

## استفاده از روش‌های ترکیبیاتی برای شمارش ساختارهای RNA

غلام‌حسن شیردل<sup>۱\*</sup>، نسرين كهكشانی<sup>۲</sup>

۱. استادیار گروه ریاضی و علوم کامپیوتر، گروه ریاضی، دانشگاه قم، قم، ایران

۲. دانشجوی کارشناسی ارشد ریاضی، گروه ریاضی، دانشگاه قم، قم، ایران

مکان انجام تحقیق: گروه ریاضی، دانشگاه قم

مسئول مکاتبات: دکتر غلام‌حسن شیردل، گروه ریاضی، دانشگاه قم، پست الکترونیکی: shirdel@khayam.ut.ac.ir

تاریخ پذیرش: ۸۹/۴/۱۳

تاریخ دریافت: ۸۸/۹/۳

### چکیده

تحقیق در زیست‌شناسی غالباً بر پایه آزمایش با مواد استوار است؛ در حالی که در ریاضیات زیستی، آزمایش‌ها دارای ماهیت نظری است. یک ضرورت اصلی به کار بردن ریاضیات در سیستم‌های زیستی، توانایی ساخت مدل‌های ریاضی است. این مدل‌ها، سیستم‌های ریاضی هستند که برهم‌کنش‌های پیچیده دستگاه‌های زیستی را با روشی ساده ارائه می‌کنند تا تحلیل ویژگی‌های آن راحت‌تر صورت گیرد. در حقیقت، ریاضیات در اکثر شاخه‌های زیست‌شناسی به کار می‌رود. شمارش تعداد ساختارهای RNA، یک روش ترکیبیاتی در ریاضیات زیستی است. در این مقاله، ساختارهای RNA به صورت گراف، مدل‌سازی شده است. دو پارامتر روی این مدل در نظر می‌گیریم. یکی از آن‌ها، طول یال است که متناظر با پیوند هیدروژنی است. دیگری، تعداد یال‌های دو به دو متقاطع است که تقاطع یال‌ها متناظر با وجود شبه‌گره در ساختار RNA است. حال با توجه به این دو پارامتر، ساختارهای RNA را طبقه‌بندی و شمارش می‌کنیم.

واژه‌های کلیدی: نمودار  $k$  غیرمتقاطع، شبه‌گره، شمارش ترکیبیاتی، اتاق وایل

### مقدمه

اسیدهای نوکلئیک ریبوزدار (RNA) و اسیدهای نوکلئیک د.کسی ریبوزدار (DNA) دیده می‌شوند. در ساختمان RNA، همیشه بازهای آلی  $A$ ،  $C$ ،  $G$ ،  $U$  و قند پنتوز (ریبوز)، در حالی که در DNA، بازهای آلی  $A$ ،  $T$ ،  $C$ ،  $G$  و قند پنتوز (د.کسی ریبوز) حضور دارند. برای تشکیل زنجیرهای طولی اسیدهای نوکلئیک، واحدهای نوکلئوتید، از محل گروه فسفات خود به هم می‌پیوندند. یک رشته اسیدنوکلئیک می‌تواند از هزاران نوکلئوتید به ترتیب‌های مختلف از نظر بازهای آلی چهارگانه‌ای که در ساختمان DNA و RNA شرکت می‌نمایند، به وجود آید. در سال ۱۹۵۳، جیمز واتسن و فرانسیس کریک، مدلی برای ساختار سه

اسیدنوکلئیک منبع اطلاعات وراثتی موجود زنده است، به عبارت دیگر، عاملی است که باعث انتقال خصوصیات یک نوع جاندار از نسلی به نسل دیگر می‌شود. اسیدهای نوکلئیک از تکرار چهار واحد مختلف تشکیل می‌شوند که هر یک را نوکلئوتید می‌نامند. بنیان‌های سازنده هر واحد نوکلئوتید را یک باز آلی نیتروژن‌دار، یک قند پنتوز و یک بنیان فسفات تشکیل می‌دهد. بازهای آلی شرکت کننده در ساختمان اسیدهای نوکلئیک عبارتند از: آدنین ( $A$ )، تیمین ( $T$ )، سیتوزین ( $C$ )، اوراسیل ( $U$ ) و گوانین ( $G$ ). دو نوع قند پنتوز در ساختمان اسیدهای نوکلئیک وجود دارد: د.کسی ریبوز و ریبوز. بنابراین، اسیدهای نوکلئیک در بدن موجود زنده به صورت

