

Capítulo 2

Biologia molecular

Emanuele Amorim Alves

Daniel Santos Souza

2.1 História da biologia molecular

2.1.1 A descoberta do DNA

Em muitos aspectos a história da biologia molecular se confunde com a descoberta do ácido desoxirribonucleico (DNA) e de sua importância na transmissão de informações entre gerações. Não é de hoje que os cientistas tentam entender as diferentes características de cada espécie. A enorme variedade de seres, desde indivíduos minúsculos até animais de grandes dimensões, tem intrigado e maravilhado muitos pesquisadores. Porém o mundo científico teve de esperar até 1858 para que Charles Darwin (1809-1882) e Alfred Russel Wallace (1823-1913) descrevessem, separada e concomitantemente, suas **teorias da evolução das espécies**. Essas teorias concebem a evolução das espécies como um processo pelo qual o indivíduo mais adaptado ao meio sobrevive, e aquele que não consegue se adaptar a

determinada mudança do ambiente termina por se extinguir. Tais mudanças seriam geradas por mutações adaptativas?

Somente após a segunda metade do século XIX surgiram as primeiras conclusões sobre a perpetuação das características hereditárias dos seres, baseadas no estudo desenvolvido pelo monge austríaco Gregor Mendel (1822-1884). Suas teorias evolutivas, resultado de estudos realizados com ervilhas, constituem o marco inicial da genética, e seus conceitos são até hoje estudados. Mendel introduziu o conceito de genes – chamados por ele de fatores – e a ideia de que eles eram herdados em pares: um gene materno e outro paterno, sendo as características da prole recessivas ou dominantes em relação às características dos pais.

Anos mais tarde, Friedrich Miescher (1844-1895) isolou um tipo de molécula encontrada no núcleo de linfócitos, ao qual denominou nucleína. A nucleína tinha características ácidas e era solúvel em soluções alcalinas diluídas. Miescher também determinou a composição química dessa molécula, que seria rica em oxigênio, nitrogênio e fósforo. Mais tarde, a natureza ácida da nucleína serviria de base para a denominação ácidos nucleicos.

Entre 1882 e 1885, Walther Flemming (1843-1905) e Eduard Strasburger (1844-1912), em estudos sobre estruturas celulares, descobriram estruturas em forma de bastão no núcleo das células e as chamaram de **cromossomos**, pois eram estruturas intensamente coradas. Poucos anos mais tarde, Theodor Boveri (1862-1915) observou que o número de cromossomos das células germinativas, em determinado estado de sua maturação, era reduzido à metade. Essa observação permitiu elucidar o fenômeno de união dos gametas para criarem uma célula somática do organismo em formação, reforçando ainda mais a teoria de Mendel.

Em 1909, o termo **gene** foi introduzido por Wilhelm Johannsen (1857-1927) para designar a unidade mendeliana antes conhecida como fator, designando os termos **genótipo** para as características genéticas do indivíduo e **fenótipo** para se referir ao seu aspecto externo.

Baseado no estudo que fez com a mosca-das-frutas (*Drosophila melanogaster*), o pesquisador Thomas Hunt Morgan (1866-1945) publicou, em 1915, o livro *O mecanismo da hereditariedade mendeliana*, em que afirmava

representarem os genes regiões dos cromossomos responsáveis pelas características inatas do indivíduo. Curiosamente, esse trabalho teve o auxílio de Alfred Sturtevant (1891-1970), um aluno de Morgan que, ao ler os trabalhos de Mendel, inicialmente não lhes dera muito crédito. A partir do trabalho de Thomas Hunt Morgan, passou-se a aceitar que os cromossomos estocavam muitos genes, e que os genes guardavam as informações passadas à progênie.

Em 1923, Robert Feulgen (1884-1955) demonstrou, por meio de técnicas de coloração específicas para o DNA, que o mesmo estava ligado intimamente aos cromossomos. Contudo, até então não se sabia da importância da molécula do DNA na transmissão das informações de um indivíduo para outro.

Até o início do século XX, as biomoléculas mais estudadas foram as **proteínas**, pois seu isolamento era mais simples do que o de outras biomoléculas, e a cada ano eram obtidas mais informações sobre proteínas do que sobre qualquer outra biomolécula existente na célula. Por isso, era de se esperar que a transmissão de informações de pais para filho, processo muito importante para a manutenção da espécie, fosse feita por meio dessas moléculas tão complexas, as proteínas.

Esse paradigma só foi quebrado em 1928, quando Frederick Griffith (1879-1941) introduziu o conceito de **transformação** com base em seus experimentos com pneumococos, bactérias causadoras da pneumonia. Nesse experimento, Griffith utilizou dois tipos de bactérias: o *Pneumococcus* do tipo S (do inglês *smooth*, que significa liso), que possui superfície lisa em decorrência de uma cobertura de polissacarídeo – o que o torna extremamente virulento –, e o *Pneumococcus* do tipo R, assim chamado por possuir a superfície rugosa (*rough* em inglês) e não ter cobertura polissacarídica, não sendo virulento.

Ao inocular a bactéria R viva e a bactéria S morta, o camundongo morria. Griffith concluiu então que havia algum fator na bactéria S que “transformava” a bactéria R em S. Essa conclusão foi confirmada quando Griffith observou, nos camundongos mortos, a presença de colônias de bactérias S vivas. Como elas eram mortas por meio do uso de calor, que desnatura proteínas, era lógico supor que o material que transmitia essa informação não era uma proteína. Pesquisadores contemporâneos de Griffith concluíram que a substância capaz de estimular essa “transformação” devia ser o material genético dos *Pneumococci* do tipo S, que readquiriam a sua virulência.

Alguns anos após o experimento de Griffith, pesquisadores repetiram os seus experimentos utilizando separações que continham apenas DNA, carboidratos, proteínas ou lipídios. Constatou-se que somente a porção com DNA era capaz de transformar as bactérias, e que o DNA perdia essa capacidade quando tratado com enzimas **DNAases**. Tais experimentos confirmaram a teoria de Griffith e deram grande credibilidade aos seus experimentos.

Em 1931, Phoebus Aaron Theodor Levene (1869-1940) identificou que **bases nitrogenadas, açúcar e fosfato** formavam as estruturas básicas dos ácidos nucleicos, afirmando que esses ácidos eram polímeros. Foi ele também que descreveu as diferenças entre o ácido ribonucleico (RNA) e o DNA, sendo essa nomenclatura amplamente utilizada após suas pesquisas.

Em 1952, Alfred Hershey (1908-1997) e Martha Chase (1927-2003) estavam utilizando em seus experimentos um tipo de vírus, denominado fago, capaz de infectar bactérias. No experimento, que ficou conhecido como “experimento do liquidificador”, eles marcaram duas culturas de fagos. Uma delas foi marcada com enxofre radioativo, que se incorporou às proteínas, e a outra, com fósforo radioativo, que se incorporou ao DNA. Posteriormente, incubaram culturas de *Escherichia coli* com os fagos para que as bactérias se contaminassem com os vírus; em seguida, liberaram os vírus das células bacterianas, utilizando um liquidificador. Com isso, as cápsulas virais, por serem mais leves, localizavam-se no sobrenadante e as bactérias, no precipitado. Hershey e Chase perceberam que grande parte do fósforo se encontrava no precipitado e que a maior parte do enxofre estava no sobrenadante, e concluíram que o material infectante era o DNA do vírus injetado dentro das células bacterianas. Após esse experimento, houve uma grande “corrida científica” para descobrir a estrutura do DNA.

Utilizando a técnica de cromatografia, Erwin Chargaff (1905-2002) verificou que havia uma relação quantitativa entre as bases nitrogenadas do DNA: a quantidade de adenina era proporcional à quantidade de timina (T), assim como a de citosina (C), à de guanina (G). Esses dados foram posteriormente denominados **postulados de Chargaff** e seriam muito importantes na descoberta da estrutura do DNA.

Em 1953, os pesquisadores James Dewey Watson (1928) e Francis Harry Compton Crick (1916-2004) publicaram trabalho em que elucidaram

a estrutura do DNA, utilizando para isso dados das pesquisas de Rosalind Franklin (1920-1958) e Linus Pauling (1901-1994). Por meio da difração de raios X, Rosalind obteve excelentes imagens do DNA que mostravam com grande precisão as dimensões da molécula: a estrutura era formada por uma espiral de 20 Å, a distância entre as bases dispostas paralelamente ao eixo helicoidal era de 3,4 Å e o valor da altura de um giro da espiral cristalizada era de 34 Å. Pauling descreveu inicialmente um modelo para o DNA no qual as bases nitrogenadas estavam voltadas para o lado externo da molécula e os fosfatos, voltados para o interior da espiral.

Watson e Crick viram que o DNA era uma espiral, mas as bases não poderiam ser externas à molécula por proporcionar características básicas ao DNA. Além disso, se os grupos fosfato estivessem tão próximos uns dos outros no interior da estrutura seriam repelidos, fazendo-a instável.

Utilizando essas informações, Watson e Crick descreveram a molécula de DNA como uma espiral na qual as bases nitrogenadas são ligadas entre si internamente (timina com adenina e citosina com guanina) e os grupos fosfato, carregados negativamente, estão voltados para fora da molécula. O corpo da molécula é constituído por polímeros de base nitrogenada açúcar-fosfato, e as ligações entre eles, denominadas ligações fosfodiéster, ocorrem na direção 3'-5'.

A elucidação da estrutura do DNA representou um marco na biologia, sendo responsável pelo surgimento da biologia molecular. Originada na bioquímica e na genética, a biologia molecular, principal ferramenta na atualidade para a elucidação e o aperfeiçoamento da vida, é a ciência do século XXI e seguramente ainda tem potencial para auxiliar o ser humano.

2.1.2 A estrutura dos ácidos nucleicos

O DNA, assim como o RNA, é um polímero formado por vários nucleotídeos ligados entre si. Os nucleotídeos são formados por uma base nitrogenada, um açúcar e um radical de ácido fosfórico (fig. 2.1). Graças ao ácido fosfórico, os ácidos nucleicos têm propriedades ácidas e possuem carga negativa quando em pH neutro ou alcalino.

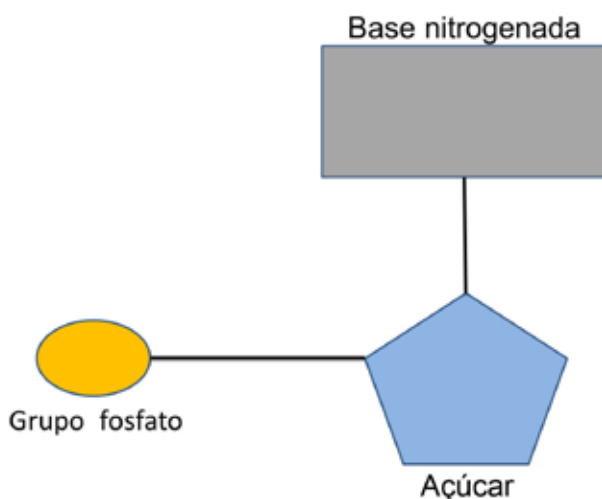


Figura 2.1. Esquema da estrutura de um nucleotídeo.

As bases nitrogenadas podem ser purinas – adenina e guanina – e pirimidinas – citosina, timina e uracila (presente somente no RNA). A estrutura das bases nitrogenadas encontra-se na figura 2.2.

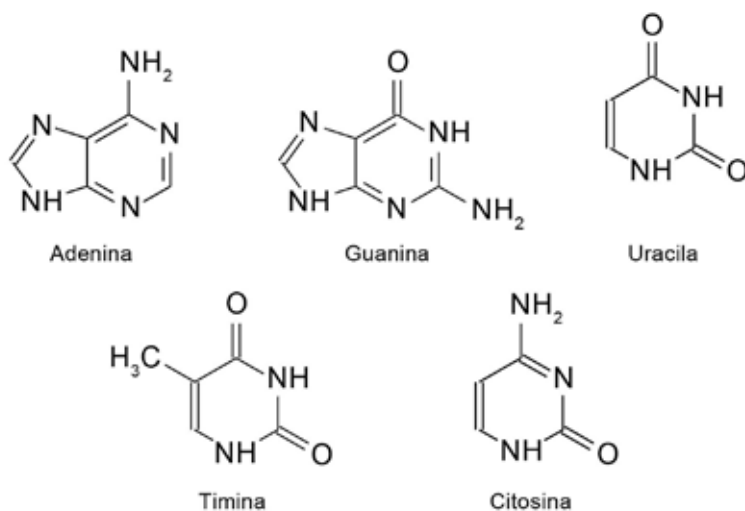


Figura 2.2. Estrutura das bases nitrogenadas do DNA e do RNA: a molécula de DNA apresenta as bases adenina, timina, citosina e guanina; o RNA contém adenina, uracila, citosina e guanina.

Os nucleotídeos se ligam por meio de ligações fosfodiéster que ocorrem entre a hidroxila do carbono 5' da pentose e a hidroxila do carbono 3' do nucleotídeo seguinte.

O DNA é formado por duas fitas de nucleotídeos; já o RNA é formado por uma única fita. Existe interação entre as duas fitas de DNA: por intermédio de pontes de hidrogênio entre as bases e por interações hidrofóbicas. As pontes de hidrogênio ocorrem entre as bases timina (T) e adenina (A) e entre a citosina (C) e a guanina (G). Entre a timina e a adenina, existem duas pontes de hidrogênio; entre a citosina e a guanina existem três. Essas ligações ocorrem por causa das dimensões moleculares das bases nitrogenadas e de seus radicais, assim como em decorrência do posicionamento dos grupos que formam as pontes de hidrogênio. Isso significa que as duas fitas de DNA são complementares entre si, sendo esse fato importante no momento da transmissão de informação. Outra forma de interação ocorre por meio das interações hidrofóbicas entre os anéis heterocíclicos das bases. Essa interação faz que a parte interna das fitas seja apolar e, em um ambiente aquoso, ela favorece a interação entre a dupla fita.

