

J. MARTINS E SILVA

# **Bioquímica da Informação Genética**



*A presente impressão foi possibilitada  
com a participação de  
Angelini Farmacêutica, Lda*



# **Bioquímica da Informação Genética**

Reedição dos textos de apoio às aulas da disciplina de Química Fisiológica (ano lectivo de 1972-1973), na Faculdade de Medicina da Universidade de Lourenço Marques, Moçambique.

No prefácio são referidos os principais acontecimentos que ocorreram no âmbito da genética molecular, até ao presente.

Lisboa, 2006

### **Título**

*“Bioquímica da Informação Genética”*

Texto de apoio à disciplina de Química Fisiológica, no ano lectivo de 1972-1973  
(Faculdade de Medicina da Universidade de Lourenço Marques, Moçambique, 1973)  
Reedição – 2006, acrescida no prefácio pela referência cronológica aos principais desenvolvimentos científicos sobre identificação do genoma e principais precursores utilizados.

### **Editor e Proprietário**

J. Martins e Silva

### **Endereço**

Instituto de Biopatologia Química  
Faculdade de Medicina da Universidade de Lisboa  
Avenida Professor Egas Moniz  
1640-028 Lisboa  
Telef. (351 217 985 136)  
Fax: (351 217 999 477)  
<http://www.fm.ul.pt>

### **Redacção**

Emília Alves

### **Tiragem**

250 Exemplares

### **Composição Montagem e Impressão**

Publicações Ciência e Vida, Lda  
Apartado 44 – 2676-901 Odivelas  
Telef. 21 478 13 12 Fax: 21 478 13 16  
E-mail: [pubcienciaevida@sapo.pt](mailto:pubcienciaevida@sapo.pt)  
[pub@cienciaevida.com](mailto:pub@cienciaevida.com)  
ISBN 972-590-082-0  
Depósito Legal n.º ?????

## ÍNDICE

<b>PREFÁCIO DA REEDIÇÃO</b> .....	11
Sobre as origens do texto .....	11
Evolução e consolidação da genética molecular .....	12
Era genómica .....	20
Localização de genes e suas mutações .....	26
Sequenciação genómica de microorganismos e parasitas .....	27
Perspectivas futuras e conclusões .....	39
Notas .....	41
<b>CAPÍTULO 1</b>	
<b>CROMOSSOMAS, GENES E HEREDITARIEDADE</b> .....	55
Conservação e transmissão da informação genética .....	55
Cromossomas .....	57
Ciclo celular e tipos de divisão celular .....	59
Mitose .....	60
Meiose .....	63
Conclusões .....	67
<b>CAPÍTULO 2</b>	
<b>NUCLEÓTIDOS E ÁCIDOS NUCLEICOS</b> .....	69
Estrutura química .....	69
Conclusões .....	75

**CAPÍTULO 3****CONFORMAÇÃO, REPLICAÇÃO, REPARAÇÃO E RECOMBINAÇÃO**

<b>DO ÁCIDO DESOXIRRIBONUCLEICO</b> .....	77
Estrutura molecular do duplex .....	77
Replicação do DNA .....	82
Reparação do DNA .....	98
Mutações .....	100
Recombinação cromossômica .....	105
Conclusões .....	109

**CAPÍTULO 4****ÁCIDO RIBONUCLEICO: ESTRUTURA, SÍNTESE E TRANSCRIÇÃO REVERSA** ..... 111 |

Estrutura .....	111
Transcrição do DNA .....	112
Regulação da transcrição do DNA .....	118
Transcrição reversa .....	119
Conclusões .....	124

**CAPÍTULO 5****TRADUÇÃO DA INFORMAÇÃO GENÉTICA** ..... 125 |

Significado, localização e componentes .....	125
Fases da tradução .....	131
Síntese polipeptídica em polirribosomas .....	139
Conclusões .....	140

**CAPÍTULO 6****CÓDIGO GENÉTICO** ..... 143 |

Definição do código .....	143
Iniciação e terminação de leitura do código genético .....	145
Degenerescência do código e mutações .....	146
Conclusões .....	149

**CAPÍTULO 7**

<b>MECANISMOS DE REGULAÇÃO DA SÍNTESE PROTEICA</b> .....	151
Regulação nos procariotas .....	151
Controlo em sistemas repressíveis .....	158
Regulação nos eucariotas .....	162
Conclusões .....	164

**CAPÍTULO 8**

<b>ESPECULAÇÕES SOBRE A ORIGEM DA VIDA</b> .....	165
Teoria criacionista .....	165
Teoria organicista .....	166
Teoria mecanicista .....	166
Vitalismo e panspermia .....	168
Teoria cosmozoica .....	170
Evolução química e biológica .....	170
Teoria materialista .....	173
A primeira “matéria” viva .....	178
Conclusões .....	181

**ANEXO I**

<b>CRONOLOGIA DAS PRINCIPAIS ETAPAS E INVESTIGADORES COM INTERVENÇÃO NO DESENVOLVIMENTO DA GENÉTICA E DA BIOLOGIA MOLECULAR, ATÉ 1973</b> .....	183
Era pré-genética .....	183
Era genética .....	184
Era do DNA .....	187
Era genómica .....	191

**ANEXO II**

<b>GLOSSÁRIO ESSENCIAL</b> .....	193
----------------------------------	-----

<b>BIBLIOGRAFIA</b> .....	203
---------------------------	-----



# PREFÁCIO DA REEDIÇÃO

## Sobre as Origens do Texto

O presente volume reedita as aulas de genética molecular que constituíram parte do programa da disciplina de Química Fisiológica que regi na Faculdade de Medicina da Universidade de Lourenço Marques, no ano lectivo de 1972 a 1973. Aquelas aulas foram reeditadas recentemente<sup>1</sup>.

A importância da descoberta da estrutura do ácido desoxirribonucleico, a progressão quase geométrica dos conhecimentos sobre o fluxo e regulação da informação genética a que dera origem, e as perspectivas de aplicação extensivas à Medicina que já então se auguravam justificaram que então lhes desse particular destaque no programa do curso e consolidasse o seu ensino, sob a forma de um texto policopiado. Circunstâncias da vida pessoal (designadamente, a mobilização militar para uma área de intervenção no norte de Moçambique) não permitiram que aqueles apontamentos contivessem os esquemas definitivos nem algumas das figuras projectadas nas aulas. Decidi, portanto, proceder à sua reedição em livro, com os elementos em falta, aproveitando a oportunidade para evocar neste Prefácio numa base cronológica as principais fases do progresso mais relevantes desde então ocorridas no âmbito da Genética Molecular.

A actual reedição mantém a disposição inicial do texto, organizado em oito capítulos: 1. *Cromossomas, Genes e Hereditariedade*; 2. *Nucleótidos e Ácidos Nucleicos*; 3. *Conformação, Replicação, Reparação e Recombinação do Ácido Desoxirribonucleico*; 4. *Ácido Ribonucleico: Estrutura, Síntese e Transcrição Reversa*; 5. *Tradução da Informação Genética*; 6. *Código Genético*; 7. *Mecanismos de Regulação da Síntese Proteica*; 8. *Especulações sobre a Origem da Vida*.

Os esquemas desenhados à mão na edição original foram refeitos, alguns dos quais pelo nosso antigo e dedicado técnico, Chim W. San.

As fotografias que completam a série de figuras foram generosamente oferecidas pelos seus autores, ou obtiveram a sua permissão para serem utilizadas na impressão, pelo que reitero o meu reconhecimento aos Drs. Alexander Rich, Arthur Kornberg, Albrecht K. Kleinschmidt, J. Herbert Taylor, Jack D. Griffith, M. Nascimento Ferreira, Sydney Brenner e R.W. Horne.

À nossa funcionária de longa data, Senhora D. Emília Alves, que num trabalho exaustivo e metucioso reorganizou a parte dactilográfica e esquemas restantes para a presente reedição, expresso os maiores agradecimentos pelo trabalho havido. Fraternal agradecimento a António Duarte pelo apoio em algumas das figuras mais complexas.

## **Evolução e Consolidação da Genética Molecular**

Nos trinta anos passados a investigação prosseguiu entusiasticamente e expandiu-se em múltiplas direcções temáticas. Os conceitos clarificaram-se, o conhecimento sobre algumas etapas e mecanismos foi esclarecido, a nomenclatura ficou mais rica, as perspectivas ampliaram-se a níveis originalmente insuspeitáveis (designadamente no diagnóstico e na terapêutica), as aplicações tecnológicas com base na análise da sequência dos ácidos nucleicos possibilitaram avanços essenciais no despiste dos polimorfismos humanos e na instituição de terapêuticas génicas<sup>2</sup>. Nesta ordem de progresso é legítimo especular sobre o futuro, em que a clínica médica será profunda e decisivamente modificada pela utilização da genómica<sup>3</sup>, ainda que esse dia nos pareça ainda longínquo e também incerto<sup>4</sup>.

Adicionalmente, a confirmação de que o “dogma central” do código genético inclui variantes que justificam a diversidade dos organismos, explicam as diferentes funções codificadas por tecido distintos a partir de um conjunto de genes comuns, e o facto de um mesmo gene poder determinar a síntese de múltiplas proteínas com actividades também diferentes, entre outras situações, não contradiz a essência da genética molecular na sua simplicidade estrutural<sup>5</sup>.

**Do DNA duplex ao código genético** – Constitui um facto assente e amplamente confirmado que a informação genética está contida na disposição se-

quencial das bases nucleotídicas de cada molécula do ácido desoxirribonucleico (DNA).

A síntese de proteínas com as mais diversas características estruturais e funcionais culmina num conjunto de etapas, através das quais a informação genética em cada duplex de DNA é, através da complementariedade das suas bases, replicada, transcrita e, por fim, traduzida<sup>6</sup>. Poderá afirmar-se que o modelo em dupla hélice para a estrutura molecular do ácido desoxirribonucleico resistiu à prova de tempo, justificando que o seu cinquentenário fosse comemorado com grande júbilo, em 2003 por todo o mundo, científico e cultural.

Aquela importante contribuição, tradicionalmente atribuída a James Watson e Francis Crick, foi publicada no número de 25 de Abril de 1953 da *Nature*, em artigo de uma página, precedendo outras duas contribuições de diferentes autores. De comum, os três artigos tiveram o mesmo título (*molecular structure of nucleic acids*), embora com subtítulos diferentes<sup>7</sup>. Tão peculiar coincidência pretendia indicar, com elegância, que o assunto havia sido estudado e resolvido por três grupos de trabalho embora, na prática, o segundo e terceiro artigos dessem a precedência ao de Watson e Crick. Sabe-se que a inclusão dos três artigos no mesmo número da *Nature* resultou de um acordo do editor do jornal com os principais responsáveis pelos dois grupos de trabalho envolvidos, Sir Lawrence Bragg (director do Laboratório Cavendish, de Cambridge, onde trabalhava Francis Crick e estagiou o norte americano James Watson) e John Randall (director do Departamento Wheatston de Física, do King's College, em Londres). Por detrás daquela decisão havia um passado de polémica e algumas *desvirtudes* em que a natureza humana é tão pródiga, nomeadamente a inveja e a deslealdade.

É um facto que o artigo de Watson e Crick atingiu a notoriedade, enquanto dos outros quase ninguém recorda. Wilkins ainda partilhou com aqueles seus contemporâneos o Prémio Nobel em Fisiologia e Medicina em 1962. Rosalind Franklin, por ter falecido entretanto aos trinta e oito anos de idade, não beneficiou de idêntico reconhecimento. Porém, é indiscutível que foram os contributos de Rosalind Franklin<sup>8</sup>, através da difracção com raios X, que esclareceram o arranjo estrutural do DNA e a disposição das bases púricas e pirimídicas no interior da dupla hélice, que possibilitaram a Watson e Crick avançarem com a hipótese publicada em 1953. É de notar que nenhum destes dois últimos autores executou trabalhos experimentais conducentes à sua conclusão. O seu contributo foi fundamentalmente conceptual. Para tal, mantiveram-se bem infor-

mados sobre os resultados e progressos registados por outros investigadores, utilizando para a formulação das suas hipóteses um gabinete cedido por Lawrence Bragg, equipado com um quadro e onde construía os esquemas tridimensionais (em arame e bolas de plástico e, por fim, com recortes em cartão das bases de purina e pirimidinas). Deste trabalho persistente resultou o modelo de DNA que ficou para a história, desenhado no artigo da *Nature* pela mulher de Crick, Odile.

Ressalvando a questionável metodologia que utilizaram para os seus objetivos, ficou a dever-se a Watson e Crick a *resolução* genial do enigma da informação hereditária e do mecanismo codificador das proteínas celulares que lhe está dependente.

Criado o modelo de DNA, o conceito ficou aberto para Crick propor, em 1955, junto de 20 colegas biólogos moleculares do “RNA Tie Club”, a hipótese da existência de uma molécula intermediária (que designou por “adaptador”) que reconheceria e localizaria cada aminoácido na sequência proteica, de acordo com a sequência de um código existente na molécula de RNA mensageiro, transcrita de uma das faixas do DNA codificador. Aquela proposta viria a ser publicada somente três anos depois<sup>9</sup>.

A equipa liderada por Paul Zamecnick, na sequência da descoberta de que os aminoácidos utilizados na síntese proteica eram pré-activados a nível do seu grupo carboxilo<sup>10</sup>, reforçou a hipótese de Crick e Cols<sup>9</sup> de uma espécie solúvel de RNA (RNA<sub>s</sub>) a que se uniriam aqueles aminoácidos<sup>11</sup>. Posteriormente, ao ser-lhe reconhecida funções de transferência de aminoácidos, aquela molécula passou a ter a designação de RNA<sub>t</sub> (*de transferência*).

Dois anos depois, Brenner e Cols<sup>12</sup> revelaram que a informação genética era transportada do DNA para os ribossomas por uma molécula instável de RNA, que designaram de RNA mensageiro (RNA<sub>m</sub>). Em 1965 a equipa de Brenner viria a caracterizar no DNA os codões de terminação da síntese proteica<sup>13</sup>.

Crick e Cols<sup>14</sup> conceptualizaram a natureza geral do código genético, a ser “escrito” em “palavras” (*codões*) constituídas por três das quatro bases diferentes do DNA, prevendo um total de 64 possíveis codões. As moléculas de RNA<sub>t</sub> interviriam na descodificação do DNA.

Hoagland, no artigo com que se associou às celebrações do cinquentenário do duplex do DNA, expressou bem a epopeia entusiástica que foi a dos primórdios da descodificação dos mecanismos que determinaram a síntese proteica<sup>15</sup>.

Ainda em 1961, Mary Lyon propôs que apenas um dos cromossomas X femininos seria aleatoriamente inativado<sup>16</sup>. Este fenómeno, que consagrou no termo “lionização” o nome daquela investigadora, explica que homens e mulheres possuem conteúdos equivalentes de DNA não obstante, em termos funcionais terem número “diferente” de cromossomas.

A definição do código genético, iniciada por Nirenberg e Matthaei em 1961<sup>17</sup>, foi essencial para o esclarecimento da síntese proteica. O código foi decifrado em cerca de seis anos, passando por duas fases experimentais. Na primeira, a composição das bases dos codões e a natureza de cada codão foram estudados com polímeros de nucleótidos, utilizando diferentes combinações de bases num sistema de síntese proteica acelular. Na segunda fase foi definida a sequência de bases em cada codão, de modo que a informação genética dos ácidos nucleicos tivesse uma correspondência específica com o ordenamento dos aminoácidos em cada tipo de cadeias polipeptídicas. Naquele estudo foi evidenciado que o RNA actuava como molde para a síntese polipeptídica, em que, por exemplo, aos resíduos de uracilo numa cadeia sintética poli U correspondia o aminoácido fenilalanina nas proteínas. À data era ainda controversa a função do RNA<sub>i</sub>, por não ter sido ainda demonstrada a sua participação na síntese proteica. A colinearidade entre a sequência dos genes no DNA e a estrutura proteica foi confirmada através de mutações genéticas, por Yanofsky e Cols<sup>18</sup>.

Khorana e Cols<sup>19</sup> deram relevante contribuição à clarificação do código genético, através de uma sequência de experiências com polinucleótidos, utilizando a técnica de Leder e Cols<sup>20</sup> para determinarem a ligação dos compostos aminoacil-RNA<sub>i</sub> aos ribossomas, em função dos codões apresentados. Pelos resultados da síntese polipeptídica foram sucessivamente deduzidas as sequências codificadoras dos aminoácidos essenciais, o que confirmou, na generalidade, as anteriores observações do grupo de Nirenberg<sup>16</sup> e a degenerescência do código, incidente na terceira posição do codão<sup>21</sup>.

Holley e Cols<sup>22, 23</sup> demonstraram, através da identificação na *E. coli* de duas moléculas de RNA<sub>i</sub> para a leucina, que a degenerescência do código seria atribuível a diferenças na sequência de nucleótidos que constituíam o anticodão nas moléculas que intervinham como “adaptadores ao código”<sup>22</sup>. O anticodão estaria localizado a meio da molécula de RNA<sub>i</sub>, cuja estrutura secundária teria a forma de uma “folha de trevo”<sup>24</sup>. No seguimento da purificação de compostos aminoacil-RNA<sub>i</sub><sup>23</sup> para a alanina, histidina, tirosina e valina na levedura<sup>25</sup>, foi definida a composição do RNA<sub>i</sub> para a alanina, a qual foi a primeira sequência

nucleotídica de um ácido nucleico a ser conhecida. Indirectamente, aquela sequência identificaria também a primeira sequência conhecida de um gene, determinante para a transferência de alanina nas células de levedura.

Em 1968, Marshall W. Nirenberg, Robert W. Holley e Har G. Khorana foram galardoados com o Prémio Nobel em Medicina, “pela forma como interpretaram o código genético e a sua função na síntese proteica”. Na conferência que proferiu na cerimónia onde lhe foi atribuído aquele Prémio, Nirenberg concluiu (a propósito do código genético) que “os ácidos nucleicos actuam quer como molde para outras moléculas quer como relógio biológico”<sup>26</sup>.

No ano anterior May Ptashne<sup>27</sup> e Walter Gilbert<sup>28</sup> haviam identificado e isolado moléculas proteicas com actividade reguladora da expressão genética; aquelas moléculas, ao unirem-se a sequências nucleotídicas próximas de genes estruturais, reprimiam a respectiva expressão.

**A recombinação de DNA e o início da engenharia genética** – Nos estudos que realizaram entre 1968 e 1970, em culturas de *Haemophilus influenzae*, Hamilton Smith e Cols<sup>29,30</sup> isolaram (por acaso) a primeira das *endonucleases de restrição*. Esta enzima (rapidamente purificada), que evidenciava a propriedade de segmentar o duplex em sítios com determinada sequência, foi também objecto de estudo por dois outros grupos de investigadores, liderados por Daniel Nathans<sup>31</sup> e Werner Arber<sup>32,33</sup>. Em 1978, Smith, Arber e Nathans foram galardoados com o prémio Nobel em Medicina pelo contributo que deram à “descoberta das enzimas de restrição e à sua aplicação aos problemas da genética molecular”.

Utilizando um conjunto de endonucleases de restrição, cada uma com especificidade para determinado sítio da sequência do duplex, passou a ser possível a obtenção de fragmentos de DNA com dimensões conhecidas, a serem utilizados nos diversos tipos de estudos, designadamente: sequenciação de DNA, isolamento de genes, construção do DNA recombinante e análise cromossómica. A tecnologia com endonucleares alterou profundamente os estudos subsequentes em biologia molecular e deu início a um novo campo de actuação, que seria a engenharia genética. Ainda em 1971, Ray Wu e Ellen Taylor<sup>34</sup> deduziram pela primeira vez uma sequência de DNA, representadas pelas 12 bases terminais do genoma do fago lambda.

Além das capacidades referidas, as enzimas de restrição seriam relevantes para as trocas genéticas *in vivo*, inibindo-as ou promovendo-as<sup>33</sup>. Adicionalmente, a transposição de segmentos de DNA seria atribuível a sequências de

inserção (*insertion sequences*, IS) do fago  $\lambda$ , incluídas no cromossoma da *E. coli* ou de outras bactérias hospedeiras. Os segmentos de DNA delimitados por sequências IS (denominadas *transposões*) poderiam ser translocados por um vector para o genoma de outro organismo, além de participarem em outros mecanismos de recombinação genética<sup>33,35</sup>.

A endonuclease isolada da *H. influenzae* e as enzimas de restrição isoladas das estirpes B, K e KCP1 da *E. coli* também clivavam o DNA do vírus SV40<sup>31</sup>. Destes estudos foi elaborado um “mapa de clivagem” para cada enzima de restrição<sup>36</sup>, o qual passou a ser uma referência para o mesmo tipo de trabalhos em outros laboratórios.

Sensivelmente na mesma época, Weiss e Richardson<sup>37</sup> identificaram uma outra enzima, a DNA ligase, cuja função consistia em catalisar a junção de fragmentos separados de DNA. A utilização da DNA ligase possibilitou a síntese da primeira molécula de DNA recombinante pela equipa de Berg, em 1972<sup>38</sup>.

Entretanto, os grupos de Howard Temin<sup>39</sup> e David Baltimore<sup>40</sup> haviam demonstrado, em simultâneo, que alguns dos vírus RNA (designados por retrovírus) tinham uma enzima (*transcritase reversa*) com a capacidade de copiar o seu RNA numa cadeia de DNA que, por replicação, originava a formação de um duplex DNA (recombinante). Com esta descoberta, não só o dogma central da biologia (DNA  $\rightarrow$  RNA  $\rightarrow$  proteína) passou a ter uma alternativa válida como, também, contribuiu para reforçar um novo campo de desenvolvimento tecnológico.

A transcritase reversa, tal como havia sucedido com as endonucleases de restrição, situa-se na origem da engenharia genética. Conjugando as duas metodologias, cópias de DNA (conhecidas por *DNA recombinante* ou *cDNA*) passaram a ser facilmente produzidas a partir de fragmentos identificados de DNA ou de RNA.

Em 1975, David Baltimore, Howard Martin Temin e Renato Dulbecco foram galardoados com o Prémio Nobel em Medicina pela “descoberta sobre a interacção entre vírus tumorais e o material genético da célula”.

A construção de moléculas de DNA por recombinação foi iniciada pelo método bioquímico que havia sido desenvolvido pelo grupo de Berg<sup>38</sup>. Nos estudos que conduziram aquela descoberta foram obtidas moléculas de DNA do vírus SV40, previamente integradas em segmentos com os três genes do metabolismo da galactose da *E. coli*. O objectivo expresso daqueles trabalhos era o desenvolvimento de um método que possibilitasse a inserção de segmen-

tos definidos de DNA (contendo nova informação) em células de mamíferos. Havia sido demonstrada a facilidade com que o SV40 provocava alterações transmissíveis no crescimento de células humanas em cultura, conferindo-lhes características tumorais<sup>41</sup>. O projecto apoiava-se na demonstrada associação (covalente, estável e transmissível) que o genoma viral adquiria junto dos genomas de outras células de mamíferos<sup>42,43</sup>.

O entusiasmo inicialmente suscitado pela técnica de recombinação do DNA junto de diversos cientistas foi seguido de perto por recomendações de prudência por parte de outros investigadores, que alertavam para o perigo potencial de serem criados novos tipos de moléculas de DNA com propriedades infectantes ou outras, de consequências imprevisíveis.

Daquele movimento resultou a proposta para uma moratória voluntária ao desenvolvimento de algumas experiências de recombinação do DNA<sup>44</sup>, a qual culminou na “Conferência Asilomar”<sup>45</sup>, realizada em 1975 no Asilomar Conference Center (Califórnia, USA) em que foram aprovadas, com exemplar consenso, recomendações fundamentais para o prosseguimento de investigações com recombinações do DNA<sup>46</sup>.

Em 1980, Paul Berg partilhou com Walter Gilbert e Frederick Sanger, o Prémio Nobel em Química, “pelos estudos fundamentais na bioquímica dos ácidos nucleicos, em particular sobre o DNA recombinante”.

A tecnologia da recombinação genética foi o ponto de partida para os processos de clonagem e outras relevantes contribuições, algumas das quais com aplicação médica.

A obtenção da primeira molécula de DNA recombinante clonada e, por conseguinte, a fundação do método de clonagem foi referida por Cohen e Cols, em 1973<sup>47</sup>. Para o efeito foi demonstrado que uma molécula de plasmídeo podia ser replicada numa bactéria. O plasmídeo era o vector do DNA que se pretendia clonar em culturas de *E. coli*. A replicação de DNA exógeno na bactéria originava quantidades substanciais de pequenos fragmentos clonados<sup>48</sup>.

No mesmo ano, recorrendo a enzimas de restrição, Khoury e Cols<sup>49</sup> obtiveram o primeiro mapa da molécula de DNA do vírus SV40.

No ano seguinte, Robert Kornberg<sup>50</sup> demonstrou que a cromatina se organizava em unidades estruturais designadas *nucleossomas*, constituídos por regiões repetitivas de DNA em cerca de 200 pares de bases, estreitamente unidas a histonas e separadas entre si por segmentos de DNA, em parte associadas a

proteínas não-histonas ou sem proteínas envoltantes, o que conferia ao conjunto uma configuração em “colar de pérolas”.

Entre outros exemplos daquela aplicação prática destaca-se a produção de vacinas de uso humano. A primeira vacina, contra a hepatite B, também considerada a primeira vacina anti-cancro (hepatocarcinoma), foi patenteada por Blumberg e Millman em 1972<sup>51</sup>. Esta vacina consistia numa suspensão viral inactivada pelo calor.

Baruch Blumberg havia descoberto o vírus da hepatite B em 1965, que ficou conhecido originalmente por “Antigénio Austrália”<sup>52</sup>. Em colaboração com Irving Millman desenvolveu um teste diagnóstico para o vírus da hepatite B<sup>53</sup> que, desde 1971, passou a ser utilizado por rotina em todos os doadores de sangue<sup>54</sup>. Num ensaio realizado por Maupas e Cols<sup>55</sup> em chimpanzés e humanos, foram obtidos resultados favoráveis para aquele tipo de recombinação. Por fim, em 1981 foi aprovada a vacina produzida a partir dos anticorpos do plasma de doentes<sup>56</sup>, embora a sua comercialização viesse a ser suspensa nos Estados Unidos da América em 1990.

Na tentativa de se produzirem vacinas mais eficazes contra diversas doenças infecciosas, foram desenvolvidos os primeiros estudos por recombinação genética. Em 1979 e 1980 os grupos de Valenzuda<sup>57</sup>, Galibert<sup>58</sup> e Charnay e Cols<sup>59</sup> apresentaram os primeiros resultados sobre a construção de vectores (plasmídeos) clonados em *E. coli*, que incluíam sequências nucleotídicas do gene que codifica a principal proteína do antigénio superficial do vírus da hepatite B (HbsAg).

Seguiu-se em 1984 a preparação, por McAleer e Cols<sup>60</sup>, da primeira vacina produzida por recombinação genética que se revelou eficaz contra uma infecção viral humana. Nesta vacina, a sequência de DNA para o antigénio HbsAg incorporada no vector (*Saccharomyces cerevisiae*) oferecia a vantagem (relativamente à de origem plasmática) de não conter nenhuma proteína humana. A vacina demonstrou eficácia em chimpanzés e também em humanos, sendo equivalente à produzida do plasma<sup>61,62</sup>.

Ainda no mesmo ano Paoletti e Cols<sup>63</sup> apresentaram resultados sobre uma vacina produzida por engenharia genética. Este processo consistia na translocação (por “splicing”) de determinado gene do agente infeccioso codificador (HbsAg do vírus da hepatite, HSV-g do vírus do herpes simplex, ou HA do vírus da *influenza*) para o vírus da vacina da varíola. Os vírus recombinantes obtidos, depois de injectados em animais de laboratório, induziram a produ-

ção de anticorpos e revelaram-se eficazes na protecção de animais re-infectados com doses que seriam letais. Em alternativa às formas injectadas, Morin e Cols<sup>64</sup> desenvolveram uma vacina de adenovírus recombinante, portadora da sequência codificadora do antigénio de superfície do vírus da hepatite B. Em animais (hamster-modelo) especialmente adaptados à vacina recombinante administrada por via oral, induziu a produção de anticorpos contra os adenovírus e contra o HbsAg.

A vacina contra a hepatite B produzida por engenharia genética começou a ser comercializada a partir de 1987.

Southern apresentou em 1975<sup>65</sup> a técnica de “blotting” (que viria a ser designada em sua homenagem (“Southern blotting”) para detecção de genes. Uma variante deste método para a detecção de RNA em lugar de DNA (que recebeu a designação de “Northern blotting”), foi desenvolvida por Alwine e Cols, dois anos mais tarde<sup>66</sup>.

Entretanto em 1976 é anunciada a descoberta de proto-oncogenes<sup>67</sup>, e estabelecida a sua relação com os oncogenes<sup>43</sup>. Nesse trabalho foi sugerido a possibilidade da participação da sequência *src* (presente no DNA das células normais e relacionada com DNA de retrovírus) na regulação do crescimento, desenvolvimento e transformação das células normais por agentes físicos, químicos ou virais.

Não só aquelas previsões se revelaram correctas como também vieram a exercer notável influência em praticamente todos os aspectos da regulação celular e também, na investigação do cancro. Pelo seu contributo neste domínio, J. Michael Bishop e Harold E. Varmus foram galardoados em 1989 com o Prémio Nobel em Medicina.

## **Era Genómica**

**Metodologias de sequenciação e identificação do genoma** – A *Era Genómica* terá sido iniciada em 1976, com a primeira sequenciação completa de um genoma, do bacteriófago RNA MS2, por Fiers e Cols<sup>68</sup>. Todavia, três anos antes Sanger e Cols tinham já sequenciado 50 bases do genoma do bacteriófago<sup>69</sup> e, em 1975, haviam apresentado um método de sequenciação de DNA<sup>70</sup>.

Uma versão mais avançada daquele método, proposta em 1977<sup>71</sup>, veio conferir um significativo impulso aos estudos daí em diante desenvolvidos sobre o

genoma. Recorrendo àquela metodologia, Seeburg e cols<sup>72</sup> sequenciaram o primeiro gene humano, referente à hormona somatomotropina coriônica e, no ano imediato, apresentaram a síntese da somatotropina em cultura bacteriana<sup>73</sup>. Na mesma época Itakuta e Cols<sup>74</sup> divulgaram a obtenção da primeira proteína humana sintetizada *in vitro*, a partir da expressão do gene da somatostatina em culturas de *E. coli*. Dois anos mais tarde aquele grupo obteve as cadeias A e B da insulina humana, por expressão dos respectivos genes (de síntese química) em *E. coli*<sup>75</sup>. No mesmo ano, Goeddel e Cols obtiveram a somatotropina humana em *E. coli* com DNA recombinante<sup>76</sup> e Roskam e Rougeon<sup>77</sup> apresentaram a clonagem quase completa do gene humano da hormona de crescimento. Rougeon e Cols<sup>78</sup> já haviam obtido, quatro anos antes, o gene da cadeia  $\beta$  da globina do coelho que foi a primeira sequência genética de um mamífero a ser clonada. Em 1980 Derynck e Cols<sup>79</sup> comunicam a produção do gene interferão humano pelo mesmo processo.

Entretanto, em 1977, os grupos de Phillip Sharp<sup>80</sup> e Richard Roberts<sup>81</sup> constataram que os genes de organismos superiores e respectivos vírus frequentemente tendiam a separar-se, aparecendo associados no RNA<sub>m</sub>. Ao analisarem a replicação do adenovírus em células humanas verificaram que no mesmo gene co-existiam sequências de DNA que não codificavam (*intrões*), intervaladas por sequências codificadoras (*exões*). Embora todo o gene fosse transcrito, os intrões eram removidos por “splicing”, pelo que o RNA<sub>m</sub> continha somente os exões transcritos.

No mesmo ano em que Maniatis e Cols<sup>82</sup> apresentaram a técnica de rastreio para elaboração de uma biblioteca genómica de eucariotas, foi fundada a primeira empresa privada de biotecnologia (a Genentech Inc), e ocorreu o nascimento do primeiro “bebé-proveta”<sup>83</sup>.

Dois estudos independentes deram renovado contributo à identificação de genomas, um dos quais foi o do grupo de Frederick Sanger<sup>84</sup> (ao identificar o genoma completo do bacteriófago f-X174), sendo o outro o de Walter Gilbert e Allan Maxam<sup>85</sup>, ao proporem um método de sequenciação química utilizável em DNA (duplex ou em cadeia simples) de qualquer proveniência.

Em 1980 Sanger e Cols<sup>86</sup> apresentaram mais um método (que ficou conhecido por “shotgun method”) para a preparação de modelos de DNA de sequenciação, utilizando bacteriófagos para desenvolvimento de fragmentos de DNA. Dois anos depois, utilizando aquele método, o mesmo grupo apresentou o genoma completo do bacteriófago lambda<sup>87</sup>.

Ainda em 1980 Walter Gilbert e Frederick Sanger foram galardoados com o Prémio Nobel da Química, pela “contribuição que deram à determinação da sequência de bases nos ácidos nucleicos”.

A produção de hibridomas a partir de linhagens celulares de mieloma-baço, com a capacidade de formar anticorpos com actividade pré-determinada, pelos métodos descritos em 1975 por Köhler e Milstein<sup>88</sup>, foi posteriormente melhorada e expandida por David Parks e Cols<sup>89</sup>, em 1979. A partir daqui estavam criadas as condições para a produção de anticorpos para investigação e aplicação médica, de que resultou uma nova área biotecnológica.

Num intervalo muito curto sucederam-se os primeiros estudos de aplicação biotecnológica em medicina molecular. Em 1976, Kan e Cols<sup>90</sup> obtiveram o primeiro diagnóstico pré-natal de alfa-talassémia pelo exame de DNA no líquido amniótico; dois anos depois propuseram a utilização daquela metodologia no diagnóstico das drepanocitoses<sup>91</sup>. Em 1979, novamente Kan e Chan identificaram a mutação responsável pela beta-talassémia<sup>92</sup>. Alguns anos mais tarde, Saiki e Cols<sup>93</sup> deram novo impulso ao diagnóstico da drepanocitose, com a identificação do defeito genómico da cadeia beta da globina.

A conceptualização de tecnologia transgénica, iniciada na década de 80, teve como modelo de escolha preferencial o murganho (*Mus musculus*)<sup>94</sup>. O primeiro animal (murganho) transgénico, apresentado por Jon Gordon e Cols<sup>95</sup> resultou da microinjecção de duas estirpes recombinadas em plasmídeo dos vírus herpes simplex e de símio 40 (SV40) no pronúcleo de óvulos fertilizados de murganhos, seguidamente reimplantados em fêmeas pseudográvidas. Os resultados obtidos demonstraram que aqueles genes virais poderiam ser introduzidos no genoma de organismos recipientes.

Alguns anos mais tarde começou a ser desenvolvida a produção, por engenharia genética, de murganhos mutantes para estudo comparativo e investigação fisiopatológica<sup>96</sup>.

Putney e Cols<sup>97</sup>, em 1983, tendo como ponto de partida o método da sequenciação “shotgun”, desenvolveram um novo processo que possibilitava a sequenciação de fragmentos genéticos para selecção de clones de DNAC para a síntese das proteínas em geral. Este método veio a ser conhecido pelo acrónimo EST (**Expressed Sequence Tag**)

**Da diversidade genómica aos primeiros marcadores patológicos** – A recombinação, factor importante na diversidade genómica, ocorre mais frequen-

temente nos designados “pontos quentes de recombinação”, cujos determinantes moleculares e de actividade continuam por esclarecer. Outra característica da recombinação é a sua variabilidade, o que poderá justificar a ocorrência de *polimorfismos proteicos*<sup>98</sup>.

O primeiro marcador polimórfico de DNA associado ao locus genético da doença de Huntington foi identificado por Gusella e Cols, em 1983<sup>99</sup>. O genótipo, localizado no braço curto do cromossoma 4 (4p16.3), foi isolado e clonado dez anos depois pelo Huntington’s Disease Collaborative Research Group<sup>100</sup>. O defeito consiste numa sequência de três nucleótidos (CAG), multirrepetida, cuja extensão parece relacionar-se com a gravidade da doença.

Merecem referência os seguintes outros exemplos de recombinações entretanto identificadas: recombinação não uniforme do agregado genético para a beta-globina humana (descrita em 1984 por Chakravate e Cols<sup>101</sup>), a da fosfoglicomutase (identificada por Yip e Cols<sup>102</sup> em 1999), e a do complexo do antígeno da histocompatibilidade (HLA) em leucócitos humanos (apresentada em 2000 e 2001 por Jeffrey e Cols<sup>103,104</sup>). São bem conhecidas duas situações patológicas humanas resultantes da recombinação em “pontos quentes”, a doença de Charcot-Marie-Tooth, em 1992, por Patel e Cols<sup>105</sup> e duas neuropatias paralisantes periféricas, por Reiter e Cols<sup>106</sup>, em 1996.

Monaco e Cols<sup>107</sup> isolaram, em 1986, fragmentos de DNA que seriam parte do gene relacionado com a distrofia muscular dos tipos Duchenne e Becker. No ano seguinte Hoffman e Cols<sup>108</sup> identificaram a proteína expressa pelo locus genético da distrofia muscular de Duchenne, que designaram por *distrofina*. Esta proteína estrutural do músculo estriado, e também do músculo liso, estava ausente nas amostras colhidas de rapazes com aquela deficiência, bem como em murganhos com alterações homólogas.

Em 1982, Tabin e Cols<sup>109</sup> evidenciaram pela primeira vez que o DNA de células do tumor da bexiga de humanos induzia transformação maligna em culturas celulares de murganho normal, o que seria compatível com a presença de oncogenes activos no homem. No mesmo ano, o grupo de Gallo<sup>110</sup> demonstrou que o DNA de células leucémicas e de linhas hematopoiéticas humanas expressavam sequências homólogas de retrovírus tumorais em animais de experiência.

Num intervalo muito curto assistiu-se à identificação dos genomas de diversos vírus, responsáveis por considerável morbidade e mortalidade humana a nível mundial. Entre esses agentes destacam-se os prováveis causadores do

síndrome da imunodeficiência adquirida (AIDS), nas diversas variantes sucessivamente identificadas.

Em 1983, Seiki e Cols<sup>111</sup> evidenciaram a sequência completa do genoma do retrovírus HTLV-I (*human T-cell leukemia vírus*), com mais de 9.000 pares de bases, pb), associado à leucemia das células T de adulto. O genoma viral integrava-se na DNA das células leucémicas. Uma variante da leucemia das células T foi detectada no mesmo ano, por Chen e Cols<sup>112</sup> num outro retrovírus humano, o HTLV-II (com 8800pb), cuja sequência era diferente do HTLV-I. No ano seguinte, Shaw e Cols<sup>113</sup>, ao compararem os vírus HTLV-I e HTLV-III, verificaram a existência de extensas regiões homólogas em ambos, a justificarem semelhanças nalgumas propriedades, como o tropismo e a transformação. Em 1985, Wain-Hobson e Cols<sup>114</sup> sequenciaram o genoma do LAV (*lymphadenopathy-associated vírus*), enquanto Ratner e Cols<sup>115</sup> identificaram dois DNA pró-vírus do tipo III da leucemia humana de células T. Dois anos mais tarde, novamente Ratner e Cols<sup>116</sup> progrediram no esclarecimento dos mecanismos da linfocitotoxicidade dos vírus HTLV-III/LAV, a qual seria dependente de determinadas sequências nucleotídicas responsáveis pela replicação, antigenicidade e actividade linfocitopática.

No mesmo período foram definidas as sequências genómicas de retrovírus considerados responsáveis pelo síndrome da imunodeficiência adquirida em macacos (SAIDS). O genoma viral SRV-1, com pouco mais de 8.000 pares de bases, além de apresentar uma estrutura típica de retrovírus, não evidenciava semelhanças especiais com os retrovírus AIDS<sup>117</sup>. Porém, essas semelhanças foram notadas por Franchini e Cols<sup>118</sup> em estirpes virais que reproduziam sintomatologia equivalente à da AIDS em macacos (STLV-III MAC), enquanto outras, que não pareciam patogénicas, haviam sido isoladas em humanos saudáveis (STLV-III AGM). Apesar de haver uma semelhança quase total entre humanos e chimpanzés, esta espécie não desenvolve a doença, ao contrário dos macacos e outros símios. Os vírus do tipo STLV-III, que induzem em macacos rhesus um quadro patológico muito semelhante ao da AIDS humana, têm um genoma quase sobreponível ao do tipo HIV2<sup>119</sup>.

Chee e Cols<sup>120</sup> apresentaram em 1990 a sequência completa do duplex DNA da estirpe AD 169 do *Cytomegalovírus* humano, sendo referida a sua ocorrência sob múltiplas formas mutantes. Esta característica veio a ser confirmada por Dolan e Cols<sup>121</sup> em 2004, tomando como base o tipo “selvagem” que, quase

com 250.000 pares de bases, é o maior genoma de todas as espécies virais da herpes humana.

**O início da terapêutica génica** – A terapia génica das doenças humanas passou a ser um objectivo premente, em especial para as situações de natureza hereditária que, frequentemente, conduzem a complicadas limitações na qualidade de vida ou morte precoce<sup>122</sup>. A deficiência em adenosina desaminase (ADA), que está na origem de uma imunodeficiência combinada grave e rapidamente fatal, foi uma daquelas situações em que urgiam medidas terapêuticas. Em 1987 Palmer e Cols<sup>123</sup> propuseram a utilização da fibroblastos cutâneos de doentes, que, depois de enriquecidos (por transferência do retrovírus) com a enzima deficiente, seriam reintroduzidos no organismo. Foi demonstrado *in vitro* que as células infectantes passavam a produzir enzimas em quantidades dez ou mais vezes superiores às dos fibroblastos de humanos normais, assegurando igualmente a metabolização de adenosina e desoxiadenosina adicionadas ao meio.

Entre as primeiras soluções terapêuticas utilizadas incluíam-se a substituição enzimática (de ADA), preconizada por Hershfield e Cols, em 1988<sup>124</sup>, e a transplantação da medula com HLA idêntico, por O'Reilly e Cols, em 1989<sup>125</sup>. No ano seguinte foi definido o primeiro protocolo de terapêutica para a deficiência em ADA<sup>126</sup>, a ser aplicado em doentes que não poderiam receber transplante medular. Para o efeito, era proposta a infusão de linfocitos T daqueles doentes, que seriam previamente infectados por um vector retroviral contendo duas cópias do gene normal para a ADA. O método proposto foi ensaiado em 1993, por Ferrari e Cols<sup>127</sup>, em murganhos imunodependentes, injectados com linfocitos de doentes em que havia sido restaurada a actividade enzimática por transferência genética, sendo constatada a recuperação da função imunológica específica. Em 1995, o grupo de Blaese<sup>128</sup> apresentou os resultados satisfatórios verificados em duas crianças com aquela deficiência enzimática, que haviam recebido terapêutica genética durante dois anos. No mesmo ano, Bordignon e Cols<sup>129</sup> apresentaram resultados igualmente positivos, em dois doentes a que havia sido administrada, por igual período de tempo, uma terapêutica semelhante. Além dos linfocitos (T e B), também foram administradas células de medula óssea de doentes, corrigidas em ambos os casos com o gene para a ADA. Verificou-se que os linfocitos provenientes da medula óssea substituíam progressivamente os linfocitos periféricos que haviam sido administrados.

## Localização de Genes e suas Mutações

Barbara McClintock recebeu o Prémio Nobel em Medicina em 1983, pelo conjunto dos seus trabalhos sobre a transposição de genes com partes de cromossomas (“elementos genéticos móveis”). O primeiro daqueles estudos havia sido apresentado pela primeira vez, em colaboração com a sua doutoranda Harriet Creighton, mais de cinquenta anos antes<sup>130</sup>. Além de destacar a disposição linear dos genes na estrutura dos cromossomas, aquele estudo conduziu a conclusões que B. McClintock, na sua conferência Nobel, resumiu na seguinte frase: “o genoma poderia em situações de stress re-estruturar-se e renovar-se”. Não havendo dúvidas sobre a troca física de genes entre os cromossomas<sup>131</sup>, a proposta de McClintock foi retomada e desenvolvida por Pennisi<sup>131</sup> e Chicurel<sup>132</sup>, no sentido de que o genoma poderia controlar a sua própria mutação<sup>133,134</sup>.

Entretanto, em 1983 McGinnis e Cols<sup>135</sup> identificaram os genes reguladores homeóticos (Hox), que se verificou serem responsáveis pelo plano corporal da generalidade dos animais<sup>136</sup> e, aparentemente também, pelo sua macroevolução<sup>137</sup>.

Em 1985, Kary Bank Mullis e Cols<sup>138</sup> desenvolveram uma tecnologia simples e renovadora “*polymerase chain reaction*” (ou “reação em cadeia da polimerase”), que, desde então, é universalmente conhecida pela sua abreviatura (PCR). A técnica, tornou possível ampliar, com precisão e rapidez as moléculas de DNA em estudo. Em complemento, Alex Jeffreys e Cols<sup>139</sup> divulgaram no mesmo ano uma técnica que permite a identificação de marcadores genéticos sob a forma de sequências repetidas de DNA (“minissatélites”), específicas de cada indivíduo. Esta técnica, conhecida por “*genetic fingerprinting*”, veio a ter grande utilidade em medicina forense, para o esclarecimento de paternidades. As possibilidades do método foram actualizadas muito recentemente pelo seu autor principal<sup>140</sup>.

Em 1989 Bale e Cols<sup>141</sup> localizaram o gene do melanoma cutâneo maligno e do nevus displásico no cromossoma 1p. Porém, o gene da fibrose cística (ou “gene CF”, de *cystic fibrosis* ou “gene CFTR”, de *cystic fibrosis transmembrane regulator*) foi o primeiro gene de uma doença humana a ser sequenciado<sup>142</sup>, depois de White e Cols<sup>143</sup>, em 1985, haverem localizado o marcador genético daquela doença no cromossoma 7.

Merece particular destaque a anterior participação do grupo de Lap-Chee Tsui<sup>144</sup>, na abordagem metodológica que utilizou para a sequenciação daquele

gene. Em artigo de revisão foram referidas as dificuldades encontradas pelos investigadores e os progressos sucessivamente obtidos na identificação do gene CFTR. A variabilidade clínica da doença (e que é a doença recessiva mais grave na população caucasiana, em que prevalecem a doença pulmonar crónica obstrutiva e a insuficiência pancreática) resultaria da distribuição heterogénea dos alelos entre a população<sup>145</sup> com fibrose cística, em que o defeito na síntese da proteína CFTR (por mutação do respectivo gene) determinaria as condições de transporte do anião cloreto (Cl<sup>-</sup>) nos epitélios<sup>146</sup>.

O primeiro ensaio terapêutico da fibrose cística foi iniciado em 1993<sup>147</sup>.

## Sequenciação genómica de microorganismos e parasitas

A sequenciação do genoma da *S. cerevisiae* foi o primeiro dos muitos projectos desenvolvidos a partir de 1989, visando a identificação genómica completa de determinados organismos modelo (em que se incluíam também, p.ex., a *Escherichia coli* e o *Caenorhabditis elegans*), para antecipar com novas metodologias o estudo do genoma humano<sup>148</sup>.

O projecto da sequenciação da *S. cerevisiae* foi activado em 1989 por Andre Goffeau<sup>149</sup>, reunindo inicialmente 74 laboratórios da União Europeia organizados em consórcio. Esta associação contribuiu com mais de 55% para a identificação completa do genoma da levedura, juntamente com dois grupos norte-americanos.

Dois anos mais tarde tiveram início quatro projectos, dos quais três de laboratórios norte-americanos (para o estudo do genoma da *S. cerevisiae*, *E. coli* e *Mycoplasma capricolum*), decorrendo o quarto estudo (para a *Caenorhabditis elegans*) em colaboração euro-americana<sup>150</sup>. A utilização do DNA recombinante conferiu enormes possibilidades não só à compreensão da biologia molecular como também às tecnologias utilizadas no estudo do genoma, designadamente no isolamento e sequenciação dos genes. A clonagem e a sequenciação conjuntas de genomas completos passaram a ser realizáveis. Para aquela melhoria contribuíram substancialmente as inovações introduzidas na técnica de sequenciação de Sanger<sup>87</sup>.

O grupo de Hood<sup>151</sup> havia apresentado, em 1986 um processo de fixação dos marcadores fluorescentes nos nucleótidos, que possibilitava a leitura e registo imediato das bandas electroforéticas por computador. No ano seguinte Gorayne

e Cols<sup>152</sup>, recorrendo pela primeira vez a um sequenciador automático, determinaram a estrutura genética dos dois genes que codificavam os receptores b-adrenérgicos e colinérgicos do coração de rato. No mesmo ano Burke e Cols<sup>153</sup> propuseram um novo método para a clonagem de segmentos volumosos de DNA (entre 100.000 a 200.000 pb).

O mapeamento e a sequenciação do cromossoma III da *S. cerevisiae* constituiu um importante avanço na identificação do genoma dos eucariotas<sup>154</sup>. Seguiu-se, com intervalo de meses, a sequenciação completa dos cromossomas II<sup>155</sup>, III<sup>156</sup> e XI<sup>157</sup>. Finalmente, em 1996 foi identificado o genoma completo da levedura<sup>158</sup>, o qual contém cerca de 6.000 genes, em que aproximadamente 75% é constituído por sequências codificadoras de proteínas. Não obstante ser o genoma mais simples dos eucariotas conhecidos à data da sua identificação, possui  $1,4 \times 10^7$  pares de bases de DNA, distribuídas por 16 cromossomas. Por se replicar rapidamente e responder com facilidade às manipulações genéticas, a *S. cerevisiae* (foi e continua a ser) um bom e reprodutível modelo de estudo dos processos genéticos em eucariotas superiores e mamíferos.

O genoma de outra levedura (*Schizosaccharomyces pombe*), sequenciado por Wood e Cols em 2002<sup>159</sup>, revelou a particularidade de ter o número de genes (4.824) entre os eucariotas estudados, dos quais cerca de cinquenta evidenciam semelhanças com genes de doença humana.

Entretanto, o primeiro genoma de um ser vivo independente a ser completamente sequenciado (pelo grupo de Fleischmann e Cols<sup>160</sup>) foi a da bactéria *Haemophilus influenzae*. A sequência nucleotídica dos  $1,8 \times 10^6$  pares de bases que constituem o genoma da *H. influenzae* (inferior a metade do genoma da *E. coli*) dispõe-se numa molécula circular, em que mais de 85% se destina a codificar proteínas.

No mesmo ano foi apresentado por Fraser e Cols<sup>161</sup> o genoma completo do *Mycoplasma genitalium*, constituído por cerca de  $6 \times 10^5$  pares de bases. Este genoma é o mais pequeno de todas as células conhecidas; do mesmo modo os micoplasmas são, entre as bactérias actualmente existentes, as que têm menores dimensões. As diferenças profundas quanto à fisiologia e actividade metabólica de ambas as bactérias reflectem-se na complexidade relativa dos seus genomas.

O genoma de *E. coli*, circular e constituído por mais de 4.000 genes, fora já parcialmente mapeado em 1992<sup>162</sup>. Apesar da sua simplicidade aparente, e de a *E. coli* constituir uma das bactérias preferencialmente utilizadas nos estudos

de biologia molecular, a sequenciação do genoma da estirpe não patogénica K-12 da *E. coli* foi concluída somente em 1997, por Blattner e Cols<sup>163</sup>. Cerca de 2/3 dos genes identificados não possuíam função conhecida e, entre os caracterizados, mais de 85% codificavam polipéptidos. Por outro lado, muitas das sequências genéticas eram comuns aos genomas de outras bactérias. De igual modo, nos genomas das outras três estirpes patogénicas de *E. coli* (O157:H7, contaminante comum da água e alimentos; CFT073, uropatogénica e CLD933, entero-hemorragica), sequenciadas posteriormente<sup>164</sup>, foram observadas grandes diferenças, quer entre si quer em relação à estirpe não patogénica, o que estaria relacionada com o seu modo de acção<sup>165</sup>.

Em 1998, Cole e Cols<sup>166</sup> apresentaram a sequenciação completa do genoma da estirpe melhor caracterizada do *Mycobacterium tuberculosis* (H37Rv), que inclui mais de 4 milhões de bases e cerca de 4.000 genes. A maioria das sequências codificantes do genoma bacteriano determina a síntese de enzimas do metabolismo lipídico, o que, por si, constitui uma característica ímpar entre outras bactérias.

Ainda naquele ano foi anunciada a sequenciação completa do genoma de dois agentes infecciosos causadores de substancial índice de morbidade e também mortalidade humanas. O genoma da espiroqueta *Treponema pallidum*, responsável pela sífilis, foi determinado por um grupo liderado por J. Craig Venter<sup>167</sup>; aquele espiroqueta engloba um total de  $1,1 \times 10^6$  pares de bases e cerca de 1.000 genes previsíveis, com particular destaque para os que expressam proteínas de membrana e a hemolisina, (potencialmente indutora de virulência). Em comparação com outra espiroqueta patogénica (*Borrelia burgdorferi*), causadora da doença de Lyme e cujo genoma fora identificado em 1997 pelo mesmo grupo<sup>168</sup>, existem diferenças genéticas relevantes. Além do genoma de *Borrelia* ter menor extensão ( $0,9 \times 10^6$  pb) e número de genes previsíveis (853), também não possui cerca de 50% das proteínas sem funções identificadas que existem nas espiroquetas.

Ainda em 1998 foi constituído um consórcio internacional para a sequenciação do genoma do *Plasmodium falciparum*<sup>169</sup>, que é o principal causador das formas mais graves de malária no homem. O genoma da *P. falciparum*, distribuído por 14 cromossomas, tem a dimensão de 23 Mb e inclui 5.300 genes identificados. Em 2002 foi sequenciado o genoma completo da estirpe 3D7 daquele plasmódio<sup>170</sup>, assim como a sequência genómica do *Plasmodium yoeii*<sup>171</sup> que é um parasita dos roedores-modelo utilizados nos estudos da malária. O

genoma do mosquito da malária *Anopheles gambiae* foi determinado no mesmo ano<sup>172</sup>. Em 2005 foi a vez de outro parasita, o *Trypanosoma brucei*, causador da doença do sono, ser codificado por Berriman e Cols.<sup>173</sup>. Entre os 9.068 genes previsíveis, num genoma com 26 Mb, destacavam-se os responsáveis pela codificação das proteínas que afectam o sistema imunitário humano, uma característica também observada nos plasmódios<sup>170</sup>.

Em 2000 Parkhill e Cols<sup>174</sup> revelaram o genoma completo da estirpe Z2491 da *Neisseria meningitidis*, responsável pelas maiores e mais graves epidemias de meningite. A variabilidade antigénica da bactéria seria atribuível ao grande número de sequências repetitivas do genoma.

No mesmo ano Heidelberg e Cols.<sup>175</sup> apresentaram a sequência genómica completa da estirpe El Tor do vibrião (*Vibrio cholerae*), agente patogénico da cólera. O genoma é constituído por  $4 \times 10^7$  pb, distribuídos por dois cromossomas circulares de dimensões desiguais. As principais funções celulares e patogénicas são referidas aos genes existentes no cromossoma maior, enquanto a composição dos outros cromossomas se assemelha à dos plasmídeos.

**O genoma dos primeiros eucariotas-modelo** – Entre o genoma das leveduras e o do homem situam-se, como exemplos de complexidade intermédia, o da planta *Arabidopsis thaliana*, o do nemátodo *Caenorhabditis elegans* e o da mosca da fruta, *Drosophila melanogaster*.

A sequenciação do genoma do nemátodo *C. elegans* foi iniciado na década de 90 pelo grupo liderado por John Sulston<sup>176,177</sup> e Robert Waterston<sup>178</sup>. Em 1992 estava concluído o primeiro mapa genético daquele nemátodo<sup>179</sup>. Em 1998, um consórcio de laboratórios<sup>180</sup> deu quase por concluído aquele trabalho, o qual representava o primeiro genoma de um eucariota multicelular a ser sequenciado. Também pela primeira vez tornava-se possível relacionar o genoma de um animal minúsculo (cerca de 1mm de comprimento) com diversos conjuntos de funções importantes, particularmente no âmbito do desenvolvimento e comportamento do sistema nervoso<sup>181</sup>. O *C. elegans* tem sido igualmente fundamental para a elucidação dos mecanismos da apoptose<sup>182</sup>. Mais recentemente foi apresentada uma revisão do genoma do *C. elegans* que abre perspectivas à compreensão molecular de diversos mecanismos no homem<sup>183</sup>.

As conclusões obtidas foram o culminar de uma previsão de Sydney Brenner<sup>184</sup>, cerca de trinta anos antes (quanto ao futuro da Biologia Molecular e, em particular, sobre a importância da sua envolvimento no estudo do sistema nervoso, utilizando o *C. elegans* como modelo electivo). Entre as dezenas de genes

que o grupo liderado por Brenner identificou, e dos quais obteve mutações, o das cadeias pesadas de miosina na *C. elegans* foi o primeiro a ser identificado num organismo vivo<sup>185</sup>.

Na opinião dos mentores do Consórcio FlyBase<sup>186</sup>, a *Drosophila* será eventualmente o eucariota mais estudado em termos genéticos, além do Homem. À data calculava-se em 3.000 novas publicações anuais sobre a *Drosophila*. A sequenciação do genoma da *Drosophila* foi concluída em 2000 por dois consórcios, independentes entre si<sup>187, 188</sup> que utilizaram como base de trabalho o mapeamento obtido da *Drosophila* pelos grupos de Garza<sup>189</sup> e Merriam<sup>190</sup>, obtido cerca de dez anos antes. O genoma, constituído por cerca de 180Mb (distribuídos por quatro cromossomas) foi determinado por sequenciação “shotgun” e clonagem. Na fracção eucromática do genoma virtualmente completo da *Drosophila*, Adam e Cols<sup>187</sup> identificaram cerca de 13.600 genes. Em 2002, Peter e Cols<sup>191</sup> apresentaram o mapeamento do cromossoma X da *Drosophila*, junto com os seus principais genes.

Entre outras vantagens da *Drosophila*, como meio de estudo genético, destaca-se a particularidade de possuir cromossomas gigantes nas células que não se dividem (p.ex., glândulas salivares das larvas). Aquelas dimensões resultam de sucessivas replicações em cromossomas indivisos, de que resultou o mapeamento de centenas de moléculas idênticas de DNA, que evidenciam múltiplas bandas, dispostas paralelamente entre si e perpendiculares ao comprimento. Este aspecto, visível em microscópico óptico, ao ser alinhado com o mapa genético, permite sem dificuldade visualizar, p.ex., a ausência de bandas, que correspondem à localização de deleções.

Através do estudo da *Drosophila* procurava-se, também, preparar a abordagem metodológica e estratégica do genoma humano<sup>187</sup>. Por esse motivo, o projecto da *Drosophila* foi incluído e subsidiado através de fundos governamentais norte-americanos, sendo concretizado em cerca de um ano. Entre outras conclusões, o conteúdo genómico permitiria o reexame de diversas doenças humanas, relacionadas com processos de replicação, reparação e tradução (p.ex., doenças neurológicas e cancro) e, ainda, com o metabolismo de drogas e toxinas, receptores nucleares hormonais e o metabolismo do ferro.

Comparando o genoma da *Drosophila* com os do *C. elegans* e da *S. cerevisiae* verificou-se, com alguma surpresa, que um organismo multicelular com razoável grau de complexidade (como a mosca), possuía somente o dobro da quantidade das proteínas expressas da levedura, que é um organismo unicelu-

lar. Daqui foi concluído que a complexidade de um organismo não é proporcional ao número dos seus genes. Um outro facto saliente consiste em que ambos os metazoários possuem proteínas mais complexas e extracelulares do que a levedura. Por sua vez, as proteínas com vários domínios existem em maior quantidade e diversidade nas moscas do que nos nemátodos.

O genoma da *Arabidopsis* foi o primeiro modelo a ser definido para uma planta (mais de 90%, num total aproximado de 125 Mb, distribuídos por 5 cromossomas)<sup>192</sup>, a que se seguiu muito recentemente o do arroz (*Oryza sativa*), também identificado em cerca de 95%, num total de 389Mb<sup>193</sup>.

**A identificação do genoma humano** – A proposta para a sequenciação do genoma humano foi apresentada numa reunião realizada na Universidade da Califórnia em Santa Cruz, em Junho de 1985. Desde esse momento e até ao projecto ser iniciado decorreram cinco anos, durante os quais investigadores e instituições debateram estratégias e metodologias e, também, dirimiram opiniões e responsabilidades. Numa dessas reuniões, realizada em Cold Spring Harbor, em Abril de 1988, foi fundada uma associação internacional de investigadores que se propunham participar no projecto do genoma humano<sup>194</sup>.

Aquela associação, que ficou conhecida pelo acrónimo HUGO (de “Human Genoma Organization”), propunha-se, entre outros objectivos, a apoiar a coordenação das actividades científicas e os recursos disponíveis para o desenvolvimento do projecto<sup>195</sup>. James Watson foi nomeado coordenador do projecto<sup>196</sup>, o qual tinha duas finalidades principais (i) definir a estrutura do DNA nos cromossomas humanos e em outros organismos, e (ii) desenvolver novas tecnologias e o mapeamento da sequenciação do genoma.

Para o impulso inicial do projecto muito contribuiu o mapa genético desenvolvido por marcadores cromossómicos, que havia sido publicado em 1987 pelo grupo de Donis Keller<sup>197</sup>, bem como um conjunto de técnicas e processos inventado desde os anos 70, designadamente a transcrição reversa do RNA<sup>39,40</sup>, a clivagem do DNA por endonucleases<sup>29-31</sup>, a clonagem molecular (que constituiu o primeiro passo para a recombinação de DNA)<sup>32</sup> iniciada pela inserção de fragmentos do DNA que se pretendia replicar num DNA neutro ou reinstalar numa célula-hospedeiro<sup>153</sup>, a separação electroforética de fragmentos com dezenas de milhões de bases nucleotídicas<sup>198</sup> e a ampliação de regiões de DNA pelo PCR<sup>138</sup>.

Em 1990, após serem ultrapassadas algumas divergências institucionais mais críticas, o US Department of Energy e o National Institute of Health apresenta-

ram em conjunto, ao governo norte-americano (numa atitude até então inédita), uma proposta de projecto (designado “Human Genome Project”) que visava a sequenciação do genoma humano, a ser financiado durante quinze anos<sup>199</sup>.

Entretanto, também em Itália (1987), Reino Unido (1989), França (1990), Rússia (1990) e Comunidade Europeia (1990) foram apresentados projectos similares.

Em 1995 foi definida uma estratégia para a sequenciação completa do genoma humano em cinco anos. No programa era prevista a conjugação de esforços multicêntricos e apoios financeiros vultuosos<sup>200,201</sup>.

No mesmo ano, Hudson e Cols<sup>202</sup> definiram um mapa genético baseado numa biblioteca cromossómica artificial de leveduras, que abrangia cerca de 94% do genoma. Em 1998 foi apresentado o mapa físico de mais de 30.000 genes referentes às proteínas mais comuns<sup>203</sup>.

À data daqueles estudos, o genoma humano estava sequenciado numa extensão de 110 Mb. Tendo em conta que o genoma se distribui por 24 cromossomas (2 cromossomas sexuais e 22 autossomas), cada um contendo entre  $5 \times 10^4$  a  $26 \times 10^4$  Mb de DNA, previa-se que houvesse 100.000 genes. A dimensão da sequência do DNA a determinar e a rapidez pretendida para a sua conclusão motivaram o desenvolvimento de técnicas mais eficazes, designadamente o desenvolvimento de identificadores genéticos EST<sup>97</sup>. A sua utilização em 1991, por Adams e Cols<sup>204</sup>, abriu caminho, rapidamente, à descoberta e mapeamento dos genes teciduais<sup>205</sup> e, ao mesmo tempo, criou uma premente necessidade em novos recursos informáticos, que impulsaram o posterior desenvolvimento da bioinformática.

Em 1993, White e Cols<sup>206</sup> apresentaram novos algoritmos adequados à análise de centenas de milhares de sequências EST. Entre outras adaptações metodológicas, em 1996 foi introduzido por Schmidt e Cols<sup>207</sup> o processo de mapeamento e sequência simultâneas do genoma humano, utilizando cromossomas artificiais bacterianos (BACs). Seguidamente, foi proposto por Weber e Myers, em 1997<sup>208</sup>, a sequenciação “shotgun” total do genoma, a qual não teve aceitação imediata. Após algumas modificações e respectiva automatização, aquele processo veio a ser aplicado<sup>209</sup>. O processo de sequenciação do genoma humano fora previamente ensaiado, com êxito, no estudo da *Drosophila melanogaster*, como foi já referido<sup>187</sup>.

A constituição da empresa Celera, em 1998<sup>209</sup>, para intervir num projecto de âmbito público, modificou a planificação e as metodologias a serem seguidas<sup>210</sup>.

Então previa-se que a sequenciação do genoma humano estivesse concluída em 2001. Depois, por motivos diversos, foi considerada a conveniência em antecipar aquela meta (como sucedeu), para o que foram utilizados financiamentos suplementares, disponibilizados sobretudo por dois organismos, o National Human Genome Research Institute e o Wellcome Trust.

