

Introduction aux acides nucléiques

1-1- Les acides nucléiques

Les acides nucléiques sont constitués par l'enchaînement d'éléments chimiques, les nucléotides (figure 1.1.). Un nucléotide est composé d'une base aromatique hétérocyclique reliée par une liaison N-glycosidique à la position 1' d'un sucre pentose cyclique (β -D-2'-désoxyribose pour l'ADN et β -D-ribose pour l'ARN), phosphorylé en position 5'.

Les bases classiquement rencontrées sont classées en deux catégories :

- les purines : adénine (A) et guanine (G)
- les pyrimidines : cytosine (C) et thymine (T) (ou uracile (U) pour l'ARN).

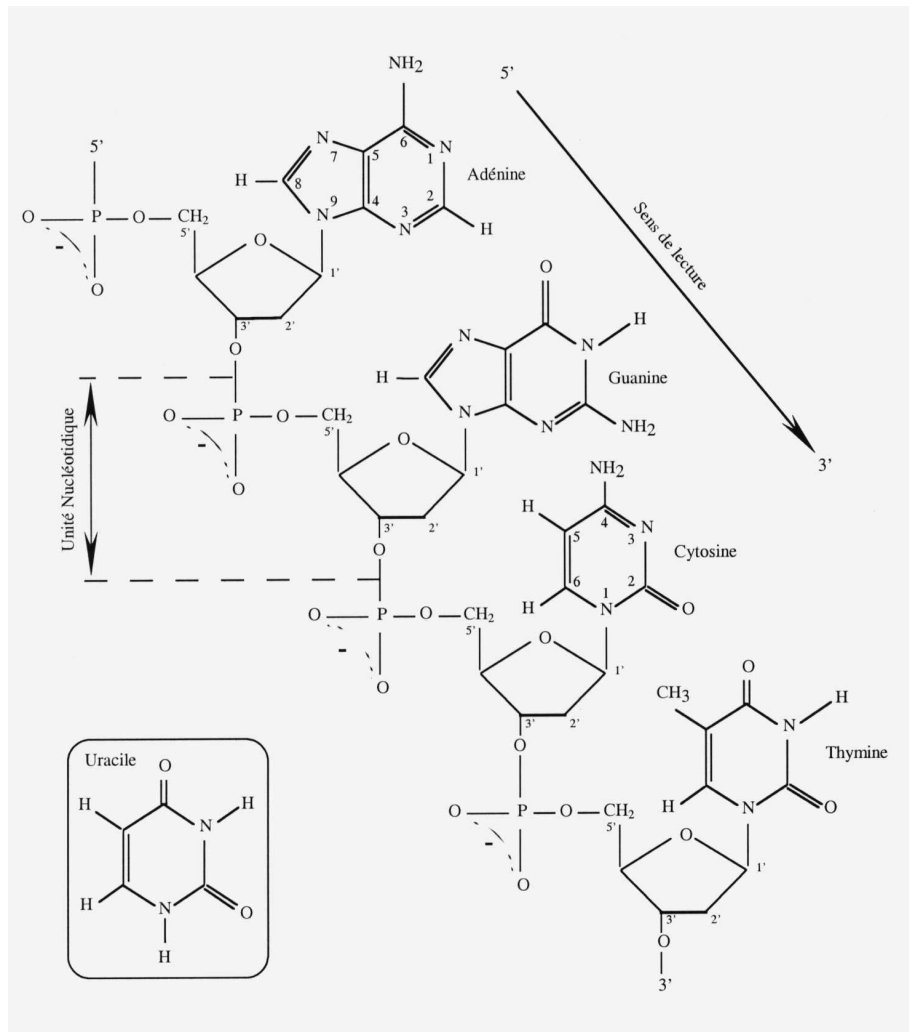


Figure 1.1. : Schéma d'un fragment d'ADN. Pour l'ARN, la base uracile et le sucre β -D-ribose remplacent la thymine et le sucre β -D-2'-désoxyribose de l'ADN.

Dans le cas des purines, c'est l'azote N₉ qui participe à la liaison N-glycosidique et dans le cas des pyrimidines, l'atome N₁.

Les sucres possèdent plusieurs degrés de liberté autour de différentes liaisons, et les angles dièdres définis par ces liaisons dépendent les uns des autres. Les conformations des sucres sont essentiellement C₂'-endo et C₃'-endo pour les doubles hélices d'ADN et d'ARN respectivement (figure 1.2.).

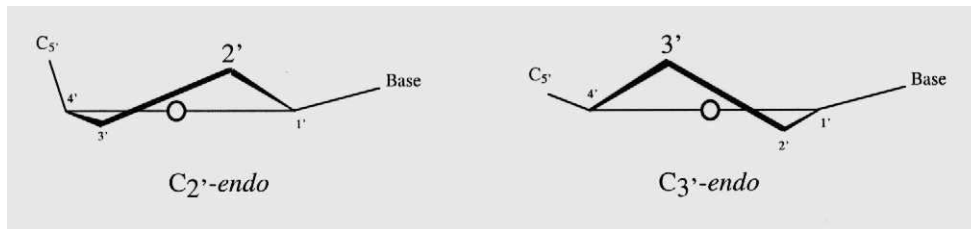


Figure 1.2. : Conformations C₂'-endo et C₃'-endo des sucres

La base peut tourner autour de la liaison N-glycosidique. L'espace conformationnel se réduit principalement à deux orientations, appelées *syn* et *anti* définies par l'angle χ , par suite de contrainte d'encombrement stérique (figure 1.3.).

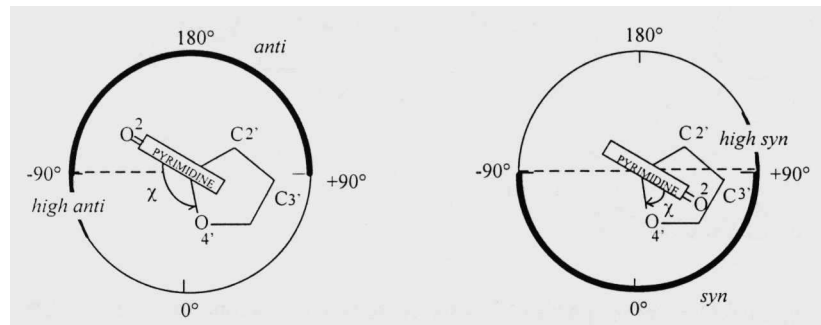


Figure 1.3. : Conformations *syn* et *anti* possibles autour de la liaison N-glycosidique pour une pyrimidine (Saenger, 1984)

Les acides nucléiques ont un caractère polyanionique dû à la présence de phosphates. Ceci permet des interactions avec des cations, modifie le pK_a apparent des groupements acido-basiques et intervient dans les phénomènes de répulsion électrostatique.

