Jalview and JPred4
Day 2: Annotation, Structures, cDNA, RNA, and Jalview clinic

supported by wellcome trust

BBSRC
20 Years of Pioneering
Great British Bioscience

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

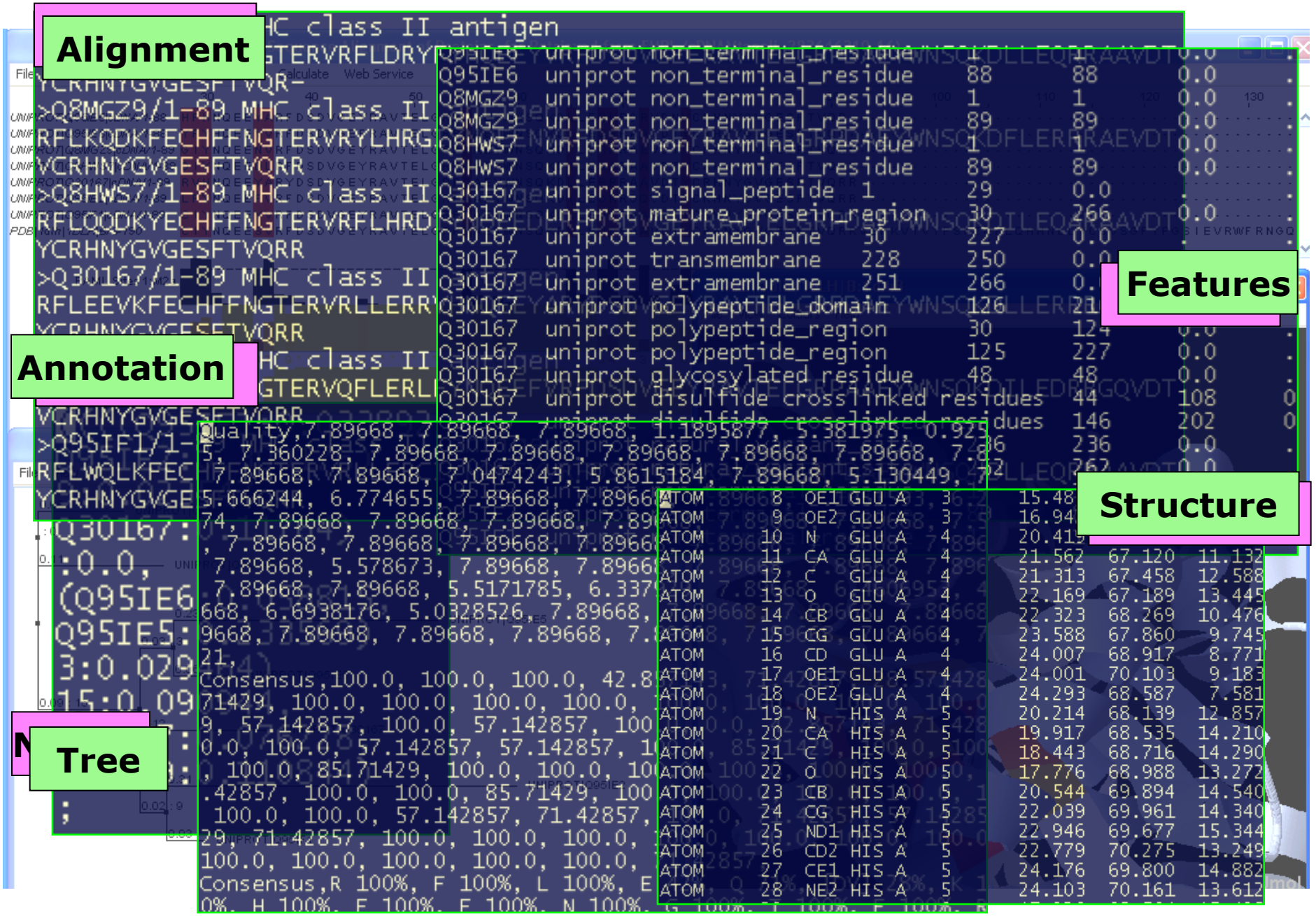
Lunch: 12.30am-1.30pm

Session 7: RNA, cDNA and Jalview

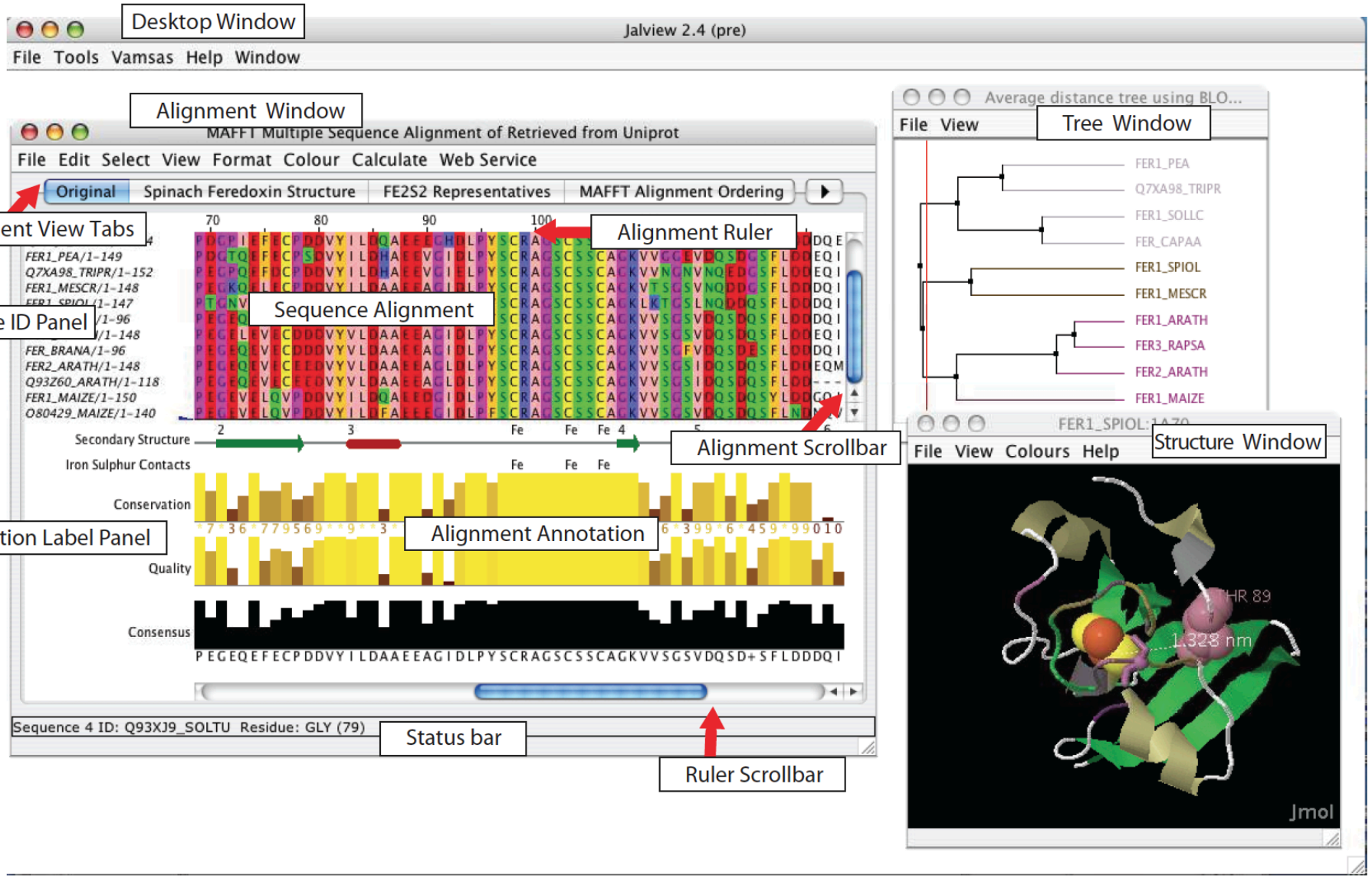
- Viewing RNA structure & Predicting RNA srtructure in VARNA and RNAAliFold
- cDNA from the European Nucleotide Archive
- Aligning by protein and nucelotide

Coffee: 3.30pm – 4pm

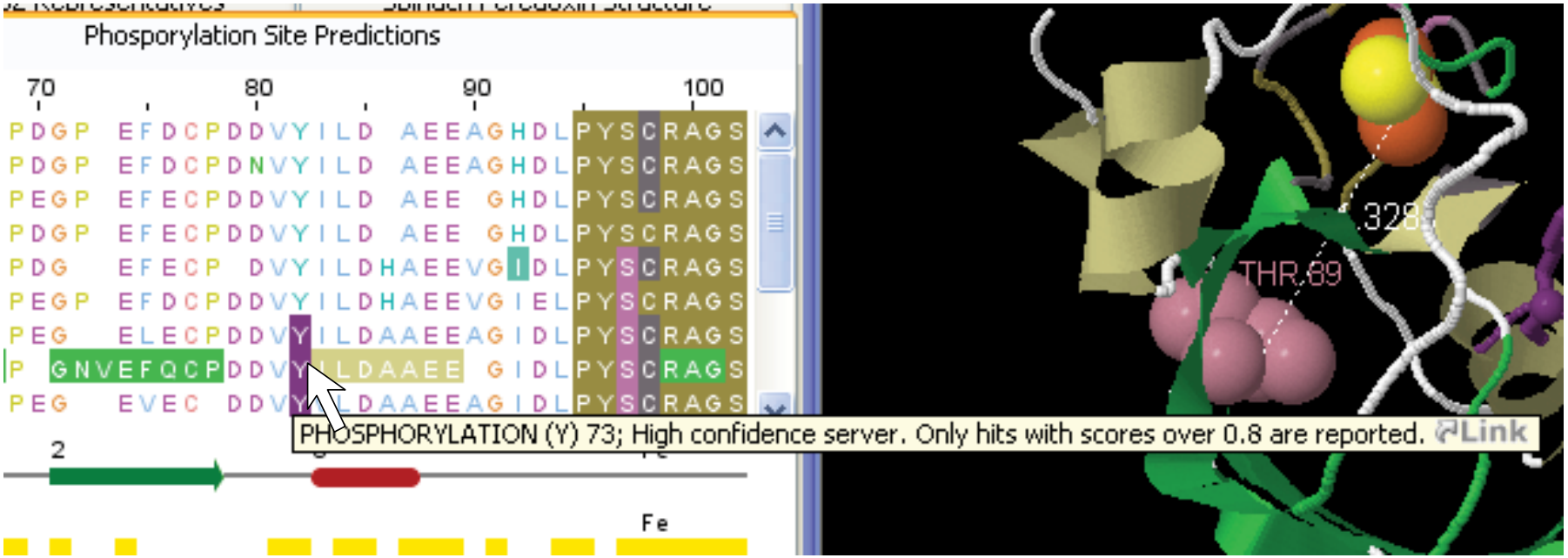
Session 8: Jalview clinic


Alignment
HC class II antigen
>Q8MGZ9/1-89 MHC class II
RFLKQDKFECHFFNGTERVRYLHRG
VCRHNYGVGESFTVQRR
>Q8HWS7/1-89 MHC class II
RFLQQDKYECHFFNGTERVRFLHRD
YCRHNYGVGESFTVQRR
>Q30167/1-89 MHC class II
RFLEEVKFECHFFNGTERVRLLERR
VCRHNYGVGESFTVQRR
HC class II
GTERVQFLERL
VCRHNYGVGESFTVQRR
>Q95IF1/1-
RFLWQLKFEC
YCRHNYGVGE
Annotation
Q95IE6 uniprot non_terminal_residue 1 1 0.0
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 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
Features
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, 7.8
21,4)
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 20.419
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

