

Introduction to Balvies

a guided tutorial

EMBL-EBI, Jan 21st, 2013.

Jim Procter & Geoff Barton

College of Life Sciences

University of Dundee, UK.

jprocter@dundee.ac.uk

Plan for the Day

08:30-09:00	Arrival and Registration
09:00-09:10	Welcome & Safety Notices
09:10-10:30	<i>Introduction to Jalview</i> – Manual Chapter 1
10:30-11:00	Coffee/Tea Break
11:00-12:30	<i>Introduction to Jalview</i> – Manual Chapter 1
12:30-13:30	Lunch
13:30-14:30	Geoff Barton Multiple sequence alignment and analysis
14:30-15:00	<i>Alignment and analysis with Jalview</i>
15:00-15:30	Coffee/Tea Break
15:30-16:00	Penny Coggil Jalview and Pfam
16:00-16:50	Alignment and analysis with Jalview continued
16:50-17:00	Day 1 Wrap-up
17:00	Departure to accommodation
18:30-19:00	Meet for evening dinner @7 in Red Lion, Hinxton.

Day 2 - Tuesday

09:00-10:30	Advanced Jalview and Jalview Hackathon
10:30-11:00	Coffee/Tea Break
11:00-12:30 continued	Advanced Jalview and Jalview Hackathon
12:30-13:30	Lunch
13:30-15:00	Jalview Clinic and Jalview Hackathon
15:00-15:30	Coffee/Tea Break
15:30-15:50	Jalview Clinic and Jalview Hackathon continued
15:50-16:15	Wrap-up
16:15	Departure

Course materials

Everything is (or will) be online at

**[http://www.jalview.org/tutorial/training-materials/
2013/EBI/Jan/](http://www.jalview.org/tutorial/training-materials/2013/EBI/Jan/)**

- These slides**
- Jalview v2.8 Tutorial (v1.4.1)**
- A copy of the tutorial is also on Penelope**

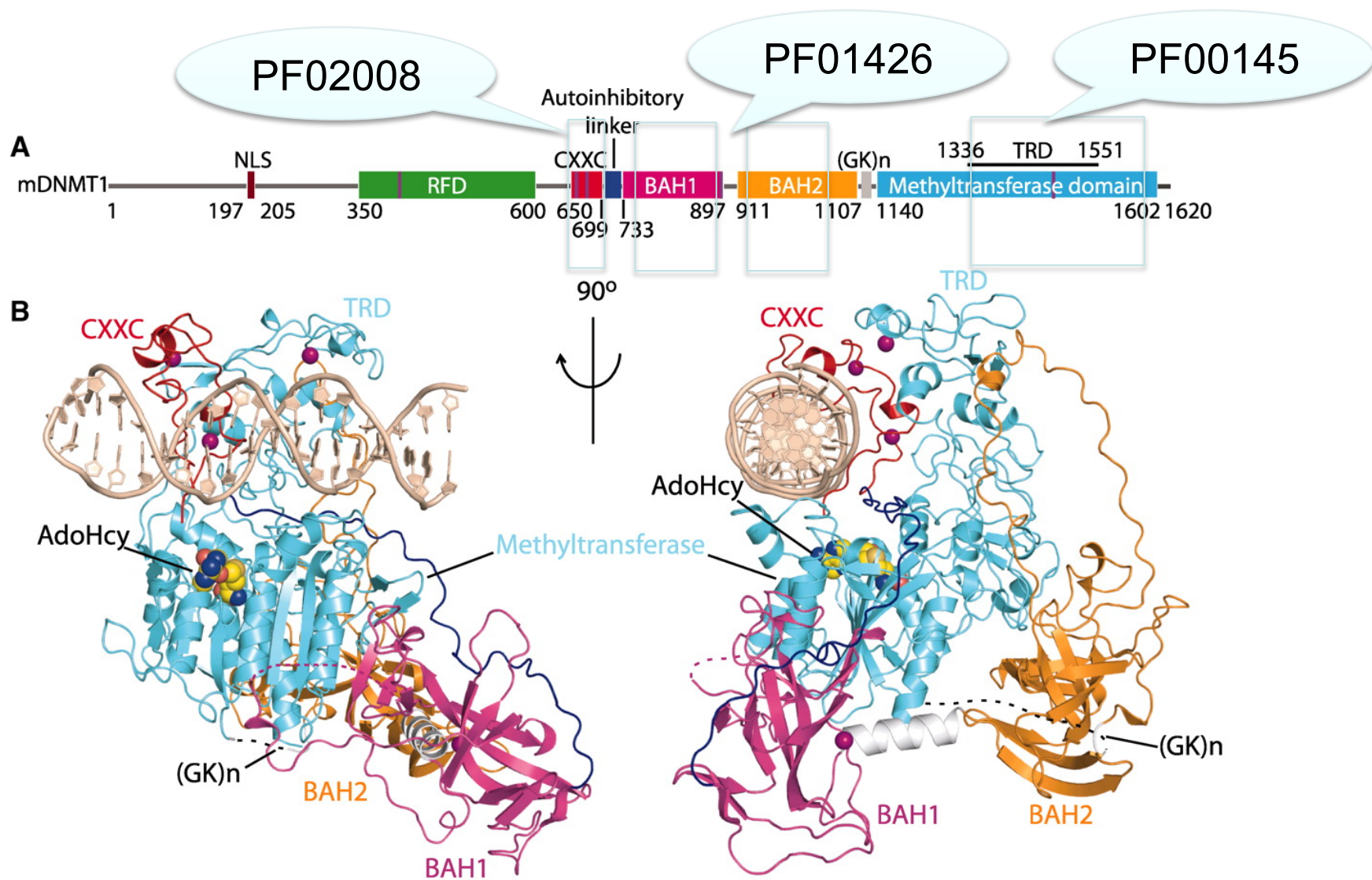


Fig. 1 Structural overview of mDNMT1(650–1602)–DNA 19-nucleotide oligomer complex with bound AdoHcy.

J Song et al. Science 2011;331:1036-1040

FASTA

```
HC class II antigen
>Q8MGZ9/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q8HWS7/1-89 MHC class II
RFLQQDKYECHFFNGTERVRFLHRD
YCRHNYGVGESFTVQRR
>Q30167/1-89 MHC class II
RFLEEVKFECHFFNGTERVRLLERR
YCRHNYGVGESFTVQRR
>Q95IE2/1-89 MHC class II
RFLWQGKYKCHFFNGTERVQFLERL
YCRHNYGVGESFTVQRR
>Q95IE6/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q95IE5/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q95IE4/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q95IE3/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q95IE2/1-89 MHC class II
RFLWQGKYKCHFFNGTERVQFLERL
YCRHNYGVGESFTVQRR
>Q95IE1/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q95IE0/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q95IE6:0.058815,
Q95IE5:0.233569,
Q95IE4:0.029464,
Q95IE3:0.029464,
Q95IE2:0.090944,
Q95IE1:0.076228,
Q95IE0:0.110844);
Consensus,R 100%, F 100%, L 100%, E
0%, H 100%, F 100%, F 100%, M 100%, G 100%, T 100%, F 100%, R
```

Newick

CSV

GFF

PDB

Bioinformatics
data is not fun to
read.....

Alignment

MHC class II antigen

```
Q95IE6 uniprot non_terminal_residue 88 88 0.0
Q8MGZ9 uniprot non_terminal_residue 1 1 0.0
Q8MGZ9 uniprot non_terminal_residue 89 89 0.0
Q8HWS7 uniprot non_terminal_residue 1 1 0.0
Q8HWS7 uniprot non_terminal_residue 89 89 0.0
Q30167 uniprot signal_peptide 1 29 0.0
Q30167 uniprot mature_protein_region 30 266 0.0
Q30167 uniprot extramembrane 30 227 0.0
Q30167 uniprot transmembrane 228 250 0.0
Q30167 uniprot extramembrane 251 266 0.0
```

Features

Annotation

Graphical Tools:

- Visualize data and results
- Access to analysis programs

So generally...