1-2- Structures de l'ADN

Maximisant les interactions hydrophobes et réduisant au minimum le nombre de molécules d'eau en son cœur, la structure de l'ADN est constituée d'un empilement hélicoïdal de plateaux de bases (nature aromatique des bases, molécules planes, hydrophobes et rigides). Les bases s'égrainent donc le long de la chaîne sucre-phosphate, colonne vertébrale de l'ADN. Classiquement, le sens de lecture d'une séquence d'ADN s'effectue de l'extrémité 5' vers l'extrémité 3' de cette chaîne (figure 1.1.).

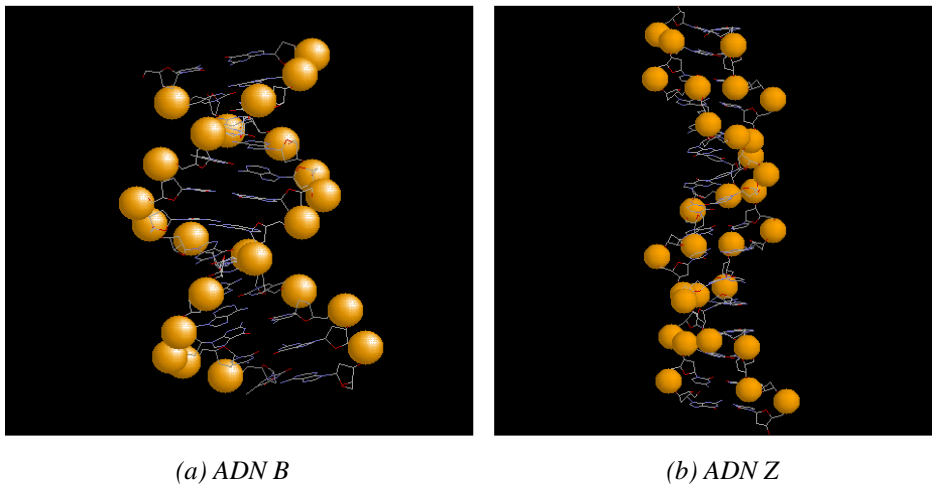


Figure 1.4. : Formes B et Z de l'ADN (<http://tome.cbs.univ-montp1.fr>)

La structure de l'ADN se caractérise par l'existence de deux chaînes polynucléotidiques appariées antiparallèlement l'une à l'autre en une structure hélicoïdale régulière. La forme dite « ADN B » découverte en 1953 par Watson et Crick (Watson & Crick, 1953 a et b) (figure 1.4.), est la structure stable la plus fréquemment rencontrée dans les conditions naturelles. Dans cette structure, une base pyrimidique est appariée à une purine (figure 1.5.).

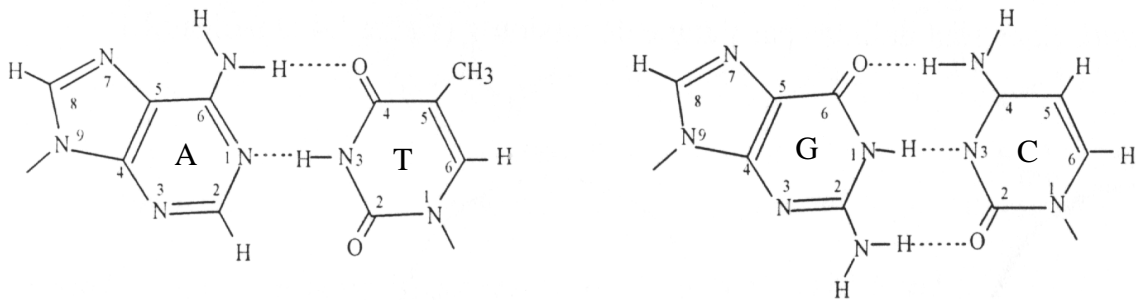


Figure 1.5. : Appariements Watson-Crick A-T et G-C

Deux autres configurations, l'ADN A et l'ADN Z ont été observées par cristallographie. Les conformations A et B montrent une forme hélicoïdale avec pas à droite, alors que l'hélice d'ADN Z (figure 1.4.) tourne dans l'autre sens. Des variations structurales dépendant de la séquence ont été découvertes, par exemple une courbure de l'hélice d'ADN (cas de l'ADN B').

Il existe encore d'autres structures dites non canoniques de l'ADN : ces structures ne s'appuient pas sur les paires de bases classiques Watson-Crick ou ne sont pas en double hélice anti-parallèle. Citons l'ADN π , un duplexe parallèle, les repliements en épingle à cheveux, et enfin les structures d'ADN à plusieurs brins telles que les hélices triples ou encore les quadruplexes de guanines et le motif i.

1-3- Les quadruplexes de guanines

Quatre guanines provenant de quatre brins distincts peuvent s'associer et former un plateau de guanines ou « tétrade de G » (figure 1.6.a.). La structure obtenue est appelée quadruplexe de guanines (figure 1.6.b.).