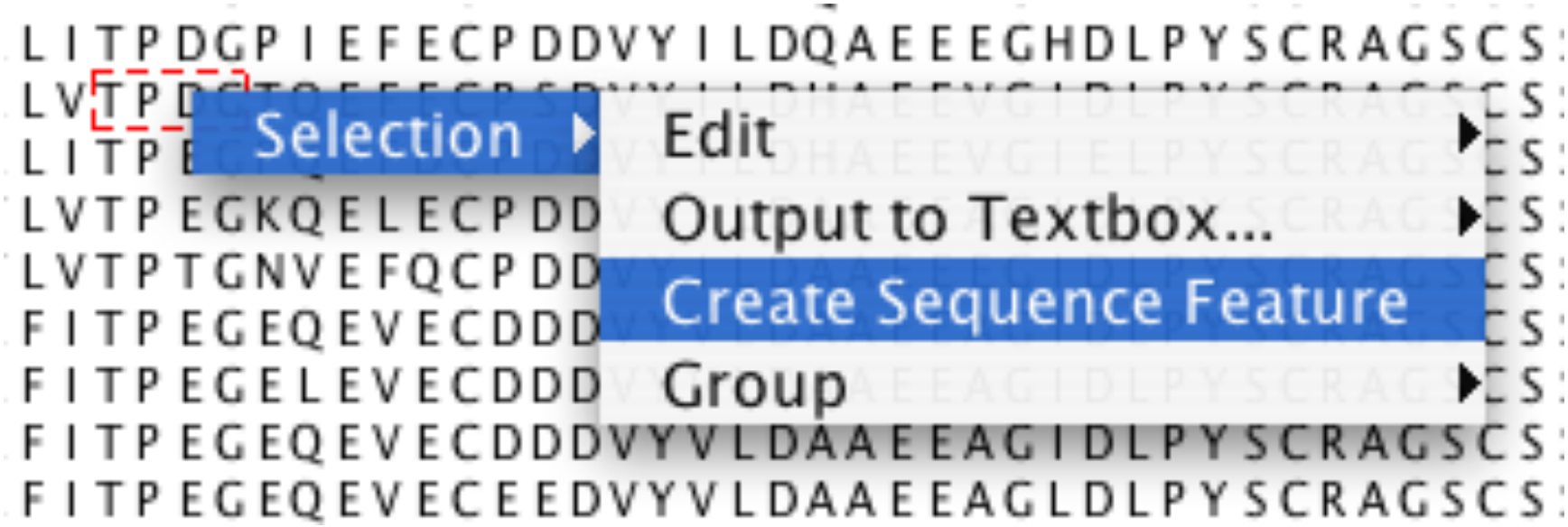
Anatomy of Jalview: Figure 1.7



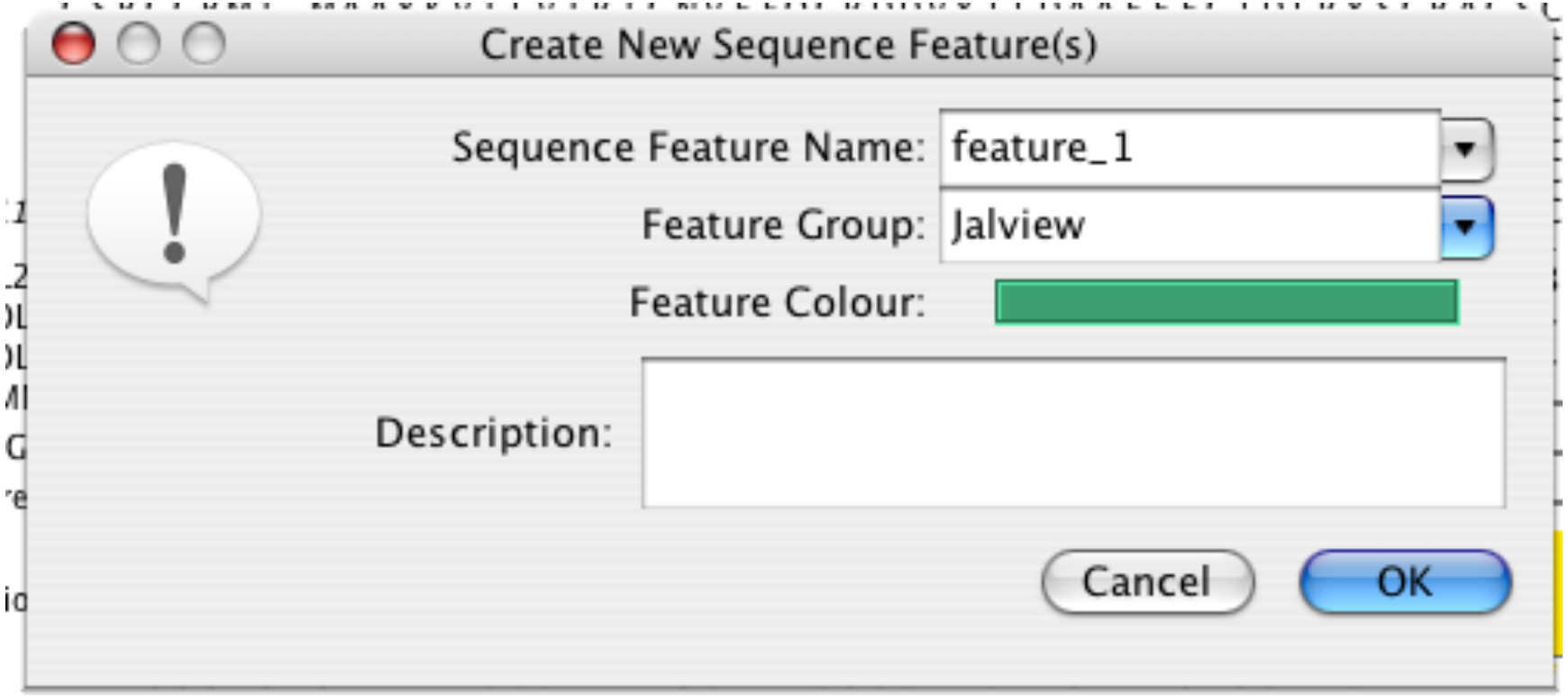
Sequence Features

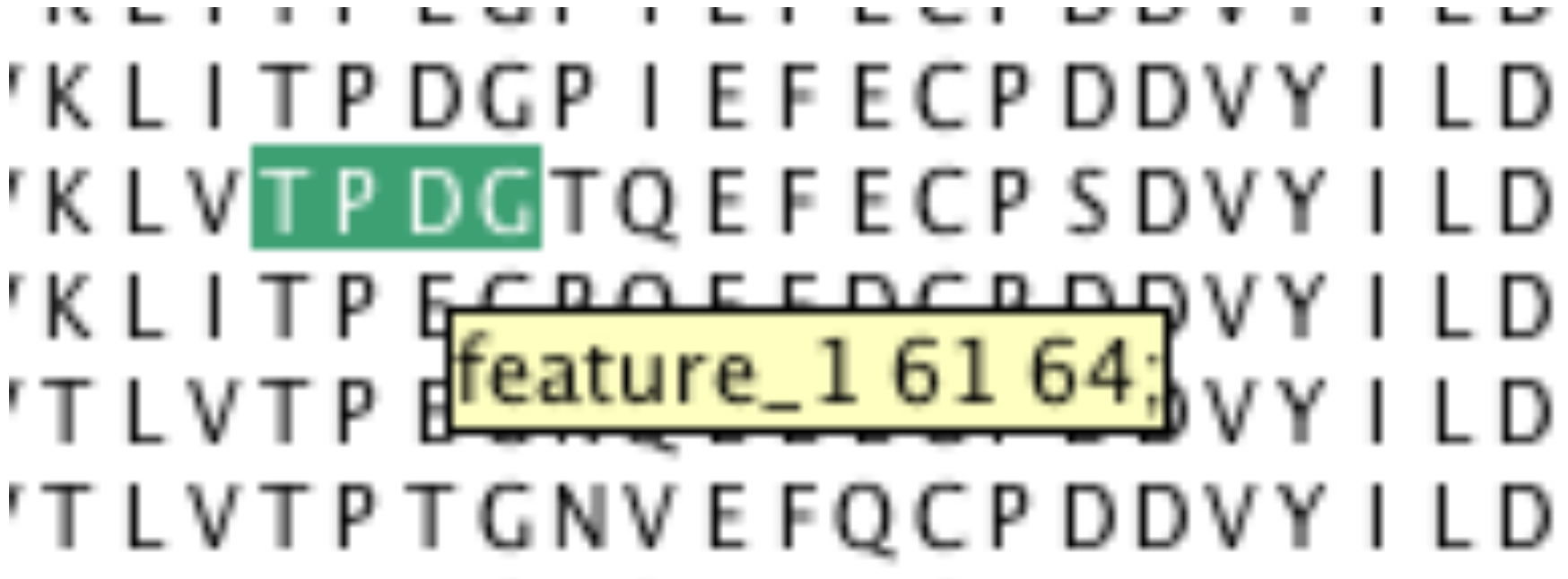


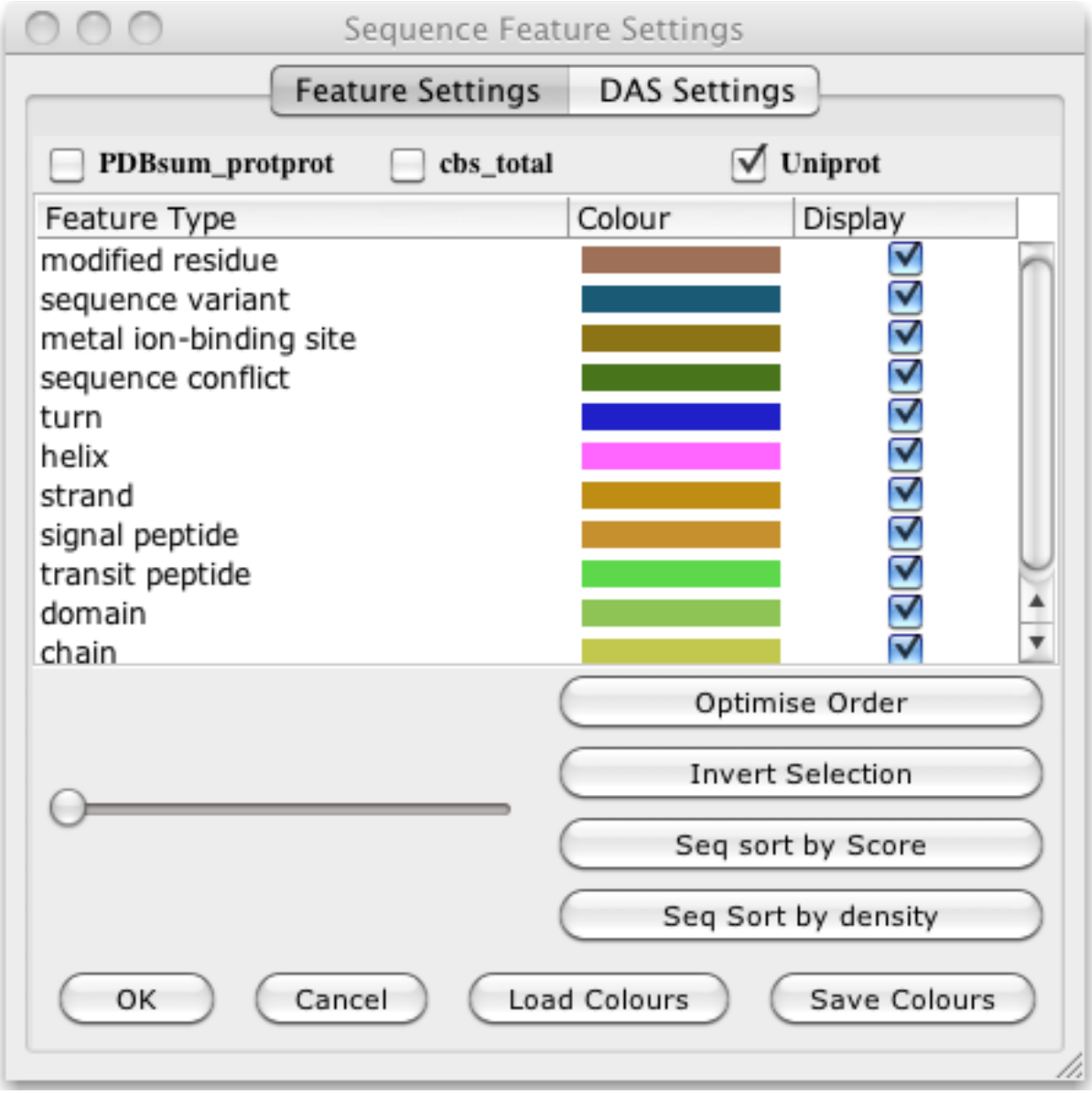
Features from selected region



Feature creation dialog box




KLITPDGPIEFECPDDVYILD
KLVTPDGTQEFECPSDVYILD
KLITP
feature_1 61 64;
TLVTP
TLVTPTGNVEFQCPDDVYILD


Sequence Feature Settings
Feature Settings
DAS Settings
PDBsum_protprot
