Part III - Bioinformatics Study of Aminoacyl tRNA Synthetases

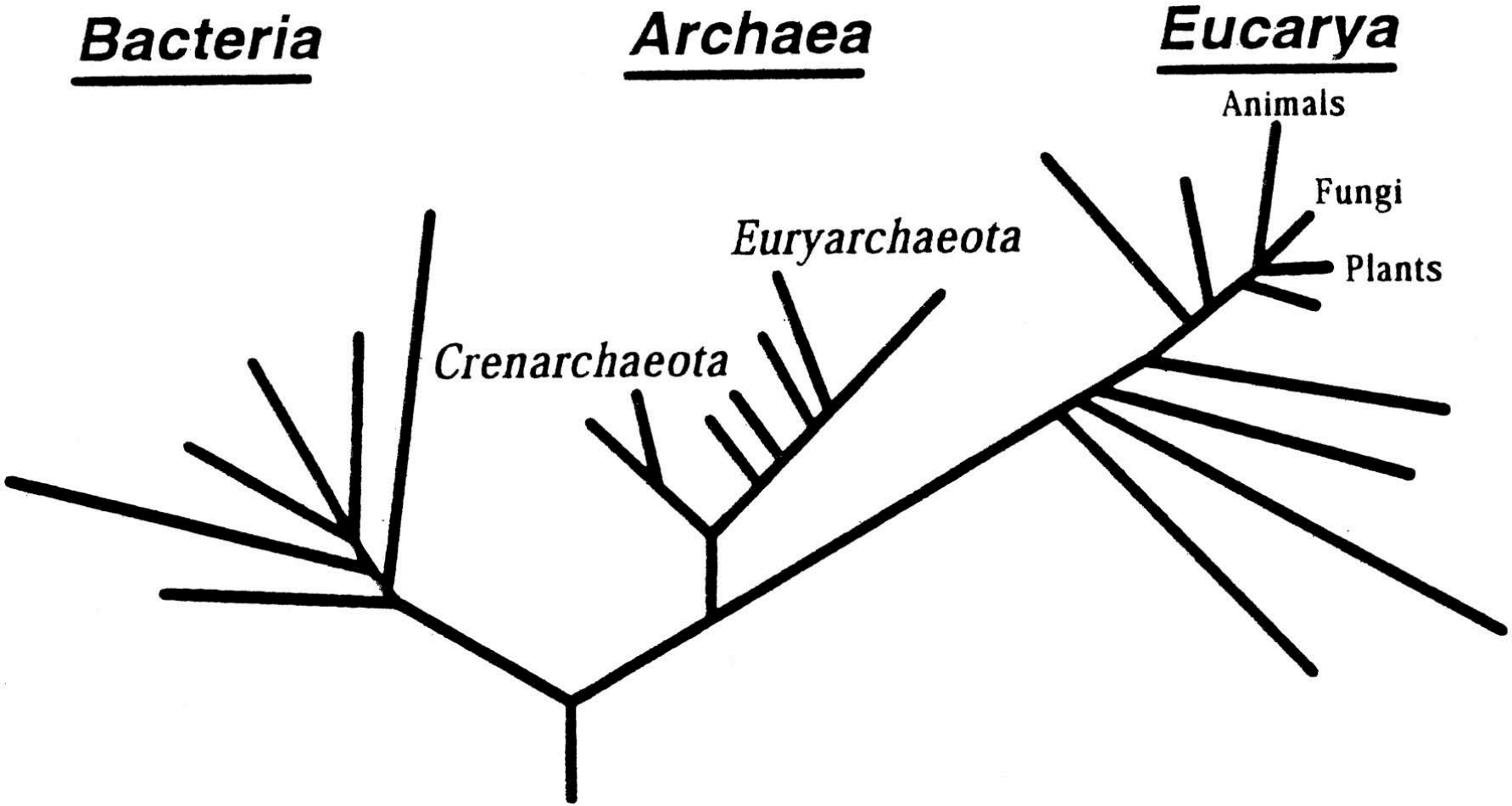
- VMD Multiseq Tutorial
- Web tools

Perth, Australia 2004 Computational Biology Workshop

Multiple Sequence Alignments

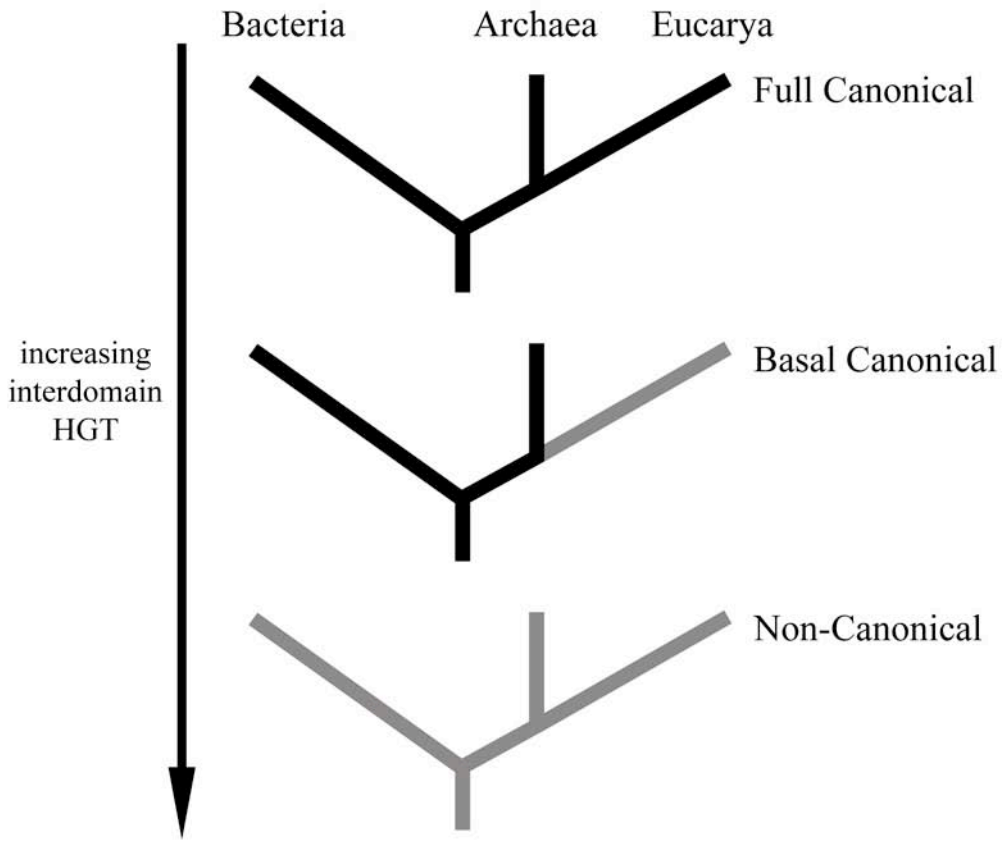
- “The aminoacyl-tRNA synthetases, perhaps better than any other molecules in the cell, eptiomize the current situation and help to under standard (the effects) of HGT” Woese (PNAS, 2000; MMBR 2000)

Universal Tree



The Universal Phylogenetic Tree inferred from comparative analyses of rRNA sequences: Woese(PNAS, 1990)

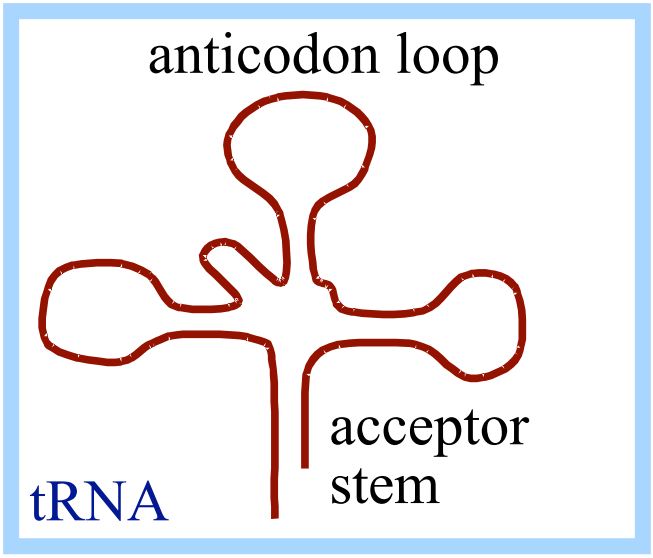
Horizontal Gene Transfer



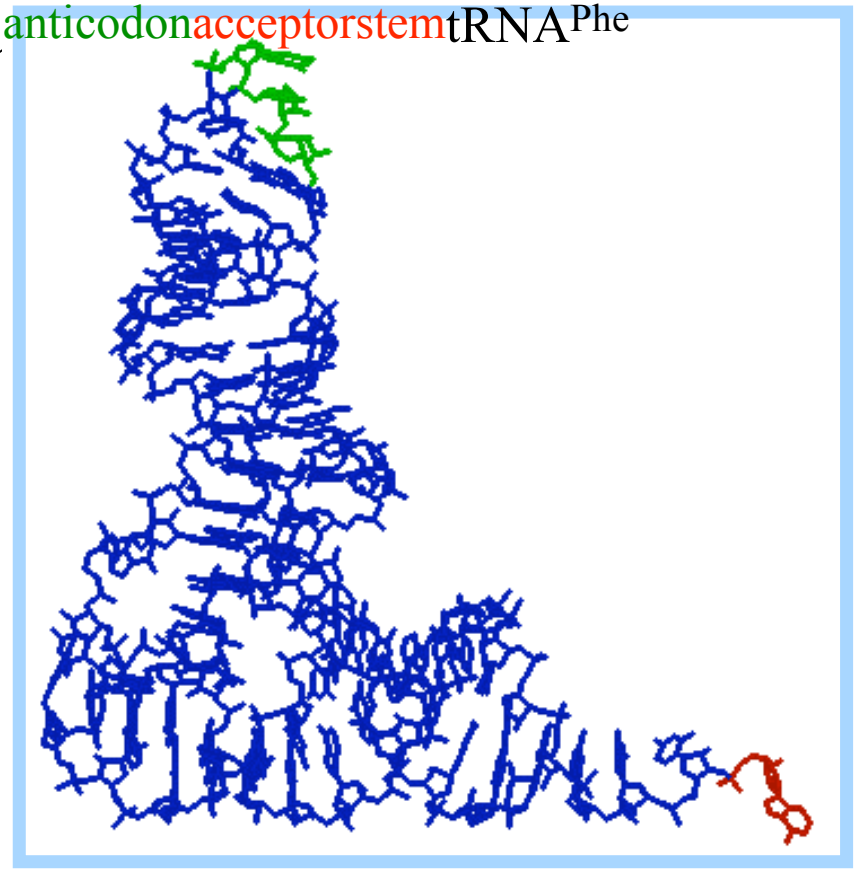
O'Donoghue and Luthey-Schulten, MMBR, 2004

