زیست‌فناوری دانشگاه تربیت مدرس
دورهٔ 3، شمارهٔ 2، پاییز و زمستان 1391

پیش‌گویی ساختار دوم RNA با روش اکتشافی

سهیلا منتصری[1]، نصرالله مقدم‌چرکری[2*]، فاطمه زارع میرک‌آباد[3]

1- کارشناس ارشد، گروه علوم کامپیوتر، دانشکده علوم ریاضی، دانشگاه تربیت مدرس، تهران، ایران
2- استادیار، گروه مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، ایران
3- استادیار، گروه علوم کامپیوتر، دانشکده ریاضی و علوم کامپیوتر، دانشگاه امیرکبیر، تهران، ایران

*تهران، کدپستی 194-14115
charkari@modares.ac.ir
(دریافت مقاله: 90/8/30، پذیرش: 91/4/20)

چکیده- RNAها در بسیاری از فرایندهای زیستی و پزشکی نقش حیاتی دارند و کارکرد RNA به طور مستقیم به ساختارش وابسته است. طراحی ساختارهای RNA مسئله‌ای اصلی در زمینه‌ی زیست‌شناسی است که در درمان و نانوتکنولوژی اهمیت دارد. به همین دلیل الگوریتم‌هایی برای پیش‌گویی ساختار دوم RNA ایجاد شده است. در این مقاله الگوریتمی برای پیش‌گویی دقیق ساختار دوم RNA بر اساس کمترین میزان انرژی آزاد و بیشترین تعداد جفت‌بازهای مجاور ارائه می‌دهیم. این الگوریتم بر پایه‌ی روش اکتشافی است که از یک ماتریس نقطه‌ای برای نشان‌دادن همه‌ی جفت‌بازهای ممکن RNA استفاده می‌کند. سپس استم‌ها[1] از ماتریس‌نقطه‌ای استخراج می‌شوند و براساس طولشان به ترتیب نزولی و سپس استم‌های با طول برابر براساس میزان انرژی آزاد به ترتیب صعودی مرتب می‌شوند. سرانجام استم‌ها به‌ترتیب برای تشکیل ساختار دوم انتخاب می‌شوند. الگوریتم پیشنهادی روی تعدادی از داده‌ها از جمله CopA، CopT، R1inv، R2inv، Tar، *Tar، DIS، $IncRNA_{54}$ و RepZ در باکتری‌ها اجرا می‌گردد. نتایج آزمایش‌ها دقت بالای الگوریتم پیشنهادی را که 95/71 درصد است، نشان می‌دهد. این الگوریتم در زمان کمتری نسبت به سایر روش‌ها اجرا می‌شود.

کلید واژگان: کمینه‌ی انرژی آزاد، جفت‌باز، ماتریس نقطه‌ای، استم

1- مقدمه

مولکول‌های RNA نقش اساسی و مهمی در سلول‌های زنده و در بسیاری از فرایندهای زیستی و پزشکی دارند. شناخت ساختار RNA در درک کارکرد آن اهمیت زیادی دارد. در واقع مولکول‌های RNA مانند بسیاری از پروتئین‌ها، نقش اساسی خود را در سلول با یک ساختار سه بعدی مجزا نشان می‌دهند [1].

تلاش‌هایی برای پیش‌گویی ساختار دوم مولکول‌های RNA با بیشینه کردن تعداد جفت‌بازها و با استفاده از

