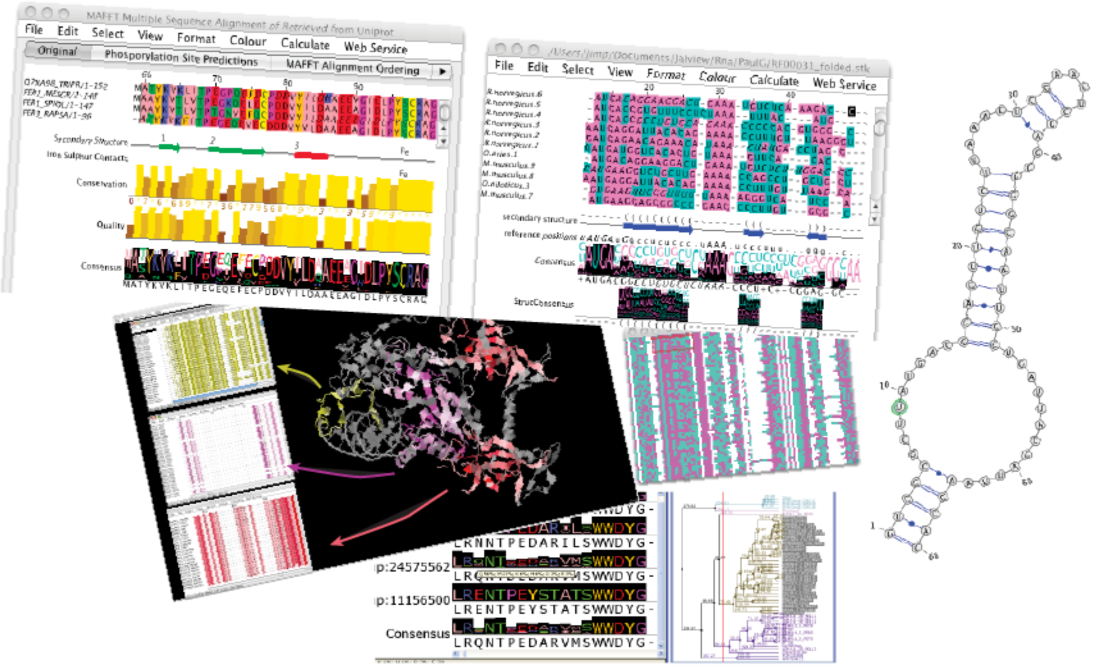
Practical Jalview



Jim Procter
University of Dundee
11th April 2014

THE BARTON GROUP

supported by
wellcome trust
BBSRC bioscience for the future

Jalview
www.jalview.org

UNIVERSITY OF DUNDEE

open source initiative

9.00-9.15am. Overview of the day

9.15am - 10.30am. Session 1. Introduction to Jalview

- starting the application, importing alignments, basic editing and creating figures.

10.30-11am. Coffee

11am - 12.30pm. Session 2: Alignment & alignment analysis

- Creating sequence alignments, importing and calculating trees, tree based alignment analysis

12.30pm to 1.30pm. Lunch

1.30pm – 3.00pm. Session 3: Annotating sequences & alignments

- Creating and viewing sequence annotation
- Protein Secondary structure prediction

3.00pm – 3.30pm Coffee

3.30pm – 4.30pm. Session 4: Working with molecular structures

- Viewing 3D Structures, superimpositions, mapping disorder and alignment quality
- Viewing RNA Secondary Structure

4.30pm – 4.45pm. Wrapup – what we didn’t cover today

~5pm .. And then to Duke’s Corner!

Course materials

Everything is online

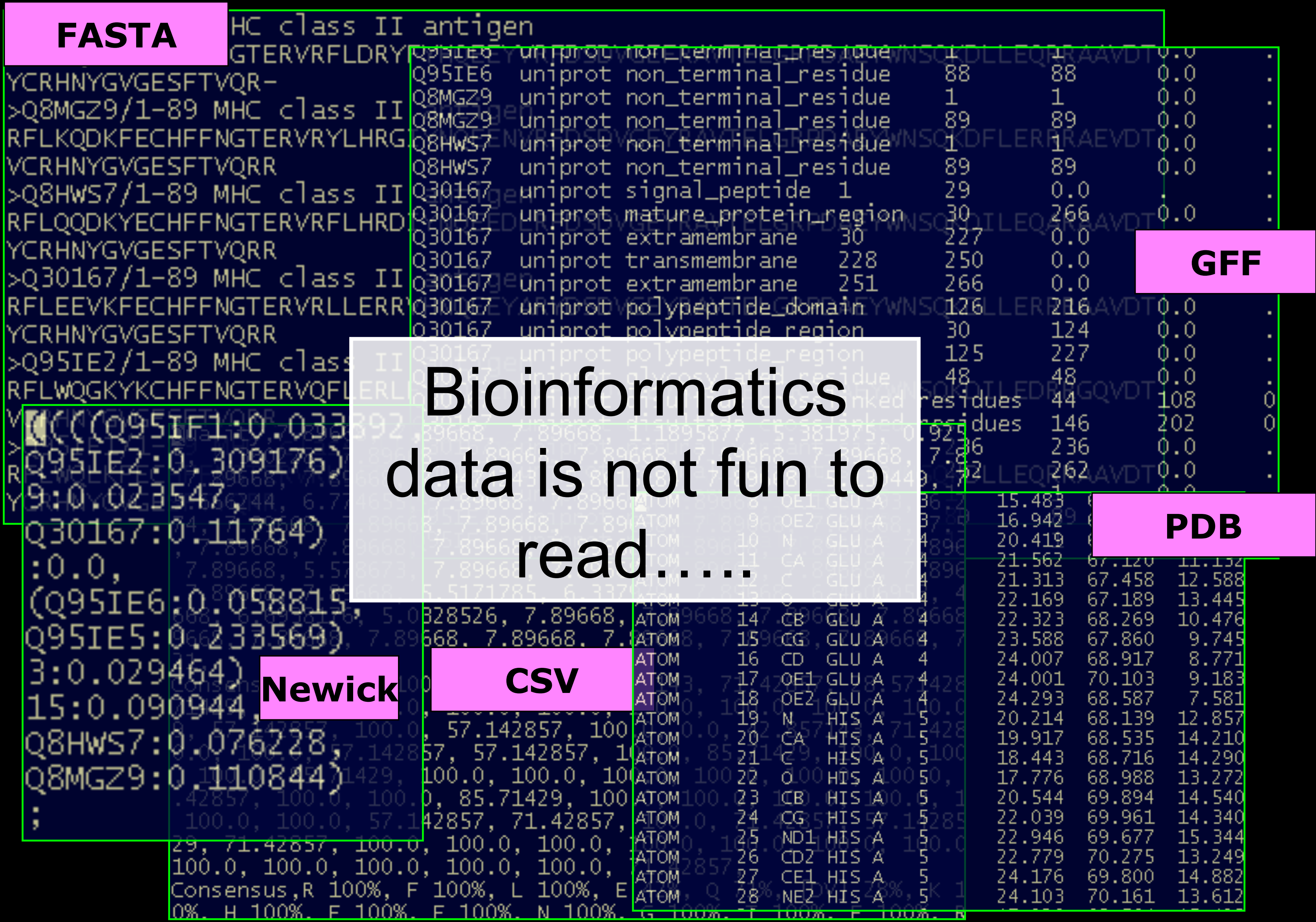
http://www.jalview.org/tutorial/training-materials/2014/Dundee/

- **These slides**

- **Jalview v2.8 Manual (v1.4.1)**
  - **Log in and**
  - **Open the manual in your PDF Viewer NOW**

- **Additional exercises + ‘advanced topics’**

Bioinformatics data is not fun to read…..



FASTA

```
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
RFLWQGKYKCHFFNGTERVQFL
```

GFF

```
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     .
```

Newick

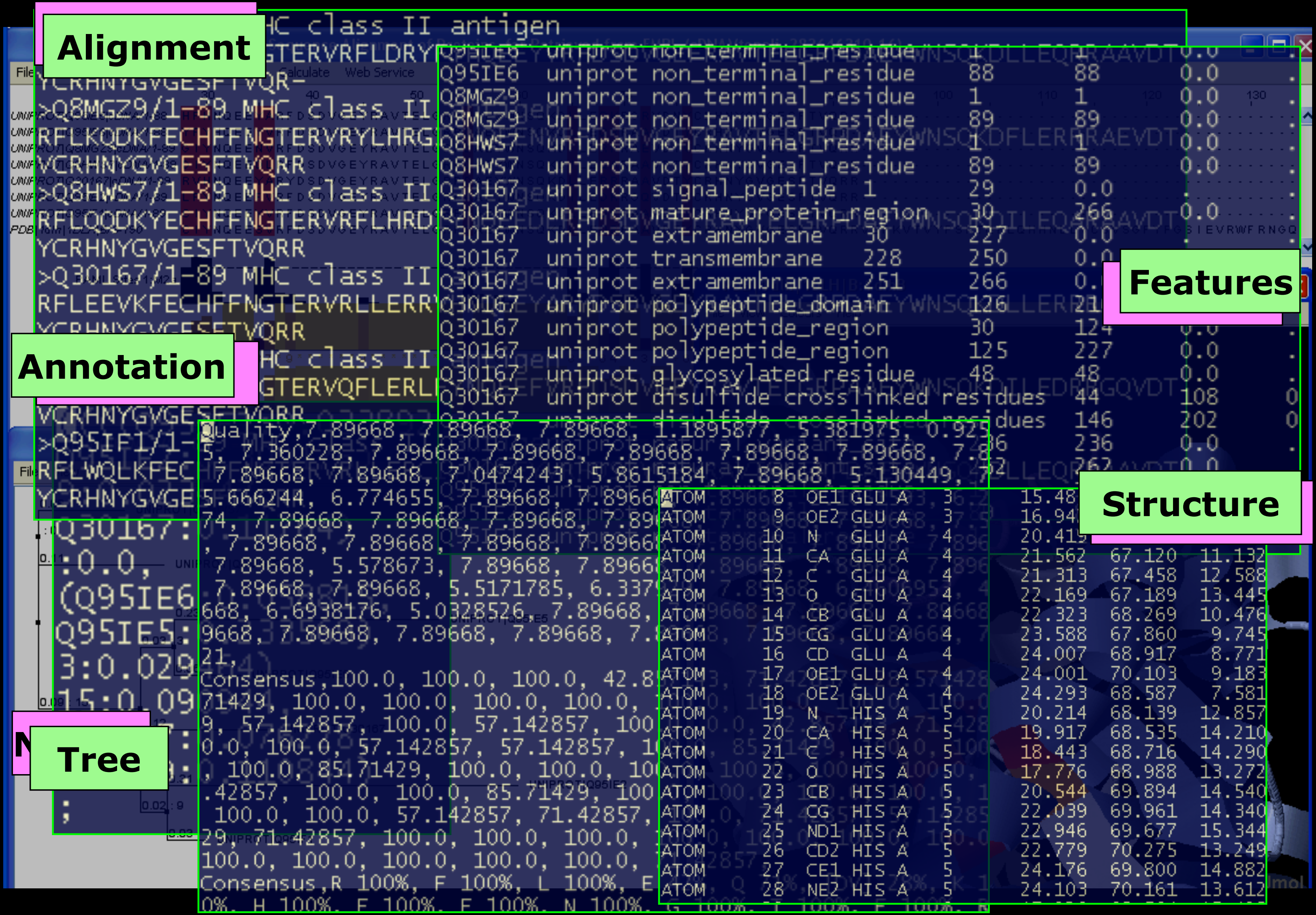
```
(((Q95IF1:0.033892,
Q95IE2:0.309176)
9:0.023547,
Q30167:0.11764)
:0.0,
(Q95IE6:0.058815,
Q95IE5:0.233569)
3:0.029464)
15:0.090944,
Q8HWS7:0.076228,
Q8MGZ9:0.110844)
;
```

CSV

```
57.142857, 100
57, 57.142857, 1
100.0, 100.0, 10
0, 85.71429, 100
42857, 71.42857,
29, 71.42857, 100.0, 100.0, 100.0,
100.0, 100.0, 100.0, 100.0, 100.0,
Consensus,R 100%, F 100%, L 100%, E
0%, H 100%, F 100%, F 100%, N 100%,
```

PDB

```
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
```



Alignment

```
HC class II antigen
GTERVRFLDRY
YCRHNYGVGESFTVQR-
>Q8MGZ9/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
YCRHNYGVGESFTVQRR
>Q8HWS7/1-89 MHC class II
RFLQQDKYECHFFNGTERVRFLHRD
YCRHNYGVGESFTVQRR
>Q30167/1-89 MHC class II
RFLEEVKFECHFFNGTERVRLLERR
YCRHNYGVGESFTVQRR
HC class II
GTERVQFLERL
YCRHNYGVGESFTVQRR
>Q95IF1/1-
RFLWQLKFEC
YCRHNYGVGE
```

Features

```
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.
Q30167  uniprot polypeptide_domain      126  21
Q30167  uniprot polypeptide_region      30   124
Q30167  uniprot polypeptide_region      125  227  0.0 .
Q30167  uniprot glycosylated residue    48   48   0.0 .
Q30167  uniprot disulfide crosslinked residues  44   108  0
Q30167  uniprot disulfide crosslinked residues  146  202  0
```

Annotation

```
Quality,7.89668, 7.89668, 7.89668, 1.1895877, 5.381975, 0.925
5, 7.360228, 7.89668, 7.89668, 7.89668, 7.89668, 7.89668, 7.8
 7.89668, 7.89668, 7.0474243, 5.8615184, 7.89668, 5.130449, 7
5.666244, 6.774655, 7.89668, 7.8966
74, 7.89668, 7.89668, 7.89668, 7.89
, 7.89668, 7.89668, 7.89668, 7.8966
 7.89668, 5.578673, 7.89668, 7.8966
 7.89668, 7.89668, 5.5171785, 6.337
668, 6.6938176, 5.0328526, 7.89668,
9668, 7.89668, 7.89668, 7.89668, 7.
21,
Consensus,100.0, 100.0, 100.0, 42.8
71429, 100.0, 100.0, 100.0, 100.0,
9, 57.142857, 100.0, 57.142857, 100
0.0, 100.0, 57.142857, 57.142857, 1
, 100.0, 85.71429, 100.0, 100.0, 10
.42857, 100.0, 100.0, 85.71429, 100
 100.0, 100.0, 57.142857, 71.42857,
29, 71.42857, 100.0, 100.0, 100.0,
100.0, 100.0, 100.0, 100.0, 100.0,
Consensus,R 100%, F 100%, L 100%, E
0%, H 100%, F 100%, F 100%, N 100%,
```

Structure

```
ATOM   8  OE1 GLU A  3    15.48
ATOM   9  OE2 GLU A  3    16.94
ATOM  10  N   GLU A  4    20.41
ATOM  11  CA  GLU A  4    21.562  67.120  11.132
ATOM  12  C   GLU A  4    21.313  67.458  12.588
ATOM  13  O   GLU A  4    22.169  67.189  13.445
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
```

Tree

```
Q30167:
:0.0,
(Q95IE6
Q95IE5:
3:0.029
15:0.09
```

Jalview

Java

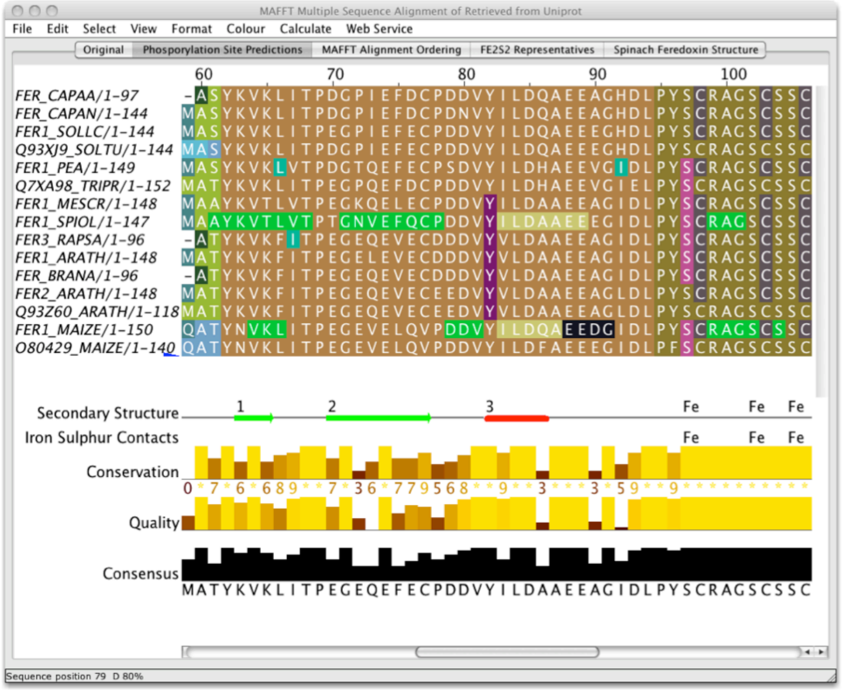
Standalone or web based

alignment
&
annotation,
tree, and
structure

viewer
and
editor

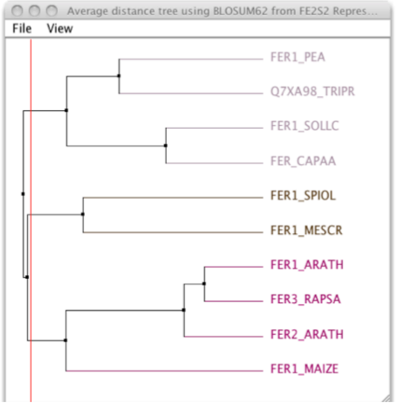
Available at

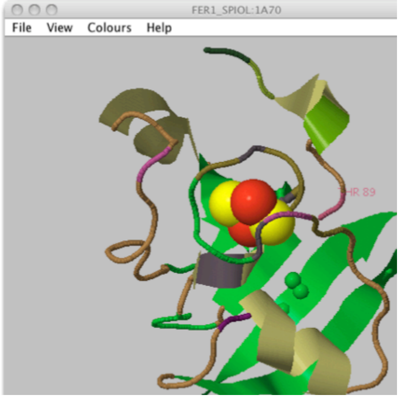
www.jalview.org



```
                  60        70        80        90        100
FER_CAPAA/1-97     -ASYKVKLITPDGPIEFDCPDDVYILDQAEEAGHDLPYSCRAGSCSSC
FER_CAPAN/1-144    MASYKVKLITPDGPIEFDCPDNVYILDQAEEAGHDLPYSCRAGSCSSC
FER1_SOLLC/1-144   MASYKVKLITPEGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSC
Q93XJ9_SOLTU/1-144 MASYKVKLITPDGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSC
FER1_PEA/1-149     MASYKVKLVTPDGTQEFECPSDVYILDHAEEVGIDLPYSCRAGSCSSC
Q7XA98_TRIPR/1-152 MATYKVKLITPEGPQEFDCPDDVYILDHAEEVGIELPYSCRAGSCSSC
FER1_MESCR/1-148   MAAYKVTLVTPEGKQELECPDDVYILDAAEEAGIDLPYSCRAGSCSSC
FER1_SPIOL/1-147   MAAYKVTLVTPTGNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSC
FER3_RAPSA/1-96    -ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER1_ARATH/1-148   MATYKVKFITPEGELEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER_BRANA/1-96     -ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER2_ARATH/1-148   MATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSC
Q93Z60_ARATH/1-118 MATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSC
FER1_MAIZE/1-150   QATYNVKLITPEGEVELQVPDDVYILDQAEEDGIDLPYSCRAGSCSSC
O80429_MAIZE/1-140 QATYNVKLITPEGEVELQVPDDVYILDFAEEEGIDLPFSCRAGSCSSC

Secondary Structure       1       2                3                Fe    Fe  Fe
Iron Sulphur Contacts                                               Fe    Fe  Fe
Conservation
                   0*7*6*689**7*36*779568**9**3***3*59**9**********
Quality
Consensus
                   MATYKVKLITPEGEQEFECPDDVYILDAAEEAGIDLPYSCRAGSCSSC
```

Sequence position 79 D 80%



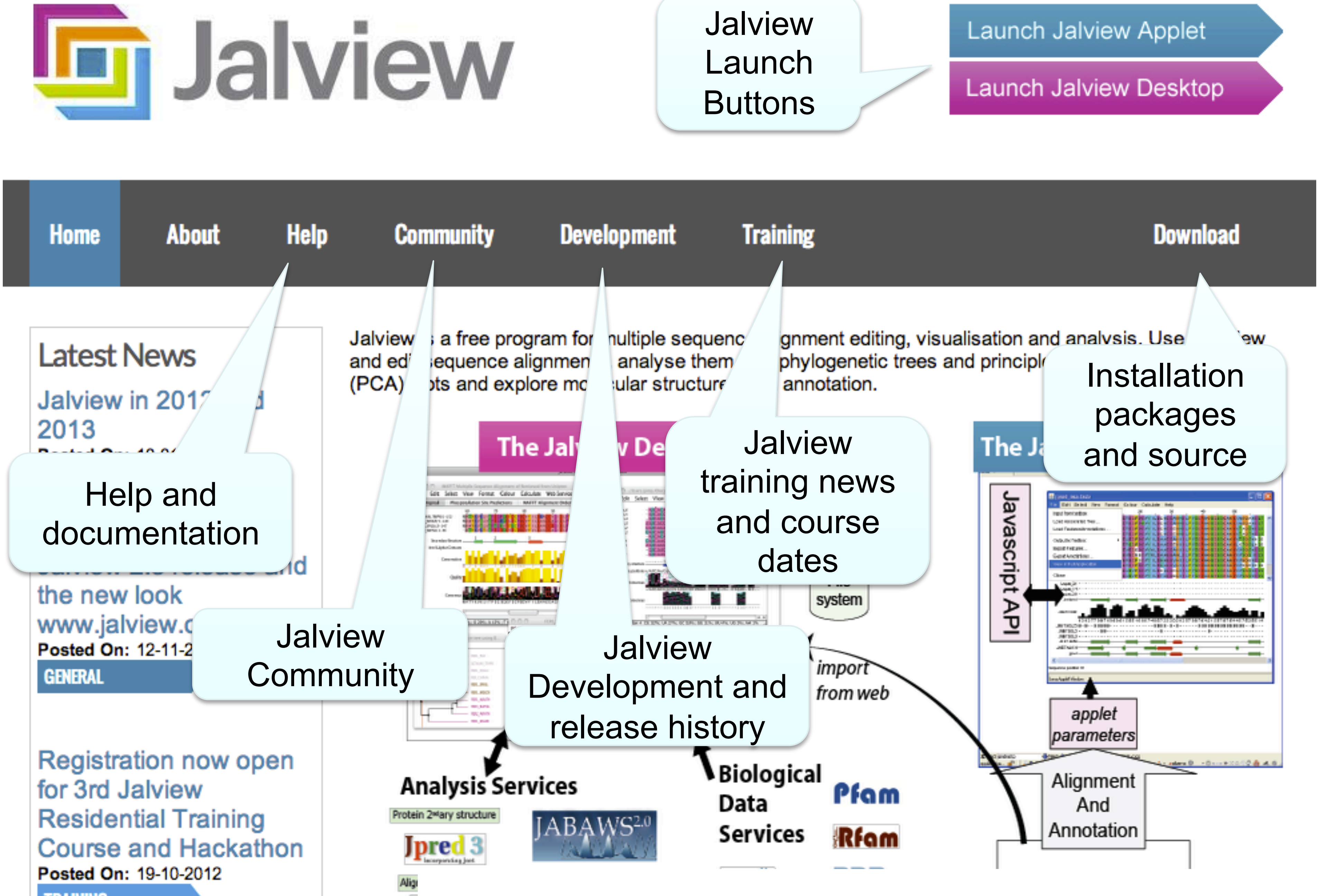


GPLv3 Free Software

UNIVERSITY OF DUNDEE

The Barton Group

School of Life Sciences, University of Dundee, Dow St, Dundee DD1 5EH, Scotland, UK



Jalview Launch Buttons

Latest News

Jalview in 201 d 2013

Posted On: 19-0

the new look

www.jalview.c

Posted On: 12-11-2

GENERAL

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

Posted On: 19-10-2012

TRAINING

Jalview s a free program for nultiple sequenc gnment editing, visualisation and analysis. Use ew and edi sequence alignmen analyse them phylogenetic trees and principl (PCA) ots and explore mo ular structure annotation.

