion de quelques bases de matériel génétique terminal à chaque cycle répliatif (**figure 14**).



Toutefois chez l'homme, la perte télomérique à chaque cycle réplcatif est d'environ 100 à 200 pb, c'est-à-dire beaucoup plus importante que la taille d'une amorce ARN. Par ailleurs, la réplcation du brin avancé (meneur) est complète mais génère un double brin à bout franc. Or, il a été observé la présence d'un brin 3' sortant à toutes les extrémités télomériques. Ces deux observations nous ont fait émettre l'hypothèse d'un mécanisme post-réplcatif, faisant intervenir une activité nucléase 5' qui génère une nouvelle extrémité 3' sortante par résection du brin C. Ce mécanisme est régulé par les protéines du télosome, qui contrôlent la détermination stricte du nucléotide terminal, et expliquerait ainsi l'attrition (érosion) plus rapide des télomères (**figure 14**).



- (a) La réplcation incomplète du brin suiveur est associée à la perte cyclique de nucléotides,
- (b) L'extrémité 3' sortante est générée par l'action d'une exonucléase 5'-3',
- (c) Dans certaines cellules, cette perte est comblée par l'action de la télomérase.

Face au problème de réplication des extrémités, la plupart des eucaryotes ont développé un mécanisme, conservé au cours de l'évolution, pour lutter contre l'attrition télomérique : l'enzyme télomérase. La télomérase est un complexe ribo-nucléoprotéique composé d'une sous-unité catalytique transcriptase inverse « télomérase » (hTERT), d'un ARN (hTR) associé à des protéines accessoires qui facilitent l'assemblage et la stabilité du complexe catalytique. La télomérase utilise la molécule d'ARN hTR comme matrice (AAUCCCAAUC), la complémentarité entre l'ARN hTR et l'ADN télomérique permettant un alignement correct. La télomérase peut ainsi catalyser plusieurs cycles d'alignement extension permettant d'ajouter des centaines de nucléotides à la même extrémité par transcription inverse.

La régulation de la télomérase s'effectue au niveau transcriptionnel. Son expression est faible ou absente dans la plupart des cellules somatiques mais présente dans les cellules souches et germinales. La plupart des tumeurs expriment le gène de télomérase ; c'est ce qu'on désigne par la dérégulation du gène de la télomérase.

#### **1-4- Fonctions des télomères**

##### **a- Échappement aux systèmes de réparation de l'ADN**

D'un point de vue mécanique, les extrémités chromosomiques correspondent à des cassures d'ADN double brin. Les dommages de l'ADN induisent habituellement une série de réponses cellulaires qui activent les points de contrôle d'arrêt du cycle cellulaire pour permettre la réparation de l'ADN. En cas de dysfonction télomérique, l'activation de ces mécanismes de réparations à l'extrémité des chromosomes peut avoir des conséquences délétères : fusions télomériques, réarrangements chromosomiques et instabilité génomique globale. Les télomères possèdent donc des propriétés de protection spécifiques qui permettent d'échapper non seulement à la réponse aux dommages à l'ADN mais aussi aux mécanismes de réparation.

##### **b- Dysfonctionnement des télomères**

Des dysfonctions des télomères et de la télomérase ont été décrites dans de nombreuses pathologies tant constitutionnelles qu'acquises.

**D'une part**, l'activation anormale de la télomérase peut être à l'origine de pathologies tumorales (cancéreuses ou pathologies acquises). Du fait de l'absence de télomérase dans les cellules somatiques, les télomères raccourcissent à chaque cycle cellulaire, jusqu'à atteindre une taille critique. Les cellules entrent alors en sénescence, une forme d'arrêt permanent du cycle cellulaire. Ce mécanisme se comporte comme une barrière anti-cancer en limitant la prolifération des cellules somatiques ayant potentiellement acquis des anomalies génomiques. Ce point d'entrée en sénescence correspond à la « limite de *Hayflick* », qui définit le nombre maximal de divisions d'une cellule somatique primaire en culture.

Dans 85 à 90 % des cas, la télomérase est réactivée (dépression du gène de la télomérase). Mais dans 10 % des cas, un mécanisme alternatif appelé ALT (Alternative Lengthening of Telomeres).

