Evolution of Translation: Dynamics of Recognition in RNA:Protein Complexes
Part II - Applications of MultiSeq

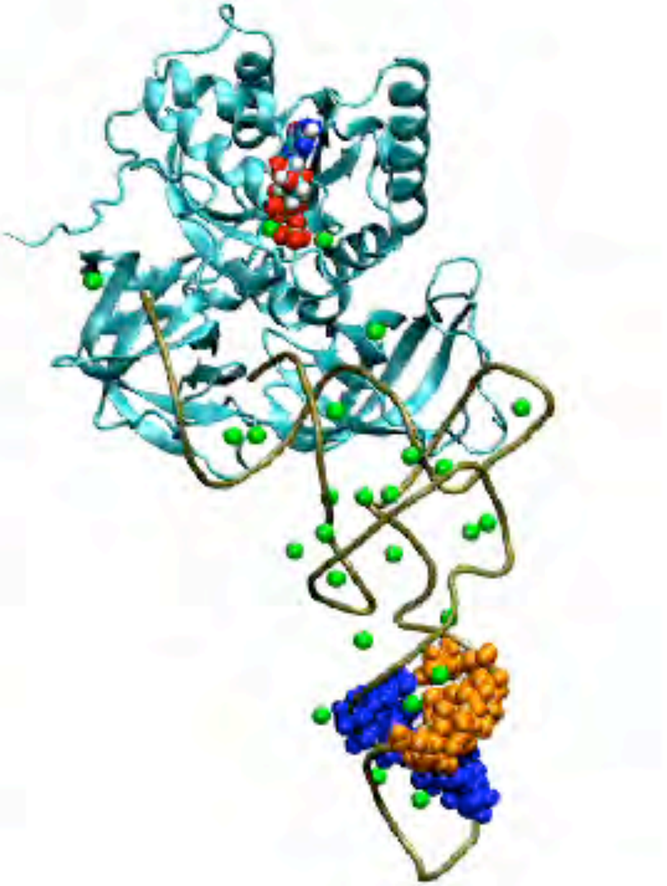
Zaida (Zan) Luthey-Schulten
Dept. Chemistry, Physics, Beckman Institute, Biophysics,
Institute of Genomics Biology

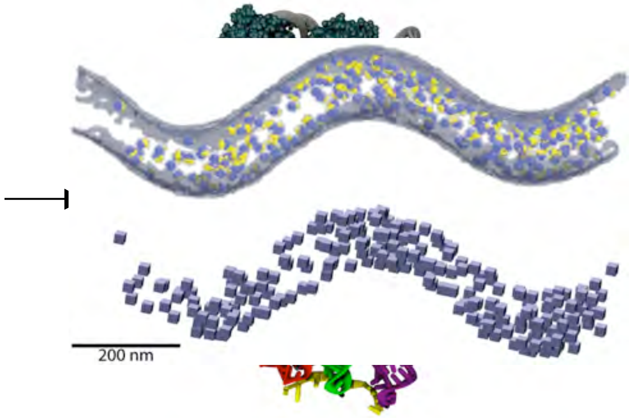
NIH Workshop 2010

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Protein:RNA Complexes in Translation

Evolution, Dynamics, Analysis





Proteins/RNA
Polyribosomes
Ribosome

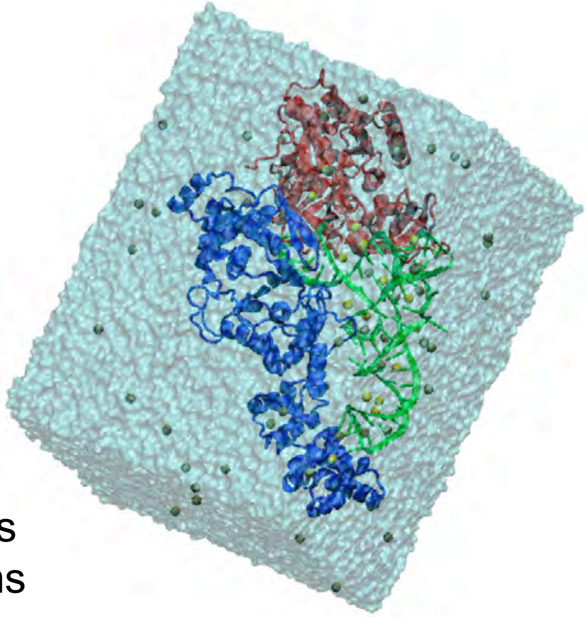
"Evolution SepRS/CysRS", ***PNAS**,2005*
"Dynamic Signaling Network", ***PNAS*** 2009
"Exit Strategy Charged tRNA" ***JMB*** 2010

"Dynamical Recognition Novel Amino Acids", ***JMB*** 2008

"Signatures ribosomal evolution"
***PNAS** 2008,* ***BMC** 2009*
"Whole cell simulations on GPUs" ***IEEE** 2009*
"Dynamics of tRNA" ***FEBS*** 2010

Evolutionary Analysis & Molecular Dynamics

MD performed with NAMD2 (1) - System Setup



Simulation Parameters

Minimization: 290,000 steps
Production run: 108 ns
Forcefields: CHARMM27 (2),
AMBER (3)
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(4)

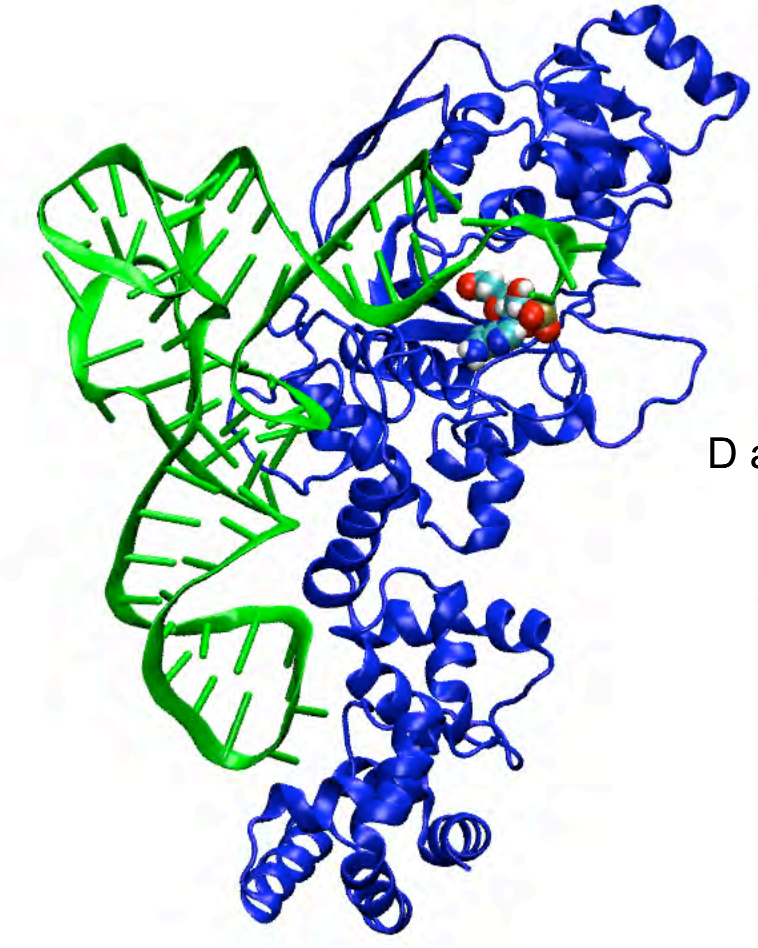
GluRS
Glu-tRNA^{Glu}
EF-Tu
GTP
Ions: Mg^{2+}, K^+
H_2O: ~27,000 molecules
System: ~130,000 atoms

System Perturbations

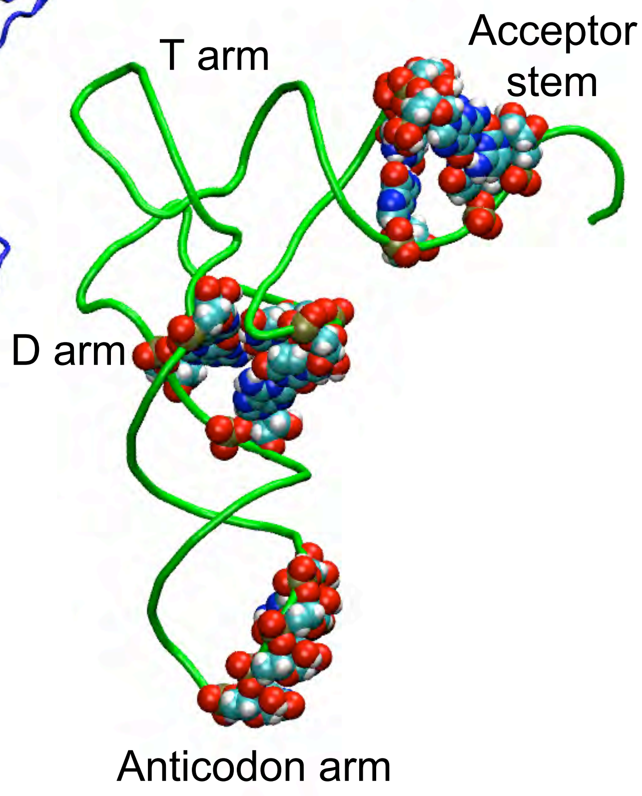
Deprotonation/protonation of reactants (aa,NTs)
Changing rotamer states of residues (5)
involved in long-lived salt bridges

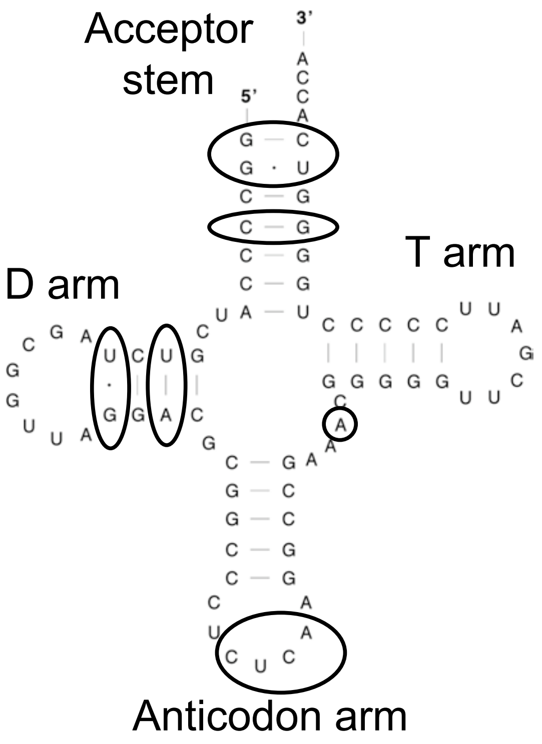
(1) Phillips, J.C. et al. *J. Comput Chem*, (2005); (2) MacKerell, A. et al. *Biopolymers* (2001);
(3) Case, D. et al. *J. Comput. Chem.* (2005); (4) Eargle, J. et al. JMB 2010, FEBS Let. 2010;
(5) Dunbrack Jr. and Cohen. *Protein Sci.* (1997)

