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 P	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 Res	sults	61
5.3.1 Bla	$stForModeller\_AIST \rightarrow HtmlView$	61
5.3.2 Hit	$RegionSelector\_AIST \rightarrow HtmlView$	62
5.3.3 Ter	nplateSelector_AIST → PDBjMineWeb	63
5.3.4 Mo	$deller\_AIST \rightarrow JmolForModeller$	64
5.3.5 CP	$Hmodels\_DTU \rightarrow JmolForModeller$	65
5.3.6 CP	$Hmodels\_DTU \rightarrow HtmlView$	66
5.3.7 Psi	$Pred\_AIST \rightarrow AISTViewer \dots$	.67
5.3.8 Poo	$dle\_AIST \rightarrow AISTViewer$	. 69
5.3.9 Tm	$Hmm_DTU \rightarrow HtmlView$	.71
5.3.10 S	$\operatorname{SignalP}_{\operatorname{DTU}} \rightarrow \operatorname{HtmlView}$	.72
5.3.11 C	$ChloroP_DTU \rightarrow HtmlView$	.73
5.3.12 L	$ipoP_DTU \rightarrow HtmlView$	.74
5.3.13 N	$MetCTL_DTU \rightarrow HtmlView$	.75
5.3.14 N	${\rm NetChop\_DTU} \rightarrow {\rm HtmlView} \ldots$	.76
5.3.15 N	${\rm NetNES\_DTU} \rightarrow {\rm HtmlView} \dots$	.77
5.3.16 N	$etPhosK_DTU \rightarrow HtmlView$	.78
5.3.17 N	${\tt NetPhos\_DTU} \rightarrow {\tt HtmlView}$	.79

	5.3.18	$NetPicoRNA\_DTU \rightarrow HtmlView \dots$	.80
	5.3.19	$DictyOGlyc\_DTU \rightarrow HtmlView$	.81
	5.3.20	$WolfPsort\_AIST \rightarrow HtmlView$	. 82
	5.3.21	$TargetP_DTU \rightarrow HtmlView$	.83
	5.3.22	SecretomeP_DTU $\rightarrow$ HtmlView	.84
	5.3.23	$DisoPred\_AIST \rightarrow AISTViewer$	.85
	5.3.24	$Memsat\_AIST \rightarrow AISTViewer$	.86
6	Phyl	ogenetic 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	$ClustalW\_AIST \rightarrow AISTViewer$	. 97
	6.3.2	$Mafft\_AIST \rightarrow AISTViewer \dots$	. 99
	6.3.3	PhylogeneticTreeView	101
7	Mole	ecular 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
	7.2.2.1	0 MMPrep_AIST	121
	7.2.2.1	1 InitMinMM_AIST	122

7.2.2.12	MoltrecMD_AIST	
7.2.2.13	ResultPathSetter	
7.2.2.14	Ammos_AIST	
7.2.2.15	fpocket2_AIST	
7.2.2.16	PdbFileReader	
7.2.2.17	Mol2FileReader	
7.3 Re	sults	
7.3.1 Bla	$astForModeller\_AIST \rightarrow HtmlView$	
7.3.2 Hi	$tRegionSelector\_AIST \rightarrow HtmlView$	
7.3.3 Te	$mplateSelector\_AIST \rightarrow PDBjMineWeb$	
7.3.4 Mo	$deller\_AIST \rightarrow JmolForModeller$	
7.3.5 Au	$toDock\_AIST \rightarrow JmolForModeller$	
7.3.6 Ini	$itMinMM\_AIST \rightarrow JmolForModeller$	
7.3.7 Mo	$pltrecMD\_AIST \rightarrow HtmlView$	
7.3.8 Re	$sultPathSetter \rightarrow JmolForModeller, HtmlView$	
7.3.9 fpc	$ocket2\_AIST \rightarrow JmolForModeller$	
8 SPAR	JT	
8.1 No	odes	
8.1.1 No	ode list	
8.1.2 Sp	arql_AIST	
8.1.2.1	Node setting	
8.1.3 Sp	arql_AIST_Adv	
8.1.3.1	Node setting	
8.1.4 Se	quenceSelector	
8.1.4.1 I	Result	
8.1.5 Ht	mlView	
8.1.5.1 I	Result	
9 Append	dix	145
9.1 LS	DBCrossSearch	
10 Contac	et	
	7.2.2.12         7.2.2.13         7.2.2.14         7.2.2.15         7.2.2.16         7.2.2.17         7.3         7.3.1         Bla         7.3.2         11         7.3.3         7.3.4         7.3.5         Au         7.3.6         7.3.8         7.3.8         7.3.9         7.3.9         8.1.1         Not         8.1.2         Sp         8.1.1         8.1.2         8.1.3         Sp         8.1.4         8.1.5         9         Appen         9.1         10	7.2.2.12MoltrecMD_AIST7.2.2.13ResultPathSetter7.2.2.14Ammos_AIST7.2.2.15fpocket2_AIST7.2.2.16PdbFileReader7.2.2.17Mol2FileReader7.3Results7.3.1BlastForModeller_AIST $\rightarrow$ HtmlView7.3.2HitRegionSelector_AIST $\rightarrow$ HtmlView7.3.3TemplateSelector_AIST $\rightarrow$ PDBjMineWeb7.3.4Modeller_AIST $\rightarrow$ JmolForModeller7.3.5AutoDock_AIST $\rightarrow$ JmolForModeller7.3.6InitMinMM_AIST $\rightarrow$ JmolForModeller7.3.7MoltrecMD_AIST $\rightarrow$ HtmlView7.3.8ResultPathSetter $\rightarrow$ JmolForModeller7.3.9fpocket2_AIST $\rightarrow$ JmolForModeller8SPARQL8.1Nodes8.1.1Node list8.1.2Sparql_AIST8.1.3Sparql_AIST8.1.4SequenceSelector8.1.5.1Result8.1.5.1Result9Appendix9.1LSDBCrossSearch10Contact

#### 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 : <u>http://www.knime.org/</u>

# 2 Combination types of Active workflow

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

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

2-1 Combination types of Active workflow

## 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.

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

3-3 Node execution conditions

#### 4. Node menu

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



3-4 Node menu

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	The node cannot be
	View window.	used unless the node
		status is yellow.
Cancel	Cancel the execution.	-
Reset	Reset the node progress.	-
Edit Node Name and	Edit the node name and	-
Description	Description.	
New Workflow Annotation	Write annotations in a text	-
	box.	
View : [viewer name]	Open a View window.	-
Cut	Cut the node.	-
Сору	Copy the node.	-
Paste	Paste the node.	-
Undo	Undo	-
Redo	Redo	-
Delete	Delete the node.	-

3-5 Command list

#### 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))

# $\odot$

#### 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.

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

3-6	Error	and	alert	messages
-----	-------	-----	-------	----------

## 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	

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

1       SetVariable       Y       Control CASE Switch         012       0130g:       0: first outport         1       Node 1       1: second outport         2       CASE Switch       CASE Switch         Variable(Start)       CASE Switch       Control CASE Switch         Variable(Start)       CASE Switch       Control CASE Switch         Variable(Start)       Image: SetVariable (Start)       Image: SetVariable (Start)         0       Image: SetVariable (Start)       Image: SetVariable (Start)         1       FastaFileReader       Y       Set FASTA file.         1       Image: SetVariable (Start)       Image: SetVariable (Start)       Image: SetVariable (Start)         1       Image: SetVariable (Start)       Image: SetVariable (Start)       Image: SetVariable (Start)	No	Name	Icon	Set	Description
Image: searchImage: searchVariable. Open Dialog: 0: first outport 1: second outport 2: third outport2CASE Switch Variable(Start)CASE Switch Variable (Start)Control CASE Switch using workflow variables.3FastaFileReaderFastaFileReader Node 2YSet FASTA file.4LSDBCrossSearch Node 4Image: search.Execute LSDB cross search.Execute LSDB cross search.	1	SetVariable	SetVariable	Y	Control CASE Switch
Image: second outport     Dialog:     0: first outport       Node 1     0: first outport     1: second outport       2     CASE Switch     CASE Switch     Control CASE Switch       Variable(Start)     Image: second outport     2: third outport       3     FastaFileReader     FastaFileReader     Y       8     LSDBCrossSearch     Image: second outport       9     Image: second outport     Image: second outport       9     FastaFileReader     Y       9     Set FASTA file.			012		Variable. Open
Image: Second outport       0: first outport         Node 1       1: second outport         2       CASE Switch         Variable(Start)       Control CASE Switch         Variable(Start)       Second Start)					Dialog:
Node 1       1: second outport         2       CASE Switch       CASE Switch         Variable(Start)       CASE Switch       Control CASE Switch         Variable(Start)       Image: second outport       Control CASE Switch         Variable(Start)       Image: second outport       Second outport         3       FastaFileReader       FastaFileReader       Y         Set FASTA file.       Image: second outport       Set FASTA file.         Image: second outport       Image: second outport       Set FASTA file.         Image: second outport       Image: second outport       Set FASTA file.         Image: second outport       Image: second outport       Set FASTA file.         Image: second outport       Image: second outport       Image: second outport         Image: second outport       Image: second outport       Image: second outport         Image: second outport       Image: second outport       Image: second outport         Image: second outport       Image: second outport       Image: second outport         Image: second outport       Image: second outport       Image: second outport         Image: second outport       Image: second outport       Image: second outport         Image: second outport       Image: second outport       Image: second outport			Node 1		0: first outport
Image: constraint of the second sec					1: second outport
2       CASE Switch Variable(Start)       CASE Switch Variable (Start)       Control CASE Switch using workflow variables.         3       FastaFileReader       FastaFileReader       Y         3       FastaFileReader       Y       Set FASTA file.         4       LSDBCrossSearch       LSDBCrossSearch       Execute LSDB cross search.					2:third outport
Variable(Start)       Variable (Start)       using workflow variables.         3       FastaFileReader       Y       Set FASTA file.         3       FastaFileReader       Y       Set FASTA file.         1       LSDBCrossSearch       LSDBCrossSearch       Execute LSDB cross search.         1       Node 4       Node 4       Image: Search search search search search search	2	CASE Switch	CASE Switch		Control CASE Switch
3       FastaFileReader       Y       Set FASTA file.         3       FastaFileReader       Y       Set FASTA file.         1       LSDBCrossSearch       LSDBCross Search       Execute LSDB cross search.         1       Node 4       Image: Search search search search.       FastaFileReader search.		Variable(Start)	Variable (Start)		using workflow
Image: Second			e case ≠V		variables.
3     FastaFileReader     Y     Set FASTA file.       3     FastaFileReader     Y     Set FASTA file.       1     Image: Set FastafileReader     Y     Set FASTA file.       1     Image: Set FastafileReader     Y     Set Fastafile.       4     Image: Set FastafileReader     Image: Set FastafileReader     Y       4     Image: Set FastafileReader     Image: Set FastafileReader     Image: Set FastafileReader       4     Image: Set FastafileReader     Image: Set FastafileReader     Image: Set FastafileReader       5     Image: Set FastafileReader			Node 2		
3       FastaFileReader       Y       Set FASTA file.         1       Image: Set FastaFileReader       Y       Set FASTA file.         1       Image: Set FastaFileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Set Fasta file.         1       Image: Set Fasta fileReader       Y       Y         1       Image: Set Fasta fileReader			10062		
4       LSDBCrossSearch       Execute LSDB cross search         Image: search       Image: search       Image: search <td< th=""><th>3</th><th>FastaFileReader</th><th>FastaFileReader</th><th>Y</th><th>Set FASTA file.</th></td<>	3	FastaFileReader	FastaFileReader	Y	Set FASTA file.
Image: A second seco					
A     LSDBCrossSearch     Execute LSDB cross search       Image: search     Image: search     Image: search					
4 LSDBCrossSearch LSDBCrossSearch search.			Node 3		
Node 4	4	LSDBCrossSearch	LSDBCross Search		Execute LSDB cross
Node 4			► ISCE		search.
			Node 4		

#### 4.2.1-1 RNA Structure Prediction Active Workflow node list

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

12	RactIP_AIST	RactIP_AIST	Y	Execute RactIP.
		<mark>,</mark> <mark>8° </mark> ►		
		Node 12		
13	Rascal_AIST	Rascal_AIST	Y	Execute Rascal.
		▶ 👯 ►		
		Node 13		
14	Sparql_AIST_Adv	Sparql_AIST_Adv	Y	Execute SPARQL
		<b>53</b>		against public
				enapoints.
		Node 14		
15	AISTViewer	AISTViewer		Display prediction
				results.
		Node 15		
16	HtmlView	HtmlView		Display results.
		Node 16		~
17	JmolforModeller	JmolForModeller		Start Jmol.
		► Unit		
		Node 17		
		Node 17		

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.

