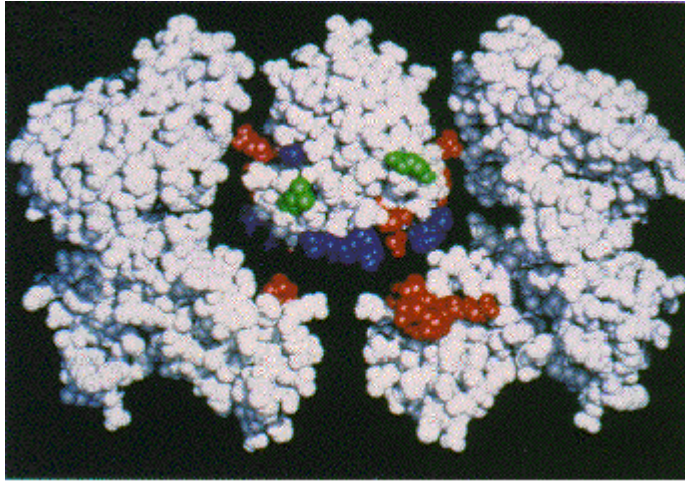


# A GENÉTICA MOLECULAR EM análises clínicas



Doutor José M. B. Cabeda  
Dr<sup>a</sup> Alexandra Estevinho  
Dr<sup>a</sup> Maria Luis Amorim

Porto, 1999

<b>1</b>	<b>A GENÉTICA MOLECULAR DA CÉLULA NÃO</b>	
<b>PATOLÓGICA</b>		<b>1.1</b>
1.1	CONCEITOS BÁSICOS DE GENÉTICA MOLECULAR	1.3
1.1.1	<i>O material genético</i>	1.3
1.1.1.1	Ácidos Nucleicos (DNA e RNA)	1.3
1.1.1.1.1	A estrutura do DNA	1.5
1.1.1.1.2	A estrutura do RNA	1.16
1.1.1.2	A ORGANIZAÇÃO DOS GENES NO GENOMA	1.17
1.1.1.3	Estrutura de um Gene	1.19
1.1.1.4	O CÓDIGO GENÉTICO	1.27
1.1.1.5	Mutações e Polimorfismos	1.29
1.1.2	- A FISIOLOGIA DO GENE	1.31
1.1.2.1	Transcrição e transcrição reversa	1.31
1.1.2.2	Tradução	1.36
1.1.2.3	Replicação	1.42
1.1.2.4	Mecanismos de Mutação do DNA	1.46
1.1.2.4.1	Erros na replicação do DNA	1.46
1.1.2.4.2	Lesões espontâneas	1.48
1.1.2.4.3	Recombinação genética	1.49
1.1.2.4.3.1	Recombinação por crossing-over	1.49
1.1.2.4.3.2	Recombinação somática	1.52
1.1.2.4.4	Mutações mediadas por elementos de transposição	1.52
1.1.2.5	Mecanismos de reparação do DNA	1.60
1.1.2.5.1	Mecanismos não excisativos	1.61
1.1.2.5.2	Mecanismos excisativos	1.63
1.1.2.5.3	Mecanismos pós-replicativos	1.65
1.2	O CICLO CELULAR	1.66
1.2.1	<i>Divisão celular e anomalias genéticas</i>	1.66
1.2.2	<i>Ciclins e CDK</i>	1.68
1.2.3	<i>Apoptose</i>	1.75
1.3	ACTIVIDADE CELULAR	1.77
1.3.1	<i>Componentes dos sistemas de sinalização</i>	1.77
1.3.2	<i>Regulação da sinalização por modulação da conformação proteica</i>	1.78
1.3.3	<i>A Superfamília dos receptores das hormonas esteróides</i>	1.78
1.3.4	<i>Receptores transmembranares; vias de transdução de sinal</i>	1.79

<b>2</b>	<b>ÁREAS DE INTERVENÇÃO DA GENÉTICA MOLECULAR EM</b>	
	<b>IMUNOLOGIA .....</b>	<b>2.2</b>
2.1	O RECEPTOR DA CÉLULA T (TCR) .....	2.3
2.1.1	<i>Estrutura somática dos genes do TCR</i> .....	2.4
2.1.2	<i>Mecanismo de rearranjo somático dos genes do TCR</i> .....	2.4
2.1.3	<i>POLIMORFISMOS DO TCR</i> .....	2.8
2.1.4	<i>O TCR EM SITUAÇÕES PATOLÓGICAS</i> .....	2.10
2.2	AS IMUNOGLOBULINAS (IG) .....	2.12
2.3	O MHC .....	2.14
2.4	O COMPLEMENTO .....	2.19
2.4.1	<i>As cascatas do complemento</i> .....	2.20
2.4.2	<i>O MHC III</i> .....	2.23
2.4.2.1	A proteína.....	2.23
2.4.2.2	O gene .....	2.25

<b>3</b>	<b>ÁREAS DE INTERVENÇÃO DA GENÉTICA MOLECULAR EM</b>	
	<b>MICROBIOLOGIA .....</b>	<b>3.1</b>
3.1	A GENÉTICA NA VIROLOGIA.....	3.2
3.1.1	- <i>O genoma viral</i> .....	3.3
3.1.2	- <i>O ciclo de vida Viral</i> .....	3.5
3.1.2.1	Passos iniciais da multiplicação viral .....	3.5
3.1.2.2	Fase da Multiplicação viral .....	3.6
3.1.2.3	O ciclo de vida dos retrovírus.....	3.14
3.1.3	- <i>Virologia em Medicina Transfusional</i> .....	3.21
3.1.3.1	- O vírus da Hepatite B (HBV).....	3.21
3.1.3.2	- O Vírus da Hepatite C (HCV).....	3.24
3.1.3.3	- O vírus Linfotrópico Humano (HTLV-I e HTLV-II) .....	3.25
3.1.3.4	- O vírus do síndrome da imunodeficiência adquirida (HIV) .....	3.27
3.1.4	- <i>Papel dos vírus em oncologia</i> .....	3.32
3.1.4.1	Oncogenes virais .....	3.34
3.1.4.2	Conversão de proto-oncogenes celulares em oncogenes .....	3.37
3.1.4.3	A translocação t(8;14)(q24;q32) e o EBV.....	3.38

<b>4</b>	<b>ÁREAS DE INTERVENÇÃO DA GENÉTICA MOLECULAR EM</b>	
	<b>HEMATO-ONCOLOGIA .....</b>	<b>4.1</b>
4.1	O CANCRO: UMA DOENÇA GENÉTICA DE CÉLULAS SOMÁTICAS.....	4.3
4.1.1	<i>A origem genética dos tumores .....</i>	<i>4.3</i>
4.1.1.1	Proto-oncogenes e oncogenes .....	4.5
	Factores de crescimento.....	4.6
4.1.1.1.2	Receptores para factores de crescimento.....	4.7
4.1.1.1.3	Transdutores intracelulares de sinal.....	4.7
4.1.1.1.4	Factores de transcrição nuclear.....	4.10
4.1.1.1.5	Proteínas de controlo do ciclo celular .....	4.11
4.1.1.2	Genes de Supressão tumoral .....	4.11
4.1.2	<i>Os carcinogénios como mutagénios.....</i>	<i>4.13</i>
4.1.3	<i>Agentes virais na origem dos tumores .....</i>	<i>4.15</i>
4.1.3.1	Oncogenes virais .....	4.15
4.1.3.2	Conversão de proto-oncogenes celulares em oncogenes .....	4.18
4.1.3.2.1	Activação de proto-oncogenes .....	4.18
4.1.3.2.2	Amplificação de proto-oncogenes .....	4.19
4.1.4	<i>A "estatística" na origem dos tumores: mutações espontâneas.....</i>	<i>4.19</i>
4.2	ANOMALIAS GENÉTICAS EM DOENÇAS HEMATO-ONCOLÓGICAS.....	4.20
4.2.1	<i>Translocações cromossómicas em hemato-oncologia.....</i>	<i>4.20</i>
4.2.1.1	Translocações envolvendo genes das Ig ou TCR .....	4.20
4.2.1.1.1	translocação t(8;14)(q24;q32) e o EBV.....	4.22
4.2.1.1.2	A t(14;18) - BCL2/IgH.....	4.23
4.2.1.2	Translocações que originam genes híbridos .....	4.25
4.2.1.2.1	t(9;22) - bcr/abl .....	4.25
4.2.1.2.2	t(15;17) - pml/rar .....	4.27
4.2.2	<i>Delecções cromossómicas em hemato-oncologia.....</i>	<i>4.30</i>
4.2.3	<i>Mutações pontuais em hemato-oncologia .....</i>	<i>4.30</i>
4.2.3.1	mutações no gene p53.....	4.30
4.3	PATOLOGIAS HEMATO-ONCOLÓGICAS.....	4.30
4.3.1	<i>LMC.....</i>	<i>4.30</i>
4.3.2	<i>LMA.....</i>	<i>4.30</i>
4.3.3	<i>LLA.....</i>	<i>4.30</i>
4.3.4	<i>LLC.....</i>	<i>4.30</i>
4.4	UTILIDADE CLÍNICA DA GENÉTICA MOLECULAR EM HEMATO- ONCOLOGIA.....	4.30
4.4.1	<i>Aprofundamento de conhecimentos.....</i>	<i>4.31</i>
4.4.2	<i>Escolha de tratamentos específicos.....</i>	<i>4.31</i>
4.4.3	<i>Monitorização da doença.....</i>	<i>4.31</i>

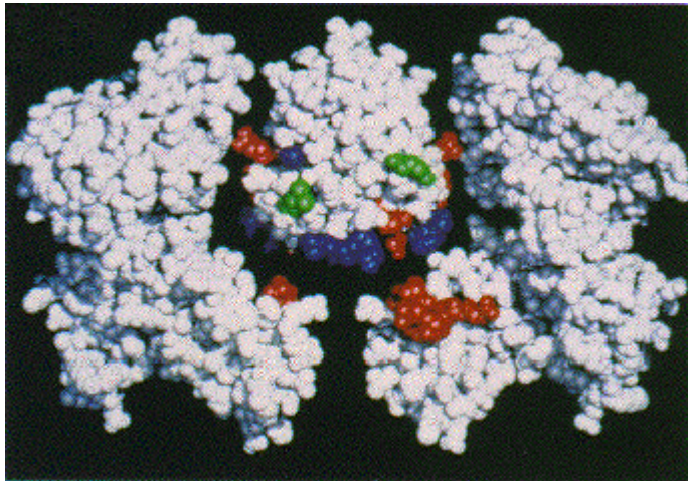
<b>5</b>	<b>ÁREAS DE INTERVENÇÃO DA GENÉTICA MOLECULAR EM</b>	
	<b>HEMATOLOGIA .....</b>	<b>5.1</b>
5.1	- DOENÇAS GENÉTICAS DO GLÓBULO RUBRO .....	5.3
5.1.1	- <i>Anemias Não esferocíticas Congénitas</i> .....	5.3
5.1.1.1	- Deficiência em Glucose-6-fosfato desidrogenase.....	5.3
5.1.1.2	- Deficiência em piruvato quinase.....	5.4
5.1.1.3	- Deficiência em $\delta$ -aminolevulinato sintetase (Anemia sideroblástica).....	5.5
5.1.2	- <i>Talassémias (anomalias das <math>\alpha</math> e <math>\beta</math>-globinas)</i> .....	5.6
5.1.3	<i>Drepanocitose</i> .....	5.11
5.1.4	- <i>Anemias Hemolíticas esferocíticas</i> .....	5.12
5.1.4.1	- Esferocitose Hereditária (HS) .....	5.13
5.1.4.2	- Deficiência grave de proteína 4.2.....	5.14
5.1.4.3	- Eliptocitose e poiquilocitose Hereditária .....	5.15
5.2	DOENÇAS GENÉTICAS EM HEMOSTASE.....	5.16
5.2.1	- <i>Resistência à proteína C Activada (mutação FV-Leiden)</i> .....	5.16
5.2.2	- <i>Mutações no gene da Protrombina</i> .....	5.18
5.2.3	- <i>Doença de von Willebrandt (Mutações no gene do vWF)</i> .....	5.19
5.2.4	- <i>Trombose familiar (Mutações nos genes da Antitrombina III, Proteína C e Proteína S)</i> .....	5.19
5.2.5	- <i>Hemofilias (Mutações nos genes dos factores VIII e IX)</i> .....	5.21
5.3	- HEMOCROMATOSE.....	5.22
5.3.1	- <i>Estudos de marcadores genéticos no locus do HLA</i> .....	5.23
5.3.2	- <i>Estudos dos IRE e IRP</i> .....	5.24
5.3.3	- <i>Estudos Genéticos do Repertório da Célula T</i> .....	5.26

## 6 MÉTODOS DE ESTUDO EM GENÉTICA MOLECULAR.....6.1

6.1	MÉTODOS NÃO COMERCIAIS .....	6.3
6.1.1	<i>Preparação de DNA e RNA</i> .....	6.3
6.1.2	<i>ELECTROFORESE</i> .....	6.4
6.1.2.1	Géis de Agarose .....	6.5
6.1.2.2	Géis de acrilamida.....	6.5
6.1.3	<i>Reacção de polimerase em cadeia (PCR)</i> .....	6.6
6.1.3.1	Princípios do método .....	6.6
6.1.3.2	RT-PCR.....	6.8
6.1.4	<i>Southern Blot</i> .....	6.9
6.1.4.1	Princípios do método .....	6.9
6.1.4.2	Endonucleases de restrição .....	6.10
6.1.4.2.1	Funções biológicas dos sistemas R-M .....	6.10
6.1.4.2.2	Sistemas R-M tipo II.....	6.11
6.1.4.2.3	Sistemas R-M tipo IIs .....	6.12
6.1.4.2.4	Montar uma reacção de restrição .....	6.12
6.1.4.3	Métodos de marcação da sonda .....	6.14
6.1.4.3.1	RADIOISÓTOPOS .....	6.14
6.1.4.3.2	POLIMERASES DO DNA .....	6.16
6.1.5	<i>Dot-Blot</i> .....	6.16
6.2	MÉTODOS COMERCIAIS.....	6.17
6.2.1	<i>Amplicor</i> .....	6.18
6.2.2	<i>"Branched DNA"</i> .....	6.19
6.2.3	<i>NASBA/Nuclisens</i> .....	6.21
6.2.4	<i>Ligase Chain Reaction</i> .....	6.23
6.2.5	<i>DEIA</i> .....	6.25

# Capítulo 1

## 1 A genética Molecular da Célula não Patológica





## 1.1 Conceitos básicos de Genética Molecular

### 1.1.1 O material genético

#### 1.1.1.1 Ácidos Nucleicos (DNA e RNA)

Foi apenas em 1944 que Griffith demonstrou que a hereditariedade era transmitida pelos ácidos nucleicos. A experiência realizada demonstrou que a capacidade de matar um ratinho era conferida a uma estirpe bacteriana não virulenta, pelo DNA de uma outra estirpe bacteriana virulenta (fig.1.1).

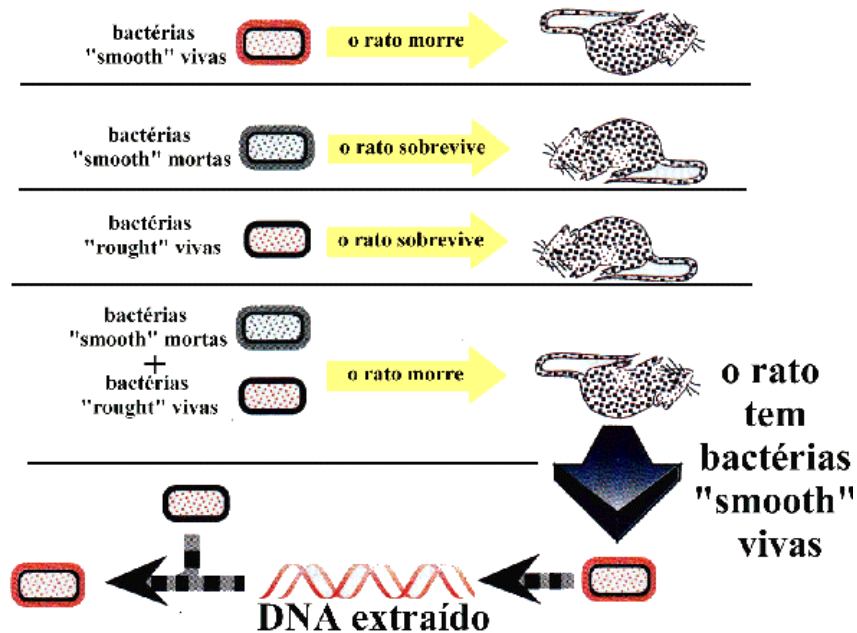


Fig. 1.1 – Experiência de Griffith demonstrando que a virulência bacteriana era uma característica genética transmitida pelos ácidos nucleicos.

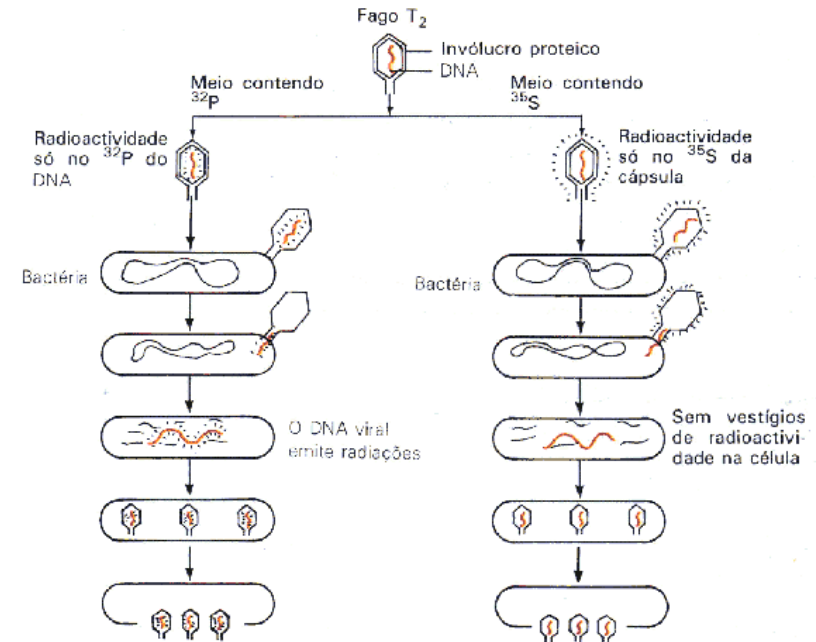


Figura 1.2 – Experiência de Hershey e Chase demonstrando que a transmissão genética nos fagos T2 estava associada ao DNA.

Em 1952, Hershey e Chase demonstraram que o fago T2 (um vírus que infecta bactérias) transmite seu código genético à bactéria infectada, através da injeção do seu DNA na bactéria, alargando assim o número de organismos que demonstradamente utilizam o DNA como registro genético (fig.1.2).

Sabemos hoje que com a exceção de alguns tipos de vírus, todos os organismos utilizam o DNA como portador da sua informação genética. Os vírus que fogem a esta regra utilizam o RNA para o mesmo efeito.

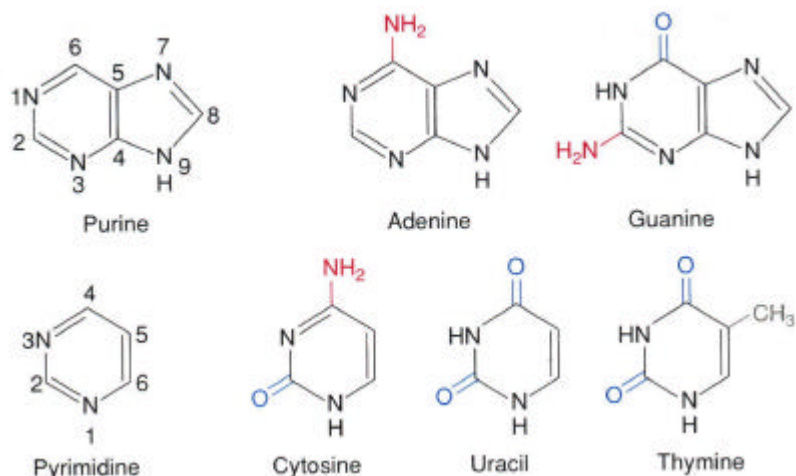


Figura 1.3 - As bases azotadas que entram na composição dos nucleótidos.

#### 1.1.1.1 A estrutura do DNA

Com a aceitação generalizada por volta dos anos 50 de que a informação genética residia no DNA, a grande questão passou a ser o mecanismo de armazenamento dessa informação. Com efeito, nesta altura não se compreendia como é que um polímero tão simples (apenas constituído por 4 tipos de unidades diferentes) e que se pensava ser homogêneo em toda a sua extensão podia codificar a enorme variedade de proteínas que compunham os organismos. Para tal, houve necessidade de elucidar de forma precisa a estrutura dos ácidos nucleicos.

Sabemos hoje que os ácidos nucleicos são polímeros de nucleótidos. Cada nucleótido contém um anel heterocíclico de carbono com 5 átomos de azoto (a base nitrogenada), 1 anel de 5 carbonos (uma pentose) e um grupo fosfato. As bases nitrogenadas são de 2 tipos: purinas e pirimidinas, sendo o número total de bases disponível de cinco (fig. 1.3). No entanto, cada tipo de ácido nucleico utiliza apenas

1.5

4 das cinco bases: o DNA contém Adeninas (A), Timidinas (T), Guaninas (G) e Citosinas (C); enquanto o RNA contém Adeninas (A), Uracilos (U), Guaninas (G) e Citosinas (C). As pentoses encontradas nos ácidos nucleicos são de 2 tipos: 2-desoxirriboses e riboses (Fig. 1.4) dando origem ao Ácido Desoxirribonucleico (DNA) e ao Ácido Ribonucleico (RNA).

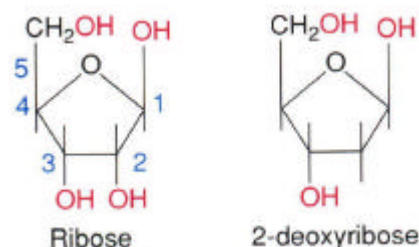


Figura 1.4 - As pentoses são o componente dos ácidos nucleicos que definem o seu tipo. A desoxirribose entra na composição do DNA, enquanto a ribose compõe o RNA

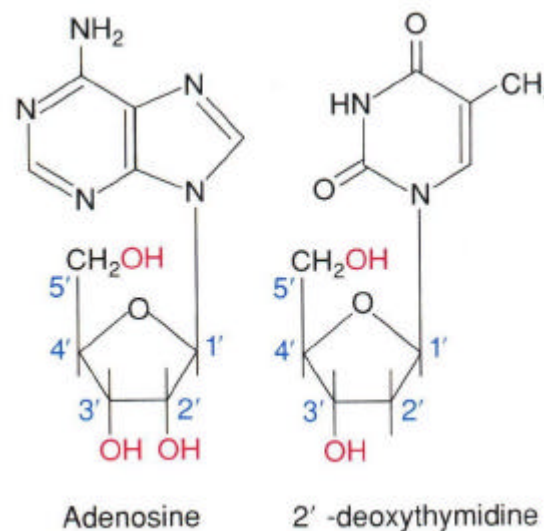


Figura 1.5 - Os nucleósidos são compostos por uma base e uma pentose.

1.6

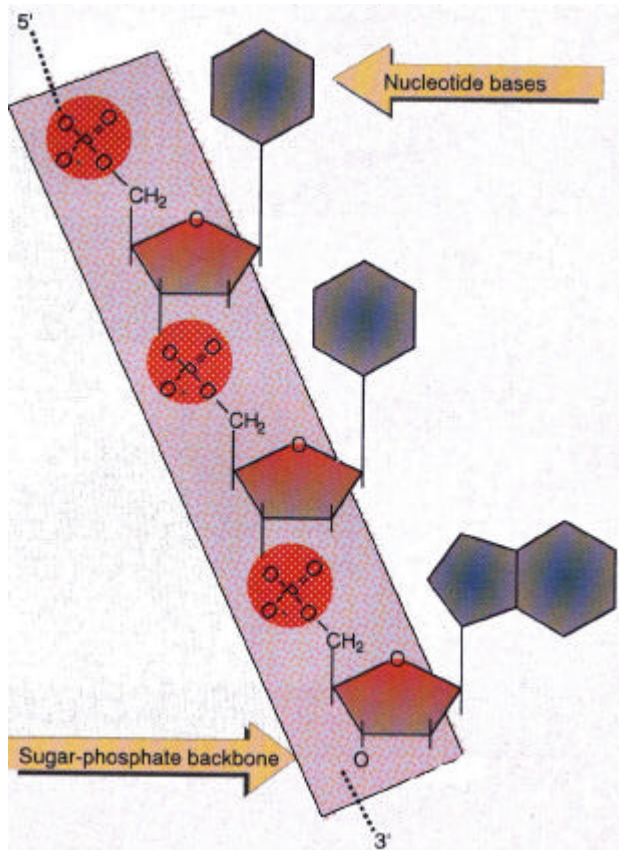


Figura 1.6 - Os nucleótidos unem-se por ligações fosfodiéster para dar origem aos ácidos nucleicos.

Os nucleósidos são compostos por uma base nitrogenada e uma pentose (fig. 1.5). Um nucleósido com um grupo fosfato no carbono 5 chama-se nucleótido. Os nucleótidos são as unidades de construção dos ácidos nucleicos, sendo unidas por uma ligação 5'-3': o carbono 5' da pentose de um nucleótido une-se ao carbono 3' da pentose do nucleótido seguinte por uma ponte fosfodiéster, ficando a base nitrogenada exterior ao esqueleto da ligação (Fig.1.6).

Em 1953, uma importante descoberta realizada por Watson & Crick transformou a visão do material genético. Dados de difracção de raios X mostraram que o DNA tem a forma de uma hélice regular. Das dimensões obtidas para a hélice na difracção de raios X, e da densidade do DNA, inferiu-se então que a hélice era composta por duas cadeias polinucleotídicas, com as bases de cada cadeia viradas para o interior da hélice. As bases de cada hélice emparelham de tal modo que uma purina se opõe sempre a uma pirimidina. Estes dados, conjugados com a observação anterior de Chargaff indicando que independentemente da quantidade de cada base, a proporção G:C e A:T é sempre a mesma no DNA, indicam que G emparelha com C e A com T na dupla hélice do DNA. Watson & Crick propuseram que o emparelhamento não se realizava por ligação covalente, mas por pontes de hidrogénio entre as bases nitrogenadas (Fig. 1.7). Para tal, as 2 cadeias devem orientar-se de modo antiparalelo (Fig. 1.7). Obteve-se assim para o DNA o modelo ilustrado na figura 1.8.

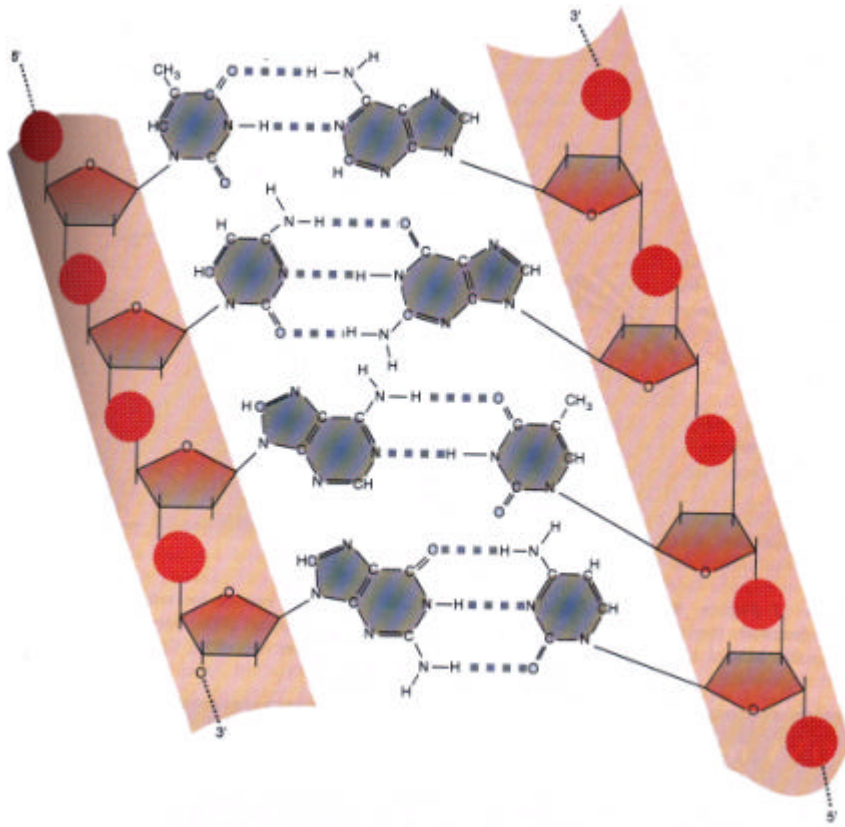


Figura 1.7 - O DNA é formado por duas cadeias com orientação antiparalela, com as bases de cada uma das cadeias a hibridizarem entre si.

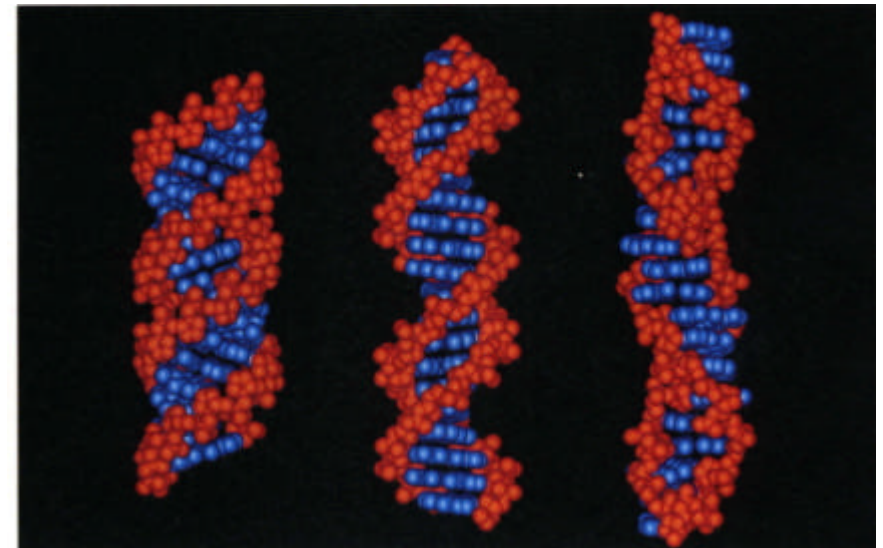


Figura 1.8 - Estruturas possíveis para a dupla hélice do DNA. O modelo da esquerda é o da forma A, o do meio o da forma B e do da direita o da forma Z.

A estrutura do DNA identificada por Watson & Crick, e ilustrada na figura 1.8B (forma B) é a que em situações fisiológicas é mais frequente. No entanto nem todo o DNA da célula se encontra nesta estrutura, e certamente *in vitro* é possível manipular as condições do meio, favorecendo outras conformações. Na tabela 1 encontram-se sumariadas as características das 4 conformações teoricamente possíveis para a conformação dos ácidos nucleicos, podendo na figura 1.8 ver-se comparativamente a conformação prevista.

Tabela 1.1 - Características dos tipos de hélice que o DNA pode tomar

Tipo de Hélice	N.º Bases por Volta	Rotação por par de Bases	Elevação por par de bases	Diâmetro da hélice
A	11	+32.7°	2.56 Å	23 Å
B	10	+36°	3.38 Å	19 Å
C	9.33	+38.6°	3.32 Å	19 Å
Z	12	-30°	3.71 Å	18 Å

A dimensão do material genético no Homem coloca o problema de como conseguir compactar 1,8m de DNA num núcleo que pode ser tão pequeno como 6µm ( $6 \times 10^{-6}m$ ). Este empacotamento tem ainda que ser flexível já que deve mudar ao longo do ciclo celular, aumentando durante as mitoses de tal modo que os cromossomas se tornam individualizados e visíveis ao microscópio óptico.

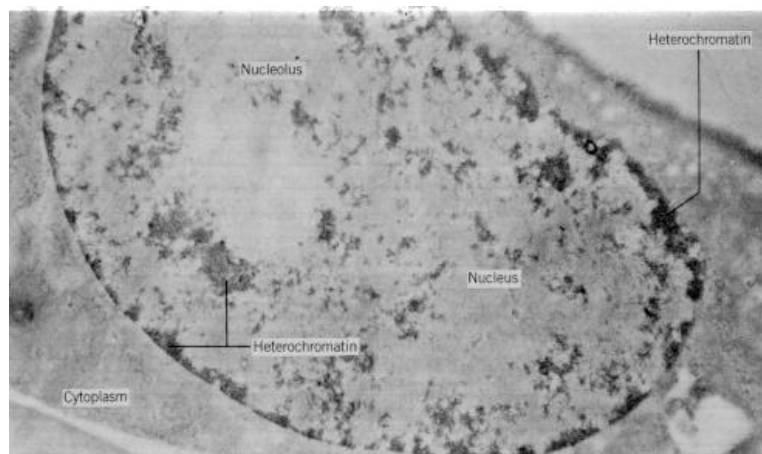


Figura 1.9 – Tipos de cromatina observáveis em interfase.

1.11

Durante a maior parte do ciclo celular a cromatina pode ser dividida em dois tipos de material genético (Fig. 1.9): a **euromatina** é a que ocupa a maior região do núcleo, sendo composta por material muito menos compactado que os cromossomas; a **heterocromatina** é composta por material muito compactado, formando fibras (encontra-se num estado intermédio entre a compactação dos cromossomas e a relativa descompactação da euromatina). A heterocromatina e a euromatina não representam fibras de DNA diferentes, já que as mesmas fibras passam pelas duas zonas do núcleo. Constituem assim partes das fibras com diferentes estados de condensação: a heterocromatina é constituída por regiões do DNA que não são habitualmente expressas na célula em causa, enquanto os genes expressos se localizam na euromatina (muito embora os genes na euromatina não estejam todos a ser expressos).

Em 1974, foi descoberta a estrutura básica de organização da cromatina em todos os eucariotas.

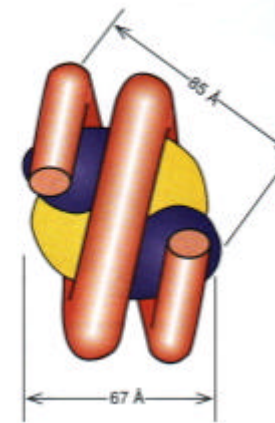


Figura 1.10 - A dupla hélice de DNA dá duas voltas ao núcleo central de proteínas do nucleosoma

1.12

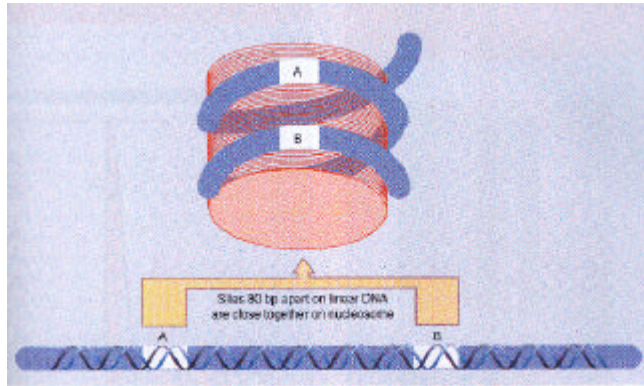


Figura 1.11 - A organização do DNA nos nucleosomas coloca próximas seqüências de DNA distantes na seqüência linear

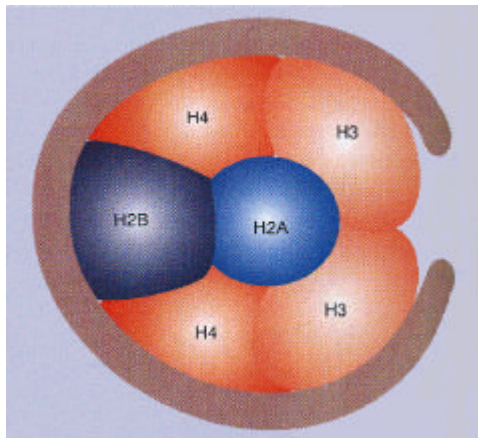


Figura 1.12 - Vista de cima da organização das histonas no nucleosoma. Existe ainda um par H2A/H2B por baixo, e por vezes uma histona H1 exterior ao nucleosoma

Esta subunidade organizativa básica (nucleosoma) contém cerca de 200 bp de DNA, organizados por um octâmero de proteínas pequenas e básicas (histonas) numa estrutura tipo rosário em que o DNA se localiza no exterior das “contas”, e as proteínas no seu interior (Fig.1.10). Esta organização explica porque os locais de ligação a proteínas se encontram por vezes tão espaçados na seqüência do DNA (Fig. 1.11). O octâmero de histonas é constituído por 2 cadeias de cada uma das histonas H1, H2A, H2B e H3, existindo ainda, por vezes, uma 5ª histona (H1) a estabilizar as 2 voltas de DNA ao octâmero (Fig. 1.12).

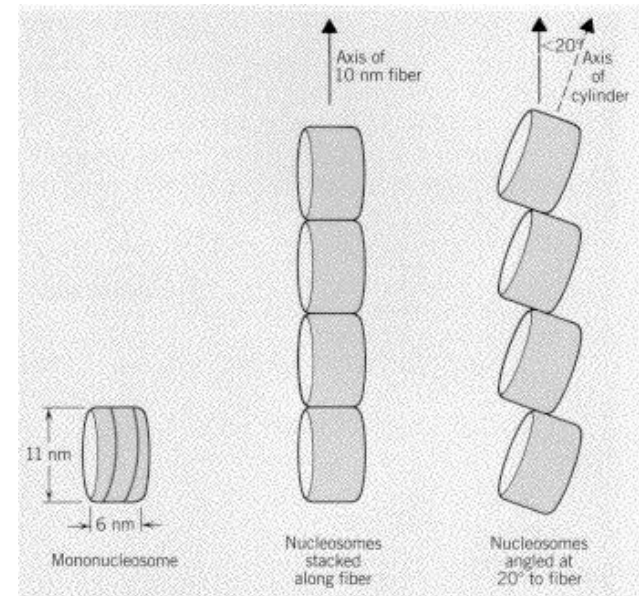


Figura 1.13 - A compactação das histonas na fibra de DNA de 10nm

A análise da cromatina ao microscópio electrónico revelou a existência de 2 tipos de fibras: a fibra de 10nm e a 30nm. A fibra de 10nm é essencialmente 1 sequência continua de nucleosomas (Fig. 1.13). Esta fibra ocorre em condições de baixa força iónica, e na ausência de histonas H1. Em condições de alta força iónica e na presença da histona H1, forma-se a fibra de 30nm, a qual é essencialmente um enrolamento de 6 nucleosomas por volta (Fig. 1.14).

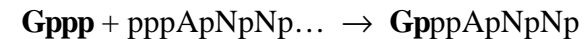
A transcrição (cópia dos genes em mRNA), como veremos no capítulo 1.1.2.1, envolve a deslocação no DNA de uma complexa maquinaria enzimática, e inclui a abertura da dupla cadeia do DNA. Este facto, não é compatível com um elevado grau de empacotamento das fibras do DNA, pelo que se compreende que os genes transcripcionalmente activos se localizem na eucromatina. No entanto, os resultados experimentais indicam que a estrutura dos genes transcripcionalmente activos envolve o empacotamento em nucleosomas, ainda que seja necessário admitir que durante a

transcrição estes sejam temporariamente “desmontados” pela maquinaria enzimática.

### 1.1.1.1.2 A estrutura do RNA

Na célula existem várias formas de RNA, as quais possuem estruturas e funções diferentes:

- **mRNA** - O mRNA ou RNA mensageiro, é a espécie de RNA que transporta a informação para a síntese das proteínas no ribossoma. O mRNA é formado no núcleo na transcrição do DNA, passando ainda por uma fase de processamento antes de atingir o citoplasma na forma madura (mRNA). O processamento efectuado inclui o “Splicing”, isto é a remoção das sequências não codificantes ou introns. Outras alterações introduzidas no processamento que ocorre no núcleo consistem na adição de uma cauda poli-adenina à extremidade 3’, e metilação CAP da extremidade 5’. A estrutura CAP resulta da ligação de um G à purina com que a transcrição habitualmente se inicia, ficando este G na orientação inversa, e ligado pelo trifosfato deixado livre pela purina:



G sofre então uma ou mais metilações.

- **tRNA** - o tRNA ou RNA de transporte é um tipo de RNA que se encontra covalentemente ligado a um aminoácido, tendo como função o transporte do aminoácido para o ribossoma, onde este vai ser posicionado com precisão, sempre que o ribossoma estiver a ler um codão complementar do tripleto que o tRNA possui (anticodão). As 64 espécies de tRNA (correspondentes aos 64 codões), possuem uma estrutura básica semelhante.

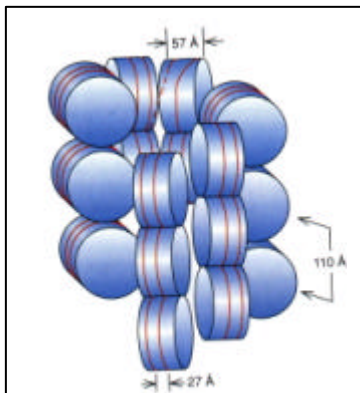


Figura 1.14 - A organização dos nucleosomas na fibra de DNA de 30nm.

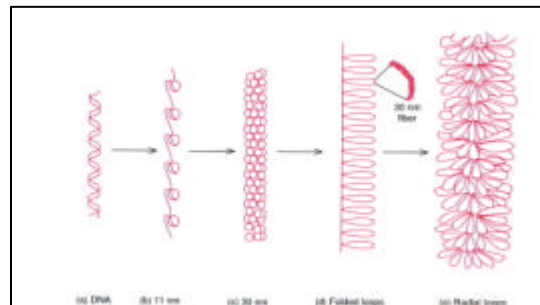


Figura 1.15 – hierarquia de condensação do DNA na cromatina.

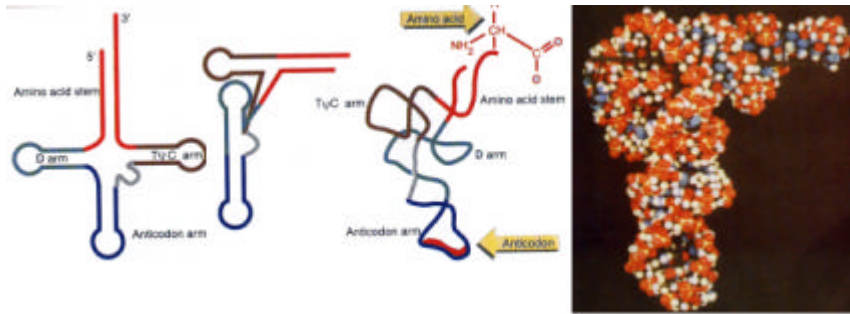


Figura 1.16 – Estrutura do tRNA

- **rRNA** - trata-se do RNA ribossomal, o qual é como o nome indica um dos componentes dos ribossomas. O rRNA constitui a maior parte da massa do ribossoma, e provavelmente todas as proteínas do ribossoma se associam ao rRNA. Assim, o rRNA forma como que o esqueleto do ribossoma, determinando a posição das várias subunidades proteicas.

#### 1.1.1.2 A ORGANIZAÇÃO DOS GENES NO GENOMA

O genoma pode, de uma forma genérica, ser classificado em DNA não repetitivo e DNA repetitivo. A abundância relativa dos dois tipos de DNA podem ser experimentalmente determinados, tendo por base a diferente cinética de re-hibridação (DNA repetitivo encontra mais rapidamente uma sequência complementar com quem pode hibridar). O DNA não repetitivo representa sequências únicas, ou seja genes de cópia única no genoma. O DNA repetitivo é constituído por DNA moderadamente repetitivo, representando genes com várias cópias no genoma, e DNA altamente repetitivo. A função do DNA altamente repetitivo permanece até ao momento uma incógnita. Como já vimos, um exemplo deste tipo de DNA é o que existe nos telómeros, onde provavelmente tem a função de estabilizar o cromossoma. Existem no entanto, repetições de pequenas unidades de sequências espalhadas pelo genoma (mini e microssatélites), os quais constituem em pequenas sequências, repetidas um determinado número de vezes. O número de repetições é em muitos casos altamente polimórfico, pelo

que estas sequências têm sido utilizados como marcadores no mapeamento genético.

**Sequências moderadamente repetitivas:** Nos genomas eucarióticos, os genes que existem em cópia única são poucos. Na maior parte dos casos, existem sequências com alguma similaridade, algumas das quais não funcionais (os pseudogenes). A vantagem da existência de mais que uma cópia dos genes é óbvia já que assim os organismos podem conservar uma cópia intacta do gene, mutando a outra, numa tentativa de evoluir. Neste processo de evolução, algumas cópias ficam com a sua funcionalidade comprometida, tornando-se **pseudogenes**. No entanto, uma vez que uma outra cópia funcional existe, nenhum efeito nefasto daí ocorre para o organismo.

Um conjunto de genes que descende por duplicação e variação de um gene ancestral é chamado de família génica. Os seus membros podem estar arrançados em grupos sequenciais (“gene clusters”), dispersos no genoma (muitas vezes mesmo em cromossomas diferentes), ou numa combinação de ambos os arranjos. Os “gene clusters” podem conter desde 2 até centenas de genes idênticos, alinhados em sequência. A dispersão dos genes ocorre por translocação de um gene após a duplicação. Os membros de um “gene cluster” têm função similar, mas podem ser expressos em tipos celulares diferentes ou em diferentes condições (Ex. Gene da globina). Em alguns casos, o gene cluster responde à grande necessidade de proteínas ou de RNA (ex.: rRNA e histonas).

**Sequências altamente repetitivas:** As sequências altamente repetitivas tomam a forma de sequências muito curtas, repetidas muitas vezes em sequência. Formam-se assim blocos de material genómico, consistindo cada bloco em longas repetições de uma unidade. Em alguns casos as unidades são rigorosamente iguais, noutros são relacionadas. A repetição sequencial de unidades de sequência forma blocos de DNA com características físicas distintas do resto do genoma, o que pode ser utilizado para as isolar. Uma das



propriedades físicas do DNA que depende da sequência é a densidade, a qual depende do conteúdo GC. A densidade é habitualmente determinada mediante a centrifugação do DNA num gradiente de Cloreto de Césio (CsCl). O DNA forma assim bandas correspondentes a sua própria densidade. Quando este procedimento é realizado para DNA genómico eucariota, forma um pico algo largo, consistindo numa mistura de sequências com densidades próximas (a banda principal). Por vezes forma-se ainda um ou mais picos adicionais, de menor intensidade. A este material chama-se o DNA satélite.

O DNA satélite existe no genoma de várias espécies eucariotas, pode ter uma densidade superior ou inferior à banda principal, mas representa habitualmente menos de 5% do genoma total.

O DNA satélite encontra-se frequentemente localizado na heterocromatina, não sendo habitualmente possível encontrar as suas sequências entre o RNA.

Nos mamíferos, as sequências que compõem cada satélite mostram divergência apreciável entre as repetições de cada. Habitualmente existem sequências curtas predominantes, mas outras relacionadas com estas, mas contendo adições, substituições e deleções formam o restante satélite. Frequentemente, pode observar-se uma hierarquia nas repetições dos satélites, com uma sequência base a repetir-se, a qual por vezes sofre modificações, as quais por sua vez se repetem também de forma mais ou menos cíclica. Este facto originou uma hierarquia de nomenclatura: DNA satélite, minisatélites, microsátélites.

### 1.1.1.3 Estrutura de um Gene

A comparação directa entre a sequência do DNA de um gene, e a sequência da proteína respectiva, permite determinar se o gene e a proteína são ou não colineares: se a sequência do gene corresponde exactamente a sequência de aminoácidos da proteína. Nas bactérias e

vírus, a equivalência é perfeita: cada gene contém uma sequência continua de nucleótidos, cuja sequência e comprimento está directamente relacionada com a da proteína. Quando falamos em correspondência entre o gene e a proteína, estamos no entanto a simplificar o que realmente se passa. Como veremos mais tarde, um gene não codifica directamente uma proteína, já que a informação tem que passar por um estado intermédio: o RNA. Assim, mesmo o mais simples dos genes tem que conter sequências de vários tipos:

- **sequências reguladoras ou não codificantes:** sequências que permitem à célula controlar que genes estão activos em cada momento, possibilitando assim uma resposta diferenciada dependente das necessidades de cada momento. As sequências reguladoras podem existir em cada extremidade do gene, e em alguns casos estar mesmo bastante distanciadas das sequências codificantes.
- **sequências codificantes:** sequências que são directamente transcritas para RNA, e deste codificadas em proteínas. Note-se que enquanto o DNA é de cadeia dupla, o RNA é de cadeia simples, pelo que apenas uma das cadeias do DNA pode ser idêntica a do RNA (codificante ou +), sendo a outra cadeia complementar do RNA (-).

Como acima foi dito, o gene não é no entanto tão simples nos eucariotas. Ao contrario das bactérias e vírus, nos organismos eucariotas, os genes e as proteínas não são colineares, isto é, a região codificante dos genes (exons) é interrompida a espaços irregulares por sequências não codificantes (introns). Este facto faz com que nos eucariotas, a expressão genica envolva um passo adicional: o splicing do RNA, ou processamento do RNA (com exons e introns) em mRNA.

Um promotor é uma sequência de DNA, habitualmente na extremidade 5' de um gene, com a função de se ligar a proteínas, e controlar a iniciação da transcrição. As proteínas a que um promotor se deve ligar, são várias, disso dependendo a sua dinâmica funcional. Genericamente pode falar-se de proteínas repressoras, proteínas activadoras, e da RNA polimerase. As proteínas repressoras, ao ligarem-se ao promotor impedem a ligação da RNA polimerase, impedindo assim o início da transcrição, enquanto a ligação das proteínas activadoras tem o efeito inverso. As propriedades do promotor que lhe conferem afinidade para as diversas proteínas dependem da sua sequência, pelo que esta varia de gene para gene, conferindo aos diversos genes características de regulação diferentes. No entanto, a ligação a polimerase do RNA é universalmente necessária, pelo que deve ser possível encontrar uma sequência "consenso" para os promotores. Esta sequência consenso consiste na sequência mínima comum entre os vários promotores, e deve incluir a sequência absolutamente necessária para a ligação a polimerase do RNA.

Para os procariotas foi possível definir a região 44-50bp "upstream" do ponto de iniciação até 20bp "downstream" com sendo a região que interacciona com a polimerase do RNA, tendo sido definida uma sequência consenso consistindo de vários padrões:

**Pribnow box ou sequência -10-** imediatamente upstream do ponto de iniciação (-18 a -12) existe uma região com a sequência T<sub>80</sub>A<sub>95</sub>T<sub>45</sub>A<sub>60</sub>A<sub>50</sub>T<sub>96</sub> (os números representam a frequência com que as bases ocorrem). A função desta sequência parece ser a de permitir que após a ligação da polimerase do RNA esta possa iniciar a sua evolução ao longo do gene, possivelmente por permitir a iniciação da abertura da cadeia do DNA (o facto de ter alto conteúdo AT facilita a abertura da dupla hélice).

**Sequência de reconhecimento ou Sequência -35 -** O seu nome deriva do facto de esta ser parte da sequência que a polimerase tem que reconhecer, mas que não fica fortemente ligada a esta. A

sequência consenso é: T<sub>82</sub>T<sub>84</sub>G<sub>78</sub>A<sub>65</sub>C<sub>54</sub>A<sub>45</sub>. A função desta região parece ser a de conferir a capacidade de ligação a polimerase do RNA.

No caso de organismos eucariotas, o estudo dos promotores é bem mais complexo, já que existem não uma RNA polimerase, mas três. A acrescentar a esta dificuldade, está o facto de não se conhecer com precisão todos os componentes da maquinaria de transcrição eucariota, pelo que os estudos *In viro* ficam comprometidos.

À partida 2 particularidades existem nos eucariotas, relativamente ao que se passa nos procariotas: 1) o promotor da polimerase III fica localizado downstream do gene; 2) não é possível conhecer as particularidades do promotor da polimerase I, já que esta transcreve apenas os genes dos rRNA os quais são todos idênticos.

No entanto o promotor da RNA polimerase II, a responsável pela transcrição da maioria dos genes nos eucariotas são conhecidos com alguma profundidade. As principais sequências consenso identificadas nos promotores da RNA polimerase II dos eucariotas são:

**TATA BOX** - sequência consenso:  $T_{82}A_{97}\frac{A_{63}}{T_{37}}A_{83}\frac{A_{50}}{T_{37}}$  Também

conhecida por Hogness box. Trata-se de uma sequência quase universalmente presente em mamíferos, aves, anfíbios e insectos. Posiciona-se a uma distância do ponto de iniciação entre 19 e 27bp. Como pode ver-se da sequência consenso, a TATA Box é constituída quase exclusivamente por AT, sendo as mutações que inserem um GC muito raras. Esta sequência é habitualmente rodeada por sequências ricas em GC, o que pode ser importante para a sua função.

**CAAT BOX** - sequência consenso  $GG\frac{T}{C}CAATCT$ . Esta sequência esta presente em alguns promotores, mas não em todos. A sua distancia ao ponto de iniciação ronda os 70 a 80bp.

As análises *In vitro* identificaram uma estrutura semelhante ao promotor bacteriano, imediatamente upstream do ponto de iniciação. No entanto, estudos *In vivo* revelaram a dependência de zonas ainda mais upstream da TATA box. Este componente pode consistir em duas regiões, uma entre -80 e -110 e a outra entre -50 e -70. Esta última pode ou não conter a CAAT box. Juntos, estas duas regiões têm uma forte influência na frequência de iniciação, possivelmente por influenciar a ligação da RNA polimerase II.

Junto ao ponto de iniciação, em redor da TATA box existe um componente que parece não ter influência na frequência de iniciação, antes de terminando o ponto de iniciação. Na ausência deste elemento, a transcrição tem uma iniciação errática.

Os promotores eucarióticos são bem mais complexos que os procariotas. Ao contrário dos promotores procarióticos, e contrariamente ao que até agora assumimos, um promotor eucariótico não funciona só. A sua actividade é enormemente aumentada de acordo com a regulação efectuada por outro tipo de sequências reguladoras: os “enhancers”. Estas sequências são distinguíveis dos promotores devido a duas características essenciais:

- a sua posição relativamente ao promotor é muito variável, podendo ser considerável, e funcionando em qualquer sentido (“upstream” ou “downstream”) e orientação.
- Um enhancer não actua apenas num promotor, podendo interactuar com qualquer promotor colocado na sua área de influência.

Vários vírus contêm enhancers. Destes, os mais perigosos para a célula que o vírus infecta são os enhancers presentes nos retrovírus. Como estes vírus se integram no genoma da célula infectada, a presença de enhancers pode levar à activação de um ou mais genes celulares que de outra forma estariam silenciosos. Desta forma, os retrovírus podem de forma indirecta, originar patologias, mesmo no seu estado “dormente”, já que mesmo na ausência de transcrição viral, podem induzir alterações no programa genético da célula infectada.

O modo de funcionamento dos enhancers permanece desconhecido. Foram no entanto colocadas várias possibilidades, entre as quais:

- **Formação de estrutura no DNA em cadeia Z**(ver figura 5). Os enhancers contêm habitualmente uma sequência alternada de pirimidinas-purinas. Esta sequência tem elevada probabilidade de formar uma estrutura em z-DNA. Se, por um lado, o modo como esta estrutura poderia afectar a transcrição não está esclarecido, por outro lado, este mecanismo poderia explicar porque os enhancers funcionam independentemente da sua orientação.
- Ligação do DNA a uma estrutura como a matriz nuclear
- ligação directa à polimerase

Quando a polimerase do RNA inicia a transcrição, esta prossegue com o complexo enzimático a percorrer o DNA, até que a enzima encontra um sinal para cessar a actividade. Neste ponto, a enzima pára de adicionar nucleótidos, liberta a cadeia de RNA nascente e dissocia-se do DNA. Assim, a terminação envolve a quebra de todas as pontes de hidrogénio entre o DNA e o RNA, e a reassociação da dupla hélice do DNA. A sequência de DNA que dá o sinal para que este processo ocorra chama-se terminador (ou abreviadamente t). Em alguns genes procarióticos, existem factores denominados anti-

terminadores, que permitem à polimerase continuar a transcrição passando por um terminador, num processo chamado de “read-through”). Assim, a terminação não constitui simplesmente uma forma de terminar a transcrição, mas também uma forma de controlar esta, já que a existência dos anti-terminadores pode determinar a transcrição ou não de determinados genes que se encontrem após o terminador.

Pouco se sabe dos terminadores dos genes eucarióticos. A principal dificuldade no estudo dos terminadores em eucarióticos é a incerteza quanto ao local de terminação da transcrição. Ainda que a maior parte das espécies de mRNA eucarióticas conhecidas possuam extremidades 3’ bem definidas, é muito difícil saber se esta extremidade foi produzida por terminação ou por processamento. No caso dos produtos da polimerase II, o problema é exacerbado pelo extenso processamento que ocorre com a adição da cauda poli-A. Pelo menos em alguns casos foi possível determinar que a extremidade 3’ observada no RNA é de facto originada por corte de uma cadeia de RNA mais longa.

