

# Jalview and JPred4

## Day 2: Annotation, Structures, cDNA, RNA, and Jalview clinic



supported by  
**wellcome**trust



# Day 2

9.30am.- 9.40am *Where did we get to ?*

## **Session 4 Annotating sequences and alignments**

- Database references, sequence features and DAS

*Coffee: 10.30am - 11am*

**Talk from Geoff: A quick intro to protein structure**

## **Session 5: Working with Structures**

- Jmol and Chimera: Viewing 3D structures and superposing them using the alignment
- Alignment annotation tracks from PDB data

## **Session 6: Disorder prediction**

- Protein disorder prediction

*Lunch: 12.30am-1.30pm*

## **Session 7: RNA, cDNA and Jalview**

- Viewing RNA structure & Predicting RNA srtructure in VARNA and RNAAliFold
- cDNA from the European Nucleotide Archive
- Aligning by protein and nucelotide

*Coffee: 3.30pm – 4pm*

## **Session 8: Jalview clinic**

# Alignment

MHC class II antigen

Q95IE6	uniprot	non_terminal_residue	88	88	0.0
Q8MGZ9	uniprot	non_terminal_residue	1	1	0.0
Q8MGZ9	uniprot	non_terminal_residue	89	89	0.0
Q8HWS7	uniprot	non_terminal_residue	1	1	0.0
Q8HWS7	uniprot	non_terminal_residue	89	89	0.0
Q30167	uniprot	signal_peptide	1	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	21	
Q30167	uniprot	polypeptide_region	30	124	0.0
Q30167	uniprot	polypeptide_region	125	227	0.0
Q30167	uniprot	glycosylated_residue	48	48	0.0
Q30167	uniprot	disulfide_crosslinked_residues	44	108	0

# Features

# Annotation

Q95IE6	uniprot	disulfide_crosslinked_residues	146	202	0
Q95IE5	uniprot	disulfide_crosslinked_residues	146	202	0
Q95IE6	uniprot	quality	7.89668	7.89668	7.89668
Q95IE5	uniprot	quality	7.89668	7.89668	7.89668

# Structure

ATOM	8	OE1	GLU	A	3	36.2	15.48		
ATOM	9	OE2	GLU	A	4	37.9	16.94		
ATOM	10	N	GLU	A	4	37.9	20.415		
ATOM	11	CA	GLU	A	4	38.96	21.562	67.120	11.132
ATOM	12	C	GLU	A	4	38.96	21.313	67.458	12.588
ATOM	13	O	GLU	A	4	38.96	22.169	67.189	13.445
ATOM	14	CB	GLU	A	4	38.96	22.323	68.269	10.476
ATOM	15	CG	GLU	A	4	38.96	23.588	67.860	9.745
ATOM	16	CD	GLU	A	4	38.96	24.007	68.917	8.771
ATOM	17	OE1	GLU	A	4	38.96	24.001	70.103	9.183
ATOM	18	OE2	GLU	A	4	38.96	24.293	68.587	7.581
ATOM	19	N	HIS	A	5	38.96	20.214	68.139	12.857
ATOM	20	CA	HIS	A	5	38.96	19.917	68.535	14.210
ATOM	21	C	HIS	A	5	38.96	18.443	68.716	14.290
ATOM	22	O	HIS	A	5	38.96	17.776	68.988	13.272
ATOM	23	CB	HIS	A	5	38.96	20.544	69.894	14.540
ATOM	24	CG	HIS	A	5	38.96	22.039	69.961	14.340
ATOM	25	ND1	HIS	A	5	38.96	22.946	69.677	15.344
ATOM	26	CD2	HIS	A	5	38.96	22.779	70.275	13.249
ATOM	27	CE1	HIS	A	5	38.96	24.176	69.800	14.882
ATOM	28	NE2	HIS	A	5	38.96	24.103	70.161	13.612

# Tree

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
0	100.0	85.71429	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 100%
0%	H 100%	F 100%	F 100%	N 100%

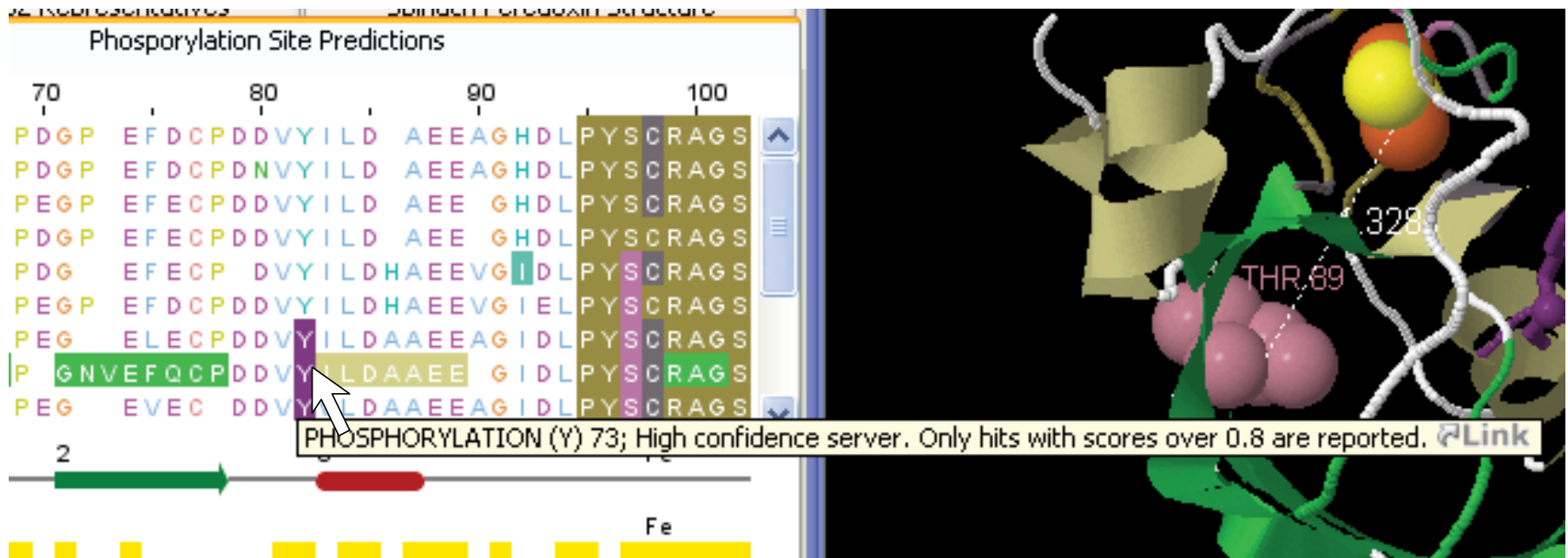
# Anatomy of Jalview: Figure 1.7

The image displays the Jalview 2.4 (pre) Desktop Window, which is divided into several functional areas:

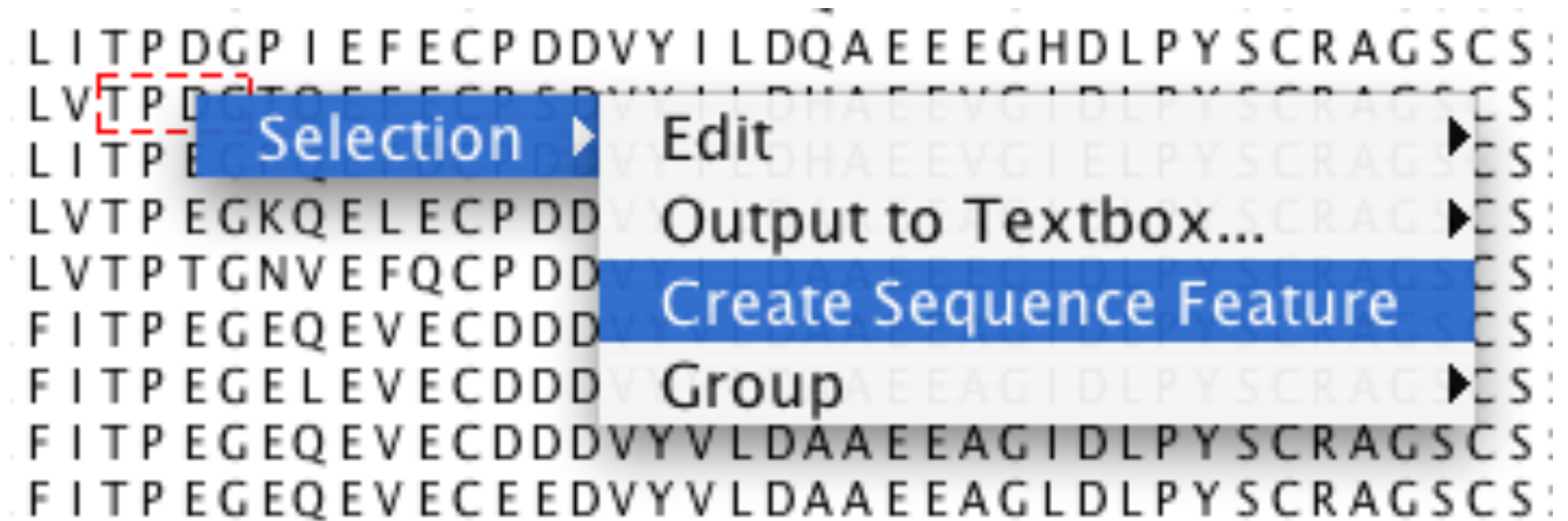
- Desktop Window:** The main application frame with a menu bar (File, Tools, Vamsas, Help, Window) and a title bar (Jalview 2.4 (pre)).
- Alignment Window:** The central workspace for sequence alignment, titled "MAFFT Multiple Sequence Alignment of Retrieved from Uniprot". It features a menu bar (File, Edit, Select, View, Format, Colour, Calculate, Web Service) and tabs for "Original", "Spinach Ferredoxin Structure", "FE2S2 Representatives", and "MAFFT Alignment Ordering".
  - Alignment View Tabs:** Located at the top left of the alignment window.
  - Sequence ID Panel:** A list of sequence identifiers on the left side, including FER1\_PEA/1-149, Q7XA98\_TRIPR/1-152, FER1\_MESCR/1-148, FER1\_SPIOI/1-147, FER1\_ARATH/1-148, FER2\_ARATH/1-148, Q93Z60\_ARATH/1-118, FER1\_MAIZE/1-150, and O80429\_MAIZE/1-140.
  - Sequence Alignment:** A multi-colored grid representing the alignment of amino acid residues across the sequences.
  - Alignment Ruler:** A horizontal scale at the top of the alignment grid, with markers at 70, 80, 90, and 100.
  - Alignment Annotation:** A series of horizontal bars below the alignment grid, including "Secondary Structure" (with green and red arrows), "Iron Sulphur Contacts" (with "Fe" labels), "Conservation" (yellow and brown bars), "Quality" (yellow and brown bars), and "Consensus" (black bars).
  - Annotation Label Panel:** A panel on the left side of the annotation bars.
  - Alignment Scrollbar:** A vertical scrollbar on the right side of the alignment grid.
  - Status bar:** Located at the bottom of the alignment window, displaying "Sequence 4 ID: Q93XJ9\_SOLTU Residue: GLY (79)".
  - Ruler Scrollbar:** A horizontal scrollbar at the bottom of the alignment window.
- Tree Window:** A separate window titled "Average distance tree using BLO..." showing a phylogenetic tree with branches and labels for sequences like FER1\_PEA, Q7XA98\_TRIPR, FER1\_SOLLC, FER1\_CAPAA, FER1\_SPIOI, FER1\_MESCR, FER1\_ARATH, FER3\_RAPSA, FER2\_ARATH, and FER1\_MAIZE.
- Structure Window:** A window titled "FER1\_SPIOI:1A70" showing a 3D ribbon structure of a protein. It includes a menu bar (File, View, Colours, Help) and a label "HR 89" with a distance of "1.328 nm". The Jmol logo is visible in the bottom right corner.