Ian Dunham liderou o grupo do Sanger Center que sequenciou, pela primeira vez, um cromossoma humano, o 22, cujos resultados foram apresentados em finais de 1999<sup>211</sup>. A sequência da eucromatina do cromossoma 22 incluiria, em cerca de 33,4 megabases, pelo menos 545 genes e 134 pseudogenes. No ano seguinte foi sequenciado<sup>212</sup> o cromossoma 21 por um consórcio de laboratórios japoneses e norte-americanos. Em cerca de 99,7% do DNA replicado foram definidas mais de 33 Mbases, de que se identificaram 127 genes concluídos, 59 pseudogenes e 98 genes previsíveis. No cromossoma 21, que é o autossoma mais pequeno, foram mapeados genes determinantes de diversas situações patológicas. O síndrome de Down está relacionado com uma cópia extra (trissomia 21).

Em 2001, Tilford e Cols<sup>213</sup> determinaram o mapa físico do cromossoma Y humano, localizando 758 marcadores genéticos, dos quais 136 existentes na região NRY (região não recombinante). Dois anos mais tarde foi constatado por Slaletsky e Cols<sup>214</sup> que aquela região NRY (que abrange cerca de 95% de todo o cromossoma) é específica e determinante do sexo masculino.

No mesmo ano foram apresentados, por Bentley e Cols<sup>215</sup>, os mapas físicos de seis autossomas (1, 6, 9, 10, 13 e 20) e o do cromossoma X, os quais representam cerca de 1/3 do genoma humano. Foi também publicado por Deloukas e Cols<sup>216</sup> a sequenciação do cromossoma 20 humano, que inclui quase 60 Mb em cerca de 99,4% do DNA eucromático, com 727 genes e 168 pseudogenes identificados.

O cromossoma 7 humano foi o quarto autossoma a ser completamente sequenciado. Os resultados, obtidos por processos diferentes, foram apresentados em 2003 por Sherer e Cols<sup>217</sup> e Hillier e Cols<sup>218</sup>, estando ambos os grupos organizados em consórcio internacional. O cromossoma 7 contém entre 153 a 158 milhões de pares de bases, o que representa cerca de 5% de todo o genoma humano. Na fracção sequenciada (cerca de 99,4%) foram identificados entre mil a dois mil genes responsáveis por doenças humanas. Os genes determinantes da fibrose cística, de várias formas de cancro e, eventualmente, do autismo, estão localizados no cromossoma 7. Uma característica do

cromossoma 7, também identificada no cromossoma Y, é a de que contém a maior quantidade de duplicações do genoma humano, o que poderá predispor a mutações frequentes<sup>219</sup>.

Finalmente, em 2001 foi apresentado por McPherson e Cols<sup>220</sup> o mapa físico do genoma humano, desenvolvido em meses por um consórcio internacional (International Human Genome Mapping Consortium), a partir de um cromossoma bacteriano artificial integrado em regiões específicas cromossômicas. Entretanto, Lander e Cols<sup>221</sup>, do mesmo consórcio, apresentaram a sequência inicial do genoma. Simultaneamente, Sachidanandam e Cols<sup>222</sup>, que integravam um outro consórcio (International SNP MaP Working Group), apresentaram o mapa da sequência genômica baseado em polimorfismos simples, que incluía todos os genes e outras sequências genômicas então conhecidas. Ainda no mesmo mês de 2001 o grupo liderado por Venter, da Celera Genomics, divulgou a sequência quase completa da parte eucromática do genoma humano<sup>223</sup>, obtida por dois processos com resultados semelhantes. No total, foram identificadas cerca de 98,9% da eucromatina, de que 5,5% serão sequências de origem não humana. Após sucessivas correções, de um valor de 39.144 genes previstos foram confirmados cerca de 23.000 presumíveis genes, e considerada a possibilidade de alguns outros milhares para novas proteínas por identificar, perfazendo um valor máximo de 38.000 (incluindo todos os genes previstos) ou 26.000 (se forem excluídas as homologias com outras espécies ou fragmentos sobrepostos). Em média, cada gene abrangeria cerca de 27.000 bases. Aproximadamente 35% do genoma identificado era constituído por sequências repetitivas, destacando-se o cromossoma 19, por ter a maior densidade genética e também a maior densidade de sequências repetitivas.

Foi sugerido que a quantidade de genes identificados, menor do que havia sido previsto (entre 50.000 a 140.000), seria compensada pela diversidade de conformações proteicas e mecanismos de transcrição, tradução e regulação pré-tradução proteica. Estas particularidades da regulação justificariam a individualidade funcional e comportamental de cada ser humano, apesar de se saber que 99,9% do genoma é idêntico ao de outros humanos.

**Genes e pseudogenes do genoma humano** – No pressuposto<sup>224</sup> de que cerca de  $\frac{1}{4}$  do genoma humano seria representado por *pseudogenes* (isto é, genes não funcionais derivados por transcrição reversa de RNAm e subsequente integração no DNA do genoma, ou resultante da duplicação de genes, com disfunção da cópia inútil), Harrison e Cols<sup>225</sup> calcularam a quantidade daquelas sequên-

cias nos cromossomas humanos 21 e 22. De acordo com as observações realizadas, haveria cerca do dobro dos pseudogenes previamente anunciados no conjunto dos dois cromossomas. Ao extrapolar esta nova realidade para todos os cromossomas, propuseram a existência de mais de 20.000 pseudogenes em todo o genoma humano. Estas sequências acumular-se-iam com particular abundância na proximidade dos centrómeros. Alguns dos pseudogenes de diferentes idades evolutivas que foram sequenciados expressavam proteínas semelhantes, justificando a proposta de que seriam “fósseis” moleculares presentes no genoma humano<sup>226</sup>.

Na realidade, é nos centrómeros e telómeros que reside a maior parte das sequências repetitivas e aparentemente inúteis, não codificantes<sup>133</sup>. Em estudos comparativos do genoma humano e de murganhos foi constatado que aquelas regiões se encontravam entre as mais conservadas do DNA<sup>133</sup>. Igualmente em ambas as espécies existe grande variabilidade da densidade genética, havendo regiões com muitos e outras com poucos genes. Cerca de 9% do genoma humano é constituído por regiões “desertas” de genes, cuja sequência de bases é, na maioria, semelhante à dos murganhos. O aumento da dimensão do genoma, a par com a diferenciação evolutiva das espécies, decorreria à custa do DNA redundante, sendo propostas várias hipóteses sobre a sua origem (adaptação ou selecção) e funções<sup>227,228-229</sup>. Em qualquer dos casos, o aumento do DNA redundante provoca a condensação da cromatina intranuclear<sup>229</sup>.

É interessante referir uma interpretação relativamente semelhante apresentada em 2005 por de Parseval e Cols<sup>230</sup>, a propósito da presença no genoma humano de uma elevada quantidade de sequências móveis, que replicam por transcrição reversa. Esta ocorrência sugere que tenham origem retroviral, resultante de infecções por vírus activos em épocas remotas da evolução do Homem. Aquelas sequências, denominadas *retrovírus endógenos humanos*, apresentam grande semelhança com os actuais retrovírus. Por ser a sua funcionalidade também equivalente, admite-se que algumas daquelas sequências contenham genes activos que, tendo sido incorporados no genoma humano, exercem funções de interesse para o hospedeiro, ou estão potencialmente implicadas em determinadas patologias.

Aproximadamente 40% das proteínas previsíveis não têm função identificada. Comparado aos genomas da *Drosophila* e *C. elegans*, o humano apresenta maior predomínio nos domínios da regulação do desenvolvimento e das actividades celulares, em particular as funções neuronais, vasculares, hemos-

tase, resposta imunológica adquirida e complexidade do citoesqueleto<sup>223</sup>. Por sua vez, comparativamente aos outros genomas já sequenciados<sup>231</sup>, poderá concluir-se que nem o número de genes nem a dimensão do genoma justificam as diferenças entre o homem e o chimpanzé, cujo volume cerebral e organização neuroanatômica são também indistinguíveis<sup>232</sup>. As diferenças deverão ser atribuídas a uma maior capacidade de interacção nessas estruturas, determinadas por um complexo sistema regulador de actividade genética, que se expressa também numa muito elevada quantidade de receptores e proteínas mediadoras de sinal celular, e amplifica a rede interactiva do organismo humano<sup>232, 233</sup>.

Já em 1975, King e Wilson<sup>234</sup> haviam proposto que as reduzidas diferenças verificadas entre os genomas do Homem e do chimpanzé, resultariam de mecanismos diferentes de controlo dos períodos de activação-desactivação dos respectivos genes estruturais.

Quanto aos chimpanzés em particular, e a todos os mamíferos e vertebrados em geral, pretende-se esclarecer a natureza dos genes, mecanismos de regulação e outras semelhanças e diferenças genómicas que justifiquem o processo evolutivo da espécie e as divergências que apresentam relativamente ao património genético do homem<sup>231</sup>.

Apesar das sequências de DNA do homem e chimpanzé serem idênticas em cerca de 99% a sua extensão, apresentam regiões de recombinação desproporcionadamente elevadas e com variação rápida, que poderão justificar as diferenças genéticas que se acumularam desde que ambas as espécies divergiram dos seus antepassados comuns<sup>233</sup>. Thomas e Cols<sup>235</sup>, que estudaram determinadas regiões genómicas em diversos vertebrados, verificaram que uma sequência de DNA com 12 Mb em doze daquelas espécies correspondia a uma região com 1,8 Mb do cromossoma 7 em dez espécies (entre os quais figura por ex., o gene da fibrose cística). O estudo confirmou ainda, por curiosidade, que os roedores representam um grupo-irmão dos primatas.

Também em 2003 Guyon e Cols<sup>236</sup> construíram o mapa genético do cão, identificando mais de 3.000 marcadores. O genoma preliminar do cão doméstico (*Canis familiaris*) foi apresentado em 2005 por Lindblad-Toh e Cols<sup>237</sup>. Paralelamente, foi identificado o mapa de polimorfismos em função das raças, a qual terá potencial utilidade para o estudo de doenças, quer do cão quer do homem.

Em 2004 o grupo liderado por Venter<sup>238</sup> divulgou um sequência mais completa e ordenada do genoma humano, enquanto Schmutz e Cols<sup>239</sup> apresentaram uma avaliação qualitativa da sequência completa do DNA. Atendendo aos

elevados custos de finalização da sequência genómica, Blakesley e Cols<sup>240</sup> propuseram, no mesmo ano, uma alternativa metodológica (de “grau intermédio”) que, no entanto, asseguraria grande qualidade ao acabamento de 99% da sequência total, com custos muito inferiores aos dos dois processos previstos no Human Genome Project. Palmer e McCombie<sup>241</sup> já haviam referido, em 2002 que a necessidade de finalização da sequência genómica dos diversos organismos estudados e futuros, constituía um empreendimento muito dispendioso.

Em 2002, Aparicio e Cols<sup>242</sup> divulgaram a sequência do DNA do peixe *Fugu rubripes*, o qual passou a ser o 2.º vertebrado com o genoma completamente identificado, um ano e meio depois da apresentação do genoma humano. Relativamente ao homem, o genoma do *Fugu* tem somente 1/9 das dimensões (com 365 Mb) mas quase igual número de genes, o que será atribuível a uma muita reduzida quantidade de cistrões e de sequências repetitivas. O facto de o *Fugu* pertencer a uma espécie evolutivamente muito afastada do homem (mas em que se presume haver identidade de sítios conservados nas sequências proteicas) confere grande interesse ao seu estudo, para a compreensão das bases genéticas e polimorfismos associados a doenças humanas.

A par com os estudos da sequenciação e comparação do genoma humano<sup>223, 238</sup>, foram organizados novos consórcios internacionais que, a partir de 2004, começaram a apresentar resultados consistentes sobre a identificação dos genomas de murganhos (*Mus musculus*)<sup>243</sup>, rato (*Rattus norvegicus*)<sup>244</sup> e chimpanzés (*Pan troglodytes*)<sup>232</sup>.

Relativamente aos murganhos de ratos, mais utilizados como animais de laboratório, destaca-se o propósito da obtenção de informações aplicáveis em trabalhos de investigação biomédica e que, por consequência, conduzem a resultados benéficos para a saúde humana.

A conclusão dos genomas humano e de outras espécies abriu perspectivas para estudos comparativos que, entre outros registos, depende de bases de dados disponíveis e centralizadas pelos identificadores e, também, da anotação das proteínas codificadas por cada um dos genes identificados. A definição de uma associação que congregue o conhecimento existente sobre os genomas de eucariotas e as sequências peptídicas que determinam constitui o objectivo do “Peptid Atlas Project”, que Desiere e Cols<sup>245</sup> apresentaram como primeira etapa para uma Era Pós-Genómica (neste caso, Proteómica), que venha a ter êxito equivalente à fase Genómica que concretizou a identificação do genoma humano.

## Perspectivas Futuras e Conclusões

Desde os trabalhos de Gregório Mendel com as ervilhas (por longo tempo ignorados), até à identificação da estrutura e finalidade potencial do DNA como sede da informação genética, decorreram cerca de oito décadas. Porém, após a conceptualização da estrutura de DNA, foi necessário cerca de meio século mais para serem descobertos, em rápida sucessão, os processos de descodificação do DNA, da identificação e reprodução de genes isolados e, por fim, com o “Human Genoma Project”, um largo número dos genomas (uns completos e outros em via de o serem) de procariotas e eucariotas.

A decifração do genoma humano e ou dos organismos vivos em geral transformou-se num empreendimento multidisciplinar, que envolve recursos humanos, técnicos e financeiros gigantescos, à escala mundial, inicialmente da iniciativa de instituições universitárias e públicas e, nos últimos anos, também com a participação de organismos privados. Biólogos, médicos, químicos, físicos, matemáticos, informáticos, engenheiros e um conjunto de outras formações técnicas, biotecnológicas e científicas juntaram esforços, conhecimentos e capacidades para a decifração dos segredos da vida e da sua evolução e, muito pragmativamente, para melhorar e prolongar a longevidade humana. A exemplar partilha de esforços e tecnologias daqueles investigadores, que inclusivamente disponibilizam os seus resultados antes da publicação, tem sido um ponto alto da intercolaboração no mundo da Ciência.

Em pouco mais de uma década foram identificados os genomas completos de organismos muito simples, procariotas e eucariotas, até se conhecer o do homem. Foi comprovada uma grande homogeneidade genómica entre o homem e outras espécies, o que sugere que, além dos genes, existem outros factores que controlam a informação genética e, portanto, a especificidade de cada organismo. Entretanto, também a identificação de famílias de genes actuantes e as características que modulam a expressão genética em cada tecido são alvo de investigação intensa.

Novos campos de investigação são perspectivados para as próximas décadas. À Genómica segue-se a Proteómica, e já se anuncia a Metabolómica, como passos de desenvolvimento inevitável.

Sucedem-se as descobertas de mutações genéticas, presumivelmente responsáveis por doenças humanas que integram o conjunto das fatalidades hereditárias. Começam também a ser reveladas outras deficiências conducentes a

situações degenerativas e cancerígenas, e cresce a investigação sobre microorganismos e parasitas, em que se incluem os causadores de epidemias mortíferas como a meningite, a cólera e o paludismo.

A terapêutica génica dá os seus primeiros passos e, simultaneamente, também crescem os problemas éticos relativos à intervenção no genoma humano e à sua clonagem. A cooperação entre instituições públicas e privadas para a sequenciação genómica, apesar dos benefícios que trouxe à evolução do processo, não oculta algum mercantilismo, numa actividade que se pretendia essencialmente interessada no desenvolvimento do conhecimento para um mundo melhor e, pelo menos, para a revelação da verdade científica.

Porém, é um factor incontestável que a evolução vertiginosa dos acontecimentos referidos alterou, profunda e rapidamente, os conceitos e sobre a vivência actual. Sobretudo, abriu caminho para horizontes desconhecidos, não só sobre a natureza humana como ainda sobre a Biosfera no seu conjunto.

## Notas

- <sup>1</sup> Martins e Silva J. – “Lições de Química Fisiológica”, Fac. Med. Univ. Lourenço Marques 1972/73 (reedição *fac-simile* em cinco volumes, Publ. Ciência e Vida, Lisboa, 2004.
- <sup>2</sup> Kelley W.N. – From the double helix to genomic medicine. *Ann. Int. Med.* 2003; 138:603-604.
- <sup>3</sup> Cooper R.S., Psaty B.M. – Genomics and Medicine: distraction, incremental progress, of the dawn of a New Age?. *Ann Intern. Med.* 2003; 138:576-580.
- <sup>4</sup> Holtzman N.A, Marteau T.M. – Will genetics revolutionize medicine? *New Engl. J. Med.* 2000; 343:141-144.
- <sup>5</sup> Gravele B.R. – Alternative splicing: increasing diversity in the proteomic world. *Trends in Genetics* 2001; 17:100-107.
- <sup>6</sup> Ast G. – The alternative genome. *Sci. Amer* 2005; 292:40-47.
- <sup>7</sup> Watson, J.D., Crick F.H.C. – Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* 1953; 171:737. Wilkins M.H.F, Stokes A.R., Wilson H.R. – Molecular structure of nucleic acids. Molecular structure of desoxyribose nucleic acids. *Nature* 1953; 171:738. Franklin R.F, Gosling R.G. – Molecular structure of nucleic acids. Molecular configuration in sodium thymonucleate. *Nature* 1953; 171:740.
- <sup>8</sup> Klug A. – Rosalind Franklin and the discovery of the structure of DNA. *Nature* 1968; 219:808-810, 843-844.
- <sup>9</sup> Crick F. – On protein synthesis. *Symp. Soc. Exp. Biol.* 1958; 12:138-163.
- <sup>10</sup> Hoagland M.B., Keller E.B., Zamecnick P.C. – Enzymatic carboxyl activation of amino acids. *J Biol Chem.* 1956; 218:345-358.
- <sup>11</sup> Zamecnick P.C., Stephenson M.L., Hecht L.I. – Intermediate reactions in amino acid incorporation. *Proc Natl Acad Sci USA.* 1958; 44:73-78.
- <sup>12</sup> Brenner S., Jacob F, Meselson M. –An unstable intermediate carrying information from gene to ribosomes for protein synthesis. *Nature* 1960; 190:576.
- <sup>13</sup> Brenner S., Stretton A.O.W., Kaplan S. – Genetic code: the “nonsense” triplets for chain termination and their suppression. *Nature.* 1965; 206:994-998.
- <sup>14</sup> Crick F.H.C., Barnett L., Brenner S., Watts-Tobin R. – General nature of the genetic code for proteins. *Nature* 1961; 192:1227-1232.
- <sup>15</sup> Hoagland M.B. – Celebrating complementary. *Ann. Intern. Med.* 2003; 138:583-586.
- <sup>16</sup> Lyon M.F. – Gene action in the X-chromosome of the mouse (*Mus musculus* L). *Nature* 1961; 190:372-373;
- <sup>17</sup> Nirenberg M.W., Matthaei J.H. – The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring on synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* 1961; 47:1588-1602.
- <sup>18</sup> Yanofsky C, Carlton C.C., Guest J.R., Helinski D.R., Henning U. – On the colinearity of gene structure and protein structure. *Proc. Natl. Acad. Sci. USA* 1964; 51:266-272.
- <sup>19</sup> Khorana H.G. – Polynucleotide synthesis and the genetic code. *Fed. Proc.* 1965; 24:1473-1487.
- <sup>20</sup> Leder P, Clark B.F.C., Sly W.S., Pestka S., Nirenberg N.W. – Cell free peptide synthesis upon synthetic oligodeoxynucleotides. *Proc. Natl. Acad. Sci. USA* 1963; 50:1135-1143.
- <sup>21</sup> Söll D., Ohtsuka E., Jones D.S., Lohermann R., Hayatsu H., Nishimura S., Khorana H.G. – Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA'S to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 aminoacids. *Proc. Natl. Acad. Sci. USA* 1965; 54:1378-1385.

- <sup>22</sup> Weisblum B., Benzer S., Holley R.W. – A physical basis for degeneracy in the amino acid code. *Proc. Natl. Acad. Sci. USA* 1962; 48:1449-1454.
- <sup>23</sup> Holley R.W., Everett G.A., Madison J.T., Zaur A. – Nucleotide sequences in the yeast alanine transfer ribonucleic acid. *J. Biol. Chem.* 1965; 240:2122-2128.
- <sup>24</sup> Nelson J.A., Ristow S.C., Holley R.W. – Studies on the secondary structure of yeast alanine tRNA: reaction with N-bromosuccinimide and nitrous acid. *Biochim. Biophys. Acta* 1967; 149:590-593.
- <sup>25</sup> Apgar J., Holley R.W., Merrih S.H. – Purification of the alanine-, valine- histidine – and tyrosine- acceptor ribonucleic acids from yeast. *J. Biol. Chem.* 1962; 237:796-802
- <sup>26</sup> Nirenberg M. – The genetic code. Nobel Lecture, 1968.
- <sup>27</sup> Ptashne M. – Isolation of the lambda phage repressor. *Proc. Natl. Acad. Sci. U S A.* 1967; 57:306-313.
- <sup>28</sup> Gilbert W., Muller-Hill B - The lac operator is DNA. *Proc. Natl. Acad. Sci. U S A.* 1967; 58:2415-2421.
- <sup>29</sup> Smith H.O., Wilcox K.W. – A restriction enzyme from *Haemophilus influenzae*. I. Purification and general properties. *J. Molec. Biol.* 1970; 51:379-391.
- <sup>30</sup> Kelly T.J.Jr., Smith H.O. – A restriction enzyme from *Haemophilus influenzae*. II. *J. Molec. Biol.* 1970; 51:393-409.
- <sup>31</sup> Danna K.J., Nathans D. – Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* 1971; 68:2913-2917
- <sup>32</sup> Arber W. – Host-controlled modification of bacteriophage. *Ann. Rev. Microbiol.* 1965; 19:365-378.
- <sup>33</sup> Arber W. – Promotion and limitation of genetic exchange. *Science* 1979; 205:361-365.
- <sup>34</sup> Wu R., Taylor E. – Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage lambda DNA. *J Mol Biol.* 1971; 57:491-511.
- <sup>35</sup> Cohen S.N. – Transposable genetic elements and plasmid evolution. *Nature* 1976; 263:731-738.
- <sup>36</sup> Kelly T.J.Jr, Nathans D. – The genome of simian virus 40. *Adv. Virus Res.* 1977; 21:85-173.
- <sup>37</sup> Weiss B., Richardson C.C. – Enzymatic breakage and joining of deoxyribonucleic acid. I. Repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage. *Proc. Natl. Acad. Sci. USA* 1967; 57:1021-1028.
- <sup>38</sup> Jackson D.A., Symons R.H., Berg P. – Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 1972; 69:2904-2909.
- <sup>39</sup> Temin H.M., Mizutani S. – RNA-dependent DNA polymerase in virions of ronssarcoma virus. *Nature* 1970; 226:1211-1213.
- <sup>40</sup> Baltimore D. – Viral RNA-dependent DNA polymerase. *Nature* 1970; 226:1209-1211
- <sup>41</sup> Shein H.M., Enders J.F. – Multiplication and cytopathogenicity of Simian vacuolating virus 40 in cultures of human tissues. *Proc. Soc. Exp. Biol. Med.* 1962; 109:495-500.
- <sup>42</sup> Vogt M., Dulbecco R. – Steps in the neoplastic transformation of hamster embryo cells by polyoma virus. *Proc. Natl. Acad. Sci. USA* 1963; 49:171-179.
- <sup>43</sup> Dulbecco R. – Cell transformation by viruses. *Science* 1969; 166:962-968
- <sup>44</sup> Committee on Recombinant DNA molecules (Chairman: Paul Berg) – Potential biohazards of recombinant DNA molecules. *Proc. Natl. Acad. Sci. USA* 1974; 71:2593-2594
- <sup>45</sup> Berg P., Baltimore D., Brenner S., Roblin R.O.III, Singer M.F. – Summary statement of the Asilomar Conference on recombinant DNA molecules. *Proc. Natl. Acad. Sci. USA* 1975; 72:1981-1984
- <sup>46</sup> Guidelines for research involving recombinant DNA molecules. *Federal Register* 1976; 41:29911-27943.
- <sup>47</sup> Cohen S.N., Chang A.C.Y., Boyer H.W., Helling R.B. – Construction of biological functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. USA* 1973; 70:3240-3244.

- <sup>48</sup> Morrow J.F., Cohen S.N., Chang A.C., Boyer H.W., Goodman H.M., Helling R.B. – Replication and transcription of eukaryotic DNA in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 1974; 71:1743-1747.
- <sup>49</sup> Khoury G., Martin M.A., Lee T.N., Danna K.J., Nathans D. – A map of simian virus 40 transcription sites expressed in productively infected cells. J. Mol. Biol. 1973; 78:377-389.
- <sup>50</sup> Kornberg R.D. – Chromatin structure: a repeating unit of histones and DNA. Science. 1974; 184:868-871.
- <sup>51</sup> Blumberg B.S., Millman I. – Vaccine Against. Viral Hepatitis and Process, 1972 (U.S. Patent Office no 3636191).
- <sup>52</sup> Blumberg B.S., Alter H.J., Visnich S. – A “new” antigen in leukaemia sera. J. Amer. Med. Ass. 1965; 191:141
- <sup>53</sup> Millman I., Loeb L.A., Bayer M.E., Blumberg B.S. – Australia antigen (a hepatitis – associated antigen): purification and physical properties. Exp. Med. 1970; 13:1190-1199.
- <sup>54</sup> Lustbader E.D., London W.T., Blumberg B.S. – Study design for a hepatitis B vaccine trial. Proc. Natl. Acad. Sci. USA 1976; 73:955-959.
- <sup>55</sup> Maupas P., Goudeau A., Coursaget P., Drucker J., Bagros P. – Immunisation against hepatitis B in man. Lancet 1976; 1:1367-1370.
- <sup>56</sup> Szmunes W., Stevens C.E., Zang E.A., Harley E.J., Kellner A. – A controlled clinical trial of the efficacy of the hepatitis B vaccine (Hepatavax B): a final report. Hepatology 1981; 1:377-385.
- <sup>57</sup> Valenzuela P., Gray P., Quiroga M., Zaldivar J., Goodman H.M., Rutter W.J. – Nucleotide sequence of the gene coding for the major protein of hepatitis B virus surface antigen. Nature 1979; 280:815-819.
- <sup>58</sup> Galibert F., Mandart E., Fitoussi F., Tiollais P., Charnay P. – Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. Nature 1979; 281:646-650.
- <sup>59</sup> Charnay P., Gervais M., Louise A., Galibert F., Tiollais P. – Biosynthesis of hepatitis B virus surface antigen in *Escherichia coli*. Nature 1980; 286:893-895.
- <sup>60</sup> McAleer W.J., Buynak E.B., Maigetter R.Z., Wampler D.E., Miller W.J., Hilleman M.R. – Human hepatitis B vaccine from recombinant yeast. Nature 1984; 307:178-180.
- <sup>61</sup> Heijink R.A., Kruining J., Bakker M., Schalm S.W. – Immune response after vaccination with recombinant hepatitis B vaccine as compared to that after plasma-derived vaccine. Antiviral Res. 1985; suppl. 1:273-279.
- <sup>62</sup> Jilg W., Lorbeer B., Schmidt M., Wilske B., Zoulek G., Deinhardt F. – Clinical evaluation of a recombinant hepatitis B vaccine. Lancet 1984; 2:1174-1175.
- <sup>63</sup> Paoletti E., Weinberg R.L., Davis S.W., Davis M. – Genetically engineered poxviruses: a novel approach to the construction of live vaccines. Vaccine 1984; 2:204-208.
- <sup>64</sup> Morin J.E., Lubeck M.D., Barton J.E., Conley A.J., Davis A.R., Hung P.P. – Recombinant adenovirus induces antibody response to hepatitis B virus surface antigen in hamsters. Proc. Natl. Acad. Sci. USA 1987; 84:4626-4630.
- <sup>65</sup> Southern E.M. – Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Molec. Biol. 1975; 98:503-507.
- <sup>66</sup> Alwine J.C., Kemp D.J., Stark G.R. – Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. USA 1977; 74:5350-5354.
- <sup>67</sup> Stehelin D., Varmus H.E., Bishop J.M., Vogt P.K. – DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature 1976; 260:170-173.
- <sup>68</sup> Fiers W. e 11 Cols. – Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 1976; 260:500-507.

- <sup>69</sup> Sanger F, Donelson J.E., Coulson A.R., Kossel H., Fischer D. – Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage fl DNA. *Proc. Natl. Acad. Sci. USA* 1973; 70:1209-1212.
- <sup>70</sup> Sanger F, Coulson A.R. – A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Molec. Biol.* 1975; 94:441-448.
- <sup>71</sup> Sanger F, Nicklen S., Coulson A.R. – DNA sequencing with chain-terminating inhibitor. *Proc. Natl. Acad. Sci. USA* 1977; 74:5463-5467.
- <sup>72</sup> Seeburg P.H., Shine J., Martial J.A., Ullrich A., Goodman H.M., Baxter J.D. – Nucleotide sequence of a human gene coding for a polypeptide hormone. *Trans. Assoc. Am. Physicians* 1977; 90:109-116.
- <sup>73</sup> Seeburg P.H., Shine J., Martial J.A., Ivarie R.D., Morris J.A., Ullrich A., Baxter J.D., Goodman H.M. – Synthesis of growth hormone by bacteria. *Nature* 1978; 276:795-798.
- <sup>74</sup> Itakura K., Hirose T., Crea R., Riggs A.D., Heyneker H.L., Bolivar F., Boyer H.W. – Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 1977; 198; 1056-1063.
- <sup>75</sup> Goeddel D.V., Kleid D.G., Bolivar F., Heyneker H.L., Yansura D.G., Crea R., Hirose T., Kraszewski A., Itakura R., Riggs A.D. – Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. USA* 1979; 76:106-110.
- <sup>76</sup> Goeddel D.V., Heyneder H.L., Hozumi T., Arentzen R., Itakura K., Yansura D.G., Ross M.J., Miozzari G., Crea R., Seeburg P.H. – Direct expression in *Escherichia coli* of a DNA sequence coding for human growth hormone. *Nature* 1979; 281:544-548.
- <sup>77</sup> Roskam W.G., Rougeon F. – Molecular cloning and nucleotide sequence of the human growth hormone structural gene. *Nucleic Acids. Res.* 1979; 7:305-320.
- <sup>78</sup> Rougeon F., Kouridsky P., Mach B. – Insertion of a rabbit beta-globin gene sequence into an *E. coli* plasmid. *Nucleic Acid Res.* 1975; 2:2365-2378.
- <sup>79</sup> Derynck R., Renaut E., Saman E., Stanssens P., De Clercq E., Content J., Fiers W. – Expression of human fibroblast interferon gene in *Escherichia coli*. *Nature* 1980; 287:193-197.
- <sup>80</sup> Berget S.M., Moore C., Sharp P.A. – Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* 1977; 74:3171-3175.
- <sup>81</sup> Chow L.T., Gelinis R.E., Broker T.R., Roberts R.J. – An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 1977; 12:1-8.
- <sup>82</sup> Maniatis T., Hardison R.C., Lacy E., Lauer J., O'Connell C., Quon D., Sim G.K., Efstratiadis A. – The isolation of structural genes from libraries of eukaryotic DNA. *Cell* 1978; 15:687-701.
- <sup>83</sup> Gwynne P., Collings A., Gestel B. – The test tube baby. *Newsweek* 1978; Jul. 24, 92(4):76.
- <sup>84</sup> Sanger F, Air G.M., Barrel B.G., Brown N.L., Coulson A.R., Fiddes C.A., Hutchison C.A., Slocombe P.M., Smith M. – Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977; 265:687-695.
- <sup>85</sup> Maxam A.M., Gilbert W. – A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* 1977; 74:560-564.
- <sup>86</sup> Sanger F, Coulston A.R., Barrel B.G., Smith A.J., Roe B.A. – Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Molec. Biol.* 1980; 143:161-178.
- <sup>87</sup> Sanger F, Coulston A.R., Hong G.F., Hill D.F., Petersen G.B. – Nucleotide sequence of bacteriophage lambda DNA. *J. Molec. Biol.* 1982; 162:729-773.
- <sup>88</sup> Köhler G., Milstein C. – Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975; 256:495-497.
- <sup>89</sup> Parks D.R., Bryan V.M., Oi V.T., Herzerberg L.A. – Antigen-specific identification and cloning of hybridomas with a fluorescence cultivated cell sorter. *Proc. Natl. Acad. Sci. Usa* 1979; 76:1962-1966.

- <sup>90</sup> Kan Y.W., Golbus M.S., Dozy A.M. – Prenatal diagnosis of alpha-thalassemia. Clinical application of molecular hybridization. *N. Engl. J. Med.* 1976; 1165-1167.
- <sup>91</sup> Kan Y. W., Dozy A.M.- Antenatal diagnosis of sickle cell anaemia by DNA analysis of amniotic-fluid cells. *Lancet* 1978; 2:910-912.
- <sup>92</sup> Chang J.C., Kan Y.W. – Beta O thalassemia, a nonsense mutation in man. *Proc. Natl. Acad. Sci. USA* 1979; 76:2886-2889.
- <sup>93</sup> Saiki R.K., Sharf S., Faloona F., Mullis K.B., Hom G.T., Erlich H.A., Arnheim N. – Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle anemia. *Science* 1985; 230:1350-1354.
- <sup>94</sup> Gordon J.W. – Studies of foreign genes transmitted through the germ lines of transgenic mice. *J. Exp. Biol.* 1983; 228:313-324.
- <sup>95</sup> Gordon J.W., Scangos G.A., Plotkin D.J., Barbosa J.A., Ruddle F.H. – Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. USA* 1980; 77:7380-7384.
- <sup>96</sup> Roths J.B., Foxworth W.C., McArthur M.J., Montgomery C.A., Kier A.B. – Spontaneous and engineered mutant mice as models for experimental and comparative pathology: history, comparison and developmental technology. *Lab Anim. Sci.* 1999; 49:12-34.
- <sup>97</sup> Putney S.D., Herlihy N.C., Schimmel P. – A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* 1983; 302:718-721.
- <sup>98</sup> Berg P. – Dissections and reconstructions of genes and chromosomes. *Science* 1981; 213:296-303.
- <sup>99</sup> Gusella J.F., Wexler N.S., Conneally P.M., Naylor S.L., Anderson M.A., Tanzi R.E., Watkins P.C., Ottina R., Wallace M.R., Sakaguchi A.Y. et al. – A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983; 306:234-238.
- <sup>100</sup> Huntington's Disease Collaborative Research Group – A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993; 72:971-983.
- <sup>101</sup> Chakravarti A., Bultow K.H., Antonarakis S.E., Waber P.C., Boehm C.D., Kazazian H.H. – Nonuniform recombination within the human beta-globin gene cluster. *Am. J. Hum. Genet.* 1984; 36:1239-1258.
- <sup>102</sup> Yip S.P., Lovegrove J.U., Rana N.A., Hopkinson D.A., Whitehouse D.P. – Mapping recombination hotspots in human fosglucomutase (PGM1). *Hum. Molec. Genet.* 1999; 9:1699-1706.
- <sup>103</sup> Jeffreys A.J., Ritchie A., Neuman R. – High resolution of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. *Hum. Molec. Genet.* 2000; 9:725-733.
- <sup>104</sup> Jeffreys A.J., Kauppi L., Neuman R. – Intensely punctuate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* 2001; 29:217-222.
- <sup>105</sup> Patel P.I., Roa B.B., Welcher A.A., Schoener-Scott R., Trask B.J., Pentao L., Snipes G.J., Garcia C.A., Francke U., Shooter E.M., Lupski J.R., Suter U. – The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A. *Nat. Genet.* 1992; 1:159-165.
- <sup>106</sup> Reiter L.T., Murakami T., Koeuth T., Pentao L., Muzny D.M., Gibbs R.A., Lupski J.R. – A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. *Nat. Genet.* 1996; 12:288-297.
- <sup>107</sup> Monaco A.P., Neve R.L., Colleti-Feener C., Bertelson C.J., Kurnit D.M., Kunkel L.M. – Isolation of candidate cDNAs from portions of the Duchenne muscular dystrophy gene. *Nature* 1986; 323:646-650.
- <sup>108</sup> Hoffman G.P., Brown R.H.J., Kunkel L.M. – Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 1987; 51:919-928.
- <sup>109</sup> Tabin C.J., Bradley S.M., Bargmann C.I., Weinberg R.A., Papageorge A.G., Scolnick E.M., Dhar R., Lowy D.R., Chang E.H. – Mechanism of activation of a human oncogene. *Nature* 1982; 300:143-149.

- <sup>110</sup> Westin E.H., Wong-Staal F., Gelmann E.P., Dalla-Favera R., Papas T.S., Lautenberger J.A., Eva A., Reddy E.P., Tronick S.L., Aaronson S.A., Gallo R.C. – Expression of cellular homologues of retroviral onc genes in human hematopoietic cells. *Proc. Natl. Acad. Sci. USA* 1982; 79:2490-2494.
- <sup>111</sup> Seiki M., Hattori S., Hirayama Y., Yoshida M. – Human adult T-cell leukaemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA – *Proc. Natl. Acad. Sci. USA* 1983; 80:3618-3622.
- <sup>112</sup> Chen I.S., McLaughlin J., Gasson J.C., Clark S.C., Golde D.W. – Molecular characterization of genome of a novel human T-cell leukaemia virus. *Nature* 1983; 305:502-505.
- <sup>113</sup> Shaw G.M., Gonda M.A., Flickinger G.H., Hahn B.H., Gallo R.C., Wong-Staal F. – Genomes of evolutionarily divergent members of the human F-cell leukemia virus family (HTLV-I and HTLV-II) are highly conserved, especially in pX. *Proc. Natl. Acad. Sci. USA* 1984; 81:4544-4548.
- <sup>114</sup> Wain-Hobson S., Sonigo P., Danos O., Cole S., Alizon M. – Nucleotide sequence of the AIDS virus, LAV. *Cell* 1985; 40:9-17.
- <sup>115</sup> Ratner L., Haseltine W., Patarca R., Livak K.J., Starcich B., Josephs S.F., Doran E.R., Rafalski J.A., Whitehorn E.A., Baumeister K. et al. – Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 1985; 313:277-284.
- <sup>116</sup> Ratner L., Fisher A., Jagodzinski L.L., Mitsuya H., Liou R.S., Gallo R.C., Wong-Staal F. – Complete nucleotide sequences of functional clones of the AIDS virus. *AIDS Res. Hum. Retroviruses* 1987; 3:57-69.
- <sup>117</sup> Power M.D., Marx P.A., Bryant M.L., Gardner M.D., Barr P.J., Luciw P.A. – Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. *Science* 1986; 231:1567-1572.
- <sup>118</sup> Franchini G., Gurgo C., Guo H.G., Gallo R.C., Collalti E., Fargnoli K.A., Hall L.F., Wong-Staal F., Reitz M.S.Jr. – Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses. *Nature* 1987; 328:539-543.
- <sup>119</sup> Chakrabarti L., Guyader M., Alizon M., Daniel M.D., Desrosiers R.C., Tiollais P., Sonigo P. – Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature* 1987; 328:543-547.
- <sup>120</sup> Chee M.S., e 14 Cols. – Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD 169. *Curr. Top. Microbiol. Immunol.* 1990; 150:125-169
- <sup>121</sup> Dolan A. e 15 Cols. – Genetic content of wild-type human cytomegalovirus. *J. Gen. Virol.* 2004; 85:1301-1312.
- <sup>122</sup> Scriver C.R. – Treatment of inherited disease: realized and potential. *Med. Clin. North. Amer.* 1969; 53:941-963.
- <sup>123</sup> Palmer T.D., Hock R.A., Osborne W.R., Milles A.D. – Efficient retrovirus-mediated transfer and expression of a human adenosine deaminase gene in diploid skin fibroblasts from an adenosine deaminase-deficient human. *Proc. Natl. Acad. Sci. USA* 1987; 84:1055-1059.
- <sup>124</sup> Hershfield M.S., Buckley R.H., Greenberg M.L., Melton A.L., Schiff R., Hatem C., Kurtzberg J., Markert M.L., Kobayashi R.H., Kobayashi A.L., Abuchowski A. – Treatment of adenosine deaminase deficiency with polyethylene glycol-modified adenosine deaminase. *N. Engl. J. Med.* 1987; 316:589-596.
- <sup>125</sup> O'Reilly R.J., Keever C.A., Small T.N., Brochstein J. – The use of HLA-non identical T- cell-depleted marrow transplant for correction of severe combined immunodeficiency disease. *Immunodef. Rev.* 1989; 1:273-309.
- <sup>126</sup> The DNA Human Gene Therapy Protocol. *Hum. Gen. Ther.* 1990; 1:327-362.
- <sup>127</sup> Ferrari G., Rossini S., Nobili N., Maggioni D., Garofalo A., Giavazzi R., Mavilio F., Bordignon C. – Transfer of the ADA gene into human ADA – deficient T lymphocytes reconstitutes specific immune functions. *Blood* 1992; 80:1120-1124.

- <sup>128</sup> Blaese R.M. e 18 Cols. – T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years. *Science* 1995; 270:475-480.
- <sup>129</sup> Bordignon C. e 11 Cols. – Gene therapy in peripheral blood lymphocytes and bone marrow for ADA-immunodeficient patients. *Science* 1995; 270:470-475.
- <sup>130</sup> Creighton H.B., McClintock B. – The correlation of cytological and genetic crossing-over in *Zea mays*. *Proc. Natl. Acad. Sci. USA* 1931; 17:492-497.
- <sup>131</sup> Pennisi E. – How the genome readies itself for evolution. *Science* 1998; 281:1131, 1933-1134.
- <sup>132</sup> Chicurel M. – Genetics. Can organisms speed their own evolution? *Science* 2001; 292:1824-1827.
- <sup>133</sup> Pennisi E. – Genomics. Charting a genome's hills and valleys. *Science* 2002; 296:1601-1603.
- <sup>134</sup> Coe E., Kass L.B. – Proof of physical exchange of genes on the chromosomes. *Proc. Natl. Acad. Sci. USA* 2005; 102:6641-6646.
- <sup>135</sup> McGinnis W., Levine M.S., Hafen E., Kuroiwa A., Gehring W.J. – A conserved DNA sequence in homoeotic genes of the *Drosophila antennapedia* and *bithorax* complexes. *Nature* 1984; 308:428-433.
- <sup>136</sup> Ronshaugen M., McGinnis N., McGinnis W. – Hox protein mutation and macroevolution of the insect body plan. *Nature* 2002; 415:914-917.
- <sup>137</sup> Tour E., Hittinger C.T., McGinnis W. – Evolutionary conserved domains required for activation and repression functions of the *Drosophila* hox protein ultrabithorax. *Development* 2005; 132:5271-5281
- <sup>138</sup> Mullis K., Faloona F., Scharf S., Saiki R., Horn G., Erlich H. – Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 1986; 51 (pt 1):263-273.
- <sup>139</sup> Jeffreys A.J., Wilson V., Thein S.L. – Individual-specific "fingerprints" of human DNA. *Nature* 1985; 316:76-79.
- <sup>140</sup> Jeffreys A.J. – Genetic fingerprinting. *Nat. Med.* 2005; 11:1035-1039.
- <sup>141</sup> Bale S.J., Dracopoli N.C., Tucker M.A., Clark W.H.Jr., Fraser M.S., Stanger B.Z., Green P., Donis-Keller H., Housman D.E., Greene M.H. – Mapping the gene for hereditary cutaneous malignant melanoma-dysplastic nevus to chromosome 1p. *N. Engl. J. Med.* 1989; 320:1367-1372.
- <sup>142</sup> Rommens J.M. et 14 Cols. – Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 1989; 245:1059-1065.
- <sup>143</sup> White R., Woodward S., Leppert M., O'Connell P., Hoff M., Herbst J., Lalouel J.M., Dean M., Vande Woude G. – A closely linked genetic marker for cystic fibrosis. *Nature* 1985; 318:382-384.
- <sup>144</sup> Tsui L.-C., Rommens J.M., Burns J., Zengerling S., Riordan J.R., Carlock L.R., Grgeschik K-H, Buchwald M. – Progress towards cloning the cystic fibrosis gene. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 1988; 319:263-273.
- <sup>145</sup> Buchwald M., Tsui L-C, Riordan J.R. – The search for the cystic fibrosis gene. *Am. J. Physiol.* 1989; 257:L47-L52.
- <sup>146</sup> Collis F.S. – Cystic fibrosis molecular, biology and therapeutic implications. *Science* 1992; 256:774-779.
- <sup>147</sup> Davies K., Williamson B. – Gene therapy begins. *Br. Med. J.* 1993; 306:1625-1626
- <sup>148</sup> Thomas K. – Yeasties and beasties: 7 years of genome sequencing. *FEBS Lett* 1996; 396:1-6.
- <sup>149</sup> Goffeau A. – Yeast. Genes in search of functions. *Nature* 1994; 369:101-102.
- <sup>150</sup> Roberts L. – Large-scale sequencing trials beging. *Science* 1990; 250:1336-1338.
- <sup>151</sup> Strauss E.C., Kobori J.A., Siu G., Hood L.E. – Specific primer - directed DNA sequencing. *Anal. Biochem.* 1986; 154:353-360.
- <sup>152</sup> Gocayne J., Robinson D.A., FitzGerald M.G., Chung F.-Z., Kerlavage A.R., Lentes K.-U., Lai J., Wang C.-D., Fraser C.M., Venter J.C. – Primary structure of rat cardiac b-adrenergic and muscarinic cholinergic receptors obtained by automated DNA sequence analysis: further evidence for a multigene family. *Proc. Natl. Acad. Sci. USA* 1987; 84:8296-8300.

- <sup>153</sup> Burke D.T., Carle G.F., Olson M.V. – Cloning of large segment of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 1987; 236:806-812.
- <sup>154</sup> Oliver S.G. e 146 Cols. – The complete DNA sequence of yeast chromosome III. *Nature* 1992; 357:38-46.
- <sup>155</sup> Feldmann H. e 96 Cols. – Complete DNA sequence of yeast chromosome II. *EMBO J.* 1994; 13:5795-5809.
- <sup>156</sup> Johnston M. e 34 Cols. – Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. *Science* 1994; 265:2077-2082.
- <sup>157</sup> Dujon B. e 107 Cols. – Complete DNA sequence of yeast chromosome XI. *Nature* 1994; 369:371-378.
- <sup>158</sup> Goffeau A, e 15 Cols. – Life with 6000 genes. *Science* 1996; 274:546-563-567.
- <sup>159</sup> Wood V. e 127 Cols. – The genome sequence of *Schizocaccharomyces pombe*. *Nature* 2002; 415:871-880.
- <sup>160</sup> Fleischmann R.D., e 39 Cols. – Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269:496-512.
- <sup>161</sup> Fraser C.M. e 28 cols. – The minimal gene complement of *Mycoplasma genitalium*. *Science* 1995; 270:397-403.
- <sup>162</sup> Daniels D.L., Plunkett G.3rd, Burland V., Blattner F.R. – Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. *Science* 1992; 257:771-778.
- <sup>163</sup> Blattner F.R., e 16 Cols. – The complete genome sequence of *Escherichia coli* K-12. *Science* 1997; 277:1453-1474.
- <sup>164</sup> Lim A., e 18 Cols. – Shotgun optical maps of the whole *Escherichia coli* O157:H7 genome. *Genoma Res.* 2001; 11:1584-1593
- <sup>165</sup> Welch R.A., e 18 Cols. – Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 2002; 99:17020-17024.
- <sup>166</sup> Cole S.T., e 40 Cols. – Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998; 393:537-544.
- <sup>167</sup> Fraser C.M. e 32 Cols. – Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* 1998; 281:375-388.
- <sup>168</sup> Fraser C.M. e 37 Cols. – Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 1997; 390:580-586.
- <sup>169</sup> Carucci D.J., Gardner M.J., Tettelin H., Cummings L.M., Smith H.O., Adams M.D., Hoffman S.L., Venter J.C. – The malaria genome sequencing project. *Exp. Rev. Molec. Med.* 1998; 1:1-9.
- <sup>170</sup> Gardner M.J. e 44 Cols. – Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002; 419:498-511
- <sup>171</sup> Carlton J.M. e 43 Cols. – Genome sequence and comparative analysis of the model rodent malaria parasite. *Plasmodium yoelii yoelii*. *Nature* 2002; 419:512-519.
- <sup>172</sup> Holt R.A. e 123 Cols. – The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 2002; 298:129-149.
- <sup>173</sup> Berriman M. e 101 Cols. – The genome of the African trypanosome *Trypanosoma brucei*. *Science* 2005; 309:416-422.
- <sup>174</sup> Parkhill J. e 27 Cols. – Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 2000; 404:502-506.
- <sup>175</sup> Heidelberg J.F. e 31 Cols. – DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 2000; 406:477-483.
- <sup>176</sup> Horvitz H.R., Sulston J.E. – Joy of the worm”. *Genetics.* 1990; 126:287-292.
- <sup>177</sup> Sulston J. e 18 Cols. – The *C. elegans* genome sequencing project: a beginning. *Nature* 1992; 356:37-41.
- <sup>178</sup> Williams B.D., Schrank B., Huynh C., Shownkeen R., Waterston R.H. – A genetic mapping – system in *Caenorhabditis elegans* based on polymorphic sequence – tagged lines. *Genetics* 1992; 131:609-624.

- <sup>179</sup> Barstead R.J., Kleiman L., Waterston R.H. - Cloning, sequencing, and mapping of an alpha-actinin gene from the nematode *Caenorhabditis elegans*. *Cell Motil Cytoskeleton*. 1991;20:69-78.
- <sup>180</sup> C. elegans Sequencing Consortium – Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 1998; 282:2012-2018.
- <sup>181</sup> Chalfie M., Jorgensen E.M. – *C. elegans* neuroscience: genetics to genome. *Trends Genet* 1998; 14:506-512.
- <sup>182</sup> Ellis R.E., Jacobson D.M., Horvitz H.R. – Genes required for the engulfment of cell corpses during programmed cell death in *Caenorhabditis elegans*. *Genetics* 1991; 129:79-94.
- <sup>183</sup> Hillier L.W., Coulson A., Murray J.I., Bao Z., Sulston J.E., Waterston R.H. – Genomics in *C. elegans*: so many genes, such a little worm. *Genome Res*. 2005; 15:1651-1660.
- <sup>184</sup> Brenner S. – The genetics of *Caenorhabditis elegans*. *Genetics* 1974; 77:71-94.
- <sup>185</sup> MacLeod A.R., Karn J., Brenner S. – Molecular analysis of the unc-54 myosin heavy-chain gene of *Caenorhabditis elegans*. *Nature* 1981; 291:386-390.
- <sup>186</sup> FlyBase Consortium – FlyBase- the *Drosophila* database. *Nucleic Acids Res*. 1994; 22:3456-3458.
- <sup>187</sup> Adams M.D. e 193 Cols. – The genome sequence of *Drosophila melanogaster*. *Science* 2000; 287:2185-2195.
- <sup>188</sup> Meyers E.W. e 28 Cols. – A whole genome assembly of *Drosophila*. *Science* 2000; 287:2196-2204.
- <sup>189</sup> Garza D., Ajioka J.W., Burke D.T., Hartl D.L. – Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science* 1989; 246:641-646.
- <sup>190</sup> Merriam J., Ashburner M., Hartl D.L., Kafatos F.C. – Toward cloning and mapping the genome of *Drosophila*. *Science* 1991; 254:221-225.
- <sup>191</sup> Peter A. e 53 Cols. – Mapping and identification of essential gene functions on the X chromosome of *Drosophila*. *EMBO Rep*. 2002; 3:34-38.
- <sup>192</sup> Arabidopsis Genome Initiative – Analysis of the genome sequence of the flowering plant *Arabidopsis Thaliana*. *Nature* 2000; 408:796-815.
- <sup>193</sup> International Rice Genome Sequencing Project – The map-based sequence of the rice genome. *Nature* 2005; 436:793-800.
- <sup>194</sup> Watson J.D., Cook-Deegan R.M. – Origins of the Human Genome Project. *FASEB J*. 1991; 5:8-11.
- <sup>195</sup> Bodmer W.F. – HUGO: the Human Genome Organization. *FASEB J*. 1991; 5:73-74.
- <sup>196</sup> Palca J. – James Watson to head NIH human genome project. *Nature* 1988; 335:193.
- <sup>197</sup> Donis-Keller H., Green P., Helms C., Cartinhour S., Weiffenbach B., Stephens K., Keith T.P., Browden D.W., Smith D.R., Lander E.S., et al – A genetic linkage map of the human genome. *Cell* 1987; 51:319-337.
- <sup>198</sup> Schwartz D.C., Cantor C.R. – Separation of yeast chromosome-sized DNAs by pulsed field gel electrophoresis. *Cell* 1984; 37:67-75.
- <sup>199</sup> Venter J.C., Adams M.D., Sutton G.G., Kerlavage A.R., Smith H.O., Hunkapiller M. – Shotgun sequencing of the human genome. *Science* 1998; 280:1540-1542.
- <sup>200</sup> Marshall E. – A strategy for sequencing the genome 5 years early. *Science* 1995; 267:783-784.
- <sup>201</sup> Marshall E. – Human genome project. Emphasis turns from mapping to large-scale sequencing. *Science* 1995; 268:1270-1271.
- <sup>202</sup> Hudson T.J., e 49 Cols. – An STS-based map of the human genome. *Science* 1995; 270:1945-1954.
- <sup>203</sup> Deloukas P., e 64 Cols. – A physical map of 30,000 human genes. *Science* 1998; 282:744-746.
- <sup>204</sup> Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., Kerlavage A.J., McCombie W.R., Venter J.C. – Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991; 252:1651-1656.

- <sup>205</sup> Adams M.D., Dubnick M., Kerlavage A.R., Moreno R., Kelley J.M., Utterback T.R., Nagle J.W., Fields C., Venter J.C. – Sequence identification of 2,375 human brain genes. *Nature* 1992; 355:632-634.
- <sup>206</sup> White O., Dunning T., Sutton G., Adams M., Venter J.C., Fields C. – A quality control algorithm for DNA sequencing projects. *Nucleic Acid. Res.* 1993; 21:3829-3838.
- <sup>207</sup> Schmitt H., Kim, U.J., Slepak T., Blin N., Simon M.I., Shizuya H. – Framework for a physical map of the human 22q13 region using bacterial artificial chromosomes (BACs). *Genomics* 1996; 33:9-20
- <sup>208</sup> Weber J.L., Myers E.W. – Human whole-genome shotgun sequencing. *Genome* 1997; 7:401-409.
- <sup>209</sup> Venter J.C., Adams M.D., Sutton G.J., Kerlavage A.R., Smith H.O., Hunkapiller M. – Shotgun sequencing of the human genome. *Genomics* 1998; 280:1540-1542.
- <sup>210</sup> Collins F.S., Patrinos A., Jordan E., Chakravarti A., Gesteland R., Walters L.R. – New goals for the US Human Genome Project 1998-2003. *Science* 1998; 282:682-689.
- <sup>211</sup> Dunham I. e 24 Cols. – The DNA sequence of human chromosome 22. *Nature* 1999; 402:489-495.
- <sup>212</sup> Hattori M., e 64 Cols. – The chromosome 21 mapping and sequencing consortium. The DNA sequence of human chromosome 21. *Nature* 2000; 405:311-319.
- <sup>213</sup> Tilford C.A. e 11 Cols. – A physical map of the human Y chromosome. *Nature* 2001; 409:943-945.
- <sup>214</sup> Skaletsky H. e 38 Cols. – The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003; 423:825-837.
- <sup>215</sup> Bentley D.R. e 100 Cols. – The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* 2001; 409:942-943.
- <sup>216</sup> Deloukas P. e 126 Cols. – The DNA sequence and comparative analysis of human chromosome 20. *Nature* 2001; 414:865-871.
- <sup>217</sup> Scherer S.W. e 89 Cols. – Human chromosome 7: DNA sequence and biology. *Science* 2003; 300:767-772.
- <sup>218</sup> Hillier L.W. e 106 Cols. – The DNA sequence of human chromosome 7. *Nature* 2003; 424:157-164.
- <sup>219</sup> Scherer S.W., Green E.D. – Human chromosome 7 circa 2004: a model for structure and functional studies of the human genome. *Hum. Molec. Genet.* 2004; 13:R303-R313.
- <sup>220</sup> McPherson J.D. e 112 Cols. – A physical map of the human genome. *Nature* 2001; 409:934-941.
- <sup>221</sup> Lander E.S. e 254 Cols. – Initial sequencing and analysis of the human genome. *Nature* 2001; 409:860-921.
- <sup>222</sup> Sachidanandam R. e 40 Cols. – A map of human genome sequence variation containing 1,42 million single nucleotide polymorphisms. *Nature* 2001; 409:928-933.
- <sup>223</sup> Venter J.C. e 274 Cols. – The sequence of the human genome. *Science* 2001; 291:1304-1351.
- <sup>224</sup> Yeh R-F, Lim L.P., Burge C. – Computational inference of homologous gene structures in the human genome. *Genome Res.* 2001; 11:803-816.
- <sup>225</sup> Harrison P.M., Hegyi H., Balasubramanian S., Luscombe N.M., Bertone P., Echols N., Johnson T., Gerstein M. – Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res.* 2002; 12:272-280.
- <sup>226</sup> Rubin G.M., e 54 Cols. – Comparative genomics of the eukaryotes. *Science* 2000; 287:2204-2215.
- <sup>227</sup> von Sternberg R – On the roles of repetitive DNA elements in the context of a unified genome-epigenetic system. *Ann. N.Y. Acad. Sci.* 2002; 981:154-188.
- <sup>228</sup> Vinogradov A.E. – Evolution of genome size: multilevel selection, mutation bias or dynamical chaos? *Curr. Opin. Genet. Dev.* 2004; 14:620-626.
- <sup>229</sup> Lynch M., Conery J.S. – The origin of genome complexity. *Science* 2003; 302:1401-1404.
- <sup>230</sup> de Parseval N., Heidmann T. – Human endogenous retroviruses: from infectious elements to human genes. *Cytogenet Genome Res.* 2005; 110:318-332.

- <sup>231</sup> Blair Hedges S., Kumar S. – Vertebrate genomes compared. *Science* 2002; 297:1283-1285.
- <sup>232</sup> Chimpanzee Sequencing and Analysis Consortium – Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 2005; 437:69-87.
- <sup>233</sup> Winckler W., Myers S.R., Richter D.J., Onofrio R.C., McDonald G.J., Bontrop R.E., McVean G.A., Gabriel S.B., Reich D., Donnelly P., Altshuler D. – Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 2005; 308:107-111.
- <sup>234</sup> King M.C., Wilson A.C. – Evolution at two levels in human and chimpanzees. *Science* 1975; 188:107-116.
- <sup>235</sup> Thomas J.W. e 70 Cols – Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* 2003; 424:788-793.
- <sup>236</sup> Guyon R. e 16 Cols. A 1-Mb resolution radiation hybrid map of the canine genome. *Proc. Natl. Acad. Sci. USA* 2003; 100:5296-5301.
- <sup>237</sup> Lindblad-Toh K. e 236 Cols. – Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 2005; 438:803-819.
- <sup>238</sup> International Human Genome Sequencing Consortium. – Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431:931-945.
- <sup>239</sup> Schmutz J. e 24 Cols. – Quality assessment of the human genome sequence. *Nature* 2004; 429:365-368.
- <sup>240</sup> Blakesley R.W. e 21 Cols; NISC Comparative Sequencing Program - An intermediate grade of finished genomic sequence suitable for comparative analyses. *Genome Res.* 2004; 14:2235-2244.
- <sup>241</sup> Palmer L.E., McCombie W.R. – On the importance of being finished. *Genome Biol.* 2002; 3:2010.1-2010.4.
- <sup>242</sup> Aparicio S. e 40 Cols. – Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 2002; 297:1301-1310.
- <sup>243</sup> Waterson R.H. e 218 Cols; Mouse Genome Sequencing Consortium – Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; 420:520-562.
- <sup>244</sup> Gibbs R.A. e 228 Cols. – Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 2004; 428:475-476.
- <sup>245</sup> Desiere F., Deutsch E.W., King N.L., Nesvizhskii A.I., Mallick P., Eng J., Chen S., Eddes J., Loevenich S.N., Aebersold R. – The Peptide Atlas project – *Nucleic Ac. Res.* 2006; 34:D655-D658.



J. MARTINS E SILVA

# **Bioquímica da Informação Genética**

**(Texto de apoio à disciplina de Química Fisiológica,  
no ano lectivo de 1972-1973)**

Faculdade de Medicina da Universidade de Lourenço Marques,  
Moçambique, 1973



# CAPÍTULO 1

## CROMOSSOMAS, GENES E HEREDITARIEDADE

Entre as propriedades fundamentais dos organismos vivos sobressai a capacidade de reprodução, pela qual cada ser pode multiplicar-se perpetuando a espécie através de gerações sucessivas de descendentes semelhantes.

### Conservação e transmissão da informação genética

A manutenção das características próprias de cada espécie é regulada por informações presentes em todas as células de cada organismo, transmitidas às células-filha por divisão celular. Este processo, que constitui o esquema básico da regulação dos seres vivos, traduz-se na formação de duas células idênticas à célula-mãe original, quer do ponto de vista morfológico ou funcional. A informação contida na célula-mãe constitui o *material genético* transmitido às células-filha.

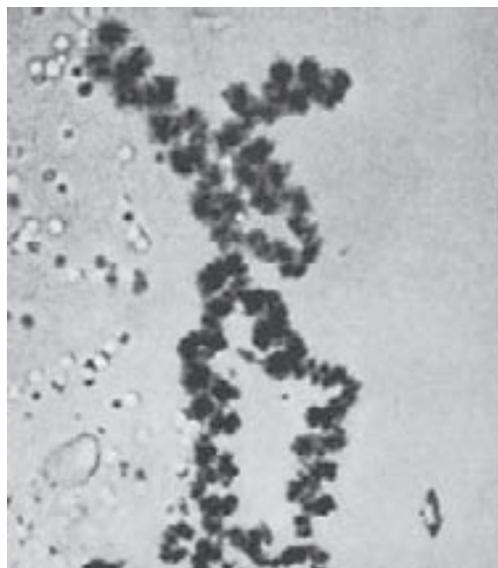
A identidade entre os indivíduos da mesma espécie não é, contudo, absoluta. Existem pequenas diferenças (morfológicas, funcionais ou de comportamento), atribuíveis também a esse material genético, que asseguram a conservação das características fundamentais da espécie. Essas diferenças entre os indivíduos da mesma espécie desenvolvem-se ao longo de gerações sucessivas, em consequência das modificações fortuitas e raras do material genético. Logo que uma nova característica aparece, independentemente do meio, passa a ser transmitida aos descendentes viáveis.

A expressão de caracteres diferentes dos iniciais resulta de modificações pontuais do material genético, transmitidas também hereditariamente. Desig-

nam-se por *mutações* as alterações da informação inicial. O aparecimento de uma mutação não influencia a restante carga genética. Mutações que afectam caracteres distintos são, assim, probabilidades independentes. A característica resultante de uma mutação poderá ser reversível, tão bruscamente como foi originada, com retorno à sua expressão inicial quando o sector alterado recupera a estrutura original.

As mutações tendem a acumular-se ao longo das gerações. A frequência das mutações varia com os organismos em que ocorre, ainda que seja sempre bastante rara. Se as mutações fossem frequentes conduziriam rapidamente a alterações profundas na espécie. Explica-se assim que as mutações viáveis constituam uma vantagem biológica, assegurando a variabilidade genética necessária à adaptação das espécies ao meio que vivem (vide Cap. III).

A expressão de cada uma das características celulares depende dos segmentos específicos do material genético, que são os *genes*. É a alteração isolada de cada um desses sectores que origina uma mutação. Os genes são, portanto, a mais pequena fracção de material genético necessário para a expressão de características específicas. O conjunto dos genes celulares constitui o *genoma* (ou genómio). Na divisão celular cada célula filha recebe uma cópia do genoma inicial. A existência de mutações pressupõe que cada gene possa ocorrer sob formas diferentes – *alelos* – com expressão variável.



**Fig. 1.1.** Cromossoma com o respectivo par de cromátidos enroscados entre si (4000 diâmetros, aproximadamente). Microfotografia amavelmente cedida pelo Dr. J. Herbert Taylor (Institute for Molecular Biophysics of Florida State University, Tallahassee, USA)

Designa-se por *genótipo* a constituição genética do organismo, correspondendo o *fenótipo* à sua aparência física (expressão). O fenótipo resulta da interacção entre o genótipo e o ambiente em que o organismo vive. O genótipo encontra-se em evolução constante, através de mutações ocasionais.

A genética actual fundamenta-se em estudos com microorganismos, que oferecem uma dupla vantagem: são haploides e multiplicam-se com extraordinária rapidez. As leveduras, fungos, bactérias e vírus constituem, assim, o material genético mais utilizado nesses estudos. Têm o inconveniente de se reproduzirem de forma diferente à dos seres superiores. Contudo, a substância química que constitui o material genético é, na maior parte dos microorganismos, idêntica à dos organismos mais evoluídos, em que se inclui o homem. Essa substância – o *ácido desoxirribonucleico (DNA)*<sup>1</sup> – é o principal constituinte dos cromossomas celulares, os quais são os portadores da informação genética.

