Published in final edited form as:

Methods Mol Biol. 2011 ; 714: 447–466. doi:10.1007/978-1-61779-005-8_27.

Identifying and searching for conserved RNA localization signals

Russell S. Hamilton and Ilan Davis

Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU

Abstract

RNA localization is an important mode of delivering proteins to their site of function. *Cis*-acting signals within the RNAs, which can be thought of as zip-codes, determine the site of localization. There are few examples of fully characterised RNA signals, but the signals are thought to be defined through a combination of primary, secondary and tertiary structure. In this chapter we describe a selection of computational methods for predicting RNA secondary structure, identifying localization signals and searching for similar localization signals on a genome wide scale. The chapter is aimed at the biologist rather than presenting the details of each of the individual methods.

Keywords

Bioinformatics; mRNA localization; RNA secondary structure

1. Introduction

In the past few decades, RNA has taken an increasingly central position beside DNA in biology. Indeed, from a catalytic and structural point of view, RNA is a much more interesting molecule than DNA. RNA plays key roles in all aspects of biology, ranging from transcription of genes into nascent RNA and then mRNA that is often localized to particular parts of the cytoplasm. In a screen of 3370 RNAs in the *Drosophila* embryo using fluorescent *in situ* hybridisation (FISH) it has been shown that 71% show some level of localization (1). RNA localization in the nervous system is of particular functional importance (2, 3), as defects in RNA localization have been implicated in Fragile-X syndrome (4), spinal muscular atrophy (5) and spinocerebellar ataxia (6). Therefore, the bioinformatics of RNA primary, secondary and tertiary structures is of great importance to the understanding of the wide variety of functions that RNA can perform.

Intracellular mRNA localization is a very common means of delivering proteins to their site of function, through targeted translation at the site of localization. This mode of post-transcriptional regulation of gene expression has been identified in most of the model organisms and is known to be very common in *Drosophila*, where it has been tested systematically (7, 8). There are several mechanisms by which mRNAs localize including diffusion, localized synthesis or degradation and active transport (9). Here we discuss mRNAs that localize by the most common of these mechanisms, actively transported mRNAs by molecular motors. Localized transcripts contain discrete *cis*-acting signals that fold and associate with a variety of *trans*-acting proteins that dictate their intracellular destinations. Amongst other proteins, the *trans*-acting factors associate with molecular

motors, anchoring complexes and translational regulatory proteins (Figure 1A). The molecular motors with their RNA cargoes then travel to their destination along microtubule or actin filaments. In general, the RNA signals to which these proteins associate are best thought of as being identified through a combination of primary sequence and structural features. However, the localization signals themselves have been defined in detail in only very few examples. In most cases, little is known about the primary sequence or structural features that direct the mRNA to a particular part of the cell. Very few localization signals have been mapped to the minimally required motif so there is little known about what confers localization of an mRNA to a particular location in the cell. The *gurken* and *I* factor have been mapped to their minimally required localization signals, however the exact sequence and structure requirements determining localization have yet to be elucidated (10). Other Dynein mediated localized mRNAs to have their signals mapped include: *bicoid* (11), *oskar* (12, 13), *fs(1)K10* (14), *orb* (15), *nanos* (16), *eve* (17), *hairy* (18) and *wingless* (19).

mRNAs contain primary, secondary and tertiary structures. The primary sequences are comprised of four nucleotides, the purines adenosine (A) and guanine (G) and the pyrimidines cytosine (C) and uracil (U). RNA secondary structure consists of base pairings between the nucleotides, which can be canonical (A:U, U:A, G:C, C:G) or non-canonical (e.g. G:U)(20-22). The base-pairings then give rise to features such as stems, loops, bulges, internal loops and pseudoknots (Figure 1B). Bases important for maintaining mRNA secondary structure are under evolutionary pressure so are usually well conserved between homologous sequences. Base pairings can be maintained through compensatory mutations, where the base pair is in the same position but the sequence differs (AU to GC). Consistent mutations are where a base pair is maintained, but by a weaker non-canonical interaction (GC to GU) or *vice-versa*. The tertiary structure of RNA contain all the secondary structure features organised in three dimensions, such as stem regions that come together to form helices in three dimensions. Full length RNA structures can be divided into structural motifs, which are important for the function of the mRNA and can be defined at the primary, secondary and/or tertiary level (21, 23). The most interesting of these motifs are those that are conserved across different species, as they are likely candidates for structural features that provide the specificity for biological function, including determining the site of localization of the RNA.

In this chapter, we focus on the computational tools that are used for the secondary structure prediction, alignment and identification of RNA signals involved in intracellular mRNA localization. However, the same tools are generally applicable to the analysis of RNA signals in any kind of biological context. The aim of this chapter is to survey the array of tools that are currently available for searching, aligning and characterising RNA localization signals. Most of the papers describing the methods are aimed at computational biologists, and there is a distinct lack of reviews that are accessible to biologists with a non-computational background. Therefore, this review concentrates on intuitive explanations of the different tools and their application, and where available are illustrated with examples of their usage in the mRNA localization field.

2. Materials

2.1 Computer environments

As most command line tools are released as open source, this tends to favour the Linux and Mac OSX operating systems. The command line versions of the programs usually provide the most flexibility for use with a full range of customisable options. However many of the tools are available as web servers and so can be accessed from any computer with a web browser and internet access. A further issue is the computational complexity of many of the algorithms requiring access to high performance computers (HPC). For example, a typical

desktop would be unable to perform Infernal (24) searches across a genome within a practical time frame, so an HPC cluster would be more appropriate. However, in most universities there are HPC facilities available for use by academic researchers, which also provide help on how to use the resources.

2.2 Software

In addition to a web browser for using the online versions of the tools, it is recommended to use a text editor. Most of the tools described in this chapter require the input RNA sequences to be in plain text rather than the default document formats of the common word processor programs. Sequences can be saved in plain text with most word processing programs, however there are more suitable programs for editing sequence data. For example, the RALEE (RNA Alignment Editor in Emacs) text editor is designed for editing RNA secondary structure alignments on multiple platforms (Linux, Mac OS and Windows)(25) (Table 1). The colour highlighting of secondary structure elements on each of the sequences makes alignment editing a much more straight-forward task. There are a variety of file formats for RNAs and their secondary structures, the most common of which are FASTA, Stockholm and connect (CT) (Figure 1C).

