

NUCLEOTIDE SEQUENCES AND JALVIEW

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 (European Nucleotide Archive) records
 - From protein to DNA
 - Recover DNA for proteins using EMBL cross references
- Defer till later: New features Jalview 2.9!

Introduction to RNA structure

Slides from Yann Ponty

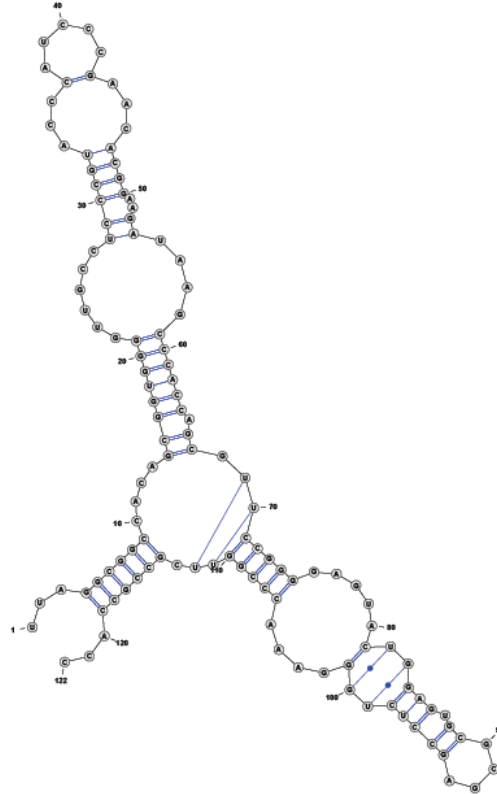
VIZBI RNA Visualisation tutorial 2012

*[www.lix.polytechnique.fr/~ponty/talks/
VIZBI-2012-Tutorial-RNA.pptx](http://www.lix.polytechnique.fr/~ponty/talks/VIZBI-2012-Tutorial-RNA.pptx)*

RNA structure(s)

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC
```

Primary structure



Secondary structure



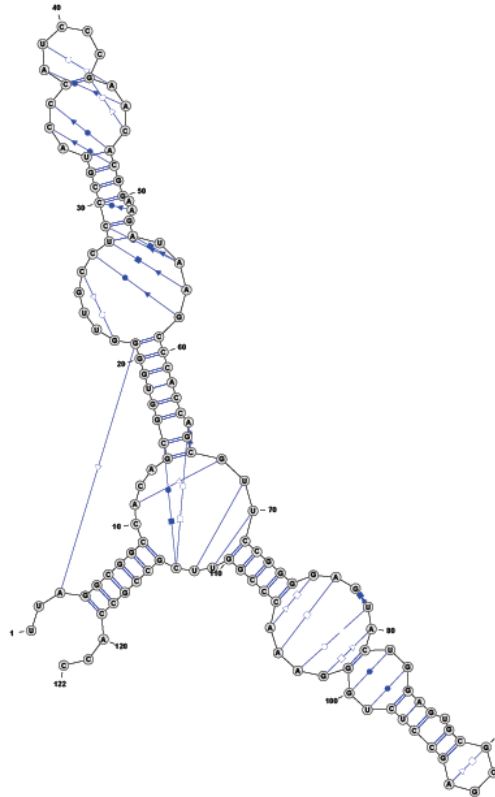
Tertiary structure

Source: 5s rRNA (PDBID: 1K73:B)

RNA structure(s)

