

# Apèndix

*[Aquesta pàgina ha estat deixada en blanc intencionadament]*

## **DNA, GENS, PROTEÏNES I MUTACIONS**

La informació hereditària de tots els organismes vius, amb l'excepció d'alguns virus, està continguda en l'àcid desoxiribonucleic (ADN). Normalment l'ADN consisteix en dues cadenes complementaries enrotllades sobre elles mateixes formant una hèlix dextrògira. Cada cadena és un polinucleòtid lineal compost per 4 nucleòtids diferents. Aquests nucleòtids estan composts per un sucre de 5 baules (desoxiribosa) un grup fosfat i una base nitrogenada que és el que els diferencia. D'aquestes bases, dues són purines, adenina (A) i guanina (G) i les altres dues són pirimidines timina (T) i citosina (C).

Les dues cadenes s'uneixen al llarg de la seqüència per ponts d'hidrogen entre les bases dels nucleòtids. Aquests ponts sempre són els mateixos i una purina s'uneix a una pirimidina. l'adenina s'uneix a la timina per mitjà de dos ponts d'hidrogen i la guanina a la citosina per mitjà de tres ponts d'hidrogen.

Cada nucleòtid s'uneix entre ells dins d'una cadena per enllaços covalents de fosfodiéster asimètrics 5'-3'. La cadena doncs és polaritzada, un extrem conté un radical fosforil al carboni 5' del sucre del nucleòtid terminal. L'altre extrem conté un grup hidroxil lliure en el carboni 3' de la desoxiribosa terminal. Així la direcció dels enllaços fosfodiéster determina el caràcter de la molècula. Per convenció els nucleòtids en una seqüència de ADN es nomenen en l'ordre de la transcripció de l'extrem 5' a l'extrem 3'. D'aquesta manera la doble hèlix de DNA conté dues seqüències antiparal·leles i complementaries.

### **GENS**

La dotació completa de material genètic contingut pel nucli d'una cèl·lula eucariota és el genoma i la porció de genoma que conté gens s'anomena ADN gènec. La resta de material genètic s'anomena no gènec. El conjunt de seqüències

transcrites per un genoma és el transcriptoma i el conjunt de proteïnes codificades pel genoma s'anomena el proteoma.

El genoma es duplica abans de cada etapa de duplicació en la cèl·lula.

Tradicionalment s'ha definit com a gen el segment de ADN que codifica per una cadena polipeptídica o que transcriu una molècula funcional de ARN. Tanmateix recents estudis moleculars han mostrat una idea més vague del que és un gen, així un gen es una seqüència de ADN genòmic o ARN que és essencial per una funció específica. Per dur a terme la funció pot no requerir la traducció o fins hi tot la transcripció.

D'aquesta definició es classifiquen els gens en tres grups diferents. Els gens que codifiquen per proteïnes que son transcrits a ARN i després traduïts a proteïna. Els gens que codifiquen per ARNs funcionals, que només es transcriuen. Finalment en el tercer grup són gens que no s'arriben a transcriure.

Els gens que codifiquen per proteïnes consisteixen en parts transcrites i parts no transcrites. Les parts no transcrites es designen segons la posició que ocupen en relació amb la part transcrita, així tenim les regions 5' i 3' flanquejants. La regió 5' conté diferents seqüències que s'anomenen senyals o regions promotores i que determinen la iniciació, el temps d'iniciació, la durada i la especificitat de teixit de la transcripció. Les regions 3' contenen senyals per la terminació del procés de transcripció així com altres elements reguladors. Tanmateix el poc coneixement actual de tots aquests elements fa molt difícil actualment de saber quan i com es transcriurà un gen coneixent les seves regions promotores.

El transcrit de ARN obtingut es coneix com a ARN missatger precursor o pre-mARN. Aquest posteriorment se li afegeix la cua de poliadenines o poli-A per obtenir el ARN missatger ja madur (mARN).

El pre-mARN conté exons i introns. Els introns són les seqüències que són eliminades en el processat del pre-mARN o *splicing*. La resta de seqüència que

forma el mARN són els exons. Els exons que són traduïts s'anomenen exons codificants i les regions a 5' i a 3' són les regions no traduïdes.