3. Methods

Unless an RNA localization signal is solely defined at the primary sequence level then methods more sophisticated than **BLAST** or **MEME** must be employed (26). If the RNAs are distantly related then it is necessary to also take into account the secondary structure. The accuracy of secondary structure prediction from single sequences can be improved by utilising covariance information from homologous sequences. The sequences must contain enough variation to show the covariance so that selection of sequences for prediction is crucial. This is of particular importance in the single stranded regions of the RNA secondary structures as this is where specific RNA:protein interactions are likely to occur. The covariance between sequences is calculated from alignments of the homologous RNAs. The alignments can be achieved by several methods. First, structures can be aligned from “single sequence secondary structure prediction” methods (e.g. **MFOLD**). Second, “align then fold” methods can be used. Or third, “simultaneous folding and alignment” methods provide another alternative. Each of these methods are outlined in the sections below, followed by methods involving RNA tertiary structure prediction and searching.

3.1 Single sequence secondary structure prediction

The prediction of secondary structure from sequence alone has been developed around maximising the number of the base pairings in the structure. The structures are most commonly scored by the minimum free energy (MFE) giving the thermodynamic stability of the overall secondary structures (27). Table 1 has a full list of the methods outlined in this section. **MFOLD/UNAFold** (28, 29) and **RNAfold** (30) are derived from the same Zuker-Steigler algorithm (31) for calculating the MFE structures for single RNA sequences, so therefore produce comparable results. A dynamic programming algorithm utilises experimentally determined energy parameters to calculate secondary structures with the lowest free energies and an RNA secondary structure prediction algorithm based on probabilistic models (32). **contraFOLD** improves on the accuracy of both thermodynamic and stochastic context free grammars (SCFG) based algorithms (See Notes 2). **Sfold** assigns probabilities to representative secondary structure motifs from a Boltzmann probability distribution. The motifs are then clustered by structural similarity into ensembles of similar structures (33). The MFE structure is not guaranteed to be included in the representative set but can be added manually during the analysis. **Sfold** is particularly useful in determining if protein binding sites are accessible throughout the ensemble of structures. **Pfold** predicts the

secondary structure of RNAs using SCFGs from an alignment of related RNA sequences (34). **RNAKinetics** calculates the kinetics of folding for a growing RNA sequence thus taking into account that as RNA molecules are produced they form dynamic rather than static structures (35). The result is an ensemble of structures assigned with probabilities. **Vienna+P** is a modification of the **RNAVienna** package, and models the secondary structure of RNAs in the presence of single stranded RNA binding proteins (30, 36). The method applies general models for RNA binding proteins and calculates the probability of a nucleotide binding each of the nucleotides in the RNA. This method does not make use of biological data such as known RNA binding motifs or the number of nucleotides typically involved in RNA:protein interactions. However the binding of proteins is very likely to alter the secondary structure *in vivo*, so represents very important progress in RNA secondary structure prediction. The method also calculates FRET values for fluorophores bound to end nucleotides in RNAs.

There have been several studies assessing the accuracy of RNA secondary structure approaches using experimentally derived structures. Mathews found **Mfold** to have an accuracy of 73% for a large database of 700 nt or less sequences (37). Doshi *et al* found an accuracy of 41% for a reference set of 16S and 23S RNA (38). Dowell and Eddy found 56% accuracy for **Mfold**, 55% for **RNAfold**, 50% for **Pknots**, and 39% for **Pfold** for a reference set of RNase P, SRP and tmRNAs (39). There is also a database of RNA structure alignments, **BRALIBASE**, for benchmarking RNA structure prediction methods (40, 41) (Table 1). Single sequence RNA secondary structure prediction is always the most appropriate for initial assessment of the RNA of interest. Once the secondary structure has been determined this guides the choice of the next methods to employ (See Figure 2).

3.2 Pseudoknot secondary structure prediction

Pseudoknots provide a challenge in the field of RNA secondary structure prediction due to the added level of complexity (Figure 1B). The Fragile-X mental retardation protein (FMRP) is a RNA binding protein (KH and RGG domains) and has been proposed as a component of mRNP transport granules (Reviewed in (42)). FMRP was recently shown to specifically target mRNA with “kissing-loop” pseudoknots. The loss of FMRP function in Fragile-X syndrome is likely to be due to the inability of the kissing-loop mRNAs to bind FMRP (43). **Pknots** uses a dynamic programming algorithm and thermodynamic energy models for canonical base pairings and ones specific to pseudoknots (44). **Hotknots** uses a heuristic algorithm to predict non-pseudoknot substructures, which are built up to form a series of secondary structure predictions (45). **Hotknots** utilises an extended version of the standard thermodynamic parameters used in MFE predictions to assess the addition of pseudoknots to the series of predicted structures. The iterative loop matching (**ILM**) approach first predicts stems using a standard non-pseudoknot MFE method and then iteratively predicts further helices (46). The algorithm can either predict the secondary structure of single sequences or using comparative information in a set of aligned homologous sequences. See Table 1 for a list of the methods outlined in this section.

3.3 Measuring secondary structure similarity

To determine if two RNAs share a common secondary structure or structural motif it is possible to calculate a similarity measure between structures (Figure 2). Although it is trivial to calculate sequence identity to assess the similarity of two sequences, the assessment of the similarity or dissimilarity of RNA secondary structures is more complex. **RNAdistance** encodes RNA secondary structure as tree representations and then performs a global comparison and scores the dissimilarity of the two structures (30). **RNAforester** also encodes structures as trees, but compares structural similarity on a local scale, the advantage being that the algorithm does not get trapped trying to find similarity across the entire length

of the sequences being compared (47)(See Notes 1 and Table 1). **RNAforester** was recently used in a study of the *internal transcribed spacer 2* (ITS2) gene which is used as a marker for the classification of new species. ITS2 is a fast evolving gene, but contains conserved secondary structures. The core of the RNA secondary structure was shown to contain a large stem-loop structure, made up from four smaller stems and is conserved across all eukaryotes (48, 49).