```
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAAGCC
CACCAGCGUUCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC
```

Primary structure



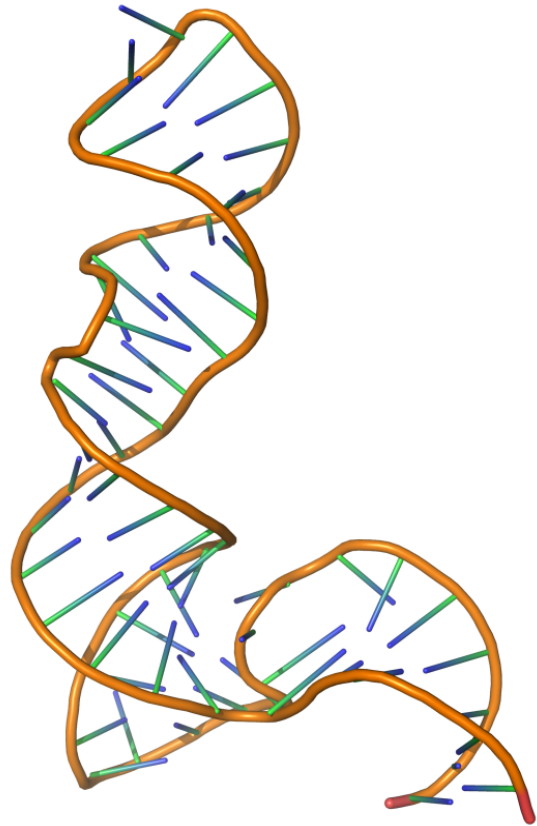
Secondary⁺ structure



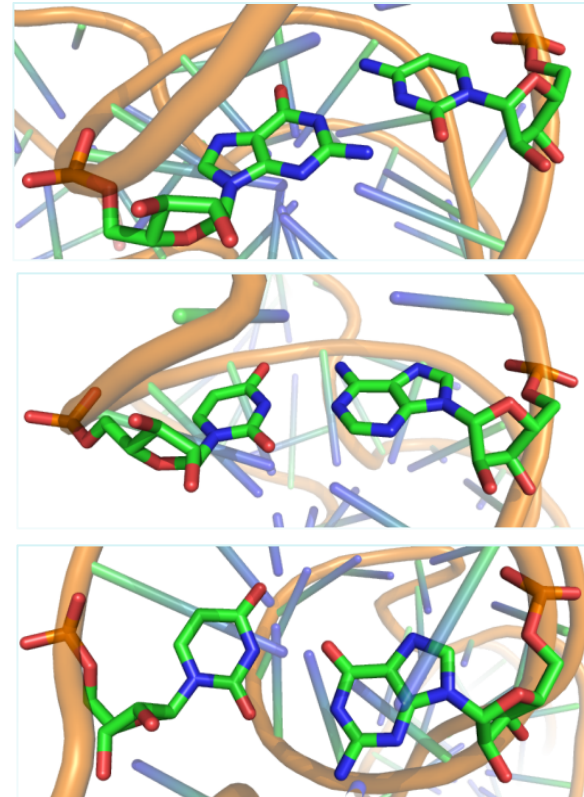
Tertiary structure

Source: 5s rRNA (PDBID: 1K73:B)

How RNA folds



5s rRNA (PDB ID: 1UN6)



G/C

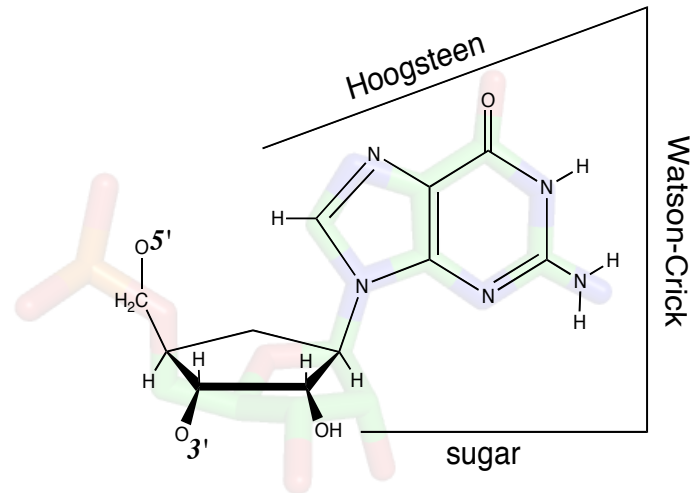
U/A

U/G

Canonical base-pairs

RNA folding = Hierarchical stochastic process driven by/ resulting in the pairing (hydrogen bonds) of a subset of its bases.

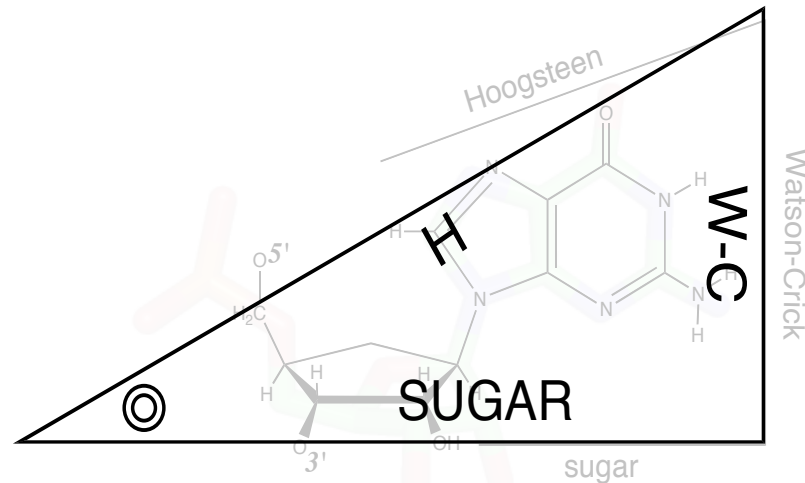
Non canonical interactions



RNA nucleotides bind through edge/edge interactions.

Non canonical are weaker, but cluster into modules that are **structurally constrained**, **evolutionarily conserved**, and **functionally essential**.

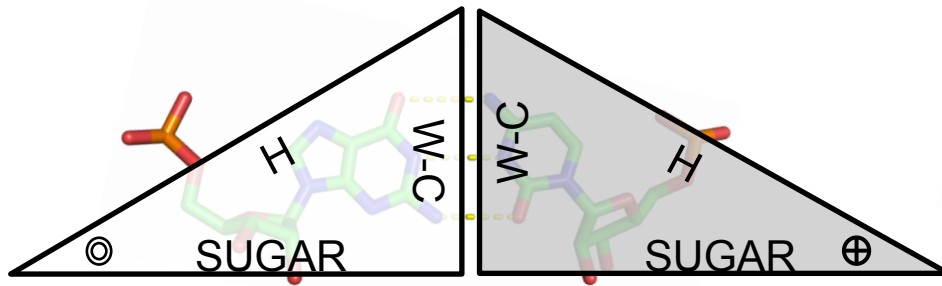
Non canonical interactions



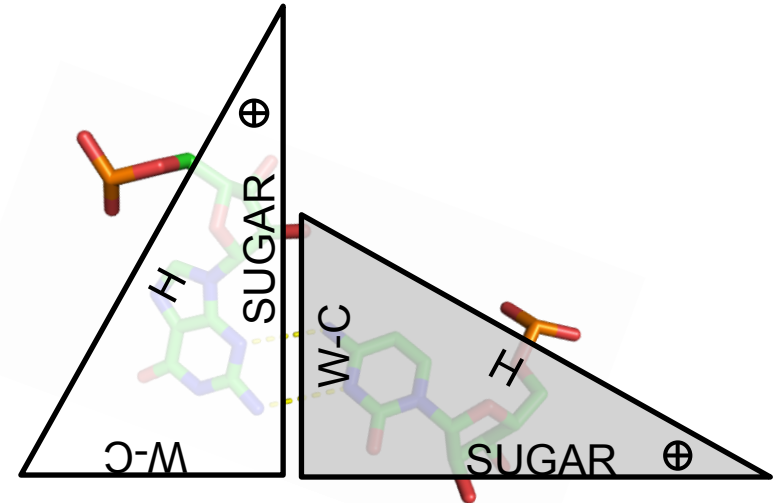
RNA nucleotides bind through edge/edge interactions.

Non canonical are weaker, but cluster into modules that are **structurally constrained**, **evolutionarily conserved**, and **functionally essential**.

Non canonical interactions



Canonical G/C pair
(WC/WC cis)

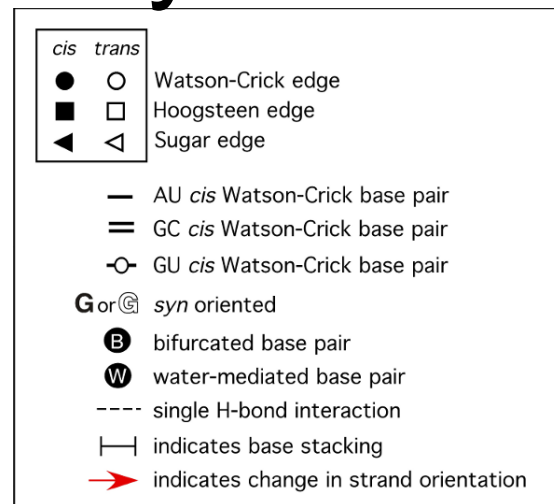
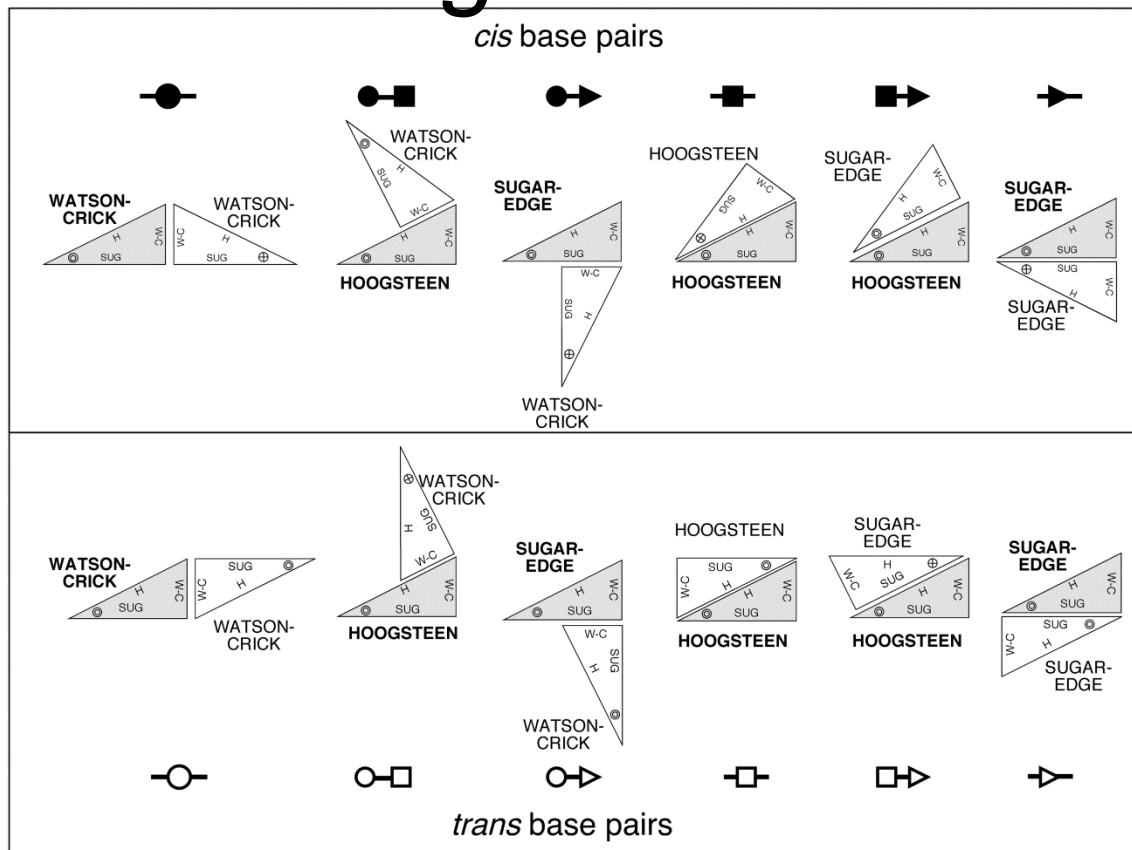


Non Canonical G/C pair
(Sugar/WC trans)

RNA nucleotides bind through edge/edge interactions.

Non canonical are weaker, but cluster into modules that are **structurally constrained**, **evolutionarily conserved**, and **functionally essential**.