مکمل را قرار می‌دهد و هر ریبونوکلئوتید جدید را به ریبونوکلئوتید قبلی وصل می‌کند. با پایان یافتن عمل رونویسی، RNA تک‌رشته‌ای از RNA جدا می‌شود. RNA مانند پروتئین، سه ساختار اصلی اول، دوم و سوم دارد. ساختار اول، شامل توالی خطی RNA است که از بازهای  $A$ ،  $C$ ،  $U$  و  $G$  تشکیل شده است. ساختار دوم، نمایش سطحی مولکول RNA است که شامل برقراری رابطه مکملی بین نواحی تک‌رشته‌ای است. برقراری رابطه مکملی بر اساس مدل واتسون و کریک، بین بازهای  $A-U$  و  $G-C$  صورت می‌گیرد، البته تشکیل پیوند هیدروژنی، بین جفت بازهای  $G-U$  نیز انجام می‌شود که پیوندی ضعیف است. ساختار سوم، آرایش سه بعدی بازهای مولکول RNA است. بر اساس آرایش جفت بازها در ساختارهای دوم، چهار زیرگروه اصلی از ساختارهای دوم، شامل حلقه‌های سنجاق‌سری، حلقه‌های برآمده، حلقه‌های داخلی و حلقه‌های چندشاخه هستند. در ساختار دوم، ریبونوکلئوتیدهایی که در پیوندهای هیدروژنی شرکت دارند، ساقه ساختار را به وجود می‌آورند. تعداد نوکلئوتیدها در حلقه‌های سنجاق‌سری، حداقل سه و تعداد پیوندهای هیدروژنی در ساقه، حداقل دو است. علاوه بر این، جفت باز بین حلقه‌ها در ساختار دوم نیز رخ می‌دهد که در این مقاله به‌طور کلی آن‌ها را شبه‌گره می‌نامیم. ساختار دوم RNA، به‌صورت گراف مسطح در صفحه رسم می‌شود. در این مقاله، ساختار دوم به‌صورت نمودار نمایش داده می‌شود. برای انجام این کار به صورت زیر عمل می‌کنیم:

۱- بازهای  $A$ ،  $U$ ،  $C$  و  $G$  در ساختار دوم را به ترتیب روی یک خط افقی یادداشت می‌کنیم و به‌ازای هر یک از این بازها یک راس در نظر می‌گیریم.

۲- رئوس متناظر با هر دو بازی را که پیوند هیدروژنی با هم برقرار کرده‌اند، به‌وسیله یک یال به‌هم وصل می‌کنیم.

۳- رئوس را از چپ به راست، شماره‌گذاری می‌کنیم.

در شکل ۱، یک ساختار RNA و نمودار متناظر آن، نمایش داده شده است. حال، رئوس و یال‌ها در

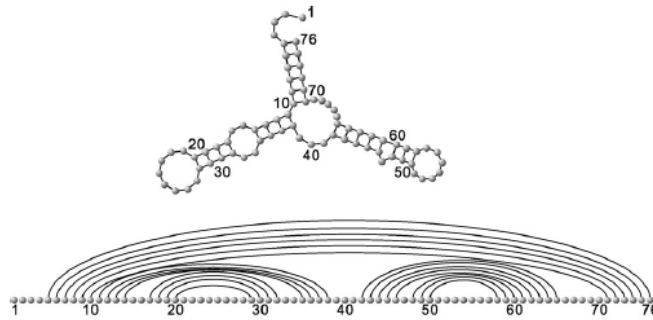
بعدی DNA ارائه دادند DNA بر طبق این مدل، از دو رشته پلی‌نوکلئوتیدی تشکیل شده است که مقابل هم قرار می‌گیرند و حول یک محور فرضی پیچ می‌خورند. گروه‌های قند و فسفات که آب دوست هستند، ستون هر کدام از دو رشته را تشکیل می‌دهند و در تماس با محیط آبی قرار دارند. بازها که تمایل کمتری به آب دارند، در داخل مارپیچ و عمود بر محور آن قرار می‌گیرند؛ به‌طوری که همواره  $A$  در مقابل  $T$  و  $C$  در مقابل  $G$  قرار می‌گیرد. این رابطه، به رابطه مکملی بین نوکلئوتیدها معروف است. بین بازهای مقابل هم، پیوندهای هیدروژنی تشکیل می‌شود که به ساختار دو رشته‌ای DNA پایداری می‌بخشد. اطلاعات در DNA به صورت رمز ذخیره شده‌اند. از اطلاعات موجود در DNA برای ساختن پروتئین‌ها استفاده می‌شود. DNA به‌طور مستقیم در سنتز پروتئین دخالت ندارد. زمانی که سلول به یک پروتئین خاص احتیاج داشته باشد، ناحیه رمزگذار این پروتئین در DNA شناسایی و به صورت RNA کپی‌برداری می‌شود. این RNA را که از روی قسمت کوچکی از DNA کپی‌برداری شده است و به‌عنوان الگو به‌طور مستقیم در سنتز پروتئین دخالت دارد RNA پیام‌بر می‌نامند. در هر ثانیه، هزاران فرایند تبدیل اطلاعات از DNA به پروتئین در هر سلول بدن ما در حال رخ دادن است. شرح مکانیسم‌هایی را که به‌وسیله آن اطلاعات از روی RNA به‌صورت RNA کپی‌برداری می‌شود، رونویسی می‌گویند. تمام RNAهای موجود در سلول از طریق رونویسی ساخته می‌شوند. آنزیم‌هایی که عمل رونویسی را انجام می‌دهند، به RNA پلی‌مراز معروف هستند. مراحل رونویسی به‌صورت زیر است:

۱- رونویسی با اتصال RNA پلی‌مراز به قسمتی از ژن به نام راه‌انداز ژن شروع می‌شود. راه‌انداز، قسمتی از RNA است که به RNA پلی‌مراز امکان می‌دهد رونویسی را از محل صحیح آغاز کند.

۲- RNA پلی‌مراز، دو رشته RNA را از یکدیگر باز می‌کند.

۳- RNA پلی‌مراز در طول نوکلئوتیدهای RNA به حرکت در می‌آید و در مقابل هر یک از د. کسی ریبونوکلئوتیدهای RNA، ریبونوکلئوتید

نمایش نمودار به ترتیب با نوکلئوتیدها و پیوند هیدروژنی بین آن‌ها متناظر هستند.



شکل ۱- ساختار دوم و نمایش نمودار آن (۵).

ساختار  $RNA$  را به صورت تابلو نوسانی و یک گام که در داخل اتاق وایل قرار دارد، نمایش می‌دهیم.

تعریف (۱) نمودار یانگ، مجموعه‌ای از مربع‌ها است که در سطرهایی از سمت چپ مرتب شده‌اند؛ به طوری که تعداد مربع‌ها در هر سطر به طور نه لزوماً اکید، کاهش می‌یابند. تابلو یانگ، با قرار دادن اعداد در نمودار به دست می‌آید؛ به طوری که اعداد باید در هر سطر از چپ به راست، به طور نه لزوماً اکید و در هر ستون از بالا به پایین، به طور اکید، افزایش یابند. یک تابلو، استاندارد نامیده می‌شود اگر هر عدد دقیقاً یک بار رخ دهد (۷).

تعریف (۲) تابلو نوسانی به طول  $n$ ، یک دنباله  $\phi = \mu^1, \mu^2, \dots, \mu^n = \phi$  از نمودارهای یانگ است؛ به طوری که برای هر  $1 \leq i \leq n$ ،  $\mu^i$  از  $\mu^{i-1}$  با اضافه کردن یک مربع یا حذف یک مربع یا بدون تغییر به دست می‌آید (۷).

الگوریتم  $RSK$ ، فرایند درج سطری عناصر داخل یک تابلو است. تابلو استاندارد یانگ  $\lambda$  را در نظر بگیرید. فرض کنید  $j, \lambda_j$ ، عنصر سطر  $i$  و ستون  $j$  در تابلوی  $\lambda$  را نشان دهد. می‌خواهیم عدد  $k$  را در این تابلو وارد کنیم. مراحل الگوریتم به صورت زیر است:

۱- قرار دهید  $i = 1$ .

۲- اگر در سطر  $i$ ، مربعی وجود ندارد، عدد  $k$  را در ابتدای این سطر قرار دهید و توقف کنید. در غیر این صورت، به مرحله ۳ بروید.

در ساختار  $RNA$ ، هر نوکلئوتید حداکثر با یک نوکلئوتید دیگر پیوند هیدروژنی برقرار می‌کند؛ پس درجه هر راس در نمایش نمودار، حداکثر یک است. اگر در ساختار دوم  $RNA$ ، شبه‌گره وجود داشته باشد، در نمایش نمودار متناظر آن، یال‌های دو به دو متقاطع وجود خواهند داشت. حال، نمودار  $k$ - غیرمتقاطع، نموداری خواهد بود که شامل یک مجموعه  $k$ -تایی از یال‌های دو به دو متقاطع نشود. به عبارت دیگر:

طول یال  $(i, j)$ ،  $j - i$  است. یک راس را هرگاه درجه آن صفر باشد، تنها می‌نامیم.