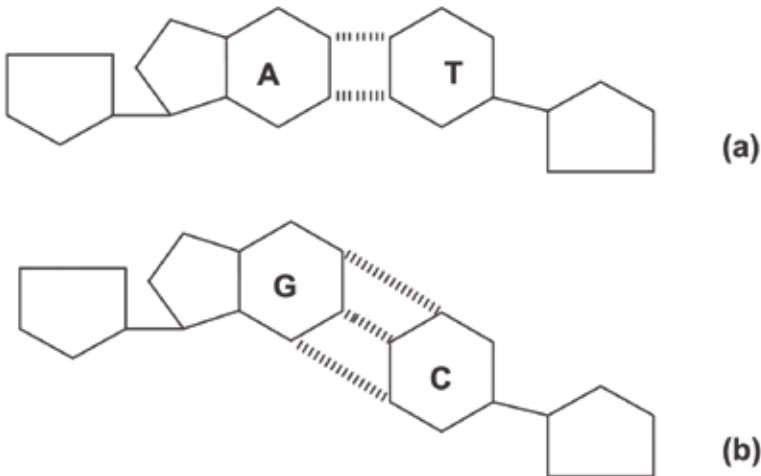


Figura 2.3. Esquema simplificado das interações de hidrogênio que ocorrem entre as bases nitrogenadas do DNA. Em (a), as duas pontes que ocorrem entre a timina e a adenina; em (b), as três pontes de hidrogênio entre a guanina e a citosina.

A espiral formada pelas fitas de DNA também possui outras características. Seu formato favorece a formação de duas reentrâncias, denominadas sulco menor e sulco maior, que auxiliam a interação entre as proteínas e a molécula do DNA. A espiral do DNA também apresenta mais duas conformações, que são variações e dependem da composição de bases e do meio em que o DNA se encontra. Essas formas foram analisadas por meio de estudos de cristalografia.

A espiral descrita por Watson e Crick é denominada forma B do DNA. Essa é a forma mais abundante nas condições fisiológicas e em soluções de baixa força iônica. Ainda existem as formas A e Z. A forma A é mais curta e mais grossa do que a forma B, e é encontrada em meios de baixo conteúdo salino. A forma Z é mais longa e fina do que a forma B. Não se conhecem ainda as vantagens da estrutura Z, mas é fato que há conversão entre B e Z no organismo.

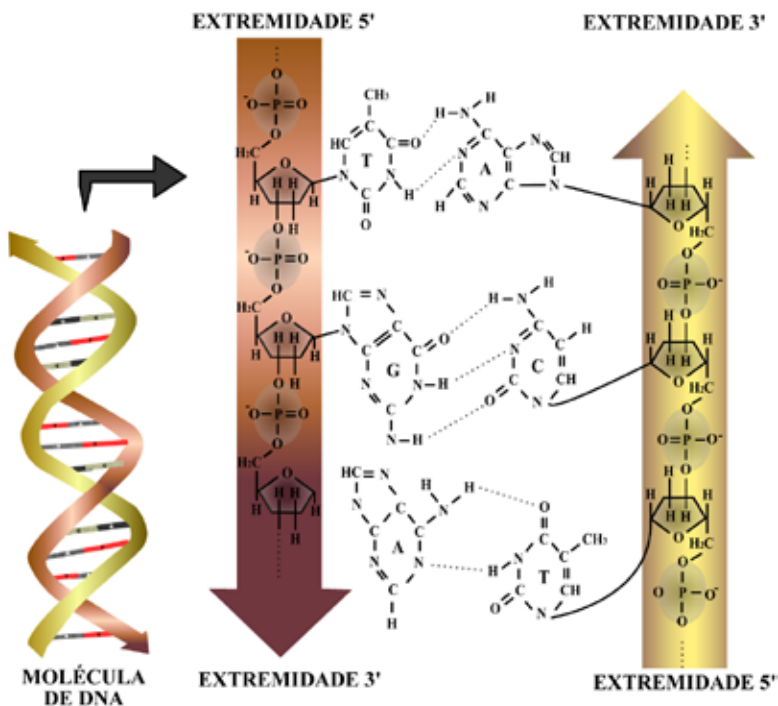


Figura 2.4. Representação esquemática da molécula dupla fita de DNA.

O RNA é semelhante ao DNA, porém não forma fita dupla. Além disso, o açúcar do RNA é uma ribose, diferente do DNA, que é uma desoxirribose. Outra característica particular do RNA é que as bases nitrogenadas que formam seus nucleotídeos apresentam pequena diferença em relação ao DNA. No RNA, a timina é substituída pela uracila (U), que também parecia com a adenina.

O RNA possui três classes principais: RNA mensageiro (mRNA), que contém a informação genética para a produção de proteínas; RNA transportador (tRNA), responsável por carrear o aminoácido específico para o sítio de síntese proteica; e RNA ribossomal (rRNA), que forma grande parte da massa do ribossomo. É no ribossomo que ocorre a síntese proteica.

2.1.3 Características físicas da molécula de DNA

Uma característica dos ácidos nucleicos é a sua capacidade de absorver luz ultravioleta no comprimento de onda de 260 nm. Isso ocorre por causa das ligações duplas dos anéis que formam as bases nitrogenadas. Essa característica favorece medições que utilizam a espectrofotometria como ferramenta e que permitem quantificar as moléculas de DNA presentes em determinada solução e também obter informações sobre as características da fita. Como as bases se encontram no interior da fita, sua absorvância é menor quando comparada à mesma molécula desnaturada, pois, quando as fitas estão separadas, elas expõem as bases, permitindo a adsorção de maior quantidade de luz ultravioleta (UV).

Uma das formas mais comuns de separar as fitas de DNA é aumentando-se a temperatura. Isso induz o aumento da energia cinética na molécula, causando o rompimento das pontes de hidrogênio entre as bases nitrogenadas, além de romper as interações hidrofóbicas – processo denominado **desnaturação**. A desnaturação também ocorre a partir da alteração de alguns parâmetros físicos, como o pH do meio, entre outros.

A capacidade de aumentar a absorção da luz UV pela molécula de DNA, utilizando o calor para desnaturá-la, é denominada **hipercromia**. Esse fenômeno é reversível: com a diminuição da temperatura, as fitas de DNA se unem lentamente, fazendo a absorção de luz voltar à absorção inicial. A reorga-

nização das cadeias é evidenciada pela diminuição da leitura de absorção, fenômeno chamado **hipocromia**.

Ao se avaliar a relação entre a temperatura e a absorvância de uma amostra, observa-se que, conforme a temperatura aumenta, também aumenta a absorvância. Isso acontece porque o aumento da temperatura das fitas abre as duplas, permitindo maior interação das bases com a luz UV e aumentando sua absorção. O ponto médio de absorção, denominado T_m , representa o ponto médio da curva de aumento de temperatura *versus* absorção da luz UV e define a temperatura na qual a metade das cadeias está dissociada.

O T_m é importante, pois permite avaliar a quantidade de citosina e guanina presente no DNA – quanto mais C e G, maior a interação, por causa da formação de três pontes de hidrogênio. Assim, em moléculas com alto teor de C e G, o T_m é mais alto. Essa característica distingue as moléculas de diferentes DNAs.

Para a manutenção da estrutura da dupla fita, também existem as interações hidrofóbicas dos anéis das bases. Essas interações são evidenciadas quando se coloca o DNA em uma solução contendo compostos que desestruturam as interações, como detergentes ou o sal trifluoroacetato. Quando essas substâncias são acrescentadas, o T_m diminui; quando elas são retiradas, ocorre a reassociação, assim como ocorre com o aquecimento e o resfriamento da molécula. Nesse caso, o tempo de reassociação depende do tamanho da molécula, o que permite calcular o tamanho do DNA por meio de cálculos utilizando dados desses experimentos. Essa análise é importante ao se tratar de DNAs desconhecidos.

Além das forças de interação, também há forças de repulsão, exercidas pelos grupos fosfatos dos nucleotídeos, na dupla fita. Assim, se o DNA for inserido em uma solução que contenha íons que minimizem a repulsão, haverá maior estabilidade da molécula. Isso ocorre quando acrescentamos à solução cátions como o Na^+ .

2.2 Duplicação da molécula de DNA

As informações contidas no DNA residem na ordem das bases nitrogenadas que o compõem. No momento da duplicação, essas informações

devem ser passadas para as fitas filhas. Porém de que forma é possível preservar essa informação?

Desde a publicação do trabalho de Watson e Crick sobre a estrutura do DNA, concebeu-se a possibilidade de duplicação do DNA. Nesse trabalho, além da descrição da dupla hélice, os cientistas ressaltaram que a duplicação do DNA ocorria de forma semiconservativa, ou seja, a dupla hélice recém-formada seria um híbrido entre uma nova cadeia e uma cadeia do DNA que dera origem a ela. Assim, a complementaridade entre as bases é uma forma de preservar as informações contidas no DNA.

A duplicação de forma semiconservativa foi comprovada nos experimentos de Matthew Meselson (1930) e Franklin Stahl (1929), em trabalho publicado em 1958. Meselson e Stahl incubaram bactérias *Escherichia coli* em um meio onde a única fonte de nitrogênio eram os isótopos de N^{15} , não radioativos, mas mais pesados que o N^{14} . As bactérias eram mantidas durante várias gerações nesse meio, a fim de incorporar nitrogênio em seu DNA. Posteriormente, eram retiradas do meio e colocadas em meio contendo N^{14} . Com esse experimento, Meselson e Stahl desejavam avaliar se as novas moléculas de DNA eram híbridas, ou seja, parte sintetizada e parte originária da fita antiga.

Os pesquisadores lisaram as membranas das células e isolaram o seu DNA; esse material foi posteriormente centrifugado em gradiente de densidade. Meselson e Stahl observaram que havia moléculas de DNA leves na parte superior do tubo, moléculas com peso intermediário e moléculas mais pesadas. Assim, concluíram que a duplicação do DNA era semiconservativa, conforme suposto anos antes por Watson e Crick, porque as moléculas leves só continham N^{14} , as moléculas pesadas continham somente N^{15} , e as moléculas intermediárias eram formadas por fitas híbridas contendo os dois isótopos do nitrogênio.

2.2.1 A duplicação do DNA e o ciclo celular

A necessidade de duplicação do DNA é algo que ocorre quando a própria célula precisa duplicar todo o seu material para formar uma nova célula. A

duplicação celular é parte integrante do ciclo celular e deve estar relacionada à sinalização recebida pela célula em determinado momento de sua vida.

O ciclo celular é formado por cinco fases: G₁, S, G₂, M e G₀. Na fase G₀, a célula encontra-se em período de dormência, com o maquinário celular voltado para as funções celulares. O ciclo de replicação se inicia mediante sinalizações que induzem a célula a entrar no período G₁. Nessa fase, ocorre a síntese de proteínas, enzimas e RNA. Em seguida, a célula entra na fase S, na qual há a duplicação do material genético. Na fase G₂, ocorre a síntese do material necessário para a duplicação celular. Já na fase M, ocorre a duplicação da célula como um todo.

A duplicação do DNA na fase S deve ter uma velocidade tal que acompanhe todo o processo do ciclo celular da célula. Células procarióticas promovem sua duplicação em tempos muito mais curtos do que as células de organismos eucarióticos. O tempo de geração da *E. coli*, por exemplo, é de aproximadamente 20 minutos; já as células animais em cultura apresentam ciclos celulares de até 24 horas.

A duplicação se inicia em pontos específicos do DNA, denominados **ORIs** (origens de replicação), uma região rica em **sequências GATC**, com aproximadamente 15 a 20 nucleotídeos. As sequências GATC são sítios de metilação do DNA nas duas fitas responsáveis pela ligação do DNA à membrana celular. É na região de ORI que se inicia a forquilha de replicação. Essa forquilha pode ser ativada mais de uma vez nos procariotos e somente uma vez nos eucariotos. No DNA humano podem existir vários ORIs. Quando há a ativação, verifica-se a abertura das fitas de DNA na região de ORI, formando-se a forquilha de replicação. Isso ocorre com o auxílio da enzima helicase, que é capaz de desfazer as pontes de hidrogênio entre as bases nitrogenadas, com gasto de trifosfato de adenosina (ATP).

Após a abertura, as fitas são estabilizadas pela ação das topoisomerasas. As proteínas estabilizadoras de fita simples, ou SSB (do inglês *single strand binding proteins*), protegem a fita aberta. Essas proteínas se ligam ao DNA de forma cooperativa e impedem que o DNA aberto se feche, mantendo o DNA de fita simples numa conformação ideal para a duplicação, além de protegerem a fita de ser degradada por nucleases.

Com as fitas abertas, a duplicação se inicia pela ação da enzima primase, responsável pela inserção de um RNA iniciador, chamado *primer*. O *primer* é um conjunto de nucleotídeos com cerca de 20 pares de bases; ele serve para promover uma terminação 3'-OH livre. A terminação livre é importante para o início da síntese da fita de DNA.

A enzima responsável pela síntese da cadeia complementar é a DNA polimerase. Essa enzima necessita da extremidade 3'-OH livre para iniciar o processo de alongamento da fita, ou seja, a adição de nucleotídeos na extremidade 3'-OH livre de uma região pareada do DNA, possibilitando o crescimento da cadeia no sentido 5'-3'. A DNA polimerase possui três funções distintas: polimerização (inserção de nucleotídeos na direção 5'-3'), função exonuclease no sentido 3'-5' (edição e correção) e função exonuclease no sentido 5'-3' (retirada dos *primers*).

