



دانشگاه تهران

پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر
پروژه کارشناسی

شمارش ساختارهای دوم دوبخشی توالی‌های RNA

نگارنده

بنیامین دلشاد ممقانی

استاد راهنما

دکتر محمد گنج‌تابش

پایان‌نامه برای دریافت درجه کارشناسی
در رشته علوم کامپیوتر

بهمن ماه ۱۳۹۶

چکیده

RNA¹ها، درشت مولکول‌هایی ضروری هستند که همانند DNA²ها از زنجیره ای طولانی نوکلئوتیدها تشکیل شده‌اند. با واسطه‌گری این مولکول‌ها، اطلاعات موجود در DNAها به پروتئین تبدیل می‌شوند. این درشت مولکول‌ها ساختارهایی با قواعد خاص دارند که می‌توان آن‌ها را مدل‌سازی ریاضی کرد. از طرفی، عملکرد RNAها در بیشتر مواقع توسط ساختار آن‌ها مشخص می‌گردد. لذا شناخت ساختارهای مختلف RNAها اهمیت زیادی دارد. در این پروژه، با استفاده از مدل‌سازی ساختار دوبخشی و دوم دوبخشی توالی‌های RNA و با استفاده از ابزارهای ترکیبیاتی، سعی در شمارش دقیق یا تخمینی تعداد ساختارهای ممکن می‌کنیم که در مسائل دیگر علوم زیستی و بیوانفورماتیک کاربردهایی دارد.

¹RiboNucleic Acid

²DeoxyriboNucleic Acid

پیش‌گفتار

یکی از شاخه‌های مهم علوم زیستی، ژنتیک است که شاید بتوان گفت ۱۵۰ سال قدمت دارد. در سال ۱۸۶۶، مندل، پدر علم ژنتیک، که خود یک کشیش بود در باغچه‌ی خود بر روی گیاه نخودفرنگی کار دست به آزمایش‌هایی زد که منجر به کشف قوانین وراثت یا قوانین مندل شد. تحقیقات ژنتیکی در این ۱۵۰ سال، منجر به کشف وجود ژن، کروموزوم، DNA، RNA، ارتباط وراثتبا هریک از ژن‌ها، نقش این مواد در بدن موجودات زنده، ارتباط آن‌ها با فعالیت پروتئین‌ها در بدن و بسیاری مسائل دیگر شده است. با کشف این پدیده‌ها مجدداً مسایل جدیدی مطرح شدند که ساختار سه‌بعدی پروتئین‌ها یکی از مثال‌های آن است. زیست‌شناسان، برای حل برخی از این مسائل به دلیل حجم بالای داده‌های مسئله و زمان‌بر بودن کار آزمایشگاهی، به ریاضیات، آمار و علوم کامپیوتر روی آوردند. این‌گونه شاخه‌ی بیوانفورماتیک مطرح گشت. این‌گونه مسائلی در این حوزه مطرح شد و باعث پیشرفت علم زیست‌شناسی گردید. در این پایان‌نامه، به بررسی یکی از همین مسائل می‌پردازیم.

فهرست مطالب

۱	مفاهیم اولیه	۱
۱	نظریه گراف ها	۱.۱
۲	تعاریف ریاضی گراف	۱.۱.۱
۳	مقدمات زیستی	۲.۱
۳	سلول	۱.۲.۱
۶	همانندسازی DNA	۲.۲.۱
۷	چگونگی به وجود آمدن پروتئین ها	۳.۲.۱
۷	مولکول RNA	۴.۲.۱
۸	ساختار RNA	۵.۲.۱
۸	ساختار اول RNA	۶.۲.۱
۹	ساختار دوم RNA	۷.۲.۱
۱۰	ساختار سوم RNA	۸.۲.۱
۱۰	نحوه‌ی نمایش ساختارهای مختلف RNA	۹.۲.۱
۱۱	مولفه‌های ساختار دوم RNA	۱۰.۲.۱
۱۴	جمع‌بندی	۳.۱
۱۶	مسئله‌ی مورد بحث	۲
۱۶	جزئیات مسئله	۱.۲
۱۸	مدل‌سازی‌ها	۲.۲
۱۸	مدل درختی	۱.۲.۲
۱۸	مدل درختی کاهش یافته	۲.۲.۲
۱۸	مدل رشته‌ای	۳.۲.۲
۱۹	مدل رشته‌ای برای ساختار دوم دوبخشی	۴.۲.۲
۲۰	بررسی روش‌های حل مسئله	۳
۲۰	بازگشتی پایه	۱.۳

۲۱	مسائل ترکیبیات شمارشی متناظر	۲.۳
۲۱	مسئله‌ی میز دایره‌ای	۱.۲.۳
۲۱	مسئله‌ی پیمایش مسیر در صفحه‌ی دکارتی	۲.۲.۳
۲۱	اعداد موتزکین	۳.۲.۳
۲۴	تلاش‌هایی برای ساختار دوم دوبخشی	۳.۳

فصل ۱

مفاهیم اولیه

در این فصل، مفاهیم اولیه‌ی مورد نیاز برای این پایان‌نامه ارائه خواهند شد. در بخش نخست، گراف‌ها و تعاریف اولیه‌ی آن معرفی می‌شوند. در بخش دوم نیز به مفاهیم زیستی مورد نیاز پرداخته می‌شود.

۱.۱ نظریه گراف‌ها

واژه‌ی گراف در ریاضیات حداقل دارای دو معنی می‌باشد. در ریاضیات ابتدایی گراف به نمودار تابع اشاره دارد و در اصطلاح ریاضی‌دانان گراف مجموعه‌ای از نقاط و خطوط متصل به هم هستند. در واقع گراف مدلی ریاضی برای یک مجموعه‌ی گسسته است که اعضای آن به طریقی به هم مرتبط هستند. اعضای این مجموعه می‌توانند انسان باشند و ارتباط آن‌ها با هم دست دادن باشد. اعضا می‌توانند اتم‌ها در یک مولکول باشند و ارتباط آن‌ها اتصال‌های شیمیایی باشد یا اعضا می‌توانند قسمت‌های مختلف زمین و ارتباط بین آن‌ها پل‌هایی باشد که آن‌ها را به هم مرتبط می‌کند (همانند مسئله کونیگسبرگ). نظریه گراف یکی از موضوع‌های مهم در ریاضیات گسسته است که به مطالعه‌ی گراف‌ها و مدل‌بندی مسائل به وسیله‌ی آن‌ها می‌پردازد. لئونارد اویلر در سال ۱۷۳۶ با حل مسئله‌ی پل‌های کونیگسبرگ نظریه گراف‌ها را بنیان گذاشت. اما جیمز جوزف سیلوستر نخستین کسی بود که در سال ۱۸۷۸ از واژه‌ی گراف برای نامیدن این مدل‌های ریاضی

استفاده کرد.

۱.۱.۱ تعاریف ریاضی گراف

$G_{V,E}$ را یک گراف گوییم هرگاه V مجموعه‌ی رأس‌ها و $E \subseteq V \times V$ مجموعه‌ی یال‌ها باشد.

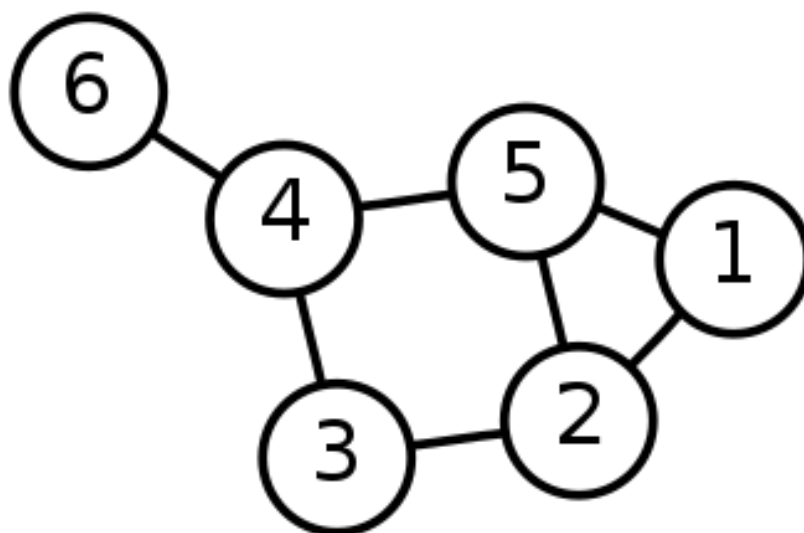
در نمایش گراف از نقطه و خم‌ها بهره می‌جوییم و هرگاه رأس‌ها متمایز باشند یا اصطلاحاً شماره‌گذاری شده باشند، به گراف حاصل، گراف برجسب‌دار می‌گوییم. یک مثال از این گراف‌ها در شکل ۱.۱ نشان داده شده است.

