Practical Jalview

a guided tutorial

Edinburgh, June 27th, 2012.

Jim Procter & Saif Shehata
College of Life Sciences
University of Dundee, UK.
jprocter@dundee.ac.uk

Plan for the Day

- **9am-12pm** *Introduction to Jalview*
  - **Starting the application, importing, editing, colouring and creating alignment figures.**
- **12.00pm – 1pm Lunch**
- **1pm - 2pm Seminar – Geoff Barton**

  ‘Alignment visualisation and Analysis with Jalview’
- **2.15pm – 5pm** *Alignment and Analysis*
  - **creating sequence alignments, importing and calculating trees, tree based alignment analysis, viewing sequence annotation, secondary structure prediction, viewing structures.**
- **9am-12pm Day 2 - Thursday 28th**
  - **Advanced Topics ‘Choose your own adventure’**

Day 2 – ‘Advanced Topics’

- **9am-12pm** *‘Choose your own Adventure’*
  - *Practice Using Jalview on your own data*
  - *Continue with tutorial exercises*
    - **Working with structure data in Jalview**
    - **Installing your own JABAWS server**
    - **Deploying JalviewLite on a web page**
    - **Jalview alignment annotation files**
    - **Jalview and nucleic acid sequences**

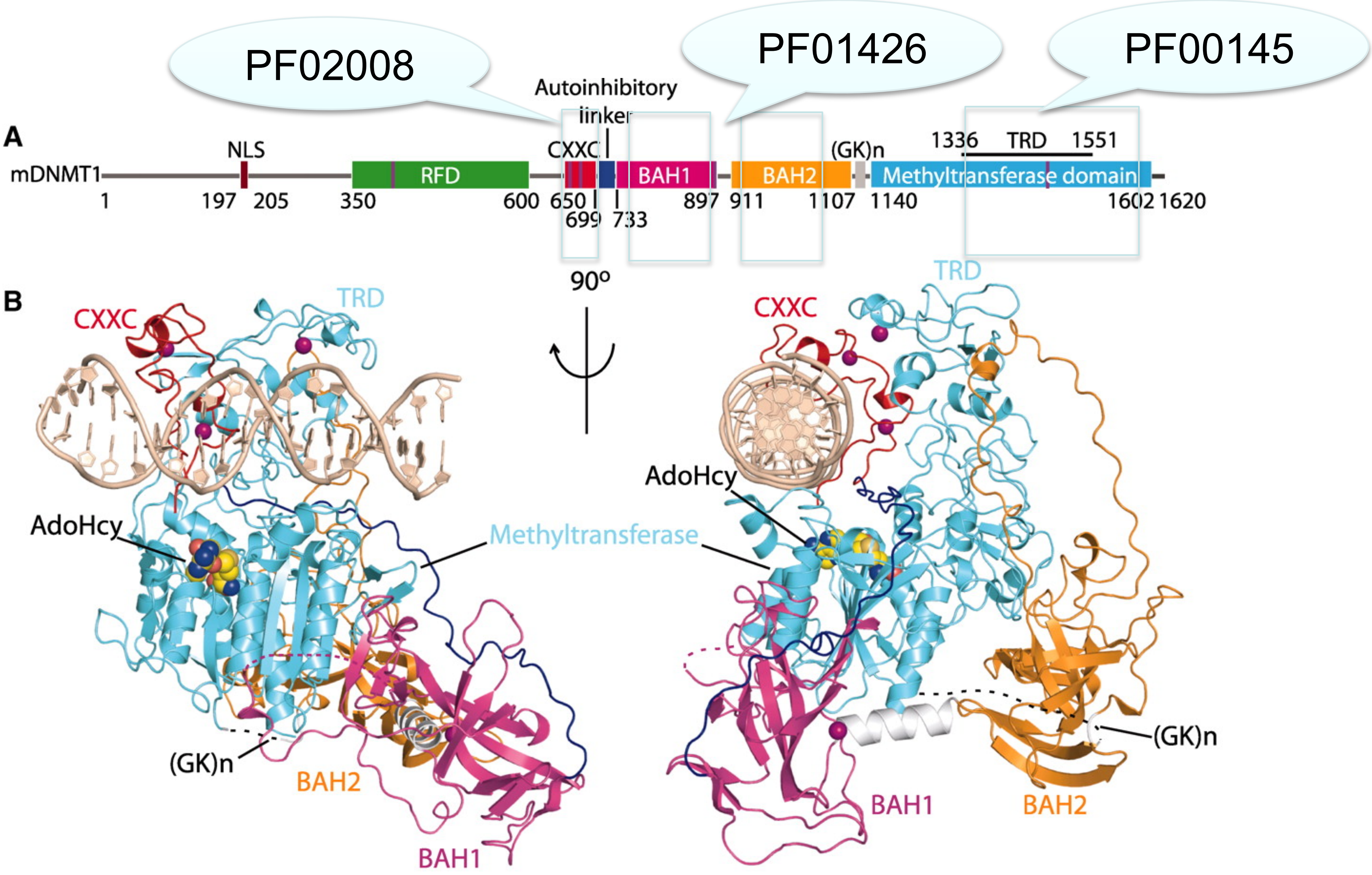
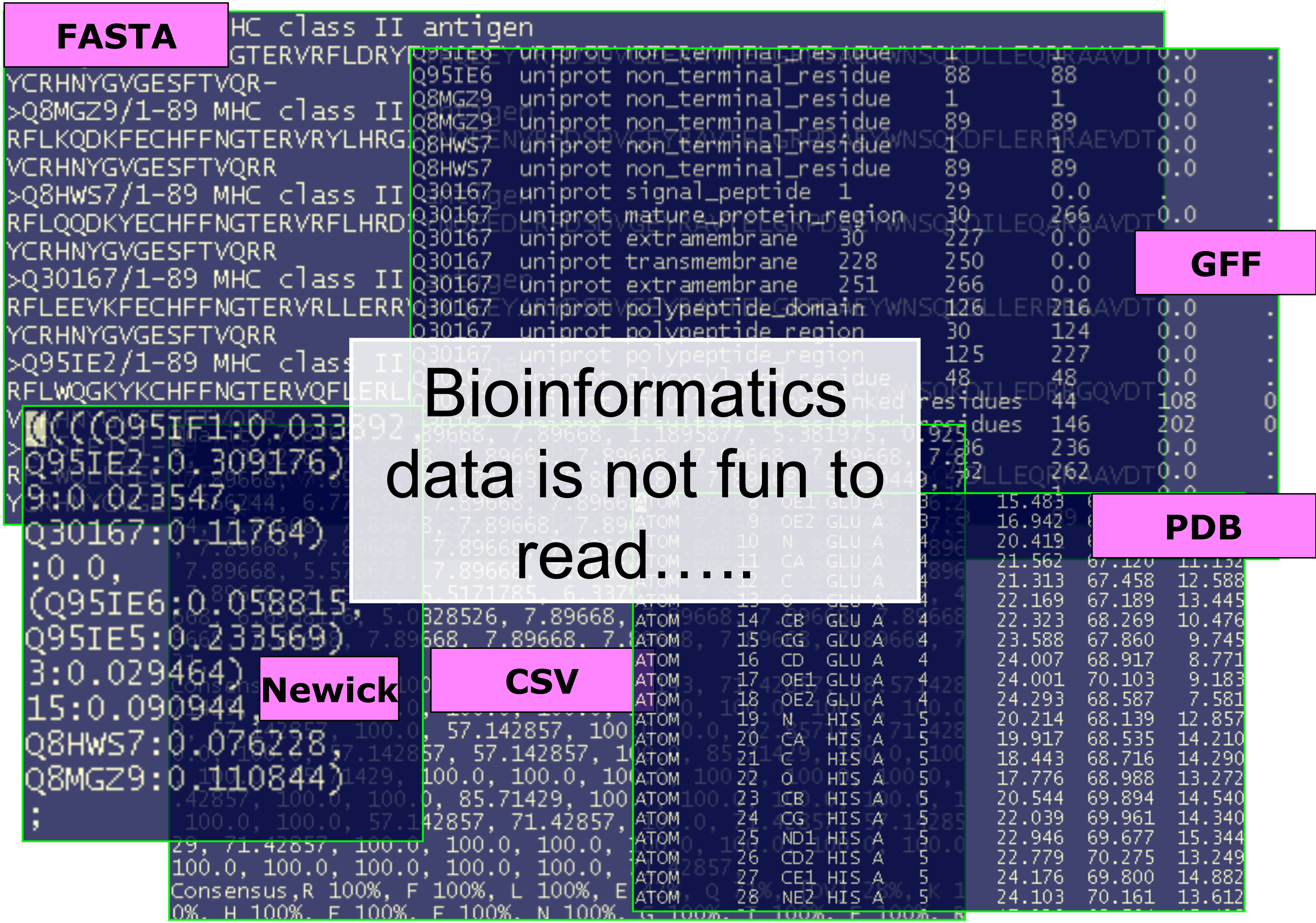


Fig. 1 Structural overview of mDNMT1(650–1602)–DNA 19-nucleotide oligomer complex with bound AdoHcy.

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

Published by AAAS



```
HC class II antigen
GTERVRFLDRY
YCRHNYGVGESFTVQR-
>Q8MGZ9/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
VCRHNYGVGESFTVQRR
>Q8HWS7/1-89 MHC class II
RFLQQDKYECHFFNGTERVRFLHRD
YCRHNYGVGESFTVQRR
>Q30167/1-89 MHC class II
RFLEEVKFECHFFNGTERVRLLERR
YCRHNYGVGESFTVQRR
>Q95IE2/1-89 MHC clas
RFLWQGKYKCHFFNGTERVQFL
```

```
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
Q30167  uniprot polypeptide_domain      126     216     0.0 .
Q30167  uniprot polypeptide_region      30      124     0.0 .
Q30167  uniprot polypeptide_region      125     227     0.0 .
```

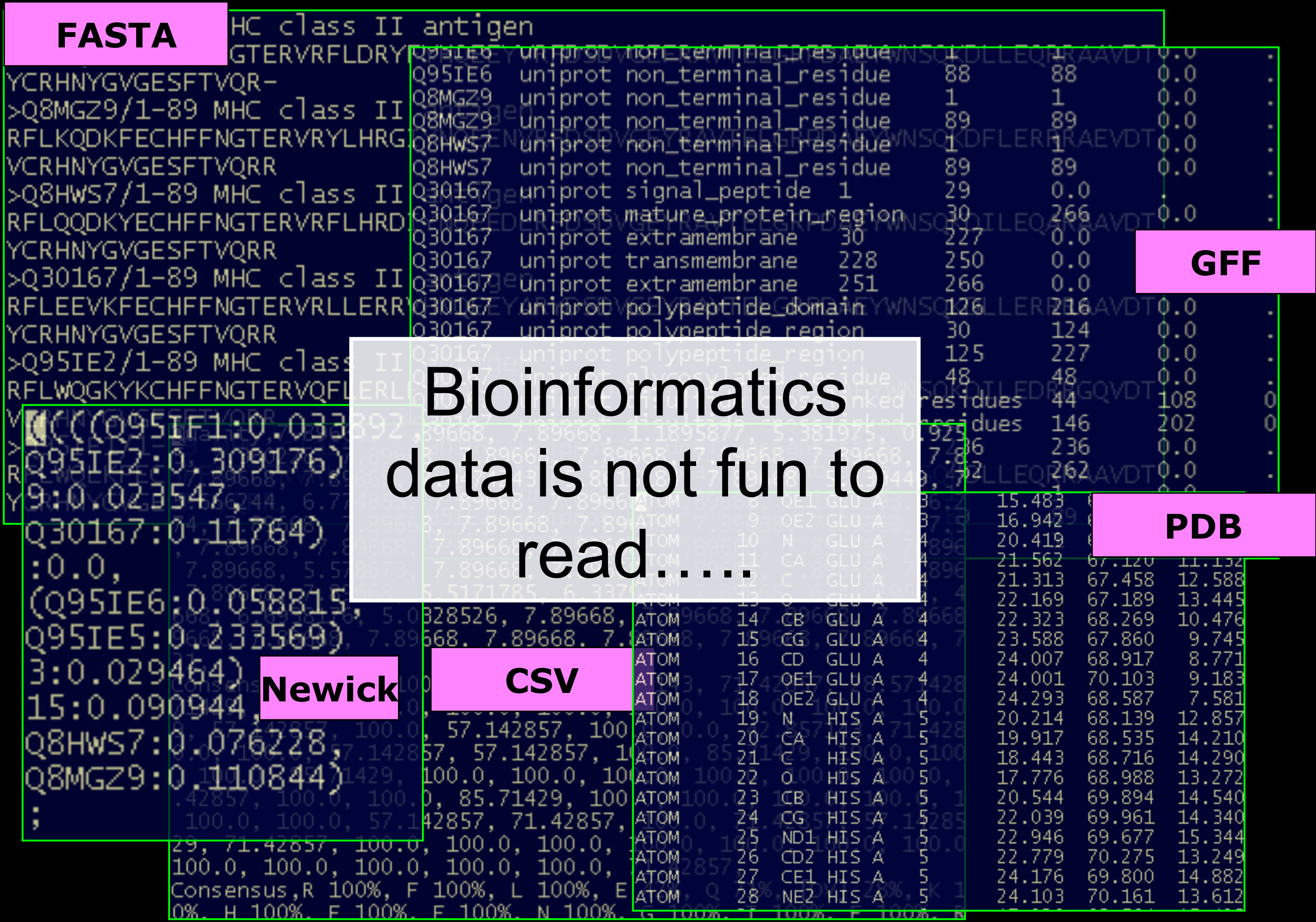
```
((((Q95IF1:0.033892,
Q95IE2:0.309176),
9:0.023547,
Q30167:0.11764)
:0.0,
(Q95IE6:0.058815,
Q95IE5:0.233569)
3:0.029464)
15:0.090944
Q8HWS7:0.076228,
Q8MGZ9:0.110844)
;
```

```
100.0, 100.0, 100.0, 100.0, 100.0,
Consensus,R 100%, F 100%, L 100%, E
```

```
ATOM     14  CB  GLU A   4      22.323  68.269  10.476
ATOM     15  CG  GLU A   4      23.588  67.860   9.745
ATOM     16  CD  GLU A   4      24.007  68.917   8.771
ATOM     17  OE1 GLU A   4      24.001  70.103   9.183
ATOM     18  OE2 GLU A   4      24.293  68.587   7.581
ATOM     19  N   HIS A   5      20.214  68.139  12.857
ATOM     20  CA  HIS A   5      19.917  68.535  14.210
ATOM     21  C   HIS A   5      18.443  68.716  14.290
ATOM     22  O   HIS A   5      17.776  68.988  13.272
ATOM     23  CB  HIS A   5      20.544  69.894  14.540
ATOM     24  CG  HIS A   5      22.039  69.961  14.340
ATOM     25  ND1 HIS A   5      22.946  69.677  15.344
ATOM     26  CD2 HIS A   5      22.779  70.275  13.249
ATOM     27  CE1 HIS A   5      24.176  69.800  14.882
ATOM     28  NE2 HIS A   5      24.103  70.161  13.612
```

Bioinformatics data is not fun to read…..

Bioinformatics data is not fun to read…..



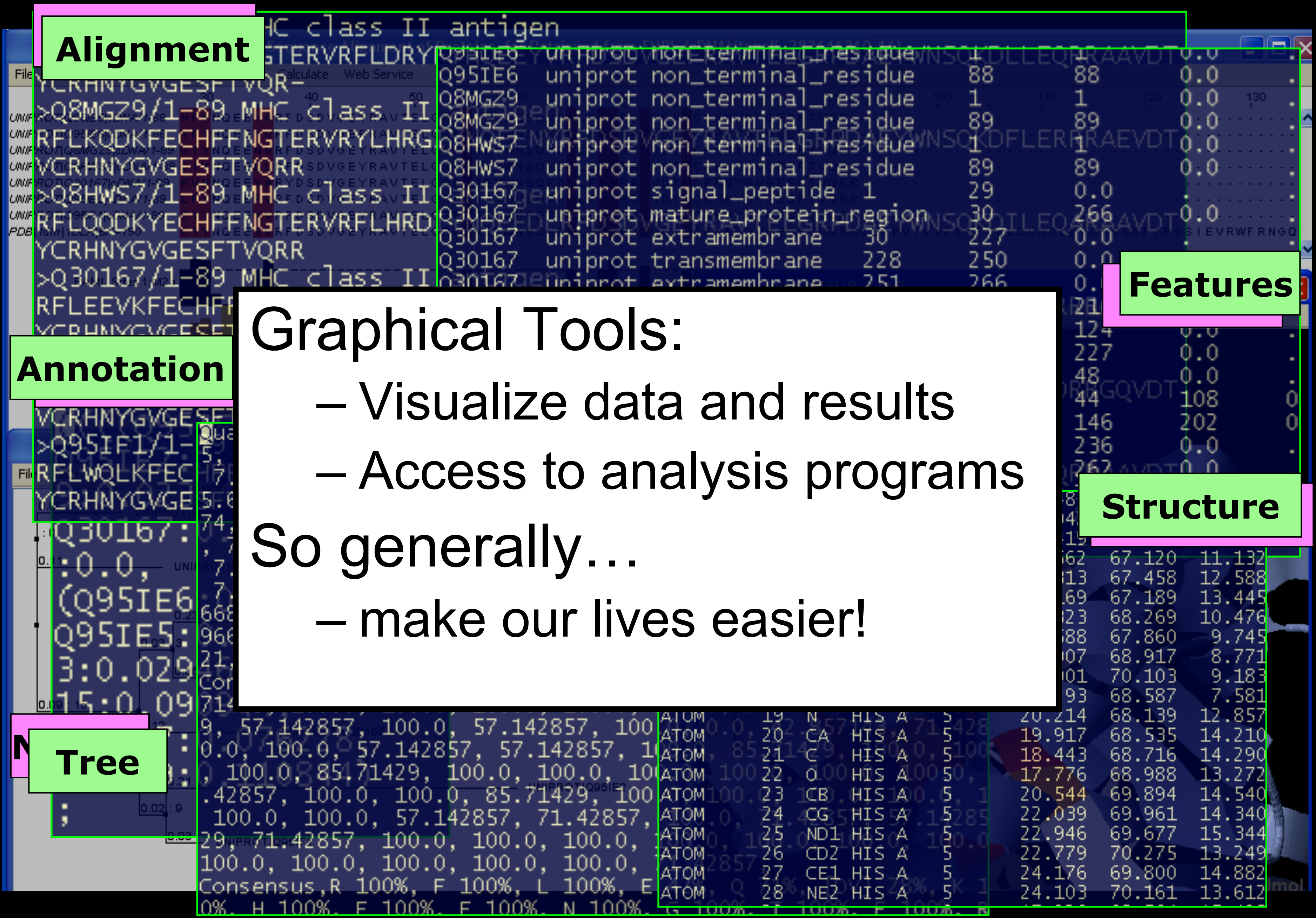
```
HC class II antigen
GTERVRFLDRY
YCRHNYGVGESFTVQR-
>Q8MGZ9/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
VCRHNYGVGESFTVQRR
>Q8HWS7/1-89 MHC class II
RFLQQDKYECHFFNGTERVRFLHRD
YCRHNYGVGESFTVQRR
>Q30167/1-89 MHC class II
RFLEEVKFECHFFNGTERVRLLERR
YCRHNYGVGESFTVQRR
>Q95IE2/1-89 MHC class II
RFLWQGKYKCHFFNGTERVQFLERL
```

```
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
Q30167  uniprot polypeptide_domain      126     216     0.0     .
Q30167  uniprot polypeptide_region      30      124     0.0     .
Q30167  uniprot polypeptide_region      125     227     0.0     .
```

```
((((Q95IF1:0.033892
Q95IE2:0.309176)
9:0.023547,
Q30167:0.11764)
:0.0,
(Q95IE6:0.058815,
Q95IE5:0.233569)
3:0.029464)
15:0.090944,
Q8HWS7:0.076228,
Q8MGZ9:0.110844)
;
```

```
57.142857, 100
100.0, 100.0, 100.0, 100.0, 100.0,
Consensus,R 100%, F 100%, L 100%, E
```

```
ATOM     14  CB  GLU A   4      22.323  68.269  10.476
ATOM     15  CG  GLU A   4      23.588  67.860   9.745
ATOM     16  CD  GLU A   4      24.007  68.917   8.771
ATOM     17  OE1 GLU A   4      24.001  70.103   9.183
ATOM     18  OE2 GLU A   4      24.293  68.587   7.581
ATOM     19  N   HIS A   5      20.214  68.139  12.857
ATOM     20  CA  HIS A   5      19.917  68.535  14.210
ATOM     21  C   HIS A   5      18.443  68.716  14.290
ATOM     22  O   HIS A   5      17.776  68.988  13.272
ATOM     23  CB  HIS A   5      20.544  69.894  14.540
ATOM     24  CG  HIS A   5      22.039  69.961  14.340
ATOM     25  ND1 HIS A   5      22.946  69.677  15.344
ATOM     26  CD2 HIS A   5      22.779  70.275  13.249
ATOM     27  CE1 HIS A   5      24.176  69.800  14.882
ATOM     28  NE2 HIS A   5      24.103  70.161  13.612
```



Graphical Tools:

- Visualize data and results
- Access to analysis programs

So generally…

- make our lives easier!

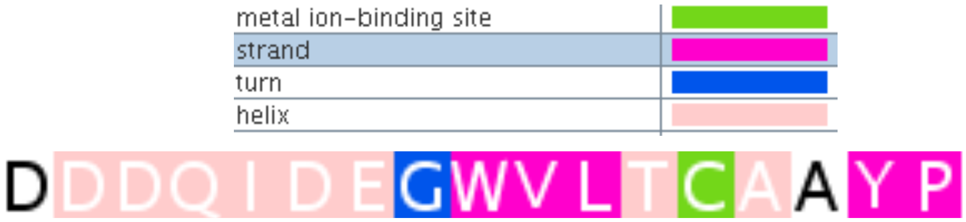
1D Protein Visualization

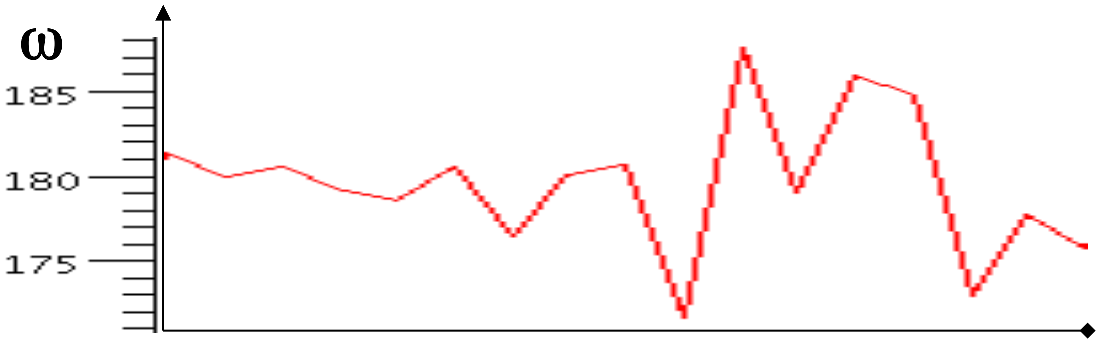
- Visualize information at each position in sequence
- Amino Acid Properties
  - Discrete colour schemes
  - Quantitative Scales
    - Hydrophobicity, polarity, size.
- Protein Features
  - Discrete annotations
  - Observations & Predictions
  - Measurements

    e.g. Omega torsion

    (reflects planarity

    of peptide bond)





Multiple Alignments

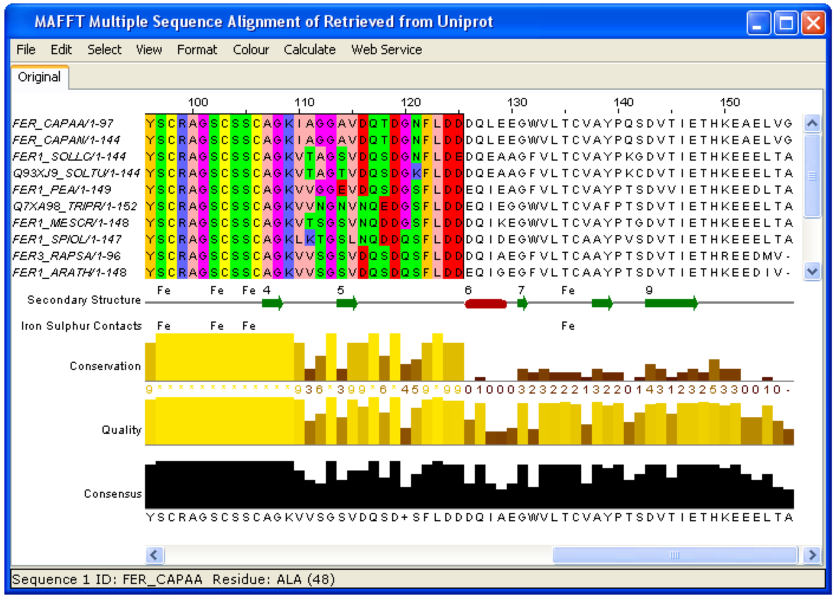
- Generated by heuristic optimization algorithm
  - Alignments sometimes need manual curation
- Correct alignments conserve common properties
- Shading can highlight differences

```
FER_CAPAA/28-82      AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLTCVAYPQ
FER_CAPAN/75-129     AEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLTCVAYPQ
FER1_LYCES/75-129    AEEEGHDLPYSCRAGSCSSCAGKVTAGSVDQSDGNFLDEDQEAAGFVLTCVAYPK
Q93XJ9_SOLTU/75-129  AEEEGHDLPYSCRAGSCSSCAGKVTAGTVDQSDGKFLDDDQEAAGFVLTCVAYPK
FER1_PEA/80-134      AEEVGIDLPYSCRAGSCSSCAGKVVGGEVDQSDGSFLDDEQIEAGFVLTCVAYPT
Q7XA98_TRIPR/83-137  AEEVGIELPYSCRAGSCSSCAGKVVNGNVNQEDGSFLDDEQIEGGWVLTCVAFPT
FER1_MESCR/79-133    AEEAGIDLPYSCRAGSCSSCAGKVTSGSVNQDDGSFLDDDQIKEGWVLTCVAYPT
FER1_SPIOL/78-132    AEEEGIDLPYSCRAGSCSSCAGKLKTGSLNQDDQSFLDDDQIDEGWVLTCAAYPV
FER3_RAPSA/28-82     AEEAGIDLPYSCRAGSCSSCAGKVVSGSVDQSDQSFLDDDQIAEGFVLTCAAYPT
FER_ARATH/80-134     AEEAGIDLPYSCRAGSCSSCAGKVVSGSVDQSDQSFLDDEQIGEGFVLTCAAYPT
FER_BRANA/28-82      AEEAGIDLPYSCRAGSCSSCAGKVVSGFVDQSDESFLDDDQIAEGFVLTCAAYPT
FER2_ARATH/80-134    AEEAGLDLPYSCRAGSCSSCAGKVVSGSIDQSDQSFLDDEQMSEGYVLTCVAYPT
FER1_MAIZE/80-134    AEEDGIDLPYSCRAGSCSSCAGKVVSGSVDQSDQSYLDDGQIADGWVLTCHAYPT
O80429_MAIZE/72-126  AEEEGIDLPFSCRAGSCSSCAGKVVSGSVDQSDQSFLNDNQVADGWVLTCAAYPT
```

What is Jalview ?