- make our lives easier!

Structure

Tree

1D Protein Visualization

- Visualize information at each position in sequence

- Amino Acid Properties

- Discrete colour schemes
- Quantitative Scales
 - Hydrophobicity, polarity, size.

DDDQ I D E G W V L T C A A Y P

DDDQ I D E G W V L T C A A Y P

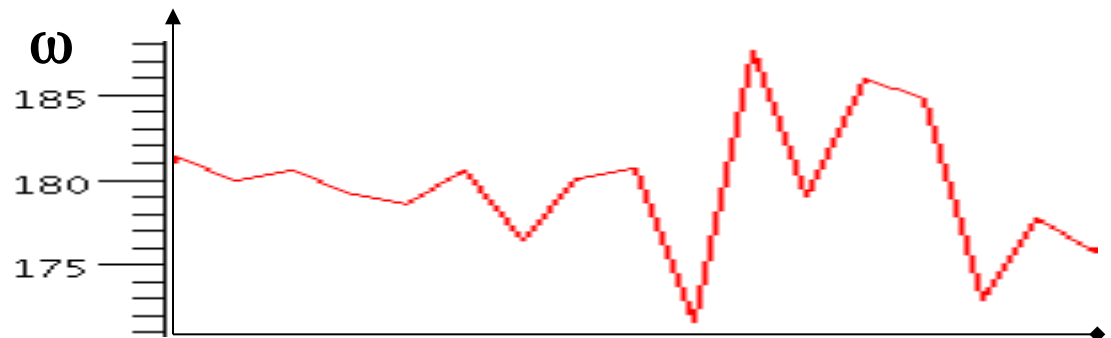
DDDQ I D E G W V L T C A A Y P

- Protein Features

- Discrete annotations
- Observations & Predictions
- Measurements
 - e.g. Omega torsion
 - (reflects planarity
 - of peptide bond)

metal ion-binding site	
strand	
turn	
helix	

D DDQ I D E G W V L T C A A Y P



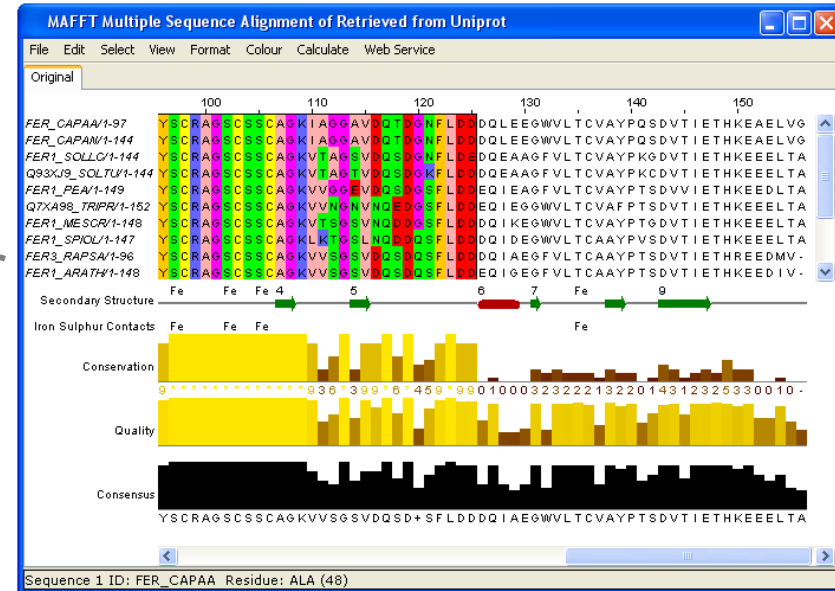
Multiple Alignments

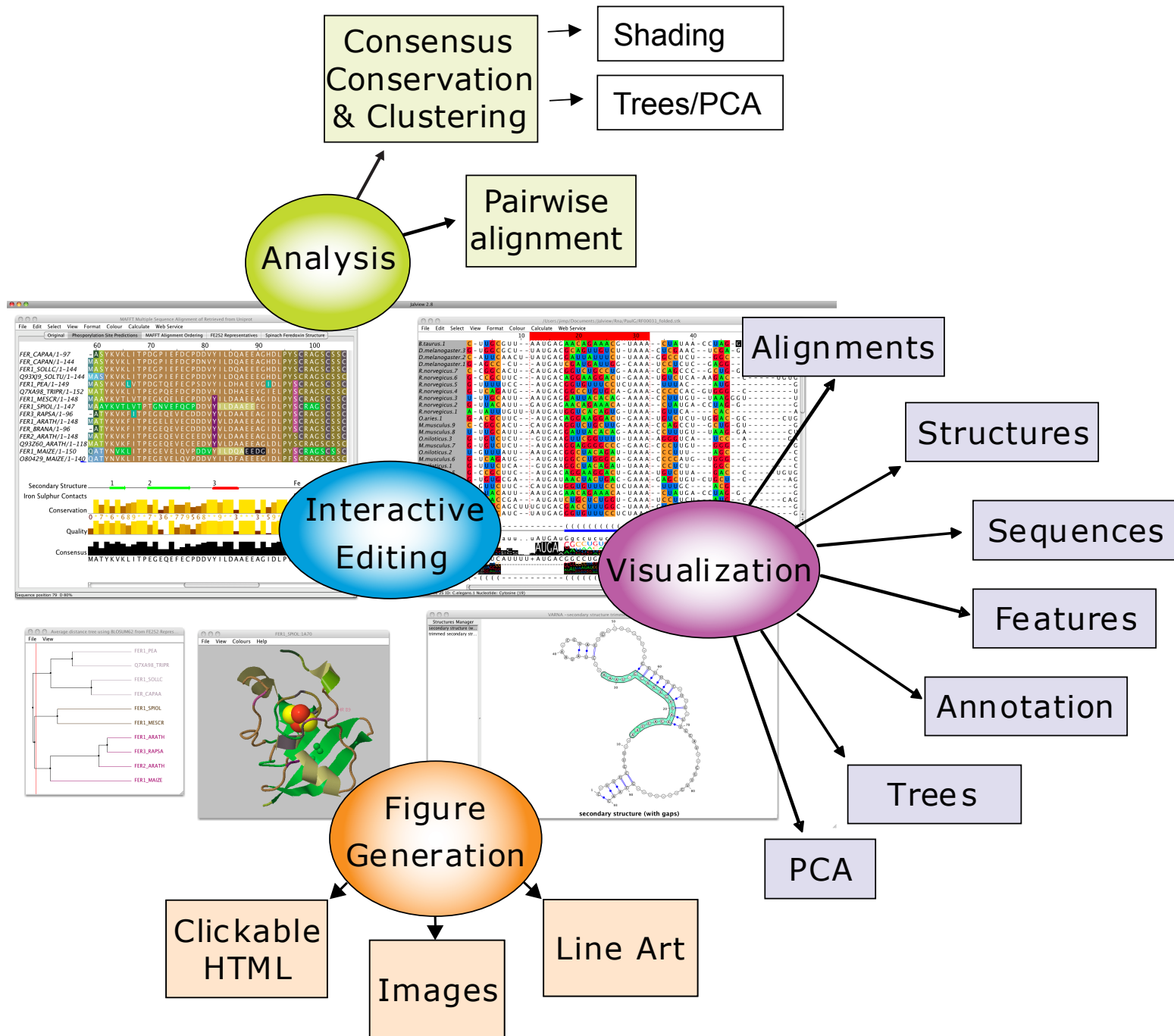
- Generated by heuristic optimization algorithm
 - Alignments sometimes need manual curation
- Correct alignments conserve common properties
- Shading can highlight differences