# Sequence Features



# Features from selected region



# Feature creation dialog box

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

Create New Sequence Feature(s)

Sequence Feature Name: feature\_1

Feature Group: jalview

Feature Colour:

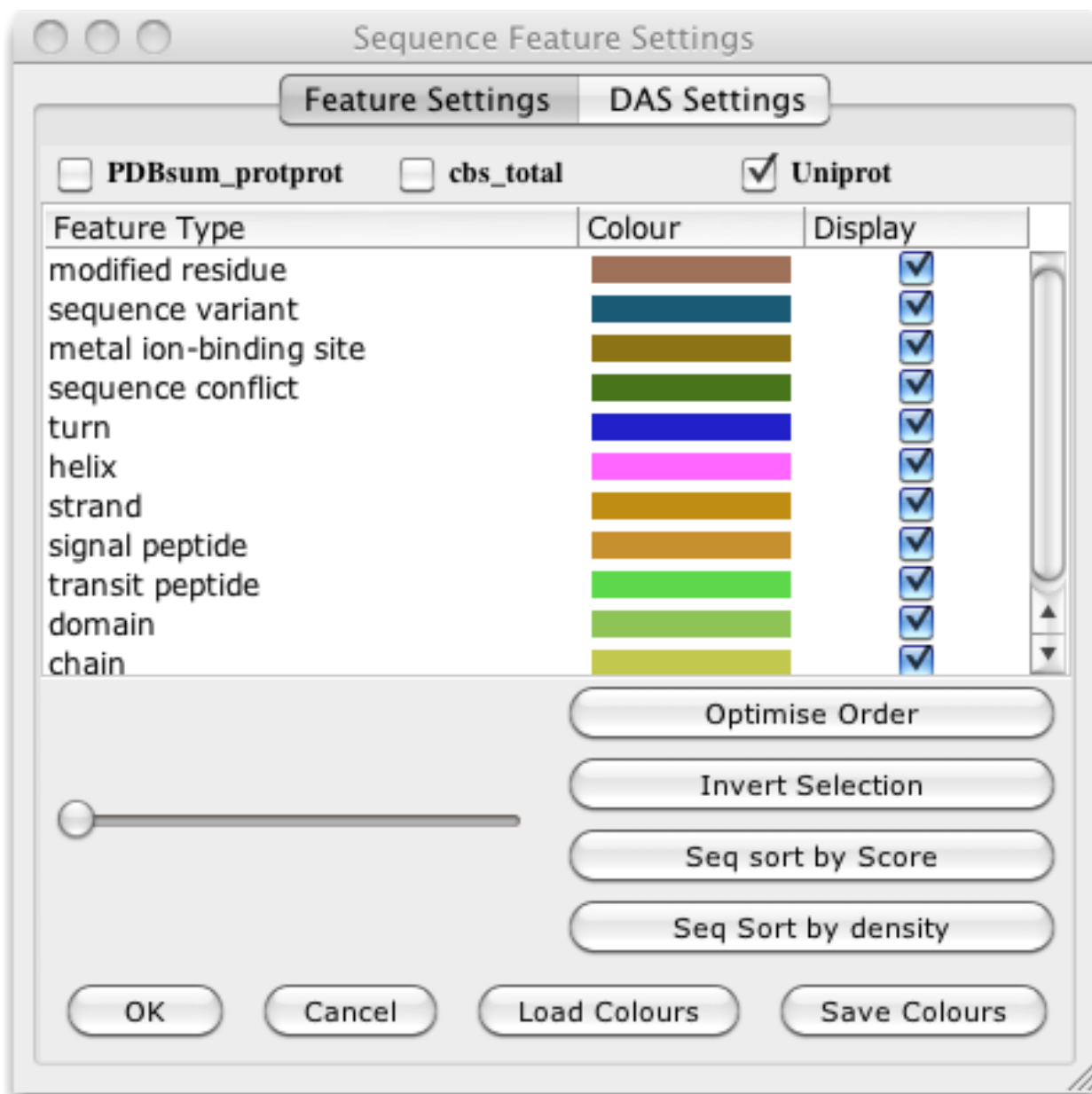
Description:

Cancel OK

'K L I T P D G P I E F E C P D D V Y I L D  
'K L V T P D G T Q E F E C P S D V Y I L D  
'K L I T P E C R O E E D C R D D V Y I L D  
'T L V T P E C R O E E D C R D D V Y I L D  
'T L V T P T G N V E F Q C P D D V Y I L D

T P D G

feature\_1 61 64;





# Sequence Features

## Section 2.8.1-3 & Ex 28

- Annotate the whole or part of a sequence
- Tasks
  - Visualise, create, modify, import and export features.

# Sequence Features

## Section 2.8.1-3 & Ex 28

- Annotate the whole or part of a sequence
- Tasks
  - Visualise, create, modify, import and export features.
- Questions
  - What are the different types of file formats available for import and export ?
  - How can you create features based on sequence motifs ?

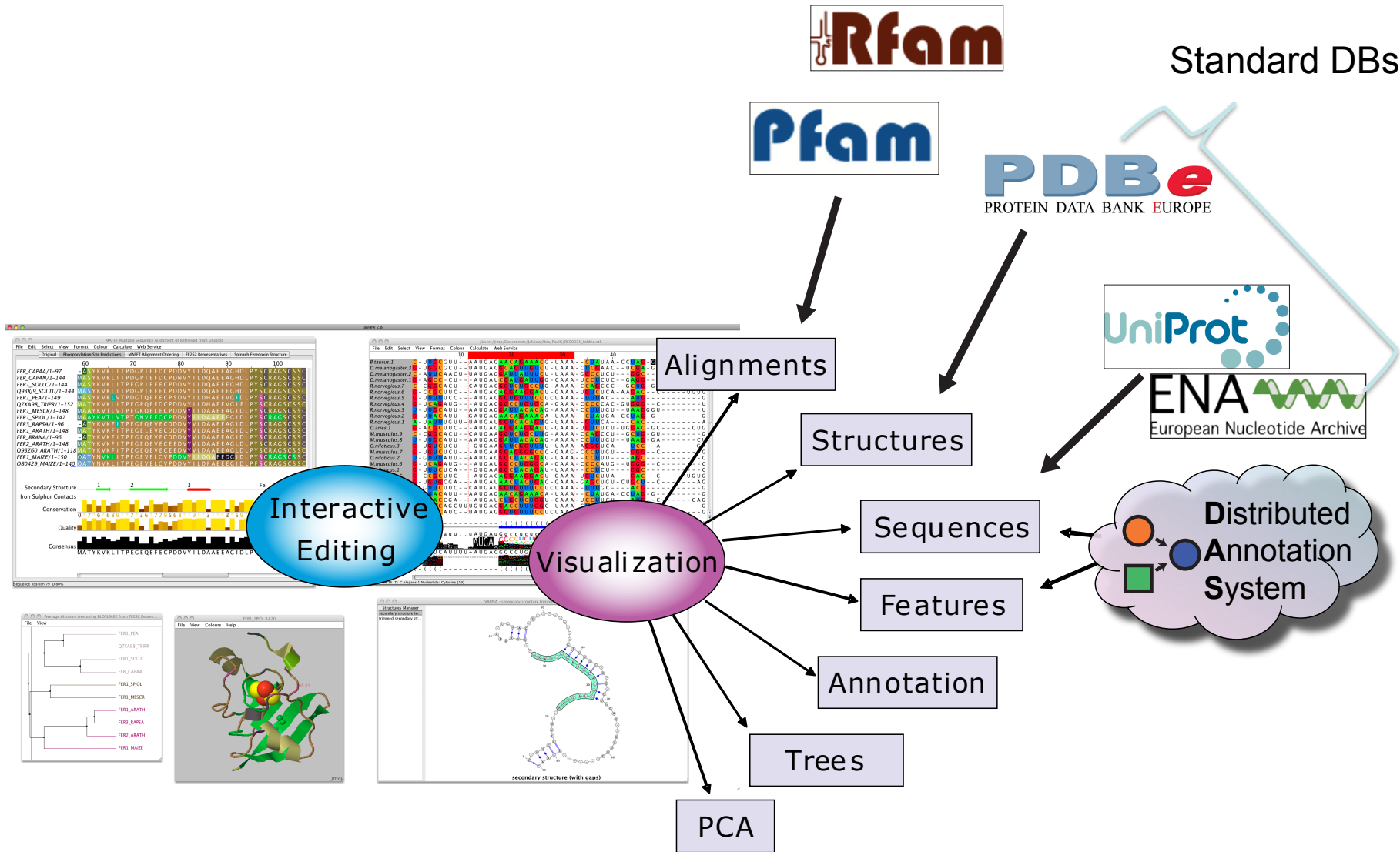
# Getting and working with sequence features and annotation

