

## **INTRODUÇÃO À BIOLOGIA CELULAR E MOLECULAR PARA O ENSINO – PARTE II**

Prof<sup>a</sup>. Adriana Freitas Neves<sup>1</sup>

Caro leitor (a), seja bem-vindo ao capítulo **Introdução à Biologia Celular e Molecular para o Ensino – Parte II**, que foi pensada em você e organizada de forma bem sintética e estruturada para melhor lhe auxiliar na compreensão da Biologia Molecular, cujos estudos nessa área vão desde contribuir para o conhecimento na ciência básica quanto para a ciência aplicada.

Serão apresentados conteúdos com foco na Biologia Molecular, tais como, uma breve abordagem do histórico do DNA como molécula da hereditariedade; seguida do dogma central da Biologia Molecular, com base na estrutura e função dos ácidos nucleicos (replicação e transcrição) e proteínas (tradução), como os genes são regulados, técnicas moleculares e suas aplicações, e aspectos da biotecnologia moderna. As unidades estão organizadas como segue:

Unidade I - Breve histórico: a descoberta do DNA como molécula da hereditariedade

Unidade II - Replicação do DNA

Unidade III – Síntese de RNA – transcrição

Unidade IV – Síntese de proteínas – tradução

Unidade V - Regulação da expressão gênica em procariotos

Unidade VI - Regulação da expressão gênica em eucariotos

Unidade VII - Bases genéticas do câncer

Unidade VIII - Biologia molecular

Espero que a organização desse capítulo possa auxiliá-lo(a) num melhor entendimento das bases dessa grande área do conhecimento científico e sua aplicação na atualidade, a partir da qual você poderá ter uma melhor compreensão da dinâmica celular, seus aspectos moleculares, bem como as técnicas aplicadas ao diagnóstico das mais diversas doenças e aos estudos evolutivos, ampliando suas capacidades na organização de suas ideias, contribuindo para a sua formação continuada com foco nas Ciências Biológicas e para aprofundamento nessa área, cuja evolução é constante!

Excelente leitura e estudos!

---

<sup>1</sup> Doutora em Genética e Bioquímica - Docente da Unidade Acadêmica Especial – Instituto de Biotecnologia, Área: Ciências Biológicas, Regional Catalão, Universidade Federal de Goiás.

**Alguns Conceitos Clássicos:**

**Alelo:** forma alternativa de um gene que pode existir um único loco.

**Alélismo Múltiplo:** conhecidos vários alelos de um gene.

**Bases Complementares:** são bases que seguem um padrão de pareamento, onde A paira com T ou U (RNA) e G com C.

**Base Nitrogenada:** estrutura em anel, representada pelas pirimidinas (C, T, U) e purinas (A, G)

**Carga Genética:** conjunto total de alelos deletérios contidos no genoma de um indivíduo.

**Cariótipo:** conjunto cromossômico de um indivíduo visto durante a metáfase mitótica. A organização dos cromossomos por tamanho e posição do centrômero é chamada de idiograma.

**Célula Diploide:** células com dois conjuntos cromossômicos (2n).

**Células Filhas:** duas células idênticas formadas a partir da divisão de uma célula.

**Célula Haplóide:** célula com metade do conjunto cromossômico (n).

**Célula Somática:** células não destinadas a serem gaméticas. Soma significa corpo.

**Centrômero:** sequência de DNA específica que separa os braços cromossômicos eucarióticos e onde se ligam as proteínas associadas ao processo de divisão de uma célula.

**Cromossomos:** o genoma contém os cromossomos: molécula longa e contínua de DNA. Cromossomos são estruturas de DNA que são vistos durante a divisão celular (Cromossomo metafásico).

**Cromossomos Autossômicos:** qualquer cromossomo que não é sexual.

**Cromossomos Homólogos:** cromossomos que fazem pares entre si na meiose.

**Cromossomos Sexuais:** cromossomos não autossômicos cuja presença ou ausência está associada ao sexo do portador.

**DNA:** ácido desoxirribonucleico, material genético básico que constitui os genes.

**Fenótipo:** características manifestadas pelo indivíduo (pode mudar durante a vida): morfológicas, fisiológicas, comportamentais, outras. É determinado em sua maioria pela interação do genótipo com o ambiente.

**Gametas:** célula haplóide especializada (sexual) contendo metade dos cromossomos. Exemplos: espermatozóide, óvulo e pólen.

**Genes:** unidade estrutural e funcional, constituída por um segmento de DNA contendo uma sequência a ser transcrita e uma sequência regulatória que possibilita a transcrição.

**Genética:** estudo dos genes e da hereditariedade. Dividida em Genética Clássica ou Mendeliana e Genética Molecular.

**Genética Molecular:** estuda a estrutura e função gênica.

**Genótipo:** refere-se ao conjunto de genes herdados por um indivíduo; correspondente genético da característica associada ao fenótipo; comparando ao fenótipo, o genótipo permanece "constantes" através da vida e fenótipos mudam continuamente, pois pode sofrer efeito ambiental.

**Genoma:** complemento inteiro de material genético em um conjunto cromossômico. Varia de espécie para espécie.

**Loco (*locus*) ou locos (*loci* - plural):** posição que determinado gene ocupa num par de cromossomos homólogos.

**RNA:** ácido ribonucleico, material genético transcrito a partir do DNA.

**Nucleotídeo:** monômero que associado forma o polímero de ácido nucleico. Constituído de

um grupo fosfato, um pentose e uma base nitrogenada.

**Nucleosídeo:** uma base nitrogenada ligada a uma pentose.

**Mutação:** alteração na sequência de DNA, podendo produzir um gene. Cromossômica quando ocorre mudança na estrutura ou número de cromossomos na célula, cuja nomenclatura correta são alterações cromossômicas e não mutações.

**Simbologia Mendeliana para a composição genética dos indivíduos:**

Aa = heterozigoto (para um dado gene, o indivíduo porta um alelo dominante e um recessivo)

AA = homozigoto dominante (para um dado gene, o indivíduo porta dois alelos dominantes)

aa = homozigoto recessivo (para um dado gene, o indivíduo porta dois alelos recessivos)

a = alelo recessivo somente se expressa quando em homozigose

## **UNIDADE I - BREVE HISTÓRICO: A DESCOBERTA DO DNA COMO MOLÉCULA DA HEREDITARIEDADE**

Nesse item serão apresentados alguns experimentos importantes na descoberta do DNA como molécula da hereditariedade. Em 1865, o monge Gregor Mendel (1822-1884) publicou seu trabalho sobre experimentos com ervilhas que propõe as leis da hereditariedade (“leis de Mendel”) e supôs que as características hereditárias são transmitidas em unidades. O trabalho permaneceu ignorado até 1900. Em 1909, o dinamarquês Wilhelm Johannsen introduziu o termo “gene” para descrever a unidade mendeliana da hereditariedade. Ele também usou os termos “genótipo” e “fenótipo” para diferenciar as características genéticas de um indivíduo e sua aparência externa. O norte-americano Thomas Hunt Morgan, juntamente com três alunos, publicou o livro “O Mecanismo da Hereditariedade Mendeliana”, no qual relata os experimentos com drosófilas (moscas-das-frutas), e mostra que os genes estão linearmente dispostos nos cromossomos: teoria cromossômica da herança.

Paralelamente às publicações sobre heranças mendelianas, o bioquímico suíço Johann Friederich Miescher, em 1869, trabalhando com células de pus de bandagem de feridos durante a guerra da Criméia, isolou uma substância até então desconhecida, a qual ele chamou de nucleína. Essa substância continha fósforo e nitrogênio. Em 1889, seu discípulo Richard Altmann mudou o nome para ácido nucleico. Entre os anos de 1903 e 1910, o bioquímico Phoebus Aaron Levene descobriu que em alguns ácidos nucleicos também havia um açúcar do tipo ribose, e em outros um açúcar que faltava um átomo de oxigênio em um dos carbonos, a desoxirribose. Experimentos posteriores permitiram o conhecimento do monômero de nucleotídeo e cada uma das quatro bases: adenina, guanina, citosina e timina.

O inglês Frederick Griffith, em 1928, publicou resultados de experimentos de transformação utilizando pneumococos encontrados em pacientes com pneumonia lobar. Nesse estudo, ele utilizou duas diferentes estirpes dessas bactérias, em separado e em associação, as quais ele chamou de virulentas e não virulentas, aquecimento da cultura virulenta, centrifugação da cultura para concentração da mesma e inoculação subcutânea dessas linhagens em camundongos, a partir de, em síntese, quatro experimentos sequenciais (Figura 1).

As bactérias não virulentas (pneumococos tipo R, do inglês *rough*, que significa rugosa) podiam matar camundongos e se tornarem encapsuladas se fossem injetadas com bactérias virulentas (tipo S, do inglês *smooth*, que significa lisa) mortas pelo calor. Isso mostrou que poderia haver transformações entre tipos de bactéria *Streptococcus pneumoniae*, que pode ser letal em camundongos. De acordo com Griffith, essa transformação seria devido a um antígeno S, que seria uma estrutura protéica específica, e quando liberado, as bactérias do tipo R poderiam utilizar para construir na sua estrutura tipo S. Ou seja, assim como todos na época, as proteínas continuavam a ser entendidas como sendo os contribuintes responsáveis pelas transmissões das características genéticas.

A natureza do princípio transformante de Griffith foi determinada pelos experimentos seguintes, publicados em 1944 por Oswald Avery, Colin MacLeod e Maclin McCarty, do Instituto Rockefeller (EUA). Dando seguimento ao trabalho na área microbiológica, com bactérias que acometem a saúde humana, esses pesquisadores trabalharam com estudos químicos de substâncias solúveis, as quais acreditavam serem importantes na patogenicidade desse microorganismos.

Interessados em isolar e se possível caracterizar quimicamente o suposto antígeno S, Avery, MacLeod e McCarty trabalharam com extração de frações separadas de substâncias químicas já conhecidas na época, tais como, os polissacarídeos, lipídeos, proteínas e ácidos nucleicos de bactérias do tipo S mortas pelo calor e as incubaram juntamente com as bactérias do tipo R vivas e as injetaram nos camundongos. Como resultado, somente algumas das frações contendo ácidos nucleicos foram capazes de matar os camundongos. O experimento seguinte envolveu o tratamento das frações com duas enzimas, uma que degrada o ácido ribonucleio e outra que degrada proteínas (tripsina quimiotripsina), seguida de incubação com bactérias R e inoculação nos camundongos.

Nesse estudo, esses pesquisadores observaram que a substância S não poderia ser RNA ou proteínas. Também fizeram testes combinatórios de atividade enzimática para enzimas, tais como, a fosfatase e a DNase (chamada de despolimerase na época) em preparações de frações de tecidos como a mucosa intestinal de cachorro, rim suíno, e observaram que havia a inativação do princípio transformante sempre nas combinações que havia a enzima DNase, independente da ação das demais enzimas. Assim, nas frações de ácidos nucleicos em que restava apenas o RNA quando incubada com bactérias R e injetadas nos camundongos não levavam à morte; no entanto, a morte era causada pela incubação com o DNA.

Assim, a partir desses experimentos sequenciais, esses pesquisadores demonstraram que o ácido nucleico, do tipo desoxirribose, seria a unidade básica capaz de transformar bactérias não-patogênicas em patogênicas, e não as proteínas. No entanto, esses pesquisadores estavam cercados, na época, por químicos e físico-químicos envolvidos no estudo de proteínas, bem como por microbiologistas e virologistas com mais visibilidade no meio científico, que era difícil que esses “modestos” cientistas pudessem ter seus trabalhos bem aceitos, incluindo a percepção de que a comunidade científica da época não estava pronta para aceitar o DNA como molécula associada à hereditariedade. O próprio Avery não era geneticista e também não era conhecido pelos geneticistas, bem como a revista que escolheram para publicarem seus trabalhos não era muito conhecida ou lida por esse grupo de cientistas. Assim, continuaram acreditando que as proteínas, um dos componentes dos cromossomos, seria a molécula da hereditariedade e não o DNA, outro componente dos cromossomos; talvez, por existir traços de impurezas por proteínas que não teria sido completamente eliminado (Figura 2).

O austríaco Erwin Chargaff foi um dos poucos cientistas a reconhecer a importância dos trabalhos desenvolvidos pelo grupo de Avery. Em 1949, trabalhando com DNA de diferentes espécies e tecidos, publicou um trabalho conhecido como “regra de Chargaff” estabelecendo uma relação quantitativa entre as bases de DNA, observando uma proporção (concentrações molares) equivalente entre adenina (A) e timina (T) e o mesmo entre guanina (G) e citosina (C). Observou ainda que a relação A+T/G+C era variável entre espécies, mas a mesma entre os diferentes tecidos da mesma espécie (Figura 3).

O ano de 1952 representou dois marcos importantes, tais como, a obtenção de imagens de DNA utilizando técnicas de difração de raios X pela britânica Rosalind Franklin e os experimentos dos norte-americanos Alfred Hershey e Martha Chase (*Cold Spring Harbor Laboratory*), os quais, utilizando marcadores radioativos, mostraram que o DNA de um bacteriófago era quem programava as células para fazer cópias do vírus e não a proteína.

O experimento de Hershey & Chase reforçou a ideia de que os genes estão contidos no DNA. Como na época já tinham o conhecimento de que o DNA continha fósforo, mas não enxofre, enquanto algumas proteínas continham enxofre, mas nenhum fósforo, esses pesquisadores marcaram o DNA pelo cultivo do fago em meio de cultura contendo o isótopo

radioativo  $P^{32}$ , enquanto a cápsula do bacteriófago foi marcada por cultivo em meio de cultura contendo  $S^{35}$ . Por centrifugação e obtenção de duas frações, eles observaram que a maior radioatividade por  $P^{32}$  foi encontrada no precipitado de células bacterianas, enquanto a maior radioatividade por  $S^{35}$  foi encontrada no sobrenadante do tubo contendo capsídeos dos fagos, indicando que as proteínas não entravam nas bactérias. Esses dados confirmaram os achados do grupo de Avery em que o material viral injetado em *E.coli* foi o DNA sendo o material responsável por transferir a informação genética através das gerações, enquanto as proteínas da cápsula não eram injetadas na célula. Além disso, a prole viral continha parte de  $P^{32}$ , mas nenhum  $S^{35}$ . Em 1969, Hershey juntamente com Max Delbruck (coordenador do Grupo Phago) e Salvador Luria, receberam o Prêmio Nobel de Medicina (Figura 4).

Até esse momento, estava então, experimentalmente, demonstrado que o DNA seria a molécula da hereditariedade, no entanto, não havia ainda o entendimento de outras importantes questões, como qual a estrutura dessa molécula, como a informação genética pode ser estocada no DNA e como essa informação é passada para a prole.

Na tentativa de responder a essas questões, em 1953, os pesquisadores James Watson e Francis Crick publicaram na revista *Nature* dois trabalhos sobre a estrutura do DNA em dupla hélice (*A Structure for Deoxyribose Nucleic Acid* e *Genetical Implications of the Structure of Deoxyribonucleic Acid*), na qual são divulgados também outros dois artigos sobre DNA, ambos descrevendo resultados experimentais de difração de raios-X de Rosalind Franklin que eram compatíveis com a estrutura proposta por Watson e Crick. Assim, a partir de experimentos prévios realizados por outros pesquisadores, e por parcerias, revisão de trabalhos e montagem de modelos prévios não aprovados pelos pesquisadores envolvidos no estudo da estrutura do DNA, como o proposto pelo químico Linus Pauling (três hélices), Watson e Crick realizaram a montagem do quebra-cabeça.

A primeira peça era o conhecimento dos nucleotídeos. De estrutura bem simples, os nucleotídeos apresentam três componentes principais, sendo um fosfato, um açúcar desoxirribose, e qualquer uma das quatro bases nitrogenadas, como bases purinas (adenina e guanina) com estrutura em dois anéis e as pirimidinas com um anel (timina e citosina) (Figura 5).

A segunda peça foi os resultados provenientes dos trabalhos apresentados por Chargaff sobre a composição de bases nitrogenadas do DNA de diferentes espécies.

A terceira peça veio dos dados de difração de raios-X do DNA, da física britânica Rosalind Franklin (1952). A famosa foto 51 de uma molécula de DNA mais hidratada (DNA-B) conseguida por Rosalind Franklin, bem como suas análises das medidas estruturais, foi entregue por seu chefe Raymond Gosling ao pesquisador Maurice Wilkins, do mesmo laboratório *King's College* em que ela trabalhava e que também trabalhava com estrutura do DNA por difração de raios-X, o qual mostrou a Watson. A pesquisadora Franklin faleceu de câncer em 1958, possivelmente ocasionado pela exposição aos raios-X. Na elucidação da dupla hélice também foram importantes as considerações feitas pelo cristalógrafo Jerry Donohue a Watson & Crick de que as bases deveriam estar na sua forma 'ceto' e não 'enol' como comumente se apresentavam nos livros.