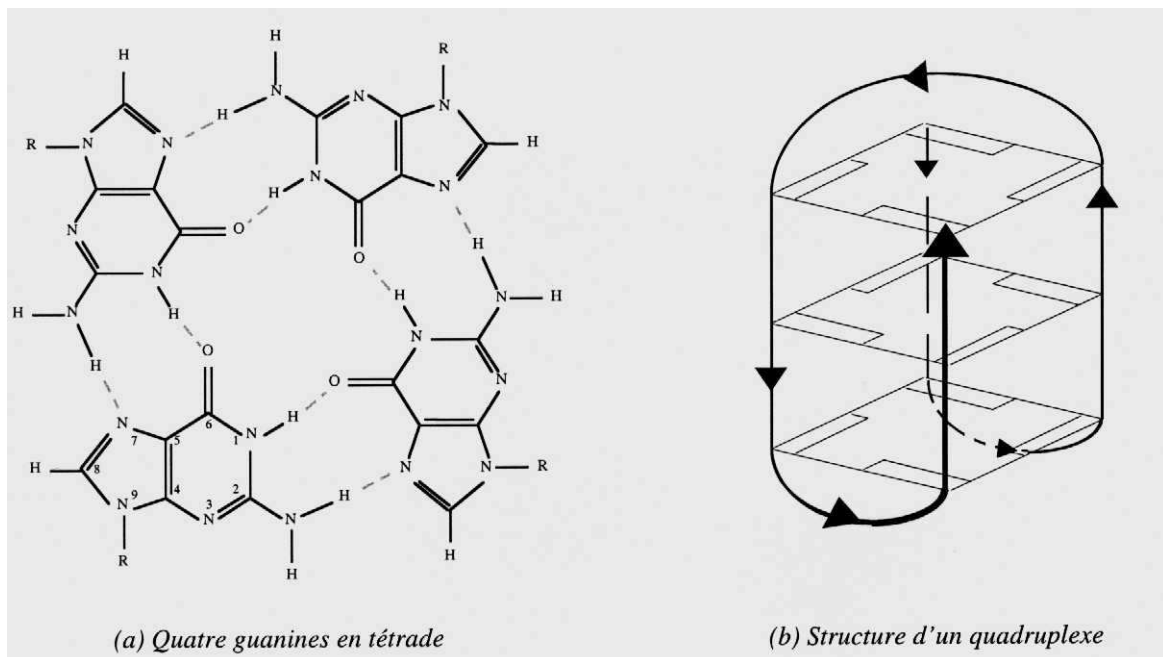


Figure 1.6. : (a) Tétrade de guanines. (b) Schéma du quadruplexe formé par le fragment du brin télomérique humain d[A(GGGTTA)₃GGG] (Wang & Patel, 1993).

L'empilement de tétrades est favorisé par la présence de cations monovalents s'insérant entre les plateaux (Patel et *al.*, 1999), entraînant une dépendance de l'architecture du quadruplexe en fonction de la nature du cation inséré (Sen & Gilbert, 1990; Miura et *al.*, 1999). Les structures les plus stables ont été observées quand des cations K^+ s'insèrent entre les tétrades de G.

Cette structure peut être formée de quatre brins pouvant provenir de fragments différents (quadruplexes dimériques et tétramériques) ou d'un seul (quadruplexe monomérique). Plusieurs orientations des brins sont possibles. Les brins d'un quadruplexe tétramérique sont orientés parallèlement, alors qu'il existe plusieurs conformations possibles pour les quadruplexes formés par repliement de un ou deux brins.

1-4- Le motif i

Première structure de l'ADN découverte par la RMN, le motif i est une structure reposant sur l'intercalation tête-bêche de deux duplexes parallèles formés de paires de bases C·CH⁺ (figure 1.7.a. et 1.7.b.) (Gehring et *al.*, 1993). Ses caractéristiques structurales principales peuvent se résumer ainsi :

- les 2 brins d'un même duplexe sont parallèles et reliés entre eux par des paires C·CH⁺
- la distance entre deux paires de bases intercalées est de ~ 0.31 nm, soit 0.62 nm entre deux bases consécutives d'un même brin (contre 0.34 nm dans le cas de l'ADN B)
- l'angle de rotation de l'hélice est compris entre 12° et 23° (contre 36° dans le cas de l'ADN B)
- le motif i présente deux sillons étroits et deux sillons larges
- l'angle de torsion glycosidique est toujours *anti* (figure 1.3.)
- la conformation des sucres est *C_{3'}'-endo*

Deux topologies d'intercalation sont possibles selon que les désoxycytidines situées à l'extérieur du motif i sont localisées à l'extrémité 3' ou à l'extrémité 5'. Comme dans le cas des quadruplexes de guanines, le motif i peut être intermoléculaire (motif tétramérique ou dimérique) ou intramoléculaire (motif i monomérique), mais l'orientation des brins doit obéir à l'intercalation tête-bêche des deux duplexes. La stabilité du motif i dépend des conditions expérimentales : pH, température, concentration en brins (pour les formes multimériques) et force ionique.

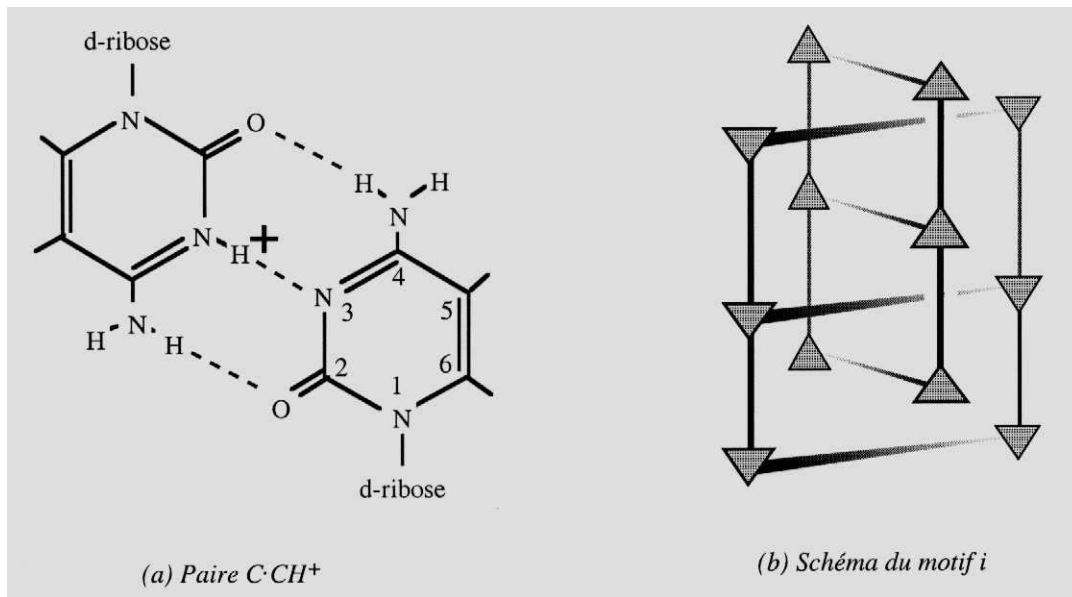


Figure 1.7. : Paire de bases C·CH⁺ et schéma du motif i adopté par le tétramère d^{(m)CCC}₄, (Leroy et al., non publié)

Malgré l'obligation d'hémi-protonation des paires C·CH⁺, le motif i intramoléculaire formé par la séquence naturelle télomérique humaine, d[(CCCTAA)₃CCCT] persiste à pH 7 (Phan et al., 2000), et celui formé par la séquence d(CCTTTCCTTTACCTTTCC) persiste même à pH 7.4 (Han et al., 1997 ; Guéron & Leroy, 2000 pour revue). Ces observations laissent supposer une existence possible du motif i *in vivo*.

Depuis la découverte du motif i, plusieurs études par cristallographie sont venues confirmer ces caractéristiques structurales (Berger et al., 1995 ; Chen et al., 1994 ; Kang et al., 1994 ; Kang et al., 1995).

