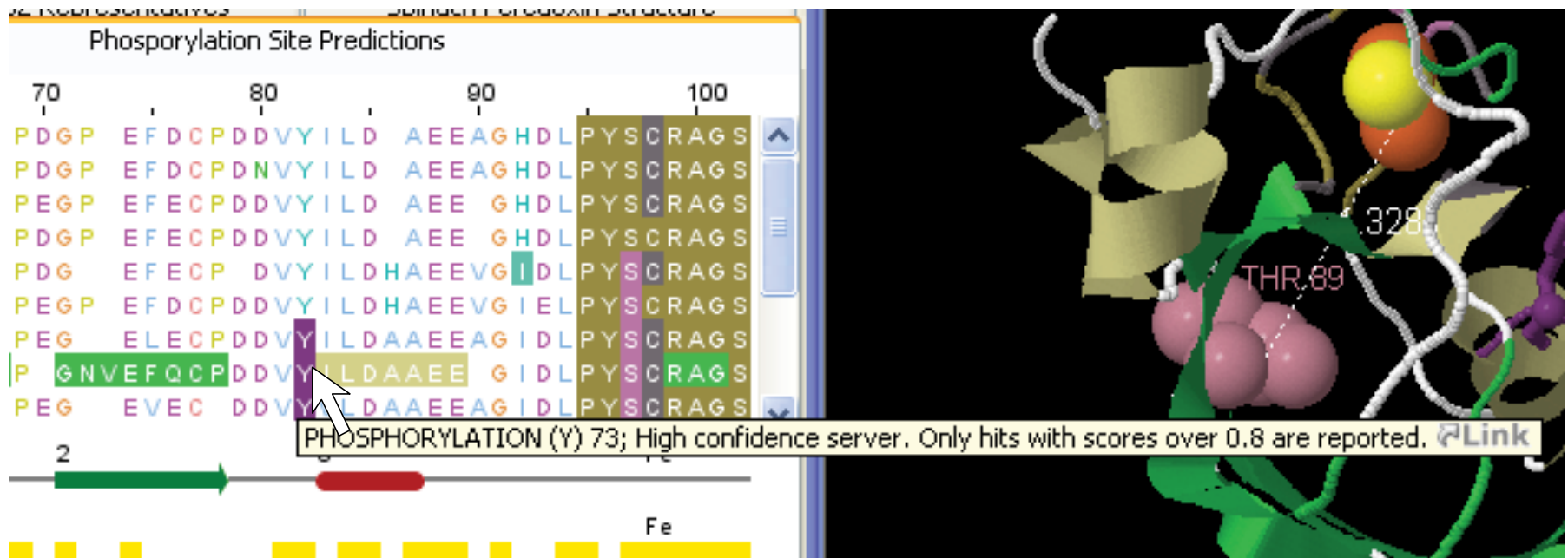


- End of Session 2
 - Alignment & analysis

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

Sequence Features

Manual section 2.8



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"/>
NES-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"/>
COMPBias		<input checked="" type="checkbox"/>
PROSITE		<input checked="" type="checkbox"/>
PRINTS		<input checked="" type="checkbox"/>
PROPEPTIDE		<input type="checkbox"/>
OUTSIDE		<input type="checkbox"/>
CHAIN		<input type="checkbox"/>
Pfam		<input type="checkbox"/>
ProDom		<input type="checkbox"/>

Invert Selection

OK
Cancel
Load Colours
Save Colours

```

P17870 P - - P N L - P C S V T L Q P G P E D T G K A
P51485 A - - P K S - A S S V T L Q P A P G D T G K F
P51466 P - - Q N L - P C S V T L Q P G P E D T G K A
P51487 P - - P N S - P S S V T L Q Q E G D D L G K F
P51477 P - - D F L - P C S V S L Q P A P S D V G K A
P51484 P - - E N A - P P S V T L Q P G S E D Q G R F
P32122 P - - A S S - P S S V T L Q P G D D D Q G K F
P08168 P - - D Y L - P C S V M L Q P A P Q D V G K S
P53179 P R G R G M - L S S I - - - - - - - - K F E F
Q09885 P - - P D I - P D S I - - - - - - - - E G I F
P30647 P - - L N C - P S S Y - - - - - - - - E S Q F
O45782 P - - K S L - P S S F - - - - - - - - E G E F
O76685 P - - I N V - P P S F - - - - - - - - E G K F
O17812 P - - E N I - P Q S F - - - - - - - - E G P F
    
```