1. Stem

برنامه‌نویسی پویا آغاز شد که در آن بهترین ساختار برای هر زیردنباله‌ی داده شده محاسبه می‌شود [2]. هم‌زمان الگوریتم مشابهی ارائه شد که از مقادیر انرژی آزاد جفت‌بازها برای محاسبه‌ی ساختاری با کمترین انرژی آزاد استفاده می‌کند [3 و 4]. در رویکردهایی انرژی آزاد با استفاده از مدل ترمودینامیکی نزدیک‌ترین همسایه تعیین می‌شود. در این مدل، انرژی آزاد ساختار به عنوان مجموع انرژی‌های آزادشده از هر استم و حلقه با استفاده از داده‌های ترمودینامیکی به‌دست‌آمده محاسبه می‌شود [5 و 6]. روش دیگر توابع پارتیشن مولکول‌های RNA را برای پیش‌گویی ساختار دوم آن‌ها حساب می‌کند که در آن از برنامه‌نویسی پویا استفاده شده است [7]. روشی بر اساس گرامرهای مستقل از متن پیشنهاد شد که در آن از الگوریتم‌های آماری برای ایجاد ساختار دوم استفاده می‌شود [8]. ابزارهای MFold در مرجع شماره‌ی [9] و RNAfold در مرجع شماره‌ی [10] با استفاده از پارامترهای انرژی ایجادشده در مرجع شماره‌ی [11] ساختارهای دوم RNA را پیش‌گویی می‌کند.

یکی از موضوعاتی که درباره‌ی RNA گفته می‌شود این است که توالی‌هایی وجود دارند که هنوز ساختار آن‌ها مشخص نشده و هیچ نمونه‌ای برای آن‌ها در پایگاه داده‌ها وجود ندارد. بنابراین پیش‌گویی ساختارها ایده‌ی خوبی برای حل این مسئله است [12]. موضوع دیگری که کاربرد مهمی در طراحی ساختارهای RNA دارد، برهم‌کنش دو مولکول RNA است [13]. پیش‌گویی ساختارهای RNA مقدمه‌ای برای تعیین ساختار برهم‌کنش دو RNA است.

در این مقاله الگوریتم اکتشافی دقیقی برای پیش‌گویی ساختار دوم مولکول RNA ارائه می‌شود. در این الگوریتم یک ماتریس نقطه‌ای[1] نشان‌دهنده‌ی همه‌ی جفت‌بازهای ممکن در RNA تشکیل می‌شود؛ سپس زیرقطرهای[2] ماتریس که می‌توانند به عنوان استم در نظر گرفته شوند استخراج شده و کمترین انرژی آزاد آن‌ها محاسبه می‌شود. در ادامه، زیرقطرها بر اساس طول به ترتیب نزولی و سپس زیرقطرهای با طول برابر به ترتیب صعودی براساس انرژی آزاد مرتب می‌شوند. در پایان، زیرقطرها به ترتیب برای تشکیل ساختار دوم RNA انتخاب می‌شوند.

2- مواد و روش‌ها

2-1- انتخاب RNA

مجموعه‌ای از RNAهای استاندارد شامل CopA، CopT، R1inv، R2inv، Tar، Tar*، DIS، $IncRNA_{54}$ و RepZ موجود در مرجع شماره‌ی [14] استفاده می‌شوند.

قبل از بیان الگوریتم، مفاهیم مربوط به RNA را مرور می‌کنیم. RNA زنجیره‌ای طولانی از نوکلئوتیدها است. سه جز ساختاری در ساختمان نوکلئوتیدهای RNA وجود دارد: باز نیتروژنی، قند و گروه فسفات. چهار باز نیتروژنی وجود دارد: آدنین (A)، سیتوزین (C)، گوانین (G) و یوراسیل (U). هر دنباله‌ی RNA دو انتهای مجزا دارد که به عنوان انتهای 5' و 3' شناخته می‌شود. دنباله‌ای که مولکول RNA را توصیف می‌کند به عنوان ساختار اولیه‌ی آن مولکول شناخته می‌شود. پس ساختار اولیه‌ی RNA با $R = r_1 r_2 \dots r_n$ در جهت 5' به 3' نشان داده می‌شود که $|R| = n$ و برای هر $1 \le i \le n$، $r_i \in \{A, C, G, U\}$. از طرفی معکوس R با