- 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

# Sources of sequence feature data

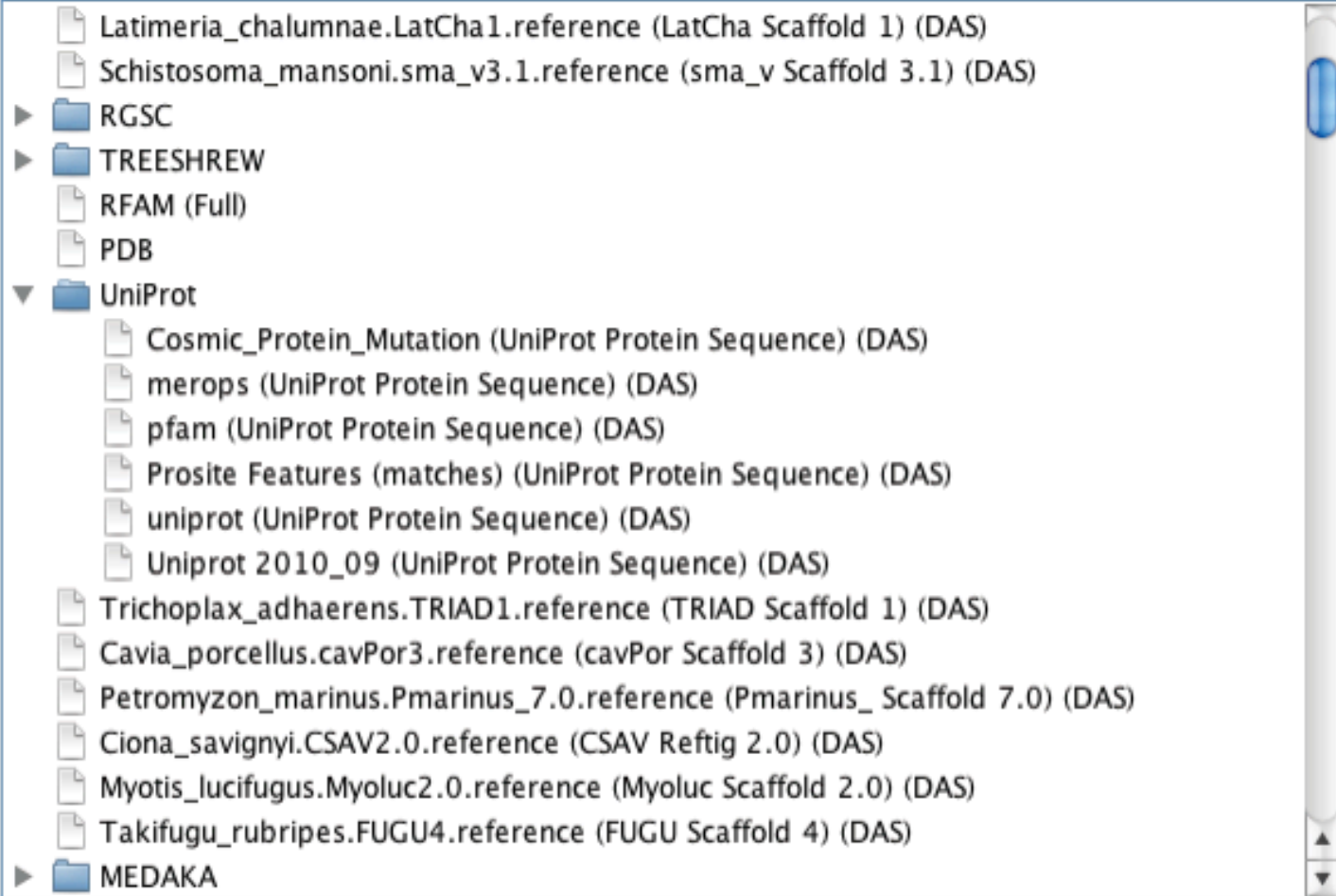
- Jalview sequence annotation files
- DAS sources
- GFF files
- Certain 'rich' alignment formats
  - Stockholm
  - AMSA

# Retrieval from External Databases





# 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

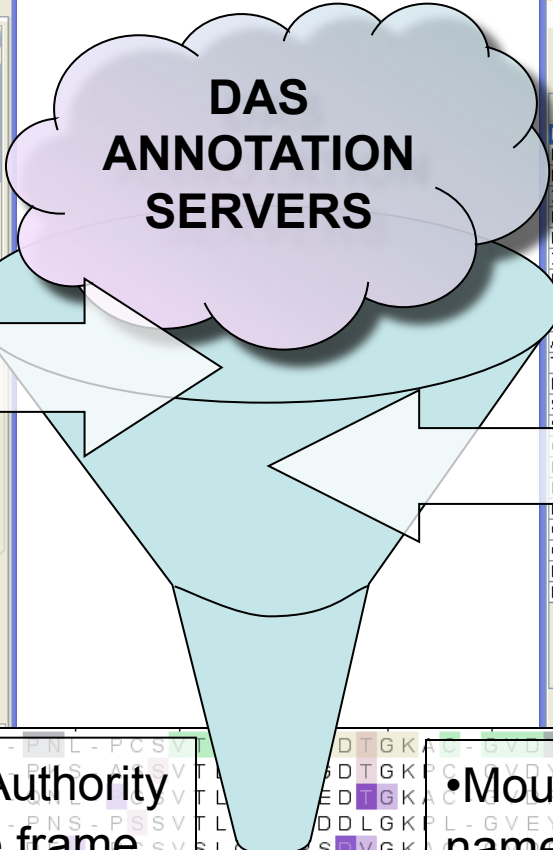
**Sequence Feature Settings**

Feature Settings | DAS Settings

Nickname: uniprot  
 URL: <http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/>  
 Admin Email: [uniprot-das@ebi.ac.uk](mailto: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/> [Reset]

[Refresh Available Sources] [Add Local Source]



**Sequence Feature Settings**

Feature Settings | DAS Settings

uniprot  Pfam Other Features  
 PDBsum\_protprot  cbs\_total

Feature Type	Colour	Display
DISULFID		<input checked="" type="checkbox"/>
Protein-protein contact		<input checked="" type="checkbox"/>
MOD_RES		<input checked="" type="checkbox"/>
ISOFORM		<input checked="" type="checkbox"/>
PHOSPHORYLATION (S)		<input checked="" type="checkbox"/>
PHOSPHORYLATION (Y)		<input checked="" type="checkbox"/>
INIT_MET		<input checked="" type="checkbox"/>
PHOSPHORYLATION (T)		<input checked="" type="checkbox"/>
ES-SIGNAL		<input checked="" type="checkbox"/>
CONFLICT		<input checked="" type="checkbox"/>
ACETYLATION		<input checked="" type="checkbox"/>
TURN		<input checked="" type="checkbox"/>
HELIX		<input checked="" type="checkbox"/>
STRAND		<input checked="" type="checkbox"/>
SIGNAL		<input checked="" type="checkbox"/>
COMPLAS		<input checked="" type="checkbox"/>
PROSITE		<input checked="" type="checkbox"/>
PRINTS		<input checked="" type="checkbox"/>
PROPERT		<input checked="" type="checkbox"/>
OUTSIDE		<input checked="" type="checkbox"/>
CHAIN		<input checked="" type="checkbox"/>
Pfam		<input checked="" type="checkbox"/>
ProDom		<input checked="" type="checkbox"/>

[Invert Selection] [Load Colours] [Save Colours]

• 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

UniProt/Swiss-Prot|P32122 P - - A S S - P S S V T L Q P G D D D Q G K P L - G V E Y S V K T W V G D H A - -  
 UniProt/Swiss-Prot|P08168 P - - D Y L - P C S V M L Q P A P Q D V G K S C - G V D F E I A A T H S T D -  
 UniProt/Swiss-Prot|P53179 P R G R G M - L S S I - - - - - K F E R G - S I T Y F L S C T L E S L N N I  
 UniProt/Swiss-Prot|Q09889 P - - P D I - P D S I - - - - - E G I P G C H I I Y T L T A S L E R A T Q -  
 UniProt/Swiss-Prot|P30647 P - - L N C - P S S Y - - - - - E S Q F G - S I R Y Q M K V E L R A S T D -  
 UniProt/Swiss-Prot|O45782 P - - K S L - P S S F - - - - - E G E F G - H I R Y T C K A I C E R P W D -  
 UniProt/Swiss-Prot|O76685 P - - I N V - P P S F - - - - - E G K Y G - Y L R Y S V T A E V D R P W R -  
 UniProt/Swiss-Prot|O17812 P - - E N I - P Q S F - - - - - E G P F G - F I R F Y I K V H M D R P H A -