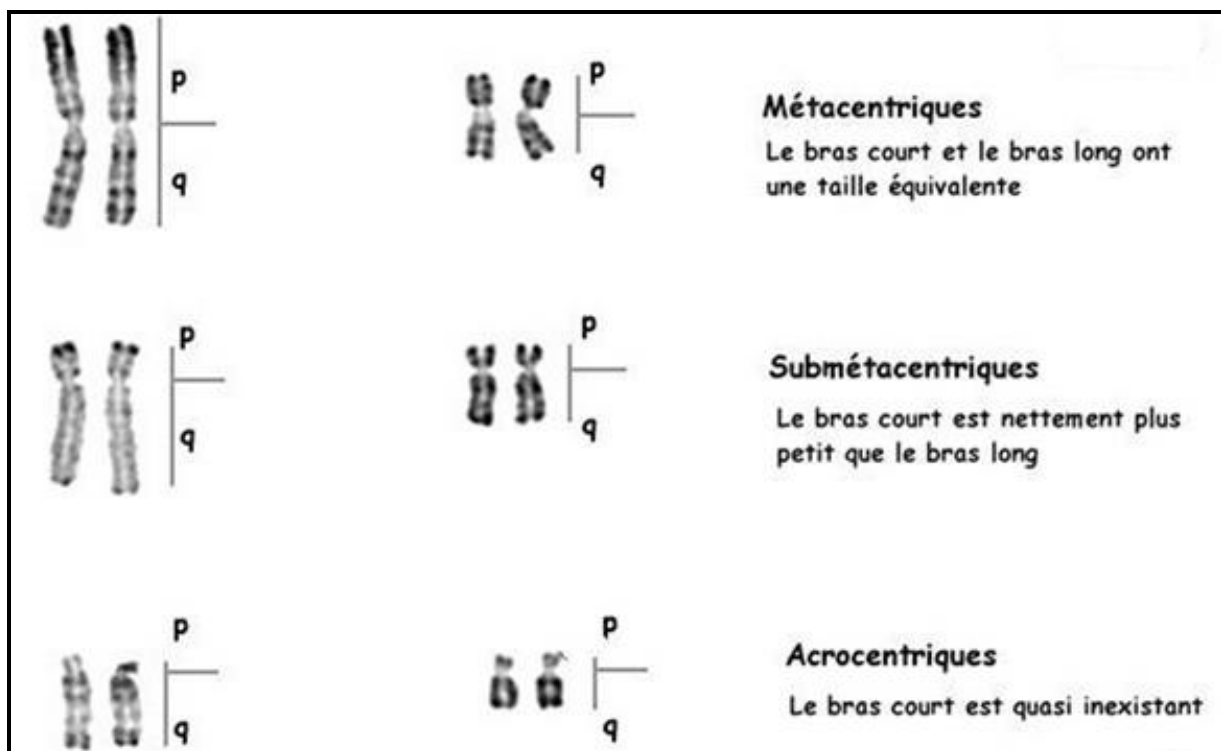
**D'autre part**, des mutations constitutives entraînant une perte de fonction de la télomérase sont à l'origine de pathologies constitutionnelles, regroupées sous le nom de syndromes de raccourcissement télomérique. Les conséquences se manifestent tout d'abord dans les tissus au renouvellement rapide, où les capacités répliquatives des cellules souches sont affectées. Les principaux signes caractéristiques sont un vieillissement prématuré (cheveux gris, dysplasie unguéale), une fibrose pulmonaire et/ou hépatique, des dysfonctions immunitaires et hématologiques, une augmentation du risque de cancer, une sensibilité accrue à la chimiothérapie et à la radiothérapie. La pathologie la plus anciennement connue est la dyskératose congénitale secondaire à des mutations des sous-unités de la télomérase ou des protéines associées. Plus récemment, le syndrome de *Hoyeraal-Hreidarsson* et des pathologies débutant chez l'adulte comme la fibrose pulmonaire idiopathique ou l'anémie aplasique ont été rattachées à des anomalies de la télomérase en 2009.

## 2- Centromères

### 2-1- Historique et définition

Trois éléments sont indispensables pour que les chromosomes puissent se répliquer et ségréger normalement : les origines de réplication, les télomères et les centromères. Les origines de réplication existent en grand nombre chez les mammifères et sont partiellement caractérisées et existent en grande partie au niveau des SAR. Il en va de même pour les télomères. Les centromères sont, quant à eux, encore loin d'avoir livré leurs secrets malgré le nombre considérable d'études qui leur ont été consacrées ces dernières années. Ils jouent pourtant un rôle essentiel dans la ségrégation correcte du matériel génétique au cours de générations et leur dysfonctionnement peut conduire à des anomalies chromosomiques importantes, dont la trisomie 21 est l'exemple le plus connu.