1. Dot matrix
2. Sub-diagonals

$r_1 \dots r_n r_{n-1}$ در جهت '3 به '5 مشخص می‌گردد. زیردنباله‌ای از توالی R که از موقعیت i شروع شده و به موقعیت j ختم می‌شود با $r_{ij} = r_i r_{i+1} \dots r_j$ نشان داده می‌شود و معکوس زیردنباله‌ی r_{ij}، $r_i \dots r_j r_{j-1}$ است. زنجیره‌ی تک‌رشته‌ای RNA با تشکیل پیوند هیدروژنی بین بازهای مکمل، به خود متصل می‌شود. بیشتر این پیوندها بین بازهای مکمل واتسون-کریک[1] (C ≡ G) و (U=A) تشکیل می‌شود و همچنین جفت‌باز وابل (G-U) نیز وجود دارد. با ایجاد این پیوندها ساختار دوم و سوم RNA ایجاد می‌شود.

ساختار دوم RNA ترکیبی از استم‌ها و نواحی منفرد[2] است. هر استم مجموعه‌ای از جفت‌بازهای کنار هم است؛ مانند (r_i, r_j) و $(r_{i'}, r_{j'})$ به طوری که i, i', j, j' دارای یکی از شرایط $i < i' < j' < j$ یا $i < i' < j' < j$ باشد. فرض کنید r_{ij} و r_{kl} دو زیردنباله از توالی RNA باشند که تشکیل استم می‌دهند. بنابراین زیر دنباله‌ی r_{ij} به عکس r_{kl} متصل می‌شود یعنی بین هر یک از بازهای r_{ij} و معکوس r_{kl} پیوند هیدروژنی برقرار می‌شود. برای نشان دادن استم در ساختار دوم RNA، هر باز در r_{ij} با '(' و هر باز در r_{kl} با ')' مشخص می‌شود. نواحی منفرد به عنوان حلقه‌ها یا single-strandedها در نظر گرفته می‌شوند. لازم است گفته شود که دو انتهای هر ناحیه‌ی حلقه به استم‌ها متصل می‌شود در حالی که تنها یک انتهای هر single-stranded به یک استم پیوند می‌خورد. هر باز در نواحی منفرد با '.' نشان داده می‌شود. بنابراین $S = s_1 \dots s_n$ ساختار دوم RNA R را نشان می‌دهد که در آن برای هر $1 \le i \le n$، $s_i \in \{'(', ')', '.'\}$.

2-2 ایجاد ساختار دوم RNA

فرض کنید RNA $R = r_1 r_2 \dots r_n$ ورودی داده‌شده باشد. الگوریتم جدیدی به نام HRNA برای پیش‌گویی ساختار دوم R در ادامه ارائه می‌دهیم:

1. ماتریس نقطه‌ای استم[3] $M^R_{n\times n}$ برای دنباله‌ی R ایجاد می‌شود و مقدار آن در موقعیت (i,j) به شکل زیر محاسبه می‌شود:

$$M^R[i,j] = \begin{cases} 1 & if\ (r_i, r_{n-j+1}) \in \{(A,U),(U,A),(C,G),(G,C)\} \\ 0 & else, \end{cases}$$

r_i و r_{n-j+1} به ترتیب باز i ام و $n-j+1$ ام در دنباله‌ی R برای هر $1 \le i, j \le n$ است.

2. در ماتریس نقطه‌ای $M^R_{n\times n}$، همه مقادیر مورب متوالی 1 که روی قطر اصلی یا موازی با آن قرار داشته باشند به عنوان زیرقطر استم[4] انتخاب می‌شود. مجموعه‌ای از زیرقطرهای استم برای R در زیر تعریف می‌شود:

$$D^R = \{< i,j,k,l > | 1 \le i \le k \le n \ \&\ 1 \le j \le l \le n\}$$

(i,j) و (k,l) به ترتیب موقعیت شروع و پایان زیرقطر استم است. فرض کنید $d^R \in D^R$ و $d^R =< i,j,k,l >$؛ این زیرقطر نشان‌دهنده‌ی این است که زیردنباله‌ی r_{ik} در R به زیر دنباله‌ی r_{jl} در معکوس R متصل می‌شود. اگر i' و j' با شرایط $i < i' < k$، $j < j' < l$ و $i' + j' = n + 1$ موجود باشند، آنگاه d^R باید از مجموعه‌ی D^R حذف و دو

