

A história da descoberta dos Introns

I. A descoberta dos genes interrompidos e dos introns de pré-mRNAs

The history of the discovery of Introns
I. The discovery of split genes and the introns of pre-mRNAs

Mario Mayer¹

1. Mario Gustavo Mayer, mestre em Morfologia na área de genética pela UNIFESP. Pesquisador Científico do Laboratório Especial de História da Ciência, do Instituto Butantan. Contato: mario.mayer@butantan.gov.br.

Resumo

A partir da determinação da estrutura da molécula informacional ácido desoxirribonucleico (DNA) feita no ano de 1953 por Watson e Crick, a biologia molecular teve um desenvolvimento fantástico que culminou, nos tempos atuais, com a possibilidade de determinação da sequência de nucleotídeos que correspondem à totalidade do conteúdo informacional de diversos organismos. A descoberta da configuração de genes interrompidos, com a discriminação de sequências denominadas exons e introns, foi um passo fundamental para o entendimento da arquitetura, funcionamento e regulação da expressão dos genes. Este é um artigo inicial de uma série que tem como objetivo fazer um levantamento dos experimentos fundamentais da descoberta dos vários tipos de introns. Descrevemos os antecedentes e a descoberta dos genes interrompidos de pré-mRNAs, os quais dão início à descrição desse tipo de organização gênica. O resgate da história da descoberta dos genes interrompidos e a descrição dos processos e métodos implicados têm sua importância na origem do vasto arsenal de técnicas e análises utilizadas nos estudos

dos mecanismos de regulação gênica hoje em dia, a qual tem profundo impacto em várias atividades humanas.

Palavras-chave

Genes interrompidos, “splicing”, RNA, intron, exon.

Abstract

Since the determination of the informational molecule deoxyribonucleic acid (DNA) structure by Watson and Crick in 1953, molecular biology had a fantastic development that culminated in recent times with the possibility of determining the nucleotide sequence of whole information content of various organisms. The discovery of interrupted (split) genes configuration and the discrimination of sequences called exons and introns were a key step to understanding the architecture, functioning and regulation of gene expression. This is the first article of a series that aims to survey the fundamental experiments of the discovery of various introns types. In this initial article, we describe the antecedents and the discovery of interrupted genes of pre-mRNAs that accounted for the initial mark of the description of this type of gene organization. The rescue of the history of interrupted genes discovery and description of the processes and methods involved has its importance in the origin of a wide source of techniques and analysis used in the studies of gene regulation mechanisms nowadays which has profound impact in various human activities.

Keywords

Split genes, “splicing”, RNA, intron, exon.

O cenário anterior à descoberta dos genes interrompidos

Na segunda metade da década de 1970, a partir de estudos em organismos procarióticos, a estrutura física do gene parecia consolidada. A colinearidade entre o gene e seus produtos, RNA mensageiro e proteína, já era estabelecida (Lewin, 1982). Como a genética clássica sugeria um comportamento

2.

O valor C é a quantidade de DNA em massa de uma célula haploide, como, por exemplo, um gameta. A origem da utilização da letra C para o conteúdo de DNA de uma célula haploide é referência da quantidade constante de DNA para um determinado genótipo.

semelhante entre genes de organismos procarióticos e eucarióticos, generalizava-se que a estrutura e os mecanismos de regulação de um gene bacteriano fossem universais.

No entanto, estudos relacionados ao material genético de células eucarióticas indicavam que a regulação da expressão gênica nessas células era diferente daquela determinada para procariotos. Padrões estruturais que diferenciam essas células já sugeriam diferentes padrões regulatórios de expressão gênica. A diferença morfológica fundamental entre células procarióticas e eucarióticas é a presença de membrana nuclear nessas últimas. Devido à ausência desse compartimento em procariotos, a síntese proteica, tradução do RNA mensageiro (mRNA) nos ribossomos, é iniciada antes mesmo que a transcrição do mRNA tenha terminado, ou seja, a tradução e transcrição são concomitantes, espacial e temporalmente acopladas. Já nos eucariotos, a presença da membrana nuclear separa os processos de transcrição e tradução, que acontecem respectivamente no núcleo e no citoplasma. Poderiam ser supostas, conseqüentemente, etapas adicionais de processamento da mensagem em eucariotos, nas quais o núcleo teria um papel de concentrador dos componentes necessários a esse processamento hipotético adicional (Crick, 1979).