Charging tRNA through allosteric signaling



Identity Elements

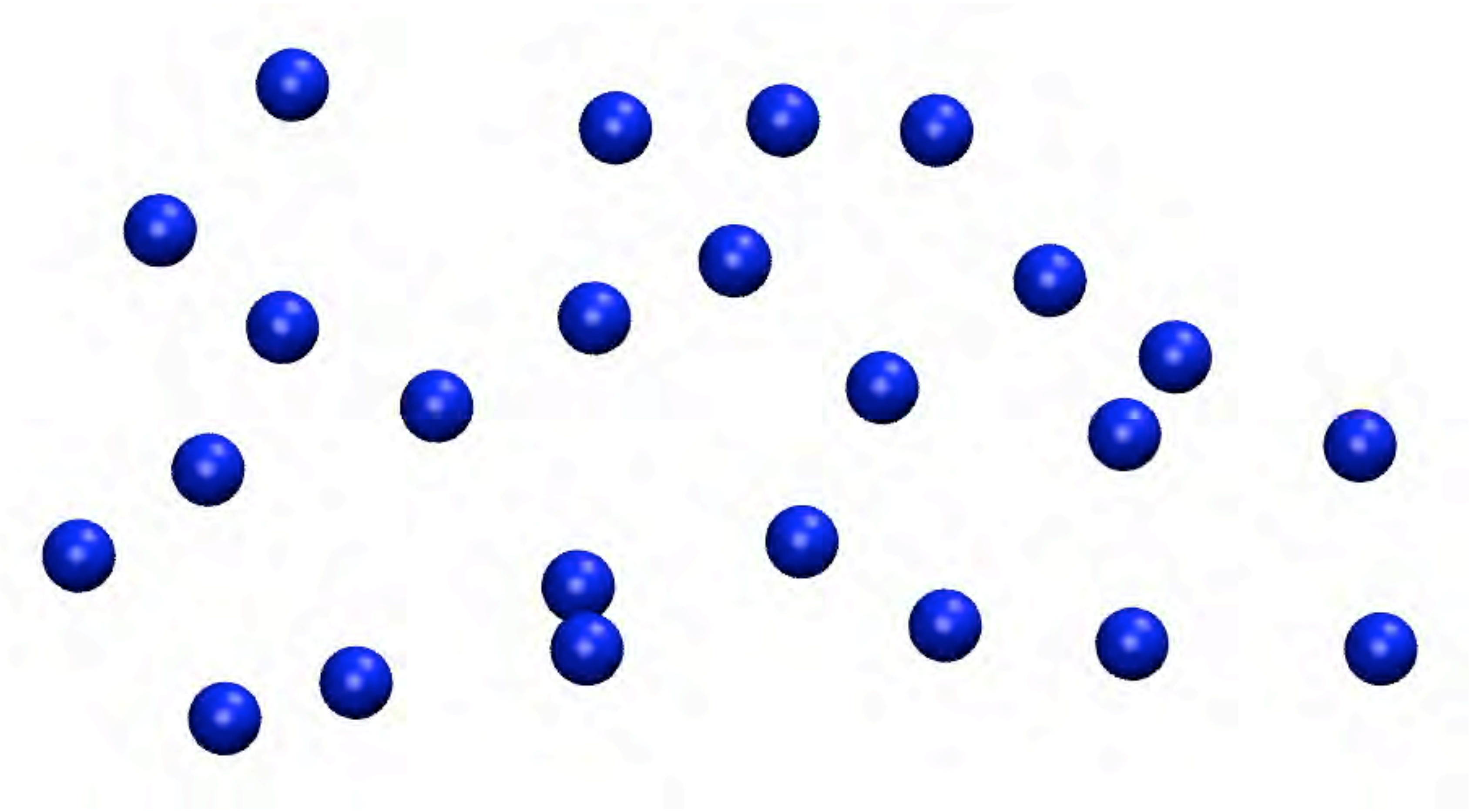




Overall Charging Reaction

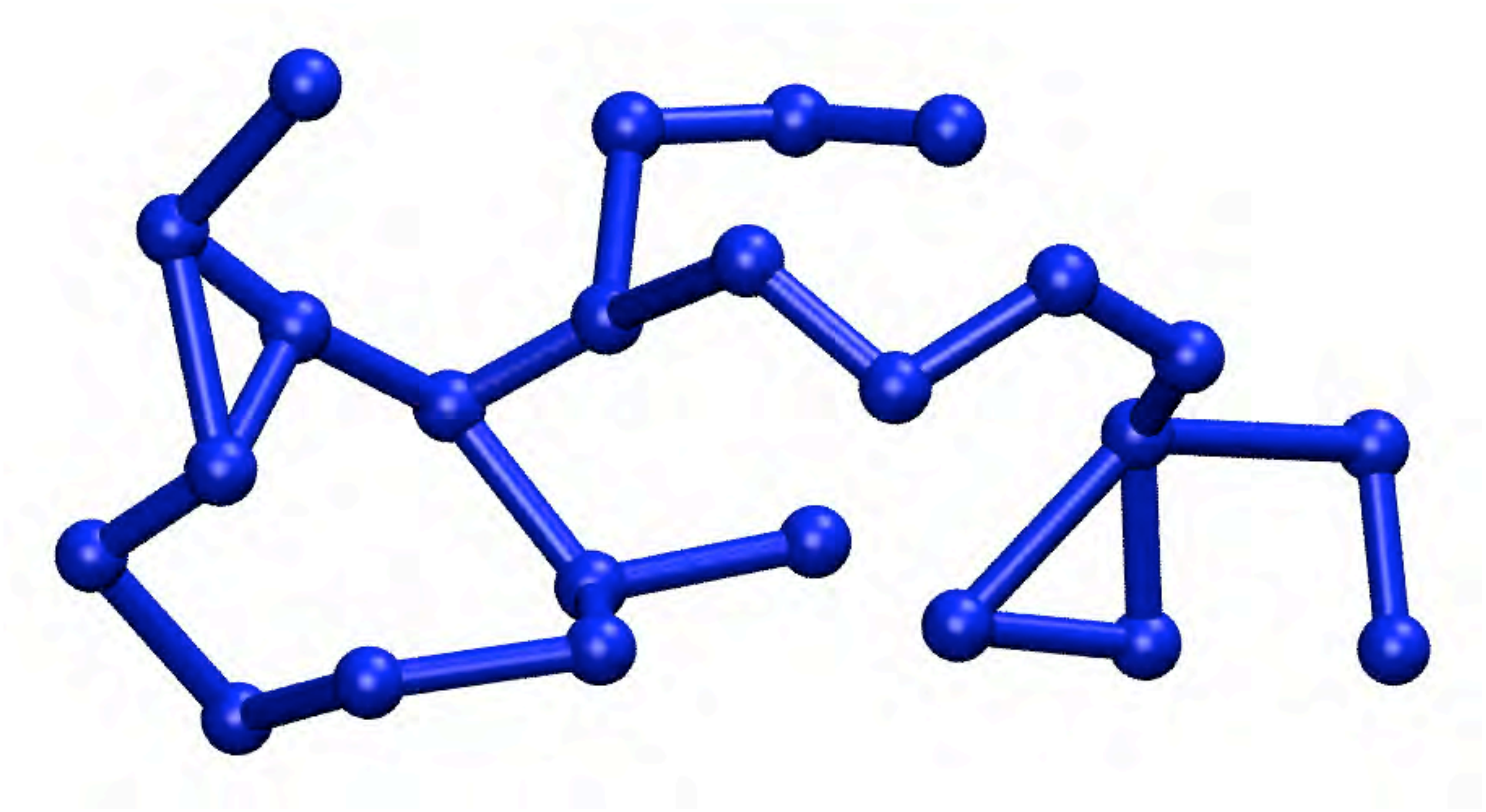
$$\text{ATP} + \text{Glu} + \text{tRNA} \xrightarrow{\text{GluRS}} \text{Glu-tRNA} + \text{AMP} + \text{PP}_i$$

How to Construct a Network?



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

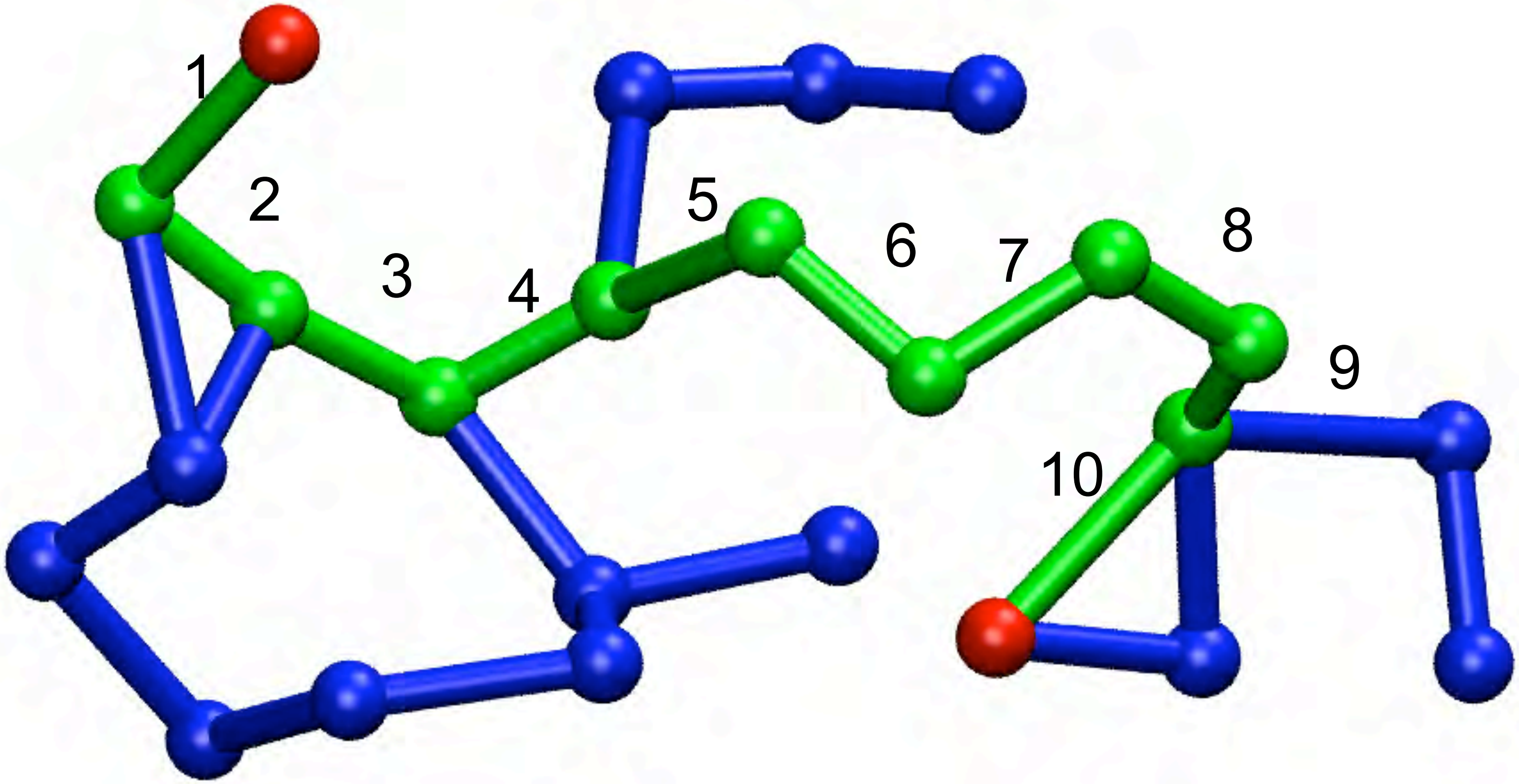
How to Construct a Network?



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

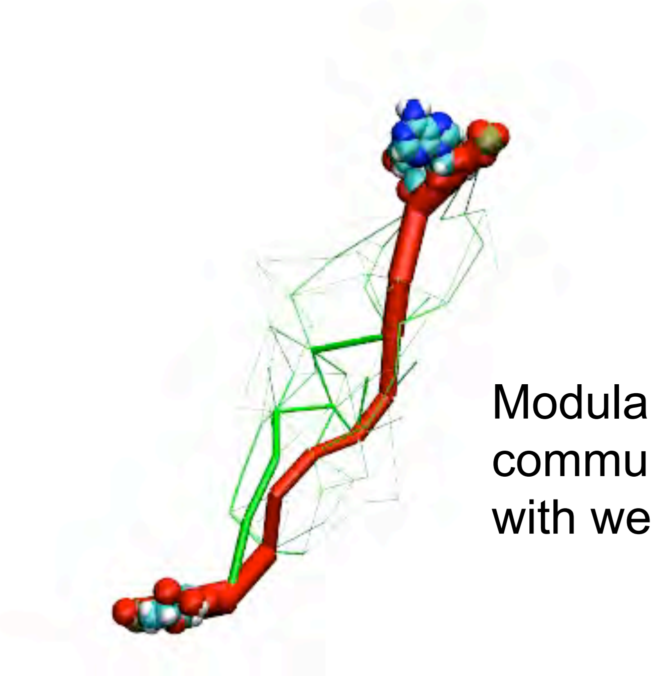
Different Paths can have Different Lengths



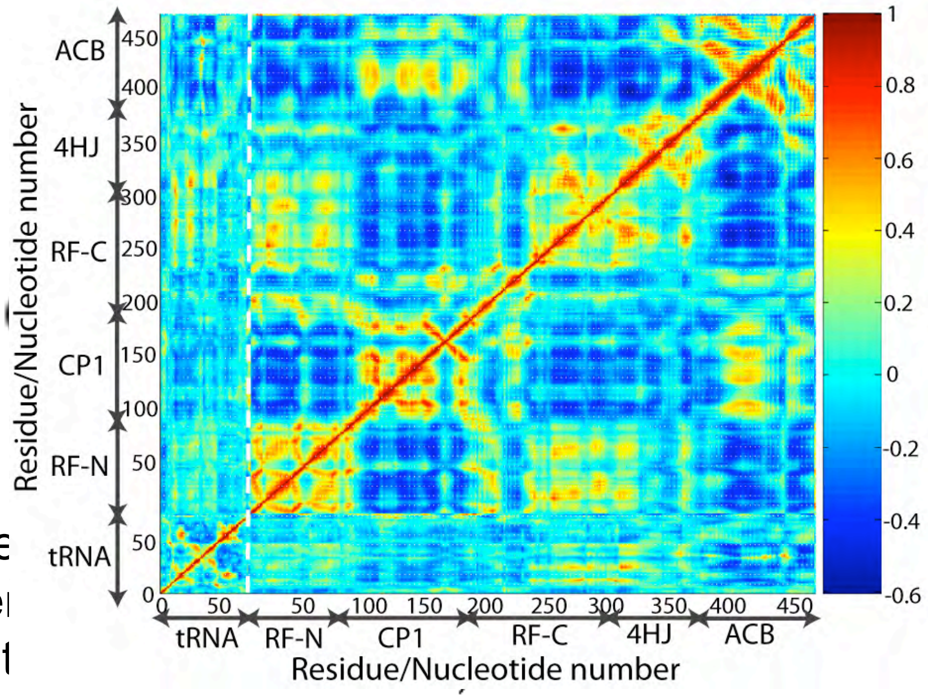
Path length = 10 (*unweighted*)

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

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



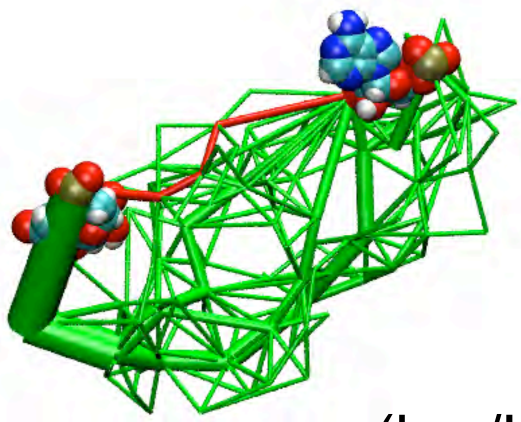
Modular structure
communities see
with weighted ne

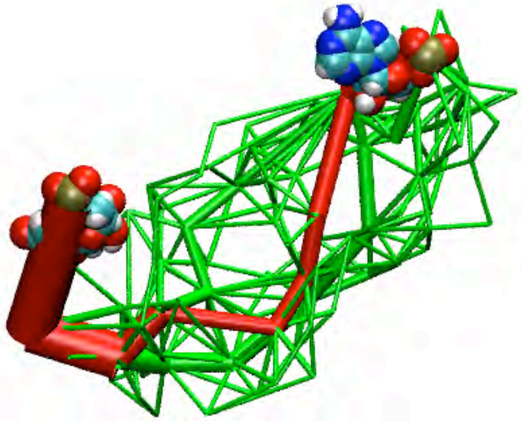


- MD simulations run with NAMD
- NTP ensemble with explicit solvent
- Neutralized with Mg^{2+} and K^+
- C_{ij} values calculated over a 16-ns window

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)

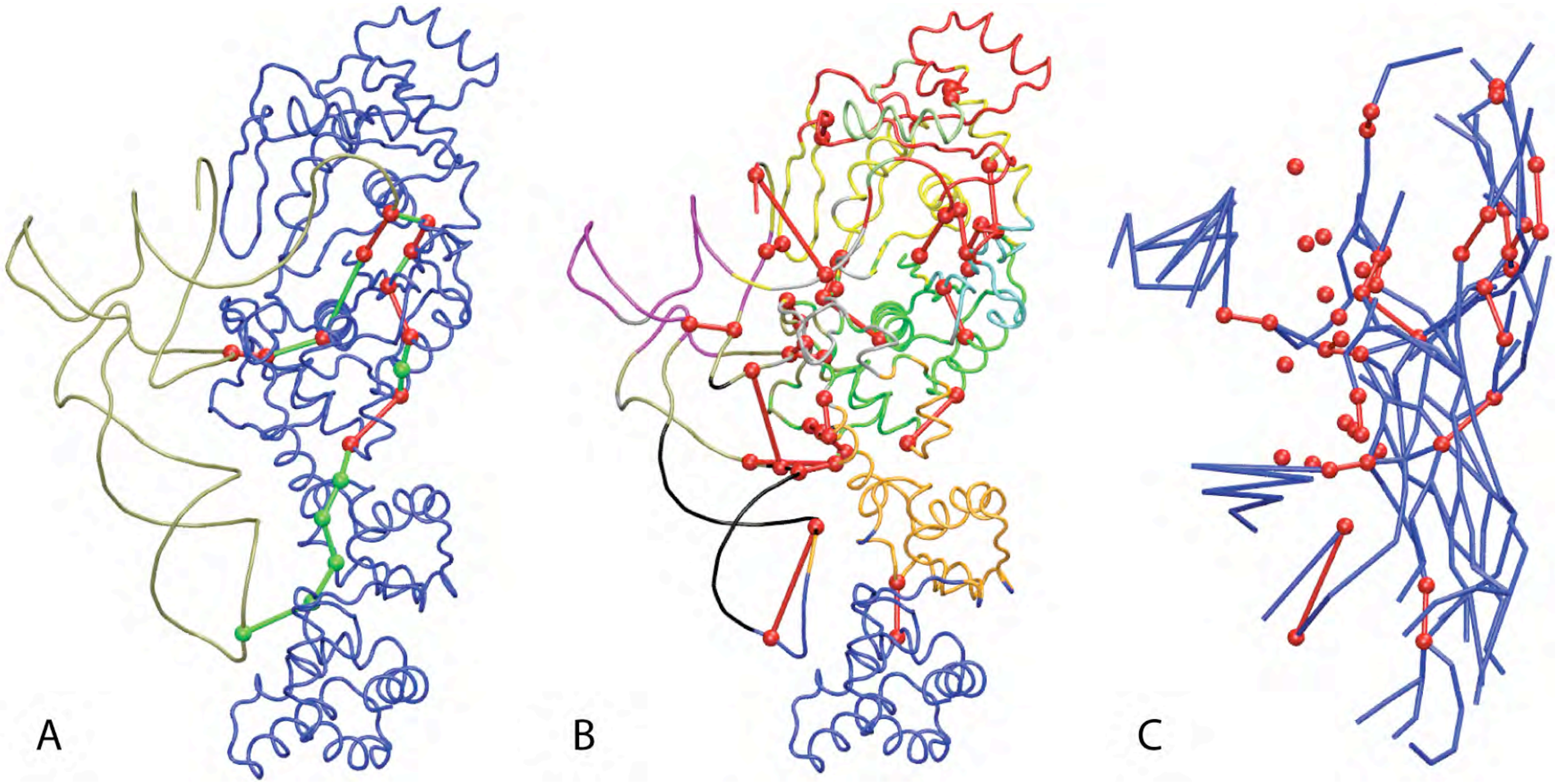
$$\frac{\exp(-\text{Sum of shortest path distances})_{U13mutant}}{\exp(-\text{Sum of shortest path distances})_{wt}} \approx 0.02$$