O procedimento das análises de difração de raios-X não é fácil de fazer nem de se interpretar e requer um tratamento matemático para a interpretação do padrão de pontos no filme. Os dados disponíveis sugeriram que o DNA é longo e fino e tem duas partes similares paralelas ao longo da molécula, mostrando que o DNA é uma molécula helicoidal (hélice/espiral). Assim, a partir dessa terceira peça foi possível deduzir a estrutura tridimensional do DNA, com pareamento entre purina e pirimidina compatível com os dados

de raios-X; enquanto que o pareamento entre purina com purina geraria um DNA largo e no pareamento entre pirimidina com pirimidina o DNA seria muito estreito, ambos não compatíveis com os dados de raios-X.

Tendo como base esses experimentos descritos anteriormente, Watson e Crick apresentaram a estrutura primária do DNA, como sendo organizado na forma de dupla hélice, sendo que cada hélice corresponde a uma cadeia de nucleotídeo disposta na polaridade inversa. Os nucleotídeos são mantidos juntos por ligações fosfodiéster entre o carbono 3' do açúcar de um nucleotídeo e o grupo fosfato do próximo nucleotídeo, enquanto as duas hélices são mantidas juntas por pontes de hidrogênio entre as bases nitrogenadas. O pareamento de bases é específico, onde adenina (A) se pareia com timina (T), guanina (G) se pareia com citosina (C), e as ligações das purinas com as pirimidinas se dão por duas e três pontes de hidrogênio, respectivamente. A complementaridade dos filamentos da dupla hélice e sua estrutura peculiar é que permite que o DNA seja o estoque da informação genética e permite que essa informação genética seja transmitida de geração a geração (Figura 6).

Em 1962, Watson & Crick, juntamente com Maurice Wilkins, foram agraciados com o Prêmio Nobel de Medicina, mas não a Rosalind Franklin, já que o Nobel não é conferido postumamente.

Entre outras propriedades apresentadas pela molécula de DNA, destacam-se a sua aparência altamente viscosa em soluções de DNA com pH 7,0 e na temperatura ambiente. Além disso, em altas temperaturas ou pH extremos o DNA sofre desnaturação pela ruptura das pontes de hidrogênio entre os pares de bases. A desnaturação do DNA não desfaz ligação covalente ficando, portanto, as duas fitas de DNA apenas separadas. Quando o pH e a temperatura voltam ao normal, as duas fitas de DNA espontaneamente se enrolam formando novamente o DNA dupla fita. Com base nas características de hidrofobicidade das moléculas, a estrutura de dupla hélice do DNA fica com parte hidrofílica, grupo fosfato e açúcar, localizados na parte externa da molécula e as bases nitrogenadas estão localizadas na parte interna da molécula. Ou seja, essas interações hidrofóbicas forçam as bases a se "esconderem" dentro da dupla hélice.

A estrutura do DNA duplex pode tomar diferentes conformações, conforme observado por Rosalind Franklin. A estrutura mais comum do DNA, semelhante ao encontrado sob condições fisiológicas e em solução, é o chamado DNA-B que se apresenta mais hidratado, com giro da hélice voltado para a direita, dupla hélice mais longa e mais fina, e 10 pares de bases são necessários para completar uma volta na hélice. O DNA assume a conformação DNA-A quando há interação DNA-RNA ou RNA-RNA, apresentando uma menor quantidade de água e, portanto, uma forma mais curta e mais grossa que o DNA-B, mas também possui giro da hélice para a direita e 11 pares de bases são necessários para completar uma volta na hélice. A relação espacial entre as duas fitas cria um sulco principal e um sulco secundário, com diferentes interações entre ligantes, tais como, proteínas. O Z-DNA tem o sentido de rotação para a esquerda, uma conformação mais alongada e mais fina do que o DNA-B, para completar uma volta na hélice são necessários 12 pares de bases e assume essa conformação em solução com altas concentrações de cátions e com sequências alternadas entre purinas e pirimidinas (por exemplo: ...GCGCGCGC...) (Figura 7).

De acordo com a proposição do "Dogma Central da Biologia Molecular", feita em 1958 por Francis Crick, e publicada em 1970 na revista *Nature (Central Dogma of Molecular Biology)*, a especificidade de um fragmento de ácido nucléico depende apenas da sequência de suas bases e que essa sequência seria a chave para a disposição dos aminoácidos em

uma proteína particular. Nesse sentido, propõe que o fluxo de informação vai do DNA para o RNA e depois para a proteína, e que não pode retornar.

Considerando esse contexto histórico sobre o fluxo da informação genética, em 1958 os norte-americanos Mattheu Meselson e Franklin Sthal demonstraram, experimentalmente, que o DNA se replica de maneira semiconservativa, em que os dois filamentos da molécula de origem se separam e cada um deles passa a se emparelhar com um filamento novo. Na época, outros modelos foram descritos, como o modelo conservativo, em que sugeria que ambas as fitas parentais serviam como molde para a formação de duas fitas novas de DNA formando uma molécula de DNA inteiramente nova. Outro modelo, o dispersivo, sugeria que as fitas de DNA eram quebradas frequentemente em que a nova fita de DNA era formada por fragmentos novos e parentais (Figura 8).

Nesse modelo, Meselson & Stahl trabalharam com células de *E. coli*, centrifugações, meios de cultura e isótopos radioativos. Inicialmente, as células de *E. coli* foram colocadas em um meio para crescimento contendo nitrogênio pesado ( $N^{15}$ ). Assim, a partir da síntese de bases nitrogenadas havia incorporação desse nitrogênio e durante o processo de replicação todo o DNA ficou marcado com este isótopo radioativo. Numa ultracentrifugação por gradiente de densidade utilizando CsCl, haveria marcação mais ao fundo do tubo. Na primeira geração de transferência dessas células para o meio contendo nitrogênio leve ( $N^{14}$ ), assim como na segunda geração também em meio contendo o  $N^{14}$  foram observadas marcações de densidade intermediária ao  $N^{14}$  e  $N^{15}$  (1ª. Geração) e novamente marcações de densidade intermediária entre  $N^{14}$  e  $N^{15}$  e marcações somente na densidade de  $N^{14}$  (2ª. Geração). Conforme pode ser observado na figura, a interpretação das bandas distintas, formadas em cada geração, indicaram um Modelo de Replicação Semiconservativa. Essas marcações poderiam ser visualizadas por luz ultravioleta (Figura 9).

Em 1956, o norte-americano e bioquímico Arthur Kornberg identificou a polimerase, enzima que catalisa (polimeriza) a síntese de DNA utilizando extratos bacterianos livres de células, demonstrando que os nucleotídeos componentes do DNA são precursores ricos em energia (dATP, dGTP, dCTP e dTTP) e são os blocos necessários à ação dessa enzima polimerizadora (Figura 10). Estudos posteriores demonstraram que a enzima DNA Polimerase tipo I promove a ligação dos nucleotídeos precursores através de ligações fosfodiéster e, além disso, ela atua apenas na presença de DNA, o qual é necessário para ditar a ordem dos quatro nucleotídeos no produto polinucleotídico e na depende de uma pequena cadeia pré-existente de nucleotídeos.

Em 1963, John Cairns demonstrou que o início da replicação do DNA se dá pela forquilha em que a dupla hélice se desenrola para produzir os dois filamentos moldes para a cópia. Utilizou timidina tritiada ( $^3H$ ), uma timina marcada com o isótopo radioativo de hidrogênio trítio. Com base no experimento de Meselson e Stahl, esperava-se uma fita quente (radioativa) e outra fria (não radioativa). Nesse experimento, células de bactérias foram crescidas por várias gerações nesse isótopo e após lisá-la foi feita a sua transferência para um slide, análise por microscópio de luz e fez também autoradiografia. Após exposição em emulsão fotográfica de um ciclo de replicação, fez a autoradiografia obtendo uma imagem com pontos pretos em forma circular, demonstrando que o DNA bacteriano é circular. Os dados dos pontos pretos (“quentes”) foi interpretado como um filamento radioativo recém-sintetizado e outro não radioativo. Num segundo ciclo de replicação, as forquilhas foram observadas, sendo que nessa a densidade de pontos escuros permitiram a interpretação de dois filamentos radioativos na forquilha (Figura 11).



## UNIDADE II - REPLICAÇÃO DO DNA

A estrutura do DNA é simples quando comparado à informação por ele carregada necessária para a manutenção de todas as atividades celulares. Como vimos na unidade anterior, a molécula de DNA é um polímero formado por monômeros chamados nucleotídeos (ou mais especificamente desoxirribonucleotídeos). A fita de DNA é dupla, com exceção de certos vírus. As duas fitas são antiparalelas entre si e as bases de cada fita são complementares. As interações entre bases acontecem pela interação do tipo ligações de hidrogênio (“pontes”), sendo que a base A (adenina) interage com T (timina) se for DNA-DNA, ou com U (uracila) se for DNA-RNA por duas pontes de hidrogênio; a G (guanina) interage com a C (citosina) por três pontes de hidrogênio.

Na replicação do DNA, ambas as fitas, paralela e antiparalela, devem ser replicadas ao mesmo tempo e, portanto, em direções opostas. Dessa forma, a partir do processo de replicação, que ocorre bidirecionalmente, são formadas duas fitas novas, uma contínua (*leader*) e uma descontínua (*lagging*), em função da sua visualização a partir de um ponto de simetria. Ou seja, a partir desse ponto de simetria ocorre a inversão entre as fitas, sendo que a líder para ser descontínua, e vice-versa. Cada fita alvo (*template*) age como molde para a síntese da nova fita, sendo a replicação do tipo semiconservativa. O início do processo se dá pela ação da DNA helicase, a qual desenovela o DNA na origem ou forquilha de replicação utilizando energia do ATP, transformando o DNA de dupla fita em DNA de fita simples (*single strand DNA* – ssDNA) e permitindo que cada uma das fitas simples seja copiada.

Assim, a replicação requer atividade coordenada de diversos tipos de enzimas, dentre estas, estão as com atividades de topoisomerases (promovem quebras temporárias para remover o superenovelamento do DNA no local da replicação, deixando-o mais relaxado), helicases (levam à deselicoidez da dupla hélice na forquilha de replicação e rompimento das ligações de hidrogênio, que unem os dois filamentos de DNA na dupla hélice), primases (adicionam os “*primers*” ou iniciadores - oligonucleotídeos de RNA - fornecendo a porção 3'-OH livre para a incorporação de novos nucleotídeos por complementaridade de bases pela DNA Polimerase), DNA polimerases (do tipo III adiciona nucleotídeos aos primers preexistentes para a formação das novas fitas e adiciona os nucleotídeos na fita retardatária após a retirada dos primers pela DNA Polimerase do tipo I), proteínas SSB (*single strand binding protein* – proteínas de ligação ao DNA de fita simples, que tem função de estabilizar o DNA deselicoideado se ligando ao DNA unifilamentar e retardando a reconstituição de dúplice), DNA ligases (fazem a ligação fosfodiéster entre os fragmentos de Okazaki da fita descontínua para formar uma fita contínua, após remoção dos primers e adição de nucleotídeos pela DNA Polimerase III).

Como dito anteriormente, no cromossomo de procaríoto o processo se inicia a partir de uma origem de replicação chamada de ori C, que apresenta um tamanho aproximado de 245-pb cuja sequência é rica em bases A e T (Figura 12); enquanto em eucariotos existem várias origens, por vezes citadas como ARS (*Autonomolously Replicating Sequence* - sequências de replicação autônoma). Essas sequências têm sido mais bem compreendidas em leveduras (*Saccharomyces cerevisiae*), mas é assumido que existe em todos os cromossomos eucarióticos. A replicação pode se iniciar simultaneamente em nos ARS dispostos ao longo dos cromossomos durante o processo de replicação, formando um processo cooperativo e rápido de replicação bidirecional (Figura 13).

Em síntese, uma vez abertas as fitas do DNA, é necessário que essas permaneçam abertas para que a replicação ocorra. As proteínas responsáveis pela prevenção do

reanelamento das fitas de DNA são chamadas de SSB mantendo o DNA em fita simples durante o processo de replicação. A primase sintetiza um primer de RNA, e, juntamente, com a DNA helicase e outros polipeptídeos forma o primossomo (Figura 14). O primer servirá como material inicial para a adição de nucleotídeos pela DNA Polimerase III, fornecendo uma extremidade com 3'-OH livre. Por ser um processo bidirecional, depois de separadas, a replicação de uma das fitas (*leading - líder*) se dá de forma contínua no sentido do movimento da forquilha de replicação 5'→3', e da outra fita (*lagging*) de forma descontínua e em direção oposta ao movimento da forquilha, formando os Fragmentos de Okazaki. Para completar os segmentos da fita descontínua, a DNA ligase une os fragmentos de Okazaki fazendo ligação fosfodiéster e formando uma fita contínua. A DNA polimerase III exerce um papel primário na catálise de nucleotídeos para formar a nova fita de DNA (Figura 15). A replicação do DNA é o processo responsável pela duplicação do material genético durante a divisão celular e formação de novas células (passagem da informação genética de célula a célula) associadas ao desenvolvimento do indivíduo, bem como no processo de reprodução (passagem da informação genética de geração a geração).

As células eucarióticas apresentam várias DNA Polimerases, sendo três dessas polimerases com suas subunidades essenciais à replicação, são estas a DNA Polimerase  $\alpha$ /primase, DNA Polimerase  $\delta$  e DNA Polimerase  $\epsilon$ . No que se refere às funções, a DNA Polimerase  $\alpha$ /primase, com quatro subunidades, está envolvida ao início do processo de replicação para formação das novas fitas por meio da síntese do primer iniciador (duas subunidades primase) seguida do início da polimerização de nucleotídeos pelas duas subunidades  $\alpha$ . As DNA Polimerases  $\delta$  e  $\epsilon$  substituem a DNA Polimerase  $\alpha$  aumentando a velocidade de síntese, por apresentarem alta processividade. Esse processo de troca de polimerases é chamado de *polymerase switching* atuando na origem de replicação de eucariotos. Como em procariotos, outras DNA polimerases também têm função envolvida no reparo do DNA que apresentam bases incorporadas erroneamente causando distorção na molécula. Essas atividades podem ser de exonuclease 5'→3' ou 3'→5'. O aumento na processividade na replicação pelas DNAs Polimerases é devido ao ancoramento a outras proteínas que formam um grampo/anel deslizante sobre a fita de DNA, circundando-a e permitindo que a DNA Polimerase faça a síntese da nova fita de maneira mais eficiente (Figura 16).

A maquinaria replicacional não consegue replicar completamente os cromossomos lineares de eucariotos, os quais apresentam regiões específicas, tais como, os centrômeros e os telômeros, importantes no processo de divisão celular e na estabilidade cromossômica (Figura 17). Os telômeros se localizam na porção terminal dos cromossomos e apresentam uma composição de bases com sequências repetitivas em *tandem* (sequencial) de aproximadamente 10 à 15Kb (1 Kb corresponde a 1000 pares de bases – pb), variável entre os diferentes organismos, rica em GT. Como exemplo, em humanos a sequência telomérica é (5'-TTAGGG-3')<sub>n</sub>.

A falta de síntese de DNA das pontas na fita tardia ou descontínua a cada ciclo de replicação acarreta a perda de algumas unidades de repetições que formam os telômeros. Em processos de várias divisões celulares seria esperado a perda progressiva dos segmentos terminais da fita de DNA o que poderia atingir sequências codificadoras, levando à perda de informações genéticas essenciais à célula. O comprimento do telômero está associado à contribuição no envelhecimento celular normal. Algumas células, no entanto, exibem a ação da enzima telomerase de eucariotos; uma enzima com ação de transcriptase reversa diferenciada, que contém uma pequena molécula de RNA, que serve de molde para a extensão dos telômeros se ligando ao ssDNA (Figura 18). Essa atividade tem sido

observada como sendo necessária para a imortalização celular no crescimento tumoral, com atividade altamente regulada durante a oncogenese, assim como no desenvolvimento. Nesse sentido, a modulação da ação da enzima telomerase pode ter implicações nas terapias anticancer e antienvhecimento.

Outro modelo peculiar de replicação é a chamada círculo rolante, utilizada por algumas moléculas circulares de DNA como os plasmídios e em alguns tipos de vírus. Nesse modelo de replicação, o DNA circular parece girar à medida que desenrola um filamento contínuo pela ação de endonucleases. O corte feito por essas enzimas em uma das fitas de DNA fornece uma extremidade 3'-OH para ação da DNA Polimerase. À medida que a fita nova vai aumentando em tamanho, a molécula de DNA vai girando (Figura 19).

### **UNIDADE III – SÍNTESE DE RNA - TRANSCRIÇÃO**

A transcrição consiste na síntese de uma molécula de RNA a partir das informações contidas na sequência de nucleotídeos de uma molécula de DNA. A transcrição é um processo extremamente variável, sendo a célula quem controla quando determinadas sequências devem ser transcritas e o quanto de RNA deve ser sintetizado. Esse mecanismo é conhecido como Controle ou Regulação da Expressão Gênica. Assim como no processo de replicação, a transcrição ocorre no sentido 5' → 3' e é baseada na complementaridade de bases, sendo que na formação do RNA apenas uma fita de DNA é usada como molde, o açúcar da pentose é uma ribose (hidroxila no carbono 2 da pentose), o nucleotídeo uracila (U) substitui a timina no pareamento com a base adenina e o RNA formado é unifilar (fita simples), com a mesma sequência da fita não-molde do DNA, também chamada de codificante, trocando o nucleotídeo T por U no RNA (Figura 20).