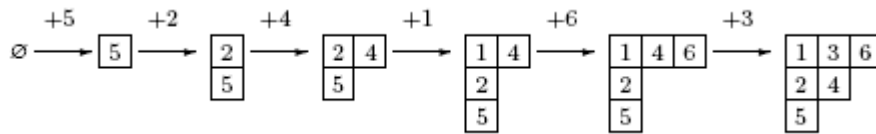
دو پارامتر  $k$  و  $\lambda$  را روی نمودارها در نظر می‌گیریم. در این مقاله، با توجه به این دو پارامتر، ساختارهای  $RNA$  را طبقه‌بندی می‌کنیم و با به کار بردن مفاهیم ترکیبیاتی، روابطی را برای محاسبه تعداد ساختارهای  $RNA$  به دست می‌آوریم (۴). برای اولین بار، بررسی ترکیبیات ساختارهای  $RNA$  توسط میشل واترمن آغاز شد (۸، ۹). واترمن، بررسی ساختارهای  $RNA$  با  $k = 2$  را آغاز کرد. اما هنوز تلاش‌ها برای بررسی ترکیبیات ساختارهای شبه‌گره  $RNA$  ادامه دارد.

### تناظر بین ساختارها و گام‌ها

راه‌های متفاوتی برای نمایش ساختارهای  $RNA$  وجود دارد که در مقدمه، به نحوه نمایش آن به صورت نمودار اشاره شد. حال در این بخش، یک

در نتیجه، یک تابلو یانگ جدید به دست می آید که شامل عدد  $k$  می شود.

مثال (۱) دنباله اعداد ۳، ۶، ۱، ۴، ۲، ۵ را طبق الگوریتم  $RSK$  و با شروع از نمودار  $\phi$  وارد می کنیم که دنباله ای از تابلوهای استاندارد یانگ حاصل می شود. این دنباله در شکل ۲ نشان داده شده است.



شکل ۲- دنباله ای از تابلوهای استاندارد یانگ (۴)

مبدا یک یال بود، این مبدا را از تابلو حذف می کنیم و اگر راس مورد نظر، تنها بود، تابلو بدون تغییر تکرار می شود. بنابراین، یک تابلو نوسانی به وجود می آید (توجه داشته باشید که از نمودار  $\phi$  شروع می کنیم).

حال گام  $\gamma_{a,a}^n$  از تابلو نوسانی بالا به صورت زیر به دست می آید:

از  $\phi$  در سمت چپ تابلو نوسانی شروع می کنیم. متناظر با هر تابلو، یک قدم وجود دارد که مختص  $x_i$  در قدم  $s$  برابر یک است اگر مربعی به سطر  $i - s$  ام تابلو اضافه شده باشد و برابر  $(-1)$  است، اگر مربعی از سطر  $i$  ام حذف شده باشد. در غیر این صورت،  $x_i$  برابر صفر است. توجه داشته باشید که فقط یکی از سه حالت فوق در  $s$  اتفاق می افتد. بنابراین با قراردادن  $a = (k-1, \dots, 1)$  گام مورد نظر به دست می آید.

برای درک بهتر مطالب بیان شده، مثال زیر را ارائه می کنیم.

مثال (۲) نمودار ۵-غیرمقاطع شکل ۳ را در نظر بگیرید. تابلو نوسانی متناظر با آن، در شکل ۴ آورده شده است. گام متناظر با آن به صورت زیر است:

- (4,3,2,1), (5,3,2,1), (6,3,2,1), (6,4,2,1), (6,4,2,1), (6,4,3,1), (6,4,3,1), (6,4,3,2), (5,4,3,2), (5,4,3,1), (6,4,3,1), (6,4,3,1), (6,4,2,1), (6,4,2,1), (6,3,2,1), (5,3,2,1), (4,3,2,1).

۳- بزرگترین عدد صحیح،  $j$ ، را به گونه ای پیدا کنید که  $\lambda_{i,j-1} \leq k$ . اگر  $\lambda_{i,j} > k$ ، قرار دهید  $j = 1$ .

۴- اگر  $\lambda_{i,j}$  وجود دارد، عدد  $k$  را به جای  $\lambda_{i,j}$  قرار دهید و سپس به ترتیب با قرار دادن  $k = \lambda_{i,j}$  و  $i = i + 1$  به مرحله ۲ بروید. اگر  $\lambda_{i,j}$  وجود ندارد، عدد  $k$  را در انتهای سطر  $i$  قرار دهید و توقف کنید.