Help and documentation

Jalview Community

Jalview Development and release history

Jalview training news and course dates

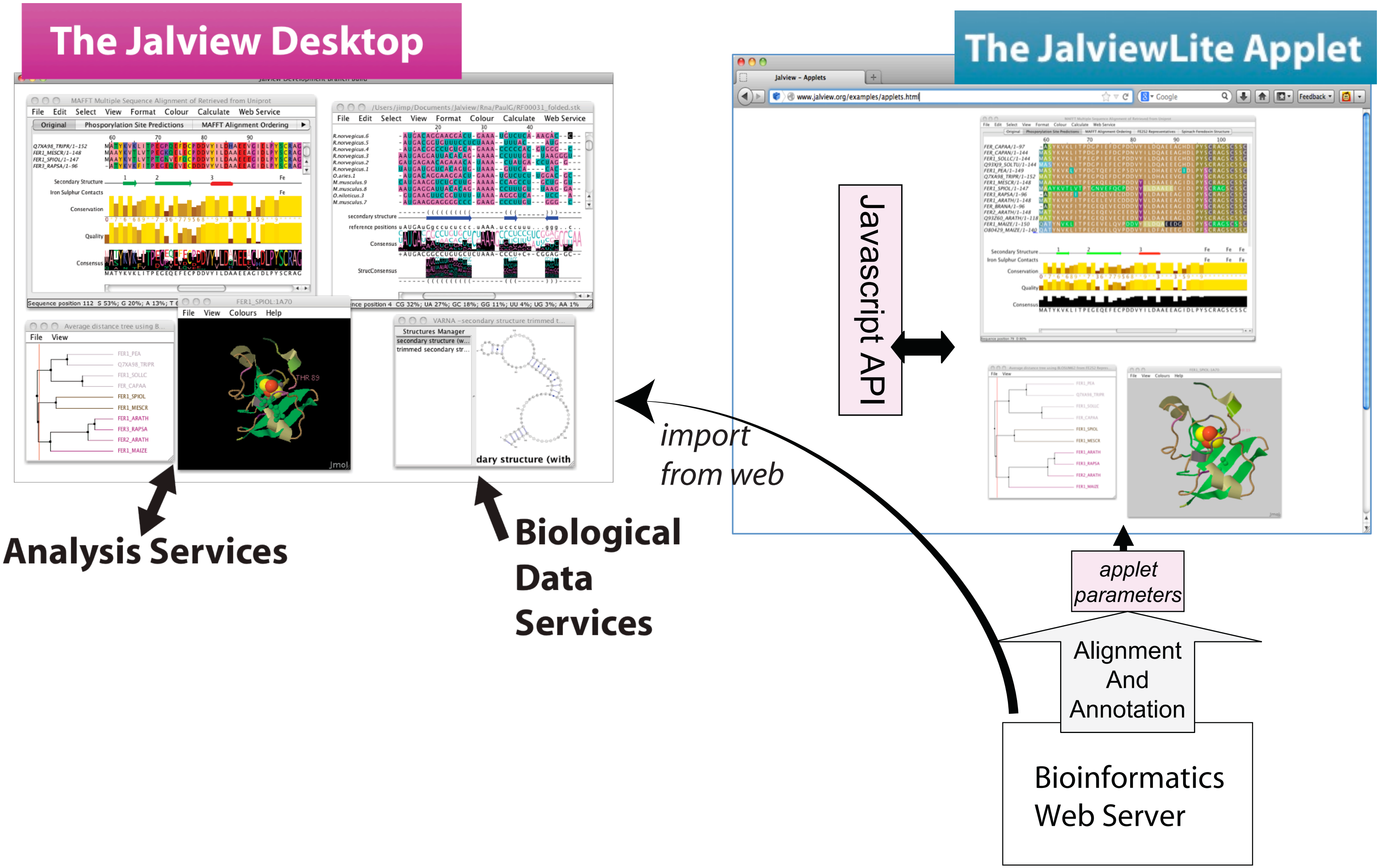
Installation packages and source

http://www.jalview.org

Jalview comes in two flavours

Launch Jalview Applet

Launch Jalview Desktop



The Jalview Desktop

The JalviewLite Applet

Javascript API

import from web

Analysis Services

Biological Data Services

applet parameters

Alignment And Annotation

Bioinformatics Web Server

Ex 1 – starting The Jalview Dekstop

PAGE 7

Use the ‘Latest Build of Current Release’

via the development page

http://www.jalview.org/development/development-builds

Webstart launch link is

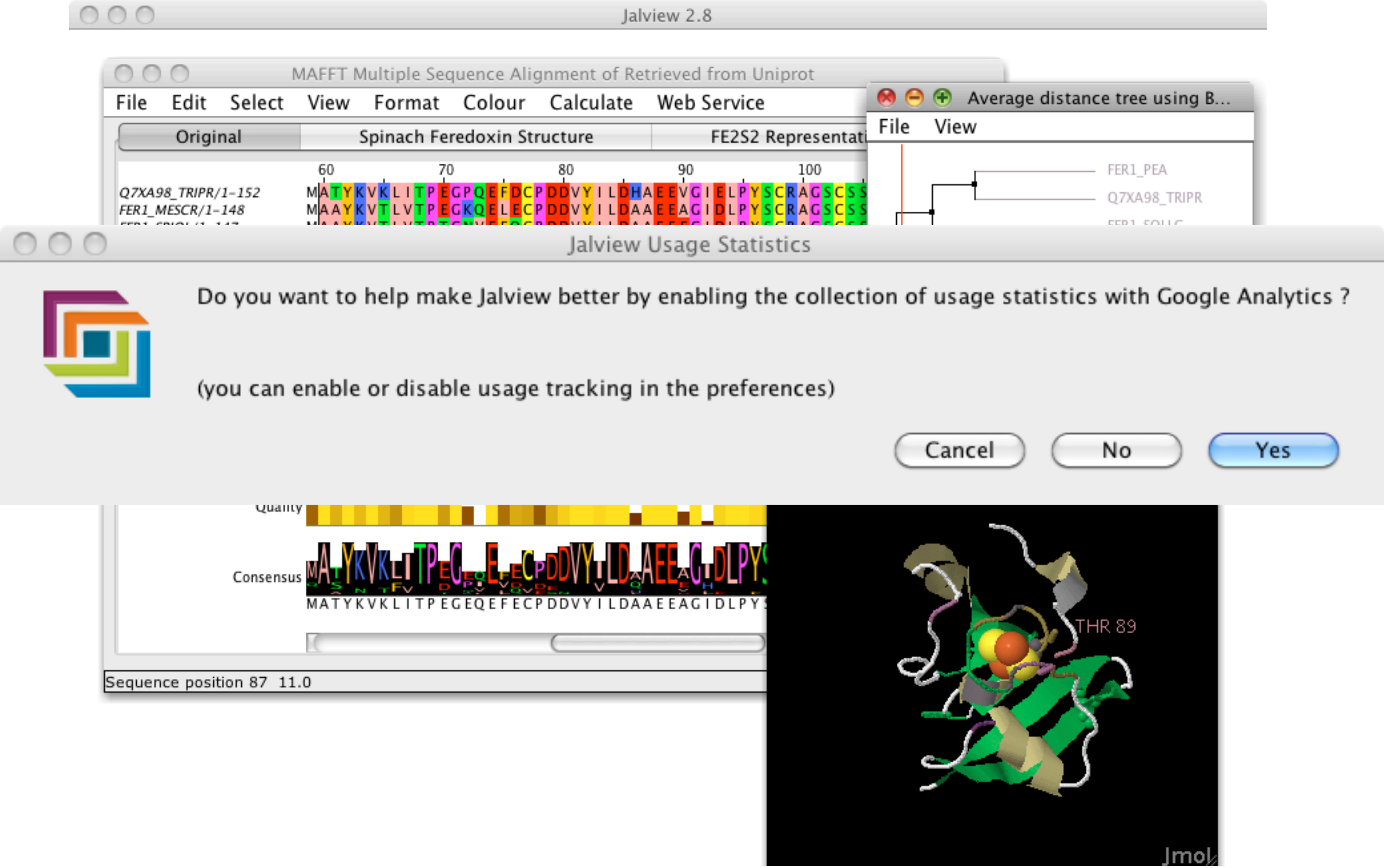
http://www.jalview.org/builds/release/webstart/jalview_2G.jnlp

Launching the jalview desktop

Version: 2.8.0b1

Last Updated: ***31 January 2014***
...Checking latest version...
Authors: Jim Procter, Andrew Waterhouse, Jan Engelhardt, Lauren Lui, Michele Clamp, James Cuff, Steve Searle, David Martin & Geoff Barton
Development managed by The Barton Group, University of Dundee, Scotland, UK.

For help, see the FAQ at www.jalview.org and/or join the jalview-discuss@jalview.org mailing list
If you use Jalview, please cite:
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



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)

Sequence position 87 11.0

Jalview News

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

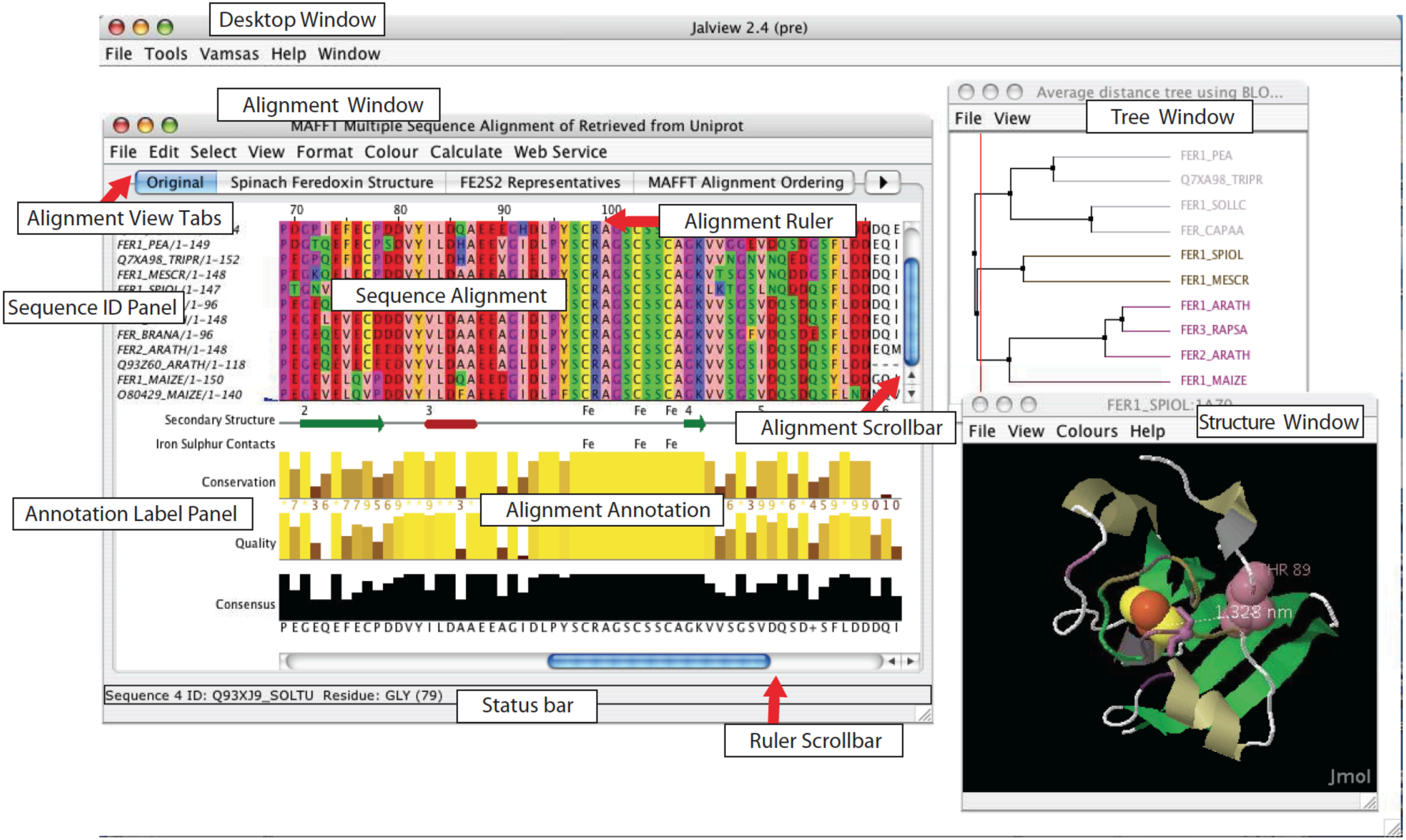
- Jan 4, 2013 Jalview in 2012 and 2013
- Nov 12, 2012 Jalview 2.8 release and the new look www.jalview.org
- Oct 18, 2012 Registration now open for 3rd Jalview Residential Training Course anc
- Sep 22, 2011 Welcome to the Jalview Desktop news channel

Jalview in 2012 and 2013
http://www.jalview.org/General/General-news/Jalview-in-2012-and-2013

2012 was quite a year here at jalview.org. A number of long running projects finally bore fruit with the launch of our new website and logo, and the release of Jalview Version 2.8. The November release of Jalview was the first to support JABAWS 2, which was launched in December 2011, and to include RNA visualization features developed by our 2010 and 2011 Google Summer of Code students: Lauren Lui and Jan Engelhart.

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

Anatomy of Jalview: Figure 1.7



Sequence 4 ID: Q93XJ9_SOLTU Residue: GLY (79)

Desktop Window

Alignment Window

Tree Window

Alignment View Tabs

Alignment Ruler

Sequence ID Panel

Sequence Alignment

Alignment Scrollbar

Structure Window

Annotation Label Panel

Alignment Annotation

Status bar

Ruler Scrollbar

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

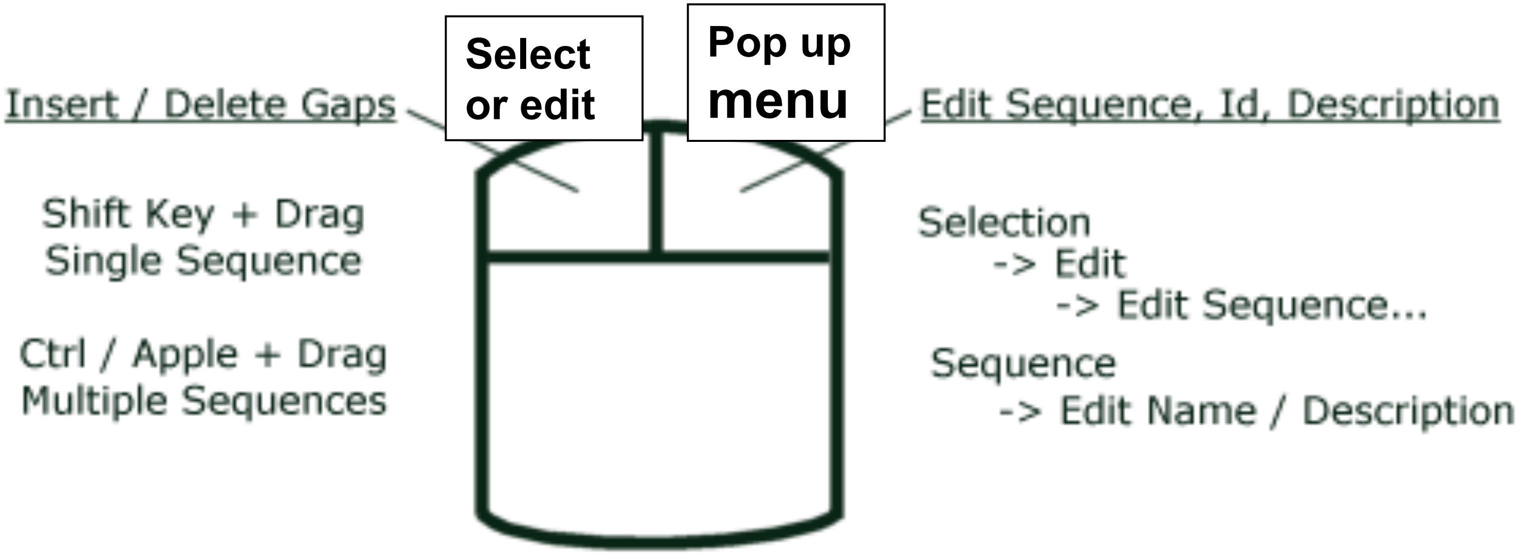
- Tasks
  - Save alignments in different formats
- Questions
  - What’s the biggest difference between a BLC file and a pileup file ?
  - Why are Jalview projects useful ?

Ex **5**,6,**7**,8 and 9 selecting, reordering, hiding/ showing and editing

- 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 at lunch.

How do I edit sequences in Jalview
http://www.jalview.org/examples/editing.html



Select or edit

Pop up menu

Insert / Delete Gaps

Shift Key + Drag
Single Sequence

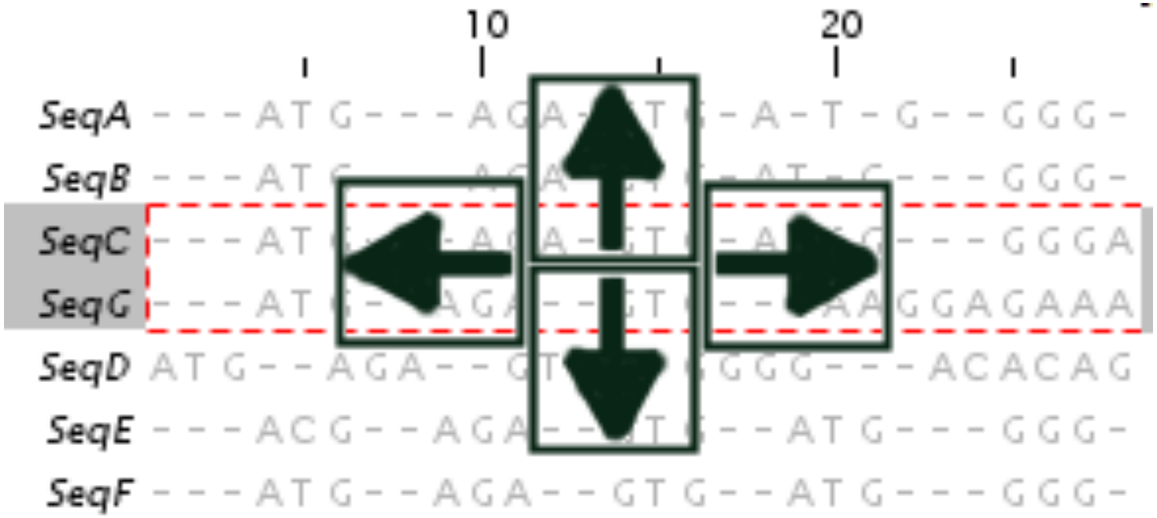
Ctrl / Apple + Drag
Multiple Sequences

Edit Sequence, Id, Description

Selection
-> Edit
-> Edit Sequence...

