

UNIVERSITAT DE BARCELONA

FACULTAT  
FARMÀCIA

DEPARTAMENT  
BIOQUÍMICA I BIOLOGIA MOLECULAR

LES PROPIETATS FÍSiques DE L'ADN EN ESCALA GENÒMICA

Josep Ramon Goñi Macià 2008

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA

DEPARTAMENT  
BIOQUÍMICA I BIOLOGIA MOLECULAR

# 1 INTRODUCCIÓ I

Ambiciosos projectes de seqüenciació (The Genome Sequence of *Drosophila melanogaster*, 2000a; The *C. elegans* Sequencing Consortium, 1998; Lander et al., 2001; J C Venter et al., 2001; Sequencing Consortium, 2002; Mark D. Adams et al., 2000) han revelat la seqüència primària del genoma de quasi 50 eucariotes, incloent molts mamífers, i molts altres estaran disponibles en els pròxims anys. L'anàlisi massiu de tanta quantitat de dades està donant com a fruit la localització de gens, i la seva estructura primària, en els mapes genòmics plans. No podem oblidar però que, dins la cèl·lula, l'ADN no és un polímer estàt uniformement on tots els segments són accessibles de forma igual. Tot al contrari, es posen de relleu complexes estructures, altament dependents de la seqüència de dinucleòtids, de modificacions epigenètiques, de canvis de l'entorn o de la presència de petites molècules o proteïnes.

La dificultat per determinar experimentalment les regions reguladores fa que les dades sobre els mecanismes que controlen l'expressió de gens sigui limitada. Es coneix que la majoria d'elements reguladors (Solovyev & Shahmuradov, 2003a) es troben en la regió immediatament anterior (centenars de posicions abans) a l'inici del gen (factors de transcripció, TF), però altres senyals es troben en posicions molt llunyanes (potenciadors o *enhancers*). Com es produeix el flux d'informació d'aquestes regions distals i el promotor proper, o dins de la regió promotora és desconegut, encara que es pensa en la formació de complexes multiprotèics. Tampoc queda clar quin mecanisme permet als factors de transcripció distingir una seqüència diana (sempre relativament curtes) en un determinat context gènic i no en un altre.

És en tots aquests aspectes on pensem que l'estructura tridimensional de la molècula d'ADN (i la seva capacitat per deformar-se) pot jugar un factor determinant en possibilitar que aquests elements llunyans en un pla unidimensional (seqüència) estiguin molt propers en l'espai tridimensional, que el posicionament dels nucleosomes permeti la unió efectiva de factors de transcripció i les seves seqüències dianes i que la deformabilitat de l'ADN permeti només en el cas biològicament rellevant la formació d'estructures quaternàries DNA-proteïnes (Shpigelman, E. N. Trifonov, & A. Bolshoy, 1993; Pedersen, Jensen, Brunak, Staefeldt, & Ussery, 2000; Josep Ramon

Goñi, Vaquerizas, J. Dopazo, & Modesto Orozco, 2006; J Ramon Goñi, Pérez, Torrents, & Modesto Orozco, 2007).

Un cas dramàtic en que les propietats físiques modulen la funcionalitat de l'ADN és quan es formen estructures inusuals a les regions promotores que poden interferir la funcionalitat del gen, reduint-la (*knowdown*) o anul·lant-la (*knockout*). Això es el que ha estat descrit per diferents gens quan es formen triplexes (hèlix de tres cadenes) en la regió promotora (Majumdar et al., 1998; Aviñó, Elena Cubero, González, Ramon Eritja, & Modesto Orozco, 2003; Spacková, Elena Cubero, Jirí Sponer, & Modesto Orozco, 2004a). Aquest mecanisme de control de la expressió gènica, pot tenir un interès biomèdic en tant que afegint oligonucleòtids d'una cadena de seqüència adequada es pot arribar a bloquejar un gen patògen. L'estudi, la localització i anotació de les propietats físiques de DNA o d'elements estructurals del genoma tenen doncs un fort impacte no només en el coneixement bàsic del genoma, sinó en la recerca de noves dianes terapèutiques.

## 1.1 OBJECTIUS

L'objectiu principal de la tesis ha estat aprofundir en el coneixement de les propietats físiques de l'ADN en mecanismes cel·lulars. Tres blocs principals han centrat la major part de la recerca realitzada:

1. L'anàlisi exhaustiu en el genoma humà de dianes antigèniques mitjançant la formació de tríplex.
  - a. Localització i identificació de seqüències formadores de tríplex.
  - b. Estudi del paper biològic d'aquestes dianes.
  - c. Estudi comparatiu de dianes de tríplex en altres espècies.
2. Estudi de propietats helicoïdals de l'ADN derivades a partir de models teòrics

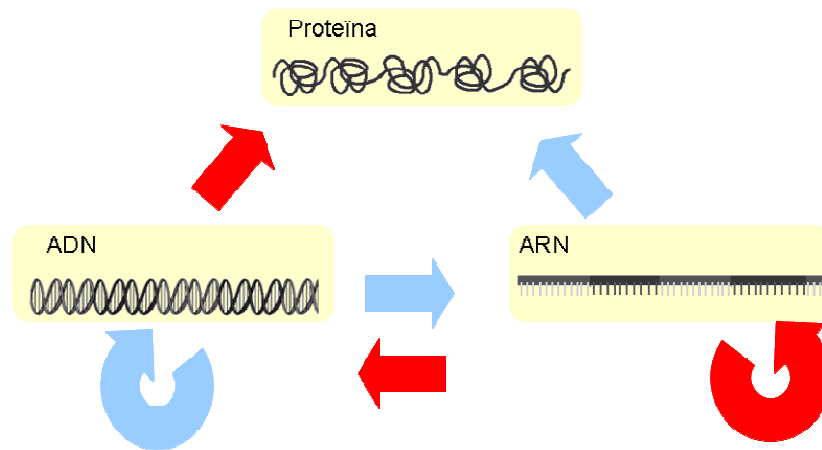
en la regulació dels gens.

- a. Identificació de propietats físiques de l'ADN amb un patró diferenciat en regions reguladores de vertebrats.
  - b. Demostració del paper regulador de les propietats físiques de la molècula d'ADN desenvolupant un mètode predictor basat exclusivament en aquest component.
3. Desenvolupament d'eines bioinformàtiques per a l'estudi de les propietats físiques de l'ADN.
- a. Implementació de desenes de mètodes predictius de propietats físiques d'ADN
  - b. Creació d'una plataforma per la visualització d'aquest propietats físiques en mapes bidimensionals i tridimensionals
  - c. Desenvolupament d'un mètode per la predicció del moviment de la molècula d'ADN
  - d. Integració del les eines anteriors en una plataforma basada en Web i en la llibreria de serveis-web del Institut Nacional de Bioinformàtica (INB)



## 2 INTRODUCCIÓ II: REGULACIÓ GENÈTICA

Tot i que els gens són els elements principals del genoma, es coneix que en organismes eucariotes, hi ha molt més ADN no-codificant que ADN que codifiqui proteïnes, tot i que sabem que al menys, part d'aquest ADN no-codificant té una funció important (Alberts et al., 1999). Per exemple, la regió immediatament superior al inici dels gens serveix en molts casos per regular la seva expressió. Amb aquest ADN regulador, les cèl·lules eucariotes han evolucionat de diferents maneres de controlar quan i on un gen s'ha d'activar, el que resulta clau en la formació d'organismes multicel·lulars complexos.



**Figura 2.1** Flux de transferència biològica. En fletxes blaves apareixen les transferències existents en la majoria d'organismes (en vermell, les especials).

### 2.1 EL DOGMA CENTRAL DE LA BIOLOGIA MOLECULAR

El dogma és un marc per entendre la transferència de les dades genètiques entre biopolímers transportadors d'informació, en la majoria d'organismes vius (F Crick, 1958, 1970). Hi ha tres classes principals de biopolímers: l'ADN, l'ARN (els dos àcids

nuclèics) i les proteïnes. Segons el dogma existeixen tres tipus de transferències d'informació en la majoria de les cèl·lules (veure la figura 2.1). Es coneix també que sota condicions especials (virus o en laboratoris) es poden donar tres tipus més de transferències (veure la figura 2.1). Totes elles es descriuen en les seccions següents.

### **2.1.1 Transferències de la informació biològica**

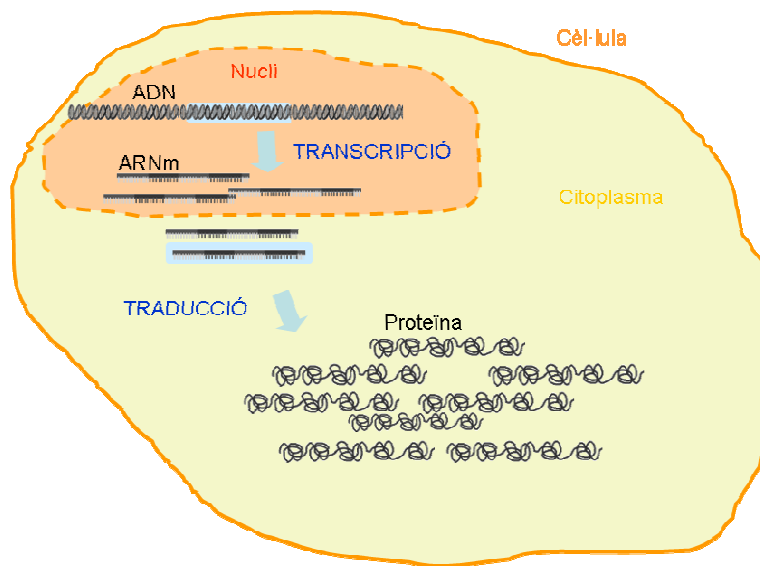
El grup de transferències generals existents en la majoria de cèl·lules descriuen el flux normal d'informació genètica (veure figura 2.2): l'ADN es copia a ADN (replicació d'ADN), la informació de l'ADN pot ser copiada a ARN missatger o ARNm (transcripció) i finalment les proteïnes poden ser sintetitzades (F Crick, 1970) usant la informació transportada en l'ARNm (traducció).

#### *2.1.1.1 Replicació de l'ADN*

La replicació de l'ADN és el procés de copiar una molècula de doble cadena d'ADN per formar dues molècules de doble cadena idèntiques a la primera (Alberts et al., 1999). El procés de replicació d'ADN és un procés fonamental usat en tots els organismes vius com a base de la herència biològica. Com que cada cadena d'ADN conté la mateixa informació genètica, les dues cadenes poden servir com a motlle per la reproducció de la cadena complementaria. Aquest procés, que es coneix com a replicació semi-conservativa es completa a partir del manteniment de la cadena motlle i l'ensamblatge de nous nucleòtids (veure el capítol Estructura de l'ADN i la figura 2.3). Per garantir que la replicació de l'ADN doni com a resultat dues dobles cadenes idèntiques existeixen mecanismes biològics d'autenticació de lectura i de revisió d'errors.

En una cèl·lula la replicació ha de produir-se abans de la divisió cel·lular. La síntesis de l'ADN comença en llocs específics del genoma, coneguts com orígens, on dos cadenes d'ADN són separades. El des-enrosament de la hèlix d'ADN i la síntesis de les noves cadenes formen una forqueta de replicació (veure figura 2.3). Les dues cadenes d'ADN tenen sentit oposat ( $5' \rightarrow 3'$  i  $3' \rightarrow 5'$ ; veure capítol 3). La còpia de les cadenes d'ADN és sempre en direcció de l'extrem 5' a l'extrem 3' de la cadena i per tant la síntesis de

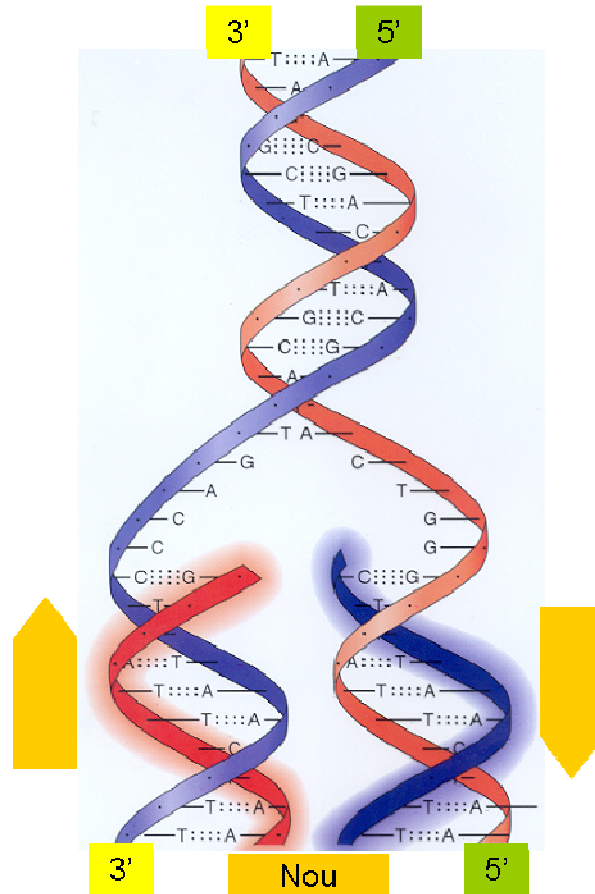
les dues cadenes es fa en dues direccions diferents (veure la figura 2.3). Un component essencial en aquest procés és l'enzima ADN polimerasa que sintetitza de forma independent a partir de cada cadena, una nova cadena complementaria afegint els nucleòtids que encaixin amb els de la cadena original. Apart de l'ADN polimerasa, altres proteïnes estan associades a la forqueta i ajuden en la iniciació i continuació de la replicació de l'ADN.



**Figura 2.2** Dogma central de la biologia molecular. L'ADN, que pot replicar-se a si mateix, guarda la informació biològica per codificar proteïnes. Aquesta informació és transcrita a través de l'ARN missatger que el transporta fins als mecanismes de traducció a proteïna.

És possible aconseguir de forma artificial la replicació d'ADN usant les mateixes enzimes que usen les cèl·lules. L'ADN polimerasa i iniciadors artificials són usats, juntament amb els quatre nucleotids tri-fosfats per iniciar la síntesis d'ADN de seqüències d'interès. La tècnica de laboratori usada és la reacció en cadena per la polimerasa (PCR de les sigles en anglès de *Polymerase Chain Reaction*).





**Figura 2.3** Replicació d'ADN. En la figura una representació de la forqueta que es forma durant la replicació, després de que la hèlix d'ADN hagi estat desenroscada i oberta. Les dues cadenes obertes es repliquen simultàniament i sempre en sentit 3' → 5' (fletxa).

### 2.1.1.2 Transcripció

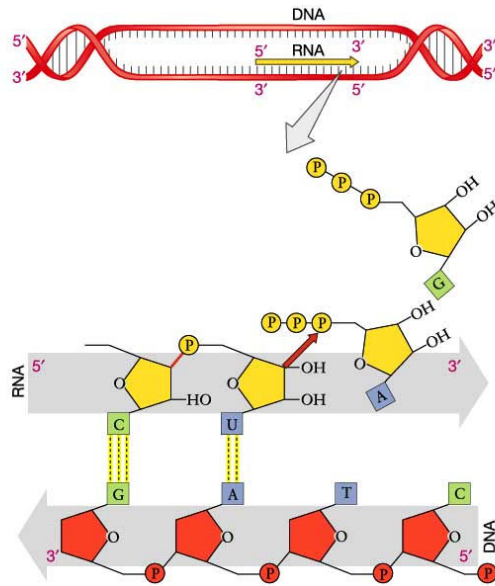
La transcripció és la síntesis de ARN a partir d'ADN. Els dos tipus de seqüències d'àcids nucleics usen el mateix llenguatge i la informació es simplement transcrita o copiada d'una molècula a l'altra. La seqüència d'ADN és copiada per l'ARN polimerasa per produir una cadena complementària d'ARN, coneguda com ARN missatger (ARNm), perquè transporti el missatge genètic des de l'ADN a la maquinària de síntesis de proteïnes de la cèl·lula. En el cas de ADN codificant de proteïnes, la transcripció és el primer pas en la expressió dels gens. Els segments de d'ADN que són transcrits en ARNm s'anomenen unitats de transcripció. Una unitat de transcripció no

està composta solament per la informació codificant de la proteïna, també inclou la seqüència que adreça i regula la síntesis de la proteïna (veure la secció Estructura dels gens en organismes eucariotes). Existeixen també en la cèl·lula mecanismes de validació de lectura en la transcripció, però són menys i no tant efectius com els controls de la replicació d'ADN; per tant la transcripció té una fidelitat de copia menor que la replicació d'ADN (Berg, Tymoczko, & Stryer, 2006).

