

Session 2: 3D Structure Viewing and Structure Prediction Methods

<http://www.jalview.org/tutorial/training-materials/2020/Edinburgh/>



The image shows a screenshot of a 'Jalview Installer' window. The window title bar reads 'Jalview Installer'. The main content area features the Jalview logo (a colorful geometric shape) and the word 'Jalview' in a large font. To the right, there is a blue arrow pointing right with the word 'Training' inside it. Below this, a calendar-style box shows 'Wed Apr 29 2020'. To the right of the calendar box, the text 'Online 1.5 days' is displayed. At the bottom of the window, there are logos for 'THE UNIVERSITY of EDINBURGH' and 'edinburgh genomics.'.



THE UNIVERSITY
of EDINBURGH

edinburgh
genomics.



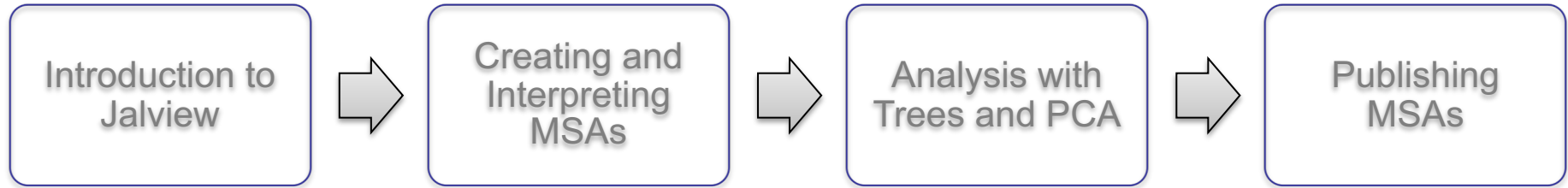
University of Dundee

supported by
wellcome trust

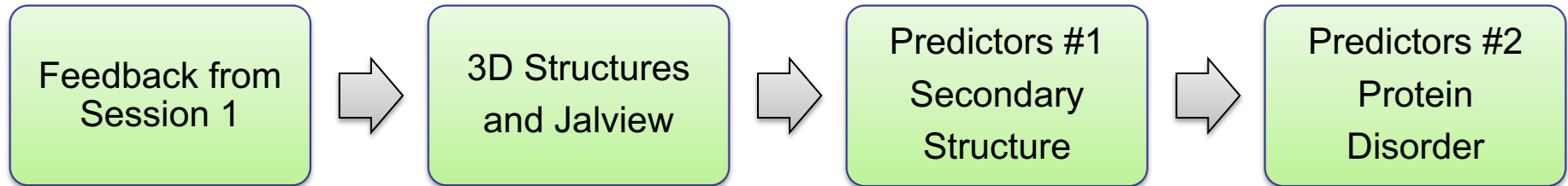


Day 2: Schedule

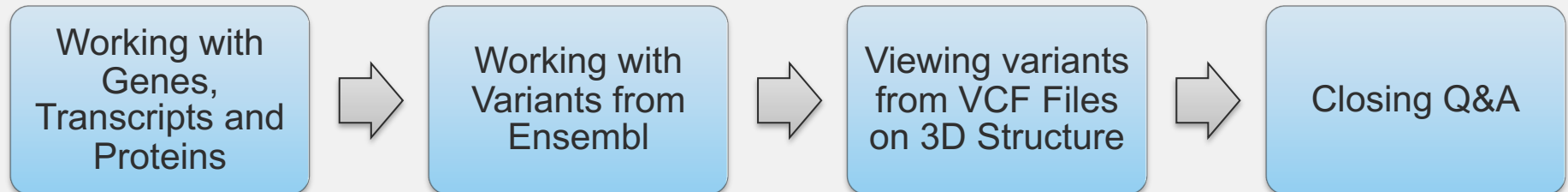
Wednesday: Core Jalview: Multiple alignment creation and MSA figure generation

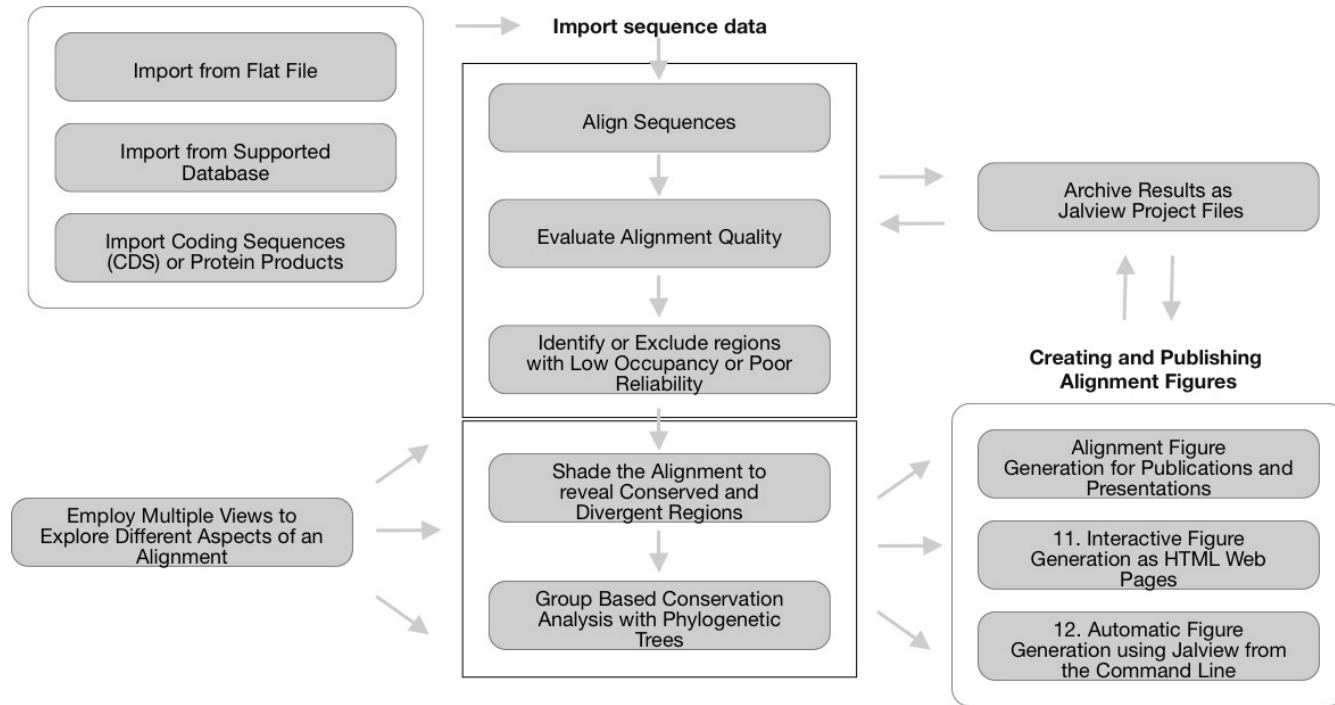


Thursday: 3D Structures in Jalview, and sequence and MSA based structure prediction



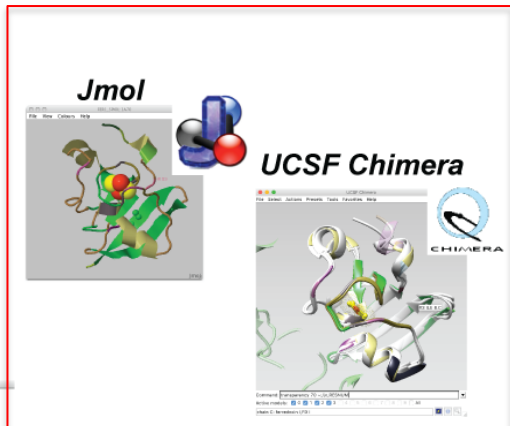
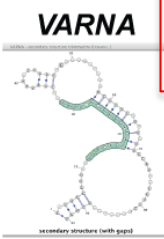
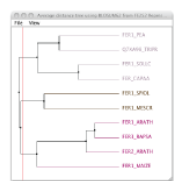
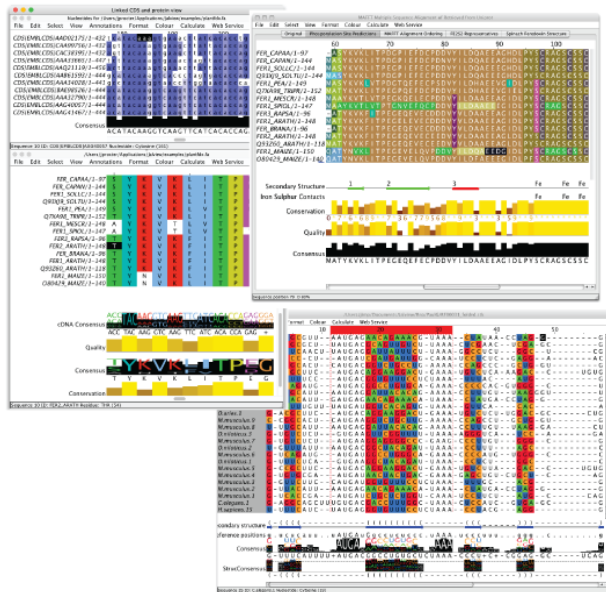
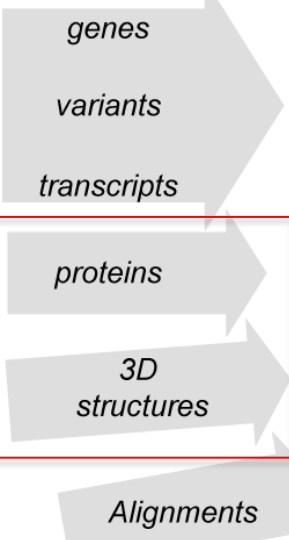
Friday: From Genome to Function – SNPs on transcripts, proteins and 3D structure





COMMENTS AND QUESTIONS FROM SESSION 1 ?

Biological Data retrieval



Interoperation with visualisation tools for Trees, 2D & 3D Structure.

Figure Export
 Interactive HTML
 BioJS msviewer
 EPS, PNG, SVG

Web based Analysis Services

JABAWS^{2.2}

Multiple Alignment

Alignment Analysis

Secondary Structure

ViennaRNA

Protein Disorder

3D Structure & Prediction

Exercises in The Jalview Tutorial

- Basics
- Selecting and Editing Sequences
- Colouring Sequences and Figure Generation
- Annotation and Features
- Multiple Sequence Alignment
- Analysis of Alignments
- Working with 3D structures**
- Protein sequence analysis and structure prediction**
- DNA and RNA Sequences

Working with 3D Structures

Superimposing Structures

Secondary Structure Prediction

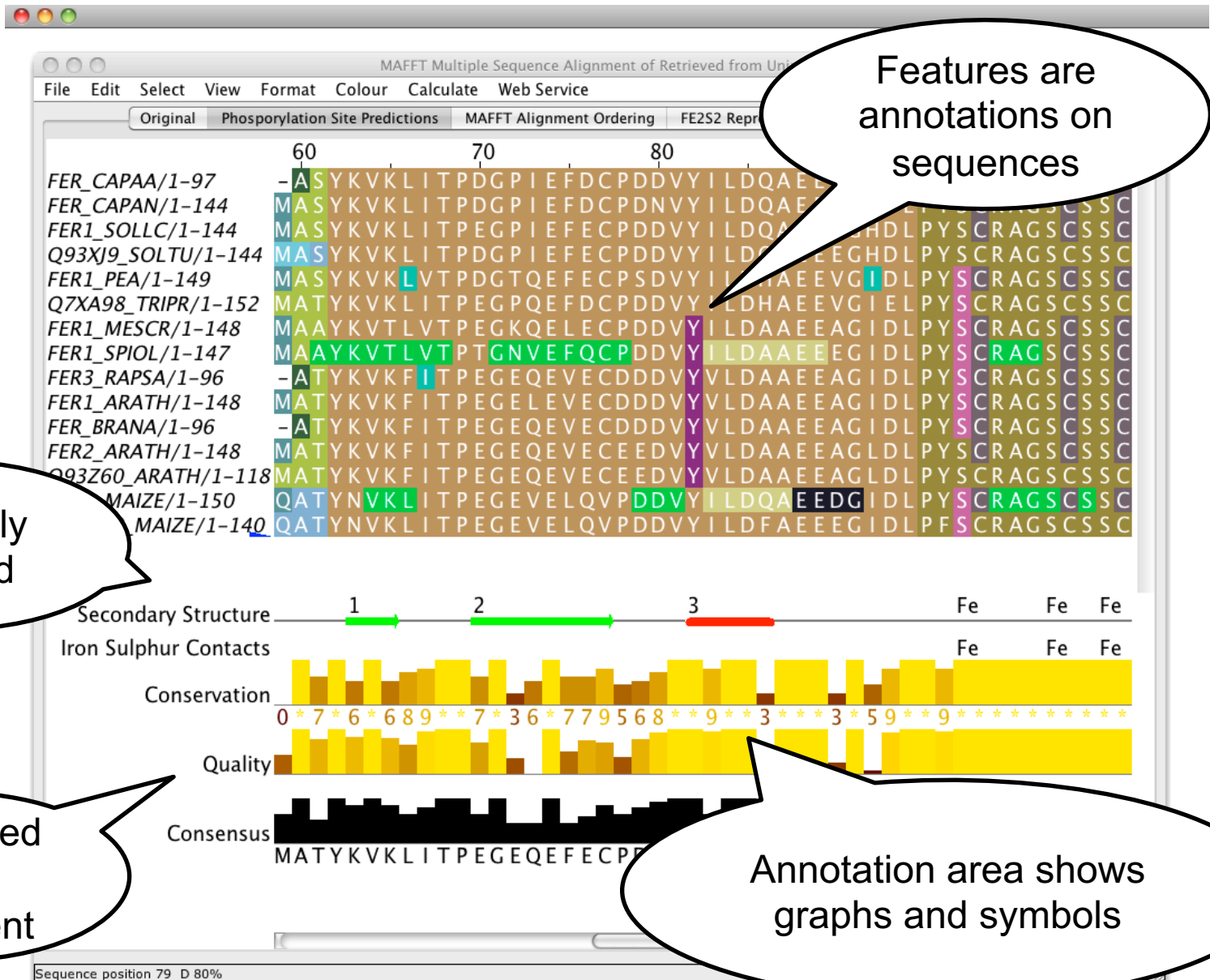
Disorder Prediction

Session 6:

3D structures in Jalview

Learning objective:

- Viewers supported by Jalview
- Configuring default structure viewer
- Methods of importing structures in Jalview
- Structure to Sequence mapping and methods
- Structure coloring and annotations
- Superimposing structures with MSAs



Features are annotations on sequences

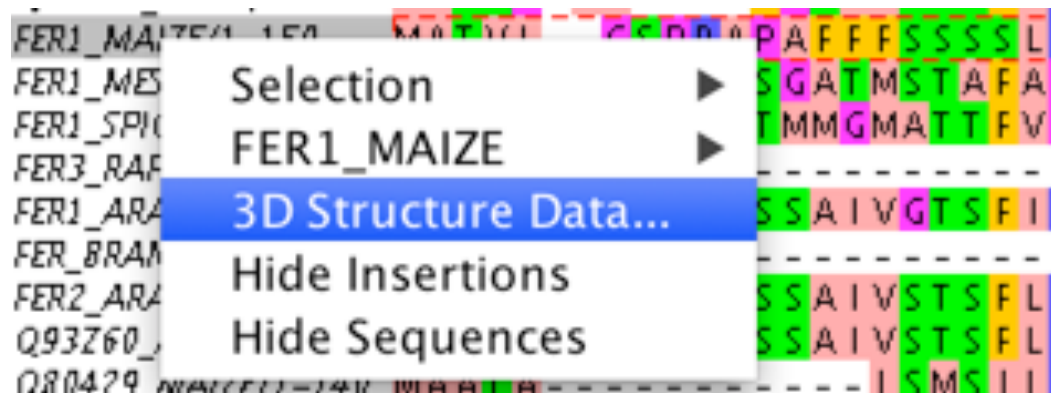
Manually created

Calculated from alignment

Annotation area shows graphs and symbols

Associating structures with sequences

- Structures in the PDB database
 - Auto-discovered using database references
 - Provide PDB ID (and chain) for sequence
- Local PDB file
 - Drag and drop to map files & sequences by ID
 - Attach PDB file to sequence manually

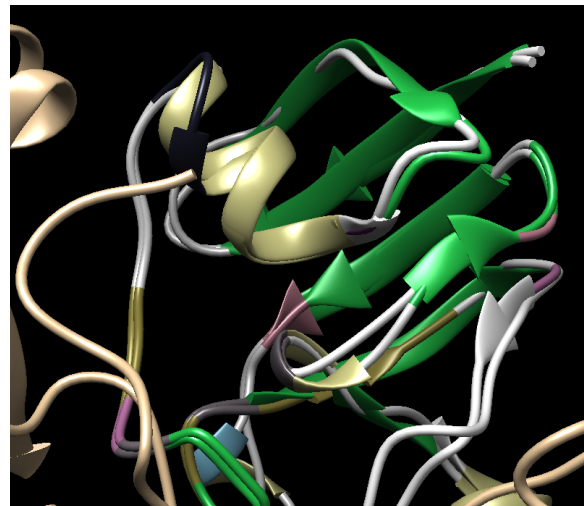
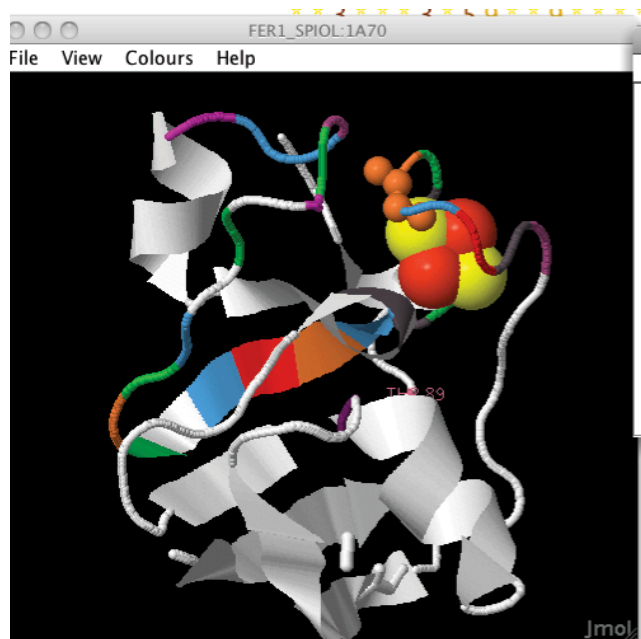


3D Structure Visualization in the Jalview Desktop



jmol.sourceforge.net

Integrated with Jalview



UCSF CHIMERA

an Extensible Molecular Modeling System

www.rbvi.ucsf.edu

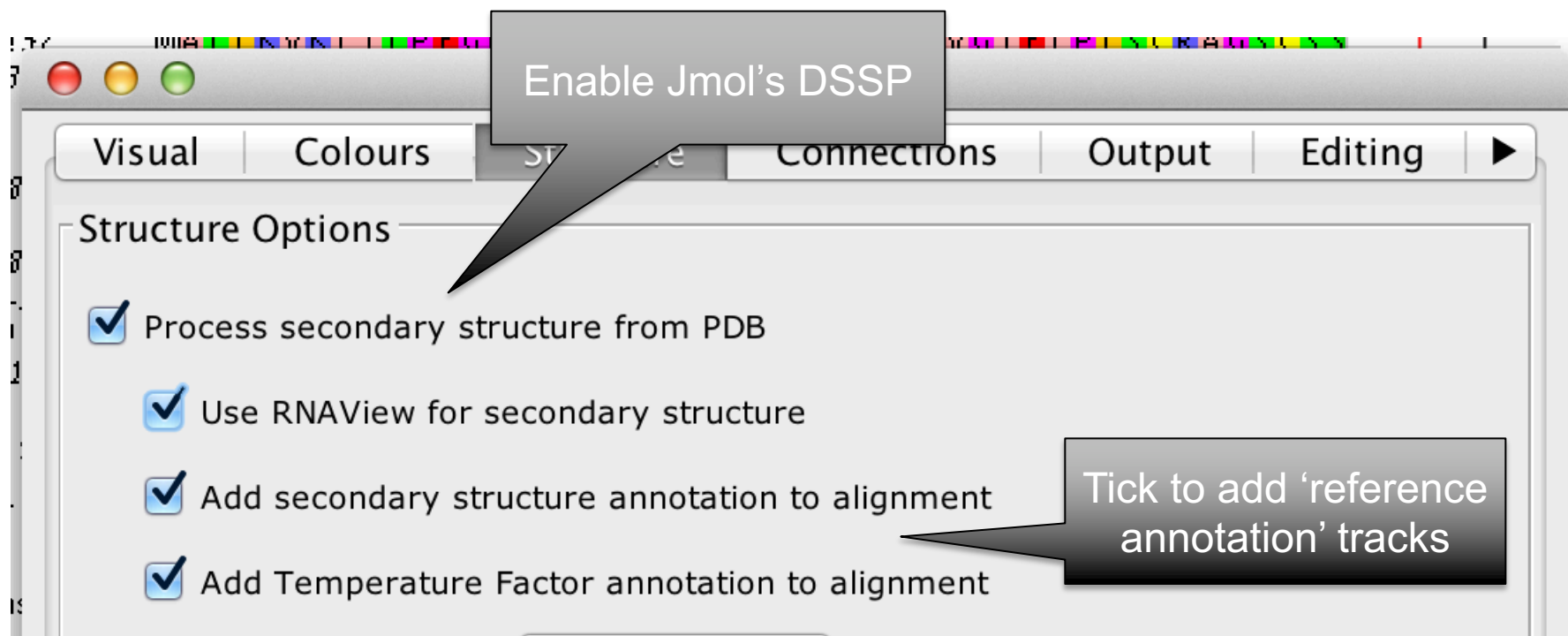
External program called
by Jalview

Which viewer to use for the exercises ?

- Both support:
 - Colour by alignment
 - Superposition
 - Mouseover highlighting
 - Saving to Jalview Projects
- Jmol is easier if you are new to structure
- If you have Chimera – see Exercise 24!

Secondary structure annotation tracks from 3D Structure

- Jalview 'Structure' preferences tab



Protein Structures in Jalview

Exercise 22

Section 7.3
Page 71

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

Remember to configure secondary structure preferences !

Protein Structures in Jalview

Exercise 22

Section 7.3
Page 71

- Task
 - Discover PDB structures for ferredoxin sequence(s)
 - Save and load structures and manipulate colouring
- Questions
 - What happens if there are lots of structures ?
 - How does Jalview match up sequence data to structural data
 - What annotation & features come from structure ?

Choosing a structure to view

Selection criteria

Toggle to inverse auto-selected option

Show additional info from PDB

Structure Chooser - Filter time (641 milli secs)

Best Quality

Invert

Found Structures Summary

Customise Displayed Options

Ref Sequence	PDB Id	Title	Experimental M...	Resolution
FER1_MAIZE	5h92	Crystal structure of the complex ...	X-ray diffraction	2.080
FER1_MAIZE	3b2f	Maize Ferredoxin 1	X-ray diffraction	1.700
FER1_MAIZE	5h8y	Crystal structure of the complex ...	X-ray diffraction	2.200
FER1_MAIZE	3w5u	Cross-linked complex between F...	X-ray diffraction	2.700
FER1_MAIZE	1gaq	CRYSTAL STRUCTURE OF THE C...	X-ray diffraction	2.590
FER1_MAIZE	3w5v	Cross-linked complex between F...	X-ray diffraction	3.810

Create New 3D View for selected structure(s) or Add to existing 3D view

Jmol view for FER1_SPIOL-1A70

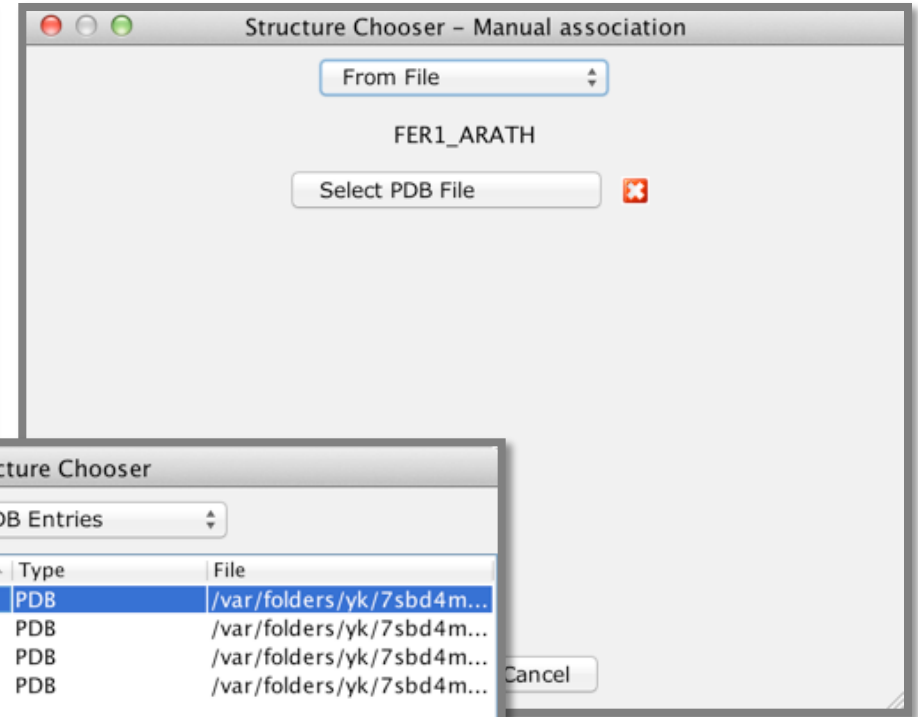
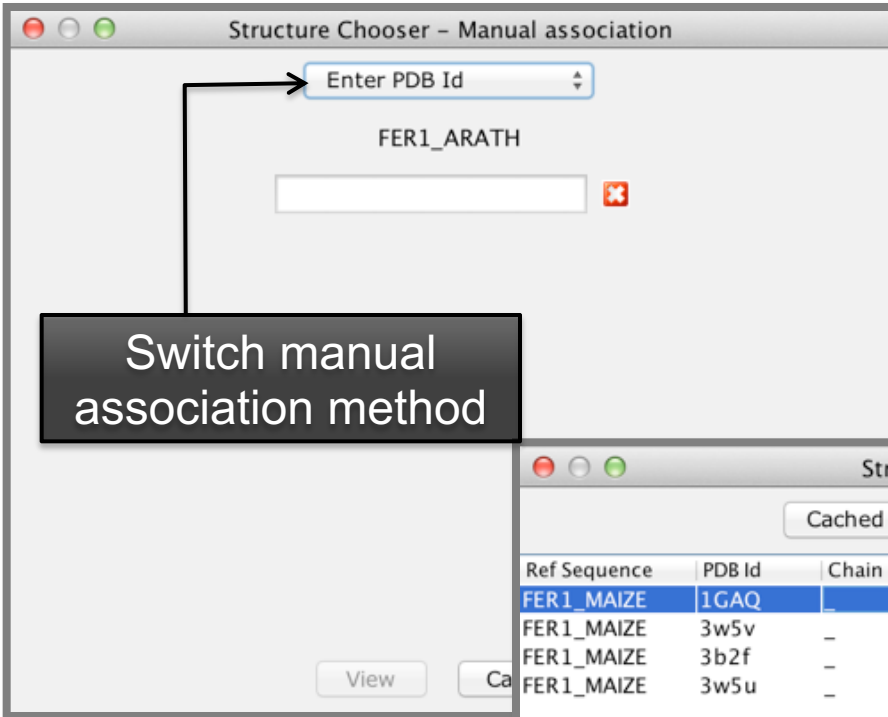
Add

Superpose Structures

New View

Cancel

Structure unavailable?



Switch manual association method

Structure Chooser

Cached PDB Entries

Ref Sequence	PDB Id	Chain	Type	File
FER1_MAIZE	1GAQ	-	PDB	/var/folders/yk/7sbd4m...
FER1_MAIZE	3w5v	-	PDB	/var/folders/yk/7sbd4m...
FER1_MAIZE	3b2f	-	PDB	/var/folders/yk/7sbd4m...
FER1_MAIZE	3w5u	-	PDB	/var/folders/yk/7sbd4m...