Dentre os aspectos diferenciais entre replicação e transcrição, destacam-se que esse último processo não requer um primer iniciador da reação de polimerização, apenas uma das fitas de DNA é copiada e a enzima responsável pela transcrição é a RNA Polimerase, que tem por função desnaturar as fitas de DNA, manter as fitas separadas, estabilizar o híbrido DNA/RNA, renaturar as fitas de DNA e terminar a síntese de RNA.

O processo de transcrição envolve três fases. Na fase de iniciação em bactérias, a RNA Polimerase inicia a cadeia nova por meio da formação de uma “bolha” de separação das fitas de DNA de forma dependente de sequências promotoras. Os promotores são sequências específicas de reconhecimento para a RNA Polimerase, que sinalizam exatamente onde a síntese de RNA deve ser iniciada, sendo altamente conservadas entre diferentes genes e espécies. O promotor antecede o sítio de iniciação de transcrição, que por convenção, a primeira base a ser transcrita é numerada como +1 (Figura 21).

Nessa fase de iniciação, a RNA Polimerase, uma holoenzima composta por cinco subunidades, faz contato com as regiões promotoras e deslicoidiza o DNA pela ação de sua subunidade dissociativa, o fator sigma. O fator sigma reconhece a região promotora, formando o Complexo Promotor Fechado. Após promover a deslicoidização de bases próximas à região -10 da fita de DNA que contém a região promotora, dá-se a formação do Complexo Promotor Aberto pela liberação do fator sigma e início da transcrição (Figura 22).

Após dissociação do fator sigma se inicia a segunda fase do processo de transcrição, o alongamento (ou elongamento), que é baseado na incorporação de nucleotídeos pela RNA polimerase no sentido 5' → 3' por complementaridade de bases. A “bolha de transcrição” é mantida e se move ao longo do duplex de DNA enquanto a RNA Polimerase vai adicionando os ribonucleotídeos e “construindo” a fita simples de RNA (Figura 23).

O término da transcrição intrínseco se dá pela formação de motivos estruturais envolvendo cerca de 40 bases terminais, baseados em sequências de bases repetidas e invertidas ricas em Cs e Gs e à formação no RNA nascente de uma alça ou grampo de terminação, seguido de uma sequência rica em U. Essa sequência de poli Us pode facilitar o desligamento da RNA Polimerase do complexo transcricional devido à fraca ligação entre A e U (Figura 24).

Para RNAs que não apresentam esses motivos, existe um segundo tipo de mecanismo de término, conhecido como *Rô-Rut*. A proteína denominada *Rô* reconhece sinais de término, os quais geralmente não têm sequências capazes de formar grampo na extremidade 3' do RNA nascente, sendo que esses RNAs podem conter uma sequência rica em C mas pobre em G, com segmento posterior chamado denominado *Rut* (do inglês, *rho*

*utilization*). A proteína *Rô*, dependente de ATP, move-se mais rapidamente sobre a fita de RNA nascente do que a RNA Polimerase, e ligando a sequência *Rut* leva à dissociação do RNA do complexo de transcrição (Figura 25).

Considerando que os genomas eucarióticos são maiores e com mais genes a serem reconhecidos e transcritos, existe uma divisão do trabalho de transcrição entre três RNA polimerases, sendo que de forma simplificada, a RNA Polimerase I transcreve RNA ribossomal (RNAr), a do tipo II transcreve RNA mensageiro (RNAm) e alguns snRNAs (pequenos RNAs nucleares) e a do tipo III transcreve RNA transportador (RNAt) e alguns snRNAs. Os promotores eucarióticos não são capazes de fornecer sinais de reconhecimento suficientes para que a RNA Polimerase inicie a transcrição. Os promotores eucarióticos são inicialmente reconhecidos por proteínas regulatórias chamadas de Fatores Gerais de Transcrição (GTFs ou simplesmente TFs – do inglês, *transcription factor*).

Na fase de iniciação de transcrição os TFs reconhecem as sequências promotoras eucarióticas, em que o motivo TATA boxe quando presente no gene ocupa a posição relativa de -25 pb do sítio +1, além de existir outras sequências proximais CCAAT, Gs/Cs ou motivos em octâmeros ocupando variadas posições relativas a partir do sítio +1. Para transcrição de RNAm, os TFs recebem a denominação de TFII em função do tipo de RNA Polimerase responsável pela síntese do RNAm. Os TFIIs que interagem com a RNA Polimerase II é representado por seis proteínas principais, sendo todas essas um complexo multiprotéico.

Em ordem, o primeiro evento na formação do complexo de pré-iniciação de transcrição consiste na ligação da proteína de ligação ao TATA (TBP, do inglês *TATA binding protein*), uma subunidade do complexo TFIID. Após essa ligação ao TATA, os outros TFIIs (TFIIA, TFIIB, TFIIF, TFIIE e TFIIH) e a polimerase são recrutados, juntamente com outras proteínas, se ligando de forma coordenada para haver desnaturação do DNA, por meio da ação da TFIIH e hidrólise de ATP (semelhante à ação da DNA helicase). Para passar para a segunda fase do processo de transcrição, o alojamento, a RNA Polimerase II é fosforilada no seu domínio carboxi terminal (CTD) que apresenta sequência repetida composta por aminoácidos, tais como, a serina e a prolina, os quais apresentam sítios de fosforilação por quinases específicas, função que pode ser atribuída a uma das subunidades do TFIIH que é uma quinase. Assim, o TFIIH é o mais complexo comparado aos demais fatores, apresentando subunidade que atuam como ATPases e outra como proteinaquinase. O estado fosforilado no CTD da RNA Polimerase II permite a liberação dessa enzima do complexo de proteínas de pré-iniciação, a qual deixa o promotor e inicia o alongamento (Figura 26).

A fase de alongamento ocorre na bolha de transcrição e diferente do que ocorre com procariotos, em que à medida que o RNA é transcrito vai sendo traduzido; em eucariotos, alguns eventos terão que ocorrer com o RNAm no núcleo celular, antes da tradução no citoplasma. As etapas associadas à maturação do RNAm formado é conhecida como Processamento do RNA.

As etapas do processamento do RNA heterogêneo nuclear (RNAhn) precursor do RNAm envolvem a modificação do RNAm nascente na extremidade 5' pela adição de um capacete de 7-metil-guanosina, que tem por função a proteção desse RNA contra o ataque de nucleases e para o direcionamento dos ribossomos sobre essa extremidade do RNAm quando está no citoplasma. Na região intermediária do RNAm também ocorre alterações importantes na função que essa molécula irá exercer na célula. Essa modificação envolve um processo chamado de *splicing*, que envolve a retirada de regiões não codificantes ou intercalares chamadas de introns e junção das sequências codificantes, os exons. A última

modificação ocorre na extremidade final do RNAm, a região 3', na qual ocorre a adição de uma cauda poli A (cerca de 200 As) pela ação da enzima Poli A Polimerase (Figura 27).

O *splicing* é realizado por um complexo formado por pequenos RNAs nucleares (snRNAs), associados a proteínas (snRNPs), chamado de spliceossomo. O spliceossomo ajuda a clivar o RNAm nos sítios doadores e aceptores de *splicing*, remove os introns, impede o afastamento dos exons, e, por fim, une os exons. A junção por meio do *splicing* requer a presença de sequências conservadas de bases AG no término do exon, GU no início e AG no término do intron, iniciando o próximo exon com GU. Nessas junções alguns nucleotídeos específicos foram encontrados sendo quase idênticos entre os genes e entre as espécies. O sítio A, chamado de ponto de ramificação, situa-se a cerca de 15-45 nucleotídeos antecedentes à 3' do ponto de corte no intron (Figura 28). A retirada do intron se dá pelo reconhecimento das regiões terminais dos exons e dos introns, formando uma dobra em sua própria molécula (alça), com participação do ponto de ramificação e das snRNPs que se associam ao mRNA propiciando o corte e novamente a junção.

A etapa de processamento é dita co-transcricional, pois ocorre nas três regiões do RNA à medida que a síntese vai acontecendo. Além do processamento que se dá por spliceossomo, existem moléculas de RNAs capazes de fazer a sua própria recomposição, chamadas de introns de auto-recomposição (*self-splicing*). Estudos recentes demonstram que os snRNAs e não o componente protéico do spliceossomo são os responsáveis pela reação de catálise para remoção dos íntrons. Os RNAs com essa capacidade catalítica, tais como, os RNAs do spliceossomo e várias outras moléculas de RNAs são chamados de ribozimas, pois apresentam muitas das características de uma enzima clássica como sítio ativo, sítio de ligação para substrato e sítio de ligação para um cofator (como um íon metálico).

No que se refere ao término da transcrição nos eucariotos em muitos genes pela presença de um sinal composto pelas bases AAUAAA ou AAUUAAA (sinal de poliadenilação) na extremidade 3' do RNAm. Como será observado na unidade seguinte, após sintetizado, o RNAm não apresenta qualquer associação com a tradução, ou seja, o tamanho do RNAm não corresponde ao tamanho da proteína a ser traduzida, pois além do *splicing*, o RNAm apresenta sequências transcritas, mas não traduzidas tanto na extremidade 3' como na 5' (UTRs, do inglês *untranslated region*). Outras características apresentadas por alguns tipos de RNAs, é a possibilidade de sofrer *splicing* alternativo, que se refere à recomposição sequencial alternativa dos exons e ocorre para atender uma necessidade tecido-específica.

Um exemplo de *splicing* alternativo pode ser observado para o gene da calcitonina. O RNAm recomposto para a expressão da proteína calcitonina, que atua nas células da tireoide, apresenta exons diferentes da calcitonina (CGRP – *calcitonin gene related peptide*) expressa pelas células do hipotálamo (Figura 29).

## **UNIDADE IV – SÍNTESE DE PROTEÍNAS - TRADUÇÃO**

A síntese proteica envolve a ação dos três tipos de RNAs principais: o RNAm, o RNA transportador (RNAt) e o RNA ribossomal (RNAr). O RNAt possui uma estrutura tipo trevo de três folhas, com dobras em sua própria molécula formando em uma extremidade a dobre o anticódon, que corresponde à trinca de nucleotídeos que fará o pareamento com o códon do RNAm (Figura 30a). Considerando o código genético que será descrito posteriormente nessa unidade, o códon dita o tipo de aminoácido a ser acoplado na região 3' (término CCA) do RNAt, sendo que cada RNAt é específico para um determinado aminoácido. A interação entre RNAt (anticódon) e RNAm (códon) se dá por complementaridade de bases, sendo que a sequência de aminoácidos para um polipeptídeo (ou proteína) é dada pelo códon presente no RNAm.

Diferentes tipos de RNAr compõem os ribossomos, a organela responsável pela síntese proteica. Em eucariotos, os RNAr de coeficiente de sedimentação 28 S e 18 S são os principais na composição do ribossomos associados a outros RNAr menores e proteínas formam as subunidades ribossomais. Os principais em procariotos são do 26 S e 13 S. No início da formação do complexo para a síntese proteica ocorre a interação da subunidade ribossomal menor com o RNAm e, na subunidade maior, formam-se o peptídeo pela carregamento de aminoácidos pelo RNAt, que ocupam os sítios aminoacil (A) seguido do peptidil (P) presentes nessa subunidade. Em procariotos, à medida que o RNAm vai sendo transcrito, também vai sendo traduzido. Já em eucariotos isso não ocorre, pois o RNAm deve migrar do núcleo para o citoplasma para que seja traduzido, ou seja, não são eventos acoplados.

No RNAm existem sequências que sinalizam para início e término de tradução. O códon iniciador de tradução é o AUG, que codifica para uma metionina em eucariotos e para uma formilmetionina em procariotos. Em procariotos, o RNAm apresenta uma sequência localizada a cerca de sete nucleotídeos acima do códon AUG que se pareia com sequências consenso do RNAr 16 S, chamada de Shine-Dalgarno (ou RBS, *ribosome binding site*). O primeiro RNAt liga-se à subunidade menor e o conjunto desliza pelo RNAm até encontrar o códon AUG. Para cada RNAm existem potencialmente três diferentes leituras codificantes (já que as bases são lidas em trincas), porém só uma das combinações de leitura (ORF, do inglês *open reading frame*) é correta sendo que o provável produto proteico frequentemente é a cadeia de polipeptídeos mais longa. Assim, a primeira trinca AUG abaixo da sequência de Shine-Dalgarno será o início da síntese proteica (Figura 30b).

Ao conjunto subunidade ribossomal menor-RNAt na posição exata para início da síntese proteica liga-se a subunidade maior, completando o ribossomo. O sítio A do ribossomo vazio aguarda a chegada do próximo RNAt, transportando o aminoácido correspondente ao códon apresentado após o códon AUG. Quando o RNAt traz esse novo aminoácido e se liga ao RNAm acontece a reação química chamada de ligação peptídica entre o primeiro aminoácido e o segundo. O primeiro RNAt que ocupava o sítio P do ribossomo se desliga, podendo transportar um novo aminoácido. O ribossomo então desloca-se no mesmo sentido 5'→3' do RNAm ocupando um novo codón o que libera o sítio A para ser ocupado por novo RNAt trazendo o aminoácido de acordo com a sequência do códon (Figura 31).

O código genético responsável por elucidar qual aminoácido deve ser trazido pelo RNAt de acordo com a sequência codificante presente no RNAm foi completamente elucidado em 1966, pelos pesquisadores Marshall Nirenberg e Heinrich Matthaei, trabalhando com triplets de UUU, com a maquinaria de tradução de *E.coli* e com

aminoácidos radioativos observaram que poli Us codificava para o aminoácido fenilalanina. Os experimentos posteriores, envolvendo diferentes combinações de bases nos códons, permitiram decifrar o código genético. Dentre as características apresentadas pelo código genético estão o fato de ser degenerado ou redundante, ou seja, um mesmo aminoácido pode ser codificado por códons diferentes; é universal, com poucas exceções é o mesmo nos mais diversos organismos e não é superposto, por ser lido em trincas sequencialmente (Figura 32).

O processo de movimento do ribossomo pela fita de RNAm se repete até que um dos três (UAA, UAG, UGA) sítios de terminação (códons de término ou *stop codons*) seja encontrado nessa fita. Neste caso, o ribossomo espera não um RNAt mas uma proteína conhecida como fator de liberação (RF, do inglês *releasing factor*), que se liga no sítio A ao códon de terminação, desestabiliza o ribossomo desfazendo a maquinaria traducional. Há vários fatores proteicos chamados fatores de iniciação e de alongamento que também colaboram no processo de síntese proteica. O intervalo de sequência do RNAm, localizado entre a posição +1 do RNAm transcrito e o códon AUG de início de tradução, é chamado de região 5'-UTR e a porção que não será traduzida, localizada após o *stop* códon, é chamada de 3'-UTR. Ambas as sequências não traduzidas podem estar presentes em transcritos de genes de procariotos e eucariotos. Considerando a sequência codificante no RNAm descrita na Figura 33, observe as possibilidades de leitura. A provável ORF é aquela mais longa começando por um AUG até um *stop* códon. A Figura 34 apresenta uma visão geral do fluxo da informação genética, considerando as sequências sinais específicas.



## **UNIDADE V - REGULAÇÃO DA EXPRESSÃO GÊNICA EM PROCARIOTOS**

As células de todo organismo, sejam uni ou pluricelulares, apresentam genes que possuem níveis constantes de expressão, chamados de constitutivos ou genes “de economia doméstica” (*housekeeping*) e seus produtos são constantemente requeridos pela célula, os quais permanecem “ligados”. Existem os genes cuja expressão é variável e estão sob influência de sinais moleculares específicos, tais como, fatores de crescimento, hormônios, e outros para que seus produtos sejam produzidos. Se todos os seus genes estivessem “ligados” a todo tempo na célula, haveria um desperdício de proteínas que não estariam sendo utilizadas o tempo todo. O mecanismo de “ligar” e “desligar” genes é conhecido como Regulação da Expressão Gênica e é de importância vital para melhor utilização da energia disponível na célula, dado o alto custo da síntese protéica.

### **EXPRESSÃO GÊNICA EM PROCARIOTOS**

O termo expressão gênica refere-se ao processo em que a informação codificada por um determinado gene é funcional na forma de RNA ou será decodificada em uma proteína. Teoricamente, em qualquer uma das etapas do fluxo da informação genética, que vai do DNA à proteína pode levar a uma expressão gênica diferencial. Os genes podem ser regulados por sinais intrínsecos às células, bem como por sinais extracelulares. Para as bactérias, cujo crescimento pode ser realizado em meios de cultura específicos, esses sinais externos estão disponíveis no meio, e podem desencadear uma série de eventos controlados por ligação a proteínas reguladoras. Essas proteínas podem ser reguladoras positivas da transcrição (ativadores) ou reguladoras negativas (repressores).