دو رأس مجاورند هرگاه با یک یال به یکدیگر وصل شده باشند. هرگاه در یک گراف، بتوان از یک رأس شروع به گذر کرد و بدون عبور از رأس و یال تکراری به همان رأس اولیه رسید، می‌گوییم گراف دور دارد. اگر بتوان با شروع از هر راسی در گراف و گذر از رأس‌ها و یال‌ها، به هر رأس دیگری در گراف رسید، آن گراف را همبند گوییم و گرنه ناهمبند است. به هر گرافی که همبند و بدون دور باشد، درخت گوییم.

اگر درختی را از یک رأس مشخص به اصطلاح آویزان کنیم و در بالا قرار دهیم، آن رأس مشخص را ریشه گوییم و به ازای هر رأس، رئوس مجاوری که در پایین آن رأس قرار دارند، بچه‌های آن رأس محسوب می‌شوند. به چنین درختی نیز، درخت ریشه‌دار می‌گوییم.

نیم‌راس نوع خاصی از رأس می‌باشد که میدانیم با رأس تفاوت دارد، و گرنه در عمل و کاربرد تفاوت خاصی ندارد و به جای نقطه، آن را با مربع توپر نشان می‌دهیم.

نمایش ماتریس مجاورت از یک گراف بدین صورت است که به ازای یک گراف n راسی، یک ماتریس $n \times n$ در نظر می‌گیریم که مولفه‌ی (i, j) آن ۱ است اگر بین رئوس i و j یال باشد و گرنه ۰ است.



شکل ۱.۱: یک مثال از گراف برجسب‌دار

۲.۱ مقدمات زیستی

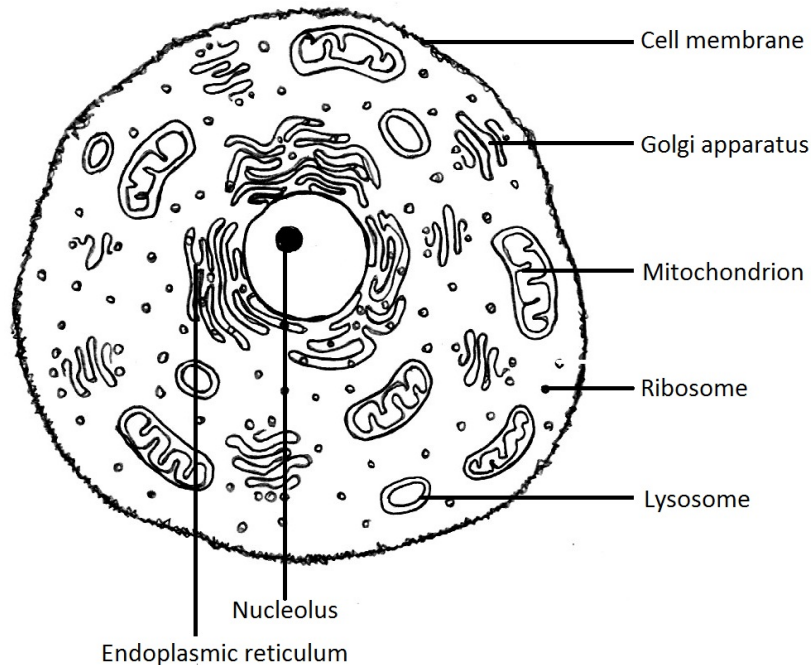
۱.۲.۱ سلول

هر موجود زنده از ساختارهای کوچکی به نام سلول تشکیل شده است. این سلول‌ها، اغلب بسیار کوچکند و با چشم غیرمسلح دیده نمی‌شوند. سلول‌ها به دو دسته تقسیم می‌شوند:

- رده‌ی اول، سلول‌های یوکاریوت می‌باشند که بسیار سازمان‌یافته و پیچیده‌اند؛ مانند سلول‌های انسانی، حیوانی، گیاهی و قارچی.

- رده‌ی دوم، سلول‌های پروکاریوت می‌باشند که ساختمان ساده‌تری دارند و در واقع ارگانیسم‌های تک‌سلولی هستند. این نوع سلول‌ها، هزاران بار کوچک‌تر از سلول‌های یوکاریوتی هستند.

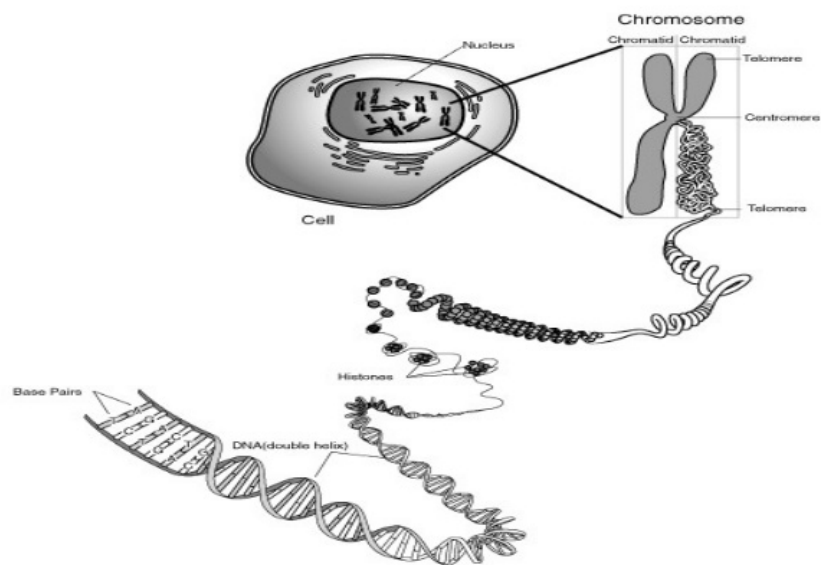
سلول‌های یوکاریوت دارای سه بخش اصلی‌اند: غشای سلولی، سیتوپلاسم و هسته. غشای سلولی، پرده‌ای نرم است که پیرامون سلول را می‌پوشاند و از آن در مقابل محیط خارجی محافظت می‌کند. فضای درون سلول را ماده‌ای به نام سیتوپلاسم پر می‌کند که ترکیب اصلی آن، آب و پروتئین‌های



شکل ۲.۱: ساختار کلی سلول.

محلول در آب هستند. در مرکز سلول، هسته قرار دارد و توسط پوسته‌ای از بقیه‌ی سلول جدا می‌شود. در داخل هسته، کروموزم‌ها وجود دارند که تمام اطلاعات ژنتیکی موجودات زنده، روی آن‌ها ذخیره می‌شود. این سه بخش در شکل ۲.۱ قابل مشاهده هستند.

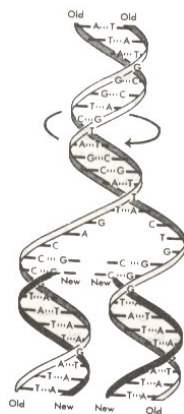
هر کروموزوم از دو نوع درشت‌مولکول به نام‌های اسید نوکلئیک و پروتئین تشکیل شده است. ریزمولکول‌ها به عنوان اجزای تشکیل دهنده‌ی درشت‌مولکول‌ها عمل می‌کنند و یا این که خودشان به تنهایی در سلول، نقش مستقل دارند. از میان ریزمولکول‌ها می‌توان به اسیدهای چرب، قندها، اسیدهای آمینه و نوکلئوتیدها اشاره کرد. در شکل ۳.۱ رابطه‌ی بین کروموزوم و سلول نمایش داده شده است.



شکل ۳.۱: چگونگی جای‌گیری کروموزوم، DNA و ژن در سلول.

