`Rtools`: a web server for various secondary structural analyses on single RNA sequences

Michiaki Hamada[1,2], Yukiteru Ono[3], Hisanori Kiryu[4], Kengo Sato[5], Yuki Kato[6], Tsukasa Fukunaga[4], Ryota Mori[4] and Kiyoshi Asai[2,4,*]

[1]Department of Electrical Engineering and Bioscience, Faculty of Science and Engineering, Waseda University, 55N–06–10, 3–4–1, Okubo Shinjuku-ku, Tokyo 169–8555, Japan, [2]Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST), 2–4–7 Aomi, Koto-ku, 135–0064 Tokyo, Japan, [3]IMSBIO Co., Ltd, 4–21–1–601 Higashi–Ikebukuro, Toshima-ku, Tokyo 170–0013, Japan, [4]Graduate School of Frontier Sciences, University of Tokyo, 5–1–5 Kashiwanoha, Kashiwa 277–8562, Japan, [5]Department of Biosciences and Informatics, Keio University, 3–14–1 Hiyoshi, Kohoku–ku, Yokohama 223–8522, Japan and [6]Center for iPS Cell Research and Application (CiRA), Kyoto University, Shogoin, Sakyo-ku, Kyoto 606–8507, Japan

ABSTRACT

The secondary structures, as well as the nucleotide sequences, are the important features of RNA molecules to characterize their functions. According to the thermodynamic model, however, the probability of any secondary structure is very small. As a consequence, any tool to predict the secondary structures of RNAs has limited accuracy. On the other hand, there are a few tools to compensate the imperfect predictions by calculating and visualizing the secondary structural information from RNA sequences. It is desirable to obtain the rich information from those tools through a friendly interface. We implemented a web server of the tools to predict secondary structures and to calculate various structural features based on the energy models of secondary structures. By just giving an RNA sequence to the web server, the user can get the different types of solutions of the secondary structures, the marginal probabilities such as base-paring probabilities, loop probabilities and accessibilities of the local bases, the energy changes by arbitrary base mutations as well as the measures for validations of the predicted secondary structures. The web server is available at http://rtools.cbrc.jp, which integrates software tools, `CentroidFold`, `CentroidHomfold`, `IPKnot`, `CapR`, `Raccess`, `Rchange` and `RintD`.

INTRODUCTION

The primary sequences of RNAs are important for their functions because the base-parings of the reverse com-plementary sub-sequences identify the target RNAs. At the same time, those RNAs frequently conserve their secondary structures. The secondary structures reflect the tertiary structures and also related to the free sub-sequences which can interact with the reverse complementary sub-sequences of the other RNAs. While high-throughput experimental methods have still limitations, it is desirable to characterize the secondary structures of RNAs *in silico*.

In silico prediction of RNA secondary structures is often referred as 'unreliable,' often failing to predict well-known secondary structures like the clover-leaf of tRNA e.g. (1). Although the prediction gives approximately 'correct' secondary structures in considerable frequency, the predicted structures almost always include many errors.

The probability distribution of the secondary structure θ of an RNA sequence x is denoted as Boltzmann distribution (2):

$$p(\theta|x) = \frac{1}{Z(x)} \exp\left(\frac{-E(\theta|x)}{RT}\right), \qquad (1)$$

where $E(\theta|x)$ is the free energy of the secondary structure θ, R is the ideal gas constant, T is a temperature and $Z(x)$ is the partition function,

$$Z(x) = \sum_{\theta} \exp\left(\frac{-E(\theta|x)}{RT}\right). \qquad (2)$$

The nearest neighbor energy model defines $E(\theta|x)$ as the summation of the free energy of all the loops, which is calculated according to the combinations of the nearest nucleotides of the loops. The experimentally identified parameters are used for this calculation.

The number of possible secondary structures of an RNA is huge, and the probability of the predicted structure (with respect to the Boltzmann distribution) is very small, around

*To whom correspondence should be addressed. Tel: +81 3 3599 8089; Fax: +81 3 3599 8089; Email: asai@k.u-tokyo.ac.jp

1% even in the fortunate cases. For example, it is known that the structure of an rRNA with the highest probability is about 10^{-22} (3). Therefore, it is dangerous to proceed to further biological inferences based on the predicted structures.