La position des centromères sur les chromosomes est exploitée en cytogénétique pour leurs classifications. Dans un caryotype, on peut distinguer les différents chromosomes selon deux critères : la taille et l'indice centromérique permettant ainsi de faire la distinction entre chromosomes métacentrique, sub-métacentriques et acrocentriques (**figure 15**).



**Figure 15 :** Classification des chromosomes en fonction de l'indice centromérique (Shaffer *et al.*, 2013).

Le centromère est un domaine chromosomique spécialisé sur lequel s'assemble le kinétochore, complexe protéique couplant le fuseau mitotique au chromosome au cours de la mitose. Le centromère joue donc un rôle crucial dans la ségrégation des chromatides sœurs. Sur le chromosome mitotique, le centromère correspond à la constriction primaire du chromosome. Les régions flanquant du centromère, nommées régions péri-centromériques, sont impliquées dans la cohésion des chromatides sœurs. On désigne par Région Centromérique (RC) le domaine constitué du centromère et des régions péri-centromériques.

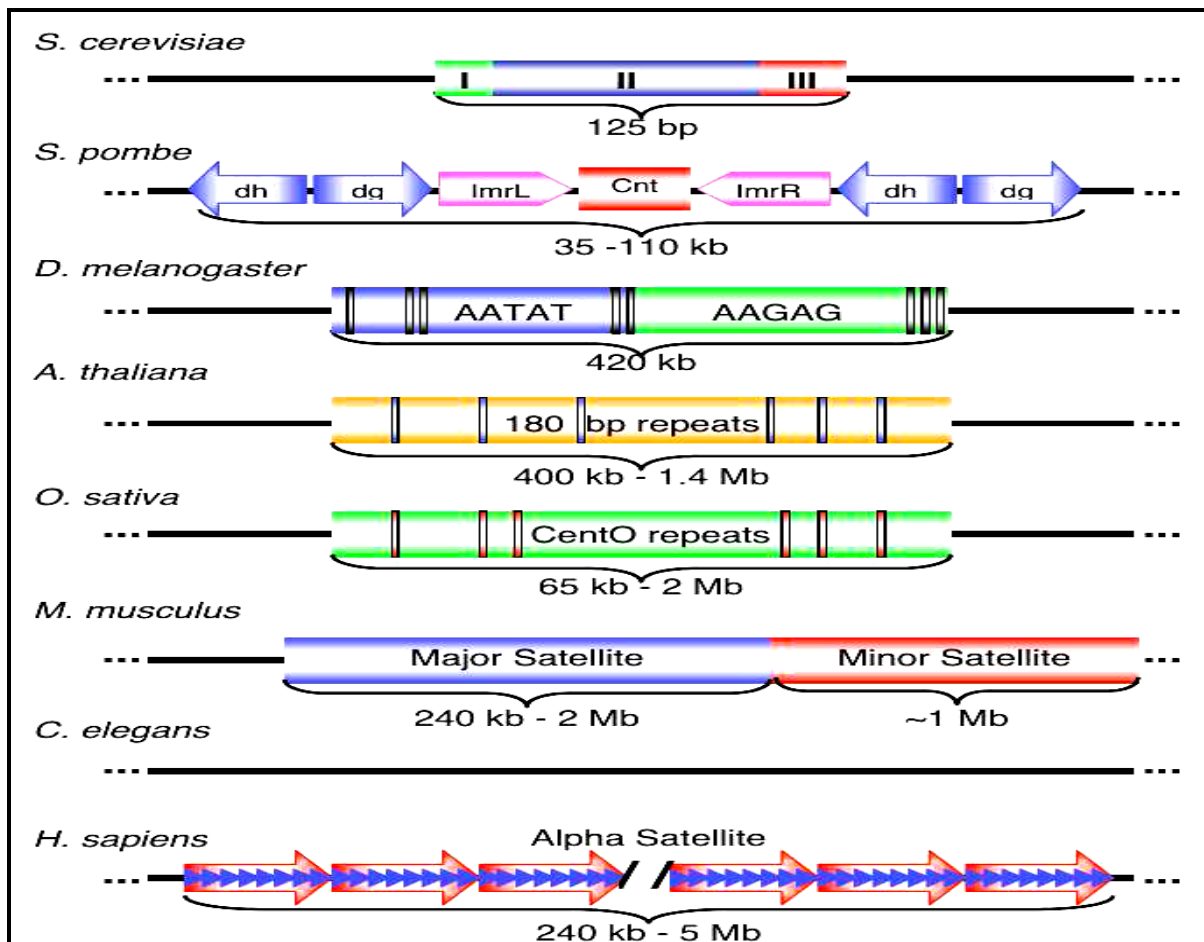
## 2-2- Structure des régions centromériques