Um fato que intrigava os cientistas até esse momento era a discrepância entre as quantidades de DNA de células eucarióticas e procarióticas. Além disso, chamava atenção a variação significativa na quantidade de DNA de células haploides (C)² entre os vários organismos eucarióticos, sem que houvesse uma correlação com a complexidade do organismo ou com a variação no número de seus genes (paradoxo do valor de C). Uma explicação para essas discrepâncias seria supor que a maior parte do genoma de eucariotos não seria transcrita, o que sem dúvida levaria à investigação do seu papel funcional. Uma suposição alternativa, e igualmente satisfatória, seria a de que a unidade de transcrição seria bem maior do que a sequência representada no mRNA (Lewin, 1980). Estudos bioquímicos do metabolismo

3. Células HeLa são células de uma linhagem celular estável, ou seja que podem ser mantidas indefinidamente em cultura, oriundas de um carcinoma de colo de útero da paciente Henrietta Lacks.

4. Cauda de poli A é uma adição de 100-200 resíduos de adenosina (A) na extremidade 3' de um mRNA após sua transcrição.

de RNA em células eucarióticas contribuíram muito para o encaminhamento da resolução desse problema (Darnell, 1973). Quando células HeLa³ eram incubadas em meio de cultura contendo uridina tritiada por 5 minutos ou menos, a maior parte do RNA marcado localizava-se numa fração nucleoplasmática constituída de moléculas de tamanho heterogêneo denominada RNA heterogêneo nuclear (hnRNA). O tamanho das moléculas componentes dessa fração variava de centenas a dezenas de milhares de nucleotídeos. Considerando-se a substituição de uridina por timidina, para efeito comparativo, a composição do hnRNA mostrou-se semelhante à do DNA celular total. Sua composição também era semelhante à do RNA de polissomos – no entanto, bastante diferente da composição de RNA ribossômico (rRNA). Estudos da estabilidade do hnRNA mostraram que a meia-vida dessa classe de RNA é bem menor do que a de mRNA. Assim, o hnRNA, por se tratar de uma fração composta por moléculas de tamanho longo e variável, de meia-vida curta e composição semelhante ao mRNA, foi indicado como o provável precursor dos mRNAs (Darnell, 1973).

A identificação da cauda de poli A⁴ na extremidade 3' dos mRNAs como uma modificação pós-transcrição levou imediatamente à verificação da existência da mesma modificação na extremidade 3' dos hnRNAs (Darnell et al., 1971; Lee et al., 1971; Edmonds et al., 1971). Da mesma forma, a verificação de outra modificação pós-transcrição, essa na extremidade 5' dos mRNAs, a estrutura 5' "cap" (7mGpppX), também levou à constatação de que a mesma estrutura estava presente em ambas as classes de RNA (Wei; Moss, 1974; Rottman et al., 1974 8; Furuichi et al., 1975). Surge então a seguinte questão: se a população de hnRNA fosse precursora de mRNA, como as extremidades 5' e 3' do hnRNA seriam preservadas durante o processamento na geração de moléculas menores de mRNA?

O modelo de adenovírus. Solução simplificada e essencial para a descoberta dos genes interrompidos