شبه  $z^{k-1}$  و  $a, b \in z^{k-1}$  را در نظر بگیرید. یک گام از  $a$  تا  $b$  به طول  $n$ ،  $\gamma_{a,b}^n$ ، تایی  $(s_1, \dots, s_n)$  است؛ به طوری که  $s_h \in \{0, \pm e_i \mid 1 \leq i \leq k-1\}$

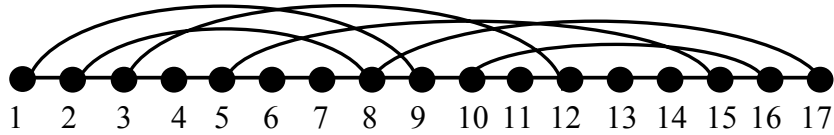
و  $b = a + \sum_{h=1}^n s_h$  هر  $s_h$  را یک قدم می نامیم.  $e_i$ ،  $i$ -امین بردار واحد را نشان می دهد. اتاق وایل را به صورت زیر تعریف می کنیم:

$C_0 = \{x \in z^{k-1} \mid 0 \leq x_{k-1} \leq x_{k-2} \leq \dots \leq x_1\}$   
متناظر با هر نمودار  $k$ -غیرمقاطع، یک تابلو نوسانی و در نتیجه یک گام  $\gamma_{a,a}^n$  در  $z^{k-1}$  وجود دارد که به صورت زیر به دست می آید:

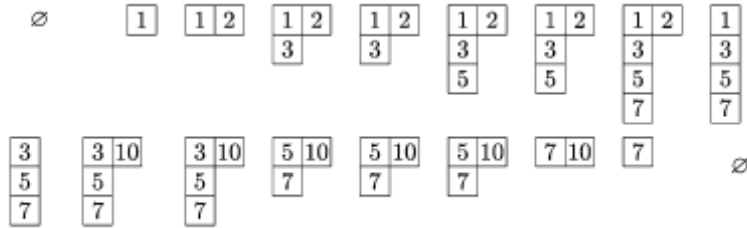
$$\exists (i_1, j_1), (i_2, j_2), \dots, (i_r, j_r);$$

$$i_1 < i_2 < \dots < i_r < j_1 < j_2 < \dots < j_r.$$

نمودار  $k$ -غیرمقطعی را در نظر بگیرید. از راست به چپ، یعنی از آخرین راس نمودار شروع می کنیم. اگر به راسی برخورد کردیم که انتهای یک یال بود، مبدا متناظر آن را توسط الگوریتم  $RSK$  وارد نمودار می کنیم. اگر به راسی برخورد کردیم که



شکل ۳- نمودار ۵-غیرمتقاطع (۴)



شکل ۴- تابلو نوسانی به طول ۱۷ (۴)

$g_{n,k}$  و مجموعه گام‌های  $\gamma_{a,a}^n$  که درون اتاق وایل قرار می‌گیرند، وجود دارد.  $g_{n,k} \rightarrow \Gamma_n^+(a,b)$  قضیه ۲ (۲) فرض کنید  $a, b \in C_0$ . در این صورت داریم:

برای  $a, b \in C_0$ ، فرض کنید  $\Gamma_n^+(a,b)$ ، تعداد گام‌های  $\gamma_{a,b}^n$  را نشان دهد که با هیچ دیواری برخورد نمی‌کنند، یعنی درون  $C_0$  قرار می‌گیرند. به‌طور دقیق، رابطه دوسویی بالا ثابت می‌شود. پس قضیه زیر را بیان می‌کنیم.  
قضیه ۱ (۱) فرض کنید  $a = (k-1, k-2, \dots, 1)$ . یک دوسویی بین مجموعه نمودارهای  $k$ -غیرمتقاطع روی  $n$  راس،

$$\sum_{n \geq 0} \Gamma_n^+(a,b) \frac{x^n}{n!} = e^x \det [I_{a_i - b_j}(2x) - I_{a_i + b_j}(2x)] \Big|_{i,j=1}^{k-1},$$

نمودار  $k$ -غیرمتقاطع با طول یال دلخواه را یک انطباق  $k$ -غیرمتقاطع می‌نامیم.

فرض کنید که  $a = b = (k-1, \dots, 2, 1) \in C_0$  باشد. حال طبق قضیه ۱ داریم:

$$|g_{n,k}| = \Gamma_n^+(a,b)$$

از طرفی با توجه به  $|g_{n,k}| = \sum_{\ell=0}^n f_k(n, \ell)$  و قضیه ۳، روابط زیر به‌دست می‌آیند:

$$\sum_{n \geq 0} \sum_{\ell=0}^n f_k(n, \ell) \frac{x^n}{n!} = e^x \det [I_{i-j}(2x) - I_{i+j}(2x)] \Big|_{i,j=1}^{k-1},$$

$$\sum_{n \geq 0} f_k(n, \cdot) \frac{x^n}{n!} = \det [I_{i-j}(2x) - I_{i+j}(2x)] \Big|_{i,j=1}^{k-1}.$$

$$I_r(2x) = \sum_{j \geq 0} \frac{x^{2j+r}}{j!(j+r)!}$$

تعریف ۳) یک نمودار  $k$ -غیرمتقاطع با طول یال دلخواه که دقیقاً  $\ell$  راس تنها دارد را یک انطباق جزئی  $k$ -غیرمتقاطع می‌نامیم (۶). تعداد انطباق‌های جزئی  $k$ -غیرمتقاطع روی  $n$  راس را با  $f_k(n, \ell)$  نشان می‌دهیم. اگر  $\ell = 0$  باشد،

$$f_{\nu}(n, \ell) = \binom{n}{\ell} \left[ C_{\frac{n-\ell}{\nu} + \nu} - C_{\frac{n-\ell}{\nu}} - C_{\frac{n-\ell}{\nu} + 1}^{\nu} \right],$$