Standard Dogma Molecular Biology

- DNA → RNA → Proteins
- Role of AARS?
- Charging of t-RNA

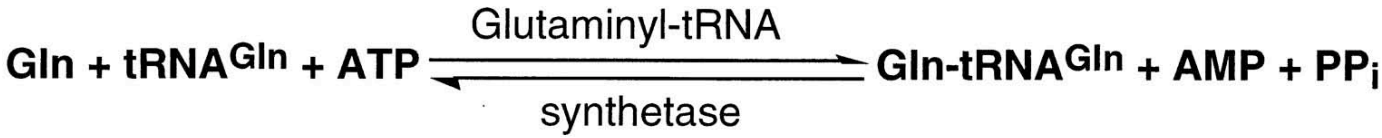


www.rpi.edu/dept/bcbp/molbiochem/MBWeb/mb2/part1/25-trna.ppt

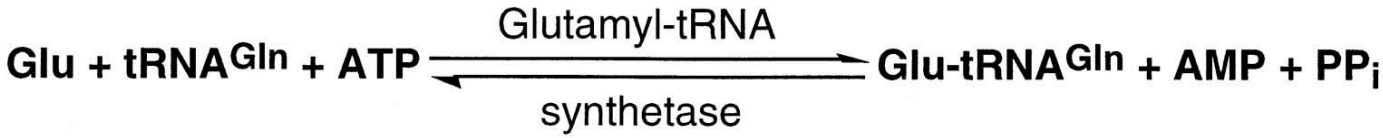


Charging the tRNA

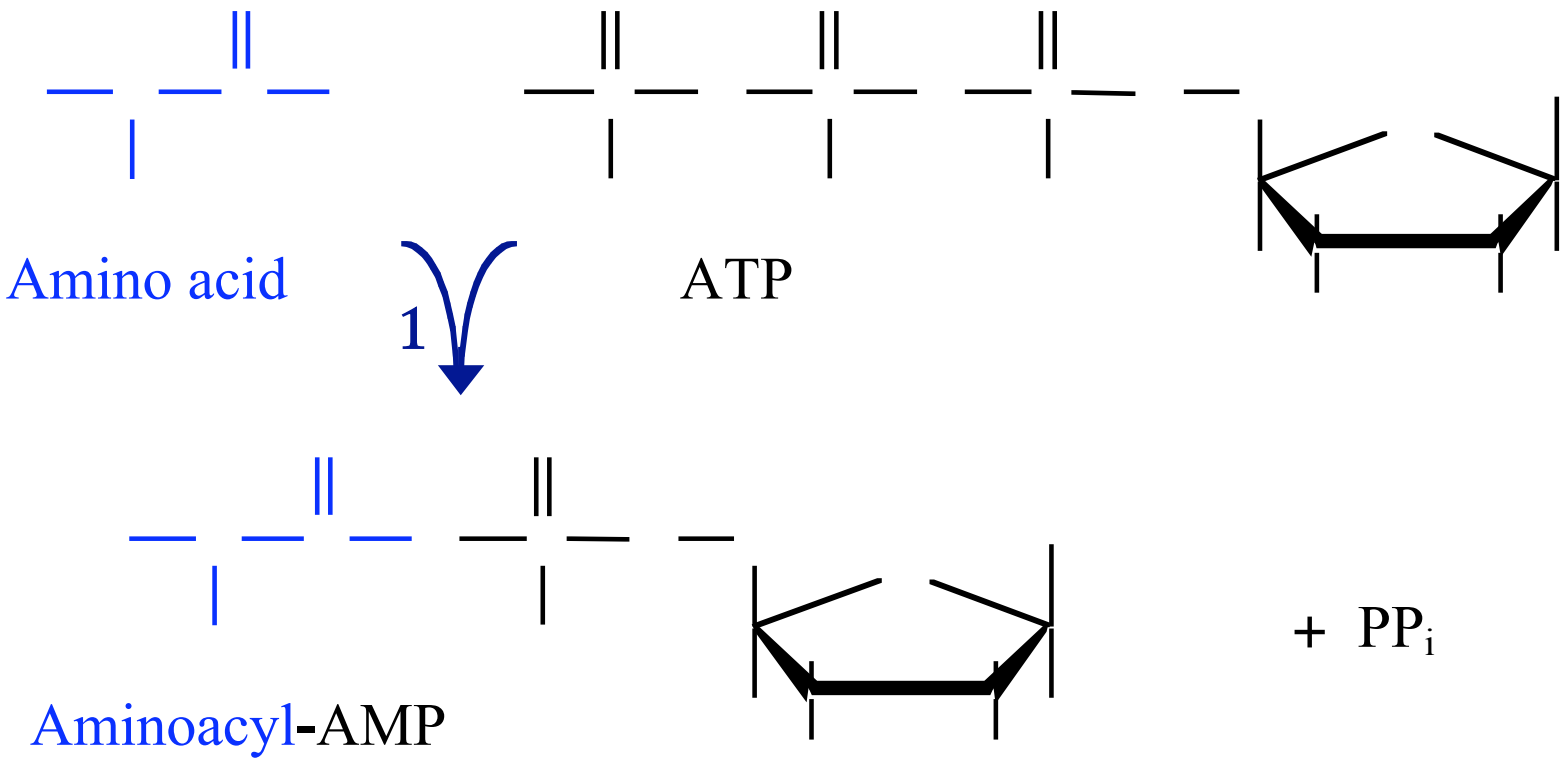
Direct acylation



tRNA-dependent amino acid modification



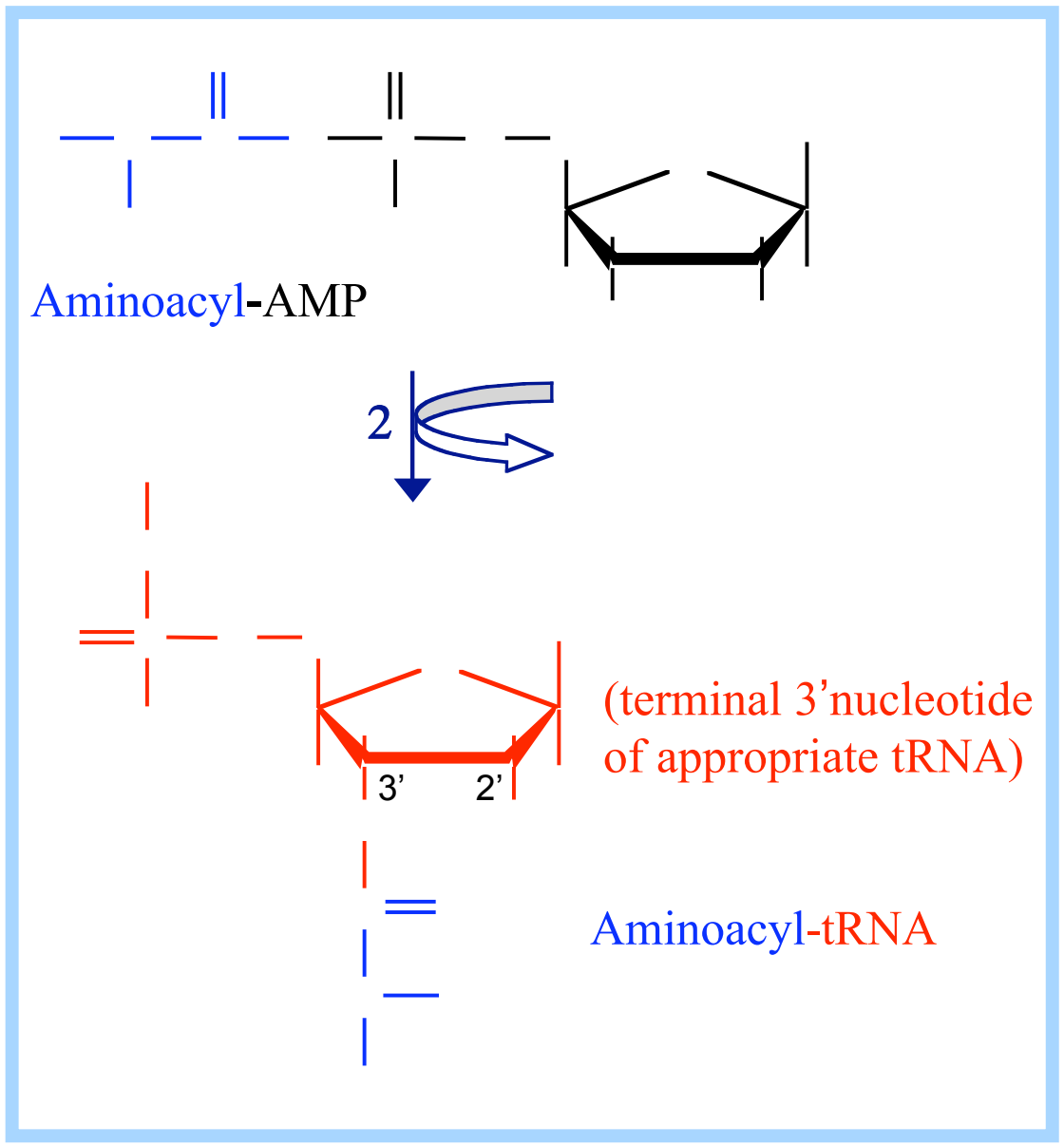
Woese, Olsen (UIUC), Ibba (Panum Inst.), Soll (Yale) *Micro. Mol. Biol. Rev.* March 2000..



Aminoacyl-tRNA Synthetases catalyze linkage of the appropriate amino acid to each tRNA. The reaction occurs in two steps.

In **step 1**, an O atom of the amino acid α-carboxyl attacks the P atom of the initial phosphate of ATP.

In **step 2**, the 2' or 3' OH of the terminal adenosine of tRNA attacks the amino acid carbonyl C atom.



Aminoacyl-tRNA Synthetase

Summary of the 2-step reaction:

1. **amino acid + ATP → aminoacyl-AMP + PP_i**

2. **aminoacyl-AMP + tRNA → aminoacyl-tRNA + AMP**

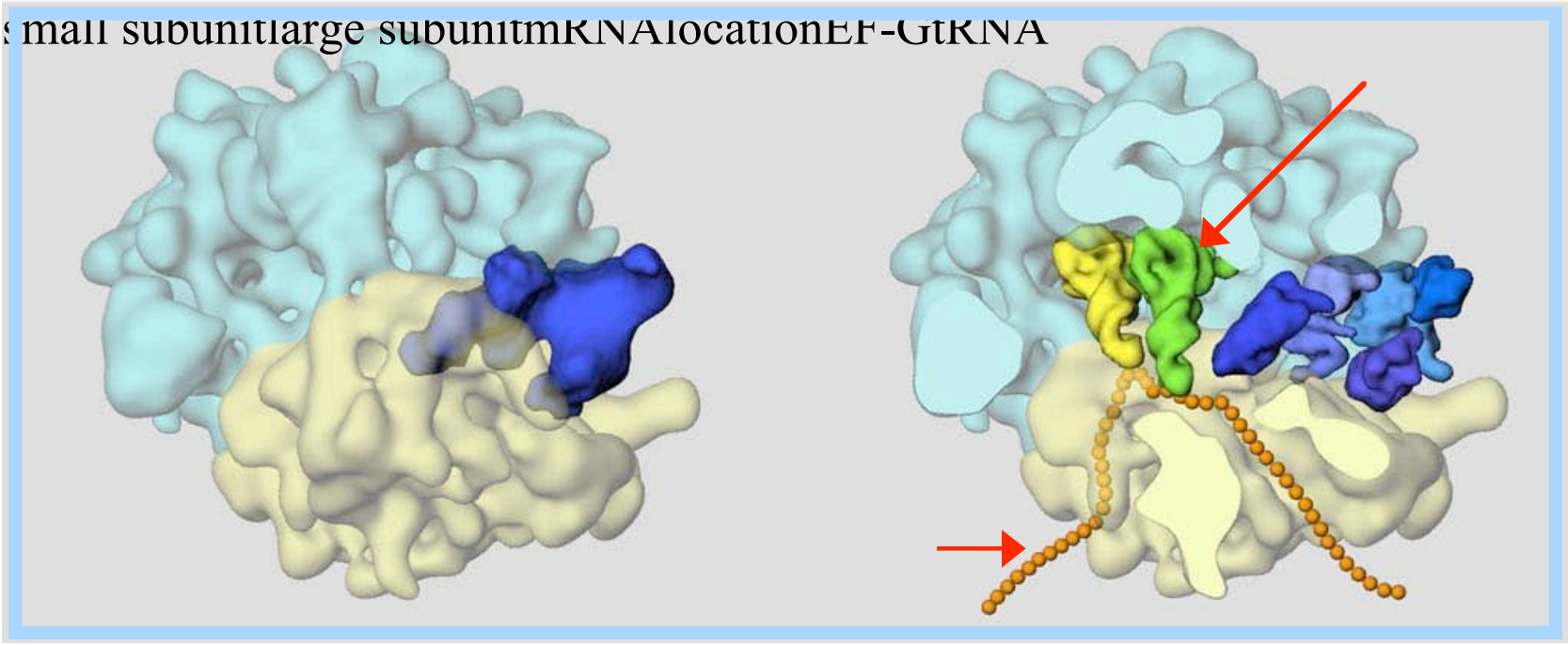
Overall Reaction:

amino acid + ATP + tRNA → aminoacyl-tRNA + AMP

+ PP_i

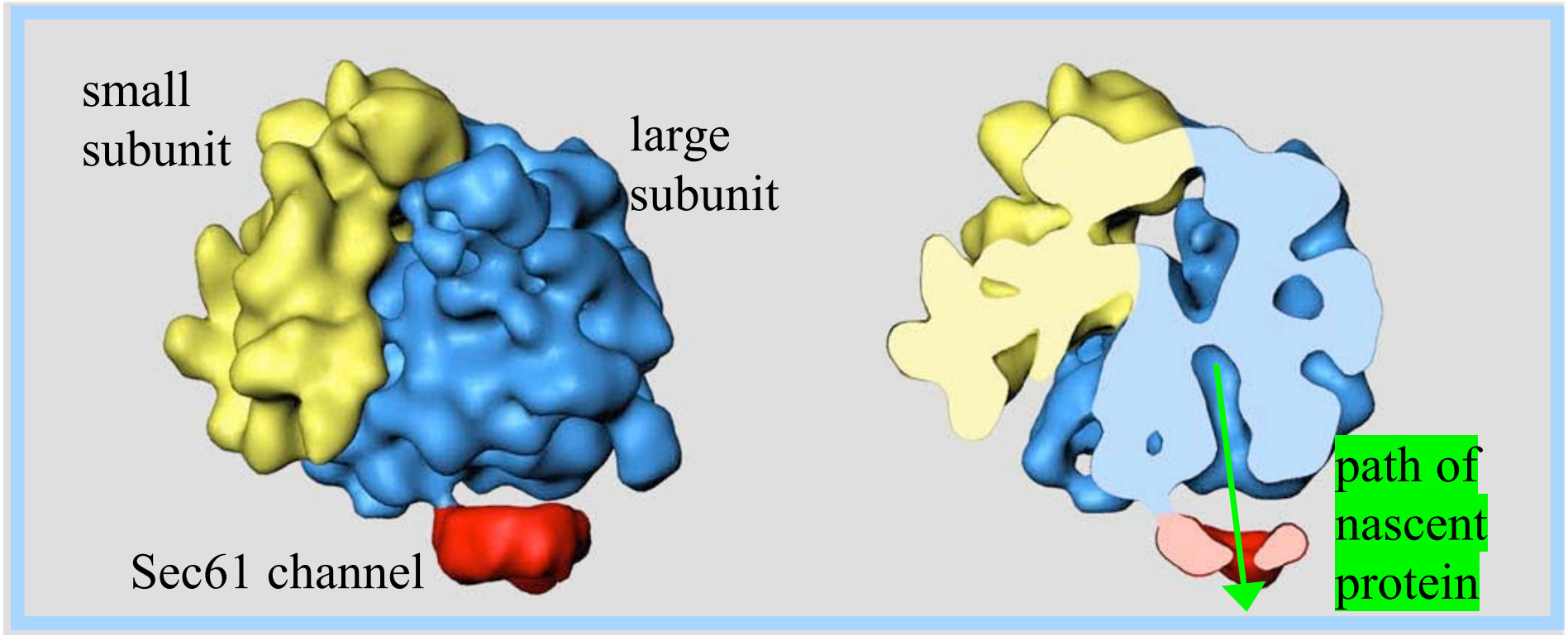
Next step: EF and Ribosome for Protein Synthesis

Structure of the *E. coli* Ribosome



The cutaway view at right shows positions of tRNA (P, E sites) & mRNA (as orange beads).

Figure: Laboratory of Joachim Frank, Wadsworth Center cryo-EM and 3D image reconstruction



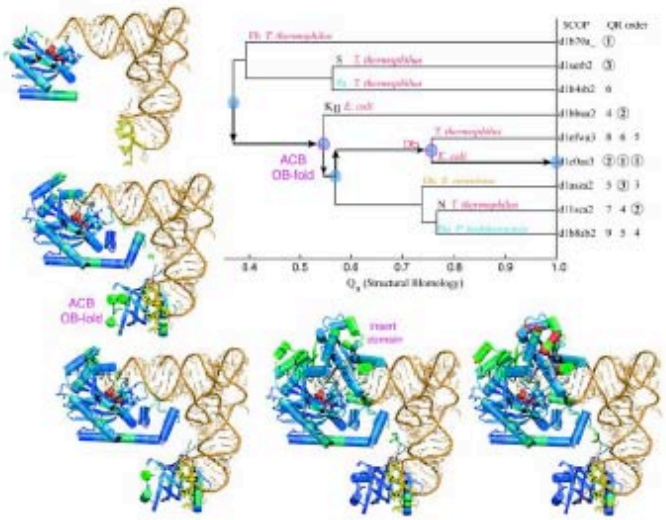
The cutaway view at right shows that the **tunnel** in the yeast large ribosome subunit, through which nascent polypeptides emerge from the ribosome, **lines up** with the lumen of the ER **Sec61 channel**.

Figure provided by Joachim Frank, whose lab carried out the cryo-EM & image reconstruction on which these images are based.

University of Illinois at Urbana-Champaign
NIH Resource for Macromolecular Modeling and Bioinformatics
School of Chemical Sciences and Beckman Institute

Evolution of Protein Structure

Aspartyl-tRNA Synthetase



VMD Developers:

Dan Wright

John Eargle

John Stone

Dr. Zan Luthey-Schulten

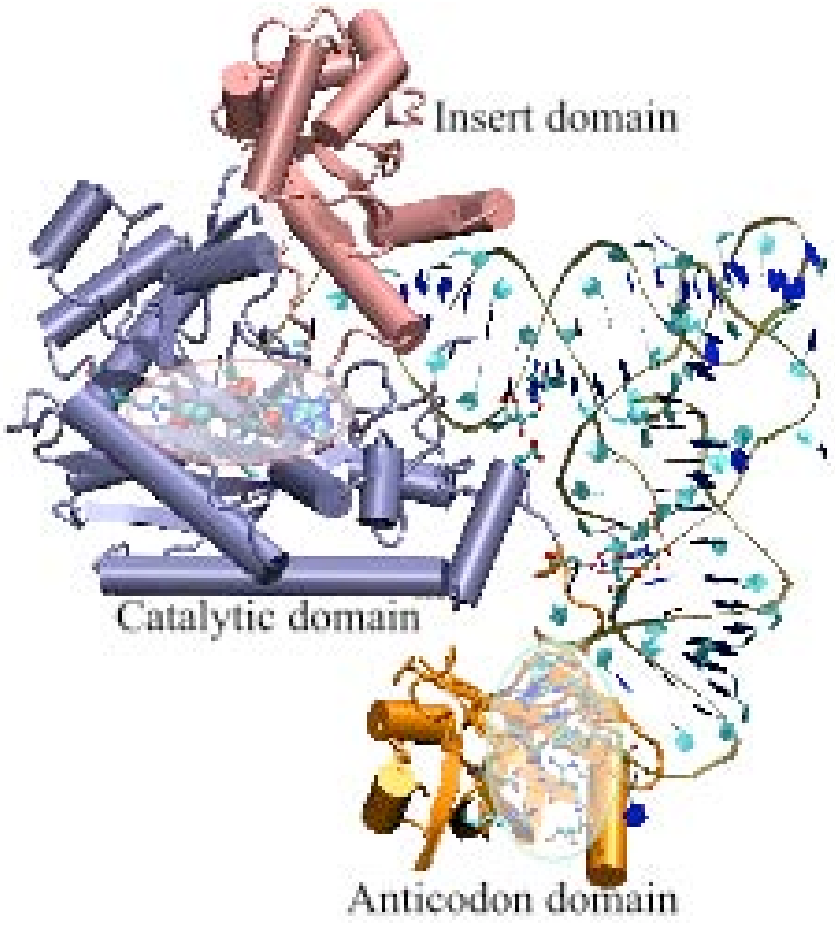
Brijeet Dhaliwal

Patrick O'Donoghue

Rommie Amaro

April 2004.

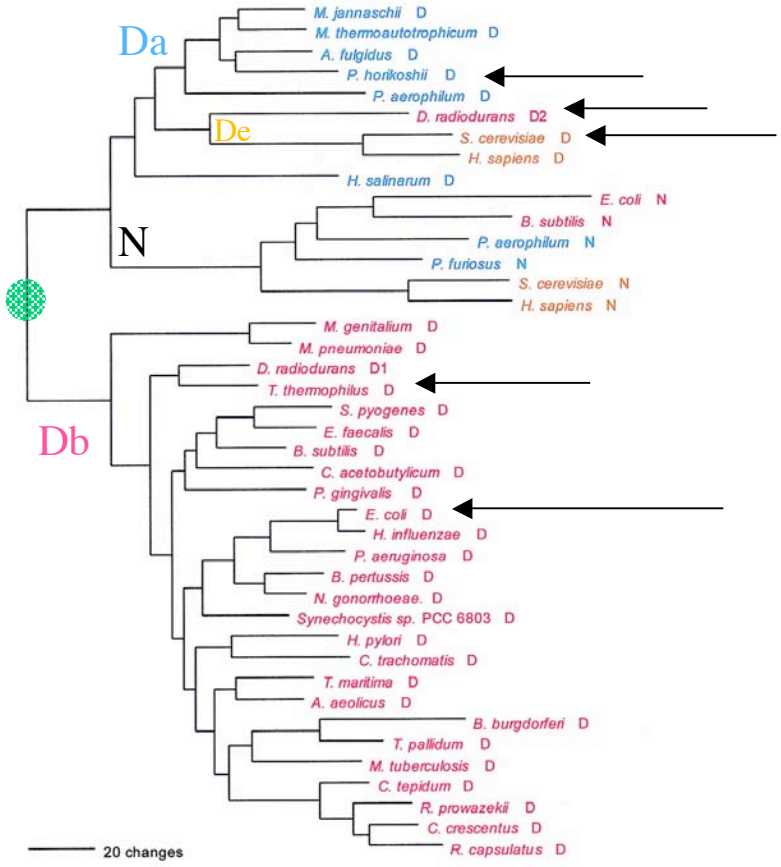
Step 1: Explore active site in catalytic domain and anticodon domain.