The *in silico* secondary structural analyses are, however, not useless but productive if the combination of the tools are appropriately utilized. Although the total probability of a precise secondary structure of an RNA sequence is very small, various *marginal* probabilities with respect to the distribution $p(\theta|x)$, which are not necessarily very small, can be introduced. For example, the 'base-pairing probability matrix' (BPPM), where each component p_{ij} is the marginal probability that x_i and x_j form a base-pair (called 'base-pairing probability') reflects the landscape of the probability distribution of the secondary structures. BPPM can be computed efficiently in $O(L^3)$ time (L is the length of RNA sequence) by using a dynamic programming (DP) based on the McCaskill model (2) (cf. Figure 2). Similar algorithms give us the marginal probabilities of a variety of local structural features, such as loop probabilities and accessibilities (4,5).

The extensions of DP algorithms also enable us to extract different types of secondary structural information. For example, the energy change by all the possible base mutations (including the combination of base mutations) can be calculated (6). The validation of the predicted structure is also important. Because the probability of the structure itself is very small, more reasonable measures are required. The credibility limit (7) is the measure of the concentration of the probabilities around the predicted structures, which reflects the stability and the reliability of the predicted structure.

DESIGN OF `Rtools` WEB SERVER

In `Rtools` web server, we implemented seven tools described in the next section, all of which focus on analyzing secondary structures for single RNA sequences, based on probability distribution of RNA secondary structures. Due to the reason, an interface of `Rtools` is quite simple, where users paste an RNA sequence in the text-form and press 'Submit' button; then `Rtools` quickly returns various analyses for the sequence. The results (e.g. base-pairing probability matrix and graphs) are offered as the following formats: EPS, PNG and JPEG, which can be easily utilized in the research paper. Additionally, users can easily adjust internal parameters in each tool through web interface (Figure 1). In `Rtools`, we implemented a routine for visualizing base-pairing probability matrix in colors. Additionally, users obtain 'Request ID' in the results page and can always access the results by using the ID.

ALGORITHMS AND SOFTWARE

Given an RNA sequence x, a set of possible secondary structures of x is denoted by $\mathcal{S}(x)$, where a secondary structure $\theta \in \mathcal{S}(x)$ is represented as a (upper-)triangular binary-valued matrix $\{\theta_{ij}\}_{i < j}$: $\theta_{ij} = 1$ if the *i*-the nucleotide in x (denoted by x_i) forms a base-pair with x_j. In addition, a probability distribution of RNA secondary structure of x

is denoted by p($\theta|x$) ($\theta \in \mathcal{S}(x)$), where $\sum_{\theta \in \mathcal{S}(x)} p(\theta|x)$ holds. Several models for $p(\theta|x)$ have been proposed (as described in Introduction section). By using this notation, a marginal probability that x_i and x_j form a base-pair (a base-pairing probability described in Introduction) is written as follows:

$$p_{ij} = \sum_{\theta \in \mathcal{S}(x)} I(\theta_{ij} = 1)p(\theta|x) \tag{3}$$

where $I(\cdot)$ is the indicator function and θ_{ij} is the (*i, j*)-element of $\theta \in \mathcal{S}(x)$ i.e., $I(\theta_{ij} = 1)$ is equal to 1 only when x_i and x_j form a base-pair ($I(\theta_{ij} = 1) = 0$ otherwise).

Each tool implemented in `Rtools` web server is closely related to marginal probabilities (including base-pairing probability) with respect to a distribution of RNA secondary structures, as described in the following.

CentroidFold

`CentroidFold` (8) predicts a pseudo-knot-free RNA secondary structure from an individual RNA sequence. The algorithm of `CentroidFold` is based on γ-centroid estimators (9), a type of maximizing expected accuracy (MEA) estimators. The γ-centroid estimators are known to have better expected accuracy than the minimum free energy (MFE) predictions, in terms of the number of true positives and true negatives of base pairs (8). On average, it is known that CentroidFold achieves better performance with respect to accurate predictions of base-pairs than other tools (10).

