BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 16 2010, pages 2057–2059
doi:10.1093/bioinformatics/btq321

Structural bioinformatics

Advance Access publication June 18, 2010

Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels

Fabrice Jossinet*, Thomas E. Ludwig and Eric Westhof
Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire du CNRS, F-67084 Strasbourg, France

Associate Editor: Anna Tramontano

ABSTRACT

Summary: Assemble is an intuitive graphical interface to analyze, manipulate and build complex 3D RNA architectures. It provides several advanced and unique features within the framework of a semi-automated modeling process that can be performed by homology and *ab initio* with or without electron density maps. Those include the interactive editing of a secondary structure and a searchable, embedded library of annotated tertiary structures. Assemble helps users with performing recurrent and otherwise tedious tasks in structural RNA research.

Availability and Implementation: Assemble is released under an open-source license (MIT license) and is freely available at http://bioinformatics.org/assemble. It is implemented in the Java language and runs on MacOSX, Linux and Windows operating systems.

Contact: f.jossinet@ibmc-cnrs.unistra.fr

Received on April 22, 2010; revised on June 8, 2010; accepted on June 10, 2010

1 INTRODUCTION

RNA molecules are able to adopt intricate 3D folds. The number of RNA tertiary structures has increased dramatically in recent years. Together with genomic sequence data, these structural data have sharpened our understanding of RNA structure and folding. By exploiting this body of knowledge, 3D architectures of RNA molecules can be produced using various molecular modeling strategies. Such theoretical approaches have proven to be valuable in the past for understanding folding and function. We describe here a new graphical tool, Assemble, which combines automated and manual protocols within an iterative modeling process. The modeling can be performed, either *ab initio* or by homology, on the basis of sequence alignment, chemical probing data and electron density maps derived by crystallography or cryo-electron microscopy.

2 RESULTS

2.1 General description of Assemble

An RNA architecture is achieved through two levels of organization: (i) the RNA secondary structure constraining and (ii) a tertiary structure stabilized by recurrent tertiary modules and long-range interactions. The manipulation and the construction of the RNA model can be done in a coherent fashion both at the 2D and 3D levels through two synchronized windows.

2.2 Construction of the secondary structure component

Since an RNA secondary structure produces the skeleton constraining the tertiary structure, its definition constitutes a first and essential step. Assemble is able to load any secondary structure described with the CT and BPSEQ file formats or stored in a FASTA file using the bracket notation. It can also compute a secondary structure for an RNA sequence stored in a FASTA file (see Section 2.8 for details).

The secondary structure displayed by Assemble is made of two important elements that can be changed interactively. First, the helices can be selected and moved to modify the 2D plot. They can also be deleted and created to fit the secondary structure according to the user's assumptions. If a complete or partial 3D structure exists, the tertiary base–base interactions are described with the geometric symbols of the Leontis–Westhof classification (Leontis and Westhof, 2001). They can also be edited, deleted and created interactively. They play key roles during the construction process, as described in the next sections.

2.3 Building the first draft of the tertiary structure component

The helical regions and the non-helical linking elements constitute the building blocks that can be selected in the 2D panel and translated into the 3D scene with a regular A-form helical fold. These building blocks are exported side-by-side in the 3D scene, but can be reorganized manually or automatically. If the user alters the underlying secondary structure during the modeling process, the corresponding residues can be removed from the 3D scene and new ones can be re-created. The default helical fold can be altered using a manual and an automated approach. For manual intervention, a sliding button panel allows to modify the torsion angles of any single residue present in the 3D scene. The rotation will be applied to all the residues linked through the sugar–phosphate backbone (in the 3′ or the 5′ direction) and to the residues paired to them. Consequently, the user can define the scope of this rotation by cutting/linking the molecular chains in the 3D scene and/or by editing the base–base interactions in the 2D panel. The automated approach is described in the next section.

2.4 Extraction and application of local RNA folds

RNA architectures are constituted of recurrent folds observed in various RNA molecules playing different biological functions.

*To whom correspondence should be addressed.

Consequently, Assemble provides the ability to extract and apply these 3D modules to selected regions made in the 2D/3D model. Assemble provides an embedded and extensible library of high-resolution structures derived from the Protein Data Bank (PDB) (Berman *et al*., 2000). This library is available through the 'MyPDB' sliding panel and is provided with each 3D structure pre-annotated with the secondary structure. A second sliding panel allows the user to query this library for specific RNA modules. Each hit can be displayed in the 2D and 3D scenes. The module can then be extracted and saved in a local RNA motifs repository with the 'Create RNA Motif' panel.

The application of an RNA module will thread a selection of the same number of residues in the 3D model into the original 3D fold. The base–base interactions stabilizing the original module will be added automatically to the secondary structure.

