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

Active Workflow Combination Type

User Manual

MolProf, AIST

2016/04/25

Contents

1 Introduction 1
2 Combination types of Active workflow 2
3 Common rules 3
4 RNA Structure Prediciton 7
4.1 Preparation 7
4.2 Nodes 8
4.2.1 Node list 8
4.2.2 Node setting 11
4.2.2.1 SetVariable 11
4.2.2.2 FastaFileReader 12
4.2.2.3 Sparql_AIST 13
4.2.2.4 CentroidFold_AIST 15
4.2.2.5 IPknot_AIST 16
4.2.2.6 RactIP_AIST 17
4.2.2.7 Rascal_AIST 18
4.2.2.8 RASSIE_AIST 19
4.3 Results 20
4.3.1 IPknot_AIST 20
4.3.2 RactIP_AIST 21
4.3.3 Rascal_AIST 22
4.3.4 RASSIE_AIST 23
4.3.5 CentroidFold_AIST 24
5 Protein Structure Prediction 25
5.1 Preparation 26
5.2 Nodes 26
5.2.1 Node list 26
5.2.2 Node setting 33
5.2.2.1 SetVariable 33
5.2.2.2 FastaFileReader 34
5.2.2.3 Sparql_AIST 35
5.2.2.4 BlastForModeller _AIST 37
5.2.2.5 HitRegionSelector_AIST 38
5.2.2.6 TemplateSelector_AIST 40
5.2.2.7 Modeller_AIST 41
5.2.2.8 CPHmodels_DTU 42

5.2.2.9 PsiPred_AIST 43
5.2.2.10 Poodle_AIST 44
5.2.2.11 TmHmm_DTU 45
5.2.2.12 SignalP_DTU 46
5.2.2.13 ChloroP_DTU 47
5.2.2.14 LipoP_DTU 48
5.2.2.15 NetCTL_DTU 49
5.2.2.16 NetChop_DTU 50
5.2.2.17 NetNES_DTU 51
5.2.2.18 NetPhosK_DTU 52
5.2.2.19 NetPhos_DTU 53
5.2.2.20 NetPicoRNA_DTU 54
5.2.2.21 DictyOGlyc_DTU 55
5.2.2.22 WolfPsort_AIST 56
5.2.2.23 TargetP_DTU 57
5.2.2.24 SecretomeP_DTU 58
5.2.2.25 DisoPred_AIST 59
5.2.2.26 Memsat_AIST 60
5.3 Results 61
5.3.1 BlastForModeller_AIST → HtmlView 61
5.3.2 HitRegionSelector_AIST → HtmlView 62
5.3.3 TemplateSelector_AIST → PDBjMineWeb 63
5.3.4 Modeller_AIST → JmolForModeller 64
5.3.5 CPHmodels_DTU → JmolForModeller 65
5.3.6 CPHmodels_DTU → HtmlView 66
5.3.7 PsiPred_AIST → AISTViewer 67
5.3.8 Poodle_AIST → AISTViewer 69
5.3.9 TmHmm_DTU → HtmlView 71
5.3.10 SignalP_DTU → HtmlView 72
5.3.11 ChloroP_DTU → HtmlView 73
5.3.12 LipoP_DTU → HtmlView 74
5.3.13 NetCTL_DTU → HtmlView 75
5.3.14 NetChop_DTU → HtmlView 76
5.3.15 NetNES_DTU → HtmlView 77
5.3.16 NetPhosK_DTU → HtmlView 78
5.3.17 NetPhos_DTU → HtmlView 79

5.3.18 NetPicoRNA_DTU → HtmlView 80
5.3.19 DictyOGlyc_DTU → HtmlView 81
5.3.20 WolfPsort_AIST → HtmlView 82
5.3.21 TargetP_DTU → HtmlView 83
5.3.22 SecretomeP_DTU → HtmlView 84
5.3.23 DisoPred_AIST → AISTViewer 85
5.3.24 Memsat_AIST → AISTViewer 86
6 Phylogenetic Tree 88
6.1 Preparation 89
6.2 Nodes 89
6.2.1 Node list 89
6.2.2 Node setting 91
6.2.2.1 FastaFileReader 91
6.2.2.2 ClustalW_AIST 92
6.2.2.3 Mafft_AIST 93
6.2.2.4 PhylogeneticTree_AIST 95
6.3 Results 97
6.3.1 ClustalW_AIST → AISTViewer 97
6.3.2 Mafft_AIST → AISTViewer 99
6.3.3 PhylogeneticTreeView 101
7 Molecular Simulation 102
7.1 Preparation 103
7.2 Nodes 103
7.2.1 Node list 103
7.2.2 Node setting 108
7.2.2.1 SetVariable 108
7.2.2.2 FastaFileReader 109
7.2.2.3 Sparql_AIST 110
7.2.2.4 BlastForModeller _AIST 112
7.2.2.5 HitRegionSelector_AIST 113
7.2.2.6 TemplateSelector_AIST 115
7.2.2.7 Modeller_AIST 116
7.2.2.8 CompoundQuery_AIST 117
7.2.2.9 AutoDock_AIST 120
7.2.2.10 MMPrep_AIST 121
7.2.2.11 InitMinMM_AIST 122

7.2.2.12 MoltrecMD_AIST 123
7.2.2.13 ResultPathSetter 124
7.2.2.14 Ammos_AIST 125
7.2.2.15 fpocket2_AIST 126
7.2.2.16 PdbFileReader 127
7.2.2.17 Mol2FileReader 128
7.3 Results 129
7.3.1 BlastForModeller_AIST → HtmlView 129
7.3.2 HitRegionSelector_AIST → HtmlView 130
7.3.3 TemplateSelector_AIST → PDBjMineWeb 131
7.3.4 Modeller_AIST → JmolForModeller 132
7.3.5 AutoDock_AIST → JmolForModeller 133
7.3.6 InitMinMM_AIST → JmolForModeller 135
7.3.7 MoltrecMD_AIST → HtmlView 136
7.3.8 ResultPathSetter → JmolForModeller、HtmlView 136
7.3.9 fpocket2_AIST → JmolForModeller 137
8 SPARQL 138
8.1 Nodes 138
8.1.1 Node list 138
8.1.2 Sparql_AIST 139
8.1.2.1 Node setting 139
8.1.3 Sparql_AIST_Adv 142
8.1.3.1 Node setting 142
8.1.4 SequenceSelector 143
8.1.4.1 Result 143
8.1.5 HtmlView 144
8.1.5.1 Result 144
9 Appendix 145
9.1 LSDBCrossSearch 145
10 Contact 147

1 Introduction

This manual describes Combination types of Active workflow that have developed at Molecular Profiling Research Center for Drug Discovery (MolProf), Advanced Industrial Science and Technology (AIST).
For the installation procedure, please refer to the installation manual available in a TOGO Web site (http://togo.medals.jp).

The Active workflows run on KNIME platform.
Please refer to the KNIME site for further information.
This manual explains how to use Active workflows.

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

2 Combination types of Active workflow

There are four Combination types of Active workflow which are available.

2-1 Combination types of Active workflow

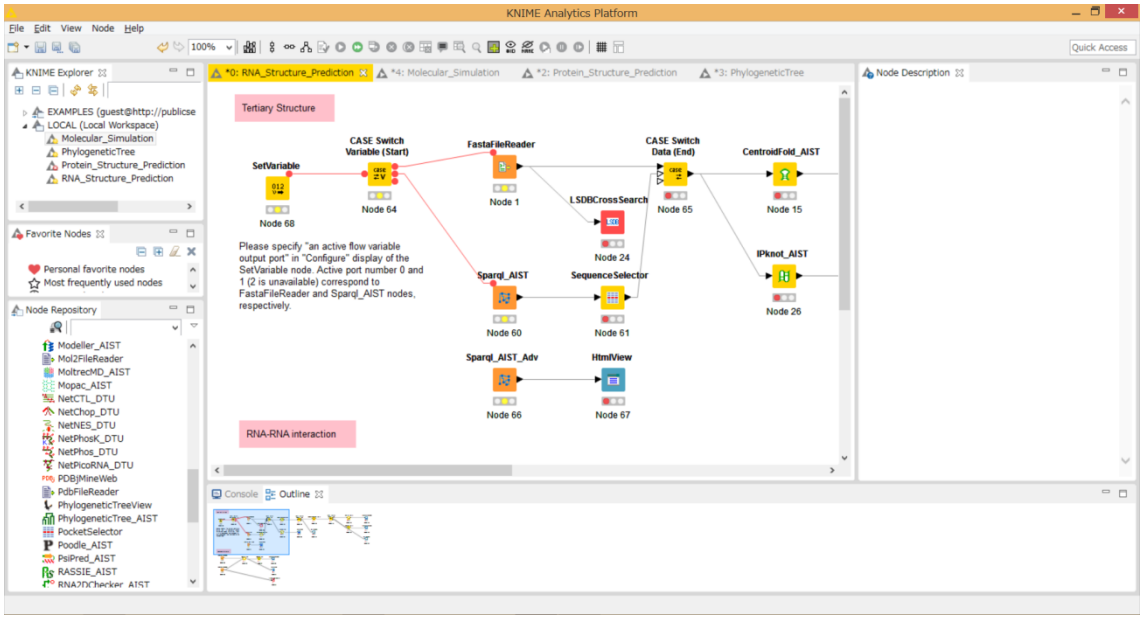
No.	Active workflow	OS	Description
1	RNA Structure Prediction Active Workflow	Windows32bit Linux 64bit MacOS (OS X 10.7 and over)	RNA second-tertiary structure prediction, RNA-RNA interaction prediction
2	Protein Structure Prediction Active Workflow	Windows32bit Linux 64bit MacOS (OS X 10.7 and over)	Protein function prediction, modelling
3	PhylogeneticTree Active Workflow	Windows32bit Linux 64bit MacOS (OS X 10.7 and over)	Multiple alignment, Phylogenetic tree
4	Molecular Simulation Active Workflow	Windows32bit Linux 64bit MacOS (OS X 10.7 and over)	Protein modelling, docking, and molecular simulation

3 Common rules

Common rules in all Active workflows are as follows:

1. Starting Active workflows

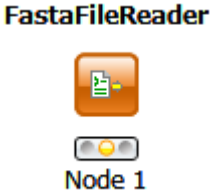
On KNIME Explorer menu on upper left corner of the screen, user can select a workflow by specifying each of four Active workflows.



3-1 RNA_Structure_Prediction workflow

2. Nodes

A node deals with an assigned task as a functional unit.



3-2 FastaFileReader

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 condition

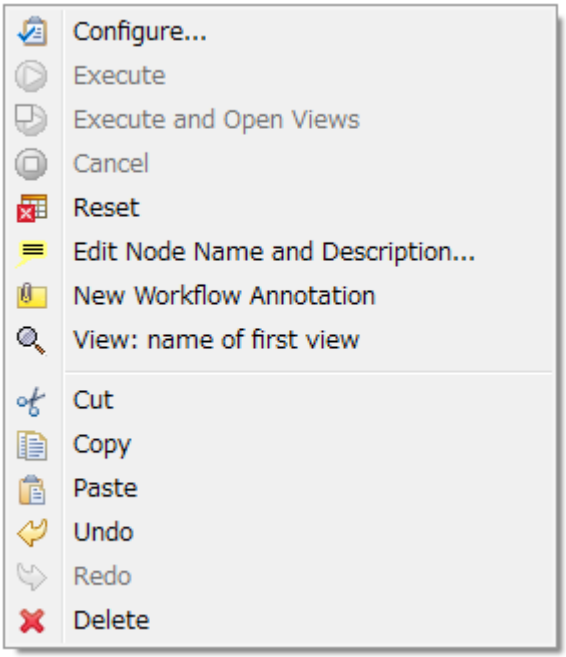
A node has a signal bar which shows the node execution condition. Types of sign are shown in the below table.

3-3 Node execution conditions

Signal bar	Color	Description
	red	To be prepared
	yellow	Preparation completion
	green	Execution termination
	blue	Running
	queued	Waiting

4. Node menu

By right-clicking on a node, KNIME commands are displayed as shown in the menu below.



3-4 Node menu

3-5 Command list

Menu command	Action	Note
Configure…	Open a Dialog window	Specify execution condition of the node.
Execute	Execute the node.	The node cannot be used unless the node status is yellow.
Execute and Open Views	Execute the node and open a View window.	The node cannot be used unless the node status is yellow.
Cancel	Cancel the execution.	-
Reset	Reset the node progress.	-
Edit Node Name and Description…	Edit the node name and Description.	-
New Workflow Annotation	Write annotations in a text box.	-
View : [viewer name]	Open a View window.	-
Cut	Cut the node.	-
Copy	Copy the node.	-
Paste	Paste the node.	-
Undo	Undo	-
Redo	Redo	-
Delete	Delete the node.	-

5. **Execution of all executable nodes**

By clicking the below icon on the top menu, user can execute all executable nodes. In case, user should set execution conditions for all nodes before executions (Execute all executable nodes (Shift+F7))

3-5 Execute all executable nodes

6. **Alert and error messages**

When an error or alert has occurred during executions, an error or alert message is displayed on the pop-up window and a KNIME console dialog on the bottom of the screen.

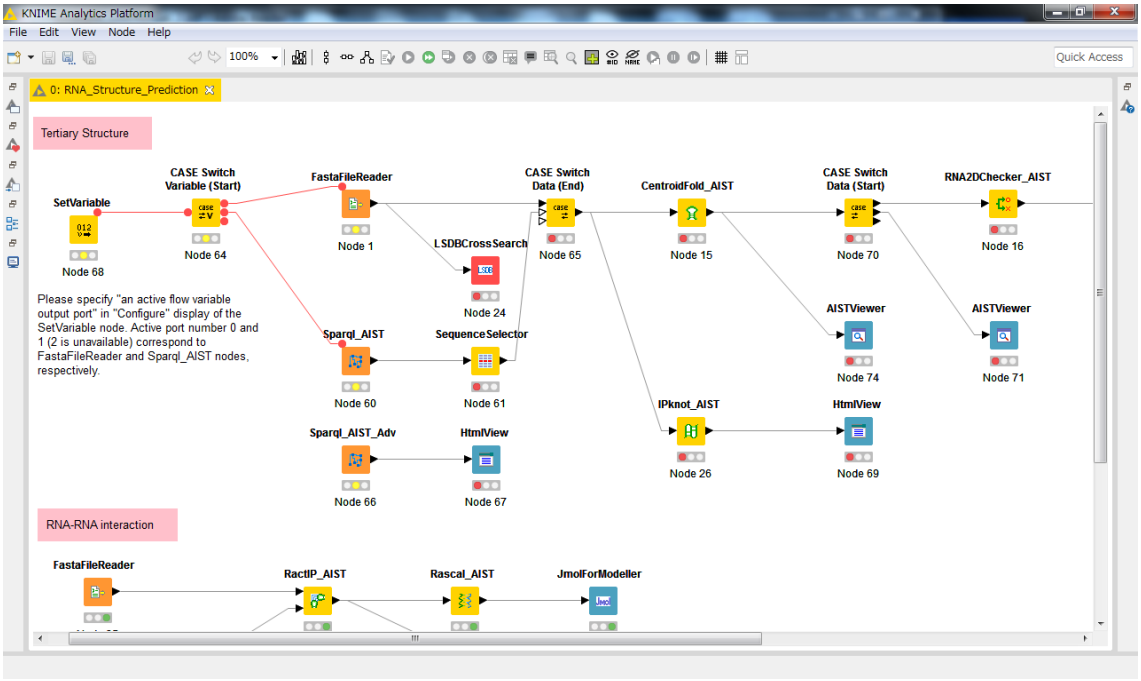
Please check the error or alert message to solve the problem.
Types of error and alert message are shown in the below table.

3-6 Error and alert messages

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

4 RNA Structure Prediciton

RNA Structure Prediction is a workflow which predicts RNA secondary-tertiary structures and RNA-RNA interactions, via SOAP.



4-1 RNA_Structure_Prediction workflow

4.1 Preparation

This workflow needs a RNA FASTA sequence file.

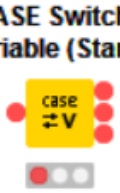
File type
RNA FASTA sequence file

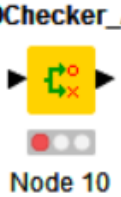
4.2 Nodes

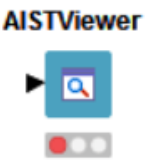
RNA Structure Prediction Active workflow has 17 types of KNIME node. The nodes are shown in the below table. Please check each KNIME node description.

4.2.1 Node list

4.2.1-1 RNA Structure Prediction Active Workflow node list

No	Name	Icon	Set	Description
1	SetVariable		Y	Control CASE Switch Variable. Open Dialog: 0: first outport 1: second outport 2:third outport
2	CASE Switch Variable(Start)			Control CASE Switch using workflow variables.
3	FastaFileReader		Y	Set FASTA file.
4	LSDBCrossSearch			Execute LSDB cross search.

5	Sparql_AIST		Y	Execute SPARQL against AIST endpoints.
6	SequenceSelector			Select a sequence from SPARQL results.
7	CASE Switch Variable (End)			CASE Switch end node.
8	CentroidFold_AIST		Y	Execute CentroidFold.
9	IPknot_AIST		Y	Execute IPknot.
10	RNA2DChecker_AIST			Execute RNA2DChecker.
11	RASSIE_AIST		Y	Execute RASSIE.

12	RactIP_AIST		Y	Execute RactIP.
13	Rascal_AIST		Y	Execute Rascal.
14	Sparql_AIST_Adv		Y	Execute SPARQL against public endpoints.
15	AISTViewer			Display prediction results.
16	HtmlView			Display results.
17	JmolForModeller			Start Jmol.

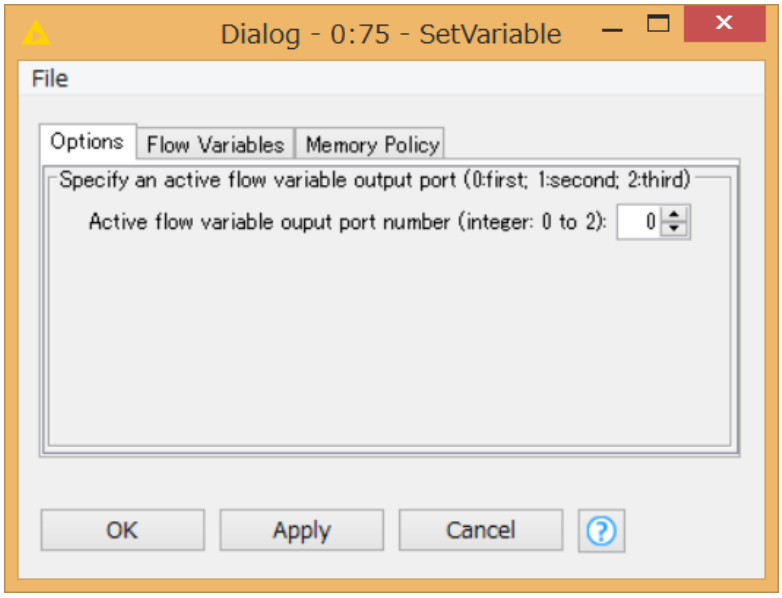
4.2.2 Node setting

4.2.2.1 SetVariable

Set a number to control CASE Switch Variable (Start) outports using workflow variables.

1) Right-clicking on the SetVariable node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.1-1 SetVariable : Configure…

0: CASE Switch Variable(Start) first outport

1: CASE Switch Variable(Start) second outport

2: CASE Switch Variable(Start) third outport

In the workflow, set 0 and 1 for FastaFileReader and Sparql_AIST, respectively.

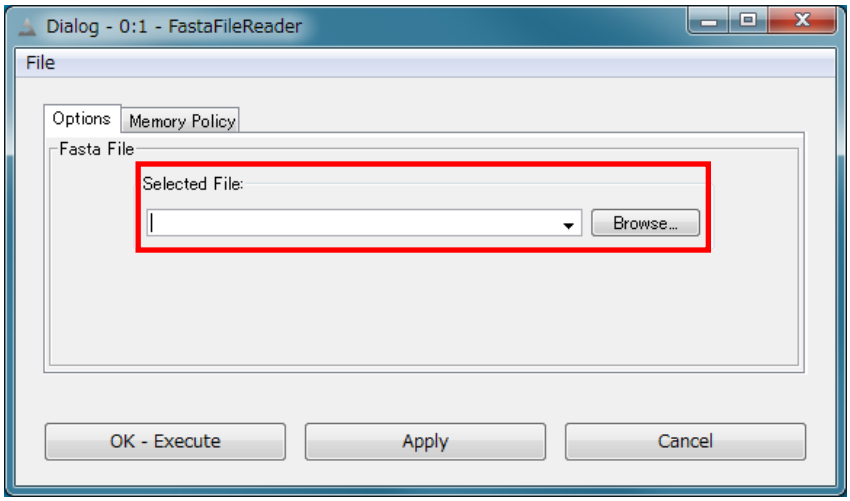
Click “OK” button after setting the conditions.

4.2.2.2 FastaFileReader

Set an RNA FASTA sequence file.

1) Right-clicking on the FastaFileReader node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.2-1 FastaFileReader : Configure…

・Options → Fasta File → SelectedFile:

Specify an RNA FASTA sequence file in a text box or “Browse” (red open rectangular)

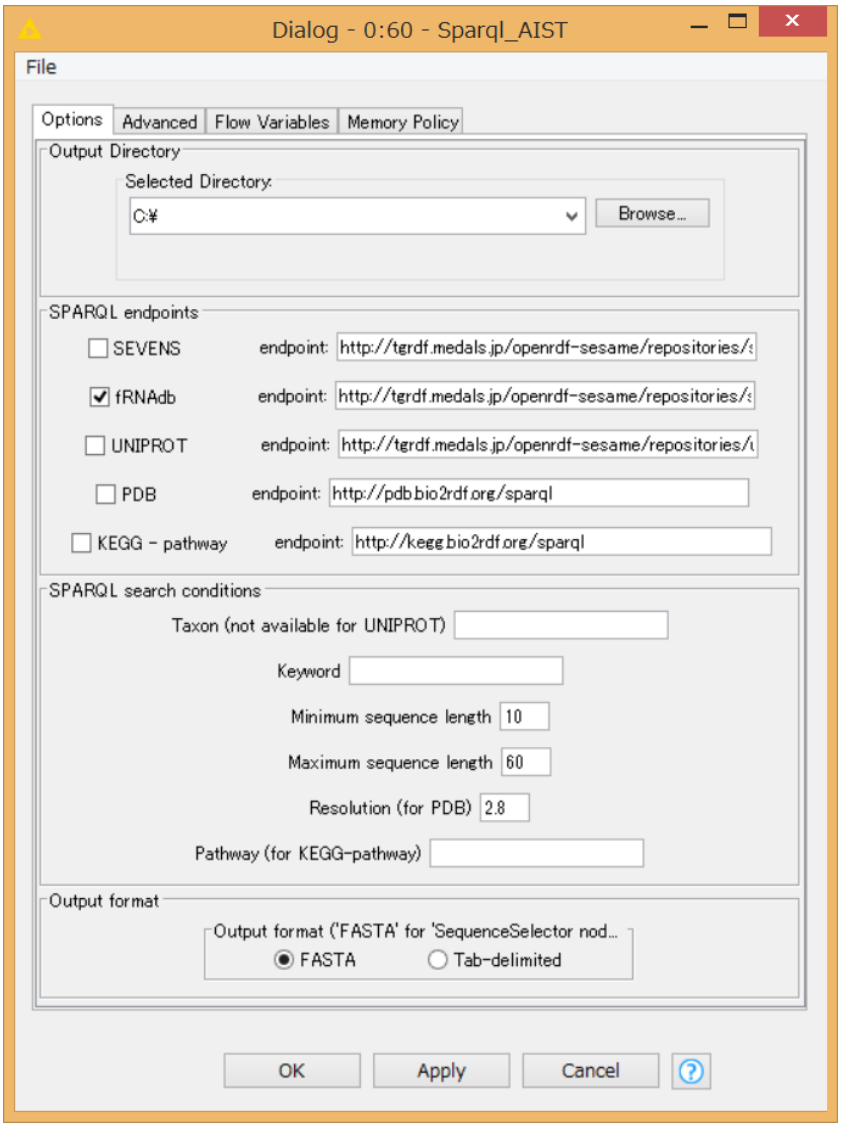
Click “OK” button after setting the conditions.

4.2.2.3 Sparql_AIST

Set AIST SPARQL endpoints and SPARQL conditions.

1) Right-clicking on the Sparql_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.3-1 Sparql_AIST : Configure…

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

Specify an output directory in a text box or “Browse”.

・**Options → SPARQL endpoints:**

Endpoints: SEVENS, fRNAdb, UNIPROT, PDB and KEGG-pathway

Check fRNAdb in this workflow.

・**Options → SPARQL search conditions:**

Specify Taxon, Keyword, Minimum sequence length, Maximum sequence length, Resolution and Pathway. In this workflow, Keyword, Minimum and Maximum sequence length parameters are effective.

・**Options → Output format:**

Specify FASTA or Tab-delimited radio buttons.

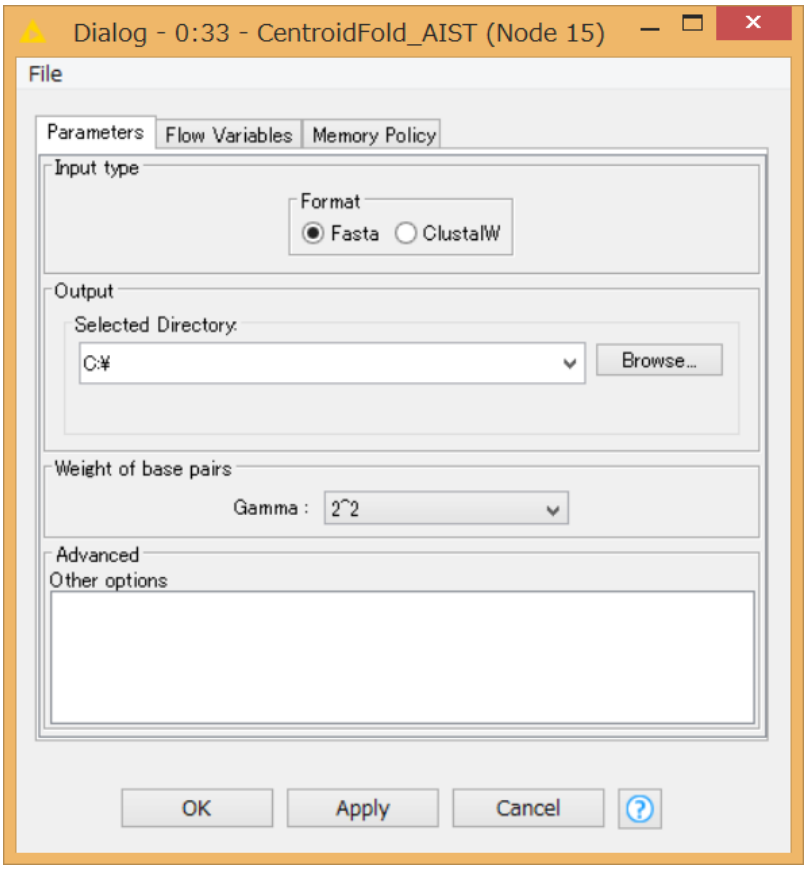
Click “OK” button after setting the conditions.

4.2.2.4 CentroidFold_AIST

Set an output directory and execution parameters.

1) Right-clicking on the CentroidFold_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.4-1 CentroidFold_AIST : Configure…

- **Options → Input type → Format**

  Specify Fasta or ClustalW format.

- **Options → Output → Selected Directory:**

  Specify an output directory in a text box or “Browse”.

- **Options → Weight of base pairs → Gamma:**

  Change Gamma value (default: 2^2).

・**Options → Advanced → Other options**

Specify CentroidFold_AIST execution parameters.

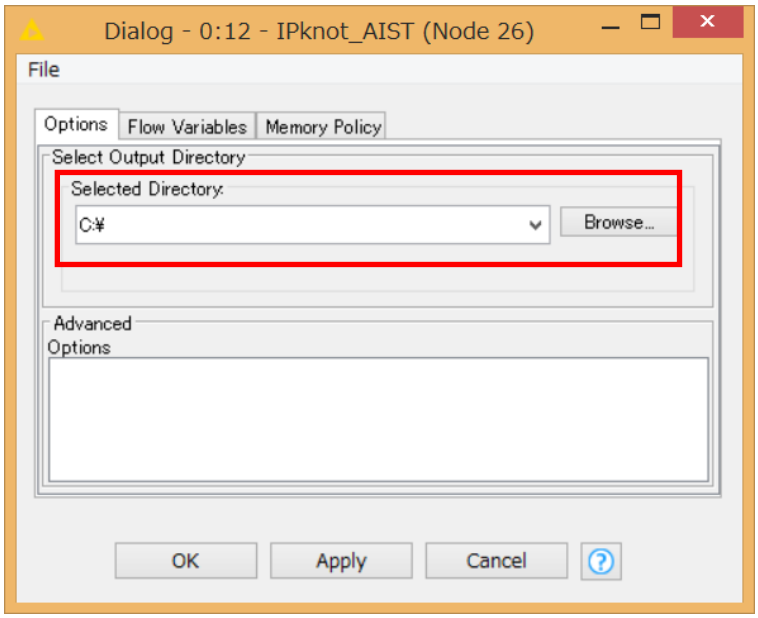
Click “OK” button after setting the conditions.

4.2.2.5 IPknot_AIST

Set an output directory.

1) Right-clicking on the IPknot_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.5-1 IPknot_AIST : Configure…

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

Specify an output directory in a text box or “Browse” (red open rectagular).

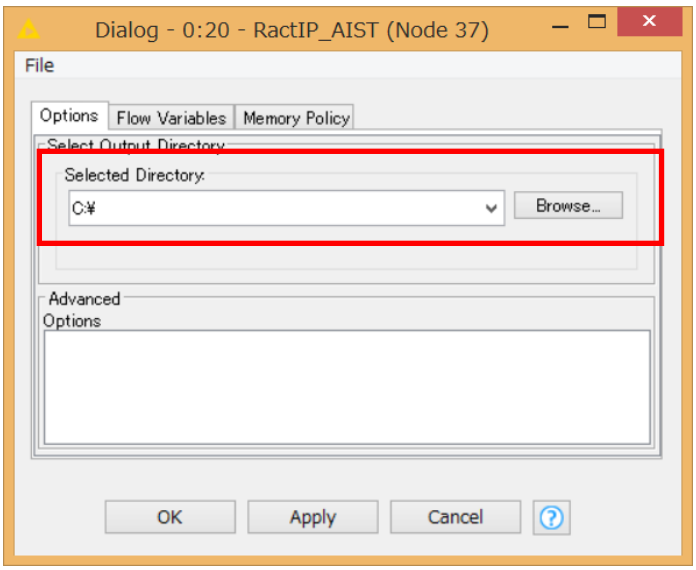
Click “OK” button after setting the conditions.

4.2.2.6 RactIP_AIST

Set an output directory.

1) Right-clicking on the RactIP_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.6-1 RactIP_AIST : Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectagular).

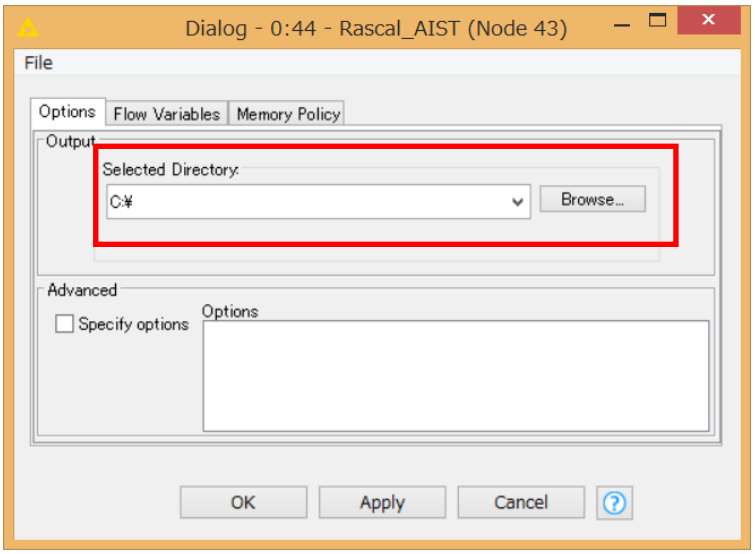
Click “OK” button after setting the conditions.

4.2.2.7 Rascal_AIST

Set an output directory and execution parameters.

1) Right-clicking on the Rascal_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.7-1 Rascal_AIST : Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectagular).

・Options → Advanced:

Specify Rascal execution parameters. Check “Specify options” in case activates Options text box

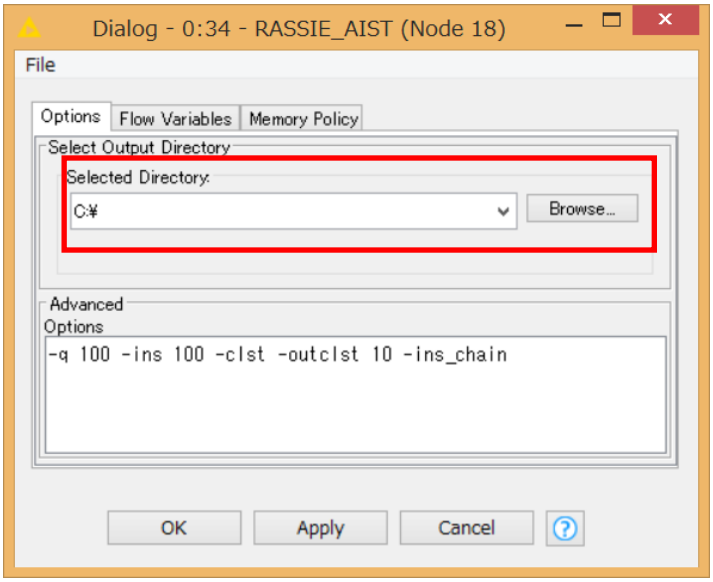
Click “OK” button after setting the conditions.

4.2.2.8 RASSIE_AIST

Set an output directory.

1) Right-clicking on the RASSIE_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



4.2.2.8-1 RASSIE_AIST ：Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectagular).

・Options → Advanced:

Specify RASSIE execution parameters.