3.4 Obtaining homologous sequences

In order to improve on the secondary structure prediction of the single sequence methods, it is necessary to make an alignment of homologous sequences (Figure 2). The alignment can then exploit the covariance of the sequences to determine the nucleotides important for conserving secondary structure. **BLAST** performs pairwise local sequence alignments, where seed alignments are made against a sequence database and are lengthened upon matching (50). BLAST returns a list of sequence hits matching the query sequence ranked in order of their statistical significance. Probabilistic models can be used to search for homologous sequences by finding similarity in both sequence and secondary structure. **Infernal** utilises covariance models, to describe both the sequence and the secondary structure of the RNAs (24)(See Notes 2). **Infernal** is used in the **Rfam** project to annotate the RNAs (51). **Rfam** is a database for RNA families which are stored as sequence alignments and covariance models (51). Seed alignments annotated with a consensus secondary structure are created for each family, which are used to build covariance models. **RNApromo** utilises SCFGs (See Notes 2) to encode an experimentally determined or MFE predicted secondary structure to search for local structure motifs (52). The method was recently applied to the Fly-FISH database of localization patterns in *Drosophila* embryos and found motifs in 9 of the 94 sets of colocalized mRNAs (1, 52). These covariance models are then used to retrieve further members of the RNA family. **UTRdb** contains the mRNA sequences of the untranslated regions (UTRs) of eukaryotic genomes (53). RNA localization signals are most commonly, but not exclusively found in the UTRs of the mRNAs as the coding regions are conserved for proteins that they encode (8). Therefore, well conserved secondary structures in UTRs are likely indicators of functional importance. The Nucleic acid database (**NDB**) contains the three dimensional coordinates for 4406 deposited RNA structures (54). The **Ensembl** project is an integrated resource for the genome annotations and sequence databases for model organisms (55). *Drosophila* is a model organism with a wealth of collated information. One major advantage of using *Drosophila* is the availability of the genome sequences for **12 *Drosophilids***. This allows the covariance information for sequences to be observed by searching across all 12 genomes (56). See Table 1 for a full list of the resources outlined in this section.

3.5 Searching for motifs – sequence and structure

RNA sequences involved in mRNA localization contain motifs which bind the proteins involved in their translational regulation, transport and anchoring. The following methods are applicable to searching for motifs at the primary and secondary structure level (See Figure 2 and Table 1). **REPFIND** finds repeats clustered in nucleotide sequences and has been used to find such repeats in *Vg1* mRNA, which is localized to the vegetal cortex of oocytes in *Xenopus laevis* (57). The *Vg1* localization signal has been mapped to 340 nt and contains four repeat motifs (E1, E2, E3, E4) which abolish or reduce localization if deleted (58, 59). The E2 motif, when present as two copies, has been shown to be necessary for localization and a sequence motif of UUUCAC was found to be crucial (60). The E2 motif was also found to be present in *VegT*, also localized to the vegetal cortex in *Xenopus* oocytes. The E2 motif occurs as five copies in the *VegT* 3'UTR and was also found to be sufficient for localization (61). **REPFIND** was developed to search for other RNAs containing the E2 motif, specifically the CAC repeat. The same searches were not possible

using BLAST (50) due to the short repeat sequences. The CAC motif was found to be present in all the RNAs localizing to the vegetal cortex and is conserved across all chordates (57). **MEMERIS** searches for sequence motifs in the context of single stranded regions and is an extension of the **MEME** algorithm for sequence only motif searching (62, 63). The rationale for the searches in single stranded regions is that some RNA-binding proteins are known to target these single stranded regions. For example, RNA-recognition motifs (RRM) and K-homology (KH) domains specifically bind RNA via approximately four nucleotides in single stranded regions (64). **RNAalifold** predicts consensus secondary structures consistent with a set of aligned homologous RNA sequences (65). **RNAalifold** was recently improved with the implementation of the **RIBOSUM** (66) substitution matrices and improved gap scoring in the alignments. **RNAalifold** was used to identify possible binding sites in the GRIK4 3'UTR, which has been implemented in bipolar disorder (67).

RNAmotif is an RNA motif search algorithm able to search at the primary, secondary and tertiary structure level (68). The algorithm allows any base interaction, including non-canonical base pairings to be defined. A recursive algorithm then searches the motif descriptor against a set of sequences to find matches. **RNAmotif** also provides the option of defining a scoring system for the matches, which can also be used to assess the similarity between sequences. **RNAfold** predicts locally stable RNA secondary structures from within long sequences such as whole genomes (69). The predicted stable structures are calculated for a user defined length by folding the entire sequence in overlapping windows of the defined length (See Notes 1). **RNAfold** was utilised as part of a pipeline, **RNA2DSearch**, for the identification of localization signals in transposable elements similar to *gurken* and the *I* factor in *Drosophila*. The transposons, *G2* and *Jockey* were identified in the search and were found to localize in a similar manner (70). **ERPIN** identifies conserved RNA motifs in sets of unaligned sequences annotated with an RNA secondary structure (71). Each single stranded or double stranded region is given a profile score, and a dynamic programming algorithm searches for the occurrence of these regions in databases of sequences. The advantage of this method is the ability to search for less rigorously defined regions and pseudoknots. The myelin basic protein (MBP) mRNA is localized to the myelin compartment of oligodendrocytes via an 11 nucleotide hnRNP A2 response element (A2RE) (72). The A2RE is necessary and sufficient for localization and is present in other mRNAs localizing to the myelin compartment (α CaMKII, neurogranin, ARC, Gag, Vpr, and PKM ζ) (73). However a simple sequence search for the A2RE reveals many instances for the A2RE that do not localize. In a recent study, it was found that the A2RE motif contains a tertiary structure kink-turn motif where a bulge in a stem presents the single stranded bulge nucleotides for protein binding (74). The study used a modified version of **DNAMAN**, however **RNAmotif** would also be able to search for kink-turn motifs with both sequence and secondary structure definitions. An investigation into the localization signal common to the *Drosophila* genes, *K10* and *Orb* used **RNABob** to show that the signal is not found in other *Drosophila* genes when defined at the secondary structure level (75).

3.6 Align then fold methods

The align then fold methods rely on accurate multiple sequence alignments and a recent evaluation of various alignment methods by Gardner *et al* revealed that the **Clustal** multiple sequence alignment program consistently performs well for sets of sequences above 55% sequence identity (41, 76). Below the 50-60% sequence identity “twilight zone” then algorithms using the simultaneous folding and alignment method, incorporating structural information, are required to achieve accurate alignments (See Figure 2). The **QRNA** non-coding gene finding algorithm uses a comparison of regions of two aligned sequences containing synonymous substitutions, indicating a coding region, against compensatory mutations that test for the presence of stems in a conserved RNA secondary structure (77). **QRNA** utilises hidden Markov models where the input is two aligned sequences (See Notes

2). **RNAz** finds thermodynamically stable and conserved RNA secondary structure motifs in sets of aligned sequences (78). A structure conservation index is calculated through the use of a support vector machine (SVM) and the method is suitable for scanning on a genome wide scale. **CMFinder** finds motifs in sets of unaligned RNA sequences using covariance models (79). One key advantage of **CMFinder** is that the method is applicable to unrelated sequences. Most align-then-fold algorithms assume the input sequences are evolutionarily related. **PETfold** predicts the structure of aligned RNA sequences by combining evolutionary and thermodynamic information (80). **PETfold** is an extension of the conserved base pair identification of **Pfold** (34) and adds thermodynamic energy parameters (See Table 1).