(Eargle, et al. PNAS 2009)

wild type

U13 mutant

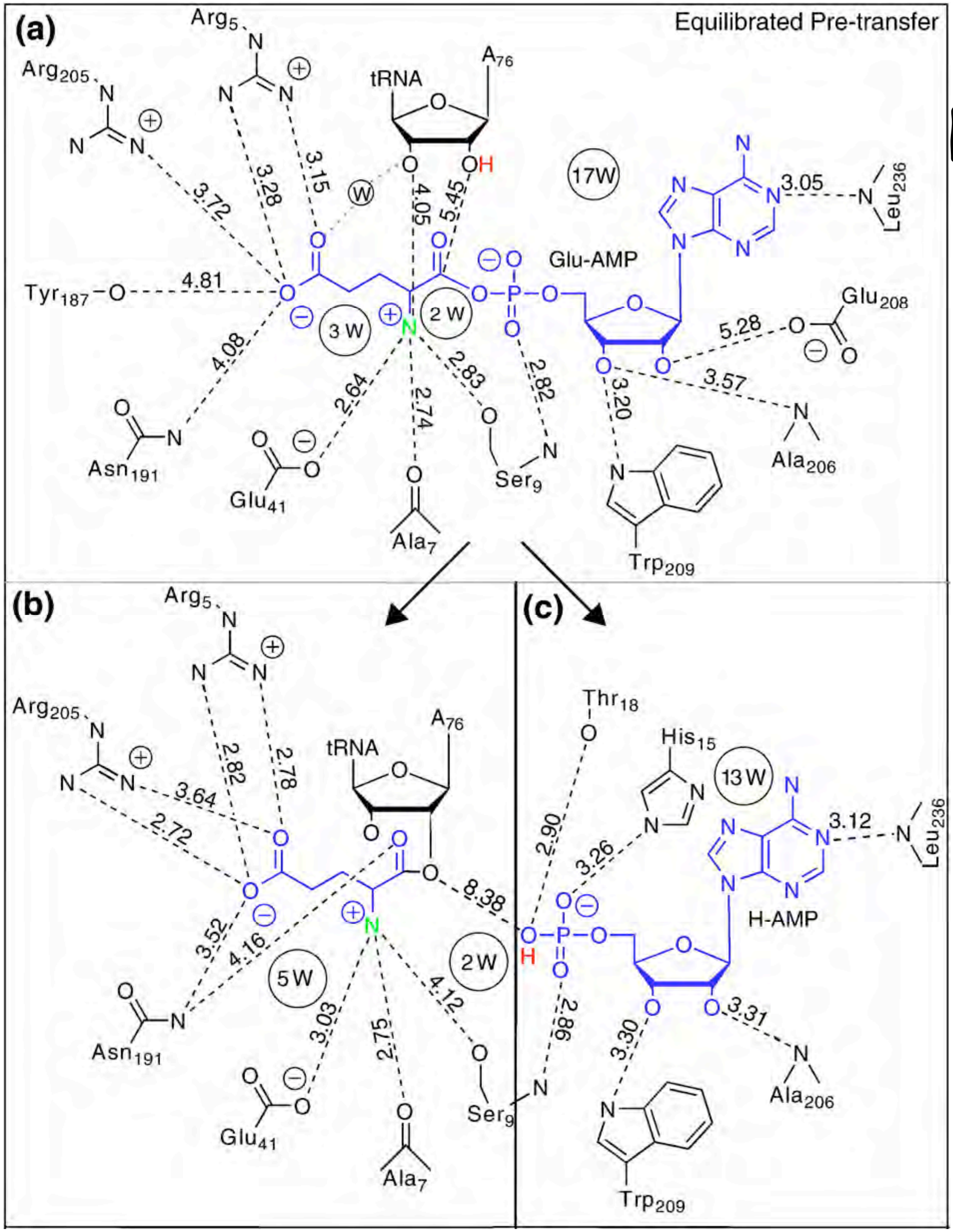
Dynamical Network, Conservation, and Betweeness



Optimal signal pathways: U13, U35 to A76

Critical(conserved) nodes connecting communities

Betweeness routes - highest density pair optimal paths



Reaction Mechanism

“Exit Strategy for Charged tRNA”

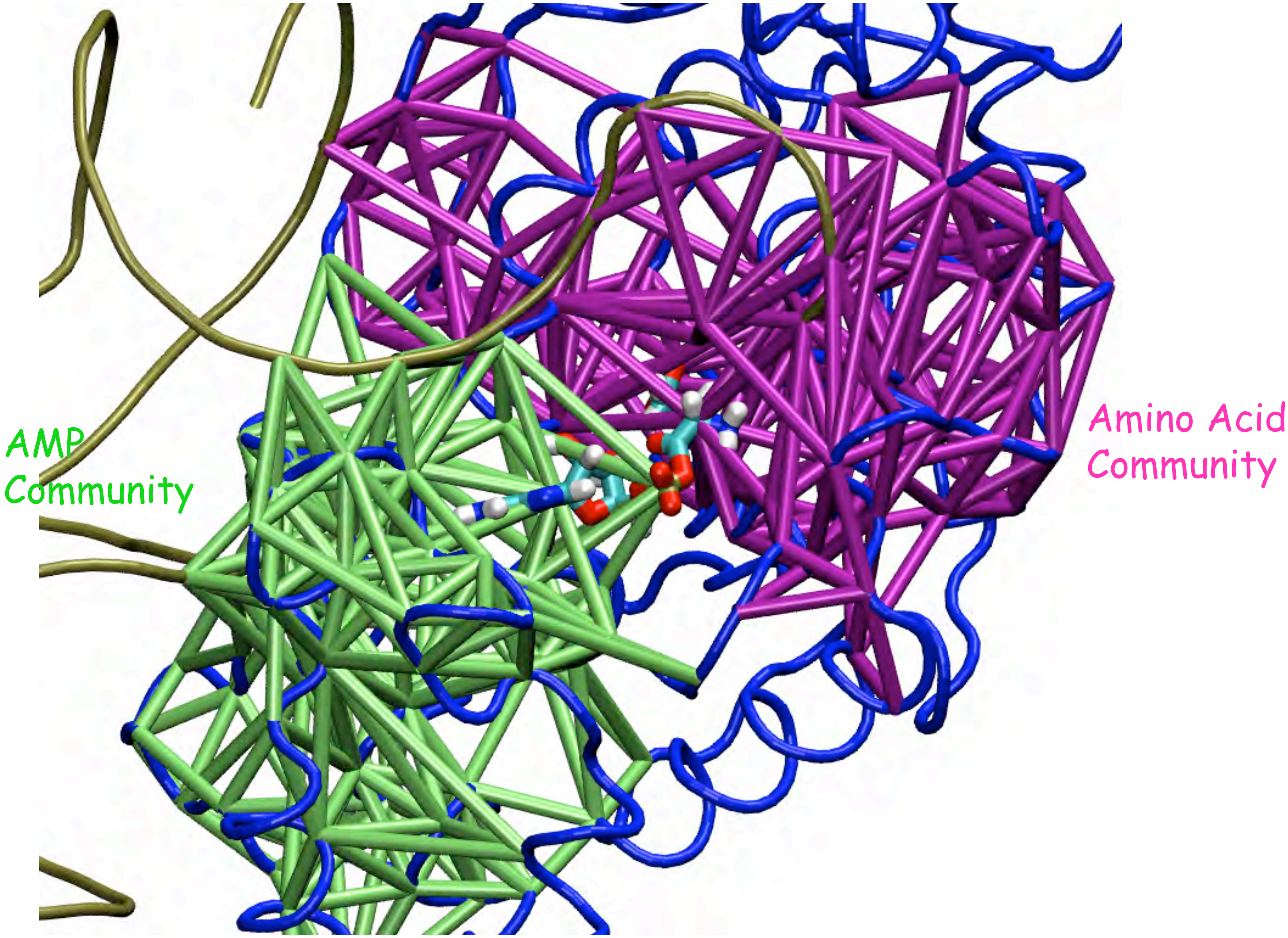
100 ns MD simulations

- Protonation states active site - PKa
- Presence/absence AMP
- FE - MM-PBSA calculations
- Shifts in Network from ID elements
- Presence of EF-TU

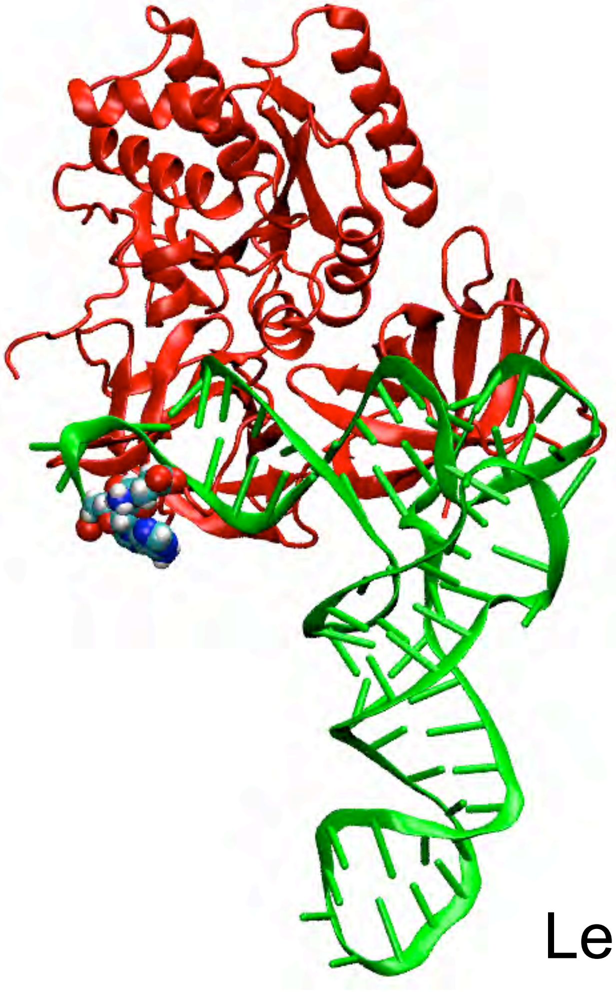
JMB 2010

Adenylate Network Communities

Reflects the mechanism



Changing Networks: tRNA Migration from GluRS to EF-Tu



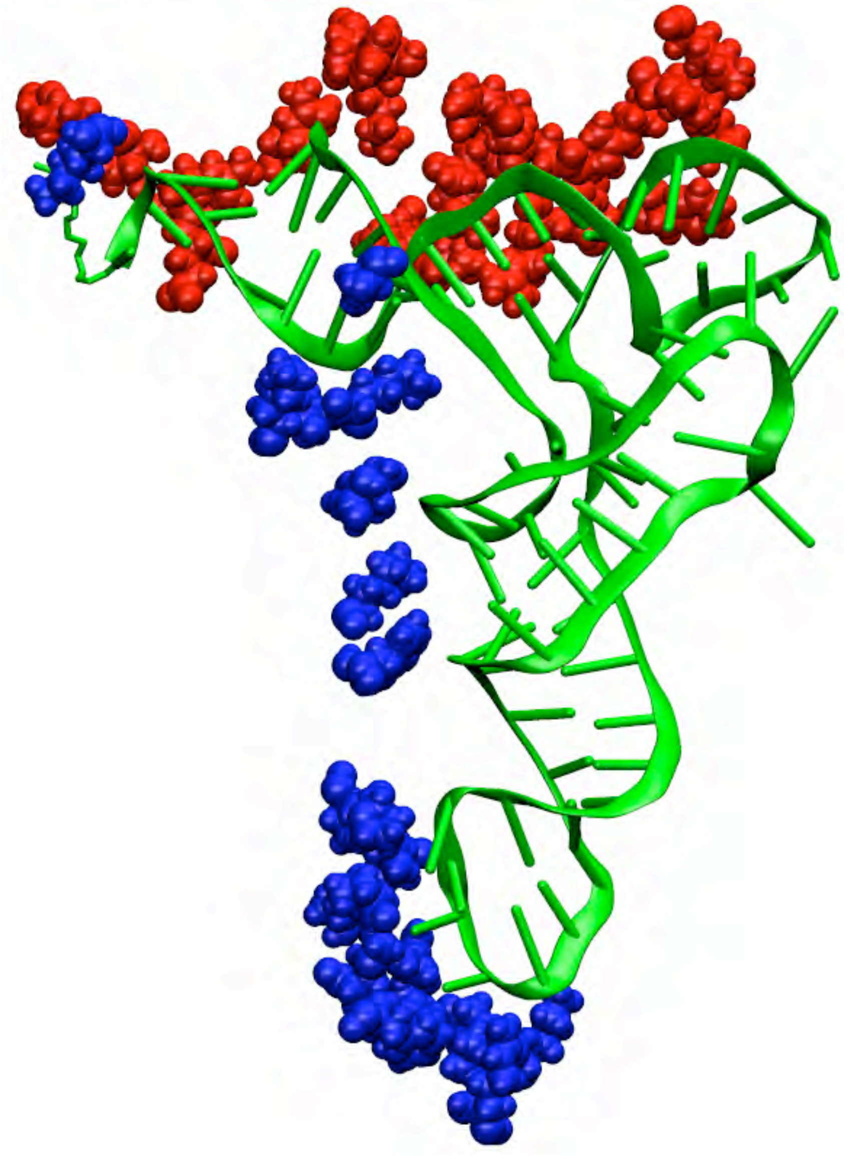
>2-fold increase in k_{cat} for CysRS:tRNA:EF-Tu

