Evolution of Translation: Dynamics of Recognition in RNA:Protein Complexes

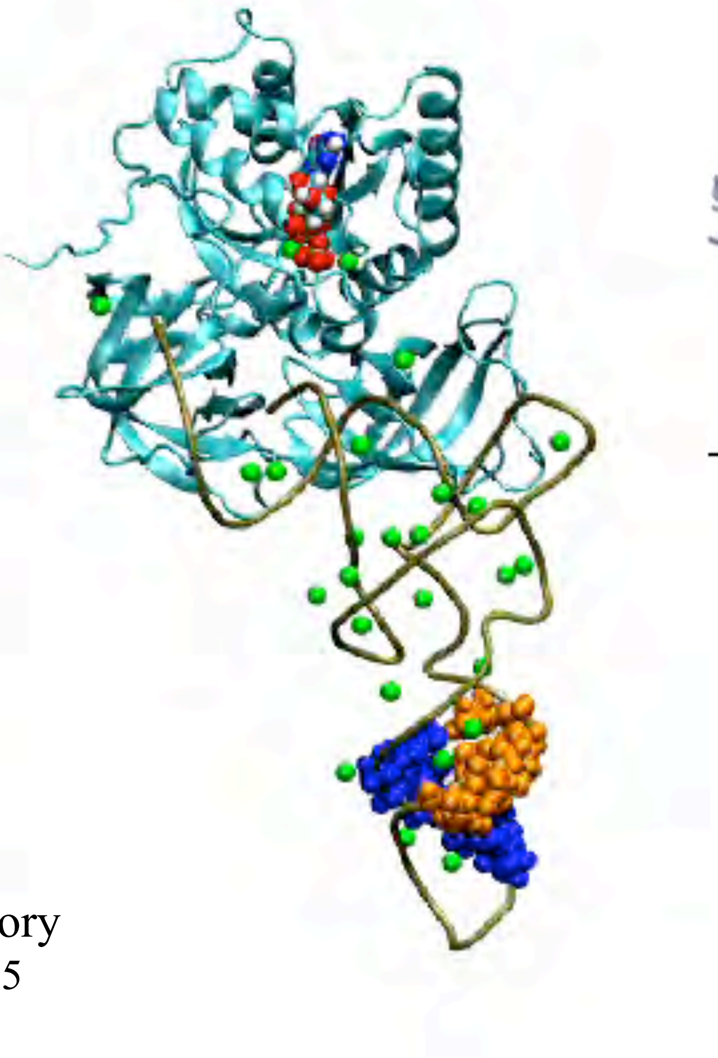
Zaida (Zan) Luthey-Schulten
Dept. Chemistry, Beckman Institute, Biophysics, Institute of Genomics Biology, & Physics

NIH Workshop 2009

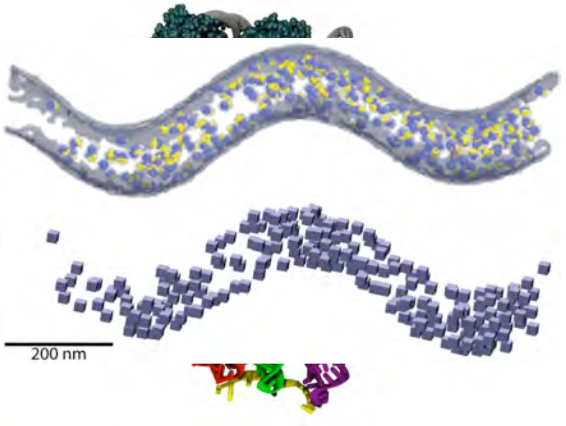
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

On the Evolution of Translation in the Modern World

SepRS/CysRS Evolutionary History
Sethi, O’Donoghue, ZLS, PNAS 2005
O’Donoghue, Sethi, Woese, ZLS,
PNAS 2005; Signaling Network
Sethi,Eargle,Black, ZLS, *PNAS* 2009



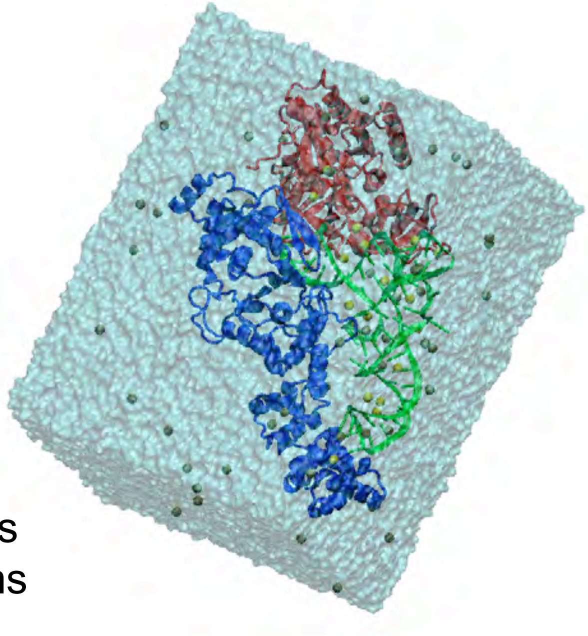
Dynamical Recognition Novel Amino Acids
Eargle, Sethi, Black, ZLS, JMB 2008



Proteins/RNA
Polyribosomes
Ribosome

Evolutionary Analysis & Molecular Dynamics

MD performed with NAMD2 - System Setup



Simulation Parameters

Minimization: 290,000 steps
Production run: 108 ns
Forcefield: CHARMM27(1)
Time step: 1 fs
VdW frequency: 2 fs
VdW cutoff: 12 Å
Switching distance: 10 Å
Pair list distance: 14 Å
Particle Mesh Ewald
Full electrostatic update: 4 fs
Ensemble: NPT
Langevin temperature: 298.15 K
Langevin pressure: 1 atm
Periodic boundary conditions

Contents of System

GluRS
Glu-tRNA^{Glu}
EF-Tu
GTP
Ions: Mg^{2+}, K^+
H_2O: ~27,000 molecules
System: ~130,000 atoms

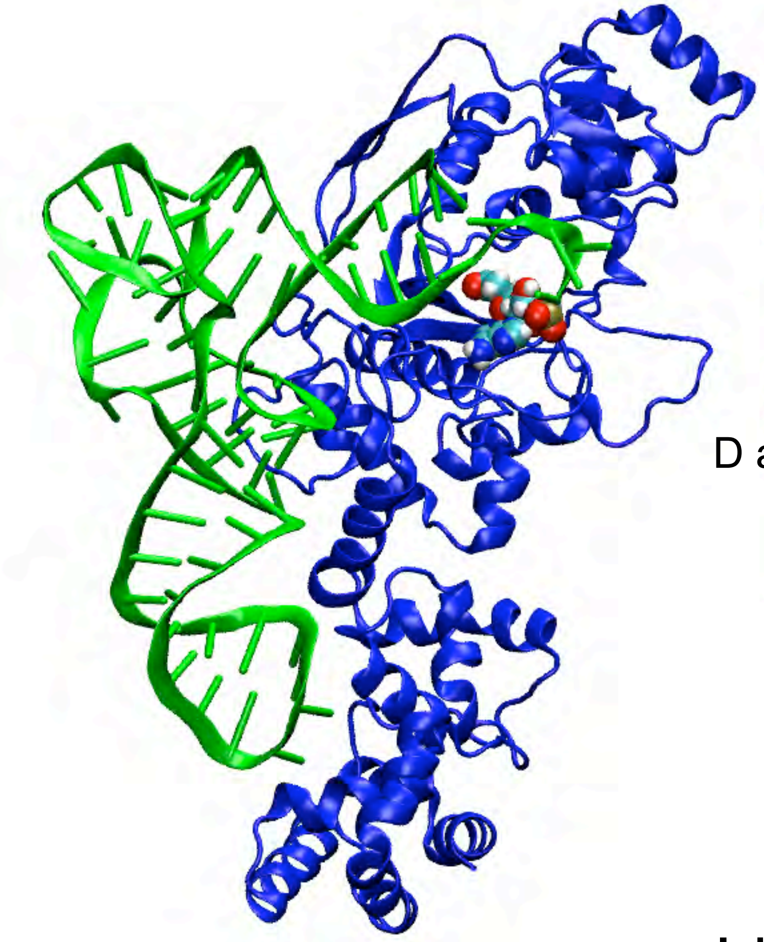
System Perturbations

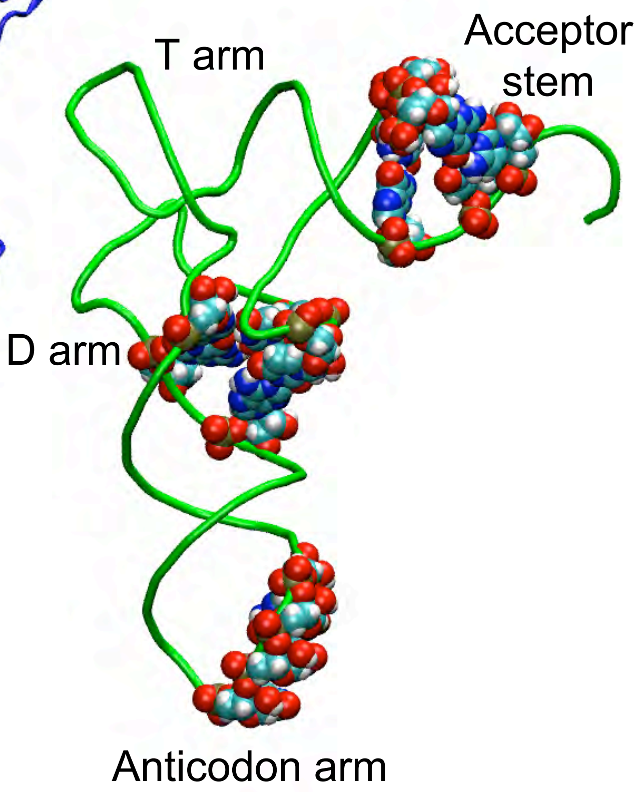
Deprotonation/protonation of charged Glu backbone amine

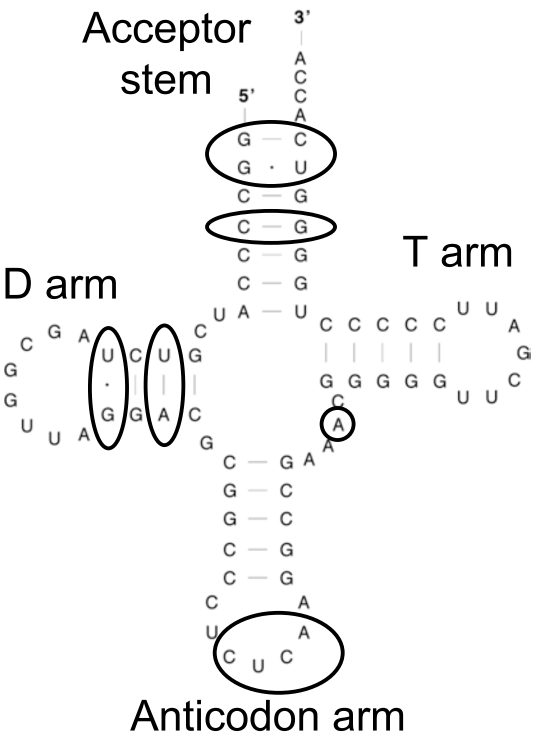
Changing rotamer states of residues (2) involved in long-lived salt bridges

(1) MacKerell, A. et al. *Biopolymers* 56:257-265 (2001)
Phillips, J.C. et al. *J. Comput Chem*, 26(16):1781-1802 (2005)
(2) Dunbrack Jr. and Cohen. *Protein Sci.* 6:1661-1681 (1997)