Experimentos com o objetivo de verificar a hipótese de que os hnRNAs como precursores de mRNAs eram difíceis de serem conduzidos à época devido à complexidade do genoma de eucariotos e das ferramentas disponíveis. No entanto, a escolha de um modelo simplificado, mais especificamente o adenovírus, foi adequada para o início da resolução do problema. O adenovírus do tipo 2 (Ad-2) é um vírus de DNA de dupla fita com aproximadamente 36 kbp, estando as regiões codificadoras de suas proteínas estruturais localizadas nos últimos dois terços do DNA viral. A transcrição de Ad2 é dirigida por RNA polimerase II e, as características dos produtos de transcrição são semelhantes às dos produtos celulares, ou seja, transcritos longos e com cauda poli A são detectados no núcleo da célula hospedeira, e parte desse RNA nuclear aparece como mRNA maduro e poliadenilado no citoplasma (Philipson et al., 1971). Outra semelhança importante refere-se à presença de estruturas 5' "cap" na extremidade 5' dos mRNAs, o que valida ainda mais o modelo de adenovírus como uma plataforma simplificada para os estudos dos mRNAs celulares (Moss e Koczot, 1976). É clássica a utilização de organismos modelo na área de genética e biologia molecular como organismos que satisfazem o estudo de determinado fenômeno agregando simplificação ao estudo para posterior generalização.

A determinação da sequência de nucleotídeos da extremidade 5' de vários mRNAs de Ad2 mostrou que a maior parte desses mensageiros possuía a mesma sequência de onze nucleotídeos iniciais (Gelinás; Roberts, 1977). Em híbridos mRNA:DNA, essa sequência de onze nucleotídeos era suscetível à digestão branda por ribonuclease T1 (Gelinás; Roberts, 1977), sugerindo que essa sequência de nucleotídeos não fosse codificada por uma região do DNA que estivesse localizada imediatamente a 5' da codificadora do restante do mRNA. Parecia que as sequências de RNA e DNA em questão não eram complementares ao longo de toda extensão. A descrição de uma técnica denominada "R-loop mapping"

(Thomas et al., 1976) contribuiu muito para a elucidação da estrutura da extremidade 5' desses transcritos. Se moléculas de RNA são incubadas com DNA, em condições nas quais híbridos RNA:DNA são favorecidos em relação a híbridos DNA:DNA, a fita de DNA com a mesma sequência de nucleotídeos que a do RNA será deslocada, formando uma alça que pode ser observada ao microscópio eletrônico, possibilitando o acompanhamento da colinearidade ou não entre sequências de RNA e DNA. Assim, a formação de uma alça corresponde à ausência de colinearidade entre os dois tipos de moléculas. Quando mRNAs de proteínas estruturais de Ad2 foram incubados com o DNA viral em condições para formação de alças de deslocamento, alças eram observadas nas regiões correspondentes às extremidades 5' dos mRNAs, assim como uma extensão na região correspondente à extremidade 3', onde a cauda de poli A é adicionada (Berget et al., 1977; Chow et al., 1977), mostrando que não havia colinearidade entre os dois tipos de moléculas nessas regiões.

O mapeamento físico do gene por meio da utilização de enzimas de restrição (mapeamento de restrição) foi fundamental na determinação das posições das alças de deslocamento. O mapeamento de restrição requer o uso de enzimas de restrição, as quais são enzimas específicas para a clivagem de uma molécula de DNA em determinados sítios específicos. O uso combinado de enzimas de restrição gera um mapa de um determinado fragmento de DNA onde distâncias relativas dos sítios de restrição podem ser medidas. Dessa forma, seria possível dar uma assinatura com um mapa de restrição às alças de deslocamento observadas. Fragmentos de outras regiões de Ad2 cuja posição fora determinada pelo mapeamento de restrição foram utilizados como referenciais em experimentos de alça de deslocamento, sendo assim possível a determinação das posições relativas onde ocorriam as projeções da extremidade 5'. Os resultados mostraram que vários mRNAs de Ad2 possuem extremidades 5' derivadas das mesmas três regiões localizadas em regiões que não são contíguas no genoma de Ad2. Esses três

5. População de mensageiros refere-se ao total de moléculas de mRNA de uma célula. Numa célula com função especializada, determinados mRNAs relacionados à função específica desse tipo celular são mais frequentes e presentes em grande quantidade.

pequenos blocos pareciam ser reunidos num único bloco, durante a biossíntese do mRNA, com eliminação das sequências interpostas (Berget et al., 1977; Chow et al., 1977).

