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

Part III – Towards in silico Cells: Simulating processes in entire cells

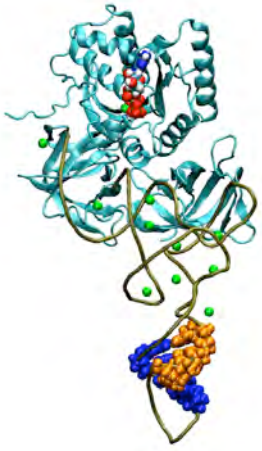
Zaida (Zan) Luthey-Schulten
Dept. Chemistry, Physics, Beckman Institute, Biophysics,
Institute of Genomics Biology

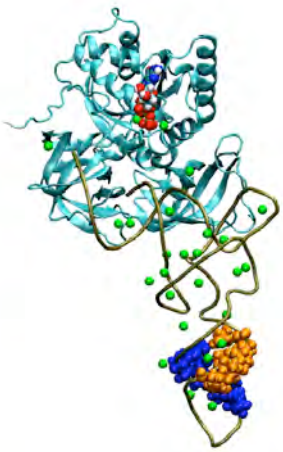
NIH Resource Macromolecular Modeling and Bioinformatics
Atlanta Workshop 2011

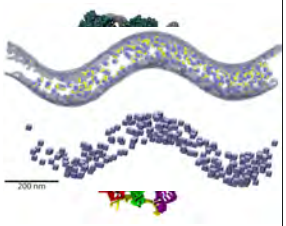
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Protein:RNA Complexes in Translation

Evolution, Dynamics, Analysis







Proteins/RNA
Polyribosomes
Ribosome

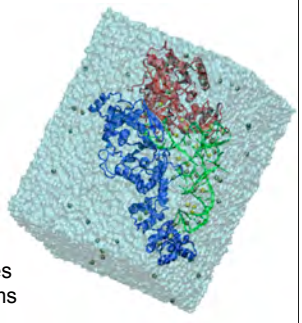
"Evolution SepRS/CysRS", ***PNAS***,*2005*
"Dynamic Signaling Network", ***PNAS*** 2009
"Exit Strategy Charged tRNA" ***JMB*** 2010

"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*

Molecular Dynamics Simulations

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)

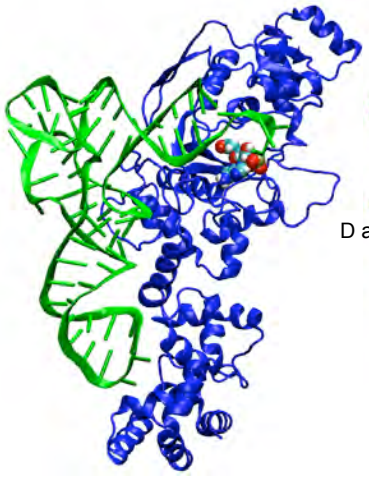
GluRS
Glu-tRNA^{Glu}
EF-Tu
GTP
Ions: Mg^{2+}, K^{+}
H_2O: ~27,000 molecules
System: ~130,000 atoms

System Perturbations

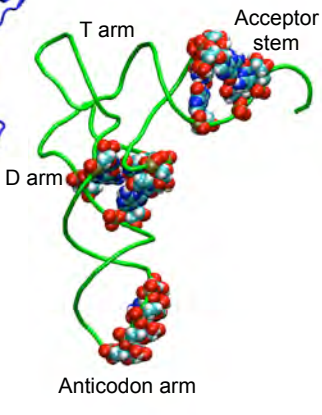
Deprotonation/protonation of reactants (aa,NTs)
Changing rotamer states of residues (5) involved in long-lived salt bridges

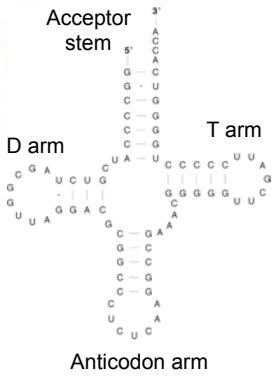
(1) Phillips, J.C. et al. *J. Comput Chem*, (2005); (2) MacKerell, A. et al. *Biopolymers* (2001);
(4) Case, D. et al. *J. Comput. Chem.* (2005); (4) Eargle, J. et al. JMB 2010, FEBS Let. 2010;
(5) Dunbrack Jr. and Cohen. *Protein Sci.* (1997)