Charging tRNA through allosteric signaling



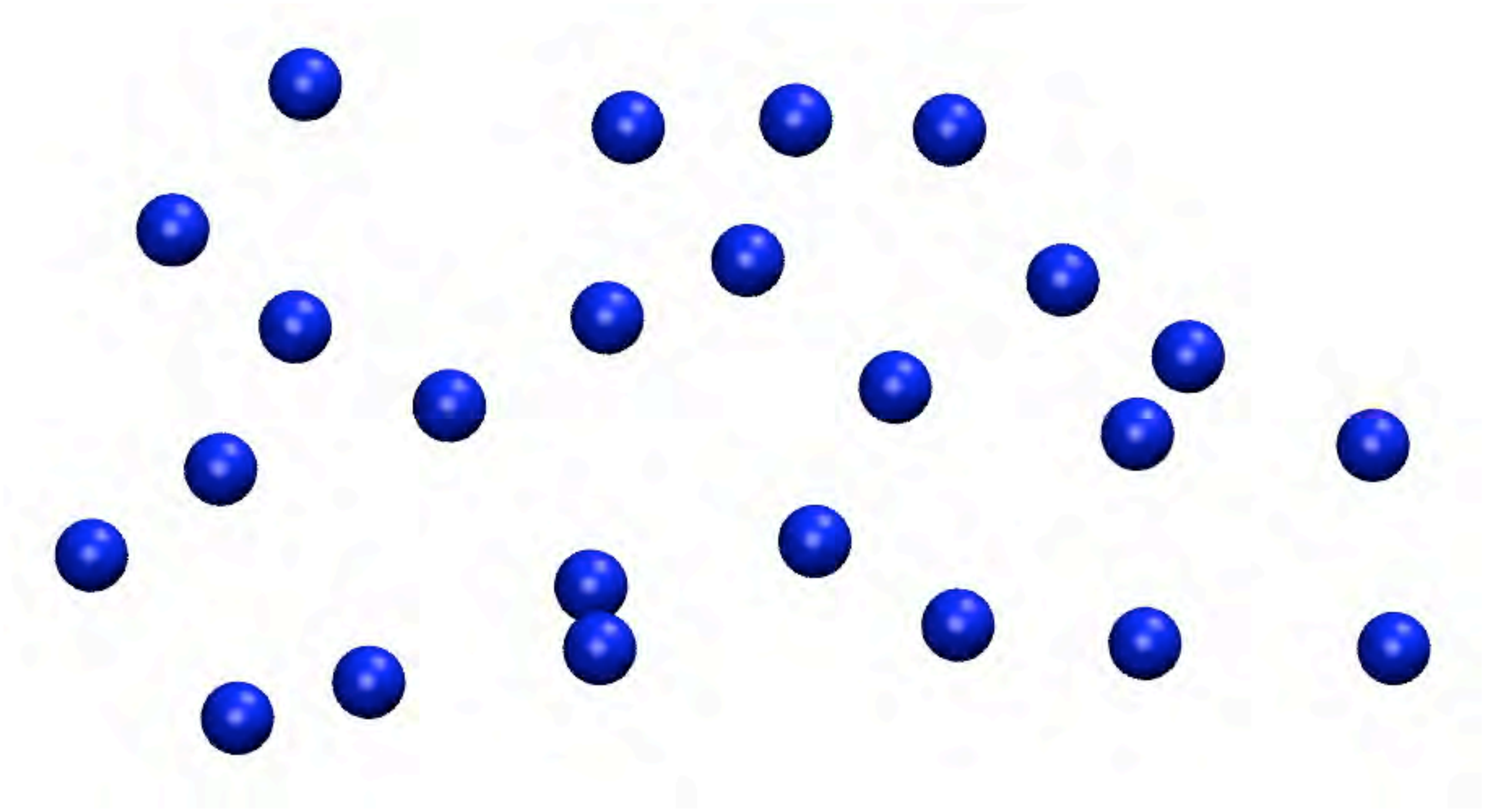




Identity Elements

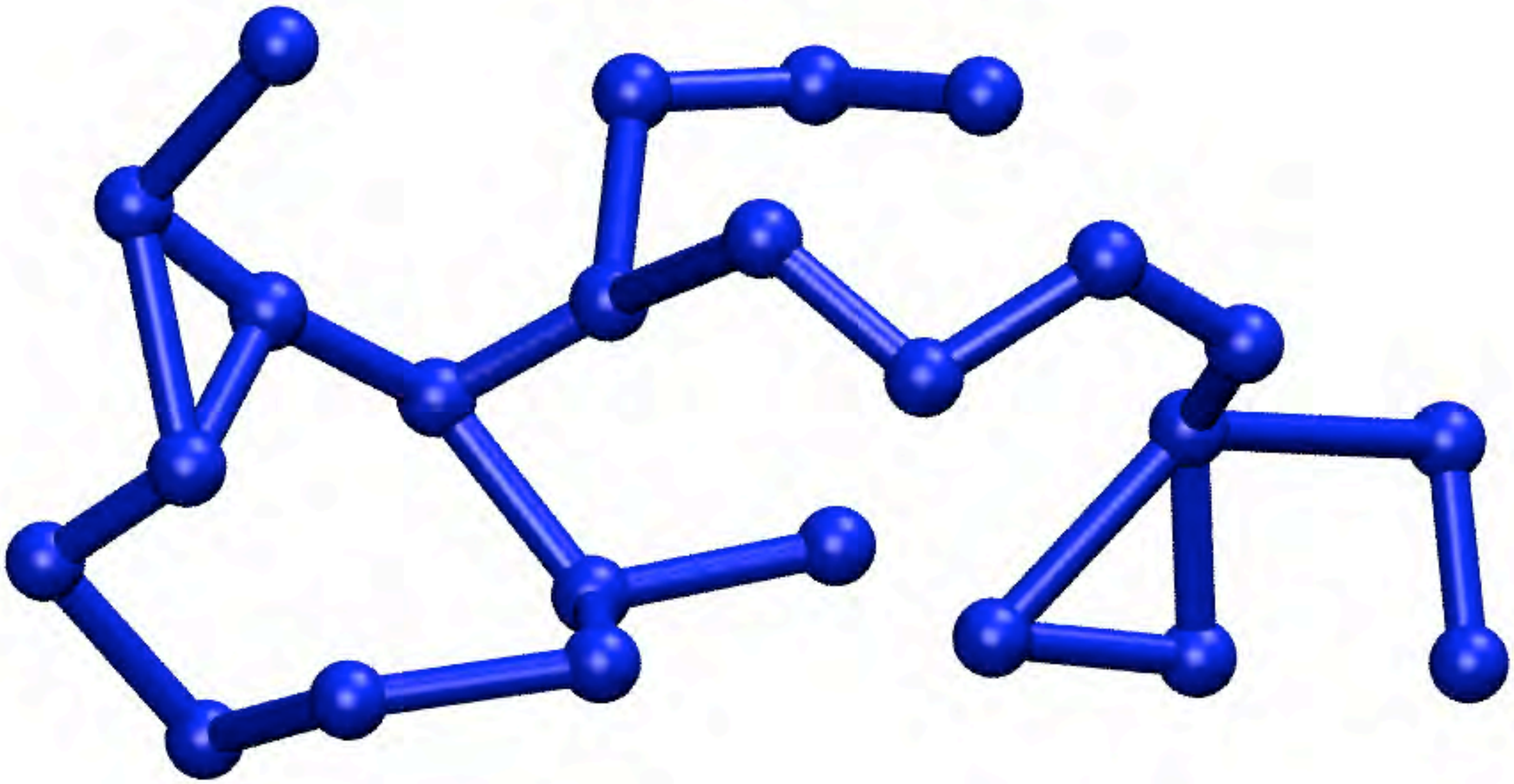
$$\text{ATP} + \text{Glu} + \text{tRNA} \xrightarrow{\text{GluRS}} \text{Glu-tRNA} + \text{AMP} + \text{PP}_i$$

How to Construct a Network



Nodes - defined at C_{α} (protein) and P (nucleotide) atoms

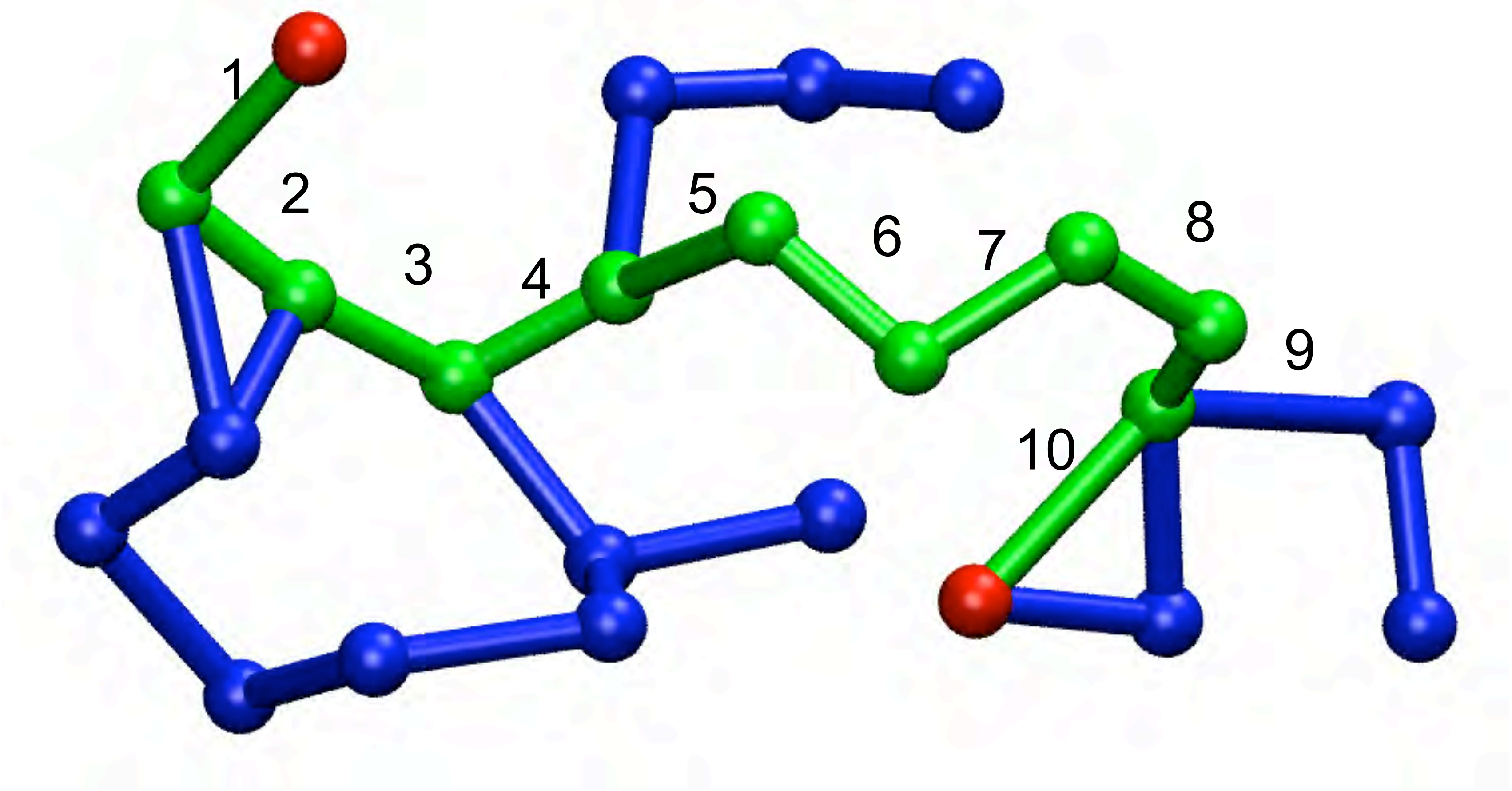
How to Construct a Network



Edges - connect nodes that are within a contact distance threshold for more than 75% of an MD trajectory

Communication between Identity Elements and Site of Chemistry

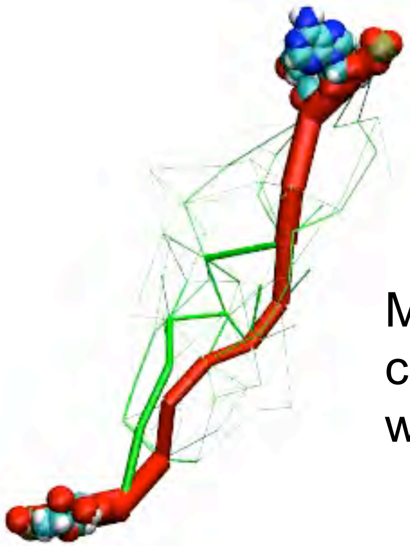
Different Paths can have Different Lengths



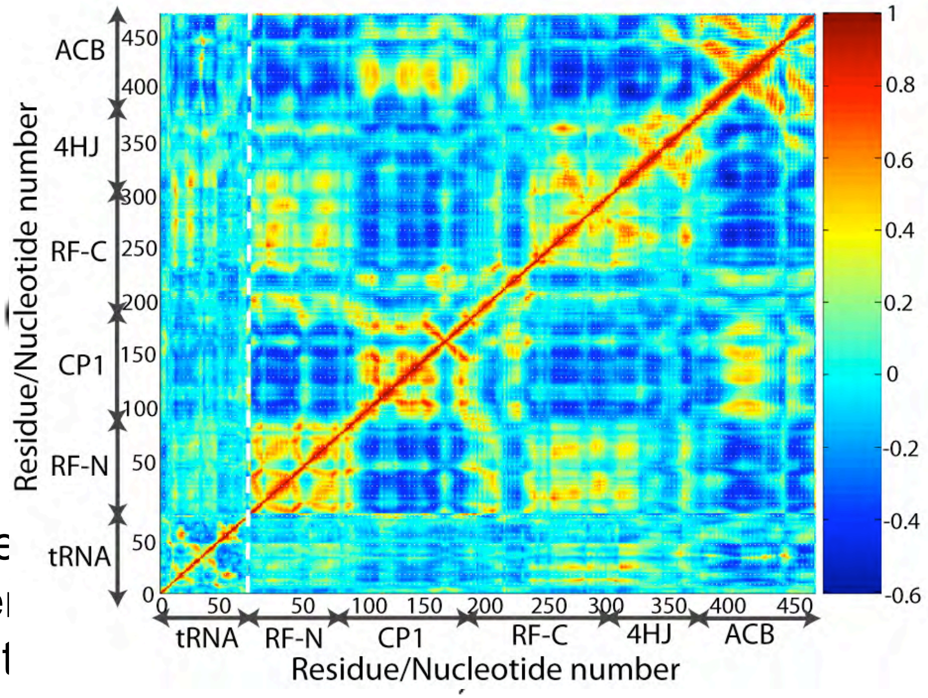
Path length = 10 (*unweighted*)