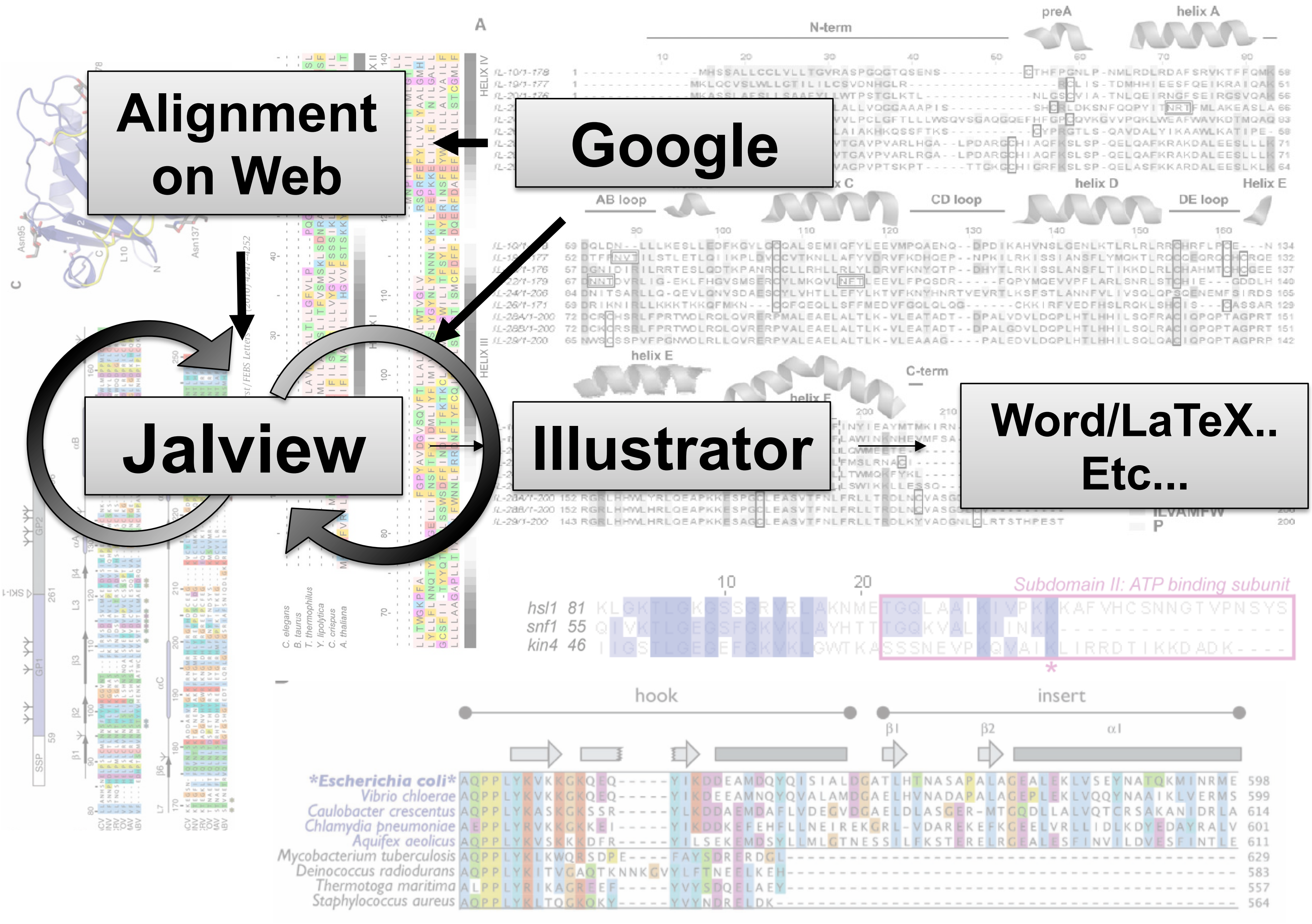
- A java alignment viewer
  - ***j***ava ***al***ignment ***view***er

    .. but not just for viewing..
- Java ?
  - – Programming language
    - Platform independence
    - Standalone or web based tool



```
FER_CAPAA/1-97      YSCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLTCVAYPQSDVTIETHKEAELVG
FER_CAPAN/1-144     YSCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLTCVAYPQSDVTIETHKEAELVG
FER1_SOLLC/1-144    YSCRAGSCSSCAGKVTAGSVDQSDGNFLDEDQEAAGFVLTCVAYPKGDVTIETHKEEELTA
Q93XJ9_SOLTU/1-144  YSCRAGSCSSCAGKVTAGTVDQSDGKFLDDDQEAAGFVLTCVAYPKCDVTIETHKEEELTA
FER1_PEA/1-149      YSCRAGSCSSCAGKVVGGEVDQSDGSFLDDEQIEAGFVLTCVAYPTSDVVIETHKEEDLTA
Q7XA98_TRIPR/1-152  YSCRAGSCSSCAGKVVNGNVNQEDGSFLDDEQIEGGWVLTCVAFPTSDVTIETHKEEELTA
FER1_MESCR/1-148    YSCRAGSCSSCAGKVTSGSVNQDDGSFLDDDQIKEGWVLTCVAYPTGDVTIETHKEEELTA
FER1_SPIOL/1-147    YSCRAGSCSSCAGKLKTGSLNQDDQSFLDDDQIDEGWVLTCAAYPVSDVTIETHKEEELTA
FER3_RAPSA/1-96     YSCRAGSCSSCAGKVVSGSVDQSDQSFLDDDQIAEGFVLTCAAYPTSDVTIETHREEDMV-
FER1_ARATH/1-148    YSCRAGSCSSCAGKVVSGSVDQSDQSFLDDEQIGEGFVLTCAAYPTSDVTIETHKEEDIV-

Conservation        9*************936*399*6*459*990100032322213220143123253300010-
Consensus           YSCRAGSCSSCAGKVVSGSVDQSD+SFLDDDQIAEGWVLTCVAYPTSDVTIETHKEEELTA
```

Sequence 1 ID: FER_CAPAA Residue: ALA (48)



Alignment on Web

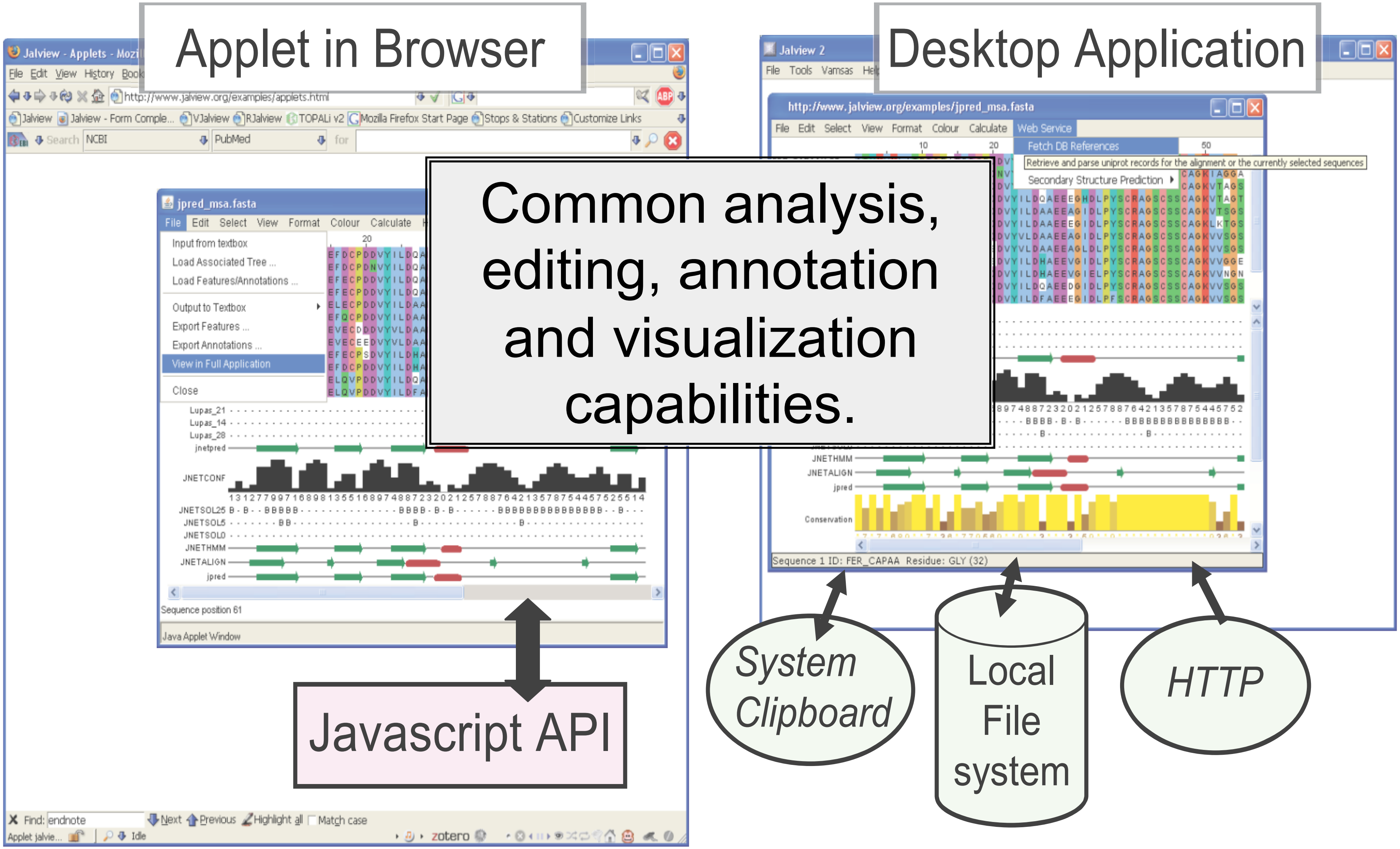
Google

Jalview

Illustrator

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

Jalview Flavours



Applet in Browser

Desktop Application

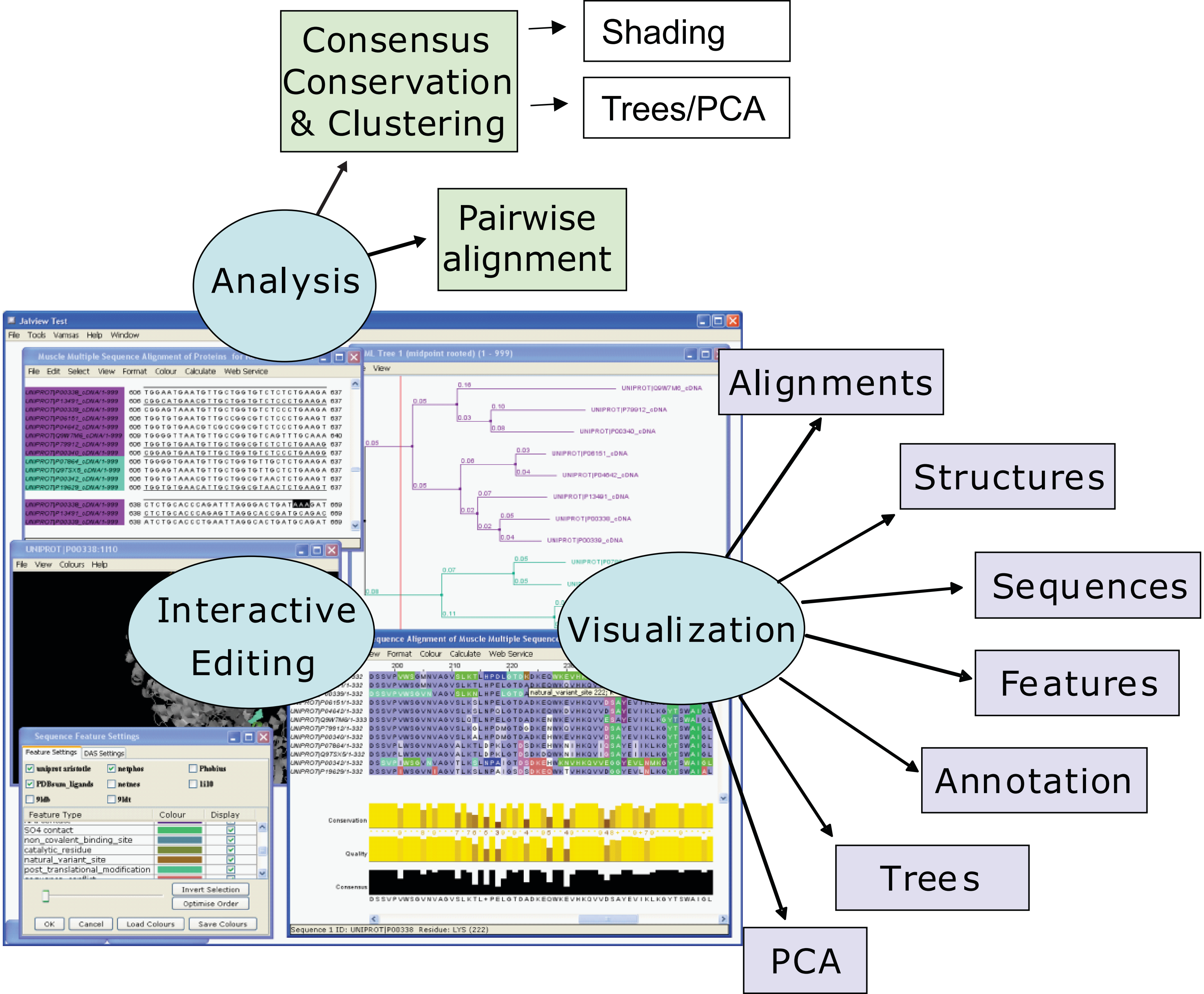
Common analysis, editing, annotation and visualization capabilities.

Javascript API

System Clipboard

Local File system

HTTP



Analysis

Consensus Conservation & Clustering

Shading

Trees/PCA

Pairwise alignment

Interactive Editing

Visualization

Alignments

Structures

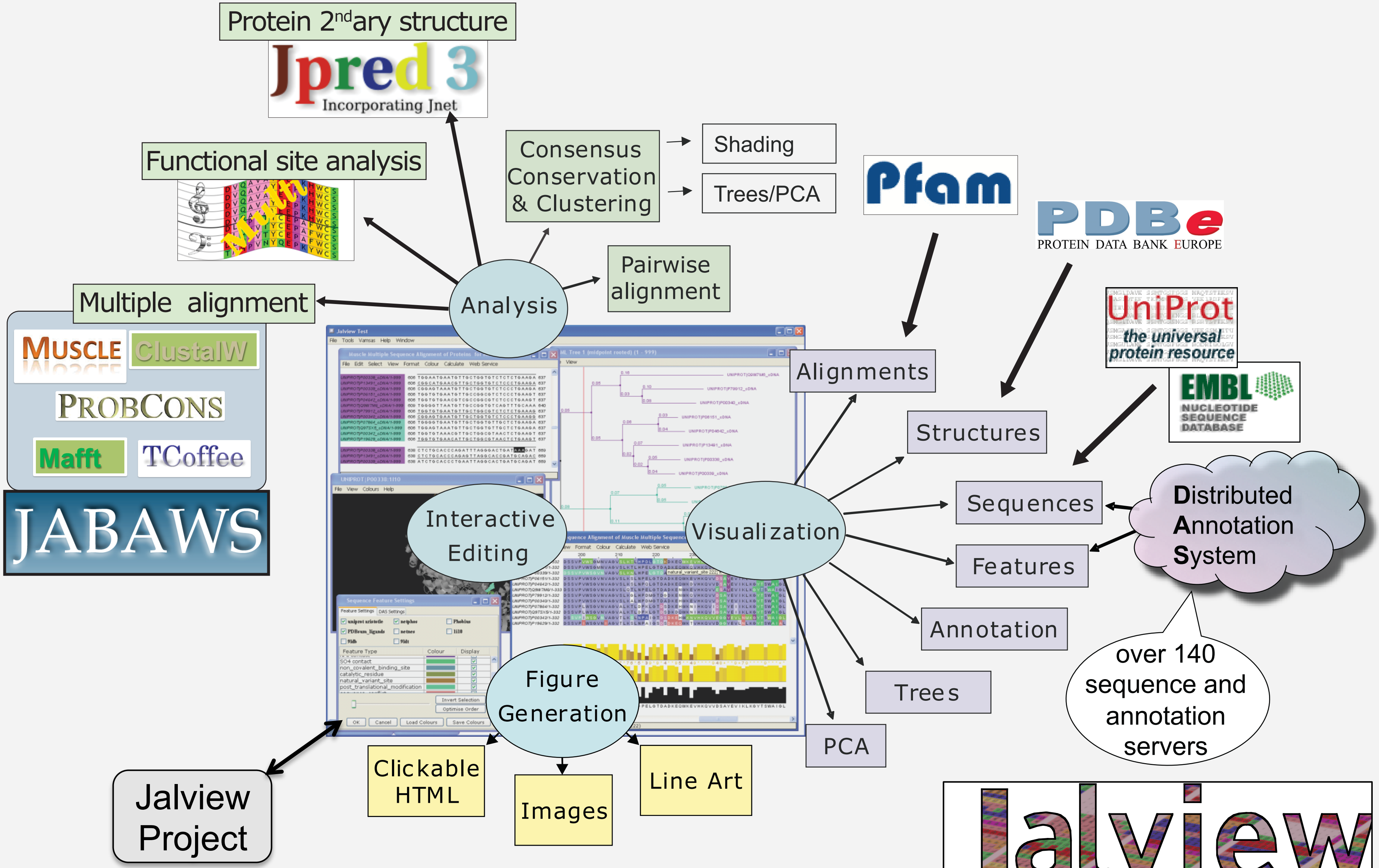
Sequences

Features

Annotation

Trees

PCA



Protein 2nd ary structure

Jpred 3
Incorporating Jnet

Functional site analysis

Multiple alignment

MUSCLE
ClustalW
PROBCONS
Mafft
TCoffee

JABAWS

Analysis

Consensus
Conservation
& Clustering

Shading

Trees/PCA

Pairwise
alignment

Pfam

PDBe
PROTEIN DATA BANK EUROPE

UniProt
the universal protein resource

EMBL
NUCLEOTIDE SEQUENCE DATABASE

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Visualization

Interactive
Editing

Figure
Generation

Distributed
Annotation
System

over 140
sequence and
annotation
servers

Clickable
HTML

Images

Line Art

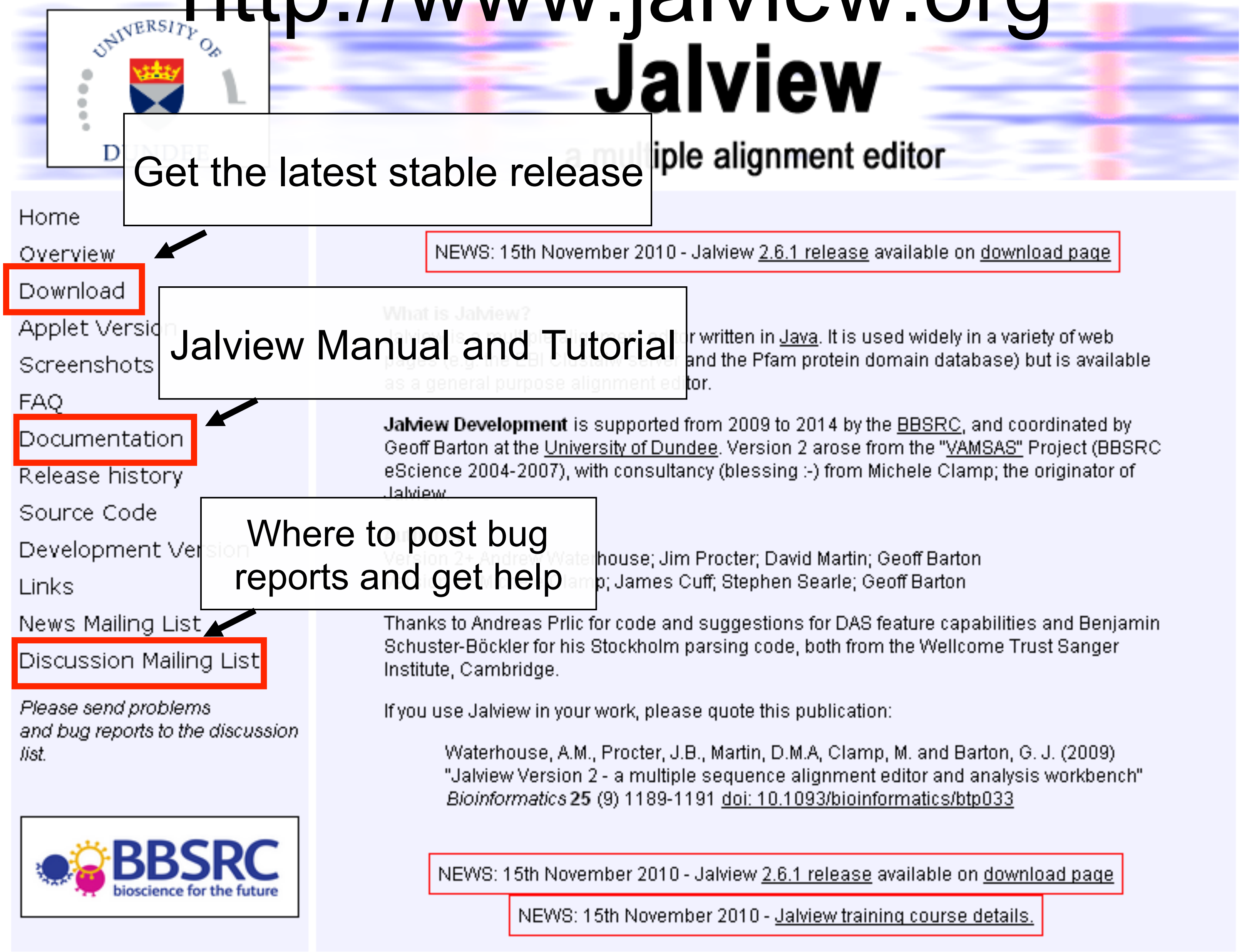
Jalview
Project

Jalview 2.7

Ex 1 – starting Jalview

For today – use the latest build of v2.7 from the release branch.

http://www.jalview.org



Get the latest stable release

Jalview Manual and Tutorial

Where to post bug reports and get help

- Home
- Overview
- Download
- Applet Version
- Screenshots
- FAQ
- Documentation
- Release history
- Source Code
- Development Version
- Links
- News Mailing List
- Discussion Mailing List

Please send problems and bug reports to the discussion list.

