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 Introduction 3
2 About the workflow of RNA related analysis 4
3 Common rules 5
4 Usage of RNA Structural Prediction 9
4.1 Preparation 9
4.2 Node 9
4.2.1 Node list 10
4.2.2 Node setting 14
4.2.2.1 SetVariable 14
4.2.2.2 Case Switch Variable 15
4.2.2.3 FastaFileReader 15
4.2.2.4 CentroidFold_AIST 16
4.2.2.5 IPknot_AIST 17
4.2.2.6 RNA2Dchecker_AIST 18
4.2.2.7 RASSIE_AIST 19
4.2.2.8 Rascal_AIST 20
4.2.2.9 RactIP_AIST 21
4.2.2.10 fpocket2_AIST 22
4.2.2.11 PocketSelector 23
4.2.2.12 FragmentSelector 24
4.2.2.13 AutoDockVina_AIST 25
4.2.2.14 DockingAnalyzer_AIST 27
4.2.2.15 Rebuild_AIST 27
4.2.2.16 MinMM_AIST 28
5 Examples for Execution and Result Viewing 30
5.1 Prediction of Tertiary Structures 30
5.1.1 Prediction of Tertiary Structures using the ReadFastaReader node 30
5.1.2 Prediction of Tertiary Structures using the Sparql_AIST node 35
5.2 Prediction of RNA-RNA interactions 39
5.3 Prediction of Protein-RNA interactions 43
6 SPARQL node 49
6.1 Sparql related node 49
6.1.1 Node list. 49
6.1.2 Sparql_AIST 51
6.1.2.1 Setting execution environment 51

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

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

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

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/

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

2 About the workflow of RNA related analysis

2-1 Active workflow combination type list

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

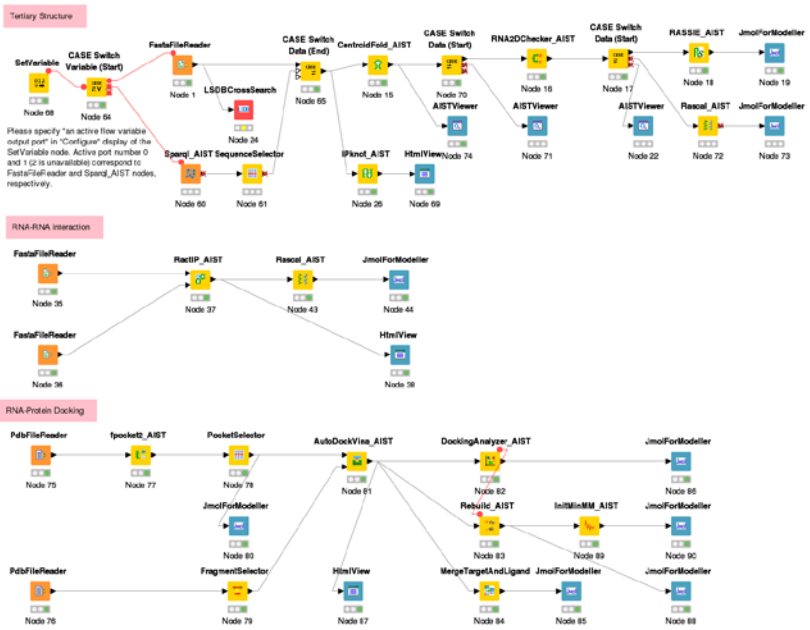
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

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;

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.

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

3. Node progress

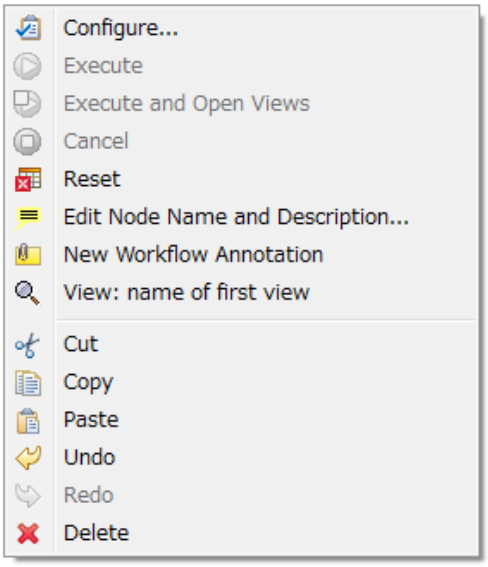
Signals below a node indicate the progress status, as shown below.

3-3 Signal of node progress list

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

4. Node menu

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



3-4-Node menu

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

3-4-2 Node menu list

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 that displays the result window. Execute a node.	The node cannot be used unless the node status is yellow.
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 Description…	Use to change the node name or description.	Another window is 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.	–
Copy	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.	–

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

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

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:

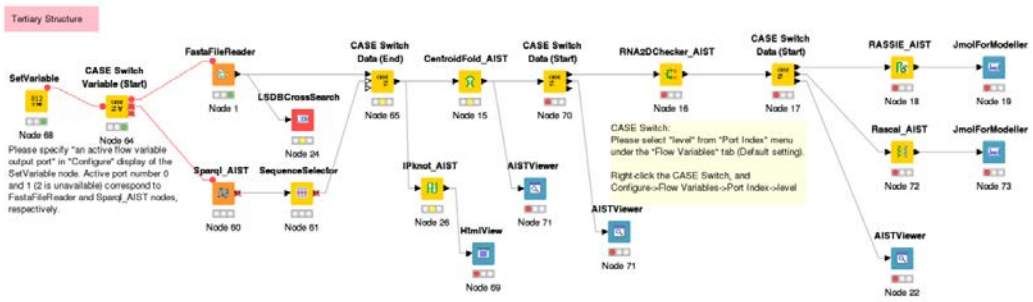
3-6 Alert messages (examples)

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

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

4 Usage of RNA Structural Prediction

The user can download combination types of RNA workflow packages (AIST-knime packages) from the TOGO web site: (http://togo.medals.jp/active_local_rna_prediction.eng.html). 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.

Preparation file
FASTA format file

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

Preparation file
PDB format file

4.2 Node

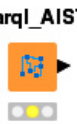
There are various nodes in the RNA Structure Prediction Active Workflow. The nodes are shown below and should be configured accordingly.

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

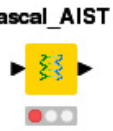
4.2.1 Node list

※ In the following list, “Y” indicates a node where the setting is necessary.

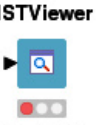
4.2.1-1 RNA_Structure_Prediction Active workflow node list

No	Node name	Icon	set	explanation
1	SetVariable		Y	Output port by specifying integer 0, 1, or 2 in this node's Configure dialog.
2	Case Switch Variable			The flow variables from the input port are moved to exactly one active output branch.
3	FastaFileReader		Y	The FASTA format file is read.
4	LSDBCrossSearch		-	LSDB cross-search is executed.
5	Sparql_AIST		Y	SPARQL searches for endpoint (microfRNAdb)
6	SequenceSelector			Selection of sequence executed

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

7	CentroidFold_AIST		Y	CentroidFold is executed.
8	IPknot_AIST		Y	IPknot is executed.
9	RNA2DChecker_AIST		Y	RNA2DChecker is executed.
10	RactIP_AIST		Y	RactIP is executed.
11	RASSIE_AIST		Y	RASSIE is executed.
12	Rascal_AIST		Y	Rascal is executed.
13	fpocket2_AIST		Y	Execute fpocket2.
14	PocketSelector			Select a pocket from the fpocket2 results.

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

15	FragmentSelector		Y	Generate a PDB ATOM file that consists of the RNA fragment selected by the user.
16	AutoDockVina_AIST		Y	Execute AutoDock Vina.
17	MergeTargetAndLigand			Merge receptor (protein) and ligand (RNA) file.
18	DockingAnalyzer_AIST		Y	Execute cluster analysis (k-means) and principal component analysis (PCA)
19	Rebuild_AIST		Y	Rebuild an RNA ligand using a complete RNA structure.
20	MinMMCandidateSelector			Selection of one model for energy minimization.
21	MinMM_AIST		Y	Energy minimization by using molecular mechanics.
22	AISTViewer			The prediction result is displayed.

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

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

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

4.2.2 Node setting

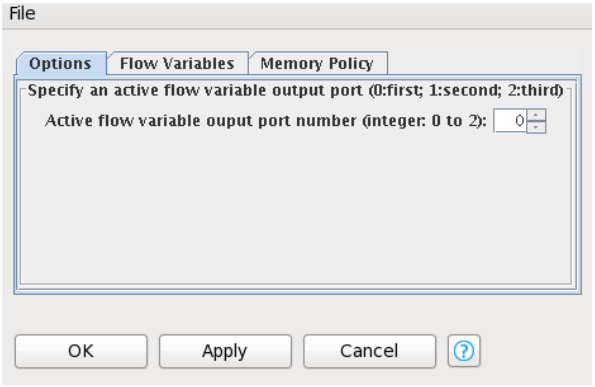
4.2.2.1 SetVariable

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.

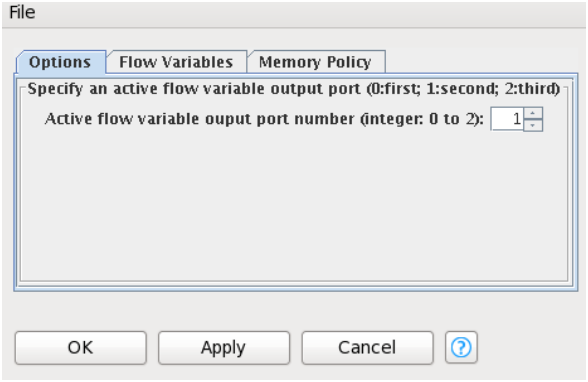
4.2.2.1-1 SetVariable icon

2) Select “Configure” from the display menu.