Information transfer? ***Weight contacts/links by correlations !***

Correlations (C_{ij}) define signaling pathways (shortest) in GluRS:tRNA



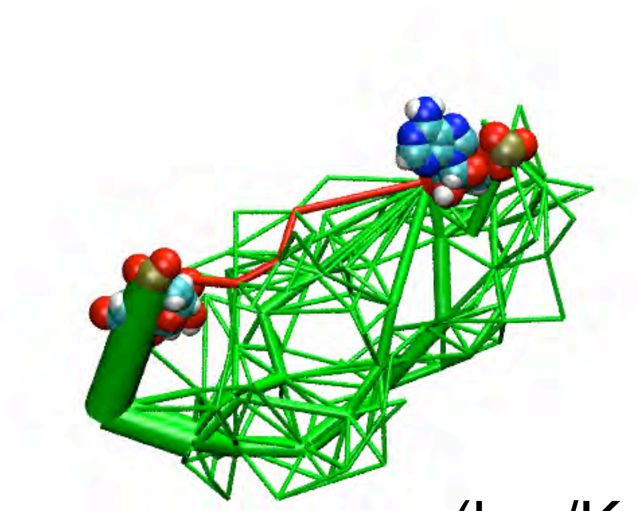
Modular structure
communities see
with weighted ne

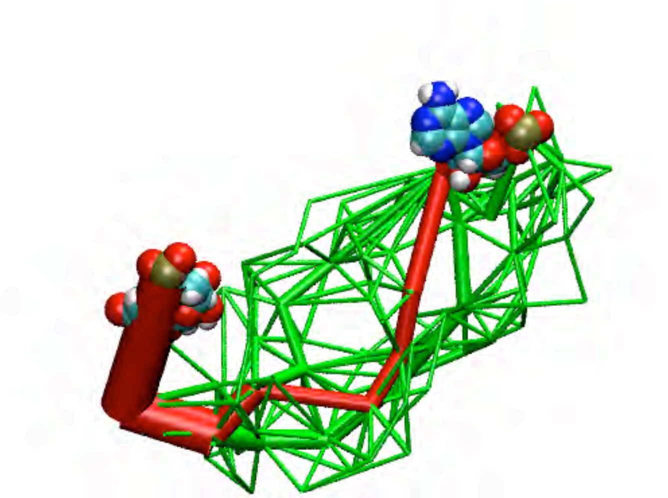


- MD simulations run with NAMD
- NTP ensemble with explicit solvent
- Neutralized with Mg^{2+} and K^+
- C_{ij} values calculated over a 16-ns window

J. Eargle, A. Sethi, A. Black, Z. Luthey-Schulten, *PNAS* 2009

Mutation Changes Communication Pathways





$$\frac{(k_{cat}/K_m)_{U13C}}{(k_{cat}/K_m)_{wt}} \approx 0.02$$ (experiment, Sekine, et al. JMB 1996)

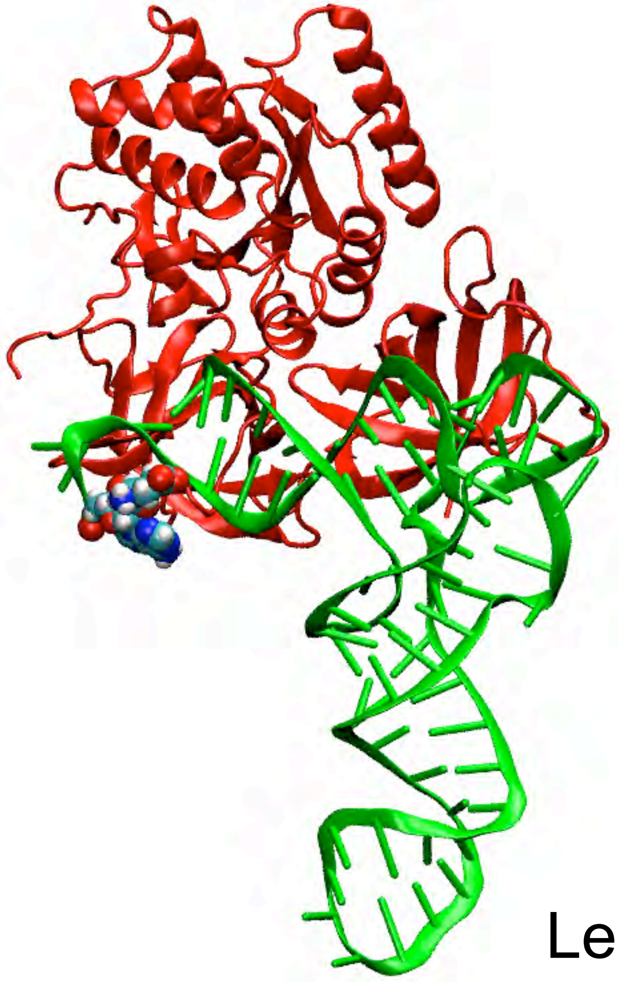
$$\frac{\exp(\text{-Sum of shortest path distances})_{U13mutant}}{\exp(\text{-Sum of shortest path distances})_{wt}} \approx 0.02$$

(Eargle, et al. PNAS 2009)

wild type

U13 mutant

Change in the Network in Glu-tRNA Migration from GluRS to EF-Tu



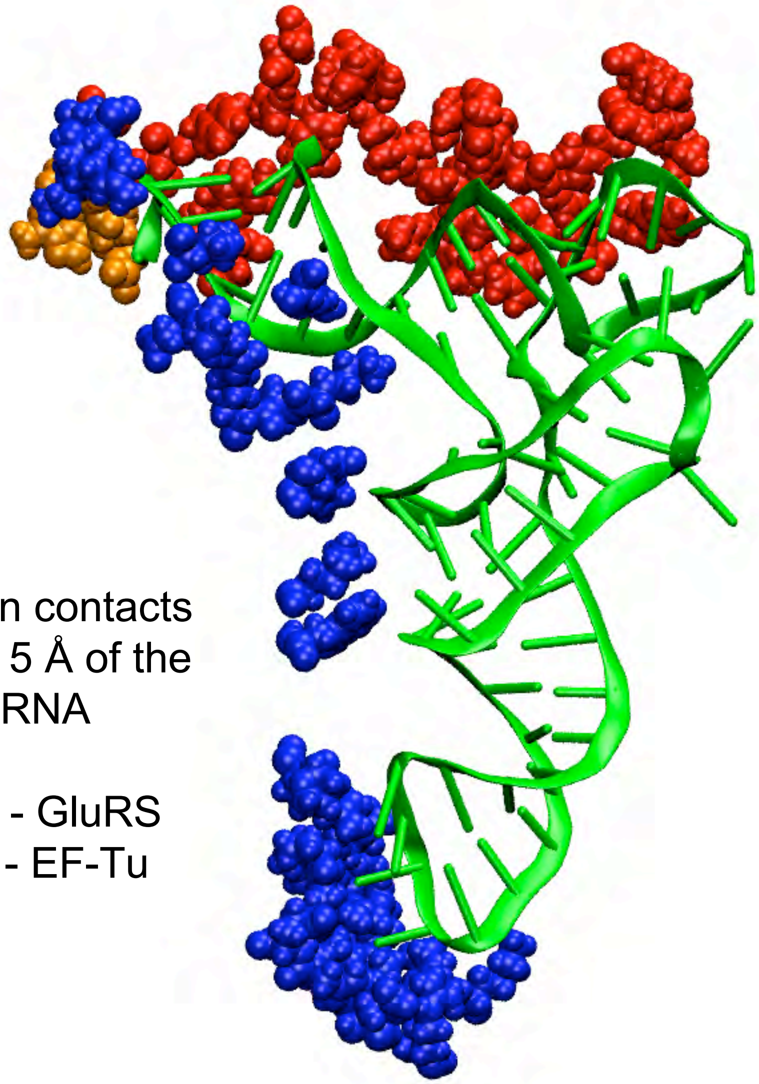
>2-fold increase
in k_{cat} for
CysRS:tRNA:EF-Tu

8-fold increase
in k_{cat} for
LeuRS:tRNA:EF-Tu

Zhang C, Perona J, Kang R,
Francklyn C, Hou Y. *JMB* 2006

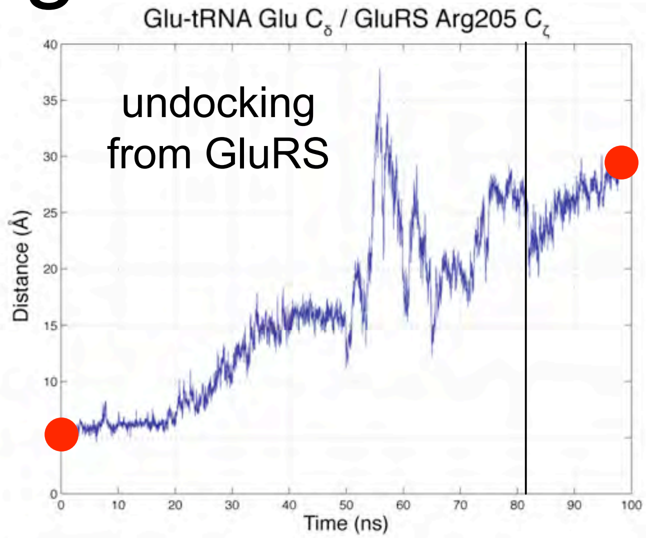
Hausmann C, Praetorius-Ibba M,
Ibba M. *NAR* 2007