Leontis/Westhof nomenclature: A *visual grammar* for tertiary motifs



Leontis/Westhof,
NAR 2002

+ Tools to infer base-pairs from experimentally-derived 3D models

RNAView, MC-Annotate...

Jalview 2.8 and RNA 2nd-ary

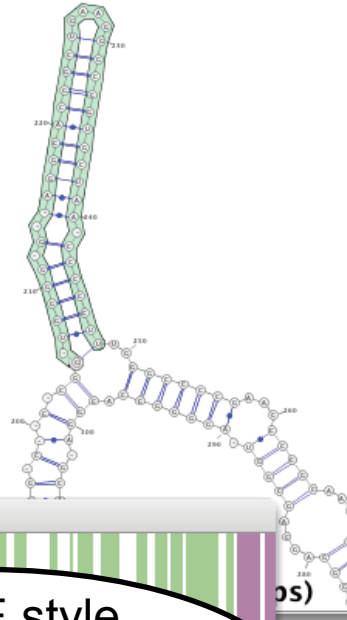
RF01857 Retrieved from RFAM (Full)

	210	220	230	240
<i>Pyrococcus_furiosus</i>	G	- UGGGG - G - -	A GGCACGGCUGAAG	GGCCGUGC UAA - CCCC
<i>Methanoculleus_ma</i>	G	- AAAGG - G -	AUCAAU GCGCCGGAG	GGGGGCGGCGA - CCU - -
<i>F.acidarmanus.1</i>	G	- UUUGC - A -	UAUGCAACUUC AAAG	GAUGAUGCAUA - UACA
<i>Thermofilum_pende</i>	G	- CCGUG - C -	ACGG - CUCCUCGUAG	GAGGAGGCCGA - GCAC
<i>Thermosiphonia_melan...</i>	U	- - - - - - - -	- - - - - - - -	- - - - - - - -
		- UCCGU - U - -	GACGCUGACC	GCGAGGGGUCAGUCCGA - GACC
		- ACAGG - A -	UUCAAUAC	
		- UUAUG - G -	CCGGGCUG	
		- UGGCG - G - -	AGCCCCC	
		- AUGUG - A -	UGGAL	
		- UGGGC - G -	GGGU	
secondary structure	((((((((
StrucConsensus	GC	GU	GC	GC
Consensus	GC	GU	GC	GC

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 VARNAs RNA Secondary Structure viewer and editor.



Overview RF01857 Retrieved from RFAM (Full)



RALEE style colouring highlights distinct stems and helices

RNA 2nd-ary Structure Prediction

ViennaRNA

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

JNet Secondary Structure Prediction

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

RNAAliFold Prediction

Change RNAAliFold settings...

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

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

RNAAlifold: A Consensus Method

ViennaRNA