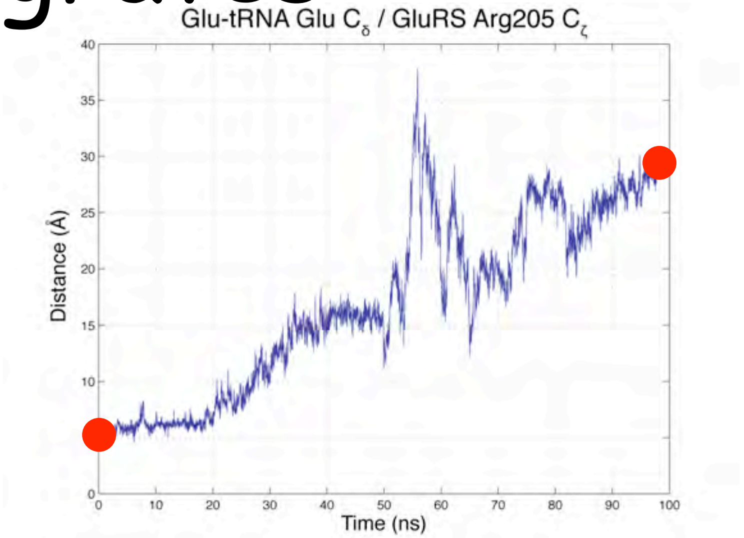
8-fold increase in k_{cat} for LeuRS:tRNA:EF-Tu

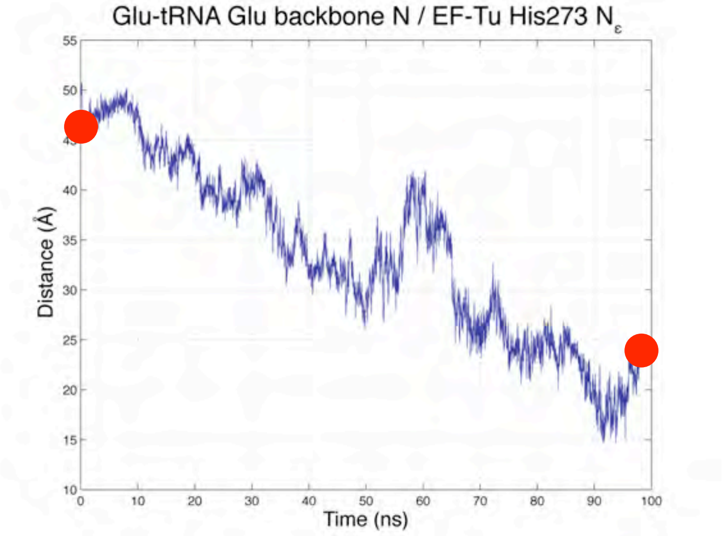
Zhang C, Perona J, Kang R, Francklyn C, Hou Y. *JMB* 2006

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

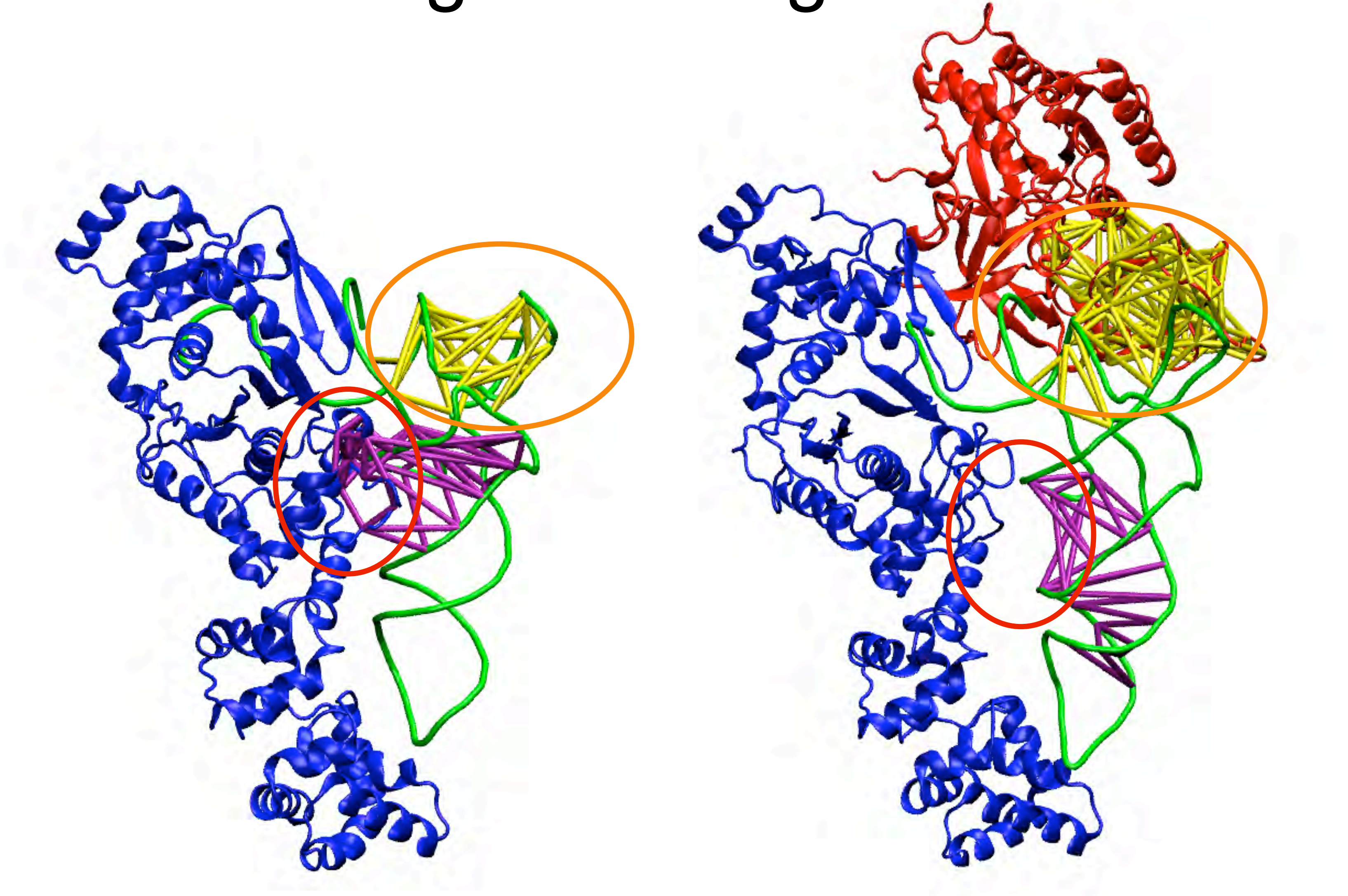
Change in Protein:RNA Contacts as tRNA Migrates







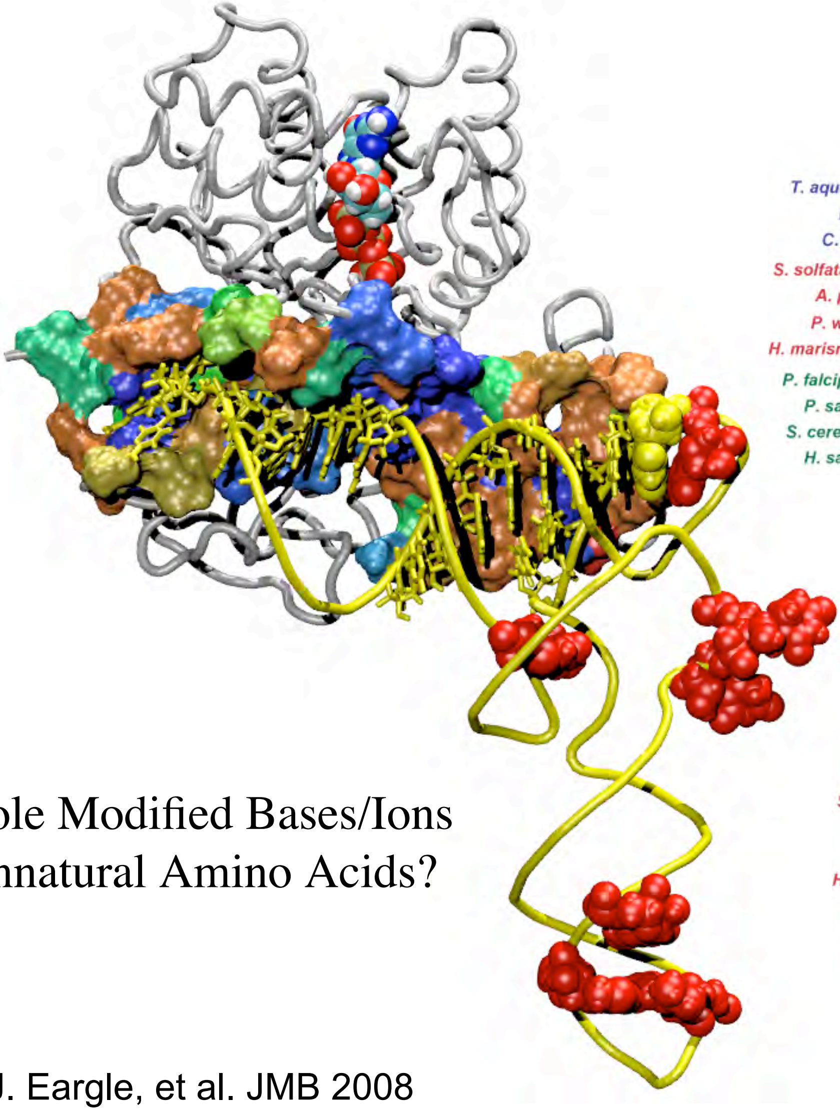
Change in Protein:RNA Contacts During tRNA Migration



purple - D arm community

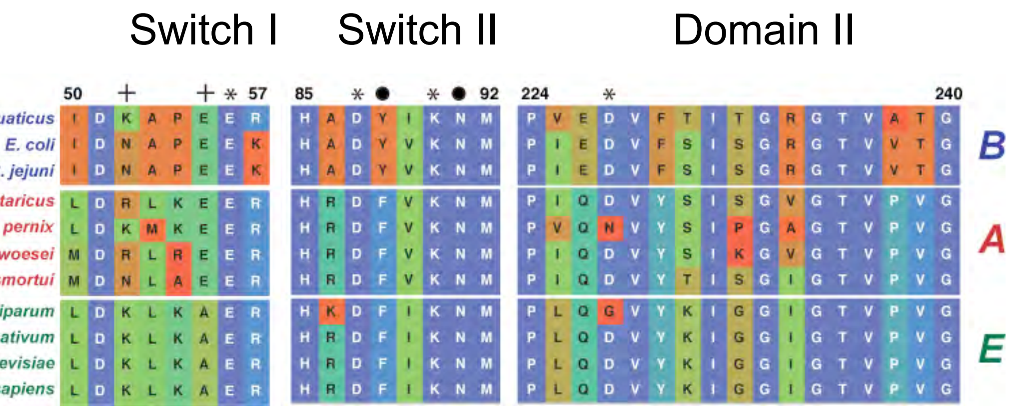
yellow - T arm community

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



Role Modified Bases/Ions
Unnatural Amino Acids?

J. Eargle, et al. JMB 2008



Switch I

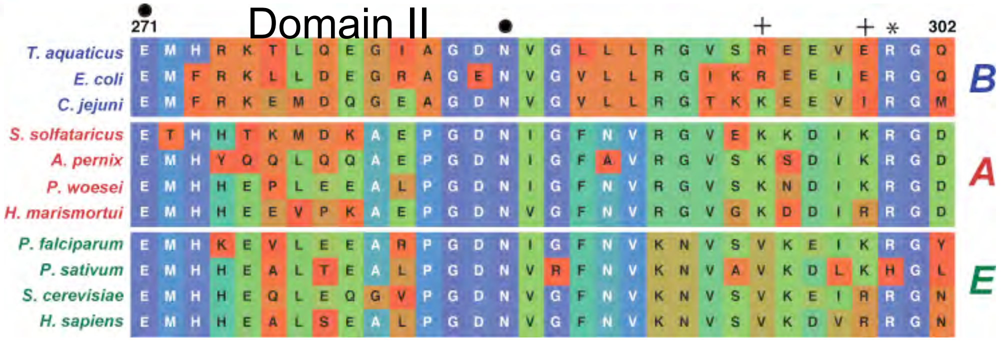
	50		+			+	*	57
T. aquaticus	I	D	K	A	P	E	E	R
E. coli	I	D	N	A	P	E	E	K
C. jejuni	I	D	N	A	P	E	E	K
S. solfataricus	L	D	R	L	K	E	E	R
A. pernix	L	D	K	M	K	E	E	R
P. woesei	M	D	R	L	R	E	E	R
H. marismortui	M	D	N	L	A	E	E	R
P. falciparum	L	D	K	L	K	A	E	R
P. sativum	L	D	K	L	K	A	E	R
S. cerevisiae	L	D	K	L	K	A	E	R
H. sapiens	L	D	K	L	K	A	E	R

Switch II

	85		*	●		*	●	92
T. aquaticus	H	A	D	Y	I	K	N	M
E. coli	H	A	D	Y	V	K	N	M
C. jejuni	H	A	D	Y	V	K	N	M
S. solfataricus	H	R	D	F	V	K	N	M
A. pernix	H	R	D	F	V	K	N	M
P. woesei	H	R	D	F	V	K	N	M
H. marismortui	H	R	D	F	V	K	N	M
P. falciparum	H	K	D	F	I	K	N	M
P. sativum	H	R	D	F	I	K	N	M
S. cerevisiae	H	R	D	F	I	K	N	M
H. sapiens	H	R	D	F	I	K	N	M

Domain II

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



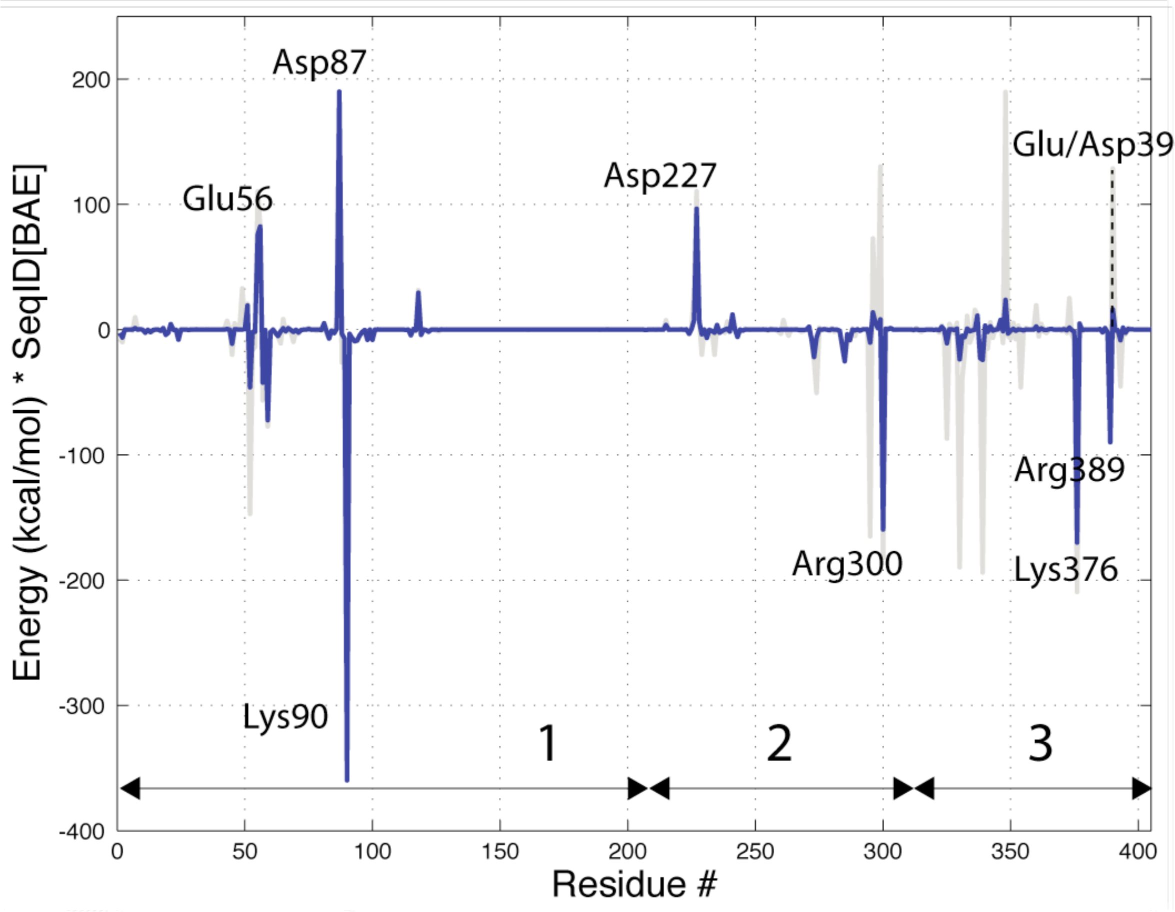
Domain II

	● 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	
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	B
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. 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	A
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	
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	E
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	

