9.00-9.15am. **Overview of the day**

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

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

10.30-11am. ***Coffee***

11am – 11.30am **Geoff Barton: Multiple alignment and Analysis**

11.30am - 12.30pm. **Session 2: Alignment & alignment analysis**

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

12.30pm to 1.30pm. ***Lunch***

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

- Creating and viewing sequence annotation
- Protein Secondary structure prediction

3.00pm – 3.30pm ***Coffee***

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

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

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

Course materials

Everything is online

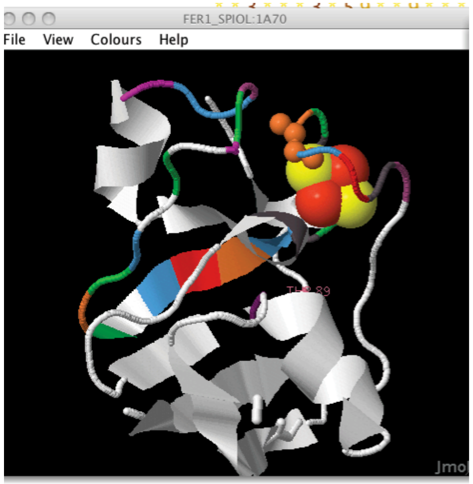
http://www.jalview.org/tutorial/training-materials/2014/London/Dec/

- **These slides**
- **Jalview v2.8.2 Manual (v1.5.0)**
  - **Log in and**
  - **Open the manual in your PDF Viewer NOW**

- Session 4
  - Working with structures
    - Viewing PDB data
    - Superimposing structures
  - Mapping data onto structure
    - Disorder prediction
    - RNA helix shading

Desktop Structure Visualization
3D structures and 2D RNA diagrams

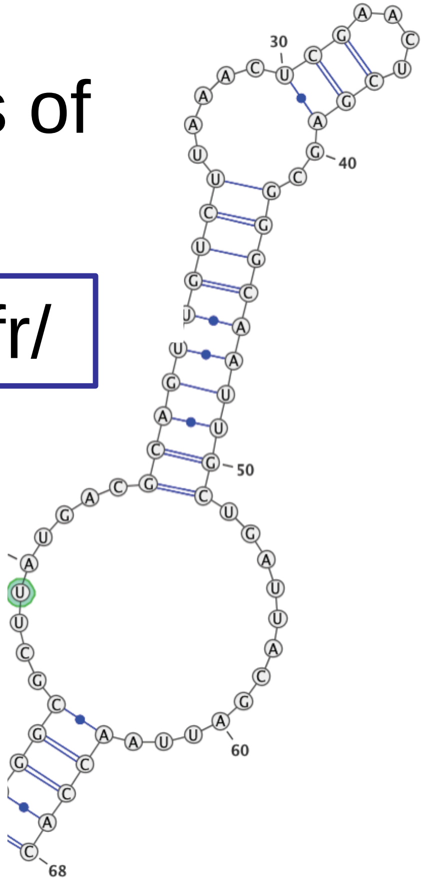
http://jmol.sourceforge.net/

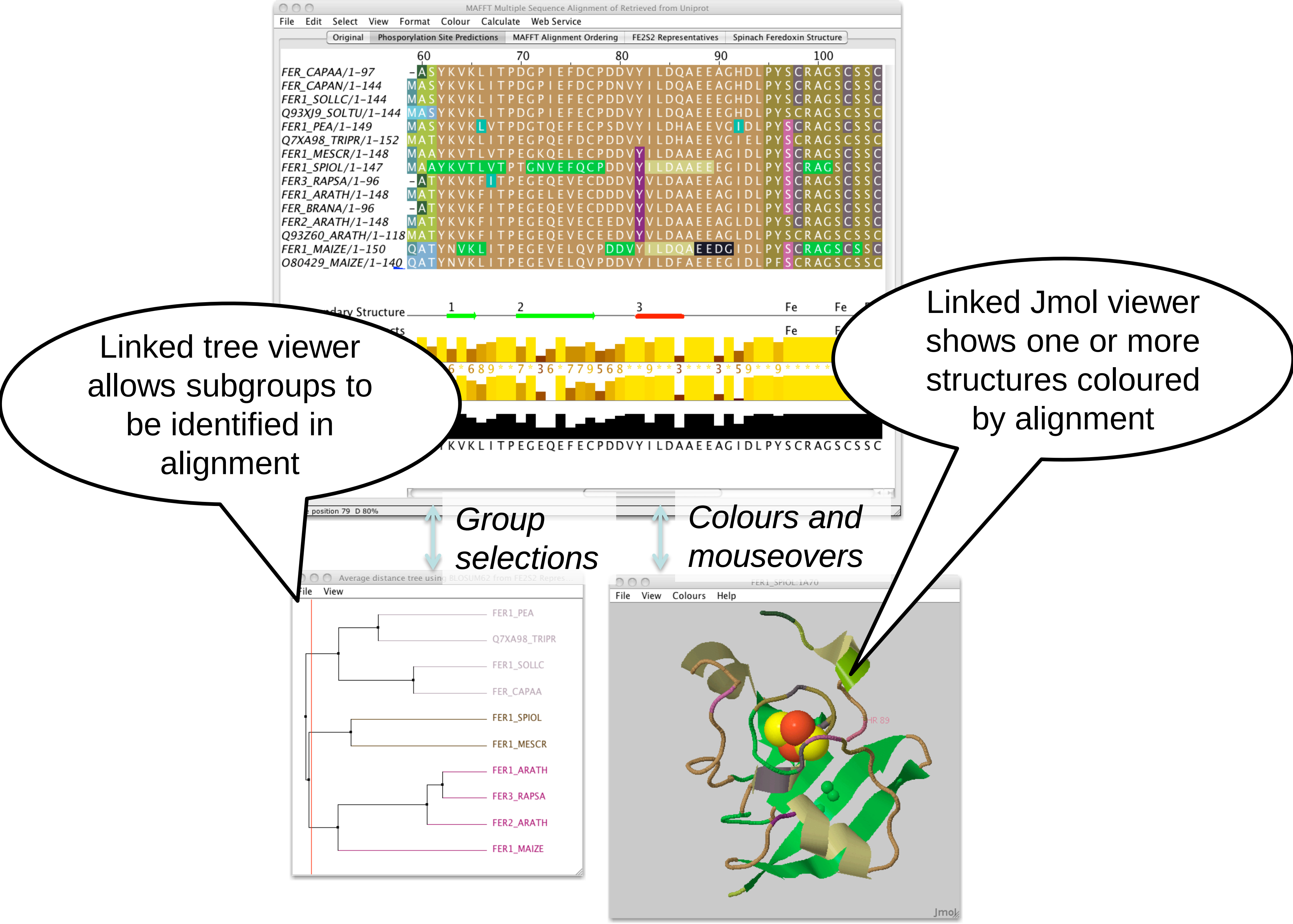


VARNA

Visual Analysis of RNA

http://varna.lri.fr/





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

Linked tree viewer allows subgroups to be identified in alignment

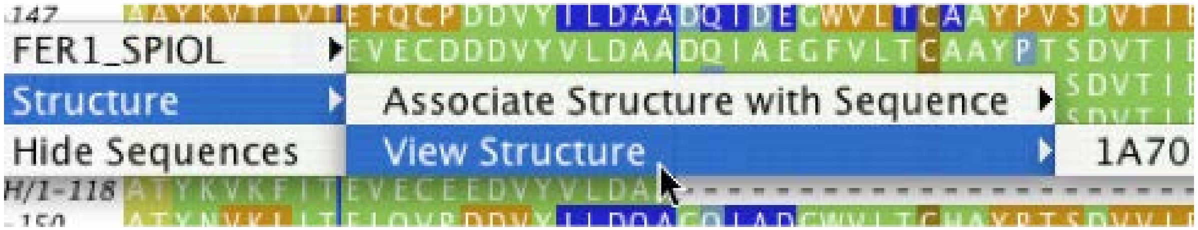
Group selections

Colours and mouseovers

Linked Jmol viewer shows one or more structures coloured by alignment

Associating structures with sequences

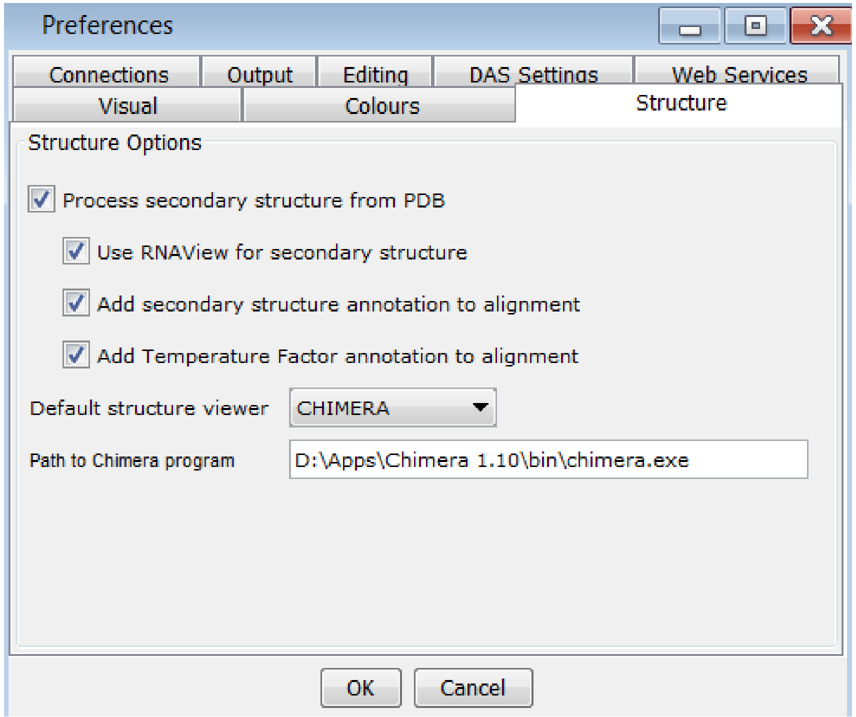
- Local PDB file
  - – Attach PDB file to sequence manually
  - – drag and drop to match files to sequences by ID
- Structures in the PDB database
  - – Provide PDB id (and chain) for sequence
  - – Discover references *via* sequence database



New Structure tab in Jalview Preferences

Checkboxes for:

- Secondary structure
  - Protein DSSP built in
  - RNAView web service
- Temperature factor
- Set default 3D viewer
  - Jmol
  - Chimera



☑ Process secondary structure from PDB

☑ Use RNAView for secondary structure

☑ Add secondary structure annotation to alignment

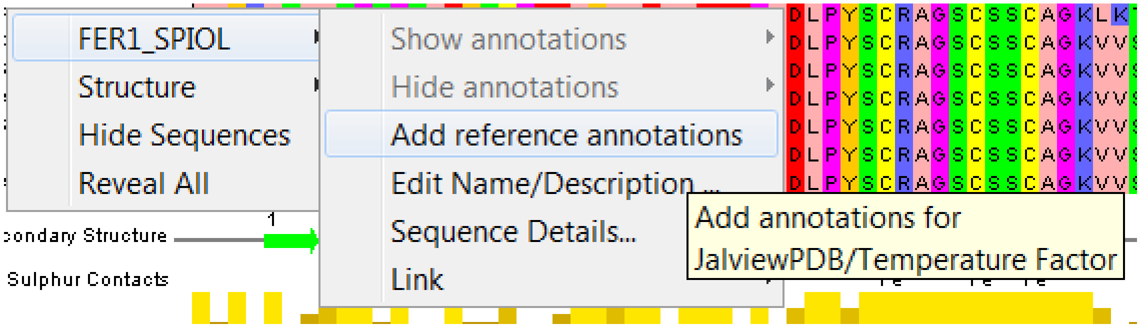
☑ Add Temperature Factor annotation to alignment

Default structure viewer: CHIMERA

Path to Chimera program: D:\Apps\Chimera 1.10\bin\chimera.exe

When Protein annotation is enabled

- Add Reference Annotations
  - Available when any sequence associated annotation can be added to the current view
  - Access via Sequence ID and ‘Selection’ submenu