FER_CAPAA/28-82	AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDQLEE ⁺ GWVLT ⁺ CVAYPQ
FER_CAPAN/75-129	AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDQLEE ⁺ GWVLT ⁺ CVAYPQ
FER1_LYCES/75-129	AE ⁺ EEGHDLPYSCRAGSCSSCAGKV ⁺ TAGSVDQSDGNFLDEDQEAAGFVLT ⁺ CVAYPK
Q93XJ9_SOLTU/75-129	AE ⁺ EEGHDLPYSCRAGSCSSCAGKV ⁺ TAGTVDQSDGKFLDDQEAAGFVLT ⁺ CVAYPK
FER1_PEA/80-134	AEEVGIDLPYSCRAGSCSSCAGKVVGGEVDQSDGSFLDDEQIEAGFVLT ⁺ CVAYPT
Q7XA98_TRIPR/83-137	AEEVGIELPYSCRAGSCSSCAGKV ⁺ VNGNVNQEDGSFLDDEQIEGGWVLT ⁺ CVAFPT
FER1_MESCR/79-133	AEEAGIDLPYSCRAGSCSSCAGKV ⁺ TSGSVNQDDGSFLDDQIKEGWVLT ⁺ CVAYPT
FER1_SPIOL/78-132	AE ⁺ EEGIDLPYSCRAGSCSSCAGKLKTGS ⁺ LNQDDQSF ⁺ LDDQIDEGWVLT ⁺ CAAYPV
FER3_RAPSA/28-82	AEEAGIDLPYSCRAGSCSSCAGKV ⁺ VSGSVDQSDQSF ⁺ LDDQIAEGFVLT ⁺ CAAYPT
FER_ARATH/80-134	AEEAGIDLPYSCRAGSCSSCAGKV ⁺ VSGSVDQSDQSF ⁺ LDDQIEGEGFVLT ⁺ CAAYPT
FER_BRANA/28-82	AEEAGIDLPYSCRAGSCSSCAGKV ⁺ VSGFVDQSD ⁺ ESFLDDQIAEGFVLT ⁺ CAAYPT
FER2_ARATH/80-134	AEEAGLDLPYSCRAGSCSSCAGKV ⁺ VSGSIDQSDQSF ⁺ LDDQMSEGYVLT ⁺ CVAYPT
FER1_MAIZE/80-134	AEEDGIDLPYSCRAGSCSSCAGKV ⁺ VSGSVDQSDQSY ⁺ LDDGQIADGWVLT ⁺ CHAYPT
O80429_MAIZE/72-126	AE ⁺ EEGIDLPFS ⁺ CRAGSCSSCAGKV ⁺ VSGSVDQSDQSF ⁺ LNDNQVADGWVLT ⁺ CAAYPT

What is Jalview ?

- A java alignment viewer
 - *j*ava *al*ignment *view*er
 - .. but not just for viewing..
- Java ?
 - Programming language
 - Platform independence
 - Standalone or web based tool





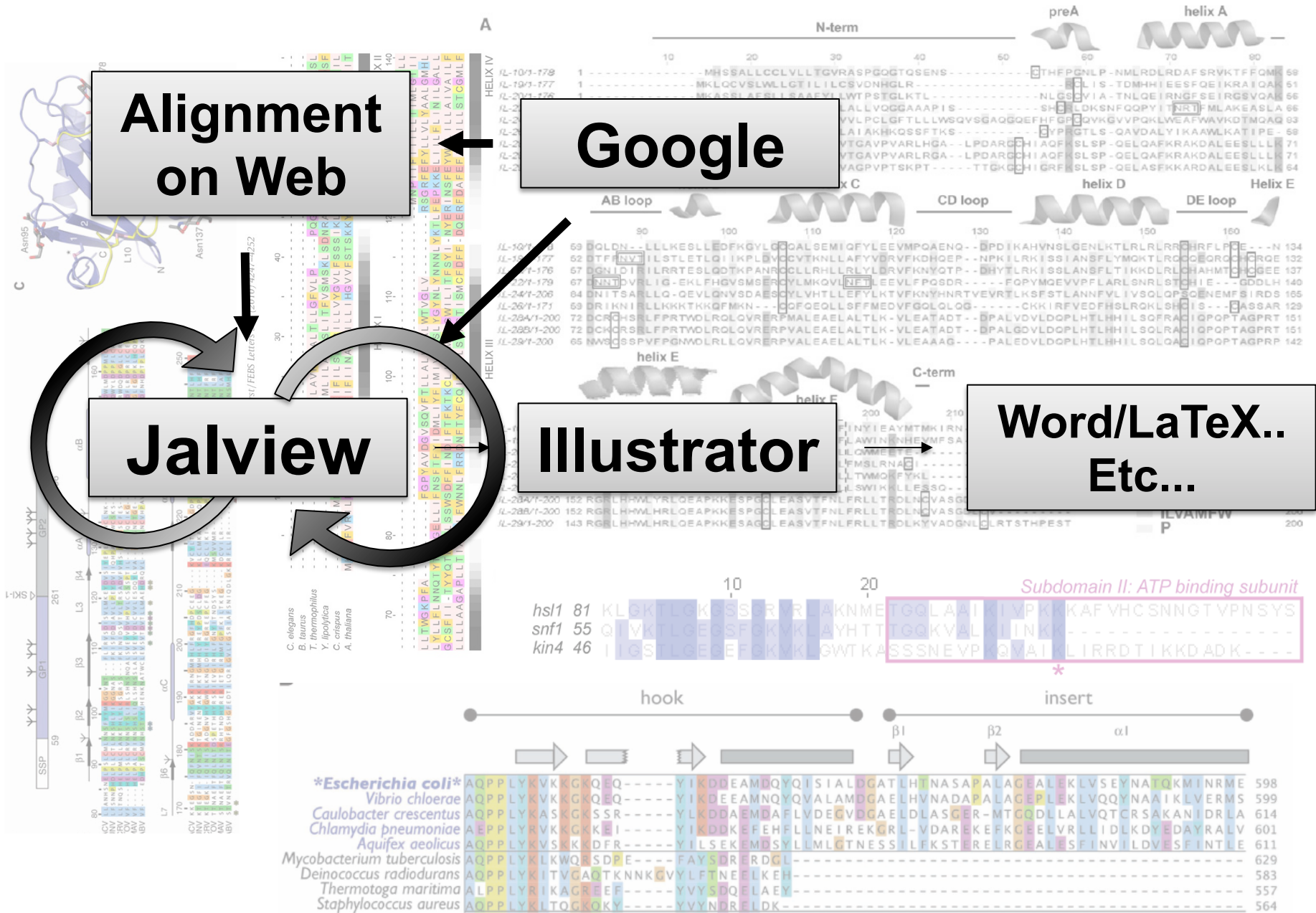
Alignment
on Web

Google

Jalview

Illustrator

Word/LaTeX..
Etc...