```

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 - UUUU UCC
R.norvegicus.4/1-61  G - UCAGAUG
    
```

```

Consensus G UUUC
           UG ACCG
           G - UUUCAUU
    
```

RNAalifold Consensus B - H B K S W Y Y

Contact Probabilities

Centroid Structure

```

( . ( ( ( ( . . .
| | | | |
| | | | |
    
```

MFE Structure

```

( . ( ( ( ( . . .
| | | | |
| | | | |
    
```

StrucConsensus

```

GC UACCUACC
CG UAGCC
UA AGCC
UU GCAUUA
( - ( ( ( - - -
    
```

- Quick
- Predictions update when alignment changes
- settings & results saved in Jalview project

- Secondary Structure Prediction
- Protein Disorder
- Analysis
- Conservation
- Fetch DB References

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

- ✓ RNAAliFold Prediction
- Change RNAAliFold settings...

JPred Conserved
Change JPred

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

```

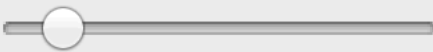
-CCUUUGU--UAAGGGU-----UUG
--CUAUGA-CCUAG-G-----GGU
--GUUCA---CAC-----AGC
-UUUUUUU-UUUUUUU-UUUUUUU
  
```

When 'partition function' enabled, the contact probabilities are displayed

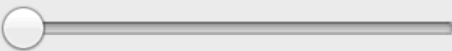
- MEA Structure
- No Closing GU
- No GU
- old
- d2
- Partition Function
- Ribosum Scoring
- Most Informative Sequence
- No LP
- Endgaps
- G-Quadruplex
- Circular

Parameters

Temperature