The time complexity of `CentroidFold` using base-pairing probability matrix $\{p_{ij}\}_{i < j}$ (Equation 3) is $O(L^3)$ for an RNA sequence of length L. As probabilistic models for RNA secondary structures, `CentroidFold` can utilize one of the following models: McCaskill model (2) with Turner 2004 energy parameters (11) that were determined experimentally; McCaskill model with Boltzmann Likelihood (BL) parameters (12) obtained by retraining energy parameters using machine learning procedures; CONTRAfold model (13) based on a machine learning based model called conditional random fields (CRFs). The interface allows us to select the energy model between McCaskill and CONTRAfold. In `CentroidFold`, ViennaRNA package (14) is utilized to compute base-pairing probability matrix for McCaskill model.

CentroidHomfold

To improve the accuracy of secondary structure predictions, comparative approaches with homologous sequences are often utilized because important structures are conserved evolutionarily. `CentroidHomfold` (15) aims to predict secondary structures more accurately (than `CentroidFold`) by using the information of the *homologous* sequences. The initial version of `CentroidHomfold` requires not only target RNA sequence whose secondary structure is to be predicted but also its homologous sequences. However, `CentroidHomfold` in `Rtools` does not need homologous sequences because `Rtools` automatically collects homologous sequences of the input from RNA sequence database (constructed based on Rfam database) (16).

Figure 1. A snapshot of the `Rtools` web server (http://rtools.cbrc.jp/). Users can paste RNA sequence in the text form (located in the top part) and press the submit button, then the server quickly returns various analyses based on secondary structures for the sequence (not shown); the results are downloadable with a few formats: TEXT, PNG, PDF and EPS. Additionally, users can easily adjust several internal parameters in each tool through web interface (shown in the bottom part).

IPKnot

`IPKnot` (17) predicts RNA secondary structures including pseudo-knot, while `CentroidFold` predicts only pseudo-knot-free RNA secondary structures. As with `Centroid-Fold`, `IPKnot` employs the maximum expected accuracy estimators in combination with integer programming for optimizing the score function with base-pairing probabilities derived from the principle of MEA. In general, computational costs should be much higher than secondary structure prediction without pseudo-knot. Although the worst case time complexity of `IPKnot` is not polynomial order, it has been empirically proven that `IPKnot` is fast enough for long RNA sequences.

A predicted secondary structure by `IPKnot` is visualized using VARNA (18). ViennaRNA package (14) are utilized to compute base-pairing probability matrix for McCaskill model. Additionally, NUPACK (Dirks-Pierce model) (19) is utilized for computing base-pairing probability matrix with considering pseudo-knot structures.

CapR

`CapR` (4) efficiently computes marginal probabilities that each RNA base position is located within each secondary structural context, where six categories of RNA structures were taken into account: (i) stem, (ii) hairpin, (iii) bulge, (iv) internal loop, (v) multi-branch loop and (vi) exterior loop (cf. Figure 3). In `CapR`, those marginal probabilities are computed according to a probability distribution of secondary structures of a given RNA sequence, provided by McCaskill (energy) model.

In order to reduce computational time, `CapR` introduces a maximal span of base-pairs, leading to the computational time of $O(Lw^2)$, where w is a maximal span of base-pair and L is the length of input sequence.