View

Cancel

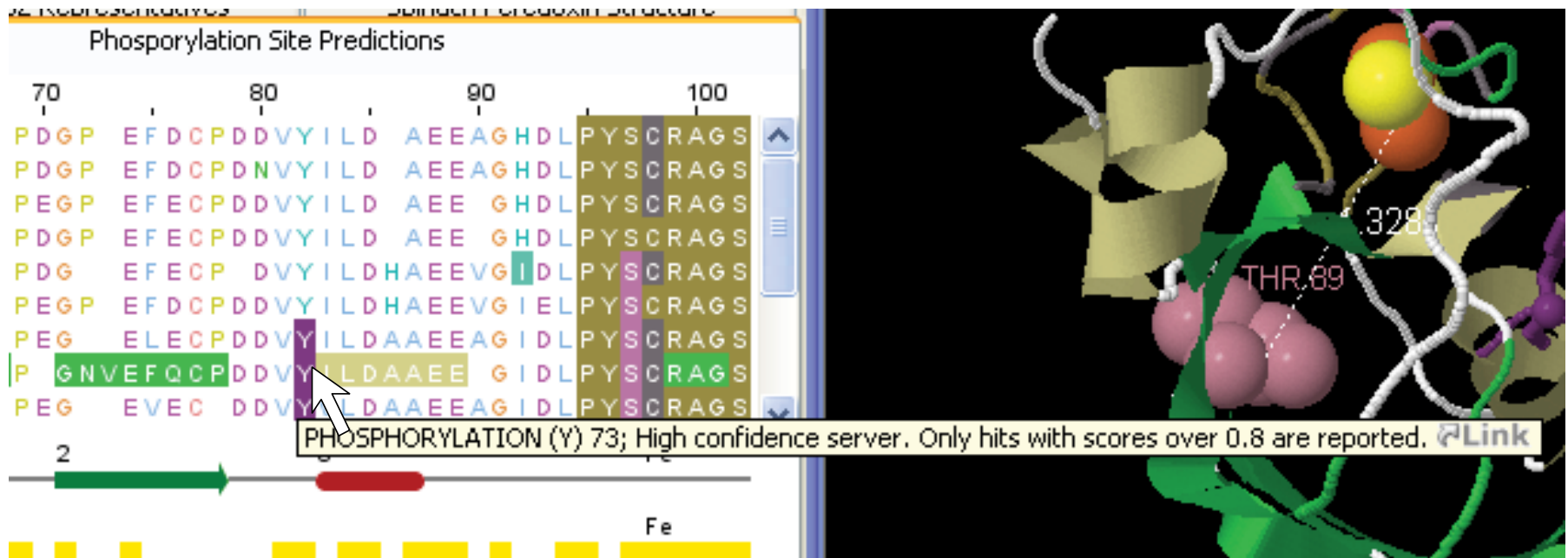
Protein Structures in Jalview

Exercise 22

Section 7.3
Page 71

- Task
 - Discover PDB structures for ferredoxin sequence(s)
 - Save and load structures and manipulate colouring
- Questions
 - What happens if there are lots of structures ?
 - How does Jalview match up sequence data to structural data
 - What annotation & features come from structure ?

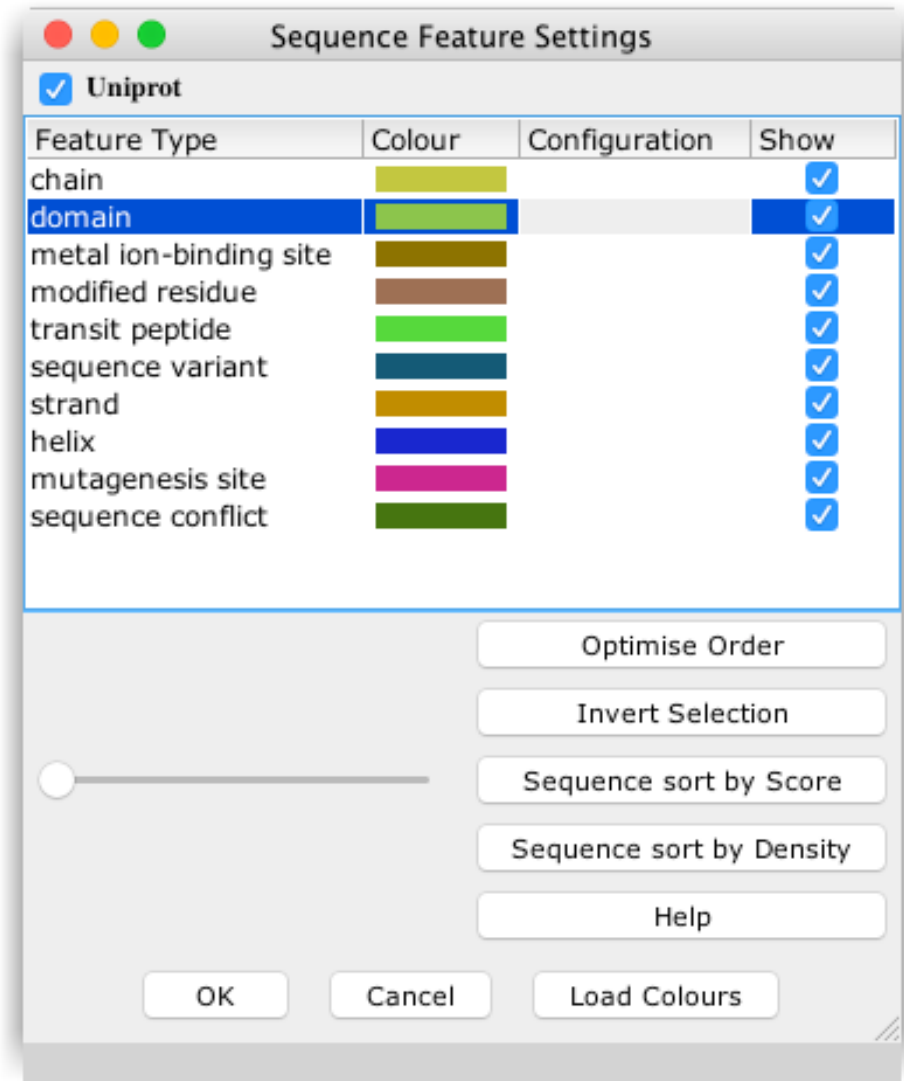
Sequence Features are visualised over the alignment



Customising Feature Display

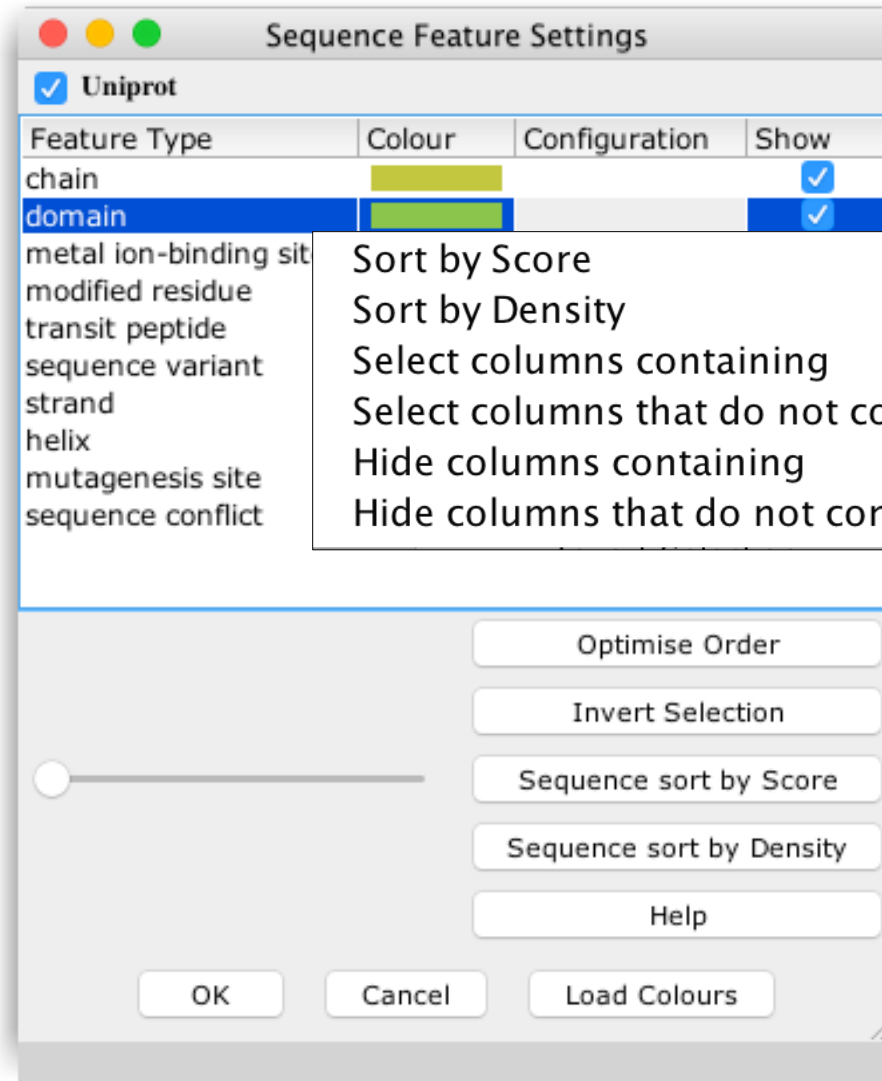
View => Feature Settings...

- Drag features to reorder on alignment
 - Or make transparent
- Change colour
- Enable/disable by name and group



Customising Feature Display

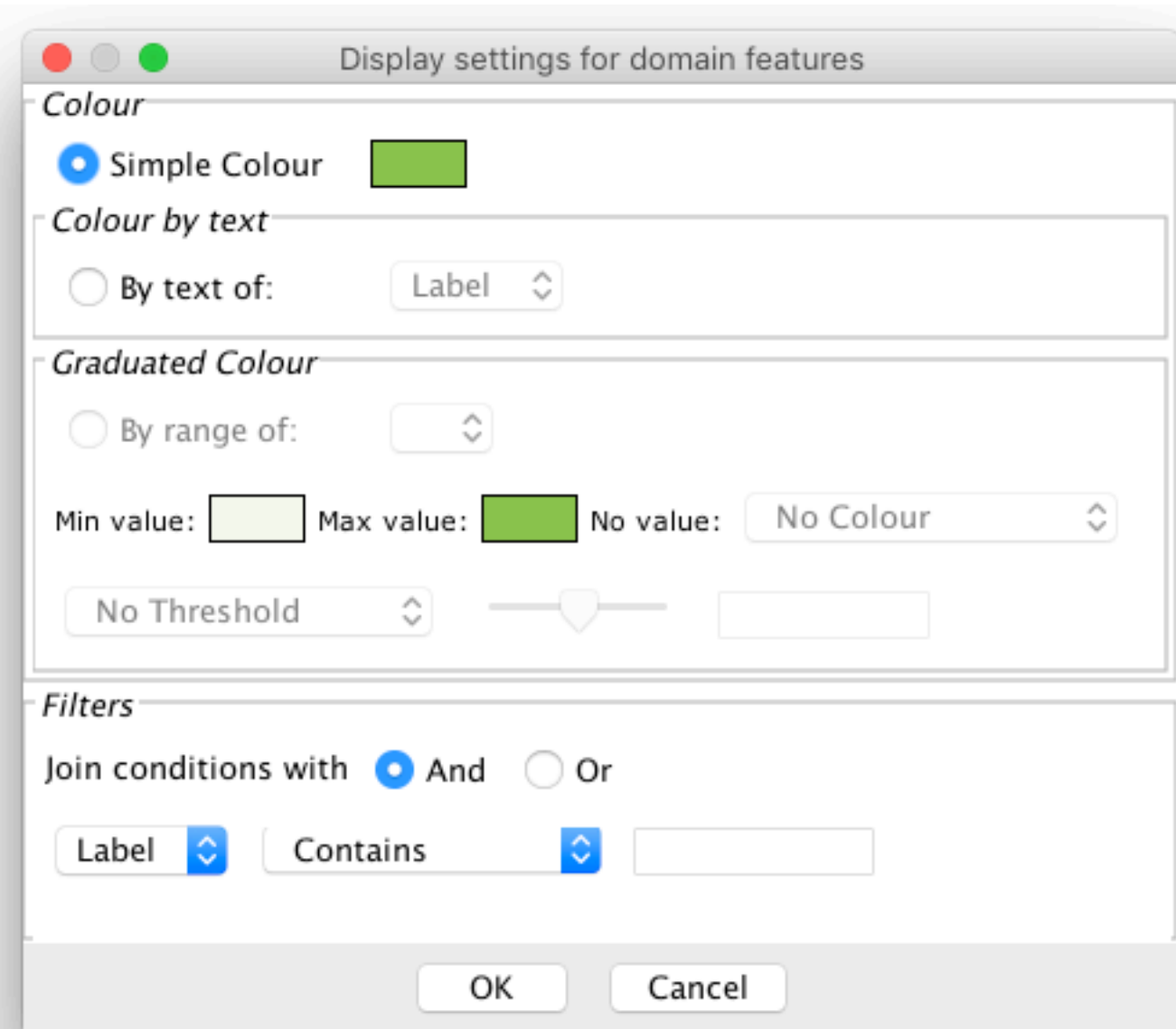
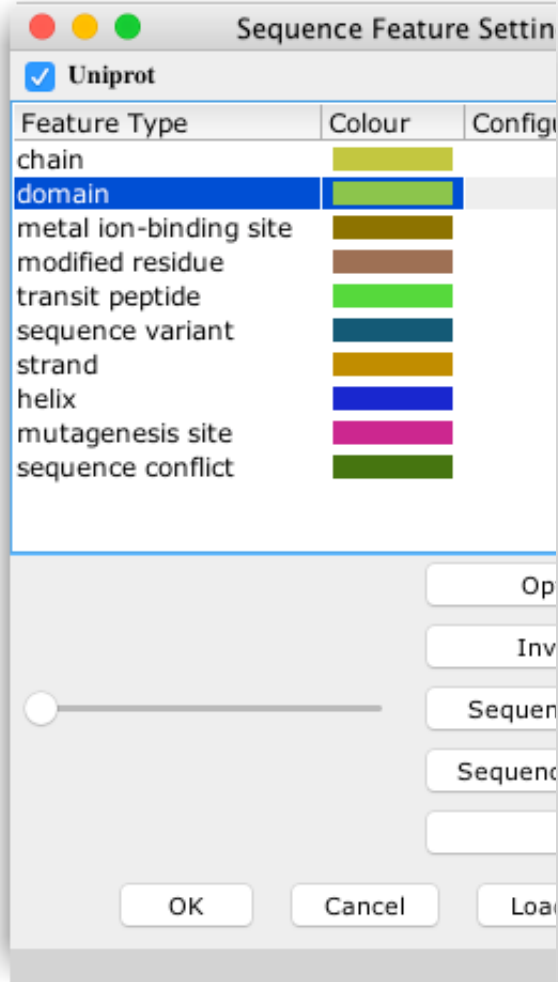
Right-click on feature => Popup menu