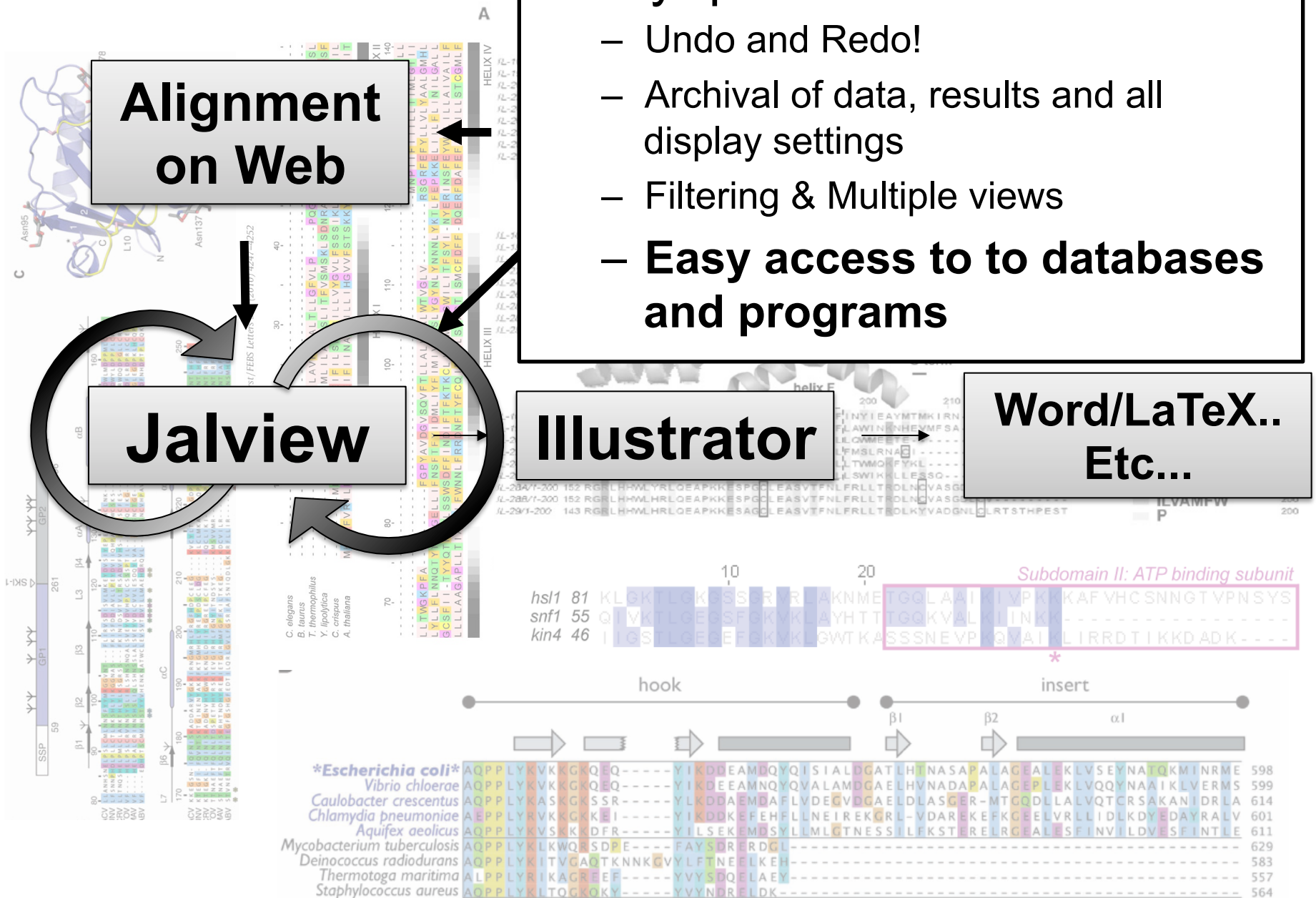
**Alignment
on Web**

Jalview

Illustrator

**Word/LaTeX..
Etc...**

- Key qualities for a workbench
 - Undo and Redo!
 - Archival of data, results and all display settings
 - Filtering & Multiple views
 - Easy access to databases and programs



Jalview Flavours

Multi-windowed Web UI
Visualization & Analysis

Common Data & Analysis
Editing, messaging and
File Input/Output

Lightweight UI
Integrate with web sites

The Jalview Desktop



Figures

HTML
EPS
PNG

Local
File
system

import
from web

Analysis Services

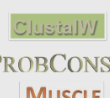
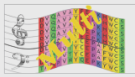
Protein 2ndary structure