A DNA polimerase é dimérica e capaz de sintetizar as duas fitas ao mesmo tempo; porém, como a sua síntese se dá na direção 5'-3', uma das fitas é polimerizada continuamente, ao passo que a polimerização da outra se dá de forma fragmentada, uma vez que as fitas do DNA possuem polaridades opostas. A fita sintetizada de forma contínua é denominada fita *leading* e a fita fragmentada é denominada fita *lagging*. Na fita fragmentada, é necessário que haja constante inserção de *primers* para haver extremidades 3'-OH livres; assim, a atuação da primase é contínua nessa fita. Os fragmentos obtidos pela síntese descontínua são denominados fragmentos de Okazaki. Eles possuem cerca de 100 a 200 nucleotídeos de comprimento, sendo intermediados pelos *primers* sintetizados pelas primases. Com o final da síntese e a substituição dos *primers* por DNA, a enzima DNA ligase liga os fragmentos que foram formados na fita contínua e na fita descontínua para que haja, no final, uma fita inteira de DNA complementar ao seu molde.

A terminação da duplicação ocorre de diversas formas. No DNA circular, de procariotos e mitocôndrias, o término da síntese se dá quando as duas forquilhas de replicação se encontram numa região denominada **Ter**. Para terminar a replicação, ela foi arranjada de maneira a formar uma espécie de armadilha onde a forquilha de replicação entre e não possa sair. Isso é obtido pela ligação da proteína **Tus** à região **Ter**.

Nas células de DNA linear, há um problema na replicação na extremidade da fita tardia. Para solucioná-lo, ocorre a síntese dos telômeros. Os telômeros são porções terminais dos cromossomos que não contêm informação para as células e que possuem várias cópias de uma sequência consenso que são adicionadas, pela telomerase, à fita tardia no final de replicação. Na fita *leading*, a terminação ocorre naturalmente ao final do molde parental.

2.3 Transcrição

Existem semelhanças e diferenças entre o processo de duplicação do DNA e a síntese de RNA. Na duplicação, deve-se ter o máximo de cuidado para que as informações sejam transferidas de forma fidedigna, ou seja, sem que nenhuma informação seja perdida. Na transcrição, que é a síntese de RNA utilizando a molécula de DNA como modelo, retratam-se as necessidades da célula em determinado momento de sua vida. Somente serão produzidos RNAs para a síntese de proteínas necessárias para a célula naquele momento determinado de sua existência. É na transcrição que a célula regula sua expressão gênica.

A transcrição é um processo executado pela enzima RNA polimerase e assemelha-se muito à duplicação, sendo a fita resultante única e formada por nucleotídeos de RNA. Somente uma das fitas do DNA serve como molde para a fita de RNA que será formada. A transcrição segue as mesmas regras de pareamento da duplicação do DNA, mas a timina é substituída pela uracila, que parecia com a adenina.

Assim como o DNA, o RNA é formado por nucleotídeos compostos por um radical de ácido fosfórico, uma base nitrogenada e um açúcar. No RNA, esse açúcar é a ribose. As bases nitrogenadas do RNA são as mesmas do DNA, com exceção da timina, que, no RNA, é substituída pela uracila.

Em eucariotos, existem três RNA polimerases diferentes. A RNA polimerase I é responsável pela síntese de RNA ribossomal; a RNA polimerase II sintetiza o RNA que origina as proteínas (RNA mensageiro); e a RNA polimerase III é responsável pela síntese de RNAs menores e de RNA transportador.

Sabe-se que grande parte do genoma humano não é funcional, ou seja, não gera nenhuma proteína. Assim, é de se esperar que grande parte do DNA não seja transcrita. Essa região não transcrita é denominada *intron*, e a região funcional é denominada *éxon*.

Para que a RNA polimerase possa iniciar a transcrição e dissociar-se depois do gene, é necessário um estímulo. No caso de genes constitutivos não há esse tipo de regulação, pois eles sempre estão sendo transcritos. A regulação ocorre principalmente em genes que se expressam somente em determinados momentos da vida da célula. Esse processo é chamado regulação da expressão gênica. A RNA polimerase não necessita de *primer* para iniciar a transcrição.

No DNA existe uma região, que antecede a região a ser transcrita, denominada região promotora. Nela existem sistemas que auxiliam a RNA polimerase a reconhecer o local onde deve iniciar a transcrição. Nos procariotos, essa região é denominada **caixa de Pribnow** e, nos eucariotos, é denominada **caixa de TATA**, por causa das repetições TATA presentes nela.

A transcrição se inicia na região promotora, e o primeiro nucleotídeo a ser transcrito é denominado +1 (sítio de início). Regiões que antecedem ao sítio de início recebem números negativos e crescentes, ao passo que as regiões posteriores recebem números positivos e crescentes.

A fase de reconhecimento da região promotora é crítica na regulação da expressão gênica, porque define se um gene será transcrito ou não. Existem duas sequências altamente conservadas nos genes. A primeira localiza-se na região -10 e tem a sequência TATAAT em repetições. A segunda está na região -35 e possui a sequência TTGACA. A RNA polimerase reconhece a sequência em -35 e inicia a sua interação com a fita de DNA. Essa interação se fortalece no momento em que a RNA polimerase chega à região -10.

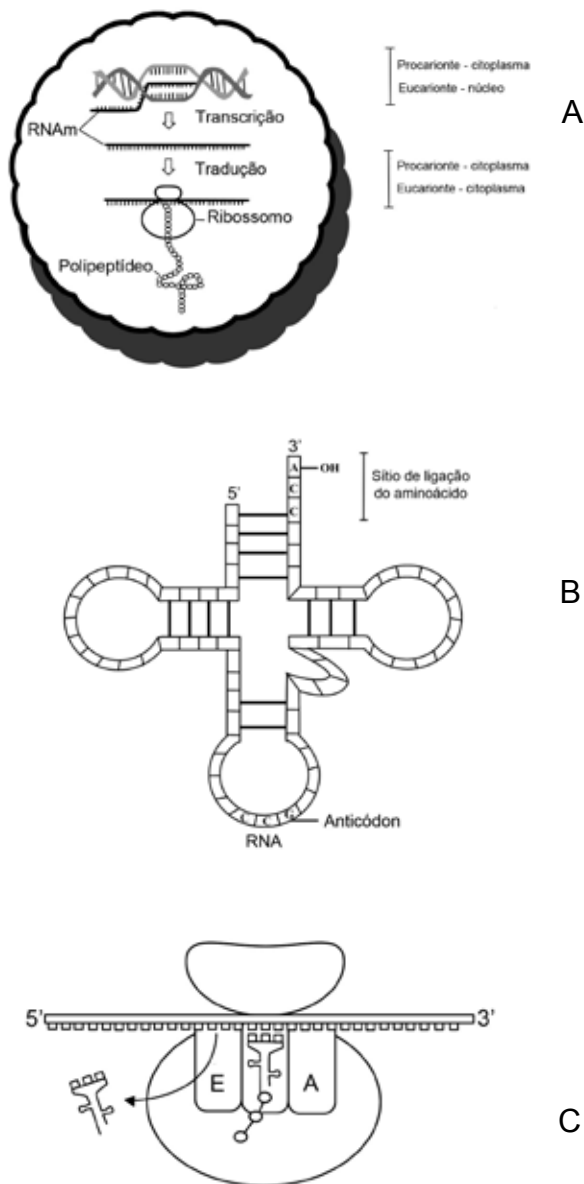


Figura 2.5. Representação esquemática dos processos de transcrição e tradução: A) visão geral do fenômeno dentro de células eucariotas e procariontes; B) RNA transportador, evidenciando a região onde encontramos a trinca de bases nitrogenadas chamada anticódon; C) interação do RNA mensageiro com o RNA transportado e com as subunidades ribossomais, representando a síntese de uma cadeia polipeptídica.

2.3.1 Tipos de RNA

2.3.1.1 RNA mensageiro

O RNA mensageiro de procariotos, ou mRNA, é policistrônico, isto é, uma molécula possui informações de mais de um gene controlado pelo mesmo promotor. Também existem sequências, denominadas sequências líderes, que são transcritas, mas não traduzidas. São sequências importantes no momento do reconhecimento do mRNA pelo ribossomo. Esse mecanismo de transcrição é fundamental nas bactérias por aumentar a velocidade da transcrição, com várias proteínas podendo ser sintetizadas ao mesmo tempo. Geralmente os RNAs traduzem várias cópias da mesma proteína ou de proteínas diferentes que participam de uma mesma via metabólica. Em procariotos, o mRNA tem vida curta, não é processado, mas é transcrito, traduzido e logo degradado.

Nos eucariotos, o mRNA é um pouco diferente, sofrendo processamento de perda dos íntrons e adição de proteções que aumentam o tempo de vida do RNA. Essas modificações são chamadas modificações pós-transcricionais ou processamento do RNA. O RNA de eucarioto, quando transcrito, não está completamente maduro e é chamado transcrito primário ou pré-RNA. Para que o RNA possa exercer sua função biológica, é necessário que ele passe pelo processamento.

No mRNA de eucariotos, essas modificações constituem principalmente a colocação de estruturas protetoras e sinalizadoras, e na retirada de partes que não gerarão proteínas (*splicing*), sendo a principal modificação a inserção de nucleotídeos na extremidade 5' (*cap*) e 3' (poliadenilação).

O *splicing*, retirada de íntrons do transcrito primário, é realizado com o auxílio de um complexo de ribonucleoproteínas denominado spliceossomo. Esse processo ocorre em duas etapas e após duas reações de transesterificação, nas quais há a liberação do íntron na forma de laço. Também há íntrons que participam de sua própria retirada, num processo denominado *autosplicing*.

A adição do *cap* 5 se dá pela ação da guanilil-transferase, quando se insere no DNA um resíduo de guanina por meio de ligação covalente. O *cap* protege o RNA da degradação por exonucleases, além de auxiliar no reconhecimento pelo ribossomo do local onde se iniciará a síntese proteica.

A poliadenilação ocorre ao término da transcrição e é auxiliada pela enzima poli-A polimerase. Além de proteção, a cauda poli-A auxilia no processo de término da transcrição.

2.3.1.2 RNA ribossômico

O RNA ribossômico, ou rRNA, é um tipo de RNA que participa da estrutura do ribossomo. Os ribossomos são estruturas mistas de rRNA e proteínas nas quais ocorre a síntese de proteínas.

Nos procaríotos, o gene que codifica o rRNA se encontra em sequência, e os genes são transcritos em *tandem*. Os três tipos de rRNA são 16S, 23S e 5S. Logo após a sua transcrição, os RNAs são processados pela enzima **Rnase II** – responsável por uma série de clivagens –, resultando na liberação dos rRNAs. O processamento até a forma madura dos rRNAs é feito por outras ribonucleases.

Em eucariotos, o processo é muito semelhante, havendo quatro tipos de rRNA: 5S, 5,8S, 18S e 28S. Como nos procaríotos, seus genes são transcritos de forma a fazer parte de uma mesma e longa unidade de transcrição, produzida pela RNA polimerase I. Somente o 5S não se encontra próximo dos outros, e seus genes estão espalhados pelo genoma do indivíduo. Sua transcrição é feita pela RNA polimerase III e não pela I.

2.3.1.3 RNA transportador

Nos procaríotos, os genes que codificam os RNAs transportadores, ou tRNAs, podem ou não estar agrupados e podem codificar o mesmo tRNA ou tRNAs diferentes. Quando estão unidos, são separados no processamento. O gene do tRNA não é policistrônico, e esses genes até podem ser transcritos juntos, mas a ação de nucleases altamente específicas os separa, e eles geram extremidades 5' e 3'. A molécula de tRNA é transcrita ao mesmo tempo, sendo metilada e adquirindo estrutura secundária, muito importante no seu processamento. Na extremidade 3', sempre há a adição da sequência de nucleotídeos ACC, que irá formar o primeiro braço do tRNA – no total, o tRNA apresenta cinco braços, sendo quatro deles fixos e um variável. Cada braço é responsável por uma ação específica que irá auxiliar na função do

tRNA de transportar o aminoácido certo para o sítio de síntese proteica. No braço que contém ACC, existe o encaixe do aminoácido que será transportado. Esse encaixe é auxiliado pela enzima aminoacil-tRNA sintetase. O segundo braço reconhece o ribossomo, o terceiro contém o código complementar à sequência do mRNA (anticódon) e o quarto contém o encaixe da aminoacil-tRNA sintetase no momento em que carrega o aminoácido. O quinto braço variável não possui função específica.

Nos eucariotos, os tRNAs são muito semelhantes; sintetizados separadamente, seus braços são específicos para cada tipo de aminoácido e apresentam estrutura terciária.

2.3.2 Regulação da expressão gênica

Todos os organismos possuem formas de regulação de quando e quais genes serão transcritos em determinado momento da vida. Alguns genes, ditos constitutivos, são transcritos continuamente, pois são fundamentais na formação de estruturas celulares básicas para o perfeito funcionamento da célula, além de participarem de rotas metabólicas indispensáveis, como a glicólise. Outros genes, ditos induzíveis, somente são transcritos quando a proteína codificada por eles é necessária à célula. Esse processo economiza a energia gasta pela célula na síntese e na degradação de proteínas.

Para as proteínas constitutivas, existem sequências que antecedem a seus genes no DNA, denominadas promotores fortes. Essas sequências são facilmente reconhecidas pela RNA polimerase e muito preservadas. Nos genes induzíveis, esses promotores não estão acessíveis à RNA polimerase a todo momento; estão acessíveis apenas quando há um estímulo externo ou interno para a síntese da proteína.

Em eucariotos multicelulares, a resposta direta às condições externas é limitada. A maioria das células encontra-se em meios uniformes, e existe uma ação coordenada que possibilita a sua diferenciação, culminando no desenvolvimento do organismo. Em determinado momento, as células sintetizam e acumulam diferentes mRNAs e proteínas que possibilitarão a diferenciação.