### a- ADN centromérique

Les séquences des RCs sont très divergentes entre les espèces. Le centromère le plus simple est trouvé chez la levure *Saccharomyces cerevisiae*, constitué d'un domaine de 125 pb, suffisant pour faciliter l'assemblage du kinétochore. Contrairement à ce centromère dit « ponctuel », la plupart des eucaryotes possèdent des centromères dit « régionaux ». Ces régions sont composées d'ADN répété en tandem ; la séquence et la taille de l'unité de répétition ainsi que le nombre de répétitions sont variables entre les espèces.

On observe des tailles de centromères allant de 35 à 110 Kb chez *Saccharomyces pombe* jusqu'à 0,3 à 5 Mb chez l'homme. On trouve plusieurs familles de séquences d'ADN satellites (séquences d'ADN répétées groupées en tandem) au niveau des RCs humaines. La situation se complique déjà chez *S. pombe* puisque ses trois chromosomes portent des séquences d'ADN centromériques plus complexes et différentes d'un chromosome à l'autre, bien que formées d'éléments d'ADN répétés, de même type mais arrangés de façon variable. Leur simplicité, d'une part, et une approche génétique de la fonction centromérique, d'autre part, ont permis la caractérisation des séquences d'ADN centromérique chez *S. cerevisiae*. Elles sont constituées par deux courtes régions d'ADN conservées (8 pb pour CDE I et 25 pb pour CDE III) séparées par CDE II (78 à 86 pb > 90 % A+ T). L'analyse de nombreuses délétions et insertions dans CDE II a montré que seule la richesse en A+T de cet élément était nécessaire au bon fonctionnement du centromère de levure, sans que la séquence elle-même ait besoin d'être particulièrement stricte. Une délétion complète de CDE I et/ou III n'a qu'un effet réduit sur la fidélité de la ségrégation chromosomique. Par digestion à la nucléase, on a pu montrer que l'ADN centromérique de *S. cerevisiae* était engagé dans un complexe nucléoprotéique de 200 à 250 pb.

À la différence de la levure, les séquences d'ADN qui assurent la fonction centromérique chez les autres espèces eucaryotes ne sont pas caractérisées. Cependant, chez tous ces organismes, les régions centromériques contiennent des quantités importantes de séquences répétées en tandem : les ADN satellites. Leur propriété principale est ce caractère répété en tandem. D'une espèce à l'autre, ils sont extrêmement variables en quantité, nombre, séquence et taille d'unité de répétition (**figure 16**).

