Molecular Profiling Research Center for Drug Discovery (molprof), AIST

RNA Structural Workflow

User Manual

AIST 2019/02/18

Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

Contents

| 1 | Inti | roduction3 |
|---|---------|---|
| 2 | Abc | out the workflow of RNA related analysis4 |
| 3 | Con | nmon rules5 |
| 4 | Usa | ge of RNA Structural Prediction9 |
| | 4.1 | Preparation |
| | 4.2 | Node9 |
| | 4.2.1 | Node list10 |
| | 4.2.2 | Node setting14 |
| | 4.2.2.1 | SetVariable14 |
| | 4.2.2.2 | 2 Case Switch Variable15 |
| | 4.2.2.3 | B FastaFileReader15 |
| | 4.2.2.4 | CentroidFold_AIST16 |
| | 4.2.2.5 | 5 IPknot_AIST17 |
| | 4.2.2.6 | 8 RNA2Dchecker_AIST18 |
| | 4.2.2.7 | 7 RASSIE_AIST19 |
| | 4.2.2.8 | B Rascal_AIST20 |
| | 4.2.2.9 | P RactIP_AIST |
| | 4.2.2.1 | 0 fpocket2_AIST |
| | 4.2.2.1 | 1 PocketSelector |
| | 4.2.2.1 | 2 FragmentSelector |
| | 4.2.2.1 | 3 AutoDockVina_AIST |
| | 4.2.2.1 | 4 DockingAnalyzer_AIST27 |
| | 4.2.2.1 | 5 Rebuild_AIST |
| | 4.2.2.1 | .6 MinMM_AIST |
| 5 | Exa | mples for Execution and Result Viewing |
| | 5.1 | Prediction of Tertiary Structures |
| | 5.1.1 | $\label{eq:prediction} Prediction of Tertiary \ Structures \ using \ the \ ReadFastaReader \ node \ \ 30$ |
| | 5.1.2 | Prediction of Tertiary Structures using the Sparql_AIST node35 |
| | 5.2 | Prediction of RNA-RNA interactions |
| | 5.3 | Prediction of Protein-RNA interactions |
| 6 | SPA | ARQL node |
| | 6.1 | Sparql related node |
| | 6.1.1 | Node list |
| | 6.1.2 | Sparql_AIST51 |
| | 6.1.2.1 | Setting execution environment |

| | 6.1.3 Sparql_AIST_Adv | 52 |
|---|---|----|
| | 6.1.3.1 Setting execution environment | 52 |
| | 6.1.4 SequenceSelector | 54 |
| | 6.1.4.1 Result | 54 |
| | 6.1.5 HtmlView | 55 |
| | 6.1.5.1 Result | 55 |
| 7 | Appendix | 55 |
| | 7.1 Appendix A: LSDBCrossSearch | 56 |
| 8 | Contact | 58 |
| 9 | Terms and Conditions of License for Use | 58 |

1 Introduction

This manual describes the KNIME workflow of RNA structural analysis, which was developed at the Molecular Profiling Research Center for Drug Discovery (molprof), Advanced Industrial Science and Technology (AIST), Tokyo, JAPAN.

For installation of the workflow combination type, please refer to the installation manual available at the Life Science Database Integration website.

The Active workflows run on the KNIME platform. Please refer to the KNIME site for details about KNIME. This manual explains how the user can use the workflows of RNA structural analysis.

KNIME: http://www.knime.org/

2 About the workflow of RNA related analysis

| No. | Active workflow | Operating | Explanation |
|-----|---------------------------------|----------------|-----------------------------|
| | combination type name | system | |
| 1 | RNA Tertiary Structure | Windows 62 bit | Workflow that predicts |
| | Workflow | Linux 64 bit | the 3D structure of RNA |
| | | MacOS | molecules |
| 2 | RNA-RNA Interaction | Windows 62 bit | Workflow that predicts |
| | Workflow | Linux 64 bit | the 3D structure of |
| | | MacOS | RNA-RNA interactions |
| 3 | RNA -Protein Interaction | Windows 62 bit | Workflow that predicts |
| | Workflow | Linux 64 bit | the 3D structure of |
| | | MacOS | RNA-protein interactions |

2-1 Active workflow combination type list

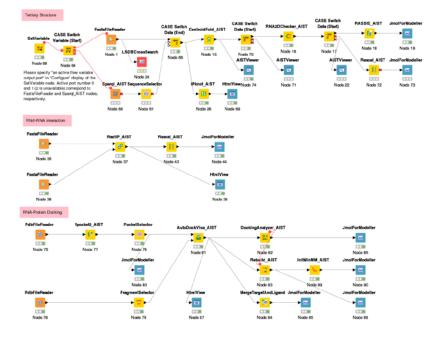
3 Common rules

Common rules in all Active workflows are as follows.

1. Starting Active workflow

Users can download the KNIME package for the RNA structure prediction workflow from the website

(http://togo.medals.jp/active_local_rna_prediction.eng.html). The workflow is then shown and ready to use.



3-1 Starting the RNA_Structure_Prediction Active workflow

2. Node

A node is an icon that is shown in a workflow screen as follows;

| File | Reader |
|------|--------|
| | |



3-2 FASTA File Reader Node (as an example)

When a node is selected, the explanation of each node is displayed in the "Node Description" column on the right of the KNIME screen.

3. Node progress

Signals below a node indicate the progress status, as shown below.

| signal color | color | Progress message | | |
|--------------|--------|-------------------------|--|--|
| e | Red | Preparing for execution | | |
| | Yellow | Stand-by | | |
| Green | | Complete | | |
| | Thick | In progress | | |
| | blue | | | |
| queued | queued | Queued | | |

3-3 Signal of node progress list

4. Node menu

A node menu is shown when right-clicking on a node, as shown below.

| ø | Configure |
|--|--------------------------------|
| \bigcirc | Execute |
| Ð | Execute and Open Views |
| | Cancel |
| × | Reset |
| ﹐₹ | Edit Node Name and Description |
| 0 | New Workflow Annotation |
| Q | View: name of first view |
| ot | Cut |
| Ð | Сору |
| Ē | Paste |
| \checkmark | Undo |
| $\langle \! \! \! \! \! \! \! \! \! \! \! \! \! \! \! \! \! \! \!$ | Redo |
| × | Delete |

3-4-Node menu

| Menu command | Action | Note |
|-------------------------|-------------------------------|-----------------------|
| Configure | Various settings of a node. | Another window is |
| | | started. |
| Execute | Execute the node. | The node cannot be |
| | | used unless the node |
| | | status is yellow. |
| Execute and Open Views | This is an active display for | The node cannot be |
| | the node that displays the | used unless the node |
| | result window. | status is yellow. |
| | Execute a node. | |
| Cancel | Cancel the execution. | The node cannot be |
| | | used unless the node |
| | | status is deep blue. |
| Reset | The setting is reset. | The node is active if |
| | | the node status is |
| | | green. |
| Edit Node Name and | Use to change the node name | Another window is |
| Description | or description. | started. |
| New Workflow Annotation | Use to insert some comment. | The comment column |
| | | is displayed. |
| View : [viewer name] | Use to display results. | Another window is |
| | | started. |
| Cut | The node and the comment, | _ |
| | etc. are cut. | |
| Сору | The node and the comment, | _ |
| | etc. are copied. | |
| Paste | The node and the comment, | _ |
| | etc., which are copied, are | |
| | pasted. | |
| Undo | Use to undo cut, copy or | _ |
| | paste. | |
| Redo | Use to cancel the action | _ |
| | undone. | |
| Delete | The node and the comment, | - |
| | etc. are deleted. | |

3-4-2 Node menu list

5. Execute all executable nodes

All the nodes can be executed simultaneously when all configurations of the nodes are complete.

In that case, click on the icon at the top of the KNIME screen (shown below) after selecting the node, which is a starting point. (Execute all executable nodes (Shift+F7).)

3-5 Execute all executable nodes

 \odot

6. Alert messages and error messages

If an alert or an error occur after a node is executed, a pop-up screen will appear along with messages in the Console on the KNIME screen. Those should be checked to resolve problems.

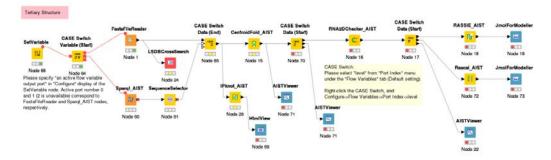
Examples of messages and measures include:

| No | Messages | Cause and measures |
|----|--|---------------------------------|
| 1 | Console: | Cause: |
| | WARN FastaFileReader 0:2:1 | The file is not specified. |
| | failed to apply settings: Please specify | Measures: |
| | a filename. | Specify the file. |
| 2 | Pop up: | Cause: |
| | SOAP execution error. | An error occurred when SOAP was |
| | Please resubmit again later. | executed. |
| | Console: | Measures: |
| | ERROR CentroidFold_AIST Execute | Execute it again later. |
| | failed: Error occurred. | |

3-6 Alert messages (examples)

4 Usage of RNA Structural Prediction

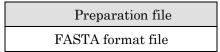
The user can download combination types of RNA workflow packages (AIST-knime packages) from the TOGO web site: (<u>http://togo.medals.jp/active local rna prediction.eng.html</u>). After installation, workflows that predict RNA structures, RNA-RNA interactions and protein-RNA complexes can be used.