## Cromossomas

Os *cromossomas*<sup>2</sup> (Fig. 1.1.), são constituintes próprios dos núcleos (zonas nucleares) de todas as células. Também foram individualizados cromossomas nas mitocôndrias, com funções ainda por esclarecer.

Cada célula contém um ou mais cromossomas, sendo o seu número fixo para cada espécie (Quadro 1.1.). Os *procariontes* (organismos unicelulares sem membrana nuclear nem outras membranas internas) por conterem um cromossoma desemparelhado, têm também a designação de células *haploides*.

Os *eucariotes* (organismos com núcleo celular bem demarcado por uma membrana, entre outras particularidades) têm bastante mais cromossomas, em número par. Nas células somáticas dos organismos superiores existem duas

---

<sup>1</sup> Faz-se uso no texto às abreviaturas da nomenclatura anglo-saxónica e internacional. Por exemplo, o acrónimo ADN, da nomenclatura portuguesa do ácido desoxirribonucleico é no texto preterido por DNA (desoxirribonucleic acid). Idêntica opção foi seguida para o ácido ribonucleico (RNA em lugar de ARN), e assim sucessivamente.

<sup>2</sup> O termo “cromossoma” advém da propriedade de um segmento (do grego *soma*, parte ou totalidade do corpo) tomar cor (do grego *khroma*) por acção de substâncias corantes. Os cromossomas tornam-se visíveis ao microscópio um pouco antes de cada divisão celular. Nesse instante, o cromossoma apresenta-se como um par de bastonetes emparelhados. Quando a célula se divide, cada um desses bastonetes, designados *cromátídeos*, separa-se para uma das células-filha.

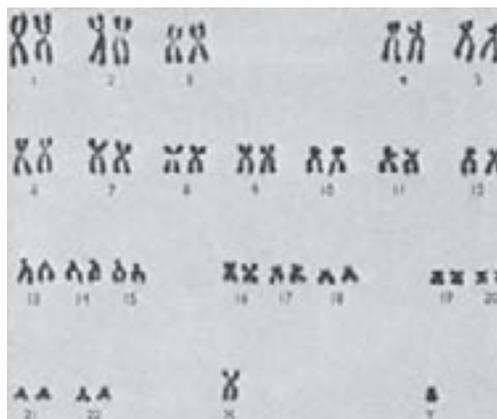
cópias de cada tipo de cromossomas – *cromossomas homólogos* – um fornecido pelo progenitor macho outro pela fêmea, de que resulta a *diploidia*. Nas células sexuais existe apenas um cromossoma de cada tipo (*haploidia*).

A constituição dos cromossomas dos procariotas é diferente da observada nos eucariotas. Naqueles, os cromossomas contêm unicamente ácidos nucleicos, em geral DNA. Contudo, alguns vírus têm como material genético o ácido ribonucleico (RNA) em lugar do DNA.

**Quadro 1.1.** Número de cromossomas em algumas espécies.

<b>Organismos procariotas (haploides)</b>	
Bactérias	1
<b>Organismos eucariotas (diploides)</b>	
Drosófila	8
Rã	26
Gato	38
Murganho	40
Rato	42
Coelho	44
Homem	46

Cada espécie (e indivíduo) possui um determinado padrão ou “carta cromossômica”, genericamente designada de cariótipo, o qual caracteriza o número, forma, dimensões e disposição dos cromossomas do núcleo celular (Fig. 1.2.).



**Fig. 1.2.** Cariótipo de cromossomas humanos na metafase. Os 46 cromossomas estão separados em pares homólogos numerados (ampliação: 1800 x) Microfotografia amavelmente cedida pelo Dr. M. Nascimento Ferreira (Laboratório de Histologia da Faculdade de Medicina da Universidade de Lourenço Marques, Moçambique).

No homem existem 46 cromossomas em cada núcleo das células somáticas, que se agrupam por características (morfológicas e estruturais) homólogas. Exceptuam-se neste caso os cromossomas sexuais. Os 44 cromossomas comuns aos dois sexos, são designados de *autossomas*. O par restante constitui os *heterossomas* ou, *cromossomas sexuais*, (XX no sexo feminino e XY no masculino). As células somáticas, que possuem os 23 pares de cromossomas, são diploides (2n), enquanto os gâmetas, que contêm apenas um cromossoma de cada um daqueles pares, são haploides (n).

Os cromossomas dos eucariotas são constituídos por *cromatina*<sup>3</sup>. O DNA e proteínas básicas (histonas), em proporções sensivelmente iguais, são os principais componentes da cromatina, a que acresce uma pequena fracção de RNA, proteínas ácidas, algumas enzimas e lípidos.

Cada cromossoma contém uma enorme molécula de DNA. Calcula-se que o comprimento total do DNA do bacteriófago T<sub>2</sub> seja de 50mm, enquanto o DNA de uma única célula de mamífero terá cerca de dois metros de comprimento e 2nm de diâmetro. As histonas são essenciais para a estrutura e função do DNA. Mantêm-lhe o enrolamento tridimensional e condicionam a transmissão da informação genética, ao inactivarem temporariamente determinados segmentos do DNA.

## Ciclo celular e tipos de divisão celular

Todas as células provêm de unidades idênticas por divisão celular. Neste processo cada célula origina duas células semelhantes. Nos eucariotas, o núcleo da célula-mãe subdivide-se em núcleos iguais, destinados às duas células-filha. Cada núcleo contém uma quantidade constante de cromossomas que, antes da divisão celular (na *interface*), se subdividem longitudinalmente em duas metades idênticas ao original, designadas *cromatídeos*. Este processo du-

---

<sup>3</sup> A cromatina apresenta-se no núcleo interfásico sob a forma condensada (heterocromatina, que fixa fortemente a coloração específica) ou difusa (eucromatina); ambas as formas coexistem no mesmo núcleo. Em microscopia electrónica, a cromatina apresenta-se como um filamento (com diâmetro entre 300 e 500 Å) dobrado muitas vezes sobre si próprio e com enrolamento helicoidal (já visível por microscopia óptica). Na interfase, o número de filamentos cromatínicos é igual ao número de cromossomas. Na metafase, o número de filamentos de cromatina por cromossoma duplica, devido à replicação do DNA, de que resultam dois cromatídeos por cromossoma unidos entre si a nível do centrómero.

plica o número de cromossomas nucleares, destinando-se metade de cada par às duas novas células.

A organização ultraestrutural do DNA no núcleo das células eucariotas varia consideravelmente durante o ciclo celular. Entretanto, também é muito variável o período de crescimento e divisão celular nos diferentes eucariotas. Enquanto as células de levedura, p.ex, se dividem em cada 120 minutos, a maior parte das células de animais e plantas superiores demora entre 10 a 20 horas, ou mais, a duplicarem-se. Alguns tipos celulares deixam de se dividir quando atingem o estado adulto, como é o caso dos miócitos do músculo estriado e as células nervosas.

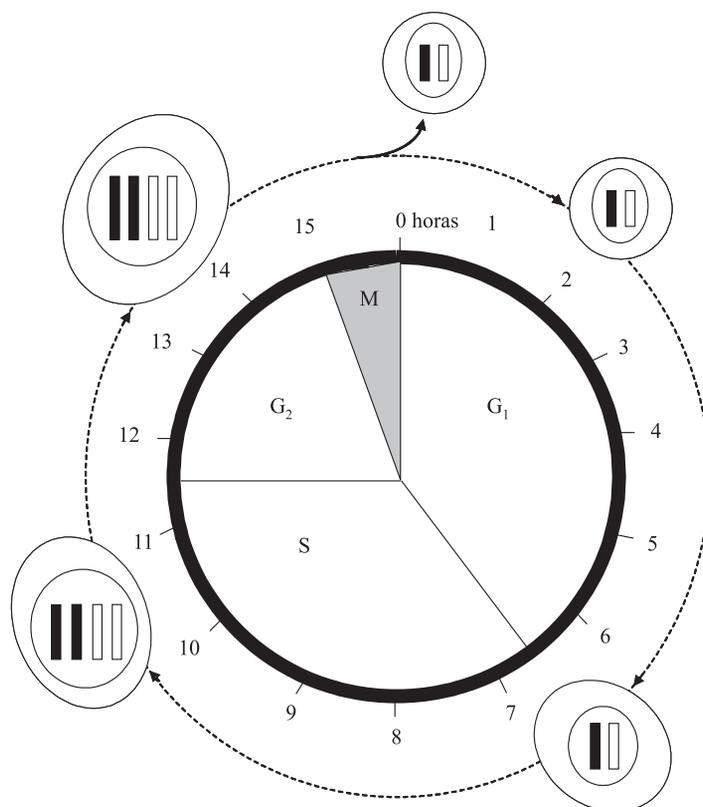
Nas células procariotas o genoma é constituído por um único cromossoma circular cujo componente é uma molécula de ácido desoxirribonucleico. A replicação desta molécula de DNA é seguida imediatamente pela divisão celular (que, por sua vez, demora muito pouco mais que o tempo requerido para a síntese de DNA). A divisão é concretizada pela formação de um septo a meio da célula original, de que resultam duas metades iguais, cada uma das quais contendo uma réplica do cromossoma inicial.

## Mitose

Por princípio, a divisão mitótica (*mitose*) das células eucariotas traduz-se na produção de duas células filhas com igual conteúdo genético. Em contraste com as células procariotas, nos eucariotas a replicação de DNA verifica-se muito antes da divisão celular.

Pressupondo que o ciclo celular (Fig. 1.3.) de uma célula de mamífero decorre em 16h, o período de *divisão celular* (fase M, de mitose) abrange os 60 minutos finais, enquanto o período maior (com cerca de 15 horas) é constituído pela fase  $G_1$  (intercalar), fase S (síntese de DNA) e fase  $G_2$  (intercalar). As fases  $G_1$ , S e  $G_2$  correspondem ao período de interfase. Por conseguinte, entre a síntese de DNA (fase S) e a fase M (mitose) existe um intervalo ( $G_2$ ), observando-se outro intervalo após a síntese e antes de nova replicação de DNA.

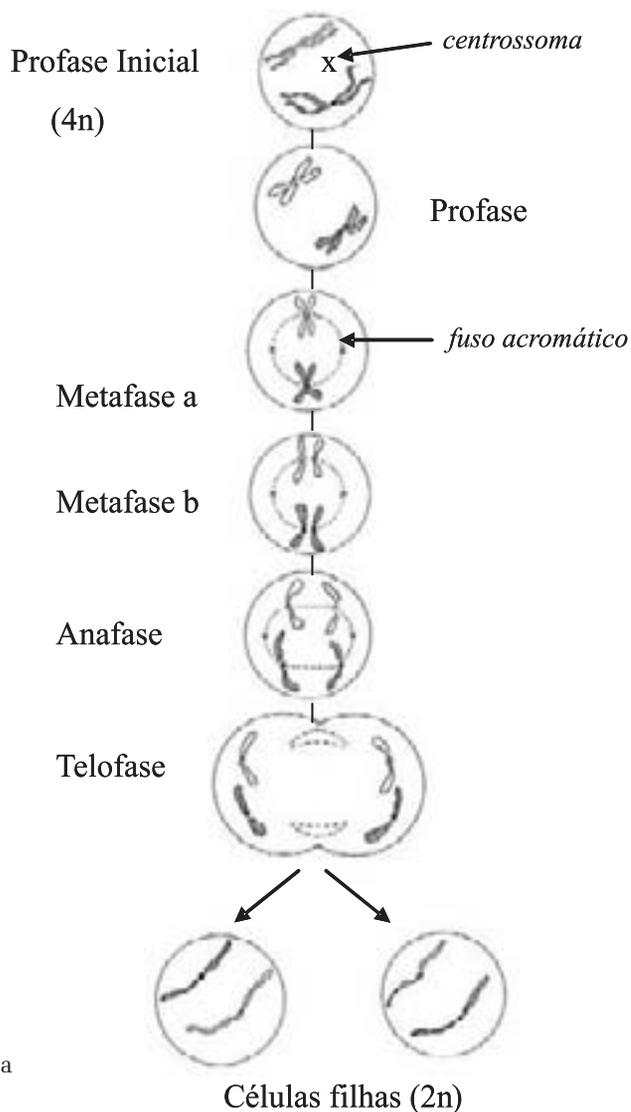
Determinadas células, como os fibroblastos, dividem-se somente quando a sua multiplicação é necessária (p.ex., na cicatrização de feridas). Normalmente, mantêm-se em fase quiescente após a mitose, sem desenvolvimento da fase S.



**Fig. 1.3.** Fase do ciclo celular numa célula hipotética de eucariota, que cresce e se divide uma vez em cada 16h. As divisões celulares de todos os organismos vivos caracterizam-se pela formação de duas células filhas de cada célula-mãe no termo de cada ciclo celular, a que se segue de imediato a cinocinese. O ciclo celular da maior parte dos eucariotas inclui duas principais fases: fase M (mitose) e interfase. Esta fase abrange três períodos:  $G_1$  (intercalar entre a mitose antecedente e a replicação seguinte do DNA), S (**S**, da **s**íntese do DNA) e  $G_2$  (intercalar entre a síntese de DNA e a mitose). Em geral, a duração da mitose é de cerca de uma hora, sendo o restante tempo ocupado pela interfase; nesta existe grande variação, de tipo para tipo celular, na duração dos respectivos componentes; com frequência, as fases  $G_1$  e S duram entre 8 a 10h.

Nos períodos intermitose (*interfase*) a cromatina dispõe-se numa rede difusa de densidade variável dificilmente observável por microscópio óptico, enquanto se distingue bem o nucléolo (um ou vários). Na fase S ocorre a replicação do DNA, originando duas cópias por cromossoma, pelo que o núcleo celular na fase  $G_2$  (pré-mitótica), apresenta o estado tetraploide ( $4n$ ).

A mitose (fase M) evolui em quatro etapas principais (Fig. 1.4.). Na primeira – *profase* – cada cromossoma torna-se visível sob a forma de dois filamentos curtos e espessos, os cromátídeos, cada um dos quais contém uma das duas moléculas-filha de DNA sintetizadas na fase S. Os cromátídeos permanecem unidos aos pares, a nível do centrómero. Este sector, com localização específica em cada tipo de cromossomas, controla-lhes os movimentos durante a divisão celular. No termo



**Fig. 1.4.** Representação esquemática das principais fases da mitose

da profase sucede a fragmentação da membrana do nuclear e o desaparecimento do nucléolo, com formação de uma nova estrutura – fibras contínuas – que, a partir do centrómero de cada cromossoma, se alonga para os centrossomas polares.

Na segunda etapa (*metafase*) cada célula ainda está no estado  $4n$ ; os pares de cromátídeos dispõem-se na região equatorial da célula próximos entre si, e esboça-se a formação de um fuso acromático. Este é constituído por fibras finas e extensas (microtúbulos e proteínas) que orientam os cromátídeos para os polos das células, onde se ligam ao respectivo centrossoma. Os polos são as regiões envolventes dos centrossomas<sup>4</sup> donde irradiam os microtúbulos (fibras aterianas e fibras contínuas, estas unindo dois centrossomas). O centrómero divide-se ainda na metafase.

A divisão do centrómero transforma os cromátídeos em cromossomas, dando início à terceira etapa – *anafase*. Nesta etapa, cada um dos cromátídeos dos pares originais desloca-se para polos opostos das células.

Por fim, na última etapa – *telofase* – os cromossomas, já perfeitamente definidos e distintos, são envolvidos por uma membrana nuclear, o que conclui a formação das duas células filhas com estrutura e composição iguais às da célula mãe, e a sua imediata separação (*citocinese*). De imediato, as células filhas ( $2n$ ) ingressam no período  $G_1$  da interfase, com subsequente repetição do processo até à divisão celular seguinte.

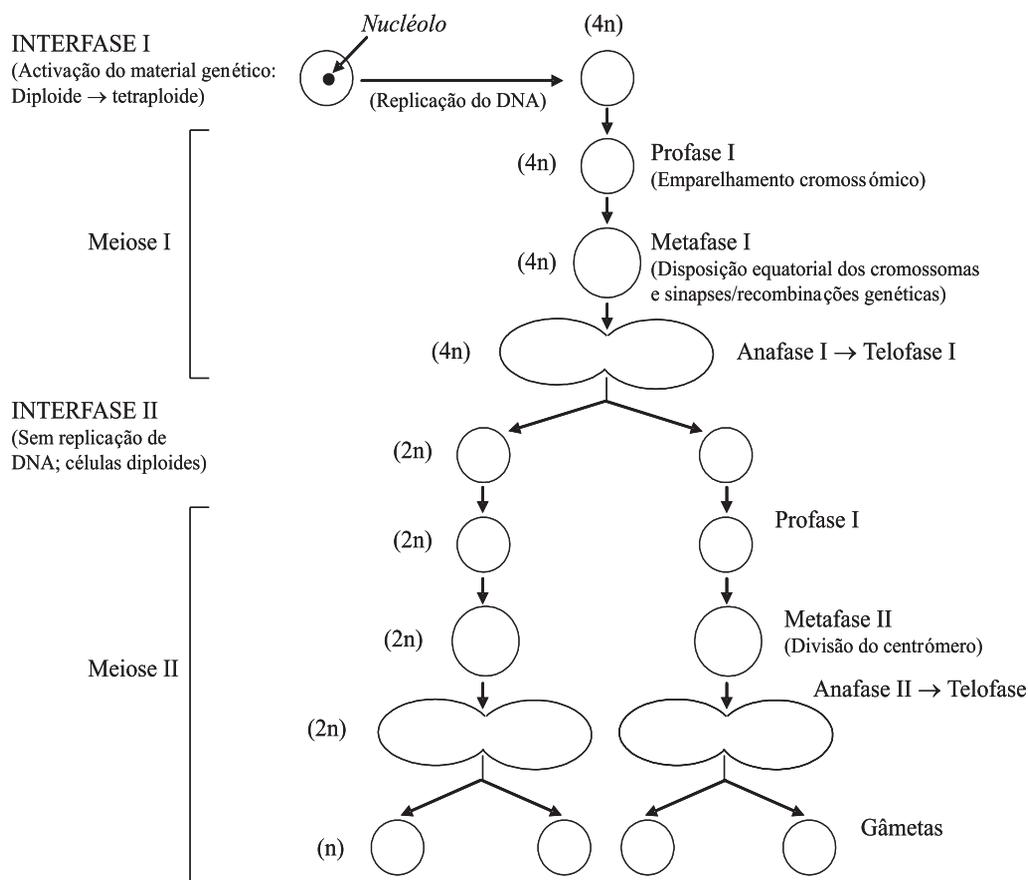
## Meiose

A *meiose* é um tipo de divisão em que as células diploides originam formas haploides, como é o caso das células germinais. Neste processo, cada um dos gametas (espermazoide e óvulo) dos animais e plantas superiores recebe um cromossoma de cada par homólogo da célula original, diploide. Cada um dos cromossomas de cada par homólogo provém dos progenitores (materno e paterno), pelo que são semelhantes (morfologica e geneticamente) mas não idênticos.

A meiose inclui dois tipos de divisão, intercalados por uma interfase (Fig. 1.5.). Após a replicação do DNA (na interfase I, em que cada célula passa de  $2n$  para  $4n$ ) a meiose I evolui sucessivamente pela profase I, metafase I, anafase I e telofase I. Neste ciclo, as alterações decorrentes na interfase I e profase I são semelhantes às

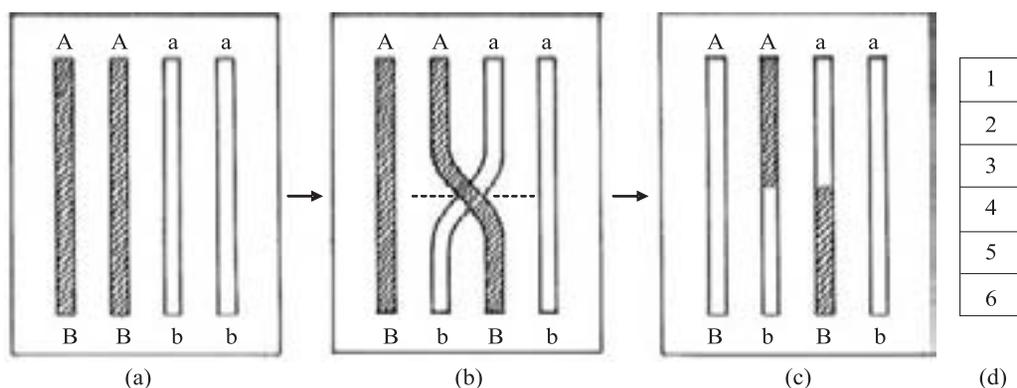
---

<sup>4</sup> Os centrossomas presentes nos polos derivam da duplicação de uma formação existente nas proximidades do núcleo, interfásico, observável por microscopia óptica como uma zona clara, no início da profase.



**Fig. 1.5.** Esquematisação das diversas fases da meiose. Nos animais e plantas superiores a meiose está restringida às células germinais. Após a replicação de DNA, sucedem-se duas divisões sucessivas (meiose I e meiose II) que reduzem o número de cromossomas para metade nas células filhas. Deste modo, as células germinais ( $2n$ ) originam gâmetas haploides (espermatozoides e óvulos) que contêm somente um cromossoma de cada par homólogo presente nas células progenitoras diploides. A meiose I é iniciada após a conclusão da interfase, durante a qual há replicação do material genético. Daqui resulta a replicação dos cromossomas homólogos nos cromatídeos. Durante a metafase I os cromossomas homólogos emparelham-se na região equatorial das células (sinapse) e permutam material genético (recombinação genética). Na anafase dá-se a segregação de cada cromossoma homólogo (constituído por dois cromatídeos unidos pelo centrômero) para polos opostos da célula-mãe, após o que ocorre a separação da célula-filha. Segue-se a meiose II (precedida de replicação do DNA), muito semelhante à mitose, pois que a divisão dos centrômeros possibilita aos cromatídeos de cada par separarem-se para as diferentes células filhas, como cromossomas individualizados. No termo da meiose II são formadas quatro células haploides, cada uma com uma cópia de cada cromossoma.

da mitose. Na metafase I verificam-se as primeiras diferenças: (a) o centrómero de cada par de cromátídeos irmãos não se divide; (b) os dois cromátídeos de cada cromossoma permanecem unidos pelo centrómeros e alinhados no plano equatorial das células (fenómeno designado de *sinapse*). Daqui resulta a permuta de segmentos entre os cromossomas por recombinação genética (*crossing-over*). Na anafase I, cada cópia do mesmo tipo de cromossomas (representado pelo par de cromátídeos unidos pelo centrómero comum) é segregada para polos opostos da célula original, o que culmina na separação em duas células, na telofase I (cada uma das quais recebe os dois cromátídeos de cada tipo de cromossomas homólogos). Devido à segregação do material genético entre os cromossomas de origem materna e paterna na metafase I, as características de cada progenitor, misturadas ao acaso entre os cromátídeos recombinados de cada par homólogo (Fig. 1.6.), são redistribuídos na anafase I, e depois na telofase I, pelas células filhas.

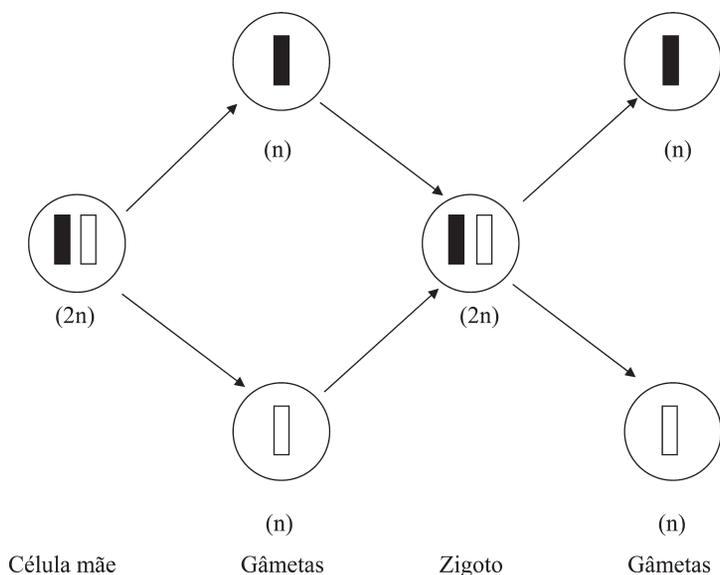


**Fig. 1.6.** Sinapse de recombinação cromossómica (e genética).

(a) A partir do estágio zigótico da profase I, e antecedendo a condensação da cromatina, os cromossomas homólogos replicam-se no respectivo par de cromátídeos, após o que começam a associar-se até ficarem com o maior eixo paralelo e aderentes entre si (o que é designado por *sinapse*) no estágio paquíteno. (b) Segue-se de imediato a recombinação entre os cromátídeos de origem paterna e materna; dois dos cromátídeos (um de cada cromossoma) enrolam-se entre si, segmentando-se em posições variáveis. (c) Os quatro fragmentos resultantes ligam-se dois a dois, reconstruindo dois cromátídeos diferentes dos originais. Neste processo resulta a troca física e uma permuta genética entre cromátídeos de origem materna e paterna, de que resultam cromossomas diferentes do par homólogo progenitor. (d) Supondo que cada algarismo corresponde à posição de um gene diferente, é evidente que para a separação dos genes 1 e 2 durante a recombinação, o cromátídeo teria de ser segmentado exactamente entre ambas as posições; porém, para separar o gene 1 do gene 5, a segmentação poderia decorrer em qualquer posição entre aqueles genes, pelo que a permuta seria mais frequente.

Segue-se a interfase II, em que, por não haver replicação do DNA, é mantido o estado de diploidia celular.

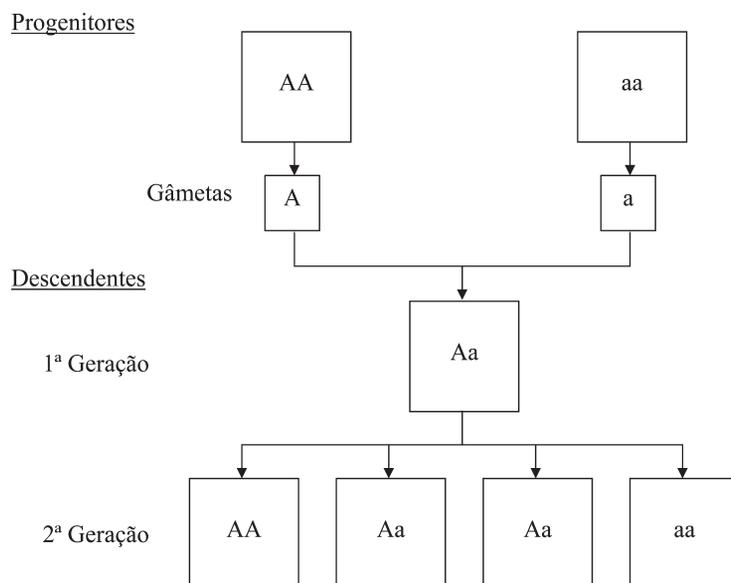
Na segunda divisão meiótica (repartida pela profase II, metafase II, anafase II e telofase II), os centrómeros dividem-se (na metafase II, tal como na mitose), donde sucede a separação de cada cromatídeo (agora cromossoma) pelas quatro células germinais resultantes. Cada uma destas células recebe, portanto, um só cromossoma de cada par homólogo (Fig. 1.7.). A reprodução sexual implica em dois tempos: (a) primeiro ocorre a passagem do estado diploide a haploide pela meiose, a que se segue, no sentido inverso, (b) a fecundação, com conversão das células haploides (do espermatozoide e do óvulo) em ovo diploide (zigoto).



**Fig. 1.7.** Representação do ciclo sexual, com alternância do estado diploide (2n) e haploide (n)

Por fusão dos gametas durante a fertilização, o ovo resultante contém 23 pares de cromossomas, em que cada um é de origem materna ou paterna. Por mitoses sucessivas, as células resultantes mantêm a diploidia.

Devido à segregação e subseqüentes recombinações genéticas, cada um dos gametas adquire uma mistura variável e indeterminada de características maternas e paternas, que está da origem na diversidade dos organismos em gerações sucessivas (Fig. 1.8.). Poderá suceder que a segregação de alguns genes decorra quase sempre com outros tipos de genes (genes associados, *linked genes*).



**Fig. 1.8.** Representação da primeira lei de Mendel, em que é evidenciada a segregação independente dos genes e a distribuição das características dominantes (AA, Aa) e recessivas (aa), através de um par de alelos homólogos. A 1.<sup>a</sup> geração (híbridos Aa; heterozigotos) tem o fenotipo Aa e características (fenotipo) semelhantes a um dos progenitores (AA). Na 2.<sup>a</sup> geração, a proporção entre dominantes e recessivos é de 3/1, dos quais um dos dominantes é homozigoto (AA) e dois são heterozigotos (Aa). Portanto, os genes dominantes e recessivos são transmitidos independentemente, com segregação independente durante a formação dos gâmetas.

## Conclusões

A teoria da hereditariedade baseia-se nos cromossomas celulares.

Cada cromossoma de eucariota é constituído por uma molécula extensa e linear de DNA (com cerca de  $2 \times 10^8$  pares de bases nos cromossomas humanos), onde se localizam o centrómero e os telómeros (extremidades). A molécula de DNA é envolvida quase completamente por proteínas (na maior parte histonas, cujas características básicas se devem a uma constituição muito rica em lisina, arginina e alanina), o que dá ao conjunto uma conformação em “super-hélice”.

A mistura de DNA e proteínas cromossómicas é designada de cromatina, a qual no mesmo núcleo pode apresentar-se sob a forma condensada (hetero-

cromatina) ou difusa (eucromatina). A irregularidade do diâmetro dos filamentos de cromatina resulta da fase de actividade do DNA: durante a replicação e ou na transcrição, o DNA desenrola-se e separa-se das proteínas nos segmentos em que decorrem aquelas funções. Na estrutura dos cromossomas são identificáveis ao longo da sua extensão, por coloração adequada, múltiplas bandas escuras e claras, as quais parecem corresponder às sequências replicantes do DNA. É possível que nem todo o DNA seja transcrito.

Os cromossomas de eucariotas alojam os segmentos de informação hereditária em sequência linear. Essas unidades representam os genes, cujo total constitui o genoma (celular). Cada gene (ou cistrão) é um segmento de DNA (de uma das cadeias do duplex) que codifica uma cadeia polipeptídica (ou molécula de RNA). Cada gene pode existir como duas ou mais estruturas alternativas (alelos). Os genes definem as características hereditárias do organismo portador (genotipo), as quais podem ser ou não expressas. Por conseguinte, o genotipo de um indivíduo representa a combinação dos alelos que possui. O fenotipo corresponde às características do genotipo que têm expressão efectiva em cada indivíduo, em consequência da interacção entre o genotipo e o ambiente em que se situa.

A mitose é a divisão celular pela qual uma célula (haploide ou diploide) origina células filhas idênticas à progenitora, possuindo igual número de cromossomas. A meiose é o mecanismo em que a célula diploide origina quatro células filhas haploides. Para tal, a meiose requer duas divisões sucessivas: na primeira há duplicação dos cromossomas antecedendo a profase I mas os centrómeros não se subdividem, pelo que cada célula filha recebe o par de cromátídeos irmãos de uma das duas cópias de cromossomas homólogo, após recombinação genética. A meiose II não é precedida de replicação do DNA.

A partir da célula inicial todas as divisões originam células filhas com o mesmo número de cromossomas da célula mãe. Exceptuam-se as células sexuais, que contêm apenas metade dos cromossomas. No homem, o estado haploide é muito curto, limitando-se ao tempo de diferenciação dos gâmetas, sem divisão subsequente. A fusão do óvulo e do espermazoide gera um ovo no estado diploide o qual, por misturas sucessivas, origina um novo organismo. A maioria das células no adulto diferenciam-se nos tecidos e órgãos (células somáticas). As células germinais originam gâmetas após meiose.

## CAPÍTULO 2

### NUCLEÓTIDOS E ÁCIDOS NUCLEICOS

#### Estrutura Química

Os mononucleótidos constituem as moléculas-base dos ácidos nucleicos e intervêm de múltiplas formas no metabolismo celular.

Cada *mononucleótido* é constituído por três tipos de moléculas: uma base azotada, uma pentose e uma de fosfato.

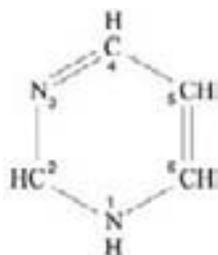
As duas espécies de bases azotadas dos nucleótidos (púricas e pirimídicas) derivam de compostos aromáticos heterocíclicos, respectivamente, a purina e a pirimidina. A purina é um derivado da pirimidina com mais um anel imidazólico do que este composto (Fig. 2.1.).

A ligação glicosídica entre o carbono 1 da D-ribose ou da 2-desoxi-D-ribose (Fig. 2.2.) e o N<sub>1</sub> da pirimidina ou o N<sub>9</sub> da purina origina um *nucleósido*, designado, respectivamente, por ribo- ou 2'-desoxirribonucleósido (Fig. 2.3.). Ambos os tipos são hidrolisáveis por enzimas específicas, as *nucleosidases*. Os nucleósidos são, portanto, constituídos por uma pentose e uma base (púrica ou pirimídica). Os nucleótidos resultam da fosforilação dos nucleósidos, através de uma ligação éster entre uma molécula de ácido fosfórico (fosfato) e um dos grupos hidroxilo (OH) livres da pentose.

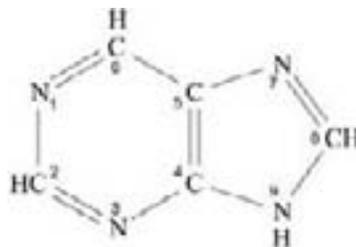
São também significativas as quantidades de nucleótidos livres existentes nas células que são produtos da hidrólise enzimática parcial dos ácidos nucleicos, por acção das enzimas *nucleases*.

A esterificação dos desoxirribonucleótidos pode ser feita apenas nas posições 3' e 5' da pentose. Ambos os tipos existem na Natureza. São três as posições possíveis de esterificação nos ribonucleótidos (carbonos 2', 3' e 5' da pentose).

(a)

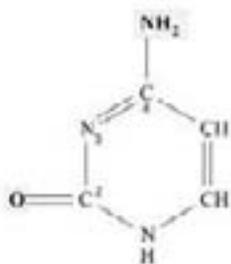


Pirimidina

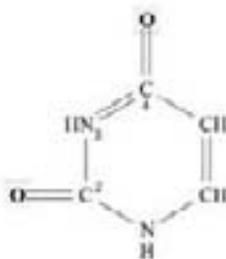


Purina

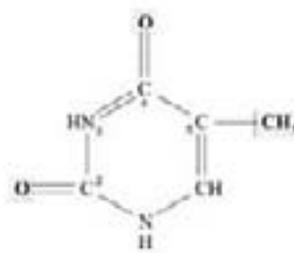
(b)

Derivados da pirimidina

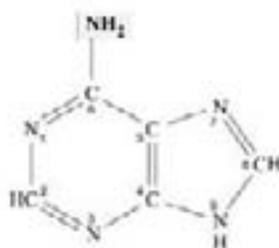
Citosina (2-oxi-4-aminopirimidina)



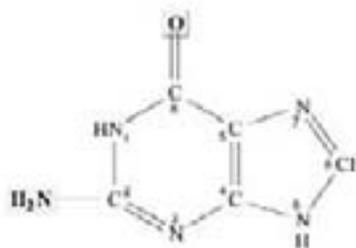
Uracilo (2,4-dioxipirimidina)



Timina (5-metil-2,4-dioxipirimidina)

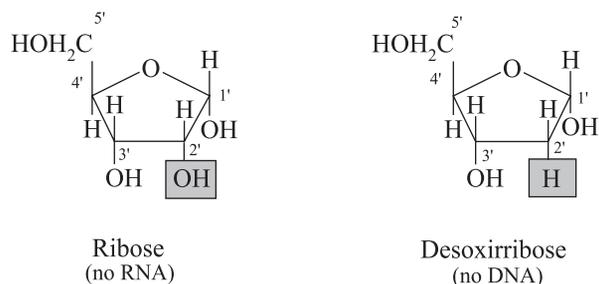
Derivados da purina

Adenina (6-aminopurina)

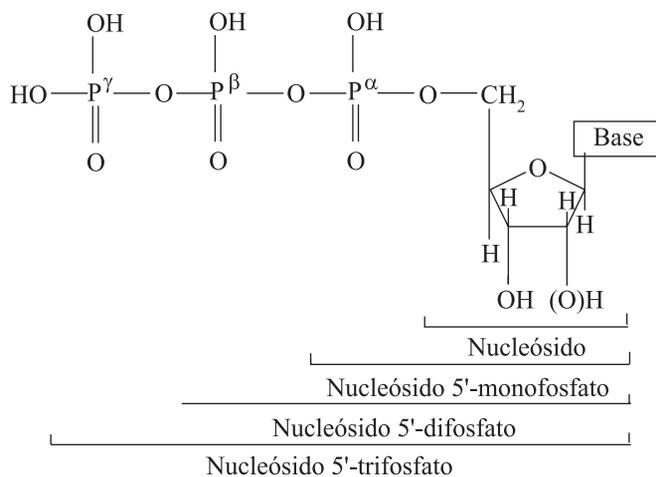


Guanina (2-amino-6-oxipurina)

**Fig. 2.1.** (a) Representação estrutural dos compostos heterocíclicos aromáticos pirimidina e purina, de que derivam as duas classes de bases azotadas constituintes dos nucleósidos e nucleótidos. (b) Principais bases púricas e pirimídicas dos nucleósidos e nucleótidos. Derivados da purina: adenina (6-aminopurina) e guanina (2-amino-6-oxipurina); derivados da pirimidina: citosina (2-oxi-4-aminopirimidina), uracilo (2,4-dioxipirimidina) e timina (5-metil-2,4-dioxipirimidina).



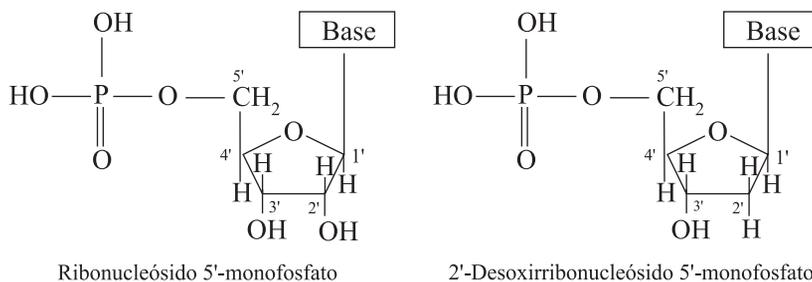
**Fig. 2.2.** Representação estrutural da D-ribose e da D-desoxirribose. As moléculas diferem entre si na posição 2', a qual, na desoxirribose, está desprovida do átomo de oxigênio existente no grupo hidroxilo da ribose.



**Fig. 2.3.** Esquema geral dos principais ribonucleósidos e desoxirribonucleósidos 5'-fosfato.

Nas células predominam os nucleótidos com um grupo fosfato na posição 5', os quais são intermediários normais no metabolismo dos ácidos nucleicos e em outras actividades metabólicas (Fig. 2.4.; Quadro 2.1.).

Também existem nucleótidos com 2 e 3 grupos fosfato na cadeia lateral ligada à posição 5' da pentose. São os ribonucleósidos ou 2' desoxirribonucleósidos 5' di- e 5'- trifosfato, cujas ligações podem ser selectivamente hidrolisadas por enzimas específicas celulares.



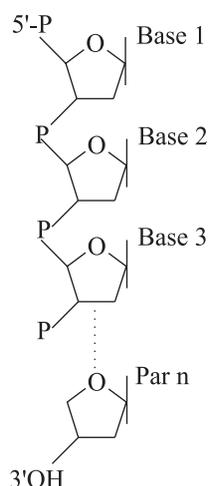
**Fig. 2.4.** Esquema geral dos nucleósidos e nucleótidos (nucleósidos 5'-mono-, di- e trifosfato). As bases mais comuns são as purinas adenina e a guanina, e as pirimidinas citosina, uracilo ou timina.

**Quadro 2.1.** Nomenclatura dos nucleósidos e nucleótidos mais frequentes.

Bases	Nucleósidos		Nucleótidos	
	Com D-ribose	Com 2'-desoxi D-ribose	Com D-ribose	Com 2'-desoxi D-ribose
Adenina (A)	Adenosina	2' Desoxiadenosina	Ácido adenílico ou Adenosina 5'-monofosfato (AMP)	Ácido desoxiadenílico ou Desoxiadenosina 5'-monofosfato (d AMP)
Guanina (G)	Guanosina	2'-Desoxiguanosina	Ácido guanílico ou Guanosina 5'-monofosfato (GMP)	Ácido desoxiguanílico ou Desoxiguanosina 5'-monofosfato (d-GMP)
Citosina (C)	Citidina	2'-Desoxicitidina	Ácido citidílico ou Citidina 5'-monofosfato (CMP)	Ácido desoxicitidílico ou Desoxicitidina 5'-monofosfato (d CMP)
Timina (T)	–	2'-Desoxitimidina	–	Ácido desoxitimidílico ou Desoxitimidina 5'-monofosfato (d TMP)
Uracilo (U)	Uridina	–	Ácido uridílico ou Uridina 5'-monofosfato (UMP)	–

Uma das funções mais importantes dos nucleósidos 5'-trifosfato é a de formar precursores de mononucleótidos ricos em energia química, indispensável ao processo das biossínteses, e outras actividades endergónicas celulares, em que se inclui a síntese enzimática dos ácidos nucleicos.

Os nucleótidos que constituem os ácidos nucleicos unem-se entre si, em regra, por pontes de ácido fosfórico, entre as posições 5' e 3' da D-ribose ou 2-desoxi-D-ribose, ainda que possam existir outras pontes de união em nucleótidos estranhos aos ácidos nucleicos, mas ainda assim existentes sob a forma livre nas células (p.ex. as coenzimas NAD, NADP e FAD, entre outras). Por conseguinte, os ácidos nucleicos são polímeros constituídos por cadeias lineares de muitos mononucleótidos, ligados por pontes fosfodiéster entre as posições 5' e 3' das respectivas pentoses (Fig. 2.5.). Este processo de união ocasiona que ambos os tipos de ácido (desoxirribonucleico e ribonucleico) possuam propriedades semelhantes.

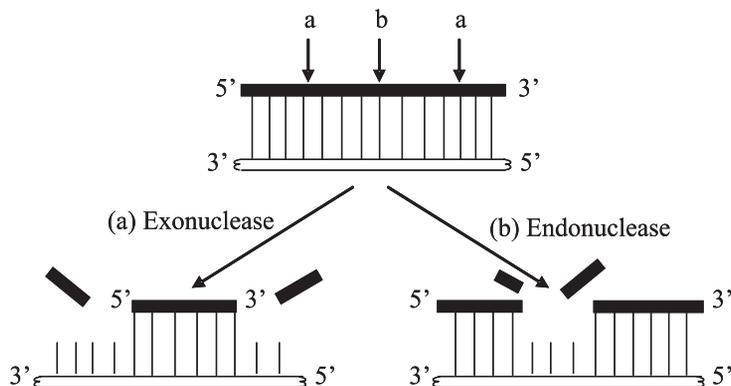


**Fig. 2.5.** Estrutura geral de um polinucleótido. Os nucleótidos são unidos entre si através de um resíduo de fosfato, por ligações 3',5' fosfodiéster entre o grupo 3'-hidroxilo de um nucleótido e o 5'-hidroxilo do seguinte. Deste modo, cada polinucleótido é uma sequência de resíduos alternados da pentose e fosfato, com uma cadeia lateral (a base) unida à pentose de cada mononucleótido.

Um modo de identificar os polinucleótidos consiste em anotar a sequência das suas bases (p.ex. A, G, U ou T, C) e o tipo de pentoses que os compõem.

A ligação éster 3', 5' é facilmente hidrolisada, libertando nucleótidos de várias dimensões. Igual efeito é obtido por *nucleases* ou *fosfodiesterases* específicas.

Há 2 tipos de fosfodiesterases (*a* e *b*), que actuam entre o hidroxilo 3' e o grupo de ácido fosfórico, (*a*), ou entre este e o grupo hidroxílico 5', (*b*). Conforme hidrolisam as pontes fosfodiéster a partir do extremo das cadeias polinucleotídicas (onde existem grupos hidroxilo livres) ou actuam em qualquer sector (independentemente dos grupos terminais), estas enzimas tomam o nome de *exo-* ou *endonucleases* (Fig. 2.6.).



**Fig. 2.6.** Representação de acção enzimática de uma exonuclease (a) e de uma endonuclease (b) numa das cadeias de moléculas de DNA. Ambas as nucleases actuam na ligação fosfodiéster do esqueleto dos polinucleótidos. As exonucleases removem sucessivos resíduos de desoxiribonucleótido das extremidades da cadeia de DNA em que actuam, enquanto as endonucleases actuam em ligações internas próximas de sequências específicas de nucleótidos.

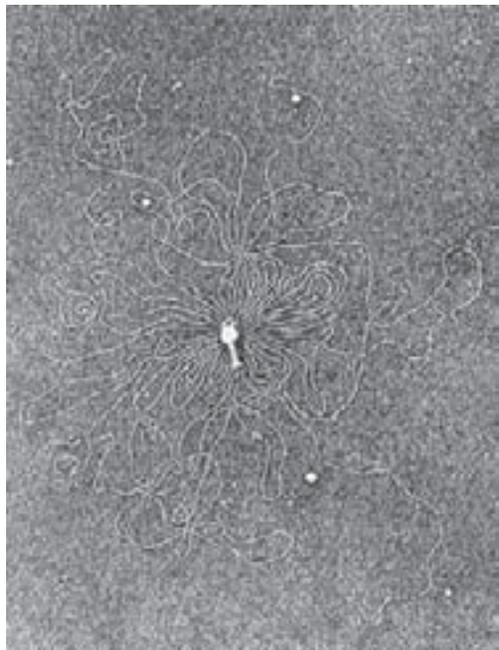
Os ácidos nucleicos constituem moléculas de DNA e de RNA. Alguns associam-se a proteínas específicas, formando complexos funcionais muito activos (p.ex. ribossomas e vírus).

As moléculas de DNA de qualquer tipo celular incluem sempre os mesmos tipos de mononucleótidos (dAMP, dGMP, dTMP e dCMP, onde *d* simboliza a molécula de desoxirribose). As proporções e sequência destes mononucleótidos na molécula de DNA variam com a espécie biológica, ainda que obedçam às seguintes relações fundamentais:

- 1 – A soma das bases púricas é igual à das pirimídicas
- 2 – A proporção molar da adenina iguala a da timina, enquanto a guanina equivale à citosina.

Além destas bases principais, comuns a todos os tipos de DNA, existem (em especial nos vírus) algumas bases especiais, que são quase sempre derivados metilados daquelas.

O peso molecular é variável. Quase todo o DNA existente nos organismos procariontes localiza-se num único cromossoma (Fig. 2.7). Nos eucariotes há vári-



**Fig. 2.7.** Molécula de DNA (extensa faixa desdobrada) expelida da cabeça de um bacteriófago T2 da *Escherichia coli*, após o fago ser submetido a shock osmótico. Reprodução autorizada da micrografia electrónica amavelmente cedida pelo autor, Dr. Alberch K. Kleinschmidt (Department of Biochemistry, New York University School of Medicine, USA).

as ou muitas moléculas de DNA, dependendo do número de cromossomas de cada célula. Por esse motivo, a maioria do DNA celular situa-se no núcleo, onde se combina com as histonas, ainda que exista uma pequena fracção citoplásmica que difere da nuclear nas bases e no peso molecular dos seus constituintes.

Existem três tipos principais de RNA celular (*mensagemiro*, *ribossómico* e de *transferência*), cada um com características próprias quanto à composição de bases e peso molecular, mas constituídos por uma só faixa de polinucleótidos. A maior percentagem localiza-se no citoplasma celular.

## Conclusões

Os mononucleótidos são compostos por três moléculas unidas entre si por ligações covalentes: uma base azotada (púrica ou pirimídica), uma pentose (D-ribose ou 2-desoxi-D-ribose) e uma de ácido fosfórico ionizado (fosfato). Os ribonucleótidos e os desoxiribonucleósidos diferem também nas bases pirimídicas uracilo (que existe somente nos ribonucleótidos) e timina (nos

desoxirribonucleótidos). Na ausência de grupo fosfato, a molécula constituída por uma base azotada ligada a uma pentose é designada de mononucleósido. Existem dois tipos de mononucleósido consoante o tipo de pentose que as constitui: ribonucleósidos (contêm D-ribose) e desoxirribonucleósidos (com 2-desoxi-D-ribose).

Nos mononucleótidos mais comuns o grupo fosfato está unido por uma ligação éster ao grupo 5'-OH da pentose. Os mononucleótidos são componentes monoméricos do ácido nucleico e participam em outras estruturas e funções importantes, designadamente coenzimas e transporte energético (p.ex., ATP), transporte de moléculas (p.ex., UDT-glicose), transferência de electrões nas reacções de oxidação-redução (p.ex., NAD, FAD).

Os ácidos nucleicos são polímeros lineares de mononucleótidos, unidos entre si por pontes fosfodiéster entre o grupo 3'-OH da pentose de um nucleótido e o grupo 5'-OH do seguinte, e que confere ao ácido nucleico uma polaridade química, 3'→5'. As bases estão unidas ao resíduo da pentose, orientando-se na perpendicular do eixo pentose-fosfato. No estado nativo, o DNA tem duas cadeias complementares enroladas em dupla hélice (duplex). O RNA tem apenas uma cadeia.

As bases dos monómeros constituintes dos ácidos nucleicos são, quase sempre, a adenina (A), guanina (G), citosina (C) e a timina (T; nos desoxirribonucleósidos e ácidos desoxirribonucleicos) ou o uracilo (U; nos ribonucleósidos e ácidos ribonucleicos). As bases mais comuns podem ser substituídas excepcionalmente por bases raras. As ligações internucleotídicas DNA e RNA podem ser hidrolisadas por enzimas (fosfodiesterases ou nucleases) de que resultam 5' fosforilo-nucleósidos ou 3' fosforilo-nucleósidos (consoante o local de hidrólise da ligação 3', 5'-fosfodiéster). O DNA é uma molécula de elevado peso molecular, em que a composição e sequência das bases, variáveis de molécula para molécula, corresponde à memória da informação genética celular. No mesmo organismo existe um tipo de molécula de DNA e três de RNA (mensageiro, transferência e ribossómico).

## CAPITULO 3

# CONFORMAÇÃO, REPLICAÇÃO, REPARAÇÃO E RECOMBINAÇÃO DO ÁCIDO DESOXIRRIBONUCLEICO

### Estrutura molecular do duplex

Entre a descoberta do DNA nos cromossomas, no início deste século, e a afirmação de que era o centro da informação genética celular, mediaram cerca de 50 anos (vide Anexo II – Cronologia). O problema principal parecia quase intransponível. Consistia em saber como era guardada a informação e depois transferida a nível molecular. Watson e Crick resolveram a questão ao apresentarem uma teoria fascinante, mediada por um mecanismo químico muito simples que, progressivamente, tem vindo a ser confirmado.

Vários factos experimentais evidenciaram o DNA como o material celular que determina as características hereditárias, sendo as proteínas cromossómicas desprovidas, em absoluto, desta função. Os trabalhos de Avery e Cols com pneumococos estiveram na origem daquela ideia, ao induzirem a virulência nas estirpes primitivamente não patogénicas pela administração de DNA extraído e purificado de estirpes patogénicas. O facto de uma infecção celular por vírus resultar da entrada do DNA (cromossoma viral; em alguns é constituído por RNA), enquanto as proteínas de revestimento ficam retidas no espaço extracelular, é outro exemplo de que a informação genética reside no DNA e não nas proteínas (Fig. 3.1.).

A quantidade de DNA de determinada espécie é constante para todo o organismo e independente de alterações exteriores, do metabolismo celular e das condições nutritivas do meio. Quanto mais complexa é a célula maior é a proporção de DNA que contém. Assim, p.ex., enquanto nos mamíferos cada célula

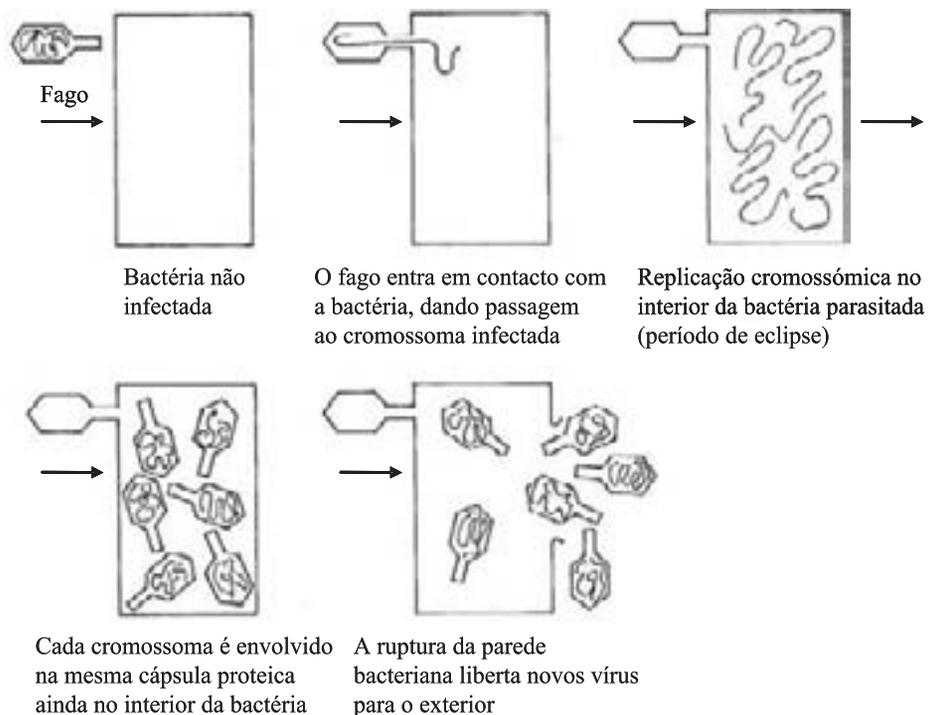


Fig. 3.1. Multiplicação de vírus no interior da célula parasitada.

possui cerca de  $6 \times 10^{-6}$  mg de DNA, nos peixes é de  $2 \times 10^{-6}$  e nos moluscos é de  $1,2 \times 10^{-6}$  mg. As bactérias, muito mais simples, possuem aproximadamente  $0,02 \times 10^{-6}$  de DNA por unidade celular. As células haploides dos eucariotas têm metade do conteúdo de DNA nas células somáticas (diploides).

O DNA localiza-se quase exclusivamente nos cromossomas nucleares. Em solução apresenta-se como uma substância com elevada viscosidade. Por ser muito estável, e não estar sujeito a síntese e degradação constantes, como a maioria dos outros substractos metabólicos, o DNA evidencia-se como a substância ideal para a conservar fielmente a informação genética das espécies.

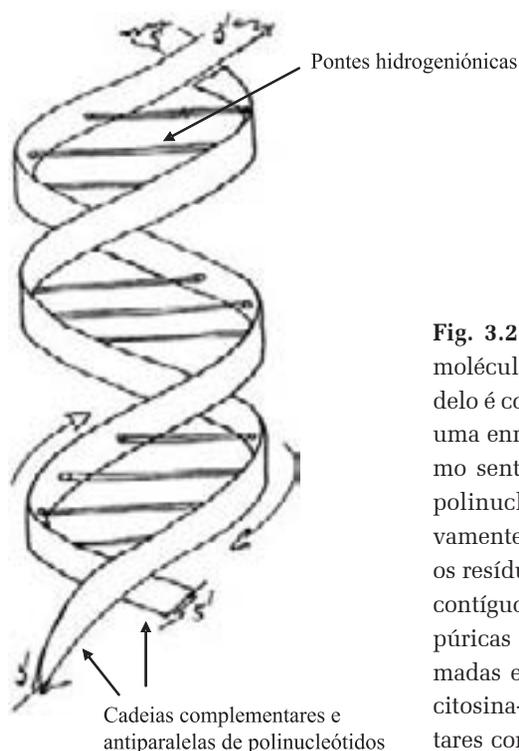
Foi sugerida a importância que teria a equivalência e o emparelhamento das bases nucleotídicas na organização do DNA e nas respectivas funções.

Era conhecido que o DNA celular tinha idêntica composição de bases em tecidos diferentes da mesma espécie, mas que variava de espécie para espécie.

Outro facto conhecido era a existência de uma relação constante entre o conjunto de bases do mesmo DNA, de modo que

$$A + G = C + T, \text{ sendo } A = T \text{ e } G = C.$$

Suspeitava-se de que o DNA tinha estrutura tridimensional. Por difracção de raios X foram observadas na molécula de DNA periodicidades variáveis com o estado de hidratação (entre duas formas extremas, A e B), sugerindo uma conformação helicoidal. Estas observações conduziram à concepção do esquema tridimensional em duplex, apresentado por Watson e Crick em 1953. De acordo com aquela hipótese a molécula de DNA seria constituída por duas cadeias extensas e finas de polinucleótidos, enroladas em torno de si próprias e de um eixo comum, formando uma dupla hélice regular (*duplex*). As bases estariam localizadas na perpendicular do eixo maior do duplex, unidas a cada uma das cadeias e paralelas entre si (Fig. 3.2.).

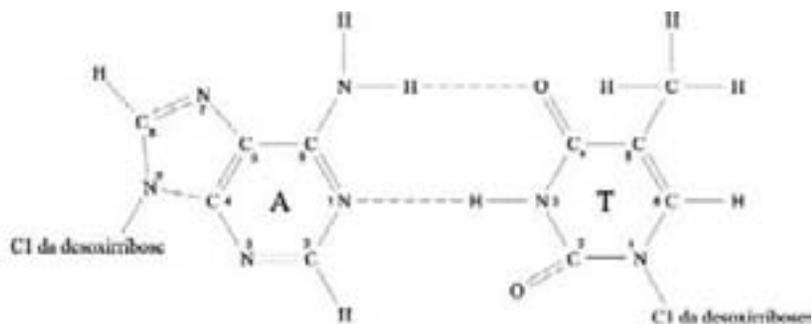


**Fig. 3.2.** Modelo tridimensional em dupla-hélice da molécula de DNA, definido por Watson e Crick. O modelo é constituído por duas cadeias de nucleótidos, cada uma enrolada para a direita e ambas enroladas no mesmo sentido em torno de um eixo comum. As cadeias polinucleotídicas do duplex são anti-paralelas (relativamente à direcção das ligações 3', 5'-fosfodiéster, entre os resíduos de fosfato e da desoxirribose de nucleótidos contíguos) e complementares (quanto ao tipo de bases púricas e pirimídicas). As ligações hidrogeniônicas formadas entre os pares adenina-timina (duas ligações) e citosina-guanina (três ligações) das cadeias complementares contribuem para a estabilidade da duplex.

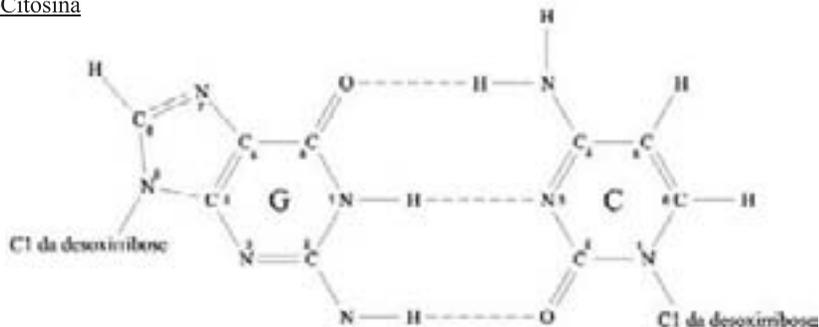
O diâmetro maior do duplex foi calculado em cerca de 20 Å; cada cadeia concluiria um enrolamento completo em cada 34 Å, em que cada segmento teria 10 nucleótidos. Deste modo, as bases estariam afastadas entre si cerca de 3,4 Å. Normalmente, as cadeias de polinucleótidos do duplex seriam constituídas por quatro tipos de nucleótidos, dispostos em sequência variável; cada nucleótido seria constituído por um resíduo de desoxirribose, um grupo fosfato e uma base púrica (adenina, A ou guanina, G) ou pirimídica (timina, T ou citosina, C). Cada cadeia seria constituída por uma disposição sequencial constante (desoxirribose-fosfato-desoxirribose-fosfato, etc.).

Em contraste com aquela regularidade do “esqueleto” de DNA, a sequência relativa e a ligação das bases à respectiva cadeia varia bastante de molécula para molécula. No entanto o emparelhamento das bases entre ambas as cadeias decorrerá se forem complementares entre si (A-T ou G-C) (Fig. 3.3.). As ligações hidrogé-

#### Adenina-Timina



#### Guanina-Citosina



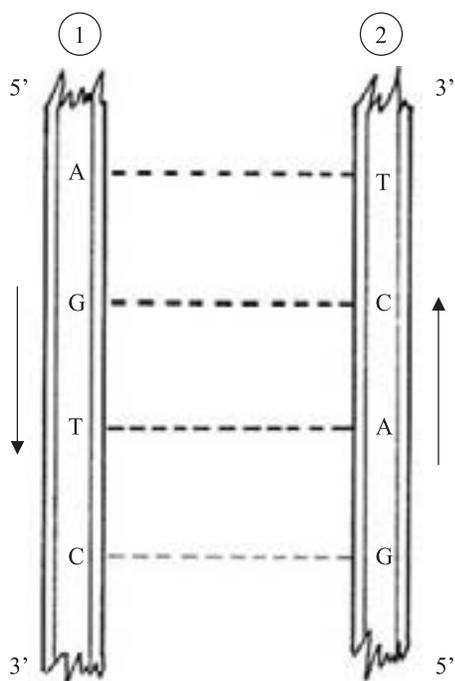
**Fig. 3.3.** Disposição complementar e antiparalela de um segmento do duplex de DNA, evidenciando as bases emparelhadas (A – adenina; T – timina; G – guanina; C – citosina) entre as cadeias 1 e 2.

niónicas entre ambos aqueles grupos são muito mais fortes do que seriam para os pares A-G e C-T; por outro lado, o par A-G seria demasiado volumoso para ocupar o espaço calculado que existe entre ambas as cadeias do duplex, enquanto os constituintes do par C-T, pelas suas reduzidas dimensões, estariam demasiado afastados entre si, não permitindo a formação de ligações hidrogeniónicas estáveis.

Ambos os pares de bases, hidrofóbicas, como que se “refugiam” no espaço formado pelo duplex, protegidos pelas moléculas de desoxirribose e pelos grupos fosfato periféricos, aqueles hidrofílicos e estes com carga negativa.

Para que a complementariedade se verifique entre as bases de ambas as cadeias (p.ex. à sequência AGTC em uma das cadeias tem de corresponder na outra cadeia do duplex a sequência TCAG) é indispensável que os dois polímeros estejam orientados em sentidos opostos (Fig. 3.4.).

É muito importante para a estabilidade da molécula de DNA, que os átomos de hidrogénio que a constituem não variem nas suas posições relativas, ainda que isto possa suceder muito raramente (p.ex., modificação do grupo amina  $NH_2$  das bases em grupo imino, NH).



**Fig. 3.4.** Disposição complementar e antiparalela de um segmento do duplex de DNA, evidenciando as bases emparelhadas (A – adenina; T – timina; G – guanina; C – citosina) entre as cadeias 1 e 2.

A relação  $A + T / G + C$  varia consoante os organismos considerados. Enquanto no Homem é de 1,4, nos seres procariotas tanto pode ser 1,84 (no bacteriófago  $T_2$ ) como 0,97 (na *Escherichia coli*<sup>1</sup>). A explicação para este facto é um pouco difícil e especulativa. No entanto é sabido que organismos semelhantes têm proporção aproximada de bases, a qual é constante na mesma espécie. Experimentalmente é possível produzir híbridos de DNA a partir de cadeias de DNA desnaturadas de duas espécies diferentes. Por arrefecimento lento, constatou-se que uma cadeia de DNA de murganho tende a ligar-se complementarmente a 25% de uma cadeia de DNA humano, o que demonstra a existência de algumas semelhanças de constituição entre ambos os tipos de DNA, provavelmente entre os segmentos que codificam proteínas com funções semelhantes no murganho e no homem.

## Replicação do DNA

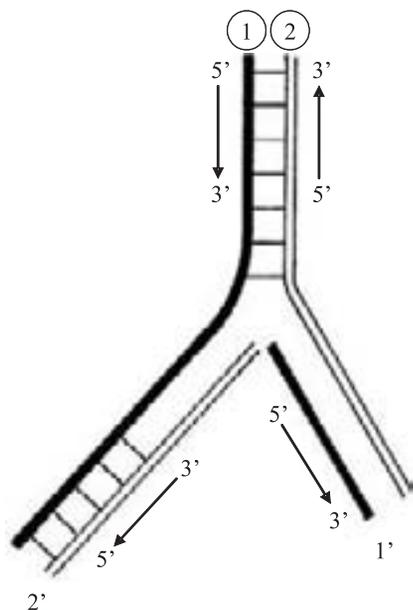
A replicação do DNA consiste na duplicação da duplex-mãe em duas moléculas de DNA idênticas, entre si e à original. As cadeias de cada duplex são complementares e antiparalelas entre si, quanto ao tipo e sequências das bases nucleotídicas que as constituem (Fig. 3.4.).