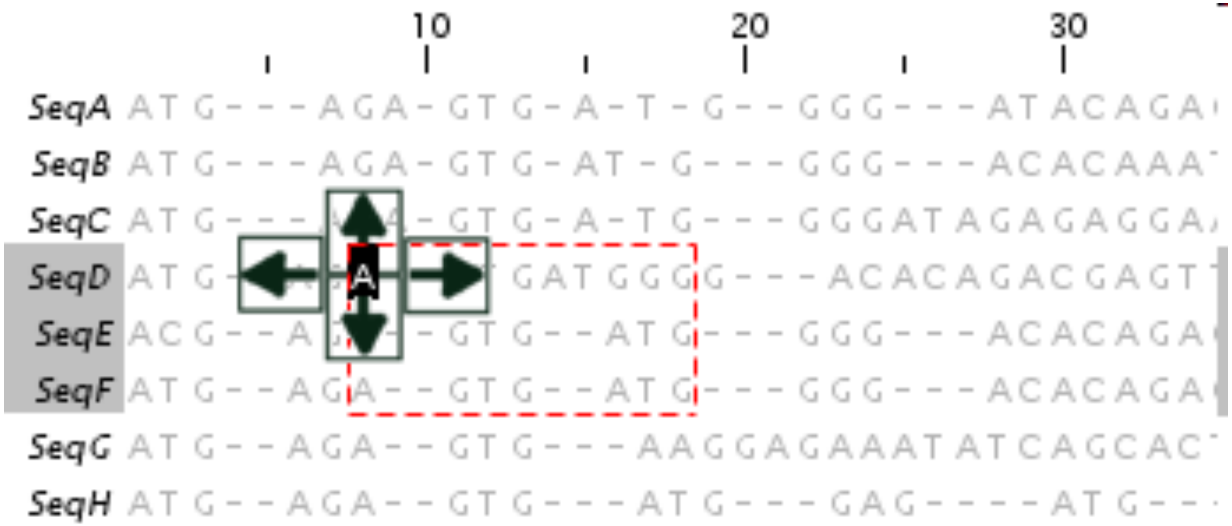
Sequence
-> Edit Name / Description

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

<u>Windows 7 Users:</u>

Group Insert - use
Shift+Space
(not CTRL+Space)

Ex 5,6,7,8 and 9 selecting, reordering, hiding/ showing and editing

- 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

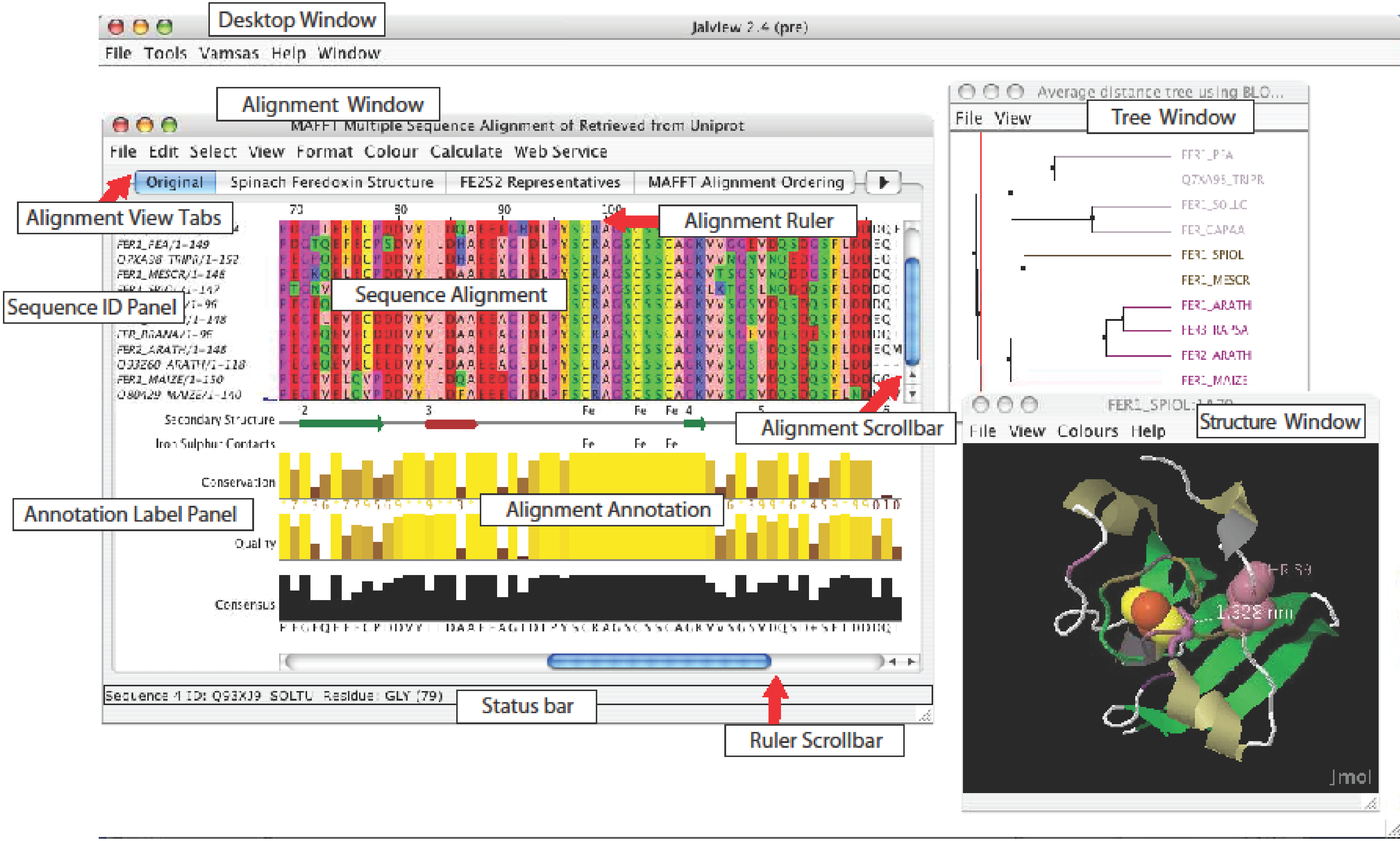
- Tasks
  - Adjust the alignment formatting options
    - Wrap
    - Sequence id margin
  - Export the alignment as a figure
    - HTML, EPS and PNG
- Questions
  - How do you control the number of columns shown in wrapped mode ?
  - How can you easily experiment with different alignment figure layouts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

- End of Session 1
  - Loading/saving
  - Navigation/Editing
  - Colouring & Figures

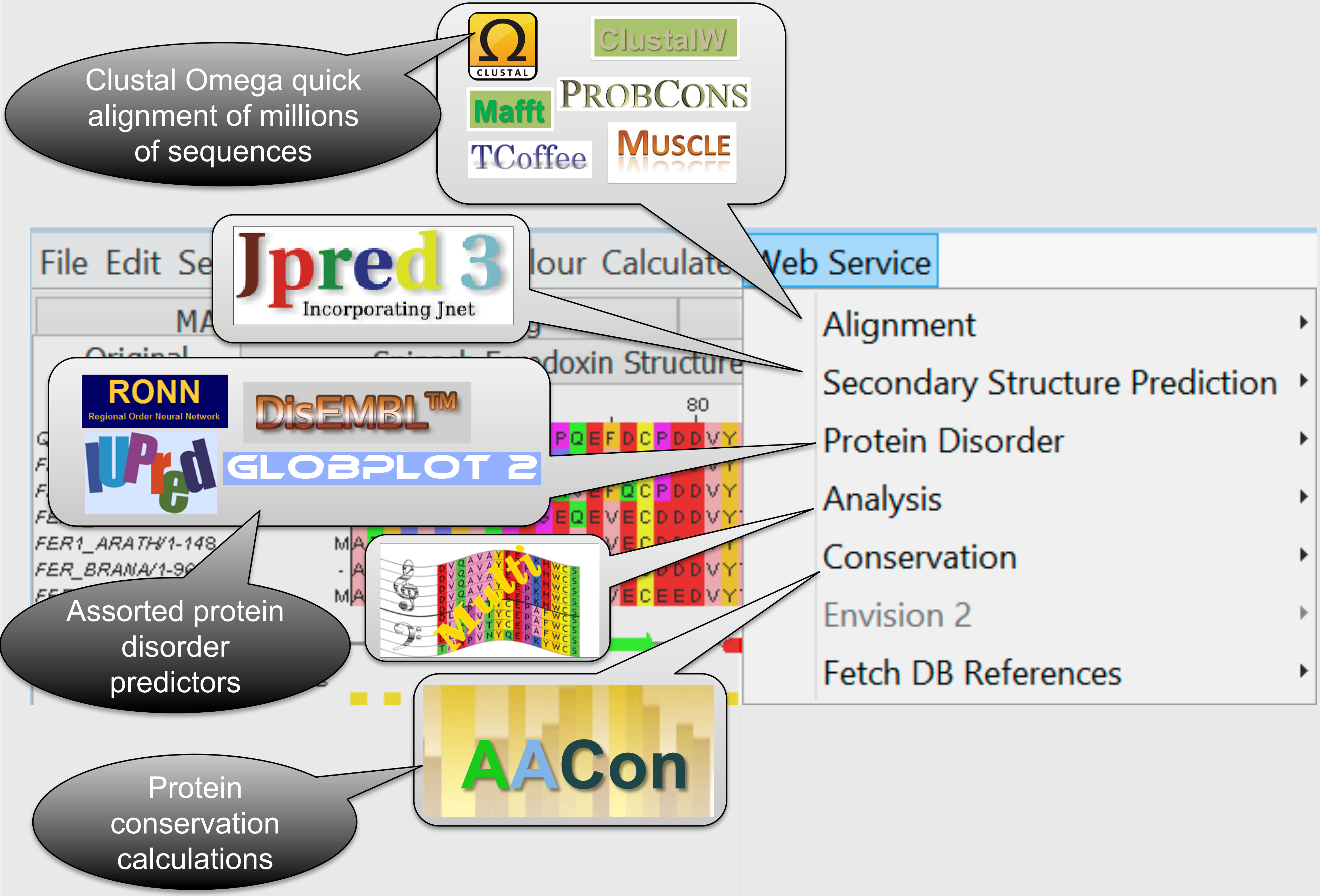
- Session 2
  - Alignment with JABAWS

    PAGE 58 IN MANUAL
  - Alignment analysis
    - Trees
    - PCA
    - Subfamily analysis

Anatomy of Jalview: Figure 1.7



Sequence 4 ID: Q93XJ9_SOLTU Residue: GLY (79)



Clustal Omega quick alignment of millions of sequences

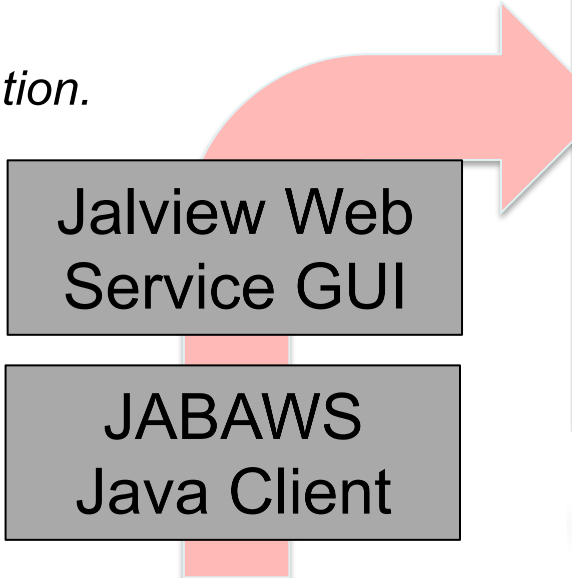
Assorted protein disorder predictors

Protein conservation calculations

www.compbio.dundee.ac.uk/jabaws

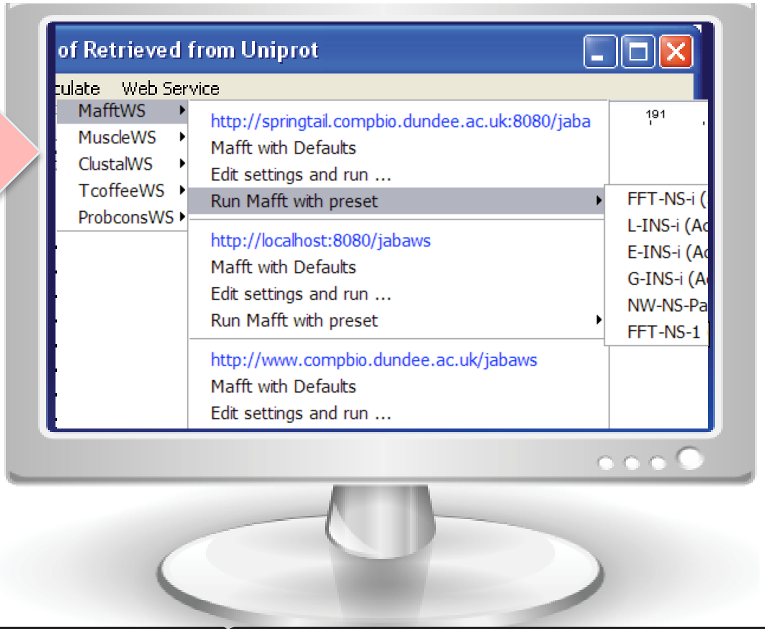
JABAWS:MSA Troshin et al. 2011, *Bioinformatics.*

JABAWS 2 In Preparation.



Jalview Web Service GUI

JABAWS Java Client



JABAWS Virtual Appliance for your private use.

vmware

Native JABAWS installs on a range of platforms

Apache Tomcat

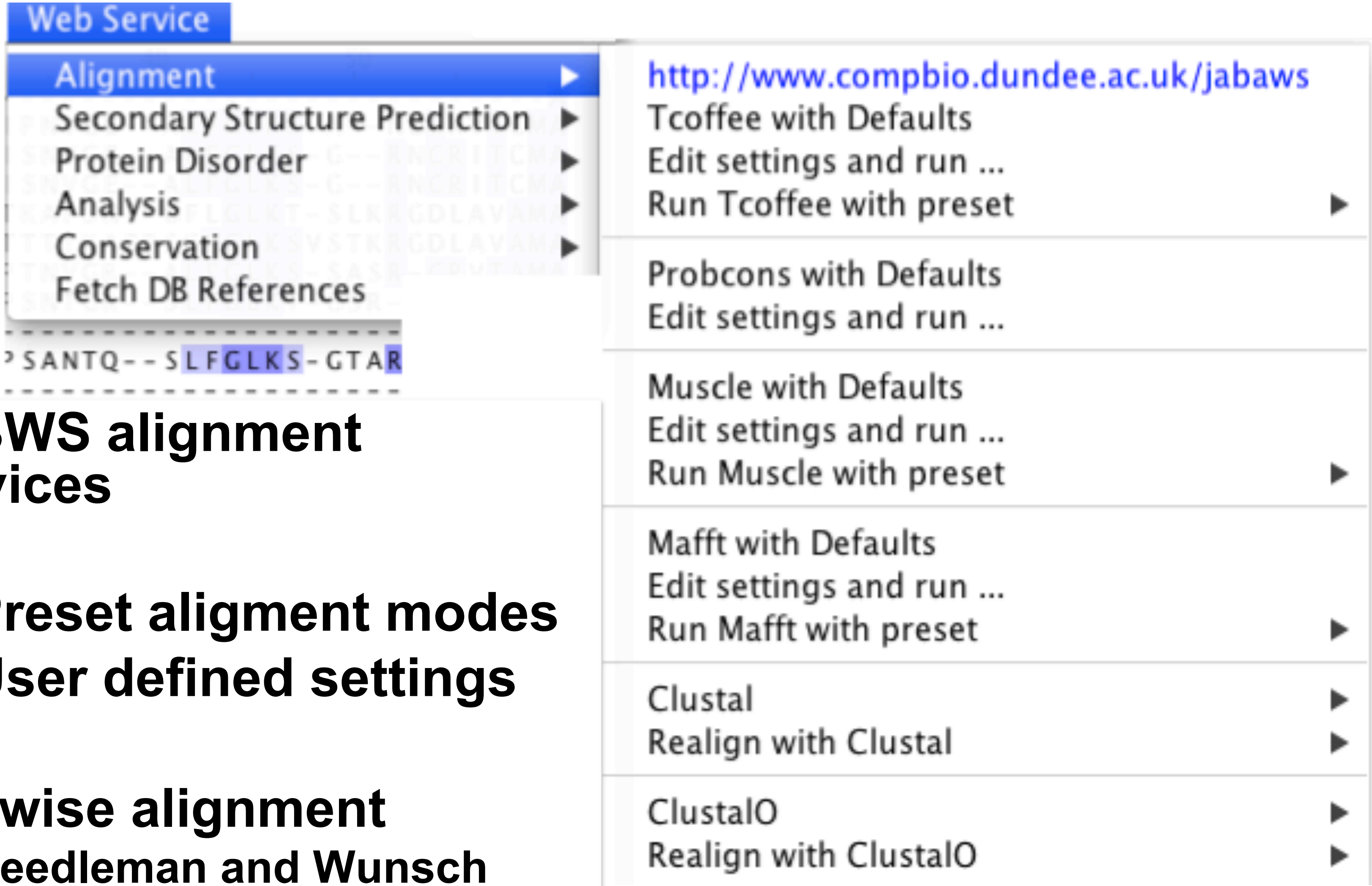
DRMAA
Distributed Resource Management Application API — www.drmaa.org

JABAWS Amazon Machine Image on EC2

powered by amazon web services™

the cloud market

Jalview's Alignment Methods



- **JABWS alignment services**
  - **Preset aligment modes**
  - **User defined settings**
- **Pairwise alignment**
  - **Needleman and Wunsch**
    - Mostly used internally

Common types of alignment algorithm

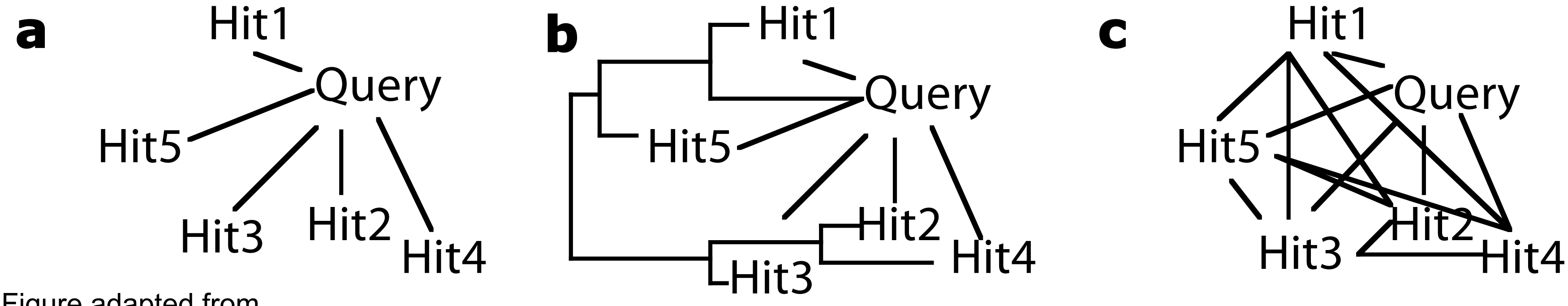


Figure adapted from
Procter et al. (2010) *Nature Methods* **7** S16 - S25

a. Sequence database searches – optimal alignment between query and hit

   e.g. **Blast** (single sequence), **PSI-Blast** and **HMMER**

b. Progressive – optimise alignment between branches on guide tree

   e.g. **ClustalW**

c. Transitive – optimise MSA to maximise consistency between pairs

   e.g. **T-COFFEE, ProbCons**

Profile methods – e.g. Muscle and MAFFT are hybrid of **b** and **c**.

Latest methods, e.g. **ClustalO**, also employ sampling strategies to speed up tree building & refinement.

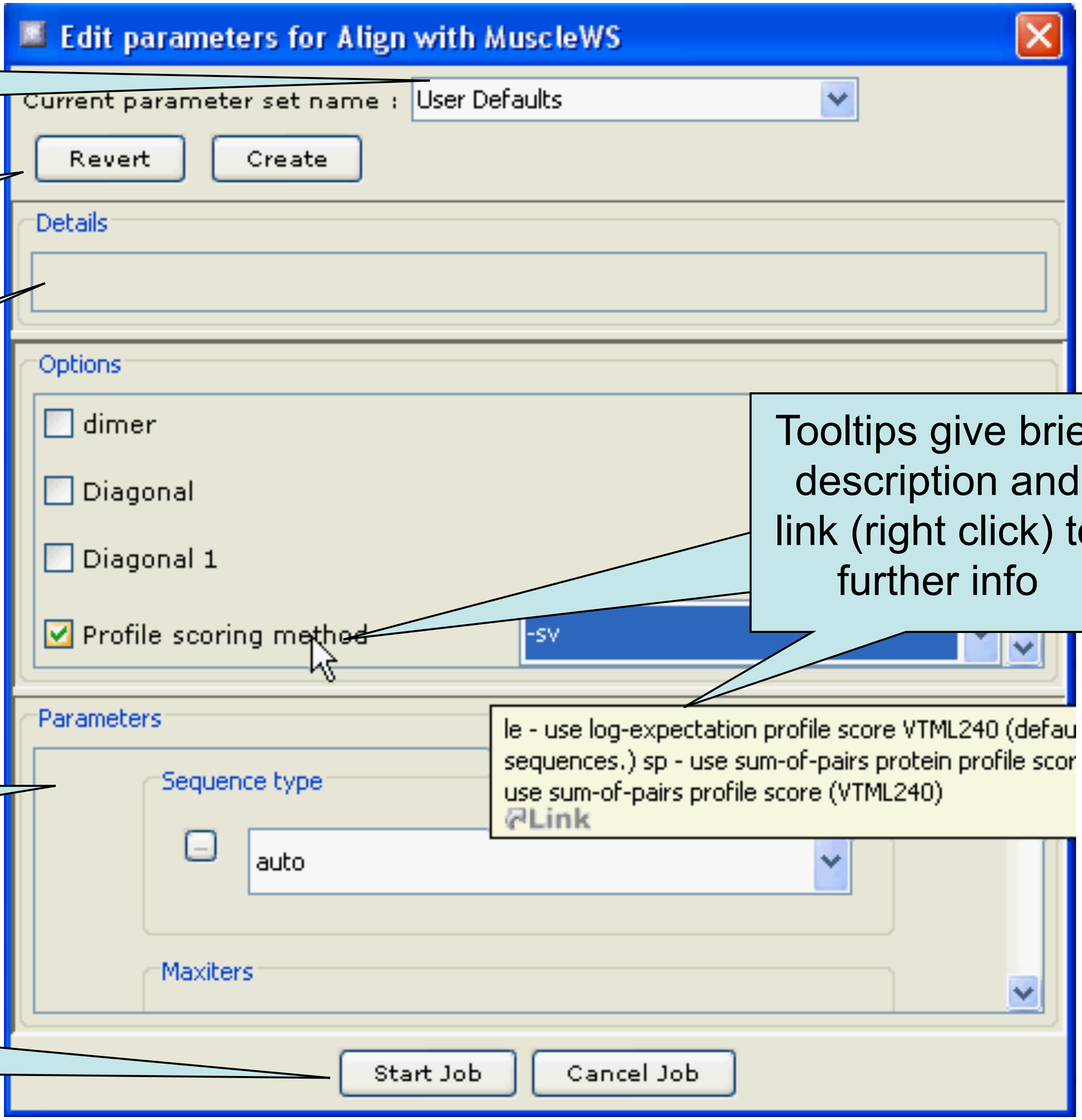
Jalview alignment exercise 25 (sect. 2.4)

- 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



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

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

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

Parameters contains more complex settings

Start job with current settings or cancel.

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 ?

- Session 2
  - Alignment with JABAWS

    PAGE 58 IN MANUAL
  - Alignment analysis
    - Trees
    - PCA
    - Subfamily analysis

Alignment analysis – Section 2.2

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

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 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.