Por meio de ligação ao DNA em sítios específicos, as proteínas reguladoras podem assim controlar a expressão de genes específicos, aumentando, bloqueando ou reduzindo a síntese de RNA, sendo que um ativador aumenta as taxas de transcrição do gene que controla, enquanto os repressores reduzem ou anulam a transcrição gênica.

Nas bactérias, o controle da expressão gênica serve, principalmente, para permitir que as células se ajustem às mudanças nutricionais no ambiente, de forma que seu crescimento e divisão sejam otimizados. Muitos genes estão em complexos chamados operons; e nestes complexos, normalmente, existe uma proteína repressora e uma ativadora da transcrição de vários genes que, atuando em “parceria”, determinam certas características.

### **OPERON LAC**

O Operon Lac foi descoberto por François Jacob e Jaques Monod, em 1960, como um mecanismo responsável pela codificação de enzimas, que absorve e digere a lactose do meio. O Operon Lac consiste de um operador, um promotor e três genes estruturais. O RNAm do operon é chamado de policistrônico ou poligênico, pois possui informação para codificar as três proteínas, cujos genes beta-galactosidase, permease e transacetilase são transcritos em um único filamento de RNAm.

Assim, no metabolismo da lactose por *E. coli* a primeira enzima produzida pelo operon é a beta-galactosidase, uma enzima capaz de clivar a lactose em glicose e galactose, que assim servirão como fonte de carbono para a célula (Figura 35).

A beta-galactosidase é uma enzima indutível, ou seja, sua expressão varia de acordo com as necessidades celulares e, caso a bactéria esteja crescendo em meio rico em lactose, sua expressão será alta; no entanto, caso a fonte de carbono seja outro carboidrato,

sua expressão será reduzida. A outra enzima do operon é a permease, que, como seu próprio nome indica, é a enzima responsável pelo transporte de lactose do meio extracelular para o meio intracelular através da membrana da célula bacteriana, pois a lactose, como a maioria dos carboidratos, não é capaz de atravessar a bicamada lipídica sem uma proteína carreadora.

A última enzima, a tiogalactosídeo transacetilase tem função de inativar galactosídeos tóxicos, que também acabam sendo transportados para dentro da célula. A expressão dessas três proteínas é regulada por um outro gene (gene I) que codifica para um regulador Lac. O gene I é capaz de codificar um repressor que está faltando ou está inativado nos organismos com o gene I<sup>-</sup>. A região promotora (P), como visto anteriormente, é o local onde a RNA Polimerase se liga para começar a transcrição. Num operon, além da região promotora existe também o operador (O), os quais formam os locais de controle do operon. Na presença do produto do gene I, o operon é incapaz de ser codificado, pois o repressor está ligado ao operador (Figura 36), no entanto, caso o indutor, a própria lactose esteja presente, essa se liga ao repressor o qual não se liga ao DNA, não bloqueando a síntese das proteínas envolvidas no metabolismo da lactose (Figura 37).

O isopropil-b-D-tiogalactosídeo (IPTG) é um indutor artificial do operon lac. Foi utilizando este composto que se descobriu o repressor lac codificado pelo gene I, que na ausência de um indutor liga-se ao operador e bloqueia a transcrição.

### **Relação Glicose e AMPc**

Se a glicose estiver no meio de cultura da *E. coli*, ela será preferencialmente usada como fonte de energia, isto é, enzimas utilizadas no metabolismo dos outros carboidratos serão pouco ou nada expressos. Os efeitos negativos da glicose sobre os níveis de muitas enzimas são chamados de repressão catabólica. Esse efeito é moderado pela ação da proteína CAP (proteína ativadora de catabólitos), também chamada de CRP (proteína receptora de AMPc) e o metabólito celular AMPc (AMP cíclico). Em bactérias, o AMPc liga-se ao CAP, sendo que esse complexo CAP-AMPc é capaz de estimular a transcrição por se ligarem a determinados promotores. No operon lac, CAP se liga em regiões adjacentes ao promotor em que se liga a RNA polimerase, potencializando as taxas de transcrição.

A concentração de AMPc é controlada indiretamente pela concentração de glicose no meio. Foi observado que quando os níveis de glicose estão altos, os níveis de AMPc estão baixos e não é capaz de se ligar ao DNA. Quando são adicionados AMPc exógeno ao meio, a célula é capaz de metabolizar outros açúcares, como a arabinose e a lactose. É interessante notar que o AMPc em procariontes tem função semelhante ao AMPc em eucariotes. Nesses, quando os níveis de glicose circulante estão baixos, há o estímulo para a produção de glucagon (“hormônio da fome”), o que leva a aumentar os níveis de AMPc dentro das células sensíveis a hormônios.

A ligação da CAP tem dois efeitos, cria uma curvatura no DNA e reforça a fraca ligação entre polimerase e promotor (Figura 38), sendo que ambos os efeitos são importantes para o aumento nas taxas de transcrição. Dessa forma, quando os níveis de glicose estão baixos, a concentração do AMPc é alta e o complexo CAP-AMPc se liga ao DNA melhorando a síntese pelo operon lac (Figura 39).

Os operons podem ser apresentados com regulação positiva e/ou com regulação negativa. Um operon está sob controle regulatório positivo quando os genes são expressos (transcritos) na presença de uma proteína que se liga ao operador e sob controle negativo, quando os genes são expressos na ausência de uma proteína repressora ligada ao operador. O operon lac apresenta os dois modelos regulatórios, sendo o primeiro

representado pela ligação do AMPc-CAP potencializando a transcrição e o segundo representado pela proteína repressora ligada à lactose, a qual se desligará do operador permitindo a ação da RNA polimerase na transcrição do genes Z, Y e A (Figura 40).

### **OPERON ARA**

Assim como a lactose, a arabinose não é utilizada diretamente como fonte de carbono, ou seja, ela precisará ser modificada a xilulose-5-fosfato para ser utilizada no metabolismo. Os genes *araBAD* não são expressos na ausência de arabinose. O operon arabinose (*ara*) codifica para três enzimas necessárias para converter arabinose em xilulose 5-fosfato, um produto intermediário da via pentose fosfato. O operon da arabinose é formado por três genes estruturais que codificam:

- *araA*: arabinose isomerase (converte arabinose em ribulose)
- *araB*: ribulocinase (fosforila ribulose)
- *araD*: ribulose-5-fosfato epimerase (converte ribulose-5-fosfato a xilulose-5-fosfato a qual pode ser metabolizada pela via das pentoses)

O operon *ara* também está sobre controle duplo, sendo que é preciso a presença do complexo CAP-AMPc e a ligação da arabinose à proteína C para que a transcrição seja eficiente. Além do gene regulador *ara C*, que codifica a proteína C, existe uma região controladora com operador (O) e outra reguladora (*ara I*). No operon *ara*, a proteína C apresenta regulação positiva, atuando como ativador, e negativa, inibindo o *araBAD*, dependendo das condições celulares. O fator da chave liga/desliga para esse operon é a própria arabinose. Diferente do Operon Lac, na presença de arabinose, esta se ligará à proteína C a qual se ligará à região I do DNA, permitindo a transcrição gênica também na presença do complexo CAP-AMPc. Na ausência de arabinose, a proteína C também se liga à região I do DNA, fazendo com que ocorra uma dobra (alça) na molécula para que a mesma proteína C interaja com o operador O, não permitindo que a RNA polimerase inicie a transcrição do Operon Ara (Figura 41).

### **OPERON TRP**

No Operon Trp, diferente dos dois tipos de regulação apresentados anteriormente, a ligação do triptofano torna o repressor ativo para controlar a síntese das cinco enzimas envolvidas na biossíntese do triptofano a partir do corismato. Para melhor entender esse modelo de operon é importante lembrar que em procariotos a transcrição e a tradução são processos acoplados, sendo que a tradução começa antes do término da transcrição. Os genes estruturais E, D, C, B, A codificam as enzimas que atuam na biossíntese do aminoácido triptofano. Além desses genes, a sequência líder, localizada imediatamente antes do gene E, atuará de forma a regular a transcrição das cinco enzimas. Como os demais, esse operon apresenta a sequência promotora seguida da sequência operadora (Figura 42a).

O repressor *trp* codificado pelo gene *trp* é uma proteína alostérica de ligação ao DNA. No controle alostérico ocorre alteração na estrutura terciária ou quaternária de uma enzima induzida pela ação de uma molécula ligante, que pode ser um ativador, um inibidor, um substrato, ou os três. Quando não há triptofano, o repressor está em uma conformação que é incapaz de se ligar ao DNA, no entanto, quando o repressor se liga de forma reversível ao triptofano, este adquire uma conformação que faz com que se liga firmemente ao sítio operador. Desse modo, os níveis de triptofano na célula podem controlar a transcrição de genes necessários para sua própria síntese:

↓ [triptofano]: proteína repressora em conformação incapaz de se ligar ao DNA.

↑ [triptofano]: ligação do triptofan à proteína repressora, que será capaz de se ligar ao sítio operador no DNA pela mudança conformacional após ligação ao ligante, levando à redução da síntese dos genes estruturais.

Assim, o triptofano controla a sua própria síntese. Se há excesso de triptofano, este se liga ao repressor, o qual adquire uma mudança conformacional ligando-se ao operador. O operador se superpõe ao promotor para o início de transcrição, ou seja, o repressor, que está ligado ao operador, impede que a RNA polimerase se ligue eficientemente ao promotor *trp* impedindo a transcrição dos genes estruturais envolvidos na biossíntese do aminoácido triptofano (Figura 42b).

### **Regulação por atenuação: importância da sequência líder**

Uma característica importante dos procariontes é a ausência de compartimentalização nas células, ou seja, não existe separação das estruturas celulares por membranas. Sendo assim, o DNA procariótico está solto no citoplasma permitindo que os processos de transcrição e tradução ocorram simultaneamente. A atenuação faz uso desse rígido acoplamento entre esses dois processos para controlar a expressão gênica.

A regulação de uma terminação prematura da transcrição, por meio de uma estrutura secundária alternativa do transcrito proveniente da sequência líder, é chamada de atenuação e regula a expressão de operons envolvidos na biossíntese de aminoácidos. Esses operons codificam sequências líderes no RNAm produzido, que são traduzidas formando estruturas secundárias dependentes da tradução. No caso do Operon *Trp*, esse peptídeo líder de 14 aminoácidos possui dois códons para triptofano, no início da sequência, que exercem um papel muito significativo na regulação. Quando há pouco triptofano, o ribossomo fica parado nos códons UGG repetidos aguardando a chegada de um novo RNAt ligado ao triptofano. O ribossomo parado no início da molécula de RNAm (região denominada 1) permite que os nucleotídeos das regiões 2 e 3 façam estruturas em alças, não ocorrendo o término da transcrição pela falta de pareamento entre 3 e 4. Ao contrário, em altas concentrações de triptofano, as sequências 3 e 4 fazem uma estrutura secundária em grampo (pareamento Gs e Cs) seguida de Us, indicando término da transcrição (Figura 43). A sequência líder do Operon *Trp* apresenta 162 nucleotídeos e está localizada antes do códon iniciador de *trpE*.

Vários outros operons para a biossíntese de aminoácidos em *E. coli* são hoje conhecidos como contendo sequências atuando como “atenuadores”. O peptídeo líder de cada um deles contém uma abundância de aminoácidos do tipo controlado pelo operon. Por exemplo, o operon de treonina codifica enzimas que sintetizam tanto treonina quanto isoleucina. O peptídeo líder contém oito treoninas e quatro isoleucinas em uma sequência de 16 aminoácidos. Sete fenilalaninas estão presentes no líder de 15 aminoácidos do operon da fenilalanina. Ainda mais marcantes são as sete histidinas em fila encontradas no peptídeo líder do operon de histidina.

## **UNIDADE VI - REGULAÇÃO DA EXPRESSÃO GÊNICA EM EUCARIOTOS**

Os organismos multicelulares eucariotos, como os animais, por exemplo, passam por diversas etapas de diferenciação desde a formação do zigoto até que o indivíduo se torne adulto. A diferenciação ocorre também em nível gênico, pois apesar de todas as células possuírem os mesmos cromossomos e; portanto, os mesmos genes, as necessidades de produtos gênicos são variáveis nos diversos pontos de diferenciação, havendo também a necessidade de mecanismos para “ligar” e “desligar” genes em épocas do desenvolvimento, condições fisiológicas e ambientes diferentes.

Assim, da mesma forma que as necessidades celulares para proteínas diferentes variam, os mecanismos pelos quais os genes são regulados também variam, sendo de vital importância o mecanismo de regulação do metabolismo celular para que a célula utilize da melhor forma a energia disponível.

Em eucariotos, o mecanismo de regulação dos genes é mais complexo, comparado aos procariotos. Isso é explicável quando se observa o maior tamanho do genoma de eucariotos e a compactação do DNA, que é compartimentado em um núcleo. Como dito anteriormente, alguns genes requisitados mais constantemente nas células são os genes *housekeeping*, responsáveis pela rotina metabólica, ou seja, em funções como a respiração, e que sejam comuns a todas as células.

Outra classe de genes são os expressos quando a célula entra em uma via metabólica específica de diferenciação. Nessa classe, por exemplo, estão os genes associados ao desenvolvimento encontrado desde *Drosophila* até humanos, e parece ser o gene principal responsável pela regulação de vários outros genes, sendo também descrito como *gene master*. Esses genes regulados foram primeiramente identificados por mutações em drosófila, sendo responsáveis pela transformação de uma parte do corpo em outra (por exemplo, uma mutação no DNA que leve à formação de pés no lugar de antenas). Em drosófila, a partir de estudos mutacionais, foram caracterizados dois grupos principais de genes homeóticos: o complexo antenapedia (regula os segmentos da cabeça e tórax) e bitórax (regula os segmentos torácicos posteriores e abdominais). Os genes homeoboxes, em humanos chamados *HOX*, possuem 60 a 80% de homologia entre organismos diferentes e codifica um domínio protéico (homeodomínio), com cerca de 60 aminoácidos (Figura 44).

Algumas outras classes de genes são expressas a “todo o tempo” apenas naquelas células que já se diferenciaram, como as células do plasma que expressam continuamente genes que sintetizam os anticorpos. Existem também genes que expressam por alterações nas condições no microambiente celular, como, por exemplo, a chegada de um hormônio por associação a seu receptor pode ligar ou desligar certos genes naquela célula específica.

Após o sequenciamento do Genoma Humano (PGH), o número estimado de genes de uma célula eucariótica é de cerca de 25 mil genes. O número de genes obtidos pelo PGH é muito inferior ao esperado, em que deveria haver um gene para cada uma das 100 mil proteínas. Com base nessa quantidade de genes, uma pergunta que surge é como a expressão do gene é regulada em eucariotos para atender as necessidades celulares? O mecanismo de *splicing* alternativo de um gene está implicado nessa resposta pela produção de uma grande diversidade de moléculas de RNAm variantes, que codificarão as diferentes isoformas de proteínas para atender as necessidades de tecidos e células diferentes.

Além do mecanismo de controle de expressão dada pela correta recomposição das moléculas de RNAm, sendo esse mecanismo regulatório em nível de pós-transcrição; a

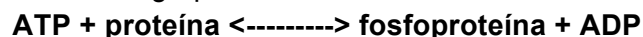
regulação da expressão gênica pode ser potencial também nos níveis transcricionais pelo correto posicionamento e ligação dos GTFs e de outras proteínas; na degradação do RNAm pós-splicing para reciclagem de nucleotídeos; em níveis traducionais para a produção de proteínas; pós-traducional e na degradação protéica pós-tradução para reciclagem de aminoácidos (Figura 45).

A partir desse esquema geral de expressão dentro da célula, podem ser definidos alguns termos provenientes da era Genômica e que estão diretamente ligados às novas ciências ômicas: i) proteoma: conjunto completo de proteínas que podem ser expressas pelo material genético; ii) transcriptoma: coleção completa de sequências transcritas do genoma, compreendendo os RNAsm e ncRNAs (RNAs não codificantes, do inglês “*non coding RNA*”) e por fim; iii) genoma: conjunto de material genético em um conjunto cromossômico

## EVENTOS DE REGULAÇÃO PÓS-TRADUCIONAIS

O ambiente aquoso dentro da célula não favorece o *dobramento correto das diferentes proteínas* recém-sintetizadas (*protein folding*). Essas proteínas requerem o dobramento pelas interações dos diferentes aminoácidos, para que adquira sua conformação tridimensional nativa. Dentre as atividades da proteína nativa está a sua atividade enzimática, de ligação ao DNA e papéis estruturais na célula. Nessas situações, o dobramento é feito pela ajuda de outras proteínas, chamadas de chaperonas (Figura 46).