A complementaridade das duas faixas de DNA explica ainda a sua replicação durante a divisão celular, cada uma sendo o modelo necessário para a síntese de uma nova cadeia que lhe é complementar. Para que o duplex seja replicado, as duas cadeias constituintes têm de se afastar entre si, o que sucede após a rotura das ligações hidrogeniónicas entre as respectivas bases centrais. À medida que o processo de replicação progride, as cadeias separam-se de modo a cada uma ser replicada numa nova faixa complementar adquirindo cada novo par o enrolamento em duplex. As moléculas resultantes são duplex híbridos, constituídos, cada um por uma das cadeias de DNA-mãe e por uma cadeia replicada, complementar daquela. (Fig. 3.5.).

Esta replicação “semi-conservadora”, que já havia sido proposta por Watson e Crick na sua hipótese original sobre a transmissão de informação genética, foi confirmada por Meselson e Stahl, em culturas de *E. coli*.

---

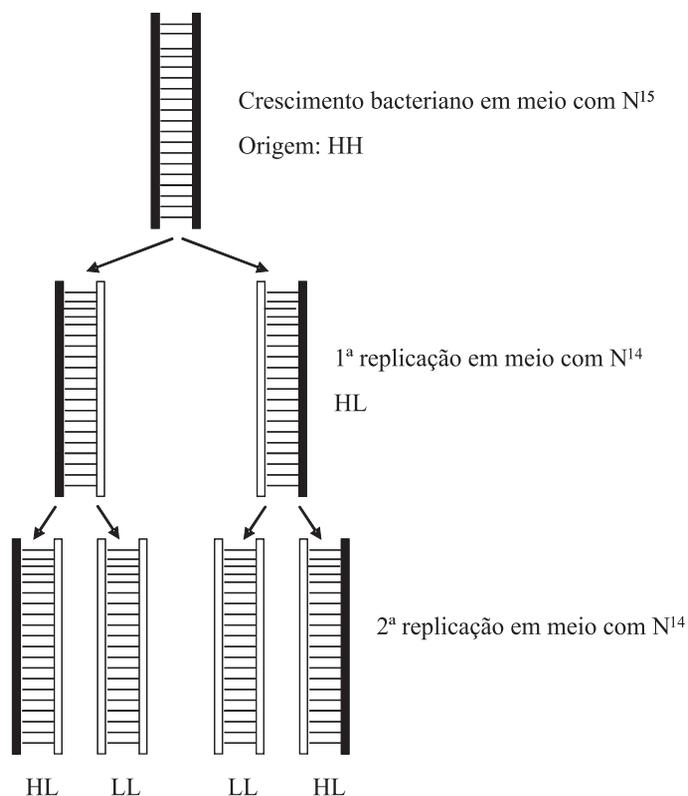
<sup>1</sup> Abreviatura utilizada no texto: *E. coli*



**Fig. 3.5.** Esquema de replicação em que cada cadeia (1 e 2) origina o seu respectivo complemento (2' e 1').

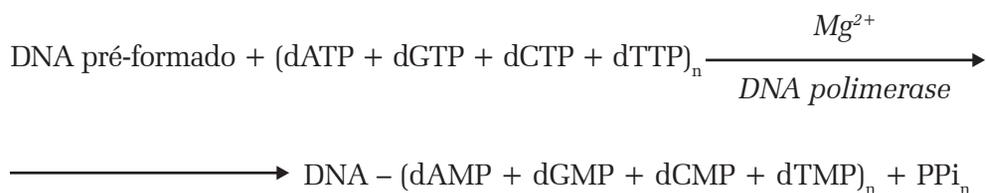
A experiência consistiu em fazer crescer gerações de *E. coli* num meio com azoto radioactivo, entre os nutrientes necessários. Consoante era usado  $N^{14}$  ou  $N^{15}$  assim as bactérias continham DNA com particularidades diferentes. As bactérias com DNA “pesado” (com  $N^{15}$ ), ao serem transferidas para um meio com  $N^{14}$ , evidenciavam uma alteração da densidade relativa dos respectivos DNA em gerações subsequentes. Concluiu-se que, desde a 1.<sup>a</sup> geração (duplicação do número de células), o DNA possuía características híbridas próprias de uma estrutura com duas faixas de densidade diferente (uma incorporando  $N^{15}$  e a outra  $N^{14}$ ). Na 2.<sup>a</sup> geração, duas moléculas de DNA eram iguais ao tipo “leve” inicial (com  $N^{14}$ ) enquanto as outras duas eram híbridas semelhantes às da 1.<sup>a</sup> geração (Fig. 3.6.). Admitiu-se que este tipo de replicação (semi-conservadora) também existiria nos eucariotas.

No seguimento daquelas experiências foi verificado que a replicação das (novas) cadeias complementares do DNA parental era catalisada por uma polimerase específica. Esta enzima promovia o crescimento do polímero de desoxirribonucleótidos, unidos entre si em sequência, por pontes 3', 5'-fosfodiéster. Para exercer a sua actividade, a *DNA polimerase* requeria a presença de um molde de DNA,  $Mg^{2+}$  e todos os principais desoxirribonucleósidos-5'-trifosfato



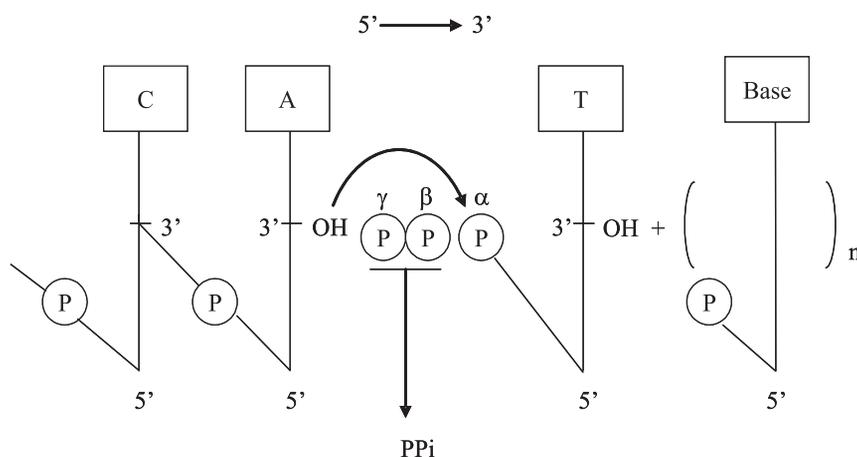
**Fig.3.6.** Replicação conservadora de DNA. Utilizando um isótopo pesado ( $N^{15}$ ), a *E. coli* incorpora-o nas duas cadeias de DNA (forma HH). Quando as bactérias contendo DNA “pesado” são transferidas para um meio com azoto “leve” ( $N^{14}$ ), deixando-as replicar uma vez, o DNA resultante terá densidade intermédia entre as formas de DNA pesado e leve. O DNA desta geração será híbrido (HL). Numa 2.ª geração, metade do DNA será híbrido (HL) e a outra metade terá DNA leve (LL).

de que derivam os constituintes comuns nas moléculas de DNA. A omissão de qualquer um deles não permitia a formação do DNA:



A polimerização ocorria no sentido 5'→3', ou seja, cada mononucleótido ligava-se através de um grupo fosfato em 5' ao grupo 3'-hidroxilo livre do mononucleótido precedente na cadeia de DNA (Fig. 3.7.).

O pirofosfato (PPi), libertado após a formação de cada ligação fosfodiéster, era hidrolisado por acção de uma *pirofosfatase* em ortofosfato (Pi); aquela hidrólise, juntamente com o decréscimo de energia livre resultante da subsequente formação de um duplex, tornaria irreversível a reacção no sentido da síntese.



**Fig. 3.7.** Formação de uma ligação fosfodiéster por acção da DNA polimerase. Na presença de um segmento pré-formado de DNA, a DNA polimerase catalisa a união de um molécula de desoxirribonucleósido-trifosfato (em que o grupo trifosfato está unido à posição 5' da respectiva desoxirribose) ao grupo 3'-hidroxilo livre da cadeia. Por deslocação nucleofílica, aquele grupo 3'-hidroxilo ataca o grupo fosfato α do novo nucleósido-trifosfato (P). Daqui resulta uma ligação fosfodiéster (5' → 3'), com remoção de um grupo pirofosfato (PPi) e o alongamento da cadeia de DNA em mais uma unidade incorporada. A diminuição da energia livre, por subsequente hidrólise do pirofosfato em duas moléculas de fosfato inorgânico, a par com a formação da dupla hélice entre a cadeia de DNA que está a ser alongada e a que lhe serve de modelo complementar, confere à reacção de polimerização características exergónicas que favorecem a sua continuação.

Posteriormente, foi demonstrado que a DNA polimerase, por si só, não possuía a capacidade de iniciar cadeias de DNA mas somente a de promover o alongamento da cadeia pré-existente, quer de DNA (nos procariontes) ou de DNA ou RNA iniciador (nos eucariotes).

A DNA polimerase purificada revelava outra actividade catalítica, semelhante à das exonucleases. Em consequência da remoção hidrolítica de mononucleótidos a partir de qualquer extremidade da cadeia de DNA, seria possível isolar sucessivos resíduos de desoxirribonucleósido-monofosfato.

Foi detectada a existência de DNA polimerase em muitos tipos celulares, quer de organismos procariotas quer nos eucariotas; nestes últimos, a polimerase aparece localizada sobretudo no núcleo, ainda que exista alguma quantidade nas mitocôndrias, possivelmente relacionada com a replicação do respectivo DNA<sup>2</sup>.

A DNA polimerase que foi purificada a partir de *E. coli* apresentava peso molecular de 109.000 dalton, era constituída por uma cadeia polipeptídica com conformação esférica, e tinha um diâmetro cerca de três vezes superior ao de um duplex de DNA.

Outra característica da DNA polimerase, muito possivelmente relacionada com a sua grande especificidade de acção, relaciona-se com os pontos de ligação que possui, separados uns dos outros, cada um com funções diferentes.

A enzima possui um sítio de ligação no centro activo, a que se fixam, em sequência, as moléculas de desoxirribonucleótido-5'-trifosfato a serem utilizadas na formação das cadeias replicadas. Próximo daquele sítio catalítico, a molécula enzimática apresenta outro sítio para a fixação das cadeias simples lineares ou fechadas de DNA, que funcionam de molde à replicação. Foi demonstrado que a DNA polimerase actuava melhor quando o DNA fornecia não só o modelo mas também uma faixa iniciadora, à qual se juntavam os sucessivos mononucleótidos que constituíam a sequência complementar à do molde. Num tipo muito purificado da polimerase foi identificado um sítio específico para a sua união à faixa de DNA a ser replicada.

Para esclarecer o mecanismo de acção da DNA polimerase foram admitidas duas hipóteses: ou cada molécula de DNA constituía um “núcleo de cresci-

---

<sup>2</sup> As mitocôndrias são como que estruturas autónomas dentro da própria célula, no que concerne à síntese das próprias proteínas. Possuem DNA circular, RNA dependente DNA-polimerase, moléculas de RNA<sub>i</sub> e aminoacil-RNA<sub>i</sub>-sintetase específicas e utilizam também o N-formil metionina como aminoácido inicial. Podem sintetizar RNA<sub>m</sub> a partir do próprio DNA, o que é também inibido pela actinomicina. Os ribossomas parecem estar quase sempre dissociados nas suas subunidades, assemelhando-se aos do tipo 70S das bactérias. Talvez por esse motivo a síntese proteica mitocondrial possa ser inibida pelo cloroanfenicol, o que não acontece com a síntese mediada pelos ribossomas 80S. Aquele antibiótico actua ao nível das subunidades 50S, impedindo a formação das ligações peptídicas (ver Capítulo 4). No conjunto, as mitocôndrias, assim como os cloroplastos, também possuidores de DNA, assemelham-se a bactérias.

mento” que, por adição sucessiva de mononucleótidos, conduzia à cadeia completa, ou, em alternativa, representaria apenas um “modelo” a partir do qual a enzima copiava uma nova sequência complementar.

Em experiências com *E. coli*, utilizando DNA desnaturado de diversas proveniências (animais, plantas, vírus, bactérias), foi verificado que a DNA polimerase daquela bactéria formava novo DNA, sendo tanto maior a actividade daquela enzima e a quantidade do DNA sintetizado quanto mais extensa fosse a molécula de DNA utilizada como modelo; contudo, a cópia de DNA nunca excedia as dimensões do modelo utilizado.

Outro processo consistiu em observar a composição de bases do DNA polimerizado em meios celulares preparados a partir de culturas de *E. coli* (contendo misturas de dATP, dGTP, dCTP, dTTP, enzima purificada e moléculas de DNA desnaturado de várias espécies). Verificou-se que o produto obtido mantinha a proporção  $(A + T)/(G + C)$  entre as suas bases, bem como respectiva composição, muito idêntica à do DNA modelo.

Entretanto, foi identificada uma outra enzima em extractos de *E. coli*, a *DNA ligase*. Esta enzima tem a propriedade de unir entre si as extremidades (ou segmentos constituintes) de duas cadeias de DNA ou de uma única cadeia, no caso de ser uma molécula circular.

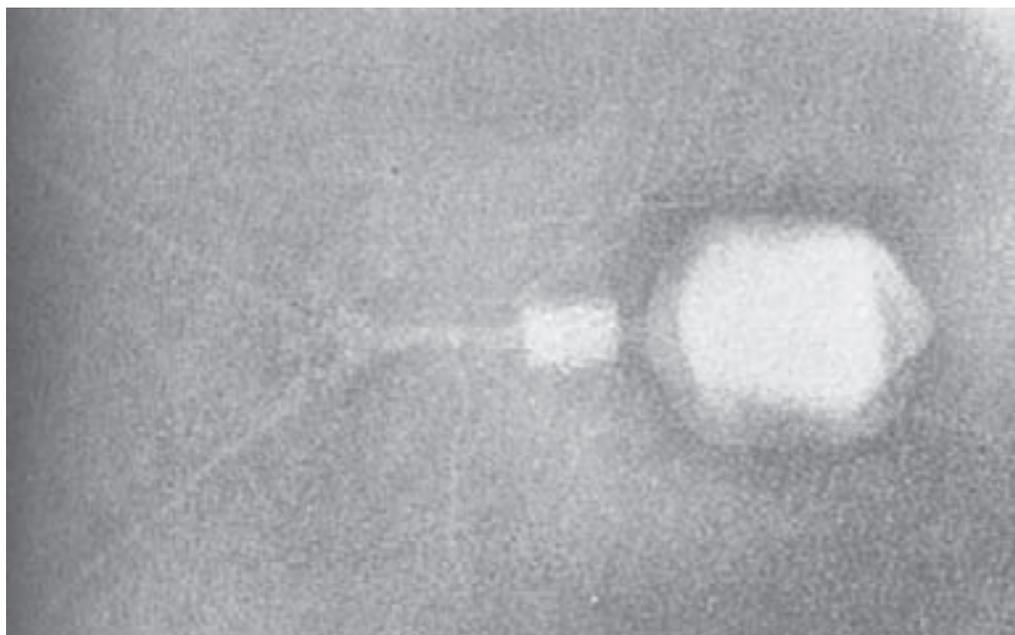
A reacção decorre em três fases. Na primeira (i) é formado um complexo entre a enzima (E) e o grupo adenílico do ATP, de que resultam as moléculas de pirofosfato (PPi) e o complexo enzima-adenosina-monofosfato (E-AMP); na segunda (ii), o grupo adenílico liga-se à extremidade livre 5'-fosfato do DNA, com o qual forma uma ligação fosfodiéster, após o que (iii) se liga ao grupo 3'-hidroxilo livre de outra cadeia de DNA, com recuperação final da enzima e AMP livre:



Como resultado, as extremidades dos fragmentos de DNA são unidas entre si desde que esteja presente uma cadeia complementar intacta de DNA (que lhes serve de modelo).

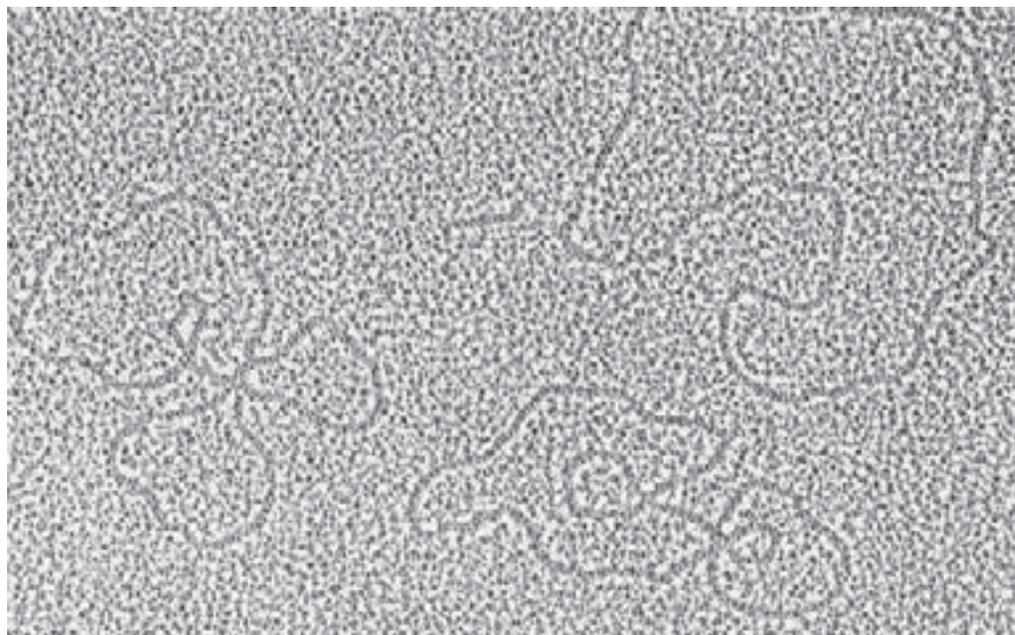
A reconhecida simplicidade dos vírus e a particularidade de acção dos bacteriófagos estimula a sua utilização preferencial nos estudos de identificação dos mecanismos e informação genética. Por observação ultramicroscópica, ve-

rificou-se que as moléculas circulares de DNA viral estavam dispostas habitualmente na região da “cabeça” do vírus. A disposição circular forneceria ao DNA alguma protecção contra replicações indesejadas (Fig. 3.8.). Para penetrar no hospedeiro a forma circular de DNA viral transforma-se em linear o que também sucede durante a divisão celular, em que cada cromossoma só se replica uma vez.



**Fig. 3.8.** Bacteriófago  $T_2$ , ampliado 500.000 diâmetros. A cabeça, que constitui o segmento mais volumoso, contém toda a informação genética (sob a forma de uma molécula gigante de DNA) necessária para a sua replicação na célula hospedeira. Logo que a cauda do vírus adere à parede da bactéria hospedeira, esta é invadida pelo DNA viral que utiliza imediatamente o equipamento do hospedeiro para reproduzir cerca de 100 cópias de fago completo, as quais passam para o meio externo por lise da célula parasitada. Reprodução autorizada pelos autores, Dr. S. Brenner e Dr. R.W. Horne (University of Cambridge, UK).

Utilizando como modelo uma das cadeias (+) de DNA circular do bacteriófago  $\Phi X174$ , marcada com trítio foi possível sintetizar em tubo de ensaio uma molécula de DNA com enzimas e substractos muito purificados (Fig. 3.9.). Na presença de DNA polimerase e de uma mistura de nucleósidos-trifosfato

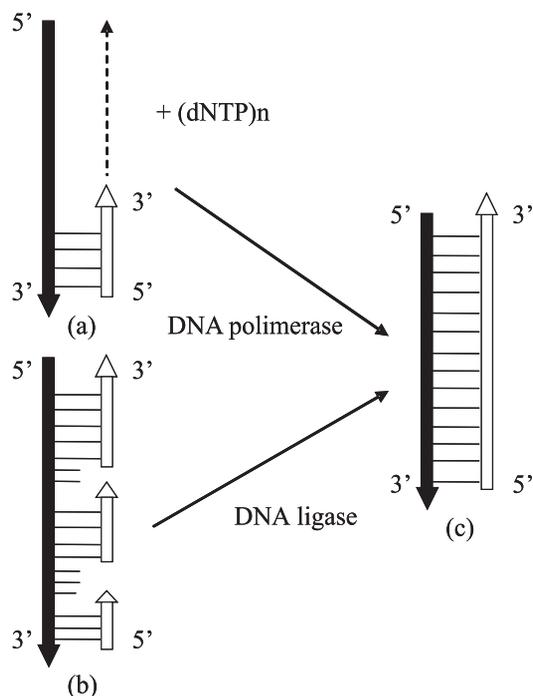


**Fig. 3.9.** Três duplex de DNA circular, em que uma das cadeias de cada par foi extraída do bacteriófago  $\Phi$  x174 (ampliação em cerca de 200.000 diâmetros). Em meio com precursores e enzimas adequadas à replicação, a cadeia simples do DNA do fago serviu de modelo à síntese de uma cadeia complementar. Os duplex híbridos resultantes (cada um com cerca de 5.000 pares de bases) evidenciaram actividade biológica.

Reprodução autorizada pelo autor, Dr. Arthur Kornberg (Department of Biochemistry, Stanford University School of Medicine, USA).

marcados com  $P^{32}$  (dATP, dCTP, dGTP e dBUTP<sup>3</sup>, foi obtida uma cadeia de DNA linear complementar (-), em que todos os resíduos de timina estavam substituídos por 5-bromo-uracilo. Por acção de outra enzima específica, a *DNA ligase*, a faixa (-) linear era encerrada em ansa fechada, originando um duplex de DNA constituído pela faixa (+) marcada com trítio e a faixa (-) com  $P^{32}$ . A união ocorria entre a extremidade 3'-hidroxilo de um fragmento (ou extremidade de cadeia) e o grupo 5'-fosfato do fragmento (ou extremidade) adjacente. A enzima permanecia transitoriamente unida a esta última extremidade, activando o grupo fosfato enquanto decorria a reacção (Fig. 3.10.).

<sup>3</sup> dBUTP-2-desoxi-5-bromo-uridina 5'-trifosfato, análogo do dTTP, com 5-bromo-uracilo no lugar da timina.



**Fig. 3.10.** A replicação de cada cadeia de DNA é assegurada por duas enzimas, a DNA polimerase (que sintetiza fragmentos de polinucleótidos complementares) e a DNA ligase (que, ao unir aqueles fragmentos sequenciais, estabelece a continuidade da cadeia complementar).

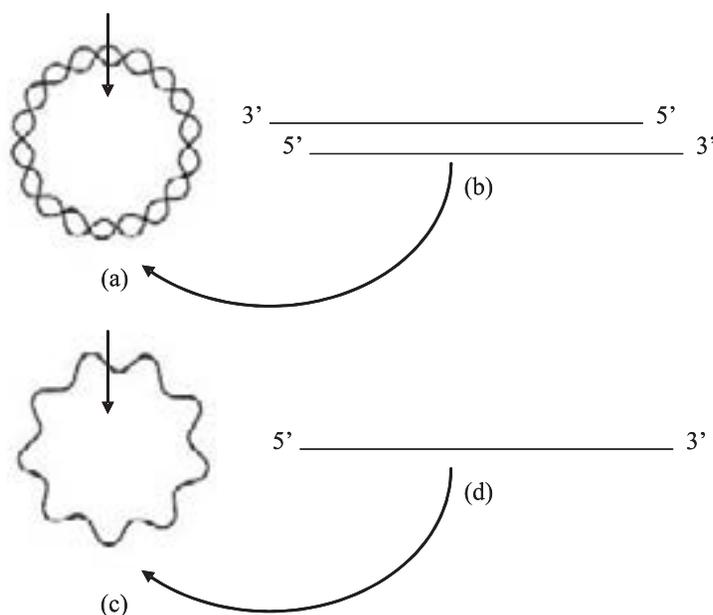
Seguidamente, foi induzida a rotura enzimática (pela desoxirribonuclease) em pontos vários da dupla hélice circular, de que resultaram algumas cadeias intactas, (-) ou (+), circulares ou lineares, e fragmentos daquela formação.

Isolando as faixas (-) circulares intactas de modo a servirem de modelo à síntese das faixas complementares (+), foi verificado (num meio com DNA polimerase, DNA ligase e os quatro tipos de desoxiribonucleósido 5-fosfato habituais) a formação de uma cadeia (+) circular, idêntica à faixa (+) original, que também mantinha a capacidade infectante.

Não obstante ter sido possível demonstrar a síntese *in vitro* do DNA pela polimerase específica em organismos com um só cadeia de DNA circular, continuava por esclarecer o processo de replicação da molécula duplex de DNA.

Um pormenor significativo a salientar daquelas experiências havia sido a necessidade de romper a estrutura circular do (-) DNA antes de viabilizar a obtenção da cadeia (+) final; enquanto se mantivesse em anel fechado não

havia replicação da cadeia (-). Daqui foi concluído haver necessidade de enzimas específicas, as *endonucleases*, que, ao actuarem em um ou mais pontos do cromossoma circular, o transformavam na forma linear por rotura de algumas pontes fosfodiéster (Fig. 3.11.). Porém, para actuarem, as nucleases teriam de reconhecer a sequência específica de nucleótidos próxima do seu local de acção. Este mecanismo funcionaria igualmente na forma linear do DNA. Em qualquer dos casos, a polimerase de DNA começaria a actuar somente nos locais de rotura das pontes fosfodiéster, onde se inicia a replicação. Não havendo ruptura na(s) cadeia(s) de DNA não existe replicação.



**Fig. 3.11.** Disposição circular de moléculas de DNA em duplex (a) ou monocadeia (c). A forma circular (duplex ou simples) converte-se em linear (b e d) por ruptura das ligações fosfodiéster em determinado local da sequência (assinalado com seta). A forma linear constitui uma adaptação necessária à sua passagem para a célula hospedeira, após o que é retomada a forma circular.

Por autorradiografia e microscopia electrónica não se confirmou que as duas faixas de DNA duplex tivessem de se separar completamente uma da outra para serem replicadas; na realidade, à medida que se replicavam, as cadeias afastavam-se entre si, reaproximando-se depois da replicação estar concluída.

Experiências engenhosas permitiram concluir que a replicação do duplex circular da *E. coli* começa num determinado local (*origem* da replicação), a partir do qual progride bidireccionalmente até ao ponto de terminação, a par com o desenrolamento do duplex. O ponto de origem da replicação é sinalizado por uma incisão no esqueleto fosfato-ribose, com exposição de um grupo 3'-hidroxilo livre ou pelo desenrolamento da duplex, induzida por determinada molécula de natureza proteica. Em ambos os casos é estimulada a síntese de um fragmento de RNA complementar de determinado segmento da cadeia a replicar.

Ainda que inicialmente fosse sugerido que a replicação decorreria sempre na mesma direcção enquanto o núcleo de crescimento se desloca, passou a admitir-se que o crescimento das cadeias copiadas decorre simultaneamente em sentidos opostos (horário e anti-horário) a partir da origem da replicação, à medida que a duplex se desenrola (Fig. 3.12.). Na *E. coli*, cada nova cadeia de DNA é formada a velocidade aproximada de  $27\mu\text{m}^{-1}$ , a que correspondem cerca de  $9 \times 10^4$  ligações fosfodiéster por minuto.

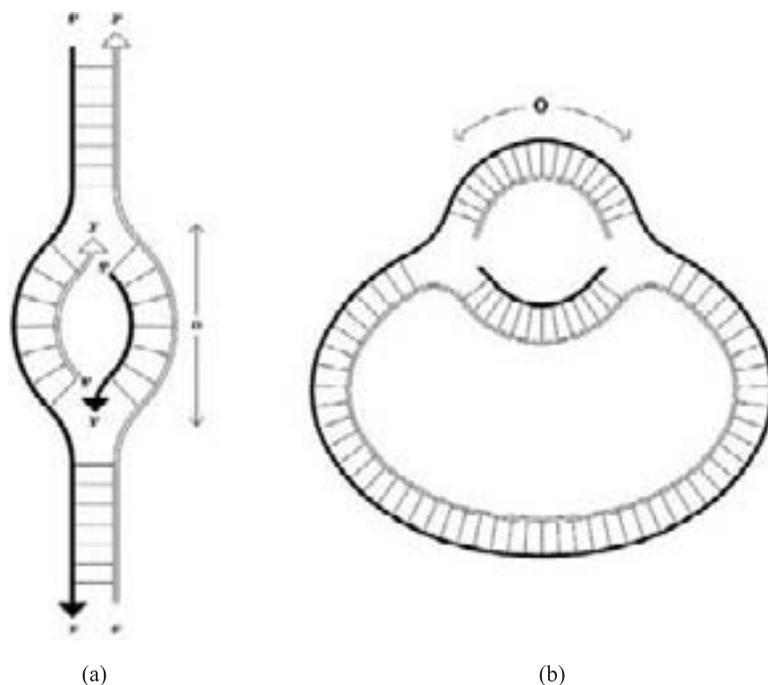
Durante uma única divisão celular (que decorre em cerca de 30 minutos) a separação de ambas as cadeias prossegue de modo divergente (em Y) no ponto de origem da replicação (ou *garfo replicativo*), até que sejam formados dois duplex circulares por cada uma das cadeias iniciais.

Um dos problemas que se colocava era o de saber onde e como se iniciaria a réplica de cada cadeia. É fácil de entender o crescimento da cadeia complementar que progride na direcção 5'→3' mas não o é tanto explicar a replicação (5'→3') da outra nova cadeia (Fig. 3.5.).

Para que o duplex seja replicado simultaneamente em ambas as suas faixas é necessário, em princípio, que o crescimento de uma das duas cadeias complementares decorra num sentido e o da outra no sentido inverso mas, em qualquer dos casos, sempre na direcção 5'→3'. Não foi identificada nenhuma enzima capaz de replicar de 3' para 5'.

Considerando que a replicação é simultânea nas duas cadeias replicadas e atendendo à direcção única em que decorre (de 5' para 3'), tornava-se difícil encontrar um mecanismo consentâneo, accionado por uma única molécula de DNA polimerase. A alternativa consistiria em admitir duas moléculas daquela enzima que actuassem em simultâneo mas em sentidos opostos.

A resposta aquele problema foi sugerida em mutantes de *E. coli* deficientes em DNA ligase, nos quais se observou acumulação de pequenos fragmentos de DNA nas proximidades dos pontos de crescimento; muitos dos nucleótidos que seriam



**Fig. 3.12.** Representação esquemática da replicação do duplex linear (a) circular (b). A replicação decorre simultaneamente em direcções opostas em ambas as cadeias (do duplex e também do DNA circular simples) a partir de um ou mais pontos de origem (O) iniciais. Havendo mais de uma origem, a replicação progride até encontrar outro “garfo replicativo”, e assim sucessivamente. Em qualquer dos casos, a DNA polimerase catalisa a união dos novos nucleótidos à extremidade 3'-hidroxilo da cadeia em crescimento.

incorporados normalmente nas cadeias polinucleotídicas, agregavam-se em fragmentos curtos e complementares a uma ou a outra das cadeias-mãe. Por outro lado, as manipulações experimentais do DNA, por muito cuidadosas que fossem, não evitavam o aparecimento de pequenos fragmentos semelhantes aos observados naqueles mutantes. Foi então sugerido que a replicação de ambas as cadeias de duplex original envolvia a formação de pequenos polinucleótidos como precursores do DNA complementar; os fragmentos dispunham-se ao longo da cada uma das cadeias a ser replicada, as quais seriam complementares e antiparalelas mas, necessariamente, orientadas na direcção 5'→3'. Por acção da DNA ligase aqueles fragmentos acabariam por ser unidos, progressivamente e em continuidade, dando origem a uma cadeia completa e complementar.

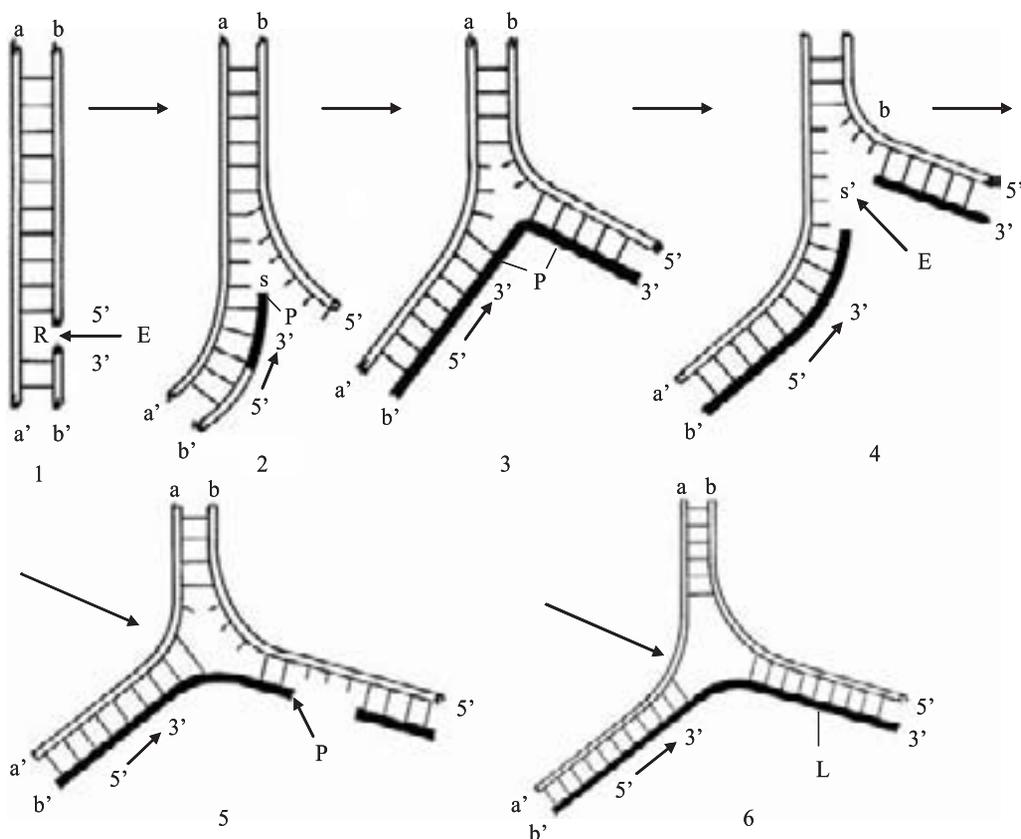
Em estudos subsequentes, Kornberg e Cols elaboraram uma hipótese de replicação do DNA que ocorreria não só *in vitro* mas também em *in vivo*, em qualquer forma de DNA (circular ou linear, duplex ou em cadeias simples). O mecanismo basear-se-ia em três tipos de enzima: DNA polimerase, DNA ligase e endonuclease. O processo da replicação seria activado pela secção de uma das cadeias de DNA por uma endonuclease; no local de rotura do esqueleto fosfato-pentose, seria iniciado imediatamente o crescimento da cadeia interrompida, a partir do grupo 3'-hidroxilo do sector seccionado, pela adição de sucessivos resíduos de mononucleótido, complementares aos da cadeia intacta, e unidos entre si por ligações 3',5'-fosfodiéster. A DNA polimerase, ao ligar-se ao grupo 3'-hidroxílico livre da secção interrompida, adicionaria na direcção 5'→3' sucessivos resíduos de mononucleótidos, complementares aos da cadeia intacta e unidos entre si por ligações 3',5'-fosfodiéster. Entretanto, a outra extremidade (5') da cadeia seccionada tenderia a afastar-se cada vez mais do alinhamento original à medida que a (nova) cadeia replicada avança. Enquanto o duplex se desenrolava, o segmento 5' da cadeia seccionada prepara-se-ia para ser também replicado (Fig. 3.13). O local em que as cadeias de DNA duplex se desenrolam, ou em que os dois segmentos da cadeia seccionada se afastam para serem replicados, é denominado *garfo de replicação ou garfo replicativo*.

Seguidamente, a DNA polimerase transitará da extremidade 3' da cadeia **b'** em replicação para o fragmento **b** (em s, na Fig. 3.13.) onde continua a adicionar mononucleótidos no sentido 5'→3', utilizando a faixa seccionada **b** como modelo. A cópia termina quando o novo segmento está completamente replicado e unido à outra extremidade **b'**.

Uma endonuclease rompe novamente a faixa recém-formada no ponto de separação (s'), sendo o processo repetido: replicação da cadeia a partir da extremidade 3' de **b'** e, ao atingir o garfo de replicação, prossegue a replicação do fragmento **b** até se aproximar do segmento sintetizado anteriormente. A DNA ligase acaba por unir os dois fragmentos replicados na faixa **b**.

O processo seria repetido o número de vezes necessário até que as duas cadeias do duplex original estivessem integralmente replicadas por acção das três enzimas (DNA polimerase, DNA ligase e endonuclease).

O modelo descrito apenas explica a replicação num só sentido, não esclarecendo como é que a forma linear de extremidade livre 5' recupera a configuração no tipo de DNA em ansa fechada.



**Fig. 3.13.** Hipótese de modelo de replicação do DNA. O processo iniciar-se-ia (1) pela clivagem da ligação fosfato-ribose em determinado sector de uma das cadeias do duplex DNA, catalisada por uma endonuclease (E); (2) As cadeias do duplex separam-se entre si, por ruptura das ligações hidrogeniónicas entre as bases complementares, após o que cada cadeia serve de modelo à síntese de uma nova cadeia, que lhe é também complementar. A extremidade 5' do segmento seccionado afasta-se do duplex enquanto que a polimerase do DNA (P), de imediato unida à extremidade 3' hidroxilo do sector seccionado, replica a cadeia intacta contra-lateral, na direcção 5' → 3', adicionando e unindo entre si sucessivos resíduos de mononucleótidos até encontrar o ponto de separação (s) do outro segmento da cadeia seccionada; (3) A DNA polimerase (P) muda a sua actividade catalítica da cadeia a para a b, replicando-a até à extremidade 5'; (4) A endonuclease (E) secciona novamente a cadeia replicada, no garfo replicativo (s'); (5) A DNA polimerase (P) regressa à cadeia para a prolongar a extremidade 3' da cadeia já replicada, e, depois do garfo, sintetiza um outro segmento complementar da cadeia b até à proximidade do segmento anterior; (6) Segue-se a união entre os dois segmentos de cadeia b' por acção da enzima DNA ligase (L).

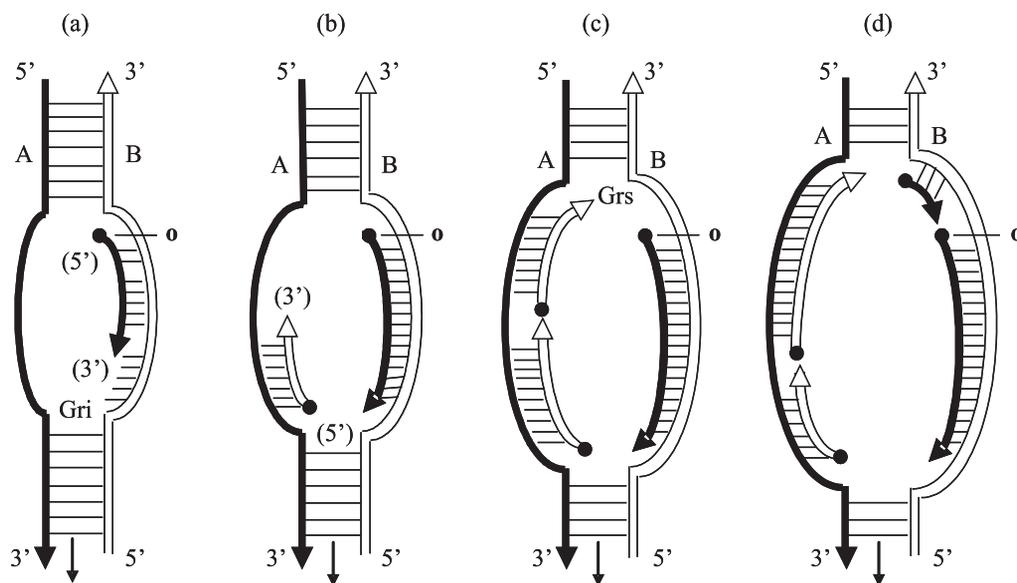
Porém, supondo que aquele mecanismo ocorre em sentido oposto, obtêm-se soluções viáveis para o problema. Há que admitir que a faixa aa' é cortada por uma endonuclease específica num ponto muito próximo ao da primeira nucleasa. O crescimento da extremidade 3' poderá ocorrer da mesma maneira, mas em sentido oposto. À medida que a extremidade 5' se afasta progressivamente do seu plano habitual juntam-se fragmentos nucleotídicos complementares, que acabam por ser unidos pela DNA ligase. As duas extremidades soltas têm sequências complementares de modo que, a certa altura, também se ligam através das ligações hidrogeniônicas habituais. Nucleases específicas asseguram o seccionamento de sucessivas ligações, indispensáveis à separação recíproca das cadeias, para que a replicação possa prosseguir. No fim do processo, resulta um duplex com o dobro da dimensão original que, por acção de endonucleases, é transformado em duas faixas duplas, lineares e complementares entre si, que poderão ou não fechar-se.

Outra hipótese admite que a replicação iniciada por determinadas proteínas que sinalizariam as regiões onde tem origem e onde termina a replicação. Nestes locais de origem assinalados, o duplex começaria a desenrolar-se (desnaturação transitória do DNA) com separação das duas cadeias, originando um garfo replicativo. Nos cromossomas lineares de eucariotas a replicação do DNA tem múltiplas origens. Observações ultramicroscópicas de DNA extraído de núcleos em divisão rápida da *Drosophila melanogaster* revelaram a existência de múltiplas "bolsas" de replicação (*replicões*), representando os sectores de encontro das duas cadeias em replicação em sentido oposto, delimitadas por dois garfos replicativos.

Em cada garfo replicativo coexistem duas novas cadeias em crescimento, em sentido oposto mas, naturalmente, na direcção 5' → 3' (Fig. 3.14.).

Uma das cadeias replicadas cresce rápida e continuamente (*cadeia líder*) acompanhando a progressão do garfo de crescimento. Em contrapartida, a outra cadeia (*cadeia atrasada*) com progressão mais lenta e descontínua cresce sob a forma de pequenos segmentos de DNA, também na direcção 5' → 3' mas em sentido oposto ao da progressão do garfo replicativo próximo da sua origem. Os segmentos serão unidos entre si pela DNA ligase.

Enquanto a cadeia que foi replicada continuamente (*cadeia líder*) é, em princípio, originada a partir do segmento pré-existente de DNA (*DNA nascente*), a molécula iniciadora da cadeia atrasada é representada por um segmento de RNA (com cerca de 4 a 12 nucleótidos de extensão) sintetizado a partir da



**Fig. 3.14.** Hipótese de replicação bidireccional em eucariotas a partir de um segmento de RNA iniciador. (a) A replicação começa em determinados locais ao longo do duplex de DNA, por acção de determinado sinal que começa por induzir a desnaturação do DNA. No esquema é assinalado somente um ponto de origem (o), em que a replicação é iniciada por pequenas moléculas de RNA (●), após o que a DNA polimerase adiciona sucessivas moléculas de mononucleósido-trifosfato (com perda de uma molécula de PPI por cada ligação fosfodiéster formada), sempre na direcção 5'→3'. A replicação forma uma cadeia contínua na direcção do garfo replicativo inferior (Gri). (b) A definição de um ponto de origem numa das cadeias transmite-se imediatamente à cadeia contralateral, na qual a DNA polimerase sintetiza, com atraso relativo e contra o sentido de crescimento inicial, pequenos fragmentos de DNA complementares aos daquela cadeia, também iniciados por uma molécula de RNA iniciador (c) Progressivamente, os novos segmentos de DNA replicados da cadeia A são unidos entre si pela DNA ligase. Assim que a frente de replicação ultrapassa o ponto de origem daquela cadeia é criado um novo garfo replicativo superior (Grs). (d) Simultaneamente, a DNA polimerase recomeça a sintetizar novamente na cadeia B, originando um novo segmento de DNA, também iniciado por RNA que se alonga até à proximidade da cadeia replicada anteriormente.

cadeia-mãe do DNA replicada. A formação desta molécula de RNA (*RNA iniciador*) DNA-dependente ocorre ao longo da cadeia a ser replicada, originando sucessivos pontos de origem (o) da cadeia atrasada de DNA. Esta é sintetizada pela DNA polimerase por adição sucessiva de desoxirribonucleótidos a partir do RNA iniciador, originando segmentos que, no total, atingem cerca de 1.000

nucleótidos. As moléculas constituídas pelo RNA iniciador e DNA nascente foram observadas pela primeira vez por Reiji Okazaki e Cols, sendo desde então designados por *fragmentos de Okazaki*.

Assim que a extremidade 3' de cada fragmento em crescimento atinge a proximidade do grupo 5'-fosfato do que o antecede, a DNA polimerase, pela sua outra actividade catalítica própria das exonucleases, remove o RNA iniciador que é de seguida degradado. De imediato, o espaço entre os sucessivos fragmentos de DNA é preenchido, pela DNA polimerase, com nucleótidos complementares da cadeia original, após o que a DNA ligase estabelece a continuidade entre os sucessivos fragmentos complementares de DNA.

Observações subsequentes revelaram a existência na *E. coli* parasitada com fagos M13 e  $\phi$ X174, não uma mas três isoformas enzimáticas da DNA polimerase (tipos I, II e III). A enzima funcional no garfo replicativo seria a DNA polimerase III. O tipo I (que possui acção semelhante à do tipo III mas é menos eficiente) parece evidenciar particular especificidade para o preenchimento de intervalos vazios na cadeia em replicação. Está por esclarecer a função do tipo I. A acção do tipo III da DNA polimerase dependeria da presença de uma outra proteína não catalítica, designada *co-polimerase III*. O “sistema polimerase III” (enzima e proteína não catalítica) iniciaria a polimerização logo que uma RNA polimerase sintetizasse a molécula de RNA que actua como seu iniciador. Outras proteínas não-enzimáticas com acção pouco conhecida participariam também na replicação do DNA, conferindo a esta operação características de grande complexidade.

Muito provavelmente, aquela complexidade dever-se-á à necessidade de um controlo rigoroso sobre os locais de iniciação e terminação da replicação do DNA, de modo a garantir a homogeneidade da produção em sucessivas gerações celulares e, portanto, para que o risco de mutações seja reduzido ao mínimo possível.

## **Reparação do DNA**

Todas as células, de bactérias a organismos superiores, estão constantemente sujeitas a diversos tipos de radiação e a outros agentes agressivos.

Quanto mais preparada estiver a célula para responder aquelas agressões maiores possibilidades terá de lhes sobreviver, o que deve ter sido a base de

selecção durante o processo evolutivo das espécies. Em contrapartida, não sendo corrigidos, aqueles erros tendem a acumular-se nas células somáticas ao ponto destas perderem a sua funcionalidade, enquanto as células germinais serão a fonte de múltiplas mutações que acabam por inviabilizar a procriação.

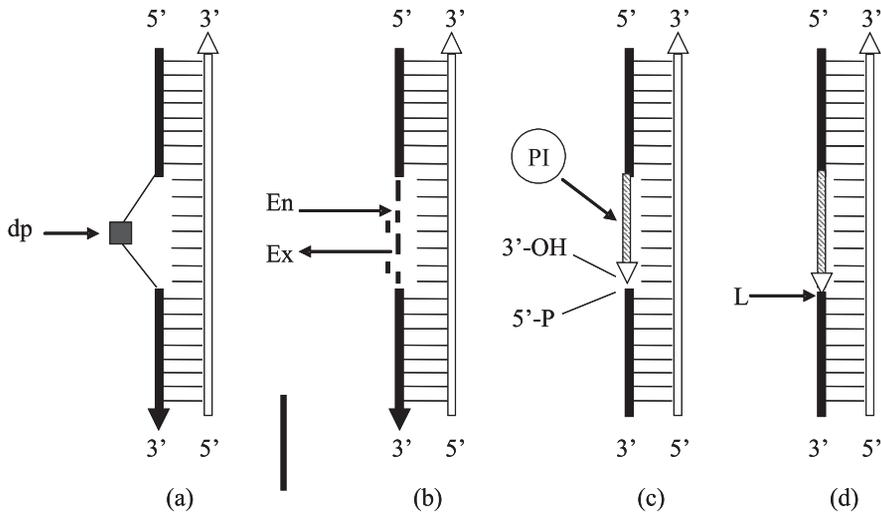
Entre as radiações mais lesivas do DNA contam-se os ultravioletas e os raios X, embora pequenas modificações de ambiente intracelular também possam alterar a estrutura de DNA.

Nas bactérias e outros eucariotas mais simples as “lesões” que não interferem na composição de nucleótidos parecem ser imediatamente reparadas por sistemas fotolíticos, que removem directamente as ligações anormais que originam dímeros primários.

As radiações ultravioleta são muito lesivas. Pela sua acção, as bases pirimídicas adjacentes (sobretudo a timina) poderão unir-se por ligações covalentes, originando dímeros inactivos na replicação, com consequências potencialmente letais.

A maioria das lesões de DNA podem ser directamente corrigidas com a intervenção de sistemas enzimáticos intracelulares, que actuam numa sequência interventiva de excisão-reparação. Desde que uma das cadeias continue intacta, a que estiver lesada poderá ser fácil e completamente reparada, não restando qualquer anomalia. Primeiro intervêm as endonucleases nas alterações reconhecidas, removendo-as por hidrólise das ligações internucleotídicas adjacentes aos dímeros. Seguidamente actuam as exonucleases, cuja acção visa remover os dímeros anormais, a par de alguns dos nucleótidos intactos adjacentes a que se encontram ligados. A actividade catalítica das exonucleases e da DNA polimerase I parece ser fundamental nesta operação. É possível que, na presença de uma base incorrectamente posicionada durante a síntese de DNA, a polimerase I depois de a remover, recopie o segmento de forma correcta. O espaço resultante da excisão é preenchido por resíduos de desoxirribonucleósido 5'-trifosfato, complementares da outra cadeia, por acção da DNA polimerase, a partir da extremidade-3' livre. A DNA ligase conclui a reparação da cadeia lesada (Fig. 3.15.).

O isolamento de um mutante de *E. coli* sem DNA polimerase, mas ainda assim capaz de replicar normalmente o seu DNA, suscitou dúvidas sobre a sua especificidade. Porém, sendo aquele mutante muito susceptível às radiações ultravioletas, foi referida a existência de pelo menos duas enzimas, uma interveniente na replicação e outra na reparação de DNA, ou apenas uma para assegurar a viabilidade celular.



**Fig. 3.15.** Mecanismos de reparação do DNA, por intervenção sequencial de três tipos de enzimas: nucleases (endo- e exo-), DNA polimerase e DNA ligase.

(a) Em uma das cadeias do duplex foi formado (por exemplo) um dímero de pirimidinas (dp) entre dois nucleótidos adjacentes, que altera a disposição estrutural e estabilidade no sector adjacente do duplex, além de ser uma fonte potencial de erros de replicação; (b) Uma endonuclease (En) ao reconhecer o sector lesado, corta ligações fosfodiéster a montante e a juzante, isolando um segmento de polinucleótidos; (c) Os nucleótidos excisados são removidos por uma exonuclease (Ex), (que é também uma das actividades da DNA polimerase I); (d) Seguidamente, a DNA polimerase I (PI) reconstrói na cadeia sectionada o segmento correcto, replicado da cadeia contralateral intacta); (d) a DNA ligase (L) une a extremidade 3'-hidroxilo do novo segmento com a extremidade 5'-fosfato da porção sectionada.

## Mutações

Como já foi referido, os genes são constituídos por sequências variáveis de mononucleótidos ligados entre si por pontes fosfodiéster. Tendo como base os estudos realizados em cromossomas virais e bacterianos, calcula-se que um gene contenha (em média) entre 900 a 1.500 pares de mononucleótidos. Sendo os resíduos de fosfato e desoxirribose constantes ao longo das moléculas de DNA, é de supor que sejam as bases (A, G, C, T) que lhes conferem características específicas, designadamente na codificação de polipéptidos pré-determinados. Ainda que sejam apenas quatro os tipos de mononucleótidos principais, são inúmeras

ras as possíveis mensagens genéticas que podem codificar, consoante a extensão da molécula de DNA e as respectivas sequências nucleotídicas (ver Capítulo 6).

As alterações hereditárias do material genético são designadas de *mutações*. Numa perspectiva global, as mutações possibilitam a evolução dos organismos, a sua adaptação a condições diferentes do meio em que vivem, e aumentam a probabilidade da sua sobrevivência. Na origem das mutações podem estar os efeitos de agentes físicos ou químicos ou, ainda, defeitos de recombinação cromossómica. As mutações génicas são classificadas em dois grandes tipos: as *mutações pontuais*, que resultam de modificações na estrutura de um gene, em geral restrictas a um ou dois pares de nucleótidos, e as *mutações (ou aberrações) cromossómicas*, que afectam segmentos ou toda a estrutura dos cromossomas (Quadro 3.1.).

As mutações pontuais têm, por conseguinte, uma base molecular, habitualmente resultante de uma falha no processo de replicação e ou de reparação do DNA. Ao perpetuarem-se, originam anomalias nas sequências dos nucleótidos (do DNA replicado e dos diversos tipos de RNA transcrito) e dos polipéptidos sintetizados, o que, na generalidade, é a causa de limitações funcionais ou, em situações extremas, inviabilizam a sobrevivência dos organismos.

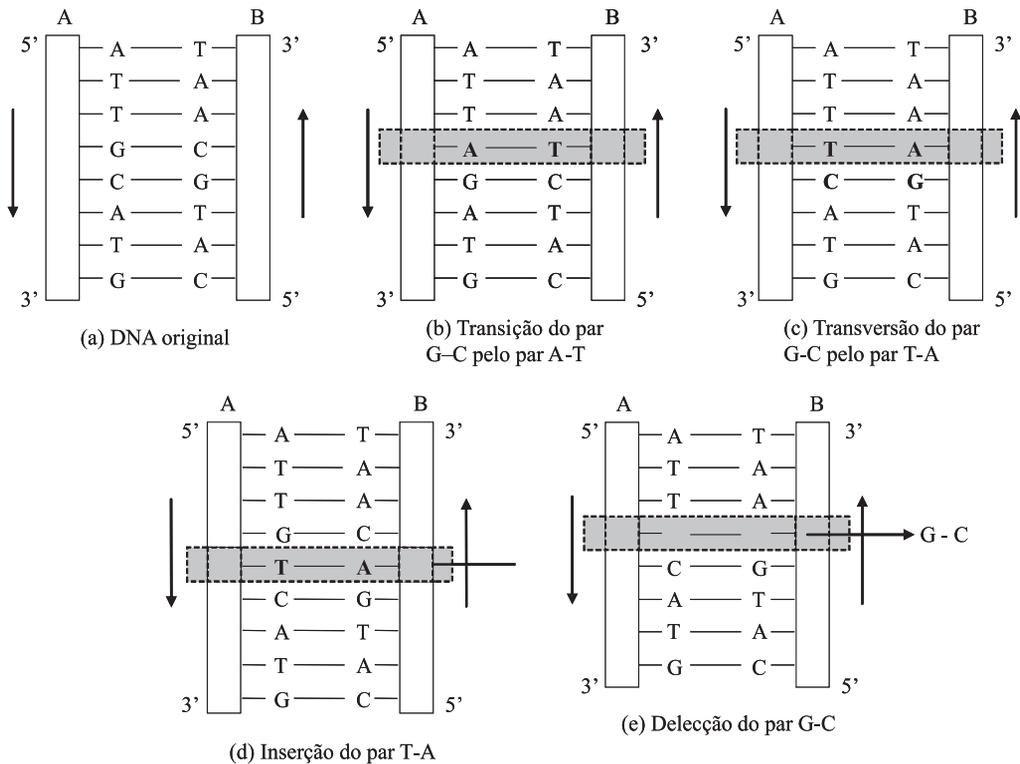
**Quadro 3.1.** Tipos principais de mutações

<b>Mutações genéticas (pontuais)</b>	
Substituição:	Transcrição Transversão
Alteração do total de nucleótidos:	Delecção Inserção
<b>Aberrações cromossómicas mais frequentes</b>	
Alterações estruturais:	Delecção Duplicação Inversão Translocação
Alterações no número de cromossomas:	Poliploidia Aneuploidia

Uma fracção importante das mutações pontuais parece resultar da instabilidade das bases nucleotídicas do DNA que, espontaneamente ou por tautomerização, originam modificações na sua estrutura (p.ex., a forma cetónica normal da timina

ou da guanina é convertida na forma enólica, rara; ou a forma amina habitual da adenina ou da citosina, é transformada em forma imina, rara) e, por consequência, originam desvios no emparelhamento (p. ex., G-T ou A-C). Por exemplo, se ao ser replicada, a cadeia que contém adenina em lugar da guanina originar um par mutante A-T em lugar do G-C, passa a haver um par errado de nucleótidos numa das duas moléculas do DNA replicado. Se estas modificações isoméricas não forem reparadas antes da replicação do DNA podem converter-se em hereditárias.

A substituição de bases na estrutura de DNA replicado por aquele tipo de alterações isoméricas origina dois tipos de mutações, as *transições* (substituição de uma purina por outra purina ou de uma pirimidina por outra pirimidina) e as *transversões* (quando uma purina é substituída por uma pirimidina ou vice-versa (Fig. 3.16.).



**Fig. 3.16.** Principais tipos de mutação pontual, relativamente ao modelo (a) “selvagem” (DNA original): (b) transição, (c) transversão, (d) inserção e (e) delecção.

Outro tipo de mutações pontuais, frequentemente provocados por agentes físicos e químicos e que também tende a transmitir-se de modo hereditário, é o das *inserções* e *deleções*. As inserções resultam da adição (intercalação) de um ou mais mononucleótidos (bases) suplementares em determinado ponto da sequência do DNA. A deleção (também denominada *contração*), caracteriza-se pelo remoção de um ou mais nucleótidos na molécula de DNA. As inserções e as deleções inseridas na estrutura de determinado gene alteram a sequência de leitura do cistrão e, por consequência, a sequência dos aminoácidos no polipéptido a ser codificado na ausência deste tipo de mutação (mutação *frameship*, isto é mudança da “janela de leitura”). As consequências deste tipo de mutação “frameship” são potencialmente mais graves do que as anteriores, ao originar genes parcial ou completamente inoperacionais, com resultados por vezes letais.

Nas mutações pontuais por substituição de um nucleótido em determinado gene são identificáveis três tipos de consequências; ou não há alterações (mutação *samesense*), ou as consequências são irrelevantes (*nonsense*) ou o polipéptido produzido conduz a actividades distintas (*missense*). Em geral, as referidas mutações por substituição têm consequências benignas ou passam despercebidas, daí serem designadas por *mutações silenciosas*.

Nas mutações *samesense* está alterada a terceira base de um determinado tripleto que, como será demonstrado, tem menor significado na leitura do código genético (ver Capítulo 6). As mutações *nonsense* originam um (ou mais) dos codões de terminação (UAG, UAA, UGA), o que se traduz na síntese de polipéptidos mais curtos do que o previsto e, por isso, desprovidos de actividade normal. As mutações *missense* alteram o significado de um codão, do que resulta a síntese de um polipéptido com iguais dimensões ao normal mas funcionalmente diferente. O exemplo mais referido é o da hemoglobina S (que está na origem da anemia de células falciformes), a qual difere da hemoglobina A (do adulto normal) pela substituição do ácido glutâmico na 6.<sup>a</sup> posição da cadeia  $\beta$  por um resíduo de valina. Esta alteração deve-se à permuta de um nucleótido com timina por outro com adenina na 2.<sup>a</sup> posição do 6.<sup>o</sup> codão da cadeia de DNA transcrita.

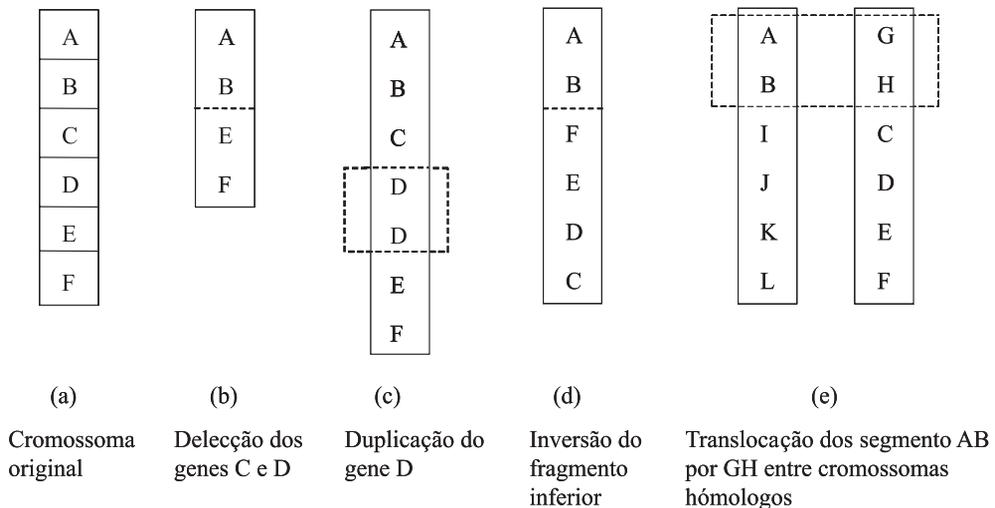
As aberrações cromossómicas são representadas por alterações estruturais no número de cromossomassendo as mais frequentes as seguintes (Fig. 3.17.): (i) *deleção* (por perda de um ou mais genes no segmento cromossómico), (ii) *duplicação* (por coexistência de mais de uma cópia do mesmo gene, no mesmo

ou diferentes cromossomas), (iii) *inserção* (colocação invertida de um segmento cromossômico, por defeito da recombinação) e (iv) *translocação* (por troca de segmentos entre cromossomas não homólogos).

As alterações no número de cromossomas subdividem-se em dois tipos: *poliploidia* (quando as células têm mais de dois cromossomas) e *aneuploidia* (por distribuição desigual de cromossomas nas células, devido à não-disfunção cromossômica ou cromatídica durante a meiose).

Um gene pode conter muitos pontos mutáveis, teoricamente tantos quanto o número de mononucleótidos que o constituem. Nem todas as mutações originam alterações visíveis nas proteínas que codificam e, portanto no fenotipo. Porém, exceptuando as originadas por anomalias cromossômicas e as originadas em sectores predispostos, o total das mutações espontâneas (pontuais) é reduzido, cerca de  $10^{-6}$  cada vez que um gene é duplicado, ou  $10^{-9}$  se se considerar o erro a nível de cada nucleótido. Na presença de agentes químicos (p.ex., agentes alquilantes) ou físicos (p.ex., radiações ultravioletas e ionizantes), que são genericamente designadas de *mutagêneos*, torna-se muito mais elevada a taxa de mutação induzidas.

As mutações cromossômicas são frequentemente provocadas por erros de *recombinação cromossômica* (“crossing-over”).



**Fig. 3.17.** Principais tipos de aberrações cromossômicas; (a) localização e sequência dos genes de um cromossoma hipotético; (b) delecção; (c) duplicação; (d) inversão e (e) translocação.

## Recombinação cromossômica

A recombinação cromossômica é um processo natural decorrente na divisão celular que, em geral, beneficia a espécie ao aumentar a diversidade do genotipo dos descendentes. Os genes favoráveis tendem a associar-se e a serem perpetuados enquanto o meio não se alterar. Adicionalmente, a recombinação cromossômica na meiose poderá separar ou reconstruir fragmentos de DNA original.

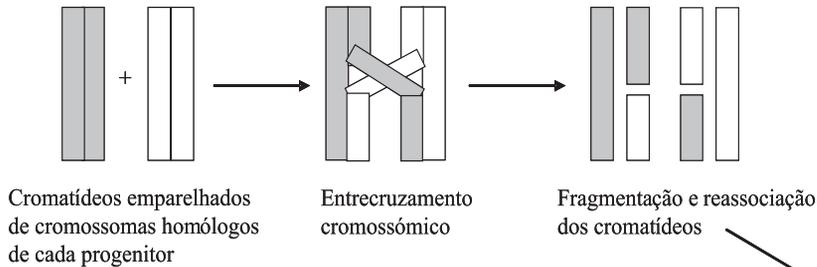
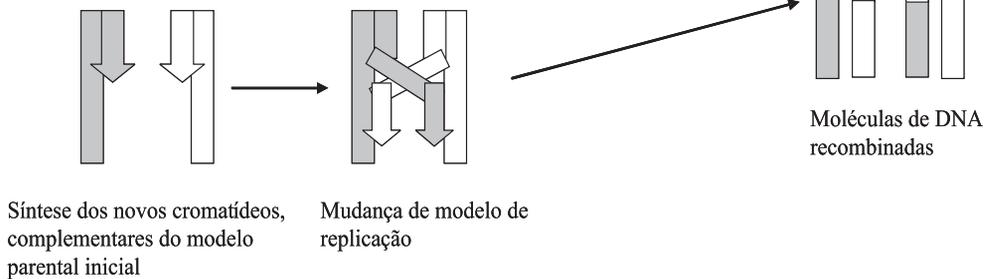
As recombinações são de dois tipos principais: as que modificam a sequência relativa nos fragmentos permutados entre cromossomas diferentes que possuem locais específicos de recombinação (*recombinação sitio-dependente*) e as que ocorrem entre fragmentos homólogos sem localização específica, embora mantendo o ordenamento dos genes nos respectivos cromossomas (*recombinação sitio-independente, homóloga ou generalizada*).