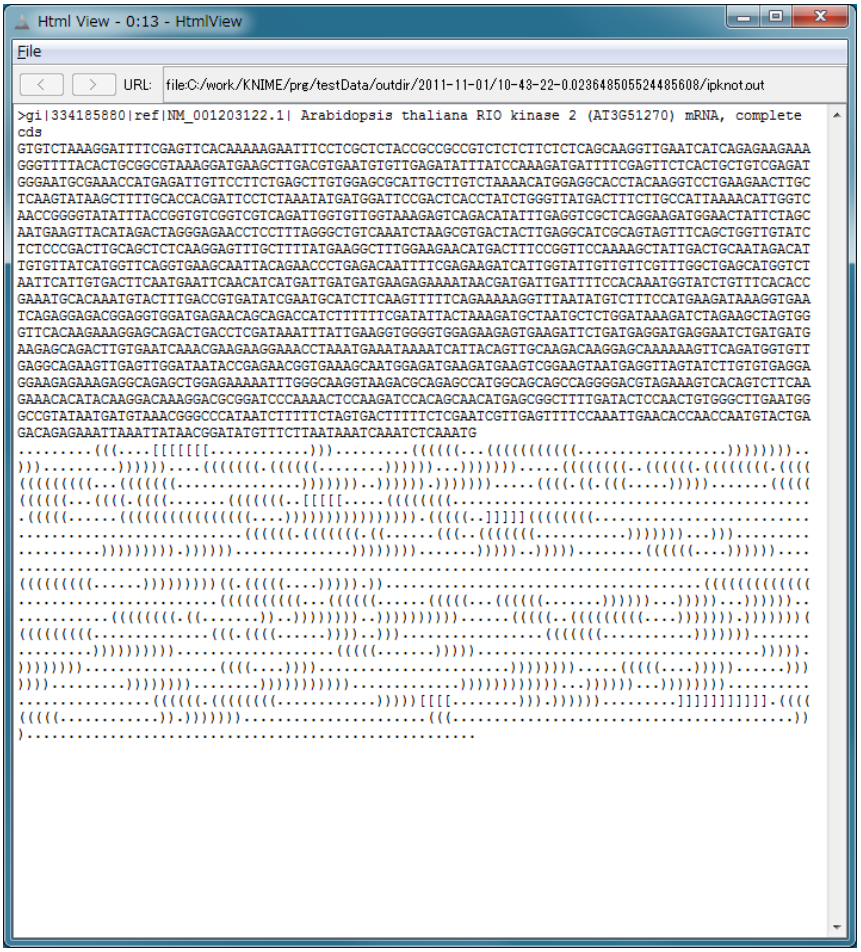
Click “OK” button after setting the conditions.

4.3 Results

4.3.1 IPknot_AIST

1) HtmlView

Results of IPknot_AIST (text format) can be displayed using the HtmlView node.

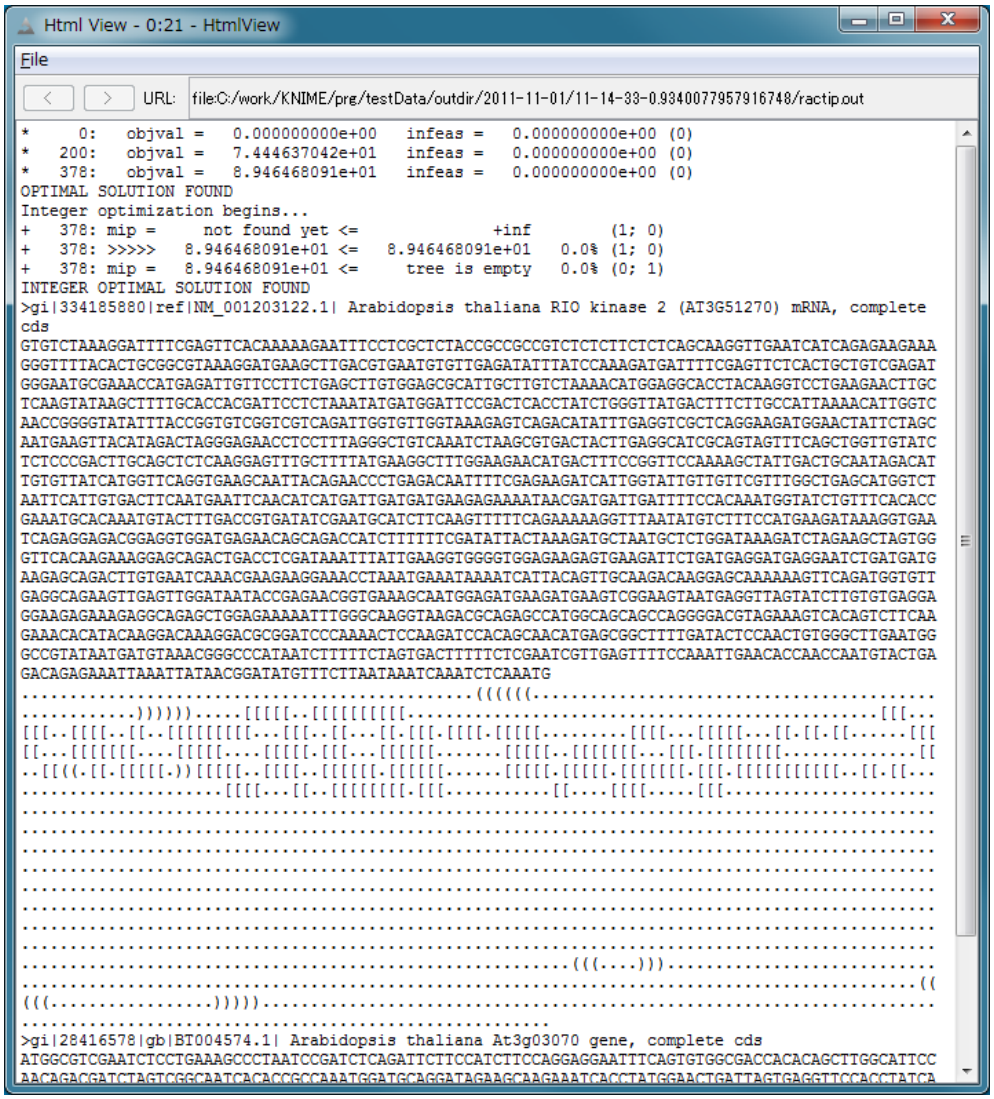


4.3.1-1 HtmlView– IPknot Results

4.3.2 RactIP_AIST

1) HtmlView

Results of RactIP_AIST (text format) can be displayed using the HtmlView node.



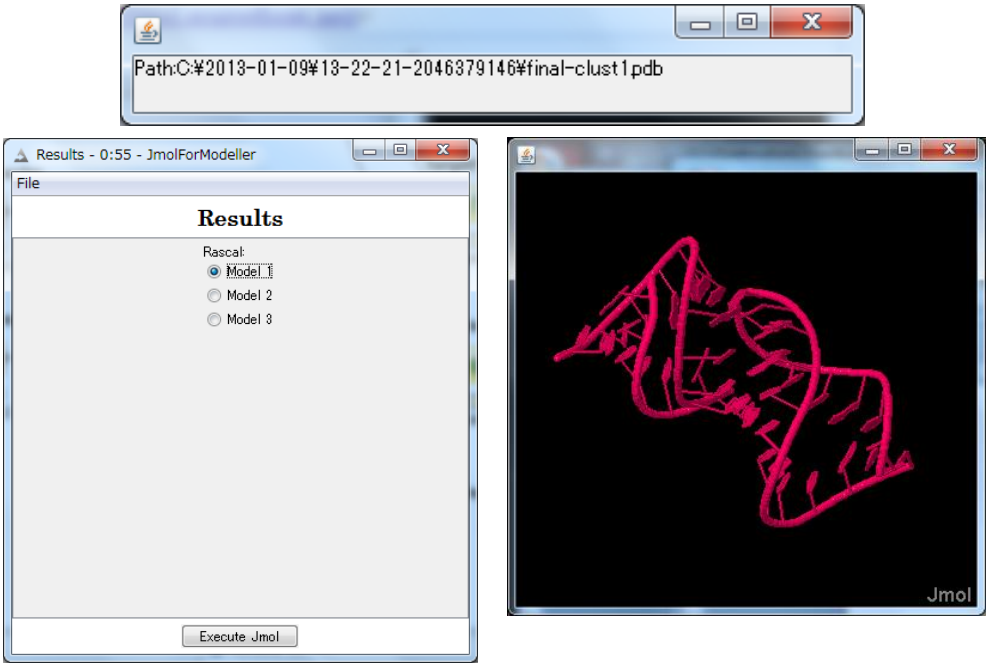
4.3.2-1 HtmlView– RactIP Results

4.3.3 Rascal_AIST

1) JmolForModeller

Results of Rascal_AIST can be displayed using the JmolForModeller node.

Please visit a Jmol web for further information.
Jmol : http://jmol.sourceforge.net/



4.3.3-1 JmolForModeller –Results

Specify a Model number radio button and click [Execute Jmol] button on the bottom of the dialog. After clicking, Jmol starts and displays RNA-RNA interaction structures specified in the dialog.

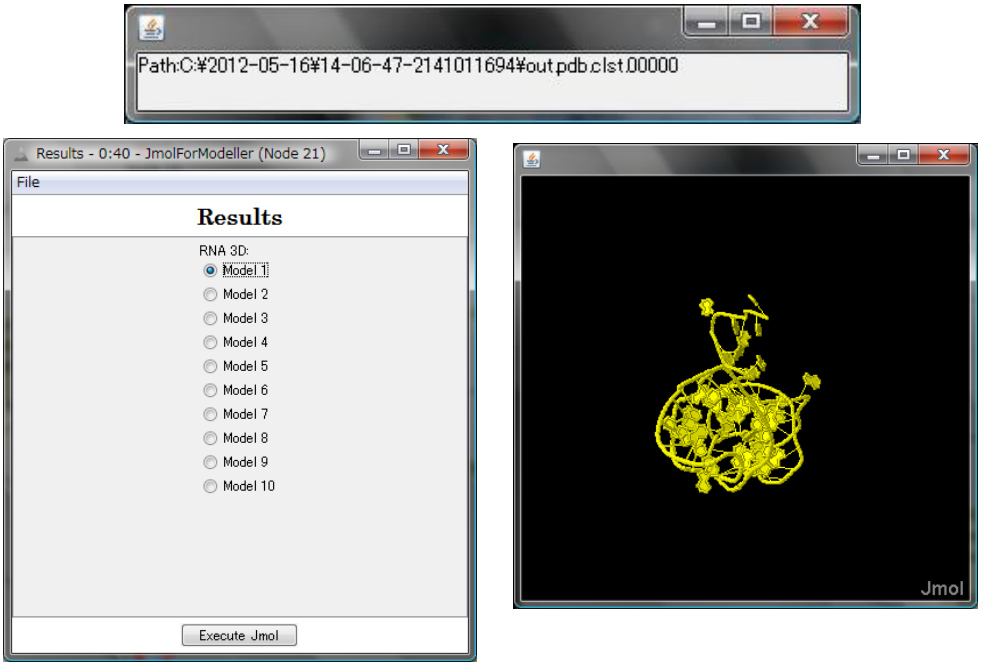
4.3.4 RASSIE_AIST

1) JmolForModeller

Results of RASSIE_AIST can be displayed using the JmolForModeller node.

Please visit a Jmol web for further information.

Jmol : http://jmol.sourceforge.net/



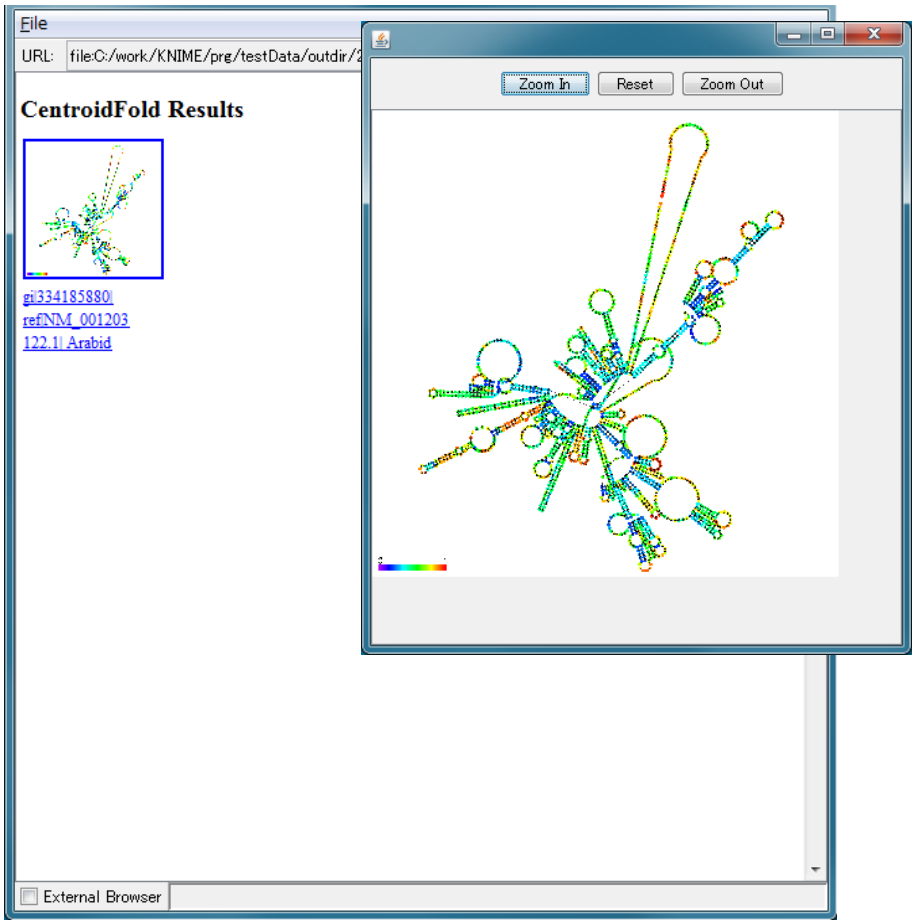
4.3.4-1 JmolForModeller –Results

Specify a Model number radio button and click [Execute Jmol] button on the bottom of the dialog. After clicking, Jmol starts and displays an RNA tertiary structure specified in the dialog.

4.3.5 CentroidFold_AIST

1) AISTViewer

Results of CentroidFold_AIST can be displayed using the AISTViewer node.



4.3.5-1 AISTViewer – CentroidFold Results

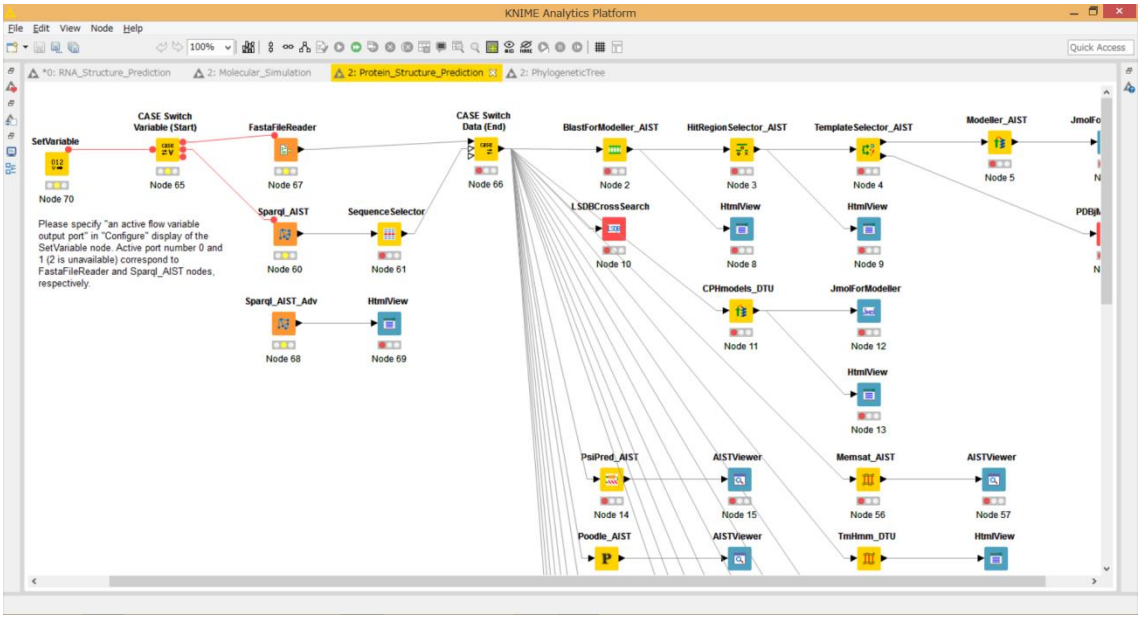
5 Protein Structure Prediction

Protein Structure Prediction is a workflow that performs protein function prediction such as modelling, disorder prediction, post-translational modification prediction, subcellular localization prediction and so on. These analyses are executed through SOAP.

Modelling procedures are as follows:

1) BLAST or PSI-BLAST is executed for a query sequence against PDB sequences detected from ATOM lines of PDB entry files.
2) Appropriate hit regions of PDB sequences are chosen based on E-value, amino acids identity and coverage.
3) MODELLER (http://salilab.org/modeller/) is executed to model protein structures of these hit regions. In this case, if hit regions show high similarities which meet user setting conditions to existing PDB structures, registered structure information of PDBjMine Web site is displayed (http://legacy.ipr.pdbj.org/mine/index.html) without modelling.
4) Modelled structures can be viewed by Jmol (http://jmol.sourceforge.net/).

*MODELLER requires a license key. Please visit a MODELLER web site (http://salilab.org/modeller/registration.html) to get it in advance.



5-1 Protein_Structure_Prediction

5.1 Preparation

This workflow needs a protein FASTA sequence file.

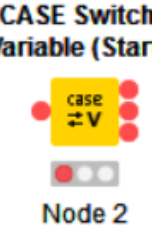
File type
protein FASTA sequence file

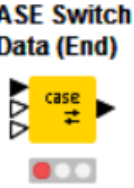
5.2 Nodes

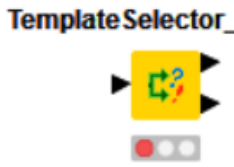
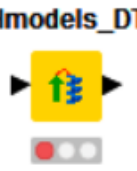
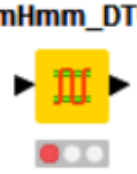
Protein Structure Prediction Active workflow has 34 types of KNIME node. The nodes are shown in the below table. Please check each KNIME node description.

5.2.1 Node list

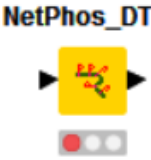
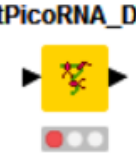
5.2.1-1 Protein Structure Prediction Active Workflow node list

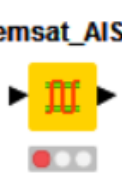
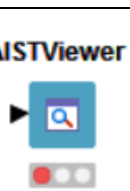
No	Name	Icon	Set	Description
1	SetVariable		Y	Control CASE Switch Variable. Open Dialog: 0: first outport 1: second outport 2: third outport
2	CASE Switch Variable(Start)			Control CASE Switch using workflow variables.

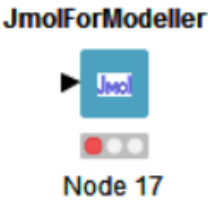
3	FastaFileReader		Y	Set FASTA file.
4	LSDBCrossSearch			Execute LSDB cross search.
5	Sparql_AIST		Y	Execute SPARQL against AIST endpoints.
6	SequenceSelector			Select a sequence from SPARQL results.
7	CASE Switch Variable (End)			CASE Switch end node.
8	BlastForModeller_AIST		Y	Execute BLAST (PSI-BLAST).
9	HitRegionSelector_AIST		Y	Detect hit regions from BLAST (PSI-BLAST) execution results.

10	TemplateSelector_AIST		Y	Choose a modelling template.
11	Modeller_AIST		Y	Execute MODELLER.
12	CPHmodels_DTU		Y	Execute homology modelling.
13	PsiPred_AIST		Y	Execute PsiPred.
14	Poodle_AIST		Y	Execute POODLE.
15	TmHmm_DTU		Y	Execute TMHMM.
16	SignalP_DTU		Y	Execute SignalP.

17	ChloroP_DTU		要	Execute ChloroP.
18	LipoP_DTU		Y	Execute LipoP.
19	NetCTL_DTU		Y	Execute NetCTL.
20	NetChop_DTU		Y	Execute NetChop.
21	NetNES_DTU		Y	Execute NetNES.
22	NetPhosK_DTU		Y	Execute NetPhosK.
18	LipoP_DTU		Y	Execute LipoP.

19	NetCTL_DTU		Y	Execute NetCTL.
20	NetChop_DTU		Y	Execute NetChop.
21	NetNES_DTU		Y	Execute NetNES.
22	NetPhosK_DTU		Y	Exeucte NetPhosK.
23	NetPhos_DTU		Y	Execute NetPhos.
24	NetPicoRNA_DTU		Y	Execute NetPicoRNA.
25	DictyOGlyc_DTU		Y	Execute DictyOGlyc.

26	WolfPsort_AIST		Y	Execute WoLF PSORT.
27	TargetP_AIST		Y	Execute TargetP.
28	SecretomeP_DTU		Y	Execute SecretomeP.
29	DisoPred_AIST		Y	Execute DISOPRED.
30	Memsat_AIST		Y	Execute MEMSAT.
31	Sparql_AIST_Adv		Y	Execute SPARQL against public endpoints.
32	AISTViewer			Display prediction results.

33	HtmlView			Display prediction results.
34	JmolForModeller			Execute Jmol.

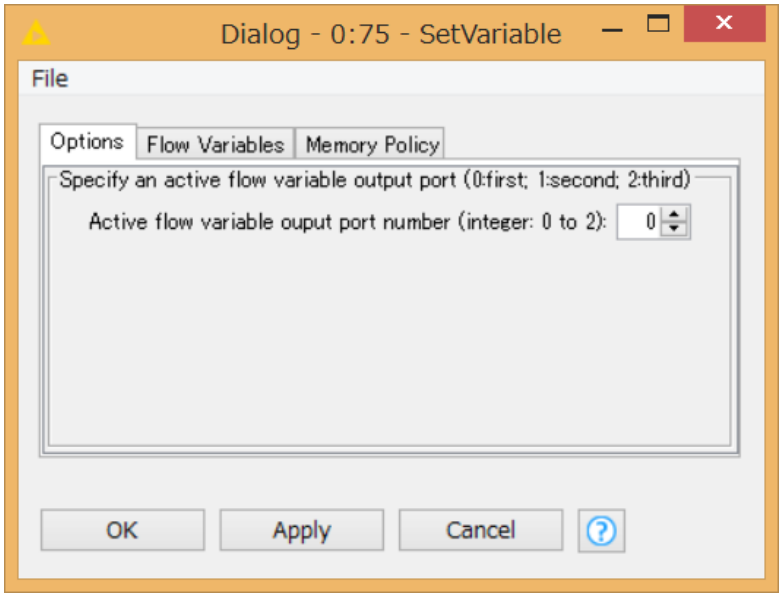
5.2.2 Node setting

5.2.2.1 SetVariable

Set a number to control CASE Switch Variable (Start) outports using workflow variables.

1) Right-clicking on the SetVariable node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.1-1 SetVariable : Configure…

0: CASE Switch Variable(Start) first outport

1: CASE Switch Variable(Start) second outport

2: CASE Switch Variable(Start) third outport

In the workflow, set 0 and 1 for FastaFileReader and Sparql_AIST, respectively.

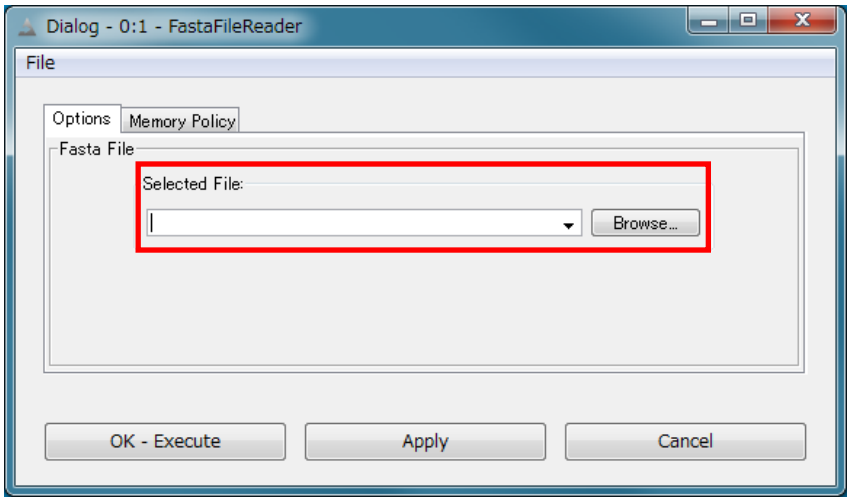
Click “OK” button after setting the conditions.

5.2.2.2 FastaFileReader

Set a protein FASTA sequence file.

1) Right-clicking on the FastaFileReader node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.2-1 FastaFileReader : Configure…

・Options → Fasta File → SelectedFile:

Specify a protein FASTA sequence file in a text box or “Browse” (red open rectangular)

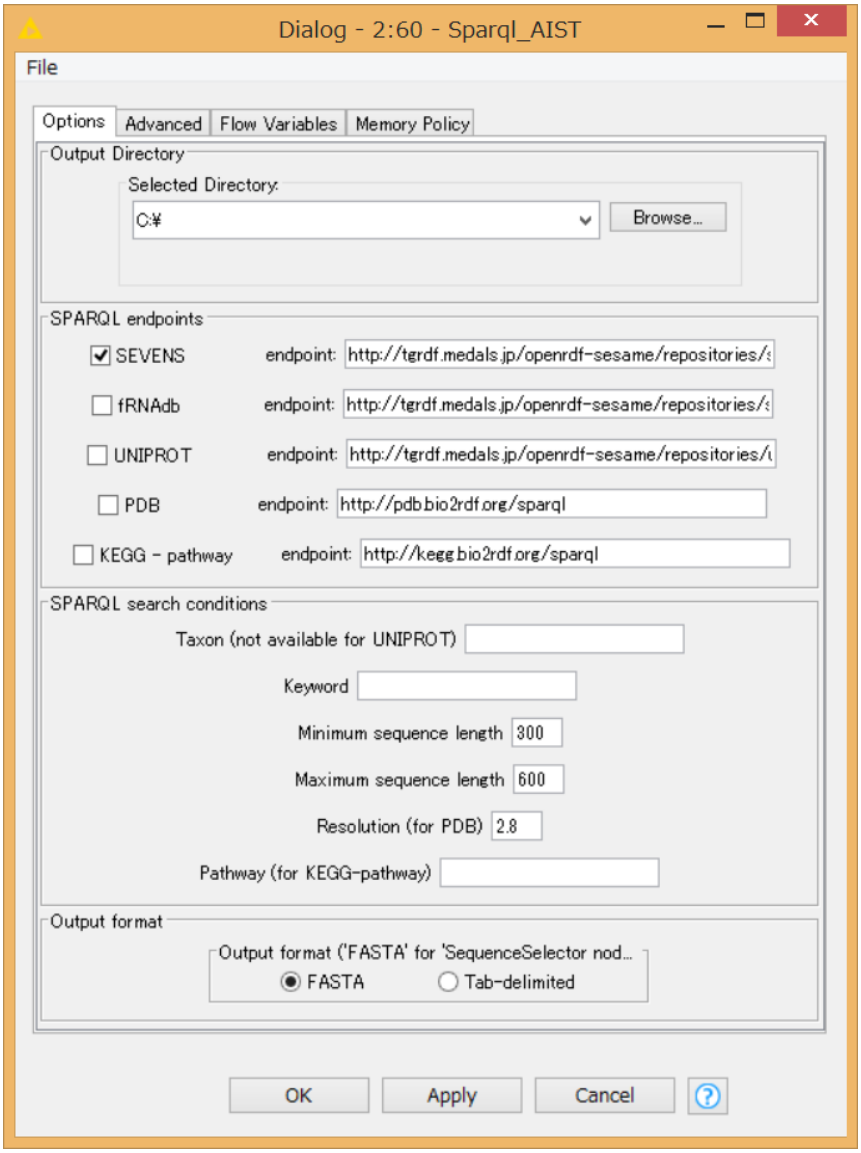
Click “OK” button after setting the conditions.

5.2.2.3 Sparql_AIST

Set AIST SPARQL endpoints and SPARQL conditions.

1) Right-clicking on the Sparql_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.3-1 Sparql_AIST : Configure…

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

Specify an output directory in a text box or “Browse”.

・**Options → SPARQL endpoints:**

Endpoints: SEVENS, fRNAdb, UNIPROT, PDB and KEGG-pathway

Check SEVENS, UNIPROT and/or PDB in this workflow.

・**Options → SPARQL search conditions:**

Specify Taxon, Keyword, Minimum sequence length, Maximum sequence length, Resolution and Pathway. In this workflow, Keyword, Minimum and Maximum sequence length, Resolution parameters are effective.

・**Options → Output format:**

Specify either FASTA or Tab-delimited radio button.

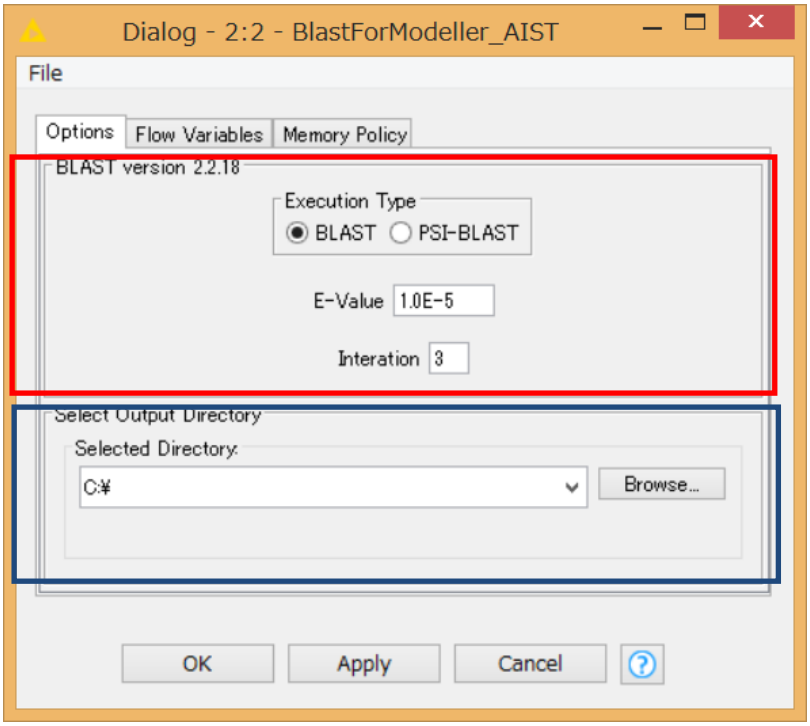
Click “OK” button after setting the conditions.

5.2.2.4 BlastForModeller _AIST

Specify either BLAST or PSI-BLAST program and set an output directory.

1) Right-clicking on the BlastForModeller_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.4-1 BlastForModeller_AIST : Configure…

- **Options → BLAST version 2.2.18 → Execution Type**

  Specify Execution Type (BLAST or PSI-BLAST) (open red rectangular)

- **Options → BLAST version 2.2.18 → E-Value**

  Specify E-value (default: 1.0E-5) (open red rectangular)

- **Options → BLAST version 2.2.18 → Iteration**

  Specify max number of Iterations for PSI-BLAST (default: 3) (open red rectangular).

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

Specify an output directory in a text box or “Browse” (blue open rectangular).

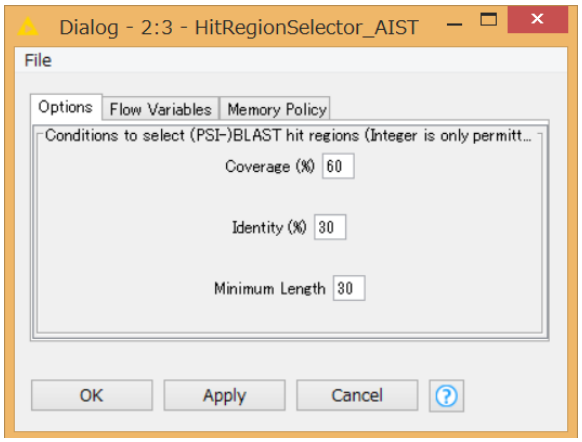
Click “OK” button after setting the conditions.

5.2.2.5 HitRegionSelector_AIST

Specify the conditions for hit regions detected by BLAST (PSI-BLAST) to use modelling.

1) Right-clicking on the HitRegionSelector_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.5-1 HitRegionSelector_AIST : Configure…

・**Options → Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) → Coverage(%)**

Specify Coverage (%) (default: 60).

・**Options → Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) → Identity(%)**

Specify Identity (%) (default: 30)

・Options → Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) → Minimum Length

Specify minimum length of amino acid sequence (default: 30).

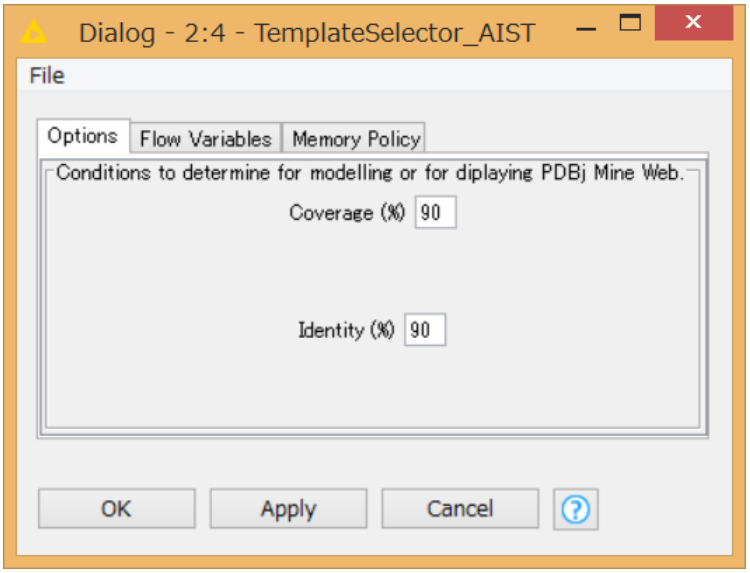
Click “OK” button after setting the conditions.

5.2.2.6 TemplateSelector_AIST

Specify Coverage and Identity.

1) Right-clicking on the TemplateSelector_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.6-1 TemplateSelector_AIST : Configure…

・Options → Condition to determine for modelling or for displaying PDBj Mine Web. → Coverage(%), Identity(%)

Specify Coverage and Identity (default: Coverage: 90, Identity: 90).

Determine a template for protein modelling. If the coverage and identity of hit regions don’t meet the template selecting conditions specified by user, protein modelling is executed by MODELLER. If not so, a web browser opens a PDBjMine web site and displays PDB structure information for the hit regions.