Creating Sequence Features

Section 2.8.1-3 & Ex 28

- Tasks
 - Visualise, create, modify, import and export features.

Sequence Features

Section 2.8.1-3 & Ex 28

- Tasks
 - Visualise, create, modify, import and export features.
- Questions
 - What are the different types of file formats available for import and export ?

Files supporting sequence feature data

- Jalview sequence feature files
- GFF files
- Certain 'rich' alignment formats
 - Stockholm
 - AMSA

Sequence Features

Section 2.8.1-3 & Ex 28

- **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 and sequence features for sequence ?

Retrieval from External Databases



Standard DBs



Visualization

Alignments

Structures

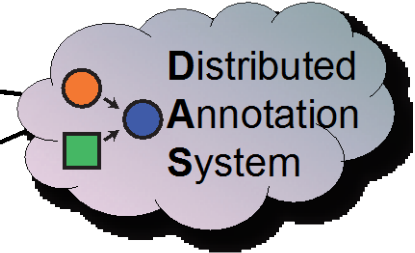
Sequences

Features

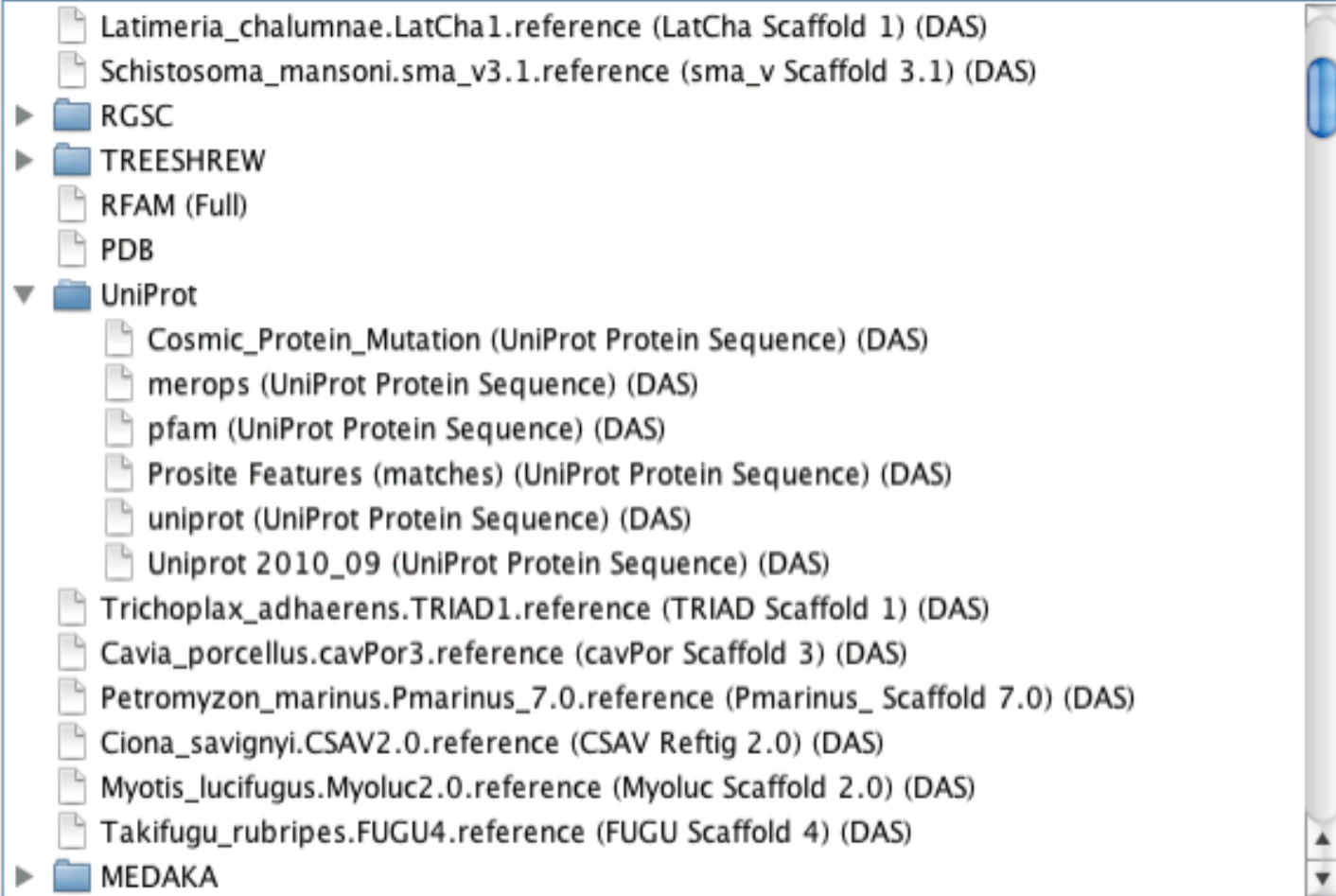
Annotation

Trees

PCA



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

Cut & Paste input - FASTA

File Edit Select View Format

Original View 1