Change in Protein:RNA Contacts During tRNA Migration



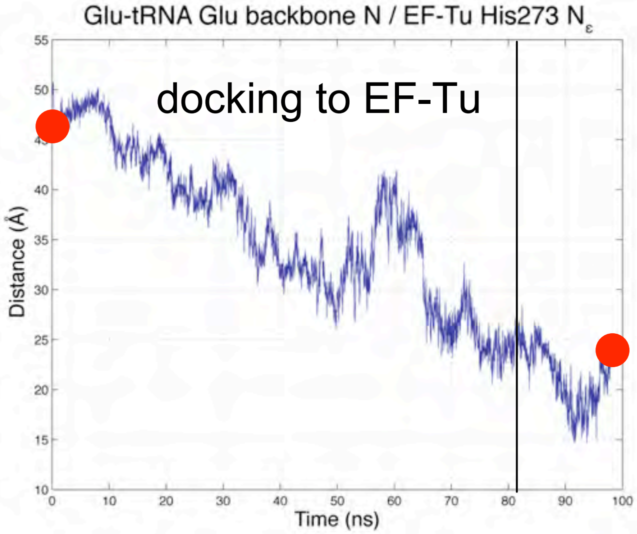
Protein contacts within 5 Å of the tRNA

blue - GluRS
red - EF-Tu



Glu-tRNA Glu C_δ / GluRS Arg205 C_ζ

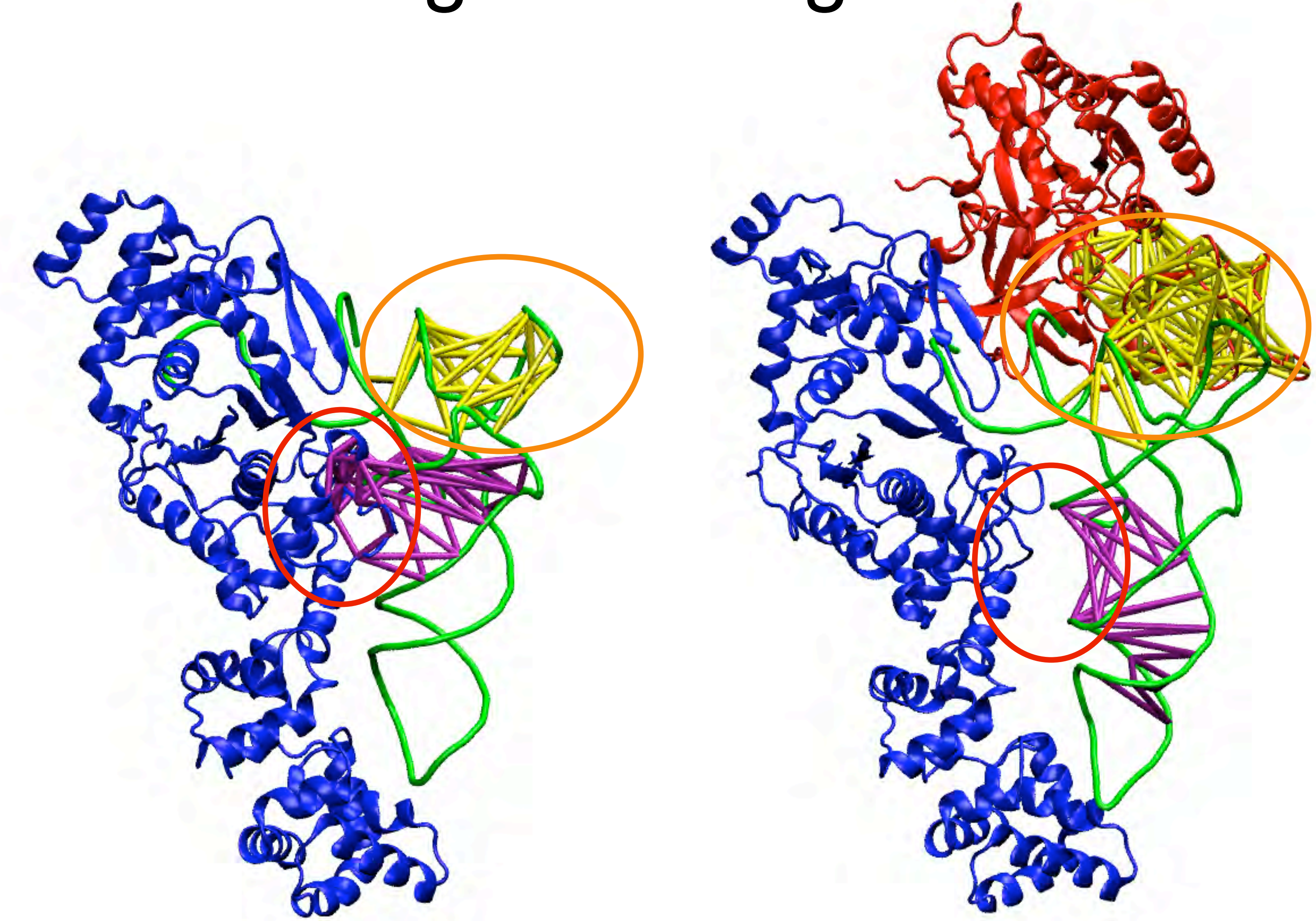
undocking from GluRS



Glu-tRNA Glu backbone N / EF-Tu His273 N_ε

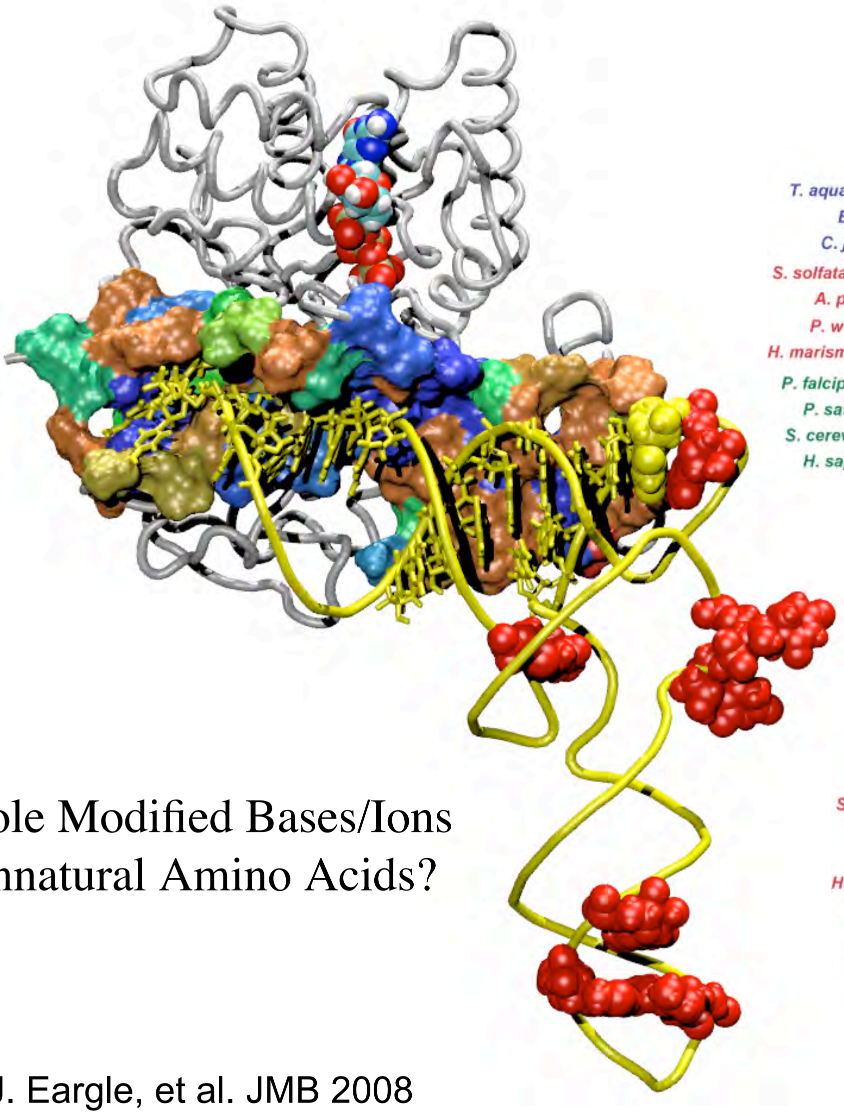
docking to EF-Tu

Change in Protein:RNA Contacts During tRNA Migration

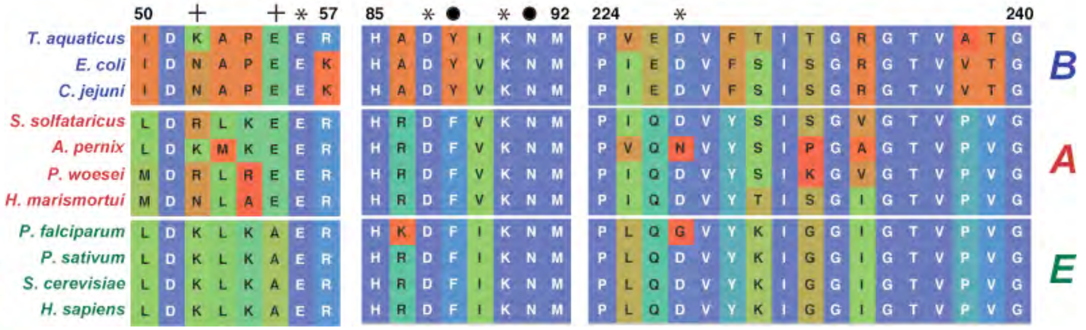


purple - D arm community   yellow - T arm community

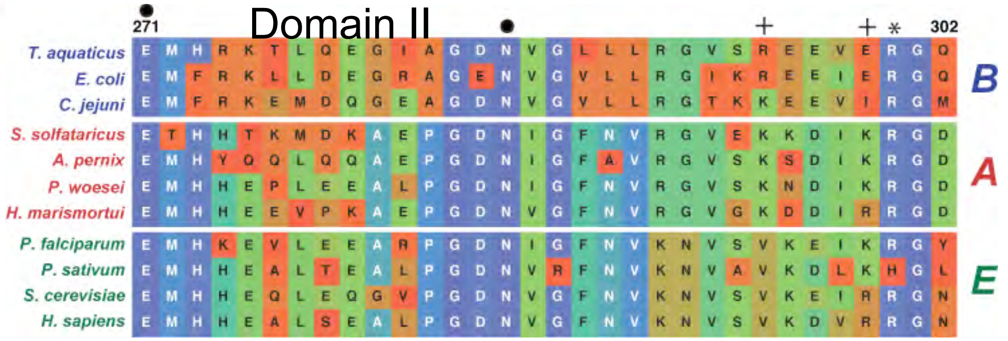
Evolution of Protein/RNA Interfaces: Dynamics of EF-Tu/tRNA Recognition



Switch I Switch II Domain II



	50		+			+	*	57	85		*	●		*	●	92	224			*													240	
T. aquaticus	I	D	K	A	P	E	E	R	H	A	D	Y	I	K	N	M	P	V	E	D	V	F	T	I	T	G	R	G	T	V	A	T	G	*B*
E. coli	I	D	N	A	P	E	E	K	H	A	D	Y	V	K	N	M	P	I	E	D	V	F	S	I	S	G	R	G	T	V	V	T	G	
C. jejuni	I	D	N	A	P	E	E	K	H	A	D	Y	V	K	N	M	P	I	E	D	V	F	S	I	S	G	R	G	T	V	V	T	G	
S. solfataricus	L	D	R	L	K	E	E	R	H	R	D	F	V	K	N	M	P	I	Q	D	V	Y	S	I	S	G	V	G	T	V	P	V	G	*A*
A. pernix	L	D	K	M	K	E	E	R	H	R	D	F	V	K	N	M	P	V	Q	N	V	Y	S	I	P	G	A	G	T	V	P	V	G	
P. woesei	M	D	R	L	R	E	E	R	H	R	D	F	V	K	N	M	P	I	Q	D	V	Y	S	I	K	G	V	G	T	V	P	V	G	
H. marismortui	M	D	N	L	A	E	E	R	H	R	D	F	V	K	N	M	P	I	Q	D	V	Y	T	I	S	G	I	G	T	V	P	V	G	
P. falciparum	L	D	K	L	K	A	E	R	H	K	D	F	I	K	N	M	P	L	Q	G	V	Y	K	I	G	G	I	G	T	V	P	V	G	*E*
P. sativum	L	D	K	L	K	A	E	R	H	R	D	F	I	K	N	M	P	L	Q	D	V	Y	K	I	G	G	I	G	T	V	P	V	G	
S. cerevisiae	L	D	K	L	K	A	E	R	H	R	D	F	I	K	N	M	P	L	Q	D	V	Y	K	I	G	G	I	G	T	V	P	V	G	
H. sapiens	L	D	K	L	K	A	E	R	H	R	D	F	I	K	N	M	P	L	Q	D	V	Y	K	I	G	G	I	G	T	V	P	V	G	