۲.۲.۱ همانندسازی DNA

ژنوم یا همان DNA در حقیقت ماده‌ی وراثتی می‌باشد و بنابراین برای این که بتواند خاصیت وراثت داشته باشد باید راهی برای ازدیاد آن وجود داشته باشد تا DNA ی یک سلول مادر که سلول اصلی است به دو سلول فرزند که از تقسیم سلول مادر به وجود می‌آیند، منتقل شود. این ازدیاد باید بدون تغییر کمی یا کیفی در ماده‌ی وراثتی انجام گیرد. مدل واتسون و کریک یک مکانیزم همانندسازی پیشنهاد می‌کند. در این مدل، هر رشته‌ی موجود در مولکول DNA ی دو رشته‌ای می‌تواند برای ساخت رشته‌ی مکمل خود، الگو باشد. در واقع در زمان همانندسازی DNA، ضمن این که دو رشته‌ی DNA ی مادر به تدریج از یکدیگر جدا می‌شوند، هر رشته نقش الگو را ایفا می‌کند و رشته‌ی جدیدی از روی آن ساخته می‌شود. این به صورتی است که نوکلئوتیدهای A، C، G و T در رشته‌ی جدید، مقابل نوکلئوتیدهای مکمل در رشته‌ی الگو قرار می‌گیرند و دو رشته‌ی جدید با پیوندهای فسفودی‌استر به یکدیگر متصل می‌شوند. جهت ساخت رشته‌ی نوساز، از ۵' به ۳' است به این مفهوم که انتهای ۵' قبل از انتهای ۳' ساخته می‌شود و انتهای ۵' نوکلئوتیدی که وارد می‌شود به انتهای ۳' آزاد رشته‌ی نوساز اضافه می‌شود. پیوندهای هیدروژنی، بین هر رشته‌ی نوساز و رشته‌ی الگو وجود دارند و لذا از یک مولکول DNA ی دورشته‌ای، دو مولکول DNA ی دو رشته‌ای حاصل می‌شود. از آنجایی که یک رشته در DNA ی دورشته‌ای فرزند، از قبل ساخته شده و در واقع، یک رشته‌ی آن، مولکول DNA ی مادر است، به این روند همانندسازی، نیمه حفاظتی می‌گویند. شکل ۴.۱ نمایی از این همانند سازی را نشان می‌دهند. اگرچه مکانیسم عمومی همانندسازی DNA به نظر ساده می‌رسد، ولی فرآیند واقعی، از نظر بیوشیمی، بسیار پیچیده است. مقداری از پیچیدگی همانندسازی DNA مربوط به پیچش DNA است که ابتدا باید باز شود و سپس، DNA همانندسازی کند و همچنین دو عمل کپی از روی رشته‌ی DNA باید تقریباً به صورت همزمان انجام شوند. مشکل دیگر، سرعت فرآیند همانندسازی است که حدود ۵۰ نوکلئوتید در ثانیه در پستانداران و ۵۰۰ نوکلئوتید در ثانیه در باکتری‌ها می‌باشد.



شکل ۴.۱: نحوه‌ی همانندسازی DNA

۳.۲.۱ چگونگی به وجود آمدن پروتئین‌ها

پروتئین‌ها در حقیق با استفاده از DNA، در یک فرآیند دو مرحله‌ای به وجود می‌آیند. در اولین مرحله که رونویسی نام دارد، نوع دیگری از نوکلئیک اسید به نام ریبونوکلئیک اسید (یا به اختصار RNA ساخته می‌شود. RNA، تک‌رشته‌ای است و با استفاده از یکی از رشته‌های DNA به عنوان الگو و رعایت روابط مکملی بین نوکلئوتیدها ساخته می‌شود. ساختار شیمیایی RNA شبیه به DNA است با این تفاوت که قند در نوکلئوتیدهای آن، ریبوز است و به جای باز آلی تیمین، از باز آلی یوراسیل در آن استفاده می‌شود. یوراسیل به اختصار با حرف U نشان داده می‌شود.

۴.۲.۱ مولکول RNA

ساختارهای متفاوتی از RNA در سلول وجود دارند که این ساختارها به توالی نوکلئوتیدهای RNA بستگی دارند. RNA های با ساختارهای مختلف، فعالیت های متفاوتی در سلول انجام می‌دهند. لذا فعالیت RNA های مختلف به توالی نوکلئوتیدهای آن‌ها بستگی دارد. این وابستگی درجه‌ی اهمیت تشخیص ساختار RNA را بالا می‌برد. از طرف دیگر، یکی از روش‌های گروه‌بندی RNA ها بر اساس ساختار آنان می‌باشد که این خود یکی دیگر از دلایل اهمیت تشخیص ساختار RNA است.

۵.۲.۱ ساختار RNA

در سلول، RNA به صورت یک رشته است که انعطاف پذیری بالایی دارد و می تواند به راحتی تا بخورد و پیوندهای هیدروژنی درون مولکولی ایجاد کند. همان طور که می دانیم، آدنین، مکمل یوراسیل است که دو پیوند هیدروژنی بین آن ها ایجاد می شود و سیتوزین مکمل گوانین است که سه پیوند هیدروژنی بین آن ها ایجاد می شود. این دو نوع پیوند را، پیوندهای واتسون- کریک یا پیوندهای کانونی نیز می نامند. پیوند دیگری که می تواند ایجاد شود، پیوند بین گوانین و یوراسیل است که ضعیف تر از دو پیوند دیگر است و جفت وابل نامیده می شود. برای RNA، سه نوع ساختار تعریف می شود که به ترتیب، ساختار اول، ساختار دوم و ساختار سوم نام دارند. هر کدام از این ساختارها بخشی از اطلاعات شکل فضایی RNA را در بر دارند. نشان داده شده است که روند شکل گیری یک مولکول RNA، یک روند یک مرحله ای می باشد. یعنی ابتدا ساختار اول RNA، سپس ساختار دوم و در انتها ساختار سوم آن شکل می گیرد.

۶.۲.۱ ساختار اول RNA

ساختار نوع اول که بیشتر برای نمایش رشته ی RNA به کار می رود، یک توالی از نوکلئوتیدهاست. این توالی به صورت یک رشته از چهار حرف A، C، G و U به که ترتیب حروف اول آدنین، سیتوزین، گوانین و یوراسیل می باشند است. بنا به قرارداد، رشته ی RNA، با ۵' در سمت چپ رشته شروع شده و به ۳' در سمت راست رشته ختم می شود. به عبارت دیگر می توان گفت یک مجموعه ی الفبای $\Sigma = A, C, G, U$ داده شده است و با توجه به آن، رشته ی RNA به صورت زیر نمایش داده می شود:

$$R = b_1 b_2 b_3 \dots b_n \quad b_i \in \Sigma \quad (1.1)$$

این ساختار فقط مولفه های سازنده ی مولکول را نشان می دهد و معمولاً برای یافتن الگوهایی خاص در رشته، مورد استفاده قرار می گیرد. نمونه ای از این الگویابی، همترازی رشته است که برای پیدا کردن الگوهای تکراری در رشته های چند RNA ای مختلف ولی مشابه مورد استفاده قرار می گیرد.

۷.۲.۱ ساختار دوم RNA

وقتی یک رشته‌ی RNA به روی خود تا می‌خورد، پیوندهای هیدروژنی، بین نوکلئوتیدهای آن رشته‌ی RNA برقرار می‌شوند. هر نوکلئوتید می‌تواند فقط با یک نوکلئوتید دیگر جفت شود. مستحکم‌ترین پیوندهایی که بین نوکلئوتیدها تشکیل می‌شوند عبارتند از GU، AU و CG و همچنین مکمل آن‌ها یعنی UA، UG و GC. استحکام بالای این پیوندها باعث شده که به طور معمول فقط این نوع از پیوندها در یک مولکول RNA مشاهده شود. برای نمایش پیوند بین نوکلئوتید i ام و نوکلئوتید j ام یک رشته‌ی RNA می‌توان از نمایش (i, j) استفاده کرد، به طوری که $1 \leq i < j \leq n$. فرض کنید برای یک رشته‌ی RNA، مجموعه‌ی تمام پیوندهای ممکن آن، A نامیده شود. ساختار دوم یک رشته‌ی RNA، یک زیرمجموعه مانند S از مجموعه‌ی A است اگر شروط زیر برقرار باشند:

- به ازای هر دو پیوند (i_1, j_1) و (i_2, j_2) در S و با توجه به این که در مجموعه‌ها عضو تکراری وجود ندارد، این رابطه‌ها صادق باشند: $i_1 \neq i_2$ و $i_1 \neq j_2$ و $i_2 \neq j_1$ و $i_2 \neq j_2$. در واقع، برقراری این شرط یعنی هیچ کدام از چهار نوکلئوتید مربوط به این دو پیوند در بیشتر از یک پیوند شرکت نمی‌کنند.