Estudos efectuados com sequências de histonas (não poliadeniladas), permitiram verificar que o mRNA termina numa estrutura semicircular (“stem-loop”). Com efeito, mutações que impeçam a formação desta estrutura, impedem a terminação, enquanto que outras mutações que revertam a mesma estrutura, embora com uma sequência diferente, restauram a terminação. Assim, a estrutura parece mais importante que a sequência que a determina.

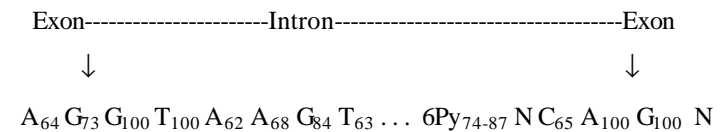
Os genes eucarióticos e procarióticos diferem numa característica essencial. Ao contrário dos genes procarióticos, os genes dos organismos eucarióticos não são contínuos, mas interrompidos. Significa isto, que no meio das sequências codificantes, surgem sequências que têm que ser retiradas do RNA, antes de este poder servir de molde à construção das proteínas. Este processo de transformação que o RNA sofre nos organismos eucarióticos é

chamado de processamento, ocorre no núcleo, e como veremos envolve não apenas a remoção das sequências extra (“splicing”) como outras transformações químicas.

Os genes eucarióticos são assim formados por dois tipos de sequências transcritas (isto é copiáveis para RNA) os exons e os introns (também chamados de intervening sequences). Os primeiros compõem as sequências que estarão presentes no RNA maduro, sendo os segundos as sequências que serão removidas durante o splicing.

A comparação das sequências nucleotídicas nas extremidades dos exons permite descrever as suas características:

- Não existe homologia extensa entre as duas extremidades de um intron, o que exclui a possibilidade da formação de uma estrutura secundária que determine os pontos de corte.
- As junções possuem uma sequência consenso conservada mas curta, a qual pode estar envolvida no processo de splicing:



### 1.1.1.4 O CÓDIGO GENÉTICO

A descoberta do código genético pretendeu responder à questão já por nós formulada (Cap. 1.1) sobre o mecanismo que permite aos ácidos nucleicos, com uma estrutura baseada em apenas quatro tipos de nucleótidos, conter a informação que codifica um enorme número de proteínas, as quais possuem 20 tipos de aminoácidos. A elucidação do código genético pretendeu ainda explicar como é que a expressão génica é regulada. No entanto, antes de esta questão poder ser estudada, era necessário estabelecer definitivamente a veracidade do dogma central da genética: Um gene - uma cadeia polipéptidica. Uma característica essencial do DNA é que a sua estrutura básica é independente da sequência (ao contrário das proteínas cuja conformação é directamente dependente da sequência). Assim, a sequência do DNA não parece ser importante devido à conformação, mas porque codifica uma sequência bem definida de aminoácidos. Note-se que o próprio conceito de que uma proteína contém sequências bem definidas de aminoácidos data dos anos 50 (a caracterização da insulina por Sanger), e portanto é estabelecida sensivelmente na mesma altura que se estuda a informação genética. A esta relação entre a sequência do DNA e a sequência proteica correspondente chamou-se código genético.

Como vimos, a sequência nucleotídica tem que conter informação suficiente para codificar aminoácidos diferentes. Como só há quatro tipos de nucleótidos no DNA, um cálculo simples indica que são necessários 3 nucleótidos (um triplete ou codão) para codificar um aminoácido. As combinações possíveis com três nucleótidos são  $4^3=64$ , pelo que o código genético é degenerado, isto é, mais do que um triplete deve codificar o mesmo aminoácido (tabela 1.2).

Tabela 1.2 - Código genético: significado dos 64 codons

		SEGUNDA BASE			
		U	C	A	G
PRIMEIRA BASE	U	UUU } <i>Phe</i> UUC } UUA } <i>Leu</i> UUG }	UCU } UCC } <i>Ser</i> UCA } UCG }	UAU } <i>Tyr</i> UAC } UAA } <i>STOP</i> UAG }	UGU } <i>Cys</i> UGC } UGA → <i>STOP</i> UGG → <i>Trp</i>
	C	CUU } CUC } <i>Leu</i> CUA } CUG }	CCU } CCC } <i>Pro</i> CCA } CCG }	CAU } <i>His</i> CAC } CAA } <i>Gln</i> CAG }	CGU } CGC } <i>Arg</i> CGA } CGG }
	A	AUU } AUC } <i>Ile</i> AUA } AUG → <i>Met</i>	AAU } AAC } <i>Thr</i> AAA } AAG }	AAU } <i>Asn</i> AAC } AAA } <i>Lys</i> AAG }	AGU } <i>Ser</i> AGC } AGA } <i>Arg</i> AGG }
	G	GUU } GUC } <i>Val</i> GUA } GUG }	GCU } GCC } <i>Ala</i> GCA } GCG }	GAU } <i>Asp</i> GAC } GAA } <i>Glu</i> GAG }	GGU } GGC } <i>Gly</i> GGA } GGG }

Podem agrupar-se os codões segundo o aminoácido que codificam (tabela 1.2). Quando tal é realizado, pode observar-se que com frequência, a base na 3ª posição não é significativa, porque os 4 codões com as mesmas 1ª e 2ª bases codificam o mesmo aminoácido (Tabela 1.2). Por vezes apenas distingue entre uma pirimidina e uma purina a 3ª posição. A esta especificidade reduzida na 3ª base chama-se degenerância da 3ª base. Esta característica, em conjunto com a tendência para aminoácidos semelhantes (isto é polares, hidrofóbicos, etc.) serem codificados por codões relacionados minimiza o efeito das mutações.

Três codões não codificam aminoácidos. Como se pode observar na tabela 2, estes codões (UUA, UAG e UGA) indicam o fim da sequência génica, sendo por isso chamados de codões stop.

O código genético foi inicialmente estudado na bactéria *E.Coli*, pelo que a universalidade deste necessitou de extenso estudo. Sabemos hoje, que genericamente o código genético é similar em todos os organismos vivos estudados. As exceções conhecidas são representadas por pequenas alterações em algumas espécies de microorganismos, e no código genético mitocondrial, o qual possui algumas particularidades em alguns organismos (Tabela 1.3).

Tabela 1.3 - Exemplos de exceções à universalidade do código genético

Organismo	Codon	Significado Provável na mitocôndria	Significado habitual
Todos	UGA	Triptofano	Terminação
Levedura	CUA	Treonina	Leucina
Mosca da fruta	AGA	Serina	Arginina
Mamíferos	AGA	Terminação	Arginina
	AUA	Metionina	Isoleucina

### 1.1.1.5 Mutações e Polimorfismos

Uma vez que o código genético é lido em tripletos não sobreponíveis, a inserção ou remoção de um nucleótido causa uma alteração na fase de leitura, alterando os codões subsequentes. Este tipo de mutação é denominado em Inglês “**frameshift**”. Mutações deste tipo são susceptíveis de reverterem através da mutação inversa, isto é, se a primeira mutação foi uma inserção e a segunda uma deleção, ou vice-versa, apenas a zona do gene situada entre as duas mutações se encontra mutada. A segunda mutação, é denominada supressora, já que suprime o efeito da primeira, limitando a zona atingida (fig. 1.17)

As **mutações pontuais** são mutações que ocorrem devido à substituição de um nucleótido por outro. A forma mais frequente de mutações pontuais é a **transição**, a qual ocorre quando uma pirimidina é substituída por outra, ou uma purina por outra. A **transversão** é menos frequente, e implica a substituição de uma pirimidina por uma purina, ou vice-versa. As mutações pontuais podem ser de 3 tipos, de acordo com o efeito que provocam no aminoácido codificado. Se não afectam o aminoácido codificado são chamadas **silenciosas**, se mudam o aminoácido codificado são chamadas **missense**, e se transformam o codão num codão stop são chamadas **nonsense**.

As mutações pontuais foram durante muito tempo consideradas as principais causas de mutações. Sabe-se no entanto hoje, que as deleções são também muito frequentes, representando uma significativa porção das mutações identificadas.

<b>Selvagem</b>	GCU	GCU	GCU	GCU	GCU	GCU	GCU	GCU	GCU	
	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	
<b>Inserção (+)</b>	GCU	GCU	AGC	UGC	UGC	UGC	UGC	UGC	UGC	U
	Ala	Ala	Ser	Cys	Cys	Cys	Cys	Cys	Cys	
<b>Deleção (-)</b>	GCU	GCU	GCU	GCU	GCU	_CUG	CUG	CU		
	Ala	Ala	Ala	Ala	Ala	Leu	Leu			
<b>Duplo mutante (+ -)</b>	GCU	GCU	AGC	UGC	UGC	_UCU	GCU	GCU		
	Ala	Ala	Ser	Cys	Cys	Ser	Ala	Ala		
<b>triplo mutante (+ + +)</b>	GCU	GAC	UGC	AUG	CUG	CAU	GCU	GCU	GCU	
	Ala	Asp	Cys	Met	Leu	His	Ala	Ala	Ala	
<b>triplo mutante (- - -)</b>	GCU	_CUG	CU_C	UGC	U_CU	GCU	GCU			
	Ala	Leu	Leu	Cys	Ser	Ala	Ala			

Figura 1.17 - Mutações frameshift e seus efeitos. Note-se que as inserções e as deleções podem anular-se mutuamente, fora da zona entre as duas mutações

As mutações podem ser vantajosas, desvantajosas ou neutras, segundo as consequências funcionais que provocam. As mutações neutras, apesar de causarem alteração na sequência não ocasionam

mudança funcional. Neste caso, deve falar-se em polimorfismo e não em mutação.

### 1.1.2 - A FISILOGIA DO GENE

#### 1.1.2.1 Transcrição e transcrição reversa

O RNA é a espécie de ácido nucleico com um papel mais alargado na genética molecular dos organismos. Não só o RNA tem o papel mais “mediático” de mensageiro, mas é também a espécie que assegura a descodificação da informação genética (o rRNA e o tRNA). Para além destes papéis centrais em todos os organismos, existem ainda vírus que utilizam o RNA como material de armazenamento de informação genética (os retrovírus). A produção do RNA tem habitualmente uma origem comum: a transcrição do DNA. No caso do mRNA, o produto formado é um intermediário cuja função requer ainda a tradução. No caso do tRNA e do rRNA, o produto formado é o efector da função a que se destina.

A transcrição é talvez o passo por excelência para a regulação

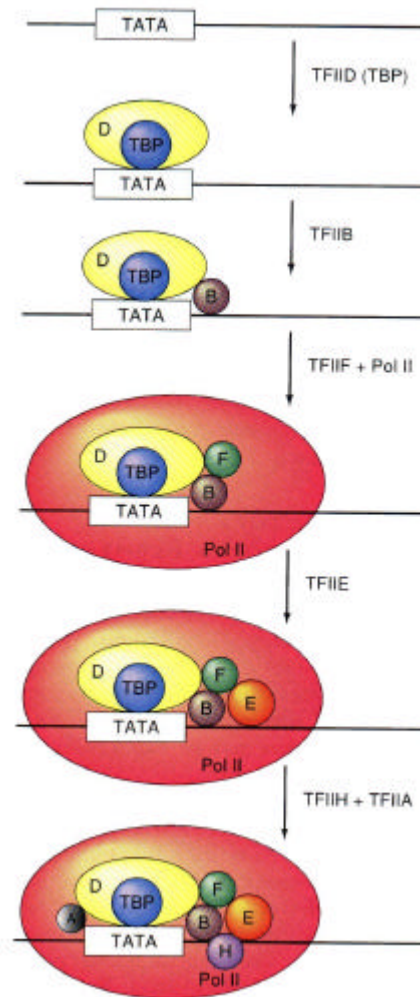


Figura 1.18 – Passos da iniciação da transcrição

da expressão génica. A decisão principal na regulação de um gene, é habitualmente a decisão de transcrever ou não esse mesmo gene. O que se traduz possivelmente numa necessidade de economia de energia e materiais por parte da célula.

A transcrição é catalisada pela RNA polimerase, e envolve a síntese de uma cadeia de RNA complementar da cadeia molde do DNA (a outra cadeia do DNA é a imagem do RNA, isto é a sua sequência é equivalente à do RNA, excepto no facto de em vez de possuir Uracilos possui Timidinas). A transcrição ocorre pelo processo habitual de emparelhamento de bases num processo altamente regulado e encadeado. Em primeiro lugar, a polimerase deve ligar-se ao DNA de cadeia dupla num processo complexo que envolve co-factores (fig. 1.18). Em seguida, as duas cadeias do DNA devem ser separadas (fig.1.19), para tornar a cadeia complementar acessível à maquinaria de transcrição. A abertura da hélice do DNA é um processo localizado, e à medida que a transcrição prossegue, novas zonas do DNA vão ficando acessíveis, enquanto as zonas já transcritas se vão emparelhando de novo, por forma a preservar a dupla hélice. A fase de **iniciação da transcrição** envolve assim, o reconhecimento do DNA pela polimerase, a abertura da hélice do DNA, e a incorporação do primeiro nucleótido na cadeia do RNA nascente. O local do gene onde se processa todo este processo é naturalmente o **promotor** (fig. 1.19). O local da incorporação do

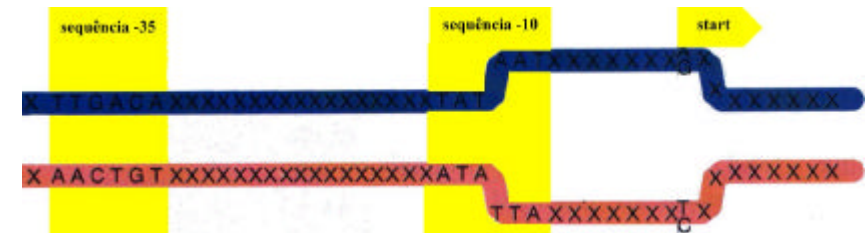


Figura 1.19 – Elementos reguladores da iniciação da transcrição eucariótica

primeiro nucleótido é designado “**start site**” ou “**startpoint**” (fig.1.19).

Depois da fase de iniciação inicia-se a **fase de alongação**, a qual produz a extensão da cadeia de RNA nascente, originando um híbrido de emparelhamento DNA-RNA. No entanto, à medida que a alongação se processa, a polimerase caminha para novas regiões do DNA, abrindo a hélice noutras zonas do gene, e fechando nas regiões já transcritas, o que implica o desemparelhamento DNA-RNA.

A **terminação** envolve o reconhecimento de um sinal indicando que não devem ser adicionados mais nucleótidos. Nesta fase, termina a ligação DNA-RNA da cadeia nascente, com libertação da polimerase e da molécula de RNA.

Desta descrição se pode inferir que a polimerase do RNA ( a enzima que catalisa a adição de nucleótidos à cadeia de RNA nascente) não funciona só, necessitando de um conjunto de outros componentes com funções essencialmente reguladoras e acessórias. Assim, quer a iniciação quer a abertura do DNA, quer a terminação são exemplos de processos em que intervêm outros factores para a progressão organizada e controlada da expressão génica. A maquinaria de transcrição das células eucarióticas é mais complexa e menos bem definida que a dos procariotas. Existem 3 polimerases nucleares, as quais ocupam diferentes locais do núcleo, e são cada qual composta por várias subunidades. Para complicar ainda mais o problema, existem ainda outras polimerases do RNA em mitocondrias e cloroplastos.

A maior parte da actividade de polimerase do RNA é realizada, nos eucariotas, pela RNA polimerase I, a qual se encontra no nucléolo, e é responsável pela transcrição dos genes codificando os rRNA (cerca de 50-70% do RNA total sintetizado). A segunda enzima, é a RNA polimerase II (20-40% da actividade total de síntese de RNA), e é responsável pela síntese do RNA heterogéneo (hnRNA), o precursor

do mRNA. A RNA Polimerase III é responsável pela restante actividade de produção de RNA (até 10% do total), tem localização nucleoplasmática e é responsável pela produção dos tRNA e muitos dos “small nuclear RNA” (snRNA).

Nenhum mecanismo de controlo da transcrição pode ser uma forma eficaz de controlar a expressão génica, se o produto da transcrição (o mRNA) não tivesse uma vida curta. Se assim não fosse, previsivelmente ocorreria uma acumulação de mensageiro, ou pelo menos o mensageiro formado permaneceria activo tanto tempo que não seria possível parar de sintetizar a respectiva proteína. Na realidade, a instabilidade do mRNA é muito acentuada. As duas formas de determinar a instabilidade do DNA baseiam-se ambas no bloquear da síntese *de novo* do mRNA (transcrição), medindo então a sua capacidade para servir na síntese proteica (semi-vida funcional), ou a sua capacidade para hibridar com uma sonda (semi-vida química). De modo geral, a semi-vida funcional é ligeiramente inferior à semi-vida química, o que sugere que pequenas degradações como um simples corte poderão ser suficientes para a inactivação biológica do mRNA. Verifica-se que este primeiro passo inicial é seguido da degradação do mRNA nos seus nucleótidos componentes, de forma mais ou menos sequencial na direcção 5'→3'.

O dogma central da genética molecular afirma que os genes são unidades que se perpetuam a si próprios, e que funcionam através da sua expressão em proteínas, através de um intermediário de RNA. Note-se que o dogma, na sua versão original define um paradigma que considera que a informação genética é transmitida unidirecionalmente: DNA→RNA→Proteína.

Hoje em dia, sabemos que a restrição do dogma central não é absoluta. Efectivamente, a informação genética pode ser transmitida de forma diferente da acima prevista. Alguns vírus de RNA, utilizam o RNA para a propagação da sua informação genética. Se esta pode parecer uma extensão relativamente pequena do dogma central, já a



existência nos retrovírus (vírus de RNA de cadeia simples que utilizam o DNA de cadeia dupla como intermediária na sua replicação) de transcriptases reversas constitui uma grande mudança no paradigma da genética molecular. As transcriptases reversas são enzimas que catalisam a síntese de um DNA de cadeia simples a partir de uma cadeia de RNA. Esta cadeia de DNA pode então ser utilizada para sintetizar DNA de cadeia dupla, utilizando a maquinaria habitual da célula, efectivamente revertendo um dos passos acima indicado: RNA→DNA. Este facto tem implicações profundas não só na forma de pensar a genética, mas também na biologia da infecção viral, já que este DNA de cadeia dupla formado, e que é uma cópia do RNA viral, vai agora integrar-se no genoma da célula, fazendo com que a infecção se propague de forma mais ou menos inofensiva à progenia da célula infectada ( a integração no genoma celular é uma parte normal do ciclo de vida do vírus sendo necessária à transcrição dos genes virais). Uma outra implicação deste mecanismo é a possibilidade de uma infecção de vírus deste tipo poder mediar a inserção de mRNA celular no genoma, como se de RNA viral se tratasse, originando duplicação génica, e/ou inserção de uma cópia do gene sob a acção de um promotor diferente, efectivamente alterando o programa genético da célula. Uma outra implicação da infecção por este tipo de vírus, foi já por nós abordada aquando da discussão da existência de enhancers, e constitui na possibilidade de colocar genes celulares sob a acção de enhancers virais, uma vez mais alterando o programa genético da célula infectada.

Os tipos de retrovírus de que existe mais informação disponível são os que originam as **partículas tipo Cem** aves e mamíferos. Estes vírus contêm duas cópias de RNA em cada virião. Assim, quando uma célula é infectada por dois viriões diferentes, podem-se originar viriões heterozigóticos, o que pode ser importante na aquisição de sequências celulares por parte do vírus, já que mesmo que em contrapartida perca algumas sequências do seu genoma, a restante

cópia do RNA viral permite-lhe continuar a ser capaz de efectuar uma infecção eficaz.

### 1.1.2.2 Tradução

A síntese proteica efectua-se no citoplasma, envolvendo uma complexa maquinaria genética centrada no ribossoma (fig. 1.20). Esta maquinaria genética pode ser vista como migrando ao longo do mRNA, lendo-o e utilizando a informação nele contida para alinhar com precisão cada aminoacil-tRNA, promovendo a ligação peptídica entre este e a cadeia peptídica nascente.

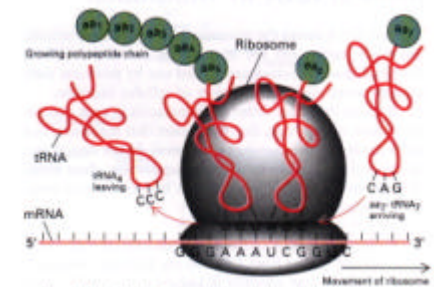


Figura 1.20 – O ribossoma interactuando com o mRNA, o tRNA e a cadeia polipeptídica nascente.

O ribossoma é assim um altamente elaborado e preciso complexo enzimático com diversificados componentes e vários centros activos, que requer vários cofactores para a sua actividade, e que obtém a energia química que necessita com a hidrólise de GTP.

Um ribossoma é composto por duas unidades (60S e 40S nos eucariotas) as quais, apesar de funcionarem em conjunto medeiam reacções diferentes na síntese proteica. O mRNA associa-se à subunidade menor, ficando associado a este por cerca de 30-40 nucleótidos. Apenas 2 moléculas de tRNA se podem associar ao ribossoma em cada momento, pelo que apenas 2 dos cerca de 30 codons associados ao ribossoma se encontram a ser processados em cada momento.

Cada tRNA liga-se ao ribossoma num local diferente deste, tendo cada um dos dois locais de ligação propriedades diferentes. Apenas o

Local A (local de entrada) pode receber um aminoacil-tRNA. Antes da entrada do aminoacil-tRNA, este local expõe o codon a ser decodificado. O último dos codons já decodificados encontra-se no local P (local dador), sendo este local ocupado pelo peptidil-tRNA (um tRNA contendo o aminoácido já covalentemente ligado por uma ligação peptídica à restante cadeia polipeptídica nascente). Quando estes locais (A e P) estão ambos ocupados ocorre a formação da ligação peptídica com transferência do polipéptido nascente para o tRNA do local A. O ribossoma desloca-se então no mRNA libertando o tRNA do local P e transferindo para este local o peptidil-tRNA do local A, e expondo um novo codon no local A.

A síntese proteica pode ser dividida em várias fases:

**Iniciação:** envolve as reacções que precedem a formação da ligação peptídica. Requer a ligação do ribossoma ao mRNA, a formação de um complexo de iniciação contendo o primeiro aminoacil-tRNA. Trata-se de um processo relativamente lento em comparação com as restantes fases da síntese proteica.

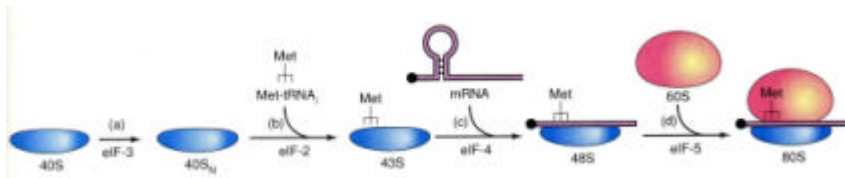


Figura 1.21 – Passos da fase de iniciação e respectivos co-factores envolvidos

Resumidamente pode dizer-se que nos eucariotas a iniciação começa com a ligação de GTP a um factor de iniciação denominado eIF-2 (eucariotic initiation factor 2). De seguida efectua-se a ligação de um N-formil-metionil-tRNA a este complexo. É o conjunto de factores assim formado que se liga então à subunidade 40S do ribossoma, a qual com o auxílio de outros factores de iniciação reconhece então a

extremidade 5' do mRNA (na qual se encontra a estrutura conhecida como CAP) por parte da subunidade 40S do ribossoma.

A subunidade 40S migra então no mRNA até encontrar um codon de iniciação. Neste ponto, liga-se a subunidade 60S, após a remoção de eIF-2 do complexo de iniciação.

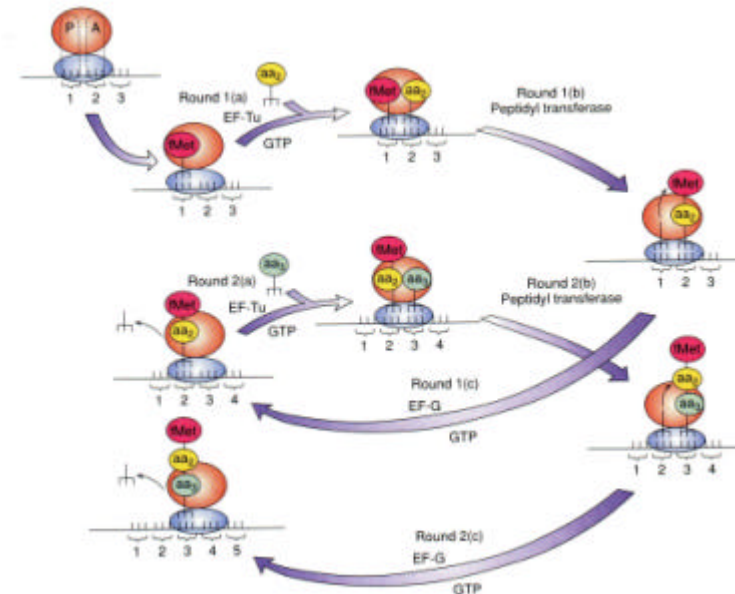


Figura 1.22 – Passos da fase de elongação da tradução eucariótica

**Elongação:** inclui todas as reacções desde a síntese da primeira ligação peptídica, até à adição do último aminoácido da cadeia polipeptídica. Os aminoácidos são adicionados um a um, naquele que constitui o processo mais rápido da síntese proteica.

Assim que a subunidade 60S se liga ao complexo de iniciação, o ribossoma fica pronto a iniciar a elongação. Para tal necessita de aminoacil-tRNA, o qual entra o local A, num processo mediado pelo factor eEF-1 (eucariotic elongation factor 1). Assim que o aminoacil-tRNA se encontra correctamente posicionado no local A, a peptidil transferase (uma função da subunidade 60S) catalisa a formação da ligação peptídica entre os aminoácidos dos locais P e A.

O último passo na elongação é a translocação, processo em que o ribossoma avança três nucleótidos de forma concertada( e que requer o factor adicional eEF-2). Este processo envolve a libertação do tRNA do

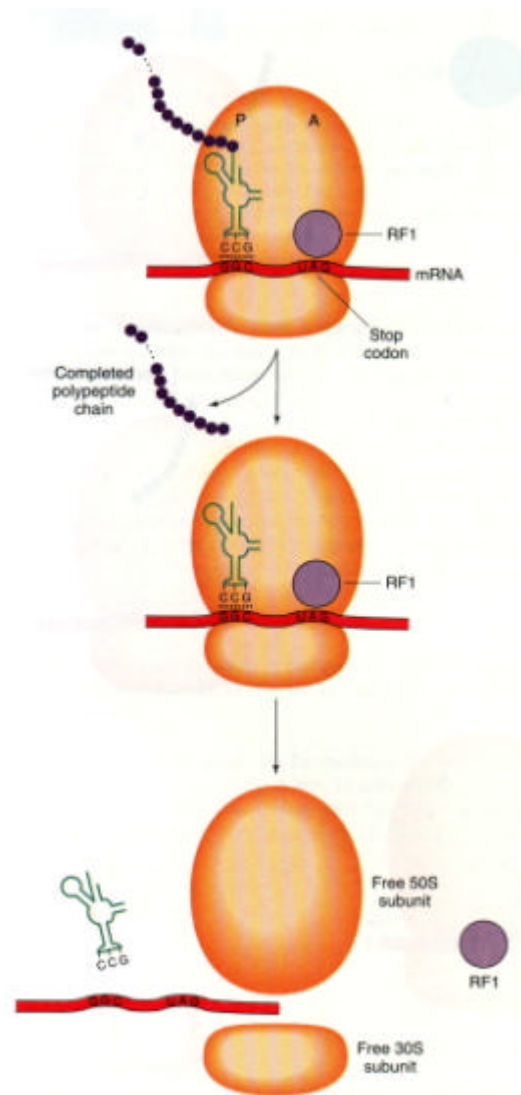


Figura 1.23 – Passos da fase de terminação da tradução eucariótica

local P, a passagem do peptidil-tRNA do local A para o local P, e a exposição do próximo codon no local A agora vazio.

**Terminação:** inclui todos os passos necessários para a libertação da cadeia polipeptídica formada, bem como a dissociação do ribossoma do mRNA. Este é um processo lento, em comparação com o tempo necessário para adicionar um aminoácido na fase de elongação.

Dos 64 tripletos, apenas 61 codificam para aminoácidos, sendo os restantes três codons stop, ou de terminação. Qualquer destes três codons (UAG, UAA e UGA) é suficiente para terminar a síntese proteica. Aos codons de terminação não corresponde nenhum tRNA, sendo estes reconhecidos directamente pelo factor proteico eRF (eucariotic release factor).

A reacção de terminação envolve a libertação do polipéptido do último tRNA, a expulsão do tRNA do ribossoma, e a dissociação deste do mRNA.

A célula eucariótica é uma estrutura finamente organizada, cujas funções são efectuadas em locais celulares definidos. A síntese proteica não constitui excepção, podendo os polirribossomas ser classificados em 2 tipos (livres e ligados a membranas), aos quais corresponde a síntese de diferentes grupos de proteínas .

Os polirribossomas livres sintetizam proteínas que não interagem com membranas, enquanto os que se encontram associados às membranas sintetizam proteínas cuja futura localização depende da sua capacidade para se ligarem às membranas. Note-se no entanto que a denominação polirribossomas livres não significa que estes se encontrem livres em solução no citoplasma. Estes polirribossomas encontram-se associados ao citoesqueleto para o que provavelmente dependem do mRNA.

Os polirribossomas tendem a estar localizados perto de núcleos, nos locais de entrada do mRNA no citoplasma. A maior parte das

proteínas sintetizadas são solúveis, e uma vez libertadas rapidamente difundem para longe do local de síntese. As proteínas que irão compor o citoesqueleto, tendem a integrar-se neste num local não muito distante do ponto de síntese.

As proteínas sintetizadas pelos ribossomas ligados a membranas têm vários destinos. Algumas são sequestradas em compartimentos celulares, outras são componentes membranares, e outras ainda são proteínas que se destinam a ser secretadas. Na maior parte dos casos das proteínas de membrana, a sua futura localização não depende da sequência da proteína madura, mas antes de uma sequência denominada “leader”, e que se localiza na zona terminal da cadeia polipeptídica nascente. Esta sequência, depois de ter determinado o destino da proteína será excisada do resto da proteína, originando a proteína madura.

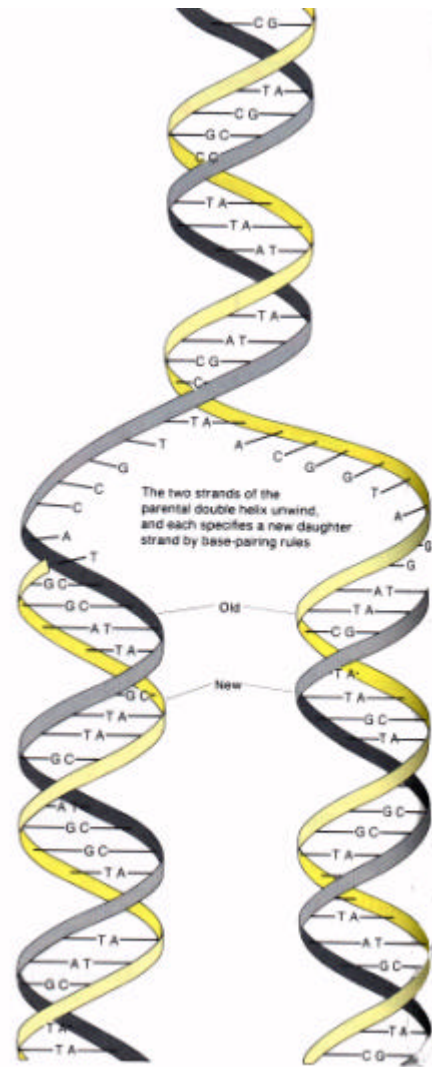


Figura 1.24 – A replicação é semiconservativa, tendo como base o emparelhamento de nucleótidos

### 1.1.2.3 Replicação

A descoberta da natureza de dupla hélice do DNA foi o ponto de partida para a compreensão do mecanismo de duplicação deste ácido nucleico, o qual é designado por replicação. Com efeito, o princípio básico do emparelhamento dos nucleótidos, em que assenta a estrutura do DNA é também o princípio essencial para a compreensão da replicação do DNA, já que esta assenta na noção de que cada cadeia de nucleótidos funciona como padrão na síntese da uma nova cadeia (fig. 1.24). Assim, a replicação do DNA é semiconservativa, isto é, após a duplicação, cada molécula de DNA tem uma cadeia original, e uma cadeia sintetizada de novo (fig. 1.24).

Como a síntese se processa sempre no mesmo sentido ( $5' \rightarrow 3'$ ) e as cadeias de DNA têm orientação invertida, numa das cadeias, a síntese é contínua, sendo na cadeia complementar descontínua, a partir de vários pontos de iniciação. Forma-se assim, em cada ponto de replicação uma estrutura em forma de Y (“Forks”). Como a replicação é bidireccional (excepto nos vírus de DNA linear), cada ponto de iniciação da replicação dá origem a dois “Forks” invertidos ( $\leftarrow \text{C} \rightarrow$ ). Para que a mesma polimerase possa sintetizar ambas as cadeias, forma-se uma estrutura local em “loop” na cadeia de síntese descontínua (fig. 1.26).

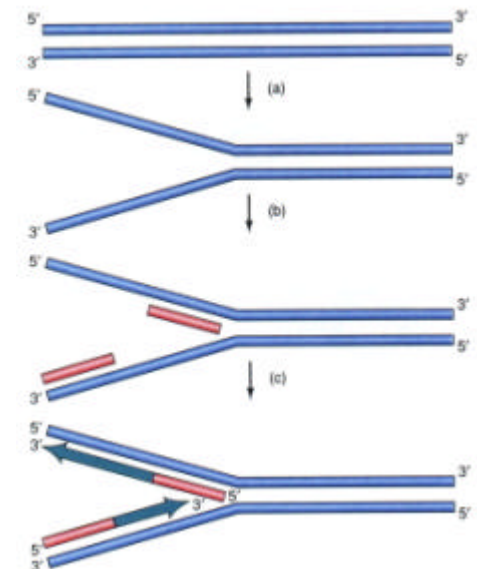


Figura 1.25 – A replicação é contínua numa cadeia (cadeia “leading”) e descontínua na outra (Cadeira “lagging”).

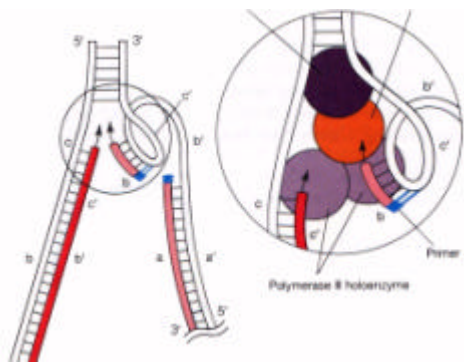


Figura 1.26 – Estrutura em loop que permite que uma polimerase sintetize as duas cadeias de DNA.

Em bactérias de DNA circular, habitualmente apenas um ponto de iniciação é suficiente para replicar todo o genoma, mas em eucariotas, cada cromossoma possui múltiplos pontos de iniciação, que cooperam para a rápida replicação

do DNA. Estes pontos possuem uma sequência semelhante entre si. Conhecem-se 3 tipos principais destas sequências: a OriC da *E.coli* a sequência do vírus SV40, e a sequência de replicação autónoma em leveduras. A OriC (bem como a maior parte das sequências equivalentes em bactérias) é composta por repetições de sequências com 9 e 13 pares de bases (9-mers e 13-mers) ao longo de uma sequência com cerca de 240bp, sendo estes 9-mers e 13-mers locais de ligação da proteína DnaA, a qual inicia a replicação. A origem de replicação do SV40 possui cerca de 65bp. Os 17 cromossomas da levedura *S.cerevisiae* possuem mais de 400 sequências de iniciação da replicação. Cada uma destas sequências (denominadas ARS de *autonomously replicating sequence*) confere a um plasmídeo a capacidade de se dividir na levedura. Estudos mutagénicos efectuados ao ARS1 (cerca de 180bp) revelaram a existência de apenas uma região essencial, com cerca de 15bp, designada elemento A. Três outros elementos (B<sub>1</sub>, B<sub>2</sub> e B<sub>3</sub>) são necessários para o funcionamento efectivo do ARS1. Um complexo de proteínas chamado ORC (“*origin recognition complex*”) liga-se ao ARS1 de modo dependente do ATP. Adicionalmente a estas sequências, uma sequência a elas adjacente, rica em Adenosinas e Timidinas facilita a abertura da dupla hélice.

A replicação eucariótica (aqui descrita para o SV40 – o modelo mais

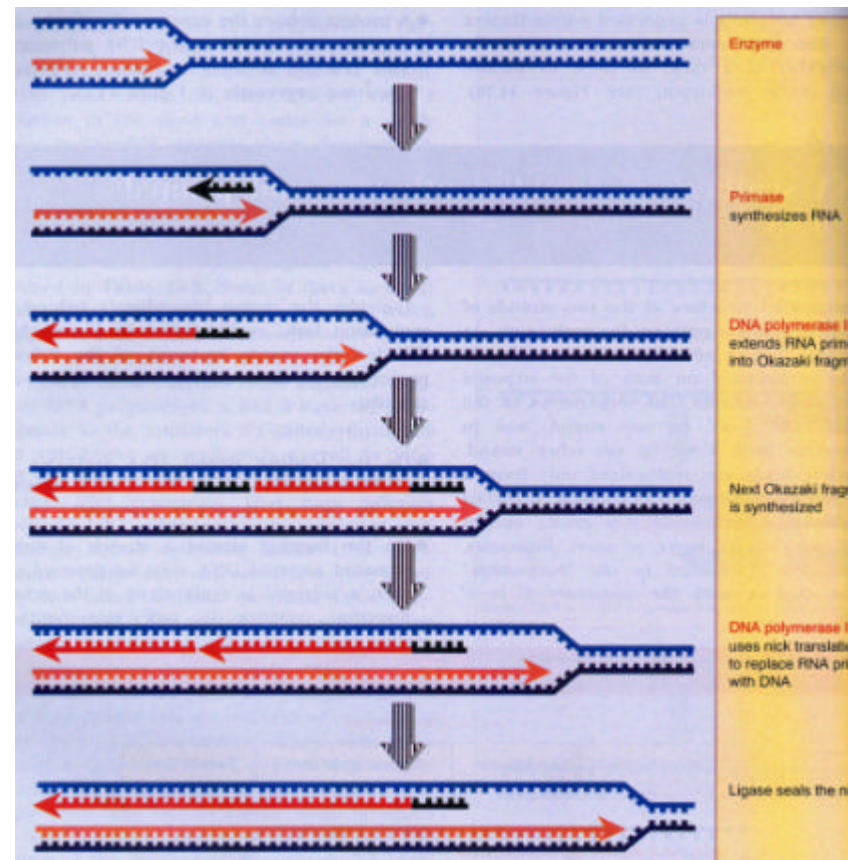


Figura 1.27 – A síntese descontínua da cadeia "lagging" exige primers de RNA sintetizados pela primase, formando-se temporariamente uma molécula mista RNA/DNA (fragmento de okasaki).

bem conhecido) inicia-se com a ligação de uma proteína viral chamada T antigen ao DNA. Esta proteína multifuncional desenrola o DNA com a sua actividade de helicase, com a ajuda de ATP e um factor denominado RFA (“*replicating factor A*”), o qual é uma

proteína que se liga a DNA de cadeia simples. A replicação inicia-se então bidirecionalmente a partir de primers de RNA produzidos por uma primase.

Duas polimerases distintas ( $\alpha$  e  $\delta$ ) trabalham no “fork” eucariótico: a pol  $\delta$  produz a cadeia “*leading*”, cuja síntese é contínua; a síntese da cadeia “*lagging*”, realizada de modo descontínuo é da responsabilidade da pol  $\alpha$ , a qual se encontra fortemente associada a uma primase. Na realidade, a síntese da cadeia “*leading*” é na realidade o trabalho sequencial das polimerases  $\alpha$  e  $\delta$ . O complexo

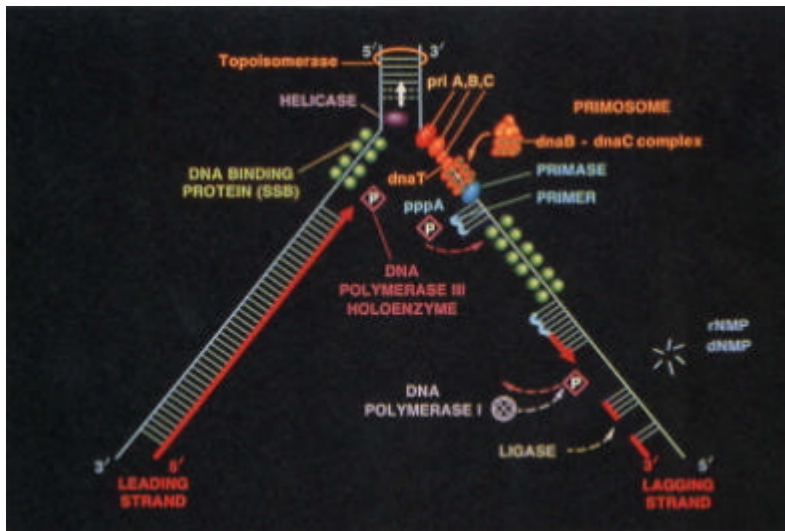


Figura 1.28 – Representação esquemática dos vários componentes que intervêm na replicação eucariótica.

primase-pol  $\alpha$  sintetiza o primer de RNA, e inicia a extensão do DNA, numa actividade estimulada pelo factor de replicação C. A ligação de PCNA (“*proliferating cell nuclear antigen*”) liberta a pol  $\alpha$ . A pol  $\delta$  liga-se então ao PCNA que activa a polimerase a qual sintetiza então a cadeia “*leading*”. Finalmente enzimas chamadas

topoisomerasas executam o importante trabalho de libertar as tensões de torção introduzidas pelo desenrolar da hélice do DNA padrão.

No final de cadeias de DNA linear, existe o risco de em cada síntese, a cadeia “*lagging*” se tornar progressivamente mais curta, pois a síntese descontínua necessita sempre de um “*template*” maior que a sequência a amplificar. Para obviar a este risco, a célula possui nas extremidades do cromossoma estruturas especiais denominadas telómeros e compostas por repetições de sequências oligoméricas (ex. na levedura é: 5’-G<sub>1-3</sub>T-3’). Os telómeros são sintetizados por um processo diferente do acima descrito por enzimas especializadas: as telomerasas. Estas enzimas são transcriptases reversas modificadas, produzindo DNA a partir de um “*template*” de RNA que a elas se encontra associado, e que determina portanto a sequência do telómero.

Finalmente um outro aspecto da síntese do DNA deve ser alvo da nossa atenção: a reparação de erros. As polimerases, não são enzimas isentas de erros, e a incorporação de nucleótidos incorrectos por vezes ocorre. Vários mecanismos de reparação destes erros existem, e deles depende a transmissão do património genético, sem mutações. O primeiro mecanismo de reparação do DNA está incorporado nas próprias polimerases. Estas enzimas, não só sintetizam DNA (actividade polimerase 5’→3’) como exercem uma actividade de verificação (“*proofreading*”) das bases recentemente adicionadas, removendo as bases incorrectas (actividade de exonuclease 3’→5’).

#### 1.1.2.4 Mecanismos de Mutação do DNA

##### 1.1.2.4.1 Erros na replicação do DNA

Quando, durante a replicação do DNA a polimerase incorpora nucleótidos errados, produz-se uma mutação. Na maior parte dos casos, tal não se verifica, apenas porque a maioria das polimerases, possuem não apenas actividade de síntese na direcção 5’→3’, mas também actividade de *proofreading* na direcção 3’→5’. Assim, se o

nucleótido incorporado for o errado, a polimerase usa a sua actividade exonucleolítica, e volta a tentar.

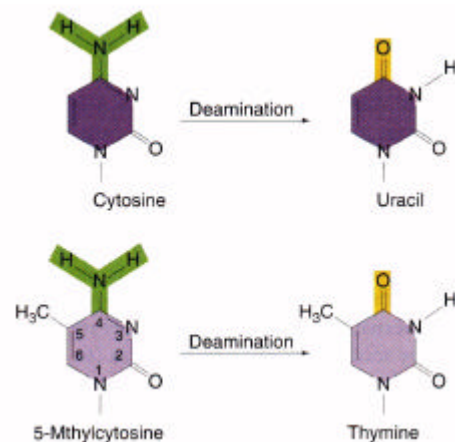
Cada base do DNA existe em uma de várias formas possíveis, chamadas tautómeros. Trata-se de isómeros, em que a posição relativa e as ligações entre os átomos variam. Normalmente apenas a forma ceto aparece no DNA, enquanto as formas enol e imino são raras. No entanto, os tautómeros enol e imino têm a capacidade de emparelhar com bases diferentes das que a forma ceto emparelha, levando a erros na replicação. Assim, por exemplo a forma imino da citosina pode emparelhar com a adenina, a forma enol da timina pode emparelhar com a guanina, a forma imino da adenina pode emparelhar com a citosina e a forma enol da guanina pode emparelhar com a timina.

Uma outra forma de incorporação estável de bases erradas ocorre com a replicação de bases ionizadas, a qual é possivelmente mais frequente do que a mediada pelos tautómeros, e justifica os efeitos mutagênicos das radiações ionizantes.

Os emparelhamentos de purinas com purinas e pirimidinas com pirimidinas, ocasionam distorções na dupla hélice, já que forçam alterações na dimensão exterior da hélice. No entanto, mesmo este tipo de mutações (transversões) ocorrem, tendo inclusivamente sido observado por difracção de raios X estruturas de DNA albergando este tipo de emparelhamentos.

Também mutações frameshift podem originar-se por erros de replicação. este tipo de erros ocorre preferencialmente em zonas de repetição de nucleótidos (ex: AAAAAA → AAAA; CTCTCT → CTCT), e pensa-se ser o resultado de um “deslizar” excessivo da polimerase, sem que este provoque uma desestabilização do emparelhamento imediatamente adjacente ao local de polimerização. Com o mesmo mecanismo se pode explicar

1.47



mecanismos de reparação do DNA, para a manutenção do código genético, garantindo a viabilidade celular e a da prole. Em certas condições, uma base pode ser adicionada ao nucleótido apurínico (como um último recurso de reparação), existindo neste processo 3 hipóteses em 4 de o resultado se traduzir numa mutação.

A **desaminação** de uma citosina origina um uracilo. Se a base assim alterada não for reparada, numa replicação posterior ela vai emparelhar com uma adenina, originando a substituição de um C por um G. Note-se que se o nucleótido onde ocorreu a desaminação possuir uma 5-metil-citosina, a base desaminada é uma 5-metil-uracilo, ou seja timina, a qual não é reconhecida pela maquinaria de recombinação, resultando numa mutação não reparável. Por este motivo as 5-metil-uracilo são “*hot spots*” de recombinação.

Uma forma menos frequente de mutação espontânea são os danos oxidativos. Espécies de oxigênio altamente reactivas como os radicais superóxido ( $O_2^{\bullet}$ ), peróxido de hidrogénio ( $H_2O_2$ ) e o seu derivado radical hidróxilo ( $OH^{\bullet}$ ) são produtos secundários do metabolismo aeróbico, e podem causar danos oxidativos de bases nucleotídicas. Por exemplo a 7-hydroxymethylguanine (GO) frequentemente emparelha com a timina, o que resulta num elevado número de mutações G → T.

1.48

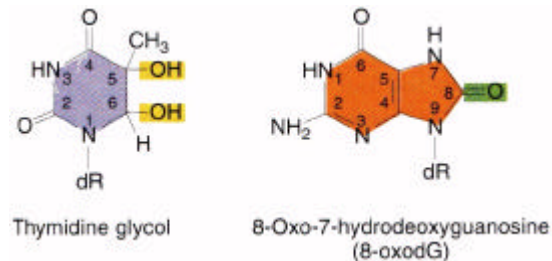


Figura 1.30 – Formação do 8-oxidG (GO)

### 1.1.2.4.3 **Recombinação genética**

#### 1.1.2.4.3.1 Recombinação por crossing-over

O crossing-over é um mecanismo normal de troca de material genético entre dois cromossomas homólogos, durante a meiose. Trata-se de um fenómeno complexo em que habitualmente é conseguida uma extraordinária precisão, mas de onde podem surgir mutações quando algo de errado acontece. Existem hoje vários modelos de recombinação por crossing-over:

#### Recombinação homóloga Geral

A recombinação homóloga geral não envolve nenhuma especificidade em termos de sequência do DNA, para além de se verificar apenas entre moléculas de DNA homólogo. Esta recombinação envolve a formação de uma estrutura em que uma molécula de DNA (dupla hélice) emparelha parcialmente com outras

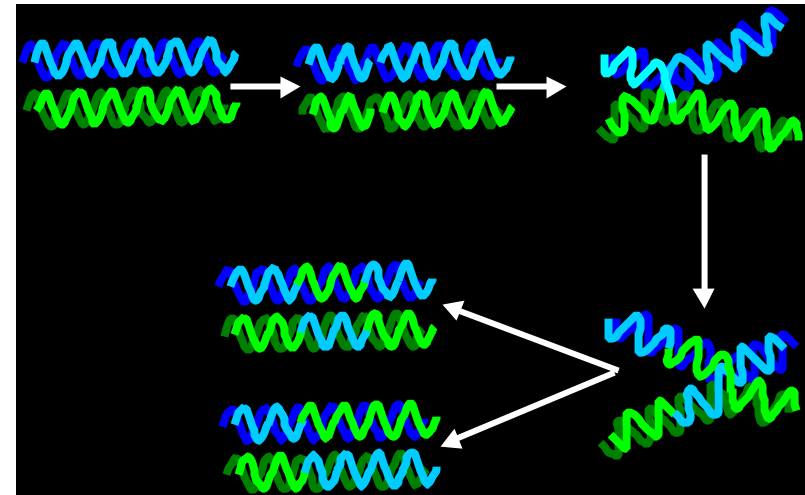


Figura 1.31 - Modelo de Holliday para a recombinação homóloga

homóloga, formando uma estrutura em quiasma. Note-se que de cada cromossoma, apenas um dos dois cromátídeos (ou uma das duas moléculas de DNA de cadeia dupla) toma parte na recombinação. O local de corte e junção ocorre então no centro do quiasma, resultando dois cromossomas híbridos, mas intactos em termos do seu conteúdo genético.

O modelo mais plausível para explicar os passos necessários para a recombinação foi apresentado por Robin Holliday (fig.1.31) para explicar o fenómeno de conversão génica. Como se pode ver na fig. 1.31, o modelo de Holliday prevê que num dos passo iniciais se verifique um corte por uma endonuclease numa das cadeias de cada molécula de DNA, a que se segue a troca de extremidades de uma das



cadeias de cada dupla hélice. As cadeias são então ligadas pela ligase formando-se cadeias híbridas. A este passo segue-se a migração do ponto de troca, terminando o processo com a resolução, a qual pode ocorrer por um de dois mecanismos (Fig. 1.31): 1) cortando e ligando mantendo as cadeias originais no resto do DNA ou; 2) cortando e ligando as cadeias novas, efectivamente trocando todo o material cromossómico deste o ponto de cross-over até ao telómero.

Este modelo, se bem que explique os resultados obtidos com fungos, não permite explicar todas as observações efectuadas com leveduras, o que levou Meselson & Radding a propor um mecanismo diferente. Este mecanismo, parte não de um corte de uma cadeia em cada molécula de DNA, mas apenas numa delas, seguindo-se a polimerização a partir de um dos terminais. Esta polimerização provoca o desemparelhamento de parte da cadeia em síntese, originando uma molécula de DNA parcialmente de cadeia simples, a qual pode então substituir a sua homóloga no outro cromossoma. Os últimos passos neste modelo consistem na ligação das cadeias com terminais, seguida da migração do ponto de junção, e resolução horizontal ou vertical. Note-se que este modelo explica a formação de produtos não passíveis de serem produzidos pelo modelo de Holliday, como é o caso de cromossomas irmãos que numa divisão possuem trocas cromossómicas desiguais entre si. Um outro tipo de implicação deste modelo é a possibilidade de ocorrência de conversão de cromatídeos e de conversão de meios-cromatídeos. Estes fenómenos ocorrem quando se formam cromatídeos com cadeias de DNA não homólogas, isto é quando, por crossing-over, o produto final é uma molécula de DNA com cadeias pontualmente não emparelhadas (locais G:T, G:A, C:T, C:A). Quando isto sucede, os mecanismos de correcção do DNA podem corrigir o mal-emparelhamento, do que resulta uma molécula de DNA convertida (conversão de cromatídeos), ou igual à molécula inicial. Pode ainda não ter ocorrido correcção, e a molécula de DNA possuir o desemparelhamento local (conversão de meio-cromatídeo). Assim se obtêm as proporções

1.51

menos frequentes, e não iguais no número de cromossomas resultantes de um crossing-over.

#### 1.1.2.4.3.2 *Recombinação somática*

Trata-se de um processo enzimático complexo e altamente regulado, que permite a uma célula T ou B adquirir um gene específico para o TCR ou para uma imunoglobulina respectivamente.

O mecanismo de recombinação (ou rearranjo) somática (descrito em pormenor na secção 2.1) é habitualmente restrito às células T e B, e apenas aos genes do TCR e Ig. No entanto, a existência de sequências de recombinação ditas crípticas (sequências de recombinação imperfeitas) pode originar a que em casos raros, nestas células, os genes do TCR ou Ig sejam refranjados indevidamente com outros loci genéticos, do que resulta uma translocação intra ou intercromossómica com perturbação do programa genético. O programa genético da célula em que ocorrem estas anomalias pode ser perturbado de uma de duas formas: 1) a formação de uma proteína quimera pode desregular a função dos genes em questão; 2) a colocação de um gene habitualmente silencioso nos linfócitos debaixo da acção do enhancer do TCR ou das Ig, faz com que este seja expresso de forma permanente. Exemplos de anomalias destes tipos, e suas consequências clínicas serão apresentadas na secção 4.2.

#### 1.1.2.4.4 *Mutações mediadas por elementos de transposição*

Os elementos de transposição são elementos genéticos com a propriedade de serem capazes de se mobilizarem, e moverem-se para outro local do genoma, ou de criarem cópias de si próprios noutra local do

1.52

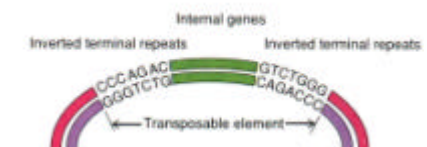


Figura 1.32 – Elemento de transposição simples ou sequência de inserção (IS do inglês *insertion sequences*).

genoma. Este conceito, criado por resultados experimentais produzidos por Marcus Rhoades e Barbara McClintock em milho, foram inicialmente encarados com grande ceticismo. Sabe-se no entanto hoje que estes elementos são bastante representados em todo o mundo biológico, dos Procariotas aos Eucariotas superiores.

As seqüências de inserção (fig.1.32) são os elementos de transposição mais simples, sendo constituídos por seqüências de nucleótidos que se podem agrupar em famílias, de acordo com a sua seqüência. Assim, a IS1 foi historicamente a primeira a ser descrita, possuindo cerca de 800bp, mas muitas outras foram entretanto descritas. Uma das características das IS é a presença de uma pequena seqüência que aparece invertida em cada uma das extremidades (IR do inglês *inverted repeats*). Os *inverted repeats* possuem entre 16 nucleótidos e algumas dezenas, sendo o seu comprimento e seqüência específico de cada IS.

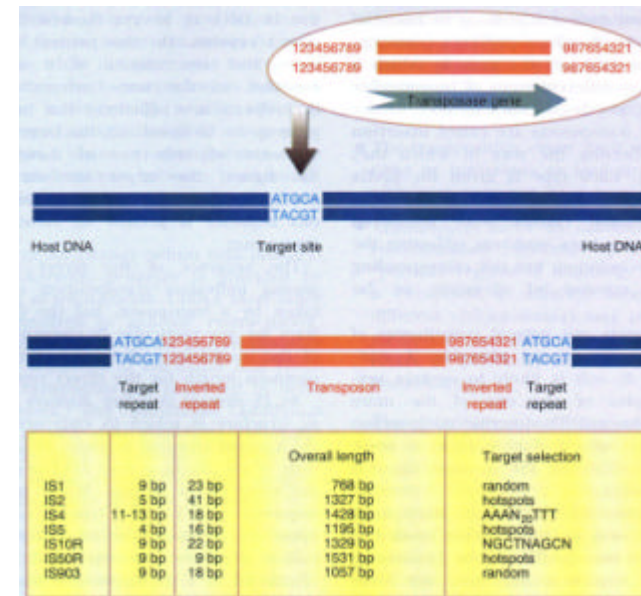


Figura 1.33 – características de alguns transposons e consequências da sua inserção

Um transposon (Tn1, Tn2, etc.) é um elemento de transposição mais complexo, contendo dois IS separados por uma seqüência de DNA que pode conter um ou mais genes. Os dois IS nas extremidades são idênticos, e podem estar na mesma orientação, ou em orientações invertidas. Se estiverem na mesma orientação os IR presentes são os IR de cada um dos IS. Se estiverem em orientações invertidas, todo o IS funciona como um IR. Os transposons herdam dos IS que os constituem a capacidade de se deslocarem no genoma, mas com uma habilidade extra que é a de transportarem consigo o(s) gene(s) que os constituem. Na verdade, em bactérias, este é o mecanismo que muitas vezes se encontra por detrás da aquisição de resistência a antibióticos num único plasmídeo.