2- Fonctions des acides nucléiques

2-1- Transmission et expression de l'information génétique contenue dans les acides nucléiques

L'information nécessaire à la synthèse des protéines est contenue dans des segments de chromosomes, les gènes. Pour les cellules eucaryotes, ces gènes peuvent être interrompus par des séquences non codantes, les introns. Les segments codants séparés par ces introns sont appelés exons. L'élaboration d'une protéine est réalisée à partir d'un gène lors de deux étapes distinctes : la transcription et la traduction.

La transcription est le transfert de l'information génétique contenue dans l'ADN d'un gène à un ARN messenger (ARN_m). Cette transcription est initiée par l'association d'ARN polymérase avec un site promoteur de l'ADN et conduit à la synthèse de plusieurs copies d'ARN_m.

Lors de la traduction, le ribosome, un complexe ARN-protéines, traduit l'information contenue par l'ARN_m en polymères linéaires d'acides aminés, les protéines. Cette traduction fait intervenir les ARN de transfert (ARN_t). Ceux-ci reconnaissent des segments de trois bases consécutives, les codons, en y associant les acides aminés correspondants (pour chaque acide aminé, il existe au moins un codon), les insérant à la position adéquate dans la séquence polypeptidique. La traduction de l'ARN_m aboutira à la fabrication des protéines correspondantes.

2-2- Mécanisme de réplication

La réplication génère une copie fidèle des chromosomes d'une cellule et s'effectue dans les deux directions à partir d'un point origine. Chaque brin sert de modèle à un nouveau brin. L'ADN est synthétisé avec une grande fidélité par une ADN polymérase qui fonctionne uniquement dans le sens 5' → 3' (figure 1.8.). Au niveau de la fourche de réplication, la synthèse de la nouvelle molécule d'ADN procède de façon asymétrique : le brin directeur est synthétisé de façon continue dans le sens de progression de la fourche de réplication, alors que l'autre brin est synthétisé de façon discontinue, par séquences de 1000 à 2000 bases, appelées « fragments d'Okazaki » (Okazaki & Okazaki, 1969).

La synthèse du brin discontinu s'appuie sur des amorçages répétés orchestrés par une primase générant des amorces d'ARN. Après chaque amorçage, il y a élimination de l'amorce d'ARN et reprise du processus de synthèse par une polymérase d'ADN à partir de l'extrémité 3' de l'ADN issue du processus d'initiation suivant. Ce mécanisme permet la synthèse de l'ensemble du brin à

synthèse discontinue, à l'exception du dernier segment (long d'une quinzaine de nucléotides) pour lequel il n'y a pas de synthèse possible. En l'absence d'autre mécanisme, il en résulte que la réplication de l'ADN s'accompagne à chaque fois d'une perte à l'extrémité 5' du chromosome.

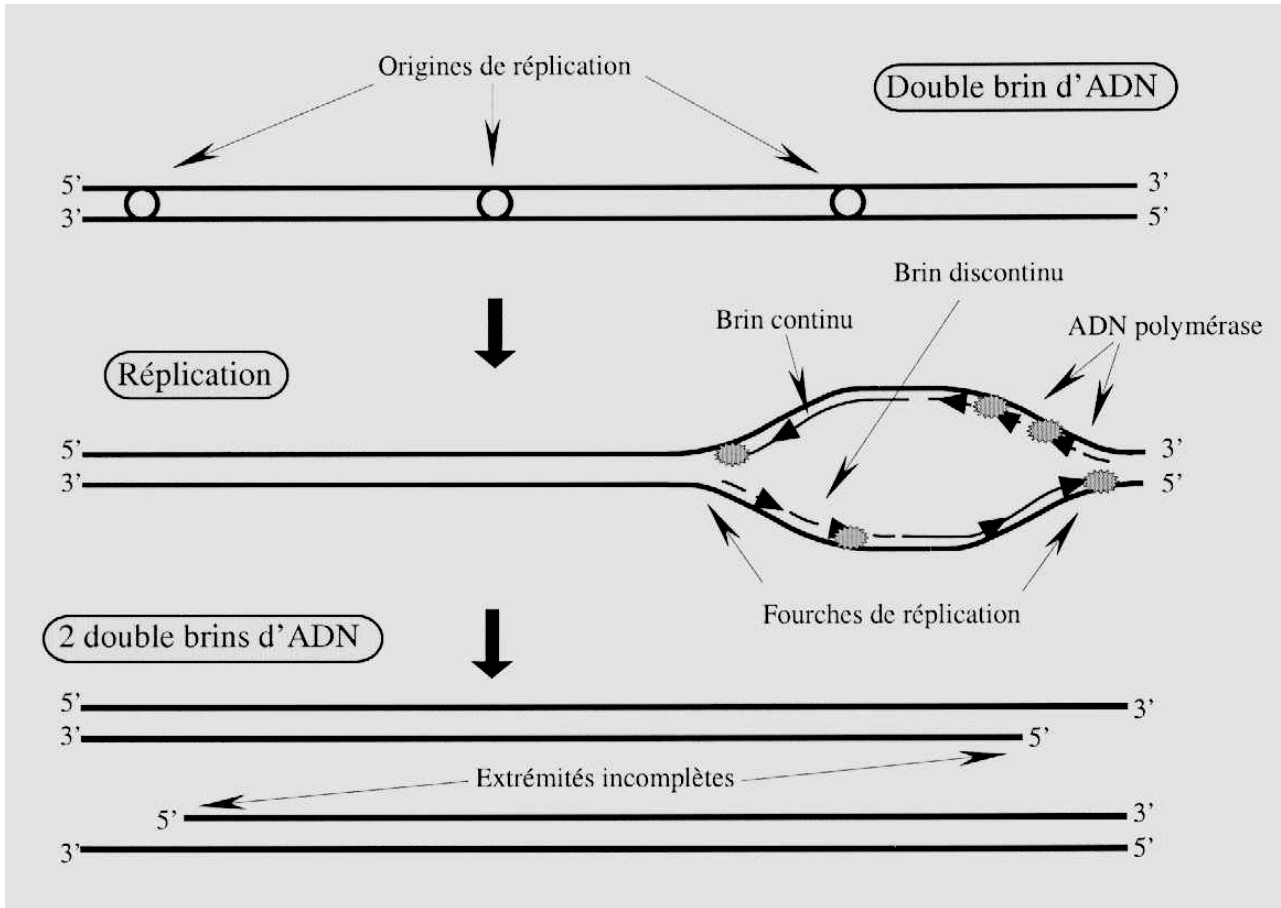
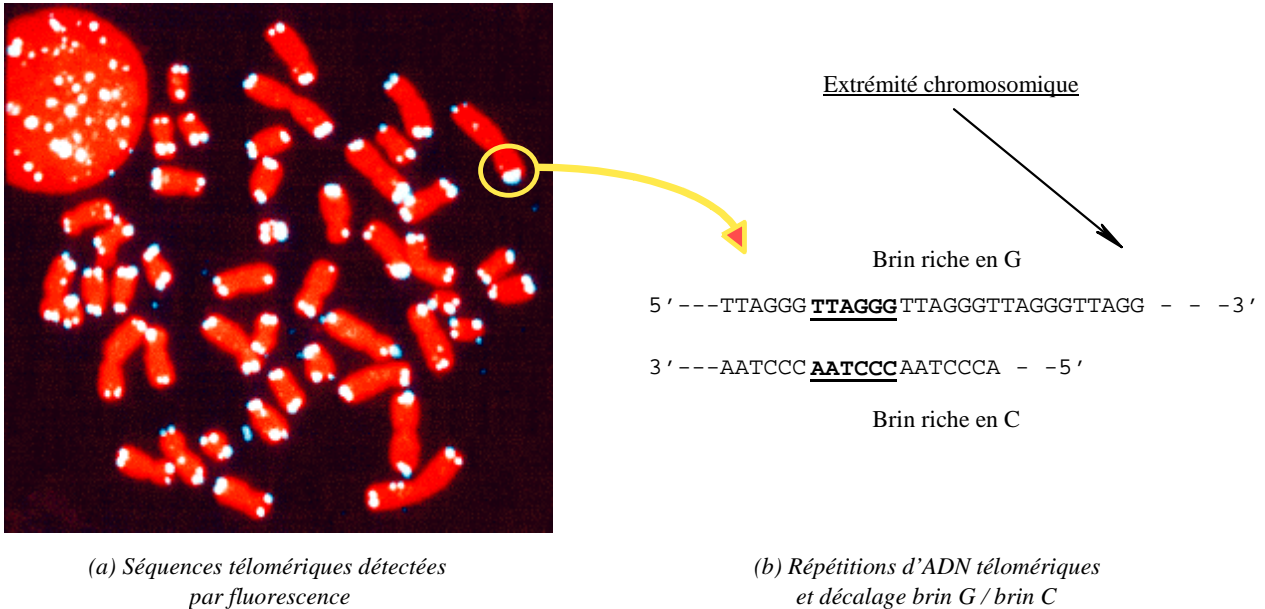