3.7 Simultaneous folding and alignment methods

These methods utilise the Sankoff algorithm (81) and are applicable when the sequence identity of the set of sequences is below the “twilight zone” of 55-60% identity (See Figure 2). **FOLDALIGN** simultaneously aligns and predicts the secondary structure of two RNA sequences, then scans the sequences for common local structural motifs (82). The algorithm can also be configured to align the sequences locally or globally across the entire length of the sequences. There is also a multiple sequence alignment version called **FOLDALIGNM**. **Dynalign** aligns and predicts the MFE secondary structure of two unaligned sequences and the predictions can be constrained to experimentally characterised base pairings (83). **CARNAC** predicts secondary structures common to a set of homologous sequences, calculating the structures of all the sequences as stable stems which are then assessed to whether they are common to all the sequences (84). The computational demands of the simultaneous folding and alignment algorithms mean that they should only be realistically considered with alignments below the “twilight zone”, as the speed and accuracy of **Clustal** are sufficient for more similar sequences. See Table 1 for a full list of the methods outlined in this section.

3.8 3D structure searching and prediction

Despite the advances in sequence searching and secondary structure prediction algorithms they fail to take the three dimensional structure of the RNAs into account. Methods predicting and searching for three dimensional RNA localization signals incorporate primary, secondary and tertiary structure information and will provide the most insight into the parts of the motif determining localization (See Figure 2). The searching for structural motifs has been successfully applied in *Saccharomyces cerevisiae* where *ASH1* mRNA is localized to the bud tip of daughter cells and binds the localization machinery via the RNA-binding protein She2p. There are four known localization signals, three from the coding region and one from the 3'UTR (85-87), each forming a stem loop. Olivier *et al* identified a consensus of two cytosine bases within the stem-loops separated by six other nucleotides corresponding to approximately 28Å (See Figure 1B). Furthermore, they used **MC-SEARCH** and **RNA motif** (68), and were able to search through a database of RNA 3D structures to find further examples containing the motif. Two of which were found to localize the bud tip in the same manner as *ASH1* mRNA (88). The computational prediction of RNA tertiary structure is a key goal in the study of RNA localization signals. The experimental determination of RNA tertiary structure by NMR and X-ray crystallography can be an expensive and time consuming affair. The **MC-FOLD/MC-SYM** pipeline predicts RNA tertiary structure from sequence (89) predicting all possible “nucleotide cyclic motifs” describing the basic building blocks of the secondary structure. These building blocks are then matched to a database of experimentally determined RNA three dimensional structures, also processed into these building blocks. A predicted three dimensional structure is then built up using the **MC-SYM** algorithm. In a recent study comparing the NMR structure determination and molecular dynamics (MD) simulation of an RNA tetraloop

(UUUU). The conclusion of the study was that the MD simulation and force-fields used in the MD were in good agreement with the MR structure (90). Although the tetraloop is a fairly simple example of RNA tertiary structure prediction the force-fields utilised in the MD are likely to be applicable to larger RNA structures such as stem loops involved in RNA localization signals (See Table 1).

3.9 Study the RNA-binding proteins

Most of the methods described in this chapter do not take into account the proteins that bind the RNA signals. By studying known binding partners of the RNAs, it is possible to deduce a vast wealth of information about the likely RNA signals. For example whether proteins binds single or double stranded RNA and whether the interactions are specific or non-specific. This is particularly relevant to the development of biochemical pull-down assays to identify proteins binding the RNA signals (91). In a recent study of the prediction of protein-RNA interactions by Shulman-Peleg *et al*, it was observed that the nucleotides interacting with the protein are formed by at least two consecutive bases (92). In addition, over 30% of the interacting bases are π -stacked, allowing the bases in hairpin loops to be available for aromatic interactions while also increasing the stability of the loop. The Amino Acid–Nucleotide inTeraction database (AANT), is a resource for searching and visualising experimentally determined interactions (93). The MS2 system for visualising native mRNAs *in vivo*, was originally developed by Bertrand *et al* in *Saccaromyces cerevisiae* to label *ASH1* mRNA (94) and has since been used to study many other transcripts in a variety of models, including, *nanos*, *bicoid*, *oskar* and *gurken* mRNA localization in *Drosophila* (95, 96). mRNAs are engineered to contain multiple copies of the MS2 loop, which then bind an MS2-coat protein fused to green fluorescent protein (GFP). The MS2-coat protein specifically binds the MS2 loops in the RNA, and is able to discriminate against the similarly structured Q β loops. The Q β -coat protein, shares a low sequence identity to the MS2-coat protein (~20%), however the tertiary structures of the binding domains are very similar; both are alpha/beta 2-layered sandwiches (97). A study by Horn *et al*, investigated the structural basis for the specific binding of MS2 and Q β coat proteins to their respective RNA loops (98). It was shown that the MS2-coat protein can be engineered to specifically bind the Q β loops by mutating just two key residues (Asn87 and Asp91). It was also shown that the Q β -coat protein could specifically bind the MS2 loops by mutating the equivalent residues. This study also gave some insight into which parts of the RNA secondary and tertiary structures are important for specific binding. In the Q β loops, mutational analysis shows that the last A in the hairpin loop and the bulged A are the only bases that are important for binding affinity when mutated.

3.10 Closing remarks

There are now a bewildering and comprehensive range of routines developed and used to search and align RNA structure at all levels. The choice of routine for a particular application is not always straight forward. To some extent the applications of some of the routines describe here are overlapping. Nevertheless, it is important to enter into a detailed and continuous dialogue between the biologists and bioinformaticians. The most important issues to include in this dialogue is to consider examples of previous applications of the rivaling method and the available bioinformatics resources including staff time and computer infrastructure.

4. Notes

4.1 Importance of local vs global secondary structure

Within cells, RNAs are always extensively decorated with proteins, so global secondary structure prediction is unlikely to represent the full picture for predicting protein binding. It

is therefore important to study the locally stable secondary structure predictions for the RNA of interest. Long *et al* proposed a two step binding of mRNAs to their targets where a short region of RNA sequence with a key secondary structure is first bound, followed by a second step of secondary structure relaxation and binding elongation (99). Furthermore, Heale *et al* showed that the target substructures of target RNAs are also important for the interaction with siRNA in mammalian cells (100).

