RNA3D
A Tool to Visualize Three-Dimensional Structure of rRNA

The RNA3D tool was developed to have a system that is capable of displaying structural information as well as interacting between the applications of ARB software in a more intuitive and flexible way. The main purposes of the RNA3D tool are to achieve the following capabilities –

- To merge the structural and phylogenetic information of underlying sequences with the structural data of rRNA.
- To allow dynamic mapping of rRNA sequence data along with the mutation, insertion and deletion information with respect to the master sequence onto the molecule in real time.
- To interact with the primary structure of rRNA sequences (ARB Primary Structure Editor) in order to facilitate the fine tuning of alignments.
- To visualize any column statistics performed on the rRNA sequences (multiple alignments) contained in the ARB databases.
- To combine the secondary structure information of rRNA (in the form of loops and helices) with the three-dimensional structure of rRNA.
- To be able to display the oligonucleotide probe targets designed (using ARB Probe Design tool) or imported, in the three-dimensional spatial conformation of rRNA molecule.
- To display base positions with respect to the master sequence within the molecule.
- To provide the user with more customizable tool which can be used according to his/her needs.

The Interface

The annotation of RNA three-dimensional structures consists of a preprocessing of the information embedded in their 3-D coordinates. Since the structure of *Escherichia coli* ribosome has not yet been resolved, the homology model of the atomic structure of the *E. coli* 30S ribosomal subunit (Tung et al. 2002) which is determined using the crystal structure of the *Thermus thermophilus* 30S ribosomal subunit (Wimberly et al. 2000) as template (PDB entry 1J5E), is retrieved from the protein data bank (PDB) (PDB entry 1M5G) and used as a reference structure for the RNA3D tool.

In order to objectively represent the structural knowledge of three-dimensional ribosomal RNA structure, the respective 3-D coordinates were extracted from the PDB file (1M5G) and used for further structural analysis and searches.

The RNA3D tool uses the popular OpenGL graphics library combined with Open Motif user interface for achieving more intuitive rendering and manipulation of the rRNA molecule with in the ARB environment. It processes PDB structural information stored in PDB file (1M5G) into the annotated structures and renders (draws) them into the virtual space using OpenGL routines. In order to provide user with the more detailed perspective of 16S rRNA structure, structural information corresponding to the ribosomal proteins were not included in the annotation. The extracted structural information is then fed to OpenGL engine, where it is further transformed into a hierarchy of OpenGL objects, which encode molecule chains, residues and base positions.

Further processing is done in order to combine other structural or sequence associated data with the three-dimensional structural data. For example, combining secondary structural information (in the form of loops and stems) with the rRNA 3D structure.

User customization

Since the user customization is an important consideration in graphical user interface (GUI) design, RNA3D tool provides the individual users with more possibilities to customize the interface to suit their particular purpose and preferences.

As a first step toward enhancing the user customization capability of RNA3D tool, any form of annotation and information overlay can be toggled on and off. This feature allows users to focus on annotations they consider important without being distracted by information irrelevant to their particular needs. Viewing all of the structural and overlay information at once result in overwhelming displays. The feature, to "toggle on and off" the information displayed, becomes more essential for viewing the molecule more clearly.

Additionally, the user is provided with more customization capabilities in the form of specifying different colors, shapes, letters, and the thickness of the objects rendered onto the scene at any time using Color Palette, Bases, Helix, Molecule and Mapping buttons of RNA3D tool. For example, the user can colorize the entire molecule based on the residues that are participating in loop or stem formation in the accepted secondary structure model of 16S rRNA.

Furthermore, a range of colors is included in order to overlay any sequence associated information onto the molecule and also a separate range of colors to visualize the series of search strings or probe targets with in the molecule.

Navigation

The entire set of visualized objects can easily be rotated, translated and scaled at the user's wish. Navigation through the molecule is basically bound to the standard mouse buttons and mapped to simple keys on the keyboard. Rotation of the molecule is achieved by moving the mouse whilst holding the left mouse button. The molecule can be rotated in any desired direction (360 degrees).

Additionally, the molecule can be made to rotate automatically by pressing the space bar on the keyboard. Pressing again the space bar will stop the automatic rotation of the molecule.