Select the Active port number 0 for FastaFileReader

or



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

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

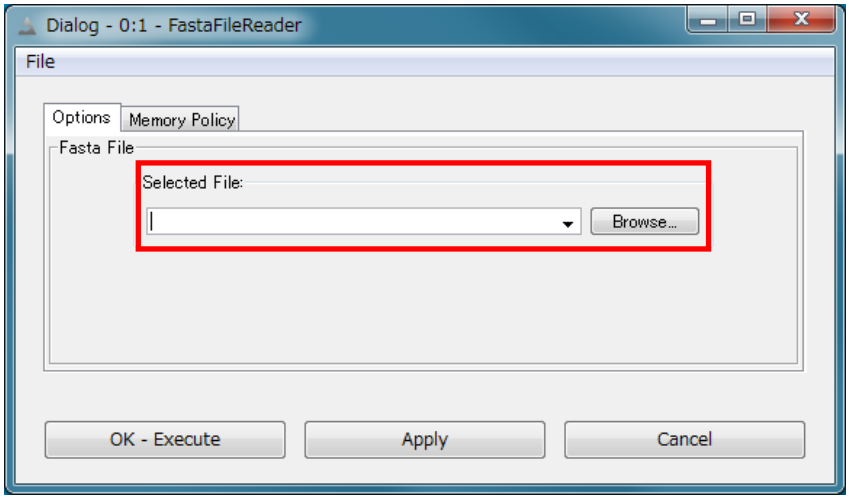
An RNA sequence in FASTA format needs to be specified.

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.



4.2.2.3-2 FastaFileReader: Configure

Press “OK” after selecting a file.

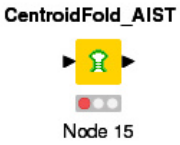
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

4.2.2.4 CentroidFold_AIST

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

Set an output directory and options.

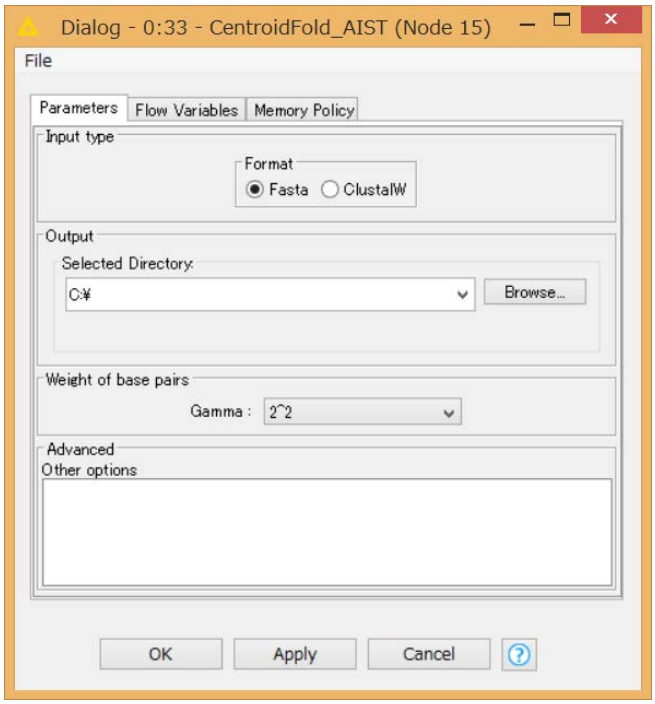
1) Select the CentroidFold_AIST icon followed by right-clicking.



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.



4.2.2.4-2 CentroidFold_AIST: Configure

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

- **Output → Select Directory → 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 → Advanced → Other options**
  Enter options if necessary.

4) Press “OK” when you have finished entering the required information in the dialog fields.

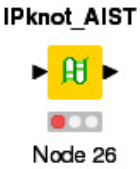
Please refer to the following website for details about CentroidFold.
http://rtools.cbrc.jp/

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.

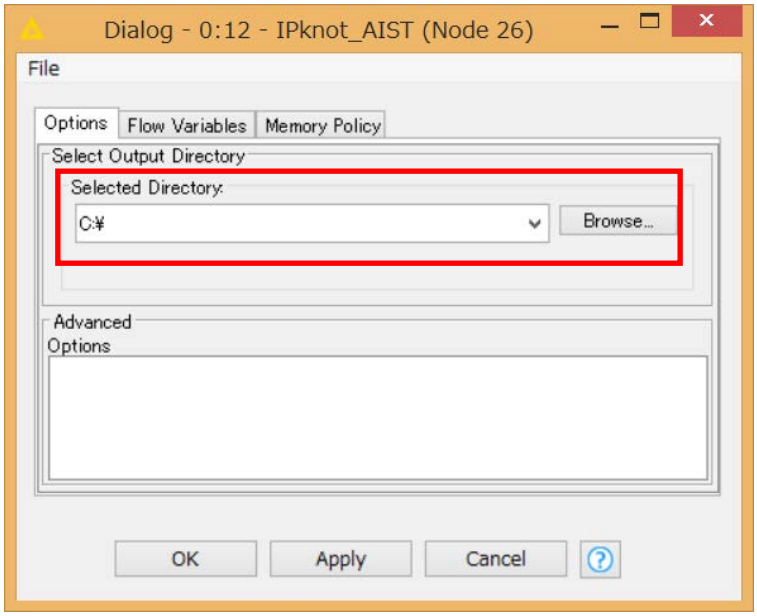


4.2.2.5-1 IPknot_AIST Icon

2) Select “Configure” from the display menu.

3) Select a directory using “Browse” in the pop-up dialog.

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



4.2.2.5-2 IPknot_AIST: Configure

Press “OK” after completion.

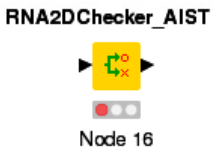
The reference website for details about IPknot is
http://rtips.dna.bio.keio.ac.jp/ipknot/

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.

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

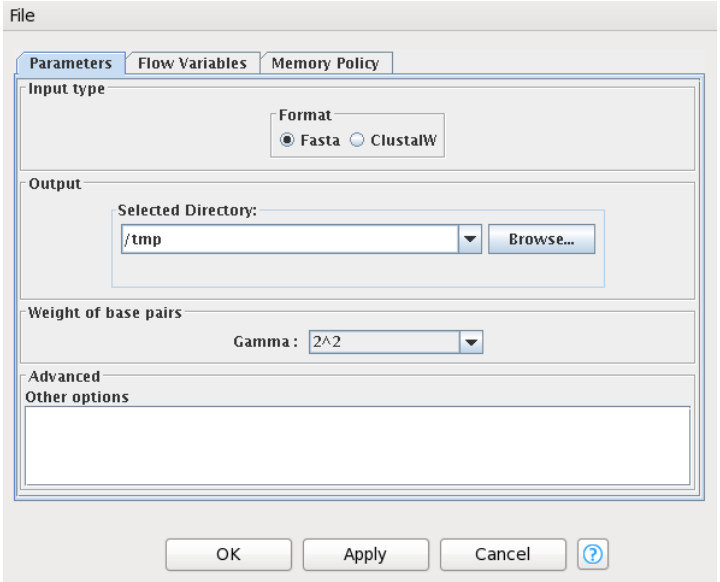
1) Select the IPknot_AIST icon followed by right-clicking.



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.



4.2.2.7 RASSIE_AIST

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

Set an output directory.

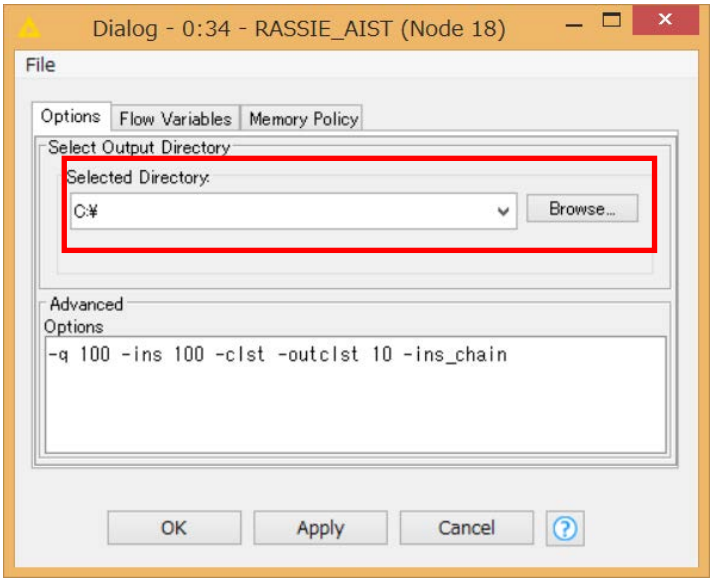
1) Select the RASSIE_AIST icon followed by right-clicking.

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

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.



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.

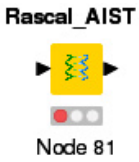
4.2.2.8 Rascal_AIST

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.

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

Set an output directory.

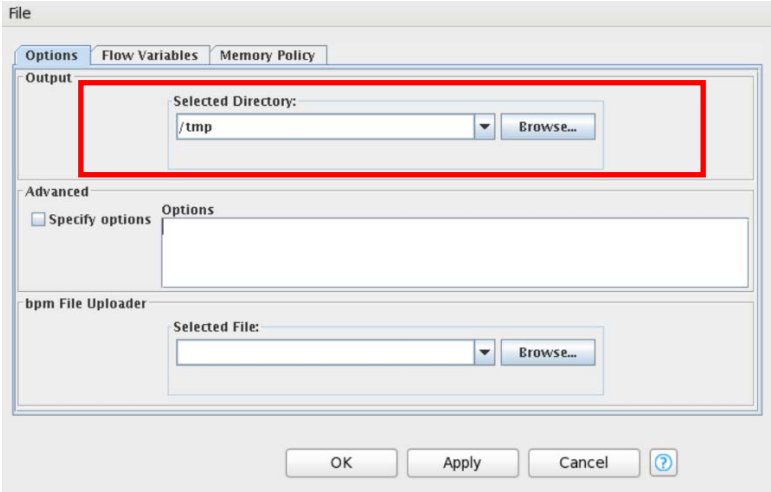
1) Select the Rascal_AIST icon followed by right-clicking.



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.



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.

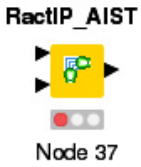
4.2.2.9 RactIP_AIST

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

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

Set an output directory.

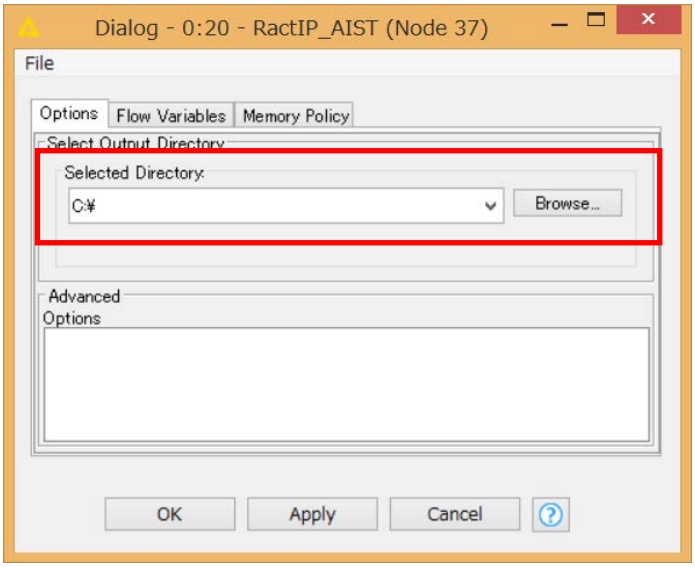
1) Select the RactIP_AIST icon followed by right-clicking.