Dynamical Recognition EF-Tu/tRNA (E.coli)

Combining MD with Evolutionary Analysis (MultiSeq)

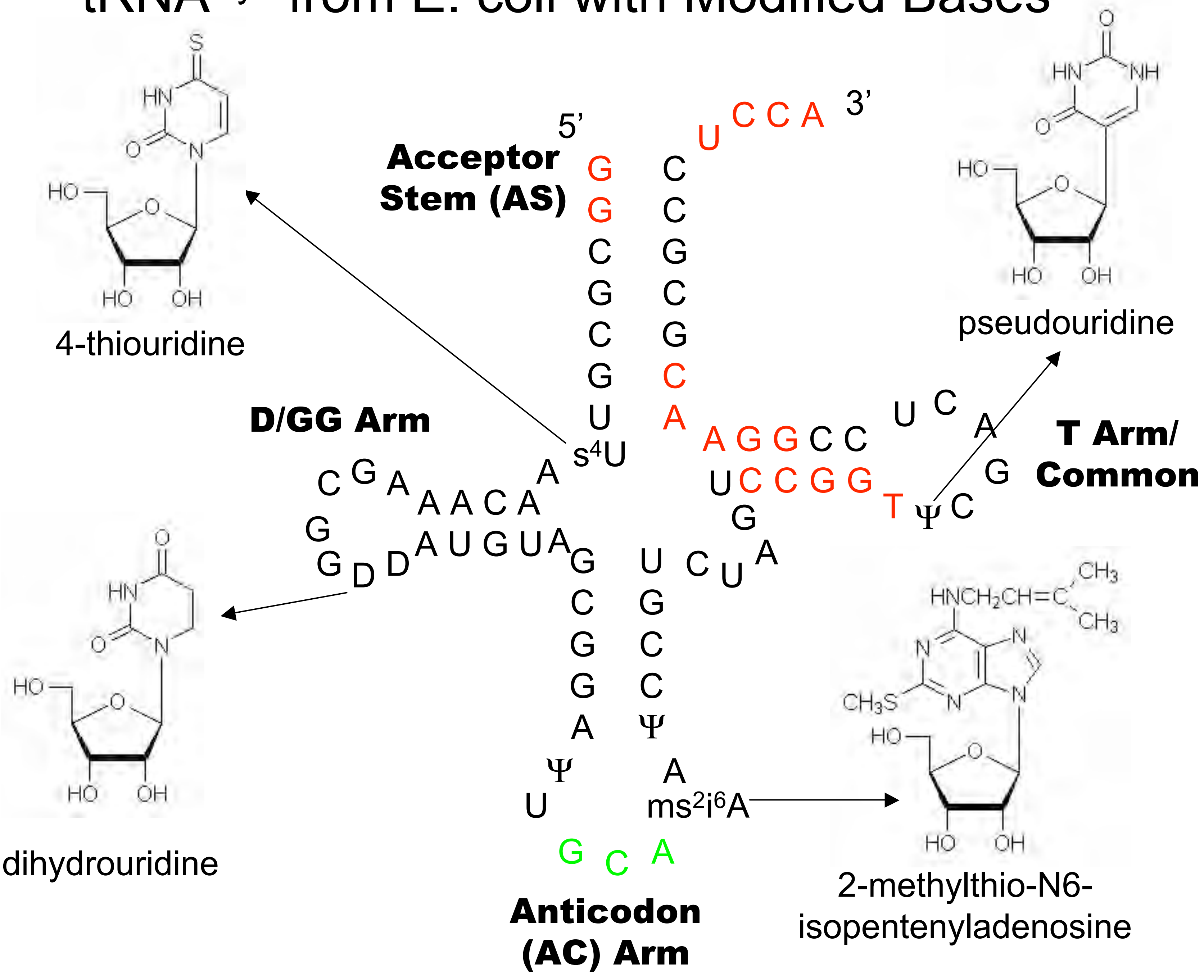
Non-bonded Interaction Energy



BAE

J. Eargle, A. Sethi, A. Black, L. Trabuco & Z. Luthey-Schulten. JMB
“Dynamics of Recognition in EF-Tu/tRNA Complex” (2008)

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



Acceptor Stem (AS)

D/GG Arm

T Arm/ Common

Anticodon (AC) Arm

4-thiouridine

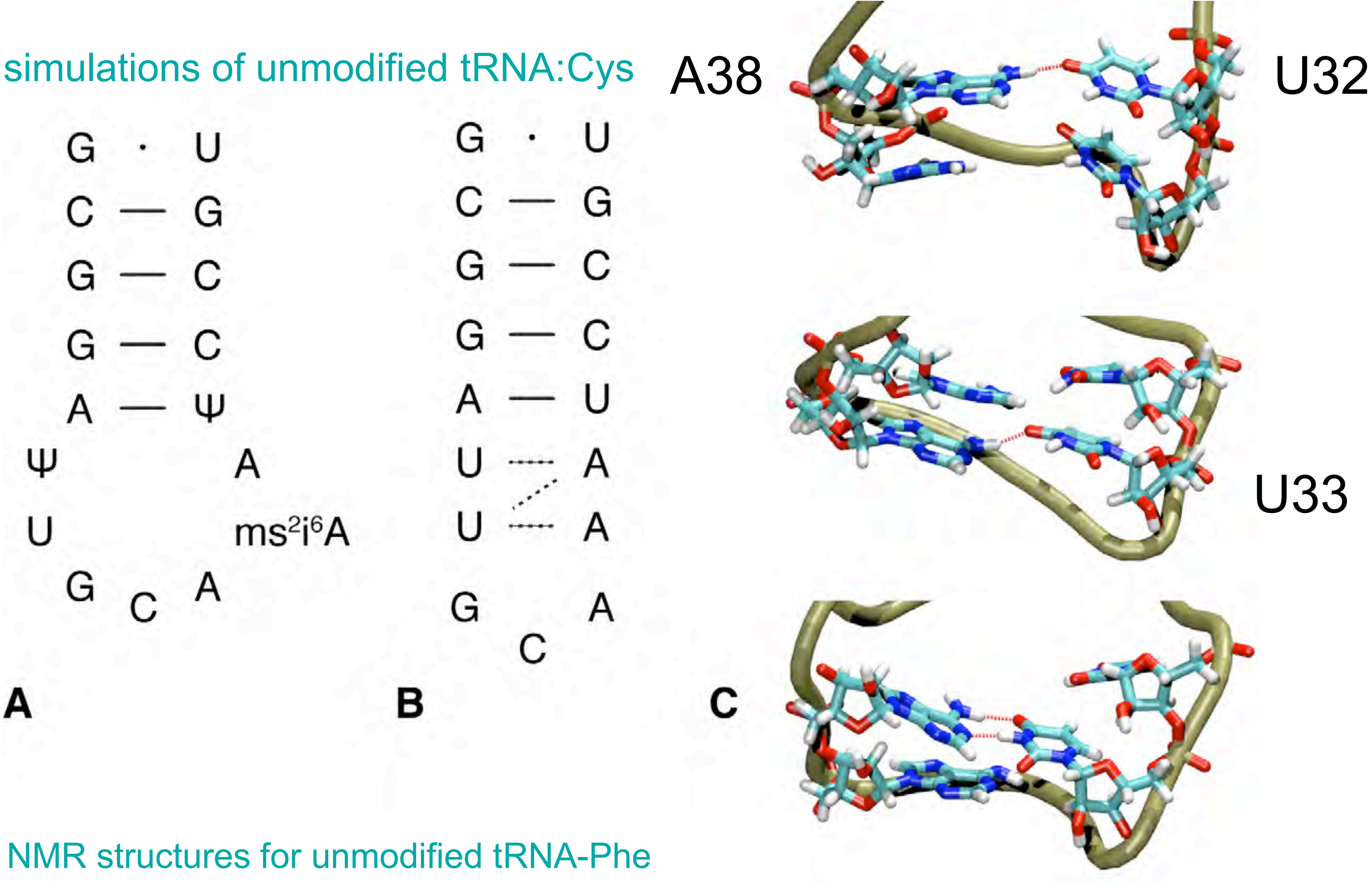
pseudouridine

dihydrouridine

2-methylthio-N6-isopentenyladenosine

Effects of Modified Bases on Dynamics of AC Stem/Loop

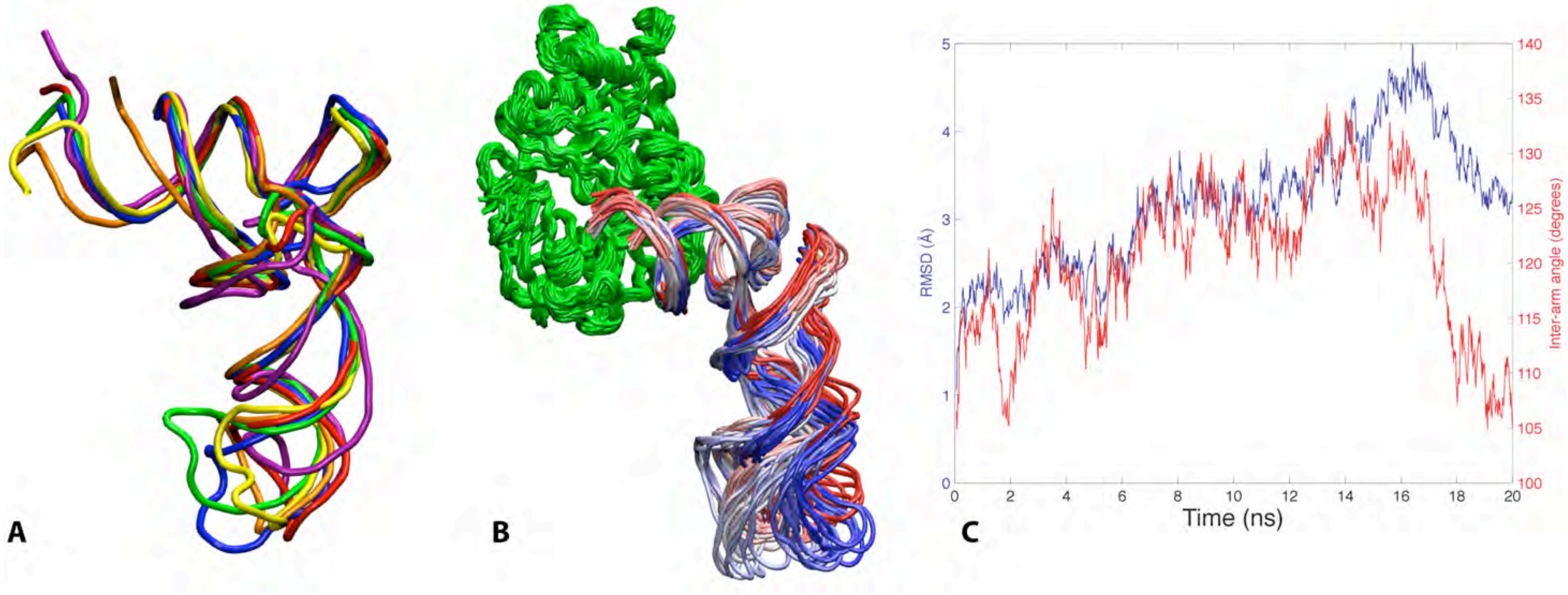
MD simulations of unmodified tRNA:Cys



Similar NMR structures for unmodified tRNA-Phe
Nikonowicz, *JMB* 2002, *personal communication*

Eargle, et al. *JMB* (2008))