Alignment Analysis

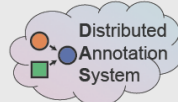
Protein Disorder

Multiple alignment



GLOBPLOT 2

Biological Data Services



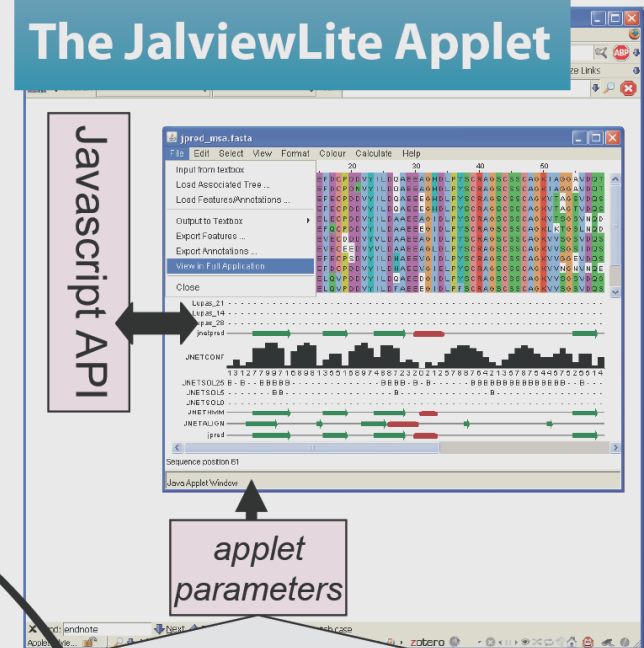
The JalviewLite Applet

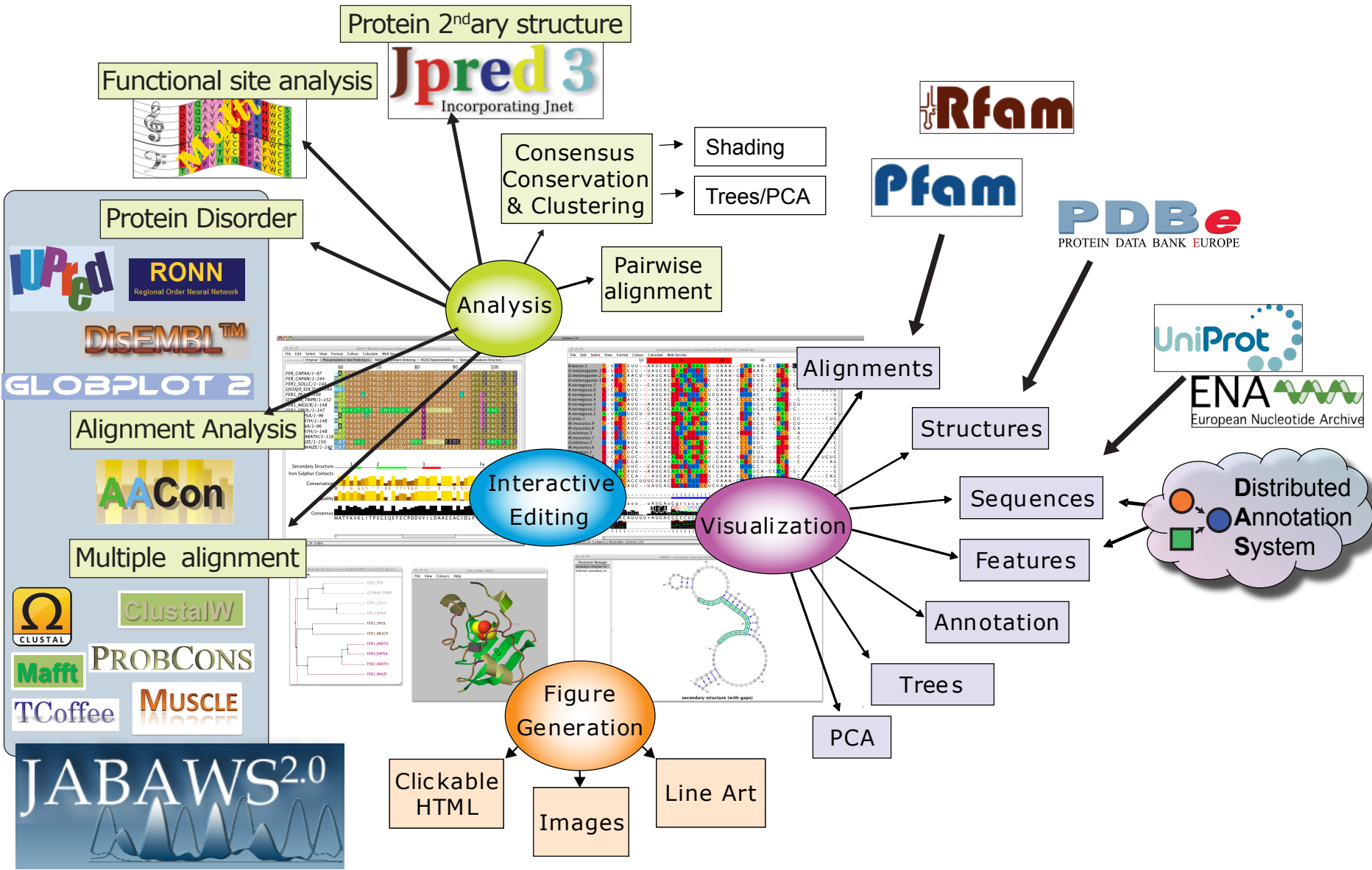
Javascript
API

applet
parameters

Alignment
And
Annotation

Bioinformatics
Web Server





Ex 1 – starting Jalview

PAGE 7

Use the Development version via the
development page



Jalview

Jalview
Launch
Buttons

Launch Jalview Applet

Launch Jalview Desktop

New logo

Home

About

Help

Community

Development

Training

Download

Latest News

Jalview in 2017
2013

Posted On: 10-07-2017

Help and
documentation

the new look

www.jalview.org

Posted On: 12-11-2012

GENERAL

Registration now open
for 3rd Jalview
Residential Training
Course and Hackathon

Posted On: 19-10-2012

TRAINING

Jalview is a free program for multiple sequence alignment editing, visualisation and analysis. Use Jalview to view sequence alignment, analyse them using phylogenetic trees and principal component analysis (PCA) plots and explore molecular structure and annotation.

Installation
packages
and source

Jalview
training news
and course
dates

Jalview
Community

Jalview
Development and
release history

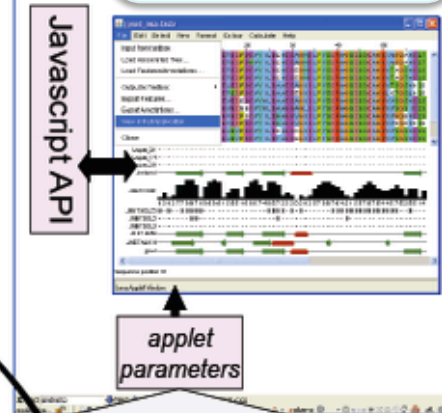
Analysis Services



Biological
Data
Services



The Jalview
Javascript API



Alignment
And
Annotation

<http://www.jalview.org>



Jalview Usage Statistics



Do you want to help make Jalview better by enabling the collection of usage statistics with Google Analytics ?

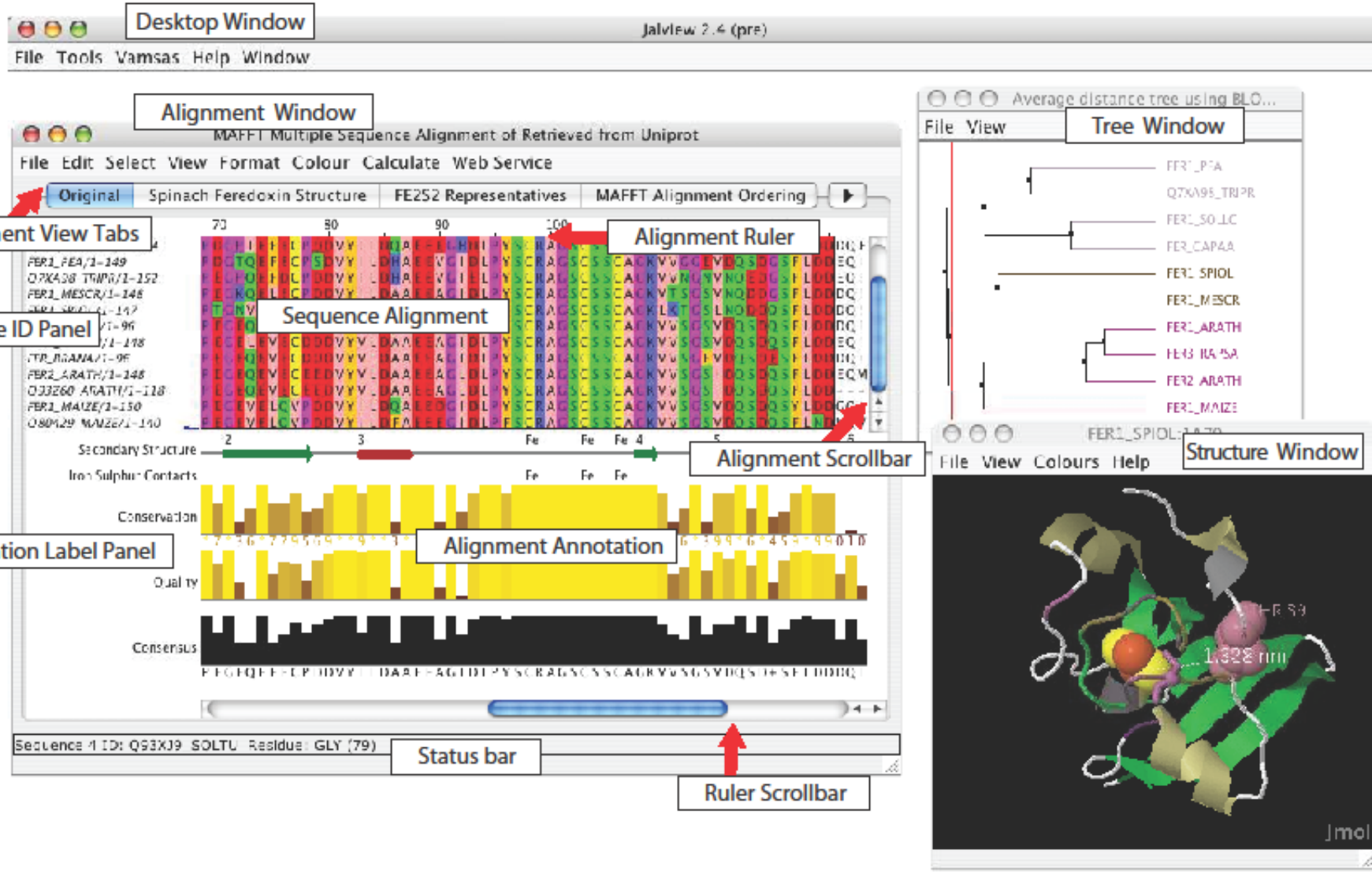
(you can enable or disable usage tracking in the preferences)

**Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009)
Jalview Version 2 - a multiple sequence alignment editor and analysis workbench
Bioinformatics doi: 10.1093/bioinformatics/btp033**

Jalview News



Anatomy of Jalview: Figure 1.7



Ex 1 – starting Jalview

- Tasks
 - Modify user preferences
 - Test that you can load the example file manually
- http://www.jalview.org/examples/exampleFile_2_7.jar

Ex 1 – starting Jalview

- Tasks

- Modify user preferences
- Test that you can load the example file manually

http://www.jalview.org/examples/exampleFile_2_7.jar

- Questions

- Where to find help ?
- How to report a bug ?

Jalview Community

- Mailing lists
 - Discussion forum and developers forum
 - Links from <http://www.jalview.org/community>
- Jalview bug database
 - <http://issues.jalview.org>
 - Also indexed on google
- Jalview development info
 - <http://www.jalview.org/development>

Ex 2 - Navigation

- Tasks
 - Open the overview window for a view
 - Jump to a specific row and column with keyboard mode

Ex 2 - Navigation

- Tasks
 - Open the overview window for a view
 - Jump to a specific row and column with keyboard mode
- Questions
 - How do you locate a sequence or sequence position if you don't know its row/column ?
 - How do you find a sequence motif ?

Ex 3 Getting data into Jalview

- Tasks
 - Importing an alignment via a url, local file, or cut' n' paste
 - Getting an alignment from Pfam

Ex 3 Getting data into Jalview

- Tasks
 - Importing an alignment via a url, local file, or cut' n' paste
 - Getting an alignment from Pfam
- Questions
 - What happens when you drag a file onto an existing alignment ?
 - What is different about the alignment retrieved from Pfam ?
 - What if you want to load a *really* big alignment ?