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.



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

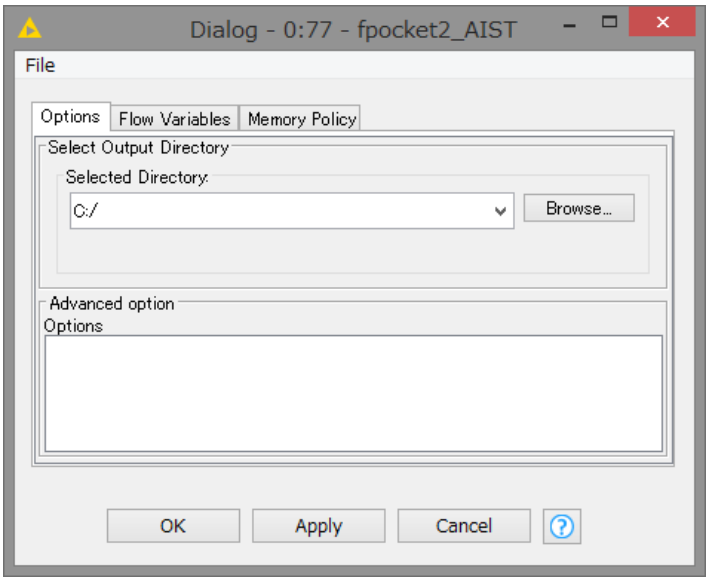
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

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.



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.

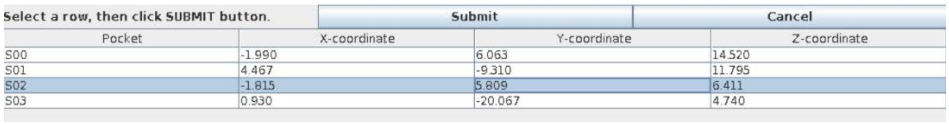
Reference: Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories, *Bioinformatics*, **27**, 3276–3285.

4.2.2.11 PocketSelector

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

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.



4.2.2.11-1 PocketSelector: Submission

2) Select “OK” to submit.

4.2.2.12 FragmentSelector

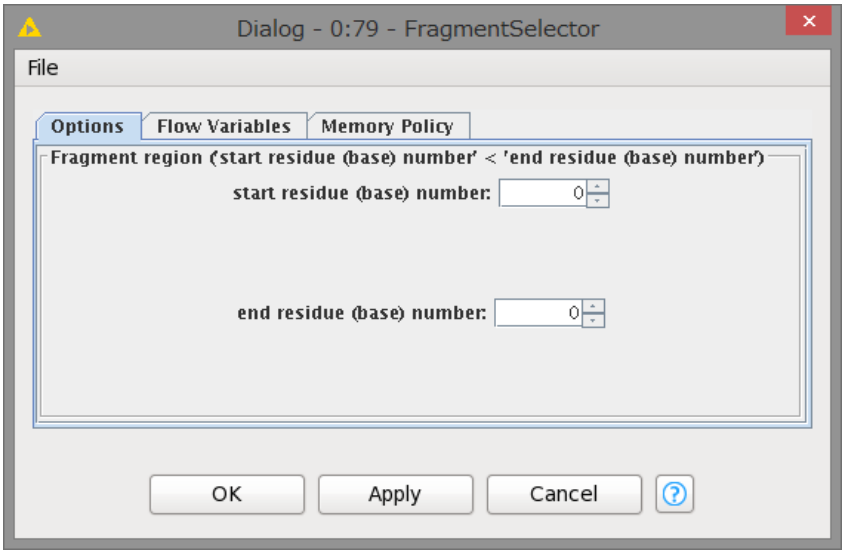
This program selects fragments of RNA in a PDB formatted file.

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.

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



4.2.2.12-1 FragmentSelector: Configure

・**Options → Fragment region → start residue (base) number:**
Specify a start residue (base) number.
・**Options → Fragment region → end residue (base) number:**
Specify an end residue (base) number.

Click the “OK” button after setting the conditions.

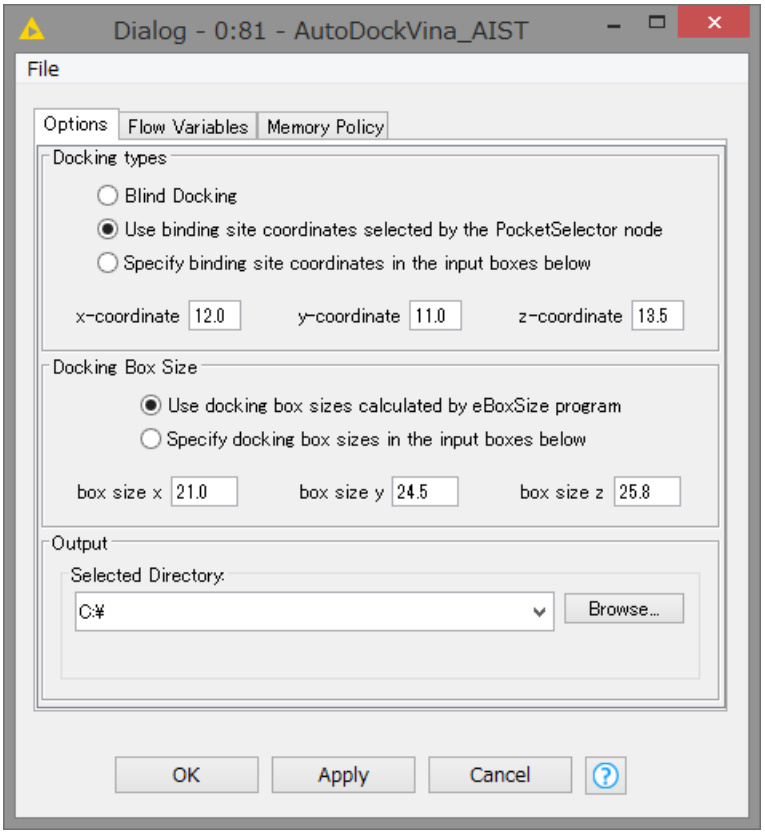
4.2.2.13 AutoDockVina_AIST

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.

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



4.2.2.13-1 AutoDockVina_AIST: Configure

・**Options → 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 → 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 → Output → Selected Directory:**

Specify an output directory path in the textbox or “Browse”.

Click the “OK” button after setting the conditions.

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

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.

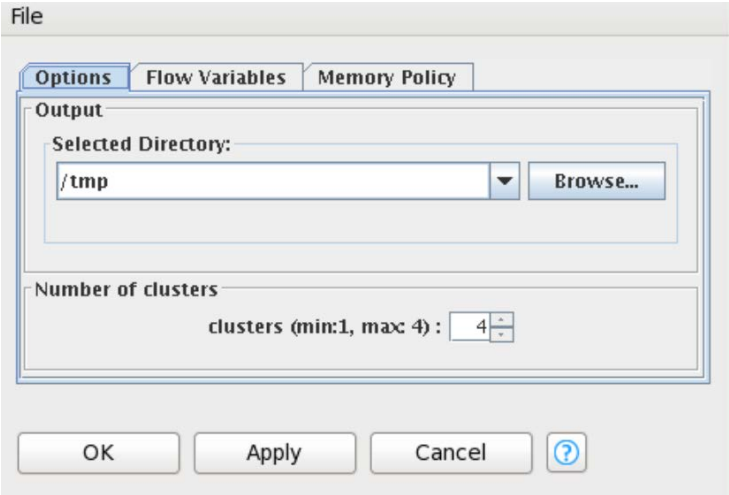
4.2.2.14 DockingAnalyzer_AIST

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.



4.2.2.14-1 DockingAnalyzer_AIST: Configure

・Options → Output → 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

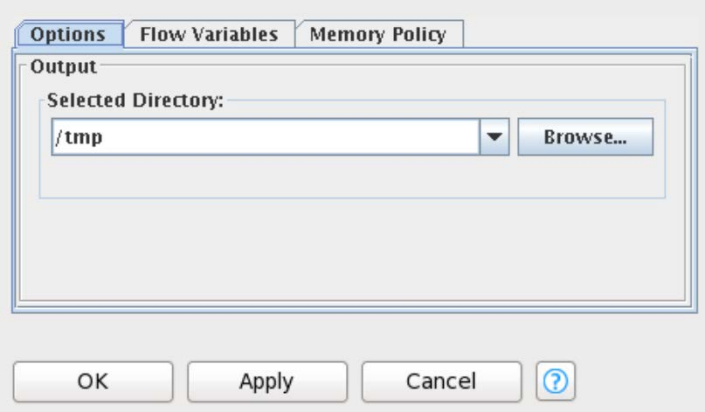
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

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.



4.2.2.15-1 Rebuild_AIST: Configure

・Options → Output → 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

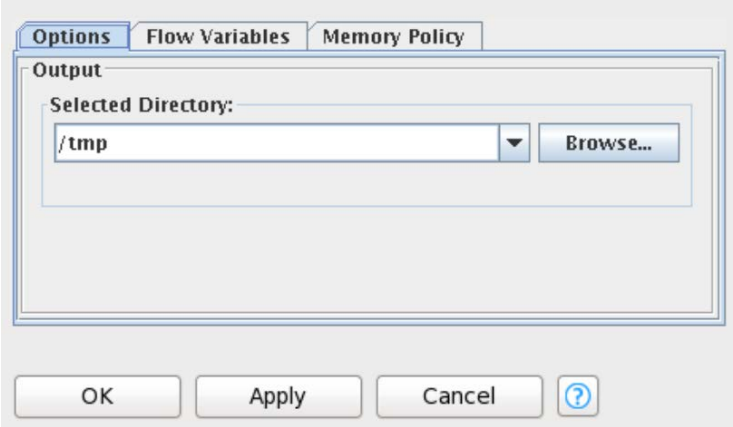
This program performs energy minimizations using molecular mechanics. If the user connects Rebuild to DockingAnalyzer via flow variables ports (red ports), this program does not open the pop-up window because candidates are already selected for each cluster by DockingAnalyzer, and this is followed by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

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



4.2.2.16-1 MinMM_AIST: Configure

・**Options → Output → Selected Directory:**

Specify an output directory path in the textbox or “Browse”.