2.5 Fitting of RNA 3D model into electron density maps

The progress in cryo-electron microscopy techniques has led to density maps of large RNA architectures at resolution around or below 7 Å (Becker *et al*., 2009; Schuler *et al*., 2006). Consequently, we have added within Assemble the ability to display such density maps along with the current 3D model. Assemble can load density maps described with the XPLOR or MRC file formats. Small-angle X-ray scattering (SAXS) data can also be used by converting them to the XPLOR format with tools like the Situs program package (Wriggers, 2010).

2.6 Geometric refinement of the RNA 3D model

Once a first 3D model is established, several geometric and structural deficiencies can subsist. Consequently, Assemble provides a geometric refinement function to optimize structural parameters like nucleotide stereochemistry, all the base–base interactions, the sugar pucker and atoms distances. The structural constraints used during this refinement step are deduced from the set of base–base interactions defined in the secondary structure displayed in the 2D panel. By increasing the number of iterations during the refinement, Assemble converges to a state close to the structure described in the 2D panel. The refinement is achieved by geometrical least squares using the Konnert–Hendrickson algorithm (Hendrickson and Konnert, 1980) as implemented in the program Nuclin/Nuclsq (Westhof *et al*., 1985).

2.7 The complementarity between Assemble and the automated methods

A couple of automated methods have been published recently, generally limited in the sizes and resolutions of the produced models (Das *et al*., 2010; Jonikas *et al*., 2009; Parisien and Major, 2008). With its ability to load tertiary structures described in PDB files and to annotate them automatically with a secondary structure, Assemble can consequently be used to improve 3D models produced automatically.

2.8 The distributed architecture of Assemble

Several tasks of Assemble are delegated to RNA algorithms available as web services:

- Contrafold (Do *et al*., 2006) and RNAfold (Hofacker, 2003) for the 2D predictions;
- RNAplot (Hofacker, 2003) for the 2D plots; and
- RNAVIEW (Yang *et al*., 2003) for the 3D annotations.

These web services are hosted by our own laboratory server and are attached to the following website: http://paradise-ibmc.u-strasbg.fr/. They have been implemented as independent modules that can be used without Assemble. The web site provides several examples of usage with command-line tools like wget, curl or our own dedicated java client. The RNAfold and RNAplot algorithms are also provided by the European Bioinformatics Institute (McWilliam *et al*., 2009). This loose coupling between the graphical interface of Assemble and its algorithms will allow us to easily include new automated tasks in the framework.

2.9 The coupling of S2S and Assemble to construct a 3D model by homology

Among the different kind of usable pieces of information to construct a 3D model, the availability of a solved tertiary structure for at least one RNA molecule within a family is the richest. Since a molecular 3D architecture evolves much more slowly than sequences, structural data can be inferred for all the other members of an RNA family by homology. More importantly, because RNA modules are recurrent and occur across the phylogenetic kingdoms, once a motif has been recognized, its sequence can be easily threaded onto the known 3D fragment.

In 2005, we have released the S2S application with the initial goal to find the conserved structural core within a multiple alignment (Jossinet and Westhof, 2005). During the development of Assemble, we have updated S2S to be able to infer a 3D model for any sequence within this structural alignment. Once a 3D model is inferred from S2S, it is saved in the directory of the structural alignment, where it can be loaded by Assemble to pursue the modeling process. Consequently, S2S and Assemble can be used independently or as two complementary steps of a modeling workflow needing a solved tertiary structure and an orthologous sequence to model.

ACKNOWLEDGEMENT

We would like to thank the S2S-Assemble community for help and support.

Funding: Human Frontier Science Program (RGP0032/2005-C to E.W., in part); French National Research Agency (AMIS-ARN, NT09_519218 to E.W. and F.J.).

Conflict of Interest: none declared.

REFERENCES

Becker,T. *et al*. (2009) Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science*, **326**, 1369–1373.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Das,R. *et al*. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.

Do,C.B. *et al*. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–e98.

Hendrickson,W.A. and Konnert,J.H. (1980) Diffraction analysis of motion in proteins. *Biophys. J.*, **32**, 645–647.

Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.

Jonikas,M.A. *et al*. (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*, **15**, 189–199.

Jossinet,F. and Westhof,E. (2005) Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics*, **21**, 3320–3321.

Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.

McWilliam,H. *et al*. (2009) Web services at the European Bioinformatics Institute-2009. *Nucleic Acids Res.*, **37**, W6–W10.

Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.

Schuler,M. *et al*. (2006) Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat. Struct. Mol. Biol.*, **13**, 1092–1096.

Westhof,E. *et al*. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.*, **184**, 119–145.

Wriggers,W. (2010) Using Situs for the integration of multi-resolution structures. *Biophys. Rev.*, **2**, 21–27.

Yang,H. *et al*. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.