A regulação da expressão gênica em eucariotos se dá, na maioria das vezes, pela ação de hormônios, principalmente esteroides, sendo pouco afetada por mudanças ambientais.

Nos procaríotos, a ausência ou a presença de nutrientes no meio influencia muito a expressão proteica. Na presença de determinado nutriente, aumenta a expressão das enzimas responsáveis pelo metabolismo desse nutriente.

Em 1961, surgiu o modelo de **óperon**. Existem vários óperons, mas o mais estudado é o óperon lac. Nele, o RNA é policistrônico e responsável pela sequência das três proteínas envolvidas no metabolismo da lactose. O gene lac Z codifica a enzima β -galactosidase, responsável pela quebra da lactose em glicose e galactose. O gene lac Y é responsável por codificar uma proteína permease que permite a entrada da lactose na célula. O gene lac A codifica a enzima transacetilase, que possui função desconhecida.

Constitutivamente, a célula sintetiza um repressor que se sobrepõe ao promotor no mRNA da célula, impedindo que ele seja transcrito. No momento em que há ausência de glicose e presença de lactose, ocorre a ligação da lactose com o repressor, inativando-o, a fim de que não haja nenhum impedimento para que o promotor impeça a transcrição dos genes lac Z, lac Y e lac A. Assim, a célula sintetiza o maquinário que permitirá o aumento da entrada de lactose nela e a quebra da lactose em galactose e glicose, para ser posteriormente usada na via glicolítica, visando à obtenção de energia.

As vias degradativas são induzidas, em geral, quando há a presença do seu substrato. Nas vias biossintéticas, pelo contrário, a ausência de determinado nutriente estimula a sua síntese, como é o caso da via de síntese do triptofano, que se assemelha à do lac óperon.

Nos eucariotos, muitas são as formas de regulação da expressão gênica. No término da transcrição de determinado gene, a expressão gênica pode não terminar ali e continuar, transcrevendo outro gene. Outra forma de regulação é a que ocorre com algumas proteínas ribossomais que, quando em excesso, ligam-se aos genes que as originariam e impedem a sua transcrição. Também é possível encontrar partes do DNA onde ocorre metilação. A presença da metila inibe a transcrição do gene daquela região. A regulação também se dá no processamento do RNA, pois, com a adição da cauda poli-A do cap 5, o

tempo de vida de determinado mRNA aumenta, tornando possível a tradução de mais proteínas a partir dele.

2.4 Tradução

2.4.1 O código genético

Como já mencionado, a descoberta do DNA ocasionou uma grande corrida para decifrar as informações contidas nessa molécula e como elas eram transmitidas. Sabe-se que as estruturas de constituição e de atividade em uma célula são formadas por proteínas. Como então relacionar a sequência do DNA de um indivíduo com as proteínas necessárias à sua sobrevivência? Para tal, deve-se levar em consideração que as proteínas já haviam sido bem estudadas na época de descoberta do DNA e que, desde a década de 1950, já se sabia que cada proteína era constituída de sequências de aminoácidos.

É a sequência de nucleotídeos que determina a sequência dos aminoácidos que formará a proteína. A relação entre o DNA e os aminoácidos é denominada código genético, que é lido em trinca denominadas códons. Cada códon, ou seja, três nucleotídeos, corresponde a determinado aminoácido. Na síntese de uma proteína, os nucleotídeos de DNA são transcritos para RNA, e a informação contida no RNA gera a sequência de aminoácidos da proteína a ser traduzida.

Como existem quatro nucleotídeos diferentes, sua combinação dois a dois resultaria em uma probabilidade de 16 aminoácidos diferentes. Como existem 20 aminoácidos, essa probabilidade não era viável. Por isso, percebeu-se que, na verdade, cada aminoácido é gerado por uma trinca de nucleotídeos, o que resulta na probabilidade de formar 64 aminoácidos. Cada trinca, ou códon, gera um aminoácido; nos humanos, somente 61 códons geram aminoácidos – os três restantes são códons de terminação, não especificando nenhum aminoácido. Em eucariotos, os códons de terminação são, na sua grande maioria, UAA, UGA e UAG. O códon de iniciação, tanto em procariotos quanto em eucariotos, é o AUG, que codifica a metionina, sendo esse aminoácido,

portanto, o primeiro em todas as proteínas. Em *E. coli*, o códon GUG também pode ser utilizado.

Como se explica o fato de haver 61 códons diferentes e somente 20 aminoácidos? Em 1961, Marshall Nirenberg e colaboradores iniciaram pesquisas utilizando sistemas *in vitro* que permitiam fazer combinações e determinar quais aminoácidos correspondiam a códons específicos. Com base nesses achados, novos estudos descobriram os códons e seus aminoácidos correspondentes, percebendo-se que mais de um códon gerava um mesmo aminoácido, o que já havia sido previsto. Essa característica do código genético é chamada degeneração, por isso diz-se que o código genético é degenerado. A partir desses estudos, montou-se o quadro 2.1.

Quadro 2.1. Código genético.

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

Apenas a metionina (Met) e o triptofano (Trp) possuem somente um códon; todos os outros aminoácidos são gerados por mais de um códon. A degeneração do código genético permite que exista mais de um tRNA para o mesmo aminoácido, ou até mesmo que um mesmo tRNA possa parear com códons diferentes. Os dois processos são possíveis, pois, caso haja apenas

uma mudança no último nucleotídeo do códon, existe a ligação do tRNA. Ocorre o que se chama pareamento oscilante, pois o pareamento com a última base é menos rígido.

O código genético, além de degenerado, possui muitas outras características: sendo um código não ambíguo – ou seja, um códon codifica apenas um único aminoácido –, o código é universal, pois é igual em todos os seres, desde bactérias até os seres humanos, o que comprova ser essa forma de armazenamento de informações muito eficiente e ter-se mantido ao longo de milhões de anos de evolução.

2.4.2 Síntese proteica

A síntese proteica é um processo complexo que envolve grande variedade de macromoléculas. Ela se dá nos ribossomos do citoplasma e nos ribossomos aderidos ao retículo endoplasmático rugoso, cujo nome deriva da presença dos ribossomos ligados à sua membrana.

Após a transcrição nos procaríotos e a transcrição e o processamento nos eucaríotos, o mRNA maduro está pronto para transmitir, na forma de uma proteína, as suas informações. Quando se encontra no citoplasma, o mRNA é reconhecido pelo ribossomo. Isso decorre da interação do mRNA com o rRNA do ribossomo, mediante uma sequência no mRNA, denominada sítio de ligação dos ribossomos (RBS), de 30 a 40 nucleotídeos. Essa região contém o códon AUG de iniciação e uma sequência capaz de fazer pareamento com o rRNA, chamada sequência de Shine-Delgarno. Nos eucaríotos, também ocorre o reconhecimento do *cap* do mRNA, que auxilia no reconhecimento do início da tradução.

O ribossomo possui dois sítios diferentes: A e P. No sítio A, dá-se a ligação do aminoacil tRNA; no sítio P, ocorre a ligação tanto do aminoacil tRNA quanto do peptidil tRNA. Após o reconhecimento da sequência de Shine-Delgarno, o mRNA se liga ao ribossomo, ocorrendo a interação entre as duas subunidades ribossomais. A partir dessa interação, o códon de iniciação vai se localizar no sítio A do ribossomo, permitindo que se dê a interação

dele com o metionil tRNA. Após a ligação do tRNA ao códon do mRNA, o ribossomo desliza sobre o mRNA, de forma que o códon inicialmente localizado no sítio A passe a se localizar no sítio P, deixando o sítio A livre. Os ribossomos deslocam-se na direção 5'-3', sintetizando a proteína no sentido amino-terminal para carboxi-terminal. Quando o sítio A se encontra vazio, o ribossomo permite a interação de outro aminoacil tRNA, que conterà o anticódon específico do códon do mRNA que se encontra no sítio A. Assim, o segundo aminoácido é incorporado ao peptídeo em formação. A ligação entre os aminoácidos é realizada pela ação da peptidil transferase, que liga o aminoácido, ou a sequência já formada, ao aminoácido preso ao tRNA no sítio A, alongando a cadeia. O tRNA desativado, que ocupava o sítio P, deixa o complexo ribossômico para ser novamente ativado no citoplasma. Esse processo acontece incessantemente e aos poucos forma-se a estrutura primária da proteína que está sendo sintetizada. O último códon lido no mRNA deve ser um fator de término (UAA, UAG ou UGA), que não codifica nenhum aminoácido mas indica o final da síntese proteica.

A síntese proteica termina com a desativação do complexo ribossomo-mRNA, a desativação do tRNA e a formação da proteína em sua estrutura primária.

2.5 Métodos de extração de DNA

Os protocolos utilizados para a extração de DNA são muito parecidos, diferindo apenas em alguns detalhes relativos à natureza do material biológico de onde o DNA será extraído. Basicamente, consistem na lise celular e na purificação do DNA obtido.

A lise celular ocorre mediante o rompimento das membranas celulares, que libera o DNA contido no núcleo. Para esse processo, são utilizados detergentes que desestabilizam os lipídios das membranas. Nessa fase, é comum o uso de ácido etilenodiaminotetracético (EDTA), que quelar íons bivalentes que funcionam como cofatores de DNAses, inibindo-as. Após a liberação do material genético, seguem-se as fases de purificação do DNA.

A fase de purificação consiste em reações enzimáticas ou interações entre macromoléculas que favorecem a retirada de todos os contaminantes do DNA. As principais moléculas contaminantes são as proteínas, principalmente as histonas, por estarem intimamente ligadas ao DNA e ao RNA.

O RNA é geralmente retirado através do uso de RNAses que irão degradá-lo, permitindo a sua retirada, juntamente com moléculas menores, por meio de diálise.

Os contaminantes proteicos podem ser digeridos pela ação da proteinase K; porém, o agente desproteinizante mais comum é o fenol, um excelente desnaturador de proteínas. Por ser um composto orgânico, o fenol atua desnaturando as proteínas e transferindo-as, ao mesmo tempo, para uma fase orgânica, enquanto o DNA, que não interage com o fenol, permanece na fase aquosa. O clorofórmio também é um excelente desnaturante e pode atuar em conjunto com o fenol como estabilizador da junção entre as fases aquosa e fenólica. O uso da mistura fenol/clorofórmio diminui a quantidade de solução aquosa retida na fase orgânica, melhorando o rendimento da extração. Nessa fase de desnaturação, pode ocorrer formação de espuma por causa das proteínas desnaturadas. A formação de espuma é impedida com o uso de álcool isoamílico no processo.

Após a desnaturação das proteínas, a solução é centrifugada para a total separação das fases aquosa e orgânica. A fase aquosa contém o DNA; já as proteínas desnaturadas localizam-se em um anel intermediário entre as duas fases. O DNA contido na fase aquosa é precipitado com etanol. Na presença de altas concentrações de cátions monovalentes, o etanol induz uma mudança estrutural transitória nas moléculas dos ácidos nucleicos, ocasionando sua agregação e precipitação. Os cátions utilizados na precipitação, assim como pequenas moléculas orgânicas ainda presentes na preparação, podem ser removidos por lavagem com etanol 70%. Ao final do processo, o DNA é solubilizado em água Milli-Q autoclavada e livre de nucleases, sendo armazenado em baixa temperatura. A seguir, apresenta-se um protocolo de extração de DNA genômico de células sanguíneas.

2.5.1 Extração de DNA genômico de sangue humano

Preparação do sangue

1) Transferir 300 mL de sangue total para cada tubo de centrífuga. Adicionar 900 mL de tampão para lise de células sanguíneas em cada tubo e inverter o tubo cuidadosamente, a fim de homogeneizar a mistura. Incubar a mistura em temperatura ambiente por 10 minutos, invertendo-se o tubo de vez em quando.

2) Centrifugar os tubos em uma microcentrífuga por 20 segundos, em velocidade máxima, em temperatura ambiente.

3) Descartar todo o sobrenadante, deixando cerca de 20 mL.

4) Ressuspender o *pellet*¹ de células brancas na pequena quantidade de sobrenadante deixada em cada tubo. Juntar as células ressuspensas em um único tubo.

5) Centrifugar novamente e retirar todo o sobrenadante.

2.5.2 Extração do DNA genômico

1) Ressuspender as células sedimentadas em 567 mL de Tris EDTA (TE).

2) Adicionar 3 mL de proteinase K² (20 mg/mL) e 30 mL de SDS³ 10%.

3) Incubar a mistura a 37°C por 1 hora.

4) Adicionar 100 mL de NaCl 5 M.

5) Em seguida, adicionar 80 mL de solução CTAB⁴/NaCl (CTAB 10%, NaCl 1,5 M), misturando os componentes suavemente.

6) Incubar por 10 minutos, a 65°C.

7) Adicionar 1 V (~800 mL) de clorofórmio/álcool isoamílico (24:1).

8) Misturar as fases e centrifugar por 5 minutos, a 1.400 rpm, em microcentrífuga tipo Eppendorf.

9) Transferir a fase aquosa para um novo tubo.

¹ Precipitado de células.

² A proteinase K auxilia na lise celular.

³ Detergente que lisa as membranas celulares.

⁴ Brometo de trimetilamônio, um complexante de proteínas que auxilia na precipitação.

- 10) Adicionar 1 V de fenol/clorofórmio/álcool isoamílico (25:24:1).
- 11) Misturar as fases e centrifugar por 5 minutos a 1.400 rpm em microcentrífuga tipo Eppendorf.
- 12) Transferir a fase aquosa para um novo tubo.
- 13) Adicionar 0,6 V de isopropanol; inverter o tubo várias vezes.
- 14) Centrifugar por 5 minutos, a 1.400 rpm, em microcentrífuga tipo Eppendorf.
- 15) Lavar o material com etanol 70% gelado e centrifugar por 5 minutos, a 1.400 rpm, em microcentrífuga tipo Eppendorf.
- 16) Ressuspender o *pellet* em 50 mL de água Milli-Q.