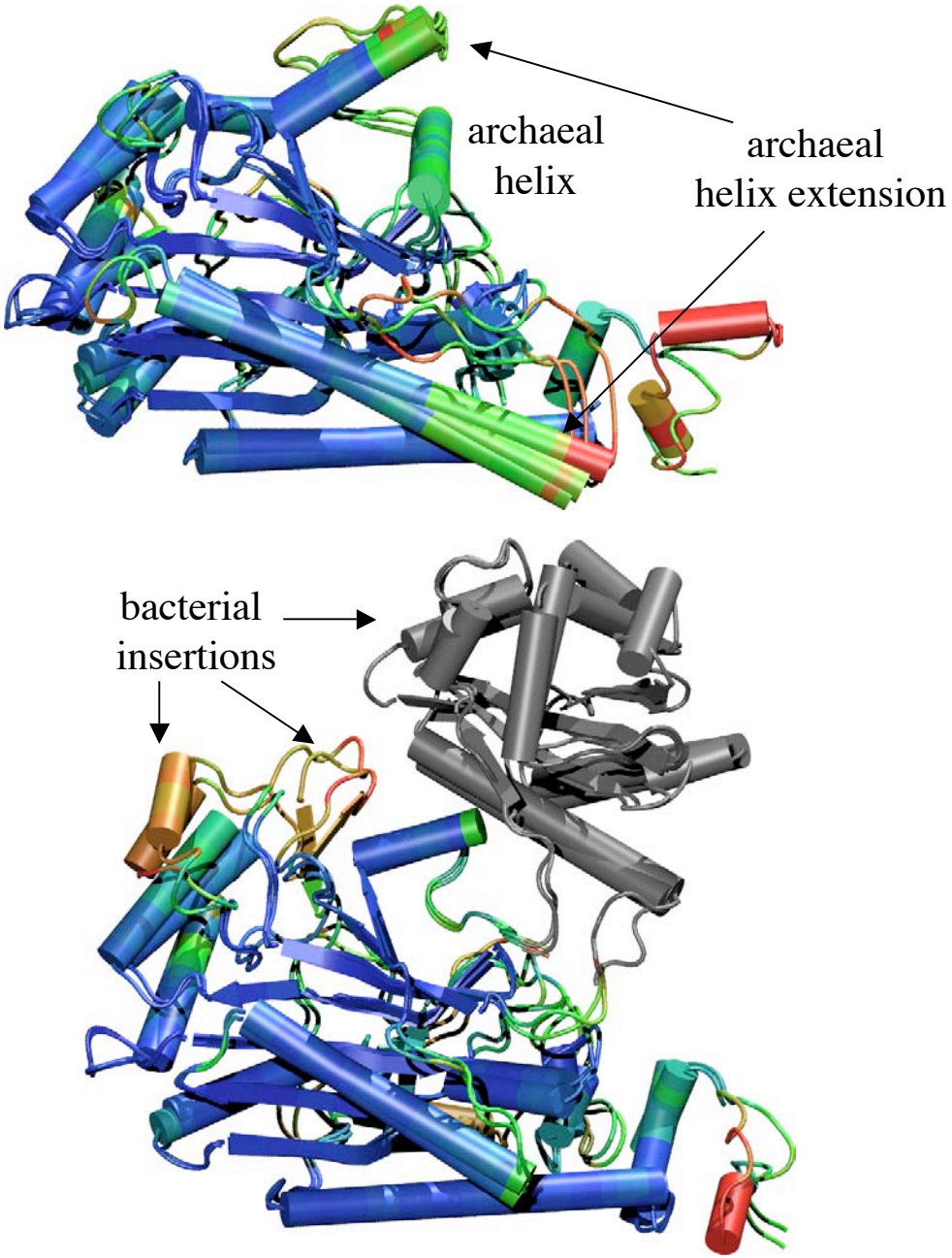


Horizontal Gene Transfer in Protein Structure

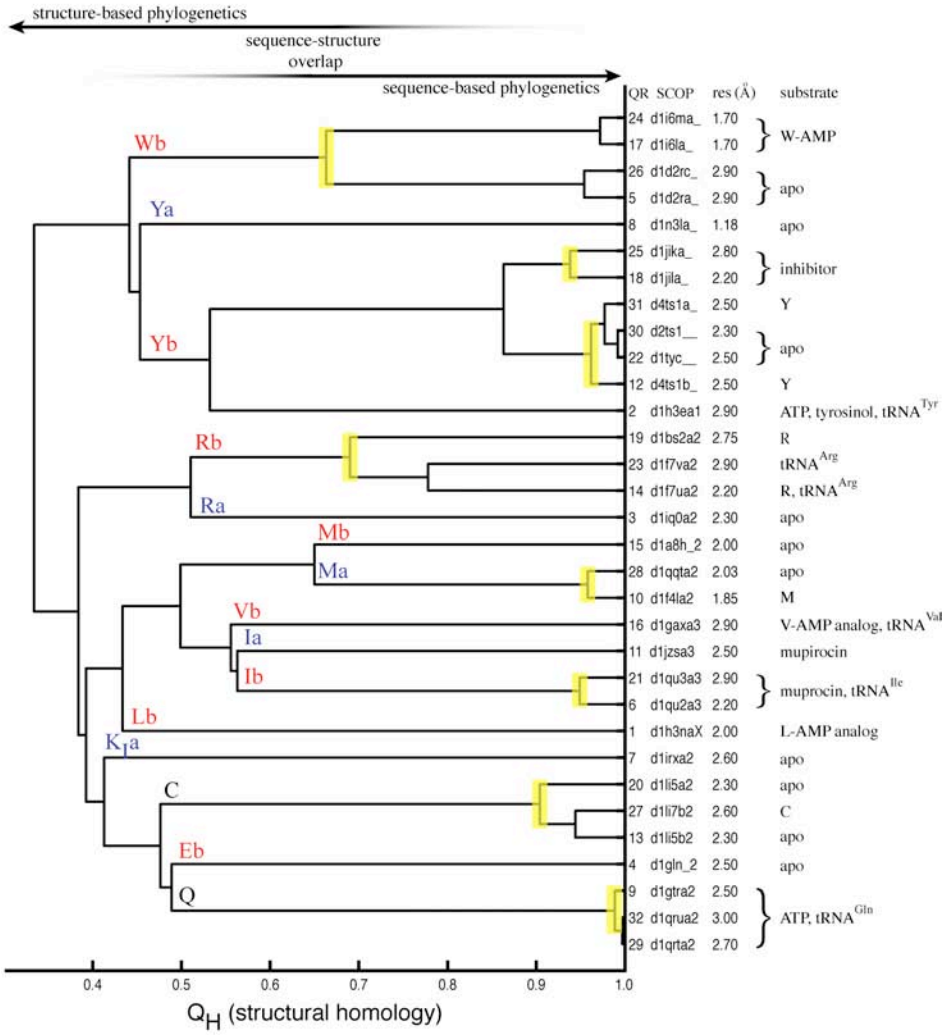
Sequence Phylogeny

AspRS-AsnRS Group

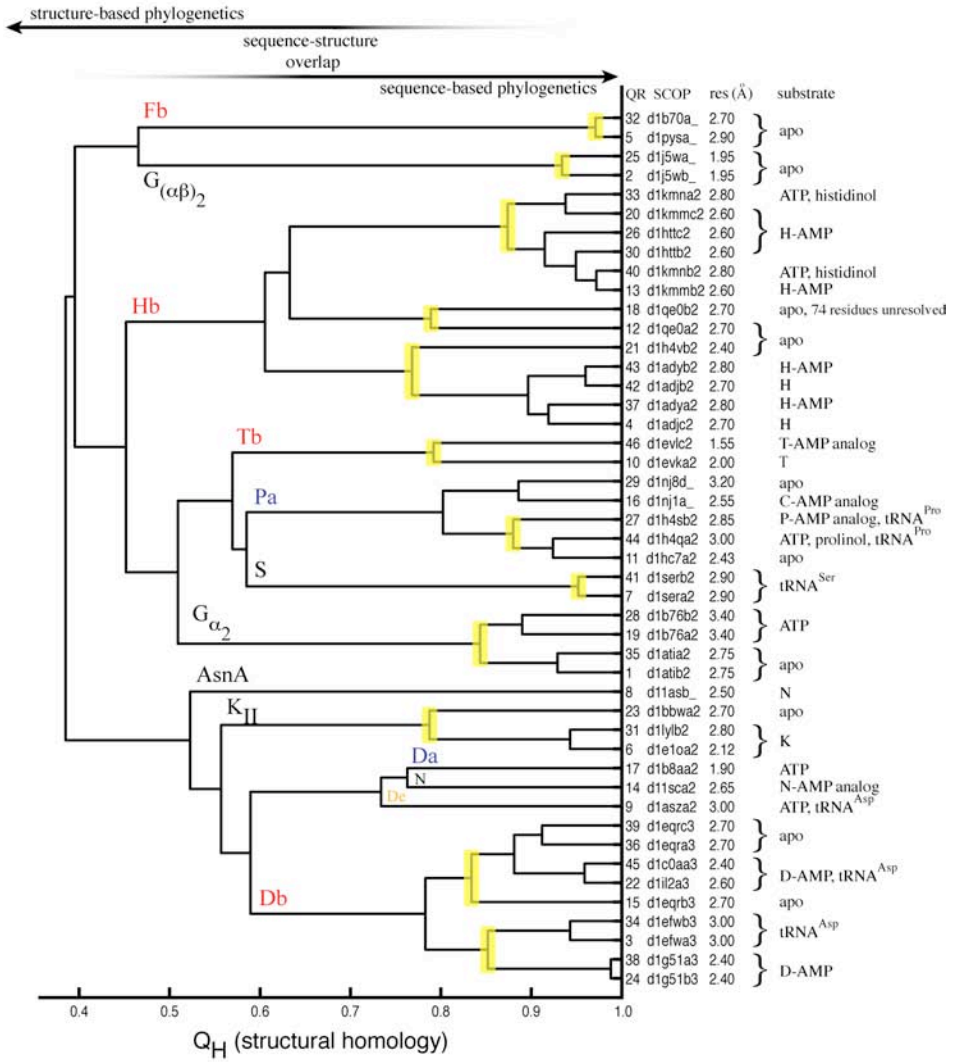




Structure Phylogeny
Class I AARSs

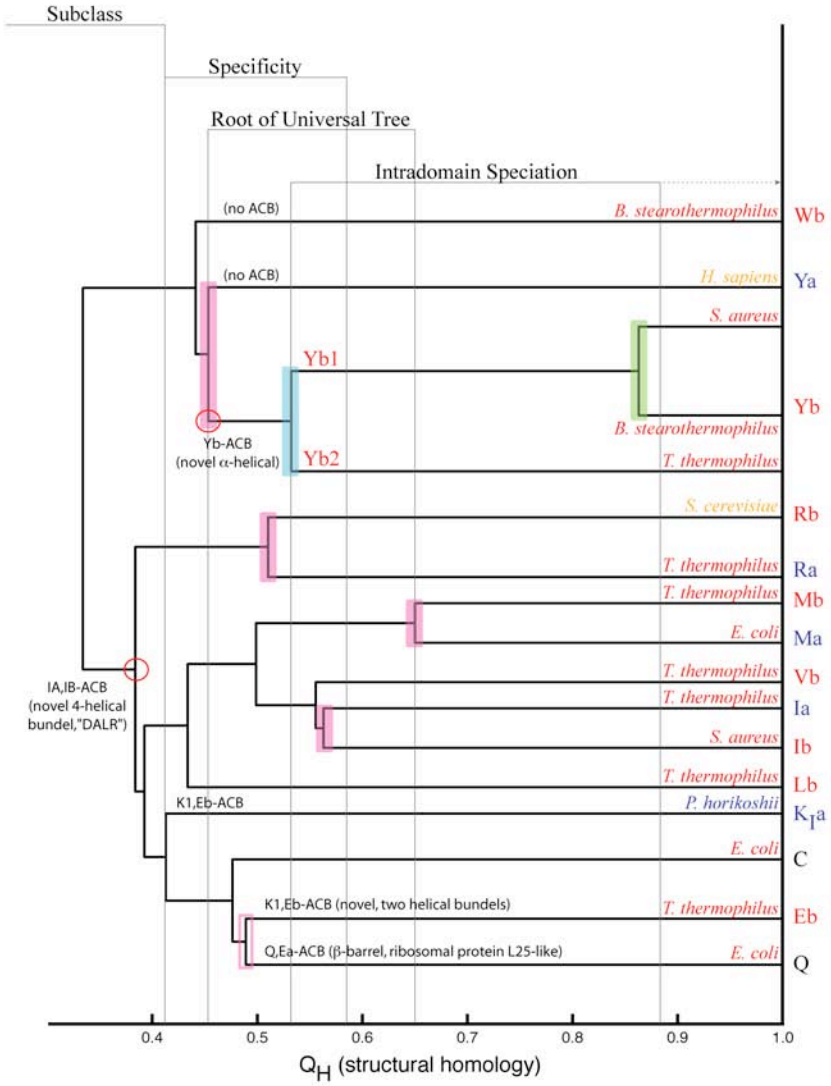


Structure Phylogeny
Class II AARSs

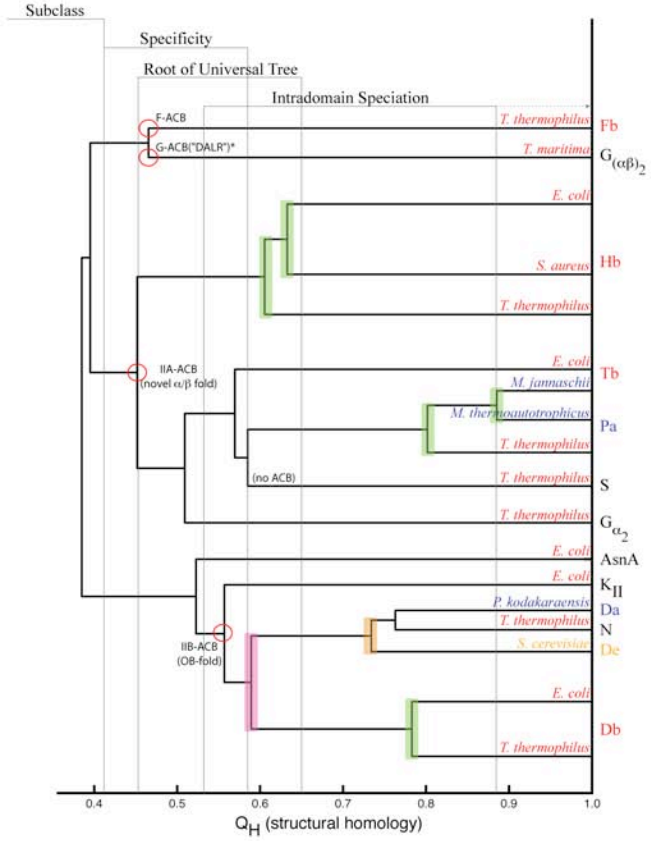


O'Donoghue and Luthey-Schulten, MMBR 2004

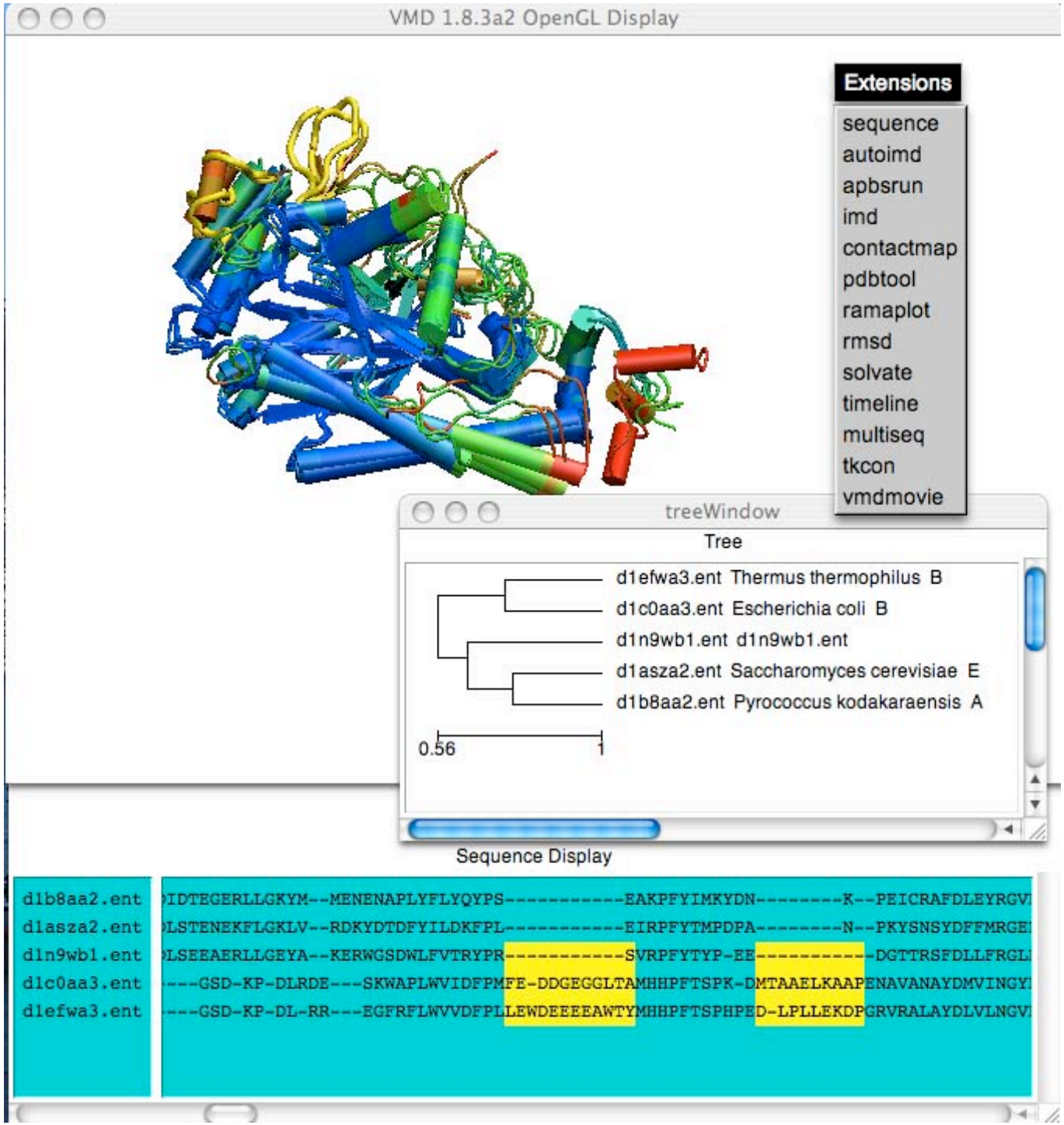
Structure Phylogeny
Class I AARSs