# 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

# Cut & Paste input - FASTA

File Edit Select View Format

Original View 1

UniProt/Swiss-Prot|P1787  
UniProt/Swiss-Prot|P5148  
UniProt/Swiss-Prot|P5146  
UniProt/Swiss-Prot|P5148  
UniProt/Swiss-Prot|P5147  
UniProt/Swiss-Prot|P5148  
UniProt/Swiss-Prot|P3212  
UniProt/Swiss-Prot|P0816  
UniProt/Swiss-Prot|P5317  
UniProt/Swiss-Prot|Q0988  
UniProt/Swiss-Prot|P3064  
UniProt/Swiss-Prot|O4578  
UniProt/Swiss-Prot|O7668  
UniProt/Swiss-Prot|O1781

Conservation

Quality

Consensus

Sequence 1 ID: UniProt/Swiss

### Sequence Feature Settings

Feature Settings DAS Settings

Nickname	Us...	
Canis_familiaris....	<input type="checkbox"/>	
Canis_familiaris....	<input type="checkbox"/>	
Gallus_gallus.WA...	<input type="checkbox"/>	
Gallus_gallus.WA...	<input type="checkbox"/>	
Ornithorhynchus_...	<input type="checkbox"/>	
Ornithorhynchus_...	<input type="checkbox"/>	
Saccharomyces_c...	<input type="checkbox"/>	
Saccharomyces_c...	<input type="checkbox"/>	
Phobius	<input type="checkbox"/>	
uniprot	<input checked="" type="checkbox"/>	
cbs_total	<input type="checkbox"/>	
PDBsum_protpro	<input type="checkbox"/>	
Pfam Other Feat	<input type="checkbox"/>	

Authority:

- Any
- UniProt
- PDBresnum
- NCBI
- Ensembl
- SGD
- MOZ2a
- RGSC
- MGI

Clone	Gene_ID	eFamily
Chromosome	Predicted	
Contig	ZFMODELS	Computationally inferred
Scaffold	eProtein	
Volume Map	Manually curated	

Nickname: uniprot  
URL: <http://www.ebi.ac.uk/das-srv/uniprot/das/aristotle/>  
Admin Email: [uniprot-das@ebi.ac.uk](mailto: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

### Find Uniprot Accession Ids

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

20

K A F C A E N L - -  
K T F V A V T D G -  
R A F C A K S I - -  
R A Y V A D S E - -  
K A F S T N N L - -  
K L F I A E T E - -  
K T W V G D H A - -  
K A F A T H S T D -  
S C T L E S L N N I  
T A S L E R A T Q -  
K V E L R A S T D -  
K A I C E R P W D -  
T A E V D R P W R -  
K V H M D R P H A -

4 5 1 4 3 2 3 3 0 -

# New in Jalview 2.8.2

## Select column by feature

The screenshot displays the Jalview 2.8.2 interface. On the left, a sequence alignment is shown with columns color-coded by features. A green arrow points to column 3. Below the alignment is a bar chart representing a feature score. The right side shows the 'Sequence Feature Settings' dialog box with the 'Feature Settings' tab selected. A context menu is open over the 'PHOSPHORYLATION (Y)' row, showing options to sort or select columns based on the feature.

Sequence Feature Settings

Feature Settings | DAS Settings

uniprot  netphos

Ia70

Feature Type	Colour	Display
RESNUM		<input type="checkbox"/>
MOD_RES		<input checked="" type="checkbox"/>
VARIANT		<input checked="" type="checkbox"/>
METAL		<input checked="" type="checkbox"/>
PHOSPHORYLATION (S)		<input checked="" type="checkbox"/>
PHOSPHORYLATION (T)		<input checked="" type="checkbox"/>
PHOSPHORYLATION (Y)		<input checked="" type="checkbox"/>
CONFLICT		<input type="checkbox"/>
TURN		<input type="checkbox"/>
HELIX		<input type="checkbox"/>
STRAND		<input type="checkbox"/>

Sort by Score

Sort by Density

Graduated Colour

Select columns containing

Select columns that do not contain

Invert Selection



# Flanking alignments and 'Trim retrieved sequences'

- New feature for Proteomic hit analysis

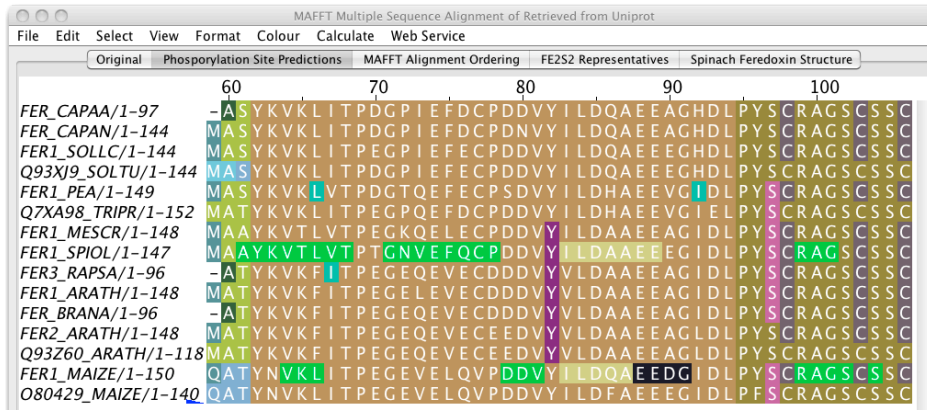
First create a new FASTA file subsequences

```
>FER1_PEA
```

```
ASYKVKLVTPDGTQEFECPSDVYILDHAEVGGIDLPYSCRAGSCSSCAGKVVGGGEVDQS  
DGSFLDD
```

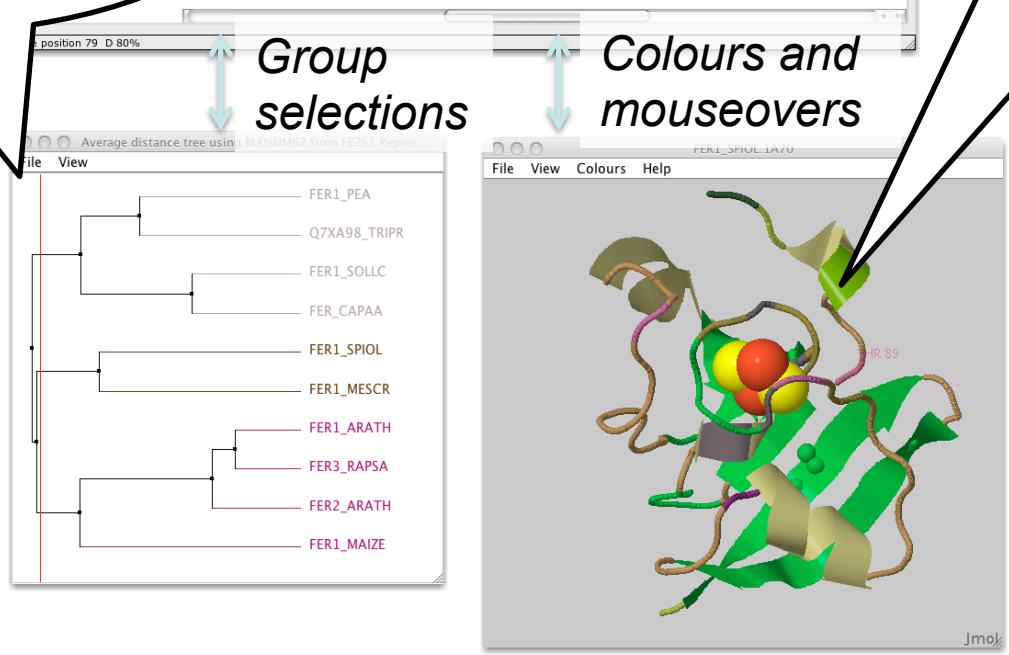
1. Make sure 'Trim retrieved sequences' is **unticked**
2. Hit the 'Retrieve from standard databases' option
3. Once finished, apply the 'Calculations->Show flanking regions' option

# **3D STRUCTURE IN JALVIEW**



Linked tree viewer allows subgroups to be identified in alignment

Linked Jmol viewer shows one or more structures coloured by alignment



# Biomolecular Structure Visualization in the Jalview Desktop



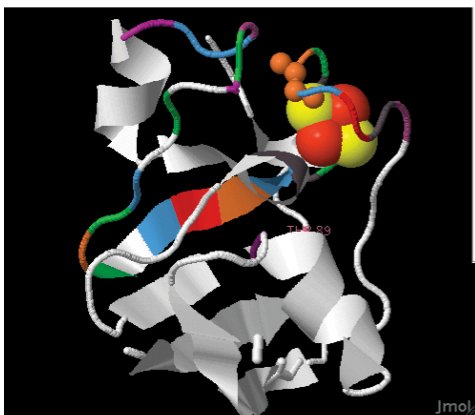
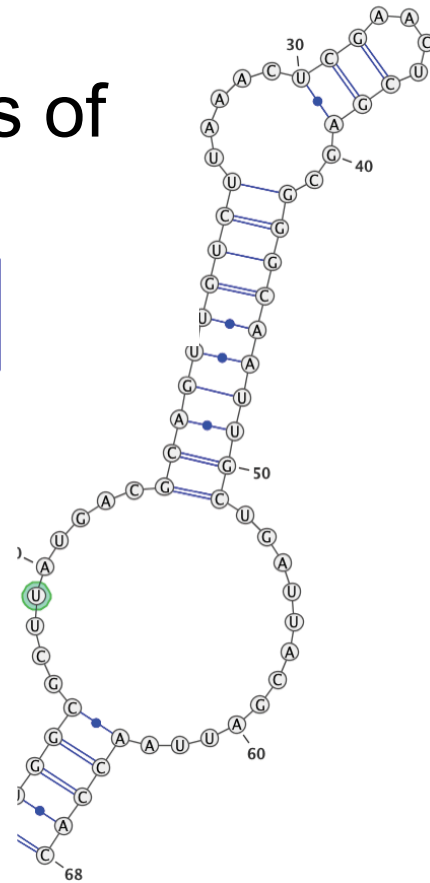
Jmol