- اگر $h < i < j < k$ ، مجموعه‌ی S نمی‌تواند شامل هر دو پیوند (h, j) و (i, k) باشد. این شرط، یک شرط لازم برای ساختارهای دوم نیست و همان‌طور که در ادامه بیان خواهد شد، ساختارهایی نیز وجود دارند که این شرط برای آن‌ها برقرار نیست. در واقع در ادامه در حالتی که این شرط برقرار باشد، به آن ساختار دوم گوییم و اگر نوکلئوتیدها را بتوان طوری به دودسته افزار کرد که هر دسته، این شرط را داشته باشد و خود ساختار اصلی این شرط را نداشته باشد، به آن ساختار دوم دوبخشی گوییم. این پایان‌نامه با ساختارهای دوم و دوم دوبخشی سروکار دارد.

- اگر S ، پیوندی به صورت (i, j) داشته باشد، باید رابطه‌ی $|i - j| > 3$ برقرار باشد. البته برای مدل‌سازی‌ای که در آینده می‌کنیم، به جای ۳ از پارامتر m استفاده می‌کنیم و به بررسی مسئله‌مان بر پایه‌ی m می‌پردازیم.

با توجه به قوانینی که برای ایجاد پیوند بین نوکلئوتیدها ارائه شد، می‌توان

تمام حالات بالقوه برای S را لیست کرد؛ در واقع می‌توان گفت که مسئله‌ی اصلی پیش‌بینی ساختار دوم RNA این است که بتوان تشخیص داد کدام پیوند در ساختار دوم شرکت می‌کند و کدام پیوند شرکت نمی‌کند. از دیگر ویژگی‌های ساختار دوم، تمایل پیوندهای آن برای تشکیل بخش‌هایی است که آن‌ها را مولفه‌های ساختاری می‌نامند. این مولفه‌ها عبارتند از حلقه‌های داخلی، هرپین‌ها، چندحلقه‌ها یا حلقه‌های چندشاخه، بالچه‌ها و حلقه‌های خارجی. پیوندهای متوالی، استحکام پیوندها را افزایش می‌دهند و بنا به اهمیتی که دارند، نام آن‌ها را استم یا هلیکس می‌نامند. پیوندهای متوالی زمانی رخ می‌دهند که نوکلئوتیدهای پیوندها همسایه باشند. در ادامه این مولفه‌ها بررسی خواهد شد.

۸.۲.۱ ساختار سوم RNA

ساختار سوم RNA در واقع ساختار سه‌بعدی آن می‌باشد که با مختصات اتمی مشخص می‌شود. عملکردهای مختلفی که RNA قادر به انجام آن‌ها هستند، به ساختار فضایی آن‌ها وابسته است. همانند ساختار دوم، ساختار سوم RNA نیز از مولفه‌هایی تشکیل شده است که به عنوان بلوک‌های سازنده‌ی مولکول در نظر گرفته می‌شوند. مشخص کردن ساختار سه‌بعدی مولکول RNA چه با استفاده از روش‌های آزمایشگاهی و چه با استفاده از روش‌های محاسباتی، کار بسیار مشکلی است و به همین دلیل، اغلب پژوهش‌ها بر روی ساختار دوم متمرکز شده‌اند.

۹.۲.۱ نحوه‌ی نمایش ساختارهای مختلف RNA

برای راحتی کار با انواع ساختارهای RNA و ارائه‌ی روش‌های محاسباتی مختلف برای مسائل مربوط به RNA، هریک از ساختارها را به روش‌های مختلفی نمایش می‌دهند. ساختار اول به صورت یک رشته از حروف A، C، G و U نمایش داده می‌شوند. ساختار دوم را به اشکال مختلفی می‌توان نمایش داد که از مهم‌ترین آن‌ها می‌توان به نمایش پرائتز-نقطه‌ای، فین من خطی، فین من دایره‌ای و گراف مسطح اشاره کرد. ساختار سوم را نیز می‌توان با کمی تغییر در روش‌های بالا و یا به صورت سه‌بعدی نمایش داد. برای هر

یک از این نمایش‌ها یک مثال در شکل ۵.۱ آورده شده است. در شکل ۵.۱ (آ) ساختار اول یک RNA دیده می‌شود. در شکل ۵.۱ (ب)، ساختار دوم به صورت پُرانتز-نقطه‌ای مشاهده می‌شود. در این نوع نمایش، از نقطه برای نمایش مکان‌هایی از ساختار که در هیچ پیوندی شرکت نکرده‌اند (بازهای موجود در این مکان‌ها، بازهای آزاد نامیده می‌شوند) استفاده می‌شود. پُرانتز باز و بسته نیز به ترتیب بیانگر بازهای اول و دوم یک پیوند هستند. در شکل ۵.۱ (ج)، ساختار دوم به فین من خطی، در شکل ۵.۱ (د)، ساختار دوم به روش فین من دایره‌ای و در شکل ۵.۱ (ه)، ساختار دوم با گراف مسطح آورده شده است. در شکل ۵.۱ (و) یک نمونه ساختار سوم نشان داده شد است.

۱۰.۲.۱ مولفه‌های ساختار دوم RNA

از آن‌جا که ساختار دوم RNA از اهمیت بالایی برخوردار است، دانشمندان، مولفه‌هایی که معمولاً در تمام ساختار دوم‌ها وجود دارند را نام‌گذاری کرده‌اند و به دلیل نقش این مولفه‌ها در پیش‌بینی ساختار دوم RNA، در این بخش به طور دقیق تعریف می‌شوند. این مولفه‌ها عبارتند از: استم، هیرپین، حلقه‌ی داخلی، بالچ، حلقه‌ی خارجی، حلقه‌ی چندشاخه و شبه‌گره.

استم یا هلیکس چندجفت نوکلئوتید را که با هم به صورت متوالی، پیوند ایجاد کرده‌اند، استم یا هلیکس می‌نامند. به طور دقیق‌تر، اگر برای نمایش پیوند بین نوکلئوتید i ام و نوکلئوتید j ام یک رشته‌ی RNA از نمایش (i, j) استفاده شود، به توالی پیوندهای $(i, j), (i+1, j-1), \dots, (i+h, j_h)$ در ساختار دوم، استم گفته می‌شود. معمولاً استم‌های با طول بیشتر، از پایداری بیشتری برخوردار هستند. استم از مهم‌ترین مولفه‌های ساختار دوم RNA است. استم‌های شامل تنها دو جفت پیوندی، استک نامیده می‌شوند.

هرپین اگر تعدادی نوکلئوتید متوالی از یک RNA، در هیچ پیوندی شرکت نکنند و نوکلئوتیدهای اطراف آن‌ها با یکدیگر، پیوند تشکیل دهند به طوری که یک استم بسازند، به حلقه‌ی ایجادشده هیرپین گفته می‌شود. به طور دقیق‌تر، اگر پیوند (i, j) وجود داشته باشد و هیچ یک از نوکلئوتیدهای

UGAGCACGCUUGUGCAAUUAACGCUAAGUUAAGGGUCUACGGAUGA

(آ) نمایش ساختار اول

UGAGCACGCUUGUGCAAUUAACGCUAAGUUAAGGGUCUACGGAUGA

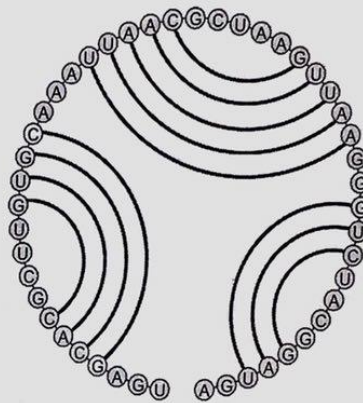
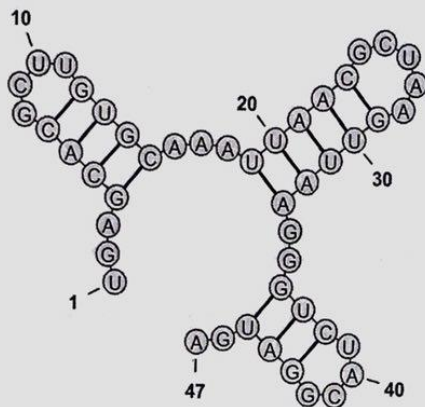
... (((((.....)))) ... (((((.....)))) ... (((.....))) ...