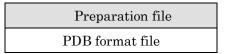
4 RNA Tertiary Structure prediction workflow

4.1 Preparation

Prepare RNA sequences in FASTA format for RNA and RNA-RNA interactions.



Prepare an RNA sequence in FASTA format and a protein PDB file for the protein-RNA interaction.



4.2 Node

There are various nodes in the RNA Structure Prediction Active Workflow. The nodes are shown below and should be configured accordingly.

4.2.1 Node list

% $\,$ In the following list, "Y" indicates a node where the setting is necessary.

| No | Node name | Icon | set | explanation |
|----|----------------------|---------------------|-----|------------------------|
| 1 | SetVariable | SetVariable | Y | Output port by |
| | | 0 <u>12</u> | | specifying integer 0, |
| | | | | 1, or 2 in this node's |
| | | Node 75 | | Configure dialog. |
| 2 | Case Switch Variable | CASE Switch | | The flow variables |
| | | Variable (Start) | | from the input port |
| | | ≠v | | are moved to exactly |
| | | Node 76 | | one active output |
| | | | | branch. |
| 3 | FastaFileReader | FastaFileReader | Y | The FASTA format |
| | | <u>Ľ</u> . | | file is read. |
| | | | | |
| | | Node 1 | | |
| 4 | LSDBCrossSearch | LSDBCrossSearch | - | LSDB cross-search is |
| | | ► <mark>1508</mark> | | executed. |
| | | | | |
| | | Node 24 | | |
| 5 | Sparql_AIST | Sparql_AIST | Y | SPARQL searches for |
| | | 1 | | endpoint |
| | | 0.0 | | (microfRNAdb) |
| | | Node 85 | | |
| 6 | SequenceSelector | SequenceSelector | | Selection of sequence |
| | | ▶■▶ | | executed |
| | | | | |
| | | Node 61 | | |

4.2.1-1 RNA_Structure_Prediction Active workflow node list

10

| ~ | Control IE-11 AICT | | v | Que tracid Fall is |
|----|--------------------|---------------------|---|-----------------------|
| 7 | CentroidFold_AIST | CentroidFold_AIST | Y | CentroidFold is |
| | | ▶ <mark> </mark> | | executed. |
| | | | | |
| | | Node 15 | | |
| 8 | IPknot_AIST | IPknot_AIST | Y | IPknot is executed. |
| | | N DE N | | |
| | | | | |
| | | Node 26 | | |
| 9 | RNA2DChecker_AIST | | Y | RNA2DChecker is |
| 5 | | RNA2DChecker_AIST | 1 | executed. |
| | | ► <mark>C</mark> × | | executed. |
| | | Node 16 | | |
| | | | | |
| 10 | RactIP_AIST | RactIP_AIST | Y | RactIP is executed. |
| | | <mark></mark> ≻ | | |
| | | | | |
| | | Node 37 | | |
| 11 | RASSIE_AIST | | Y | RASSIE is executed. |
| | | RASSIE_AIST | | |
| | | | | |
| | | | | |
| | | Node 18 | | |
| 12 | Rascal_AIST | Rascal_AIST | Y | Rascal is executed. |
| 12 | hastal_AIST | | 1 | Rascal is executed. |
| | | ► <mark>××</mark> ► | | |
| | | Node 91 | | |
| | | Node 81 | _ | |
| 13 | fpocket2_AIST | | Y | Execute fpocket2. |
| | | | | |
| | | | | |
| | | | | |
| 14 | PocketSelector | PocketSelector | | Select a pocket from |
| | | ▶ 🔳 ▶ | | the fpocket2 results. |
| | | | | |
| | | Node 78 | | |
| | | Node / 8 | | |

| 15 | FragmentSelector | FragmentSelector | Y | Generate a PDB |
|----|-------------------------|---------------------------------------|---|-----------------------|
| 10 | rragmentSelector | · · · · · · · · · · · · · · · · · · · | I | |
| | | × [™] | | ATOM file that |
| | | | | consists of the RNA |
| | | Node 79 | | fragment selected by |
| | | | | the user. |
| 16 | AutoDockVina_AIST | AutoDockVina_AIST | Y | Execute AutoDock |
| | | Node 81 | | Vina. |
| 17 | MergeTargetAndLigan | MergeTargetAndLigand | | Merge receptor |
| | d | | | (protein) and ligand |
| | | | | (RNA) file. |
| | | Node 84 | | |
| 18 | DockingAnalyzer_AIST | DockingAnalyzer_AIST | Y | Execute cluster |
| 10 | Dooming mary zor_rine r | | - | analysis (k-means) |
| | | P <mark>_I</mark> ≧_P | | and principal |
| | | | | component analysis |
| | | Node 82 | | (PCA) |
| 19 | Rebuild_AIST | Rebuild_AIST | Y | Rebuild an RNA |
| 10 | Nebulia_AIST | = 0 | 1 | ligand using a |
| | | | | |
| | | | | complete RNA |
| | | Node 83 | | structure. |
| 20 | MinMMCandidateSelec | MinMMCandidateSelector | | Selection of one |
| | tor | ▶ | | model for energy |
| | | | | minimization. |
| | | Node 91 | | |
| | | | | |
| 21 | MinMM_AIST | MinMM_AIST | Y | Energy minimization |
| | | ► 🗛 ► | | by using molecular |
| | | | | mechanics. |
| | | Node 91 | | |
| 22 | AISTViewer | AISTViewer | | The prediction result |
| | | | | is displayed. |
| | | | | |
| | | Node 22 | | |

| 23 | HtmlView | HtmlView | The prediction result is displayed. |
|----|-----------------|-----------------|--|
| 24 | JmolForModeller | JmolForModeller | The tertiary structure of the RNA or protein is displayed by using Jmol. |

or

4.2.2.1 SetVariable

A SetVariable node sets an active flow variable output port by specifying an integer 0, 1, or 2 in this node's Configure dialog. Please specify "an active flow variable output port" in the "Configure" display of the SetVariable node. An active port number of 0 and 1 (2 is unavailable) corresponds to FastaFileReader and Sparql_AIST nodes, respectively.

1) Select the SetVariable icon followed by right-clicking.



4.2.2.1-1 SetVariable icon

2) Select "Configure" from the display menu.

| File |
|---|
| Options Flow Variables Memory Policy Specify an active flow variable output port (0:first; 1:second; 2:third) Active flow variable ouput port number (integer: 0 to 2): 0 ÷ |
| OK Apply Cancel 🕖 |

Select the Active port number 0 for FastaFileReader

| File |
|---|
| Options Flow Variables Memory Policy |
| Specify an active flow variable output port (0:first; 1:second; 2:third) |
| Active flow variable ouput port number (integer: 0 to 2): $1 + \frac{1}{2}$ |
| |
| |
| |
| |
| |
| |
| OK Apply Cancel |

Select the Active port number 1 for Sparql_AIST.

4.2.2.2 Case Switch Variable

This node takes the flow variable 0 or 1 from SetVariable. The parameter is controlled by the previous node.

4.2.2.3 FastaFileReader

An RNA sequence in FASTA format needs to be specified.

1) Select the FastaFileReader icon followed by right-clicking.



4.2.2.3-1 FastaFileReader icon

- 2) Select "Configure" from the display menu.
- 3) Select a file using "Browse" in the pop-up dialog.

| 🛓 Dialog - 0:1 - FastaFileReader | x |
|----------------------------------|---|
| File | |
| | |
| Options Memory Policy | |
| Fasta File | |
| Selected File: | |
| | |
| ▼ Browse | |
| | |
| | |
| | |
| | |
| | |
| | |
| OK - Execute Apply Cancel | |
| | |

4.2.2.3-2 FastaFileReader: Configure

Press "OK" after selecting a file.

CentroidFold predicts the secondary structure of RNA using generalized centroid estimators.

Set an output directory and options.

1) Select the CentroidFold_AIST icon followed by right-clicking.



4.2.2.4-1 CentroidFold_AIST icon

- 2) Select "Configure" from the display menu.
- 3) Select a directory using "Browse" in the pop-up dialog.

| | ● Fasta ○ ClustalW |
|---------------------------|--------------------|
| Output Selected Direct | 21% |
| C:¥ | V Browse |
| Advanced | |
| Other options | |
| | |
| | |

4.2.2.4-2 CentroidFold_AIST: Configure

16

• Output \rightarrow Select Directory \rightarrow Browse:

Select output directory for the calculated results.