Qualquer dos grandes tipos de recombinação cromossômica envolve o seccionamento e a permuta de fragmentos do duplex; não depende da replicação mas necessita de um complexo sistema enzimático (designado por *complexo sinaptonémico*) para ser produzido.

A ultra-estrutura daquele complexo inclui uma região central e dois componentes laterais, cada um dos quais está firmemente unido a um dos cromatídeos do par de cromossomas homólogos. Ao longo do complexo existem, distanciados entre si, os *módulos de recombinação*, ao nível dos quais é processado o cruzamento e corte dos cromatídeos internos, com subsequente permuta de segmentos. Entre as proteínas que actuam no complexo sinaptonémico e modulam a recombinação destacam-se enzimas que intervêm na replicação e reparação do DNA. Este fenómeno parece ser semelhante nos organismos procaríotas e nos eucariotas superiores.

Na sequência da observação de que genes em cópias diferentes de cromossomas homólogos tendem a ser segregados durante a meiose, foram propostos dois modelos para explicar o processo de recombinação (Fig. 3.18.).

A primeira daquelas hipóteses admitia que, após emparelharem entre si na profase da meiose, os cromatídeos partiam-se em determinado local para, de imediato, os fragmentos se unirem entre si, com reconstrução de cromatídeos híbridos, constituídos por partes de origem materna e paterna. Entre-

**Modelo I (Fragmentação e ressociação)****Modelo II (Cópia selectiva)**

**Fig.3.18.** Dois possíveis modelos de “crossing over” para a recombinação cromossómica. No modelo I há troca de segmentos pré-formados entre os cromatídeos de ambos os progenitores. No modelo II, o DNA de cada cromatídeo é replicado inicialmente a partir de cópias de um dos progenitores e termina com a replicação do molde do outro progenitor.

tanto, como as recombinações poderiam ocorrer não só entre cromossomas homólogos mas também entre segmentos do mesmo gene, foi proposto um segundo mecanismo, em que as moléculas de DNA recombinado seriam geradas pela replicação de uma das cadeias parentais que, a determinado momento e sem interrupção, mudava de molde, transitando para uma das cadeias de DNA do segundo progenitor (cópia selectiva). A recombinação decorreria juntamente com a duplicação dos cromossomas, a partir das cadeias de DNA de ambos os progenitores.

A principal diferença entre ambos os mecanismos assentava na origem dos componentes dos cromossomas reconstruídos: enquanto na primeira hipótese (fragmentação e reassociação) havia permuta do material genéti-

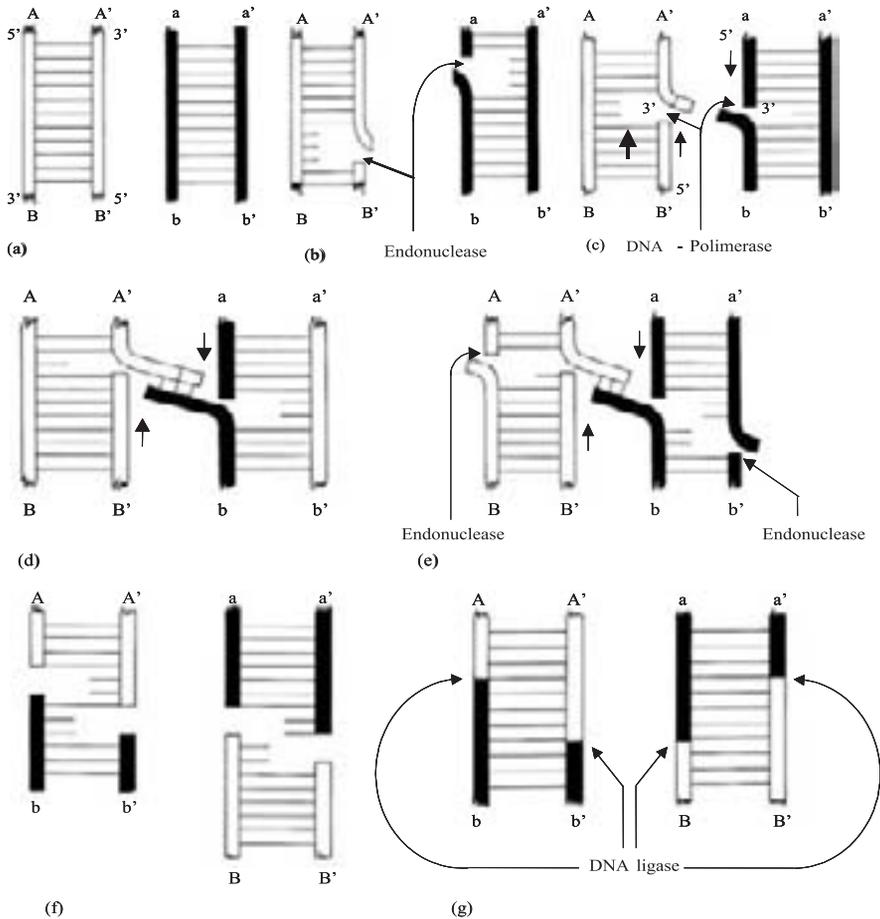
co pré-existente nos cromossomas originais, na segunda hipótese (*cópia selectiva*) os cromatídeos recombinados seriam híbridos sintetizados a partir da transcrição de cadeias complementares, parte materna e parte paterna.

Mais uma vez, foram os fagos  $T_4$  e  $\lambda$  em células de *E. coli* que forneceram a demonstração concludente de que na recombinação cromossômica intervêm nucleases, cada uma codificada por genes específicos do cromossoma viral. Na infecção pelo  $T_4$  foi ainda possível identificar a presença de uma enzima que se assemelhava à DNA polimerase I. Adicionalmente, foi comprovado que a recombinação ocorreria entre moléculas não replicantes, isto é, a partir de segmentos pré-formados, o que veio em apoio da primeira daquelas hipóteses referidas (fragmentação e reassociação) como mecanismo principal do “crossing-over”

De acordo com aquele modelo, o “crossing-over” inicia-se pelo emparelhamento de dois duplex diferentes e lineares. Seguidamente, uma endonuclease catalisa a hidrólise de uma ligação fosfodiéster em uma das cadeias de cada duplex, após o que a DNA polimerase adiciona sucessivos resíduos de mononucleótido às extremidades 3' livres das cadeias seccionadas, deslocando os segmentos pré-existentes da posição que ocupavam no respectivo duplex. Formam-se assim cadeias simples que, no momento em que ficam emparelhadas com o da outra faixa seccionada, estabelecem ligações hidrogeniônicas entre pares de bases complementares, dando origem a uma pequena cadeia dupla entre moléculas diferentes de DNA (Fig. 3.19.).

Assim que a endonuclease seccionar as cadeias ainda intactas, os fragmentos resultantes recombinam-se entre si pela intervenção de uma exonuclease sobre estes últimos, de modo a poderem emparelhar-se e a formar ligações hidrogeniônicas entre bases complementares. As cadeias são completadas por adição de alguns mononucleótidos pela DNA polimerase, sendo finalmente “selados” pela DNA ligase, dando origem a dois duplex *recombinantes* com estrutura genética recíproca.

No entanto, para que a recombinação cromossômica se dê de modo eficiente e sem erros, é necessário que haja correspondência exacta entre as bases complementares, durante a formação das pontes hidrogeniônicas. A possibilidade de ser produzida uma mutação (por deleção ou inserção) aumenta à medida que se repetem bases iguais lado a lado.



**Fig. 3.19.** Hipótese de mecanismo para a recombinação cromossômica: (a) duas moléculas de DNA emparelhadas (modelo Holliday, que consagra o seu proponente, Robin Holliday); (b) corte de uma ligação fosfodiéster em dois cromátídeos adjacentes, um de cada duplex e na mesma posição relativa, por uma endonuclease, com afastamento progressivo da extremidade 5'-P de cada cromátídeo seccionado para o espaço intercalar; (c) uma molécula de DNA, polimerase (provavelmente o tipo I) sintetiza cada um dos segmentos seccionados (b e B'), a partir da extremidade livre 3'; (d) continua a sintetizar uma cadeia complementar do fragmento seccionado (A' e b); (e) uma endonuclease secciona ligações fosfodiéster nos cromátídeos intactos (AB e a'b') de ambos os duplex, o que, em conjunto com as transformações anteriores, origina quatro fragmentos de duplex incompletos e recombinantes; (f) da colisão ao acaso entre as extremidades das seqüências emparelhadas resulta a formação de dois duplex lineares, completados pela DNA polimerase nos sectores em que existe somente uma das cadeias e/ou faltam seqüências de mononucleótidos excedentários, removidos por uma exonuclease, deste modo emparelhando regiões homólogas; (g) ao ser finalizado o emparelhamento das bases complementar entre cada conjunto de novas cadeias reconstituídas, a DNA ligase une os segmentos respectivos, originando duas moléculas de DNA recombinadas.

## Conclusões

A quantidade de DNA celular aumenta a par com a sua diferenciação e posição na escala evolutiva das espécies. Os eucariotas unicelulares (alguns dos quais são preferencialmente utilizados como meio de estudo em biologia bacteriana) têm três a cinco vezes mais DNA por célula do que os procariotas. Os genomas bacterianos possuem muito poucas sequências repetitivas ou sem função aparente: virtualmente todo o DNA é constituído por genes operacionais. Pelo contrário, nos organismos superiores, somente uma muito pequena fracção de DNA parece ter funções úteis, sendo a grande maioria representada por sequências repetitivas

O DNA cromossómico contém a informação genética que determina o fenotipo do organismo. Aquela informação depende da quantidade de DNA celular e da sequência de bases dos nucleótidos que o constituem, resultante da permutação desses monómeros ao longo das cadeias de DNA, específicas para cada tipo celular, organismos e espécies. A disposição relativa dos genes e dos aminoácidos nas cadeia polipeptídicas que codificam é co-linear.

A replicação de DNA é semi-conservadora. Presume-se que o mecanismo de replicação seja idêntico nos procariotas e eucariotas. Em qualquer dos casos a replicação é catalisada pela DNA polimerase, na presença de uma cadeia de DNA a ser replicada e dos quatro principais tipos de desoxirribonucleótido, sob a forma trifosfato. A replicação decorre sempre na direcção 5'→3', com a formação de uma ligação 3',5'-fosfodiéster por cada novo mononucleósido adicionado, a par com a saída de uma molécula de pirofosfato. A replicação é iniciada num duplex intacto pelo seccionamento de uma das ligações fosfodiéster em uma das cadeias (o que sucede por acção da enzima endonuclease, ao reconhecer o sítio de iniciação da DNA polimerase) ou por outro estímulo (eventualmente uma proteína) que desnatura o duplex no local de iniciação da replicação (sequência específica de nucleótidos). A partir desse local de iniciação, a polimerização (pelo menos na *E. coli*) pode decorrer por segmentos curtos. Na hipótese de não haver interrupção da cadeia a ser replicada, no local de iniciação da replicação é sintetizada uma molécula pequena de RNA iniciador, cujo alongamento é continuado pela DNA polimerase em curtos fragmentos de RNA – DNA complementares (segmentos de Okasaki), que se mantêm dissociados entre si até serem por fim unidos entre si por uma DNA ligase. As duas cadeias de DNA

são replicadas quase simultaneamente sempre na direcção 5'→3' mas em sentido antiparalelos.

Todas as células somáticas (diploides) de cada espécie ou organismo superior contêm a mesma quantidade de DNA, independentemente do meio ou da dieta.

O DNA nativo é constituído por uma duplex (duas cadeias complementares enroladas entre si e num eixo comum). A complementaridade resulta do emparelhamento das bases púricas e pirimídicas pertencentes a cada cadeia, A-T e G-C, localizadas no interior da dupla cadeia e unidas entre si por pontes hidrogeniónicas. O total de bases de adenina iguala o de timina, e o total de guanina é também igual ao de citosina. Na generalidade dos vírus e bactérias o DNA forma um círculo fechado, pelo que o respectivo mapa genético é também circular.

A localização dos genes (e das mutações) nos cromossomas é determinado por segregação genética e recombinação de alelos. Experiências de recombinação (em vírus e bactérias) entre cromossomas homólogos possibilitam a determinação posicional relativa dos genes nesses cromossomas (mapa cromossómico ou genético). Cada uma das cadeias de DNA pode replicar-se noutra cadeia complementar.

As mutações podem ser pontuais (por troca, inserção ou deleção de uma ou mais bases nucleotídicas na sequência de DNA) ou afectar um segmento mais ou menos extenso do DNA (aberração cromossómica, em número e forma).

As mutações, tal como a localização dos genes, podem ser determinadas em sequência linear. Cada gene (ou cistrão) codifica a síntese de uma cadeia polipeptídica. As mutações podem originar polipéptidos diferentes do normal (por substituição de uma ou mais aminoácidos) com funções semelhantes e distintas ou a sua síntese é inviabilizada. Em geral as mutações genéticas são recessivas relativamente ao tipo "selvagem".

## CAPÍTULO 4

### ÁCIDO RIBONUCLEICO: ESTRUTURA, SÍNTESE E TRANSCRIÇÃO REVERSA

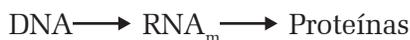
#### Estrutura

A síntese de polipéptidos por adição sucessiva de aminoácidos requer que a informação contida no DNA seja veiculada ao local de síntese. Essa tarefa é executada por um tipo especial de moléculas de ácido ribonucleico, o *RNA mensageiro* (RNA<sub>m</sub>).

Nas células nucleadas quase toda a síntese proteica ocorre no citoplasma, onde praticamente não existe DNA. Fazendo uso dos aminoácidos marcados com isótopos radioactivos verificou-se que quanto mais activa é a síntese proteica num determinado tecido mais ricas são as suas células em RNA. A incorporação de aminoácidos nas proteínas em formação decorre sempre no mesmo subsector celular, os ribossomas.

O ácido ribonucleico mensageiro seria o molde da síntese proteica, ao transportar, dos cromossomas até aos ribossomas, a informação genética existente no DNA. Uma vez cumprida a sua missão, o RNA mensageiro seria degradado nos seus mononucleótidos constituintes.

No conjunto, a síntese proteica decorreria com a seguinte sequência:



O RNA<sub>m</sub>, além de constituir uma pequena fracção da totalidade do RNA celular, teria duração muito fugaz (calcula-se que nas bactérias do RNA tenha meia-vida de cerca de dois minutos).

A estrutura do RNA<sub>m</sub> é, quimicamente, muito semelhante à do DNA, sendo constituídos por quatro tipos de nucleótido que têm como base a adenina, guanina, citosina ou o uracilo. Esta simplicidade não obsta a que cada um dos milhares de proteínas diferentes existentes nos organismos seja codificada por um RNA<sub>m</sub> específico.

Entretanto, a partir de reticulocitos e bactérias infectadas por fagos (aqueles mobilizando o seu mecanismo para formar hemoglobina e estas obedecendo à informação transmitida pelo DNA do vírus) foram isoladas moléculas de RNA<sub>m</sub> específico e identificada a sequência das suas bases, complementares com uma das cadeias de DNA da mesma célula.

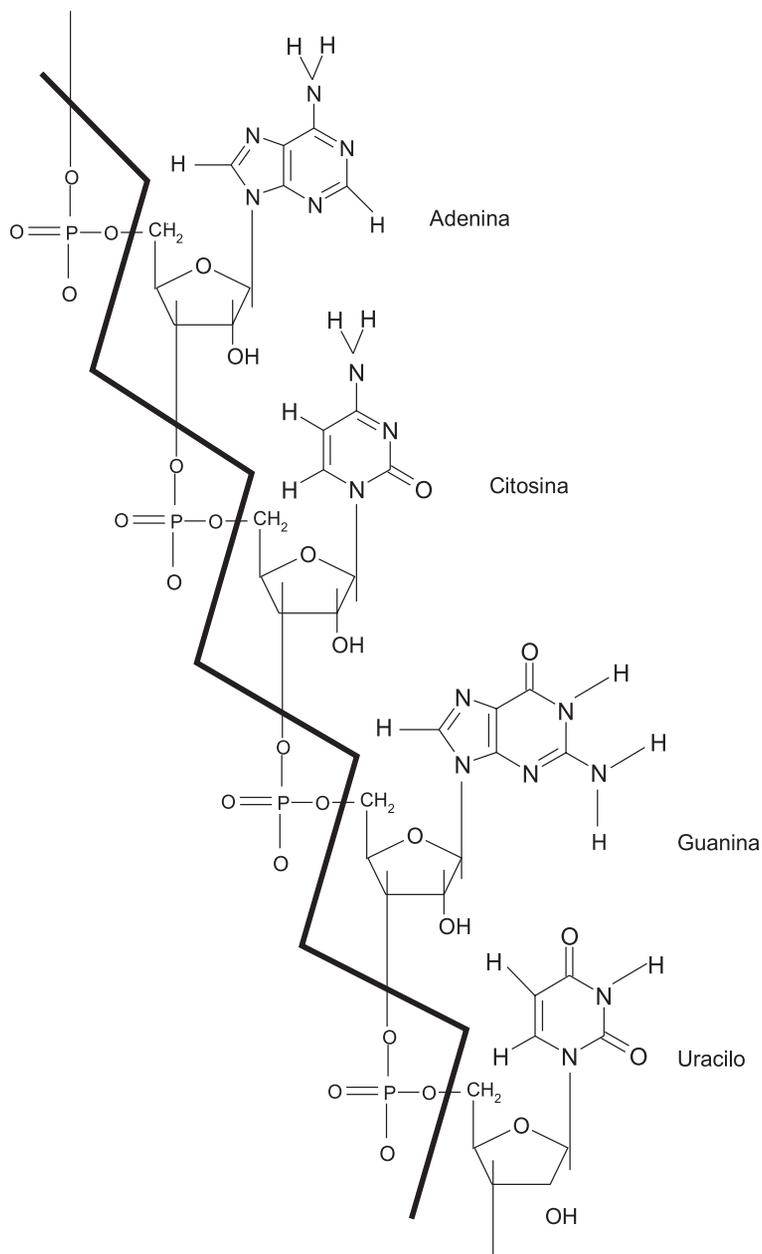
Além de conterem uracilo em vez de timina, também a ribose substitui em todos os tipos de RNA a desoxirribose própria das moléculas de DNA (Fig. 4.1.).

Os ribonucleótidos unem-se entre si por ligações 3', 5'-fosfodiéster, formando longas cadeias lineares. Apenas se conhece um exemplo de RNA<sub>m</sub> circular, identificado no vírus da encefalomiocardite dos murganhos, cuja extensão varia entre 75 a mais de 10.000 mononucleótidos.

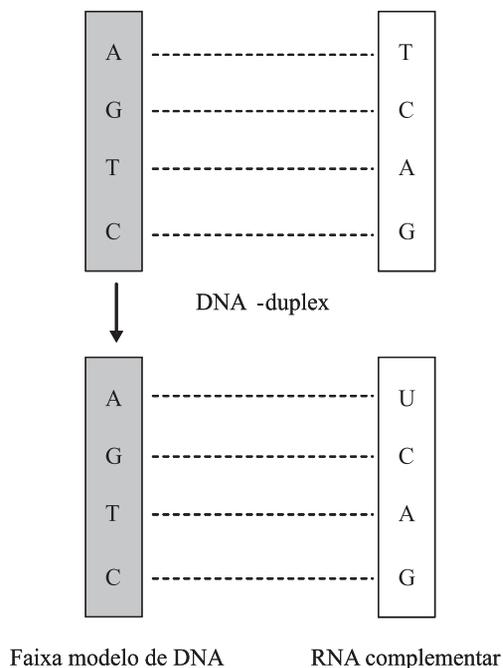
O número de bases adenina numa molécula de RNA<sub>m</sub> é diferente do total de bases uracilo, enquanto a quantidade de nucleótidos com guanina também difere das que têm citosina. Todavia, devido ao enrolamento da cadeia de RNA sobre si própria, formam-se pequenos sacos onde existe um número aproximadamente igual de resíduos de adenina e uracilo ou de guanina com citosina, que possibilitam o emparelhamento transitório entre si (AU, CG), com a formação de pontes hidrogeniônicas semelhantes às do DNA.

## **Transcrição do DNA**

Sendo o DNA e o RNA estruturalmente semelhantes é admissível a existência de um mecanismo que, em determinadas fases de divisão celular, permita a uma das cadeias da molécula de DNA (*cadeia-molde ou modelo*) transcrever a informação para o RNA, em vez de se replicar. A transcrição seria mediada pela complementariedade entre as bases das faixas polinucleotídicas de um e outro, de modo a que à adenina correspondesse a timina (ou o uracilo no RNA) e a guanina emparelhasse com a citosina (no DNA e no RNA) (Fig. 4.2.).



**Fig. 4.1.** Estrutura linear de um polirribonucleótido. Os nucleótidos ligam-se entre si por ligações 3',5'-fosfodiéster, entre o grupo hidroxila do carbono 3' de um mononucleótido e o grupo hidroxila do carbono 5' do mononucleótido seguinte. A linha a cheio indica esquematicamente a intercepção das ligações fosfodiéster da cadeia.

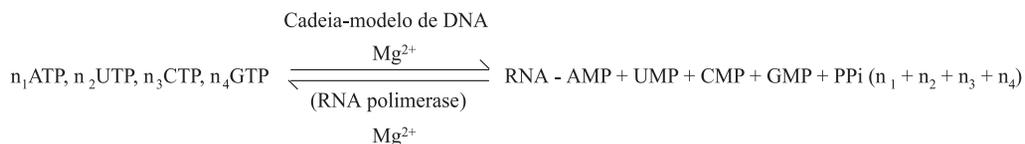


**Fig. 4.2.** Representação esquemática da síntese de RNA, em que é utilizada uma das cadeias de DNA como modelo complementar

Seria de supor que a replicação fosse catalisada por enzima(s) semelhante(s) à que catalisa o alongamento dos desoxirribonucleótidos (a DNA polimerase)

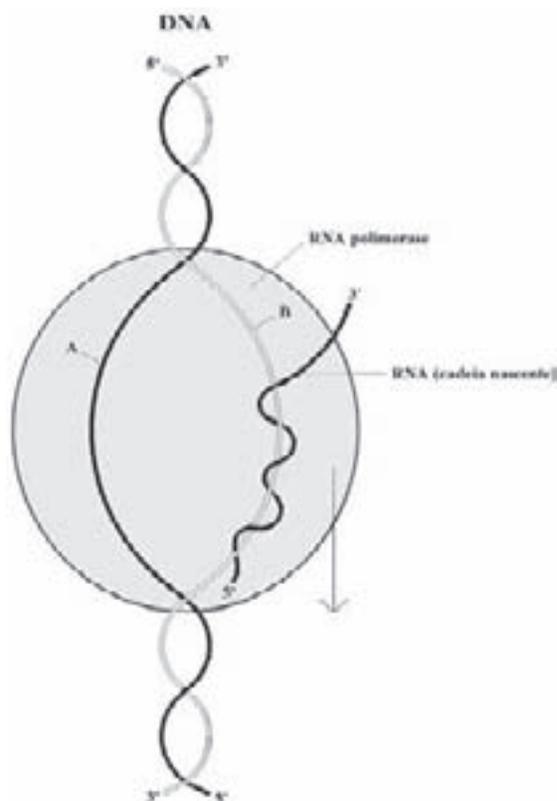
Em finais da década de 60 foi descoberta a *RNA polimerase* que, para actuar (tal como o DNA), requer  $Mg^{2+}$ , um molde (molécula de DNA) e todos os quatro tipos de ribonucleósidos 5'-trifosfato descritos. A reacção, muito semelhante à de replicação do DNA, baseia-se na adição de sucessivos resíduos de ribonucleótidos à extremidade 3' hidroxilo do ribonucleótido (ou do segmento de RNA) pré-existente. A enzima catalisa a formação, através de uma molécula de fosfato, de uma ligação 3', 5'-fosfodiéster entre aquele grupo 3'-OH e o 5'-OH do novo nucleótido a incorporar, pelo que a polimerização sucede na direcção 5'→3'. A reacção é completada pela remoção de uma molécula de pirofosfato (PPi) em cada ligação formada, o que confere elevadas características exergónicas ao RNA em formação. A hidrólise do PPi pela pirofosfatase assegura a continuidade da sua síntese na direcção estabelecida.

A RNA polimerase catalisaria a adição de sucessivas moléculas de ribonucleótidos ao grupo 3'-OH livre da cadeia de RNA que, como o DNA, cresce na direcção 5' → 3':



A RNA polimerase tem sido, desde então, isolada a todos os níveis celulares (procariotas ou eucariotas), localizando-se nos organismos superiores sobretudo no núcleo, ainda que também exista nas mitocôndrias.

Uma vez formado, o RNA destaca-se rapidamente da faixa da molécula de DNA que lhe serviu de modelo, permitindo que o duplex se reconstitua (Fig. 4.3.).



**Fig. 4.3.** Representação sequenciada da transcrição de uma molécula de RNA a partir de uma das cadeias (B) do duplex de DNA. A ligação da RNA polimerase ao local de iniciação da cadeia-modelo de DNA, induz o desenrolamento local do duplex (numa extensão de 10 a 20 nucleótidos, enquanto é formado um híbrido DNA-RNA entre a cadeia de DNA transcrito (B) e a molécula nascente de RNA. À medida que a RNA polimerase avança (na direcção 5'→3', indicado pela seta), progride também o sector desenrolado, enquanto a cadeia transcrita se reúne ao seu par complementar, retomando a forma em duplex, com restabelecimento das pontes hidrogeniónicas. Ao ser activado o sítio de terminação da replicação (na cadeia de DNA transcrito), a polimerase dissocia-se juntamente com a cadeia de RNA e moléculas de pirofosfato (PPi) que resultaram da formação de ligações fosfodiéster entre os ribonucleósidos trifosfato utilizados.

É muito raro haver erros de transcrição mas, se os houver, o curto período de meia-vida da molécula de RNA<sub>m</sub> não possibilita a manutenção da anomalia decorrente. Além do RNA<sub>m</sub>, são constituintes comuns das células dois outros tipos de ácidos ribonucleicos, o de transferência (RNA<sub>t</sub>) e o ribossômico (RNA<sub>r</sub>).

O RNA<sub>t</sub> e o RNA<sub>r</sub> são sintetizados do mesmo modo que o RNA<sub>m</sub>, por transcrição do DNA. No entanto, ainda que aqueles dois tipos constituam, em conjunto, a maior parte do RNA celular (98% nas bactérias), somente 1% do total de DNA lhes serve de modelo. Um gene parece codificar 30 a 40 diferentes espécies de RNA<sub>t</sub> enquanto as moléculas de RNA<sub>r</sub> parecem ser codificadas por vários genes.

A partir de diferentes moléculas de DNA com determinada proporção de bases (A + T)/(G + C) obtêm-se moléculas de RNA em que a relação (A + U)/(G + C) é aproximadamente comparável às que lhes serviu de modelo.

Experiências com qualquer uma das cadeias de DNA do vírus  $\phi$ X174 comprovaram que a RNA polimerase sintetizava moléculas de RNA com sequência de bases complementar à da cadeia transcrita. Usando ambas as cadeias de DNA obtiveram-se duas moléculas complementares (entre si) de RNA.

Admitindo que cada uma das cadeias de DNA originava uma molécula complementar de RNA, em vez de uma molécula peptídica haveria dois péptidos diferentes, o que não sucede *in vivo*. Nos vírus SP<sub>8</sub> e T<sub>7</sub>, foi demonstrado que a cadeia de DNA copiada era sempre a mesma, quer *in vivo* ou *in vitro*.

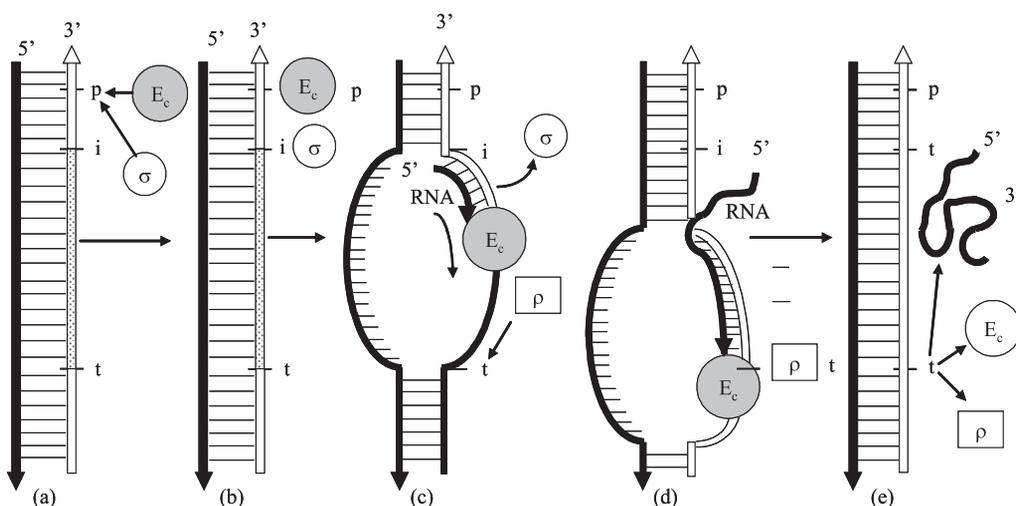
Quando ambas as cadeias do duplex são replicadas, como sucede nos fagos T<sub>4</sub> ou  $\lambda$ , decerto existem diferenças na distribuição de genes em cada uma, que são necessárias à multiplicação viral.

A estrutura da RNA polimerase é bastante complexa: a forma activa contém cinco polipéptidos diferentes ( $2\alpha$ ,  $\beta$ ,  $\beta'$ ,  $\omega$ ,  $\sigma$ ), agregados entre si por ligações secundárias. Uma das cadeias polipeptídicas ( $\alpha$ ) existe em duplicado na molécula enzimática, ao contrário das restantes, monoméricas. A cadeia  $\sigma$  (sigma) não possui actividade catalítica. A sua união às restantes cadeias do modelo é mais fraca, pelo que se destaca facilmente do agregado  $\alpha_2$ ,  $\beta$ ,  $\beta'$  e  $\omega$ , o qual constitui a componente activa da enzima. Este *núcleo enzimático* catalisa a formação das ligações fosfodiéster dos polinucleótidos, com ou sem a cadeia  $\sigma$  presente. Aparentemente, a cadeia  $\sigma$  selecciona a cadeia de DNA que servirá de modelo ao RNA e interpreta o sinal de início de transcrição. Uma vez iniciada a polimerização do RNA, a cadeia  $\sigma$  é

libertada para se unir a outro núcleo enzimático. Em mutantes com deficiência da cadeia  $\sigma$ , o núcleo enzimático pode formar RNA na presença de sinais iniciadores falsos mas, em lugar de ser seleccionada uma das cadeias de DNA, a transcrição decorre em ambas.

Todas as cadeias normais de RNA começam por mononucleótidos com adenina ou guanina. Por esse motivo, o “sinal de início” deverá seleccionar uma base pirimídica (T ou C) do DNA para que o RNA comece a ser transcrito.

Ao atingir o comprimento ideal, são activados sinais de “fim de transmissão” que interrompem a polimerização do RNA. Estes sinais são traduzidos por proteína(s) específica(s) – o factor  $\rho$  (rho) – com peso molecular de aproximadamente 60.000 dalton, cujo o mecanismo de acção exacto continua por esclarecer (Fig. 4.4.).



**Fig. 4.4.** Representação esquemática da transcrição pelo RNA polimerase de um gene delimitado pelo local de iniciação (i) e local de terminação (t). (a) união inespecífica da RNA polimerase ( $E_c$ , fracção catalítica; factor sigma,  $\sigma$ ) a uma das cadeias do duplex de DNA no local definido pelo promotor (p); (b) Reconhecimento (pelo factor  $\sigma$ ) da ligação da RNA polimerase ao sítio de iniciação da transcrição; (c) a polimerase induz o desenrolamento progressivo da dupla hélice (desde um pouco antes e até um pouco depois do sector catalisado) com a progressão da transcrição, de uma das cadeias-molde catalisada pela fracção catalítica (E), com libertação do factor  $\sigma$  desde a sua iniciação, e reconstituição da dupla hélice após a transcrição; (d) ao ser atingido o local de terminação (t), sinalizado pelo factor  $\rho$ , cessa a transcrição, estando concluída uma nova cadeia de RNA; (e) segue-se a dissociação imediata do RNA da cadeia molde, enzima e factor  $\rho$ .

Há razões para crer que nas bactérias (p.ex., na *E. coli*) um único tipo de RNA polimerase forma todos os tipos de ácido ribonucleico ( $\text{RNA}_m$ ,  $\text{RNA}_t$  e  $\text{RNA}_r$ ). Eventualmente o mesmo sucederá nos eucariotas, requerendo que as codificações de muitos tipos de RNA diferentes seja regulada de modo a que a respectiva transcrição se processe em períodos pré-definidos de desenvolvimento e actividade dos organismos, ou de acordo com as solicitações ou limitações do meio.

Quase todo o RNA bacteriano é citoplásmico. Contudo, nos mamíferos (p.ex. nos hepatocitos) a proporção é muito diferente: 11% do RNA total localiza-se no núcleo (sendo quase todo do tipo  $\text{RNA}_m$ ), 15% nas mitocôndrias ( $\text{RNA}_r$  e  $\text{RNA}_t$ ), mais de 50% nos ribossomas (praticamente só  $\text{RNA}_r$ ) e 24% no citosol (sobretudo  $\text{RNA}_t$ ).

Cada molécula de RNA polimerase transcreve a informação centrada em um ou mais genes estruturais (ou cistrões), a que corresponde a codificação de uma cadeia de  $\text{RNA}_m$  no segmento numa cadeia polipéptidica. As moléculas de  $\text{RNA}_m$  nas bactérias são facilmente policistrónicas, enquanto as dos eucariotas são monocistrónicas, isto é, codificam um gene.

## Regulação da transcrição do DNA

Ainda que o mecanismo da síntese proteica pareça ser, em termos gerais, semelhante em todas as formas de vida, é evidente que a composição do DNA bem como a regulação da sua expressão na síntese proteica sejam substancialmente diferentes entre procariotas e eucariotas. Nos mamíferos, em particular, o DNA é heterogéneo e contém sequências muito repetitivas entre outras que o não são, está envolvido por proteínas e localiza-se num sector específico da célula – o núcleo – com particularidades distintas do citoplasma e restantes organitos.

A transcrição é regulada por proteínas reguladoras, por sua vez codificadas por genes reguladores.

Nos procariotas a transcrição é regulada por controlo positivo (factor da RNA polimerase) e negativo (proteínas repressão). Foram identificadas proteínas reguladoras que induzem (o início) da transcrição (*factores ou activadores de transcrição*) e outros que a reprimem (*factores repressores da transcrição*). Outras proteínas sinalizam a terminação da transcrição. Enquanto os activado-

res sinalizam a sequência de nucleótidos onde deve ser iniciada a transcrição (designado *promotor*) estimulando a união da RNA polimerase no local, os repressores impedem essa ligação. Em termos práticos, o factor s funciona como um activador da transcrição. Por sua vez, o factor r funciona como um repressor, pois que, ao ligar-se à RNA polimerase, suspende a transcrição.

O promotor é constituído por 15 a 20 pares de mononucleótidos que, além de assegurarem eficazmente a ligação de RNA polimerase, determinam se a transcrição vai ou não acontecer, bem como a frequência com que ocorre.

O mecanismo regulador da síntese do DNA nos eucariotas será semelhante mas, considerando a sua maior complexidade, disporá de outros tipos de controlo. Neste âmbito cabe destacar as principais proteínas da cromatina, exclusivas dos eucariotas (as histonas), que se dispõem espaçadamente ao longo de cerca de 10% da extensão do DNA, unindo sectores de carga negativa desta molécula. Adicionalmente, está identificada uma larga variedade de proteínas não-histónicas (com massa equivalente à das histonas) que parecem participar nas diversas actividades, designadamente na expressão genética e na replicação do DNA. Aparentemente, a união das proteínas ao DNA tem por função primordial proteger esta molécula da acção das nucleases, impedindo ainda a transcrição em determinados períodos de inactividade.

Na generalidade, os locais de controlo estão restritos a curtas sequências com poucas dezenas de nucleótidos onde se ligam as moléculas de RNA polimerase e as proteínas reguladoras (ver Capítulo 7, para maior desenvolvimento sobre os mecanismos de regulação da síntese proteica).

## **Transcrição reversa**

A transcrição nem sempre se passa como foi descrito atrás, ou seja, por passagem da informação genética do DNA ao RNA e, deste, para a síntese proteica. De facto, são conhecidos vírus em que todos os ácidos nucleicos são polímeros de ribonucleótidos. Os primeiros genomas RNA foram identificados em vírus de plantas, na generalidade constituídos por RNA e proteínas. Os bacteriófagos  $F_2$  e  $R_{17}$  são outros exemplos de vírus com RNA. Desde então, os vírus foram classificados em dois grupos, consoante possuem DNA ou RNA.

Qualquer dos tipos de ácidos nucleicos (DNA ou RNA), isolados e purificados dos vírus que os contêm como material genético, têm a capacidade de

infectar células e induzir nestas a produção de novos vírus, idênticos aos infectantes que forneceram o molde de ácido nucleico.

Se bem que na maioria dos vírus com DNA a estrutura desta molécula seja em duplex, há alguns tipos constituídos por uma faixa simples. Por sua vez, o RNA em cadeia única é típico na maior parte dos vírus que o contêm (p.ex., vírus da influenza, da poliomielite,  $F_2$ ,  $R_{17}$ ). Nos reovírus, contudo, o RNA apresenta-se em duplex, sendo ambas as cadeias complementares entre si. Este pormenor tem pouca importância, pois que ao penetrar na célula a única cadeia a entrar é replicada de imediato. O mecanismo seguido depende do tipo de ácido nucleico do vírus infectante.

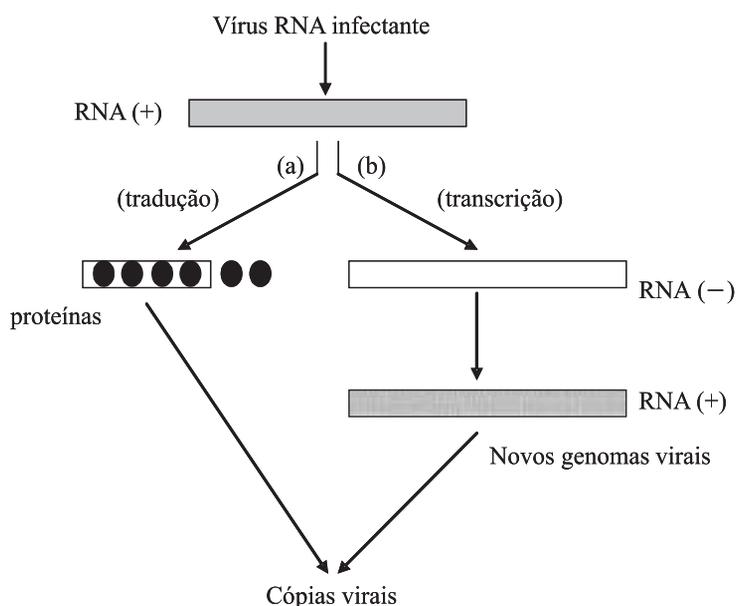
Se o material genético for o DNA, as suas (ou única) cadeia(s) serve(m) de molde quer para a replicação quer a transcrição de síntese das proteínas específicas do vírus no organismo hospedeiro.

No caso de vírus RNA, este funciona como molde para a formação de mais moléculas de RNA e ainda codifica a síntese das proteínas que lhe são específicas. Em qualquer dos casos o vírus infectante (DNA ou RNA) obtém grande número de cópias das proteínas e ácido nucleico específicos, os quais lhe possibilitam a sua multiplicação no organismo hospedeiro em partículas virais, amadurecidas, com capacidade infectante idêntica à da forma progenitora. A demonstração de que o RNA actua como material genético dos vírus das plantas foi revelada nas células infectadas com o vírus do mosaico do tabaco, o qual originava múltiplas cópias com estrutura e capacidade infectante idênticas

Estudos subsequentes em *E. coli* infectados por fagos RNA elucidaram o processo da síntese do RNA dependente do RNA, baseado no emparelhamento de bases complementares; o processo, muito semelhante ao da replicação e transcrição de RNA, é catalisado por uma enzima específica com actividade semelhante às polimerases de DNA e RNA.

Os fagos RNA mais simples que estão caracterizados (p.ex.,  $R_{17}$ ,  $F_2$ , Q $\beta$ ) contêm uma cadeia de RNA com cerca de 3.000 nucleótidos; esta sequência codifica os 1.100 aminoácidos que constituem, pelo menos, três tipos de proteínas diferentes. Duas das cadeias polipeptídicas que formam a camada proteica protectora (cápside) mobilizam, na sua síntese, aproximadamente metade do material genético. A outra proteína, com cerca de 600 aminoácidos e actividade catalítica, é a enzima *RNA replicase* (ou sintetase), que induz a auto-replicação da cadeia de RNA no hospedeiro, imediatamente após este ser invadido pela RNA viral; se aquela cadeia RNA actuar como RNA<sub>m</sub>, é designada por RNA (+), pelo que a RNA replicase

catalisará a formação de uma cadeia RNA complementar (-) a à do modelo invasor. Este mecanismo requer um perfeito emparelhamento das bases de ambas para que a réplica seja exacta. Após serem formados, os duplex RNA servem de modelo a cadeias (+) complementares, umas aproveitadas como material genético dos novos vírus e outras como molde para a síntese de proteínas específicas (Fig. 4.5.).

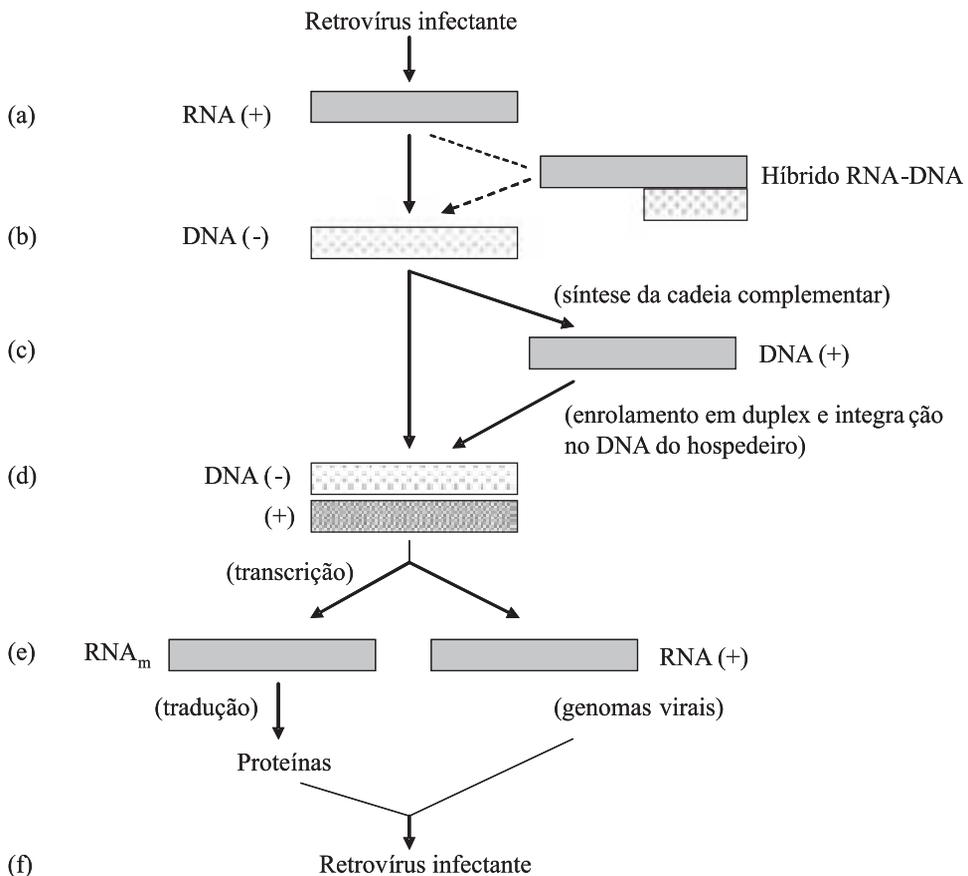


**Fig. 4.5.** Replicação do vírus RNA da cadeia simples. O RNA actua em dois processos simultâneos: (a) ao funcionar como RNAm, origina a síntese de poliproteínas, de que resultam (por acção das proteases intracelulares) diversos tipos de moléculas proteicas específicas dos vírus; (b) por acção da RNA replicase (ou RNA polimerase dependente do RNA) são formadas cadeias RNA (-), complementares das moléculas de RNA (+) infectante e, novamente por transcrição a partir de RNA (-), são formados novos genomas virais (RNA+) prontos a reconstituírem, com as proteínas, novas cópias virais.

Mais recentemente, foi verificado por Temin e Cols que o RNA do vírus do sarcoma de Rous tinha a propriedade de sintetisar uma cópia de DNA nas células infectadas. Este facto, que estava em desacordo com o mecanismo de transcrição conhecido até então, seria atribuível a um novo tipo de enzima activo em determinados vírus animais RNA com capacidade cancerígena.

A *transcrição reversa* (como foi designada) seria catalisada por uma DNA polimerase RNA-dependente (ou *transcritase reversa*) presente no vírus infec-

tante. O material genético do vírus, incubado em meio celular enriquecido com desoxirribonucleótidos, originava moléculas de DNA (equivalente a provírus). Estas experiências demonstraram que o RNA poderia servir de modelo ao DNA por acção de uma polimerase específica. Incubações prévias com ribonuclease impediam aquela reacção (Fig.4.6.).



**Fig. 4.6.** Replicação de retrovírus. (a) O RNA viral, depois de penetrar na célula hospedeira (b) serve de molde à síntese de uma cadeia complementar de DNA (com formação de híbridos RNA-DNA transitórios), por acção de uma enzima também viral, a transcriptase reversa, (c) que catalisa também a síntese de outras cadeias complementares de DNA; (d) as duas cadeias dispõem-se em duplex, que é recombinado com o DNA do hospedeiro, (e) a que se segue a sua transcrição em RNA, com características de RNA<sub>m</sub> e também com capacidades provirais; (f) o processo culmina com a regeneração de muitas cópias idênticas às do retrovírus infectante.

Quase em simultâneo, outro grupo de investigadores liderados por Baltimore detectou o mesmo tipo de polimerase em preparações de vírus do sarcoma de Rous e do vírus de leucemia de murganho. Foi então sugerido que a polimerase seria exclusiva dos vírus causadores de tumores e que, portanto, teria qualquer relação directa com o problema do cancro.

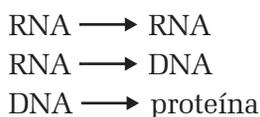
Posteriormente, foi demonstrado que o novo DNA era complementar do RNA do vírus. Durante a replicação dos vírus formam-se híbridos RNA-DNA. Através de uma mutante termolábil do vírus RNA do tumor de Rous verificou-se ser necessária a presença do genoma RNA do vírus para que a célula normal malignizasse.

Ainda que aparentemente revolucionária, a transcrição reversa está implícita no “dogma central” de Watson e Crick. Em 1970, Crick reafirmou aquela ideia ao propor três tipos de transferência da informação genética.

O tipo geral incluía:



enquanto o “tipo especial” poderia abranger circunstâncias muito raras ou ainda desconhecidas:



Um terceiro tipo, presumiria situações “improváveis”, p.ex., uma proteína servir de modelo a ácido nucleicos.

Resultados experimentais mais recentes sugerem que o “tipo especial” estaria largamente distribuído, pelo menos em tecidos diferentes. De facto, utilizando modelos naturais e sintéticos foi detectada actividade típica da polimerase RNA-dependente em células de leucemia humana. Recorrendo somente a modelos sintéticos, Spiegelman detectou aquela actividade enzimática em mais de 100 tipos de leucemias e tumores sólidos humanos, em tecidos embrionários (humanos e de animais de experiência), e em fígados em regeneração.

Quer em cultura de células de murganhos ou em cultura de fibroblastos diploides humanos, foi também observada actividade enzimática própria de transcriptase reversa. No entanto, pela fraca especificidade dos modelos sintéticos utilizados, aqueles resultados aguardam confirmação.

## Conclusões

A transcrição do DNA em RNA é catalisada por uma RNA polimerase, tendo como molde uma das cadeias de DNA e na presença dos principais quatro ribonucleósidos-5'-trifosfato, sendo o processo muito semelhante ao da replicação. A RNA polimerase coloca e une entre si cada um dos precursores monoméricos das cadeias de RNA, assegurando a complementaridade das bases com a cadeia modelo do DNA, a ser transcrita na direcção 5'→3'. A RNA polimerase é constituída por cinco tipos de cadeias polipeptídicas, das quais uma (que não tem actividade catalítica, ( $\sigma$ )) reconhece a sequência de mononucleótidos de DNA onde começa a transcrição, enquanto o núcleo enzimático (com as restantes quatro cadeias) catalisa ligações fosfodiéster. A transcrição termina ao ser atingida pela enzima uma sequência de nucleótidos de DNA que sinaliza a terminação, que é reconhecida por outra proteína, o factor  $\rho$ .

Em células infectadas por vírus RNA a transcrição de RNA decorre a partir da molécula-modelo de RNA, por acção de uma polimerase RNA-dependente. Em vírus oncogénicos, o RNA viral pode originar por acção de uma polimerase especial, também RNA-dependente (a transcriptase reversa), a formação de uma cadeia de DNA (complementar), a qual promove a síntese de outra cadeia complementar, regenerando um duplex DNA recombinante; deste duplex é sintetizado RNA com carácter viral e também com actividade de RNA mensageiro. Em ambos os casos, a RNA viral utiliza a estrutura e componentes de células hospedeiras para se multiplicar em cópias idênticas, a partir de novos genomas virais (RNA) e proteínas da cápside

# CAPÍTULO 5

## TRADUÇÃO DA INFORMAÇÃO GENÉTICA

### Significado, localização e componentes

A *tradução* consiste na transferência da informação contida no RNA<sub>m</sub> para o sistema formador de proteínas, activo nos ribossomas celulares. Pode dizer-se que os mais significativos progressos neste campo surgiram com a utilização de isótopos radioactivos em estudos com animais e células intactas e, posteriormente, em homogeneizados celulares, em geral obtidos de fígado de rato. A bactéria *E. coli*, pelo seu fácil manejo e grande rapidez de multiplicação, teve também grande utilidade no esclarecimento das complexas etapas da síntese proteica.

Nos estudos inicialmente realizados em animais e células intactas foram obtidas as seguintes conclusões determinantes:

- a) A síntese proteica requer energia;
- b) Os aminoácidos são os precursores iniciais das proteínas;
- c) As proteínas teciduais estão em permanente renovação;
- d) O fígado dos vertebrados é muito activo na síntese proteica.

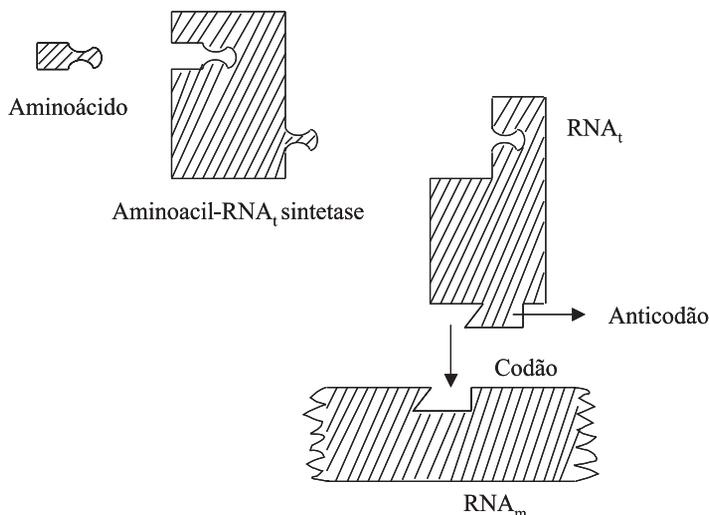
Seguidamente, foi demonstrado que a síntese proteica ocorria em partículas nucleoproteicas intracelulares, (que receberam a designação de *ribossomas*), após o que as proteínas sintetizadas naqueles componentes eram transferidas para as restantes estruturas celulares. Numa fase posterior, foi verificado que a síntese proteica ocorreria em meio celular desde que este contivesse, além dos aminoácidos essenciais, *microssomas* (fracção que incluía os

ribossomas), substratos solúveis, ATP (proveniente das mitocôndrias, activadas por substratos respiratórios) e magnésio. Na fracção solúvel destacavam-se dois factores essenciais, um termolábil (com actividade catalítica própria das sintetases) e outro termoestável (identificado como ácido ribonucleico), que recebeu a designação inicial de *RNA solúvel*, depois modificada para a de *RNA de transferência* ( $\text{RNA}_t$ ).

No prosseguimento foi verificado que a *sintetase* tinha por função unir (com ligação éster) cada molécula de aminoácido a uma molécula de  $\text{RNA}_t$ . Na realidade, foi elucidado que a cada tipo de aminoácido corresponde uma sintetase específica. Verificou-se também que existiam numerosos tipos diferentes de  $\text{RNA}_t$ , cuja função era a de transportar aminoácidos para a sua posição correcta na cadeia polipeptídica, cuja sequência era definida por uma molécula específica de RNA mensageiro. As diferentes moléculas de  $\text{RNA}_t$  actuariam como “adaptadores moleculares” (dos correspondentes aminoácidos) na sequência polipeptídica (Fig. 5.1.). Não havendo afinidade química entre a estrutura química dos aminoácidos e a sequência de bases púricas e pirimídicas que codificam a mensagem genética nos  $\text{RNA}_m$ , concluiu-se que seria o  $\text{RNA}_t$  o intermediário adequado. Seria energeticamente mais económico que cada molécula de  $\text{RNA}_t$  transferisse o respectivo aminoácido (inalterado) do que submeter este a modificações químicas. Adicionalmente, havia a vantagem de uma única enzima (a sintetase) ser suficiente para unir o aminoácido à molécula de  $\text{RNA}_t$ , em simultâneo com o aporte endergónico ao conjunto.

O peso molecular de quase todas as sintetases é aproximadamente 100.000 dalton, sendo a sua actividade fortemente dependente dos grupos sulfidrílicos que contêm. Como foi dito, é fundamental para a reacção que catalisam a sua grande especificidade, quer para os aminoácidos quer para a molécula de  $\text{RNA}_t$  correspondente. Esta especificidade implica a existência, na molécula da sintetase, de dois pontos de ligação essenciais, um para o aminoácido e o outro para o  $\text{RNA}_t$ , além de um terceiro para o ATP. Esta especificidade é de tal ordem que a frequência de um aminoácido errado numa dada cadeia polipeptídica em crescimento é muito inferior a 1/1000.

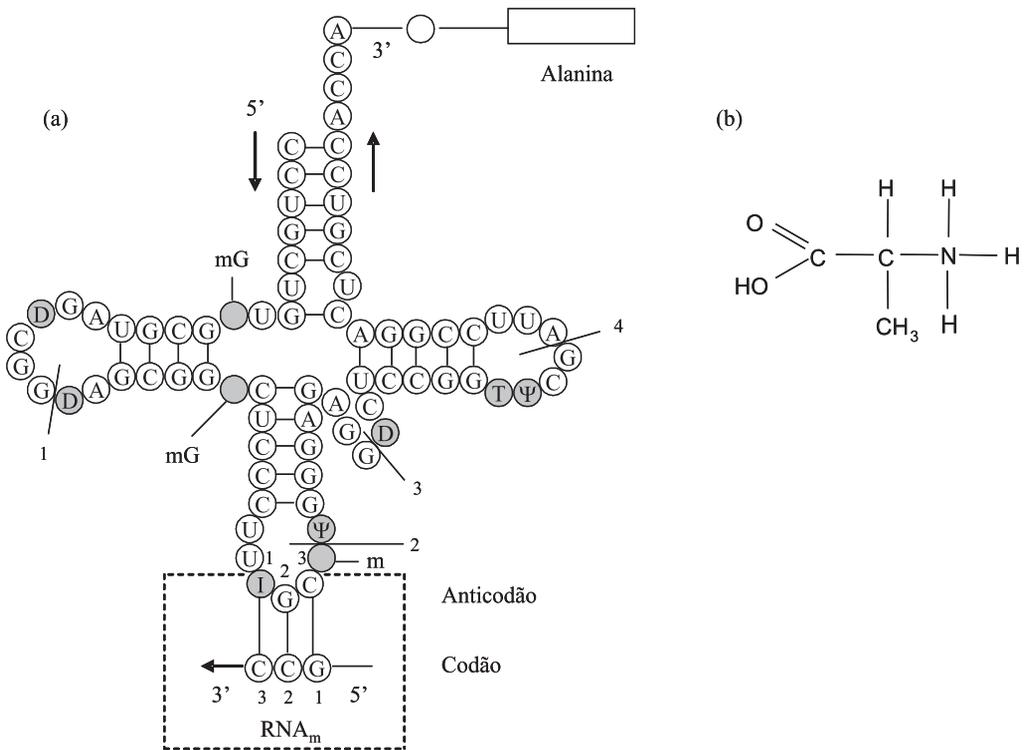
Seguidamente, havia que esclarecer como é que o  $\text{RNA}_t$  reconhecia um determinado aminoácido para o incorporar no ponto exacto da cadeia peptídica.



**Fig. 5.1.** Representação esquemática da interdependência entre cada aminoácido e a molécula transportadora (RNA<sub>t</sub>), cuja união é catalisada por uma enzima (aminoacil sintetase) que lhe é específica. Por sua vez, a selecção de cada aminoácido para a respectiva posição na sequência polipeptídica a ser sintetizada é determinado por um tripleto (codão) definido na cadeia de RNA<sub>m</sub> que está a ser traduzido. Deste modo, o posicionamento de cada aminoácido é assegurado pelo RNA<sub>t</sub> através de um anticodão que serve de “adaptador” ao codão de RNA<sub>m</sub>.

Os RNA<sub>t</sub> são moléculas relativamente pequenas, cada uma com peso molecular entre 23.000 e 300.000, constituídas por 75 a 90 mononucleótidos. Todas as moléculas de RNA<sub>t</sub> na conformação nativa têm estrutura tridimensional, com forma aproximada à de uma “folha-de-trevo”, apesar de serem constituídas por uma única cadeia de RNA, que se dobra sobre si própria em determinados sectores (Fig. 5.2.). O emparelhamento das bases complementares permite a formação de ligações hidrogeniónicas, em que o total de resíduos de adenina equivale ao de uracilo, enquanto a quantidade de resíduos de guanina é muito aproximada à de citosina. O enrolamento resulta do emparelhamento máximo das bases complementares constituintes, exceptuando os sectores que, devido ao predomínio de bases ou nucleótidos raros, dão origem a quatro ou cinco ansas principais.

A disposição e forma de ansas do RNA<sub>t</sub>, reveladas por difracção de raios X em cristais de RNA<sub>t</sub> de muitas espécies, sugeriam que a conformação de RNA<sub>t</sub> mudava quando “recebia” o aminoácido específico.



**Fig. 5.2.** Representação estrutural da estrutura secundária da molécula de RNA<sub>t</sub>-alanina (tRNA<sup>Ala</sup>), isolada pela primeira vez em leveduras.

(a) Modelo em “folha-de-trevo”, característico de todos os tipos de RNA<sub>t</sub>. As zonas de emparelhamento intracadeia são designadas *hastes* e as não emparelhadas *ansas*. As bases raras que existem intercaladas na sequência da cadeia de RNA<sub>t</sub>, originam em média quatro ansas por molécula. Cada letra assinala a base do respectivo nucleótido. A escuro são sinalizadas as bases raras: D= dihidrouridina; I= inosina; T= timina (não é habitual a existência em RNA<sub>t</sub>); y= pseudouridina; m= grupo metilo. Na ansa 2 localiza-se um tripleto, denominado anticodão, que especifica o tipo de aminoácido a que se liga e transfere para a posição correcta na sequência polipeptídica em formação. A informação sobre o tipo de aminoácido e sua localização no polipeptídeo está representada na sequência de tripletos de RNA<sub>m</sub>. Cada tripleto do RNA<sub>m</sub>, denominado codão reconhece o tripleto (anticodão) do RNA<sub>t</sub> como sequência complementar de bases. O emparelhamento codão-anticodão é transitoriamente estabilizado por ligações hidrogeniónicas entre as bases complementares, enquanto decorre o posicionamento e formação de ligações peptídicas no polipeptídico em formação.

(b) Fórmula molecular da alanina. A união com o resíduo de adenosina 5' monofosforil (A) terminal ocorre por uma ligação éster entre o grupo 3'-OH da ribose de (A) e o grupo carboxílico (COOH) da alanina.

A sequência de bases (de nucleótidos) constituintes do RNA<sub>t</sub> varia de molécula para molécula, embora sem alteração na sua conformação geral. Igualmente, todas as moléculas de RNA<sub>t</sub> possuem idêntica sequência e tipo de mononucleótidos (CCA) na extremidade 3' livre. O nucleótido terminal na outra extremidade (5') da cadeia é, quase sempre, o ácido guanílico (G). Além das bases habituais (adenina, guanina, citosina e uracilo), estão identificados na molécula de RNA<sub>t</sub> mais de 10% de *bases raras* (ou “menores”, que são em geral formas metiladas das bases normais ou derivados), além de outros mononucleótidos raros, como o ácido pseudouridílico ou ácido ribotimidílico.

O resíduo adenílico (A) na extremidade 3' constitui o ponto de ligação éster (na posição 2' ou 3') a um aminoácido específico, por acção da respectiva aminoacil-RNA<sub>t</sub> sintetase. Por sua vez, o reconhecimento do aminoácido pelo RNA<sub>t</sub> é determinado por uma sequência de três mononucleótidos (diferentes para cada RNA<sub>t</sub>) existente no segmento oposto aquela extremidade da molécula. Aquele tripleto de reconhecimento contém um *anticodão* complementar a um dos tripletos (*codão*) que constituem a molécula do RNA<sub>m</sub>, o qual determina o local de fixação sequencial de cada aminoácido activado (com o respectivo RNA<sub>t</sub>).

Cada tipo de RNA<sub>t</sub> tem um anticodão próprio, delimitado por bases raras; estas bases parecem isolar o tripleto e impedir a sua hidrólise por nucleases. Considerando que o mesmo aminoácido pode ser unido a diferentes moléculas de RNA<sub>t</sub>, e sendo a sua união catalisada unicamente por uma sintetase específica, poderá concluir-se que a cada aminoácido correspondem anticodões (no RNA<sub>t</sub>) e codões (no RNA<sub>m</sub>) diferentes; a precisão com que o processo decorre deve depender da especificidade das sintases respectivas, quer para os aminoácidos quer para as moléculas de RNA<sub>t</sub>.

Os tripletos referidos nas moléculas de RNA<sub>m</sub> e RNA<sub>t</sub> (codão e anticodão) deverão ter sequências de bases complementares que possibilitem a formação das ligações hidrogeniónicas indispensáveis à inserção correcta de cada aminoácido na sequência peptídica em crescimento. Por conseguinte, cada molécula de RNA<sub>t</sub> traduziria a mensagem de código genético (referente a cada codão) ao seleccionar e posicionar cada aminoácido naquela sequência.

A metilação de algumas das bases do RNA<sub>t</sub> requer S-adenosil metionina (como dador de grupos metilo) e a respectiva transferase. Outras bases também contêm enxofre (p.ex., os ácidos pseudo-uridílico e ribotimidílico). A função des-

tas bases raras parece ser, também, a de evitar o emparelhamento normal e, portanto, a formação de duplex, de modo a manter expostos grupos químicos necessários à formação de ligações secundárias com o RNA<sub>m</sub>, com os ribossomas e com as moléculas de aminoacil-RNA<sub>t</sub> sintetase.

Na extremidade da ansa 1 das moléculas de RNA<sub>t</sub> (Fig. 5.2) existem duas daquelas bases (metilguanina). Admite-se que o RNA<sub>t</sub> se une à superfície ribossômica por esta ansa. Na ansa 2 onde se localiza o anticodão, três dos sete nucleótidos desemparelhados também são do tipo raro: pseudouridina, metil-inosina e inosina. Na ansa 3, constituída por oito a doze nucleótidos, localiza-se a di-hidro-uridina, à qual parece unir-se a aminoacil sintetase. Muito provavelmente as bases raras são adicionadas ou sintetizadas depois da transcrição do (RNA<sub>t</sub>).