Add annotations for JalviewPDB/Temperature Factor

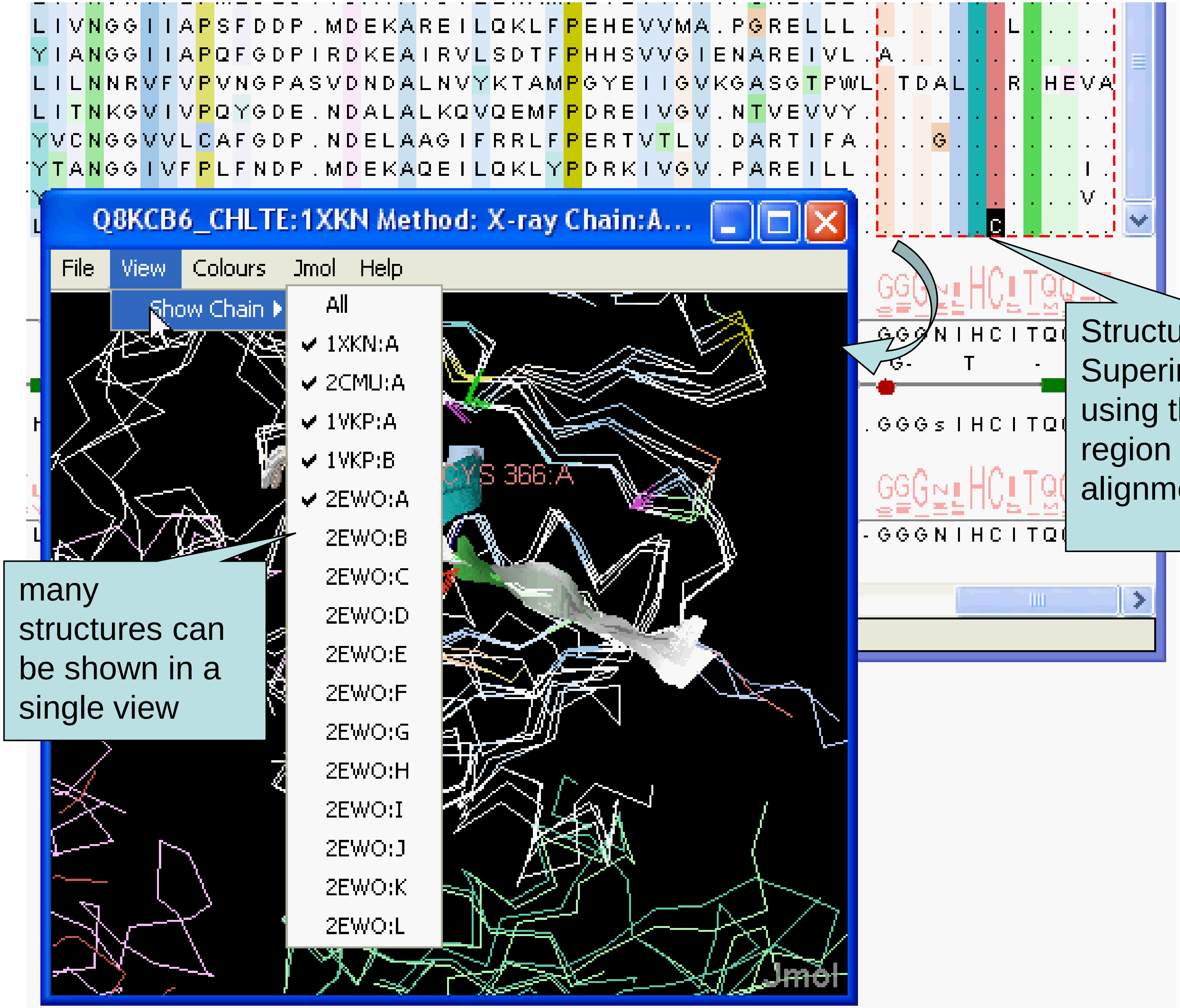
Protein Structures in Jalview
Sec 2.1. Exercise 14

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

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

Protein Structures in Jalview
Sec 2.1. Exercise 14

- Task
  - Discover PDB structures for ferredoxin sequence(s)
  - Save and load structures and manipulate colouring
- Questions
  - How does Jalview match up sequence data to structural data
  - Did anyone try ‘Add Reference Annotation’ ?



Structures can be Superimposed using the visible region of alignment

many structures can be shown in a single view

Superposing Structures using Alignments
Sec 2.1.4 – Exercise 15

- Task
  - Align structures using the ferredoxin alignment

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

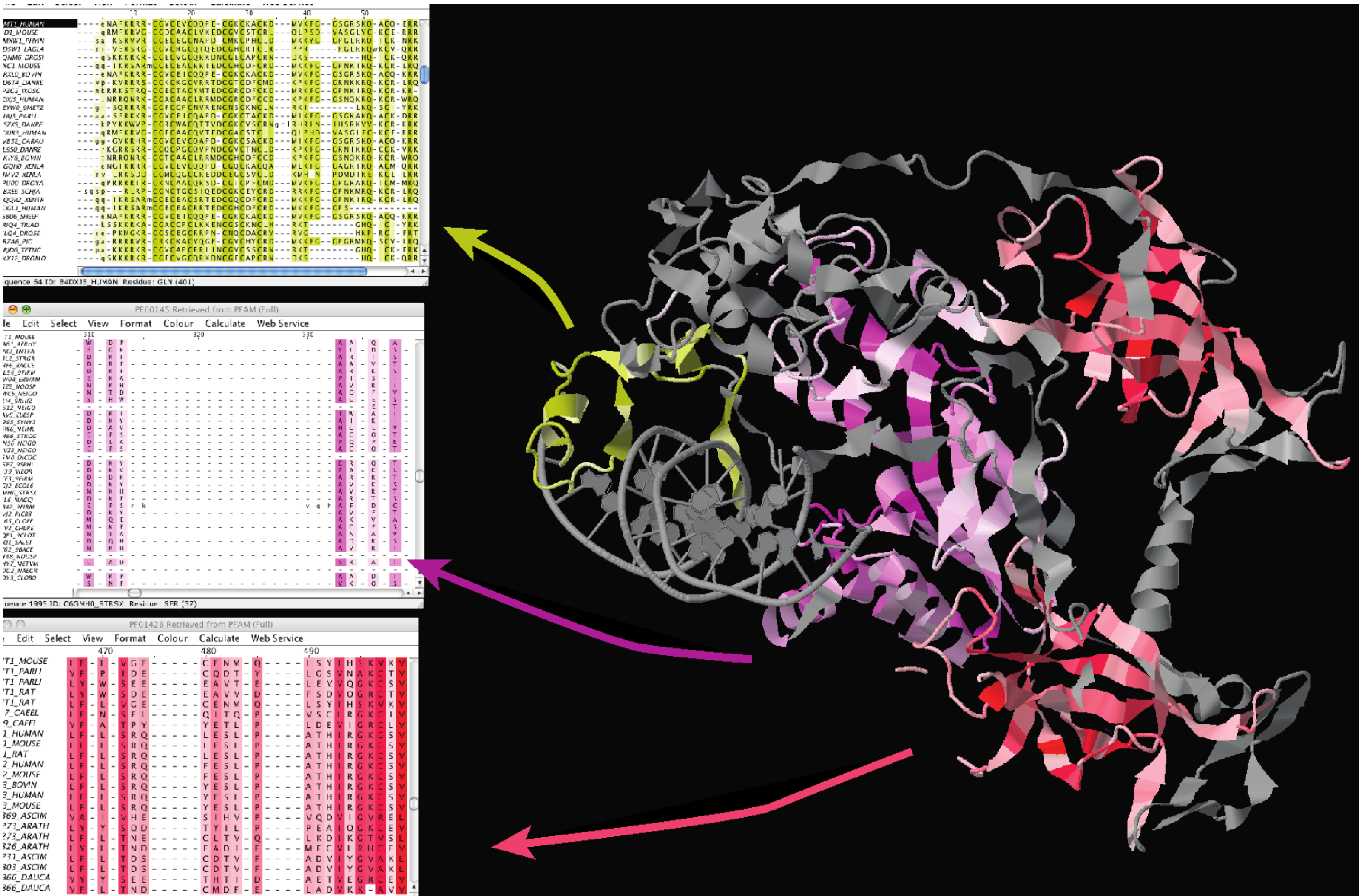
- **Use Jmol->Align menu**

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

Superposing Structures using Alignments
Sec 2.1.4 – Exercise 15

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

Colouring structures using many multiple alignments
Sect 2.1.5. Exercise 16



Protein Disorder prediction

Exercise 27

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

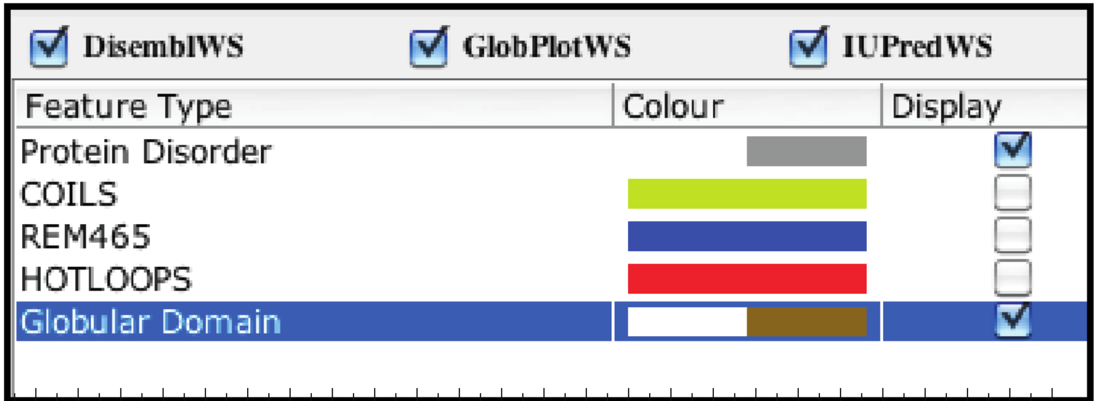
- Programs provided by JABAWS 2 employ
  - Machine learning methods (**DisEMBL**)
  - Similarity to disordered sequences (**RONN**)
  - Empirical amino acid statistics (**IUPred**, **GlobPlot**)