1. Watson-Crick
2. Single regions
3. Stem dot-matrix
4. Stem sub-diagonal

زیرقطر (i,j,i',j') و $(i'+1,j'+1,k,l)$ به مجموعه‌ی مورد نظر اضافه شوند.

3. طول و کمینه‌ی انرژی آزاد[1] هر یک از زیرقطرهای استم محاسبه می‌شود. طول زیرقطر استم تعداد 1های موجود در آن است. کمترین انرژی آزاد هر زیرقطر استم s به عنوان مجموع انرژی آزادشده از دو جفت باز مجاور آن، به شکل زیر تعیین می‌شود:

$$MFE(s) = \sum_{(r_i,r_j),(r_{i+1},r_{j-1})\in s} e(r_{i,i+1}, r_{j-1,j}),$$

$e(r_{i,i+1}, r_{j-1,j})$ انرژی آزادشده از دو جفت‌باز مجاور (r_i, r_j) و (r_{i+1}, r_{j-1}) را نشان می‌دهد. کمینه‌ی انرژی آزادشده از همه‌ی دو جفت‌بازهای مجاور در جدول 1 نشان داده شده است.

4. زیرقطرهای استم به ترتیب نزولی طولشان در D^R مرتب می‌شوند و سپس زیرقطرهای با طول برابر براساس کمینه‌ی انرژی آزاد به ترتیب صعودی مرتب می‌شوند.

5. زیرقطرها به ترتیب قرارگیری در مجموعه‌ی D^R برای تشکیل ساختار دوم R انتخاب می‌شوند. توجه کنید که زیرقطرهای انتخابی نباید هم‌پوشانی[2] داشته باشند. فرض کنید $d_1, d_2 \in D^R$ باشند؛ به‌طوری‌که $d_1 = (i_1,j_1,k_1,l_1)$ و $d_2 = (i_2,j_2,k_2,l_2)$. هم‌پوشانی دو زیرقطر استم d_1 و d_2 در ادامه تعریف می‌شود:

$$Overlap(d_1,d_2) = \begin{cases} 1 \text{ if } \exists\, p: i_1 \le p \le k_1 \ \&\ i_2 \le p \le k_2, \\ 1 \text{ if } \exists\, p: j_1 \le p \le l_1 \ \&\ j_2 \le p \le l_2, \\ 0 \ else. \end{cases}$$

جفت‌بازها در ساختار دوم RNA با ')' و '(' نشان داده می‌شوند و بازهایی که جفت‌نشده باقی ماندند با '.' مشخص می‌شوند. پیچیدگی زمان و فضای محاسباتی این الگوریتم به ترتیب $O(n^2 \log n^2)$ و $O(n^2)$ است، (n طول RNA است). همان‌طور که پیش‌تر گفته شد زیرقطرها به‌عنوان مناطق ممکن برای استم در نظر گرفته می‌شوند. بنابراین بکارگیری ماتریس‌نقطه‌ای و یافتن زیرقطرها روشی مناسب برای کاهش زمان محاسباتی است.

مثال. توالی $R = GGAACUUAAGUCC$ و ساختار دوم آن در ادامه نشان داده شده است:

$$R = GGAACUUAAGUCC$$
$$(((\ .(((\ )))\ )))$$

پرانتزهای باز متوالی و پرانتزهای بسته‌ی متناظر آن‌ها یک استم را مشخص می‌کنند. در این مثال دو استم وجود دارد. اولی از اتصال بین GGA و معکوس UCC و دومی از اتصال بین CUU و معکوس AAG ساخته می‌شود.