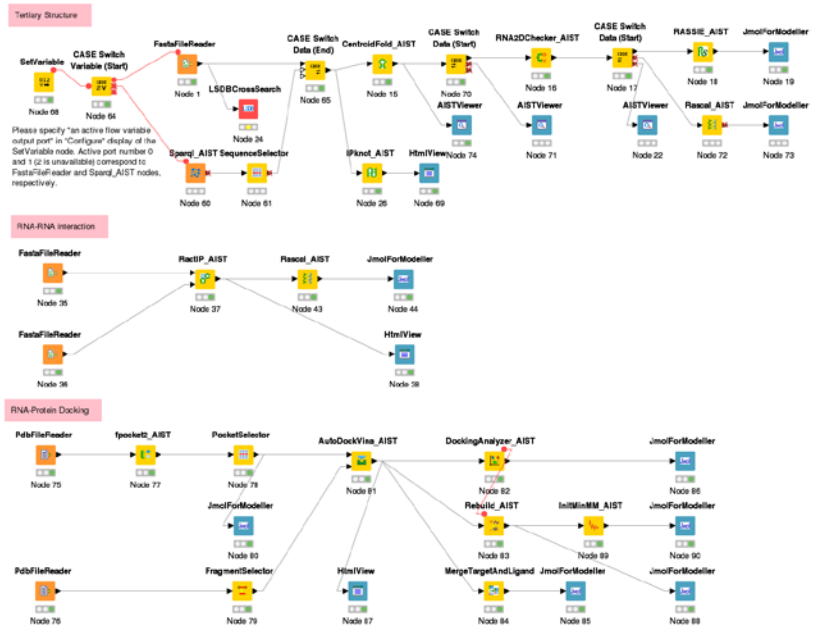
Click the “OK” button after setting the conditions.

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

5 Examples for Execution and Result Viewing

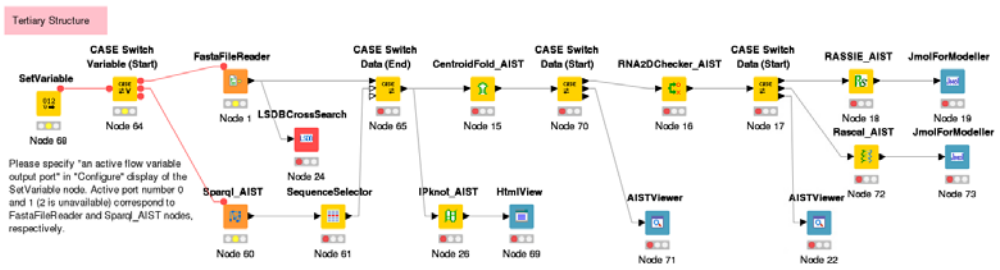
The user can download the combination types of RNA workflow packages (AIST-knime packages) from the ToGo-WF website:
(http://togo.medals.jp/active_local_rna_prediction.eng.html).
After installation, the workflow of RNA Tertiary Structure Prediction, which performs RNA 3D structure and RNA-RNA/protein 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



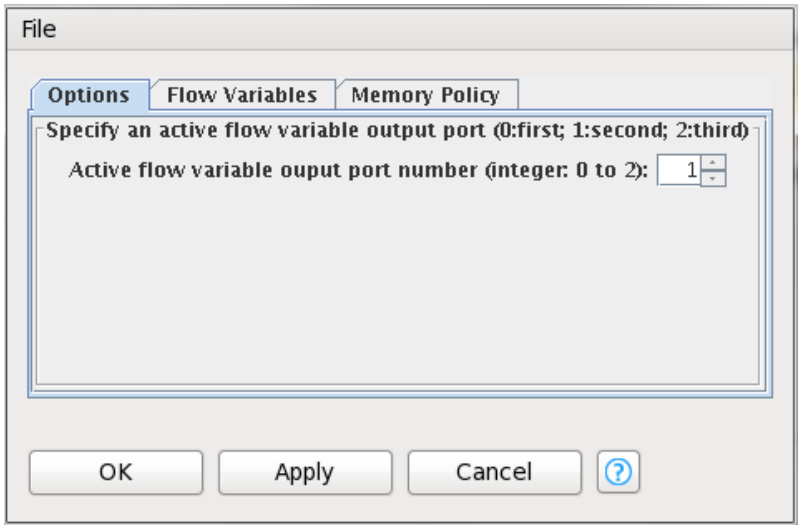
5.1 Prediction of 3D RNA structures

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.



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

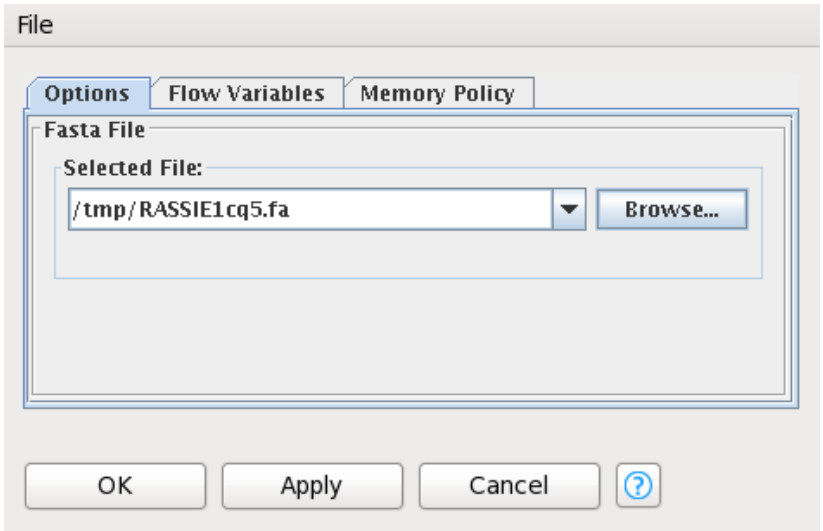
FASTA file example:

>1CQ5

GGCGUUUACCAGGUCAGGUCCGGAAGGAAGCAGCCAAGGCGCC

Example of an RNA sequence:

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

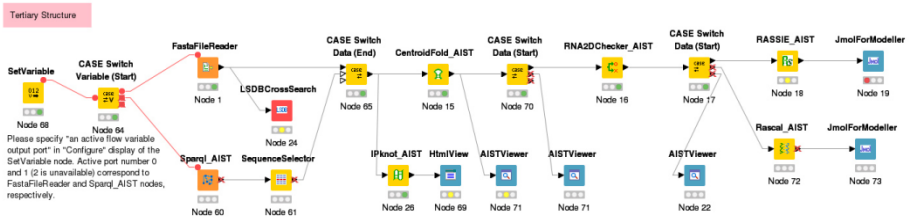


3) Click the CASE Switch Data (Start) node.

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

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.

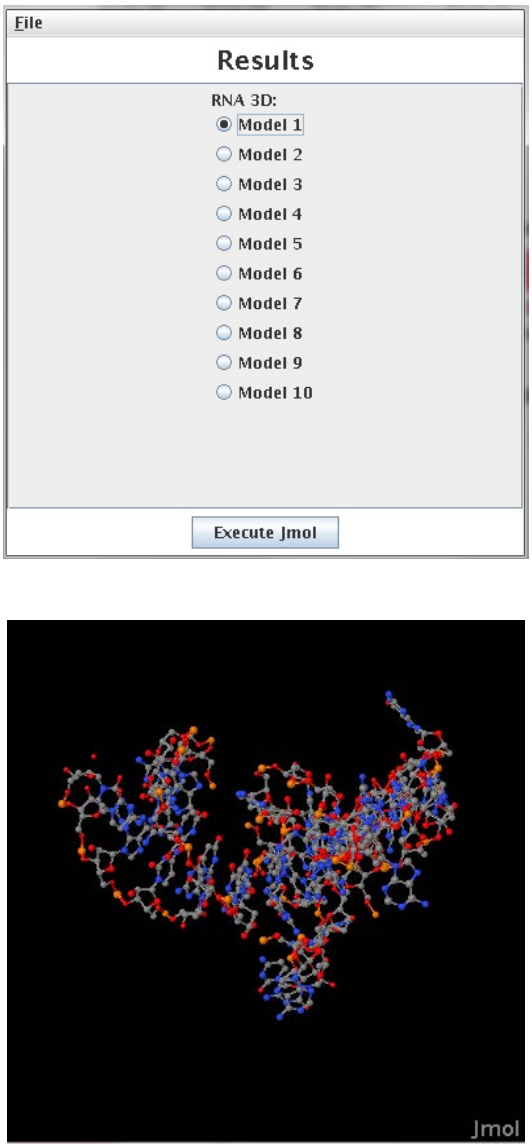
4) Execute RASSIE and JmolForModeller nodes.

5) JmolForModeller

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/

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



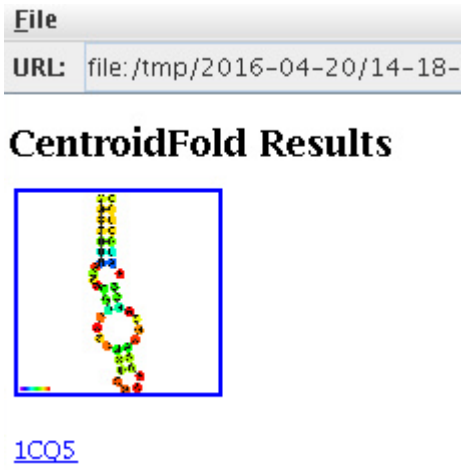
5.1.1-2 JmolForModeller –Predicted 3D RNA structure

6) AISTViewer

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.

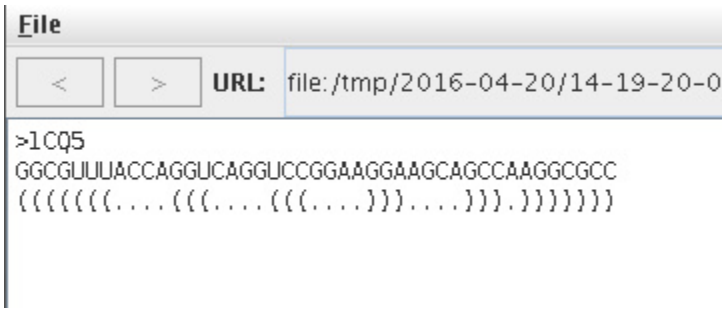
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved



5.1.1-3 AISTViewer – CentroidFold results

7) HTMLViewer

The results of IPknot_AIST can be viewed by HtmlViewer.