Esses resultados levaram à formulação da hipótese de que os genes de Ad2 têm estrutura interrompida, ou seja, um DNA não codificador separaria regiões codificadoras (Berget et al., 1977; Chow et al., 1977). Mais ainda, esses resultados poderiam explicar a dinâmica da associação entre genes celulares, hnRNA e mRNAs. O hnRNA transcrito a partir de vários genes poderia ser processado a partir da remoção de sequências interpostas, gerando moléculas de mRNA citoplasmáticas menores (Berget et al., 1977; Chow et al., 1977).

Genes celulares interrompidos.

O modelo de ovalbumina

O caráter genérico desse tipo de padrão estrutural em eucariotos foi inicialmente obtido pela descrição do padrão em genes celulares cuja expressão fosse majoritária na população de mensageiros⁵ de um determinado tipo celular. Além dessa estratégia, foi fundamental o desenvolvimento de novas ferramentas de análise, como as técnicas de “Southern blot”, clonagem de cDNAs, construção de bibliotecas de DNA genômico e determinação da sequência de nucleotídeos de DNA.

O gene de ovalbumina de galinha foi o primeiro gene celular no qual a estrutura de genes interrompidos foi descrita (Breathnach et al., 1977). A ovalbumina é constituída de uma cadeia polipeptídica de 386 aminoácidos, e é sintetizada somente por células da glândula tubular do oviduto quando a ave está em fase de postura. A transcrição do gene que codifica ovalbumina é regulada por hormônios sexuais femininos. Na ausência desses hormônios o gene não é transcrito em mRNA e a proteína não é sintetizada. Com o objetivo de estudar essa regulação, foi fundamental a análise comparativa do funcionamento de genes de ovalbumina em dois contextos diferentes, no contexto de uma célula especializada na sua expressão e no de uma célula onde ovalbumina

não é expressa. Como o mRNA de ovalbumina representa 40% do mRNA total de células da glândula tubular do oviduto, a estratégia utilizada foi a obtenção de clones de moléculas de cDNA sintetizadas a partir do mRNA de ovalbumina isolado. Como esse mRNA era majoritário nessas células, seu isolamento tornava-se uma tarefa menos árdua, utilizando-se para esse procedimento a imunoprecipitação de RNA total de polissomo com anticorpos antiovalbumina (Humphries et al., 1977). Vários clones de cDNA de ovalbumina foram obtidos e, com a utilização de uma bateria de enzimas de restrição, um mapa de restrição desses cDNAs foi gerado. O próximo passo seria comparar a estrutura do gene de ovalbumina no DNA de células do oviduto de galinha com a estrutura do mesmo gene no contexto de células não relacionadas à expressão dessa proteína, tais como eritrócitos (Breathnach et al., 1977). A obtenção de bibliotecas de DNA genômico era uma tarefa que estava em desenvolvimento, e a análise diferencial da estrutura do gene de ovalbumina nos dois tipos celulares foi feita com a técnica de "Southern blot". Essa técnica consiste na hibridação de moléculas de DNA a ser investigado marcadas radioativamente, no caso um cDNA de ovalbumina, com as moléculas alvo que consistem nos produtos de digestão dos DNAs genômicos com enzimas de restrição imobilizados em substrato sólido. As enzimas escolhidas para a digestão dos DNAs genômicos foram EcoRI e *HindIII*, uma vez que o cDNA de ovalbumina isolado não possuía sítios de restrição para essas enzimas. Essa estratégia possibilitaria a detecção do gene de ovalbumina intacto. A hibridação com a sonda de cDNA de ovalbumina deveria resultar na detecção de uma única banda para esse gene de cópia única. Surpreendentemente, quatro bandas foram observadas em digestões com EcoRI e três bandas em digestões com *HindIII*. O padrão observado era exatamente o mesmo para DNAs de células de oviduto e de eritrócitos. Esses resultados indicavam que a estrutura do gene de ovalbumina deveria ser diferente da estrutura do cDNA de ovalbumina, apontando