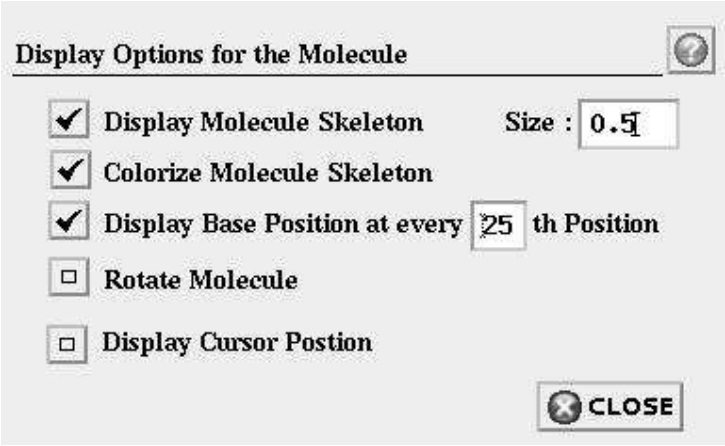
Translation (horizontal and vertical movement of the molecule in RNA3D Window) of the molecule is performed using left, right, up and down arrow keys on the keyboard.

Molecule can be scaled using the zoom function of RNA3D tool. For easy navigation, the zoom function is bound to wheel of the mouse. The user can zoom in to or out of the molecule by performing upward or downward motion of the wheel, respectively.

Easy rotation, translation and scaling of the molecule enable the user to observe the buried and exposed molecule sections. This feature is more useful when the user wishes to see the binding sites of rRNA targeted oligonucleotide probes in the molecule.

Furthermore, the user can display the current cursor position in the linear structure of sequence alignments (ARB Primary structure editor) in the molecule. Using this feature, the user can quickly locate the region of rRNA sequence he/she is examining in the both primary and three-dimensional structure of the rRNA molecule.

General Display Options for the Molecule



The following display options can be set to the three-dimensional structure of small subunit rRNA -

Display Molecule Skeleton
Enabling this will display the entire molecule skeleton in a user-defined color. By setting a grey or light color you can achieve transparent contours of the molecule displayed avoiding its interference with the mapping information.

Size
The size or thickness of the skeleton can be set by specifying the desired thickness in the "size" box. By default it is set to decimal 5.

Colorize Molecule Skeleton
Based on the residues participating in secondary structural motifs (loops, helices, bulges) the molecule skeleton can be colored. Color settings with respect to secondary structural motifs can be changed using the "Color Palate" of the RNA3D interface.

Display Base Position
Base positions corresponding to the reference sequence (*Escherichia coli*) can be displayed by checking this check box. The interval of positions to be

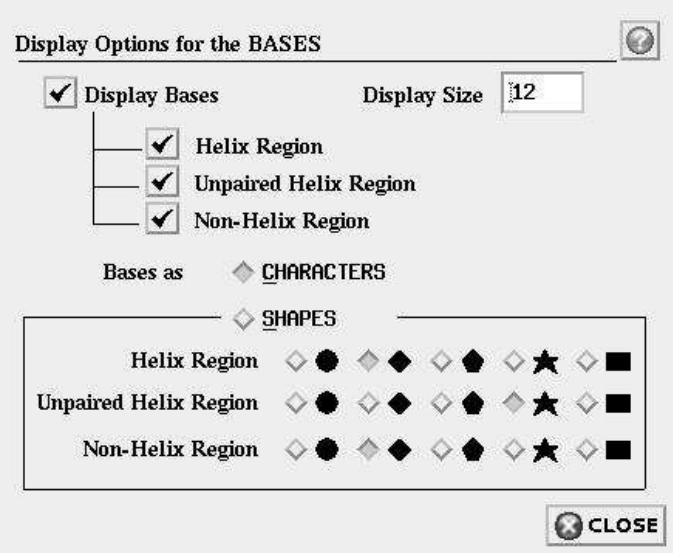
displayed can be changed by specifying the desired "interval size" at the included box. Displaying the base positions helps i) to locate probe binding sites within the molecule, ii) to refine the sequence alignments according to the molecule structure, and also iii) to identify the exact position in the primary sequence, where insertions, deletions and base substitutions occur with respect to the template sequence when a different rRNA sequence is mapped onto the master structure.