که  $C_m = \frac{1}{m+1} \binom{2m}{m}$  -امین عدد کاتالان را نشان می‌دهد.

$$S_k(n, \ell) = \sum_{b=0}^{\nu} (-1)^b \binom{n-b}{b} f_k(n - \nu b, \ell),$$

$$S_k(n) = \sum_{b=0}^{\lfloor \frac{n}{\nu} \rfloor} (-1)^b \binom{n-b}{b} \sum_{\ell=0}^{n-\nu b} f_k(n - \nu b, \ell).$$

اگر در ساختار دوم  $RNA$ ، بین اولین و آخرین نوکلئوتید پیوند فسفری ایجاد شود، در نمایش نمودار آن، بین دو راس ۱ و  $n$  یالی وجود نخواهد داشت. بنابراین، تعریف زیر را ارائه می‌کنیم:

تعریف (۵) یک ساختار  $k$ -غیرمقاطع مدور  $RNA$ ، یک ساختار  $k$ -غیرمقاطع  $RNA$  است که شامل یال  $(1, n)$  نمی‌شود (۴).

تعداد ساختارهای  $k$ -غیرمقاطع مدور  $RNA$  با دقیقاً  $\ell$  راس تنها روی  $n$  را با  $S_k^{(c)}(n, \ell)$  نمایش می‌دهیم و تعداد ساختارهای  $k$ -غیرمقاطع مدور  $RNA$  روی  $n$  را با  $S_k^{(c)}(n)$  نشان می‌دهیم.

قضیه (۴) فرض کنید  $k \in N$  و  $k \geq 2$  باشد؛ در این صورت داریم:

$$S_k^{(c)}(n, \ell) = \sum_{b=0}^{\frac{n-\ell}{\nu}} (-1)^b \left[ \binom{(n-\nu) - (b-1)}{b-1} + \binom{n-b}{b} \right] f_k(n - \nu b, \ell),$$

$$S_k^{(c)}(n) = \sum_{b=0}^{\lfloor \frac{n}{\nu} \rfloor} (-1)^b \left[ \binom{(n-\nu) - (b-1)}{b-1} + \binom{n-b}{b} \right] \sum_{\ell=0}^{n-\nu b} f_k(n - \nu b, \ell).$$

نداشته باشد؛ به عبارت دیگر، طول کمان در آن، بزرگ‌تر یا مساوی ۳ است (۴). تعداد ساختارهای  $k$ -غیرمقاطع محدود  $RNA$  با دقیقاً  $\ell$  راس تنها روی  $n$  را با  $S_k^{(r)}(n, \ell)$  نمایش می‌دهیم و تعداد

با قرار دادن  $k = 2$  و  $k = 3$  در رابطه اول، نتیجه زیر به دست می‌آید.

نتیجه (۱) اگر  $(n - \ell)$  زوج باشد، داریم:

$$f_{\nu}(n, \ell) = \binom{n}{\ell} C_{\frac{n-\ell}{\nu}},$$

### طبقه‌بندی و شمارش ساختارهای $RNA$

تعریف (۴) یک ساختار  $k$ -غیرمقاطع  $RNA$  یک نمودار  $k$ -غیرمقاطع با طول یال بزرگ‌تر مساوی ۲ است (۴). تعداد ساختارهای  $k$ -غیرمقاطع  $RNA$  روی  $n$  را با  $S_k(n)$  و تعداد ساختارهای  $k$ -غیرمقاطع  $RNA$  با دقیقاً  $\ell$  راس تنها روی  $n$  را با  $S_k(n, \ell)$  نشان می‌دهیم.

به ساختارهای  $k$ -غیرمقاطع  $RNA$  که در آن‌ها  $k \geq 3$  است، ساختارهای شبه‌گره  $RNA$  هم گفته می‌شود.

در شکل ۱، یک ساختار ۲-غیرمقاطع  $RNA$  و در شکل ۳، یک ساختار ۵-غیرمقاطع  $RNA$  نشان داده شده است.