Raccess

`Raccess` (5) computes accessibility (the marginal probability of forming a single strand with respect to a probability distribution of secondary structures) of every region of RNA sequences. In `Raccess`, McCaskill model is uti-
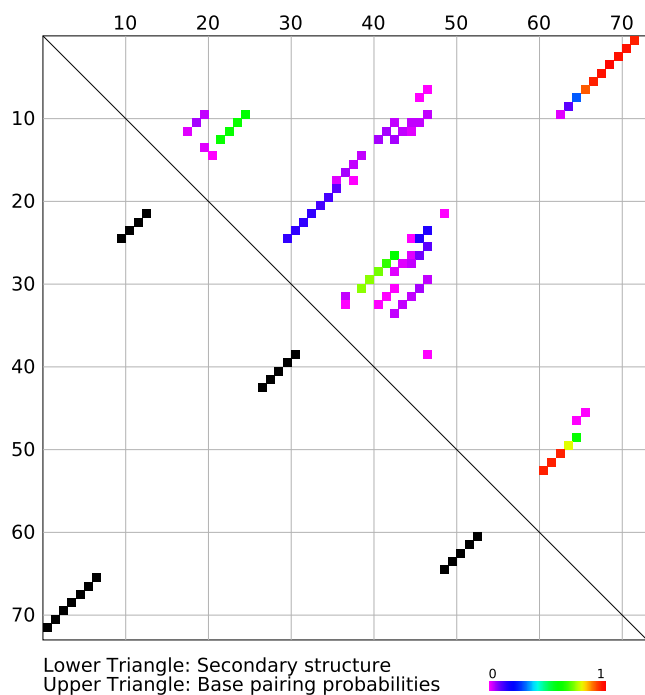
Figure 2. The upper triangular matrix shows a base-pairing probability matrix (BPPM) of a tRNA sequence (having a 'clover-leaf' secondary structure), produced by `Rtools` web server (with `CentroidFold`). A base-pairing probability is a marginal probability (with respect to a probability distribution of secondary structures) that a pair of nucleotides form a base-pair. The colors of plots show the values of probabilities. On the other hand, the lower triangle shows the optimal RNA secondary structure predicted by `CentroidFold`, using the base-pairing probability matrix shown in the upper triangle.
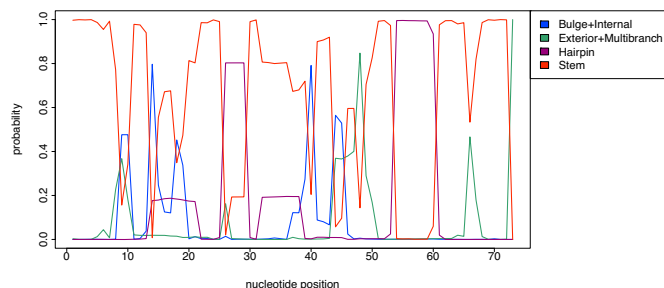

Figure 3. Results of `CapR` for a tRNA sequence. The horizontal axis shows nucleotide position of the input sequence and the vertical axis shows marginal probabilities for structural contexts (bulge, exterior, hairpin, internal, multi-branch and stem) with respect to a probability distribution of secondary structures of the sequence. In `CapR`, the distribution is provided by McCaskill model (Equation 1), which is based on experimentally determined energy parameters.

lized for a probability distribution of secondary structures. Although the accessibility for every nucleotide is computed by using probability of being unpaired, which is easily calculated by base-pairing probability matrix, `Raccess` enables users to specify the accessible length (Figure 4), which would be useful for miRNA target predictions.

Given an input sequence x and a segment $[a, b]$, which means the sub-sequence $x_a x_{a+1} \cdots x_b$ of a given RNA se-
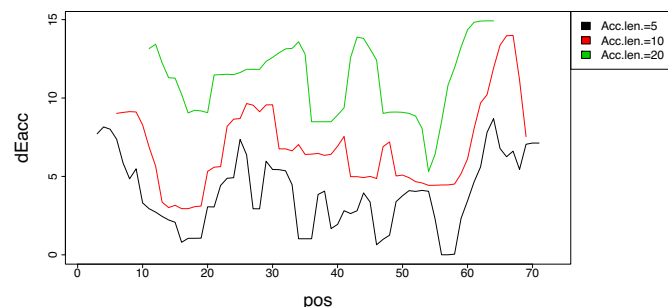

Figure 4. Results of `Raccess` for a tRNA sequence, where the colors of graph show accessible length. In this figure, `Raccess` computes the accessibility of a segment $[a, b] = [x, x + 1 - 1]$ in the transcript for all the positions x with fixed length l of 5, 10 and 20. Thermodynamic energy that is required to keep range $[a, b]$ being accessible is given by Equation 4. In the figure, these access_energy($[a, b]$) are plotted at $(x + l/2)$.

quence x, the accessibility of the segment is defined by

$$\text{access\_energy}([a, b]) = -RT \log(p([a, b])) \quad (4)$$

where

$$p([a, b]) = \frac{\sum_{\theta \in S[a,b]} \exp(-E(\theta)/RT)}{\sum_{\theta \in S(x)} \exp(-E(\theta)/RT)} \quad (5)$$

In the above, $S[a, b]$ is all the secondary structures having range $[a, b]$ as loop region (Clearly, $S[a, b] \subseteq S(x)$ holds). Accessible energy is considered to be energy that is required to keep range $[a, b]$ being accessible. Note that accessibility is closely related to marginal probabilities with respect to a probability distribution of secondary structures (cf. Equation 5).