2.6 Métodos de análise e quantificação de DNA

2.6.1 Eletroforese

A **eletroforese** é a principal técnica utilizada em biologia molecular para a análise de DNA. Ela permite separar, identificar e purificar os fragmentos de DNA quando não é possível utilizar outras técnicas, como os gradientes de centrifugação. É uma técnica rápida, que permite excelente resolução dos fragmentos de DNA.

A eletroforese é utilizada na separação de várias macromoléculas e consiste na migração das partículas – nesse caso, de fragmentos de DNA – através de um gel durante a aplicação de uma diferença de potencial. Assim, em um polo da eletroforese, é fornecida carga negativa e, no outro, positiva. Como os grupos fosfatos possuem carga negativa, o DNA corre, então, do polo negativo para o positivo. Essa “corrida” se dá através de uma malha, que auxiliará na separação das moléculas. Para formar essa malha, utilizam-se géis que, quando gelificados – ou seja, quando se tornam sólidos –, formam as malhas dentro de suas estruturas. Os géis mais utilizados em biologia molecular são os de **poliacrilamida** e os de **agarose**.

O gel de poliacrilamida é formado pela reação da acrilamida com a bisacrilamida. A acrilamida é uma molécula linear, e a bisacrilamida apresenta forma de “T”. Misturando essas duas moléculas, forma-se a malha. Diferentes relações entre as concentrações dessas moléculas permitem criar diferentes gradientes de separação. Para preparar o gel de poliacrilamida, devem-se misturar as duas substâncias formadoras, nas concentrações desejadas, e adicionar o Temed, que atua como catalisador da polimerização.

O gel de agarose é formado somente por agarose. A agarose é um polissacarídeo e forma uma malha que retém as moléculas durante a migração. Dependendo da concentração de agarose, há diferenças no gradiente de separação. Quanto maior a concentração, mais fechados são os buracos da malha e maior a retenção de moléculas. Para preparar um gel de agarose, faz-se a mistura do pó de agarose com uma solução tampão. Após realizar a fusão da agarose, coloca-se brometo de etídio, um corante capaz de se intercalar entre as bases do DNA e que fluoresce quando excitado com luz UV. Por sua capacidade de interação com as bases do DNA, o brometo de etídio é considerado um corante carcinogênico, devendo o seu manuseio ser efetuado com grande cuidado. Quando a mistura esfria, o gel está sólido. Esse endurecimento é feito diretamente na cuba de corrida da amostra. Antes da total solidificação, deve-se colocar um pente de eletroforese, pois ele cria poços que serão utilizados para a colocação das amostras.

A “corrida” de DNA se dá em conjunto com duas moléculas, adicionadas à amostra de DNA. Tais moléculas são denominadas monitores de corrida. Os monitores mais utilizados são o **azul de bromofenol** e o **xilenocianol**. O azul de bromofenol possui cor azulada e corre como um DNA pequeno, ao passo que o xilenocianol possui cor esverdeada e corre como um DNA maior.

Após a corrida, o DNA é visualizado mediante transiluminação de luz UV, que excita o brometo de etídio, o qual fluoresce com uma cor alaranjada (590 nm). Para o armazenamento das informações contidas no gel, existem aparelhos com câmeras fotográficas acopladas que permitem fotografar e transferir para o computador a fotografia do gel.

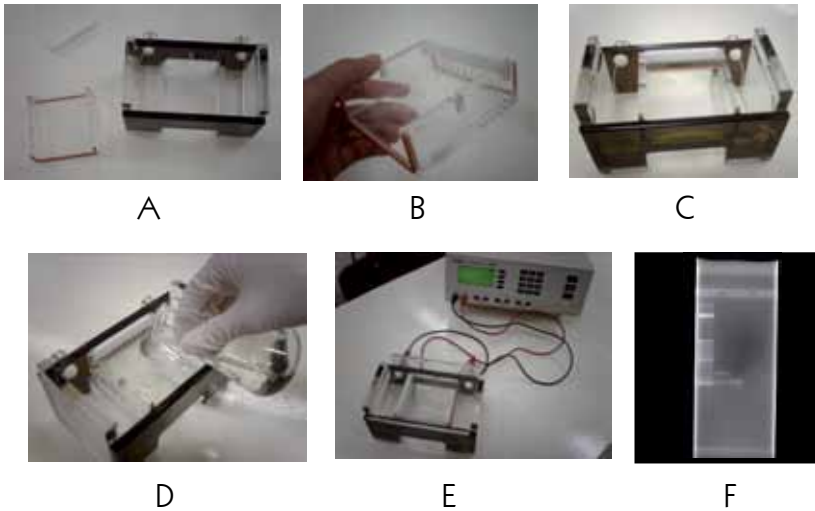


Figura 2.6. Etapas do preparo de uma eletroforese em gel de agarose: A) cuba de eletroforese horizontal com o pente e o suporte para preparo do gel; B) suporte com pente pronto para receber a agarose aquecida; C) cuba preparada para a gelificação do gel; D) agarose aquecida para preparo do gel; E) aparato de eletroforese preparado para a corrida; F) fotografia de um gel de eletroforese mostrando uma corrida de fragmentos de DNA (padrão de 50pb).

2.6.2 Espectrofotometria

A espectrofotometria é uma técnica muito utilizada na biologia, na química e na física e utiliza um aparelho denominado espectrofotômetro, instrumento que permite comparar a radiação absorvida por uma solução ao incidirmos radiação na amostra.

Cada substância é capaz de absorver uma quantidade de luz específica quando excitada com determinada quantidade de energia. A energia inserida eleva os elétrons da molécula para níveis energéticos mais altos. A energia necessária para essa elevação de nível eletrônico é específica de cada molécula e vai variar segundo as características das ligações presentes na mesma.

Em decorrência das ligações duplas presentes nas bases purínicas e pirimidínicas do DNA e do RNA, esses, quando são excitados com uma luz no

comprimento de onda de 560 nm, emitem fluorescência e podem ser detectados por **espectrofotometria**. Quanto maior a quantidade de DNA e RNA presente em determinada amostra, maior será a absorvância lida.

Para diminuir ao máximo a interferência das proteínas que ainda podem estar presentes na amostra, deve-se fazer uma leitura a 280 nm – comprimento de onda em que as ligações peptídicas das proteínas fluorescem. Para avaliar o nível de pureza dos ácidos nucleicos contidos em determinada extração, fazem-se leituras a 260 e 280 nm, analisando-se a relação 260/280. Se o resultado obtido estiver entre 1,8 e 2,0, a amostra se encontra em boas condições de análise, contendo pouca interferência de proteínas. Caso a leitura seja menor do que 1,6, será necessário novo processo de purificação dos ácidos nucleicos. Essa relação para a análise de pureza não é confiável se a amostra estiver contaminada com fenol, pois a mistura de água e fenol fluoresce a 270 nm, um comprimento de onda muito próximo dos 280 nm utilizados na medição. Para utilizar tal relação, deve-se garantir que todo o fenol tenha sido retirado da amostra no momento de seu preparo.

A espectrofotometria não indica se a amostra contém somente DNA, somente RNA, ou ambos, nem se eles estão íntegros, mas proporciona uma boa noção da quantidade na amostra, em mg, do DNA e do RNA. Para se obterem dados quantitativos, é preciso levar em consideração algumas relações matemáticas. No caso de um DNA de fita simples, como uma sonda ou um *primer*, cada unidade de absorvância a 260 nm equivale a 32,7 mg de DNA. Numa amostra de extração de DNA genômico, estima-se que em uma absorvância igual a 1 haja uma concentração de 50 mg de DNA/mL.

Em algumas preparações, a quantidade de DNA pode ser pequena e limitar bastante o uso da espectrofotometria como técnica de quantificação. Nesses casos, a técnica escolhida é a **fluorimetria**. A fluorimetria detecta nanogramas de DNA. Nesse método, a amostra é incubada com o fluorocromo Hoechst 33258 que se liga à *minor groove* do DNA. Essa característica do corante permite que ele core somente DNA íntegros e não core o RNA. É, portanto, uma técnica seletiva. Para efetuar a análise, deve-se construir uma curva de calibração do aparelho e avaliar a amostra desconhecida, interpolando os resultados lidos com a reta a fim de encontrar o valor correspondente à concentração do DNA desconhecido.

2.7 Endonucleases de restrição

As endonucleases de restrição são enzimas capazes de reconhecer e cortar o DNA em locais precisos, permitindo a obtenção de fragmentos específicos de DNA. Na década de 1950, cientistas observaram que algumas cepas de *E. coli* eram resistentes à infecção por bacteriófagos. Avanços nos estudos indicaram, e posteriormente confirmaram, que a resistência aos parasitas ocorre pela existência de um sistema de enzimas na parede da bactéria que reconhece e elimina seletivamente o DNA dos bacteriófagos.

As endonucleases de restrição pertencem a um grupo maior de enzimas, denominadas nucleases, que em geral atuam clivando as ligações fosfodiéster – ligantes de nucleotídeos adjacentes no DNA. O corte na molécula de DNA é feito mediante o reconhecimento, por parte das enzimas, de sequências específicas de 4 a 8 pares de base (pb). Essas sequências de reconhecimento variam de acordo com a enzima; uma vez identificadas, é feito um corte duplo na molécula de DNA: um em cada fita.

Existem dois tipos distintos de clivagem do DNA: 1) cortes em um mesmo eixo de simetria, gerando extremidades abruptas; e 2) cortes simétricos, que apresentam sequenciais de bases complementares, porém sem apresentar o mesmo eixo, gerando extremidades coesivas. O quadro 2.2 mostra os dois tipos de cortes e as extremidades formadas.

Quadro 2.2. Tipos de clivagem do DNA.

Extremidades coesivas (corte pela enzima EcoRV)	Extremidades abruptas (corte pela enzima BamHI)
5' G↓GATCC 3' 3' CCTAG↓G 5'	5' GAT↓ATC 3' 3' CTA↓TAG 5'

Mais de cem tipos de enzimas de restrição foram identificados e purificados. A identificação de cada uma dessas enzimas é feita pela abreviação do nome do microrganismo do qual a enzima foi isolada, seguida de algarismos romanos (ou outras letras) que representam a ordem da descoberta ou a linhagem à que a bactéria utilizada pertence. Por exemplo, a enzima Hind III foi

isolada da linhagem d III da bactéria *Haemophilus influenzae*. O quadro 2.3 traz a relação de algumas das principais enzimas de restrição, o organismo de origem e o local de clivagem.

Quadro 2.3. Relação das principais enzimas de restrição.

Enzima	Organismo fonte	Local de clivagem
BamHI	<i>Bacillus amyloliquefaciens</i> H	5' G ↓ GATCC 3'
BglII	<i>Bacillus globigii</i>	5' A ↓ GATCT 3'
EcoRI	<i>Escherichia coli</i> RY13	5' G ↓ AATTC 3'
EcoRV	<i>Escherichia coli</i> R321	5' GAT ↓ ATC 3'
HaeIII	<i>Haemophilus aegyptius</i>	5' GG ↓ CC 3'
HindIII	<i>Haemophilus influenzae</i> Rd	5' A ↓ AGCTT 3'
HpaII	<i>Haemophilus parainfluenzae</i>	5' C ↓ CGG 3'
NotI	<i>Nocardia otitidiscaviarum</i>	5' GC ↓ GGCCGC 3'
PstI	<i>Providencia stuartii</i> 164	5' CTGCA ↓ G 3'
SmaI	<i>Serratia marcescens</i> Sb	5' CCC ↓ GGG 3'

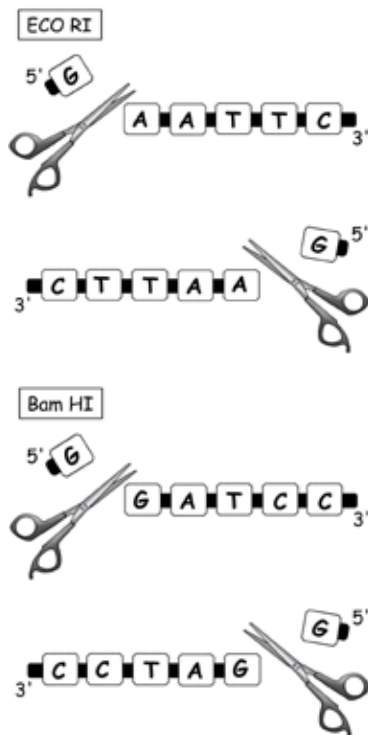


Figura 2.7. Funcionamento das endonucleases de restrição.

2.8 Hibridação molecular

O processo de hibridação do DNA é um dos principais processos para detecção de um gene em particular ou de um segmento específico de ácido nucleico. Apesar de existirem algumas variações do método clássico de hibridação, os procedimentos básicos desses métodos se assemelham em muitos pontos.

Na técnica, um papel de nitrocelulose é prensado numa placa de ágar com colônias individuais de bactérias oriundas de uma biblioteca, contendo cada bactéria um DNA recombinante diferente. Ao tocar a placa com o papel, algumas bactérias ficam aderidas, fornecendo uma “cópia” da placa. O papel é, então, tratado com substâncias alcalinas que provocam a destruição das células e a desnaturação do DNA presente, o qual permanece aderido ao papel no local da colônia de origem.