قضیه (۳) فرض کنید  $k \in N$  و  $k \geq 2$  باشد؛ در این صورت داریم:

ساختارهای  $RNA$  واقعی، یال‌هایی به شکل  $(i, i+1)$ ،  $(i, i+2)$ ،  $(i, i+3)$  را ندارند. حال ایده‌هایمان را به شمارش ساختارهای  $RNA$  محدود تعمیم می‌دهیم.

تعریف (۶) یک ساختار  $k$ -غیرمقاطع  $RNA$  را محدود می‌نامیم اگر یالی به شکل  $(i, i+2)$  را

قضیه (۵) (۴) فرض کنید  $k \geq 2$  و  $k \in N$  باشد؛ در این صورت داریم:

ساختارهای  $k$ -غیرمقاطع محدود  $RNA$  روی  $n$  را با  $S_k^{(r)}(n)$  نشان می‌دهیم.

$$S_k^{(r)}(n, \ell) = \sum_{b_1 \geq 0, b_2 \geq 0} (-1)^{b_1 + b_2} \lambda(n, b_1, b_2) f_k(n - 2(b_1 + b_2), \ell),$$

$$S_k^{(r)}(n) = \sum_{b_1 \geq 0, b_2 \geq 0}^{[\frac{n}{2}]} (-1)^{b_1 + b_2} \lambda(n, b_1, b_2) \sum_{\ell=0}^{n-2(b_1+b_2)} f_k(n - 2(b_1 + b_2), \ell),$$

$n$  های مختلف است. ابتدا با استفاده از فرمول‌هایی که برای انطباق‌ها به دست آمد، تعداد انطباق‌ها را محاسبه کرده، سپس با قرار دادن در فرمول‌های بخش ۳، تعداد ساختارها را به دست می‌آوریم. حال با استفاده از این روابط، ۲۰ عدد نخست تعدادی از این ساختارها را محاسبه می‌کنیم که نتیجه، در قالب جدول ۱ آورده شده است.

که  $\lambda(n, b_1, b_2)$  تعداد راه‌های انتخاب دقیقاً  $b_1$  تا یال به طول ۱ و  $b_2$  تا یال به طول ۲ روی  $n$  را نشان می‌دهد. در بخش ۳، ساختارهای  $RNA$  را طبقه‌بندی کرده و سپس روابطی را برای محاسبه تعداد این ساختارها به دست آوردیم. حال نتیجه اصلی این مقاله، به دست آوردن تعدادی از این ساختارها به‌ازای

جدول ۱- بیست عدد نخست برخی از ساختارهای  $RNA$

$S_2^{(r)}(n)$	$S_2^{(c)}(n)$	$S_2(n)$	$S_2(n)$	$n$
۱	۱	۱	۱	۱
۱	۱	۱	۱	۲
۱	۱	۲	۲	۳
۲	۴	۵	۴	۴
۵	۱۱	۱۳	۸	۵
۱۴	۳۱	۳۶	۱۷	۶
۴۰	۹۲	۱۰۵	۳۷	۷
۱۱۹	۲۸۵	۳۲۱	۸۲	۸
۳۶۴	۹۱۳	۱۰۱۸	۱۸۵	۹
۱۱۴۵	۳۰۱۳	۳۳۳۴	۴۲۳	۱۰
۳۶۸۸	۱۰۱۹۸	۱۱۲۱۶	۹۷۸	۱۱
۱۲۱۳۹	۳۵۳۰۱	۳۸۶۳۵	۲۲۸۳	۱۲
۴۰۷۳۴	۱۲۴۶۱۹	۱۳۵۸۳۵	۵۳۷۳	۱۳
۱۳۹۰۷۱	۴۴۷۷۰۲	۴۸۶۳۳۷	۱۲۷۳۵	۱۴
۴۸۲۲۱۴	۱۶۳۳۶۶۵	۱۷۶۹۵۰۰	۳۰۳۷۲	۱۵
۱۶۹۵۴۶۹	۶۰۴۵۴۵۹	۶۵۳۱۷۹۶	۷۲۸۳۲	۱۶
۶۰۳۶۷۶۸	۲۲۶۵۶۲۵۸	۲۴۴۲۵۷۵۸	۱۷۵۵۰۲	۱۷
۲۱۷۴۰۹۶۹	۸۵۸۸۸۲۳۰	۹۲۴۲۰۰۲۶	۴۲۴۷۴۸	۱۸
۷۹۱۱۷۸۲۲	۳۲۹۰۱۸۴۶۰	۳۵۳۴۴۴۲۱۸	۱۰۳۲۰۰۴	۱۹
۲۹۰۶۷۴۴۷۰	۱۲۷۲۵۱۳۶۹۳	۱۳۶۴۹۳۳۷۱۹	۲۵۱۶۳۴۷	۲۰

## نتایج و بحث

حال گراف‌هایی را که روی  $n$  راس با یک سری شرایط، قابل رسم هستند، شمارش می‌کنیم. با انجام این کار در واقع تعداد ساختارهای دوم را که روی  $n$  نوکلئوتید وجود دارند، به دست می‌آوریم. با این روش، تعداد ساختارها با نرخ رشد کمتری به دست می‌آیند و در نتیجه، روش جدیدتری برای پیش‌بینی ساختارهای  $RNA$  ارائه می‌شود (۳).