	Dialog - 0:75 - SetVariable 🛛 🗖 🗙
File	
Options Flow Specify an acti Active flow	Variables Memory Policy ve flow variable output port (0:first; 1:second; 2:third) variable ouput port number (integer: 0 to 2): 0 ÷
ОК	Apply Cancel 🕜

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.

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.

🛓 Dialog - 0:1 - FastaFileReader	x
File	
Easta File	
Selected File:	
- Browse	
↓ Druwse	
OK - Execute Apply Cancel	

4.2.2.2-1 FastaFileReader : Configure...

```
• Options \rightarrow Fasta File \rightarrow SelectedFile:
```

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

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.

🔺 Dialog - 0:60 - Sparql_AIST 🚽 🗖 🗙
File
Options Advanced Flow Variables Memory Policy
Output Directory
Selected Directory.
C:¥
SPARQL endpoints
SEVENS endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/s
✓ fRNAdb endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/s
UNIPROT endpoint http://tgrdf.medals.jp/openrdf-sesame/repositories/i
PDB endpoint: http://pdb.bio2rdf.org/sparql
KEGG - pathway endpoint: http://keggbio2rdf.org/sparql
SPARQL search conditions
Taxon (not available for UNIPROT)
Keyword
Minimum sequence length 10
Maximum sequence length 60
Resolution (for PDB) 2.8
Pathway (for KEGG-pathway)
Output format
Output format ('FASTA' for 'SequenceSelector nod
● FASTA ◯ Tab-delimited
OK Apply Cancel

4.2.2.3-1 Sparql\_AIST : Configure...

#### • Options $\rightarrow$ Output $\rightarrow$ Selected Directory:

Specify an output directory in a text box or "Browse".

#### • Options $\rightarrow$ SPARQL endpoints:

Endpoints: SEVENS, fRNAdb, UNIPROT, PDB and KEGG-pathway Check fRNAdb in this workflow.

#### • Options $\rightarrow$ 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 $\rightarrow$ Output format:

Specify FASTA or Tab-delimited radio buttons.

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.

🔺 Dialog - 0:33 - CentroidFold_AIST (Node 15) 🛛 🗖 💌
File
Parameters Flow Variables Memory Policy
Format • Fasta O ClustalW
Output
Selected Directory
C.¥ V Browse
Weight of base pairs Gamma: 2^2 v
Advanced Other options
OK Apply Cancel 🕐

4.2.2.4-1 CentroidFold\_AIST : Configure...

- Options  $\rightarrow$  Input type  $\rightarrow$  Format Specify Fasta or ClustalW format.
- Options  $\rightarrow$  Output  $\rightarrow$  Selected Directory:

Specify an output directory in a text box or "Browse".

• Options  $\rightarrow$  Weight of base pairs  $\rightarrow$  Gamma: Change Gamma value (default: 2^2).

#### • Options $\rightarrow$ Advanced $\rightarrow$ 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.

🔥 Dialog - 0:12 - IPknot_AIST (Node 26) 🛛 🗕 🗖 💌
File
Options Flow Variables Memory Policy Select Output Directory C# Browse
Advanced Options
OK Apply Cancel 🕐

4.2.2.5-1 IPknot\_AIST : Configure...

#### • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

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.

△ Dialog - 0:20 - RactIP_AIST (Node 37) - □ ×
File
Options Flow Variables Memory Policy
Selected Directory
C.¥ ✓ Browse
Advanced Options
OK Apply Cancel ?

4.2.2.6-1 RactIP\_AIST : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory:

Specify an output directory in a text box or "Browse" (red open rectagular). Click "OK" button after setting the conditions. 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.

	Dialog - 0:44 - Rascal_AIST (Node 43) 🛛 🗕 🗙
File	
Options	Flow Variables Memory Policy
Output	Selected Directory.
- Advanci	ed Options
	OK Apply Cancel 🕐

4.2.2.7-1 Rascal\_AIST : Configure...

#### • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

#### • Options $\rightarrow$ Advanced:

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

#### 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.

🔥 Dialog - 0:34 - RASSIE_AIST (Node 18) 🛛 🗕 🗖 🗙
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory
C:¥ V Browse
Advanced
-q 100 -ins 100 -clst -outclst 10 -ins_chain
OK Apply Cancel 🕡

4.2.2.8-1 RASSIE\_AIST : Configure...

#### • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

#### • Options $\rightarrow$ Advanced:

Specify RASSIE execution parameters.

## 4.3 Results

## 4.3.1 IPknot\_AIST

## 1) <u>HtmlView</u>

Results of IPknot\_AIST (text format) can be displayed using the HtmlView node.

🛓 Html View - 0:13 - HtmlView		x
Eile		
URL: file:C:/work/KNIME/prg/testData/outdir/2011-11-01/10-43-22-0.023648505524485608/ipk	not.out	
>gi 334185880 ref NM_001203122.1  Arabidopsis thaliana RIO kinase 2 (AT3G51270) mRNA, cds	complete	*
GTGTCTAAAGGATTTTCGAGTTCACAAAAAGAATTTCCTCGCTCTACCGCCGCCGTCTCTCTC	AGAGAAGAAA	
GGGATGCGAAACCATGAGATTGTTCCTTCTGAGCTTGTGGAGCGCATTGCTTGTCTAAAACATGATGTTTCGAGATGCTCCTG GGGAATGCGAAACCATGAGATTGTTCCTTCTGAGCTTGTGGAGCGCATTGCTTGTCTAAAACATGGAGGCACCTACAAGGTCCTG	AAGAACTTGC	-
TCAAGTATAAGCTTTTGCACCACGATTCCTCTAAATATGATGGATTCCGACTCACCTATCTGGGTTATGACTTTCTTGCCATTAA	AACATTGGTC	:
AACCGGGGTATATTTACCGGTGTCGGTCGTCAGATTGGTGTTGGTAAAGAGTCAGACATATTTGAGGTCGCTCAGGAAGATGGAA AATGAAGTTACATAGACTAGGGAGAACATCCTTTTAGGGCTGTCAAATTAAGGGTCAGCATACTTGAGGCATCGCAGTAGTTTCAGG	CTATTCTAGO	
TCTCCCCGACTTGCAGCTCTCAAGGAGTTTGCTTTTATGAAGGCTTTGGAAGAACATGACTTTCCCGGTTCCAAAAGCTATTGACTG	CAATAGACAT	
TGTGTTATCATGGTTCAGGTGAAGCAATTACAGAACCCTGAGACAATTTTCGAGAAGATCATTGGTATTGTTGTTCGTTTGGCTG	AGCATGGTCI	1
GAATICATIGIGACIICAAIGAATICAACAICAIGATIGAIGAIGAGAAGAAGAAGAAGAIGAIGAIIGAII	TAAAGGTGAA	( I
TCAGAGGAGACGGAGGTGGATGAGAACAGCAGACCATCTTTTTTCGATATTACTAAAGATGCTAATGCTCTGGATAAAGATCTAG	AAGCTAGTGO	;
GTTCACAAGAAAGGAGCAGACTGACCTCGATAAATTTATTGAAGGTGGGGTGGAGAAGAGTGAAGATTCTGATGAGGATGAGGAA	TCTGATGATG	1
GAGGCAGAAGTTGAGTTGGATAATACCGAGAAGGAGAGCGAGAAGCAATGGAGATGAAGTGAGTG	TGTGTGAGGA	
GGAAGAGAAAGAGGCAGAGCTGGAGAAAAATTTGGGCAAGGTAAGACGCAGAGCCATGGCAGCAGCAGGGGACGTAGAAAGTCA	CAGTCTTCAA	1
GAAACACATACAAGGACAAAGGACGCGGATCCCAAAACTCCAAGATCCACAGCAACATGAGCGGCTTTTGATACTCCAACTGTGG	GCTTGAATGG	ŧ
GCCGFATAAFGATGATGATAGCGGCCCCATAAFCTTTTTCTAGTGACTTTTTCTCGAAFCGTTGAGTTTTCCCAAAFTGAACACCCAACC	ARIGIACIOR	
	))))))))	
$(1)) \cdots \cdots ((((((((((((((((((((((((((((((($		
$\cdot ((((((((((((((((((((((((((((((((($		
((((((((((((((((((((((((((((((((((((	)	
((((((((((((((((((((((((((((((((((((		(
((((((((((((((((((((((((((((((((((((	··//////··	6
((((((((((((((((((((((((((((())))))))))	)))	1
$((((((\dots, \dots))))))))) \dots (((((((\dots, \dots))))))) \dots (((((((\dots, \dots)))))))))$	····)))));	
$(((((\cdots,)))))$		
·······(((((((((((((((((((((((((((((((	]]]]].((((	1
(((((()).))))))	))	
/		
		-

4.3.1-1 HtmlView-IPknot Results

## 1) <u>HtmlView</u>

Results of RactIP\_AIST (text format) can be displayed using the HtmlView node.