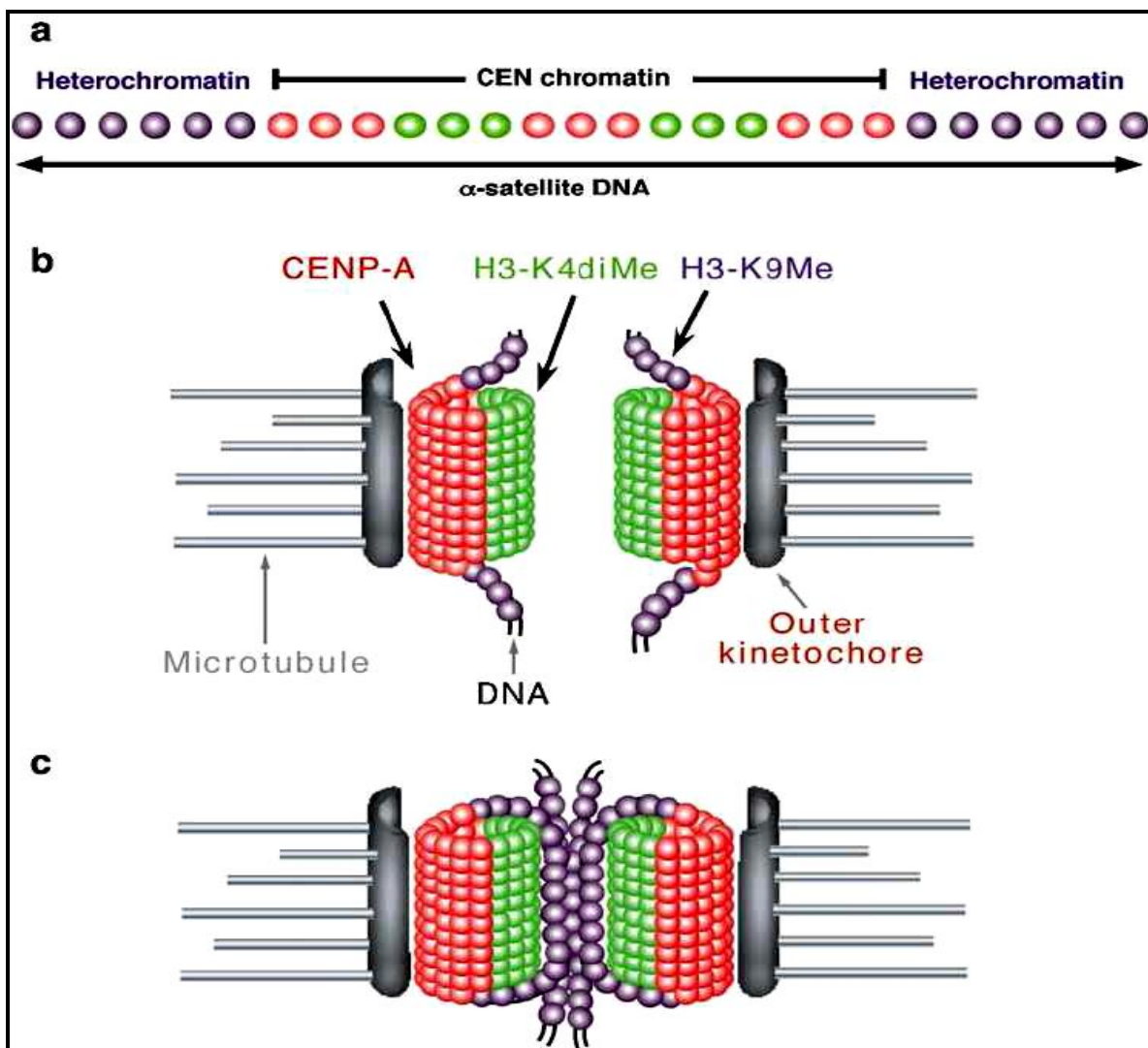


**Figure 16** : Organisation des RCs entre les espèces (Rudd, 2005).

Chez les mammifères, l'ADN satellite le plus abondant et le plus étudié est l'ADN dit « alpha satellite » qui représente une portion importante du génome humain. Il est constitué de répétitions de 171 pb, organisées en tandem, et il est présent sur tous les chromosomes dans les régions centromériques. Chaque chromosome est équipé d'une sous-famille spécifique du même ADN alpha satellite qui se distingue des autres. Cette homologie réduite de séquence peut être exploitée en cytogénétique par hybridation *in situ* dans diverses situations.

### b- Chromatine centromérique

Les RCs constituent, avec les télomères, l'hétérochromatine constitutive caractérisée par des marques d'histones répressives telles que H3K9me3, H3K27me3 et H4K20me3, et par un fort taux de méthylation de l'ADN ainsi que par leur réplication tardive. Les centromères de toutes les espèces sont caractérisés par la présence du variant de l'histone H3 CenP-A (parfois nommé CenH3), très conservé et qui se trouve exclusivement au niveau de ces centromères. Cette dernière caractéristique est nécessaire et suffisante à la formation du centromère ; c'est pourquoi elle est considérée comme définissant le centromère, à défaut d'un consensus de séquences d'ADN (**figure 17**).



**Figure 17** : Modèle de structure secondaire de la chromatine centromérique (Rudd, 2005).

### 2-3- Paradoxe du centromère

Le rôle joué par le centromère dans l'assemblage du kinétochore et la division cellulaire est très conservé au cours de l'évolution, alors que la séquence d'ADN est très divergente entre les espèces. Cela semble paradoxal car en général, la conservation de la fonction d'un locus au cours de l'évolution est corrélée avec la conservation de sa séquence. Cette singularité est connue sous le nom de paradoxe du centromère. En revanche deux caractéristiques sont conservées au cours de l'évolution : d'une part, on trouve un variant spécifique de l'histone H3 (CenpA chez l'homme) au niveau de tous les centromères. D'autre part, le caractère répété et riche en A/T des séquences est présent chez quasiment tous les eucaryotes.