[jmol.sourceforge.net](http://jmol.sourceforge.net)

**VARNA**

Visual Analysis of  
RNA

[varna.lri.fr](http://varna.lri.fr)



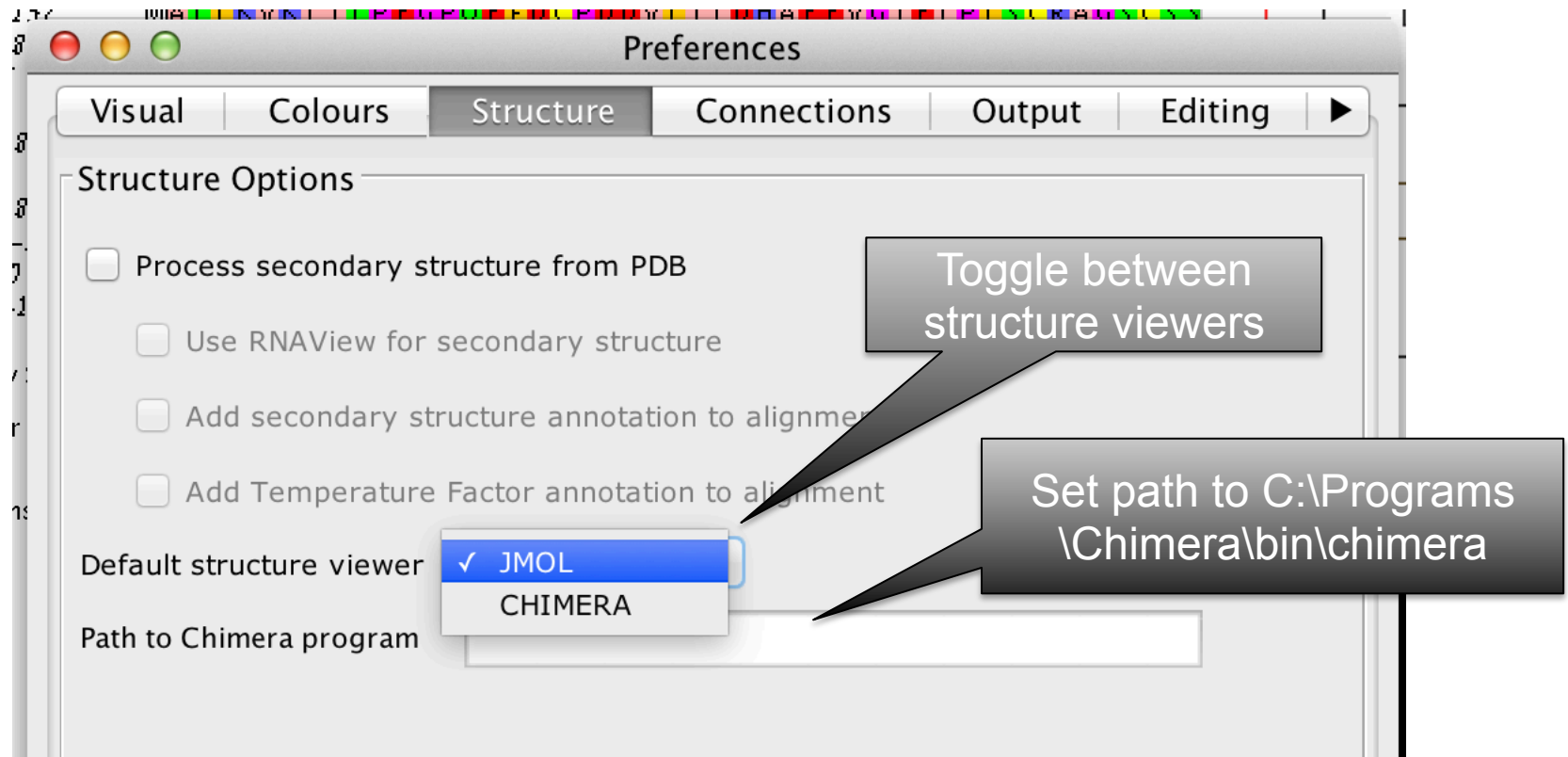
**UCSF CHIMERA**

an Extensible Molecular Modeling System

[www.rbvi.ucsf.edu](http://www.rbvi.ucsf.edu)

# Configure Jalview and Chimera

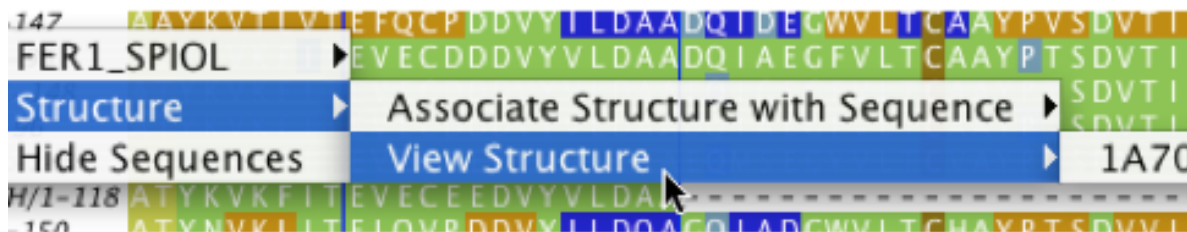
- Jalview 'Structure' preferences tab



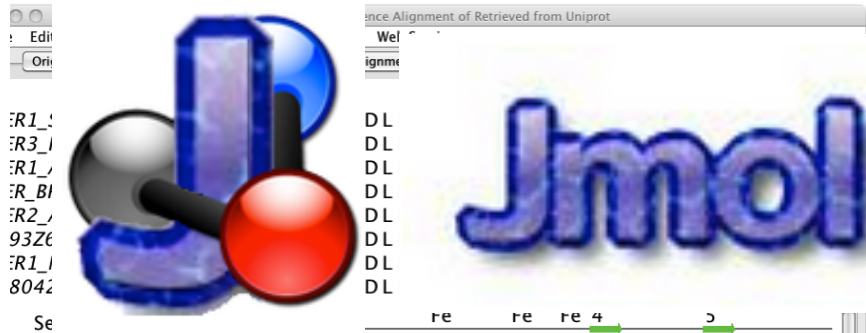


# Associating structures with sequences

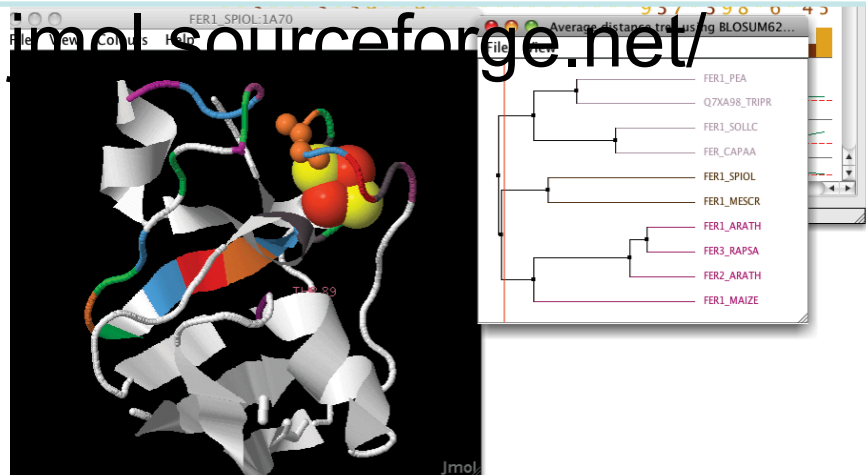
- Local PDB file
  - Attach PDB file to sequence manually
  - drag and drop to match files to sequences by ID
- Structures in the PDB database
  - Provide PDB id (and chain) for sequence
  - Discover references *via* sequence database