- Options tab → Weight of base pairs → Gamma:
 Change the parameter from the pull-down menu if necessary.
 The default value is "2^2".
- Options tab \rightarrow Advanced \rightarrow Other options Enter options if necessary.
- Press "OK" when you have finished entering the required information in the dialog fields.

Please refer to the following website for details about CentroidFold. http://rtools.cbrc.jp/

4.2.2.5 IPknot_AIST

IPknot predicts RNA secondary structures, including a wide range of pseudoknots.

Set an output directory.

1) Select the IPknot_AIST icon followed by right-clicking.



4.2.2.5-1 IPknot_AIST Icon

- 2) Select "Configure" from the display menu.
- 3) Select a directory using "Browse" in the pop-up dialog.

| le | Dialog - 0:12 - IPknot_AIST (Node 26) | _ 🗆 💌 |
|-------------------|---------------------------------------|--------|
| Options | Flow Variables Memory Policy | |
| Select | Dutput Directory | |
| Selec | ted Directory | |
| C:¥ | ~ | Browse |
| Advanc Options | | |
| | OK Apply Cancel | |

4.2.2.5-2 IPknot_AIST: Configure

Press "OK" after completion.

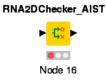
The reference website for details about IPknot is http://rtips.dna.bio.keio.ac.jp/ipknot/

4.2.2.6 RNA2Dchecker_AIST

This node checks whether the RNA 2D structure is suitable for computing tertiary structures by RASSIE and RASCAL. This makes a branch point to perform the prediction efficiently by checking the complexity of the 2D structure. Queries with simple 2D structures will be sent to RASSIE, whereas other queries with complex 2D structures will be sent to Rascal. Queries without any meaningful 2D structure will be aborted.

18

1) Select the IPknot_AIST icon followed by right-clicking.



4.2.2.6-1 RNA2Dchecker_AIST icon

- 2) Select "Configure" from the display menu.
- 3) Select a directory using "Browse" in the pop-up dialog.

| nput type | 1 1 |
|------------------------|-------------------------------|
| | Format Fasta ClustalW |
| | |
| Dutput | |
| | Selected Directory: |
| | /tmp Trowse |
| | |
| Veiaht of | base pairs |
| rengine of | Gamma: 2/2 |
| | |
| Advanced Ither opti | |
| cher oper | |
| | |
| | |
| | |

4.2.2.7 RASSIE_AIST

RASSIE performs rapid predictions of simple structures using fragment structures derived from 2D structures.

Set an output directory.

1) Select the RASSIE_AIST icon followed by right-clicking.



4.2.2.7-1 RASSIE_AIST icon

- 2) Select "Configure" from the display menu.
- 3) Select a directory using "Browse" in the pop-up dialog.

| Options | |
|-------------------|---|
| | Dutput Directory ted Directory |
| C:¥ | V Browse |
| | |
| Options -a 100 |) -ins 100 -clst -outclst 10 -ins chain |
| |) -ins 100 -clst -outclst 10 -ins_chain |
| |) -ins 100 -clst -outclst 10 -ins_chain |

4.2.2.7-2 RASSIE_AIST: Configure

Press "OK" after selecting.

Reference: Yamasaki, S., Nakamura, S. and Fukui, K. (2012) Prospects for Tertiary Structure Prediction of RNA Based on Secondary Structure Information, *J Chem Inf Model*, **52**, 557–567.

4.2.2.8 Rascal_AIST

In Rascal, single strand fragments that are three nucleotides in length are used to solve complicated structures. This method also predicts tertiary structures of RNA–RNA complexes using secondary structure information and a fragment assembly algorithm. Set an output directory.

1) Select the Rascal_AIST icon followed by right-clicking.



4.2.2.8-1 Rascal_AIST icon

- 2) Select "Configure" from the display menu.
- 3) Select a directory using "Browse" in the pop-up dialog.

| Jutput | |
|-------------|---------------------|
| | Selected Directory: |
| | /tmp Browse |
| _ L | |
| dvanced | |
| Specif | options |
| | |
| | |
| | |
| nm File I | Iploader |
| | |
| pm rnc (| |
| pin riic (| Selected File: |
| ,pin the s | |
| ,pin rine (| Selected File: |

4.2.2.8-2 Rascal_AIST: Configure

Press "OK" after selecting.

Reference: Yamasaki, S., *et al.* (2014) Tertiary Structure Prediction of RNA-RNA Complexes Using a Secondary Structure and Fragment-Based Method, *J Chem Inf Model*, **54**, 672–682.

4.2.2.9 RactIP_AIST

RactIP provides services for predicting RNA joint secondary structures under general types of interactions, including kissing hairpins.

Set an output directory.

1) Select the RactIP_AIST icon followed by right-clicking.



4.2.2.9-1 RactIP_AIST icon

- 2) Select "Configure" from the display menu.
- 3) Select a directory using "Browse" in the pop-up dialog.

|)ptions Select (| Flow Variables Memory Policy |
|---------------------|--------------------------------|
| | ted Directory. |
| C:¥ | V Browse |
|)ptions | |

4.2.2.9-2 RactIP_AIST: Configure

Press "OK" after selecting.

The reference site for details about RactIP is http://rtips.dna.bio.keio.ac.jp/ractip/

4.2.2.10 fpocket2_AIST

This node executes fpocket2, which is a protein pocket (cavity) detection algorithm based on Voronoi tessellation.

Set an output directory and options.

- 1) Right-click on the fpocket2_AIST node and specify "Configure".
- 2) Open the below pop-up window and set execution conditions.

| 👃 Dialog - 0:77 - fpocket2_AIST 🛛 🗖 🔜 | |
|---|--|
| File | |
| Options Flow Variables Memory Policy Select Output Directory C:/ Browse | |
| - Advanced option Options | |
| OK Apply Cancel 🕐 | |

4.2.2.10-1 fpocket2_AIST: Configure

Options → Select Output Directory → Selected Directory:
Specify an output directory path in the text field or "Browse".
Options → Advanced option → Options:
Specify execution options in the text field.

Click the "OK" button after setting the conditions.

Reference: Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories, *Bioinformatics*, **27**, 3276–3285.

4.2.2.11 PocketSelector

PocketSelector launches a viewer to select a pocket site. The user can select only one pocket or interaction site on the viewer.

1) Select one of the coordinates.

| Select a row, then click SUB | MIT button. | | Submit | Cancel |
|------------------------------|-------------|--------------|--------------|--------------|
| Pocket | | X-coordinate | Y-coordinate | Z-coordinate |
| 500 | -1.990 | | 6.063 | 14.520 |
| 501 | 4.467 | | -9.310 | 11.795 |
| 502 | -1.815 | | 5.809 | 6.411 |
| 503 | 0.930 | | -20.067 | 4.740 |

4.2.2.11-1 PocketSelector: Submission

2) Select "OK" to submit.

| 🛃 Confi | irm | X | | | |
|---------|--|--------|--|--|--|
| i | i Selected rows will be set to an out-port. OK? | | | | |
| | ОК | Cancel | | | |

4.2.2.12 FragmentSelector

This program selects fragments of RNA in a PDB formatted file.

Set start and end residue (base) numbers for generating a fragment RNA ATOM file.

- 1) Right-click on the FragmentSelector node and specify "Configure".
- 2) Open the below pop-up window and set execution conditions.

| | Dialog - 0:79 - FragmentSelector | x | | | |
|------|--|---|--|--|--|
| File | 2 | | | | |
| | Options Flow Variables Memory Policy | | | | |
| F | Fragment region (start residue (base) number' < 'end residue (base) number') | | | | |
| | start residue (base) number: 0 + | | | | |
| | end residue (base) number: | | | | |
| | OK Apply Cancel 🕐 | | | | |

4.2.2.12-1 FragmentSelector: Configure

• Options \rightarrow Fragment region \rightarrow start residue (base) number: Specify a start residue (base) number.

• Options \rightarrow Fragment region \rightarrow end residue (base) number: Specify an end residue (base) number.

Click the "OK" button after setting the conditions.

4.2.2.13 AutoDockVina_AIST

Set XYZ coordinates and an output directory path.

