

# Part III - Evolutionary Studies Using Multiseq in VMD

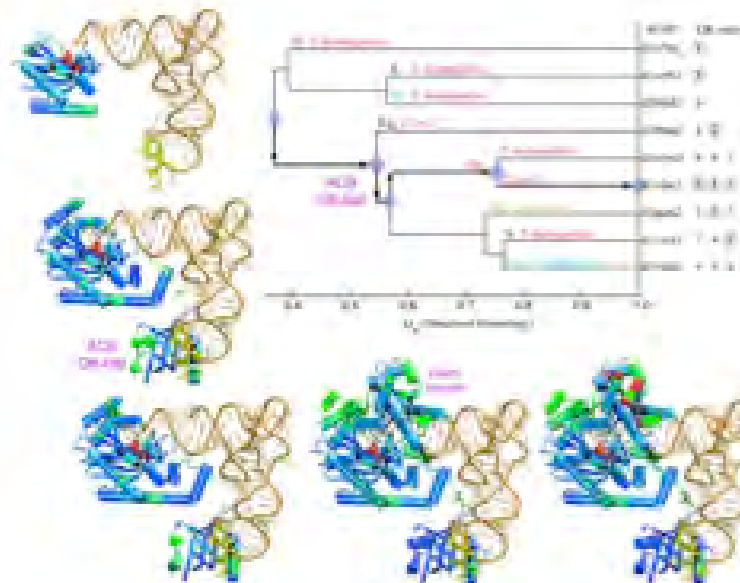
- Aminoacyl tRNA Synthetases
- tRNA
- Aquaporins

Frankfurt, 2006, Computational Biology Workshop

# Evolution of Biomolecular Structure

Class II tRNA-Synthetases and tRNA

---



MultiSeq Developers:

Elijah Roberts

John Eargle

Dan Wright

Prof. Zan Luthey-Schulten

Patrick O'Donoghue

Anurag Sethi

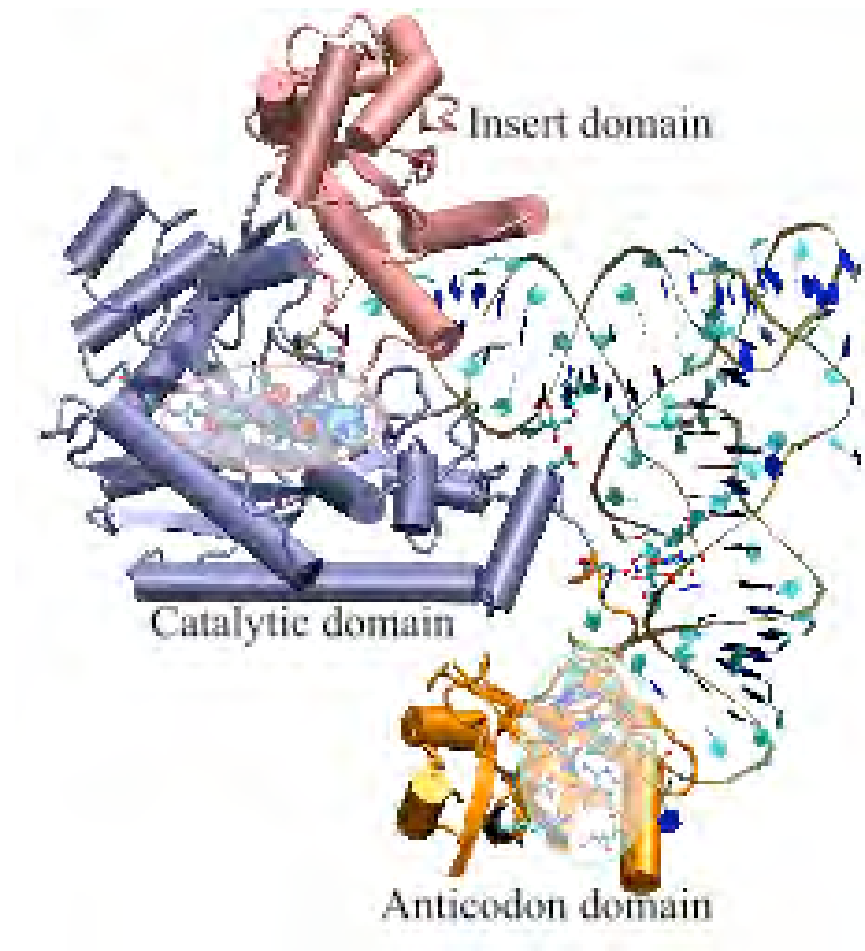
Brijeet Dhaliwal

March 2006.

# Canonical Pattern & Horizontal Gene Transfer

- “The aminoacyl-tRNA synthetases, perhaps better than any other molecules in the cell, optimize the current situation and help to understand the effects of HGT” Woese (PNAS, 2000; MMBR 2000)
- Carl Woese - Crafoord Prize 2003

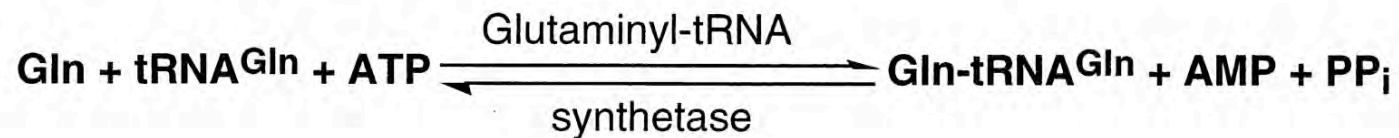
Step 1: Explore active site in catalytic domain and anticodon domain in AspRS from Ecoli