bppmThreshold

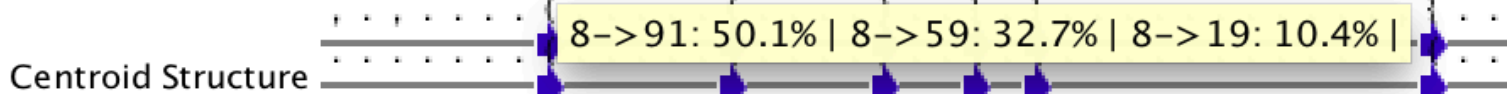
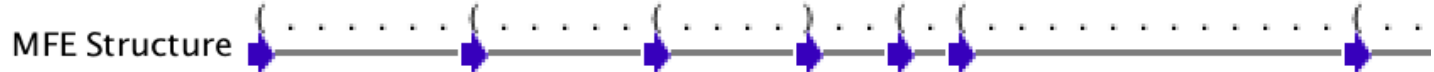


Tooltips provide additional RNAAlifold information

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



RNAalifold Consensus G_UUUCAUU__AUGACGGCCUGUGCU_UAAA__CCUCC

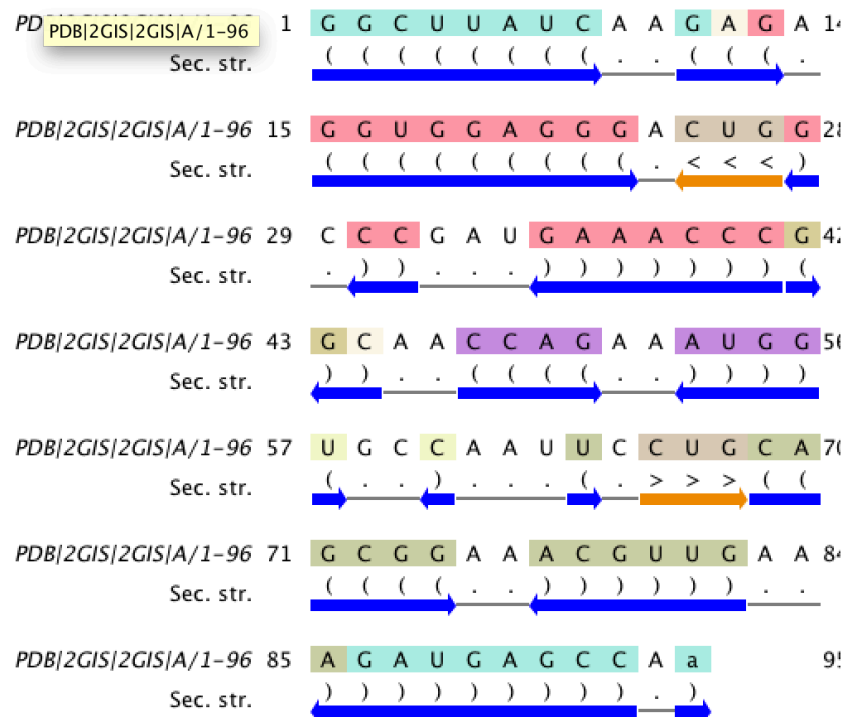


Sequence position 8 8->91: 50.1% | 8->59: 32.7% | 8->19: 10.4% |

“2.5D” RNA structure

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



* RNAView will shortly be replaced by **DSSR** (Xiang-Jun Lu)

<http://x3dna.org/index.php>

Exercise 33

- Viewing RNA structures
 - RFAM alignment
 - Colour by helices
 - Apply RNAAlifold
 - View 3D structure and 2.5D structure

Protein Feature visualization on DNA

Section 2.10, exercise 33

- Task
 - Retrieve a DNA contig and visualize features from UNIPROT at their coding positions.
- Question
 - What fields in an EMBL sequence record can Jalview use ?

Semantic Processing: Database Reference Tracing

'get me the sequences from database blah for the selected sequences'

The screenshot shows a window titled "Retrieved from EMBL" with a menu bar including File, Edit, Select, View, Format, Colour, Calculate, and Web Service. The "Calculate" menu is open, showing options like Sort, Calculate Tree, Pairwise Alignments..., Principal Component Analysis, Translate cDNA, Get