Igual que en el cas de la replicació de l'ADN, la transcripció es processa sempre en el sentit 5' → 3'. Només una de les dues cadenes d'ADN és transcrita. La cadena d'ADN transcrita fa de model per ordenar la seqüència de nucleòtids per l'ARNm. La cadena model d'ADN es llegeix 3' → 5' per la ARN polimerasa i la nova cadena d'ARNm es sintetitzada en sentit 5' → 3' (veure la figura 2.4). L'ARN polimerasa s'enganxa a l'extrem 5' de la unitat de transcripció i es mou fins a final 3'. Excepte pel fet de que les timines en l'ADN són representades per uracils en l'ARNm, la cadena d'ARN sintetitzada contindrà la mateixa seqüència que la cadena codificant (la cadena completaria a la model) de l'ADN.

### 2.1.1.3 Traducció

Després de la transcripció de l'ADN, l'ARN missatge troba la via per arribar al ribosoma, on es tradueix. En les cèl·lules procariotes, que no tenen nucli, el procés de transcripció i traducció ocorren en el mateix espai. En cèl·lules eucariotes, el lloc de la transcripció (el nucli) està normalment separat del lloc de la traducció (el citoplasma), per tant l'ARNm ha de ser transportat fora del nucli cap al citoplasma, on finalment s'uneix als ribosomes. L'ARNm és llegit pel ribosoma com una seqüència de triplets de nucleòtids coneguts com codons. Cada codó codifica per un aminoàcid segons la taula genètica (veure taula 2.1). El codó inicial de l'ARNm i reconegut pel ribosoma és normalment AUG. En el procés de traducció intervé també un tipus especial d'ARN, l'ARN transferència (ARNt) que és portadors d'aminoàcids, els elements dels que es componen les proteïnes. Els tRNAs tenen un lloc on subjectar l'aminoàcid i un altre lloc anomenat anti-codó. L'anti-codó es un triplet d'ARN complementari a l'ARNm que correspon a l'aminoàcid que transporta segons la taula del codi genètic (veure figura 2.2).



**Figura 2.4** Transcripció d'ADN a ARN. Només una de les dues cadenes de l'ADN es transcriu (la cadena model). La transcripció es fa en sentit 3' → 5' i la cadena d'ARN resultant és idèntica a la cadena complementària de l'ADN model (exceptuant el canvi de timines per uracils) .

Un cop format el complex ribosoma-ARNm, els codons de l'ARNm encaixen amb els anti-codons de l'ARNt i d'aquesta manera s'aconsegueix codificar la proteïna a partir de la seqüència continguda en l'ARNm. A mesura que els aminoàcids s'encadenen en la seqüència del pèptid creixent, es comencen a plegar en la conformació adequada. Aquest plegament continua fins que el naixent polipèptid es alliberada del ribosoma com a proteïna madura. En alguns casos la nova cadena de polipèptids necessita un processament posterior per arribar a ser una proteïna madura. El plegament pot ser un procés complex i pot requerir de la intervenció d'altres proteïnes anomenades proteïnes de plegament (en anglès *chaperone*).

El codi genètic és un conjunt de normes per les quals el material genètic (seqüències d'ADN o d'ARN) es tradueix en proteïnes (seqüències d'aminoàcids). Específicament el codi defineix les correspondències entre triples de nucleòtids (codons) i aminoàcids. Cada codó es correspon a un sol aminoàcids, no així al contrari, ja que cada aminoàcid pot correspondre a un o més codons. Cal dir que no tots els organismes, o no tots els

orgànuls cel·lulars, codifiquen amb la mateixa taula genètica, però la gran majoria comparteixen el mateix codi genètic, conegut com estàndard o canònic (veure taula 2.1).

| <b>Codó</b> | Aminoàcid | <b>Codó</b> | Aminoàcid | <b>Codó</b> | Aminoàcid | <b>Codó</b> | Aminoàcid |
|-------------|-----------|-------------|-----------|-------------|-----------|-------------|-----------|
| <b>TTT</b>  | Phe       | <b>TCT</b>  | Ser       | <b>TAT</b>  | Tyr       | <b>TGT</b>  | Cys       |
| <b>TTC</b>  | Phe       | <b>TCC</b>  | Ser       | <b>TAC</b>  | Tyr       | <b>TGC</b>  | Cys       |
| <b>TTA</b>  | Leu       | <b>TCA</b>  | Ser       | <b>TAA</b>  | Ter       | <b>TGA</b>  | Ter       |
| <b>TTG</b>  | Leu       | <b>TCG</b>  | Ser       | <b>TAG</b>  | Ter       | <b>TGG</b>  | Trp       |
| <b>CTT</b>  | Leu       | <b>CCT</b>  | Pro       | <b>CAT</b>  | His       | <b>CGT</b>  | Arg       |
| <b>CTC</b>  | Leu       | <b>CCC</b>  | Pro       | <b>CAC</b>  | His       | <b>CGC</b>  | Arg       |
| <b>CTA</b>  | Leu       | <b>CCA</b>  | Pro       | <b>CAA</b>  | Gln       | <b>CGA</b>  | Arg       |
| <b>CTG</b>  | Leu       | <b>CCG</b>  | Pro       | <b>CAG</b>  | Gln       | <b>CGG</b>  | Arg       |
| <b>ATT</b>  | Ile       | <b>ACT</b>  | Thr       | <b>AAT</b>  | Asn       | <b>AGT</b>  | Ser       |
| <b>ATC</b>  | Ile       | <b>ACC</b>  | Thr       | <b>AAC</b>  | Asn       | <b>AGC</b>  | Ser       |
| <b>ATA</b>  | Ile       | <b>ACA</b>  | Thr       | <b>AAA</b>  | Lys       | <b>AGA</b>  | Arg       |
| <b>ATG</b>  | Met       | <b>ACG</b>  | Thr       | <b>AAG</b>  | Lys       | <b>AGG</b>  | Arg       |
| <b>GTT</b>  | Val       | <b>GCT</b>  | Ala       | <b>GAT</b>  | Asp       | <b>GGT</b>  | Gly       |
| <b>GTC</b>  | Val       | <b>GCC</b>  | Ala       | <b>GAC</b>  | Asp       | <b>GGC</b>  | Gly       |
| <b>GTA</b>  | Val       | <b>GCA</b>  | Ala       | <b>GAA</b>  | Glu       | <b>GGA</b>  | Gly       |
| <b>GTG</b>  | Val       | <b>GCG</b>  | Ala       | <b>GAG</b>  | Glu       | <b>GGG</b>  | Gly       |

**Taula 2.1.** Codi genètic estàndard. La taula mostra els 64 codons possibles i la seva correspondència amb un aminoàcid. La direcció de l'ARN és 5' → 3'

Per últim, es important saber que no tota la informació genètica està guardada en el codi genètic. Tal com s'explica en seccions següents, l'ADN dels organismes contenen també diferents elements estructurals que poden contribuir de manera rellevant al fenotip (veure secció Genotips i fenotips) però que operen usant un conjunt de regles diferent i molt mes complex.

#### 2.1.1.4 *Transferències especials de la informació biològica*

La resta de mecanismes de transferència que es coneixen (veure figura 2.1) no són pas generals entre els organismes. Dos d'elles, la transcripció reversa i la replicació d'ARN es coneix que es donen de forma natural. La tercera, la traducció directa d'ADN a proteïna no es coneix que es produeixi en la natura i és fruit d'un procés artificial induït per l'home.

#### 2.1.1.4.1 Transcripció reversa

La transcripció reversa es una transferència d'informació de l'ARN a l'ADN (el revers d'una transcripció normal). Se sap que es produeix en alguns retrovirus, com Virus de la Immunodeficiència Humana (VIH), o en eucariotes avançats, en el cas dels retrotransposons.

#### 2.1.1.4.2 Replicació de l'ARN

La replicació d'ARN és la còpia d'un ARN a un altre. Aquest és el mecanisme pel qual alguns virus d'ARN es poden replicar (F Crick, 1970)

#### 2.1.1.4.3 Traducció directa de l'ADN a proteïna

La traducció directa d'ADN a proteïna ha estat demostrada *in vitro* usant extractes cel·lulars que contenen ribosomes però no la cèl·lula intacta. Aquest fragments poden expressar proteïnes a partir d'ADN extern (McCarthy & Holland, 1965).

## **2.2 ESTRUCTURA DELS GENS EN ORGANISMES EUCARIOTES**

El desenvolupament de l'organisme pot estar determinat a través de la interacció dels productes dels gens entre ells i el seu entorn. El gen és una unitat localitzada en el genoma que està associada a regions reguladores, regions transcrites i altres seqüències funcionals (Pearson, 2006). El concepte de gen ha evolucionat en els últims anys tenint en compte els complexos patrons de regulació, la conservació gènica i l'ARN no-codificant (Gerstein et al., 2007).

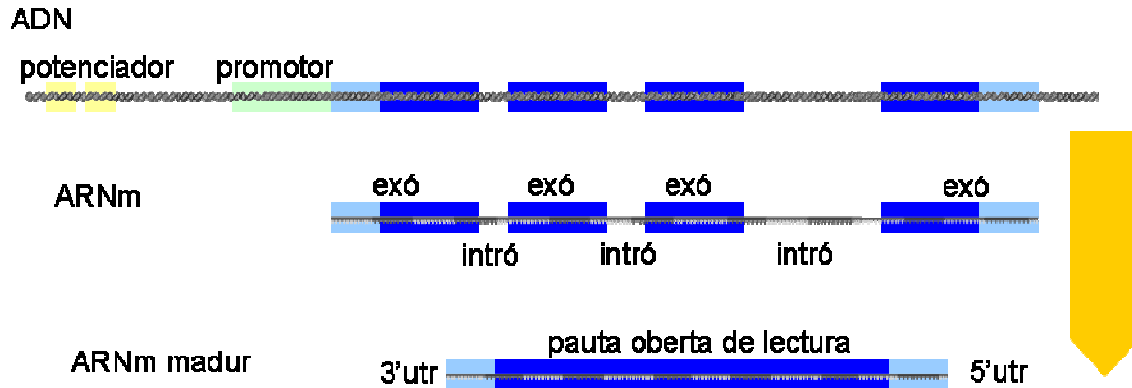
En les cèl·lules, els gens consisteixen en llargues tires d'ADN que contenen regions promotores (veure secció Promotors de gens) que controlen l'activitat del gen i les seqüències codificants i no codificants. Les seqüències codificants dels gens, conegudes com exons, determinen el producte del gen, mentre les no codificants (introns) poden regular la seva expressió. Quan un gen és actiu, les seqüències codificants i no

codificants es copien en el procés de transcripció, produint la còpia d'ARN.

En general les parts no codificants dels gens són eliminades en l'ARNm en un procés d'entroncament conegut com *splicing*. En eucariotes un gen pot codificar múltiples proteïnes creades a partir de diferents ensamblatge dels exons a través d'un mecanisme conegut com *splicing* alternatiu. En procariotes, els introns son menys comuns i els gens normalment contenen una cadena ininterrompuda d'ADN, anomenada cistró, que codifica la proteïna. Els gens procariotes s'agrupen normalment en unitats anomenades operons amb una sola regió reguladora en comú.

El conjunt de tots els gens d'un organisme és conegut com a genoma. La mida del genoma d'un organisme es normalment inferior en procariotes (tant en nombre de gens com en el nombre de nucleòtids). Tot i així, no hi ha una clara relació entre el la mida del genoma i la complexitat d'un organisme eucariota. Un dels genomes coneguts més llargs és el de l'ameba (*Amoeba dubia*), un organisme unicel·lular amb 670 bilions de parells de bases, 200 vegades superior al genoma humà (J C Venter et al., 2001). El nombre estimat de gens en l'home ha estat contínuament revisat des de que el projecte de seqüenciació va ser completat (J C Venter et al., 2001; I. H. G. S. Consortium, 2001) deixant aquest nombre actualment entre 20,000 i 25,000 (Human Genome Sequencing ConsortiumInternational, 2004). Com a referència, la planta Arabidopsis Thaliana té un nombre similar de gens (~20,000) en un genoma 270 vegades (The Genome Sequence of Drosophila melanogaster , 2000b) més petit (157 milions de parells de bases).

La cerca computacional de gens en genomes eucariotes és molt complexa, donada la presència d'ADN no codificant com els introns, les regions UTR o els pseudo-gens (gens no operatius però d'estructura gènica similar als gens operatius). En moltes espècies eucariotes, només una petita part de l'ADN es coneix que codifiqui proteïnes i la distància entre gens sol ser molt llarga i molt poc predictable. La major part de les regions no codificants aparentment no té una funcionalitat coneguda. En l'home es pensa que existeixen 12-15 gens per cada milió de parells de bases (James D. Watson et al., 2003), el que significa que trobar el seus extrems és una tasca molt costosa.



**Figura 2.5** Elements estructurals dels gens. Els gens eucariotes es fragmenten en blocs codificants (exons) i blocs no codificants (introns) que en un procés post-transcripcional són re-arranjats. Els extrems de l'ARNm madur (ja manipulat) que no codifiquen tenen funcions reguladores. Els gens tenen associat una regió promotora que també regula la expressió del producte genètic juntament amb elements potenciadors (*enhancer*) o silenciadors distants.

Tal com s'ha comentat, l'element principal del gen, el que és responsable més directe del producte funcional que se'n desprèn (la proteïna en la majoria dels casos) és l'exó, però aquest és complementat per altres elements claus dels mecanismes biològics (veure figura 2.5; comentats breument en les següents seccions).

### 2.2.1 Exons

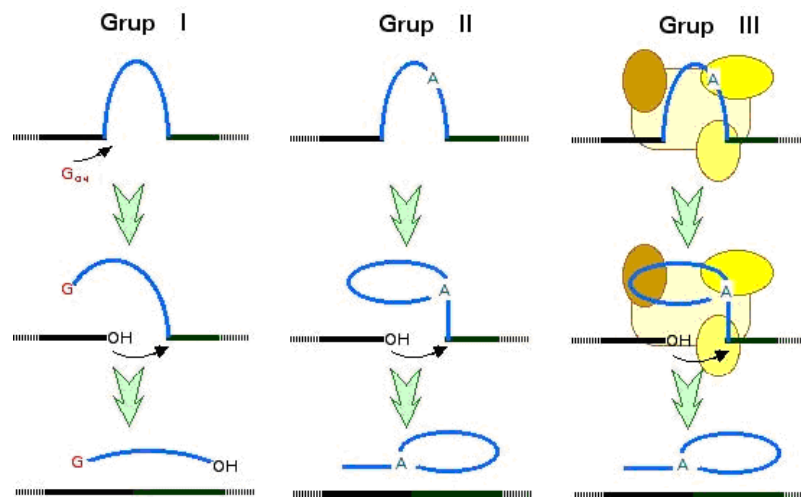
Els exons són cada una de les porcions o blocs dels gens dels organismes eucariotes que són transcrits i posteriorment es tradueixen. Els exons no són continus en el gen sinó que s'alternen amb els introns; en relació amb aquests, els exons són molt més curts.

En molts gens, cada exó conté una part del que es coneix com pauta oberta de lectura (en anglès *open reading frame*; ORF) que codifica per a una porció específica de la proteïna completa. Tot i que sovint s'entén que els exons fan referència a les regions codificants dels gens això no és del tot precís ja que es coneixen molts exons que no

codifiquen (M Q Zhang, 1998).

## 2.2.2 Introns

Els introns són cadascuna de les porcions dels gens dels organismes eucariotes que són transcrits, però no traduïdes a proteïnes que són eliminats o estroncats en l'ARN madur. Els introns poden constituir segments llargs (entre 500 i 1,500 nucleòtids cadascun). La seva funció, encara en discussió, podria ésser la de separadors dels exons, encara que no es pot descartar alguna funció reguladora en alguns cassos. La presència d'introns en els gens minva en descendir l'escala filogenètica: és freqüent en els animals superiors, rara en la mosca i el llevat, i nul·la en els bacteris.



**Figura 2.6** Els introns, per definició, són els segments separadors dels exons d'un mateix gen i que s'eliminen en la formació d'ARNm madur. Segons la manera com s'eliminen es defineixen 3 tipus d'introns: grup I, II, III.