cbs_total
Uniprot
Feature Type
Colour
Display
modified residue
sequence variant
metal ion-binding site
sequence conflict
turn
helix
strand
signal peptide
transit peptide
domain
chain
Optimise Order
Invert Selection
Seq sort by Score
Seq Sort by density
OK
Cancel
Load Colours
Save Colours

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 ?

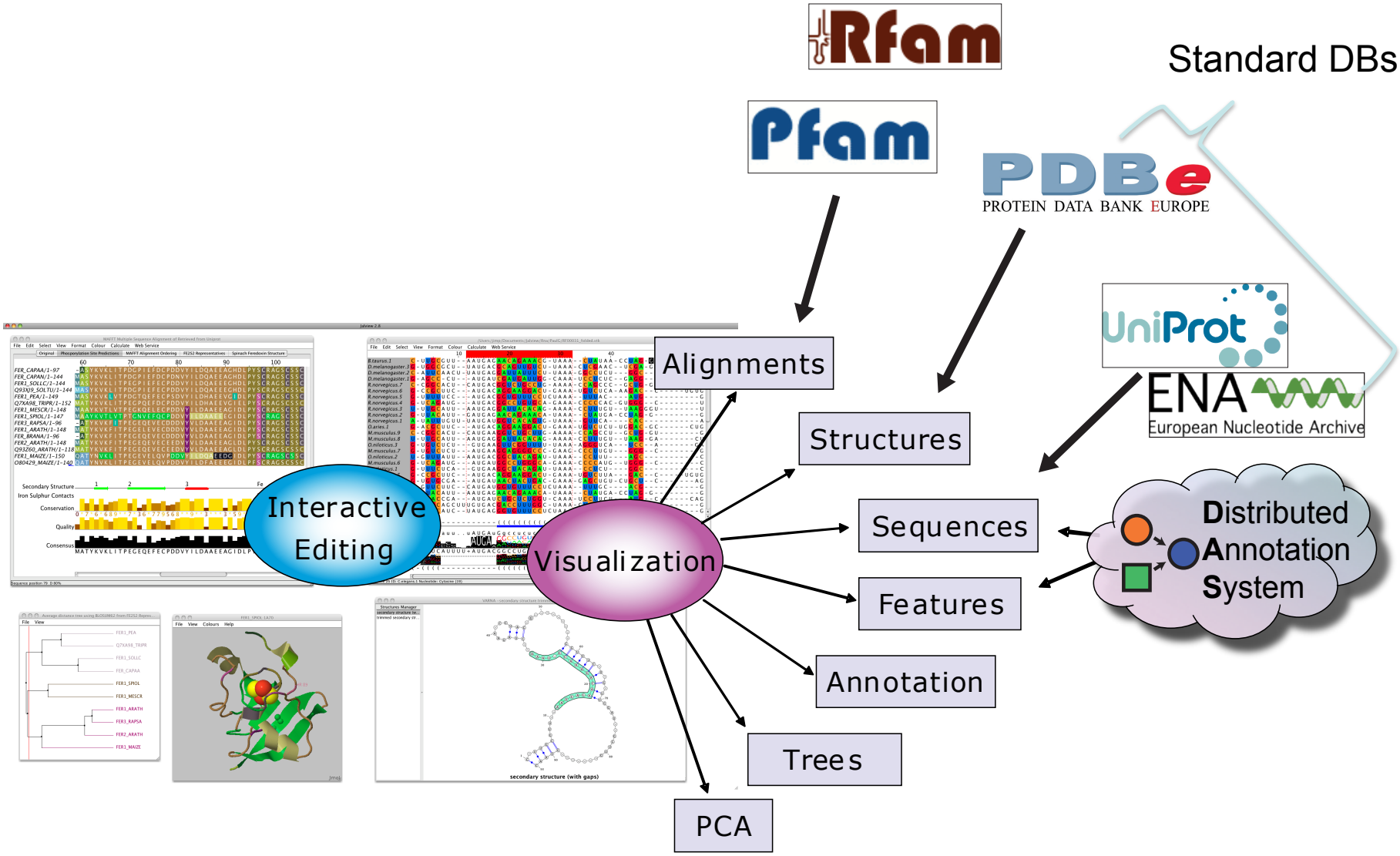
Getting and working with sequence features and annotation