(ب) نمایش ساختار دوم به صورت پرانتز-نقطه‌ای



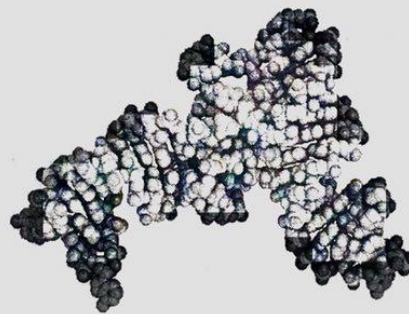
UGAGCACGCUUGUGCAAUUAACGCUAAGUUAAGGGUCUACGGAUGA

(ج) نمایش ساختار دوم به روش فین من خطی



(ه) نمایش ساختار دوم با گراف مسطح

(د) نمایش ساختار دوم به روش فین من دایره‌ای



(و) نمایش ساختار سوم به صورت سه‌بعدی

شکل ۵.۱: نحوه‌ی نمایش ساختارهای مختلف RNA

$i+1$ تا $j-1$ در هیچ پیوندی شرکت نکنند، به ناحیه‌ی به دست آمده هیرپین گفته می‌شود. میزان عدم پایداری یک هیرپین، بیشتر به طول قسمتی از آنکه فاقد پیوند است بستگی دارد. طول یک هیرپین برابر است با تعداد نوکلئوتیدهای آزاد موجود در داخل حلقه.

حلقه‌ی داخلی یا بالج وقتی دو استم متوالی شکل می‌گیرند، قسمت‌هایی از رشته‌ی RNA که در بین این دو استم قرار دارند، با هم جفت نمی‌شوند. به این ناحیه که در واقع دو زیررشته از دو قسمت مختلف رشته‌ی RNA هستند و با یکدیگر پیوند ایجاد نکرده‌اند، حلقه‌ی داخلی گفته می‌شود. به طور دقیق‌تر، اگر $i+1 < p < q < j-1$ و یک ساختار دوم، پیوندهای (i, j) و (p, q) را داشته باشد ولی نوکلئوتیدهای بین i و p و همچنین طور نوکلئوتیدهای بین j و q در هیچ پیوندی شرکت نکرده باشند، به این دو ناحیه که در پیوند شرکت نکرده‌اند حلقه‌ی داخلی گفته می‌شود. میزان عدم پایداری حلقه‌ی داخلی به طول رشته‌های تشکیل‌دهنده‌ی آن (که در پیوند شرکت نکرده‌اند) بستگی دارد. وقتی یکی از دوناحیه‌ی یک حلقه‌ی داخلی که در پیوند شرکت نکرده، تهی باشد ناحیه‌ی ایجاد شده را بالج می‌نامند. پس بالج، حالت خاصی از حلقه‌ی داخلی است. طول حلقه‌ی داخلی و بالج، همانند طول هیرپین تعریف می‌شود. حلقه‌های داخلی که طول ناحیه‌ی اول آن‌ها برابر با m و طول ناحیه‌ی دوم آن‌ها برابر با n باشد را حلقه‌های داخلی $m \times n$ نیز می‌نامند.

حلقه‌ی چندشاخه به ناحیه‌هایی از رشته‌ی RNA که یک استم به آن‌ها وارد شده و حداقل ۲ استم دیگر از آن‌ها خارج می‌شوند حلقه‌ی چندشاخه گفته می‌شود. برای این که بتوان این مولفه را توضیح داد، تعریف می‌شود پیوند (i, j) ، پیوند (r, s) را احاطه می‌کند اگر $i < r < s < j$. حال اگر (i, j) ، دو یا چند پیوند دیگر مانند $(p, q), \dots$ را احاطه کند، ولی این پیوندها یکدیگر را احاطه نکنند، گفته می‌شود که یک حلقه‌ی چندشاخه شکل گرفته است.

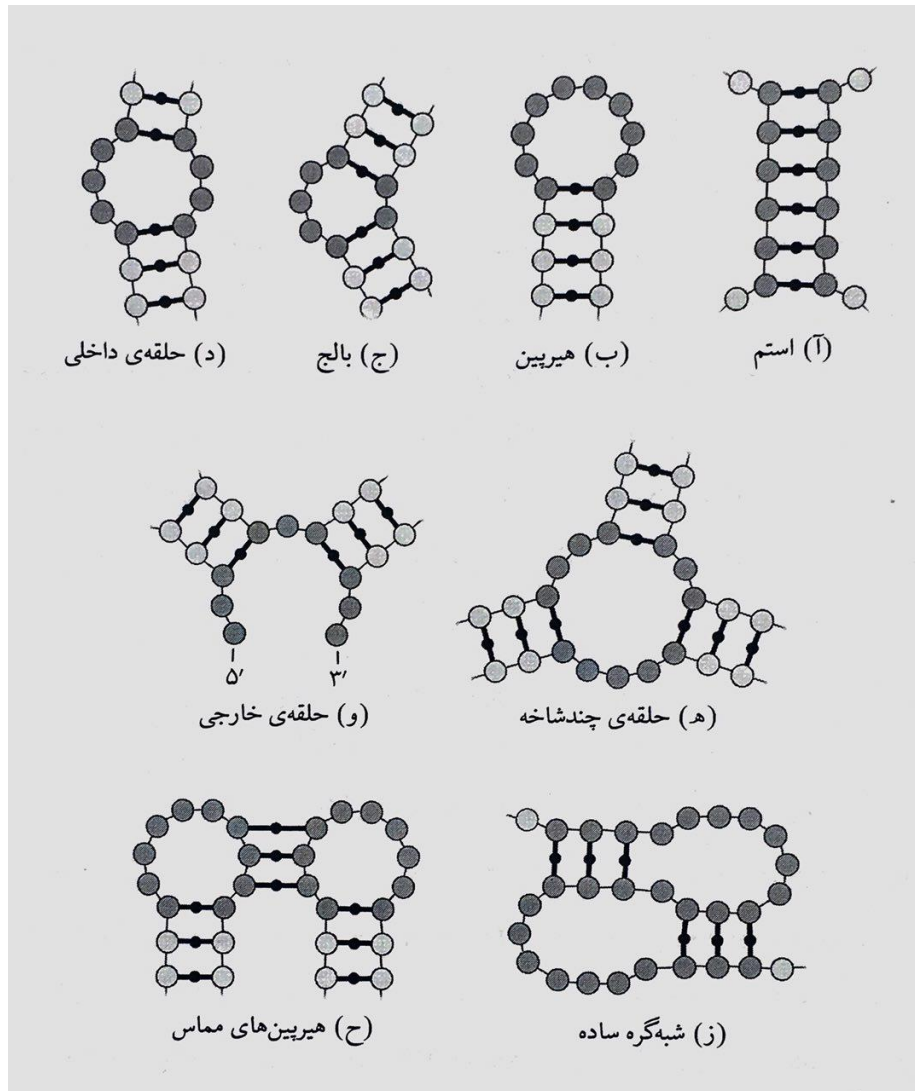
حلقه‌ی خارجی به مجموعه‌ی بیرونی‌ترین بازهای آزاد و پیوندها، حلقه‌ی خارجی گفته می‌شود. برای تعریف دقیق‌تر، گفته می‌شود پیوند (r, s) از پیوند

(i, j) دسترس پذیر است اگر پیوند (i, j) پیوند (r, s) را احاطه کند ولی هیچ پیوند دیگری مانند (p, q) وجود نداشته باشد به طوری که $i < p < r < s < j$ می توان تعریف مشابهی را برای دسترس پذیری یک باز آزاد ارائه داد. مجموعه‌ی تمامی بازهای آزاد و پیوندهایی که از پیوند (i, j) دسترس پذیر هستند را حلقه‌ی بسته شده توسط پیوند (i, j) می نامند. توجه داشته باشید که خود این پیوند، جزئی از این حلقه در نظر گرفته نمی شود. به مجموعه‌ی تمامی بازها و پیوندهایی که از هیچ پیوندی، دسترس پذیر نیستند، حلقه‌ی خارجی گفته می شود. هر ساختار دوم، تنها یک حلقه‌ی خارجی دارد و آن حلقه، حتما شامل دو انتهای ۵' و ۳' خواهد بود.