Ex 4. Saving alignments

- Tasks
 - Save alignments in different formats

Ex 4. Saving alignments

- Tasks
 - Save alignments in different formats
- Questions
 - What's the biggest difference between a BLC file and a pileup file ?
 - Why are Jalview projects useful ?

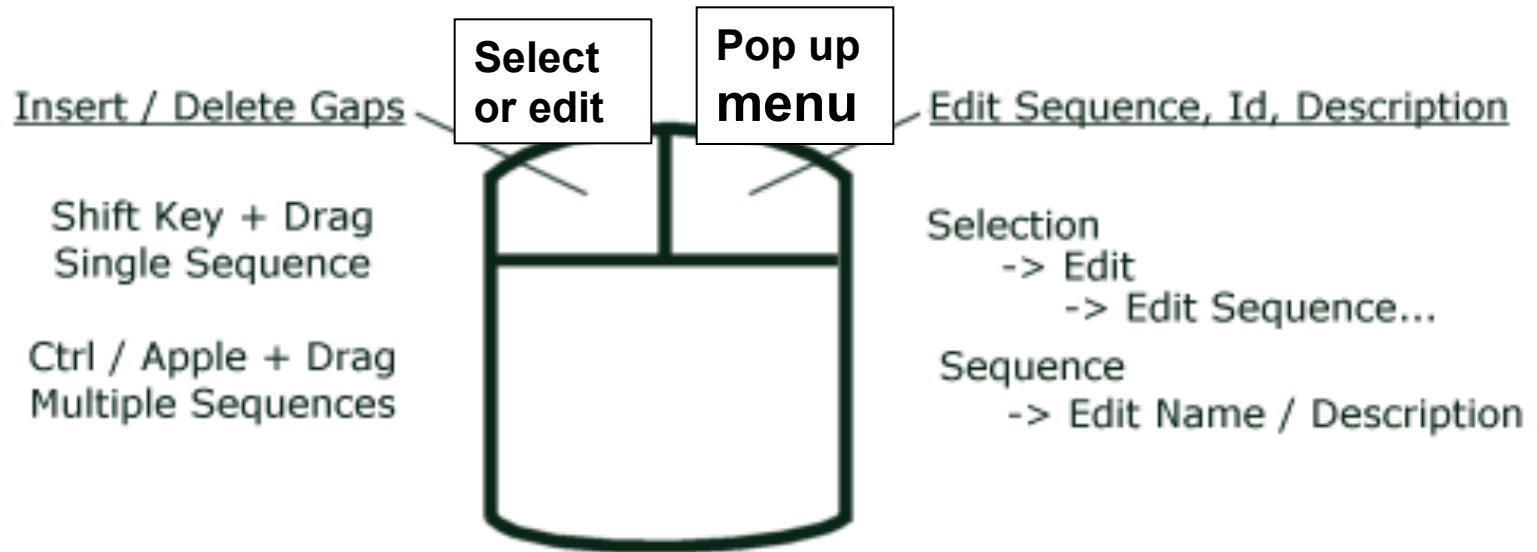
Ex 5,6,7,8 and 9

selecting, editing, hiding and showing

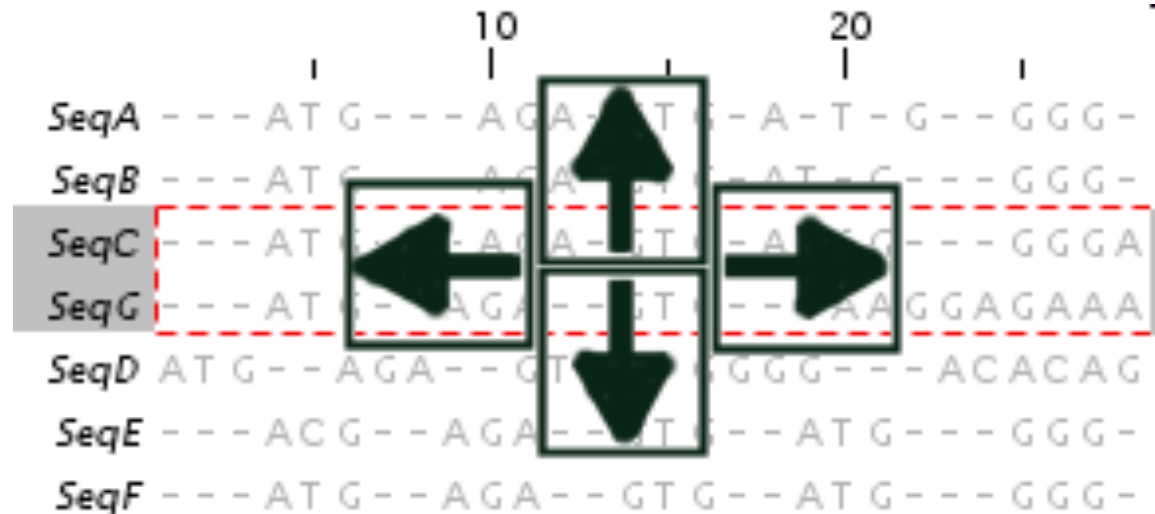
- Tasks
 - Get used to the mouse and keyboard based selection and alignment editing controls
 - Learn how to work on specific parts of an alignment
- Exercise 8 and 9 let you practice mouse and keyboard based editing techniques
 - If you don't finish them now, do them when you have a spare 15 mins.

How do I edit sequences in Jalview

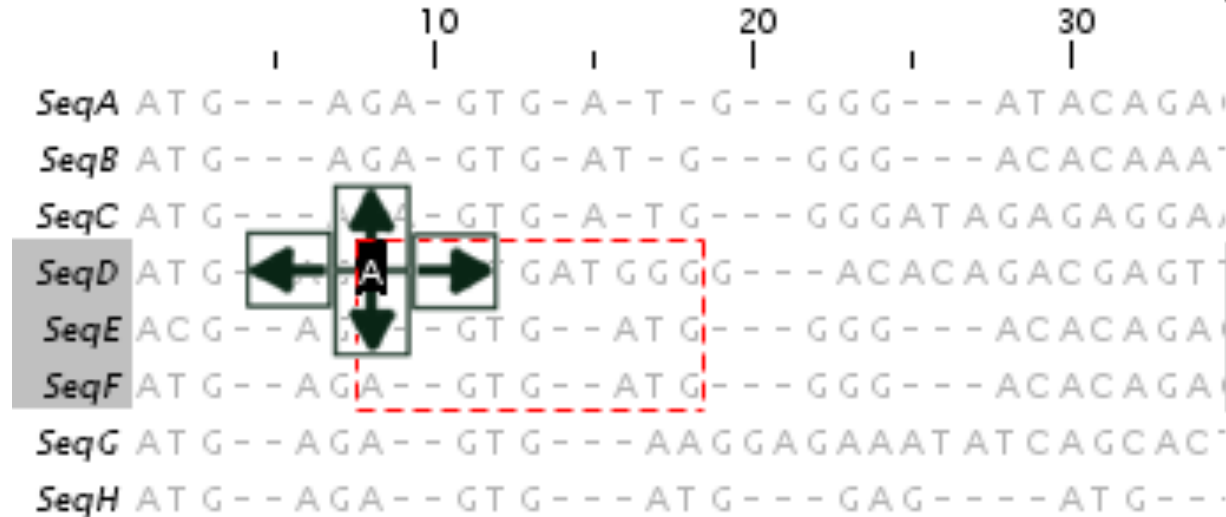
<http://www.jalview.org/examples/editing.html>



Selected sequences can be moved up and down or slid from left to right



F2 enables/disables keyboard mode



Cursor Keys - Move Cursor

Alt + Cursor Keys - Move Sequence

[X] Space - Insert [X] gap(s)

[X] Delete / Backspace - Delete [X] gap(s)

8 C - Move to Column 8

4 S - Move to Sequence 4

8,4<return> - Move to column 8, sequence 4

6 P - Move to Position 6

Q - Define the top left corner of selection area

M - Define the bottom right corner of selection area

Windows 7 Users:

Group Insert - use

Shift+Space

(not CTRL+Space)

Ex 5,6,7,8 and 9

selecting, editing, hiding and showing

- Tasks
 - Get used to the mouse and keyboard based selection and alignment editing controls
 - Learn how to work on specific parts of an alignment
- Questions
 - Why would you create representative sequences ?
 - How do you insert a gap in the middle of a sequence without affecting the rest of its alignment ?