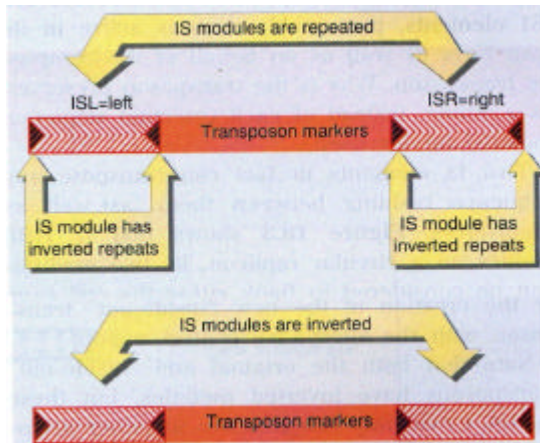


Figura 1.34 – Os transposons são elementos de transposição mais complexos. São constituídos por uma sequência central contendo genes ou marcadores, e sequências laterais que podem ser repetições directas ou invertidas de IS.

Os mecanismos de transposição podem ser conservativos ou replicativos. No primeiro ocorre a deslocação do elemento de transposição de um local para outro, sem que o numero de elementos tenha variado. No segundo caso, é deixada uma cópia do elemento de transposição no local original, isto é o elemento de transposição é replicado para um novo local.

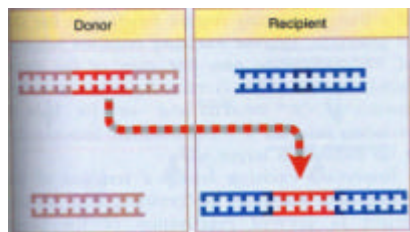


Figura 1.35 – Mecanismo de transposição conservativa

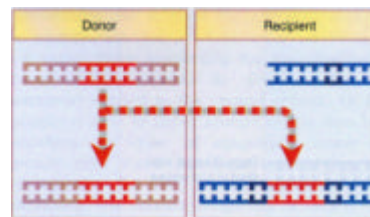


Figura 1.36 – Mecanismo de transposição replicativa

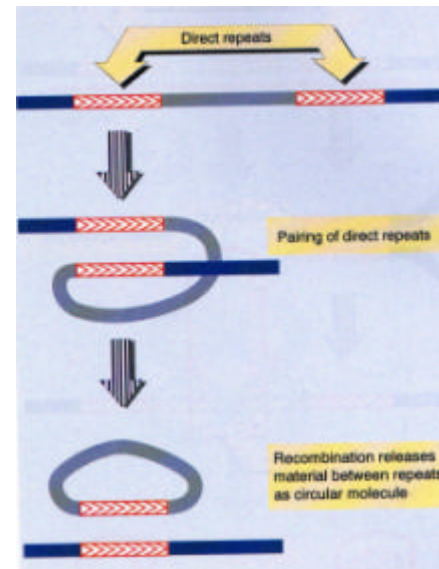


Figura 1.37 – Os transposons podem ser excisados por recombinação homóloga entre os direct repeats, deixando uma cópia dos repeats no local de inserção

O mecanismo de transposição replicativa é catalizado por uma transposase, e envolve um intermediário que possui todo o genoma do plasmídeo dador integrado no genoma do plasmídeo receptor, possuindo este intermediário duas cópias do elemento de transposição (uma em cada extremidade do genoma dador). Em seguida, ocorre como que um cross-over (mediado pela sequência do elemento de transposição), que é depois resolvida pela transposase (a qual pode ou não ser codificada nos elementos de transposição, pelo que a sua acção é efectuada em trans ou cis respectivamente). Uma outra consequência da transposição é a ocorrência de uma duplicação do DNA receptor em cada extremidade da inserção. Esta duplicação resulta do facto de o corte em cada cadeia do DNA receptor ser efectuada em locais diferentes, e de a sequência de DNA de cadeia simples ser depois resolvida pela polimerase.

A presença de elementos de transposição origina uma alta frequência de deleções na sua vizinhança. Estas deleções ocorrem como eventos de transposição aberrantes e podem incluir ou não parte do elemento de transposição. A própria mobilidade dos elementos de transposição pode ser causa de mutações, já que quando um IS se posiciona no meio de um gene, produz uma interrupção deste. Se altera a sequência codificante, ou a separa do promotor, o resultado pode ser um gene não funcional, com funcionalidade alterada, ou com baixa ou nula expressão.

Os elementos de transposição acima descritos, são relativamente simples e pequenos, existindo nos procariontes. Nos Eucariotes foram também descritos elementos de transposição como os Ty (levedura), os *Copia like elements* (drosofila), os *FB elements* (drosofila), os *P elements* (drosofila), e os retrovírus.

Os Ty elements possuem uma sequência altamente variável, mas terminada por uma sequência repetida (a sequência  $\delta$ ), a qual é

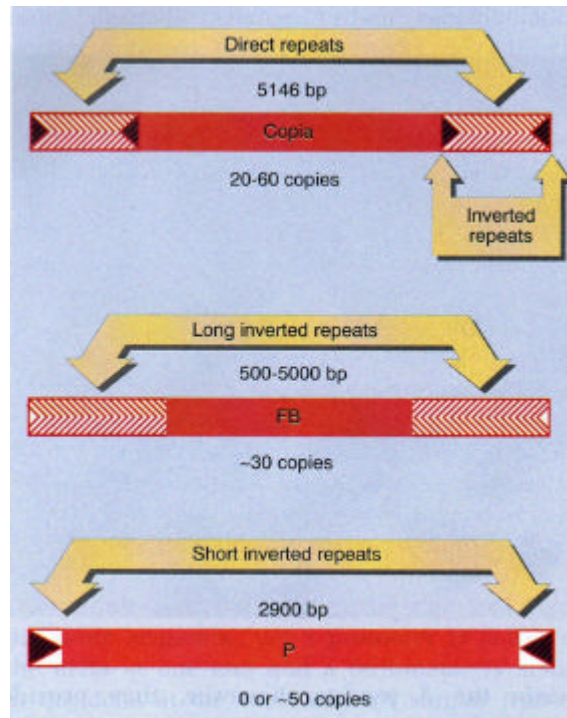


Figura 1.38 – Elementos de transposição dos eucariotes e suas características

também variável de elemento para elemento. Os Ty podem ter vários Kb de comprimento, e as sequências  $\delta$  cerca de 300bp. Uma particularidade das sequências  $\delta$  é que não se constituem em repetições invertidas, mas repetições directas.

Cerca de 10% do genoma da drosofila é composto por famílias de elementos de transposição disseminados, e que se movem como elementos discretos no genoma. Na drosofila foram caracterizados 3 tipos de elementos de transposição distintos: (*Copia like*, *Fold Back* ou *FB* e os *P elements*).

Os *Copia like* são compostos por 7 famílias diferentes, variando em tamanho de 5 a 8.5Kb. Membros de cada família são representados com 10-100 elementos no genoma da drosofila. Cada elemento possui uma repetição directa longa, e uma repetição invertida imperfeita e curta, sendo estruturalmente semelhante ao Ty da levedura.

Os elementos FB (*Fold Back*) variam em tamanho de algumas centenas de pares de bases a alguns Kb. Possuem entre si sequências semelhantes mas não iguais. Possuem terminais longos em repetições invertidas. Por vezes todo o elemento é constituído pelas repetições invertidas, com excepção de uma pequena sequência entre estas. O seu nome deriva do facto de em virtude de possuírem estas repetições tão longas estes elementos poderem hibridizar consigo próprios, como que dobrando-se sobre si mesmos. As propriedades observadas em elementos FB indicam que estes podem excisar-se a si próprios do genoma, provocando rearranjos cromossómicos com elevada frequência.

Os elementos P variam em tamanho de 0.5 a 2.9 Kb (os elementos mais pequenos derivam de deleções parciais num elemento P completo), tendo sempre uma sequência repetida invertida perfeita com 31 bp nas suas extremidades. Na região central dos elementos P

podem existir até 3 *open reading frames*, sugerindo que os elementos completos codificam para três proteínas.

Uma propriedade surpreendente dos elementos P é que se uma fêmea sem elementos P for cruzada com um macho com P+, a progenia apresenta mutagênese por inserção de elementos P (isto é os elementos P apresentam-se activos). Em contraponto, se a fêmea for P+ e o macho P-, a progenia não apresenta mutagênese por inserção os elementos P.

O modelo correntemente aceite para explicar as propriedades dos elementos P baseia-se na existência no interior do elemento de uma transposase e de produtos repressores dos elementos P. Se a fêmea possuir elementos P, então o óvulo possui proteínas repressoras dos elementos P, pelo que não ocorre transposição. Se no entanto a fêmea não possuir estes elementos, os elementos P do macho vão estar activos, inserindo-se no genoma, do que resulta mutagênese por inactivação de genes.

Os provírus (DNA de cadeia dupla inserida no genoma do hospedeiro) dos retrovírus podem ser considerados elementos de transposição já que podem efectivamente transportar-se de um lugar para outro no genoma da célula alvo. Com efeito, o genoma dos retrovírus tem algumas semelhanças com os elementos de transposição habituais. Em particular possuem um LTR (*long terminal repeat*) em cada extremidade semelhantes aos observados nos elementos Ty1 da drosófila.

A inserção dos retrovírus, tal como a dos elementos de transposição resulta na duplicação da zona de DNA onde se inserem. Uma outra forma de descrever esta similaridade entre os retrovírus e os elementos de transposição como os Ty1 e os *copia-like* é a de dizer que estes apresentam sequências que indicam serem reminiscências de retrovírus. Esta forma de ver o problema recebeu suporte experimental de experiências efectuadas por Jef Boeke and Gerald

Fink, demonstrando a existência de um intermediário de RNA na inserção destas sequências.

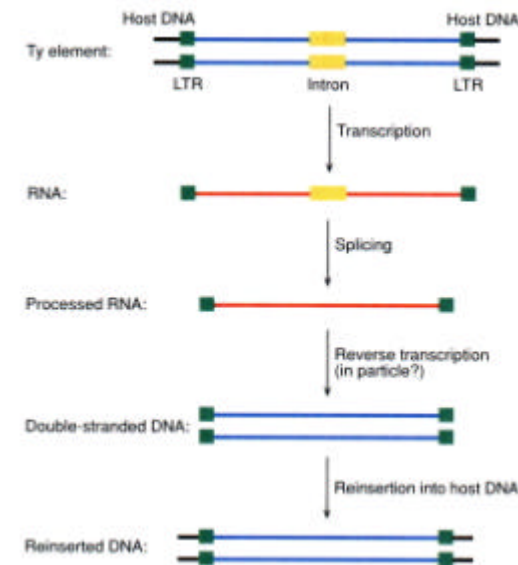


Figura 1.39 – Os retrotransposons e os retrovírus possuem semelhanças na forma de se integrar no genoma celular.

Os elementos de transposição que utilizam este intermediário de RNA (ex. Ty1 e *copia-like*) são denominados retrotransposons, e são frequentes entre os eucariotas, sendo geralmente divididos em duas classes: Os retrotransposons virais (como os Ty1 e os *copia-like*) e os retrotransposons não virais (os mais frequentes entre os mamíferos). Exemplos de retrotransposons não virais entre os mamíferos são os LINES (Long interspersed elements; 1-5Kb) e os SINES (short interspersed elements; ex. as sequências alu).

#### 1.1.2.5 Mecanismos de reparação do DNA

As células eucarióticas desenvolveram uma série de mecanismos integrados de reparação de danos no DNA, os quais podem ser

essencialmente de três tipos: mecanismos excisativos, não excisativos e mecanismos pós replicativos.

#### 1.1.2.5.1 Mecanismos não excisativos

- **Evitar erros antes de surgirem:** alguns sistemas enzimáticos neutralizam compostos potencialmente perigosos antes de poderem reagir com o DNA. Um bom exemplo é o mecanismo de eliminação de radicais superóxido, catalizado pela superóxido dismutase e pela catalase ( $2 \text{HO}\cdot \rightarrow \text{H}_2\text{O}_2$  ;  $2 \text{H}_2\text{O}_2 \rightarrow \text{H}_2\text{O} + \text{O}_2$ ). Um outro exemplo é o do sistema catalizado pela proteína do gene *mut*, o qual previne a incorporação do 8-oxodT no DNA hidrolizando o trifosfato de 8-oxodT em monofosfato.
- **Reversão directa do dano:** A forma mais directa de reverter uma lesão é efectuar o seu processo inverso, restaurando assim a sequência alterada. No entanto, a reversão directa nem sempre é possível, já que algumas reacções são no essencial irreversíveis.

Um dos casos em que a reversão é possível é o da fotodimerização por luz UV. O fotodímero de ciclobutano pirimidínico pode ser reparado por uma fotoliase (foi encontrada em eucariotas inferiores e bactérias, mas não em humanos).

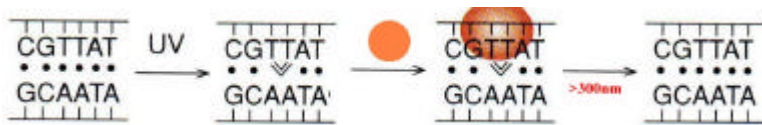


Figura 1.40 – Esquema de acção da fotoliase que por acção da luz reverte a formação dos anéis de ciclobutano pirimidínico formados por acção da luz ultravioleta.

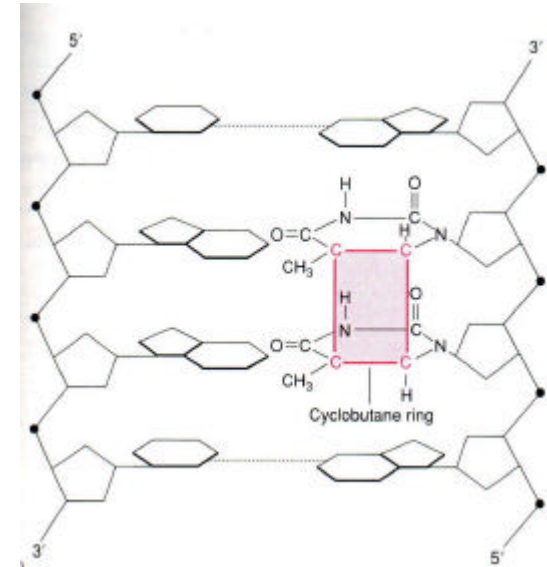


Figura 1.41 – dímero de ciclobutano pirimidínico formado por acção da luz ultravioleta.

Esta enzima liga-se ao dímero, catalisando a sua cisão em monómeros, na presença de luz visível de certos comprimentos de onda.

Um outro mecanismo de reversão directa é catalisado pelas alquiltransferases. Estas enzimas removem certos grupos alquilo adicionados às guaninas na posição O-6, por certos agentes químicos. Note-se no entanto que este é um processo enzimático em que as alquiltransferases ficam inactivadas, do que pode resultar uma saturação deste mecanismo em situações de grande alquilação do DNA.

### 1.1.2.5.2 Mecanismos excisivos

- **Mecanismo de reparação geral por excisão:** Também denominado *nucleotide excision repair*, envolve a quebra da ligação fosfodiéster de ambos os lados da lesão, mas apenas na cadeia nucleotídica lesada. A sequência interrompida daí resultante é reparada por uma polimerase e uma ligase, tendo como padrão a cadeia nucleotídica complementar. Este mecanismo é no Homem bastante complexo, envolvendo pelo menos 17 proteínas

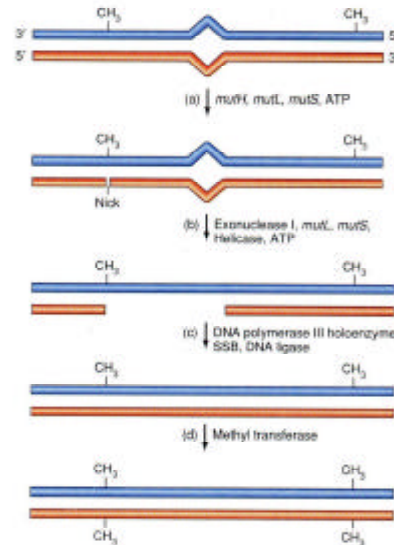


Figura 1.42 – Mecanismo geral de reparação por excisão.

- **Mecanismo de reparação ligado à transcrição:** A transcrição e reparação estão também ligadas, o que o demonstra o envolvimento de TFIIH (um factor de transcrição) na reparação, bem como o facto de nos eucariotas e nos procariotas, a cadeia transcrita é, nos genes activos, preferencialmente reparada, em desfavor da cadeia complementar.
- **Mecanismo excisativo específico:** Algumas lesões provocam danos geométricamente mais subtis, pelo que a sua detecção é mais difícil (não causam grandes distorções no DNA). Para estes

casos, as células desenvolveram mecanismos específicos de detecção e reparação dos danos:

- ↗ **Reparação de DNA Glicosilase (*base excision repair*):** As DNA glicosilase não clivam as pontes fosfodiéster, mas antes as ligações N-glicosilicas (base-pentose), gerando um nucleótidoapurínico ou apirimidínico (locais AP). Esta lesão é posteriormente reparada pelo mecanismo de reparação endonucleolítica AP (ver abaixo). Existem numerosas glicosilases, como por exemplo Uracil-DNA glicosilase, a hipoxantina glicosilase (a hipoxantina é o resultado de uma desaminação da adenina), 3-metiladenina glicosilase, 3-metilguanina glicosilase, a 7 metilguanina glicosilase, etc.

- ↗ **Reparação endonucleolítica de AP:** Todas as células têm endonucleases que atacam os locais purínicos e pirimidínicos (AP endonucleases). A acção destas enzimas é vital já que como vimos a despurinação espontânea é muito elevada. Estas enzimas produzem quebras das ligações fosfodiéster junto ao nucleótidoapurínico. A este corte endonucleolítico segue-se um corte exonucleolítico, o preenchimento do vazio pela polimerase, e finalmente o restabelecimento de continuidade pela ligase.

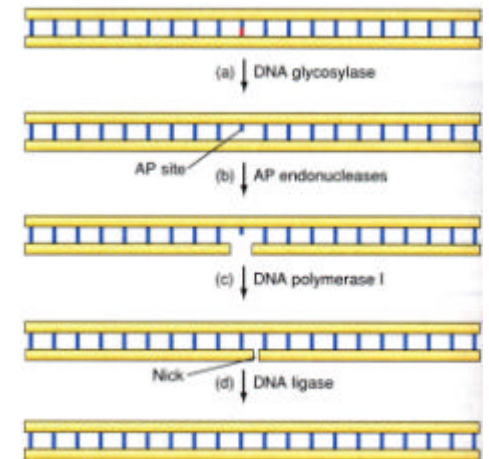


Figura 1.43 – Mecanismo de reparação endonucleolítica de locaisapurínicos

☞ **O sistema GO:** Para prevenir lesões mediadas pela 8-oxodG, duas glicosilases codificadas pelos genes *mutM* e *mutY* trabalham em conjunto com o produto do gene *mutT* já referido. Quando surge um dano oxidativo produz lesões tipo GO, a glicosilase codificada pela *mutT* remove a lesão. No entanto, por vezes as lesões passam despercebidas, emparelhando com adenina. Neste caso, o produto do gene *mutY* remove a adenina levando à sua substituição pela citosina correcta por reparação de síntese, a que se segue a remoção do produto GO pelo produto *mutM*.

### 1.1.2.5.3 Mecanismos pós-replicativos

Alguns mecanismos de reparação são capazes de reconhecer os erros no DNA, mesmo após ou durante a sua replicação:

- **Reparação de *Mismatches*:** Este mecanismo detecta a existência de bases mal emparelhadas (*mismatches*) durante a replicação. Este processo envolve três passos essenciais: 1) a detecção da existência de um par de bases mal emparelhado; 2) o reconhecimento de que a base mal incorporada tem que ser a que existe na cadeia a ser sintetizada; 3) excisar a base incorrecta e reparar o corte. Assim, o passo crítico é o reconhecimento da cadeia sintetizada de novo, o que é conseguido recorrendo ao atraso na metilação relativamente à síntese de DNA. Assim, a cadeia de DNA não metilada é reconhecida como sendo a que deve ser processada.
- **Reparação por recombinação:** O gene *recA* envolvido no

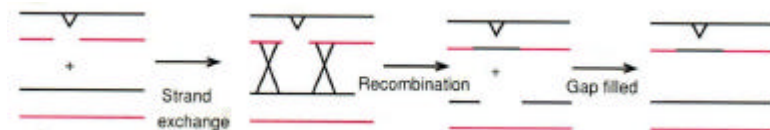


Figura 1.44 – Reparação por recombinação

1.65

bypass SOS está também envolvido na reparação por recombinação. Neste mecanismo, o sistema de replicação pára junto a um dano, continuando depois da lesão, deixando uma quebra de alguns nucleótidos na cadeia sintetizada de novo. Neste

mecanismo de reparação, o DNA da zona de quebra é cortado da outra molécula de DNA homóloga, sendo esta depois reparada por síntese normal. Assim, este mecanismo de reparação difere do SOS, já que neste são adicionados nucleótidos de forma aleatória na zona de quebra (efectivamente criando mutações), enquanto na reparação por recombinação não são efectuadas nenhuma alterações ao programa genético da célula.

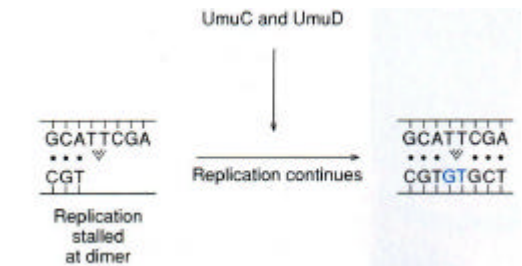


Figura 1.45 – Mecanismo de reparação SOS. este mecanismo apesar de permitir a continuação da replicação, ao introduzir aleatoriamente nucleótidos, pode produzir mutações.

## 1.2 O ciclo celular

### 1.2.1 Divisão celular e anomalias genéticas

De acordo com a sua actividade reprodutora, a célula pode encontra-se em interfase (a fase entre duas divisões celulares), ou em fase M (a fase em que se encontra em mitose). A maior parte do ciclo celular passa-se portanto em interfase. Durante períodos específicos da interfase a célula integra estímulos e toma decisões centrais para a entrada na fase de mitose. De acordo com a actividade específica que se efectua, a interfase subdivide-se em fases, com actividades e papéis específicos na divisão celular, e que culminam na mitose:

1.66



- **Fase G1** (de “GAP” 1; 6 a 12 horas). Não há qualquer síntese de ácidos nucleicos, mas uma intensa actividade de síntese proteica prepara a maquinaria necessária às fases subsequentes. A decisão de avançar para a duplicação do DNA ocorre durante a fase G1 (ponto de restrição).

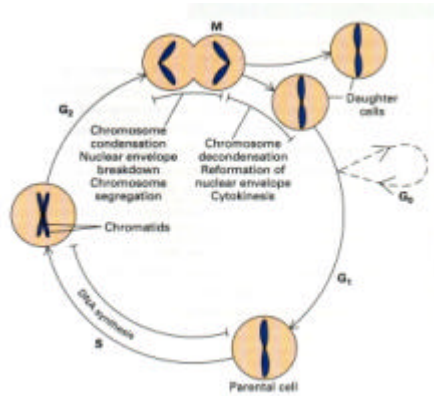


Figura 1.46 – Fases do ciclo celular

- **Fase S** (6 a 8 horas). Inicia-se com a replicação do DNA, e continua até que todo o DNA se encontra duplicado (conteúdo em DNA 4n).
- **Fase G2** (de “GAP” 2; 3 a 4 horas). Decorre entre o fim da fase S e o início da mitose, durante o qual a célula organiza os seus 2 conjuntos completos de cromossomas diplóides. No final desta fase ocorre a decisão de avançar para a mitose
- Durante a **mitose** (1 hora), um conjunto diplóide de cromossomas é segregado para cada núcleo. Apenas nesta fase, os cromossomas são observáveis como entidades discretas, sendo a organização celular destruída, e organizado o fuso acromático entre os dois futuros núcleos. Toda a actividade sintética pára durante a mitose.
- **Fase G0**. Uma fase específica de células maduras, que pararam a divisão celular. No entanto alguns tipos de células, podem ser estimulados a deixar a fase G0 e entrar em G1.

1.67

### 1.2.2 Ciclinas e CDK

Durante todo o ciclo celular, existem pontos de decisão, os quais verificam se a célula se encontra já preparada para proceder para a fase seguinte. Alguns dos mais importantes são: O DNA já se encontra todo duplicado, e apenas duplicado? Existem erros na duplicação do DNA? A massa celular está já duplicada? Estes “checkpoints” existem sob a forma de factores activadores e inibidores. Destes, alguns encontram-se já identificados enquanto que a existência de outros se encontra apenas deduzida do comportamento observado em fusões de células em fases diferentes do ciclo celular. Nomeadamente o factor que determina a decisão de passar o ponto de restrição é totalmente desconhecido, enquanto que o factor que determina a entrada na fase S se encontra identificado como sendo uma proteína cinase, relacionada com a cinase que inicia a mitose. Esta última (MPF – M phase promoting Factor ou M phase Kinase), a primeira a ser identificada é a mais bem estudada, é um dímero composto por duas subunidades: uma a subunidade catalítica (p34) é activada no início da fase S, a outra (p45) é uma ciclina (acumula-se por síntese contínua durante a interfase), sendo destruída durante a mitose, o que é responsável pela inactivação da cinase da fase M, o que sinaliza a saída da mitose.

O estudo cristalográfico do dímero de MPF revelou que a ligação da ciclina provoca uma alteração conformacional na p34, a qual é essencial para a formação do centro catalítico. Uma MPF tem apenas um tipo de p34, mas pode conter uma de vários tipos de ciclinas. Basicamente dois tipos de ciclinas pouco relacionadas (A e B) podem fazer parte da MPF. A similaridade entre estas duas ciclinas centra-se numa estrutura com cerca de 150 aminoácidos chamada a cyclin box. Em mamíferos existem ainda dois tipos de ciclinas B (B1 e B2).

A p34 encontra-se em níveis constantes ao longo de todo o ciclo celular, e como a síntese de ciclina é constante ao longo de toda a fase G1, a sua presença não constitui o sinal de activação da MPF. Na

1.68

realidade, o dímero acumula-se numa forma inactiva, sendo a modificação da p34 que constitui o evento activador do dímero. A activação da p34 é efectuada por fosforilação/desfosforilação, já que para esta se encontrar activa é necessária a presença de grupos fosfatos em alguns resíduos e a sua ausência noutros. Como a actividade cinase da MPF é auto-catalítica, basta a activação de uma pequena porção, para que rapidamente toda a MPF disponível seja activada. Como a MPF não é uma fosfatase, e é necessária actividade de cinase e fosfatase para a activação da MPF, esta deve activar directamente a fosfatase que a activa, criando um circulo auto-catalítico completo.

O evento inactivador é constituído pela destruição proteolítica da ciclina, a qual ocorre durante a fase M.

Na realidade, o controlo da divisão celular é constituído por uma intricada rede de reacções de cinase e fosfatase, que desencadeiam a activação/inactivação de uma série de factores, permitindo o desencadear temporalmente organizado das várias fases do ciclo celular. Toda esta série de reacções culminam na activação da MPF com a concomitante entrada na fase M. Dois modelos gerais podem explicar a actividade da MPF: 1) Pode ser um regulador central que fosforila proteínas alvo que por sua vez actuam para regular outras actividades, i.e., pode tratar-se de uma reacção clássica em cascata; 2) pode ser um regulador central, que activa ele próprio uma serie de substratos cruciais indispensáveis para realizar trabalhos regulatórios ou reorganizacionais envolvidos no ciclo celular.

Os substratos da MPF têm em comum uma estrutura proteica constituída por Ser-Pro flanqueada por resíduos básicos (habitualmente na forma Ser-Pro-X-Lys). Substratos potenciais incluem a histona H1 (possivelmente necessária para condensar os cromossomas), lamininas (possivelmente necessárias para desorganizar o envelope nuclear), nucleolina (possivelmente envolvida na paragem da actividade ribossomal), bem como outras

proteínas estruturais e enzimáticas mais provável é que a MPF (maioria) destes substratos.

A proteína p34 foi pela primeira vez identificada pela dimensão do seu ovo (modelo para purificar proteínas). Posteriormente descobriu-se que a proteína p34 é uma cinase. Dado o grande paralelo em mamíferos, o homólogo animal é a ciclina B, e o nome do exemplo mais bem conhecido é a ciclina B das células animais. Na levedura o cdc2 é um homólogo de cdc13 formando uma ciclina B das células animais. cdc2 tem um par diferente de resíduos de serina e treonina na ciclina-B. Note-se que as 2 fosforilações são diferentes e a activação celular pode também ser definida em cada momento.

Na levedura o cdc2 é um homólogo de cdc13 formando uma ciclina B das células animais. cdc2 tem um par diferente de resíduos de serina e treonina na ciclina-B. Note-se que as 2 fosforilações são diferentes e a activação celular pode também ser definida em cada momento.

O produto de cdc25 é necessário para a activação da MPF por fosforilação da Y15 e T161 de Cdc2 respectivamente. Apesar de não possuir actividade fosforiladora, a cdc25 é efectivamente uma fosfatase mediada pela Wee1 e CAK, e pela sua actividade de desfosforilação que constitui o evento chave na activação da M phase da cinase. Como mutantes de cdc25 não param a mitose se a duplicação

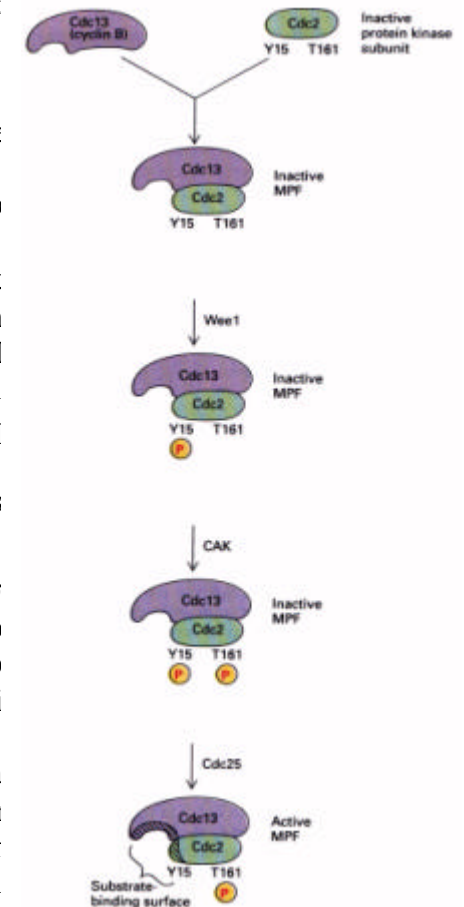
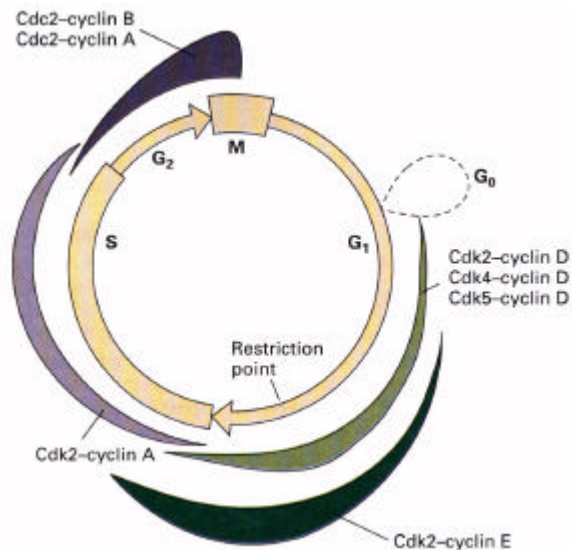


Figura 1.47 – via de activação da MPF por fosforilação da Y15 e T161 de Cdc2 respectivamente. Apesar de não possuir actividade fosforiladora, a cdc25 é efectivamente uma fosfatase mediada pela Wee1 e CAK, e pela sua actividade de desfosforilação que constitui o evento chave na activação da M phase da cinase. Como mutantes de cdc25 não param a mitose se a duplicação

de DNA não tiver sido conseguida, esta fosfatase é importante para assegurar que a fase S se completa antes de se iniciar a fase M.

Mutantes do gene *wee1* permitem que a fase M se inicie sem que o crescimento necessário tenha ocorrido. Assim, este gene normalmente impede o início da fase M, se a massa celular crítica não tiver sido atingida. *Wee1* codifica para uma cinase pouco usual. Pode fosforilar em serina/treoninas e tirosinas. Inibe *cdc2* por fosforilação em *tyr-15*. Um outro gene *mik1* tem efeito semelhante. Este controlo da *cdc2* na transição G2/M parece ter sido preservado ao longo da evolução, já que genes homólogos são encontrados em vários tipos de leveduras, em anfíbios e em mamíferos.

A activação da transição G1/S requer activação da *cdc2/cig2* (em *S. Pombe*), mas também a inactivação da *cdc2/cdc13* é indispensável. Com efeito, mutantes de *cdc13* não só não entram em fase M, como prosseguem por vários ciclos da fase S, sugerindo que a *cdc2/cdc13* é um inibidor da fase S. Assim, activação de *cdc2/cd13* durante a fase G2 impede a continuação da fase S, e estimula o início da fase M. A destruição da *cd13* no fim da fase M pára a fase M e permite um novo ciclo de fase S antes de se iniciar nova divisão. O funcionamento



1.71

Figura 1.48 – Intervenção das várias ciclinas em diferentes fases do ciclo celular.

deste ciclo depende, provavelmente de *cdc18*. Esta proteína é activada após a passagem do ponto START, sendo necessária para entrar na fase S. Para *cdc18* ser activa é necessário que *cdc2/cdc13* esteja inactiva. A activação deste dímero na fase M inactiva a *cdc18*, impedindo novo ciclo de síntese de DNA. A actividade de *cdc2/cdc13* é dependente do factor *rum1*. Sem este factor, as células entram em mitose prematuramente. Assim, *rum1* deve ser um inibidor de *cdc2/cdc13* (M phase kinase) é expresso entre G1 e G2 mantendo a M phase kinase inactiva ( o que é particularmente importante em G1 antes da fosforilação em *tyr-15*), e reprimindo o nível de *cdc13*.

Desta imagem, dois conceitos gerais emergem para o controlo da divisão celular: 1) a presença de loops de feedback de controlo e; 2) a presença de duplicidade de efeitos, em que um factor activa a fase seguinte e reprime a anterior.

A passagem do ponto START está também dependente de *cdc2*. No entanto, a ciclina que se associa a *cdc2* é diferente quando a cinase controla o ponto start e a transição G2M.

Os mecanismos envolvidos na regulação do ciclo celular em leveduras e em animais são semelhantes, ainda que o número de subunidades presentes em animais seja maior, havendo subunidades específicas para o controlo das transições G1/S e G2/M.

A comparação das subunidades envolvidas nos diferentes seres estudados estão descritas na tabela 1.4.

Como se pode ver, nos animais, a cinase dependente de ciclinas que regula a transição G2/M é única, á semelhança do que se passa com as leveduras. No entanto, existem várias ciclinas que com ela podem interactivar. O mesmo não se passa na transição G1/S, a qual é regulada por várias cinases e várias ciclinas.

1.72

**Tabela 1.4** - comparação das subunidades envolvidas nas várias transições do ciclo celular em diferentes organismos

	Transição			
	G1/S		G2/M	
	subunidade catalítica	subunidade reguladora	subunidade catalítica	subunidade reguladora
<i>S.Pombe</i>	cdc2	cig 1,2	cdc2	cdc13
<i>S.cerevisiae</i>	cdc28	CLN1-3	cdc28	CLB1-4
<i>Mamíferos</i>	cdk2(p33), cdk4	ciclinas A, D1, D2, D3, E	p34(cdc2)	ciclinas A, B1, B2

Cdk = cyclin dependent kinases

Uma outra diferença tem a ver com as ciclinas C e E, as quais acumulam durante a fase G1 como as restantes, mas não são destruídas na fase M.

Assim, como vimos, nos animais existem essencialmente 2 tipos de cinases : a cdk2 e a cdc2, respectivamente responsáveis pela transição G1/S e G2/M. Ainda que se trate de componentes diferentes, as estruturas regulatórias conhecidas são similares em ambas, pelo que se deduz que são similarmente regulados. Assim, a síntese de ciclinas na fase G1, vai progressivamente aumentando a sua concentração, até que se chega a um ponto em que se inicia a formação de dímeros. No entanto estes não são ainda activos, sendo necessárias reacções de fosforilação nos resíduos 15 e 161. A primeira fosforilação ocorre por acção da Wee1 cinase, enquanto a segunda é catalisada pela CAK (cdc2-activating kinase). O dímero activo promove então a entrada em mitose, activando concomitantemente por fosforilação a fosfatase cdc25. Esta por sua vez promove a remoção do grupo fosfato na tirosina 15 da cdc2, inactivando-a. A ciclina é por sua vez alvo de degradação proteolítica, resultando apenas a cdc2 fosforilada na thr-161, terminando assim a mitose. Não foram identificadas cinases que actuem separadamente em thr-14, em alguns casos a mesma enzima actua na thr-14 e tyr-15.

1.73

A mitose ocorre como um processo temporal controlado. A mitose, como vimos, é iniciada pela activação da MPK. O progresso da mitose requer a degradação de ciclinas e outras proteínas. Por exemplo a separação dos cromossomas na anafase requer a actividade proteolítica, mas não directamente de ciclinas. Existem pelo menos 3 alvos proteolíticos: o primeiro evento é a degradação da ciclina A na metafase, seguindo-se a degradação de 2 alvos na anafase. Uma proteína desconhecida tem que ser degradada para a separação das cromatídeas irmãs e a degradação de ciclina B é necessária para a inactivação da MPK. No final da mitose as fosforilações efectuadas pela MPK têm que ser revertidas.

A síntese de ciclinas D é activada pela acção de factores de crescimento. Estas ciclinas têm uma semi-vida muito curta, o que permite à célula responder a factores externos e à sua remoção. Estas ciclinas estão provavelmente envolvidas na reentrada na divisão celular de células em G0.

O produto do gene de supressão tumoral RB (gene do retinoblastoma) é um substracto dos complexos cdk-ciclinas D, e exerce o seu efeito durante a fase G1 que precede o ponto de restrição. O produto deste gene, na ausência de fosforilação liga-se ao regulador genético E2F, actuando como supressor de alguns genes, bloqueando a transição G1/S. Quando por acção de cdk4,6-ciclina D1,2,3 é fosforilado, liberta a E2F, a qual passa a actuar como um activador genético, promovendo a entrada na fase S. RB é o alvo de várias vias que inibem o crescimento celular, pelo que pode ser uma forma importante de vários sinais manterem a célula em G1 ou G0. Alguns destes sinais (incluindo o TGFβ) actuam através de inibidores de cinases cdk (chamadas ckis). Exemplos de ckis são a p15/p16 (ink4), p21(c1p1/wAF1) e p27(kip1). A importância das ckis é realçada pelo facto de p16 ser também um gene de supressão tumoral, o que indica que a p16 e possivelmente todas as outras ckis são necessárias para parar o crescimento celular. Na verdade, parece que a via das ckis até

1.74

RB é uma via ventral para o bloqueio do crescimento celular, já que são conhecidos genes de supressão tumoral em todos os seus passos.

### 1.2.3 Apoptose

Durante a vida de um organismo multicelular, algumas células morrem num processo natural designado por Apoptose ou morte celular programada. Nos vertebrados, os exemplos mais visíveis de apoptose ocorrem no sistema imune e no sistema nervoso. O processo de morte é característico, envolvendo compactação celular, fragmentação membranar, condensação da cromatina e fragmentação do DNA. Este processo é activo, dependendo do RNA e da síntese proteica.

São conhecidas várias formas de activar a apoptose. Estas envolvem insultos moleculares (retirada de factores de crescimento, tratamento com glucocorticoides, irradiação  $\gamma$ ), bem como estímulos específicos como os produzidos pelas células T citotóxicas ou activação de p53. Assim, a apoptose é importante na embriogénese e contenção do crescimento tecidual, bem como na resposta imune e na contenção de cancro.

Foram já descritas várias mutações recessivas associadas com o estímulo ou o bloqueio da apoptose. Alguns exemplos são: 1) a *lpr* no ratinho (esta mutação leva a deficiência em *fas*) que causa a proliferação excessiva levando a autoimunidade; 2) a *gld* (generalized lymphoproliferative disease) cuja mutação ocorre no gene que codifica para o receptor de *fas*. A proteína *fas* é um receptor membranar relacionado com o receptor para o TNF. Tanto o *fas* como o TNF são capazes de estimular a apoptose. Ao nível da porção transmembranar destes receptores existe uma sequência de 80 aminoácidos essencial para o envio do sinal apoptótico, pelo que é designado por “domínio da morte”. Pouco se sabe sobre o mecanismo de sinalização do receptor *fas* ou TNF para o interior da célula. Contudo, foram identificadas algumas proteínas que interagem com o

“domínio da morte”, as quais curiosamente também os possuem. Pensa-se assim que este domínio seja importante para promover a dimerização destas proteínas. Outros compostos importantes na via da apoptose são a família de proteases designada por ICE. O protótipo destas proteases é a “*IL2 $\beta$ -converting enzyme*”: uma protease de cisteína que por proteólise converte o precursor da IL2 $\beta$  na sua forma activa. O processo de apoptose desencadeado pelas ICE é inibido por *crmA* e pelo *bcl-2*.

Uma outra via apoptótica é desencadeada pelos linfócitos Tc, os quais matam as células alvo libertando na sua superfície grânulos contendo proteases de serina e outras proteínas líticas como a perforina, abrindo buracos na superfície das células alvo. As proteases de serina dos grânulos são chamadas granzimas, e na presença de perforina induzem morte por apoptose nas células alvo.

O *bcl-2* inibe a apoptose. Foi originalmente descrito como um protooncogene activado em linfomas por translocação causando a sua hiper-expressão. O *bcl-2* tem uma sequência de ancoragem membranar na zona c-terminal, tendo sido encontrado nas membranas externas de mitocôndria, núcleo e retículo-endoplasmático. Um outro gene semelhante ao *bcl-2* nos mamíferos é o gene *Bax*. O produto do gene *Bax* pode dimerizar com *bcl-2*. O dímero *bcl-2*, bem como o dímero *Bax* bloqueiam a apoptose, mas o heterodímero (*bcl-2/Bax*) não o faz. Assim, a susceptibilidade de uma célula à apoptose é proporcional à proporção *bcl-2/Bax*.

## 1.3 Actividade Celular

### 1.3.1 Componentes dos sistemas de sinalização

Todos os sistemas de comunicação intercelular têm vários componentes. Tipicamente, uma molécula denominada ligando é libertada pela célula sinalizadora. Alguns ligandos são proteínas, enquanto outros são pequenas moléculas como péptidos, esteróides ou vitamina D. O ligando liga-se ao receptor, habitualmente uma

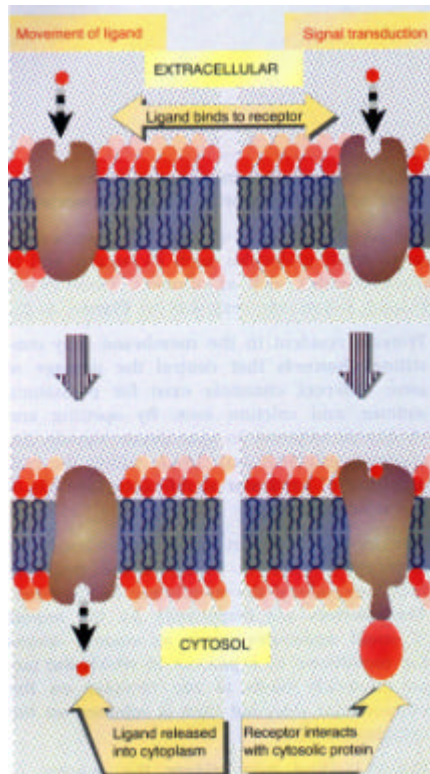


Figura 1.49 – Os sinais podem ser recebidos na membrana celular por um receptor o qual envia o sinal para o interior da célula. Em alguns casos a molécula sinalizadora é ela própria enviada para o citoplasma, enquanto noutros casos é o receptor que interage com componentes intracelulares para assegurar a passagem do sinal.

proteína, na membrana celular ou no interior da célula alvo. Alguns tipos de complexo ligando-receptor são capazes de alterar a expressão

genética directamente, enquanto outros necessitam de enviar sinais através de uma cadeia de transdução, levando o sinal da membrana celular até ao núcleo (fig.1.49).

Alguns ligandos, chamados hormonas viajam a grandes distâncias na circulação sanguínea, antes de interagirem com a sua célula alvo. Estas moléculas podem actuar como sinais mestres para a actividade de órgãos diferentes, os quais podem assim responder de forma coordenada. Outros ligandos não actuam à distância, mas apenas nas células vizinhas das que os produziram.

### 1.3.2 Regulação da sinalização por modulação da conformação proteica

Interações de pequenas moléculas com as proteínas receptoras podem provocar intensas alterações conformacionais nestas. Por exemplo, alterações conformacionais estão na origem da regulação mediada por algumas proteínas cinases. A alteração da conformação coloca ou não amino-ácidos em posições favoráveis à fosforilação. Este mecanismo é extremamente eficiente, não só porque permite a rápida activação dos sinais, mas também porque permite revertê-los rápida e facilmente, reciclando os componentes de sinalização.

### 1.3.3 A Superfamília dos receptores das hormonas esteróides

Um grupo diverso de ligandos, incluindo várias hormonas esteróides, bem como outras pequenas moléculas como a hormona tiróide e a vitamina D, devido à sua estrutura não polar, são capazes de atravessar a membrana citoplasmática, actuando directamente no interior da célula alvo. Os receptores para estes ligandos chamam-se globalmente a superfamília dos receptores esteróides, e estão estrutural e funcionalmente relacionados. A ligação destes ligandos ao seu receptor, causa uma alteração conformacional, a qual provoca a sua libertação de uma proteína sequestradora. Nestas condições, o complexo receptor-ligando desloca-se para o núcleo, onde em

conjunto com outros reguladores da transcrição vão activar ou reprimir directamente a expressão génica, ao ligar-se a sequências reguladoras chamadas HRE (“*hormone response elements*”).

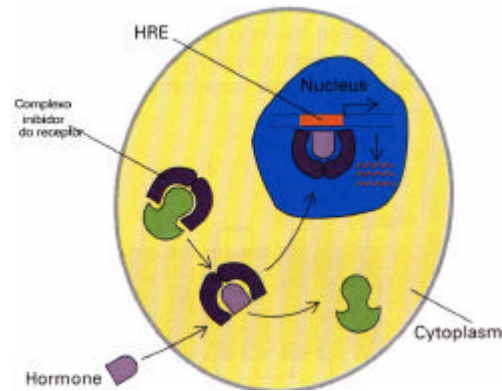


Figura 1.50 – As hormonas esteróides não possuem um receptor membranar, mas um receptor citoplasmático, já que são capazes de atravessar livremente a membrana plasmática.

### 1.3.4 Receptores transmembranares; vias de transdução de sinal

A maior parte dos ligandos são moléculas demasiado volumosas e/ou com carga, para poderem atravessar a membrana citoplasmática. Assim, os seus receptores são habitualmente proteínas de membrana, as quais medeiam a passagem de um sinal para o interior da célula. Os receptores transmembranares são assim compostos por 3 domínios principais: o extracelular, que se liga ao ligando; o transmembranar, que pode atravessar uma ou mais vezes a membrana; e o citoplasmático, que medeia a transmissão do sinal.

Muitos ligandos são dímeros, podendo assim ligar a mais do que um receptor. Desta forma, os ligandos colocam a porção citoplasmática de dois receptores fisicamente próximos, activando a via de sinalização destes receptores (Fig. 1.51). Alguns destes receptores são cinases, tendo por isso a habilidade de fosforilar certos resíduos em proteínas alvo, enquanto outros são serinas/treoninas cinases, e outros não têm qualquer actividade enzimática.

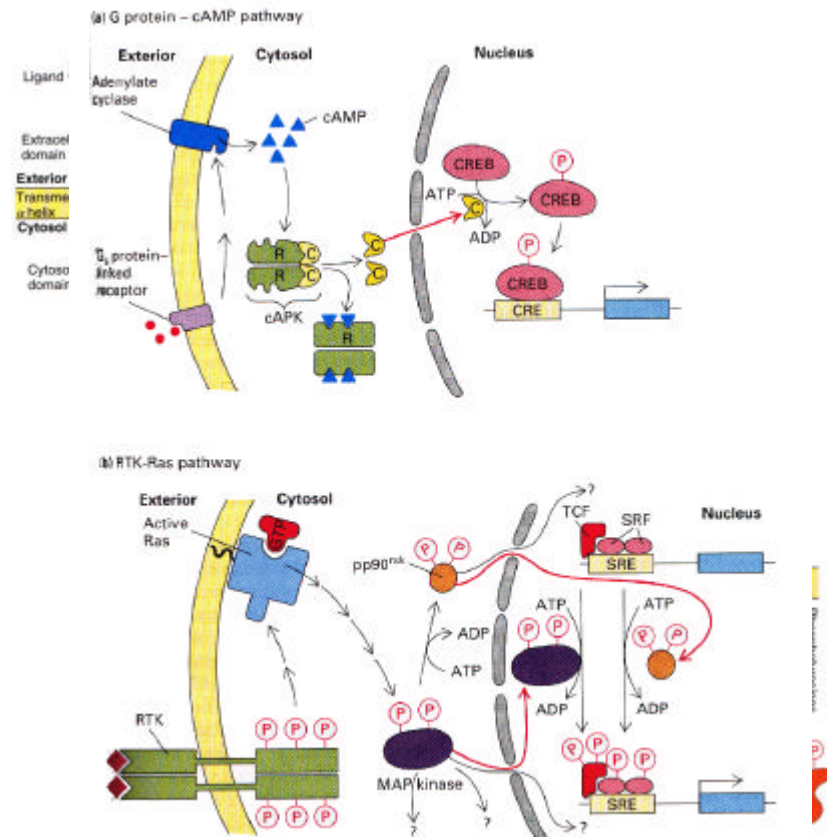


Figura 1.52 – Com frequência a propagação do sinal no interior da célula envolve a utilização de mensageiros secundários como a) cAMP ou b) proteínas G

Um dos mais bem conhecidos exemplos de receptores deste tipo são os receptores tirosina quinase, nomeadamente os receptores para os factores de crescimento. Estes receptores, após ligação do ligando dimerizam, originando a sua auto-fosforilação (fig. 1.51). Esta por sua vez activa a actividade de cinase, que vai fosforilar outras proteínas, iniciando uma cascata transdutora do sinal em que a

fosforilação causa conformação proteica com consequente activação/inactivação proteica. Eventualmente a cascata de activação leva à fosforilação de factores de transcrição, com consequente alteração do estado de activação de um ou mais genes. Mas a autofosforilação não causa apenas a fosforilação de proteínas alvo.

Também a ligação a proteínas adaptadoras é estimulada pela autofosforilação. Interações entre o complexo e outras moléculas causam a propagação do sinal. (fig. 1.51) Com frequência a propagação deste sinal envolve proteínas G (fig. 1.52). As proteínas G são proteínas cuja vida é um ciclo entre a forma ligada a GDP (o estado inactivo) e a forma ligada a GTP (o estado activo); um exemplo particularmente importante de uma proteína G é a *ras*, a qual como veremos está envolvida na carcinogénese. A propagação do sinal das RTK leva à activação de uma proteína que se liga a proteínas-G inactivas, alterando a sua conformação, levando-a a ligar-se ao GTP. A proteína G assim activada liga-se então a uma proteína cinase citoplasmática, alterando a sua conformação, activando-a, o que a leva a fosforilar outras proteínas, incluindo proteínas cinases e factores de transcrição.

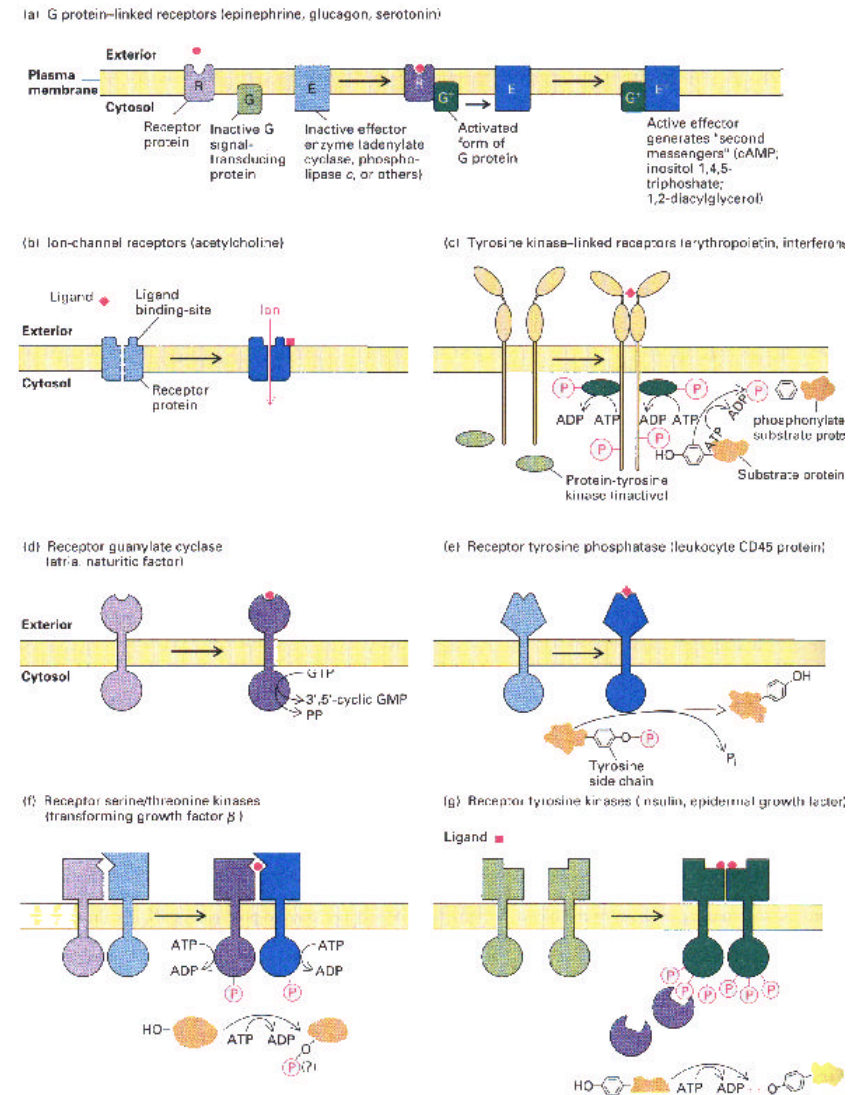
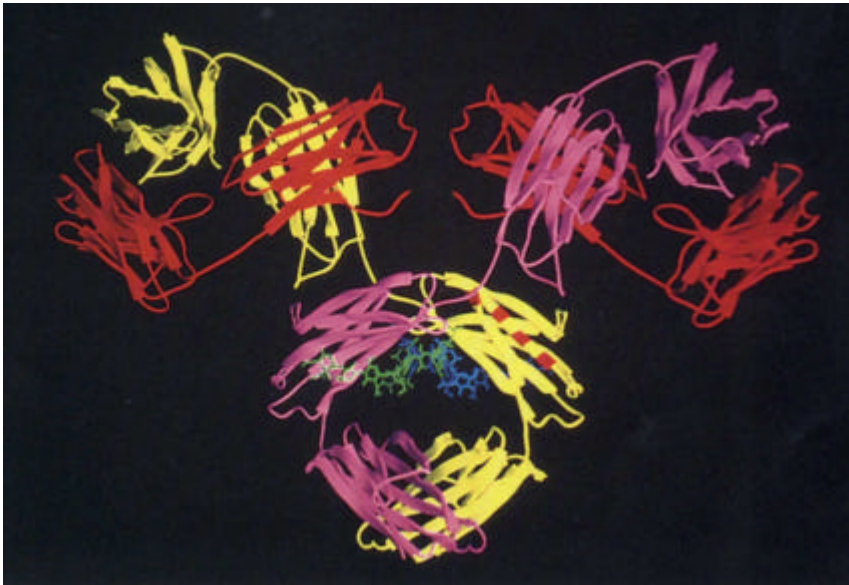


Figura 1.53 – Sumário dos vários tipos de receptores e vias de transdução de sinal.



# Capítulo 2

## 2 Áreas de intervenção da genética molecular em Imunologia



## 2.1 O Receptor da Célula T (TCR)

Ao efectuar a evolução do sistema imune para funções de reconhecimento específico, a natureza teve que resolver o problema gigantesco de codificar num genoma limitado, um número suficiente de genes capaz de reconhecer todo o mundo exterior e interior do organismo. A solução encontrada é se bem que económica, bem complexa, como o revela o facto de não ser ainda possível construir sistemas de recombinação *in vitro* isentos de células. Estes receptores do antigénio são de 2 tipos: as imunoglobulinas, produzidas pelos linfócitos B, das quais existem formas solúveis e formas membranares, e que reconhecem o antigénio na forma nativa; e o receptor da célula T, do qual apenas existem fisiologicamente formas membranares, e que reconhece o antigénio depois de processado por uma célula apresentadora do antigénio (APC), e apresentado no contexto do MHC da APC.