Es distingeixen tres tipus d'introns (Bonen & J. Vogel, 2001) segons la forma en que son eliminats durant l'estruncament (veure figura 2.6). Els introns del grups I es caracteritzen per l'absència d'un motiu en comú en els llocs de connexió (tot i que si poden tenir-lo a la part interior). La eliminació dels introns del grup I és un procés



auto-catalític estretament relacionat amb la estructura secundària i terciària de la molècula precursora d'ARN. Els introns del grup II també s'elimina per un procés auto-catalític dependent d'estructura, però aquest si tenen una seqüència consens (amb la regla 5' GUGYG...AY 3'). El introns del grup III són introns més curts molt semblants al del grup II amb certa conservació del motiu consens però en aquest no formen cap estructura secundària específica i requereixen de proteïnes externes (*spliceosome*) per la seva eliminació.

### 2.2.3 Entroncament (*splicing*) alternatiu

El fenomen del *splicing* alternatiu de l'ARNm permet que una seqüència primària del gen transcrit (ARNm pre-madur, o pre-ARNm) sigui separada i re-connectada per produir productes alternatius. Les proteïnes variants que comparteixen un origen comú (gen o pre-ARNm) és coneixen pel terme "isoforma".

El *splicing* alternatiu té una gran rellevància en la genètica invalidant la hipòtesi de que un gen equival a una sola proteïna de les cèl·lules procariotes. S'ha proposat que per les cèl·lules eucariotes aquest mecanisme és un salt quantitatiu en la eficiència dels genomes, ja que la informació pot guardar-se de manera molt més econòmica. Moltes proteïnes poden ser codificades en l'ADN en una mida que en procariotes només serviria per dues proteïnes.

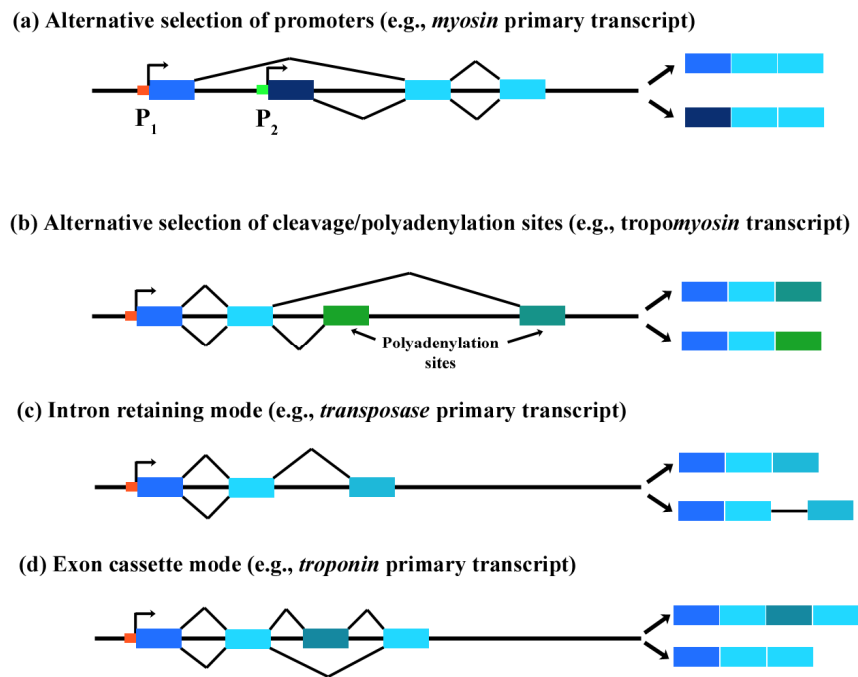
La duplicació de gens és al igual que el *splicing* alternatiu un procés per diversificar el repertori de proteïnes. Curiosament aquest dos fenòmens estan inversament relacionats, o sigui que les grans famílies de gens (on hi ha hagut molta duplicació) es caracteritzen per una absència de variants de *splicing*, i vice versa (Talavera, C. Vogel, Modesto Orozco, Teichmann, & de la Cruz, 2007). La duplicació en gens amb múltiples isoformes podria ser desfavorable donat que en aquest casos s'amplificaria l'impacte en un sistema biològic estrictament regulat (Talavera et al., 2007).

Existeixen quatre modes coneguts de *splicing* (veure figura 2.7): la selecció alternativa de promotors, la selecció alternativa d'un exó, la retenció d'un intró o la eliminació d'un exó. Una conseqüència molt important dels gens amb promotors alternatius és

que un mateix gen pot tenir unitats de regulació independents que codifiquen isoformes de la proteïna. Aquest model de gens complica encara més la cerca i l'estudi dels elements reguladors dels gen eucariota.

## 2.2.4 Regions no traduïdes

Una part de l'ARNm ja madur (sense introns) no acaba essent traduïda. Concretament aquestes parts no traduïdes (UTR, de l'anglès Untranslated Region) es localitzen als dos extrems de l'ARNm. Si es troba en el inici (en la part 5'; veure capítol Estructura de l'ADN) es coneix com la regió 5'UTR, mentre que el que es troba en la part final (3') es coneix com 3'UTR.



**Figura 2.7** Modes de *splicing* coneguts: a) selecció alternativa de promotor, b) selecció alternativa d'exó, c) retenció d'intró o d) eliminació d'exó. Font: <http://www.scfbio-iitd.res.in/tutorial/splicing.html>

#### 2.2.4.1 Regions 5'UTR

La regió 5'UTR és una secció del gen que comença en la posició +1 a partir del inici de transcripció (TSS; *transcription start site*) i acaba justament abans del primer codó (normalment AUG) de la regió codificant. Conté normalment un lloc d'unió al ribosoma (RBS; *ribosome binding site*). La mida de la regió 5'UTR es variable, en general de cent (o mes) parells de bases.

- En general els elements que podem trobar en les regions 5'UTR són seqüències que poden promoure l'inici de traducció, llocs d'unió per proteïnes, que poden afectar l'estabilitat del ARNm o la seva traducció i elements reguladors que no depenen de proteïnes

#### 2.2.4.2 Regions 3'UTR

La regió 3'UTR és el segment no traduït de l'ARNm situat en l'extrem final (3') i per tant, que comença a partir del final del marc traduït. En els gens humans aquest regió fa de mitja uns 700 nucleòtids (Lodish et al., 2003) i el seu estudi desperta actualment un interès especial per els elements reguladors que s'hi troben, especialment els llocs d'unió dels microRNAs, veure secció de Regulació epigenètica (Lim et al., 2005). Apart aquesta regió també té la senyal de poliadenilació, normalment AAUAAA, o una variant propera. Aquesta marca senyala el final de transcripció (aproximadament 30 bases mes tard). Seguidament es troba una seqüència d'adenines (cua poli-A). Finalment, la regió 3'UTR és coneguda pels llocs d'unió per proteïnes que poden afectar la estabilitat de l'ARNm. Aquestes proteïnes també poden determinar la localització de la molècula en la cèl·lula. També en relació a aquests elements trobem elements rics en AU (Alberts et al., 1999).

### 2.2.5 Promotors de gens

El promotor es la regió reguladora que es localitza en el DNA, generalment en la part superior de la regió 5'UTR del gen que normalment promou l'expressió d'aquest gen.

Segons la ARN-polimerasa que activen parlem de promotors Tipus I (aquells que

controlen l'activació de gens ribosomals), de Tipus II (els que regulen gens que codifiquen per mARN i per diversos ARN nuclears petits) i els de Tipus III (aquells que regulen els gens dels tRNA i d'alguns ARN petits). Amb diferència els promotors de Tipus II son els més comuns.

Com que els promotors estan en mols casos localitzats immediatament adjacents al gen en qüestió, les posicions del promotors han estat històricament designades de forma relativa al inici de transcripció (TSS; on el gen es comença a transcriure). Per exemple, la posició negativa -100 fa referència al nucleòtid anterior al TSS (seguint el sentit de la cadena on està anotat el gen) i a una distància de 100 parells de bases.

Tal com s'ha comentat en la secció de Estroncament (*splicing*) alternatiu, es dóna el cas de gens amb múltiples promotors alternatius. L'ús de promotors alternatius és un mecanisme versàtil per aconseguir diversitat i flexibilitat en la regulació dels gens. L'ús de promotors alternatius pot influenciar l'expressió d'un gen de diverses maneres. El nivell d'inducció a la transcripció entre els diferents promotors pot ser molt diferent, així com la eficiència de la traducció de les diferents isoformes. Els promotors alternatius també poden ser sensibles a diferents senyals externes el que obre un ventall de possibilitats per adaptar l'expressió gènica a les necessitats de l'organisme (Ayoubi & Van De Ven, 1996)

El promotor es compon principalment per un nucli, situat aproximadament entre la posició -34 i el TSS (Smale & Kadonaga, 2003). Al nucli s'hi uneix alguna de les classes d'ARN polimerasa encarregada de transcriure l'ADN. També és el lloc on s'uneixen alguns dels factors de transcripció més comuns (com la proteïna d'unió a la caixa TATA; veure la secció Factors de transcripció). El promotor pròxim, que conté els elements primaris de regulació del gen, ocupa la regió entre -250 (nucleòtids abans del TSS) i el TSS. S'hi uneixen factors de transcripció més específics (Smale & Kadonaga, 2003). Finalment el promotor distal, es una seqüència sense una mida determinada, sempre anterior al TSS. Conté llocs d'unió a elements reguladors normalment amb una influència menor que els del promotor pròxim sobre la regulació de l'expressió.

### 2.2.5.1 *Promotors en procariotes*

En procariotes, el promotor consisteix en dues seqüències curtes situades en les posicions -10 i -35. L'element més proper al TSS (-10) és essencial per la transcripció del gen i té un motiu de sis nucleòtids amb seqüència consens TATAAT (TATA-box). El segon motiu (TTGACA) situat uns nucleòtids més llunyà (-35) millora molt l'eficiència de transcripció, però no es essencial (Jacques, Rodrigue, Gaudreau, Goulet, & Brzezinski, 2006).

### 2.2.5.2 *Promotors en eucariotes*

La regulació de la expressió de gens eucariotes és molt complexa i està lluny de ser completament caracteritzada. La majoria dels gens son transcrits per la ARN polimerasa de tipus II una gran complex proteic amb més de 10 subunitats que precisa a més de proteïnes addicionals com els factors de transcripcions generals (GTF: TFIIA, TFIIB, IFIID, TFIIF and TFIIH) per crear un gran cluster proteic unit al DNA (el complex de pre-inicialització). La regió 200-300 bp per sobre del TSS, a on s'uneix el complex es denomina “core proximal del promotor” i es un lloc ple de dianes per factors de transcripció. Senyals addicionals poden estar ubicades a regions molt distants en seqüència (probablement properes en estructura) denominades “enhancers”.

Al contrari de la situació en procariotes, no son clares les senyals que identifiquen el core proximal d'un promotor. La TATA-box (Smale & Kadonaga, 2003) que reconeix el factor de transcripció proximal TFIID amb la seqüència consens TATAAAA es troba a la regió -25 -30 d'alguns gens, però no en gens importants com els de manteniment (“house-keeping”), oncogens, o els factors de creixement. Altres gens presenten l'element d'inicialització (Inr) que pot treballar independent o coordinat amb la TATA-box. Els requeriments de seqüència de l'Inr son molt febles, amb una seqüència consens PyA(-1)NT/APyPy. No està clara la prevalència dels elements Inr en mamífers, encara que es comú a mosca. L'element Inr està sovint associat a un element promotor localitzat després del TSS (el downstream promoter element; DPE) que te una seqüència consens A/GGA/TC/TG/A/C a les posicions +28...+32 i que es creu que interacciona també amb el TFIID. Una última, però molt difusa senyal associada a promotors eucariotes es la presència de illes CpG, que son detectades en al voltant

del 50% dels gens humans, incloent-hi la majoria dels de manteniment. No existeix una seqüència consens per les illes CpG, però són detectades com fragments llargs (0.5-2 kbp) amb contingut anormalment llarg de C+G i una gran població de passos CpG. Es creu que els promotors associats a les illes CpG tenen TSS múltiples i que no tenen senyals addicionals com TAT-box, Inr o DPE.

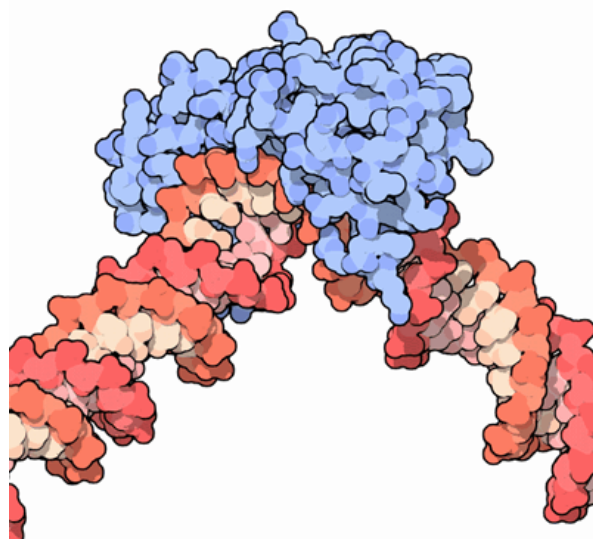
## 2.2.6 Factors de transcripció

Els factors de transcripció (TF) són proteïnes que s'uneixen a una seqüència específica d'ADN ajudant al control de la seva transcripció (Latchman, 1997). Els factors reconeixen i s'enganxen a un segment d'ADN d'acord amb un patró propi (seqüència o estructura de la molècula) de reconeixement. Quan això passa, la proteïna pot arribar a deformar notablement la hèlix d'ADN (veure figura 2.7). El lloc on la proteïna s'uneix a l'ADN es coneix com lloc d'unió al factor de transcripció (TFBS; *transcription factor binding site*).

Els factors de transcripció poden desenvolupar la seva funció de forma individual o més comunament format un complex superior amb altres proteïnes. Això és possible gràcies a una estructura que combina un domini d'unió al DNA que reconeix aquest amb interaccions normalment via solc ample, i un domini de trans-activació que és el responsable de transmetre la senyal del reconeixement a altres proteïnes. Resultat de la unió els factors de transcripció poden incrementar (activadors) o disminuir (repressor) la presència d'ARN polimerasa, l'enzim que activa la transcripció dels gens (Nikolov & Burley, 1997).

La funció biològica relacionada amb els factors de transcripció és la de la regulació bàsica de la transcripció. En eucariotes existeix una classe de transcriptors que són indispensables per que la transcripció sigui efectiva. Molts d'aquests factors no s'enganxen a l'ADN, però són part d'un complex que interactua directament amb l'ARN polimerasa. Entre aquests factors hi figura el que s'uneix a la caixa TATA, tal com es veu en la figura 2.8. Els TF també juguen un paper molt important la resposta a senyals intracel·lulars i del medi, el control de cicle cel·lular i el desenvolupament.

Les cèl·lules es comuniquen entre si alliberant molècules que produeixen una senyal en cascada dins una altra cèl·lula. Al final d'aquest camí d'esdeveniments estan els factors de transcripció encarregats de potenciar o reprimir un conjunt de gens determinats. És el cas de la senyalització dels estrògens, dels que se'n coneix el paper crític en el desenvolupament, maduració i activació del sistema reproductiu femení. De vegades els estímuls de la cèl·lula són ambientals i no induïts per una cèl·lula veïna. És el cas dels factors d'impacte de calor. Aquests factors potencien l'expressió de gens necessaris per la supervivència de la cèl·lula en altes temperatures.



**Figura 2.8** Factor de transcripció unit a ADN. En aquest cas s'aprecia com la el factor de transcripció de la capsa TATA (en blau) deforma notablement la hèlix d'ADN pel lloc on s'ha unit (TFBS). Font: [www.pdb.org](http://www.pdb.org)

Molts factors de transcripció, especialment els onco-gens (gens relacionats amb el càncer) o els supressors tumorals, ajuden a regular el cicle cel·lular controlant quant pot créixer una cèl·lula, quan es pot reproduir i quan ha de morir de manera programada (apoptosis). Altres TF tenen la funció d'apagar o encendre els gens apropiats provocant canvis morfològics en la cèl·lula i forçant la seva diferenciació. Un cas molt estudiat d'aquesta classe de factors són els de la família Hox, proteïnes altament conservades entre grans distàncies evolutives. Els gens Hox regulen el gens realitzadors (gens que

produïxen proteïnes que fan teixits, òrgans i estructures com els ulls o les ales) unint-se a un motiu de seqüència consens TAAT en la regió promotora.