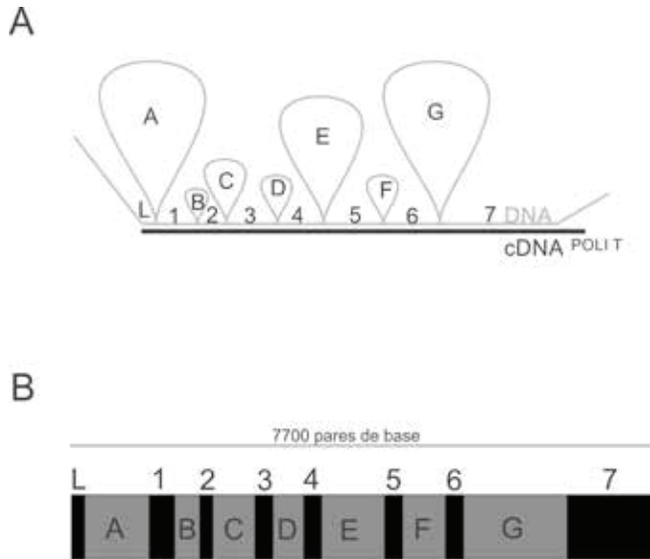
para explicação de genes interrompidos inicialmente observada em adenovírus (Breathnach et al., 1977).

Numa outra série de experimentos, a utilização de enzimas de restrição com sítios presentes tanto no cDNA quanto no DNA genômico permitiu a determinação de fragmentos correspondentes nos dois tipos de DNA. Foram obtidos mapas de restrição, tanto do DNA genômico quanto do cDNA de ovalbumina, e foram descritos vários segmentos presentes no DNA genômico, mas ausentes no cDNA (Mandel et al., 1978). Nessa mesma época, a análise de outros genes celulares, tais como os genes de globina (Jeffreys; Flavel, 1977; Tilghman et al., 1978b) e de imunoglobulinas (Tonegawa et al., 1978), agregou novas evidências para a generalização do padrão de genes interrompidos em eucariotos.

Com o objetivo de obter um mapa mais preciso, e já dispondo de técnicas mais avançadas na construção de bibliotecas de DNA genômico, assim como na detecção de clones, foi possível o isolamento de fragmentos do gene de ovalbumina. Esses fragmentos, em conjunto, compunham a totalidade do gene (Garapin et al., 1978a; Garapin et al., 1978b; Dugaiczky et al., 1978). De posse desses fragmentos de DNA genômico clonados, assim como do cDNA de ovalbumina, foi possível a aplicação da técnica de “R-loop mapping”. Sete alças de deslocamento no DNA genômico foram observadas, as quais corresponderiam a sete sequências interpostas (Garapin et al., 1978a; Garapin et al., 1978b; Dugaiczky et al., 1978). No entanto, a técnica de “R-loop mapping” possui limite de detecção de no máximo 50-100 bases, o que significaria a imprecisão na determinação do número de sequências interpostas com menos de 50-100 bases (Figura 1A).

Mais uma vez um grande avanço técnico na área, a determinação da sequência de nucleotídeos de DNA, possibilitou que a estrutura do gene de ovalbumina fosse determinada definitivamente. Sequências de nucleotídeos do cDNA e do DNA genômico de ovalbumina foram comparadas e os resultados confirmaram a presença de sete sequências interpostas no DNA genômico, as quais

Figura 1. Organização do gene de ovalbumina de galinha. (A) Representação esquemática de micrografia eletrônica de experimento de "R-loop" onde foram utilizados cDNA de ovalbumina e um fragmento de DNA que contém o gene de ovalbumina. Note que sete alças de deslocamento são observadas (A-G), e correspondem a regiões interpostas às oito regiões comuns ao DNA genômico e cDNA (L,1-7). A cauda de poli T, produto de transcrição reversa da cauda de poli A do mRNA, está representada. (B) Representação esquemática do gene de ovalbumina mostrando as regiões que estão presentes no mRNA (pretas). Ilustração de Henrique Nicoletto Mayer.



não apresentam correspondência no cDNA, e oito blocos de sequências do DNA genômico também presentes no mRNA. Esses resultados mostraram também que a ordem das sequências presentes em ambos, cDNA e DNA genômico, é a mesma (O'Hare et al., 1979). No que se diz respeito ao tamanho total do gene que codifica a ovalbumina, esse é de 7.700 pares de base, ou seja, cerca de quatro vezes maior que o mRNA maduro (1.872 pares de base) e sete vezes maior que o número de nucleotídeos do mesmo mRNA utilizados na tradução da proteína (1.158 pares de pares de base para 386 aminoácidos) (Figura 1B).