Os ribossomas têm conformação semelhante em todas as células, ainda que variável quanto à estrutura e dimensão; estão localizados sobretudo no citoplasma (livres ou unidos ao retículo endoplasmático), além de existirem (em quantidades muito limitadas) no núcleo, mitocôndrias e cloroplastos.

Cada ribossoma é dissociável por centrifugação em duas subunidades, com coeficientes de sedimentação S (de Svendberg<sup>1</sup>) de 50S e 30S nos procariotas, e 60S e 40S nos eucariotas. Ambas as subunidades contêm RNA<sub>t</sub> e proteínas, na proporção de 1/1 em muitos organismos, mas cerca de 2/1 na *E. coli* (Quadro 5.1.).

**Quadro 5.1.** Composição dos ribossomas citoplasmáticos de procariotas e eucariotas, com referência a coeficientes de sedimentação (S), tipo de moléculas proteicas e de ácido ribonucleico ribossômico (RNA<sub>r</sub>) em cada subunidade.

Estruturas	Procariotas			Eucariotas		
Ribossoma completo	70S			80S		
Sub-unidades ribossômicas	30S	e	50S	40S	e	60S
Composição das subunidades:	↓		↓	↓		↓
RNA <sub>r</sub>	16S		5S e 23S	18S		5S e 28S23
Proteínas (total)	21		34	23		50

<sup>1</sup> O termo “unidade de Svendberg (S)” provém do nome do químico sueco Theodor Svendberg (1884-1971), galardoado com o prêmio Nobel da Química em 1926 pelo mérito dos seus trabalhos sobre a sedimentação de substâncias coloidais e de elevado peso molecular em solução. Para aqueles estudos o Dr. Svendberg desenvolveu o primeiro modelo de ultracentrifuga, tendo concluído que S (equivalente a sedimentação, posteriormente definida em unidades S) era proporcional à velocidade de sedimentação num dado campo gravitacional e dependente da dimensão e forma das moléculas centrifugadas.

Nos ribossomas bacterianos foram identificados dois grandes tipos de RNA<sub>r</sub>, sendo necessário que ambos coexistam para que o ribossoma mantenha a sua estrutura. Apesar de as moléculas de RNA<sub>r</sub> serem formadas por uma cadeia simples de polirribonucleóticos, há suficiente equivalência de bases complementares para que algumas se emparelhem de modo semelhante às do RNA<sub>r</sub>, dando ao conjunto uma dimensão tri-dimensional irregular. A maioria dos constituintes proteicos dos ribossomas tem funções estruturais.

Apesar de ter sido possível, em estudos com bactérias, reconstruir a subunidade de 30S a partir dos seus constituintes (proteínas e RNA), originando partículas activas na síntese proteica, continua obscura a função exercida pelo RNA<sub>r</sub>. É sugerido que as zonas não emparelhadas favorecem a colocação correcta das moléculas de RNA<sub>m</sub> e RNA<sub>t</sub> intervenientes na síntese da cadeia peptídica, a qual cresce por adição consecutiva de novos resíduos de aminoácidos ao segmento pré-existente. Deste modo, cada cadeia começa por um grupo aminado (terminal) de um determinado aminoácido, cujo grupo carboxílico forma uma ligação peptídica com o grupo aminado do aminoácido seguinte, até surgir o sinal de terminação.

Cada ribossoma fornece superfícies de contacto ao RNA<sub>m</sub>, às moléculas de aminoacil-RNA<sub>t</sub> e à cadeia polipeptídica em formação. A fracção de RNA<sub>m</sub> ligada aos ribossomas é muito reduzida (aproximadamente 1%), unindo-se, de modo reversível, à subunidade mais pequena dos ribossomas.

Somente um polipéptido pode ser sintetizado de cada vez no mesmo ribossoma. Na *E. coli* bastam 10 segundos, em condições favoráveis, para que se obtenha um polipéptido com o peso molecular de 40.000; o ribossoma fica apto a iniciar a síntese de outra cadeia peptídica assim que se liberta do polipéptido recém-sintetizado. Nos reticulocitos de coelho, cada cadeia  $\alpha$  da hemoglobina (que possui cerca de 150 resíduos de aminoácidos) é completada em três minutos, a 37°C.

## Fases da tradução

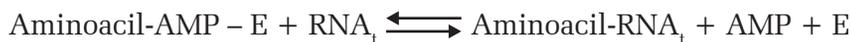
No seu conjunto, a tradução da mensagem genética no RNA<sub>m</sub> processa-se através de um sistema complexo que inclui os seguintes intervenientes principais: aminoácidos essenciais, ribossomas, aminoacil-sintetases, três tipos de ácido ribonucleico (RNA<sub>m</sub>, RNA<sub>t</sub> e RNA<sub>r</sub>) e energia química. O processo de tradução envolve quatro fases principais: (a) *activação dos aminoácidos*, (b) *iniciação da cadeia peptídica*, (c) *alongamento da cadeia peptídica* e (d) *terminação da cadeia peptídica*.

**(a) Activação dos aminoácidos** – A etapa decorre na fracção solúvel do citoplasma, onde cada um dos vinte aminoácidos diferentes que constituem normalmente as proteínas são activados por enzimas específicas de cada um deles (pelo menos nas bactérias), as *aminoacil RNA<sub>t</sub> sintetases*.

Estas enzimas catalisam duas etapas sequenciais (activação e transferência de aminoácidos). Na primeira decorre a activação do aminoácido (AA) a ser incorporado na cadeia polipeptídica. Dessa reacção endergónica, catalisada pela sintetase (E) específica do AA, resulta o composto aminoacil-AMP-E:



Na reacção seguinte, a sintetase transfere o radical aminoacilo (a que se encontra ligado) para a extremidade 3' do respectivo RNA<sub>t</sub>; o aminoácido é unido por uma ligação éster à correspondente molécula de RNA<sub>t</sub>, numa reacção também com características endergónicas:



A reacção global da activação de cada aminoácido em condições celulares é representada do seguinte modo:



em que



A hidrólise do pirofosfato (PPi) pela pirofosfatase em duas moléculas de ortofosfato (Pi) determina o sentido da reacção, que é o da união do aminoácido activado a uma molécula de RNA<sub>t</sub>.

A fidelidade da cópia polipeptídica resultante é, em parte, explicada pela dupla especificidade de cada molécula de aminoacil-RNA<sub>t</sub>.sintetase: por um lado reconhece apenas o seu aminoácido, entre os vinte aminoácidos essenciais e, por outro, reconhece um ou alguns dos RNA<sub>t</sub> que transferem aquele aminoácido.

**(b) Iniciação da cadeia polipeptídica** – Uma das questões inicialmente mais enigmáticas foi a das características do primeiro aminoácido de cada cadeia polipeptídica, e de que modo seria reconhecido pelo complexo ribossômico. Teoricamente, havia um codão próprio para aquele aminoácido ou, em alternativa, o primeiro aminoácido teria estrutura distinguível. Estudos realizados em culturas bacterianas (e posteriormente também em cloroplastos e mitocôndrias de eucariotas) revelaram que a generalidade das cadeias polipeptídicas era iniciada por um resíduo de metionina formilada (N-formilmetionina; fMet). O grupo *formilo* provém de uma das formas de ácido fólico (N<sup>10</sup>-formil-tetrahidrofolato; N<sup>10</sup>-formil-FH<sub>4</sub>), sendo a sua ligação à metionina catalisada por uma *transformilase*. Aquela reacção ocorre somente depois da metionina estar unida a uma molécula específica de RNA<sub>t</sub> (RNA<sub>tF</sub>):



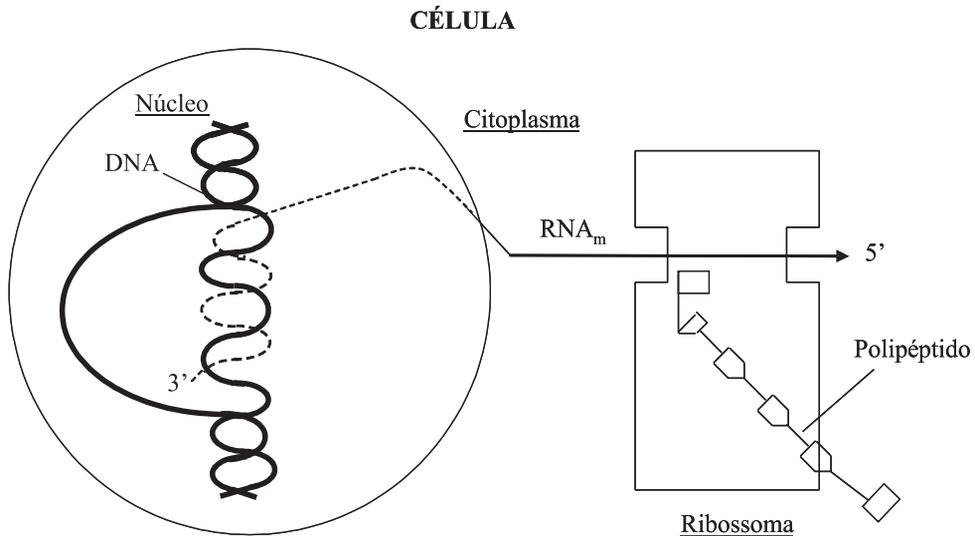
Entretanto, foi verificada a existência de dois tipos de RNA<sub>t</sub> específicos para a metionina; um dos tipos (RNA<sub>tF</sub>) fixa a metionina com aptência para a formilação (e que, portanto, constitui o aminoácido iniciador), enquanto o segundo tipo transfere a metionina *não formilada* para outras posições na sequência polipeptídica.

Uma vez iniciada a síntese polipeptídica, e antes da sua conclusão, o grupo N-formilo é removido enzimaticamente por uma *desformilase*; por vezes, é também removida a metionina inicial por uma *aminopeptidase*, juntamente com alguns resíduos de aminoácido que lhe estão imediatamente ligados.

Na iniciação da cadeia polipeptídica são identificadas três etapas sucessivas que culminam na formação dos designados complexos I, II e III.

No *complexo I* participam os seguintes componentes principais: o aminoácido iniciador unido ao respectivo RNA<sub>t</sub> (formil-metionil- RNA<sub>t</sub>; fMet- RNA<sub>t</sub>), cadeia de RNA<sub>m</sub>, subunidade 30S dos ribossomas, energia química (sob a forma de guanosina 5'-trifosfato; GTP) e três proteínas que actuam como factores de iniciação (F<sub>1</sub>, F<sub>2</sub> e F<sub>3</sub>) da cadeia polipeptídica.

A molécula de RNA<sub>m</sub> desloca-se sempre no mesmo sentido (5' → 3') assim que se une ao ribossoma. A extremidade 5', sintetizada em primeiro lugar, também é a primeira a ser “lida”, podendo suceder que o resto da cadeia de RNA<sub>m</sub> ainda esteja a ser transcrita pelo DNA enquanto, simultaneamente, a sua primeira posição já está a ser traduzida nos ribossomas (Fig. 5.3.).



**Fig. 5.3.** A transdução do RNA<sub>m</sub> nos ribossomas conduz à síntese de um polipéptido específico codificado a partir de uma das cadeias de DNA duplex, cuja transcrição da cadeia de RNA<sub>m</sub> poderá estar a decorrer simultaneamente.

Nos ribossomas completos (70S), existem dois sítios para a inserção das moléculas de aminoacil- RNA<sub>t</sub> a serem traduzidas, um para o RNA<sub>t</sub> unido ao péptido em formação (sítio “P”, de **p**éptido) e outro para o RNA<sub>t</sub> com o aminoácido seguinte (sítio “A”, de **a**cilo). Aquelas posições contactam com as duas subunidades ribossómicas e com os codões do RNA<sub>m</sub> a serem transcritos.

Durante a síntese proteica observa-se uma continuada dissociação dos ribossomas 70S nas subunidades 30S e 50S, a que se segue a sua reassociação na estrutura 70S. Na sequência dos estudos realizados em *E. coli* concluiu-se que, para ser iniciada a transcrição, era indispensável que os ribossomas inativos (70S) se dissociassem nas suas duas subunidades. Somente depois, o segmento de RNA<sub>m</sub> a ser transcrito e o fMet- RNA<sub>t</sub><sub>if</sub> poderiam unir-se à subunidade ribossómica 30S, originando o *complexo iniciador* (ou *complexo I*).

Aparentemente, a formação do complexo I resulta da junção de dois agregados, um representado pela subunidade 30S e pelo segmento do RNA<sub>m</sub> a ser transcrito (unidos entre si por pontes de hidrogénio entre bases complementares), sendo o outro constituído pelo composto fMet-RNA<sub>t</sub><sub>if</sub>. O factor de iniciação 3 (F<sub>3</sub>) será indispensável para que a subunidade 30S se una ao RNA<sub>m</sub>,

após o que é fixado também o factor  $F_1$ . Do mesmo modo, a ligação de Fmet-RNA<sub>t</sub> à subunidade ribossómica 30S requer a presença de GTP e de  $F_2$ . Eventualmente, a junção dos dois agregados é activada de um lado por  $F_1$  e do outro por  $F_2$ .

Na sequência, o complexo I une-se à subunidade 50S, daí resultando a re-associação do ribossoma 70S completo (*complexo II*), no qual o fMet-RNA<sub>t</sub> parece estar posicionado no sítio “A”.

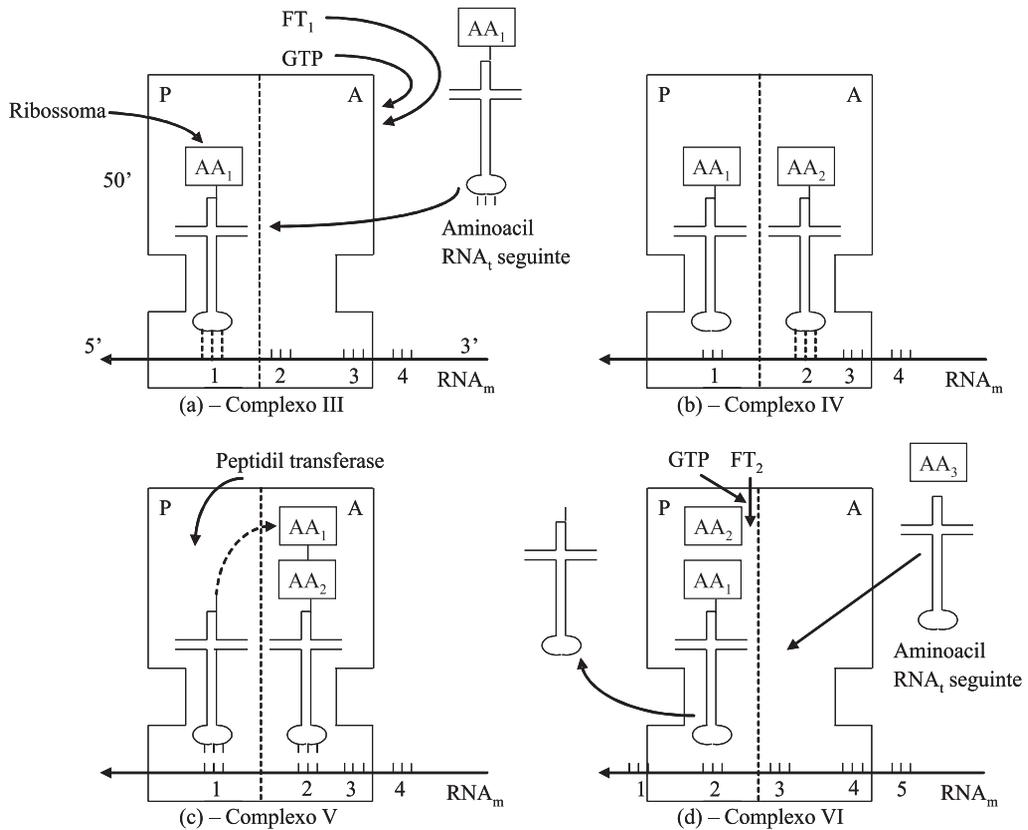
A formação do *complexo III* consiste na transferência do complexo iniciador do polipéptido (Fmet-RNA<sub>tF</sub>) do sítio “A” para o sítio “P” que lhe está adjacente na subunidade 50S. Esta transferência requer energia química (resultante de hidrólise do GTP em GDP e Pi) e decorre com a libertação dos três factores de iniciação, os quais ficam disponíveis para o recomeço de outra cadeia peptídica.

Na realidade, não está bem esclarecido se o F-met RNA<sub>t</sub> se fixa primeiro no sítio “P” ou no sítio “A” (e daqui é transferido para o “P”), mas um facto é certo: aquele conjunto tem de estar no sítio “P” para que a segunda molécula de amino acil-RNA<sub>t</sub> possa ligar-se ao ribossoma.

**(c) Alongamento da cadeia peptídica** – As cadeias polipeptídicas crescem por adição sucessiva de aminoácidos, de modo que, iniciado pelo grupo aminado do primeiro aminoácido, cada polipéptido termine com o grupo carboxílico do último dos resíduos. À medida que é transportado para os ribossomas, cada aminoácido conjuga o seu grupo aminado com o grupo carboxilo do resíduo antecedente, formando uma ligação peptídica (Fig. 5.4.).

Este processo foi bastante bem ilustrado em reticulocitos incubados com aminoácidos marcados por isótopos radioactivos. À medida que o tempo de incubação aumentava, isolavam-se cadeias de hemoglobina em que a radioactividade abrangia cada vez maior quantidade do polipéptido, com aumento gradual da extremidade aminada para a carboxílica.

São identificáveis três etapas distintas durante o alongamento da cadeia polipeptídica. Na primeira (que corresponde à formação do designado *complexo IV*), o ribossoma recebe o aminoácido seguinte ( $AA_2$ ). Segue-se a formação do *complexo V*, representada pela ligação peptídica entre os dois aminoácidos presentes no ribossoma, após o que ocorre a translocação (*complexo VI*) do RNA<sub>t</sub> do dipéptido para a posição “P” que estava ocupada pela molécula antecedente de RNA<sub>t</sub> desacilado, entretanto expelida para o citosol.



**Fig. 5.4.** Esquema representativo do alongamento da cadeia polipeptídica.

(a) O aminoácido iniciador, sob a forma de composto formilmietionina- $\text{RNA}_t$ , está localizado no sítio peptidil; (b) fixação do 2.º aminoácido ( $\text{AA}_2$ ) transportado como molécula de aminoacil  $\text{RNA}_t$ , no sítio aminoacilo (“A”) do complexo ribossómico com a participação do factor de transferência ( $\text{FT}_1$ ); (c) formação de uma ligação peptídica entre o grupo carboxilo terminal do aminoácido pré-existente (ligação a uma molécula específica de  $\text{RNA}_t$  alojada no sítio peptidil (“P”) do complexo ribossómico) e o grupo aminado do novo aminoácido; (d) por acção da enzima peptidil-transferase, que também intervém na formação da ligação peptídica com a participação do factor de transferência 2 ( $\text{FT}_2$ ) e consumo de energia (da hidrólise do GTP), da molécula de  $\text{RNA}_t$  (que perdeu o respectivo aminoácido ou péptido) é removido do sítio peptidil, dando lugar à molécula peptidil- $\text{RNA}_t$  que se encontrava no sítio “A”, enquanto o terceiro aminoácido está em vias de se fixar no sítio “A”. As sucessivas alterações de posição decorrem a par com o movimento da cadeia de  $\text{RNA}_m$  na interface ribossómica, com exposição dos codões específicos para cada conjunto aminoacil- $\text{RNA}_t$  (posicionamento numerado).

Na formação do complexo IV cada novo aminoacil-RNA<sub>t</sub> (excepto o primeiro) liga-se originalmente ao sítio aminoacilo (“A”) do complexo III. As novas moléculas do aminoacil-RNA<sub>t</sub> são posicionadas de acordo com o anticódon do RNA<sub>t</sub>, complementar ao codão de RNA<sub>m</sub> a ser transcrito. Para que esta união ocorra é requerida energia química (derivada da hidrólise da GTP em GDP + Pi) e uma proteína específica (*factor de alongamento ou de transferência 1*; FT<sub>1</sub>). Nem esta nem uma outra proteína funcionalmente semelhante (*factor de transferência 2*; FT<sub>2</sub>) são proteínas ribossómicas, o que aliás também sucede com os factores de iniciação.

A reacção entre AA-RNA<sub>t</sub>, GTP e FT<sub>1</sub> origina um agregado que, ao transferir o componente AA-RNA<sub>t</sub> para o sítio “A”, liberta o GDP-FT<sub>1</sub> e fosfato inorgânico (Pi). Desta sequenciação resulta o complexo IV, constituído por um ribossoma com duas moléculas de RNA<sub>t</sub> aminoaciladas, uma na posição “A” (AA-RNA<sub>t</sub>) e outra na posição “P” (fMet-RNA<sub>f</sub>).

O complexo V caracteriza-se pela formação de uma ligação peptídica, entre aqueles dois grupos aminoacilo, nas posições “A” e “P” dos ribossomas. Na reacção, o grupo aminado da molécula de AA<sub>2</sub>-RNA<sub>t</sub> em “A” rompe (por deslocação nucleofílica) a ligação éster existente entre o grupo carboxilo (do aminoácido anterior) e o grupo hidroxilo 2’ ou 3’ do resíduo de adenilato, da extremidade 3’ do respectivo RNA<sub>f</sub>. A enzima específica desta etapa, *peptidil-transferase*, que parece existir nas subunidades 50S, não requer grupos fosfato de alta energia (ATP ou GTP) para actuar, sendo-lhe aparentemente suficiente a energia fornecida pela clivagem daquela ligação éster. Ao ser formada a ligação peptídica, o aminoácido (fMet) unido ao RNA<sub>f</sub> na posição “P” é transferido para a posição “A”, cujo RNA<sub>t</sub> passa a ligar-se a dois grupos aminoacilo (dipeptidil-RNA<sub>t</sub>), enquanto a molécula de RNA<sub>f</sub> na posição “P” do ribossoma fica desacilada, sendo depois expelida.

Para que o alongamento da cadeia polipeptídica prossiga torna-se necessário que a molécula peptidil-RNA<sub>t</sub> seja translocada da posição “A” para a “P”, de modo a dar entrada a nova molécula de RNA<sub>t</sub> aminoacilada. Esta fase de translocação (que corresponde à formação do complexo VI) requer que a posição “P” seja libertada do RNA<sub>f</sub>; seguidamente, o dipeptidil-RNA<sub>t</sub> (fMet-AA<sub>2</sub>-RNA<sub>2</sub>) situado na posição “A” é translocado para a posição “P” livre, com a intervenção do segundo factor de transferência (FT<sub>2</sub> ou translocase) e consumo energético (por hidrólise de GTP em GDP e Pi).

Em simultâneo com a translocação, decorre a deslocação gradual da cadeia de RNA<sub>m</sub> ao longo do ribossoma, de modo a posicionar o codão para o peptidil-RNA<sub>t</sub> em “P” e o novo tripleto de nucleótidos para o aminoacil-RNA<sub>t</sub> seguinte em “A”. O RNA<sub>m</sub> como que desliza protegido num “canal” formado pelas duas subunidades ribossómicas que, em cada momento, contém segmentos com cerca de sessenta mononucleótidos por ribossoma.

O processo descrito anteriormente para alongamento da cadeia peptídica repete-se para cada novo aminoácido que é transportado pelo respectivo RNA<sub>t</sub> para o sítio “A” do ribossoma, até à sua finalização. Neste período, a cadeia polipeptídica que se une ao ribossoma (do tipo Fmet-AA<sub>2</sub>-AA<sub>3</sub>—AA<sub>n</sub>-RNA<sub>t</sub>) cresce a partir da extremidade aminada para a carboxílica.

**(d) Terminação da cadeia peptídica e conformação final** – A cadeia polipeptídica separa-se do ribossoma depois de conter todos os seus aminoácidos, pré-determinados pelo respectivo gene. A finalização da cadeia é sinalizada por três codões especiais do RNA<sub>m</sub> (*codões de finalização*; ver Capítulo 6 – Código Genético). Só então o polipéptido completo se separa do seu RNA<sub>t</sub> final, por hidrólise da ligação éster entre o grupo carboxilo terminal e o RNA<sub>t</sub>. Esta fase é mediada por proteínas específicas dos ribossomas (*factores de libertação*,  $R_1$ ,  $R_2$  e  $R_3$ ), em resposta aos codões do RNA<sub>m</sub> que anunciam o “fim de cadeia”. Ao ligarem-se ao ribossoma, os factores de libertação induzem a translocação do peptidil-RNA<sub>t</sub> final do sítio “A” para o sítio “P”, após o que é hidrolisada a nova ligação entre a cadeia polipeptídica e aquele RNA<sub>t</sub>. O polipéptido é libertado, seguindo-se a dissociação do agregado, constituído por ribossoma, molécula de RNA<sub>m</sub> e RNA<sub>t</sub>.

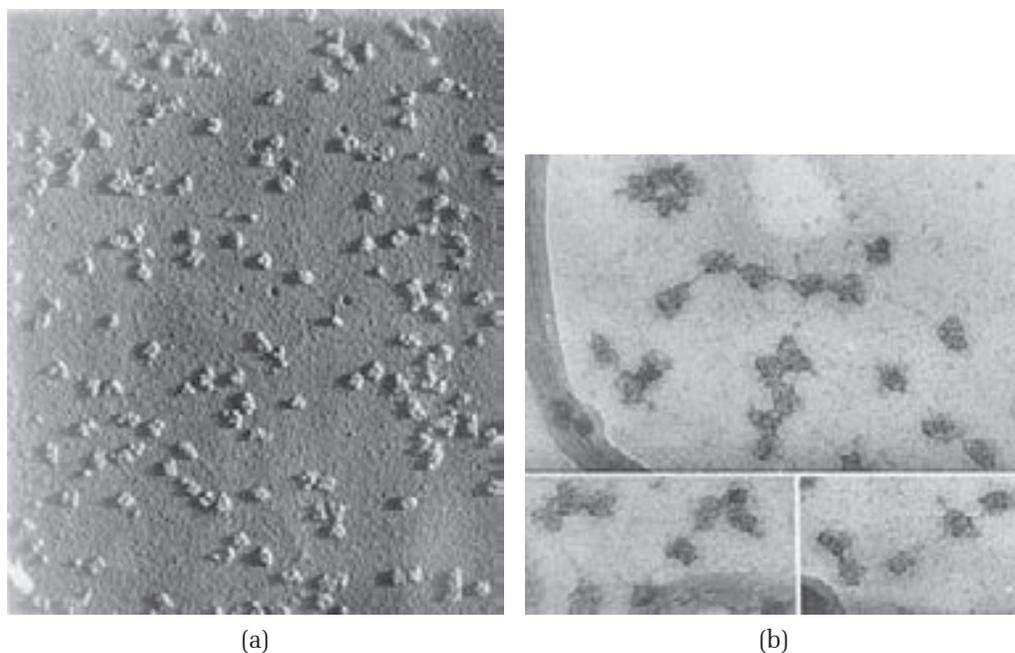
À medida que a cadeia polipeptídica é formada e se alonga decorre o seu enrolamento numa estrutura tridimensional, que será semelhante à da conformação definitiva depois de se libertar do RNA<sub>t</sub> final e ser remodelada. Esta remodelação na forma definitiva traduz-se pela remoção de determinados aminoácidos (que não fazem parte do polipéptido final), que poderão estar localizados no início (p.ex., o formilmetionina iniciador) ou no fim da cadeia (por exopeptidases), ou no seu interior (por acção da endopeptidases).

Aparentemente as proteínas dissociam-se do ribossoma com a sua estrutura final completa e conformação final, isto é, aptas a entrar em actividade.

## Síntese polipeptídica em polirribosomas

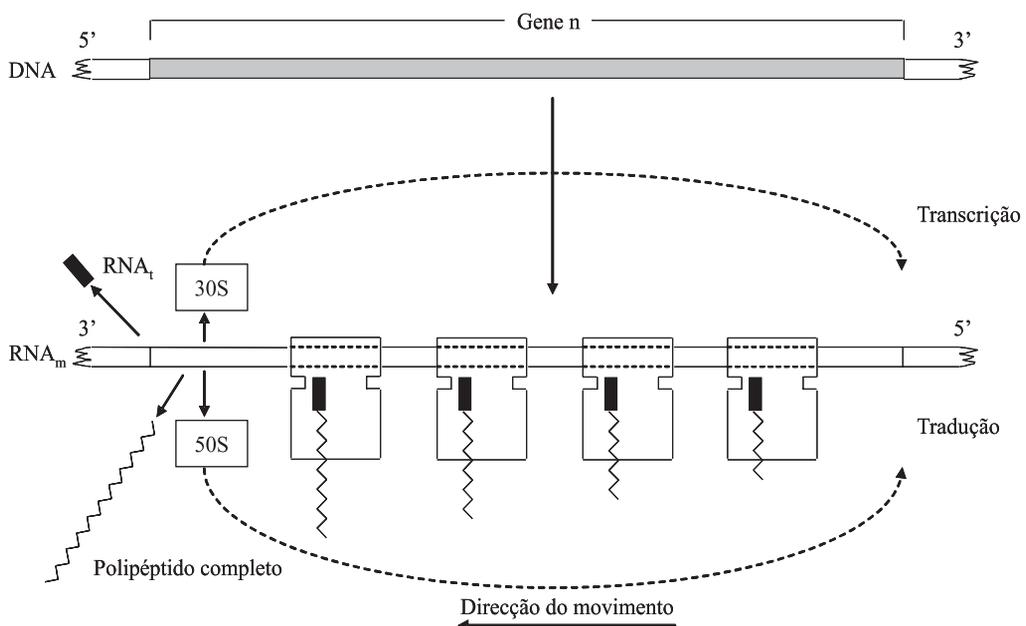
Consoante as proteínas sejam requeridas para a renovação da estrutura celular ou para formar secreções assim a síntese é mediada por ribossomas livres ou ligados ao retículo-endoplásmico, respectivamente.

Ainda que cada ribossoma isolado possa formar uma cadeia polipeptídica, a associação de ribossomas em grupos ligados pela mesma cadeia de RNA<sub>m</sub> (*polissoma* ou *polirribossoma*) aumenta a eficiência da transdução genética (Fig. 5.5.). O número de ribossomas de um polissoma é proporcional, aproximadamente, ao número de aminoácidos que constituem o péptido a ser sintetizado. À medida que



**Fig. 5.5.** Micrografia electrónica de polirribosomas de reticulocito de coelho. Em (a) são visíveis os aglomerados de número variável de ribossomas sombreados com platina, em alguns dos quais parece evidenciar-se um segmento de interligação (ampliação de 100.000 diâmetros); (b) aquele segmento constituído por uma molécula de RNA<sub>m</sub>, conecta conjuntos de 4 a 6 ribossomas que ficam transitoriamente unidos, como um colar de pérolas, enquanto neles decorrer a síntese das cadeias polipeptídicas (coloração de acetato de urânio; ampliação de 400 000 diâmetros). Reprodução autorizada das fotografias amavelmente oferecidas pelo Dr. Alexander Rich (Department of Biophysics, Massachusetts Institute of Technology, USA).

aumenta o número de aminoácidos da cadeia peptídica, o polirribosoma tende a organizar-se tri-dimensionalmente. Esta organização possibilita a formação simultânea de várias cadeias polipeptídicas ao longo da mesma cadeia de RNA<sub>m</sub> (Fig. 5.6.), no sentido 5' → 3', em que não só a função de cada um dos ribossomas está sincronizada em relação aos outros, como também o RNA<sub>m</sub> sintetizado é muito melhor aproveitado do que sucederia se os ribossomas actuassem isolados.



**Fig. 5.6.** Representação esquemática de um polirribossoma onde se sintetiza determinado polipéptido em cada uma das unidades ribossómicas. A sequência deste é codificada pela mensagem que o DNA transcreve na RNAm.

## Conclusões

O DNA não é o molde directo da síntese proteica. A informação genética do DNA é primeiramente transcrita para moléculas de RNA mensageiro, as quais determinam a constituição de cada polipéptido. A síntese das proteínas decorre nos ribossomas (constituídos por proteínas específicas e outro tipo de RNA, o

RNA<sub>t</sub>, que não contém informação genética) a partir de aminoácidos activados no citoplasma por sintetases específicas de cada aminoácido essencial e da molécula de RNA<sub>t</sub> que lhe serve de “adaptador molecular”, originando os compostos aminoacil-RNA<sub>t</sub>. O reconhecimento do aminoácido transportado pela RNA<sub>t</sub> é determinada pela sintetase. Em geral, existe mais de uma molécula de RNA<sub>t</sub> por cada aminoácido. A especificidade das moléculas de RNA<sub>t</sub> é determinada por um triplete de mononucleótidos (anticodão) localizado numa das suas ansas. Cada anticodão é complementar de outro triplete complementar da molécula de RNA<sub>m</sub>, a qual codifica a selecção do aminoácido (codão). Em cada ribossoma estão identificados dois sítios (“A”, de aminoácido e “P”, de péptido) de interacção entre codão e anticodão. Quando um codão de RNA<sub>m</sub> é posicionado em “A”, o RNA<sub>t</sub> com anticodão complementar e que transporta o respectivo aminoácido une-se (por ligações hidrogeniónicas transitórias) naquela posição, sendo depois transferido para a posição “P”, a par com a deslocação de RNA<sub>m</sub> que, deste modo, faz chegar sucessivos codões à posição “A” da interface ribossómica.

O crescimento da cadeia polipeptídica decorre por etapas sucessivas: activação dos aminoácidos, iniciação da cadeia, alongamento da cadeia e terminação da cadeia péptida. Nos procariotas (e presume-se que também sucede nos eucariotas) cada cadeia é iniciada pela metionina formilada, sob a forma de N-formilmetionina-RNA<sub>t</sub> (Fmet-RNA<sub>tf</sub>), quando o codão iniciador AUG no RNAm se une à subunidade 30S dos ribossomas. O ribossoma, que se encontrava dividido nas suas duas subunidades (30S e 50S), é restaurado na forma completa 70S, a qual é mantida enquanto decorrer a síntese completa de cada polipéptido, codificado pela molécula de RNA<sub>m</sub>. A formação do complexo iniciador descrito tem a intervenção sequencial de três proteínas (F<sub>1</sub>, F<sub>2</sub> e F<sub>3</sub>) não ribossómicas. No decurso desta fase, o composto Fmet-RNA<sub>tf</sub> está colocado no sítio “P” até à chegada do aminoácido seguinte transportado pela respectiva molécula de RNA<sub>t</sub> o subsequente alongamento da cadeia decorre no seguinte modo: fixação do novo aminoácido RNA<sub>t</sub> em “A” (na presença de uma proteína de transferência FT<sub>1</sub> e GTP), formação de ligação peptídica com o aminoácido precedente (Fmet) pela peptidil-transferase, com o dipéptido na posição “A” e o RNA<sub>tf</sub> livre em “P”; a translocação de “A” para “P” decorre com a expulsão de RNA<sub>tf</sub> e “P” requer uma segunda proteína de transferência (FT<sub>2</sub>) e energia (proveniente da hidrólise do GTP).

O crescimento da cadeia decorre no mesmo modo por cada novo aminoácido a ser incorporado e à medida que novos codões de RNA<sub>m</sub> são posicionados na posição “A” do ribossoma. A síntese do polipéptido termina ao serem reco-

nhecidos os tripletos de terminação (UAA, UAG e AGA) na cadeia de RNA<sub>m</sub> em tradução; três proteínas ribossômicas específicas sinalizam aqueles codões (factor de libertação R<sub>1</sub>, R<sub>2</sub> e R<sub>3</sub>) e promovem a separação do polipéptido finalizado do ribossoma, dissociação deste nas sub-unidades 30S e 50S, e a separação do RNA<sub>m</sub>-molde. Em cada ligação peptídica formada é utilizada energia química equivalente à hidrólise de, pelo menos, três ligações de alta energia. Os ribossomas podem actuar em conjunto na tradução da mesma cadeia de RNAm, formando polirribossomas, em que cada ribossoma sintetiza independentemente a sua cadeia polipeptídica.

Nas células de eucariotas, a síntese das proteínas de secreção decorre no reticulo-endoplasmático. Também nas mitocôndrias são sintetizados alguns tipos próprios de proteínas, as quais são codificadas de uma molécula de DNA circula por polimerases (do DNA e RNA) e tipos específicos de RNAt, enzimas e ribossomas semelhantes aos que actuam no citosol.

# CAPÍTULO 6

## CÓDIGO GENÉTICO

### Definição do código

São conhecidas as sequências de mononucleótidos que codificam cada um dos aminoácidos (Quadro 4.1.). No conjunto, os quatro tipos de base que constituem os mononucleótidos incorporados na estrutura dos ácidos nucleicos determinam, em (diversas) combinações de três mononucleótidos em sequência (tripletos), o posicionamento de cada um dos vinte aminoácidos comuns na sua posição correcta na estrutura peptídica codificada.

A diversidade de cadeias peptídicas sintetizadas, cada qual com características estruturais e funções específicas, resulta da sequência dos aminoácidos posicionados, a qual é determinada pela sequência dos codões traduzidos. Através de cálculos matemáticos simples quanto ao número de arranjos das quatro bases que poderiam codificar os vinte aminoácidos essenciais, concluiu-se que para combinações duas a duas seriam codificados somente 16 aminoácidos ( $4^2$ ), o que era insuficiente; entretanto, combinações de três bases ( $4^3 = 64$ ) excediam largamente o número de aminoácidos diferentes. Este excedente fazia pressupor a possibilidade de mais de um tripleto codificar o mesmo aminoácido, o que realmente foi confirmado.

A constituição de cada unidade de codificação nucleotídica (codão) para cada aminoácido foi esclarecida com a utilização sistemática de polinucleótidos sintéticos com composição ou proporção variáveis que, adicionados a ribossomas isolados, conduziam à formação de sequências de determinado(s) aminoácido(s).

**Quadro 4.1.** Código Genético

Disposição convencional do ordenamento dos nucleótidos constituintes de cada tripleto em cada molécula de RNA<sub>m</sub> determinado pela respectiva base (letra). A cada tripleto (codão) corresponde um aminoácido. A sequência (da 1.<sup>a</sup> para a 3.<sup>a</sup> letra) corresponde à direcção da leitura da extremidade 5' para a 3'. A sequência AUG define o sinal de iniciação (SI) da tradução; as sequências UAA, UAG e UGA são sinais de terminação (ST). O codão GUG da valina por vezes actua como sinal de iniciação da cadeia, ao codificar a metionina. (Nota: é mantida a nomenclatura anglo-saxónica na abreviatura dos aminoácidos).

Código genético					
1. <sup>a</sup> letra	2. <sup>a</sup> letra				3. <sup>a</sup> letra
	U	C	A	G	
U	PHE	SER	TYR	CYS	U
	PHE	SER	TYR	CYS	C
	LEU	SER	(ST)	(ST)	A
	LEU	SER	(ST)	TRY	G
C	LEU	PRO	HIS	ARG	U
	LEU	PRO	HIS	ARG	C
	LEU	PRO	GLN	ARG	A
	LEU	PRO	GLN	ARG	G
A	ILEU	THR	ASN	SER	U
	ILEU	THR	ASN	SER	C
	ILEU	THR	LYS	ARG	A
	MET SI	THR	LYS	ARG	G
G	VAL	ALA	ASP	GLY	U
	VAL	ALA	ASP	GLY	C
	VAL	ALA	GLU	GLY	A
	VAL (SI)	ALA	GLU	GLY	G

A primeira experiência do género foi conduzida com um polímero de ácido uridílico (poli U), utilizando ribossomas isolados de *E. coli* e todos os meios necessários à mistura proteica, distribuídos por uma série de tubos, cada um com um aminoácido diferente marcado. Ao ser analisada a fracção polipeptídica formada verificou-se que apenas a fenilalanina havia sido incorporada na sua estrutura. Concluiu-se que o codão para a fenilalanina seria constituído por um tripleto de resíduos de ácido uridílico. Ensaio idênticos esclareceram ser a prolina, a lisina e a glicina, respectivamente, codificadas por polímeros

de ácidos citidílico (codão CCC), adenílico (codão AAA) e guanílico (codão GGG).

Na fase seguinte recorreu-se ao uso de polímeros mistos (p.ex., poli UG). Ainda que fosse desconhecida a sequência das bases, sabendo-se a relação mútua entre ambas, era possível prever os tripletos viáveis e em que posição se situavam.

Utilizando proporções diferentes de dois tipos de mononucleótidos (p.ex., duas vezes mais moléculas de ácido uridílico do que ácido guanílico, resultavam sequências repetidas variáveis (tais como: UUU, UUG, UGG, GGU, GUG, GGG) e grande diversidade de aminoácidos formados. Embora a composição em aminoácidos fosse conhecida, a dos tripletos codificadores continuava por esclarecer.

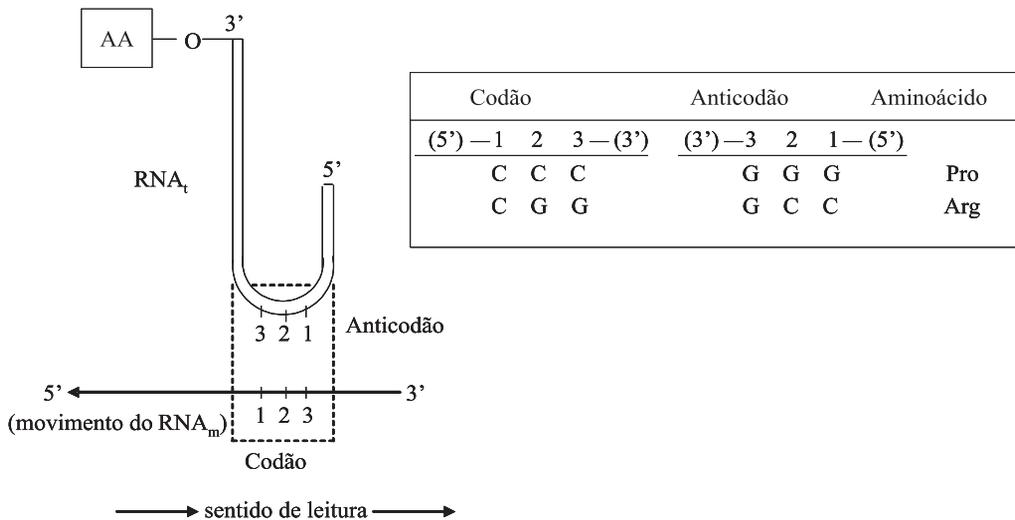
Comparando aqueles dados com o número de vezes em que determinados aminoácidos eram incorporados nas proteínas, foi possível estabelecer um certo número de relações constantes entre tripletos e aminoácidos codificados. Nestas experiências foi ainda confirmado o que havia sido inicialmente uma hipótese: o mesmo aminoácido poderia ser codificado por mais de um triplete.

## Iniciação e terminação de leitura do código genético

Para a mensagem ser entendida e, portanto, haver a incorporação de determinado aminoácido na sequência polipeptídica em formação, os trinucleótidos constituintes de cada triplete têm de ser “lidos” num sentido específico (5' → 3'). Justifica-se, por isso que a sequência CCCCGG codifique para prolina-arginina e não para arginina-prolina (Fig. 6.1.).

Para que assim suceda, os grupos terminais de triplete deverão ter um grupo hidroxilo livre terminal na extremidade 3' e um grupo fosforilado na extremidade 5'. A fosforilação da extremidade-3' inactiva o trinucleótido.

Os ensaios em ribossomas isolados de *E. coli* permitiu a identificação progressiva de vários codões, o que foi reforçado com a utilização de polirribonucleótidos sintéticos em que eram repetidas determinadas sequências (p.ex., AUGAUGAUGAUGAUG etc), dos quais se individualizam tripletos diferentes consoante o ponto de início da leitura (p.ex., AUG, UGA, GAU); estes tripletos indicam a fixação nos ribossomas de moléculas aminoacil-RNA<sub>t</sub> diferentes (p.ex., metionina, codão de finalização, aspartato).



**Fig. 6.1.** Selecção de uma molécula de aminoacil-RNA<sub>t</sub> por um codão do RNA<sub>m</sub>. A molécula de RNA<sub>t</sub> (que possui um anticodão específico numa das extremidades) transporta um determinado aminoácido, ligado na extremidade 3' oposta, para a posição que vai ocupar na sequência do polipéptido em formação. Esse posicionamento é determinado pelo codão do RNA<sub>m</sub> a ser traduzido. O emparelhamento do codão do RNA<sub>m</sub> com o anticodão do RNA<sub>t</sub> é complementar e antiparalelo. A leitura do codão decorre da esquerda (extremidade 5') para a direita (extremidade 3'), sendo a disposição do codão em sentido inverso.

O sinal de *início* e *fim* de determinado código deverá estar colocado correctamente, sem o que resultará numa proteína anómala. Os três tripletos (de um total de sessenta e quatro possíveis) que não codificam aminoácidos conhecidos (UAG, UAA, UGA) simplesmente assinalam o “*fim da leitura*” (*codões de finalização*) da cadeia peptídica.

O *início da codificação* é determinado pelo codão AUG, a que corresponde a formilmetionina (associada ao RNA<sub>tr</sub>), mas que também codifica a metionina (transferida pela RNA<sub>tr</sub>) para posições no interior da cadeia polipeptídica.

## Degenerescência do código e mutações

Quase todos os aminoácidos (exceptuando o triptofano e a metionina) são codificados por mais de um triplete, embora com participação variável. Crê-se que esta *degenerescência de código* tenha vantagens biológicas.

O terceiro nucleótido de cada código é menos específico do que os dois primeiros. Assim, dois aminoácidos diferentes podem ser codificados por tripletos que difiram entre si somente na terceira posição: XYA ou XYG, XYU ou XYG.

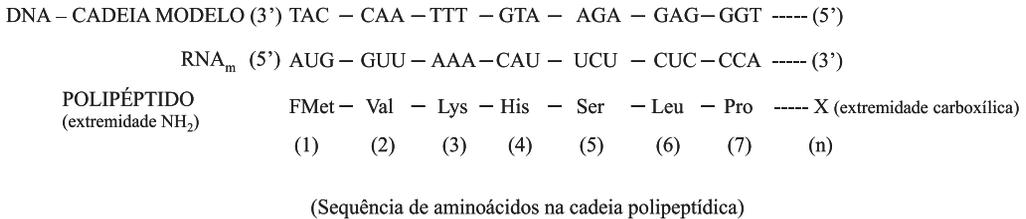
Os anteriores ensaios *in vitro* foram posteriormente confirmados através de mutações induzidas em bactérias e em mutações espontâneas detectadas em humanos, o que indica ser o código genético universal. Esta característica foi amplamente demonstrada quando, ao incubar ribossomas de *E. coli* e polinucleótidos sintéticos com várias moléculas de aminoacil-RNA<sub>i</sub> extraídas do fígado de cobaias, se verificou haver leitura correcta do código. Pequenas diferenças quanto à fixação de alguns aminoacil-RNA<sub>i</sub> no código, consoante são de *E. coli* ou de cobaia, sugerem haver diferenças evolutivas respeitantes às proteínas em ambas as espécies.

O estudo das variantes de hemoglobina humana, quer das cadeias a ou b, forneceu exemplos claros sobre processo genético. Progressivamente, foram identificados os pontos de substituição dos aminoácidos ao longo das cadeias de hemoglobina, resultantes da alteração de um único nucleótido, contudo suficiente para modificar o sentido codificador do tripleto. Estabeleceu-se assim que a maioria das mutações espontâneas seriam provocadas pela troca de uma só base no tripleto (*mutações pontuais*), mesmo que continuasse a ser uma base do mesmo tipo (púrica ou pirimídica).

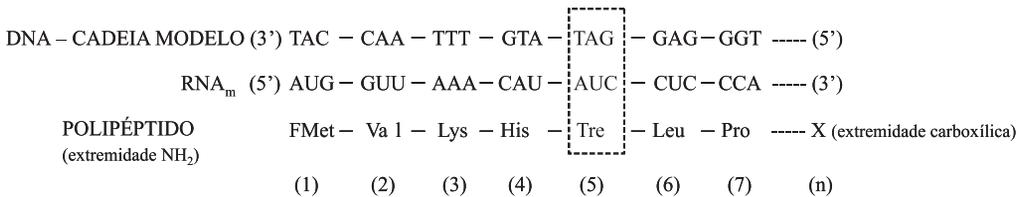
Por análise seriada dos aminoácidos de ambas as cadeias de hemoglobina em doentes com vários tipos de hemoglobinopatias, foi verificado que a substituição de um aminoácido num determinado ponto de qualquer das cadeias altera as características globais da hemoglobina (Fig. 6.2.).

Para esclarecer se haveria alguma relação entre o aminoácido substituído e a localização relativa da mutação no respectivo gene, e por ser muito difícil fazer pesquisas deste género em mamíferos, recorreu-se novamente a microorganismos. Na *E. coli*, a disposição dos pontos mutantes na estrutura linear do gene que codifica uma das cadeias polipeptídicas de triptofano-sintetase estava exactamente correlacionada com a posição relativa dos aminoácidos substituídos na sequência normal. Esta observação concorda em absoluto com a correspondência linear existente entre a sequência dos codões no gene e a sequência de aminoácidos na cadeia polipeptídica que codifica. Entretanto, experiências em mutantes de *E. coli* sujeitos a substâncias mutagénicas demonstraram não ser necessária a mesma sequência de aminoácidos em determinadas zonas da cadeia polipeptídica A, para que a triptofano sintetase fosse activa. Assim, é possível

(a)



(b)



**Fig. 6.2.** Colinearidade de uma determinada sequência nucleotídica na cadeia molde de DNA e o posicionamento de cada aminoácido na sequência polipeptídica normal; (a) sequência normal e (b) mutação pontual na posição 5 do polipéptido.

que existam, por vezes, substituições ao longo dos polipéptidos sem que se alterem as suas propriedades funcionais, originando formas de vida limitada, a menos que sejam adaptáveis às necessidades do ambiente (Vide Capítulo 4).

Resta saber se haverá alguma vantagem biológica para o excesso de tripletos codificadores dos 20 aminoácidos naturais. Pelo menos, é de admitir uma razão para que a estrutura química, as propriedades dos aminoácidos e os tripletos correspondam entre si. Assim, p.ex., os codões em que U é a letra intermédia seleccionam aminoácidos com cadeias não polares (fenilalanina, leucina, isoleucina, valina), enquanto os codões para os ácidos aspártico e glutâmico, ambos de carga negativa, são iniciados por GA.

O facto de vários tripletos possíveis para o mesmo aminoácido diferirem somente numa letra (como acontece para a terceira posição dos códigos de prolina, treonina, alanina) permite que, havendo mutações, elas não provoquem efeitos (*mutações silenciosas*) e, até, venham a originar proteínas com maior capacidade de sobrevivência futura.

Por outro lado, existem proteínas com aminoácidos que não estão previstos no código genético. É o que acontece, p.ex., com a hidroxiprolina e a hidroxili-

sina do colagénio, que resultam de modificações químicas, respectivamente da prolina e da lisina, após serem codificadas e posicionadas numa determinada sequência peptídica.

Não foi ainda possível isolar os fenómenos bioquímicos das suas origens genéticas. Afinal, é de crer que o funcionamento das múltiplas vias metabólicas nas espécies biológicas tenha sido submetido, desde sempre, à interacção com o meio ambiente, o qual estaria na origem de constantes adaptações selectivas, provavelmente essenciais para a sua sobrevivência.

O mecanismo daquelas variações constantes, que vem sendo analisado a nível molecular há quase duas décadas, abriu extraordinárias possibilidades à compreensão de muitos fenómenos vitais.

Grande parte do conhecimento alcançado apoiou-se em experiências com organismos de vida muito curta, como os vírus e as bactérias, o que não impediu que fossem demonstrados processos semelhantes a nível de animais superiores, incluindo o homem. Apesar do código genético ser universal, detectaram-se algumas diferenças quanto a pormenores relacionados com a diferenciação evolutiva das espécies.

Basta lembrar que, enquanto as bactérias e vírus possuem um único cromossoma, os organismos eucariotas têm uma ou mais dezenas. Acresce que a reprodução sexual dos eucariotas permite a recombinação dos genes dos progenitores, o que não sucede nos procariotas, que têm reprodução assexuada. Estas diferenças explicam a grande variabilidade da expressão genética nos descendentes de eucariotas, ainda que, no conjunto, a informação transmitida seja a mesma.

## Conclusões

O código genético é universal, sendo determinado por tripletos de mononucleótidos que integram a estrutura do DNA e do RNA<sub>m</sub>. As sequências das bases (púricas e/ou pirimídicas) de cada triplete codifica determinados aminoácidos. Cada triplete codificador é designado por codão. Para os vinte aminoácidos essenciais são conhecidos sessenta e um codões, o que significa que mais de um codão pode codificar o mesmo aminoácido (degenerescência do código). Os três tripletos não codificados sinalizam as terminações da tradução. A leitura do polímero de nucleótidos que serve de molde à sequência polipeptí-

dica faz-se da esquerda (extremidade 5', fosforilada) para a direita (extremidade 3', hidroxílica). Os dois primeiros mononucleótidos de cada codão são mais importantes para a determinação do respectivo aminoácido, sendo a terceira posição muito menos específica (o que justifica frequentemente a degenerescência do código). O código (no ácido nucleico que está a ser traduzido, codão a codão) é lido em sequência, não é virgulado nem sobreposto. A formação de cada polipéptido requer sinais iniciadores (AUG, para a formilmetionina e idêntico para a metionina nas posições intermédias) e sinais de finalização (UAA, UAG e UGA) do código.

## CAPÍTULO 7

### MECANISMOS DE REGULAÇÃO DA SÍNTESE PROTEICA

A existência de DNA nos procariotas e eucariotas representa um capital de informação genética armazenado nas respectivas células, a ser expresso quando necessário. Em qualquer dos organismos, e em particular nos eucariotas, a informação que permanece silenciosa tende a ser muito superior à evidenciada em dado momento. Nas células diferenciadas dos eucariotas, conforme o tipo assim varia a quantidade e proporção relativa entre as suas múltiplas proteínas. Por ex., uma enzima predominante nos hepatocitos poderá ser irrelevante no tecido cerebral, e vice-versa. Nos procariotas ou nos eucariotas determinados substratos metabólicos específicos podem estimular centenas de vezes a actividade catalítica de determinada enzima que, em repouso, seria insignificante. Conclui-se das anteriores considerações que a síntese proteica está sujeita a mecanismos reguladores de notável precisão, com acção particular no metabolismo e também na diferenciação celular.

Apesar destes fenómenos terem sido evidenciados a vários níveis da escala biológica, a maioria do que se conhece é fundamentado por estudos em culturas bacterianas, sobretudo de *E. coli*. Progressivamente, as conclusões daí resultantes têm sido extrapoladas, com naturais reservas, para os organismos superiores.

#### Regulação nos procariotas

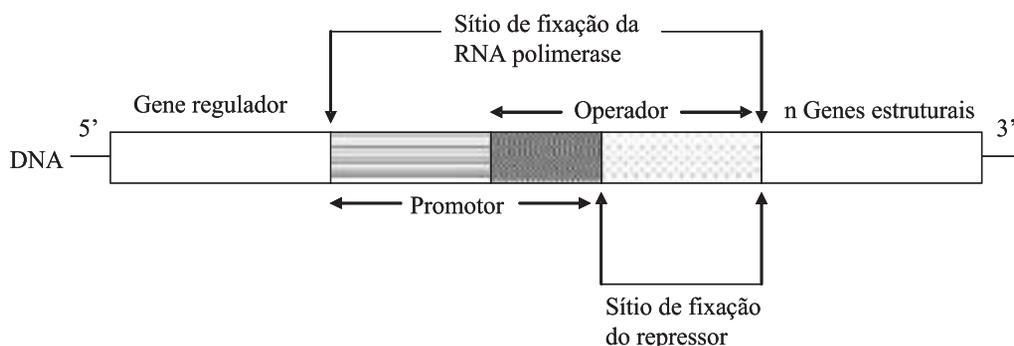
Nos procariotas, além dos mecanismos de *regulação da concentração proteica* (regulação negativa, regulação positiva da transcrição e regulação da tradução) também estão identificados processos de controlo e da *actividade enzimática* em que permanece inalterável a concentração proteica (regulação pós-tradução).

**(a) Regulação negativa da transcrição** – A regulação negativa da transcrição verifica-se em sistemas indutíveis e em sistemas repressíveis. Nas bactérias (em que a regulação da transcrição foi analisada com particular minúcia), verificou-se que a síntese de RNA<sub>m</sub> e a dos produtos finais proteicos que codifica era induzida ou reprimida por nutrientes específicos. Este controlo exerce-se essencialmente sobre os genes estruturais, ou que asseguram a informação genética para a síntese da maior parte das proteínas que intervêm no metabolismo, divisão e diferenciação celulares. A par daqueles existem genes estruturais que transcrevem proteínas constitutivas, isto é, cuja síntese se mantém constante e independente dos estímulos indutores ou repressores existentes. Um outro grupo de proteínas, sintetizadas em quantidades muito mais reduzidas, é constituído pelas *proteínas reguladoras* que, ao actuarem sobre o DNA em locais definitivos, modulam a transcrição genética, inibindo-a (controlo negativo) ou activando-a (controlo positivo). As moléculas reguladoras (activadoras ou repressoras) são codificadas por *genes reguladores*.

Os *repressores* actuam na sequência (ou na proximidade) onde se fixa a molécula de RNA polimerase, bloqueando-lhe fisicamente o acesso e, portanto, inibem a síntese de RNAm. A sequência de DNA onde se une o repressor é designada de *operador*; este situa-se próximo ou em sobreposição à sequência onde a polimerase se liga e começa a transcrição (*promotor*). Os genes estruturais, em conjunto com o respectivo promotor, definem uma unidade de acção coordenada designada *operão*.

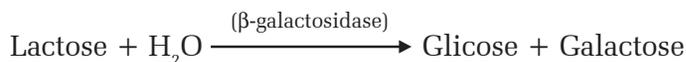
A capacidade de fixação das moléculas repressoras aos respectivo operadores é substancialmente afectada por outras moléculas de dimensões reduzidas a que se combinam (*efectores*). Existem dois principais tipos de efectores. Um dos tipos (designado *indutor*), ao unir-se ao repressor provoca a sua inactivação ou seja, dissocia-o do promotor, o que provoca um efeito indutor sobre a transcrição (Fig. 7.1 ); o outro tipo de efector (designado *co-repressor*) induz efeito contrário, ao transformar uma molécula apo-repressora inactiva num composto activo (constituído pelo *apo-repressor* a que se uniu o efector), que bloqueia a transcrição de DNA.

O segundo tipo de moléculas proteicas reguladoras – *activadores* – ao unirem-se ao operador aumentam a frequência de ligação da RNA polimerase ao promotor e, por consequência, aumentam a transcrição dos genes operadores. Este aspecto será analisado adiante, a propósito da regulação positiva da transcrição.



**Fig. 7.1.** Representação de um operão bacteriano (teoricamente similar ao dos eucariotas). Cada cistão (ou gene estrutural) codifica uma cadeia polipeptídica específica. Os mesmos promotor, operador e gene regulador adjacentes aos genes estruturais controlam a transcrição dos diversos cistrões. O operador define as sequências em que se fixa o repressor enquanto o promotor representa o sítio onde se fixa e é iniciada a transcrição pela RNA polimerase. As sequências do promotor e do operador podem sobrepor-se nalgumas extensões. Esta sobreposição justifica o controlo negativo que o repressor, ao unir-se ao operador, exerce sobre a iniciação da transcrição.

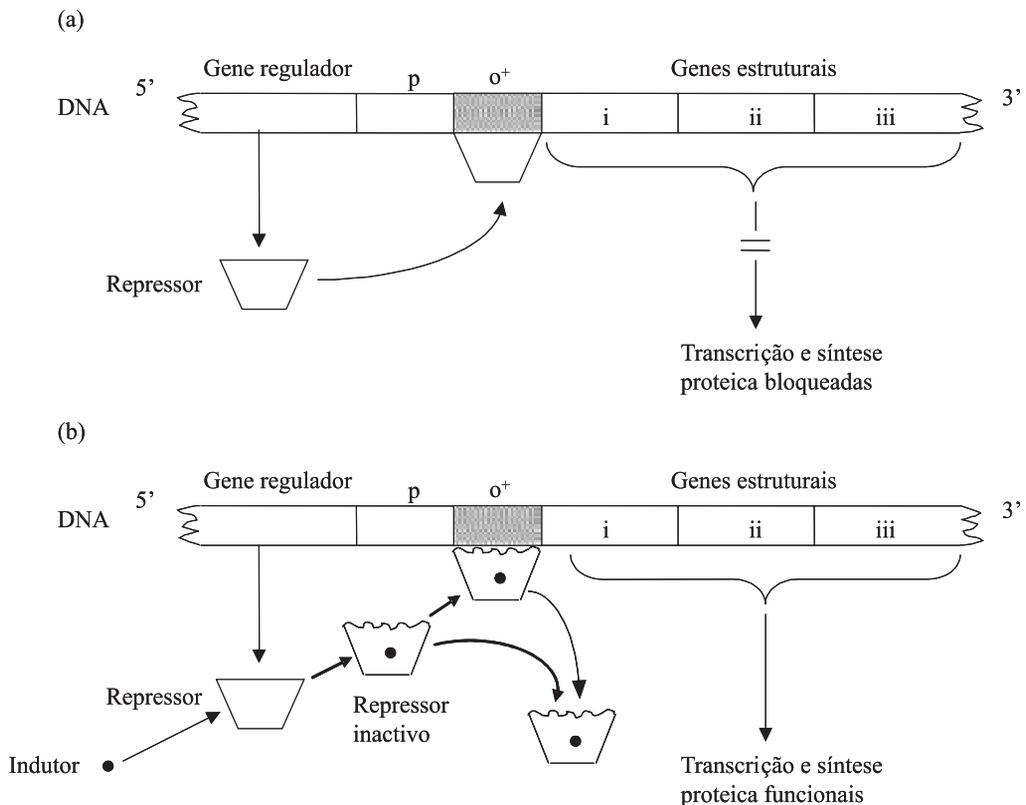
**Controlo em sistemas inductíveis** – Foi o primeiro mecanismo evidenciado, tendo por base os estudos pioneiros de Jacob e Monod em culturas de *E. coli*. Este tipo de controlo verifica-se em actividades catabólicas. Observações preliminares revelaram que a glicose e galactose eram as únicas fontes de carbono e hidrogénio utilizáveis pela *E. coli* no meio de crescimento. O tipo “selvagem” da *E. coli* não utiliza lactose se dispuser de um substrato mais simples, como a glicose. A utilização da lactose requer que a *E. coli* disponha de uma enzima, a  $\beta$ -galactosidase, que hidrolise a lactose em glicose e galactose:



Mudando a bactéria para um meio em que a lactose seja o único substrato com carbono verifica-se que, após um curto período sem resposta, a concentração da  $\beta$ -galactosidase aumenta rapidamente até níveis muito superiores aos basais. Na presença de lactose calcula-se que a  $\beta$ -galactosidase perfaça cerca de 3% das proteínas totais, ao que corresponde mais de  $3 \times 10^3$  cópias enzimáticas por célula. Nesta situação a síntese da  $\beta$ -galactosidase foi induzida pelo seu substracto, a lactose. Se no meio de cultura a glicose substituir a lactose assis-

te-se ao inverso, isto é, à diminuição rápida da concentração de  $\beta$ -galactosidase para níveis mínimos. Neste enquadramento, a  $\beta$ -galactosidase é uma *enzima indutível*, e a lactose é o estímulo *indutor* num processo de *indução enzimática* (Fig. 7.2.). A indução seria a cessação de um estado de repressão natural.

Enquanto na presença de lactose as células dispõem de abundante quantidade de RNA<sub>m</sub> específico para a  $\beta$ -galactosidase, num meio desprovido de lactose a quantidade daquele RNA<sub>m</sub> é irrelevante o que significa que, em condições naturais, a transcrição está bloqueada por um repressor.



**Fig. 7.2.** Representação de um operão indutível.

(a) O repressor, ao ligar-se ao operador (o<sup>+</sup>), bloqueia a transcrição dos genes estruturais, pelo que não há síntese proteica; (b) Na presença de um indutor com afinidade para o repressor este torna-se inativo, pelo que não se liga ou dissocia-se do operador. A RNA polimerase passa a poder ligar-se ao promotor (b) e a iniciar a transcrição dos genes estruturais, originando uma molécula de RNA policistronico, que codifica três tipos de polipéptidos.

Sucedem fenômenos idênticos com a quantidade de ribossomos existentes na *E. coli*. Se a bactéria crescer rapidamente, os ribossomos poderão constituir cerca de 20% da sua massa celular; em condições desfavoráveis, o ritmo de crescimento baixa e o seu número poderá reduzir-se até 5%, a par com a diminuição da síntese proteica.

Seguidamente, foi demonstrado que a utilização da lactose pelo *E. coli* requer três enzimas ( $\beta$ -galactosidase, galactósido permease e tiogalactósido transacetilase). A permease é a enzima membranar que activa a entrada da lactose na célula contra um gradiente de concentração, enquanto a transacetilase parece catalisar a transferência de um grupo acetilo (do acetil-CoA) para o respectivo substrato. Cada uma daquelas enzimas é modificada por um gene estrutural específico (**z** para a  $\beta$ -galactosidase, **y** para a permease e **a** para a transacetilase). Foram isoladas estirpes mutantes deficientes no gene para a  $\beta$ -galactosidase (*z'*), para a permease (*y'*) e para a transacetilase (*a'*).