### **2.2.7 Potenciadors (*enhancers*) i silenciadors de la regulació**

Els *enhancers* són regions curtes d'ADN que poden ser el lloc d'unió a proteïnes (essencialment semblants als factors de transcripció). Així com els factors de transcripció actuen de manera local (*cis-acting*), unint-se a ADN proper al inici de transcripció (o en introns) per regular localment el gen, en el cas dels *enhancers*, el lloc d'unió que no necessiten estar proper al promotor (*trans-acting*) per incrementar (o reprimir) bruscament l'activitat de transcripció de la del gen. Les estructures superiors de l'ADN (veure capítol Estructura de l'ADN) fan que aquest es plegui de manera que els *enhancers* estiguin geomètricament propers al promotor i al gen. Això els permet interactuar amb els factors de transcripció. Cal destacar que no tots els enhancer actuen com activadors de transcripció, un bon nombre d'ells son de fet silenciadors, evitant amb la seva presencia la unió productiva dels factors de transcripció.

## **2.3 GENOTIPS I FENOTIPS**

El fenotip és una característica observable en un organisme, com per exemple la seva morfologia, el seu desenvolupament o les seves propietats bioquímiques. Els fenotips són influenciats per la combinació del la genètica de l'organisme i els factors ambientals. Això contrasta amb el genotip de l'organisme que ve definit per la seva herència genètica, que pot ser que causi un efecte en l'organisme. Així doncs la relació entre genotip i fenotip no sempre es directa. Alguns gens només estan expressats en algun fenotip donades certes condicions ambientals i per altra banda, alguns fenotips poden ser el resultat de múltiples genotips (Churchill, 1974).

La majoria d'organismes més evolucionats són diploides, que significa que tenen dos còpies de cada cromosoma (Alberts et al., 1999). Els organismes diploides tenen per tant dos còpies homòlogues de cada gen i si aquestes són idèntiques (tenen el mateix al·lel) el genotip de l'organisme es homozigot per aquest gen. Si existeix alguna



diferència entre les dues còpies, l'organisme té dos al·lells per aquest gen i es coneix com heterozigot. Tot i que, existeixen nombroses excepcions, els fenotips associats a cada al·lel poden ser dividits entre dominants o recessius. Un fenotip dominant és expressat quan només fa falta que una de les còpies del gen tingui l'al·lel associat al fenotip. En canvi un fenotip recessiu necessita que les dues còpies dels gen siguin del mateix al·lel associat.

El polimorfisme d'un sol nucleòtid (SNPs) és una variació de la seqüència d'ADN en un sol nucleòtid entre dos genomes de la mateixa espècie (Alberts et al., 1999) (o entre còpies dels cromosomes d'un mateix organisme diploide). Els SNPs poden estar en regions codificants, en regions no codificants de gens o en regions inter-gèniques. Una mutació en la regió codificant d'un gen no té com a conseqüència sempre una modificació de la composició de la proteïna produïda i moltes mutacions són doncs silencioses. En cas contrari, si les dues variants del SNP generen diferències en la proteïna es refereixen com SNPs *no-sinònims* i poden tenir conseqüències fenotípiques, fins i tot (Ferrer-Costa et al., 2005; Ferrer-Costa, Modesto Orozco, & de la Cruz, 2002, 2005).

Encara que els SNPs no estiguin en regions codificants la diferència entre les dues variants pot tenir conseqüències en altres mecanismes moleculars (i un gran impacte en el fenotip) com el *splicing* del gen, la modificació de llocs d'unió de factors de transcripció o en l'activitat d'ARN no codificant i resultar en diferències fenotípiques importants, algunes d'elles patològiques o fins i tot letals.

## **2.4 REGULACIÓ EPIGENÈTICA**

La epigenètica fa referència als canvis en la expressió dels gens, que es mantenen després de la divisió cel·lular i fins i tot entre generacions, però que no involucren cap canvi en la seqüència de l'ADN de l'organisme. El terme fa referència a la idea de que factors ambientals poden causar una diferència en com els gens s'expressen sense que la composició genètica hagi sofert cap canvi (Jaenisch & Bird, 2003). Un cas molt il·lustratiu de la regulació epigenètica és el procés de diferenciació de les cèl·lules en organismes superiors. En humans totes les cèl·lules comparteixen el material genètic,

però les diferències que hi ha entre les que pertanyen a teixits diferents són mantingudes de forma estable. Existeixen diferents tipus de mecanismes de regulació epigenètica que tot i ser considerats dinàmics són capaços de ser heretats entre generacions cel·lulars.

#### **2.4.1 Metilació de l'ADN i la remodelació de la cromatina**

Existeixen diverses capes de regulació en la expressió genètica i una d'elles és la remodelació de la cromatina, el complex que formen l'ADN i les proteïnes anomenades histones (veure capítol 3). La remodelació de la cromatina pot ser iniciada per una modificació post-transcripcional dels aminoàcids que canvien les histones o per afegir grups metil a l'ADN, en dinucleòtids CG (CpG), per convertir la citosina a 5-metil-citosina.

En les histones, la part terminal (coneguda com la cua la histona) que està poc estructurada és on es solen acumular els canvis més representatius que estan altament associades a competències transcripcionals. Cal emfatitzar que diferents modificacions en histones poden funcionar de maneres diferents com per exemple la acetilació, que pot jugar papers diferents segons en la posició on es produeixi. També es cert que múltiples modificacions poden donar-se, treballant de forma conjunta. Es coneix com el codi de la histona la idea de que les diverses modificacions dinàmiques poden regular la transcripció d'un gen de forma sistemàtica i reproduïble. Així sabem que la fosforilació de la histona H2A (veure capítol 3) està relacionada amb el cicle cel·lular i la dinàmica dels cromosomes (Ito, 2007). També que per exemple la fosforilació de la H3 que s'ha observat en l'activació de la transcripció i la condensació del cromosoma durant la mitosis i la meiosis o la fosforilació de la H4 juga un paper clau en la resposta a trencament de l'ADN de doble cadena, la progressió del cicle cel·lular i la expressió dels gens (Ito, 2007).

La metilació de l'ADN es dona de forma freqüent en organismes superiors a la seqüència repetida de dinucleòtids CG (coneguts com CpG per distingir del parell de bases C:G; la "p" fa referència al pont de fosfat que uneix la citosina amb la guanina) probablement com un mecanisme per evitar la desaminació química de la citosina

que produiria la transducció C→U i preservar així la integritat de l'ADN. La excepció a aquesta regla general són les illes CpG, que estan sense metilar. Una metilació d'aquestes illes produeix la inactivació del gen associat el que ha estat descrit com un mecanisme de control addicional per alguns gens. La pauta de metilació dels gens humans sembla ben conservada i alteracions en la mateixa (per exemple metilació de promotors associats a gens d'apoptosis) ha estat relacionada amb malalties com el càncer. La metilació/desmetilació de l'ADN sembla un procés molt ben regulat a nivell cel·lular, amb enzims específics encarregats de controlar-la, però en qualsevol cas no queda clara en últim termini la raó física que fa que una seqüència metilada no reconegui de manera productiva als factors activadors de la transcripció.

Cal senyalar, que la metilació de l'ADN o remodelació de la cromatina per modificacions en les histones no es donen en tots els organismes, ni tant sols en els eucariotes, i es propi d'espècies superiors (normalment en plantes i animals). La capacitat de la cèl·lula per metilar una regió promotora és un important element de regulació que pot activar o desactivar un grup de gens.

#### **2.4.2 Silenciament de gens basat en àcids nucleics**

El descobriment de mecanismes de silenciament de gens per mitjà d'àcids nucleics es remunta a finals dels 70 (Zamecnik & Stephenson, 1978) amb el descobriment de que una seqüència de 13 parells de bases, complementària a un segment del virus del sarcoma de Rous inhibia la seva replicació. Avui en dia es coneixen altres processos que es difereixen d'aquest però que es basen en el reconeixement d'una seqüència de nucleòtids d'un gen (ADN o ARN) per part d'una altra seqüència complementària o antisentit, que indueix al silenciament del gen.

Aquest descobriment ha obert la porta a l'ús dels oligonucleòtids com a molècules terapèutiques, i al mateix temps, l'exploració dels motius de seqüència en ADN i ARN com a dianes mèdiques. La idea es crear un cadena d'àcids nucleics capaç de reconèixer de forma única mitjançant l'aparellament de bases el gen específic que es vol reprimir, sense que afecti l'expressió de la resta de gens. La modificació de la molècula d'àcids nucleics pot incrementar la efectivitat de la reacció i facilitar que no

sigui destruïda per mecanismes de control de la cèl·lula. Algunes de les estratègies antisentit poden reprimir la traducció del gen, mentre que altres actuen sobre l'ARNm (post-transcripcionals)

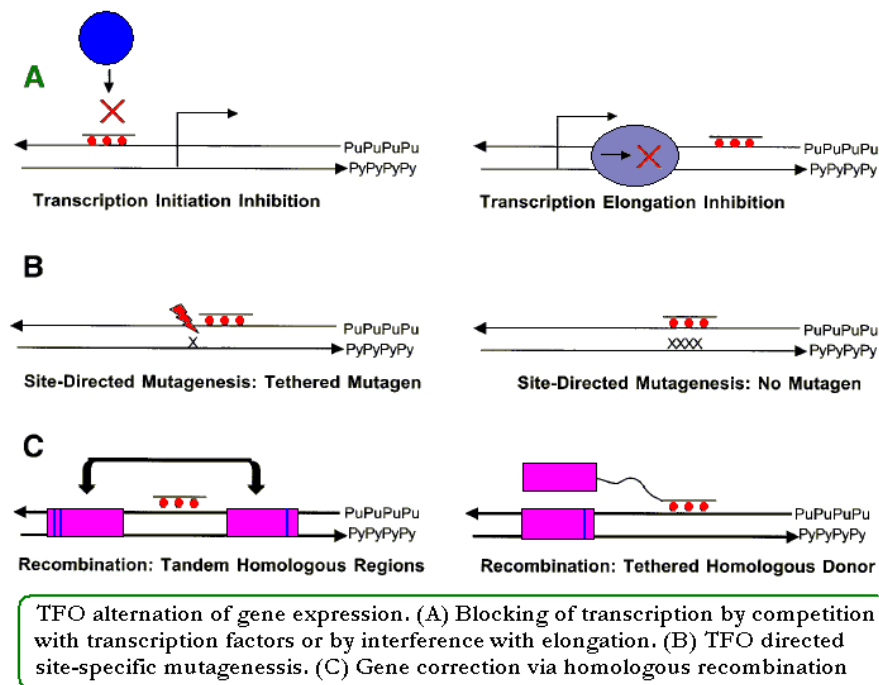
#### 2.4.2.1 Regulació transcripcional

La interrupció de la expressió d'un gen a nivell de la seva transcripció es refereix com a mecanisme aniti-gen (Vasquez & J. H. Wilson, 1998). Per aquest mètode l'oligonucleòtid antisentit reconeix una doble cadena d'ADN (o d'ARN o d'un híbrid entre els dos) formant una triple hèlix o tríplex (veure capítol Estructura de l'ADN). Es per això, que la molècula antisentit es coneix en aquest cas com a seqüència formadora de tríplex (TFO; en anglès *triplex forming oligonucleotide*) i la doble cadena complementaria com a seqüència diana de tríplex (TTS; en anglès *triplex target sequence*). La inhibició de la transcripció d'un gen per la formació d'un tríplex es pot donar de dues maneres diferents. Si el tríplex es forma en la regió reguladora del gen (veure figura 2.9), aquest pot competir amb els factors de transcripció envaint el seu lloc d'unió (Knauert & Glazer, 2001). Una segona possibilitat es que el tríplex es formi en la regió transcrita del gen. La formació d'aquesta molècula inhibiria la transcripció per bloqueig estèric (Knauert & Glazer, 2001).

Tot i que, la formació de tríplex es produeix de forma natural en les cèl·lules, no es coneix actualment que existeixi un mecanisme generalitzat de regulació genètica associat a aquesta molècula. No obstant si que es possible fer servir tríplexs per modular externament la funció de gens i hipotèticament per dissenyar nous fàrmacs basats en TFOs absolutament específics. L'inconvenient que tenen venen per: la seva reduïda estabilitat en condicions fisiològiques (veure el capítol Estructura de l'ADN), les restriccions en els motius de seqüència capaços de formar aquesta estructura (Gowers & Fox, 1999), la susceptibilitat a degradació per nucleases del TFO, i la dificultat per introduir-lo en la regió nuclear de la cèl·lula. Tot això fa que avui en dia encara no existeixin drogues basades en aquest via.

### 2.4.2.2 Regulació post-transcripcional

El silenciament dels gens en la fase post-transcripcional (entre la transcripció del gen i la traducció de l'ARNm) és un procés que es creu que s'origina com mecanisme de defensa contra transposons, virus o altres tipus de ARN/ADN invasiu (Fagard, Boutet, Morel, Bellini, & Vaucheret, 2000). Existeixen tres tipus de regulació post-transcripcional basada en àcids nucleics (veure la figura 2.10): la antisentit, la basada en ribozims i l'ARN interferència.

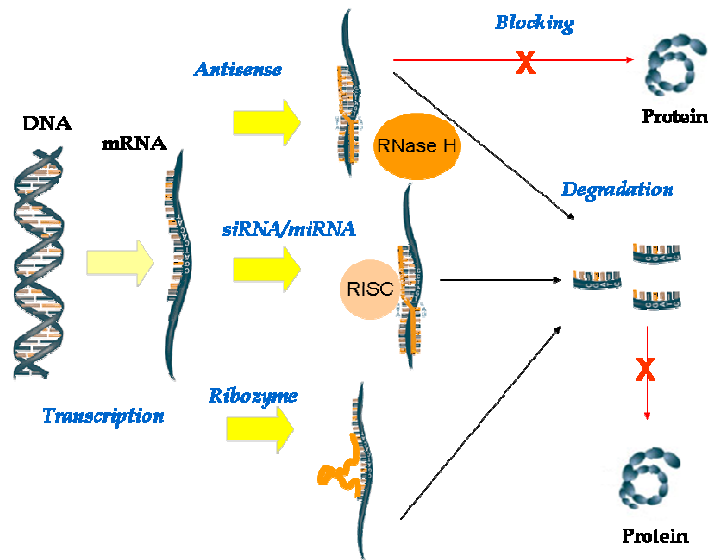


**Figura 2.9** Entre els efectes biològics que s'associen a la formació de tríplex destaca (A) el bloqueig de la transcripció. La formació de tríplex també pot ser usada per induir mutacions (B) en el genoma i per potenciar la recombinació (C) de regions (Knauert & Glazer, 2001).

#### 2.4.2.2.1 Antisentit

- Les molècules antisentit a nivell pots-transcripcional (els TFOs son antisentit a nivell transcripcional) reconeixen i interactuen amb l'ARNm amb una seqüència complementària a la seva, modificant la expressió del gen.

Existeixen diverses formes d'antisentit que es pot subdividir en dos grups (Prescott, Harley, & Klein, 2004) segons siguin o no dependents d'enzim. Els primers depenen de l'activitat degradadora d'ARN de la RNasa H. La RNasa H és una enzima que degrada la cadena d'ARN de hèlix híbrides entre ARN i ADN. Els no dependents d'enzim es basen en el bloqueig estèric del nucleòtid antisentit que no permet la traducció de l'ARNm.



**Figura 2.10** Estratègies de silenciament post-transcripcional. Els oligonucleòtids antisentit anul·len la transcripció de l'ARNm per bloqueig estèric o per la degradació induïda per RNasa H (part superior). El mecanisme de interferència d'ARN a partir de molècules siARN o miARN que estan enllaçades al complex RISC i indueixen a la degradació (part central). Els ribozims netegen de forma catalítica l'ARNm diana (part inferior).

#### 2.4.2.2.2 Ribozims

Un ribozim és una cadena d'ARN que té funció biocatalitzadora amb capacitat d'unir-se a un ARN per reconeixement de seqüència i degradar-lo de manera catalítica. Entre els Ribozims coneguts, el *Hammerhead* pot induir a la degradació dependent de RNasa d'ARN de doble cadena (dsRNA). Aquest tipus de ribozim es va descobrir a

mitjanats dels 80 en viroides de plantes (Prody, Bakos, Buzayan, Schneider, & Bruening, 1986). L'ús de ribozims modificats ha estat també fortament investigat com possible font de nous medicaments, però encara no ha estat desenvolupat cap fàrmac basat en ells.