NEWS: 15th November 2010 - Jalview 2.6.1 release available on download page

What is Jalview?

Jalview is a multiple alignment editor written in Java. It is used widely in a variety of web pages (e.g. the EBI Clustalw server and the Pfam protein domain database) but is available as a general purpose alignment editor.

Jalview Development is supported from 2009 to 2014 by the BBSRC, and coordinated by Geoff Barton at the University of Dundee. Version 2 arose from the "VAMSAS" Project (BBSRC eScience 2004-2007), with consultancy (blessing :-) from Michele Clamp; the originator of Jalview.

Authors

Version 2+ Andrew Waterhouse; Jim Procter; David Martin; Geoff Barton
Version 1: Michele Clamp; James Cuff; Stephen Searle; Geoff Barton

Thanks to Andreas Prlic for code and suggestions for DAS feature capabilities and Benjamin Schuster-Böckler for his Stockholm parsing code, both from the Wellcome Trust Sanger Institute, Cambridge.

If you use Jalview in your work, please quote this publication:

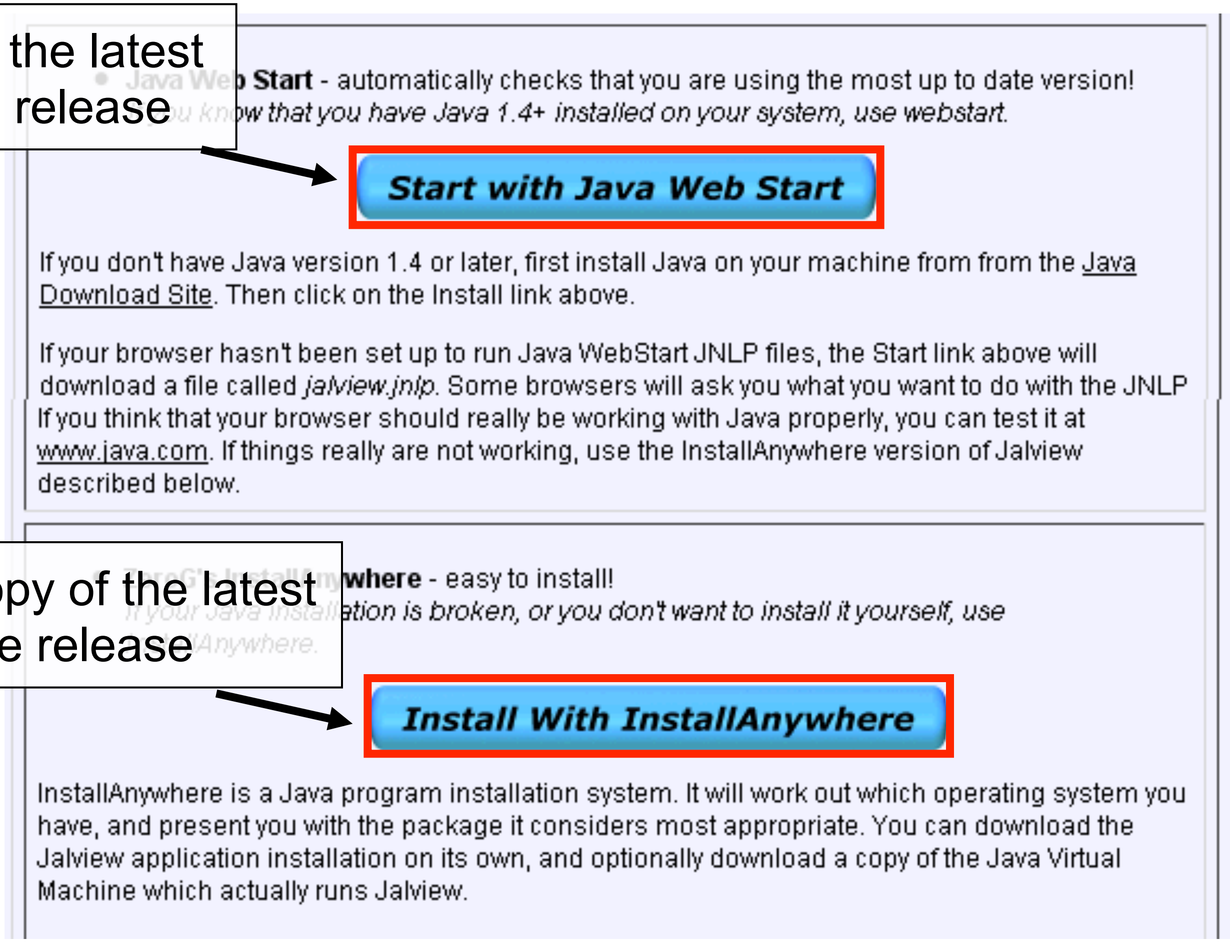
> Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench" *Bioinformatics* **25** (9) 1189-1191 doi: 10.1093/bioinformatics/btp033

NEWS: 15th November 2010 - Jalview 2.6.1 release available on download page

NEWS: 15th November 2010 - Jalview training course details.

Starting The Jalview Desktop

http://www.jalview.org/download.html



Launch the latest stable release

- **Java Web Start** - automatically checks that you are using the most up to date version!
  ou know that you have Java 1.4+ installed on your system, use webstart.

If you don't have Java version 1.4 or later, first install Java on your machine from from the Java Download Site. Then click on the Install link above.

If your browser hasn't been set up to run Java WebStart JNLP files, the Start link above will download a file called *jalview.jnlp*. Some browsers will ask you what you want to do with the JNLP

If you think that your browser should really be working with Java properly, you can test it at www.java.com. If things really are not working, use the InstallAnywhere version of Jalview described below.

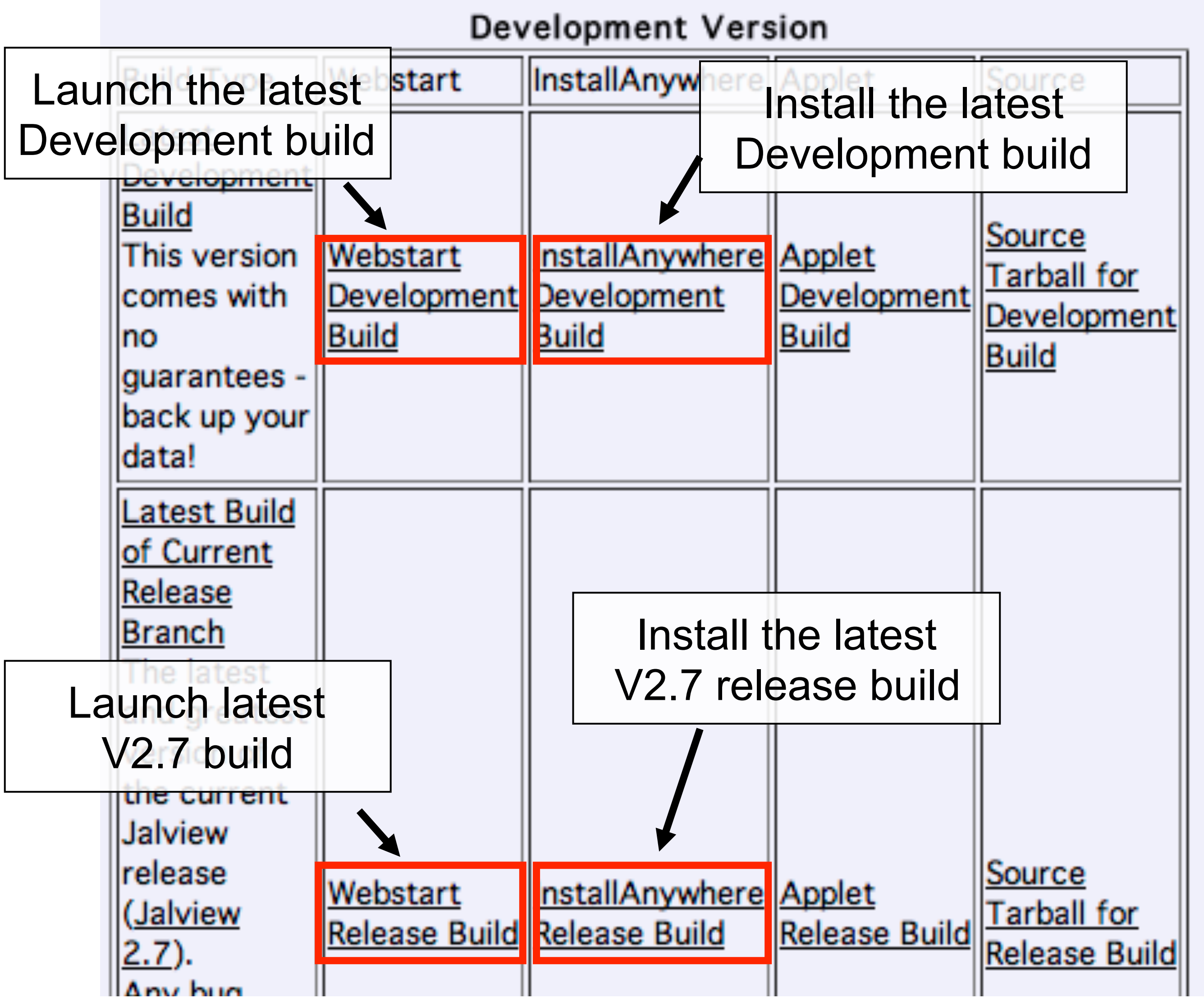
Install a copy of the latest stable release

- **ZeroG InstallAnywhere** - easy to install!
  If your Java installation is broken, or you don't want to install it yourself, use InstallAnywhere.

InstallAnywhere is a Java program installation system. It will work out which operating system you have, and present you with the package it considers most appropriate. You can download the Jalview application installation on its own, and optionally download a copy of the Java Virtual Machine which actually runs Jalview.

Starting The development builds

http://www.jalview.org/versions.html



Development Version

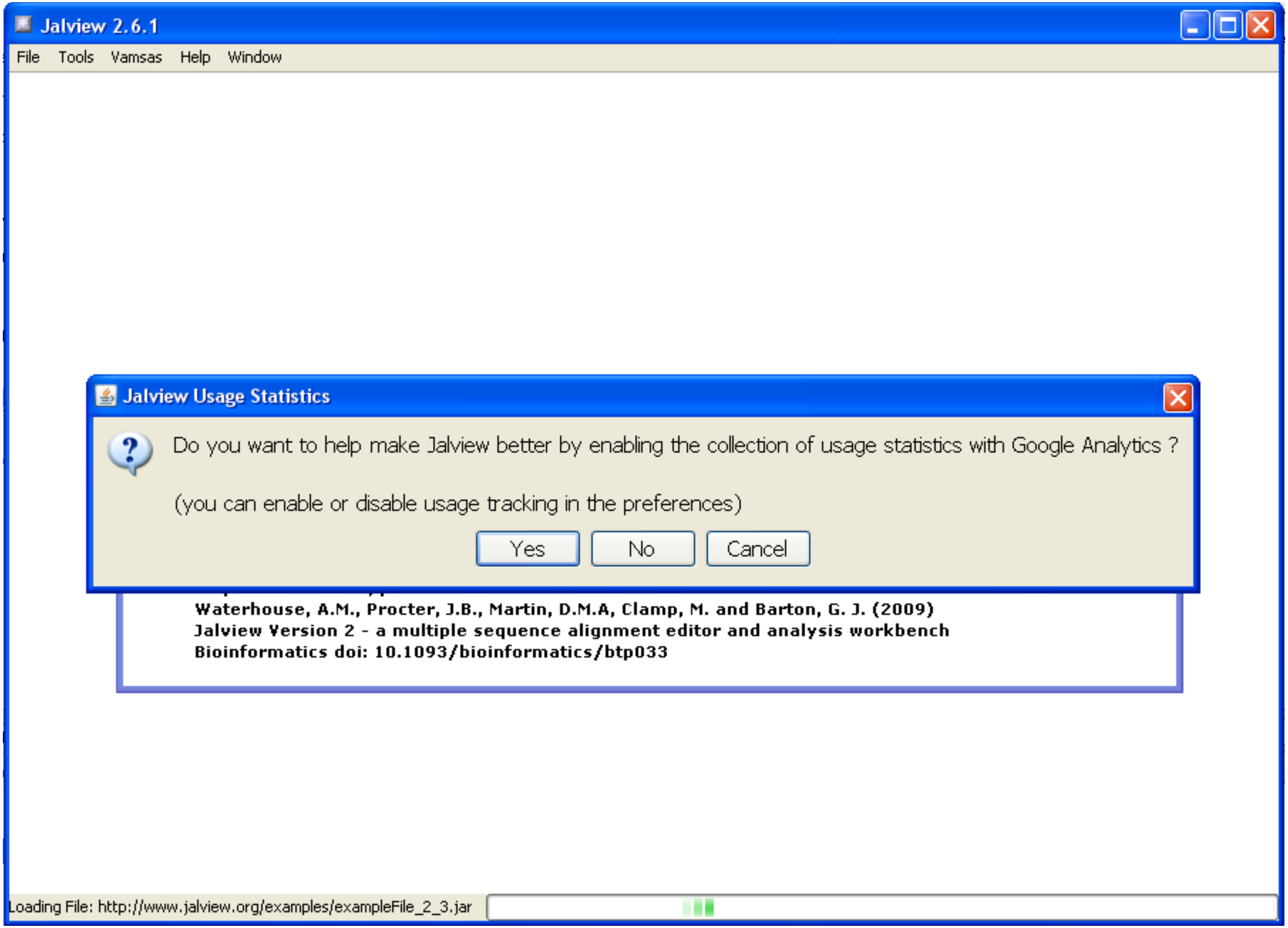
Build Type	Webstart	InstallAnywhere	Applet	Source
Latest Development Build This version comes with no guarantees - back up your data!	Webstart Development Build	InstallAnywhere Development Build	Applet Development Build	Source Tarball for Development Build
Latest Build of Current Release Branch The latest and greatest version of the current Jalview release (Jalview 2.7). Any bug	Webstart Release Build	InstallAnywhere Release Build	Applet Release Build	Source Tarball for Release Build

Launch the latest Development build

Install the latest Development build

Launch latest V2.7 build

Install the latest V2.7 release build



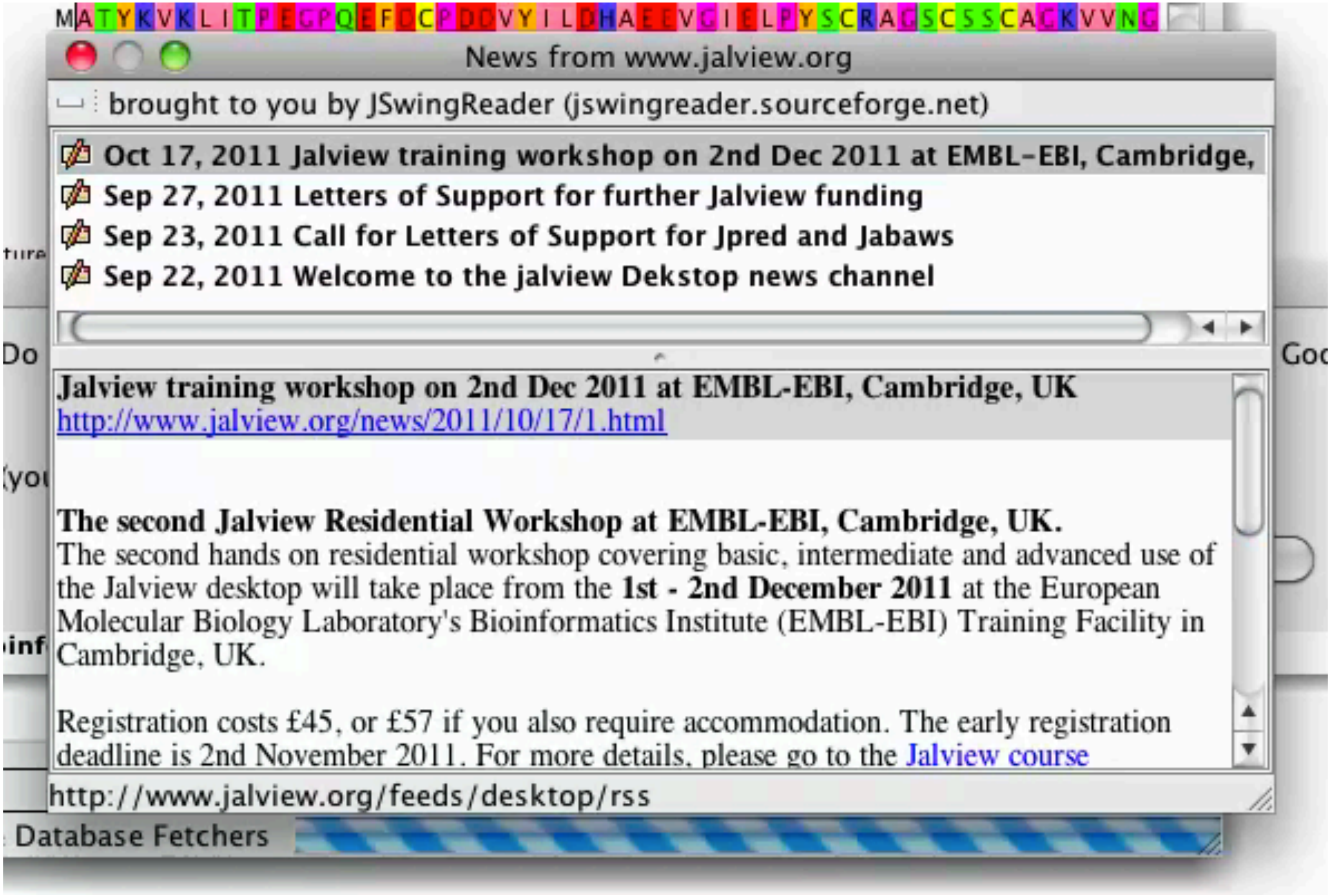
Do you want to help make Jalview better by enabling the collection of usage statistics with Google Analytics ?

(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

Loading File: http://www.jalview.org/examples/exampleFile_2_3.jar

Jalview News



News from www.jalview.org

brought to you by JSwingReader (jswingreader.sourceforge.net)

- Oct 17, 2011 Jalview training workshop on 2nd Dec 2011 at EMBL-EBI, Cambridge,
- Sep 27, 2011 Letters of Support for further Jalview funding
- Sep 23, 2011 Call for Letters of Support for Jpred and Jabaws
- Sep 22, 2011 Welcome to the jalview Dekstop news channel

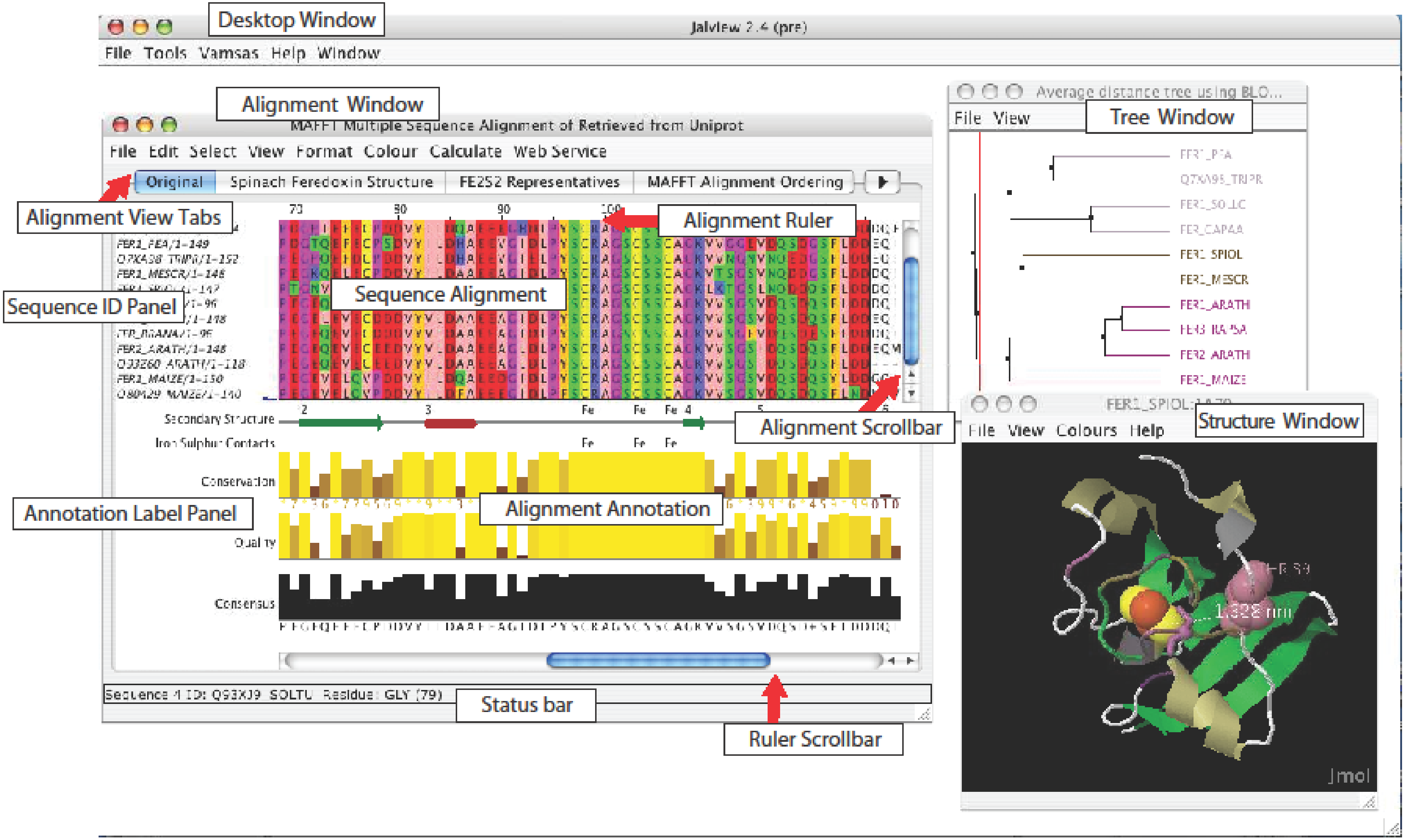
Jalview training workshop on 2nd Dec 2011 at EMBL-EBI, Cambridge, UK
http://www.jalview.org/news/2011/10/17/1.html

The second Jalview Residential Workshop at EMBL-EBI, Cambridge, UK.
The second hands on residential workshop covering basic, intermediate and advanced use of the Jalview desktop will take place from the **1st - 2nd December 2011** at the European Molecular Biology Laboratory's Bioinformatics Institute (EMBL-EBI) Training Facility in Cambridge, UK.