Charging tRNA through allosteric signaling



Identity Elements

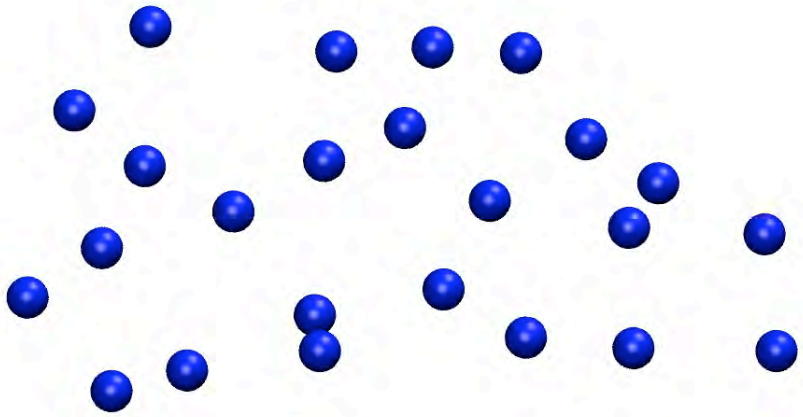




Overall Charging Reaction

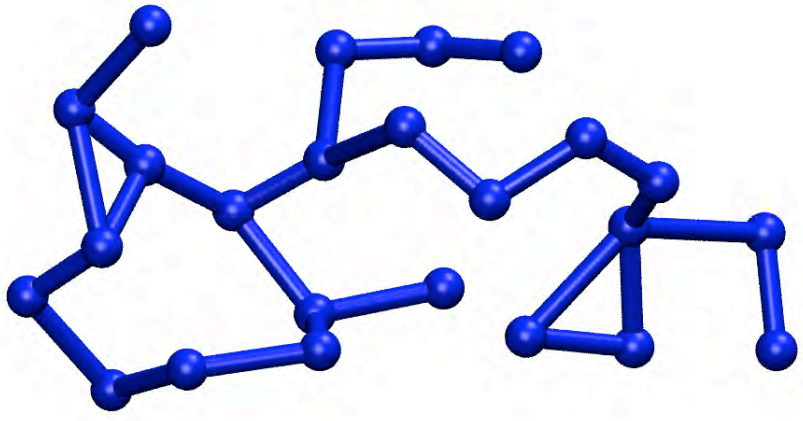
$$ATP + Glu + tRNA \xrightarrow{GluRS} Glu\text{-}tRNA + AMP + PP_i$$

How to Construct a Network?



Nodes - defined at C_a (protein) and P (nucleotide) atoms

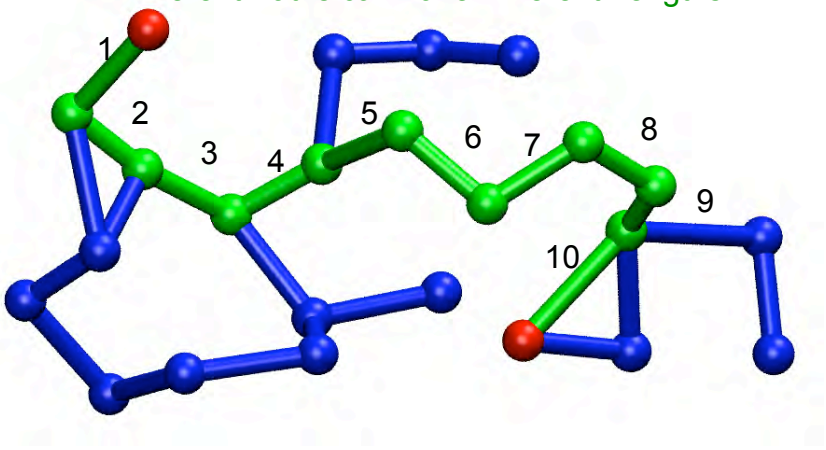
How to Construct a Network?



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

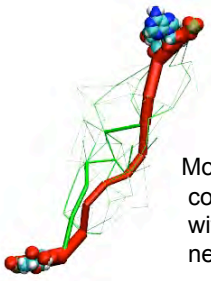
Different Paths can have Different Lengths



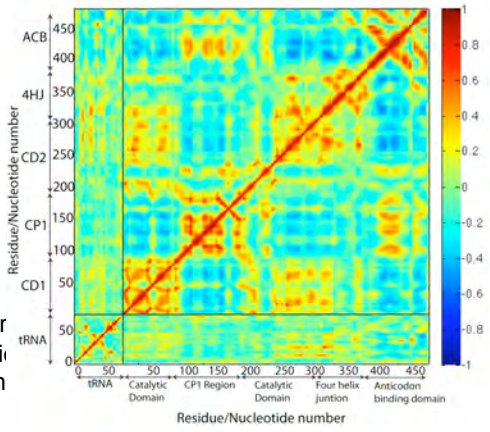
Path length = 10 (*unweighted*)