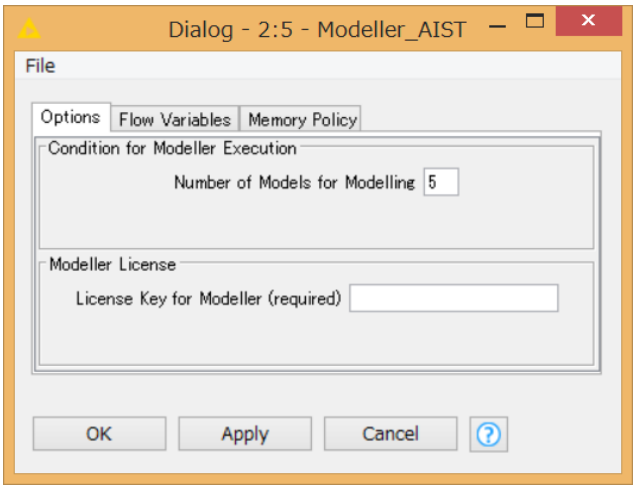
Click “OK” button after setting the conditions.

5.2.2.7 Modeller_AIST

Specify maximum number of models and input a MODELLER license key.

1) Right-clicking on the Modeller_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.7-1 Modeller_AIST : Configure…

・Options → Condition for Modeller Execution → Number of Models for Modelling

Specify maximum number of models (default: 3)

Range: 1 ~ 10

・Options → Modeller License → License Key for Modeller (required)

Input a MODELLER license key (required).

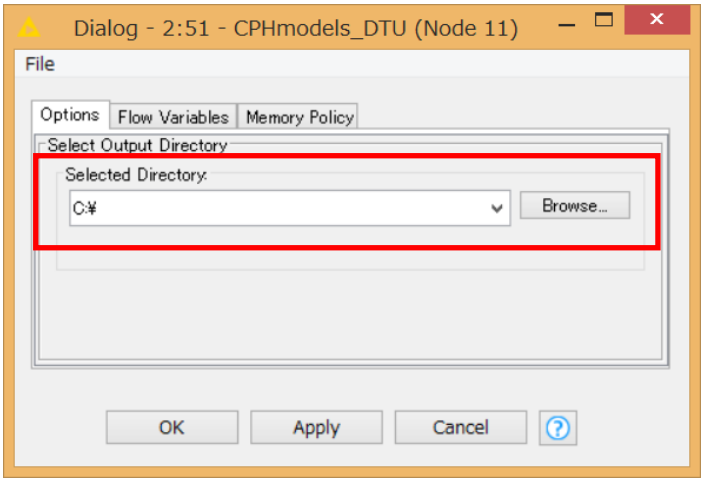
Click “OK” button after setting the conditions.

5.2.2.8 CPHmodels_DTU

Set an output directory.

1) Right-clicking on the CPHmodels_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.8-1 CPHmodels_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

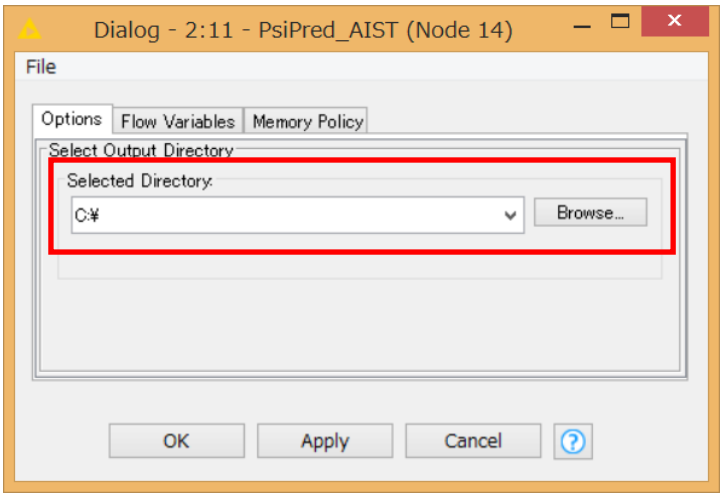
Click “OK” button after setting the conditions.

5.2.2.9 PsiPred_AIST

Set an output directory.

1) Right-clicking on the PsiPred_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.9-1 PsiPred_AIST : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

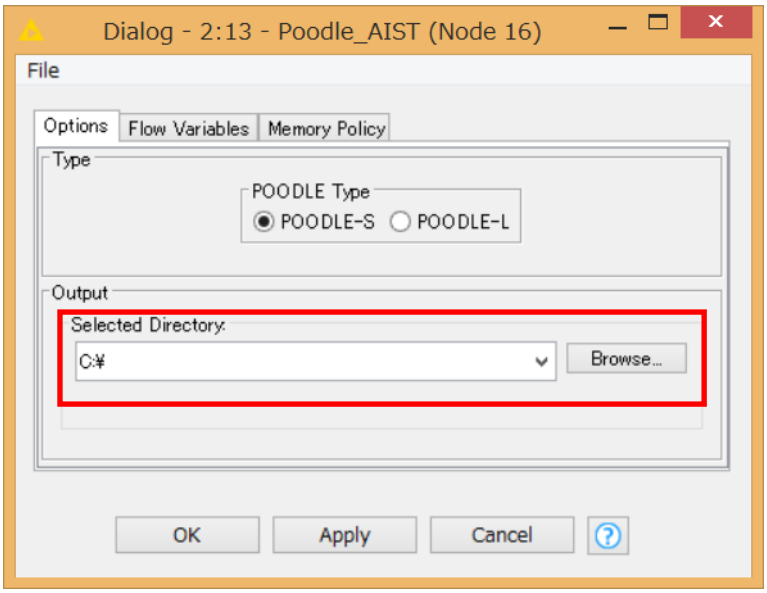
Click “OK” button after setting the conditions.

5.2.2.10 Poodle_AIST

Set an output directory.

1) Right-clicking on the Poodle_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.10-1 Poodle_AIST : Configure…

・Options → Type → POODLE Type

Select POODEL program.

POODLE-S: for short disorder region prediction.

POODLE-L: for long (40 aa and over) disorder region prediction.

・Options → Output → Selected Directory :

Specify an output directory in a text box or “Browse” (red open rectangular).

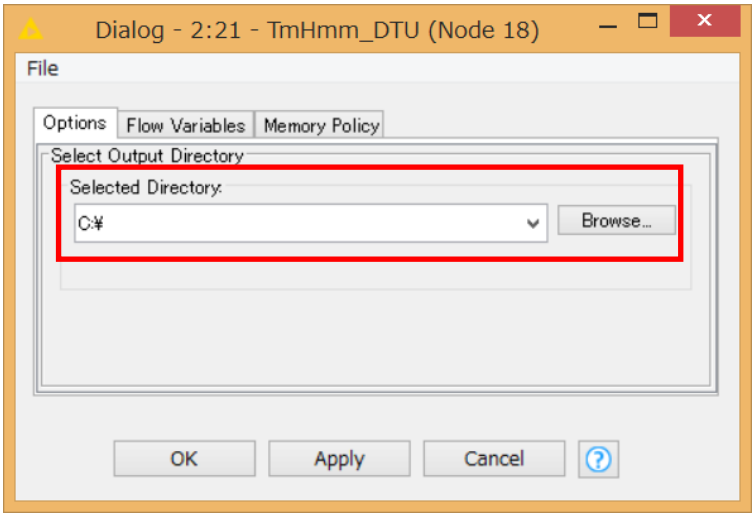
Click “OK” button after setting the conditions.

5.2.2.11 TmHmm_DTU

Set an output directory.

1) Right-clicking on the TmHmm_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.11-1 TmHmm_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

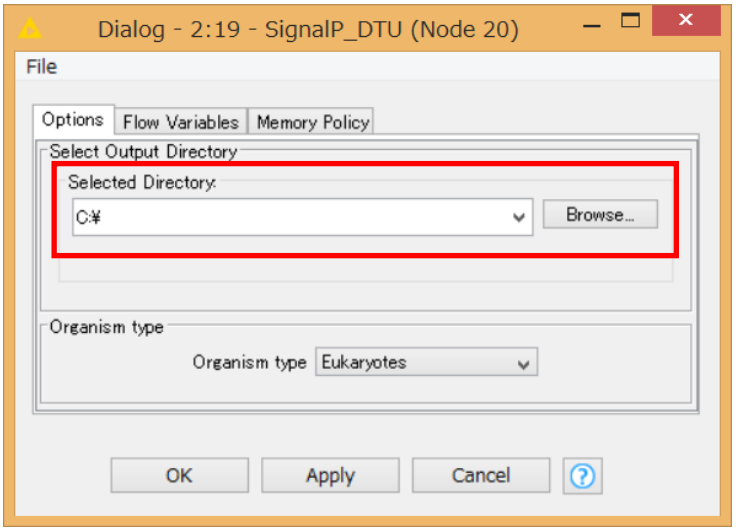
Click “OK” button after setting the conditions.

5.2.2.12 SignalP_DTU

Set an output directory.

1) Right-clicking on the SignalP_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.12-1 SignalP_DTU : Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectangular).

・Options → Organism Type:

Selectg Eukaryotes, Gram-negative bacteria or Gram-positeve bacteria.

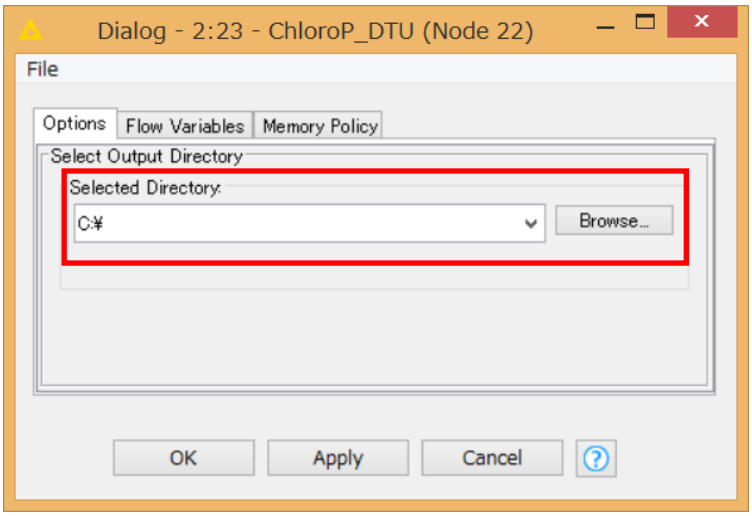
Click “OK” button after setting the conditions.

5.2.2.13 ChloroP_DTU

Set an output directory.

1) Right-clicking on the ChloroP_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.13-1 ChloroP_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

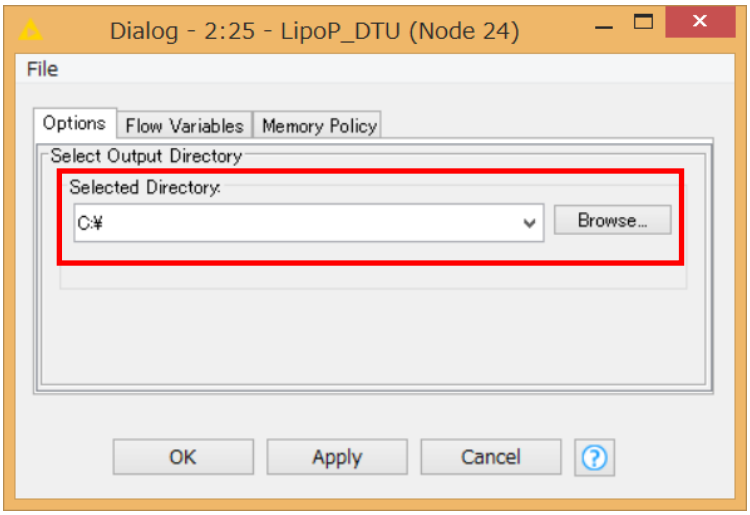
Click “OK” button after setting the conditions.

5.2.2.14 LipoP_DTU

Set an output directory.

1) Right-clicking on the LipoP_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.14-1 LipoP_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

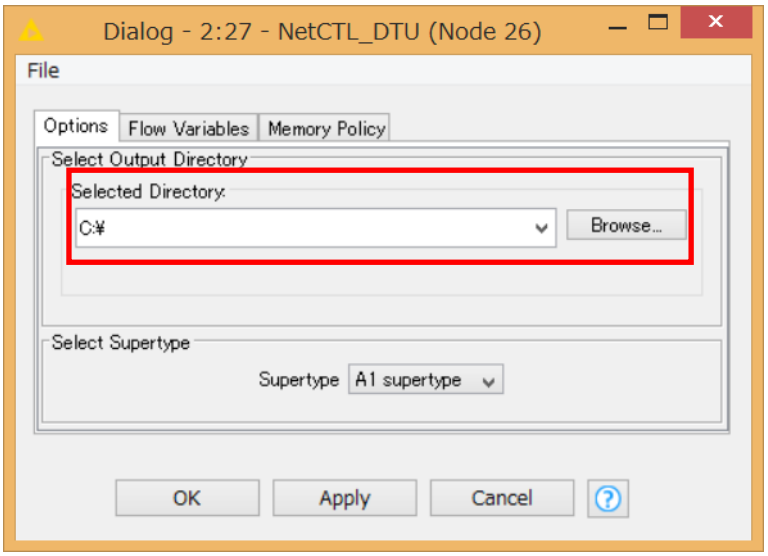
Click “OK” button after setting the conditions.

5.2.2.15 NetCTL_DTU

Set an output directory and ”Supertype”.

1) Right-clicking on the NetCTL_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.15-1 NetCTL_DTU : Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectangular).

・Options → Select Supertype:

Choose a Supertype (A1~A3, A24, A26, B7, B8, B27, B39, B44, B58 or B62)

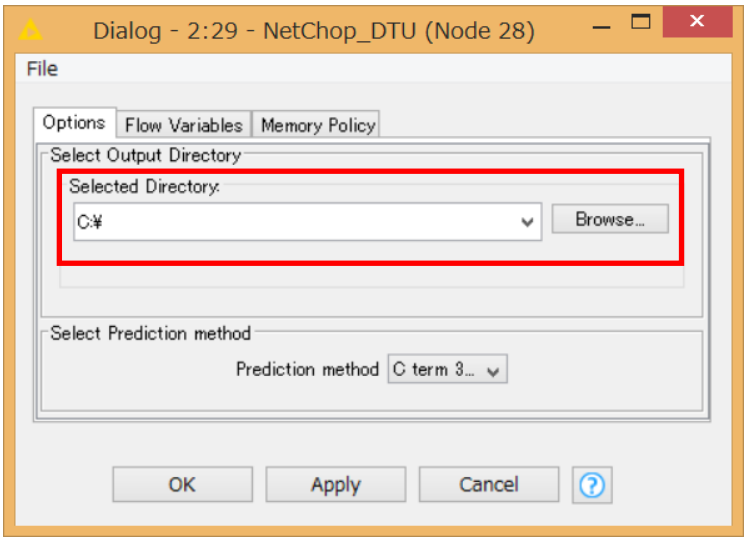
Click “OK” button after setting the conditions.

5.2.2.16 NetChop_DTU

Set an output directory and ”Prediction method”.

1) Right-clicking on the NetChop_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.16-1 NetChop_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

・Options → Select Prediction method:
Choose either “C term 3.0” or “20S 3.0”.

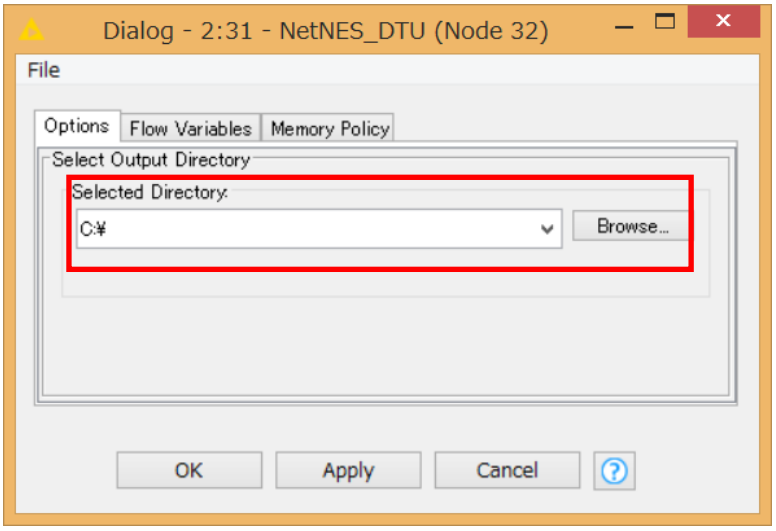
Click “OK” button after setting the conditions.

5.2.2.17 NetNES_DTU

Set an output directory.

1) Right-clicking on the NetNES_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.17-1 NetNES_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

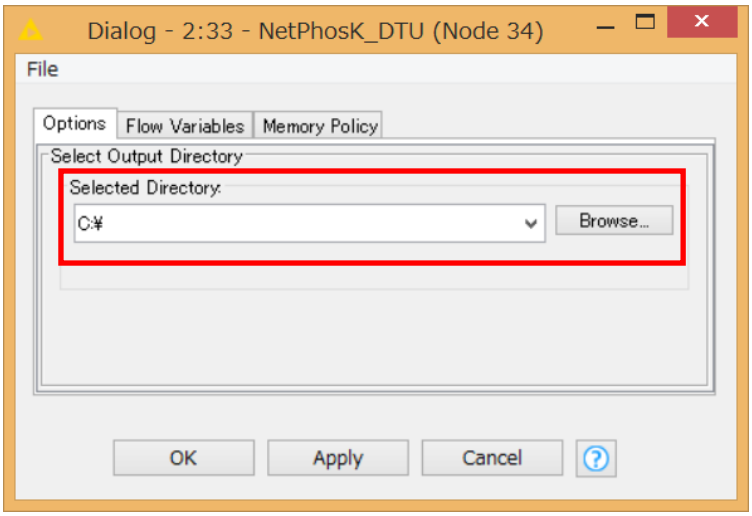
Click “OK” button after setting the conditions.

5.2.2.18 NetPhosK_DTU

Set an output directory.

1) Right-clicking on the NetPhosK_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.18-1 NetPhosK_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

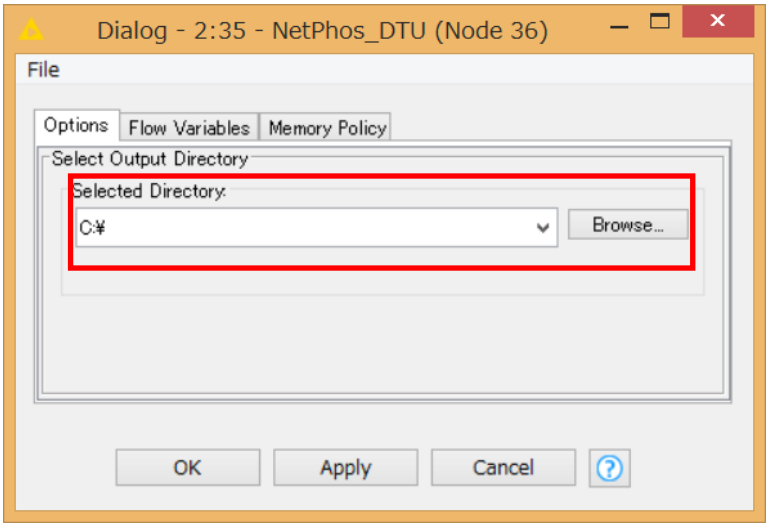
Click “OK” button after setting the conditions.

5.2.2.19 NetPhos_DTU

Set an output directory.

1) Right-clicking on the NetPhos_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.19-1 NetPhos_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

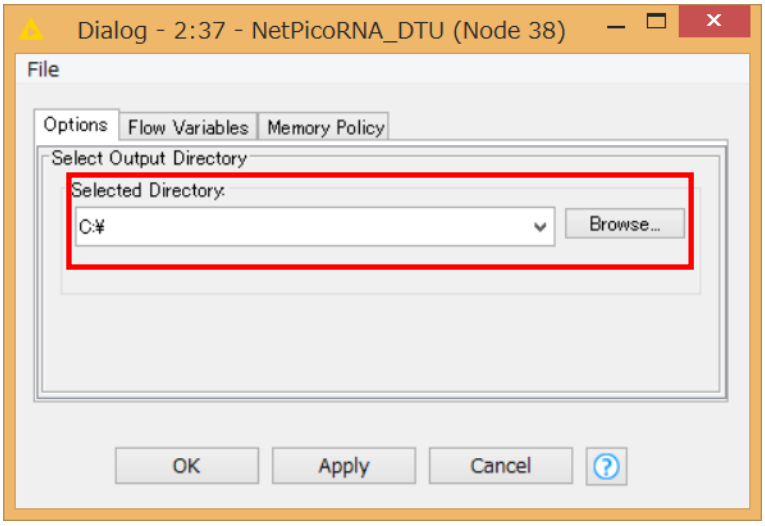
Click “OK” button after setting the conditions.

5.2.2.20 NetPicoRNA_DTU

Set an output directory.

1) Right-clicking on the NetPicoRNA_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.20-1 NetPicoRNA_DTU ：Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

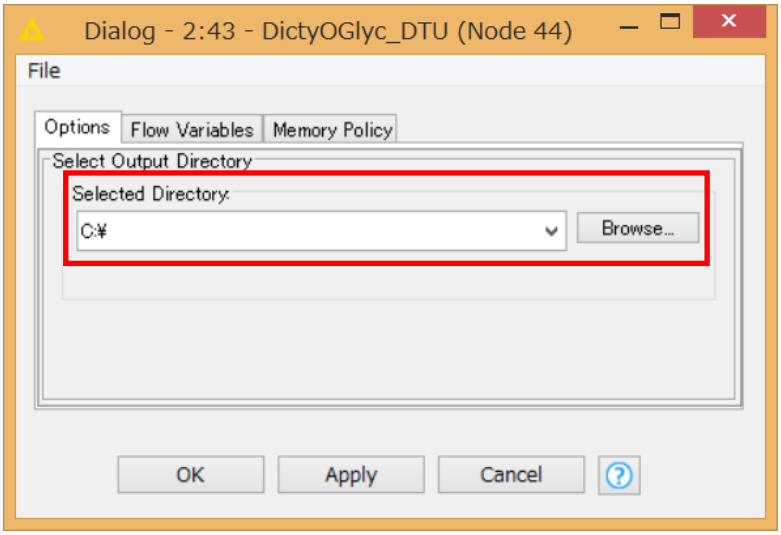
Click “OK” button after setting the conditions.

5.2.2.21 DictyOGlyc_DTU

Set an output directory.

1) Right-clicking on the DictyOGlyc_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.21-1 DictyOGlyc_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

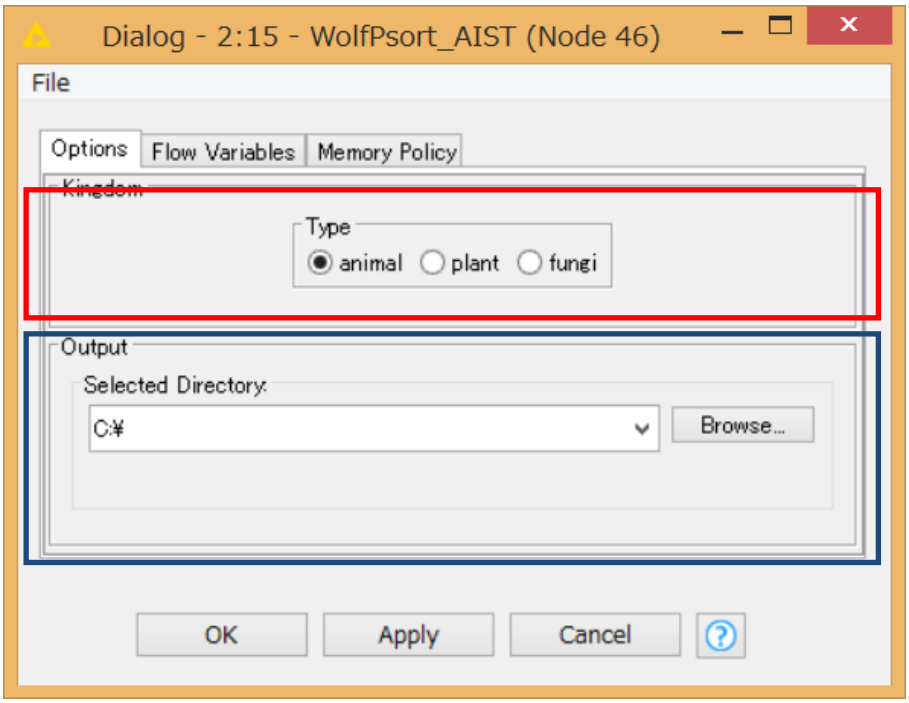
Click “OK” button after setting the conditions.

5.2.2.22 WolfPsort_AIST

Set an output directory and kingdom.

1) Right-clicking on the WolfPsort_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.22-1 WolfPsort_AIST ：Configure…

・**Options → Kingdom → Type**

Choose animal, plant or fungi (red open rectangular).

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

Specify an output directory in a text box or “Browse” (blue open rectangular).

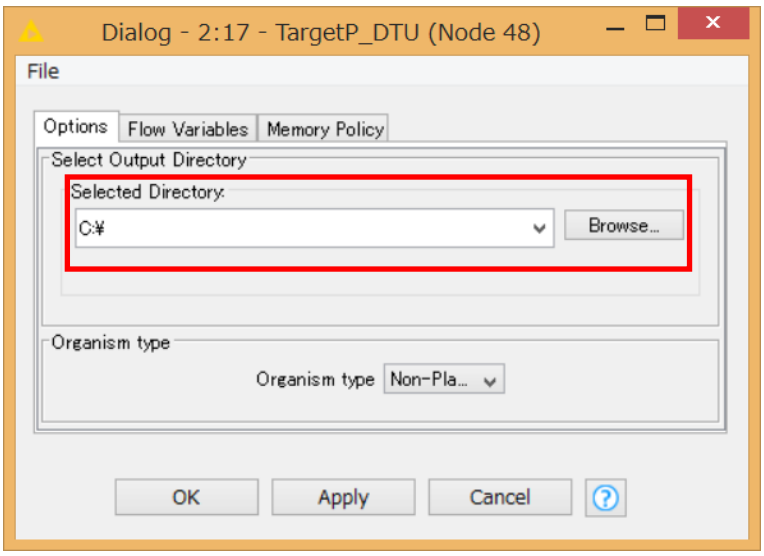
Click “OK” button after setting the conditions.

5.2.2.23 TargetP_DTU

Set an output directory and kingdom.

1) Right-clicking on the TargetP_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.23-1 TargetP_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

・Options → Organism type:
Choose either Non-Plant or Plant.

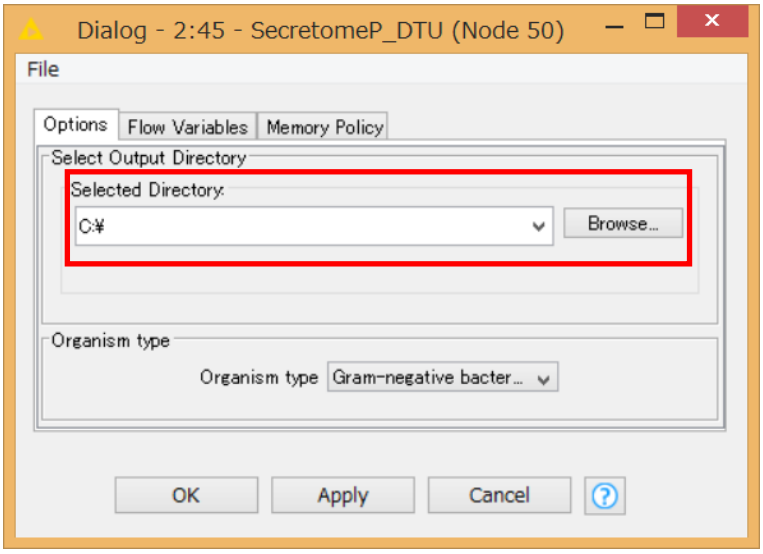
Click “OK” button after setting the conditions.

5.2.2.24 SecretomeP_DTU

Set an output directory and organism type.

1) Right-clicking on the SecretomeP_DTU node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.24-1 SecretomeP_DTU : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

・Options → Organism type:
Choose Gram-negative bacteria, Gram-positive bacteria or Mammalian.

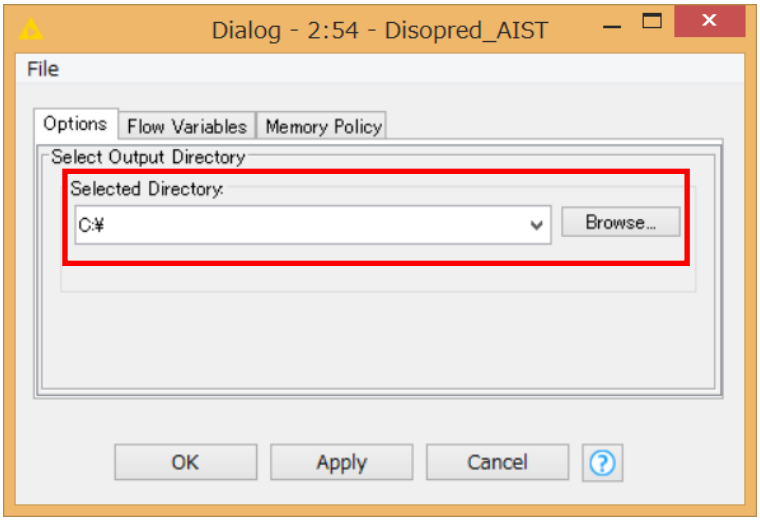
Click “OK” button after setting the conditions.

5.2.2.25 DisoPred_AIST

Set an output directory.

1) Right-clicking on the DisoPred_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.25-1 DisoPred_AIST : Configure…

・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

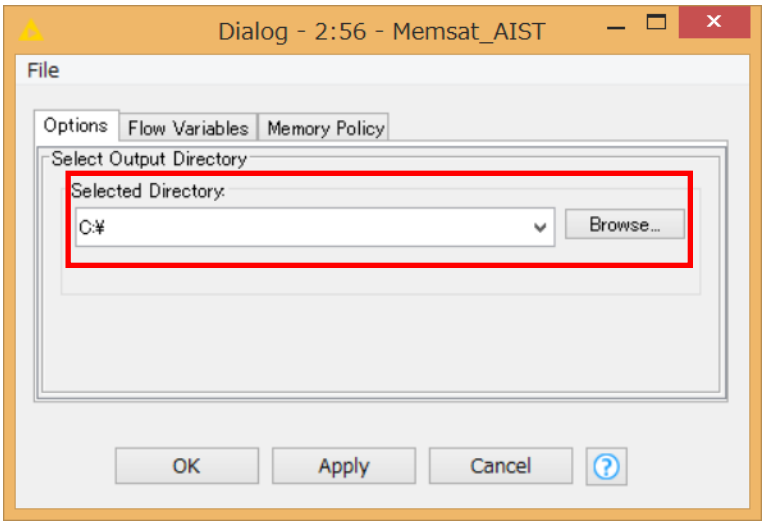
Click “OK” button after setting the conditions.

5.2.2.26 Memsat_AIST

Set an output directory.

1) Right-clicking on the Memsat_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



5.2.2.26-1 Memsat_AIST : Configure…

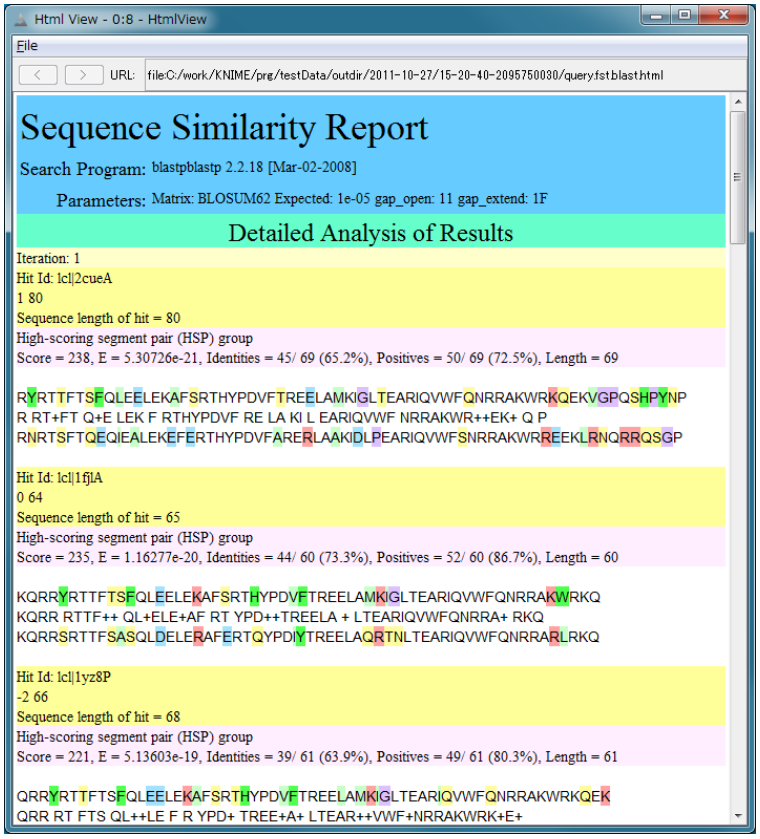
・Options → Select Output Directory → Selected Directory:
Specify an output directory in a text box or “Browse” (red open rectangular).

Click “OK” button after setting the conditions.

5.3 Results

5.3.1 BlastForModeller_AIST → HtmlView

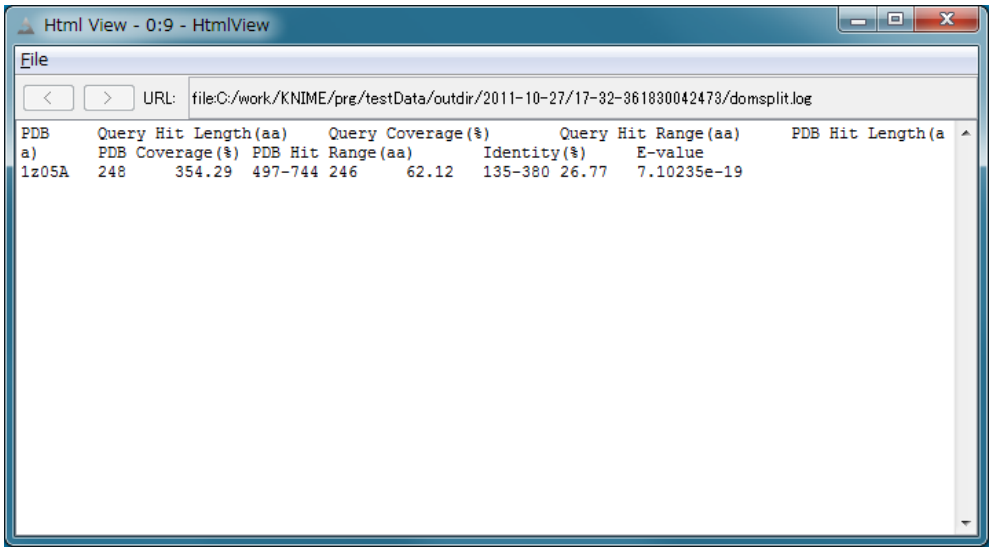
Results of BlastForModeller_AIST can be displayed using the HtmlView node.