Rchange

`Rchange` (6) computes entropy and internal energy changes of secondary structures for mutated sequences. Given an input RNA sequence x and its mutated sequence x', the relative change of ensemble free energy is defined by

$$\frac{dF}{|F|} = \frac{F(x) - F(x')}{|F(x)|} \quad (6)$$

where

$$F(x) = -RT \log(Z(x)).$$

In the above, $Z(x)$ is the partition function (defined in Equation 2). In `Rtools`, the maximum and minimum values of Equation 6 with respect to SNPs for each position in the sequence are provided (cf. Figure 5).

Recently, 'riboSNitch' has been receiving great attentions (20) and `Rchange` will be useful for computational analysis for riboSNitch.

RintD

`RintD` (1) provides an efficient method to summarize the probability distribution $p(\theta|x)$, which cannot be summarized (visualized) easily due to its high dimensionality of each solution.
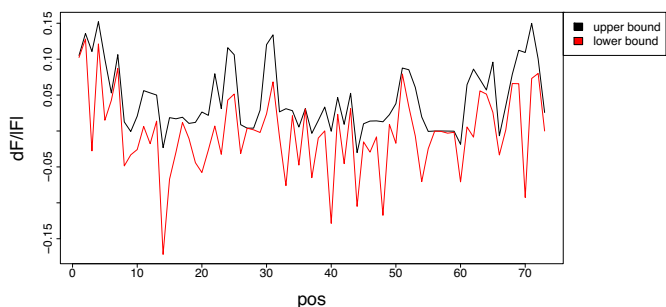
Figure 5. Results of Rchange for a tRNA sequence. The upper and lower bounds of the relative changes of the ensemble free energy ($dF/|F|$ in Equation 6. Because there are three mutated nucleotides for each position, the upper bound values (shown in black line) indicate the largest energy increase caused by a single mutation, and the lower bound values (shown in red line) indicate the largest energy decrease caused by a single mutation.
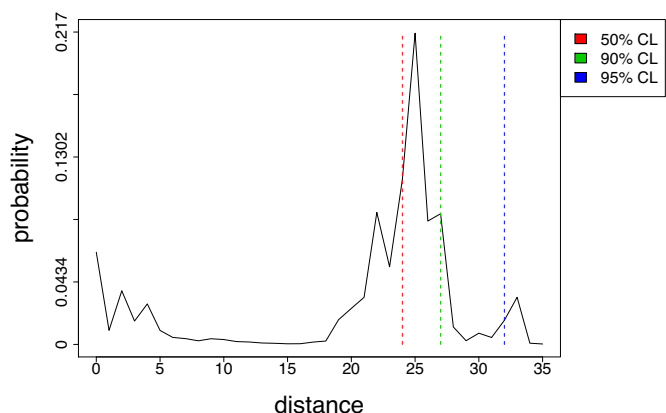

Figure 6. The reference secondary structure was obtained by Centroid-Fold. The horizontal axis indicates the (hamming) distance of the reference secondary structure (d in Equation 7), and the vertical axis shows marginal probabilities p_d in Equation 7. The dashed lines indicate α%-credibility limits (7), $CL(\alpha)$ in Equation 8: $CL(50)$ (red), $CL(90)$ (green) and $CL(95)$ (blue).

Given a *reference* (or *target*) secondary structure $y \in \mathcal{S}(x)$ of interest, $\mathcal{S}_d$ denotes a set of RNA secondary structures whose hamming distance from y is equal to d. More formally, we define

$$\mathcal{S}_d = \left\{ \theta \in \mathcal{S}(x) \middle| d = \sum_{i<j} I(\theta_{ij} \neq y_{ij}) \right\}$$

where $I(\cdot)$ is the indicator function. Then, RintD can efficiently compute $\{p_d\}_d$ ($d = 0, 1, \ldots$) where p_d is a marginal probability as follows:

$$p_d = \sum_{\theta \in \mathcal{S}_d} p(\theta|x) \qquad (7)$$

In RintD, McCaskill model is utilized as $p(\theta|x)$.