UniProt/Swiss-Prot|P17870
 UniProt/Swiss-Prot|P51480
 UniProt/Swiss-Prot|P51460
 UniProt/Swiss-Prot|P51480
 UniProt/Swiss-Prot|P51470
 UniProt/Swiss-Prot|P51480
 UniProt/Swiss-Prot|P32120
 UniProt/Swiss-Prot|P08160
 UniProt/Swiss-Prot|P53170
 UniProt/Swiss-Prot|Q09880
 UniProt/Swiss-Prot|P30640
 UniProt/Swiss-Prot|O45780
 UniProt/Swiss-Prot|O76680
 UniProt/Swiss-Prot|O17810

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"/>

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:

Authority:

Authority	Clone	eFamily
Any	Gene_ID	Predicted
UniProt	Chromosome	ZFMODELS
PDBresnum	Contig	Computationally inferred
NCBI	Scaffold	eProtein
Ensembl	Volume Map	Manually curated
SGD		
MOZ2a		
RGSC		
MGI		

Use Registry

Find Uniprot Accession Ids

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

0

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 -

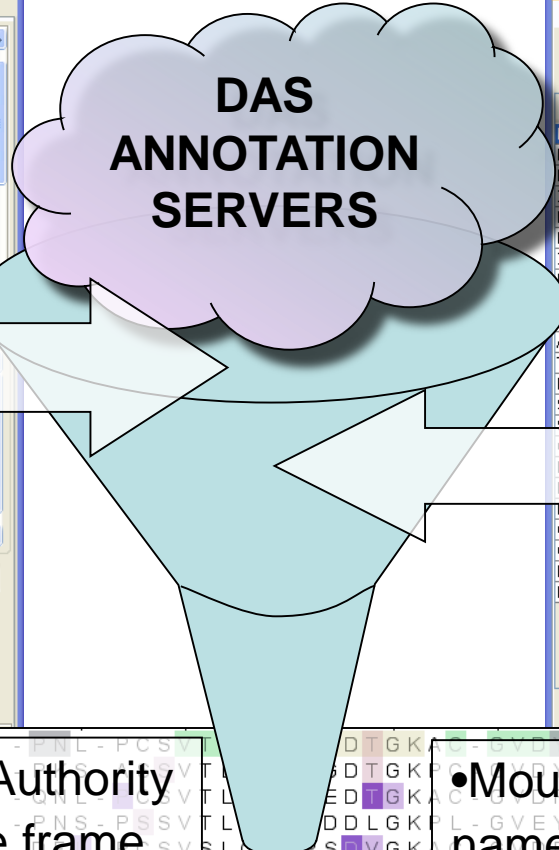
Sequence Features Dialog box

Sequence Feature Settings (DAS Settings)