Cross References, Autocalculate Consensus, and Extract Scores... The "Get Cross References" sub-menu is also open, showing PDB and UNIPROT. A list of sequences is visible on the left, with EMBL|U07178/1-1668 highlighted. A yellow box highlights the details for this sequence, including its description and various database references. A status bar at the bottom shows "Sequence 8 ID: EMBL|M22585 Nucleotide: Guanine (1)".

Sequence ID	Nucleotide
EMBL X04752/1-1167	caacagttatcctg
EMBL U07177/1-1068	aaacatgtccactg
EMBL AF070998/1-1305	tttccgggtgccgn
EMBL U95378/1-1267	gtaaacactgataga
EMBL U13680/1-1254	cggcaaccgctcgac
EMBL U07178/1-1668	aggcaacgacgtgcc
EMBL X02152/1-1661	tgctgcagccgctg
EMBL M22585/1-1615	ggaccgagcgcacc
EMBL U13687/1-1681	ggagcaacttggcg
EMBL X01964/1-1609	gtgtgctggagcca
EMBL X53828/1-1575	gaattccgccccgc

EMBL|U07178/1-1668
Sus domesticus lactate dehydrogenase-A (LDH-A) mRNA, complete cds.
EMBL U07178
EMBL-ALIGN ALIGN_000494
GOA P00339
InterPro IPR001236
InterPro IPR001557
InterPro IPR011304
PDB 9LDB
PDB 9LDT
UNIPROT P00339

Sequence 8 ID: EMBL|M22585 Nucleotide: Guanine (1)

1. Is this reference a cross reference ?
2. Is there already a sequence associated with this reference ?
If not: Retrieve it.
3. Copy associated sequence to new alignment.

Supplementary exercises

- cDNA and Protein splitframe