Rotate Molecule
Enabling this check box rotates the molecule automatically. The direction and speed of the rotation can be changed by using left mouse button and mouse movement, respectively. Alternatively, molecule can also be rotated by pressing "space bar" on the keyboard.

Display Cursor Position
Checking this box will enable the cursor position to be displayed in the molecule. Cursor position is directly connected to primary structure editor (ARB_EDIT4) and any movement of cursor in ARB_EDIT4 is instantly updated in the RNA3D window.

Display Options for the Residues / Bases

To achieve more performance and dynamic overlay of any sequence associated information, rendering (drawing) was simplified to chain display with the capacity to display residues in the form nucleotides – Adenosine (A), Guanine (G), Cytosine (C) and Uracil (U) at the respective coordinates in the molecule. Also viewing the entire chemical structure in the molecule's 3D structure is less readable.



Display Bases
By enabling the check box the corresponding residues in the rRNA sequence can be displayed on the rRNA 3D structure. Disabling this check box will display the molecule skeleton without residues.

Helix Region
Enabling this check box will display residues that are participating in Helix formation in the secondary structure models of small subunit rRNA.

Unpaired Helix Region
Enabling this check box will display residues that are participating in bulge (unpaired helix) formation in the secondary structure models of small subunit rRNA.

Non-Helix Region
Enabling this check box will display residues that are participating in loop (non-helix) formation in the secondary structure models of small subunit rRNA.

Display Size
The size of the residues displayed can be changed by specifying the desired size in "Display size" box.

CHARACTERS
The corresponding residues are displayed with the actual nucleotides – Adenosine (A), Guanine (G), Cytosine (C) and Uracil (U).

SHAPES
The corresponding residues are displayed with the respective shapes specified for different structural motifs.

By setting different colors for the secondary structural motifs (loops, stems and bulges) using "Color Settings", the respective regions can be easily recognized in the rRNA 3D structure.

Mapping Secondary Structural Information

Secondary and tertiary structure interactions of the well established comparative structure models of rRNA which are used in SECEDIT to generate 2D structure models are fitted to the three-dimensional structure of *E. coli* master sequence.

The preprocessing was done on the secondary structure models in order to differentiate stems (helices), bulges (unpaired bases in the helix region) and loops in the rRNA crystal structure. When variability maps are overlaid onto the structure, this feature enables the user to identify the conserved and variable regions in the small subunit of the ribosomal RNA (see "overlaying of sequence associated information" section).

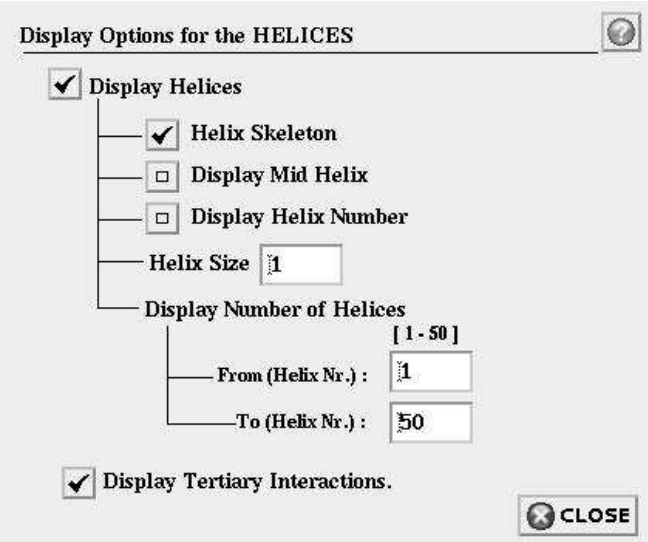
Mutations of single nucleotide with respect to loop and stem regions of the rRNA structure can be seen when mutation information (calculated for the overall sequences in the database) is superimposed.

Optionally, the user can choose the number of stems (helix regions) to be displayed in the crystal structure along with the helix numbers (1-50 for 16S rRNA). Helix numbers are displayed according to the ARB numbering scheme.

Tertiary interactions observed (Gautheret et al. 1995) in the small subunit of rRNA can also be displayed in the rRNA structure.

Different colors can be defined for all of the secondary and tertiary interactions, which help the user to immediately see the differences and distribution of different interactions in the small subunit of rRNA crystal structure.