Flexibility in tRNA structures observed in crystallography and simulations



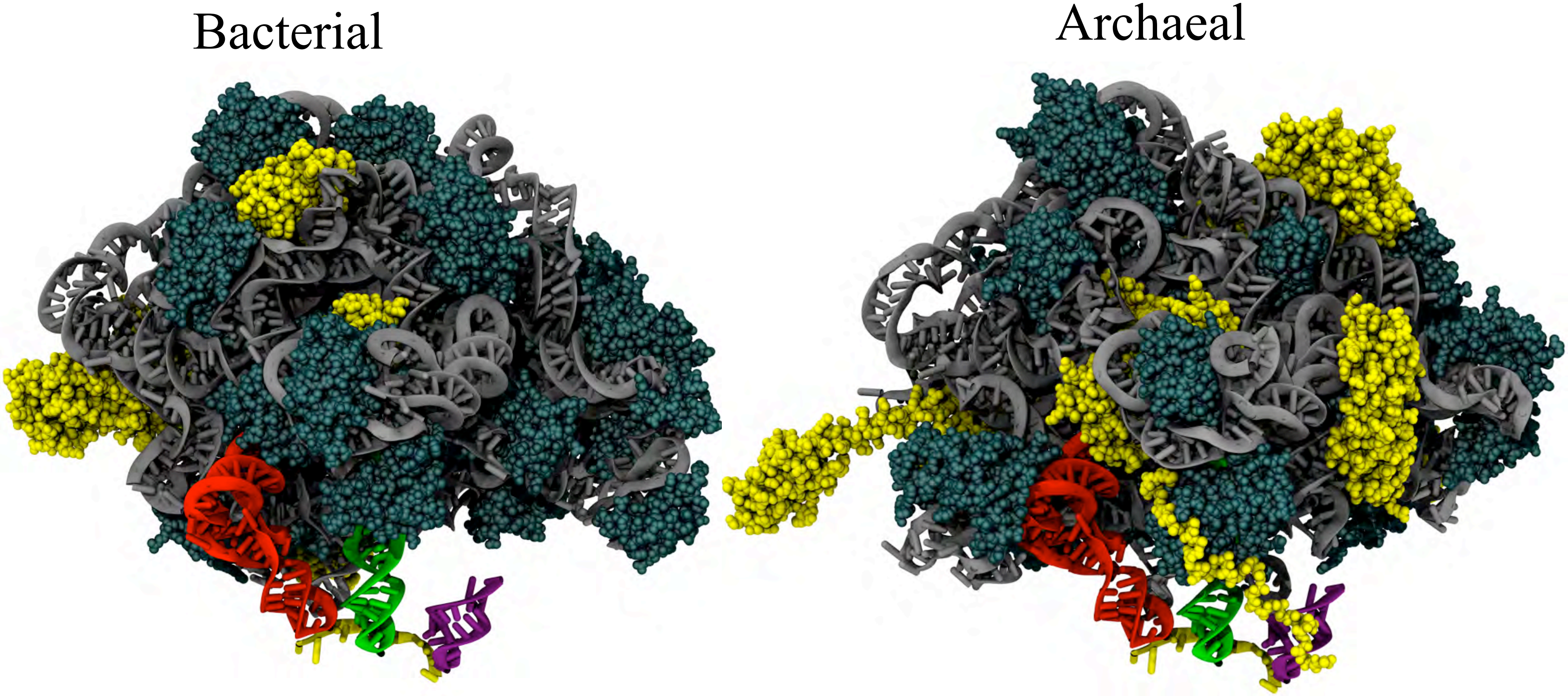
X-ray 5 tRNAs

MD EF-Tu:tRNA:cys

MD: RMSD and Interarm
Angle in tRNA:cys

JMB 2008 & FEBS Let. 2010

Molecular Signatures in Ribosome Evolution

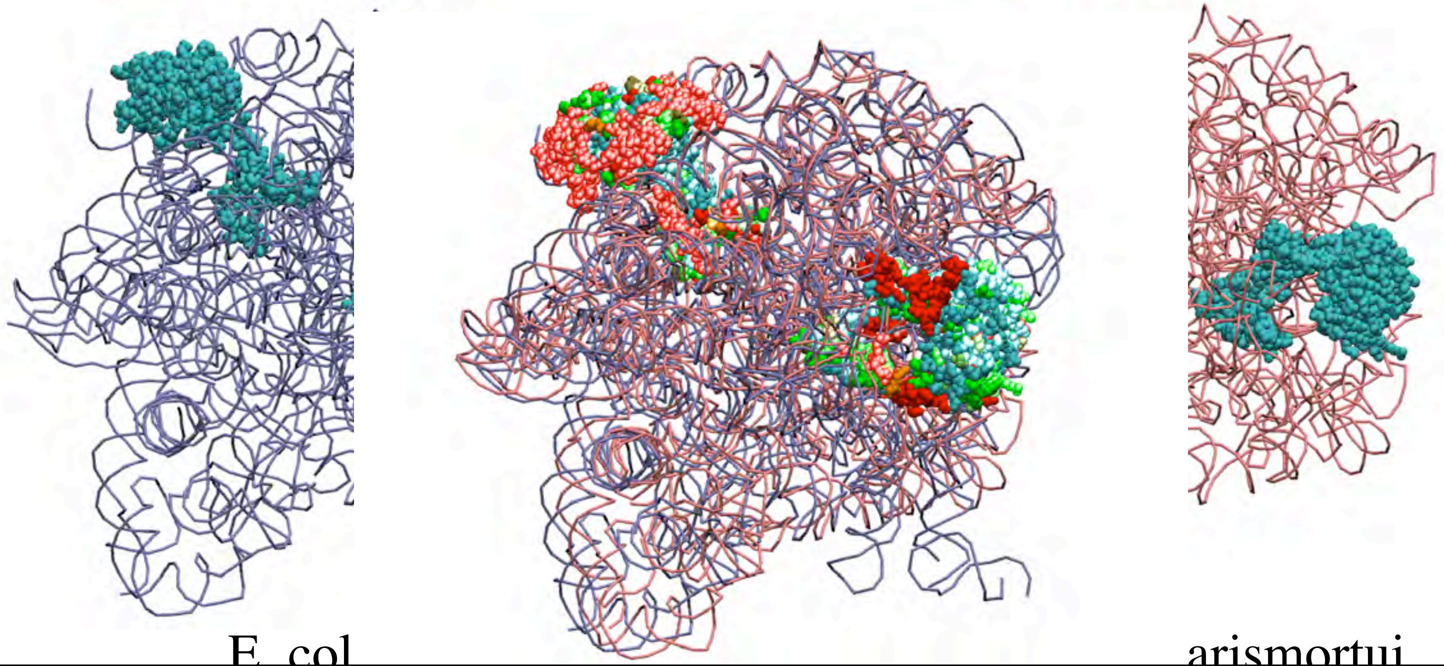


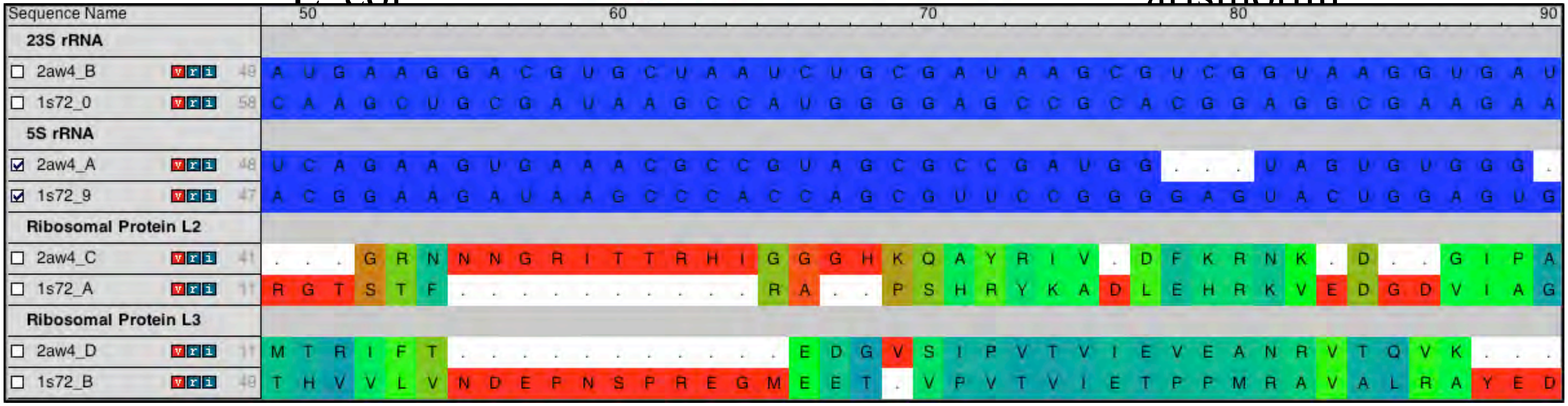
Ribosomal signatures: idiosyncrasies in ribosomal RNA and/or proteins characteristic of the domains of life

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

Structural Overlaps with STAMP

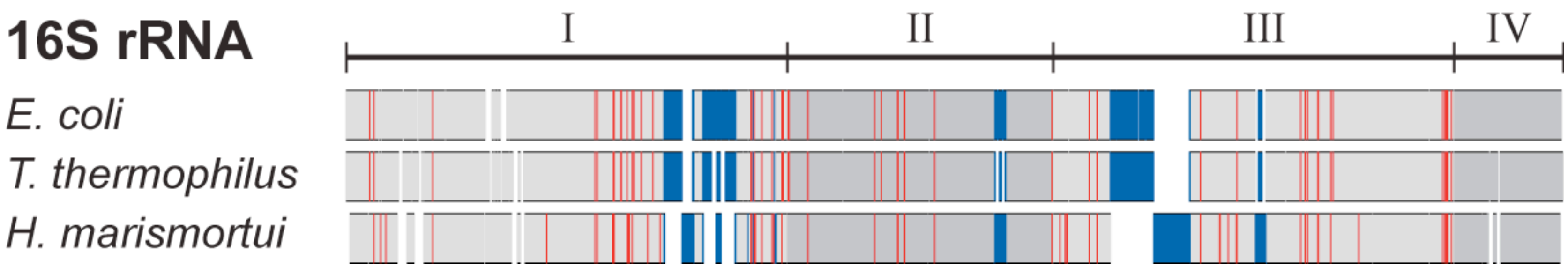
Comparing ribosomal large subunit with r-proteins L2 and L3
180,000 atoms in 4 rRNAs and 58 proteins





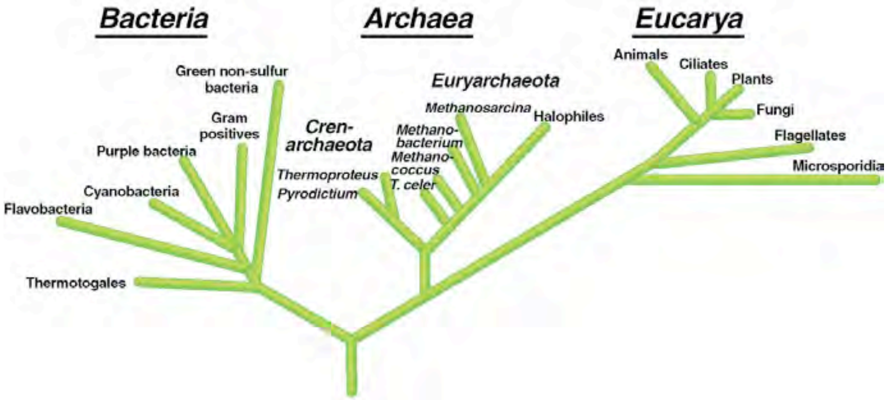
Sequence Name		50 · 60 · 70 · 80 · 90
23S rRNA		
☐ 2aw4_B	49	AUGAAGGACGUGCUAAUCUGCGAUAAGCGUCGGUAAGGUGAU
☐ 1s72_0	58	CAAGCUGCGAUAAGCCAUGGGGAGCCGCACGGAGGCGAAGAA
5S rRNA		
☑ 2aw4_A	48	UCAGAAGUGAAACGCCGUAGCGCCGAUGG...UAGUGUGGG.
☑ 1s72_9	47	ACGGAAGAUAAGCCCACCAGCGUUCCGGGGAGUACUGGAGUG
Ribosomal Protein L2		
☐ 2aw4_C	41	...GRNNNGRITTRHIGGGHKQAYRIV.DFKRNK.D..GIPA
☐ 1s72_A	11	RGTSTF..........RA..PSHRYKADLEHRKVEDGDVIAG
Ribosomal Protein L3		
☐ 2aw4_D	11	MTRIFT...........EDGVSIPVTVIEVEANRVTQVK...
☐ 1s72_B	49	THVVLVNDEPNSPREGMEET.VPVTVIETPPMRAVALRAYED

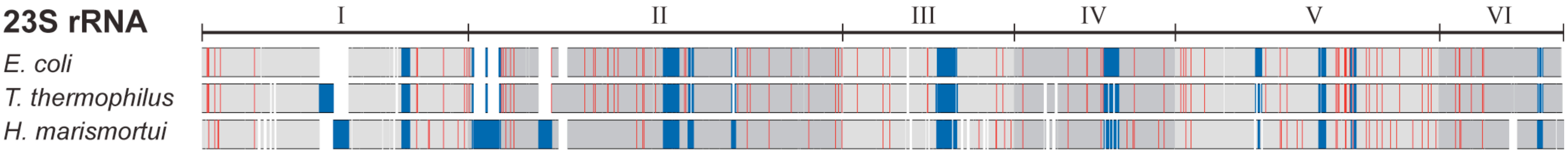
Sequence and Structure Signatures in Ribosomal RNA



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

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

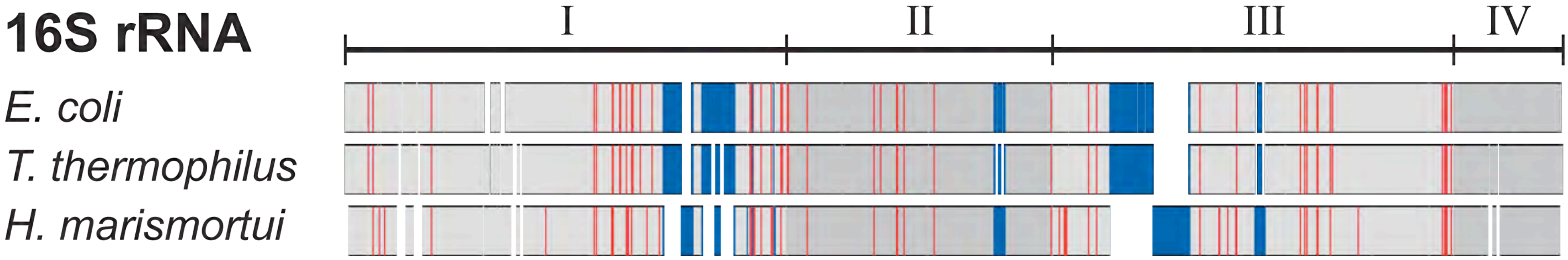




Roberts, Montoya, Sethi, Woese, ZLS, *PNAS* 2008

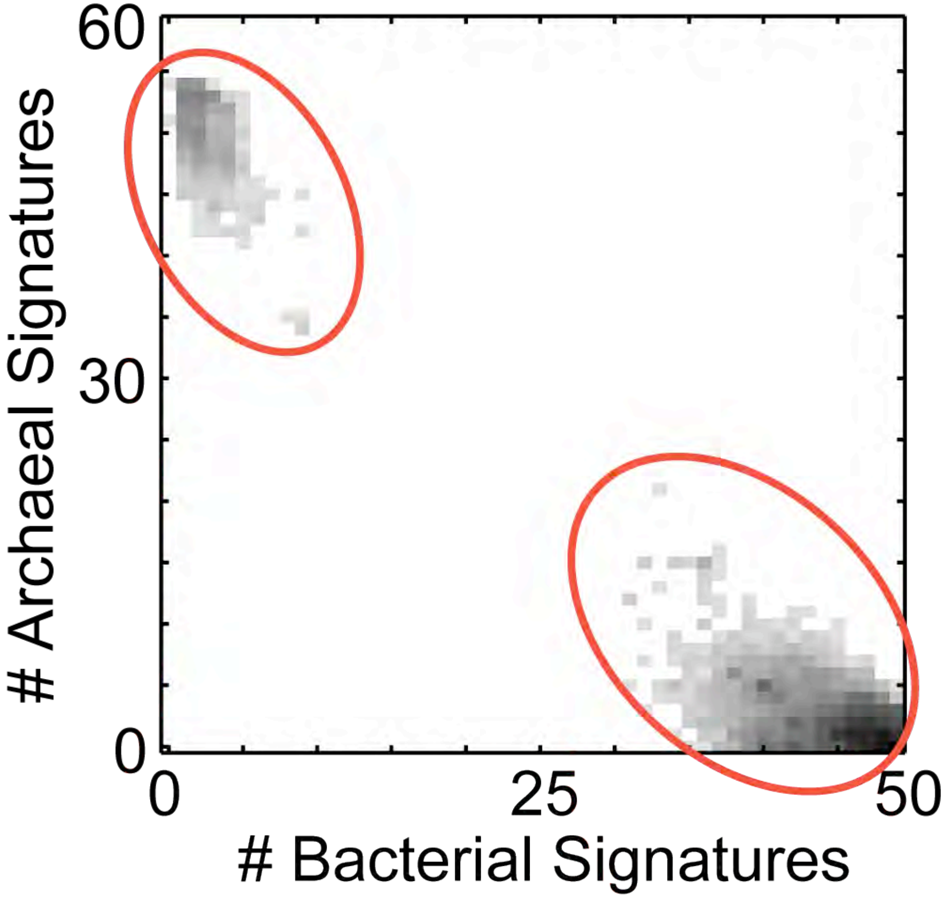
Draw signature figure (cont)

```
proc drawSequencePositions {g x y1 y2 sequenceID firstPosition lastPosition positionWidth sequenceColor borderColor
                                                                                                 borderWidth}
{
    set startPosition ""
    for {set j $firstPosition} {$j <= $lastPosition} {incr j} {
        set element [::SeqData::getElement $sequenceID $j]

        # If this is the beginning of a sequence block, record that one started.
        if {$element != "-" && $startPosition == ""} {
            set startPosition $j
        }

        # If this is the end of a sequence block, draw the block.
        if {($element == "-" || $j == $lastPosition) && $startPosition != ""} {
            if {$element == "-"} {
                set endPosition [expr $j-1]
            } else {
                set endPosition [expr $j]
            }
            set x1 [expr $x+($startPosition*$positionWidth)]
            set x2 [expr $x+(($endPosition+1)*$positionWidth)]
            $g create rectangle $x1 $y1 $x2 $y2 -fill $sequenceColor -outline $sequenceColor -width $borderWidth
            $g create line $x1 $y1 $x2 $y1 -fill $borderColor -width $borderWidth
            $g create line $x1 $y2 $x2 $y2 -fill $borderColor -width $borderWidth
            set startPosition ""
        }
    }
}
```