Day 1

9.30-9.40am. Overview of the Course

Session 1: Introduction to Jalview

- starting the application, importing and exporting alignments and sequence data

Coffee: 10.30am - 11am

Session 1: Continues

- Selection, colouring, basic editing and creating figures.

12.30pm to 1.30pm. Lunch

1.45pm to 2.30pm Geoff Barton – Multiple alignment and analysis

Session 2: Creating, editing and analysing alignments.

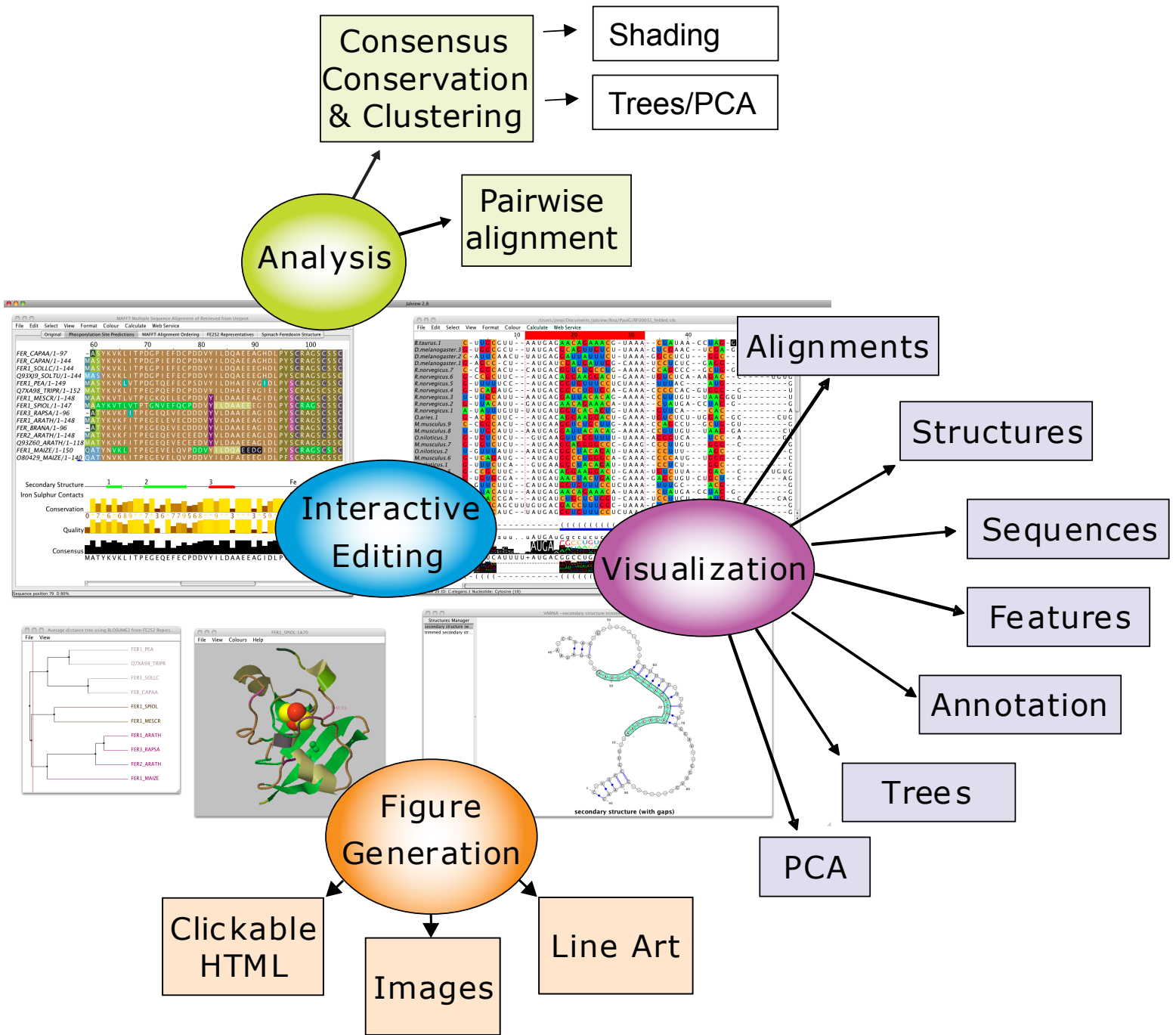
- Jalview alignment methods
- Creating sequence alignments

Coffee: 3.30pm – 4pm

Session 3: Alignment and analysis

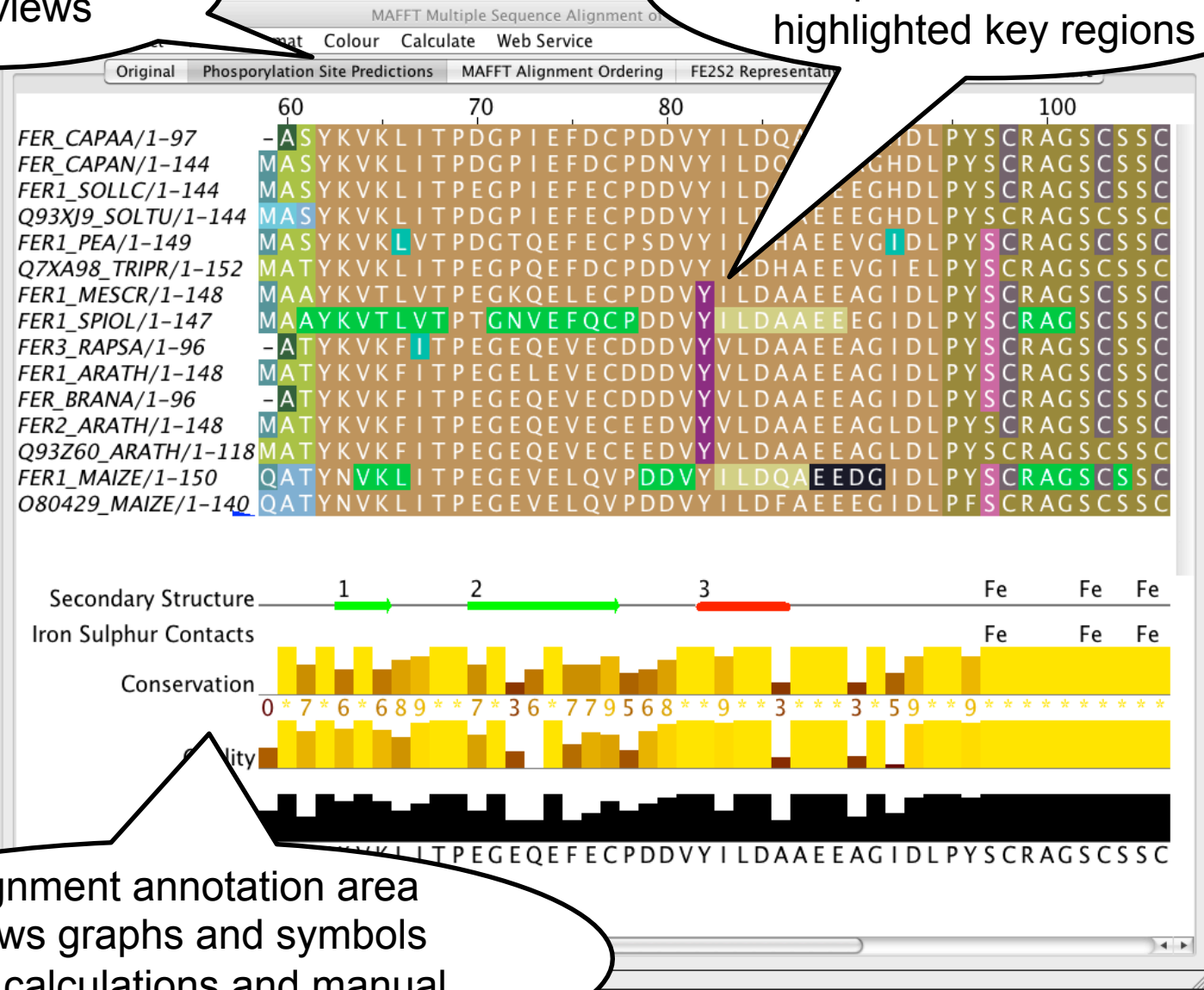
- importing and calculating trees and PCA
- Tree based alignment analysis
- working with group associated alignment annotation.

5.15pm – 5.30pm: Wrap up and plan for tomorrow

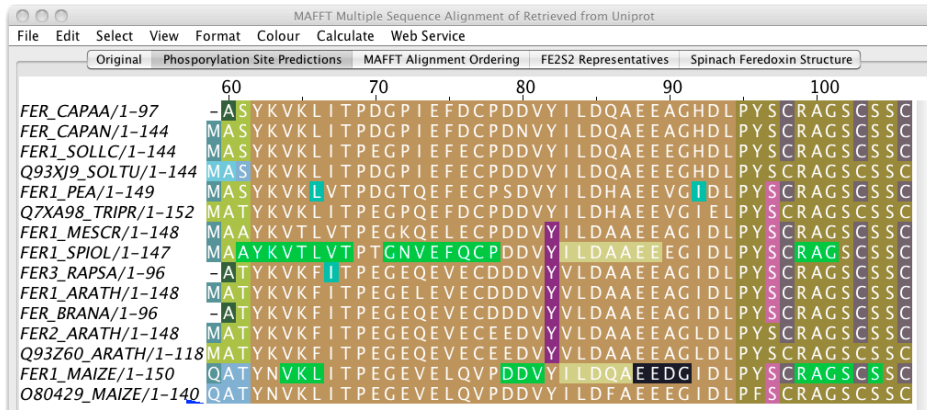


One alignment,
many views

Sequence features on
sequences shown to
highlighted key regions

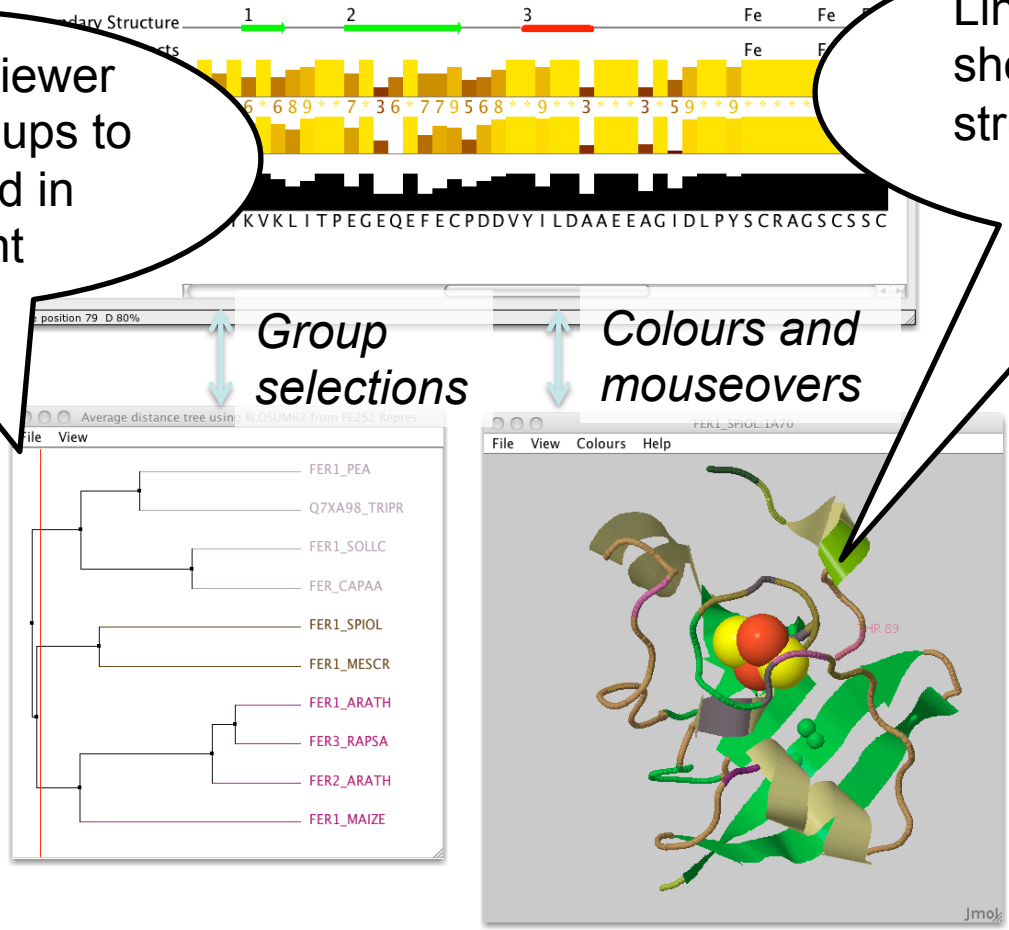


Alignment annotation area
shows graphs and symbols
from calculations and manual
curation



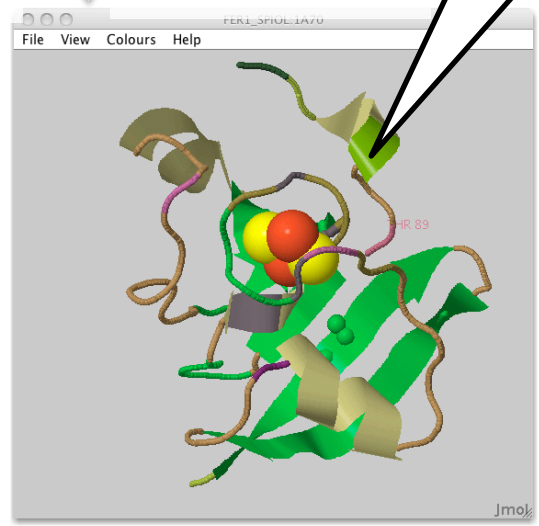
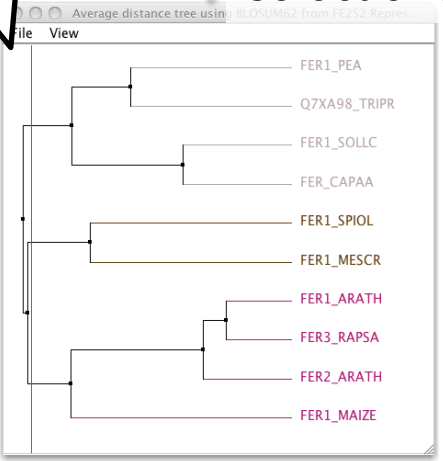
Linked tree viewer allows subgroups to be identified in alignment

Linked Jmol viewer shows one or more structures coloured by alignment



Group selections

Colours and mouseovers