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/> [Reset]

[Refresh Available Sources] [Add Local Source]



Sequence 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"/>
TURNOVER		<input checked="" type="checkbox"/>
HELI		<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|P17870P - - PNL - PCSV - - - - - GVDL - EVRAPCAENL - -
 UniProt/Swiss-Prot|P08168P - - DYL - PCSV - - - - - GVDL - EVRAPCAENL - -
 UniProt/Swiss-Prot|P51487P - - PNS - PSSV - - - - - GVDL - EVRAPCAENL - -
 UniProt/Swiss-Prot|P51484P - - ENA - PPSV - - - - - GVDL - EVRAPCAENL - -
 UniProt/Swiss-Prot|P32122P - - ASS - PSSV - - - - - GVDL - EVRAPCAENL - -
 UniProt/Swiss-Prot|P08168P - - DYL - PCSV - - - - - GVDL - EVRAPCAENL - -
 UniProt/Swiss-Prot|P53179P - - RGR - LSS - - - - - KFERG - SITYFLSCTLES - -
 UniProt/Swiss-Prot|Q09889P - - PDI - PDS - - - - - EGIPGCHI - IYTLTASLERATQ - -
 UniProt/Swiss-Prot|P30647P - - LNC - PSS - - - - - ESQFG - SIRYQMKVELRASTD - -
 UniProt/Swiss-Prot|O45782P - - KSL - PSS - - - - - EGDFG - HIRYTKCAICERPWD - -
 UniProt/Swiss-Prot|O76685P - - INV - PPS - - - - - EGKYG - YLRYSVTA - EVDRPWR - -
 UniProt/Swiss-Prot|O17812P - - ENI - PQS - - - - - EGDFG - FIRFYIKVHMDRPHA - -

The Distributed Annotation System

Section 2.9.2, Exercise 31

- Task
 - Browse available DAS sources for protein sequences
 - Retrieve annotation for the ferredoxin alignment.

The Distributed Annotation System

Section 2.9.2, Exercise 31

- Task
 - Browse available DAS sources for protein sequences
 - Retrieve annotation for the ferredoxin alignment.
- Question
 - What does the ‘optimise order’ button do ?

Shading, thresholding, colour by label.

```

... MASTALSSAIVSTSF LRRQQTPI SLRSLPFANT... QSLFGLKS...
... MAATTTTMMGMATTFVVPKQAPPMM AALPSNTG... RSLFGLKT...
... MASISGTMISTSF LPRKPAVTS LKAIS... NVG... EALFGLKS...
... MASTALSSAIVSTSF LRRQQTPI SLRSLPFANT... QSLFGLKS...
... MASISGTMISTSF LPRKPVVTS LKAIS... NVG... EALFGLKS...
ATVLSGSPRAPAFFFS SSSSLRAAPAPTAVALPAAKVG... IMGRSA...
... MASVSATMISTSF MPRKPAVTS LKPI... NVG... EALFGLKS...
... MATTPALYGTA VSTSF MRRQVPVMSVATT TTKAFPSGGFGLKSV...
... MAATTAALSGATMSTAFAPKT... PPMTAALPTNVG... RALFGLKS...
... MASTALSSAIVGTSF IRRSPAPIS LRSLSANT... QSLFGLKS...
... MAATALSMSILRAPP... PCFSSPLRLRVAVAKPLAAPM...
... MATTPALYGTA VSTSF LRTQPMMSVT TTKAFSN... GFLGLKT...