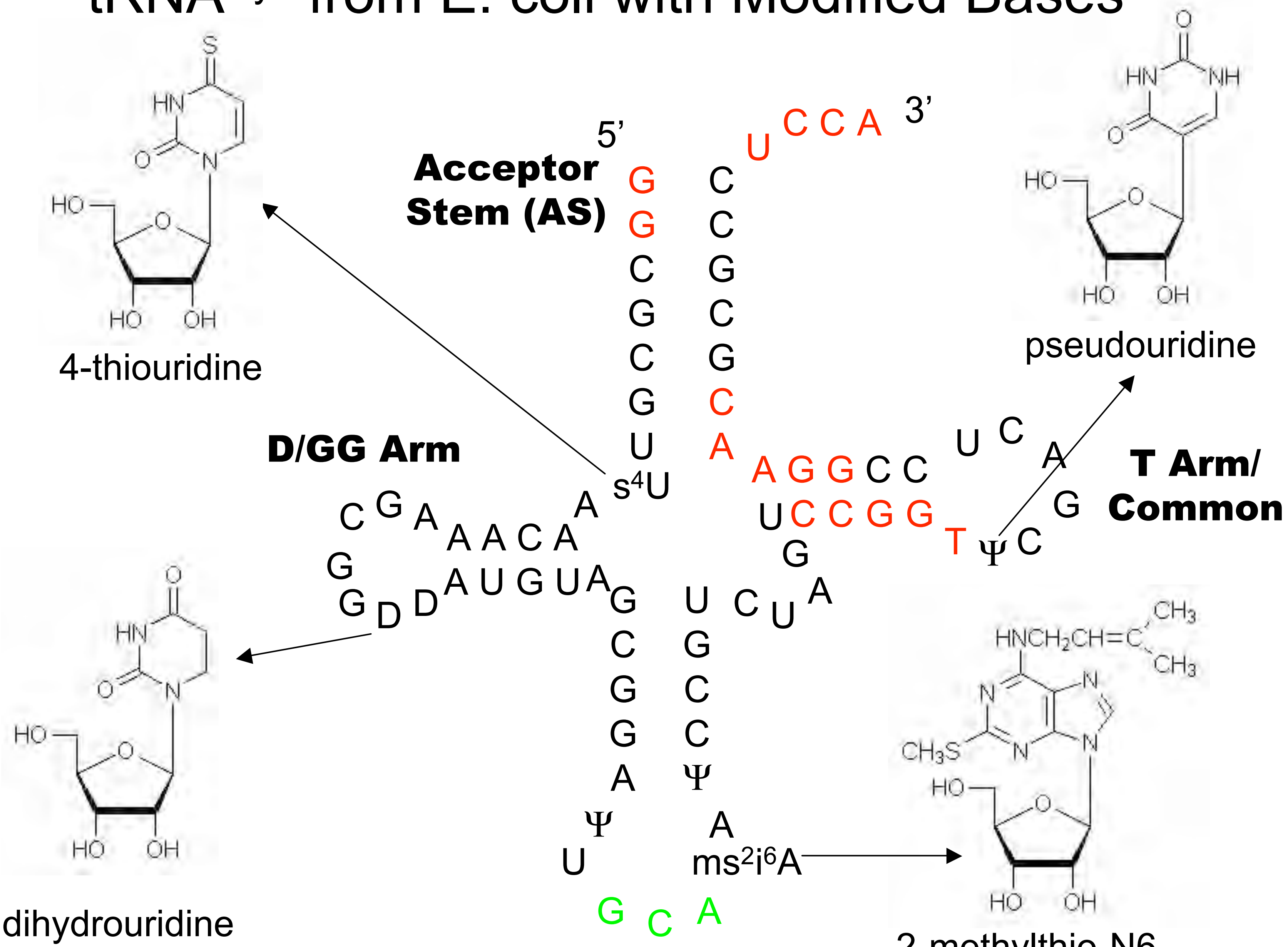
Domain II

	● 271														●										+				+	*		302	
T. aquaticus	E	M	H	R	K	T	L	Q	E	G	I	A	G	D	N	V	G	L	L	L	R	G	V	S	R	E	E	V	E	R	G	Q	*B*
E. coli	E	M	F	R	K	L	L	D	E	G	R	A	G	E	N	V	G	V	L	L	R	G	I	K	R	E	E	I	E	R	G	Q	
C. jejuni	E	M	F	R	K	E	M	D	Q	G	E	A	G	D	N	V	G	V	L	L	R	G	T	K	K	E	E	V	I	R	G	M	
S. solfataricus	E	T	H	H	T	K	M	D	K	A	E	P	G	D	N	I	G	F	N	V	R	G	V	E	K	K	D	I	K	R	G	D	*A*
A. pernix	E	M	H	Y	Q	Q	L	Q	Q	A	E	P	G	D	N	I	G	F	A	V	R	G	V	S	K	S	D	I	K	R	G	D	
P. woesei	E	M	H	H	E	P	L	E	E	A	L	P	G	D	N	I	G	F	N	V	R	G	V	S	K	N	D	I	K	R	G	D	
H. marismortui	E	M	H	H	E	E	V	P	K	A	E	P	G	D	N	V	G	F	N	V	R	G	V	G	K	D	D	I	R	R	G	D	
P. falciparum	E	M	H	K	E	V	L	E	E	A	R	P	G	D	N	I	G	F	N	V	K	N	V	S	V	K	E	I	K	R	G	Y	*E*
P. sativum	E	M	H	H	E	A	L	T	E	A	L	P	G	D	N	V	R	F	N	V	K	N	V	A	V	K	D	L	K	H	G	L	
S. cerevisiae	E	M	H	H	E	Q	L	E	Q	G	V	P	G	D	N	V	G	F	N	V	K	N	V	S	V	K	E	I	R	R	G	N	
H. sapiens	E	M	H	H	E	A	L	S	E	A	L	P	G	D	N	V	G	F	N	V	K	N	V	S	V	K	D	V	R	R	G	N	

Role Modified Bases/Ions
Unnatural Amino Acids?

J. Eargle, et al. JMB 2008

tRNA^{Cys} from E. coli with Modified Bases



Acceptor Stem (AS)