Dos dois tipos de linfócitos, a célula T é a responsável pela resposta imune dita celular. Para tal, estas células estão equipadas à sua superfície com um receptor para o antigénio (TCR do inglês T-Cell-Receptor), através do qual a célula madura recebe um estímulo de activação quando encontra o antigénio para o qual é específica. Este receptor é composto por um de dois tipos de heterodímeros ( $\alpha\beta$  ou  $\gamma\delta$ ). São portanto 4 os genes do TCR, dos quais apenas 2 estarão a ser transcritos em cada célula T. Cada um dos genes é composto por um máximo de 4 tipos de segmentos (V ou variável, D ou de diversidade, J ou de junção, C ou constante). Cada um destes segmentos é composto por mais que um elemento génico, dos quais cada clone celular escolherá um e apenas um para ser utilizado no TCR que irá expressar. As células T diferem assim de todas as restantes células do organismo (com excepção dos linfócitos B), pois o conteúdo genético da célula madura é diferente do de qualquer outra célula que não pertença ao mesmo clone.

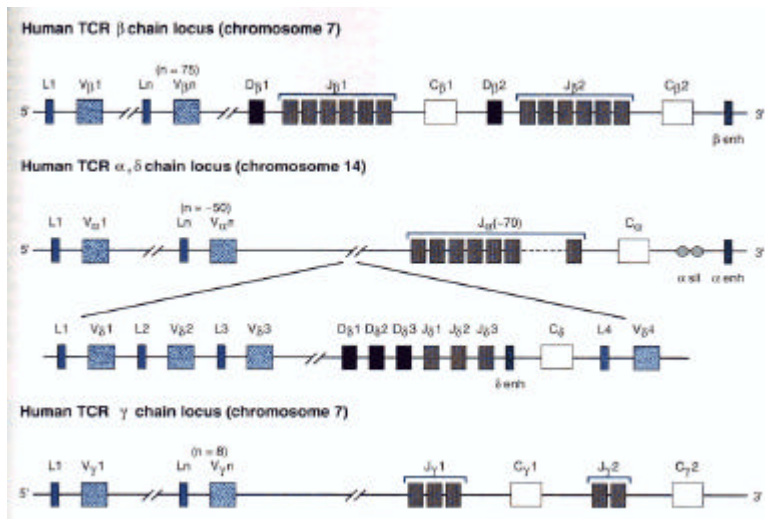
### 2.1.1 Estrutura somática dos genes do TCR

Os genes do TCR, tal como os das imunoglobulinas possuem uma configuração somática, igual em todas as células não linfóides. Nos linfócitos, a configuração destes genes é alterada no processo denominado recombinação, para dar origem a um gene funcional.

Os 4 genes do TCR existem em 3 locus cromossómicos, já que o gene  $\delta$  está localizado no interior do gene  $\alpha$  (Fig. 1). Os locus  $\beta$  e  $\delta$  possuem 4 classes de segmentos (V,D,J,C), e os locus  $\alpha$  e  $\gamma$  apenas 3 (V,J,C). Como se pode ver na fig. 1, no Homem, a organização básica dos locus do TCR é a dita estendida, em que os vários tipos de segmentos se organizam separadamente no genoma (V.V. (etc.) .D.D. (etc.) .J.J. (etc.)). No caso do locus  $\alpha\delta$ , uma variação a esta configuração permite ao gene  $\delta$  partilhar segmentos V com o gene  $\alpha$  (Lewis, 1994).

### 2.1.2 Mecanismo de rearranjo somático dos genes do TCR

O mecanismo de rearranjo somático dos genes do TCR não é diferente do observado para as imunoglobulinas. Na verdade, foi possível clonar células B com os genes do TCR rearranjado (O'Connor et al., 1985), sugerindo que ambos os receptores são substractos do mesmo conjunto de enzimas. O processo de recombinação quer do TCR quer das imunoglobulinas (doravante denominada recombinação V(D)J) depende primariamente de sequências sinal que flanqueando os segmentos a recombinar constituem todos os elementos necessários para indicar aos componentes enzimáticos onde efectuar a recombinação (Lewis et al., 1985; Akira et al., 1987; Hesse et al., 1987). Estes sinais de junção variam em sequência, mas seguem de muito perto o consenso heptamero-espçador-nonamero, em que as sequências consenso do heptamero e do nonamero são respectivamente CACAGTG e ACAAAAACC. O espçador tem uma sequência muito variável, mas o seu comprimento tem 12 ou 23 pares de bases (bp)(Max et al., 1979; Sakano et al., 1979,1981; Kurosawa et



al., 1981). A regra base que dita a orientação dos rearranjos é a de que apenas podem rearranjar elementos com espaçadores diferentes, isto é um elemento com uma sequência sinal composta por um espaçador de 12 bp apenas rearranja com uma outra cujo espaçador for de 23 bp e vice-versa. Desta forma rearranjos envolvendo elementos do mesmo grupo (V com V; J com J) são impedidos. O mecanismo molecular que origina esta restrição é no entanto ainda hoje completamente desconhecido (Lewis, 1994).

Quando dois segmentos genicos se envolvem no processo de recombinação, é feito um corte na fronteira entre a sequência sinal e a sequência codificante, em cada um. As quatro extremidades assim formadas são então ligadas formando uma “junção codificante”, e uma “junção sinal” (fig. 2). Devido à configuração cromossômica, as sequências codificantes são retidas no genomas, sendo as “Junções

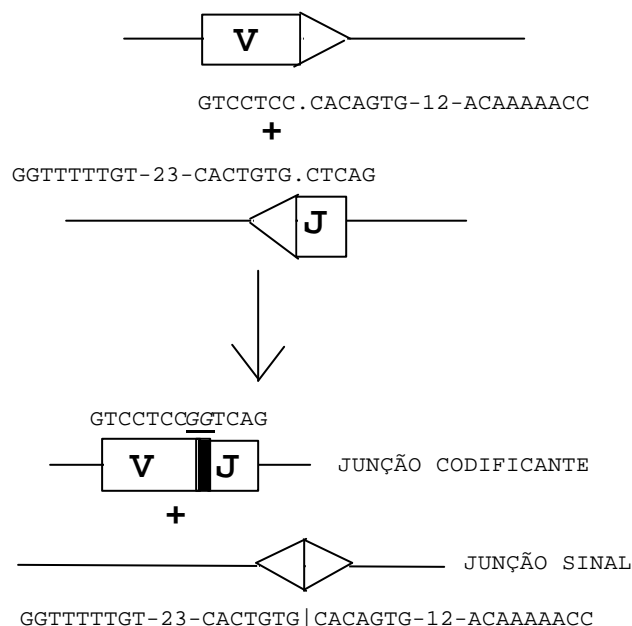
sinal” excisadas sob a forma de DNA circular extracromossômico (Fujimoto et al., 1987; Okazaki et al., 1987).

A junção codificante, não ocorre no entanto sempre numa posição fixa. Por um lado a quantidade de material genético com que cada elemento contribui pode variar em até 10 nucleótidos (Max et al, 1979; Sakano et al 1979; Weiggert et al 1980). Por outro lado, resíduos extra não incluídos na configuração “germline” podem ser incluídos (Sakano et al., 1981; Lafaille et al., 1989; McCormack et al., 1989). Estes resíduos extra podem ser de dois tipos fundamentais: os “resíduos N” (do Inglês Non-germline-regions) e os “resíduos P” (de Palindromicos).

Os “resíduos N” têm tipicamente um elevado conteúdo G/C (Alt et al., 1982; Roth et al., 1989), não ultrapassam os 15 nucleótidos, e ocorrem mais frequentemente nas junções codificantes que nas junções de sinal (Lewis, 1994). Estes resíduos são adicionados pela enzima TdT (do Inglês Terminal deoxynucleotidil transferase) como o demonstram os modelos de animais transgênicos com inativação do gene desta enzima (Gilfillan et al, 1993; Komori et al., 1993). No entanto, o facto de estes modelos resultarem em uma muito grande, mas não completa abolição da frequência de “resíduos N” parece indicar a existência de um mecanismo alternativo, independente da expressão de TdT (Lewis, 1994). A regulação de TdT na ontogenia, origina a menor frequência de “resíduos N” em animais fetais ou neonatais, possivelmente para permitir o domínio de alguns receptores com especificidades necessárias numa fase mais precoce da ontogenia (Gu et al., 1990; Feeney, 1991, 1992).

Os “resíduos P” parecem ter origem numa molécula intermediária tipo “hairpin” gerada (após o corte na sequência sinal) pela ligação covalente das duas cadeias da dupla hélice do DNA, a qual seria posteriormente clivada num ponto diferente do inicial (Fig. 3.; Lieber, 1991; Roth et al., 1992).

O agente ou agentes de recombinação permanecem ainda largamente desconhecidos, ou incompletamente caracterizados e purificados (Lewis, 1994). A tendência actual é no entanto no sentido de aceitar que a recombinação V(D)J se realiza não por um factor, mas por uma colecção de factores com actividades pouco relacionadas. Os factores já identificados incluem RAG-1 e RAG-2 (do inglês Recombination activating Gene; Schatz et al., 1988, 1989; Oettinger et al., 1990;), NBP (do inglês nonamer binding protein; Halligan et al., 1987; Li et al., 1989), T-160 (Shirakata et al., 1991), Rc (Wu et al., 1993), RBP-Jk (Hamaguchi et al., 1989), Rp (do inglês recognition protein; Muegge et al., 1993). Dos factores identificados com base na sua capacidade para produzir cortes no DNA, nenhum apresentava a especificidade necessária (Desiderio et al., 1984; Kataoka et al., 1984; Hope et al., 1986). Apenas um factor foi identificado com base na sua actividade



**Figura 2** - Equação padrão para a recombinação V(D)J. Os sinais de junção são indicados por triângulos e os segmentos codificantes por quadrados. Extraído de Lewis, 1994.

de ligase, tendo sido denominado VDJP (do inglês V(D)J Joining Protein; refered in Lewis, 1994). A actividade de ligase desta proteína só pôde ser observada em fragmentos contendo sinais de ligação, pelo que possui a especificidade necessária para estar envolvida na recombinação V(D)J (Lewis, 1994).

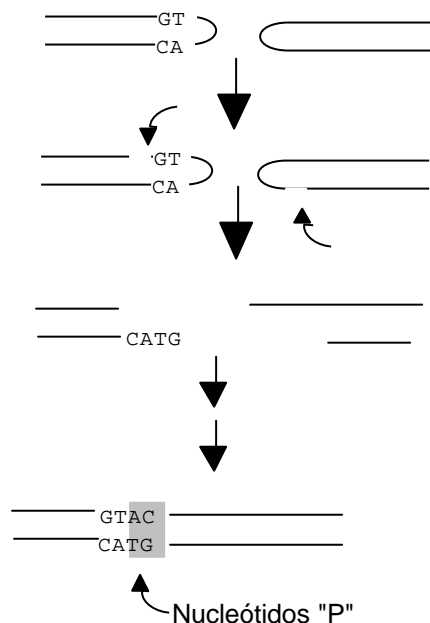
### 2.1.3 POLIMORFISMOS DO TCR

As deleções de regiões variáveis foram dos primeiros polimorfismos a serem detectados no genoma do TCR, tanto em ratinhos de laboratório (Behlke et al., 1986; Haqqi et al., 1989a., 1989b) como em ratinhos selvagens (Pullen et al., 1990; Jouvin-Marche et al., 1989). Polimorfismos mais pontuais foram no entanto também detectados no gene de V $\beta$ 17 de ratinho, verificando-se que as 2 substituições de aminoácidos afectavam a especificidade final do receptor (Cazenave et al., 1990).

No Homem, apenas uma deleção de V $\beta$ 's foi documentada, consistindo na deleção de V $\beta$ 6.2 (mas não de qualquer outro dos genes de V $\beta$  testados) num único indivíduo venezuelano pertencente à tribo índia waraos (Concanon et al., 1987). No entanto os polimorfismos das regiões variáveis do TCR parecem ser quase universalmente representados, ainda que não frequentes na população (Concanon et al., 1987). Com efeito, uma busca sistemática por RFLP indicou a existência de polimorfismos em 12 das 14 famílias de V $\beta$ 's estudadas (Concanon et al., 1987). Alguns destes polimorfismos podem constituir variações silenciosas, como é o caso de um polimorfismo encontrado em V $\beta$ 12.2 (Day et al., 1992), outras no entanto afectam a expressão do gene em linfócitos T maduros, como são os casos dos polimorfismos de V $\beta$ 1 (Robinson, 1989), V $\beta$ 18 (Charmley et al., 1993), V $\beta$ 3 (Posnett et al., 1994) e V $\beta$ 6.7 (Posnett et al., 1986; Li et al., 1990; Prashar et al., 1991). Este último com a particularidade de ser detectável com um anticorpo (Posnett et al., 1986), o que permitiu mapear o epítopo de ligação do anticorpo numa zona de hipotética ligação a superantígenos (Prashar et al., 1991).

Também o polimorfismo descrito para V $\beta$ 3 é único, já que este polimorfismo localiza-se no espaçador, constituindo assim, o único exemplo conhecido de uma mutação numa zona não codificante do TCR, que afecta a expressão do respectivo gene (Posnett et al., 1994). Finalmente, a variação alélica identificada no V $\beta$ 18 é a única que introduz um codão stop, originando um “buraco” no repertório presente em 11% dos indivíduos estudados (Charmley et al., 1993).

Estes dados indicam que mesmo variações moderadas de apenas 1 ou 2



**Figura 3** - Mecanismo proposto para a origem dos nucleótidos P (Adaptado de Lewis, 1994).

pares de bases nas sequências codificantes ou não codificantes do genoma do TCR podem ter repercussões significativas no repertório do TCR (Vissinga et al., 1994).

## 2.1.4 O TCR EM SITUAÇÕES PATOLÓGICAS

Qualquer processo que induza rearranjos cromossómicos tem o potencial de desorganizar o genoma de forma nociva. Desde muito cedo se reconheceu que o sistema de recombinação V(D)J poderia ter um papel pelo menos em alguns casos de malignidade de células T e B (Tycko and Sklar, 1990; Reis et al., 1991; Korsmeyer, 1992; Lieber, 1993). Uma possibilidade de mecanismo consiste na participação da maquinaria de rearranjo no potenciar de rearranjos oncogénicos (Boehm and Rabbits, 1989; Tycko and Sklar, 1990), quer por reconhecimento indevido de alvos de recombinação imperfeitos ou crípticos (Aplan et al., 1990; Brown et al., 1990), quer pela doação de um terminal de uma cadeia de DNA correctamente cortada (Bakhshi et al., 1987). Exemplos de ambas as hipóteses existem (Lewis, 1994), tornando difícil o estabelecimento de um mecanismo único de envolvimento da maquinaria de rearranjo V(D)J na geração de malignidade. A sua implicação parece no entanto indubitável em alguns casos de leucemias linfoblásticas das células T (Lewis et al., 1994), bem como em muitos outros tumores linfoides (Tycko and Sklar, 1990; Rabbits, 1991; Reis et al., 1991; sawyers et al., 1991; korsmeyer, 1992; Magrath, 1992).

O carácter monoclonal do rearranjo dos genes do TCR e das imunoglobulinas, permitiu que as técnicas de análise destes genes se tornassem ferramentas úteis na identificação de tumores de células T e B (Rabbits et al., 1985; O'connor et al., 1985; Paslier et al., 1987; Rambaldi et al., 1985; Loughran et al., 1988; Tawa et al., 1987; Loughran et al., 1988; Furon et al., 1989; Flug et al., 1985), particularmente na definição da linhagem (O'connor et al., 1985; Hare et al., 1989), avaliação da progressão da doença e recidiva (Minden et al., 1985; Minden and Mak, 1986; Griesser et al., 1989; Lee et al., 1987). Dois tipos de abordagens foram utilizados: a utilização de anticorpos idiotípicos, reconhecendo na sua maioria epitopes das regiões variáveis do TCR (Royer et al., 1987; Smith et al., 1988; Janson et al., 1991); e a detecção de rearranjos clonais por southern

Blotting (Flug et al., 1985), o que permite a detecção de clones representando pelo menos 1% da população celular em análise (Griesser et al., 1989; Minden et al., 1985). Em alguns casos, como a investigação da linhagem em leucemias linfoblásticas, parece mesmo haver vantagem na utilização conjugada de ambas as técnicas (Hare et al., 1989). Refira-se no entanto que a identificação de monoclonalidade não deve ser imediatamente interpretada como sinónimo de malignidade. Com efeito, em casos de hiperlinfocitoses CD8 ou CD4 (situação com curso benigno), foi possível encontrar monoclonalidade (Minden and Mak, 1986; Griesser et al., 1989; Cabeda resultados não publicados) sem que tal significasse uma posterior evolução maligna. Alguns autores sugerem mesmo que esta patologia possa ter início policlonal evoluindo posteriormente para uma situação monoclonal (Minden and Mak, 1986) não se observando no processo qualquer sinal de evolução para malignidade (Cabeda resultados não publicados). No entanto, a detecção de uma população clonal de células em proliferação pode ajudar a discriminar entre um processo reactivo e uma lesão neoplásica ou pré-neoplásica (Griesser et al., 1989).

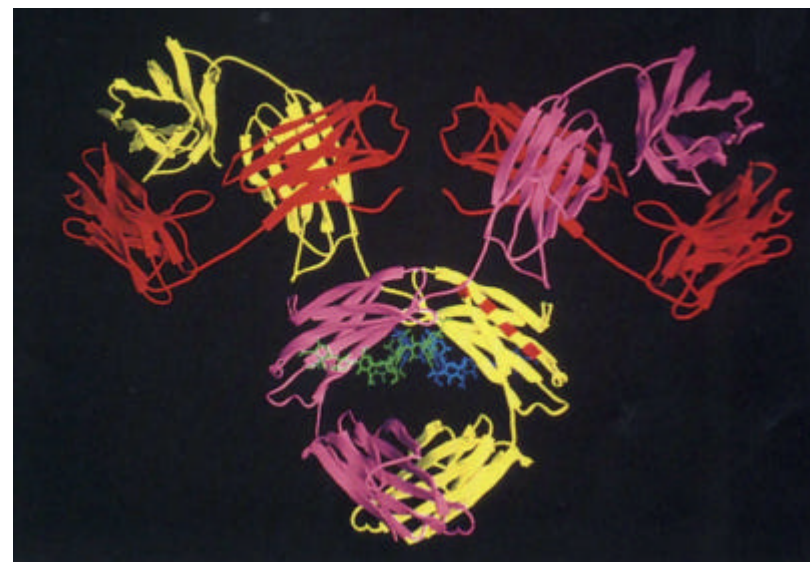
As mesmas técnicas utilizadas na investigação de clonalidade em doenças linfoproliferativas foram aplicadas à investigação de doenças autoimunes, imunodeficiências e outras (Royer et al., 1987; Posnett et al., 1988; Goldman et al., 1992). Exemplos destes estudos são os realizados para a sarcoidose pulmonar (Tamura et al., 1991; Forrester et al., 1994; Forman et al., 1994), doença de Chrons (Posnett et al., 1990), esclerose múltipla (Wucherpfenning et al., 1990; Ben-Nun et al., 1991; Nick et al., 1995; Droogan et al., 1994), Doença de Kawasaki (Abe et al., 1992), Miastenia Gravis (Infante et al., 1992), Artrite Reumatoide (Sottini et al., 1991; Howell et al., 1991; Uematsu et al., 1991; Haqqi et al., 1992; Posnett et al., 1988; Zagon et al., 1994), encefalomielite (Urban et al., 1988), doença tiroideia autoimune (Davies et al., 1992), doença de Gravis (Posnett et al., 1988), Trombocitopenia purpura Idiopática (Posnett et al., 1988), Psoríase (Posnett et al., 1988), Diabetes (Posnett et al., 1988; Wong et al.,

2.11

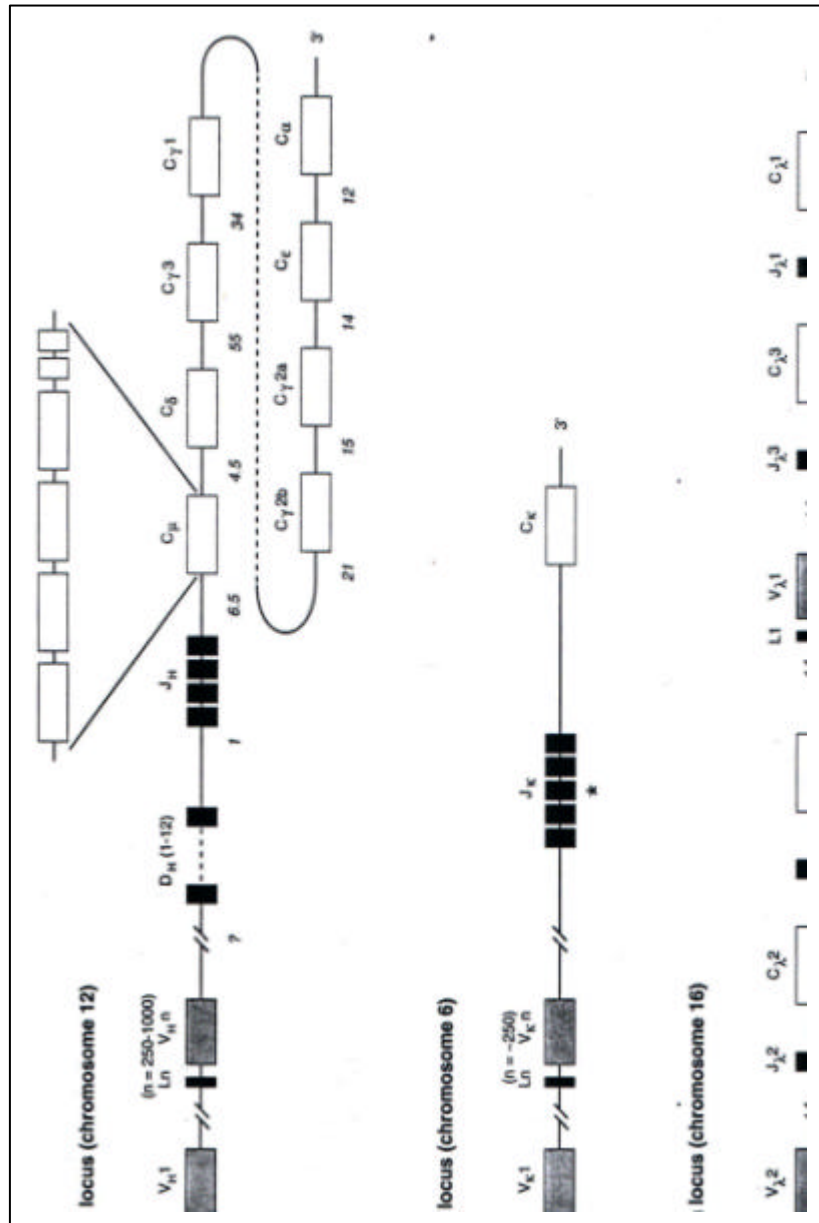
1994) Lupus eritematoso (Olive et al., 1994; Rozzo et al., 1994; Okubo et al., 1994 e “Coeliac Disease” (Falk et al., 1994).

Em todos estes estudos, anomalias do repertório do TCR foram encontradas quer nas células CD4 quer nas células CD8, devidas quer a diminuição (Goldman et al., 1992; Posnett et al., 1988) quer a aumento da frequência de utilização de determinados genes variáveis (restantes referências acima mencionadas). No entanto, nos casos em que a mesma patologia foi investigada por mais que um laboratório, os resultados nem sempre foram concordantes (Tabela 1), o que pode ser devido quer à falta de técnicas padrão, quer à extensa utilização de tecnologia de PCR quantitativo, a qual não se encontra ainda hoje suficientemente padronizada.

## 2.2 As Imunoglobulinas (Ig)



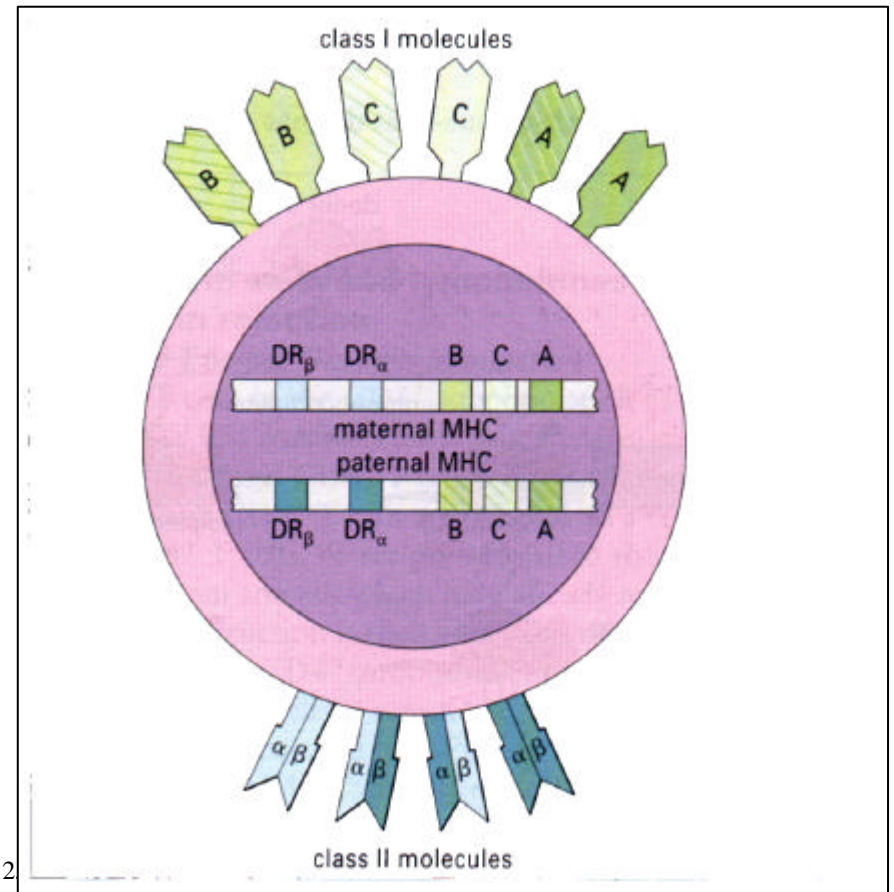
2.12



2.13

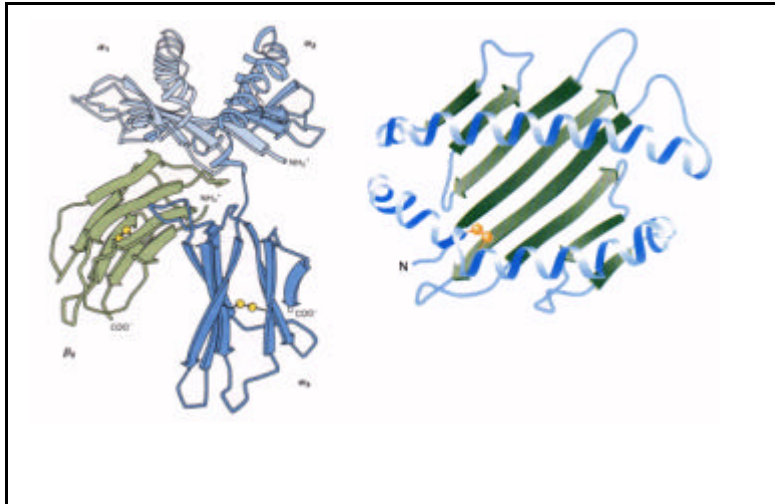
## 2.3 O MHC

Similarmente ao que sucede com os grupos sanguíneos, que determinam a compatibilidade das transfusões de sangue, um grupo de genes, globalmente conhecido como o Complexo Major de Histocompatibilidade (MHC) controla a compatibilidade dos tecidos em transplantes de órgãos. Estes genes, localizados no homem no braço curto do cromossoma 6, encontram-se concentrados numa única região ou locus, existindo genes homólogos em todas as espécies de mamíferos.



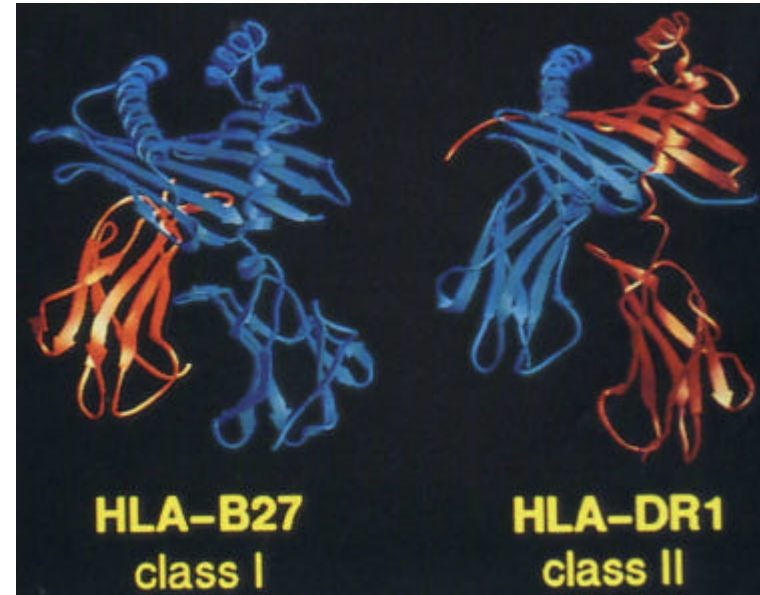
2

No homem, os antígenos produzidos por estes genes são globalmente designados por Antígenos de Histocompatibilidade Humana (HLA). Deve notar-se no entanto, que neste locus genético se encontram genes não envolvidos na histocompatibilidade, como é por exemplo o caso de genes do complemento.



A principal característica genética dos genes do MHC é o seu elevado grau de polimorfismo. Como já dissemos, um gene é polimórfico, quando existem alterações na sua sequência que, ou não alteram a sequência proteica, ou a alteram sem que a sua funcionalidade seja diminuído. No caso dos genes do MHC, os polimorfismos existem não só a nível genético, mas também a nível proteico, sem que a função destas proteínas saia diminuída, mas antes reforçada. Com efeito, como veremos, sendo a função dos genes do MHC a de apresentar aos linfócitos T o maior número possível de péptidos diferentes, esta só beneficia com a existência de um elevado número de alelos (a consequência do elevado grau de polimorfismo), já que os péptidos que um alelo não apresentar podem ser apresentados por outro alelo.

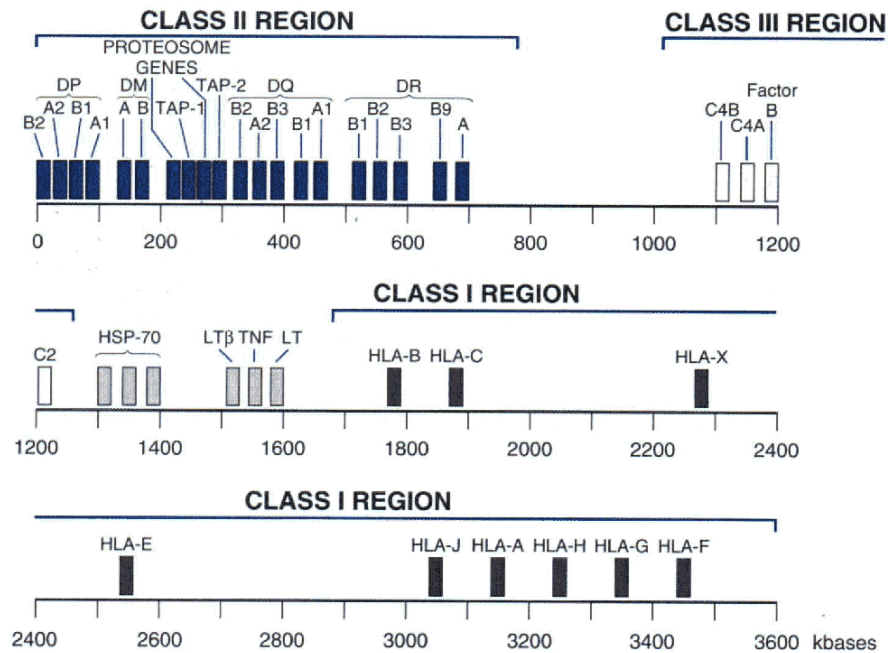
A hereditariedade dos genes do MHC efectua-se em bloco, já que estes genes se encontram agrupados num mesmo locus cromossómico. No entanto, ocorrência de recombinação entre os genes deste locus faz



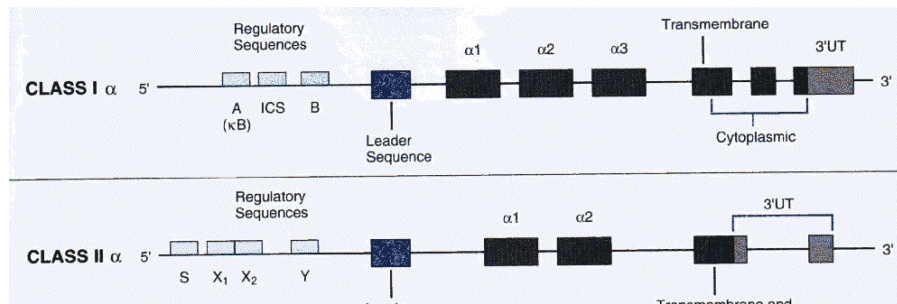
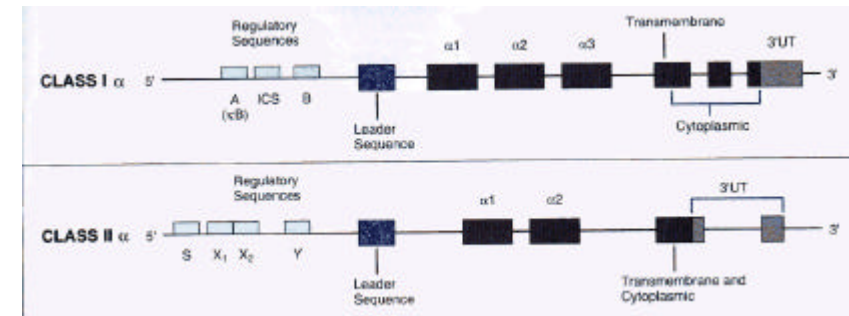
com que nem sempre a regra da transmissão em bloco se verifica, já que em alguns casos raros, a existência de recombinação entre os dois cromossomas homólogos de uma célula cria novas associações entre os genes.

Como a individualidade de um indivíduo, em termos de transplante de órgão, é definida pelo conjunto de alelos que possui para cada um dos genes do MHC, o elevado número de alelos de cada gene, e a esporádica ocorrência de recombinação concorrem para criar um elevadíssimo número de haplótipos (sequência alélica dos vários genes do MHC), criando as imensas combinações que constituem a individualidade dos órgãos para transplante.





imunológicas contra bactérias e parasitas. Entre estes dois grupos de genes, localizam-se os genes da classe III. Este loci, alberga genes do complemento e outros que não participam nas reacções de histocompatibilidade.



Os vários genes do MHC foram agrupados em 3 grupos, de acordo com a sua função e localização genómica. Os genes da classe I (no homem HLA-A, HLA-B, e HLA-C) constituem um grupo de genes localizados na zona 5' do complexo, e participam nas reacções imunológicas contra tumores e vírus. Os genes da classe II (DP, DQ e DR) localizam-se na zona 3' do complexo, e participam nas reacções

## 2.4 O Complemento

O sistema do complemento é constituído por mais de 30 proteínas séricas e da superfície das células, que interagem com resto do sistema imune otimizando as condições para que se possam realizar muitas das funções para a imunidade humoral. O próprio nome: complemento, foi atribuído graças à capacidade destas proteínas complementarem a acção dos anticorpos.

Este sistema tem várias funções no sistema imune :

- Algumas proteínas do complemento são intermediários da citólise levando à formação de poros na camada bilipídica de microorganismos da membrana celular, que provoca a lise osmótica dos microorganismos.
- Promove a opsonização de microorganismos ou de partículas estranhas. Esta função é exercida por proteínas específicas para esta função: as opsoninas. Os leucócitos fagocíticos têm receptores para as opsoninas promovendo a ligação dos leucócitos aos microorganismos e facilitando a fagocitose.
- Alguns fragmentos que resultam da fragmentação de certas proteínas do complemento activam a inflamação. Estes péptidos são chamados anafilotoxinas já que por vezes provocam uma resposta tão forte que se assemelha a uma reacção alérgica.
- O sistema do complemento é o responsável pela solubilização e pela destruição fagocítica dos complexos imunes que se poderiam depositar tanto nas paredes do vasos como nos órgãos.
- Finalmente, uma das funções mais importantes do complemento é promover a resposta humoral facilitando a apresentação do antigénio e baixando o nível de sensibilidade da activação das células B pelos antigénios.

Assim, o complemento desempenha uma papel muito importante no a ataque aos microorganismos, quer na activação da imunidade humoral quer directamente, levando á lise osmótica .

### 2.4.1 As cascatas do complemento

O complemento actua por cascatas enzimáticas que permitem amplificar a acção de uma só molécula, permitindo uma resposta imune mais eficaz. Algumas das proteínas do sistema do complemento são enzimas proteolíticas que actuam concertadamente levando à formação final do complexo MAC (Membrane attack complex) que provoca a lise osmótica dos microorganismos patogénicos (Tabela 2-1).

As cascatas têm duas vias de actuação: a via clássica e a via alternativa. Uma das proteínas chave nas duas cascatas é a proteína C3 (Figura 2.1.1-1). A proteína C3 é clivada em C3a que é uma anafilotoxina, e em C3b que funciona como convertase do C5. A partir da clivagem do C5 as duas cascatas são idênticas e culminam na formação do complexo MAC que se vai inserir na membrana dos microorganismos e provocar a lise osmótica (Figura 2.1.1-1). No entanto, antes da clivagem do C5 as duas cascatas seguem caminhos diferentes e são activadas por mecanismos diferentes.

A via alternativa é a mais antiga filogeneticamente mas assumiu este nome já que foi descoberta depois da via clássica. Esta via é activada na presença de microorganismos e não precisa de outras respostas imunes específicas, sendo portanto um mecanismo de imunidade inata. Os passos iniciais da via alternativa têm como objectivo a formação da C3 convertase que como já referi vai clivar o C3 e dar origem à C5 convertase. A C3 convertase da via alternativa é um complexo proteico formado por um produto da clivagem proteolítica do factor B (Bb) e de um fragmento do C3 que existe na circulação (C3b ou C3i). É o complexo C3Bb que actua como C3 convertase.

Tabela 2-1- alguns componentes proteicos das cascatas do complemento (continua na pag. seguinte)

Componente	Tamanho /kD	Concentração no soro/mg/mL	Produtos de activação	Funções
<b>Via alternativa</b>				
Factor B	93	200	Ba	Bb é uma serina protease que faz parte das C3 e C5 convertases.
Factor D	25	1-2		Protease que circula na forma activa: cliva o factor B ligado ao C3b.
Properdina	220	25		Estabiliza a C3 convertase da via alternativa
C3	195	550-1200	C3a C3b	C3a é uma anafilotoxina. C3b liga-se covalentemente a superficies activantes onde faz parte das C3 e C5 convertases e também actua como opsonina.

Tabela 2-2- alguns componentes proteicos das cascatas do complemento (continuação).

Componente	Tamanho /kD	Concentração no soro/mg/mL	Produtos de activação	Funções
<b>Via Clássica</b>				
C1	750			Inicia a via clássica: Liga-se a porção Fc das Ig
C1q	410	75		
C1r	85	50	C1r	Serina protease: cliva o C1s.
C1s	85	50	C1s	C1s é uma serina protease: cliva O C4 e C2.
C4	210	200-500	C4a C4b	C4a é uma anafilotoxina. C4b liga-se covalentemente a superficies activantes onde faz parte da C3 convertase e também age como opsonina.
C2	110	20	C2a C2b	Serina protease que faz parte das C3 e C5 convertases.

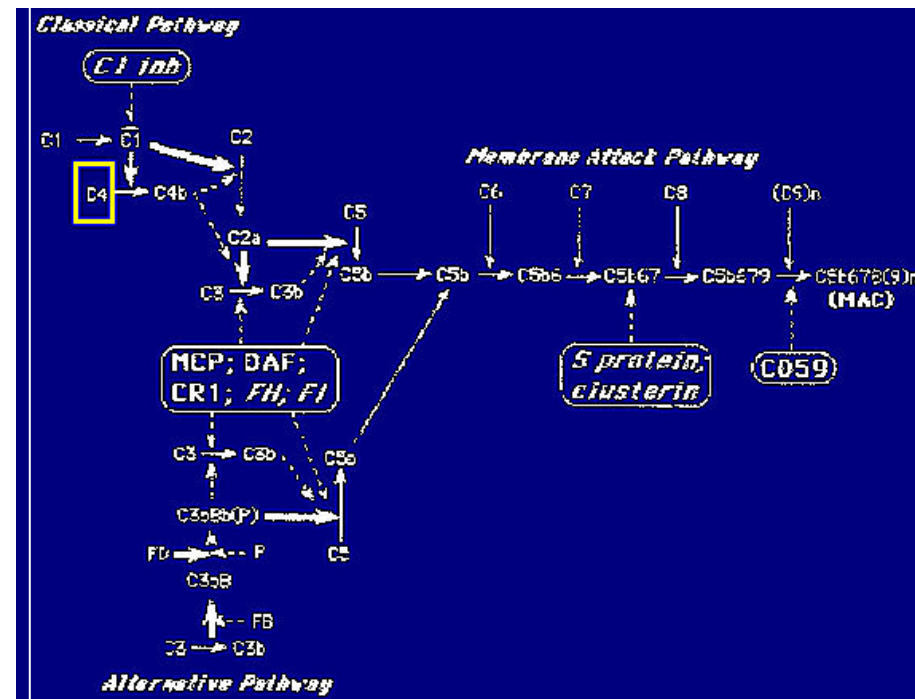


Figura 2.4.1-1As duas cascatas do complemento. As proteínas do complemento vão reagindo entre si para formarem o complexo MAC, que se liga aos microorganismos. A via alternativa é activada na ausência de anticorpos sendo um mecanismo relacionado com a imunidade inata. A via clássica por sua vez é um dos mecanismos essenciais na imunidade humoral.

Na via clássica a C3 convertase é formada na presença do complexo antígeno-anticorpo e depende da acção de uma série de proteínas chave (C1, C2 e C4) e é um mecanismo da imunidade humoral.

O C1 é proteína iniciadora da cascata e permite a activação da via clássica. O complexo antígeno-anticorpo liga-se à molécula C1 que

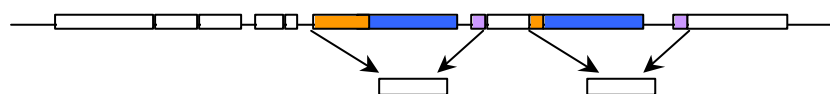
depois de activada, vai clivar o C4 em C4a e C4b. O C4b por sua vez vai clivar o C2 e dar origem à C3 convertase da via clássica-o complexo C4bC2a.

### 2.4.2 O MHC III

Os genes do complemento localizam-se num cluster do MHC III que, por sua vez se situa entre o MHC classe II e o MHC classe I.

Os genes do cluster do MHC III codificam para um série de proteína entre as quais as proteínas do complemento: factor B, C2 e C4; para enzimas: Citocromo P<sub>450</sub> 21- Hidroxilase, para a RP1. Estes genes estão organizados em módulos: clusters RCCX (Figura 2.1.2-2) que são formados pelos genes RP, C4 e a Citocromo P<sub>450</sub> 21-Hidroxilase. Estes clusters resultam de duplicação dos genes C4, RP, da enzima Citocromo P<sub>450</sub> 21-Hidroxilase e Tenascina X. Tal como aconteceu anteriormente, as mutações nestes genes normalmente envolvem todos os genes do cluster. Como vamos ver adiante estes acontecimentos podem ser importantes.

Figura 2.4.2-1 Organização dos genes do MHC III



#### 2.4.2.1 A proteína

O C4 é uma das proteínas da via clássica do complemento que participa na formação da C3 convertase, além de dar origem a uma anafilotoxina que promove a inflamação. A proteína é constituída por três cadeias polipeptídicas  $\alpha$ ,  $\beta$  e  $\gamma$  (figura 2.1.3-1). A pré-proteína é

constituída por 1744 aminoácidos, mas sofre uma clivagem posterior originando a proteína sérica activa.



Figura 2.4.2-2 Mapa do gene do C4. A zona assinalada com C4d compreende a zona de diferenciação dos dois isotipos (C4A e C4B):

AC CTC TCT CCA GTG ATA **CAT** AGG

Existem duas isoformas da proteína activa: o C4A e a C4B. As duas isoformas têm reactividades diferentes com grupos básicos ou ácidos. O C4A reage mais rapidamente com grupos amino, enquanto que o C4B reage mais facilmente com o grupo hidroxilo. Além disso a actividade hemolítica do C4B é três a quatro vezes maior que a do C4A. A diferença nas reactividades deve-se ao facto de existirem aminoácidos diferentes no local reactivo da proteína. No resíduo 1106 do C4B existe uma Histidina enquanto que no C4A há uma asparagina.

No caso do C4B o aminoácido nucleofílico (His) ataca o grupo tiol de um resíduo de cisteína, formando um intermediário ácido (figura xxx). No caso do C4A a asparagina, não sendo nucleofílica não origina a formação do intermediário pelo que a reacção de ligação às outras moléculas é mais lenta (figura).

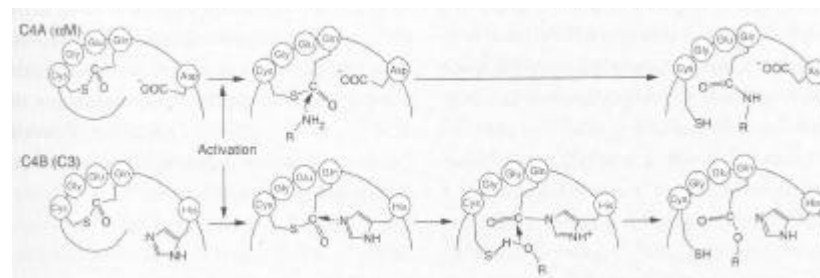


Figura xxx – Mecanismo de acção do C4A e do C4B.

As duas isoformas do C4 diferem apenas em quatro aminoácidos entre os resíduos 1101 a 1106.

A diferenciação funcional entre as duas isoformas dá-se mesmo no aminoácido 1106, já que mutações pontuais neste resíduo dão origem a conversões de C4B a C4A. No fundo estas diferenças fazem com que o C4A reaja mais facilmente com péptidos enquanto que o C4B reage com carboidratos. Devido a estas reactividades diferentes o C4b do C4A consegue opsonizar os fragmentos imunes mais facilmente que o C4b do C4B (Vaishnav *et al.*, 1998). Esta função é muito importante no caso de doenças imunes em que há uma grande acumulação de complexos imunes nos órgãos e nos vasos.

#### 2.4.2.2 O gene

- Localização e organização

O C4 é codificado por dois genes: C4A (ácido) e C4B (básico), localizados no braço curto do cromossoma 6, mais exactamente no locus 6p21.3. Os genes são constituídos por 41 exões com aproximadamente 21 Kb. O tamanho dos genes depende da existência do intrão 9 que tem 7 Kb. Assim quando o intrão está presente o gene tem 21 Kb enquanto que quando não existe intrão o gene tem 16 Kb.

- Polimorfismos e mutações

Tal como os genes do MHC III os genes do C4 são muito polimórficos e apresentam grandes variações quer em tamanho quer em número.

As mutações mais frequentes são os alelos nulos e são definidos pela ausência de proteína no plasma. Cerca de 30% da população possui alelos nulos para o C4: 10% apresentam o haplotipo C4AQ0 e 16% C4BQ0. Já a ausência total de C4 (C4AQ0 + C4BQ0) é muito rara e

está associada a formas graves de doenças imunes. Os alelos nulos podem resultar de deleções do gene ou de mutações que dão origem a codões STOP (1).

Estão referidas inserções de 2bp (TC) que dão origem a pseudogenes do C4 e originam genes nulos (Figura 2.1.3-2). Quando se dá uma deleção completa do gene normalmente ainda são abrangidos os genes vizinhos como é o caso da Citocromo P<sub>450</sub> 21-Hidroxilase (13).

```
Inserção TC L Y W G S Q S I V I R A.....STOP
5667      CTG TAC TGG GGC TCT CAG TCA CTG ATT CTC AGA GCA..... CGC TGA
          CTG TAC TGG GGC TCA GTC ACT GAT TCT CAG AGC AT.....
C4 Normal L Y W G S I
```

Figura 2.4.2-3 Inserção de TC duas bases que dão origem a uma proteína truncada do C4 não funcional.

Como já foi referido anteriormente, as diferenças de tamanho devem-se ao intrão 9 que tem 6.5Kb e que está presente em todos os genes do C4A (dando origem a um gene com 22Kb), mas que só está presente em alguns C4B fazendo com que o tamanho deste gene varie entre 22 e 16Kb. O intrão 9 contém um retrovírus endógeno (HERV-K) que está inserido no locus C4A e só em alguns C4B e que está espalhado por todo o genoma humano.

OS dois isotipos: C4A e o C4B, resultaram duma duplicação dos genes C4 (e provavelmente os genes anexos dando origem aos clusters RCCX). O retrovírus não se encontra nos macacos do Mundo Novo mas está presente nos macacos do Velho Mundo, revelando que a inserção do retrovírus se deu depois da divergência entre os macacos do novo mundo e nos macacos do mundo velho. Como este intrão existe em todos os genes do C4A mas apenas em alguns do C4B isto leva supor que a inserção do vírus se tenha dado depois da duplicação

dos genes. A existência do vírus em genes C4B pressupõe conversões génicas ou mecanismos de “crossing over” desigual.

Outro tipo de mutações a assinalar são também as interconversões génicas. Estão referidas na literatura casos em que mutações de uma só base transformam um dos genes do C4 no outro (Figura 2.1.3-3). Estas mutações são detectáveis por Southern blot já que o tamanho dos fragmentos que são reconhecidos como C4A ou C4B não corresponde ao valor esperado.

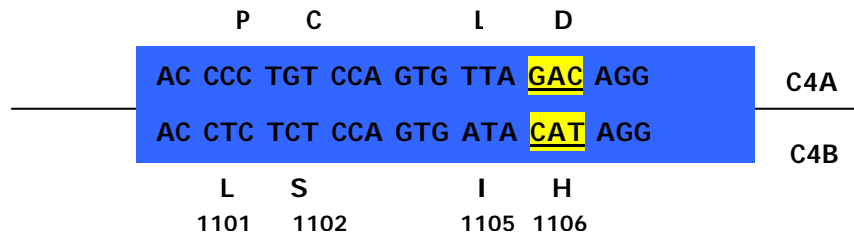
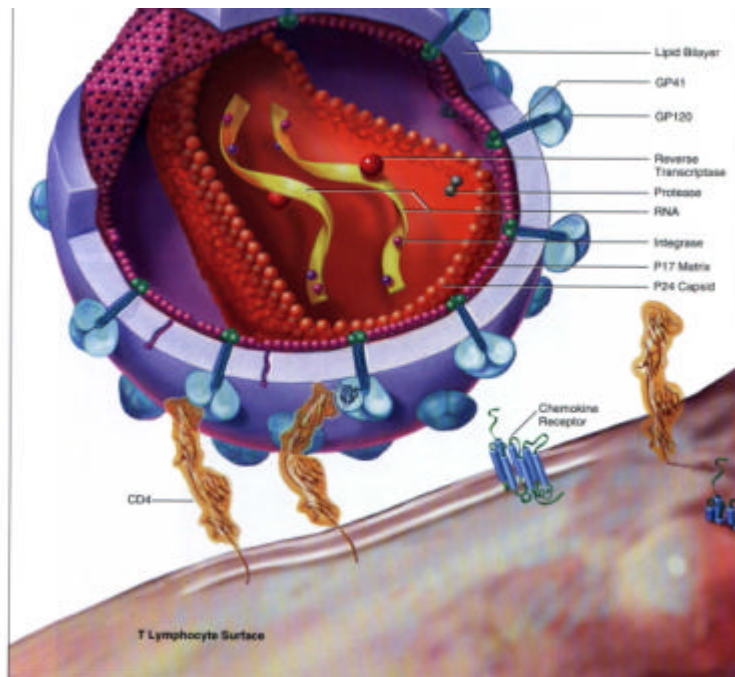


Figura 2.4.2-4 Mutação no gene C4B que o pode transformar no gene C4A. Apesar de existirem mais pontos de discordância entre os dois genes como nos aminoácidos 1101, 1102 e 1105, basta uma substituição de um aspartato por uma histidina (aminoácido 1106) para transformar funcionalmente o C4A em C4B.

# Capítulo 3

## 3 Áreas de intervenção da genética molecular em Microbiologia



3.1

### 3.1 A Genética na Virologia

As partículas virais completas (viríões) são constituídas por dois tipos fundamentais de material: um genoma e uma cápside. O genoma pode ser de DNA ou RNA, e encontrar-se ou não associado a proteínas. A cápside constitui um invólucro proteico que envolve o genoma e possui simetria helicoidal ou icosaédrica (fig. 3.1). A nucleocápside é por vezes revestida por uma envelope membranar. Alguns viríões possuem ainda enzimas, as quais são essenciais para as fases iniciais da replicação do genoma viral.

Os vírus têm um modo de multiplicação único, que os torna parasitas intracelulares obrigatórios. Na verdade, os vírus não são capazes de replicação autónoma, já que não possuem a maior parte da maquinaria enzimática necessária para esse efeito. Para obviar esta limitação, os vírus socorrem-se da maquinaria enzimática da célula hospedeira, introduzindo-lhe modificações que sequestram a sua actividade para o seu próprio genoma, em desfavor do genoma celular.

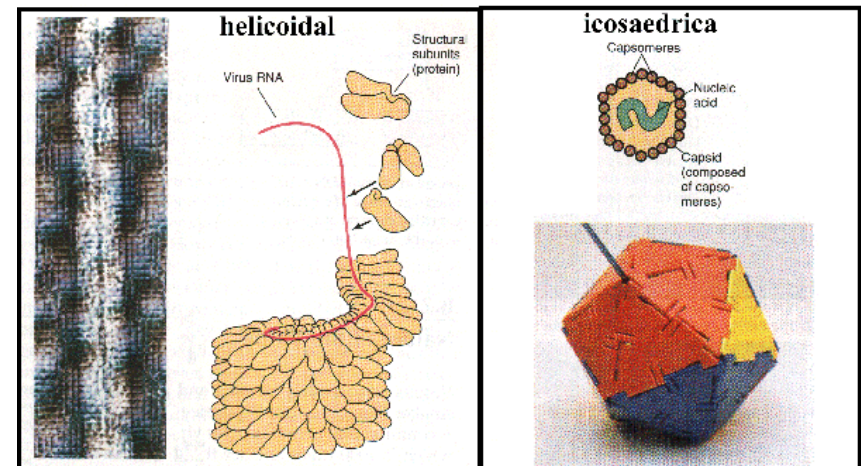


Figura 3.1 – Tipos de simetria viral, e respectivo empacotamento do material genético

3.2

Os vírus parasitam todo o mundo vivo, desde as bactérias às plantas e aos animais. Cada espécie de vírus possui no entanto uma especificidade de hospedeiro, a qual é extensível não só à espécie mas ainda ao tecido ou órgão que infecta. Esta especificidade de células hospedeiras, é determinada por um lado pela existência de receptores apropriados na membrana das células, e por outro pela existência de factores celulares necessários para a replicação viral.

### **3.1.1 - O genoma viral**

O genoma dos vírus é constituído por DNA ou RNA. O ácido nucleico viral pode ainda ser de cadeia simples ou dupla. A quantidade de material genético por virião varia entre 3 e 300 Kb. Assim, os vírus mais pequenos contêm apenas 3 a 4 genes, enquanto os maiores contêm algumas centenas. Exceptuando os retrovírus, os viriões contêm apenas uma cópia do genoma, isto é, são haplóides. Em alguns casos de vírus de plantas, cada virião contém apenas uma parte do genoma, sendo necessário a infecção de vários viriões numa mesma célula para, por complementarização, se reunir todo o genoma viral na célula infectada.

A dificuldade de replicar os terminais de uma molécula de DNA linear é ultrapassada em alguns vírus com a adopção de DNA de cadeia dupla circular como material genético, enquanto outros vírus, possuindo moléculas de DNA linear, circularizam-no após a infecção. Outros vírus possuem redundâncias terminais, que permitem que umas moléculas de DNA completem outras por recombinação pós-replicação. Uma outra estratégia adoptado por alguns vírus consiste em possuir sequências palindrómicas nas extremidades, as quais actuam como primers durante a replicação.

Alguns vírus possuem sequências génicas que reflectem a sua semelhança em comportamento aos transposons, como por exemplo a presença de "*terminal repeats*".