**Select
or Hide
columns**

Customising Feature Display

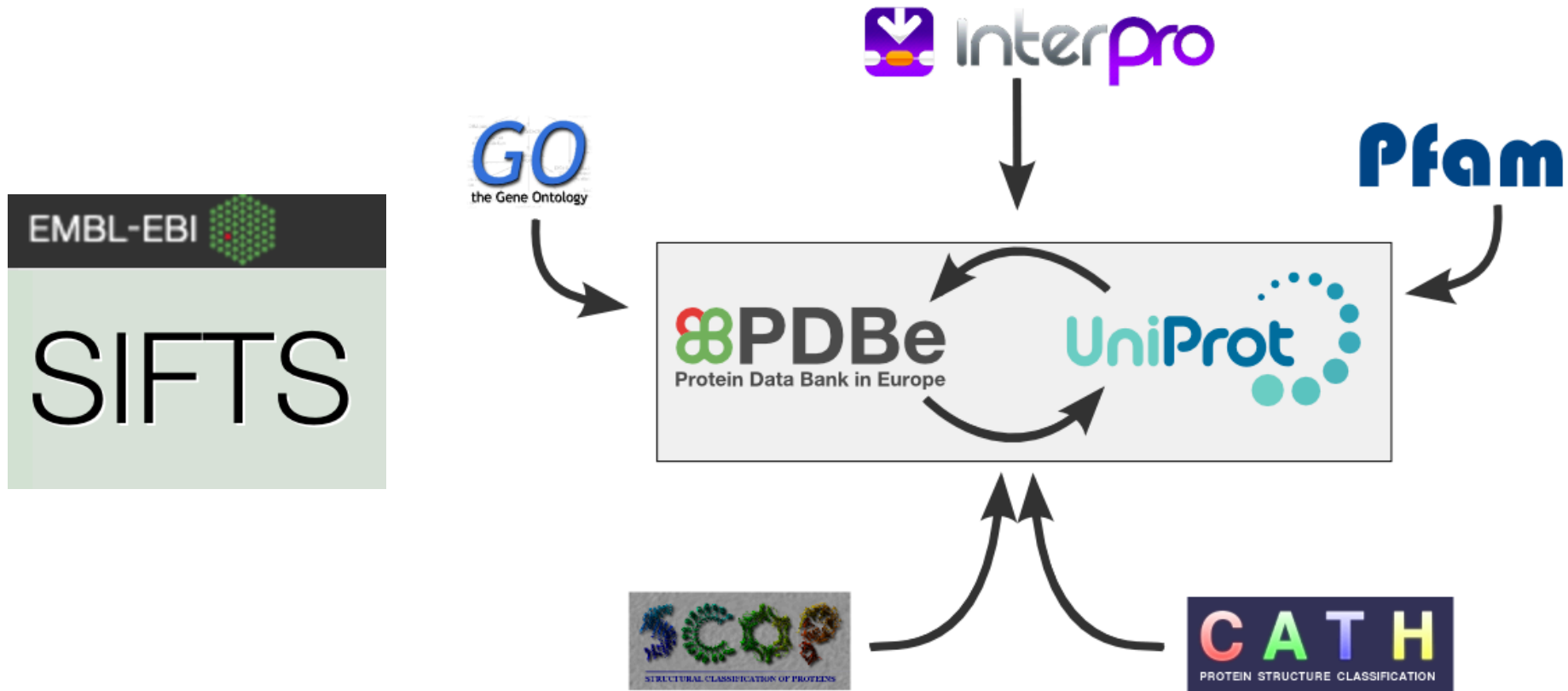
Complex Colour Schemes & filters



Sources of sequence features

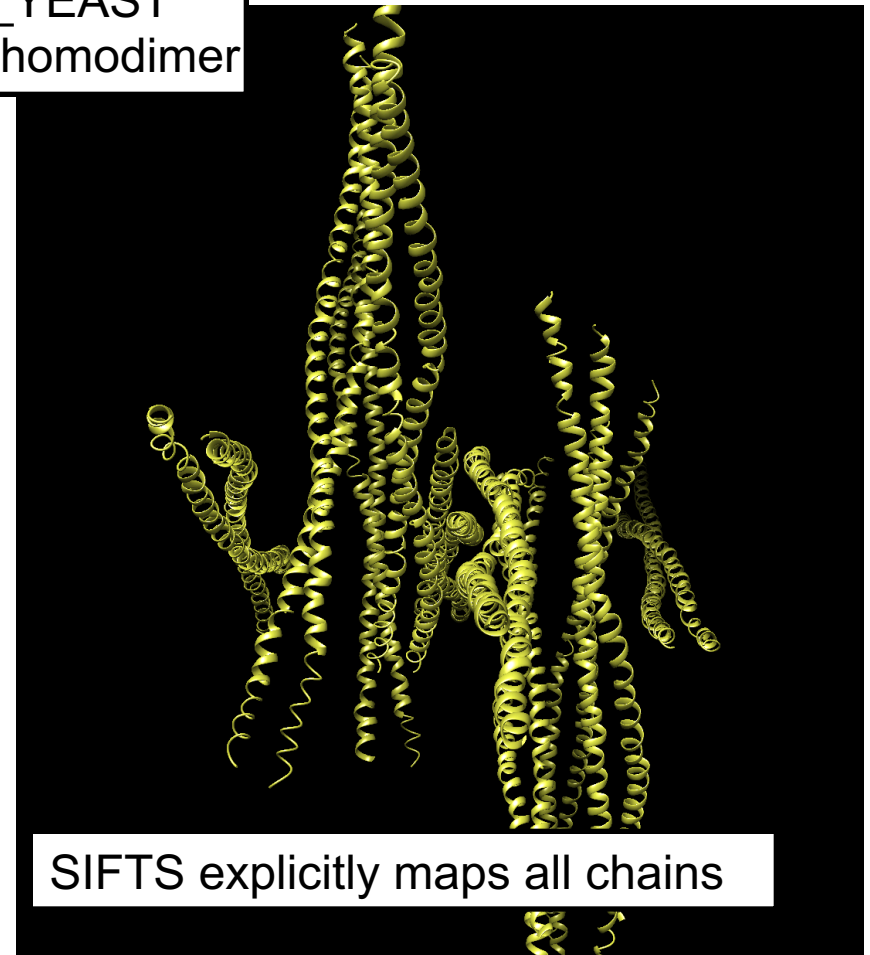
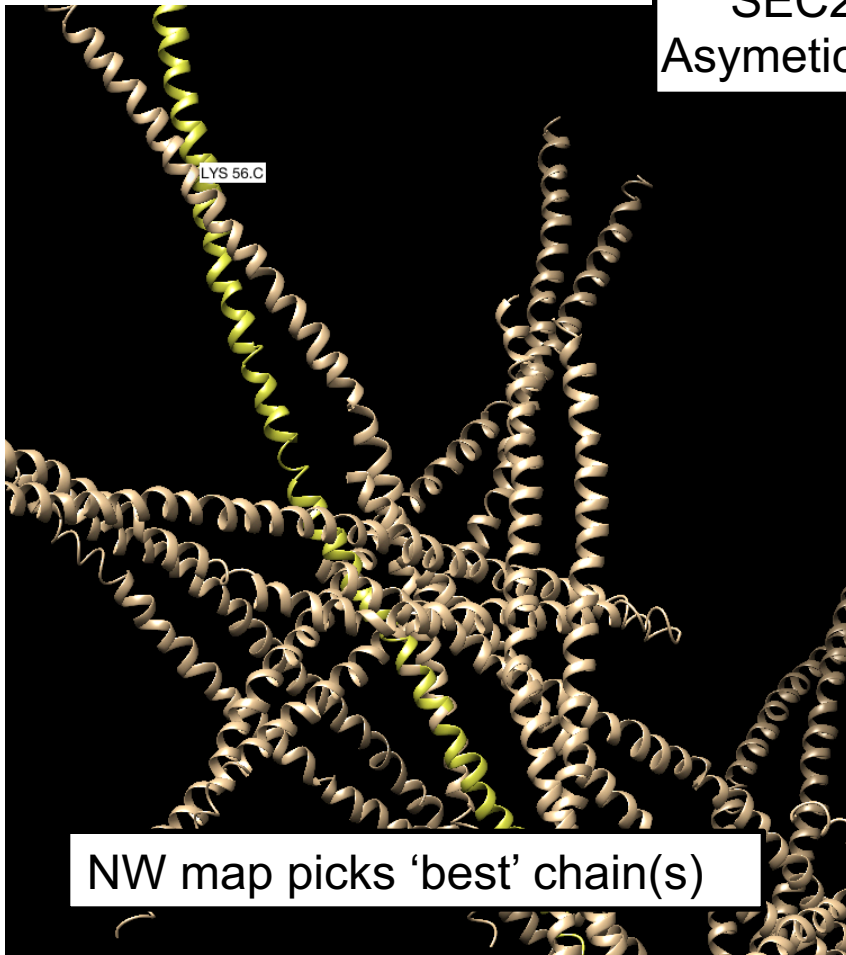
- **Sequence databases**
- 3D Structure associations
- Stockholm & AMSA Alignment files
- **GFF2/3 (*Generalised Feature Format*)**
- **Jalview format feature files**
 - Like GFF
 - can create GFF in Excel
 - Allows scores to be added

Protein sequence -> 3D Structure



2.10 – Accurate Uniprot <> PDB mapping for multimers

SEC2_YEAST
Asymetic homodimer



Protein Structures in Jalview

Exercise 22

Section 7.3
Page 71

- Task
 - Discover PDB structures for ferredoxin sequence(s)
 - Save and load structures and manipulate colouring
- Questions
 - What happens if there are lots of structures ?
 - How does Jalview match up sequence data to structural data
 - What annotation & features come from structure ?



File Edit Select View Annotations Format Colour Calculate Web Service

3W5V|A/18-314 18 KCSKKQEEGVVTNLYKPKEPYVGRCLLNTKITGDDAPGETWHMVFSTEGKIP 69
3W5V|B/1-96 1 ATYNVKLITPEGEVELQVPDDVYILDQAEEDGIDLPHYSCRAGSCSSCAGKVV 52

3W5V|A/18-314 70 YREGQSIGVIADGVDKNGKPHKVRLYSIASSAIGDFGDSKTVSLCVKRLIYT 121
3W5V|B/1-96 53 SGSVDQSDQSYLDDGQICDGWVLTCHAYPTSDVVIETHKEEELT 96

3W5V|A/18-314 122 NDAGEIVKGVCSNFLCDLQPGDNVQITGPVGKEMLPKDPNATIIMLATGTG 173
3W5V|B/1-96

3W5V|A/18-314 174 IAPFRSFLWKMF FEKHDDYKFENGLGWFLFGVPTSS SLLYKEEFGKMKERAPE 225
3W5V|B/1-96

3W5V|A/18-314 226 NFRVDYAVSREQTNAAGERMYIQTRMAEYKEELWELLKKDNTYVYMCGLKGM 277
3W5V|B/1-96

3W5V|A/18-314 278 EKGIDDIMVSLAEKDGIDWFDYKKQLKRGDQWNVEVY 314
3W5V|B/1-96

10 20 30 40

3W5V|A/18-314 K C S K K O F E G W T N L Y K P K E P Y W G R C L L N T K I T G D D A P G F T W H M V F
3W5V|B/1-9 R A G S C S

- Selection ▶
- 3W5V|A ▶
- 3D Structure Data...
- Hide Insertions
- Hide Sequences

- Sequence Details
- Show annotations ▶
- Hide annotations ▶
- Add reference annotations
- Edit ▶
- Output /Temperature Factor
- Create Sequence Feature...
- Create Group
- Edit New Group ▶

Add annotations for
jalview.ext.jmol.JmolParser/Secondary Str

Conservation



Copied sequences- from /Users/jprocter/git/jalview/examples/3W5V.pdb

File Edit Select View Annotations Format Colour Calculate Web Service

3W5V A/18-314	18 KCSKKQEEGVVTLNLYKPKPEYVGRCLLNTKITGDDAPGETWHMVFSTEGKIP	69
3W5V B/1-96	1 ATYNVKLITPEGEVELQVPDDVYILDQAEEDGIDLPHYSCRAGSCSSCAGKVV	52

Temperature Factor

Secondary Structure



Temperature Factor

Secondary Structure



3W5V A/18-314	70 YREGQSIGVIADGVDKNGKPKHVRLYSIASSAIGDFGDSKTVSLCVKRLIYT	121
3W5V B/1-96	53 SGSVDQSDQSYLDDGQICDGWVLTCHAYPTSDVVIETHKEEELT	96