جدول 1 کمینه‌ی انرژی آزادشده از همه‌ی دو جفت‌بازهای مجاور

5'->3'	AA	AC	AG	AU	CA	CC	CG	CU	GA	GC	GG	GU	UA	UC	UG	UU
AA	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-0.9
AC	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-2.2	.
AG	.	.	.	.	.	.	.	.	.	.	.	.	.	-2.1	.	-0.6
AU	.	.	.	.	.	.	.	.	.	.	.	.	-1.1	.	-1.4	.
CA	.	.	.	.	.	.	.	.	.	.	.	-2.1	.	.	.	.
CC	.	.	.	.	.	.	.	.	.	.	-3.3	.	.	.	.	.
CG	.	.	.	.	.	.	.	.	.	-2.4	.	-1.4	.	.	.	.
CU	.	.	.	.	.	.	.	.	-2.1	.	-2.1	.	.	.	.	.
GA	.	.	.	.	.	.	.	-2.4	.	.	.	.	.	.	.	-1.3
GC	.	.	.	.	.	.	-3.4	.	.	.	.	.	.	.	-2.5	.
GG	.	.	.	.	.	-3.3	.	-1.5	.	.	.	.	.	-2.1	.	-0.5
GU	.	.	.	.	-2.2	.	-2.5	.	.	.	.	.	-1.4	.	1.3	.
UA	.	.	.	-1.3	.	.	.	.	.	.	.	-1	.	.	.	.
UC	.	.	-2.4	.	.	.	.	.	.	.	-1.5	.	.	.	.	.
UG	.	-2.1	.	-1	.	.	.	.	.	-1.4	.	0.3	.	.	.	.
UU	-0.9	.	-1.3	.	.	.	.	.	-0.6	.	-0.5	.	.	.	.	.

3- یافته‌ها

برای ارزیابی دقت پیش‌گویی HRNA از سه معیار حساسیت[3]، برجستگی ویژه[4] و میزان F[5] که در زیر تعریف شده است، استفاده می‌کنیم:

1. Minimum free energy
2. Overlapping

3. Sensitivity
4. Specificity
5. F-measure

$$\text{حساسیت} = \frac{\text{تعداد جفت بازهای به طور صحیح پیشگویی شده}}{\text{تعداد جفت بازها در ساختار مرجع}} \quad (1)$$

$$\text{برجستگی ویژه} = \frac{\text{تعداد جفت بازهای به طور صحیح پیش‌گویی شده}}{\text{تعداد جفت بازها ی پیش‌گویی شده}} \quad (2)$$

$$\text{میزان F} = \frac{\text{برجستگی ویژه}*\text{حساسیت}*2}{(\text{برجستگی ویژه}+\text{حساسیت})} \quad (3)$$

برای نمونه، CopA را در نظر بگیرید. شکل 1 نتیجه‌ی پیش‌گویی ساختار دوم CopA را که بسیار به ساختار واقعی نزدیک است نشان می‌دهد. جدول 2 دقت پیش‌گویی HRNA در حساسیت، برجستگی ویژه و میزانF را بر روی داده‌های آزمایشی نشان می‌دهد. برای RNAهای R1inv، Tar، *Tar، R2inv، DIS و CopT دقت پیش‌گویی 100/00 درصد در هر سه معیار است. دقت پیش‌گویی CopA در حساسیت، برجستگی ویژه و میزان F، به‌ترتیب 92/3 ، 100/00 و 96/00 درصد است. برای $IncRNA_{54}$، RepZ، میزانF به ترتیب 87/91 و 75/85 درصد به دست آمده است. همان‌طور که مشاهده می‌شود، دقت متوسط الگوریتم پیشنهادی روی داده‌های آزمایشی به‌ترتیب 96/76، 94/66 و 95/71 درصد در حساسیت، برجستگی ویژه و میزانF است. بنابراین پیش‌گویی با HRNA، نتایج دقیقی می‌دهد. در این جدول زمان اجرای الگوریتم روی داده‌ها بر حسب ثانیه نشان داده شده است. پیچیدگی زمانی تعدادی از روش‌های موجود در مراجع [2،7،15،16،17،18،19،20] در جدول 3 نشان داده شده است. طبق این جدول کمینه‌ی زمان محاسباتی این الگوریتم‌ها $O(n^3)$ است. همان‌طور که بیان شد، HRNA در زمان $O(n^2 \log n^2)$ اجرا می‌شود. بنابراین الگوریتم پیشنهادی در مقایسه با سایر روش‌ها به زمان محاسباتی کمتری نیاز دارد.