D/GG Arm

T Arm/ Common

Anticodon (AC) Arm

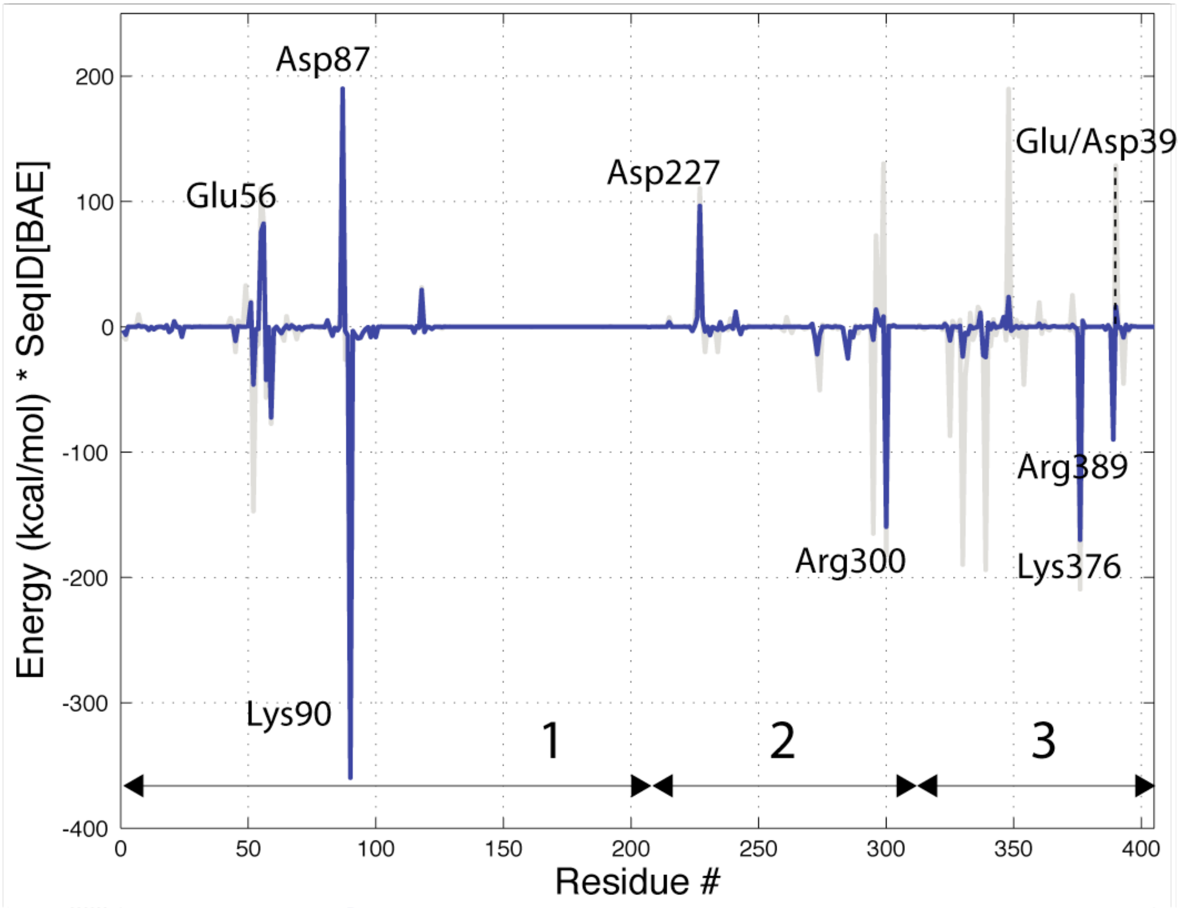
4-thiouridine

pseudouridine

dihydrouridine

2-methylthio-N6-isopentenyladenosine

Interaction Energy at Interface EF-Tu$^{T.th}$/RNA$^{\mathit{Ecoli}}$

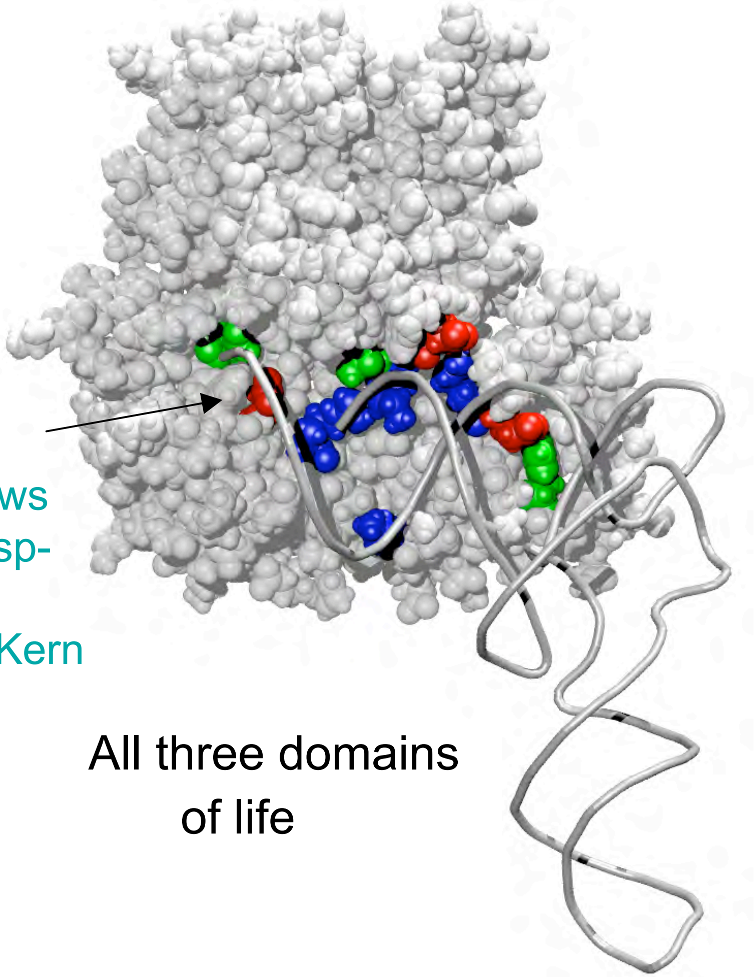


BAE

Averaging over last 16 ns/20ns MD trajectories

J. Eargle, et.al. JMB 2008

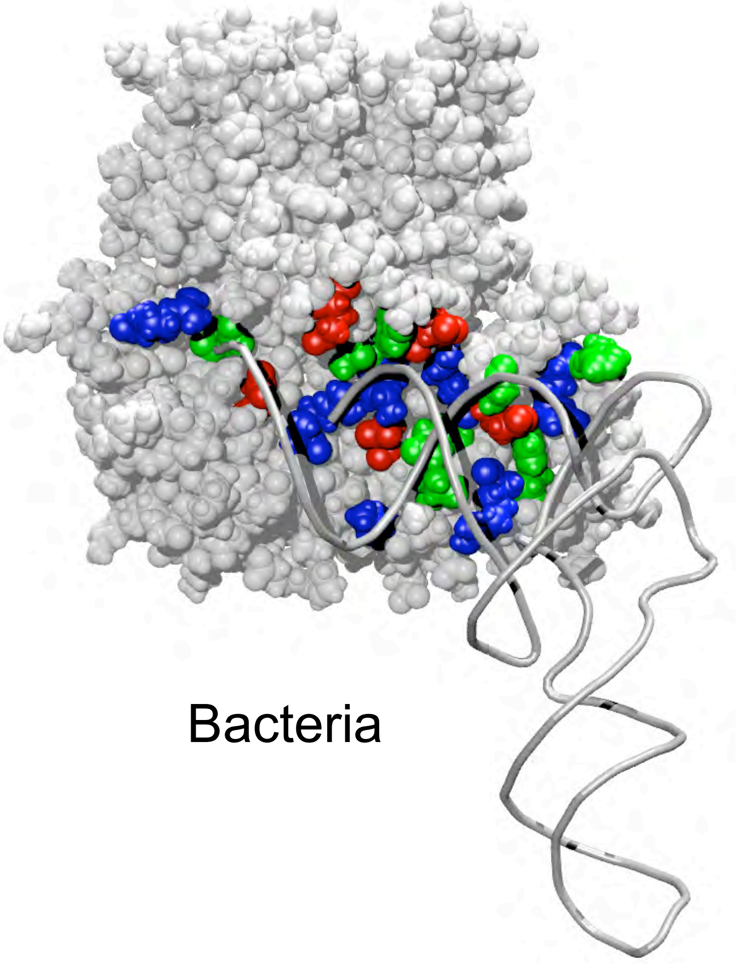
EF-Tu Interface Conservation



Asp227

Deletion allows docking of asp-tRNA:Asn Roy,Becker,Kern 2007

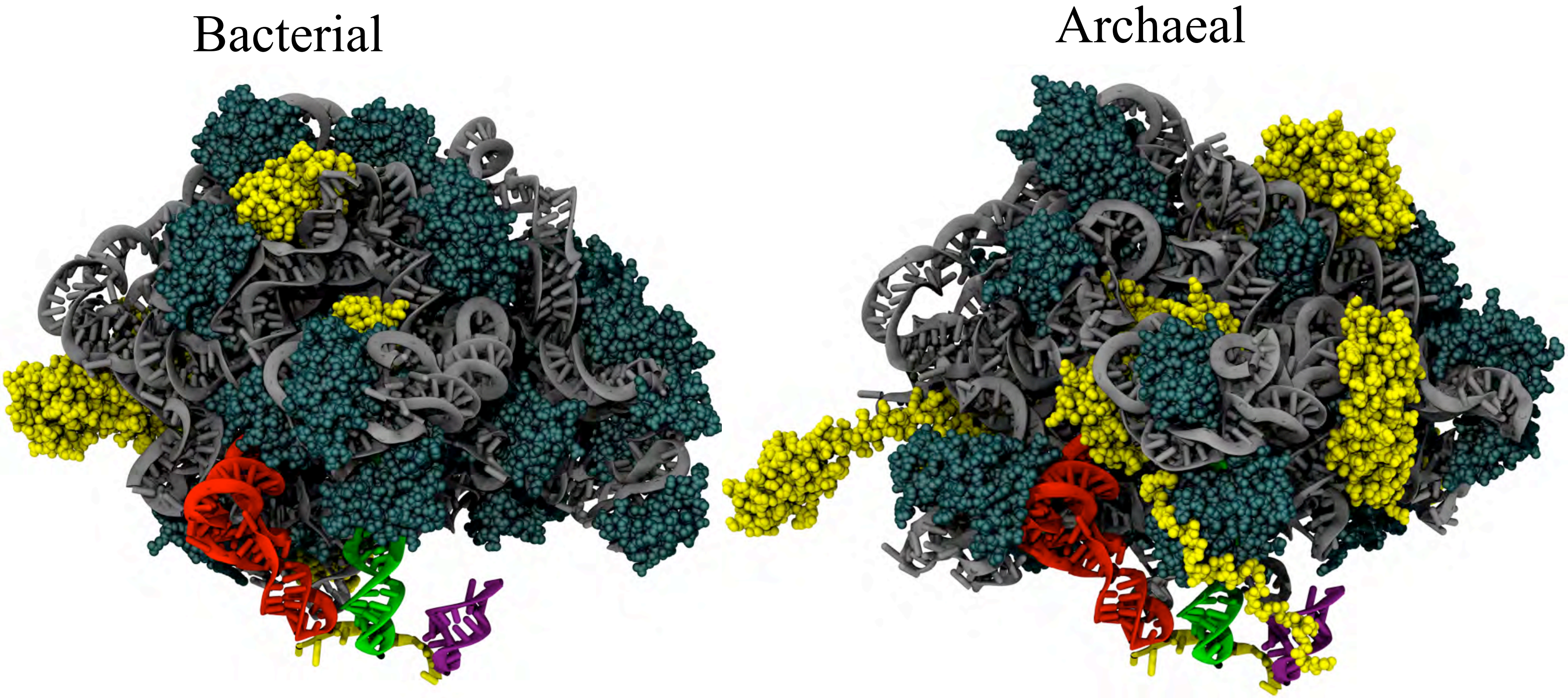
All three domains of life



Bacteria

- Residues colored by type.
- Positive: blue, Negative: red, Polar: green

Molecular Signatures in Ribosome Evolution

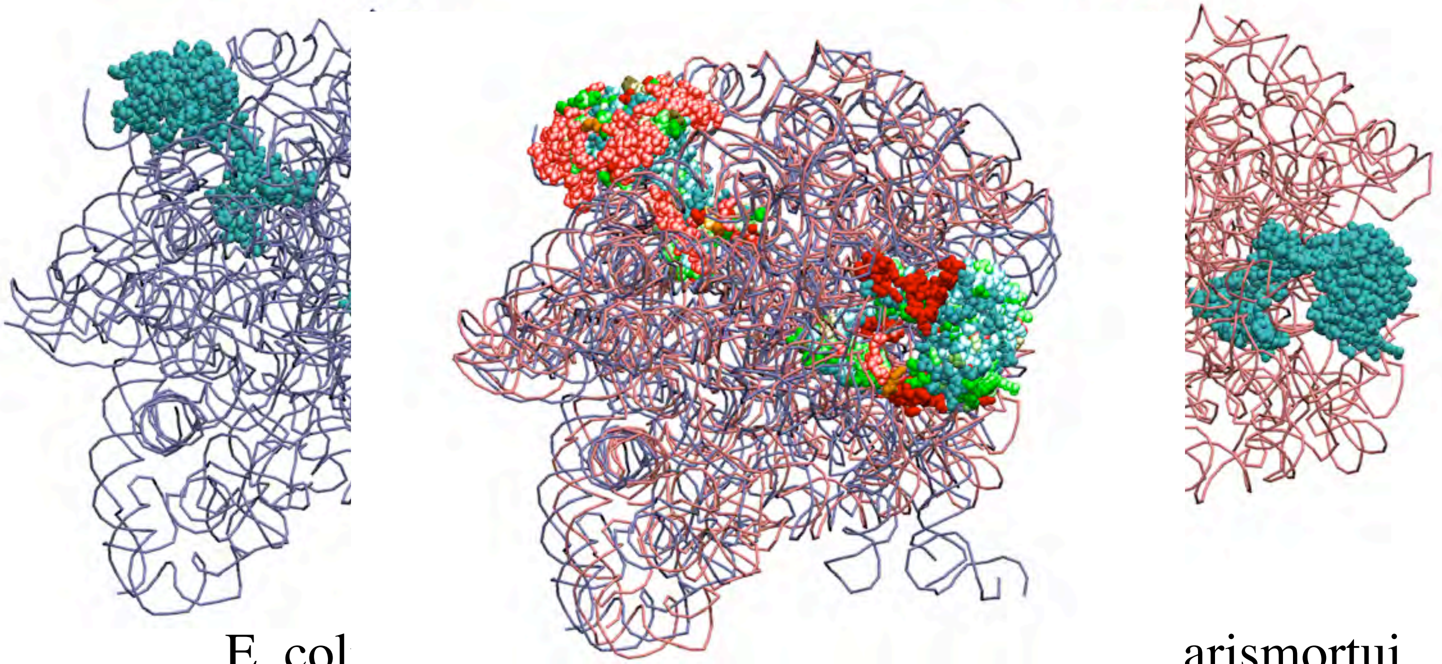


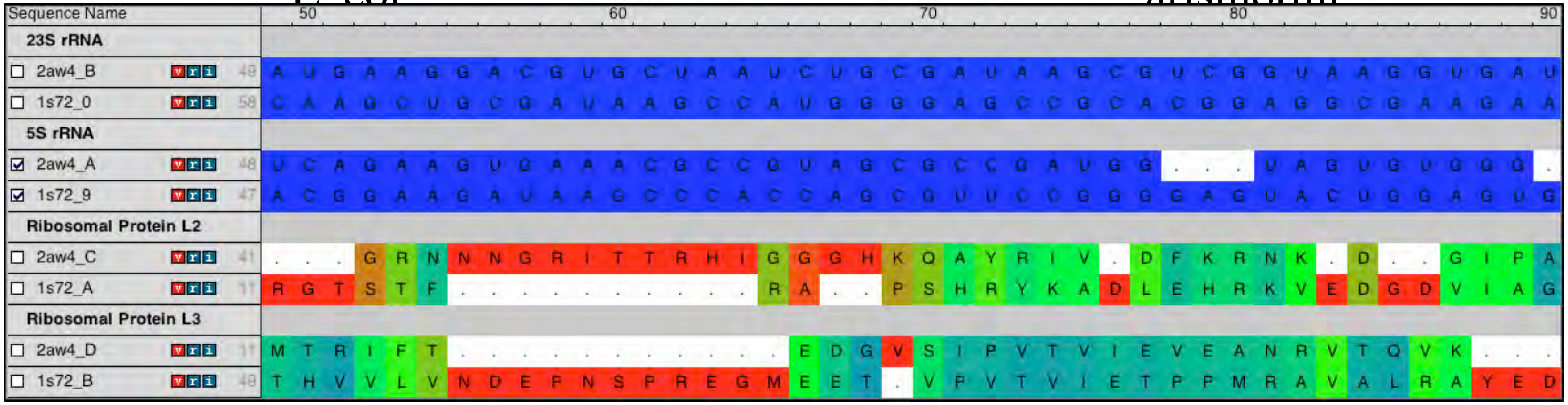
Ribosomal signatures: idiosyncrasies in ribosomal RNA and/or proteins characteristic of the domains of life

Roberts, Sethi, Montoya, Woese, Luthey-Schulten,
“Molecular Signatures of Ribosomal Evolution”,
Proc. Natl. Acad. Sci. USA, 2008, 105:13953

Comparing Ribosomal Structures

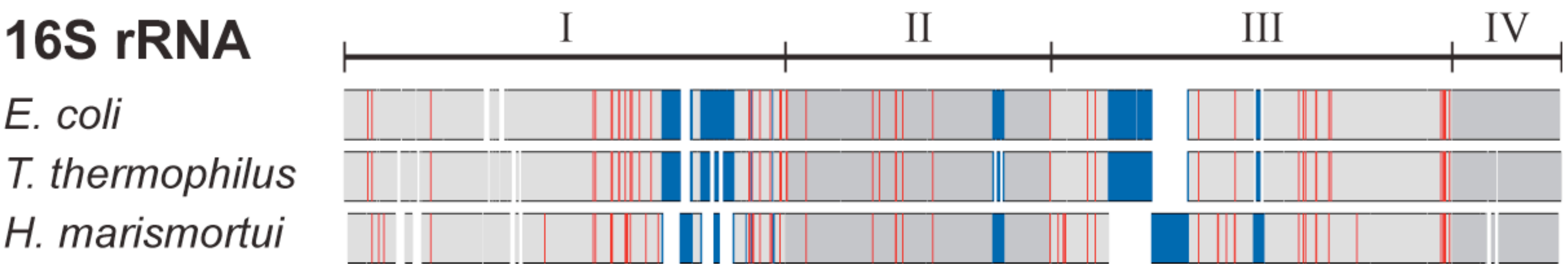
Ribosome large subunit showing ribosomal proteins L2 and L3
180,000 atoms in 4 rRNAs and 58 proteins





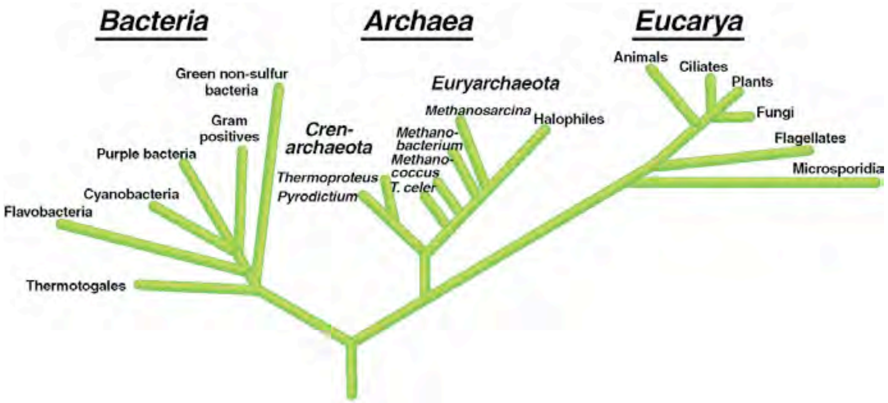
Sequence Name			50 … 60 … 70 … 80 … 90
23S rRNA			
☐ 2aw4_B	v r i	49	A U G A A G G A C G U G C U A A U C U G C G A U A A G C G U C G G U A A G G U G A U
☐ 1s72_0	v r i	58	C A A G C U G C G A U A A G C C A U G G G G A G C C G C A C G G A G G C G A A G A A
5S rRNA			
☑ 2aw4_A	v r i	48	U C A G A A G U G A A A C G C C G U A G C G C C G A U G G . . . U A G U G U G G G .
☑ 1s72_9	v r i	47	A C G G A A G A U A A G C C C A C C A G C G U U C C G G G G A G U A C U G G A G U G
Ribosomal Protein L2			
☐ 2aw4_C	v r i	41	. . . G R N N N G R I T T R H I G G G H K Q A Y R I V . D F K R N K . D . . G I P A
☐ 1s72_A	v r i	11	R G T S T F . . . . R A . . P S H R Y K A D L E H R K V E D G D V I A G
Ribosomal Protein L3			
☐ 2aw4_D	v r i	11	M T R I F T . . . . E D G V S I P V T V I E V E A N R V T Q V K . . .
☐ 1s72_B	v r i	49	T H V V L V N D E P N S P R E G M E E T . V P V T V I E T P P M R A V A L R A Y E D

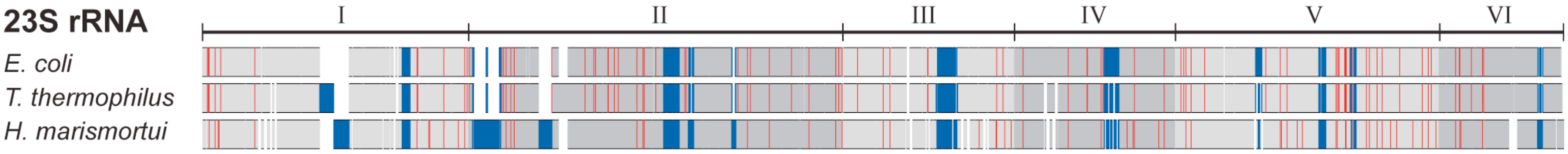
Sequence and Structure Signatures in Ribosomal RNA



Ribosomal Signatures: Idiosyncrasies in rRNA and/or rproteins characteristic of the domains of life

69 (119) & 6 (14) in 16S (23S)