🛓 Html View - 0:21 - HtmlView	X
Eile	
URL: file:C:/work/KNIME/prg/testData/outdir/2011-11-01/11-14-33-0.9340077957916748/ractip.out	
<pre>* 0: objval = 0.00000000e+00 infeas = 0.000000000e+00 (0) * 200: objval = 7.444637042e+01 infeas = 0.000000000e+00 (0) * 378: objval = 8.946468091e+01 infeas = 0.000000000e+00 (0)</pre>	Î
Integer optimization begins         + 378: mip = not found yet <= +inf (1; 0)	
<pre>Jqi 334185880 ref NM_001203122.1  Arabidopsis thaliana RIO kinase 2 (AT3G51270) mRNA, compl cds GTGTCTAAAGGATTTTCGAGTTCACAAAAGAATTTCCTCGCTCTACCGCCGCCGCCGTCTCTCTC</pre>	ete GAAA AGAT TIGC
TCAAGTATAAGCTTTTGCACCACGATTCCTCTAAATATGATGGATTCCGACTCACCTATCTGGGTTATGACTTTCTTGCCATTAAAAACATT AACCGGGGTATATTTACCGGTGGTCGGGCGTCAGATTGGTGTTAGAAGCACGACATATTTGAGGCGCCCCAGGAAGATGGAACTATTC AATGAAGTTACATAGACTAGGGAGAACCTCCTTTAGGGCTGCTAAATCTAAGCGTGACTACTGAGGCACCGCAGTAGTTCAGCTGGT TCCCCCGACTTGCAGGCTCTCAAGGAGTTGCTTTATGAAGGCTTTGGAAGAACATGACTTCCGGTTCCAAAAGCTATTGACTGCAATAG TGTGTTATCATGGTCAGGTGAAGCAATTACAGAACCCTGGACACAATTTCGAGATCATTGGTATTGTTGTTCGTCAGCAGCAC AATTCATTGTGACTTCAATGAATCCAACATCATGATGATGATGAAGAAAATAACGATGATTTCCACAAATGGTATCCATGTTTCCA	GGTC TAGC TATC ACAT GTCT CACC
GAAATGCACCAAATGTACTTTGACCGTGATATCGAATGCATCTTCAAGTTTTTCGGAAAAAGGTTTAATATGTCTTTCCAAGAGTAAAGG TCAGAGGACGGAGGGGGGGGGG	TGAA GTGG ≡ GATG TGTT AGGA TCAA LATGG LCTGA
GACAGAGAAATTAAATTATAACGGATATGTTTCTTAATAAATCAAATCTCAAATG	
	·····
······································	·····
>gi 28416578 gb BT004574.1  Arabidopsis thaliana At3g03070 gene, complete cds ATGGCGTCGAATCTCCTGAAAGCCCTAATCCGATCTCAGATTCTTCCATCTTCCAGGAGGAATTTCAGTGTGGCGACCACACAGCTTGGCA AACAGACGATCTAGTCGGCAATCACCCCCCAAATGGATGCAGGATAGAAGCAAGAAATCACCTATGGAACTGATTAGTGAGGTTCCACCT	TTCC

4.3.2-1 HtmlView-RactIP Results

## 4.3.3 Rascal\_AIST

1) <u>JmolForModeller</u>

Results of Rascal\_AIST can be displayed using the JmolForModeller node.

Please visit a Jmol web for further information.

o mor · noto n more our our gome
----------------------------------



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) <u>JmolForModeller</u>

Results of RASSIE\_AIST can be displayed using the JmolForModeller node.

Please visit a Jmol web for further information.

Jmol : <u>http://jmol.sourceforge.net/</u>



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) <u>AISTViewer</u>

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

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

5.2.1-1 Protein Structure Prediction Active Workflow node list

3	FastaFileReader	FastaFileReader	Y	Set FASTA file.
		18- ►		
		Node 3		
4	LSDBCrossSearch	L SDBCross Search		Execute LSDB cross
				search.
		Node 4		
5	Sparql_AIST	Spargl AIST	Y	Execute SPARQL
		<b>1</b>		against AIST
				endpoints.
		Node 5		
6	SequenceSelector	Sequence Selector		Select a sequence
		▶ ▦ ▶		from SPARQL
				results.
		Node 6		
7	CASE Switch Variable	CASE Switch		CASE Switch end
	(End)	Data (End)		node.
		Node 7		
8	BlastForModeller_AIST		Y	Execute BLAST
				(PSI-BLAST).
		Node 8		
9	HitRegionSelector AIST		Y	Detect hit regions
-				from BLAST
		ra, ≤		(PSI-BLAST)
				execution results.
		Node 9		

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

17	ChloroP_DTU	ChloroP_DTU	要	Execute ChloroP.
		▶ ≫ ▶		
		Node 17		
18	LipoP_DTU	LipoP_DTU	Y	Execute LipoP.
		► <mark>⊯</mark> ►		
		Node 18		
19	NetCTL_DTU	NetCTL_DTU	Y	Execute NetCTL.
		►		
		Node 19		
20	NetChop_DTU	NetChop_DTU	Y	Execute NetChop.
		r 🔸 🕨		
		Node 20		
21	NetNES_DTU	NetNES_DTU	Y	Execute NetNES.
		▶ <mark>ॠ</mark> ►		
		Node 21		
22	NetPhosK_DTU	NetPhosK_DTU	Y	Execute NetPhosK.
		► <mark>核</mark> .►		
		Node 22		
18	LipoP_DTU	LipoP_DTU	Y	Execute LipoP.
		► <mark>⊯</mark> ►		
		Node 18		

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

26	WolfPsort_AIST	WolfPsort_AIST	Y	Execute WoLF PSORT.
27	TargetP_AIST	TargetP_DTU	Y	Execute TargetP.
28	SecretomeP_DTU	SecretomeP_DTU	Y	Execute SecretomeP.
29	DisoPred_AIST	Disopred_AIST کک Node 29	Y	Execute DISOPRED.
30	Memsat_AIST	Memsat_AIST	Y	Execute MEMSAT.
31	Sparql_AIST_Adv	Sparql_AIST_Adv	Y	Execute SPARQL against public endpoints.
32	AISTViewer	AISTViewer		Display prediction results.
33	HtmlView	HtmlView	Display prediction	
----	-----------------	-----------------	--------------------	
			results.	
		Node 16		
34	JmolForModeller	JmolForModeller	Execute Jmol.	
		► <u>Unci</u>		
		Node 17		

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.

	Dialog - 0:75 - SetVariable 🛛 🗖 🗙
File	
Optiona	
Specify	an active flow variable output port (Offirst: 1:second: 2:third)
Activ	ve flow variable ouput port number (integer: 0 to 2):
Ok	C Apply Cancel ?

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.

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.

🛓 Dialog - 0:1 - FastaFileReader	- • ×
File	
Options Memory Policy	
	-
	_
▼ Browse.	
OK - Execute Apply Ca	ancel

5.2.2-1 FastaFileReader : Configure...

```
• Options \rightarrow Fasta File \rightarrow SelectedFile:
```

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

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.

📐 🛛 Dialog - 2:60 - Sparql_AIST 🛛 🗕 🗖		
File		
Options Advanced Flow Variables Memory Policy		
Output Directory		
Selected Directory.		
C.¥		
SPARQL endpoints		
SEVENS endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/s		
fRNAdb endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/s		
UNIPROT endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/c		
PDB endpoint: http://pdb.bio2rdf.org/sparql		
KEGG - pathway endpoint: http://keggblo2rdr.org/sparqi		
SPARQL search conditions		
Taxon (not available for UNIPROT)		
Kewword		
Minimum sequence length 300		
Maximum sequence length 600		
Resolution (for PDR) 28		
Pathway (for KEGG-pathway)		
Output format		
Output format ('FASTA' for 'SequenceSelector nod		
FASTA Tab-delimited		
OK Apply Cancel (?)		

5.2.2.3-1 Sparql\_AIST : Configure...

#### • Options $\rightarrow$ Output $\rightarrow$ Selected Directory:

Specify an output directory in a text box or "Browse".

#### • Options $\rightarrow$ SPARQL endpoints:

Endpoints: SEVENS, fRNAdb, UNIPROT, PDB and KEGG-pathway Check SEVENS, UNIPROT and/or PDB in this workflow.

#### • Options $\rightarrow$ 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 $\rightarrow$ Output format:

Specify either FASTA or Tab-delimited radio button.

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.

🔺 🛛 Dialog - 2:2 - BlastForModeller_AIST 🚽 🗖 💌
File
Options Flow Variables Memory Policy
BLAST version 2.2.18
E-Value 1.0E-5
Literation (I)
Interation 3
Select Output Directory
Selected Directory
C:¥ V Browse
OK Apply Cancel 🕐

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  $\rightarrow$  BLAST version 2.2.18  $\rightarrow$  E-Value

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

## • Options $\rightarrow$ BLAST version 2.2.18 $\rightarrow$ Iteration

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

#### • Options $\rightarrow$ Select Output Directory $\rightarrow$ 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.

- Right-clicking on the HitRegionSelector\_AIST node and specify "Configure...".
- 2) Open the below pop-up window and set execution conditions.

🝐 Dialog - 2:3 - HitRegionSelector_AIST 🛛 🗕 🗙
File
Options Flow Variables Memory Policy
Conditions to select (PSI-)BLAST hit regions (Integer is only permitt
Coverage (%) 60
Identity (%) 20
Identity (wy 30
Minimum Length 30
·
OK Apply Cancel 🕐

5.2.2.5-1 HitRegionSelector\_AIST : Configure...

• Options  $\rightarrow$  Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input)  $\rightarrow$  Coverage(%)

Specify Coverage (%) (default: 60).

• Options  $\rightarrow$  Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input)  $\rightarrow$  Identity(%)

```
Specify Identity (%) (default: 30)
```

# • Options $\rightarrow$ Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) $\rightarrow$ Minimum Length

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

Specify Coverage and Identity.

- Right-clicking on the TemplateSelector\_AIST node and specify "Configure...".
- 2) Open the below pop-up window and set execution conditions.

🝐 Dialog - 2:4 - TemplateSelector_AIST 🛛 🗖 💌
File
Options Flow Variables Memory Policy Conditions to determine for modelling or for diplaying PDBj Mine Web.
Identity (%) 90
OK Apply Cancel 🕐

5.2.2.6-1 TemplateSelector\_AIST : Configure...

## • Options $\rightarrow$ Condition to determine for modelling or for displaying PDBj Mine Web. $\rightarrow$ 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.

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.

🔥 Dialog - 2:5 - Modeller_AIST 😑 🗖	x
File	
Options Flow Variables Memory Policy Condition for Modeller Execution Number of Models for Modelling 5	
Modeller License License Key for Modeller (required)	]
OK Apply Cancel 💽	

5.2.2.7-1 Modeller\_AIST : Configure...

 $\boldsymbol{\cdot}$  Options  $\rightarrow$  Condition for Modeller Execution  $\rightarrow$  Number of Models for Modelling

Specify maximum number of models (default: 3) Range:  $1 \sim 10$ 

• Options  $\rightarrow$  Modeller License  $\rightarrow$  License Key for Modeller (required) Input a MODELLER license key (required).

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

△ Dialog - 2:51 - CPHmodels_DTU (Node 11)		x
File		
Options Flow Variables Memory Policy		=
Selected Directory		
C¥	Browse	
OK Apply Cancel	0	

5.2.2.8-1 CPHmodels\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:11 - PsiPred_AIST (Node 14) 🛛 🗕 💌
File Options Flow Variables Memory Policy Select Output Directory C.¥ Browse
OK Apply Cancel 🕐

5.2.2.9-1 PsiPred\_AIST : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

△ Dialog - 2:13 - Poodle_AIST (Node 16) - □ ×
File
Options Flow Variables Memory Policy
POODLE Type POODLE-S O POODLE-L
Output Selected Directory.
OK Apply Cancel 🕐

5.2.2.10-1 Poodle\_AIST : Configure...

## • Options $\rightarrow$ Type $\rightarrow$ POODLE Type

Select POODEL program.

POODLE-S: for short disorder region prediction.

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

## • Options $\rightarrow$ Output $\rightarrow$ Selected Directory :

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

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

🔥 Dialog - 2:21 - TmHmm_DTU (Node 18) 🛛 🗕 💌
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory.
C¥ Browse
OK Apply Cancel 📀

5.2.2.11-1 TmHmm\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:19 - SignalP_DTU (Node 20) 🛛 🗕 🗖	<b>K</b>
File	
Options Flow Variables Memory Policy Select Output Directory	ī
Selected Directory.	
C¥ V Browse	
Organism type	1
Organism type Eukaryotes 🗸 🗸	
OK Apply Cancel ?	

5.2.2.12-1 SignalP\_DTU : Configure...

## • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

#### • Options $\rightarrow$ Organism Type:

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

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

🔥 Dialog - 2:23 - ChloroP_DTU (Node 22) 🛛 🗕 💌
File
Options Flow Variables Memory Policy Select Output Directory
C.¥ Browse
OK Apply Cancel

5.2.2.13-1 ChloroP\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:25 - LipoP_DTU (Node 24) — 🗖 🗙
File Options Flow Variables Memory Policy Select Output Directory
Selected Directory. C.¥ V Browse
OK Apply Cancel 🕐

5.2.2.14-1 LipoP\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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.

🔥 Dialog - 2:27 - NetCTL_DTU (Node 26) 🛛 🗕 💌
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory.
C¥
Select Supertype Supertype A1 supertype v
OK Apply Cancel 🕜

5.2.2.15-1 NetCTL\_DTU : Configure...

## • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

#### • Options $\rightarrow$ Select Supertype:

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

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.

🔥 Dialog - 2:29 - NetChop_DTU (Node 28) 🛛 🗕 💌
File
Options Flow Variables Memory Policy
Select Output Directory
C# Browse
Select Prediction method
Prediction method C term 3 v
OK Apply Cancel 🔇

5.2.2.16-1 NetChop\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

• Options  $\rightarrow$  Select Prediction method: Choose either "C term 3.0" or "20S 3.0".

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

🔺 Dialog - 2:31 - NetNES_DTU (Node 32) 🛛 🗕 🗖 💌
File
Options Flow Variables Memory Policy
Select Output Directory
Selected Directory
C.¥ V Browse
OK Apply Cancel 🕐

5.2.2.17-1 NetNES\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:33 - NetPhosK_DTU (Node 34) 🛛 🗕 💌
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory.
C.* Drowse
OK Apply Cancel

5.2.2.18-1 NetPhosK\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:35 - NetPhos_DTU (Node 36) 🛛 🗕 🗙
File
Options Flow Variables Memory Policy
Select Output Directory
Selected Directory.
C¥
OK Apply Cancel 🕐

5.2.2.19-1 NetPhos\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:37 - NetPicoRNA_DTU (Node 38) 🛛 🗕 💌	
File	
Options Flow Variables Memory Policy	
- Select Output Directory Selected Directory	
C# Browse	
OK Apply Cancel 🕜	

5.2.2.20-1 NetPicoRNA\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:43 - DictyOGlyc_DTU (Node 44) 🛛 🗖 🗙
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory
C:¥ ✓ Browse
OK Apply Cancel 🕐

5.2.2.21-1 DictyOGlyc\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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.

🔺 Dialog - 2:15 - WolfPsort_AIST (Node 46) 🛛 🗖 💌
File
Options Flow Variables Memory Policy
Type
Output Selected Directory. C:¥
OK Apply Cancel 🕐

5.2.2.22-1 WolfPsort\_AIST : Configure...

• Options  $\rightarrow$  Kingdom  $\rightarrow$  Type

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

```
• Options \rightarrow Select Output Directory \rightarrow Selected Directory:
```

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

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.

🔺 Dialog - 2:17 - TargetP_DTU (Node 48) 🛛 🗕 💌
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory.
C.¥ Browse
Organism type Organism type Non−Pla ↓
OK Apply Cancel 🔇

5.2.2.3-1 TargetP\_DTU : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

#### • Options $\rightarrow$ Organism type:

Choose either Non-Plant or Plant.

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.

🔺 Dialog - 2:45 - SecretomeP_DTU (Node 50) 🛛 🗕 💌
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory.
C:¥ V Browse
Organism type Organism type Gram-negative bacter v
OK Apply Cancel ၇

5.2.2.24-1 SecretomeP\_DTU : Configure...

## • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

#### • Options $\rightarrow$ Organism type:

Choose Gram-negative bacteria, Gram-positive bacteria or Mammalian.

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

🔥 Dialog - 2:54 - Disopred_AIST 🛛 🗕 🔍
File
Options Flow Variables Memory Policy
Selected Directory.
C.¥ ✓ Browse
OK Apply Cancel

5.2.2.25-1 DisoPred\_AIST : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

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

🔥 Dialog - 2:56 - Memsat_AIST 🚽 🗖 🗙
File
Options Flow Variables Memory Policy
Selected Directory.
C:¥
OK Apply Cancel 🕜

5.2.2.26-1 Memsat\_AIST : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory: Specify an output directory in a text box or "Browse" (red open rectangular).

## 5.3.1 BlastForModeller\_AIST $\rightarrow$ 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 $\rightarrow$ HtmlView

Results of HitRegionSelector\_AIST can be displayed using the HtmlView node.

🛓 Html View - 0:9 - HtmlView	- • ×
Eile	
URL: file:C:/work/KNIME/prg/testData/outdir/2011-10-27/17-32-361830042478/domsp	olit.log
PDB Query Hit Length(aa) Query Coverage(%) Query Hit Range(aa) a) PDB Coverage(%) PDB Hit Range(aa) Identity(%) E-value 1205A 248 354.29 497-744 246 62.12 135-380 26.77 7.10235e-19	PDB Hit Length(a 🔺

5.3.2-1 HitRegionSelector\_AIST execution results (HtmlView)

## 5.3.3TemplateSelector\_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.

🗻 PDBj Mine - 0:7 - PDBjMineWeb	_ <b>D</b> _ X _	
Eile		
A similar region to an existing PDB structure was found.		
PDB code: ItwfC hit region: 384-649		
Open PDBj Mine Web		

5.3.3-1 PDBjMineWeb – PDBj Mine



5.3.3-2 PDBjMineWeb – PDBj Mine

## 5.3.4 Modeller\_AIST $\rightarrow$ JmolForModeller

🔔 Results - 0:6 - JmolForModeller	é
File	
Results	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sequence Region (aa): 1-326 Model 1 : Objective Function = 14423853 Model 2 : Objective Function = 14659855 Model 3 : Objective Function = 1478.6036 Model 4 : Objective Function = 1497.5812 Model 5 : Objective Function = 1760.7408 Sequence Region (aa): 327-381 Model 1 : Objective Function = 221.4748 Model 2 : Objective Function = 235.2573 Model 3 : Objective Function = 260.7991 Model 4 : Objective Function = 285.5612 Model 5 : Objective Function = 548.8173	Indexes and a second seco
Sequence Region (aa): 384-649 PDB: ali/1twfC	
Execute Jmol	
Path:C:¥2012-09-05¥16-09-01-13297947	56¥domain_1¥ali/usersequence.1-326

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 : <u>http://jmol.sourceforge.net/</u>

## 5.3.5 CPHmodels\_DTU $\rightarrow$ JmolForModeller

🛓 Results - 0:5 - JmolForModeller	
Eile  CPHmodels:  Model: 1	Jmol
Execute Jmol	
Path:C:¥2013-01-15¥18-39-10-11132594	- C X X X X X X X X X X X X X X X X X X

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 : <u>http://jmol.sourceforge.net/</u>

## 5.3.6 CPHmodels\_DTU $\rightarrow$ 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 $\rightarrow$ 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 : <u>http://bioinf.cs.ucl.ac.uk/psipred/</u>



5.3.7-1 AISTViewer – PSIPRED Result
🛓 PSI	IPRED						×
# PSI	PRED HFORMAT (P	SIPRED V2.	ō by David	Jones)			^
Conf:	98887871787507	3677111675	3541000371	1357799999	999882376	2159999	
Pred:	CCCCCCCCEEEEC			ЖННННННН		CEEEEEE	
AA:	MNGIEGENETVEESI 10	20 20 vrigversperi	SAPUTTLAEF	WUFSMLAATN 40	FLLIMLGFP 50	INFLILY 60	-
	10	20	00	10	00	00	=
Conf:	87520001450567	7778778878	3872233010	4430113687	658567600	0001116	
Pred:	EEEECHHCCCCCHHH	ННННННННН	HHHCCCHHE	LEEECCCEEEE	ECCCCCCCH	CELATIC	
AA.	70	RU 80	-mvruuriii 90	100	110	120	
	10	00				120	
Conf:	72377888877330	267862011	0314676303	1015888887	752577444	2122076	
Pred:	ССНИНИНИНИИ	CEFEFEECCO	COCCOCCEE	EHHHHHHHH	HHHCCCCCC	CHHHHCC	
AA:	GETALWOLVVLATER	1.40	1EO	MGYAFTWYMA 100	170	100 ton	
	150	140	150	100	170	100	
Conf:	57410124441167	5547877267	7223689999	9999998610	120001453	2011210	
Pred:	CCCEECCCCCCCCC	CCCCCCCEE	ЕЕНННННН	ННННННН	НННННННН	НННННН	
AA:	EGMQCSCGIDYYTP	HEETNNESEV 200	210	PLIVIFFCYG	QLVFTVKEA	AAQQQES 2.40	
	190	200	210	220	230	240	
Conf:	11023344457778	39999999997:	2133555899	19840688744	1224530354	3001100	
Pred:	ССНННННННН	ннннннн	HCCCCCEEEE	EEECCCCCCC	CEEEECHHH	HHCCCCC	 -

5.3.7-2 PSIPRED Result – TextView

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

🔹 POODLE	
PFRMAT DR	<u>^</u>
REMARK K. Shimizu, Y. Muraoka, S. Hirose, and T. Noguchi	
REMARK "Feature Selection Based on Physicochemical Properties of	
REMARK Redefined N-term Region and C-term Regions for Predicting Disorder"	
REMARK Proc. of IEEE CIBCB 2005, pp262-267.	
METHOD Prediction for short disorder using modified PSSM	=
METHOD	
K D 0.712	
R D 0.696	
E D 0.632	
V D 0.555	
R O D.466	
A D 0.035	
F N N 709	
S D 0.729	
R D 0.718	
R D 0.693	
K D 0.629	-

5.3.8-2 POODLE Result - TextView

### 5.3.9 TmHmm\_DTU $\rightarrow$ 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 $\rightarrow$ HtmlView

Results of SignalP\_DTU can be displayed using HtmlView node.



5.3.10-1 SignalP\_DTU execution results (HtmlView)

 $Results \ of \ ChloroP\_DTU \ can \ be \ displayed \ using \ HtmlView \ node.$ 

A Html View - 0:9	9 - HtmlView	-				
Eile						
	L: file:C:/2013-01-15	/19-17-23-0.1	0762894	92128335/	ltu-result.html	
ChloroP re	sults					
## chlorop v1.	1 prediction res	ults ######			#######	
umber of query	sequences: 1					
ame	Length	Score	cTP	-CS	cTP-	
				score	length 	
equence	118	0.509	Y	1.130	16	

5.3.11-1 ChloroP\_DTU execution results (HtmlView)

# 5.3.12 LipoP\_DTU $\rightarrow$ HtmlView

Results of LipoP\_DTU can be displayed using HtmlView node.