Temperature Factor

Secondary Structure



Temperature Factor

Secondary Structure



3W5V A/18-314	122 NDAGEIVKGVCSNFLCDLQPGDNVQITGPVGKEMLPKDPNATIIMLATGTG	173
---------------	---	-----

Exercise 23 Superposing Structures using Alignments

Section 7.3
Page 72

- Task
 - Align structures using the ferredoxin alignment
Typo: For FER1_MAIZE structure use 1GAQ
 - Experiment with views to control what part of the alignment is used to superimpose the structures

```
LIVNGGIIAPSFDDP.MDEKAREILQKLFPEHEVVMA.PGRELLL...L...
YIANGGIIAPQFGDPIRDKEAIRVLSDTFPHHSVVG IENAREIVL.A...
LILNNRVFVVPVNGPASVDNDALNVYKTAMPGYEIIIGVKGASGTPWL.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

Exercise 23 Superposing Structures using Alignments

Section 7.3
Page 72

- Task
 - Align structures using the ferredoxin alignment
Typo: For FER1_MAIZE structure use 1GAQ
 - Experiment with views to control what part of the alignment is used to superimpose the structures

Exercise 23 Superposing Structures using Alignments

Section 7.3
Page 72

- Task
 - Align structures using the ferredoxin alignment
- **Typo: For FER1_MAIZE structure use 1GAQ**
 - 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 ?

For Chimera Users (Ex 24)

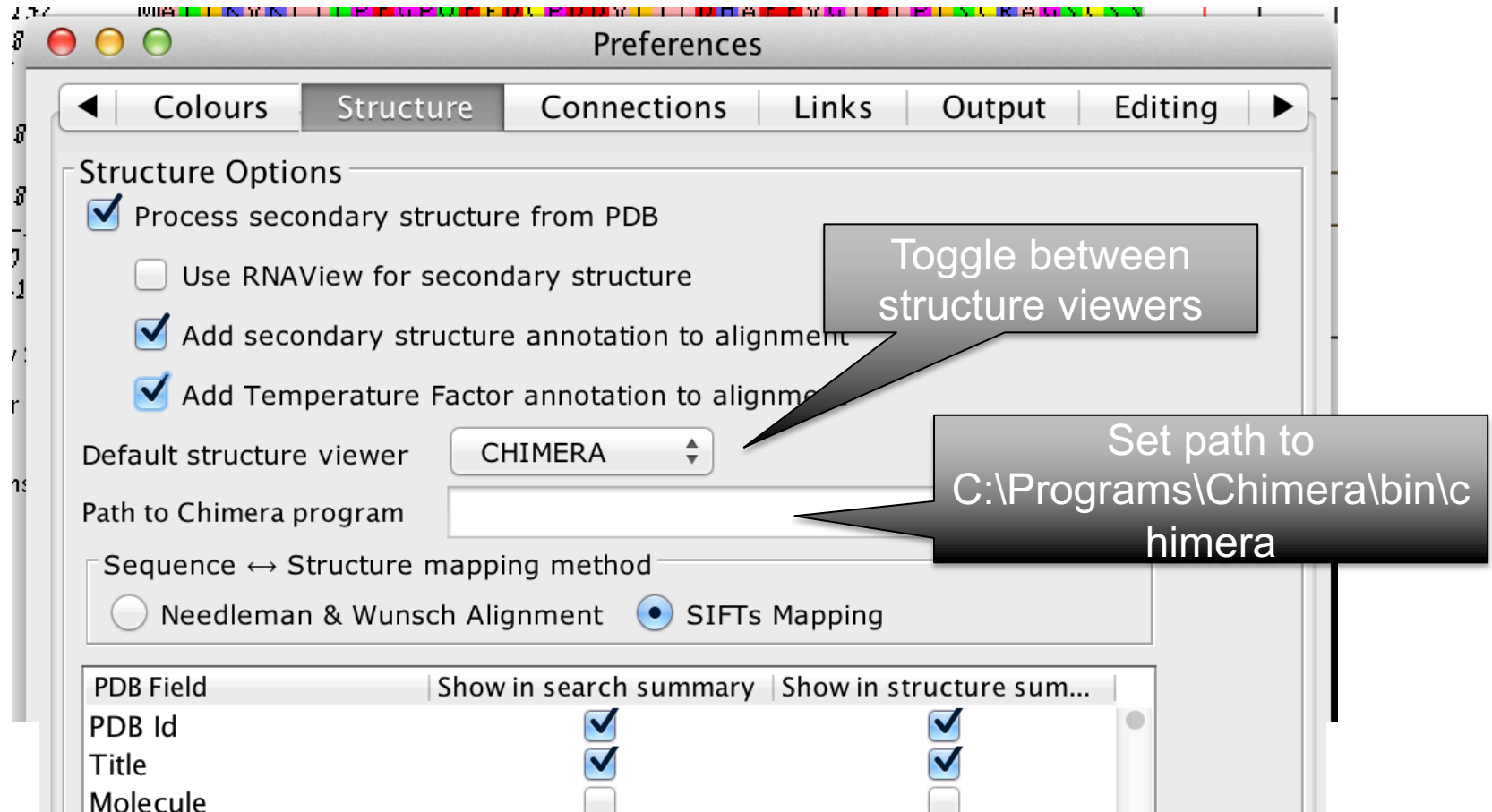


UCSF CHIMERA

an Extensible Molecular Modeling System

Configure Jalview and Chimera

- Jalview 'Structure' preferences tab



Chimera selection highlighting

UCSF Chimera

File Select Actions Presets Tools Favorites Help

Command: `select :FES z <2.9`

Active models: 0 1

/Users/jprocter/git/jalview/examples/plantfdx.fa

Select View Annotations Format Colour Calculate Web Service