```

Graduated Feature Colour for hydrophobic_region

Colour by Label Min: Max:

No Threshold Threshold is Min/Max

Adjust threshold

```

... MMAS+ALSGT+VSTSF LRRQPAPTSLAALP+NVG... +SLFGLKS... STA

```

```

VYILDQAAEEAGHDL PYS CRAGSCSSCAGKIAGGAVDQTDGNFLDDQLE
VYVLDAAEEAGLDLPYS CRAGSCSSCAGKVVVSGSIDQSDQS... LDD...
VYILDAAEEAGIDL PYS CRAGSCSSCAGKLTGSLND... S... DQIT...
VYILDQAAEEESHDL PYS CRAGSCSSCAGKVTAGSVDDQSDGNFLDDEQEA
VYVLDAAEEAGIDL PYS CRAGSCSSCAGKVVVSGSMDQSDQS... FLDDDDQIA
VYVLDAAEEAGLDLPYS CRAGSCSSCAGKVVVSGSIDQSDQS... LDD... EQMS...
VYILDQAAEEESHDL PYS CRAGSCSSCAGKVTAGTVDDQSD... KFLDDDDQEA
VYILDQAAEEEDGIDL PYS CRAGSCSSCAGKVVVSGSMDQSDQS... SYLDDDDQIA
VYILDQAAEEAGHDL PYS CRAGSCSSCAGKIAGGAVDQTDGNFLDDQLE
VYILDHAAEEVGIEL PYS CRAGSCSSCAGKVVNGNVNDEDS... LDDDEQIE
VYVLDAAEEAGIDL PYS CRAGSCSSCAGKVVVSGFVDDQSD... FLDDDDQIA
VYILDAAEEAGIDL PYS CRAGSCSSCAGKVTSGSVNQQDDG... LDDDDQII
VYVLDAAEEAGIDL PYS CRAGSCSSCAGKVVVSGSMDQSDQS... FLDDDEQIG
VYILDFAEEEGIDL PYS CRAGSCSSCAGKVVVSGSMDQSDQS... FLNDNQVA
VYILDHAAEEVGIEL PYS CRAGSCSSCAGKVVVGGVDDQSDG... FLDDDEQIEAGFVLT... CVAYPTSDVV... IETH

```

Feature Type	Colour	Display
metal ion-binding site		<input checked="" type="checkbox"/>
O-phosphorylated L-serine		<input checked="" type="checkbox"/>
O-phosphorylated L-threonine		<input checked="" type="checkbox"/>
O4'-phosphorylated L-tyrosine		<input checked="" type="checkbox"/>
hydrophobic_region		<input checked="" type="checkbox"/>