🔬 Html View -	- 0:13 - HtmlView			-	
<u>F</u> ile					
$\langle \rangle$	URL: file:C:/2013-01-15/19-21-	18-0.1402	2694363771	405/dtu-result.html	
LipoP re	sults				
♯ Sequence C ♯ Cut−off=-3	YT score=-0.200913 margin:	=0.9720	77		
Sequence	LipoP1.0:Best CYT	1	1	-0.200913	
Sequence	LipoP1.0:Margin CYT	1	1	0.972077	
Sequence	LipoP1.0:Class TMH	1	1	-1.17299	

# 5.3.12-1 LipoP\_DTU execution results (HtmlView)

 $Results \ of \ NetCTL\_DTU \ can \ be \ displayed \ using \ HtmlView \ node.$ 

A Html View - 0:28 - HtmlView			-	
File				
	19-24-8-0.5537612516963457/	dtu-result.html		
NetCTL results				
-				
NetCTL-1.2 predictions using MH	supertype A1. Threshol	d 0.750000		
				E
1 ID Sequence pep MSGGSSCSQ	f 0.0693 aff_rescale	0.2942 cle 0.0791 tap	-0.1830 COMB	0.2969
2 ID Sequence pep SGGSSCSQT	f 0.0555 aff_rescale	0.2355 cle 0.0293 tap	-1.0460 COMB	0.1876
3 ID Sequence pep GGSSCSQTP	f 0.0546 aff_rescale	0.2319 cle 0.0824 tap	-0.2910 COMB	0.2297
4 ID Sequence pep GSSCSQTPS	f 0.0757 aff_rescale	0.3215 cle 0.0308 tap	-2.4180 COMB	0.2052
5 ID Sequence pep SSCSQTPSR	if   0.0626 aff_rescale	0.2657 cle 0.2791 tap	1.5570 COMB	0.3854
6 ID Sequence pep SCSQTPSRA	f 0.0593 aff_rescale	0.2517 cle 0.0662 tap	-0.4120 COMB	0.2410
7 ID Sequence pep CSQTPSRAI	f 0.0635 aff_rescale	0.2694 cle 0.1013 tap	0.5560 COMB	0.3124
8 ID Sequence pep SQTPSRAIP	if   0.0528 aff_rescale	0.2241 cle 0.0299 tap	0.1900 COMB	0.2381
9 ID Sequence pep QTPSRAIPA	if 0.0798 aff_rescale	0.3387 cle 0.1018 tap	-0.5050 COMB	0.3287
10 ID Sequence pep TPSRAIPAT	f 0.0541 aff_rescale	0.2295 cle 0.2279 tap	-1.0840 COMB	0.2095
11 ID Sequence pep PSRAIPATR	if 0.0480 aff_rescale	0.2037 cle 0.7916 tap	1.1040 COMB	0.3776
12 ID Sequence pep SRAIPATRR	if 0.0573 aff_rescale	0.2432 cle 0.7824 tap	1.6160 COMB	0.4413
13 ID Sequence pep KAIPAIKKV	t U.U648 att_rescale	U.2752 cle U.7819 tap	U.734U COMB	0.4291
14 ID Sequence pep AIPAIRKVV	tt U.U582 aft_rescale	U.24/2 cle U./U38 tap	U./U4U COMB	0.3879
15 ID Sequence pep IPAIRKYVL	ft U.U548 aff_rescale	0.2325 cle 0.9593 tap	U.844U COMB	0.4186
15 ID Sequence pep PAIKKYYLG	ft U.U54U aff_rescale	0.2294 cle 0.2241 tap	-1.8500 COMB	0.1741
10 ID Sequence pep AIRKVVLQD	IT U.UOUD ATT_rescale	0.2207 CIE 0.0400 tap	-1.7890 COMB	0.1741
10 ID Sequence pep IRRVVLaba	f 0.0551 all_rescale	0.2204 CTE 0.0720 tap	-1.0070 COMB	0.1000
20 ID Sequence pep RKWYLODGV	f 0.0560 off rescale	0.2416 cle 0.0000 tap	0.8330 COMB	0.4170
21 ID Sequence pep KVVLaDGVG	f 0.0503 aff_rescale	0.2410 cle 0.0701 tap 0.2430 cle 0.9199 tap	1 3050 COMB	0.2773
22 ID Sequence pep YVEdburge	f 0.0620 aff rescale	0.2430 cle 0.3133 tap 0.2631 cle 0.2107 tap	0.0700.COMB	0.9982
23 ID Sequence pep (GDGVA) PP	f 0.0663 aff rescale	0.2814 cle 0.0468 tap	-0.2870 COMB	0.2741
24 ID Sequence pep GDGVQI PPG	f 0.0486 aff rescale	0.2062 cle 0.0470 tap	-1.9970 COMB	0.1134
25 ID Sequence pep DGVOLPPGD	f 0.0461 aff rescale	0.1958 cle 0.0315 tap	-2.3500 COMB	0.0830
26 ID Sequence pep GVQLPPGDY	f 0.1097 aff_rescale	0.4658 cle 0.9745 tap	2.7130 COMB	0.7477
27 ID Sequence pep VQLPPGDYS	f 0.0509 aff_rescale	0.2159 cle 0.0298 tap	-2.1200 COMB	0.1144
	· · · · · · ·	0.0530 1.0.5000 .		

5.3.13-1 NetCTL\_DTU execution results (HtmlView)

Results of NetChop\_DTU can be displayed using HtmlView node.

4	Htr	nl Vi	ew -	0:29 - Html	View	the second s	_ <b>D</b> _ X
Ei	le						
ſ	<		>	URL: file:C:/2	2013-01-15/19-29	-27-0.5876331447090448/dtu-result.html	
F							A
N	Jet	Ch	op	results			
17							
Ne	et Ch	C ao	.0	predictions	using version	C-term. Threshold 0.500000	
							E
						-	
۶ ۲	os	AA	C	score	Ident		
						-	
	1	M	S	0.636129	Sequence		
	2	8	·	0.062696	Sequence		
	3 1	ն Հ	•	0.140400	Sequence		
	4	ն 0	•	0.050451	Sequence		
	5	0	·	0.007000	Sequence		
	0	0	•	0.027808	Sequence		
	0	U c	÷	0.039821	Sequence		
	0	0	0	0.070000	Sequence		
	9 10	Q T	•	0.0000077	Sequence		
	10	I D	•	0.0028277	Sequence		
	12	r e	•	0.002301 n nonoto	Sequence		
	12	D	•	0.030010	Sequence		
	14	n A	•	0.278127	Sequence		
	15	н Т	•	0.000138	Sequence		
	16	P	•	0.101300	Sequence		
	17	Å	•	0.101801	Sequence		
	18	T	•	0.227901	Sequence		
	19	R	s	0.791574	Sequence		
	20	R	S	0.782428	Sequence		
	21	v	S	0.781868	Sequence		
	22	Ý	S	0.703768	Sequence		
	23	L	S	0.959304	Sequence		
	24	G		0.224106	Sequence		
	25	D		0.045504	Sequence		-

5.3.14-1 NetChop\_DTU execution results (HtmlView)

### 5.3.15 NetNES\_DTU $\rightarrow$ HtmlView

Results of NetNES\_DTU can be displayed using HtmlView node.



5.3.15-1 NetNES\_DTU execution results (HtmlView)

 $Results \ of \ NetPhosK\_DTU \ can \ be \ displayed \ using \ HtmlView \ node.$ 

🔬 Htm	l View - 0:	34 - HtmlViev		×
<u>F</u> ile				
$\left[ \right]$		L: file:C:/2013	-01-15/19-36-50-0.00723448796424031/dtu-result.html	
		•.		
Netl	PhosK	results		
Method	l: NetPhos	K without E	SS filtering:	
Query:	Sequenc	е		
Site	Kinase	Score		
S-2	CKI	0.52		
S-5	cdc2	0.55		
S-6	PKC	0.61		
S-8	DNAPK	0.60		
T-10	p38MAPK	0.51		
T-10	cdk5	0.68		
T-18	PKC	0.69		=
T-37	p38MAPK	0.54		
T-37	cdc2	0.54		
T-37	cdk5	0.54		
T-45	cdc2	0.55		
T-46	p38MAPK	0.50		
T-46	cdk5	0.57		
T-70	p38MAPK	0.51		
T-70	cdk5	0.68		
8-83	cdk5	0.63		
S-85	CKII	0.62		
S-94	DNAPK	0.64		
S-94	ATM	0.61		
S-96	PKC	0.75		
8-101	GSK3	0.50		
8-112	CKII	U.65		
8-112	DNAPK	U.55		
9-115	AIM	0.62		-

5.3.16-1 NetPhosK\_DTU execution results (HtmlView)

Results of NetPhos\_DTU can be displayed using HtmlView node.

Truth view -	0:35 - 1	HtmlView					×
<u>F</u> ile							
$\langle \rangle$	URL: file	e:C:/2013-01-1	5/19-39-0	)-0.02101	7865753876763/dtu-result.html		
NetPhos	resul	ts					<u>^</u>
118 Sequend	ce						
MSGGSSCSQTPSF	RAIPATRE	\VVLGDGVQLPP	GDYSTTP	GGTLFST	TPGGTRIIYDRKFLMECRNSPVTKTPPRDLPTIPG	80	
VTSPSSDEPPME/	ASQSHERM	ISPEDKRAGGEE	SQFEMDI			160	
S.S.T	T.		YT.		ΤΤ	80	E
	s	.s	s			160	
Phosphorylati	ion site	es predicted	:	Ser:	9 Thr: 5 Tyr: 1		
	Seri	ne predicti	ons				
Name	Pos	Context v	Score	Pred			
Sequence	2	MSGGSS	0.011				
Sequence	5	MSGGSSCSQ	0.162				
Sequence	6	SGGSSCSQT	0.944	*S*			
Sequence	8	GSSCSQTPS	0.511	*S*			
Sequence	12	SQTPSRAIP	0.005				
Sequence	35	PGDYSTTPG	0.211				
Sequence	44	GTLFSTTPG	0.055	•			
Sequence	65	ECRNSPVTK	0.958	*8*			
Sequence	83	PGVTSPSSD	0.991	*8*			
Sequence	85	VTSPSSDEP	0.987	*S*			
Sequence	86	TSPSSDEPP	0.982	*S*			
Sequence	94	PMEASQSHL	0.084	•			
Sequence	96	EASQSHLRN	0.644	*S*			
Sequence	101	HLRNSPEDK	0.998	*S*			
Sequence	112	GGEESQFEM	U.862	*8*			

5.3.17-1 NetPhos\_DTU execution results (HtmlView)

Results of NetPicoRNA\_DTU can be displayed using HtmlView node.

