

## Teoria algébrica dos grafos e a estrutura secundária do RNA

Heber Cristina Teixeira<sup>1</sup>, Ana Paula de Paiva Lima<sup>2</sup>, Aline Marçal Rossinol<sup>3</sup>, Ellen Peixoto de Oliveira<sup>4</sup>

<sup>1,2,3,4</sup>Departamento de Matemática, Universidade Federal Viçosa, Viçosa-MG, Brasil

**Resumo:** O objetivo deste trabalho é estudar a predição da estrutura secundária da molécula do RNA, a partir de sua estrutura primária. A conhecida associação entre grafos e a representação bidimensional da estrutura secundária é utilizada. Por outro lado, a teoria algébrica de grafos fornece interessantes relação entre grafos e estruturas algébricas. Neste trabalho, à nível de iniciação científica, estudamos conexões entre estruturas algébricas e a estrutura secundária do RNA, via teoria de grafos.

**Palavras-chave:** RNA, Grafos.

### Introdução

O principal componente nas atividades biológicas em todo ser vivo são as proteínas. Todas as reações químicas que propiciam a existência de vida são executadas por elas, seja como enzimas que aceleram uma reação química, seja em atividades fins, tais como o movimento de um músculo. Embora existam milhares de proteínas distintas, todas são constituídas basicamente por apenas 20 aminoácidos. A informação que codifica a construção de uma proteína a partir destes 20 aminoácidos encontra-se codificada no núcleo das células, mais especificamente no DNA. Apesar de conter o código, a molécula do DNA possui uma rigidez em seu formato espacial, determinado pela dupla hélice, que por um lado propicia a estabilidade do código genético de uma espécie mas, por outro lado, prejudica a construção de proteínas, que possuem forma geométrica tridimensional difíceis de serem compiladas a partir da rigidez da dupla hélice do DNA. A transição do código contido no DNA, no núcleo da célula, para a produção da proteína, no citoplasma é feita pelo RNA. Este é o chamado dogma central da biologia, DNA→RNA→proteína. O RNA é um polímero formado por milhares de subunidades, chamados nucleotídeos. Ele é polimerizado a partir dos genes, agindo como peça fundamental na maquinaria da tradução celular. O fato de possuir apenas um filamento (ao invés da dupla hélice do DNA), chamado de estrutura primária, permite ao RNA a exibilidade necessária para moldar a forma tridimensional da proteína que irá sintetizar. A estrutura secundária do RNA, que é produzida por pontes de hidrogênio, é determinada, a partir de sua estrutura primária, ou seja, de sua sequência de nucleotídeos. Predizer as possíveis estruturas secundárias a partir de sua estrutura primária é o problema tratado neste projeto.

<sup>1</sup>heber.teixeira@ufv.br

<sup>2</sup>anapauladepaivalima@gmail.com

<sup>3</sup>alinerrossinol@gmail.com

<sup>4</sup>ellen.peixoto@ufv.br

## Desenvolvimento

Em 1978, Waterman e Smith [6] fizeram a primeira abordagem do RNA via grafos, marcando o início do uso da teoria espectral de grafos no estudo da estrutura secundária do RNA. Um grafo  $G$  consiste de um conjunto finito de vértices e de arestas. Associamos a um grafo uma matriz quadrada  $A_{n \times n}$ , onde  $n$  é o número de vértices. Cada entrada  $a_{ij}$  desta matriz expressa o número de arestas entre os vértices  $i$  e  $j$ . Como trabalhamos apenas com grafos simples, isto é, sem loops e com no máximo uma aresta entre dois vértices, a matriz de adjacências possui entradas 0 ou 1, e sua diagonal é toda nula. Seu polinômio característico  $p(x) = \sum_1^n a_{ii}x^i$  expressa em seus coeficientes algumas propriedades do grafo. Por exemplo,  $a_{n-2}$  é o número de arestas enquanto  $a_{n-3}$  o dobro do número de triângulos. A teoria espectral de um grafo é tema bastante estudado na teoria algébrica de grafos. Nesta área, várias estruturas algébricas como grupos e álgebras associativas são associadas a um grafo. Recentemente T. Schlick [5] associou o RNA a um par formado por um grafo árvore e seu dual obtendo, a partir do espectro das respectivas matrizes Laplacianas, resultado importante na predição da estrutura secundária do RNA. O desenvolvimento atual do projeto, ainda em andamento, consiste em estudar o trabalho de T. Schlick e sua equipe e, a partir dele, obter estruturas algébricas não lineares associadas ao RNA.

## Conclusões

Em nossa apresentação, exibiremos relações estabelecidas em [4] e [5] entre algumas moléculas do RNA e os correspondentes grafos árvore e dual, além das respectivas propriedades algébricas. O prosseguimento do trabalho visa, a partir da relação acima, obter resultados algébricos que expressem propriedades topológicas da estrutura secundária do RNA.

## Referências

- [1] M. Cox, J. Doudna, M. O'Donnell. *Biologia Molecular. Princípios e Técnicas*. Artmed, Porto Alegre, 2012.
- [2] C. Godsil, G. Royle. *Algebraic Graph Theory*. Springer, New York, 2001.
- [3] C. Reidys. *Combinatorial Computational Biology of RNA. Pseudoknots and Neutral Networks*. Springer, 2010.
- [4] T. Schlick. *Adventures with RNA graphs*. Methods, 2018.
- [5] H.H. Gan, D. Fera, J. Zorn, M. Tang, N. Shiffeldrim, U. Laserson, N. Kim, T. Schlick, *RAG: RNA-As-Graphics database-concepts, analysis, and features*. Bioinformatics 20 (2004) 1285-1291.
- [6] M.S. Waterman, T.F. Smith, *RNA secondary structure: a complete mathematical analysis*. Math. Biosci. 42 (1978) 257-266.