PCA and Phylogeny 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

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

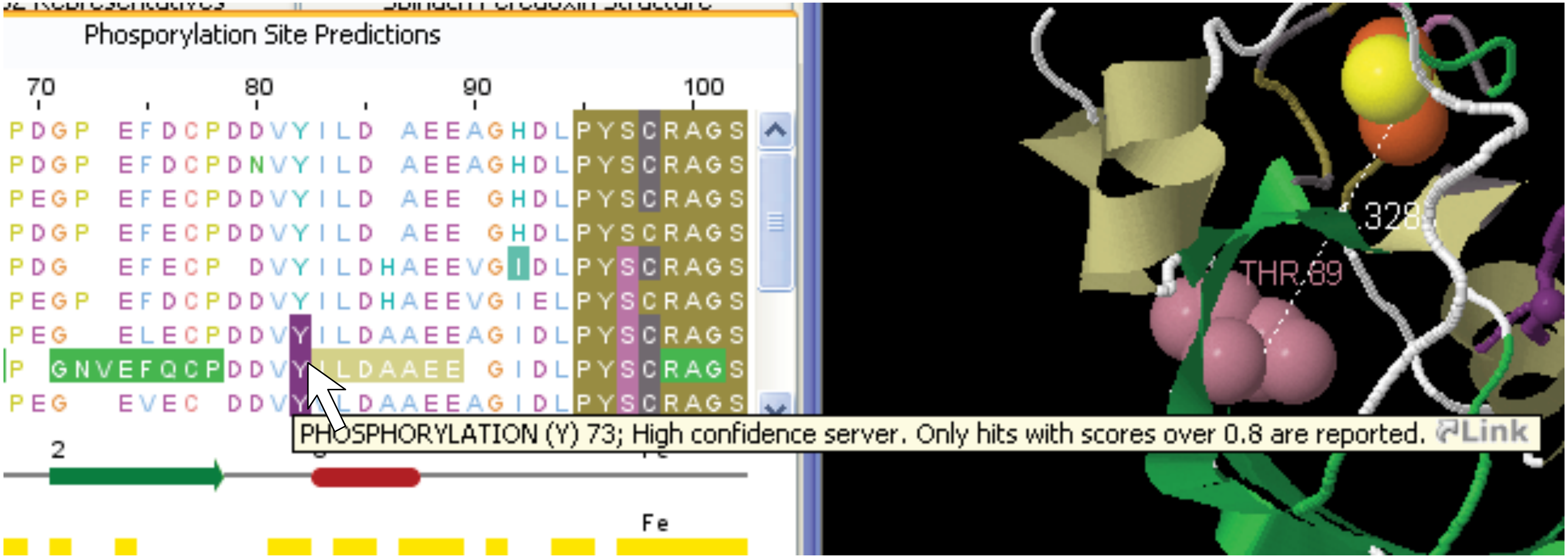
- Task
  - Use the group consensus sequence logos to more easily compare tree subgroups
  - Use ‘Make groups for selection’ to subdivide groups by specific mutation
- Questions
  - How can you work out which group is associated with which annotation row ?

- End of Session 2
    - Alignment & analysis

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

Sequence Features

Manual section 2.8

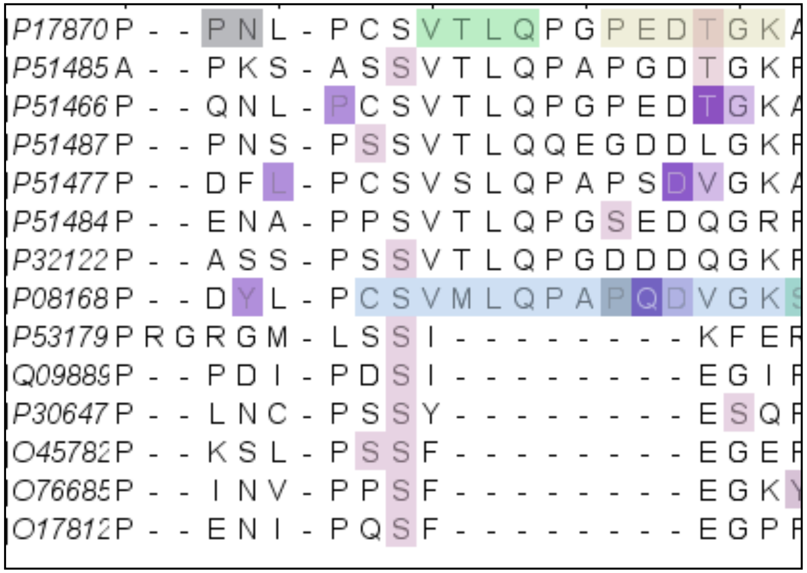


```
PDGP EFDCPDDVYILD AEEAGHDLPYSCRAGS
PDGP EFDCPDNVYILD AEEAGHDLPYSCRAGS
PEGP EFECPDDVYILD AEE GHDLPYSCRAGS
PDGP EFECPDDVYILD AEE GHDLPYSCRAGS
PDG  EFECP DVYILDHAEEVGIDLPYSCRAGS
PEGP EFDCPDDVYILDHAEEVGIELPYSCRAGS
PEG  ELECPDDVYILDAAEEAGIDLPYSCRAGS
P GNVEFQCPDDVYILDAAEE GIDLPYSCRAGS
PEG  EVEC DDVY LDAAEEAGIDLPYSCRAGS
```

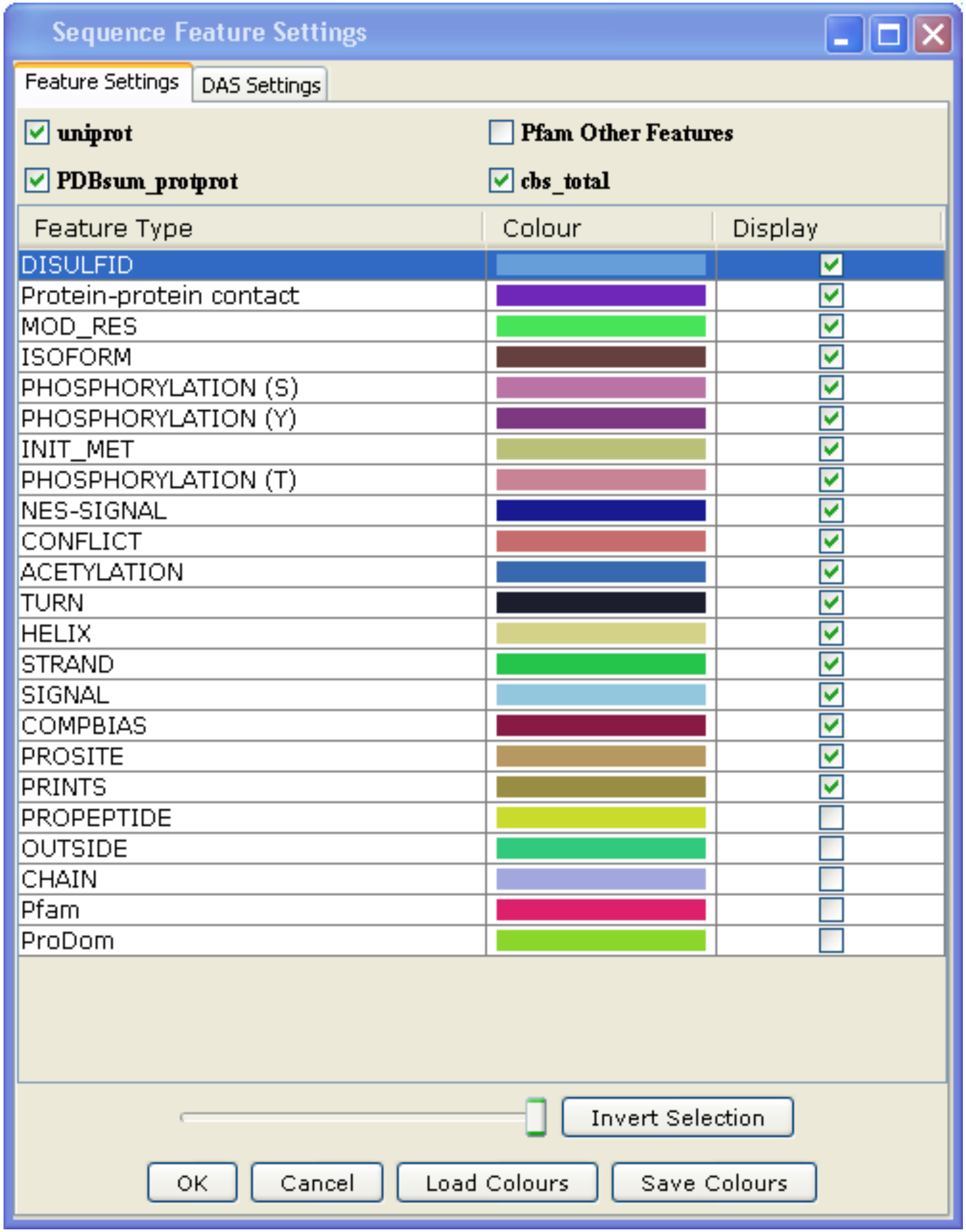
PHOSPHORYLATION (Y) 73; High confidence server. Only hits with scores over 0.8 are reported.

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



```
P17870 P - - P N L - P C S V T L Q P G P E D T G K
P51485 A - - P K S - A S S V T L Q P A P G D T G K
P51466 P - - Q N L - P C S V T L Q P G P E D T G K
P51487 P - - P N S - P S S V T L Q Q E G D D L G K
P51477 P - - D F L - P C S V S L Q P A P S D V G K
P51484 P - - E N A - P P S V T L Q P G S E D Q G R
P32122 P - - A S S - P S S V T L Q P G D D D Q G K
P08168 P - - D Y L - P C S V M L Q P A P Q D V G K
P53179 P R G R G M - L S S I - - - - - - - - - K F E
Q09889 P - - P D I - P D S I - - - - - - - - - E G I
P30647 P - - L N C - P S S Y - - - - - - - - - E S Q
O45782 P - - K S L - P S S F - - - - - - - - - E G E
O76685 P - - I N V - P P S F - - - - - - - - - E G K
O17812 P - - E N I - P Q S F - - - - - - - - - E G P
```



☑ uniprot ☐ Pfam Other Features
☑ PDBsum_protprot ☑ cbs_total

Feature Type	Colour	Display
DISULFID	🟦	☑
Protein-protein contact	🟪	☑
MOD_RES	🟩	☑
ISOFORM	🟫	☑
PHOSPHORYLATION (S)	🟪	☑
PHOSPHORYLATION (Y)	🟪	☑
INIT_MET	🟨	☑
PHOSPHORYLATION (T)	🟥	☑
NES-SIGNAL	🟦	☑
CONFLICT	🟥	☑
ACETYLATION	🟦	☑
TURN	⬛	☑
HELIX	🟨	☑
STRAND	🟩	☑
SIGNAL	🟦	☑
COMPBIAS	🟥	☑
PROSITE	🟫	☑
PRINTS	🟫	☑
PROPEPTIDE	🟨	☐
OUTSIDE	🟩	☐
CHAIN	🟪	☐
Pfam	🟥	☐
ProDom	🟩	☐

Sequence Features
Section 2.8.1-3 & Ex 27

  - Annotate the whole or part of a sequence
  - Database refs are special case.
- Tasks
  - Visualise, create, modify, import and export features.

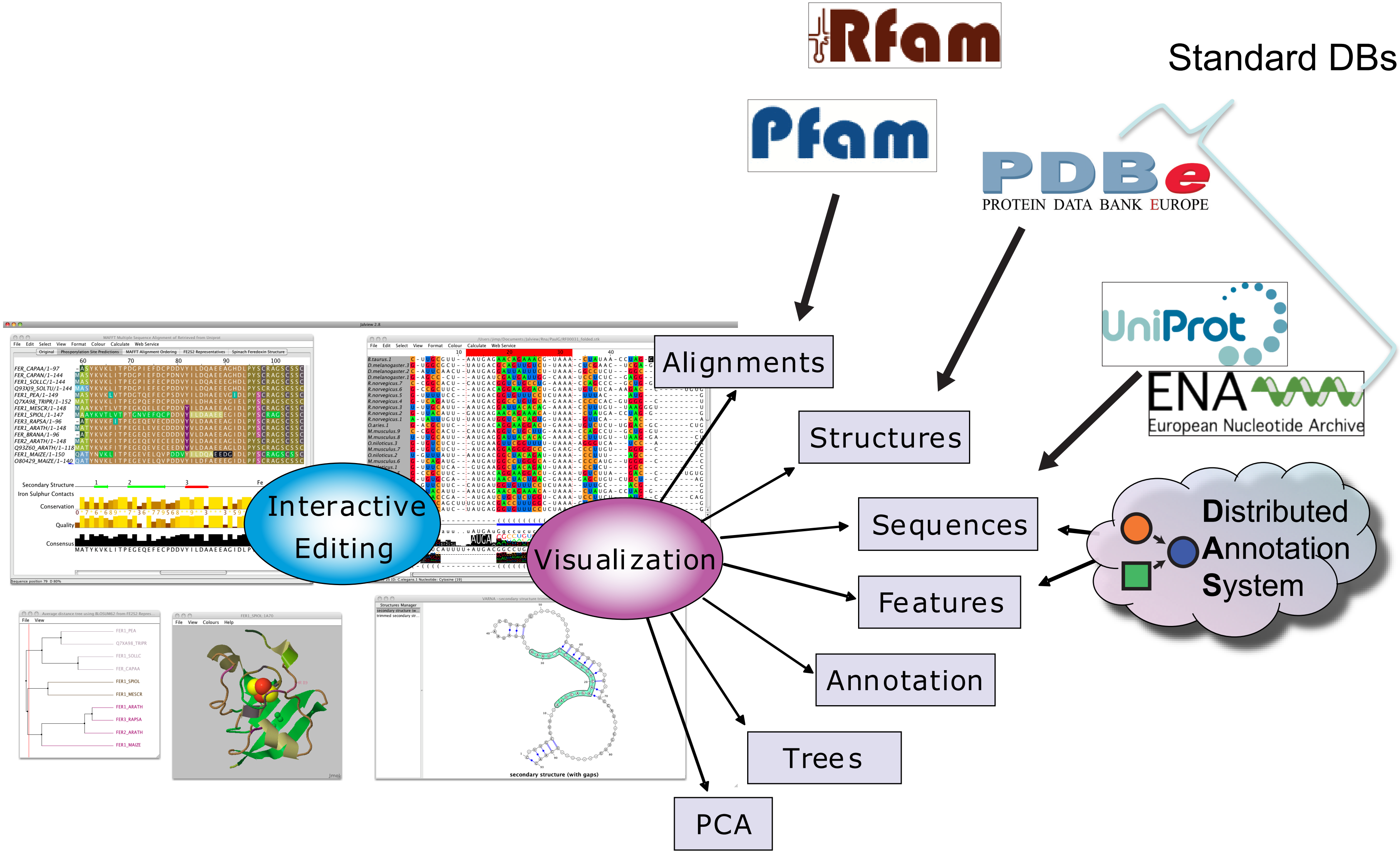
Sequence Features Section 2.6.1-3 & Ex 27

  - 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 ?

Sources of sequence feature data

- Jalview sequence annotation files
- DAS sources
- GFF files
- Certain ‘rich’ alignment formats
  - Stockholm
  - AMSA

Retrieval from External Databases



Standard DBs

Rfam

Pfam

PDBe
PROTEIN DATA BANK EUROPE

UniProt

ENA
European Nucleotide Archive

Interactive Editing

Visualization

Alignments

Structures

Sequences

Features

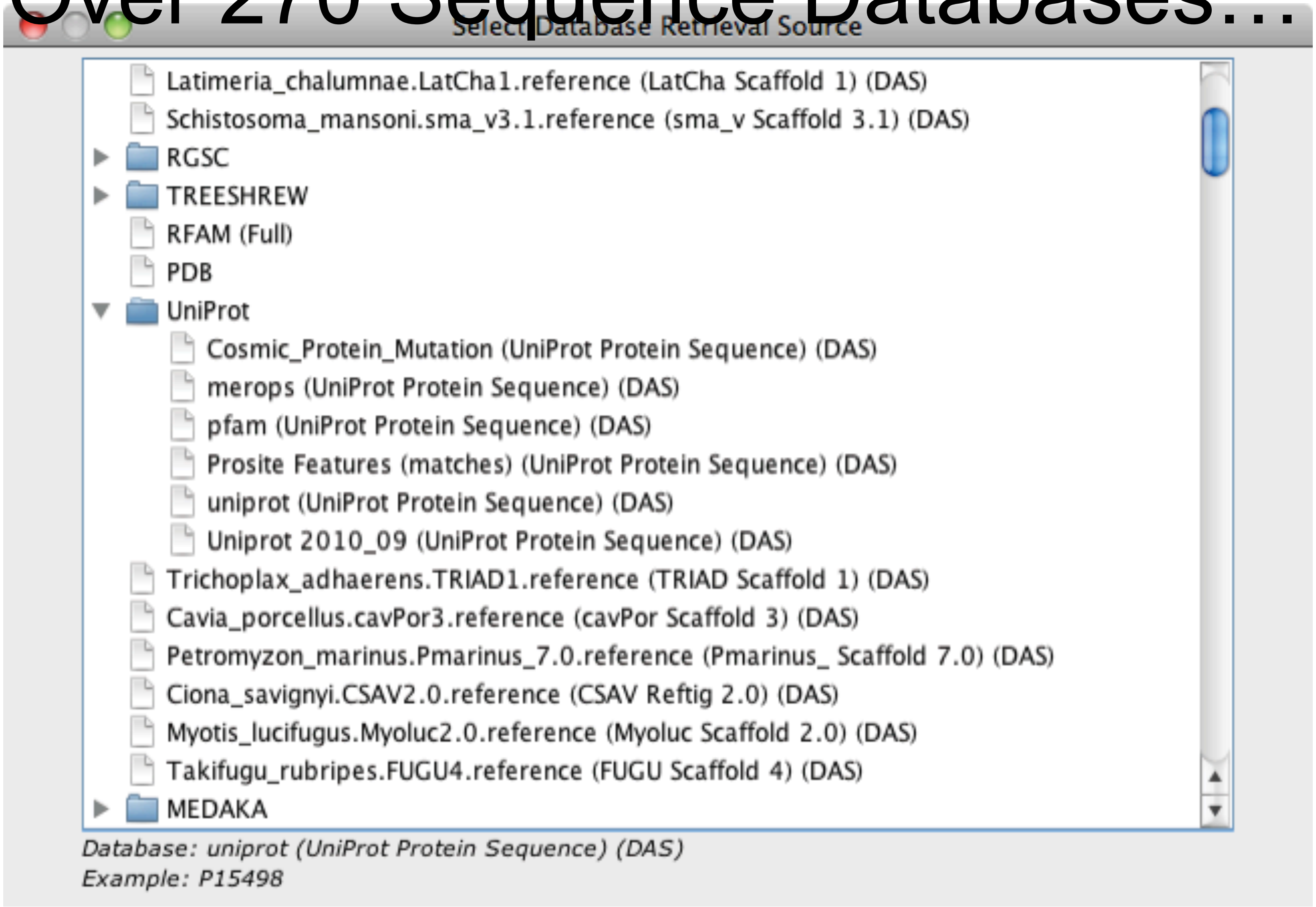
Annotation

Trees

PCA

Distributed Annotation System

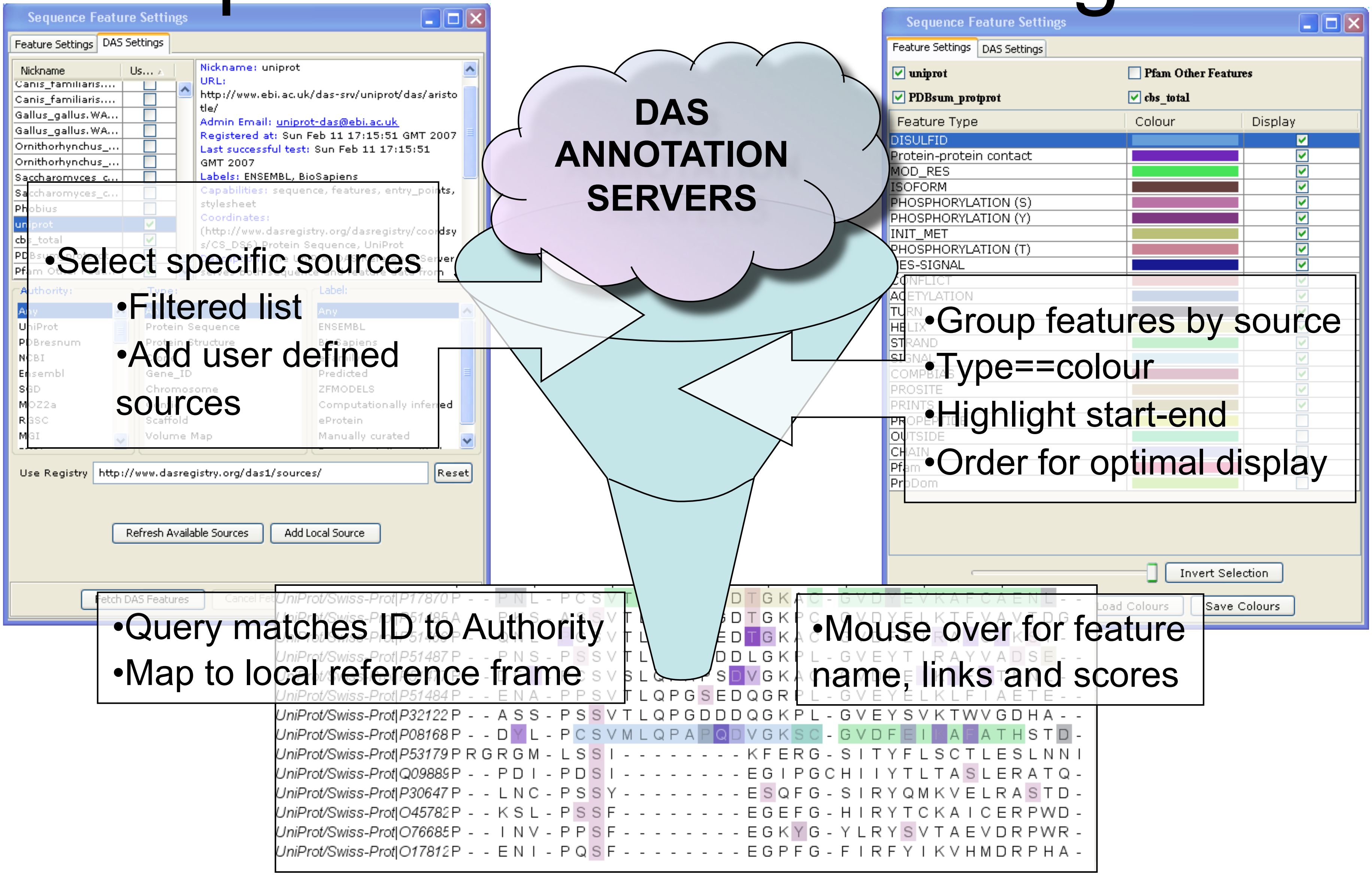
DAS allows Jalview access to Over 270 Sequence Databases…



- Latimeria_chalumnae.LatCha1.reference (LatCha Scaffold 1) (DAS)
- Schistosoma_mansoni.sma_v3.1.reference (sma_v Scaffold 3.1) (DAS)
- ▶ RGSC
- ▶ TREESHREW
- RFAM (Full)
- PDB
- ▼ UniProt
  - Cosmic_Protein_Mutation (UniProt Protein Sequence) (DAS)
  - merops (UniProt Protein Sequence) (DAS)
  - pfam (UniProt Protein Sequence) (DAS)
  - Prosite Features (matches) (UniProt Protein Sequence) (DAS)
  - uniprot (UniProt Protein Sequence) (DAS)
  - Uniprot 2010_09 (UniProt Protein Sequence) (DAS)
- Trichoplax_adhaerens.TRIAD1.reference (TRIAD Scaffold 1) (DAS)
- Cavia_porcellus.cavPor3.reference (cavPor Scaffold 3) (DAS)
- Petromyzon_marinus.Pmarinus_7.0.reference (Pmarinus_ Scaffold 7.0) (DAS)
- Ciona_savignyi.CSAV2.0.reference (CSAV Reftig 2.0) (DAS)
- Myotis_lucifugus.Myoluc2.0.reference (Myoluc Scaffold 2.0) (DAS)
- Takifugu_rubripes.FUGU4.reference (FUGU Scaffold 4) (DAS)
- ▶ MEDAKA