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

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

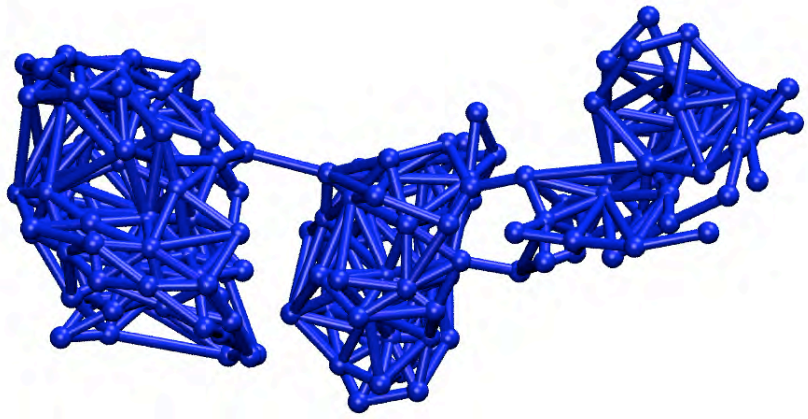


Modular str
communiti
with weigh
network

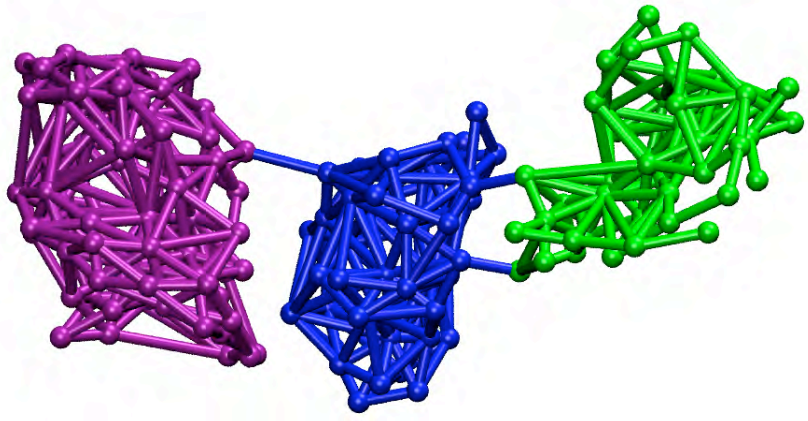


- MD simulations run with NAMD
- NTP ensemble
- Neutralized with Mg^{2+} and K^{+}
- C_{ij} values calculated over a 16-ns window

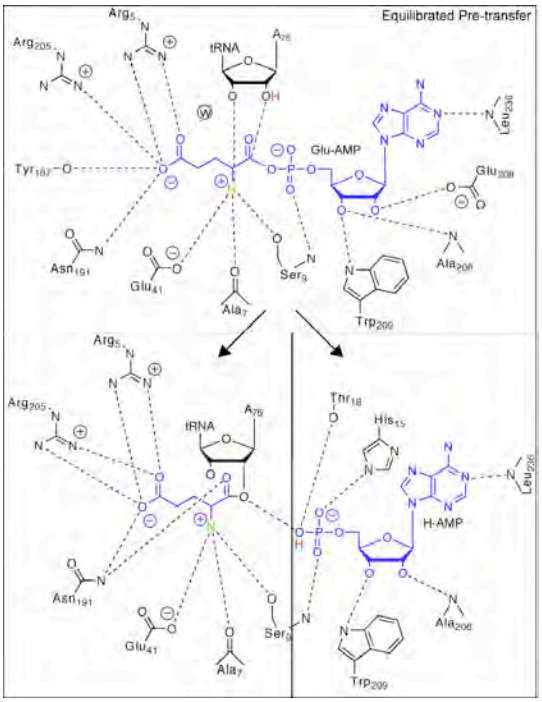
Nodes Cluster Together in Modules called Communities



Nodes Cluster Together in Modules called Communities



Communities are subnetworks with many intracommunity edges but few intercommunity edges.