Disorder Predictions from JABAWS

JABAWS Analysis Service

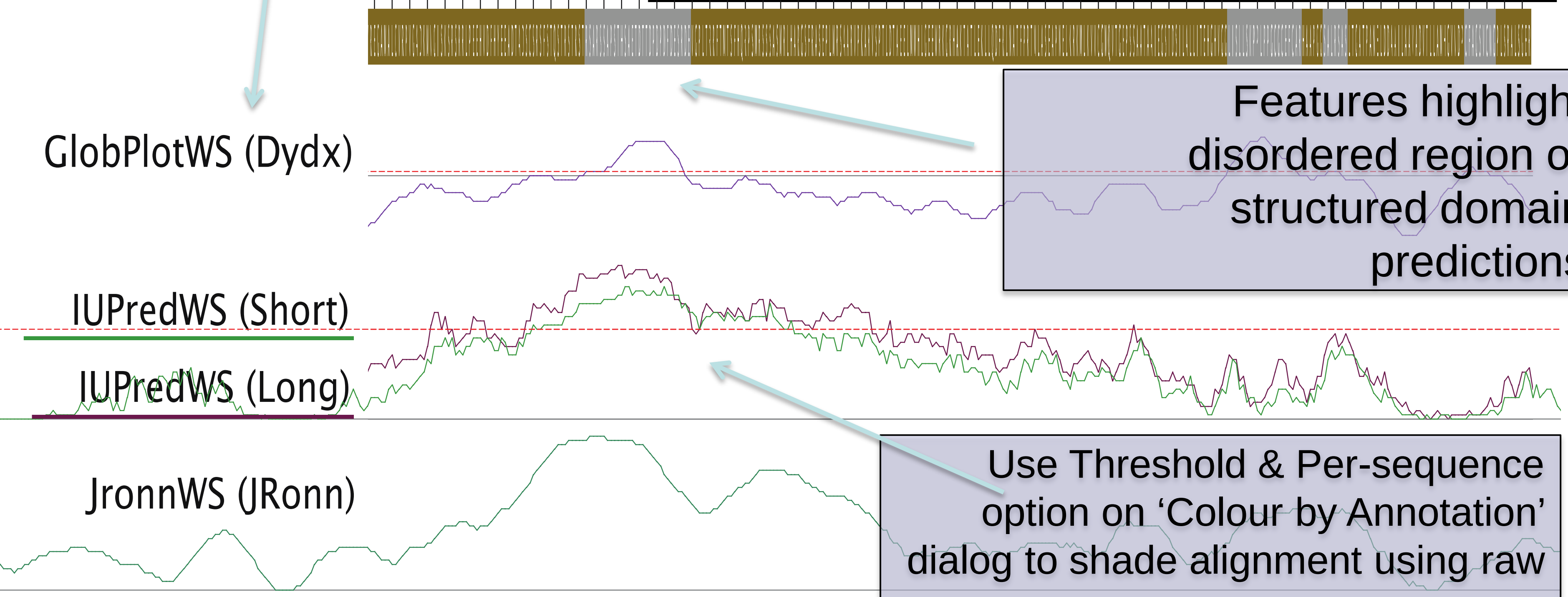
Disorder Predictions

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



☑ DisemblWS ☑ GlobPlotWS ☑ IUPredWS

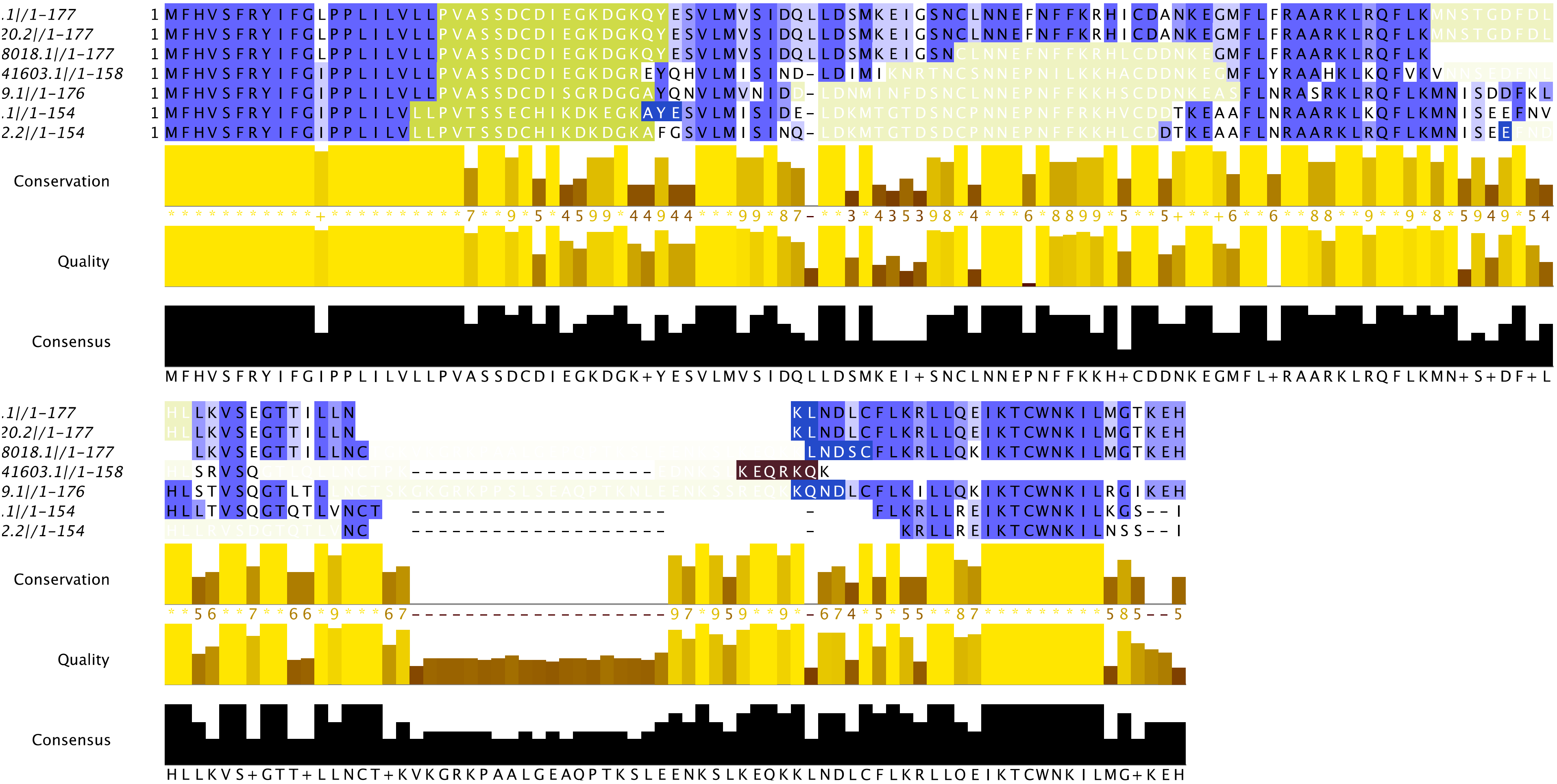
Feature Type	Colour	Display
Protein Disorder	▨	☑
COILS	🟩	☐
REM465	🟦	☐
HOTLOOPS	🟥	☐
Globular Domain	🟫	☑



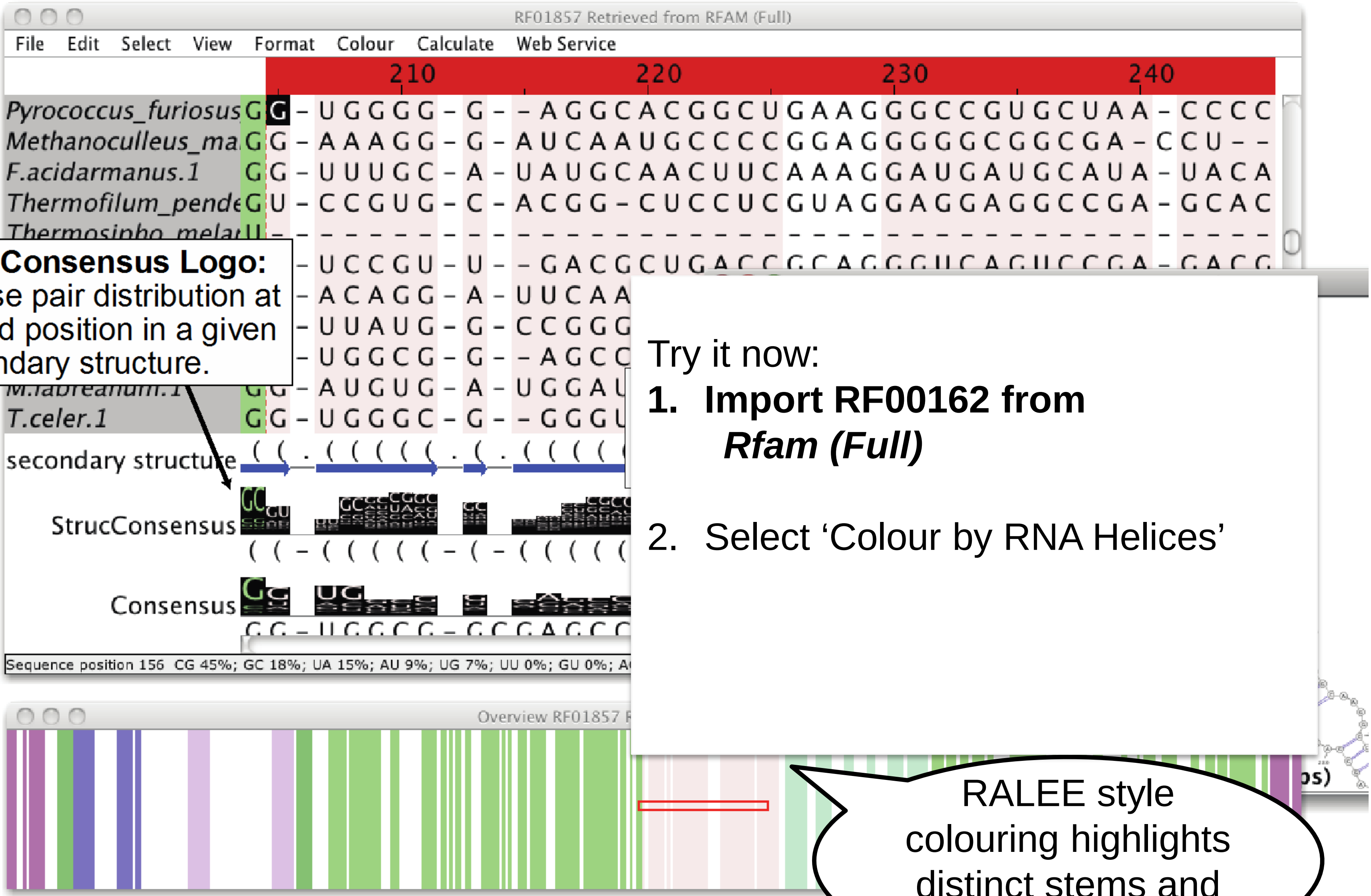
Features highlight disordered region or structured domain predictions

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

Disorder in Interleukin 7



RNA 2nd-ary Structure



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

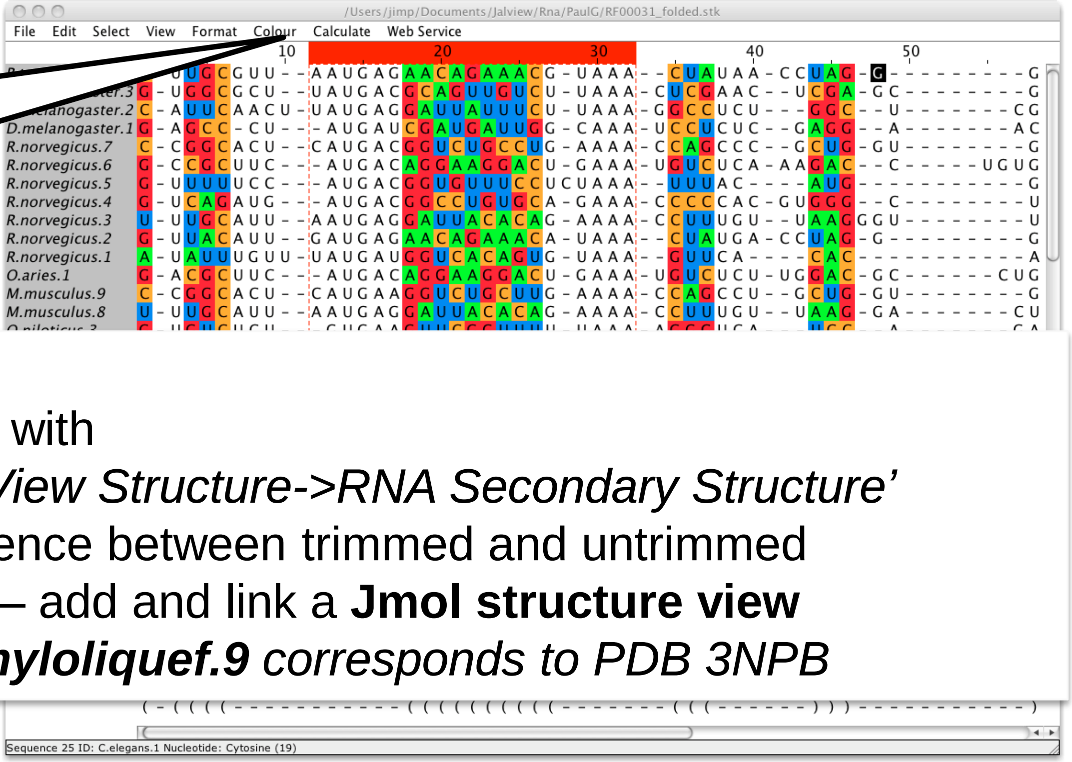
Structure Consensus Logo: Shows base pair distribution at each paired position in a given RNA secondary structure.

Try it now:

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

RALEE style colouring highlights distinct stems and helices

Selections and mouse positions shared between alignment view and VARNA



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

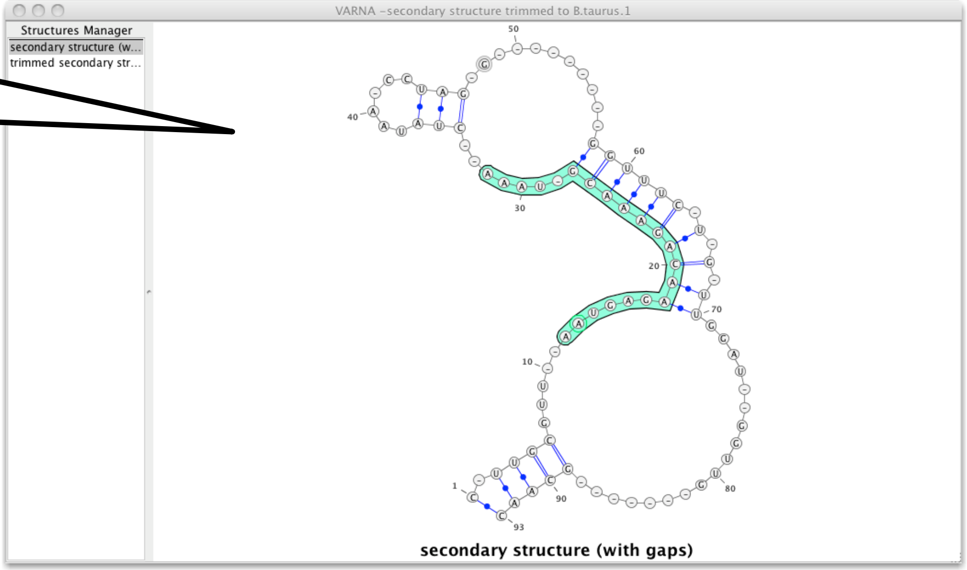
Try it now:

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

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

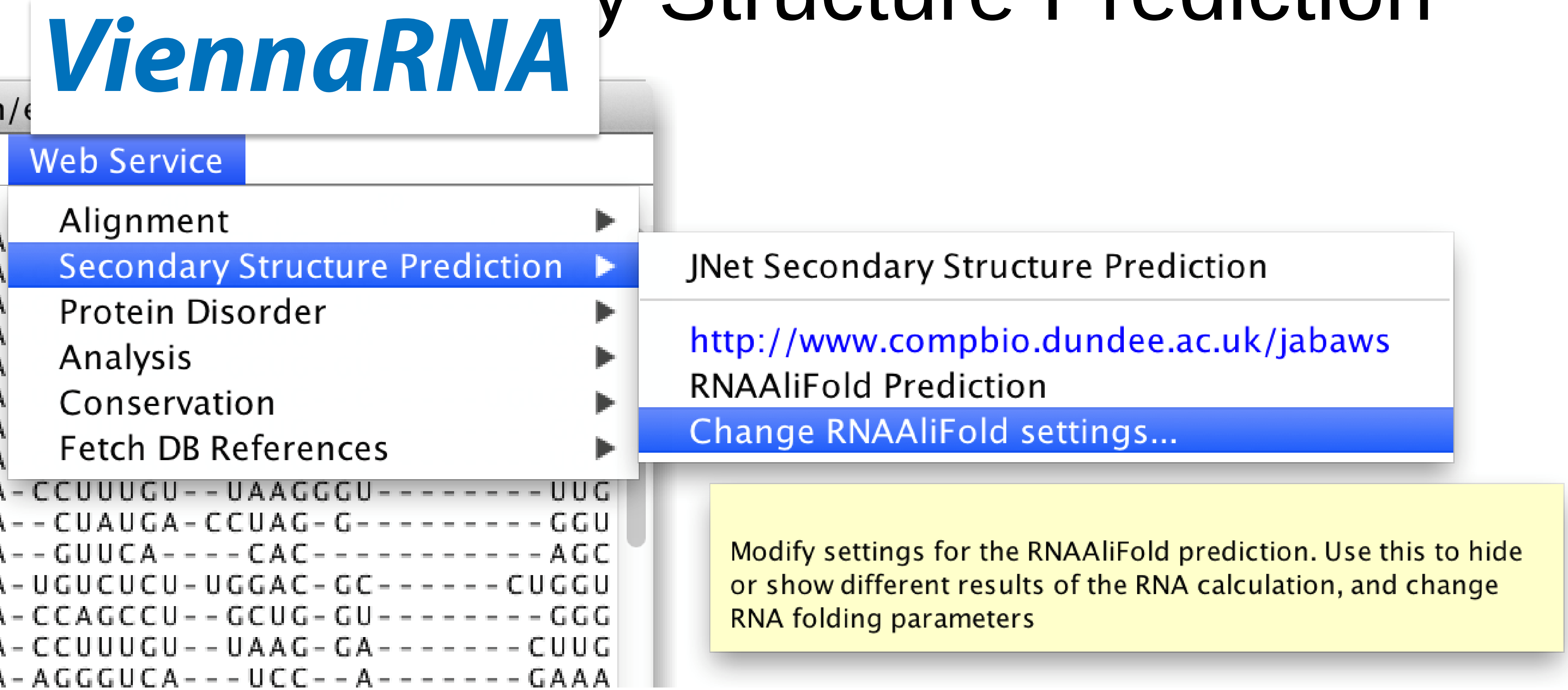
Sequence 25 ID: C.elegans.1 Nucleotide: Cytosine (19)

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



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

2.8.1 - Interactive Alignment based RNA 2nd-ary Structure Prediction

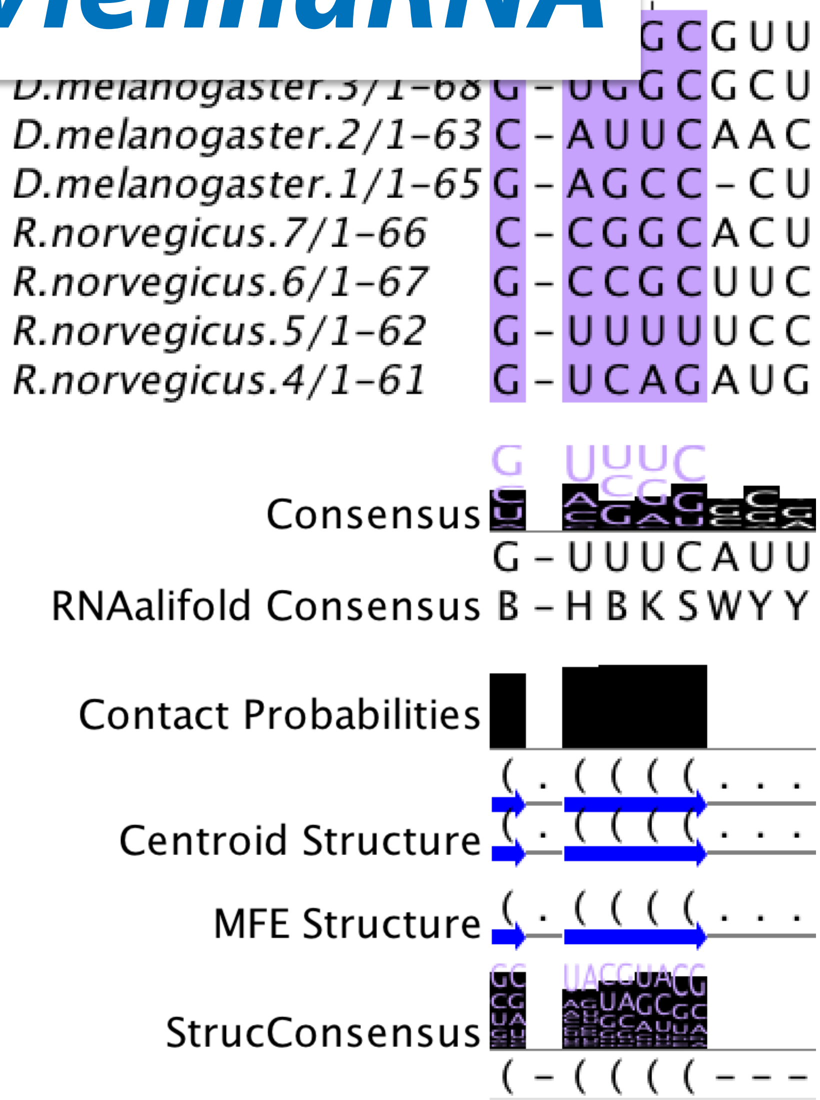


```
-CCUUUGU--UAAGGGU---------UUG
--CUAUGA-CCUAG-G---------GGU
--GUUCA----CAC-----------AGC
-UGUCUCU-UGGAC-GC------CUGGU
-CCAGCCU--GCUG-GU--------GGG
-CCUUUGU--UAAG-GA-------CUUG
-AGGGUCA---UCC--A-------GAAA
```

Modify settings for the RNAAliFold prediction. Use this to hide or show different results of the RNA calculation, and change RNA folding parameters

2.8.1 - Interactive Alignment based RNA 2^{nd}-ary Structure Prediction

ViennaRNA

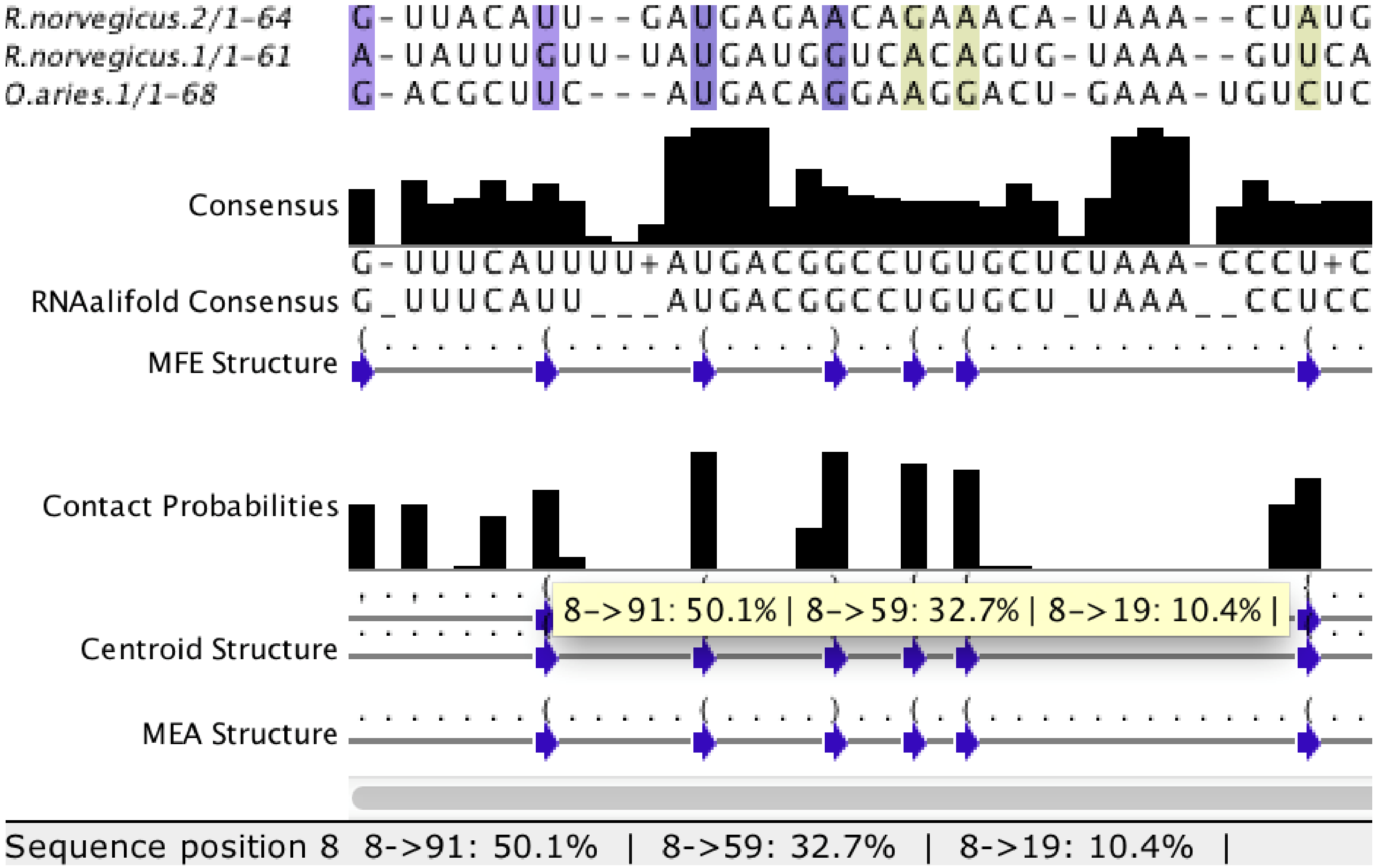


```
                                    GCGUU
D.melanogaster.3/1-68  G - UGGCGCU
D.melanogaster.2/1-63  C - AUUCAAC
D.melanogaster.1/1-65  G - AGCC - CU
R.norvegicus.7/1-66    C - CGGCACU
R.norvegicus.6/1-67    G - CCGCUUC
R.norvegicus.5/1-62    G - UUUUUCC
R.norvegicus.4/1-61    G - UCAGAUG

Consensus
                       G - UUUCAUU
RNAalifold Consensus   B - HBKSWYY

Contact Probabilities
                       ( . ( ( ( ( . . .
Centroid Structure     ( . ( ( ( ( . . .
MFE Structure          ( . ( ( ( ( . . .
StrucConsensus
                       ( - ( ( ( ( - - -
```

- Can be enabled for any view
- Updated if alignment changes
- settings & results saved in Jalview project

Implemented by our 2013 Summer student

Tooltips show alternative base pairs



```
R.norvegicus.2/1-64        G-UUACAUU--GAUGAGAACAGAAACA-UAAA--CUAUG
R.norvegicus.1/1-61        A-UAUUUGUU-UAUGAUGGUCACAGUG-UAAA--GUUCA
O.aries.1/1-68             G-ACGCUUC---AUGACAGGAAGGACU-GAAA-UGUCUC

Consensus                  G-UUUCAUUUU+AUGACGGCCUGUGCUCUAAA-CCCU+C
RNAalifold Consensus       G_UUUCAUU___AUGACGGCCUGUGCU_UAAA__CCUCC
MFE Structure              (.......(......(.....)..(.(..............(..
```