5.3.1-1 BlastForModeller_AIST execution results (HtmlView)

5.3.2 HitRegionSelector_AIST → HtmlView

Results of HitRegionSelector_AIST can be displayed using the HtmlView node.



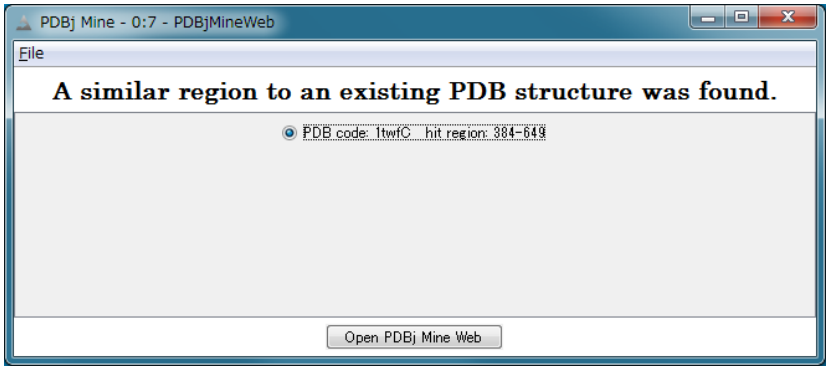
5.3.2-1 HitRegionSelector_AIST execution results (HtmlView)

5.3.3 TemplateSelector_AIST → PDBjMineWeb

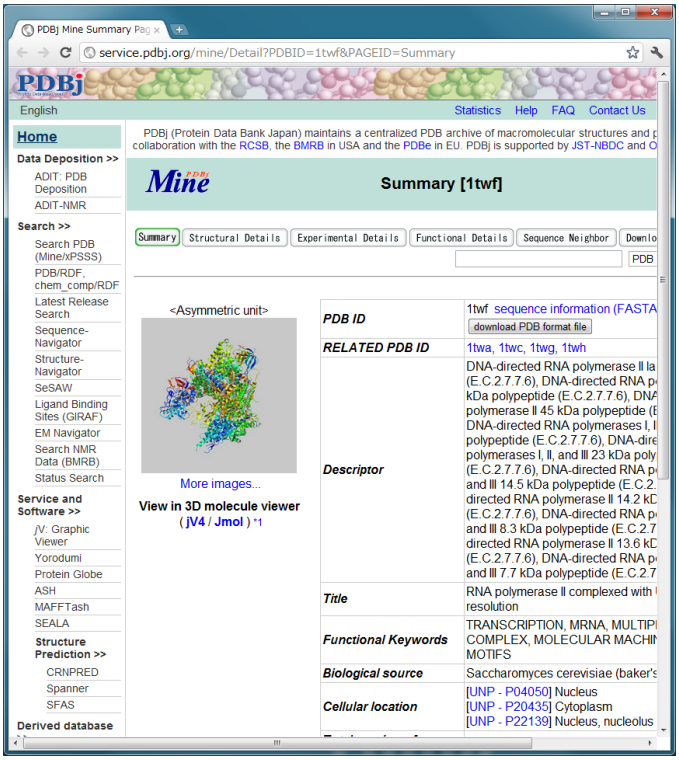
If an identity and coverage of hit regions are high compared with user setting conditions, results of TemplateSelector_AIST can be displayed using PDBjMineWeb node.

First, a dialog window is opened and PDB code and hit region range of its structure are displayed with a radio button.

By selecting a radio button and clicking “Open PDBj Mine Web” button, a PDBjMine web site is opened and displays a corresponding structure information.



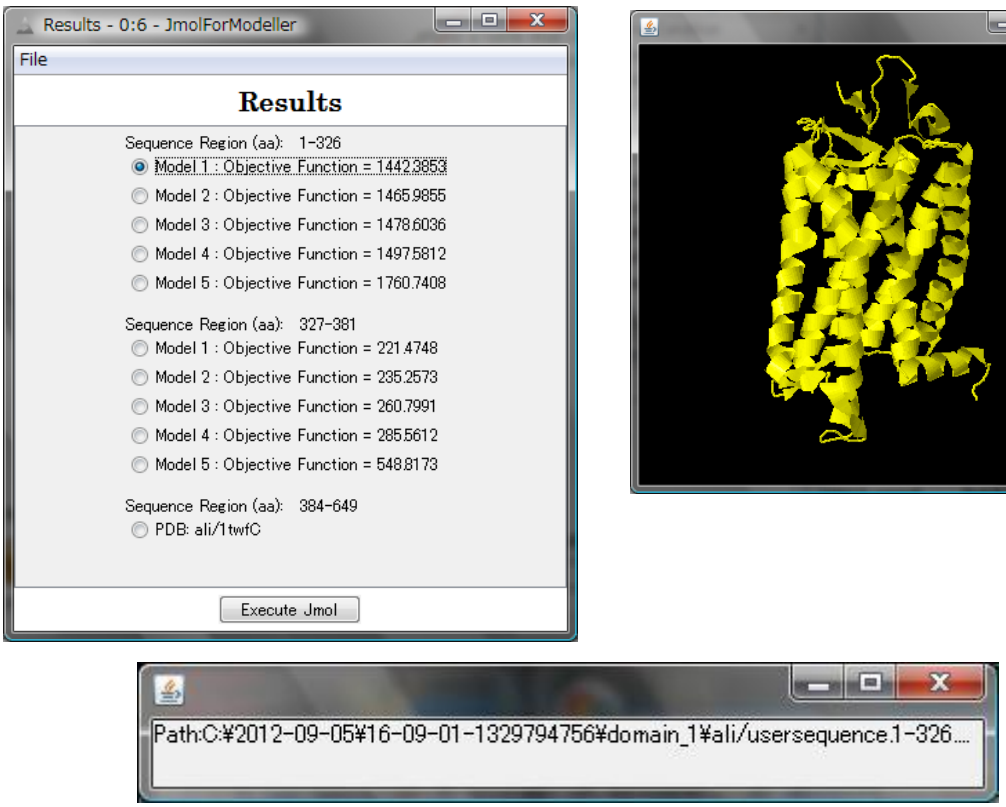
5.3.3-1 PDBjMineWeb – PDBj Mine



5.3.3-2 PDBjMineWeb – PDBj Mine

5.3.4 Modeller_AIST → JmolForModeller

Results of Modeller_AIST can be displayed using JmolForModeller node.



5.3.4-1 JmolForModeller – Modeller Results

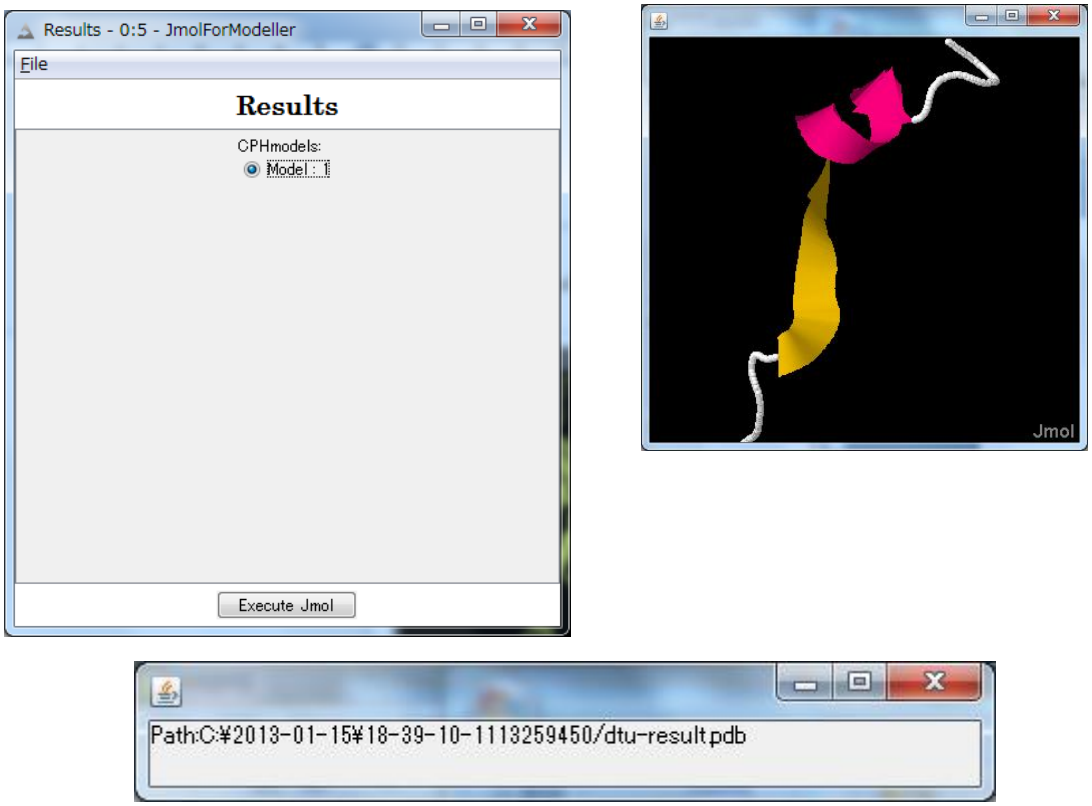
First, a dialog window is opened and models and these objective functions (modelled structure only) are displayed in every hit region category. By selecting a radio button corresponded to each model and clicking “Execute Jmol” button, Jmol starts and displays a specified structure.

Please a Jmol web site about how to operate Jmol.

Jmol : http://jmol.sourceforge.net/

5.3.5 CPHmodels_DTU → JmolForModeller

Results of CPHmodels_DTU can be displayed using JmolForModeller.



5.3.5-1 JmolForModeller – CPHmodels Results

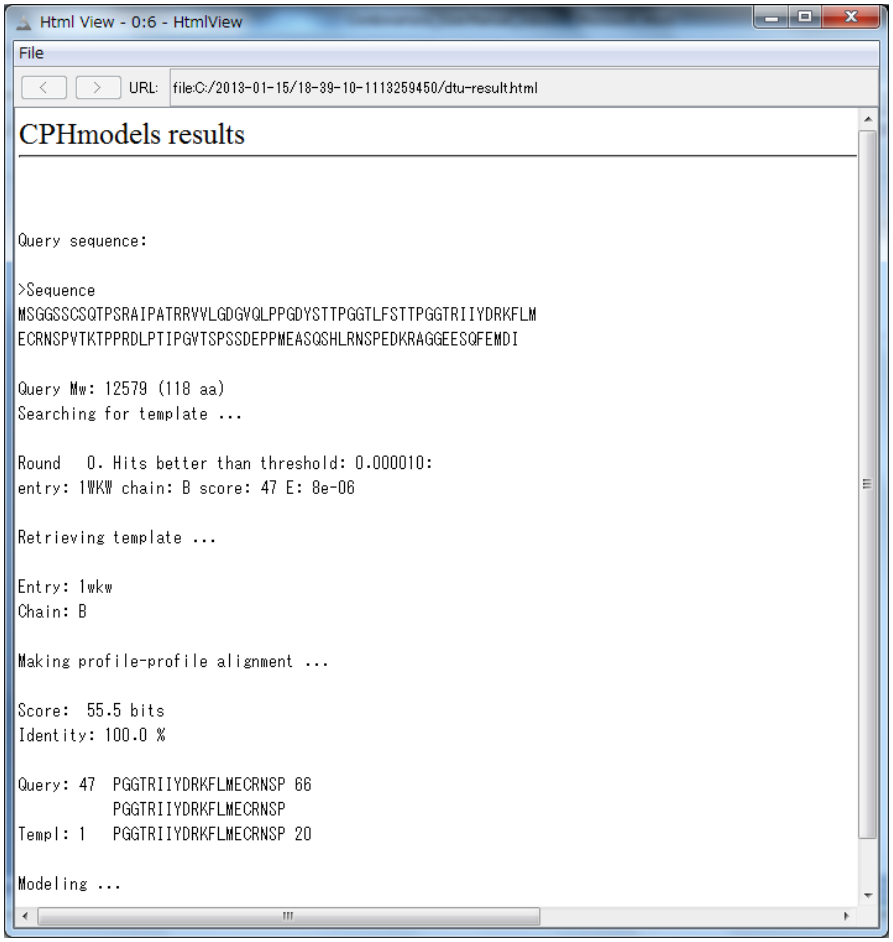
First, a dialog window is opened and a model with a radio button is displayed. By clicking a "Execute Jmol" buton, Jmol starts and displays a modelling structure.

Please a Jmol web site about how to operate Jmol.

Jmol : http://jmol.sourceforge.net/

5.3.6 CPHmodels_DTU → HtmlView

Results of CPHmodels_DTU can be displayed using HtmlView node.



5.3.6-1 CPHmodels_DTU execution results (HtmlView)

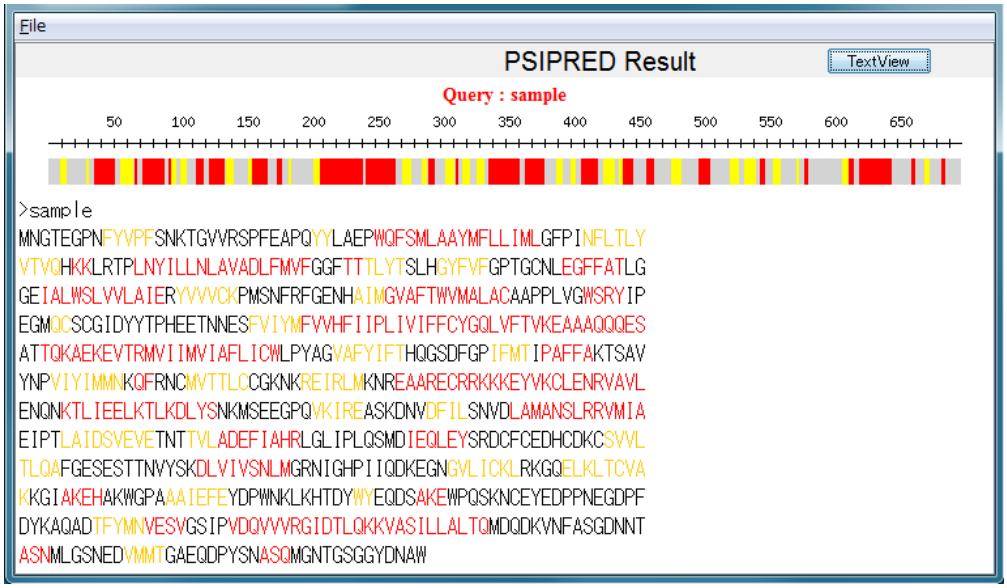
5.3.7 PsiPred_AIST → AISTViewer

Results of PsiPred_AIST can be displayed using AISTViewer node.

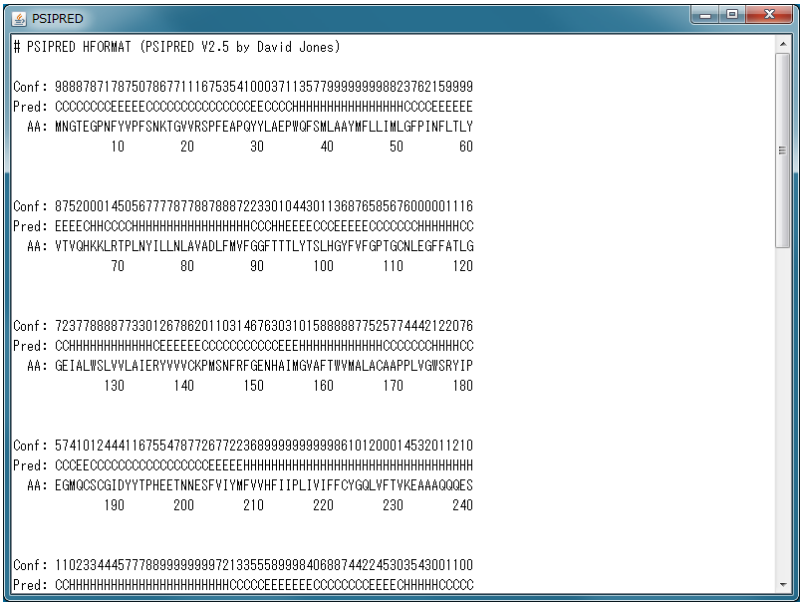
By clicking a “TextView” button on the top-right corner of the screen, a pop up window is opened and text format results are displayed.

Please visit a PSIPRED web site for further information.

PSIPRED : http://bioinf.cs.ucl.ac.uk/psipred/



5.3.7-1 AISTViewer – PSIPRED Result



5.3.7-2 PSIPRED Result – TextView

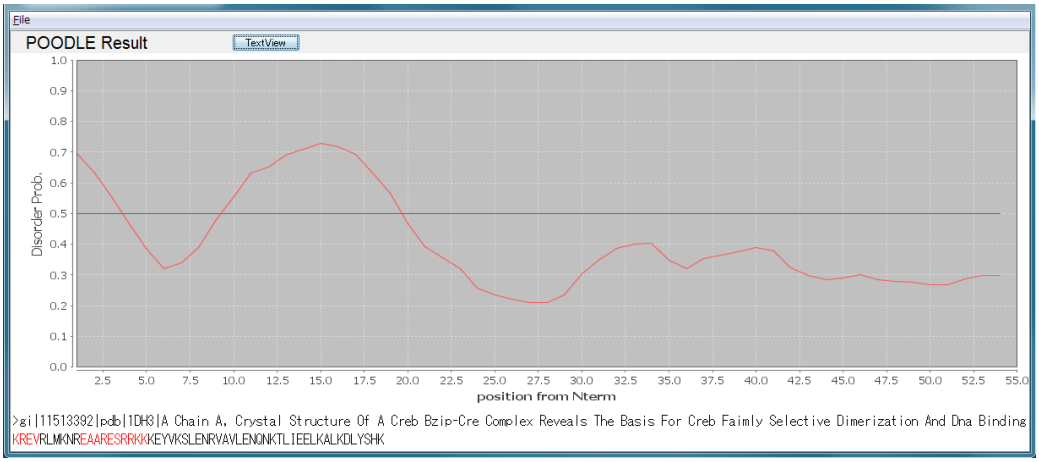
5.3.8 Poodle_AIST → AISTViewer

Results of Poodle_AIST can be displayed using AISTViewer node.

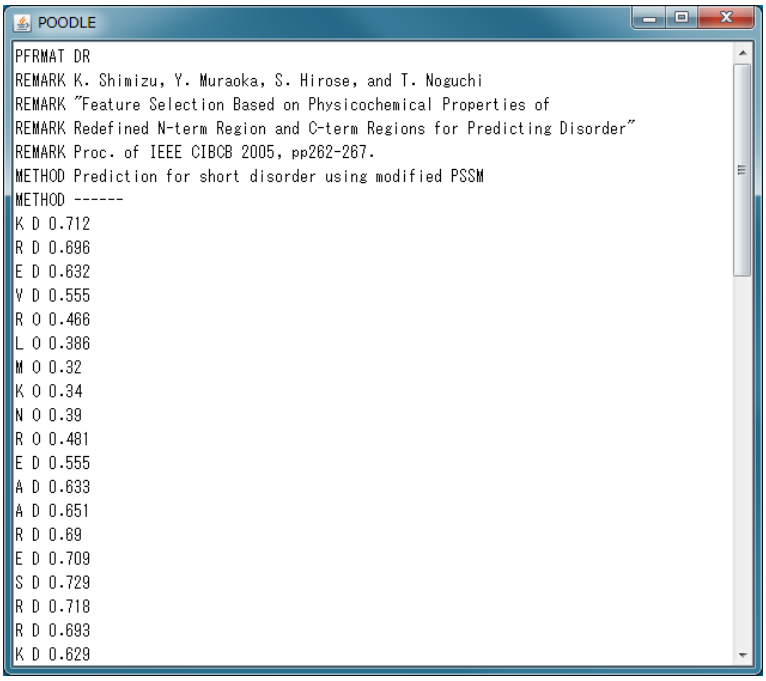
Disorder probabilities are shown in form of a line plot. Vertical and horizontal axes are disorder probabilities and amino acid residue numbers, respectively. Regions which have 0.5 and over disorder probabilities (over a blue vertical line) regard as disordered.

An amino acid sequence is also displayed below the line plot graph. Red characters correspond to disorder regions.

By clicking a "TextView" button on the top of the screen, a pop up window is opened and text format results are displayed.



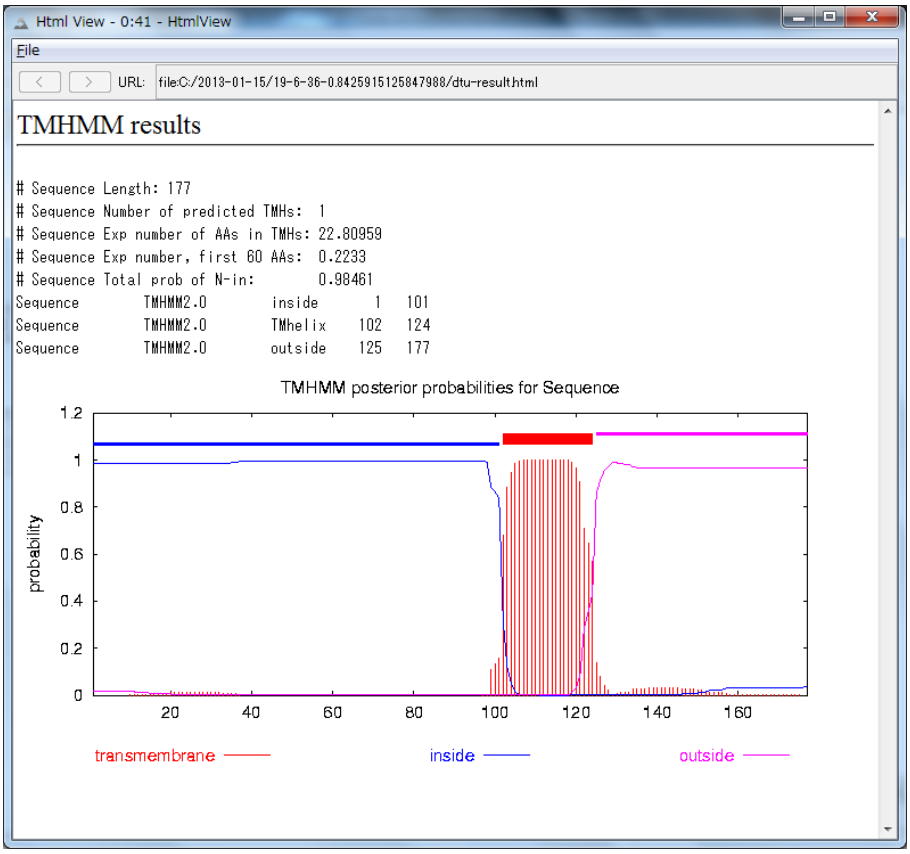
5.3.8-1 AISTViewer – POODLE Result



5.3.8-2 POODLE Result - TextView

5.3.9 TmHmm_DTU → HtmlView

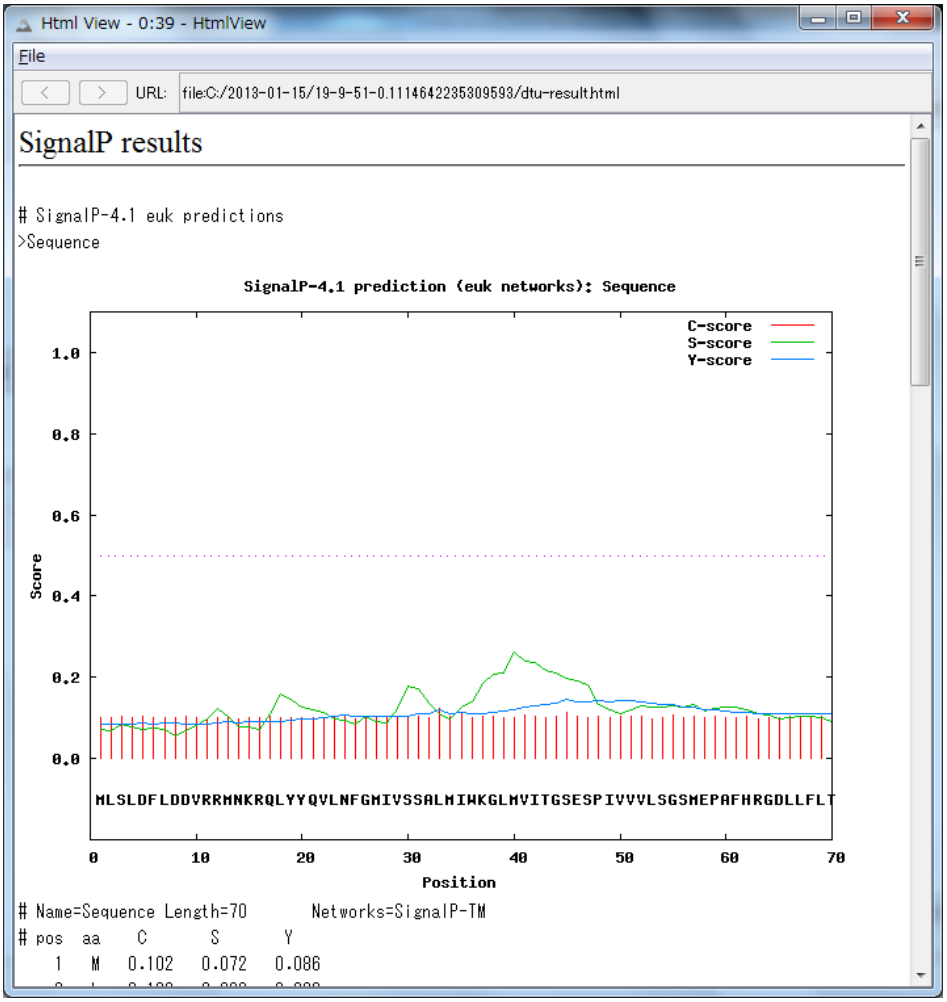
Results of TmHmm_DTU can be displayed using HtmlView node.



5.3.9-1 TmHmm_DTU execution results (HtmlView)

5.3.10 SignalP_DTU → HtmlView

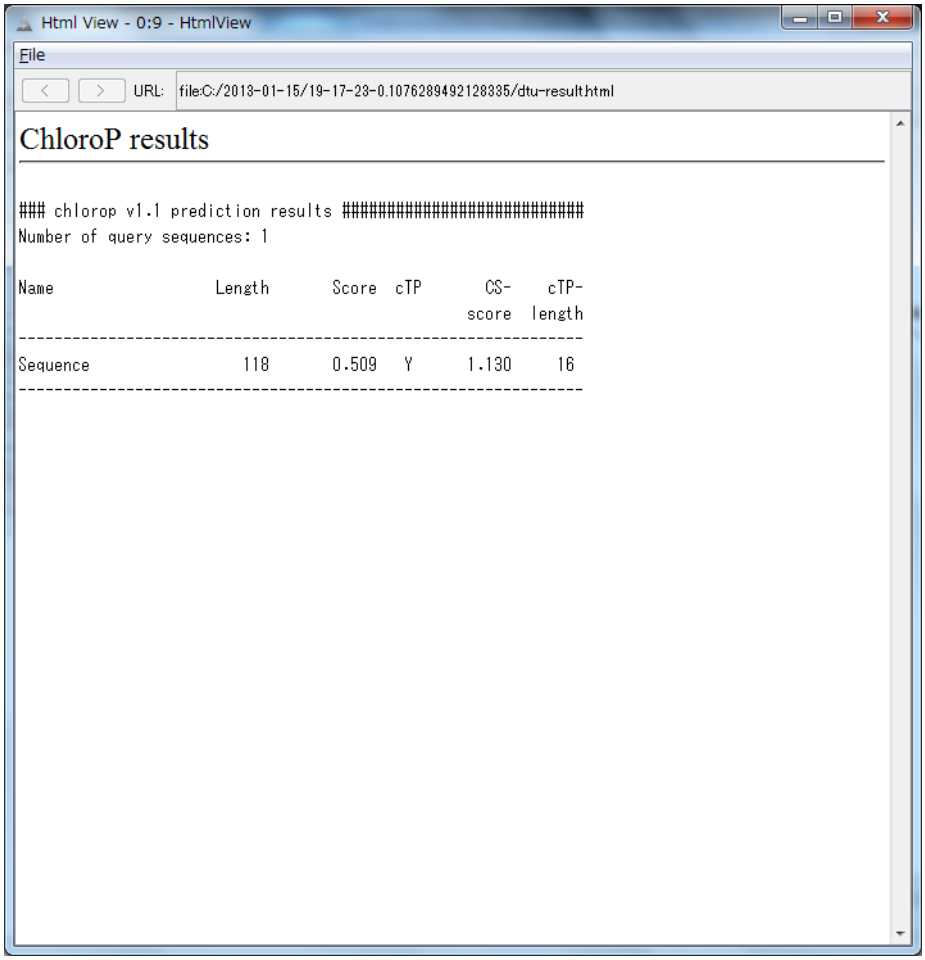
Results of SignalP_DTU can be displayed using HtmlView node.



5.3.10-1 SignalP_DTU execution results (HtmlView)

5.3.11 ChloroP_DTU → HtmlView

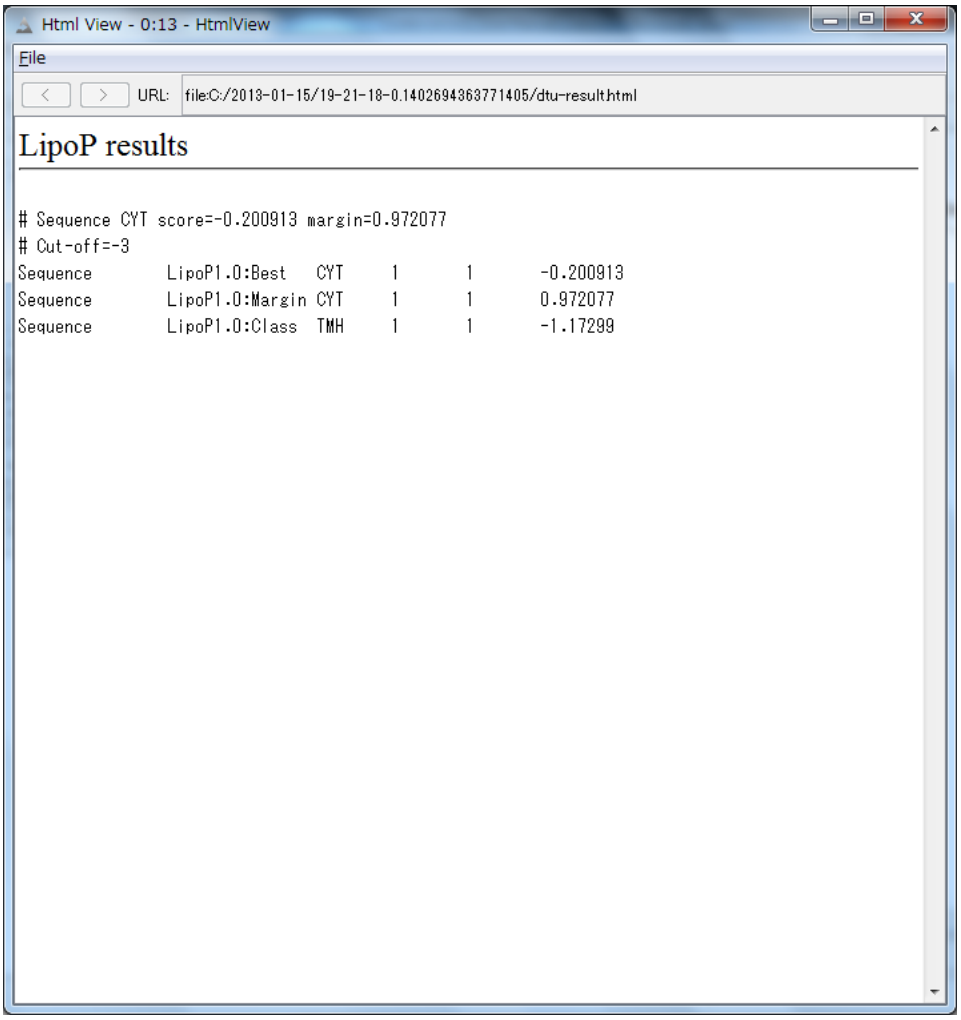
Results of ChloroP_DTU can be displayed using HtmlView node.



5.3.11-1 ChloroP_DTU execution results (HtmlView)

5.3.12 LipoP_DTU → HtmlView

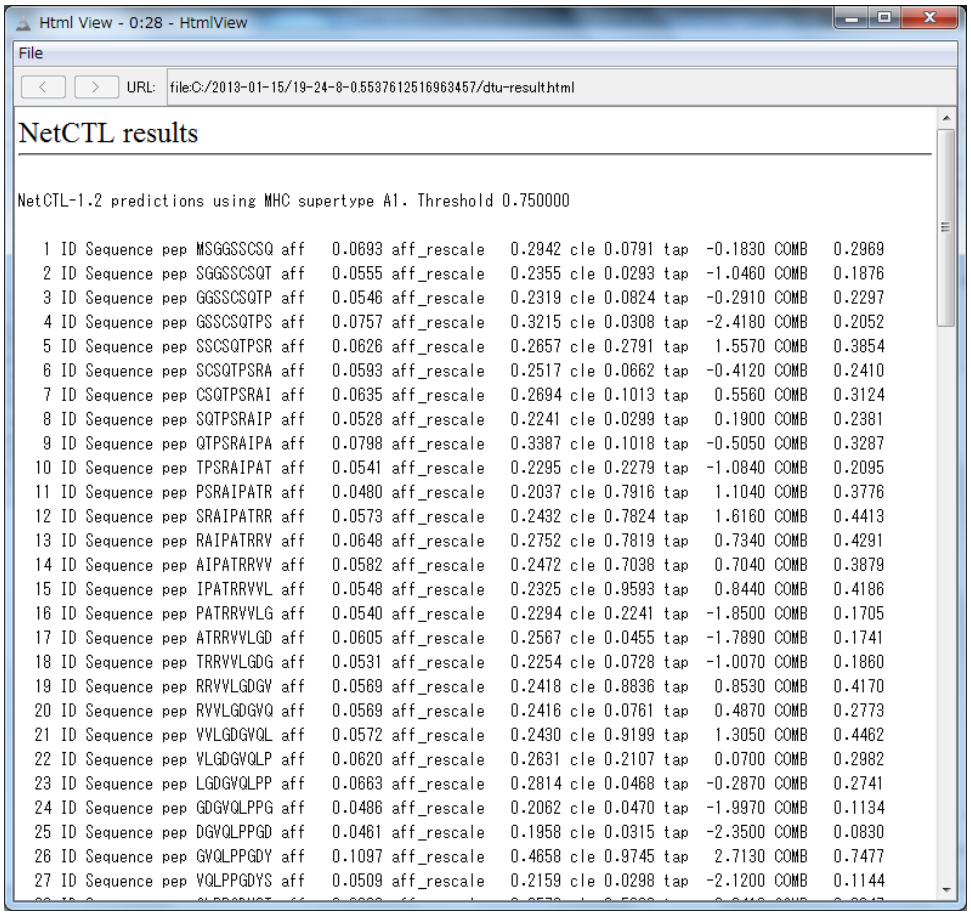
Results of LipoP_DTU can be displayed using HtmlView node.