🛓 Html View - 0:3	6 - Htm	lView			-	_ <b>D</b> X
<u>F</u> ile						
	file:C:,	/2013-01-1	5/19-46-1	2-0.0481980993836937	9/dtu-result.html	
NetPicoRN	A re	sults				<u>م</u>
>Sequence MVCLRLPGGSCMAVLT YNQEEYVRFDSDVGEF QRRVHPKVTVYPSKTC DWTFQTLVMLETVPRS FLGAGLFIYFRNQKGH	VTLMVL: RAVTEL PLQHHNI GEVYTC SGLQPR	SSPLALAGI GRPDEEYWI LLVCSVSGI QVEHPSVT: GFLS	DTRPRFLE NSQKDFLE FYPGSIEV SPLTVEWR	YSTSECHFFNGTERVRF DRRAAVDTYCRHNYGVG RWFRNGQEEKTGVVSTG ARSESAQSKMLSGVGGF	LDRYF ESFTV LIHNG VLGLL	
You have selecte - 2A prediction - 3C prediction - 3CFMDV predict - auto predictio	d the ' ion n	followin.	g option:	s:		
<i></i> 2Apro prediction	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,			****	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	
OUTPUT:						
Residue	Pos	Clv	Surf	Sequence	Comment	
G A	8 29	0.049 0.047	0.695 0.468	CLRLPGGSCMA SPLALAGDTRP		
N V	48 73	0.038 0.092	0.653 0.393	ECHFFNGTERV RFDSDVGEFRA		-

5.3.18-1 NetPicoRNA\_DTU execution results (HtmlView)

 $Results \ of \ DictyOGlyc\_DTU \ can \ be \ displayed \ using \ HtmlView \ node.$ 

🔬 Html V	/iew - 0:11	HtmlVie	W				_ 0	×
<u>F</u> ile								
	> URL:	file:C:/201	3-01-15/19-5	53-54-0.6088173	1377901589/dtu-r	esulthtml		
Dicty	OGlvc	result	s					<b>^</b>
			-					-11
Nomo · S	oquonco	L	anath• 11	2				
MSGGSSCS	QTPSRAIPA		GVQLPPGDYS:	, ITPGGTLFSTTP	GGTRIIYDRKFLN	ECRNSPVTKTPPRDLPTIPG	80	
VTSPSSDE	PPMEASQSH	LRNSPEDK	RAGGEESQFE	MD I				
	G					G	80	
Name	Residue	Number	Potential	Threshold	Assignment			E
Sequence	Ser	0002	0.0231	0.3937	•			
Sequence	Ser	0005	0.0926	0.4637	•			
Sequence	Ser	0006	0.0743	0.4726	•			
Sequence	Ser	0008	0.3551	0.4/16	•			
Sequence	Inr	0010	0.0554	0.4577				
Sequence	აer Thr	0012	0.0709	0.40// 0.5606	u			
Sequence	Inr	0010	0.0193	0.0000	•			
Sequence	oer Thr	0035	0.2023	0.4407	•			
Sequence	Thr	0037	0.0385	0.4796	•			
Sequence	Thr	0041	0.0585	0.5576				
Sequence	Ser	0044	0.2398	0.5646				
Sequence	Thr	0045	0.0229	0.5496				
Sequence	Thr	0046	0.0287	0.5286				
Sequence	Thr	0050	0.2696	0.5735				
Sequence	Ser	0065	0.2373	0.4836				
Sequence	Thr	0068	0.2909	0.4237				
Sequence	Thr	0070	0.5861	0.4037	G			
Sequence	Thr	0077	0.0833	0.4916				
Sequence	Thr	0082	0.2337	0.4786	•			
Sequence	Ser	0083	0.3293	0.4756	•			-
C	C	0005	0 1500	0 4057				

5.3.19-1 DictyOGlyc\_DTU execution results (HtmlView)

### 5.3.20 WolfPsort\_AIST $\rightarrow$ HtmlView

Results of WolfPsort\_AIST can be displayed using HtmlView node.



5.3.20-1 WolfPsort\_AIST execution results (HtmlView)

 $Results \ of \ Target P\_DTU \ can \ be \ displayed \ using \ HtmlView \ node.$ 

🛓 Html View - 0:	40 - HtmlView	_						×
<u>F</u> ile								
	RL: file:C:/2013-01-1	5/19-57-50-0	.01006785	95121699	3∕dtu-i	result	html	
TargetP re	sults							 
### targetp v1. Number of query Cleavage site p Using NON-PLANT	1 prediction re vsequences: 1 predictions inclu networks	sults #####	######			#####	######	
Name	Len	mTP	SP	other	Loc	RC	TPlen	
Sequence	266	0.064	0.935	0.026	S	1	29	
cutoff		0.000	0.000	0.000				

5.3.21-1 TargetP\_DTU execution results (HtmlView)

Results of SecretomeP\_DTU can be displayed using HtmlView node.

A Html View -	0:38 - HtmlView				
File					
$\langle \rangle$	URL: file:C:/2013-0	-15/19-59-54-0.3048	13015146292736/dtu-re	esulthtml	
Secreton	neP results				
lon-classica	lly secreted prot	eins should obt	ain an NN-score e	exceeding	
he normal ti o contain a	hreshold of <b>0.5</b> , cignal poptida	but not at the	same time be pre	dicted	
letwork 1	Network 2	Network 3	SecP score	Sequence name	
747061	n q8n <i>44</i> q	 N 745926	 П 82 <b>44</b> 79	Sequence	
	0.00010	0.140020	01027110	Soquence	

5.3.22-1 SecretomeP\_DTU execution results (HtmlView)

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

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/

File	
MEMSAT Result	TextView
Quer	y
Transmembrane (TM)	
Outside helix cap of TM	
Outside sequence	
Inside sequence	
>Querv	
MNGTEGPNEYVPESNKTGVVRSPEEAPOYYLAEPWOE	SMLAAYMELLIMI GEPINELTLY
GETAL WSL VVL A TERVVVVCKPMSNEREGENHA TMCV	AFTWAMAL ACAAPPL VOWSRYTP
ENUNKTLIEELKILKULYSNKMSEEGPUVKIREASKL	JNVDFTLSNVDLAMANSERRVMTA
EIPTLAIDSVEVETNTTVLADEFIAHRLGLIPLQSMD	)IEQLEYSRDCFCEDHCDKCSVVL
TLQAFGESESTTNVYSKDLVIVSNLMGRNIGHPIIQD	KEGNGVLICKLRKGQELKLTCVA
KKGIAKEHAKWGPAAAIEFEYDPWNKLKHTDYWYEQD	)SAKEWPQSKNCEYEDPPNEGDPF
DYKAQADTFYMNVESVGSIPVDQVVVRGIDTLQKKVA	SILLALTQMDQDKVNFASGDNNT
ASNMLGSNEDVMMTGAEQDPYSNASQMGNTGSGGYDN	IA₩

5.3.24-1 AISTViewer – Memsat Result

🛃 MEMSAT
seqfile-nn
699 residues read from file.
$u_{\rm plix}$ 1 from 252 (in) to 278 (out) • 25001
Score = 25.881000
Helix 1 from 38 (out) to 62 (in) : 30467
000000XXXXXXXXXXXXXIIIIII++++++++++
Score = 44.105000
Helix 1 from 4 (in) to 23 (out) : 4294962640
Helix 2 from 34 (out) to 54 (in) : 20607
Helix 3 from 57 (in) to 75 (out) : 4294962310
Helix 4 from 78 (out) to 96 (in) : 25178
Helix 5 from 99 (in) to 118 (out) : 10411
Helix 6 from 121 (out) to 140 (in) : 18157
Helix 7 from 152 (in) to 171 (out) : 18411
Helix 8 from 197 (out) to 221 (in) : 25851
Helix 9 from 224 (in) to 242 (out) : 4294961946
Helix 10 from 245 (out) to 263 (in) : 962
Helix 11 from 266 (in) to 284 (out) : 7003
Helix 12 from 287 (out) to 306 (in) : 11757
Helix 13 from 330 (in) to 348 (out) : 4294960008
Helix 14 from 351 (out) to 370 (in) : 4294959329
• <u> </u>

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.

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/	

Please visit below web sites for further information.



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

No	Name	Icon	Set	Description
1	FastaFileReader	FastaFileReader	Y	Read
		首		Multi-FASTA file.
		prepare Multi-FASTA		
2	Mafft_AIST	Mafft_AIST	Y	Execute Mafft.
		► <mark>=</mark>		
		Execute MAFFT via SOAP		
3	ClustalW_AIST	ClustalW_AIST	Y	Execute ClustalW.
		► <mark>==</mark> ►		
		Execute CLUSTALW via SOAP		
4	PhylogeneticTree_AIST	PhylogeneticTree_AIST	Y	Execute
		<mark>►</mark> ត្រា ►		Phylogenetic Tree.
		Create a phylogenetic tree		

5	AISTViewer	AISTViewer		Display prediction
				results.
		Display a multiple alignment		
6	PhylogeneticTreeView	<b>PhylogeneticTreeView</b>		Display
		► <b>U</b>		Phylogenetic Tree.
		Execute Archaeopteryx		
7	LSDBCrossSearch	LSDBCrossSearch	-	Execute LSDB
		► ISCE		cross search.
		LSDB cross search		

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.

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

6.2.2.1-1 FastaFileReader : Configure...

```
• Options \rightarrow Fasta File \rightarrow SelectedFile:
```

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

Click "OK" button after setting the conditions.

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.

🔥 Dialog - 3:11 - ClustalW_AIST (Execute CLUS 🗕 🗖 🗙
File
Options Flow Variables Memory Policy Type PROTEIN O DNA
Select Output Directory Selected Directory: C:¥
OK Apply Cancel ?

6.2.2.2-1 ClustalW\_AIST : Configure...

#### • Options $\rightarrow$ Type

Choose either PROTEIN or DNA based on query sequence type.

#### • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

Click "OK" button after setting the conditions.

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.

🛆 Dialog - 3:2 - Mafft_AIST (Execute MAFFT via SOAP) 🛛 🗖 💌
File
Options Flow Variables Memory Policy Select Output Directory Selected Directory: C.¥
Advanced Options retree 2maxiterate 0bl 62op 1.53ep 0.0clustalout
OK Apply Cancel

6.2.2.3-1 Mafft\_AIST : Configure...

#### • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

### $\boldsymbol{\cdot}$ Options $\rightarrow$ Advanced $\rightarrow$ 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		: 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.

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.

🔥 Dialog - 3:4 - PhylogeneticTree_AIST (Create 🗕 🗖 💌
File
Options Flow Variables Memory Policy Select Output Directory C# Browse
Phylogenetic Tree Methods NJ UPGMA BOOTSTRAP (valid for "NJ On On Off Number of BOOTSTRAP 1000
OK Apply Cancel 🖓

6.2.2.4-1 PhylogeneticTree\_AIST : Configure...

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory:

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

#### • Options $\rightarrow$ Phylogenetic Tree

Specify types of phylogenetic tree (blue open rectangular).

 $\boldsymbol{\cdot}$  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.

### 6.3.1 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

S ClustalW	
>2_gi 21406208 gb AY087471.1 _A	^
	=
 ^^^T^^^^^^^^^^^^^^^^^^^^^^	
TAPPTCTGAATTTCTCGTTGTTGGAACAATGGCGTCGAATCTCCTCGAAAG	
GAATIGCTGGTCTCTCTCCCCTTTTATTGC-ATTITTAAGTT	
TGIGTATIGITITITICIGGIGIGCCTACTACATCTICAGCTATATIATC	
TAATAAAGGATICGATCAAAGTCGGGTAAGTITGATTTTGTTTGATCTC	
ACTICAGCACTIGICATGIIGIAACATICAATCICIGATATCACTGIYII	
II	

6.3.1-2 ClustalW Result – TextView

#### $6.3.2 \text{ Mafft} \rightarrow \text{AISTViewer}$

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

No	Name	Icon	Set	Description
1	SetVariable	SetVariable	Y	Control CASE
		012		Switch Variable.
		V		Open Dialog:
				0: first outport
		Node 1		1: second outport
				2:third outport
2	CASE Switch	CASE Switch		Control CASE
	Variable(Start)	Variable (Start)		Switch using
		● case ≠V		workflow
				variables.
		Node 2		
		Node 2		
3	FastaFileReader	FastaFileReader	Y	Set FASTA file.
		<b>₿</b> - ►		
		0.0		
		Node 3		

7.2.1-1 Molecular Simulation Active Workflow node list
4	LSDBCrossSearch	LSDBCross Search		Execute LSDB
		► IST		cross search.
		Node 4		
5	Sparal AIST	Coord AICT	Y	Execute SPARQL
		Sparqi_AIST		against AIST
		103 ►		endpoints.
		C 900N		~ -
6	SequenceSelector	Sequence Selector		Select a sequence
		► <mark></mark> ►		results.
		Node 6		
7	CASE Switch Variable	CASE Switch		CASE Switch end
	(End)			node.
		Node 7		
8	BlastForModeller_AIST		Y	Execute BLAST
				(PSI-BLAST).
		► <mark></mark> ►		
		Node 9		
0	HitPogion Coloctor AIST	Node 8	v	Detect hit regions
9	THIRegionSelector_AIS1	HitRegion Selector_AIST	1	from BLAST
		► <mark>ਚ<sup>®</sup>≂</mark> ►		(PSI-BLAST)
				execution results.
		Node 9		
10	TemplateSelector_AIST	Template Selector_AIST	Y	Choose a
				modelling
				tempiate.
		Node 10		

11	Modeller_AIST	Modeller_AIST	Y	Execute
		▶ <mark>- f≩</mark> ▶		MODELLER.
		Node 11		
12	DockingTemplateSelector	DockingTemplateSelector	Y	Select a template
		▶ 🔳 ▶		structure from
				results.
		Node 12		
13	PdbFileReader	PdbFileReader	Y	Read PDB file.
		Node 13		
14	fpocket2_AIST	fpocket2_AIST	Y	Execute fpocket2.
		► <mark>E</mark> T►		
		Node 14		
15	PocketSelector	PocketSelector		Select a pocket
		▶   ►		site from
				ipocket2 results.
		Node 15		
16	AutoDock_AIST	AutoDock_AIS1	Y	Execute
				AutoDock.
		Node 16		
17	CompoundQuery_AIST	CompoundQuery_AIST	Y	Execute chemical
		32 ×		ligands search.
		Node 17		

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

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

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.