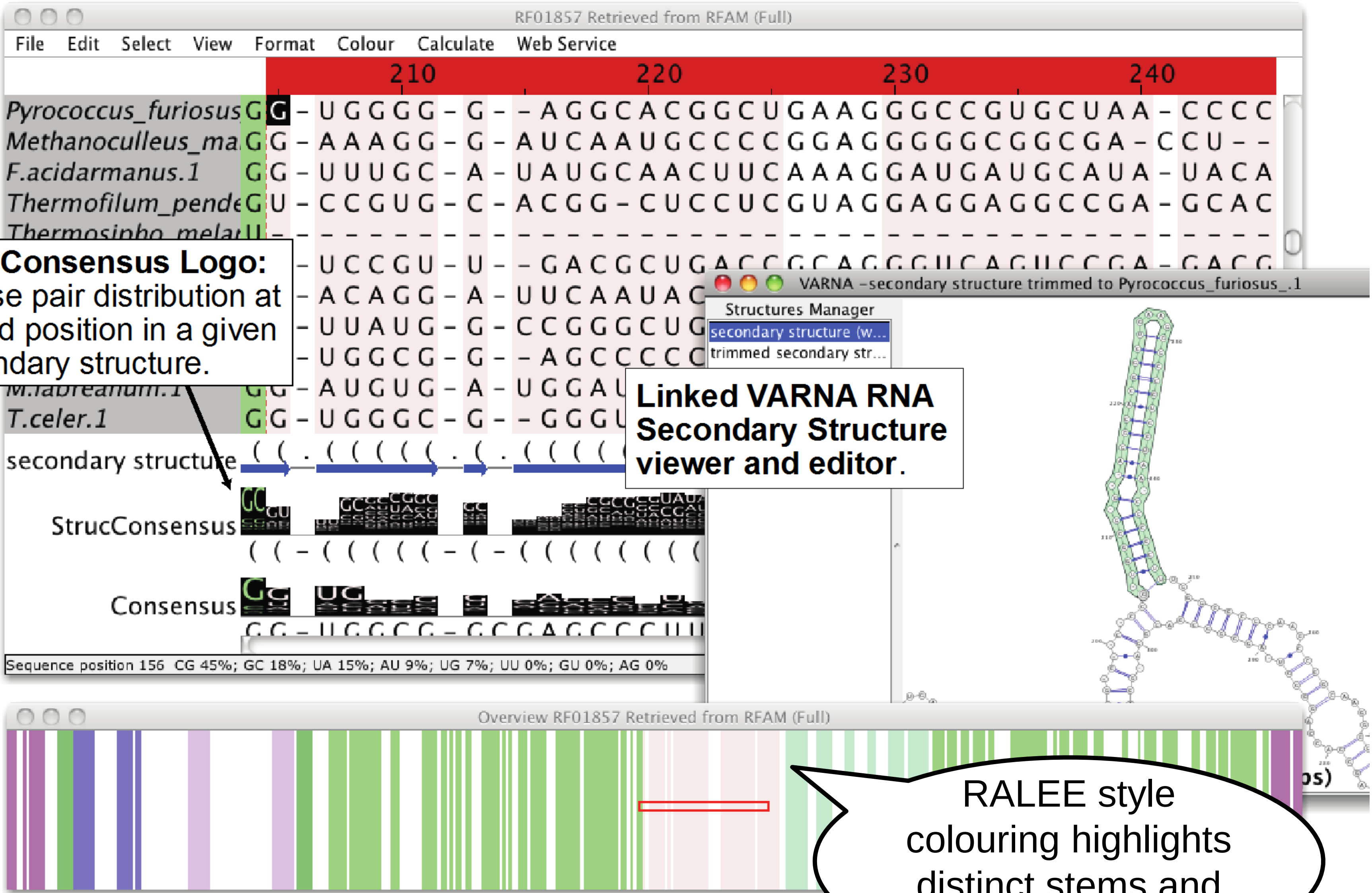
8->91: 50.1% | 8->59: 32.7% | 8->19: 10.4% |

Sequence position 8 8->91: 50.1% | 8->59: 32.7% | 8->19: 10.4% |

DNA and Protein in Jalview

- Discussed in Section 2.10 of manual
- From DNA to Protein
  - Calculations => Translate cDNA
  - View protein annotation on exons using EMBL records
- From protein to DNA
  - Recover DNA for proteins using EMBL cross references

Jalview 2.8 and RNA 2nd-ary



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

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

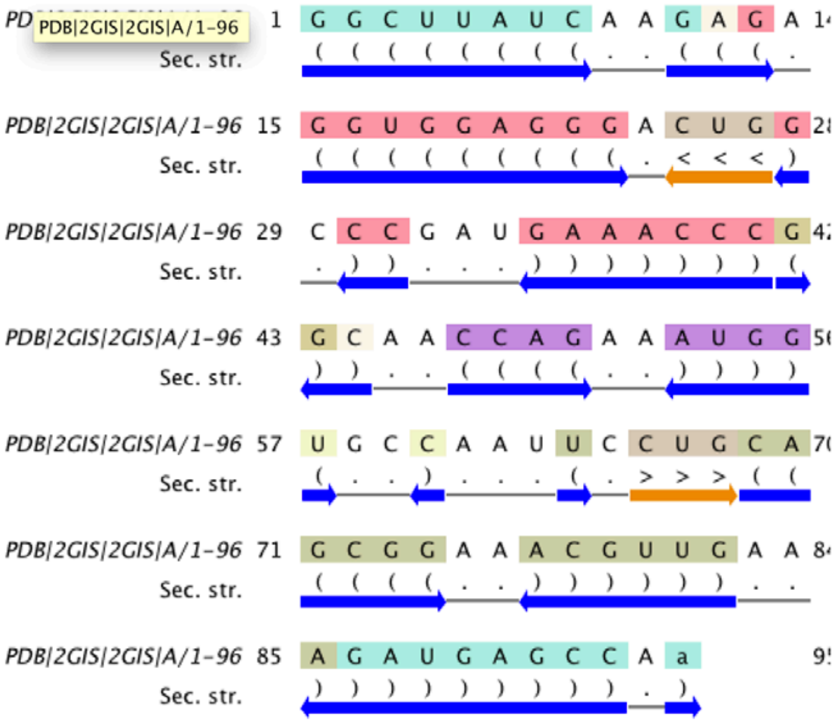
Structure Consensus Logo: Shows base pair distribution at each paired position in a given RNA secondary structure.

Linked VARNA RNA Secondary Structure viewer and editor.

RALEE style colouring highlights distinct stems and helices

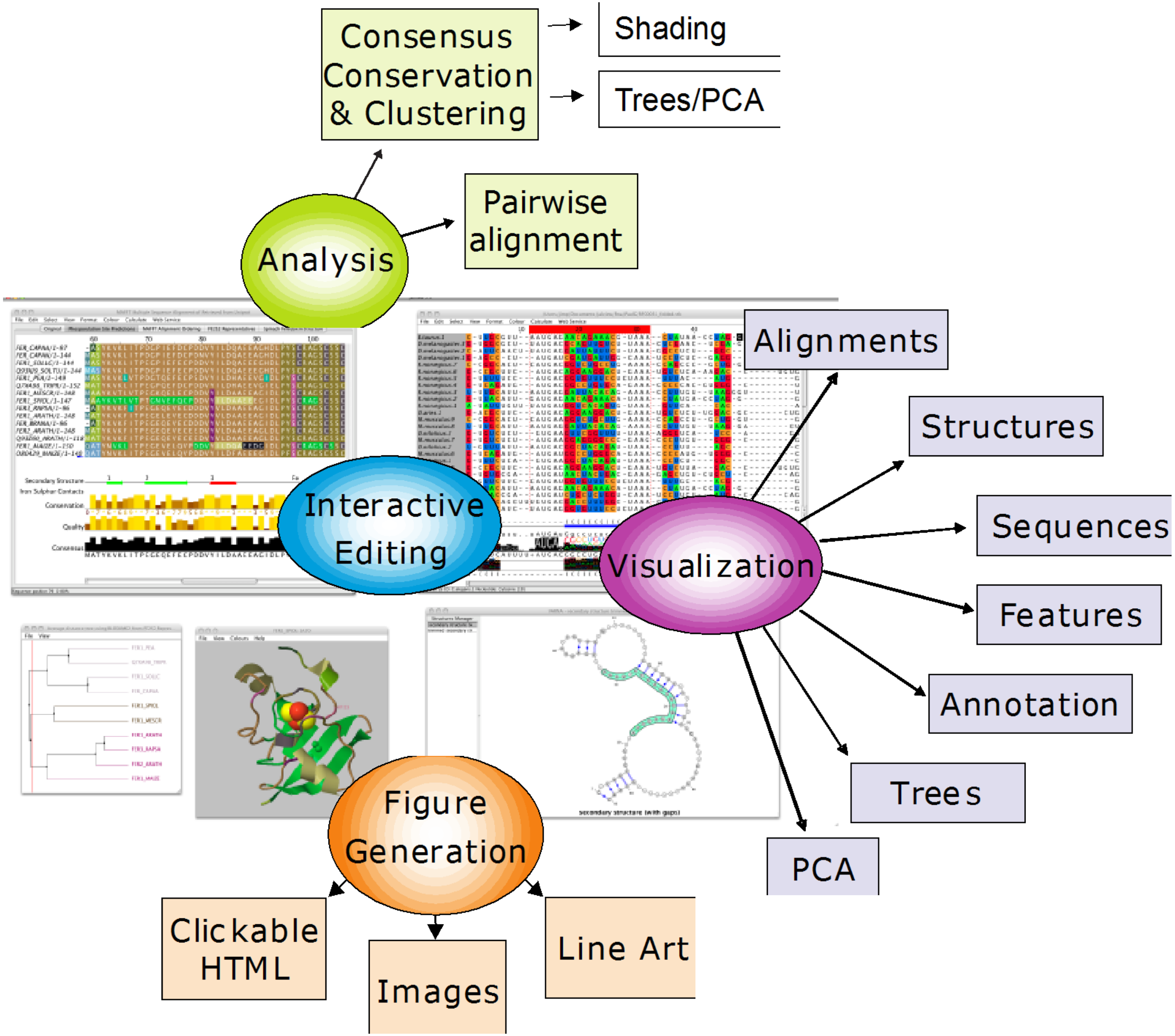
RNAView, pyRNA

- Fabrice Jossinet's pyRNA server includes RNAView*
  - Identify and characterise base pair interactions in 3D structure
- Used by Jalview to obtain secondary structure for RNA 3D data



```
PD PDB|2GIS|2GIS|A/1-96  1   G G C U U A U C A A G A G A 1
             Sec. str.       ( ( ( ( ( ( ( ( . . ( ( ( .

PDB|2GIS|2GIS|A/1-96  15   G G U G G A G G G A C U G G 2
             Sec. str.       ( ( ( ( ( ( ( ( ( . < < < )

PDB|2GIS|2GIS|A/1-96  29   C C C G A U G A A A C C C G 4
             Sec. str.       . ) ) . . . ) ) ) ) ) ) ) (

PDB|2GIS|2GIS|A/1-96  43   G C A A C C A G A A A U G G 5
             Sec. str.       ) ) . . ( ( ( ( . . ) ) ) )

PDB|2GIS|2GIS|A/1-96  57   U G C C A A U U C C U G C A 7
             Sec. str.       ( . . ) . . . ( . > > > ( (

PDB|2GIS|2GIS|A/1-96  71   G C G G A A A C G U U G A A 8
             Sec. str.       ( ( ( ( . . ) ) ) ) ) ) . .

PDB|2GIS|2GIS|A/1-96  85   A G A U G A G C C A a       9
             Sec. str.       ) ) ) ) ) ) ) ) ) . )
```

* RNAView will shortly be replaced by **DSSR** (Xiang-Jun Lu) **http://x3dna.org/index.php**