5'-GUGGGCCCCGGUAAUCUUUUCGUACUCGCCAAAGUUGAAGAAGAUUAUCGGGGUUU-3'

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

ساختار دوم واقعی CopA

5'-GUGGGCCCCGGUAAUCUUUUCGUACUCGCCAAAGUUGAAGAAGAUUAUCGGGGUUU-3'

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

ساختار دوم پیش‌گویی‌شده CopA

شکل 1

جدول 2 دقت پیش‌گویی و زمان محاسباتی HRNA

توالی RNA	طول	حساسیت (%)	برجستگی ویژه (%)	میزان F (%)	زمان (ثانیه)
Tar	16	100/00	100/00	100/00	1/09e -6
Tar*	16	100/00	100/00	100/00	1/97e-6
R1inv	21	100/00	100/00	100/00	2/01e-6
R2inv	19	100/00	100/00	100/00	1/89e-6
DIS	35	100/00	100/00	100/00	3/7e-6
CopA	56	92/30	100/00	96/00	1/35e-5
CopT	57	100/00	100/00	100/00	5/47e-6
$IncRNA_{54}$	54	100/00	78/57	87/91	5/61e-6
RepZ	61	78/57	73/33	75/85	6/86e-6
Average		96/76	94/66	95/71	4/65e-6

جدول 3 مقایسه‌ی پیچیدگی زمانی تعدادی از روش‌ها

الگوریتم پیشگویی ساختار RNA	مرجع	پیچیدگی زمانی
HRNA	-	$O(n^2 \log n^2)$
RNAfold	Hofackeret al	$O(n^3)$
Akutsu's Alg.	Akutsu	$O(n^4)$
Pknots-RE	Rivas and Eddy	$O(n^6)$
RNAalifold	Hofacker et al	$O(m \times n^4 + n^3)$
DP. Partition function alg.	McCaskill	$O(n^3)$
Zuker's Alg.	Zuker and Stiegler	$O(n^4)$
Nussinov's Alg.	Nussinov et al	$O(n^3)$
Waterman and Smith Alg.	Waterman and Smith	$O(n^3)$

4- بحث و نتیجه‌گیری

در این مقاله روش اکتشافی دقیقی برای پیش‌گویی ساختار دوم RNA معرفی شد. در این روش یک ماتریس‌نقطه‌ای ساخته شد که نشان‌دهنده‌ی همه‌ی جفت‌بازهای ممکن در RNA است. سپس زیرقطرهای ماتریس که به عنوان مناطق ممکن برای استم در نظر گرفته می‌شوند استخراج شد. در ادامه، طول هر یک از زیرقطرها و کمینه‌ی انرژی آزاد آن‌ها محاسبه شد. پس از آن، زیرقطرها بر اساس طول به ترتیب نزولی و زیرقطرهای با طول برابر، به ترتیب صعودی کمینه‌ی انرژی آزاد مرتب شدند. در پایان، زیرقطرها به ترتیب برای تشکیل ساختار دوم RNA انتخاب شدند. الگوریتم پیشنهادی روی تعدادی از RNAها در باکتری‌ها به‌کار برده شد. نتایج آزمایش‌ها دقت بالای 95/71 درصد این الگوریتم را نشان می‌دهد. الگوریتم به زمان و فضای محاسباتی $O(n^2 log\, n^2)$ و $O(n^2)$ نیاز دارد که n نشان‌دهنده‌ی طول RNA است. بنابراین زمان محاسباتی الگوریتم پیشنهادی کمتر از سایر الگوریتم‌های مشابه است.