4.2 What are covariance models (CMs) and Stochastic Context Free Grammars (SCFGs)

A covariance model is a probabilistic model simultaneously describing both the primary and secondary structure of an RNA. The models are analogous to hidden Markov models (HMMs) used for sequence only alignment and searching (101). Covariance models are a form of stochastic context free grammars (102, 103). Where the nucleotides and base-pairings are described in the form of probabilities. By aligning homologous sequences to the original RNA sequence it is then possible to determine the probabilities of insertions, deletions and nucleotide changes. These probabilities, for sequence and secondary structure, are encoded in a model. Covariance models are used in secondary structure prediction, multiple sequence and structure alignment and similarity searching against sequence databases.

Acknowledgments

RSH and ID are supported through a Wellcome Trust Senior Research Fellowship (081858) to ID.

References

1. Lecuyer E, Yoshida H, Parthasarathy N, et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell*. 2007; 131:174–187. [PubMed: 17923096]
2. Dahm R, Kiebler M, Macchi P. RNA localisation in the nervous system. *Semin Cell Dev Biol*. 2007; 18:216–223. [PubMed: 17331761]
3. Hengst U, Jaffrey SR. Function and translational regulation of mRNA in developing axons. *Semin Cell Dev Biol*. 2007; 18:209–215. [PubMed: 17478114]
4. Jin P, Warren ST. New insights into fragile X syndrome: from molecules to neurobehaviors. *Trends in Biochemical Sciences*. 2003; 28:152–158. [PubMed: 12633995]
5. Paushkin S, Gubitz AK, Massenet S, Dreyfuss G. The SMN complex, an assemblyosome of ribonucleoproteins. *Current Opinion in Cell Biology*. 2002; 14:305–312. [PubMed: 12067652]
6. Mutsuddi M, Marshall CM, Benzow KA, Koob MD, Rebay I. The Spinocerebellar Ataxia 8 Noncoding RNA Causes Neurodegeneration and Associates with Stau1 in Drosophila. *Current Biology*. 2004; 14:302–308. [PubMed: 14972680]
7. St Johnston D. Moving messages: the intracellular localization of mRNAs. *Nat Rev Mol Cell Biol*. 2005; 6:363–375. [PubMed: 15852043]
8. Palacios IM, St Johnston D. Getting the message across: the intracellular localization of mRNAs in higher eukaryotes. *Annu Rev Cell Dev Biol*. 2001; 17:569–614. [PubMed: 11687499]
9. Palacios IM. How does an mRNA find its way? Intracellular localisation of transcripts. *Semin Cell Dev Biol*. 2007; 18:163–170. [PubMed: 17368060]
10. Van De Bor V, Hartswood E, Jones C, Finnegan D, Davis I. gurun and the I factor retrotransposon RNAs share common localization signals and machinery. *Dev Cell*. 2005; 9:51–62. [PubMed: 15992540]
11. Gavis ER, Lunsford L, Bergsten SE, Lehmann R. A conserved 90 nucleotide element mediates translational repression of nanos RNA. *Development*. 1996; 122:2791–2800. [PubMed: 8787753]
12. Kim-Ha J, Webster PJ, Smith JL, Macdonald PM. Multiple RNA regulatory elements mediate distinct steps in localization of oskar mRNA. *Development*. 1993; 119:169–178. [PubMed: 8275853]

13. Munro TP, Kwon S, Schnapp BJ, St Johnston D. A repeated IMP-binding motif controls oskar mRNA translation and anchoring independently of *Drosophila melanogaster* IMP. *J Cell Biol.* 2006; 172:577–588. [PubMed: 16476777]
14. Macdonald PM, Kerr K, Smith JL, Leask A. RNA regulatory element BLE1 directs the early steps of bicoid mRNA localization. *Development.* 1993; 118:1233–1243. [PubMed: 8269850]
15. Macdonald PM, Struhl G. cis-acting sequences responsible for anterior localization of bicoid mRNA in *Drosophila* embryos. *Nature.* 1988; 336:595–598. [PubMed: 3143913]
16. Serano TL, Cohen RS. A small predicted stem-loop structure mediates oocyte localization of *Drosophila* K10 mRNA. *Development.* 1995; 121:3809–3818. [PubMed: 8582290]
17. Davis I, Ish-Horowicz D. Apical localization of pair-rule transcripts requires 3' sequences and limits protein diffusion in the *Drosophila* blastoderm embryo. *Cell.* 1991; 67:927–940. [PubMed: 1959136]
18. Bullock SL, Zicha D, Ish-Horowicz D. The *Drosophila* hairy RNA localization signal modulates the kinetics of cytoplasmic mRNA transport. *Embo J.* 2003; 22:2484–2494. [PubMed: 12743042]
19. dos Santos G, Simmonds AJ, Krause HM. A stem-loop structure in the wingless transcript defines a consensus motif for apical RNA transport. *Development.* 2008; 135:133–143. [PubMed: 18045835]
20. Stombaugh J, Zirbel CL, Westhof E, Leontis NB. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* 2009; 37:2294–2312. [PubMed: 19240142]
21. Leontis NB, Lescoute A, Westhof E. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol.* 2006; 16:279–287. [PubMed: 16713707]
22. Lemieux S, Major F. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.* 2002; 30:4250–4263. [PubMed: 12364604]
23. Holbrook SR. RNA structure: the long and the short of it. *Curr Opin Struct Biol.* 2005; 15:302–308. [PubMed: 15963891]
24. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: inference of RNA alignments. *Bioinformatics.* 2009; 25:1335–1337. [PubMed: 19307242]
25. Griffiths-Jones S. RALEE--RNA ALignment editor in Emacs. *Bioinformatics.* 2005; 21:257–259. [PubMed: 15377506]
26. Hamilton RS, Davis I. RNA localization signals: deciphering the message with bioinformatics. *Semin Cell Dev Biol.* 2007; 18:178–185. [PubMed: 17452113]
27. Eddy SR. How do RNA folding algorithms work? *Nat Biotechnol.* 2004; 22:1457–1458. [PubMed: 15529172]
28. Zuker M. Calculating nucleic acid secondary structure. *Curr Opin Struct Biol.* 2000; 10:303–310. [PubMed: 10851192]
29. Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol.* 2008; 453:3–31. [PubMed: 18712296]
30. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res.* 2003; 31:3429–3431. [PubMed: 12824340]
31. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 1981; 9:133–148. [PubMed: 6163133]
32. Do CB, Woods DA, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics.* 2006; 22:e90–98. [PubMed: 16873527]
33. Ding Y, Lawrence CE. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.* 2003; 31:7280–7301. [PubMed: 14654704]
34. Knudsen B, Hein J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.* 2003; 31:3423–3428. [PubMed: 12824339]
35. Danilova LV, Pervouchine DD, Favorov AV, Mironov AA. RNAKinetics: a web server that models secondary structure kinetics of an elongating RNA. *J Bioinform Comput Biol.* 2006; 4:589–596. [PubMed: 16819804]
36. Forties RA, Bundschuh R. Modeling the interplay of single-stranded binding proteins and nucleic acid secondary structure. *Bioinformatics.* 2009

37. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*. 2004; 101:7287–7292. [PubMed: 15123812]
38. Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics*. 2004; 5:105. [PubMed: 15296519]
39. Dowell RD, Eddy SR. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics*. 2004; 5:71. [PubMed: 15180907]
40. Gardner PP, Giegerich R. A comprehensive comparison of comparative RNA structure prediction approaches. *BMC Bioinformatics*. 2004; 5:140. [PubMed: 15458580]
41. Gardner PP, Wilm A, Washietl S. A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Res*. 2005; 33:2433–2439. [PubMed: 15860779]
42. Bassell GJ, Warren ST. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron*. 2008; 60:201–214. [PubMed: 18957214]
43. Darnell JC, Fraser CE, Mostovetsky O, et al. Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. *Genes Dev*. 2005; 19:903–918. [PubMed: 15805463]
44. Rivas E, Eddy SR. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J Mol Biol*. 1999; 285:2053–2068. [PubMed: 9925784]
45. Ren J, Rastegari B, Condon A, Hoos HH. HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *Rna*. 2005; 11:1494–1504. [PubMed: 16199760]
46. Ruan J, Stormo GD, Zhang W. An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Bioinformatics*. 2004; 20:58–66. [PubMed: 14693809]
47. Hochsmann M, Toller T, Giegerich R, Kurtz S. Local similarity in RNA secondary structures. *Proc IEEE Comput Soc Bioinform Conf*. 2003; 2:159–168. [PubMed: 16452790]
48. Schultz J, Maisel S, Gerlach D, Muller T, Wolf M. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *Rna*. 2005; 11:361–364. [PubMed: 15769870]
49. Schultz J, Muller T, Achtziger M, Seibel PN, Dandekar T, Wolf M. The internal transcribed spacer 2 database--a web server for (not only) low level phylogenetic analyses. *Nucleic Acids Res*. 2006; 34:W704–707. [PubMed: 16845103]
50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–410. [PubMed: 2231712]
51. Gardner PP, Daub J, Tate JG, et al. Rfam: updates to the RNA families database. *Nucleic Acids Res*. 2009; 37:D136–140. [PubMed: 18953034]
52. Rabani M, Kertesz M, Segal E. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proc Natl Acad Sci U S A*. 2008; 105:14885–14890. [PubMed: 18815376]
53. Mignone F, Grillo G, Licciulli F, et al. UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res*. 2005; 33:D141–146. [PubMed: 15608165]
54. Berman HM, Olson WK, Beveridge DL, et al. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys J*. 1992; 63:751–759. [PubMed: 1384741]
55. Hubbard TJP, Aken BL, Ayling S, et al. Ensembl 2009. *Nucl. Acids Res*. 2009; 37:D690–697. [PubMed: 19033362]
56. Stark A, Lin MF, Kheradpour P, et al. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature*. 2007; 450:219–232. [PubMed: 17994088]
57. Betley JN, Frith MC, Graber JH, Choo S, Deshler JO. A ubiquitous and conserved signal for RNA localization in chordates. *Curr Biol*. 2002; 12:1756–1761. [PubMed: 12401170]
58. Mowry KL, Melton DA. Vegetal messenger RNA localization directed by a 340-nt RNA sequence element in *Xenopus* oocytes. *Science*. 1992; 255:991–994. [PubMed: 1546297]

59. Deshler JO, Highett MI, Schnapp BJ. Localization of *Xenopus* Vg1 mRNA by Vera protein and the endoplasmic reticulum. *Science*. 1997; 276:1128–1131. [PubMed: 9148809]
60. Gautreau D, Cote CA, Mowry KL. Two copies of a subelement from the Vg1 RNA localization sequence are sufficient to direct vegetal localization in *Xenopus* oocytes. *Development*. 1997; 124:5013–5020. [PubMed: 9362462]
61. Kwon S, Abramson T, Munro TP, John CM, Kohrmann M, Schnapp BJ. UUCAC- and vera-dependent localization of VegT RNA in *Xenopus* oocytes. *Curr Biol*. 2002; 12:558–564. [PubMed: 11937024]
62. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994; 2:28–36. [PubMed: 7584402]
63. Hiller M, Pudimat R, Busch A, Backofen R. Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. *Nucleic Acids Res*. 2006; 34:e117. [PubMed: 16987907]
64. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*. 2007; 8:479–490. [PubMed: 17473849]
65. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*. 2008; 9:474. [PubMed: 19014431]
66. Klein RJ, Eddy SR. RSEARCH: finding homologs of single structured RNA sequences. *BMC Bioinformatics*. 2003; 4:44. [PubMed: 14499004]
67. Pickard BS, Knight HM, Hamilton RS, et al. A common variant in the 3' UTR of the GRIK4 glutamate receptor gene affects transcript abundance and protects against bipolar disorder. *Proc Natl Acad Sci U S A*. 2008; 105:14940–14945. [PubMed: 18824690]
68. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res*. 2001; 29:4724–4735. [PubMed: 11713323]
69. Hofacker IL, Priwitzer B, Stadler PF. Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics*. 2004; 20:186–190. [PubMed: 14734309]
70. Hamilton RS, Hartswood E, Vendra G, et al. A bioinformatics search pipeline, RNA2DSearch, identifies RNA localization elements in *Drosophila* retrotransposons. *Rna*. 2009; 15:200–207. [PubMed: 19144907]
71. Lambert A, Fontaine JF, Legendre M, et al. The ERPIN server: an interface to profile-based RNA motif identification. *Nucleic Acids Res*. 2004; 32:W160–165. [PubMed: 15215371]
72. Ainger K, Avossa D, Diana AS, Barry C, Barbarese E, Carson JH. Transport and localization elements in myelin basic protein mRNA. *J Cell Biol*. 1997; 138:1077–1087. [PubMed: 9281585]
73. Carson JH, Gao Y, Tatavarty V, et al. Multiplexed RNA trafficking in oligodendrocytes and neurons. *Biochim Biophys Acta*. 2008; 1779:453–458. [PubMed: 18442491]
74. Tiedge H. K-turn motifs in spatial RNA coding. *RNA Biol*. 2006; 3:133–139. [PubMed: 17172877]
75. Cohen RS, Zhang S, Dollar GL. The positional, structural, and sequence requirements of the *Drosophila* TLS RNA localization element. *Rna*. 2005; 11:1017–1029. [PubMed: 15987813]
76. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007; 23:2947–2948. [PubMed: 17846036]
77. Rivas E, Eddy SR. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*. 2001; 2:8. [PubMed: 11801179]
78. Washietl S, Hofacker IL, Stadler PF. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A*. 2005; 102:2454–2459. [PubMed: 15665081]
79. Yao Z, Weinberg Z, Ruzzo WL. CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics*. 2006; 22:445–452. [PubMed: 16357030]
80. Seemann SE, Gorodkin J, Backofen R. Unifying evolutionary and thermodynamic information for RNA folding of multiple alignments. *Nucleic Acids Res*. 2008; 36:6355–6362. [PubMed: 18836192]
81. Sankoff D. Simultaneous Solution of the RNA Folding, Alignment and Protosequence Problems. *SIAM Journal on Applied Mathematics*. 1985; 45:810–825.