Introns, exons e "splicing"

Em 1978, Walter Gilbert, considerando a estrutura interrompida dos genes de eucariotos, sugeriu os seguintes termos: "introns", para designar as sequências de DNA de um gene qualquer que não estão presentes no mRNA maduro (sequências intervenientes, int = interveniente), e "exons", para designar as sequências presentes no mRNA maduro (sequências expressas, ex = expresso) (Gilbert, 1978).

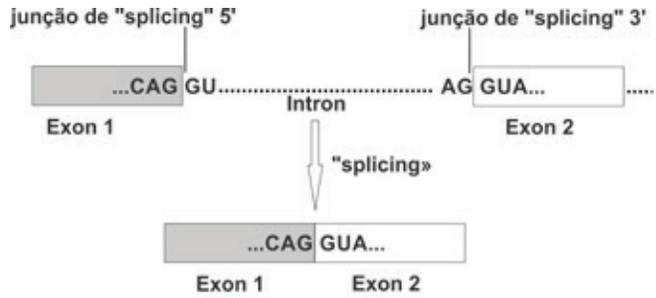
Assim, após a descoberta de introns nos genes de eucariotos, surge uma segunda questão: como, a

partir de uma estrutura de genes interrompidos, é gerado um mRNA maduro? Ou seja, como a informação contida nos introns é eliminada durante a biossíntese de mRNA? Várias hipóteses foram formuladas na época (Crick, 1979). Primeiro, o DNA na célula produtora do mRNA específico poderia ser rearranjado, de forma que os introns fossem deslocados para outra região do genoma ou até mesmo eliminados. Essa hipótese parecia ter respaldo em pelo menos um modelo, o dos genes de imunoglobulinas; no entanto, até aquele momento nenhum outro modelo fora caracterizado com a mesma configuração. Segundo, o DNA permaneceria inalterado, mas a RNA polimerase catalisaria a produção do transcrito primário, saltando os introns e gerando produtos que apenas possuísem exons. Terceiro, cada um dos exons seria transcrito independentemente, e posteriormente seriam montados na ordem correta para formar o mRNA maduro. Quarto, a RNA polimerase sintetizaria um transcrito primário onde toda sequência do gene estaria representada, seguindo-se o processamento desse transcrito, com eliminação dos introns para a geração do mRNA maduro (Crick, 1979).

Deve-se observar que, de todas as hipóteses acima relacionadas, apenas a quarta supõe a presença de sequências relativas a introns no transcrito primário. Consequentemente, a detecção dessas sequências em transcritos primários seria um argumento a favor dessa hipótese. Novamente a técnica de “R-loop” foi empregada com êxito na avaliação dessa hipótese. Nesses experimentos foi utilizado um RNA denominado 15S, que supostamente era precursor do mRNA maduro de β -globina, além do DNA codificador de β -globina. Esses experimentos mostraram que sequência correspondente ao intron estava presente no precursor (Tilghman et al., 1978a).