#### 2.4.2.2.3 Interferència d'ARN

La interferència d'ARN és un mecanisme que inhibeix la traducció del gen després de la seva transcripció o simplement interfereix en el procés de transcripció. Des de fa pocs anys sabem que la interferència d'ARN té un paper clau en la regulació dels processos cel·lulars. En aquest mecanisme es coneix com ARN interferent (ARNi) com la molècula d'ARN que reconeix específicament l'ARNm del gen a regular. Els ARNi són molècules petites (20-25 nucleòtids) que es generen per fragmentació de precursors més llargs.

Existeixen diversos tipus de ARNi, un d'ells és l'ARNs interferent petit (siARN; de l'anglès *small interference RNA*) que es compon per 2 cadenes d'ARN (20-21 nucleòtids) perfectament complementaris formant una doble cadena amb els extrems tallats. Els siARNs són el producte processat pel Dicer, una enzima que talla segments (siARN) llargs d'ARN de doble cadena (Bernstein, Caudy, S. M. Hammond, & Hannon, 2001). Una de les dues cadenes del siARN (la anti-sentit) s'acobla a un complex superior denominat RISC (de l'anglès *RNA-induced silencing complex*), que utilitza la cadena per guiar i identificar ARNm amb la mateixa seqüència. El complex RISC talla l'ARNm en dos meitats perquè posterior sigui degradat. Els siARN poden ser induïts de forma externa per mètodes de transfecció per reduir l'expressió d'un gen d'interès.

Els micro-ARNs (miRNAs) són petits ARN interferents que es generen a partir de precursors específics codificats en el mateix genoma de la cèl·lula. Els miRNAs un cop transcrits es pleguen i formen una doble hèlix en forma de forqueta (hairpin) intramolecular que conte elements de complementarietat casi perfecta (varia entre plantes i animals). De manera semblant als siARN, els miARNs també s'incorporen al RISC perquè aquest identifiqui ARNms propis de la cèl·lula. Depenent del grau de

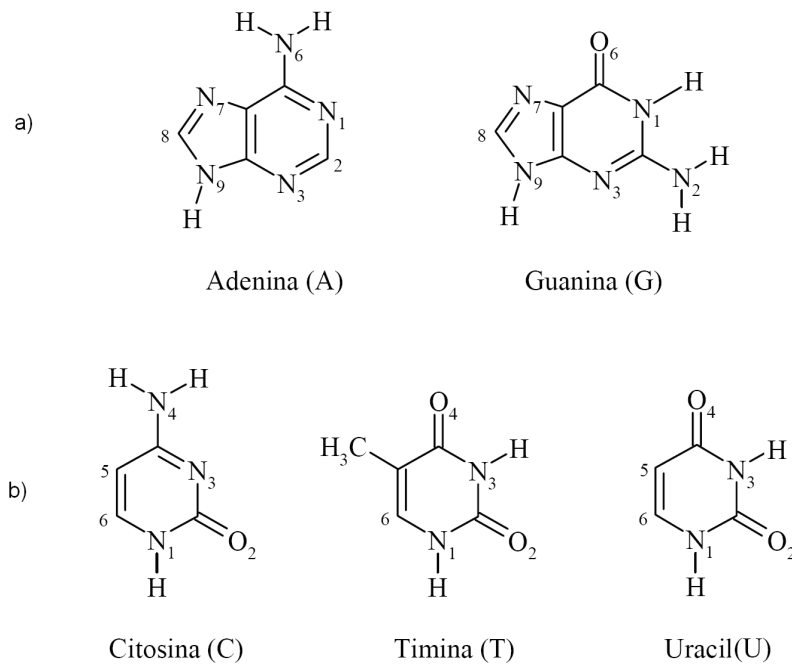
complementarietat entre el miARN i el ARNm es produirà una degradació del ARNm o solament la inhibició de la traducció (Birmingham et al., 2006).





### 3 INTRODUCCIÓ III: ESTRUCTURA DE L'ADN

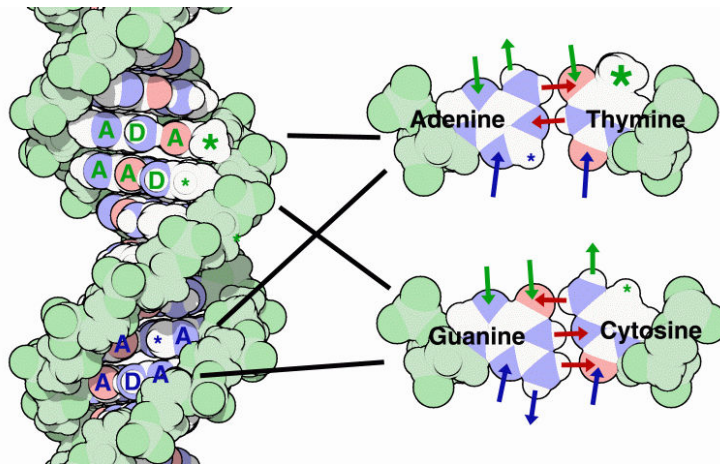
Durant els primers anys del segle XX els biòlegs van reconèixer que els gens estaven localitzats en els cromosomes, coneguts aleshores com estructures lineals presents en el nucli de les cèl·lules eucariotes (Alberts et al., 1999). Es pensava que l'ADN era la una molècula amb poca complexitat química (tot al contrari que les proteïnes). Per tant s'acceptava amb dificultat el paper principal de l'ADN com a material genètic. Només a partir dels treballs estructurals de Watson i Crick (JD Watson & FH Crick, 1953) es reconeix que l'ADN és una entitat estructural amb la capacitat de mantenir i expressar la herència genètica de la cèl·lula.



**Figura 3.1** Estructura química de les bases nitrogenades (a) purines i (b) pirimidines. En la seva nomenclatura es numeren els àtoms pesats corresponents, mentre els hidrògens adopten el mateix número que l'àtom al que està unit (Saenger, 1984)

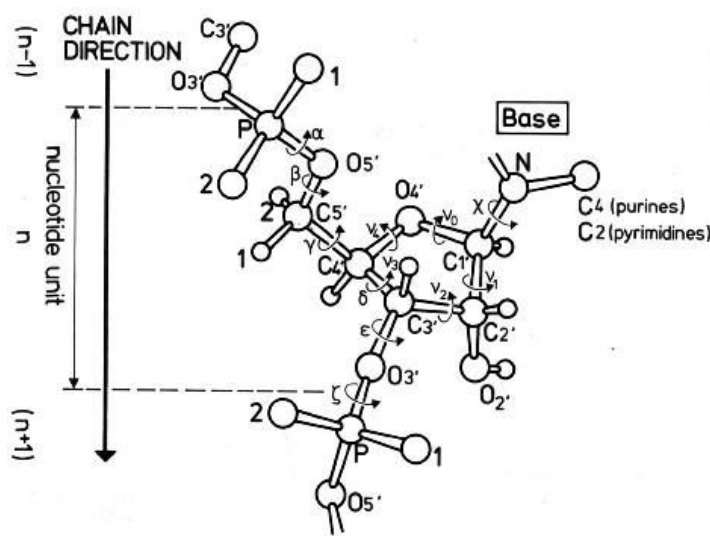
### 3.1 ESTRUCTURA PRIMÀRIA DE L'ADN

Els nucleòtids són la unitat bàsica de la que es compon la estructura molecular dels àcids nucleics. Estan compostats per una base nitrogenada, un sucre (ribosa en l'ARN i 2'deoxiribosa en l'ADN) i un fosfat. Existeixen dos tipus fonamentals de base nitrogenades: les púriques (en purines) i les pirimidíniques (en pirimidines). D'elles, cinc apareixen com a bases codificants portadores d'informació genètica: adenina (A), guanina (G), citosina (C), timina (T) i uracil (U), trobant-se aquesta última present solament en l'ARN (veure figura 3.1). Les bases estan unides al sucre mitjançant l'enllaç glicosídic i el sucre al fosfat mitjançant enllaços ester a les posicions 3' o 5'. Els nucleòtids generen polímers (els àcids nucleics) mitjançant la formació d'enllaços fosfodiester 5'→3' formant així una cadena que pot ser molt llarga (veure figura 3.2). Per convenció (Alberts et al., 1999) identificarem l'inici de la cadena com l'extrem 5' i el final com l'extrem 3'



**Figura 3.2** Hèlix de doble cadena d'ADN. L'esquelet de sucre i fosfat està representat en color verd i les bases amb el color estàndard per àtom. Es pot observar tant d'interacció d'apilament (esquerra), com la de pont d'hidrogen (dreta). Es pot observar també a la figura de la esquerra la presència dels solcs ample (o major) i estret (o menor) a la part superior i inferior de l'ADN respectivament.

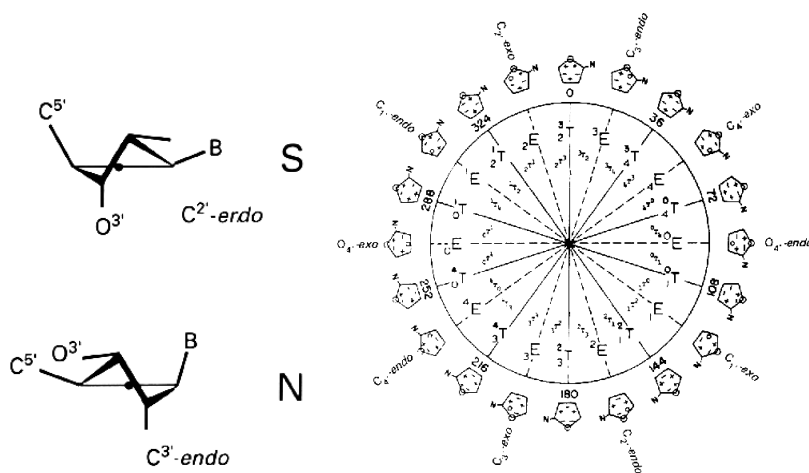
En els àcids nuclèics el paper de cada un dels seus components està perfectament marcat: i) les nucleobases doten a l'àcid nucleic de especificitat gràcies sobre tot a la capacitat de donar i rebre ponts d'hidrògen, ii) el fosfat defineix les propietats generals de poli-anió de l'ADN i iii) el sucre és el principal responsable de la flexibilitat conformacional del polímer, ja que els enllaços glicosídics, el exocíclic C4'-C5' i el moviment concertat de l'anell furanòsid (moviment pseudo-rotacional del sucre) són els punts més flexibles dels àcids nuclèics.



**Figura 3.3** Representació dels diferents graus de flexibilitat d'un element d'una cadena d'ARN.

La majoria dels àcids nuclèics presenten nucleòtids amb conformació anti respecte a l'enllaç glicosídics (i.e. el pla de la base nitrogenada està situat per fora del pla de l'anell furanòsid). La preferència conformacional respecte a l'anell furanòsid és més complexa, ja que el moviment respecte als dihedres interiors es concertat (veure figura 3.3) generant un complex moviment conformacional denominat "pseudo-rotació", representat per dos paràmetres: l'amplitud de *puckering* que informa del grau de pèrdua de planaritat de la anell i l'angle de fase que defineix com és aquesta pèrdua de planaritat (per exemple qui o quins àtoms estan fora del pla furanòsid). La conformació

d'anell (en anglès *puckering*) més important a l'ADN fisiològic es la forma C2'endo o Sur (S), mentre que a l'ARN troben la conformació C3'endo o Nord (N). El canvi entre els conformers N i S es relativament simple a nucleòtids aïllats, produint-se per la regió Est, mentre que és molt complicada en àcids nucleics, pel dramàtic canvi conformacional que produiria (veure figura 3.4).



**Figura 3.4.** Esquerra: diagrama representatiu d'una furanosa en conformació C2'endo (S) i C3'endo (N). Dreta: cercle pseudo-rotacional, la distància radial fa referència a l'amplitud de *puckering*, mentre que l'angular es refereix a l'angle de fase.

### 3.2 ESTRUCTURA SECUNDÀRIA DE L'ADN

Els nucleòtids s'agrupen formen estructures complexes la majoria de les quals es caracteritzen per tenir una naturalesa helicoidal intrínseca (veure figura 3.2). La forma helicoidal es genera, entre altres causes, en un intent de la molècula per minimitzar la forta repulsió de carga que s'origina entre els seus fosfats. Els àcids nucleics formen estructures d'una o dues cadenes. Tot i que no és tant habitual (veure secció Formes no canòniques de l'ADN) es donen casos on el nombre de cadenes és superior (tres i quatre).

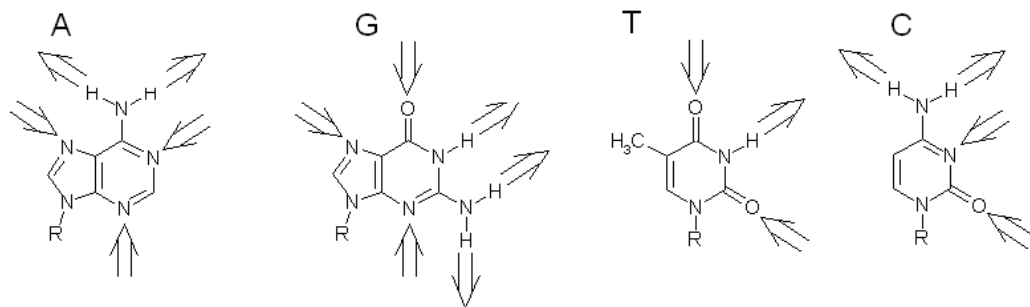
Les hèlices d'ADN han de tenir una conformació amb una relació de forces que

garanteixi la integritat estructural de la molècula, però a la vegada, aquest balanç ha de permetre flexibilitat a la molècula com per que aquesta pugui fer la seva funció biològica. Les interaccions principals que dominen la estructura de l'ADN són (veure figura 3.2):

- Les interaccions per pont d'hidrogen
- Les interaccions d'apilament o *stacking*
- L'efecte del solvent (aigua i ions) reduint les repulsions entre fosfats i facilitat l'apilament de les parts apolars de les bases.

### 3.2.1 Ponts d'hidrogen

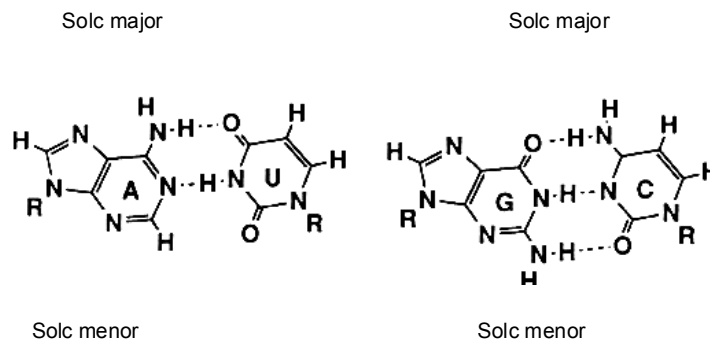
Com ja van postular Watson & Crick les bases es reconeixen específicament mitjançant els ponts d'hidrogen. Cada base nitrogenada disposa d'un perfil únic d'aquests ponts que funcionen a mode de petjada dactilar davant de les altres molècules. A nivell químic els ponts d'hidrogen són un tipus d'interaccions amb una marcada naturalesa electrostàtica entre un "acceptador" (un heteroàtom unit a un hidrogen) i un "donador" (un heteroàtom que orienta cap al hidrogen del donador un parell electrònic lliure); veure Figura 3.5.



**Figura 3.5** Patró de donadors (fletxes apuntant a l'exterior) i receptors (fletxes cap a l'exterior) de ponts d'hidrogen en les 4 bases nitrogenades de l'ADN.

Els ponts d'hidrogen mantenen la estructura interna de la hèlix fent que els nucleòtids d'una cadena reconeixin als de l'altre(s). A més són també els responsables de la síntesis complementaria de dúplexs i, per tant, són les interaccions químiques determinants pel manteniment del codi genètic.