Database: uniprot (UniProt Protein Sequence) (DAS)
Example: P15498

Sequence Features Dialog box



Nickname: uniprot
URL: http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/
Admin Email: uniprot-das@ebi.ac.uk
Registered at: Sun Feb 11 17:15:51 GMT 2007
Last successful test: Sun Feb 11 17:15:51 GMT 2007
Labels: ENSEMBL, BioSapiens

Use Registry http://www.dasregistry.org/das1/sources/

☑ uniprot ☐ Pfam Other Features
☑ PDBsum_protprot ☑ cbs_total

Feature Type	Colour	Display
DISULFID		☑
Protein-protein contact		☑
MOD_RES		☑
ISOFORM		☑
PHOSPHORYLATION (S)		☑
PHOSPHORYLATION (Y)		☑
INIT_MET		☑
PHOSPHORYLATION (T)		☑

- Select specific sources
  - Filtered list
  - Add user defined sources

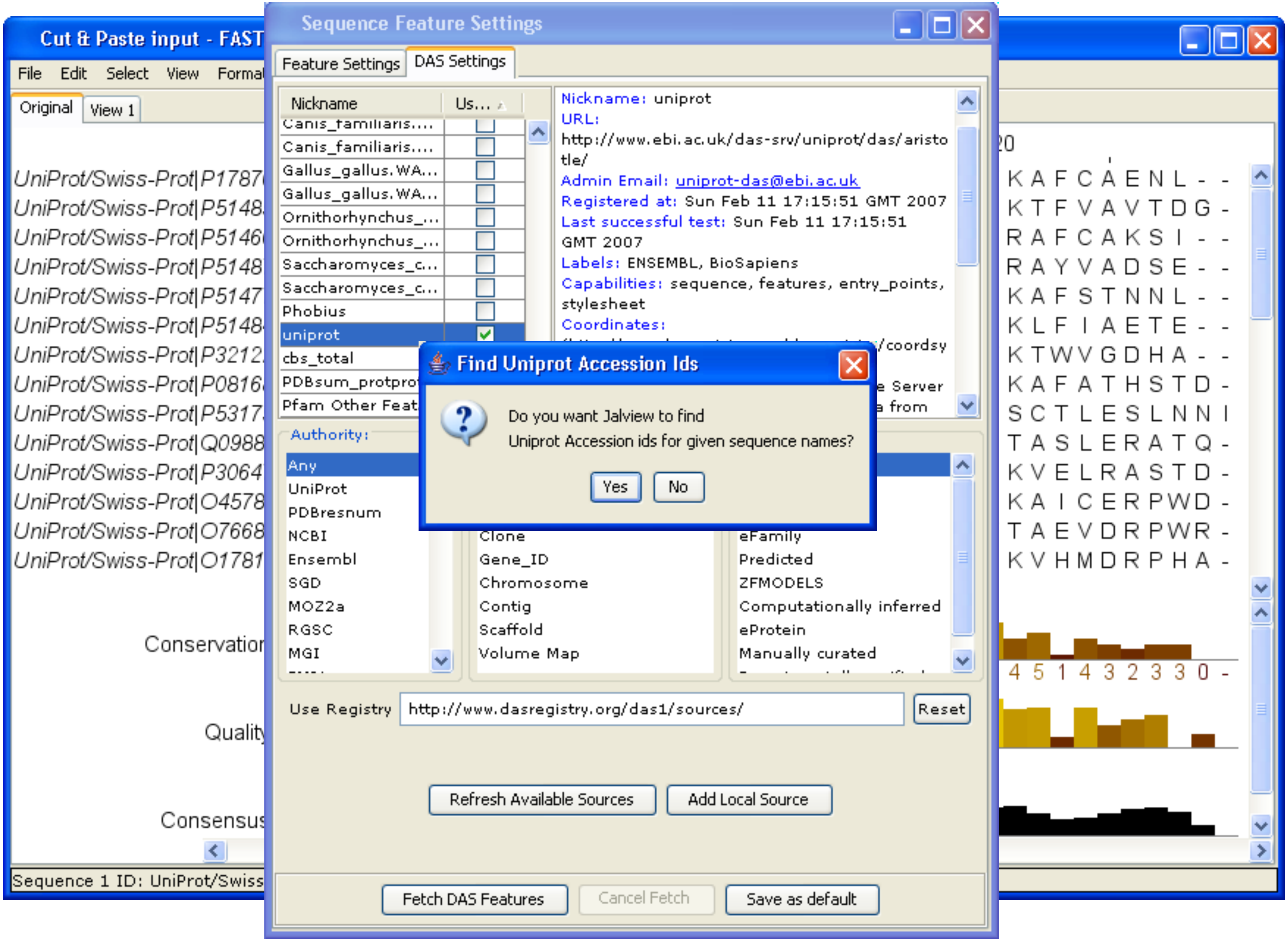
- Group features by source
- Type==colour
- Highlight start-end
- Order for optimal display

- Query matches ID to Authority
- Map to local reference frame

- Mouse over for feature name, links and scores

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
  - Verify retrieval by examining sequence annotation



Nickname: uniprot
URL:
http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/
Admin Email: uniprot-das@ebi.ac.uk
Registered at: Sun Feb 11 17:15:51 GMT 2007
Last successful test: Sun Feb 11 17:15:51 GMT 2007
Labels: ENSEMBL, BioSapiens
Capabilities: sequence, features, entry_points, stylesheet
Coordinates:

Use Registry: http://www.dasregistry.org/das1/sources/

Find Uniprot Accession Ids

Do you want Jalview to find Uniprot Accession ids for given sequence names?

The Distributed Annotation System
Section 2.9.2, 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.

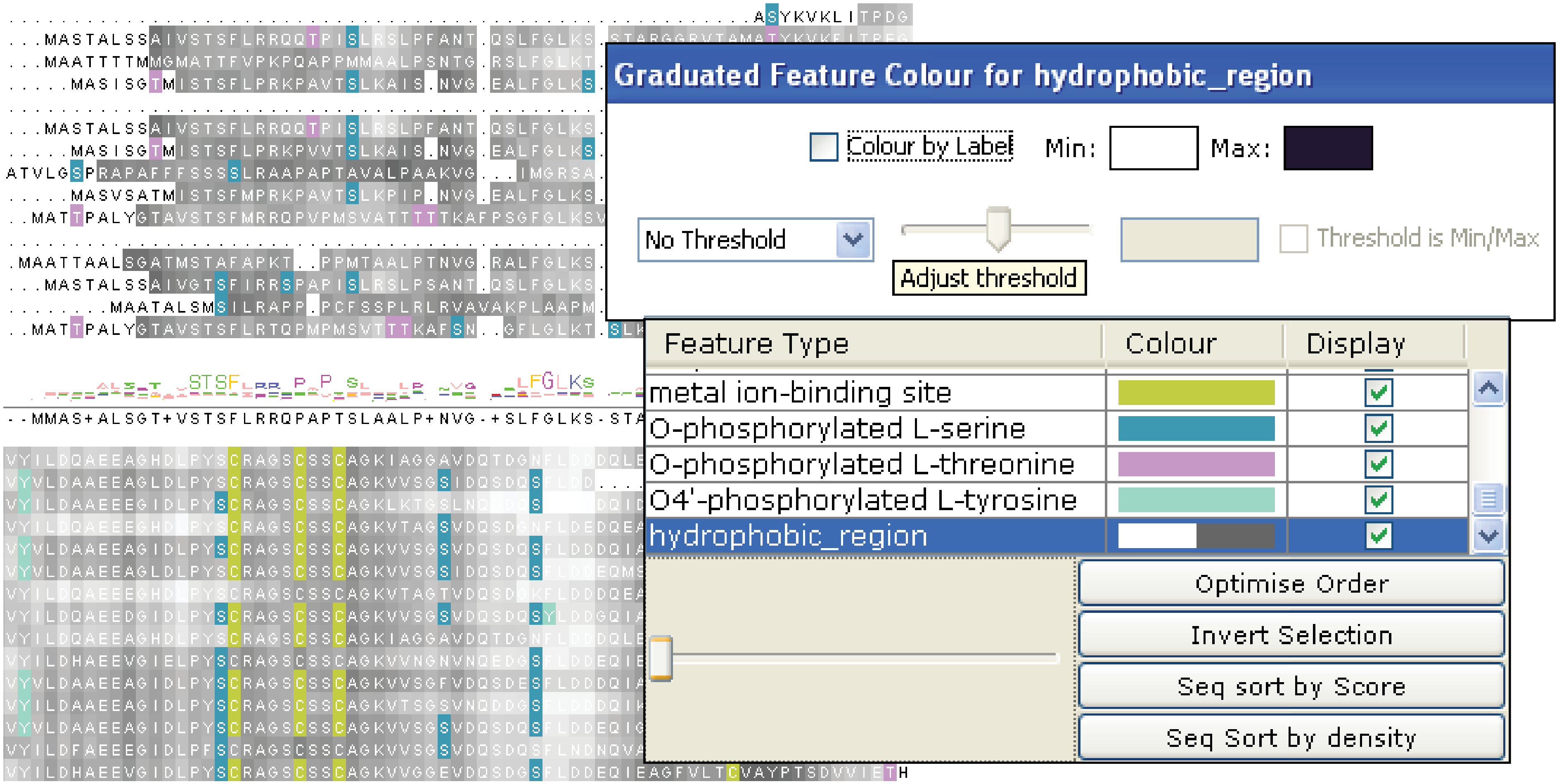
The Distributed Annotation System
Section 2.9.2, Exercise 30

  - 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 2.9.4)

- 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 ▾ Adjust 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 2.9.4)

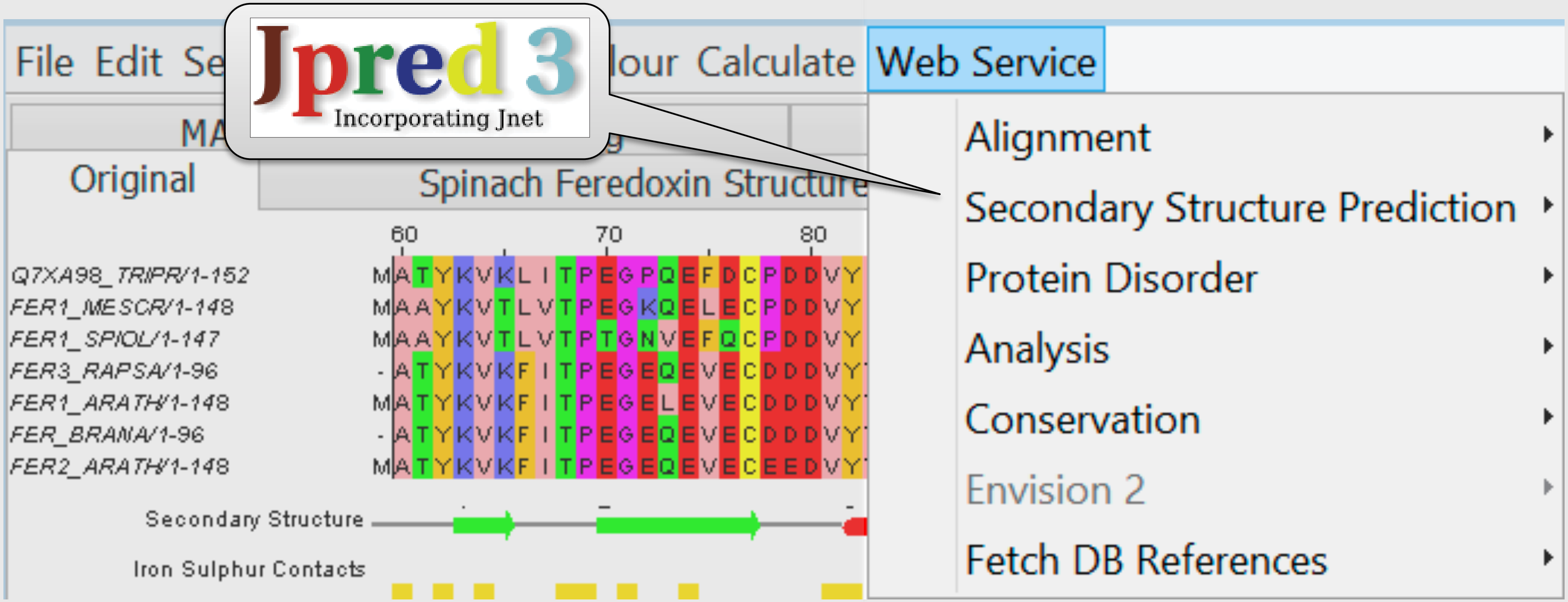
- Task
  - Shading features using labels and scores
  - Sorting alignment using feature scores
- Questions
  - What types of features are best displayed with a ‘label’ colourscheme ?
  - How would you display only the highest or lowest scoring features ?

Protein secondary structure prediction

Section 2.6 onwards

in the Manual

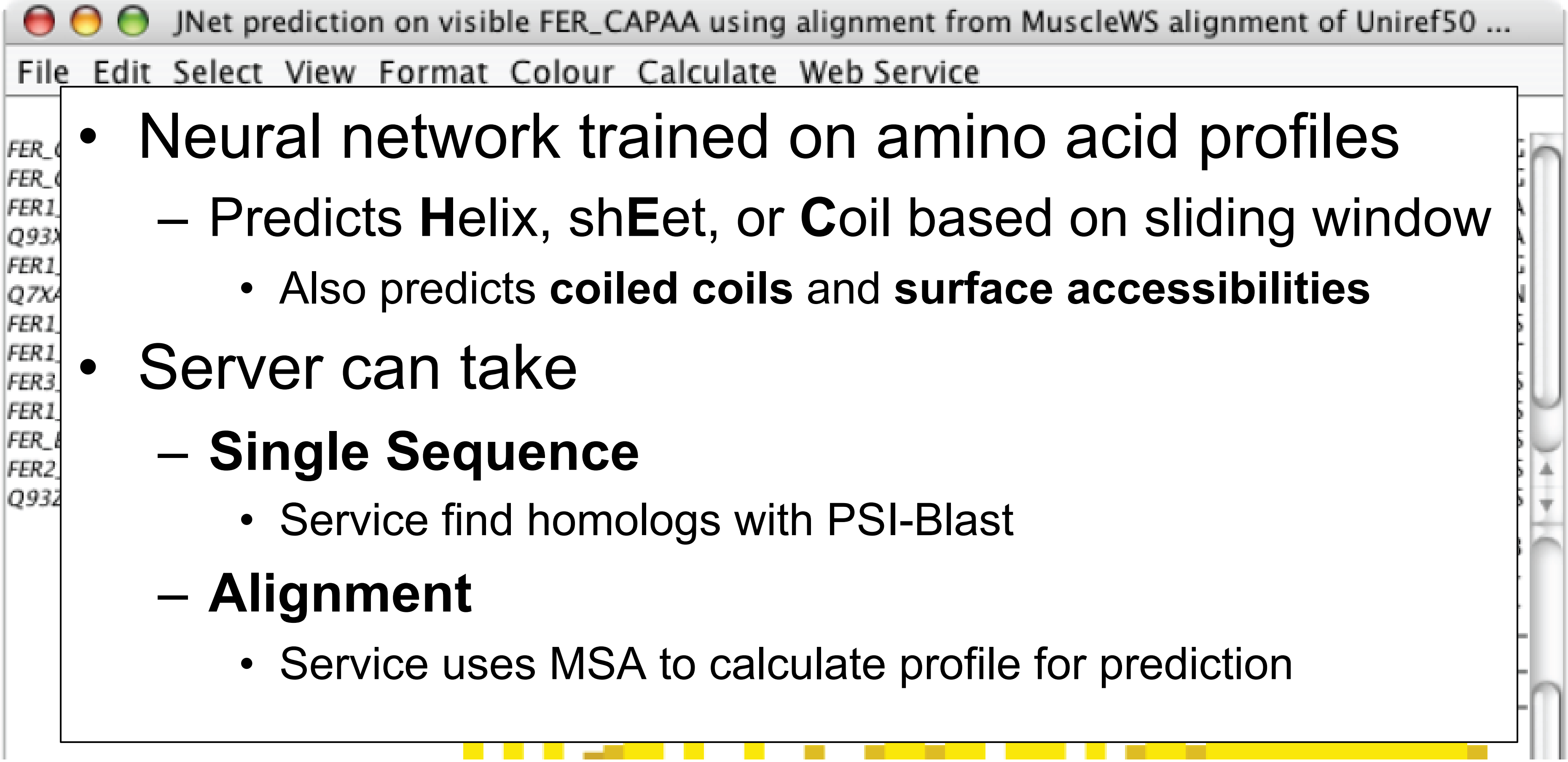
Page 65



```
Q7XA98_TRIPR/1-152  MATYKVKLITPEGPQEFDCPDDVY
FER1_MESCR/1-148    MAAYKVTLVTPEGKQELECPDDVY
FER1_SPIOL/1-147    MAAYKVTLVTPTGNVEFQCPDDVY
FER3_RAPSA/1-96     -ATYKVKFITPEGEQEVECDDDVY
FER1_ARATH/1-148    MATYKVKFITPEGELEVECDDDVY
FER_BRANA/1-96      -ATYKVKFITPEGEQEVECDDDVY
FER2_ARATH/1-148    MATYKVKFITPEGEQEVECEEDVY
```

Jpred 3
Incorporating Jnet

Protein Secondary Structure Prediction Sec. 2.6


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

- Neural network trained on amino acid profiles
  - Predicts **H**elix, sh**E**et, or **C**oil based on sliding window
    - Also predicts **coiled coils** and **surface accessibilities**
- Server can take
  - **Single Sequence**
    - Service find homologs with PSI-Blast
  - **Alignment**
    - Service uses MSA to calculate profile for prediction

Exercise 26

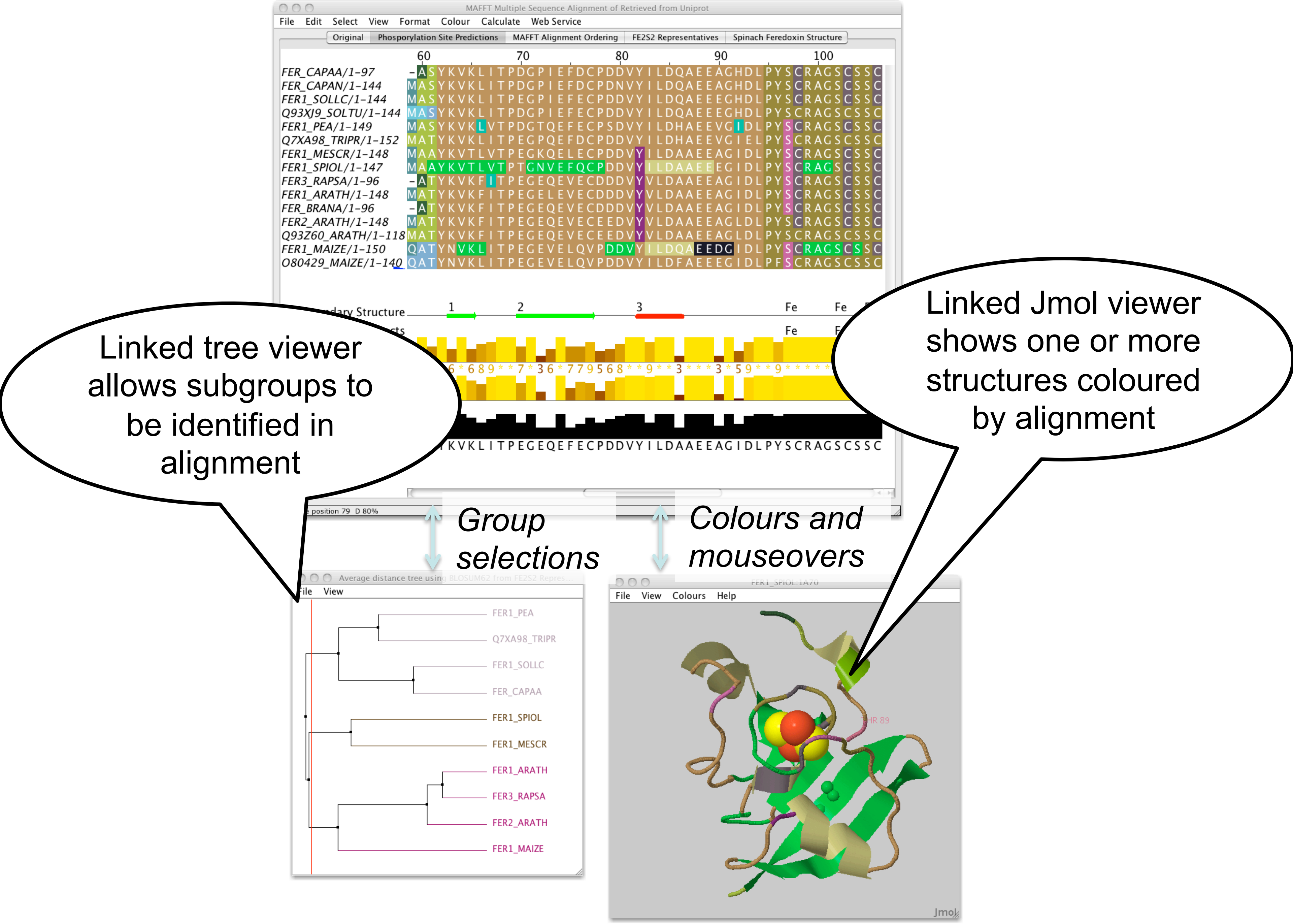
- Tasks
  - Perform a variety of Jnet predictions
    - Note the effect of hidden regions
    - Learn about sequence associated annotation
- Questions
  - …

Exercise 26

- Tasks
  - Perform a variety of Jnet predictions
    - Note the effect of hidden regions
    - Learn about sequence associated annotation

- 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 ?