82. Havgaard JH, Lyngso RB, Gorodkin J. The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search. *Nucleic Acids Res.* 2005; 33:W650–653. [PubMed: 15980555]
83. Mathews DH. Predicting a set of minimal free energy RNA secondary structures common to two sequences. *Bioinformatics.* 2005; 21:2246–2253. [PubMed: 15731207]
84. Touzet H, Perriquet O. CARNAC: folding families of related RNAs. *Nucleic Acids Res.* 2004; 32:W142–145. [PubMed: 15215367]
85. Chartrand P, Meng XH, Singer RH, Long RM. Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle in vivo. *Curr Biol.* 1999; 9:333–336. [PubMed: 10209102]
86. Chartrand P, Meng XH, Huttelmaier S, Donato D, Singer RH. Asymmetric sorting of ash1p in yeast results from inhibition of translation by localization elements in the mRNA. *Mol Cell.* 2002; 10:1319–1330. [PubMed: 12504008]
87. Gonzalez I, Buonomo SB, Nasmyth K, von Ahsen U. ASH1 mRNA localization in yeast involves multiple secondary structural elements and Ash1 protein translation. *Curr Biol.* 1999; 9:337–340. [PubMed: 10209099]
88. Olivier C, Poirier G, Gendron P, Boisgontier A, Major F, Chartrand P. Identification of a conserved RNA motif essential for She2p recognition and mRNA localization to the yeast bud. *Mol Cell Biol.* 2005; 25:4752–4766. [PubMed: 15899876]
89. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature.* 2008; 452:51–55. [PubMed: 18322526]
90. Koplin J, Mu Y, Richter C, Schwalbe H, Stock G. Structure and dynamics of an RNA tetraloop: a joint molecular dynamics and NMR study. *Structure.* 2005; 13:1255–1267. [PubMed: 16154083]
91. Czaplinski K, Kocher T, Schelder M, Segref A, Wilm M, Mattaj IW. Identification of 40LoVe, a *Xenopus* hnRNP D family protein involved in localizing a TGF-beta-related mRNA during oogenesis. *Dev Cell.* 2005; 8:505–515. [PubMed: 15809033]
92. Shulman-Peleg A, Shatsky M, Nussinov R, Wolfson HJ. Prediction of interacting single-stranded RNA bases by protein-binding patterns. *J Mol Biol.* 2008; 379:299–316. [PubMed: 18452949]
93. Hoffman MM, Khrapov MA, Cox JC, Yao J, Tong L, Ellington AD. AANT: the Amino Acid-Nucleotide Interaction Database. *Nucl. Acids Res.* 2004; 32:D174–181. [PubMed: 14681388]
94. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. *Mol Cell.* 1998; 2:437–445. [PubMed: 9809065]
95. Jaramillo AM, Weil TT, Goodhouse J, Gavis ER, Schupbach T. The dynamics of fluorescently labeled endogenous gurken mRNA in *Drosophila*. *J Cell Sci.* 2008; 121:887–894. [PubMed: 18303053]
96. Weil TT, Forrest KM, Gavis ER. Localization of bicoid mRNA in late oocytes is maintained by continual active transport. *Dev Cell.* 2006; 11:251–262. [PubMed: 16890164]
97. Orengo CA, Pearl FM, Thornton JM. The CATH domain structure database. *Methods Biochem Anal.* 2003; 44:249–271. [PubMed: 12647390]
98. Horn WT, Tars K, Grahn E, et al. Structural basis of RNA binding discrimination between bacteriophages Qbeta and MS2. *Structure.* 2006; 14:487–495. [PubMed: 16531233]
99. Long D, Lee R, Williams P, Chan CY, Ambros V, Ding Y. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol.* 2007; 14:287–294. [PubMed: 17401373]
100. Heale BS, Soifer HS, Bowers C, Rossi JJ. siRNA target site secondary structure predictions using local stable substructures. *Nucleic Acids Res.* 2005; 33:e30. [PubMed: 15722476]
101. Eddy SR. What is a hidden Markov model? *Nat Biotechnol.* 2004; 22:1315–1316. [PubMed: 15470472]
102. Eddy SR, Durbin R. RNA sequence analysis using covariance models. *Nucleic Acids Res.* 1994; 22:2079–2088. [PubMed: 8029015]
103. Durbin, R.; Eddy, SR.; Krogh, A.; Mitchison, GJ. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press; 1998.
104. Hofacker IL, Fekete M, Stadler PF. Secondary structure prediction for aligned RNA sequences. *J Mol Biol.* 2002; 319:1059–1066. [PubMed: 12079347]

(A) Schematic of the Dynein mediated RNA transport particle. The RNA binds to the motor via one or more transacting factors and linker proteins, such as Staufen, Squid and Egalitarian. **(B)** RNA secondary structure can contain several features such as stems, hairpin loops, internal loops and bulges. The single stranded regions, and the distances between them, are particularly important for specific RNA:protein interactions. **(C)** Common file formats for RNA secondary structure (FASTA, Stockholm and CT)

Figure 2.