La conservation du caractère répété des séquences au cours de l'évolution pose la question de son rôle dans le fonctionnement du centromère. Des études portant sur des chromosomes humains anormaux, caractérisés par une relocalisation du centromère, montrent que de nouveaux centromères peuvent se former sur des régions dépourvues de séquences répétées, mais présentant néanmoins une séquence riche en A/T ; ces structures sont appelées néo-centromères. Ces observations montrent que le centromère peut être fonctionnel sans la présence d'ADN répété ; cette caractéristique n'est donc, apparemment, pas nécessaire.



# Références bibliographiques

- **ADKINS NL, WATTS M, GEORGE PT et al.** 2004. To the 30-nm chromatin fiber and beyond. *Biochimica Biophysica Acta*. 1677:12-23.
- **ALBERTS B, JOHNSON A, LEWIS J et al.** 2002. Molecular Biology of the Cell. 4<sup>ème</sup> édition. *Garland Science, New York*. ISBN:0-8153-3218-1.
- **BANNISTER et KOUZARIDES.** 2011. Regulation of chromatin by histone modifications. *Cell Research*. 21:381-395.
- **BAXTER J, MERKENSCHLAGER M et FISHER AG.** 2002. Nuclear organization and gene expression. *Current Opinion in Cell Biology*. 14:372-376.
- **BECKER PB et HORZ, W.** 2002. ATP-dependent nucleosome remodeling. *Annual Review of Biochemistry*. 71:247-273.
- **BEDNAR J, HOROWITZ RA, GRIGORYEV SA et al.** 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences-USA*. 95:14173-14178.
- **CALLEN J C et PERASSO P.** 2005. Biologie cellulaire : des molécules aux organismes. 2<sup>ème</sup> édition. *Dunod, Paris*. ISBN:2-1004-9236-5.
- **CHURCH D.** 2013. Le génie dans vos gènes: médecine épigénétique et nouvelle biologie de l'intention. *Éditions Dangles*. Pagination multiple. ISBN:2-7033-0972-4.
- **CREMER T et CREMER C.** 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*. 2:292-301.
- **DANIEL L. HARTI, ELISABETH W. JONES.** 2003. Génétique : Les grands principes. 3<sup>ème</sup> édition Paris, *Dunod*. Pagination multiple. ISBN: 978-2-10-006735-0.
- **DONNAI R.** 2009. Génétique médicale : De la biologie à la pratique clinique. Édition Bruxelles, de Boeck. Pagination multiple. ISBN: 2-8041-5892-6.
- **DUNHAM I, KUNDAJE A, ALDRED S et al.** 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489(7414):57-74.
- **EBERHARTER A et BECKER PB.** 2004. ATP-dependent nucleosome remodelling: factors and functions. *Journal of Cell Science*. 117:3707-3711.
- **GRIFFITHS A, CARROLL S, WESSLER S et al.** 2013. Introduction à l'analyse génétique. 6<sup>ème</sup> édition. *Éditions De Boeck Supérieur*. Pagination multiple. ISBN:2-8041-6013-0.
- **HEARD E.** 2013. Épigénétique et mémoire cellulaire. *Éditions Fayard*. Pagination multiple. ISBN:2-2136-7948-7.

- **HURET JL, AHMAD M, ARSABAN M *et al.*** 2012. Atlas of genetics and cytogenetics in oncology and haematology in 2013. *Nucleic acids research*. 41(D1):D920-D924.
- **PÀLDI A.** 2018. L'épigénétique ou la nouvelle ère de l'hérédité. *Éditions le Pommier*. Pagination multiple. ISBN:2-7465-1684-5.
- **PALM W et DE LANGE T.** 2008. How shelterin protects mammalian telomeres. *Annual review of genetics*. 42:301-334.
- **RIDGWAY P, MAISON C ET ALMOUZNI G.** 2006. Functional organization of the genome: chromatin. *Atlas of Genetics and Cytogenetics in Oncology and Hematology*. 6(3).
- **SERRE JL.** 2006. Génétique Rappels de cours, exercices et problèmes corrigés. Dunod, Paris. Pagination multiple. ISBN : 2-1005-0524-6.
- **TURNER P.C, MCLENNAN A.G.** 2000. L'essentiel en biologie moléculaire. Port Royal Livre. *BERTI Edition*. Pagination multiple. ISBN: 9-7822-5720-433-2.