# Secondary structure from 3D data



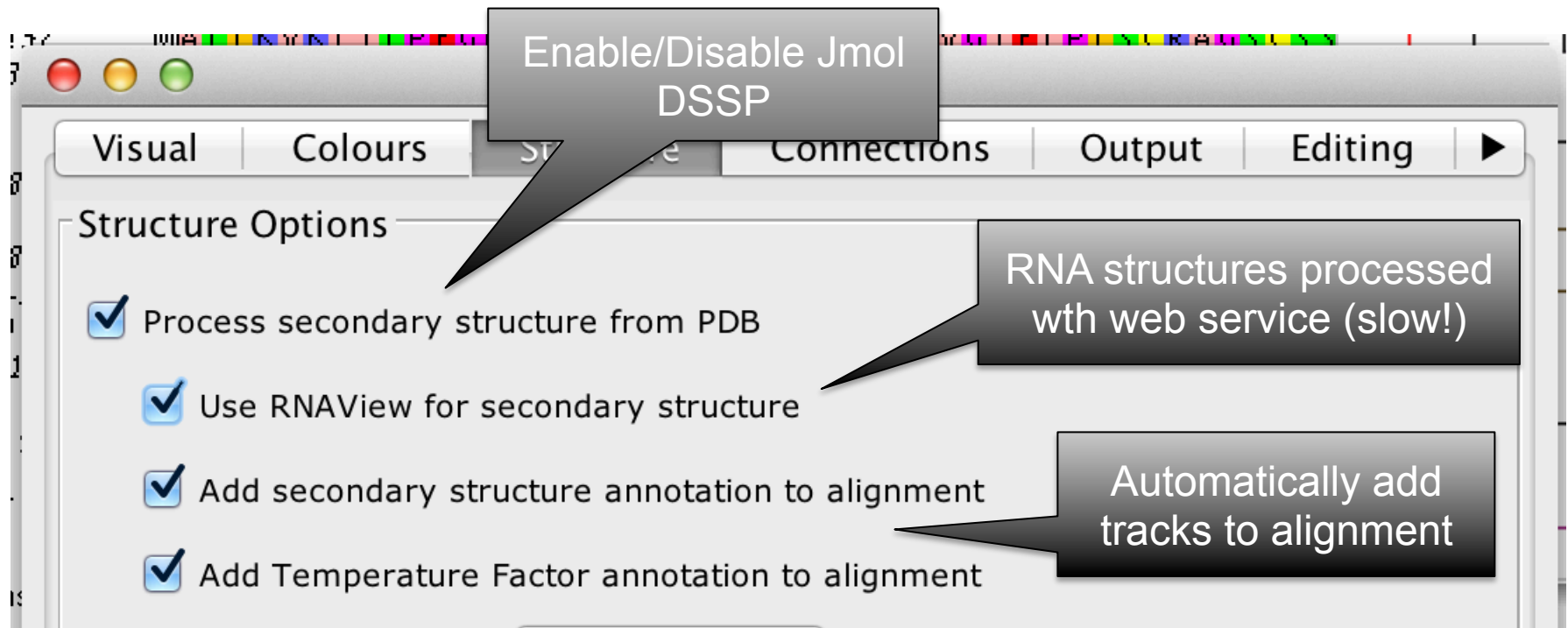
<http://>



- Jmol includes a Java port of **DSSP**
  - Courtesy of the Vriend Lab
- Jalview 2.8.2 now employs Jmol to parse PDB data
  - Display residue level structure data on sequences

# Secondary structure annotation tracks from 3D Structure

- Jalview 'Structure' preferences tab



# Protein Structures in Jalview

## Sec 2.1. Exercise 14

- Task
  - Discover PDB structures for ferredoxin sequence(s)
    - 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

# Hiding and showing sequence associated annotation

- Not yet in manual
- ‘Show reference annotation’
  - Option in Sequence ID popup menu
  - Individual sequence & Selection
- Annotation ID popup menu
  - Show/hide by type
- Annotations dropdown menu
  - Order annotations by alignment order

```
LIVNGGIIAPSFDDP.MDEKAREILQKLFPEHEVVMA.PGRELLL...L...
YIANGGIIAPQFGDPIRDKEAIRVLSDTFPHHSVVG IENAREIVL.A...
LILNNRVFVPVNGPASVDNDALNVYKTAMPGYEIIIGVKGASGTPWL.TDAL.R.HEVA
LITNKGVI VPPQYGDE.NDALALKQVQEMFPDREIVGV.NTVEVVY...
YVCNGGVVLC AFGDP.NDELAAGIFRRLFPERTVTLV.DARTIFA...G...
YTANGGIVFPLFNDP.MDEKAQEILQKLYPDRKIVGV.PAREILL...I...
L...V...
```

Q8KCB6\_CHLTE:1XKN Method: X-ray Chain:A...

File View Colours Jmol Help

Show Chain

- All
- ✓ 1XKN:A
- ✓ 2CMU:A
- ✓ 1VKP:A
- ✓ 1VKP:B
- ✓ 2EWO:A
- 2EWO:B
- 2EWO:C
- 2EWO:D
- 2EWO:E
- 2EWO:F
- 2EWO:G
- 2EWO:H
- 2EWO:I
- 2EWO:J
- 2EWO:K
- 2EWO:L

CYS 366:A

Jmol

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

If 'View all N structures' doesn't align structures:

- **Use Jmol/Structure Viewer->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

```
10 20 30 40 50
M17_HUMAN ---eNAFKRRR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GSGRSKQ-AQC-E-RR
D1_MOUSE ---qRMFKRVG-EGDAALVKEDEGVGSTERL---QLPSD-VASGLYC-KCE-RRR
M01_PHRIN ---sS-KSRVVV-EGEECGNAPD-EMKPEHLD---MKRFG--GPGLRKQ-TEK-RRK
35W1_LAGLA ---F-VERSRD-EGVEGCGQTQEDGCHER-ELR-PPR---PGLRRD-WKV-QRR
2MME_DROS ---qSKKKRRR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GSGRSKQ-AQC-E-RR
KCL_MOUSE ---qQ-IKRSARM-EGEEAERT-EDGCHDFERD---MKRFG--GPNKIRQ-KER-LRQ
KLG_BOVIN ---eNAFKRRR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GSGRSKQ-AQC-E-RR
2672_DANRE ---vD-KVRRRS-EGKCKGCVPR-TOEETEDFMD---KPFKG--GRNKKR-KER-LRQ
VCC_DROS ---hRRKSTAD-EGEETADYMTEDGGRDFEKD---MKRFG--GPNKIRQ-KER-LRQ
N05_HUMAN ---I-NRRQNRK-EGCAALLRMDGEGHDFEED---KPFKG--GSGRSKQ-AQC-E-RR
T9W2_SNETZ ---gF-SQRRRR-EGTFCPEVFKENQSEKNEIN---SK---LQD-SET-YRK
I45_PARL1 ---a-SERKKR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GSGRSKQ-AQC-E-RR
52X3_DANWE ---L-PYKRVV-EGKQWADITTYDEKGVSRNGY-IRKRLN-IHSRKYV-KER-KRR
N061_HUMAN ---hRMFKRVG-EGEAADQIT-EDGCAETELL---GLPFD-VASGLYC-KCE-RRR
852_CARAU ---gD-CVKRHR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GSGRSKQ-AQC-E-RR
300_DANRE ---L-KGRSRH-EGVEGCGQTQEDGCHER-ELR-PPR---PGLRRD-WKV-QRR
V19_BOVIN ---I-NRRQNRK-EGTCAALLRMDGEGHDFEED---KPFKG--GSGRSKQ-AQC-E-RR
D2H0_XENLA ---eNGIKRRR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GAGTKR-AEN-QRR
M12_XENLA ---F-LRKSDD-EGMQLGCRDEDEGSEVLLD---KMH-N-FDMQIK-ECL-LRR
V002_DROVA ---qPRRRRTF-EGNCAACORS-DECTEPTFMD---MVKFG--GPRAKQ-TEN-WRQ
R002_SGHA ---r9SP-RLRF-EGNCTGSLQEDGKGVSELD---KMH-N-FDMQIK-ECL-LRR
DQK2_XENTR ---qQ-IKRSARM-EGEEAERT-EDGCHDFERD---MKRFG--GPNKIRQ-KER-LRQ
XCL1_HUMAN ---qQ-IKRSARM-EGEEAERT-EDGCHDFERD---MKRFG--GPNKIRQ-KER-LRQ
S06_SHEEP ---eNAFKRRR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GSGRSKQ-AQC-E-RR
I04_TROAD ---L-SKKRKA-EGAGPFLKKEGSEKNEILH---RKT---GHQ-ICI-YRK
L04_DROSE ---F-FPMGR-EGEEGCRFPN-EMQDADRIV---SVG---HSP-REI-FRT
L2AE_PIG ---gA-RRRRVR-ERKCKAQVGE-EGVHYERD---MVKFG--GPRMKQ-SEV-LRQ
V0K_TFNG ---pA-KKKRRR-EGVEAPERLLINGVSSERN---RKT---GHQ-ICK-FRR
V122_DROMO ---qSKKKRRR-EGVEEVCDDP-E-EGKCKAKD---MVKFG--GSGRSKQ-AQC-E-RR
```

sequence 64 ID: B4DXJ5 HUMAN Residue: GLN (401)