Els gens que codifiquen per RNA funcionals no són traduïts a proteïna. En cèl·lules eucariotes sofreixen maduració via splicing. Molts d'ells sofreixen posteriorment modificacions posttranscripcionals que inclouen modificació dels nucleòtids estàndard així com addicions enzimàtiques d'altres grups funcionals als extrems.

El que es coneix fins ara dels gens no transcrits no és tan exhaustiu com dels dos grups anteriors. Dins aquests s'han identificat diferents subgrups. Gens replicadors, que especifiquen l'inici i la terminació de la replicació. Són seqüències específiques dins els replicons que determinen el punt d'inici i de terminació de la replicació, també són zones d'anclatge tant de la maquinària de la replicació com d'inhibidors i activadors de la replicació. Els gens de recombinació determinen les seqüències per la recombinació durant la meiosi. Les seqüències telomèriques també s'inclouen en aquest grup. Els gens de segregació que donen unió específica a la maquinària que separa els cromosomes durant la segregació mitòtica i meiòtica. Finalment també s'inclouen regions que determinen la unió d'elements estructurals dins la maquinària cel·lular i de les zones que confereixen estructura específica als cromosomes.

Finalment cal fer esment als pseudogens, aquests són seqüències de ADN no gènic que mostra una similaritat important amb un gen funcional però que contenen defectes que l'impedeixen d'expressar-se. Alguns són transcrits i d'altres fins hi tot traduïts. Són presents en diferent proporció en tots els organismes.

## PROTEÏNES

Els aminoàcids són els elements estructurals que componen les proteïnes, totes les proteïnes de tots els organismes es componen a partir dels 20 aminoàcids primaris. Cada aminoàcid està compost per un grup  $-NH_2$  (amino) i un grup  $-COOH$  (carboxil) units a banda i banda d'un carboni central anomenat carboni  $\alpha$ . També units a aquest carboni tenim un hidrogen (H) i un grup químic que compona el que es coneix com cadena lateral i que es generalitza com a grup R.

El carboni  $\alpha$  és enantiomèric així es poden formar dos aminoàcids isomèrics per cada un d'ells no idèntics que són imatge especular un de l'altre. La natura només usa els L aminoàcids. Les cadenes laterals dels aminoàcids varien en forma, càrrega, capacitat formadora de ponts d'hidrogen, composició i reactivitat química i això els fa diferents els uns dels altres. Segons les seves propietats es classifiquen en diferents grups que es solapen i que els donen la plasticitat suficient per donar a les proteïnes totes les propietats necessàries per la vida.

Les proteïnes són macromolècules que consisteixen en una o més cadenes polipeptídiques on cada una té una única i definida seqüència. En aquesta cadena polipeptídica un  $\alpha$ -carboxil d'un aminoàcid s'uneix a un  $\alpha$ -amino del següent aminoàcid formant un enllaç peptídic. Les proteïnes tornen a ser molècules polars amb un extrem amina lliure que és per convenció el primer de la seqüència o N-terminal i un carboxílic lliure que correspon a l'últim o C-terminal. La longitud de les proteïnes varia molt des d'uns pocs aminoàcids fins a uns quants milers en els casos més extrems. Es solen considerar fins a 4 nivells d'estructuració en les proteïnes. L'estructura primària és simplement la seqüència lineal de la proteïna. La estructura secundària correspon a l'ordenació espacial o plegament dels residus més propers en la seqüència primària. Alguns d'aquests reordenaments espacials corresponen a estructures regulars que els confereixen certa periodicitat. Dins els elements d'estructura secundària tenim l'hèlix- $\alpha$ , aquí

la cadena principal de la seqüència peptídica es plega formant una hèlix dextrogira molt compacta que exposa a l'exterior les cadenes laterals dels seus aminoàcids. L'estructura s'estabilitza per mitjà de ponts d'hidrogen que es formen entre els residus de l'hèlix a intervals de 4 aminoàcids que és el que queda just una volta per sobre en l'hèlix. Un altre element són les làmines- $\beta$  en el qual s'uneixen cadenes- $\beta$  paral·leles o antiparal·leles per mitjà de ponts d'hidrogen entre cadenes adjacents. Altres zones no estructurades es coneixen com a *random coil*.