Flow chart of computational methods representing a generalised approach for the secondary structure prediction and searching for RNA localization signals. The blue boxed text refer to the main subheadings in this book chapter. Green dashed lines represent methods that are optional as they might not be appropriate in all cases.

Table 1

Program/Resource	Brief Description	References
<i>A. Text Editors</i>		
RALEE	RNA Alignment Editor in Emacs http://personalpages.manchester.ac.uk/staff/sam.griffiths-jones/software/ralee/	(25)
<i>B. Single Sequence RNA secondary structure prediction</i>		
Mfold / UNAFold	Minimum free energy secondary structure prediction for single RNA sequences http://mfold.bioinfo.rpi.edu/	(28, 29)
RNAfold (RNAVienna)	Predicts minimum free energy secondary structures for single RNAs http://www.tbi.univie.ac.at/RNA/	(30)
contraFOLD	RNA secondary structure searching prediction based on experimentally determined parameters and probabilistic models http://contra.stanford.edu/contrafold/	(32)
Sfold	RNA secondary structure prediction using a Boltzmann statistical ensemble http://sfold.wadsworth.org	(33)
Pfold	Prediction of RNA secondary structures using stochastic context free grammars on aligned RNA sequences http://www.daimi.au.dk/~compbio/pfold	(34)
RNAkinetics	Calculated RNA secondary structure based on local sequences, rather than the global sequence http://bioinf.fbb.msu.ru/RNA/kinetics	(35)
Vienna+P	Calculates RNA secondary structure in the presence of RNA binding proteins http://bioserv.mps.ohio-state.edu/Vienna+P	(36)
<i>C. Pseudoknot secondary structure prediction</i>		
Pknots	RNA secondary structure prediction for pseudoknots http://selab.janelia.org/software.html	(44)
ILM	Prediction of RNA secondary structures with pseudoknots http://cic.cs.wustl.edu/RNA/	(46)
HotKnots	Pseudoknot RNA secondary structure prediction http://www.cs.ubc.ca/labs/beta/Software/HotKnots/	(45)
<i>D. Measuring secondary structure similarity</i>		
RNAdistance	Calculates RNA secondary structure global similarity based on tree alignment http://www.tbi.univie.ac.at/RNA/	(30)
RNAforester	Local comparison of RNA structures based on tree alignments http://bibiserv.techfak.uni-bielefeld.de/rnaforester/	(47)
<i>E. Getting homologous sequences</i>		
Rfam	Database of RNA families http://rfam.sanger.ac.uk	(51)
UTRdb	Database for eukaryotic 3' UTR and 5' UTR sequences http://utrdb.ba.itb.cnr.it	(53)
NDB	Database of RNA three dimensional structures http://ndbserver.rutgers.edu	(54)
Ensemble	Provides sequences data for a large number of model organisms http://www.ensembl.org	(55)
12× Drosophilids	The genome sequences for the twelve sequenced <i>Drosophilids</i> http://flybase.org	(56)
BRALIBASE	Database of RNA structural alignments for the benchmarking of RNA alignment programs http://projects.binf.ku.dk/pgardner/bralibase/	(41)
BLAST	Sequence searching based on local sequence alignments http://blast.ncbi.nlm.nih.gov/Blast.cgi	(50)

Program/Resource	Brief Description	References
Infernal	Models sequence and secondary structure similarity for searching sequence databases for homologous RNA structures http://selab.janelia.org/software.html	(24)
RNApromo	Uses a SCFG approach for searching for RNA motifs using experimentally determined or MFE RNA secondary structures http://genie.weizmann.ac.il/pubs/rnamotifs08/rnamotifs08_exe.html	(52)
<i>F. Search for motifs - sequence and structure</i>		
REPFIND	Finds clustered repeats in nucleotide sequences http://zlab.bu.edu/repfind/	(57)
MEME / MEMERIS	Searching for sequence motifs / searching for sequence motifs in single stranded RNA secondary structures http://meme.sdsc.edu/meme4_3_0/intro.html http://www.bioinf.uni-freiburg.de/~hiller/MEMERIS/	(62, 63)
RNAalifold	Prediction of RNA secondary structure from aligned RNA sequences utilising the covariance between sequences http://www.tbi.univie.ac.at/RNA/	(104)
RNAmotif	Searches sequence databases for user defined sequence and secondary structure motifs http://casegroup.rutgers.edu/casegr-sh-2.5.html	(68)
RNAfold (RNAVienna)	Prediction of locally stable RNA structures from within larger sequences and can be performed on a genome scale http://www.tbi.univie.ac.at/RNA	(69)
ERPIN	Identifies conserved RNA structure motifs in sets of unaligned sequences http://tagc.univ-mrs.fr/erpin/	(71)
DNAMAN	Software package for multiple sequence alignment and analysis http://www.lynnon.com	commercial
RNA2DSearch	A bioinformatics pipeline for identifying similar localization elements http://www.rna2dsearch.com	(70)
RNABob	Pattern match based searching for RNA secondary structures http://selab.janelia.org/software.html	(13)
<i>G. Align then fold methods</i>		
QRNA	Non coding gene finder, comparing synonymous vs compensatory mutations between two sequences http://selab.janelia.org/software.html	(77)
RNAz	Calculates a structural conservation index by comparing averaged MFE structures against individual MFE structures for a set of RNAs http://www.tbi.univie.ac.at/~wash/RNAz/	(78)
CMfinder	RNA secondary structure searching based on covariance models http://bio.cs.washington.edu/yzizhen/CMfinder/	(79)
PETfold	RNA secondary structure prediction based on evolutionary and thermodynamic information - the sequences need not be homologues. http://genome.ku.dk/resources/petfold/	(80)
<i>H. Simultaneous RNA Secondary structure prediction and alignment of RNAs</i>		
Dynalign	Simultaneous alignment and secondary structure prediction for two or more RNA sequences http://rna.urmc.rochester.edu/dynalign.html	(83)
FOLDALIGN	Simultaneously folds and aligns RNA sequences based on local or global alignments http://foldalign.ku.dk/	(82)
CARNAC	Predicts RNA secondary structures by simultaneously aligning and folding the sequences http://bioinfo.lifl.fr/RNA/carnac/	(84)
<i>I. RNA Tertiary Structure</i>		
AANT	Amino Acid–Nucleotide Interaction database, a resource for searching and	(93)

Program/Resource	Brief Description	References
	visualising experimentally determined interactions http://aant.icmb.utexas.edu	
MC-Fold / MC-Sym	Molecular modeling for RNA tertiary structures from sequence http://www.major.irc.ca/MC-Pipeline/	(89)