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

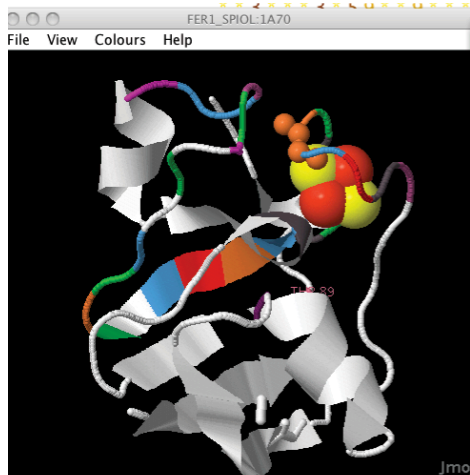
Desktop Structure Visualization

3D structures and 2D RNA diagrams



Jmol

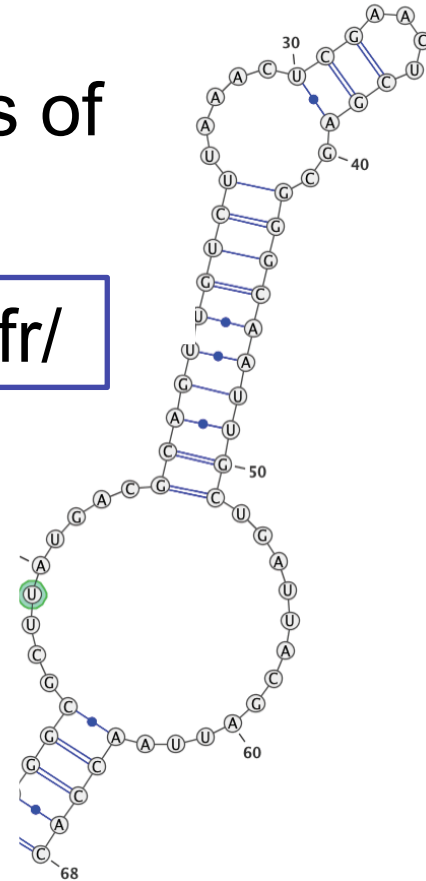
<http://jmol.sourceforge.net/>



VARNA

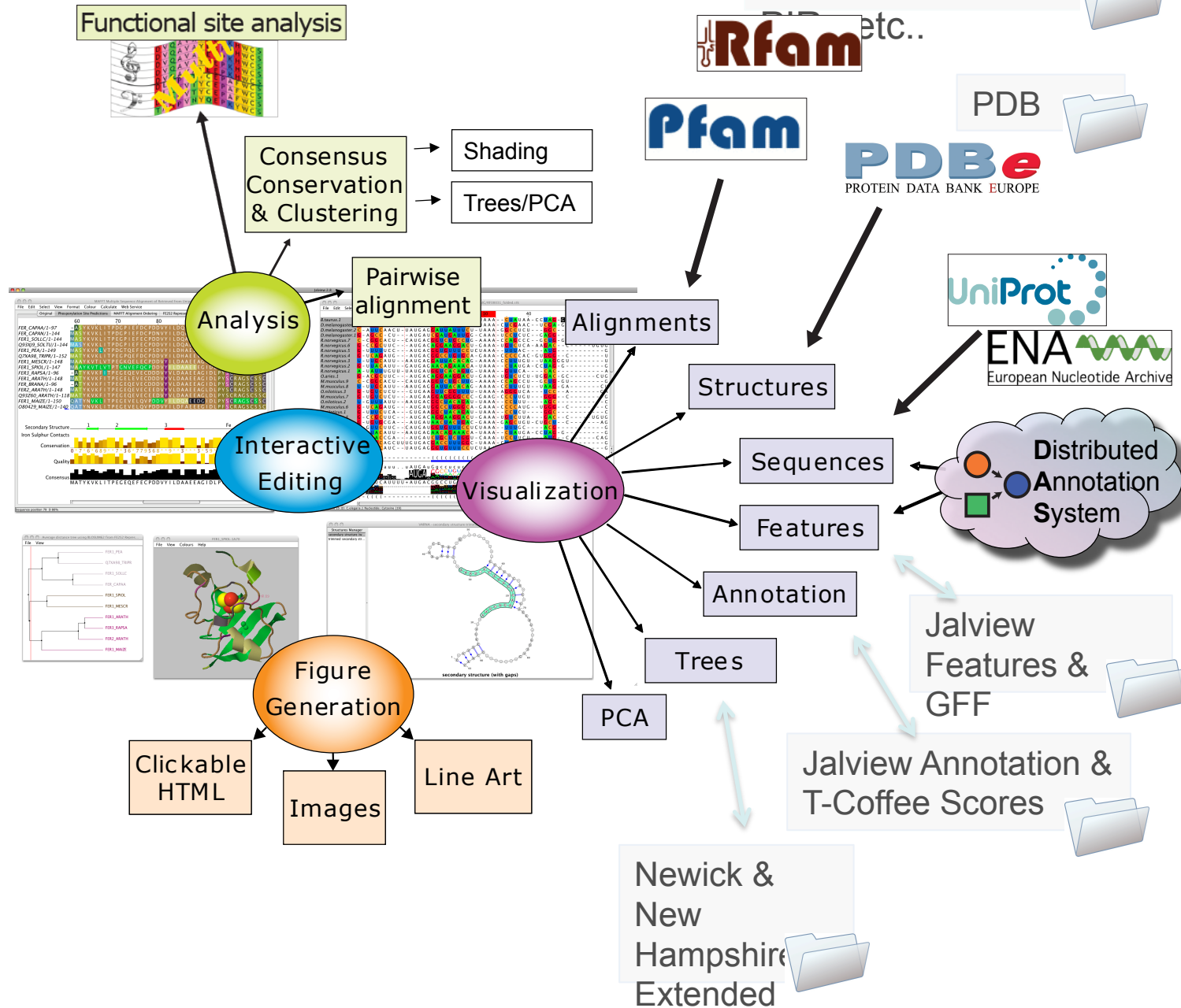
Visual Analysis of
RNA

<http://varna.lri.fr/>



The Jalview Desktop

Clustal Fasta Pile-Up
BLC Stockholm
etc..



NEW FOR 2014

Secondary structure

Jpred 3
Incorporating Jnet

ViennaRNA

Protein Disorder

UPred

RONN
Regional Order Neural Network

DisEMBL™

GLOBPLOT 2

Alignment Analysis

AACon

Multiple alignment

Ω T COFFEE
CLUSTAL

GLProbs MSAProbs

PROBCONS

Mafft MUSCLE W2
CLUSTAL

JABAWS 2.1

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

Analysis

alignment

Alignments

Structures

Interactive

Sequences

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/>

DB

EUROPE

UniProt

ENA
European Nucleotide Archive

Distributed Annotation System

new
ures &

otation &

Protein Scores

Newick &
New
Hampshire
Extended

File Edit Se



hour Calculate

Web Service

MA

Original

Spinach Ferredoxin Structure

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



Secondary Structure



Iron Sulphur Contacts



Alignment

Secondary Structure Prediction

Protein Disorder

Analysis

Conservation

Envision 2

Fetch DB References