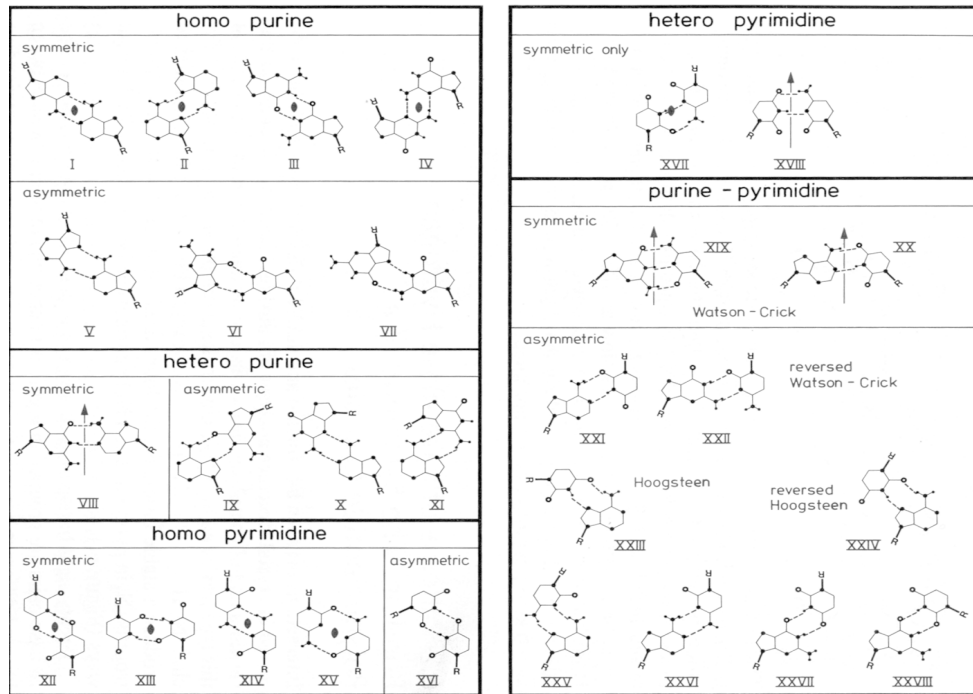
El reconeixement entre dues bases no és tan únic com inicialment es va pensar. De fet existeix una gran diversitat entre patrons d'unió, 28 d'ells amb una eficiència alta (compartint dos o més ponts d'hidrogen). La forma fisiològicament més comú de reconeixement en l'ADN és el proposat per Watson i Crick, (WC) on tal com es mostra en la figura 3.6, la guanina s'aparella amb la citosina (mitjançant tres ponts d'hidrogen) i la adenina amb la timina (mitjançant dos enllaços) (JD Watson & FH Crick, 1953). Altres modes d'aparellament són la forma WC reversa els *Hoogsteen* (en forma normal i en forma reversa) i els *wobble* (veure figura 3.7). En aquest tipus de reconeixement alternatius es dona una pèrdua de ponts d'hidrogen, però en aquest casos poden afavorir-se les interaccions d'apilament (Blackburn & Gait, 1996).



**Figura 3.6.** Esquema dels les parelles canòniques (Watson-Crick) A·U i G·C i de la definició dels solcs associats.

Cal destacar que la formació de ponts d'hidrogen canònics tipus Wason-Crick no impedeix que les bases mantinguin una pauta de lectura a traves de les regions al voltant dels N7 i N3 (Pur), veure Figura 3.6. Aquesta possibilitat és aprofitada per altres nucleòtids per formar estructures tricatenaries (descrites més endavant) i també per

proteïnes i petits lligands per reconèixer específicament l'ADN sense necessitat d'obrir-lo (veure figures 3.2 i 3.8).



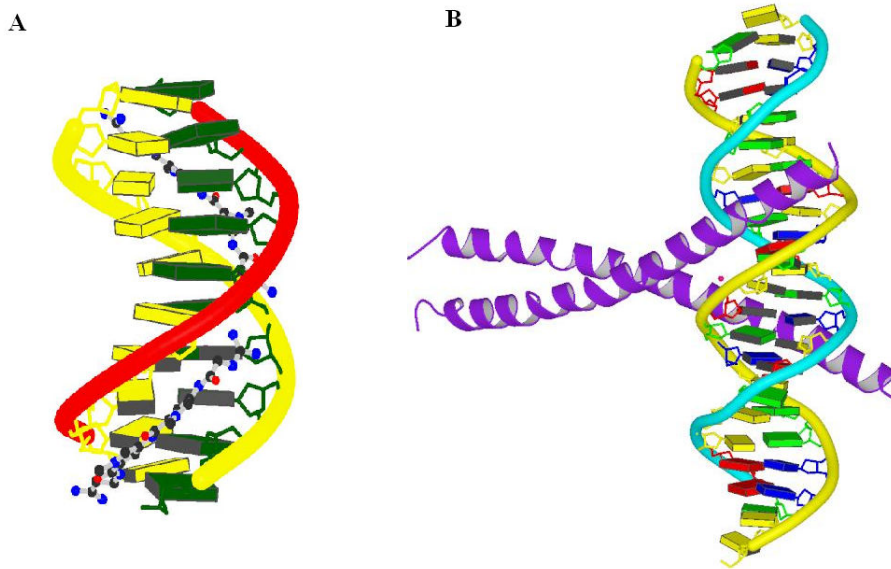
**Figura 3.7** Aparellaments entre bases amb més de dos ponts d'hidrogen. Entre ells el que més està representat en la naturalesa és el Watson-Crick que es dona sempre entre una purina i una pirimidina.

### 3.2.2 Interaccions d'apilament

En l'ADN es dona el fet de que les bases es col·loquen una sobre l'altra de forma més o menys paral·lela a una distància aproximadament de 3.4 Å donant el fenomen conegut com apilament (o *stacking*). Físicament, l'apilament té una naturalesa complexa, les forces electrostàtiques guien l'orientació, però globalment són desfavorables, mentre que les forces dispersives (van der Waals) són menys direccionals, però fortament estabilitzants (Weiner, Kollman, D. T. Nguyen, & Case, 1985). L'efecte hidrofòbic no



estabilitza especialment l'apilament de les nucleobases, però sí fa que aquesta sigui la interacció dominant en l'ADN al desestabilitzar molt la formació de ponts d'hidrogen (P Hobza & J Sponer, 1999).



**Figura 3.8** Estructures tridimensionals d'una droga i una proteïna unides a l'ADN. A) Unió de dues Netropsines a l'ADN pel solc menor i B) Factor de Transcripció CREB unit a l'ADN per solc major

S'ha demostrat experimentalment que les bases tendeixen a agregar-se en solucions aquoses, essent l'apilament la força primordial d'aquest efecte en detriment dels ponts d'hidrogen que son molt penalitzats a l'aigua (veure més endavant). També s'ha comprovat com els parells que involucren les guanines (veure figura 3.9) solen ser els més estables i els que involucren la timina els que tenen menys estabilitat. La relació d'estabilitat de l'apilament és en general: purina:purina > pirimidina:purina > pirimidina:pirimidina.

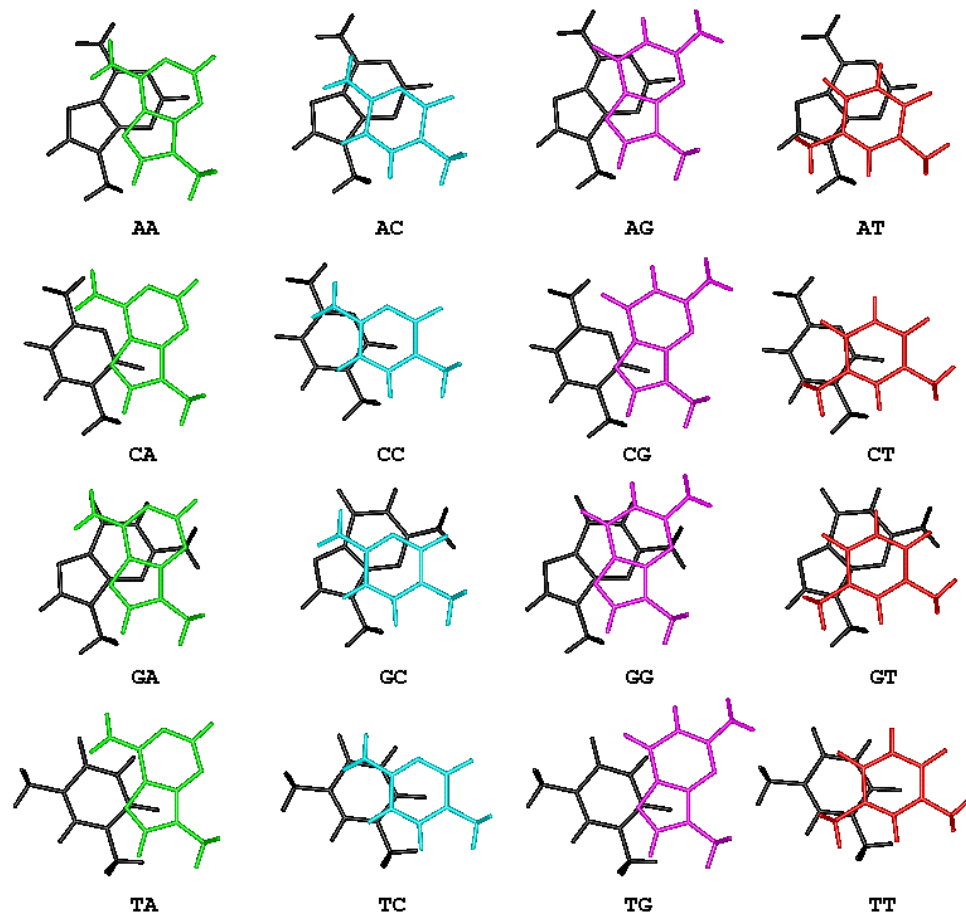


Figura 3.9. Aparellaments d'apilament entre bases canòniques d'ADN.

### 3.2.3 L'efecte del solvent

L'aigua i els ions són el solvent natural de les estructures dels àcids nuclèics. L'aigua és responsable de l'apantallament de les repulsions entre fosfats i redueix molt l'estabilitat dels ponts d'hidrogen que en altre cas serien massa estables i farien impossible transcripció i replicació. L'aigua modula també (encara que no tant) els apilaments i és responsable de detalls estructurals importants, com ara la geometria dels solcs. Fora de l'aigua les propietats del l'ADN poden canviar i cal ser prudent vers la informació derivada de d'algunes tècniques com per exemple; la espectroscòpia de masses

(Gidden, E. S. Baker, Ferzoco, & Bowers, 2005; B. S. Larsen & McEwen, 1998), que requereixen que aquest passi per un estat en fase gas (en el buit; sense solvent). S'ha demostrat que en aquest cas, la estructura de la hèlix canvia radicalment tot i que manté alguna de les seves propietats en dissolució (Manuel Rueda, Susana G Kalko, F Javier Luque, & Modesto Orozco, 2003).

### **3.3 PROPIETATS FISICOQUÍMIQUES DE L'ADN**

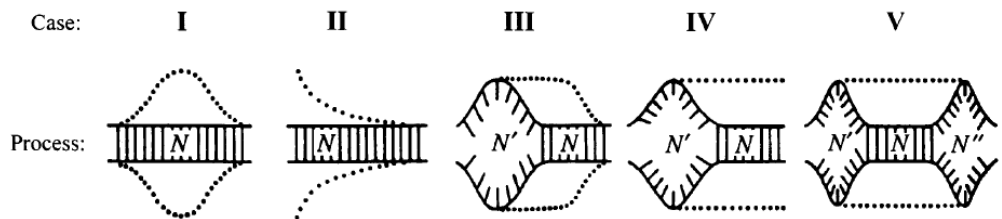
L'ADN codifica la informació genètica en la seva seqüència de bases. Per accedir a aquesta informació, la maquinària de transcripció ha de trobar la manera desmuntar les estructures d'ordre superior i finalment obrir l'hèlix fent que una de les cadenes pugui ser transcrita (veure capítol 2). Aquest és un procés complex, no directament dependent de seqüència, sinó de les propietats estructurals i físiques de l'ADN. Comprendre els factors que influeixen la estructura física i la dinàmica de la molècula de l'ADN ens ajudaria a comprendre el mecanisme cel·lular de la regulació i expressió genètica. Malauradament, mentre la informació sobre l'estructura secundària de l'ADN és ampla no passa lo mateix en els nivells d'estructura superior ni sobre les propietats físiques primàries de l'ADN. És per això, que la simulació està esdevenint l'eina fonamental per extreure simultàniament informació estructural i física sobre els àcids nucleics i en concret sobre l'ADN. En els següents punts descriurem algunes de les propietats físiques de l'ADN que es sospita que tenen un paper més rellevant per la modulació de les seves propietats biològiques.

#### **3.3.1 Propietats de fusió de l'ADN**

La desnaturalització de l'ADN, també coneguda com fusió, és el procés pel qual les cadenes de la molècula es desenrosquen i es separen pel trencament dels ponts d'hidrogen que uneixen les bases. El terme desnaturalització pot fer referència a aquest procés quan és motivat per un augment de la temperatura al medi o també quan la separació ve motivada per agents químics. Per convenció quan en el medi existeixen múltiples còpies de la mateixa molècula d'ADN, la temperatura de fusió ( $T_m$ ) es defineix com la temperatura en la qual la meitat de les molècules estan encara

formant una hèlix mentre les molècules de la segona meitat es troben ja separades (Santalucia, 1998). La temperatura de fusió (o *melting* en anglès) depèn tant de la longitud de la molècula d'ADN, com de la seva composició (seqüència de bases) i en general l'ADN guanya estabilitat amb la mida i amb el % de parells GC (Shchyolkina et al., 2000).

La desnaturalització de l'ADN en seqüències petites (~15 nts) és un procés que es pot considerar que es completa en un sol pas, però quan la mida és superior la situació es una mica més complexa. La fusió de seqüències genòmiques es dona en fases que afecten dominis (o sub-regions) que es dissocien de forma coordinada fins que l'ADN s'acaba separant en cadenes independents. L'estabilitat local de cada domini depèn de la seva seqüència, però també del seu entorn i situació (veure figura 3.10). Un domini pot derivar en la formació d'una bombolla interna (cas I) o si està en la part terminal (cas II) obrir un extrem de la seqüència. Altres possibilitats són la elongació d'una obertura (cas III) o la expansió d'una bombolla (cas IV). Finalment si la regió separa dues bombolles veïnes (cas V), la seva fusió les agruparà en una sola (RD Blake et al., 1999). Estem per tant parlant d'un procés complex amb molts intermediaris i on el significat real de la temperatura de fusió no es massa clar.



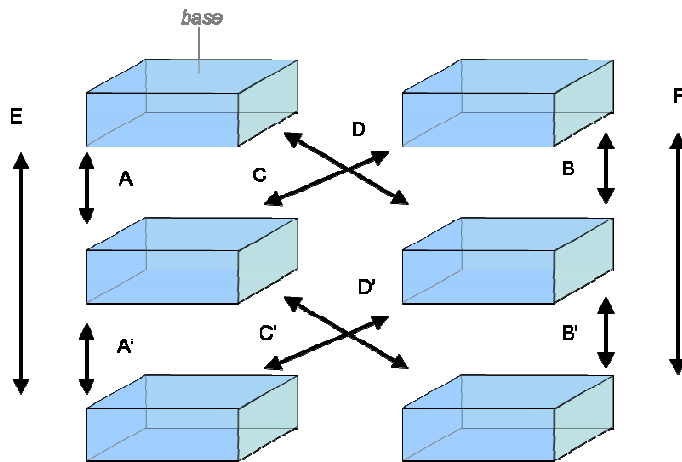
**Figura 3.10** Representació esquemàtica dels cinc processos de dissociació de l'ADN (RD Blake et al., 1999). Una mateixa seqüència pot tenir una temperatura de fusió diferent depenent de en quina d'aquestes fases es trobi.

### 3.3.2 Energies d'apilament

Tal com s'ha explicat més amunt, l'apilament de les bases en l'ADN és un dels

principals responsables de la estabilitat de la molècula. La energia total d'apilament d'una seqüència es pot descompondre de forma lineal en funció de l'energia d'apilament entre bases veïnes (A,B,C,D; veure figura 3.11) i en menor magnitud també es poden considerar les energies entre bases no contigües (E,F). Les energies d'apilament que regeixen la hèlix es poden dividir entre les que afecten una sola cadena (intra-cadena) i les que es donen entre dues cadenes (inter-cadena) que poden tenir un nivell d'intensitat similar a les inter-catenàries.

Al ser l'apilament un element que construeix en major mesura a l'estabilitat de la hèlix d'ADN, aquesta energia esta fortament relacionada amb la temperatura de fusió d'un segment curt d'ADN.