- 1) Right-click on the AutoDockVina_AIST node and specify "Configure".
- 2) Open the below pop-up window and set execution conditions.

| 🔥 Dialog - 0:81 - AutoDockVina_AIST 🛛 – 🗖 🗙 | | | | |
|--|--|--|--|--|
| File | | | | |
| Options Flow Variables Memory Policy | | | | |
| Docking types | | | | |
| Use binding site coordinates selected by the PocketSelector node | | | | |
| O Specify binding site coordinates in the input boxes below | | | | |
| x-coordinate 12.0 y-coordinate 11.0 z-coordinate 13.5 | | | | |
| Docking Box Size | | | | |
| Use docking box sizes calculated by eBoxSize program | | | | |
| O Specify docking box sizes in the input boxes below | | | | |
| box size x 21.0 box size y 24.5 box size z 25.8 | | | | |
| Output | | | | |
| Selected Directory | | | | |
| C.¥ Browse | | | | |
| | | | | |
| | | | | |
| OK Apply Cancel (| | | | |

4.2.2.13-1 AutoDockVina_AIST: Configure

• Options \rightarrow Docking Types:

There are three radial buttons: 1) Blind Docking, 2) Use binding site coordinates selected by the PocketSelector node and 3) Specify binding site coordinates in the input boxes below. The default setting is 2).

• Options \rightarrow Docking Box Size:

There are two radial buttons: 1) Use docking box sizes calculated by the eBoxSize program and 2) Specify docking box sizes in the input boxes below. The default setting is 1).

• Options \rightarrow Output \rightarrow Selected Directory:

Specify an output directory path in the textbox or "Browse".

Click the "OK" button after setting the conditions.

Reference: Trott O, Olson AJ (2010) Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J Comput Chem*, **31**, 455–461.

4.2.2.14 DockingAnalyzer_AIST

This program analyzes ligand-receptor docking structures using PCA and k-means clustering methods.

Set an output directory path.

- 1) Right-click on the node and specify "Configure".
- 2) Open the below pop-up window and set execution conditions.

| itput Selected | Directory: |
|-------------------|------------------------------|
| /tmp | ▼ Browse |
| | |
| umber o | f clusters |
| inder o | clusters (min:1, max: 4) : 4 |
| | |

4.2.2.14-1 DockingAnalyzer_AIST: Configure

• Options \rightarrow Output \rightarrow Selected Directory:

Specify an output directory path in the textbox or "Browse".

3) Set the number of clusters for k-means.

Click the "OK" button after setting the conditions.

4.2.2.15 Rebuild_AIST

This program rebuilds ligand (RNA) structure models using their fragments and an original structure

Set an output directory path.

- 1) Right-click on the node and specify "Configure".
- 2) Open the below pop-up window and set execution conditions.

| Options Flow Variables Output Selected Directory: | Memory Policy |
|---|---------------|
| /tmp | Browse |
| | |
| ОК Арріу | Cancel |

4.2.2.15-1 Rebuild_AIST: Configure

• Options \rightarrow Output \rightarrow Selected Directory:

Specify an output directory path in the textbox or "Browse". Click the "OK" button after setting the conditions.

4.2.2.16 MinMM_AIST

This program performs energy minimizations using molecular mechanics. If the user connects Rebuild to DockingAnalyzer via flow variables ports (red ports), this program does not open the pop-up window because candidates are already selected for each cluster by DockingAnalyzer, and this is followed by MinMM.

Set an output directory path.

- 1) Right-click on the node and specify "Configure".
- 2) Open the below pop-up window and set execution conditions.

| ptions Flow Varia utput | ables Memory | / Policy | |
|-------------------------------|----------------|----------|--------|
| Selected Directory: - /tmp | | • | Browse |
| | | | |
| | | | |
| | | | |
| | | | |

4.2.2.16-1 MinMM_AIST: Configure

• Options \rightarrow Output \rightarrow Selected Directory:

Specify an output directory path in the textbox or "Browse". Click the "OK" button after setting the conditions.

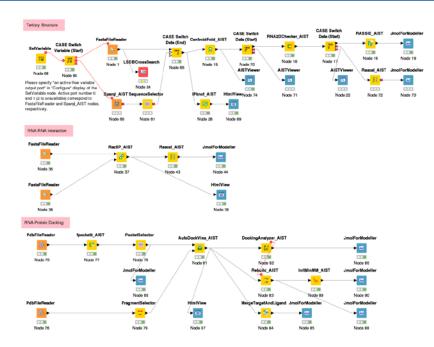
5 Examples for Execution and Result Viewing

The user can download the combination types of RNA workflow packages (AIST-knime packages) from the ToGo-WF website:

(http://togo.medals.jp/active local rna prediction.eng.html).

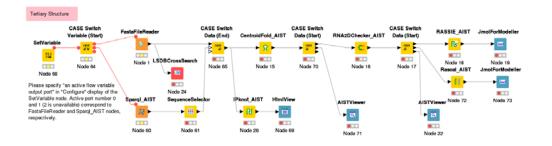
After installation, the workflow of RNA Tertiary Structure Prediction, which performs RNA 3D structure and RNA-RNA/protein interaction predictions, can be used. The configuration of all nodes is initially set to default values for typical calculations.

5.1 Prediction of Tertiary Structures



5.1 Prediction of 3D RNA structures

5.1.1 Prediction of Tertiary Structures using the ReadFastaReader node



Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

1) SetVariable = 0 for ReadFastReader.

| File |
|---|
| Options Flow Variables Memory Policy Specify an active flow variable output port (0:first; 1:second; 2:third) Active flow variable ouput port number (integer: 0 to 2): 1 Active flow variable ouput port number (integer: 0 to 2): 1 - |
| OK Apply Cancel |

2) Set a FASTA file for ReadFastReader by selecting "Configure". FASTA file example:

>1CQ5 GGCGUUUACCAGGUCAGGUCCGGAAGGAAGCAGCCAAGGCGCC

Example of an RNA sequence:

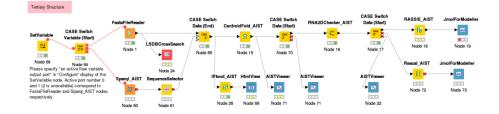
 $http://togo.medals.jp/active_local_rna_prediction.html$

| File |
|--|
| Options Flow Variables Memory Policy Fasta File Selected File: |
| /tmp/RASSIE1cq5.fa |
| OK Apply Cancel 🖓 |

3) Click the CASE Switch Data (Start) node.

5.1.1-1 Execute selected and executable nodes (F7)

The workflow can be stopped at the CASE Switch Data (Start) node and the RASSIE_AIST node is automatically selected for this computation by RNA2Dchecker.

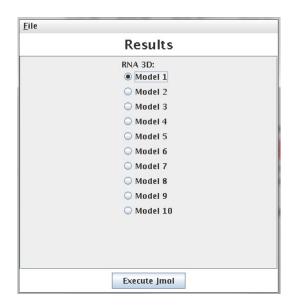


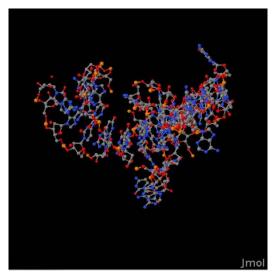
In RASSIE, the fragment structures based on 2D structures are used to predict simple structures rapidly. For this example (PDB ID: 1CQ5), it may take 3–5 minutes. In Rascal, single strand fragments that are three nucleotides in length are used to solve complex structures.

- 4) Execute RASSIE and JmolForModeller nodes.
- 5) <u>JmolForModeller</u>

Results of RASSIE_AIST can be viewed as a 3D structure in JMOL by JmolForModeller.

Please refer to the following website for details about Jmol: http://jmol.sourceforge.net/





5.1.1-2 JmolForModeller –Predicted 3D RNA structure

6) <u>AISTViewer</u>

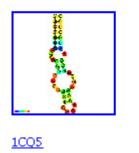
The results of CentroidFold_AIST can be viewed as CentroidFold results by AISTViewer.

A pop-up window appears by clicking the secondary structure predicted in another window.

 File

 URL:
 file:/tmp/2016-04-20/14-18

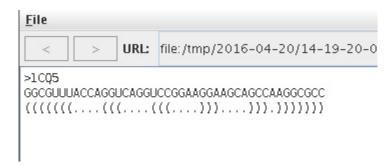
CentroidFold Results



5.1.1-3 AISTViewer - CentroidFold results

7) <u>HTMLViewer</u>

The results of IPknot_AIST can be viewed by HtmlViewer.



5.1.1-4 HTMLViewer – IPknot results

Reference: Yamasaki, S., Nakamura, S. and Fukui, K. (2012) Prospects for Tertiary Structure Prediction of RNA Based on Secondary Structure Information, *J Chem Inf Model*, **52**, 557–567.