Analysis

Consensus Conservation & Clustering

Shading

Trees/PCA

Pairwise alignment

Interactive Editing

Visualization

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Figure Generation

Clickable HTML

Images

Line Art

Jalview Flavours

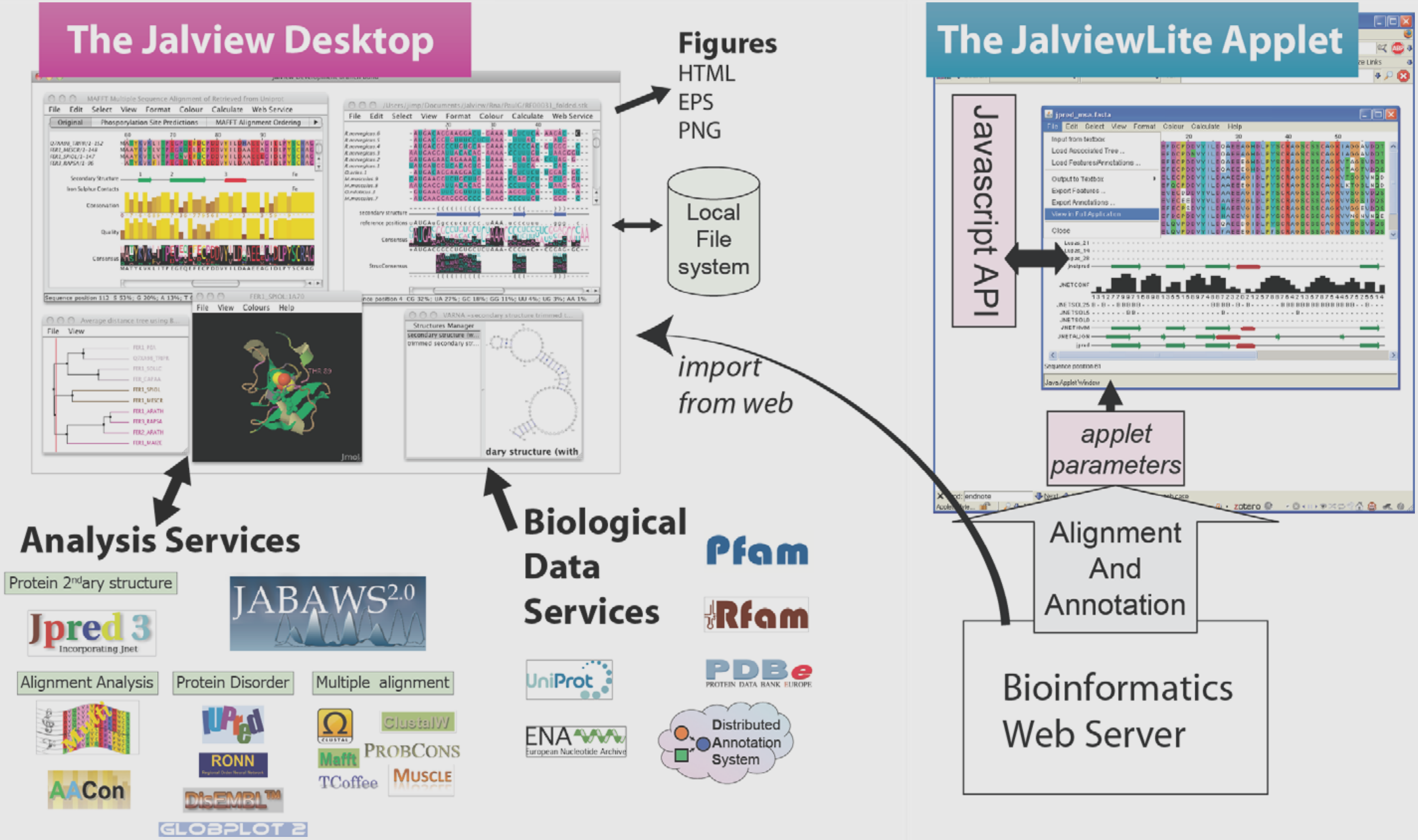
Multi-windowed UI

Visualization & Analysis

Common Data & Analysis
Editing, messaging and
File Input/Output

Lightweight UI

Integrate with web sites



The Jalview Desktop

Figures
HTML
EPS
PNG

Local File system

import from web

Analysis Services

Protein 2ndary structure

Alignment Analysis

Protein Disorder

Multiple alignment

Biological Data Services

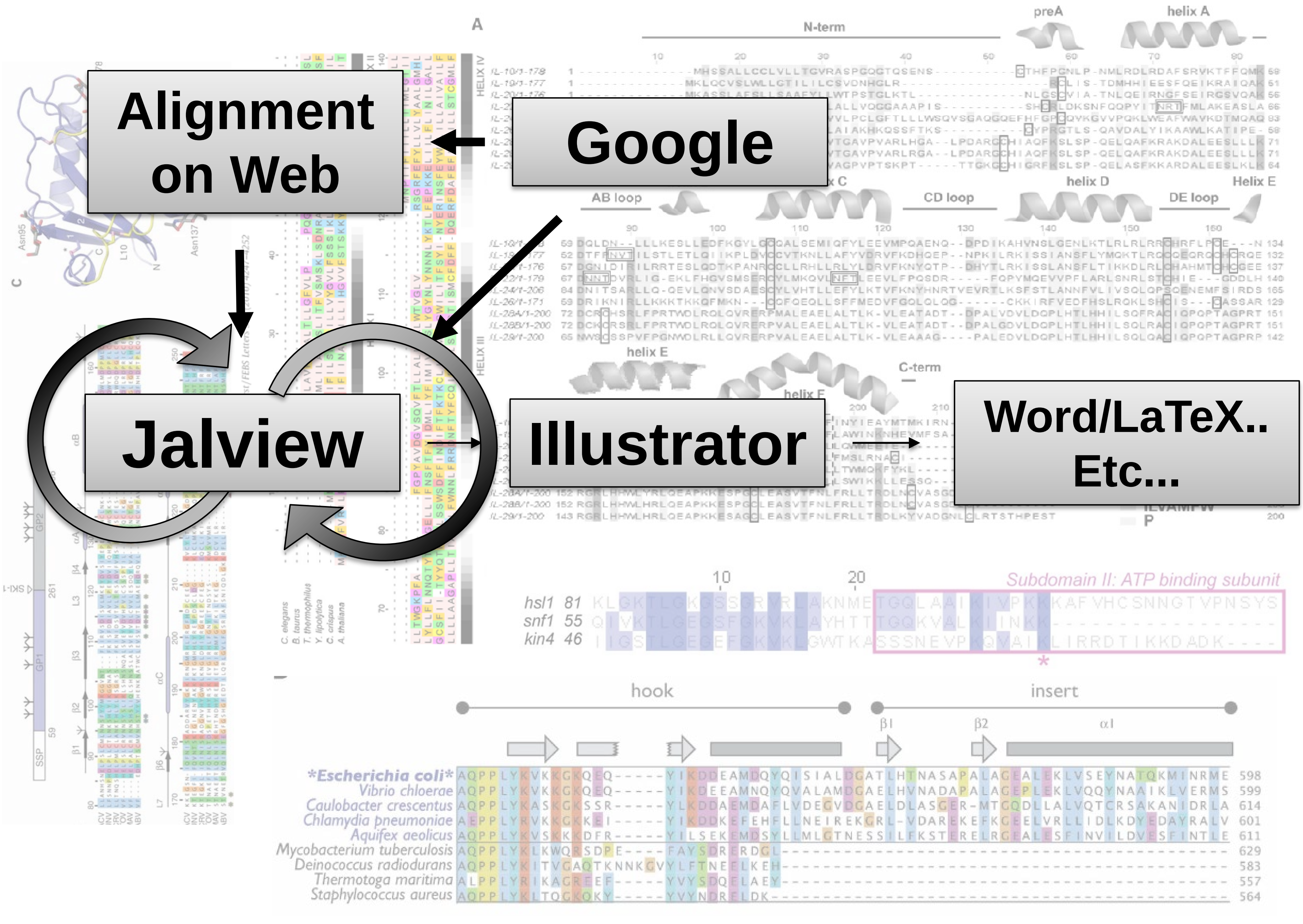
The JalviewLite Applet

Javascript API

applet parameters

Alignment And Annotation

Bioinformatics Web Server



Alignment on Web

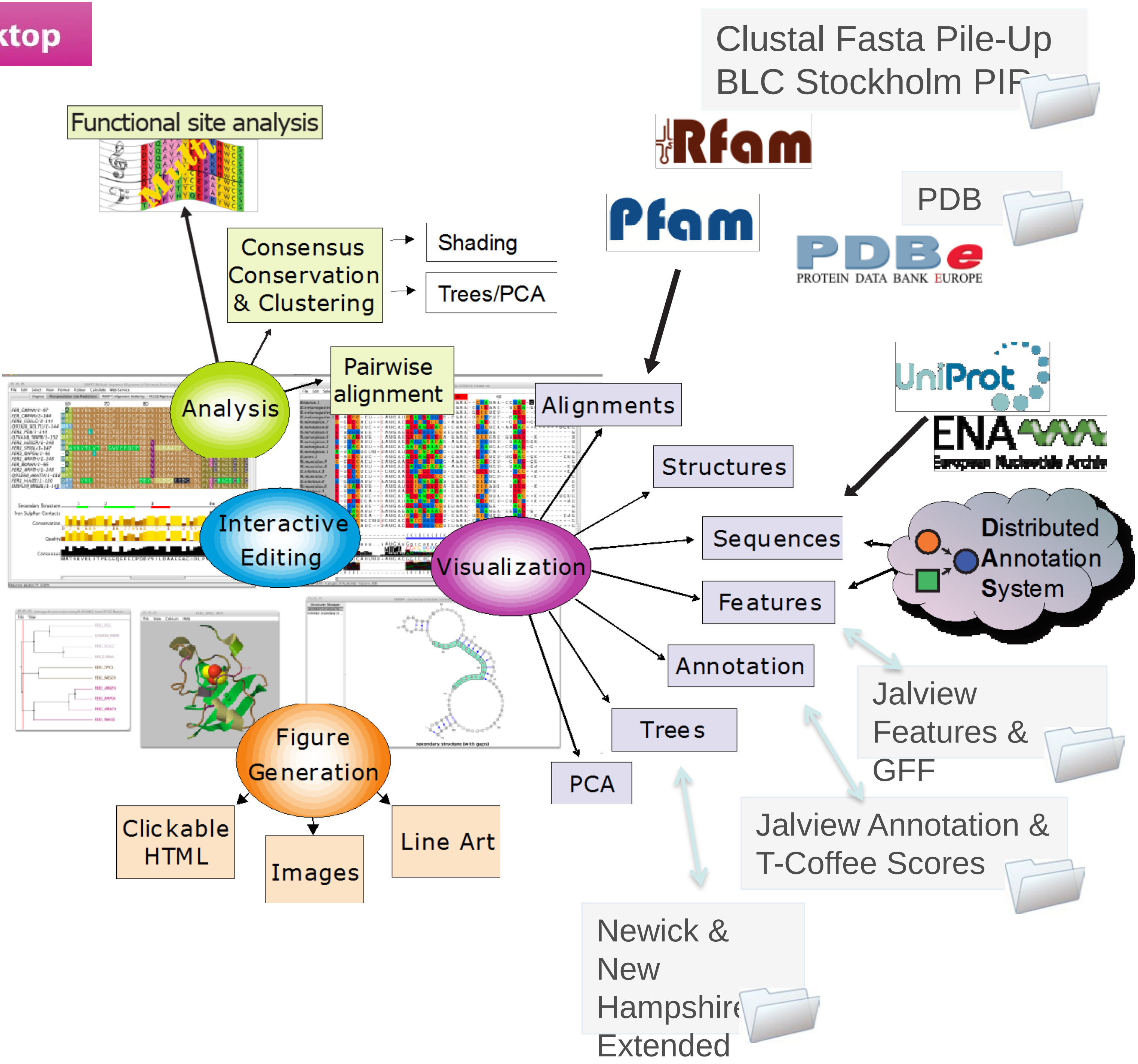
Google

Jalview

Illustrator

Word/LaTeX..
Etc...