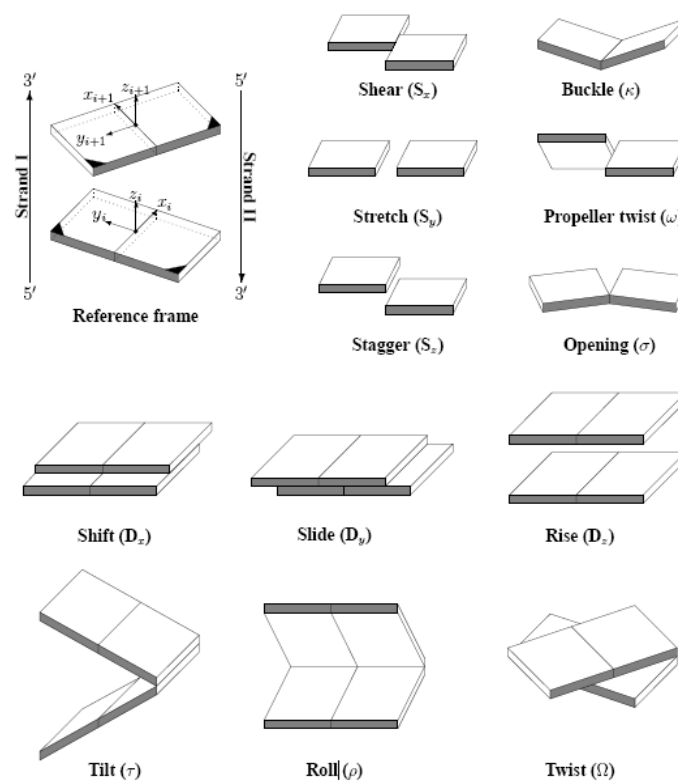
**Figura 3.11** Representació esquemàtica de les diferents contribucions energètiques que modulen l'apilament entre parells de bases.

### 3.3.3 Geometria de la hèlix d'ADN.

La geometria d'una molècula d'ADN es pot descriure de manera aproximada a partir de la col·locació de cada una de les bases que componen la seqüència a l'hèlix. Cada parell de bases es disposen entre sí com a dos blocs enfrontats entre amb una disposició específica. Agafant com a referència una de les dues bases podem definir

posicionament espacial de la segona a partir de tres paràmetres translacionals (shear[x], stretch[y], stagger[z]) i tres rotacionals (buckle[x], propeller-twist[y], opening[z]), tal com es mostra en la figura 3.12.

De manera paral·lela la geometria de dos parells de bases (o d'un dinucleòtid) pot ser enterament caracteritzada per 6 paràmetres helicoïdals (R.E. Dickerson, 1989): 3 translacionals (shift[x], slide[y], rise[z]) i 3 rotacionals (tilt[x], roll[y], twist[z]). Aquest valors defineixen de forma precisa la localització i orientació de cada parell de bases de l'ADN relativament al seu antecessor, caracteritzant així la estructura helicoïdal de la molècula. Fent servir aquest model helicoïdal és possible descriure raonablement àcids nucleics canònics amb un número molt reduït de paràmetres geomètrics i amb un nivell de resolució força acceptable.



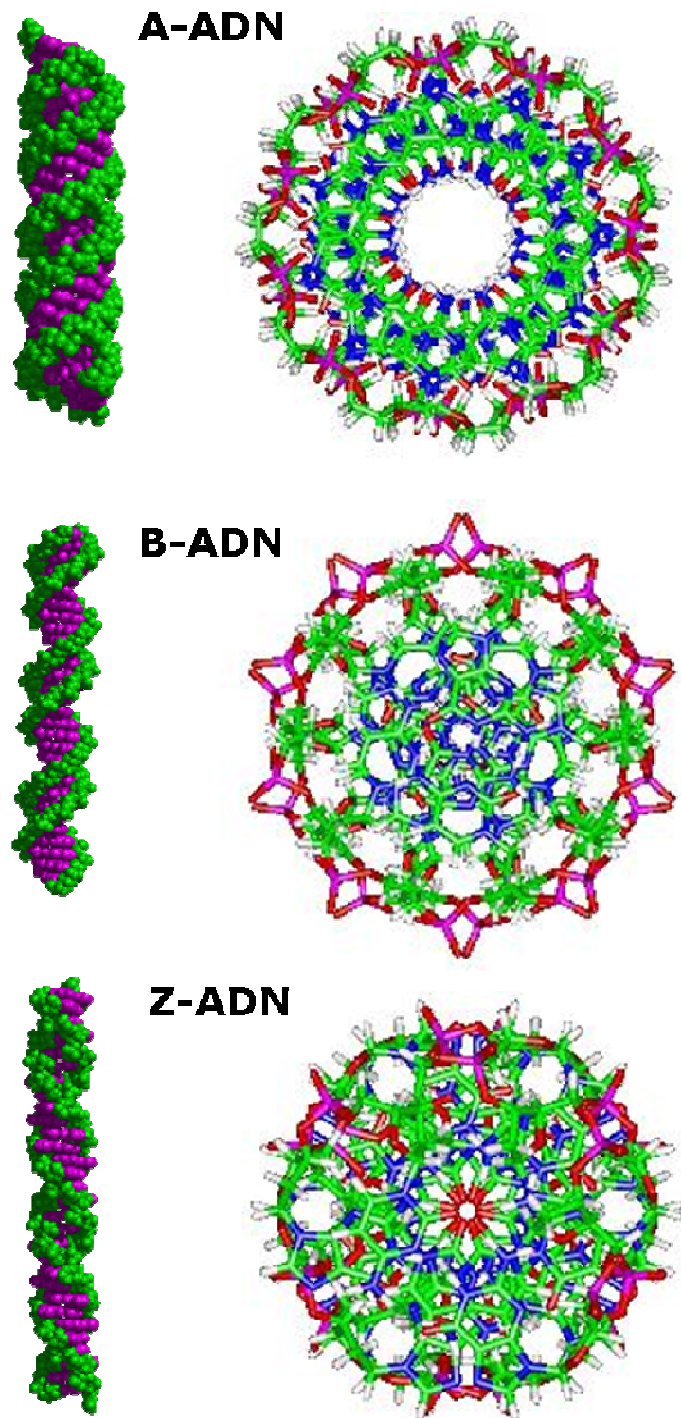
**Figura 3.12** Definició dels paràmetres que relacionen les bases complementaries i la seqüència de passos entre parells de bases (R.E. Dickerson, 1989, p. dickersondickers).

L'ús del sistema de referència helicoïdal permet definir l'existència de dos grans famílies estructurals als àcids nucleics: i) dextrogires i ii) levogires. Les primeres venen definides per valors de twist positius, mentre que les segones tenen valors de twist negatius. Dins de la família levogira trobem la forma Z de l'ADN (veure figura 3.13) mentre que dins de la família dextrogira trobem les formes més importants fisiològicament dels àcids nucleics: i) la forma A i ii) la forma B (Figures 3.13). La forma A-ADN és la majoritària en l'ARN fisiològic i es pot donar en l'ADN en condicions d'hidratació baixa o be unit a certes proteïnes, mentre que la segona es la forma majoritària en l'ADN fisiològic.

La forma A es correspon a una hèlix molt ampla, amb un forat central i amb una forta "inclinació" en el parells de bases, twist al voltant d'aproximadament 32° (el que dona 11 parells de bases per volta), tots els nucleòtids estan en conformació anti, mentre que els sucres estan en conformació Nord (C3'endo), el solc "estret" superficial i ample, mentre que el solc "ample" es estret i profund. La forma B es representa com una hèlix més estilitzada, amb un twist d'aproximadament 36° (10 parells de bases per volta), tots els nucleòtids estan en conformació anti, mentre que els sucres estan en conformació Sur (C2'endo), el solc "estret" es estret, mentre que el solc "ample" és ample i una mica mes profund (veure Taula 3.1).

| <b>Família</b>  | <b>Família B</b>  | <b>Família A</b>                                  | <b>Z-ADN</b>   |
|-----------------|---|---|--|
| Gir de la hèlix | Dextrogir   | Dextrogir   | Levogir  |
| Volta           | 8-10 nucleòtids   | 11 nucleòtids                                     | 12 nucleòtids  |
| Solc            | El major és més ample i una mica més profund que el menor | El major és més estret i més profund que el menor | El menor és molt profund i estret. El major és quasi inexistent. |

Taula 3.1. Característiques de les 3 famílies de formes d'ADN de doble cadena.

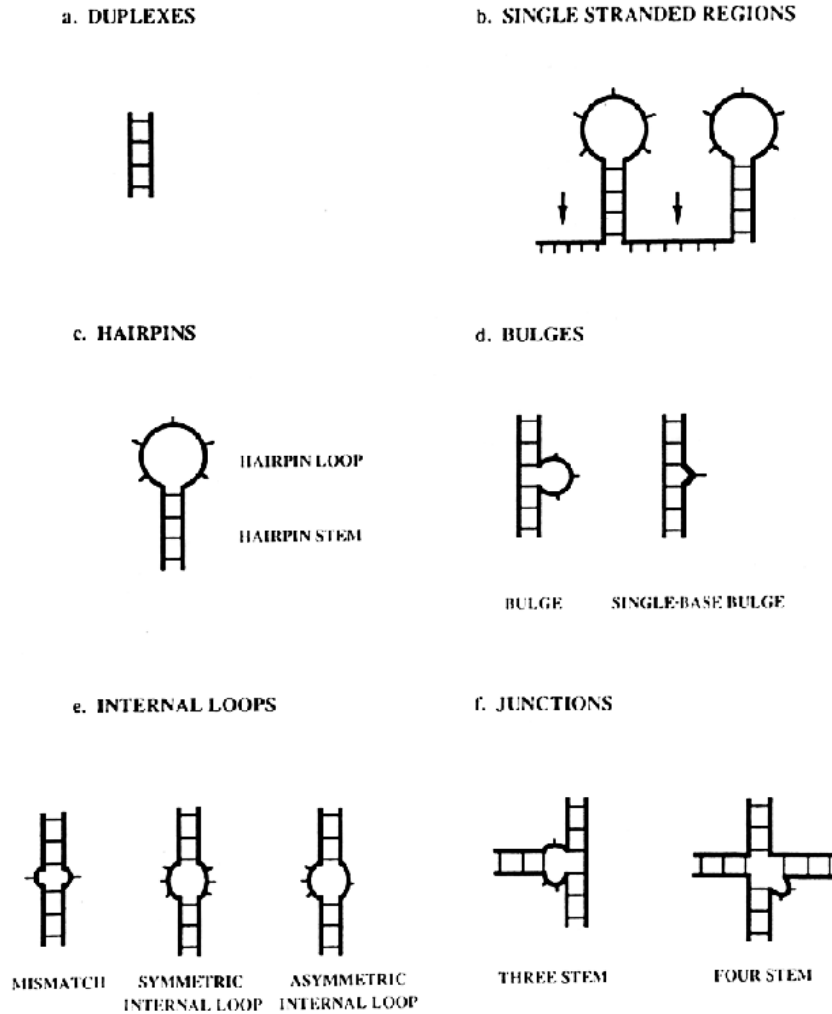


**Figura 3.13** Imatge lateral i frontal d'un A-ADN (part superior) d'un B-ADN (part central) i d'un



### **3.3.4 Propietats de flexibilitat de l'ADN**

La flexibilitat en la molècula d'ADN està estretament lligat a les propietats intrínseques de la geometria de l'ADN i és determinant en les interaccions ADN-proteïna ja que, moltes de les proteïnes que s'enganxen a l'ADN l'acaben distorsionant (Bates & Maxwell, 2005). Un exemple d'aquest fenomen el trobem en la proteïna que s'uneix a la caixa TATA (veure capítol 2), que produeix un doblegament d'uns  $100^\circ$  aproximadament (Bates & Maxwell, 2005); veure Figura 2.8. En aquest cas la seqüència té un important efecte per aconseguir aquest angle. La substitució del motiu TATAAAA per TAAAAAA redueix dramàticament l'angle i també l'activitat transcripcional (es creu que és provocat per la pèrdua d'un pas extremadament flexible com l'AT). Un segon exemple molt rellevant és el cas de les proteïnes que s'uneixen a l'ADN per embolcallar l'ADN en el nucleosomes (veure secció Estructures superiors de l'ADN). Tot i que, per la seva naturalesa els nucleosomes són capaços d'enroscar qualsevol seqüència d'ADN, els experiments han demostrat la preferència de la formació d'aquest complex en seqüències amb la propietat de tenir una flexibilitat anisotròpica (Bates & Maxwell, 2005).

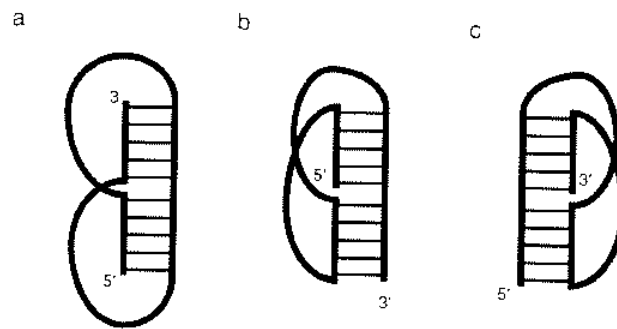


**Figura 3.14.** Esquema de formes inusuals locals que poden tenir els àcids nuclèics, molt especialment l'ARN.

### 3.4 FORMES NO CANÒNIQUES DE L'ADN

Les estructures secundàries de l'ADN adopten sempre que és possible la forma helicoïdal, en general de doble hèlix. En el cas de l'ARN la situació es depèn del grau de auto-complementarietat de la seqüència, trobant-se moltes irregularitats locals a l'estructura (veure figura 3.14) com: i) regions de cadena simple sense estructura, ii) forquetes (*hairpins*), iii) proturberàncies locals (*bulldges*), iv) bucles (*loops*) i v)

creuaments (*junctions*), que també es formen en alguns moments de la vida cel·lular a l'ADN. A banda d'aquestes estructures irregulars locals, a l'ARN es troba sovint la formació d'altres estructures irregulars no locals com són els pseudo-nusos (*pseudoknots*), gràcies als quals la estructura de l'ARN es pot compactar millor a partir d'interaccions molt lluny a la seqüència (veure Figura 3.15).



**Figura 3.15** Els tres tipus més comuns de *pseudoknots* que es troben a l'ARN (Chastain & Tinoco, 1991)

Finalment, cal senyalar que a més de les estructures de doble hèlix, tant l'ADN com l'ARN poden formar estructures helicoïdals no canòniques com són les hèlix intra- o inter-catenàries de 3 i a 4 cadenes estabilitzades per ponts d'hidrogen i interaccions d'apilament.

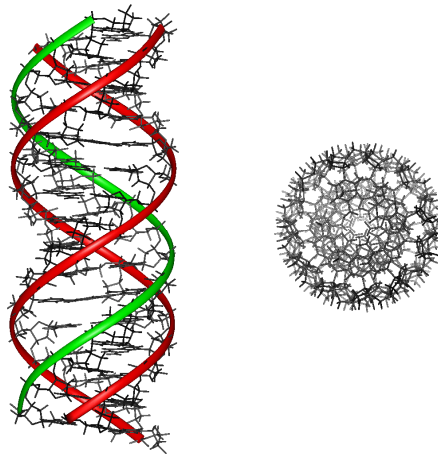
### 3.4.1 Tríplex

La facultat dels àcids nuclèics per formar triple hèlix es coneix des dels anys 50-60, però en els últims anys un gran nombre d'estudis s'han centrat de nou en aquesta estructura degut a la seva possible implicació en processos biològics i el seu potencial en aplicacions biotecnològiques (J S Sun, Garestier, & Hélène, 1996).

Una triple hèlix es forma quan una cadena de polinucleòtids es col·loca en el solc major de una doble hèlix d'ADN, al temps que les seves bases interaccionen per mitjà de ponts d'hidrogen Hoogsten (o reverse Hoogsteen) amb les purines del l'aparellament

Watson-Crick (veure figura 3.16). Les triple hèlix es poden formar en ARN en ADN on en una combinació d'aquests.

La incorporació de la tercera cadena, també denominada TFO (en anglès *triplex forming oligonucleotide*), en el solc major de la doble hèlix, coneguda com TTS (en anglès *triplex target sequence*), fa que aquest es faci més ample i es subdivideixi en dos solcs nous. Existeixen dos famílies de tríplex depenen del tipus de bases que componen el TFO: el primer tipus amb el TFO ric en pirimidines (T o C) o el segon tipus on el TFO es ric en purines (A o G). En els dos casos, es manté invariant la interacció del TFO exclusivament amb residus homo-purina de la mateixa cadena del TTS.