5.3.12-1 LipoP_DTU execution results (HtmlView)

5.3.13 NetCTL_DTU → HtmlView

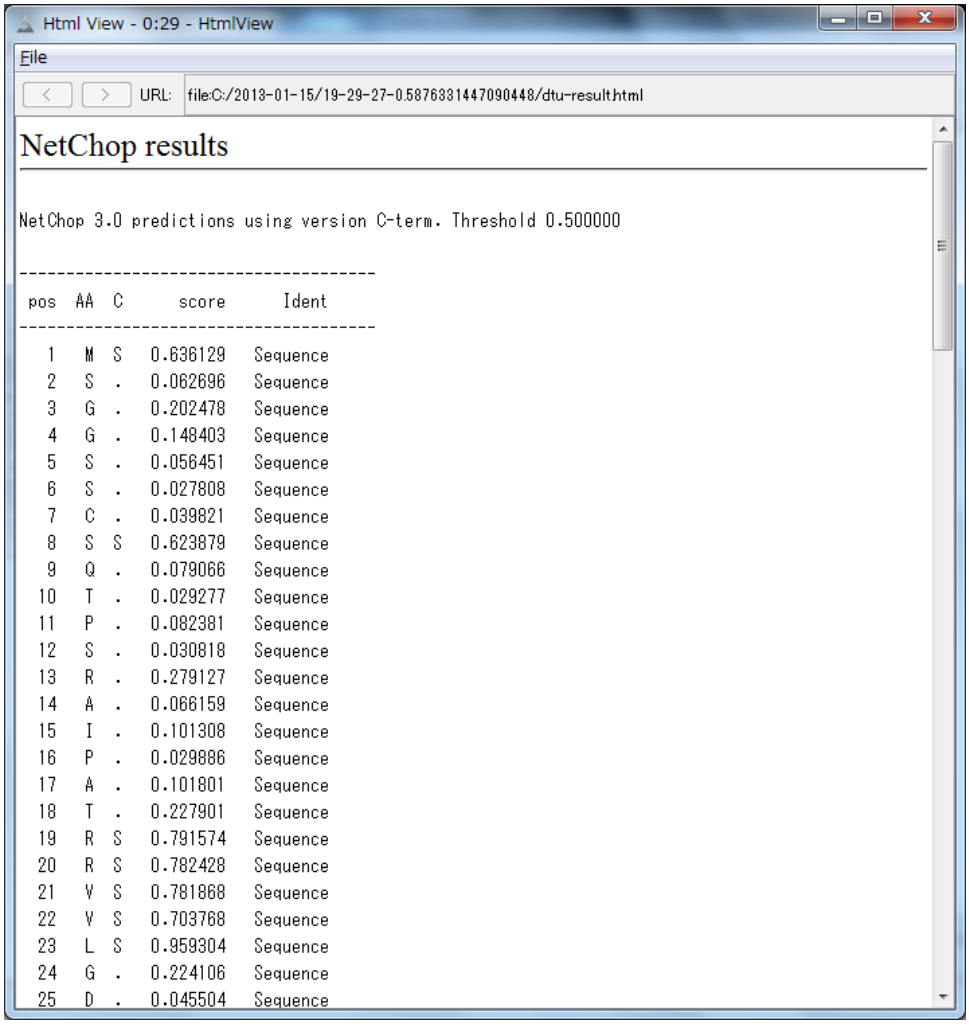
Results of NetCTL_DTU can be displayed using HtmlView node.



5.3.13-1 NetCTL_DTU execution results (HtmlView)

5.3.14 NetChop_DTU → HtmlView

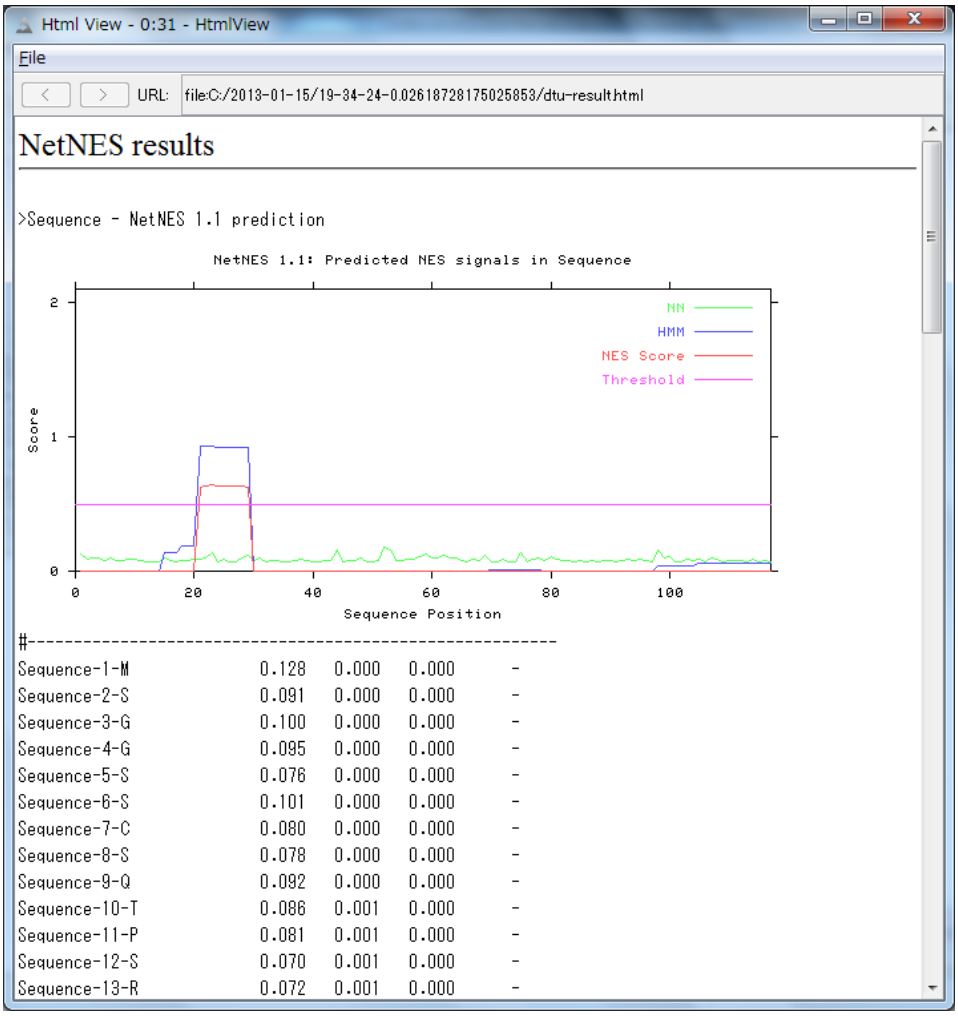
Results of NetChop_DTU can be displayed using HtmlView node.



5.3.14-1 NetChop_DTU execution results (HtmlView)

5.3.15 NetNES_DTU → HtmlView

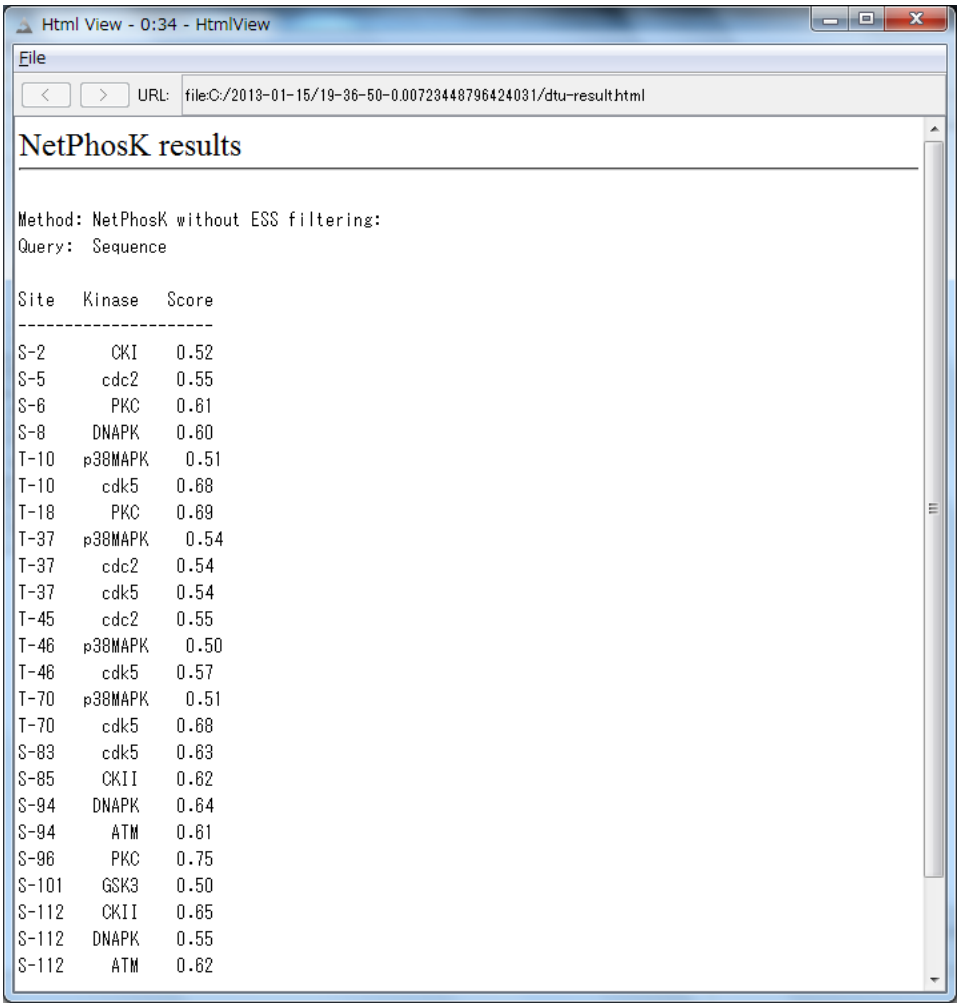
Results of NetNES_DTU can be displayed using HtmlView node.



5.3.15-1 NetNES_DTU execution results (HtmlView)

5.3.16 NetPhosK_DTU → HtmlView

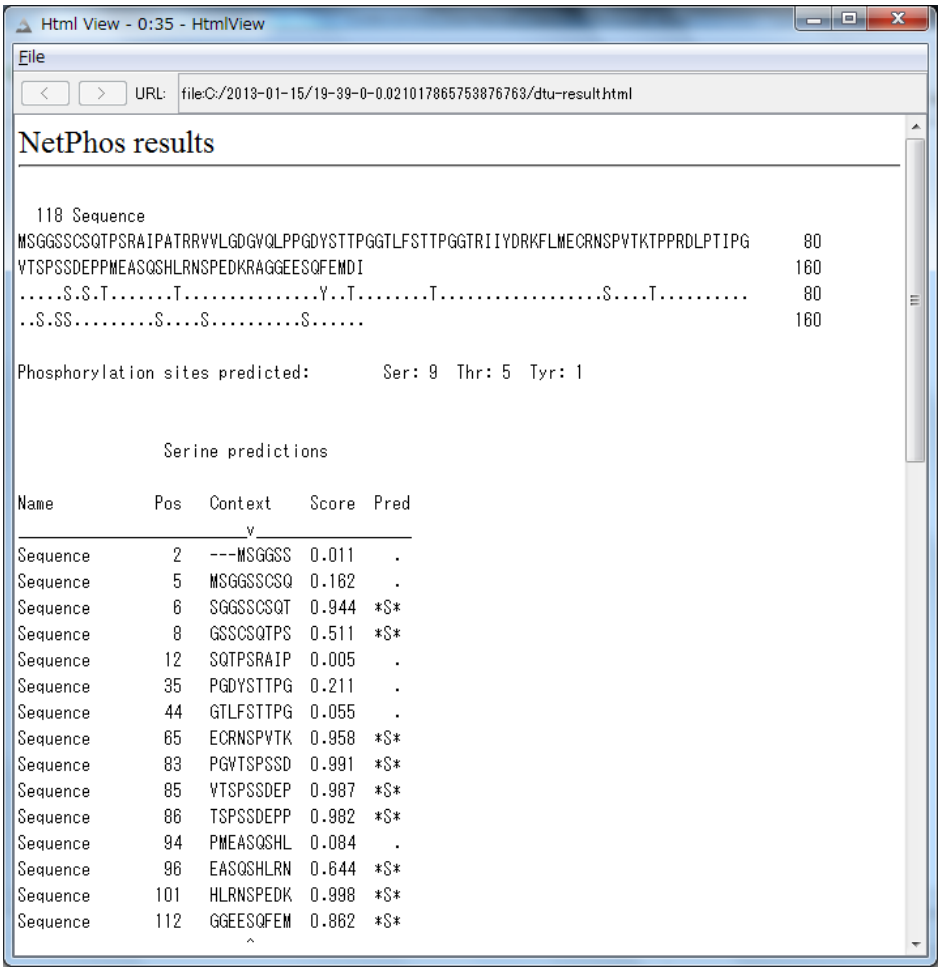
Results of NetPhosK_DTU can be displayed using HtmlView node.



5.3.16-1 NetPhosK_DTU execution results (HtmlView)

5.3.17 NetPhos_DTU → HtmlView

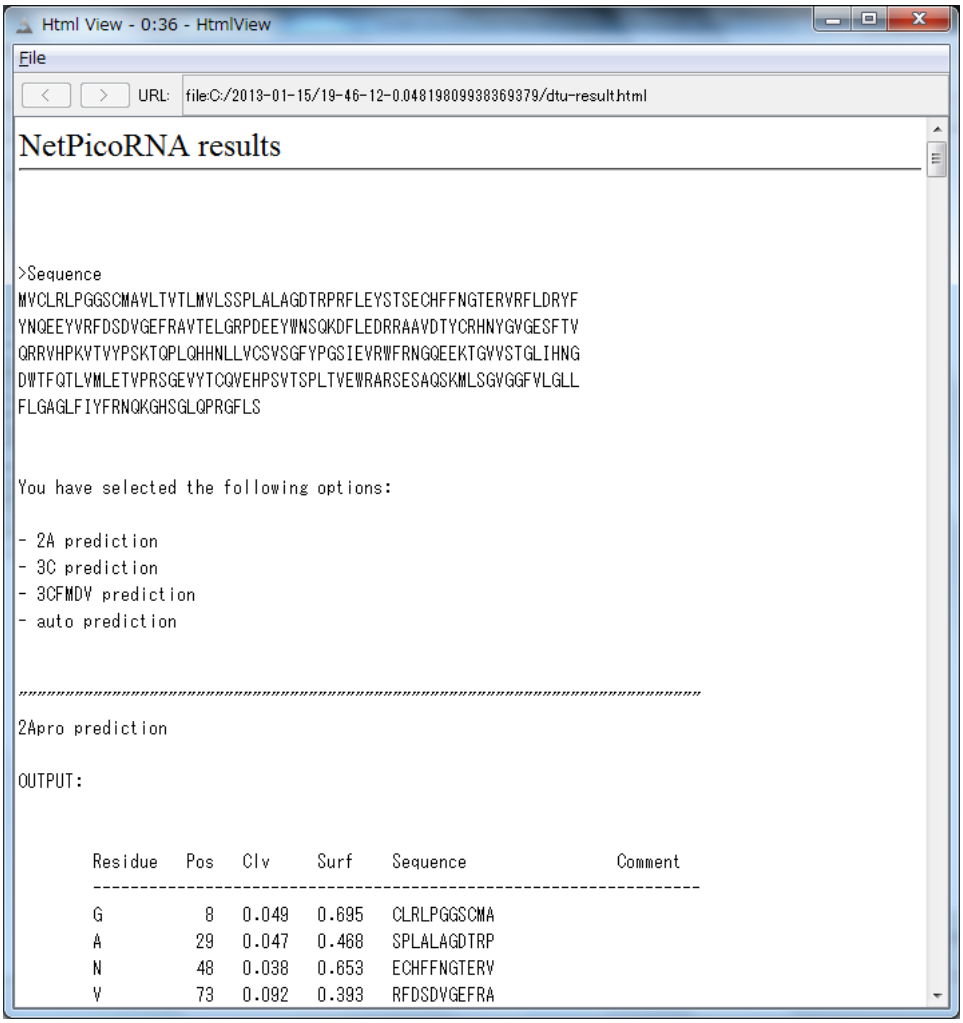
Results of NetPhos_DTU can be displayed using HtmlView node.



5.3.17-1 NetPhos_DTU execution results (HtmlView)

5.3.18 NetPicoRNA_DTU → HtmlView

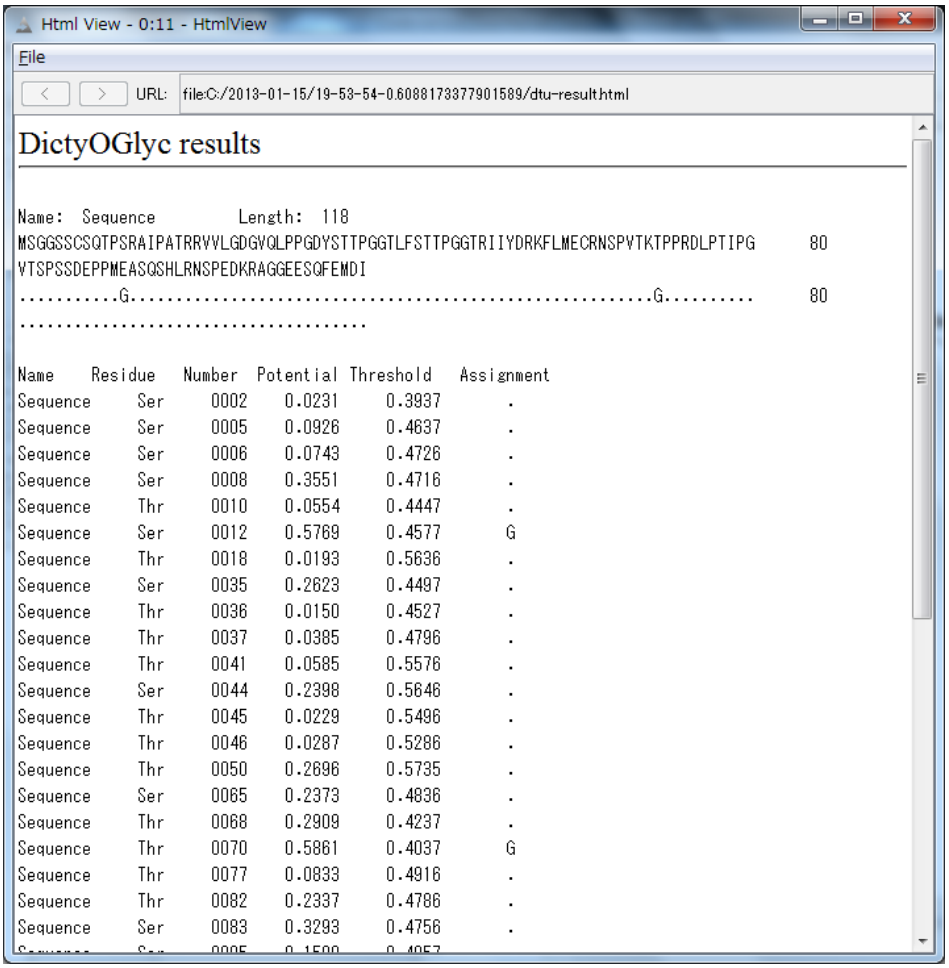
Results of NetPicoRNA_DTU can be displayed using HtmlView node.



5.3.18-1 NetPicoRNA_DTU execution results (HtmlView)

5.3.19 DictyOGlyc_DTU → HtmlView

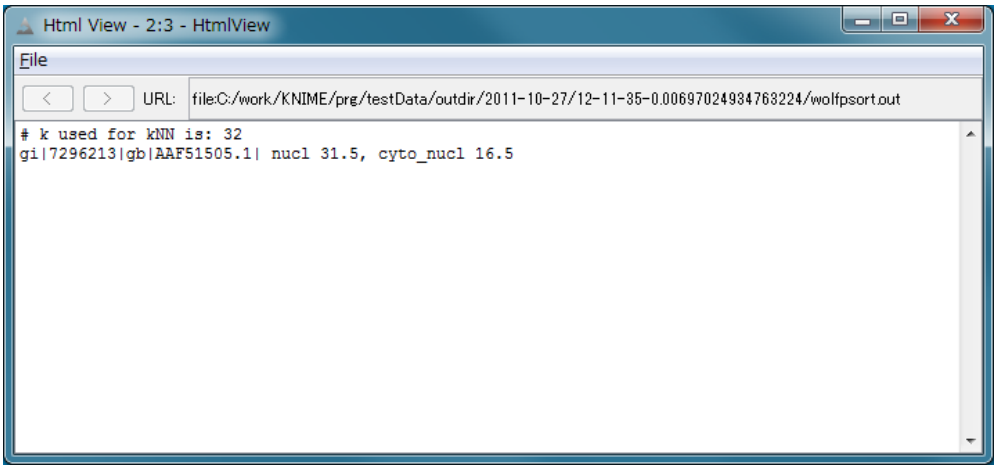
Results of DictyOGlyc_DTU can be displayed using HtmlView node.



5.3.19-1 DictyOGlyc_DTU execution results (HtmlView)

5.3.20 WolfPsort_AIST → HtmlView

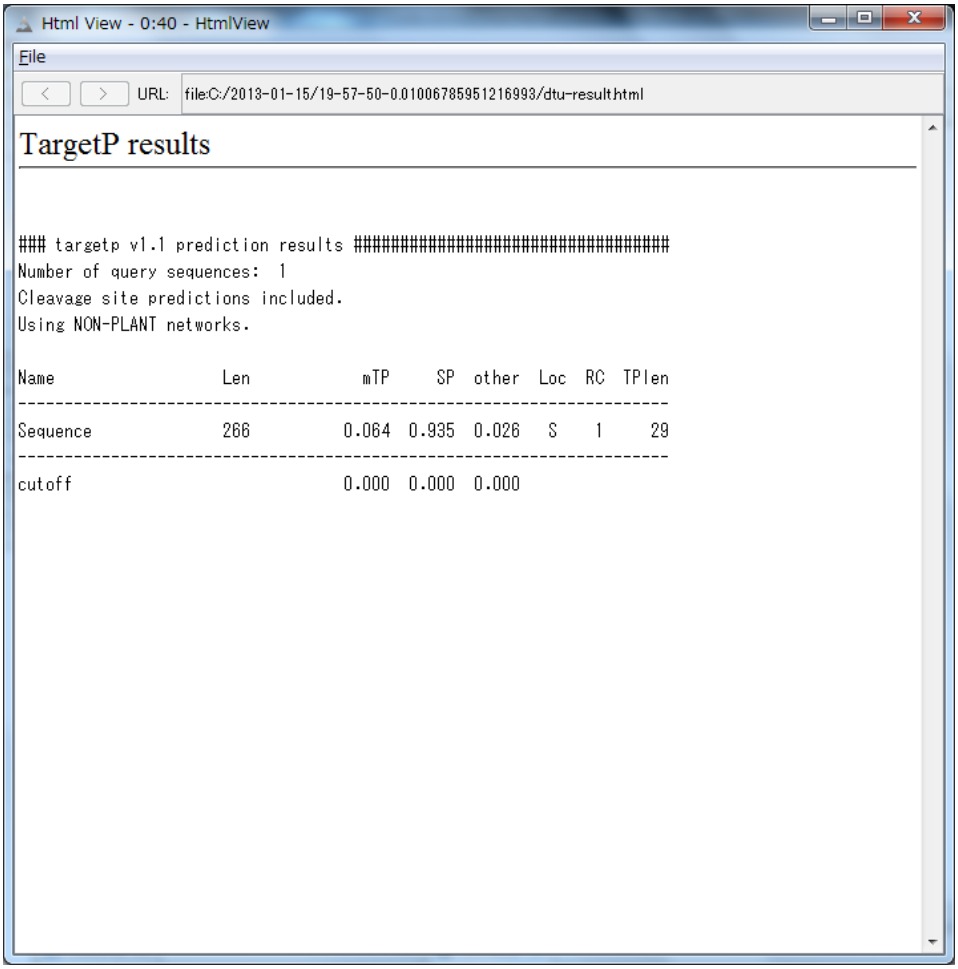
Results of WolfPsort_AIST can be displayed using HtmlView node.



5.3.20-1 WolfPsort_AIST execution results (HtmlView)

5.3.21 TargetP_DTU → HtmlView

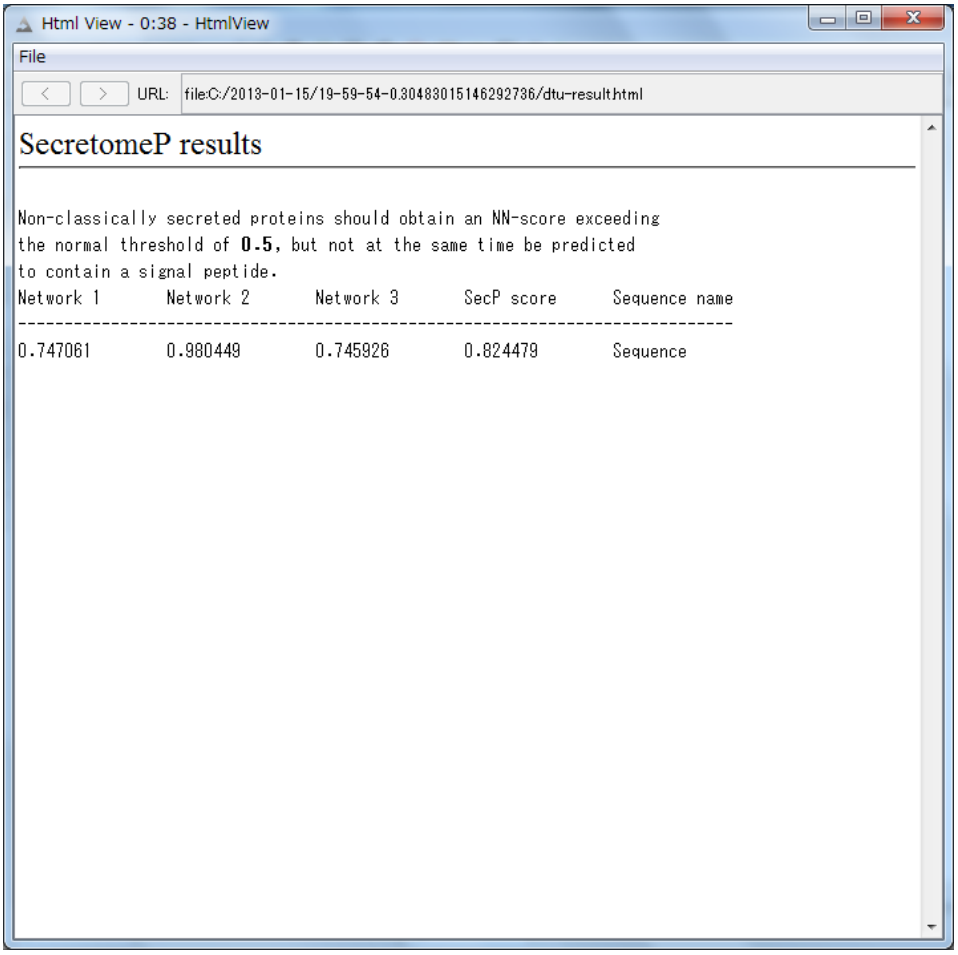
Results of TargetP_DTU can be displayed using HtmlView node.



5.3.21-1 TargetP_DTU execution results (HtmlView)

5.3.22 SecretomeP_DTU → HtmlView

Results of SecretomeP_DTU can be displayed using HtmlView node.



5.3.22-1 SecretomeP_DTU execution results (HtmlView)

5.3.23 DisoPred_AIST → AISTViewer

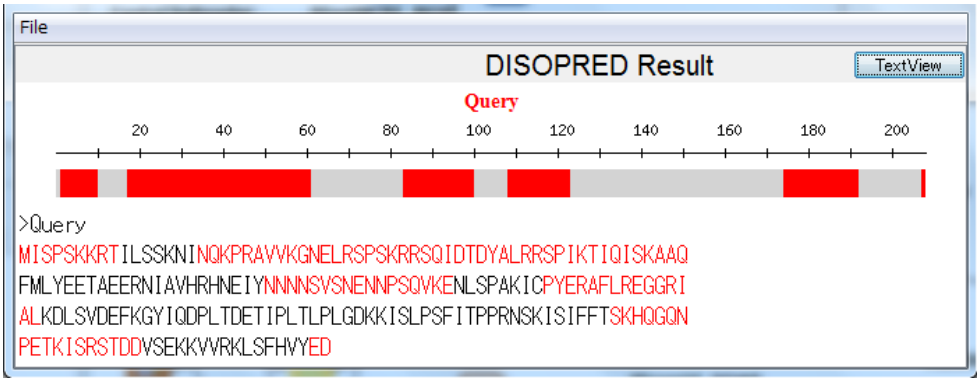
Results of DisoPred_AIST can be displayed using AISTViewer node.

Disorder regions are shown in red boxes and characters.

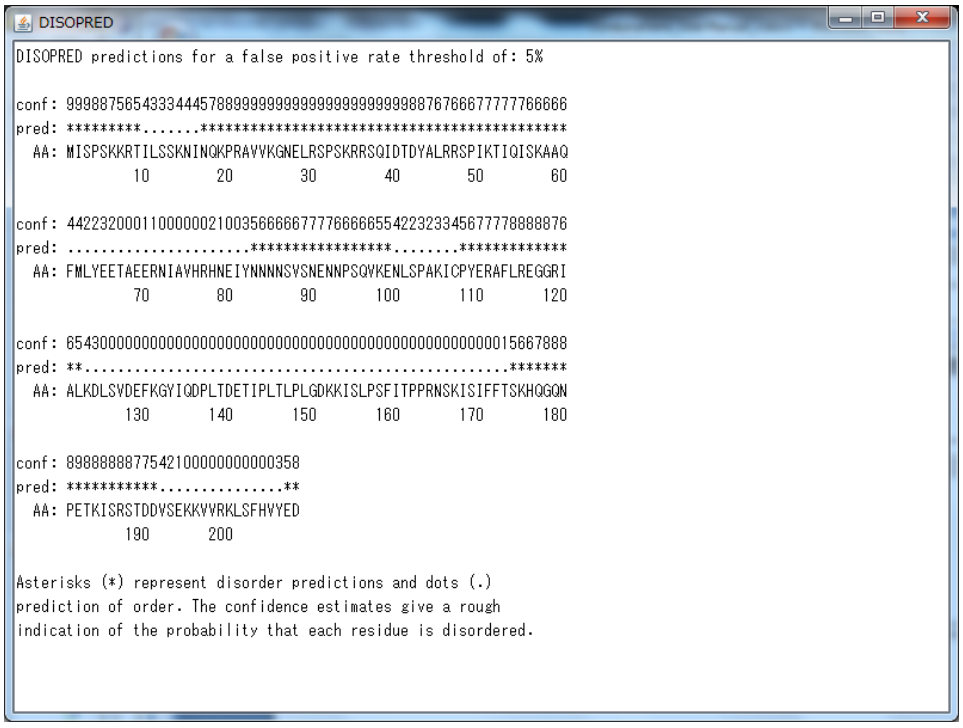
By clicking a “TextView” button on the top-right corner of the screen, a pop up window is opened and text format results are displayed.

Please visit a DISOPRED web site for further information.

DISOPRED : http://bioinf.cs.ucl.ac.uk/index.php?id=806



5.3.23-1 AISTViewer – DISOPRED Result



5.3.23-2 DISOPRED Result – TextView

5.3.24 Memsat_AIST → AISTViewer

Results of Memsat_AIST can be displayed using AISTViewer node.

Each color used in an amino acid sequence are as follows:

Transmembrane(TM) regions: red

Outside helix cap of TM: magenta

Inside helix cap of TM: orange

Outside sequence: grey

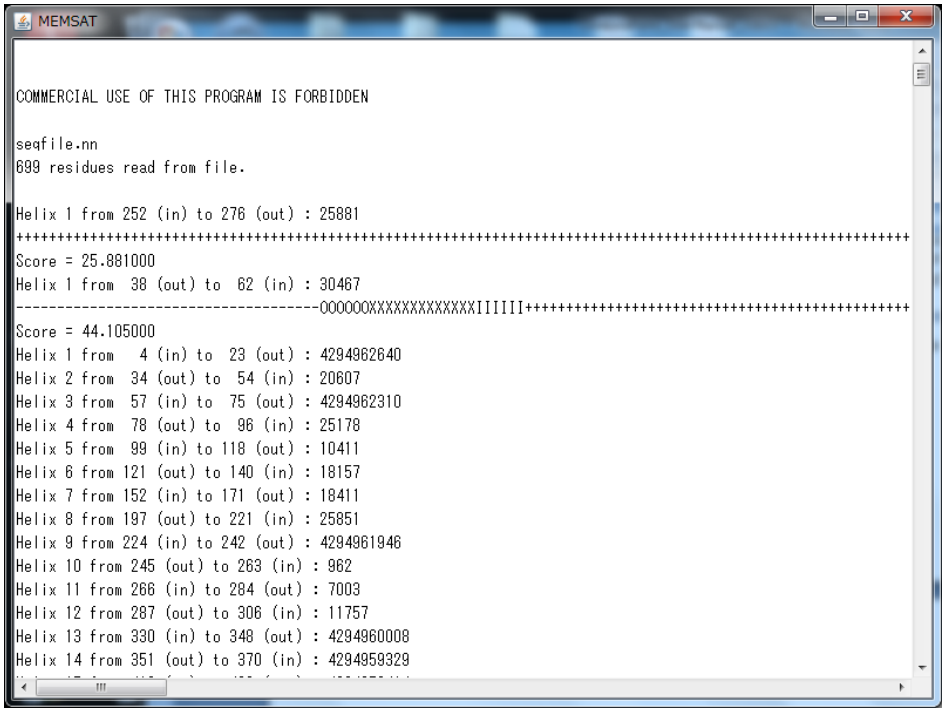
Inside sequence: black

By clicking a "TextView" button on the top of the screen, a pop up window is opened and text format results are displayed.

Please visit a MEMSAT web site for further information.