Registration costs £45, or £57 if you also require accommodation. The early registration deadline is 2nd November 2011. For more details, please go to the Jalview course

http://www.jalview.org/feeds/desktop/rss

Anatomy of Jalview: Figure 1.7



Sequence 4 ID: Q93XJ9_SOLTU Residue: GLY (79)

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/
- Jalview bug database
  - – http://issues.jalview.org
    - Also indexed on google

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

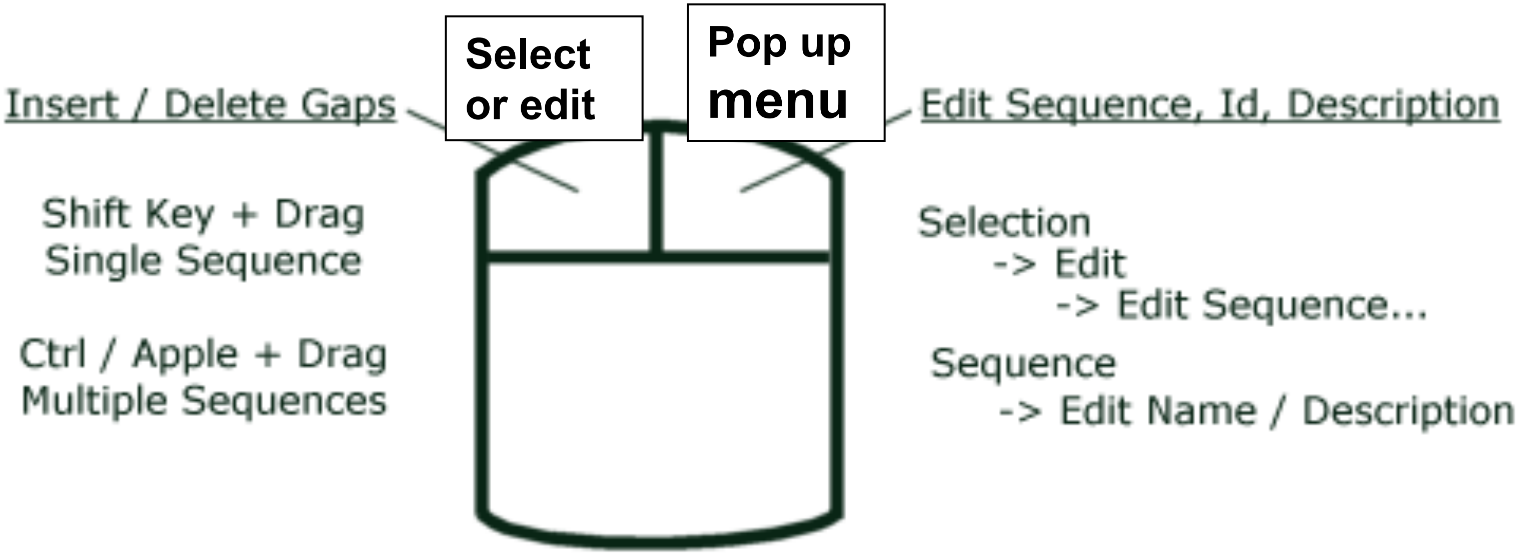
- Tasks
  - Save alignments in different formats
- Questions
  - What’s the biggest difference between a BLC file and a pileup file ?
  - Why are Jalview projects useful ?

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.

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



Select or edit

Pop up menu

Insert / Delete Gaps

Shift Key + Drag
Single Sequence

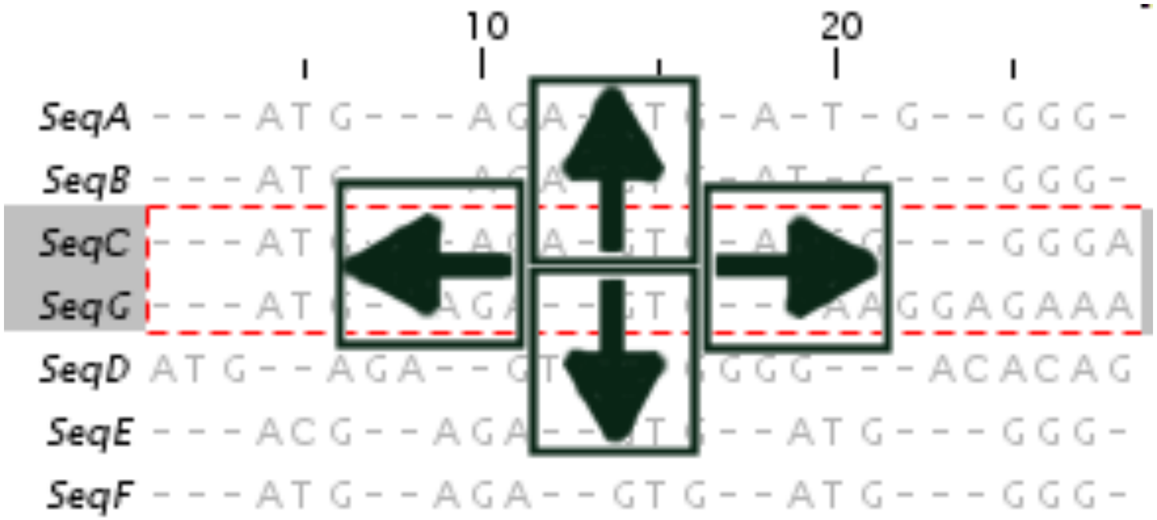
Ctrl / Apple + Drag
Multiple Sequences

Edit Sequence, Id, Description

Selection
-> Edit
-> Edit Sequence...

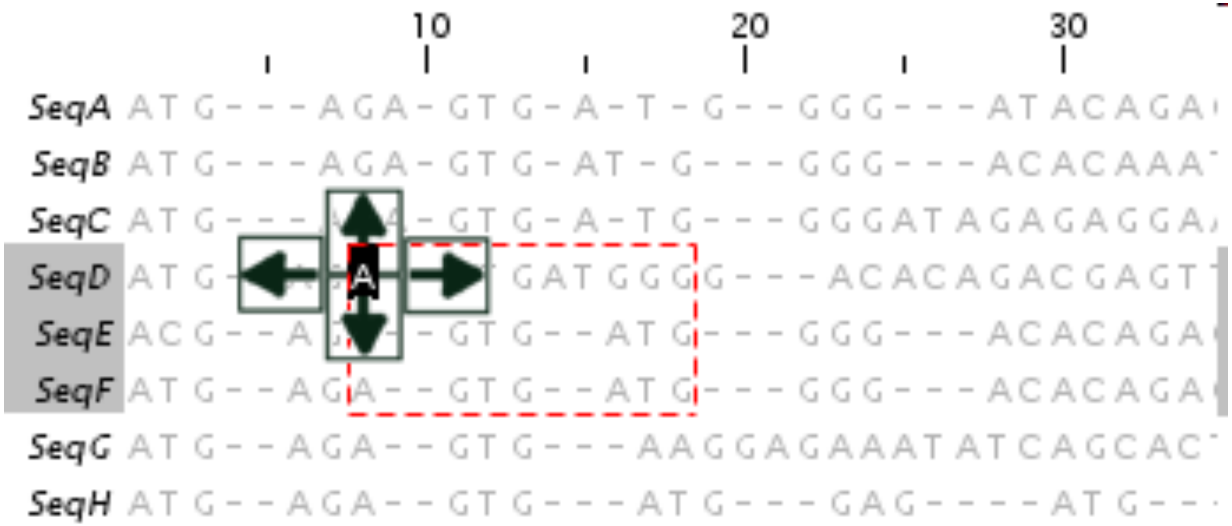
Sequence
-> Edit Name / Description

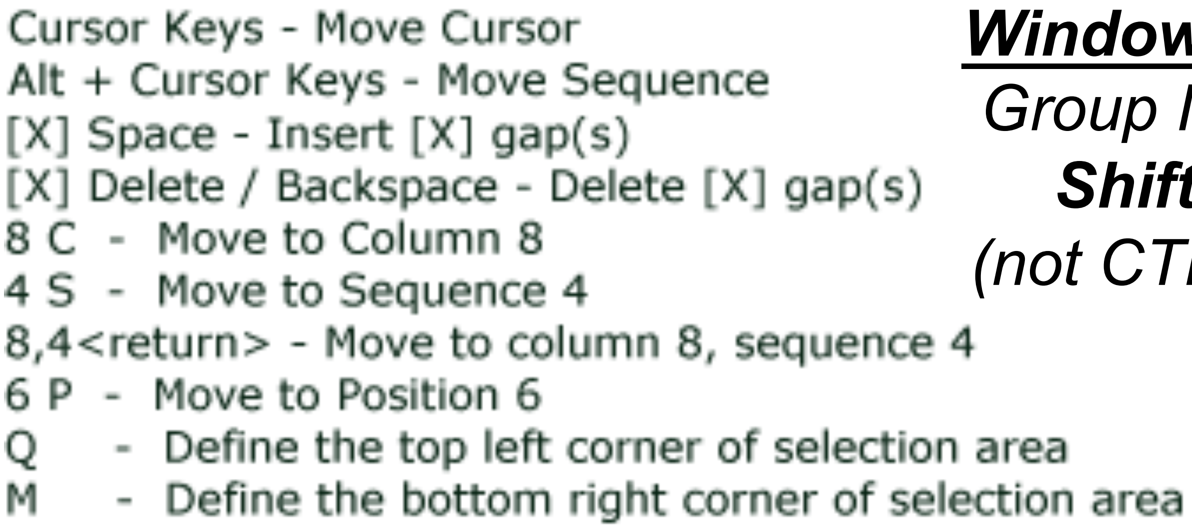
Selected sequences can be moved up and down or slid from left to right



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

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

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 & analysis
  - Annotation & secondary structure prediction
  - Working with structures

Alignment Analysis

Using jalview to analyse the relationships between aligned sequences.

Topics for Session 2

- Alignment analysis – Section 2.2
  - Principal component analysis
  - Phylogenetic trees
  - Redundancy removal
  - Tree based conservation analysis
  - Subdividing alignment by mutation
- Alignment – Section 2.3.3

Phylogenetic analysis and Jalview

- Built in tree methods
  - UPGMA
    - Fast, simple, but not reliable for phylogenetic inferrence
  - 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 feredoxin 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
- 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
- Questions
  - – How can you work out which group is associated with which annotation row ?

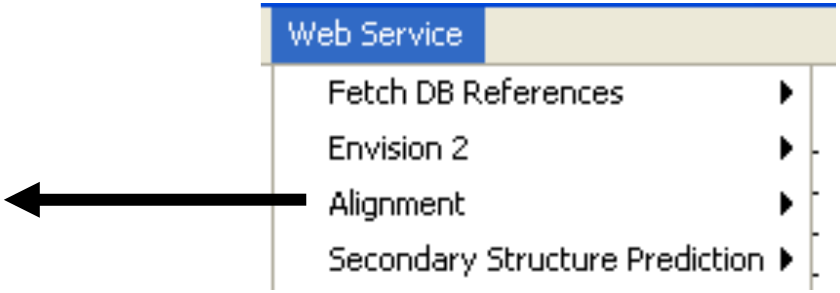
- End of Alignment analysis
  - PCA and Trees
  - Subfamily analysis

- Next – Section 2.3.3
  - Alignment with JABAWS.

PAGE 54 IN MANUAL

Jalview’ s Alignment Methods

- **Needleman and Wunsch Pairwise Alignment**
  - Global alignment of pairs of sequences
    - Mostly used internally (described in section 2.2.7)
- **Multiple Sequence Alignment Services**
  - ClustalW
  - Muscle
  - MAFFT
  - ProbCons
  - T-COFFEE



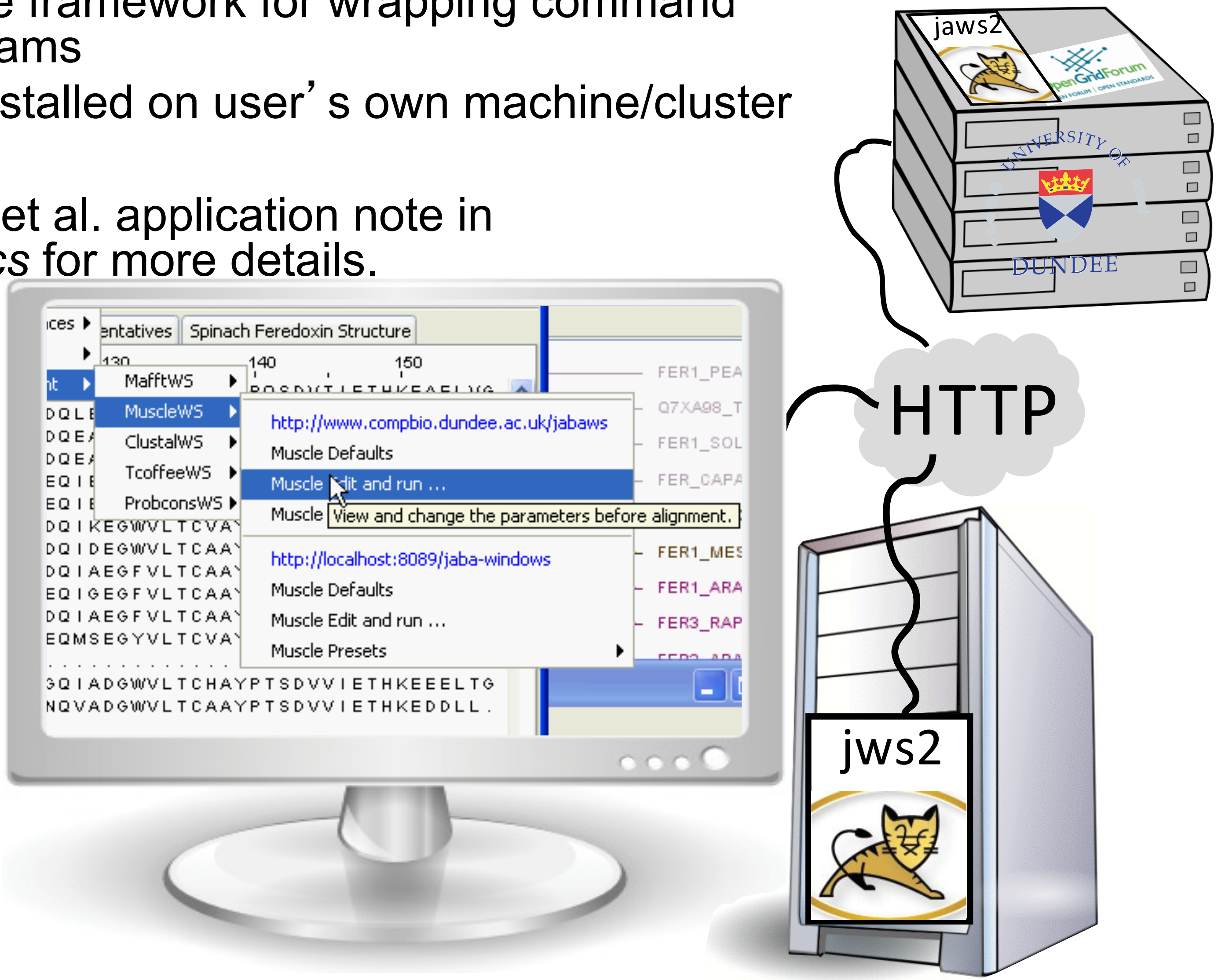
Alignment Web Services: JABAWS:MSA

Replaces original Jalview 2 services:

- Extensible framework for wrapping command line programs
- Can be installed on user's own machine/cluster

See Troshin et al. application note in *Bioinformatics* for more details.

Peter Troshin



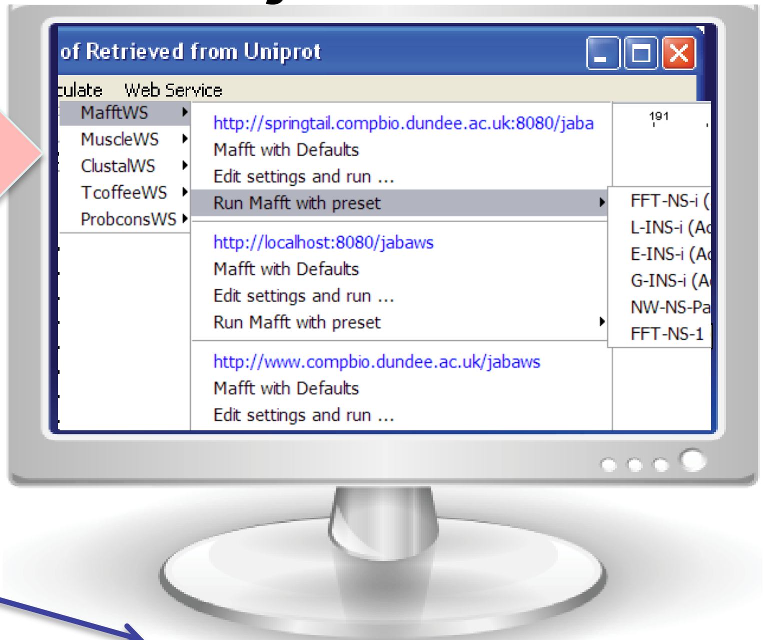
View and change the parameters before alignment.

www.compbio.dundee.ac.uk/jabaws

Troshin et al. 2011, *Bioinformatics.*

Jalview Web Service GUI

JABAWS Java Client



MUSCLE ClustalW PROBCONS Mafft TCoffee

JABAWS

JABAWS Virtual Appliance on user's own machine

VirtualBox vmware

JABAWS tomcat install configured for local or cluster execution

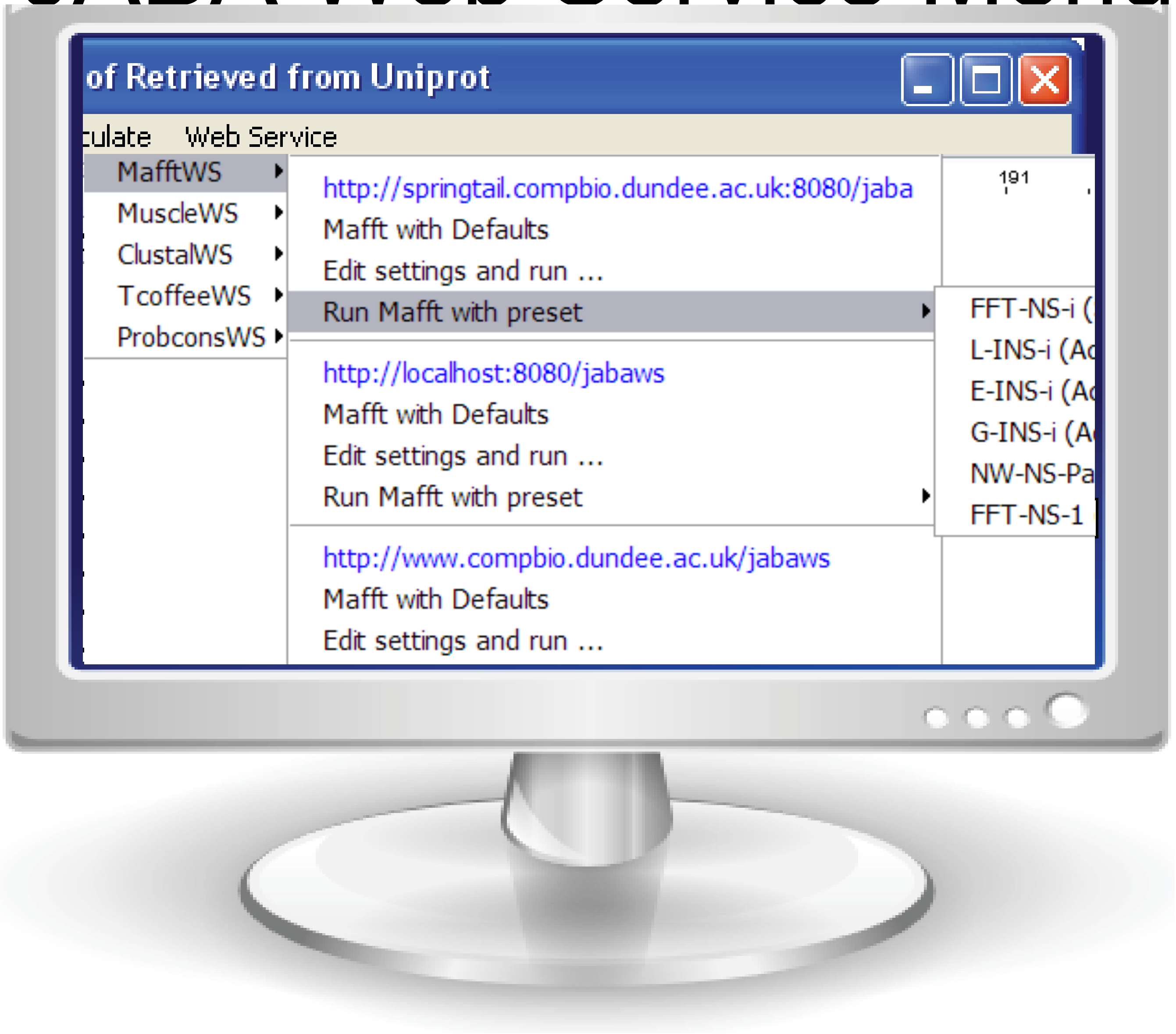
MUSCLE ClustalW PROBCONS Mafft TCoffee

JABAWS

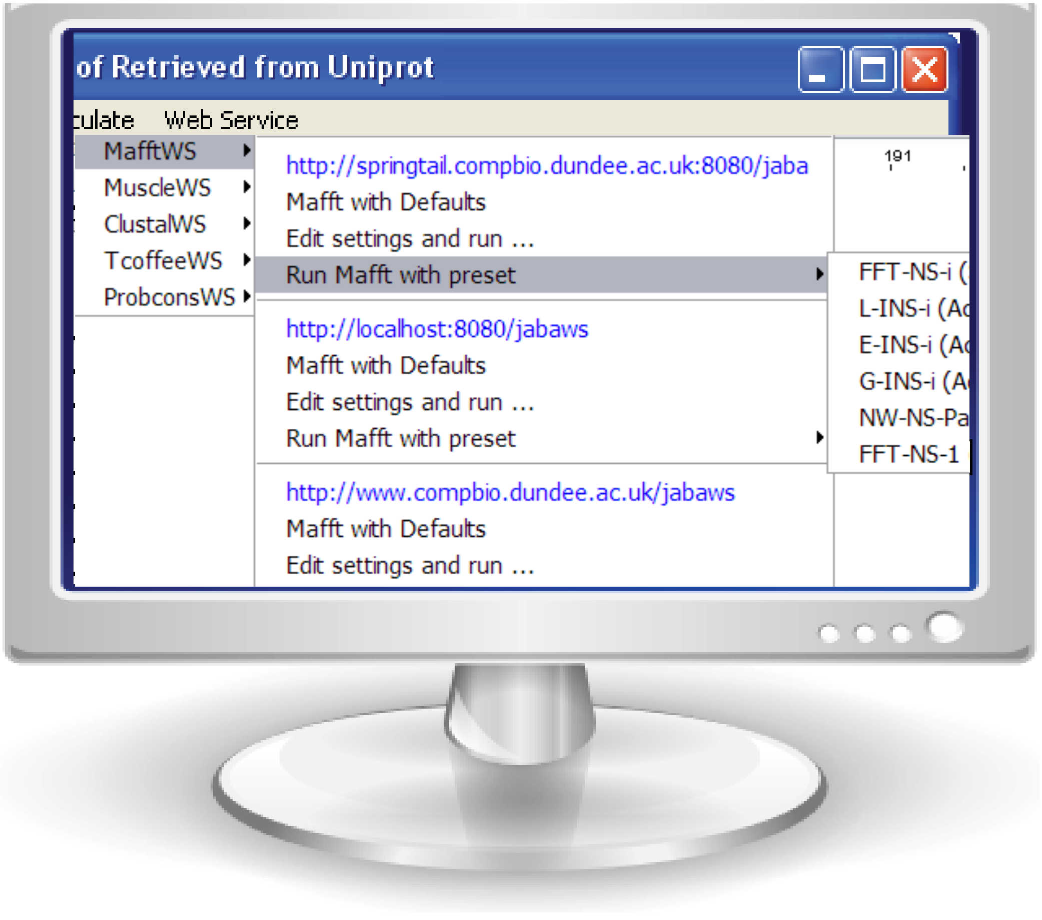
Apache Tomcat

DRMAA
Distributed Resource Management Application API — www.drmaa.org

JABA Web Service Menu



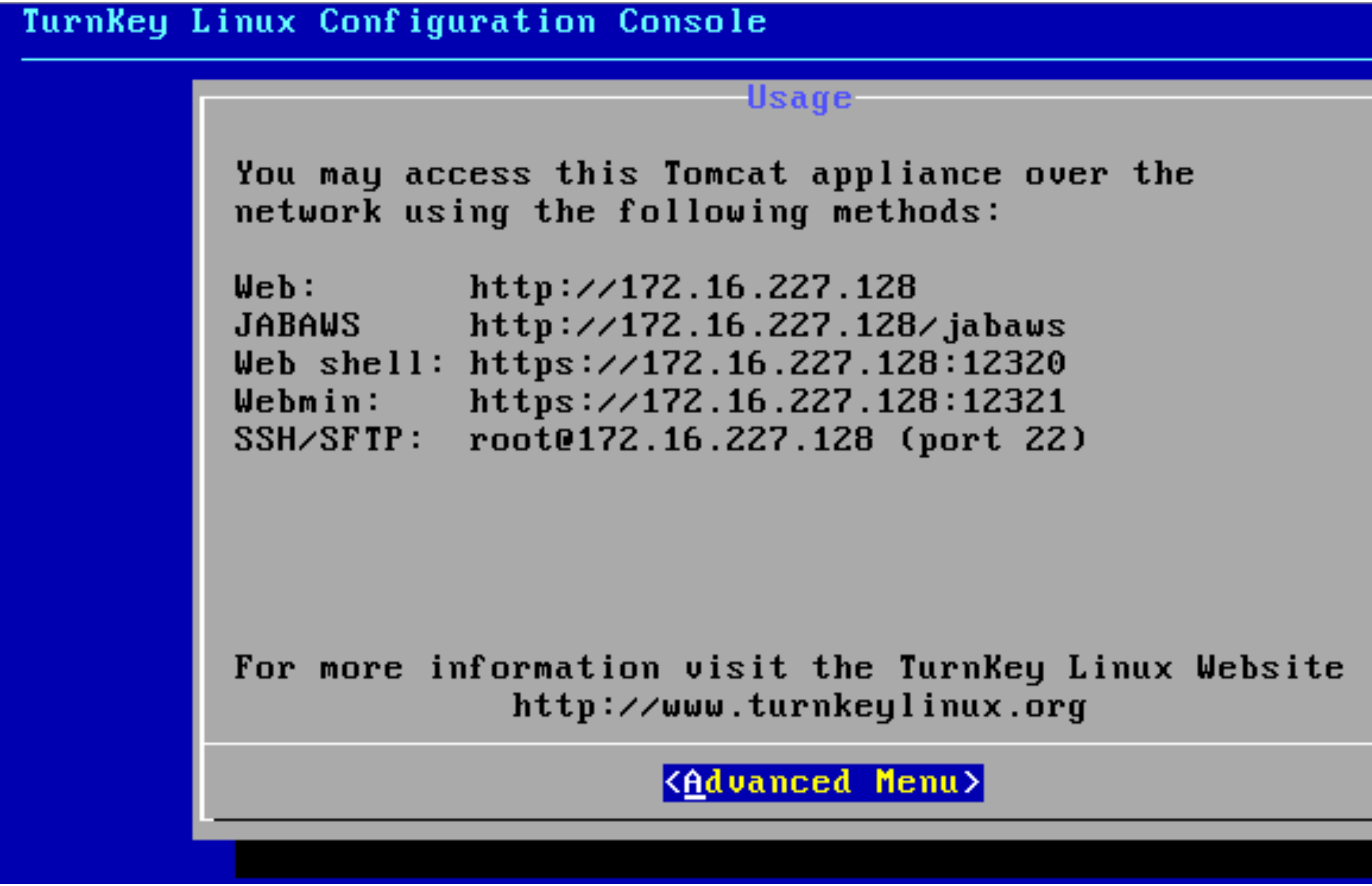
Rearranging the web services menu.. Section 2.3.4 Exercise 22