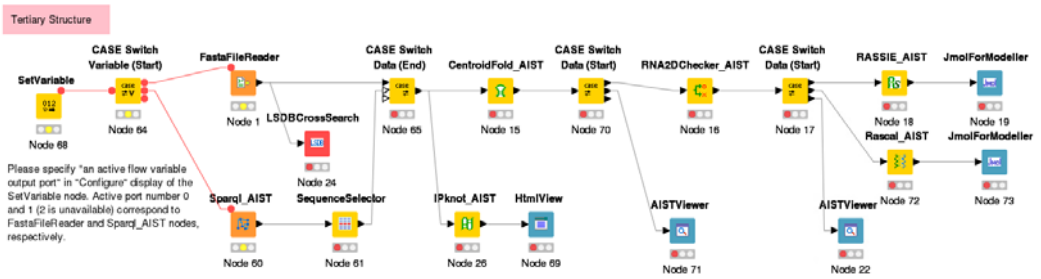


5.1.1-4 HTMLViewer – IPknot results

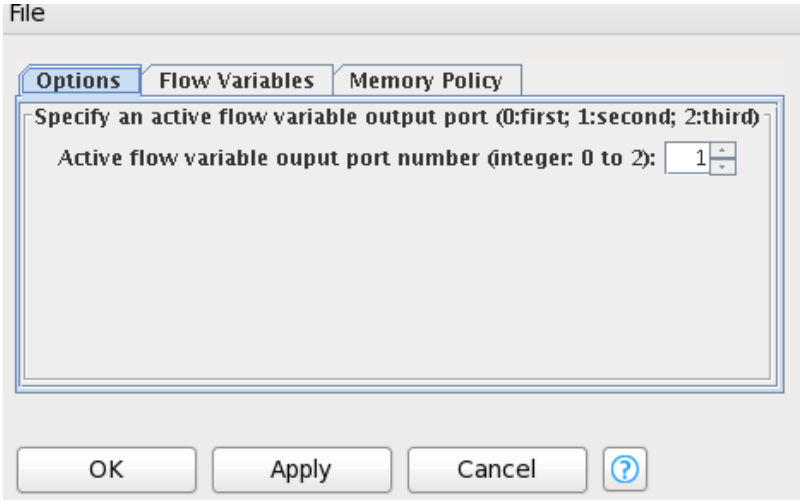
Reference: Yamasaki, S., Nakamura, S. and Fukui, K. (2012) Prospects for Tertiary Structure Prediction of RNA Based on Secondary Structure Information, *J Chem Inf Model*, **52**, 557–567.

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

5.1.2 Prediction of Tertiary Structures using the Sparql_AIST node



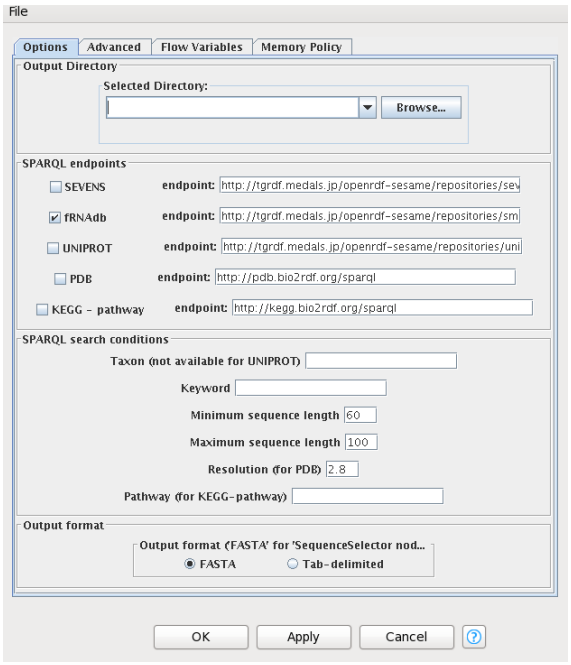
1) SetVariable = 1 for Sparql_AIST.



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.

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



3) SequenceSelector Node.

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

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)

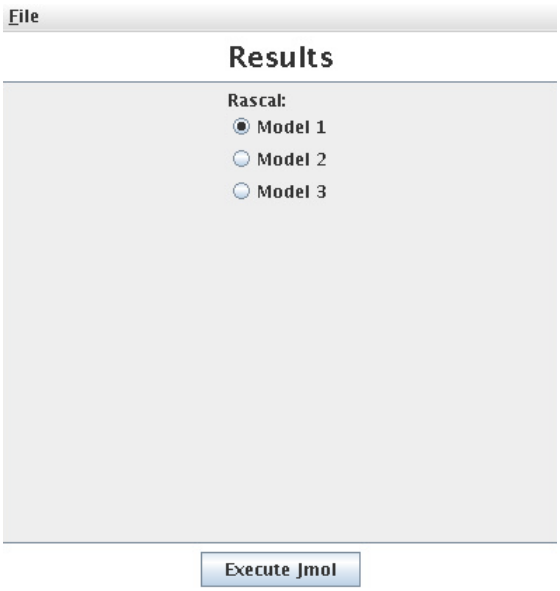
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

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), it may take about two hours. During execution, check the status of the node ("in progress") and do not close KNIME before the status reads "Complete".

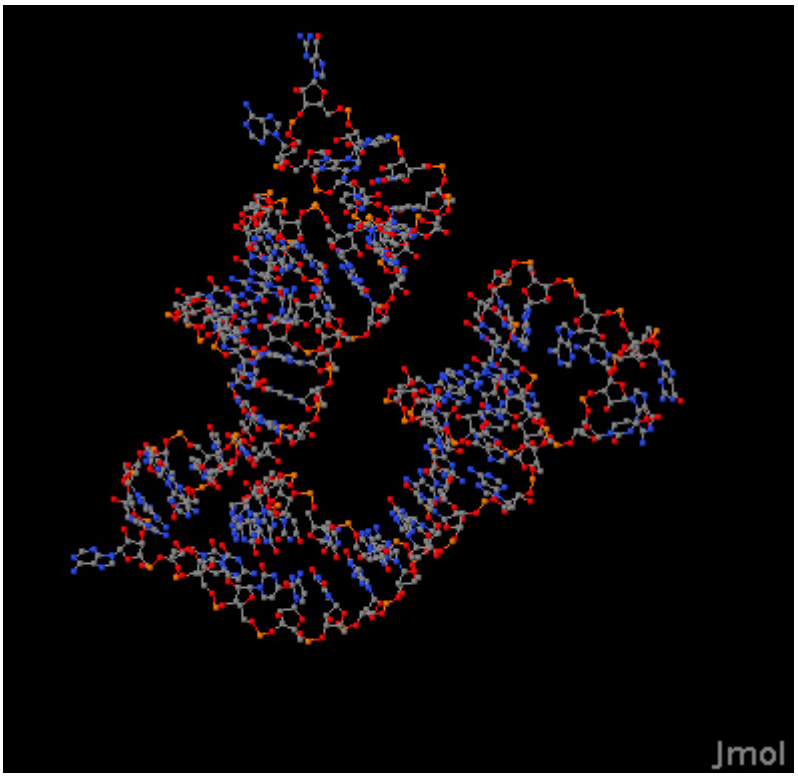
5) Execute RASSIE and JmolForModeller nodes.

6) JmolForModeller

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



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

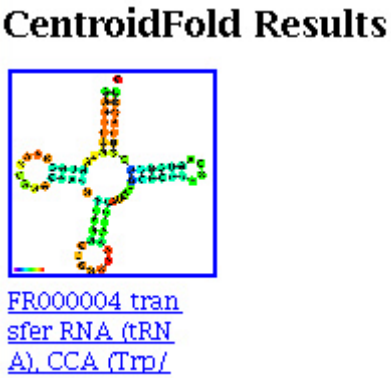


5.1.2-2 JmolForModeller –Predicted 3D RNA structure

7) AISTViewer

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.



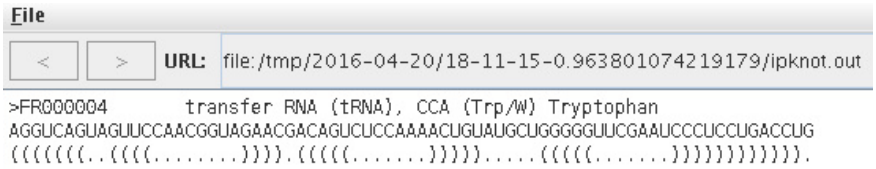
5.1.2-3 AISTViewer – CentroidFold results

8) HtmlView

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

The results of IPknot_AIST can be viewed by HtmlViewer.

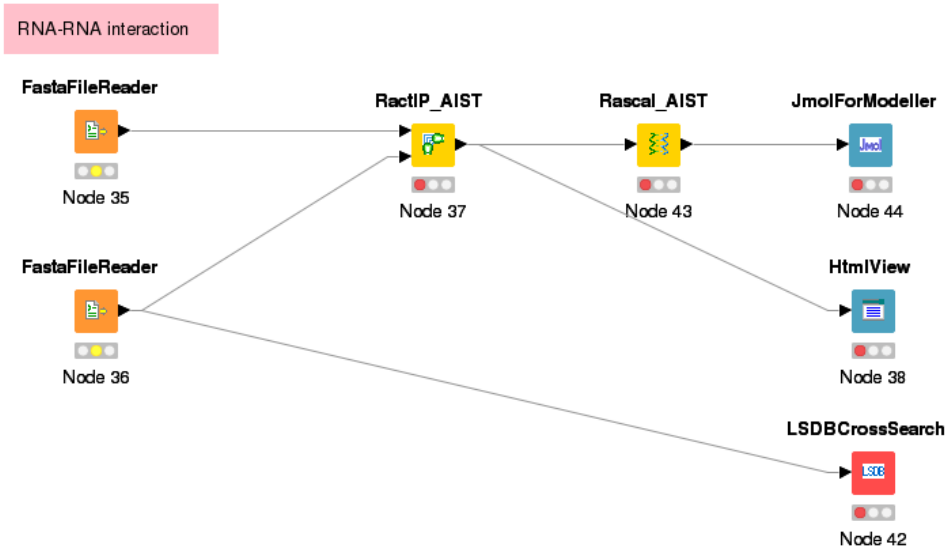
The results of Raccess_AIST can be viewed as text by HtmlView.