MEMSAT : http://bioinf.cs.ucl.ac.uk/software_downloads/memsat/

5.3.24-1 AISTViewer – Memsat Result



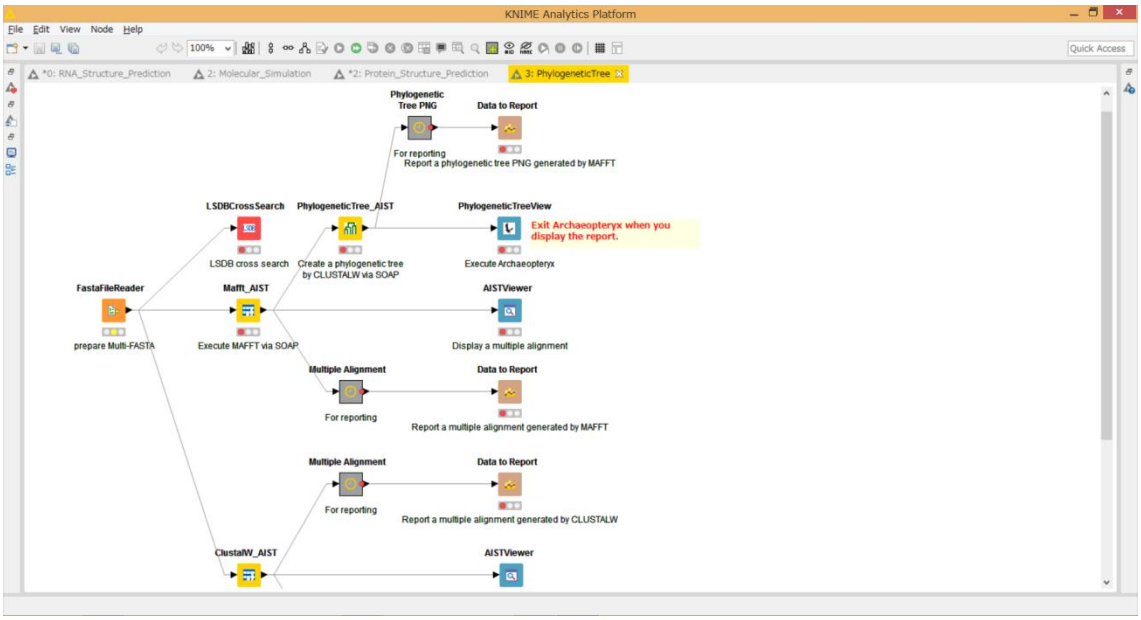
5.3.24-2 Memsat Result – TextView

6 Phylogenetic Tree

Phylogenetic Tree is a workflow that executes multiple alignment and makes phylogenetic tree. Each prediction node is executed via SOAP.

Please visit below web sites for further information.

Software	Web site
Mafft	http://mafft.cbrc.jp/alignment/software/
ClustalW	http://www.clustal.org/
Archaeopteryx	https://sites.google.com/site/cmzmasek/home/software/archaeopteryx
LSDB	http://biosciencedbc.jp/dbsearch/



6-1 Phylogenetic Tree

6.1 Preparation

This workflow requires nucleic acid or protein multi-FASTA format file.
This multi-FASTA file must contain at least four sequences.

File type
Multi-FASTA format file

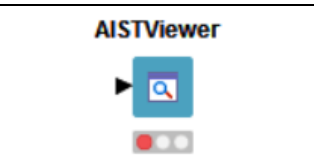
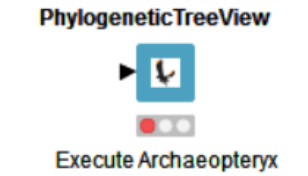
6.2 Nodes

Phylogenetic Tree Active workflow has 7 types of KNIME node. The nodes are shown in the below table. Please check each KNIME node description.

6.2.1 Node list

6.2.1-1 Phylogenetic Tree Active Workflow node list

No	Name	Icon	Set	Description
1	FastaFileReader		Y	Read Multi-FASTA file.
2	Mafft_AIST		Y	Execute Mafft.
3	ClustalW_AIST		Y	Execute ClustalW.
4	PhylogeneticTree_AIST		Y	Execute Phylogenetic Tree.

5	AISTViewer			Display prediction results.
6	PhylogeneticTreeView			Display Phylogenetic Tree.
7	LSDBCrossSearch		-	Execute LSDB cross search.

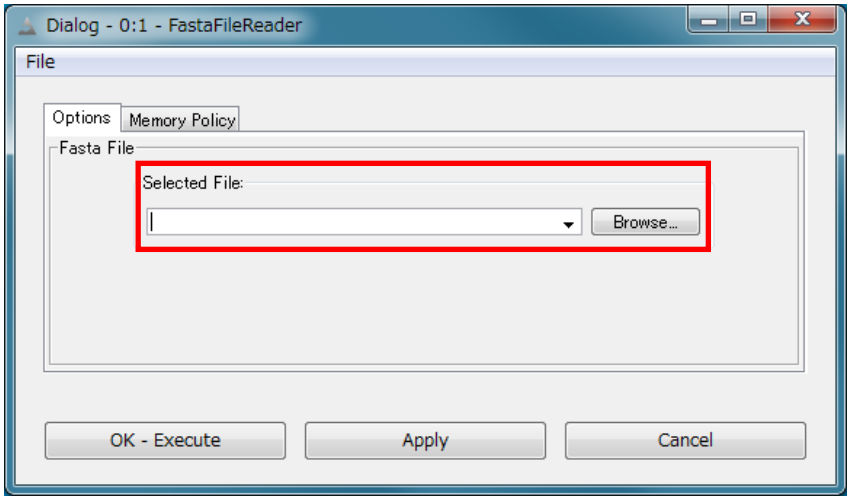
6.2.2 Node setting

6.2.2.1 FastaFileReader

Set multi-FASTA sequence file.

1) Right-clicking on the FastaFileReader node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



6.2.2.1-1 FastaFileReader : Configure…

・Options → Fasta File → SelectedFile:

Specify multi-FASTA sequence file in a text box or “Browse” (red open rectangular)

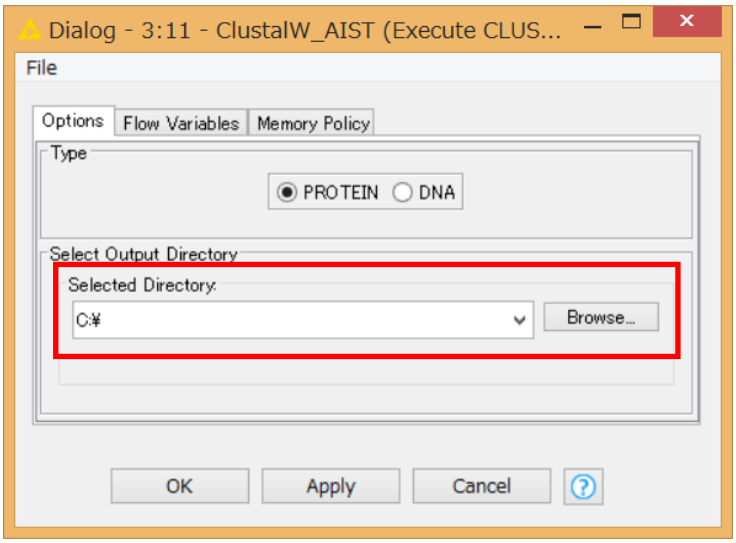
Click “OK” button after setting the conditions.

6.2.2.2 ClustalW_AIST

Set an output directory.

1) Right-clicking on the ClustalW_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



6.2.2.2-1 ClustalW_AIST : Configure…

・Options → Type

Choose either PROTEIN or DNA based on query sequence type.

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectangular)

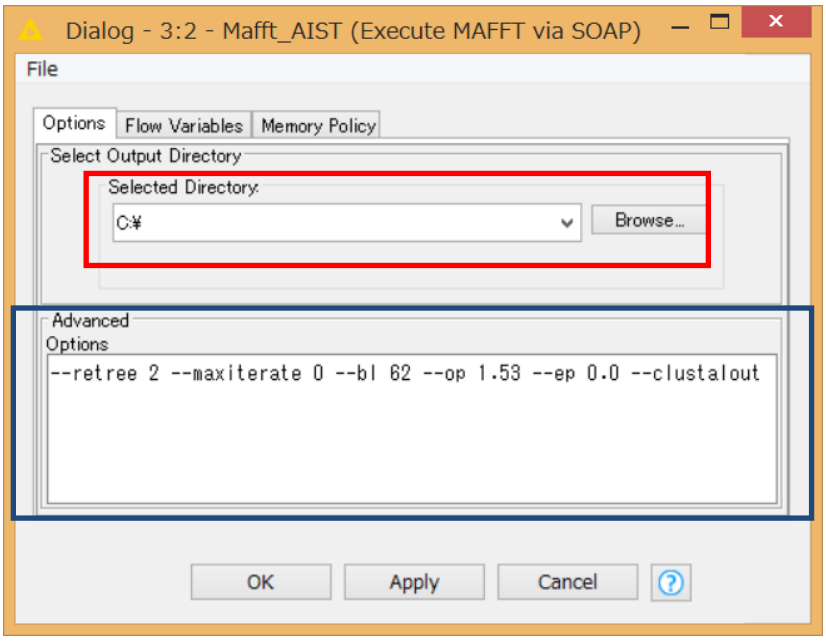
Click “OK” button after setting the conditions.

6.2.2.3 Mafft_AIST

Set an output directory and execution options.

1) Right-clicking on the Mafft_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



6.2.2.3-1 Mafft_AIST : Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectangular)

・**Options → Advanced → Options**

Specify Mafft execution parameters (blue open rectangular).

User can specify each parameter by separating a single space. Mafft options are as follows:

--op #	: Gap opening penalty, default: 1.53
--ep #	: Offset (works like gap extension penalty), default: 0.0
--maxiterate #	: Maximum number of iterative refinement, default: 0
--clustalout	: Output: clustal format, default: fasta
--reorder	: Outorder: aligned, default: input order
--quiet	: Do not report progress

Default option setting is shown in below table.

--retree 2 --maxiterate 0 --bl 62 --op 1.53 --ep 0.0 --clustalout

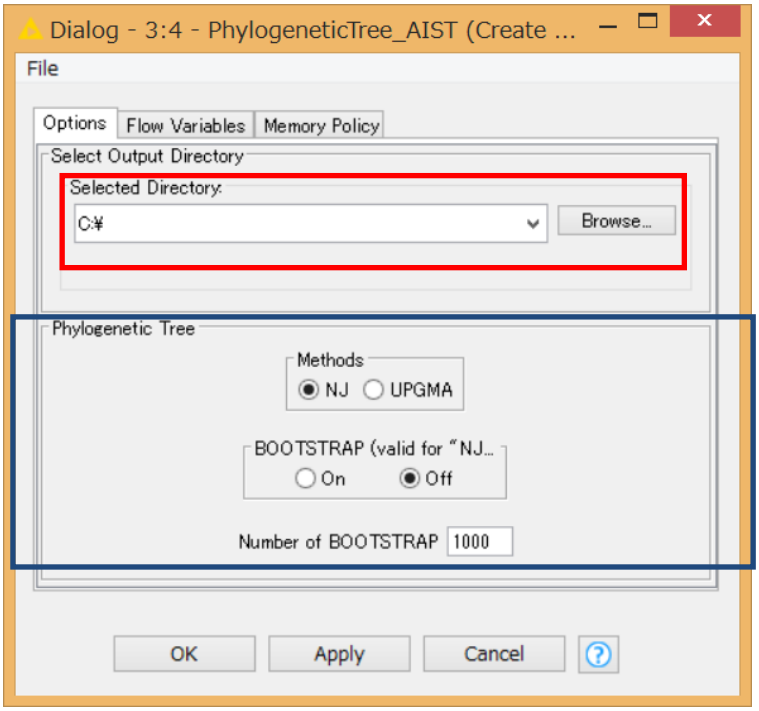
Click “OK” button after setting the conditions.

6.2.2.4 PhylogeneticTree_AIST

Set an output directory and execution parameters.

1) Right-clicking on the PhylogeneticTree_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



6.2.2.4-1 PhylogeneticTree_AIST : Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (red open rectangular)

・Options → Phylogenetic Tree

Specify types of phylogenetic tree (blue open rectangular).

・Methods :

Choose either NJ or UPGMA.

・BOOTSTRAP (valid for “NJ”) :

Choose either On or Off.

・Number of BOOTSTRAP :

Specify number of BOOTSTRAP (default: 1000)

Click “OK” button after setting the conditions.

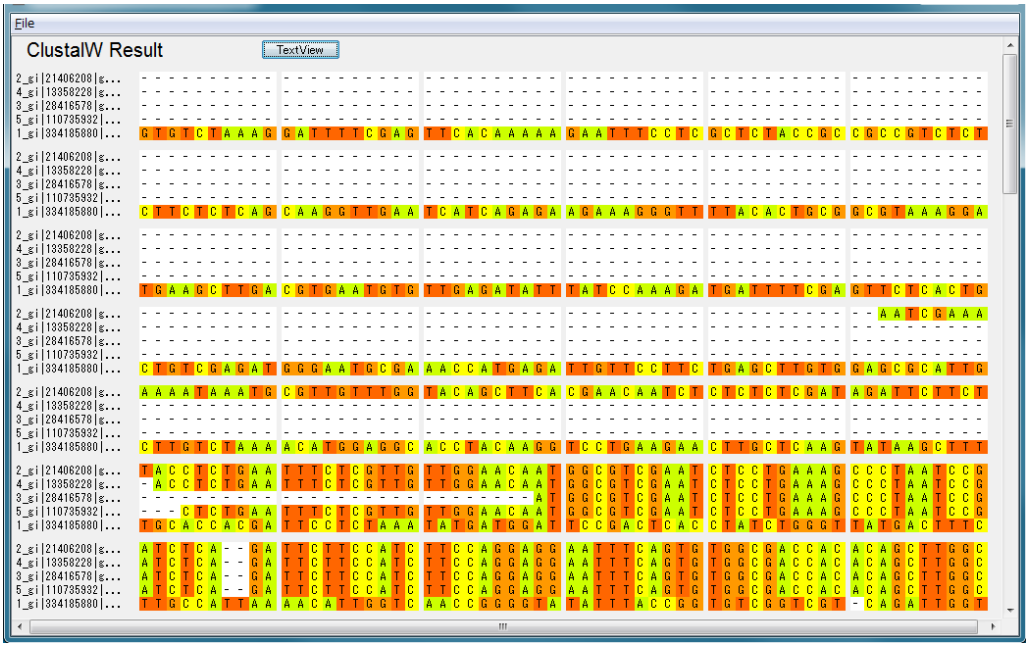
6.3 Results

6.3.1 ClustalW_AIST → AISTViewer

Results of ClustalW_AIST can be displayed using AISTViewer node.

Each sequence name and its multiple alignment result is displayed on the left and right side of the screen, respectively.

By clicking a “TextView” button on the top of the screen, a pop up window is opened and text format results are displayed.



6.3.1-1 AISTViewer – ClustalW Result

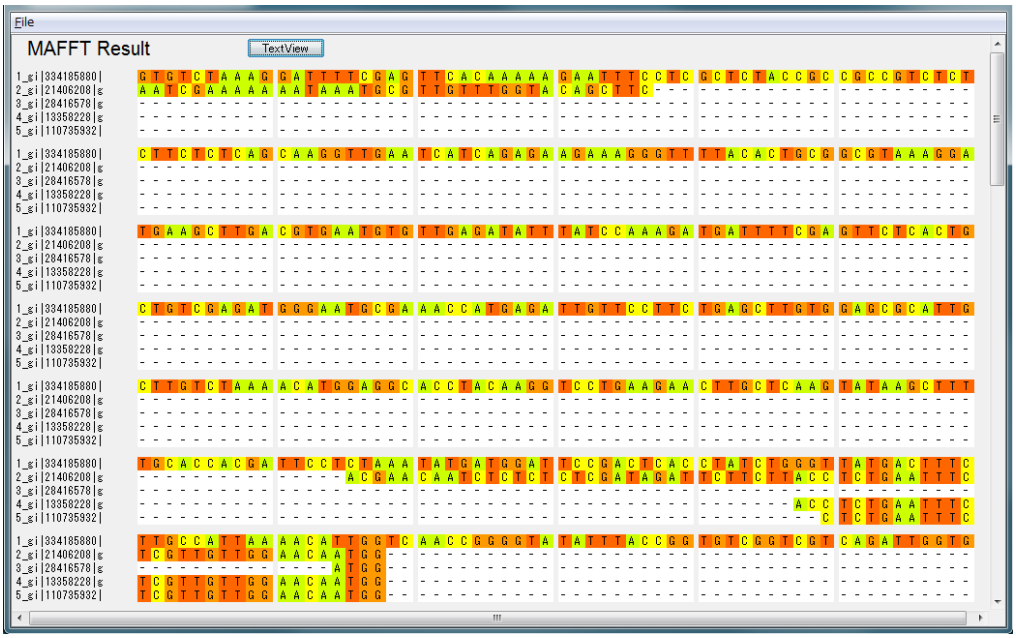
6.3.1-2 ClustalW Result – TextView

6.3.2 Mafft_AIST → AISTViewer

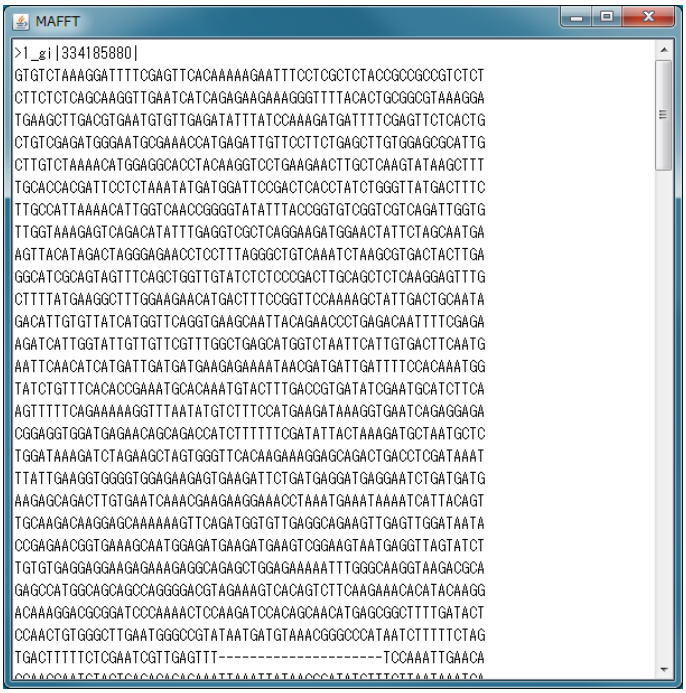
Results of Mafft_AIST can be displayed using AISTViewer node.

Each sequence name and its multiple alignment result is displayed on the left and right side of the screen, respectively.

By clicking a “TextView” button on the top of the screen, a pop up window is opened and text format results are displayed.



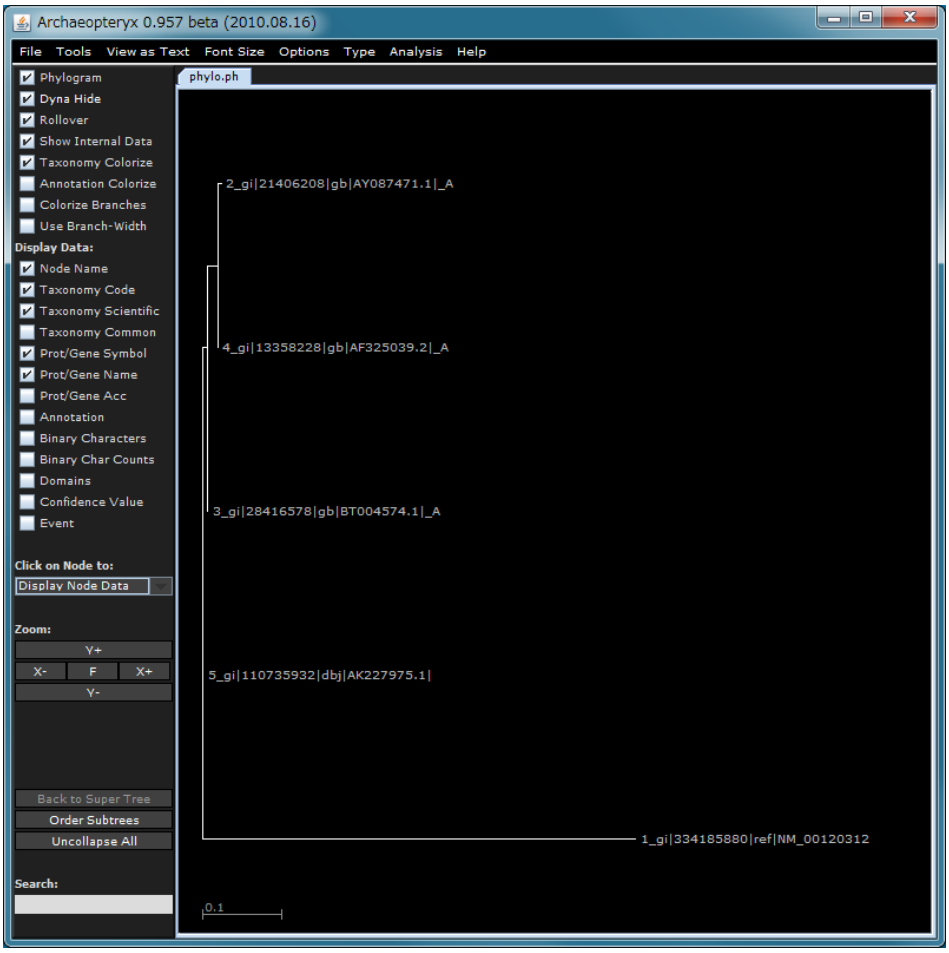
6.3.2-1 AISTViewer – MAFFT Result



6.3.2-2 MAFFT Result – TextView

6.3.3 PhylogeneticTreeView

Results of PhylogeneticTreeView can be displayed using PhylogeneticTreeView node. This node starts Archaeopteryx viewer.



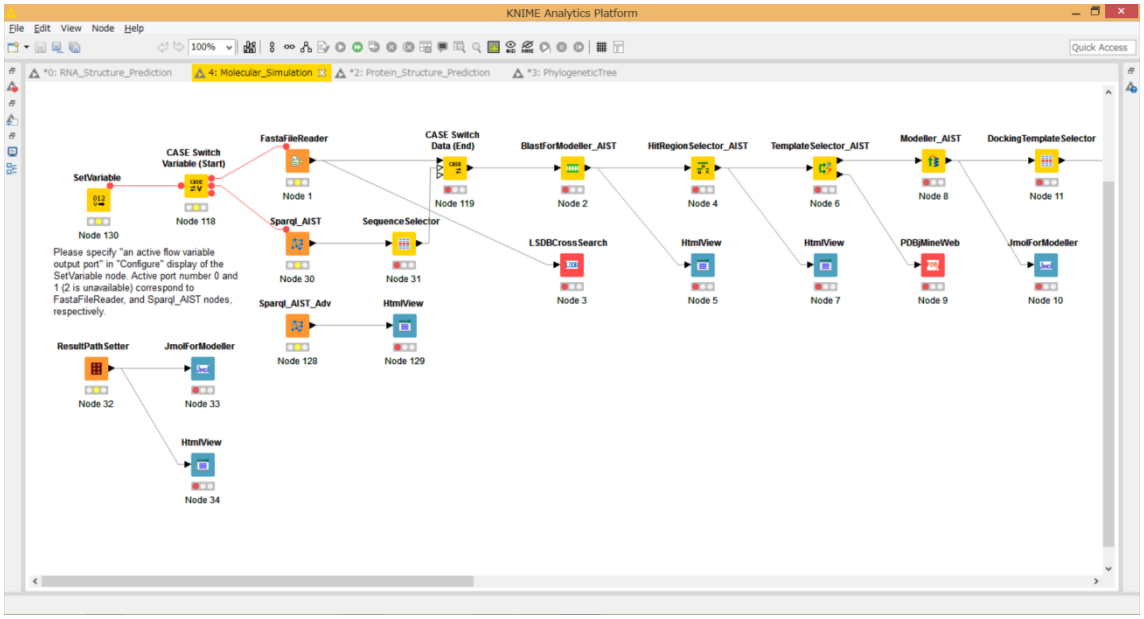
6.3.3-1 PhylogeneticTreeView

7 Molecular Simulation

Molecular Simulation is a workflow that executes molecular simulations such as protein modelling, ligand docking, molecular mechanics calculation, molecular dynamics calculation and so on.

This workflow procedure is as follows:

1) Protein modelling is executed for a query protein sequence by MODELLER.
2) Ligand docking calculation is executed for chemical ligands and the modelled structure by AutoDock program(http://autodock.scripps.edu/). These chemical ligands are registered by Namiki database (http://www.namiki-s.co.jp/) and user can select chemical ligands based on user setting conditions.
3) Energy minimization calculation is executed based on docking calcuration results, and followed by molecular mechanics and molecular dynamics calculation.



7-1 Molecular Simulation Active Workflow

7.1 Preparation

This workflow requires a protein FASTA format file.

File type
Protein FASTA format file

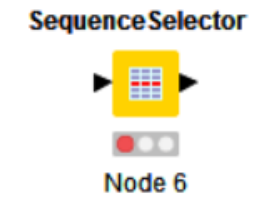
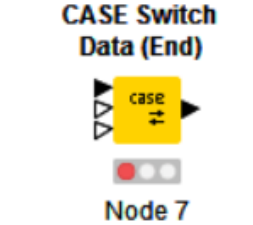
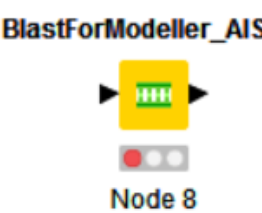
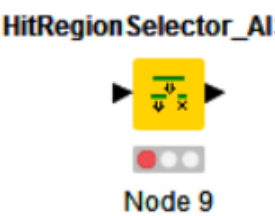
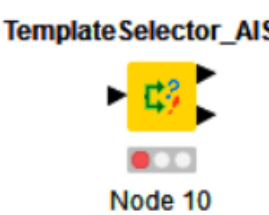
7.2 Nodes

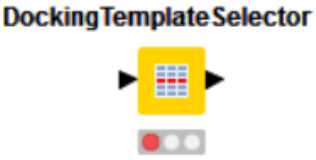
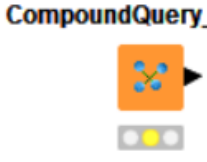
Molecular Simulation Active Workflow has 29 types of KNIME node. The nodes are shown in the below table. Please check each KNIME node description.

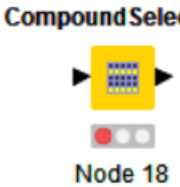
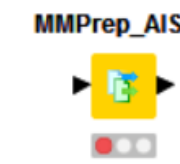
7.2.1 Node list

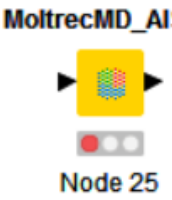
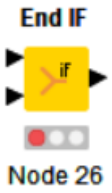
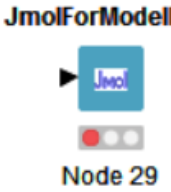
7.2.1-1 Molecular Simulation Active Workflow node list

No	Name	Icon	Set	Description
1	SetVariable		Y	Control CASE Switch Variable. Open Dialog: 0: first outport 1: second outport 2:third outport
2	CASE Switch Variable(Start)			Control CASE Switch using workflow variables.
3	FastaFileReader		Y	Set FASTA file.

4	LSDBCrossSearch			Execute LSDB cross search.
5	Sparql_AIST		Y	Execute SPARQL against AIST endpoints.
6	SequenceSelector			Select a sequence from SPARQL results.
7	CASE Switch Variable (End)			CASE Switch end node.
8	BlastForModeller_AIST		Y	Execute BLAST (PSI-BLAST).
9	HitRegionSelector_AIST		Y	Detect hit regions from BLAST (PSI-BLAST) execution results.
10	TemplateSelector_AIST		Y	Choose a modelling template.

11	Modeller_AIST		Y	Execute MODELLER.
12	DockingTemplateSelector		Y	Select a template structure from modelling results.
13	PdbFileReader		Y	Read PDB file.
14	fpocket2_AIST		Y	Execute fpocket2.
15	PocketSelector			Select a pocket site from fpocket2 results.
16	AutoDock_AIST		Y	Execute AutoDock.
17	CompoundQuery_AIST		Y	Execute chemical ligands search.

18	CompoundSelector			Select chemical compounds from search results.
19	Mol2FileReader		Y	Read Mol2 format file.
20	Ammos_AIST		Y	Execute Ammos.
21	MergeTargetAndLigand			Merge AutoDock results and template information.
22	SiteAndPoseSelector			Select a docking result.
23	MMPrep_AIST		Y	Prepare molecular mechanics calculation.
24	InitMinMM_AIST		Y	Execute energy minimize calculation and molecular mechanics calculation.

25	MoltrecMD_AIST		Y	Execute molecular dynamics calculation.
26	End IF			End IF node.
27	ResultPathSetter		Y	Specify an output directory name stored prediction results.
28	HtmlView			Display prediction results.
29	JmolForModeller			Start Jmol.

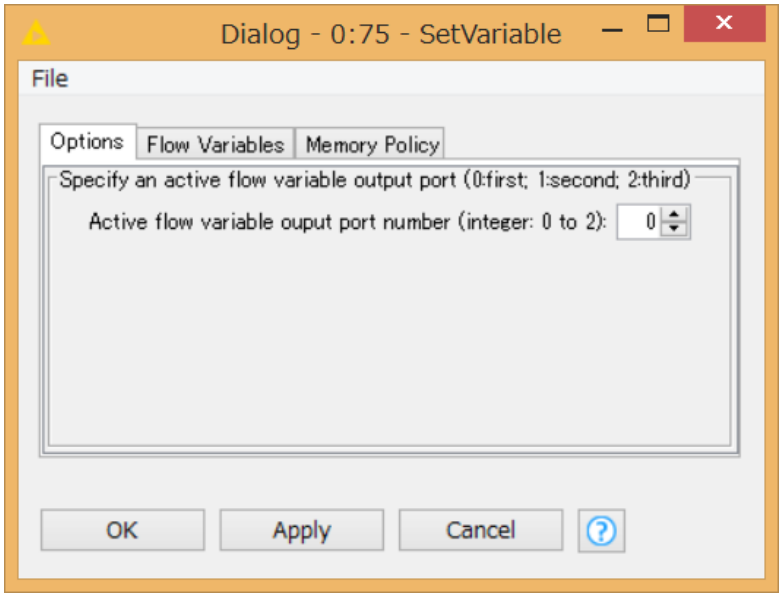
7.2.2 Node setting

7.2.2.1 SetVariable

Set a number to control CASE Switch Variable (Start) outports using workflow variables.

1) Right-clicking on the SetVariable node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.1-1 SetVariable : Configure…

0: CASE Switch Variable(Start) first outport

1: CASE Switch Variable(Start) second outport

2: CASE Switch Variable(Start) third outport

In the workflow, set 0 and 1 for FastaFileReader and Sparql_AIST, respectively.

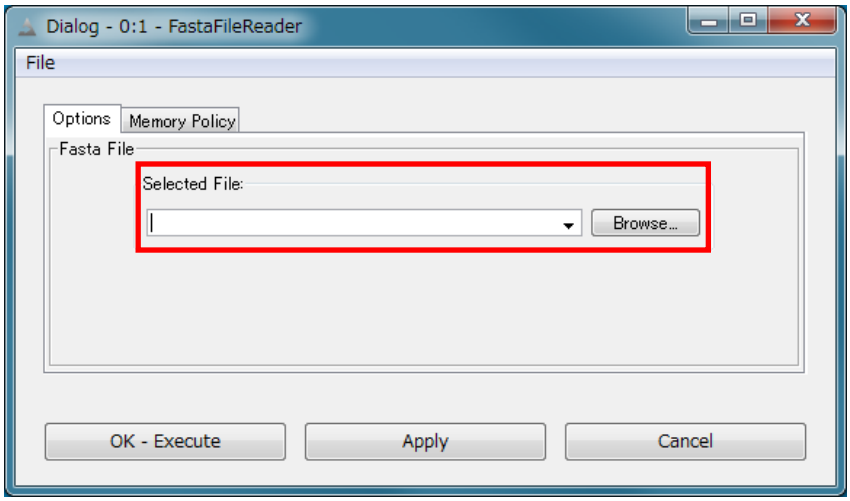
Click “OK” button after setting the conditions.

7.2.2.2 FastaFileReader

Set a protein FASTA sequence file.

1) Right-clicking on the FastaFileReader node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.2-1 FastaFileReader : Configure…

・Options → Fasta File → SelectedFile:

Specify a protein FASTA sequence file in a text box or “Browse” (red open rectangular)

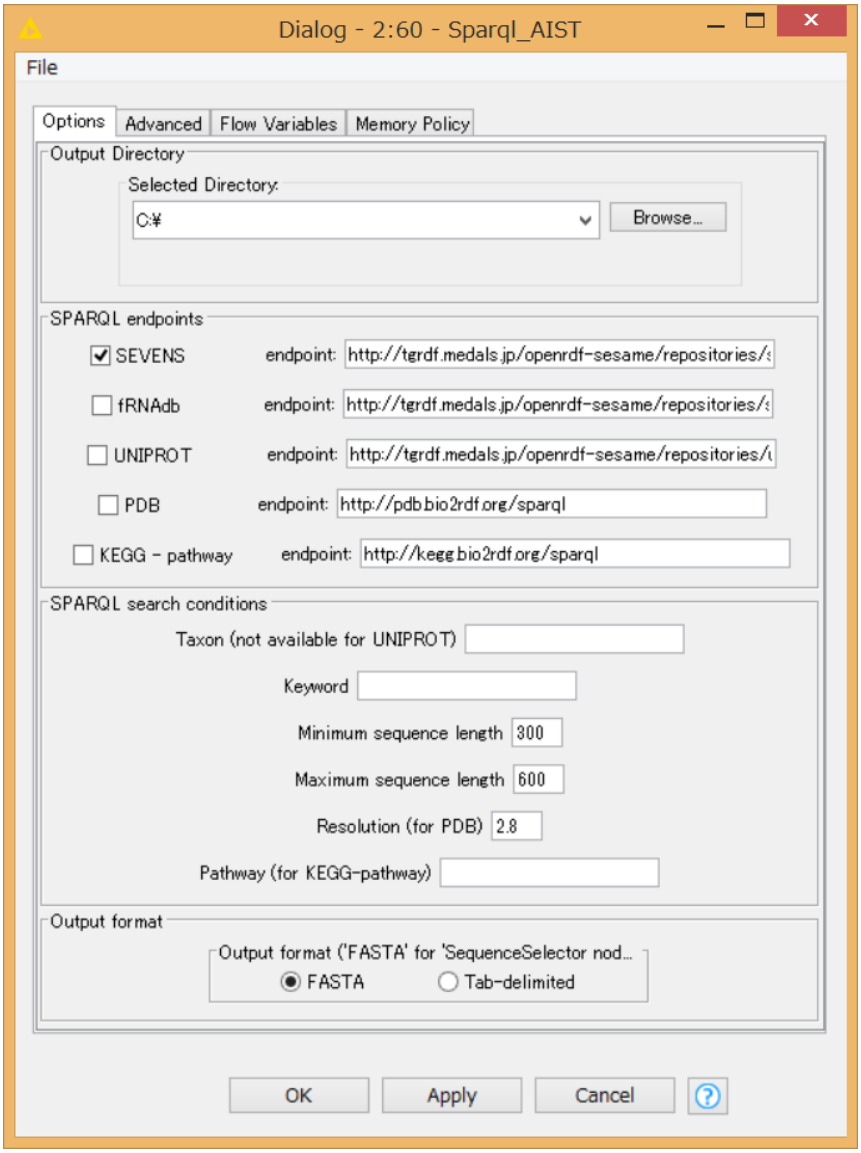
Click “OK” button after setting the conditions.