Installing your own JABA Server
Sec 2.3.5 Ex 23 and 24

NOTE: JABAWS NEEDS A 64Bit MACHINE

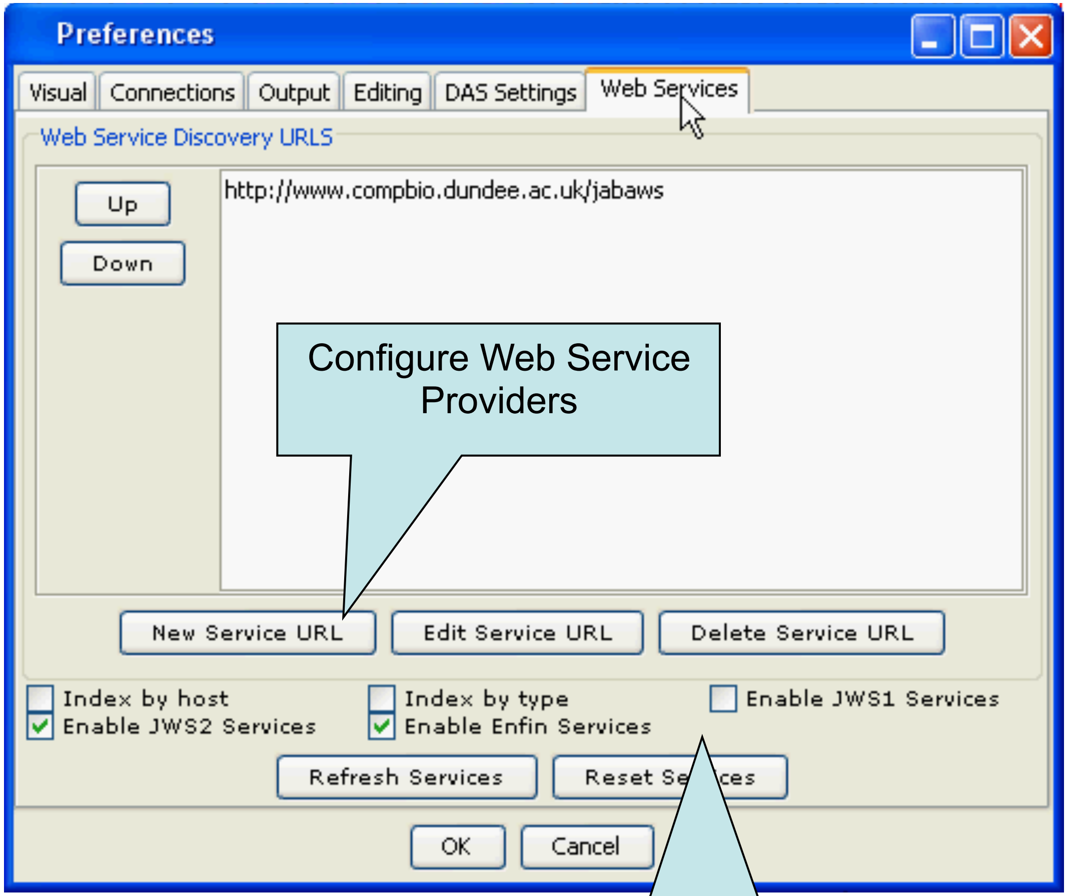
- If you have VMWare Player installed
  - Launch JABAWS on your machine and connect Jalview to it
- If you don’t have VMWare player
  - Connect Jalview to the Kings Buildings JABAWS install (thanks Alastair!)



```
TurnKey Linux Configuration Console

                         Usage

  You may access this Tomcat appliance over the
  network using the following methods:

  Web:        http://172.16.227.128
  JABAWS      http://172.16.227.128/jabaws
  Web shell:  https://172.16.227.128:12320
  Webmin:     https://172.16.227.128:12321
  SSH/SFTP:   root@172.16.227.128 (port 22)




  For more information visit the TurnKey Linux Website
               http://www.turnkeylinux.org
```

Configuring a JABA Server

Web Services Preferences



Web Service Discovery URLS

http://www.compbio.dundee.ac.uk/jabaws

☐ Index by host
☑ Enable JWS2 Services
☐ Index by type
☑ Enable Enfin Services
☐ Enable JWS1 Services

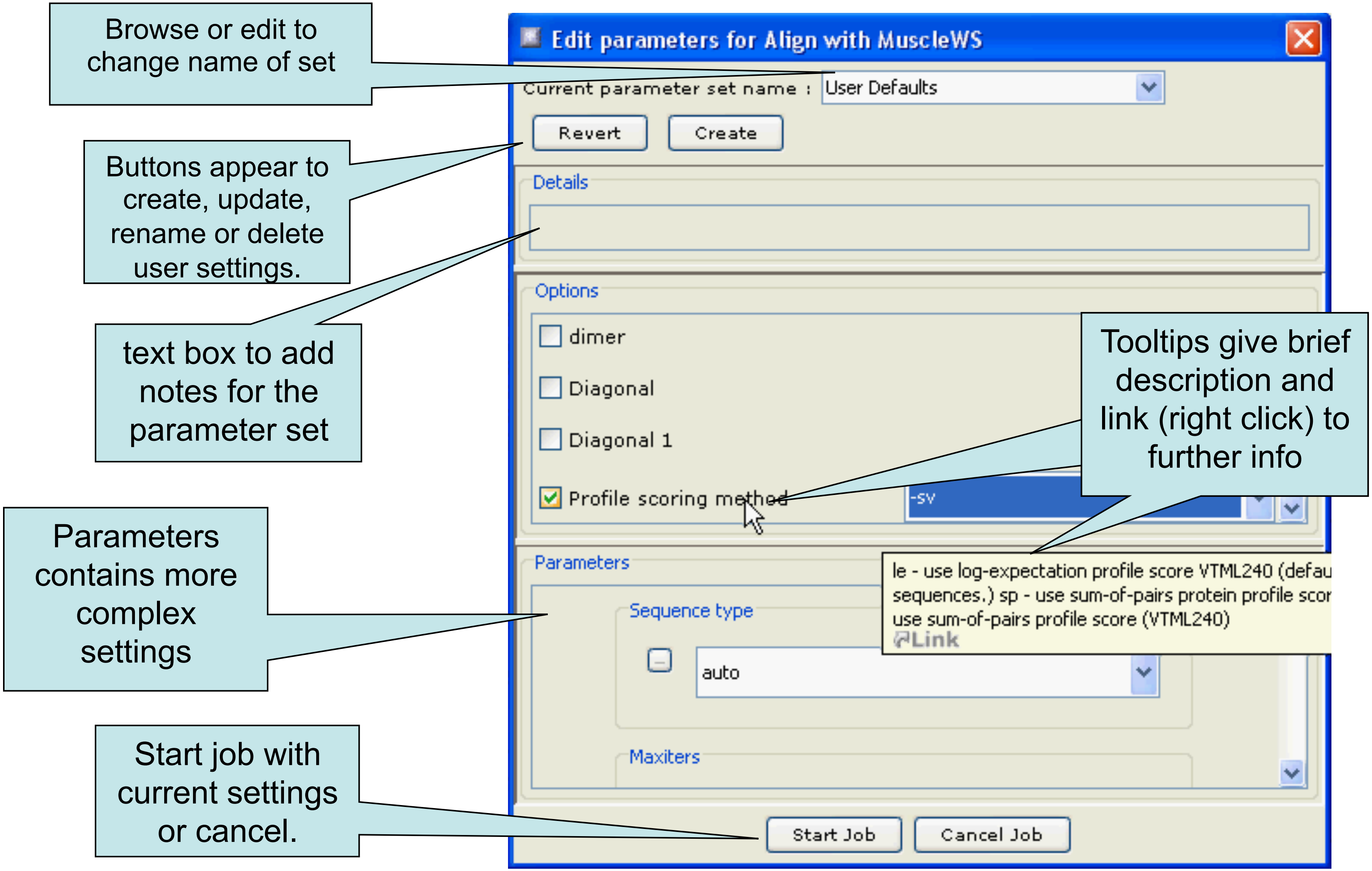
Configure Web Service Providers

Configure web services menu layout

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



Browse or edit to change name of set

Buttons appear to create, update, rename or delete user settings.

text box to add notes for the parameter set

Parameters contains more complex settings

Start job with current settings or cancel.

Tooltips give brief description and link (right click) to further info

Edit parameters for Align with MuscleWS

Current parameter set name : User Defaults

☐ dimer

☐ Diagonal

☐ Diagonal 1

☑ Profile scoring method -sv

Sequence type: auto

le - use log-expectation profile score VTML240 (defau
sequences.) sp - use sum-of-pairs protein profile scor
use sum-of-pairs profile score (VTML240)
Link

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 feredoxin alignment ?

- Next:
  annotation and analysis

Alignment and Sequence Annotation

Section 2.5 onwards

in the Manual

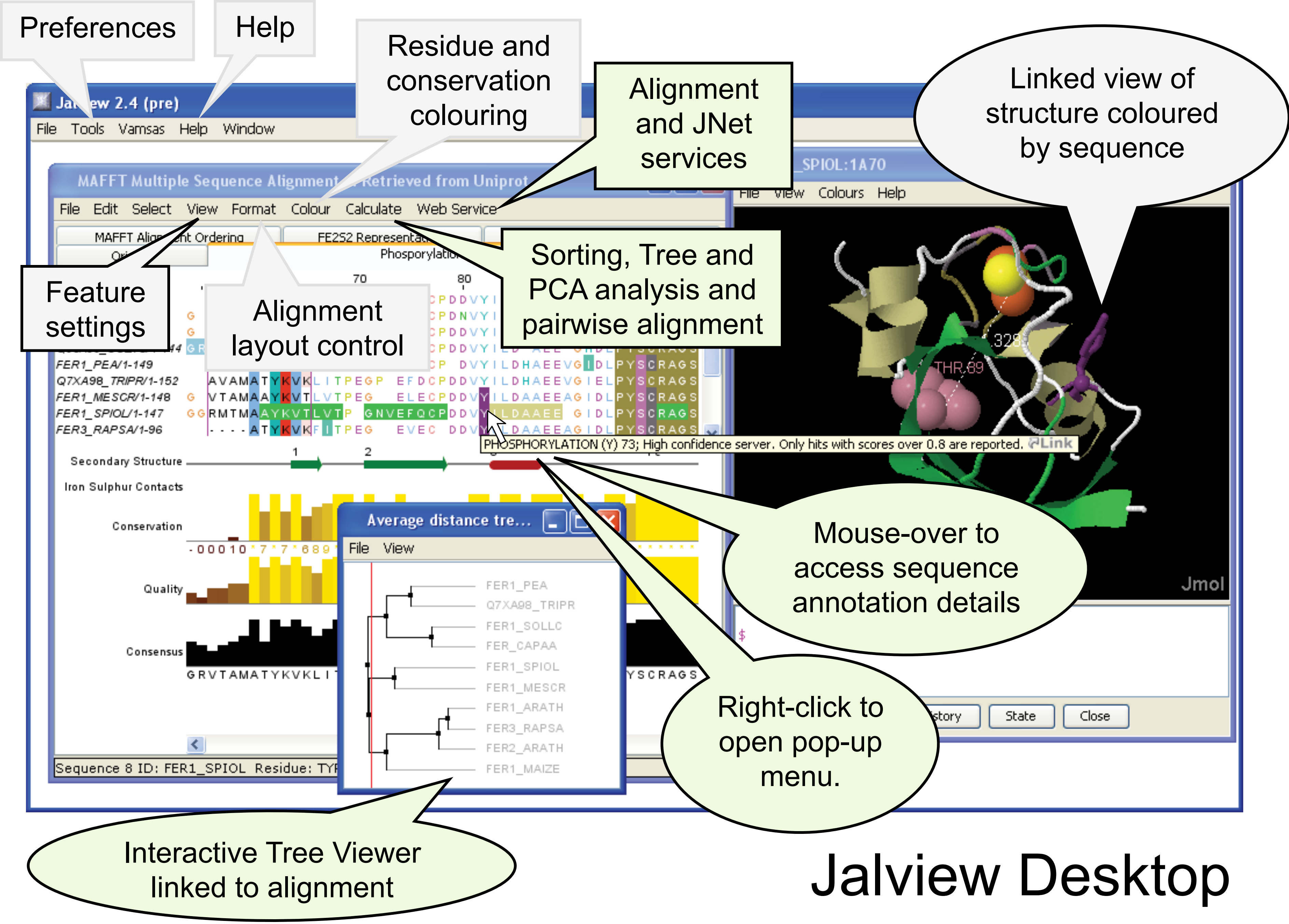
Topics for Session 4

A final meander through the material in the manual:

- 2.5 Protein secondary structure prediction
- 2.6.4 Alignment annotation
- 2.6.1-3, 2.7 Working with sequence annotation

If there's time …

- 2.1 Working with PDB structures
- 2.8 DNA and Protein sequences and Jalview



PHOSPHORYLATION (Y) 73; High confidence server. Only hits with scores over 0.8 are reported. Link

Sequence 8 ID: FER1_SPIOL Residue: TYR

Preferences

Help

Residue and conservation colouring

Alignment and JNet services

Linked view of structure coloured by sequence

Feature settings

Alignment layout control

Sorting, Tree and PCA analysis and pairwise alignment

Mouse-over to access sequence annotation details

Right-click to open pop-up menu.

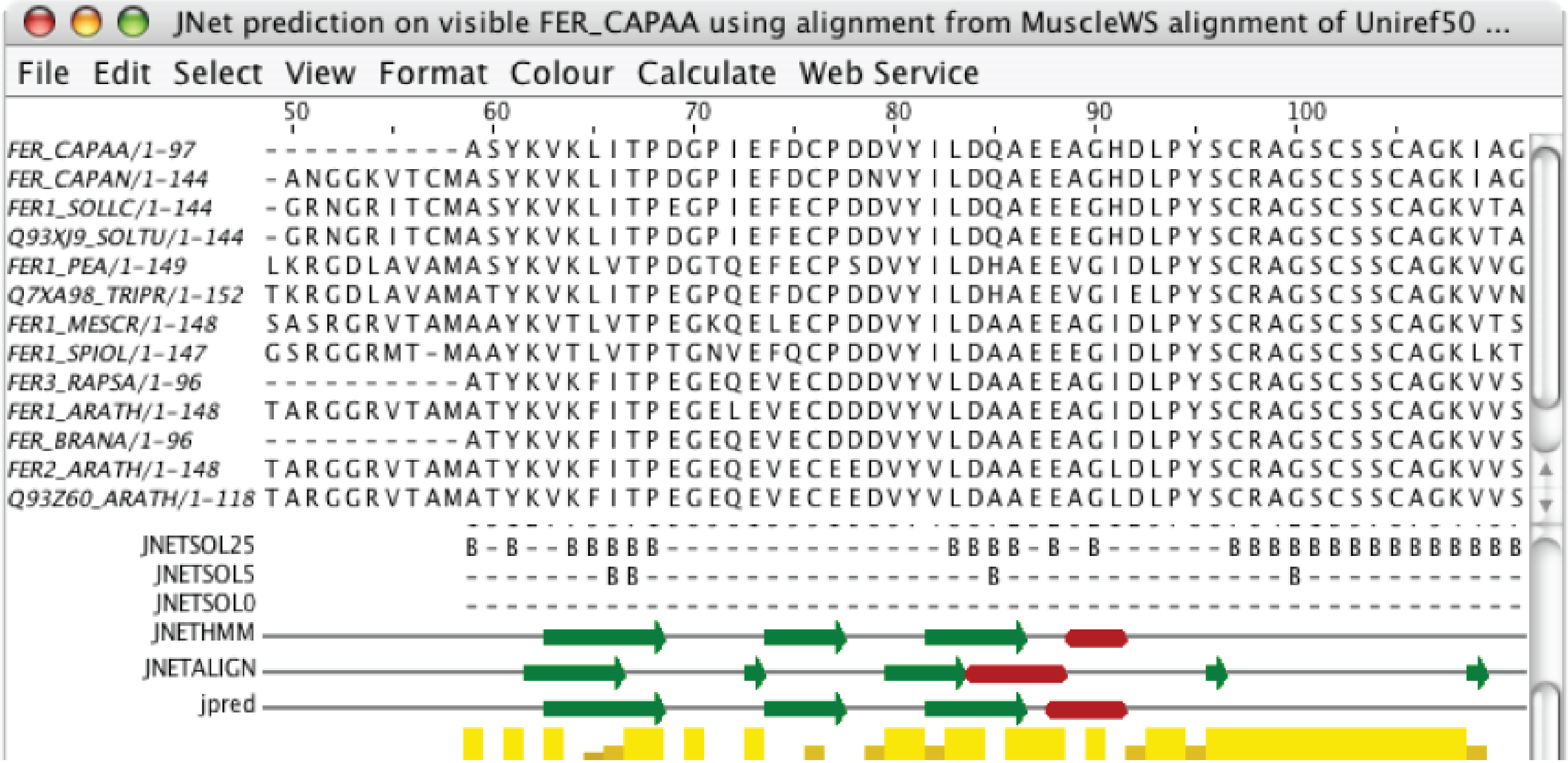
Interactive Tree Viewer linked to alignment

Jalview Desktop

Protein Secondary Structure Prediction
Sect 2.5

- Jalview interfaces with the Jpred protein secondary structure predictor
- Prediction is based on
  - Neural net which can recognise helical, coil or beta strand using amino acid patterns
  - Amino acid profile for a sequence
    - Multiple sequence alignment
    - Profile from sequence database search
      - **P**osition **S**pecific **S**ubstitution **M**atrix

Protein Secondary Structure Prediction
Sect 2.5



```
                    50        60        70        80        90        100
FER_CAPAA/1-97     ----------ASYKVKLITPDGPIEFDCPDDVYILDQAEEAGHDLPYSCRAGSCSSCAGKIAG
FER_CAPAN/1-144    -ANGGKVTCMASYKVKLITPDGPIEFDCPDNVYILDQAEEAGHDLPYSCRAGSCSSCAGKIAG
FER1_SOLLC/1-144   -GRNGRITCMASYKVKLITPEGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSCAGKVTA
Q93XJ9_SOLTU/1-144 -GRNGRITCMASYKVKLITPDGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSCAGKVTA
FER1_PEA/1-149     LKRGDLAVAMASYKVKLVTPDGTQEFECPSDVYILDHAEEVGIDLPYSCRAGSCSSCAGKVVG
Q7XA98_TRIPR/1-152 TKRGDLAVAMATYKVKLITPEGPQEFDCPDDVYILDHAEEVGIELPYSCRAGSCSSCAGKVVN
FER1_MESCR/1-148   SASRGRVTAMAAYKVTLVTPEGKQELECPDDVYILDAAEEAGIDLPYSCRAGSCSSCAGKVTS
FER1_SPIOL/1-147   GSRGGRMT-MAAYKVTLVTPTGNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSCAGKLKT
FER3_RAPSA/1-96    ----------ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSCAGKVVS
FER1_ARATH/1-148   TARGGRVTAMATYKVKFITPEGELEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSCAGKVVS
FER_BRANA/1-96     ----------ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSCAGKVVS
FER2_ARATH/1-148   TARGGRVTAMATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSCAGKVVS
Q93Z60_ARATH/1-118 TARGGRVTAMATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSCAGKVVS

         JNETSOL25           B-B--BBBBB---------------BBBB-B-B------BBBBBBBBBBBBBB
          JNETSOL5           -------BB------------------B--------------B----------
          JNETSOL0           -----------------------------------------------------
           JNETHMM
         JNETALIGN
             jpred
```

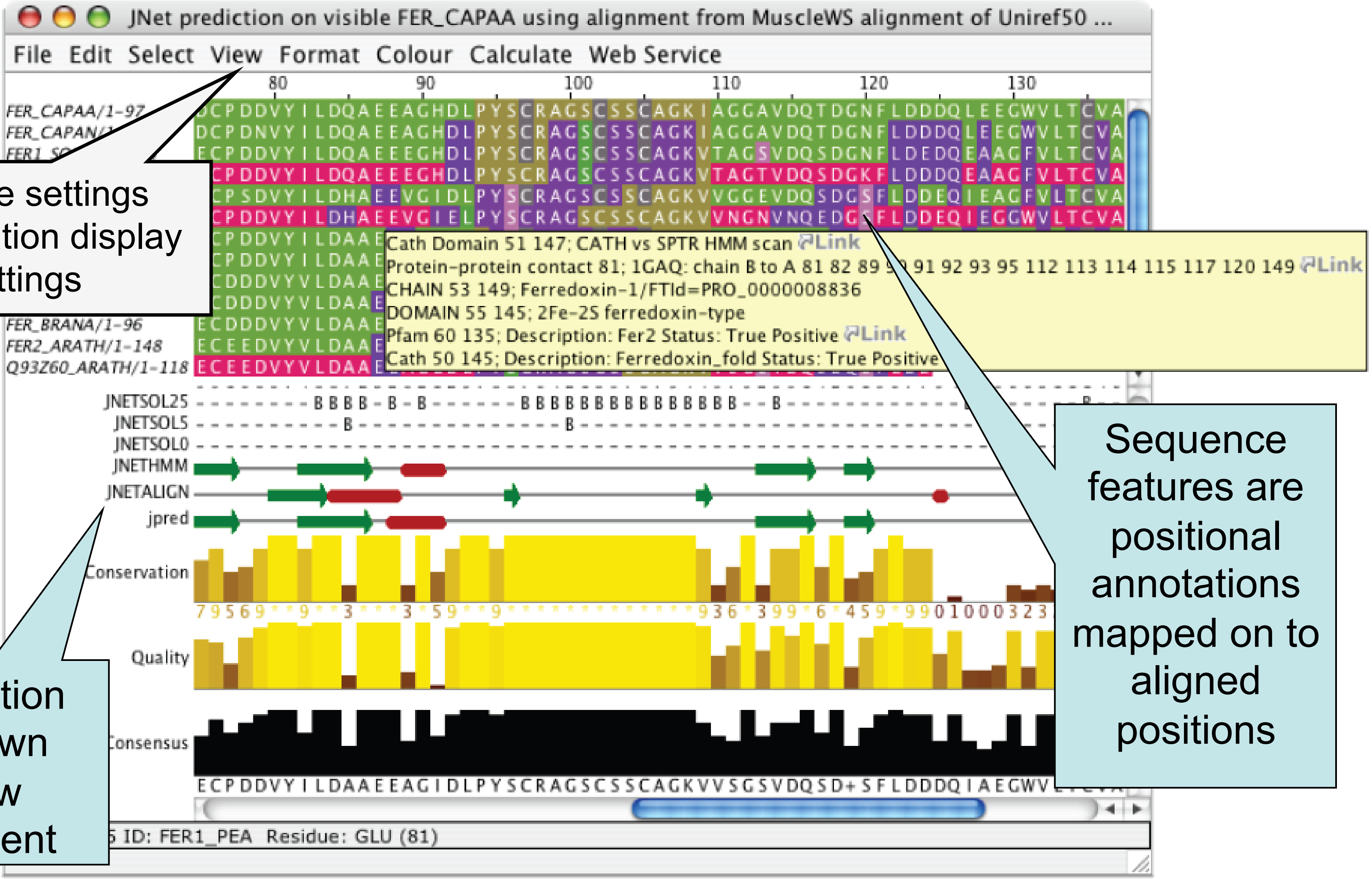
Exercise 26 (Sec 2.5)

- Tasks
  - Perform a variety of Jnet predictions
    - Note the effect of hidden regions
    - Learn about sequence associated annotation
  - Save your results for the next exercise
- Questions
  - …

Exercise 26 (Sec 2.5)

- Tasks
  - – Perform a variety of Jnet predictions
    - Note the effect of hidden regions
    - Learn about sequence associated annotation
  - – Save your results for the next exercise
- Questions
  - – What other data does Jnet provide ?
  - – Which is better – a PSI blast prediction or an MSA based prediction ?
  - – What happens when you have hidden regions ?