- 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

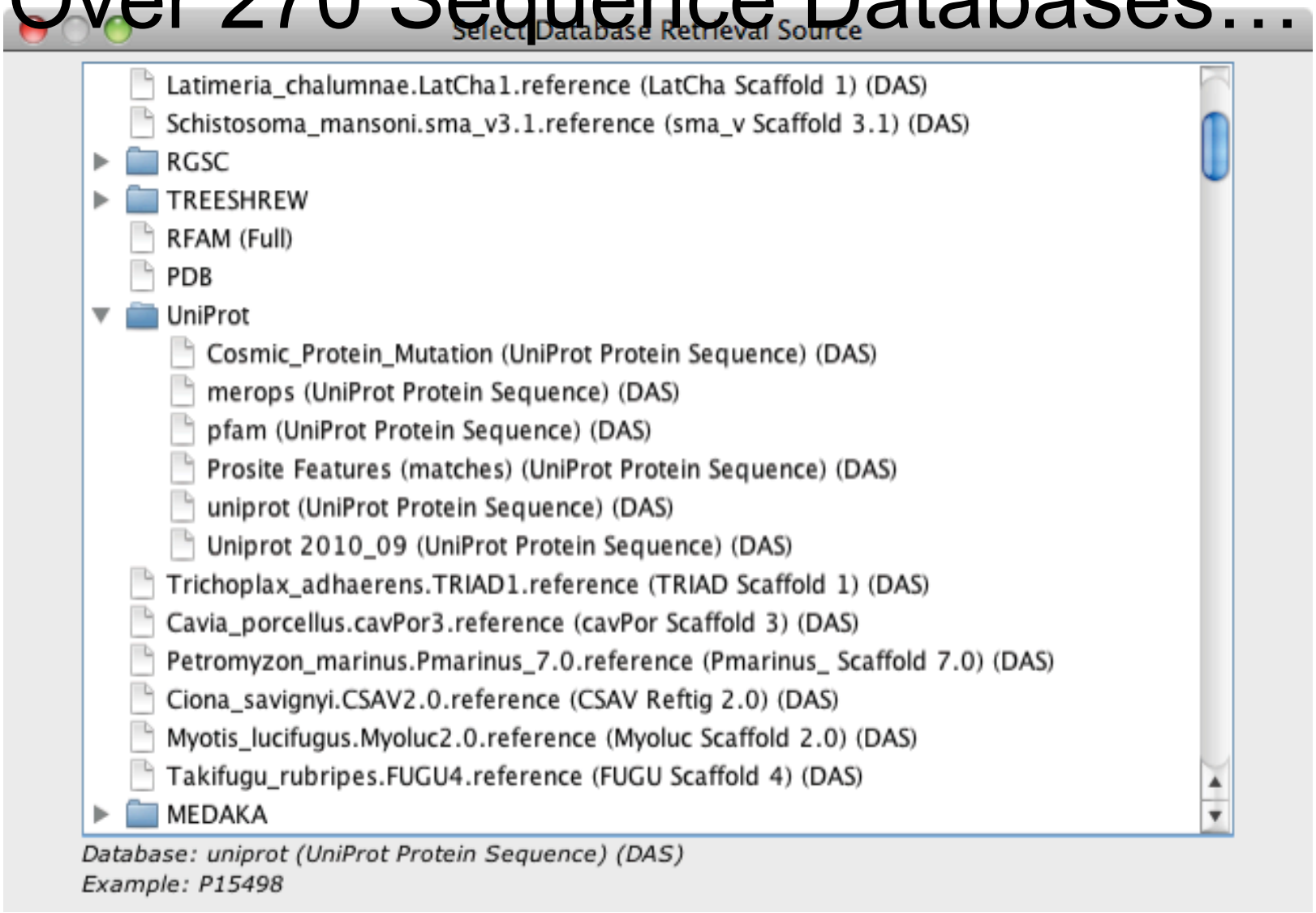
Sources of sequence feature data

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

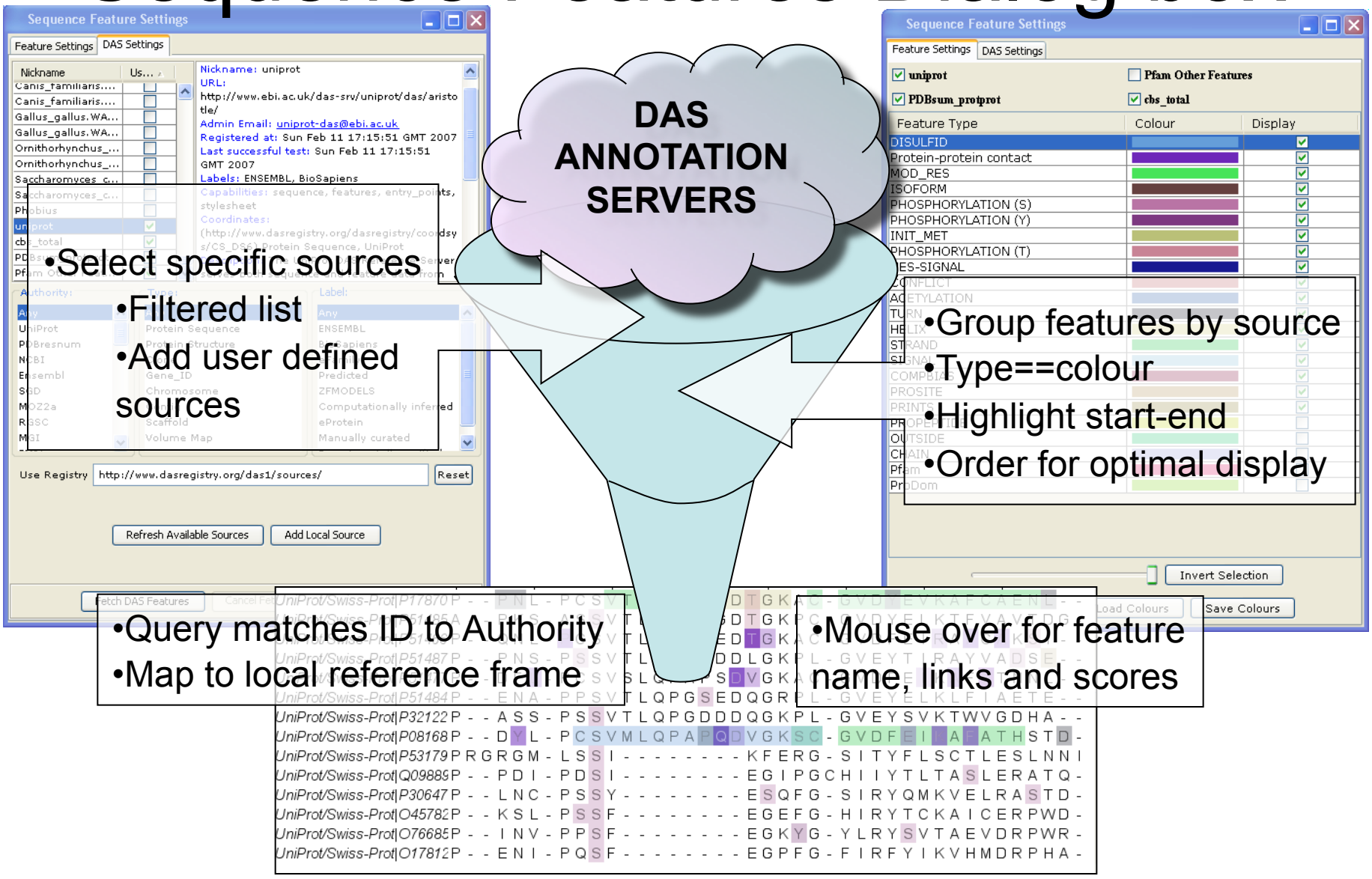
Retrieval from External Databases



DAS allows Jalview access to Over 270 Sequence Databases…

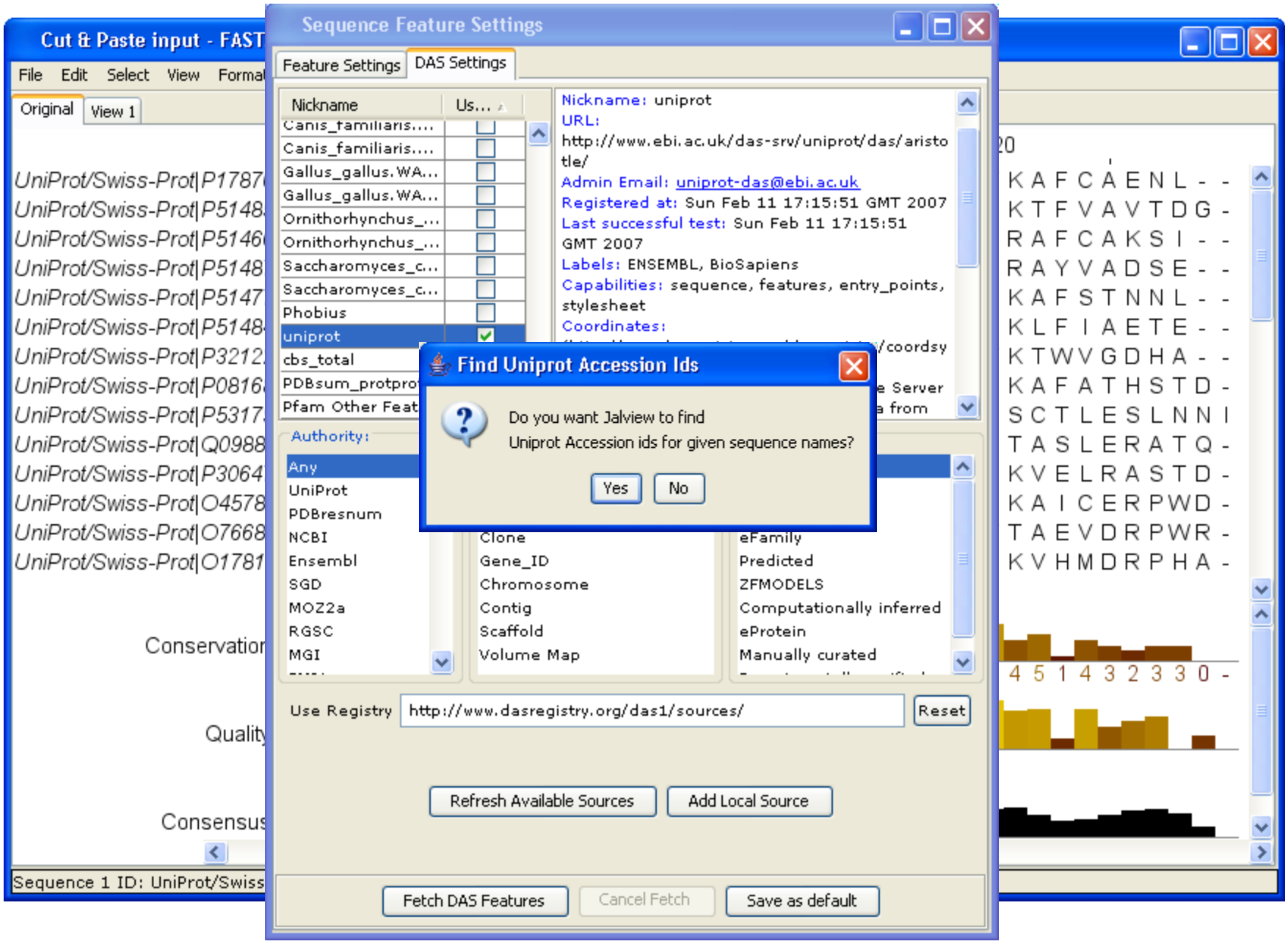


Sequence Features Dialog box

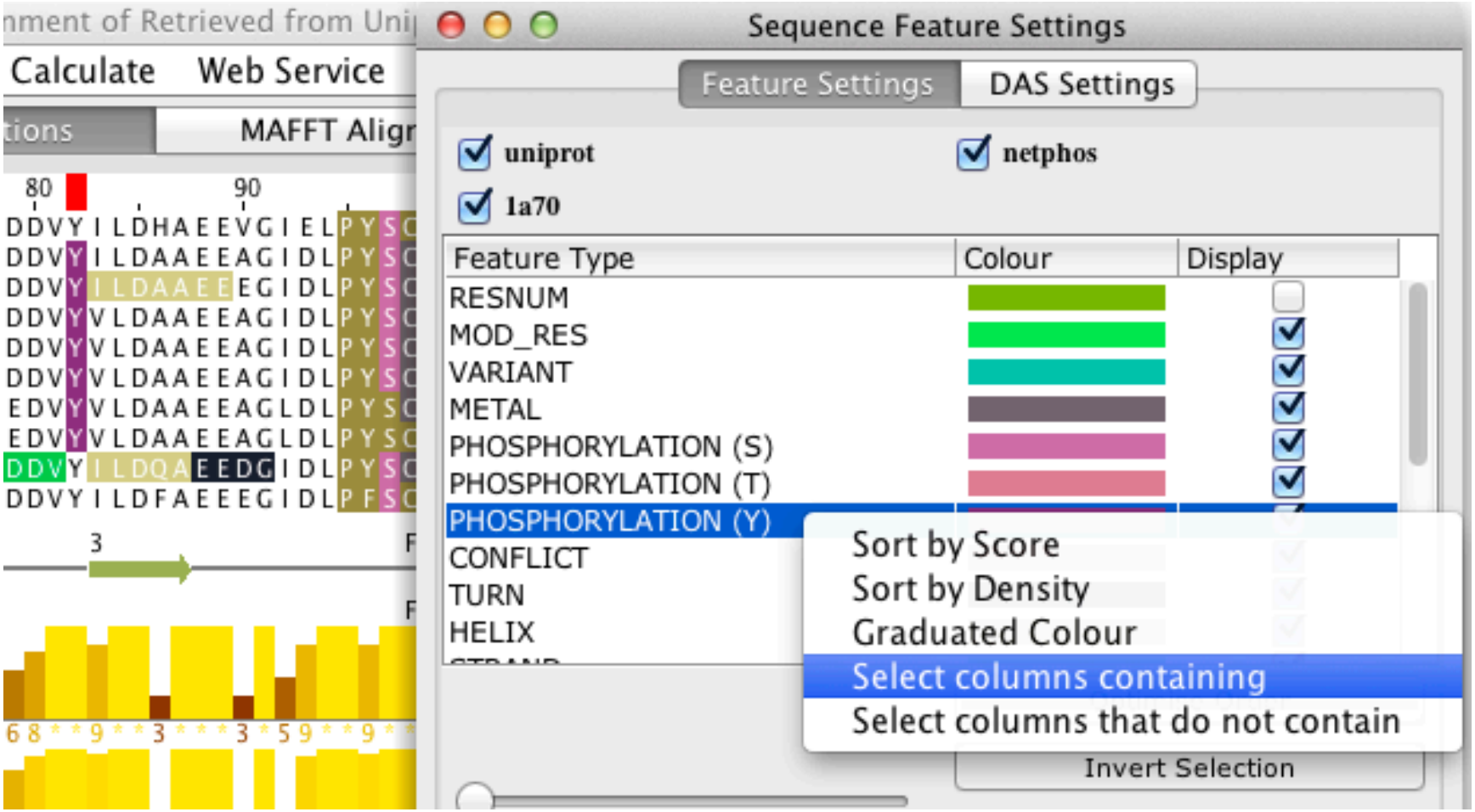


Jalview and Sequence Databases Sec 2.9.1 Ex. 29

- 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 - FAST
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....
Canis_familiaris....
Gallus_gallus.WA...
Gallus_gallus.WA...
Ornithorhynchus_...
Ornithorhynchus_...
Saccharomyces_c...
Saccharomyces_c...
Phobius
uniprot
cbs_total
PDBsum_protpro
Pfam Other Feat
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:
Find Uniprot Accession Ids
Do you want Jalview to find
Uniprot Accession ids for given sequence names?
Yes
No
Authority:
Any
UniProt
PDBresnum
NCBI
Ensembl
SGD
MOZ2a
RGSC
MGI
Clone
Gene_ID
Chromosome
Contig
Scaffold
Volume Map
eFamily
Predicted
ZFMODELS
Computationally inferred
eProtein
Manually curated
Use Registry
http://www.dasregistry.org/das1/sources/
Reset
Refresh Available Sources
Add Local Source
Fetch DAS Features
Cancel Fetch
Save as default
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



Flanking alignments and ‘Trim retrieved sequences’