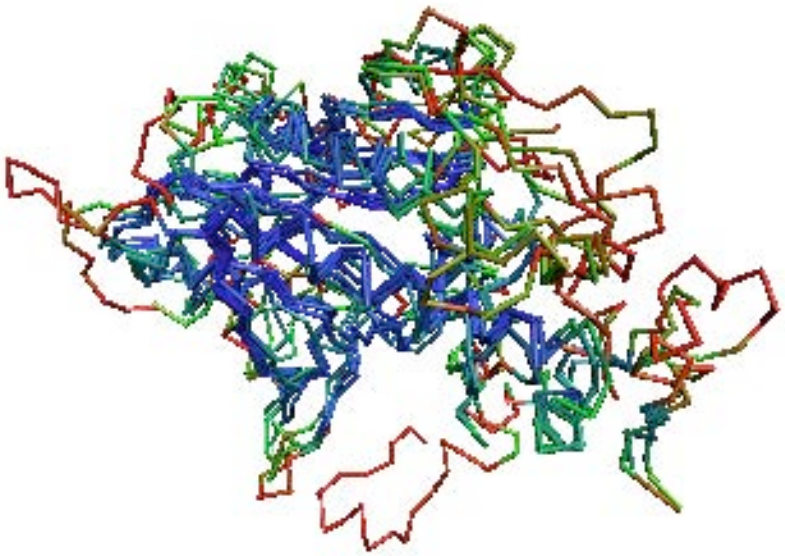
Structure Phylogeny
Class II AARSs



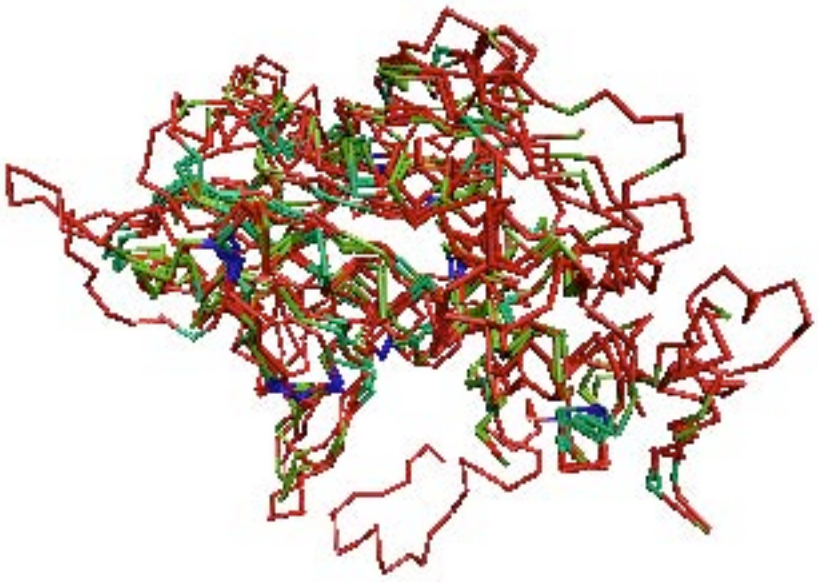
Multiseq
extension in VMD



Conservation



Core Structure Conserved



Sequence Identity of Core
Less than 15%

Useful Web Tools

- SCOP - Structure Database
- NCBI Genomes – Sequence and Gene Information
- SWISSPROT - Sequence Database
- PFAM Domain Architecture
- Clustal Multiple Sequence Alignments
- Hidden Markov Methods
- Phylip Phylogenetic Trees
- Matlab - Statistics UPGMA