Alignment Annotation and sequence features.



JNet prediction on visible FER_CAPAA using alignment from MuscleWS alignment of Uniref50 ...

Cath Domain 51 147; CATH vs SPTR HMM scan Link
Protein-protein contact 81; 1GAQ: chain B to A 81 82 89 9 91 92 93 95 112 113 114 115 117 120 149 Link
CHAIN 53 149; Ferredoxin-1/FTId=PRO_0000008836
DOMAIN 55 145; 2Fe-2S ferredoxin-type
Pfam 60 135; Description: Fer2 Status: True Positive Link
Cath 50 145; Description: Ferredoxin_fold Status: True Positive

ID: FER1_PEA Residue: GLU (81)

Feature settings & annotation display settings

Sequence features are positional annotations mapped on to aligned positions

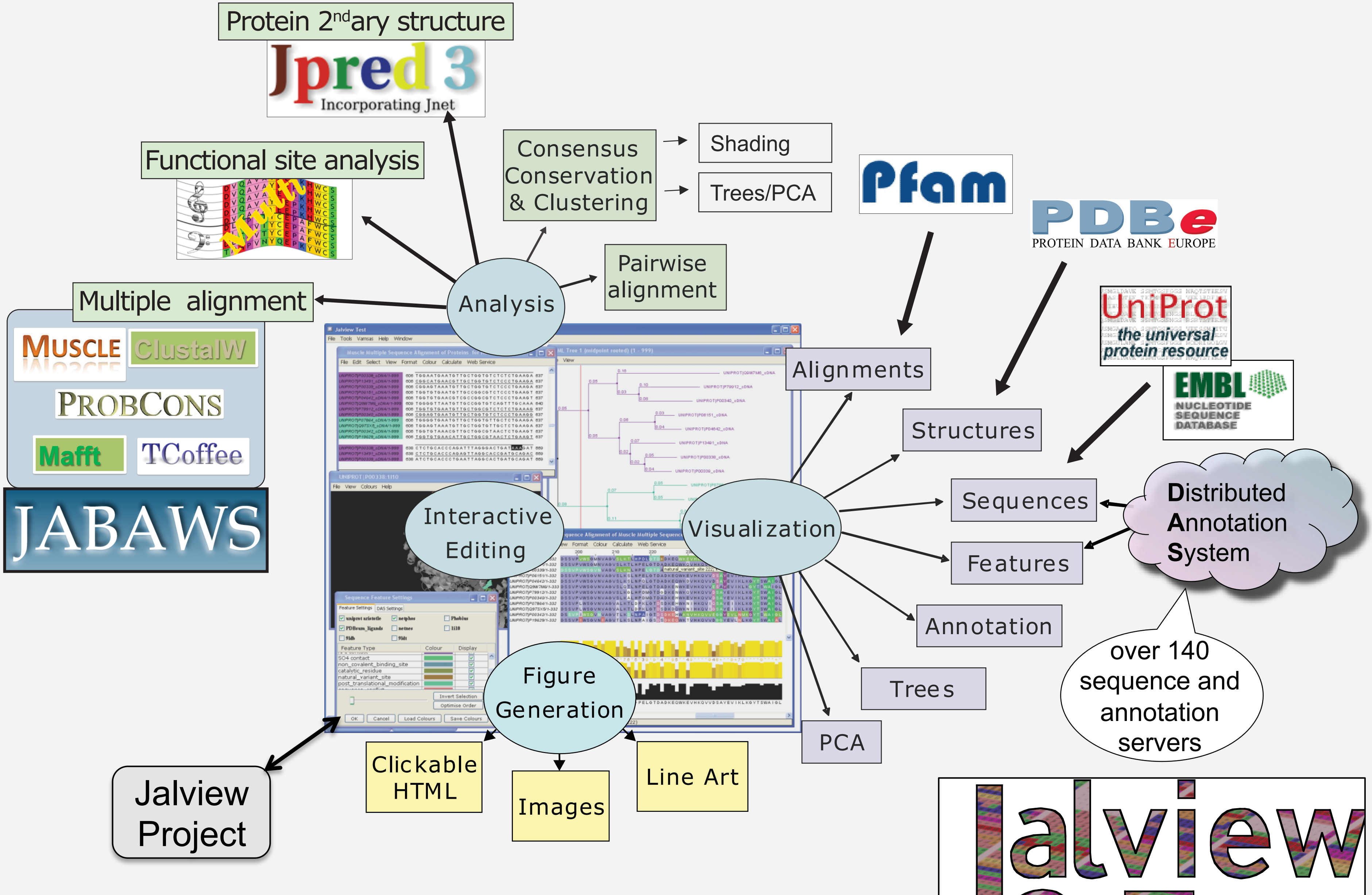
Annotation Is shown below alignment

Creating, editing and using annotation. Exercise 28 (Sec. 2.6.4)

- Tasks
  - Manually annotate some columns using the interactive editing functions
  - Learn about jalview annotation files
    - How to change the appearance of quantitative data.
    - Understand how to create sequence associated annotation
- Questions
  - What other things can be defined in jalview annotation files ?

- Next:

Sequence and alignment annotation and database references.



Protein 2ndary structure

Jpred 3
Incorporating Jnet

Functional site analysis

Multiple alignment

MUSCLE

ClustalW

PROBCONS

Mafft

TCoffee

JABAWS

Analysis

Consensus
Conservation
& Clustering

Shading

Trees/PCA

Pairwise
alignment

Pfam

PDBe
PROTEIN DATA BANK EUROPE

UniProt
the universal
protein resource

EMBL
NUCLEOTIDE
SEQUENCE
DATABASE

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Visualization

Interactive
Editing

Figure
Generation

Distributed
Annotation
System

over 140
sequence and
annotation
servers

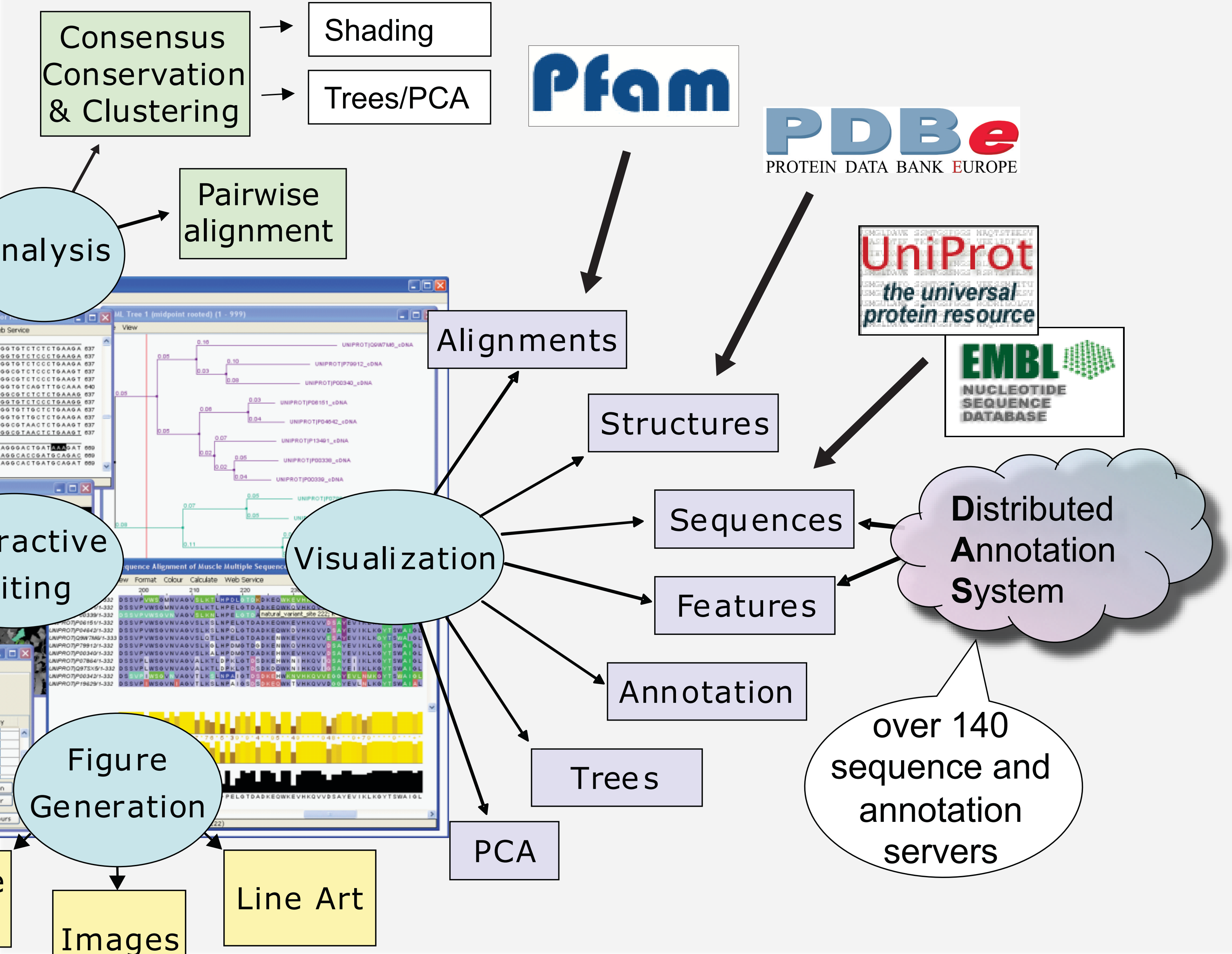
Clickable
HTML

Images

Line Art

Jalview
Project

Jalview
2.7



Consensus Conservation & Clustering

Shading

Trees/PCA

Pfam

Pairwise alignment

nalysis

PDBe
PROTEIN DATA BANK EUROPE

UniProt
the universal protein resource

EMBL
NUCLEOTIDE SEQUENCE DATABASE

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Visualization

ractive
iting

Figure Generation

e

Images

Line Art

Distributed
Annotation
System

over 140 sequence and annotation servers

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

Jalview and Sequence Databases Sec 2.7.1 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
  - Verify retrieval by examining sequence annotation
- Question
  - What happens if only a subsequence is present in the alignment ?
  - Does database annotation get shared between alignments ?

Sequence Features
Section 2.6.1-3 & Ex 27

  - Annotate the whole or part of a sequence
  - Database refs are special case.
- Tasks
  - Visualise, create, modify, import and export features.
- Questions
  - What are the different types of file formats available for import and export
  - What services allow you to discover annotation for sequence ?

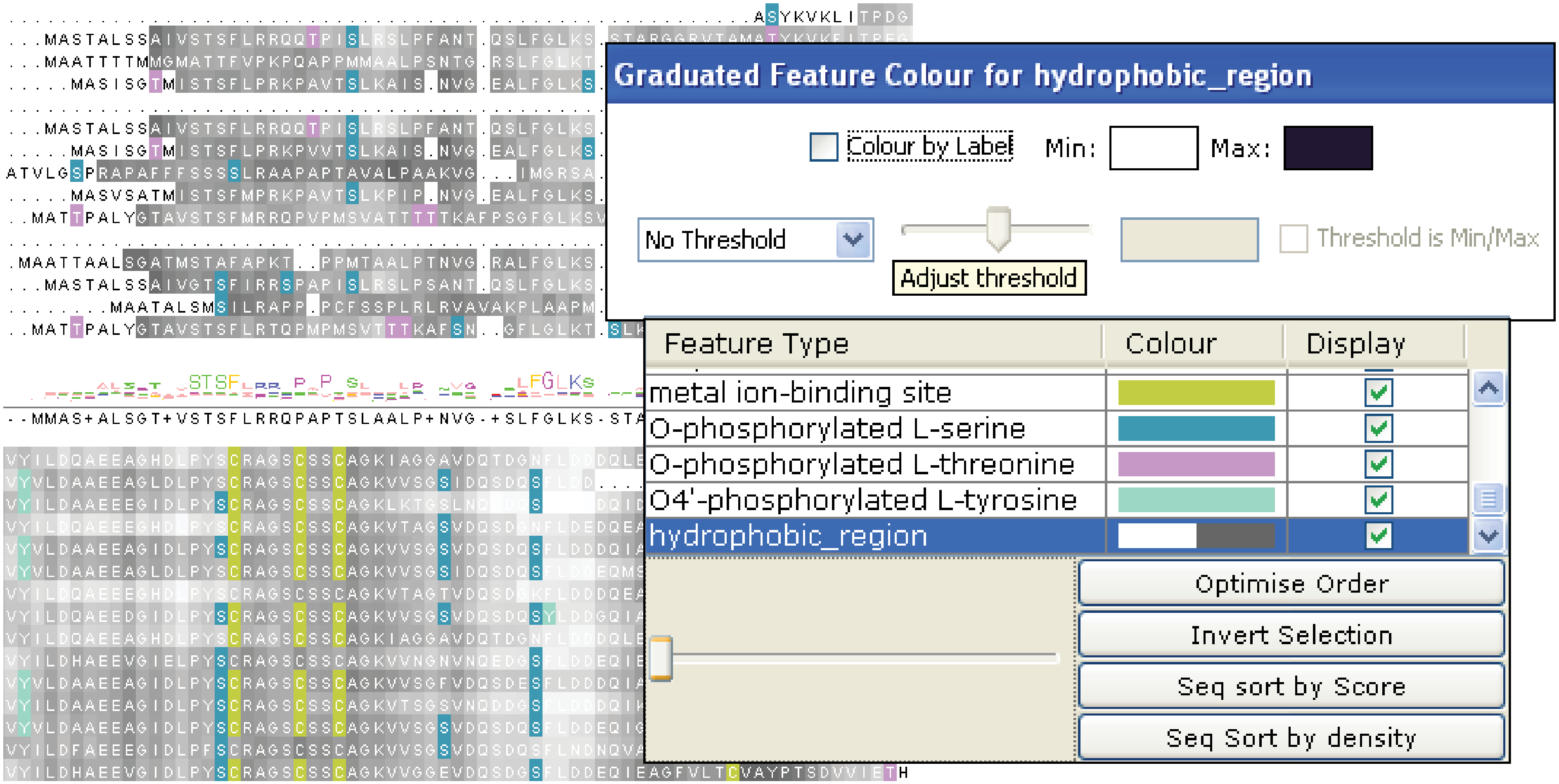
The Distributed Annotation System
Section 2.7.2, Exercise 30

  - Web servers that jalview can use to discover annotation for a sequence
- Task
  - Browse available DAS sources for protein sequences
  - Retrieve annotation for the ferredoxin alignment.
- Question
  - What does the ‘optimise order’ button do ?

Working with sequence features Ex 31 and 32 (Sec 2.7.4)

- Task
  - Shading features using labels and scores
  - Sorting alignment using feature scores

Shading, thresholding, colour by label.



Graduated Feature Colour for hydrophobic_region

☐ Colour by Label Min: Max:

No Threshold

☐ Threshold is Min/Max

Feature Type	Colour	Display
metal ion-binding site	🟩	☑
O-phosphorylated L-serine	🟦	☑
O-phosphorylated L-threonine	🟪	☑
O4'-phosphorylated L-tyrosine	🟩	☑
hydrophobic_region	▨	☑

Working with sequence features Ex 31 and 32 (Sec 2.7.4)

- Task
  - Shading features using labels and scores
  - Sorting alignment using feature scores
- Questions
  - What types of features are best displayed with a ‘label’ colourscheme ?
  - How would you display only the highest or lowest scoring features ?

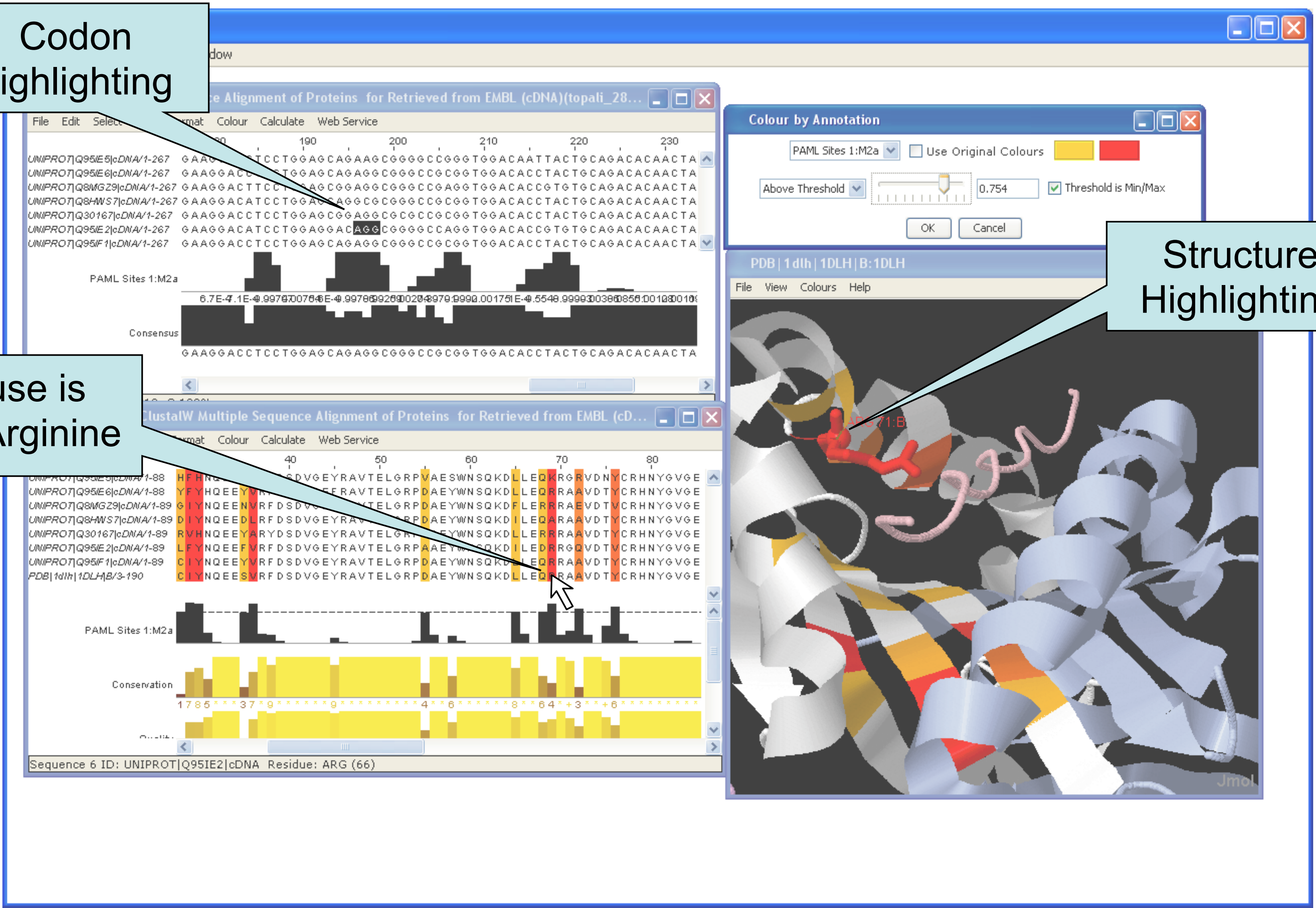
- Next:

Structures, and working with DNA and Protein.