Tertiary Structure CASE Switc Data (Start) CASE S CASE S Rs ene T case 2 <mark>۲</mark> (ase 7 • C) Node 65 Node 15 Node 70 Node 16 Node 17 ode 18 rie 19 Node 64 scal AIST + <u>88</u> + of 1 node Ac ive port num HtmlVi de 72 ISTVI ISTV d 1 (2 is u <mark>⊁ (H</mark>) Node 26 Node 61 ode 7 ode 22

5.1.2 Prediction of Tertiary Structures using the Sparql_AIST node

1) SetVariable = 1 for Sparql_AIST.

| File | |
|--|---|
| Options Flow Variables Memory Policy | |
| Specify an active flow variable output port (0:first; 1:second; 2:third) | 1 |
| Active flow variable ouput port number (integer: 0 to 2): $1 + 1$ | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| OK Apply Cancel 🕐 | |

2) Configure for Sparql_AIST.

We provide a sparql endpoint for the KNIME workflow. For the prediction of RNA structures, the endpoint is set to fRNAdb, which is based on http://dbarchive.biosciencedbc.jp/jp/frnadb/desc.html.

| Options Advanced | I Flow Variables Memory Policy | | | |
|---------------------|--|--|--|--|
| Output Directory | | | | |
| Select | ed Directory: | | | |
| Browse | | | | |
| | | | | |
| | | | | |
| SPARQL endpoints | | | | |
| SEVENS | endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/sev | | | |
| 🖌 fRNAdb | endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/sm | | | |
| | | | | |
| UNIPROT | endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/uni | | | |
| PDB | endpoint: http://pdb.bio2rdf.org/sparql | | | |
| _ | | | | |
| 🔲 KEGG - pathwa | y endpoint: http://kegg.bio2rdf.org/sparql | | | |
| SPARQL search condi | tions | | | |
| Taxo | n (not available for UNIPROT) | | | |
| | Keyword | | | |
| | | | | |
| | Minimum sequence length 60 | | | |
| | Maximum sequence length 100 | | | |
| | Resolution (for PDB) 2.8 | | | |
| Dat | thway (for KEGG-pathway) | | | |
| Fa | niway (ibi Kegg-patriway) | | | |
| Output format | | | | |
| | Output format (FASTA' for 'SequenceSelector nod | | | |
| | FASTA | | | |
| | | | | |
| | | | | |

3) SequenceSelector Node.

Select a row and click "Submit" to obtain FASTA data from the fRNAdb endpoint.

| Select a row, then click SUBMIT button. | Submit | Cancel | | |
|---|--|--|--|--|
| ID | Description | Sequence | | |
| FR000001 | Group II intron | UUGAGCCGUAUGCGAUGAAAGUUGCACGUACGGUUCUUUAAGGG . | | |
| FR000004 | transfer RNA (tRNA), CCA (Trp/W) Tryptophan | AGGUCAGUAGUUCCAACGGUAGAACGACAGUCUCCAAAACUGUAU | | |
| FR000006 | suhB | GAACUCAGCCCUUUCGGGCGUUUCCUCCCAAUGACUCGGCCGCC. | | |
| R000009 | Putative conserved nancoding region (EvoFold) | AUGAACUUGUUGUAGUUCUCAU | | |
| R000011 | Piwe-interacting RNA (piRNA) | UCGCCACUUGAUACUCUAUGACAAUGGC | | |
| FR000014 | Piwi-interacting RNA (piRNA) | UGUGCAAACCUCGGGGUGGGGU | | |
| R000015 | Piwi-interacting RNA (piRNA) | UAGAGGAUUCUAUGGCAGGACCAGAAC | | |
| FR000020 | Putative conserved noncoding region (EvoFold) | AUUGCASUUGCUGCUUUGUACAGASGUUACUGCAAU | | |
| FR000022 | Piwi-interacting RNA (piRNA) | UACUUGUUCCAACACCAUCUGAUGGUCAAU | | |
| FR000023 | Putative conserved noncoding region (EvoFold) | GAAUUUUAUCCAAGGGAUAGGGUUC | | |
| R000024 | Piwi-interacting BNA (piBNA) | UGUGGCCUCUGGACGUUACGGGAACU | | |
| R000035 | Piwi-interacting RNA (piRNA) | AAAUAUUCCAGUAGGCCUGAGGUCAAG | | |
| R000036 | transfer RNA (tRNA), TTG (Gln/Q) Glutamine | UAGGGGGUAGUAUAAUUGGAAGUACUAAAAAUUUUGAAUUUUUAA | | |
| FR000037 | HIV gag stem loop 3 (GSL3) | GGCCAGGGGGGAAAGAAACACUAUAUGCUAAAACACCUAGUAUGGG | | |
| FR000038 | Putative conserved nancoding region (EvoFold) | GGAGGUUUGCUCGGAAACUUCC | | |
| R000039 | transfer RNA (tRNA), CAT (Met/M) Methionine | GGCUGAAUAGUUUAAUUGGUUAAAACAUCGGAUUCAUAAGCCGGG | | |
| FR000042 | Piwi-interacting RNA (piRNA) | UCUGGCCUUUGGACAGUUGUGGUGUC | | |
| R000044 | Putative conserved noncoding region (EvoFold) | CUUAGCUAUGUGCAUAGUUUAAG | | |
| R000047 | Piwi-interacting RNA (piRNA) | GGAUUUCACUGGAAUCGGUCAUAAAAA | | |
| R000048 | Putative conserved nancoding region (EvoFold) | CCAAUUAAAUUUGUGUAAUUGG | | |
| R000049 | transfer RNA (tRNA), TTG (Gln/Q) Glutamine | UAGGAGAUAGAAUAAUGGAGUUCUAAAGAUUUUGAGUCUUUAUGU | | |
| FR000050 | Putative conserved noncoding region (EvoFold) | AUUUUUUGAACCAAGCAGAAAAAU | | |
| R000053 | Piwi-interacting RNA (piRNA) | UGGGAUGUUUGAUUUUAGCUUCCUCUCCGU | | |
| R000054 | Putative conserved noncoding region (EvoFold) | UUAUUUAAUAUGUAAAUUGUAUUGCUAUACAUAAAAUAA | | |
| R000056 | Putative conserved noncoding region (EvoFold) | AGAUGGAUGAGUUGGAUCCAUCU | | |
| R000059 | small nucleolar RNA (snoRNA) 2186 | GGCCGGUGAUGACAAGACCAUUGUCAUACUAGAUAUUGAUCAUAU | | |
| R000060 | Piwi-interacting BNA (piBNA) | UGUUGAAAAUGUGCCUCCUGGAGACAGC | | |
| FR000061 | transfer RNA (tRNA), TTC (Glu/E) Glutamic acid | GGCUCCUUGGUCAAGCGGUUAAGACACOGCCCUUUCACGGCGGU. | | |
| IRABARAS. | A 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 | ALCHINE AND ALCHING ALCHI | | |

For this example, FR000004 is selected for further calculations. The structure prediction is difficult for this example and takes time. The computation required for this example may be considered to be close to our system limit.

4) Click the CASE Switch Data (Start) node.

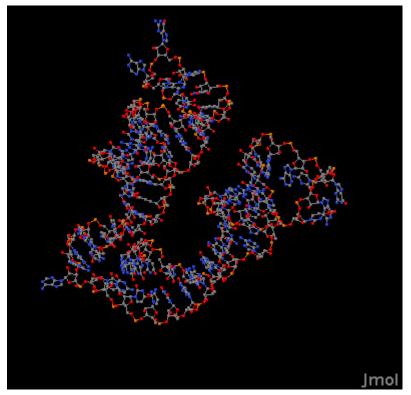
5.1.2-1 Execute selected and executable nodes (F7)

The workflow can be stopped at the CASE Switch Data (Start) node and the Rascal_AIST node is automatically selected for this computation by RNA2Dchecker. In RASSIE, the fragment structures derived from 2D structures are used to predict simple structures rapidly. In Rascal, single strand fragments that are three nucleotides in length are used to solve complex structures. There are essentially no experimental tertiary structure examples available for structures with long chain loop structures and single-strands. Thus, in principle, RASSIE cannot predict these tertiary structures. For such targets, tertiary structure prediction is performed by Rascal. For this example (FR000004), <u>it may take about two hours</u>. During execution, check the status of the node ("in progress") and do not close KNIME before the status reads "Complete".

- 5) Execute RASSIE and JmolForModeller nodes.
- 6) <u>JmolForModeller</u>

Results of RASSIE_AIST can be viewed as a 3D structure in JMOL by using JmolForModeller.

| File | |
|------|--------------------|
| | Results |
| | Rascal: Model 1 |
| | O Model 2 |
| | 🔾 Model 3 |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | Execute Jmol |
| | |

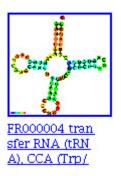


5.1.2-2 JmolForModeller –Predicted 3D RNA structure

7) <u>AISTViewer</u>

The results of CentroidFold_AIST can be viewed as CentroidFold results by AISTViewer.

A pop-up window appears by clicking on the secondary structure predicted.