Após essa etapa, é adicionada ao papel a sonda de DNA marcada com nucleotídeos radioativos, que somente vai ligar-se ao DNA complementar. Como as moléculas de DNA foram desnaturadas pela adição do álcali, elas se encontram no formato de fita simples, e isso permite o pareamento com o DNA marcado (anelamento). O papel é lavado para a retirada da sonda de DNA que não foi anelada. Ao final, o DNA hibridado (que anelou com a sonda de DNA) pode ser detectado por autorradiografia.

2.8.1 Técnicas de hibridação

Os métodos de eletroforese em gel são fundamentais para separar moléculas de DNA de diferentes tamanhos. Assim como as proteínas, os ácidos nucleicos apresentam grupamentos químicos com cargas elétricas. Porém, enquanto as proteínas apresentam grupamentos positivos e negativos, os ácidos nucleicos só apresentam cargas negativas, provenientes dos radicais fosfatos presentes em sua molécula.

Fragmentos de DNA contendo menos de 1.000 nucleotídeos são separados por géis de poliacrilamida. Entretanto, como os poros desses géis são muito pequenos para permitir a passagem de moléculas maiores de DNA, utiliza-se a eletroforese em gel de agarose, um polissacarídeo extraído de algas

marinhas. A eletroforese em gel de agarose é o método mais utilizado nas análises de fragmentos de DNA e RNA. Cromossomos inteiros, contendo milhões de nucleotídeos, podem ser separados por uma técnica de eletroforese em gel de agarose intitulada eletroforese em gel de campo pulsado (PFGE) (do inglês *pulsed field gel electrophoresis*).

As amostras formam bandas invisíveis, que ficam situadas em diferentes posições do gel de agarose ou poliacrilamida, e é necessário que as bandas de DNA sejam coradas ou marcadas. Um método de boa sensibilidade para corar o DNA é a utilização do brometo de etídio. Outro método de detecção ainda mais sensível consiste na incorporação de um radioisótopo nas moléculas de DNA. Um fragmento de restrição contendo uma sequência específica de bases pode ser identificado mediante hibridação a um filamento complementar de DNA marcado radioativamente. O radioisótopo ^{32}P , por sua capacidade de se incorporar ao fosfato do DNA, é muito utilizado para isso. Um fragmento de interesse de DNA é separado de um conjunto de moléculas por eletroforese em agarose, desnaturado para formar um único filamento e, em seguida, transferido para um suporte de nitrocelulose, onde é exposto a uma sonda marcada com ^{32}P complementar à sua sequência. O fragmento de DNA contendo a sequência é visualizado por autorradiografia. Essa técnica foi desenvolvida pelo bioquímico britânico Edwin Southern, sendo chamada *Southern blotting*. O *Northern blotting* é utilizado para detectar RNA e segue o mesmo protocolo do *Southern blotting*. A transferência de *Western blotting* é uma técnica que detecta proteínas por coloração, por meio do emprego de anticorpos específicos.

2.9 Clonagem de DNA

2.9.1 Clonagem

O fenômeno da clonagem é aquele no qual um organismo, ou uma célula, é formado a partir de outro por meio de um tipo de reprodução assexuada, mantendo-se, em geral, seu conjunto de genes.

A existência dos clones é algo comum na natureza. Muitos protozoários, bem como as bactérias e alguns tipos de fungos, reproduzem-se por clonagem. E até mesmo as nossas células, ao realizarem o tipo de divisão celular conhecida como mitose, efetuam um processo de clonagem. Com os avanços da biotecnologia, os cientistas passaram a ser capazes de clonar organismos inteiros. A ovelha Dolly foi o primeiro deles.

Por conta do desenvolvimento de novas técnicas de engenharia genética, pode-se isolar um gene específico (ou um conjunto de genes) de um organismo e introduzi-lo em outro de espécie diferente, permitindo sua multiplicação junto com a reprodução do organismo receptor.

2.9.2 Clonagem de DNA

Na biologia molecular, clonagem consiste na produção de cópias exatas de genes ou de grupo de genes (fragmentos de DNA). Nessa técnica, o DNA é retirado de uma matriz (uma suspensão celular ou um fragmento de tecido) e digerido por endonucleases de restrição. Os fragmentos gerados contêm os genes de interesse, que são ligados a um vetor de clonagem, geralmente um plasmídeo.

Um vetor de clonagem é uma estrutura de DNA com capacidade de se introduzir em células bacterianas, um processo conhecido como transfecção. Um dos vetores de clonagem mais utilizados são os plasmídeos. Os plasmídeos são moléculas circulares de DNA presentes em bactérias nas quais geralmente se encontram os genes responsáveis pela resistência a antibióticos. Entretanto, é importante lembrar que existem vários outros vetores de clonagem, bem como diferentes formas de se fragmentar e clonar o DNA.

O primeiro passo da clonagem é a extração do DNA de uma célula doadora qualquer e também do plasmídeo que servirá de vetor. A extração de DNA apresenta uma série de protocolos diferentes, sendo o fundamento básico acessar o DNA presente em determinado compartimento celular para depois separá-lo dos demais componentes celulares (lipídios, proteínas, RNA etc.).

A principal diferença entre a extração de DNA genômico e a extração do plasmídeo é a necessidade, na extração desse último, de se separar o plasmídeo do DNA bacteriano. Para que essa separação seja possível, utiliza-se a característica do plasmídeo de ser uma molécula muito menor do que os fragmentos de DNA cromossômico da bactéria.

Certa quantidade de substância alcalina (básica) é adicionada na solução de lise celular. Essa substância provoca a elevação do pH, que acaba por desnaturar as moléculas de DNA (plasmidial ou cromossômico). Em seguida, é adicionado ácido acético, que neutraliza o álcali, e acetato de potássio. Após a neutralização, os fragmentos de DNA tendem a refazer a molécula original. Entretanto, o sal (acetato de potássio) provoca a precipitação de todas as moléculas que não são muito solúveis em água. Os fragmentos grandes formados pela desnaturação do DNA cromossômico não conseguem se renaturar com rapidez suficiente e, por serem insolúveis, precipitam junto com os restos celulares e proteínas da bactéria. O DNA plasmidial, ao contrário, por causa de seu menor tamanho, regenera-se rapidamente, com o que aumenta a sua solubilidade na água e evita a sua precipitação. Dessa forma, ao se recolher o sobrenadante após uma centrifugação, teremos essencialmente o DNA plasmidial. Uma vez que se tenham os dois tipos de DNA (da célula doadora e o plasmidial), a fase seguinte consiste em clivar ambos com a enzima de restrição escolhida, misturando-os depois em uma solução contendo enzimas ligases, o que permitirá o pareamento dos fragmentos e a reconstrução das ligações fosfodiéster em cada fita de DNA. Essa união entre o fragmento de DNA (inserto) e o plasmídeo (vetor) forma o que chamamos **DNA recombinante**.

O DNA recombinante deve ser introduzido na célula hospedeira para que ela possa replicar-se e, conseqüentemente, replicar o(s) gene(s) do inserto. Isso é obtido, em um processo passivo, por meio das membranas plasmáticas das bactérias previamente tratadas com solução de cloreto de cálcio, ou ativo, pelo uso de descargas elétricas que abrem os poros das células, processo denominado eletroporação. Veremos ambos os métodos a seguir.

2.10 Transformação bacteriana

O processo de transformação bacteriana consiste na introdução de um vetor dentro da bactéria. Esse processo é chamado transformação por causa das novas características que podem ser adquiridas pela bactéria com a introdução do vetor (como a resistência a determinados antibióticos). Existem basicamente dois processos distintos: a eletroporação e a transformação com cloreto de cálcio.

A eletroporação é a união das bactérias e dos vetores em um único tubo, submetido a uma descarga elétrica com o intuito de provocar a desestabilização da membrana plasmática da bactéria e permitir a entrada do vetor. Após essa etapa, as bactérias são transferidas para um meio de cultura e incubadas a 37°C, a fim de que possam se recuperar do choque recebido.

Na transformação com cloreto de cálcio, misturam-se, em um único tubo, as bactérias, o vetor e o cloreto de cálcio. O cloreto de cálcio dissocia-se na solução, liberando íons cálcio (cátions) que irão neutralizar a carga negativa do DNA. Com o choque térmico, a membrana da bactéria desestabiliza (ele tem, portanto, o mesmo papel do choque elétrico), e o DNA, com sua carga neutralizada pelos íons de cálcio, entra facilmente na bactéria.

Após a transformação, as bactérias são colocadas em meios de cultura e incubadas para que possam se multiplicar e formar colônias. A distinção entre as bactérias que possuem o vetor e aquelas que não possuem é feita mediante as características conferidas pelos plasmídeos. Se o plasmídeo confere resistência a determinado antibiótico, o plaqueamento das bactérias em um meio com esse antibiótico selecionará somente aquelas que são resistentes (transformadas), matando as demais.

2.11 Biblioteca genômica

A biblioteca genômica é uma coleção de fragmentos de DNA que representa o genoma inteiro de um organismo ou de uma célula, coleção adquirida pelo processo de clonagem. Dependendo da fonte de DNA utilizada, há uma variedade de formas de bibliotecas genômicas. Uma das formas mais comuns

é aquela na qual a biblioteca é formada pela clivagem do genoma inteiro de determinado organismo em milhares de fragmentos que serão clonados por inserção em determinado vetor de clonagem.

Primeiramente, o DNA a ser clonado é parcialmente digerido por uma endonuclease de restrição, gerando fragmentos grandes, mas com tamanhos diferentes. Esses fragmentos de DNA de tamanhos diferentes são separados por eletroforese em gel, para se separar um tamanho específico de fragmento, que será escolhido em razão do vetor a ser utilizado – por exemplo, um bacteriófago (fago λ). Após a inserção do fragmento no vetor, promove-se a transformação bacteriana, e o lisado resultante apresenta fragmentos de DNA do organismo escolhido em um grande número de fagos, o que garante que praticamente todo o genoma esteja ali representado. Esses fagos recombinantes constituem a biblioteca genômica. O organismo pode ser propagado milhares de vezes, o que permite a utilização da biblioteca por longos períodos.

2.12 A reação em cadeia da polimerase (PCR)

O Projeto Genoma Humano (PGH) permitiu avanços tecnológicos que ampliaram a capacidade de acesso a informações relativas à sequência de genes de um organismo. Enquanto o sequenciamento do genoma completo de um organismo apresenta como etapa intermediária a criação de uma ou mais bibliotecas, a clonagem de um gene pode ser feita de maneira rápida, sem ajuda da biblioteca, visto que a sequência genômica já está completa. Se soubermos pelo menos uma parte da sequência do gene a ser clonado, podemos gerar milhares de cópias pelo processo conhecido como reação em cadeia da polimerase (PCR) (do inglês *polymerase chain reaction*).

A reação é preparada a partir de uma solução contendo a amostra de DNA que se quer amplificar, uma DNA polimerase estável ao calor (Taq polimerase), quatro tipos de deoxinucleotídeos trifosfatados (nucleotídeos com as bases nitrogenadas A, T, C e G, chamados de dNTPs), além de dois oligonucleotídeos chamados *primers*. Os *primers* são oligonucleotídeos complementares às duas extremidades (5' e 3') do fragmento a ser amplificado.

No início, as enzimas utilizadas na PCR não resistiam às altas temperaturas necessárias para a desnaturação da molécula de DNA, o que implicava a adição de grandes quantidades de enzimas a cada novo ciclo de amplificação. Com a descoberta da enzima termostável Taq DNA polimerase, isolada da bactéria de fontes termais *Thermus aquaticus*, resistente a temperaturas de até 117°C (com temperatura ótima de 72°C), foi possível a realização da reação sem que houvesse necessidade de adição de novas quantidades de enzimas. A utilização de Taq polimerase e a criação de termocicladores (equipamento que automatiza os ciclos de temperatura) mais eficientes tornaram a técnica mais fácil e barata.

Após o preparo da mistura para a reação, a amostra é colocada em um termociclador, onde será submetida a ciclos térmicos que se alternam entre temperaturas elevadas e temperaturas mais baixas. A alta temperatura causa a desnaturação da molécula de DNA, desfazendo as pontes de hidrogênio e separando as duas fitas complementares. Por outra parte, a diminuição da temperatura permite a hibridação das extremidades do fragmento de DNA com os oligonucleotídeos complementares (*primers*). A enzima Taq polimerase utiliza os nucleotídeos para polimerizar uma nova sequência de DNA complementar ao fragmento de interesse. O DNA original é utilizado como molde para a construção da nova sequência – o molde é chamado de *template*. Um novo ciclo é iniciado com a desnaturação dos fragmentos e a síntese de novos fragmentos em temperaturas baixas. Dessa maneira, a cada novo ciclo, aumenta-se a quantidade de cópias do fragmento de DNA a ser amplificado. Ao final da reação, o produto apresenta grande quantidade de DNA amplificado junto com o DNA original do início.

Em outras palavras, podemos separar as fases da reação de polimerase em três etapas: separação dos filamentos do DNA (desnaturação), hibridação de *primers* (anelamento) e síntese de DNA (extensão).

A separação dos filamentos ocorre por desnaturação do DNA. Nessa etapa, os dois filamentos da molécula de DNA são separados por aquecimento da solução a temperaturas próximas de 95°C.

A hibridação de *primers* (oligonucleotídeos) consiste no anelamento ou pareamento de cada par de *primers* a um filamento de DNA. A solução é

abruptamente resfriada para a temperatura específica de anelamento do *primer* utilizado, permitindo sua ancoragem e a delimitação da região específica a ser amplificada pela DNA polimerase.

A síntese de DNA ocorre com o aquecimento da solução a 72°C, temperatura ótima para a Taq DNA polimerase, que faz a extensão dos dois *primers* de oligonucleotídeos. A Taq polimerase sintetiza um novo fragmento de DNA a partir da fita molde original.