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

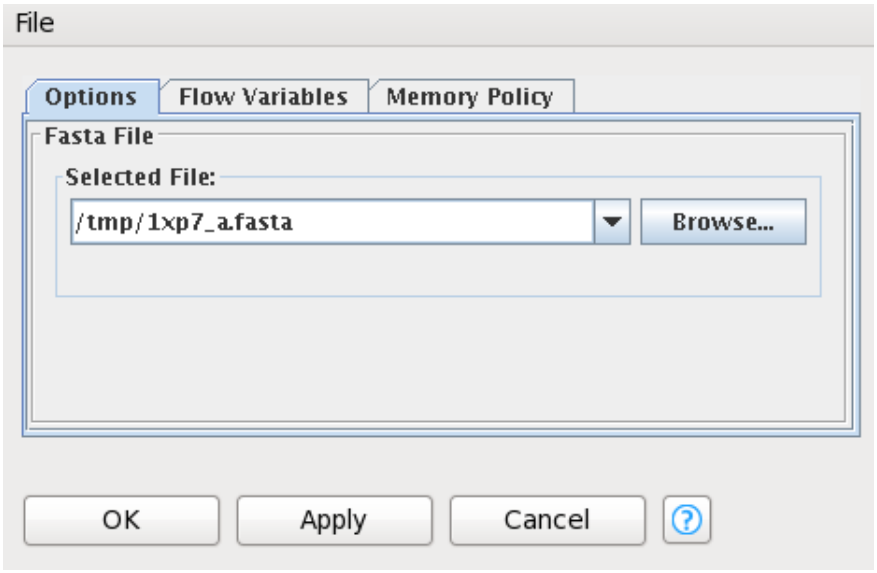
1) Set a FASTA file for each ReadFastReader by selecting “Configure”.

FASTA file example:

>1XP7:A|PDBID|CHAIN|SEQUENCE

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

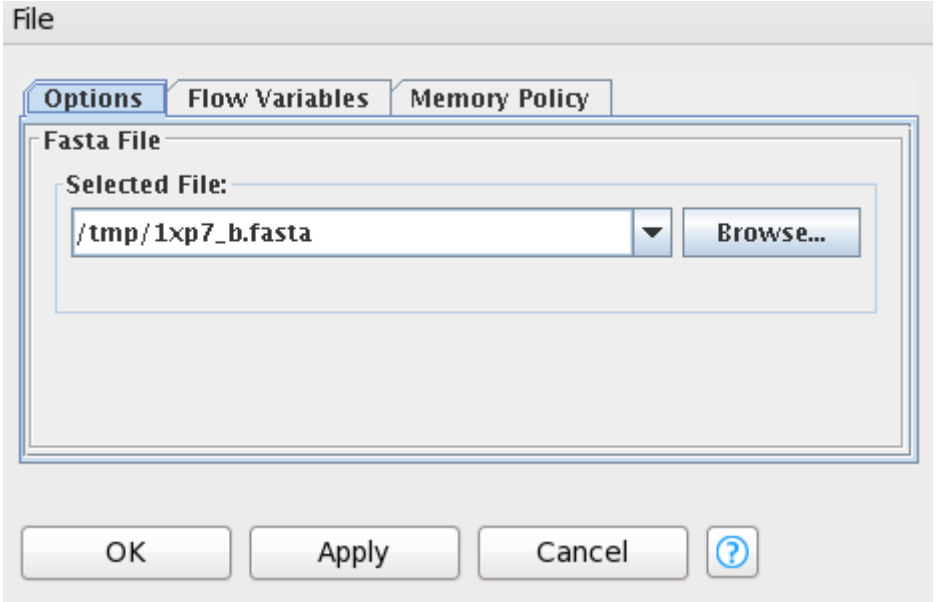
CUUGCUGAAGUGCACACAGCAAG



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



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

3) Click the RactIP_AIST node.

5.2-1 Execute selected and executable nodes (F7)

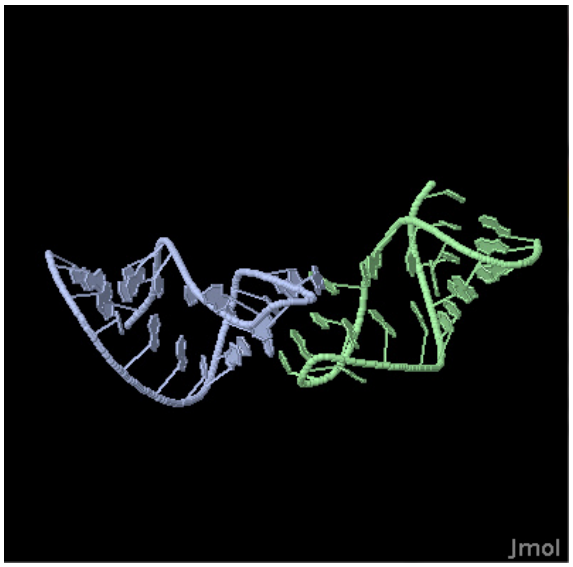
4) Execute Rascal and JmolForModeller nodes.

5) JmolForModeller

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/

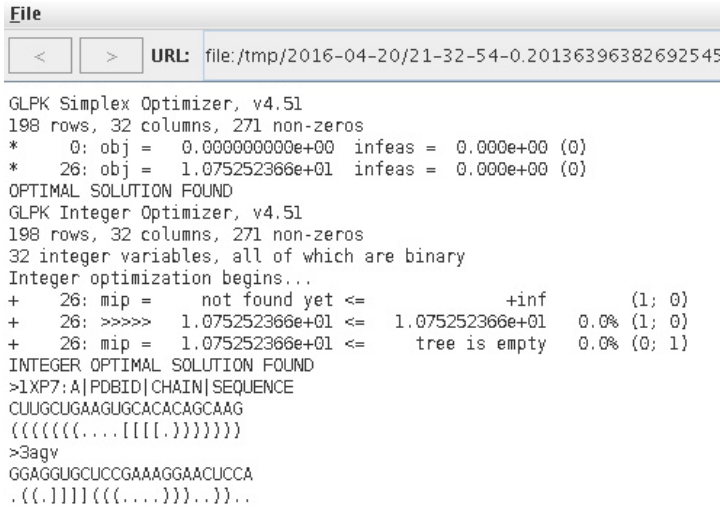
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved



5.2-2 JmolForModeller –Predicted 3D RNA structure

6) HTMLViewer

The results of IPknot_AIST can be viewed by HtmlViewer.

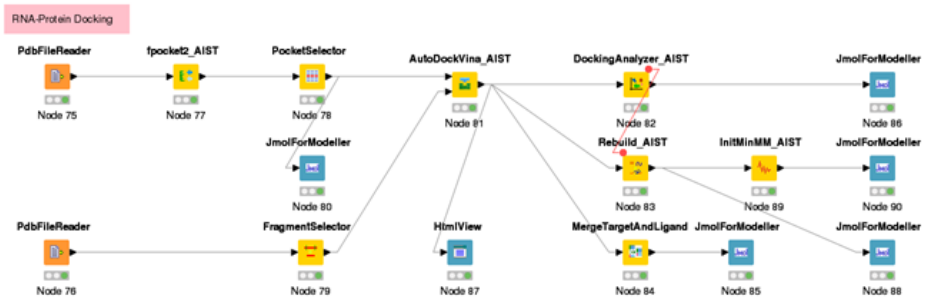


5.2-2 HtmlViewer – IPknot results

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

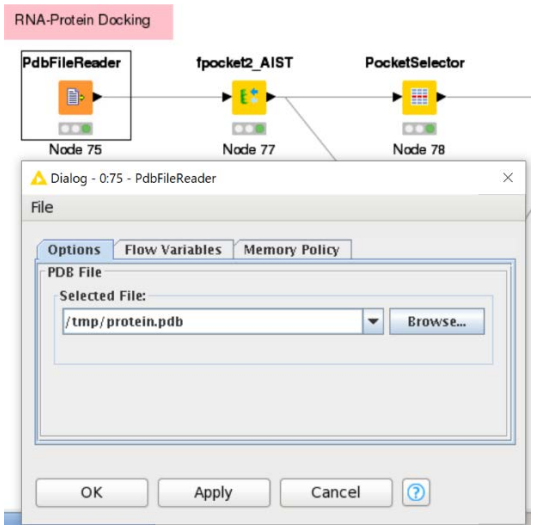
Reference: Yamasaki, S., *et al.* (2014) Tertiary Structure Prediction of RNA-RNA Complexes Using a Secondary Structure and Fragment-Based Method, *J Chem Inf Model*, **54**, 672–682.

5.3 Prediction of Protein-RNA interactions



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

Select a protein PDB file.



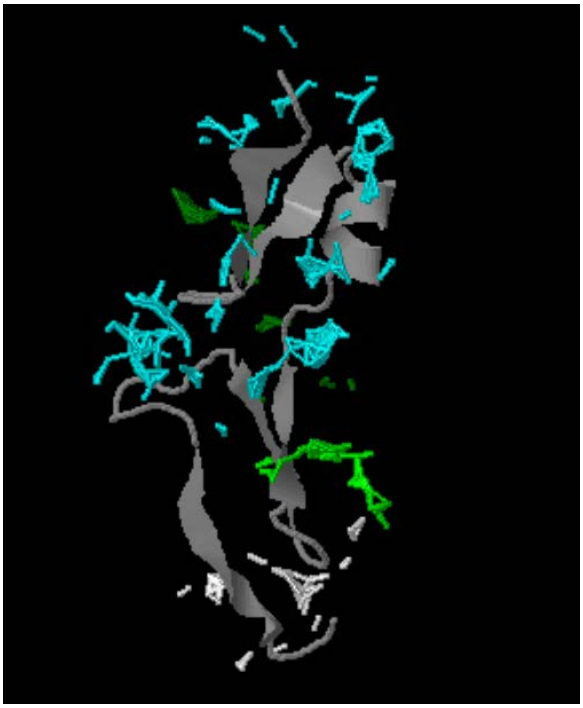
2) Search for interaction sites of the uploaded PDB file using Fpocket2_AIST.

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

Execute Jmol after the fpoeck2 calculation is finished.

Pocket	X-coordinate	Y-coordinate	Z-coordinate
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

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



5.3 JmolForModeller –Fpocket2

3) Select one of docking/interacting positions.