	Dialog - 0:75 - SetVariable 🛛 🗖 🗙
File	
Options Flow Specify an acti Active flow	Variables Memory Policy ve flow variable output port (0:first; 1:second; 2:third) variable ouput port number (integer: 0 to 2): 0 ÷
ОК	Apply Cancel

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.

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.

🛓 Dialog - 0:1 - FastaFileReader	- • ×
File	
Options Maximum Balling	
□ Fasta File	
Selected File:	
▼ Browse	
OK - Execute Apply Cance	

7.2.2.2-1 FastaFileReader : Configure...

```
• Options \rightarrow Fasta File \rightarrow SelectedFile:
```

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

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.

🛆 Dialog - 2:60 - Sparql_AIST 🗧 🗖 💌		
File		
Options Advanced Flow Variables Memory Policy		
Output Directory		
Selected Directory:		
C.¥		
SPARQL endpoints		
SEVENS endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/s		
fRNAdb endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/		
UNIPROT endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/u		
PDB endpoint: http://pdb.bio2rdf.org/sparql		
KEGO - anthuru - andraint http://kagg.bio?xdf.org/oparal		
SPARQL search conditions		
Taxon (not available for UNIPROT)		
Kewword		
Minimum sequence length 300		
Maximum sequence length 600		
Resolution (for PDB) 2.8		
Pathway (for KEGG-pathway)		
Output format		
Output format ('FASTA' for 'SequenceSelector nod		
FASTA O Tab-delimited		
OK Apply Cancel ?		

7.2.2.3-1 Sparql\_AIST : Configure...

#### • Options $\rightarrow$ Output $\rightarrow$ Selected Directory:

Specify an output directory in a text box or "Browse".

#### • Options $\rightarrow$ SPARQL endpoints:

Endpoints: SEVENS, fRNAdb, UNIPROT, PDB and KEGG-pathway Check SEVENS, UNIPROT and/or PDB in this workflow.

#### • Options $\rightarrow$ 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 $\rightarrow$ Output format:

Specify FASTA or Tab-delimited radio buttons.

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.

🔥 Dialog - 2:2 - BlastForModeller_AIST 🚽 🗖 💌
File
Options Flow Variables Memory Policy
BLAST version 2.2.18 Execution Type BLAST O PSI-BLAST
E-Value 1.0E-5 Interation 3
Select Output Directory Selected Directory: C:¥
OK Apply Cancel ?

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  $\rightarrow$  BLAST version 2.2.18  $\rightarrow$  E-Value

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

## • Options $\rightarrow$ BLAST version 2.2.18 $\rightarrow$ Iteration

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

#### • Options $\rightarrow$ Select Output Directory $\rightarrow$ 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.

- Right-clicking on the HitRegionSelector\_AIST node and specify "Configure...".
- 2) Open the below pop-up window and set execution conditions.

🝐 Dialog - 2:3 - HitRegionSelector_AIST 🛛 🗕 💌
File
Options Flow Variables Memory Policy
Conditions to select (PSI-)BLAST hit regions (Integer is only permitt
Coverage (%) 60
Identity (9) 20
Literative (AV) 30
Minimum Length 30
OK Apply Cancel 🕜

7.2.2.5-1 HitRegionSelector\_AIST : Configure...

• Options  $\rightarrow$  Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input)  $\rightarrow$  Coverage(%)

Specify Coverage (%) (default: 60).

• Options  $\rightarrow$  Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input)  $\rightarrow$  Identity(%)

```
Specify Identity (%) (default: 30)
```

# • Options $\rightarrow$ Condition to select (PSI-)BLAST hit regions (Integer is only permitted to input) $\rightarrow$ Minimum Length

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

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.

🝐 Dialog - 2:4 - TemplateSelector_AIST 🛛 🗖 💌
File
Options Flow Variables Memory Policy Conditions to determine for modelling or for diplaying PDBj Mine Web.
Identity (%) 90
OK Apply Cancel 🕐

7.2.2.6-1 TemplateSelector\_AIST : Configure...

# • Options $\rightarrow$ Condition to determine for modelling or for displaying PDBj Mine Web. $\rightarrow$ 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.

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.

△ Dialog - 2:5 - Modeller_AIST - □	x
File	
Options Flow Variables Memory Policy Condition for Modeller Execution Number of Models for Modelling 5	
Modeller License License Key for Modeller (required)	]
OK Apply Cancel 🕐	

7.2.2.7-1 Modeller\_AIST : Configure...

 $\boldsymbol{\cdot}$  Options  $\rightarrow$  Condition for Modeller Execution  $\rightarrow$  Number of Models for Modelling

Specify maximum number of models (default: 3) Range:  $1 \sim 10$ 

• Options  $\rightarrow$  Modeller License  $\rightarrow$  License Key for Modeller (required) Input a MODELLER license key (required).

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.

📥 Dialog - 4:62 - CompoundQuery_AIST (Nod 🗕 🗖 💌
File
Options Output Directory Flow Variables Memory Policy Select Database
Database Namiki
Search Words
✓ Molecular Weight from 300.0 to 305.0
✓ log P from 1.0 to 2.0
TPSA from 80.0 to 140.0
smiles
inchikey
Number of rotatable bonds from 3 to 7
Charge from 0 to 2
H-bond Acceptor from 3 to 8
H-bond Donor from 8 to 8
Number of rings from 1 to 5
Search Conditi
OK Apply Cancel 🕐

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:  $1.0 \sim 2.0$ ).

#### 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).

File
Options Output Directory Flow Variables Memory Policy
Selected Directory:
C.¥
OK Apply Cancel 🕐

7.2.2.8-2 CompoundQuery\_AIST : Configure...

# • Options $\rightarrow$ Select Output Directory $\rightarrow$ Selected Directory:

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

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

Δ.	Dialog - 4:111 - Autol	Dock_AIST (By specific site)	- 🗆 🗡
File			
Options Flow Variables Memory Policy			
Secity binding site coordinate (X, Y, Z)			
✓ use	x-coordinate 0.0	y-coordinate 0.0	z-coordinate 0.0
Output	Selected Directory		
	C¥	V Browse	
		N. N.	
jp.cbrc.molf.activeflow.soap.autode	ock.coorx", "jp.cbrc.molf.activeflow.soap.auto	dock.coory" and "jp.cbrc.molf.activeflow.soap.a	autodock.coorz" are controlled by variables.
		OK	Apply Cancel (?)

## 7.2.2.9-1 AutoDock\_AIST : Configure...

## • Options $\rightarrow$ XYZ-coordinate

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

• Options  $\rightarrow$  Select Output Directory  $\rightarrow$  Selected Directory:

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

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

🔺 Dialog - 4:65 - MMPrep_AIST (Node 23) 🛛 🗕 🗙
File Options Flow Variables Memory Policy Output Selected Directory
C¥
OK Apply Cancel

7.2.2.10-1 MMPrep\_AIST : Configure...

• Options  $\rightarrow$  Output  $\rightarrow$  Selected Directory :

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

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

🔥 Dialog - 4:66 - InitMinMM_AIST (Node 24) 🛛 🗖 💌
File Options Flow Variables Memory Policy Output Selected Directory. Ct
OK Apply Cancel 🕡

7.2.2.11-1 InitMinMM\_AIST : Configure...

• Options  $\rightarrow$  Output  $\rightarrow$  Selected Directory :

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

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

🔥 Dialog - 4:67 - MoltrecMD_AIST (Node 25) 🛛 🗕 💌
File  Options Flow Variables Memory Policy  Output  Selected Directory:  C.¥  Browse
OK Apply Cancel 🕡

7.2.2.12-1 MoltrecMD\_AIST : Configure...

• Options  $\rightarrow$  Output  $\rightarrow$  Selected Directory :

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

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.

🔥 Dialog - 4:76 - ResultPathSetter (Node 32) 🛛 🗖 💌				
File				
Options Flow Variables Memory Policy				
Specify an absolute path of directory stored prediction results				
Browse				
Select a KNIME node name corresponding to the specified directory path				
KNIME nodes Ammos_AIST 🗸 🗸				
Sequence name (for Poodle_AIST and PsiPred_AIST)				
query				
OK Apply Cancel (?)				

7.2.2.13-1 ResultPathSetter : Configure...

## $\cdot$ Options

## ${\bf Selected} \ {\bf Directory}:$

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).

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

🔥 Dialog - 4:110 - Ammos_AIST (Node 36) 🛛 🗕 🗖 💌	
File Options Flow Variables Memory Policy Select Output Directory C* Browse	
OK Apply Cancel 🕐	

7.2.2.14-1 Ammos\_AIST : Configure...

## • Options $\rightarrow$ Output $\rightarrow$ Selected Directory :

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

- 1) Right-clicking on the fpocket2\_AIST node and specify "Configure...".
- 2) Open the below pop-up window and set execution conditions.

🔥 Dialog - 4:107 - fpocket2_AIST (Node 13) 🛛 🗖 💌
File
Options Flow Variables Memory Policy Select Output Directory
Selected Directory.
C.¥
OK Apply Cancel

7.2.2.15-1 fpocket\_AIST : Configure...

## • Options $\rightarrow$ Output $\rightarrow$ Selected Directory :

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

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.

△ Dialog - 4:102 - PdbFileReader (Node 37) -	×
File	
Options Flow Variables Memory Policy	
Selected File:	
V Brow	se
OK Apply Cancel 🤇	
Options Flow Variables Memory Policy PDB File Selected File: OK Apply Cancel ?	30

7.2.2.16-1 PdbFileReader : Configure...

## • Options $\rightarrow$ PDB File $\rightarrow$ SelectedFile:

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

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.

🙏 Dialog - 4:109 - Mol2FileReader (Node 35) 🛛 🗕 💌
File
Options Flow Variables Memory Policy
Selected File:
✓ Browse
OK Apply Cancel ၇

7.2.2.17-1 Mol2FileReader : Configure...