For a given RNA secondary structure y, α% *credibility limit* (7) (denoted by $CL(\alpha)$) is defined as follows:

$$CL(\alpha) = \min \left\{ d \in \mathbb{N} \cup \{0\} \middle| \sum_{d':d'\leq d} p_{d'} \geq \frac{\alpha}{100} \right\} \qquad (8)$$

where $\mathbb{N}$ is the set of all natural numbers. In other words, α%-credibility limit is the minimum Hamming distance radius that contain α% of the possible RNA secondary structures (7). Clearly, the credibility limit for the target sequence can be computed by utilizing $\{p_d\}_{d=0, 1, \ldots}$. Note that the smaller credibility limit indicates that the reference structure is more stable (reliable) in the ensemble of secondary structures.

RintD computes $\{p_d\}_{d=0, 1, \ldots}$ and credibility limits of the reference structure. Moreover, RintD can efficiently compute the distribution from *two* reference RNA secondary structures (cf. (1)), which is also implemented in Rtools (cf. Figure 6).

AN EXAMPLE OF PERFORMING Rtools WEB SERVER

As a case study of our web server (Rtools), we took a transfer RNA (tRNA) sequence (AL138651.1/64525-64597): GGGGAUGUAGCUCAUAUGGUAGAGCG CUCGCUUUG CAUGCGAGAGGCACAGGGUUCG AUUCCCUGCAUCU CCA, and analyzed it by using Rtools. As a result, the execution time of Rtools was within 1 min.

A base-pairing probability matrix and the optimal secondary structure predicted by CentroidFold of the tRNA sequence are shown in upper and lower matrices in Figure 2, respectively. The figure shows that Centroid-Fold successfully predicted 'clover-leaf secondary structure' of tRNAs for this sequence. Although the BPPM does not predict RNA secondary structures, it includes richer information of secondary structures than the optimal secondary structure. Moreover, Figure 3 indicates the results for CapR, where the marginal probabilities of six structural contexts (Bulge, Exterior, Hairpin, Internal, Multi-blanch and stem) in each position are shown. We emphasize that those structural information includes richer information than conventional base-pairing probability matrix.

Figure 4 shows the accessibility (Equation 4) of the sequence, where three accessible lengths are shown. Users can find accessible region from this figure; in this case the figure suggests that the loop region of 'clover-leaf secondary structure' is accessible.

Figure 5 shows the results of Rchange for the same tRNA sequence, where the upper bound and lower bound of $dF/|F|$ in Equation 6 are plotted in each position. By utilizing this figure, users can easily find nucleotides whose SNPs greatly affect the RNA secondary structures. Additionally, it would be useful to see predicted secondary structures (by CentroidFold) and context profiles (by CapR) in combination with the results of Rchange.

Figure 6 (the results of RintD) shows the distribution of secondary structures from the reference structure (the optimal structure predicted by CentroidFold). Specifically, it plots (d, p_d) for $d = 0, 1, 2, \ldots$, where p_d is defined in Equation 7. The figure shows that there exists a peak of probability mass around $d = 25$. Moreover, the credibility limits of the optimal secondary structure ($CL(50)$, $CL(90)$ and $CL(95)$ in Equation 8) are relatively large, indicating that the sum of the probabilities of secondary structures near the optimal secondary structure is not dominant in the ensem-

ble distribution of secondary structures. (Note that all the figures above were generated by `Rtools` web server.)

In summary, by utilizing `Rtools`, users can obtain not only secondary structures but also various information including marginal probabilities of structural context (computed by `CapR`), accessibility (computed by `Raccess`), the influence of single nucleotide mutation (computed by `Rchange`) and the summary of the distribution with respect to the predicted secondary structure (computed by `RintD`).

DISCUSSION

There exists several related web servers such as RNAfold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) (14), RNAstructure web server (21) and Sfold web server (22). As described in previous sections, `Rtools` include several features that are different from those existing web server.

One direction to improve the web server is to incorporate experimental probing information (23) into our structural analyses. Because there exist several methods to incorporate the probing information into RNA secondary structure prediction e.g. (24), we plan to implement one (or a few) of those methods in the future.

CONCLUSION

In this study, we have developed a novel web server, `Rtools` (http://rtools.cbrc.jp/), to perform various secondary structural analyses on *single* RNA sequences by `Centroid-Fold`, `CentroidHomfold`, `IPKnot`, `CapR`, `Rchange`, `Raccess` and `RintD` (all of which were published in reliable journals). The all tools in `Rtools` are closely related to marginal probabilities with respect to probability distribution of RNA secondary structures of given RNA sequence. By 'one-click', users can conduct various analyses of single RNA sequences that they are interested in, and obtain many results which will be utilized for their research paper and/or their further study.