- Session 4
    - Working with structures
        - Viewing 3D structures
    - Mapping data onto structure
        - Disorder prediction
        - Alignment reliability
    - RNA Structure



```
FER_CAPAA/1-97      -ASYKVKLITPDGPIEFDCPDDVYILDQAEEAGHDLPYSCRAGSCSSC
FER_CAPAN/1-144     MASYKVKLITPDGPIEFDCPDNVYILDQAEEAGHDLPYSCRAGSCSSC
FER1_SOLLC/1-144    MASYKVKLITPEGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSC
Q93XJ9_SOLTU/1-144  MASYKVKLITPDGPIEFECPDDVYILDQAEEEGHDLPYSCRAGSCSSC
FER1_PEA/1-149      MASYKVKLVTPDGTQEFECPSDVYILDHAEEVGIDLPYSCRAGSCSSC
Q7XA98_TRIPR/1-152  MATYKVKLITPEGPQEFDCPDDVYILDHAEEVGIELPYSCRAGSCSSC
FER1_MESCR/1-148    MAAYKVTLVTPEGKQELECPDDVYILDAAEEAGIDLPYSCRAGSCSSC
FER1_SPIOL/1-147    MAAYKVTLVTPTGNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSC
FER3_RAPSA/1-96     -ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER1_ARATH/1-148    MATYKVKFITPEGELEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER_BRANA/1-96      -ATYKVKFITPEGEQEVECDDDVYVLDAAEEAGIDLPYSCRAGSCSSC
FER2_ARATH/1-148    MATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSC
Q93Z60_ARATH/1-118  MATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSC
FER1_MAIZE/1-150    QATYNVKLITPEGEVELQVPDDVYILDQAEEDGIDLPYSCRAGSCSSC
O80429_MAIZE/1-140  QATYNVKLITPEGEVELQVPDDVYILDFAEEEGIDLPFSCRAGSCSSC

                      KVKLITPEGEQEFECPDDVYILDAAEEAGIDLPYSCRAGSCSSC
```

Linked tree viewer allows subgroups to be identified in alignment

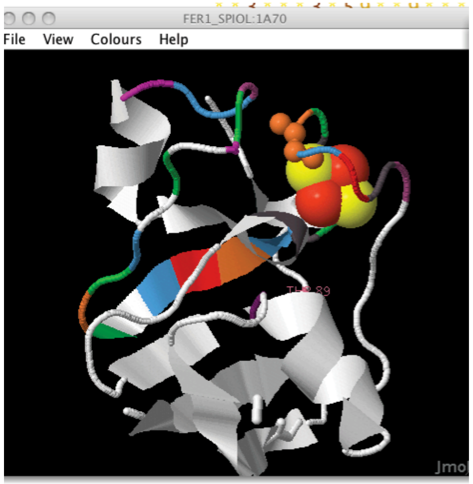
Group selections

Colours and mouseovers

Linked Jmol viewer shows one or more structures coloured by alignment

Desktop Structure Visualization
3D structures and 2D RNA diagrams

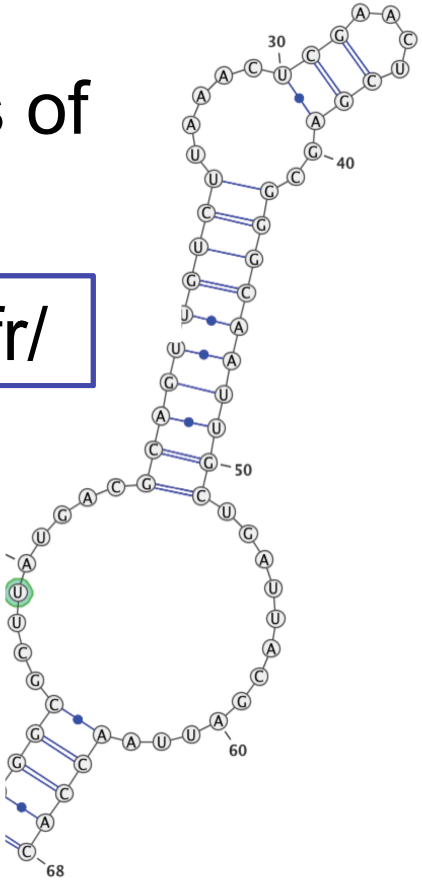
http://jmol.sourceforge.net/



VARNA

Visual Analysis of RNA

http://varna.lri.fr/



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

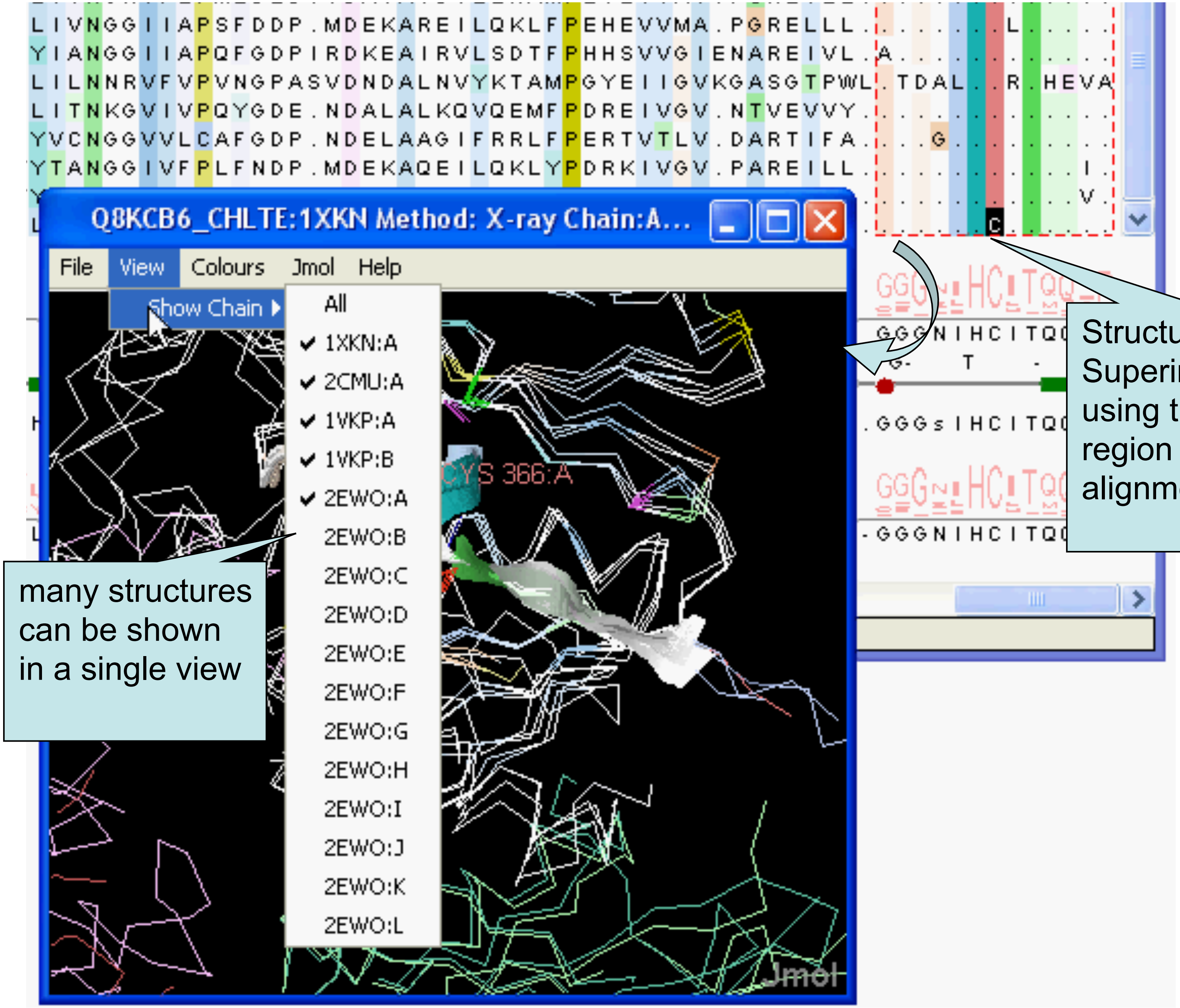
Protein Structures in Jalview
Sec 2.1. Exercise 14

- Task
  - Discover PDB structures for ferredoxin sequence(s)
    - **Note use**

      Fetch Database Refs->UNIPROT->Uniprot
  - 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



Structures can be Superimposed using the visible region of alignment

many structures can be shown in a single view

Superposing Structures using Alignments
Sec 2.1.4 – Exercise 15

- Task
  - Align structures using the ferredoxin alignment

If ‘View all N structures’ doesn’t align structures:

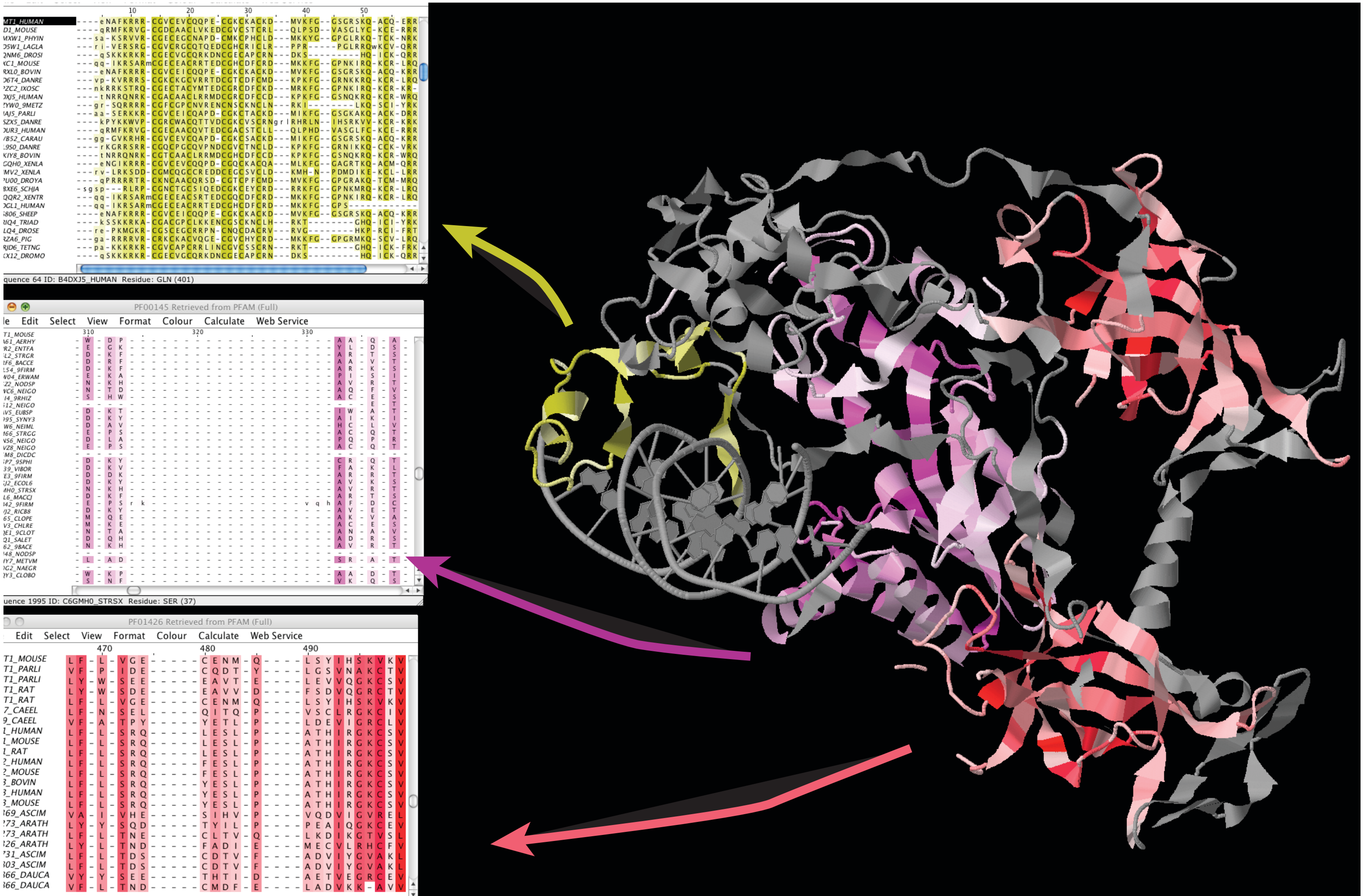
- **Use Jmol->Align menu**

  - Experiment with views to control what part of the alignment is used to superimpose the structures

Superposing Structures using Alignments
Sec 2.1.4 – Exercise 15

- Task
  - Align structures using the ferredoxin alignment
  - 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



quence 64 ID: B4DXJ5_HUMAN Residue: GLN (401)

PF00145 Retrieved from PFAM (Full)

uence 1995 ID: C6GMH0_STRSX Residue: SER (37)

PF01426 Retrieved from PFAM (Full)

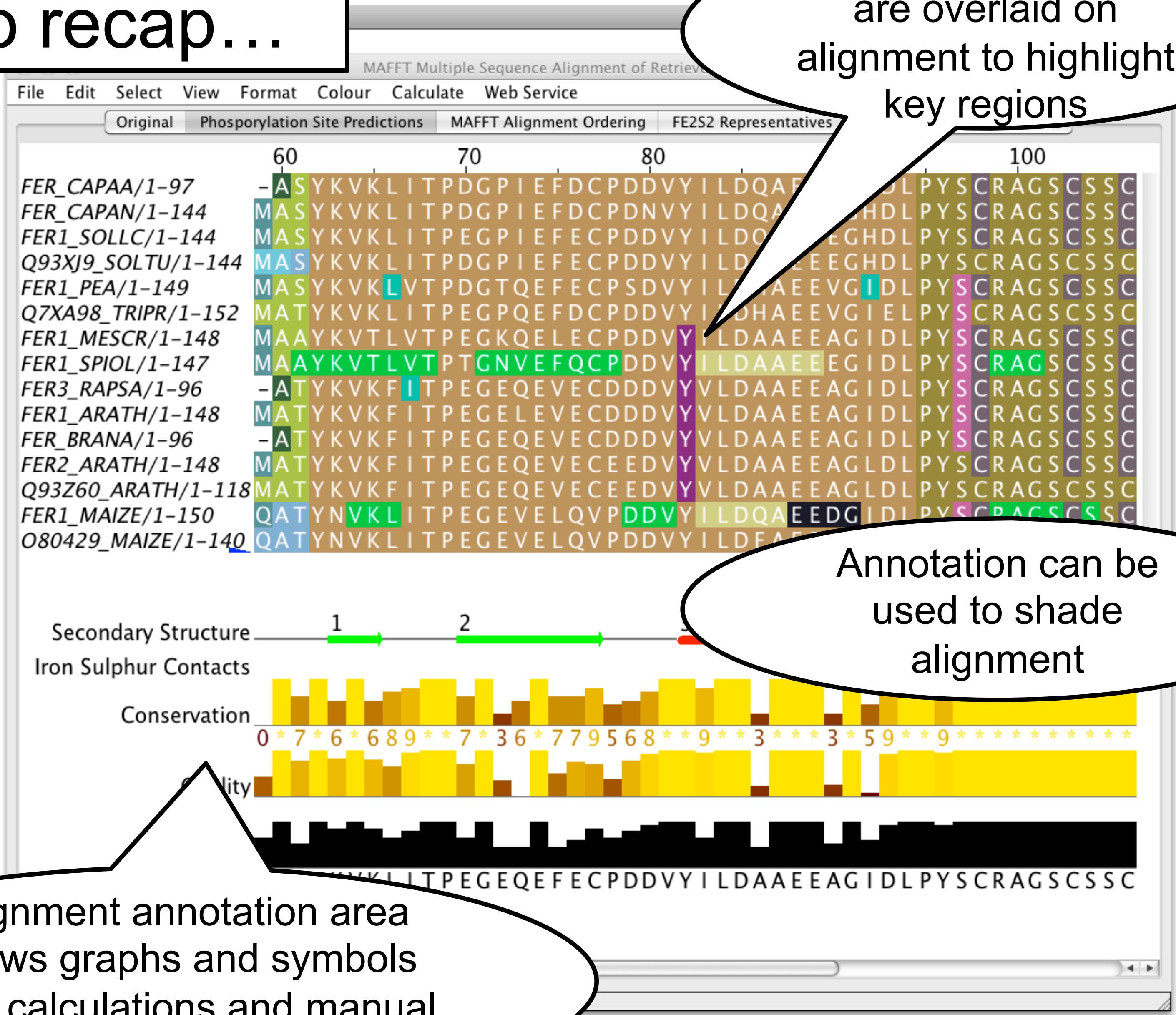
Supplementary exercises

- View the PDF from

  www.jalview.org/training/tutorial-materials/2014/Dundee/Apr

- 3 Short exercises
  - Protein Disorder prediction
  - Calculating/importing alignment quality scores
  - RNA Secondary & Tertiary structure

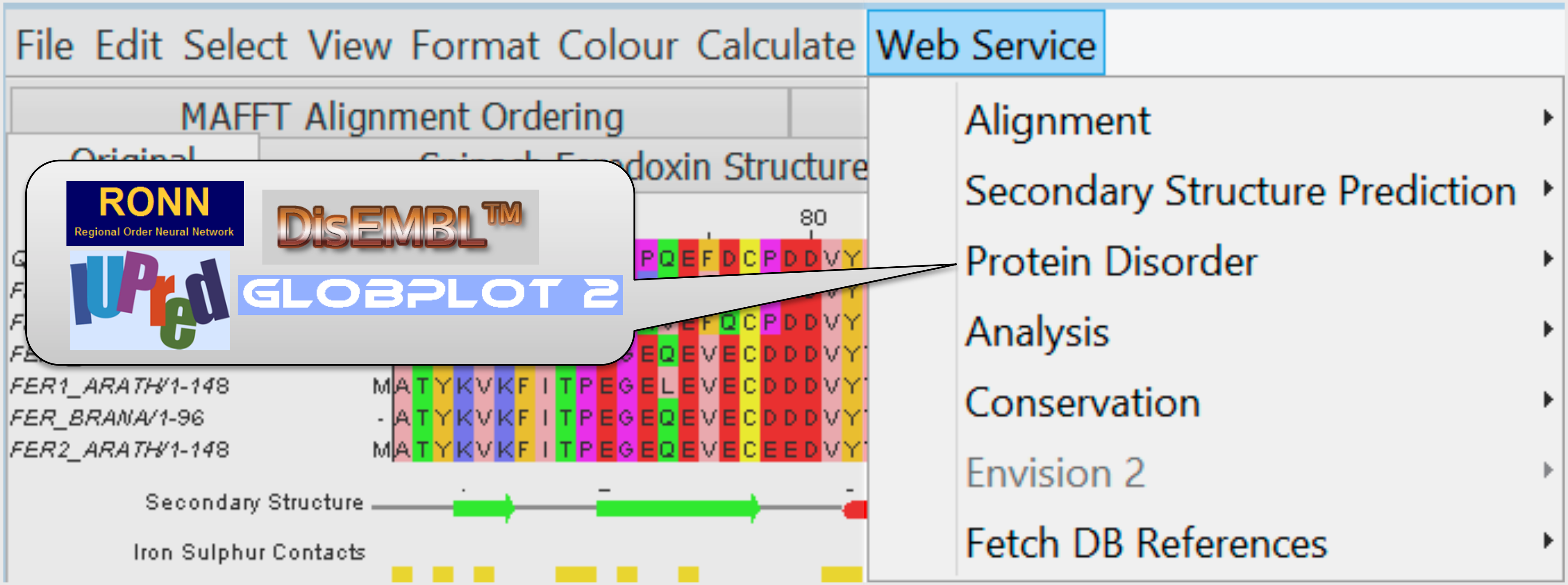
Just to recap…



Sequence features are overlaid on alignment to highlight key regions

Annotation can be used to shade alignment

Alignment annotation area shows graphs and symbols from calculations and manual curation


File
Edit
Select
View
Format
Colour
Calculate
Web Service
MAFFT Alignment Ordering
Original
Spinach Ferredoxin Structure
80
FER1_ARATH/1-148
FER_BRANA/1-96
FER2_ARATH/1-148
Secondary Structure
Iron Sulphur Contacts
RONN
Regional Order Neural Network
DisEMBL™
IUPred
GLOBPLOT 2
Alignment
Secondary Structure Prediction
Protein Disorder
Analysis
Conservation
Envision 2
Fetch DB References

Protein Disorder prediction
(Supplementary Exercise)

- Complementary problem to secondary structure prediction
  - Recognise structured & unstructured domains
  - Predict holes in density maps (REM450)
  - Detect flexible loops (‘HOTLOOPS’)

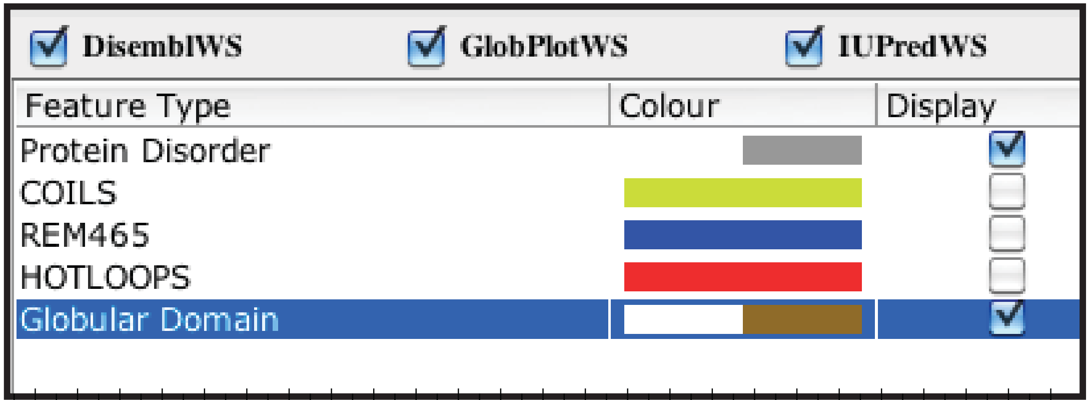
- Programs provided by JABAWS 2 employ
  - Machine learning methods (**DisEMBL**)
  - Similarity to disordered sequences (**RONN**)
  - Empirical amino acid statistics (**IUPred**, **GlobPlot**)