به خاطر استفاده از ماتریس نقطه‌ای، یافتن زیرقطرهای آن و همچنین در اولویت قرار دادن زیرقطرهای با کمینه‌ی

انرژی آزاد برای تشکیل پیوند، HRNA نتایج دقیقی را در زمان محاسباتی کم پیش‌گویی می‌کند.

5- سپاسگزاری

از دکتر محمد گنج تابش بـه علـت فـراهم‌کـردن جـدول کمینه انرژی آزاد (جدول 1) تشکر می‌شود.

6- مراجع

[1] Meyer, I.M. (2008) Predicting novel RNA-RNA interactions. Current opinion in structural biology. 18, 387-393.

[2] Nussinov, R., Pieczenik, G., Griggs, J.R., and Kleitman, D.J. (1978) Algorithms for loop matching. SIAM J.Appl.Math. 35, 68-82.

[3] Nussinov, R., and Jacobson, A.B. (1980) Fast algorithm for predicting the secondary structure of single-stranded RNA. In Proceedings of the National Academy of Sciences of the United States of American. 77, 6309-6313.

[4] Zuker, M., and Sankoff, M. (1984) RNA secondary structures and their prediction. Blletin of Mathematical of biology. 46(4), 591-621.

[5] Zuker, M., Mathews, D.H., and Turner, D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In RNA Biochemistry and Biotechnology.

[6] Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. Journal of Molecular Biology. 288, 911–940.

[7] McCaskill, J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers. 29, 1105-1119.

[8] Sakakibara, Y., Brown, M., Hughey, R., Mian, I.S., Sjolander, K., Underwood, R.C., and Hussler, D. (1999) Stochastic context-free grammars for tRNA modeling. Nucleic Acids Res. 22, 5112-5120.

[9] Zuker, M. (1994) Prediction of RNA secondary structure by energy minimization. Journal of Molecular Biology. 25, 267–94.

[10] Hofacker, I.L. (2009) Vienna RNA secondary structure server. Nucleic Acids Research. 31(13), 3429–3431.

[11] Mathews, D.H., and Turner, D.H. (2006) Prediction of RNA secondary structure by free energy minimization. 16(3), 270-278.

[12] Zvelebil, M., and Baum, J.O. (2008) Understanding Bioinformatics. Garland Science. 461-514.

[13] Salare, R., Backofen, R., and Sahinalp, S.C. (2010) Fast prediction of RNA-RNA

interaction, Algorithms for molecular Biology. 5, 5-15.

[14] 14. Kato, Y., Akutsu, T., and Seki, H. (2009) A grammatical approach to RNA-RNA interaction prediction. Pattern recognition. 42, 531-538.

[15] Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, L., Tacker, M., and Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. Monatshefte Chemie. 125, 167-88.

[16] Akutsu, T. (2000) Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. Discrete Applied Mathematics. 104, 45–62.

[17] Rivas, E., and Eddy, S.R. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. Journal of molecular biology. 285(5): 2053-68.

[18] Hofacker, I.L., Fekete, M., and Stadler, P.F. (2002) Secondary structure prediction for aligned RNA sequences. Journal of molecular biology. 319(5), 1059–66.

[19] Zuker, M., and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic acids research. 9(1): 133–48.

[20] Waterman, M., and Smith, T. (1978) RNA secondary structure: A complete mathematical analysis. Math Biosci. 42(3–4), 257–66.