## • Options $\rightarrow$ MOL2 File $\rightarrow$ SelectedFile:

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

# 7.3.1 BlastForModeller\_AIST $\rightarrow$ 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 $\rightarrow$ HtmlView

Results of HitRegionSelector\_AIST can be displayed using the HtmlView node.

🛓 Html V	/iew - 0:9 -	HtmlView					
<u>F</u> ile							
	> URL:	file:C:/work/KNI	1E/prg/testData/outdir	/2011-10-27/17-3:	2-361830042473/dom	split.log	
PDB a) 1205A	Query Hit PDB Cover 248 3	Length(aa) age(%) PDB Hi 54.29 497-74	Query Coverage ( t Range (aa) 4 246 62.12	%) Query Identity(%) 135-380 26.77	Hit Range(aa) E-value 7.10235e-19	PDB Hit 1	length (a 🔺

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.

🗻 PDBj Mine - 0:7 - PDBjMineWeb	<b>_ D _ X</b>				
Eile					
A similar region to an existing PDB structure was found.					
PDB code: Itw/C hit region: 384-649					
Open PDBj Mine Web					

7.3.3-1 PDBjMineWeb – PDBj Mine

O PDBj Mine Summar	ry Pag × 🕒		
← → C 🕓 servi	ice.pdbj.org/mine/Detail?PDBID=	=1twf&PAGEID=Summary	y 🖒
PDB	C CLARTER	2250997	REAL REAL PROPERTY
THE CAR AND THE CAR	ound have	A ADDALOGO	approver you
English			Statistics Help FAQ Contact Us
Home	PDBj (Protein Data Bank Japan) m	naintains a centralized PDB at	inchive of macromolecular structures and
Data Deposition >>		XD III OON and the r DDC 2	
ADIT: PDB Deposition	Miñe	Summary	/ [1twf]
ADIT-NMR			
Search >>		Constant Protection Constant	to total Common Nationham Down
Search PDB (Mine/xPSSS)	Summary Structural Details	Jerimental Details	nal Details Sequence weignbor Down
PDB/RDF, chem_comp/RDF			
Latest Release Search	<asymmetric unit=""></asymmetric>		1twf sequence information (FAST
Sequence- Navigator	See 2		download PDB format file
Structure-	a the top	RELATED FUBIU	DNA-directed RNA polymerase II
SeSAW	No.		(E.C.2.7.7.6), DNA-directed RNA
Ligand Binding Sites (GIRAF)			kDa polypeptide (E.C.2.7.7.0), Dr polymerase II 45 kDa polypeptide
EM Navigator			polynentide (F C.2.7.7.6), DNA-d
Search NMR Data (BMRB)	179	Descriptor	polymerases I, II, and III 23 kDa polymerases I, II, and III 24 kDa polymerases I, II, and II, and II, and III 24 kDa polymerases I, and II, a
Status Search	More images	Descriptor	and III 14.5 kDa polypeptide (E.C
Service and Software >>	View in 3D molecule viewer		directed RNA polymerase II 14.2 (E.C.2.7.7.6), DNA-directed RNA
JV: Graphic Viewer	( <b>JV4</b> / <b>JMOI</b> ) *1		and III 8.3 kDa polypeptide (E.C.2 directed RNA polymerase    13.6
Yorodumi			(E.C.2.7.7.6), DNA-directed RNA
Protein Globe			and III 7.7 kDa polypeptide (E.C.2
ASH		Title	RNA polymerase II complexed with
MAFFTash		This	resolution
SEALA			TRANSCRIPTION, MRNA, MULT
Structure Prediction >>		Functional Keywords	COMPLEX, MOLECULAR MAG
CRNPRED		Biological source	Saccharomyces cerevisiae (bake
Spanner SFAS		Cellular location	[UNP - P04050] Nucleus [UNP - P20435] Cytoplasm
Derived database			[UNP - P22139] Nucleus, nucleolu

7.3.3-2 PDBjMineWeb – PDBj Mine

# 7.3.4 Modeller\_AIST $\rightarrow$ JmolForModeller

🔔 Results - 0:6 - JmolForModeller	
File	$\sim$
Results	~ ~ ~
Sequence Region (aa): 1-326 Model 1 : Objective Function = 14423853 Model 2 : Objective Function = 14659855 Model 3 : Objective Function = 1478.6036 Model 4 : Objective Function = 1497.5812 Model 5 : Objective Function = 1760.7408 Sequence Region (aa): 327-381 Model 1 : Objective Function = 221.4748 Model 2 : Objective Function = 235.2573 Model 3 : Objective Function = 260.7991 Model 4 : Objective Function = 285.5612 Model 5 : Objective Function = 548.8173 Sequence Region (aa): 384-649	Jmol
Execute Jmol	56¥domain_1¥ali/usersequence.1-326

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 $\rightarrow$ 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 $\rightarrow$ JmolForModeller

Results - 0:334 - JmolForModeller	
File	
Results	
Init Min MM:                Model: 1          Score: -10.6764000000002            Model: 2          Score: -11.6439999999996            Model: 3          Score: -8.18949999999998            Model: 4          Score: -14.0838            Model: 5          Score: -10.1937            Model: 6          Score: -10.164400000002            Model: 7          Score: -10.164400000002            Model: 7          Score: -10.2748999999998            Model: 8          Score: -11.539299999997            Model: 9          Score: -12.6345000000001	
Execute Jmol	Jmol ¥1/PLpdb

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 $\rightarrow$ 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  $\rightarrow$  JmolForModeller, HtmlView

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

# 7.3.9 fpocket2\_AIST $\rightarrow$ JmolForModeller

Results of fpocket2\_AIST can be displayed using JmolForModeller node.



Pocket X-coordinate Y-coordinate Z-coordinate

7.118

33.766

47.385

29.645

24.018

44.893

**10.19**4

20.618

<u>\$</u>

**SO1** 

**SO**2

**SO**3

**SO**4

**SO**5

**SO**9

-3.060

-1.532

9.036

-0.178

5.684

2.260

-10.532

-5.248



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.

- - X

3.562

-3.714

-0.620

-0.247

10.578

16.355

7.775

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

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

## 8.1 Nodes

# 8.1.1 Node list

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

# 8.1.1-1 SPARQL node list

# 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.

	Dialog - 4:104 - Sparql_AIST (Node 30) 🛛 🗖 🗖	x
File		_
Options Ad	wanced Flow Variables Memory Policy	
Output Dire	interior for for a label of herios	
Se	elected Directory.	
	∑¥	
SPARQL en	ndnoints	
SEV	VENS endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/s	
	NAdb endpoint: http://tgrat.medais.jp/openrat-sesame/repositories/s	
🗌 🗆 UNI	IPROT endpoint: http://tgrdf.medals.jp/openrdf-sesame/repositories/	
PC	DB endpoint: http://pdb.bio2rdf.org/sparql	
□ KEGG	a - pathway endpoint: http://kegg.bio2rdf.org/spargl	1
-SPADOL	and any dataset	
-SPARQL se	Taxon (not available for LINIPROT)	
-SPARQL se	Taxon (not available for UNIPROT)	
⊂SPARQL se	Taxon (not available for UNIPROT)	
-SPARQL se	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300	
-SPARQL se	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600	
SPARQL se	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8	
SPARQL se	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8 Pathway (for KEGG-pathway)	
SPARQL se	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8 Pathway (for KEGG-pathway)	
- SPARQL se	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8 Pathway (for KEGG-pathway) sat	
Ouput iom	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8 Pathway (for KEGG-pathway) Not Output format ('FASTA' for 'SequenceSelector nod © FASTA Tab-delimited	
Oupur form	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8 Pathway (for KEGG-pathway) Not Output format ('FASTA' for 'SequenceSelector nod FASTA Tab-delimited	
Ouiput form	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8 Pathway (for KEGG-pathway) Pathway (for KEGG-pathway) Nat Output format ('FASTA' for 'SequenceSelector nod Image: FASTA Tab-delimited	
Ouput form	earch conditions Taxon (not available for UNIPROT) Keyword Minimum sequence length 300 Maximum sequence length 600 Resolution (for PDB) 2.8 Pathway (for KEGG-pathway) Nat Output format ('FASTA' for 'SequenceSelector nod © FASTA Tab-delimited OK Apply Cancel ?	

8.1.2.1-1 Sparql\_AIST : Configure...


8.1.2.1-2 Sparql\_AIST : Configure...

## • Options $\rightarrow$ Output Directory:

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

## • Options $\rightarrow$ 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  $\rightarrow$  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 $\rightarrow$ 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 I	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.

🛆 Dialog - 4:128 - Sparql_AIST_Adv 况 🗖 🗙
File
Options Flow Variables Memory Policy Output Directory Selected Directory: C:¥
SPARQL endpoint endpoint SPARQL
Input SPARQL Query
SELECT * WHERE {
OK Apply Cancel 🕐

8.1.3.1-1 Sparql\_AIST\_Adv : Configure...

### • Options $\rightarrow$ Output Directory:

Specify an output directory (red open rectangular).

#### • Options $\rightarrow$ SPARQL Endpoints:

Input a public SPARQL endpoint (blue open ractangular).

### • Options $\rightarrow$ 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.

SPARQL Results	_	
elect a row, then click SUBMIT button.	Submit	Cancel
ID	Description	Sequence
BRC-HSAP-02-0037	Adrenergic receptors	MDHQDPYSVQATAAIAAAITFLILFTIFGNALVILAVLTSRSLR
BRC-HSAP-04-0004	Adrenergic receptors	MASPALAAALAVAAAAGPNASGAGERGSGGVANASGASW
BRC-HSAP-05-0044	Adrenergic receptors	MGQPGNGSAFLLAPNGSHAPDHDVTQERDEVWVVGMGI
BRC-HSAP-05-0047	Adrenergic receptors	MNPDLDTGHNTSAPAHWGELKNANFTGPNQTSSNSTLP
BRC-HSAP-08-0018	Adrenergic receptors	MVFLSGNASDSSNCTQPPAPVNISKAILLGVILGGLILFGVL
BRC-HSAP-10-0036	Adrenergic receptors	MGSLQPDAGNASWNGTEAPGGGARATPYSLQVTLTLVCL
BRC-HSAP-10-0037	Adrenergic receptors	MGAGVLVLGASEPGNLSSAAPLPDGAATAARLLVPASPPA
BRC-HSAP-20-0001	Adrenergic receptors	MTFRDLLSVSFEGPRPDSSAGGSSAGGGGGSAGGAAPS

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.

## 8.1.5.1 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.

🛓 name of first view - 0:4 - LSDBCrossSearch (Node 4)	3		
Eile			
FASTA Header Lists			
1: >gi 334185880 ref NM_001203122.1 Arabidopsis thaliana RIO kinase 2: >gi 21406208 gb AY087471.1 Arabidopsis thaliana clone 35785 mRNA 3: >gi 28416578 gb BT004574.1 Arabidopsis thaliana At3g03070 gene, 4: >gi 13358228 gb AF325039.2 Arabidopsis thaliana AT3g03070 (AT3g0 5: >gi 110735932 dbj AK227975.1 Arabidopsis thaliana mRNA for hypot			
LSDB Cross Search			
Search Words : LSDB Cross Search			
How to input search keywords			
AND:->'space' e.g. network socket			
OR ->' ' eg. network∣socket			
XOR → '!' eg. network ! socket			
Wild card: -> '*' e.g. inter*, *sphere			
Priority order: ' ' > ' (space), '!			

### 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 : <u>http://biosciencedbc.jp/dbsearch/</u>

# **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