Display Options for Mapping Secondary Structural Information



Display Helices

Enabling this check box maps the secondary structural motifs (loops, helices and bulges) onto the molecule. Secondary structural information is according to the comparative models of rRNA used in primary and secondary structure editors.

Helix Skeleton

Enabling will draw a skeleton of secondary structure mask. By setting a grey or light color you can achieve transparent mask avoiding any interference with other information overlays.

Display Mid-Helix

This displays middle point of the helices.

Display Helix Number

Checking this box will display the corresponding helix numbers in the rRNA 3D structure. Helix numbers are according to ARB numbering scheme. The

small subunit (16S) rRNA of *E.coli* contains 50 helices which are numbered from 1 to 50.

Display Number of Helices
Using this you can set the number of helices you would like to be displayed in the 3D molecule. This feature is very helpful to thoroughly examine the specific helices in the structure.

Helix Size
Thickness or the size of the helices can be set by specifying the desired value in this box.

Display Tertiary Interactions
The tertiary interactions observed in small subunit rRNA can be displayed in the three-dimensional conformations of small subunit rRNA by enabling this check box. The tertiary information data is from Gautheret et al. See references at the end.

Color settings related to helix, skeleton, mid-helix, helix number and tertiary interactions can be changed using "Color Settings" of the main RNA3D window.

Superimposing rRNA sequence data

Any rRNA sequence contained in the multiple sequence alignments of ARB primary structure editor can be overlaid onto the structure of *E. coli* in the RNA3D window. Desired rRNA sequence to be mapped onto the structure can be selected by the left mouse button in the multiple sequence alignment and the selected rRNA sequence will be instantly mapped onto the master structure.

The display of full name and short name (ARB ID) within the RNA3D window, in addition to the crystal structure, immediately displays the current species (rRNA sequence) that is mapped onto the master structure. This feature is more useful when a large number of rRNA sequences are included in the multiple sequence alignments. Also the feature is later very handy, when the user walks through the constructed phylogenetic tree selecting species in different taxa. And the selected species in the tree is dynamically mapped to the molecule in the RNA3D window.

The selected rRNA sequence is annotated with mutation (base substitutions), insertion and deletion information at each site as compared to the master sequence (*E. coli*).

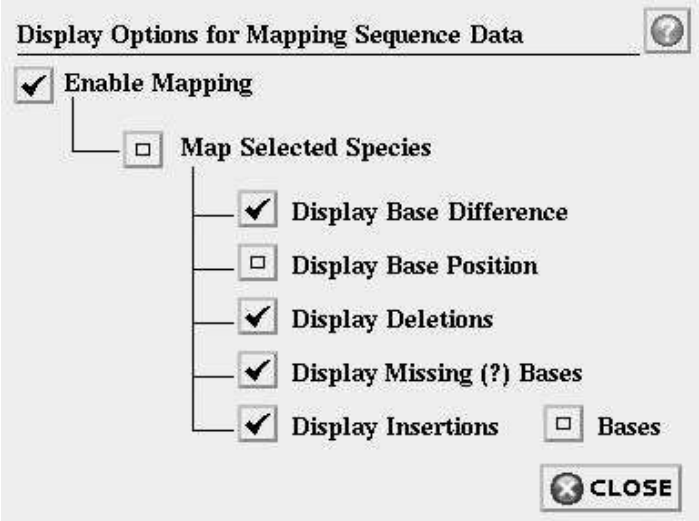
For the regions where the sequences are aligned without deletion or insertion, direct base substitution (mutation) is applied. Because the C′---C′ distance is essentially the same (~10.2 Å) in all Watson-Crick base pairs (Watson and Crick, 1953), this simple procedure preserves the base pairing and the double helical structure while substituting the bases. Although there do exist the requirement of structural adjustments for non-Watson-Crick base pairs, currently, simple base substitutions are kept because the development of new models to achieve the necessary structural adjustments is out of the scope of the RNA3D tool.

In the regions where the alignment (of selected rRNA sequence) involves insertions, the respective insertion points to corresponding *E. coli* base position in the alignment are shown as down arrows in the crystal structure.

The number of insertions and the participating nucleotides can also be displayed at the insertion points.

In the case of regions, where deletions are observed in the alignment corresponding to the master sequence (*E. coli*), respective sites in the crystal structure are indicated as deleted, using ☠ symbol.