Os genes virais, sendo transcritos com a maquinaria enzimática celular, obedecem a princípios de regulação semelhantes aos genes celulares. Assim, possuem as estruturas gerais descritas no capítulo 1.1.1.3, ainda que por vezes, as sequências reguladoras sejam ligeiramente diferentes das celulares, conferindo aos genes virais vantagem selectiva na competição para a transcrição e tradução. Por exemplo, nos *Poxvirus* as sequências reguladoras da transcrição viral não seguem o padrão celular, necessitando de elementos reguladores codificados pelo genoma viral.

Os vírus de RNA possuem RNA de cadeia simples ou dupla. Os vírus de RNA de cadeia simples podem possuir cadeias positivas (o RNA do virião pode actuar directamente como mRNA) ou cadeias negativas (o RNA do virião tem que ser transcrito em mRNA). Os vírus de RNA+ têm habitualmente as características habituais do mRNA (CAP e poliA). O RNA dos vírus de RNA- não possuem CAP ou cauda poliA, mas possuem um terminal 5' com um nucleósido trifosfato.

O genoma dos Vírus de RNA de cadeia dupla é habitualmente segmentado, ainda que um segmento possa codificar para mais que uma proteína. A segmentação do genoma favorece a recombinação genética entre os vírus, originando conseqüentemente uma enorme variabilidade genética.

O genoma viral encontra-se densamente compactado dentro da cápside dos vírus icosaédricos. O DNA associado a proteínas ou poliaminas forma um core central de voltas paralelas. Em alguns vírus animais como os poliomavirus, o DNA associa-se fortemente a histonas centrais, formando uma estrutura semelhante à cromatina. O maior nível organizativo encontra-se nos vírus de DNA dos adenovirus. Nestes vírus, o DNA em conjunto com proteínas virais específicas forma doze bolas iguais, cada localizada num vértice da cápside icosaédrica.

Os vírus que possuem envelope possuem, para além das proteínas associadas ao genoma e das da cápside, proteínas do envelope. Estas podem ser de dois tipos: as glicoproteínas e as proteínas da matriz. As



glicoproteínas são proteínas transmembranares com grandes domínios extracelulares e domínios citoplasmáticos muito pequenos. As proteínas da matriz não são habitualmente glicosiladas. Algumas possuem vários domínios transmembranares, enquanto outras apenas se ligam à face interna das membrana por um domínio hidrofóbico. Estas proteínas, reforçam o envelope, e estabelecem pontes entre este e a nucleocápside. Em alguns casos, a estrutura fluida do envelope, e a quase total ausência de conexões à cápside fornece aos viriões uma forma pleiomórfica, enquanto em outros, a conexão firme entre o envelope e a cápside lhes confere formas características. As principais proteínas do envelope, tal como as da cápside e das proteínas associadas ao material genético são codificadas no genoma viral. Na realidade, a maioria dos antigénios virais são proteínas do envelope, cuja especificidade depende da sua origem no genoma viral. Alguns vírus contêm, também alguns componentes do envelope que não são codificados no genoma viral, sendo oriundos da membrana citoplasmática da célula hospedeira.

Para além dos componentes estruturais referidos, alguns vírus contêm enzimas, que executam funções essenciais nas fases iniciais da infecção. Tipicamente são enzimas que transcrevem o RNA viral para mRNA (em vírus de RNA- e em vírus de RNA de cadeia dupla), ou que transcrevem o RNA viral para DNA (nos retrovírus). Trata-se de enzimas específicas dos vírus que não estão presentes nas células a infectar, e que portanto devem ser fornecidas pelo virião que delas necessita.

### **3.1.2 - O ciclo de vida Viral**

#### **3.1.2.1 Passos iniciais da multiplicação viral**

Nos animais, toda a nucleocápside penetra na célula hospedeira, onde liberta o material genético, tornando-o disponível para a replicação. Assim, os passos iniciais da multiplicação viral são:

- **Adsorção** (o vírus liga-se à célula, habitualmente através de ligandos no envelope ou na cápside, e de receptores na superfície

celular. Este passo é bloqueado por anticorpos dirigidos quer contra o ligando no vírus, quer contra o receptor na célula).

- **Penetração** (segue-se rapidamente à adsorção, processando-se habitualmente por pelo menos três processos: 1) endocitose (o processo mais eficaz em que a penetração é mediada por receptores); 2) fusão do invólucro com a membrana plasmática; 3) translocação.
- **Descorticação** (o mecanismo mais frequente para a libertação do material genético consiste na desestabilização de componentes do envelope e cápside por interacção com componentes celulares, um fenómeno que ocorre por exemplo com a acidificação da vesícula de endocitose. A acidificação causa a exposição de aminoácidos hidrofóbicos, os quais se vão ligar à membrana vesicular, libertando o material genético no exterior da vesícula.

#### **3.1.2.2 Fase da Multiplicação viral**

A fase que se segue à libertação do material genético no citosol da célula alvo, é a da multiplicação dos vírus. Nesta fase, os vírus virulentos bloqueiam a síntese de proteínas celulares, forçando a maquinaria celular a sintetizar exclusivamente as proteínas virais. A replicação e transcrição do DNA celular são também inibidos nos adenovírus, herpesvírus, e poxvírus, sendo ainda observados, com frequência, cortes cromossómicos.

O mecanismo de bloqueio da síntese proteica celular é muito variável. Por exemplo, o poliovírus possui uma protease viral que lisa uma proteína celular necessária para o reconhecimento do CAP. Como o RNA deste vírus não possui o CAP, não é afectado. Os Herpesvirus causam a degradação do mRNA celular, enquanto o gene E1b dos adenovirus impede o transporte do mRNA celular para o citoplasma. Na maioria dos casos, estes mecanismos estão dependentes da síntese de proteínas virais, para se iniciar o bloqueio da síntese proteica celular.

Ao contrário dos vírus virulentos, os vírus moderados não só não inibem a síntese proteica celular como podem estimulá-la.

#### 3.1.2.2.1 Fase da síntese de vírus de DNA

Durante a fase da síntese, ocorrem três processos genéticos altamente regulados: a transcrição, a tradução e a replicação. No processo de síntese dos vírus animais, e ao contrário do que sucede nos procariontes, a transcrição e a tradução não estão fisicamente ligadas, já que, com a exceção do *poxvírus*, a transcrição ocorre no núcleo, e a tradução no citoplasma.

**Transcrição:** Os mRNA produzidos são como os celulares monocentrados, sendo também sujeitos aos habituais mecanismos de processamento pós-transcricional. A transcrição tem uma organização temporal. Na maioria dos vírus de DNA apenas uma fracção do genoma é transcrito na fase inicial (*early genes*), sendo os restantes genes transcritos numa fase posterior (*late genes*). Os vírus mais complexos têm ainda os *immediate early genes*, os quais são expressos na presença de inibidores da síntese proteica, bem como *delayed early genes*, os quais requerem síntese proteica para a sua expressão. A passagem dos *early genes* aos *late genes* necessita da prévia replicação do DNA.

A regulação da transcrição é efectuada por proteínas presentes no virião, ou codificadas por genes específicos virais ou celulares. A sequência 5' de cada gene viral encontra-se envolvida nesta regulação, estendendo-se por vezes durante algumas centenas de bases a região reguladora. Estas regiões possuem semelhanças funcionais com os enhancers celulares, respondendo em *trans* a substâncias produzidas por outros genes, e em *cis* ao gene que regulam

A transcrição inicial, e o processamento da maioria do mRNA são efectuados pela maquinaria enzimática celular. A exceção são os *poxvírus* cujo genoma não migra para o núcleo. Estes vírus, cujo

genoma não possui introns, utilizam a maquinaria enzimática mais simples codificada no seu próprio genoma, e transportada no virião.

**Tradução:** As proteínas virais são sintetizadas nos polisomas citoplasmáticos, numa sequência temporal correspondente à descrita para a transcrição. Assim, numa fase inicial são transcritos essencialmente proteínas reguladoras (polimerases, inibidores da tradução e transcrição celular, activadores dos *late expressing genes*), enquanto na fase tardia, são produzidas essencialmente proteínas estruturais do virião. Na maioria das classes víricas, estas proteínas migram para o núcleo onde ocorre a reunião das nucleocápsides.

**Replicação:** A replicação utiliza percursores do meio celular. Os vírus mais simples utilizam a polimerase do DNA da célula hospedeira, enquanto os vírus mais complexos como os adenovírus, os herpesvírus e os *poxvírus* possuem as suas próprias polimerases para esse efeito. A síntese inicia-se cerca do meio do período de eclipse, após a transcrição dos *early genes*. A replicação viral é semiconservativa, mas o mecanismo preciso de replicação depende do tipo de vírus:

Adenovirus: Apresentam replicação assimétrica, que se inicia no terminar 3' de uma das cadeias de DNA. A cadeia em síntese utiliza como primer um ácido citidílico ligado ao precursor da proteína terminal. Após a completa síntese desta cadeia, a cadeia oposta é também replicada de forma semelhante, após formar uma estrutura tipo *hairpin* por emparelhamento das repetições invertidas.

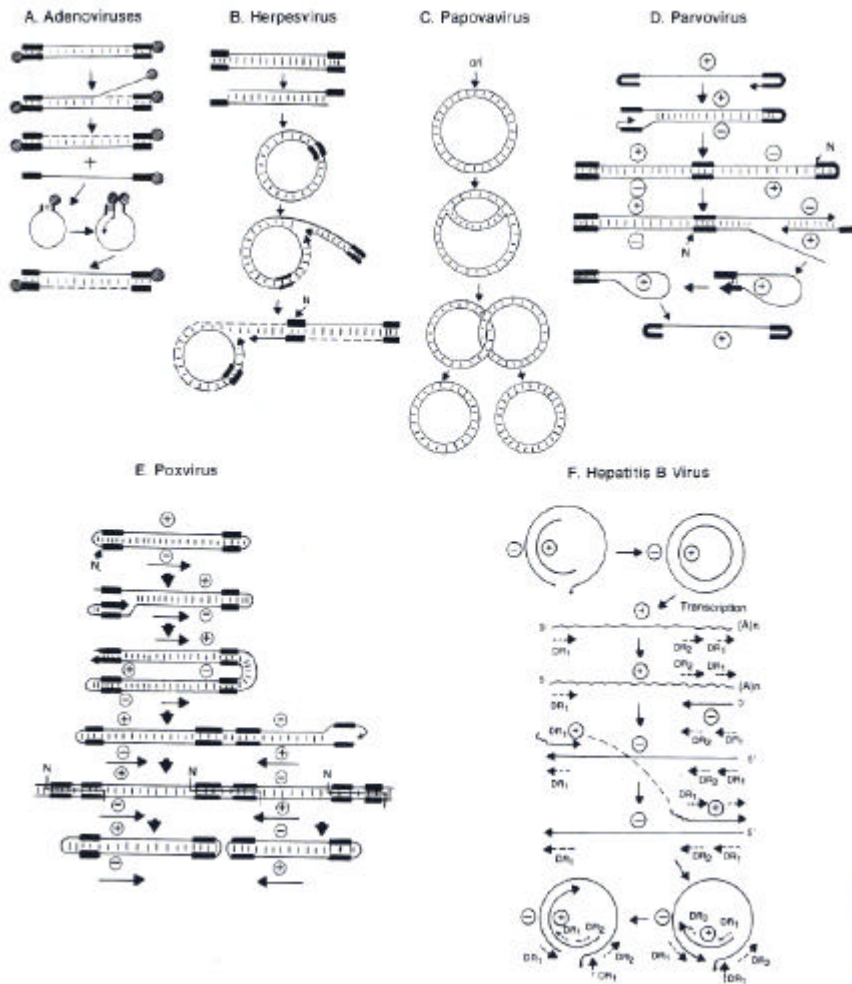


Figura 3.2 – Tipos de vírus de DNA e respectivos mecanismos de replicação

**Replicação com recurso a intermediários circulares:** Como já foi dito, alguns vírus utilizam intermediários circulares, para conseguir replicar na totalidade o seu genoma. É o caso dos *herpesvírus*, os quais possuem um genoma de DNA de cadeia dupla, com cerca de 300-400 bp de repetições directas nas extremidades. Após a

infecção, o DNA migra para o núcleo da célula hospedeira, onde sofre exonucleólise limitada, que lhe permite formar estruturas circulares por emparelhamento das extremidades. Numa fase inicial, a replicação, origina monómeros circulares, enquanto numa fase posterior se formam concatâmeros circulares, os quais são posteriormente cortados originando o DNA linear dos futuros viriões.

**Poxvírus:** O Poxvírus possui um genoma de DNA bastante extenso com uma peculiaridade única. As duas cadeias de DNA complementar são covalentemente unidas nas extremidades. A ligação ocorre no fim de duas sequências repetidas e invertidas (*inverted repeats*). Esta organização permite ao vírus replicar-se de forma contínua, originando produtos poliméricos que são posteriormente clivados em monómeros e de novo fechados nas extremidades.

**Vírus da hepatite B:** Apesar de ser um vírus de DNA, utiliza a transcrição reversa para a sua replicação. Os viriões contêm DNA circular, parcialmente de cadeia dupla. A cadeia negativa (complementar do mRNA) é completa, apresentando-se de forma circular com um corte a garantir a descontinuidade. A cadeia positiva é incompleta. Após a infecção, a cadeia positiva é completada, gerando-se uma cadeia completa circular e fechada, a qual é transcrita. O ácido nucleico é então transcrito de forma reversa para DNA por uma enzima viral, num processo complexo que envolve um salto entre dois *direct repeats*, originando a cadeia positiva incompleta.

### 3.1.2.2.2 Fase da síntese de vírus de RNA

Os vírus de RNA podem ser divididos em 7 classes, de acordo com a polaridade do seu RNA (o mRNA tem polaridade positiva), bem como o número de moléculas de RNA, e se este é de cadeia simples ou dupla. (tabela 3.1).

A estratégia de síntese destes vírus é ditada pela ausência de unidades de tradução múltiplas num mesmo mensageiro. Para obviar a esta dificuldade três estratégias foram desenvolvidas: 1) o virião de RNA actua como mensageiro, e é transcrito de forma monocistrónica num péptido gigante, que é depois cortado originando as proteínas funcionais; 2) O virião de RNA é transcrito em vários mRNA monocistrónicos por iniciação da transcrição em pontos diferentes do RNA viral; 3) O genoma do vírus é uma colecção de várias moléculas

Tabela 3.1 - Tipos de vírus de RNA

Classe	Constituição do genoma	Transcriptase no virião	Infectividade do RNA	Exemplo
I	+SS	-	+	picornavirus
II	+SS	-	+	coronavirus
III	-SS	+	-	paramixovirus
IV	-SS MM	+	-	ortomixovirus
V	PM	+	-	arenavirus
VI	DS MM	+	-	reovirus
VII	+SS MM	+	-	retrovirus

+ polaridade do mRNA

- polaridade oposta à do mRNA

CS cadeia simples

DS cadeia dupla

MM múltiplas moléculas de RNA

PM polaridade múltipla

de RNA, sendo cada uma transcrita em espécies de mRNA monocistrónicos.

Nos vírus da **classe I** o RNA actua como mensageiro. A mesma molécula tem ainda que iniciar a replicação, o que necessita de proteínas codificadas pelo RNA viral. Assim, a organização temporal tradução/replicação é imposta pela necessidade de proteínas não estruturais, mas essenciais para a replicação.

Nos vírus da **classe II**, o genoma origina primeiro um transcrito de polaridade negativa, o qual é então transcrito em mRNA monocistrónicos de diferentes tamanhos, num processo único. Cada um inicia com uma curta sequência *leader* 5', a qual é adicionada aos

transcritos correspondentes aos vários genes, continuando a sequência do mRNA até ao final 3' do genoma.

Os vírus da **classe III**, têm genoma com polaridade negativa, uma transcriptase vírica transcreve o genoma em vários mRNA monocistrónicos, a partir de um único promotor. A transcriptase pára e reinicia a transcrição em cada junção génica, originando os vários mRNA.

Os vírus da **classe IV** têm a particularidade de possuir o seu genoma disperso por várias moléculas de RNA, sendo cada segmento transcrito em moléculas de mRNA distintas. Em alguns casos, um segmento dá origem a dois produtos genéticos por splicing alternativo. Por este motivo o ciclo de vida destes vírus passa por uma fase nuclear, já que o splicing de RNA é essencial para a sua fisiologia.

Os vírus da **classe V** têm um genoma peculiar, já que metade tem polaridade negativa, sendo transcrito em mRNA por uma transcriptase vírica, enquanto a outra metade do genoma tem polaridade positiva, sendo transcrito duas vezes até se obter o mRNA.

Os vírus da **classe VI** possuem segmentos múltiplos não sobreponíveis de RNA de cadeia dupla. Cada segmento é transcrito num mRNA diferente por uma transcriptase vírica.

Os vírus da **classe VII** são os únicos vírus de RNA cujo ciclo de vida passa por uma fase em que o genoma existe sob a forma de DNA. Estes vírus possuem dois RNA de cadeia simples idênticos de polaridade positiva com uma cauda poli-A na extremidade 3', e um CAP na extremidade 5'. Cada um destes RNA é transcrito reversamente em DNA por uma transcriptase reversa presente no virião. O DNA resultante é então transcrito em mRNA para a actividade vírica.

Nos vírus de RNA não existe diferenciação entre *early genes* e *late genes*, com a possível excepção dos reovirus (classe V).

**Replicação dos vírus das classes I a V:** A replicação, que ocorre no citoplasma consiste em todos os casos em produzir uma cadeia padrão, complementar do genoma, e com o mesmo comprimento que este, a qual é depois utilizada como padrão para a síntese do RNA genómico da progenia. Em alguns casos, a replicação e a transcrição interferem-se mutuamente, pois a falta inicial de proteínas virais tem como consequência que só a transcrição é possível, mais tarde, a mesma cadeia de RNA deve mediar os dois processos. Assim, por vezes os vírus possuem proteínas específicas, que se ligam ao RNA, promovendo a replicação. Na replicação do RNA, a cadeia sintetizada permanece associada à cadeia padrão, formando uma molécula de RNA de cadeia dupla com o comprimento do genoma viral (RNA conhecido por RF de *replicating form*). A síntese de novas cadeias realiza-se de forma conservativa e assimétrica: a próxima síntese remove a cadeia formada de novo, a qual se liga a proteínas da cápside, gerando uma nova nucleocápside.

**Replicação dos vírus da classe VI:** Cada segmento de RNA do genoma destes vírus é replicado independentemente dos restantes. A replicação está intimamente ligada à transcrição pela transcriptase viral. A replicase usa o mRNA nascente como padrão, fazendo a partir deste uma cadeia de polaridade negativa, a qual actua então como padrão para a síntese do RNA genómico de polaridade positiva. As duas cadeias permanecem ligadas, sendo incorporadas na cápside viral. esta replicação é assimétrica (a cadeia negativa é a única a servir de padrão para a síntese do novo genoma) e conservativa (o RNA original não integra as novas partículas virais).

**Replicação dos vírus das classes VII:** O ciclo de vida dos retrovírus é objecto de descrição alargada na secção seguinte. Cabe aqui apenas referir que sinteticamente a replicação parte do intermediário de DNA de cadeia dupla (integrado no genoma celular), sendo a partir deste produzido o RNA por transcrição normal.

### 3.1.2.3 O ciclo de vida dos retrovírus

Pela sua importância nas infecções crónicas correntes com maior relevo em saúde pública, os retrovírus, e em particular o HIV, têm sido alvo de intensa investigação, pelo que muito se sabe já acerca das especificidades do seu ciclo de vida.

Como já vimos o genoma dos retrovírus é constituído por RNA de cadeia simples de polaridade positiva. Tal como o mRNA celular possui uma cauda poliA na extremidade 3', e uma estrutura CAP na

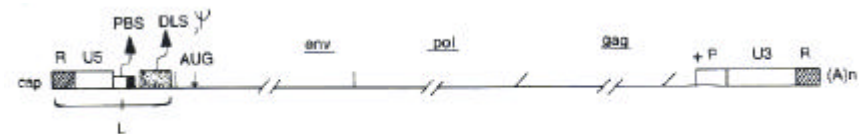


Figura 3.3 – Genoma tipo de um retrovírus.

extremidade 5'. Cada virião possui duas moléculas de RNA (é diplóide), com a mesma polaridade, juntas por uma estrutura de dimerização na extremidade 5'. O virião contém ainda alguns RNA celulares de baixo peso molecular entre os quais tRNA, os quais, como veremos, possuem funções essenciais no ciclo de vida dos retrovírus.

Os retrovírus possuem essencialmente 3 genes básicos: *gag*, *pol* e *env* respectivamente pela ordem 5'→3'. Os onco-retrovírus humanos possuem ainda um gene adicional, e os lentivirus possuem cinco genes adicionais. Para além das estruturas CAP e poli-A, as extremidades genómicas dos retrovírus possuem ainda, na ordem 5'→3', uma redundância terminal (R), a sequência U5 (do inglês *unique 5'*) e o local de ligação do primer (PBS do inglês *primer binding site*). O PBS possui 16 a 18 nucleótidos e é complementar de um da extremidade 3' de um tRNA que a ele se encontra ligado, e que funciona como primer para a transcrição reversa.

O codão de iniciação AUG do gene gag localiza-se algumas centenas de nucleótidos depois do local CAP, e neste intervalo encontram-se ainda duas sequências importantes o local de dimerização (DLS do inglês *dimer linkage site*) o qual mantém juntas as duas cópias do vírião, e o sinal de empacotamento ( $\Psi$ ) que permite o empacotamento

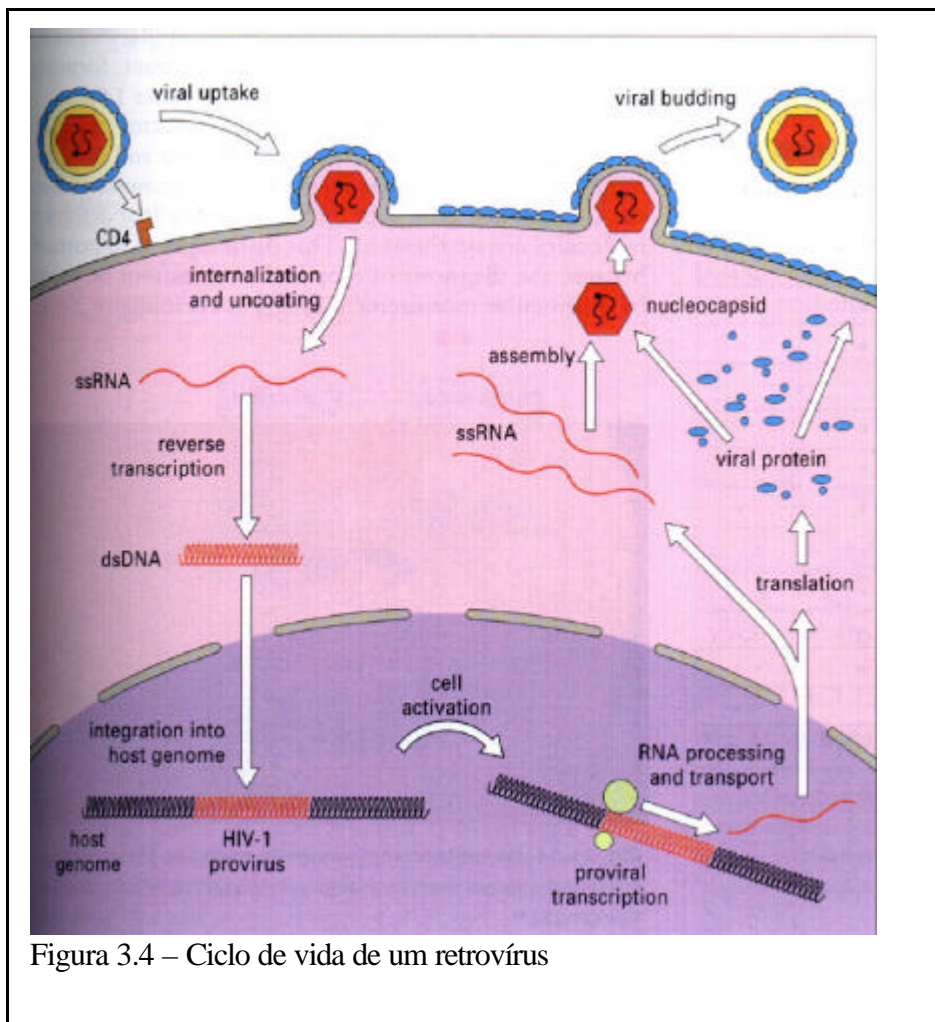


Figura 3.4 – Ciclo de vida de um retrovírus

do RNA dos víriões.

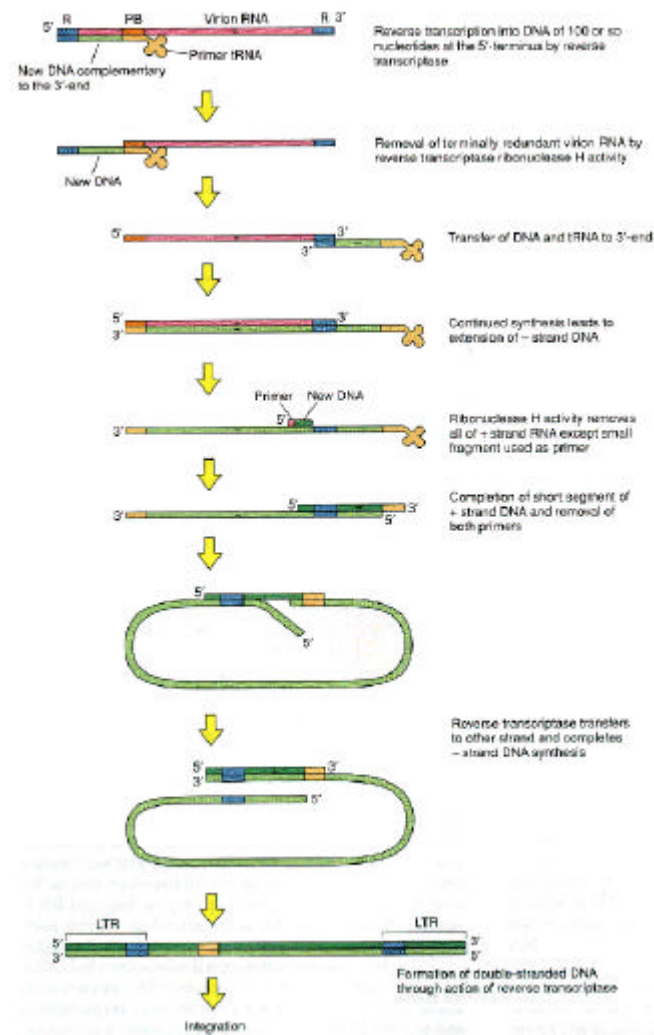


Figura 3.5 – Mecanismo de transcrição reversa nos retrovírus

A extremidade 3' contém após o final do gene *env*, a região de primer + (+P) com 12 bases ricas em purinas (intervém na transcrição reversa) a sequência U3 (*unique 3'* – contém sequências importantes para a transcrição viral), e redundância terminal (R) semelhante à da extremidade 5' e finalmente a cauda poliA.

Esta organização geral com os terminais repetitivos a circunscrever os genes é semelhante à dos transposons eucarióticos.

A adsorção e entrada dos retrovírus nas células hospedeiras não é diferente dos restantes vírus, dependendo da existência de receptores apropriados na superfície das células alvo.

A maioria dos retrovírus não são nem citopáticos, nem alteram grandemente o metabolismo das células que infectam. Em cultura as células infectadas continuam habitualmente a dividir, ao mesmo tempo que produzem e libertam partículas virais.

Pouco depois da infecção (menos de uma hora em células de galinha infectadas com um retrovírus compatível) tem lugar a síntese de uma cópia de DNA de polaridade negativa por transcrição reversa do RNA viral (utilizando a transcriptase reversa viral). Ainda antes de estar completa a cadeia de DNA de polaridade negativa, tem início a síntese da cadeia de DNA de polaridade positiva (uma reacção também mediada pela transcriptase reversa viral). Quando termina a síntese da cadeia de polaridade negativa, a parte do RNA que permanece a ela ligada é degradada pela RNase H. Inicialmente a cadeia de DNA nascente é uma cadeia dupla linear com cortes, já que a cadeia – é contínua e a + descontínua.

Cerca de 6 a 9 horas após a infecção os cortes são preenchidos, e o DNA desloca-se para o núcleo sob a forma de DNA de cadeia dupla circular. Cerca de 24 horas após a infecção várias cópias deste DNA encontram-se integradas no genoma celular como **provírus**, em locais do genoma aparentemente aleatórios.

As moléculas de RNA para os novos viriões são então produzidas por transcrição normal dos provírus integrados. Como consequência deste

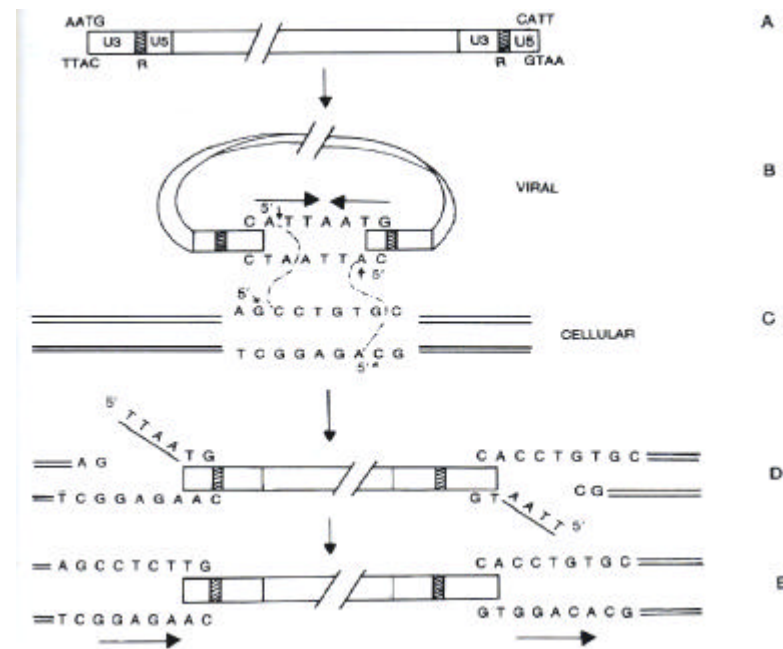


Figura 3.6 – Inserção de um retrovírus no genoma celular

método de replicação, o DNA de cadeia dupla é diferente do RNA do virião, possuindo mais cerca de 500 a 600 nucleótidos. As diferenças consistem nas duas extremidades que se tornaram iguais. As estruturas que medeiam este processo são designadas por LTR (*long terminal repeats*) consistindo cada (na direcção 5' → 3') por um U5, um R e um U3. Cada LTR contém um sinal para a adição de um CAP, uma TATA box e um local de CAP determinando a origem da transcrição, bem como um sinal para a adição de uma cauda poliA.

O provírus utiliza apenas os sinais de iniciação do LTR 5' porque a transcrição a partir deste interfere com a iniciação da transcrição a partir do LTR 3'. Este contribui com os sinais de terminação, e pode iniciar a transcrição se faltar o LTR 5'. Tal como as sequências presentes nos transposons eucariotas, os LTR virais possuem em cada extremidade *short direct repeats* do local de integração (5 a 13

nucleótidos). Uma outra importante característica dos LTR é a existência de um enhancer.

Os LTR são cruciais no processo de integração no genoma celular. Apenas moléculas circulares com os LTR ligados extremidade-com-extremidade são passíveis de integração. A integrase corta as duas cadeias de DNA na junção U5-U3, faz dois cortes no DNA genómico, em locais ligeiramente diferentes, e insere o DNA viral. Após a inserção, ficam alguns nucleótidos de DNA celular em cadeia simples, o que é reparado. Após este passo, o provírus é 4 nucleótidos mais pequeno que o DNA circular e é rodeado pelos *direct repeats* do DNA celular resultantes da reparação do DNA de cadeia simples. Estes provírus podem ser excisados por recombinação entre os LTR, mas este é um processo extremamente raro.

A produção de proteínas virais inicia-se em simultâneo com a síntese do DNA. Existem basicamente duas espécies de RNA traduzido. A mais pesada corresponde à espécie presente no virião, e origina a *gag* e em alguns casos (5%) a *pol* (o gene *pol* encontra-se out-of-frame, pelo que só em

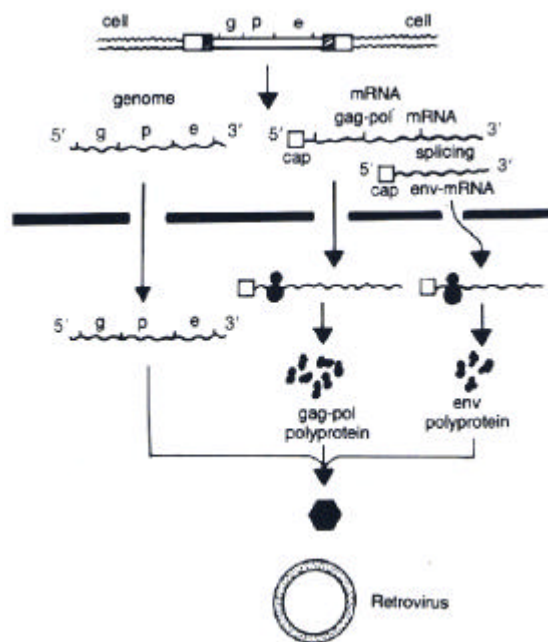


Figura 3.7 – Produção das partículas virais do retrovírus

ribossomas que executaram o frame-shift a *pol* é passível de leitura). Uma outra espécie de RNA resultante do splicing da primeira dá origem à proteína *env*. As proteínas produzidas (*gag-pol* e *env*) são na realidade poliproteínas que serão posteriormente clivadas dando origem às proteínas virais maduras. A protease que efectua este corte é uma protease viral específica codificada entre os genes *gag* e *pol*.

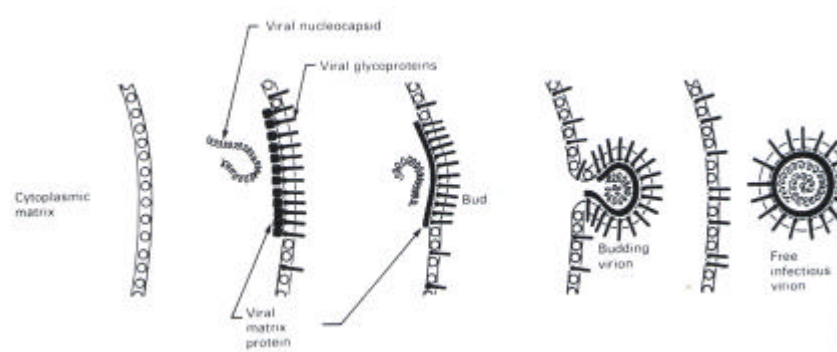


Figura 3.8 – Saída das partículas virais por evaginação.

A poliproteína *env* entra no retículo endoplasmático durante a síntese, dirigindo-se depois para o golgi onde é glicosilada, sendo então canalizada para a membrana celular. A maioria da *gag* fica retida no citosol, mas uma pequena parte segue a mesma via da *env*.

Cerca de 8 horas após a infecção, os percursores *gag* e *gag-pol* começam a juntar-se ao RNA viral, formando uma nucleocápside debaixo da membrana plasmática, onde após ocorrer a proteólise a nucleocápside se liga ao *env*. Esta poliproteína é então clivada originando uma proteína transmembranar da matriz e uma proteína externa, ligadas por pontes dissulfureto. O virião então formado é excretado por evaginação, o que permite manter a integridade celular.



### 3.1.3 - Virologia em Medicina Transfusional

#### 3.1.3.1 - O vírus da Hepatite B (HBV)

O soro de doentes com hepatite B possui habitualmente 3 estruturas diferentes com o antígeno Hb<sub>s</sub>Ag. A partícula de Dane é a menos comum, possuindo todas as estruturas observadas em vírus (core electrónicamente denso revestido por um envelope), e é infecciosa. As partículas esféricas são as mais numerosas, sendo diferentes dos restantes vírus; a sua superfície parece ter uma estrutura semelhante à da partícula de Dane, mas sem simetria aparente. As formas filamentosas são também frequentes. Quando estas estruturas filamentosas são sujeitas à acção de um detergente aniónico, originam estruturas indistinguíveis das partículas esféricas, sugerindo que esta estrutura é um aglomerado de partículas esféricas.

O genoma do HBV é constituído, como já foi referido, por DNA

circular de cadeia dupla incompleta. Assim, possui uma cadeia longa (L), de comprimento constante, com uma extremidade 3' livre, e com uma extremidade 5' ligada covalentemente a uma pequena proteína. O vírus possui ainda uma cadeia curta (S de *short*) cujo comprimento varia entre 15% e 60% do comprimento do genoma circular. A posição dos terminais 5' de ambas as cadeias são fixas, mas não coincidentes, sendo a posição do terminal 3' da cadeia S variável. O emparelhamento das duas cadeias, assegura que o genoma assume uma configuração circular, já que as cadeias não possuem terminais coincidentes.

A nucleocápside é formada por uma única proteína - o antígeno do core (AB<sub>c</sub>Ag). Esta proteína possui actividade de cinase, sendo capaz de se autofosforilar; um domínio desta proteína, rico em argininas (a

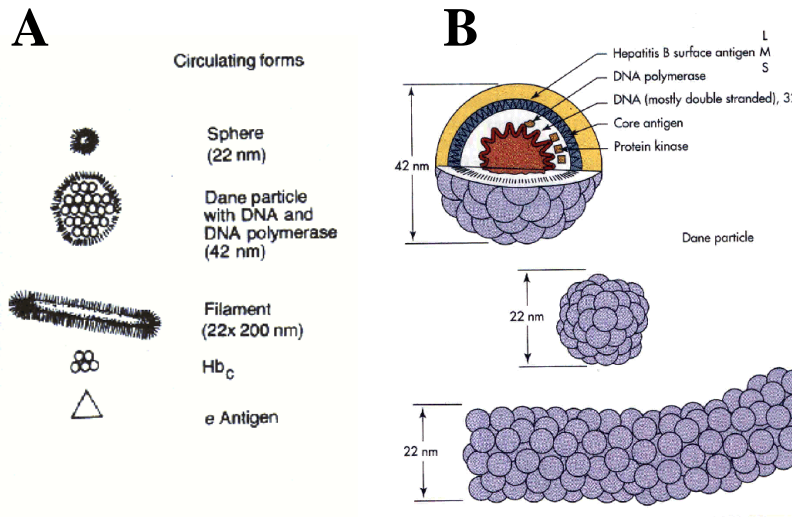


Figura 3.9 – O vírus da Hepatite B existe em circulação sob a forma de várias Partículas (A). A partícula de Dane (B) é uma partícula completa e infecciosa.

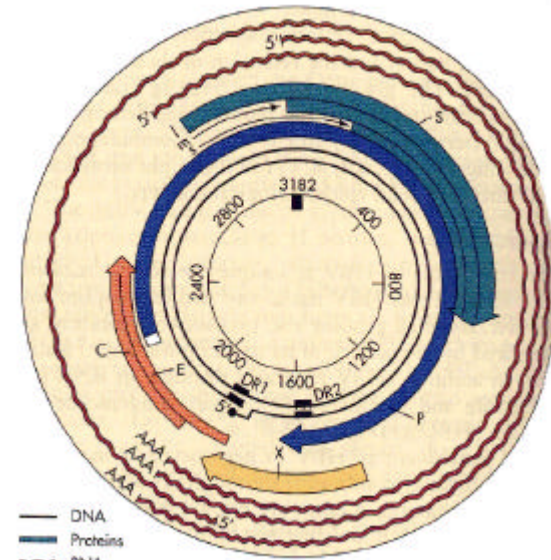


Figura 3.10 – O genoma do HBV, espécies de mRNA dele derivadas e respectivas proteínas

zona carboxi-terminal) interage com o DNA. No seu interior a nucleocápside contém ainda o DNA, a DNA polimerase e uma proteína cinase. A polimerase viral depende dos dNTP e  $Mg^{2+}$  celulares, mas é capaz de produzir as duas cadeias de DNA viral sem necessitar de qualquer primer. A polimerase tem ainda a função de transcriptase reversa, como veremos abaixo.

O envelope viral contém 3 glicoproteínas designadas *major*, *midle* e *large proteins*. O antígeno de superfície ( $HBV_sAg$ ), que é o componente principal das partículas esféricas é um dímero da *major protein*.

A entrada do virião no hepatócito é provavelmente mediada pela ligação da *middle protein* à albumina polimerizada. Após a entrada no hepatócito, e a libertação do genoma no citoplasma, este dirige-se ao núcleo, onde com a ajuda da DNA polimerase viral produz um DNA

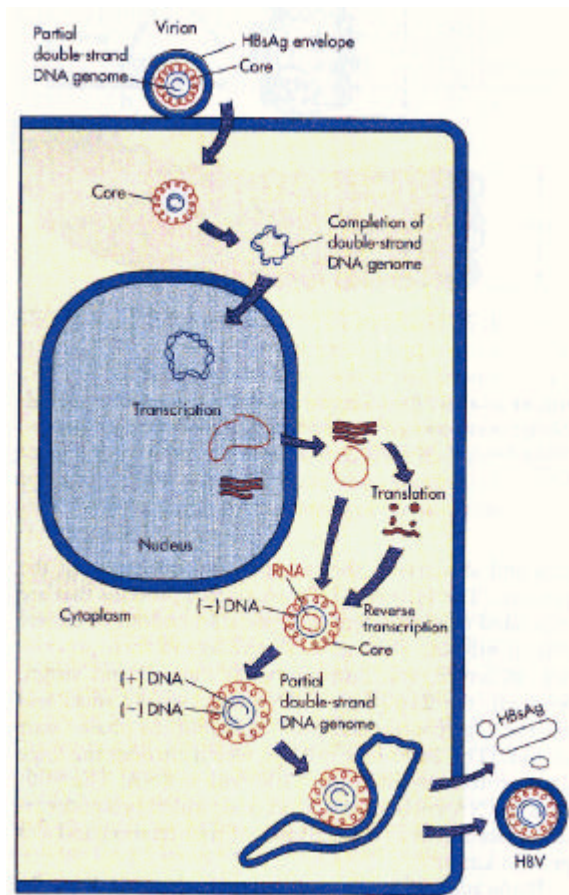


Figura 3.11 – Ciclo de vida do HBV

de cadeia dupla completa. A RNA polimerase celular produz então múltiplas cópias de RNA (RNA pregenoma) a partir da cadeia de polaridade negativa. O RNA é então transportado para o citoplasma onde é produzida a proteína do core, que rapidamente encapsula o RNA pregenoma, bem como a DNA polimerase viral produzida *de novo* e a proteína terminal do DNA. Esta proteína, substitui provavelmente o primer, levando a DNA polimerase a produzir a cadeia L do DNA viral com a sua actividade de transcriptase reversa. À medida que o DNA é sintetizado, o RNA pregenoma é degradado, com excepção de uma pequena porção derivada da zona 5', a qual serve com primer para a síntese da cadeia S. O core completo assim formado obtém o seu envelope por evaginação de parte da membrana celular contendo o  $Hb_sAg$ .

A detecção molecular do genoma viral, para além de constituir uma forma altamente específica de detecção do vírus, permite ainda a detecção de títulos virais mais baixos e patologicamente activos, bem como o acompanhamento da eficácia da acção terapêutica. Se à especificidade da reacção de hibridação, for ainda aliada a sensibilidade da reacção de PCR, obtém-se um método analítico extremamente sensível e específico, e portanto de grande utilidade clínica.

### 3.1.3.2 - O Vírus da Hepatite C (HCV)

O vírus da Hepatite C (HCV), isolado pela primeira vez em 1989, parece ser um importante factor causador de doença hepática crónica, cirrose e carcinoma hepatocelular, em todo o mundo. Hoje em dia foram já isoladas várias estirpes virais do HCV, o que permitiu desenvolver testes serológicos e de genética molecular (detecção de RNA viral) para a determinação da viremia.

O genoma do HCV possui 9379 nucleótidos, sendo constituído por uma cadeia simples de RNA com um único e longo “open reading frame”. O produto genético é uma proteína precursora com 3011 aminoácidos, que por proteólise pós-traducional origina proteínas

estruturais (core e envelope) e não estruturais (proteases, helicases, polimerases do RNA).

A detecção do HCV não é possível de ser realizada por testes padrão de detecção de antígenos no soro, já que as partículas virais circulam no soro em concentrações abaixo das detectáveis por imunoenaios. Assim, a maioria dos estudos epidemiológicos foram inicialmente baseados na prevalência de anticorpos contra o antígeno c100-3. Presentemente, é ainda testada a presença de anticorpos contra outros antígenos virais (testes ELISA de terceira geração). No entanto, estes testes originam um grande número de falsos positivos, pelo que se tornou necessário o desenvolvimento de testes confirmativos.

Como já foi referido, várias estirpes do gene da Hepatite C foram já identificadas (estirpes 1a, 1b, 2, 2a, 2b, 3a, 4, 5), tendo estirpes diferentes prognósticos e repostas terapêuticas diferentes.

### 3.1.3.3 - O vírus Linfotrófico Humano (HTLV-I e HTLV-II)

O vírus HTLV tipo I foi o primeiro retrovírus humano descrito, sendo considerado o agente causador da Leucemia/linfoma da célula T de Adultos (ATLL). Este vírus tem ainda sido associado com uma família

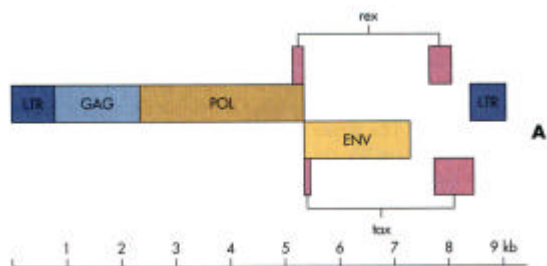


Figura 3.12 – Genoma do HTLV-1

de doenças neurológicas incluindo “spastic paraparesis” e Mielopatia associada a HTLV-I. Mais recentemente, também a poliomiositose e a poliartrite têm sido associadas a infecções por HTLV-I. As manifestações clínicas, e epidemiologia do HTLV-II são no entanto

menos claras, não havendo nenhum síndrome clínico especificamente associado à infecção por este vírus.

Estes vírus têm um muito limitado leque de hospedeiros. Com efeito, sendo o CD4 o receptor para o vírus, apenas as células que o expressam (linfócitos Th e monócitos) são passíveis de infecção.

O HTLV induz a formação de sincícios multinucleados em diversas linhas celulares (o que é um bom indicador de diagnóstico).

O papel do vírus na indução de tumores parece provável, já que as

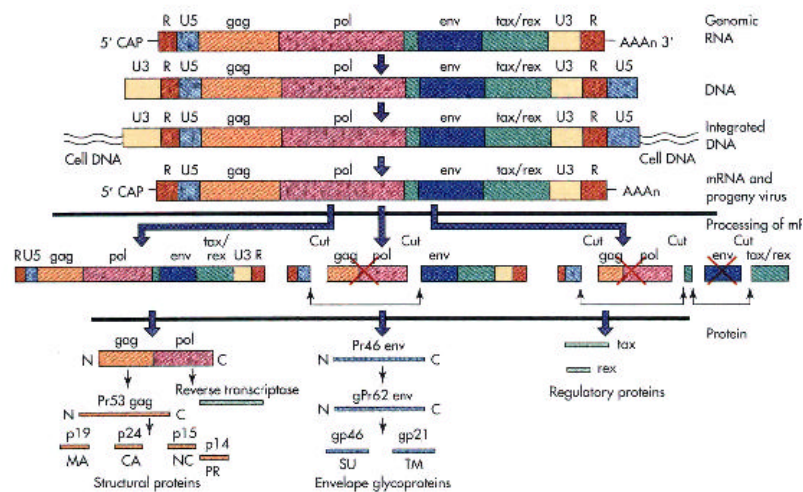


Figura 3.13 – expressão dos genes do HTLV

células leucémicas têm todas o vírus, e este aparece integrado do genoma celular de modo clonal (sempre no mesmo local, num indivíduo, mas em locais diferentes em indivíduos diferentes).

Como o vírus não possui um oncogene, o seu papel deve-se à influencia em um oncogene celular. Como o seu local de integração é diferente de indivíduo para indivíduo, este papel deve ser feito em *trans*, possivelmente através da expressão do gene viral *tax*, o qual é um activador da transcrição. O gene *tax* encontra-se na extremidade 3’

do gene *env* estendendo para a região U3, a qual é altamente conservada em diferentes isolados virais. A proteína *tax* é para o vírus um ativador de sequências *enhancer like* presentes no LTR viral. A nível celular, o principal efeito desta proteína parece ser a activação dos genes da IL-2, o que permite às células multiplicar de forma autónoma.

Devido à razão custo/benefício particularmente boa dos testes de detecção do HTLV baseados em ELISA, este é habitualmente o método de eleição para o screening de produtos sanguíneos. No entanto, e dado o elevado número de falsos positivos apresentados, um resultado positivo deve ser confirmado por um método diferente, como o western Blot. Os métodos de detecção do genoma viral baseados em PCR e hibridação constituem testes poderosos complementares. Devido à grande sensibilidade e especificidade do PCR, aliada à enorme sensibilidade da hibridação de DNA, este método permite com rapidez, sensibilidade e especificidade determinar a presença mesmo de níveis muito baixos de partículas virais, bem como seguir a resposta à terapêutica muito depois de os resultados por ELISA serem negativos.

### 3.1.3.4 - O vírus do síndrome da imunodeficiência adquirida (HIV)

O vírus da Imunodeficiência Humana (HIV) é demonstradamente o causador do Síndrome Humano de Imunodeficiência Adquirida (SIDA). Sendo um retrovírus, é capaz de inserir uma cópia de DNA do seu genoma de RNA no genoma da célula infectada, podendo manter-se desta forma num estado latente.

O genoma do HIV (cerca de 10Kb) contém 9 genes, existindo diferenças importantes entre o HIV-1 (predominante na Europa, América e África Central) e o HIV-2 (predominante na África ocidental). Como os restantes retrovírus, o seu genoma inclui os genes *gag*, *pol* e *env*. Os produtos dos genes *gag* e *pol* são processados por uma protease viral, que é parte da poliproteína *pol*. O produto do gene *env* é processado por uma protease celular, originando a glicoproteína

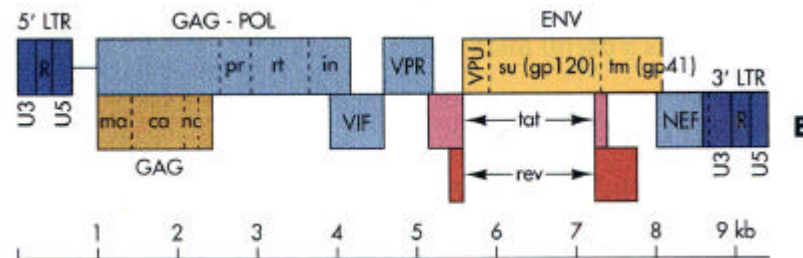


Figura 3.14 – Mapa Genómico do HIV

transmembranar gp-41, e a proteína externa gp-120. Ambas as estirpes possuem ainda os cinco outros pequenos genes: *tat*, *ver*, *nef*, *vif* e *vpr*. Adicionalmente, o HIV-1 possui ainda o gene *vpu* enquanto o HIV-2 possui o gene *vpx*.

Mutações nos genes *tat* e *rev* inibem a multiplicação viral, enquanto mutações no gene *nef* têm o efeito oposto. O gene *tat* codifica um transactivador que, em colaboração com uma proteína celular, aumenta a expressão de todos os genes virais por aumento da transcrição. O seu efeito consiste na prevenção da terminação precoce da transcrição na 59ª base após o local CAP. A sequência nesta zona (região TAR), forma uma estrutura em *hairpin* que facilita a terminação, sendo assim necessário um inibidor da terminação. A proteína *ver* é necessária para a expressão dos genes *gag*, *pol* e *env*, mas não para os genes *ver* e *tat*. Como todos os genes são transcritos num único transcrito que é depois processado, este efeito diferencial deve ser efectuado num passo pós-transcricional. No gene *env*, esta proteína, ligando-se num local na metade 3' do gene impede a inibição da sua tradução causada pela interacção do restante mRNA com factores celulares. A proteína *rev* pode ainda impedir o splicing do transcrito *env*, fazendo com que os genes *tat* e *env* sejam expressos com prejuízo do gene *env*. O modo de acção do *ver* nos genes *gag* e *pol* pode ser similar.

A proteína *nef* reprime a transcrição do genoma HIV actuando na sua zona NR (de *negative regulatory*). Esta proteína tem propriedades similares ao oncogene *ras*: é miristolado, liga-se ao GTP, é fosforilado

por uma proteína cinase, e tem actividade GTPase. Mutações neste gene aumentam a progenia viral 2 a 3 vezes, sugerindo que este gene pode estar envolvido no desenvolvimento do estado latente.

Os genes *vif*, *vpu*, *vpr* e *vpx* provavelmente codificam proteínas do virião, que aumentam a eficiência da infecção.

Como vimos, a transcrição do genoma do HIV é activada ou inibida pela ligação de factores específicos de origem viral ao LTR 5'. No entanto, estes factores são coadjuvados por factores celulares, ligando assim o estado de activação/latência viral ao próprio estado de activação da célula infectada. Assim, a região NR responde à proteína *nef*, mas também à proteína celular Ppt-1, expressa em linfócitos não activados (este gene inibe também a síntese de IL-2). Várias regiões virais (regiões de ligação do SP1, enhancers, TATA box) aumentam a transcrição em resposta a factores celulares como o NFk-B, que são produzidos em linfócitos activados.

O vírus é capaz de infectar células que possuem o CD4 na sua superfície, ainda que só esta molécula não seja suficiente para garantir a entrada do vírus (células de ratinho transfectadas, não são infectáveis, a não ser que o genoma seja injectado directamente na célula). Este receptor liga-se à gp120 (um dos dois produtos de clivagem da gp160 – o produto do gene *env*). Estudo mutacionais indicam também a necessidade do domínio extracelular amino-terminal da gp41 (o outro produto de clivagem da gp160), a qual está ligada não covalentemente à gp120. É possível que a ligação da gp120 exponha domínios da gp41, que então se liga a um outro receptor celular, fazendo com que esta proteína de membrana entre na membrana celular, arrastando a membrana viral consigo, e libertando o core no citoplasma celular. O vírus tem também sido observado em endosomas, indicando que a via endocítica é também uma via de entrada do vírus.

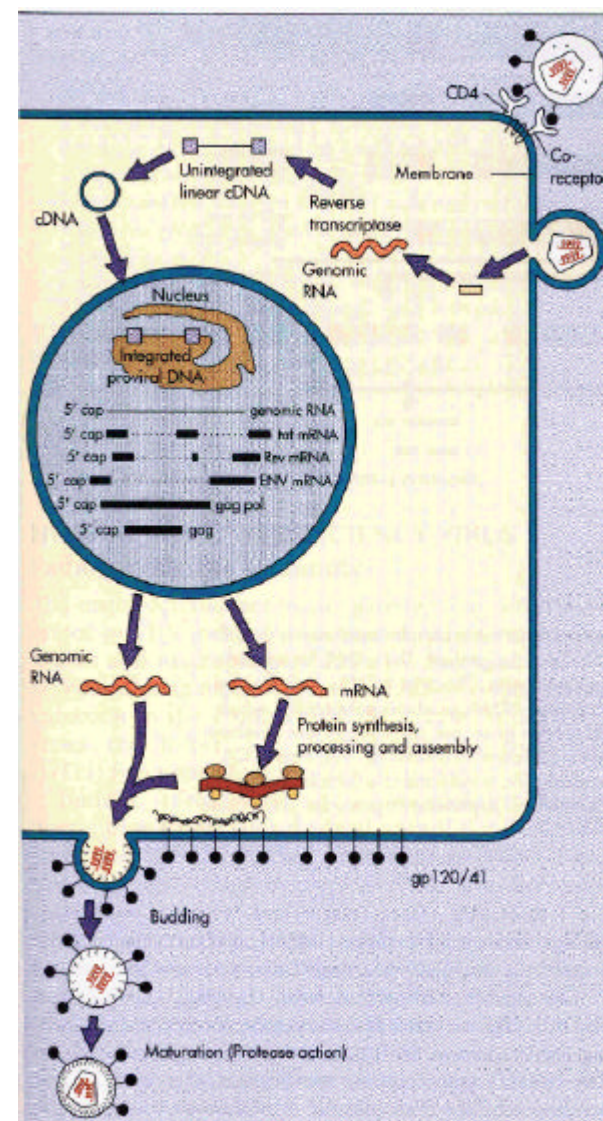


Figura 3.15 – ciclo de vida do HIV

Uma importante consequência funcional da infecção é a diminuição da expressão do CD4, em parte por interferência, isto é ligação intracelular da gp120 com o CD4, impedindo a sua migração para a superfície.

A consequência da infecção na sobrevivência celular varia, já que o vírus se pode integrar no genoma celular, entrando num estado de infecção latente, com pouca ou nenhuma produção de viriões. No entanto, a activação celular, como vimos, pode estimular a entrada do vírus numa fase de infecção produtiva, com um pico na replicação vírica, e libertação de partículas infecciosas seguida de morte celular. Note-se que algumas linhas celulares não morrem, mesmo quando o vírus está na fase produtiva (ex. H9 e linhas monocíticas). A morte celular parece ser um fenómeno activo, mediado pelo gp41 na membrana celular, já que a amputação do seu domínio intracelular impede a morte celular. Um outro factor tendente a provocar a morte celular é a acumulação de DNA não integrado.