شبه‌گره شبه‌گره‌ها که فراوانی آن‌ها نسبت به مولفه‌های قبلی کمتر است، مولفه‌هایی هستند که شرط دومی که برای ساختار RNA بیان شد، در آن‌ها نقض شده است. به عبارت دیگر، شبه‌گره، زیرساختاری مانند $i, \dots, i+m$ از یک ساختار دوم است به طوری که r و s ای وجود دارند که $i < r < s < i+m$ و پیوندهای (i, s) و $(r, i+m)$ در آن ساختار وجود دارند. این مولفه‌ها معمولا شکل‌های خاصی به خود می‌گیرند که از جمله‌ی آن‌ها می‌توان به شبه‌گره ساده و هیرپین‌های مماس اشاره کرد. در شکل ۶.۱، دو نمونه از شبه‌گره‌ها و انواع دیگر مولفه‌های ساختار دوم RNA نمایش داده شده‌اند.

۳.۱ جمع‌بندی

در این فصل، مفاهیم اولیه‌ی مورد نیاز برای این پایان‌نامه ارائه شدند. ابتدا مفاهیم پایه و تعاریف اولیه نظریه گراف‌ها مطرح شدند که برای مدل‌سازی‌ها از آن‌ها بهره می‌جوئیم. در انتها مبانی زیست‌شناسی مولکولی و به ویژه مولکول RNA و نقش مهم آن در داخل سلول به طور مفصل مورد بررسی قرار گرفتند.



شکل ۶.۱: مولفه‌های مختلف ساختار دوم RNA. در هر شکل، بازهای تیره‌تر، بیانگر مولفه‌ی متناظر هستند.

فصل ۲

مسئله‌ی مورد بحث

در این فصل، به بیان شفاف صورت مسئله پرداخته می‌شود. در واقع ما با یک مسئله‌ی شمارشی طرف هستیم که در این فصل، با استفاده از مدل‌سازی‌ها آن را تبدیل به یک مسئله‌ی کاملاً ترکیبیاتی می‌کنیم تا از ابزارهای ترکیبیات شمارشی برای حل آن، بتوان بهره جست. می‌خواهیم سعی کنیم تعداد ساختارهای ممکن برای ساختار دوم و ساختار دوم دویخشی RNA را بشماریم.

۱.۲ جزئیات مسئله

تعریف ۱.۱.۲. یک ساختار دوم، یک گراف برچسب‌دار n راسی با ماتریس مجاورت A است به طوری که:

$$(i) \quad a_{i,i+1} = 1 \text{ برای هر } 1 \leq i \leq n-1,$$

(ii) برای هر i حداکثر یک $k \neq i-1, i+1$ وجود داشته باشد به طوری که $a_{ik} = 1$

(iii) اگر $a_{ij} = a_{kl} = 1$ و $i < j < k$ و $i < l < j$.

ما یال (i, k) را ($|i - k| \neq 1$) یک جفت پایه می‌نامیم. هر راس i را که فقط به رئوس $i-1$ و $i+1$ وصل باشد، جفت نشده می‌نامیم. راس i ، یک راس داخلی برای جفت پایه‌ی (k, l) محسوب می‌گردد اگر $k < i < l$. به علاوه، اگر هیچ

جفت پایه‌ی دیگری مثل (p, q) نباشد به طوری که $k < p < i < q$ ، گوئیم راس i اکیدا داخلی است برای جفت پایه‌ی (k, l) .

تعریف ۲.۱.۲. ساختار دوم از عناصر زیر تشکیل شده است:

(i) پشته شامل جفت پایه‌های متوالی $(p-k, q+k), (p-k+1, q+k-1), \dots, (p, q)$ است به گونه‌ای که هیچ یک از $(p-k-1, q+k+1)$ و $(p+1, q-1)$ جفت پایه نباشند. طول پشته در این حالت $k+1$ است. جفت پایه‌ی پایانی پشته است.

(ii) حلقه شامل همه‌ی رئوس جفت نشده‌ای است که اکیدا داخل یک جفت پایه‌ی (p, q) باشند.

(iii) راس خارجی یک راس جفت نشده است که به هیچ حلقه‌ای تعلق ندارد. مجموعه‌ی همه‌ی رئوس خارجی که مجاور هم هستند، یک عنصر خارجی نامیده می‌شوند. یک عنصر خارجی، اگر شامل راس اول یا آخر شود، پایان باز است وگرنه مفصل است.

لم ۱.۱.۲. هر ساختار دوم S به صورت یکتا، به پشته‌ها، حلقه‌ها و عناصر خارجی تقسیم می‌شود.

برهان. هر راس جفته شده‌ای، متعلق به یک جفت پایه‌ی مشخص است که آن جفت پایه نیز متعلق به یک پشته‌ی مشخص و یکتا است. هر راس جفت نشده‌ای هم یا خارجی است یا اکیدا داخل یک جفت پایه‌ی مشخص است که به ترتیب متعلق به یک عنصر خارجی یا حلقه‌ی مشخص و یکتا است. لذا این تقسیم بندی به صورت یکتا وجود دارد.

تعریف ۳.۱.۲. پشته‌ی $[(p, q), \dots, (p+k, q-k)]$ پایانه‌ای نامیده می‌شود اگر $p-1=0$ یا $q+1=n+1$ یا دو راس $p-1$ و $q+1$ داخل هیچ جفت پایه‌ای نباشند. زیرساختاری که توسط جفت پایه‌ی پایانی (p, q) از پشته‌ی پایانه‌ای ساخته می‌شود، یک مولفه از S است.

لم ۲.۱.۲. هر ساختار دوم S به صورت یکتا، به مولفه‌ها و رئوس خارجی تقسیم می‌شود.

اثبات، بدیهی است.
طبق تعریف، یک ساختار بدون مولفه، یک ساختار باز است.

تعریف ۴.۱.۲. درجه‌ی یک حلقه، یک به علاوه‌ی تعداد جفت پایه‌های پایانی پشته‌هایی است که داخل نزدیک‌ترین جفت پایه‌ی حلقه است. حلقه‌ی درجه یک، هیرپین نامیده می‌شود. حلقه‌ی درجه‌ی دوم، بالج و درجات بالاتر، چندحلقه نامیده می‌شوند.

تعریف ۵.۱.۲. ساختار دوم دوبخشی، گراف برچسب داری است که ساختار دوم نیست ولی اگر مجموعه رئوس آن را به دو بخش افزار کنیم، زیرساختار حاصل از هر بخش ساختار دوم باشد. ضمن بین دو بخش یالی به منظور جفت شدن نباشد.

همان طور که دیده می‌شود، تعاریف طوری است که بیشترین تطابق با واقعیت داستان را داشته باشد.

۲.۲ مدل‌سازی‌ها

۱.۲.۲ مدل درختی

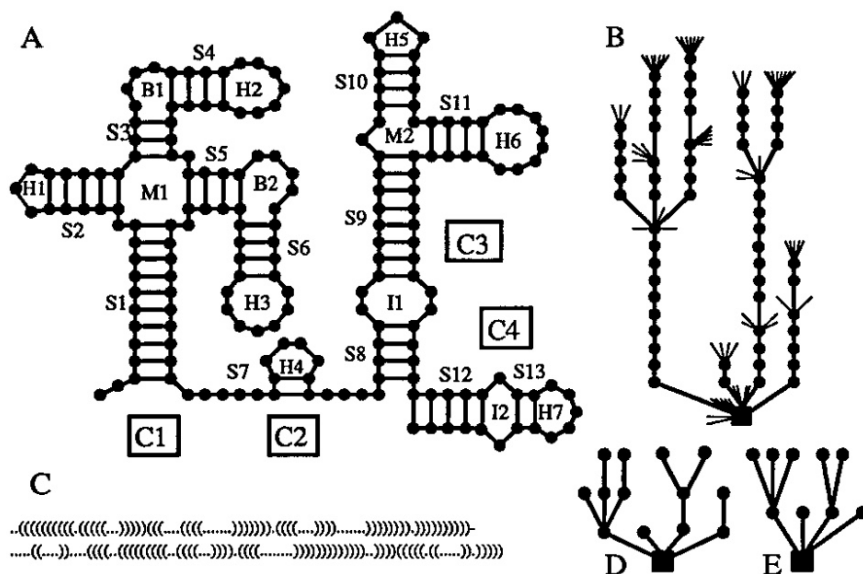
یک ساختار دوم S می‌تواند به صورت یک درخت ریشه‌دار نمایش داده شود که هر جفت پایه‌ی (p, q) با یک راس x نمایش داده می‌شود که y_1, \dots, y_k بچه‌های x هستند و هر یک نمایانگر یک جفت پایه است که اکیدا داخل (p, q) است. برای هر راس جفت نشده‌ی z نیز، یک نیم‌راس اضافه می‌کنیم به راسی که نمایانگر نزدیک‌ترین جفت پایه است به حلقه‌ای که شامل z است. رئوس خارجی و جفت پایه‌هایی که به جایی وصل نیستند، به ریشه وصل می‌گردند.