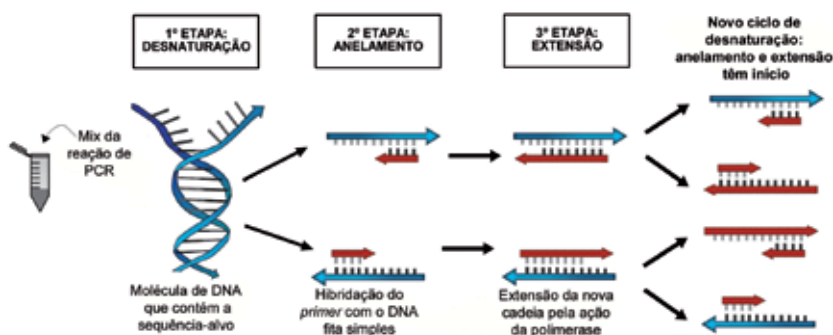


Figura 2.8. Representação das etapas sequenciais da reação em cadeia da polimerase (PCR).

2.13 Sequenciamento de DNA

Métodos de sequenciamento do DNA são importantes para a determinação das sequências de nucleotídeos de um fragmento de DNA, permitindo o sequenciamento completo de dezenas de milhares de genes. Vários organismos tiveram seus genomas completamente decifrados – inclusive o homem, no Projeto Genoma Humano (PGH).

O Projeto Genoma Humano teve início nos Estados Unidos em 1990. Seus principais objetivos foram os de identificar e mapear os genes dos 23 pares de cromossomos humanos, determinar a sequência de todas as bases do nosso genoma, armazenar essas informações em bancos de dados para analisá-las e desenvolver métodos eficientes para usar essas informações na biologia e na medicina.

No dia 6 de abril de 2000, a firma americana Celera Genomics anunciou a obtenção da sequência do genoma humano, e a publicação da análise da sequência do genoma humano foi feita em fevereiro de 2001. Por meio do PGH, foi possível descobrir que o genoma humano possui entre 30 mil e 50 mil genes. Partindo-se da ideia de que cada nucleotídeo do DNA é representado por uma letra correspondente à sua base nitrogenada (A, T, C ou G), seria possível escrever um “livro genômico humano” com aproximadamente 840 mil páginas.

Inicialmente, os processos de sequenciamento de DNA eram realizados manualmente. Isso consumia muito tempo, dinheiro e esforços dos pesquisadores envolvidos. Atualmente, com o avanço na tecnologia de sequenciamento, é possível a leitura de 500 mil nucleotídeos em apenas um dia.

2.13.1 Sequenciamento de Sanger

No ano de 1977, Frederick Sanger, na Inglaterra, desenvolveu um método que se tornou a base para todo o sequenciamento moderno de DNA. O método consiste na incorporação aleatória de dideoxinucleotídeos trifosfatos (ddNTPs), em uma fita de DNA, pela enzima DNA polimerase. As ddNTPs, ao contrário dos deoxinucleotídeos trifosfatados (dNTPs), não possuem em sua estrutura uma hidroxila na posição 3'. Assim, a síntese da fita de DNA é paralisada sempre que a DNA polimerase incorpora uma ddNTP na nova fita.

A primeira etapa da reação consiste na desnaturação da molécula de DNA, formando fitas simples que servirão de molde para a DNA polimerase. É necessária a presença de sequências iniciadoras, os *primers*, para que a DNA polimerase possa começar a atuar. Além disso, a solução deve conter baixas concentrações de ddNTP e altas concentrações de dNTP. No decorrer da reação, a DNA polimerase utiliza os dNTPs para a síntese na nova fita de DNA até que, aleatoriamente, utiliza uma ddNTP, a qual, por não possuir uma hidroxila na posição 3', interrompe a polimerização da nova cadeia.

A inclusão de marcadores fluorescentes de cores diferentes para cada ddNTP permite a identificação da cadeia truncada (que não foi capaz de

terminar a polimerização em virtude da adição da ddNTP), independentemente do tamanho do fragmento. Os fragmentos são separados por eletroforese em gel de poliacrilamida. Existem equipamentos (sequenciadores automáticos) capazes de distinguir os quatro tipos de ddNTP existentes em razão da captação de sua fluorescência. A ordem em que os diferentes fragmentos passam pelo detector de fluorescência indica a sequência dos nucleotídeos da cadeia complementar ao DNA molde, determinando assim a sequência original.

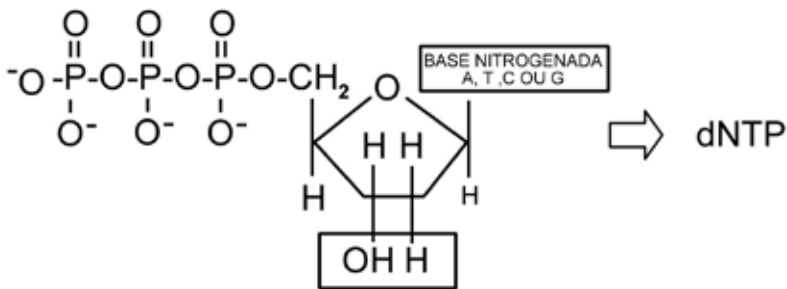


Figura 2.9. Desoxirribonucleosídeos trifosfatos (dNTP).

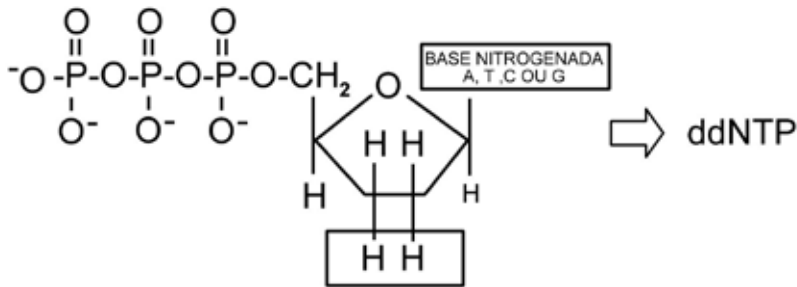


Figura 2.10. Didesoxirribonucleosídeos trifosfatos (ddNTP).

2.13.2 Exemplo de processo metodológico de análise de ácidos nucleicos

Muitos são os métodos utilizados na manipulação e análise de ácidos nucleicos. A seguir, apresentamos um exemplo de processo metodológico muito

utilizado atualmente, denominado polimorfismo no comprimento dos fragmentos de restrição (RFLP, do inglês *restriction fragment length polymorphisms*).

2.13.2.1 Polimorfismo no comprimento dos fragmentos de restrição

Apesar dos mecanismos bioquímicos existentes para que erros sejam evitados durante a duplicação do DNA, existe uma taxa de mutação constante no genoma de todas as espécies. Essas mutações, nem sempre perceptíveis — pois podem estar localizadas em regiões não codificantes (íntrons ou regiões intergênicas) —, são chamadas polimorfismos e podem ser de grande importância, por exemplo, no entendimento da evolução de determinados genes e na detecção de susceptibilidade genética a determinadas doenças.

A técnica de RFLP baseia-se na hidrólise do DNA com enzimas de restrição específicas e posterior separação dos fragmentos gerados pela digestão por eletroforese. Dessa maneira, busca-se reconhecer modificações do DNA que possam estar presentes dentro da sequência de reconhecimento das enzimas de restrição. Então, um indivíduo que apresente uma modificação no genoma que impossibilite a clivagem por determinada enzima de restrição pode ser identificado por meio da comparação com outro da mesma espécie e que não apresente essa modificação, o que não impossibilita a clivagem do DNA. Com a realização da eletroforese, pode-se comparar o tamanho das amostras de DNA e inferir sobre uma possível alteração ou não. O RFLP pode ser aplicado em DNA plasmidial ou DNA cromossômico.

- **Eletroforese**

Primeira etapa: extração de DNA

A retirada do DNA da matriz escolhida pode seguir protocolos variados. Geralmente o DNA genômico é extraído de uma camada de células brancas separadas (*buffy-coat*) ou do sangue total.

Podem ser utilizados *kits* comerciais para o processo de extração. Seja como for, o processo baseia-se em dois procedimentos básicos: rompimento das membranas celulares (geralmente com detergentes), permitindo a liberação

do DNA; e tratamentos químicos ou enzimáticos, como RNA, proteínas e outras moléculas, para purificar a amostra de contaminantes.

Segunda etapa: amplificação

Após a extração do DNA da matriz escolhida, é preciso amplificar o fragmento específico do material genético de interesse. Para isso, utiliza-se a técnica de PCR anteriormente descrita. Em geral, o procedimento é o mesmo, contendo apenas algumas modificações.

A escolha dos *primers* responsáveis por selecionar a região correta do genoma escolhido a ser amplificada dependerá do produto desejado ao final da amplificação. Da mesma maneira, as condições de temperatura do termociclador dependem das características físico-químicas dos *primers* utilizados. Após essa etapa, o analisador terá uma quantidade suficiente de DNA, e ele poderá ser quantificado por métodos espectrofotométricos.

Terceira etapa: purificação do DNA

A etapa de purificação consiste na eliminação de substâncias indesejáveis à análise, deixando-se somente o ácido nucleico de interesse. A escolha do processo de purificação depende de vários fatores, como as substâncias “contaminantes” de sua amostra, o grau de pureza de que sua análise necessita e a disponibilidade financeira e temporal do projeto desenvolvido. Alguns protocolos de purificação de ácidos nucleicos podem ser muito rápidos. No entanto, dependendo da complexidade da amostra analisada, o método mais rápido pode não ser o mais adequado. O mesmo ocorre em relação ao custo.

Quarta etapa: digestão

O tratamento com enzimas de restrição é comumente chamado digestão. A digestão gera fragmentos de tamanhos específicos de pares de base, que podem ser identificados por eletroforese em gel de agarose. Aqueles fragmentos de menor peso molecular migram mais rapidamente do que os de maior peso molecular.

Quinta etapa: eletroforese

Após as etapas de amplificação, purificação e digestão, são realizadas eletroforeses, para a análise do produto, em gel de agarose. Depois da eletroforese, o gel é corado com brometo de etídio. O tamanho dos fragmentos (bandas) é determinado pela comparação com um DNA-padrão colocado no mesmo gel em que estão as amostras. Esse DNA-padrão é comercializado por empresas especializadas, e contém moléculas que geram bandas de tamanhos específicos de pares de bases. O intervalo entre as bandas pode representar um salto de 50 ou de 100 pares de bases, dependendo do padrão adquirido.

2.14 O DNA e as ciências forenses

Cada indivíduo possui seu DNA próprio, com exceção de gêmeos idênticos. Mesmo em indivíduos da mesma espécie, o DNA é muito semelhante em vários aspectos, pois são necessárias sínteses das mesmas proteínas para as atividades de um organismo de mesma espécie.

Como descrito anteriormente, cada proteína possui um gene no DNA responsável por sua codificação. Assim, no momento da tradução, aquela sequência do mRNA obtida a partir de nucleotídeos complementares do DNA gerará a proteína necessária para o perfeito funcionamento da célula. Porém, em eucariotos, vale ressaltar a presença de íntrons, isto é, pedaços do DNA que não codificam nenhuma proteína e por isso não são utilizados. Seria lógico pensar que, caso houvesse uma mutação nesses pedaços do DNA, não haveria alteração na vida do indivíduo, pois esse íntrons não são utilizados. O que ocorre é exatamente isso. Os éxons são extremamente conservados para que se possa preservar o proteoma do indivíduo, mas os íntrons são passíveis de variadas mutações.

Esse fato foi constatado por Alec Jeffreys, um geneticista inglês. A partir de seus estudos com a mioglobina, ele percebeu que havia um pequeno trecho do DNA que se repetia continuamente. Jeffreys também percebeu que esse trecho não fazia parte somente do DNA da mioglobina, mas do genoma. Em 1980, outros dois geneticistas, Ray White e Arlene Wyman, perceberam

que o número de repetições do trecho descrito por Jeffreys variava de indivíduo para indivíduo.

O trecho descrito por Jeffreys era uma sequência curta de quatro nucleotídeos – GATA –, que se repetiam de forma variável em cada indivíduo. Surgiu, assim, a técnica de *DNA fingerprinting* (impressão digital de DNA), pois a análise do número de repetições desses trechos permite individualizar uma pessoa. Tais trechos foram denominados microssatélites ou STRs (do inglês *shot tandem repeats*, ou seja, repetições curtas consecutivas). É a análise dos STRs que permite, atualmente, correlacionar amostras de DNA a amostras de determinado indivíduo, sendo importante instrumento técnico na medicina forense.

A técnica de *DNA fingerprinting* se inicia com a extração do DNA de amostras coletadas da vítima ou de algum suspeito. Qualquer tipo de fluido biológico ou amostra de tecido que contenha células pode ser utilizado para a análise de DNA. As amostras mais utilizadas atualmente são sangue, sêmen (no caso de crimes sexuais), células do epitélio da bochecha, amostras de biópsias de tecidos moles e ossos, e pelos e cabelos que contenham células do bulbo capilar, entre outras. O DNA obtido é tratado pela técnica RFLP, método que se baseia em clivagens feitas por enzimas de restrição (enzimas que clivam o DNA em determinados pontos específicos), gerando fragmentos de DNA de diferentes tamanhos e sequências específicas. Em seguida, esses fragmentos são separados por eletroforese, marcados e analisados. O perfil formado será específico de cada indivíduo, dado que os fragmentos formados pelas enzimas de restrição, os microssatélites, têm quantidade variável de indivíduo para indivíduo.

Essa técnica revolucionou a ciência forense, pois permitia correlacionar com precisão determinado vestígio de algum suspeito. Sua única limitação era a grande quantidade de DNA necessária, só conseguida em algumas amostras. Com a descoberta da técnica de PCR, esses fragmentos puderam ser amplificados, ou seja, a quantidade de DNA obtida podia ser aumentada, permitindo que amostras com pequenas quantidades de DNA também fizessem parte do rol de amostras coletadas na cena do crime. Atualmente, uma quantidade de apenas vinte células é suficiente para a identificação do DNA. A quantidade de DNA obtida é de cerca de 1 a 20 ng.