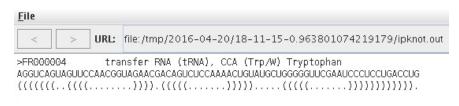
CentroidFold Results

5.1.2-3 AISTViewer - CentroidFold results

8) HtmlView

The results of IPknot_AIST can be viewed by HtmlViewer.

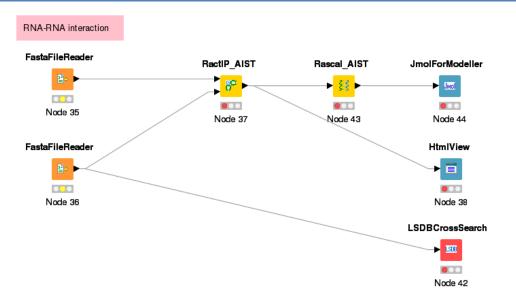
The results of Raccess_AIST can be viewed as text by HtmlView.



5.1.2-3 HtmlViewer – CentroidFold results

Reference: Yamasaki, S., *et al.* (2014) Tertiary Structure Prediction of RNA-RNA Complexes Using a Secondary Structure and Fragment-Based Method, *J Chem Inf Model*, **54**, 672–682.

5.2 Prediction of RNA-RNA interactions



5.2 Prediction of RNA-RNA interactions

1) Set a FASTA file for each ReadFastReader by selecting "Configure".

FASTA file example: >1XP7:A | PDBID | CHAIN | SEQUENCE

CUUGCUGAAGUGCACACAGCAAG

| File |
|---|
| Options Flow Variables Memory Policy Fasta File Selected File: /tmp/1xp7_afasta Browse |
| OK Apply Cancel 🕐 |

2) Set a FASTA file for ReadFastReader by selecting "Configure".

FASTA file example: >1XP7:B|PDBID|CHAIN|SEQUENCE CUUGCUGAAGUGCACACAGCAAG

Example of an RNA sequence:

 $http://togo.medals.jp/active_local_rna_prediction.html$

| File |
|--------------------------------------|
| Options Flow Variables Memory Policy |
| Fasta File |
| Selected File: |
| /tmp/1xp7_b.fasta 💌 Browse |
| |
| |
| |
| |
| |
| |
| OK Apply Cancel 🕐 |
| |

3) Click the RactIP_AIST node.

0

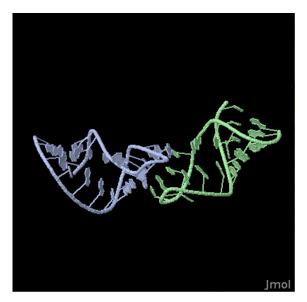
5.2-1 Execute selected and executable nodes (F7)

- 4) Execute Rascal and JmolForModeller nodes.
- 5) <u>JmolForModeller</u>

Results of RASSIE_AIST can be viewed as a 3D structure in JMOL by JmolForModeller.

Please refer to the following website for details about Jmol: http://jmol.sourceforge.net/

| Results | |
|--------------------|--|
| Rascal: Model 1 | |
| O Model 2 | |
| Model 3 | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |



5.2-2 JmolForModeller – Predicted 3D RNA structure

6) <u>HTMLViewer</u>

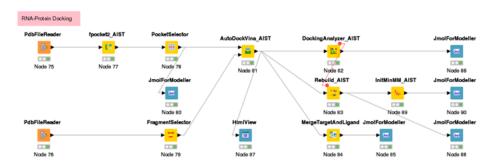
The results of IPknot_AIST can be viewed by HtmlViewer.

| < > URL: file:/tmp/2016-04-20/21-32-54-0.20 | 136396 | 3826925 |
|--|--------|---------|
| GLPK Simplex Optimizer, v4.51 | | |
| 198 rows, 32 columns, 271 non-zeros | | |
| * 0: obj = 0.000000000e+00 infeas = 0.000e+00 | | |
| * 26: obj = 1.075252366e+01 infeas = 0.000e+00 | (0) | |
| OPTIMAL SOLUTION FOUND | | |
| GLPK Integer Optimizer, v4.51 | | |
| 198 rows, 32 columns, 271 non-zeros | | |
| 32 integer variables, all of which are binary | | |
| Integer optimization begins | | |
| | | |
| + 26: >>>> 1.075252366e+01 <= 1.075252366e+01 | | (1; 0) |
| ⊦ 26: mip = 1.075252366e+01 <= tree is empty | 0.0% | (0; 1) |
| INTEGER OPTIMAL SOLUTION FOUND | | |
| >1XP7:A PDBID CHAIN SEQUENCE | | |
| CUUGCUGAAGUGCACACAGCAAG | | |
| (((((((,,,[[[[,]))])))))))))))) | | |
| >3agv | | |
| GGAGGUGCUCCGAAAGGAACUCCA | | |
| .((.]]]]((())))) | | |

5.2-2 HtmlViewer – IPknot results

Reference: Yamasaki, S., *et al.* (2014) Tertiary Structure Prediction of RNA-RNA Complexes Using a Secondary Structure and Fragment-Based Method, *J Chem Inf Model*, **54**, 672–682.

5.3 Prediction of Protein-RNA interactions



1) Set a PDB file for PdbFileReader by selecting "Configure".

Select a protein PDB file.

| PdbFileReader | fpocket2_AIST | PocketSelector |
|-----------------------|-------------------------|----------------|
| Node 75 | Node 77 | Node 78 |
| 🛆 Dialog - 0:75 - Pdb | FileReader | |
| File | | |
| Options Flow | Variables Memory Policy | |
| Selected File: | | |
| /tmp/protein. | pdb | ▼ Browse |
| | | |
| | | |
| | | |
| | | |
| | | |

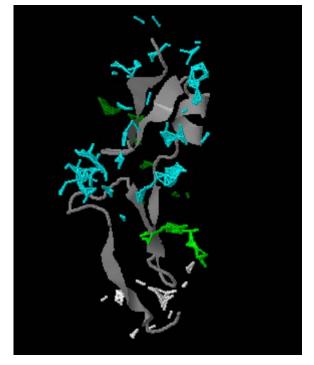
2) Search for interaction sites of the uploaded PDB file using Fpocket2_AIST.

| fpocket2 results | |
|------------------|--|
| Model: 1 | |
| | |
| | |
| | |

| Pocket | X-coordinate | Y-coordinate | Z-coordinate |
|-------------|--------------|--------------|--------------|
| SOO | -1.990 | 6.063 | 14.520 |
| SO1 | 4.467 | -9.310 | 11.795 |
| \$02 | -1.815 | 5.809 | 6.411 |
| SO 3 | 0.930 | -20.067 | 4.740 |

Execute Jmol after the fpoeck2 calculation is finished.

The colors of the coordinates correspond to the positions in the PDB.



5.3 JmolForModeller – Fpocket2

3) Select one of docking/interacting positions.

| Select a row, then click SU | | Submit | | Cancel | |
|-----------------------------|--------|--------|-------------|----------------|--|
| Pocket | X-coor | dinate | Y-coordinat | te Z-coordinat | |
| S00 | -1.990 | | 6.063 | 14.520 | |
| S01 | 4.467 | | -9.310 | 11.795 | |
| S02 | -1.815 | | 5.809 | 6.411 | |
| S03 | 0.930 | | -20.067 | 4.740 | |

| 🛃 Conf | irm | × | |
|--------|--|--------|--|
| i | i Selected rows will be set to an out-port. OK | | |
| | ОК | Cancel | |

Click "OK" to submit the selected position for AutoDockVina.

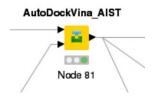
4) Select an RNA PDB file.

| DB File Selected | File: | | |
|---------------------|--------|-------|--------|
| /tmp/r | na.pdb | • | Browse |
| | | | |
| | | | |
| | | | |

5) Set the fragment of RNA for docking calculations.

| Options Flow Variables Memory Policy |
|---|
| Fragment region (start residue (base) number' < 'end residue (bas |
| start residue (base) number: 11 🗧 |
| end residue (base) number: 13 |
| OK Apply Cancel 🕐 |

6) Run docking calculations for both the protein and RNA fragment.



7) Merge the docking results for the RNA fragment and protein.

| AutoDock-VINA results |
|-----------------------|
| Model 1 |
| 🔾 Model 2 |
| 🔾 Model 3 |
| 🔾 Model 4 |
| 🔾 Model 5 |
| 🔾 Model 6 |
| 🔾 Model 7 |
| Model 8 |
| 🔾 Model 9 |
| Model 10 |
| 🔾 Model 11 |
| O Model 12 |
| 🔾 Model 13 |
| 🔾 Model 14 |
| 🔾 Model 15 |
| 🔾 Model 16 |
| Model 17 |
| Model 18 |
| Model 19 |
| O Model 20 |
| Execute Jmol |

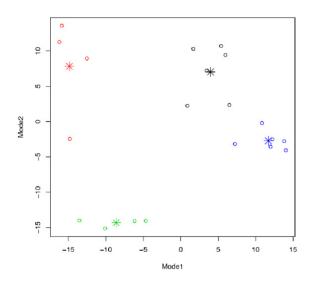
Twenty docking poses for the RNA fragment-protein complex.