7.2.2.3 Sparql_AIST

Set AIST SPARQL endpoints and SPARQL conditions.

1) Right-clicking on the Sparql_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.3-1 Sparql_AIST : Configure…

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

Specify an output directory in a text box or “Browse”.

・**Options → SPARQL endpoints:**

Endpoints: SEVENS, fRNAdb, UNIPROT, PDB and KEGG-pathway

Check SEVENS, UNIPROT and/or PDB in this workflow.

・**Options → SPARQL search conditions:**

Specify Taxon, Keyword, Minimum sequence length, Maximum sequence length, Resolution and Pathway. In this workflow, Keyword, Minimum and Maximum sequence length, Resolution parameters are effective.

・**Options → Output format:**

Specify FASTA or Tab-delimited radio buttons.

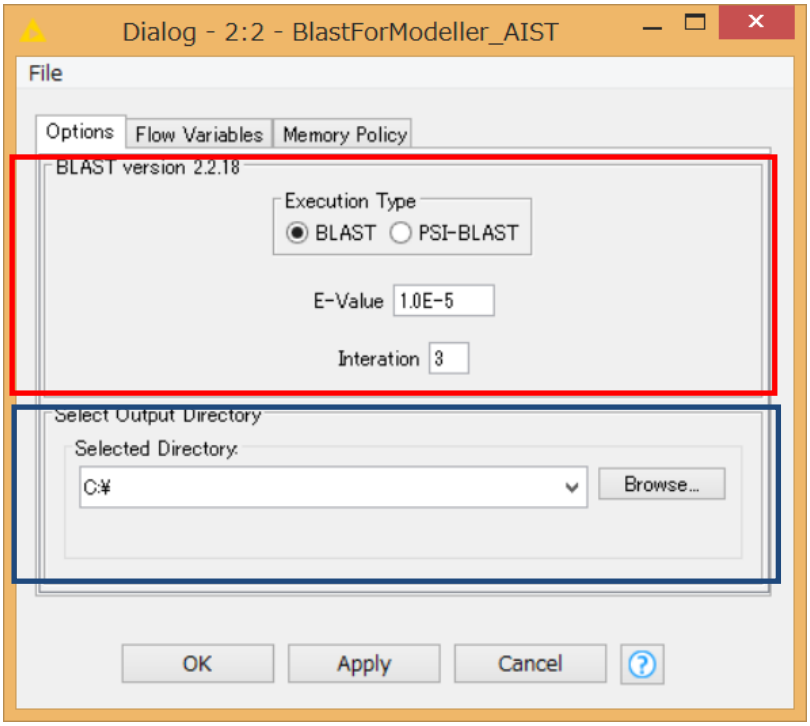
Click “OK” button after setting the conditions.

7.2.2.4 BlastForModeller _AIST

Specify either BLAST or PSI-BLAST program and set an output directory.

1) Right-clicking on the BlastForModeller_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.4-1 BlastForModeller_AIST : Configure…

- **Options → BLAST version 2.2.18 → Execution Type**

  Specify Execution Type (BLAST or PSI-BLAST) (open red rectangular)

- **Options → BLAST version 2.2.18 → E-Value**

  Specify E-value (default: 1.0E-5) (open red rectangular)

- **Options → BLAST version 2.2.18 → Iteration**

  Specify max number of Iterations for PSI-BLAST (default: 3) (open red rectangular).

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

Specify an output directory in a text box or “Browse” (blue open rectangular).

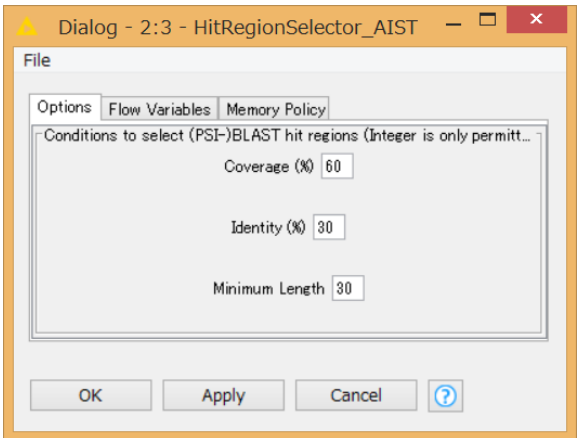
Click “OK” button after setting the conditions.

7.2.2.5 HitRegionSelector_AIST

Specify the conditions for hit regions detected by BLAST (PSI-BLAST) to use modelling.

1) Right-clicking on the HitRegionSelector_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.5-1 HitRegionSelector_AIST : Configure…

・**Options → Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) → Coverage(%)**

Specify Coverage (%) (default: 60).

・**Options → Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) → Identity(%)**

Specify Identity (%) (default: 30)

・Options → Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) → Minimum Length

Specify minimum length of amino acid sequence (default: 30).

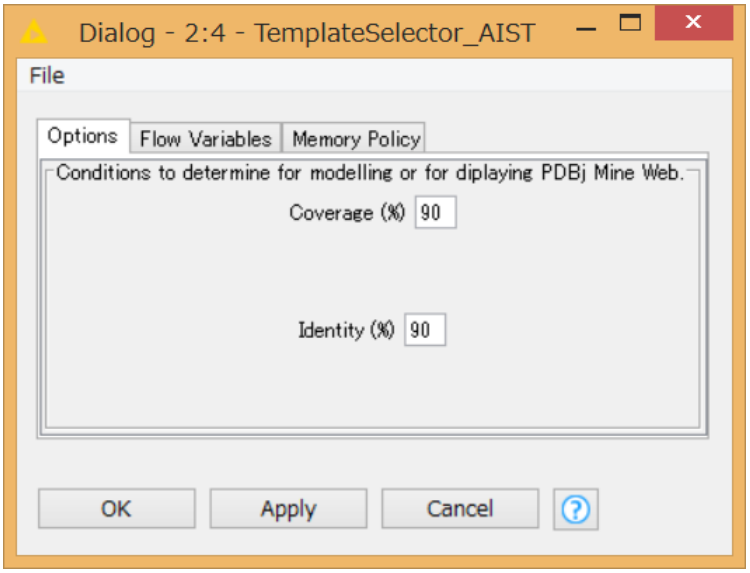
Click “OK” button after setting the conditions.

7.2.2.6 TemplateSelector_AIST

Specify Coverage and Identity.

1) Right-clicking on the TemplateSelector_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.6-1 TemplateSelector_AIST : Configure…

・Options → Condition to determine for modelling or for displaying PDBj Mine Web. → Coverage(%), Identity(%)

Specify Coverage and Identity (default: Coverage: 90, Identity: 90).

Determine a template for protein modelling. If the coverage and identity of hit regions don’t meet the template selecting conditions specified by user, protein modelling is executed by MODELLER. If not so, a web browser opens a PDBjMine web site and displays PDB structure information for the hit regions.

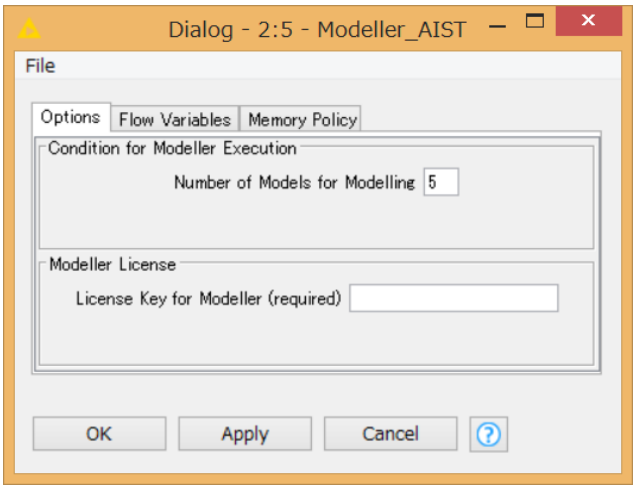
Click “OK” button after setting the conditions.

7.2.2.7 Modeller_AIST

Specify maximum number of models and input a MODELLER license key.

1) Right-clicking on the Modeller_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.7-1 Modeller_AIST : Configure…

・Options → Condition for Modeller Execution → Number of Models for Modelling

Specify maximum number of models (default: 3)

Range: 1 ~ 10

・Options → Modeller License → License Key for Modeller (required)

Input a MODELLER license key (required).

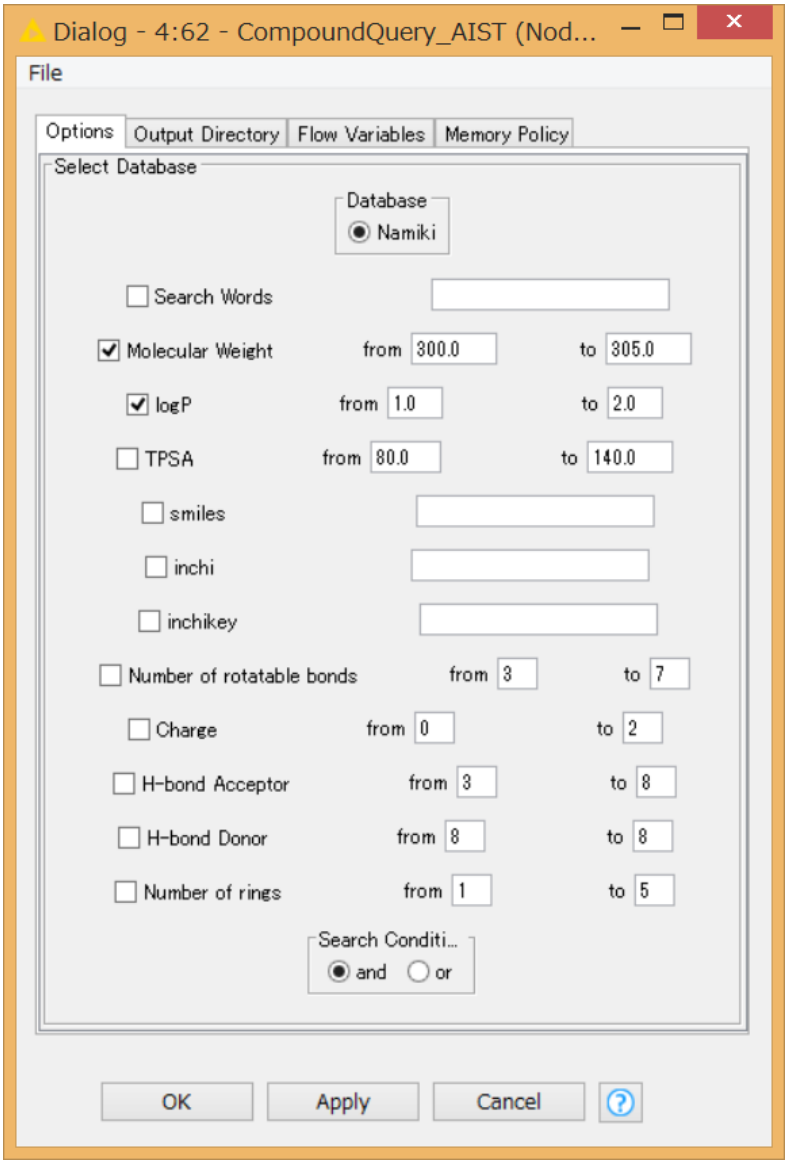
Click “OK” button after setting the conditions.

7.2.2.8 CompoundQuery_AIST

Set an output directory and chemical compound search conditions.

1) Right-clicking on the CompoundQuery_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.8-1 CompoundQuery_AIST ： Configure…

・Options

Select each checkbox if you want to use each parameter.

Database:

Specified Namiki database.

Search Words:

Input search words.

Molecular Weight:

Specify a range of molecular weight (default: 300.0~305.0).

logP:

Specify a range of logP (default: 1.0~2.0).

TPSA:

Specify a range of TPSA (default: 80.0~140.0).

smiles:

Input smiles.

inchi:

Input inchi.

inchikey:

Input inchikey.

Number of rotatable bonds:

Specify a range of rotatable bond (default: 3~7).

Charge:

Speficy a range of Charge (default: 0~2).

H-bond Acceptor:

Specify a range of H-bond Acceptor (default: 3~8).

H-bond Donor:

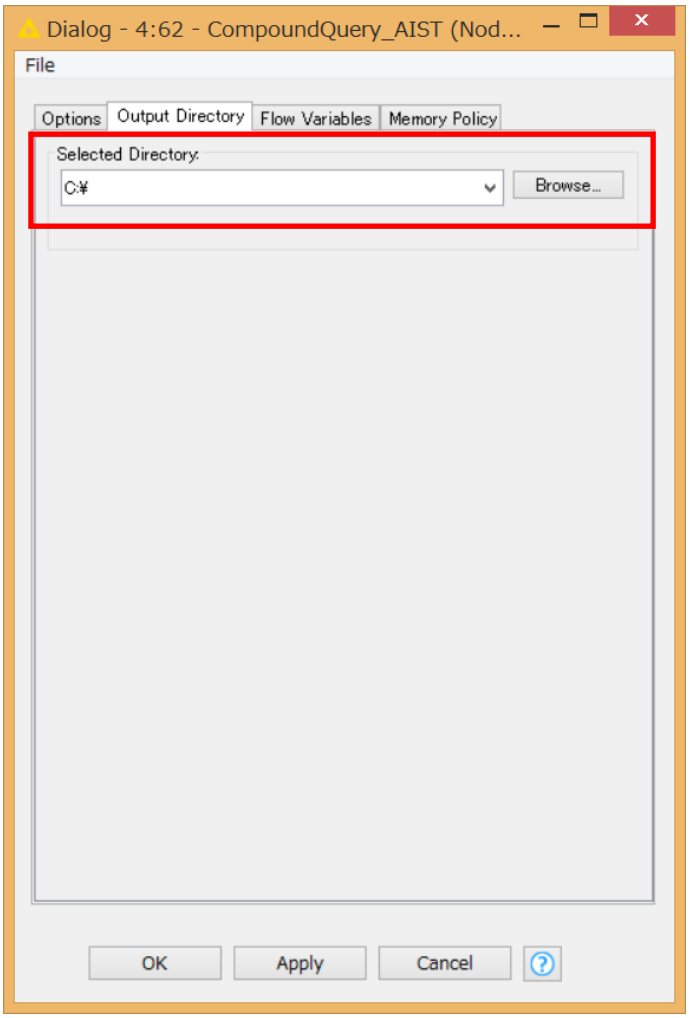
Specify a range of H-bond Donor (default: 8~8).

Number of rings:

Specify a range of ring (default: 1~5).

Search Condition:

Specify either AND or OR (default: AND).



7.2.2.8-2 CompoundQuery_AIST ： Configure…

・Options → Select Output Directory → Selected Directory:

Specify an output directory in a text box or “Browse” (blue open rectangular).

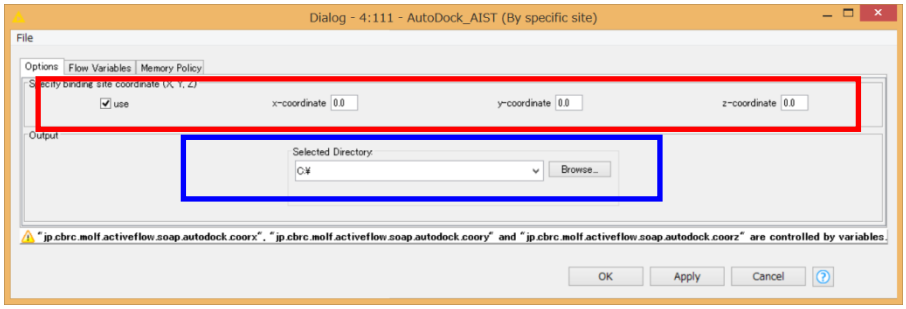
Click “OK” button after setting the conditions.

7.2.2.9 AutoDock_AIST

Set an output directory.

1) Right-clicking on the AutoDock_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.9-1 AutoDock_AIST : Configure…

- **Options → XYZ-coordinate**

  Specify X, Y, and Z coordinates (red open rectangular) (Optional).

- **Options → Select Output Directory → Selected Directory:**

  Specify an output directory in a text box or “Browse” (blue open rectangular).

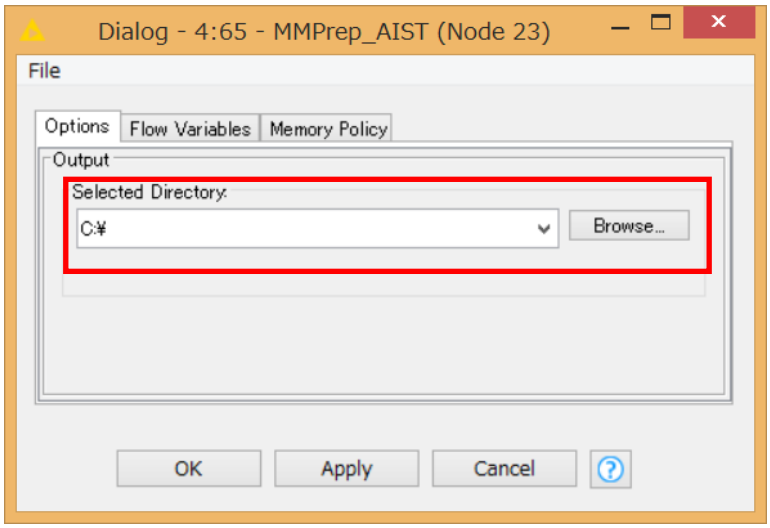
Click “OK” button after setting the conditions.

7.2.2.10 MMPrep_AIST

Set an output directory.

1) Right-clicking on the MMPrep_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.10-1 MMPrep_AIST ：Configure…

・Options → Output → Selected Directory :

Specify an output directory in a text box or “Browse” (red open rectangular).

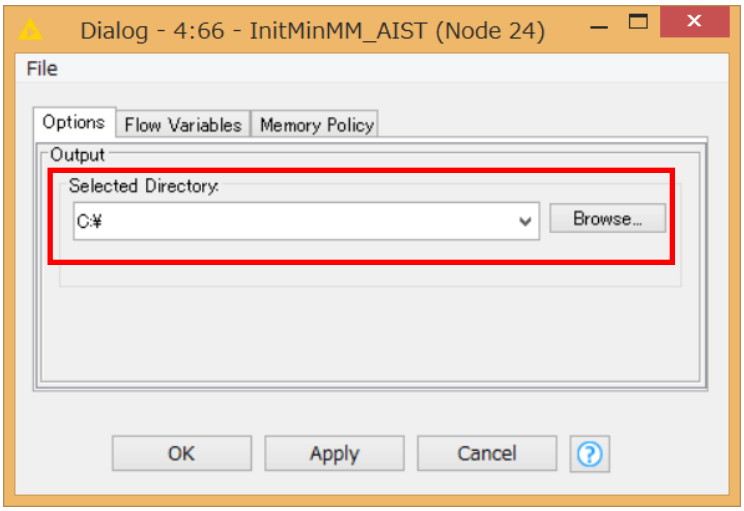
Click “OK” button after setting the conditions.

7.2.2.11 InitMinMM_AIST

Set an output directory.

1) Right-clicking on the InitMinMM_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.11-1 InitMinMM_AIST : Configure…

・Options → Output → Selected Directory :

Specify an output directory in a text box or “Browse” (red open rectangular).

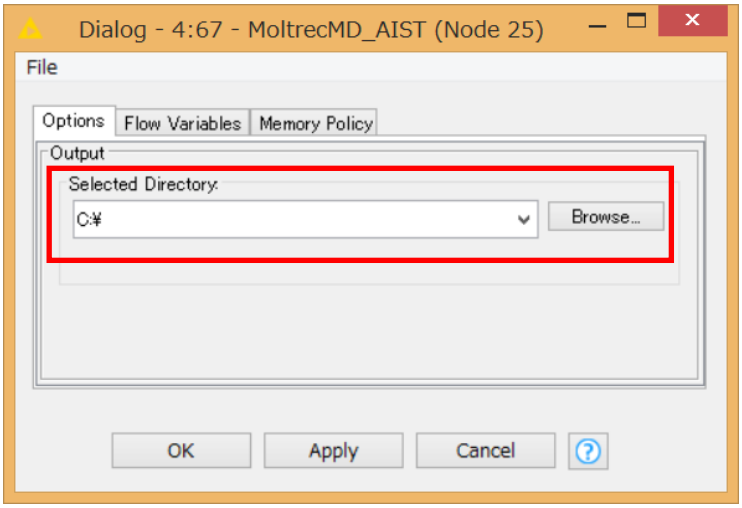
Click “OK” button after setting the conditions.

7.2.2.12 MoltrecMD_AIST

Set an output directory.

1) Right-clicking on the MoltrecMD_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.12-1 MoltrecMD_AIST ：Configure…

・Options → Output → Selected Directory :

Specify an output directory in a text box or “Browse” (red open rectangular).

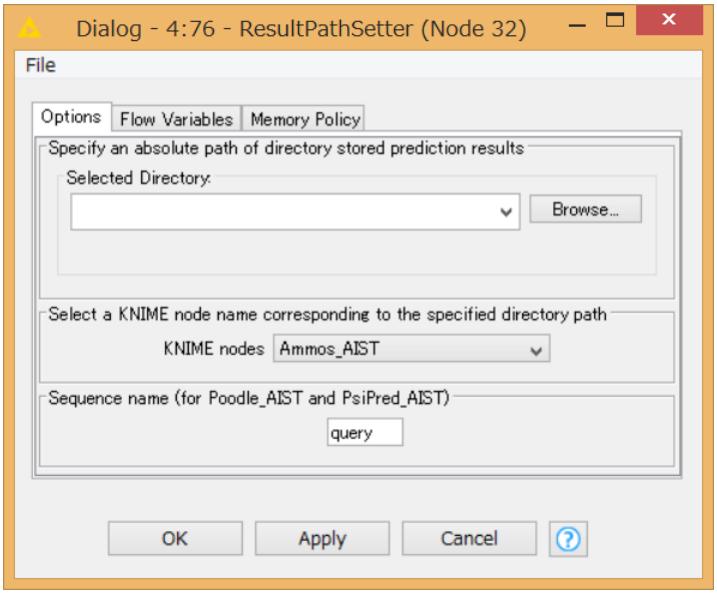
Click “OK” button after setting the conditions.

7.2.2.13 ResultPathSetter

Set an output redictory stored prediction results, KNIME node name and query sequence name (for Poodle_AIST and PsiPred_AIST).

1) Right-clicking on the ResultPathSetter node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.13-1 ResultPathSetter ：Configure…

・**Options**

Selected Directory：

Specify an output directory in a text box or “Browse” (red open rectangular).

KNIME nodes：

Select a KNIME node name corresponded to specified results.

Sequence Name：(for Poodle_AIST and PsiPred_AIST)

Input a query sequence name (default: query).

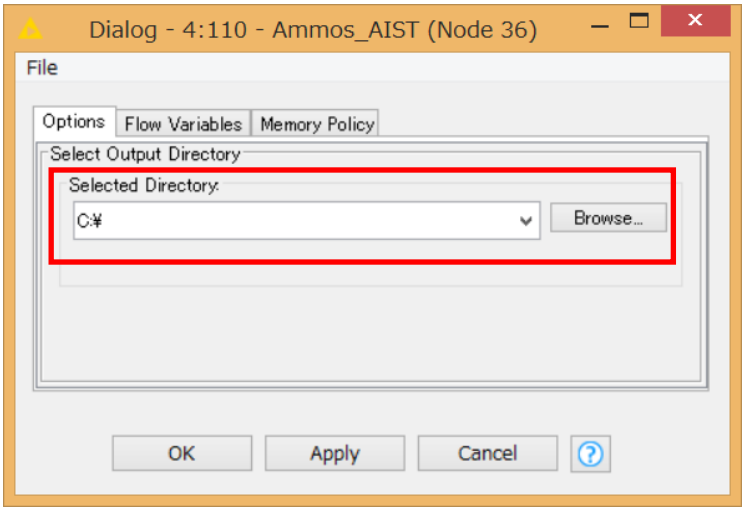
Click “OK” button after setting the conditions.

7.2.2.14 Ammos_AIST

Set an output directory.

1) Right-clicking on the Ammos_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.14-1 Ammos_AIST : Configure…

- **Options → Output → Selected Directory :**

  Specify an output directory in a text box or “Browse” (red open rectangular).

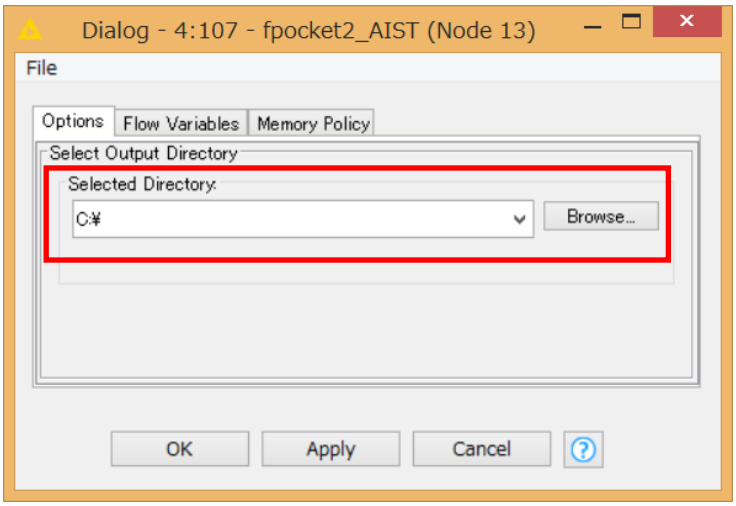
Click “OK” button after setting the conditions.

7.2.2.15 fpocket2_AIST

Set an output directory.

1) Right-clicking on the fpocket2_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.15-1 fpocket_AIST : Configure…

- **Options → Output → Selected Directory :**

  Specify an output directory in a text box or “Browse” (red open rectangular).

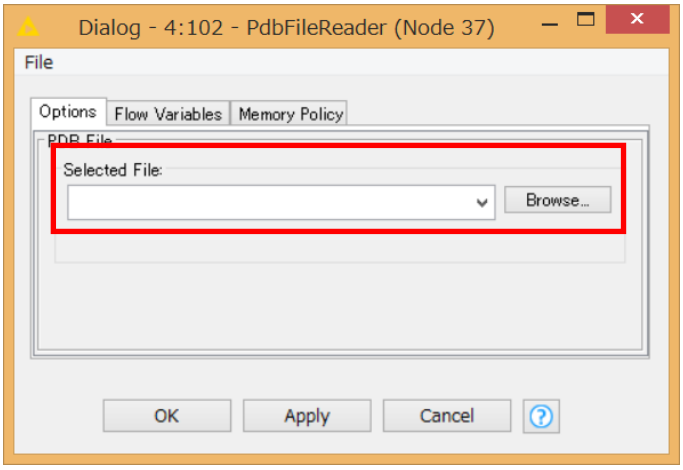
Click “OK” button after setting the conditions.

7.2.2.16 PdbFileReader

Set a PDB format ATOM file.

1) Right-clicking on the PdbFileReader node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.16-1 PdbFileReader : Configure…

- **Options → PDB File → SelectedFile:**

  Specify a PDB format ATOM file in a text box or “Browse” (red open rectangular).

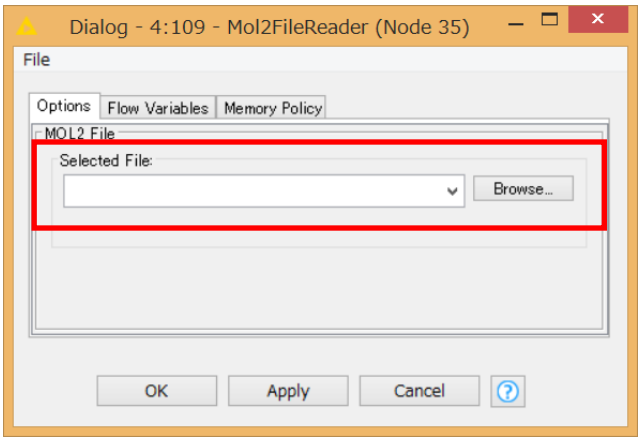
Click “OK” button after setting the conditions.

7.2.2.17 Mol2FileReader

Set a Mol2 format file.

1) Right-clicking on the Mol2FileReader node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



7.2.2.17-1 Mol2FileReader : Configure…

・Options → MOL2 File → SelectedFile:

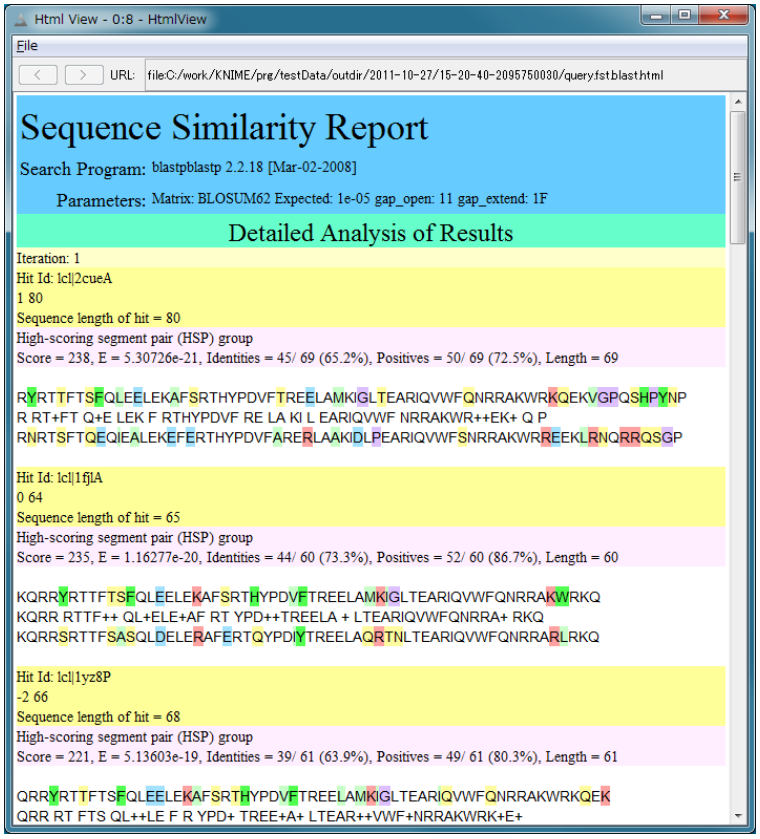
Specify a Mol2 format file in a text box or “Browse” (red open rectangular).

Click “OK” button after setting the conditions.

7.3 Results

7.3.1 BlastForModeller_AIST → HtmlView

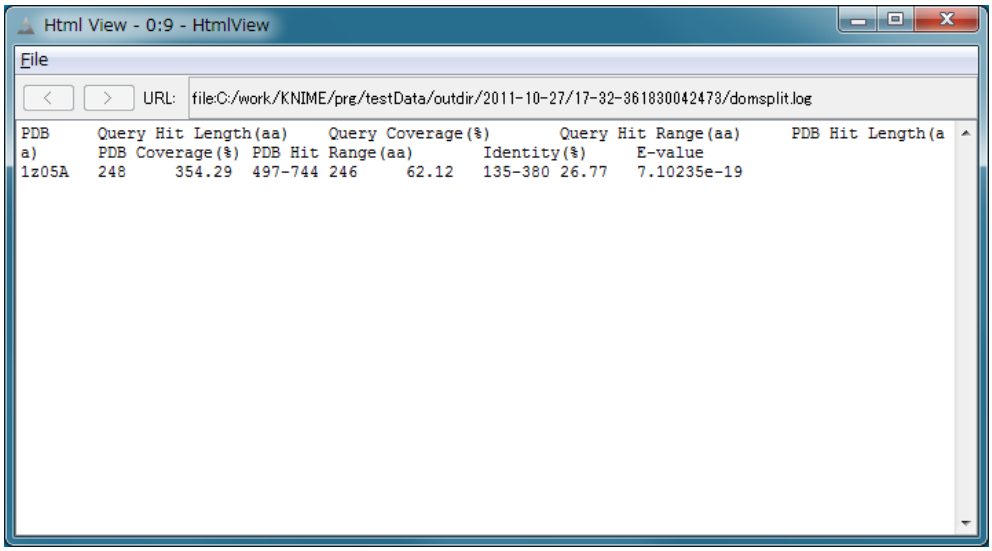
Results of BlastForModeller_AIST can be displayed using the HtmlView node.



7.3.1-1 BlastForModeller_AIST execution results (HtmlView)

7.3.2 HitRegionSelector_AIST → HtmlView

Results of HitRegionSelector_AIST can be displayed using the HtmlView node.