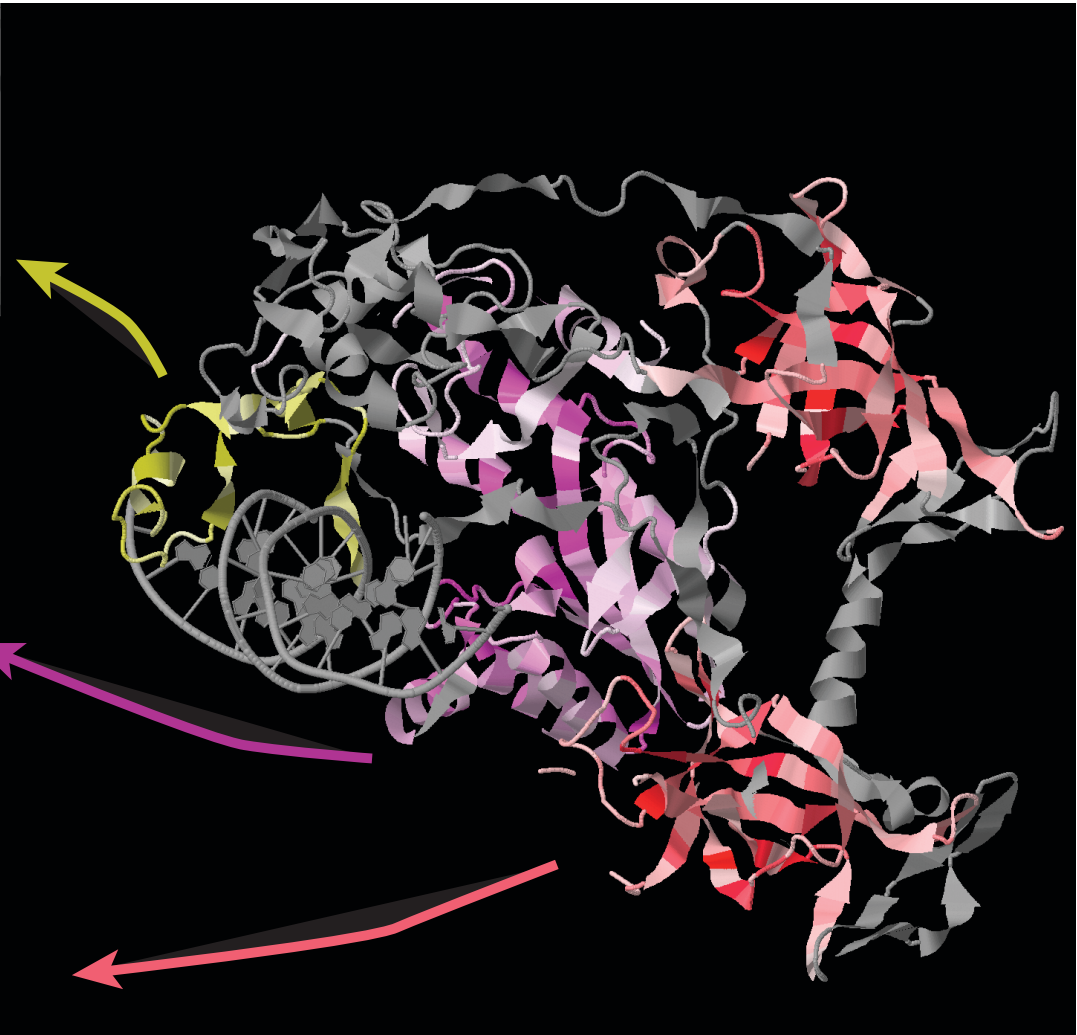
PF00145 Retrieved from PFAM (Full)

le	Edit	Select	View	Format	Colour	Calculate	Web Service
71_MOUSE	D	P	.	.	.	.	.
82_MERVI	D	P	.	.	.	.	.
R2_ZNTFA	D	G	F	.	.	.	.
12_STRC	D	R	F	.	.	.	.
176_BACCE	D	R	F	.	.	.	.
156_SFRM	D	R	F	.	.	.	.
N04_ERHAM	D	R	F	.	.	.	.
22_NH0P	D	R	F	.	.	.	.
VCE_NIGCO	D	R	F	.	.	.	.
16_SHEEP	D	R	F	.	.	.	.
112_NIGCO	D	R	F	.	.	.	.
105_SHEEP	D	R	F	.	.	.	.
195_SNYF3	D	R	F	.	.	.	.
W6_NEM1	D	R	F	.	.	.	.
M6_STRCC	D	R	F	.	.	.	.
V36_NIGCO	D	R	F	.	.	.	.
426_NIGCO	D	R	F	.	.	.	.
M8_DICDC	D	R	F	.	.	.	.
97_PHRIN	D	R	F	.	.	.	.
39_WBOW	D	R	F	.	.	.	.
13_SFRM	D	R	F	.	.	.	.
U2_ECDE	D	R	F	.	.	.	.
490_STRX	D	R	F	.	.	.	.
48_MACQ	D	R	F	.	.	.	.
42_SFRM	D	R	F	.	.	.	.
22_KCBR	D	R	F	.	.	.	.
63_GLOW	D	R	F	.	.	.	.
V3_CHRE	D	R	F	.	.	.	.
81_RLQDT	D	R	F	.	.	.	.
21_SALT	D	R	F	.	.	.	.
R2_BACE	D	R	F	.	.	.	.
48_NODDP	D	R	F	.	.	.	.
107_METOM	D	R	F	.	.	.	.
N2_NAGEB	D	R	F	.	.	.	.
193_GLOBO	D	R	F	.	.	.	.

sequence 1995 ID: C6GMH0\_STRSX Residue: SER (37)

PF01426 Retrieved from PFAM (Full)

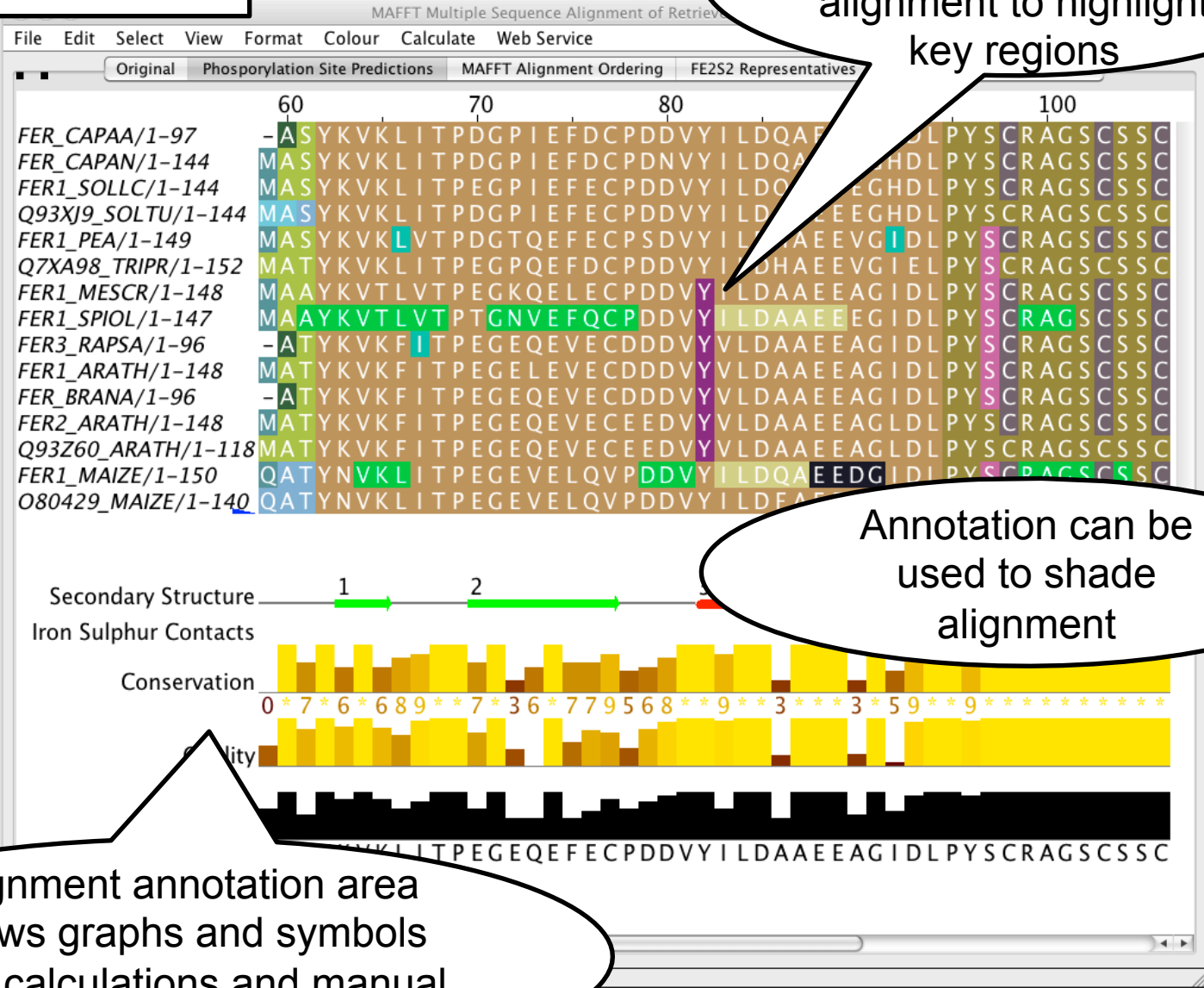
Edit	Select	View	Format	Colour	Calculate	Web Service
T1_MOUSE	F	V	C	E	N	Q
T1_PARL1	V	P	I	D	E	---
T1_PARL1	L	Y	W	S	D	E
T1_RAT	L	Y	W	S	D	E
T1_RAT	L	Y	W	S	D	E
Z_CAEL	L	F	N	S	E	L
9_CAEL	V	F	A	T	P	Y
1_HUMAN	L	F	L	S	R	Q
1_MOUSE	L	F	L	S	R	Q
1_RAT	L	F	L	S	R	Q
2_HUMAN	L	F	L	S	R	Q
2_MOUSE	L	F	L	S	R	Q
2_BOVIN	L	F	L	S	R	Q
3_HUMAN	L	F	L	S	R	Q
3_MOUSE	L	F	L	S	R	Q
169_ASCIM	V	A	L	V	H	Q
173_ARATH	L	F	L	T	N	E
173_ARATH	L	F	L	T	N	E
126_ARATH	L	F	L	T	N	D
311_ASCIM	L	F	L	T	D	S
103_ASCIM	L	F	L	T	D	S
166_DAUCA	V	F	L	S	E	E
166_DAUCA	V	F	L	T	N	D



**AND NOW: SOME DISORDER**

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 ... Ferredoxin Structure



A collection of logos for bioinformatics tools: RONN (Regional Order Neural Network), DisEMBL™, IUPred, and GLOBPLOT 2.