۲.۲.۲ مدل درختی کاهش یافته

همانند مدل درختی به دست می‌آید با این تفاوت که به جای اختصاص دادن راس به هر جفت پایه، به هر پشته یک راس اختصاص داده می‌شود و نیم‌راس‌ها حذف می‌گردند. این مدل قطعا همه‌ی اطلاعات RNA اصلی را ندارد ولی باز هم اطلاعات مهمی در بر دارد و مدلی بااهمیت است.

۳.۲.۲ مدل رشته‌ای

نمایش رشته‌ای ساختار دوم S می‌تواند بر اساس این قوانین به دست آید:



شکل ۱.۲: یک RNA دلخواه و مدل‌های درختی، درختی کاهش‌یافته و رشته‌ای برای آن

(i) اگر رشته i جفت نشده بود، $T_i = \cdot$.

(ii) اگر (p, q) یک جفت پایه باشد و $p < q$ ، آن‌گاه $T_p = "$ و $T_q = "$.

رشته‌ی T همان نمایش ساختار دوم S می‌باشد.

۴.۲.۲ مدل رشته‌ای برای ساختار دوم دوبخشی

این حالت مشابه مدل رشته‌ای است با این تفاوت که به جای بعضی پرانتزها می‌توان از براکت استفاده کرد. رشته‌ای برای ساختار دوم دوبخشی معتبر است که اولاً اگر براکت‌ها را نیز پرانتز کنیم، ساختار دوم نباشد و دوماً، زیررشته حاصل از براکت‌ها و زیررشته حاصل از پرانتزها ساختار دوم باشند.

فصل ۳

بررسی روش‌های حل مسئله

۱.۳ بازگشتی پایه

فرض کنید تعداد ساختار دوم‌های معتبر برای i راس، S_i باشد. ساختار دوم‌های $n+1$ راسی را می‌توان از روی ساختار دوم‌های n راسی ساخت. در یک حالت می‌توان یک راس جفت‌نشده به انتهای یک ساختار n تایی افزود. در حالت دیگر، راس جدید، با یکی از n تایی قبلی جفت باید شده باشد. فرض کنیم فاصله راس جدید از جفتش، k باشد که می‌دانیم $k \leq m$. در این حالت برای رئوس داخلی S_k حالت داریم و برای مابقی رئوس، S_{n-k-1} . پس داریم:

$$S_{n+1} = S_n + \sum_{k=m}^{n-1} S_k S_{n-k-1}, \quad n \leq m+1 \quad (1.3)$$

$$S_0 = S_1 = \dots = S_{m+1} = 1$$

رابطه بازگشتی بالا رابطه‌ی نسبتاً مناسبی است که با یک الگوریتم از نوع برنامه نویسی پویا^۱ با مرتبه‌ی زمانی $O(n^2)$ می‌توانیم مقادیر آن را بسازیم. در برخی روش‌های شمارشی با اعمال محدودیت‌هایی روی m یا روی تعداد

^۱Dynamic Programming

هیرپین‌ها و از این دست موارد، روابط دیگری به دست می‌آید که در بخش بعدی به مواردی از آن‌ها اشاره می‌شود.

۲.۳ مسائل ترکیبیات شمارشی متناظر

۱.۲.۳ مسئله‌ی میز دایره‌ای

در این مسئله، یک میز دایره‌ای داریم که n نفر دور آن نشسته‌اند. می‌خواهیم بعضی از آن‌ها را دو به دو به هم وصل کنیم طوری که پاره خط‌هایی که واصل جفت‌ها هستند، یک دیگر را قطع نکنند. به وضوح هر جواب برای این مسئله، متناظر یک ساختار معتبر برای یک ساختار دوم RNA است. برای ساختار دوم دوبخشی می‌توان مسئله را این طور فرض کرد که پاره‌خط‌ها طوری باشند که اگر به دو دسته افراز شوند، در هر دسته، پاره‌خط‌ها یکدیگر را قطع نکنند. مثال حالت‌های خاص این مسئله در شکل‌های ۱.۳ و ۲.۳ مشخص شده است.

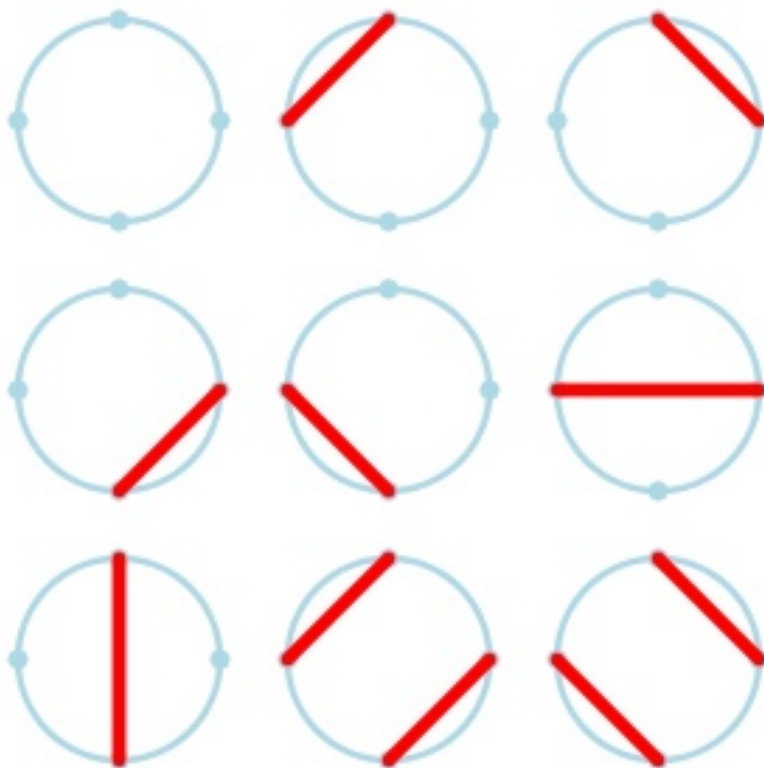
۲.۲.۳ مسئله‌ی پیمایش مسیر در صفحه‌ی دکارتی

در این مسئله، ابتدا در مختصات $(0, 0)$ قرار داریم و می‌خواهیم به مختصات $(n, 0)$ برویم. در هر مرحله، مجاز به یکی از سه حرکت $(1, 1)$ و $(1, 0)$ و $(1, -1)$ هستیم. تنها شرطی که داریم، این است که هیچگاه نباید به زیر محور X برویم. این مسئله به وضوح با تناظر حرکت‌ها به ترتیب $(, .)$ و $(,)$ به رشته‌های معتبر در مدل‌سازی رشته‌ای برای ساختار دوم تبدیل می‌شود و بالعکس. نمونه‌ای از حالت خاص مسئله در شکل ۳.۳ نشان داده شده است.

