Freiburg RNA Tools: a web server integrating IntaRNA, ExpaRNA and LocARNA

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1093/nar/gkq316">http://dx.doi.org/10.1093/nar/gkq316</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Oxford University Press</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Feb 17 17:43:52 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/72078">http://hdl.handle.net/1721.1/72078</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution Non-Commercial</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc/2.5">http://creativecommons.org/licenses/by-nc/2.5</a></td>
</tr>
</tbody>
</table>
Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LocARNA

Cameron Smith1, Steffen Heyne1, Andreas S. Richter1, Sebastian Will1,2 and Rolf Backofen1,*

1Bioinformatics Group, University of Freiburg, Freiburg 79110, Germany and 2Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received February 2, 2010; Revised March 31, 2010; Accepted April 17, 2010

ABSTRACT
The Freiburg RNA tools web server integrates three tools for the advanced analysis of RNA in a common web-based user interface. The tools INTARNA, EXPARNA and LocARNA support the prediction of RNA–RNA interaction, exact RNA matching and alignment of RNA, respectively. The Freiburg RNA tools web server and the software packages of the stand-alone tools are freely accessible at http://rna.informatik.uni-freiburg.de.

INTRODUCTION
During the last decade, the discovery of a multitude of regulatory and catalytic RNA molecules has attracted attention to RNA in biological research (1,2). Many non-coding RNAs (ncRNAs) require a specific structure to perform their functions or interact via structural base pairing (3–6). Thus, RNA analysis demands tools that take into account RNA structure. The Freiburg RNA tools web server gives access to tools for three advanced RNA analysis tasks via an integrated, easy to use interface that supports the combination of these tools.

The server offers the prediction of RNA–RNA interaction (INTARNA), exact RNA matching (EXPARNA), and the multiple alignment of RNA (LocARNA). All tools are recently developed, continuously maintained, highly accurate and among the best of their class (7–9). Consequently, the tools have been used in recent studies (10–12).

Performing complex analysis tasks, the server complements available web servers for RNA analysis such as the Vienna RNA web suite (13) and the mfold web server (14). Among other services, the Vienna RNA web suite gives access to an older version of LocARNA. In this contribution, we offer increased functionality with improved performance.

ABSTRACT
The central purpose of the web server is to provide RNA analysis tools that have been developed by the Freiburg Bioinformatics Group. To this end, the web server integrates three tools for different analysis tasks in a common framework.

Each tool accepts a set of sequences in FASTA format as its main input. These sequences can be either entered directly or uploaded. The tools allow specific extension of the FASTA format, such as the annotation of sequences by secondary structure in dot–bracket format.

Furthermore, each input page provides program-specific options, with reasonable default settings, in order that the user can configure the respective tool to their needs. For user convenience, the server distinguishes between basic options and advanced parameters. The latter are hidden by default and can be unfolded on demand. In this way the server provides broad flexibility without confusing the less experienced user. Input is validated and the user is informed of inconsistencies as early as possible. For each task, we provide example input for demonstration purposes. Online help is provided and describes each tool, its input, available options and output.

Tool output and method are specific to each task and are, therefore, described separately. Where possible the output is illustrated by graphical presentation of the results. Figures are displayed in the browser as pngs and offered for download in postscript and pdf format.

Each analysis task is processed following a general scheme: jobs are scheduled to a computing cluster in order that jobs can be computed in parallel and resources flexibly adapted to the server load. Currently, we reserve eight cores for parallel computation. After submission, the current status of the job is reported and the user receives a URL allowing access to the job status or output. Upon job completion the result page is displayed online in the web browser.
Finally, the server provides links to the source code of the tools. The stand-alone command-line versions are more convenient and appropriate for large-scale studies; however, there are no input size restrictions by our web server. All tools use the widely accepted Turner free-energy model for RNA folding with standard energy parameters (15).

**INTARNA**

**INTARNA** is a tool for fast and accurate prediction of interactions between two RNA molecules (7). It has been designed to predict mRNA target sites for ncRNAs like eukaryotic microRNAs or bacterial small RNAs (sRNAs), but it can also be used to predict other RNA–RNA interactions.