Disorder Predictions from JABAWS

JABAWS Analysis Service

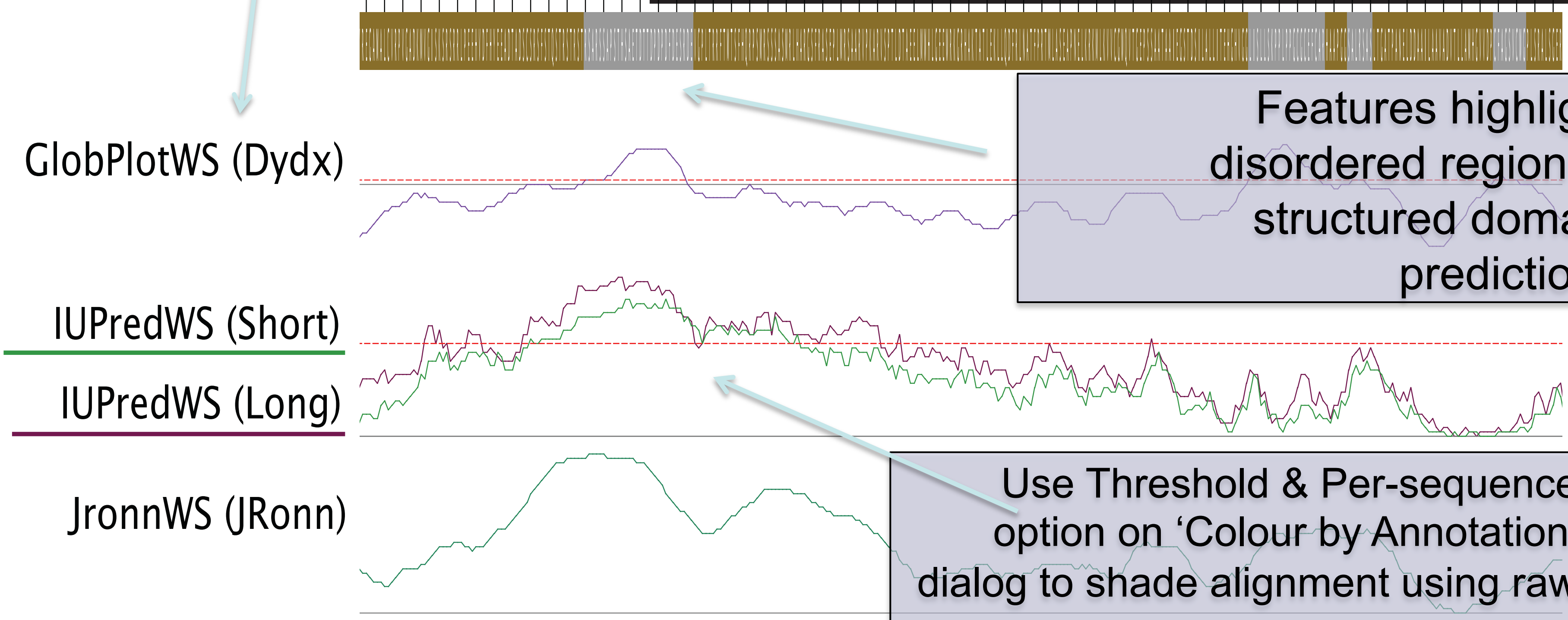
Disorder Predictions

Create **annotation** and **features**



☑ DisemblWS ☑ GlobPlotWS ☑ IUPredWS

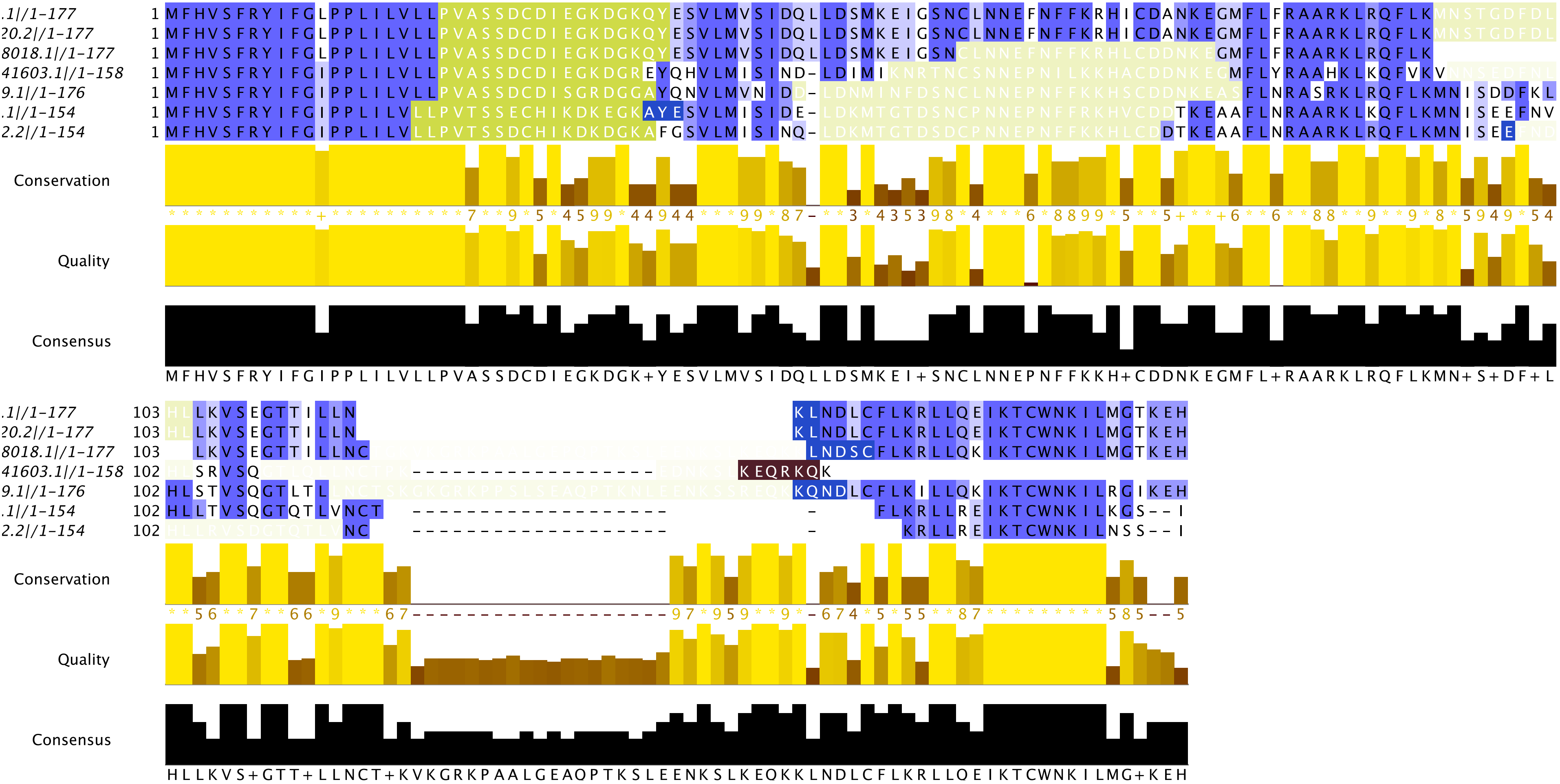
Feature Type	Colour	Display
Protein Disorder	▨	☑
COILS	🟨	☐
REM465	🟦	☐
HOTLOOPS	🟥	☐
Globular Domain	⬜🟫	☑



Features highlight disordered region or structured domain predictions

Use Threshold & Per-sequence option on ‘Colour by Annotation’ dialog to shade alignment using raw scores

Disorder in Interleukin 7



RNA 2nd-ary Structure

```
Pyrococcus_furiosus G G - U G G G G - G - - A G G C A C G G C U G A A G G G C C G U G C U A A - C C C C
Methanoculleus_ma   G G - A A A G G - G - A U C A A U G C C C C G G A G G G G G C G G C G A - C C U - -
F.acidarmanus.1     G G - U U U G C - A - U A U G C A A C U U C A A A G G A U G A U G C A U A - U A C A
Thermofilum_pende   G U - C C G U G - C - A C G G - C U C C U C G U A G G A G G A G G C C G A - G C A C
Thermosipho_melar   U - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
                        - U C C G U - U - - G A C G C U G A C C G C A G G U C A G U C C G A - G A C G
                        - A C A G G - A - U U C A A
                        - U U A U G - G - C C G G G
                        - U G G C G - G - - A G C C
M.labreanum.1       G G - A U G U G - A - U G G A U
T.celer.1           G G - U G G G C - G - - G G G U
secondary structure ( ( . ( ( ( ( ( . ( . ( ( ( ( (
StrucConsensus      ( ( - ( ( ( ( ( - ( - ( ( ( ( (
Consensus           G G - U G G C G - G C G A G C C
Sequence position 156  CG 45%; GC 18%; UA 15%; AU 9%; UG 7%; UU 0%; GU 0%; A
```

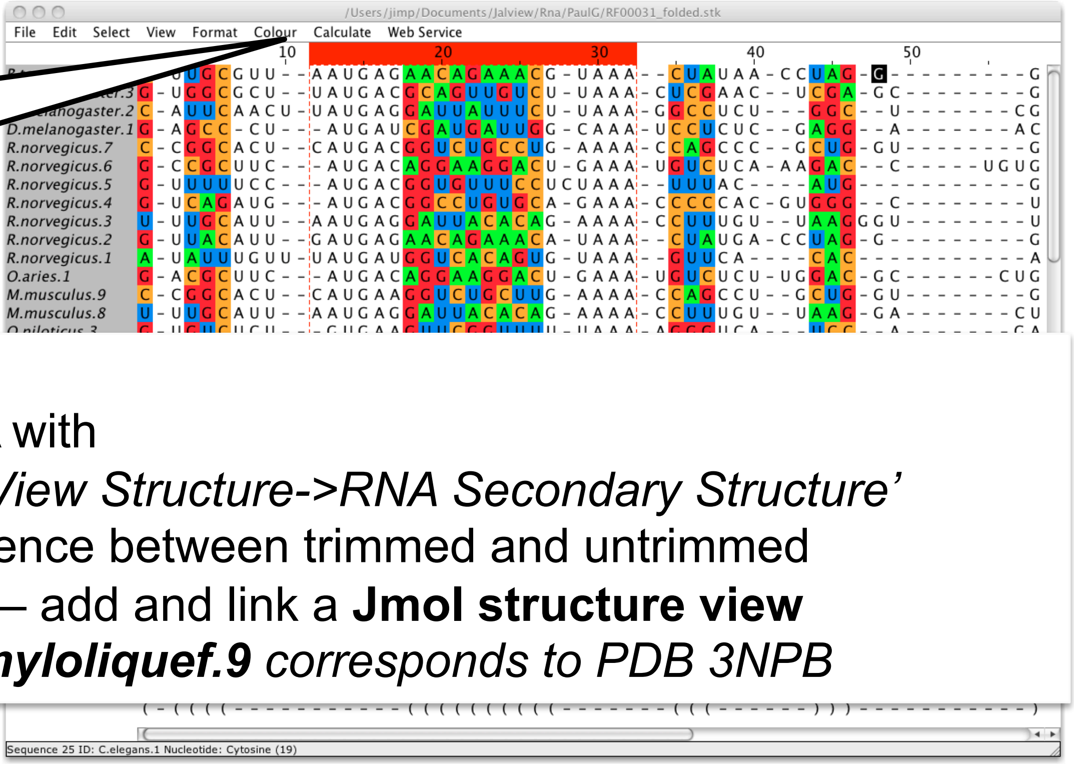
Structure Consensus Logo: Shows base pair distribution at each paired position in a given RNA secondary structure.

Try it now:

1. **Import RF00162 from *Rfam (Full)***
2. Select 'Colour by RNA Helices'

RALEE style colouring highlights distinct stems and helices

Selections and mouse positions shared between alignment view and VARNA



```
                   UUGCGUU--AAUGAGAACAGAAACG-UAAA--CUAUAA-CCUAG-G---------G
er.3             G-UGGCGCU--UAUGACGCAGUUGUCU-UAAA-CUCGAAC--UCGA-GC--------G
elanogaster.2    C-AUUCAACU-UAUGAGGAUUAUUUCU-UAAA-GGCCUCU---GGC--U-------CG
D.melanogaster.1 G-AGCC-CU---AUGAUCGAUGAUUGG-CAAA-UCCUCUC--GAGG--A-------AC
R.norvegicus.7   C-CGGCACU--CAUGACGGUCUGCCUG-AAAA-CCAGCCC--GCUG-GU--------G
R.norvegicus.6   G-CCGCUUC---AUGACAGGAAGGACU-GAAA-UGUCUCA-AAGAC--C-----UGUG
R.norvegicus.5   G-UUUUUCC---AUGACGGUGUUUCCUCUAAA--UUUAC----AUG-----------G
R.norvegicus.4   G-UCAGAUG---AUGACGGCCUGUGCA-GAAA-CCCCCAC-GUGGG--C--------U
R.norvegicus.3   U-UUGCAUU--AAUGAGGAUUACACAG-AAAA-CCUUUGU--UAAGGGU--------U
R.norvegicus.2   G-UUACAUU--GAUGAGAACAGAAACA-UAAA--CUAUGA-CCUAG-G---------G
R.norvegicus.1   A-UAUUUGUU-UAUGAUGGUCACAGUG-UAAA--GUUCA----CAC-----------A
O.aries.1        G-ACGCUUC---AUGACAGGAAGGACU-GAAA-UGUCUCU-UGGAC-GC------CUG
M.musculus.9     C-CGGCACU--CAUGAAGGUCUGCUUG-AAAA-CCAGCCU--GCUG-GU--------G
M.musculus.8     U-UUGCAUU--AAUGAGGAUUACACAG-AAAA-CCUUUGU--UAAG-GA-------CU
```

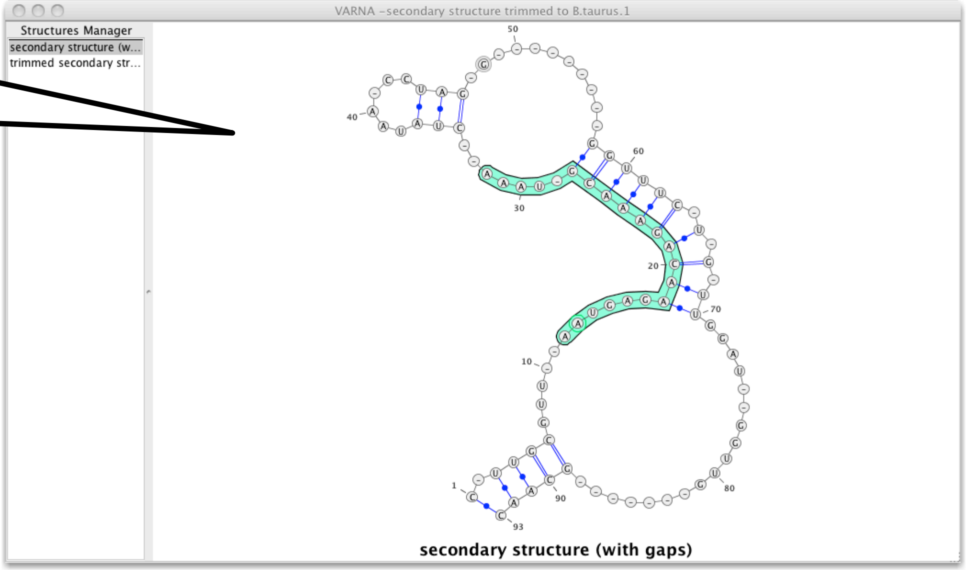
```
(-((((-----------((((((((((-------(((------)))-----------)
```

Sequence 25 ID: C.elegans.1 Nucleotide: Cytosine (19)

Try it now:

1. Open VARNA with
   'Structure->View Structure->RNA Secondary Structure'
2. Explore difference between trimmed and untrimmed
3. Bonus points – add and link a **Jmol structure view**
   Bacillus_amyloliquef.9 *corresponds to PDB 3NPB*

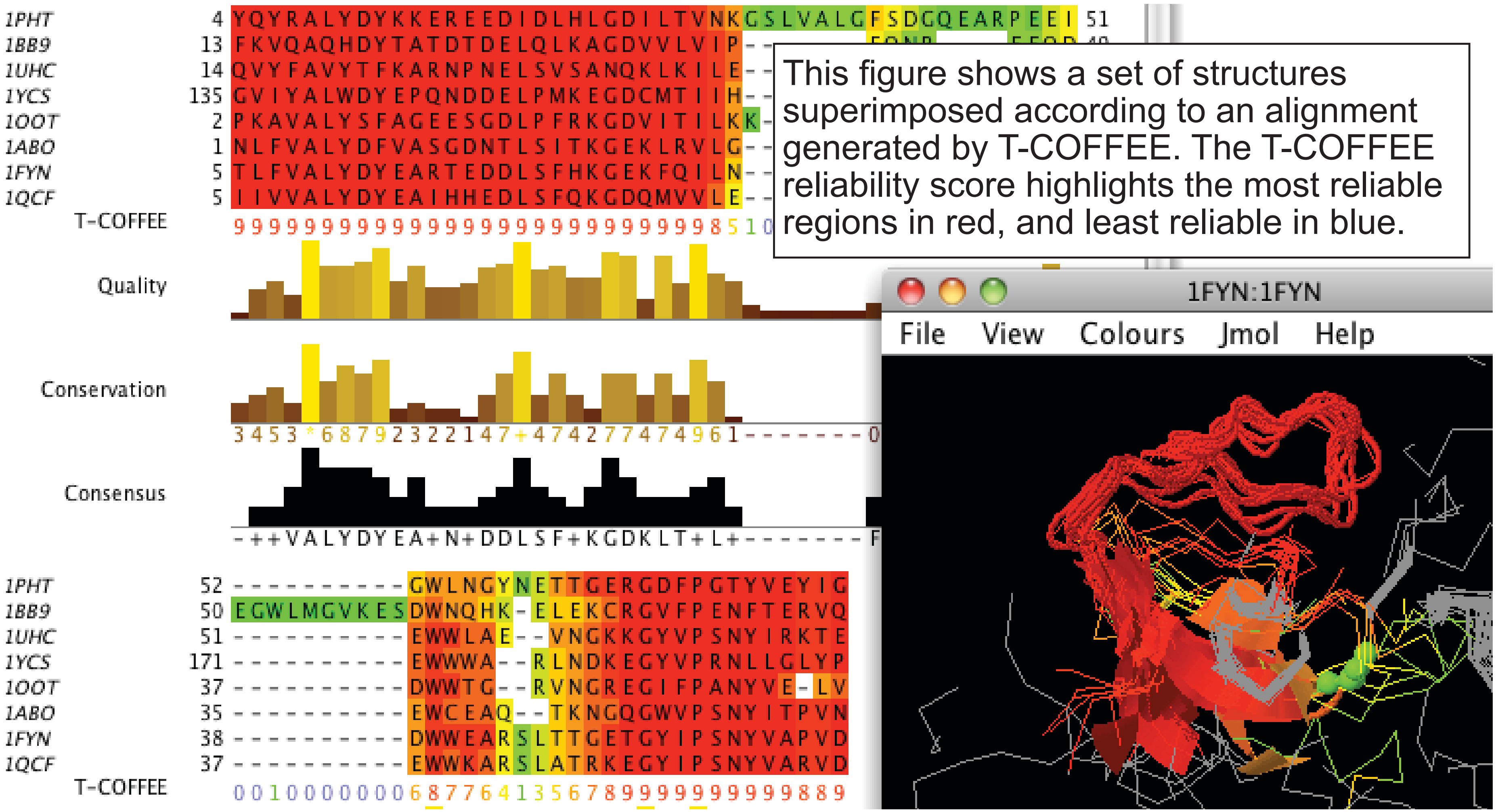
VARNA has a wide range of 2D RNA plots and supports interactive annotation



VARNA: Interactive drawing and editing of the RNA secondary structure
Kévin Darty, Alain Denise and Yann Ponty
Bioinformatics (2009) **25** 1974-1975

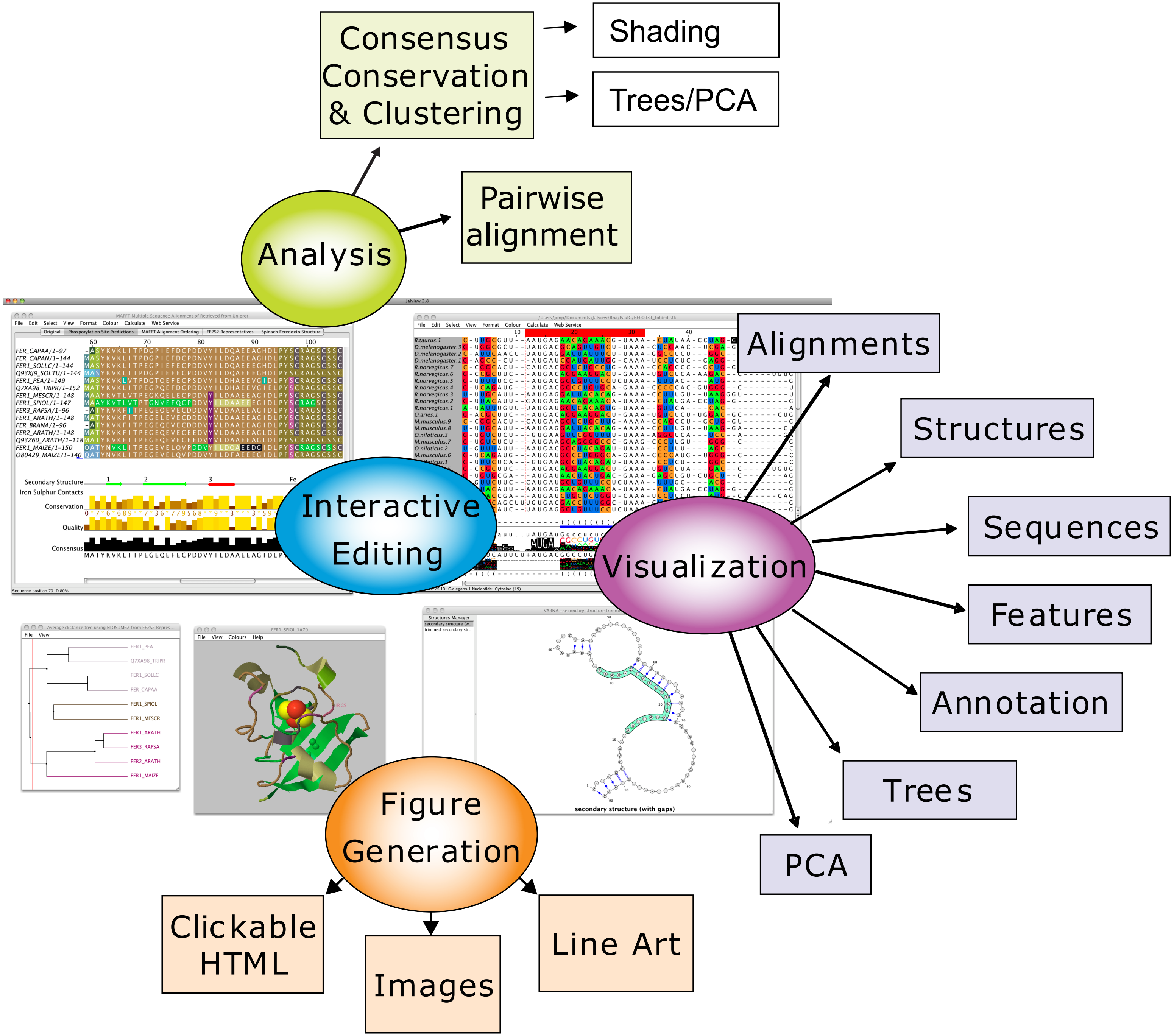
T-COFFEE alignment reliability scores

This figure shows a set of structures superimposed according to an alignment generated by T-COFFEE. The T-COFFEE reliability score highlights the most reliable regions in red, and least reliable in blue.



```
1PHT       4 YQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSDGQEARPEEI 51
1BB9      13 FKVQAQHDYTATDTDELQLKAGDVVLVIP--
1UHC      14 QVYFAVYTFKARNPNELSVSANQKLKILE--
1YCS     135 GVIYALWDYEPQNDDELPMKEGDCMTIIH--
1OOT       2 PKAVALYSFAGEESGDLPFRKGDVITILKK-
1ABO       1 NLFVALYDFVASGDNTLSITKGEKLRVLG--
1FYN       5 TLFVALYDYEARTEDDLSFHKGEKFQILN--
1QCF       5 IIVVALYDYEAIHHEDLSFQKGDQMVVLE--
     T-COFFEE 99999999999999999999999998510

     Quality

Conservation
             3453*687923221 47+4742774749 61--------0

   Consensus
             -++VALYDYEA+N+DDLSF+KGDKLT+L+--------F

1PHT      52 ----------GWLNGYNETTGERGDFPGTYVEYIG
1BB9      50 EGWLMGVKESDWNQHK-ELEKCRGVFPENFTERVQ
1UHC      51 ----------EWWLAE--VNGKKGYVPSNYIRKTE
1YCS     171 ----------EWWWA--RLNDKEGYVPRNLLGLYP
1OOT      37 ----------DWWTG--RVNGREGIFPANYVE-LV
1ABO      35 ----------EWCEAQ--TKNGQGWVPSNYITPVN
1FYN      38 ----------DWWEARSLTTGETGYIPSNYVAPVD
1QCF      37 ----------EWWKARSLATRKEGYIPSNYVARVD
     T-COFFEE 00100000006877641356789999999999889
```

9.00-9.15am.	Overview of the day
9.15am - 10.30am.	**Session 1. Introduction to Jalview** – starting the application, importing alignments, basic editing and creating figures.
10.30-11am.	*Coffee*
11am - 12.30pm.	**Session 2: Alignment & alignment analysis** – Creating sequence alignments, importing and calculating trees, tree based alignment analysis
12.30pm to 1.30pm.	*Lunch*
1.30pm – 3.00pm.	**Session 3: Annotating sequences & alignments** – Creating and viewing sequence annotation – Protein Secondary structure prediction
3.00pm – 3.30pm	*Coffee (Late! Sorry !)*
3.30pm – 4.30pm.	**Session 4: Working with molecular structures** – Viewing 3D Structures, superimpositions, mapping disorder and alignment quality – Viewing RNA Secondary Structure
4.30pm – 4.45pm.	Wrapup – what we didn’t cover today
~5pm	.. And then to Duke’s Corner!