- New feature for Proteomic hit analysis

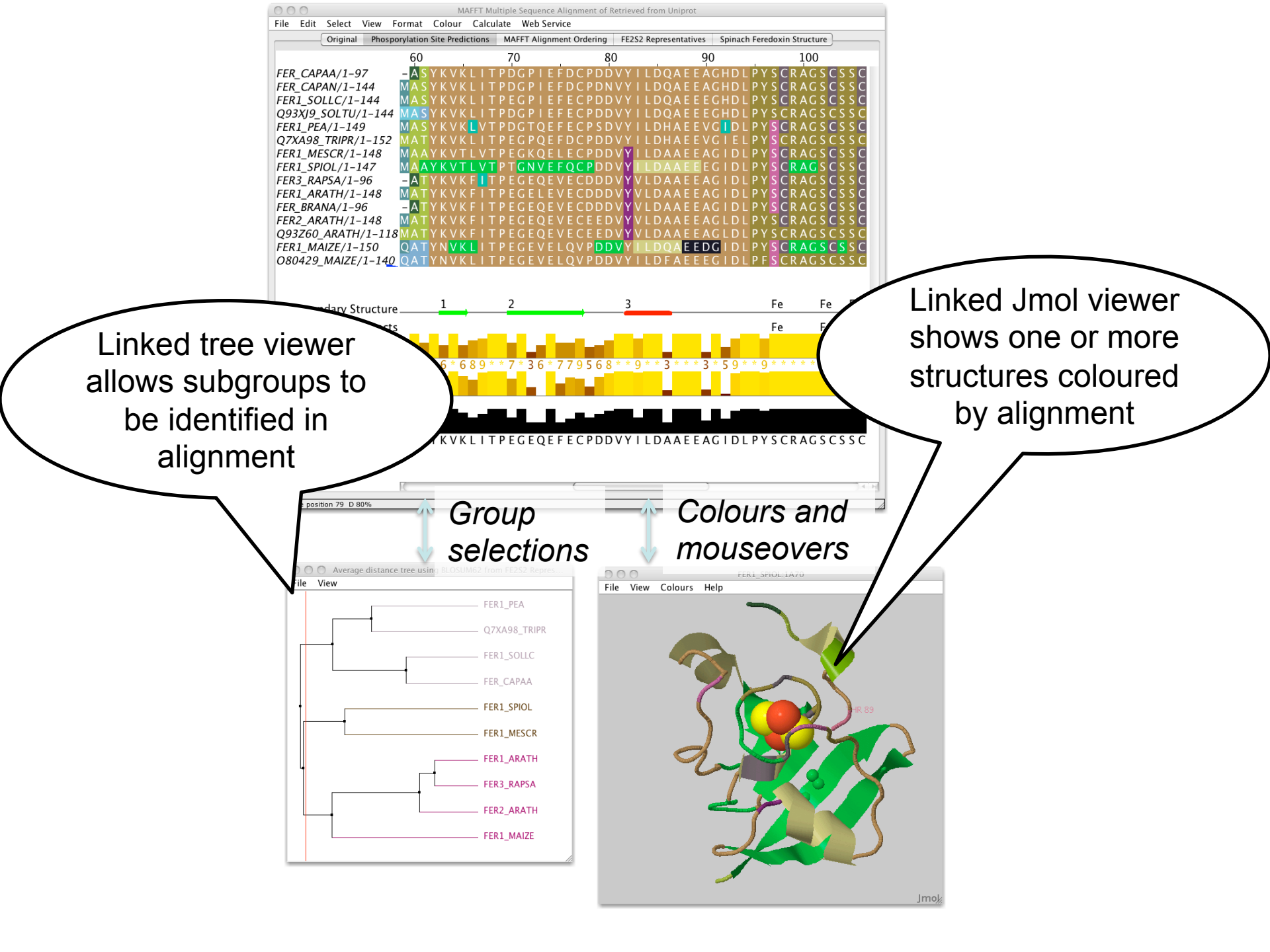
First create a new FASTA file subsequences

>FER1_PEA

ASYKVKLVTPDGTQEFECPSDVYILDHAEEVGIDLPYSCRAGSCSSCAGKVVGGEVDQS
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

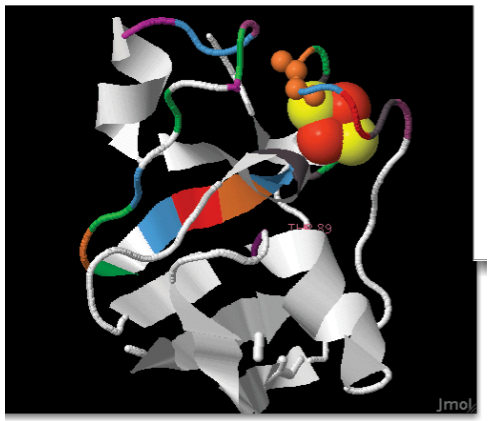
3D STRUCTURE IN JALVIEW

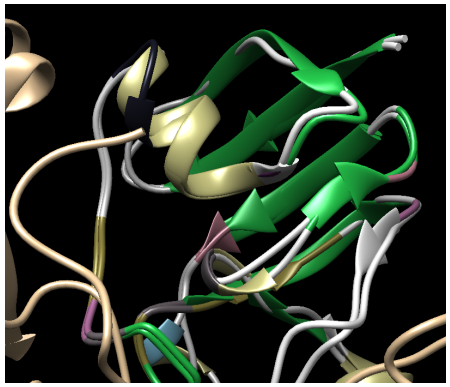

Linked tree viewer allows subgroups to be identified in alignment
Linked Jmol viewer shows one or more structures coloured by alignment
Group selections
Colours and mouseovers
MAFFT Multiple Sequence Alignment of Retrieved from Uniprot
File Edit Select View Format Colour Calculate Web Service
Original
Phosporylation Site Predictions
MAFFT Alignment Ordering
FE2S2 Representatives
Spinach Feredoxin Structure
60 70 80 90 100
FER_CAPAA/1-97
FER_CAPAN/1-144
FER1_SOLLC/1-144
Q93XJ9_SOLTU/1-144
FER1_PEA/1-149
Q7XA98_TRIPR/1-152
FER1_MESCR/1-148
FER1_SPIOL/1-147
FER3_RAPSA/1-96
FER1_ARATH/1-148
FER_BRANA/1-96
FER2_ARATH/1-148
Q93Z60_ARATH/1-118
FER1_MAIZE/1-150
O80429_MAIZE/1-140
Fe
Average distance tree using BLOSUM62 from FE2S2 Repres
File View
FER1_PEA
Q7XA98_TRIPR
FER1_SOLLC
FER_CAPAA
FER1_SPIOL
FER1_MESCR
FER1_ARATH
FER3_RAPSA
FER2_ARATH
FER1_MAIZE
File View Colours Help
Jmol