Figure 1.8. : Mécanisme de réplication; fonctionnement d'une fourche de réplication (Okazaki & Okazaki, 1969) et raccourcissement consécutif (Olovnikov, 1973).

2-3- Les télomères, extrémités des chromosomes

L'ADN codant ne représente que ~ 5% de l'ADN chromosomique humain. Le reste contient notamment des zones constituées d'un grand nombre de répétitions d'une séquence élémentaire. C'est le cas des extrémités des chromosomes, appelés télomères (figure 1.9.a.). Ces éléments structuraux, composés de répétitions d'une séquence d'ADN spécifique, associées à des protéines, ont un rôle de maintien de l'intégrité des chromosomes. Le schéma d'organisation de l'ADN télomérique de tous les organismes étudiés à ce jour présente une extension monocaténaire riche en

G à l'extrémité 3' (Greider, 1990 ; Blackburn, 1991 ; Greider, 1991 ; Blackburn, 1994 ; Fang & Cech, 1993 ; Price, 1990 ; Zakian, 1995). Par exemple, l'ADN télomérique des mammifères se compose de la séquence répétée, $d[(CCCTAA) \cdot (TTAGGG)]$ (figure 1.9.b.).



(a) Séquences télomériques détectées par fluorescence

(b) Répétitions d'ADN télomériques et décalage brin G / brin C

Figure 1.9. : (a) Chromosomes métaphasiques humain hybridés *in situ* avec la séquence humaine $d(TTAGGG)_n$ fluorescente (Kipling, 1995). (b) Schéma des deux brins télomériques des mammifères.

Les répétitions télomériques sont généralement régulières, à l'exception de celles de certains champignons et protozoaires. Leur longueur varie de quelques bases (cas des organismes ciliés) à plusieurs dizaines de kilobases (pour certains mammifères) (tableau 1.1.). A titre d'exemple, la longueur de l'ADN télomérique humain est supérieure à 10000 paires de bases (de Lange, 1994) et la longueur du monobrin riche en G est d'une centaine de bases (Blackburn, 1991).

Enfin, chacun des brins de l'ADN télomérique pris séparément peut former *in vitro*, soit un quadruplexe de guanines (Sen & Gilbert, 1988), soit un motif i (Ahmed et al., 1994 ; Leroy et al., 1994 ; Manzini et al., 1994). Par exemple, les structures par RMN du quadruplexe de guanines d'un fragment d'ADN télomérique humain (Wang & Patel, 1993) et du motif i de la séquence complémentaire (Phan et al., 2000) ont pu être déterminées.

Organismes	Longueurs (en paires de bases)	Séquences brin C 5'→3'	Séquences brin G 5'→3'
Protozoaires			
<i>Tetrahymena</i>	< 100 ^a	C ₄ A ₂	T ₂ G ₄
<i>Oxytricha, Euplotes</i>	< 100 ^{b,c}	C ₄ A ₄	T ₄ G ₄
<i>Paramecium</i>	< 100 ^a	C ₃ [C/A]A ₂	T ₂ [T/G]G ₃
<i>Plasmodium</i>	~ 1500	C ₃ T[G/A]A ₂	T ₂ [T/C]AG ₃
Champignons			
<i>Neurospora</i>		C ₃ TA ₂	T ₂ AG ₃
<i>Cryptococcus</i>		C ₃₋₅ TA ₂	T ₂ AG ₃₋₅
<i>Saccharomyces cerevisiae</i>	< 500 ^d	C ₂₋₃ A(CA) ₁₋₆	(TG) ₁₋₆ TG ₂₋₃
<i>Schizosaccharomyces</i>		C ₁₋₆ G ₀₋₁ T ₀₋₁ GTA ₁₋₂	T ₁₋₂ ACA ₀₋₁ C ₀₋₁ G ₁₋₆
Plantes			
<i>Chlamydomonas</i>	< 500 ^e	C ₃ TA ₄	T ₄ AG ₃
<i>Arabidopsis</i>	3000 ^f	C ₃ TA ₃	T ₃ AG ₃
Invertébrés			
<i>Ascaris</i>		GC ₂ TA ₂	T ₂ AG ₂ C
<i>Bombyx mori</i>	> 10000 ^g	C ₂ TA ₂	T ₂ AG ₂
Vertébrés^h			
<i>Homme</i>	> 10000 ⁱ	C ₃ TA ₂	T ₂ AG ₃
<i>Mus musculus</i>	> 50000 ^j	C ₃ TA ₂	T ₂ AG ₃

Tableau 1.1. : Séquences télomériques de quelques organismes (d'après Wellinger & Sen, 1997). ^a Blackburn & Gall, 1978 ; ^b Oka et al., 1980 ; ^c Wellinger et al., 1993 ; ^d Shampay et al., 1984 ; ^e Petracek et al., 1990 ; ^f Richards & Ausubel, 1988 ; ^g Okazaki et al., 1993 ; ^h Meyne et al., 1989 ; ⁱ de Lange, 1994 ; ^j Kipling & Cooke, 1990.