90,000 Environmental 16S rRNA Confirm Sequence Signatures

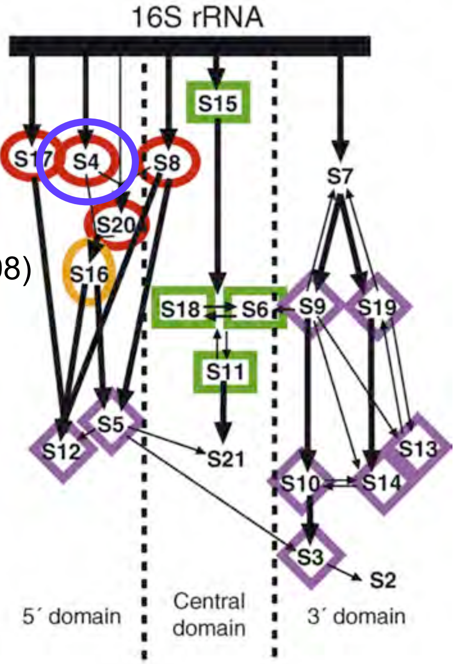
- Analysis of the ribosomal signatures in 90,000 environmental samples shows that no "gray" area exists: a ribosome is either bacterial or archaeal in nature.
- It follows then that for the domains of life there is also no continuum. The bacteria and archaea are separated by an evolutionary saltation.



Data: "Greengenes", Lawrence Berkeley

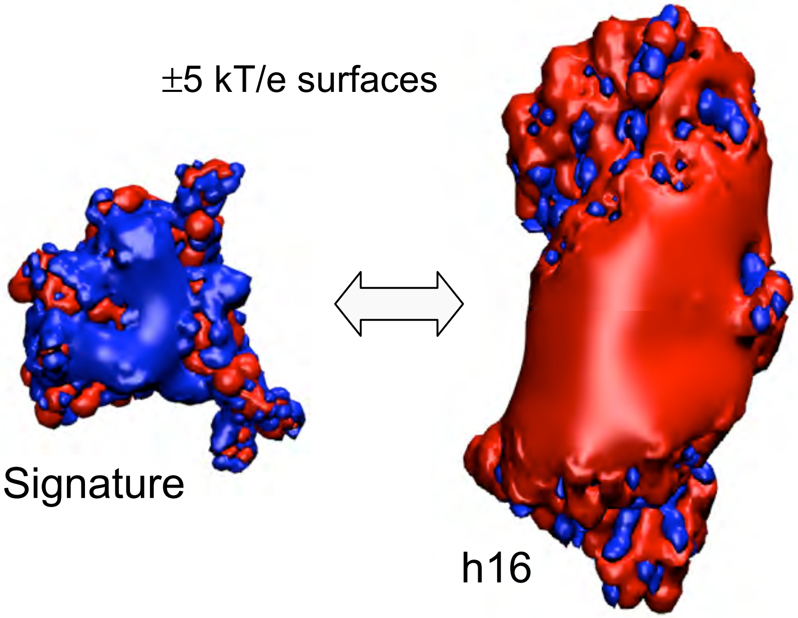
Roberts, et al. *PNAS* 2008

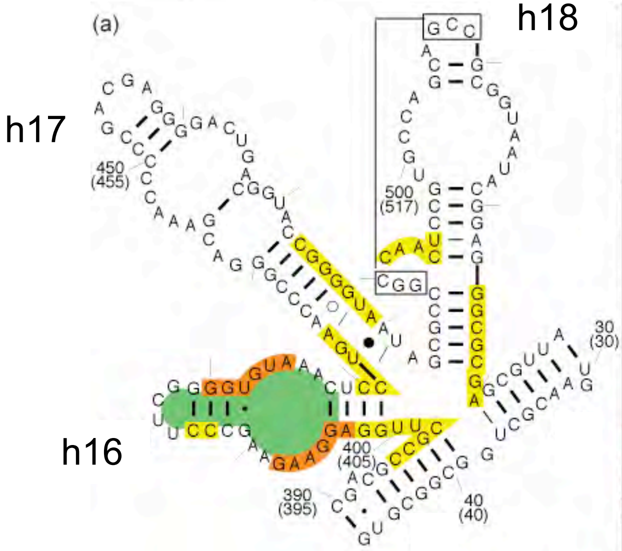
Effects of Signatures on Ribosomal Assembly?

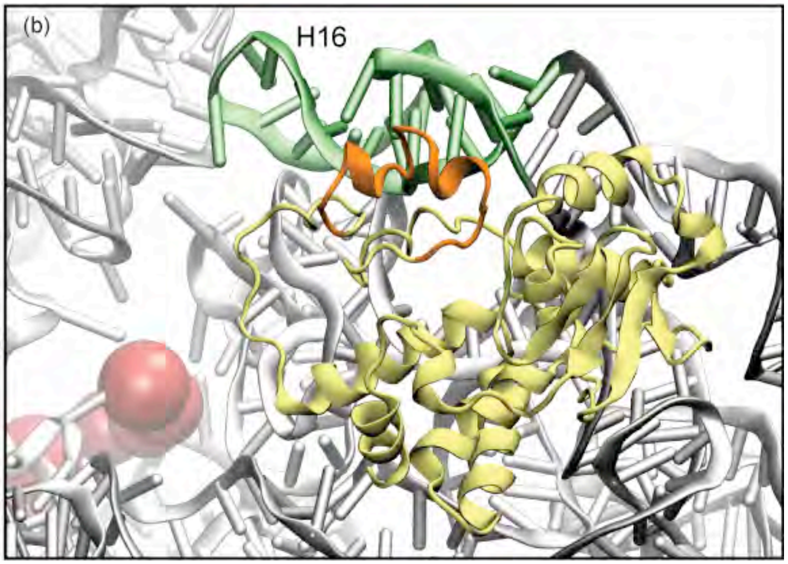


Williamson (2008)
Nomura (1970)

±5 kT/e surfaces





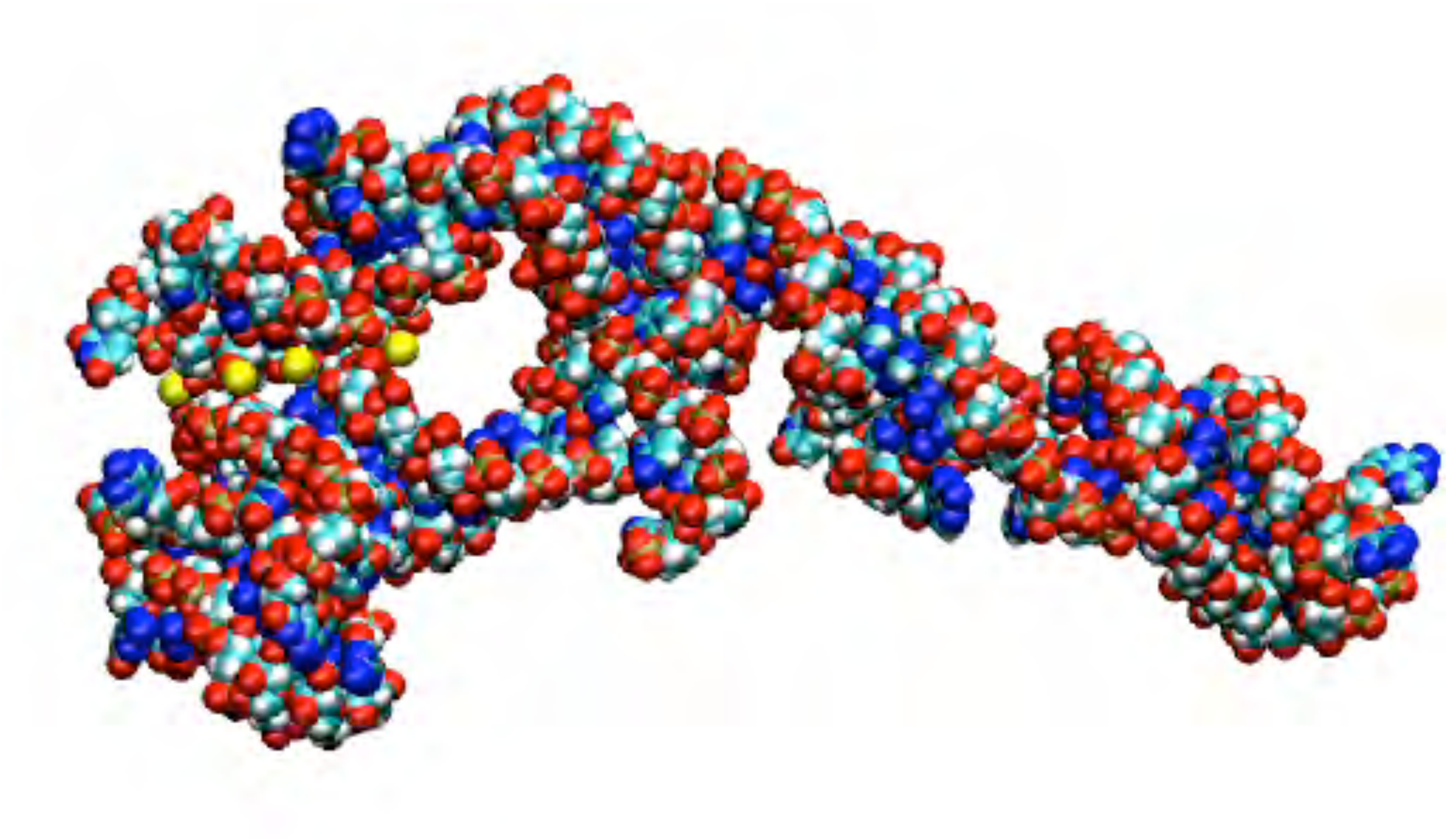


E. Roberts, A. Sethi, J. Montoya, C. Woese, Z. Luthey-Schulten, *PNAS*, (2008)

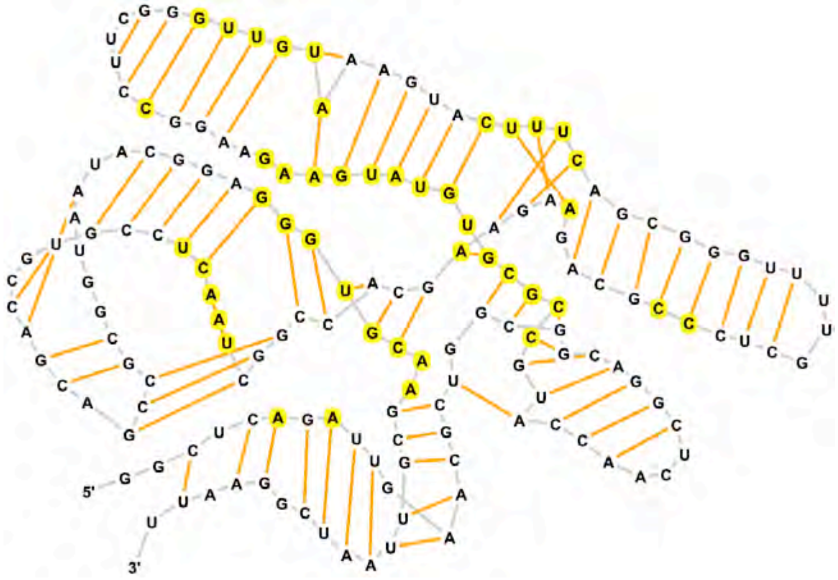
Conformational Changes in (Dis)Assembly

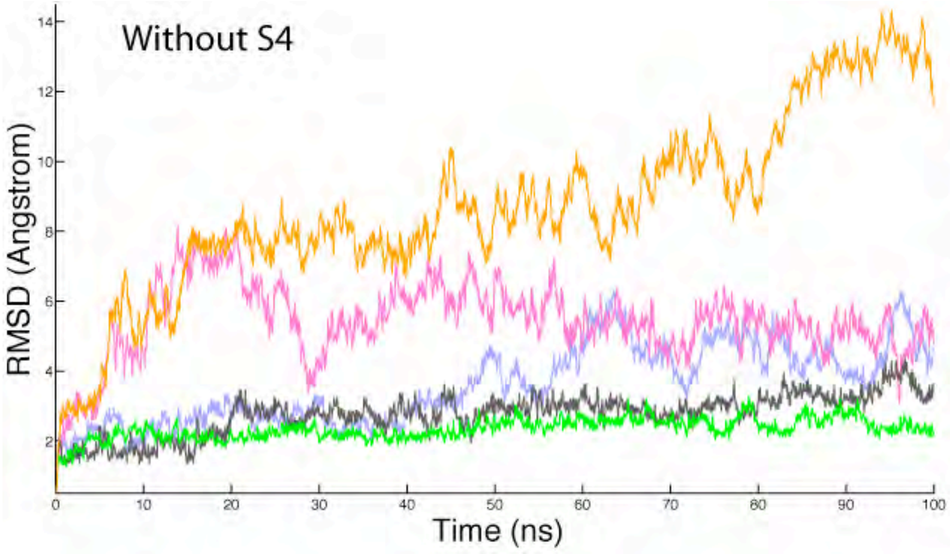
Without S4, h16 Separates from h18

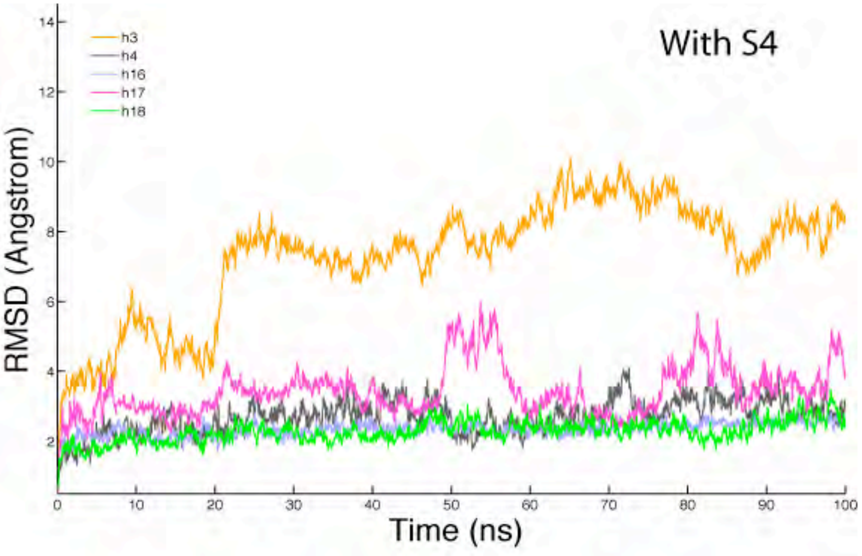
John Eargle and Ke Chen - With TJ Ha, and S. Woodson