**Input and output**

The input of **INTARNA** is a set of ncRNA sequences and a set of mRNA sequences. The output consists of a table that summarizes the results of the prediction and links to all predicted putative interactions between the ncRNAs and mRNAs. The output table can be sorted by columns to allow selection of interactions by sequence identifier, interaction energy score or interaction position in each sequence (Figure 1).

**Method**

**INTARNA** computes interactions by minimizing an interaction energy score via dynamic programming. The scoring is based on the hybridization free energy and accessibility of the interacting subsequences. The accessibility of an interaction site is defined as the free energy that is required to make the interaction site single-stranded. This is based on the thermodynamic ensemble of all secondary structures. Computation of these accessibilities is realized via the Vienna RNA library (16–18). It is assumed that ncRNAs fold globally and that mRNAs fold locally with a given maximal base pair distance. The algorithm runs in \( O(n^2) \) time and space when accessibilities are pre-computed.

Furthermore, **INTARNA** enables the inclusion of an interaction seed, i.e. an initial interaction region of (nearly) perfect sequence complementarity. The user has to specify the minimal number of perfectly paired bases and the maximal number of unpaired bases in the seed region. Other seed features, such as the seed position in the ncRNA, are optional.

In addition to the optimal solution according to the interaction energy score, **INTARNA** optionally reports also suboptimal interactions. The user can specify the maximal number of suboptimal predictions per sequence pair or restrict the reported interactions by an energy threshold.

**Prior application and evaluation**

**INTARNA** was validated on a data set of 18 experimentally verified sRNA–mRNA interactions, on which it achieved the highest accuracy, of all compared methods, in terms of sensitivity and positive predictive value (7). In a genome-wide target search, **INTARNA** showed the best prediction performance together with the comparable approach RNAUP (17), but with considerably lower computing time and memory requirement (7). Recently, **INTARNA** was applied to identify two novel mRNA targets of the cyanobacterial RNA Yfr1 (10).

**EXPARNA**

**EXPARNA** is a tool for very fast comparison of RNAs by exact local matches (8,20). Instead of computing a full sequence-structure alignment, **EXPARNA** efficiently

![Figure 1. Screenshot of the Intarna result page for a set of example sequences. The table summarizes all predicted interactions including one suboptimal interaction for mRNA seq3. It can be sorted by clicking on the header of a column. The interaction shown below the table is highlighted in green. For this prediction, additional information such as interaction positions and different contributions of the interaction energy score are given.](image-url)
computes the best arrangement of sequence–structure motifs common to two RNAs. This approach is beneficial for comparative sequence analysis in biology and in high-throughput RNA analysis tasks. E XPARNA elucidates information about identical structural motifs. This is not directly addressed by sequence–structure alignment tools and, therefore, may remain hidden. In addition, the predicted set of motifs can be used as anchor constraints to speed up and guide Sankoff-style alignment methods like LocARNA and related approaches that are in principle able to profit from alignment constraints.

Input and output

The input of E XPARNA is a pair of RNA sequences and secondary structures in dot–bracket notation using an extended FASTA format. If no secondary structure is available, the sequences are automatically folded by RNAFOLD (16). E XPARNA outputs the optimal set of exact pattern matches (EPMs) between the input RNAs. The result is presented graphically as coloured secondary structure plots (Figure 2a). Additionally, the web server allows the user to download results in different text file formats, e.g. as a structure annotated alignment or a list of (all) exact matches.

Method

E XPARNA performs a fast pre-processing step that determines the set of all possible EPMs for two given RNAs (21). An EPM is a local substructure that is identical in sequence and structure to both RNAs. EPMs are maximally extended and bond preserving, but the set of all EPMs contains overlapping and crossing EPMs. Therefore, in the next step E XPARNA computes the best set of non-crossing and non-overlapping EPMs, i.e. the longest collinear sequence of exact matching substructures for two RNAs. The dynamic programming algorithm runs in $O(H \cdot n^2)$ time and $O(n^2)$ space with $H \ll n^2$ for real RNA structures.

Prior application and evaluation

E XPARNA results agree well with existing alignment-based methods like RNAFORESTER (22), but results are obtained in a fraction of the compared run time. The performance of E XPARNA combined with LocARNA was evaluated on BRALiBASE 2.1 (23) and gave an overall speed-up of 4.25x with an alignment accuracy close to LocARNA alone (8).