Analysis

Consensus Conservation & Clustering

Shading

Trees/PCA

Pairwise alignment

Interactive Editing

Visualization

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Figure Generation

Clickable HTML

Images

Line Art

Jalview Flavours

Multi-windowed 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

Jpred 3
Incorporating Jnet

JABAWS 2.0

Alignment Analysis

AACon

Protein Disorder

IUPred
RONN
DisEMBL™
GLOBPLOT 2

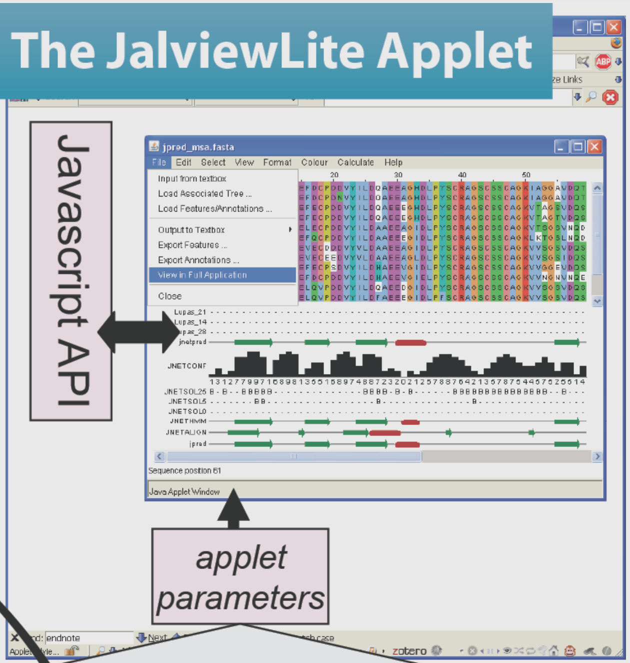
Multiple alignment

ClustalW
Mafft
ProbCons
TCoffee
Muscle

Biological Data Services

Pfam
Rfam
UniProt
PDBe
PROTEIN DATA BANK EUROPE
ENA
European Nucleotide Archive
Distributed Annotation System

The JalviewLite Applet



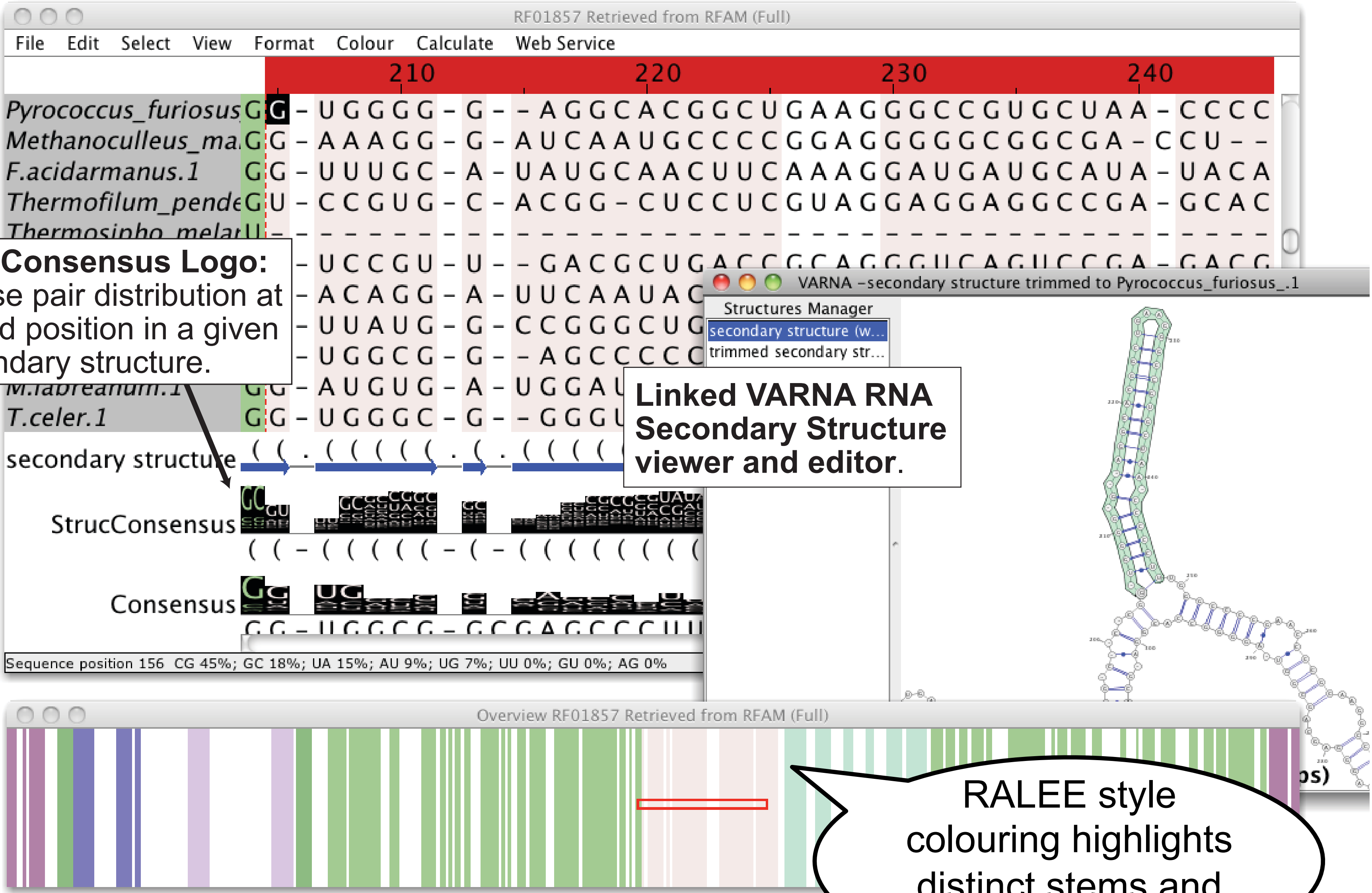
Javascript API

applet parameters

Alignment
And
Annotation

Bioinformatics
Web Server

Jalview 2.8 and RNA 2nd-ary



```
Pyrococcus_furiosus G G - U G G G G - G - - A G G C A C G G C U G A A G G G C C G U G C U A A - C C C C
Methanoculleus_ma   G G - A A A G G - G - A U C A A U G C C C C G G A G G G G C G G C G A - C C U - -
F.acidarmanus.1     G G - U U U G C - A - U A U G C A A C U U C A A A G A U G A U G C A U A - U A C A
Thermofilum_pende   G U - C C G U G - C - A C G G - C U C C U C G U A G G A G G A G G C C G A - G C A C
```

Sequence position 156 CG 45%; GC 18%; UA 15%; AU 9%; UG 7%; UU 0%; GU 0%; AG 0%

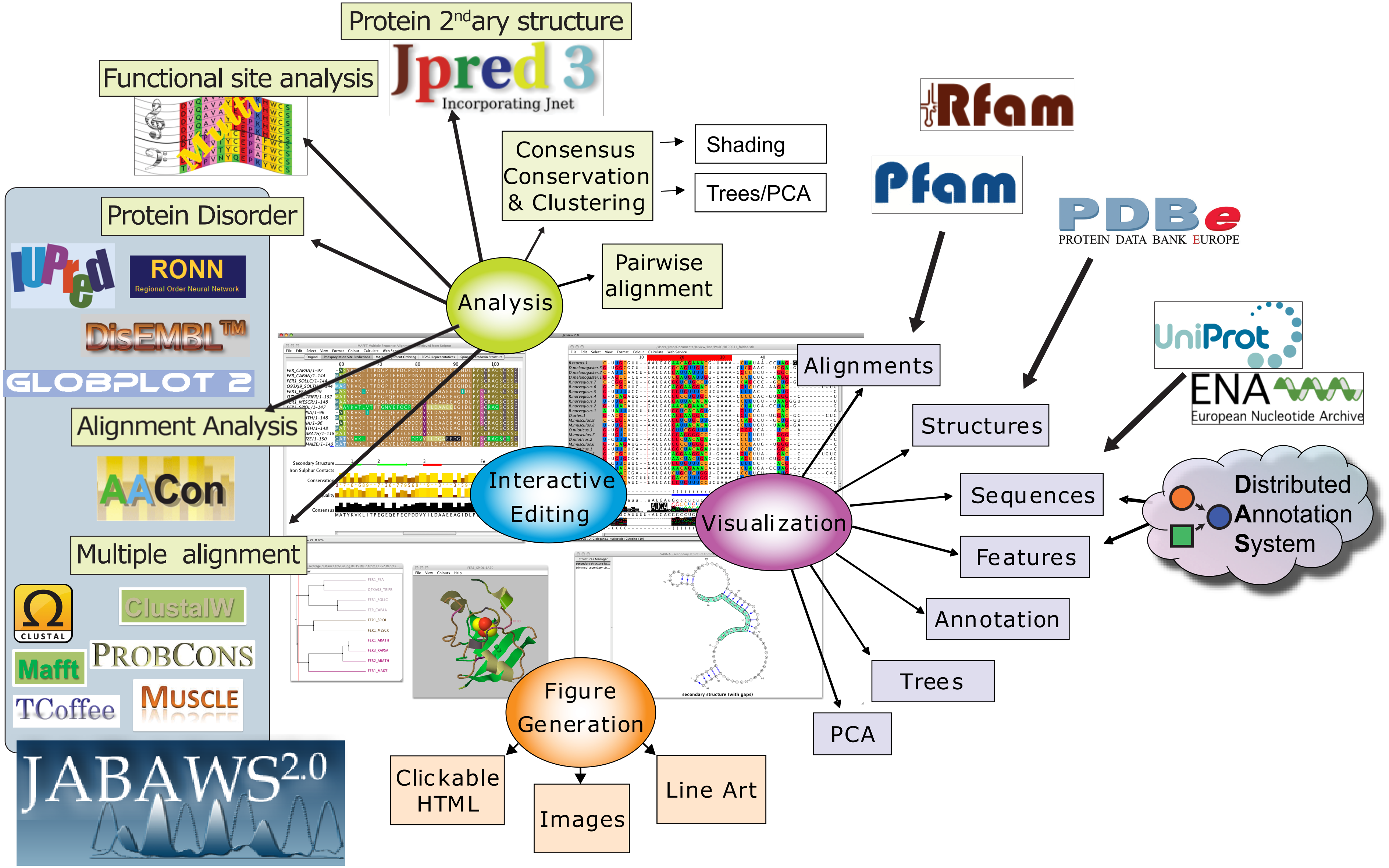
Structure Consensus Logo: Shows base pair distribution at each paired position in a given RNA secondary structure.

Linked VARNA RNA Secondary Structure viewer and editor.

RALEE style colouring highlights distinct stems and helices

DNA and Protein in Jalview

- Discussed in Section 2.10 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



Analysis

Protein 2ndary structure

Jpred 3 Incorporating Jnet

Functional site analysis

Consensus Conservation & Clustering

Shading

Trees/PCA

Pairwise alignment

Protein Disorder

IUPred

RONN Regional Order Neural Network

DisEMBL™

GLOBPLOT 2

Alignment Analysis

AACon

Multiple alignment

CLUSTAL

ClustalW

Mafft

PROBCONS

TCoffee

MUSCLE

JABAWS2.0

Interactive Editing

Visualization

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Rfam

Pfam

PDBe PROTEIN DATA BANK EUROPE

UniProt

ENA European Nucleotide Archive

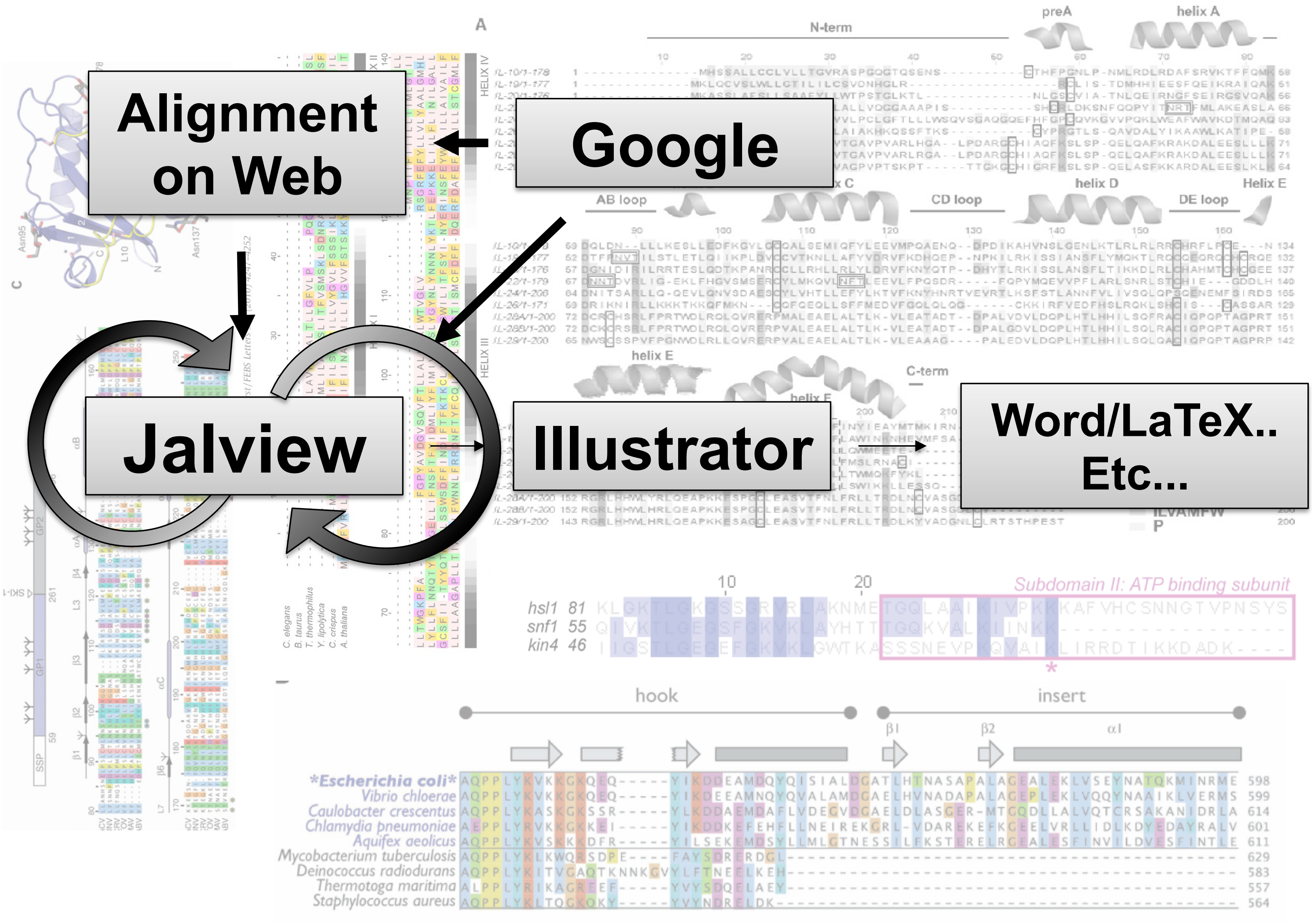
Distributed Annotation System

Figure Generation

Clickable HTML

Images

Line Art



Alignment on Web

Google

Jalview

Illustrator

Word/LaTeX.. Etc...

Coming up soon

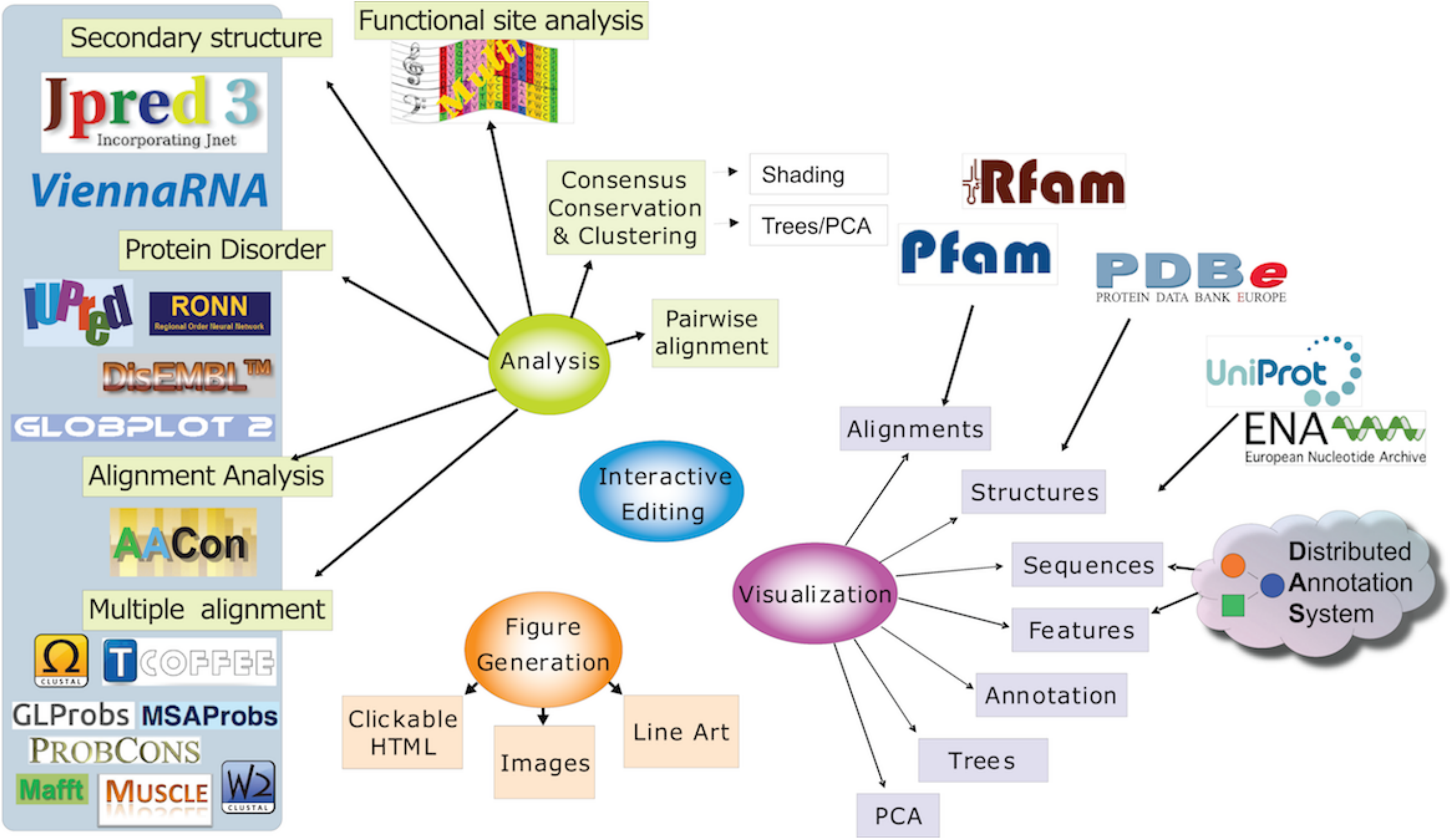
see http://issues.jalview.org

- Jalview 2.8.1
  - bugfixes, efficiency, ambiguous nucleotides
  - JABAWS 2.1
    - RNA Secondary structure prediction
    - Even more alignment methods
- Jalview 2.8.2
  - protein and RNA secondary structure from PDB files
  - JABAWS Jpred client

JABAWS 3.0a

Jpred 3

Incorporating Jnet



Secondary structure

Protein Disorder

Alignment Analysis

Multiple alignment

Functional site analysis

Consensus Conservation & Clustering

Shading

Trees/PCA

Pairwise alignment

Analysis

Interactive Editing

Figure Generation

Clickable HTML

Images

Line Art

Alignments

Structures

Sequences

Features

Annotation

Trees

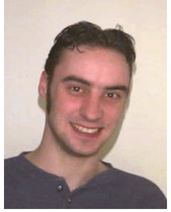
PCA

Visualization

Distributed Annotation System

The Jalview developers

Michele Clamp
Harvard & MIT.



James Cuff
Harvard & MIT

Steve Searle
Sanger, UK

Andrew Waterhouse
Basel, Switzerland.

RNA Features

Lauren Lui
UC Santa Cruz, USA.

Jan Engelhardt
Univ. Leipzig, Germany.

Yann Ponty (**VARNA**)
École Polytechnique,

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

Google SUMMER OF CODE 2011

T-COFFEE Scores

Paolo di Tomasso
Notredame Group, CRG, Spain.

Geoff Barton

David Martin (**Teaching**)

Sasha Sherstnev (**JABAWS**)

Peter Troshin (**JABAWS**)

Barry Strachan (**logo**)

Tom Walsh (**Apache**)

Ryan Maclaughlan (**CSS**)

Andrew Millar (**Drupal**)

All the Jalview users, and …

Jalview
www.jalview.org

supported by

wellcome trust

BBSRC
bioscience for the future