```
100      110      120      130
Y SCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLT CVAY
Y SCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQLEEGWVLT CVAY
Y SCRAGSCSSCAGKV TAGSVDQSDGNFLDEDQEAAGFVLT CVAY
Y SCRAGSCSSCAGKV TAGTVDQSDGKFLDDDQEAAGFVLT CVAY
Y SCRAGSCSSCAGKV VGG EVDQSDGSFLDDEQIEAGFVLT CVAY
Y SCRAGSCSSCAGKV VNGN VNQEDGSFLDDEQIEGGWVLT CVAF
Y SCRAGSCSSCAGKV TSGS VNQDDGSFLDDDQIEKGWVLT CVAY
Y SCRAGSCSSCAGK LKTGSLNQQDDQSFLDDDQIDEGWVLT CAAY
Y SCRAGSCSSCAGK VVSGSVDQSDQSFLDDDQIEAGFVLT CAAY
Y SCRAGSCSSCAGK VVSGSVDQSDQSFLDDEQIEGEGFVLT CAAY
Y SCRAGSCSSCAGK VVSGFVDQSDQSFLDDDQIEAGFVLT CAAY
Y SCRAGSCSSCAGK VVSGSIDQSDQSFLDDEQMS EGYVLT CVAY
Y SCRAGSCSSCAGK VVSGSIDQSDQSFLDD-----
Y SCRAGSCSSCAGK VVSGSVDQSDQSYLDDGQIADGWVLT CHAY
F SCRAGSCSSCAGK VVSGSVDQSDQSFLNDNQVADGWVLT CAAY
```

q*****q37*398*6*459* 100323222132

CAPAA Residue: GLU (71)

Command: `select :FES z <2.9`

Active models: 0 1

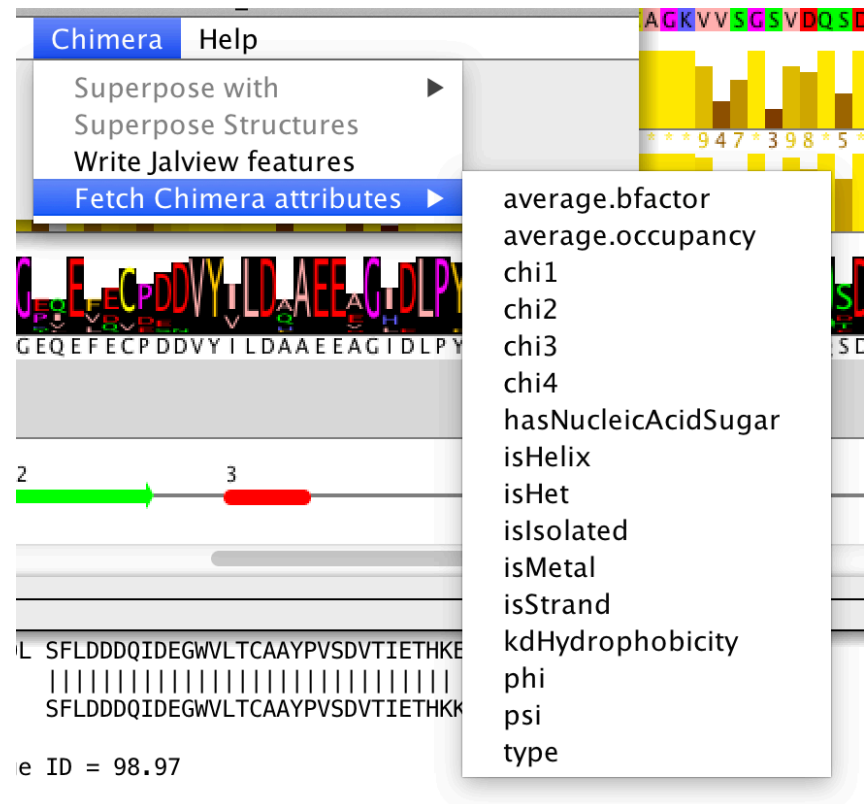
Selection in chimera

Selection highlighted in Jalview
Press B to select columns!



Pass annotation between Jalview and Chimera

- Write jalview features
 - Sequence Feature added to Chimera's Attribute set
- Fetch Attributes
 - Import as a Jalview feature



Many MSAs : 1 Structure View

	10	20	30	40	50			
T1_HUMAN	---E NAFKRRR	EGVEEVDQPE	EGEKAKCD	---MVKFG	GSGRSKQ	ACQ	ERR	
DI_MOUSE	---RMFRRVG	EGDCAALVK	EDGVESTRLL	---GLPSD	---VASGLYV	KCE	RRR	
AWI_PPHIN	---S--KSRVVE	EGDEECNAPD	ENKGPHELD	---MKKFG	GPFLRRQ	TCR	NRK	
9WK_LAGLA	---F--VERSRK	EGVEGCGQTO	EDGEHRELR	---PFR	---PGLRRQ	CVG	RRK	
INME_DROSD	---S KKKRRR	EGVEGCGKQNE	EGEAFERN	---DKS	---	HQ	ICK	QRR
CL_MOUSE	---R--VRSARME	EGEASRRT	EDGEHDFERD	---MKKFG	GPNKIRQ	KER	LHQ	
MLL_BOVIN	---E NAFKRRR	EGVEEIQDPE	EGEKAKCD	---MVKFG	GSGRSKQ	ACQ	KRR	
974_DAMRE	---V--KVKRRS	EGEKCEVVR	TDEGTFDEMD	---KPKFG	GRNKRQ	KCR	LHQ	
DCC_DROSC	---R--RKRSTR	EGEETAVMT	EDGEKDFEKL	---MRKFG	GPNKIRQ	KER	KFR	
105_HUMAN	---I--NRQRNR	EGCAACLRRM	DEGRDFECD	---KPKFG	GSNQRQ	KCR	WRQ	
796_METZ	---G--SRRRRR	EGEGCPNVR	ENSGKNEHLN	---SK	---	LQ	SEI	YRR
AIS_PARLI	---S--SERKKR	EGVEEIQAPD	EGKCTACKD	---MIKFG	GSGRKQ	ACK	DRR	
20X_DAMRE	---P--PKKVVV	EGRWACQIT	VDEGAGVSRNG	I--HRLN	I--HGRVY	KCR	KRR	
90E_HUMAN	---R--RMFRVVG	EGCAACQIT	EDGEASSTELL	---GLPSD	---VASGLEC	KCE	RRR	
852_CARAU	---G--GVKRRR	EGVEEIQAPD	EGKESACKD	---MIKFG	GSGRSKQ	ACQ	KRR	
390_DAMRE	---K--KRKRSR	EGEGPQDVP	NDEVEFTELID	---KPKFG	GRNKRQ	KCR	WRQ	
07E_BOVIN	---I--NRQRNR	EGTCAACLRRM	DEGHDFECD	---KPKFG	GSNQRQ	KCR	WRQ	
20H_XENLA	---E--NCIKRRR	EGVEEVDQPD	EGEKACQD	---MLKFG	GAGTRQ	ACM	QRR	
9VZ_XENLA	---F--LRKSDS	EGMCGCCRCR	EDGECSVELD	---KMH-N	PMDI	KRE	KCL	LRK
U00_DROVA	---G--PBRRRTR	EGNCAAQDSD	EGTDFEMD	---MVKFG	GPGRKQ	TCM	WRQ	
10E_SCHIA	---F--PMMCKR	EGEELCEFRP	ENGDADRNV	---RVG	---	HSP	RCI	FRT
2QR2_XENTR	---Q--IKRSAR	EGEAGSRT	EDGEQDFERD	---MKKFG	GPNKIRQ	KER	LHQ	
XLL_HUMAN	---Q--IKRSAR	EGEAGSRT	EDGEQDFERD	---MKKFG	GP	---	---	
90K_SHEP	---E--NAFKRRR	EGVEEIQDPE	EGEKAKCD	---MVKFG	GSGRSKQ	ACQ	KRR	
104_TROAD	---L--SKKKRKA	EGAGCP	ELKENGSGKNEHLN	---RKT	---	GH	ICI	YRK
10E_DROSE	---F--PMMCKR	EGEELCEFRP	ENGDADRNV	---RVG	---	HSP	RCI	FRT
12AG_PIC	---G--RRRRVR	ERKCKACVOE	EGVHYERD	---MKKFG	GPGRKQ	SCV	LHQ	
10K_TETNG	---P--KRRRKR	EGVEAPERLL	NEGVESSERN	---RKT	---	GH	ICK	FRK
11Z_DROMO	---S--KRRRKR	EGVEGCGKQNE	EGEAFERN	---DKS	---	HQ	ICK	QRR

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

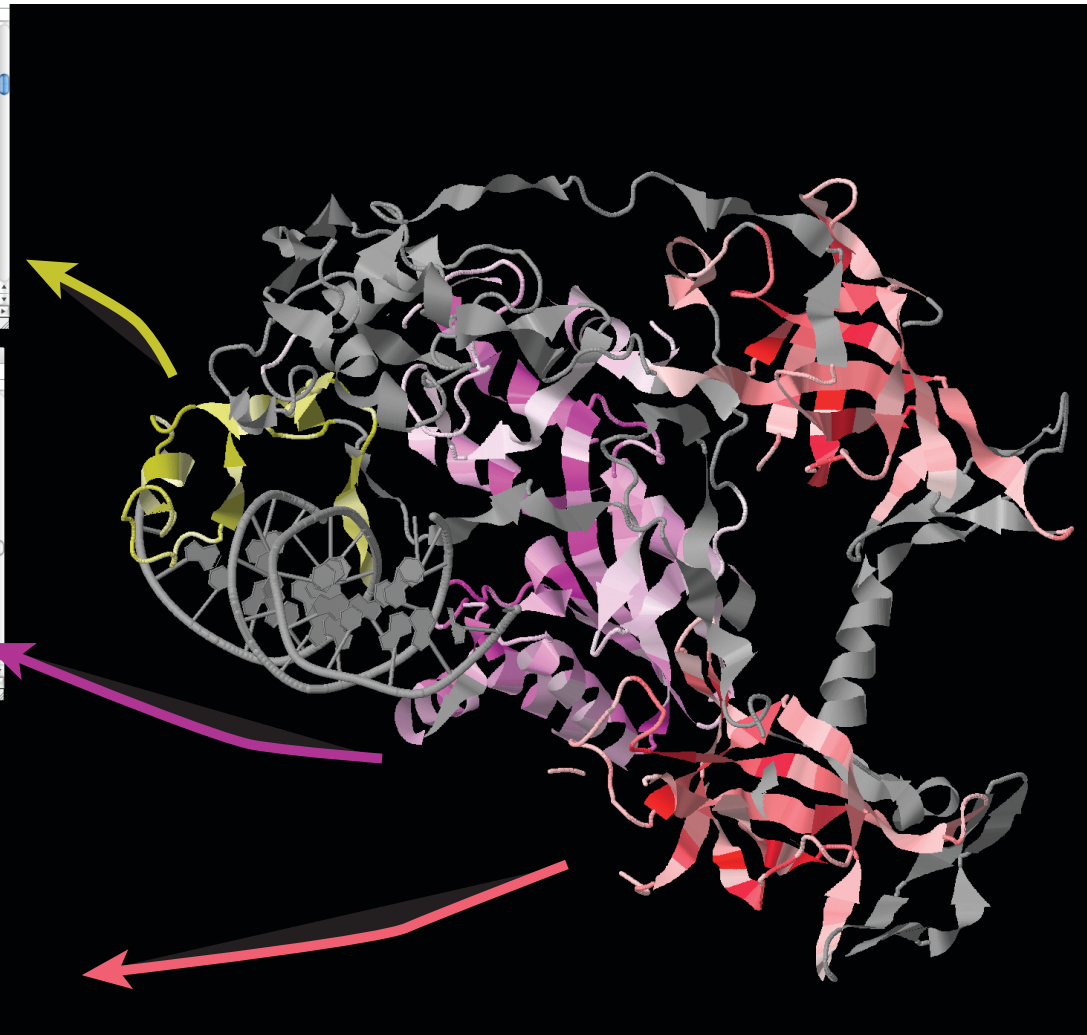
PF00145 Retrieved from PFAM (full)

	310	320	330	340	350	360	370	380	390	400	410	420	430	440	450	460	470	480	490	500
T1_MOUSE	D	P	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
R2_ENTFA	D	G	K	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
12_STRCA	D	E	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
Y6_MACCI	D	E	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
5E_SFRM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
104_ERHAM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
22_NODSP	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
KCS_NICOD	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
1E_SHEP	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
12_NICOD	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
95_ERHAM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
95_SUNY3	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
96_NEME	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
96_STRIC	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
10E_NICOD	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
28_NICOD	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
MR_DICDC	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
97_SFRM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
19_VBRM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
E3_SFRM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
12_ECOLI	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
10D_STRX	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
1E_MACCI	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
42_SFRM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
12_KICBR	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
15_GLOVE	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
43_OHRT	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
41_SLEOT	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
21_SALET	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
12_MACE	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
4E_NODSP	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
97_METRM	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