Uma das características mais notáveis do HIV é a sua hipervariabilidade. Isolados diferentes do vírus apresentam diferenças na sua composição genética, devidas a mutações (inserções, deleções e mutações pontuais). Tal facto deve-se à sua transcriptase reversa, a qual introduz mutações a um ritmo cerca de um milhão de vezes superior ao observado em vírus de DNA. Dos genes virais, o *env* parece ser o mais variável, com 5 zonas de hipervariabilidade na gp120. Os genes *gag* e *pol* são os menos variáveis, possivelmente devido a restrições nas variações viáveis. Uma importante consequência da hipervariabilidade da gp120 (o antigénio viral mais exposto) é a sua capacidade para iludir reacções imunológicas mediadas por anticorpos, fazendo com que as especificidades destes se tornem “desactualizadas”.

Dos testes disponíveis para a detecção do HIV, o ELISA é habitualmente utilizado. No entanto, e devido ao elevado número de falsos positivos, um resultado positivo deve ser confirmado por uma técnica complementar, habitualmente Southern Blot. Os métodos

baseados na reacção de PCR, ainda que de uso relativamente recente, constituem poderosos meios complementares ao ELISA, já que associam a enorme sensibilidade da reacção de PCR com a especificidade quer do PCR quer da hibridização. Estes métodos têm ainda a vantagem de detectar directamente o genoma viral, independentemente do estado imunológico ou replicativo do vírus, aumentando por isso a sua eficiência de detecção. Estes testes são ainda importantes na avaliação clínica da infecção em crianças nascidas de mães seropositivas, já que é independente da presença de anticorpos ou antigénios provenientes do sangue materno.

### **3.1.4 - Papel dos vírus em oncologia**

O papel dos vírus em oncologia foi inicialmente considerado apenas como uma curiosidade académica, sem contraponto no mundo clínico. No entanto a descoberta em 1932 do papel de vírus de DNA em tumores de coelhos selvagens, e mais tarde (Bittner, 1936) em adenocarcinoma de ratinhos e (Gross, 1951) em leucemias de ratinho levou ao estabelecimento seguro do papel dos vírus na etiologia de alguns tumores. Estes estudos e os que lhe seguiram permitiram estabelecer os mecanismos que permitem que a etiologia viral nos tumores pode permanecer insuspeita: 1) a tumorigenese pode ser um fenómeno raro em vírus largamente disseminados, e habitualmente inócuos; 2) em alguns vírus, as partículas virais são heterogéneas, sendo que a maioria infecta as células sem ocasionar tumores, o que faz com que o tumor só surja numa fase muito avançada da doença, disfarçando a sua característica infecciosa.

Posteriormente, o potencial oncogénico viral foi estudado, avaliando o seu potencial transformador em culturas celulares. Estes estudos levaram à conclusão que um vírus que induziu um tumor, deixa de ser reconhecido em cultura pela sua infectividade, mas que a sua presença pode ser confirmada a nível de DNA ou RNA.

Nos anos 60, verificou-se que a maioria das classes de vírus de DNA eram capazes de induzir tumores em animais, enquanto dos vírus de RNA apenas os retrovírus tinham essa capacidade. (tabela 3.2)

Tabela 3.2 – potencial oncogénico das várias classes de vírus

genoma de	Familia/grupo do vírus	Vírus oncogénicos
DNA	Adenovirus	muitos tipos
	Papovavirus	Polioma, SV40, papiloma
	Herpesvirus	EBV, CMV, herpes
	Hepadnavirus	HBV
	Poxvirus	fibroma vírus
	Parvovírus	nenhum
RNA	retrovírus	Leukosis viruses, Sarcoma viruses
	Picornavirus	Nenhum
	Togavirus	Nenhum
	Ortomixovirus	Nenhum
	Paramixovirus	Nenhum
	Rabdovirus	Nenhum
	Coronavirus	Nenhum
	Arenavirus	Nenhum
	Reovirus	Nenhum

A descoberta da passagem do ciclo viral dos retrovírus por uma fase de DNA, levou à unificação do papel viral na oncogénese, como sendo uma propriedade atribuível ao DNA viral. Mais recentemente, esta regra ganhou força com a descoberta que o papel oncogénico é atribuível a oncogenes transportados ou activados pelos vírus (ver capítulo 4.1.4).

### 3.1.4.1 Oncogenes virais

O estudo do papel dos vírus na etiologia dos tumores, revelou a existência de oncogenes no genoma de alguns vírus. Como vimos, trata-se de um conjunto de genes homólogos de genes celulares, mas que devido à sua desregulação adquirem a propriedade de estimular o crescimento celular de forma contínua. Os exemplos são hoje em dia muito numerosos. A título de exemplo veja-se a tabela 3.3.

Os oncogenes presentes nos retrovírus são habitualmente genes celulares alterados, não realizando funções específicas para os vírus, pelo que não conferem qualquer vantagem ao vírus para além da de lhe permitirem uma enorme multiplicação enquanto integrados no genoma da célula transformada. Os oncogenes presentes nos vírus de DNA são habitualmente de natureza diferente, realizando funções específicas do vírus. Trata-se de vírus que se replicam em células não activadas, pelo que não possuem habitualmente a maquinaria necessária para a replicação viral. Os

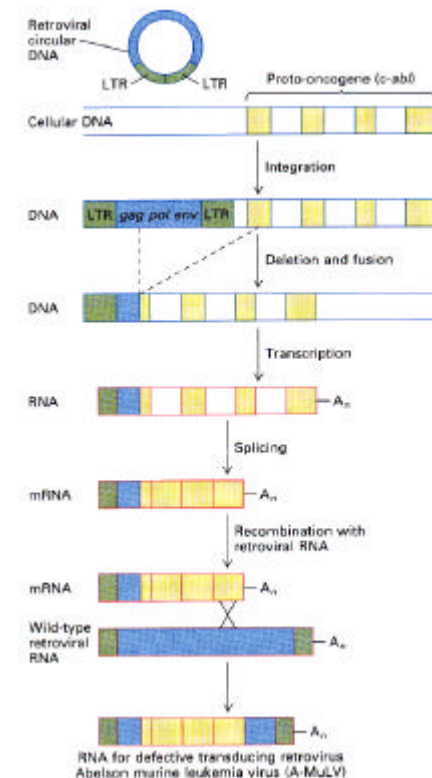


Figura 3.16 – incorporação de um proto-oncogene celular no genoma viral, com a sua transformação em oncogene

oncogenes virais resolvem este problema, estimulando a produção destas enzimas, por activação da maquinaria de replicação do DNA celular. O efeito secundário traduz-se no crescimento celular, isto é, num fenótipo transformado.

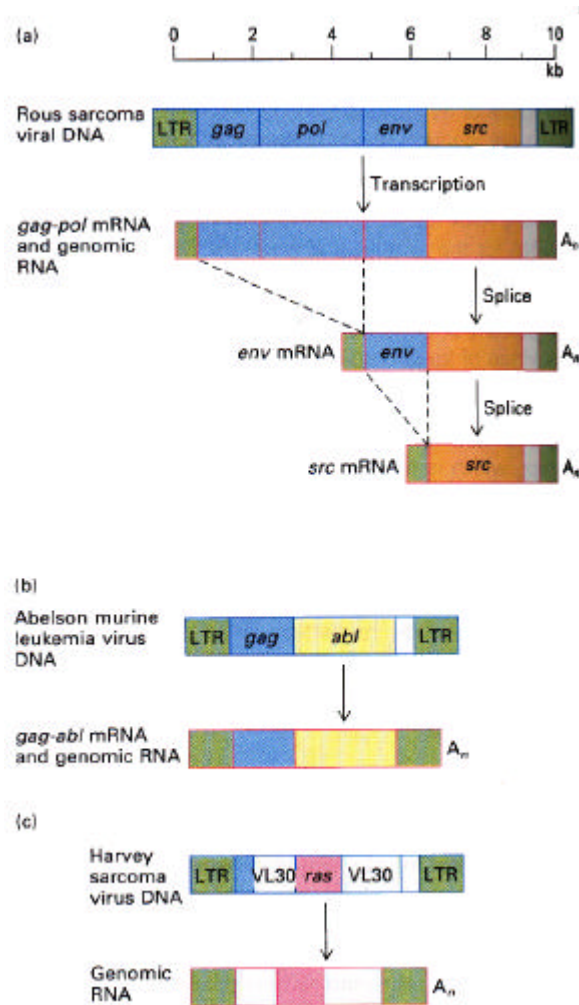


Figura 3.17 – exemplos de vírus com oncogenes no genoma viral, e sua expressão.

3.35

Tabela 3.3 – Oncogenes virais

Oncogene	Localização	Função
<b>Oncogenes presentes em vírus de DNA</b>		
E1A	Núcleo/citoplasma	Regula a transcrição
E1B	Núcleo/citoplasma	Regula a transcrição
PV-ST	citoplasma	
PV-MT	membrana citoplasmática	liga e estimula pp60 <sup>c-src</sup> e pp62 <sup>c-yes</sup>
PV-LT	núcleo	inicia a síntese de DNA e regula a transcrição
SV40-ST	citoplasma	
SV40-LT	núcleo, memb.citoplasm.	inicia a sínt.DNA, reg.transcrição, liga p53
<b>Oncogenes presentes em Retrovírus</b>		
<i>abl</i>	memb.citoplasm.	tirosina cinase
<i>erb A</i>	citoplasma	receptor da hormona tiroideia
<i>erb B</i>	membranas	tirosina cinase / EGF receptor
<i>ets</i>	núcleo	
<i>fes</i>	memb.citoplasm.	tirosina cinase
<i>fgr</i>	memb.citoplasm.	tirosina cinase
<i>fms</i>	memb.citoplasm.	receptor do CSF-1 (tirosina cinase)
<i>fos</i>	núcleo	
<i>fps</i>		
<i>kit</i>	membranas	tirosina cinase
<i>mil/raf</i>	citoplasma	serina/tirosina cinase
<i>mos</i>	citoplasma	serina cinase
<i>myb</i>	núcleo	
<i>myc</i>	núcleo	
<i>ras</i>	memb.citoplasm.	proteína G
<i>raf</i>		
<i>rel</i>	citoplasma	
<i>ros</i>	citoplasma	tirosina cinase
<i>sis</i>	citoplasma e excretado	subunidade do PDGF
<i>src</i>	memb.citoplasm.	tirosina cinase
<i>ski</i>	núcleo	
<i>yes</i>		tirosina cinase
<b>Oncogenes não presentes em vírus</b>		
<i>bcl</i> - linfoma folicular humano		<i>p53</i> – activo em células transformadas
<i>bcr</i> - LMC		<i>ret</i> – Linfoma
<i>int-1,2,3,4</i> – cancro da mama (rato)		<i>rho</i> – semelhante a <i>ras</i>
<i>met</i> – linha celular transformada		
<i>neu</i> – neuroglioblastoma de rato		

3.36



### 3.1.4.2 Conversão de proto-oncogenes celulares em oncogenes

#### 3.1.4.2.1 Activação de proto-oncogenes

O principal mecanismo que origina a conversão de um proto-oncogene num oncogene consiste na alteração da sequência genética por acumulação de eventos mutacionais. Estas mutações podem ocorrer quer na região codificante, quer na região controladora, ainda que na maior parte dos casos, as alterações se verifiquem na região

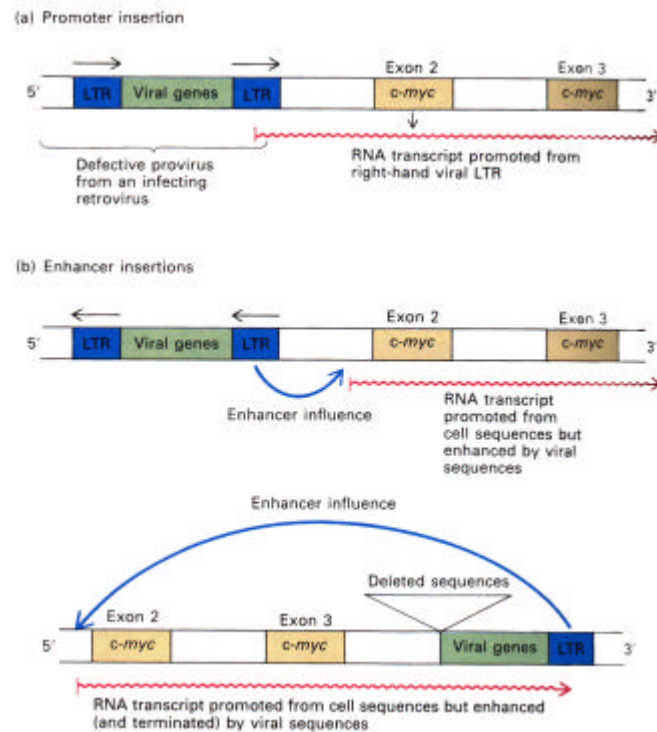


Figura 3.18 – Mecanismos de activação viral de proto-oncogenes celulares.

codificante. Assim, as mutações podem traduzir-se em alterações de um ou mais amino-ácidos com funções essenciais, ou mesmo na eliminação de extensas zonas da molécula (causada por mutações nonsense).

#### 3.1.4.2.2 Amplificação de proto-oncogenes

Os vírus necessitam habitualmente de realizar a transcrição de pelo menos parte dos seus genes a um ritmo extremamente elevado, assegurando assim uma rápida formação de novos viriões, antes de as suas proteínas serem detectadas pelo sistema imune na superfície celular. Para conseguir este efeito, os vírus possuem enhancers muito eficientes, os quais quando integrados no genoma celular podem descontrolar o programa genético da célula, levando ao aumento de expressão de alguns genes celulares, que de outra forma seriam expressos em níveis baixos. Surge assim, um oncogene por um efeito quantitativo por oposição ao efeito qualitativo descrito na secção anterior. Facilmente se percebe que se o gene cuja expressão foi alterada estimular a divisão celular, o efeito da sua desregulação quantitativa pode ser a transformação celular.

#### 3.1.4.3 A translocação t(8;14)(q24;q32) e o EBV

O vírus do Epstein-Barr (EBV) foi pela primeira vez observado por Epstein e Barr em células de doentes com mononucleose infecciosa, tendo desde então sido demonstrado ser o agente causador desta patologia. Este vírus está ainda associado ao Linfoma de Burkitt, muito embora a nível genómico, as estirpes associadas à mononucleose infecciosa e ao Linfoma de Burkitt apresentem cerca de 35% de diferenças.

O vírus pertence à família dos Herpesvirus, possuindo no entanto características peculiares: 1) o seu DNA é composto por 5 regiões de DNA de sequência única, e por 7 regiões de sequências repetitivas,

incluindo os terminais genômicos. O genoma do EBV possui ainda um elevado grau de polimorfismo entre os vários isolados.

O vírus infecta selectivamente as células B, através de um receptor relacionado com o terceiro componente do complemento (c3b). A infecção primária resulta numa fase de latência, em que apenas regiões limitadas, não contíguas do genoma são transcritas (genes de expressão latente). Os genes de expressão latente são 6 genes de expressão nuclear (EBNA- *Epstein Bar nuclear antigens*) e 3 genes codificando proteínas membranares (LMP e TP1 e 2). Todas as proteínas de expressão latente, com a excepção de EBNA1 são apresentadas pelo HLA, e reconhecidas pelas células citotóxicas, contribuindo para a manutenção da infecção num estado suportável pelo hospedeiro. Isto mesmo é exemplificado no facto de a maioria dos tumores associados a EBV que surgem nos doentes imunossuprimidos, regredirem quando a imunossupressão é reduzida ou retirada. Na fase inicial da infecção, o DNA linear do vírus circulariza servindo-se dos terminais repetitivos, seguindo-se a replicação. A maior parte do genoma viral existe intracelularmente nesta forma circular extracromossómica, tendo no entanto sido detectados ocasionalmente cópias de DNA integrado no genoma celular. A fase replicativa do ciclo é caracterizada pela activa transcrição viral, extensa replicação, e produção de proteínas dos “*late genes*”. Cada célula produz no entanto poucas partículas virais infecciosas.

Devido à existência de uma fase latente, e à manutenção da infecção num estado suportável nos indivíduos não imunocomprometidos, a infecção por EBV é muito frequente (cerca de 95% dos adultos estão infectados).

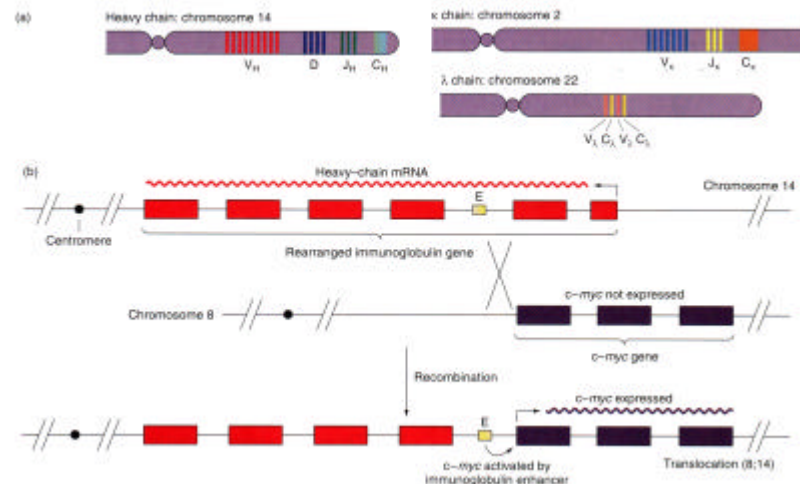


Figura 3.19 - Diagrama mostrando os eventos genéticos geradores de uma das três translocações encontradas no Linfoma de Burkitt. O oncogene c-MYC está normalmente localizado no braço longo(q) do cromossoma 8. A translocação t(8;14) coloca este oncogene junto ao locus da IgH, e por isso sob a influência do enhancer da IgH, o qual é muito activo em linfócitos B.

O EBV está classicamente associado ao Linfoma de Burkitt (BL), o qual é endémico em África e à doença linfoproliferativa do linfócito B (BLPD), característica dos indivíduos imunocomprometidos. Recentemente foram ainda identificadas associações entre o EBV e subtipos de Linfoma de Hodgkin e de Linfomas T.

O Linfoma de Burkitt (BL) é fundamentalmente diferente da Doença Linfoproliferativa do Linfócito B, já que a primeira surge num indivíduo imunocompetente. Neste caso, o vírus parece escapar à vigilância do sistema imune, expressando apenas o gene EBNA1, o qual como vimos não é apresentado pelo HLA. No caso do BL existem ainda anomalias genéticas associadas à infecção pelo EBV, conduzindo à desregulação do c-myc e ou p53, o que contribui para o fenótipo maligno da célula infectada. Ainda que a implicação definitiva

do EBV na etiologia dos linfomas que ocorrem em indivíduos infectados com o vírus da SIDA permaneça motivo de investigação, o facto de a replicação do EBV nestas células ser muito alta e de uma única estirpe, parece favorecer a implicação do EBV.

Assim, parece ser importante controlar a replicação do EBV nos indivíduos afectados por estes tumores, sendo ainda necessário dispor de tecnologias laboratoriais para detectar a existência de baixos níveis de vírus replicativamente activos nos indivíduos em tratamento.

# Capítulo 4

## 4 Áreas de intervenção da genética molecular em hemato-oncologia

## 4.1 O cancro: Uma doença genética de células somáticas

### 4.1.1 A origem genética dos tumores

Uma abordagem sistematizada e aprofundada deste tema está para além do âmbito deste manual. Procuraremos fornecer a informação suficiente para uma visão global da oncogénese hematológica, com uma descrição dos principais mecanismos nela envolvidos. Numa segunda parte faremos uma exposição das alterações genéticas mais frequentes nas principais doenças hematológicas. Finalmente abordaremos de forma breve a utilidade actual dos estudos moleculares em hemato-oncologia.

Em resposta a necessidades internas ou a factores externos a que estão submetidas, as células podem seguir várias vias: a) não reacção; b) diferenciação para adaptação às novas necessidades; c) divisão celular; d) apoptose. Estes processos, envolvem múltiplas vias enzimáticas que interagem em complexas redes ainda não completamente esclarecidas. Importa, conhecer algumas dessas vias para mais facilmente entender como a sua alteração pode desregular a capacidade proliferativa, de diferenciação ou de sobrevivência de uma célula, levando-a a adquirir características neoplásicas.

Podem considerar-se seis tipos de proteínas que participam no controlo da divisão celular e sobrevivência da célula: 1) factores de crescimento; 2) receptores de factores de crescimento; 3) transdutores intracelulares do sinal; 4) factores de transcrição nuclear; 5) proteínas de controlo do ciclo celular; 6) proteínas de controlo da apoptose. A todos estes níveis a alteração estrutural/funcional dos genes que codificam estas proteínas pode criar desequilíbrios que favoreçam a proliferação da célula, rompendo os mecanismos de controlo habituais. Na realidade, actualmente pensa-se que a maioria das neoplasias se inicia por um único evento mutacional no DNA de uma única célula. Desta forma, a oncogénese é actualmente entendida como uma doença genética

adquirida em células não somáticas. A mutação que ocorre no DNA da célula inicial, confere-lhe vantagem selectiva sobre as restantes. Esta vantagem de sobrevivência/proliferação cria condições para mutações subsequentes no património genético da sua descendência (clone), com o aparecimento de células com alterações genéticas adicionais (subclones) que lhes vão conferindo vantagens selectivas progressivas, em paralelo com uma maior agressividade clínica – noção "*multistep*" da carcinogénese.

É costume dividir os genes envolvidos no aparecimento de tumores em dois grupos:

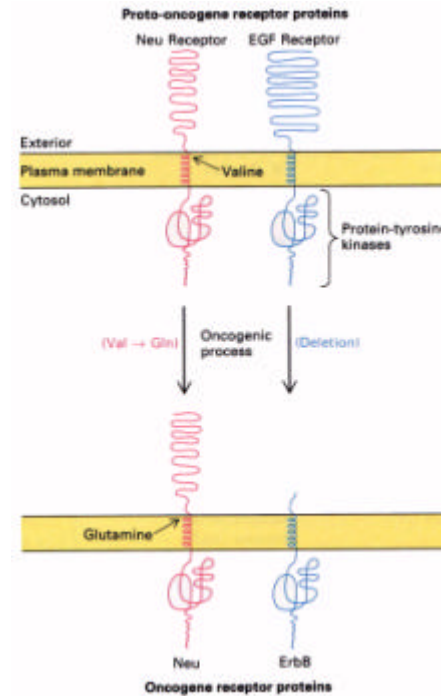
- 1) proto-oncogenes – São genes que se encontram nas mais diferentes espécies de vertebrados e invertebrados e que codificam uma proteína envolvida no controlo da divisão celular. A sua activação em oncogénese causa um efeito positivo na proliferação celular, mesmo na presença da versão não alterada do restante alelo (acção dominante);
- 2) genes de supressão tumoral – são genes que codificam proteínas envolvidas no controlo da proliferação celular e da integridade do DNA. A sua alteração cria condições para a proliferação de células com alterações do DNA. Neste caso, ambos os genes teriam que estar funcionalmente inactivos (acção recessiva). Note-se no entanto, que em alguns casos (ex. p53), o facto de a proteína funcionar como multímeros, cria condições para a existência de uma acção dominante (ver capítulo 4.3.2).

São habitualmente alterações congénitas dos genes de supressão tumoral que são responsáveis pela tendência hereditária para o desenvolvimento de tumores (ex. famílias Li-Fraumeni, Retinoblastoma, Ataxia-telangiectasia), o que se compreende, já que cria de forma transmissível uma menor resistência ao acumular de eventos mutacionais.

#### 4.1.1.1 Proto-oncogenes e oncogenes

Assumida a característica genética dos tumores, resta procurar os genes responsáveis pelo fenótipo tumoral. Uma das primeiras pistas proveio da observação que alguns vírus eram capazes de induzir o fenótipo tumoral nas células que infectavam. O estudo do conteúdo genético destes vírus permitiu identificar um número hoje bastante extenso de genes indutores de características oncológicas, genericamente designados por oncogenes. Mais surpreendente que a identificação dos oncogenes foi a descoberta de genes homólogos no genoma das células normais. Estes genes (denominados por analogia de proto-oncogenes) não são iguais, mas homólogos aos oncogenes. Na verdade, os oncogenes são cópias mutadas dos proto-oncogenes, estando as mutações (pontuais, nonsense e translocações) associadas à geração das características oncológicas, quer por alteração directa da respectiva proteína, quer por alteração do programa de expressão da proteína. Os proto-oncogenes e respectivos oncogenes são designados por um nome de três letras (ex. *src*). Se o nome é escrito só com minúsculas e em itálico, este refere-se ao oncogene, enquanto o proto-oncogene é representado sem itálico e com a primeira letra em maiúsculas. O nome do oncogene quando existente no genoma de um vírus é representado com o prefixo *v* (ex: *v-src*). Por oposição o proto-oncogene é também por vezes representado em minúsculas, com o prefixo *c* (ex: *c-src*).

#### 4.1.1.1.1 Factores de crescimento



Os factores de crescimento constituem sinais (sob a forma de hormonas ou não) que sinalizam a uma célula que deve iniciar o processo de divisão. Estas moléculas, ao ligarem ao respectivo receptor, induzem a produção de uma plethora de sinais intracelulares e respectivas respostas como a mobilização de reservas energéticas, diferenciação e entrada no ciclo de divisão celular. Como células diferentes possuem receptores diferentes, cada factor de crescimento produz respostas em alguns tipos celulares e noutros não.

O único caso conhecido em que um factor de crescimento originou um oncogene é o do PDGF (platelet derived growth factor). No entanto, foi experimentalmente confirmado o potencial destas proteínas para produzirem características oncológicas, se estimuladas em cascatas autócrinas. Assim, quando experimentalmente uma célula com um receptor para o factor de crescimento X foi induzida a a produzir este factor de crescimento, gerou-se um ciclo fechado de produção/consumo de estímulo (ciclo autócrino), que levou a célula à divisão permanente (ex. de ciclos

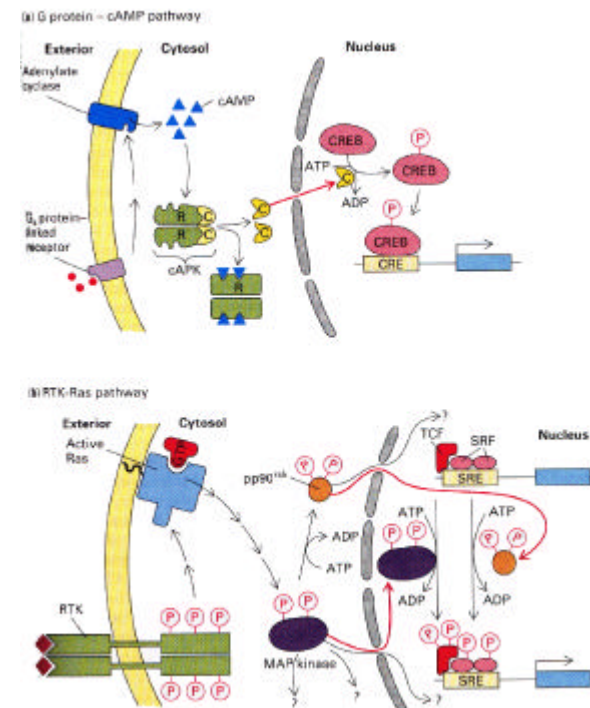
autócrinos produzidos em laboratório são o mediado pelo GM-CSF e TGF- $\alpha$ ).

#### 4.1.1.1.2 Receptores para factores de crescimento

Alguns receptores para factores de crescimento possuem na sua cauda citoplasmática actividade de tirosina cinase, transmitindo o estímulo produzido pela ligação do factor de crescimento, através da fosforilação de uma ou mais proteínas citoplasmáticas. Estes receptores tornam-se oncogenes, se uma mutação fizer com que a actividade de tirosina cinase não esteja condicionada à ligação do factor de crescimento. Por exemplo uma mudança num aminoácido localizado na zona transmembranar do gene Neu transforma-o no oncogene *neu*. Na maioria dos casos, uma grande parte do domínio extracelular do receptor é delectado originando o oncogene (ex. ErbB).

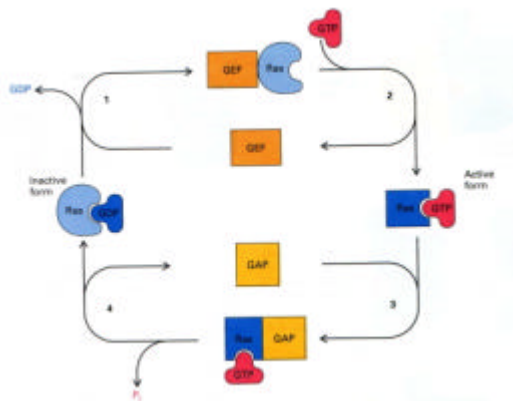
#### 4.1.1.1.3 Transdutores intracelulares de sinal

Este é o maior grupo conhecido de proto-oncogenes. É constituído por proteínas que transmitem sinais desde receptores até alvos nucleares. Os mais bem conhecidos transdutores de sinal são as proteínas G. Um bom exemplo é a proteína Gs, a qual controla a produção de cAMP. Esta proteína detecta a ligação de um ligando ao seu receptor, liga-se a GTP, activa a adenil ciclase, levando à formação de cAMP. Hidrolisa então o GTP, regressando a um estado inactivo. Uma mutação no gene da Gs, eliminando a actividade de hidrólise do GTP, transforma esta proteína num oncogene, já que o mecanismo de inactivação desta proteína fica inoperacional, do que resulta um aumento constitucional de cAMP celular, o qual leva a uma proliferação desregulada de células pituitárias, levando a tumores deste órgão.



Muitos dos proto-oncogenes desta classe codificam para tirosinas cinases não receptores (ex. Src e Abl). Trata-se portanto de proteínas citoplasmáticas ou nucleares, sem qualquer domínio transmembranar ou extracitoplasmático. Muitas destas proteínas possuem miristatos, um ácido gordo longo ligado à sua glicina N-terminal. Este facto faz com que estejam parcialmente ligados à membrana citoplasmática, colocando a sua actividade de tirosina cinase numa posição idêntica à dos receptores com esta actividade.

A alteração genética que origina o oncogene *src* foi estudada em grande detalhe. A proteína normal (pp60<sup>c-src</sup> ou c-*src*) possui múltiplos locais de fosforilação, através dos quais é controlada. Fosforilação da tyr-527 junto da sua extremidade C-terminal origina uma grande redução na sua actividade cinase. Esta zona proteica está



frequentemente alterada nas proteínas *src* que possuem actividade cinase constitutiva. Por exemplo no vírus do sarcoma Rous, a v-*src* sofreu uma deleção que eliminou os últimos 18 aminoácidos da c-*src*.

Um outro grupo de oncogenes

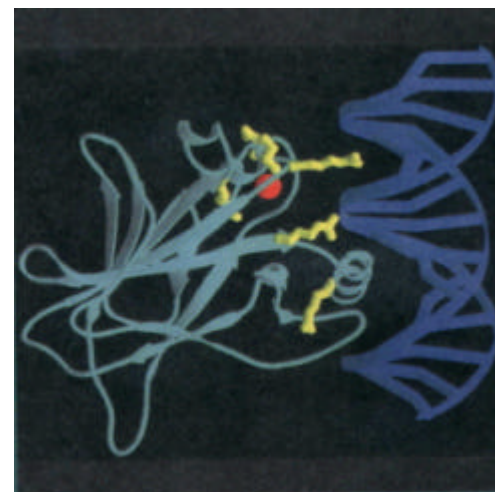
desta classe são os genes Ras, cujos produtos se ligam ao GTP, hidrolisando-o lentamente. Tal como as proteínas Src, as Ras possuem ácidos gordos ligados (um grupo farnesil), localizando-se no interior da membrana citoplasmática. Este proto-oncogene é transformado por uma simples substituição da valina-12 por glicina, o que apesar de apenas causar uma ligeiríssima alteração conformacional, impede a hidrólise do GTP, causando um fenótipo tumoral.

O produto dos genes Crk não apresenta actividade bioquímica conhecida, mas possui em comum com muitos outros oncogenes domínios designados por SH2 e SH3 (*src* homologous domains), os quais interactuam com outras proteínas. O domínio SH2 interactua

com péptidos curtos, em locais proteicos com uma tirosina fosforilada, enquanto o domínio SH3 interactua com domínios proteicos ricos em prolina. Assim, a actividade oncogénica do *crk* implica que um aspecto central na génese de tumores é não apenas a fosforilação de proteínas chave, mas também o padrão de associações proteicas mediadas pelos domínios SH, pelo padrão de fosforilações e eventualmente pela exposição de grupos ricos em prolina por modulação da conformação proteica.

#### 4.1.1.1.4 Factores de transcrição nuclear

Seja porque mecanismo for, todos oncogenes originam mudanças no programa genético das células em que existem. Estas mudanças traduzem-se na alteração do tipo e/ou quantidade de espécies de mRNA produzidas, e das respectivas proteínas. Os factores de transcrição nuclear actuam directamente nos genes, aumentando ou diminuindo a transcrição dos genes com que interactuam. Fazem-no de uma de duas formas: 1) interactuando com os promotores, e assim





ligando/desligando a transcrição; 2) interagindo com enhancers, e assim aumentando ou diminuindo a quantidade de mRNA produzido.

Um bom exemplo de proto-oncogene deste tipo são os genes Jun e fos. Os produtos destes genes, quando dimerizados constituem parte do factor de transcrição AP-1, o qual se liga a sequências nos promotores e enhancers de muitos genes. Presumivelmente, estes genes actuam como oncogenes, activando a transcrição de genes que promovem o crescimento celular, ou inibindo a transcrição de genes que reprimem a divisão celular.

Muitas proteínas nucleares codificadas por proto-oncogenes são induzidas quando as células são estimuladas a crescer. Exemplos são o aumento de c-Fos e c-Myc em células 3T3 estimuladas com PDGF. Curiosamente, o aumento fisiológico destas proteínas é rápido, transitório e moderado, enquanto a sua expressão oncogénica é prolongada e extremamente elevada. Este efeito é obtido à custa da perda de sequências génicas que instabilizam quer o mRNA quer a proteína, contribuindo assim para o seu rápido desaparecimento.

#### **4.1.1.1.5 Proteínas de controlo do ciclo celular**

Como vimos já, o ciclo celular é precisamente controlado por várias proteínas, nomeadamente por ciclinas, cdk, p53 e RB, assegurando que o crescimento, duplicação de DNA e divisão nuclear estão perfeitamente sincronizados. Se a regulação da expressão de uma ou mais ciclinas é alterada, ou se ocorrem mutações nos genes que codificam as ciclinas, o p53 ou o RB, podem surgir desregulações do ciclo celular, com consequências oncogénicas.

#### **4.1.1.2 Genes de Supressão tumoral**

A herança de certos genes aumenta para quase 100% a probabilidade de desenvolvimento de tumores. Um caso clássico é o retinoblastoma, que é como muitos outros tumores hereditários um tumor da infância. Este tumor levou à identificação do primeiro gene de supressão

tumoral, o RB (gene do retinoblastoma). Crianças que herdaram uma única cópia defeituosa do RB (frequentemente traduzida numa deleção no cromossoma 13) desenvolvem em média 3 tumores de retinoblastoma, cada qual resultado de uma única célula transformada. Apesar de parecer um elevado número, verifica-se que apenas 1 em cada  $10^6$  células da retina destas crianças desenvolve tumores, o que significa que mesmo neste caso de hereditariedade dominante, ao nível celular a expressão fenotípica é altamente recessiva, sendo necessário um segundo evento para despoletar o fenótipo. Este segundo evento é a mutação da cópia normal do gene RB.

Como vimos, o gene RB é um gene de controlo da divisão celular. Outro gene da sua classe que actua também como gene de supressão tumoral é o p53. Trata-se de genes cuja função é verificar a progressão do ciclo celular, mantendo as células num estado quiescente, ou levá-las a entrar em apoptose, se as condições não forem as adequadas para a progressão do ciclo celular. O p53 actua como um tetrâmero, ou mesmo um oligomero de ordem mais elevada, o que significa que a diminuição da concentração de proteína funcional ocasionada pela mutação de um dos dois alelos na célula pode eliminar quase por completo a actividade do p53, pois virtualmente todos os oligómeros possuirão pelo menos uma subunidade mutada. Assim, o p53 é um gene de supressão tumoral com uma hereditariedade invulgar, já que ao contrário das mutações do RB (e da maioria dos genes de supressão tumoral), as mutações do p53 funcionam de forma dominante. Cerca de metade dos tumores humanos possuem mutações do p53. Tais mutações impedem o controlo do ciclo celular e permitem a acumulação de defeitos no DNA, funcionando como um intensificador da probabilidade de ocorrência de mutações oncogénicas.

Como já vimos uma das formas de acção do p53 é induzir a apoptose. No entanto, para que tal funcione, é necessário que outros genes, que habitualmente a suprimem estejam capazes de ser desactivados. Um destes genes é o bcl-2, o qual é tem como função proteger as células

válidas da apoptose, sendo a sua expressão muito diminuída quando as células entram em apoptose. Assim, se este gene se encontrar desregulado, por forma a que a sua expressão seja elevada e constitutiva, induz um fenótipo tumoral nas células, impedindo que os genes de supressão tumoral induzam a morte celular programada.

#### **4.1.2 Os carcinogénios como mutagénios**

Pensa-se que muitos tumores humanos têm na sua origem a acção de agentes químicos. Os agentes químicos com acção carcinogénica têm uma enorme variedade de estruturas, sem uma actividade química unificadora óbvia. Podem no entanto ser agrupados em duas categorias principais: os de acção directa e os de acção indirecta. Os carcinogénios de acção directa (de que se conhecem poucos exemplos) são espécies químicas electrófilas, reagindo com compostos de carga negativa. Os carcinogénios indirectos, requerem transformação metabólica, a qual consiste na introdução de centros electrofílicos. Esta transformação é efectuada por complexos enzimáticos como o P-450, constituintes habituais dos organismos (no fígado dos mamíferos), os quais têm como principal função a destoxificação de compostos químicos tóxicos. Com efeito, algumas drogas terapêuticas, insecticidas, hidrocarbonetos policíclicos e outros compostos naturais são tão insolúveis em água, mas solúveis em gorduras, que se acumulariam nos organismos, se não sofressem transformações químicas que os tornassem solúveis em água. Estas transformações consistem na adição de grupos hidrófilicos, o que torna possível a excreção por solubilização na água. Na maioria dos casos, os altamente reactivos grupos epóxido adicionados pela P-450 são rapidamente hidrolisados em grupos hidroxilo, seguindo o processo de solubilização com a adição de ácido glucorónico ou outros grupos. No entanto, em alguns casos, por presumivelmente os grupos epóxido não se encontrarem facilmente acessíveis à epóxido hidratase, estes grupos permanecem como tal, formando-se os carcinogénios.

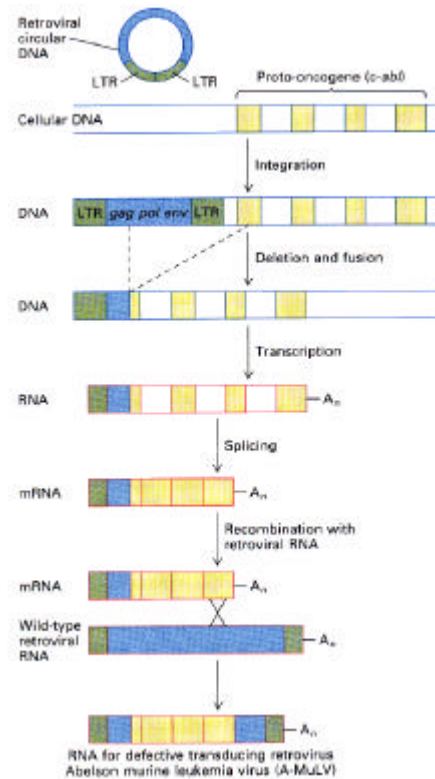
Uma vez no interior das células, os electrófilos podem reagir com compostos possuindo centros negativamente carregados. Ainda que estes centros possam existir em proteínas, RNA e DNA, é a acção sobre o DNA que torna estes compostos carcinogénios. Com efeito, estes compostos ao reagirem com as bases do DNA, em posições que dependem do seu tamanho e estrutura, causam alterações na sequência do DNA. Assim, a acção carcinogénica destes compostos é sobreponível à sua acção mutagénica.

### 4.1.3 Agentes virais na origem dos tumores

#### 4.1.3.1 Oncogenes virais

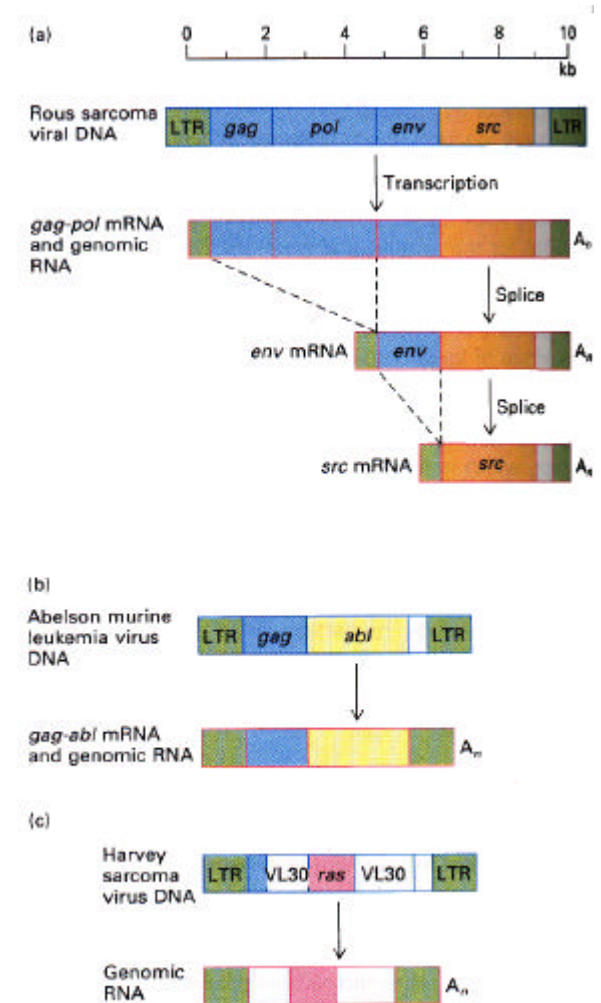
O estudo do papel dos vírus na etiologia dos tumores, revelou a existência de oncogenes no genoma de alguns vírus. Como vimos, trata-se de um conjunto de genes homólogos de genes celulares, mas que devido à sua desregulação adquirem a propriedade de estimular o crescimento celular de forma contínua. Os exemplos são hoje em dia muito numerosos. A título de exemplo veja-se a tabela yyy2.

Os oncogenes presentes nos retrovírus são habitualmente genes celulares alterados, não realizando funções específicas para os vírus, pelo que não conferem qualquer vantagem ao vírus para além da de lhe permitirem uma enorme multiplicação enquanto integrados no genoma da célula transformada. Os oncogenes presentes nos vírus



4.15

de DNA são habitualmente de natureza diferente, realizando funções específicas do vírus. Trata-se de vírus que se replicam em células não activadas, pelo que não possuem habitualmente a maquinaria necessária para a replicação viral. Os oncogenes virais resolvem este problema, estimulando a produção destas enzimas, por activação da maquinaria de replicação do DNA celular. O efeito secundário traduz-se no crescimento celular, isto é, num fenótipo transformado.



4.16

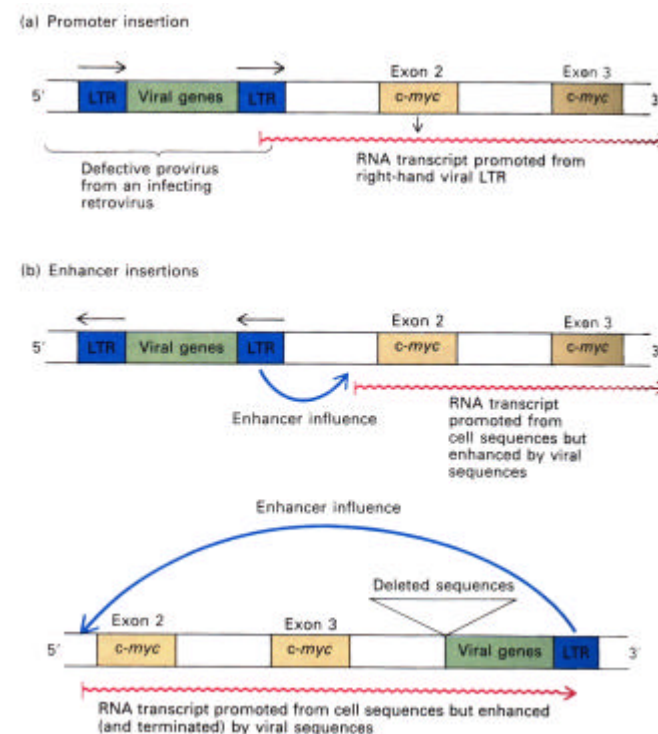
Tabela yyy2 – Oncogenes virais

Oncogene	Localização	Função
<b>Oncogenes presentes em vírus de DNA</b>		
E1A	Núcleo/citoplasma	Regula a transcrição
E1B	Núcleo/citoplasma	Regula a transcrição
PV-ST	citoplasma	
PV-MT	membrana citoplasmática	liga e estimula pp60 <sup>c-src</sup> e pp62 <sup>c-yes</sup>
PV-LT	núcleo	inicia a síntese de DNA e regula a transcrição
SV40-ST	citoplasma	
SV40-LT	núcleo, memb.citoplasm.	inicia a sínt.DNA, reg.transcrição, liga p53
<b>Oncogenes presentes em Retrovírus</b>		
<i>abl</i>	memb.citoplasm.	tirosina cinase
<i>erb A</i>	citoplasma	receptor da hormona tiroidea
<i>erb B</i>	membranas	tirosina cinase / EGF receptor
<i>ets</i>	núcleo	
<i>fes</i>	memb.citoplasm.	tirosina cinase
<i>fgr</i>	memb.citoplasm.	tirosina cinase
<i>fms</i>	memb.citoplasm.	receptor do CSF-1 (tirosina cinase)
<i>fos</i>	núcleo	
<i>fps</i>		
<i>kit</i>	membranas	tirosina cinase
<i>mil/raf</i>	citoplasma	serina/tirosina cinase
<i>mos</i>	citoplasma	serina cinase
<i>myb</i>	núcleo	
<i>myc</i>	núcleo	
<i>ras</i>	memb.citoplasm.	proteína G
<i>raf</i>		
<i>rel</i>	citoplasma	
<i>ros</i>	citoplasma	tirosina cinase
<i>sis</i>	citoplasma e excretado	subunidade do PDGF
<i>src</i>	memb.citoplasm.	tirosina cinase
<i>ski</i>	núcleo	
<i>yes</i>		tirosina cinase
<b>Oncogenes não presentes em vírus</b>		
<i>bcl</i> - linfoma folicular humano		<i>p53</i> – activo em células transformadas
<i>bcr</i> - LMC		<i>ret</i> – Linfoma
<i>int-1,2,3,4</i> – cancro da mama (rato)		<i>rho</i> – semelhante a <i>ras</i>
<i>met</i> – linha celular transformada		
<i>neu</i> – neuroglioblastoma de rato		

#### 4.1.3.2 Conversão de proto-oncogenes celulares em oncogenes

##### 4.1.3.2.1 Activação de proto-oncogenes

O principal mecanismo que origina a conversão de um proto-oncogene num oncogene consiste na alteração da sequência genética por acumulação de eventos mutacionais. Estas mutações podem ocorrer quer na região codificante, quer na região controladora, ainda que na



maior parte dos casos, as alterações se verifiquem na região codificante. Assim, as mutações podem traduzir-se em alterações de um ou mais amino-ácidos com funções essenciais, ou mesmo na eliminação de extensas zonas da molécula (causada por mutações nonsense).

#### **4.1.3.2.2 Amplificação de proto-oncogenes**

Os vírus necessitam habitualmente de realizar a transcrição de pelo menos parte dos seus genes a um ritmo extremamente elevado, assegurando assim uma rápida formação de novos viriões, antes de as suas proteínas serem detectadas pelo sistema imune na superfície celular. Para conseguir este efeito, os vírus possuem enhancers muito eficientes, os quais quando integrados no genoma celular podem descontrolar o programa genético da célula, levando ao aumento de expressão de alguns genes celulares, que de outra forma seriam expressos em níveis baixos. Surge assim, um oncogene por um efeito quantitativo por oposição ao efeito qualitativo descrito na secção anterior. Facilmente se percebe que se o gene cuja expressão foi alterada estimular a divisão celular, o efeito da sua desregulação quantitativa pode ser a transformação celular.

#### **4.1.4 A "estatística" na origem dos tumores: mutações espontâneas**

As mutações "espontâneas" surgem por uma grande variedade de mecanismos, tal como já referido na secção 1.1.2.4. Qualquer destes mecanismos contribui para que fenómenos mutacionais potencialmente oncogénicos ocorram com alguma frequência. Estes eventos não são no entanto na maior parte dos casos determinantes, graças aos mecanismos de reparação referidos em 1.1.2.5. No entanto, por vezes a reparação dos danos mutacionais não é eficaz, dando origem a mutações efectivas e potencialmente oncogénicas.

## **4.2 Anomalias Genéticas em doenças hemato-oncológicas**

### **4.2.1 Translocações cromossómicas em hemato-oncologia**

Dentro das alterações cromossómicas presentes em hemopatias malignas, as mais bem estudadas são as translocações. Em alguns casos, as translocações são específicas de um determinado tipo de leucemia, o que atesta a sua importância na patogénese da doença. Noutros casos, ainda que as translocações não pareçam ser específicas de uma patologia definida como entidade clínica autónoma, o seu estudo demonstrou o envolvimento de oncogenes, sendo em alguns casos o motor da descoberta destes.

Segundo o mecanismo de activação dos oncogenes envolvidos, podem distinguir-se dois tipos de translocações: as que envolvem um dos genes das Imunoglobulinas ou do TCR, e as que originam genes híbridos funcionais, com características oncogénicas, e em que participa pelo menos um oncogene.

#### **4.2.1.1 Translocações envolvendo genes das Ig ou TCR**

Nestas translocações, um protooncogene é colocado pela translocação debaixo da acção do potente enhancer dos genes das Ig ou do TCR. Nestas condições, a sequência do proto-oncogene não é alterada, mas a sua transcrição encontra-se muito aumentada, por acção do enhancer. Exemplos de translocações deste tipo os referidos na tabela xxx3

Como descrito anteriormente (secção 2.1.4), ainda que não se conheça com precisão o mecanismo envolvido nestas translocações, a caracterização das zonas de junção parece envolver a maquinaria fisiológica de recombinação somática.

Tabela xxx3. – Translocações envolvendo os genes do TCR e das Ig

	Translocação	Doença	Gene afectado
IgH	t(8;14)(q24;q32)	Burkitt	C-Myc
	t(11;14)(q13;q32)	LNH-manto	bcl-1 (CCND1/PRAD1)
	t(14;18)(q32;q21)	LNH-foliculares	bcl-2 (apoptose)
	t(14;19)(q32;q13)	LLC-B	bcl-3 (ciclina?)
	t(3;14)(p27;q32)		
	t(5;14)(q31;q32)	LLA-pre B	IL-3 (citocina)
IgK	t(2;8)(p12;q24)	Burkitt	
	t(2;3)(p12;q27)	LNH-célula grande	bcl-6 (zinc-finger)
Igλ	t(8;22)(q24;q11)	Burkitt	
	t(3;22)(q27;q11)		
TCR-α/δ	t(1;14)(p32;q11)		Tal-1
	t(10;14)(q24;q11)	LLA-T	TCL-3 (Hox11)
	t(11;14)(p14;q11)	LLA-T	Rhombotin 1
	t(11;14)(p13;q11)	LLA-T	Rhombotin 2 (TCL-2)
	t(8;14)(q24;q11)	LLA-T	C-Myc
TCR-β	t(1;7)(p32;q35)	LLA-T	
	t(7;9)(p34;q32)	LLA-T	Tal-2
	t(7;11)(p35;p13)		dominio LIM
	t(7;9)(q34;q34.3)	LLA-T & LNH	TAN-1
	t(;7)(q34;q34)	LLA-T	lck
	t(7;19)(q34;p13)	LLA-T	LYL1

#### 4.2.1.1.1 translocação t(8;14)(q24;q32) e o EBV

No Linfoma de Burkitt, ocorrem 3 tipos de translocações recíprocas, todas envolvendo o gene MYC (8q24), e genes dos cromossomas 2 (IgK), 14 (IgH) e 22 (Igλ). A translocação mais frequente é a t(8;14)(q24;q32), a qual ocorre em mais de 75% destes linfomas. O gene MYC é conhecido pela sua importância na proliferação celular, mas não é expresso nas células B maduras. No entanto, estas translocações colocam este gene dependente de enhancers das imunoglobulinas, pelo que o gene passa a estar activo nas células B que possuem estas translocações, dando origem a níveis de mRNA semelhantes aos encontrados nas células normais em proliferação.

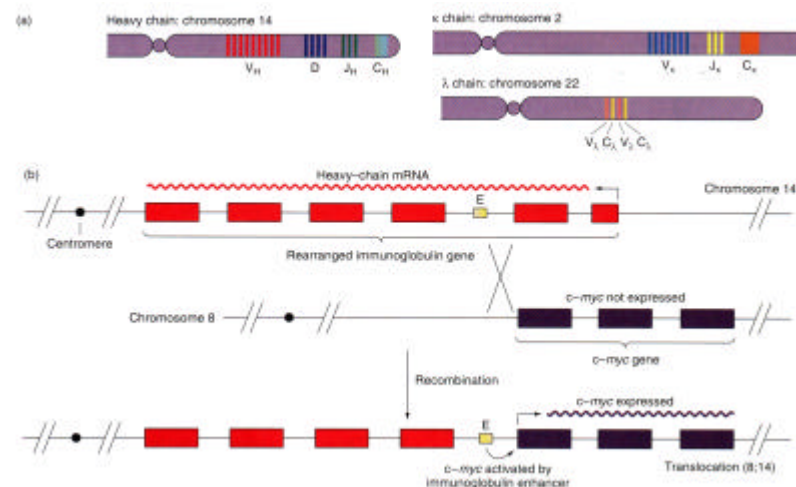


Figura 1 - Diagrama mostrando os eventos genéticos geradores de uma das três translocações encontradas no Linfoma de Burkitt. O oncogene c-MYC está normalmente localizado no braço longo(q) do cromossoma 8. A translocação t(8;14) coloca este oncogene junto ao locus da IgH, e por isso sob a influência do enhancer da IgH, o qual é muito activo em linfócitos B.

Estudos moleculares revelaram a existência de 2 mecanismos de geração da translocação t(8;14)(q24;q32). O primeiro, ocorre no, Linfoma endêmico da África equatorial, associado à infecção por EBV, o gene MYC não é rearranjado, encontrando-se intacto, se bem que próximo das regiões DH ou JH do gene IgH. Esta mutação ocorre no estágio celular pré-B, quando a maquinaria de recombinação dos genes das Imunoglobulinas está activa. O segundo mecanismo de geração desta translocação não está associado à infecção pelo EBV. Neste caso, a translocação ocorre imediatamente 3' do gene MYC, ou dentro deste, envolvendo ainda a região de "switch" do gene IgH. As células neste caso apresentam um fenótipo mais maduro, compatível com a ocorrência da mutação numa altura em que a célula efectuava o "switch" de imunoglobulinas.

#### 4.2.1.1.2 A t(14;18) - BCL2/IgH

Em cerca de 85% dos Linfomas foliculares (FL), e 25% dos Linfomas Difusos (DL), surge a translocação t(14;18)(q32;q21), envolvendo os genes BCL-2 (B-Cell Lymphoma/Leukemia-2 gene) no cromossoma 18 e um dos segmentos JH do gene da IgH no cromossoma 14. O gene BCL-2 codifica uma proteína que parece ter potencial oncogénico sendo importante na fase pré-B do desenvolvimento do linfócito B, ao prevenir a morte celular por apoptose. Duas regiões de quebra foram identificadas no cromossoma 18: 2/3 das translocações envolvem uma região de 150 bp na zona 3' não traduzida do gene (o Major Breakpoint region ou mbr). As restantes translocações envolvem o Minor cluster region (mcr) localizado cerca de 20 Kb após o início do gene. A translocação parece não afectar a sequência do BCL-2, mas tão somente os níveis de mRNA deste gene, e ocorre presumivelmente por erro na maquinaria genética de recombinação das Ig. Esta interpretação parece ser suportada pela descoberta de regiões N na junção dos "breakpoints", bem como pela existência de mutações somáticas na mesma zona.

O advento do PCR transformou o estudo molecular das mutações envolvendo o gene do BCL-2, e particularmente a translocação t(14;18), tanto a nível do mbr como do mcr. Desta forma foi possível detectar 1 célula mutante num universo de 100.000 células normais, permitindo uma nova sensibilidade na detecção de doença residual mínima.