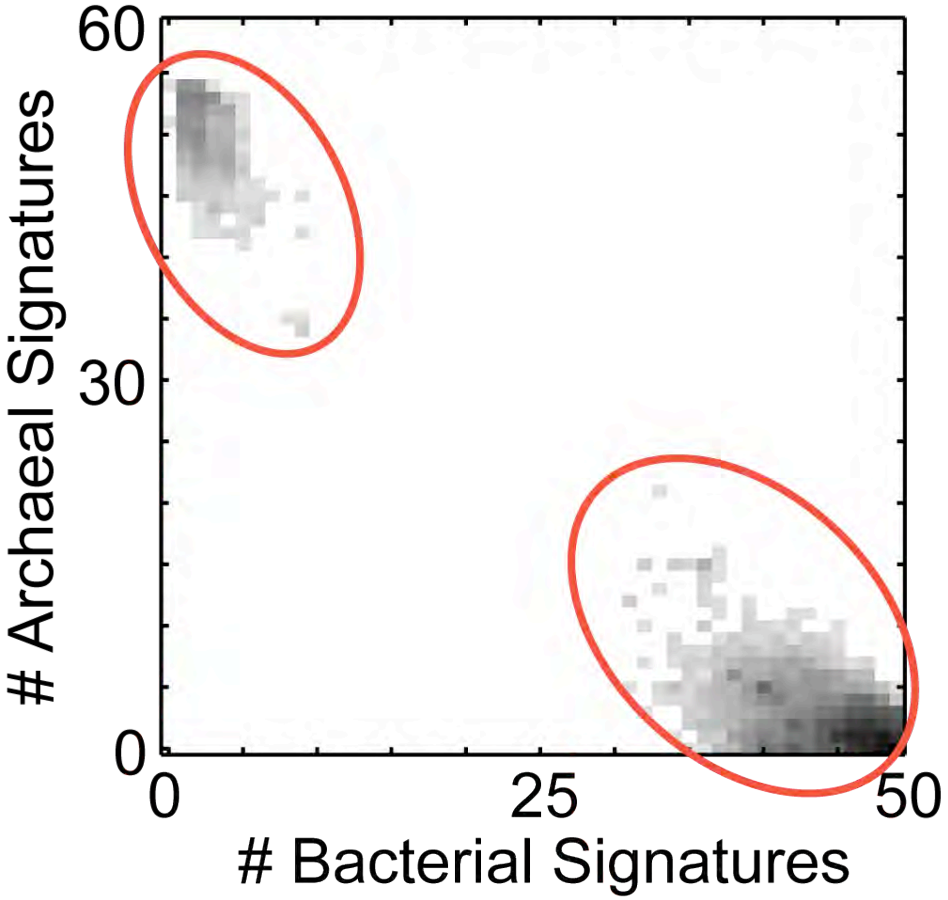




Roberts, Montoya, Sethi, Woese, ZLS, *PNAS* 2008 (in press)

90,000 Environmental 16S rRNA Sequences Confirm

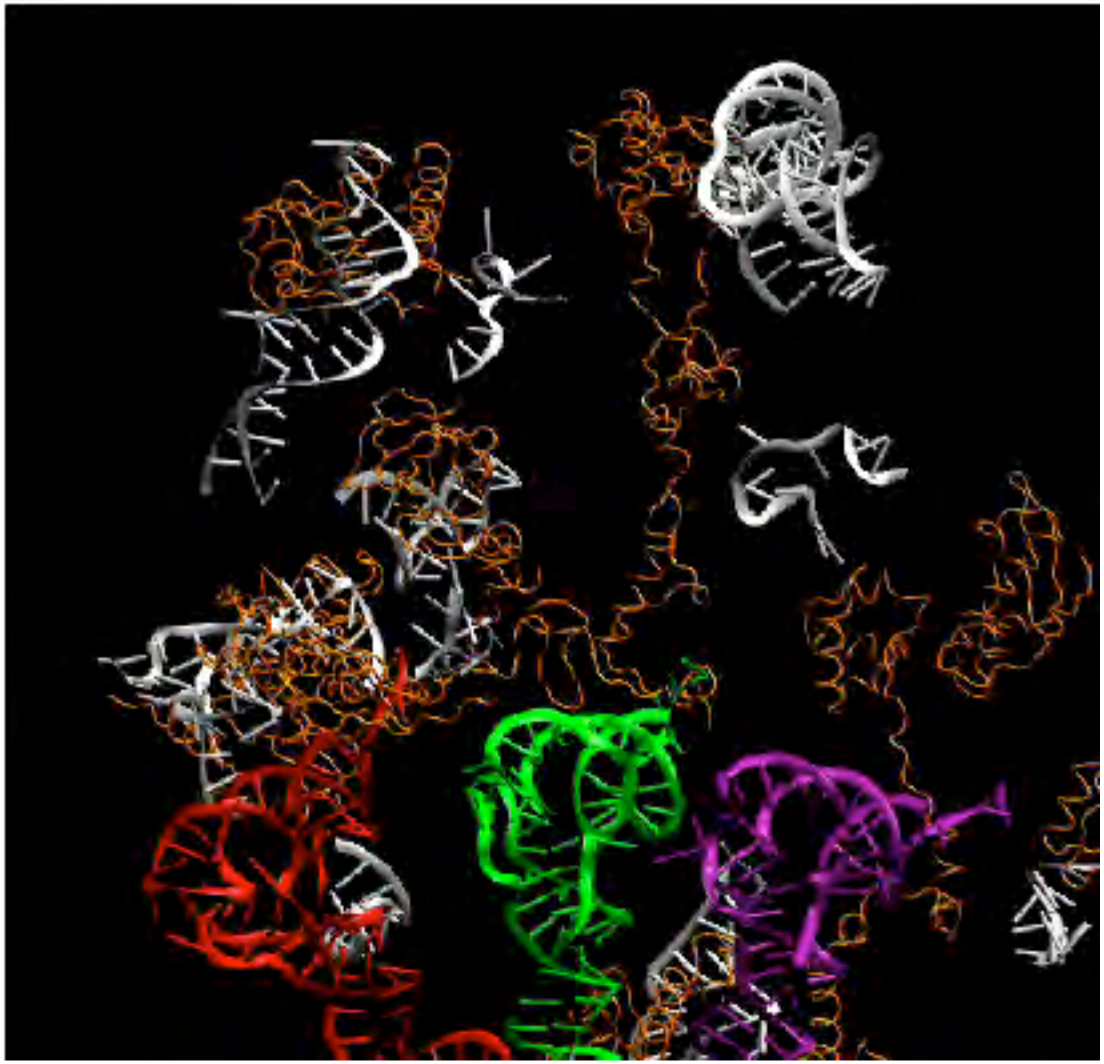
- Analysis of the ribosomal signatures in 90,000 environmental samples shows that no “gray” area exists: a ribosome is either bacterial or archaeal in nature.
- It follows then that for the domains of life there is also no continuum. The bacteria and archaea are separated by an evolutionary saltation.



Data: “Greengenes”, Lawrence Berkeley

Roberts, et al. *PNAS* 2008

Domain Specific Proteins & rRNA Structure Signatures



Archaea

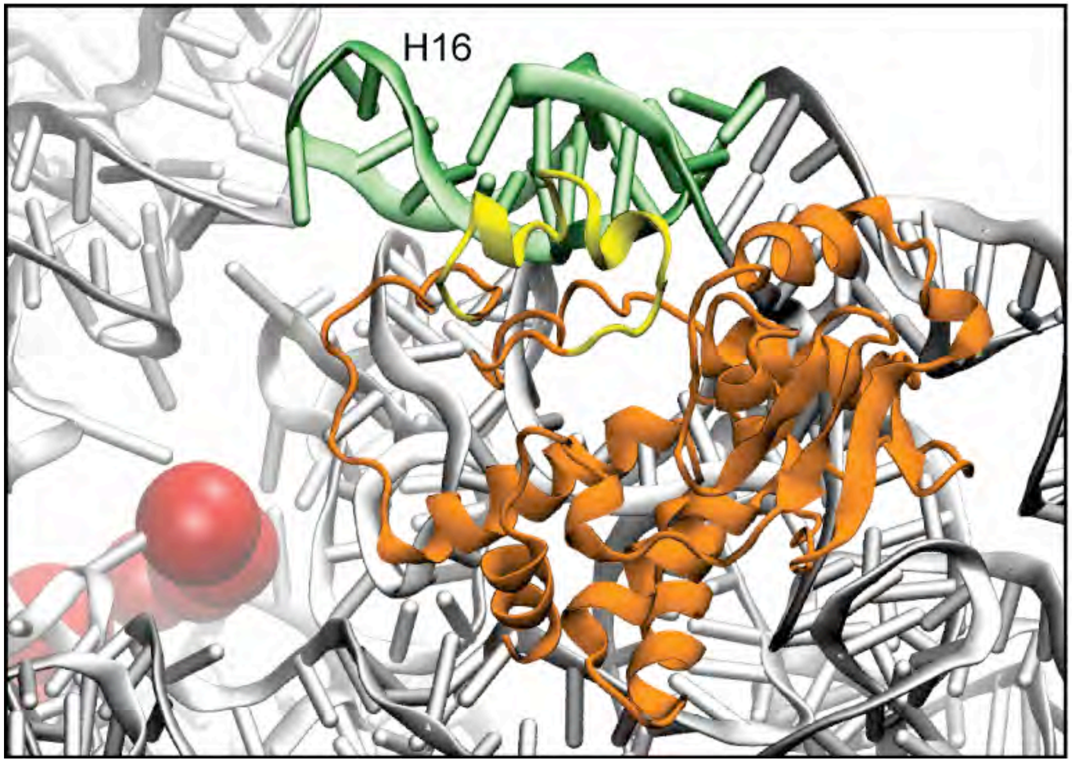
Late or Early Invention?

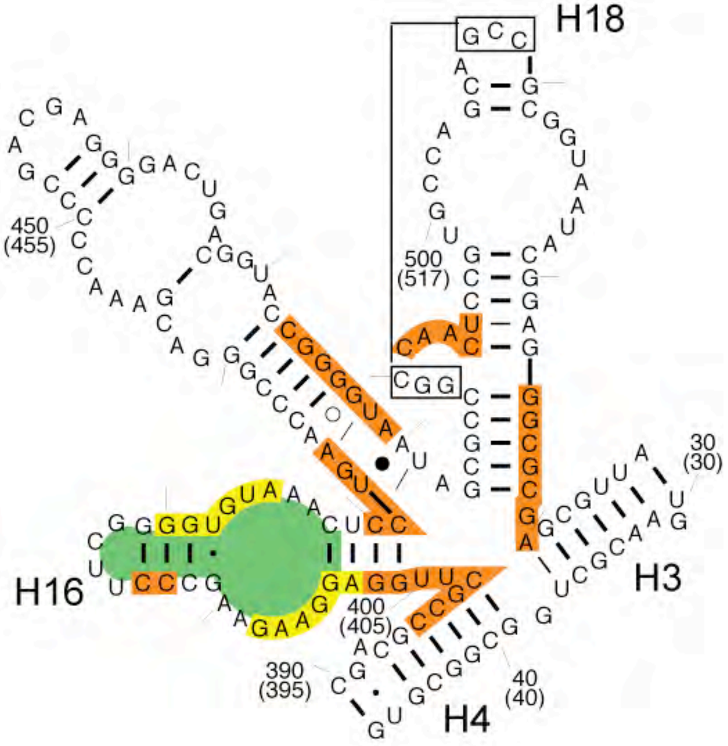
Effects on Translation?

Roberts, Montoya,Sethi,Woese ZLS, PNAS 2008

Coevolution RNA-H16 and S4 rProtein Bacterial Insert

Using Mutual Information derived from Evolutionary Profiles
*Max Entropy Cutoff for EP**

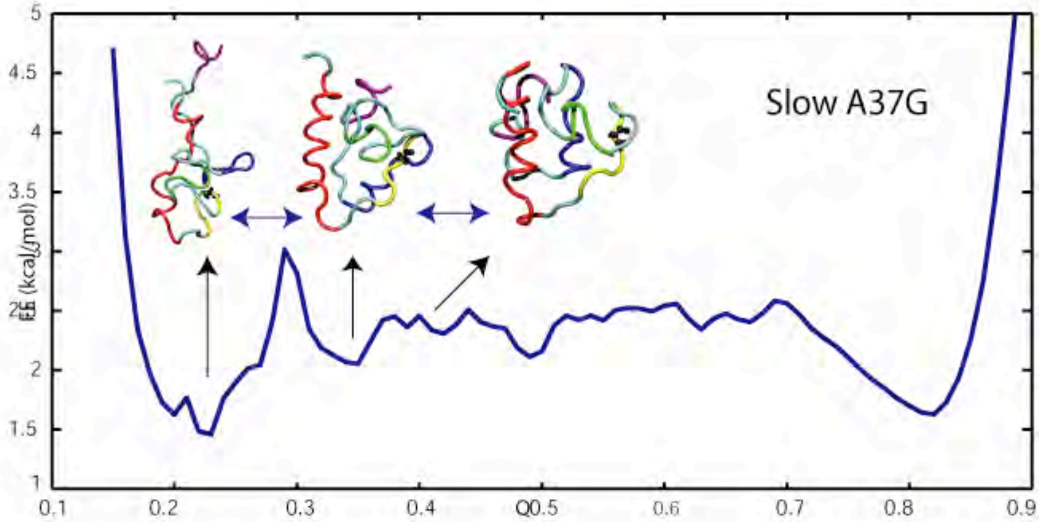




* E. Roberts, AS, JM, CW, ZLS PNAS (2009)

MI predicts signature in rRNA
Woese and Winker 1990

Protein Folding: hybrid MD Charmm + Go Potentials



Energy function:

$$E = E_{AA} + k * E_{Go} \quad E_{Go} = -\sum_{i}^{N_{C_\alpha}} \sum_{j \neq i, \pm 1, \pm 2}^{N_{C_\alpha}} \gamma_{ij} * exp\left[-\frac{\left(r_{ij} - r_{ij}^{Nat}\right)^2}{2\left(|i-j|^{0.15}\right)^2} \right]$$