Structural Classification of Proteins

Scop (Astral Database)

Protein: Aspartyl-tRNA synthetase (AspRS) from *Escherichia coli*

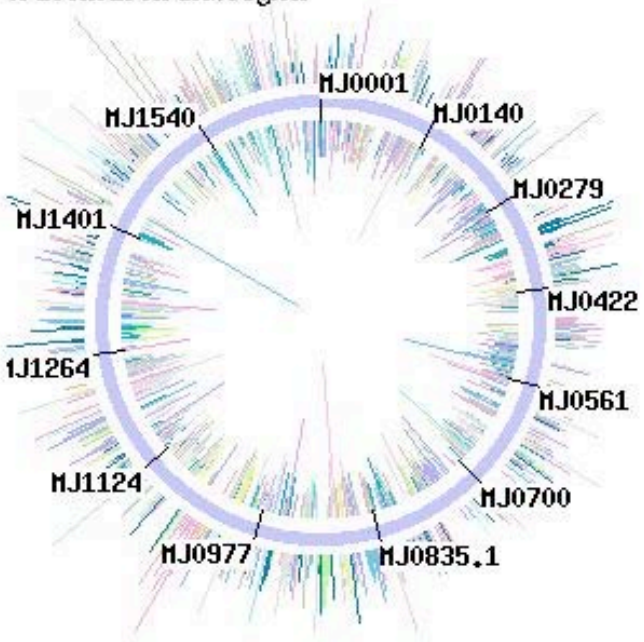
Lineage:

1. Root: scop
2. Class: All beta proteins
3. Fold: OB-fold
   barrel, closed or partly opened n=5, S=10 or S=8; greek-key
4. Superfamily: Nucleic acid-binding proteins
5. Family: Anticodon-binding domain
   barrel, closed; n=5, S=10
6. Protein: Aspartyl-tRNA synthetase (AspRS)
   this is N-terminal domain in prokaryotic enzymes and the first "visible" domain in eukaryotic enzymes
7. Species: *Escherichia coli*

PDB Entry Domains:

1. 1c0a
   1. region a:1-106
2. 1il2
   complexed with 1mg, 5mc, 5mu, amo, h2u, psu, so4
   1. region a:1-106
   2. region b:1001-1106
3. 1eqr
   complexed with mg
   1. region a:1-106
   2. region b:1-106
   3. region c:1-106

Protein coding genes distribution map
To see map locations of genes, click on a region in the map, to zoom in on that region

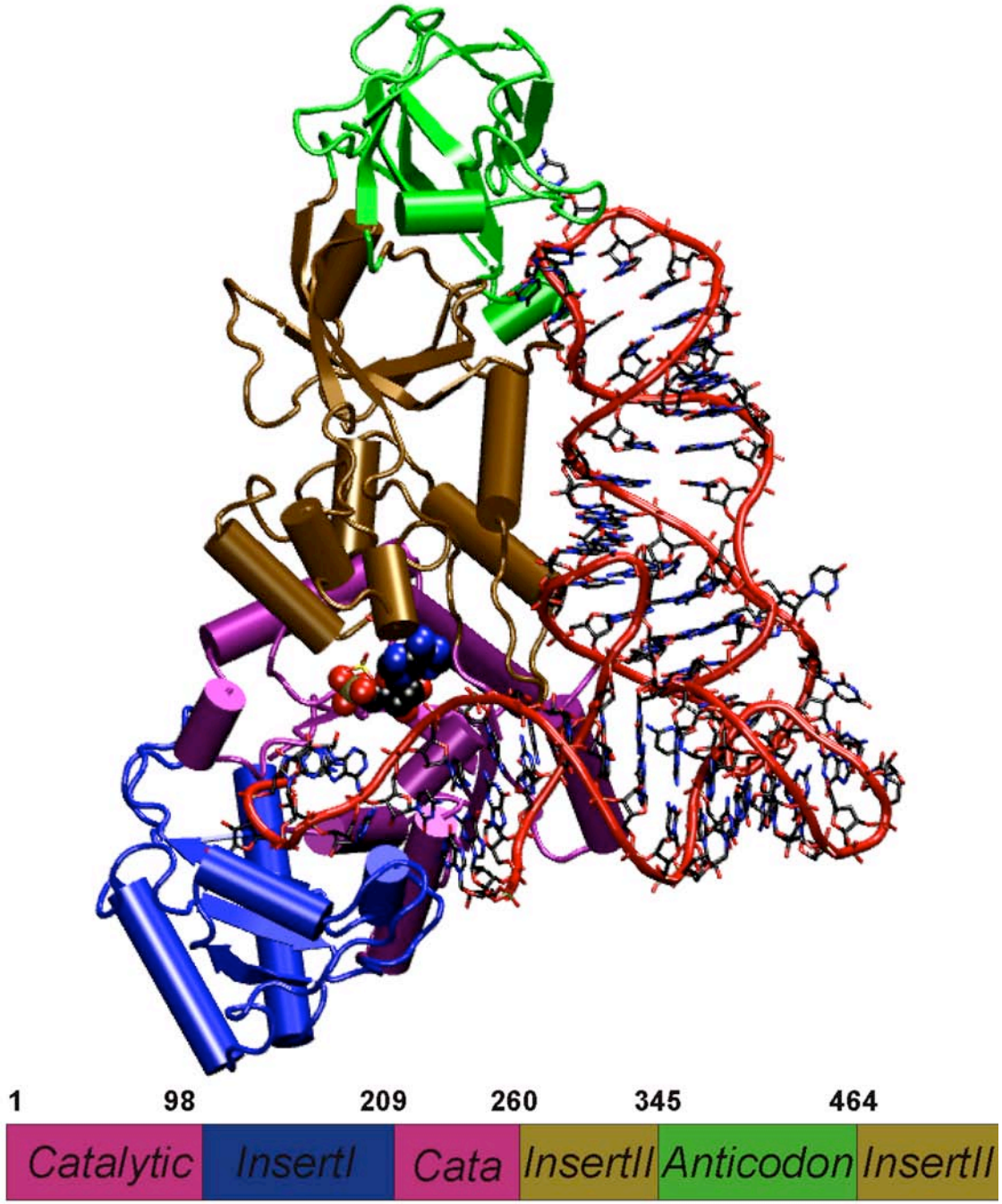


Gene Classification based on COG functional categories

- Translation, ribosomal structure and biogenesis
- Transcription
- DNA replication, recombination and repair
- Cell division and chromosome partitioning
- Posttranslational modification, protein turnover
- Cell envelope biogenesis, outer membrane
- Cell motility and secretion
- Inorganic ion transport and metabolism
- Signal transduction mechanisms
- Energy production and conversion
- Carbohydrate transport and metabolism
- Amino acid transport and metabolism
- Nucleotide transport and metabolism
- Coenzyme metabolism
- Lipid metabolism
- Secondary metabolites biosynthesis, transport ar
- General function prediction only
- Function unknown
- No COG match

Organism: Methanocaldococcus jannaschii
Genetic Code: 11
Lineage: Archaea; Euryarchaeota; Methanococci; Methanococcales; Methanocaldococcaceae; Methanocaldococcus.

Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F., Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F., Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.
Science 273 (5278), 1058-1073 (1996)
96337999


1
98
209
260
345
464
Catalytic
InsertI
Cata
InsertII
Anticodon
InsertII

NCBI 3D

NCBI

protein–protein **BLAST**

Nucleotide | Protein | Translations | Retrieve results for an RID

Search

Your Favorite Sequence in Fasta Format

Set subsequence From: To:

Choose database swissprot

Do CD-Search

Now: BLAST! or Reset query Reset all

Report from SWISS-PROT

Comments

- ***CATALYTIC ACTIVITY***: ATP + L-aspartate + tRNA(Asp) = AMP + diphosphate + L-aspartyl-tRN
- ***COFACTOR***: Binds 3 magnesium ions per subunit *(By similarity)*.
- ***SUBCELLULAR LOCATION***: Cytoplasmic.
- ***SIMILARITY***: Belongs to class-II aminoacyl-tRNA synthetase family.

Copyright

Cross-references

EMBL	AB010464; BAA31457.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence]
HSSP	Q52428; 1B8A. [HSSP ENTRY / PDB]
HAMAP	MF_00044; -; 1. PBIL [Family / Alignment / Tree]
InterPro	IPR004523; AspS_arch. IPR004364; tRNA-synt_2. IPR002312; tRNA-synt_asp. IPR004365; tRNA_anti. IPR006195; tRNA_ligase_II. Graphical view of domain structure.
Pfam	PF00152; tRNA-synt_2; 1. PF01336; tRNA_anti; 1.
PRINTS	PR01042; TRNASYNTHASP.
TIGRFAMs	TIGR00458; aspS_arch; 1.
PROSITE	PS50862; AA_TRNA_LIGASE_II; 1.
ProDom	[Domain structure / List of seq. sharing at least 1 domain]
HOBACGEN	[Family / Alignment / Tree]
BLOCKS	O24822.
ProtoNet	O24822.
ProtoMap	O24822.

PFAM Report

Representative tRNA-synt_2 family proteins

This family may contain **overlapping domains**, to change the graphical view click here

SYDC_YEAST [Saccharomyces cerevisiae (baker's yeast)] aspartyl-trna synthetase, cytoplasmic (ec 6.1.1.12) (aspartate--trnaligase) (asprs)

[556 residues]

SYD_CAEEL [Caenorhabditis elegans] aspartyl-trna synthetase (ec 6.1.1.12) (aspartate--trna ligase)(asprs)

[531 residues]

SYD_HUMAN [Homo sapiens (human)] aspartyl-trna synthetase (ec 6.1.1.12) (aspartate--trna ligase)(asprs)

[501 residues]

SYD_RAT [Rattus norvegicus (rat)] aspartyl-trna synthetase (ec 6.1.1.12) (aspartate--trna ligase)(asprs)

[501 residues]

CLUSTALW: Multiple Sequence Alignment[help]

General Setting Parameters:

Output Format: CLUSTAL

Pairwise Alignment: FAST/APPROXIMATE SLOW/ACCURATE

Enter your sequences (with labels) below (copy & paste): PROTEIN DNA

Support Formats: FASTA (Pearson), NBRF/PIR, EMBL/Swiss Prot, GDE, CLUSTAL, and GCG/MSF