6Z_NACER	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
91_GLOD	D	E	K	A	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F

Sequence 1995 ID: C6GMH0 STRSX Residue: SER (37)

PF01426 Retrieved from PFAM (full)

	470	480	490	500	510	520	530	540	550	560	570	580	590	600	610	620	630	640	650	660																
T1_MOUSE	L	F	F	V	G	E	L	---	---	E	N	M	---	---	Q	---	L	G	S	V	H	S	K	K	V	V										
T1_PARLI	V	F	F	I	D	E	---	---	---	C	Q	D	T	---	---	Y	---	L	G	S	V	N	A	K	K	T	V									
T1_PARLI	L	Y	---	---	---	---	---	---	---	E	A	V	T	---	---	E	---	L	E	V	Q	Q	G	R	K	T	V									
T1_RAT	L	Y	---	---	---	---	---	---	---	S	D	E	---	---	---	---	F	S	D	V	Q	G	R	K	T	V	V									
T1_RAT	L	F	---	---	---	---	---	---	---	V	G	N	M	---	---	---	---	H	S	V	H	S	K	K	V	V	V									
Z_CAEEL	L	F	---	---	---	---	---	---	---	S	E	L	---	---	---	---	Q	I	T	Q	---	V	S	C	L	R	G	K	T	V	V					
Z_CAEEL	V	F	---	---	---	---	---	---	---	A	T	P	Y	---	---	---	---	Y	E	T	L	---	L	D	E	V	I	G	R	C	L	V	V			
1_HUMAN	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
1_MOUSE	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
1_RAT	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
Z_HUMAN	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
1_MOUSE	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
1_BOVIN	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
1_HUMAN	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
1_MOUSE	L	F	---	---	---	---	---	---	---	L	S	R	Q	---	---	---	---	L	E	S	L	---	A	T	H	R	G	K	K	T	V	V	V			
69_ASCIM	V	A	---	---	---	---	---	---	---	V	H	---	---	---	---	---	---	S	I	H	V	---	V	Q	D	V	I	G	V	R	E	L	V	V		
73_ARATH	L	F	---	---	---	---	---	---	---	Y	---	---	---	---	---	---	---	S	I	H	V	---	V	Q	D	V	I	G	V	R	E	L	V	V		
73_ARATH	L	F	---	---	---	---	---	---	---	L	T	N	E	---	---	---	---	C	L	T	V	---	Q	---	L	K	D	K	G	T	V	S	L	V	V	
26_ARATH	L	Y	---	---	---	---	---	---	---	L	T	N	D	---	---	---	---	F	A	D	I	---	---	---	M	E	C	V	L	R	H	V	K	F	L	V
31_ASCIM	L	F	---	---	---	---	---	---	---	L	T	D	S	---	---	---	---	C	D	T	V	---	---	---	A	D	V	L	G	V	K	F	L	V	V	
103_ASCIM	L	F	---	---	---	---	---	---	---	L	T	D	S	---	---	---	---	C	D	T	V	---	---	---	A	D	V	L	G	V	K	F	L	V	V	
86_DAUCA	V	F	---	---	---	---	---	---	---	Y	---	---	---	---	---	---	---	T	H	T	I	---	---	---	A	E	T	V	E	G	R	C	E	V	V	
86_DAUCA	V	F	---	---	---	---	---	---	---	L	---	---	---	---	---	---	---	C	M	D	F	---	---	---	L	A	D	V	K	K	---	---	---	---	---	



Colouring structures using many multiple alignments

Exercise 25

Section 7.3.3
Page 76

- Get used to loading your own structural data
- Explore the **View => Colour by** submenu
- **WARNING! Instructions are out of date!**
 - Step 25.d and 25.e
 - *Use the 'Add' button to associate each domain alignment view with the MOUSE_DNMT1 structure.*

End of Session 6: 3D structures in Jalview

What we covered:

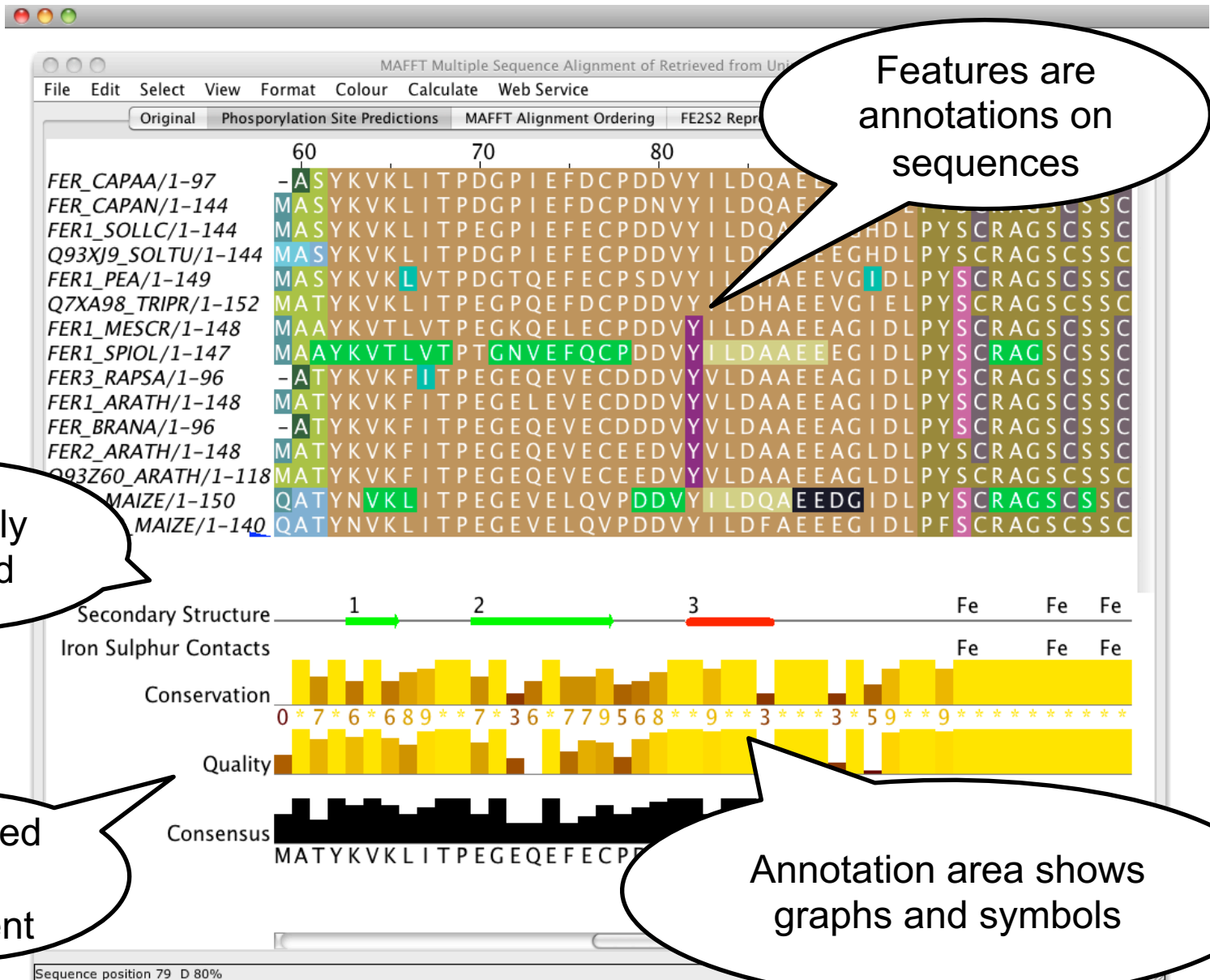
- Viewers supported by Jalview
- Configuring default structure viewer
- Methods of importing structures in Jalview
- Structure to Sequence mapping and methods
- Structure coloring and annotations
- Superimposing structures with MSAs

Session 7 & 8:

Secondary Structure and Disorder

Learning objective:

- Jalview Features and Annotation
- Secondary Structure Prediction Methods
- Running JPred Predictions
- Protein Disorder Prediction
- Different kinds of results from Disorder Predictors



Features are annotations on sequences

Manually created

Calculated from alignment

Annotation area shows graphs and symbols

Select by annotation



RF00014 Retrieved from RFAM (Seed)

File Edit **Select** View Annotations Format Colour Calculate Web Service

Find ⌘F
Select all ⌘A
Deselect all
Invert Sequence Selection ⌘I
Invert Column Selection ⌘⇧I
Create Groups ⌘G

Select/Hide Columns by Annotation

- ✓ secondary structure
- reference positions
- StrucConsensus
- Consensus

Search Filter: Label Description

Structures Filter

Select All Alpha Helix Beta Strand Turn

Select Hide

60 70 80

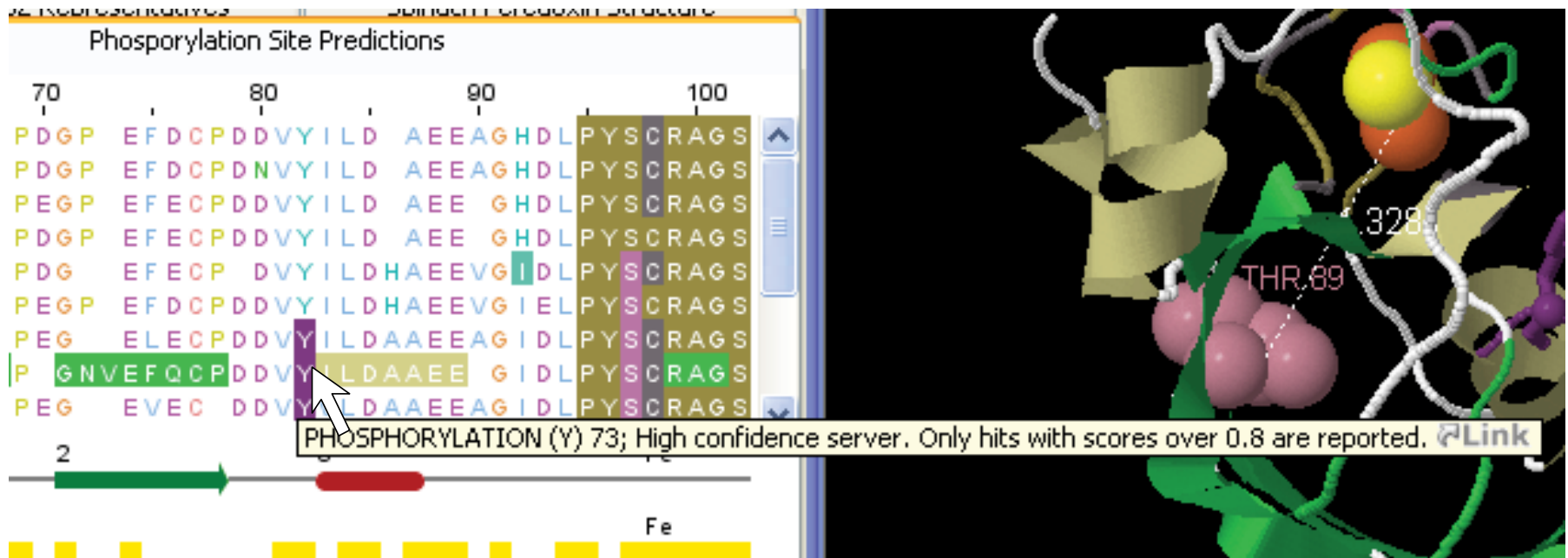
```
UUGAUCCCGA-CUCCUGC-GAGUCGGGAUU  
UCAUCCCGGU-CAUCCCC-AUGGCCGGGAU  
UUCAUCCCGA-CCCCUCAGGGUCGGGAUU  
UUGAUCCCG-AACCCGUA-GGGCCGGGAUU  
UUCAUCCCGACACCCUCA-GGUCGGGAUU
```

u c . g g G u C g G G a u U

---)) -))))))) -

+ C A G G G U C G G G A U U

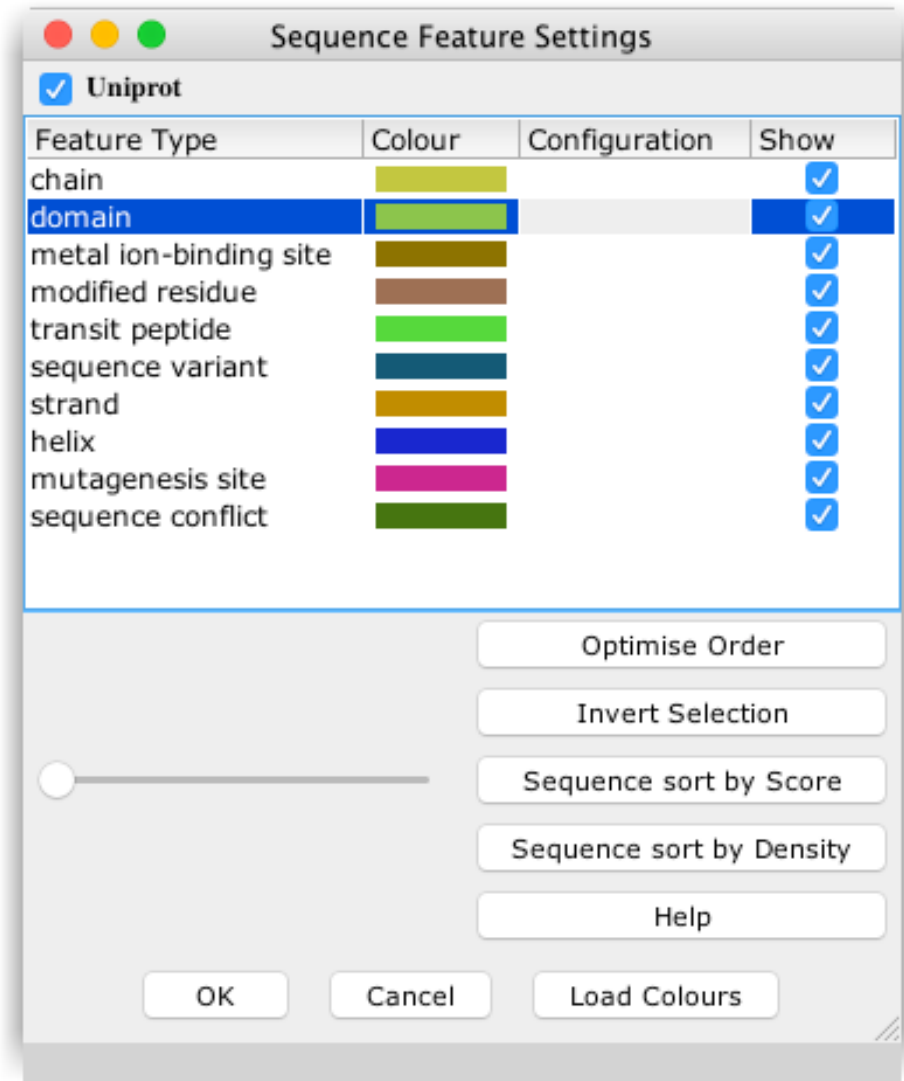
Sequence Features are visualised over the alignment



Customising Feature Display

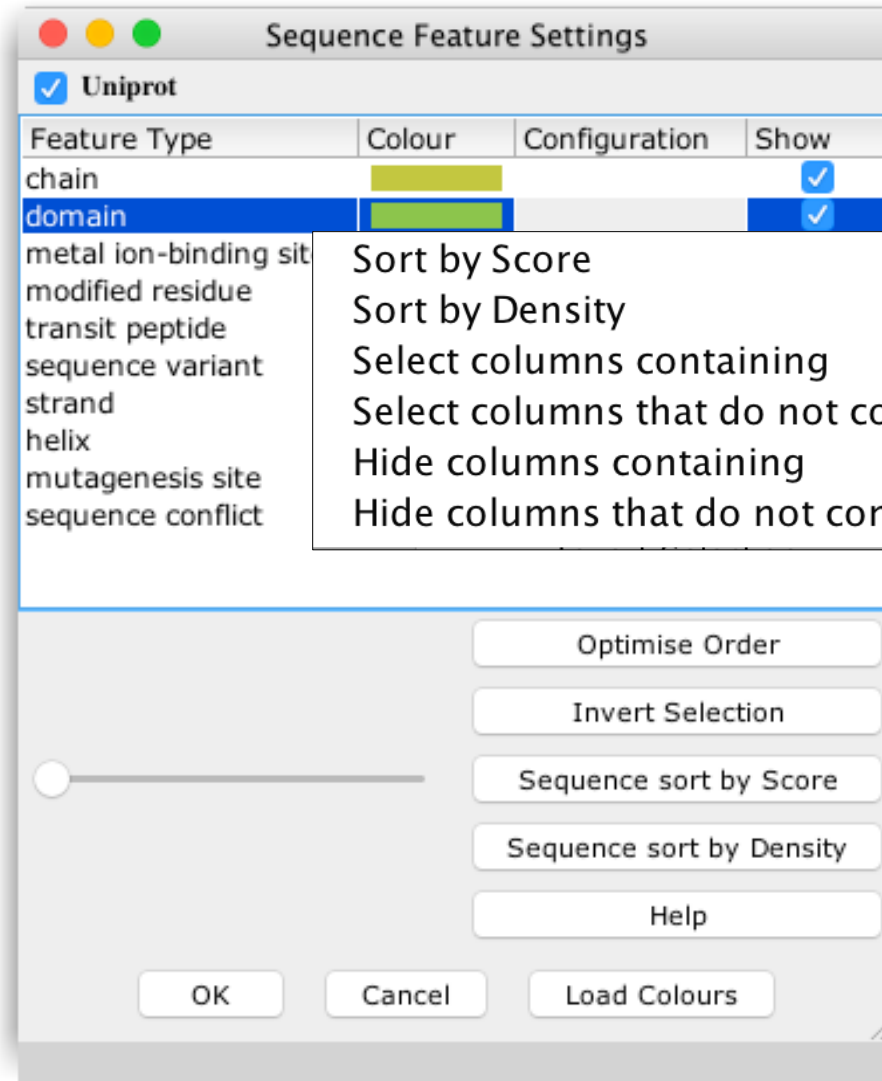
View => Feature Settings...

- Drag features to reorder on alignment
 - Or make transparent
- Change colour
- Enable/disable by name and group



Customising Feature Display

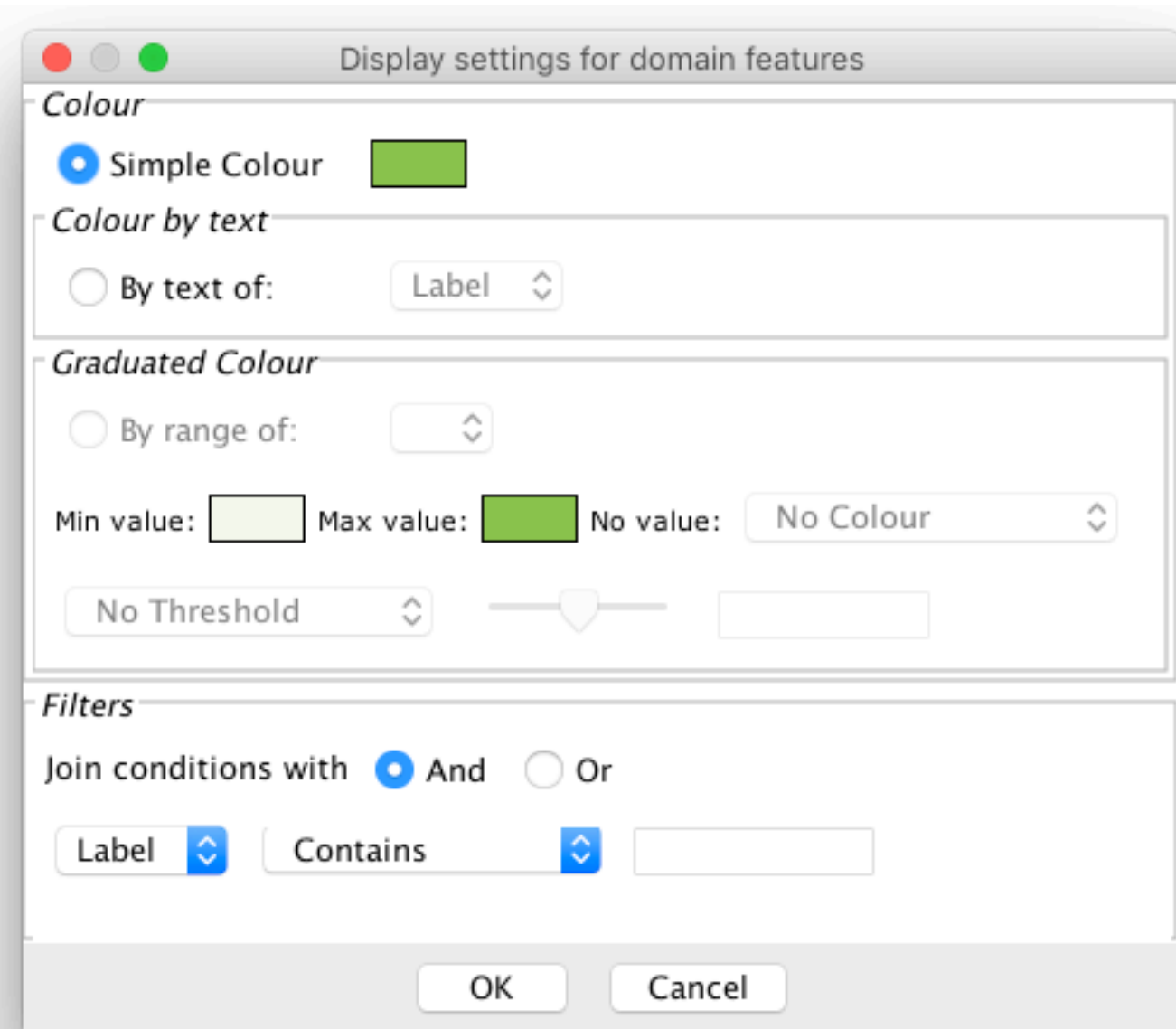
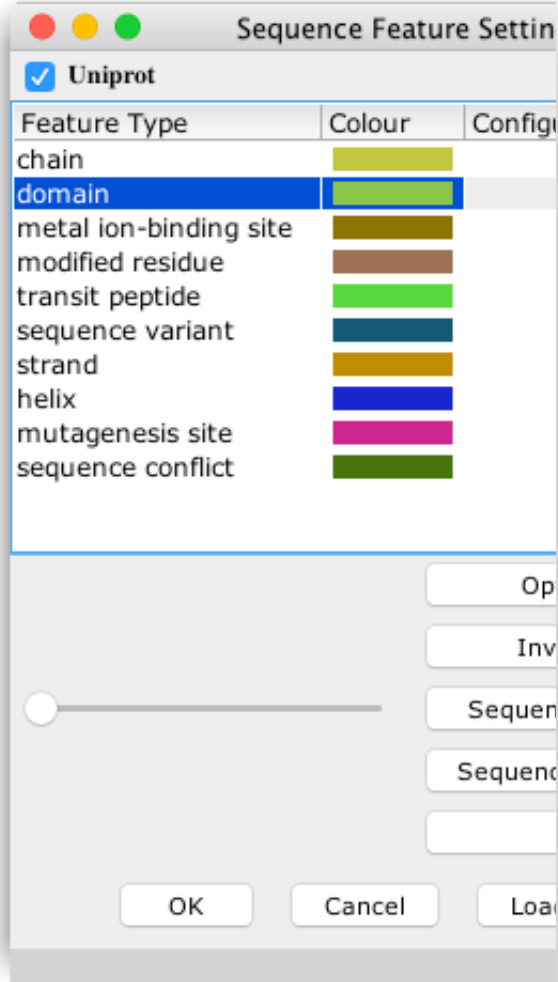
Right-click on feature => Popup menu



**Select
or Hide
columns**

Customising Feature Display

Complex Colour Schemes & filters

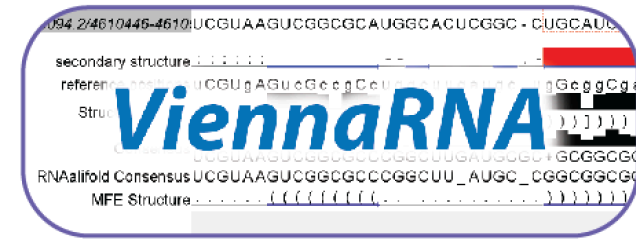


Sources of sequence features

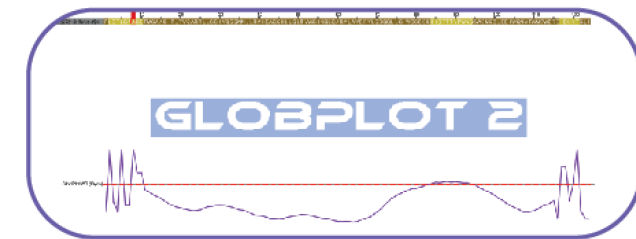
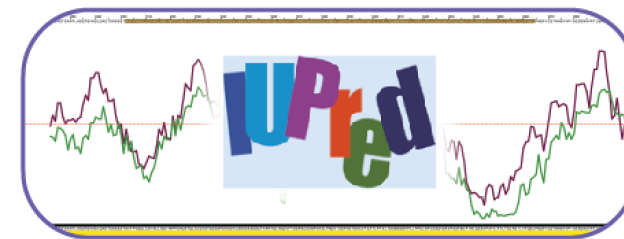
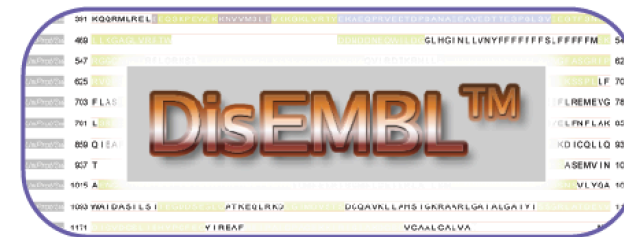
- **Sequence databases**
- 3D Structure associations
- Stockholm & AMSA Alignment files
- **GFF2/3 (*Generalised Feature Format*)**
- **Jalview format feature files**
 - Like GFF
 - can create GFF in Excel
 - Allows scores to be added

SECONDARY STRUCTURE PREDICTION

RNAAlifold
Not today!



PROTEIN DISORDER PREDICTION



<http://www.compbio.dundee.ac.uk/jpred>

Jpred 4

Incorporating Jnet



A Protein Secondary Structure Prediction Server

- Home
- REST API
- About
- News
- F.A.Q.
- Help & Tutorials
- Monitoring
- Contact
- Publications

Input sequence^(?)