ENDNOTES

In general, marginal probabilities are computed by the following form: $p = \sum_{\theta \in S} p(\theta|x)$ where S is a subset of $\mathcal{S}(x)$ ($S \subset \mathcal{S}(x)$). Various marginal probabilities are introduced according to S.

FUNDING

REFERENCES

1. Mori,R., Hamada,M. and Asai,K. (2014) Efficient calculation of exact probability distributions of integer features on RNA secondary structures. *BMC Genomics*, **15**(Suppl. 10), S6.
2. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
3. Hamada,M. (2014) Fighting against uncertainty: an essential issue in bioinformatics. *Brief. Bioinformatics*, **15**, 748–767.
4. Fukunaga,T., Ozaki,H., Terai,G., Asai,K., Iwasaki,W. and Kiryu,H. (2014) CapR: revealing structural specificities of RNA-binding protein target recognition using CLIP-seq data. *Genome Biol.*, **15**, R16.
5. Kiryu,H., Terai,G., Imamura,O., Yoneyama,H., Suzuki,K. and Asai,K. (2011) A detailed investigation of accessibilities around target sites of siRNAs and miRNAs. *Bioinformatics*, **27**, 1788–1797.
6. Kiryu,H. and Asai,K. (2012) Rchange: algorithms for computing energy changes of RNA secondary structures in response to base mutations. *Bioinformatics*, **28**, 1093–1101.
7. Webb-Robertson,B.J., McCue,L.A. and Lawrence,C.E. (2008) Measuring global credibility with application to local sequence alignment. *PLoS Comput. Biol.*, **4**, e1000077.
8. Hamada,M., Kiryu,H., Sato,K., Mituyama,T. and Asai,K. (2009) Prediction of RNA secondary structure using generalized centroid estimators. *Bioinformatics*, **25**, 465–473.
9. Hamada,M., Kiryu,H., Iwasaki,W. and Asai,K. (2011) Generalized centroid estimators in bioinformatics. *PLoS ONE*, **6**, e16450.
10. Puton,T., Kozlowski,L.P., Rother,K.M. and Bujnicki,J.M. (2013) CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Res.*, **41**, 4307–4323.
11. Turner,D.H. and Mathews,D.H. (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.*, **38**, D280–D282.
12. Andronescu,M.S., Pop,C. and Condon,A.E. (2010) Improved free energy parameters for RNA pseudoknotted secondary structure prediction. *RNA*, **16**, 26–42.
13. Do,C.B., Woods,D.A. and Batzoglou,S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–e98.
14. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
15. Hamada,M., Sato,K., Kiryu,H., Mituyama,T. and Asai,K. (2009) Predictions of RNA secondary structure by combining homologous sequence information. *Bioinformatics*, **25**, i330–i338.
16. Hamada,M., Yamada,K., Sato,K., Frith,M.C. and Asai,K. (2011) CentroidHomfold-LAST: accurate prediction of RNA secondary structure using automatically collected homologous sequences. *Nucleic Acids Res.*, **39**, W100–W106.
17. Sato,K., Kato,Y., Hamada,M., Akutsu,T. and Asai,K. (2011) IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics*, **27**, 85–93.
18. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
19. Zadeh,J.N., Steenberg,C.D., Bois,J.S., Wolfe,B.R., Pierce,M.B., Khan,A.R., Dirks,R.M. and Pierce,N.A. (2011) NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170–173.
20. Solem,A.C., Halvorsen,M., Ramos,S.B. and Laederach,A. (2015) The potential of the riboSNitch in personalized medicine. *Wiley Interdisc. Rev. RNA*, **6**, 517–532.
21. Bellaousov,S., Reuter,J.S., Seetin,M.G. and Mathews,D.H. (2013) RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*, **41**, W471–W474.
22. Ding,Y., Chan,C.Y. and Lawrence,C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, **32**, W135–W141.
23. Mortimer,S.A., Kidwell,M.A. and Doudna,J.A. (2014) Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.*, **15**, 469–479.
24. Quarrier,S., Martin,J.S., Davis-Neulander,L., Beauregard,A. and Laederach,A. (2010) Evaluation of the information content of RNA structure mapping data for secondary structure prediction. *RNA*, **16**, 1108–1117.