# Charging the tRNA

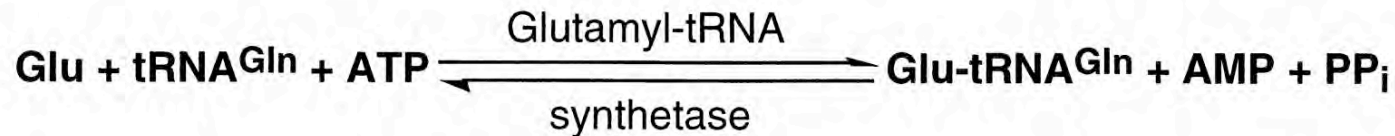
## *Direct acylation*

---



## *tRNA-dependent amino acid modification*

---

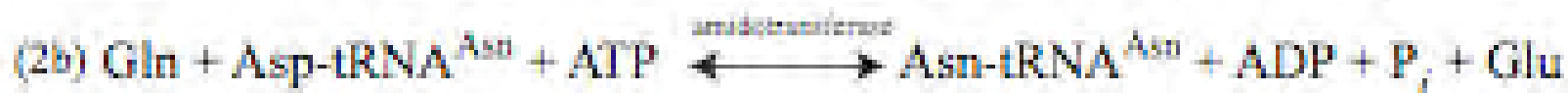


# Amino Acid Biosynthesis and tRNA Charging

## Direct Pathway

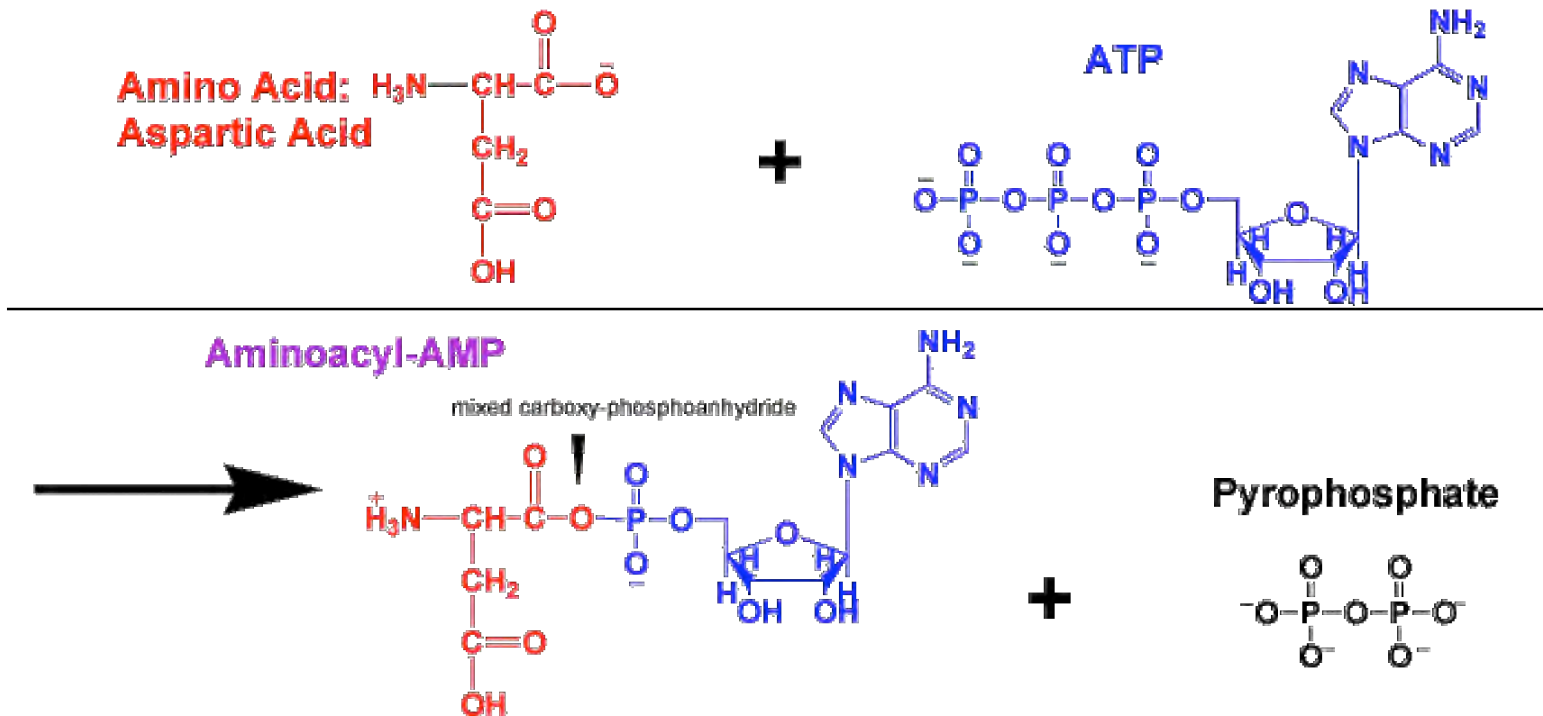


## Indirect Pathway



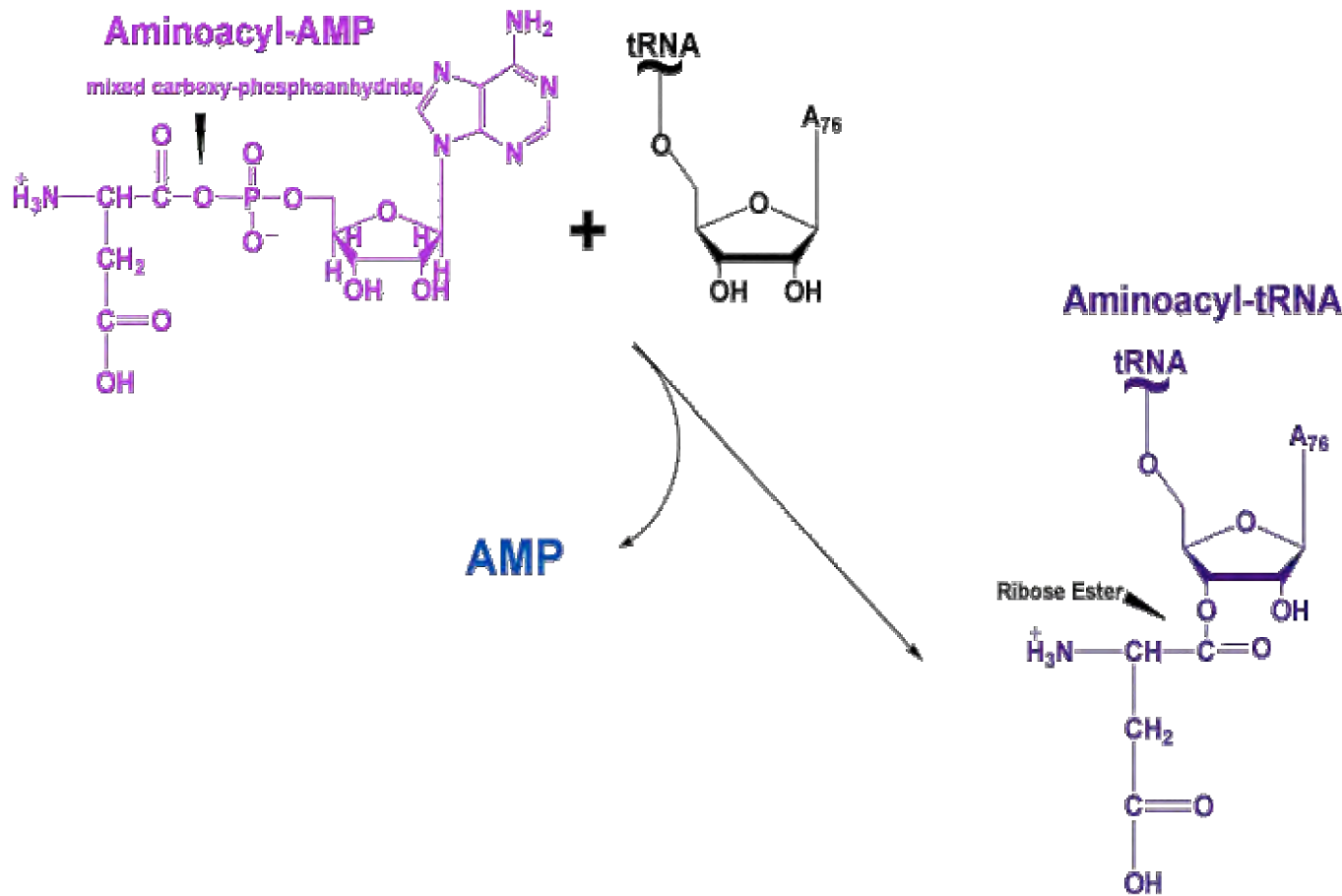
Indirect Pathway - AspRS is non-discriminating. Some organisms do not contain genes to make Asn and use Gln as source of ammonia. In that case ID pathway is the only way to obtain Asn and a direct AsnRS not found in the organism. Similar SepRS/SepCysS

# Step 1: Creation of the Aminoacyl-Adenylate Complex



In [step 1](#), an O atom of the amino acid  $\alpha$ -carboxyl attacks the P atom of the alpha phosphate of ATP. The products are Aminoacyl-AMP containing a mixed carboxy-phosphoanhydride bond and pyrophosphate.

# Step 2: Creation of the Aminoacyl-tRNA



In step 2, the 2' or 3' OH of the terminal adenosine of the 3' end of the tRNA attacks the amino acid carbonyl C atom, creating a ribose ester.