Display Options for Mapping Sequence Data



Enable Mapping
Checking this box will enable the mapping or overlaying of any information onto the molecule globally. It is very useful to swiftly switching off mapping information.

Map Selected Species
This check box will enable mapping rRNA sequence data contained in the multiple alignments onto the 3D molecule.

Display Base Difference
Enabling this check box will display the substitutions or mutations observed with respect to *E.coli* sequence onto 16S rRNA 3D structure.

Display Base Position
Base positions corresponding to the observed substitutions or mutations in the mapped rRNA sequence are displayed by enabling this check box.

Display Deletions
Enabling this check box will display deletions in mapped rRNA sequence with respect to *E.coli* reference sequence data.

Display Insertions
Enabling this check box will display insertions in mapped rRNA sequence with respect to *E.coli* reference sequence data. By checking "Bases" box, the number of insertions along with the actual bases or residues is displayed at the insertion points.

Display Missing Bases
Bases or residues which are presumed to be missing in the rRNA sequence alignments when comparing with the consensus model and/or during manual curation, can be visualized in the 3D structure. Missing bases denoted as dots (".") in the multiple sequence alignments are mapped onto the rRNA 3D structure as question marks ("?") by enabling this check box. Such missing bases are more often attributed to errors during sequencing.

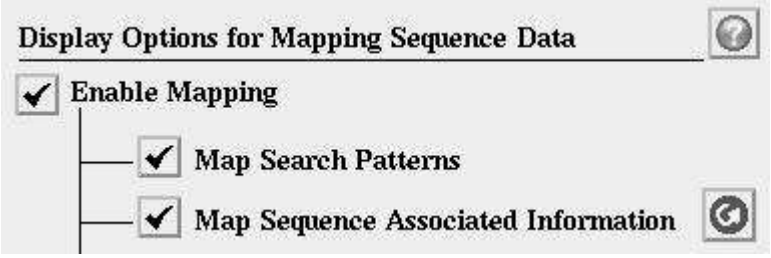
Color settings related to mapped sequence data including insertions, deletions, mutations, and missing residues can be changed using "Color Settings" of the main RNA3D window.

Mapping Probes and Overlaying Sequence Associated Information

Mapping Oligonucleotide Probes

The localization of the proposed oligonucleotide probe targets can be visualized in customizable background colors with in the rRNA crystal structure. Using the navigation capabilities of RNA3D tool (see "Navigation" section), one can get an idea about the probable binding site of the proposed probe with respect to the structural conformation of rRNA.

Oligonucleotide probes are designed using integrated Probe Design and Probe Match tools of ARB. The selected oligonucleotide probe in probe match window is directly mapped onto the rRNA 3D structure by enabling "**Map Search Patterns**" check box.



Overlaying Sequence Associated Information (SAI)

Various column statistics like sequence consensus, base frequency, positional variability based on parsimony method and any other user defined column statistics that are performed on the sequence alignments can be readily overlaid onto the 3D structure.

Once the column statistics are performed, the user can define the color translation table for the chosen SAI in the ARB primary structure editor (see "View | Visualize SAIs" menu). Different colors (up to 10 colors) can be set to the values or characters stored in the SAI to visualize in the molecular structure. The molecule can be re-colored using new settings anytime by clicking the color palate button (using Color Settings in RNA3D window).

By enabling the "**Map Sequence Associated Information**" check box, the transformed data is readily overlaid onto the rRNA 3D structure. Any change in the SAIs and respective color transformations can be reapplied by clicking "refresh" button.

By superimposing information derived form the underlying multiple sequence alignment on to the molecule dynamically helps in evaluating individual rRNA sequences and the multiple sequence alignments with respect to three-dimensional conformations of small subunit rRNA.

References

Gautheret D, Damberger SH, Gutell RR: **Identification of base-triples in RNA using comparative sequence analysis.** *J Mol Biol* 1995, **248:** 27-43.

Tung CS, Joseph S, Sanbonmatsu KY: **All-atom homology model of the Escherichia coli 30S ribosomal subunit.** *Nat Struct Biol* 2002, **9:** 750-755.

Wimberly BT, Brodersen DE, Clemons WM, Jr., Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V: **Structure of the 30S ribosomal subunit.** *Nature* 2000, **407:** 327-339.