Protein Structure and Jalview
Section 2.1

- Jalview includes the Jmol molecular graphics viewer
  - Structures can be coloured by their aligned sequences
  - Position of mouse highlighted in sequence or structure

Structure shaded by sequence



Colour by Annotation: PAML Sites 1:M2a; ☐ Use Original Colours; Above Threshold; 0.754; ☑ Threshold is Min/Max

Sequence 6 ID: UNIPROT|Q95IE2|cDNA Residue: ARG (66)

Codon highlighting

Mouse is over Arginine

Structure Highlighting

Associating structures with sequences

- Local PDB file
  - Attach PDB file to sequence manually
  - drag and drop to match files to sequences by ID
- Structures in the PDB database
  - Provide PDB id (and chain) for sequence
  - Discover references *via* sequence database

Protein Structures in Jalview
Sec 2.1. Exercise 14

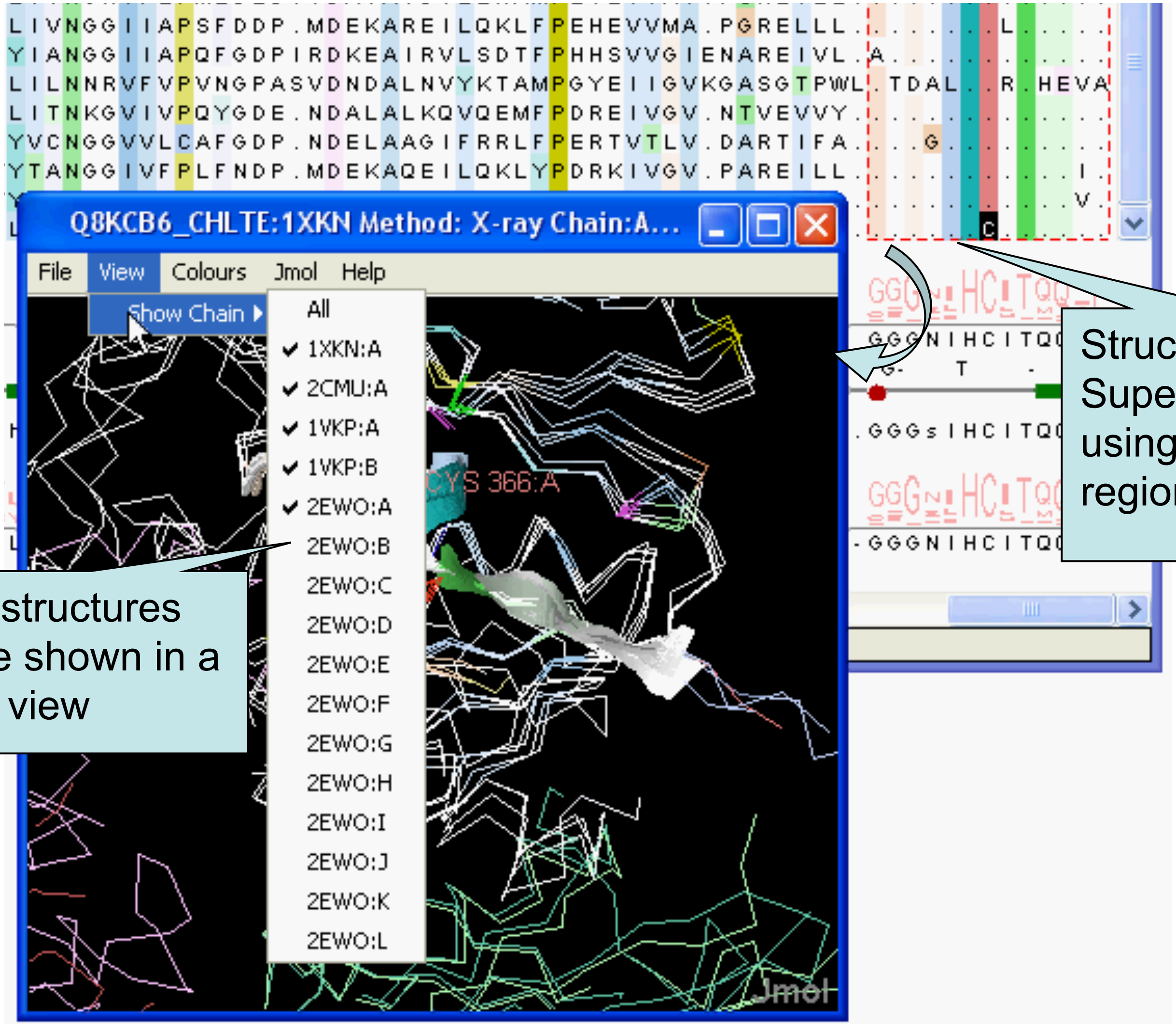
- Task
  - Discover PDB structures for ferredoxin sequence(s)
  - Save and load structures and manipulate colouring

Protein Structures in Jalview
Sec 2.1. Exercise 14

- Task
  - Discover PDB structures for ferredoxin sequence(s)
  - Save and load structures and manipulate colouring
- Questions
  - How does Jalview match up sequence data to structural data

New Features in Jalview 2.6 (Sep. ’10)

Upgrade to Jmol 12



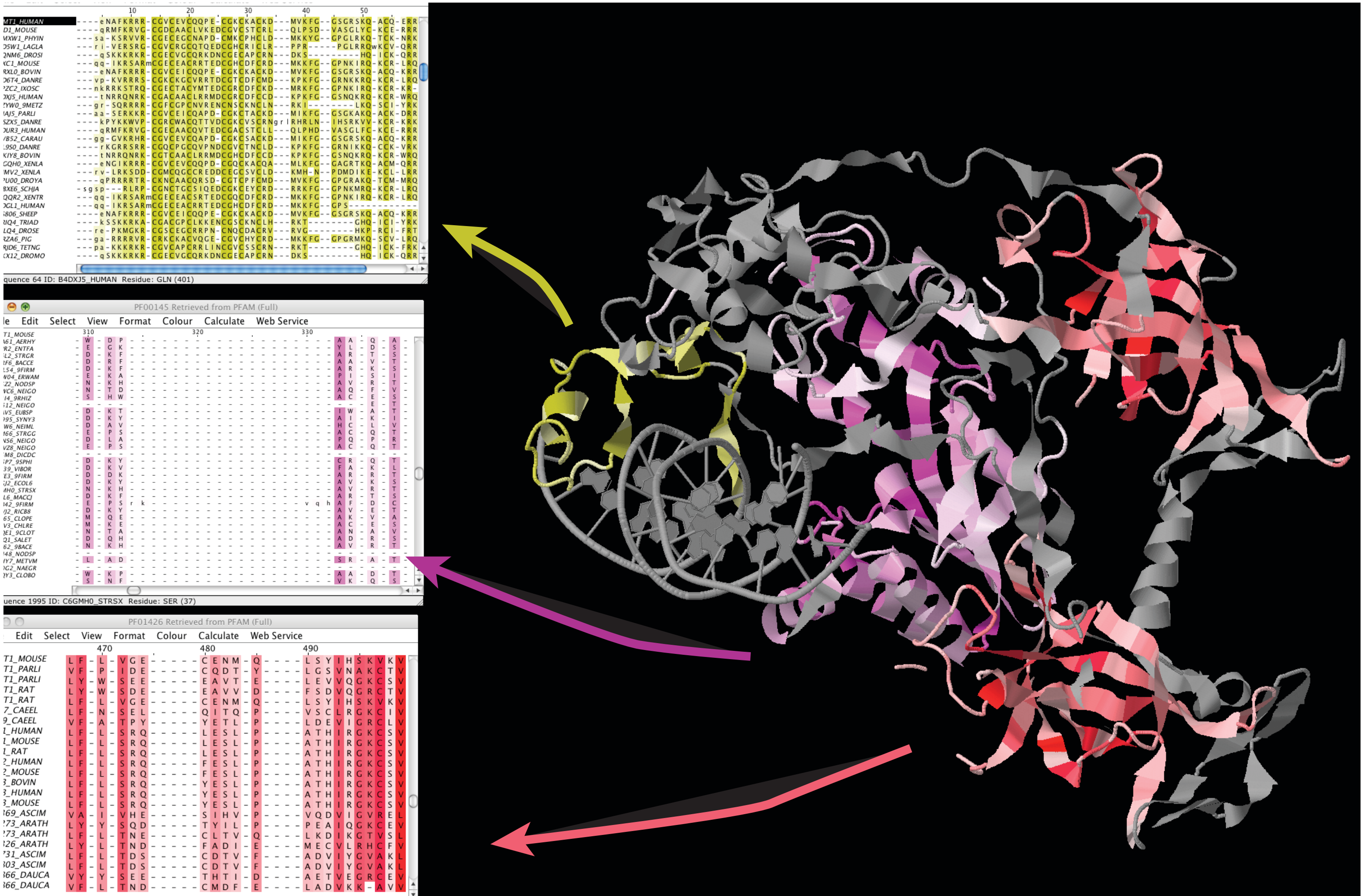
many structures can be shown in a single view

Structures can be Superimposed using the visible region of alignment

Superposing Structures using Alignments
Sec 2.1.4 – Exercise 15

- Task
  - Align structures using the ferredoxin alignment
    - Note: May hit bug in v2.7 with ‘View all N structures’
  - Experiment with views to control what part of the alignment is used to superimpose the structures
- Questions
  - What colourscheme would highlight the conserved parts of the structures ?
  - Which view gave the ‘best’ structure superposition ?
    - How did you decide this ?

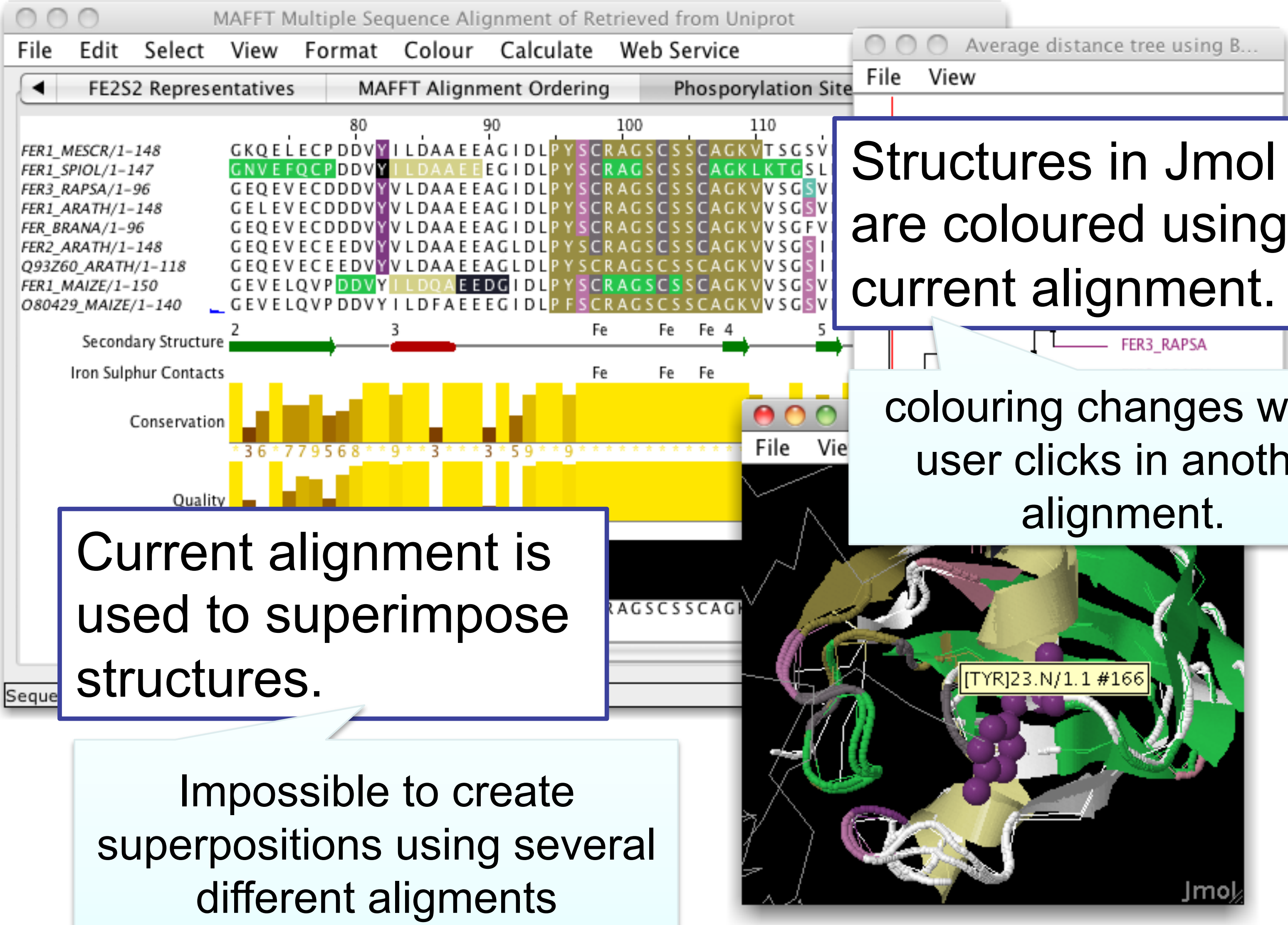
Colouring structures using many multiple alignments
Sect 2.1.5. Exercise 16



- Next:

Structures, and working with DNA and Protein.

Some problems with Structure Visualization in Jalview 2.6



```
                         80        90        100       110
FER1_MESCR/1-148      GKQELECPDDVYILDAAEEAGIDLPYSCRAGSCSSCAGKVTSGSV
FER1_SPIOL/1-147      GNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSCAGKLKTGSL
FER3_RAPSA/1-96       GEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSCAGKVVSGSV
FER1_ARATH/1-148      GELEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSCAGKVVSGSV
FER_BRANA/1-96        GEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSCAGKVVSGFV
FER2_ARATH/1-148      GEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSCAGKVVSGSI
Q93Z60_ARATH/1-118    GEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSCAGKVVSGSI
FER1_MAIZE/1-150      GEVELQVPDDVYILDQAEEDGIDLPYSCRAGSCSSCAGKVVSGSV
O80429_MAIZE/1-140    GEVELQVPDDVYILDFAEEEGIDLPFSCRAGSCSSCAGKVVSGSV
```

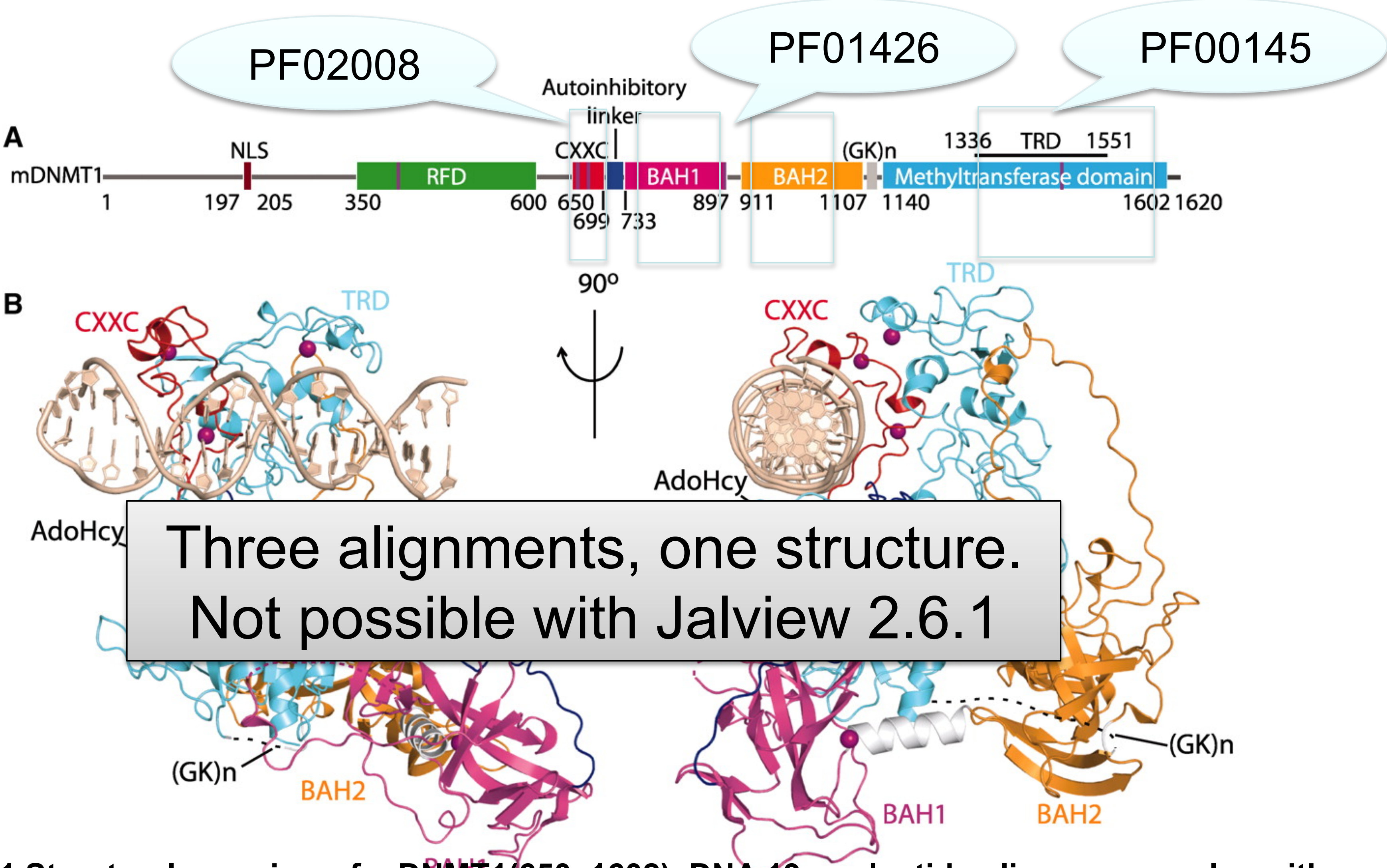
Structures in Jmol view are coloured using current alignment.

colouring changes when user clicks in another alignment.

Current alignment is used to superimpose structures.

Impossible to create superpositions using several different aligments

Case Study : Exploring functional interfaces within a multi-domain complex using PFAM alignments



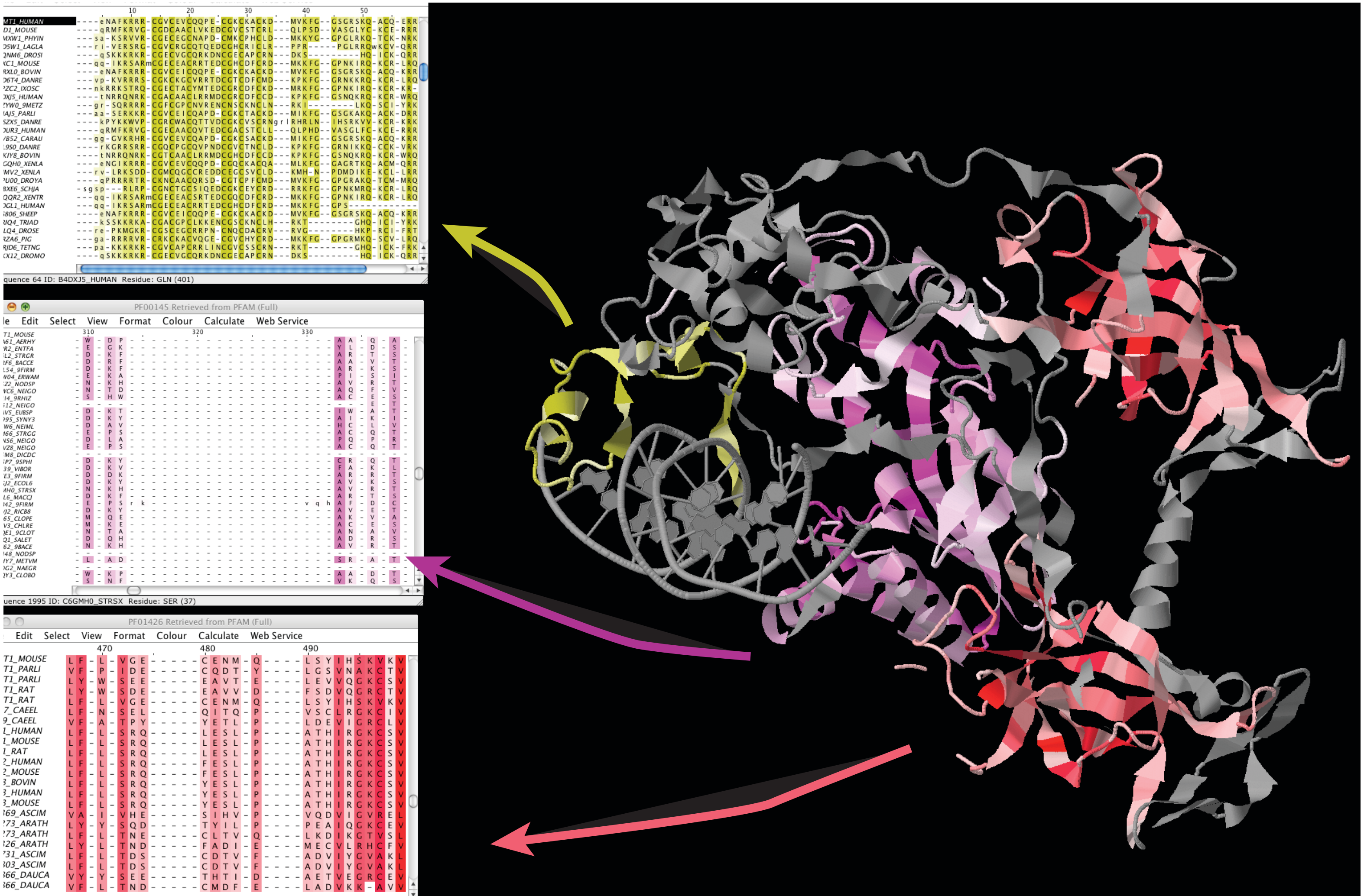
Three alignments, one structure.
Not possible with Jalview 2.6.1

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

Published by AAAS

Colouring structures using many multiple alignments
Sect 2.1.5. Exercise 16

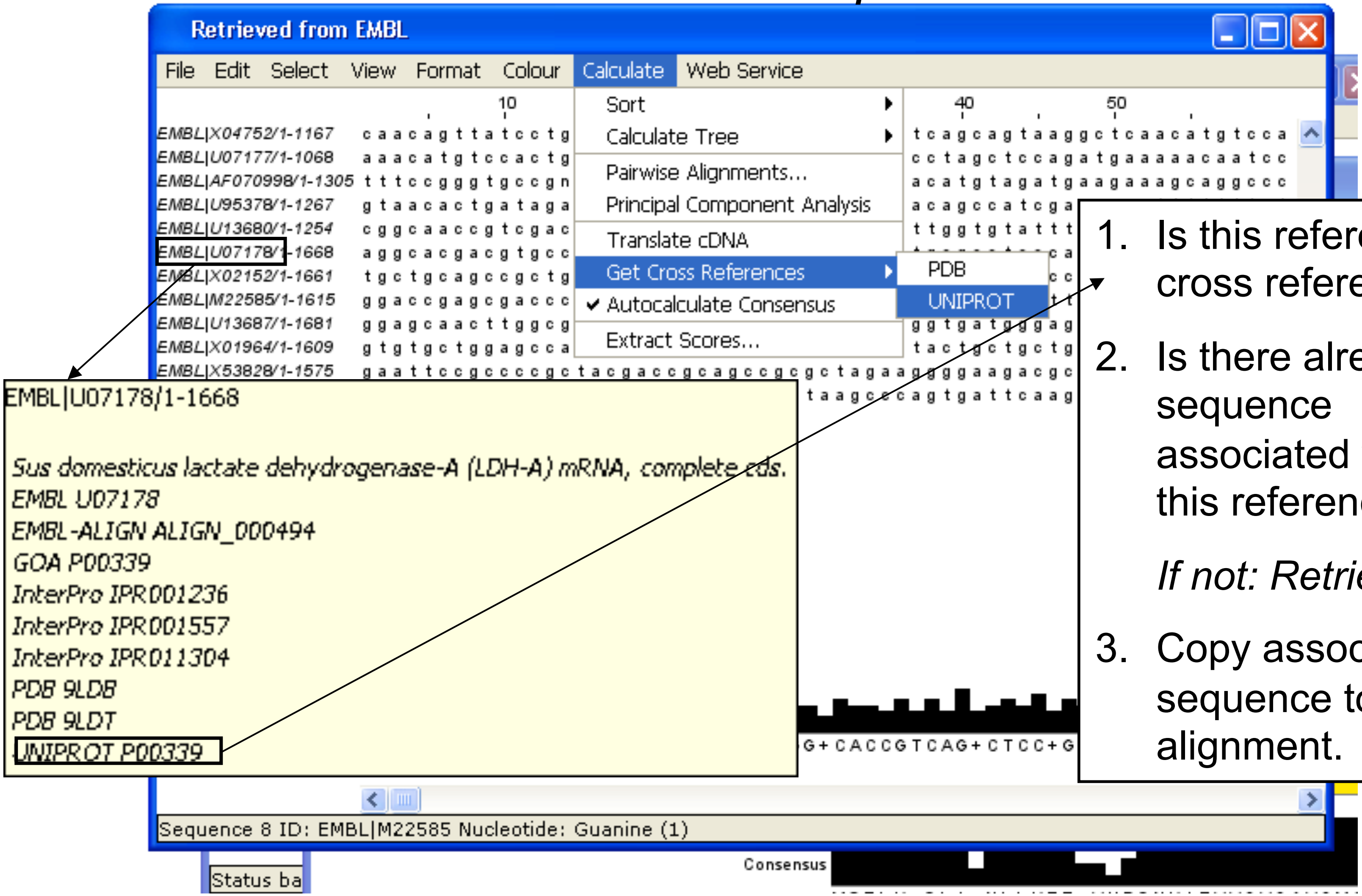


DNA and Protein in Jalview

- Discussed in Section 2.8 of manual
- From DNA to Protein
  - Calculations => Translate cDNA
  - View protein annotation on exons using EMBL records
- From protein to DNA
  - Recover DNA for proteins using EMBL cross references

Semantic Processing: Database Reference Tracing

‘get me the sequences from database ***blah*** *for the selected sequences’*



EMBL|U07178/1-1668

Sus domesticus lactate dehydrogenase-A (LDH-A) mRNA, complete cds.
EMBL U07178
EMBL-ALIGN ALIGN_000494
GOA P00339
InterPro IPR001236
InterPro IPR001557
InterPro IPR011304
PDB 9LDB
PDB 9LDT
UNIPROT P00339

Sequence 8 ID: EMBL|M22585 Nucleotide: Guanine (1)

1. Is this reference a cross reference ?
2. Is there already a sequence associated with this reference ?

   If not: Retrieve it.
3. Copy associated sequence to new alignment.

Protein Feature visualization on DNA

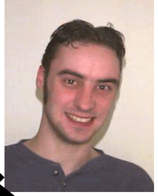
Section 2.8, exercise 33

- Task
  - Retrieve a DNA contig and visualize features from UNIPROT at their coding positions.
- Question
  - What fields in an EMBL sequence record can Jalview use ?

A brief history of Jalview

Jalview
Version 1
1997

Michele Clamp
Now at BioTeam, Boston, US.