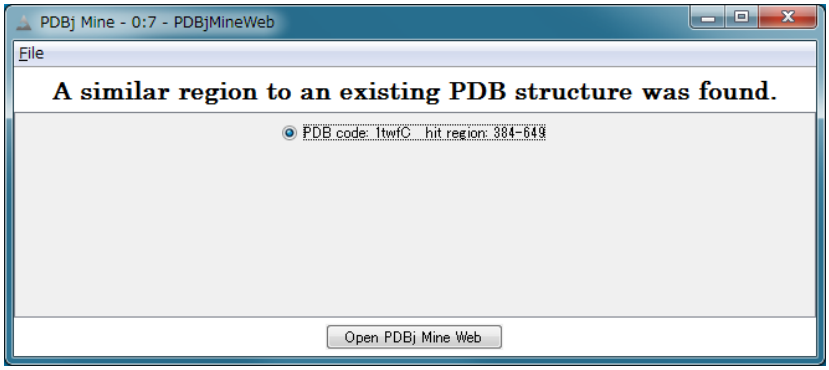
7.3.2-1 HitRegionSelector_AIST execution results (HtmlView)

7.3.3 TemplateSelector_AIST → PDBjMineWeb

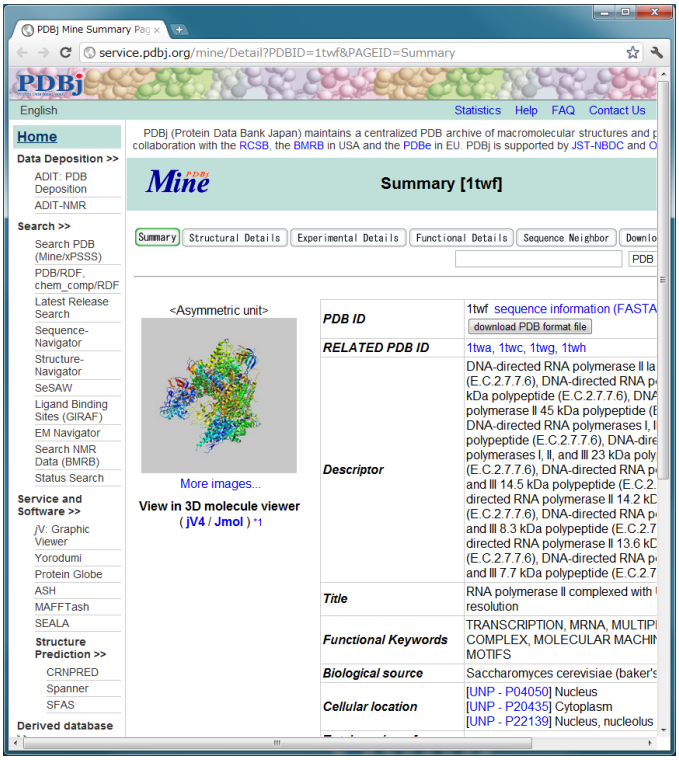
If an identity and coverage of hit regions are high compared with user setting conditions, results of TemplateSelector_AIST can be displayed using PDBjMineWeb node.

First, a dialog window is opened and PDB code and hit region range of its structure are displayed with a radio button.

By selecting a radio button and clicking “Open PDBj Mine Web” button, a PDBjMine web site is opened and displays a corresponding structure information.



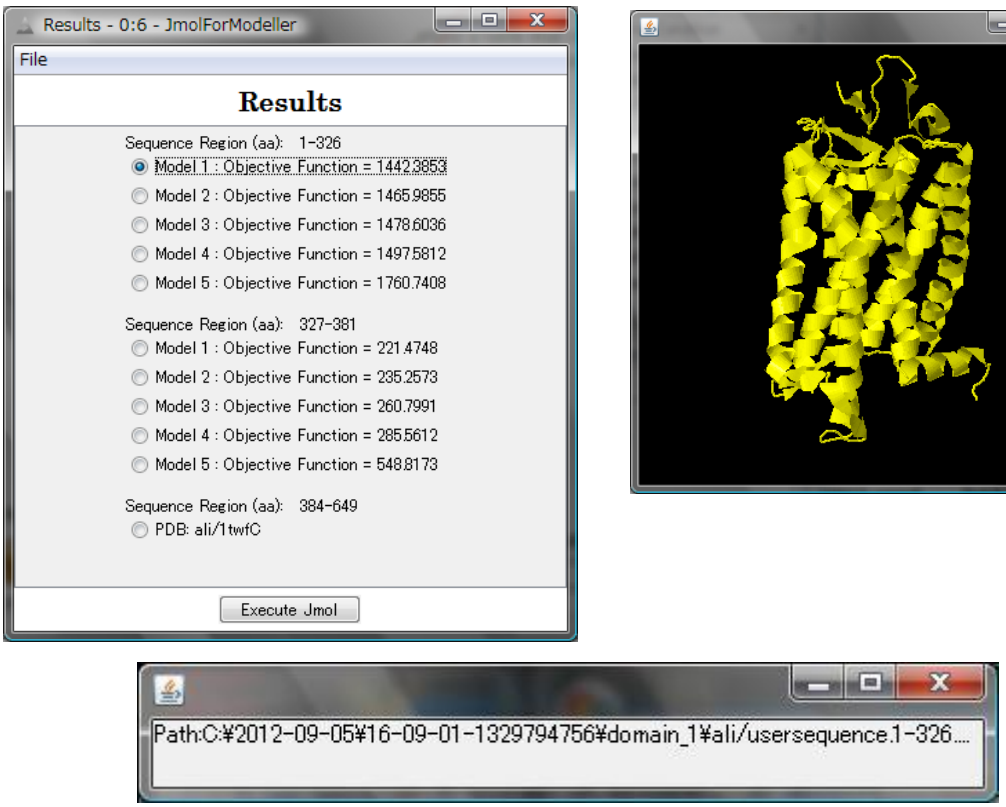
7.3.3-1 PDBjMineWeb – PDBj Mine



7.3.3-2 PDBjMineWeb – PDBj Mine

7.3.4 Modeller_AIST → JmolForModeller

Results of Modeller_AIST can be displayed using JmolForModeller node.



7.3.4-1 JmolForModeller – Modeller Results

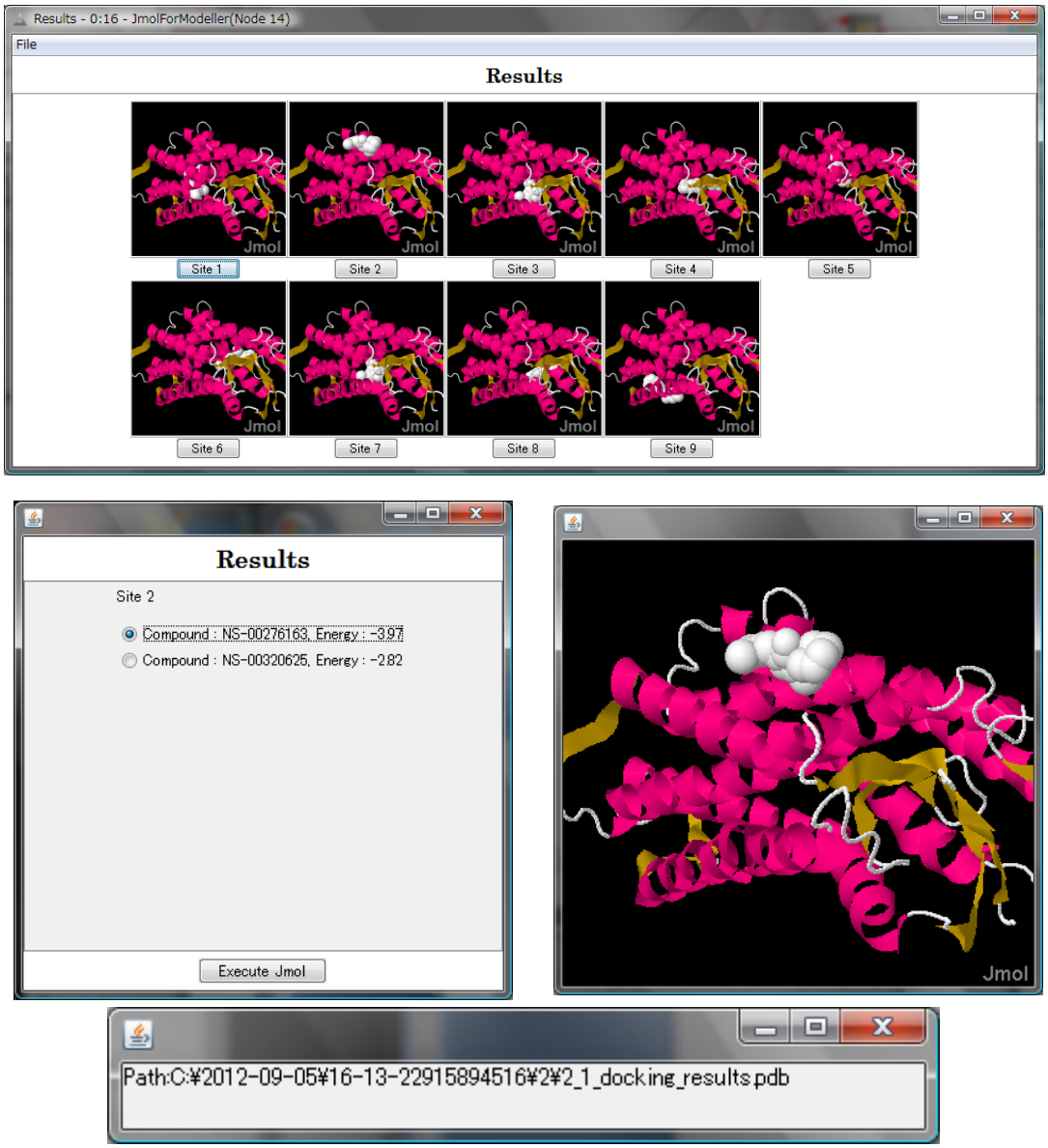
First, a dialog window is opened and models and these objective functions (modelled structure only) are displayed in every hit region category. By selecting a radio button corresponded to each model and clicking “Execute Jmol” button, Jmol starts and displays a specified structure.

Please a Jmol web site about how to operate Jmol.

Jmol : http://jmol.sourceforge.net/

7.3.5 AutoDock_AIST → JmolForModeller

Results of AutoDock_AIST can be displayed using JmolForModeller node.



7.3.5-1 JmolForModeller – Results

First, a dialog window is opened and every docking image is displayed. By clicking a button below each docking image, another dialog window is also opened. The window displays compound name and engery score with a radio button. By selecting a radio

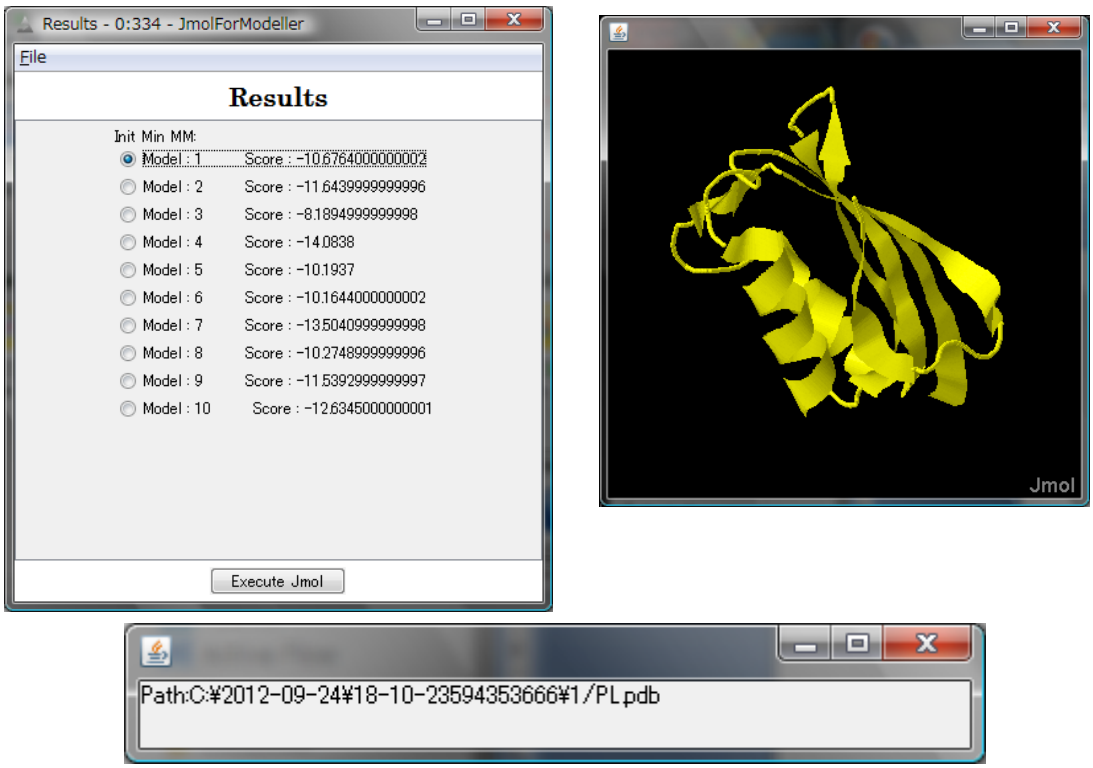
button and clicking a “Execute Jmol” button on the bottom of the screen, Jmol starts and specified docking result is displayed.

Please a Jmol web site about how to operate Jmol.

Jmol : http://jmol.sourceforge.net/

7.3.6 InitMinMM_AIST → JmolForModeller

Results of InitMinMM_AIST can be displayed using JmolForModeller node.



7.3.6-1 JmolForModeller –Results

First, a dialog window is opened and model structures and their scores are displayed with radio buttons. By selecting a model and clicking a “Execute Jmol” button on the bottom of the screen, Jmol starts and specified model structure is displayed.

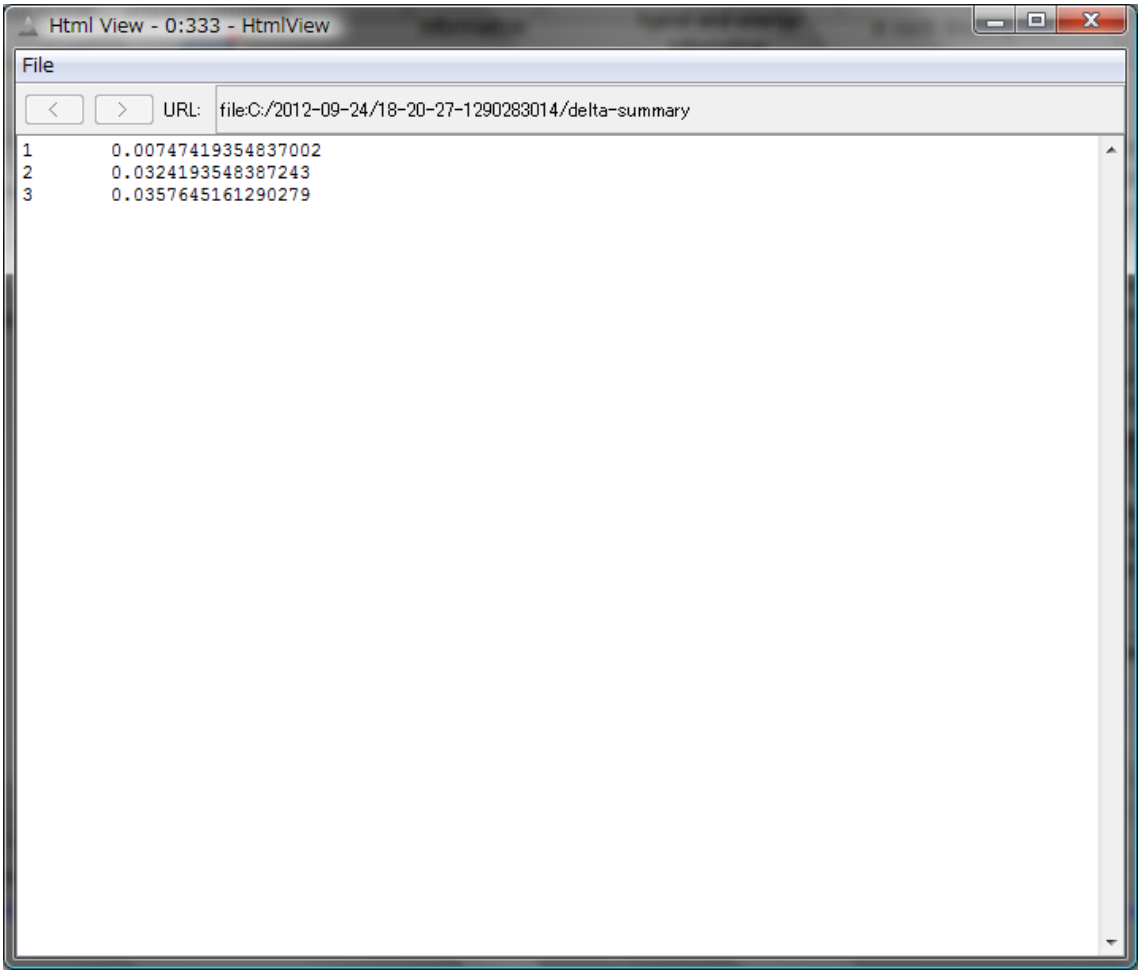
Please a Jmol web site about how to operate Jmol.

Jmol : http://jmol.sourceforge.net/

7.3.7 MoltrecMD_AIST → HtmlView

Results of MoltrecMD_AIST can be displayed using HtmlView node.

Molecular dynamics calculation scores are displayed in every model.



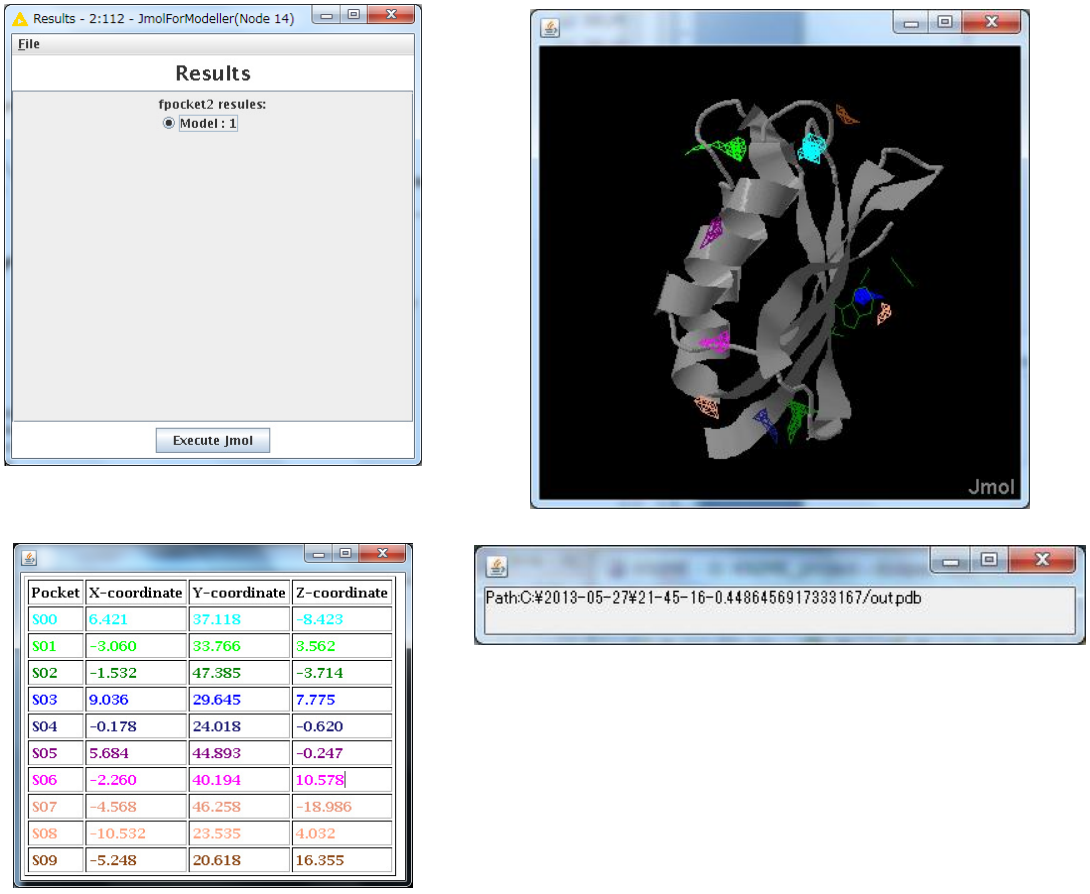
7.3.7-1 HtmlView – Results

7.3.8 ResultPathSetter → JmolForModeller、HtmlView

Results of ResultPathSetter can be displayed using JmolForModeller or HtmlView node.

7.3.9 fpocket2_AIST → JmolForModeller

Results of fpocket2_AIST can be displayed using JmolForModeller node.



Pocket	X-coordinate	Y-coordinate	Z-coordinate
$00	6.421	37.118	-8.423
$01	-3.060	33.766	3.562
$02	-1.532	47.385	-3.714
$03	9.036	29.645	7.775
$04	-0.178	24.018	-0.620
$05	5.684	44.893	-0.247
$06	-2.260	40.194	10.578
$07	-4.568	46.258	-18.986
$08	-10.532	23.535	4.032
$09	-5.248	20.618	16.355

7.3.9-1 JmolForModeller –Results

First, a dialog window is opened. By clicking a “Execute Jmol” on the bottom of the screen, Jmol starts and specified structure with pocket sites is displayed. Another popup window is also opened and displays XYZ-coordinates of each pocket site.

Please a Jmol web site about how to operate Jmol.

Jmol : http://jmol.sourceforge.net/

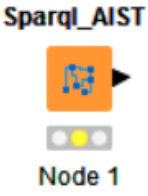
8 SPARQL

There are two types of SPARQL execution nodes and two relatives.

8.1 Nodes

8.1.1 Node list

8.1.1-1 SPARQL node list

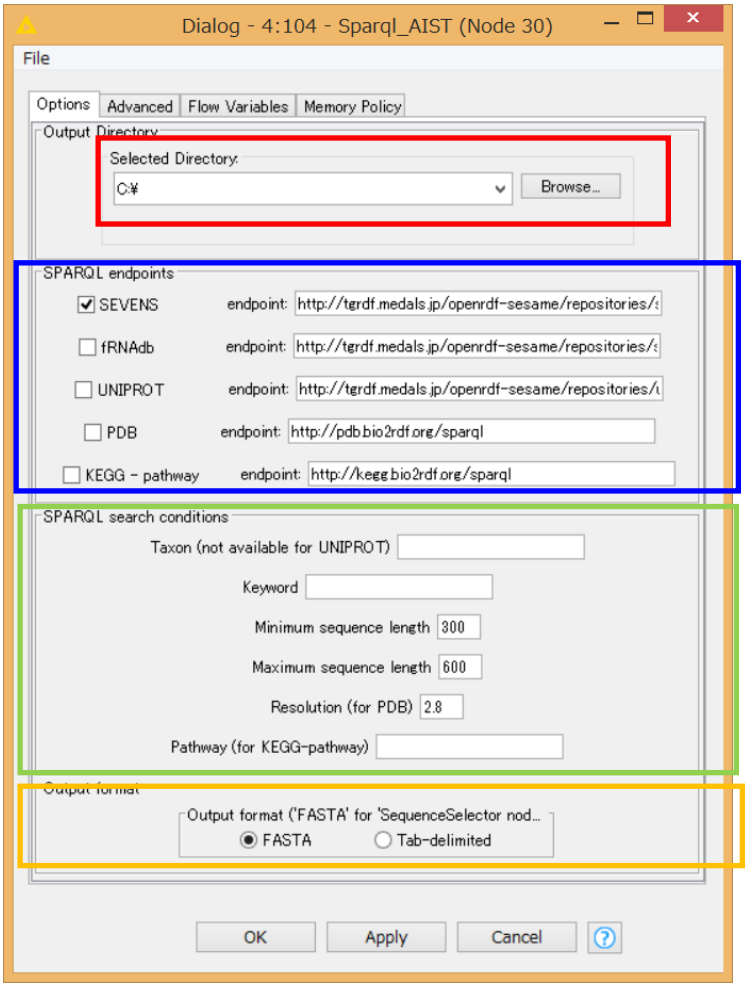
No	Name	Icon	Set	Description
1	Sparql_AIST		Y	Execute SPARQL against AIST endpoints.
2	Sparql_AIST_Adv		Y	Execute SPARQL against public endpoints.
3	SequenceSelector			Display SPARQL results and generate FASTA file for specified result.
4	HtmlView			Display SPARQL results.

8.1.2 Sparql_AIST

8.1.2.1 Node setting

1) Right-clicking on the Sparql_AIST node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



8.1.2.1-1 Sparql_AIST : Configure…

8.1.2.1-2 Sparql_AIST ：Configure…

・**Options → Output Directory:**

Specify an output directory in a text box or “Browse” (red open rectangular).

・**Options → SPARQL Endpoints:**

Specify AIST SPARQL endpoints (blue open ractangular). Types of AIST SPARQL endpoint are as follows:

SEVENS, fRNAdb, UNIPROT, PDB and KEGG-pathway

・**Options → SPARQL search conditions:**

Specify SPARQL search conditions (green open rectangular). Types of search condition are as follows:

1) Taxon (except for UNIPROT and KEGG-pathway)
2) Keyword seach (except for KEGG-pathway)
3) Minimum and maximum sequence length (except for KEGG-pathway)
4) Resolution (for PDB)
5) Pathway (for KEGG-pathway)

・Options → Output format:

Specify either FASTA or Tab-limited (orange open rectangular). Please specify FASTA if you want to use SequenceSelector node.

・Advanced：

Input SPARQL sentence. If user input something in this filed, all above SPARQL search conditions become invalid.

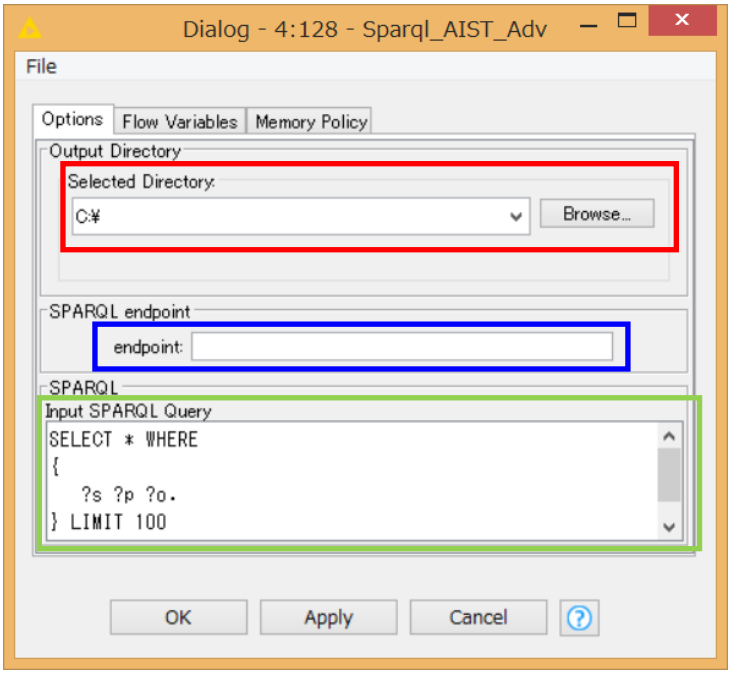
Click “OK” button after setting the conditions.

8.1.3 Sparql_AIST_Adv

8.1.3.1 Node setting

1) Right-clicking on the Sparql_AIST_Adv node and specify ”Configure…”.

2) Open the below pop-up window and set execution conditions.



8.1.3.1-1 Sparql_AIST_Adv : Configure…

- **Options → Output Directory:**

  Specify an output directory (red open rectangular).

- **Options → SPARQL Endpoints:**

  Input a public SPARQL endpoint (blue open ractangular).

- **Options → Input SPARQL Query:**

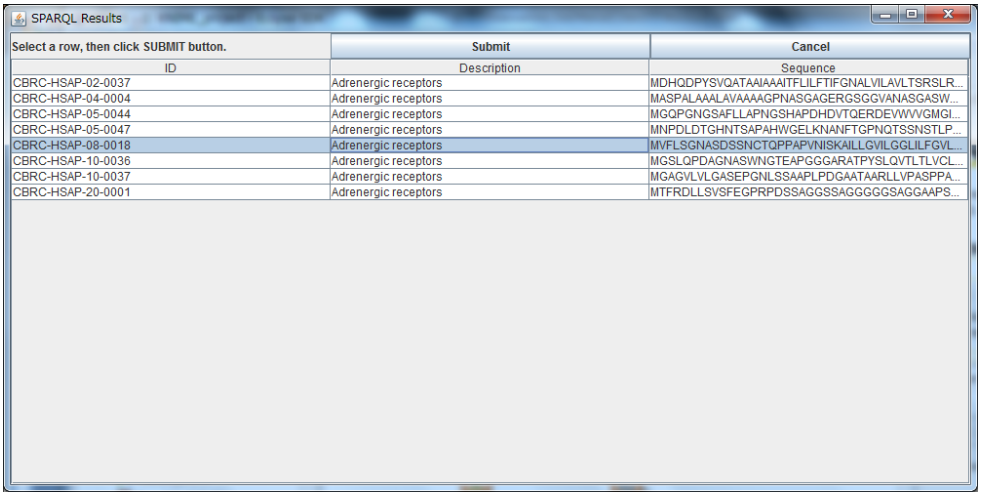
  Input SPARQL sentence (green open rectangular).

  Click “OK” button after setting the conditions.

8.1.4 SequenceSelector

8.1.4.1 Result

By right-clicking on the SequenceSelector node and select “Execute” menu, SequenceSelector is executed and displays SPARQL results. This node is only available if FATA format parameter is specified in the Sparql_AIST node.



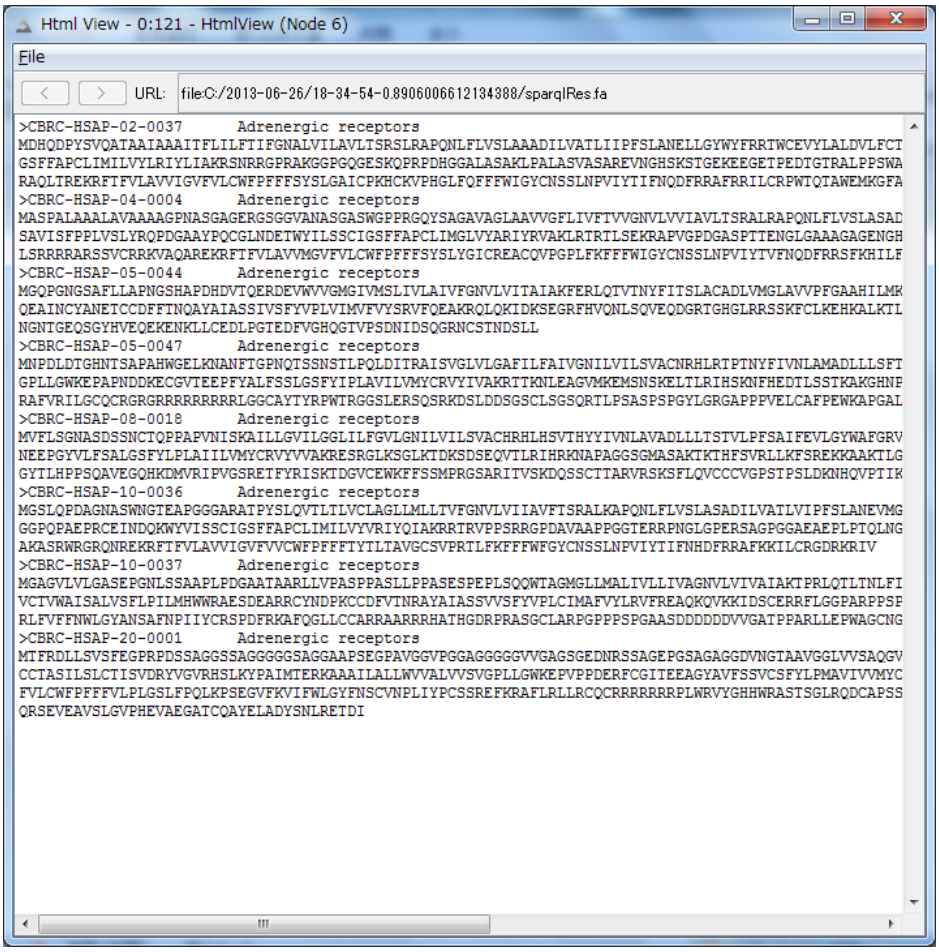
8.1.4.1-1 SequenceSelector results

User can specify one sequence information. By clicking a Submit button on the top of the screen, a FASTA file is generated and set its file path to an outport of this node.

8.1.5 HtmlView

8.1.5.1 Result

By right-clicking on the HtmlView node and select “Execute” menu, HtmlView node is executed. This node is available for displaying Sparql_AIST and Sparql_AIST_Adv nodes.

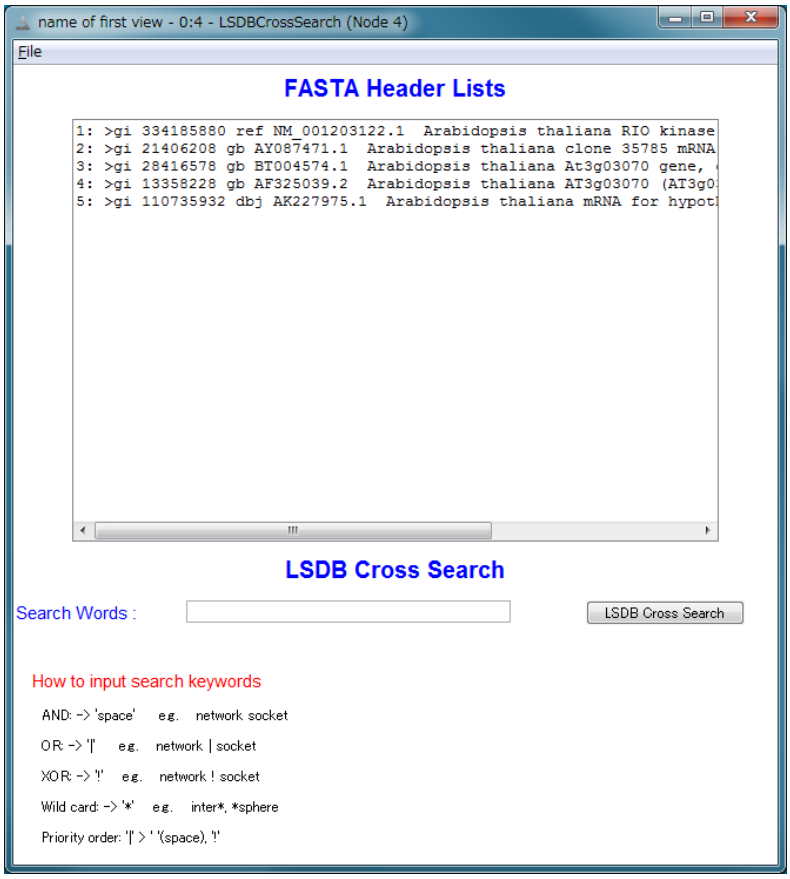


8.1.5.1-1 HtmlView

9 Appendix

9.1 LSDBCrossSearch

LSDBCrossSearch node executes LSDB cross search based on search words.



9.1-1 LSDBCrossSearch View window

FASTA header line(s) is(are) displayed in FASTA Header Lists text box.
User can input search words in a text box based on below rules.

- AND : single space
- OR : |
- XOR : !
- Wild card : *

OR is given priority over AND.

By clicking the LSDB Cross Search button, this site is opened and performs a search based on specified words.

Please visit a LSDB Cross Search web site for further information.

LSDB Cross Search site：http://biosciencedbc.jp/dbsearch/

10 Contact

Please send your queries or comments, if you have, to the address below.
workflow@molprof.jp

Molecular Profiling Research Center for Drug Discovery of AIST plans to listen to user's requests positively, and 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