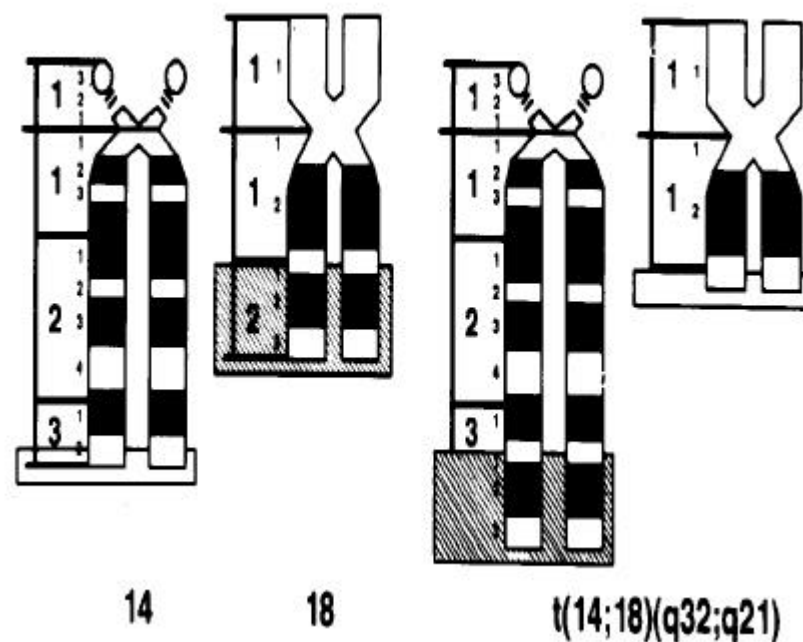


Figura 2 - Diagrama mostrando os cromossomas 14 e 18 normais, e os cromossomas resultantes da translocação t(14;18)(q32;q21), envolvendo os genes BCL-2 (18q21) e IgH (14q32).

#### 4.2.1.2 Translocações que originam genes híbridos

Estas translocações afectam um proto-oncogene, o qual é justaposto a um outro gene, originando um gene quimera funcional, que é transcrito, originando uma proteína com características oncogénicas.

Exemplos deste mecanismo são apresentados na tabela xxxy.

Tabela xxxy- Translocações que originam genes híbridos

Translocação	Doença	Genes afectados	
t(9;22)(q34;q11)	LMC;LLA adulto/LMA	abl	bcr <sup>a</sup>
t(15;17)(q22;q21)	LMA-M3	PML <sup>b</sup>	RARα <sup>d</sup>
t(11;17)(q35;q21)	LMA-M3	PLZF <sup>c</sup>	RARα <sup>d</sup>
t(5;17)(q35;q21)		NPM <sup>e</sup>	RARα <sup>d</sup>
t(8;21)(q22;q22)	LMA-M2	ETO <sup>c</sup>	AML1
Inv(16)(p13;q22)	LMA-M4Eo	MYH <sup>f</sup>	CBFB <sup>g</sup>
t(6;9)(p23;q34)	LMA,LMA-TdT+	DEK	CAN
inv(3)(q21;q26)	LMA-M1	EVII <sup>c</sup>	
t(3;3)(q21;q26)			
t(1;9)(q23;q13)	LLA-pre B	PBX1 <sup>h</sup>	E2A
t(12;21)(p13;q22)	LLA-prec.B infantis	TEL	AML1
t(2;5)(p23;q35)	LNH-anaplásico	ALK <sup>a</sup>	NPM <sup>e</sup>
t(4;11)(q21;q23)	LLA-pre B	ALL <sup>c</sup>	AF4 <sup>i</sup>
t(1;11)(q32;q23)	LAM	AF1P <sup>e</sup>	
t(6;11)(q27;q23)	LAM	AF6 <sup>i</sup>	
t(9;11)(p21;q23)	LAM-M5,prec.B	AF9 <sup>i</sup>	
t(11;17)(q23;q21)	LAM	AF19	
t(11;19)(q23;p13)	LAM,LLA-prec. B		ENL <sup>i</sup>
t(8;13)(p11;q12)	síndrome mieloproliferativo	ZNF198	FGFR-1

a) tirosina cinase

b) guanosina trifosfatase

c) zinc finger

d) receptor acido retinoico

e) fosfoproteína

f) gene da miosina

g) factor de transcrição

h) gene homeótico

i) transdução de sinal

##### 4.2.1.2.1 t(9;22) - bcr/abl

A primeira anomalia cromossómica consistente em tumores humanos foi identificada por Nowell e Hungerford na Leucemia Mielóide

Crónica (CML) em 1960. Porque este achado foi realizado em Filadélfia, utilizou-se o nome desta cidade para designar esta translocação (t(9;22)(q34;q11)) ou ainda cromossoma Ph. Estudos moleculares revelaram que esta translocação envolvia no cromossoma 9 o gene ABL(Abelson proto-oncogene) e no cromossoma 22 o gene BCR(Breakpoint Cluster Region gene) originando um gene quimera codificando 1 proteína com capacidade oncogénica. O cromossoma Ph mais frequente, surge em cerca de 90% dos casos de CML, e variantes citogenéticas surgem em mais 5%. Dos restantes 5%, cerca de metade possui rearranjos do gene ABL não detectados por cariotipagem, mas visíveis por métodos moleculares, sendo os restantes 2.5% considerados Ph<sup>-</sup>. O cromossoma Ph é ainda frequente em Leucemias Linfoblásticas Agudas (ALL; 5% das crianças e 20% dos adultos), e mais raramente em Leucemias Mieloblásticas Agudas (AML; cerca de 1%).



O cromossoma Ph é, como vimos, originado pela junção de parte dos genes BCR e ABL. Esta junção, ocorre sempre no mesmo ponto no gene ABL, mas pode ocorrer em 3 locais diferentes do gene BCR. Estas diferentes junções, dão origem a 2 tipos de proteínas: a p190, resultante da junção do exon e1 do gene BCR com o a2 do gene ABL (transcrito e1a2) e a p210 resultante da junção do exon a2 do gene ABL com os exons b2 ou b3 do gene BCR (transcritos b2a2 e b3a2). A vasta maioria dos casos de CML (95%) expressam a proteína p210. Já na ALL, cerca de 70% dos casos de ALL expressam a p190, e os restantes 30% a p210.

**4.2.1.2.2 t(15;17)**  
-  
**pml/rar**

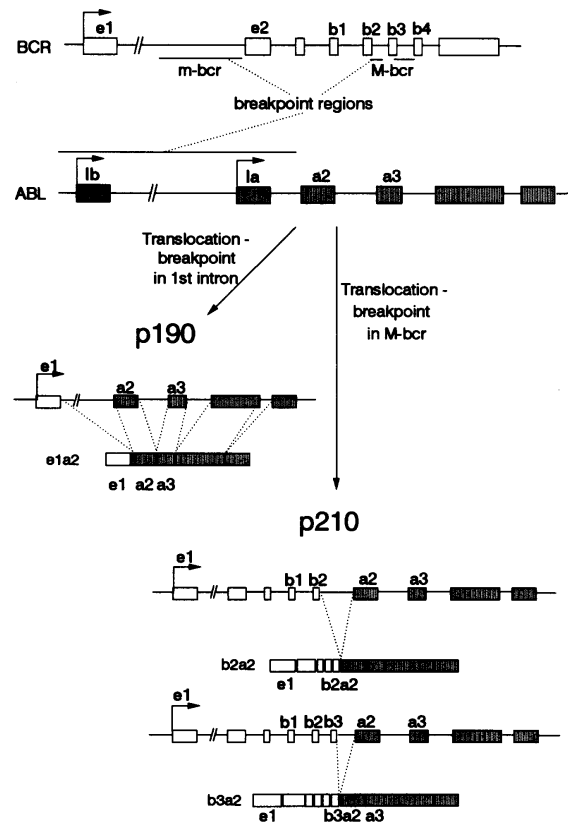


Figura 3 - Os genes BCR e ABL normais, e as translocações que originam as proteínas p190 e p210 do gene quimera BCR-ABL. A proteína p210 é característica da CML, sendo a p190 a proteína BCR-ABL encontrada na maioria dos casos de ALL (ver texto).

Cerca de 10% das FAB-M3 são negativas para a t(15;17), e não respondem clinicamente ao ATRA. Os restantes 90% foram constituem uma entidade clinica denominada Leucemia Aguda Promielocítica (APL) sendo citogeneticamente caracterizados pela presença da translocação t(15;17)(q22;q21). Os genes envolvidos na translocação são o gene PML no cromossoma 15, e o gene RAR $\alpha$  (Receptor do Acido Retinóico) no cromossoma 17. Os pontos de rotura no *locus* RAR $\alpha$  não estão distribuídos ao acaso, mas localizados numa zona de 16 Kb do intron 2. De igual modo, os pontos de rotura no *locus* PML não são aleatórios, concentrando-se em apenas 3 regiões do gene: intron 3 (bcr3: 47% dos casos), exon 6 (bcr2: 4% dos casos), e intron 6 (bcr1: 49% dos casos). Como consequência da translocação, formam-se genes quimera (PML/RAR $\alpha$  e RAR $\alpha$  /PML).

O gene quimera PML/RAR $\alpha$  é transcricionalmente funcional, pelo que origina uma espécie de mRNA passível de detecção por RT-PCR. Foi assim possível determinar a presença deste transcrito em 100% dos casos de APL ( em contraste com apenas 70% dos casos expressando o gene RAR $\alpha$  /PML), esta é uma tecnologia de grande valor na detecção de doença residual mínima nesta patologia.

Desta forma, foi sugerido que um teste positivo deve ser indicativo da continuação do tratamento, ao passo que doentes com 2 testes negativos, e mais de 2 meses de remissão completa, podem ser poupados a sessões terapêuticas.

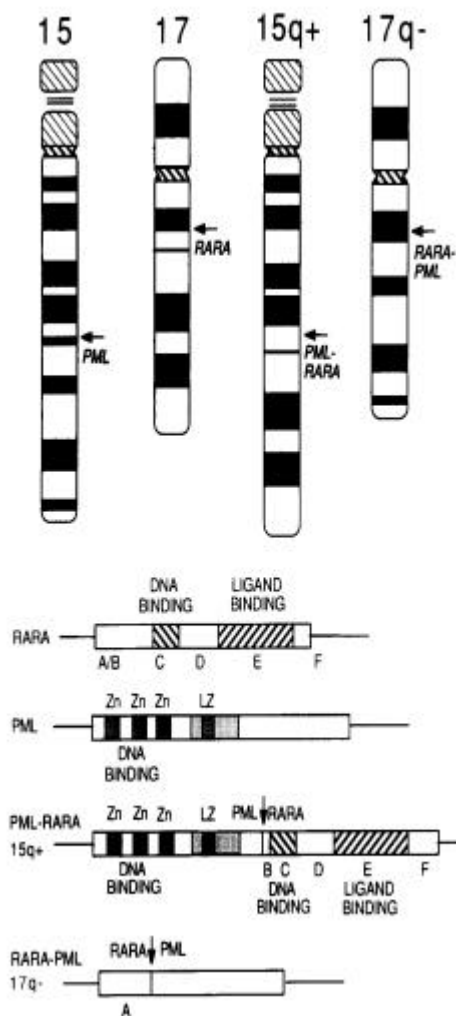


Figura 4 - A localização cromossômica e estrutura normal dos genes PML e RARA, e a translocação t(15;17)(q22;q21) que origina o gene quimera PML-RARA.

## 4.2.2 Deleções cromossômicas em hematologia

### 4.2.3 Mutações pontuais em hematologia

#### 4.2.3.1 mutações no gene p53

Apesar de anomalias citogenéticas envolvendo 17p13 serem raras, o gene p53 aí localizado encontra-se mutado (como determinado por SSCP) em 10-15% dos casos de LLC. Os doentes com deleções ou translocações que envolvem esta zona, possuem quase invariavelmente mutações deste gene. Existe ainda uma forte correlação entre a existência de mutações no gene p53, e um estadió avançado, resistente à quimioterapia, e curta sobrevivência.

## 4.3 Patologias hematológicas

### 4.3.1 LMC

### 4.3.2 LMA

### 4.3.3 LLA

### 4.3.4 LLC

## 4.4 Utilidade clínica da genética molecular em hematologia

Uma vez que o gene quimera BCR-ABL é apenas expresso nas células malignas, a sua detecção molecular constitui um poderoso método de avaliar a progressão da doença. Com efeito, vários autores servindo-se da grande sensibilidade e especificidade da metodologia de PCR (reação em cadeia de polimerase), desenvolveram estratégias para avaliar a doença residual mínima, inferindo mesmo dados válidos na avaliação de prognóstico. Foi assim possível observar que se é frequente a detecção permanente ou intermitente de células residuais BCR-ABL<sup>+</sup>, vários meses após transplante de medula e remissão citogenética completa, já a sua detecção 1 ano após o transplante é

indicadora de pior prognóstico que o dos casos em que se observe remissão por PCR. Não obstante estes dados, a validade da avaliação de prognóstico com base nos dados obtidos por PCR constitui, presentemente motivo de aceso debate e estudo.

PML/RAR

Se os estudos efectuados no final do tratamento parecem ter pouco valor prognóstico, já os estudos efectuados mais tarde parecem ter grande valor prognóstico, com resultados positivos em RT-PCR a indicarem uma recaída. Com efeito, estudos de doentes em remissão por períodos prolongados de tempo (4-12 anos) revelaram que a sobrevida está associada com a erradicação das células PML/RAR $\alpha$ , pelo que este deve ser o objectivo terapêutico.

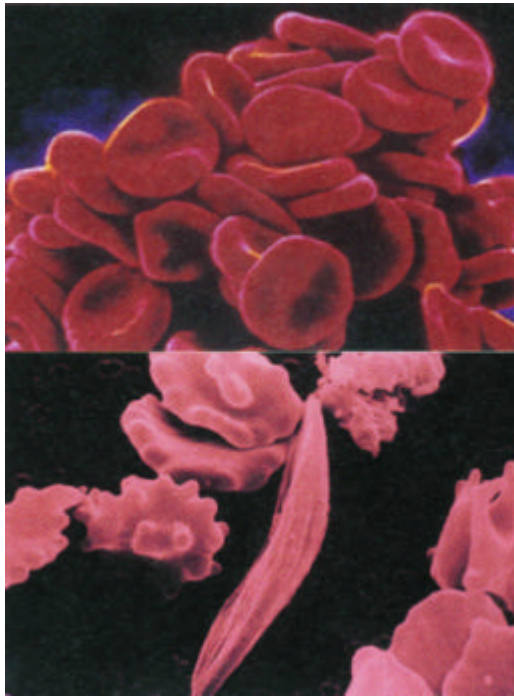
**4.4.1 *Aprofundamento de conhecimentos***

**4.4.2 *Escolha de tratamentos específicos***

**4.4.3 *Monitorização da doença***

# Capítulo 5

## 5 Áreas de intervenção da genética molecular em hematologia



5.1

5.2

## 5.1 - DOENÇAS GENÉTICAS DO GLÓBULO RUBRO

### 5.1.1 - Anemias Não esferocíticas Congénitas

#### 5.1.1.1 - Deficiência em Glucose-6-fosfato desidrogenase

A deficiência em i(G6PD) é uma anomalia genética muito frequente, sendo estimado que afecte cerca de 400 milhões de indivíduos em todo o mundo. A maioria dos portadores desta deficiência é assintomática, correndo no entanto o risco de desenvolver anemias hemolíticas agudas, quando expostos a certas infecções, drogas, ou ingestão de favas. Uma pequena porção dos portadores da deficiência sofre de uma doença mais pronunciada: anemia não-esferocítica congénita.

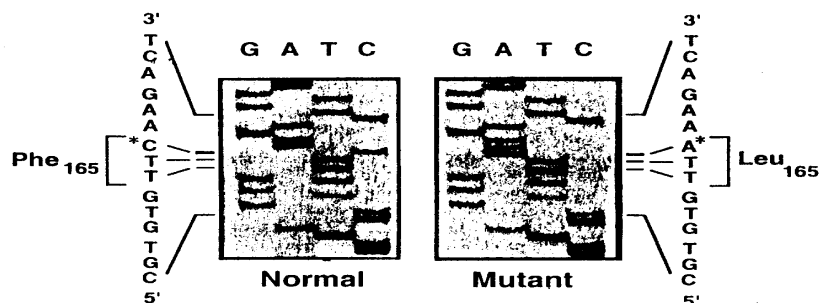


Figura 1 - Gel de sequenciamento do exon 5 do gene ALAS2, indicando uma transversão C para A na posição 547, prevendo a substituição de uma Phe por uma Leu no aminoácido 165 da enzima.

A G6PD é um homodímero. O gene que codifica o respectivo polipéptido de 514 aminoácidos é extremamente conservado na escala evolutiva, estando localizado no homem no cromossoma X

(Xq28). Este gene é expresso em todas as células do organismo, sendo essencial à viabilidade celular. Até ao momento foram identificadas 75 mutações de G6PD, correspondendo a mais de 100 variantes enzimáticas. As mutações constituem quase exclusivamente mutações Missense, causando portanto a substituição de um único aminoácido. Dois grupos de situações são de relevância clínica: a primeira é composta pelas mutações que afectando a actividade enzimática, deixam no entanto actividade suficiente para o metabolismo normal do eritrócito. Neste caso, os portadores são assintomáticos, enquanto não existirem factores externos propensores ao desenvolvimento de anemia hemolítica aguda. No segundo caso, a deficiência em G6PD é tão severa, que os indivíduos desenvolvem anemia não esferocítica congénita.

A caracterização das deficiências genéticas da G6PD permitiu verificar que mutações diferentes podem originar uma deficiência com semelhantes características bioquímicas, e vice-versa.

#### 5.1.1.2 - Deficiência em piruvato quinase

A Deficiência em Piruvato Quinase (PKD) é a causa mais comum de Anemia Hemolítica Não Esferocítica Hereditária. Geralmente assintomática nos portadores (heterozigóticos), a deficiência de Piruvato Quinase manifesta-se nos indivíduos homozigóticos (ou duplamente heterozigóticos) como anemia hemolítica crónica, com severidade variável.

No Homem existem dois genes de piruvato quinase: PKLR (codificando as isoenzimas L e R) e PKM2 (codificando as isoenzimas M1 e M2). o primeiro é o único habitualmente expresso em eritrócitos, sendo o responsável pela PKD nestas células.

A caracterização bioquímica da PK permitiu identificar cerca de 300 variantes enzimáticas, no entanto a caracterização molecular das mutações que afectam o gene da PK permitiu verificar uma heterogeneidade mais limitada, já que a mesma mutação parece estar

associada a variedades bioquímicas diferentes (ex.: PK Nagasaki, PK Tóquio e PK Beirut resultam da mutação <sup>1151</sup>ACCG-ATG, e as variantes PK Fukushima e PK Maebashi resultam da mutação <sup>349</sup>CAG - AAG).

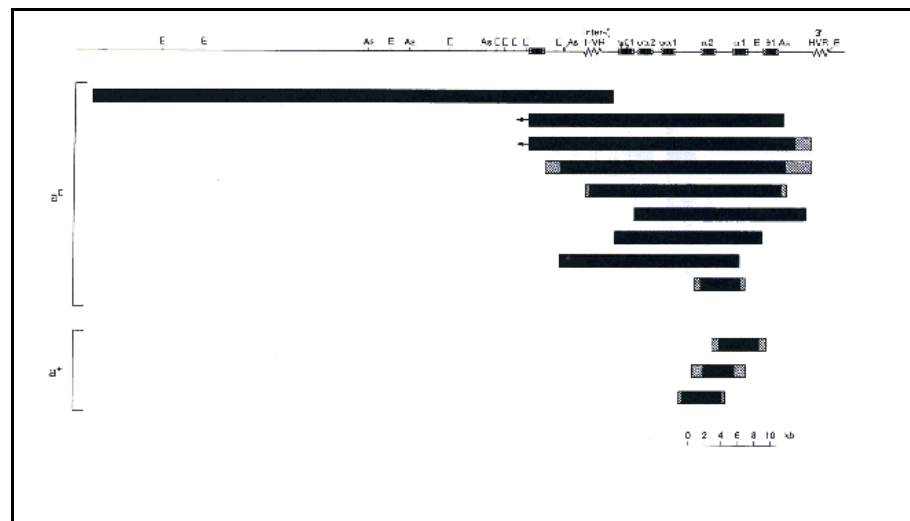
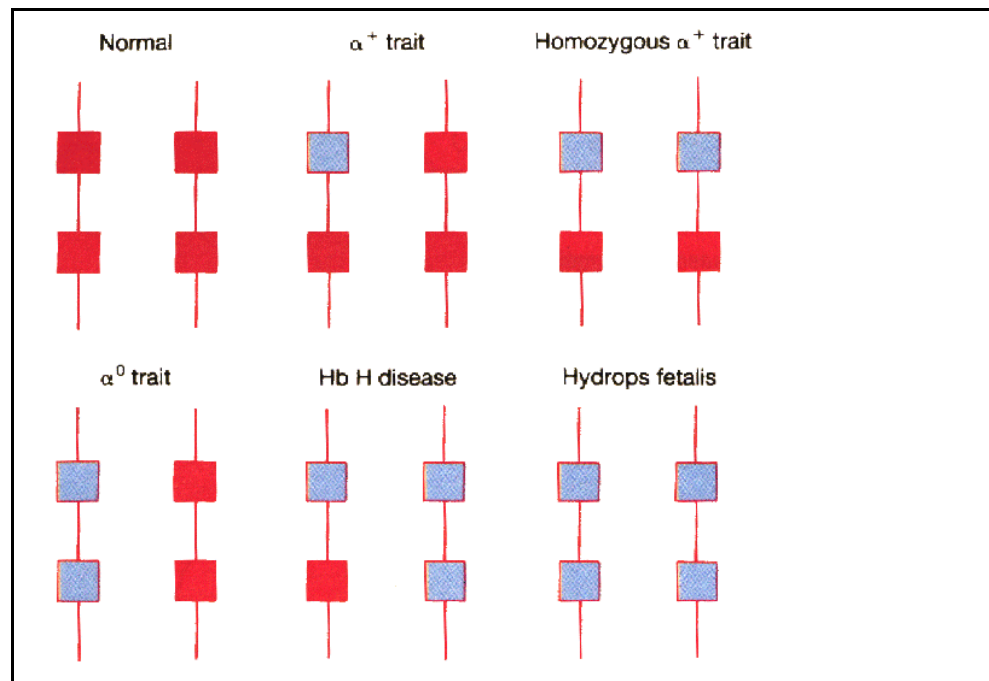
As mutações encontradas gene PKLR, responsáveis pela PKD, incluem mutações tipo “missense”, “nonsense” e inserções. Estas mutações foram encontradas quer na região codificante quer na região promotora. De grande utilidade nos estudos familiares da PKD foi a descoberta de polimorfismos de microssatélites no intron 11, bem como de um polimorfismo C/A na posição 1705, os quais podem ser utilizados para seguir os haplótipos herdados dos progenitores.

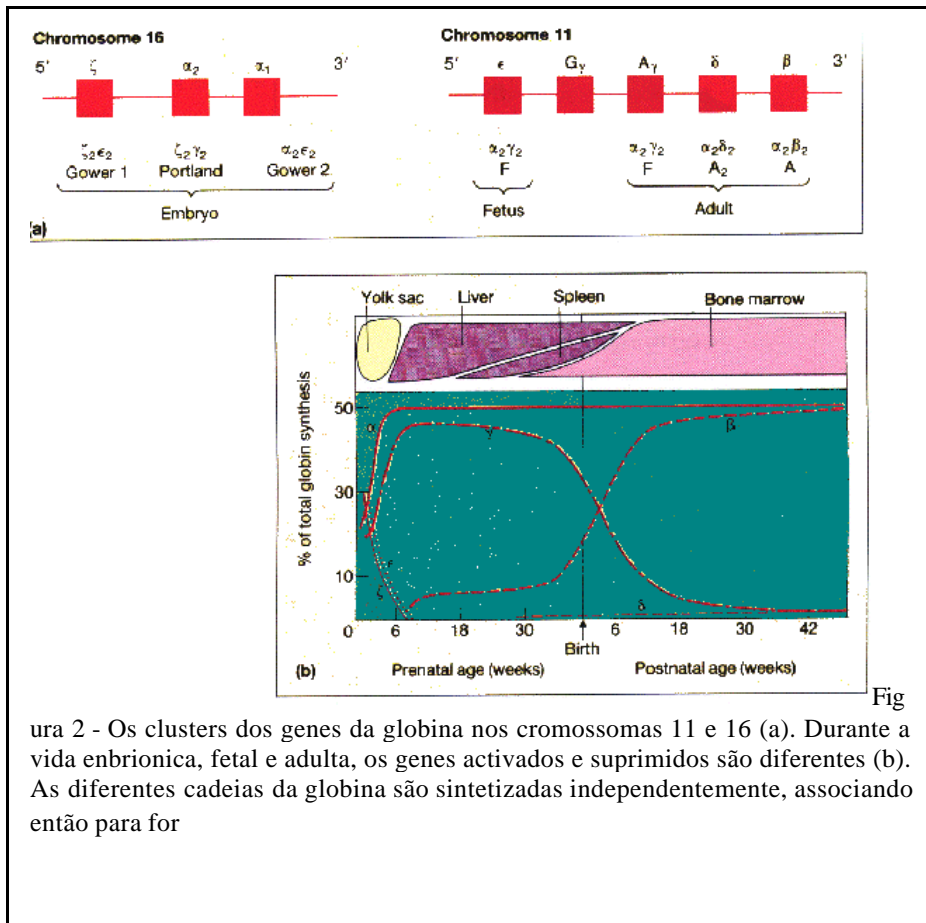
### 5.1.1.3 - Deficiência em δ-aminolevulinato sintetase (Anemia sideroblástica)

A enzima δ-aminolevulinato sintetase (ALAS) catalisa o primeiro passo da síntese do grupo heme. Nos vertebrados existem 2 formas de ALAS: uma forma transcrita constitucionalmente em todas as células, e uma forma apenas existente nas células eritróides.

No Homem, a forma eritróide (ALAS2) é codificada num gene localizado no cromossoma X (Xp11.21), estando a sua actividade reduzida na anemia sideroblástica ligada ao cromossoma X. Que a origem da anemia sideroblástica está relacionada com este gene ficou claramente demonstrado quando em 1992 Bishop descreveu a primeira mutação neste gene, e a sua segregação ao longo de 8 gerações de uma família afectada. A identificação molecular desta mutação, permitiu ainda pela primeira vez desenhar métodos de diagnóstico e estudo de transmissão familiar para esta doença.

### 5.1.2 - Talassémias (anomalias das α e β-globinas)



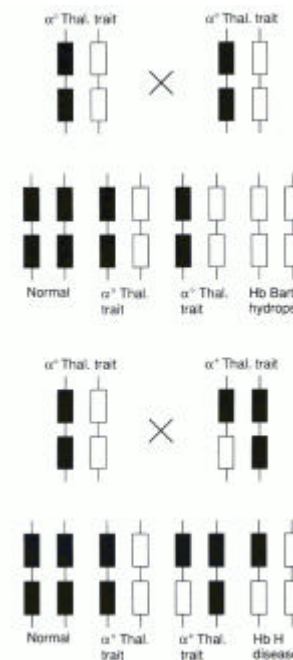


Fig

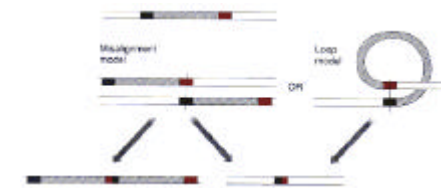
ura 2 - Os clusters dos genes da globina nos cromossomas 11 e 16 (a). Durante a vida enbrionica, fetal e adulta, os genes activados e suprimidos são diferentes (b). As diferentes cadeias da globina são sintetizadas independentemente, associando então para for

Os *loci* genéticos das  $\alpha$ - e  $\beta$ -globinas incluem vários genes arranjados em “clusters”. O cluster humano da  $\alpha$ -globina, inclui um gene enbrionico ( $\zeta_2$ ), dois genes fetais/adultos ( $\alpha_2$  e  $\alpha_1$ ), e vários pseudogenes ( $\psi\zeta_1$ ,  $\psi\alpha_2$  e  $\psi\alpha_1$ ) bem como um gene sem função conhecida ( $\theta_1$ ). Os genes da globina são expressos em níveis muito elevados nas células eritróides, mas não são expressos em nenhuma outra célula.

A  $\alpha$ -talassémia resulta da produção deficiente das cadeias  $\alpha$  da hemoglobina embriónica ( $\alpha_2\epsilon_2$ ), fetal ( $\alpha_2\gamma_2$ ) e adulta ( $\alpha_2\beta_2$ ).



As formas mais frequentes de alfa talassémia consistem na delecção de um ou ambos os genes  $\alpha$  do cromossoma 16. Desta forma, os portadores de  $\alpha$ -talassémia possuem 3 ( $-\alpha/\alpha\alpha$ ) ou dois ( $-\alpha/-\alpha$ ,  $---/\alpha\alpha$ ) genes, ao passo que os doentes possuem apenas 1 gene ( $---/-\alpha$ ). Os doentes com síndrome de “Hb Bart’s



hydrops fetalis” não possuem genes  $\alpha$  ( $---/---$ ).

Outras causas menos frequentes de  $\alpha$ -talassémia são mutações pontuais no gene da  $\alpha$ -globina, e muito raramente em elementos reguladores destes genes.

A  $\beta$ -talassémia caracteriza-se por uma síntese reduzida da  $\beta$ -globina, levando a um desequilíbrio da síntese das cadeias  $\alpha$ /não  $\alpha$  que é o factor major na gravidade da doença.

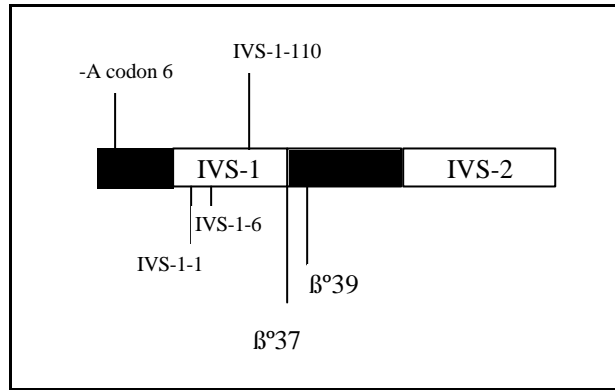
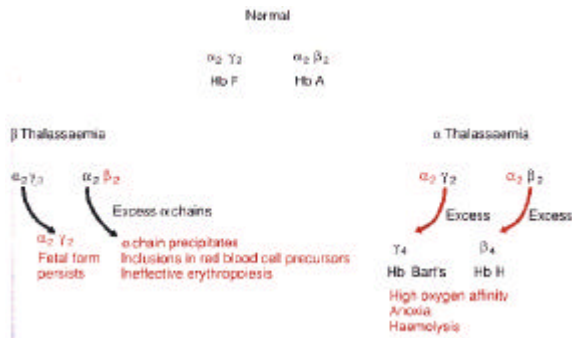


Figura 3 - Posições das mutações do gene da  $\beta$ -globina mais frequentes na zona mediterrânica.



Os genes da  $\beta$ -globina estão arranados num cluster no braço curto do cromossoma 11, na ordem 5'- $\epsilon$ - $\gamma$ -A $\gamma$ - $\psi$ - $\delta$ - $\beta$ -3'. O cluster contém muitos polimorfismos de restrição (RFLP). Uma característica destes polimorfismos, é que a sua associação às várias formas de  $\beta$ -globina não é ao acaso. Assim, em cada população foi encontrado um número limitado de haplótipos  $\beta$  (padrão de arranjos dos RFLP), pelo que a análise destes haplótipos fornece informação clínica (dependente do background genómico em que as mutações ocorrem). No início dos

anos 80 esta era a estratégia de eleição para o screening da  $\beta$ -talassémia. Esta tecnologia era então complementada com a clonagem e sequenciação dos genes mutantes, completando-se

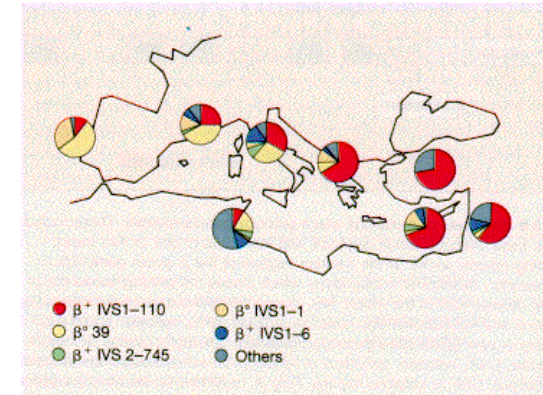


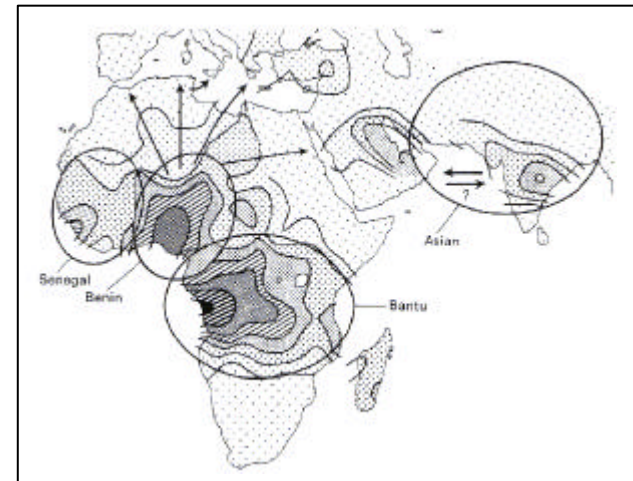
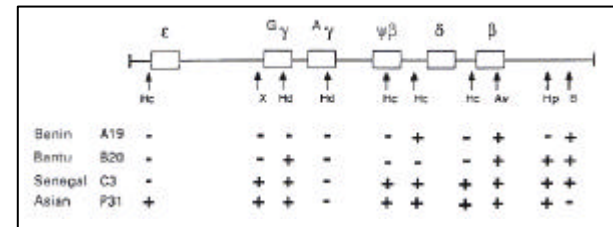
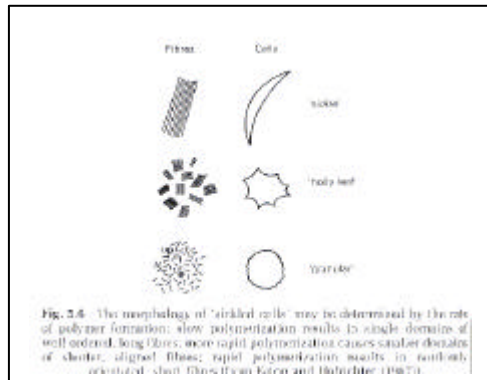
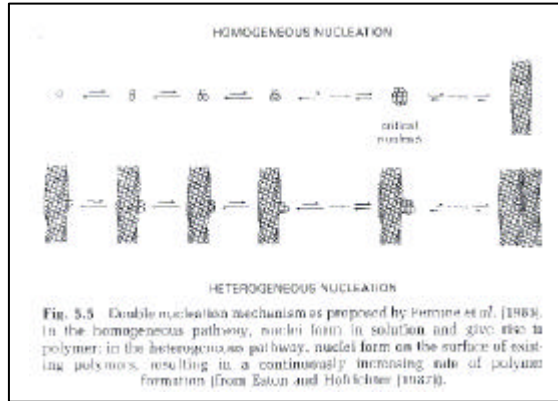
Figura 4 - Distribuição geográfica de algumas mutações na  $\beta$ -globina

assim a caracterização destes. Com o desenvolvimento da tecnologia de PCR em 1985, foi possível amplificar o gene da  $\beta$ -globina directamente a partir de DNA genómico, sequenciando directamente os produtos amplificados, facilitando deste modo a caracterização das mutações.

Ao contrário das mutações causadoras de  $\alpha$ -talassémia, as  $\beta$ -talassémias são geralmente formas não delecionadas, podendo ocorrer na região codificante do gene, ou na região promotora. Existem mutações particularmente frequentes em determinadas comunidades, o que aliado à tecnologia de PCR em combinação com a utilização de reacções de restrição, simplifica grandemente a detecção e caracterização de mutações em estudos de screening, e de modo particular em estudos familiares.



### 5.1.3 Drepanocitose

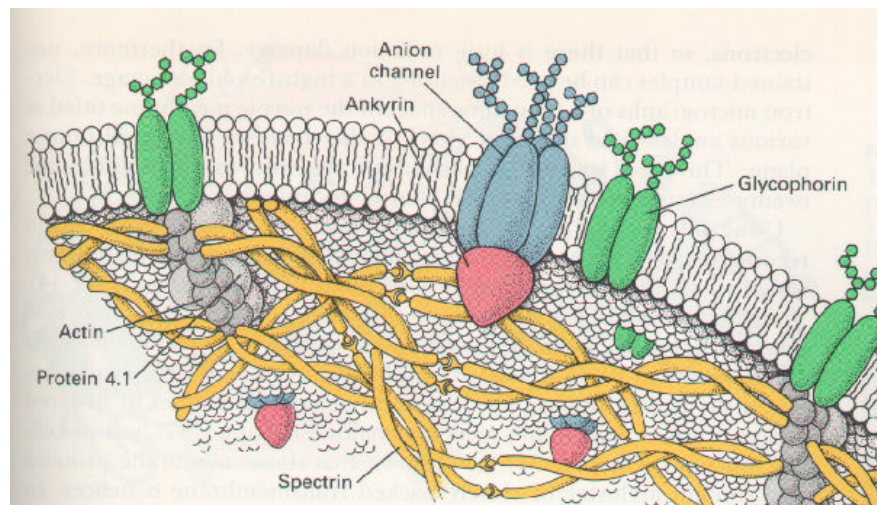


### 5.1.4 - Anemias Hemolíticas esferocíticas

As células vermelhas devem as suas propriedades mecânicas como a resistência, elasticidade e deformabilidade a uma rede proteica (citoesqueleto) que suporta o folheto bilipídico da sua membrana celular, do lado citoplasmático, e a proteínas integrais da membrana. As proteínas do citoesqueleto ligam-se às proteínas integrais da membrana, formando uma estrutura responsável pela manutenção da forma do eritrócito, e pelas suas propriedades mecânicas.

Entre as proteínas integrais de membrana do eritrócito encontram-se o transportador de aniões (banda 3), as glicoforinas, o transportador de glicose, uma variedade de transportadores de cátions, e as proteínas dos grupos sanguíneos. O citoesqueleto é essencialmente formado por

espectrina, actina, anquirina, proteína 4.1 e proteína 4.2. Os heterodímeros de espectrina organizam-se numa rede em média hexagonal, a qual se liga à banda 3 por pontes formadas pela anquirina, ou pela actina, banda 4.1 e glicoforina.



Figxxx – esquema da organização do citoesqueleto do eritrócito.

A esferocitose hereditária (HS), eliptocitose hereditária (HE) e a sua forma agravada Poiquilocitose hereditária (HP) são um conjunto heterogéneo de anemias hemolíticas congénitas. Estas patologias resultam de alterações nas proteínas integrais da membrana eritrócitária, e seu citoesqueleto.

#### 5.1.4.1 - Esferocitose Hereditária (HS)

A esferocitose hereditária é a patologia mais frequente da membrana eritrócitária. A sua frequência foi estimada em 1:5000 nascimentos, mas este valor pode ser uma estimativa por baixo se considerarmos que os portadores podem ser assintomáticos, e que cerca de 1% dos dadores de sangue apresentam fragilidade osmótica aumentada.

A esferocitose Hereditária é uma patologia heterogénea em termos da sua apresentação clínica, modo de transmissão e hereditariedade. Pode ser descrita como uma anemia hemolítica caracterizada pela presença de eritrócitos esféricos, com baixa resistência osmótica, os quais são selectivamente retidos e destruídos no baço. Estas características parecem dever-se a um enfraquecimento das ligações entre o citoesqueleto e a membrana, as quais podem ser devidas a deficiências em várias proteínas. Em cada uma destas proteínas, surgem variadíssimas mutações causadoras deste fenótipo, o que justifica a sua heterogeneidade clínica e de transmissão.

Cerca de 40% dos casos de HS apresentam mutações afectando o gene da anquirina. Mutações tipo “frameshift” causam alterações estruturais muito grandes na proteína, causando formas dominantes (anquirina de estuttgart, e anquirina de Marburg). Alterações menos severas são causadas por mutações pontuais, e originam formas recessivas (anquirina de Düsseldorf, anquirina de Walsrode). Em alguns casos, as alterações só são visíveis no mRNA, já que resultam de junções *splicing* anormais (anquirina de Praga, anquirina de Rakivnik).

Cerca de 20 a 30% dos casos de HS resultam de mutações no gene da proteína banda 3 (mutação de Praga e mutação de Coimbra). Outro tipo de mutações (mutante de Tuscaloosa e mutante de Montefiore), ocorrendo na zona citoplasmática da banda 3, causam perturbações na sua ligação à proteína 4.2, resultando num menor nível desta proteína na membrana eritrócitária.

Podem ainda causar esferocitose mutações nos genes da proteína 4.2 (ELB42), e das alfa e beta espectrinas (SPTA1 e SPTB).

#### 5.1.4.2 - Deficiência grave de proteína 4.2

A total, ou quase total ausência de banda 4.2 resulta num quadro clínico diferente da HS. Os esferócitos não existem, e a fragilidade osmótica é pouco alterada. A hemólise é muito severa, e o padrão de

transmissão recessivo. Foram identificados 4 mutantes do gene da proteína 4.2, os quais estão na origem deste quadro clínico (mutante de Nippon, mutante de Toseur, e mutante de Lisboa).

#### 5.1.4.3 - Eriptocitose e poiquilocitose Hereditária

A eliptocitose hereditária, e a sua forma agravada a poiquilocitose hereditária constituem um grupo heterogéneo de patologias caracterizadas pela presença de eritrócitos com forma elíptica. O quadro clínico é muito variado, desde condições assintomáticas até à presença de hemólise clinicamente significativa. A Eriptocitose apresenta-se habitualmente com transmissão dominante, enquanto a poiquilocitose apresenta transmissão recessiva. A poiquilocitose apresenta ainda fragmentação dos eritrócitos microesferocitose e instabilidade térmica.

Geneticamente a eliptocitose deriva de um conjunto de mutações localizadas na região de dimerização das espectrinas, ou de mutações originando uma redução de proteína 4.1.

São conhecidas 25 mutações de  $\alpha$ -espectrina, as quais se localizam em zonas onde perturbam o processo de auto-associação. Dependendo de o alelo não mutado ser de alta ou baixa expressão, assim a mutação pode ter uma representação suave ou mais acentuada.

Todas as mutações de  $\beta$ -espectrina conhecidas que originam HE localizam-se na repetição  $\beta$ 17, a qual contém o local responsável pela auto-dimerização. Mutações pontuais na hélice 2 ou mesmo na hélice 1 desta repetição possuem um padrão de transmissão recessivo. No entanto outras mutações, ocasionando truncagens de  $\beta$ -espectrina são transmitidas segundo um padrão dominante.

Cerca de 30% dos casos de HE resultam de alelos *Null* (não expressos ou não funcionais) do gene da proteína 4.1. Esta condição é clinicamente silenciosa na forma heterozigótica. Duas mutações foram descritas: uma mutação pontual no codon de iniciação

“downstream” (o único existente no precursor do eritrócito), e uma deleção de um aminoácido no local de ligação do complexo actina-espectrina. Outros exemplos de defeitos a este nível constituem os alelos que originam erros de *splicing*.

Mutações raras originam a falta de glicoforina C. Como a proteína 4.1 se liga à glicoforina, a falta desta origina a falta de proteína 4.1.

## 5.2 DOENÇAS GENÉTICAS EM HEMOSTASE

A hemostase, constitui um riquíssimo campo de intervenção da genética molecular no diagnóstico, estudos familiares e rastreio populacional, já que o número de mutações e polimorfismos associados a doenças genéticas em hemostase é relativamente elevado (tabela 1). As técnicas utilizadas para o estudo das doenças genéticas em hemostase cobrem o espectro completo das técnicas de genética molecular, sendo portanto um ótimo exemplo da larga gama de tecnologias hoje disponíveis para estes estudos (tabela 1).

### 5.2.1 - Resistência à proteína C Activada (mutação *FV-Leiden*)

Recentemente, apenas 10% dos indivíduos que apresentavam tromboembolismo venoso possuíam uma anomalia genética predispondo à doença. Este grupo de doentes tinha uma deficiência absoluta ou funcional de um dos principais componentes dos mecanismos de regulação da coagulação (ATIII, Prot.C, Prot.S). Recentemente, um novo mecanismo foi encontrado para justificar esta patologia: a Resistência à Proteína C Activada (APCR). Este mecanismo, estando implicado na origem de cerca de 50% dos casos de trombose venosa (dependendo da selecção dos casos), parece ter origem numa mutação pontual do gene do Factor V (Factor V de Leiden).

É hoje habitualmente aceite que todos os casos de APCR têm origem nesta mutação, a qual está presente em cerca de 5% da população normal, sendo cerca de 20-50% dos doentes com tromboembolismo

portadores heterozigóticos desta mutação. No entanto, não parece haver um aumento da frequência de portadores da mutação em doentes que tenham sofrido enfarte do miocárdio ou ataque cardíaco, sugerindo um papel limitado na doença arterial.

A presença da mutação, a qual pode ser facilmente detectada por PCR seguido de corte por enzima de restrição, aumenta em cerca de 8 vezes o risco de tromboembolismo. Este risco pode ainda ser aumentado por factores adicionais, que podem ser genéticos ou adquiridos.

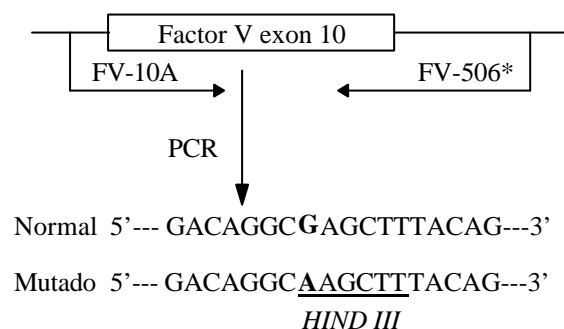


Figura 5 - Estratégia de identificação da mutação FV-Leiden. O exon 10 do factor V é amplificado por PCR. O fragmento amplificado contém um local de restrição para Hind III no caso de existir a mutação, mas não no caso do gene não mutado.

### 5.2.2 - Mutações no gene da Protrombina

Tabela 1 - Resumo de polimorfismos identificados em genes responsáveis por doenças hemorrágicas e trombóticas hereditárias

Gene	Localização <sup>a)</sup>	Tipo de polimorfismo	Enzima de restrição	Nº Alelos	Heterozigoidade (%)
F VIII	Intron 13	Repetição CA	-	8	80
	Intron 18	RFLP	BclII	2	42
	Intron 19	RFLP	Hind III	2	42
	Intron 22	RFLP	XbaI	2	48
	Intron 22	Repetição CA	-	2	44
F IX	Intron 1	RFLP	DdeI	2	36
	Intron 3	RFLP	XmnI	2	41
	Intron	RFLP	TaqI	2	45
	Extr. 3'	RFLP	HhaI	2	48
vWF	Intron 2	RFLP	SmaI	2	45
	Intron 2	RFLP	HhaI	2	45
	Exon 14	RFLP	AccI	2	46
	Exon 18	RFLP	RsaI	2	45
	Intron 19	RFLP	MspI	2	44
	Exon 28	RFLP	HphI	2	50
	Exon 28	RFLP	BstEII	2	46
	Intron 40	Repetição(TCTA) <sub>n</sub>	-	8	75
PC	7 Kb 5'	RFLP	MspI	2	42
	Exon 1	RFLP	RsaI	2	48
	Exon 6	RFLP	XbaI	2	48
	Exon 8	Sequência	-	2	45
PS	Exon 15	RFLP	BstXI	2	49
ATIII	Extr. 5'	Distância	-	2	37
	Exon 4	RFLP	PstI	2	37
		RFLP	DdeI	2	50
		Repetição (ATT) <sub>n</sub>	-	>10	28

a) Nos introns a numeração corresponde ao nucleótido, nos exons a numeração corresponde ao aminoácido; Extr.=extremidade

### **5.2.3 - Doença de von Willebrandt (Mutações no gene do vWF)**

A doença de von Willebrand (VWD) é a mais comum doença hemorrágica no homem, sendo causada por uma deficiência qualitativa ou quantitativa no factor de von Willebrand (VWF). Esta doença pode ser transmitida segundo padrões dominantes ou recessivos, de acordo com a mutação em causa.

A clonagem do gene do VWF (180 Kb, contendo 52 exons, e localizado em 12p12) permitiu a identificação molecular das mutações responsáveis pela VWD. Os estudos iniciais permitiram a identificação de deleções neste gene. Contudo, e devido à dimensão do gene do VWF, a maior parte dos estudos posteriores centraram-se em zonas do gene importantes para funções determinadas da proteína, nomeadamente a dimerização e processamento intracelular (exons 1-16), ligação ao factor VIII (exons 17-25), ligação a colagénio (exons 28-34). Neste sentido, a classificação funcional do VWF constitui um valioso auxiliar no estudo genético de cada doente, já que permite concentrar esforços numa área restrita de um gene excessivamente longo para ser estudado por inteiro.

### **5.2.4 - Trombose familiar (Mutações nos genes da Antitrombina III, Proteína C e Proteína S)**

Existem essencialmente 2 mecanismos inibidores da actividade das proteínases de serina envolvidas na coagulação: antitrombina III (ATIII), e o sistema Proteína C (PC)- Proteína S (PS)-trombomodulina (TM). A falha dos mecanismos inibidores predispõe para a trombose, e esta predisposição pode ser hereditária, como ficou demonstrado pela primeira vez em 1965 para a antitrombina (Egeberg et al., 1965). Desde então ficou demonstrada uma forte relação entre deficiências da PC e PS e tromboembolismo venoso.

A identificação de deficiências hereditárias a nível molecular da deficiência em antitrombina III ocorreu em 1984. Desde então, os avanços na genética molecular dos inibidores permitiu a identificação de outros defeitos genéticos, permitindo assim estudar a transmissão nas famílias afectadas.

O gene da antitrombina III está localizado em 1q23-25, tendo 13.5 Kb e 7 exons. A sequenciação dos exons permitiu a identificação de um número considerável de mutações pontuais originando deficiência tipo I. A mutação mais comum consiste numa alteração da *fase de leitura* (Shift mutation), a qual resulta na presença de um codão stop prematuro. Outras alterações frequentes envolvem a formação de codões stop directamente resultantes de uma mutação pontual.

A deficiência tipo II parece resultar em todos os casos de uma mutação pontual ocasionando a alteração de um aminoácido. Parece existir uma forte correlação entre a localização do aminoácido substituído, e a alteração funcional observada: alteração dos aminoácidos 24,27 e 129 (aminoácidos positivamente carregados), ou dos aminoácidos 41 e 99 (aminoácidos electricamente neutros) origina uma redução na afinidade para a heparina (negativamente carregada); mutações nos aminoácidos 382-394 resultam na deficiente inibição da trombina; mutações nos aminoácidos 402-407 e 429 produzem mutantes com alterações múltiplas.

O gene da PC (uma glicoproteína dependente da vitamina K) foi localizado em 2q13-14. Trata-se de um gene com 12 Kb e 9 exons. A sequenciação deste gene permitiu a identificação de várias mutações associadas à deficiência de Proteína C tipo I com um predomínio de substituições de aminoácidos.

O gene da PS (uma proteína do plasma dependente da vitamina K) está localizado no cromossoma 3 (3p11.1-11.2, possui cerca de 80Kb de DNA e contem 15 exons. Na sua vizinhança encontra-se ainda um pseudo-gene com uma estrutura muito homóloga. Os estudos para a

identificação de mutantes nesta proteína são ainda poucos, mas foram já encontrados vários mutantes, responsáveis pelas deficiências de PS observadas.

### **5.2.5 - Hemofilias (Mutações nos genes dos factores VIII e IX)**

A hemofilia A é uma das doenças hemorrágicas mais frequentes, sendo causada por uma deficiência do factor VIII (FVIII) circulante, cujo gene está localizado no cromossoma X (Xq28). A doença é heterogénea, tanto a nível molecular, como a nível da severidade clínica. A dimensão do gene do FVIII impediu até à pouco tempo atrás a análise genética das mutações causando o gene, dificultando a caracterização da transmissão familiar. O uso de polimorfismos intragénicos ou extragénicos permitiu no entanto recentemente iniciar uma caracterização precisa do cromossoma X dos familiares de indivíduos afectados. Esta estratégia, não pode no entanto ser universalmente empregue, já que nem sempre existem marcadores polimórficos associados à mutação em causa. Nestes casos, apenas a determinação da mutação específica permite estudar com clareza os indivíduos em causa. Neste intuito, um esforço considerável foi desenvolvido no sentido de empregar métodos de identificação de alterações genéticas pontuais, obviando à sequenciação de grandes extensões genómicas. Com este propósito, técnicas como “single strand conformation polymorphism” (SSCP), “denaturing gradient gel electrophoresis (DGGE) e chemical cleavage of mismatch (CCM) foram utilizadas com sucesso na identificação de mutações. Estes estudos permitiram determinar que a maioria dos doentes com doença moderada ou suave possuem mutações tipo missense, enquanto a maioria dos doentes graves possuem mutações tipo “frameshift”. Em alguns casos de doença grave foi ainda possível identificar um defeito no *splicing* dos exons 22-23.

A hemofilia B é uma doença hereditária de transmissão recessiva, devida à total ausência, ou quantidades reduzidas de factor IX (FIX).

O gene do FIX foi totalmente sequenciado, contendo 8 exons. Cerca de 40% dos aminoácidos do FIX parecem ser essenciais para a função da proteína, pelo que não é de admirar que a doença seja causada por um largo espectro de mutações. Assim, e uma vez que não é possível dirigir o estudo para uma região específica do gene, as técnicas de estudo que têm sido utilizadas para identificar as mutações causadoras da Hemofilia têm sido essencialmente as mesmas técnicas acima descritas.

### **5.3 - HEMOCROMATOSE**

A hemocromatose hereditária (HH) é uma doença genética de transmissão autossómica recessiva. Contrariamente ao que se pensava, sabemos hoje que a hemocromatose não é uma doença rara, já que a sua frequência é comparável à da anemia da “Sickle cell”, fibrose cística ou distrofia muscular (Nichols & Bacon, 1989). Foi estimado que só nos Estados Unidos devem existir entre 600,000 e 1,000,000 de doentes, bem como cerca de 27,000,000 de indivíduos portadores do gene (Nichols & Bacon, 1989).

A HH é caracterizada por uma falha na regulação da absorção do ferro da dieta, mantendo-se a absorção mesmo na presença de altos níveis de ferro armazenado (Alper et al., 1951; Cox & Peters, 1978; Williams et al., 1986; Lynch et al., 1989; Whittaker et al., 1989). Os indivíduos com HH possuem depósitos de ferro superiores a 4 g, não sendo raros indivíduos com depósitos superiores a 20 g, contrastando assim com os habituais 500-1000 mg nos indivíduos normais (Nichols & Bacon, 1989). O quadro completo de HH envolve a deposição de ferro em vários órgãos, nomeadamente: fígado, resultando em cirrose hepática; coração com diminuição da função cardíaca e perturbação do ritmo; articulações com formação de poliartropatia; pele, com formação de pigmentação dérmica característica; e glândulas endócrinas, originando falhas endócrinas como diabetes e gonadopatias (Milder et al., 1980).

A hemocromatose hereditária, não sendo uma doença para a qual exista uma cura é no entanto facilmente tratável por flebotomias (Bomford & Williams., 1976; Niederau et al., 1985). O tratamento ainda que seja meramente correctivo, e por isso exija intervenção durante toda a vida do doente, é bastante eficaz, consistindo numa primeira fase em 1 a 2 flebotomias semanais (tratamento intensivo), a que se seguem, após a depleção dos depósitos de ferro, flebotomias mensais ou trimestrais (tratamento de manutenção). Uma vez que a sobrecarga de ferro é passível de correcção, a hemocromatose hereditária, como modelo de estudo da interacção entre o metabolismo do ferro e o sistema imunológico, tem a grande vantagem de nos fornecer dados sobre a direcção das interacções. Com efeito, se a sobrecarga de ferro exercer um efeito fisiologicamente relevante no sistema imunológico destes indivíduos, a remoção da sobrecarga deverá corrigir esse defeito. Inversamente, se o sistema imunológico tiver uma acção relevante na homeostase do ferro, então esta deverá preceder a sobrecarga de ferro, não sendo corrigida com a remoção de ferro.

A origem precisa do erro na regulação da absorção do ferro na HH constitui ainda hoje um mistério, tal como o é a identidade do gene ou genes responsáveis pela doença. No entanto, e dada a aparente falta de relação entre a doença e perturbações da função imunológica clássica (ocorrência de infecções, a já descrita estreita associação entre o fenótipo HLA-A3 e o caractere hemocromatose (Simon, 1975, 1977a, 1977b) tem vindo a ser interpretado pelos geneticistas como indicador da estreita ligação física entre o *locus* HLA-A e o gene da hemocromatose (Simon, 1977a).

