

Análise do Splicing Alternativo do Gene Hint-1 Através do Código BCH Associado

Luiz Antônio Leandro Franco* Reginaldo Palazzo Júnior†

Faculdade de Engenharia Elétrica e de Computação, FEEC, UNICAMP,
13083-859, Campinas, SP

E-mail: franco_luiz9@hotmail.com, palazzo@dt.fee.unicamp.br,

RESUMO

Cada célula no corpo humano deve crescer, reproduzir-se, processar informações, responder a estímulos e realizar uma série considerável de reações químicas. As células são envolvidas pela membrana celular e preenchidas com uma solução aquosa concentrada de substâncias químicas, substâncias físicas e o citoplasma. O citoplasma é o material celular localizado entre a membrana celular e o núcleo, sendo o local onde se concentra o maior número de atividades celulares, [1].

Uma molécula de ácido desoxirribonucleico (DNA) consiste de duas longas cadeias polipeptídicas compostas por quatro tipos de subunidades nucleotídicas. No caso dos nucleotídeos do DNA, o açúcar é uma desoxirribose ligada a um único grupo fosfato, cujas bases são: adenina (A), citosina (C), guanina (G) ou timina (T). As sequências codificantes de genes eucarióticos são caracteristicamente interrompidas por sequências intervenientes não-codificantes (íntrons). Tanto as sequências de íntrons quanto de éxons são transcritas em RNA. As sequências dos íntrons são removidas do RNA transcrito por meio de um processo denominado *splicing de RNA*. Grande parte do *splicing de RNA* que ocorre nas células atua na produção de mRNA, sendo denominado *splicing do precursor de mRNA* (ou pré-mRNA). Somente após ter ocorrido o *splicing* e o processamento das extremidades 5' e 3' esse RNA será denominado mRNA, [2].

O gene Hint-1 do *nematoide C. elegans* (identificado por "geneID" número 184760, no NCBI) é composto por 3 éxons e 2 íntrons com os seguintes comprimentos (em termos de nucleotídeos): éxon 1 com 123, íntron 1 com 44, éxon 2 com 138, íntron 2 com 74 e éxon 3 com 132, totalizando 511 nucleotídeos. Este gene foi identificado como uma palavra-código de um código BCH com parâmetros (511, 502, 3) sobre o anel \mathbb{Z}_4 com polinômio gerador $g(x) = x^9 + 2x^7 + x^5 + 3$, via o algoritmo de geração de sequências de DNA proposto em [3]-[5]. A matriz geradora G tem dimensão 502×511 . Por outro lado, sabemos do processo de codificação que a palavra-código \mathbf{v} (sequência do gene Hint-1) resulta da seguinte operação $\mathbf{v} = \mathbf{u}.G$. Desse processo temos duas possibilidades de análise. A primeira tem a ver com a caracterização das submatrizes associadas aos correspondentes éxons e íntrons do gene Hint-1 e suas relações. A segunda tem a ver com as componentes do vetor \mathbf{u} , que chamaremos *vetor de sinalização*.

Sob o ponto de vista do vetor sinalização, notamos que existem componentes deste vetor que são comuns tanto a éxons como a íntrons, mostrando uma forte ligação na região de fronteira. Uma interpretação biológica que fazemos do vetor sinalização \mathbf{u} é a de realizar a localização/identificação no DNA da sequência precursora do RNA, pré-RNA. O próximo passo é a obtenção do mRNA associado ao correspondente gene. Para isso, é necessário que o mecanismo de *splicing* do pré-mRNA entre em ação. Isto por sua vez implica que a maquinaria de *splicing* deve reconhecer três regiões na molécula precursora

*Bolsista de Mestrado - CAPES

†Pesquisa com suporte financeiro do CNPq e FAPESP

do RNA: a região de splicing 5', a região de splicing 3' e o ponto da forquilha na sequência do íntron que forma a base do fragmento em laço a ser excisado. Cada um desses três sítios tem uma sequência nucleotídica consenso, que é similar entre os íntrons e que fornece a posição onde deve ocorrer o splicing.

Sob o ponto de vista da matriz geradora G , o espaço vetorial gerado tem dimensão 502. Todavia, as dimensões dos subespaços correspondentes aos éxon 1, íntron 1, éxon 2, íntron 2, e éxon 3 apresentam os seguintes valores 123, 53, 147, 83, 132. Note que a soma dessas dimensões vale 538, portanto ultrapassando o valor 502. Isso implica que o espaço total não é uma soma direta dos correspondentes subespaços. Mais ainda, estabelece uma dependência entre os subespaços vizinhos. Essa dependência entre subespaços vizinhos nada mais é que uma memória associada. Biologicamente podemos inferir que um íntron estabelece um processo de "amarramento" entre os éxons subsequentes e que se mostram importantes tanto no aspecto da realização do splicing alternativo como no da confiabilidade. Ambos processos de vital importância para a conservação da espécie.

A quantidade de proteínas decorrentes do realização do processo de splicing alternativo do gene Hint-1, as isoformas e suas relações bem como as análises de "inserção" e "deleção" de nucleotídeos nos limiares dos pares éxon-íntron e íntron-éxon serão apresentadas tendo como base a utilização dos códigos de Tenengolts, [6].

Palavras-chave: *Códigos Corretores de Erros, Splicing Alternativo e Hint-1*

Referências

- [1] Marieb E. N., Hoehn K., *Anatomia e Fisiologia*, Artmed, 2009.
- [2] Bruce Alberts et al., *Biologia Molecular da Célula*, 5th Edition, Artmed, 2010.
- [3] Rocha ASL, Faria LCB, Kleinschmidt JH, Palazzo Jr R, Silva-Filho MC, "DNA sequences generated by \mathbb{Z}_4 -linear codes," *IEEE Intl. Symp. on Inform. Theory, ISIT2010*, pp. 1320-1324, 2010.
- [4] Faria LCB, Rocha ASL, Kleinschmidt JH, Palazzo Jr R, Silva-Filho MC, "DNA Sequences Generated by BCH Codes over GF(4)," *Electronics Letters*, vol. 46, pp. 202-203, 2010.
- [5] Faria LCB, Rocha ASL, Kleinschmidt JH, Silva-Filho MC, Bim E, Herai RH, Yamagishi MEB, Palazzo Jr R, "Is a Genome a Codeword of an Error-Correcting Code?," *Plos ONE*, vol. 7, No. 5, pp. e36644, 2012.
- [6] Tenengolts G, "Nonbinary codes, correcting single deletion or insertion," *IEEE Trans. Inform. Theory*, vol. 30, No. 5, pp. 766-769, 1984.