80  
P Q E F D C P D D V Y  
... E Q E V E C D D D V Y  
FER1\_ARATH/1-148 M A T Y K V K F I T P E G E L E V E C D D D V Y  
FER\_BRANA/1-96 - A T Y K V K F I T P E G E Q E V E C D D D V Y  
FER2\_ARATH/1-148 M A T Y K V K F I T P E G E Q E V E C E E D V Y

Secondary Structure ————

Iron Sulphur Contacts ■ ■ ■ ■ ■ ■ ■

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

# Protein Disorder prediction






- 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**)

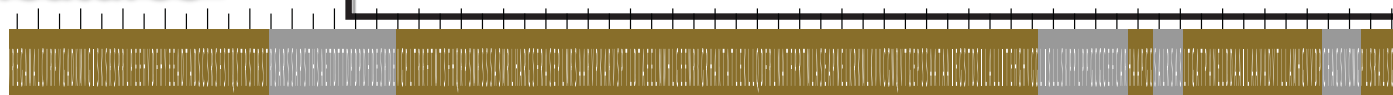
# Disorder Predictions from JABAWS

JABAWS  
Analysis  
Service

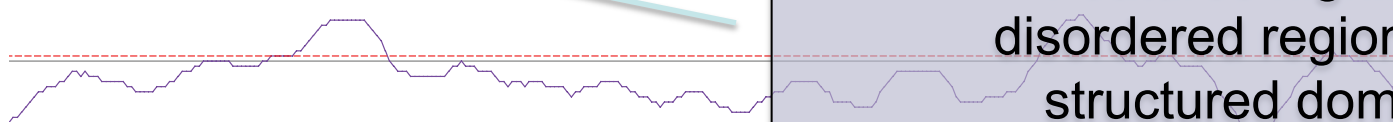
*Jalview*  
**JABAWS 2.0**  
*Client*

Process results into  
both **annotation** and  
**features**

<input checked="" type="checkbox"/> DisembIWS	<input checked="" type="checkbox"/> GlobPlotWS	<input checked="" type="checkbox"/> IUPredWS
Feature Type	Colour	Display
Protein Disorder		<input checked="" type="checkbox"/>
COILS		<input type="checkbox"/>
REM465		<input type="checkbox"/>
HOTLOOPS		<input type="checkbox"/>
Globular Domain		<input checked="" type="checkbox"/>

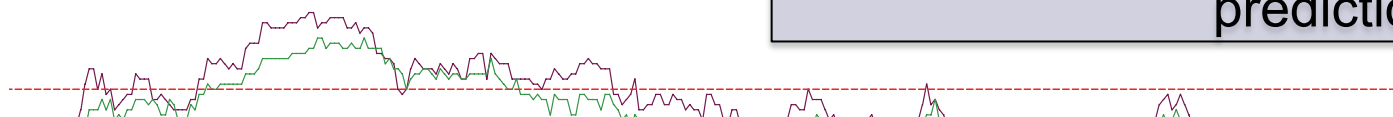


GlobPlotWS (Dydx)



Features highlight  
disordered region or  
structured domain  
predictions

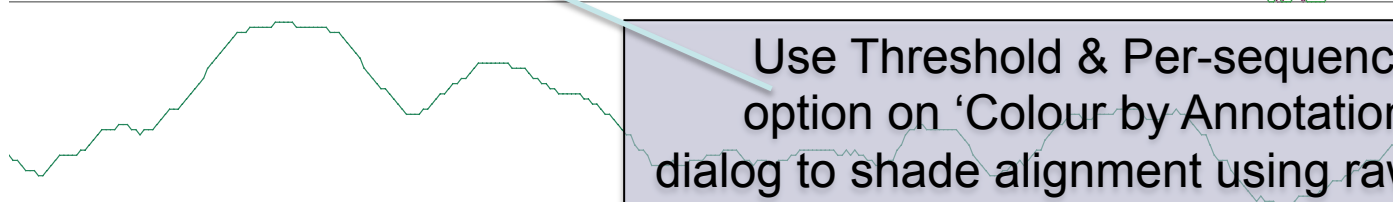
IUPredWS (Short)



IUPredWS (Long)



JronnWS (JRonn)



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

# Exercise 27 – IUPred and Disembl Disorder predictions

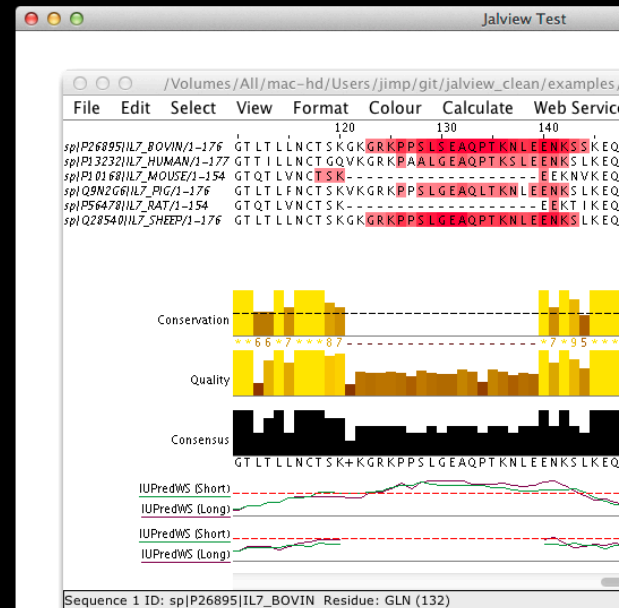
- Task
  - Apply the Disembl and IUPred disorder predictors to a set of protein sequences
  - Compare disordered predictions and available 3D data



# Exercise 27 – IUPred and Disembl Disorder predictions

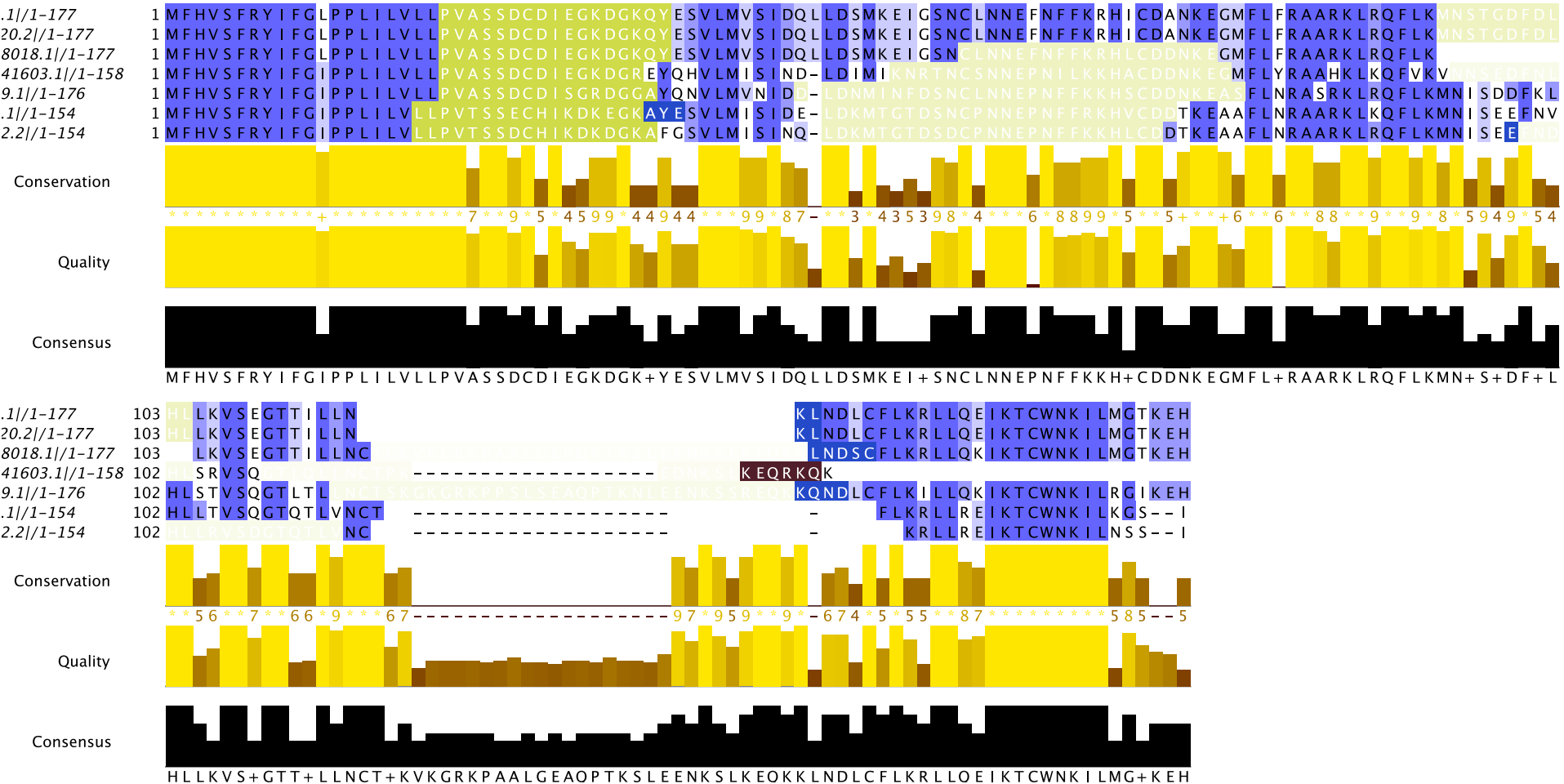
- Task
  - Apply the Disembl and IUPred disorder predictors to a set of protein sequences
  - Compare disordered predictions and available 3D data
- Question
  - How do you decide which regions are ‘significantly disordered’ ?

# Disorder in Interleukin 7



Human Interleukin-7 structure in Chimera coloured according to IUPred disorder prediction made in Jalview, with a glutamate sidechain

# Disorder prediction on orthologs



# Creating, editing and using annotation.

## Exercise 29 (Sec. 2.8.4)

- 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 ?