Biomolecular Structure Visualization in the Jalview Desktop

Jmol

jmol.sourceforge.net





UCSF CHIMERA

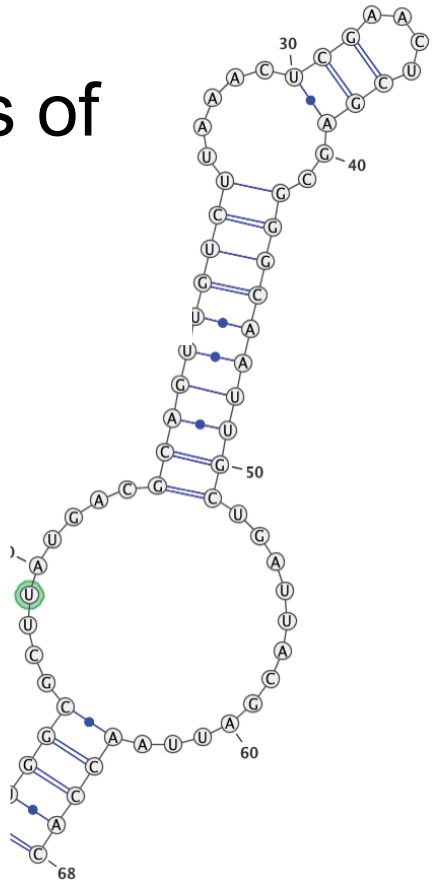
an Extensible Molecular Modeling System

www.rbvi.ucsf.edu

VARNA

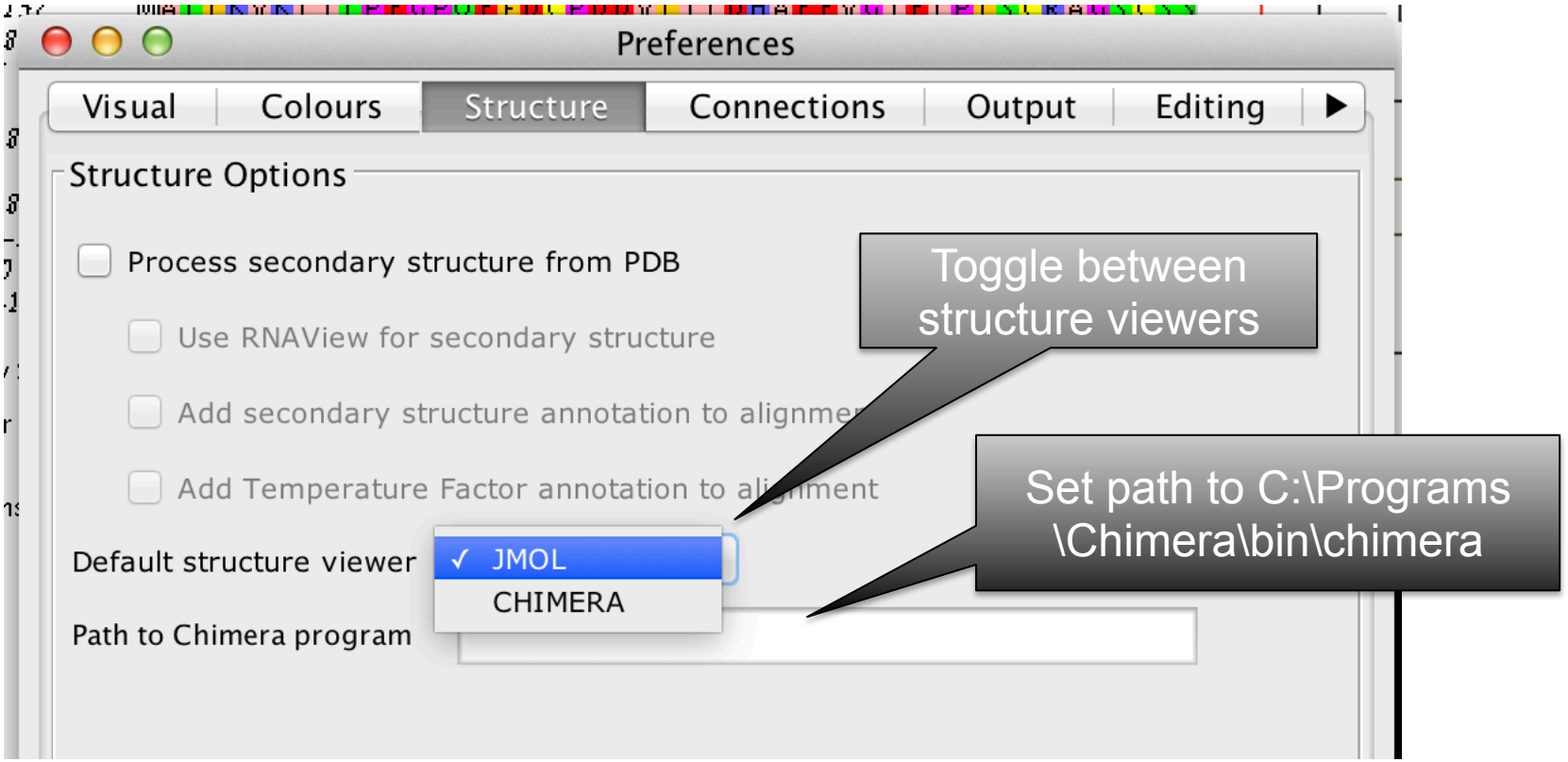
Visual Analysis of RNA

varna.lri.fr



Configure Jalview and Chimera

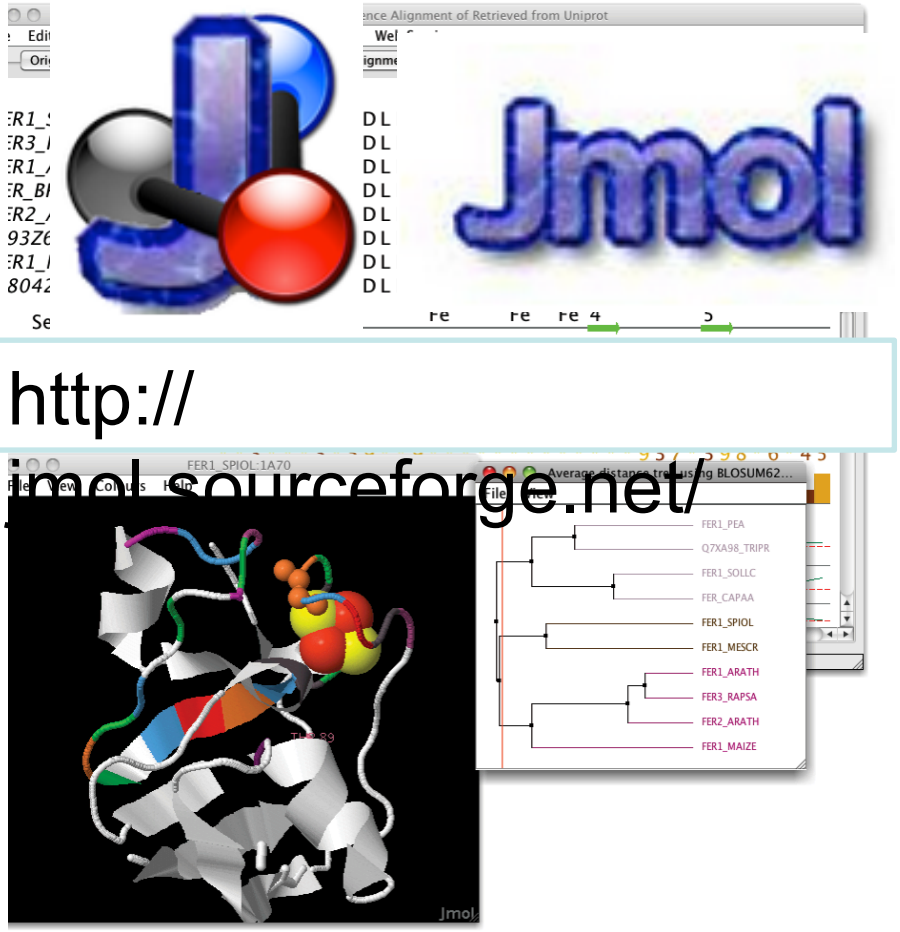
- Jalview ‘Structure’ preferences tab



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

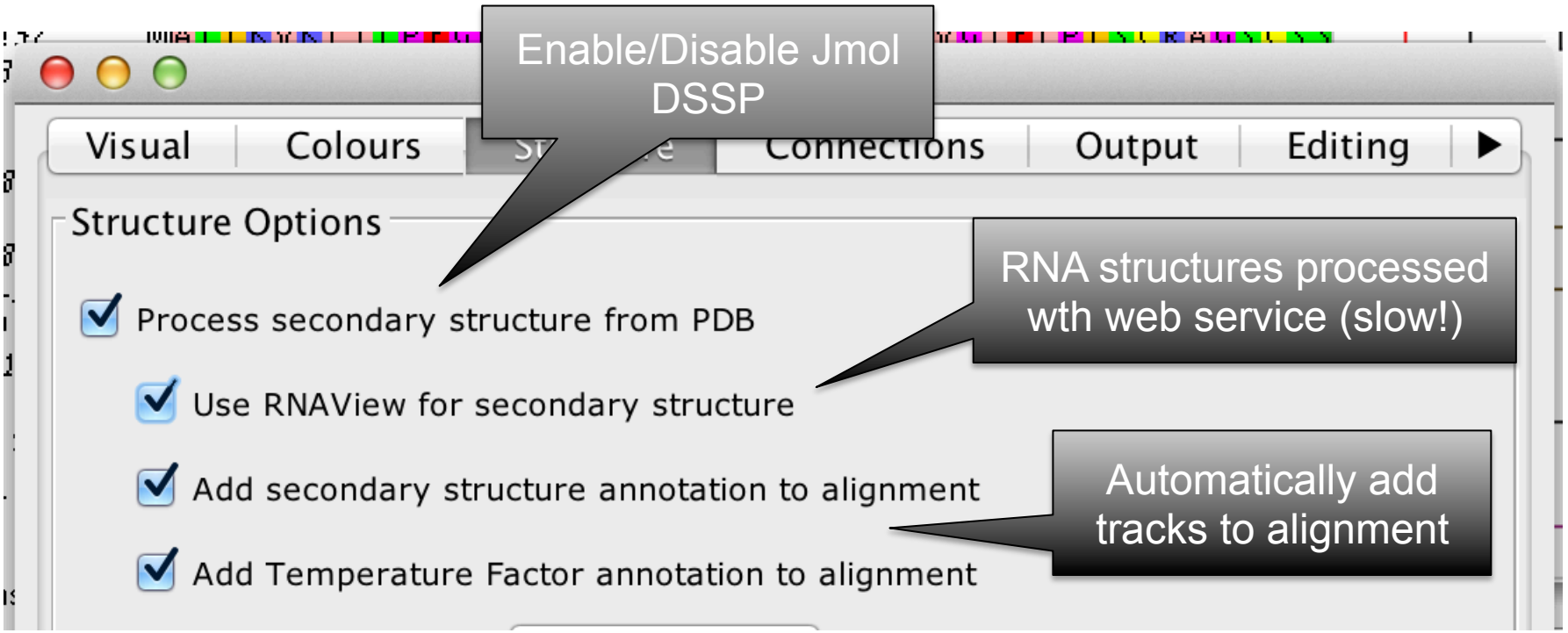


http://jmol.sourceforge.net/

- 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)

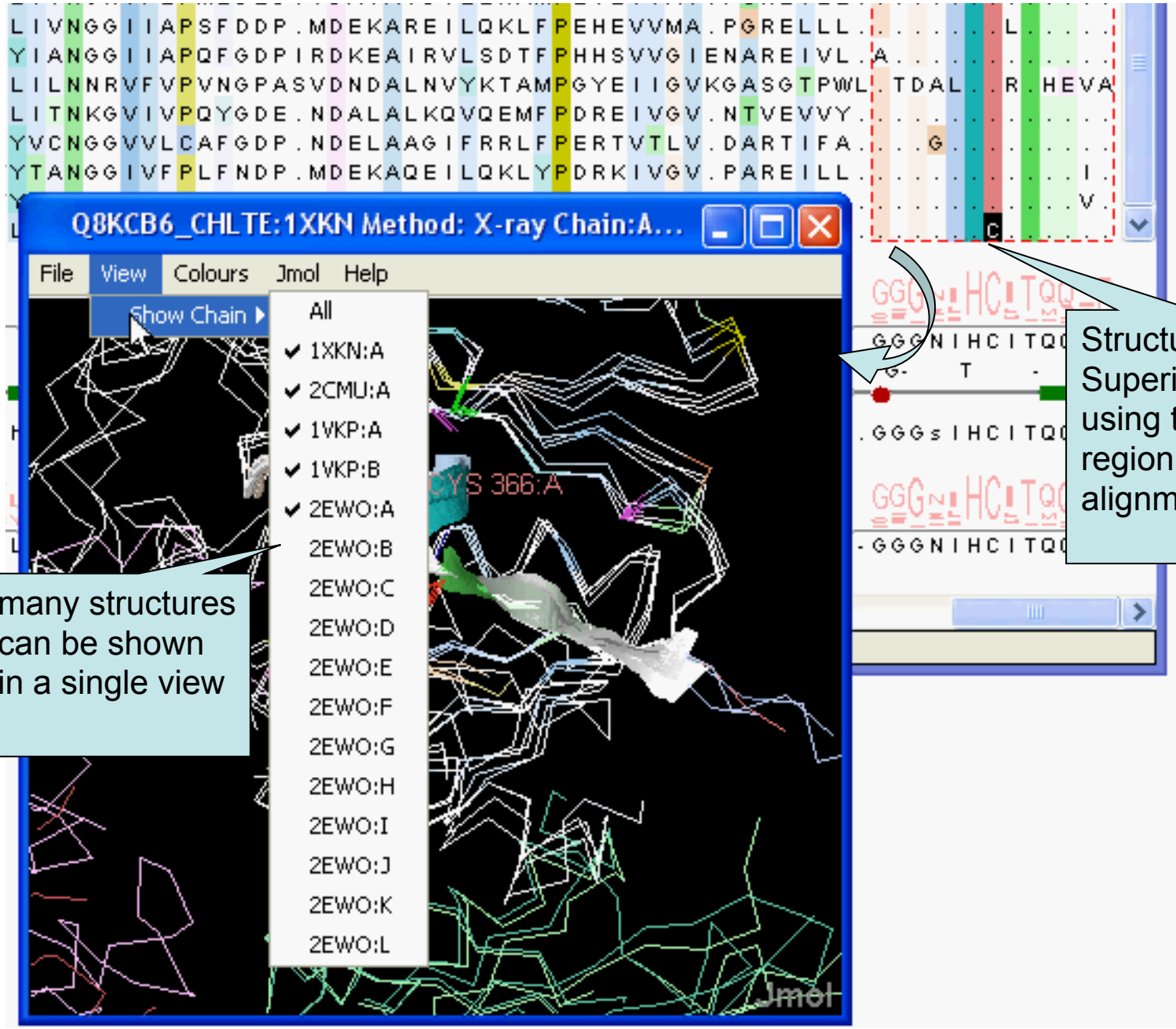
    Fetch Database Refs->UNIPROT->Uniprot
  - Save and load structures and manipulate colouring

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


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
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:

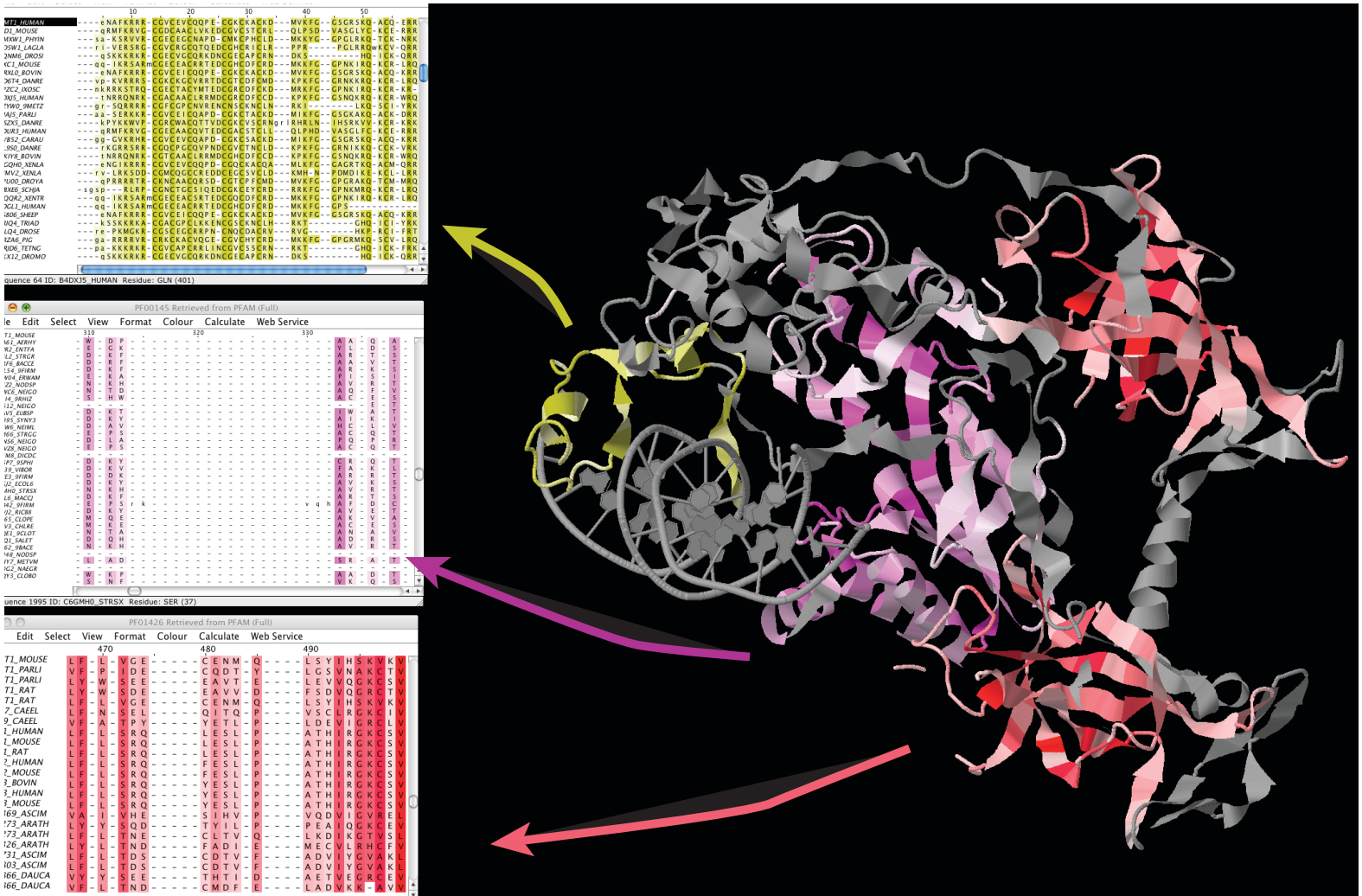
- **Use Jmol/Structure Viewer->Align menu**

  - 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

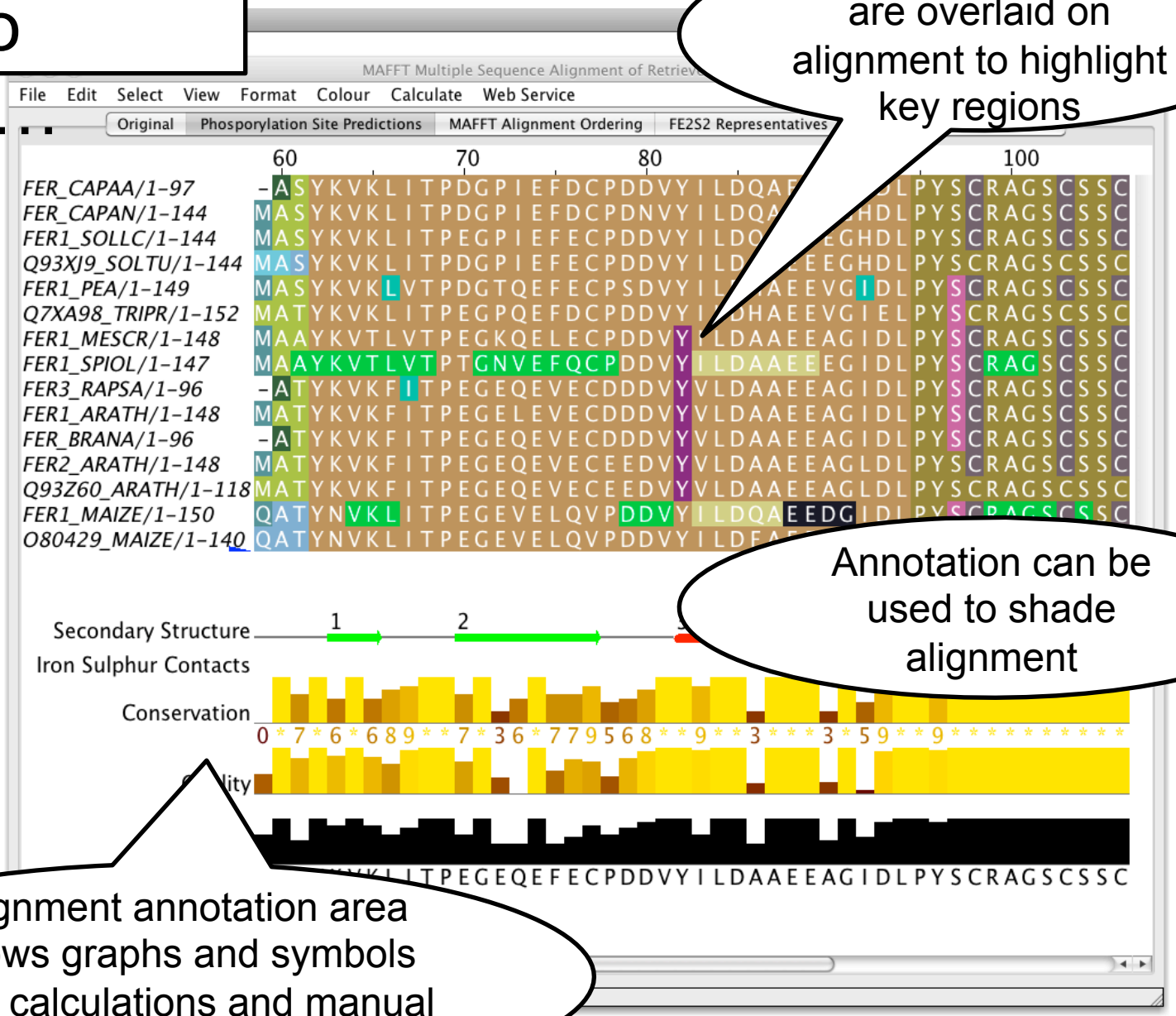
- Task
  - Align structures using the ferredoxin alignment
  - Experiment with views to control what part of the alignment is used to superimpose the structures
- Questions
  - What colourscheme would highlight the conserved parts of the structures ?
  - Which view gave the ‘best’ structure superposition ?
    - How did you decide this ?

Colouring structures using many multiple alignments
Sect 2.1.5. Exercise 16



AND NOW: SOME DISORDER

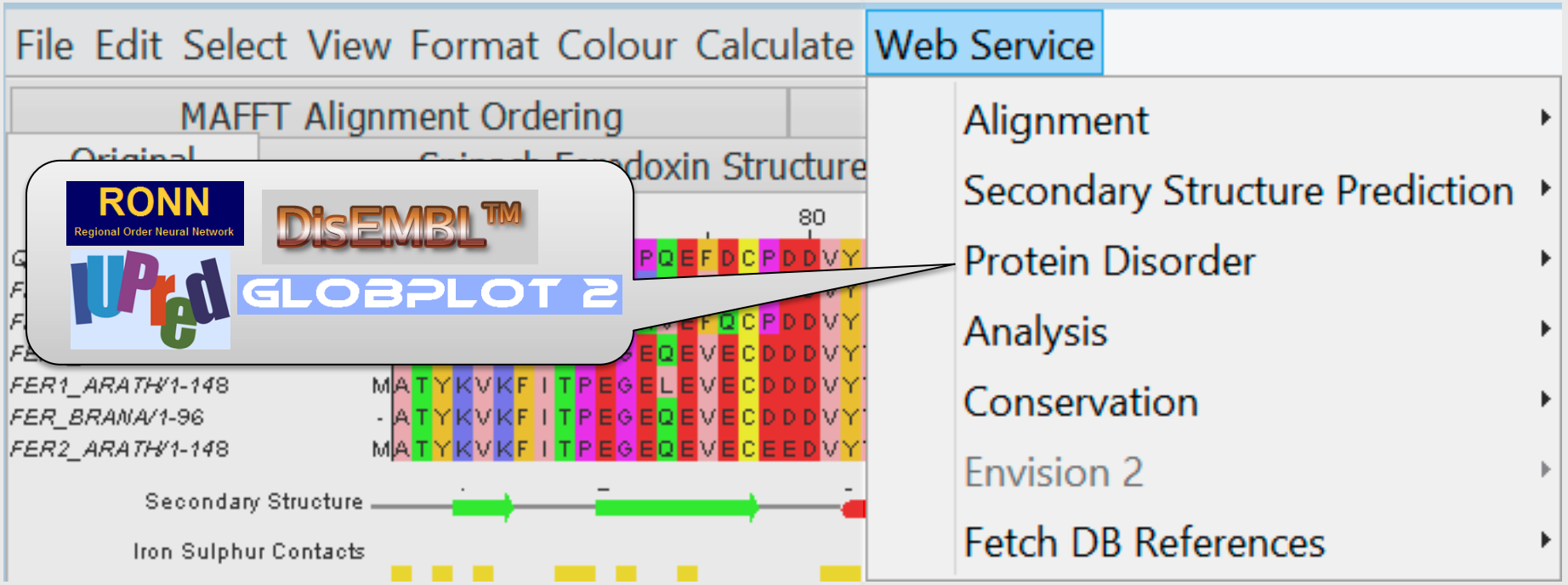
Just to recap…



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
RONN
Regional Order Neural Network
DisEMBL™
GLOBPLOT 2
80
FER1_ARATH/1-148
FER_BRANA/1-96
FER2_ARATH/1-148
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’)

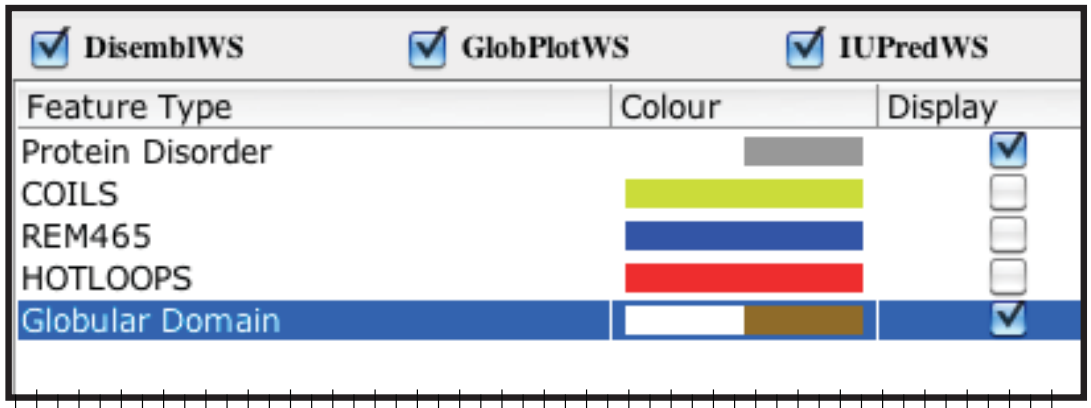
- Programs provided by JABAWS 2 employ
  - Machine learning methods (**DisEMBL**)
  - Similarity to disordered sequences (**RONN**)
  - Empirical amino acid statistics (**IUPred**, **GlobPlot**)