Ex 10 & 11 : Colouring

- Tasks
 - Learn how to colour all, or part of the alignment by
 - Amino acid property
 - Annotation

Ex 10 & 11 : Colouring

- Tasks

- Learn how to colour all, or part of the alignment by
 - Amino acid property
 - Annotation

- Questions

- Why is colouring the alignment useful ?
- How would you highlight acidic residues in your alignment ?

Ex 12,13 – alignment layout and export

- Tasks
 - Adjust the alignment formatting options
 - Wrap
 - Sequence id margin
 - Export the alignment as a figure
 - HTML, EPS and PNG

Lunch

Ex 12,13 – alignment layout and export

- Tasks
 - Adjust the alignment formatting options
 - Wrap
 - Sequence id margin
 - Export the alignment as a figure
 - HTML, EPS and PNG
- Questions
 - How do you control the number of columns shown in wrapped mode ?
 - How can you easily experiment with different alignment figure layouts ?
 - What programs can edit EPS files ?

- End of Session 1
 - Loading/saving
 - Navigation/Editing
 - Colouring/Annotation
 - Figure export

- Session 2
 - Alignment with JABAWS
- Section 2.4
PAGE 58 IN MANUAL
 - Alignment analysis

Next

- Section 2.4
 - Alignment with JABAWS.

PAGE 58 IN MANUAL

- End of Alignment analysis
 - PCA and Trees
 - Subfamily analysis

End of anli

- Section 2.4
 - Alignment with JABAWS.

PAGE 58 IN MANUAL

- End of Alignment analysis
 - PCA and Trees
 - Subfamily analysis

www.compbio.dundee.ac.uk/jabaws

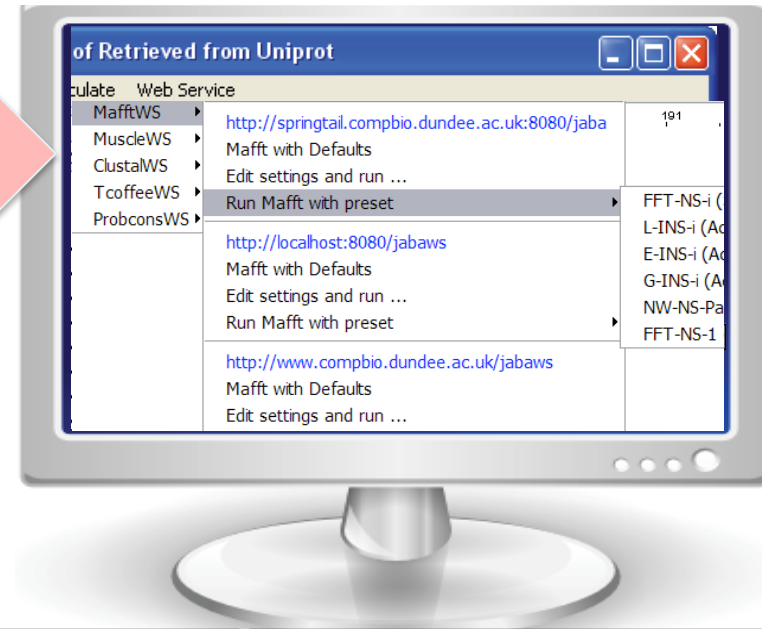
JABAWS:MSA Troshin et al. 2011,
Bioinformatics.

JABAWS 2 In Preparation.



Jalview Web
Service GUI

JABAWS
Java Client



JABAWS Virtual
Appliance for
your private use.



Native JABAWS installs
on a range of platforms



Apache
Tomcat



JABAWS Amazon
Machine Image
on EC2



the
cloud
market

Clustal Omega quick alignment of millions of sequences



ClustalW

Mafft

PROBCONS

TCoffee

MUSCLE

Jpred 3
Incorporating Jnet

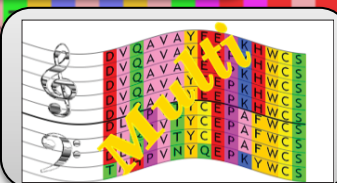
RONN
Regional Order Neural Network

DisEMBL™

UPred

GLOBPLOT 2

Assorted protein disorder predictors



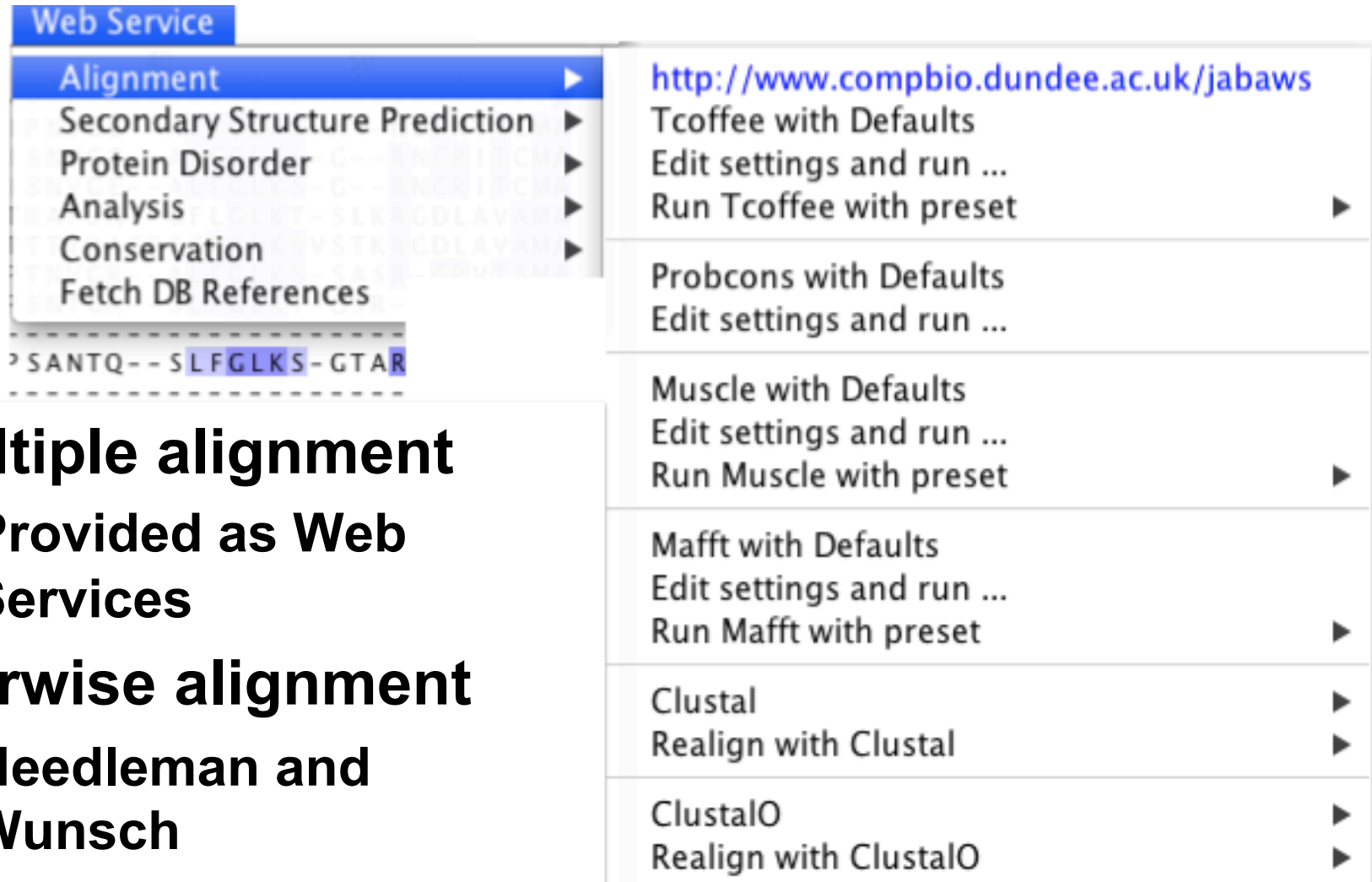
Protein conservation calculations

AACON

- Alignment
- Secondary Structure Prediction
- Protein Disorder
- Analysis
- Conservation
- Envision 2
- Fetch DB References