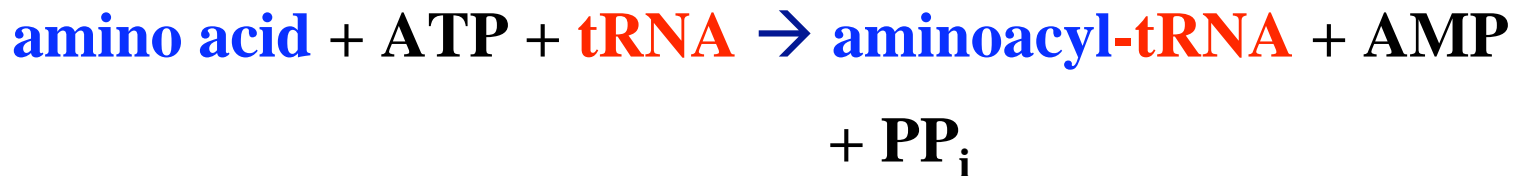
# Aminoacyl-tRNA Synthetase

---

Summary of the 2-step reaction:

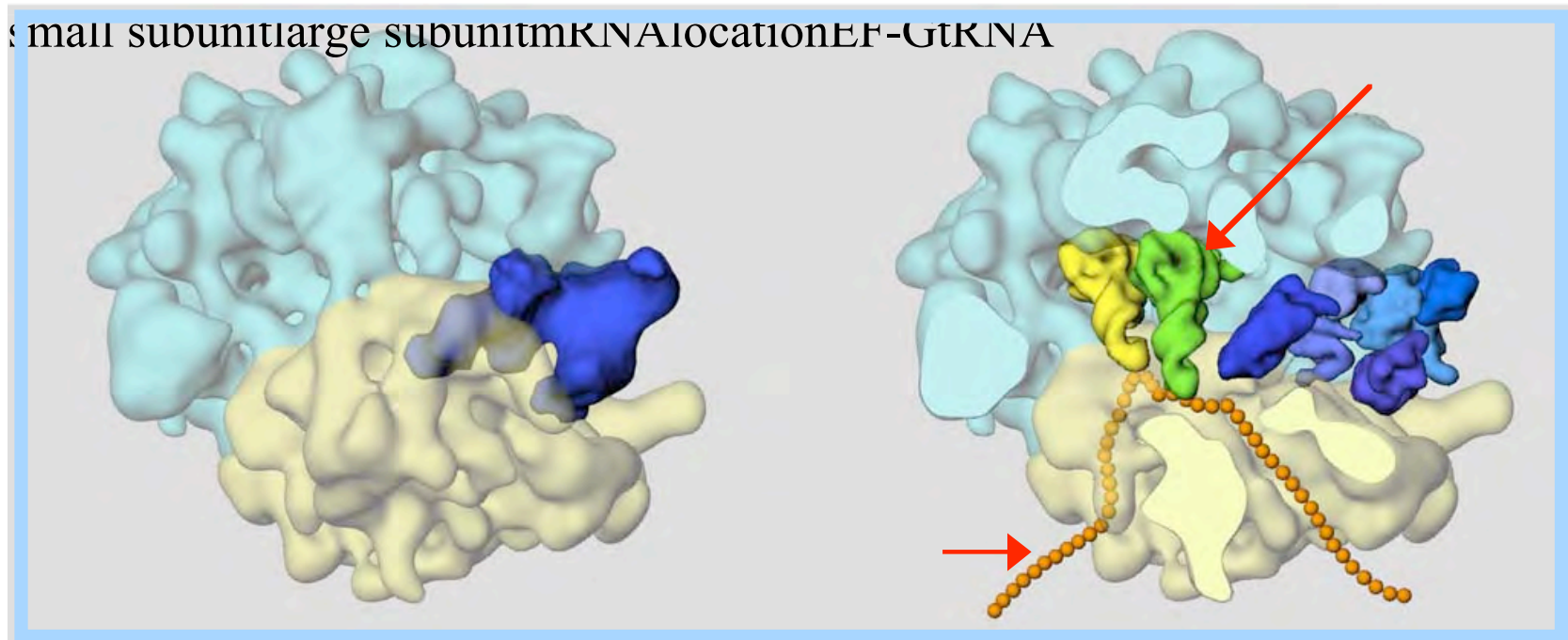


Overall Reaction:



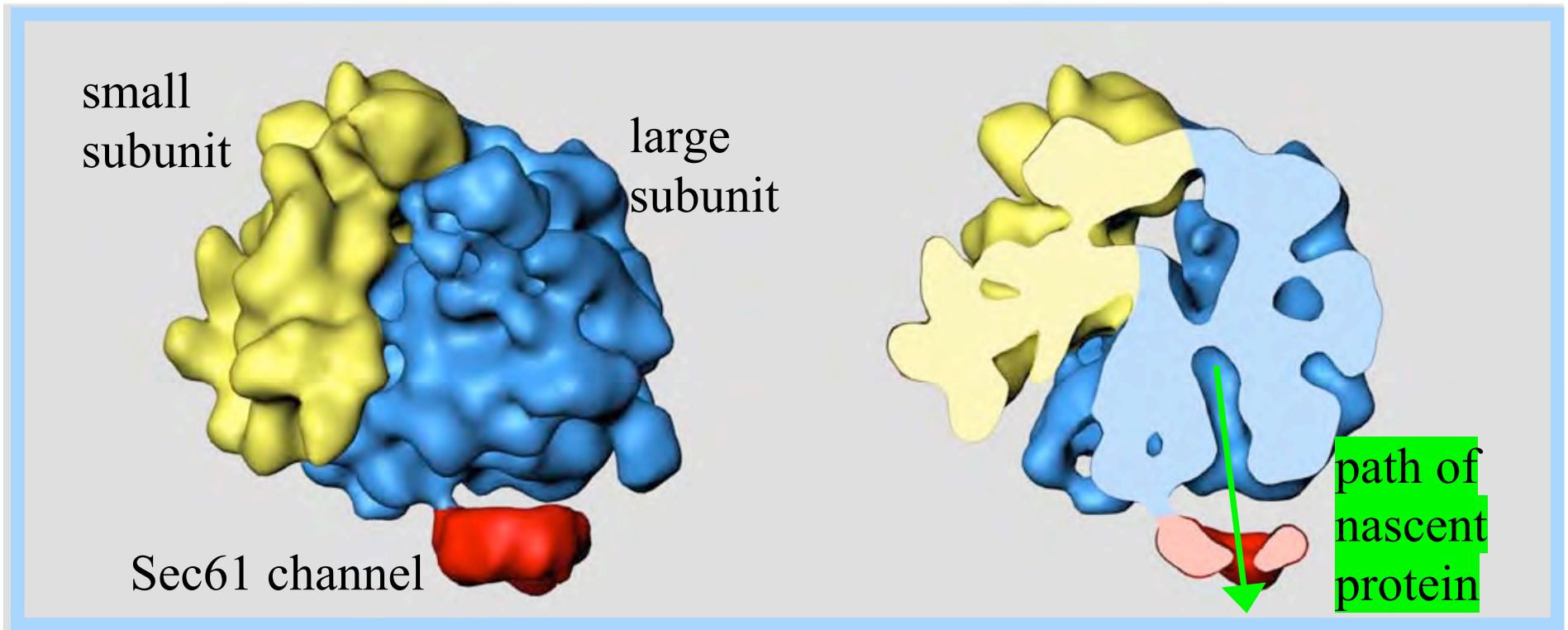
**Next step: EF and Ribosome for Protein Synthesis**