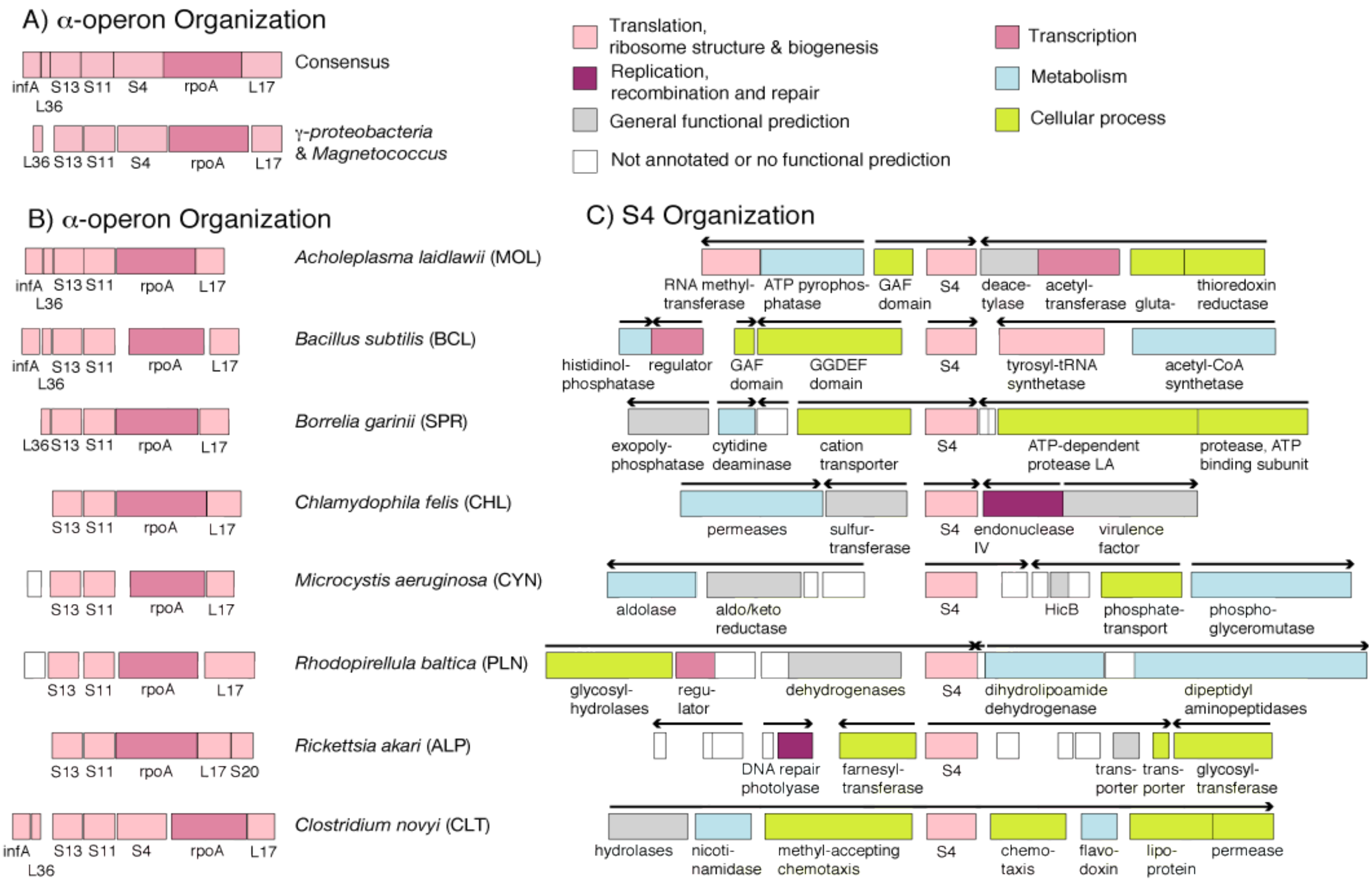
Stability of 5-way junction helices with/without S4





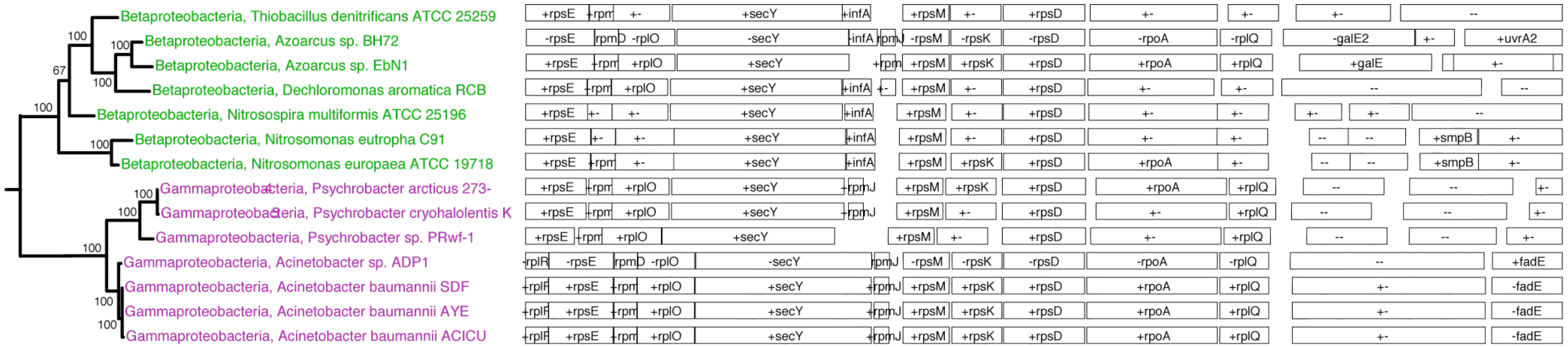


Signatures in Genomic Context



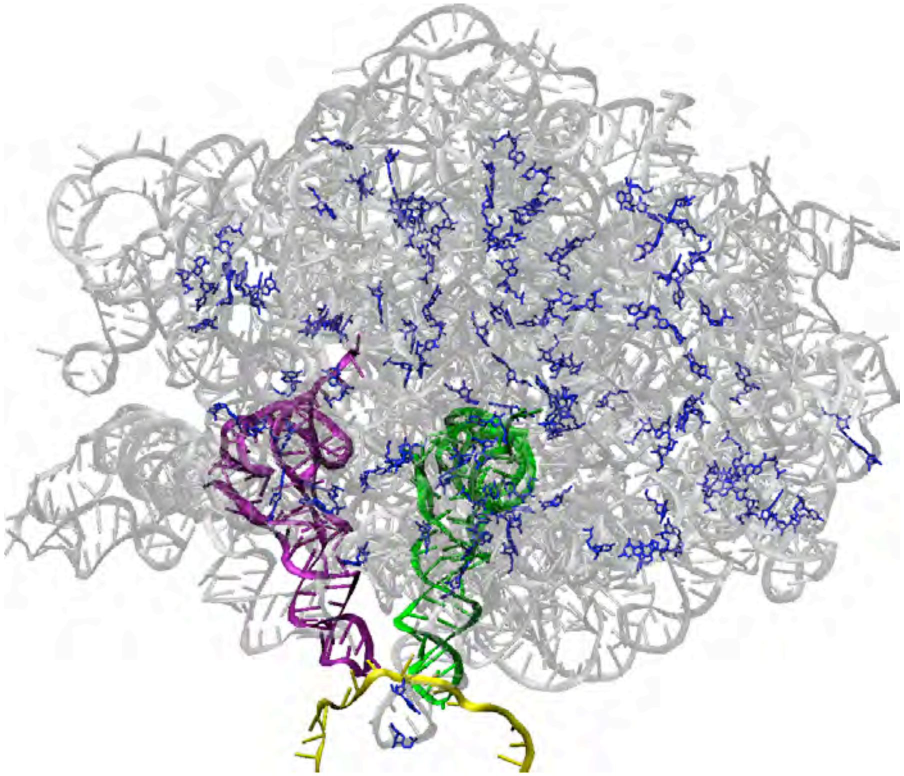
Combined genomic context/phylogenetic tree

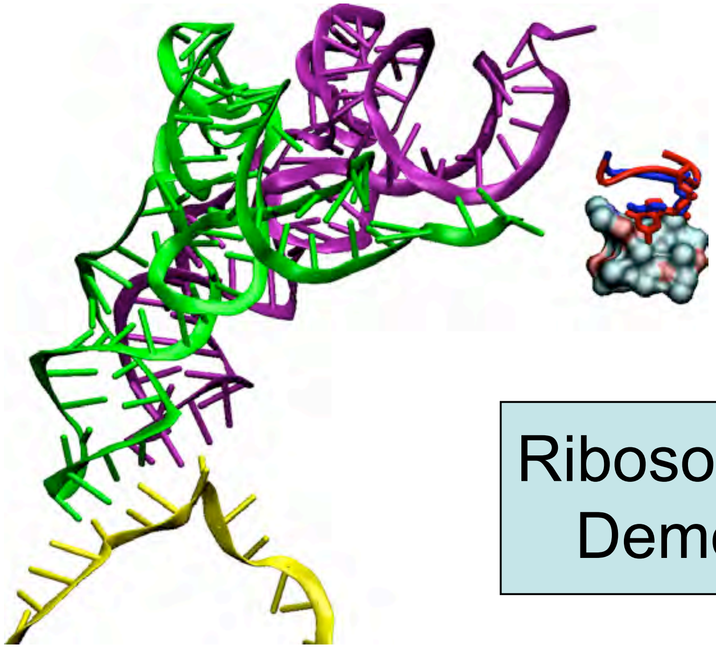
```
proc drawGenomicContextOfAlignment {outputFilename alignment contextDistance scaling genomeDirectory} {

    foreach sequence $alignment {

        # Make sure we have the GI number for this sequence.
        set giNumber [::SeqData::getSourceData $sequence "gi"]

        # Make sure we can tell which genome this sequence is from.
        set taxonomy [join [::SeqData::getLineage $sequence 1 0 1] ","]
        if {![info exists genomeTaxonomyMap($taxonomy)]} {
            error "ERROR) Unknown genome for sequence [::SeqData::getName $sequence]: $taxonomy"
        }

        # Go through each of the genome context files for the genome.
        set foundGene 0
        foreach genomeName $genomeTaxonomyMap($taxonomy) {
            ...
        }
    }

    # Draw the genomic context.
    drawMultipleGenomicContext $outputFilename $alignment $geneFiles $genePositions $geneStrands $contextDistance
}
```



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

New Tools in VMD/MultiSeq

Protein / RNA
Sequence Data

SwissProt DB (400K),
Greengenes RNA (100K)
Signatures, Zoom

Metadata Information,
Clustal &
Phylogenetic Trees

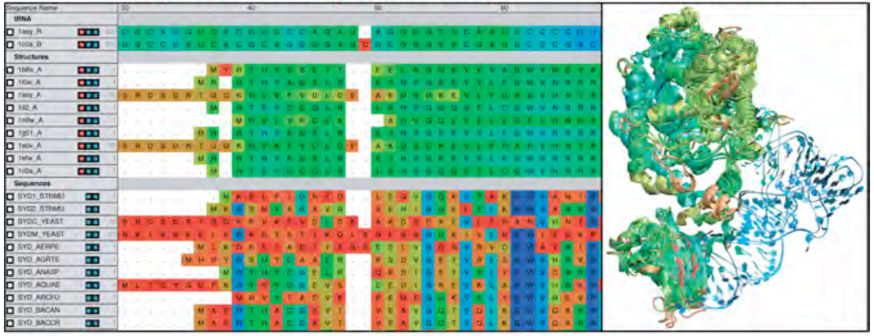
RAXml Trees,
Genomic Content,
Temperature DB

Blast & PsiBlast

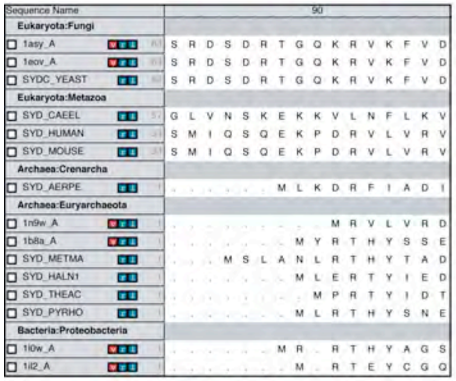
Sequence Editor

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

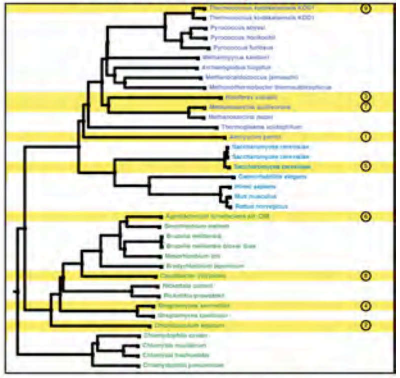
Synchronization between 1D and 3D views



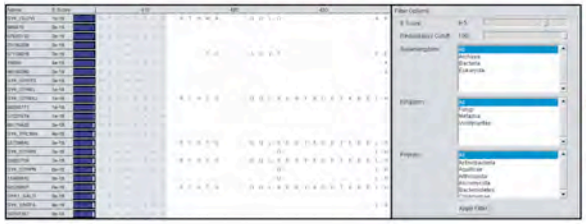
Group data by taxonomic classification



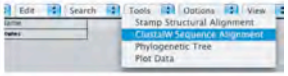
View sequence or structure phylogenies and eliminate redundancy with QR



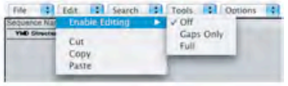
Import data directly from BLAST databases



Align sequences with Clustal



Sequence Editor: Manually adjust alignments or sequences



Sequence /Structure
Alignment

Protein & RNA
secondary structure

QR non-redundant
seq / str sets

Cluster analysis
/ Bioinformatics
scripting

Tutorials
MultiSeq/AARS
EF-Tu/Ribosome

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)

VMD/MultiSeq Tutorials

1. Analysis aquaporins in VMD/MultiSeq
2. Evolution of Translation: AARS:tRNA
3. Evolution of Translation: EF-Tu:tRNA
4. Evolution of Translation: Ribosome
5. Participant's project

Acknowledgements

<u>Collaborators (UIUC)</u>

Carl Woese

Ido Golding

Martin Gruebele

Taekjip Ha

Susan Martinis

Gary Olsen

Mike Heath

Wen-Mei Hwu

<u>Students/Postdocs</u>

Alexis Black

Ke Chen

John Eargle

Li Li

Damien Mathew

Jonathan Montoya

Elijah Roberts

Leonardo Sepulveda

Anurag Sethi

Dan Wright

NSF, NIH, BI, DOE & IGB

Thanks to John Stone, Michael Bach, Jim Philips at the NIH Resource for Macromolecular Modeling and Bioinformatics, Beckman Institute, UIUC