A lactose é um indutor natural e específico da síntese daquelas três enzimas, embora não exclusivo. O isopropil-tiogalactósido (IPTG) tem idêntica capacidade indutora e, pelo facto de não ser metabolizado, possibilitou a identificação de mecanismos de regulação da síntese da  $\beta$ -galactosidase. Designadamente, foi observado que o mesmo indutor provoca a transcrição das três enzimas requeridas para a utilização da lactose, concluindo-se que a indução era *coordenada* (indução das três enzimas) pela mesma molécula indutora, *específica* (efeito restrito à lactose) e *gratuita* (por ocorrer quer com a lactose quer com um indutor análogo, como o IPTG).

Foram identificados mutantes de *E. coli* que sintetizam a galactosidase (e outras também a permease) na ausência de moléculas indutoras, sendo por tal designadas *mutantes constitutivas* (*i*<sup>-</sup>, insensíveis ao indutor, a contrapôr às indutíveis, *i*<sup>+</sup>). Por sua vez as proteínas formadas sob concentração e velocidade constantes, independentemente dos seus substratos (*proteínas constitutivas*), são componentes estruturais fundamentais da célula.

Aparentemente, a transcrição dos genes estruturais dependeria de um mecanismo regulador que, quando inactivo, permitiria a síntese daquela enzima. A existência deste mecanismo foi confirmado com a identificação de uma proteína difusível transcrita de genes reguladores, que exercia uma acção repressora sobre os genes estruturais em referência mas não interferia na estrutura da enzima sintetizada. *Normalmente, os genes estruturais estariam reprimidos*, passando à forma activa na presença de um activador. Explicar-se-ia assim que na ausência (ou defeito, por mutação) do gene regulador, a transcrição dos genes estruturais

prosseguisse com ou sem um activador. Foi ainda verificado que o mesmo repressor impedia a transcrição dos três genes estruturais que promovem a metabolização da lactose, o que caracterizava o fenómeno da *repressão coordenada*. O repressor difundiria do seu local de síntese (os ribossomas) fixando-se num segmento de DNA, com cerca de 25 a 30 bases, algumas das quais antes e a maioria depois do sítio de iniciação da síntese pela RNA polimerase. Por conseguinte, a sequência ocupada pelo operador sobrepuja-se um pouco com a do promotor mas não com as dos genes estruturais e reguladores.

Em determinados mutantes, a galactosidase de *E. coli* poderia ser produzida a níveis muito superiores aos habituais, embora à custa da repressão de outras proteínas bacterianas, também necessárias.

Por fim, com o isolamento e purificação do repressor da síntese de galactosidase, foi confirmado que aquela molécula proteica (tetramérica) possuía dois sítios alostéricos para ligação específica, um para o DNA e outro para o indutor enzimático. Na presença do indutor o repressor era *inactivado*, não se fixando ao DNA, do que resultava o desbloqueamento da transcrição (e da tradução subsequente) do cistrão da galactosidase.

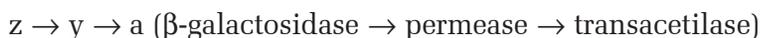
Foram isolados mutantes que possuíam repressores activos, mas que não bloqueavam a transcrição dos três genes estruturais que metabolizam a lactose. Na sequência, foi demonstrada a existência de um terceiro *locus*, o do *gene operador* (*o*), localizado em continuidade aos genes estruturais. Este gene, que não promovia a transcrição de qualquer produto difusível, seria o local para a ligação e acção do repressor. A substituição do gene normal ( $o^+$ ) por um gene constitutivo ( $o^c$ ), parcialmente insensível ao repressor, justificaria a continuidade da transcrição daqueles três genes estruturais, embora na presença de um repressor.

Num outro tipo de mutantes não havia transcrição dos três cistrões estruturais referidos, quer na presença quer na ausência de repressor e, também, sem haver alteração do operador. Desta observação resultou a identificação de um quarto gene (designado *promotor*, *p*) relacionado com o aproveitamento da lactose. O promotor define o segmento de DNA próximo a que se liga a RNA polimerase e onde é iniciada a transcrição. Nas situações em que o operador está alterado ou bloqueado pelo repressor, ou o promotor está modificado, a transcrição não se inicia, mesmo que os restantes genes estejam íntegros e potencialmente funcionais.

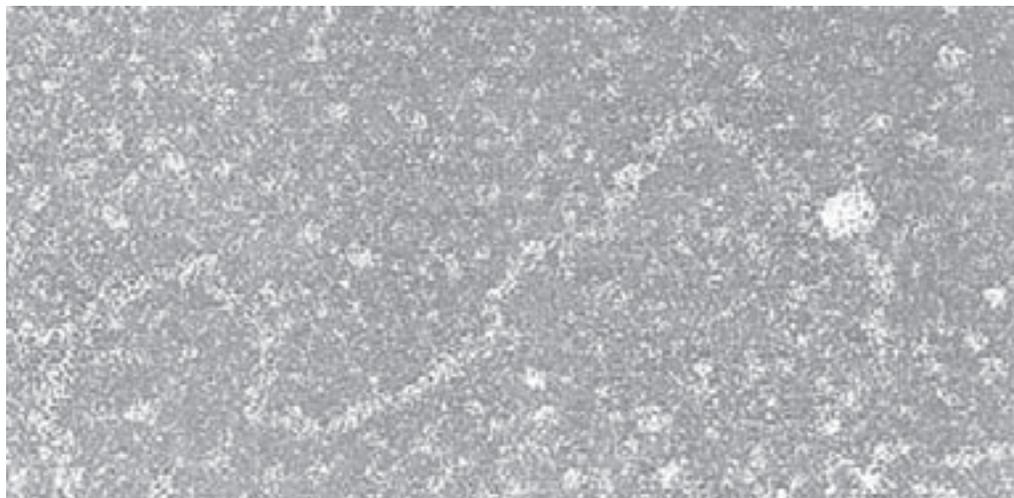
O conjunto de genes estruturais, regulador, operador e promotor (no cromossoma da *E. coli* estão próximos uns dos outros) sob regulação coordenada tem a designação de *operão*. A regulação coordenada dos genes de um determi-

nado operão significa que todos estão induzidos ou pelo contrário reprimidos. Porém, os segmentos de RNA<sub>m</sub> transcritos de cada gene podem ser traduzidos, com frequência, independentemente entre si.

O operão responsável pela metabolização da lactose, tem a designação de *lac operão* (Fig. 7.3.). Posteriormente foram identificados na *E. coli* outros operões (como o da galactose, *gal operão*), que diferem no número de genes estruturais e nem sempre estão dispostos em sequência ou próximos entre si, podendo estar dispersos pelo cromossoma. No caso do *lac operão* e de outros com disposição similar é transcrita uma única molécula de RNA<sub>m</sub> que contém todos os cistrões. O RNA<sub>m</sub> “policistrónico” é traduzido nos ribossomas na direcção habitual (5' → 3'), dando origem a uma sequência constante (*transcrição polarizada*) da cadeia polipeptídica codificada pelos genes estruturais:



É variável o número de vezes em que cada cistrão de *lac operão* (e outros) é sintetizado a partir da mesma molécula de RNA<sub>m</sub>. Está por esclarecer o mecanismo que regula aquela variabilidade.



**Fig. 7.3.** Molécula de *lac repressor* (esfera branca, à direita da micrografia) fixada num segmento de DNA ao promotor do *lac operão* (ampliação de 500.000 diâmetros; o fundo granuloso é efeito do sombreado da micrografia electrónica). Reprodução autorizada pelo autor, Dr. Jack Griffith (Cornell University, New York, USA).

Atendendo a que as ligações entre os repressores, indutores e co-repressores são simples e não covalentes, é possível regular rapidamente a formação do RNAm. A adaptação à mudança (p.ex., condições do meio) depende, sobretudo, da grande instabilidade metabólica da maioria das moléculas de RNA<sub>m</sub>. Com efeito a meia-vida de muitas das moléculas de RNA<sub>m</sub> de *E. coli* oscila pelos 2 minutos a 37°C. Por outro lado, o sistema de crescimento da célula utiliza as moléculas de RNA<sub>m</sub>. Estes dois motivos explicam que a maioria dos “moldes” de muitas proteínas da *E. coli* seja substituível em escassos minutos.

## Controlo em sistemas repressíveis

Outro processo de regulação relacionado mas de sentido contrário é a do operão repressível, o qual também exemplifica a facilidade e rapidez de adaptação da *E. coli* às condições de cultura, neste caso em situações anabólicas.

As bactérias sintetizam todos os aminoácidos essenciais que intervêm na síntese proteica, desde que o meio de cultura que lhes forneça azoto e carbono. Contudo, se o meio de cultura dispuser de aminoácidos pré-formados, a concentração das enzimas que intervêm na respectiva via da biossíntese diminui rapidamente para níveis irrelevantes, devido a um fenómeno de *repressão enzimática*. Um dos exemplos clássicos é o da biossíntese do triptofano. Em condições normais, se a *E. coli* não dispuser de uma fonte de aminoácidos, o operão de triptofano expressa a formação das enzimas que conduzem à sua síntese. Quando o triptofano é fornecido ao meio, o produto final da via (o triptofano) reprime a síntese das enzimas específicas da sua própria formação.

As enzimas que se adaptam negativamente ao estímulo *repressor* são, por isso, *catalisadores repressíveis*, e o processo é controlado por *repressão pelo produto final*.

Estudos subsequentes demonstraram que os aminoácidos actuariam somente como co-repressores (ao unir-se a uma outra molécula proteica, apo-repressora) conferindo ao complexo formado a capacidade de repressor. Nesta perspectiva o co-repressor é como que um indutor do repressor (Fig. 7.4). Enquanto no sistema inductível (p.ex., no lac operão) o repressor bloqueia o operador antes de ser iniciada a indução (pela lactose), no sistema repressível (p.ex., a do Trp operão) o repressor permanece inactivo enquanto não for activado pelo triptofano (produto final).

A propósito da regulação da biossíntese de aminoácidos pelo *E. coli*, foi sugerido que a função da molécula co-repressora não seria exercida pelo seu

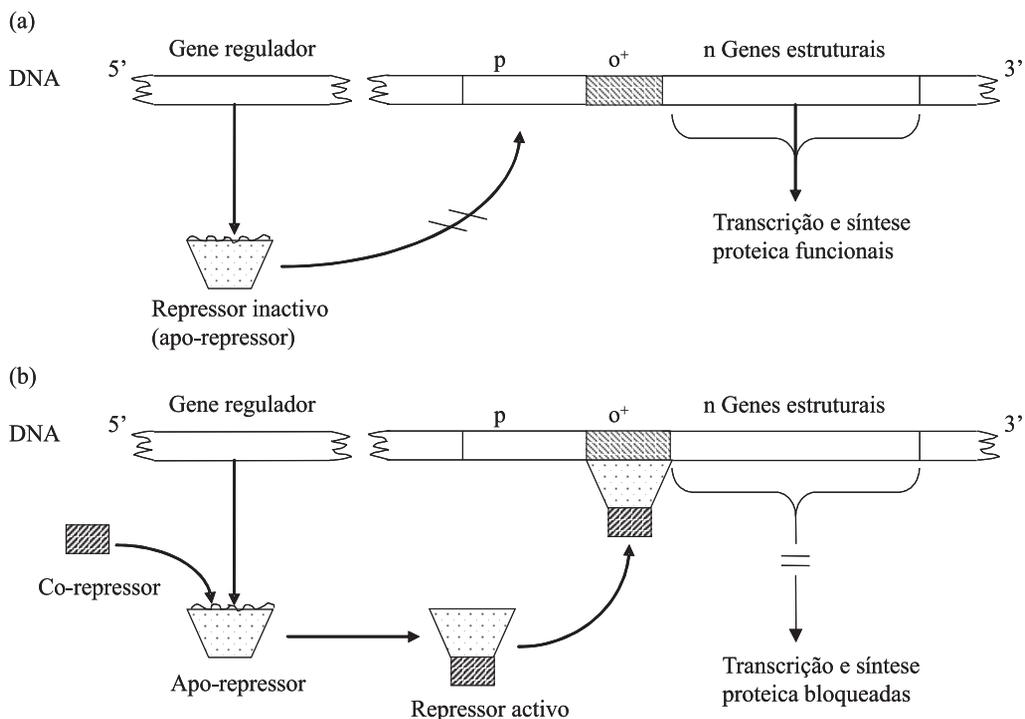


Fig. 7.4. Representação de um operon repressível. (a) A síntese proteica progride enquanto o repressor estiver inativo; (b) Quando o produto final (ou equivalente) da via atinge concentração em excesso une-se (como co-repressor) ao repressor, inativando-o.

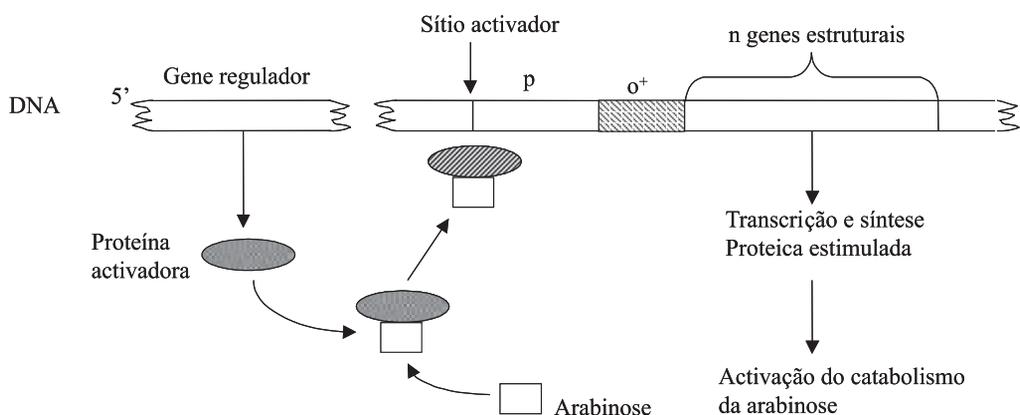
produto final (os aminoácidos) mas pelo precursor no processo, por exemplo uma molécula de aminoacil-RNA<sub>t</sub>.

Atendendo a que em diversos estudos de repressão enzimática estão envolvidas mais enzimas da via da biossíntese, admite-se que o mesmo produto final actue como repressor comum da biossíntese de várias enzimas, por repressão coordenada dos genes que codificam aquelas enzimas. A repressão coordenada pressupõe que a biossíntese também seja coordenada nos genes adjacentes ou próximos entre si.

Por conseguinte, enquanto a união (lábil, por ligações hidrogeniónicas e outras ligações fracas) do repressor (apo-repressor) ao co-repressor do sistema repressível transforma uma molécula inactiva num repressor *activo*, a ligação do repressor activo a um indutor conduz à formação de um complexo *inactivo* na repressão.

**(b) Regulação positiva da transcrição** – Alguns operões que codificam enzimas responsáveis pelo catabolismo de oses (excepto da glicose) podem ser reprimidos na presença da glicose. De facto, a glicose adicionada a meio de cultura de *E. coli* reprime a formação de muitas enzimas inductíveis (em que também se inclui a  $\beta$ -galactosidase) e constitutivas, mesmo que o meio contenha lactose. Aquele efeito parece resultar da acumulação de um produto (catabolito) da degradação da glicose, a que acresce a diminuição do conteúdo celular em AMP<sub>c</sub>. Contudo, quando a concentração daquele nucleótido cíclico aumenta, cessa o efeito repressor da glicose (*repressão por catabolito*). Este efeito resultaria da sensibilização do promotor *lac* por uma proteína desconhecida e/ou pelo AMP<sub>c</sub>, que estimulará a fixação da polimerase do RNA no início da respectiva transcrição (por regulação positiva).

Aparentemente, os operões *lac*, *gal* e também o de arabinose (*ara*) são regulados pelo repressor catabolito. A ligação de RNA polimerase ao promotor e o início da transcrição poderão ser estimuladas por proteínas activadoras. Na presença de arabinose e glicose, o *ara-operão* é regulado por uma proteína que, ao ligar-se a uma sequência nucleotídica próxima do promotor, activa a respectiva transcrição e, em consequência, a metabolização da arabinose (Fig. 7.5.)



**Fig. 7.5.** Controlo positivo da transcrição por acção de uma proteína activadora. Na presença da arabinose (que actua como efector), a proteína activadora fixa-se a uma sequência próxima do promotor, o que aumenta a afinidade do promotor para a RNA polimerase e activa a transcrição. Em consequência, é estimulada a síntese da enzima para o catabolismo da arabinose.

**(c) Regulação da tradução** – Diversas alterações da síntese proteica podem ser incluídas neste grupo. Por ex., a molécula de RNA<sub>m</sub> poderá não ser traduzida em toda a sua extensão por, a determinado momento e devido a mecanismos a identificar, se desviar dos ribossomas. Desta situação resulta a existência de mutações pontuais em determinado operão que poderiam afectar (por interrupção) a síntese de determinada(s) proteína(s) (devido à formação de codões sem sentido) e, também alterar a tradução global da respectiva RNA<sub>m</sub>, particularmente a dos cistrões seguintes à da mutação.

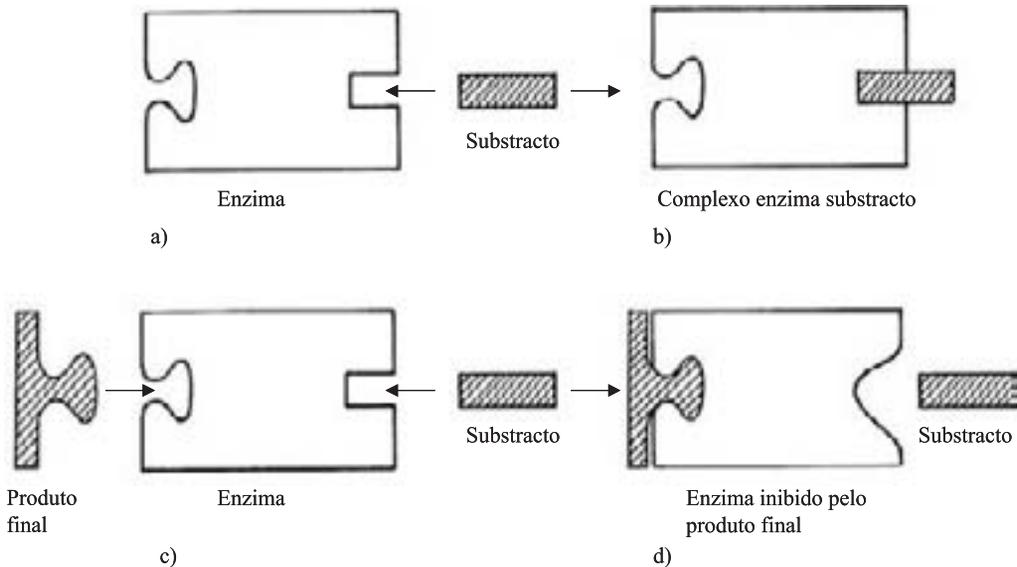
Adicionalmente, quer o início quer a terminação da tradução são pontos sensíveis, potencialmente afectados por desregulação.

**(d) Regulação da actividade catalítica (pós-tradução)** – Além dos mecanismos que afectam a quantidade de enzimas (ou de outras proteínas celulares) formadas, que demoram algum tempo para se efectivar, o controlo pode ser exercido muito mais rapidamente pela modificação da respectiva actividade. Neste processo, o produto final de determinada via metabólica presente em concentrações elevadas, ao combinar-se reversivelmente com a enzima inicial (no sítio alostérico) da biosíntese em causa, faz cessar toda a actividade desta. Sendo este mecanismo inibidor mediado por acumulação do produto final, o qual inibe a sua própria formação, é designado por *inibição retrógada* ou retro-inibição (Fig. 7.6.).

É o que sucede, p.ex., com o heme, que actuando *in vitro* ou *in vivo* em sistemas bacterianos ou animais (possivelmente também em humanos), inibe a actividade da sintetase do ácido  $\delta$ -aminolevulínico, que é a enzima inicial da biosíntese do heme. Sendo estruturalmente muito pouco semelhante ao substracto da enzima inibida, o produto final liga-se-lhe ao sítio alostérico, o que modifica a conformação da enzima e, portanto, a capacidade catalítica do seu centro activo. O produto final comporta-se como um *efector alostérico negativo*. Considerando que a ligação entre efector e enzima é estabelecida por forças fracas, assim que a sua concentração diminui a enzima liberta-se do efeito e altera a conformação primitiva o que, conseqüentemente, restaura a sua capacidade catalítica.

Foram isolados mutantes de *E. coli* que mantêm o centro activo normal embora possuam alterações no centro alostérico, de que resulta serem insensíveis a retrocontrolo inibidor, continuando a sintetizar produtos excedentes e desnecessários.

Quando a concentração do produto final é muito elevada, a inibição retrógada é reforçada pela repressão da transcrição da enzima reguladora.



**Fig. 7.6.** Inibição da actividade enzimática pelo produto final ou retro-inibição. Na biossíntese, a enzima reguladora (a) que está em geral localizada na etapa inicial da sequência, tem características alostéricas, isto é, além do sítio activo específico para o substrato; (b) apresenta um “outro” (alós) sítio para a fixação da molécula reguladora; (c) o retro controlo da sequência biossintética é em geral exercido pelo seu produto final; (d) ao fixar-se no sítio alostérico, o inibidor diminui a afinidade da enzima para o seu substrato, diminuindo ou fazendo cessar a sua actividade.

## Regulação nos eucariotas

Na generalidade, os mecanismos de controlo da síntese proteica nos eucariotas são semelhantes, embora muito mais complexos e lentos, do que nos procariotas.

Sem preocupações de exemplificação exaustiva, anotam-se as seguintes observações que poderão estar relacionadas com a especificidade daqueles mecanismos:

- (i) Ao contrário do procariotas, as células dos eucariotas têm núcleo, o que, entre outras particularidades, justifica que a transcrição e a tradução genética decorram em separado e intervaladas, enquanto nas procariotas são simultâneas.

- (ii) A maior parte do DNA dos organismos superiores existe em pares de cromossomas homólogos (diploidia), enquanto a dos procariotas existe num só cromossoma.
- (iii) Os genomas dos procariotas e eucariotas diferem em tamanho. Enquanto na *E. coli*, p.ex., existem cerca de 3.000 genes num único cromossoma ( $4 \times 10^6$  pares de bases), o humano possui mais de 1.000 vezes a quantidades de DNA ( $3 \times 10^9$  pares de bases por célula haploide).
- (iv) A maior parte dos genomas de cada tipo celular dos eucariotas permanece irreversivelmente “silencioso”, enquanto no dos procariotas a não transcrição de DNA é temporária e reversível. Esta diferença será justificada pela existência de proteínas (histonas) que formam complexos com o DNA nos núcleos de eucariotas; pelo contrário, a histona não existe nos procariotas.
- (v) A transcrição está sujeita a indutores e repressores da própria célula e de células de outros tecidos (p.ex., hormonas), por via dos receptores nucleares (ou citoplásmicos, de que emanam sinais para o núcleo) que afectam locais pré-determinados da cromatina. A transcrição resultante é referida a determinados genes, enquanto outros permanecem inactivos no mesmo instante.
- (vi) Os operões referentes a sequências metabólicas ou funções associadas não estão necessariamente contidos no mesmo cromossoma, nem em *loci* adjacentes. Esta dispersão coloca problemas quanto ao número de moléculas de RNA<sub>m</sub> transcritas e ao mecanismo de regulação coordenada. Uma das hipóteses mais consistente sugere a existência de sensores no genoma, a que estaria associada a activação coordenada dos diversos intervenientes na transcrição.
- (vii) Nos animais superiores, o fígado é o órgão privilegiado para a regulação da síntese proteica, por indução e repressão enzimáticas.
- (viii) A diferenciação nos organismos multicelulares pressupõe que os genes são activados ou desactivados em processos diferentes de evolução pré-programada no genoma, na dependência de efectores específicos e eventualmente, também por histonas (proteínas básicas que formam complexos com segmentos extensos do DNA nuclear).
- (ix) A metilação de algumas bases dos ácidos nucleicos poderá contribuir para o repressão (mantida e constante, de tipo para tipo celular) de segmentos específicos de DNA.

## Conclusões

Os organismos adaptam-se às condições externas e aos estímulos que geram, através de respostas que aumentam ou diminuem a sua actividade. A diferenciação celular e o metabolismo são exemplos de actividades dependentes de mecanismos reguladores de grande precisão baseados no controlo da síntese proteica, exercido a nível de enzimas indutíveis e/ou repressíveis. A indução enzimática constitui uma resposta adaptativa em que a quantidade de determinada(s) enzima(s) aumenta na presença de substrato(s) específico(s) ou por a carência do seu produto final; em sentido oposto, a repressão enzimática representa a diminuição da concentração da enzima, em resposta a um excesso do produto final da sua própria actividade. A indução e a repressão enzimáticas são determinadas por genes próprios: gene estrutural, gene regulador e gene operador. O gene estrutural codifica a molécula enzimática que responde à indução e à repressão. O gene regulador codifica a síntese da molécula repressora. O gene operador é o *locus* onde se fixa a molécula repressora e que, deste modo, impede a transcrição do gene estrutural.

A indução é activada quando a molécula repressora é removida do operador por um factor indutor (p.ex., aumento do substrato). Na inibição enzimática pelo produto final de uma sequência biossintética, a molécula repressora torna-se activa (impedindo aquela transcrição) somente quando se liga ao produto final da via, o qual participa como co-repressor.

A indução e a repressão enzimáticas podem afectar simultaneamente diversas enzimas, em actuação coordenada. Nestas circunstâncias, a indução coordenada e/ou repressão coordenada dependem de proteínas codificadas por genes associados, que constituem um operão. Cada operão tem um operador, também sensível a moléculas repressoras próprias e com actuação similar à da indução e repressão isoladas.

A replicação de DNA pode ser controlada por proteínas activadoras e pela repressão por catabolitos. Adicionalmente, a regulação pode ser exercida a nível de activadores catalíticos, sendo característico o processo de retro-inibição da enzima reguladora da via da biossíntese. O controlo da síntese proteica pode também ser exercida a nível da tradução.

Admite-se que virtualmente todas as células do mesmo organismo contenham o mesmo genoma, embora com expressão diferenciada devido às funções que estão definidas. Aquela delimitação funcional depende do controlo por indução/repressão da síntese proteica, em que a maior parte do genoma se encontra reprimido, com variação de tipo para tipo celular.

## CAPÍTULO 8

### ESPECULAÇÕES SOBRE A ORIGEM DA VIDA

#### Teoria Criacionista

A origem da vida terá sido uma das primeiras interrogações do *Homo Sapiens* que decerto, começou por atribuir a um acto sobrenatural.

As preocupações sobre a origem da vida foram posteriormente conceptualizadas nos textos da Génesis.

Na Antiga Grécia, Aristóteles admitia a possibilidade da vida resultar de geração espontânea. Embora as suas observações incidissem em larvas, Aristóteles admitia que um idêntico processo seria aplicável também ao homem. O conceito criacionista da origem da vida continuou a ser aceite durante os dois milénios seguintes, apoiado por teólogos medievais de nomeada, como Tomás de Aquino(século XIII) e, posteriormente, nos séculos XVI a XVIII, por vultos tão proeminentes como William Harvey e Isaac Newton. Pelo meio, na Antiguidade e na Idade Média, foram registadas superstições em que os animais teriam origem a partir de plantas e árvores, em que os insectos e larvas nasciam de carne putrefacta, as rãs da lama e os ratos do trigo pôdre, entre outras propostas imaginosas.

O primeiro a opor-se à teoria da formação espontânea foi Paracelso, nos séculos XV-XVI que, indignado, queimou os livros que a propunham.

Em 1688 o florentino Francesco Redi (1626-1697) deu o primeiro contributo experimental para a refutação da teoria da geração espontânea, ao demonstrar que as larvas que apareciam na carne resultavam de ovos depositados pelas moscas. Cerca de um século mais tarde, em 1765, o também italiano Lazzaro Spallanzani (1729-1799) igualmente refutou aquela doutrina, ao

evidenciar que, após ebulição prolongada, as soluções contidas em recipientes hermeticamente fechados não continham microorganismos. Todavia, aquelas provas continuavam a ser insuficientes para os defensores da geração espontânea, para quem (naquelas experiências) o “princípio da vida” seria destruído pelo calor.

Louis Pasteur (1822-1895) também demonstrou, em 1826, que as soluções esterilizadas poderiam ser contaminadas por microorganismos existentes no ar somente quando estivessem nos respectivos recipientes. Todavia, o próprio Pasteur não desaprovava a existência da geração espontânea, mas apenas a sua origem sob determinadas condições.

## **Teoria Organicista**

Além de advogar a geração espontânea, Aristóteles propôs que os organismos fossem entendidos como unidades funcionais no seu todo e não pelas características ou propriedades das suas partes. Este conceito unitário anti-reducionista, também conhecido por *Organicismo*, continuou defendido até à primeira metade do século XX por Herbert Spencer (1820-1903), Edward Stuart Russel (1887-1954) e Ludwig von Bertalanffy (1901-1972). De certo modo, o *organismo* seria mais do que a soma dos seus constituintes, sendo estes que, pela sua actuação, gerariam aquele “factor adicional” que, de certo modo, fazia assemelhar o Organicismo ao *Vitalismo*.

Até ao século XIX e começo do século XX foram propostas outras teorias sobre a origem da vida, em que se destacaram a *Mecanicista*, o *Vitalismo*, a *Panspermia*, e a *Cosmozoica*.

## **Teoria Mecanicista**

A doutrina mecanicista teve particular desenvolvimento a partir do século XVII, com o contributo determinante de René Descartes e também de William Harvey (1578-1657), Isaac Newton (1643-1727) e outros.

Porém, mais de dois mil anos antes, na Antiga Grécia, já alguns filósofos defendiam que toda a matéria existente na superfície terrestre e visível no firmamento era reductível a aglomerados de partículas. Para Demócrito, (c. 470

a.C. – c 380 a.C.), talvez o primeiro mecanicista conhecido e pioneiro da futura teoria atômica, essas partículas seriam infinitamente pequenas e indivisíveis (de que resultou o termo *átomo*). Os átomos teriam forma e propriedades físicas distintas, consoante estavam na constituição da água, da terra ou do fogo; podiam associar-se ou dissociar-se e, da sua associação, resultariam substâncias com propriedades diferentes; aquelas interações seriam dependentes somente das leis da Natureza e não de “deuses ou demónios”. Outros seguidores de Demócrito, contribuíram para o reforço das suas ideias, com destaque para o (também) filósofo Epicurus (c. 341 a.C. – c. 270 a.C.), que adoptou o atomismo como explicação do universo, e o médico Erasistratus (c. 304 a.C. – c. 250 a.C.), que se distinguiu nos estudos anatómicos e acreditava em que todas as forças corporais tinham origem mecânica.

Com Descartes o Mecanicismo tornou-se uma doutrina que, em sentido lato, era naturalmente reducionista: os princípios biológicos seriam idênticos às leis físicas e químicas que regem a matéria inerte, pelo que os organismos não difeririam das máquinas: o todo seria igual à soma das partes, de acordo com uma disposição adequada à execução de determinada função. Em sentido restrito, o Mecanicismo defendia que os seres vivos teriam de ser interpretados como máquinas para serem entendidos. A matéria viva resultaria da combinação de elementos inorgânicos em determinadas proporções e de acordo com as leis referidas.

Com Descartes, o ser vivo e a natureza no seu todo não eram mais do que máquinas desprovidas do que o Vitalismo propôs como sendo o princípio vital. Cada órgão funciona como uma roda dentada, de cujo funcionamento coordenado resulta o que se designa por *vida*. Igualmente, a compreensão do todo implica na sua decomposição e no estudo das peças mais elementares.

Além de defender que as actividades vitais são regidas pelas leis comuns da Física e da Química, o Mecanicismo fundamenta a origem da vida na matéria inorgânica. Este pressuposto contrariava a ideia de que as substâncias orgânicas provinham somente dos seres vivos ou, mesmo, que lhe eram essenciais. Por conseguinte, não havendo seres vivos não haveria substâncias orgânicas. O Mecanicismo ultrapassou aquele obstáculo admitindo a intervenção do acaso na combinação dos diversos elementos, de que teria resultado uma macromolécula com características vitais. Ao admitir o acaso a tese mecanicista tornou-se insusceptível de verificação experimental e, por consequência, irrelevante.

## Vitalismo e Panspermia

Etimologicamente, *Vitalismo* deriva do latim *vita* que significa “vida”. Para o Vitalismo e também para a *Panspermia*, a vida resultaria de um princípio “vital”, sobrenatural e distinto da matéria, que rege os fenômenos da vida. Para a Panspermia aquele princípio seria eterno e universal. Ambas as doutrinas, bem como o Organicismo, opunham-se a que a vida fosse reduzida às suas características físicas, químicas e também às psíquicas (quando se aplica).

As tentativas para isolar a “força vital” relacionada com os fenômenos da vida constituíram uma demanda persistente desde a Antiguidade. Os registros mais consistentes são atribuíveis a R. Fludd (1574-1637) e Jan Baptist van Helmont (1577-1644). Este último dedicou grande parte da sua vida a destilar o sangue arterial, na crença de nele encontrar o “espírito vital”, opondo-se por isso, tenazmente à prática comum da sangria como terapêutica. O seu filho, Franciscus Mercurius van Helmont (1614-1699) defendeu com eloquência aquela doutrina, conferindo-lhe um toque filosófico.

Porém, o Vitalismo, como ideia organizada, nasceu dos estudos do médico alemão Georg Ernest Stahl (1660-1734), também químico de renome que, entre 1704 e 1708, propôs que os compostos combustíveis eram ricos em “flogisto” (uma espécie de fluido com grande movimento, que se evidenciaria pela sensação de calor e luz proveniente dos corpos em combustão). Por idêntico motivo, as cinzas (que haviam perdido “flogisto”) não ardiam mas perdiam peso.

A teoria de Stahl, então também qualificada como *Animismo*, exerceu profunda influência na fisiologia da época. No seu entender, os organismos não eram sistemas mecânicos nem químicos, como outros propunham, por serem definidos por leis diferentes das que regiam o Universo inerte, e que, por tal, davam consistência ao *Vitalismo*.

A doutrina teve desenvolvimento na escola de Montpellier, em particular a partir da segunda metade do século XVIII com os trabalhos dos médicos Theophile de Bordeu (1722-1776) e Paul Joseph Barthez (1734-1806), que rejeitaram o Mecanicismo como explicação para o mistério da vida. Em alternativa, propunham uma “alma” que existiria em todos os sistemas com movimento espontâneo, em que cada órgão teria sensibilidade e ação próprias. Por conseguinte, tal como a combustão, que ocorria se houvesse a perda do princípio intrínseco (o “flogisto”), também a metafísica da vida seria explicável pelo princípio vital da “alma”.

Jons Jacob Berzelius (1779-1845), químico e médico sueco (e um dos fundadores da Química Moderna) propôs que os compostos químicos se subdividiam em orgânicos e inorgânicos, consoante eram produzidos ou não, respectivamente, pela matéria viva. As substâncias orgânicas requeriam, para serem formadas, o “princípio vital” que existiria somente na matéria viva. Por idênticas razões, as substâncias orgânicas não poderiam provir da matéria inorgânica sem a intervenção dos tecidos vivos. A interpretação de Berzelius era, de certo modo, semelhante à que Stahl propusera para os compostos combustíveis, e de que derivara o Vitalismo.

O postulado de Berzelius, que se opunha à síntese das substâncias orgânicas a partir das inorgânicas, foi contestado alguns anos mais tarde por um outro discípulo e colega alemão, Friedrich Wöhler (1800-1882), quando este sintetizou pela primeira vez, em 1828 uma substância orgânica (a ureia) a partir de cianeto de amónio aquecido. Na sequência daquela primeira contestação (ainda que indirecta) à teoria do Vitalismo, a existência de uma força vital intrínseca aos fenómenos biológicos (e na origem das substâncias orgânicas) foi progressivamente posta em causa por um conjunto de observações notáveis, verificadas até aos finais do século XIX, na generalidade protagonizadas por cientistas alemães, a maioria dos quais eram fisiologistas.

Na primeira dessas contribuições, Johannes Peter Müller (1801-1858), revelou (cerca de 1826) que o nervo óptico estimulado registava uma impressão luminosa, quer na presença quer na ausência de luz. Seguiu-se Karl Friedrich Wilhelm Ludwing (1816-1895) que, em 1847, demonstrou que a circulação e a pressão sanguíneas eram explicáveis através de forças mecânicas comuns. Nove anos mais tarde Ludwing foi pioneiro na extracção de órgãos de animais de experiência, mantendo-os vivos por perfusão de sangue bombeado do exterior. Cerca de 1843, o igualmente fisiologista Emil De Bois Reymond (1818-1896) estabeleceu os princípios da electrofisiologia, ao demonstrar que a excitação nervosa progride a par de uma alteração eléctrica ao longo do nervo, com velocidade muito semelhante à propagação de electricidade em fios de cobre e não, como defendia o Vitalismo, pela propagação de um fluxo etéreo vital.

No prosseguimento dos estudos de Willy Kühne (1837-1900) com “fermentos” pancreáticos e acção digestiva, de que isolou a tripsina, Eduard Büchner (1860-1917) demonstrou que aquelas proteínas catalíticas, definitivamente designadas por *enzimas*, também podiam actuar fora das células (em tubos de ensaio e, portanto na ausência de matéria viva) com resultados semelhantes

aos produzidos no organismo vivo. Porém, neste contexto, os vitalistas ainda defendiam que o facto da actividade desenvolvida *in vivo* decorrer muito mais rapidamente, à temperatura corporal e em condições muito mais drásticas do que *in vitro*, resultava de leis e princípios próprios. Haveria reacções inseparáveis da vida, como a transformação do açúcar em álcool, que, portanto, não poderiam ser produzidos num sistema inerte. Em 1896/7 Büchner conseguiu demonstrar o que os vitalistas afirmavam ser impossível: obter álcool a partir do filtrado acelular, de leveduras a que havia adicionado açúcar. Desde então, o Vitalismo perdeu muito dos seguidores e, os que eventualmente ainda existam, situam-se mais no campo do misticismo do que no da ciência.

### **Teoria Cosmozoica**

Esta teoria, defendida desde 1907 por Svante Arrhenius (1859-1927), admitia que a vida existiria desde sempre no Universo, migrando sob a forma de esporos de uns planetas para outros através do espaço, sob o impulso da luz solar. A ideia tornou-se rapidamente popular e o facto de ter sido admitida a existência da pressão da radiação solar dava-lhe alguma coerência. Alguns anos mais tarde, em 1910, a teoria tornou-se irrelevante ao ser verificado que a radiação ultravioleta (muito intensa na luz solar), além dos efeitos semelhantes a prever por parte de muitas outras radiações do espaço, entretanto identificadas, destruía rapidamente os esporos bacterianos.

### **Evolução Química e Biológica**

As teorias anteriormente referidas pouco contribuíram para esclarecer as questões fundamentais da origem da vida, pelo que o assunto ainda mantinha um acentuado pendor filosófico nos princípios do século XX.

Porém, progressivamente, com o contributo científico da Física, Química e Biologia, foi possibilitada uma melhor compreensão sobre a evolução da Terra e do Universo. O desenvolvimento das vôos espaciais exploratórias em anos mais recentes deu um forte impulso ao desenvolvimento de instrumentos e conceitos mais precisos sobre a origem do Cosmos, e sobre a constituição e evolução nos sistemas já explorados. Em consequência, as especulações filosó-

ficas deram lugar a elementos mais concretos, que culminaram na identificação do ácido desoxirribonucleico como a primeira substância identificadora da vida. Tornou-se claro que, sem DNA, não existe reprodução celular, nem síntese ou renovação proteica, nem perpetuação das espécies. Todavia, continua sem se conhecer bem os limites de fronteira que separam os seres vivos da matéria inerte. Também permanece em discussão a origem do próprio DNA e a dos precursores de que foi inicialmente formado.

É actualmente aceite, por consenso, que o Universo terá começado com o hidrogénio, por ser o elemento mais abundante, de que terão resultado novos elementos e compostos de peso molecular elevado. Admite-se que a Terra foi formada há  $4,6 \times 10^9$  anos, por condensação de gases e poeiras espaciais, a que se seguiu a formação de uma crosta rica em silicato de ferro e magnésio, a envolver um núcleo liquefeito em que abundavam o níquel e o ferro. O ferro da superfície terá contribuído para a retenção do hidrogénio, pelo que a atmosfera primitiva (atmosfera I) manteve características redutoras durante, talvez,  $1 \times 10^9$  anos, a par com a perda por difusão de néon e outros gases nobres. Além de ser muito redutora e não ter oxigénio livre, acredita-se que a atmosfera I contivesse gases hidrogenados, designadamente, metano, amoníaco, sulfato de hidrogénio e vapor de água.

Na sequência da fotodissociação da água pelos diversos tipos de energia radiante que incidiram na Terra primitiva, o hidrogénio, o amoníaco e o metano diminuíram progressivamente, a par com uma lenta acumulação de dióxido de carbono e azoto, a que cresceu uma camada de ozono envolvente. Esta nova atmosfera (tipo II), além de ser muito menos redutora, continha já algum oxigénio que, em conjunto com o azoto e o dióxido de carbono, terão marcado o início da *Evolução Química* da vida. O calor, descargas eléctricas e outras formas de energia radiante (como as radiações ultravioletas de baixo comprimento de onda) aparentam estar na base do fornecimento de energia que transformaram vários componentes inorgânicos (inicialmente dissolvidos nos mares que cobriam quase inteiramente a superfície terrestre) nos primeiros compostos orgânicos.

Através daqueles mecanismos favoráveis, ao longo de um período que se calcula em  $1 \times 10^9$  anos, terão sido formados, por *condensação abiótica*, os primeiros aminoácidos, purinas, pirimidinas, hidrocarbonetos, açúcares, precursores de polímeros e outras macromoléculas biológicas como os polissacáridos, polipéptidos, polinucleótidos e lípidos. Admite-se que no final daquele

período haveria condições para o aparecimento da primeira forma de vida que iniciou a fase de *Evolução Biológica*.

Tem sido sugerido que o período da evolução química se manteve ainda durante algum tempo, sobrepondo-se ao início da evolução biológica, enquanto as formas primitivas da vida (que terão sido anaeróbios heterotróficos) continuavam a extrair carbono e energia dos compostos orgânicos existentes. Assim que os compostos orgânicos foram esgotados, sobreviveram somente as células aptas a utilizar formas mais simples de carbono (como o  $\text{CO}_2$ ), as quais utilizavam o Sol como fonte energética e fixavam o azoto da atmosfera. Seguiram-se as primeiras células fotossintéticas que, em lugar de utilizarem a água como origem do hidrogénio, terão recorrido ao sulfito de hidrogénio. Os primeiros seres produtores de oxigénio terão sido as algas verde-azuis há cerca de  $3,1 \times 10^9$  anos, de que resultou uma aumento discreto do oxigénio na atmosfera que, de características inicialmente reductoras, foi evoluindo para as de oxidante.

Foram necessários entre 1,9 a  $1,7 \times 10^9$  anos para que a atmosfera contivesse oxigénio em quantidade suficiente para se desenvolverem as células eucariotas, seguidas pelos organismos multicelulares, em que se incluíam, há cerca de  $7 \times 10^8$  anos, os primeiros vertebrados aeróbios e as plantas superiores.

Os primeiros hominídeos do género *Australopithecus* terão surgido nos últimos dois a quatro milhões de anos, o *Homo erectus* terá vivido entre  $2,5 \times 10^8$  anos antes, e o *Homo sapiens* foi referenciado pela primeira vez há 3 a  $5 \times 10^4$  anos.

Os esquemas evolucionistas propostos por Charles Darwin (1809-1892) e Alfred Russell Wallace (1823-1913) têm sido confirmados, sobretudo através de sucessivos achados arqueológicos.

É muito provável que a escala crescente de complexidade física, química e biológica que foi proposta tenha progredido por sucessivas modificações dos estádios precedentes, em que cada etapa representaria uma pequena alteração probabilisticamente compatível.

A existência de bactérias muitos anos antes dos organismos superiores sugere que a essência da vida já existiria em seres unicelulares muito simples. Por exemplo, os aminoácidos e as bases púricas e pirimídicas ter-se-ão mantido como componentes celulares imutáveis (ou quase) desde o início e até aos nossos dias.

Todavia, para tal suceder, seria ainda indispensável que o ponto de partida pré-biótico fosse representado por um conjunto de compostos orgânicos parti-

cularmente abundantes no “mar primitivo”<sup>1</sup>. Seguidamente, as moléculas mais complexas terão adquirido capacidades funcionais, como as de catálise. Depois terão surgido condições para que algumas daquelas moléculas orgânicas pudessem auto-organizar-se em estruturas celulares e, por fim, foram concretizados os mecanismos de replicação celular.

Subsistem grandes dúvidas de que o esquema apresentado tivesse sido tão linear. Pelo contrário são admitidos múltiplos pontos de partida, a maioria dos quais foi descontinuado desde determinado momento. Virtualmente todos os tipos de organismos ou estruturas que já existiram estão extintas, tendo sobrevivido as linhas celulares com maior capacidade de adaptação às sucessivas mudanças da atmosfera e do “caldo pré-biótico”, e subsequente fase biótica.

Por conseguinte, os seres vivos terão resultados de moléculas anímicas, organizadas de modo mais ou menos complexo num todo funcional, em que cada componente adquiriu funções específicas para duas finalidades primordiais: *extracção de energia do ambiente e reprodução*.

Neste contexto, e no pressuposto de que a vida foi originada na Terra e não transportada de qualquer outro ponto do Universo, merecem particular destaque os valiosos contributos resultante das experiências que reproduziram em laboratório as condições (especulativas) da atmosfera e da crosta terrestre primitivas. Através desses estudos as teorias anteriormente propostas para a origem da vida foram acrescidas por novos argumentos e perspectivas. Designadamente, a continuidade implícita nos períodos da evolução química e biológica como que recuperou a ideia da geração espontânea mas com uma diferença fundamental: a de que a evolução ocorreria durante milhões de anos e muito lentamente, enquanto na proposta inicial seria quase imediata e da evolução rápida.

## **Teoria Materialista**

Desde os finais do século XIX e princípios do século XX, um notável conjunto de filósofos e cientistas como Friedrich Engels (1820-1895), Herbert Spencer (1820-1903), Ernst Haeckel (1834-1919) e Aleksandr Ivanovich Oparin (1894-1980) analisaram as teorias da geração espontânea e do Vitalismo numa pers-

---

<sup>1</sup> “Mar primitivo” e “caldo pré-biótico” são utilizados no texto em significado equivalente de termos

pectiva materialista. Além de se oporem aquelas teorias, entendiam que a vida resultara, em determinado momento, da evolução contínua da matéria, do mesmo modo que, pela combinação dos elementos e compostos mais simples, eram formadas substâncias complexas. Dessas sínteses, envolvendo substâncias mais elaboradas, eventualmente de natureza coloidal, terá emanado uma matéria com as propriedades especiais que definem a vida.

A teoria Materialista, ao contrário da Meticinista, refutava que os seres vivos dependiam exclusivamente de substâncias orgânicas: estas teriam sido formadas sem intervenção dos seres vivos, precedendo a sua existência e sem intervenção do acaso. Por influência das leis da Natureza, a vida teria surgido através de etapas contínuas de complexidade crescente a partir da matéria pré-existente, em que se evidenciariam os compostos orgânicos. Da evolução subsequente teriam resultado níveis de organização da matéria com propriedades e actividades ainda inexistentes nas etapas anteriores ou nos seus constituintes originais.

A concepção materialista da vida teve em Oparin (bioquímico russo) o seu principal mentor. Numa primeira comunicação que apresentou em 1922 à Sociedade Botânica em Moscovo, Oparin postulou que as substâncias orgânicas já existiam na Terra antes do aparecimento da vida, pelo que esta seria o resultado da evolução a partir da matéria orgânica. A matéria teria evoluído gradual e continuamente, passando pelas fases de substância inorgânica, orgânica, macromolecular, elementos pré-celulares, matéria viva primitiva e, finalmente, a das formas superiores da vida.

No início, eventualmente ainda no período em que as poeiras e gases antecederam a estabilização da crosta externa da Terra, teria havido a síntese espontânea, sob temperaturas elevadas, de compostos orgânicos a partir dos principais constituintes da atmosfera primitiva (metano, amoníaco, vapor de água e algum hidrogénio livre), por acção da energia radiante e de outros tipos energéticos; a partir dos compostos orgânicos mais simples, tais como aminoácidos e açúcares, teriam sido formados os primeiros polímeros, por mecanismos não-enzimáticos. A ausência do oxigénio na atmosfera I explicaria que os hidretos de azoto (o amoníaco,  $\text{NH}_3$ ) e os de carbono (metano,  $\text{CH}_4$ ) tivessem surgido antes dos óxidos de carbono ( $\text{CO}_2$ ) e de azoto ( $\text{NO}_2$ ).

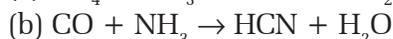
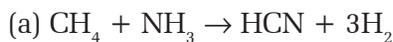
Também no pressuposto de que a atmosfera primitiva continha dióxido de carbono, amoníaco e vapor de água mas não tinha oxigénio, o geneticista e evolucionista inglês John B.S. Haldane (1892-1964) propôs, em 1928, que a

origem da vida teria sido antecedida por importante acumulação de detritos orgânicos no mar primitivo, em resultado da exposição daqueles constituintes da atmosfera à luz ultravioleta de baixo comprimento de onda.

A teoria de Oparin foi inicialmente contestada, primeiro por ser discutível a composição da atmosfera primitiva e, depois, por não haver quaisquer provas daquele tipo de evolução até ao aparecimento da vida na Terra.

Porém, alguns anos mais tarde, foi experimentalmente confirmado que os constituintes primitivos da atmosfera primitiva podiam gerar compostos orgânicos. Coube a Stanley Lloyd Miller, que em 1953 ainda era estudante graduado no laboratório de Harold Urey, na Universidade de Chicago, realizar a experiência, que se tornou clássica no género, utilizando um aparelho muito simples por si concebido. Nesse equipamento fazia passar vapor de água em ebulição por um balão que continha uma atmosfera de metano, amoníaco e hidrogénio, após o que a mistura passava por um condensador. Quando o vapor de água se misturava no balão com os restantes gases era accionada uma descarga eléctrica através de eléctrodos aí instalados, de que resultava uma faísca (reproduzindo a energia de um raio).

Depois de uma semana de experiências, sob descargas eléctricas sucessivas, foi detectada, na água do condensador, grande abundância e diversidade de compostos orgânicos hidrossolúveis, com destaque para trinta e quatro tipos de aminoácidos (dos quais somente dez existem actualmente), ácidos gordos simples, ureia e outras substâncias. Na fase gasosa foram detectados o monóxido de carbono, o dióxido de carbono e o azoto. As experiências de Miller tiveram o duplo mérito de evidenciar que a formação das moléculas orgânicas existentes nos tecidos vivos poderiam provir de misturas gasosas referidas à atmosfera I. Em segundo lugar, aquela síntese decorria com extraordinária rapidez (ao fim de uma semana) e precisão. Na opinião de Miller, as descargas eléctricas sobre a atmosfera primitiva geravam ácido cianídrico (HCN) através das seguintes reacções possíveis:



O ácido cianídrico, sujeito às condições estimuladoras da atmosfera primitiva, origina três produtos (cianamida, cianoacetileno e nitrilos) que são precursores reconhecidos de mais de cinco dezenas de compostos orgânicos (com destaque para os aminoácidos, purinas, pirimidinas, porfirinas, polipéptidos e outros polímeros).

Em 1961 o bioquímico hispânico Juan Oronin, demonstrou que partindo de uma mistura irradiada, rica em HCN ou formaldeído, obtinha grande diversidade de aminoácidos, alguns péptidos e também adenina e outros componentes dos ácidos nucleicos, como a ribose e a desoxirribose.

As experiências de Miller foram continuadas por Phillip Abelson e outros grupos de investigadores. Em 1956, Abelson, utilizando diferentes misturas de gases em combinações também distintas, verificou que, sendo a mistura redutora (contendo hidrogénio, azoto, monóxido de carbono e dióxido de carbono, mas sem metano nem amoníaco) e sujeita a energia radiante, eram obtidos compostos orgânicos, principalmente aminoácidos; pelo contrário, as misturas oxidantes não produziam compostos orgânicos. Estas experiências confirmaram, indirectamente, que a primitiva atmosfera teria de ser redutora e desprovida de oxigénio para se formarem compostos orgânicos elementares. Adicionalmente, em 1959, foi observado pelos cientistas alemães Wilhelm Groth e H. Von Weysenhoff que, além das descargas eléctricas, também a utilização de radiações ultravioletas naqueles meios experimentais origina aminoácidos, confirmando-se assim a proposta que Haldane havia sugerido mais de trinta anos antes.

Na verdade, muitos outros tipos de energia produziam substâncias orgânicas. Em sucessivas experiências promovidas por diferentes investigadores foram obtidas (na generalidade sem a intervenção de catalisadores e com composições variadas de misturas gasosas, irradiadas por determinada forma de energia) centenas de compostos orgânicos, entre os quais muitos dos constituintes celulares. Das formas de energia experimentadas destacam-se as radiações (luz visível, infravermelha, ultravioleta, raios X, raios gama), electricidade (descarga silenciosa e com ruído), energia corpuscular de alta energia (partículas alfa e beta), ultrassónica e hiperssónica. Os catalisadores mais utilizados em algumas séries de experiências foram o ferro, compostos de molibdénio e complexos de fosfato (p.ex., o ácido ortofosfórico ou os seus sais).

Melvin Calvin e Cols terão sido os primeiros investigadores a estudarem o efeito da energia ionizante na síntese abiótica dos compostos orgânicos. Para o efeito, fizeram incidir partículas emitidas por um ciclotrão numa solução com dióxido de carbono, obtendo o ácido fórmico e formaldeído. O facto de a radiação ionizante ser uma forma relativamente irrelevante da energia captada pela Terra justificou algumas críticas negativas por outros investigadores. Entre as diversas fontes energéticas que a Terra recebe do Espaço, as radiações ultravioletas e as descargas eléctricas estão entre as principais. As dificuldades

que apresenta a utilização controlada das radiações ultravioletas, ao contrário do que sucede com as descargas eléctricas, fez com que se privilegiassem estas fontes energéticas nas experiências de simulação da síntese pré-biótica.

O bioquímico cingalês naturalizado norte-americano, Ciryil Ponnampereuma, constatou em 1963, num aparelho por si concebido, que ao promover uma descarga eléctrica com que pretendia simular a acção de um raio, obtinha (em 24h), uma acumulação de hidrocarbonetos e de outras substâncias azotadas numa solução que reproduzia o “caldo pré-biótico” proposto nas experiências iniciais. No mesmo laboratório foi estudado o efeito das radiações ionizantes, com o feixe electrónico de um acelerador linear projectando sobre a atmosfera contida num tubo de vidro horizontal. Ao agir sobre a mistura de metano, amoníaco e água presentes no tubo, o feixe de electrões originou a formação de adenina. Experiências semelhantes, em que se fez incidir energia radiante ou térmica sobre um meio que simulava a atmosfera primitiva, evidenciaram a obtenção abiótica de nucleósidos (como a adenosina e a desoxiadenosina) e nucleótidos (a partir das pentoses e bases apropriadas e na presença de ésteres polifosfóricos). Além de AMP, também o ADP e o ATP foram obtidos naquelas condições experimentais.

Em 1965 Ponnampereuma observou a formação de dinucleótidos a partir de dois nucleótidos; a formação de oligómeros e polímeros orgânicos, (a partir de nucleótidos, aminoácidos, açúcares ou ácidos gordos) teria decorrido do mesmo modo, por formação de ligações covalentes e remoção duma molécula de água entre cada par de monómeros. Para que aquelas condensações pudessem suceder em meio aquoso, portanto em condições termodinamicamente desfavoráveis, haveria que as considerar em termos especiais, particularmente sob temperaturas bem superiores às da ebulição da água ou na presença de outras moléculas que removessem de imediato a água proveniente das ligações covalentes.

Os resultados apresentados em 1958 pelo bioquímico norte-americano Sidney Walter Fox, em que foi referida a formação de cadeias peptídicas a partir de misturas de aminoácidos aquecidos acima dos 100°C, sugeriam que a geração abiótica dos primeiros aminoácidos, bem como a dos seus polímeros, teriam decorrido em condições similares.

É admissível que os gases vulcânicos, ricos em  $\text{CH}_4$ ,  $\text{NH}_3$  e vapor de água, reunissem as condições indispensáveis à síntese dos primeiros aminoácidos e, de seguida, à sua polimerização no magna ou nas cinzas expelidas. Os polifosfatos, que se revelaram importantes indutores da polimerização da aminoácidos e nucleótidos em ambientes identificados com os da Terra primitiva, serão funcio-

nalmente idênticos ao ATP que é requerido para a formação das ligações covalentes entre as biomoléculas actualmente existentes. Por sua vez, as substâncias condensadoras nas sínteses abióticas seriam derivadas da carbodimida, composto formado por acção da energia radiante sobre o HCN. Nestas condições experimentais foi demonstrada a formação de péptidos a partir de aminoácidos, o que se pensa ter sucedido também na era pré-biótica na presença de polifosfatos, em locais de temperatura tão elevadas como na proximidade dos vulcões.

Idêntico mecanismo terá favorecido a formação dos polinucleótidos a partir da ribose, fosfato e bases púricas e pirimídicas, assim como a dos polissacáridos a partir da glicose.

### **A primeira “matéria” viva**

Atendendo a que, das centenas de compostos produzidos experimentalmente nas condições abióticas referidas, somente algumas dezenas fariam parte da primeira matéria viva, poder-se-á admitir que as moléculas incorporadas teriam estrutura e propriedades mais adequadas para a função biológica e, como tal, foram seleccionadas entre a maioria existente. Em alternativa, a sua utilização ter-se-á devido simplesmente ao acaso.

A primeira hipótese parece ser a mais correcta, considerando o exemplo da constituição das proteínas detectadas há de  $3,1 \times 10^9$  anos em fósseis bacterianos e nas bactérias actuais. Entre a mais de centena e meia de aminoácidos que terão existido nos primórdios da vida terrestre, somente os vinte tipos de aminoácidos então utilizados continuam a ser os precursores naturais das proteínas constituintes de todos os organismos conhecidos. Por conseguinte, a selecção preferencial de determinada molécula em desfavor de muitas outras que foram formadas terá resultados de múltiplas tentativas (ao acaso), até ser encontrada a conformação e a estrutura mais adequada.

Eventualmente, a fronteira entre a matéria inerte e a que adquiriu características vitais terá derivado do mesmo processo, por tentativas de adaptação a uma complexidade crescente.

Para Oparin, a formação das células primitivas (que designou por *protobiontes*) começou pela separação das substâncias coloidais do “caldo pré-biótico”, o que ocorreria sob a forma de gotas microscópicas (“*coacervatos*”). Aquele processo de separação do coloide (denominado *coacervação*) fora identificado

anteriormente em soluções com elevada concentração de polímeros hidratados em que, por acção espontânea, a solução inicial se separava em duas fases aquosas, uma com elevada concentração de polímeros e outra muito mais diluída. Oparin admitiu ainda que, ao serem formadas, as gotas microscópicas poderiam incorporar biomoléculas de pequenas dimensões, tais como a glicose ou aminoácidos, pelo que a concentração e a estrutura dos coacervatos seria potencialmente diversificada, dependendo do meio de que proviessem.

A formação e diversidade da constituição das gotículas sugeria a existência de *membranas* delimitativas do meio.

A diversidade das gotículas seria essencial para a sua selecção natural entre as estruturas pré-bióticas; as mais estáveis e equilibradas teriam melhores condições de sobrevivência e de desenvolvimento, em natural dependência do ambiente. Admitindo que os coacervatos teriam catalisadores, aqueles que melhor coordenassem as reacções exergóticas e endergóticas internas a favor dos mecanismos de síntese tenderiam a adaptar-se e a desenvolver-se. Ou seja, era admitido já um esboço do *metabolismo unidireccional*. Porém, ao aumentarem de volume, acentuar-se-ia o risco da sua fragmentação em gotas de menores dimensões, das quais algumas (ou todas) que tivessem catalisadores poderiam dar continuidade ao processo, no sentido da adaptação ao ambiente e crescimento.

Eventualmente, aquele mecanismo de fragmentação terá evoluído para a *divisão celular*. Oparin e colaboradores conseguiram obter uma grande variedade de gotículas com volume, concentração e composição química muito diversas. Observaram ainda desigualdades na distribuição de determinados componentes (p.ex., ácido nucleico) dentro da mesma gotícula, enquanto outras moléculas simples (p.ex., açúcares e mononucleótidos) estariam igualmente distribuídos nas duas fases da suspensão.

As gotas do coacervato, possivelmente formado no “caldo pré-biótico”, terão sido o ponto de partida para o desenvolvimento da primeira célula. A natureza coloidal das microvesículas impediria a dispersão dos constituintes, favorecendo a sua aglutinação e interacção reactiva. Oparin conseguiu demonstrar que a velocidade de determinada reacção química poderia ser diferente quando decorria dentro das gotículas ou no meio externo. Adicionalmente (à semelhança do que sucede nas células actuais, cuja sobrevivência e actividade dependem das trocas constantes da matéria e energia com o meio externo), também existia permuta de substâncias entre os coacervatos e o meio de solução. A experimentação foi alargada a reacções bioquímicas simples. Por exemplo, quando micromoléculas enri-

quecidas com a fosforilase do glicogénio eram adicionadas a uma solução de glicose-1-fosfato, constatava-se a síntese e a acumulação intravesicular de glicogénio.

Muito recentemente, em 1969, Fox propôs uma outra hipótese para as estruturas celulares que seriam geradas espontaneamente durante o *arrefecimento* lento das soluções de proteínoides formados a elevadas temperaturas. Por terem a forma em gota esférica, de dimensões razoavelmente uniformes (em média cerca de  $2,0\mu$ ), aquelas estruturas foram designadas de *microesferas*. Nestas formações foi detectada, em pH estável, uma camada dupla de revestimento externo semelhante às membranas biológicas, embora não contivessem lípidos. À excepção deste pormenor, as microesferas revelam muitas das propriedades das células biológicas, designadamente comportamento osmótico (em solução hipertónica mirravam, enquanto em meio hipotónico inchavam, pressupondo a existência de uma membrana semi permeável), capacidade de regulação da velocidade das reacções químicas internas e divisão em duas partes sob determinado estímulo químico (sendo visível uma divisória entre as duas metades que, nalguns casos, se separavam por completo). Em geral, era detectável no interior das microesferas um constante movimento “browniano” das partículas constituintes.

À semelhança das leveduras, as microesferas em solução proteinoide concentrada formavam protuberâncias dispersas no seu contorno, as quais, por vezes se destacavam formando como que microesferas-filha. Acresce que a microesfera revelava tendência para formar agregados (por vezes lineares) e capacidade para se auto-organizarem.

Em 1966, A.E. Smith comunicou que havia obtido microesferas semelhantes às de Fox; ao submetê-las a desidratação seguida por re-hidratação, obtivera uma estrutura mais volumosa (entre 10 e  $30\mu$  de diâmetro), semelhante à dos coacervatos. Experiências posteriores com metodologia semelhante confirmaram a formação de microesferas e apoiaram a hipótese de que os percursores das células mais primitivas, no decurso da evolução biológica teriam sido microestruturas com características semelhantes. Por conseguinte, as microesferas seriam *protocélulas*.

As observações anteriores de Oparin, Fox e outros sugerem que as primeiras proteínas continham capacidade informativa, pelo que o desenvolvimento das células primitivas estaria dependente daquelas moléculas, sem necessidade da existência ou intervenção dos ácidos nucleicos. De cada interacção química resultaria uma nova molécula com particularidades estruturais e reactivas que influenciavam, de modo mais ou menos acentuado, as reacções seguintes. Quando este tipo de acontecimentos se tornou restrito (p.ex., em microesferas), terá sucedido uma progressiva

diferenciação entre o meio interno e o externo. No meio interno haveria reacções com formação de produtos até então inexistentes ou que se encontravam demasiado dispersos no exterior. A fragmentação dessas estruturas instáveis dispersaria os constituintes, misturando-os com outros já existentes no meio. Ao serem formadas novas microvesículas, os seus componentes representariam a mistura molecular disponível no meio. Num determinado momento da evolução o conjunto de moléculas reactivas existentes nas microvesículas terá originado uma variação brusca no produto originado, com repercussões na estrutura e nas condições interna. Daquele tipo de modificações bruscas terá sido originado, em dado instante da Evolução Química, um organismo celular primitivo detentor das características dos seres vivos.