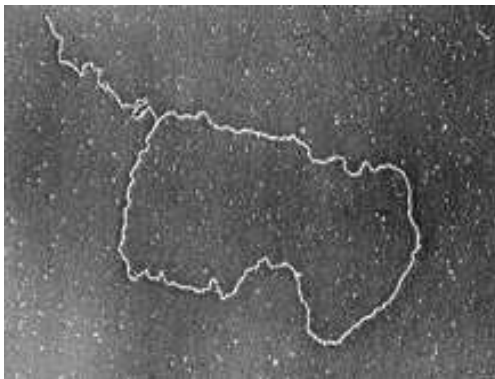
2-4- Mécanismes liés à la régulation des télomères

Lors de la réplication de la cellule, la longueur des télomères diminue (environ 50-200 bases par division pour les mammifères) selon le principe proposé par Olovnikov (Olovnikov, 1996 pour revue). En l'absence de mécanisme compensateur, ce phénomène limite le nombre possible de réplifications cellulaires. A partir d'une longueur critique, un mécanisme conduit à la mort de la cellule. Mais le raccourcissement peut être corrigé par l'intermédiaire d'une transcriptase inverse spécialisée, la télomérase (Blackburn, 1992 ; Lingner et al., 1995 ; Counter, 1996 ; Chiu & Harley, 1997 ; O'Reilly et al., 1999 pour revue récente). Elle se compose d'une partie protéique et d'un ARN servant de matrice pour la synthèse de l'ADN télomérique. La région de l'ARN télomérase humain servant de modèle est le fragment r(CUAACCCUAAC) codant pour la répétition d'ADN télomérique d(TTAGGG).

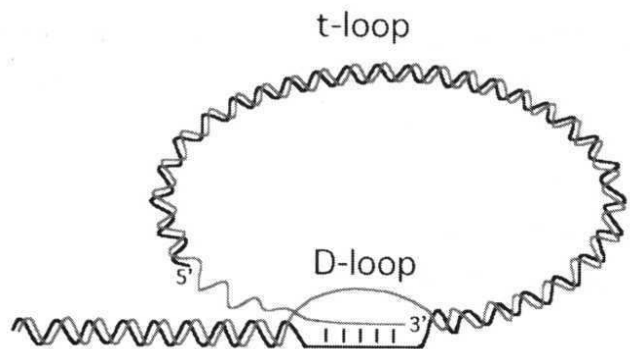
L'environnement nucléoprotéique d'un télomère régule l'activité de la télomérase. Des expériences ont montré *in vivo* que l'activité de la télomérase était régulée par la présence de protéines s'associant au duplexe (Bianchi *et al.*, 1997 ; Billaud *et al.*, 1997 ; Van Steensel & de Lange, 1997 ; Ancelin *et al.*, 1998 ; Van Steensel *et al.*, 1998) et qu'elle nécessitait la présence de certaines protéines affines du brin G ainsi qu'une dégradation sélective du brin C (Wellinger *et al.*, 1996 ; Makarov *et al.*, 1997).

Un dérèglement de l'activité de la télomérase pourrait être responsable de l'apparition de certaines tumeurs (Shay, 1997 ; Holt & Shay, 1999). Ainsi, il a été récemment montré que l'inactivation de la télomérase jouait un rôle limitant dans la croissance des cellules tumorales, mais son activation seule ne suffit pas pour immortaliser toutes les cellules primaires ou les transformer en cellules tumorales (Hahn *et al.*, 1999a et b).

L'extrémité télomérique des chromosomes présente une caractéristique structurale récemment mise en évidence (Griffith *et al.*, 1999). Les télomères forment une boucle (observation par microscopie électronique (figure 1.10.a.) sur des extrémités télomériques de clone de cellule *HeLa*), la t-boucle. Au niveau de la fermeture de cette t-boucle, le monobrin G s'associerait avec la double hélice télomérique pour former une D-boucle (figure 1.10.b). Cette D-boucle pourrait se déplacer le long du télomère et sa structure serait induite par la protéine TRF2 (Greider, 1999).



(a) Photographie d'une extrémité télomérique d'un chromosome humain observée par microscopie électronique (Griffith *et al.*, 1999)



(b) Schéma de la structure possible de l'extrémité des télomères (Lundblad, 2000)

Figure 1.10.: (a) Observation par microscopie électronique d'une t-boucle (ADN télomérique d'un clone de cellule *HeLa*) ; (b) Schéma proposé par Greider (Greider, 1999) de la structure du télomère.

Ces observations placent dans un contexte nouveau le suicide programmé d'une cellule (appelé apoptose). Ce phénomène reste mal compris et les mécanismes biologiques permettant de distinguer un télomère trop court, non élucidés. On pourrait imaginer que la t-boucle ne puisse plus se former si le télomère est trop court. La cellule pourrait le détecter alors comme un chromosome cassé et déclencher l'apoptose.

Plusieurs hypothèses structurales sont envisagées concernant les mécanismes associatifs intervenant au sein de la D-boucle :

- le monobrin G s'associe au brin C en s'insérant en lieu et place du brin G normalement impliqué dans le duplexe d'ADN télomérique. Celui-ci peut alors soit rester libre, soit se replier en quadruplexe de G ;
- les deux brins G forment chacun un quadruplexe ou s'associent en un quadruplexe multimérique et le brin C reste libre ou forme un motif i ;
- les trois brins s'associent et forment une triple hélice C·G×G selon un modèle similaire à celui proposé pour *Tetrahymena* (Veselkov et al., 1993).

Actuellement, aucune des solutions proposées n'a été confirmée. La compréhension des interactions entre les brins télomériques et les facteurs extérieurs intervenant dans la régulation des télomères laissent envisager de nouveaux champs d'application thérapeutique concernant le vieillissement cellulaire et pourraient permettre de comprendre les étapes liées à la cancérogenèse. Différentes études sont en cours concernant, par exemple la possible formation d'une triple hélice C·G×G à partir de fragments d'ADN télomérique humain, ou encore, la stabilisation de quadruplexe de G par des ligands.