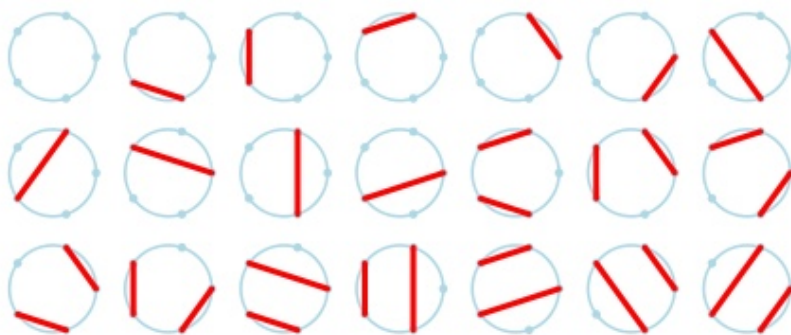
۳.۲.۳ اعداد موتزکین

این اعداد، همه مسائل گفته شده را در حالت $m = 0$ می‌شمارد. چند صورت دارد. در یکی، همان رابطه S است که m را در آن صفر قرار داده باشیم:

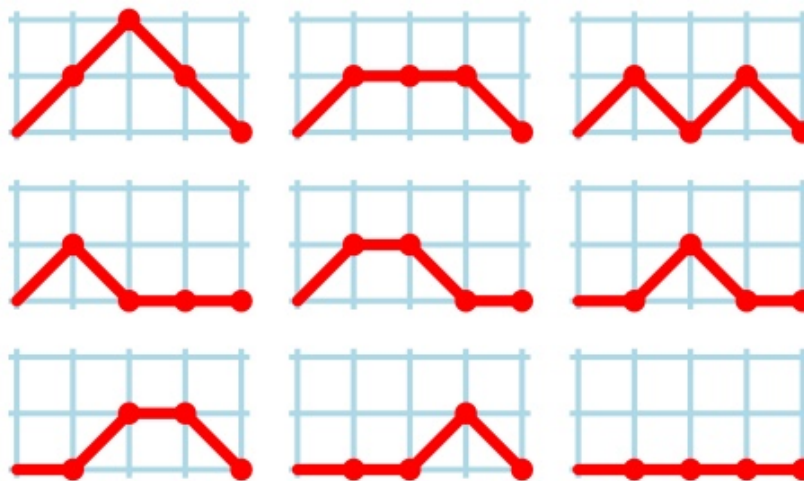
$$M_n = M_{n-1} + \sum_{i=0}^{n-2} M_i M_{n-2-i}, \quad n \leq 1 \quad (۲.۳)$$



شکل ۱.۳: حالات مختلف مسئله‌ی میز دایره‌ای برای $n = 4$



شکل ۲.۳: حالات مختلف مسئله‌ی میز دایره‌ای برای $n = 5$



شکل ۳.۳: حالات مختلف مسئله‌ی پیمایش مسیر در صفحه‌ی دکارتی برای $n = 4$

در دیگری، بر اساس دو جمله‌ی قبلی محاسبه شده است:

$$M_n = \frac{2n-1}{n+2}M_{n-1} + \frac{3n-3}{n+2}M_{n-2} \quad (۳.۳)$$

همچنین با استفاده از اعداد کاتالان^۲، می‌توان آن‌ها را توصیف کرد:

$$M_n = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2k} C_k \quad (۴.۳)$$

دنباله‌ی اعداد موتزکین به صورت زیر است:

۱، ۱، ۲، ۴، ۹، ۲۱، ۵۱، ۱۲۷، ۳۲۳، ۸۳۵، ۲۱۸۸، ۵۷۹۸، ۱۵۵۱۱،
 ...، ۴۱۸۳۵، ۱۱۳۶۳۴

^۲Catalan Numbers

۳.۳ تلاش‌هایی برای ساختار دوم دوبخشی

به دست آوردن روشی بازگشتی برای این مسئله، کاری دشوار است و هنوز مقدار دقیقی برای آن وجود ندارد. حتی الگوریتم کارایی برای شمارش این ساختارها وجود ندارد. یکی از این الگوریتم‌ها که مرتبه زمانی بالایی دارد، استفاده از روش پس‌گرد^۳ است. به گونه‌ای که افزارهای مختلف رئوس را ساخته و برای هر دسته یک ساختار دوم قرار می‌دهیم و حاصل را در مجموعه پاسخ‌ها اضافه می‌کنیم. هم‌چنین می‌توان باندهایی برای مسئله پیدا کرد. تعداد ساختارهای دوم دوبخشی را R_n می‌نامیم برای n راس. هم‌چنین قرار می‌دهیم $B_n = R_n + S_n$. در واقع B تعداد اجتماع ساختار دوم و دوم دوبخشی است.

لم ۱.۳.۳

$$R_n \leq S_n^2 - S_n \quad (۵.۳)$$

برهان. به ازای هر افزایی و ساختار دوم یا دوم دوبخشی معتبری، دو دسته را می‌توان این‌گونه فرض کرد که شامل همه رئوس هستند ولی همه‌ی رئوسی که متعلق به دسته‌ی مربوطه نیستند، جفت نشده قرار می‌دهیم. اکنون هر دسته حکم یک ساختار دوم را دارد و همه‌ی ترکیب‌های این دو، همه‌ی حالات ساختار دوم و دوم دوبخشی معتبر رو پوشش می‌دهد. پس:

$$B_n \leq S_n^2$$

و از رابطه‌ی مربوط به B داریم:

$$R_n + S_n \leq S_n^2$$

لذا با یک جابه‌جایی ساده، حکم لم ثابت می‌گردد.

^۳Backtracking

کتاب نامه

- [1] Hofacker, P. Schuster, P. Stadler, Combinatorics of RNA secondary structures, *Discrete Appl. Math.*, 88 (1-3) (1998), pp. 207-237.
- [2] GanjTabesh, M., Steyaert, J.-M., Enumerating RNA structures, including pseudoknots of any topology, (2011) *Match*, 66 (1), pp. 399-414.
- [3] Bernhart, Frank R. (1999), "Catalan, Motzkin, and Riordan numbers", *Discrete Mathematics*, 204 (1-3): 73–112.
- [4] Donaghey, R.; Shapiro, L. W. (1977), "Motzkin numbers", *Journal of Combinatorial Theory, Series A*, 23 (3): 291–301.
- [5] Guibert, O.; Pergola, E.; Pinzani, R. (2001), "Vexillary involutions are enumerated by Motzkin numbers", *Annals of Combinatorics*, 5 (2): 153–174.
- [6] Motzkin, T. S. (1948), "Relations between hypersurface cross ratios, and a combinatorial formula for partitions of a polygon, for permanent preponderance, and for non-associative products", *Bulletin of the American Mathematical Society*, 54 (4): 352–360.
- [7] J. Tao, Y. Xu, and Q. M.zhang. *Current Topics in Computational Molecular Biology*. The MIT Press, 2002.
- [8] I. L.Hofacker, M.Fekete, and F. P. Stadler. Secondary structure prediction for aligned rna sequences. *Journal of Molecular Biology*, 319(5):1059-1066, 2002.

واژه‌نامه

Vertex-labeled graph	گراف برچسب‌دار
Base pair	جفت پایه
Interior	داخلی
Immediately interior	اکیدا داخلی
Unpaired	جفت نشده
Stack	پشته
Loop	حلقه
Terminal base pair	جفت پایه‌ی پایانی
External vertex	راس خارجی
Free end	پایان باز
Joint	مفصل
Component	مولفه
Open Structure	ساختار باز
Hairpin	هیرپین
Multiloop	چندحلقه
Bulge	بالج
Dynamic Programming	برنامه‌نویسی پویا
Backtracking	پس‌گرد
Motzkin Numbers	اعداد موتزکین
Catalan Numbers	اعداد کاتالان
Degree	درجه
Secondary structure	ساختار دوم
Bi-secondary Structure	ساختار دوم دوبخشی

Abstract

Secondary and bi-secondary structures of polynucleotides can be view as a certain class of planner vertex-labeled graphs. They are also similar to another combinatorial problems. We construct recursion formulae enumerating secondary structures and another formulaes with other perspectives. Also, we try to find some bond for the number of bi-secondary structures.

Key words. Planner graphs. Asymptotic Enumeration. Secondary structures. Bi-secondary Structures



Faculty of Science
School of Mathematics, Statistics and Computer Science

Enumeration of RNA secondary & bi-secondary structures

By

Benyamin Delshad Mamaghani

Supervisor

Dr. Mohammad GanjTabesh

Project for receiving B.Sc. degree
Computer Science

February 2018