The Jalview Desktop



Functional site analysis

Consensus
Conservation
& Clustering

Shading

Trees/PCA

Pairwise
alignment

Analysis

Interactive
Editing

Visualization

Figure
Generation

Clickable
HTML

Images

Line Art

Alignments

Structures

Sequences

Features

Annotation

Trees

PCA

Clustal Fasta Pile-Up
BLC Stockholm PIR

PDB

PDBe
PROTEIN DATA BANK EUROPE

Rfam

Pfam

UniProt

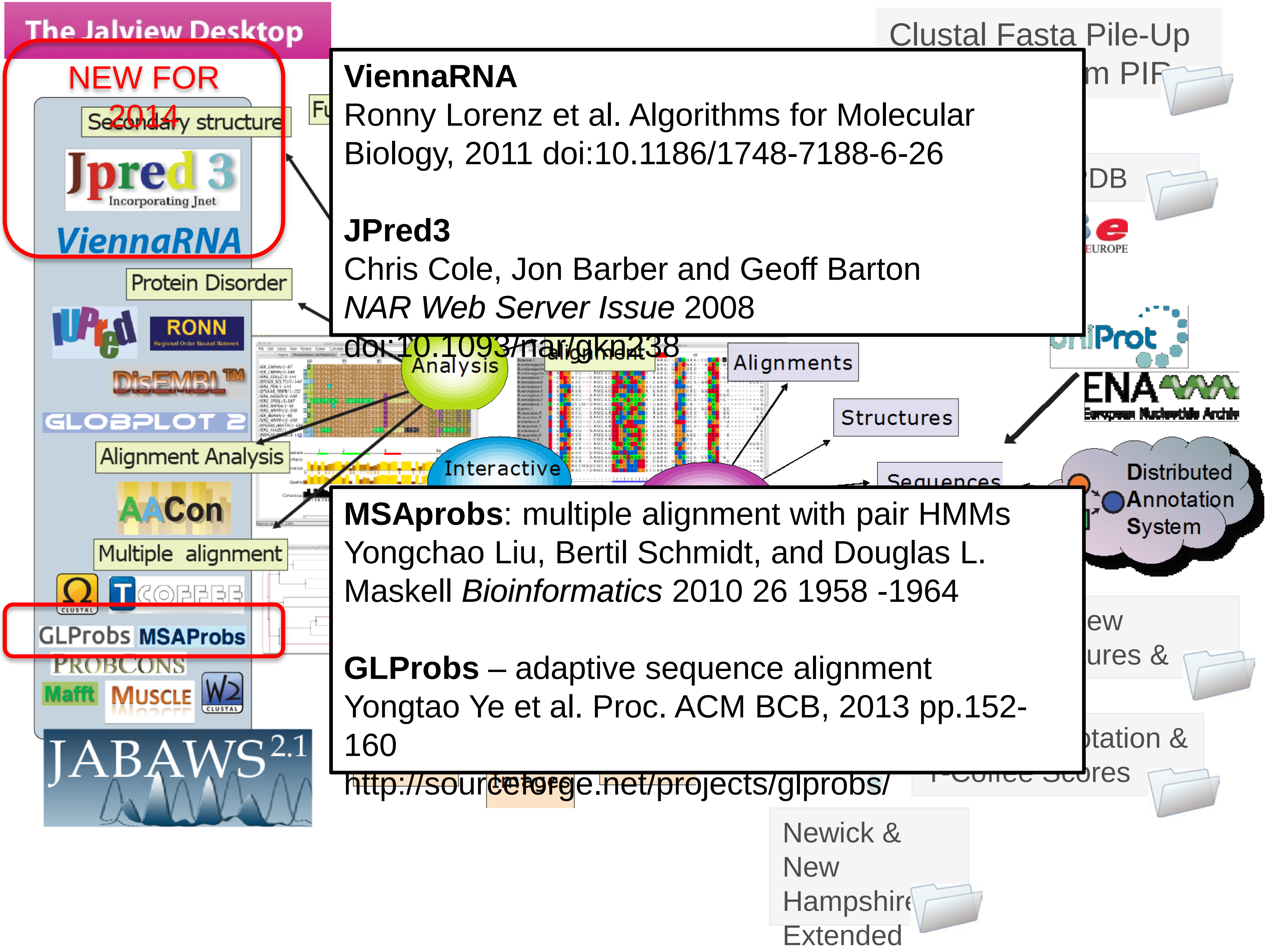
ENA
European Nucleotide Archive

Distributed
Annotation
System

Jalview
Features &
GFF

Jalview Annotation &
T-Coffee Scores

Newick &
New
Hampshire
Extended



ViennaRNA
Ronny Lorenz et al. Algorithms for Molecular Biology, 2011 doi:10.1186/1748-7188-6-26

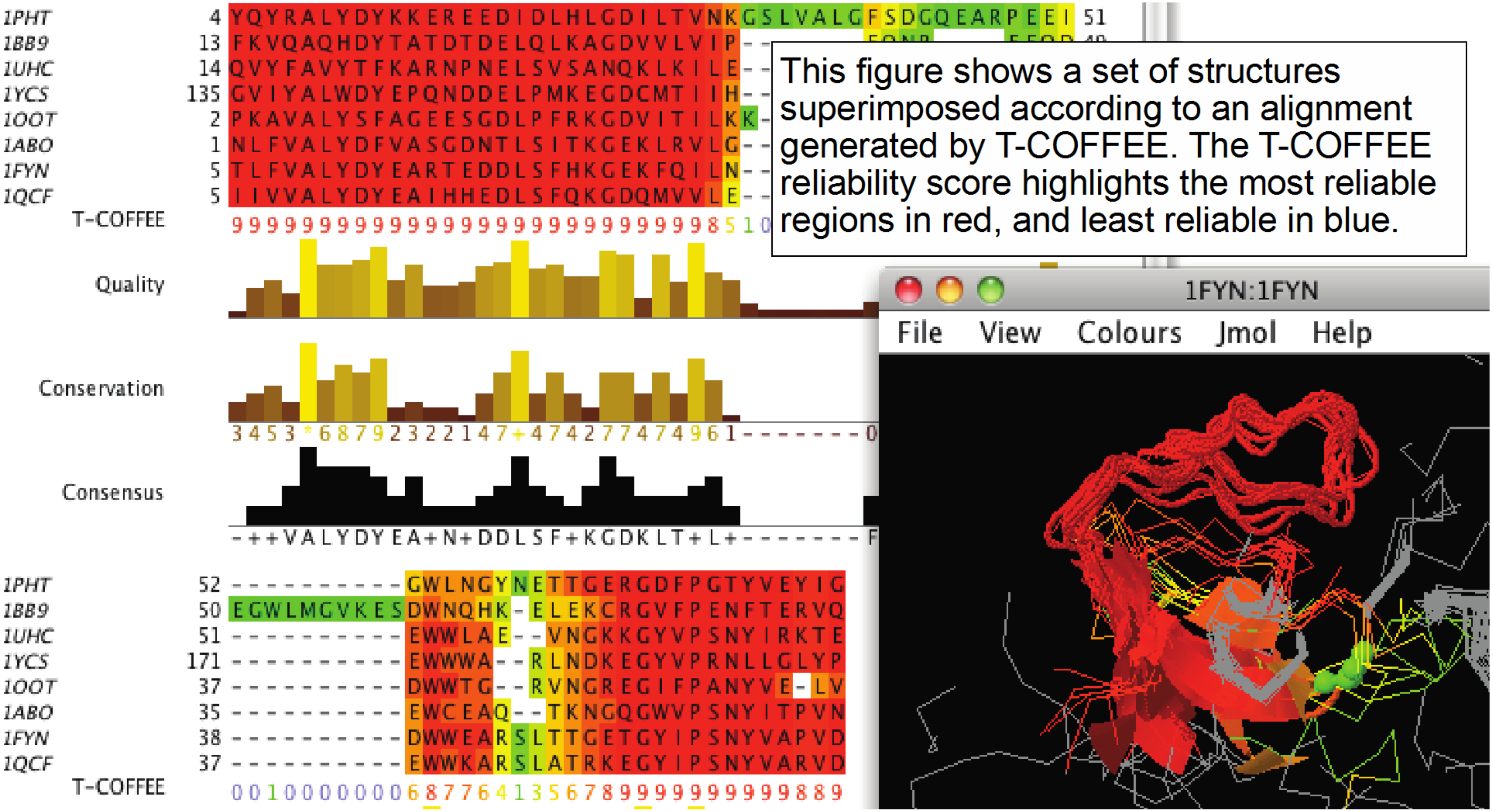
JPred3
Chris Cole, Jon Barber and Geoff Barton
NAR Web Server Issue 2008
doi:10.1093/nar/gkn238

MSAprobs: multiple alignment with pair HMMs
Yongchao Liu, Bertil Schmidt, and Douglas L. Maskell *Bioinformatics* 2010 26 1958 -1964

GLProbs – adaptive sequence alignment
Yongtao Ye et al. Proc. ACM BCB, 2013 pp.152-160
http://sourceforge.net/projects/glprobs/