# 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

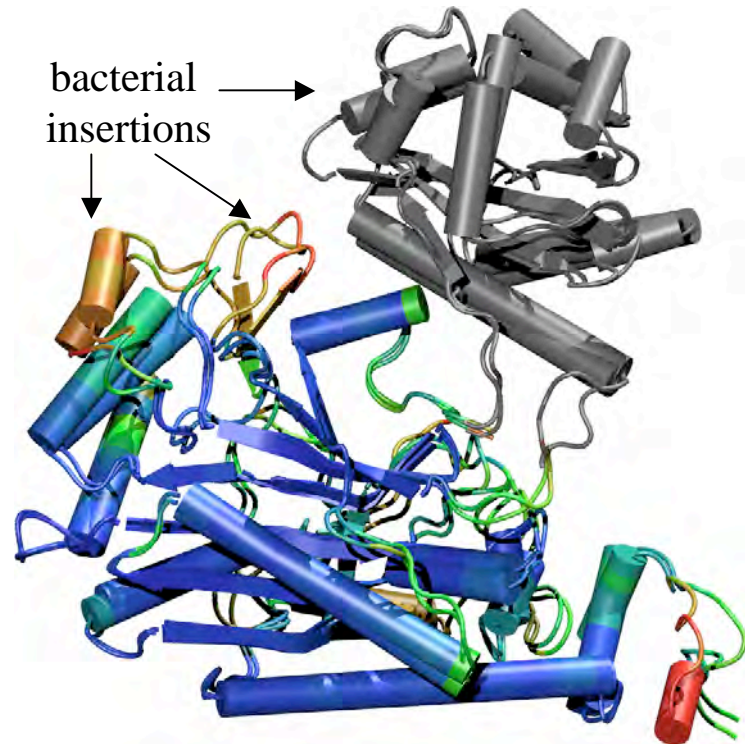
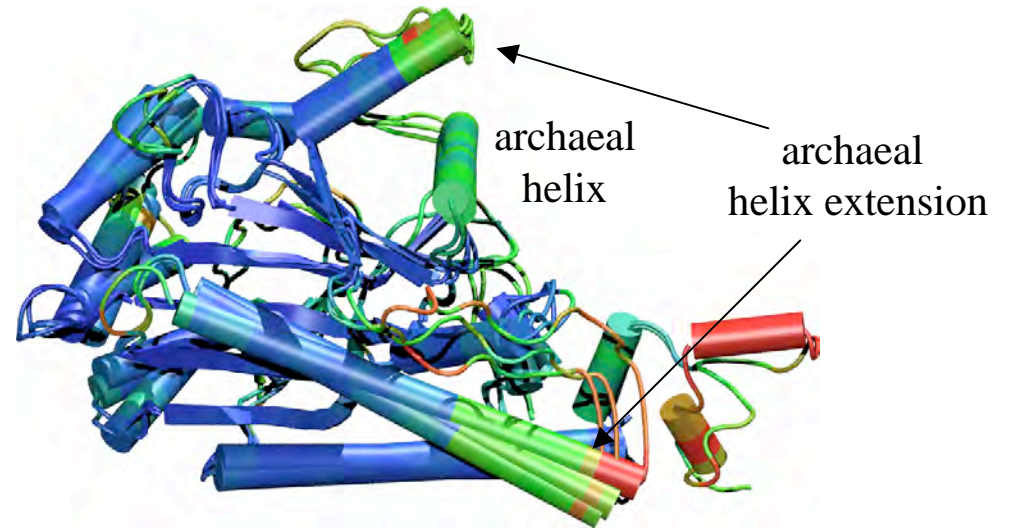
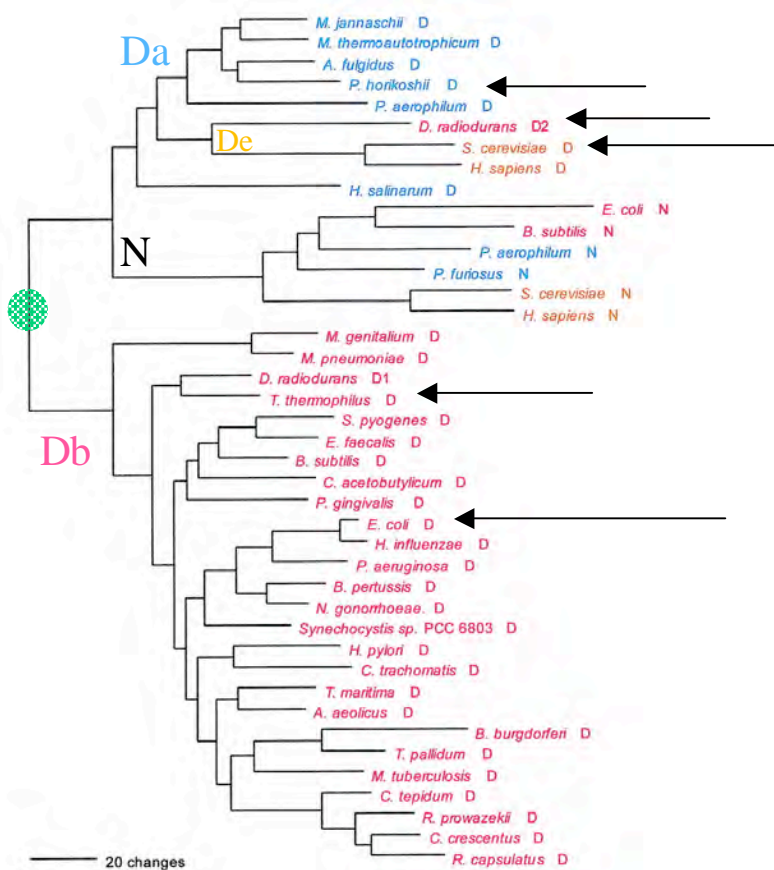


The cutaway view at right shows that the **tunnel** in the yeast large ribosome subunit, through which nascent polypeptides emerge from the ribosome, **lines up** with the lumen of the ER **Sec61 channel**.

# Horizontal Gene Transfer in Protein Structure

## Sequence Phylogeny

### AspRS-AsnRS Group



# Multiseq extension in VMD

VMD 1.8.3a2 OpenGL Display

**Extensions**

- sequence
- autoimd
- apbsrun
- imd
- contactmap
- pdbtool
- ramaplot
- rmsd
- solvate
- timeline
- multiseq
- tkcon
- vmdmovie

**treeWindow**

Tree

```

graph LR
    A[d1efwa3.ent Thermus thermophilus B] --- B(( ))
    C[d1c0aa3.ent Escherichia coli B] --- B
    B --- D(( ))
    E[d1n9wb1.ent d1n9wb1.ent] --- D
    D --- F(( ))
    G[d1asza2.ent Saccharomyces cerevisiae E] --- F
    H[d1b8aa2.ent Pyrococcus kodakaraensis A] --- F
  
```

0.56

**Sequence Display**

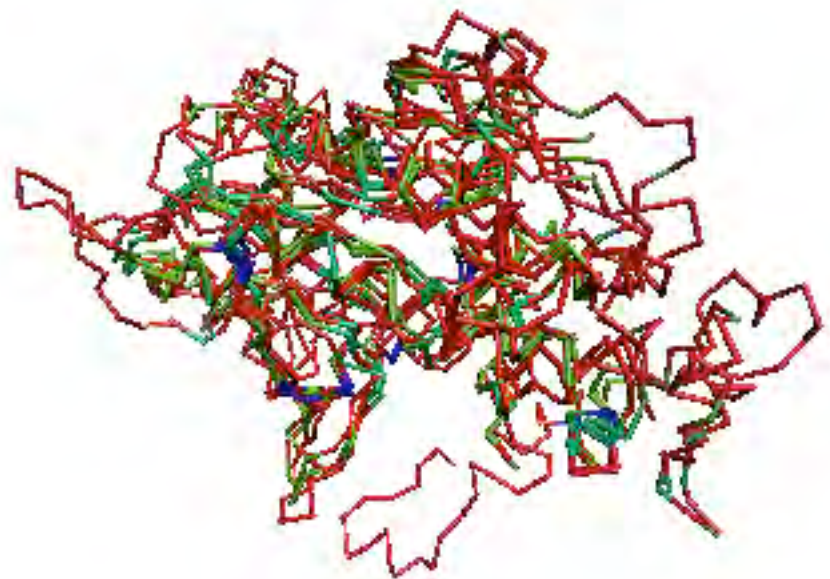
```

d1b8aa2.ent  IDTEGERLLGKYM--MENENAPLYFLYQYPS-----EAKPFYIMKYDN-----K--PEICRAFDLEYRGM
d1asza2.ent  LSTENEKFLGKLV--RDKYDTDFYILDKFPL-----EIRPFYTMPDPA-----N--PKYSNSYDFFMRGEI
d1n9wb1.ent  LSEEAERLLGEYA--KERWGSDFWLVTRYPR-----SVRPFYTYTYP-EE-----DGTTRSFDLLFRGLI
d1c0aa3.ent  ---GSD-KP-DLRDE---SKWAPLWVIDFPMFE-DDGEGGLTAMHHPFTSPK-DMTAAELKAAPENAVANAYDMVINGYI
d1efwa3.ent  ---GSD-KP-DL-RR---EGFRFLWVVDFFLLEWDEEEEAWTYMHHHPFTSPHPED-LPLLEKDPGRVRLAYDLVLNGVI
  