No que se refere às *modificações da proteína dentro da célula*, existem mais de 300 modificações de cadeias laterais de aminoácidos que podem ocorrer após a tradução da proteína. As mais comuns são a fosforilação e ubiquitização. A fosforilação é um evento reversível, cuja ação é atribuída às enzimas cinases (quinases) que adicionam grupos fosfatos em resíduos de serina, treonina e tirosina de proteínas específicas. Ao contrário, as enzimas fosfatases removem os grupos fosfatos.



Esse processo está envolvido numa variedade de eventos celulares, como, por exemplo, na interação proteína-proteína (interactoma, que é o conjunto completo de interações entre proteínas); na atividade enzimática; na interação proteína-DNA/RNA (spliceossomo, ribossomo, replissomo, primossomo, dentre outras).

Outra classe de modificações que podem acontecer nas proteínas é a *ubiquitinação*. Esse evento corresponde à adição de uma proteína ubiquitina (altamente conservada entre plantas e animais) em aminoácidos lisina de outras proteínas. Após essa adição a proteína fica marcada para degradação por proteases dentro da maquinaria do proteassomo (Figura 47).

## PROTEÍNAS REGULATÓRIAS

As proteínas regulatórias podem atuar como coativadores, correpressores e transativadores. Os coativadores correspondem a uma proteína ou complexo protéico em que outras proteínas regulatórias, como os GTFs da RNA Polimerase, podem se ligar aumentando as taxas de transcrição. Os correpressores atuam em conjunto com uma proteína repressora para reprimir as taxas de transcrição, não sendo ligados diretamente ao DNA. Os fatores de transcrição (TFs), por sua vez, atuam por se ligarem a elementos reguladores em cis (descrito abaixo) e direta ou indiretamente afetam o início da transcrição, sendo denominados de transativadores.

Para aquelas proteínas que se ligam diretamente ao DNA, os mecanismos de interação envolvem fracas ligações entre as cadeias laterais dos aminoácidos e as bases nitrogenadas. As proteínas que se ligam ao DNA podem ser agrupadas em um pequeno número de famílias, com base nas suas características estruturais. Os motivos que dão nomes a essas famílias correspondem a pequenas seções das proteínas.

As proteínas normalmente se ligam ao DNA na forma de homo ou heterodímeros, que correspondem a subunidades idênticas ou diferentes subunidades de proteínas, respectivamente. Entre os domínios de ligação ao DNA apresentado pelas proteínas regulatórias estão:

i) hélice volta hélice (*helix-turn-helix*): foi o primeiro tipo identificado ocorrendo comumente nas proteínas de procarionotes e apresenta em sua estrutura duas  $\alpha$  hélices ligadas por uma volta  $\alpha$ ;

ii) dedo de zinco (*zinc-finger*): são comuns em proteínas de regulação gênica de eucariotos, que apresentam vários desses domínios e apresenta um átomo de zinco ligado a resíduos de histidina e/ou cisteína;

iii) zíper de leucina: encontrados em muitos fatores de transcrição de eucariotos, a molécula é uma  $\alpha$  hélice, com resíduos de leucina sequenciais e duas subunidades semelhantes (homodímeros) ou diferentes (heterodímeros), interagem por forças hidrofóbicas entre as cadeias laterais do aminoácido leucina formando uma estrutura semelhante a zíper ou pinça que segura o DNA inserindo suas  $\alpha$ -hélices (Figura 48).

## REGULAÇÃO DA EXPRESSÃO GÊNICA EM NÍVEL PÓS-TRANSCRICIONAL

Dentro os mecanismos de regulação em nível de pós-transcrição estão o processamento do RNAm, necessário para que a molécula de RNAm final seja funcional, a recomposição alternativa que ocorre em muitos tecidos para atender as diferentes necessidades dos tecidos para as proteínas, já estudados na Unidade III. Dentre outros eventos que ocorrem em nível de regulação pós-transcricional estão a edição do RNA (*RNA editing*) e o silenciamento gênico pós-transcricional por RNAs de interferência.

A edição do RNA (*RNA editing*), assim como o *splicing* alternativo, é importante na produção de proteínas que exercem função tecido-específica para a produção de diferentes isoformas a fim de atender a necessidade tecidual. No entanto, essas isoformas são produzidas pela atuação de uma enzima que atua na edição do RNAm maduro alterando codons na molécula de RNA. Um exemplo pode ser dado para as proteínas de baixa densidade como a apolipoproteína (apo), as quais são requeridas pelo fígado e pelo intestino, ambas envolvidas no metabolismo de lipídeos. O RNAm no fígado codifica para a proteína apoB-100 (sem edição), envolvida no transporte do colesterol e dos triglicerídeos produzidos por síntese endógena. No intestino, esse RNAm é editado pela ação da enzima citidina-desaminase, que converte a base "C" em "U" levando à formação de um códon de parada no meio do RNA e a proteína formada é uma apoB-48 (mais curta), envolvida no transporte dos lipídeos provenientes da dieta para os diversos tecidos. A desaminação ocorre assim, de modo tecido-específico e a mensagem é editada nas células intestinais, mas não nas células hepáticas. Como produto final de tradução, o códon CAA no fígado é traduzido como glutamina e no intestino os nucleotídeos "C" sofre desaminação, gerando um códon de término de tradução UAA (Figura 49).

Os silenciadores gênicos pós-transcricionais (PTGS) constituem numa molécula de RNA que apresenta homologia com outra molécula de RNA, transcrita por um determinado gene. Assim, quando a molécula de RNA é transcrita, a outra molécula homóloga se liga ao RNA, impedindo que o gene se expresse. Em experimentos realizados com o nematódeo *C.*

*elegans*, pelos pesquisadores norte-americanos Andrew Fire e Craig Mello, contemplados com o Nobel de medicina/fisiologia em 2006, foi observado que a introdução de RNAds (fita dupla; abreviação de *double-stranded*) na célula reduzia a expressão de genes que continham sequências idênticas (ou muito semelhantes) a uma das fitas desse RNAds. Esse mecanismo ficou conhecido como regulação da expressão gênica por RNAs de interferência (RNAi) e os detalhes envolvidos nessa regulação foram posteriormente descobertos. Os RNAi endógenos, ou seja, produzidos pela própria célula foram chamadas de miRNAs (microRNAs) para diferenciar dos RNAi que têm o mesmo efeito supressor de expressão gênica, no entanto, de origem exógena, chamados de siRNAs (*short interfering RNA*).

Os siRNAs provenientes de origem exógena (vírus ou artificial) apresentam um completo pareamento com o RNAm alvo frequentemente em apenas um sítio, enquanto os miRNAs são formados de um gene próprio a partir de exon, intron ou de síntese de RNA de regiões intergênicas e apresentam um pareamento incompleto podendo ser complementar a múltiplos sítios de um alvo. No núcleo, os precursores de miRNAs são longos, com cerca de 70 a 90 nucleotídeos de comprimento, e após serem processados em precursores menores (pré-miRNA) pela enzima Drosha (função de RNase) são transportados para o citoplasma pela proteína exportina. Os siRNAs não sofrem esse transporte. Os siRNAs não têm sido encontrado naturalmente em mamíferos, contudo a molécula relacionada, miRNA, tem sido amplamente descrita em vários organismos e tipos celulares.

No citoplasma, as moléculas de dsRNA são reconhecidas por uma enzima com função de RNase, a *Dicer*, a qual produz os siRNAs/miRNAs. Os RNAi clivados apresentam 21-23 nucleotídeos de dsRNA. Esses RNAi são incorporados em um complexo multi-protéico, o RISC (complexo de silenciamento gênico induzido por RNA), levando ao rompimento da dupla fita do dsRNA e uma das fitas é guiada pelo RISC a seu alvo de RNA, que apresenta homologia para clivagem endonucleolítica ou impedimento da passagem dos ribossomos durante o processo de tradução.

O que podemos observar aqui, é que, embora as proteínas sejam efetoras fundamentais da função celular, a base da complexidade e variação fenotípica pode primariamente ser devido a um sistema paralelo de informação necessária para a coordenação e modulação da expressão gênica por RNAs. Dentre dessa complexidade de RNAs vistos até agora estão os RNAs não codificantes (RNAnc) funcionais, como os RNAr, RNAt, snRNAs, ribozimas, miRNAs e siRNAs. Em humanos, aproximadamente 98% de todos os produtos transcritos correspondem a RNAnc.

O mecanismo da RNAi pode ter evoluído para proteger as células de elementos infecciosos ou desestabilizadores de genes, como certos vírus e transposons, que utilizam em seu sistema de replicação um dsRNA intermediário. Alguns vírus de ambos, plantas e animais, possuem um genoma de dsRNA. Em muitos outros vírus seus genomas são de RNA, que ao entrar na célula hospedeira é rapidamente convertida em dsRNA. Então o RNAi pode ser uma arma poderosa para conter a infecção por estes vírus através da destruição dos RNAm e assim bloquear a síntese de proteínas virais essenciais. Os transposons consistem em um tipo de DNA localizado no genoma de organismos complexos, com propriedades de saída do DNA e reinserção em outro local, podendo causar danos se eles se inserem em genes ou sequências regulatórias importantes. Os retrotransposons tem seu DNA copiado em RNA, os quais são transcritos reversamente em DNA e parte das moléculas dsRNAs são, assim, alvo de interferência pelas enzimas DICER e RISC, e os RNAi gerando podem destruir estes transposons impedindo sua inserção no genoma que pode provocar ruptura em genes essenciais (Figura 50).



## REGULAÇÃO TRANSCRICIONAL

O mecanismo de regulação em eucariotos está basicamente sob um controle positivo. Isso porque, se os genes estivessem sob um controle negativo, haveria necessidade da síntese de aproximadamente 100 mil repressores para se ligarem aos operadores gênicos e impedirem a transcrição das cerca de 100 mil proteínas, que seria o número estimado de proteínas anteriormente à conclusão do Projeto Genoma Humano. Considerando um controle positivo, as proteínas regulatórias podem se ligar aos sítios reguladores, estimular a transcrição e se desligar ligando-se em outro local.

Além dos eventos necessários à transcrição de um gene descritos na Unidade III, importante aqui é a observação de que existem nesse mecanismo regulatório os elementos de atuação em “cis” e os elementos regulatórios de ação em “trans”. Os elementos regulatórios em “cis” são aqueles que desempenham funções regulatórias da expressão gênica no mesmo cromossomo da sequência de um gene que irá ser transcrito, como as regiões promotoras dos genes, seus elementos proximais e os distais, destacando-se os reforçadores ou *enhancers*. Os silenciadores também são sequências em “cis” localizadas na molécula de DNA, cuja função é reduzir ou reprimir as taxas de transcrição de um gene. Os elementos em “trans” são aqueles que desempenham funções regulatórias a distância da sequência pela qual foram codificados, atuando normalmente em cromossomos diferentes dos quais foram produzidos. Os *transativadores* correspondem às proteínas produzidas por RNA expressos de cromossomos diferentes daqueles em que irão regular, por isso a denominação “trans”. Dentro dessa classe estão as proteínas que atuam como fatores de transcrição (TFs), que são capazes de recrutar outras proteínas, como os coativadores (Figura 52), visto que a afinidade da RNA polimerase ao promotor é baixa, havendo a necessidade de ligação das proteínas regulatórias para aumentar a afinidade da RNA polimerase ao promotor e iniciar a transcrição. Para regular os níveis de transcrição em eucariotos, o modelo é complexo envolvendo a regulação gênica por cascata protéica dada pelos fatores gerais de transcrição (GTFs), como visto na Unidade III.

Em síntese, os *enhancers* funcionam como intensificadores, promovendo a transcrição de modo mais eficiente, e os silenciadores, que são as regiões onde proteínas repressoras se ligam inibindo a transcrição dos genes correspondentes. Nos *acentuadores* se ligam as proteínas coativadoras, que estimulam a transcrição dos genes correspondentes, sendo alvo de fatores específicos. Os *acentuadores* podem estar localizados acima do gene (*upstream*), abaixo do gene (*downstream*), ou mesmo dentro do gene. Frequentemente, os *enhancers* levam a uma curvatura do DNA devido a interações com as proteínas ligadas à região promotora. Em leveduras, esses elementos estimuladores de transcrição foram descritos como UAS (*upstream activating sequences*). Essas sequências estão cerca de 100 bases acima do início da transcrição e são equivalentes aos *acentuadores* dos eucariotos mais complexos (Figura 51).

O balanço entre a ligação de ativadores e de repressores aos *enhancers* e aos silenciadores, respectivamente, auxilia no controle das taxas de transcrição dos genes de eucariotos. O que realmente faz com que o gene seja transcrito é a combinação única e particular do promotor e de fatores de transcrição.

Os hormônios e seus fatores de transcrição representam uma classe de reguladores da expressão gênica em “trans” de ligação a sequências denominadas elementos de resposta a hormônio (HRE) (cis), que correspondem aos sítios de ligação no DNA dos receptores proteicos de ligação a hormônios. As proteínas receptoras de hormônio se localizam no citoplasma, mas atuam no núcleo da célula, possuindo domínio de ligação a ligante e domínio de ligação ao DNA, levando à transcrição de genes de resposta a

hormônios. Dentre os exemplos dessa classe estão as proteínas receptoras de andrógeno (AR). Essa proteína pertence à família dos receptores nucleares e possui dois domínios dedos de zinco de ligação ao DNA. No citoplasma está na sua forma inativa associada às chaperonas que impedem a sua degradação. Quando a testosterona entra na célula, no citoplasma é convertida em DHT (dihidrotestosterona) pela ação da enzima 5-alfa-redutase, e a DHT se liga ao domínio de ligação a ligante do receptor de andrógeno. O AR complexado à DHT passa para o núcleo e na forma de homodímero levam à transcrição dos genes regulados por hormônio pela ligação ao ARE (elemento de resposta a andrógeno) (Figura 53).

### COMPACTAÇÃO DO DNA EUCARIÓTICO

A associação do DNA de eucariotos às proteínas básicas histonas (H1, H2A, H2B, H3 e H4) dá ao material genético da célula uma maior condensação, formando a sua estrutura secundária. A união de duas proteínas H2A, duas H2B, duas H3 e duas H4 envolvidas por duas voltas de DNA, formam a unidade chamada de nucleossomo, que são ligados entre si pela histona H1. Assim, o enrolamento do DNA ao redor das histonas contribui para o empacotamento, reduzindo a sua extensão linear. A formação de uma estrutura mais condensada na metáfase mitótica, contendo várias voltas de DNA ao redor de várias histonas, resulta na formação de um solenoide (estrutura de 30nm) tendo em vista a associação dos nucleossomos (Figura 54). Em condições experimentais, o desprendimento da maioria das proteínas do DNA nos cromossomos mitóticos favorece a visualização “direta” de *loops*. Quando o núcleo está desprovido de proteínas, o DNA lança os *loops* que permanecem fixados a uma matriz proteica residual, denominada arcabouço (*scaffolding*) que confere aos cromossomos a sua estrutura peculiar vista na metáfase. A formação desses *loops* parece ser necessária para a ocorrência de transcrição.

A cromatina pode ser dividida em eucromatina e heterocromatina. A eucromatina é a região no cromossomo rica em genes e transcricionalmente ativas (*eu*=verdadeira). Outras regiões da cromatina mais densamente condensada recebem o nome de heterocromatina, com pouca ou nenhuma atividade transcricional. Por sua vez, a heterocromatina é dividida em constitutiva, contendo regiões não expressas, como, por exemplo, algumas sequências repetidas no DNA, que desempenham papel estrutural no cromossomo, e em heterocromatina facultativa, que pode ser expressa em certos tipos celulares e em outros não.

### MODELAGEM DE NUCLEOSSOMOS

A condensação do material genético está relacionada com a inativação de expressão nos cromossomos. As primeiras evidências de modelagem da cromatina associada à expressão gênica vieram dos estudos com dípteros e anfíbios. A modelagem da cromatina corresponde à modificação da estrutura local da cromatina ao redor das sequências reguladoras de genes.

Nos dípteros os cromossomos dos tecidos larvares, como glândulas salivares, ajudaram a esclarecer a regulação gênica em eucariotos, pois as células nesses tecidos não se dividem e, assim, seus cromossomos se replicam, mas permanecem pareados aos homólogos, produzindo os cromossomos politênicos. As análises desses cromossomos em vários estágios do desenvolvimento do inseto permitiram a observação de áreas específicas de espessamento formando os puffs ou anéis de Balbini. O Puff é descrito como um afrouxamento do DNA altamente helicoidizado em estruturas longas em formas de alças, que surgem e desaparecem em determinados tecidos em locais específicos do cromossomo

à medida que ocorre o desenvolvimento, indicando que cada Puff é um sítio ativo de transcrição. A “puffagem” temporal indica claramente mudanças na atividade gênica ao longo do tempo.

Um fenômeno similar de “ligar” e “desligar” genes foi evidenciado em cromossomos dos ovócitos de anfíbios, que foram chamados de cromossomos plumosos. Esses cromossomos foram isolados na prófase meiótica, onde os homólogos eram mantidos juntos por quiasma, e não se condensavam, ao contrário, eram bastante longos e distendidos. Cada par de cromátides irmãs forma uma série de pares de alças conectada por um fino eixo central. As alças dos plumosos estão ativamente envolvidas na síntese do RNA. Como tal, elas são estrutural e funcionalmente similares aos Puffs observados nos cromossomos politênicos. Em ambos os casos, a compactação do cromossomo foi relaxada, permitindo assim que a transcrição ocorra.