Com o desenvolvimento da técnica de “Northern blot” (Alwine et al., 1977), ou seja, transferência de RNA previamente fracionado por eletroforese em gel para um substrato sólido, foi possível a demonstração de que todas as sequências de um gene com estrutura interrompida, incluindo-se os introns, são transcritas em uma única molécula precursora

Figura 2. Representação esquemática mostrando a organização de um intron hipotético padrão. A junção de "splicing 5'" com o exon 1 a montante e a junção de "splicing 3'" com o exon 2 a jusante estão assinaladas. Note os dinucleotídeos conservados GU na extremidade 5' do intron e AG na extremidade 3' do intron. Com o processamento por "splicing", o intron é descartado e os exons reunidos. Ilustração de Henrique Nicoletto Mayer.



longa, a qual é processada para gerar o mRNA maduro (Roop et al., 1978). Assim, com o acúmulo dessa evidência para outros genes de eucariotos, ficou estabelecido que a grande maioria desses genes possui estrutura interrompida, e que a geração de mRNA ocorre por meio do processamento do precursor de mRNA (pré-mRNA), com a eliminação dos introns e reunião dos exons na mesma ordem em que estão representados no DNA genômico. Esse processamento de RNA foi denominado "splicing", e os limites entre os exons e introns foram denominados junções de "splicing". Considerando-se um intron a junção entre esse e o exon, a montante foi denominada junção de "splicing" 5', enquanto a junção entre o mesmo intron e um exon a jusante foi denominada junção de "splicing" 3'. Essas junções ainda podem ser denominadas sítios de "splicing", respectivamente sítio de "splicing" 5' ou sítio doador e, sítio de "splicing" 3' ou sítio aceptor, ainda que o termo "sítio" possa ser utilizado para designar regiões um pouco mais extensas que a própria junção, ou seja, a junção em questão mais sequências vicinais (Figura 2).

A estrutura de genes interrompidos e o processamento por "splicing" foram inicialmente descritos em genes que codificam proteínas. No entanto, quase que imediatamente após essa descoberta, introns foram descritos em genes cujo produto final são RNAs, tais como tRNA (Goodman et al., 1977; Valenzuela et al., 1978) e rRNA (Wild; Gall, 1979; Cech; Rio, 1979; Din et al., 1979; Din; Enberg, 1979),

dando um caráter mais amplo a esse tipo de estrutura gênica e de processamento.

Nos próximos artigos, pretendemos discorrer sobre a história da descoberta do mecanismo de “splicing” dos pré-mRNAs, dos introns autocatalíticos do grupo I, II, III e “twintrons”, trans-“splicing” e introns de tRNAs.

Agradecemos a Henrique Nicoletto Mayer pelas ilustrações.

Referências

- Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzoyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA*, 1977, n(74): 5350-5354.
- Berget SM, Moore C, Sharp PA. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA*, 1977, n(74): 3171-3175.
- Breathnach R, Mandel JL, Chambon P. Ovalbumin gene is split in chicken DNA. *Nature*, 1977, n(270): 314-319.
- Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 1977, n(12): 1-8.
- Cech TR, Rio DC. (1979). Localization of transcribed regions on extrachromosomal ribosomal RNA genes of *Tetrahymena thermophila* by R-loop mapping. *Proc. Natl. Acad. Sci. USA*, 1979, n(76): 5051-5055.
- Crick F. Split genes and RNA splicing. *Science*, 1979, n(204): 264-271.
- Darnell JE. The origin of mRNA and the structure of the mammalian chromosome. *Harvey Lect*, 1973, n(69): 1-47.
- _____, Wall R, Tushinski RJ. An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. *Proc. Natl. Acad. Sci. USA*, 1971, n(68): 1321-1325.