Select one of the 20 poses and execute Jmol to view the structure.

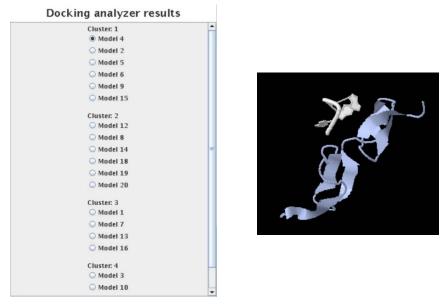


5.3 AutoDock Vina – Docking structure of RNA fragment and protein

Analyze the docking outputs by clustering the results.
 The results can be viewed by using HtmlViewer and JmolForModeller.

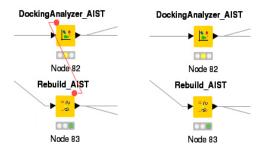


5.3 Analyzer - PCA analysis and clustering of structures of the RNA fragment and protein

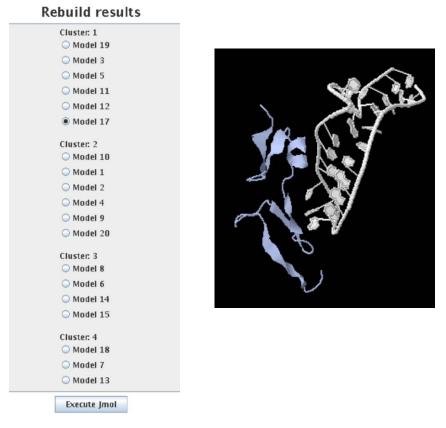


5.3 Analyzer – Tertiary clustering structures of the RNA fragment and protein

When you link the nodes between "DockingAnalyzer" and "Rebuild" (red line), the results of the docking analysis by "DockingAnalyzer" is passed to the "Rebuild" node. Analysis of the structures using your own approach can be achieved by deleting the red line. Thus, all docked results are simply rebuilt without clustering the structures.



9) Rebuild the RNA structure based on the fragment poses.

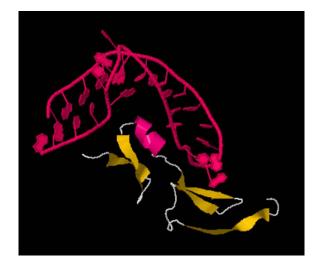


5.3-5 Rebuild –Rebuilt structures of the RNA-protein complex

10) Minimize the structure.

Minimization takes time and is dependent on the size of the structure of the complex.

During execution, check the status of the node ("in progress") and do not close KNIME before the status reads "Complete".



5.3 MinMM – Energy minimized structures of RNA and protein

6 SPARQL node

SPARQL nodes execute SPARQL against endpoints and there are two nodes.

6.1 Sparql related node

6.1.1 Node list

There are two SPARQL nodes and two relative nodes as follows:

| No | Name | Icon | Set | Explanation |
|----|-------------|-------------|-----|----------------|
| 1 | Sparql_AIST | Sparql_AIST | Y | Execute |
| | | | | SPARQL simply |
| | | BJ | | against |
| | | | | endpoints AIST |
| | | | | released. |

6.1.1-1 SPARQL node list

| 2 | Sparql_Adv | Sparql_AIST_Adv | Y | Execute |
|---|-------------------|------------------|---|------------------------------|
| | | 53 - | | SPARQL simply against |
| 3 | SequenceSelector | SequenceSelector | | endpoints. Display SPARQL |
| 0 | Sequences elector | | | results and |
| | | | | select a sequence. |
| 4 | HtmlView | | | Display SPARQL |
| | | | | results. |

6.1.2 Sparql_AIST

6.1.2.1 Setting the execution environment

1) Select the Sparql (Simple) icon and right-click.

Sparql_AIST



6.1.2.1-1 Sparql (Simple) icon

- 2) Select "Configure" from the menu.
- 3) Setup execution conditions.

| 🛓 Dialog - 0:114 - Sparql (Simple) |
|---|
| File |
| Options Advanced Flow Variables Memory Policy Output Directory |
| Selected Directory. |
| C:¥ |
| SPARUL endpoints |
| SEVENS endpoint http://terdf.cbrc.jp/openrdf-sesame/repositories/sev |
| fRNAdb endpoint: http://lod.dbcls.jp/openrdf-sesame/repositories/frna |
| UNIPROT endpoint http://lod.dbcls.jp/openrdf-sesame/repositories/unip |
| PDB endpoint: http://pdb.bio2rdf.org/sparql |
| KEGG - pathway endpoint: http://pdb.bio2rdf.org/sparql |
| SPARQL search conditions |
| Taxon (not available for UNIPROT) |
| Keyword |
| Minimum sequence length 300 |
| Maximum sequence length 600 |
| Resolution (for PDB) 2.8 |
| Pathway (for KEGG-pathway) small |
| Output format |
| Output format ('FASTA' for 'SequenceSelector node') |
| FASTA |
| |
| OK Apply Cancel |
| |

6.1.2.1-2 Sparql (Simple): Configure

• Options tab \rightarrow Output Directory:

Red frame: specify an absolute path of a directory to store a SPARQL result file.

• Options tab \rightarrow SPARQL Endpoint:

Blue frame: select a SPARQL endpoint. The user can select endpoints as follows: SEVENS (GPCR database: http://sevens.cbrc.jp/) fRNAdb (functional RNA database: https://dbarchive.biosciencedbc.jp/jp/frnadb/desc.html) UNIPROT (only human data, http://www.expasy.org/) PDB (100% non-redundant, x-ray, protein data, http://www.rcsb.org/) KEGG pathway (pathway: http://www.genome.jp/kegg/pathway.html

• Options tab \rightarrow SPARQL search conditions:

Green frame: specify SPARQL search conditions as follows:

- 1) Taxon (except for UNIPROT and the KEGG pathway)
- 2) Keyword (except for the KEGG pathway)
- 3) Sequence length (max, min) (except for the KEGG pathway
- 4) Resolution (PDB only)
- 5) Pathway (only the KEGG pathway)

• Options tab \rightarrow Output format:

Orange frame: select either "FASTA" or "Tab-limited".

FASTA – generates a result file as FASTA format.

Tab-limited – generates a result file as Tab-limited format.

Please select "FASTA" if you are going to use SequenceSelector.

• Options tab \rightarrow Advanced:

Green frame: specify SPARQL query. If you specify, the above search conditions are neglected.

Then, click the "OK" button.

6.1.3 Sparql_AIST_Adv.

6.1.3.1 Setting the execution environment

1) Select the Sparql (Simple) icon and right-click.



6.1.3.1-1 Sparql_AIST_Adv. icon

2) Select "Configure" from the menu.

| Options Flow Variables | Memory Policy |
|------------------------------|--|
| Output Directory | elected Directory: |
| | tmp Browse |
| -SPARQL endpoint | ndpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/sm |
| SPARQL Input SPARQL Query | |
| SELECT * WHERE { | |
| | |
| | |
| | OK Apply Cancel 🕡 |

• SPARQL endpoint

Specify endpoint.

Examples:

http://tgrdf.medals.jp/openrdf-sesame/repositories/sevens

http://tgrdf.medals.jp/openrdf-sesame/repositories/smallRNAdb

http://tgrdf.medals.jp/openrdf-sesame/repositories/uniprot-reviewed-hum and

http://pdb.bio2rdf.org/sparql

http://kegg.bio2rdf.org/sparql

Input SPARQL query

Specify SPARQL query. Then, click the "OK" button.

6.1.4 SequenceSelector

6.1.4.1 Result

By right-clicking the SequenceSelector node and selecting "Execute" from the menu, the SequenceSelector node is executed and SPARQL results are displayed. This node is available for Sparql (AIST) and Sparql(Simple) nodes if user selects "FASTA" in the search condition at the nodes.