Free energy profiles: umbrella sampling & WHAM

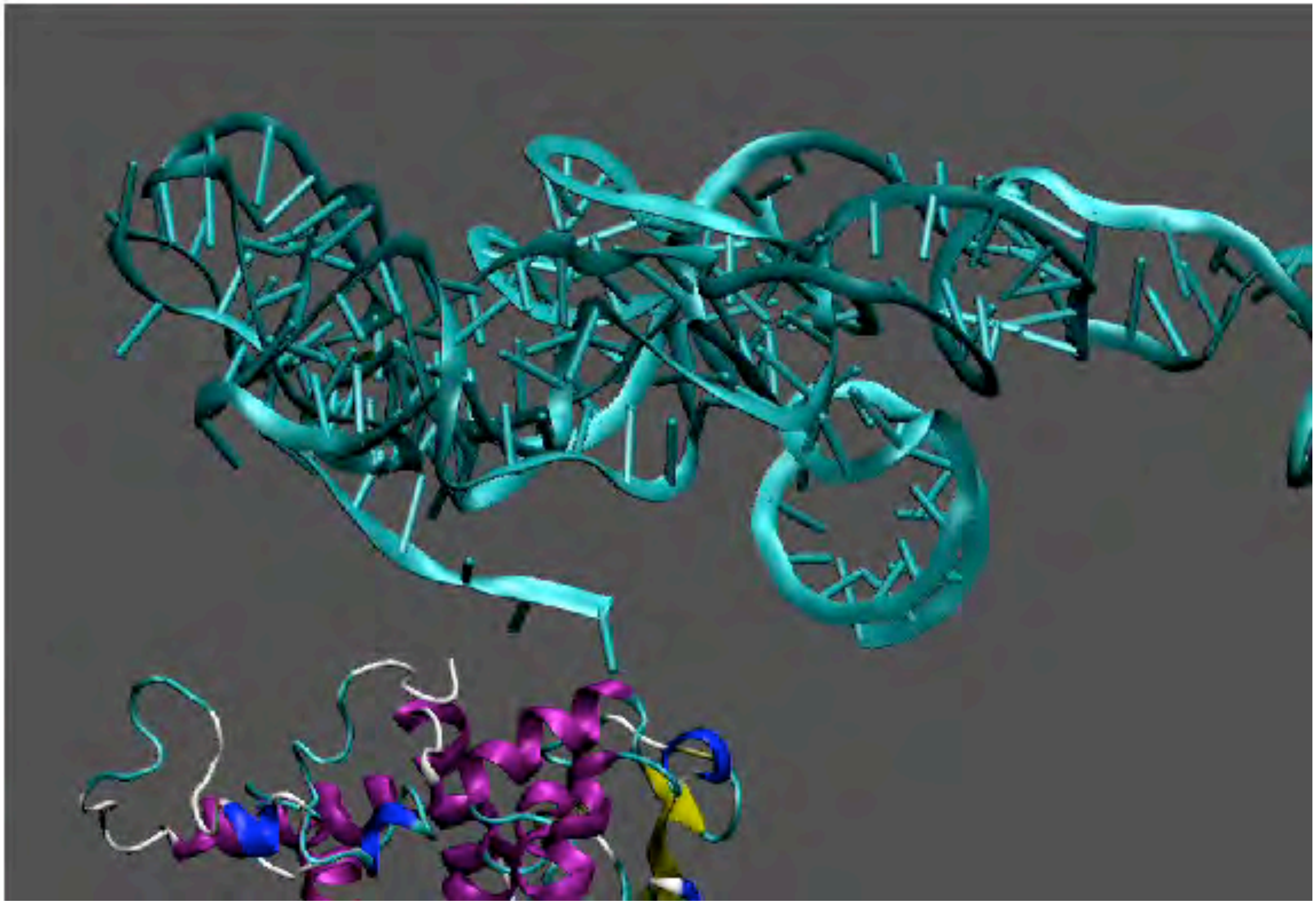
$$E_i = E_{CHARMM} + V_i(Q) \quad V_i(Q) = k_u(Q - Q_i)^4 \quad \Delta FE(Q) = -k_B T \ln p_0^{est}(Q)$$

Mean first--passage time:

$$\tau_{x_1 \to x_2} = \int_{x_1}^{x_2} dy \frac{\exp(\beta U(y))}{D(y)} \int_{\sim 0}^{y} exp(-\beta U(z)) dz$$

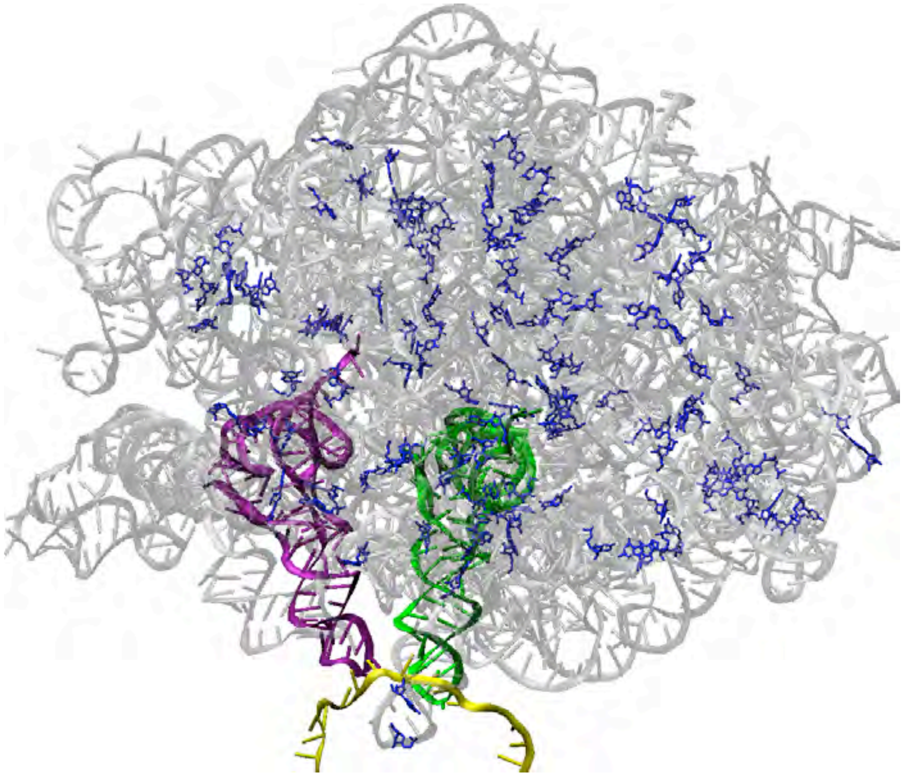
With T. Pogorelov , Biophys. J. 2004
Jonathan Montoya, 2007

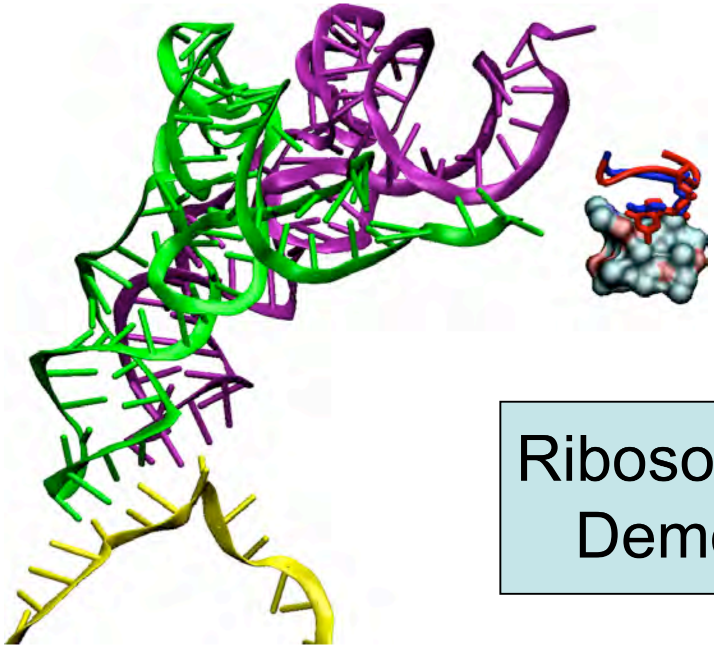
R-protein S4 Docking



23S rRNA Sequence Signatures - Effects on Translation?

- Signatures are nucleotides that are conserved in and characteristic of a domain of life.
- 5% of 2900 nucleotides differentiate Bacteria from Archaea/Eukarya
- Signature A 2058 (*E. coli*) in exit channel is target of natural antibiotics like erythromycin
- In resistant bacterial strains, A2058DMA is post-translationally modified (riboswitch)





Ribosome Demo

Evolutionary Analysis Tools in MultiSeq/VMD

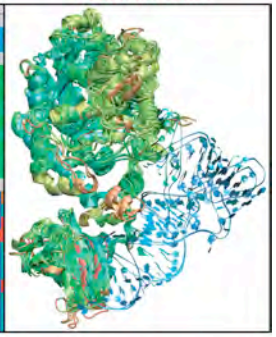
J. Eargle, D. Wright, Z. Luthey-Schulten, *Bioinformatics,* 22:504 (2006)
E. Roberts, J. Eargle, D. Wright, Z. Luthey-Schulten, *BMC Bioinformatics,* 7:382 (2006)

Protein Sequence / Structure Data

Entire SwissProt Seqs

View structural data colored by structural conservation and sequence data colored by sequence identity

Synchronization between 1D and 3D views

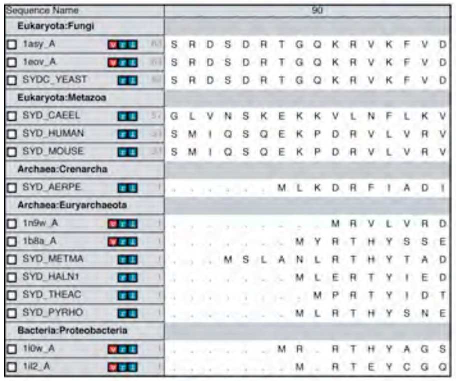


RNA Sequence / Structure Data

90,000 Envir. Seqs

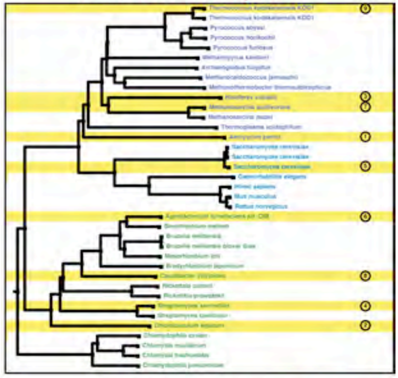
Metadata Information, Clustal & Phylogenetic Trees, Signature Analysis

Group data by taxonomic classification



```
Sequence Name                    90
Eukaryota:Fungi
☐ 1asy_A         S R D S D R T G Q K R V K F V D
☐ 1eov_A         S R D S D R T G Q K R V K F V D
☐ SYDC_YEAST     S R D S D R T G Q K R V K F V D
Eukaryota:Metazoa
☐ SYD_CAEEL      G L V N S K E K K V L N F L K V
☐ SYD_HUMAN      S M I Q S Q E K P D R V L V R V
☐ SYD_MOUSE      S M I Q S Q E K P D R V L V R V
Archaea:Crenarcha
☐ SYD_AERPE      . . . . . . M L K D R F I A D I
Archaea:Euryarchaeota
☐ 1n9w_A         . . . . . . . . . M R V L V R D
☐ 1b8a_A         . . . . . . . M Y R T H Y S S E
☐ SYD_METMA      . . . . M S L A N L R T H Y T A D
☐ SYD_HALN1      . . . . . . . M L E R T Y I E D
☐ SYD_THEAC      . . . . . . . . M P R T Y I D T
☐ SYD_PYRHO      . . . . . . . M L R T H Y S N E
Bacteria:Proteobacteria
☐ 1l0w_A         . . . . . . M R . R T H Y A G S
☐ 1il2_A         . . . . . . . M . R T E Y C G Q
```

View sequence or structure phylogenies and eliminate redundancy with QR

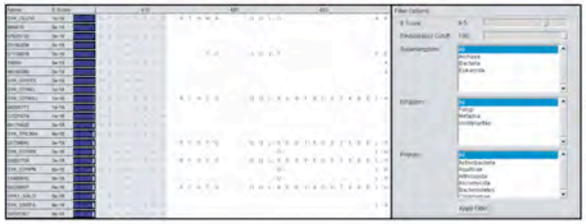


QR non-redundant seq/struct sets

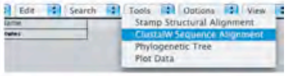
Cluster analysis / Bioinformatics scripting

Blast & PsiBlast Sequence Editor

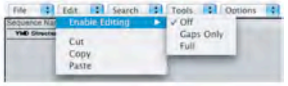
Import data directly from BLAST databases



Align sequences with Clustal



Sequence Editor: Manually adjust alignments or sequences



Tutorials
MultiSeq/AARS
EF-Tu/Ribosome

MultiSeq Tutorials