```

# Conservation

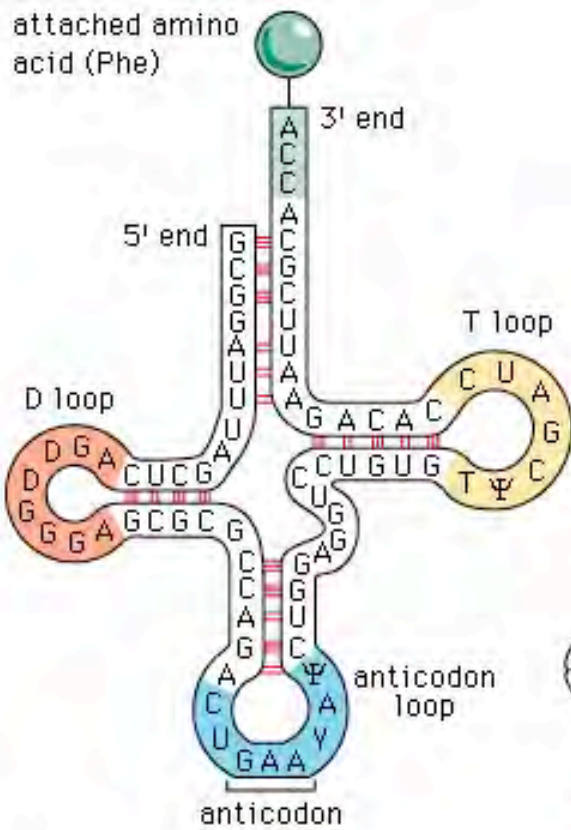


Core Structure Conserved



Sequence Identity of Core  
Less than 15%

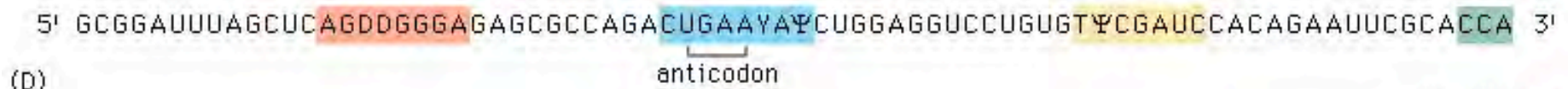
# tRNA Structure



(A)

(B)

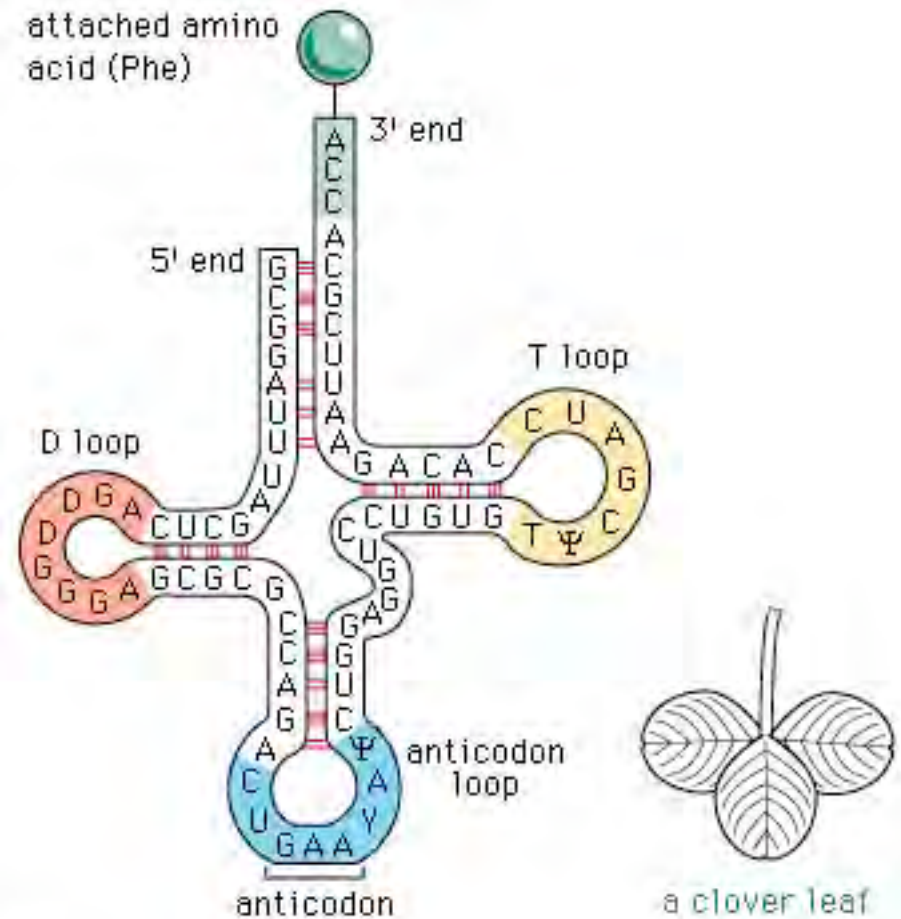
(C)



# tRNA Secondary Structure

Most RNAs have **secondary cloverleaf structure**, consisting of stem & loop domains.

Double helical **stems** arise from **base pairing** between complementary stretches of bases within the same strand.

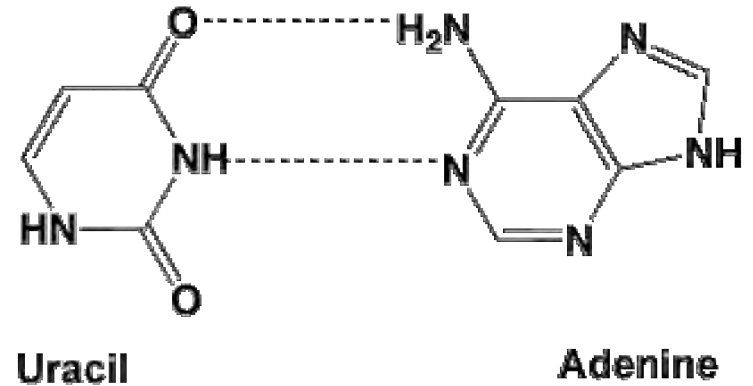


(A)

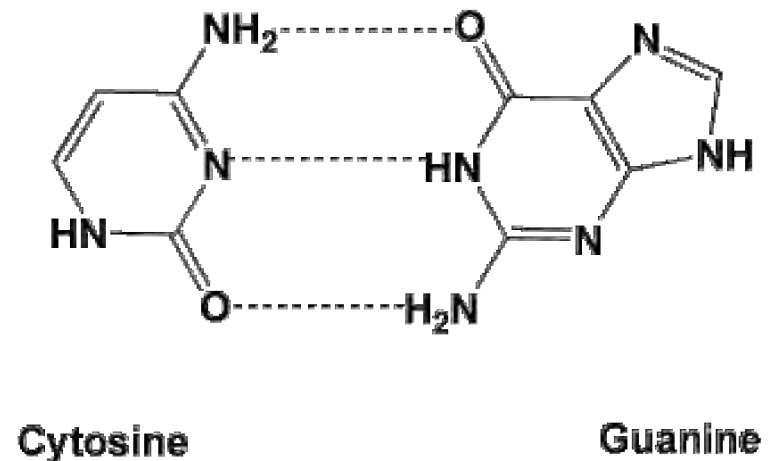
**Loops** occur where **lack of complementarity**, or the presence of **modified bases**, prevents base pairing.



**Hydrogen bonds** link 2 complementary nucleotide bases on separate nucleic acid strands, or on complementary portions of the same strand.



Conventional **base pairs**: **A & U (or T)**; **C & G**.



# Genetic code

The **genetic code** is based on the sequence of bases along a nucleic acid.

Each **codon**, a sequence of **3 bases** in mRNA, codes for a particular amino acid, or for chain termination.

Some amino acids are specified by 2 or more codons.

**Synonyms** (multiple codons for the same amino acid) in most cases differ only in the 3<sup>rd</sup> base. Similar codons tend to code for similar amino acids. Thus effects of mutation are minimized.