Optimise Order

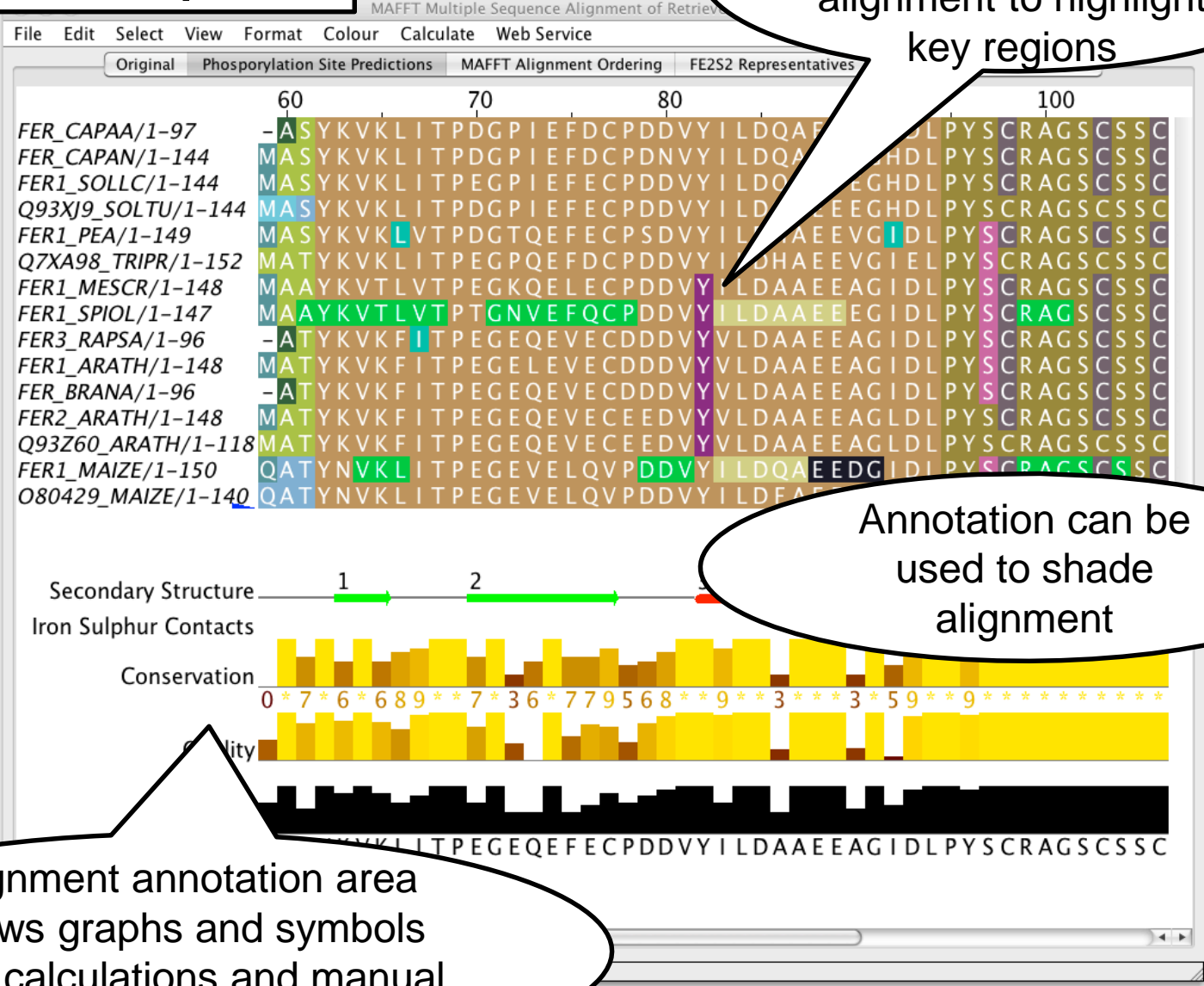
Invert Selection

Seq sort by Score

Seq Sort by density

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 Se



hour Calculate

Web Service

MA

Original

Spinach Ferredoxin Structure

Q7XA98_TRIPR1-152

FER1_MESCR1-148

FER1_SPIOL1-147

FER3_RAPSA1-96

FER1_ARATH1-148

FER_BRANA1-96

FER2_ARATH1-148

	60		70		80																			
Q7XA98_TRIPR1-152	M	A	T	Y	K	V	K	L	I	T	P	E	G	P	D	E	F	D	C	P	D	D	V	Y
FER1_MESCR1-148	M	A	A	Y	K	V	T	L	V	T	P	E	G	K	Q	E	L	E	C	P	D	D	V	Y
FER1_SPIOL1-147	M	A	A	Y	K	V	T	L	V	T	P	T	G	N	V	E	F	Q	C	P	D	D	V	Y
FER3_RAPSA1-96	-	A	T	Y	K	V	K	F	I	T	P	E	G	E	Q	E	V	E	C	D	D	D	V	Y
FER1_ARATH1-148	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_BRANA1-96	-	A	T	Y	K	V	K	F	I	T	P	E	G	E	Q	E	V	E	C	D	D	D	V	Y
FER2_ARATH1-148	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

File Edit Select View Format Colour Calculate

Web Service

MAFFT Alignment Ordering

Original ... Ferredoxin Structure



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 secondary structure prediction

Section 2.6 onwards
in the Manual

Page 65



Protein Secondary Structure Prediction Sec. 2.6

• Neural network trained on amino acid profiles

- Predicts **Helix**, **Sheet**, or **Coil** 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 ?