### **5.3.1 - Estudos de marcadores genéticos no locus do HLA**

Devido à estreita associação entre o *locus* HLA-A no cromossoma 6 (6p21.3) o gene da hemocromatose, nos últimos anos tem-se assistido a um crescente ritmo de estudo de marcadores polimórficos tipo

microsatélite nesta região do genoma humano, procurando-se definir marcadores mais estreitamente ligados à hemocromatose que o próprio HLA-A3. No momento presente estão disponíveis vários microsatélites, cuja segregação familiar permite definir haplótipos estendidos, dando maior rigor que a determinação HLA à classificação dos familiares como homozigóticos ou heterozigóticos, bem como na determinação de indivíduos potencialmente portadores do gene em estudos de “screening” da população. Deve no entanto notar-se que para o estudo da população normal, os estudos genéticos na hemocromatose são muito limitados, servindo apenas de elemento de apoio aos dados bioquímicos mais relevantes como os níveis séricos de ferritina, de transferrina, de ferro, e a taxa de saturação da ferritina

### **5.3.2 - Estudos dos IRE e IRP**

Os IRE (do inglês Iron Responsive Elements) são sequências existentes nos mRNA do receptor da transferrina, ferritina, aconitase mitocondrial e eALAS (5-aminolevulinato sintetase). Trata-se de sequências muito conservadas na evolução (95% entre o homem, o rato e a galinha no caso do receptor da transferrina), e que possivelmente permitem a formação de uma estrutura no mRNA em forma de ansa (Kühn, 1994). Estas sequências são as principais responsáveis pela regulação pós-translacional dos genes a que pertencem, conferindo-lhes a capacidade de se modularem dependendo da concentração intracelular de ferro (Kühn, 1994). Esta regulação é mediada por uma proteína citoplasmática denominada IRP (do Inglês Iron Responsive Protein), mas previamente também conhecida como IRF, IRE-BP e FRP (Müllner et al., 1989; Rouault et al., 1988; Walden et al., 1988).

O efeito regulador do IRE (e consequentemente da IRP) depende da localização do IRE no mRNA. No mRNA do receptor da transferrina, os 5 IREs estão presentes na região 3' não traduzida, pelo que a ligação da IRP aumenta a estabilidade do mRNA, prolongando o seu

tempo de semi-vida (Kühn, 1994). No caso do mRNA dos genes da ferritina, da aconitase mitocondrial e da eALAS, o IRE encontra-se situado na região 5' não traduzida do mRNA, a curta distância da sequência CAP, pelo que a ligação da IRP bloqueia a iniciação da tradução pelos ribossomas (Kühn, 1994).

A IRP é uma proteína bi-funcional (Klausner et al., 1993; Hirling et al., 1994). Na presença de ferro, a sua capacidade de ligar aos IRE é mínima, o que pode derivar da incorporação de 1 grupo 4Fe-4S na proteína. Nestas condições, a IRP é capaz de funcionar como aconitase citoplasmática, catalisando a formação de isocitrato (Emery-Goodman et al., 1993). Pelo contrário, na ausência de ferro, a IRP liga-se com grande especificidade aos IRE, perdendo a actividade de aconitase.

O sistema imunológico parece utilizar este duplo papel do IRP para controlar a sua actividade, e assim controlar o metabolismo intracelular do ferro.

Experiências descritas em 1993 por Drappier et al., comprovam a existência de efeitos reguladores de citocinas, nomeadamente do IFN- $\gamma$  e TNF- $\alpha$  na indução da ligação de IRP a IRE em macrófagos em cultura devido à modulação da síntese de óxido nítrico proveniente da via da L-arginina (Drappier et al., 1993).

O óxido nítrico, mercê da sua reduzida dimensão tem acesso directo ao núcleo 4Fe-4S da IRP, onde induz a sua degradação (Kühn, 1994). Assim, a presença de óxido nítrico (regulada pelo IFN- $\gamma$  e TNF- $\alpha$ ) é um factor tendente a aumentar a "pool" de IRP capaz de ligar ao IRE, diminuindo a "pool" capaz de funcionar como aconitase (Kühn, 1994).

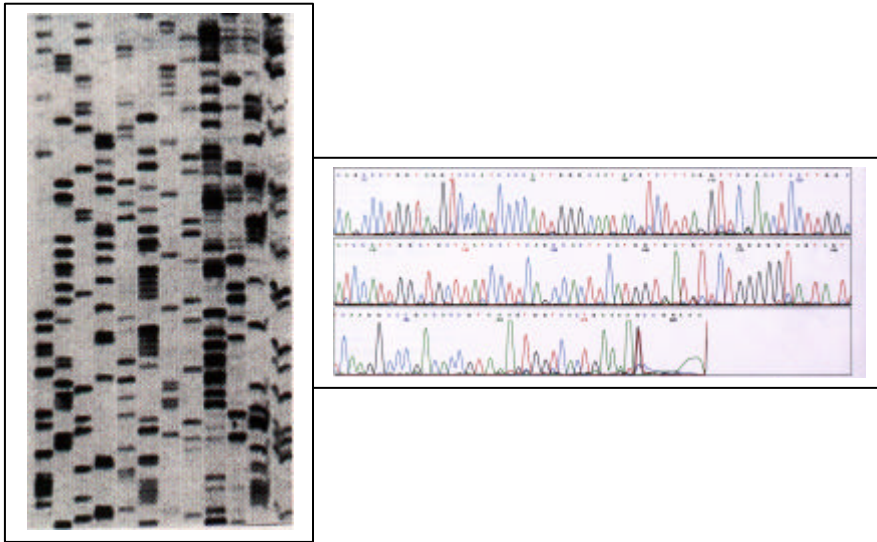
### **5.3.3 - Estudos Genéticos do Repertório da Célula T**

Os resultados do grupo da Prof. Maria de Sousa, revelando a existência de uma correlação entre a razão CD4/CD8 e os níveis de ferro dos doentes, levou este mesmo grupo a procurar anomalias no repertório das células T nesta patologia. Este estudo revelou a existência de uma correlação entre anomalias na representação do TCR V $\beta$ 6.7 entre as células CD8+ e a severidade da expressão clínica da doença, nomeadamente o desenvolvimento de cirrose hepática. Estes resultados, sugerindo intervenções do sistema imune na regulação de sistemas internos, reforçam a visão crescente do sistema imune como um sistema regulador, não só das potencialmente nocivas interferências externas, mas também das frequentemente fatais perturbações dos delicados equilíbrios homeostáticos internos do organismo. Esta interpretação faz com que estes resultados tenham implicações clínicas de grande relevo não só na hemocromatose hereditária mas também em todas as formas de sobrecarga de ferro (Porto, 1993; Cabeda, 1995).



# Capítulo 6

## 6 MÉTODOS DE ESTUDO EM GENÉTICA MOLECULAR



## 6.1 MÉTODOS NÃO COMERCIAIS

### 6.1.1 Preparação de DNA e RNA

Qualquer análise em genética molecular requer, obviamente, o estudo do DNA ou do RNA, pelo que o primeiro passo em qualquer técnica genética consiste no isolamento e purificação de uma ou mesmo das espécies de ácidos nucleicos. As variadas técnicas disponíveis para esse efeito, as quais originam DNA ou RNA com diferentes propriedades de pureza e integridade física, possuem princípios semelhantes. Todas se iniciam com uma lise suave das células a estudar, seguida de ataques enzimáticos e/ou químicos para destruir os componentes proteicos da mistura. Finalmente, a purificação do DNA ou RNA faz-se por um de vários métodos, de acordo com os objectivos pretendidos.

A purificação e correcto manuseamento de RNA é bem mais difícil que para o DNA. Este facto não resulta de uma maior complexidade de procedimentos, mas da maior estabilidade das RNAses. Com efeito, ao contrário das DNAses, as RNAses são extremamente estáveis, e não necessitam de cofactores para funcionarem. Desta forma, não é possível inactivá-las com a adição de quelantes do magnésio (EDTA) como acontece para as DNAses. A inactivação



Figura 6.1 – A Purificação de DNA ou RNA por adsorção em coluna é o método mais rápido disponível. O DNA ou RNA produzido é de qualidade suficiente para a realização de PCR ou RT-PCR, mas devido à fragmentação que ocorre não é utilizável para Southern Blot.

das RNAses é eficiente com a utilização de dietilpirocarbonato (DEPC), mas a alta toxicidade deste composto, aliada à necessidade da sua eliminação por autoclavagem, impede a sua utilização em todas as soluções. Em reacções enzimáticas é possível utilizar inibidores específicos de RNAses (RNase inibidor ou abreviadamente RNAsin), como o extraído do tecido placentário, para inactivar as RNAses provenientes do material celular donde é extraído o RNA. Este não é, no entanto, um método prático para o tratamento generalizado dos reagentes e material de plástico do laboratório. Se aliarmos a este facto, a presença em grandes quantidades de DNAses e RNAses nas mãos humanas, facilmente se compreende a imperiosa necessidade de utilizar luvas no laboratório de Genética Molecular, não para a protecção do operador, mas para proteger a amostra do ataque das RNAses e DNAses do manuseador da amostra e restantes materiais de laboratório.

### 6.1.2 ELECTROFORESE

A grande maioria dos métodos de genética molecular requer num determinado momento o fraccionamento de ácidos nucleicos segundo o seu comprimento. Para tal utilizam-se as técnicas de electroforese, que consistem na separação dos ácidos nucleicos

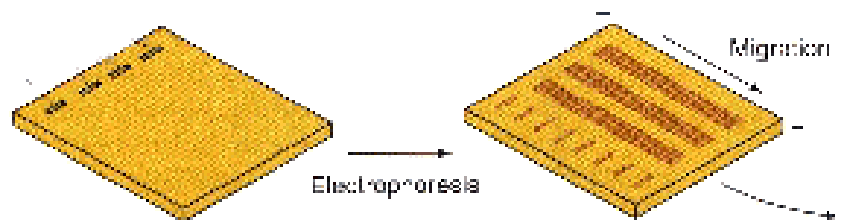


Figura 6.2 – Separação de ácidos nucleicos por electroforese em gel horizontal.

numa matriz porosa (habitualmente géis de agarose, ou acrilamida) sob a força de um campo eléctrico (os ácidos nucleicos têm carga negativa, pelo que migram em direcção ao polo positivo).

A matriz na qual os ácidos nucleicos devem ser separados depende essencialmente do tamanho dos fragmentos a separar, mas também do destino final a dar a estes uma vez separados. Existem diferentes tipos matrizes:

#### 6.1.2.1 Géis de Agarose

- **agarose** - A agarose tem relativamente à acrilamida a vantagem de constituir uma matriz não tóxica de muito fácil preparação (basta solubilizar a agarose em pó num tampão (TAE ou TBE) a quente, e deixar então arrefecer. Várias agaroses existem, as quais permitem separar fragmentos com mais ou menos nucleótidos. A agarose normal, permite uma boa resolução para fragmentos relativamente grandes, utilizando-se em baixas concentrações (0.8-2%). Já a agarose Nusieve (FMC-bioproducts) permite separar com grande resolução fragmentos com menos de 1000bp, pelo que se adapta melhor aos fragmentos habitualmente obtidos por PCR. Uma variação destas duas agaroses (Nusieve 3:1), não é mais que uma mistura de 3 partes de Nusieve com uma parte de agarose normal, o que origina um gel com uma viscosidade aceitável em altas concentrações (normalmente até 4%, tal como a Nusieve), mas com um poder de resolução superior quer à Nusieve, quer à agarose normal. Existem ainda agaroses com baixa temperatura de fusão. Estas agaroses têm a desvantagem de ser mais sensíveis a aumentos de temperatura durante a electroforese, mas a vantagem de facilitarem a purificação do DNA separado.

#### 6.1.2.2 Géis de acrilamida

- **Acrilamida** - Os géis de acrilamida baseiam-se na formação de uma matriz porosa, de polímeros de acrilamida. Para a

fazer, utilizam-se monómeros de acrilamida, um reagente bifuncional (bis-acrilamida), e um gerador de radicais livres (iniciador da reacção de polimerização; normalmente peróxido de amónio), bem como um catalisador da polimerização (temed). A resolução dos géis de acrilamida é muito grande podendo facilmente separar fragmentos com apenas um nucleótido de diferença, pelo que habitualmente se utiliza como matriz nos géis de sequenciação. A grande desvantagem deste tipo de matriz consiste na sua grande fragilidade, toxicidade e complexidade de preparação.

### 6.1.3 Reacção de polimerase em cadeia (PCR)

#### 6.1.3.1 Princípios do método

O PCR (Polimerase Chain Reaction) é um procedimento rápido para a amplificação enzimática *in vitro* de segmentos específicos de DNA. A descoberta desta tecnologia teve um enorme impacto na genética molecular, provocando uma revolução de tal ordem que em 1994 foi atribuído ao “inventor” do PCR um Prémio Nobel.

A base teórica do PCR é muito simples, baseando-se na propriedade das polimerases do DNA para catalisar a formação de uma cópia de uma cadeia de DNA, apenas quando encontram uma extremidade 3' livre. Desta forma, foi possível partir de uma molécula de DNA de cadeia dupla, desnaturá-la pelo calor, baixando de seguida a temperatura até um valor que permita a ligação específica de um oligonucleótido sintético, específico para a região 5' do segmento a amplificar. Depois desta hibridação específica, a polimerase inicia então a síntese da cadeia complementar ao “molde”. Entretanto, um processo semelhante deverá ter ocorrido em simultâneo para a restante cadeia da dupla hélice inicial, pelo que no fim deste ciclo, efectivamente foi duplicada a quantidade de DNA da zona de interesse. O processo prossegue com nova desnaturação pela temperatura, repetindo-se este ciclo um número definido de vezes.

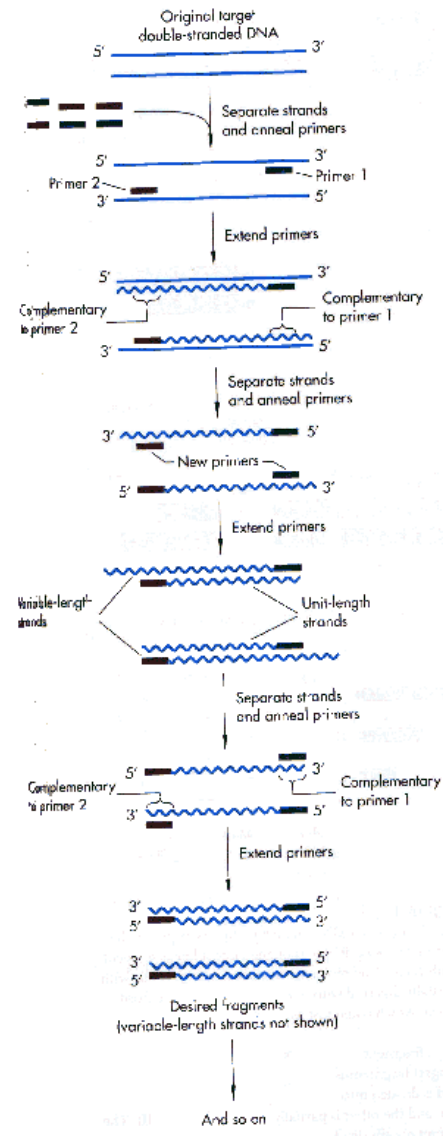


Figura 6.3 – Esquema de funcionamento de um PCR

Como em cada ciclo se duplica a quantidade de DNA de interesse que existia no início do ciclo, no final do processo amplificamos  $2^n$  vezes o segmento de DNA em que se estava interessado. Na maior parte dos casos, utilizam-se entre 24 e 35 ciclos de temperatura, pelo que no final existem  $2^{24}=16,777,216$  a  $2^{35}=34,359,738,368$  vezes mais cópias do segmento de interesse que inicialmente.

Para implementar este procedimento é necessário incluir na mistura de reacção não só o DNA a estudar, mas também os oligonucleótidos específicos (primers), desoxinucleótidos trifosfatados (dNTP's), e uma polimerase do DNA termoestável (para resistir às flutuações de temperatura necessárias para realizar as várias fases da reacção). Na prática outros componentes são também adicionados, para que as condições de reacção serem as ideais para a enzima utilizada. Um dos componentes que todas as enzimas até agora descobertas utilizam é o  $MgCl_2$ , de cuja concentração dependente em larga medida a especificidade, e

qualidade do DNA amplificado.

Os primers são utilizados num largo excesso relativamente ao DNA a ser amplificado, já que são necessárias pelo menos tantas moléculas de primer quantas as cadeias de DNA que se deseja formar. Os primers são desenhados por forma a que um tenha a sequência complementar invertida da extremidade 3' do segmento a amplificar (pelo que hibridiza directamente com esta zona do molde, deixando livre uma extremidade 3' para que a polimerase inicie o seu trabalho produzindo uma nova cadeia na direcção 5'→3'). O restante primer é desenhado por forma a ter a sequência da extremidade 5' do fragmento a amplificar. Desta forma este primer hibridiza com esta extremidade da cadeia complementar do DNA, deixando livre uma extremidade para a cópia da respectiva cadeia. Como a polimerização só termina quando a temperatura se eleva para a fase de desnaturação, no primeiro ciclo produzimos 2 tipos de fragmentos, ambos com início bem definido, mas com terminação incerta. No entanto, como os fragmentos produzidos no primeiro ciclo vão ser os moldes para a polimerização da cadeia complementar no ciclo seguinte, aos poucos, o produto preponderante tem ambas as extremidades determinadas pelos locais de ligação de ambos os primers.

### 6.1.3.2 RT-PCR

A reacção de RT-PCR (reverse transcriptase polymerase chain reaction), não é mais do que um PCR normal realizado a partir de um DNA sintetizado por transcrição reversa a partir de RNA. Assim, são necessários neste processo dois tipos de polimerases do DNA: primeiro uma DNA polimerase dependente do RNA (isto é a transcriptase reversa), e depois a polimerase normal. Este processo pode assim ser efectuado quer pela utilização sequencial de 2 enzimas diferentes, ajustando as condições da reacção à enzima,

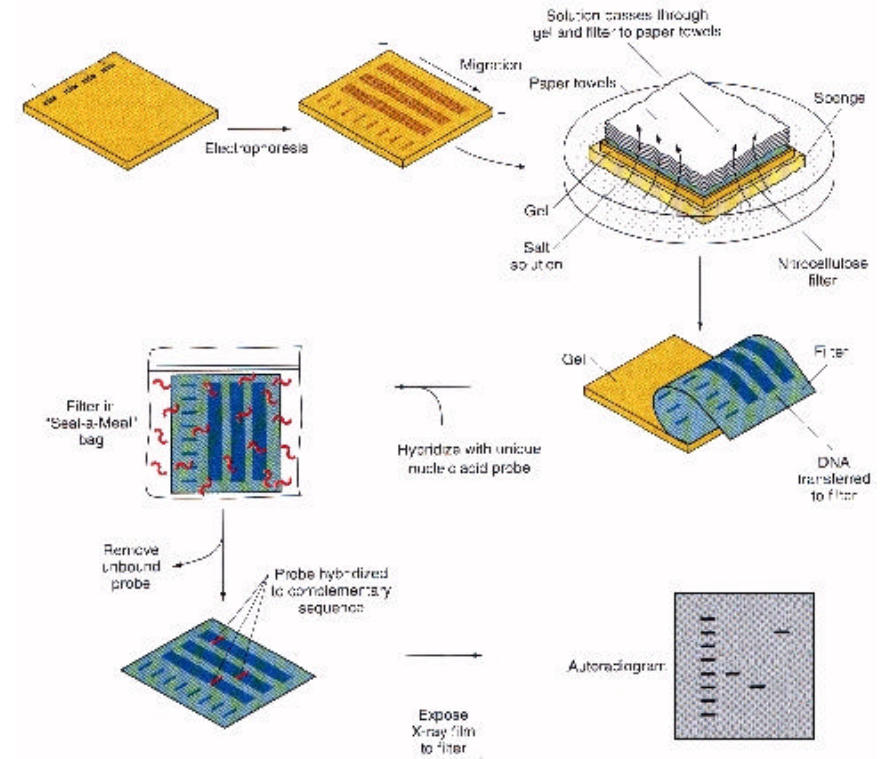


Figura 6.4 – Esquema de funcionamento de um Southern Blot

quer através da utilização de enzimas especiais que possuem ambas as actividades, embora em condições diferentes de reacção.

## 6.1.4 Southern Blot

### 6.1.4.1 Princípios do método

Frequentemente, após a separação dos ácidos nucleicos, torna-se necessário identificar o fragmento de interesse, numa mistura complexa de fragmentos separados.

A técnica de eleição para esse efeito é a técnica de Southern Blotting (ou Northern Blotting, conforme o ácido nucleico seja

DNA ou RNA respectivamente), seguida de hibridação com uma sonda específica para o fragmento de interesse. Esta técnica consiste na passagem dos fragmentos de DNA (ou RNA) separados, depois de desnaturados, para uma membrana de Nylon ou celulose, por capilaridade ou por vácuo, seguida da fixação do ácido nucleico à membrana. Esta é então utilizada numa reacção de hibridação, em que é colocada uma sonda de cadeia simples de DNA (ou de RNA) em contacto com a membrana, em condições químicas e de temperatura que asseguram que a sonda hibridiza apenas com o fragmento complementar. A ligação da sonda é então revelada por autorradiografia (no caso de sondas radioactivas), ou por quimoluminescência (ECL; Amersham).

#### 6.1.4.2 Endonucleases de restrição

##### 6.1.4.2.1 Funções biológicas dos sistemas R-M

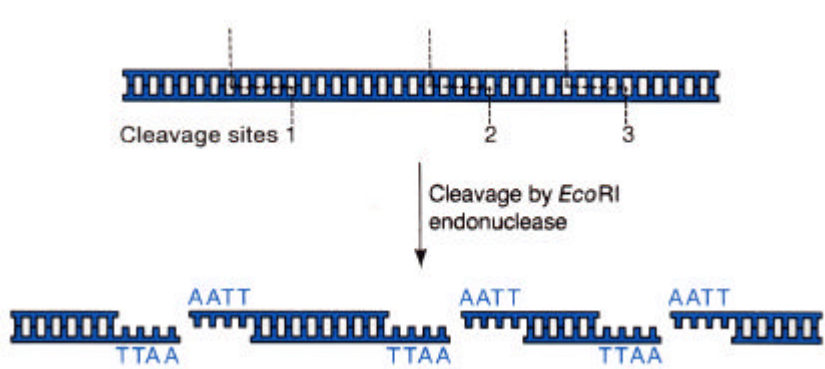


Figura 6.5 - As endonucleases de restrição fragmentam o DNA de forma específica. A especificidade é determinada pela sequência do DNA.

Pensa-se que a função biológica das endonucleases de restrição é a protecção das células contra DNA externo à célula. Esta assunção

tem como base o facto de a grande maioria das enzimas descobertas efectivamente cortarem o DNA extracelular que in vitro se insere. Simultaneamente o DNA endógeno é protegido devido à metilação específica de nucleótidos na sequência reconhecida, efectuada por uma metiltransferase específica. O conjunto da endonuclease de restrição e respectiva metiltransferase formam o que se designa por sistema de modificação de restrição (R-M system).

Existem pelo menos 4 tipos de sistemas R-M, distinguidos pela composição das suas subunidades, pelo tipo de sequências reconhecidas, e pelos cofactores necessários para a sua actividade. Cerca de 93% das enzimas caracterizadas pertencem ao tipo II. Juntamente com as enzimas da classe IIs (cerca de 5% das enzimas descritas) constituem o grosso das enzimas comercialmente disponíveis. As enzimas de tipo I (cerca de 1%) e as de tipo III (<1%) são relativamente pouco frequentes. Algumas outras enzimas existem que não podem ser incluídas em nenhuma destas classes.

##### 6.1.4.2.2 Sistemas R-M tipo II

As enzimas de tipo II são as mais simples. Reconhecem sequências de DNA simétricas, cortando entre as sequências, deixando um terminal 3' hidroxil e um terminal 5'fosfato. Requerem apenas magnésio para a sua actividade, e as metiltransferases respectivas requerem apenas s-adenosylmetionina. Reconhecem uma variedade de sequências quase ilimitada, mas poucas reconhecem sequências com menos de 4 ou mais de 8 bp.

Estas enzimas são habitualmente compostas por um homodímero, pelo que necessariamente interactivam com uma sequência de repetição invertida, já que cada subunidade reconhece o mesmo motivo em cadeias de DNA opostas.

As metiltransferases do tipo II são habitualmente compostas por um monómero, o que pode reflectir a necessidade de metilar apenas

uma das cadeias de DNA nascentes (durante a duplicação do DNA, uma das cadeias já se encontra metilada), ao contrario das endonucleases de restrição que tem que reconhecer e cortar as duas cadeias do DNA.

#### **6.1.4.2.3      *Sistemas R-M tipo IIs***

As enzimas do tipo IIs utilizam geralmente os mesmos cofactores que as enzimas tipo II, mas as suas sequências de reconhecimento são assimétricas e ininterrompidas, tendo 4 a 7 bp. O local de corte não se situa no interior da sequência de reconhecimento, mas a uma distância de até 20bp num dos sentidos. Nestes sistemas, a metilação é efectuada por duas metiltransferases (uma para cada cadeia), sendo em alguns sistemas metiladas bases diferentes em cada cadeia do DNA.

#### **6.1.4.2.4      *Montar uma reacção de restrição***

##### **6.1.4.2.4.1    *Estabilidade térmica***

As enzimas de restrição (como todas as enzimas) devem ser sujeitas a menor variação térmica possível. A temperatura a que as enzimas são habitualmente conservadas é -20°C, pelo que quando se transportam para a bancada, devem permanecer em gelo, ou idealmente num congelador de bancada, os quais mantêm uma temperatura de -20°C durante cerca de 2 horas (depende do fabricante). Devido a instabilidade das enzimas a temperatura ambiente, estas devem ser os últimos componentes da mistura de reacção a adicionar, para minimizar quer o choque entre a composição do tampão de conservação e a da mistura, quer o tempo de permanência a temperatura ambiente.

##### **6.1.4.2.4.2    *Concentração de glicerol***

Para aumentar o tempo de conservação, as enzimas de restrição são habitualmente conservadas em 50% de glicerol. No entanto, um excesso de glicerol na mistura de reacção (>5%) ocasiona um comportamento errático da enzima. Assim, o volume de enzima adicionado nunca pode exceder os 10% do volume total da reacção.

##### **6.1.4.2.4.3    *Tampão de reacção***

As diferentes enzimas tem actividades diferentes em determinados tampões. Assim, as companhias que as fornecem estudaram um conjunto de tampões concentrados (habitualmente 10X), os quais são optimizados para a actividade das varias enzimas. Estes tampões devem sempre que possível ser utilizados com as respectivas enzimas, pois a actividade de uma enzima num tampão diferente do sugerido pode ser quase nula. Se a experiência obrigar a utilização de varias enzimas no mesmo tubo de reacção, deve ser escolhido o tampão que apresentar o melhor compromisso entre a actividade das duas enzimas (os fornecedores fornecem habitualmente uma tabela com a %actividade de cada enzima em cada um dos tampões que fornecem). Finalmente, algumas enzimas necessitam da adição de componentes extra aos tampões padrão (ex. BSA). Também neste caso, soluções concentradas destes compostos são fornecidas com a enzima.

**Actividade das enzimas:** Por definição, 1 unidade de enzima de restrição digere completamente 1µg de DNA, num volume de 50µl, ao fim de 1 hora. No entanto esta actividade é apenas indicativa, já que tipos diferentes de DNA podem possuir conformações diferentes, e um número diferente de locais de restrição. Assim, utiliza-se de modo geral 2 a 3 vezes mais enzima, e entre 3 a 16 horas de incubação. O volume da reacção não deve ser inferior a 50µl, já que aumentam os erros de pipetagem, e a probabilidade de a concentração de glicerol ser superior a 5%.

Um factor crítico na boa execução de qualquer reacção enzimática e a homogeneidade da mistura. Deve-se homogeneizar a mistura de reacção por inversão e pipetagem repetida, mas nunca utilizar o vortex já que a violência deste pode desnaturar a enzima, deixando-a inactiva.

#### 6.1.4.2.4.4 *Temperatura de reacção*

A temperatura de incubação da maioria das endonucleases de restrição é de 37°C, mas algumas enzimas, isoladas de bactérias termofilicas necessitam de incubações entre 50-60°C (verifique a temperatura ideal para cada enzima, junto do fornecedor).

### 6.1.4.3 Métodos de marcação da sonda

#### 6.1.4.3.1 RADIOISÓTOPOS

##### 6.1.4.3.1.1 <sup>32</sup>P

O isótopo mais comumente utilizado para a marcação radioactiva de ácidos nucleicos é o <sup>32</sup>P. Este radioisótopo emite partículas β e tem uma actividade específica elevada (9200 Ci/mmol na sua forma pura) e um tempo de semi-vida relativamente curto (14 dias). Existem comercialmente disponíveis todas as espécies de trifosfatos de nucleótidos marcados com <sup>32</sup>P, e com variadíssimas actividades específicas. Note-se que o átomo radioactivo do dNTP, para a marcação de ácidos nucleicos deve ser o γ (os átomos α e β são libertados na formação da ligação com o nucleótido seguinte).

##### ➤ <sup>33</sup>P

O <sup>33</sup>P emite menos energia que o <sup>32</sup>P, mas possui idêntica capacidade de penetração. Devido à sua menor energia, este radionucleótido origina bandas mais bem definidas que o <sup>32</sup>P.

##### ➤ <sup>35</sup>S

O <sup>35</sup>S emite partículas com energia ainda mais baixa que a do <sup>33</sup>P (a sua actividade específica é de 1500 Ci/mmol na forma pura), tendo no entanto um tempo de semi-vida mais longo (87 dias). Os nucleótidos marcados com <sup>35</sup>S possuem um grupo tiol em substituição de um oxigénio no grupo fosfato, o que pode inibir a actividade de algumas enzimas. Por outro lado, uma vez que a sua energia é menor, este radioisótopo induz menos danos no DNA, pelo que as sondas com ele marcadas são mais estáveis. A menor energia deste nucleótido permite também obter bandas ainda mais bem definidas que as obtidas com <sup>33</sup>P, muito embora possa levar mais tempo a imprimir o filme fotográfico. É de salientar ainda o facto de este radioisótopo ser menos nocivo para o operador de laboratório, pese no entanto o facto de também ser mais difícil de detectar contaminações com o auxílio de um contador Geiger.

##### ➤ <sup>3</sup>H

O trítio é o radioisótopo de menor energia que é utilizado para a marcação de ácidos nucleicos. Com a sua actividade específica de apenas 29 Ci/mmol na forma pura, e um tempo de semi-vida de 12 anos, é o isótopo mais fraco de todos os procedimentos autorradiográficos. A sua baixa capacidade de penetração torna-o também o isótopo menos perigoso no laboratório, mas também mais difícil de detectar com contadores portáteis tipo Geiger.

### ➤ Outros radioisótopos

Ainda que tal aconteça com muito menos frequência, também é possível utilizar <sup>14</sup>C e <sup>125</sup>I para a marcação de ácidos nucleicos.



### 6.1.4.3.2 POLIMERASES DO DNA

Estão hoje em dia disponíveis uma vasta gama de polimerases do DNA, com propriedades e aplicações diferentes. Os principais factores a ter em conta na escolha da polimerase certa para cada tipo de trabalho são:

- **remoção de nucleótidos existentes:** da existência desta actividade pode depender a fidelidade do produto formado. Todas as polimerases cometem erros. Da sua capacidade de verificar o trabalho realizado, e remover os nucleótidos erroneamente incorporados depende a fidelidade do produto final. Obviamente, a existência desta actividade também resulta numa menor velocidade de reacção, o que pode dificultar a obtenção de produtos longos. Pode assim concluir-se que a opção pela existência ou não desta actividade na enzima escolhida deve ser realizada com base no resultado pretendido. Tipicamente, as reacções que se destinam a sequenciação, ou a clonagem devem ser sempre realizadas por enzimas com esta actividade.
- **Estabilidade térmica:** também esta característica pode ser benéfica ou prejudicial, dependendo do objectivo e protocolo específicos a utilizar. Por exemplo, numa reacção de PCR, a utilização de enzimas termoestáveis evita a destruição da enzima no passo de desnaturação. No entanto, se o protocolo envolver a posterior inactivação da enzima pelo calor, é necessário escolher uma enzima termossensível.

### 6.1.5 Dot-Blot

Uma variante à técnica do Southern Blot consiste no dot-blot. Trata-se também de uma técnica em que o DNA é covalentemente ligado a uma membrana, sendo posteriormente hibridizado com uma sonda. A diferença fundamental entre o dot-blot e o Southern-Blot

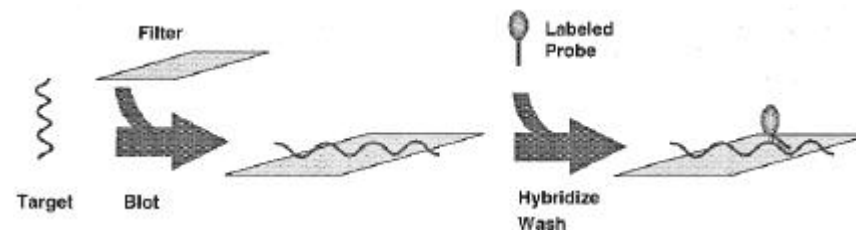


Figura 6.6 – Esquema simplificado de funcionamento do dot-blot / slot-blot

consiste na separação electroforética que ocorre no Southern, mas que não existe no dot-blot. Assim, o dot-blot não proporciona qualquer tipo de informação quanto ao peso molecular e numero de espécies de DNA responsáveis pelo sinal. Assim, este método é particularmente útil em experiências em que a discriminação qualitativa é menos importante que a quantitativa.

Uma variação ao dot-blot consiste no slot-blot. A única diferença entre estas duas técnicas consiste na forma que o sinal de hibridização origina. No dot-blot o DNA é aplicado numa fenda em círculo, originando um sinal circular. No slot-blot o DNA é aplicado numa fenda rectangular, em que uma das dimensões é muito reduzida (cerca de 1 mm), originando um sinal tipo banda. Assim, a quantificação do sinal originado pelo slot-blot é mais fácil, fiável e reprodutível.

## 6.2 Métodos Comerciais

A tecnologia de PCR encontra-se patenteada pela Roche, pelo que só esta empresa e suas subsidiárias se encontram autorizadas a desenvolver "kits" comerciais baseados nesta técnica.

Em resposta, várias outras companhias (Chiron, Abbot, Organnon Teknika) desenvolveram técnicas alternativas independentes da amplificação por PCR. Outras companhias (Sorin) optaram por

licenciar o PCR à Roche, desenvolvendo apenas técnicas de detecção próprias.

### 6.2.1 *Amplicor*

O Amplicor da Roche é uma técnica inteiramente dependente do PCR para a detecção e/ou quantificação de espécies de RNA ou DNA em amostras biológicas.

Assim, tudo o que foi dito na secção sobre PCR se aplica a este método comercial, o qual se encontra já disponível em versão semi-automática.

Deve ser realçado que este método inclui um controlo interno de amplificação, o qual consiste numa espécie de ácido nucleico introduzida na mistura de reacção, e que vai co-amplificar com o

assim produzido vai no entanto ser independentemente analisado, pois possui uma sequência interna diferente do DNA alvo, pelo que hibridiza com sondas diferentes.

O amplicor incorpora ainda um método de eliminação de contaminação proveniente de reacções de PCR anteriores. Com efeito, em vez de dATP, o kit utiliza dUTP, pelo que se produz uma molécula de DNA com uridina. Esta é eficientemente hidrolizada pela enzima Uracil-N-glicosilase (o nome comercial da Roche é Amperase), a qual é incluída na mistura de reacção. Esta enzima tem a particularidade de não funcionar a temperatura elevada, pelo que só se encontra activa antes de se iniciar a reacção (período durante o qual elimina possíveis contaminações). No final da reacção a Amperase é inactivada por meios químicos, assegurando a preservação do produto do PCR a ser analisado.

O sistema amplicor encontra-se disponível em formato manual ou semi-automático (Cobas-Amplicor), os quais diferem no método de detecção. No primeiro caso, o amplimer encontra-se biotilado (um dos primers é biotilado), o que lhe permite ligar-se a um poço de microplaca coberto com estreptavidina. No caso do formato semi-automático, a biotina é do amplimer liga-se a uma partícula magnética, o que permite ao aparelho reter os amplimers por meio de um campo magnético. Em ambos os casos a detecção é mediada por uma sonda, e a quantificação é realizada por leitura de absorvância.

### 6.2.2 "*Branched DNA*"

A sistema da Chiron (hoje Bayer) opta não por amplificar as moléculas de ácido nucleico a serem detectadas, mas por amplificar o sinal obtido de cada molécula presente na amostra. Para tal utiliza ácidos nucleicos sintéticos com derivações ("*branches*") em número preciso e bem determinado, o que através de hibridizações sucessivas produz o aumento de sinal necessário.

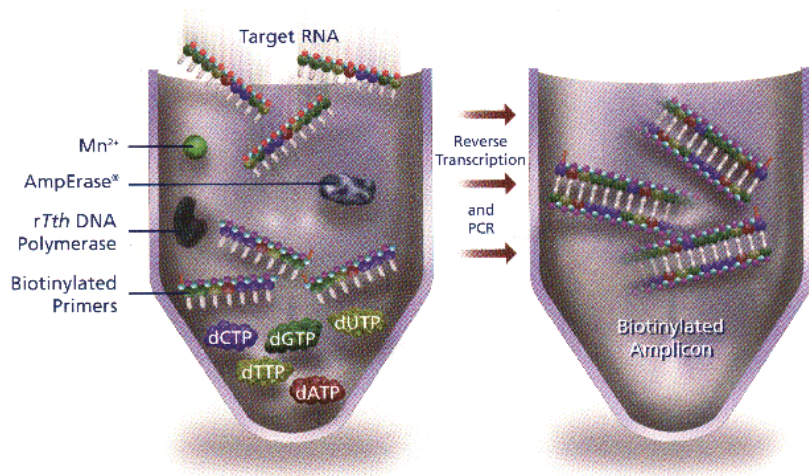


Figura 6.7 – A mistura de reacção da amplicor incorpora Amperase e dUTP para eliminação de contaminações.

ácido nucleico alvo, já que possui a mesma sequência terminal (o que permite aos primers hibridizarem). O DNA de controlo interno

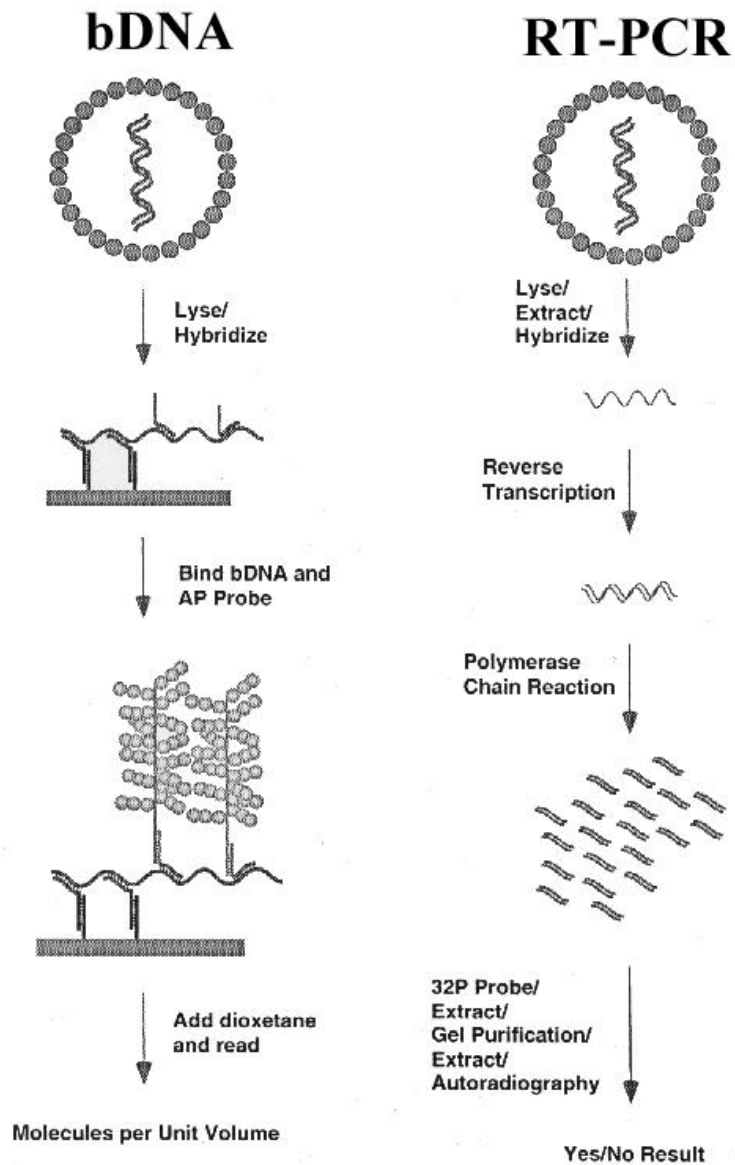


Figura 6.8 – Comparação entre o PCR e o bDNA

### 6.2.3 NASBA/Nuclisens

A Organon Teknika produziu um método em que com uma única temperatura, mas utilizando três enzimas diferentes se realiza à semelhança do PCR uma grande amplificação do número de moléculas de ácido nucleico em estudo. Para tal foi buscar a inspiração à própria célula, já que esta produz a partir de um único gene, milhares de cópias de transcritos.

Os métodos NASBA e o seu sucessor mais recente o NucliSens realizam assim uma série de reacções sucessivas (ver figura 6.9), que permitem produzir uma molécula de DNA a partir de uma molécula de RNA (utilizando uma RT), seguidamente a molécula de RNA inicial é degradada pela RNase H, seguindo-se a produção da cadeia complementar pela mesma RT que produziu a primeira cadeia de DNA. O facto de o primeiro primer ter incorporado uma sequência 5' não homóloga contendo o promotor da T7 RNA polimerase permite então a esta enzima produzir várias cópias de transcrito do gene alvo. Cada uma destas moléculas de transcrito pode então por sua vez ser replicada pela RT, amplificando-se o processo que passa a ser cíclico (fig. 6.9).

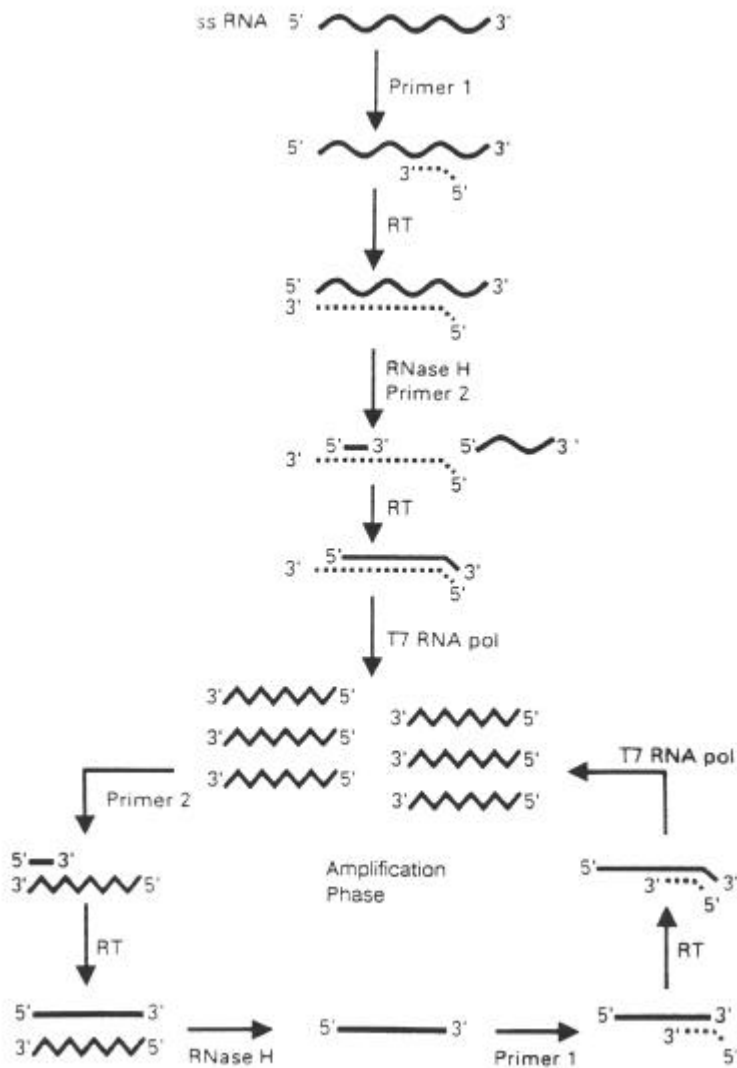


Figura 6.9 – esquema de funcionamento do NASBA/ NucliSens

### 6.2.4 Ligase Chain Reaction

A técnica desenvolvida pela Abbot é em tudo semelhante ao PCR, mas não se limita a utilizar uma polimerase para produzir a amplificação da sequência alvo. Nesta técnica são utilizados primers muito longos que cobrem a quase totalidade da sequência a estudar. A parte não coberta pelos primers é uma zona central, a qual após a hibridização dos primers é sintetizada por uma polimerase, e finalmente os dois fragmentos unidos por uma ligase. Assim, a existência do gene pode ser detectada pela detecção da união dos dois primers, a qual é dependente da hibridização, e consequentemente da existência do gene alvo (fig.6.10).

A detecção foi também desenvolvida pela Abbot e denomina-se MEIA (microparticle immuno sorbent assay). Nesta técnica, micropartículas cobertas com um anticorpo reactivo contra um hapteno existente no primer A são incubadas com os "amplimers" (fig. 6.10b). Se a reacção de ligase ocorreu, esta incubação "prende" não só o primer A, mas também o primer B, o qual está conjugado com um outro hapteno identificável por um anticorpo conjugado com uma enzima que decompõe um substrato, a qual decompondo um composto e emitindo fluorescência.

Tanto o LCR como o MEIA se encontram automatizados.

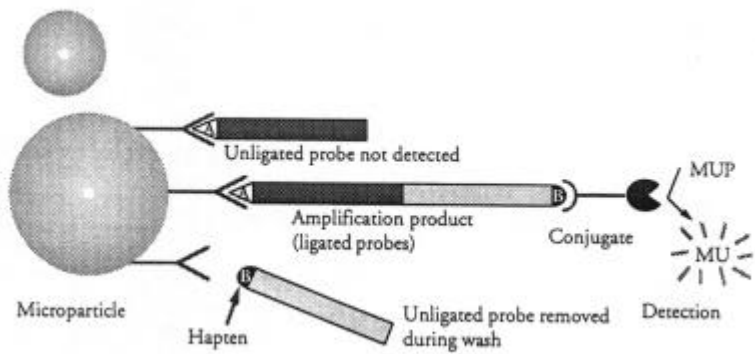
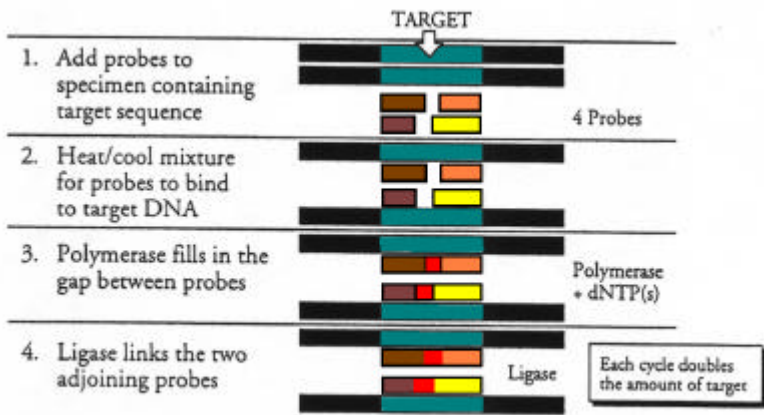


Figura 6.10 – Esquema de funcionamento da Ligase Chain Reaction (LCR) e do Microparticle Enzyme Immuno-sorbent assay (MEIA).

### 6.2.5 DEIA

O sistema DEIA da Sorin é um sistema genérico de detecção e quantificação da existência de moléculas de DNA complementares a uma sonda. Para o seu desenvolvimento, esta empresa isolou um anticorpo de rato reativo contra DNA de cadeia dupla, mas não reativo contra DNA monocatenário. Desta forma, foi possível criar um sistema em tudo idêntico a um ELISA convencional, em que o antígeno a ser estudado é a existência de DNA de cadeia dupla, ou seja DNA que hibridizou com uma sonda específica.

O desenho do sistema (figura 6.11) parte da utilização de uma placa revestida com estreptavidina, à qual é ligado uma sonda conjugada com biotina. É então efectuada uma reacção de hibridização na própria placa, que a ser eficaz produz moléculas de DNA de cadeia dupla, originando a presença do antígeno reconhecido pelo anticorpo a utilizar em seguida. Todo o resto do processo é sobreponível a um vulgar ELISA.

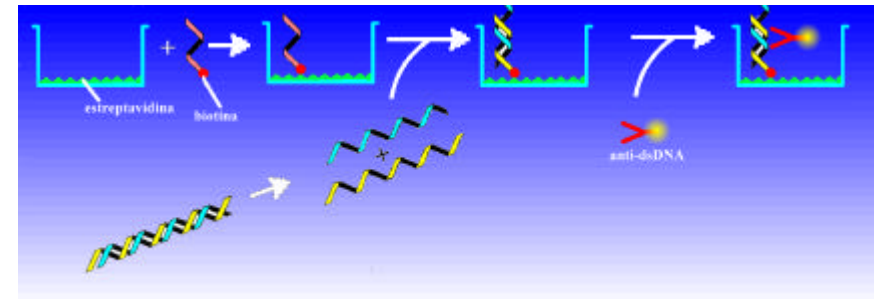


Figura 6.11 – Esquema de funcionamento do DNA Enzyme Immuno-sorbent assay (DEIA)

# Capítulo 7

## 7 Índice Remissivo

### 2

2-desoxirriboses, 2.2

### A

ácidos nucleicos. *Consulte* DNA e RNA

adenil ciclase, 4.5

Adeninas, 2.2

alelos, 2.18, 4.8, 4.45

aminoácido, 2.11, 2.21, 2.22, 2.24, 2.29, 2.30, 4.5, 4.36, 4.37, 4.45, 4.48, 4.50

anemia não-esferocítica congênita, 4.36

anemias hemolíticas agudas, 4.36

apoptose, 2.46, 2.47, 4.8, 4.9, 4.34

### B

bases azotadas, 2.3

BCL-2, 4.35

### C

cAMP, 4.5

carcinogénios, 4.9, 4.10, 2

cdk, 2.45, 4.7

células citotóxicas, 4.33

células tumorais, 4.2, 1

ciclinas, 2.40, 2.42, 2.44, 2.45, 4.7, 1

Citosinas, 2.2

complemento, 2.17, 4.33

Crk, 4.6

### D

DNA, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.11, 2.12, 2.13, 2.14, 2.15, 2.16, 2.18, 2.19, 2.21, 2.25, 2.26, 2.27, 2.28, 2.32, 2.33, 2.34, 2.39, 2.42, 2.43, 2.46, 2.4, 2.5, 2.6, 2.9, 2.12, 2.13, 2.14, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.9, 3.10, 3.12, 3.13, 3.24, 4.7, 4.9, 4.10, 4.41, 4.50, 5-2, 5-4, 5-5, 5-6, 5-7, 5-8, 5-9, 5-10, 5-11, 5-12

### E

enhancers, 2.18, 2.28, 3.7, 3.26, 4.7, 4.13, 4.31

### F

factores de crescimento, 2.45, 2.50, 4.2, 4.4, 4.5

Factores de crescimento, 4.4

Factores de Crescimento

PDGF, 4.7

Factores de Transcrição

AP-1, 4.7

factores de transcrição nuclear, 4.4, 4.7

Factores de transcrição nuclear, 4.7

fosforilação, 2.40, 2.42, 2.43, 2.44, 2.45, 2.48, 2.50, 4.5, 4.6

### G

glicose-6-fosfato desidrogenase, 4.36

GM-CSF, 4.5

GTP, 2.29, 2.30, 2.50, 3.26, 4.5, 4.6

Guaninas, 2.2

## H

HEMATOLOGIA, ii  
HLA, 2.11, 2.17, 2.19, 4.33, 4.34, 4.53

## I

Imortalização, 4.2  
imunoglobulinas  
  Igλ, 4.31  
imunoglobulinas, 2.2, 2.9, 4.31  
  IgH, 4.31, 4.31, 4.34, 4.35, 3  
Imunoglobulinas  
  IgK, 4.31  
IMUNOLOGIA, ii

## L

Linfoma de Burkitt, 4.31, 4.32, 4.33  
Linfoma de Hodgkin, 4.33  
Linfomas Difusos, 4.34  
Linfomas foliculares, 4.34  
Linfomas T, 4.33

## M

mbr, 4.34, 4.35  
mcr, 4.34, 4.35  
Metástase, 4.2  
MICROBIOLOGIA, ii  
mRNA, 2.10, 2.11, 2.15, 2.19, 2.25,  
  2.27, 2.28, 2.29, 2.30, 2.31, 2.32,  
  2.14, 3.3, 3.5, 3.6, 3.7, 3.10, 3.11,  
  3.12, 3.13, 3.25, 4.7, 4.29, 4.31,  
  4.34, 4.43, 4.54  
mutações, 2.17, 2.19, 2.22, 2.23, 2.24,  
  2.34, 2.38, 2.46, 3.25, 3.27, 4.3, 4.7,  
  4.8, 4.13, 4.14, 4.15, 4.16, 4.28,  
  4.35, 4.37, 4.38, 4.40, 4.41, 4.43,  
  4.44, 4.45, 4.46, 4.49, 4.50, 4.51,  
  4.52, 2  
mutações oncogénicas, 4.9  
mutagénios, 4.9, 2

## N

nonsense  
  Mutações nonsense, 2.24, 4.3, 4.13,  
  4.38  
nucleótido, 2.2, 2.4, 2.23, 2.26, 4.48, 5-  
  4, 5-11  
nucleótidos, 2.2, 2.3, 2.4, 2.14, 2.19,  
  2.21, 2.26, 2.27, 2.30, 2.31, 2.32,  
  2.34, 2.5, 2.8, 3.22, 5-3, 5-7, 5-11, 5-  
  12

## O

oncogenes, 2.13, 3.30, 4.3, 4.4, 4.5, 4.5,  
  4.6, 4.7, 4.10, 4.13, 2  
  bcl-2, 2.47, 4.9  
  p53, 2.46, 4.7, 4.8, 4.9, 4.12, 4.28,  
  4.34, 2  
Oncogenes  
  Abl, 4.5  
  BCL-2, 4.34  
  myc, 4.31, 4.31, 4.32  
  ras, 4.6  
  RB, 2.45, 2.46, 4.7, 4.8  
  Src, 4.5, 4.6

## P

PDGF, 4.4, 4.12  
pentose, 2.2, 2.4  
pentoses, 2.2, 2.3  
Piruvato Quinase, 4.37  
pontuais  
  Mutações pontuais, 2.23, 2.24, 2.7,  
  3.27, 4.3, 4.40, 4.43, 4.44, 4.50,  
  4.51, 2  
promotores, 2.15, 2.16, 2.17, 4.7  
proteína G, 2.50  
proteínas de controlo do ciclo celular,  
  4.4  
Proteínas de controlo do ciclo celular,  
  4.7  
proteínas G, 2.50, 4.5  
proto-oncogenes  
  c-Fos, 4.7  
  c-Myc, 4.7  
Proto-oncogenes, 4.3

## R

receptores de factores de crescimento,  
  4.4

retrovírus, 2.18, 2.25, 2.28, 3.2, 3.5,  
  3.13, 3.14, 3.15, 3.22, 3.24, 3.29,  
  3.30, 4.10  
riboses, 2.2  
RNA, 2.2, 2.3, 2.11, 2.13, 2.14, 2.15,  
  2.16, 2.17, 2.19, 2.20, 2.25, 2.26,  
  2.27, 2.28, 2.34, 2.46, 3.2, 3.3, 3.4,  
  3.5, 3.6, 3.10, 3.11, 3.12, 3.13, 3.22,  
  3.24, 4.10, 5-2, 5-6, 5-7, 5-13

## S

SH2, 4.6  
SH3, 4.6  
*src*, 4.3, 4.6, 4.12  
  c-src, 4.4, 4.6

## T

TGF-α, 4.5  
Timidinas, 2.2, 2.25, 2.33  
tirosina cinase, 2.50, 4.5, 4.6, 4.12  
transcrição, 2.10, 2.11, 2.15, 2.16, 2.17,  
  2.18, 2.19, 2.25, 2.26, 2.27, 2.28,  
  2.36, 2.49, 2.50, 3.3, 3.6, 3.7, 3.8,  
  3.9, 3.10, 3.12, 3.13, 3.14, 3.15,  
  3.23, 3.25, 3.26, 4.4, 4.7, 4.12, 4.13,  
  4.33, 5-6, 1  
transdutores intracelulares de sinal, 4.4  
Transdutores intracelulares de sinal, 4.5  
Transformação, 4.2  
translocação, 2.13, 2.31, 2.47, 4.19,  
  4.26, 4.29, 4.30, 4.31, 4.31, 4.32,  
  4.34, 4.35  
translocações, 4.3, 4.27, 4.28, 4.31,  
  4.32, 4.34  
tumores, 2.9, 2.19, 3.23, 3.28, 3.29, 4.3,  
  4.5, 4.6, 4.8, 4.9, 4.10, 4.14, 4.26,  
  4.33, 4.34, 1, 2

## U

Uracilos, 2.2, 2.25