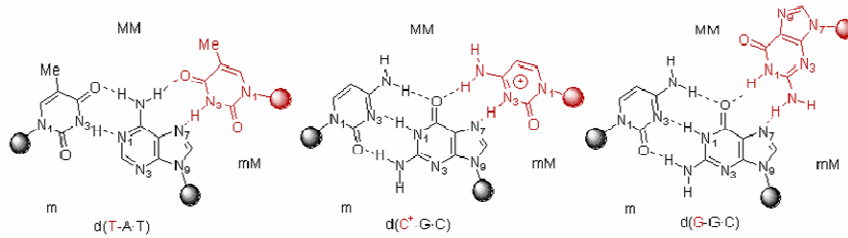


**Figura 3.16** Model de la triple hèlix en ADN. La estructura de l'ADN de triple hèlix està constituïda per dues cadenes en estructura Watson-Crick formant un dúplex (vermell), però amb una cadena extra que es disposa al llarg del solc major (verd) unida per ponts d'hidrogen del tipus Hoogsteen.

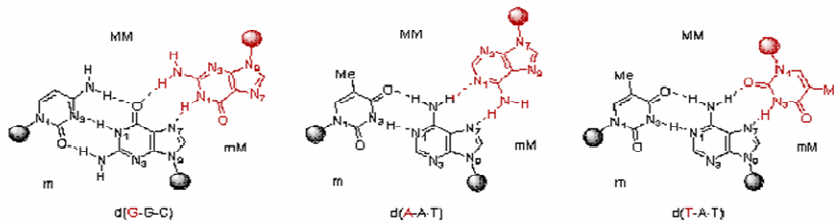
En la família de tríplex amb TFO rics en pirimidina, la tercera cadena es manté paral·lela a la cadena de homo-purines a la qual, s'uneix per ponts d'hidrogen Hoogsteen per formar tríades de  $(C \cdot G) \cdot C^+$  i  $(T \cdot A) \cdot T$  segons es mostra en la figura 3.17. El signe positiu (+) en la tripleta  $(C \cdot G) \cdot C^+$  fa referència a la protonació del N3 de la citosina del TFO que es necessària per la formació del tríplex.

El segon tipus de tríplex, amb el TFO enriquit amb purines, la tercera cadena és en aquest cas anti-paral·la a la cadena de purines del TTS. La interacció entre les dues es dona a través de punts d'hidrogen Hoogsteen revers per formar les tríades (C·G)·G, (T·A)·A i (T·A)·T tal com es mostra en la figura 3.17.

### Parallel triplex



### Anti-parallel triplex



**Figure 3.17** Possibles triplets de nucleòtids per la formació de triple hèlix. En negre el dúplex (TTS) i en vermell la tercera cadena (TFO). Les tres primeres formes son possibles quan el TFO és paral·lel a la cadena d'homo-purines del TSS. En les tres inferiors el TFO es antiparal·lel. La forma (C·G)·C<sup>+</sup> requereix la protonació de la C en el TFO, cosa que la fa dependent del pH.

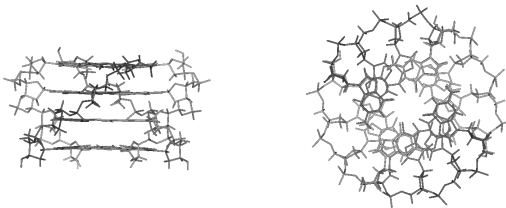
La estabilitat dels tríplex depenen de la composició de la seqüència i de possibles desencerts entre el TFO i la cadena d'homo-purines del TTS. Una interrupció en la seqüència de homo-purines per una pirimidina redueix dràsticament la estabilitat dels tríplex. En el cas dels tríplex paral·lels (seqüència timines en TFO), l'estabilitat també és dependent del pH del medi donat el requeriment de protonació de les C.

Al contrari del que inicialment es suposava degut a estudis de difracció de fibres estudis

recents de dinàmica molecular (Shields, C. A. Laughton, & M. Orozco, 1997; Soliva, C.A. Laughton, F.J. Luque, & M. Orozco, 1998) i experiments de NMR (Plum & Breslauer, 1995) suggereixen que els tríplex de DNA són més similars a la família B dels dúplex que a la A, amb tots els sucres en conformació *Sur*. La formació d'una triple hèlix en una doble hèlix d'ADN, però altera radicalment les seves propietats conformacionals, físiques i de reactivitat degut a la invasió del solc ample per part del TFO, el que dificulta que aquest tríplex pugi ser reconegut per la maquinària cel·lular de replicació y transcripció.

### 3.4.2 Quàdruplex: ADN de quatre cadenes

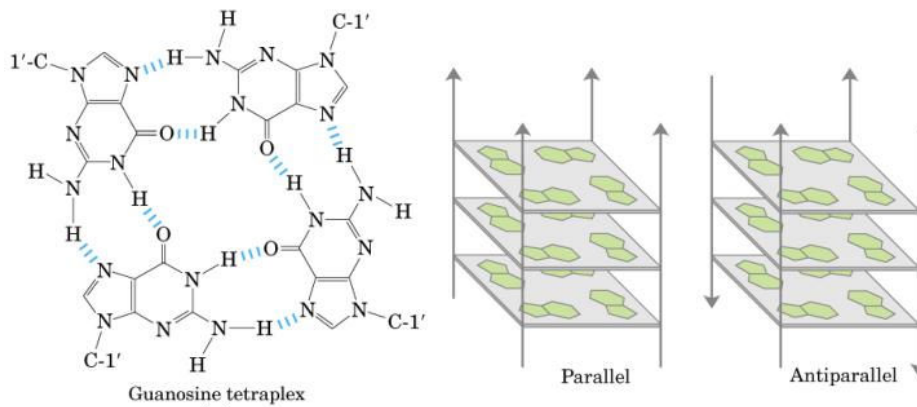
És conegut que els oligonucleòtids rics en guanines poden agregar-se junts formant una hèlix de quatre cadenes. Aquesta estructura, que es dona en presència d'un catió monovalent com  $\text{Na}^+$  o  $\text{K}^+$ , es coneix com G-ADN, quàdruplex o tétraplex (veure figura 3.18). Els quàdruplex són estructures inusuals de l'ADN que es donen *in vivo* en regions telomèriques, al final dels cromosomes (Williamson, Raghuraman, & Cech, 1989) i són claus en la seva estabilització. Durant el procés de replicació de la cèl·lula, es necessari que existeixi una porció final de l'ADN dels cromosomes extra en l'extrem 3' del cromosoma que actuï d'iniciador per la síntesis de la cadena complementaria. Aquestes regions extra son riques en guanines i la seva presència és molt important en aquest procés. El fragment extra de poli-G es replega sobre si mateix i les guanines es reconeixen mútuament per ponts d'hidrògen del tipus Hoogsteen (veure figura 3.19).



**Figura 3.18** Model de hèlix de quatre cadenes d'ADN. La estructura del quàdruplex queda estabilitzada per un ió  $\text{Na}^+$  o  $\text{K}^+$  en el seu interior.

Les quatre guanines, conegudes com quartets-G, es troben apilades una damunt de l'altra formant la hèlix de quatre cadenes (Laughlan et al., 1994). La disposició de les cadenes pot ser paral·lela o anti-paral·lela, depenent de la naturalesa de la seqüència que connecta les guanines o de dels cations que estableixen la estructura, típicament ions monovalents com  $\text{Na}^+$  o  $\text{K}^+$  (veure figura 3.19). La conformació de les guanines pot ser tota anti (cas del tetraplexe inter-catenaris paral·lels) o la meitat syn i la meitat anti (cas dels tetraplexes intra-catenaris antiparal·lels).

G tetraplex = 4 stranded DNA structure



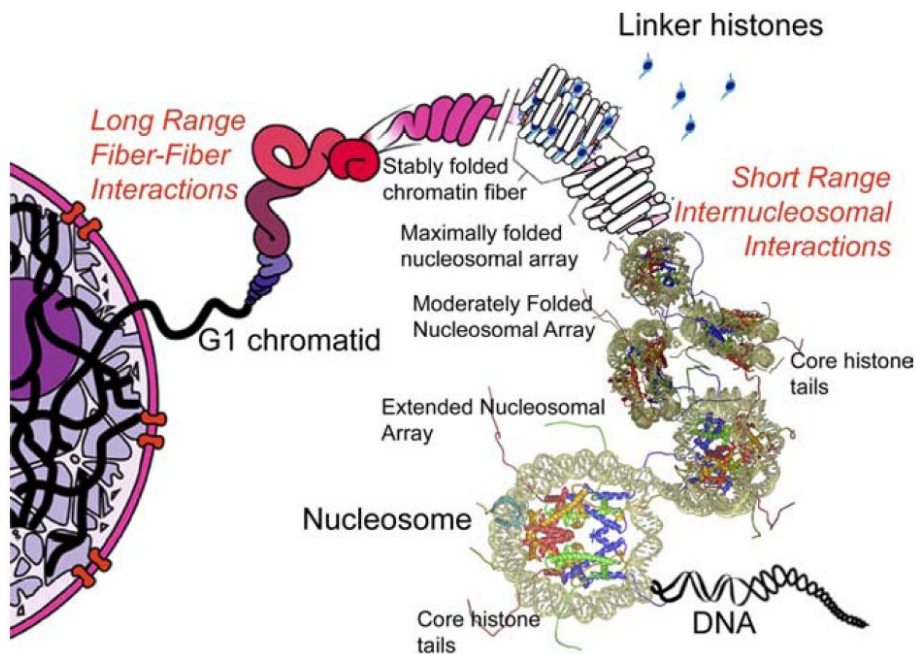
**Figura 3.19** Tétrades de les guanines en quàdruplex. Les quatre guanines s'aparellen mitjançant pont d'hidrògen del tipus Hoogsteen. Les cadenes del quàdruplex poden ser paral·leles o antiparaleles.

Els tétraplex son estructures molt important *in vivo*, però la seva formació no es desitjable quan s'està treballant amb formació de triplex, especialment amb cadenes de TFO riques en guanine, ja que si es forma un tétraplex intramolecular en la cadena TFO el triplex no es forma. Es per això que s'ha treballat molt en el disseny de bases modificades que desestabilitzin el tetraplex sense afectar, o idealment fins i tot estabilitzant, el triplex (Spacková, Elena Cubero, Jirí Sponer, & Modesto Orozco, 2004b).

### 3.5 ESTRUCTURES D'ORDRE SUPERIOR

L'ADN adopta nivells d'estructura superior a conseqüència de la necessitat de compactar-se eficientment en l'interior del nucli cel·lular. Aquests nivells de compactació no són fixes al llarg de la vida cel·lular, ja que l'ADN ha de canviar el nivell d'estructuració en funció de l'etapa del cicle en el que és trobi la cèl·lula.

L'eix de la hèlix d'ADN pot retorçar-se per formar una súper-hèlix, el que produeix un major grau d'empaquetament. Aquest enrotllament porta a la doble hèlix d'ADN a organitzar-se en cromosomes mitjançant un procés que es completa per etapes. La primera etapa és la estructuració de l'ADN per formar nucleosomes, associacions de l'ADN amb histones. La segona etapa consisteix en la condensació dels nucleosomes en fibres de cromatina de diferent amplada. Finalment la cromatina s'empaqueta en cromosomes (G R Smith, 1981), la estructura superior de l'ADN (veure figura 3.20)

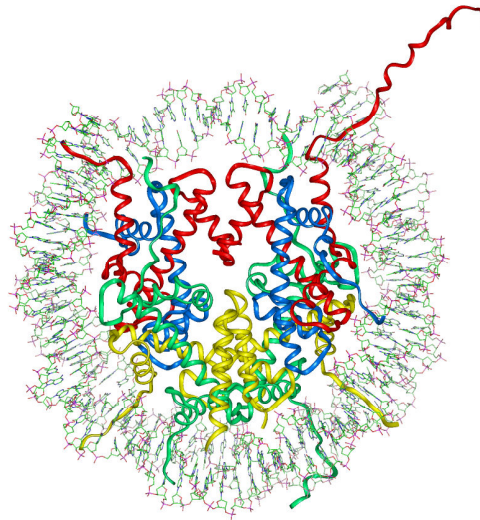


**Figura 3.20** Nivells d'empaquetament de l'ADN



### 3.5.1 Estructura del nucleosoma

La primera etapa en la condensació de l'ADN és el nucleosoma, on l'ADN cromosòmic és empaquetat en una estructura compacta amb l'ajuda d'unes proteïnes especialitzades que s'anomenen histones.



**Figura 3.21** Estructura d'un nucleosoma, la unitat fonamental d'empaquetament. Les 8 proteïnes, 2 còpies de cada tipus de histona (cada una en color diferent), es troben en la part central, mentre l'ADN es replega al voltant de les proteïnes. Estructura determinada per (Luger, Mader, R. K. Richmond, Sargent, & T. J. Richmond, 1997).

Un nucleosoma (veure figura 3.21) es forma quan un fragment de la doble hèlix d'ADN s'enrosca fortament al voltant del nucli central de les proteïnes-histones. Una histona addicional (H1) s'encarrega d'estabilitzar el segment que connecta directament amb el nucli (*core*) del nucleosoma, a on existeixen quatre histones diferents (H2A, H2B, H3 i H4) que s'agrupen per formar un octàmer (tetràmer de dímers). La unitat del nucleosoma està formada per 2 sub-bloqs d'histones (4 + 4) i el segment d'ADN que enrosquen, aproximadament 145 parells de bases, al voltant del nucli central. La

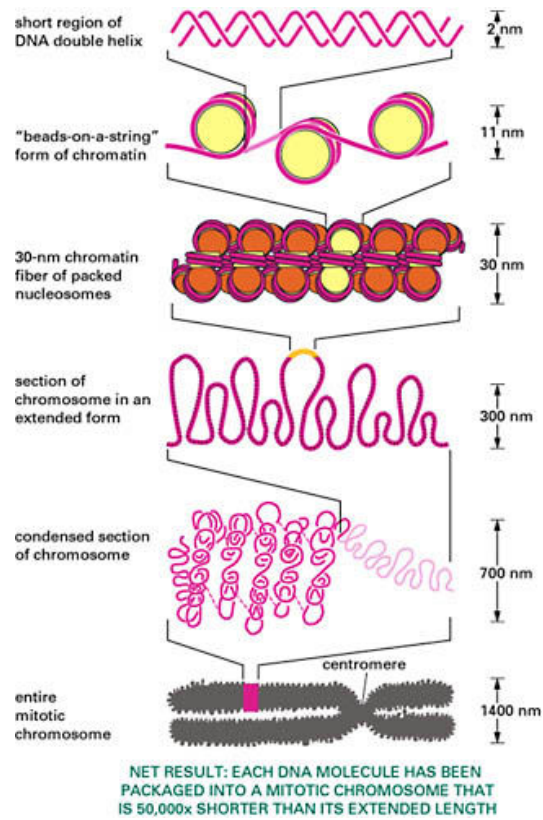
seqüència d'ADN d'un nucleosoma dona aproximadament 1.75 super-voltes levogires al voltant del nucli proteic del nucleosoma, aconseguint un empaquetament amb la relació 1:7. Els nucleosomes estan espaiats cada 200 parells de bases en organismes eucariotes i es suposa que hi ha un subtil codi, probablement relacionat amb les propietats físiques de la fibra que guia el seu posicionament, encara que aquest està encara per establir, i sembla dependre molt de factors com el grau de modificació química de les histones.

### **3.5.2 Estructures superiors al nucleosoma**

Les estructures corresponents a ordres superiors d'empaquetament cromosòmic (veure figura 3.22) són més especulatives, donat que no s'han pogut realitzar estudis d'alta resolució sobre les mateixes. Això és degut a la gran mida i heterogeneïtat de la cromatina.

L'ADN en la cromatina s'organitza en agrupacions de nucleosomes. Els nucleosomes normalment s'empaqueten junts amb l'ajuda de la histona H1 i d'un fragment separador d'ADN. D'aquesta manera es formen fibres de 10-11 nm de diàmetre que adopten una aparença de collar. Aquesta fibra s'empaqueta seguint un model conegut com "solenoides" amb uns 5 nucleosomes per volta que dona una super-hèlix de uns 30 nm de diàmetre (veure figura 3.22), amb la que aconseguim un empaquetament dels nucleosomes amb una relació de 1:40.

En fases posteriors de condensació d'ADN, un dels models proposats (Alberts et al., 1999) suggereix la formació de llaços a partir d'aquestes fibres de 30 nm. Cada una contindria aproximadament 50 voltes de solenoides i possiblement s'empaquetaria en una super-espiral, unint-se a un nucli central. L'hèlix continuada de llaços constitueix el cromosoma eucariòtic.



**Figura 3.22** Nivells d'empaquetament de l'ADN. Aquestes fases successives d'empaquetament donen una estructura més compacta al cromosoma.