El plegament espacial dels residus no contigus o l'ordenament dels elements d'estructura secundària és el que es coneix com a estructura terciària. L'estabilització d'aquest nivell estructural és per mitjà de forces covalents els ponts disulfur i també de no covalents com són ponts d'hidrogen, interaccions hidrofòbiques i ponts salins.

Algunes proteïnes estan formades per més d'una cadena polipeptídica, cada una d'aquestes cadenes s'anomena subunitat i l'estructura quaternària correspon a l'ordenació espacial d'aquestes subunitats.

La síntesi de les proteïnes es du a terme per mitjà d'un procés de decodificació, on la informació genètica que conté un mARN es tradueix als 20 aminoàcids primaris amb l'ús de ARNs de transferència (tARN). Els tARN són petites molècules de ARN, usualment entre 70 i 90 nucleòtids de llargada. Cada un dels 20 aminoàcids té com a mínim un tipus de tARN associat i en general n'hi ha varis. La traducció s'esdevé pel reconeixement seqüencial dels triplets de nucleòtids adjacents i no solapats que s'anomenen codons per la seqüència complementària de tres nucleòtids (l'anticodó) en el tARN.

La traducció comença al codó d'iniciació i arriba fins al codó stop o de terminació. Un cop acabada la traducció de la proteïna la creació d'una proteïna funcional pot dependre de modificacions postraduccionals que poden ser

modificacions d'aminoàcids, eliminació de fragments terminals en ambdós extrems, transport intra o inter cel·lular o l'addició de grups prostètics.

Taula 1

PRIMERA POSICIÓ (extrem 5')	SEGONA POSICIÓ				TERCERA POSICIÓ (extrem 3')
	U	C	A	G	
U	UUU Phe (F)	UCU Ser (S)	UAU Tyr (Y)	UGU Cys (C)	U
	UUC Phe (F)	UCC Ser (S)	UAC Tyr (Y)	UGC Cys (C)	C
	UUA Leu (L)	UCA Ser (S)	UAA Stop	UGA Stop	A
	UUG Leu (L)	UCG Ser (S)	UAG Stop	UGG Trp (W)	G
C	CUU Leu (L)	CCU Pro (P)	CAU His (H)	CGU Arg (R)	U
	CUC Leu (L)	CCC Pro (P)	CAC His (H)	CGC Arg (R)	C
	CUA Leu (L)	CCA Pro (P)	CAA Gln (Q)	CGA Arg (R)	A
	CUG Leu (L)	CCG Pro (P)	CAG Gln (Q)	CGG Arg (R)	G
A	AUU Ile (I)	ACU Thr (T)	AAU Asn (N)	AGU Ser (S)	U
	AUC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)	C
	AUA Ile (I)	ACA Thr (T)	AAA Lys (K)	AGA Arg (R)	A
	AUGMet*(M)	ACG Thr (T)	AAG Lys (K)	AGG Arg (R)	G
G	GUU Val (V)	GCU Ala (A)	GAU Asp (D)	GGU Gly (G)	U
	GUC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)	C
	GUA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)	A
	GUG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)	G

\*AUG és la senyal d'iniciació i també codifica per la resta de residus de Met interns



La correspondència entre els codons i els aminoàcids ve determinada per un grup de regles anomenada el codi genètic. Amb poques excepcions el codi genètic per les proteïnes nuclears és universal.

Donat que un codó està format per tres nucleòtids i tenim 4 nucleòtids diferents en els àcids nucleics, hi ha  $4^3=64$  possibles codons diferents. En el codi genètic universal, que podem veure a la taula 1, tenim 61 codons que codifiquen per diferents aminoàcids. Els tres restants són codons de stop o parada. Aquests tres codons són UAA, UAG, UGA.

Amb 61 codons i 20 aminoàcids primaris, molts dels aminoàcids (18 de 20) són codificats per més d'un codó, és el codi degenerat. Els diferents codons que codifiquen pel mateix aminoàcid són anomenats sinònims.



Calella/Barcelona, Juny de 2005