JABAWS 2 services
in Jalview 2.8

Jalview's Alignment Methods



- **Multiple alignment**
 - Provided as Web Services
- **Pairwise alignment**
 - Needleman and Wunsch
 - Mostly used internally (described in section 2.2.7)

Jalview alignment exercise 25 (sect. 2.4)

- Tasks
 - Align sequences using different methods
 - Use the Webservices' 'alignment' submenu
 - Explore how hidden regions affect alignment jobs.

Jalview alignment exercise 25 (sect. 2.4)

- Tasks
 - Align sequences using different methods
 - Use the Webservices' 'alignment' submenu
 - Explore how hidden regions affect alignment jobs.
- Questions
 - Why does jalview run several jobs if the input includes hidden regions ?
 - What does 're-alignment' mean ?

Alignment Job Parameter Settings

The screenshot shows a software window titled "Edit parameters for Align with MuscleWS". It contains several sections: "Details" with a text box, "Options" with checkboxes for "dimer", "Diagonal", "Diagonal 1", and "Profile scoring method" (which is checked), and "Parameters" with a "Sequence type" dropdown set to "auto". At the bottom are "Start Job" and "Cancel Job" buttons. Callouts provide the following information:

- Browse or edit to change name of set**: Points to the "Current parameter set name" dropdown menu, which currently shows "User Defaults".
- Buttons appear to create, update, rename or delete user settings.**: Points to the "Revert" and "Create" buttons.
- text box to add notes for the parameter set**: Points to the empty text box in the "Details" section.
- Parameters contains more complex settings**: Points to the "Parameters" section, which includes a "Sequence type" dropdown.
- Start job with current settings or cancel.**: Points to the "Start Job" and "Cancel Job" buttons at the bottom.
- Tooltips give brief description and link (right click) to further info**: Points to a tooltip for the "Profile scoring method" option, which lists options like "le - use log-expectation profile score VTM240" and "sp - use sum-of-pairs protein profile score", and includes a "Link" icon.

Modifying alignment parameters ?

Jaba Alignment Exercise

- Task
 - Run the alignment from step **b** of ex. 25 using the JABA clustalW service
 1. Run with default settings
 2. Use the ‘Edit parameters’ dialog to run an alignment with the following:
 - Gap opening (internal and end gaps) = 3
 - Gap Extension = 0.05
 - Compare the two alignments. You may want to save them for later, too.
- Questions
 - What effect has modifying the gap penalties had on the ferredoxin alignment ?

Alignment Analysis

Using jalview to analyse the relationships between aligned sequences.

Alignment analysis – Section 2.2

- Principal component analysis
- Phylogenetic trees
- Redundancy removal
- Tree based conservation analysis
- Subdividing alignment by mutation

=> If there's time, carry on to session 3

Phylogenetic analysis and Jalview

- Built in tree methods
 - UPGMA
 - Fast, simple, but not reliable for phylogenetic inference
 - Neighbour joining
 - Slower than UPGMA
 - Useful for a first approximation
 - NJ does not work well for very divergent sequence sets
 - » Need to add in close relatives to get an idea of topology
- Import trees from elsewhere
 - Load a Newick format tree file onto an alignment from another program

PCA and Phylogeny Exercises

Section 2.2 - Exercise 17 and 18

- Tasks
 - Calculate Principal component analyses (PCAs) and trees on the ferredoxin alignment
 - Explore the use of the interactive tree viewer
 - Use it to select subgroups on the alignment.

PCA and Phylogeny Exercises

Section 2.2 - Exercise 17 and 18

- Tasks
 - Calculate Principal component analyses (PCAs) and trees on the ferredoxin alignment
 - Explore the use of the interactive tree viewer
 - Use it to select subgroups on the alignment.
- Questions
 - What is the role of BLOSUM62 or Percentage identity in the tree building process ?

Tree based conservation analysis

Sect. 2.2.3 Exercise 19

- “Poor man’ s” character inference analysis
 - Compare conservation patterns within and between branches of a tree
- Task
 - Use interactive tree viewer to subdivide alignment and identify difference in conservation pattern

Tree based conservation analysis

Sect. 2.2.3 Exercise 19

- “Poor man’ s” character inference analysis
 - Compare conservation patterns within and between branches of a tree
- Task
 - Use interactive tree viewer to subdivide alignment and identify difference in conservation pattern
- Questions
 - How can you tell which differences are important ?
 - How can you navigate the sub-groups of a large alignment ?

Sub-groups and Sub-group Annotation

Exercise 21

- Task
 - Use the group consensus sequence logos to more easily compare tree subgroups
 - Use ‘Make groups for selection’ to subdivide groups by specific mutation

Sub-groups and Sub-group Annotation

Exercise 21

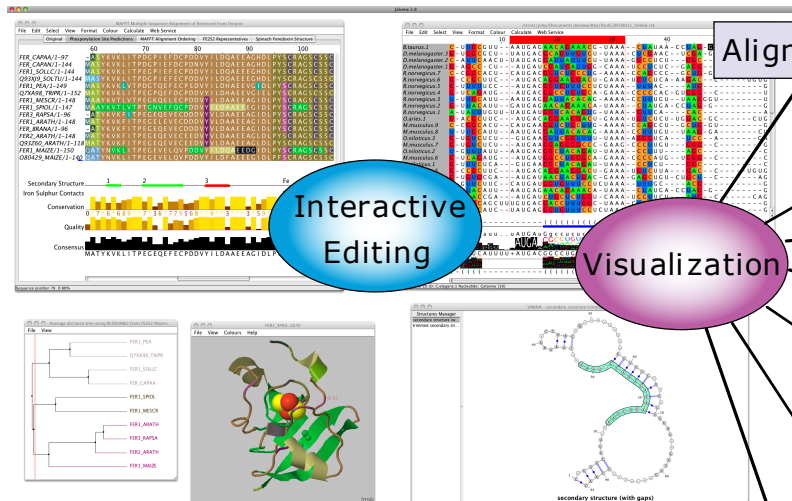
- Task
 - Use the group consensus sequence logos to more easily compare tree subgroups
 - Use ‘Make groups for selection’ to subdivide groups by specific mutation
- Questions
 - How can you work out which group is associated with which annotation row ?

- End of Session 2
 - Alignment
 - Alignment analysis

- Session 3
 - Sequence DB refs and Sequence Features
 - Protein secondary structure prediction
 - Alignment annotation

Getting and working with sequence features and annotation

- Sequence Databases
- Sequence feature sources
 - DAS Sequence feature retrieval
 - GFF and Jalview feature files
- Visualizing features
 - Highlighting annotated regions
 - Shading and reordering based on scores and labels



Alignments

Structures

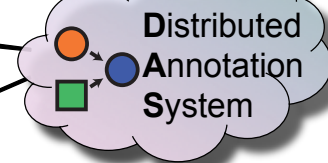
Sequences

Features

Annotation

Trees

PCA



Jalview and Sequence Databases

Sec 2.9.1 Ex. 29

- Can retrieve new sequences or match against existing records using IDs
- Task
 - Recover the Uniprot annotation for the ferredoxin sequences using their IDs
 - **make sure you are using the latest 2.8 Build**
 - Verify retrieval by examining sequence annotation