### **ACETILAÇÃO E DESACETILAÇÃO DE HISTONA**

Para que a RNA polimerase tenha acesso aos promotores dos genes alvo de transcrição, a cromatina deve sofrer modificações, de forma a mover o octâmero de histonas o que levará à mudança nas posições dos nucleossomos. Assim, os nucleossomos não são estáticos na cromatina, permitindo a sua remodelagem. Alguns coativadores podem fazer essa remodelagem. Uma das modificações que pode ocorrer na cromatina, para que os genes estejam transcricionalmente ativos, é a acetilação de histonas na sua porção N-terminal que fica fora do nucleossomo (cauda de histonas) em aminoácidos lisinas. Nas histonas hiperacetiladas, o grupo acetil é adicionado por enzimas histonas acetil transferases (HAT) resultando em genes transcricionalmente ativos, uma vez que o grupo acetil altera a interação das histonas com o DNA e o octamero desliza para uma nova posição no DNA. Ao contrário, quando as histonas estão hipoacetiladas, o grupo acetil é retirado por enzimas histonas desacetilases (HDAT) resultando em repressão gênica (Figura 55).

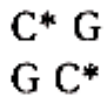
### **REGULAÇÃO EPIGENÉTICA**

A regulação epigenética pode ser definida como o conjunto de modificações nucleares que modulam a expressão gênica sem modificações na sequência de nucleotídeos do DNA, podendo ser herdada. Os fatores ambientais podem levar a mudanças nos padrões epigenéticos, tendo assim influências sobre o fenótipo. Essas modificações incluem a acetilação das histonas, bem como modificações da estrutura da própria cromatina devido à metilação no DNA, principalmente em bases citosinas.

A metilação é o principal mecanismo epigenético, resultante de uma reação química que consiste na adição de um grupo metila (-CH<sub>3</sub>) a uma molécula de DNA, em que um átomo de hidrogênio da base é substituído por um grupo metil. A metilação está assim envolvida nos processos de regulação da estrutura da cromatina, estabilidade genômica e na patogênese de doenças, especialmente o câncer. Um dos mecanismos mais conhecidos de metilação do DNA é o responsável pela inativação do cromossomo X em fêmeas de mamíferos para a compensação de dose de genes presentes no cromossomo X, conhecido como Teoria de Mary Lyon. Em alguns vertebrados e em *Drosophila*, não ocorre metilação do DNA.

No DNA, a metilação consiste na transferência de grupos metil a algumas das bases citosinas (C) situadas prévia e contiguamente a uma guanina (G). Nos mamíferos, o grupo metil é acrescentado à citosina em um dinucleotídeo CG. O padrão de metilação é denominado metilação simétrica porque os grupos metil são encontrados em ambos os

filamentos no mesmo contexto, conforme abaixo (asterisco corresponde à metilação):



Cerca de 70 a 80% de todos os dinucleotídeos CG são metilados no genoma como um todo. A maioria dos dinucleotídeos CG não metilados é encontrada em aglomerados perto dos promotores gênicos. Essas sequências de bases CGs, localizadas acima do promotor gênico e que são passíveis de metilação, são chamadas de ilhas CPGs, em que o “p” representa a ligação fosfodiéster. Estas ilhas, com exceções, parecem permanecer não-metiladas em todos os tipos de células. O resultado da metilação de Cs será o silenciamento dos genes, provocando alterações nos níveis transcricionais de um gene sem necessidade de que se produza uma alteração na sequência do DNA. Para esse processo, há a atuação das enzimas DNA-metiltransferases (adicionam metil ao DNA) e desmetilases (removem esses grupos do DNA).

### **IMPRINTING GENÔMICO**

Descoberto há quase 20 anos em mamíferos, o *imprinting* genômico faz com que alguns genes autossômicos apresentem um padrão incomum de herança. Isso se deve a um *imprinting parental*, em que os genes imprintados são expressos na célula como se fossem de cópia única, caracterizando um modelo de herança monoalélica.

Os alelos de um gene podem ser ativos ou inativos na prole devido ao DNA dos genes terem sido imprintados de forma dependente de qual genitor foi herdado. A base do *imprinting* é que as sequências de DNA de genes imprintados é metilado. O *imprinting* é um processo reversível, em que um alelo de origem paterna submetido ao *imprinting* (e que, por exemplo, não deverá ser expresso na prole), quando herdado por uma mulher, deve ser convertido na sua linhagem germinativa e não estar metilado, de modo que, quando ela transmitir esse alelo à sua prole, o mesmo seja reconhecido como sendo de origem materna (que deve ser expresso na prole).

Um exemplo desse modelo de *imprinting* foi observado no gene IGF-2 (fator de crescimento semelhante à insulina) em camundongos. Esse gene sofre *imprinting* materno, ou seja, o gene a ser expresso deve ser sempre o alelo herdado do pai para que o camundongo tenha o fenótipo normal. Embora o alelo materno seja perfeitamente normal em sequência, este será inativado por metilação, e, caso aconteça ao contrário, e esse gene de origem paterna seja imprintado e o da mãe seja expresso, o indivíduo será raquítico, crescendo menos da metade do tamanho normal (Figura 56).

Em humanos, os exemplos mais bem estudados de *imprinting* são as síndromes de Prader-Willi e Angelman, em que ocorre deleção no cromossomo 15 (15q11-q13) em 70% dos casos. De uma forma simplificada, considerando o *imprinting* materno e deleção nessa porção gênica do cromossomo paterno, ocorre a síndrome de Prader-Willi. Nesse caso, o cromossomo 15 materno deveria contribuir para o fenótipo normal; no entanto, está imprintado e o paterno apresenta uma deleção. Essa síndrome pode afetar ambos os sexos e, dentre as características dos portadores, estão a obesidade, as mãos e pés pequenos, baixa estatura e retardo mental. Ao contrário, se o *imprinting* é paterno e o gene da mãe apresenta deleção nessa mesma porção cromossômica, ocorre a síndrome de Angelman. Nesse caso, o alelo paterno normal está imprintado (silenciado) e o materno, que deveria expressar o alelo normal, está com a deleção. Dentre as características que os indivíduos com a síndrome de Angelman apresentam estão aspecto facial incomum, baixa estatura,

grave retardo mental, convulsões, disposição feliz imotivada (“marionete feliz”) e irritabilidade.

## **UNIDADE VII - BASES GENÉTICAS DO CÂNCER**

O câncer é uma doença genética, independentemente de ocorrer de forma esporádica ou hereditária, pois a carcinogênese sempre inicia com danos no DNA. Geralmente, esses danos são potencializados por agentes químicos, físicos ou virais, denominados de carcinógenos. Dentre os exemplos de vírus potencialmente promotores oncogênicos em humanos estão o HBV (hepatite B) e o HPV (papiloma vírus) associados ao carcinoma hepatocelular (tumor de fígado) e ao câncer de cólo do útero em mulheres, respectivamente. Dentre os fatores ambientais estão fumo, bebida, exposição solar, fatores hormonais, dentre outros.

Qualquer célula normal pode ser sítio de origem de um processo neoplásico, mas para que este aconteça é necessário uma série de eventos que vai sendo acumulada com o passar dos anos, por isso, a idade é um dos fatores de risco mais bem estabelecido associado a essa doença. A formação das neoplasias se dá pelo desequilíbrio entre as taxas de proliferação celular e a apoptose resultado de uma morte celular programada. Esses eventos são regulados por uma grande quantidade de genes, que, ao sofrerem mutações, podem ter seus produtos expressos de maneira alterada, iniciando a formação de um tumor. Portanto, o câncer é uma doença de múltiplas etiologias.

Um tumor corresponde a alterações celulares que levam ao aumento no volume celular e tecidual, podendo ser benigno ou maligno. No tumor benigno, as células crescem mais lentamente e são diferenciadas, se assemelhando às do tecido normal. Já no tumor maligno, as células crescem mais rapidamente e são indiferenciadas (menos especializadas), apresentam alterações morfológicas, com menor adesão célula a célula e maior invasividade com potencial metastático. A metástase é um processo pelo qual as células cancerosas invadem e se espalham para outros tecidos do corpo por via linfática ou sanguínea. A origem de um câncer é dita monoclonal, pois a célula alterada escapa ao controle regulatório, formando uma massa de células alteradas.

Os cânceres podem ocorrer em diferentes tecidos do corpo, sendo aqueles que ocorrem em células epiteliais (de revestimento de mucosas das vias aéreas, tubo digestório, glândulas como tireoide, mama e próstata) são denominados carcinomas, que incluem os adenocarcinomas. Entre aqueles que acometem as células de tecidos ósseo, conjuntivo e muscular estão os sarcomas. Para as células hematológicas brancas estão as leucemias, que acometem as células da medula óssea, os linfomas (células hematológicas do sistema linfóide) e os mielomas (raros), que começam nas células do sistema imunológico, cânceres do sistema nervoso central, que começam nos tecidos do cérebro e da medula espinhal, dentre outros tipos. De modo geral, os diferentes tipos de câncer correspondem às diferentes células do corpo. No que se refere à cura dessa patologia destaca-se a importância na realização dos exames periódicos a fim de detectá-lo precocemente, pois muitos cânceres são curáveis, desde que tratados nos estágios iniciais.

No Brasil e EUA, os cânceres mais comuns são o de próstata nos homens, mama nas mulheres e de pulmão em ambos os sexos. A predisposição para desenvolver o câncer pode ser herdada, sendo denominada de câncer familiar. Contudo, a maioria dos casos podendo atingir 99% (dependente do tipo de câncer), se deve ao acúmulo de mutações espontâneas nos tecidos somáticos, denominadas de esporádicas.

Para estudos, *in vitro*, as células tumorais podem ser removidas do tecido tumoral e

crescidas em meio contendo nutrientes adequados, às vezes indefinidamente. Como apresentam crescimento desregulado e não responsivos aos sinais químicos que inibem a divisão celular, *in vitro*, as células cancerosas apresentam perda de inibição por contato em que as células, ao se tocarem, continuam formando camadas com superposição umas às outras se acumulando no meio de cultura, formando massas; ao contrário das células normais que crescem formando monocamadas. Em geral, também apresentam um citoesqueleto desorganizado podendo produzir proteínas incomuns e apresentar anormalidades cromossômicas. Uma vez estando o DNA danificado ou alterado por mutações, há três processos que podem ocorrer na célula para correção do erro: i) o reconhecimento e parada no ciclo celular para reparo do dano; ii) a morte celular pela ativação da apoptose; iii) mais raramente, a transmissão do dano para as células descendentes por falhas nos outros mecanismos, levando ao processo de oncogênese. Existem, basicamente, duas categorias de genes envolvidos nas formações neoplásicas: os oncogenes e os genes supressores de tumor.

### **PROTO-ONCOGENES E ONCOGENES**

O controle das atividades celulares normais é feito por muitos genes, entre estes, estão os proto-oncogenes, que se transformam em oncogenes quando, ao sofrerem mutações gênicas ou alterações cromossômicas, apresentam ganho de função ou hiperexpressão de RNAs e/ou proteínas, em que um único alelo mutado é suficiente para alterar o fenótipo de uma célula normal para maligna, conhecido como efeito dominante.

Esses genes são responsáveis por aumentar as taxas de proliferação celular, ao mesmo tempo em que podem inibir a apoptose; eventos que podem dar início a uma neoplasia. Contudo, cada mutação oncogênica, por si só, raramente é capaz de induzir a um estado canceroso. Quando vários genes diferentes reguladores do crescimento em uma célula são alterados por mutações e a célula não pode compensar os efeitos individuais, o seu crescimento pode tornar-se desregulado e o estado canceroso celular pode então ocorrer.

Dentre as alterações cromossômicas pode-se destacar a translocação recíproca que ocorre entre os cromossomos 9 e 22 na Leucemia Mielóide Crônica, formando a alteração conhecida como Cromossomo Philadelphia. O gene quimérico *BCR/C-ABL* gera uma proteína de fusão com aumento de atividade (Figura 57).

### **GENES SUPRESSORES TUMORAIS**

Os supressores tumorais são genes que expressam produtos que controlam o ciclo celular (G1, S, G2, M) e quando mutados deixam de exercer seus papéis através de processos específicos para cada gene.

O ciclo celular apresenta períodos de síntese de proteínas específicas e DNA, e de divisão do material genético. Os ciclos são controlados por sinais externos ou extrínsecos à células, como fatores de crescimento e hormônios, e por sinais internos à células, principalmente por duas classes de proteínas: as ciclinas (nome proveniente da síntese e degradação ao final de um ciclo) e as proteínas quinases (com atividade de fosforilação de outras proteínas) dependentes de ciclinas chamadas de CDKs (ou Cdc). As transições entre as diferentes fases do ciclo são controladas por “pontos de controle” (*checkpoints*) que o param, por exemplo, em resposta ao DNA danificado ou até que a síntese do DNA seja completada. A formação do complexo ciclinas-CDKs é um evento crítico nesse processo.

Um dos pontos de checagem mais importantes é chamado de *start*, que está entre a metade de G1 e S. Nesse ponto, a formação do complexo ciclinas-CDKs e a fosforilação de

proteínas, como a pRB, é importante para a entrada na fase S. As células nas quais o *start* está alterado são propensas a se tornarem cancerosas, pois o DNA danificado deve ser reparado antes da replicação do DNA, com importante atuação da proteína pRB (Figura 58).

O gene RB1 está situado no cromossomo 13 e foi descoberto por meio de mutações na pRB em retinoblastoma; por isso seu gene recebeu esse nome, mas se expressa em vários tecidos. A proteína produzida pelo RB bloqueia o ciclo celular quando não está fosforilada. Nesta forma, a proteína pRB se liga ao fator de transcrição E2F. Após a fosforilação, a pRB se separa de E2F que atuará na ativação de genes-alvo de progressão do ciclo celular. Quando o RB1 está mutado, seu produto encontra-se alterado podendo ser encontrado permanentemente fosforilado, não se ligando a E2F o que resultará na progressão do ciclo celular para a fase S e, por fim, a mitose. Assim, na ausência do freio por RB e de outros freios do ciclo celular, as células se dividem incessantemente formando os tumores. As mutações no gene RB têm sido descritas em vários cânceres como os de próstata, bexiga e osteossarcomas. Estudos em camundongos demonstraram que as mutações homocigotas herdadas nesse gene levam à morte logo após o nascimento (Figura 59).

Outro importante gene supressor tumoral é o TP53 que está localizado no cromossomo 17. Quando o DNA está danificado, os níveis da proteína p53 são aumentados na célula, e o ciclo celular só passará de G1 para S se o dano for reparado. Assim, em condições normais, o nível dessa proteína é baixo, mas quando as células são tratadas com um agente que danifica o DNA, como a radiação, seus níveis são aumentados. Essa proteína atuará de duas formas como fator de transcrição: 1) pela ativação de genes envolvidos na parada no ciclo celular para que o DNA danificado seja reparado e; 2) pela ativação de genes que induzem a apoptose. No primeiro caso, a p53 liga ao promotor do gene da proteína p21. A proteína p21 é inibidora do complexo ciclina-CDK, o qual não irá fosforilar, por exemplo, a pRB. A pRB não fosforilada, não se desliga de E2F, e os genes de progressão do ciclo celular ativados por E2F não são transcritos até que o dano no DNA seja reparado. No segundo caso, a p53 ativa a transcrição de genes pro-apoptóticos como o BAX induzindo a célula à morte celular (Figura 60). A disfunção da p53 faz com que o ciclo celular prossiga com a mutação no DNA, permitindo sua transmissão às células descendentes e iniciando um processo neoplásico. Um exemplo de doença causada por alterações na p53 é a Síndrome de Li-Fraumeni, condição em que ocorre predisposição a desenvolver câncer em vários locais, como mama, ossos, sistema nervoso central, leucemias, entre outros. As mutações no gene TP53 são encontradas na maioria de todos os tumores. A perda de função da p53 é, portanto, uma etapa fundamental na carcinogênese.

Dentre os modelos que existem para explicar a carcinogênese envolvendo genes supressores tumorais, a Hipótese dos Dois Eventos, elaborada por Alfred Knudson, em 1971 (hipótese de Knudson) é a mais aceita. Nessa hipótese, as mutações devem provocar perda de função dos dois alelos. Essas mutações têm caráter recessivo, diferente das mutações em proto-oncogenes, uma vez que um único alelo mutado não é capaz de induzir uma neoplasia. Nos tumores de caráter hereditário, quando apenas uma mutação é herdada na linhagem germinativa, a outra mutação somática deverá ser adquirida ao longo da vida. Nos tumores esporádicos, as duas mutações são somáticas e adquiridas ao longo da vida. Isso explica como algumas doenças hereditárias não se manifestam em todos os indivíduos da família, uma vez que a segunda mutação ocorre ao acaso (Figura 61).