Todo exame de DNA utiliza um cálculo que determina a raridade da combinação entre perfis encontrados nas amostras. Esse cálculo determina a probabilidade de o DNA da amostra combinar com o de determinado indivíduo. Esse cálculo é baseado na comparação do padrão de polimorfismo – presença numa população de uma ou mais formas de determinado gene ou de determinada sequência do genoma – com bancos de dados de uma população específica. Para essa análise, deve-se conhecer com que frequência a combinação de fragmentos ocorre no grupo onde se insere o indivíduo suspeito, para estimar assim a confiabilidade do exame.

Atualmente, a identificação por DNA é utilizada pela perícia técnica para elucidar diversos acontecimentos. É uma técnica fundamental na identificação de suspeitos de crime sexual, pois é capaz relacionar o DNA encontrado no sêmen deixado na vítima com amostras de DNA de suspeitos. Em catástrofes, a análise de DNA permite a identificação de corpos carbonizados ou em estado muito avançado de decomposição. Para essa análise, utiliza-se o DNA mitocondrial, que possui características oriundas somente da mãe, mas que também, quando comparado ao DNA genômico, apresenta grande estabilidade. Também é possível identificar cadáveres mutilados e partes de órgãos de determinados indivíduos, e relacionar objetos a determinado crime. Uma faca contendo sangue pode ser fonte de identificação do DNA do agressor, que, ao segurá-la sem o uso de luvas, transfere elementos celulares para o cabo do instrumento.

Além da identificação, a análise de DNA também é capaz de estabelecer a paternidade de um indivíduo, técnica utilizada fora do âmbito criminal. O teste de paternidade é possível porque o DNA de um indivíduo é formado por alelos do DNA da mãe e alelos do DNA do pai. Assim, os STRs têm perfis que se assemelham, em parte, aos alelos do pai e, em parte, aos alelos da mãe. Nesse tipo de teste, procede-se à análise dos DNAs do filho, da mãe, do suposto pai e de uma mistura entre o DNA do filho e o do suposto pai. O padrão formado no gel de agarose do DNA do filho deve ter relação, em parte, com o DNA de seus pais. Caso essa relação não ocorra, a paternidade, e até mesmo a maternidade (procedimento utilizado no caso de troca de bebês em maternidades) não serão confirmadas.

Apesar de muito útil, essa técnica possui várias desvantagens. As amostras utilizadas devem ser coletadas de forma tal a não se degradarem até a chegada ao laboratório e não se misturarem com outras amostras. Além disso, é preciso impedir as trocas nos materiais coletados pelo perito. Por isso, é fundamental a preservação correta do local de crime, além do treinamento para o aprimoramento das técnicas de coleta. Outra desvantagem do método é a expressão da probabilidade de acerto em razão da averiguação em bancos de dados com estatísticas sobre o DNA de determinada população: o fato de duas amostras possuírem o mesmo perfil para um grupo de STRs não significa que tenham origem comum. A interpretação dos testes depende das frequências populacionais para cada marcador genético utilizado. Por isso, quando ocorre o mesmo perfil para um grupo de STRs, é preciso expressar numericamente a sua significância.

A maior limitação do teste de DNA é a necessidade de haver suspeitos em potencial para a comparação de amostras coletadas das vítimas ou do local do crime. Muitos países – como Estados Unidos, Canadá, Inglaterra, Alemanha e França – possuem bancos de dados genéticos de indivíduos anteriormente condenados. Assim, quando se obtém determinada amostra de DNA, é possível compará-la com todos os DNAs do banco de dados, o que permite relacionar ao crime investigado suspeitos até então inimagináveis.

Os bancos de DNA contêm informações genéticas de determinados indivíduos e, em sua grande maioria, são de caráter forense. Porém, o uso desses dados pressupõe contrabalançar os direitos do indivíduo e os interesses coletivos. Muitos são os aspectos a serem avaliados no uso de um banco de dados desse tipo, já que muitos temores o circundam. É isso porque as informações contidas no DNA são muito mais amplas do que as contidas em impressões digitais. Pelo estudo do DNA de um indivíduo, podem-se estabelecer traços de sua personalidade – um exemplo disso é o fato de homens com um cromossomo Y a mais (XYY) possuírem, na maioria dos casos, comportamento agressivo – ou revelar possíveis doenças que serão desenvolvidas por uma pessoa – como o câncer ou patologias cardíacas –, o que pode levar à discriminação genética.

No Brasil, não há lei que regule o uso do exame de DNA para fins de identificação criminal. Ainda está em discussão uma lei que permita

a inclusão da análise de DNA nos procedimentos de identificação utilizados rotineiramente, como a fotografia e a papiloscopia, na identificação de detentos indiciados ou acusados de crimes de homicídio doloso, de receptação qualificada e contra a liberdade sexual, dentre outros.

2.15 Utilizações complementares do DNA

2.15.1 *Transgênicos*

O desenvolvimento da engenharia genética permitiu a retirada de genes de uma espécie e sua introdução em um indivíduo de espécie diferente. Com essa nova ferramenta nas mãos, o ser humano foi capaz de reproduzir genes de interesse, criando o que chamamos de organismos geneticamente modificados (OGMs).

A produção de OGMs é parte da nova era da biotecnologia. No Brasil, a lei nº 11.105, de 24 de março de 2005, estabelece normas de segurança para atividades que envolvem organismos geneticamente modificados, definindo-os como organismos cujo material genético (DNA ou RNA) tenha sido modificado por qualquer técnica de engenharia genética. Esses indivíduos são constantemente confundidos com os organismos transgênicos, porém nem todos os organismos geneticamente modificados são transgênicos. Os transgênicos são organismos que receberam em sua constituição genética genes de organismos de outras espécies. Aqueles organismos que tiveram seus genes alterados apenas quanto à sua posição ou expressão não são considerados transgênicos.

Os transgênicos têm sido utilizados de forma mais significativa na agricultura, pois com a técnica da transgenia é possível criar plantas resistentes a pragas e a agrotóxicos, organismos resistentes a solos inadequados ou a condições climáticas desfavoráveis ou, até mesmo, organismos capazes de produzir nutrientes específicos de interesse alimentício ou industrial.

Muitas são as discussões acerca dos transgênicos, mas elas podem ser classificadas em quatro dimensões: saúde, agricultura, meio ambiente e ética. No que se refere à saúde, o maior problema é a possibilidade de os organismos transgênicos produzirem toxinas desconhecidas. A introdução de um novo gene pode levar à produção de uma proteína não existente antes no alimento, aumentando as chances de reações alérgicas à nova substância.

Em relação à agricultura, os problemas são vários, como a dependência dos agricultores de indústrias químicas que produzem sementes transgênicas como as agrotóxico-resistentes – por exemplo, as sementes de soja da variedade Roundup Ready®, produzidas pela Monsanto, que resiste ao herbicida Roundup, também produzido por ela. Além disso, a dispersão natural, pelo vento ou pela chuva, de sementes transgênicas traz o risco de contaminação de plantações não transgênicas, acarretando problemas financeiros e jurídicos.

Os impactos ao meio ambiente ainda são muito discutidos entre produtores de transgênicos e investigadores da área. Por exemplo, o fato de ocorrerem modificações bioquímicas do solo (visto que qualquer ser vivo é capaz de intervir e modificar o meio em que vive) causadas por organismos transgênicos e modificações nas frequências gênicas das populações primitivas.

A questão ética permeia todas as demais discussões a respeito dos transgênicos. Os motivos que movem a sua utilização são alvo de críticas científicas e religiosas, fazendo que a transgenia seja sempre um assunto polêmico.

2.15.2 Projeto Genoma Humano

Após a descoberta da molécula de DNA e de sua forma de funcionamento em processos como a síntese proteica, bem como na determinação das características herdadas, grande parcela dos cientistas se dedicou a projetos que visavam à utilização dos conhecimentos de genética para o desenvolvimento da espécie humana. Sem dúvida, o mais ambicioso deles foi o Projeto Genoma Humano, um consórcio internacional idealizado, no final da década de 1980, por James D. Watson – prêmio Nobel de Medicina e Fisiologia em 1962 e descobridor, ao lado de Francis Crick, da estrutura em dupla

hélice da molécula de DNA —, com o objetivo de caracterizar todos os seres humanos com base na sequência de nucleotídeos de seu genoma. Acreditava-se que, em virtude do seu papel essencial no processo de síntese proteica, o DNA tinha enorme importância na determinação de características bioquímicas e fisiológicas do organismo humano.

Além dos Estados Unidos, vários países participaram do projeto, entre eles Alemanha, França, Inglaterra e Japão. Nos Estados Unidos, que centralizaram as pesquisas, houve a participação de duas frentes importantes: uma pública e outra privada. A frente pública, representada pelo Consórcio Internacional para o Sequenciamento do Genoma Humano, foi dirigida por Francis Collins. A iniciativa privada ficou a cargo da empresa Celera Genomics, liderada por Craig Venter. Mesmo não tendo sido harmônica em todos os momentos, a colaboração entre essas duas frentes propiciou a finalização do projeto em 2003.

O termo genoma se refere ao conjunto de todos os genes existentes nos 23 pares de cromossomos da espécie humana. É válido lembrar que cada espécie possui um número próprio de cromossomos, bem como de genes que os compõem, o que determina um genoma diferente para cada tipo de organismo. Podemos chamar **genômica** o processo de mapeamento, sequenciamento e análise do genoma. A genômica trata de localizar os genes no conjunto de cromossomos do organismo para posterior caracterização da sequência de bases nitrogenadas dos genes e elucidação de sua função dentro do organismo.

A genômica pode ser dividida em estrutural e funcional: a genômica estrutural se refere à fase na qual são construídos mapas genéticos que fornecem a localização e a sequência dos genes do organismo; a genômica funcional está relacionada com a expressão gênica, ou seja, com as propriedades funcionais dos conjuntos de genes.

O desenvolvimento tecnológico contribuiu para o entendimento da complexa constituição do genoma humano e só foi possível com o sequenciamento dos genomas de outras espécies, principalmente microrganismos. As espécies foram selecionadas com base no interesse científico, médico ou econômico. Uma série de instrumentos e técnicas, como a PCR e os sequenciadores automáticos, surgiram nesse momento.

O Brasil não ficou de fora desses avanços da era genômica. Em 1997, a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) determinou o sequenciamento do genoma do microrganismo *Xylella fastidiosa*, bactéria responsável por certos tipos de infecção das laranjeiras, importante fonte econômica da região. Esse primeiro sequenciamento teve como principal objetivo a adequação dos laboratórios, com o fornecimento de equipamentos e treinamento técnico. Com o sucesso do projeto, outros mais ambiciosos foram iniciados, como o financiamento do projeto de definição do transcrito humano. Atualmente, o Brasil se coloca como um dos países que mais produziu sequências de organismos vivos no mundo.

Apesar de muitas das expectativas em relação a ele ainda não terem sido realizadas, o PGH gerou importante conjunto de informações para a comunidade científica. Apenas 1,1 a 1,4% do genoma realmente codifica proteínas, sendo que 75% do DNA estão localizados entre os genes. Diferentes segmentos de tamanhos variáveis se repetem no decorrer no genoma, o que talvez denuncie uma complexa história evolutiva. Segundo estimativas, o conjunto de genes da espécie humana varia entre 30 e 40 mil, número que é apenas duas vezes maior do que o número de genes da *Drosophila melanogaster*, a mosca-da-fruta. No entanto, os genes humanos parecem ser muito mais complexos.

As principais diferenças entre os genomas de duas pessoas estão nas modificações específicas de uma única base nitrogenada, chamadas polimorfismos ou SNP (do inglês *single nucleotides polymorphisms*, ou seja, polimorfismos de base única). As consequências do mapeamento genético para a medicina são profundas, pois ele permitirá o diagnóstico mais preciso e rápido, bem como o conhecimento prévio da suscetibilidade genética de uma pessoa ao desenvolvimento de determinada doença. É preciso ressaltar, no entanto, que o DNA não é o único responsável pelo desenvolvimento de doenças: os fatores ambientais e sociais possuem papel tão preponderante quanto o conjunto de genes que a pessoa apresenta. Por isso, é imprescindível, para a melhoria da saúde da população humana, a análise multifatorial dos determinantes patológicos.

Muitas perguntas ainda estão sem resposta e muitos problemas ainda precisam ser resolvidos com auxílio dos estudos e do desenvolvimento da genética e

da biologia molecular. Entretanto, é inegável que os conhecimentos adquiridos nessas áreas colaboram na melhoria da qualidade de vida do homem, e que muitas conquistas científicas serão alcançadas a curto e médio prazo por meio da engenharia genética.

Referência bibliográfica

WATSON, James D.; CRICK, Francis H. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, n. 171, p. 737-7388, April 1953.

Bibliografia complementar

ALBERTS, Dennis B. et al. *Biologia molecular da célula*. 4. ed. São Paulo: Artmed, 2004.

AZEVEDO, Maristella de Oliveira et al. *Técnicas básicas em biologia molecular*. Brasília: Editora UnB, 2003.

HAUSMANN, Rudolf. *História da biologia molecular*. 2. ed. Ribeirão Preto: Funpec–RP, 2002.

RUMJANEK, Franklin David. *Introdução à biologia molecular*. Rio de Janeiro: Âmbito Cultural, 2001.

ZAHA, Arnaldo; FERREIRA, Henrique B.; PASSAGLIA, Luciane Maria P. *Biologia molecular básica*. 3. ed. Porto Alegre: Mercado Aberto, 2003.