Em 1929 o geneticista norte-americano Herman Joseph Müller (1890-1967) propusera uma perspectiva diferente, ao sugerir que os ácidos nucleicos haviam sido formados no “caldo pré-biótico” *antes* das proteínas, organizando-se em um ou mais genes de que emanava a informação para a síntese proteica, metabolismo, auto-replicação e organização celular primitiva. Aquela hipótese genética, que apontava no sentido de os ácidos nucleicos terem sido, desde o princípio, as moléculas com o potencial requerido para gerar e desenvolver a matéria viva, veio a ser confirmada por Watson e Crick cerca de um quarto de século mais tarde.

## Conclusões

A Terra terá sido formada há cerca de 4,6 mil milhões de anos por condensação de poeiras e gases. Seguiu-se um lento arrefecimento da sua superfície, durante a qual foram formados os primeiros compostos orgânicos por efeito da energia radiante sobre a atmosfera primitiva, inicialmente redutora. Naquele período de evolução química terão sido sintetizadas centenas de compostos orgânicos, entre os quais algumas dezenas que continuam a ser constituintes comuns da matéria viva. Entre os compostos orgânicos que emergiram do “caldo pré-biótico” destacam-se os aminoácidos, bases púricas e pirimídicas, hidrocarbonetos, açúcares e ácidos gordos. O ácido cianídrico, formado a partir do metano e amoníaco e ou do monóxido de carbono, terá sido o principal precursor dos compostos azotados. Seguiu-se (na presença de agentes condensadores que promoveram a remoção de moléculas de água) a polimerização de algumas daquelas moléculas, de que resultou a síntese (não catalisada) dos polipéptidos, nucleótidos e polímeros de açúcares. A síntese destes compostos em condições abióticas (no período da evolução quími-

ca) foi confirmada em sucessivas experiências laboratoriais por diferentes grupos de investigadores, utilizando combinações gasosas que simulavam a composição da atmosfera primitiva sob tipos diversos de energia radiante. As primeiras estruturas celulares primitivas terão surgido há  $3,1 \times 10^9$  anos. Experimentalmente, foi evidenciada a possibilidade da formação de estruturas com algumas propriedades e comportamentos das células biológicas; essas estruturas foram obtidas por coacervação de soluções poliméricas coloidais e por arrefecimento lento da proteínoídes polimerizados. Em ambos os exemplos, as proteínas seriam o componente informativo que caracterizava a vida e de que teria resultado a constituição de funcionalidade das células biológicas. Outra hipótese propunha que a informação estaria nos ácidos nucleicos, organizado sob a forma de um ou mais genes. Esta hipótese veio a ser confirmada pela identificação da organização estrutural do ácido desoxirribonucleico em duplex e do mecanismo que promove a sua replicação e transcrição, através da molécula do ácido ribonucleico, em que da descodificação de cada cistrão resulta a formação de uma cadeia polipeptídica específica.

No decurso da evolução alguns dos tipos dos ácidos ribonucleicos (como o RNA<sub>r</sub> e o RNA<sub>p</sub>) terão intervindo como catalisadores enzimáticos no decurso da primitiva síntese proteica.

As características recentes dos organismos vivos serão o resultado de uma progressiva auto-organização e auto-regulação celulares, em grande parte dependente da informação genética contida no respectivo genoma.

Não parece ser provável, por exemplo, que uma via metabólica, como é conhecida actualmente, tenha existido assim desde o início, no “caldo pré-biótico”. As condições primitivas de que resultaram os primeiros compostos orgânicos e os precursores das actuais biomoléculas possibilitariam somente interacções moleculares químicas e energeticamente favoráveis. A evolução do material orgânico no sentido dos primeiros organismos vivos terá sido essencial para o desenvolvimento e aperfeiçoamento progressivo das vias metabólicas e, também, da restante constituição e funcionalidade biológicas. Parece óbvio admitir-se que a vida foi uma ocorrência inevitável ao longo de uma evolução gradual da matéria, tendo o seu aperfeiçoamento resultado de sucessivas adaptações e regulações simultâneas das suas características morfológicas, bioquímicas e fisiológicas.

Nesta perspectiva, não só a estrutura e a diferenciação das células e tecidos, por si, mas também cada uma das suas propriedades, expressa o resultado de muitos milhões de anos de evolução e de continuada adaptação ao meio e às condições externas que lhe são impostas.

## ANEXO I

### **CRONOLOGIA DAS PRINCIPAIS ETAPAS E INVESTIGADORES COM INTERVENÇÃO NO DESENVOLVIMENTO DA GENÉTICA E DA BIOLOGIA MOLECULAR, ATÉ 1973**

O mérito das “criações” científicas não dispensa nem deve ignorar os contributos de anteriores investigadores. São recordados, num breve resumo cronológico, em grandes períodos de influência conceptual, os autores que marcaram as etapas mais relevantes para o esclarecimento do DNA como portador e transmissor da memória genética de cada organismo vivo.

#### **Era Pré-Genética**

- 1859 – *Charles Robert Darwin* publicou “The Origin of Species”, no qual defendeu que a evolução se baseia num processo de selecção natural de cada espécie e indivíduo, que transmite as adaptações.
- 1865 – *Gregory Mendel* evidenciou os princípios da segregação equivalente e da distribuição independente dos genes, e definiu os conceitos de *dominância* e *recessividade*, relacionando-os com as características transmitidas em gerações sucessivas. Os traços estariam localizados (e seriam transmitidos) em unidades hereditárias, não identificadas.
- 1869 – *Friedrich Miescher* extraiu e purificou o conteúdo do núcleo celular (que designou por *nucleína*), o qual seria constituído por proteínas e por uma outra substância rica em fósforo. Aquela nucleoproteína, pelo seu elevado peso molecular, seria uma substância complexa a que Miescher atribuiu a potencialidade de transmissão das características genéticas.

- 1879 – *Albrech Kossel* verificou que a fracção não proteica da nucleoproteína continha bases púricas (adenina e guanina) e pirimídicas (citosina, timina e uracilo).
- 1880 – *Walter Flemming* descreveu a separação de cada cromatídeo para as células-filha durante a mitose.
- 1882 – *Eduard van Beneden* revelou que todas as espécies que estudara possuíam um número próprio de cromossomas.
- 1889 – *Richard Altmann* isolou da nucleoproteína (ou nucleína) a componente proteica e uma substância que denominou *ácido nucleico*.

## Era Genética

- 1902 – *Theodor Boveri* e *Walter Sutton*, trabalhando independentemente, identificaram os cromossomas como unidades da hereditariedade (teoria cromossômica da hereditariedade). *Walter Sutton* verificou que cada cromossoma se emparelhava na meiose com outro fisicamente semelhante; quando as células germinais eram produzidas, os cromossomas dividiam-se de modo a que cada gâmeta recebia somente um dos cromossomas de cada par homólogo, que seria portador de características genéticas específicas. O ovo resultante da conjugação das células funcionais passaria a ter o dobro dos cromossomas, sendo cada um dos cromossomas homólogos fornecidos pelas células progenitoras. Foi concluído por *Sutton* que a meiose celular (nas células germinais) fundamentaria a distribuição das características hereditárias mendelianas, com representação das suas possíveis combinações nos produtos da conjugação (representado pela fórmula  $2^n$ , sendo  $n$  o número de cromossomas), enquanto nas gâmetas haveria  $n$  cromossomas.
- *Archibald Garrod* estabeleceu pela primeira vez que uma doença (alcaptonúria) poderia ter causa genética (devido a uma rara mutação, recessiva), consagrando o termo por que ficaram conhecidos os *erros genéticos do metabolismo*. A doença tornava-se manifesta antes da sintomatologia articular (por artrite grave), através da excreção urinária de elevadas quantidades de homogentisato, que conferia a cor negra à urina. Considerando que aquela substância era um reconhecido produto do catabolismo da fenilalanina e da tirosina, *Garrod* propôs que o defei-

to genético derivava de uma deficiência enzimática, a nível da degradação do anel benzénico daqueles aminoácidos.

1905 – *William Bateson* propôs o termo *genética* e demonstrou que alguns genes não tinham distribuição independente, sendo transmitidos associadamente.

1908 – É derivada a lei de Hardy-Weilberg

1909 – *Wilhelm Ludwing Johanssen* cunhou o termo *gene* para determinados elementos da hereditariedade ainda mal esclarecidos.

1910 – A *Eugenia* adquire força social, sendo desencadeados movimentos racistas e pró-esterilização de determinados grupos étnicos, que se prolongaram pelas décadas seguintes.

1909 a 1930 – *Phoebus Aaron Theodor Levene* identificou a D-ribose como componente do ácido nucleico. Vinte anos mais tarde detectou outro tipo de açúcar, a 2-desoxi-D-ribose. Sugeriu que os ácidos nucleicos formavam complexos moleculares de grandes dimensões, constituídos por unidades (que designou por *nucleótidos*) de fosfato, açúcar, bases púricas e pirimídicas, e água. A simplicidade molecular, uniformidade, ubiquidade, e o facto de se organizar em polímeros semelhantes aos polissacáridos, desvalorizou, junto da comunidade científica, as potenciais funções biológicas dos ácidos nucleicos. A informação hereditária continuou a ser atribuída às proteínas cromossómicas, por serem diferentes entre indivíduos e espécies.

1910-1911 – *Thomas Hunt Morgan* demonstrou que os genes estavam localizados nos cromossomas, o que confirmaria a teoria genética da hereditariedade. Ao identificar os pares de cromossomas XY (no sexo masculino) e o XX (no sexo feminino) sugeriu que determinadas características transmitidas estariam associadas ao sexo. Posteriormente, em estudos sobre a cor dos olhos da mosca da fruta *Drosophila*, demonstrou como os genes poderiam ter mutações ou serem alterados. A distribuição das unidades hereditárias por cruzamento e recombinação dos cromossomas explicaria a segregação dos genes em quatro grupos associados, cada um dos quais representava os quatro cromossomas detectados nas células de *Drosophila*. Estas conclusões foram confirmadas seguidamente em células de milho e ervilhas, nas quais o número de grupos genéticos associados correspondia ao total de cromossomas de cada espécie (respectivamente, em número de dez ou sete). Adicionalmente,

o grupo de Morgan constatou que os genes estavam posicionados linearmente, ao longo dos cromossomas da *Drosophila*.

- 1913 – *Alfred Sturtevant* (colaborador de T.H. Morgan) apresentou o primeiro mapa genético de um cromossoma, em que os genes se dispunham em sequência linear.
- 1925 a 1927 – *Hermann Joseph Müller* demonstrou que os raios X induziam mutações em função da dose utilizada, de que estabeleceu os princípios da mutação pontual espontânea como resultado de alterações físico-químicas ultramicroscópicas. Aqueles resultados sugeriam também que os genes seriam a base da vida e da evolução.
- 1928 – *Fred Griffith* estabeleceu as bases da recombinação genética, ao demonstrar que ratos infectados com pneumococos vivos inofensivos (estirpe R, de *rough*, desprovida da cápsula externa) em conjunto com uma espécie inativada de pneumococos virulentos (estirpe S, de *smooth*, capsulada), morriam infectados pela estirpe anteriormente inofensiva (que entretanto adquirira as características da estirpe S), cuja virulência era depois transmitida nas culturas seguintes, mesmo na ausência da estirpe inactiva.
- 1931 – *Harriet Creighton* e *Barbara McClintock* demonstraram em estirpes de milho (*Zea mays*) que a recombinação genética resultava da permuta de segmentos cromossômicos. Embora os genes estivessem dispostos em sequência linear nos cromossomas, a sua posição (ordem) relativa poderia mudar. Estas observações constituíram a primeira prova microscópica da clivagem e reconstituição dos segmentos cromossômicos por transposição.
- 1939 – *William Thomas Astbury*, utilizando a cristalografia de raios X, calculou em 3,4 Å a distância entre os nucleótidos constituintes do DNA.
- 1940-1941 – *George Wells Beadle* e *Edward Lawrie Tatum*, tomando como base de estudo os resultados que o grupo de Morgan havia obtido com a côr dos olhos de *Drosophila*, procederam a enxerto do tecido ocular de larvas de mutantes (moscas com olhos vermelhos-claro) em larvas normais ou do tipo selvagem (moscas com olhos vermelhos-escuro). Nas larvas transplantadas o enxerto desenvolveu olhos suplementares de côr normal (vermelho-escuro), o que foi atribuído à existência de uma substância metabólica que favoreceria o desenvolvimento de pigmentos normais no tecido transplantado. Por outro lado, a côr clara dos

olhos dos mutantes foi atribuída a duas mutações enzimáticas que impediriam a formação do pigmento normal a partir do triptofano. Nas experiências desenvolvidas em fungo de pão (*Neurospora crassa*), aqueles investigadores verificaram que os mutantes que não se desenvolviam no meio habitual recuperavam o seu crescimento quando aquele suporte era enriquecido com aminoácidos. Igualmente, foi presumida a existência de um defeito no gene responsável pela produção de dada proteína, o qual seria normalizado pela adição de aminoácidos. Daqueles estudos foi proposta a hipótese de que cada “gene seria responsável por uma proteína”, o que estabeleceu uma relação potencial entre os genes e os processos bioquímicos.

1943 – *Erwin Schrödinger* especulou sobre o conceito de código genético, considerando-a um tipo de informação baseada em número limitado de unidades químicas que seriam transmitidas e expressas a partir de um molde inicial.

## Era do DNA

1944 – *Oswald Theodor Avery, Colin MacLeod e Maclyn McCarty*, tendo como base a descoberta de Griffith (que designaram por “princípio transformador”), demonstraram que a transformação bacteriana (em estudo com *Streptococcus pneumoniae*) seria provocada pelo DNA. Para aquela conclusão procederam a experiências muito simples, em que o DNA extraído de culturas da estirpe S, ao ser misturado num tubo de ensaio com células da estirpe R, induzia uma parte das bactérias características do tipo patogénico (S), que se mantinham em culturas sucessivas. A adição de proteases ao meio não provocava alterações naquelas características. Porém, na presença de quantidades mínimas de desoxirribonuclease, aquele princípio transformador deixava de se manifestar. Por conseguinte, a transmissão da hereditariedade dependeria do DNA e não de uma proteína, como até então fora intransigentemente defendido.

1946 – *Joshua Lederberg* demonstrou, à semelhança do que Beadle e Tatum haviam verificado, que os genes das bactérias estavam sujeitos a mutações dependentes da nutrição. Os genes poderiam ser transferidos de uma bactéria para outra, por contacto ou conjugação. Os trabalhos de

Lederberg contribuíram também para a popularização das bactérias como meio de estudo genético e bioquímico de grande simplicidade, reprodutibilidade de resultados e escassos requisitos técnicos.

- 1947 – *S. Furberg* sugeriu que o DNA, observado por difracção de raios X, teria estrutura helicoidal.
- 1949 – *Linus Carl Pauling*, utilizando a técnica de electroforese, separou a hemoglobina normal da hemoglobina S (responsável pela drepanocitose), demonstrando assim que as doenças genéticas poderiam ser interpretadas como processos de origem molecular.
- 1950 a 1953 – *Erwin Chargaff* verificou a existência de uma relação paritária entre o total de bases púricas e pirimídicas ( $A = T$  e  $G = C$ ) presentes em cada molécula de DNA de várias origens, ainda que as bases nucleotídicas não existissem em proporção constante nos ácidos nucleicos. Isto é, o total de  $A + T$  nem sempre seria igual o total de  $G + C$ , enquanto a relação  $A + T / G + C$  tenderia a variar no DNA de origens diferentes. As observações de Chargaff demonstraram que os ácidos nucleicos não eram os polímeros “monótonos” como se pensava até então.
- 1950 – *John Randall*, em colaboração com *Maurice Wilkins*, *Raymond Gosling* e *Alex Stokes*, iniciou a análise da fibra de DNA (extraída da cabeça de espermazoides) por metodologia de raios X; ao grupo juntou-se, alguns meses mais tarde, *Rosalind Franklin*, recém doutorada e perita em cristalografia de raios X.
- 1951 – *Maurice Wilkins* e *Rosalind Franklin* definiram a configuração helicoidal do DNA, utilizando métodos cristalográficos e de difracção dos raios X.
- *Linus Carl Pauling* utilizando a cristalografia de raios X em cristais de proteínas puras, propôs a estrutura  $\alpha$ -hélice como base da estrutura secundária das proteínas, a qual resultaria do enrolamento dos aminoácidos numa conformação helicoidal estabilizada por ligações hidrogeniônicas intra-cadeia. Este tipo de rearranjo molecular helicoidal veio a ser confirmado alguns anos mais tarde por *John Kendrew* e *Cols* (1957) na molécula de mioglobina, e por *Max Perutz* e *Cols* (1959) na estrutura terciária da molécula da hemoglobina.
  - *Francis Crick*, que trabalhava em estruturas proteicas e não tinha interesse no estudo do DNA, acolheu *James Dewey Watson* (recém-licenciado em Biologia) como seu estagiário. *Watson* entusiasmado por uma

- conferência de Wilkins a que assistira, mobilizou a atenção de Crick para a análise das potencialidades do DNA como molécula transmissora da informação hereditária, dando início a uma colaboração frutuosa.
- 1952 – *Alfred Hershey e Martha Chase* demonstraram em *Escherichia coli* com marcação radioactiva que o DNA de fagos (e não as suas proteínas) era a molécula mediadora da hereditariedade, o que confirmava os resultados de Avery e Cols em 1944. Nas infecções bacterianas por fagos DNA, somente a molécula do ácido nucleico viral penetrava no interior da bactéria-hospedeiro, ficando a parte proteica no exterior. Por consequência, seria o DNA e não as proteínas o portador da informação genética.
- *Frederick Sanger* determinou pela primeira vez a configuração geral e a estrutura química de uma proteína, ao sequenciar todos os aminoácidos da molécula de insulina. Para o conseguir, Sanger fraccionou a molécula em fragmentos com dois ou mais aminoácidos, os quais recolocava seguidamente na posição correcta.
- 1952 a 1954 – *Paul Zamecnik* e Cols demonstraram que a síntese proteica decorria numa partícula ribonucleoproteica (posteriormente designada por *ribossoma*) existente no citosol.
- 1953 – *James Watson e Francis Crick* apresentaram a solução para a estrutura química do DNA e respectiva participação na hereditariedade. A molécula de DNA seria constituída por duas cadeias de polinucleótidos complementares nas respectivas bases, orientadas em sentido opostos e dispostas entre si em duplex. A estabilidade do duplex era assegurada por ligações hidrogeniónicas entre as bases de cada cadeia, dispostas no interior do duplex, sendo o “esqueleto” externo constituído por unidades alternadas de fosfato e desoxirribose, unidas entre si por ligações fosfodiéster. O conceito estrutural de DNA em duplex possibilitou a replicação de DNA a partir de cada uma das cadeias. Os genes (estruturais) determinariam a síntese das proteínas específicas. Porém, ainda não estava esclarecido o processo de transferência da informação genética do DNA.
- 1955-1960 – *Harry Eagle* aperfeiçoou com sucesso o meio de cultura *in vitro* de células de mamífero, através da análise sistemática dos nutrientes requeridos para o seu crescimento. Aquele meio tornou-se o suporte básico utilizado (e divulgado com o nome do seu descobridor) nos estudos sobre o crescimento e diferenciação das células de mamíferos. Me-

recem também destaque os contributos que *Theodore T. Puck* conferiu à metodologia de culturas de células isoladas de mamíferos e o impulso pioneiro de *Renato Dulbecco* no desenvolvimento das técnicas quantitativas em virologia, estudo dos vírus DNA oncogénicos e virologia animal.

- 1956 – *Jo Hin Tjio* e *Albert Levan* definiram o número correcto de cromossomas humanos, no total de 46.
- *Arthur Koldberg* e *Colts* isolaram e purificaram a DNA polimerase, evidenciaram a participação do DNA como modelo na replicação e o mecanismo de acção conjunta de três enzimas (DNA polimerase, DNA ligase e endo-desoxirribonuclease) no processo da replicação.
- 1957-1959 – *John C. Kendrew*, *Howard M. Dintzis* e *G. Bodo*, utilizando cinco compostos cristalinos diferentes da mioglobina de baleia, obtiveram a primeira representação tri-dimensional daquela molécula, por difracção de raios X. Dois anos depois, *Max Perutz*, *Michael Rossmann*, *Ann Cullis*, *Hilary Muirhead* e *Tony North*, também utilizando a análise dos raios X, identificaram o mapa electrónico tri-dimensional da molécula da hemoglobina de cavalo, com resolução semelhante à obtida para a mioglobina.
- 1958 – *Matthew Meselson* e *Franklin Stahl* demonstraram a replicação semi-conservadora do DNA em culturas bacterianas. Uma das cadeias de DNA actua como modelo para a síntese de uma nova cadeia complementar; a replicação do novo DNA decorreria com o desenrolamento gradual do DNA original, seguida pelo re-enrolamento em hélix das cadeias híbridas.
- 1960 – *Samuel Weiss* promoveu a identificação e purificação da RNA polimerase
- 1961 – *François Jacob* e *Jacques Monod* propuseram um mecanismo de regulação genética (modelo do operão) pelo qual os produtos de determinado gene (regulador) controlariam (positivamente ou negativamente) a actividade de outros genes (estruturais).
- *Sidney Brenner*, *François Jacob* e *Matthew Meselson* confirmaram o RNA como intermediário na transmissão da informação genética, dos genes para a síntese proteica nos ribossomas.
  - *Marshall Nirenberg* e *Heinrich Matthaei* desenvolveram as primeiras tentativas de decifração experimental do código genético através da utilização de polímeros sintéticos de DNA contendo uma única base. Ficou esclarecido que o código genético se organizava em tripletos de

- mononucleótidos. A decifração completa do código genético foi completada em 1964.
- *Mary Lyon* reconheceu, pela primeira vez, o fenómeno da inactivação de um dos dois cromossomas X na mulher (identificado como heterocromatina durante o ciclo celular).
- 1964 – *Howard Temin*, a partir de estudos com vírus RNA, sugeriu a existência de excepções ao dogma central da biologia molecular.
- *Charles Yanofsky* e Cols evidenciaram que a sequência dos aminoácidos numa determinada proteína era co-linear com as mutações nos respectivos genes.
- 1965 – *David C. Phillip* e Cols obtiveram uma imagem de baixa resolução da estrutura da enzima lisozima, cristalizada da clara de ovo da galinha. Naquela determinação foi utilizado um método de cristalografia de raios X, semelhante ao que Kendrew e Perutz haviam utilizado para a determinação da estrutura espacial das moléculas de mioglobina e hemoglobina, respectivamente.
- 1968 – *R.N. Bryan*, *M. Sugira* e *M. Hayashi* descreveram pela primeira vez a transcrição completa do genoma do fago fX-174, pela RNA polimerase da *E. coli*, por hibridização DNA-RNA.
- 1970 – *Howard Temin* e *David Baltimore* demonstraram, independentemente, que os vírus tumorais constituídos por RNA sintetizavam o DNA tendo como modelo o RNA (transcrição reversa), o que seria catalisado pela enzima transcriptase reversa.
- *Hamilton O. Smith* e Cols isolaram da bactéria *Haemophilus influenzae* as endonucleases de restrição, as quais tinham a especificidade de sectionar a molécula de DNA em ligações adjacentes a determinadas sequências-tipo de nucleótidos.

## Era Genómica

- 1972 – *Paul Berg* e Cols sintetizaram, pela primeira vez em tubo de ensaio, uma molécula de DNA recombinante, por união de fragmentos diferentes de DNA.
- 1973 – *Stanley Cohen* e *Herbert Boyer*, liderando grupos independentes, obtiveram a primeira molécula de DNA clonado, a partir de genes de rã implantados em *E. coli*.



## ANEXO II

### GLOSSÁRIO ESSENCIAL

<sup>14</sup>Carbono (<sup>14</sup>C) – Isótopo radioactivo de carbono, com meia-vida de 5700 anos, emissor de radiação β fraca (electrões).

<sup>3</sup>H (trítio) – Isótopo radioactivo de hidrogénio, com semi-vida de 12,5 anos, emissor de radiação β fraca.

*Alelo* (abreviatura de *alelomorfo*) – Forma alternativa do mesmo gene.

*Aminoácidos* – Unidades estruturais das proteínas; existem 20 tipos (essenciais), que têm em comum um grupo carboxilo e um grupo amina (livres); diferem entre si na cadeia lateral, hidrofóbica ou hidrofílica (polar, sem carga eléctrica, ou com carga eléctrica, positiva ou negativa), e representam uma “letra” do alfabeto molecular nas sequências polipeptídicas.

*Aminoacil-RNAt sintetase* – Enzima que une cada aminoácido específico (activado em aminoacilo-AMP à custa do ATP) a uma molécula de RNAt (de que resulta a molécula acil-RNAt, pela libertação de AMP) que contém um anticodão específico para aquele aminoácido na sequência polipeptídica.

*Autólogo* – Referente ao mesmo indivíduo.

*Autossomas* – Cromossomas que constituem o património genético de cada indivíduo (excluindo os cromossomas sexuais). No ser humano totalizam 22 pares de cromossomas idênticos (em cada par).

*Autossómico* – Referente aos autossomas.

*Bacteriófago* (ou *fago*) – Vírus infectante de uma bactéria.

*Bases análogas*(ou *raras*) – Estruturas púricas ou pirimídicas que diferem das bases normais e que podem ser incorporadas nos ácidos nucleicos.

*Bases emparelhadas* (par de bases) – Emparelhamento (com formação de ligações hidrogeniónicas) entre as bases (púricas e pirimídicas) complementa-

res que constituem a estrutura dos nucleótidos nos ácidos nucleicos. O nucleótido com a base adenina (adenosina-monofosfato) emparelha com o da timina (desoxitimidina-monofosfato no ácido desoxirribonucleico, DNA) ou com o do uracilo (uridina-monofosfato no ácido ribonucleico, RNA), enquanto o da guanina (guanosina-monofosfato no RNA, ou desoxiguanosina-monofosfato no DNA) emparelha com o da citosina (citosina-monofosfato no RNA ou desoxicitidina, no DNA).

*Cariótipo* – Designação genérica do número, forma e dimensões dos cromossomas. Representa a “carta cromossômica celular” em que a forma e o número de cromossomas são características de cada espécie (cariótipo da espécie) e sexo (cariótipo masculino normal: 46, XY; cariótipo feminino normal: 46, XX). Pelo exame do cariótipo são detectáveis algumas anomalias, designadamente no total dos cromossomas (p.ex., a trissomia 21).

*Centrómero* – Região especializada e estrangida no centro de cada cromossoma onde se juntam (na metafase) os cromátídeos resultantes da sua divisão longitudinal (na mitose e na meiose), e onde se fixam os microtúbulos do fuso mitótico (ou acromático) que, na anafase, promove a separação de cada cromátídeo para polos opostos do fuso (onde serão localizadas as células-filha).

*Ciclo celular* – Sequência temporal das cinco principais fases que ocorrem numa célula eucariota: mitose (M), síntese de DNA (S), períodos intercalares ( $G_1$  e  $G_2$ ) e período de repouso ( $G_0$ ).

*Cistrão* (equivale a gene) – Segmento de DNA que codifica a síntese de uma cadeia polipeptídica (ou molécula de RNA transcrito).

*Codão* – Sequência de três bases nucleotídicas (triplete) na estrutura do ácido nucleico, que codifica determinado aminoácido.

*Codão de iniciação* – Triplete (AUG) posicionado no começo de uma sequência genética que codifica o início da cadeia polipeptídica.

*Código genético* – Correspondência entre os tripletos nucleotídicos e os aminoácidos que constituem cada polipéptido.

*Codões de terminação* – Codões (sequências TAA, TGA e TAG) localizadas na extremidade 3' de uma sequência genética, que codificam a terminação do polipéptido que foi sintetizado.

*Complementaridade* – União específica por pontes de hidrogénio entre pares de bases púricas e pirimídicas emparelhadas em segmentos paralelos dos ácidos nucleicos. A adenosina (base púrica) une-se a uma base pirimídica

(timina no DNA, ou uracilo no RNA), enquanto a guanina (púrica) se liga à citosina (pirimídica).

*Congénito* – Característica existente desde o nascimento.

*Constitutivos* – Genes ou proteínas presentes em quantidades fixas (sem regulação prévia e independentemente das condições do meio) geralmente em todas as células.

*Co-repressor* – Metabolito que, ao combinar-se com uma molécula repressora (apo-repressora), inibe especificamente a formação de uma determinada enzima participante na sua própria síntese (ou via metabólica)

*Cromátídeos* – Cópias idênticas duplicadas de cada cromossoma que se mantêm unidas pelo mesmo centrómero.

*Cromatina* – Mistura nucleoproteica constituinte de cada cromossoma, que adquire coloração própria quando sujeita a corantes específicos. A cromatina é composta por DNA (cerca de 30%), RNA (% insignificante) e proteínas (mais de 50% na maior parte são histonas). A proporção de cada componente varia de espécie para espécie.

*Cromossomas homólogos* – Que emparelham na meiose, têm morfologia semelhante e possuem genes que expressam as mesmas características.

*Crossing-over* (entrecruzamento cromossómico) – Mecanismo de permuta recíproca de segmentos entre cromátídeos homólogos que decorre na meiose (ver recombinação cromossómica).

*Degenerescência do código* – Substituição de uma das bases (em geral a localizada na terceira posição) num triplete que codifica o mesmo aminoácido que o triplete original.

*Delecção* – Perda de um segmento de cromossomas de DNA.

*Desoxirribonucleósido* – Composto que resulta da união de uma molécula de 2'-desoxirribose com uma base (púrica ou pirimídica).

*Desoxirribonucleótido* – Composto de três tipos de moléculas: uma base (púrica ou pirimídica) unida a uma desoxirribose, que é por sua vez esterificada a um grupo fosfato.

*Dímero* – Estrutura resultante da associação de duas subunidades idênticas.

*Diploide (2n)* – Número de cromossomas presentes nas células somáticas (total de 46 no ser humano), e que é o dobro do existente nas células germinais (haploides).

*DNA (ácido desoxirribonucleico)* – Polímero de desoxirribonucleótidos.

*DNA complementar (cDNA)* – Molécula de DNA sintetizada pela enzima transcriptase reversa a partir de molécula de RNA<sub>m</sub>.

*Dogma-central* (da biologia molecular) – Conceito em que a informação genética flui do DNA (que funciona como modelo estrutural) para o RNA, e deste para uma proteína (polipéptido).

*Dominante* – Gene (ou carácter genético) prevalente que se manifesta no fenótipo, apesar de existir apenas em um dos alelos do par de cromossomas homólogo.

*Endonuclease* – Enzima que hidrolisa as ligações internas da cadeia de DNA.

*Eucariota* – Organismo constituído por células nucleadas.

*Exonuclease* – Enzima que hidrolisa a cadeia de DNA a partir das suas extremidades.

*Factor  $\rho$*  – Proteína que intervém na terminação correcta da síntese de uma molécula de RNA.

*Factor  $\sigma$*  – Subunidade resultante da RNA polimerase que reconhece os sítios específicos do DNA onde é iniciada a síntese do RNA.

*Factor de iniciação* – Proteínas ( $F_1$ ,  $F_2$  e  $F_3$ ) que participam no início da síntese proteica.

*Factor de libertação* – Proteína específica que identifica os sinais de terminação da síntese proteica.

*Factores de transferência* – Proteínas (factores I e II) requeridas para a síntese das ligações peptídicas nos ribossomas: o factor I participa na união do aminoácido  $\text{RNA}_t$  ao sítio “A” do ribossoma onde decorre a síntese da cadeia polipeptídica; o factor II provoca a deslocação do peptidil  $\text{RNA}_t$  do sítio “A” para o “P” do ribossoma.

*Fenotipo* – Expressão das características de um gene ou do organismo, que resulta da interacção do genotipo com o meio ambiente.

*Fertilização* – Fusão dos gâmetas de sexos opostos de que resulta um ovo (haploide) fertilizado.

*Fuso mitótico (acromático)* – Conjunto de microtúbulos dispostos numa conformação geral fusiforme evidente na mitose, de que resulta a separação de cada cromátídeo do plano equatorial para os polos da célula-mãe.

$G_1$  – Fase do ciclo celular entre o fim da mitose e o início da síntese de DNA.

$G_2$  – Fase do ciclo que se segue à síntese de DNA e precede a divisão mitótica.

*Garfo de replicação (ou replicativo)* – Região (segmento de nucleótidos) do DNA onde começa o desenrolamento e a separação das cadeias parentais do duplex, para molde da síntese e alongamento de um par de novas cadeias complementares.

*Gene(ou cistrão)* – Segmento de DNA (sequência de desoxirribonucleóticos) que codifica a formação de uma cadeia polipeptídica (ou uma molécula de RNA).

*Genes reguladores* – Genes a que é atribuída a função (principal) de regular a síntese dos produtos dos outros genes.

*Genoma* – Sequência completa do DNA de um organismo.

*Genotipo* – Constituição genética de um organismo.

$G_0$  – Fase celular quiescente em que, apesar de metabolicamente activa, não há proliferação.

*Haploide (ia)* – Organismo ou célula que possui uma cópia de cada cromossoma; número de cromossomas existentes nos gâmetas (nos homens o total é de 23).

*Hélice* – Estrutura em espiral constituída por unidades que se repetem a distância certa em cada volta, em rotação ou translação.

*Hemizigoto(íco)* – Organismo, ou célula, que possui apenas uma cópia de determinado alelo dos progenitores.

*Heterocromatina* – Cromatina condensada e inactiva durante a transcrição.

*Heterocromossomas* – Cromossomas sexuais. No ser humano existe um par idêntico (X, no sexo feminino) ou diferente (X e Y, no sexo masculino) nas células somáticas. Em cada gâmeta existe um só tipo daqueles cromossomas (X na mulher; X ou Y no homem).

*Heteroduplex* – Molécula de DNA em que as duas faixas constituintes têm sequências de bases complementares.

*Heterólogo* – Substância ou estrutura com formas diferentes na mesma espécie, ou pertencentes a outra espécie.

*Heterozigoto (íco)* – Organismos ou célula cujos alelos, paterno e materno, são diferentes.

*Híbrido (DNA-RNA)* – Duplex constituído por uma cadeia de DNA unida a uma de RNA por ligações hidrogeniónicas entre as respectivas bases complementares.

*Histonas* – Proteínas ricas em aminoácidos básicos que se associam ao DNA nos cromossomas de eucariotas.

*Homólogo* – Substância ou estrutura com morfologia semelhante.

*Homozigoto(íco)* – Organismo ou célula em que os pares de alelos, paterno e materno, são idênticos.

*In vitro* (do latim, *em vidro*) – Experiências realizadas fora do organismo vivo, no meio celular ou em cultura de tecidos.

- In vivo* (do latim, *em vida*) – Experiências realizadas no organismo intacto.
- Informação genética* – Informação contida na sequência de bases nucleotídicas em cada molécula de DNA ou RNA.
- Ligase (do DNA)* – Enzima que une entre si segmentos separados de DNA, dando continuidade às respectivas cadeias.
- Locus* (plural-*loci*) – Localização de cada gene na estrutura cromossômica
- Mapa genético* – Disposição dos pontos de mutação num cromossoma, definidos por experiências de recombinação genética.
- Meiose* – Processo de divisão (em duas fases sucessivas) das células diploides em células germinais haploides.
- Mitose* – Divisão somática (celular e nuclear) de que resultam, de cada célula original, duas células-filha com o mesmo número (diploide) de cromossomas, idênticos entre si e aos da célula progenitora.
- Monómero* – Sub-unidade estrutural com características próprias que, através de reacções simples, origina compostos (polímeros) de moléculas iguais (ou diferentes) em sequência.
- Mutação* – Alteração hereditária do material genético de que resultam formas diferentes do gene afectado; varia entre a alteração numa das bases nucleotídicas de uma das cadeias do duplex até à perda de segmentos extensos de DNA.
- Mutagénico* – Factor (físico ou químico) que favorece a formação de mutações (p.ex., calor, substâncias alquilantes, radiações)
- Nucleases* – Enzimas que hidrolisam as ligações fosfodiéster 3', 5' (que unem as unidades de pentose de dois nucleótidos contíguos através do mesmo grupo fosfato) nas moléculas dos ácidos nucleicos (*desoxirribonucleases*, as que actuam no DNA; *ribonucleases*, as do RNA)
- Nucleótido* – Composto equinolar de uma pentose, um grupo fosfato e uma base (púrica ou pirimídica)
- Operador* – Gene contíguo aos genes estruturais, que serve de local de ligação e acção do repressor.
- Operão* – Conjunto dos genes estruturais, regulador, operador e promotor, com transcrição coordenada e sob regulação comum.
- Origem da replicação* – Segmento de DNA-mãe que determina o local de iniciação da replicação na fase S do ciclo celular. Nos eucariotas, ocorrem repetidamente, ao longo do duplex DNA parental, múltiplas sequências nucleotídicas para iniciação da cadeia.

*Plasmídeo* – Pequena molécula de DNA circular extracromossômica existente nas bactérias, com capacidade de replicação autónoma. Num sentido lato, o termo plasmídeo aplica-se a qualquer factor hereditário intracelular e extracromossômico, independentemente da sua identidade (quer seja viral, orgânico, autoreprodutivo, endossimbionte ou outra).

*Polimerase* – Enzima que catalisa a síntese (replicação e reparação) do DNA ou do RNA (transcrição) a partir de uma molécula isolada do DNA, ou que forma o DNA a partir de uma molécula de RNA (em transcritase reversa).

*Polímero* – Composto, geralmente de peso molecular elevado e de grandes dimensões, constituído por unidades (*monómeros*) repetidas regularmente e ligadas entre si por ligações covalentes.

*Polinucleótido* – Sequência linear extensa de nucleótidos unidos entre si por ligações fosfodiéster, pelas quais o carbono 3' da pentose (ribose ou desoxirribose) de um nucleótido se liga, através do grupo fosfato intercalado, à posição 5' da pentose do nucleótido contíguo.

*Polinucleótido ligase* – Enzima que une, por ligação covalente, os nucleótidos que constituem o polímero (DNA).

*Polipéptidos* – Polímero de aminoácidos unidos entre si por ligações peptídicas.

*Polirribossomas* – Conjunto de ribossomas associados à mesma molécula de RNA<sub>m</sub>, que participa na síntese de polipéptidos.

*Procariota* – Organismo celular de dimensões muito reduzidas e muito simples, desprovido de membrana nuclear e de organitos membranosos, que terá sido o primeiro tipo celular no decurso da Evolução Biológica.

*Promotor* – Segmento do DNA a que se fixa a RNA polimerase e que indica o local de início da transcrição.

*Quiasma* – Região em que os cromossomas homólogos parecem estar unidos entre si, originando estrutura em forma de cruz.

*Recessivo* – Alelo cuja expressão é minimizada pelo seu par dominante.

*Recombinação cromossômica* (ver *crossing-over*) – Rearranjo da sequência cromossômica (ou genes) resultante do *crossing-over* entre pares de cromossomas homólogos de progenitores diferentes, que origina uma nova combinação de marcadores genéticos nas células ou organismos-filho e respectivo fenotipo.

*Recombinação homóloga* – Recombinação entre segmentos de DNA que possuem sequências nucleotídicas homólogas.

*Relação A+T/G+C* – Proporção relativa entre o somatório das bases complementares (adenina–tiamina e guanina–citosina) numa molécula de DNA.

*Reparação* – Processo de excisão e substituição de segmentos lesados do DNA.

*Replicação* – Região (sequência de nucleótidos) do DNA em que tem origem a replicação.

*Replicação (do DNA ou dos cromossomas)* – Processo de síntese pelo qual uma molécula de DNA (ou um cromossoma) actua como molde para a sua própria duplicação; cada uma das cadeias do duplex replica uma cadeia complementar, de modo que cada nova molécula de DNA contém uma cadeia original e uma nova cadeia que lhe é complementar.

*Repressor* – Molécula reguladora de natureza proteica produzida por um gene regulador que bloqueia a transcrição, ao unir-se a um indutor (co-repressor) ou ao operador.

*Ribonucleósido* – Composto de uma base (púrica ou pirimídica) unida a uma pentose (ribose).

*Ribonucleótido* – Composto de três tipos de moléculas: uma base (púrica ou pirimídica) unida a uma pentose, por sua vez esterificada a um grupo fosfato.

*Ribossomas* – Partículas celulares de pequenas dimensões (<200Å de diâmetro), constituídas por moléculas de RNA e proteínas, que são o local da síntese proteica celular.

*RNA (ácido ribonucleico)* – Polímero de ribonucleótidos. Estão identificados três tipos principais: RNA mensageiro (RNA<sub>m</sub>), RNA ribossómico (RNA<sub>r</sub>) e RNA de transferência (RNA<sub>t</sub>). O RNA<sub>m</sub> transmite a informação genética do DNA para o local da síntese proteica (ribossomas), funcionando como modelo para a selecção e seriação dos aminoácidos que constituem cada tipo de polipéptido codificado. O RNA<sub>t</sub> funciona como adaptador entre aqueles aminoácidos e o respectivo codão na molécula de RNA<sub>m</sub>, durante a fase de tradução genética. O RNA<sub>r</sub> é um constituinte dos ribossomas com funções ainda por esclarecer.

*Sequência (de aminoácidos, polipeptídica ou proteica)* – Seriação linear de aminoácidos na estrutura primária de um polipéptido ou proteína.

*Sinapse (cromossómica)* – Emparelhamento de cromossomas (cromatídeos) homólogos durante a meiose.

*Somática* – Qualquer célula de um organismo, excepto as germinais (óvulos e espermatozóides).

*Supressor (gene)* – Gene que pode reverter o fenotipo resultante da mutação de um outro gene.

*Svendberg* – Unidade de sedimentação (S), cujo valor relaciona o peso da molécula sedimentada, por centrifugação, com a sua forma molecular.

*Telómeros* – As duas extremidades de cada cromossoma.

*Tradução* (genética) – Síntese de uma cadeia polipeptídica a partir da informação genética (sequência de bases) contida numa molécula de RNA<sub>m</sub>.

*Transição* – Síntese de uma cadeia linear de RNA a partir de uma molécula de DNA modelo.

*Transição* – Troca de bases do mesmo tipo (uma purina por outra ou uma pirimidina por outra).

*Transcritase reversa* – Polimerase que transcreve uma molécula de RNA<sub>m</sub> numa cadeia complementar de DNA.

*Transdução genética* – Transmissão de material genético de uma célula para outra, por infecção viral.

*Translocação* – Localização de um segmento cromossómico noutra cromossoma.

*Transversão* – Troca de bases de tipos diferentes (uma purina por uma pirimidina)

*Zigoto* – Ovo fertilizado (diploide) resultante da conjugação dos gametas (haploides) macho e fêmea.



## **BIBLIOGRAFIA**

### **Referências Gerais**

- DE BUSK, A.G. – “Molecular Genetics”, 2nd ed. Collier-Macmillan Canada Toronto, Ontario, 1970.
- FOX S.W. – “The Origins of Prebiological Systems and their Molecular Matrices”, Acad. Press, N.Y., 1965.
- INGRAM V.M. – “The Biosynthesis of Macromolecules”, W.A. Benjamin Inc., New York, 1965.
- KEOSIAN J. – “The Origin of Life”, 2<sup>nd</sup> ed, Reinhold Book Coop, New York, 1968.
- LEHNINGER A.L. – “Bioenergetics”. 2<sup>a</sup> ed, W.A. Benjamin Inc., Menlo Park Calif, 1971.
- MAALOE, O., KJELDGAARD N.O. – “Control of Macromolecular Biosynthesis”, W.A. Benjamin, New York, 1965.
- MARKERT C.L., URSPRUNG H. – “Developmental Genetics”, Prentice-Hall International Inc., New Jersey 1971.
- OPARIN A.I. – “The Origin of Life on Earth”, Macmillan, New York, 1938
- THOMPSON J.S., THOMPSON M.W. – “Genetics in Medicine”, W.B. Saunders Co. Phil., 1966.
- WATSON J.D. – “Molecular Biology of the Gene”, 2<sup>a</sup> ed W.A. Benjamin Inc., N.Y. 1970.
- WOESE C.R. – “The Genetic Code”, Harper and Row, New York, 1967.

## Artigos

### Capítulo 1 – Cromossomas, Genes e Hereditariedade

- BONNER J., DAHMUS M.E., FAMBROUGH D., HUANG R.-C., MARUSHIGE K., TUAN D.Y.H. – The biology of isolated chromatin. Chromosomes, biologically active in the test tube, provide a powerful tool for the study of gene action. *Science* 159:47-56, 1968.
- CAIRNS J. – The chromosome of *Escherichia coli*. *Cold Spring Harbor Symp. Quant. Biol.* 28:43-46, 1963.
- KORNBERG R.D. – Chromatin structure: a repeating unit of histones and DNA. *Science* 184:868-871, 1974.
- McKUSICK V.A. – The mapping of human chromosomes. *Sci. Amer.* 224:104-113, 1971.
- MESELSON M., WEIGLE J.J. – Chromosome breakage accompanying genetic recombination in bacteriophage. *Proc. Natl. Acad. Sci. (USA)* 47:857-868, 1961.
- SORIEUL S., EPHRUSSI B. – Karyological demonstration of hybridization cells in vitro. *Nature* 190:653-654, 1966.
- THOMAS C.A.Jr. – The genetic organization of chromosomes. *Annu. Rev. Genet.* 5:237-256, 1971.

### Capítulo 2 – Nucleótidos e Ácidos Nucleicos

- BRITTON R.J., KOHNE D.E. – Repeated sequences in DNA. *Science* 161:529-540, 1968.
- FELSENFELD G., MILES H.T. – Physical and chemical properties of nucleic acids. *Annu. Rev. Biochem.* 36:407-448, 1967.
- HEBY O., AGRELL I. – Observation of the affinity between polyamines and nucleic acids. *Hoppe-Seyler's Z. Physiol. Chem.* 352:29-38, 1971.
- HOTCHKISS R.D. – Toward a general theory of genetic reproduction in DNA. *Adv. Genetics* 16:325-348, 1971.
- KLUG A. – Rosalind Franklin and the discovery of the structure of DNA. *Nature* 219:808, 1968.
- THOMAS C.A. JR., MACHATTIE L.A. – Circular T2 DNA molecules. *Proc. Natl. Acad. Sci. (USA)* 52:1297-1301, 1964..
- WATSON, J.D., CRICK F.H.C. – Genetical implications of the structure of deoxyribonucleic acid. *Nature* 177:964-967, 1953

### Capítulo 3 – Conformação, Replicação, Reparação e Recombinação do Ácido Desoxirrinonucleico

- AUERBACH C., KILBEY B.J. – Mutation in eukaryotes. *Annu. Rev. Genet.* 5:163-218, 1971.
- BERGER H.J., HUANG R.C.C. – Purification and characterization of a deoxyribonucleic acid polymerase from rat liver. *J. Biol. Chem.* 246:7275-7283, 1971.
- BRIDGER B.A. – Mechanisms of radiation mutagenesis in cellular and subcellular systems. *Annu. Rev. Nucl. Sci.* 19:139-178, 1969.
- CLEAVER J.E. – Defective repair replication of DNA in xeroderma pigmentosum. *Nature* 218:652-656, 1968.
- HOLLIDAY R. – Induced mitotic crossing-over in relation to genetic replication in synchronously dividing cells of *Ustilago maydis*. *Genet Res.* 10:104-120, 1965.
- HUBERMAN J.A., RIGGS A.D. – On the mechanism of DNA replication in mammalian chromosomes. *J. Molec. Biol.* 32:327-341, 1968.
- HUBERMAN J.A., KORNBERG A., ALBERTS B.M. – Stimulation of T<sub>4</sub> bacteriophage DNA polymerase by the protein product of T<sub>4</sub> gene 32. *J. Molec. Biol.* 62:39-52, 1971.
- HYODO M., FLICKINGER R.A. – Replicon growth rates during DNA replication in developing frog-embryos. *Biochem. Biophys. Acta* 299:24-33, 1973.
- KORNBERG A. – Active center of DNA polymerase. *Science* 163:1410-1418, 1969.
- KORNBERG A. – The synthesis of DNA. *Sci. Amer.* 219:64-78, 1969
- LARK K.G. – Initiation and control of DNA synthesis. *Annu. Rev. Biochem.* 38:569-604, 1969.
- LIDERBERG J. – Virus, genes, and cells. *Bacteriol Rev.* 26:133-139, 1957.

### Capítulo 4 – Ácido Ribonucleico: Estrutura, Síntese e Transcrição Reversa

- BALTIMORE D. – Viral RNA-dependent DNA polymerase. *Nature* 226:1209-1211, 1970.
- COLVILL A.J. – Early in vitro and in vitro transcription of T<sub>4</sub> DNA. II The fidelity of in vitro transcription. *Biochem. Biophys. Acta* 209:106-111, 1970.
- DANIEL V., SARID S., BECHMANN J.S., LITTAUER U.Z. – In vitro transcription of a transfer RNA gene. *Proc. Natl. Acad. Sci. (USA)* 66:1260-1266, 1970.
- ERON L., BLOCK R. – Mechanism of initiation and repression of in vitro transcription of the lac operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. (USA)* 68:1828-1832, 1971.

- GREEN M.H. – Strand selective transcription of T<sub>4</sub> DNA in vitro. *Proc. Natl. Acad. Sci. (USA)* 52:1388-1395, 1964.
- HAYASHI M. – A DNA-RNA complex as an intermediate of in vitro genetic transcription. *Proc. Natl. Acad. Sci. (USA)* 54:1736-1743, 1965.
- HAYASHI M., HAYASHI M.N., SPIEGELMAN S. – Restriction of in vivo genetic transcription to one of the complementary strands of DNA. *Proc. Natl. Acad. Sci. (USA)* 50:664-672, 1963.
- HEINEMAN S.F., SPIEGELMAN W.G. – Control of transcription of the repressor gene in bacteriophage lambda. *Proc. Natl. Acad. Sci. (USA)* 67:1122-1129, 1970.
- HIDA Y., KAMEYAMA T., OSHIMA Y., HORIUCHI T. – Regulation of the lactose operon in vitro. I. Transcription of the lactose operon. *Mol. Molec. Genet.* 106:296-306, 1970.
- JONES K.W., TRUMAN D.E. – A hypothesis for deoxyribonucleic acid transcription and messenger ribonucleic acid synthesis in vivo. *Nature* 202:1264-1267, 1964.
- JUKES T.H. – Some recent advances in studies of the transcription of the genetic message. *Adv. Biol. Med. Phys.* 175:1-41, 1963.
- KAMIYAMA M., WANG T.Y. – Activated transcription from rat liver chromatin by non-histone protein. *Biochim. Biophys. Acta* 228:563-570, 1971.
- KIDSON C., KIRBY K.S. – Selective alteration of mammalian messenger-RNA-synthesis: evidence for differential action of hormones on gene transcription. *Nature* 203:599-603, 1964.
- KOSTRABA N.C., WANG T.Y. – Differential activation of transcription of chromatin by non-histone fractions. *Biochim. Biophys. Acta* 262:169-180, 1972.
- MESELSON M., STAHL F.W. – The replication of DNA. *Cold Spring Harb. Symp. Quant. Biol.* 23:9-12, 1958.
- MILANESI G., BRODY E.N., GEIDUSHEK E.P. – Sequence of the in vitro transcription of T<sub>4</sub> DNA. *Nature* 221:1014-1016, 1969.
- MOSES R.E., RICHARDSON C.C. – Replication and repair of DNA in cells of *Escherichia coli* treated with toluene. *Proc. Natl. Acad. Sci. (USA)* 67:674-681, 1970.
- ROSS J., SCOLNICK E.M., TODARO G.J., AARONSON S.A. – Separation of murine cellular and murine leukaemia virus DNA polymerases. *Nature New Biol.* 231:163-167, 1971.
- SIGAL NB., ALBERTS B. – Genetic recombination: the nature of a crossed strand – exchange between two homologous DNA molecules. *J. Molec. Biol.* 71:789-793, 1972.

- SKÖLE, O., ZETTERBERG A. – Studies on the aminoacid regulation of RNA synthesis in mammalian cells in tissue culture. *Expt. Cell. Res.* 55:289-294, 1969.
- SPIEGELMAN S., BURNY A., DAS M.R., KEYDAR J., SCHLOM J., TRÁVNÍČEK M., WATSON K. – Synthetic DNA-RNA hybrids and RNA-DNA duplexes as templates for the polymerases of the oncogenic RNA viruses. *Nature* 228:430-432, 1970.
- SPIEGELMAN S., PACE N.R., MILLS D.R., LEVISOHN R., EIKHOM T.S., TAYLOR M.M., PETERSON R.L., BISHOP H.L. – The mechanism of RNA replication. *Cold Spring Harbor Symposia in Quant. Biol.* 33:101-124, 1968.
- STADLER D.R. – The mechanism of intragenic recombination. *Annu. Rev. Genet.* 7:113-127, 1973.
- SUGINO A., HIROSE S., OKASAKI R. – RNA-linked nascent DNA fragments in *Escherichia coli*. *Proc. Natl. Acad. Sci. (USA)* 69:1863-1867, 1972.
- TAYLOR J-H., MYERS T.L., CUNNINGHAM H.L. – Programmed synthesis of deoxyribonucleic acid during the cell cycle. *In Vitro* 6:309-321, 1971.
- TEMIN H.M., MIZUTANI S. – RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226:1211-1213, 1970.
- TEVETHIA M.J., MANDEL M. – Transcription of the aligned chromosome of *Bacillus subtilis* W23. *Proc. Natl. Acad. Sci. (USA)* 58:1174-1181, 1967.
- TOMINAGA H., AKI J., NATORE Y. – Metabolic turnover of messenger ribonucleic acid in rat liver. *Biochem. Biophys. Acta.* 228:183-192, 1971.
- TONEGAWA S., HAYASHI M. – The direction of in vitro genetic transcription on circular DNA. *Biochim Biophys Acta* 123:634-637, 1966.
- TURKINGTON R.W., MAJUMDER G.C., KADOAMA N., MACINDOE J.H., FRANTZ W.L. – Hormonal regulation of gene expression in mammary cells. *Recent Progr. Horm. Res.* 29; 417-455, 1973.
- WEISS S.B. – Enzymatic incorporation of ribonucleoside triphosphates into the internucleotide linkages of ribonucleic acid. *Proc. Natl. Acad. Sci. (USA)* 46:1020-1030, 1960.
- WICKNER W., SHEKMAN R., GEIDER K., KORNBERG A – A new form of DNA polymerase III and a copolymerase replicate a long single-stranded primer-template. *Proc. Natl. Acad. Sci. (USA)* 70:1764-1767, 1973.
- YURA T., IMAI M., OKAMOTO T., HIRAGA S. – Transcription of the tryptophan operon of *Escherichia coli* in vitro. I. Detection and quantitative determination of specific RNA. *Biochim. Biophys. Acta* 169:494-510, 1968.

## Capítulo 5 – Tradução da Informação Genética

- BLOBEL G., POTTER V.R. – Studies on free and membrane bound ribosomes in rat liver. *J. Molec. Biol.* 26:293-301, 1967.
- BORST P., GRIVELL L.A. – Mitochondrial ribosomes. *Febs Letters* 13:73-88, 1971.
- CAMPBELL P.N. – Biosynthesis of proteins. *Progress in Biophys. Mol. Biol.* 15:1-38, 1965.
- ELSON D. – Ribosomal enzymes In: “Enzyme Cytology”, D.B. Roodyn (ed), Academic Press, New York 1967, pp. 407-472.
- FLESSEL C.P. – Cell-free synthesis of ribosomal proteins on polyribosomes. *J. Mol. Biol.* 58:579-594, 1971.
- GUTHRIE C., NOMURA M. – Initiation of protein synthesis: A critical test of the 30S subunit model. *Nature* 219:232-235, 1968.
- HOERZ W., MCCARTY K.S. – Initiation of protein synthesis in a rabbit reticulocyte lysate system. *Bioch. Bioph. Acta* 228:526-535, 1971.
- KAEMPFFER R.O.R., MESELSON M., RASKAS H.J. – Cyclic dissociation into stable subunits and reformulation of ribosomes during bacterial growth. *J. Mol. Biol.* 31:277-289, 1968.
- LAMBORG M.R., ZAMECNIK P.C. – Aminoacid incorporation into protein by extracts of *E. coli*. *Biochim. Biophys. Acta* 42:206-211, 1960.
- LIPMANN F. – Polypeptide chain elongation in protein biosynthesis. *Science* 164:1024-1031, 1969.
- MARSHALL R.E., CASKEY C.T., NIREMBERG M. – Fine structure of RNA codewords recognized by bacterial, amphibian and mammalian transfer RNA. *Science* 155:820-825, 1967.
- MOLDAVE K. – Nucleic acids and protein synthesis. *Annu. Rev. Biochem.* 34:419-443, 1965.
- MONRO R.E. – Protein synthesis: uncoupling of polymerization from template control. *Nature* 223:903-905, 1969.
- NICHOLS J.L. – Nucleotide sequence from the polypeptide chain termination region of the coat protein cistron in bacteriophage R17. *Nature* 225:147-151, 1970.
- NIENHUIS A.W., LAYCOCK, D.G., ANDERSON W.F. – Translation of rabbit haemoglobin messenger RNA by thalassaemic and non-thalassaemic ribosomes. *Nature New Biol.* 231:205-208, 1971.

- NISHIMURA S., WEINSTEIN I.B. – Fractionation of rat liver transfer ribonucleic acid. Isolation of tyrosine, valine, serine and phenylalanine transfer ribonucleic acids and their coding properties. *Biochemistry* 8:832-842, 1969.
- NOMURA M. – Ribosomes. *Sc.Amer.* 221:28-35, 1969.
- NOVELLI G.D. – Amino acid activation for protein synthesis. *Ann. Rev. Biochem.* 36:449-484, 1967.
- OHTA T., THACH R.E. – Binding of formylmethionyl-tRNA and aminoacyl-tRNA to ribosomes. *Nature* 229:238-243, 1968.
- ONO Y., SKOULTCHI A., WATSON J., LENGUEL P. – Stoichiometry of aminoacyl-transfer RNA binding and GTP cleavage during chain elongation and translocation. *Nature* 223:697-701, 1969.
- SCHREIR M.H., NOLL H. – Conformational changes in ribosomes during protein synthesis. *Proc. Natl. Acad. Sci.(USA)* 68:805-809, 1971.
- SINGER M.F., LEDER P. – Messenger RNA: an evaluation. *Ann. Rev. Biochem.* 35:195-230, 1966.
- SÖLL D. – Enzymatic modification of transfer RNA. *Science* 173:293-299, 1971.
- STEITZ J.A. – Polypeptides chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* 224:957-964, 1969.
- WATSON J.D. – The involvement of RNA in the synthesis of proteins. *Science* 140:17-26, 1963.
- ZAMECNIK P.C. – An historical account of protein synthesis with current overtones – a personalized view. *Cold Spring Harbor Symp. Quant. Biol.* 34:1-16, 1969.

## Capítulo 6 – Código Genético

- BRENNER S., BARNETT L., KATZ E.R., CRICK F.H.C. – UGA: A third nonsense triplet in the genetic code. *Nature* 413:449-450, 1967.
- CRICK F.H. – Codon-anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* 19:548-555, 1966.
- CRICK F.H.C. – The origin of the genetic code. *J. Mol. Biol.* 38:367-379, 1968.
- GAREN A. – Sense and nonsense in the genetic code. *Science* 160:149-159, 1968.
- HELINSKI D., YANOFSKY C. – Genetic control of protein structure. In: *The Proteins*, H. Neurath (ed), vol. 4, 1966, pp. 1-93.
- LACEY J.C., PRUITT K.M. – Origin of the genetic code. *Nature* 223:799-804, 1969.

- NIRENBERG M.W., LEDER P. – RNA codewords and protein synthesis. *Science* 145:1399-1407, 1964.
- NIRENBERG M.W., MATTHAEI J.H. – The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. (U.S.A.)* 147:1588-1602, 1961.
- SINGER M.F., LEDER P. – Messenger RNA: An evaluation. *Annu. Rev. Biochem.* 35:195-230, 1966.
- YANOFSKY C., CARLTON B.G., GUEST J.R., HELINSKI D.R., HENNING U. – On the colinearity of gene structure and protein structure. *Proc. Natl. Acad. Sci. (USA)* 51:266-272, 1964.

## Capítulo 7 – Mecanismos de Regulação da Síntese Proteica

- BRETSCHER M.S. – How repressor molecules function. *Nature* 217:509-511, 1968.
- BRITTEN R.J., DAVIDSON E.H. – Gene regulation for higher cells: A theory. *Science* 165:349-357, 1969.
- CLINE A.L., BOCK R.M. – Translational control of gene expression. *Cold Spring Harbor Symp. Quant. Biol.* 31:321-333, 1966.
- ENGLESBERG E., WILCOX B. – Regulation: positive control. *Annu. Rev. Genetics* 8:219-242, 1974.
- EPSTEIN W., BECKWITH – Regulation of gene expression. *Annu. Rev. Biochem.* 37:411-436, 1968.
- GELERHRTER T.D. – Regulatory mechanisms of enzyme synthesis: enzyme induction. In: “Enzyme Synthesis and Degradation in Mammalian Systems”, M. Rechcigl Jr (ed), S. Karger, Basel 1971.
- HENSHAW E.E., HIRSCH C.A., MORTON B.E., HIATT H.H. – Control of protein synthesis in mammalian tissues through changes in ribosomes activity. *J. Biol. Chem.* 246:436-446, 1971.
- JACOB F., MONOD J. – Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3:318-356, 1961.
- MILLER J.H., IPPEN K., SCAIFE J.G., BECKWITH J.R. – The promoter-operator region of the operon of *Escherichia coli*. *J. Mol. Biol.* 38:413-420, 1968.
- O'MALLEY B.W. – Hormonal regulation of nucleic acid and protein synthesis. *Trans. N.Y. Acad. Sc.* 31(serie II):478-503, 1969.
- OHNO S. – Gene duplication, mutation load, and mammalian. Genetic regulatory systems. *J. Med. Genet.* 9:254-263, 1972.

- PTASHNE M. – The operators controlled by the lambda phage repressor. Proc. Natl. Acad. Sci. (USA) 60:1282-1287, 1968.
- RIGGS A.D., BOURGEOIS S. – On the assay, isolation and characterization of the *lac* repressor. J. Mol Biol. 34:361-364, 1968.
- SCHIMKE R.T., DOYLE D. – Control of enzyme levels in animal tissues. Annu. Rev. Biochem. 39:929-976, 1970.
- SILVERSTONE A.E., MEGASANICK B., REZNIKOFF W.S., MILLER J.H., BECKWITH J.R. – Catabolite sensitive site of the *lac* operon. Nature 221:1012-1014, 1969.
- STENT G. – Induction and repression of enzyme synthesis. In: “The Neurociences, a Study Program”, Rockefeller University Press, New York, 1967, pp. 152-161.
- TOMKISS G.M., GELERHRTER T.D., GRANNER D., MARTIN D.JR., SAMUELS H.H., THOMPSON E.B. – Control of specific gene expression in higher organisms. Science 164:1474-1480, 1969.

## Capítulo 8 – Especulações sobre a Origem da Vida

- BARGHOORN E.S. – The oldest fossils. Sc. Amer. 224:30-42, 1971.
- EGLINTON G., CALVIN M. – Chemical fossils. Sci. American 216:32-43, 1967.
- FERRIS J.P. – Cyanovinyil phosphate: a prebiological phosphorylating agent? Science 161:53-54, 1968.
- FOX S.W. – Self-ordered polymers and propagative cell-like systems. Naturwissenschaften 56:1-9, 1969.
- FOX S.W., KRAMPITZ G. – Catalytic decomposition of glucose in aqueous solution by thermal proteinoids. Nature 203:1362-1364, 1964.
- GARRISON W.M., MORRISON D.C., HAMILTON J.G., BENSON A.A., CALVIN M. – Reduction of carbon dioxide in aqueous solutions by ionizing radiation. Science 114:416-418, 1951.
- ISAACS J.D. – The nature of oceanic life. Sc. Amer. 221:146-162, 1969.
- LACEY J.C., PRUITT K.M. – Origin of the genetic code. Nature, 223:799-804, 1969.
- LOHRMANN R., ORGEL L.E. – Prebiotic synthesis: phosphorylation in aqueous solution. Science 161:64-66, 1968.
- MATTHEWS C.N., MOSER R.E. – Peptide synthesis from hydrogen cyanide and water. Nature 215:1230-1234, 1967.

- MEINSCHEN W.G., NAGY B., HENNESSY D.J. – Evidence in meteorites of former life: the organic compounds in carbonaceous chondrites are similar to those found in marine sediments. *Ann. N. Y. Acad. Sci.* 108:553-579, 1963.
- MILLER S.L. – A production of aminoacids under possibles primitive earth conditions. *Science* 117:528-529, 1953.
- MULLER G. – Organic microspheres from the Precambrian of South-West Africa. *Nature* 235:90-95, 1972.
- OPARIN A.I. – The origin of life and the origin of enzymes. *Adv. Enzymology* 27:347-380, 1965.
- ORGEL L.E. – Evolution of the genetic apparatus. *J. Mol. Biol.* 38:381-393, 1968.
- UREY H.C. – Abundances of interstellar molecules and laboratory spectroscopy. Abundances of elements. *Ann. N. Y. Acad. Sci.* 194:35-44, 1972.
- WALD G. – The origins of life. *Proc. Natl. Acad. Sci. (USA)* 52:595-611, 1964.