Além de mutações nos genes que codificam para a pRB e p53, alterações em outros genes podem estar associadas à carcinogênese. Entre esses, estão os genes BRCA1 e

BRCA2 presentes nos cromossomos 17 e 13, respectivamente. São ativados nas fases G1 e S do ciclo celular e por apresentarem uma função que parece estar associada ao reparo de danos no DNA, se os genes BRCA1 e BRCA2 estiveremse mutados, predispondo ao aparecimento de cânceres, como os de mama e de ovário. Mulheres que apresentam mutação nos genes BRCA1 e BRCA2 têm 85% de chance de desenvolver câncer de mama antes dos 70 anos de idade.

## UNIDADE VIII - BIOLOGIA MOLECULAR

A Biologia Molecular é a área de estudo da biologia em nível molecular, com especial foco no estudo da estrutura do material genético, suas funções e seus produtos de expressão (RNAs e proteínas), estudando ainda as interações entre DNA, RNA e síntese proteica. Dentre os métodos utilizados para esse fim, estão aqueles aplicados pela Genética Molecular, como a Tecnologia do DNA Recombinante, seguida da clonagem bacteriana, a Reação em Cadeia da Polimerase, produção de organismos transgênicos, dentre outras.

### TÉCNICAS DO DNA RECOMBINANTE

Corresponde ao mecanismo de isolar e caracterizar qualquer gene ou sequências de DNA de qualquer organismo. A engenharia genética, seguida de clonagem, são métodos utilizados na tecnologia do DNA recombinante, para a construção de Bibliotecas Gênicas e Genômicas. A engenharia genética realiza a manipulação direta de ácidos nucleicos (DNA, RNA) com o objetivo de transferir uma informação genética de um organismo a outro.

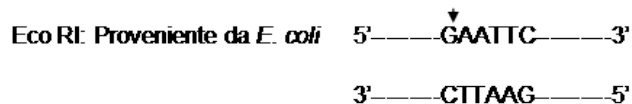
#### Clonagem

Um segmento de DNA é ligado a um pequeno cromossomo autoreplicante e inserido em uma célula hospedeira apropriada. As endonucleases de restrição, que são utilizadas nesse processo, correspondem a enzimas que fazem cortes internos nas moléculas de DNA de forma sítio específicas, sendo designadas pelo uso da primeira letra do gênero do micro-organismo que a produz e as duas primeiras letras da espécie. O número que a acompanha corresponde à ordem de identificação. Ex.: EcoR1 - E = *Escherichia* ; co = *coli*, 1 = primeira identificada.

Existem aproximadamente 400 enzimas de restrição diferentes, cuja função é de proteção do material genético da bactéria contra DNAs exógenos, como os DNAs virais. Atuam por reconhecer sequências palindrômicas no DNA, ou seja, em que a leitura é a mesma em ambos os sentidos na fita dupla. Os cortes gerados podem ser lisos e coesivos (Figura 62). As extremidades coesivas geram pontas unifilamentares complementares podendo ser reunidas com outros DNAs cortados com a mesma enzima. No caso das enzimas que geram extremidades lisas, é necessária a ligação de sequências denominadas de adaptadores (*linkers*). A ligação fosfodiéster é realizada pela enzima DNA ligase.

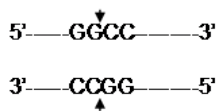


Coesivas:



Lisas:





Nesse processo são também utilizados os vetores de clonagem que são moléculas de DNA capazes de amplificar, em centenas de cópias, a informação genética que neles foi inserida. O mais comum na rotina laboratorial é o plasmídeo (Figura 63). A molécula de DNA de um plasmídeo é circular, extracromossômico e bifilamentar pequeno (cerca de 3 mil pares de bases) encontrado em procariotos, que apresenta replicação independente do genoma desses organismos (autorreplicação) devido à presença do sítio OriC, além de possuir sequências para marcadores fenotípicos como para resistência à antibióticos (Amp<sup>R</sup> - Resistência à Ampicilina) e para expressão da beta-galactosidase. O DNA plasmidial também contém sítios para clivagem com várias endonucleases de restrição (ER) presentes uma única vez no genoma plasmidial, denominado sítio de clonagem múltipla (SMC), de policlonagem ou polilinker. Assim, a seleção de *E.coli* contendo o plasmídeo com o fragmento de interesse (inserto) é feita pela análise de sobrevivência em meio contendo antibiótico e pela não expressão da beta-galactosidase, uma vez que o sítio de clonagem está dentro do sítio LacZ, que será interrompido pelo inserto (mutação). O primeiro tipo de plasmídeo descrito foi o pSC101 conferindo resistência à tetraciclina, sendo que plasmídeos mais versáteis foram posteriormente desenvolvidos.

## BIBLIOTECAS DE DNA

Os cromossomos inteiros (biblioteca genômica) ou em parte (biblioteca gênica) são clonados para isolar o DNA de uma célula, organismo, ou fragmentos de tecidos, por meio da digestão com enzimas de restrição de ambos, vetor e inserto com a mesma enzima. Após formação do DNA recombinante pela ligase, este é inserido nas células hospedeiras por meio da transformação e após vários ciclos de replicação são formadas as colônias de clones de bactérias contendo o inserto.

Um dos métodos de transformação do hospedeiro pela inserção do plasmídeo exógeno é a utilização de reagentes químicos. O tratamento da membrana com um sal, geralmente o cloreto de cálcio, faz com que a membrana de *E.coli* fique permeável ao DNA, no momento em que o plasmídeo entra na célula (Figura 64). Além da transformação química, é possível realizar a permeabilização da membrana, por meio de choques, provocados por um aparelho eletroporador.

Como dito anteriormente, após a transformação, a seleção poderá ser realizada utilizando um antibiótico e os marcadores fenotípicos para a cor. As bactérias transformadas com plasmídeo serão resistentes ao antibiótico utilizado, enquanto as demais, que não estão transformadas com o plasmídeo, morrerão. A seleção fenotípica pela cor é realizada pela adição de um substrato para a beta-galactosidase no meio de crescimento das bactérias. Num processo normal, sem inserto, o IPTG (isopropil-B-tio-galactosídeo) após ser clivado pela beta-galactosidase leva à formação de uma cor azul nas colônias, isso porque o sítio para a produção da enzima beta-galactosidase (gene lacZ), contendo o sítio SMC, é capaz de produzir a enzima. Contudo, após a ligação do inserto no SMC, o gene lacZ é interrompido e o seu produto protéico não será produzido, resultando em colônias brancas (Figura 65).

A biblioteca gênica consiste na clonagem de um gene ou segmento gênico específico por meio da extração de RNA, seguida da sua conversão em DNA complementar ao RNA (cDNA), pela enzima Transcriptase Reversa (RT extraída de vírus). O processo

envolve a utilização de um primer, por exemplo, poliT (liga à cauda poliA do RNAm) para síntese do cDNA pela RT. O cDNA é convertido em DNA dupla fita (dsDNA) pelo uso da enzima DNA polimerase e PCR descrita a seguir, e então inserido num vetor de clonagem após digestão do produto com a mesma enzima de restrição utilizada na clivagem do vetor e ligado pela DNA ligase, ou pela utilização de kits. O processo seguinte é o mesmo descrito anteriormente (Figura 66).

### **REAÇÃO EM CADEIA DA POLIMERASE (PCR)**

A PCR foi desenvolvida por Saiki (1985) e Mullis (1987) para amplificar segmentos de DNA. A PCR é considerada uma técnica revolucionária, por permitir estudar em detalhes a estrutura e expressão de genes por um processo de replicação, *in vitro*. Contudo, diferente das técnicas de Southern e Northern-Blot, que serão aqui descritas posteriormente, a sequência do gene que se pretende estudar deve ser conhecida. A PCR permite a análise de sequências de DNA ou RNA de fontes escassas de ácidos nucléicos e degradada, como sêmen e bulbo capilar, tendo, portanto, grande aplicação na área forense (identificação de pessoas) e criminalística, por uma análise de *fingerprinting* nos casos que envolvem identidade incerta ou busca de prováveis suspeitos. Atualmente, faz parte de muitos laboratórios de análises clínicas, não somente para investigação de paternidade, mas no diagnóstico de doenças genéticas hereditárias, doenças parasitárias, infecciosas, imunológicas.

Na PCR são utilizados primers que delimitam o alvo para iniciar a reação, a enzima DNA polimerase termoestável, cuja mais conhecida é a Taq DNA polimerase, que foi isolada da bactéria de fontes termais *Thermus aquaticus*, que realiza a síntese da nova fita a partir dos primers, os nucleotídeos (dNTPs),  $MgCl_2$  como cofator da Taq, um tampão para garantir a eficiência da reação e o molde de DNA extraído das células ou tecidos que se pretende estudar. A reação é realizada num aparelho denominado de termociclador, pois é pela variação de temperaturas, e não por um complexo com várias enzimas, como ocorre na replicação *in vivo*, que a PCR é feita.

A uma temperatura de cerca de 95°C ocorre a desnaturação do DNA, entre 50-65°C ocorre a ligação dos primers (tamanho variável) e a cerca de 72°C a DNA polimerase irá fazer a adição de nucleotídeos a partir da extremidade 3-OH' livre do primer. O DNA alvo poderá ser um gene ou parte dele, ou de qualquer sequência de bases genômicas do DNA que se pretenda analisar. A temperatura de anelamento dos primers será em função do tamanho e da composição de bases que possuem. A variação de ciclos para as etapas de desnaturação, anelamento/ligação do primers e extensão de nucleotídeos pela DNA polimerase é repetida por cerca de 35 ciclos para a amplificação (replicação) do DNA alvo (Figura 67).

### **TÉCNICAS DE HIBRIDAÇÃO**

#### **Transferência de Southern (Southern-blot)**

Essa técnica utiliza dois princípios dos ácidos nucléicos, que são a possibilidade de cortes no DNA por enzimas de restrição e a mobilidade do DNA por eletroforese em gel de agarose ou poliacrilamida, dada pelas cargas negativas na molécula, que submetido a um campo elétrico e em solução tampão, migrará para o polo positivo.

O DNA genômico extraído de uma célula, que deverá ser obtido em grandes quantidades e estar íntegro, após clivagem com enzimas de restrição, é submetido à eletroforese para separação das moléculas geradas pelo peso molecular dado em pares de bases (pb). Em seguida, o DNA é desnaturado em solução alcalina para ficar de fita simples

e transferido para uma membrana de náilon. Uma sonda, constituída por uma pequena sequência de DNA complementar à sequência alvo, é marcada radioativamente ou alternativamente por marcações ditas frias por colorimetria ou por fluorescência. Após marcação, a sonda formará um híbrido com o DNA alvo fragmentado, processo chamado de hibridação, e que está fixado na membrana náilon; ou seja, a sonda que encontra o alvo irá parear com o mesmo que está imobilizada na membrana. No caso de sondas quentes (radioativas), ao ser adicionado um substrato e exposição a um filme de raio-X, ocorrerá a marcação de pontos escuros (bandas) no filme (Figura 74).

Esse método foi descrito em 1975, sendo útil em vários experimentos, como na comparação inter e intra espécie, no *fingerprinting* em exames de vínculo genético, na comparação entre indivíduos de uma população, nas análises forenses, dentre outras. A base genética está no ganho ou perda de sítios de restrição nos diferentes alelos ou sequências de DNA, codificantes ou não. Devido às manchas geradas na radiografia, a técnica contém a palavra "blot" no nome.

### **Transferência por Northern-blot**

A técnica é semelhante à descrita anteriormente e o nome *Northern* (norte) veio para contrapor a transferência de *Southern* (sul), sendo utilizado nas análises de expressão gênica após extração de RNA e fixação em membrana de náilon. A diferença no método está na não utilização de enzimas de restrição. Devido à alta instabilidade do RNA e sensibilidade ao ataque de RNases, os procedimentos devem ser realizados tomando cuidado com a contaminação por essas enzimas, utilizando reagentes inibidores de RNases, luvas limpas, e reagentes desnaturantes para separar as estruturas secundárias dos RNAs. Pode ser aplicada no estudo da expressão de um gene em um tecido em comparação a outros, determinar a expressão constitutiva ou induzida do gene, padrão de expressão temporal, por exemplo, associado ao crescimento dos indivíduos ou desenvolvimento do embrião, dentre outras.

### **Transferência pela técnica de Western**

Consiste na separação e caracterização de proteínas em géis de poliacrilamida na presença de reagentes desnaturantes para desnaturação de proteínas, necessária para separação de subunidades associadas à sua função. As proteínas são transferidas para uma membrana de nitrocelulose e um anticorpo será utilizado para localização da proteína específica em um dado tecido. Um anticorpo secundário contra o primário é conjugado a um isótopo radioativo ou colorimétrico e quando o substrato é adicionado o produto resultante é visível, sendo útil nas análises de expressão de produtos gênicos codificantes específicos.

## **PROJETOS GENOMAS**

Um dos grandes avanços alcançados nos últimos anos advém do uso da Tecnologia do DNA recombinante, Clonagem, PCR e Sequenciamento do DNA para a obtenção de genomas de organismos como *H. influenzae* (1ª. bactéria com genoma completo, em 1995), *E. coli*, *S. cerevisiae*, *D. melanogaster*, *C. elegans* e outros; e dentre estes, o humano. O objetivo do Projeto Genoma Humano (PGH) foi o de construir um mapa físico detalhado de todo o genoma humano. A HUGO (*human genome organization*) coordenou os esforços dos geneticistas em todo o mundo para esse fim. O pesquisador James Watson foi o primeiro diretor do projeto. Em junho de 2000, foi anunciado pela HUGO e empresa CELERA (privada), que 96% do genoma havia sido sequenciado e mapeado, e cerca de 25 mil genes foram identificados a partir de 3,1 bilhões de bases.

O potencial uso do conhecimento das sequências levanta questões de cunho ético e social, dependendo de como serão usadas. Por exemplo, existe a possibilidade de identificação de indivíduos com doenças neurodegenerativas dominantes associadas com a expansão de nucleotídeos, normalmente CAG (X-frágil, Machado-Joseph, Doença de Huntington). Associada à fertilização *in vitro*, um casal portador de alelo poderia aceitar ou recusar a implantação de um embrião, após análise de seu DNA e presença do alelo para essas doenças. Será que seria importante que os indivíduos que já sabem que tem pais heterozigotos, poderem escolher se desejam saber se são dominantes e que por volta dos 30 anos poderão apresentar os sintomas? Como preparar psicologicamente as pessoas frente a essas descobertas?

Por outro lado, quando há a descoberta de doenças para as quais já existam medicamentos ou outros métodos de tratamento, ou que indiquem necessidade de mudança de comportamento e alimentação, seria importante o diagnóstico e para que esse seja realizado, é importante o conhecimento das mutações ou polimorfismos que causam o distúrbio. Além disso, o conhecimento da estrutura gênica é importante para o desenvolvimento de diagnóstico e terapias gênicas.

O sequenciamento de curtas sequências RNAs (EST, termo que vem do inglês *expressed sequences tags*), que foram expressos em uma determinada condição (normal e/ou alterada), é feito após clonagem e montagem de uma biblioteca gênica, ou seja, os cDNAs são utilizados para gerar as ESTs. Como os RNAs mensageiros possuem ORF, os projetos de sequenciamento de ESTs são denominados de ORESTES. Essas sequências de ESTs podem ser completas ou parciais, e, no caso dessas últimas, podem ser alinhadas sobre *contigs* genômicas.

### **Sequenciamento de DNA pelo Método de Sanger**

Na década de 1960, as pesquisas do bioquímico Fredrick Sanger permitiram iniciar o sequenciamento do RNA, o que tornou teoricamente possível entender toda a enorme quantidade de informações contidas no DNA, e não apenas exemplos isolados. Isso levou ao real interesse em conhecer a relação entre cada gene e cada característica física, inclusive as doenças de origem genética. O método original de Sanger para o sequenciamento de ácidos nucleicos é um método enzimático, utilizando reações simples e individualizadas de PCR seguida de eletroforese. Por este motivo, e porque nenhuma polimerase inicia o seu trabalho se não tiver disponível uma extremidade 3'OH livre, é necessário para a utilização deste método ter algum conhecimento da sequência da extremidade 5' da região a sequenciar. Com este conhecimento, é possível desenhar um primer complementar da cadeia a sequenciar, que por hibridização com esta vai fornecer a necessária extremidade 3'-OH livre (Figura 68).

A reação prossegue, incorporando nucleotídeos até que um nucleotídeo terminador (didesoxi) seja incorporado. Os produtos da reação de sequenciamento de Sanger eram separados em géis de poliacrilamida segundo o seu peso molecular. Se em cada reação de sequenciamento for apenas incorporado uma espécie de terminador, cada reação pode servir como revelador da posição em que aparece o respectivo nucleotídeo. A eletroforese em paralelo das quatro reações irmãs de sequenciamento permite então obter um gel clássico, que se lê seguindo os produtos com peso molecular sucessivamente crescente (Figura 69).