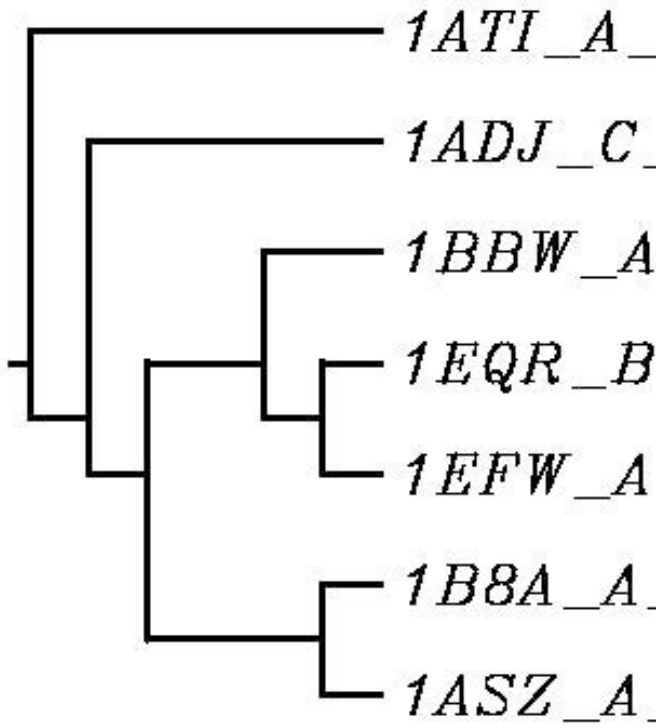
```
>1EQR:B(000:0-000)
VLPLDSNHVNTEEARLKYRYLDLRRPEMAQRLKTRAKITSLVRRFMDDHGFLDIETPMLT
KATPEGARDYLVPSRVHKGKFYALPQSPQLFKQLLMMSGFDRYYQIVKCFRDEDLRADRQ
PEFTQIDVETSFMTAPQVREVMEALVRHLWLEVKGVDLGDFPVMTFAEAERRYGSDKPDL
RDESKWAPLWVIDFPMFEDDGEGGLTAMHHPFTSPKDMTAAELKAAPENAVANAYDMVIN
```

Or give the file name containing your query

Browse...

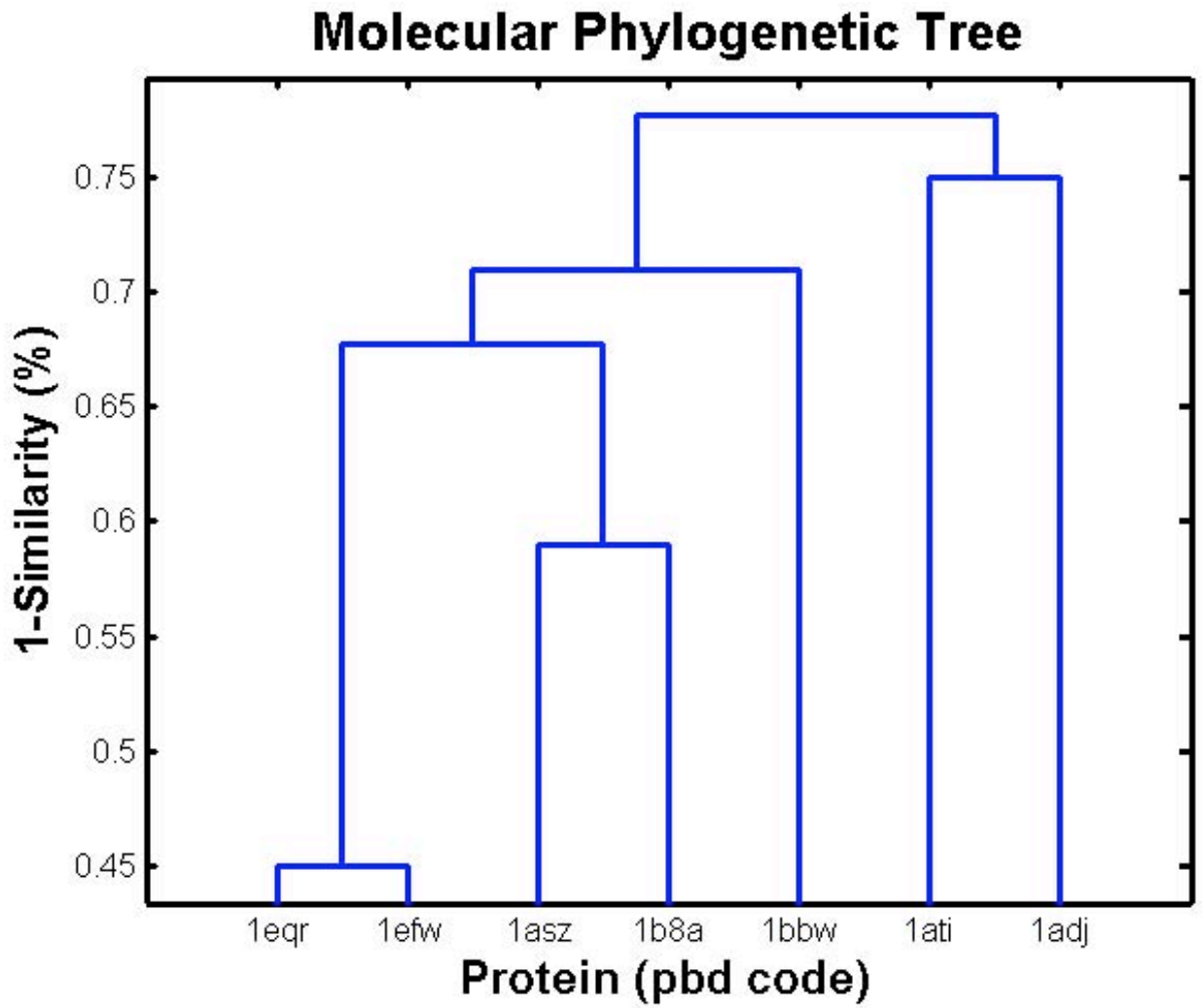
Execute Multiple Alignment Reset

Sequence UPGMA Dendrogram - Clustal



Specificity	Organism	PDB code:chain	ASTRAL catalytic domain
Aspartyl	Eubacteria	1EQR:B	d1eqrb3
Aspartyl	Archaea	1B8A:A	d1b8aa2
Aspartyl	Eukarya	1ASZ:A	d1asza2
Glycl	Archaea	1ATI:A	d1atia2
Histidyl	Eubacteria	1ADJ:C	d1adjc2
Lysl	Eubacteria	1BBW:A	d1bbwa2
Aspartyl	Eubacteria	1EFW:A	d1efwa3

Phylogenetic (UPGMA) Tree - Matlab



Pogorelov and Luthey-Schulten, UIUC 2003

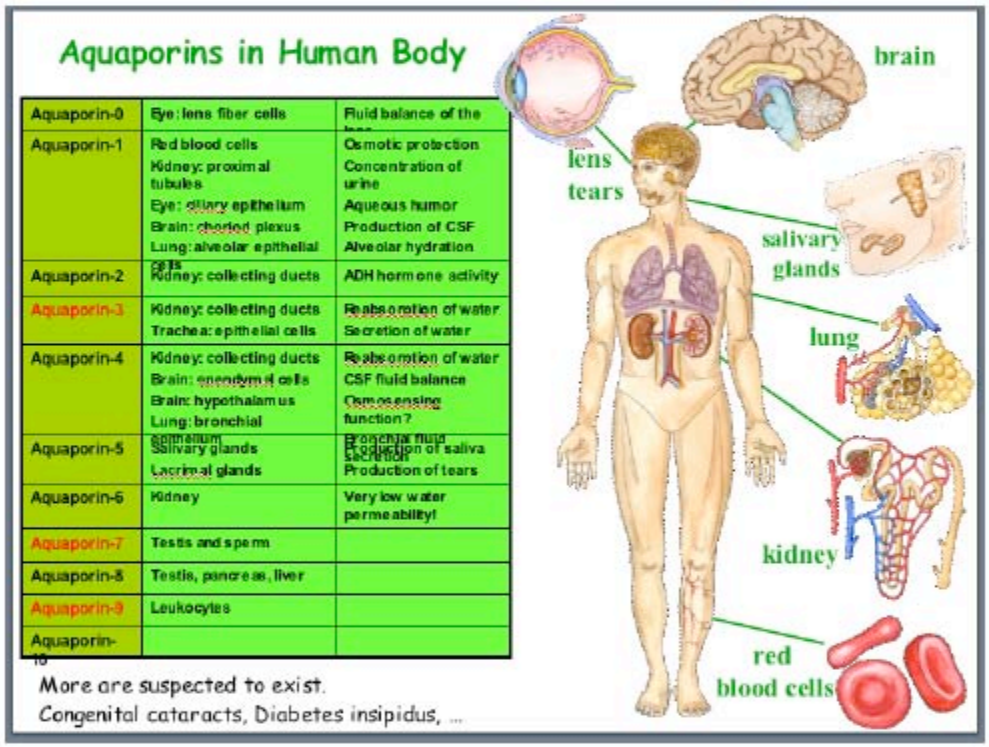
Bioinformatics of Aquaporin Tutorial Week II Perth

Highlights:

- Structural overlap
- Correlation conserved residues and mechanism

University of Illinois at Urbana-Champaign
NIH Resource for Macromolecular Modeling and Bioinformatics
Beckman Institute

Aquaporins



VMD Developers:

John Stone

Dan Wright

John Eargle

Fatemeh Khalili

Elizabeth Villa

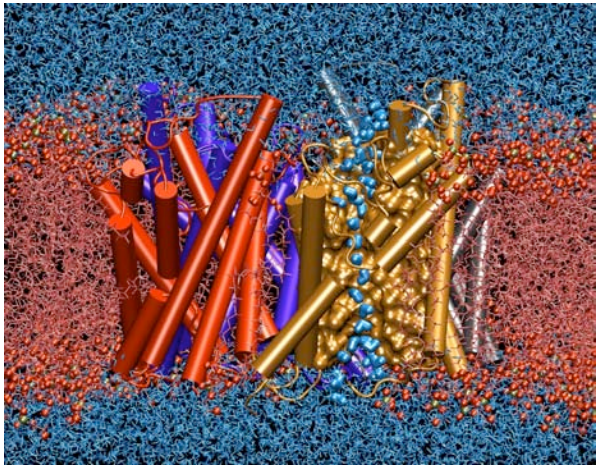
Emad Tajkhorshid

Brijeet Dhaliwal

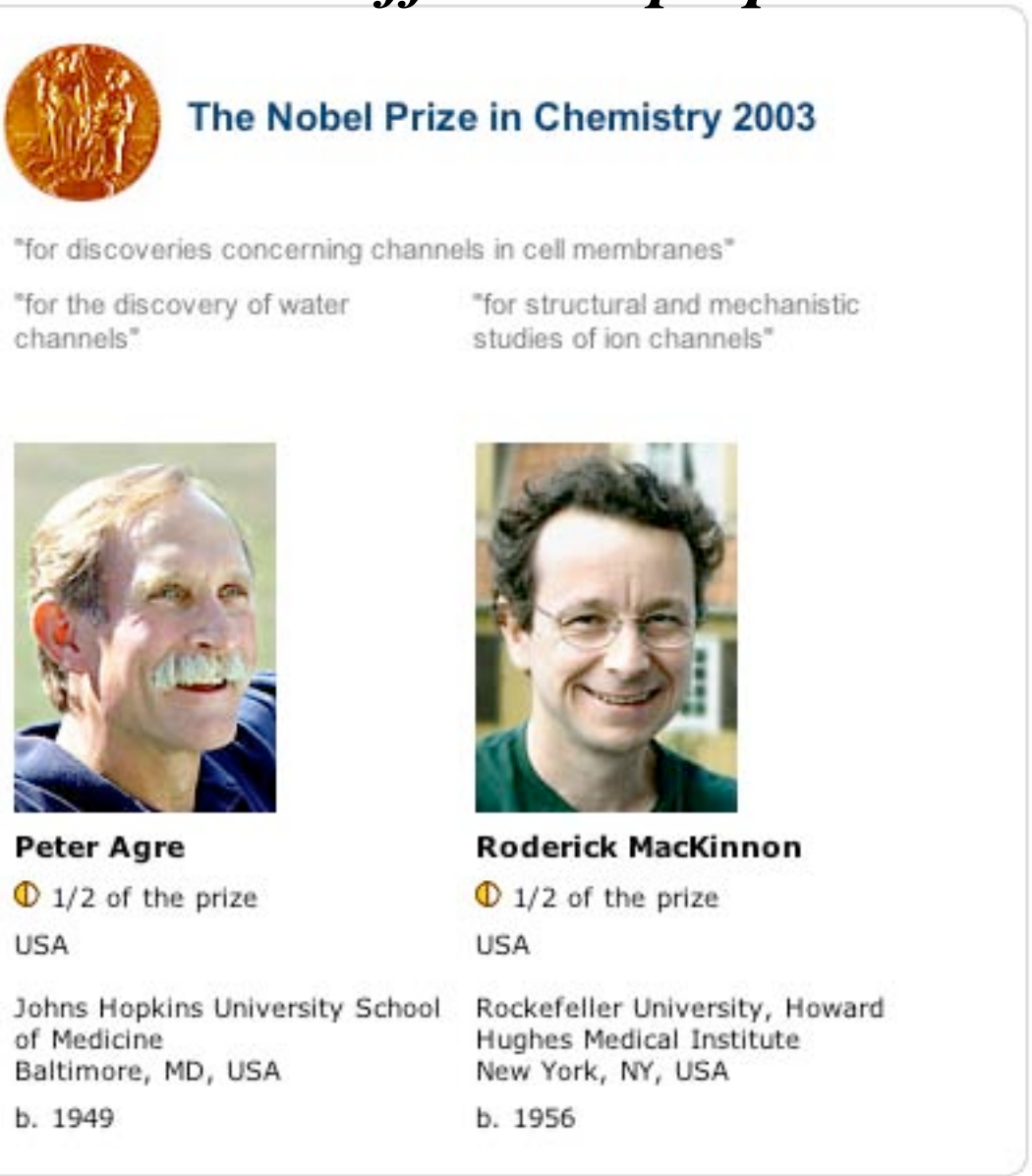
Zan Luthey-Schulten

Towards Understanding Membrane Channels

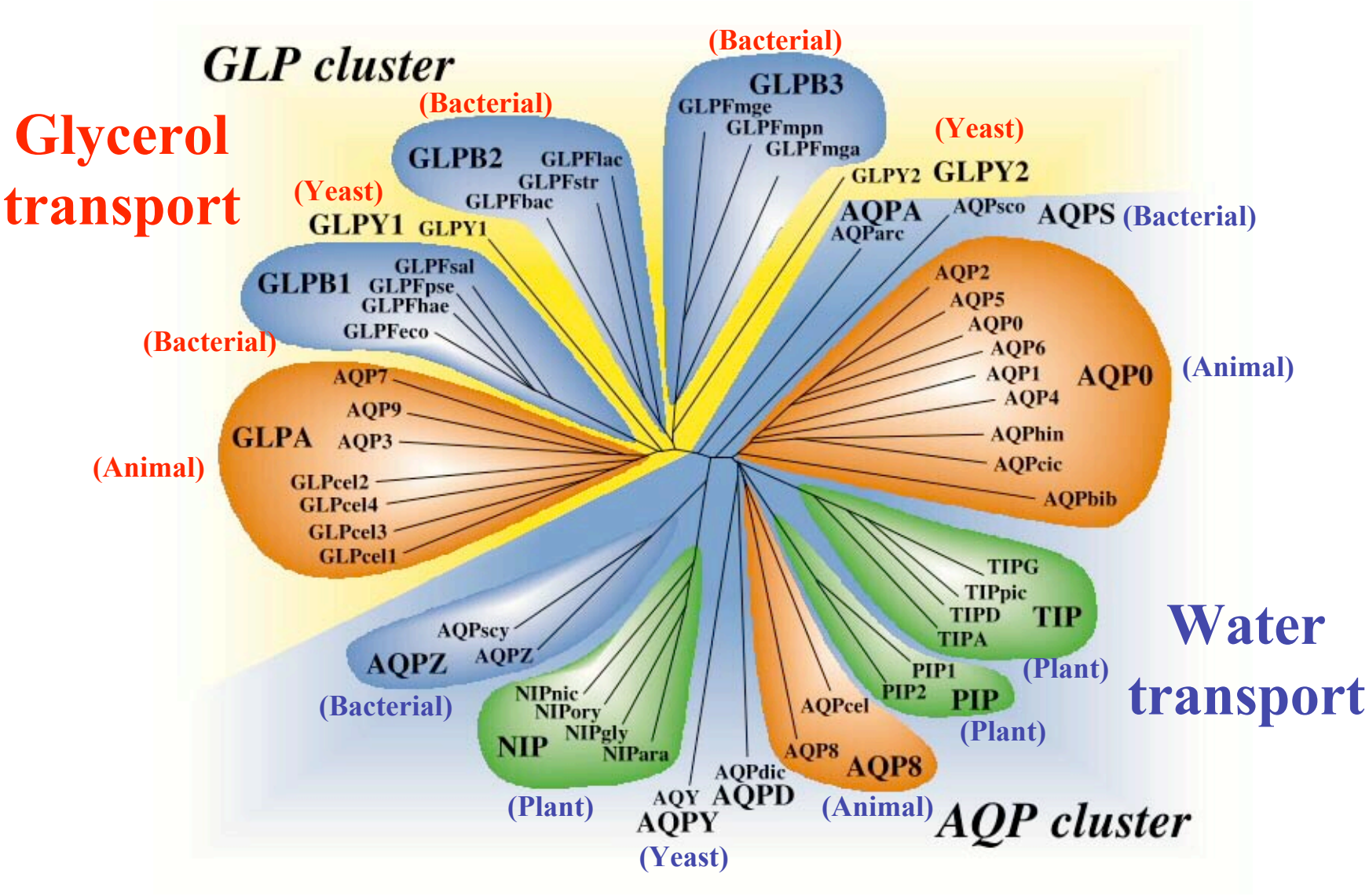
The versaltile, highly selective and efficent aquaporin



GlpF Structure (Stroud et al)
NAMD with full electrostatics
Periodic boundary conditions
NpT ensemble at 310 K
1ns equilibration
Protein: ~ 15,000 atoms
Lipids: ~ 40,000 atoms
Water: ~ 51,000 atoms
Total: ~ 106,000 atoms
4 hrs / ns – 1024 TSC CPUs

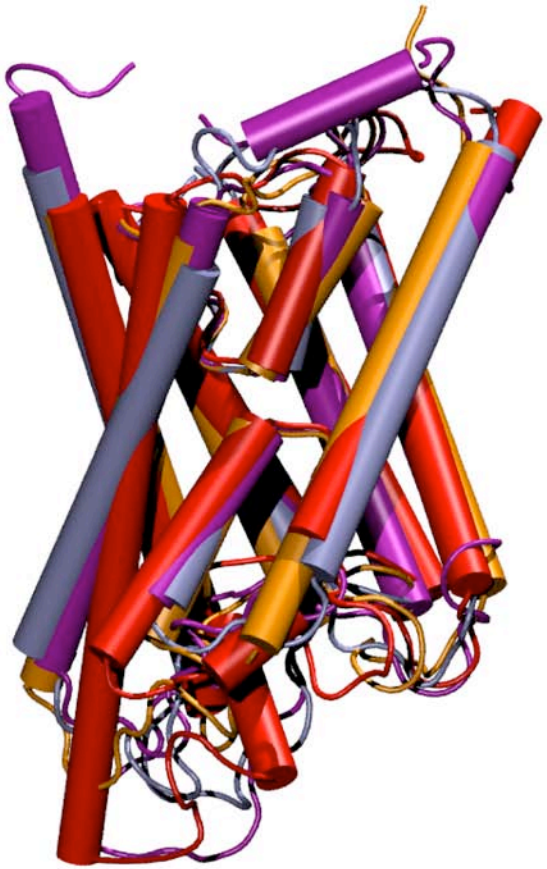


The Aquaporin Superfamily

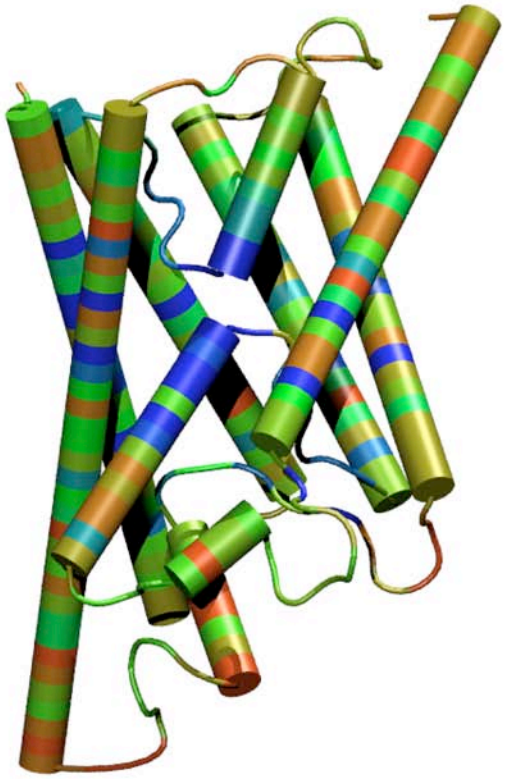


Heymann and Engel *News Physiol. Sci.* **14**, 187 (1999)

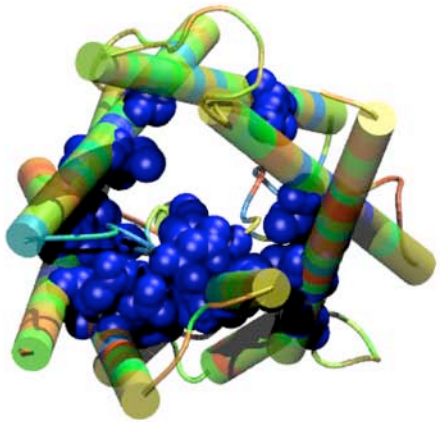
Structure and Sequence Comparisons
Water/Glycerol Channels



2 AQP1, GLPF, AQPZ
from animal and bacteria



GLPF Sequence Conservation



Top view

Acknowledgements - Tutorials

Seq Alignment

- Rommie Amaro
- Felix Autenrieth
- Brijeet Dhaliwal
- Barry Isralewitz
- Taras Pogorelov
- Anurag Sethi

Evolution AARS

- Rommie Amaro
- Patrick O'Donoghue
- Brijeet Dhaliwa

Bioinformatics Aquaporins

- Fatemeh Araghi
- Brijeet Dhaliwal
- Elizabeth Villa

VMD Developers: Dan Wright, John Eargle, John Stone