- Din N, Engberg J. Extrachromosomal ribosomal RNA genes in Tetrahymena: structure and evolution. *J. Mol. Biol.* 1979, n(134): 555-574.
- Din N, Engberg J, Kaffenberger W, Eckert WA. The intervening sequence in the 26S rRNA coding region of *T. thermophila* is transcribed within the largest stable precursor for rRNA. *Cell*, 1979, n(18): 525-532.
- Dugaiczyk A et al. The natural ovalbumin gene contains seven intervening sequences. *Nature*, 1978, n(274): 328-333.
- Edmonds M, Vaughan Jr. MH, Nakazato H. Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. *Proc. Natl. Acad. Sci. USA*, 1971, n(68): 1336-1340.
- Furuichi Y, Morgan M, Muthukrishnan S, Shatkin AJ. Reovirus messenger RNA contains a methylated, blocked 5'-terminal structure: m-7G(5') ppp(5') G-MpCp. *Proc. Natl. Acad. Sci. USA*, 1975, n(72): 362-366.
- Garapin AC et al. Electron microscopy and restriction enzyme mapping reveal additional intervening sequences in the chicken ovalbumin split gene. *Cell*, 1978, n(14): 629-639.
- Garapin AC et al. Isolation by molecular cloning of a fragment in the split ovalbumin gene. *Nature*, 1978b, n(273): 349-354.
- Gelinas RE, Roberts RJ. One predominant 5'-undecanucleotide in adenovirus 2 late messenger RNAs. *Cell*, 1977, n(11): 533-544.
- Gilbert W. Why genes in pieces? *Nature*, 1978, n(271): 501.
- Goodman HM, Olson MV, Hall BD. Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor SUP4-o. *Proc. Natl. Acad. Sci. USA*, 1977, n(74): 5453-5457.
- Humphries P et al. Molecular cloning of extensive sequences of the in vitro synthesized chicken ovalbumin structural gene. *Nucleic Acids Res*, 1977, n(4): 2389-2406.

- Jeffreys AJ, Flavell RA. The rabbit beta-globin gene contains a large insert in the coding sequence. *Cell*, 1977, n(12): 1097-1108.
- Lee SY, Mendecki J, Brawerman G. A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells. *Proc. Natl. Acad. Sci. USA*, 1971, n(68): 1331-1335.
- Lewin B. *Gene Expression*, 1980, v(2), 2nd. edn (New York, John Wiley & Sons).
- _____. (1982). *Gene Expression*, 1982, v(1), 6th. edn (New York, John Wiley & Sons).
- Mandel JL et al. Organization of coding and intervening sequences in the chicken ovalbumin split gene. *Cell*, 1978, n(14): 641-653.
- Moss B, Koczot F. Sequence of methylated nucleotides at the 5'-terminus of adenovirus-specific RNA. *J Virol*, 1976, n(17): 385-392.
- O'Hare K, Breathnach R, Benoist C, Chambon P. No more than seven interruptions in the ovalbumin gene: comparison of genomic and double-stranded cDNA sequences. *Nucleic Acids Res.*, 1979, n(7): 321-334.
- Phillipson L, Wall R, Glickman G, Darnell JE. Addition of polyadenylate sequences to virus-specific RNA during adenovirus replication. *Proc. Natl. Acad. Sci. USA*, 1971, n(68): 2806-2809.
- Roop DR et al. Transcription of structural and intervening sequences in the ovalbumin gene and identification of potential ovalbumin mRNA precursors. *Cell*, 1978, n(15): 671-685.
- Rottman F, Shatkin AJ, Perry RP. Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. *Cell*, 1974, n(3): 197-199.
- Thomas M, White RL, Davis RW. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc. Natl. Acad. Sci. USA*, 1976, n(73): 2294-2298.
- Tilghman SM et al. The intervening sequence of a mouse beta-globin gene is transcribed within

- the 15S beta-globin mRNA precursor. *Proc. Natl. Acad. Sci. USA*, 1978a, n(75): 1309-1313.
- Tilghman SM et al. Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. *Proc. Natl. Acad. Sci. USA*, 1978b, n(75): 725-729.
- Tonegawa S et al. Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. USA*, 1978, n(75): 1485-1489.
- Valenzuela P et al. Structure of yeast phenylalanine-tRNA genes: an intervening DNA segment within the region coding for the tRNA. *Proc. Natl. Acad. Sci. USA*, 1978, n(75): 190-194.
- Wei CM, Moss B. Methylation of newly synthesized viral messenger RNA by an enzyme in vaccinia virus. *Proc. Natl. Acad. Sci. USA*, 1974, n(68): 1336-1340.
- Wild MA, Gall JG. An intervening sequence in the gene coding for 25S ribosomal RNA of *Tetrahymena pigmentosa*. *Cell*, 1979, n(16): 565-573.

Data de recebimento: 04/03/2016

Data de aprovação: 20/05/2016