Com o intuito de facilitar a leitura dos géis de sequenciamento, tornando ao mesmo tempo mais precisas as diferenças de migração em gel e rápidas, as quatro reações de sequenciamento foram incorporadas numa só, utilizando, para isso, terminadores

modificados. A modificação consiste na marcação de cada espécie de terminador com um fluorocromo (sinal emitido como cores diferentes e captada pelo software), o que permite que cada produto de PCR específico emita luz num comprimento de onda bem determinado e a leitura dos géis é feita por equipamentos específicos capazes de analisar os comprimentos de onda emitidos.

No caso dos sequenciamentos em larga escala, como os que são feitos para obtenção dos genomas, foram sendo desenvolvidos sofisticados programas de computador, que foram desenvolvidos para organizar as sequências de DNA obtidas em sequências contínuas maiores, denominadas contigs. Assim, as leituras contendo sequências idênticas, são consideradas como sobreposições e são reunidas, formando contigs maiores e quanto maior o número de sequências com partes sobrepostas, maior será a sequência obtida e menor a quantidade de lacunas nessa sequência que está sendo organizada (Figura 71). Nos projetos genomas foram feitos alinhamento dos clones, ordenamento e posicionamento nos cromossomos, seguida da caracterização de sequências gênicas.

### **Bioinformática**

Atualmente o termo “Bioinformática” tem sido normalmente utilizado para se referir à análise computacional de dados genômicos, o que inclui também análises de interações DNA-DNA, DNA-RNA, RNA-RNA, proteína-proteína, proteína-ácido nucleico, sendo referida ainda como Biologia Computacional. O sentido desse termo pode ser bem mais amplo, incluindo tudo relacionado à biologia que possa ser auxiliado pela computação.

Porém, é natural relacionar análise de dados genômicos à computação, devido à enorme quantidade de informação, dificultado pelo tratamento manual. As bases atuais de dados genômicos como <http://www.ncbi.nlm.nih.gov> e <http://www.expasy.org>, dentre outros, ajudam pesquisadores a entender as funções biológicas, as estruturas químicas e a história evolutiva de organismos. Quando uma nova sequência de bases de nucleotídeos, ou aminoácidos é decifrada, esta pode ser comparada às sequências cadastradas na base, a fim de encontrar sequências relacionadas a ela, o que pode ajudar a revelar algumas propriedades da sequência em questão. Esta área de pesquisa tem se tornado cada vez mais importante devido ao rápido crescimento das bases de dados e sua análise manual pode ser impraticável.

O *blast* é uma das ferramentas presentes no site do NCBI (*National Center for Biotechnology Information*), a qual permite, a partir de uma curta sequência de nucleotídeo ou proteína (*query sequence*), realizar a busca de um ou vários genomas, fornecendo a probabilidade de que aquela seja a sequência procurada (*hits*) de todas as sequências de ácidos nucleicos e proteínas previstas relacionadas a sequência pesquisada.

### **Variabilidade Genômica, sequências repetitivas e análise de fingerprinting**

O DNA em tandem (sequencial) e hipervariável, também conhecidos como DNA satélite, envolve sequências repetidas, uma seguida de outra e agrupadas. As famílias de DNA satélite variam quanto à localização no genoma e comprimento das unidades repetidas que constituem a sequência, formando algumas classes, tais como:

- a) Minissatélites ou VNTRs (*variable number tandem repeats*) correspondem às repetições em blocos (tandem) geralmente maiores que sete (CAGT/CAG...) e representam polimorfismos abundantes nas populações. As análises de *fingerprinting* por *Southern-blot* analisam os blocos de VNTRs presentes nos indivíduos.
- b) Microssatélites ou STR (*short tandem repeats*) são pequenas sequências repetidas denominadas de mononucleotídeos se a mesma base se repete (repetições de

poliadeninas), apresentando dinucleotídeos com repetições de duas bases (citosina-adenina/C-A) e até o agrupamento curto de mais bases que estão espalhados pelo genoma, na sua maioria em regiões não codificadoras. As análises de *fingerprinting* por PCR analisam os blocos de repetições STRs entre os indivíduos, sendo atualmente, amplamente empregada nos exames de DNA, como os de paternidade.

Além do DNA em tandem, há o DNA repetitivo disperso. Nessa classe, estão os elementos de transposição chamados de transposons. Quando há um intermediário de RNA, esses elementos são chamados de retrotransposons. Nesses elementos as repetições podem ser encontradas como cópias únicas distribuídas pelo genoma, ou agrupadas, conferindo-lhes a capacidade de “se mover” de um local para outro no genoma, normalmente deixando uma cópia original no lugar, o que faz com que possam se acumular por todo o genoma, causando mutações e doenças genéticas em humanos (Figura 72). No entanto, o movimento dos transposons é relativamente raro nas células humanas, mas esses elementos foram tão bem-sucedidos na sua propagação que compreendem cerca de 50% do genoma humano. Os transposons parecem ter um papel mais estrutural, por exemplo, DNA intergênico, do que associado à codificação de proteínas (menor que 2% no homem).

Em plantas parece que esses elementos ocupam uma proporção maior no genoma enquanto em organismos mais simples, como bactérias e fungos, seu número é muito menor. A grande quantidade dessas sequências repetitivas nos organismos mais complexos indica sua importância na manutenção das funções celulares vitais, não podendo ser considerado como DNA lixo.

Os transposons podem se inserir dentro de genes, prejudicando completamente a função dos mesmos, assim como podem se inserir dentro de sequências gênicas reguladoras, onde sua presença pode provocar alterações na expressão desse gene. Foi a partir das perdas de função e alterações na expressão gênica que os transposons foram descobertos. Na década de 40, a pesquisadora Barbara McClintock evidenciou a presença dos transposons em milho por meio de variação na coloração das sementes na espiga. Existem muitos tipos diferentes de elementos de transposição e de uma forma mais didática, esses podem ser divididos em três classes por compartilharem características comuns na estrutura e em sua organização geral, bem como nos mecanismos de transposição, que são os transposons de DNA, retrotransposons semelhantes a vírus e retrotransposons com poli A.

Os transposons de DNA permanecem como DNA do começo ao fim de um ciclo de recombinação, se deslocando utilizando mecanismos que envolvem a clivagem, devido à presença de um gene par a transposase, seguido da religação das fitas de DNA pela presença dos sítios de recombinação (sequências repetidas específicas e invertidas de DNA) que delimitam o gene transposase, além de codificar outras proteínas que participam da recombinação. Os sítios de recombinação presentes nas extremidades terminais invertidas variam em extensão de 25 a algumas centenas de pares de bases, não sendo repetições exatas de sequência e possuem as sequências de reconhecimento da recombinação pelas transposases (algumas vezes chamadas de integrases). Diversos transposons de DNA bacterianos contêm genes que codificam proteínas que conferem resistência a um ou mais antibióticos. A presença do transposon, portanto, torna a célula hospedeira resistente a esse antibiótico.

Os transposons de DNA que contêm o gene da transposase são os chamados de transposons autônomos e aqueles mais simples, que possuem as sequências invertidas, mas não contêm o gene da transposase na sua sequência, são denominados transposons não autônomos. Esses elementos possuem apenas as repetições terminais invertidas, que

são as sequências que atuam em *cis*, necessárias à transposição. Em uma célula que tenha também um transposon autônomo, que codifique uma transposase, esta irá reconhecer também as repetições terminais invertidas do elemento não autônomo, permitindo sua transposição. Entretanto, na ausência desse transposon “auxiliar” (para fornecer a transposase), os elementos não autônomos permanecem estáticos, incapazes de deslocamento.

Os retrotransposons semelhantes aos vírus, também chamados de retrotransposons LTR, constituem elementos móveis que se deslocam para novos locais do DNA utilizando um intermediário temporário de RNA. Os retrotransposons também possuem sequências terminais repetidas e invertidas que são os sítios de ligação e ação da transposase. As repetições terminais invertidas estão inseridas em sequências repetidas mais longas, organizadas nas duas extremidades do elemento, como repetições diretas e são denominadas longas repetições terminais ou LTRs (*long terminal repeats*). Esses retrotransposons codificam proteínas necessárias para a sua mobilidade como a integrase/transposase e a transcriptase reversa. A transcriptase reversa (RT) é necessária à transposição, pois esta ocorre através de um intermediário de RNA. Como esses elementos convertem RNA em DNA, o reverso do trajeto normal do fluxo da informação biológica (DNA para RNA), eles são conhecidos como “retro” elementos. A distinção entre retrotransposons semelhantes aos vírus e retrovírus é que o genoma de um retrovírus é empacotado em uma partícula viral que sai da sua célula hospedeira e é capaz de infectar uma nova célula. Ao contrário, os retrotransposons podem se deslocar apenas para novos sítios no DNA dentro de uma mesma célula.

Os retrotransposons com poli(A), também descritos como retrotransposons não-LTR, não apresentam as repetições terminais invertidas características presentes nas outras duas classes de transposons, mas as duas extremidades do elemento móvel apresentam sequências distintas, sendo que uma extremidade é chamada de 5'-UTR enquanto a outra extremidade apresenta uma região chamada 3'-UTR, seguida por uma região poli(A). Também apresentam as ORFs que codificam a transposase e a RT. Normalmente, os retrotransposons com poli(A) se apresentam em ambas as formas, autônoma ou LINEs e não autônoma ou SINES.

Os LINEs foram identificados primeiramente como uma família de sequências repetidas e o termo vem do inglês como longo elemento nuclear intercalado (*Long Interspersed Nuclear Element*), são abundantes nos genomas de vertebrados, correspondendo a cerca de 20% do genoma humano, com tamanho médio de 6.000 pb. O elemento L1 é um dos LINEs mais bem conhecidos do genoma humano. Além de promover sua própria mobilidade, os LINEs também fornecem as proteínas necessárias para a transcrição reversa e para a integração de outras classes de sequências repetidas relacionadas, os retrotransposons com poli(A) não autônomos, conhecidos como curtos elementos nucleares intercalados, os SINES (*Short Interspersed Nuclear Elements*), representado pela sequência Alu, bem difundida no genoma humano e que pode ser clivada pela endonuclease de restrição denominada Alu I.

Como curiosidade, as similaridades dos retrotransposons com retrovírus vão além de sua capacidade de inserir no genoma, como visto nos elementos de transposição de leveduras, os Ty (*Transposons in Yeast*), em que o RNA de Ty, é encontrado nas células empacotado em uma espécie de partícula viral, no entanto não podem sair e infectar novas células como fazem os vírus. De forma diferente da maioria dos transposons, os Ty se integram, preferencialmente, em regiões cromossômicas específicas, como, por exemplo, os elementos Ty1 de *S.cerevisiae* quase sempre se transpõem no DNA próximos à região

promotora de genes transcritos pela RNA polimerase III, que transcreve especificamente os genes de RNAt.

No genoma, existem cópias adicionais de sequências altamente relacionadas a diversos genes celulares, além das cópias repetitivas, as quais parecem ter perdido seus promotores e introns e, frequentemente, contêm interrupções próximas das extremidades 5'. Essas sequências são conhecidas como pseudogenes processados e, normalmente, não são expressos na célula. Esses pseudogenes são, muitas vezes, flanqueados por pequenas repetições no DNA-alvo.

Por fim, existe ainda uma parte do genoma que também não está associada à expressão de proteínas ou a RNAs estruturais, chamadas de DNA intergênico, correspondendo a mais de 60% do genoma humano e não apresenta função conhecida. Assim como nas classes descritas anteriormente, existem os DNAs intergênicos únicos e os repetitivos. Cerca de 25% do DNA intergênico é único, compreendendo regiões aparentemente não funcionais, incluindo genes mutantes não funcionais, fragmentos de genes e pseudogenes. Os genes mutantes e os fragmentos de genes surgem da mutagênese aleatória simples ou erros de recombinação do DNA.

## **AVANÇOS EM BIOTECNOLOGIA**

Dentre as diferentes áreas de trabalho da Biotecnologia moderna estão a engenharia genética, cultura de células e tecidos, engenharia de proteínas e de produtos alimentares que são responsáveis por progressos em inúmeras áreas científicas, assim como por alguns dos produtos que consumimos no dia a dia. Os transgênicos, ou organismos geneticamente modificados (OGMs), são organismos que contêm moléculas de DNA recombinantes inseridas em seu genoma, mas proveniente de outro organismo. As primeiras moléculas de DNA recombinantes ocorreram nos microorganismos; no entanto existe uma série de animais e plantas transgênicos.

As técnicas de produção de transgênicos podem ser utilizadas para produção de efeito aditivo para um aumento de produtividade em animais, para produção de novas características metabólicas, inclusive resistência a patógenos, na produção de organismos resistentes a vírus, na indústria farmacêutica e medicina, como um exemplo, seria a produção de leite com outras proteínas, ovelhas que secretam fatores de coagulação no leite. A primeira patente animal foi conferida em 1988 à Universidade de Harvard (EUA) para um rato transgênico com um oncogene, que confere suscetibilidade ao câncer de mama, facilitando testes de carcinogênese, auxiliando no diagnóstico e nos testes de medicamentos contra a doença.

Dentre outros exemplos de inovação tecnológica potenciais, pelo uso de animais transgênicos feitos no Brasil e no mundo, pode-se citar o camundongo expressando o hormônio de crescimento de um rato, apresentando crescimento maior que o normal; vaca também capaz de produzir hormônio de crescimento humano (somatotropina) no seu leite, além de vaca com potencial para produzir insulina humana no leite; coelhos expressando fator IX de coagulação e hormônio folículo estimulante (FSH) para bovino; caprinos expressando a proteína do fator estimulante de colônia de granulócitos humanos (sigla em inglês hG-CSF), com aplicação para portadores de Aids, pacientes com câncer com tratamento quimioterápico. As análises dos efeitos colaterais dos transgênicos são avaliados, e, entre esses, estão fraqueza muscular, degeneração tecidual e esterilidade. Como curiosidade, no início dos estudos com transgênicos, pesquisadores da Califórnia produziram um *Sapobacter* ou *Bactosapo*, em virtude da introdução do gene para o RNAr de *Xenopus laevis* no genoma de plasmídeos de bactérias *E.coli* que passaram a expressar



esse gene exógeno.

Outro exemplo de transgênico divulgado na mídia foi a clonagem revolucionada, quando Ian Wilmut e seus colegas do Instituto Roslin, em 1997, na cidade de Edinburgo, Escócia, clonaram com sucesso uma ovelha chamada Dolly, que foi o primeiro mamífero clonado. Dolly provou ser geneticamente idêntica às ovelhas Finn Dorsett e não à ovelha Blackface, o que demonstrou claramente que era um clone bem-sucedido (Figura 75).

Diferente dos animais, as plantas transgênicas são mais fáceis de produzirem clones devido à totipotência, uma importante característica desse grupo, pois permite a regeneração de plantas inteiras a partir de células somáticas modificadas. O uso da *Agrobacterium tumefaciens* é um dos métodos que permite a produção de transgênicos em plantas. Esta bactéria do solo causa galha de coroa (tumores) em dicotiledôneas, sendo que entre a raiz e o caule são susceptíveis a ferimentos. Quem contém a informação genética para a capacidade de induzir tumor por essa bactéria é o plasmídeo Ti, capaz de se integrar ao DNA das células vegetais. O gene exógeno de interesse é inserido na região do T-DNA e o gene causador do tumor é desligado para não causar o tumor (Figura 76).

## REFERÊNCIAS

- ALBERTS, B. **Biologia molecular da célula**. 4. ed., Porto Alegre: Artes Médicas, 2004.
- BORGES-OSORIO, M. R.; ROBINSON, W. M. **Genética humana**. 2. ed., Porto Alegre: Editora Artmed S.A., 2001.
- GRIFFITHS, A. J. F.; MILLER, J. H.; SUZUKI, D. T.; LEWONTIN, R. **Introdução à genética**. 8. ed., Rio de Janeiro: Guanabara Koogan, 2006.
- MALAJOVICH, M. A. **Biotecnologia 2011**. Rio de Janeiro: Edições da Biblioteca Max Feffer do Instituto de Tecnologia ORT, 2012. Disponível em: <[http://www.bteduc.bio.br/livros/Biotecnologia\\_2012.pdf](http://www.bteduc.bio.br/livros/Biotecnologia_2012.pdf)>. Acesso em:
- NELSON, D.L.; COX, M.M. **Lehninger: princípios de bioquímica**. 5. ed. Porto Alegre: Artmed, 2011.
- NUSSBAUM, L. R.; McINNES, R. R.; WILLARD, H. F. **Thompson & Thompson: genética médica**. 7. ed., Rio de Janeiro: Editora Elsevier Ltda, 2008.
- OLIVEIRA, M. Biotecnologia-alternativa animal: cabras, vacas, galinhas e camundongos transgênicos são um novo meio para produção de medicamentos. **Pesquisa Fapesp**, Sao Paulo<sup>[AL1]</sup>: Agência Fapesp, 147: 84-87, 2008.
- WATSON, J.D. et al. **Biologia molecular do gene**. 5. ed., Porto Alegre: Artmed, 2006.