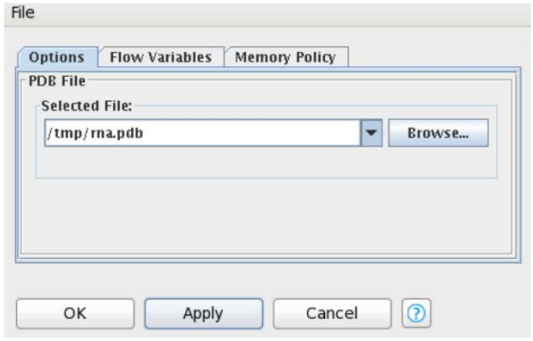
Select a row, then click SU... | Submit | Cancel

Pocket	X-coordinate	Y-coordinate	Z-coordinate
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

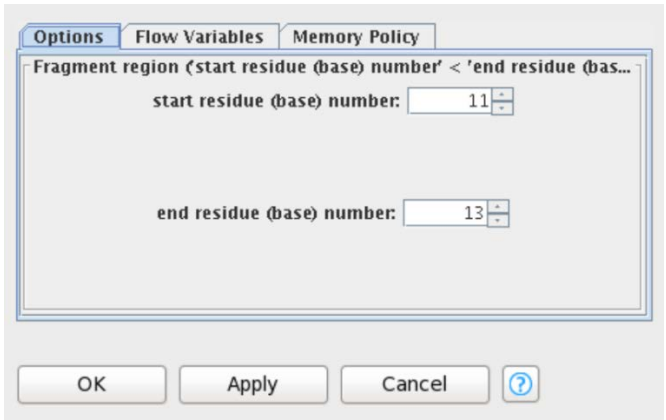
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

Click “OK” to submit the selected position for AutoDockVina.

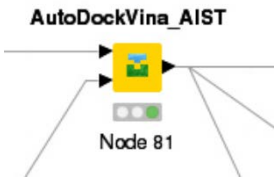
4) Select an RNA PDB file.



5) Set the fragment of RNA for docking calculations.

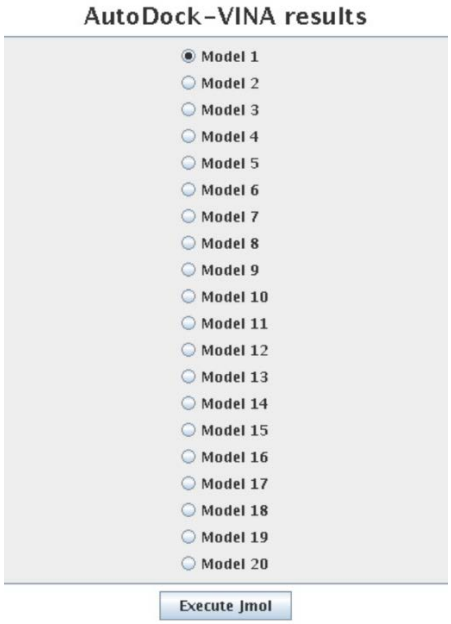


6) Run docking calculations for both the protein and RNA fragment.



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

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



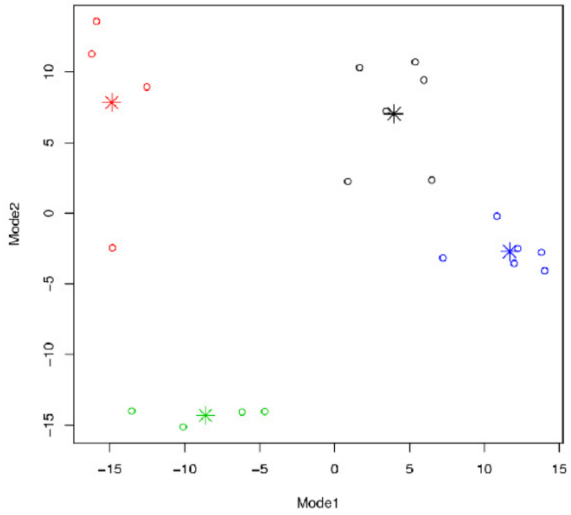
Twenty docking poses for the RNA fragment-protein complex.

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

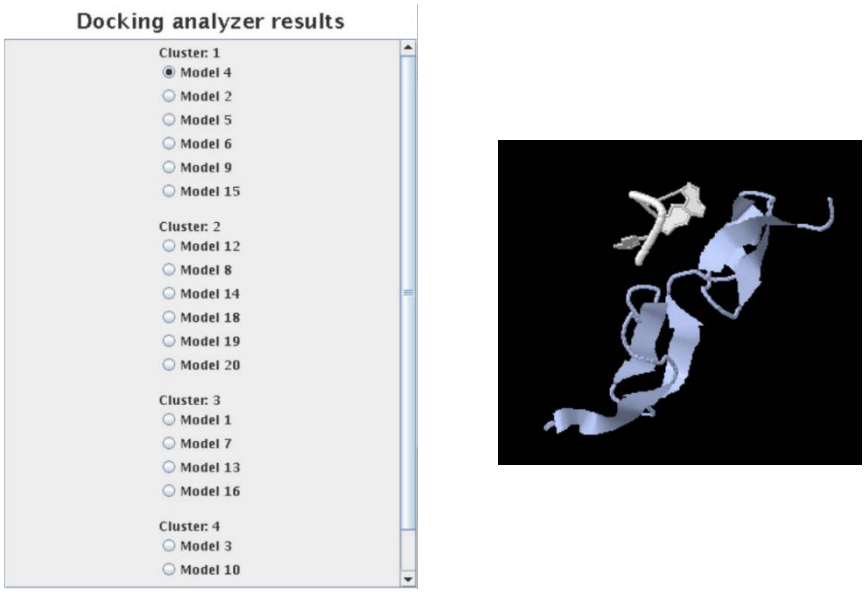
5.3 AutoDock Vina – Docking structure of RNA fragment and protein

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

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



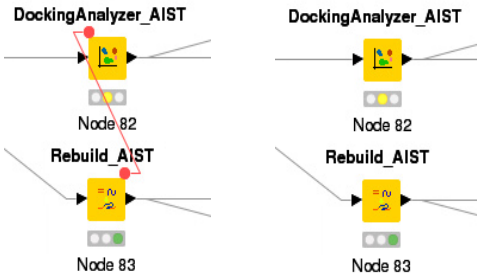
5.3 Analyzer – PCA analysis and clustering of structures of the RNA fragment and protein



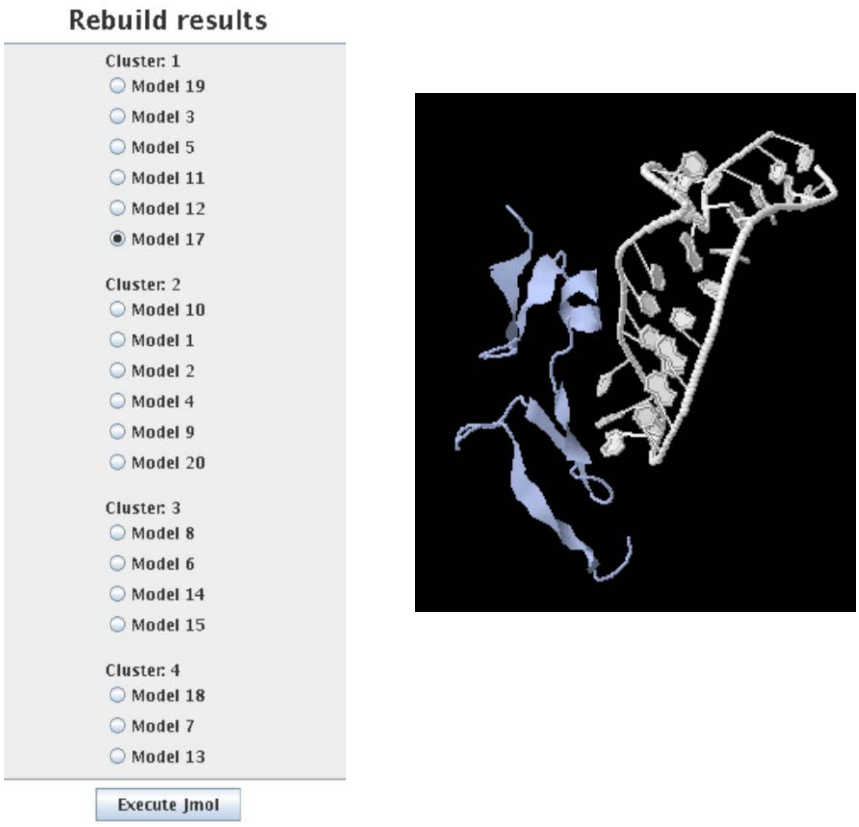
5.3 Analyzer –Tertiary clustering structures of the RNA fragment and protein

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.

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



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.

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

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:

6.1.1-1 SPARQL node list

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

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

2	Sparql_Adv		Y	Execute SPARQL simply against endpoints.
3	SequenceSelector			Display SPARQL results and select a sequence.
4	HtmlView			Display SPARQL results.

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

6.1.2 Sparql_AIST

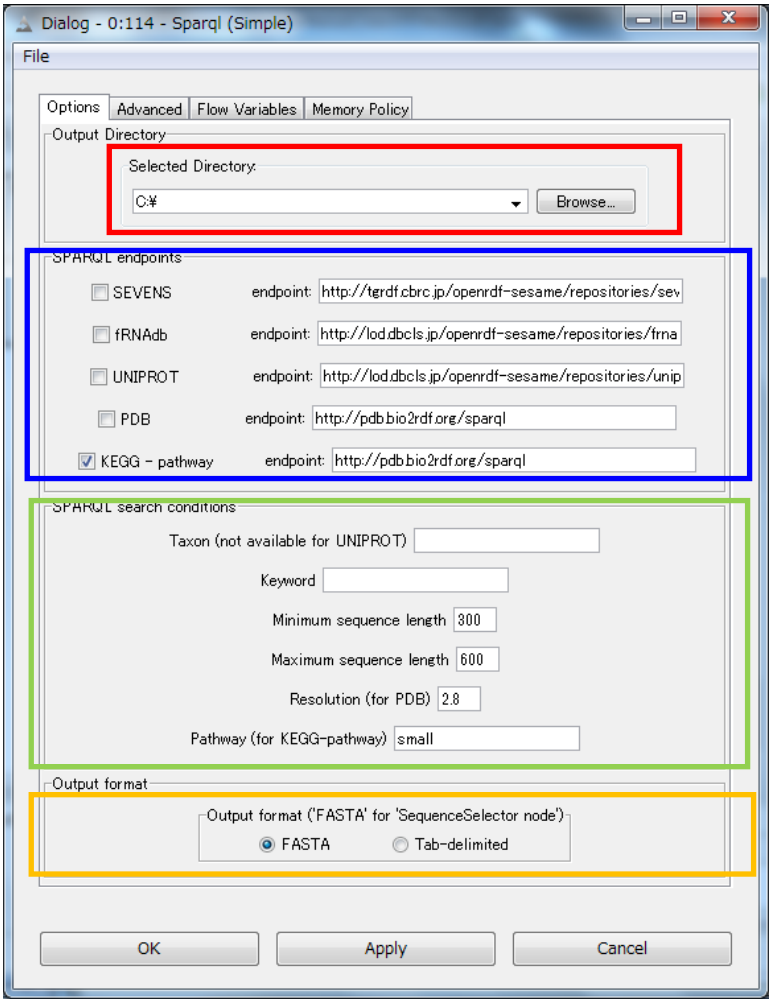
6.1.2.1 Setting the execution environment

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

6.1.2.1-1 Sparql (Simple) icon

2) Select “Configure” from the menu.