Combining E XPARNA and LocARNA

E XPARNA’s exact matches can be beneficially used as anchor constraints for a full sequence–structure alignment. This allows the calculation of a constraint alignment by LocARNA, hence enabling alignment of very large RNAs that could not otherwise be aligned in reasonable time. This approach also maintains existing structural motifs in the resulting alignment (Figure 2b). This procedure is supported by the web server with a direct link from the E XPARNA results page to the LocARNA input page.

LocARNA

LocARNA is a tool for aligning multiple RNA sequences (9). It is one of the fastest and most accurate tools for this

Figure 2. (a) Annotated structures from the E XPARNA output. Regions of exact pattern matches (EPMs) have the same colour. Shown are two bacterial RNase P RNAs (left: A-type P RNA from Escherichia coli; right: B-type P RNA from Bacillus subtilis). Structures are taken from RNase P database (19). The numbers indicate four large EPMs. (b) Workflow for combining E XPARNA with LocARNA. First E XPARNA is called on the input RNAs to predict EPMs 1–4. This information is used as anchor constraints for a complete sequence–structure alignment by LocARNA. EPM numbers and colours are taken from (a).
purpose. Comparable with programs like DYNALIGN, FOLDALIGN and LARA (24–26), LOCARNA performs Sankoff-style simultaneous alignment and folding (27). Such programs generate high quality alignments that take structural similarity into account. Notably, the structural information is not required a priori but can be inferred, in parallel to the alignment process, based on an RNA free–energy model.

**Input and output**

The input of LOCARNA is a set of sequences in FASTA format. Optionally the input can be enriched by structural constraints and anchor constraints. These constraints are useful for guiding the automatic alignment using prior knowledge and/or speeding up the computation.

The output is a multiple alignment of these sequences optimized according to the LOCARNA score, which evaluates both sequence and structural similarity (Figure 3a). The output alignment optionally satisfies given constraints. When performing local alignment, LOCARNA will allow unaligned fragments at the beginning and end of all sequences without penalty. The LOCARNA alignment is shown together with the predicted structure, using the alignment as input, by RNAALIFOLD (28) (Figure 3b).

**Method**

LOCARNA computes pairwise alignments by dynamic programming. Multiple alignments are constructed from pairwise alignments using a progressive alignment strategy. LOCARNA achieves its low time and space complexity of $O(n^4)$ and $O(n^2)$, respectively, for pairwise alignment because it needs to consider only significant base pairs (9).

**Prior application and evaluation**

A prior LOCARNA version was validated by a re-clustering of Rfam (9). At high average recall, the Rfam families were reproduced with good precision. Furthermore, LOCARNA was benchmarked using BRA LiBASE 2.1 (23) for multiple alignment. It performed better than the comparable approaches FOLDALIGN and LARA (29).

**IMPLEMENTATION**

The Freiburg RNA tools web server is based on a general framework developed for the CPSP web tools server (30) and has been continuously improved. XHTML is served by Apache Tomcat v5.5.25 that supports the use of JavaServer Pages and Java Servlets consequently allowing a large deal of dynamically generated content to be provided.

JavaScripting is used to aid the user in providing well-formed input (sequences and parameters), which is then stored in a Java Bean and processed by a Java Servlet. Jobs with valid input parameters are submitted to a computing cluster managed by Sun Grid Engine and the user is directed to the results page to wait for job completion. Upon completion of a job, the user can access result details consisting of ‘system call’ parameters saved by the Java Bean, raw tool result data and post-processed result data (e.g. images generated from raw tool results by a Java Servlet). JavaScript, JSP and JavaServlet pages are used to increase accessibility of result features. The whole web server is run on a virtual machine hosted on a server running Scientific Linux.

**ACKNOWLEDGEMENTS**

We thank Dragoş Alexandru Sorescu, Martin Mann and Stefan Jankowski for their assistance in setting up the web server. We also thank all people who have been involved in testing the robustness and functionality of the final web server.

**FUNDING**

German Research Foundation (DFG) (BA 2168/2-1 to A.S.R. and R.B., BA 2168/3-1 to R.B., BA 2168/4-1 to R.B. and WI 3628/1-1 to S.B.); German Federal Ministry of Education and Research (BMBF) (0313921 to R.B.)
and S.H.). Funding for open access charge: University of Freiburg.

Conflict of interest statement. None declared.

REFERENCES