Things I haven’t talked about …

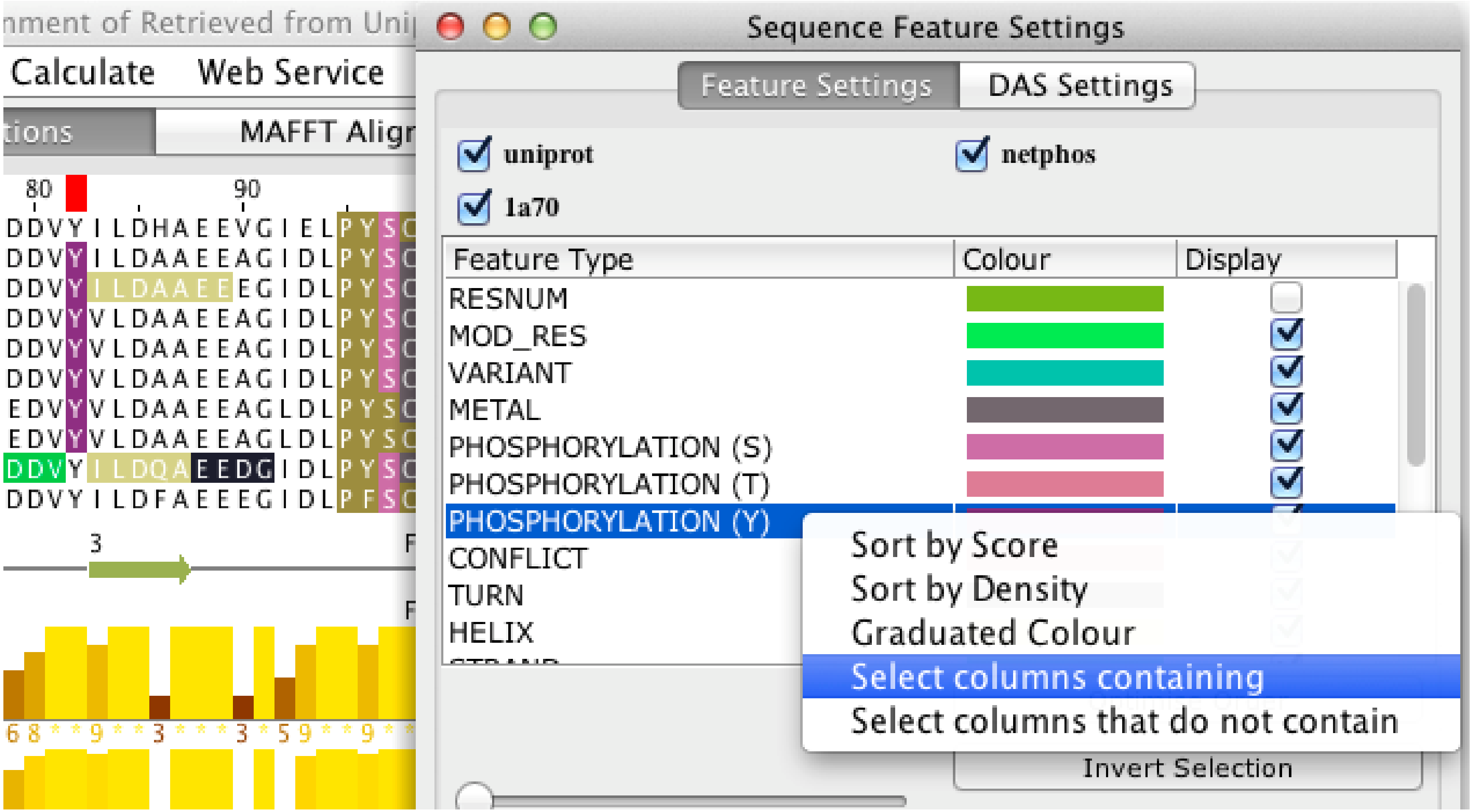
T-COFFEE alignment reliability scores



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

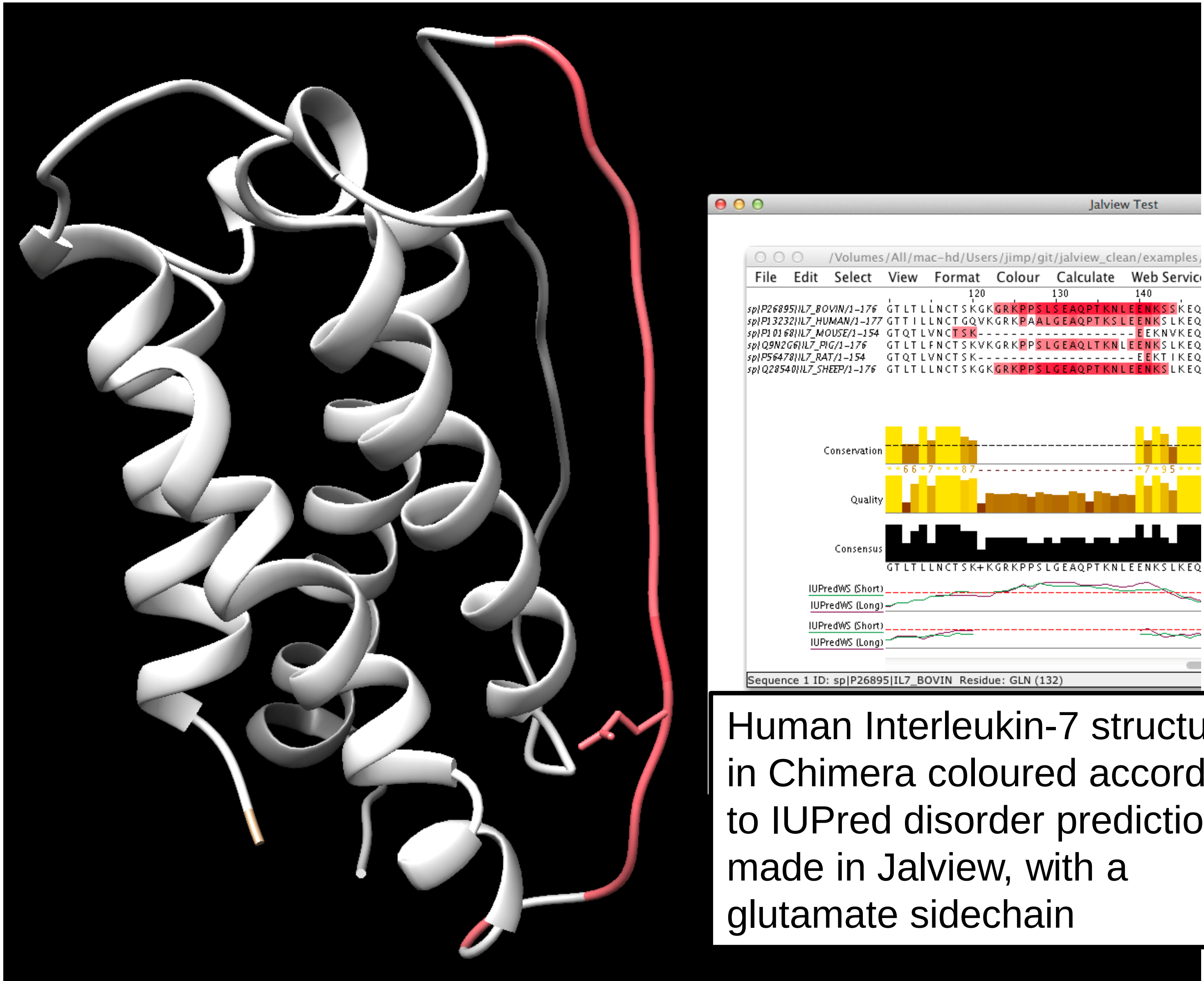
```
1PHT      4 YQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSDGQEARPEEI 51
1BB9     13 FKVQAQHDYTATDTDELQLKAGDVVLVIP--
1UHC     14 QVYFAVYTFKARNPNELSVSANQKLKILE--
1YCS    135 GVIYALWDYEPQNDDELPMKEGDCMTIIH--
1OOT      2 PKAVALYSFAGEESGDLPFRKGDVITILKK-
1ABO      1 NLFVALYDFVASGDNTLSITKGEKLRVLG--
1FYN      5 TLFVALYDYEARTEDDLSFHKGEKFQILN--
1QCF      5 IIVVALYDYEAIHHEDLSFQKGDQMVVLE--
    T-COFFEE 99999999999999999999999999851 0

     Quality

Conservation
            3453*68792322147+4742774749 61-------0

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

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

Select column by feature



Feature Type	Colour	Display
RESNUM	■	☐
MOD_RES	■	☑
VARIANT	■	☑
METAL	■	☑
PHOSPHORYLATION (S)	■	☑
PHOSPHORYLATION (T)	■	☑
PHOSPHORYLATION (Y)	■	☑
CONFLICT		
TURN		
HELIX		

Jalview and Chimera

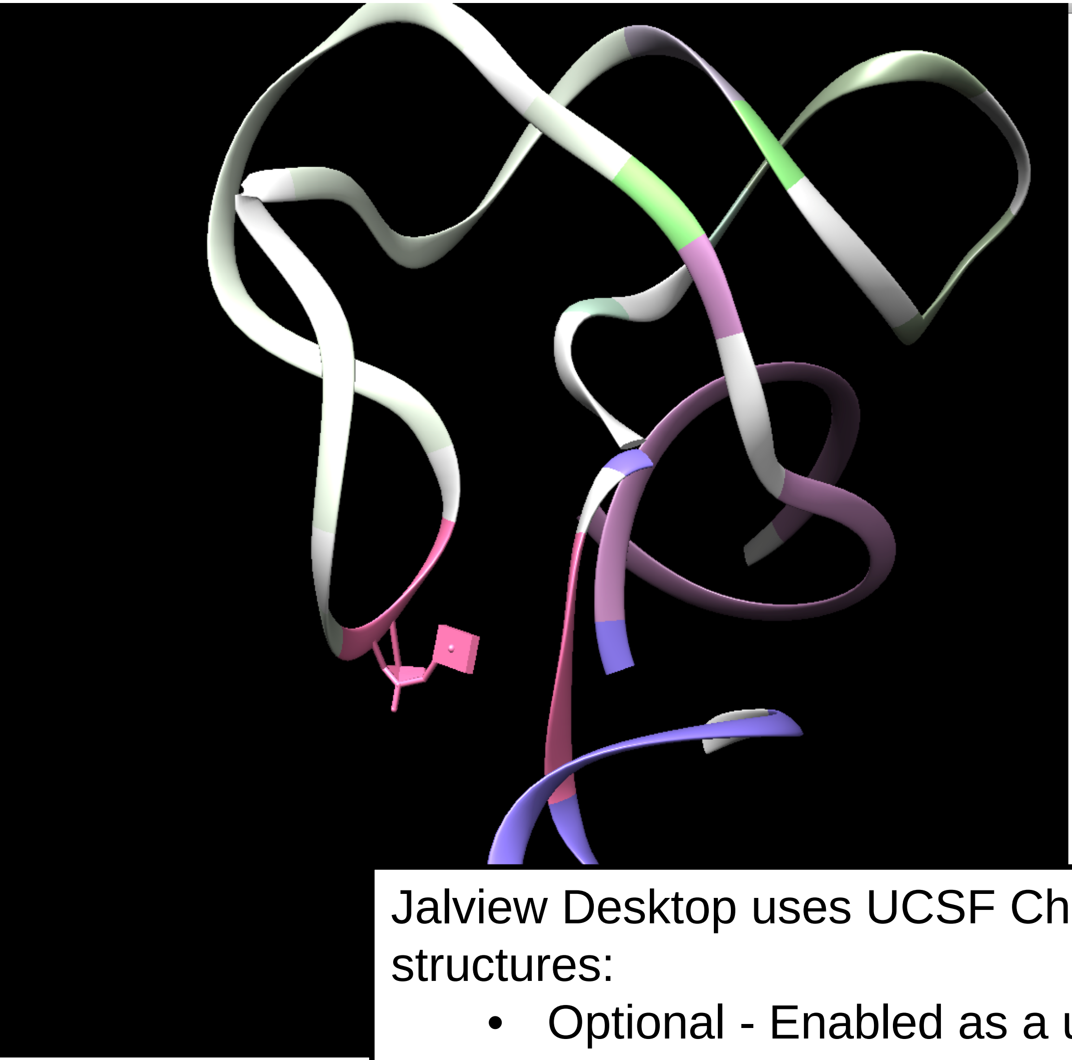


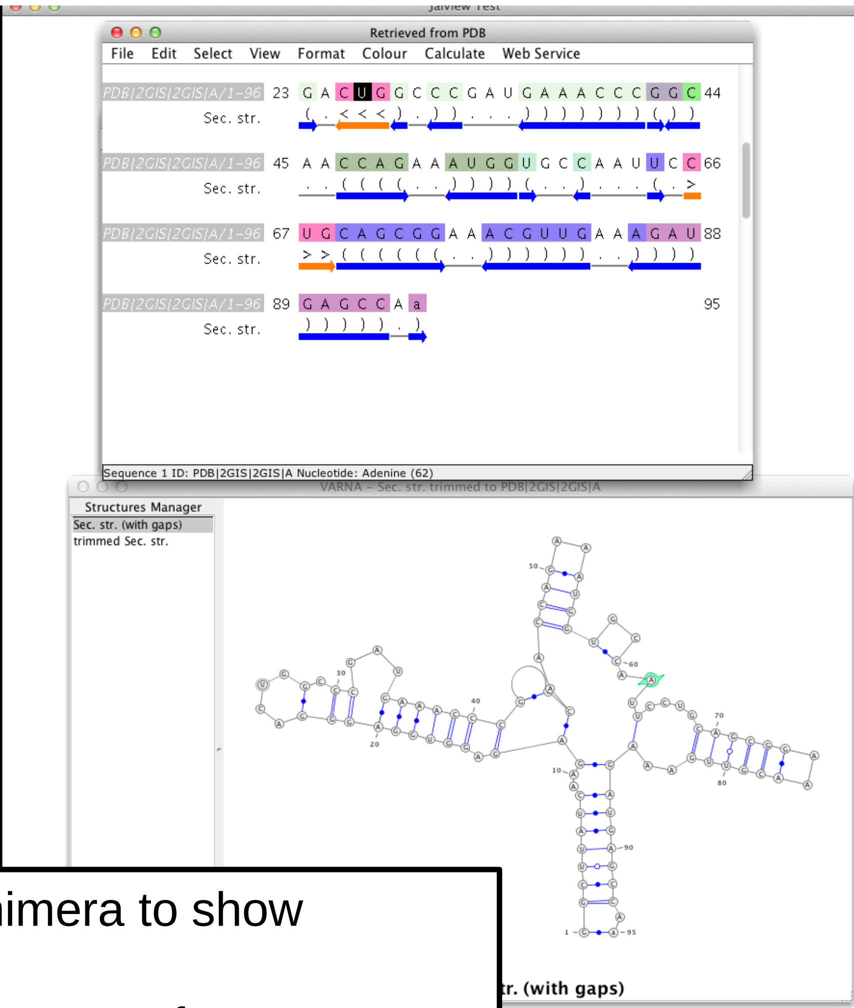
```
                          120         130         140
sp|P26895|IL7_BOVIN/1-176  GTLTLLNCTSKGKGRKPPSLSEAQPTKNLEENKSSKEQ
sp|P13232|IL7_HUMAN/1-177  GTTILLNCTGQVKGRKPAALGEAQPTKSLEENKSLKEQ
sp|P10168|IL7_MOUSE/1-154  GTQTLVNCTSK-----------------------EEKNVKEQ
sp|Q9N2G6|IL7_PIG/1-176    GTLTLFNCTSKVKGRKPPSLGEAQLTKNLEENKSLKEQ
sp|P56478|IL7_RAT/1-154    GTQTLVNCTSK-----------------------EEKTIKEQ
sp|Q28540|IL7_SHEEP/1-176  GTLTLLNCTSKGKGRKPPSLGEAQPTKNLEENKSLKEQ

**66*7***87------------------*7*95***
GTLTLLNCTSK+KGRKPPSLGEAQPTKNLEENKSLKEQ
```

Sequence 1 ID: sp|P26895|IL7_BOVIN Residue: GLN (132)

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

Jalview + VARNA + Chimera





```
PDB|2GIS|2GIS|A/1-96  23  G A C U G G C C C G A U G A A A C C C G G C 44
           Sec. str.      ( . < < < ) . ) ) . . . ) ) ) ) ) ) ) ( ) )
PDB|2GIS|2GIS|A/1-96  45  A A C C A G A A A U G G U G C C A A U U C C 66
           Sec. str.      . . ( ( ( ( . . ) ) ) ) ( . . ) . . . ( . >
PDB|2GIS|2GIS|A/1-96  67  U G C A G C G G A A A C G U U G A A A G A U 88
           Sec. str.      > > ( ( ( ( ( ( . . ) ) ) ) ) ) . . ) ) ) )
PDB|2GIS|2GIS|A/1-96  89  G A G C C A a                               95
           Sec. str.      ) ) ) ) ) . )
```

Sequence 1 ID: PDB|2GIS|2GIS|A Nucleotide: Adenine (62)

Jalview Desktop uses UCSF Chimera to show structures:

- Optional - Enabled as a user preference
- Structures coloured & superposed like with Jmol
- Positional highlighting from Jalview->Chimera

Jalview & Dundee Protein Resource Developers

Mungo
Carstairs

Tochukwu
(Charles)
Ofoegbu

Alexey
Drozdetskiy
Jnet Protein
prediction and
analysis services

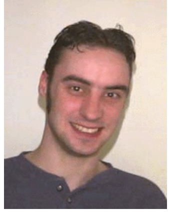
supported by
wellcome trust

Jalview
www.jalview.org

BBSRC
bioscience for the future

<u>The Jalview developers</u>

Michele Clamp
Harvard & MIT.



James Cuff
Harvard & MIT

Steve Searle
Sanger, UK

Andrew Waterhouse
Basel, Switzerland.

<u>RNA Features</u>

Lauren Lui
UC Santa Cruz, USA.

Jan Engelhardt
Univ. Leipzig, Germany.

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

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

Google
SUMMER OF CODE
2011

<u>T-COFFEE Scores</u>

Paolo di Tomasso
Notredame Group, CRG, Spain.

Geoff Barton

David Martin (**Teaching**)

Sasha Sherstnev (**JABAWS**)

Peter Troshin (**JABAWS**)

Barry Strachan (**logo**)

Tom Walsh (**Apache**)

Ryan Maclaughlan (**CSS**)

Andrew Millar (**Drupal**)

All the Jalview users, and …

Jalview
www.jalview.org

supported by
wellcome trust

BBSRC
bioscience for the future