2-5- Protéines spécifiques des télomères

Plusieurs groupes de recherches (Blackburn, 1994 ; Zakian, 1995 ; Greider, 1996 ; Holt et al., 1996 ; Olovnikov, 1996 ; Greider, 1998 ; Shore, 1998 ; Gottschling & Stoddard, 1999 ; O'Reilly et al., 1999) ont dirigé leurs efforts dans la découverte et la caractérisation de protéines se liant spécifiquement aux séquences télomériques. Des protéines ont été identifiées comme reconnaissant le double brin, par exemple Rap1p pour la levure *Saccharomyces cerevisiae* (Giraldo & Rhodes, 1994, Giraldo et al., 1994) (figure 1.12), PPT pour *Physarum* (Coren et al., 1991) et Trf1p/Trf2p

dans les cellules des mammifères (Zhong et al., 1992 ; Chong et al., 1995 ; Billaud et al., 1997) (figure 1.11.).

Le simple brin télomérique riche en G de *Saccharomyces cerevisiae* est reconnu par les protéines Est1p (Virta-Pearlman et al., 1996) et Cdc13p, qui assurent un rôle de protection du télomère (Porter et al., 1996 ; Lin & Zakian, 1996). Le simple brin est aussi reconnu par les sous-unités protéiques α et β chez *Oxytricha* (Raghuraman & Cech, 1989), TBP chez *Euplotes crassus* (Price, 1990), TEP et TGP chez *Tetrahymena* (Sheng et al., 1995 ; Schierer & Henderson, 1994), GBP chez *Chlamydomonas* (Petracek et al., 1994), XTEF chez *Xenopus* (Cardenas et al., 1993), MyoD dans certaines cellules de mammifères (Weisman-Shomer & Fry, 1993) et hnRNP A2/B1 dans les cellules *HeLa* (Ishikawa et al., 1993). D'autres protéines interagissent spécifiquement avec les tétrades de G (Frantz & Gilbert, 1995 ; Liu et al., 1995), ou encore induisent cette formation *in vitro* (Giraldo & Rhodes, 1994 ; Giraldo et al., 1994).

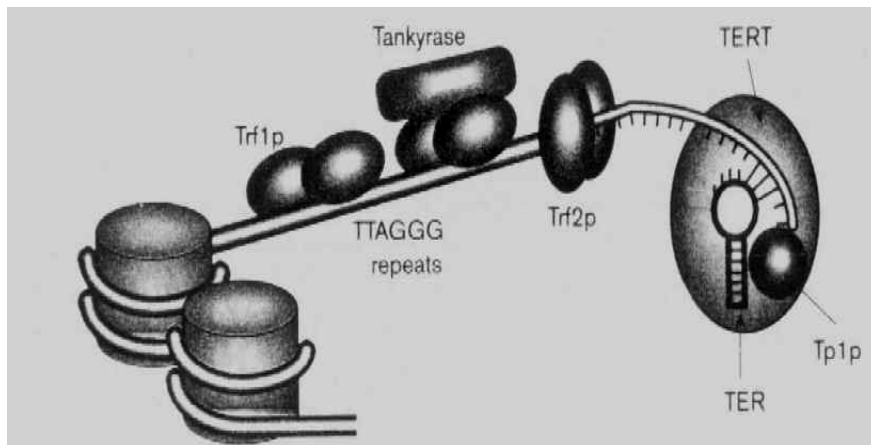


Figure 1.11. : Schéma d'une organisation fonctionnelle possible de l'extrémité télomérique humaine (O'Reilly et al., 1999).

Concernant les séquences télomériques complémentaires riches en C, une protéine se liant spécifiquement au fragment d((CCCTAA)₃CCCT) a été mise en évidence pour les cellules humaines *HeLa* (Marsich et al., 1996) et les cellules de mammifères (Marsich et al., 1998). Deux autres protéines issues de cellules *HeLa*, ASF/SF2 et hnRNP K, ont été identifiées comme reconnaissant spécifiquement la séquence naturelle d((TAACCC)₄TAA) (Lacroix et al., 2000). Ces protéines sont à même de s'associer avec d'autres séquences riches en C capables d'adopter le motif *in vitro*, mais on ne connaît pas encore la structure de l'ADN dans le complexe.

2-6- Région centromérique

Chez l'homme, la zone centrale des chromosomes, appelée centromère, présente elle aussi des répétitions de séquences élémentaires riches en cytosines sur un brin et en guanines sur le brin complémentaire. Défini par les cytologistes comme la constriction primaire des chromosomes, le centromère intervient lors de la méiose et de la mitose. Le centromère humain représente environ 10% du génome, et se compose principalement de 5 satellites, dont les plus importants sont les satellites α et III. Le premier comporte des suites répétées de 171 paires de bases, contenant une sous-séquence de 17 nucléotides, d(TCCCGTTTCCAACGAAG), (site de l'association d'une protéine, la centromeric protein B ou CEN-P B), dont une partie, d(TCCCGTTTCCA) forme *in vitro* un motif i dimérique (Gallego et al., 1997).

Le satellite III contient des millions de bases de répétitions presque parfaites de la séquence d((CCATT)(TAAGG)) (Grady et al., 1992). D'un point de vue structural, on sait que le brin riche en G peut adopter *in vitro* des structures 'tige-boucle' (Chou et al., 1994 ;Zhu et al., 1995) et que la séquence naturelle d[(CCATT)₃CC] peut former des structures en motif i monomériques ou dimériques (forme favorisée par la présence de KCl) (Nonin et al., 1997 ; Nonin-Lecomte & Leroy, 2001).

A la différence des régions télomériques et de certaines protéines centromériques (CENP-A, B et C) remarquablement conservées entre espèces différentes, les ADN centromériques des différents mammifères ne présentent quasiment aucune homologie. Dans ce contexte, la variabilité des structures en motif i (mono- ou dimériques selon les séquences de boucles) adoptées par le brin C du satellite III *in vitro*, n'exclut pas un rôle biologique. Le brin complémentaire riche en G présente une variabilité comparable.