3) Setup execution conditions.



6.1.2.1-2 Sparql (Simple): Configure

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

・**Options tab → Output Directory:**

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

・**Options tab → 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 → 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 → 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 → 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.

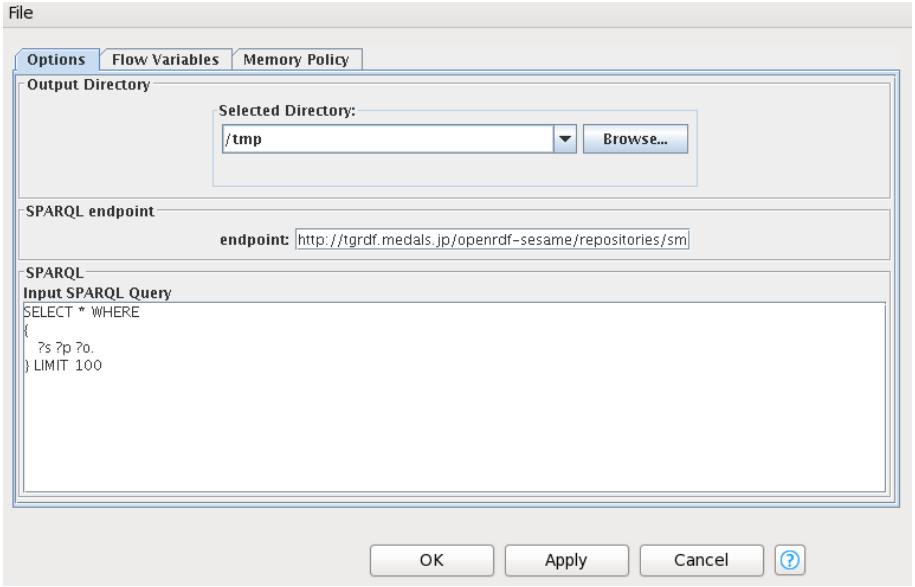
6.1.3.1 Setting the execution environment

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

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

6.1.3.1-1 Sparql_AIST_Adv. icon

2) Select “Configure” from the menu.



- **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-human

  and

  http://pdb.bio2rdf.org/sparql

  http://kegg.bio2rdf.org/sparql

- **Input SPARQL query**

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

Specify SPARQL query.

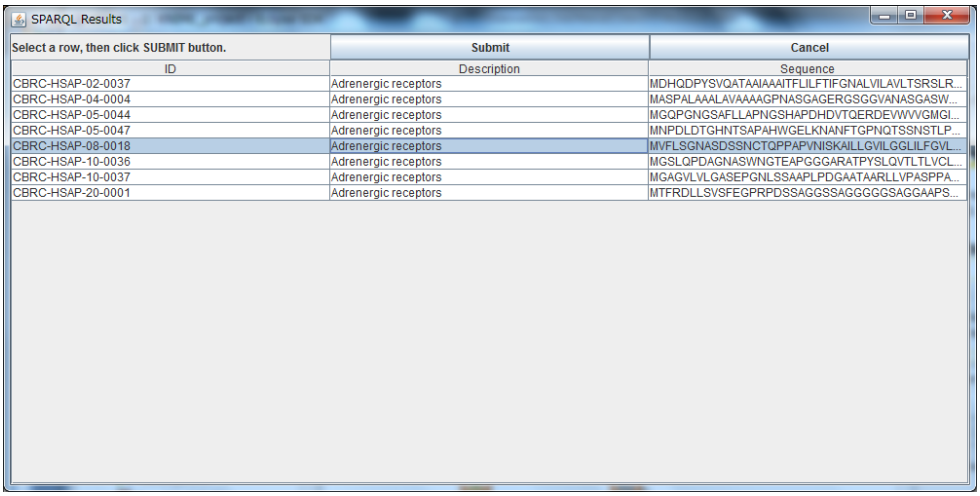
Then, click the “OK” button.

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.

6.1.4.1-1 SequenceSelector icon



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.

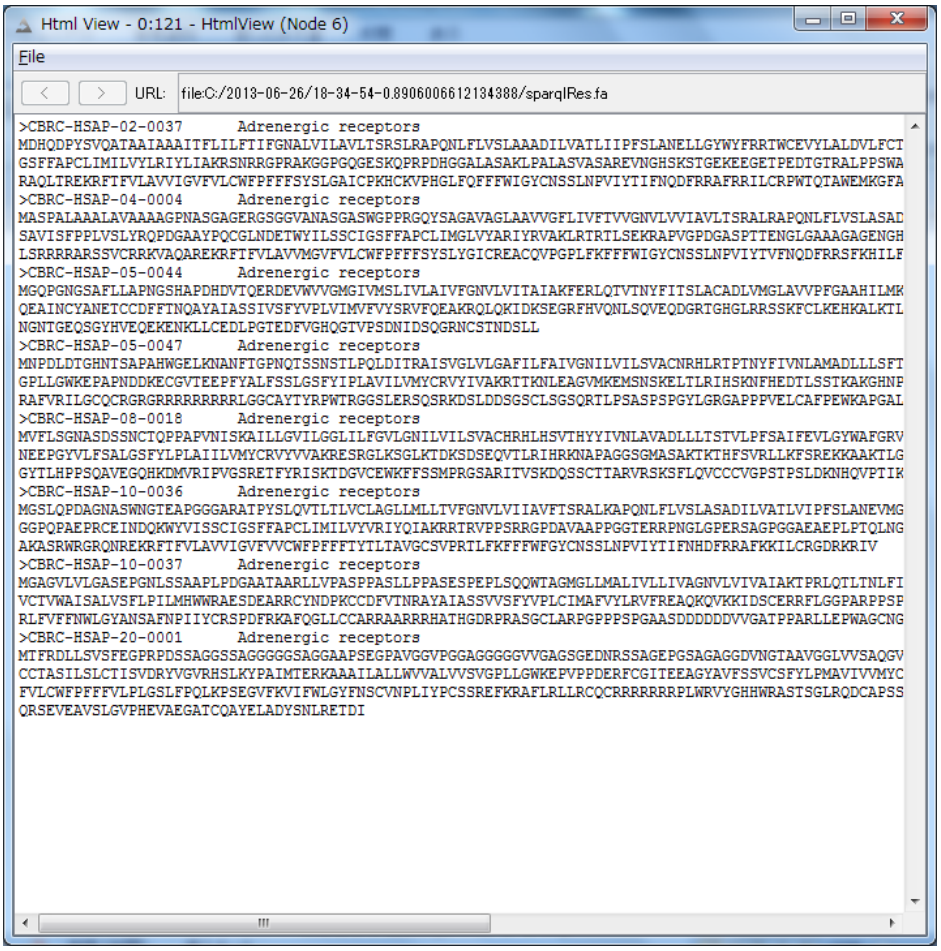
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

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



6.1.5.1-2 HtmlView results

7 Appendix

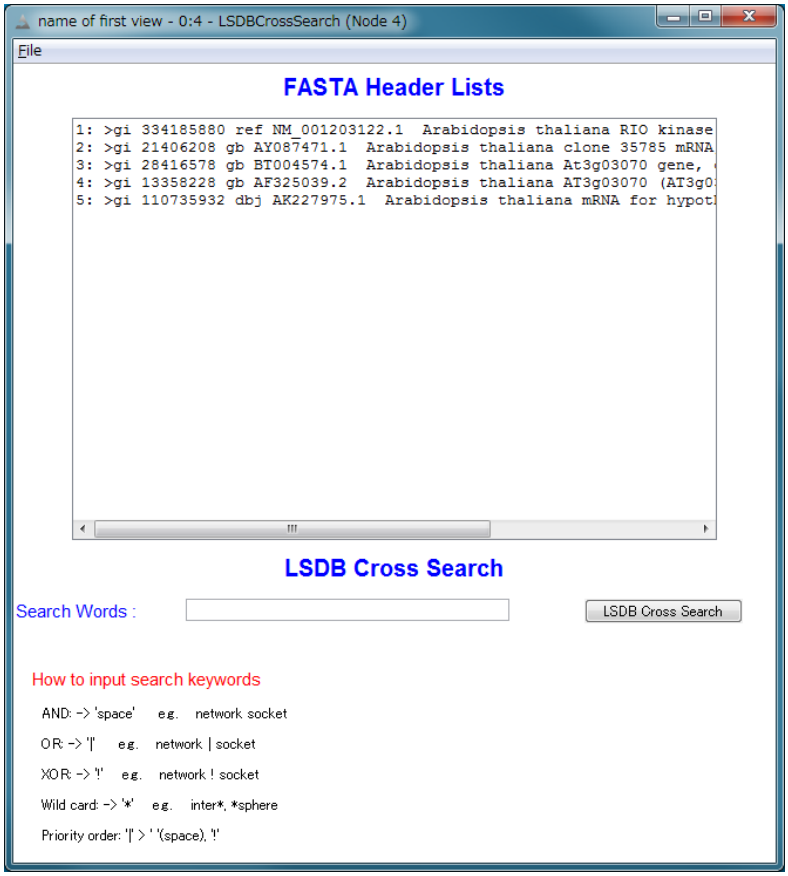
Copyright (c) Molecular Profiling Research Center for Drug Discovery, AIST 2018 All Rights Reserved

7.1 Appendix A : LSDBCrossSearch

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

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

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

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: http://biosciencedbc.jp/dbsearch/

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

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
Creative Commons Non-commercial 4.0 International License.

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

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