James Cuff
Now at Harvard.

Steve Searle
Now at Sanger, UK

2004
Jalview 1 published.

Jalview
Version 2
2005

David
Martin

2.1
2.2
2.3
2.4
2.5
2.6

BBSRC
bioscience for the future

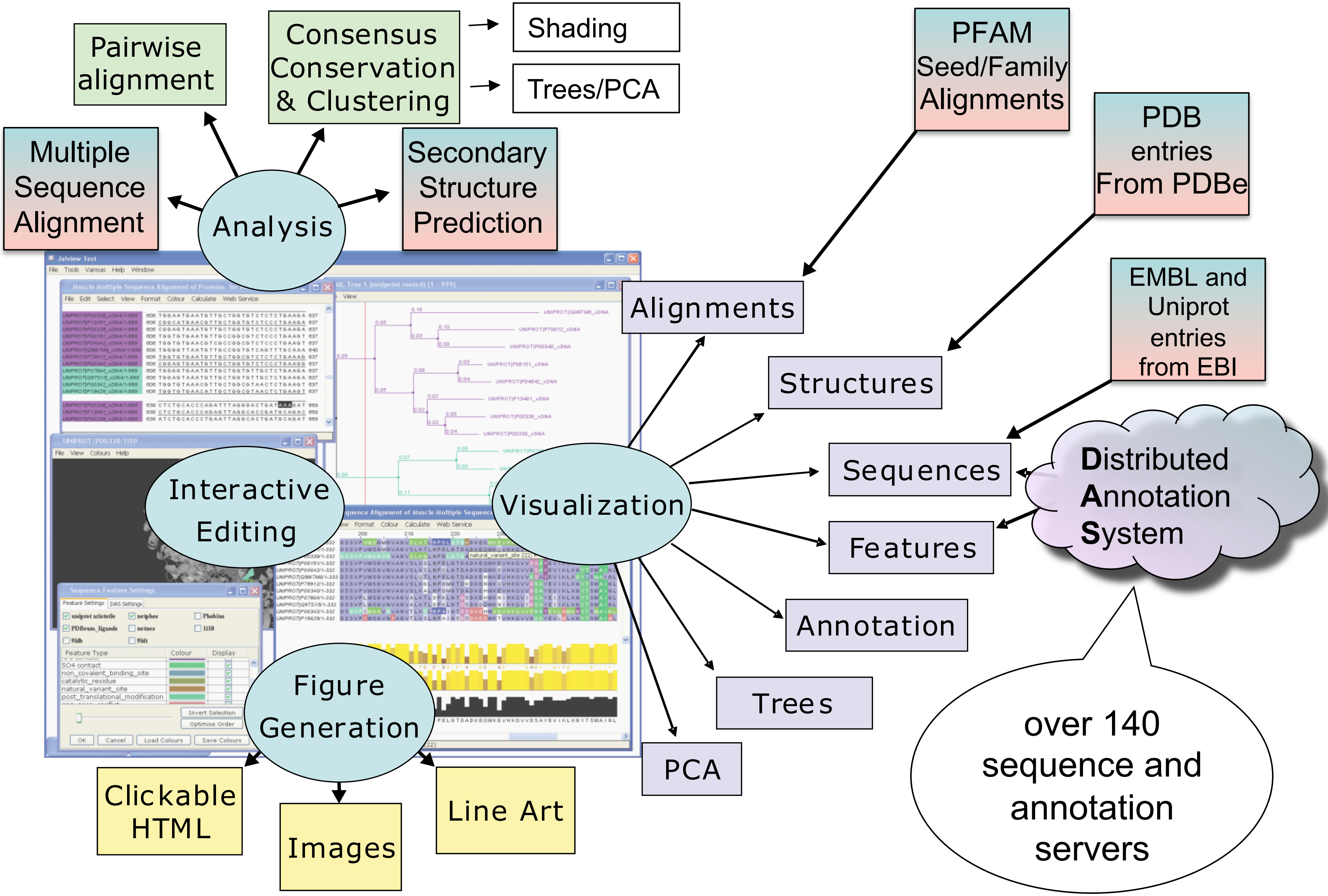
Andrew Waterhouse
Riken, Japan

Jim
Procter

Jalview 2 Paper
2009

scienceWATCH.com
TRACKING TRENDS & PERFORMANCE IN BASIC RESEARCH
New Hot Paper

WRAP UP



Pairwise alignment

Consensus Conservation & Clustering

Shading

Trees/PCA

Multiple Sequence Alignment

Analysis

Secondary Structure Prediction

PFAM Seed/Family Alignments

PDB entries From PDBe

EMBL and Uniprot entries from EBI

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Visualization

Interactive Editing

Figure Generation

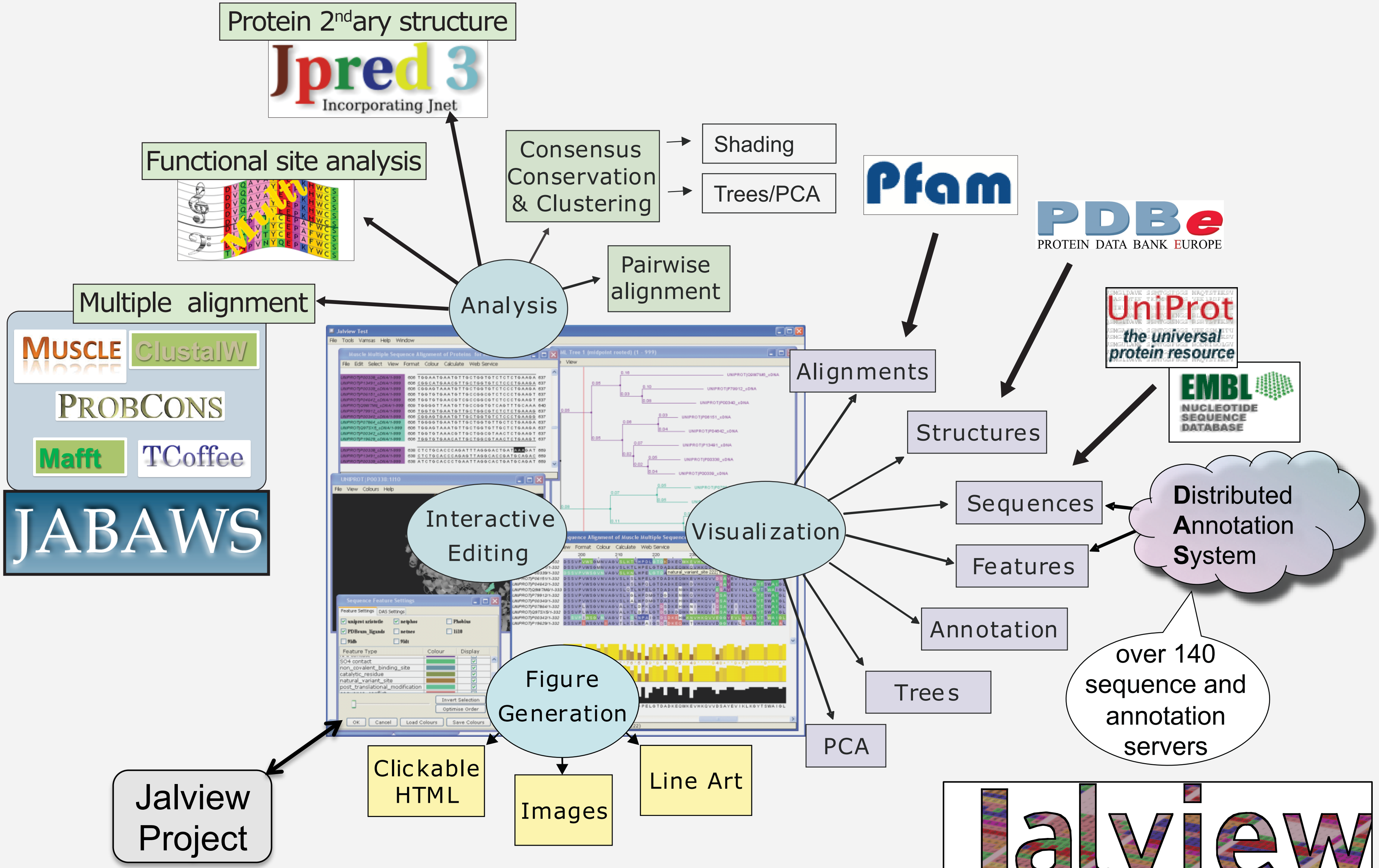
Clickable HTML

Images

Line Art

Distributed **A**nnotation **S**ystem

over 140 sequence and annotation servers



Protein 2^{nd}ary structure

Jpred 3
Incorporating Jnet

Functional site analysis

Multi

Multiple alignment

MUSCLE
ClustalW
PROBCONS
Mafft
TCoffee

JABAWS

Analysis

Consensus
Conservation
& Clustering

Shading

Trees/PCA

Pairwise
alignment

Pfam

PDBe
PROTEIN DATA BANK EUROPE

UniProt
the universal protein resource

EMBL
NUCLEOTIDE SEQUENCE DATABASE

Visualization

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Distributed
Annotation
System

over 140
sequence and
annotation
servers

Interactive
Editing

Figure
Generation

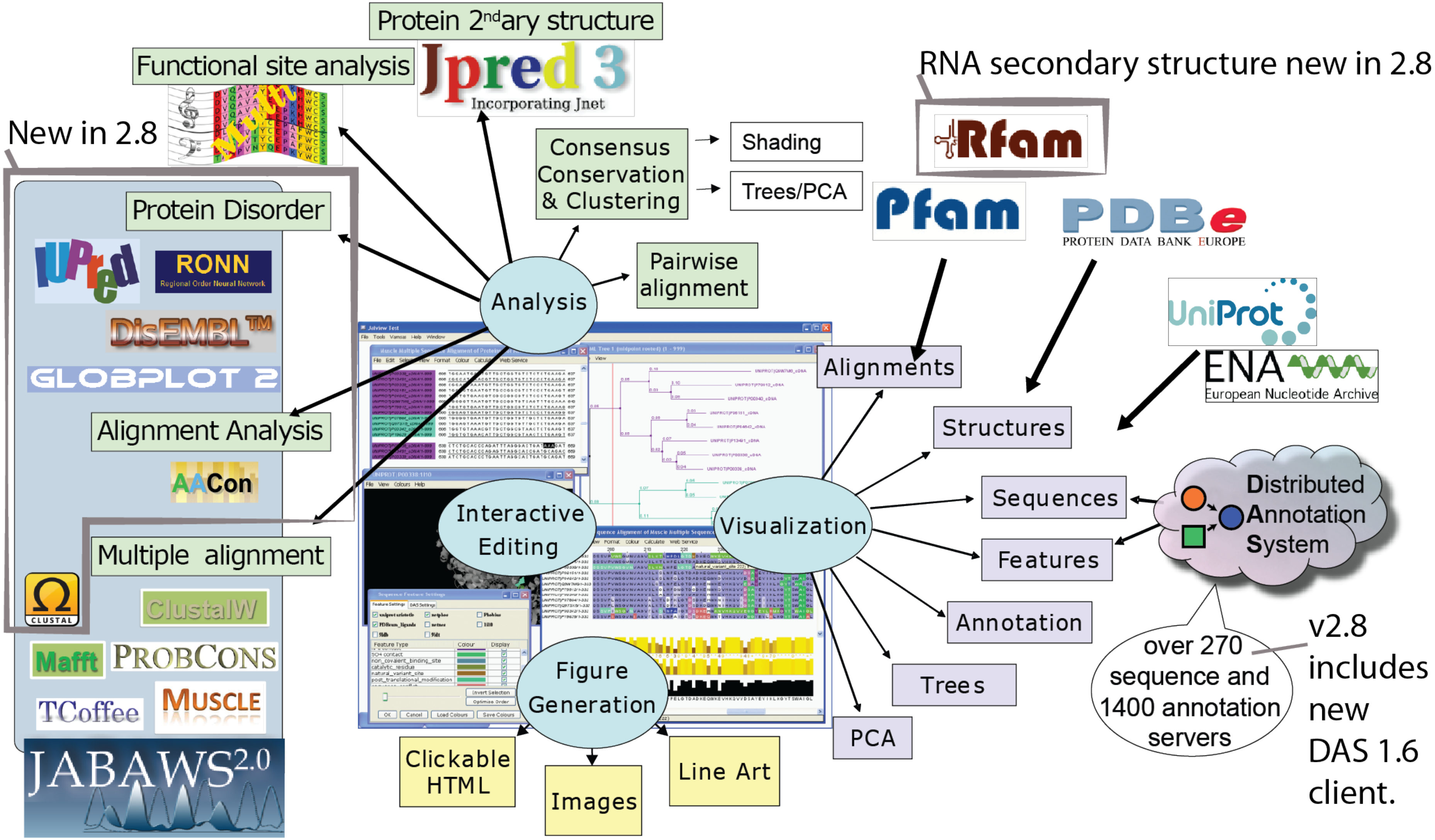
Clickable
HTML

Images

Line Art

Jalview
Project

Jalview
2.7



Protein 2ndary structure

Jpred 3
Incorporating Jnet

Functional site analysis

Multi-Harmony

RNA secondary structure new in 2.8

Rfam

New in 2.8

Consensus
Conservation
& Clustering

Shading

Trees/PCA

Pfam

PDBe
PROTEIN DATA BANK EUROPE

Protein Disorder

IUPred

RONN
Regional Order Neural Network

DisEMBL™

GLOBPLOT 2

Analysis

Pairwise
alignment

UniProt

ENA
European Nucleotide Archive

Alignments

Structures

Alignment Analysis

AACon

Interactive
Editing

Visualization

Sequences

Features

Distributed
Annotation
System

Multiple alignment

CLUSTAL

ClustalW

Mafft

ProbCons

TCoffee

Muscle

Annotation

Trees

PCA

over 270
sequence and
1400 annotation
servers

v2.8
includes
new
DAS 1.6
client.

Figure
Generation

JABAWS 2.0

Clickable
HTML

Images

Line Art

Next in Jalview… even more services!

A grab bag
of Disorder
predictors

JABAWS$^{2.0}$

RONN
Regional Order Neural Network

DisEMBL™

IUPred

GLOBPLOT 2

aligns
millions of
sequences

CLUSTAL

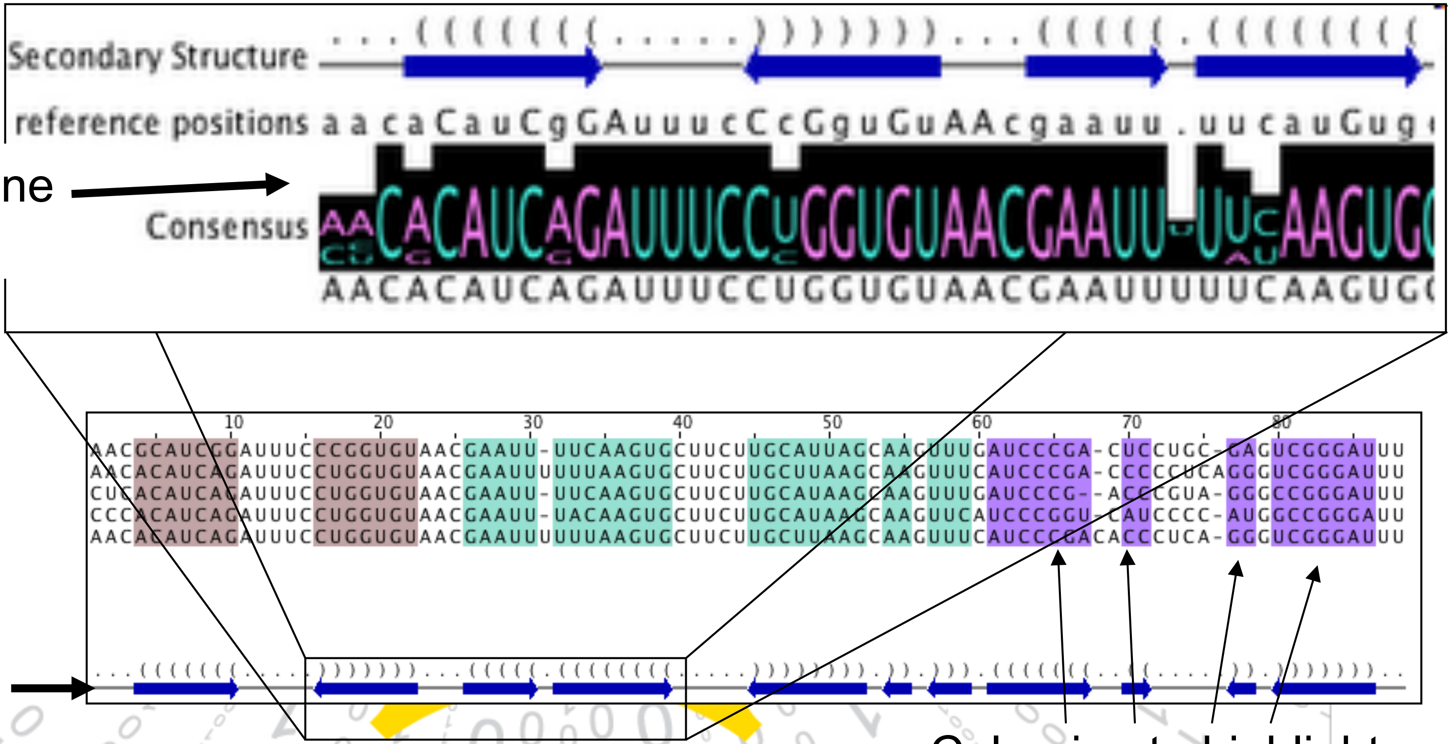
AACon

Provides Over 10 sequence
conservation measures!
Developed by Agnieska Golicz
(U. Dundee summer student)

Coming up: RNA Support for Jalview

Lauren Lui, UC Santa Cruz.

http://jalview-rnasupport.blogspot.com/



```
AACGCAUCGGAUUUCCCGGUGUAACGAAUU-UUCAAGUGCUUCUUGCAUUAGCAAGUUUGAUCCCGA-CUCCUGC-GAGUCGGGAUUU
AACACAUCAGAUUUCCUGGUGUAACGAAUUUUUUAAGUGCUUCUUGCUUAAGCAAGUUUCAUCCCGA-CCCCCUCAGGGUCGGGAUUU
CUCACAUCAGAUUUCCUGGUGUAACGAAUU-UUCAAGUGCUUCUUGCAUAAGCAAGUUUGAUCCCG--ACCCGUA-GGGCCGGGAUUU
CCCACAUCAGAUUUCCUGGUGUAACGAAUU-UACAAGUGCUUCUUGCAUAAGCAAGUUCAUCCCGGU-CAUCCCC-AUGGCCGGGAUU
AACACAUCAGAUUUCCUGGUGUAACGAAUUUUUUAAGUGCUUCUUGCUUAAGCAAGUUUCAUCCCGACACCCUCA-GGGUCGGGAUUU
```

Purine/pyrimidine colourscheme

alignment fetcher

WUSS annotation parser (from RALEE)

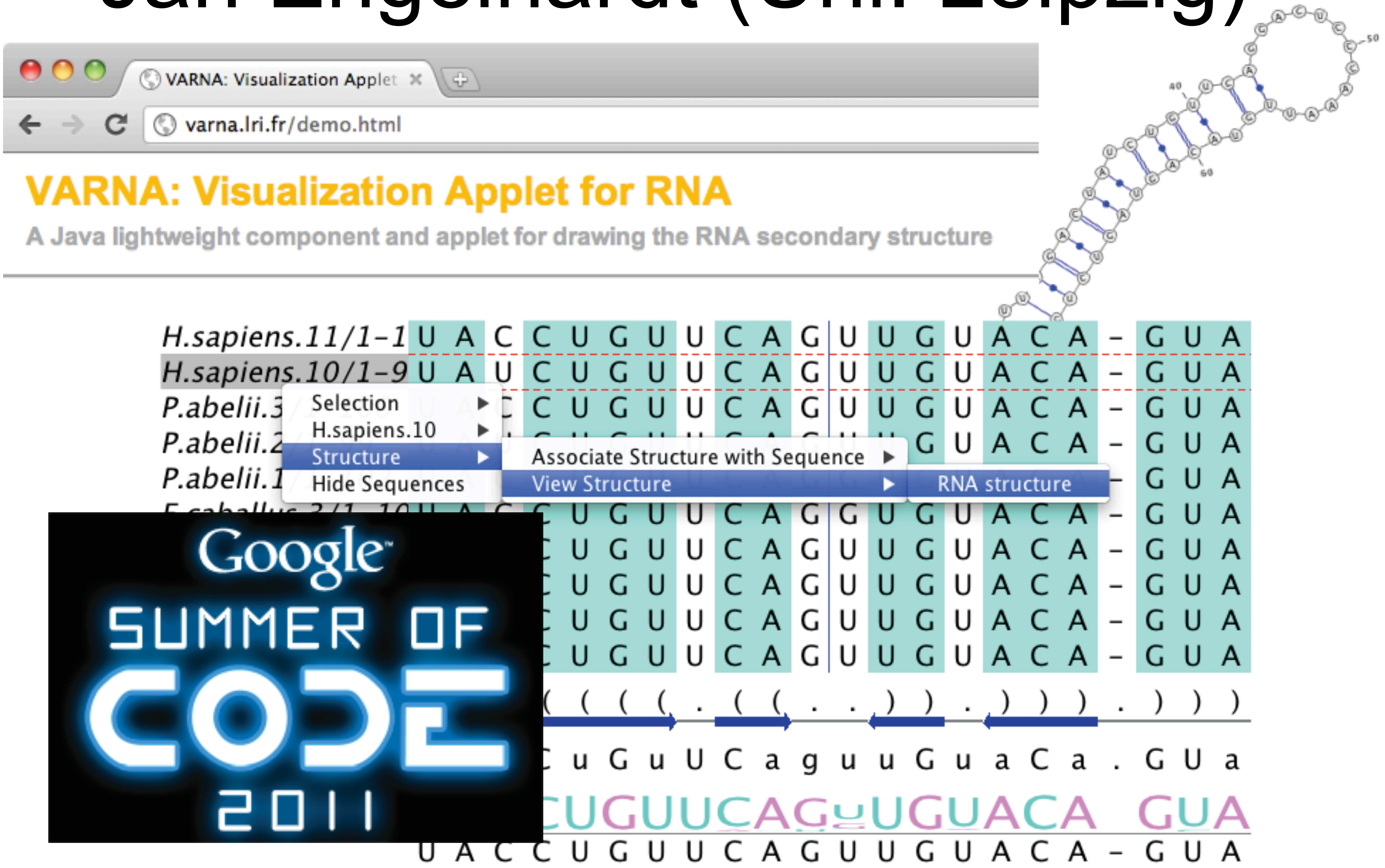
Colouring to highlight helical structure

SUMMER OF CODE // 2010

code.google.com/soc

Google

Jan Engelhardt (Uni. Leipzig)



VARNA: Visualization Applet

varna.lri.fr/demo.html

VARNA: Visualization Applet for RNA

A Java lightweight component and applet for drawing the RNA secondary structure

]

Jalview 2.8 development build

Development build installer here:

http://www.compbio.dundee.ac.uk/user/ws-dev1/jalview/develop/InstallAnywhere/Web_Installers/install.htm

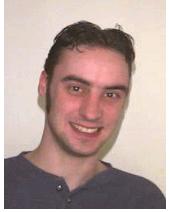
Web based applet demonstration pages:

http://www.compbio.dundee.ac.uk/user/ws-dev1/jalview/develop/applet/applets.html

Acknowledgements

The Jalview developers

Michele Clamp
Bioteam, USA.



James Cuff
Harvard, USA.

Steve Searle
Sanger, UK

Andrew Waterhouse
Riken, Japan

BBSRC
bioscience for the future

- The jalview users
- Geoff Barton
- David Martin
- Peter Troshin

ENFIN
Enabling Systems Biology

wellcome trust

Bernd Brandt
VU Amsterdam
Rafael Jiminez
EBI, UK.

RNA Experts

Google
SUMMER OF CODE
2011

SUMMER OF CODE // 2010
code.google.com/soc Google

Jan Engelhardt
Univ. Leipzig, Germany.
Yann Ponty, LIX, Fr.
Lauren Lui
UC Santa Cruz, USA.

Kersten Schroeder
U. Dundee, UK.
Paul Gardiner *Rfam,NZ*.
Albert Vilella, EBI, UK.
Hilmar Lapp & Karen Cranston, NESCent