3- Le sujet de thèse

Les télomères, comme les centromères, présentent deux brins complémentaires pouvant adopter *in vitro* deux structures non canoniques, le quadruplexe de guanines et le motif i. Ce particularisme structural et l'importance biologique des régions chromosomiques impliquées renforcent la question

du rôle biologique de ces structures. Depuis la découverte du motif i, et de protéines associées au brin G initiant la formation de quadruplexes de guanines, les régions riches en répétitions de cytosines et de guanines, et plus particulièrement, les télomères, font l'objet d'un regain d'intérêt. Afin de caractériser une éventuelle fonction biologique du motif i, nous avons entrepris de rechercher des protéines capables de reconnaître cette structure. Cette étude a été entreprise chez une levure, *Saccharomyces cerevisiae*, dont le génome présentait l'avantage d'être totalement séquencé.

3-1- Les télomères de levure

Chez la levure *Saccharomyces cerevisiae*, la longueur des télomères est d'environ 300 ± 75 paires de bases. La séquence consensus du brin C des télomères est $C_{2-3}A(CA)_{1-6}$, mais il n'existe pas de répétitions régulières comme la séquence C_3TA_2 pour les mammifères. La matrice d'ARN de la télomérase de *Saccharomyces cerevisiae*, r ($3'ACACACACCCACACCAC5'$), plus longue que celle de la plupart des autres organismes [par exemple, r($3'CAATCCCAATC5'$) chez l'homme], n'est que partiellement copiée, selon un positionnement variable de l'amorce d'ADN non encore élucidé (Nugent & Lundblad, 1998 pour revue). Le gène *EST2* code la transcriptase inverse de la télomérase. Les protéines Est1p, Est3p et Est4p/Cdc13p composant la télomérase peuvent avoir un rôle de facteur de régulation. Un ensemble spécifique de protéines (Cdc13p, Est1p, Stn1p) est probablement directement impliqué dans la protection et la régulation de la longueur des télomères par la télomérase (Lin & Zakian, 1996).

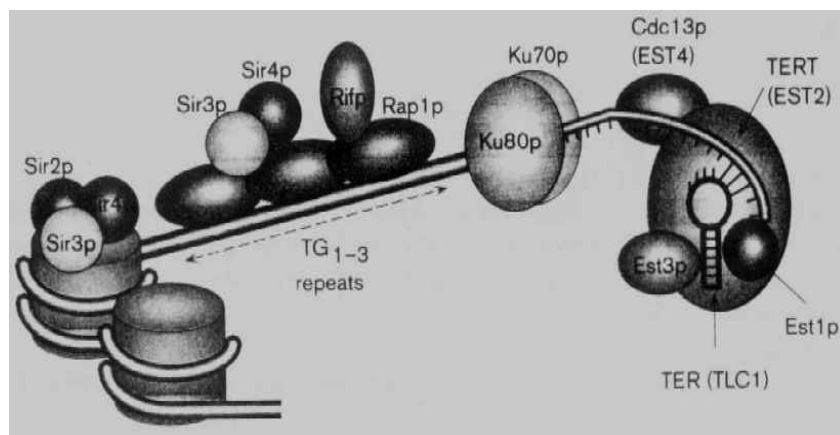


Figure 1.12. : Schéma d'une organisation fonctionnelle possible de l'extrémité télomérique de *Saccharomyces cerevisiae* (O'Reilly et al., 1999).

L'élongation du brin G des télomères de *Saccharomyces cerevisiae* est étroitement liée à celle du brin opposé par les ADN polymérases α et δ (Diede & Gottschling, 1999). Plusieurs protéines, s'associant à l'ADN télomérique de *Saccharomyces cerevisiae*, coopèrent et participent aussi à la régulation de la longueur des télomères (figure 1.12., voir tableau non exhaustif en annexes) dont :

- les protéines Sir1p, Sir2p, Sir3p et Sir4p (Bourns et *al.*, 1998) par ailleurs impliquées dans un phénomène de régulation des gènes, le « silencing »;
- les protéines Rif1p et Rif2p (Bourns et *al.*, 1998) interagissant avec Rap1p (Rap1p interacting factor). Rif2p coopère avec Rif1p pour réguler la longueur des télomères (Wotton & Shore, 1997) ;
- la protéine Rap1p dont la surproduction altère la structure du télomère (Conrad et *al.*, 1990). Cette protéine favorise aussi la formation de quadruplexe de guanines (Giraldo et *al.*, 1994). Elle peut induire une distorsion de la double hélice d'ADN télomérique (Gilson et *al.*, 1993) et est également impliquée dans l'association des deux brins d'ADN (Gilson et *al.*, 1994).
- les protéines Ku70p et Ku80p reconnaissent la transition entre le double brin télomérique et l'extension monocaténaire riche en G, plus longue (Bianchi & de Lange, 1999) et interviennent en cas de dommage sur le double brin d'ADN (Martin et *al.*, 1999).
- la protéine Tel2p interagit avec le double brin d'ADN télomérique (Kota & Runge, 1998) et se fixe spécifiquement sur le simple brin G. Elle pourrait avoir un rôle dans la régulation des télomères (Kota & Runge, 1999).

Les gènes situés à proximité des télomères peuvent être inactivés par un phénomène appelé « silencing » (Loo & Rine, 1995), où l'on retrouve impliquée la protéine Rap1p (Cockell M. et *al.*, 1995), les protéines Ku70p et Ku80p (Martin et *al.*, 1999 ; Mishra & Shore, 1999) et les protéines Sir (Ivy et *al.*, 1986). Les télomères servent alors de réservoir pour des facteurs de transcription impliqués dans la structure de la chromatine réprimée (Marcand et *al.*, 1996), ce qui permet de faire un lien entre l'activité transcriptionnelle de la cellule et l'intégrité des télomères.

Un autre phénomène appelé « Telomere Position Effect » ou TPE (Gottschling et *al.*, 1990, Kyrion et *al.*, 1993) régule l'expression des gènes placés à proximité des télomères. Plusieurs protéines déjà citées jouent un rôle dans ce phénomène, parmi lesquelles, la protéine Tel2p (Runge & Zakian, 1996 ; Zakian, 1996) et les protéines Ku70p et Ku80p (Evans et *al.*, 1998 pour revue).

3-2- Présentation du sujet de thèse

L'étude développée dans le présent ouvrage a été entreprise à partir d'une séquence télomérique et d'extraits de protéines de *Saccharomyces cerevisiae*, respectivement synthétisés et produits au laboratoire.

Ce projet comporte plusieurs étapes :

- mettre en évidence une (ou plusieurs) protéine(s) de levure se liant spécifiquement à une séquence riche en cytosines ;
- déterminer les caractéristiques et la spécificité de cette association ;
- identifier les protéines impliquées ;
- purifier les protéines identifiées et vérifier qu'elles s'associent avec le brin C de l'ADN télomérique ;
- comprendre les mécanismes de cette association.