```
MQQVPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSKWVPCLEFSKVGFFVYRENHRSPGYDGRYW  
TMWKLPMFGCTDATQVLKELEEAKKAYPDAFVRIIGFDNVRQVQLISFIAYKPPGC
```

[Advanced options \(click to show/hide\)](#)

[Make Prediction](#)

[Reset Form](#)

Primary citation: Drozdetskiy A, Cole C, Procter J & Barton GJ. Nucl. Acids Res.
(first published online April 16, 2015) doi: 10.1093/nar/gkv332 [[link](#)]
More citations: [link](#).



Protein Secondary Structure Prediction

Section 8.1

- Neural network trained on amino acid profiles
 - Predicts **H**elix, **s**h**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 Jpred4 predictions
 - Note the effect of hidden regions
 - Learn about sequence associated annotation
 - Learn about reference sequence views
- Questions
 - ...

Exercise 26

- Tasks
 - Perform a variety of Jpred4 predictions
 - Note the effect of hidden regions
 - Learn about sequence associated annotation
 - Learn about reference sequence views
- 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 ?

AND NOW: SOME DISORDER

File Edit Select View Format Colour Calculate

Web Service

MAFFT Alignment Ordering

Original ... Ferredoxin Structure



FER1_ARATH/1-148 MATYKVKF ITPEGELEVECCDDVY
FER_BRANA/1-96 -ATYKVKF ITPEGEQEVECCDDVY
FER2_ARATH/1-148 MATYKVKF ITPEGEQEVECEEVY

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

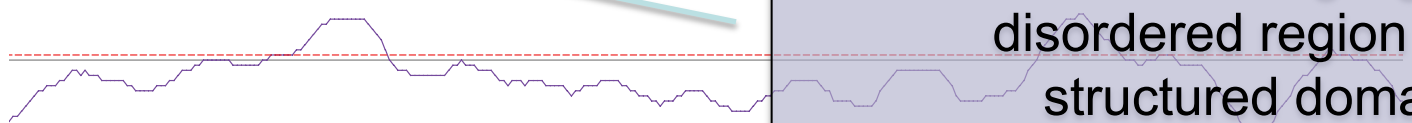
Jalview
JABAWS 2.0
Client

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"/>

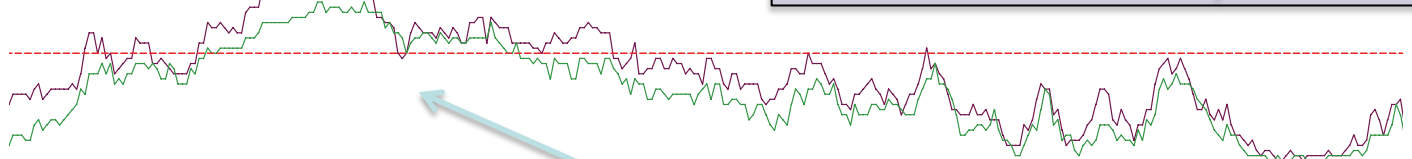


GlobPlotWS (Dydx)



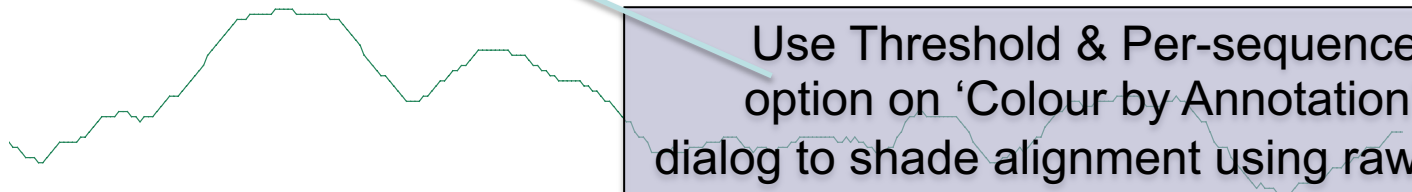
Features highlight
disordered region or
structured domain
predictions

IUPredWS (Short)



IUPredWS (Long)

JronnWS (JRonn)



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

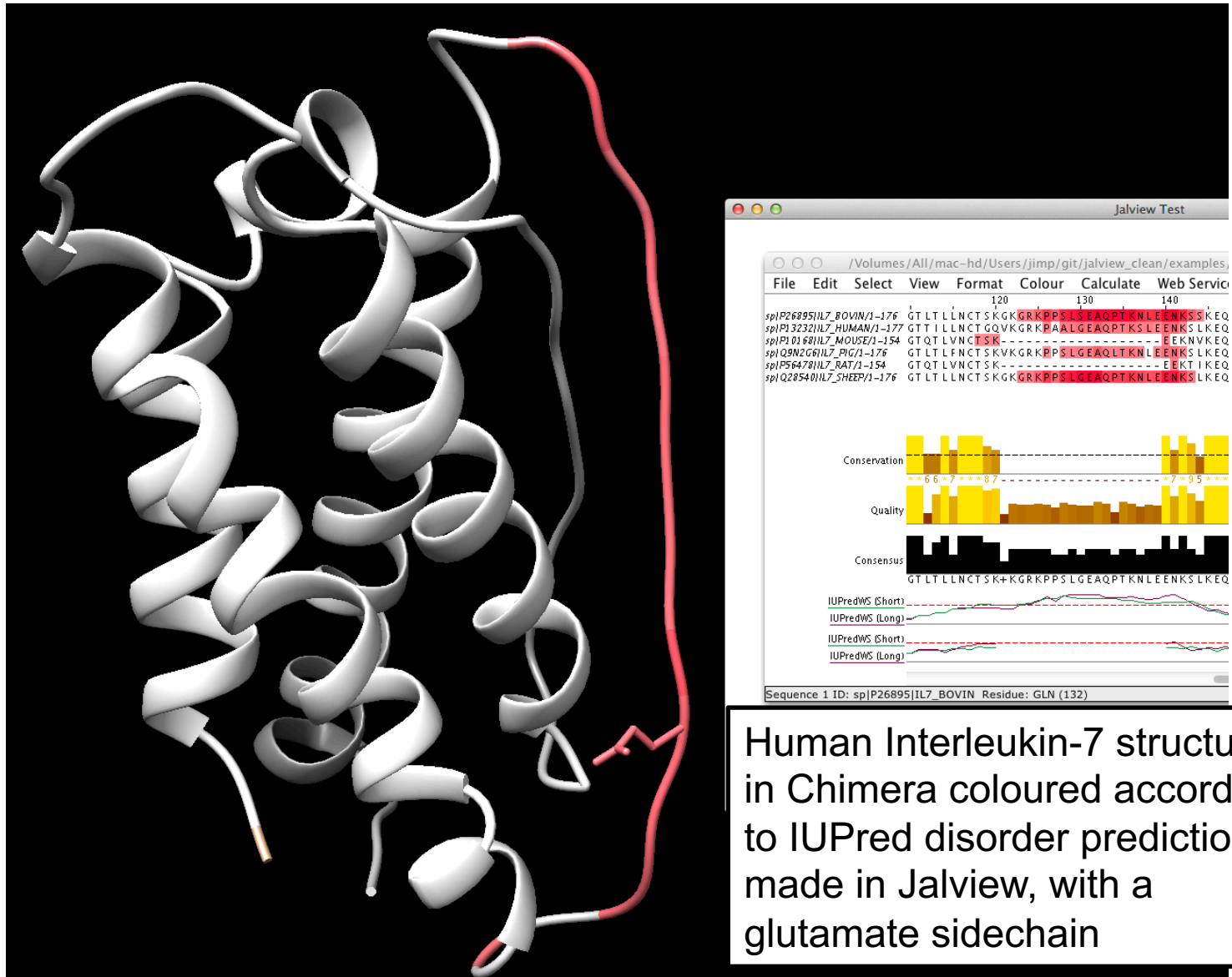
Exercise 27 – IUPred and Disembl Disorder predictions

- 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

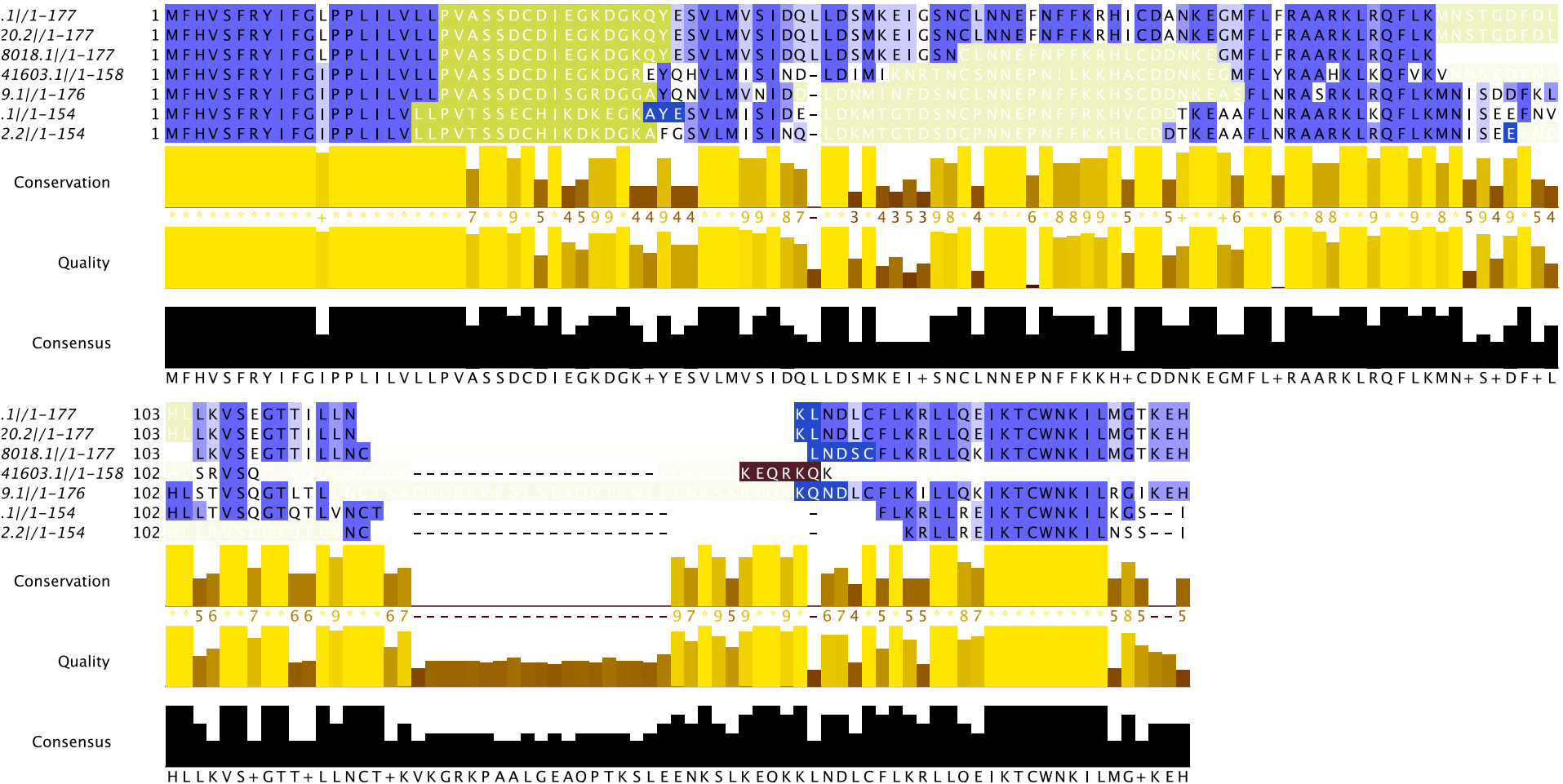
- Task
 - Apply the Disembl and IUPred disorder predictors to a set of protein sequences
 - Compare disordered predictions and available 3D data
- Question
 - How do you decide which regions are ‘significantly disordered’ ?

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



End of Session 7 & 8:

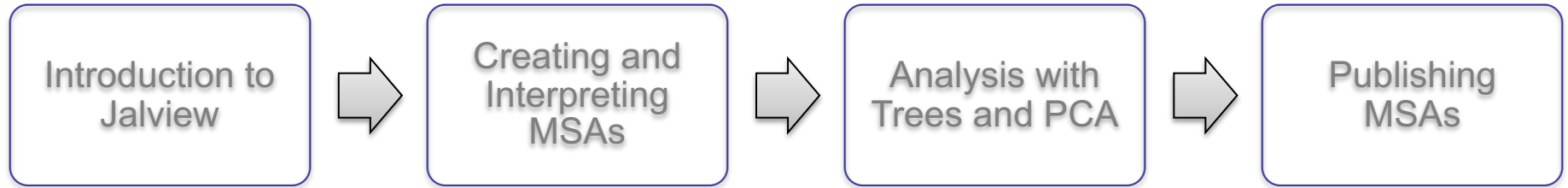
Secondary Structure and Disorder

Learning objective:

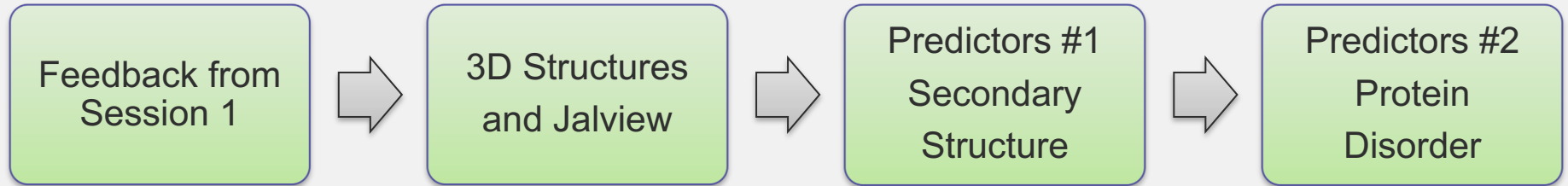
- Jalview Features and Annotation
- Secondary Structure Prediction Methods
- Running JPred Predictions
- Protein Disorder Prediction
- Different kinds of results from Disorder Predictors

Day 3: Schedule

Wednesday: Core Jalview: Multiple alignment creation and MSA figure generation



Thursday: 3D Structures in Jalview, and sequence and MSA based structure prediction



Friday: From Genome to Function – SNPs on transcripts, proteins and 3D structure