با توجه به تعاریف بیان شده، ساختارهای محدود، یک شرط اضافه بر ساختارهای  $k$ -غیرمقاطع دارند که آن، محدود بودن طول یال است. با مقایسه  $S_p(n)$  و  $S_p^{(r)}(n)$  در جدول ۱، مشاهده می‌کنیم که تعداد ساختارهای محدود، به‌طور چشمگیری کاهش یافته است. حال با اعمال شرایطی که ساختارهای  $RNA$  واقعی از لحاظ شکل دارند، سعی داریم تعداد ساختارهای  $RNA$  واقعی را شمارش کنیم.

از سویی می‌دانیم که کارکرد اکثر دنباله‌های  $RNA$ ، به ساختارهای آن‌ها بستگی دارد. بنابراین، درک ساختارهای  $RNA$  به منظور مطالعه تابع  $RNA$  زیستی و همچنین در فرایند طراحی ساختارهای  $RNA$  مصنوعی، اهمیت به‌سزایی دارد.

یک مساله اصلی در زیست‌شناسی، پیش‌بینی شکل یک مولکول است. که در این مقاله، شکل  $RNA$  و ترکیبیات مربوط به آن را مورد بررسی قرار دادیم. به دلیل پیش‌بینی بسیار مشکل ساختارهای سوم، تلاش‌ها برای پیش‌بینی ساختار مولکول  $RNA$  بیشتر روی ساختار دوم مولکول متمرکز است. بنابراین، باید اطلاعات مناسبی در مورد ساختار دوم داشته باشیم. اگر یک ساختار اول  $RNA$  را داشته باشیم، طبق الگوریتم‌های پیش‌بینی، ساختار دوم آن را می‌توان به دست آورد. این الگوریتم‌ها به روش‌های متفاوتی این کار را انجام می‌دهند، ولی سعی می‌کنند که ساختار دوم تولید شده، از لحاظ شکل و سطح انرژی به ساختارهای واقعی نزدیک‌تر باشد. ما به دنبال روش کلی‌تری برای انجام این کار هستیم. برای این منظور، ساختار دوم را به صورت یک گراف رسم می‌کنیم. این کار دو مزیت دارد:

- ۱- هر ویژگی که ساختار دوم از لحاظ شکل دارد، متناظر با یک اصطلاح در نظریه گراف خواهد بود.
- ۲- بررسی ترکیبیاتی گراف‌ها ساده‌تر از بررسی ترکیبیاتی ساختارهای دوم است.

## منابع مورد استفاده

1. Chen, W. Y. C., Deng, E. Y. P., Du, R. R. X., Stanley, R. P., Yan, C. H., 2007. Crossings and nestings of matchings and partitions. *Trans Amer Math Soc* 359:1523-1530.
2. Grabiner, D. J., Magyar, P., 1993. Random walks in Weyl chambers and the de-composition of tensor powers. *J Alg Comb* 2: 239-260.
3. Huang, F. W. D., Peng, W. W. J., Reidys, C. M., 2008. Folding 3 – noncrossing RNA pseudoknot structures. *J. Comput. Biol.* (in press).
4. Jin, E. Y., Reidys, C. M., 2008. Combinatorics of RNA structures with pseudo-knots. *Bull Math Biol* 70:34-39.
5. Jin, E. Y., Reidys, C. M., 2009. RNA pseudoknot structures with arc-length  $\geq 3$  and stack-length  $\geq \sigma$ . *Discrete Appl Math* 159: 25-36.
6. Jin, E. Y., Reidys, C. M., Wang, R., 2008. Asymptotic analysis of  $k$ -noncrossing matchings. arXiv: 0803.0848.
7. Schensted, C. E., 1961. Longest increasing and decreasing subsequences. *Canad. J Math* 13:179-190.
8. Waterman, M. S., Schmitt, W. R., 1994. Linear trees and RNA secondary structure. *Discrete Appl Math* 51:317-323.
9. Waterman, M. S., 1978. Secondary structure of single-stranded nucleic acids. *Adv Math Suppl Stud* 1: 167-212.