SequenceSelector



6.1.4.1-1 SequenceSelector icon

| select a row, then click SUBMIT button. | Submit | Cancel |
|---|----------------------|--|
| ID | Description | Sequence |
| CBRC-HSAP-02-0037 | Adrenergic receptors | MDHQDPYSVQATAAIAAAITFLILFTIFGNALVILAVLTSRSLF |
| BRC-HSAP-04-0004 | Adrenergic receptors | MASPALAAALAVAAAAGPNASGAGERGSGGVANASGASW |
| BRC-HSAP-05-0044 | Adrenergic receptors | MGQPGNGSAFLLAPNGSHAPDHDVTQERDEVWVVGMG |
| BRC-HSAP-05-0047 | Adrenergic receptors | MNPDLDTGHNTSAPAHWGELKNANFTGPNQTSSNSTL |
| BRC-HSAP-08-0018 | Adrenergic receptors | MVFLSGNASDSSNCTQPPAPVNISKAILLGVILGGLILFGV |
| BRC-HSAP-10-0036 | Adrenergic receptors | MGSLQPDAGNASWNGTEAPGGGARATPYSLQVTLTLVC |
| BRC-HSAP-10-0037 | Adrenergic receptors | MGAGVLVLGASEPGNLSSAAPLPDGAATAARLLVPASPP |
| BRC-HSAP-20-0001 | Adrenergic receptors | MTFRDLLSVSFEGPRPDSSAGGSSAGGGGGSAGGAAP |
| | | |
| | | |

6.1.4.1-2 SequenceSelector results

The user can select sequence information, and if the "Submit" button is selected (i.e., clicked) the absolute path of a results file stored by sequence information is setup and sent to the out-port of the SequenceSelector node. By connecting the out-port to the in-port of the sequence analysis nodes, the user can execute sequences obtained in the SPARQL results.

6.1.5 HtmlView

6.1.5.1 Result

The HtmlView node is executed by right-clicking the HtmlView node and selecting "Execute" from the menu. This node is available for Sparql (AIST) and Sparql (Simple) and can display results.





6.1.5.1-1 HtmlView icon

| A Html View - 0:121 - HtmlView (Node 6) |
|---|
| Eile |
| URL: file:C:/2013-06-26/18-34-54-0.8906006612134388/sparqlRes.fa |
| <pre>>CERC-HSAP-02-0037 Adrenergic receptors MHODPYSOATAAIAAAITELIAFIIFONALVILAVIJSKSIAAPONLFLVSLAAADILVAILIIPFSLANELLGYWYFRRTWCEVYLALDVLFCT GSFFAPCLIMILVYLRIYLIAKRSNRRGFRAKGGPQQESKQPRPDHGGALASAKLPALASVASAREVWGHSKSTGEKEEGETPEDIGTRALPPSNA RAQLTREKRFTFVLAVVUGVUCWFFFFFSYSLGICLCEKHCKVPHELOFFFWIGYCNSSLNPVIYTIFNQFRAFRRILCPWTQTAWEMKGFA CGRC-HSAP-04-0004 Adrenergic receptors MASPALAAALAVAAAAGNAAGAGERGSGGVANASGASWGPPSGVSAGAVAGLAAVVGFLIVFTVVGNVLVIAVLTSRALRAPONLFLVSLASAD SAVISFPIVSLYNQPDGAATPQCGINDETWYLLSSCIGSFFAPCLIMGLVYARIYKVAKLRTHTLSEKRAFVQPDGASPTIENGLGAAAGAGENCH LSRRRARSSVCRKVAQAREKRTFVLAVVMGVGVGVGMGIVMSLUVLATVGRVLVITAILAKRERLQTVTHYPITSLACADLVMGLAVVPFGAAHILKK QCEAINCYANETCCDFFTNQAYAIASSIVSFYVELVIMVFVSKVFQEAKGLQKIDKSEGRFHVONLSQVEQDGRTGHGLRRSSKFCLKEHKALKTL MGNTGEGSSYNVEQEKENNLCEDLFGTDFVCHQGTVPSDNIDSQGRNCSTNDSLL XCBRC-HSAP-05-0047 Adrenergic receptors MNPDDLOFGNUTSSAPLMWGELKNANFGFGNSSNSTLPGLDITHAISVGLVLGAFILFAIVGNILVILSVACNRHLRFPTNYFIVNIAMADLLLSTT GPLLGMKEPAPNDDKCGVTEEPTALFSSLGSFYTPLAVINVFVVSKVFQEAKGLQKIDKSEGRFHVONLSQVEQDGRTGHGLRRSSKFCLKEHKALKTL MRNTEQGSSYNVEQEKENNLCEDLFGTDFVCHQGTVPSDNIDSQGRNCSTNDSLL CCBRC-HSAP-05-0047 Adrenergic receptors MNFDLDTGHUKGGRRGRRRRRRREGGAYTYRDWTRGSLERSQSRKDSLDDSGSCLSGSORILFSASPSGYLGRGAPPVELCAFPEWKAPGAL 2CBRC-HSAP-08-0018 Adrenergic receptors MVFLISGNASDSSNCTOPPAPVNISKAILLGVLIGGLLIGGVLINIVILSVACHRHLHSVHYIVINLAVADLLISTVLPFSAIFEVLGYWAFGK SCBRC-HSAP-08-0018 Adrenergic receptors MSLQFDAGNASMWGTLAFGGGAARTPYSLQVTILTWVLXVSKSKSLKSLKTLRHKNAPAGGSGRASANTHTHFSVLLKFSBRKKAAKTLG GYTLHPSSOAVEGOKHKMVRTVVSKETTYHISKDCVCEWKFFSSMFRGSARITISKDQSSCTTARVRSKSFLQVCCCVGPSFSLDRNHQVPTIK 2CBRC-HSAP-10-0036 Adrenergic receptors MGSLQFDAGNASMWGTLAFGGGAARTPYSLQVTILIVVLAKKNTRVPESSMERGSARITIFKSVGLVSASGUCCSGFGRERARARTHRFTWSLKKAARTUG GYTLHPSSOAVEGOKHKWKTVVLAVVGFFFFTTYLIAKCSVFFSSMFRGSARITIFKSVFVLCIMASSLLKFSBRKKAARTUG GYTLHPSDAVEGOKNDWRTPSDFRARGAAAPLPSDFARCHNDFYLCUFFFTFWFGYCNSSLSLPVLITINGAASANTHFFSVLAKKAARTUG CCTASILSLCTISVDRYGVRHEKXAEDARCYNDPKCCCFVTNRAVALISSVSFTVPLCIMAFVHYLVIAAPSLLSABDILVATLVIFFSLAFFFF MGSQUVUGASEPGNLSAAPLPDGAATPSD</pre> |
| |
| |
| • |

6.1.5.1-2 HtmlView results

7 Appendix

7.1 Appendix A : LSDBCrossSearch

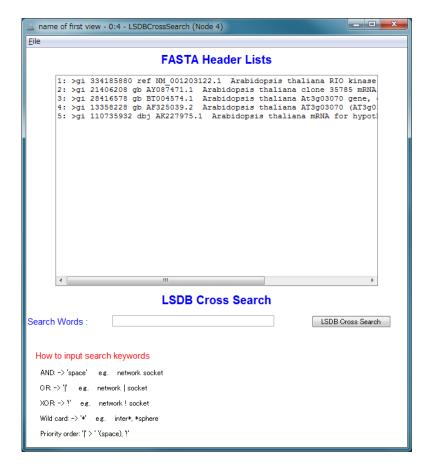
A Life Science Database cross-search can be executed in the green node status after executing the LSDBCrossSearch node.

LSDBCrossSearch



The Life Science Database cross-search site was developed as part of the Database Integration project, which was promoted by the Ministry of Education, Culture, Sports, Science and Technology.

If "View" is selected in the right-click-menu of the LSDBCrossSearch node, the View window of the LSDBCrossSearch node will appear.



7.1-1 LSDBCrossSearch View window

Headers of the FASTA file used for the LSDBCrossSearch node are shown in FASTA Header Lists.

A keyword(s) for cross-search should be entered in the text box.

For a combined search, the following symbols should be used:

·AND retrieval: Space " "

•OR retrieval: Pipe "| "

·Exclusive-OR retrieval: Exclamation mark ""

•Wildcard search: Asterisk "* "

OR has the highest priority.

Cross-search will be carried out by clicking the LSDB Cross Search button, and a Web browser of the Life Science Database cross-search will appear as shown below.



7.1-2 LSDB window

Please refer to the Life Science Database cross-search site for details. Life Science Database website: <u>http://biosciencedbc.jp/dbsearch/</u>

8 Contact

Please send your queries or comments to the email address below. workflow@molprof.jp

Molecular Profiling Research Center for Drug Discovery of AIST will listen to requests made by users positively and aims to make the system better.

Molecular Profiling Research Center for Drug Discovery (MolProf) Advanced Industrial Science and Technology (AIST) http://togo.medals.jp

AIST Tokyo Waterfront Bio-IT Research Building 2-4-7 Aomi, Koto-ku, Tokyo, 135-0064, Japan

9 Terms and Conditions of License for Use



RNA Structure Prediction is licensed under a <u>Creative Commons Non-commercial 4.0 International License</u>.

If you want to take advantage of this workflow for commercial purposes, please contact us for details (<u>workflow@molprof.jp</u>).