The Standard Genetic Code

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

# tRNA Databases and Web Resources

**MFOLD : Prediction of RNA secondary structure (M. Zuker)**

<http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html>

**Vienna RNA Package (Ivo Hofacker)**

<http://www.tbi.univie.ac.at/~ivo/RNA/>

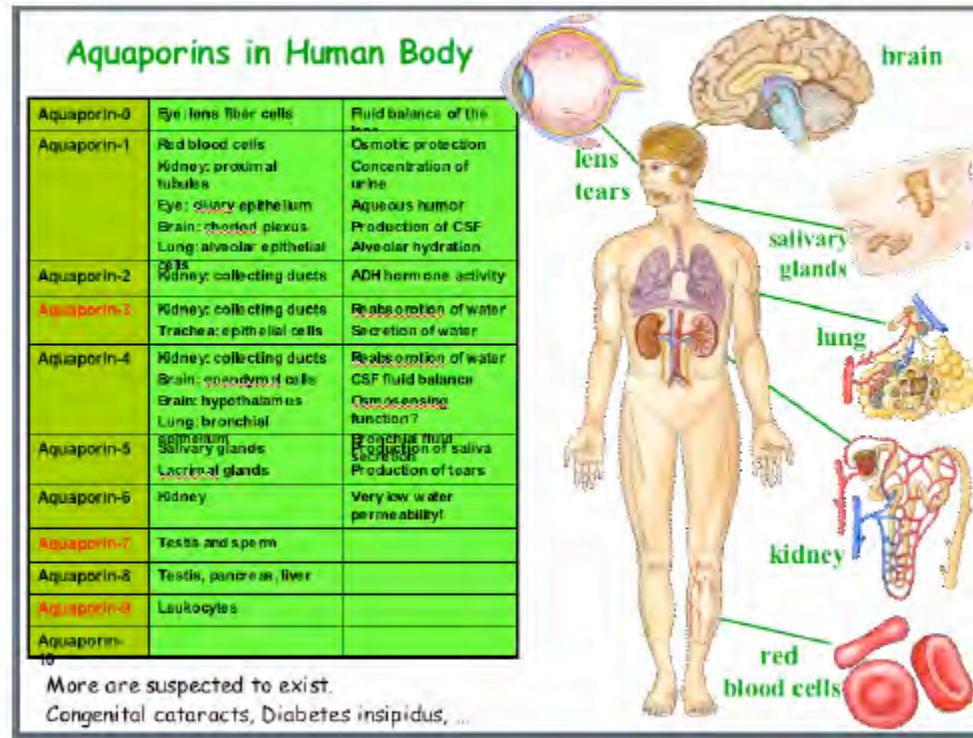
**DOE Joint Genome Institute**

<http://www.jgi.doe.gov/>

**Compilation of tRNA sequences and sequences of tRNA genes (Mathias Sprinzl)**

<http://www.uni-bayreuth.de/departments/biochemie/trna/>

# Aquaporins



VMD Developers:

John Stone

Dan Wright

John Eargle

Fatemeh Khalili

Elizabeth Villa

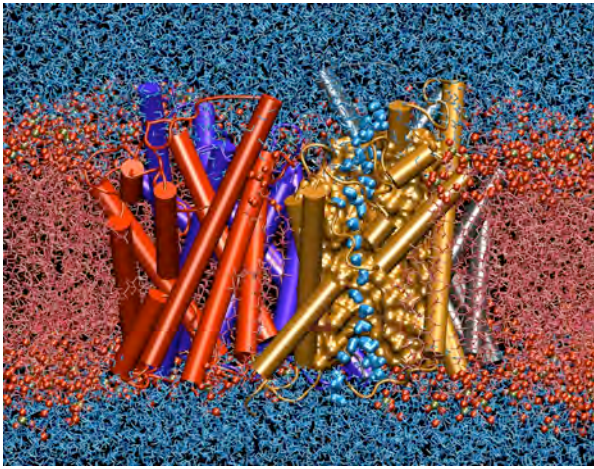
Emad Tajkhorshid

Brijeet Dhaliwal

Zan Luthey-Schulten

# Towards Understanding Membrane Channels

*The versatile, highly selective and efficient 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 Nobel Prize in Chemistry 2003

"for discoveries concerning channels in cell membranes"

"for the discovery of water channels"

"for structural and mechanistic studies of ion channels"