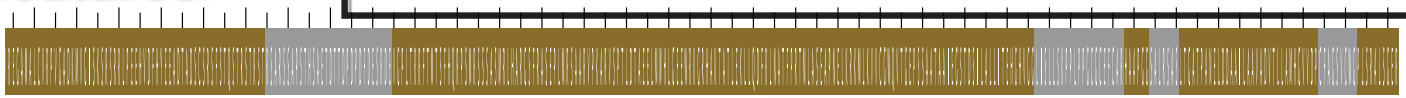
Disorder Predictions from JABAWS

JABAWS Analysis Service

Jalview JABAWS 2.0 Client

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



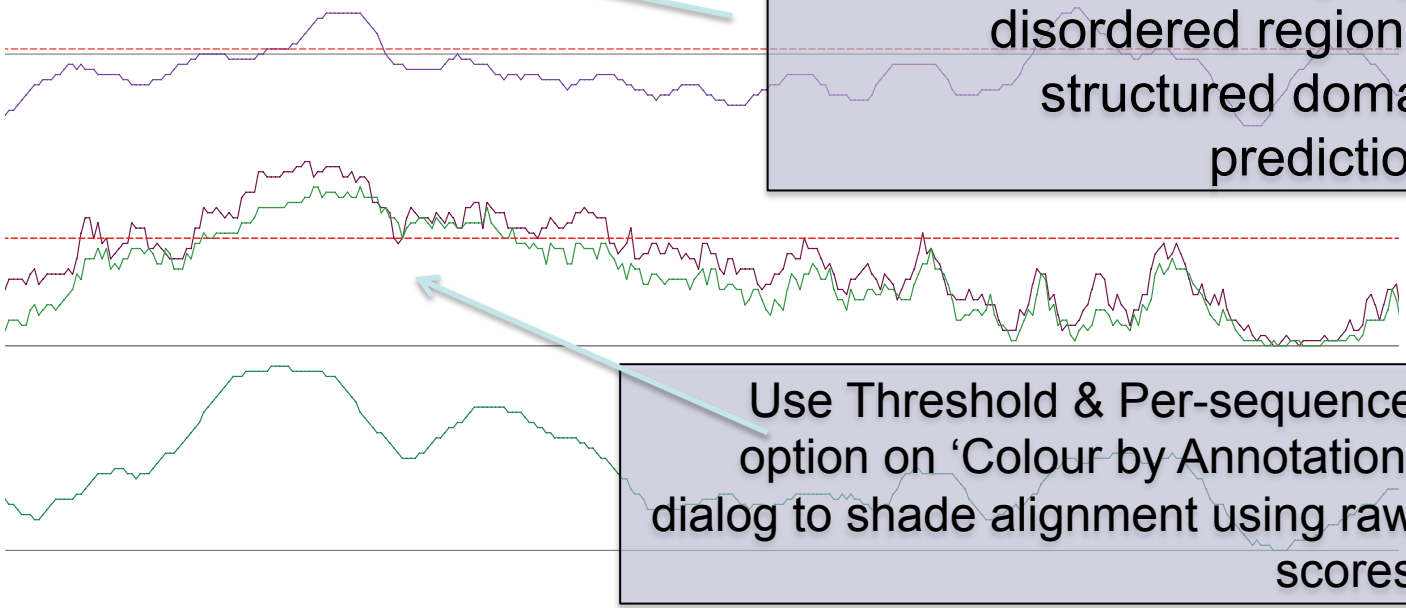


GlobPlotWS (Dydx)

IUPredWS (Short)

IUPredWS (Long)

JronnWS (JRonn)



Features highlight disordered region or structured domain predictions

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

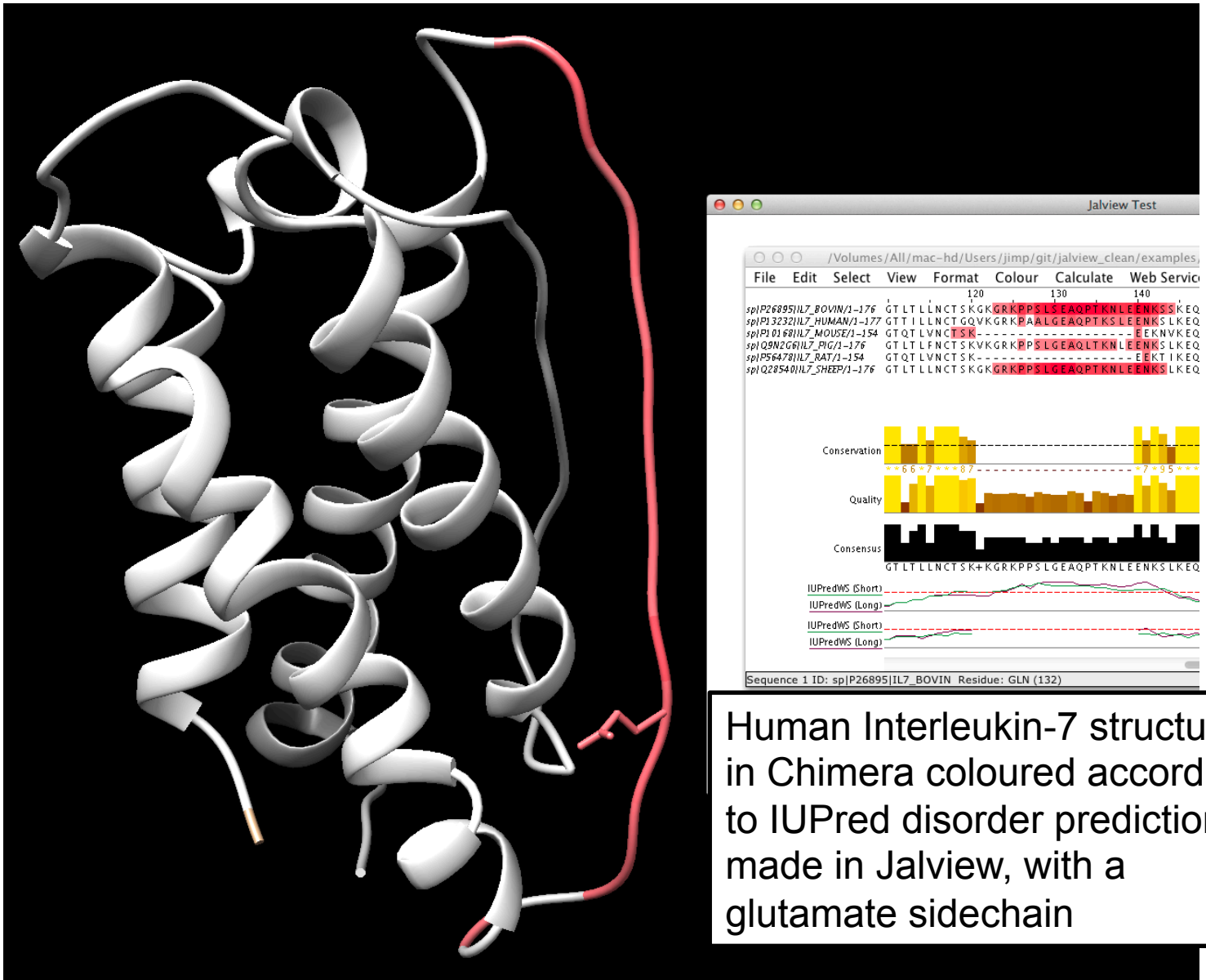
Exercise 27 – IUPred and Disembl Disorder predictions

- 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

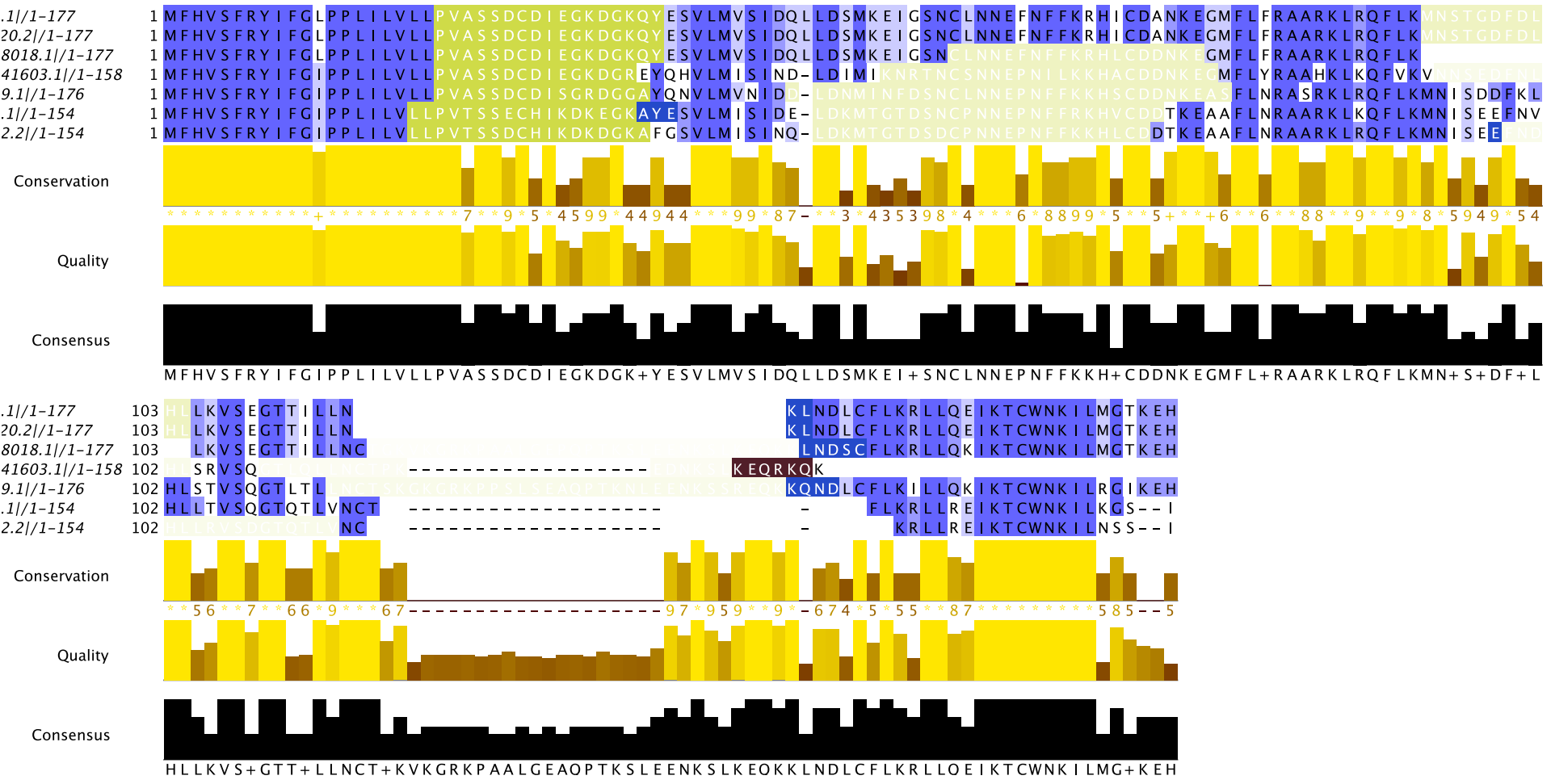
- Task
  - Apply the Disembl and IUPred disorder predictors to a set of protein sequences
  - Compare disordered predictions and available 3D data
- Question
  - How do you decide which regions are ‘significantly disordered’ ?

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 ?