**Peter Agre**

🏆 1/2 of the prize

USA

Johns Hopkins University School of Medicine  
Baltimore, MD, USA

b. 1949



**Roderick MacKinnon**

🏆 1/2 of the prize

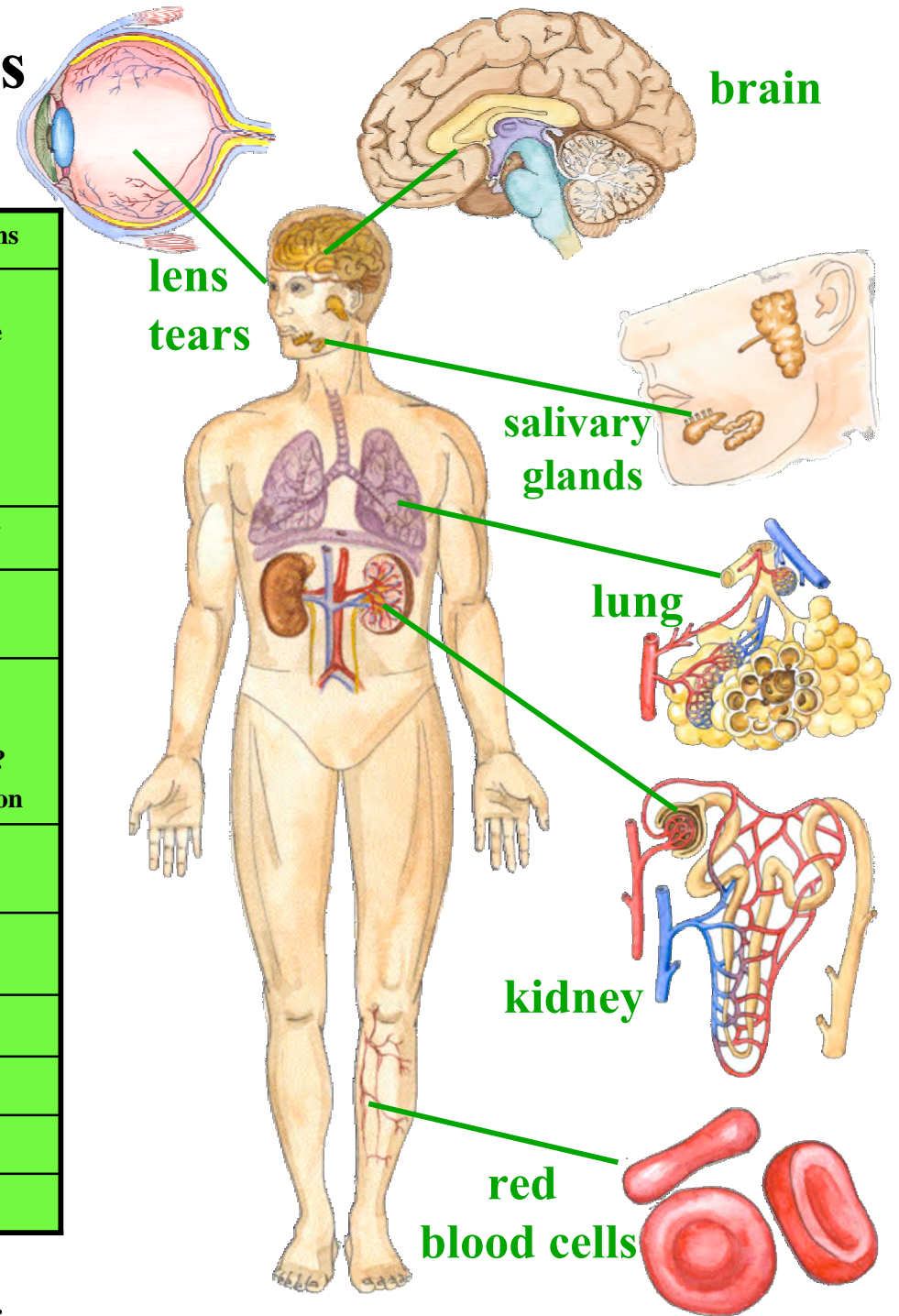
USA

Rockefeller University, Howard Hughes Medical Institute  
New York, NY, USA

b. 1956

# Water and **Glycerol** Channels in the Human Body

<b>Aquaporin-0</b>	Eye: lens fiber cells	Fluid balance of the lens
<b>Aquaporin-1</b>	Red blood cells Kidney: proximal tubules Eye: ciliary epithelium Brain: choroid plexus Lung: alveolar epithelial cells	Osmotic protection Concentration of urine Aqueous humor Production of CSF Alveolar hydration
<b>Aquaporin-2</b>	Kidney: collecting ducts	ADH hormone activity
<b>Aquaporin-3</b>	Kidney: collecting ducts Trachea: epithelial cells	Reabsorption of water Secretion of water
<b>Aquaporin-4</b>	Kidney: collecting ducts Brain: ependymal cells Brain: hypothalamus Lung: bronchial epithelium	Reabsorption of water CSF fluid balance Osmosensing function? Bronchial fluid secretion
<b>Aquaporin-5</b>	Salivary glands Lacrimal glands	Production of saliva Production of tears
<b>Aquaporin-6</b>	Kidney	Very low water permeability!
<b>Aquaporin-7</b>	Testis and sperm	
<b>Aquaporin-8</b>	Testis, pancreas, liver	
<b>Aquaporin-9</b>	Leukocytes	
<b>Aquaporin-10</b>		

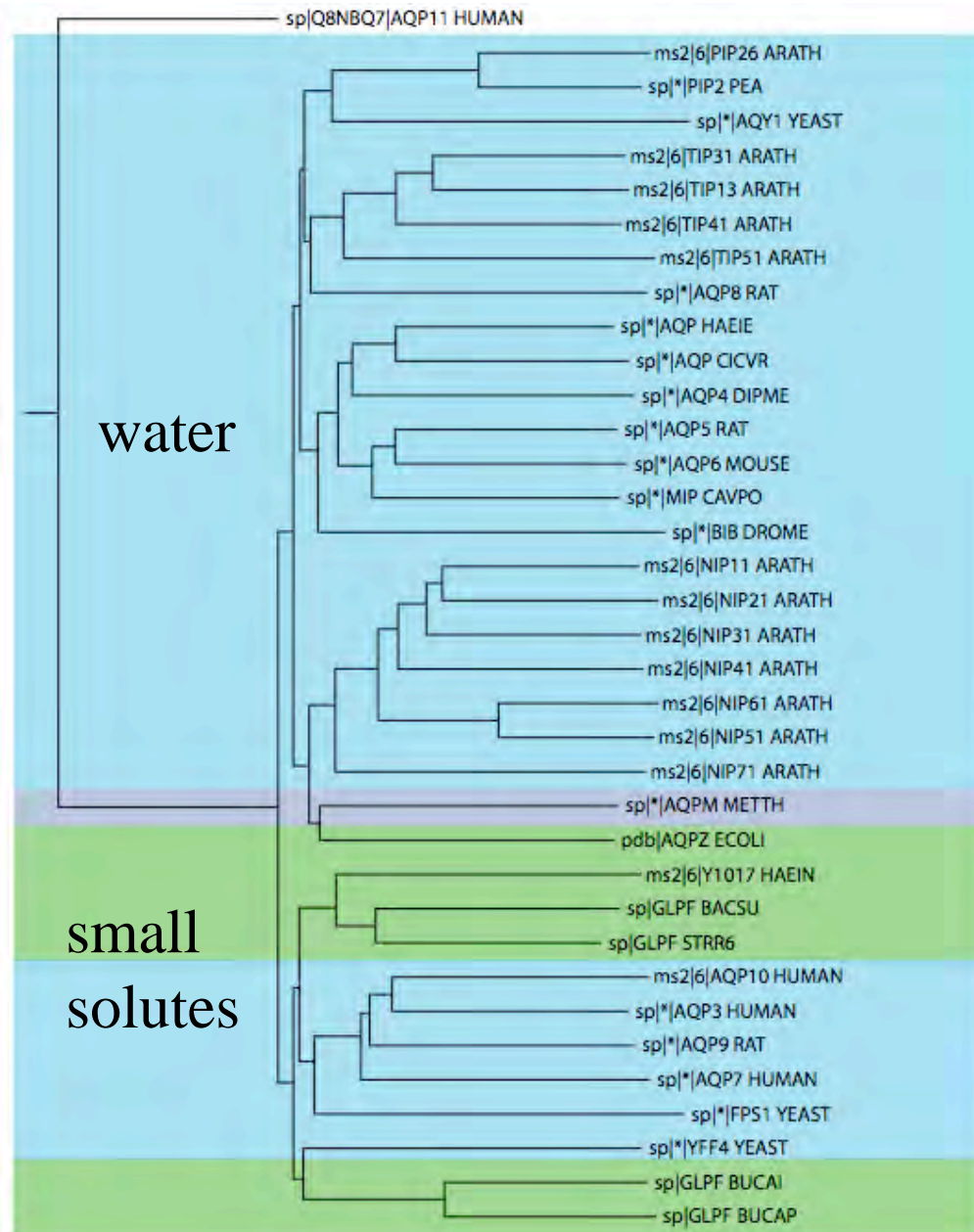


Additional members are suspected to exist.



# Phylogenetic Analysis of Aquaporins & Aquaglycerol Porins

Rooted AQP11



PIP, TIP, MIP  
AQP0, 1, 4, 5, 6, 8

AQPM, AQPZ

GLPF  
AQP3, 7, 9, 10

Yi Wang, A. Sethi  
ZLS - unpublished



# Functionally Important Features of Aquaporins

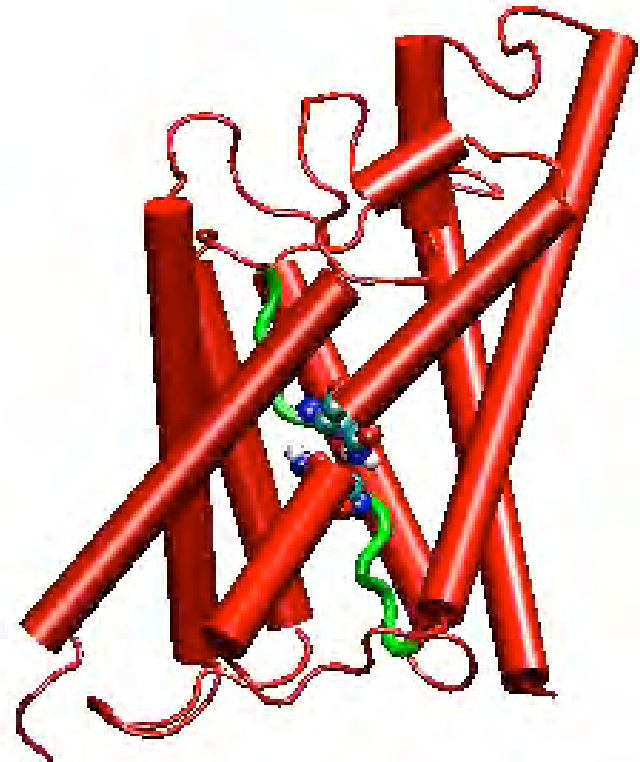
- Water and glycerol transport
- Exclusion of ions and protons
- Tetrameric arrangement in membrane

Aquaporins of known structure:

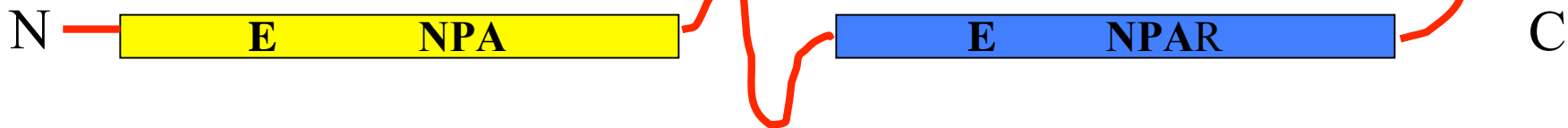
**GlpF** – E. coli glycerol channel (aquaglyceroporin)

– Fu, et al., Science (2000)

**AQP1** – Mammalian aquaporin-1 (pure water channel) -Sui et al, Nature (2001)



**~100% conserved -NPA- signature sequence**





# Load Aquaporins 1j4n, 1fqy, 1lda, 1rc2 into VMD

The screenshot displays the VMD (Visual Molecular Dynamics) software interface. The main window, titled "VMD 1.8.2b7 OpenGL Display", shows a 3D ribbon representation of four aquaporin proteins: 1j4n (grey), 1fqy (red), 1lda (blue), and 1rc2 (yellow). The "VMD Main" window at the top left contains a table of loaded molecules:

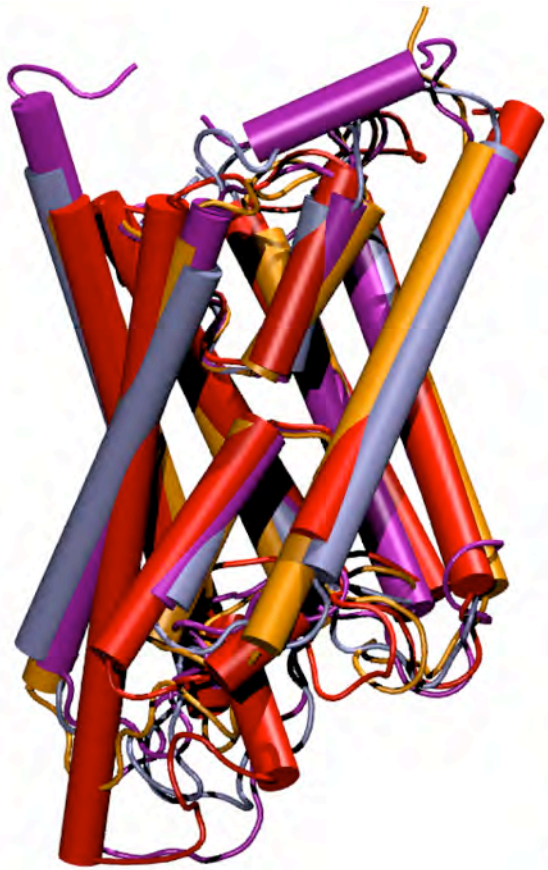
ID	T	A	D	F	Molecule	Atoms	Frames	Vol
1	A	D	F		1J4N	2029	1	0
2	A	D	F		1FQY	1661	1	0
3	A	D	F		1lda	1997	1	0
5	T	A	D	F	1rc2	3530	1	0

The "Graphical Representations" window on the left shows the configuration for the selected molecule "5: 1rc2". The "Style" is set to "Tube", "Color" is "ColorID 3", and "Selection" is "chain A". The "Drawing Method" is also set to "Tube". The "Multiple Sequence Alignment" window at the bottom right displays the following alignment:

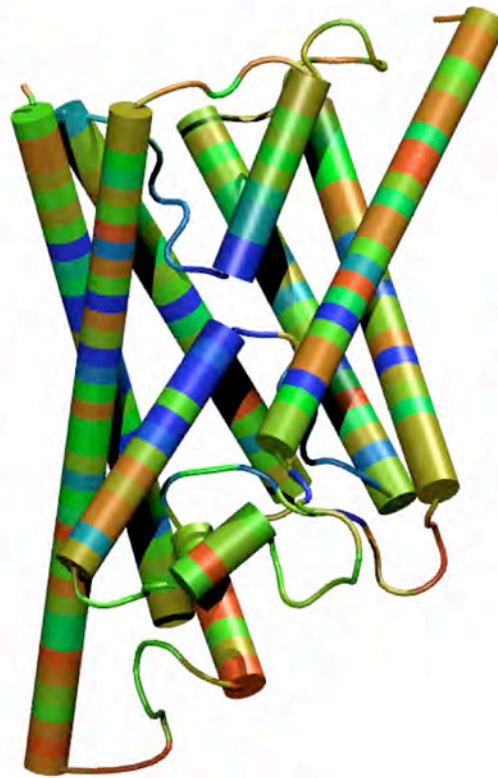
PDB code	Description
1j4n	Bovine AQP1
1fqy	Human AQP1
1lda	E. coli Glycerol Facilitator (GlpF)
1r2c	E. coli AqpZ

```
d1fqya_.ent  KLFWRVAVAEFLATLTFVFIISIGSALGFRYPVGNHOTAVQDNVKSFLAFGLSIATLA
d1j4na_.ent  MASEFRKKLFWRAVVAEFLAHILFIPISIGSALGFHYPIKSNQTTGAVQDNVKSFLA
d1lda_.ent   TLNGQCIAEFLGTGLLIFFGVGCVAALKVAGASFGQWEISVINGLGVAMAYLTAGV
d1rc2a_.ent  HFRKLAABCFTFWLVPGGCSAVLAAGFPPELIGIFAGVALAFGLTVLTHAFVGH
```

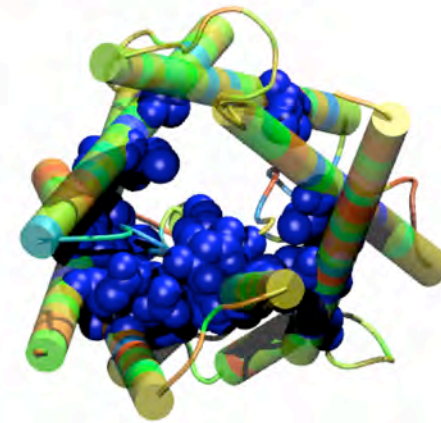
# Structure and Sequence Comparisons Water/Glycerol Channels



2 AQP1, GLPF, AQPZ  
from animal and bacteria



GLPF Sequence Conservation



Top view

# Showing Conserved Residues - Tetramer

The image displays the VMD (Visual Molecular Dynamics) interface. The top-left window, titled "VMD Main", contains a table of loaded molecules and playback controls. The table lists four molecules: 1J4N (2029 atoms), 1FQY (1661 atoms), 1lda (1997 atoms), and 1rc2 (3530 atoms). The bottom-left window, "Graphical Representations", shows settings for the selected molecule "5: 1rc2", including a table for "Selected Atoms" (chain A) and drawing options like "Coloring Method" (ColorID 3) and "Drawing Method" (Tube). The main display window, "VMD 1.8.2b7 OpenGL Display", shows a 3D ribbon representation of a tetramer protein structure. The structure is composed of four subunits, each represented by a different color: red, blue, orange, and grey. Conserved residues are highlighted in yellow. Below the ribbon representation, a blue surface representation of the same tetramer is shown, with the conserved residues also highlighted in yellow.

ID	T	A	D	F	Molecule	Atoms	Frames	Vol
1	A	D	F		1J4N	2029	1	0
2	A	D	F		1FQY	1661	1	0
3	A	D	F		1lda	1997	1	0
5	T	A	D	F	1rc2	3530	1	0

Style	Color	Selection
Tube	ColorID 3	chain A

Selected Atoms: chain A

Coloring Method: ColorID 3, Material: Opaque

Drawing Method: Tube

Radius: 0.5, Resolution: 11

Apply Changes Automatically: Apply

# Acknowledgements - Tutorials

## Evolution AARS/tRNA

- John Eargle
- Elijah Roberts
- Patrick  
O'Donoghue
- Rommie Amaro

## Bioinformatics Aquaporins

- Fateme Araghi
- Elizabeth Villa
- Yi Wang
- Anurag Sethi

VMD Developers: Dan Wright, John Eargle, John Stone  
Editorial Assistant: Brijet Dhaliwal