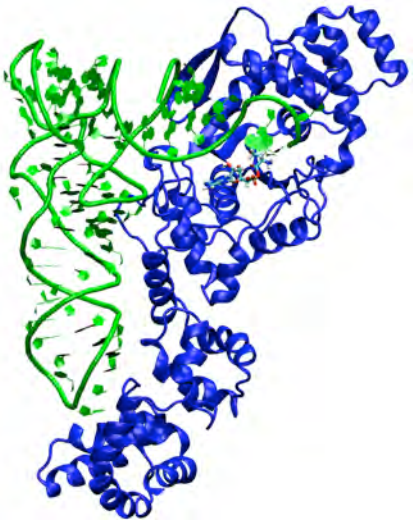


Reaction Mechanism for the Transfer of Glu to $tRNA^{Glu}$

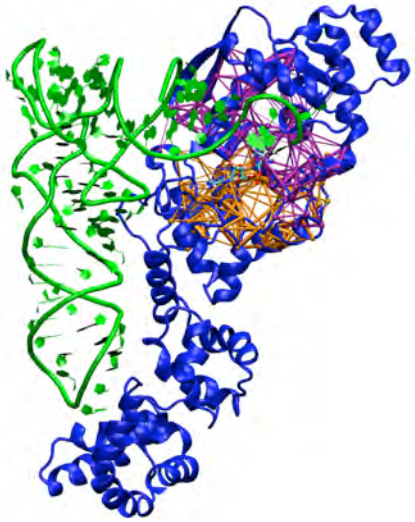
Perona JJ, Rould MA, Steitz TA
Biochemistry 1993

Black A, Eargle J, Sethi A,
Luthey-Schulten Z. *JMB* 2010
100s ns MD simulations

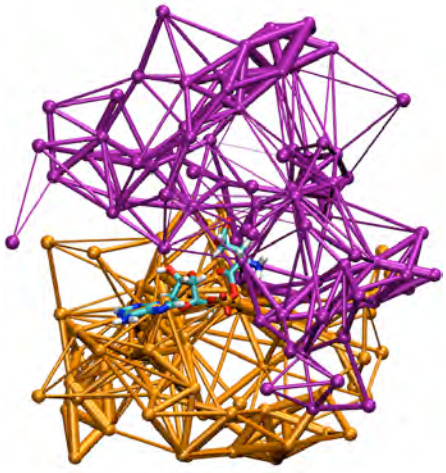
Community Structure Reveals Function in GluRS



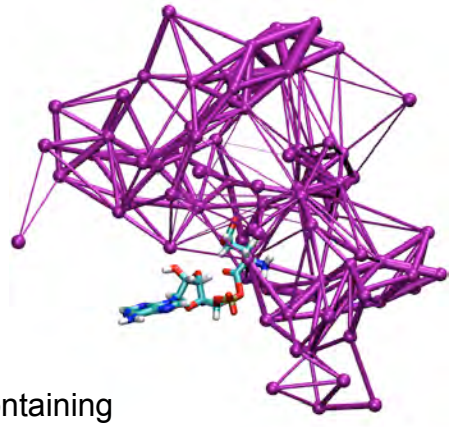
Community Structure Reveals Function in GluRS



Community Structure Reveals Function in GluRS



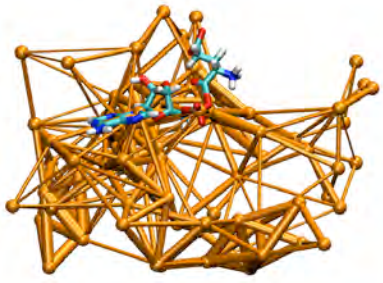
Community Structure Reveals Function in GluRS



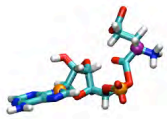
community containing amino acid moiety

Community Structure Reveals Function in GluRS

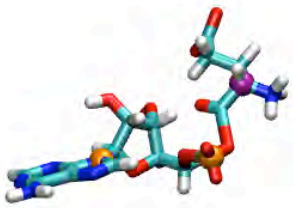
community containing AMP moiety



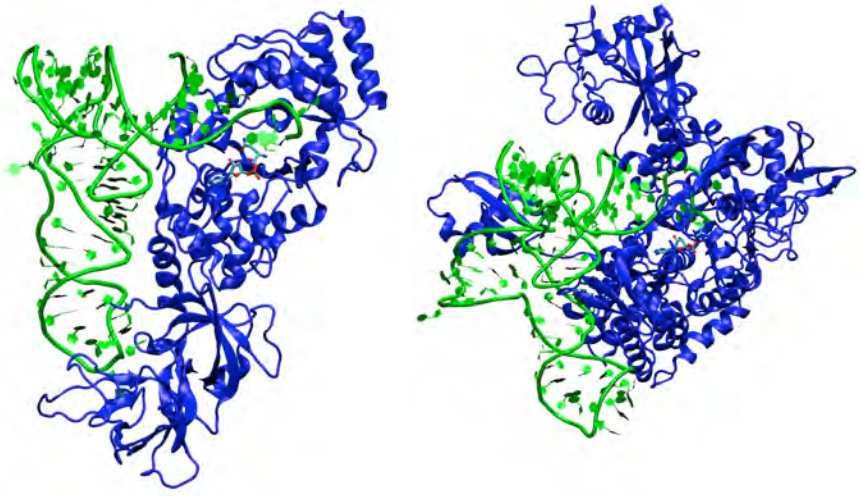
Community Structure Reveals Function in GluRS



Community Structure Reveals Function in GluRS

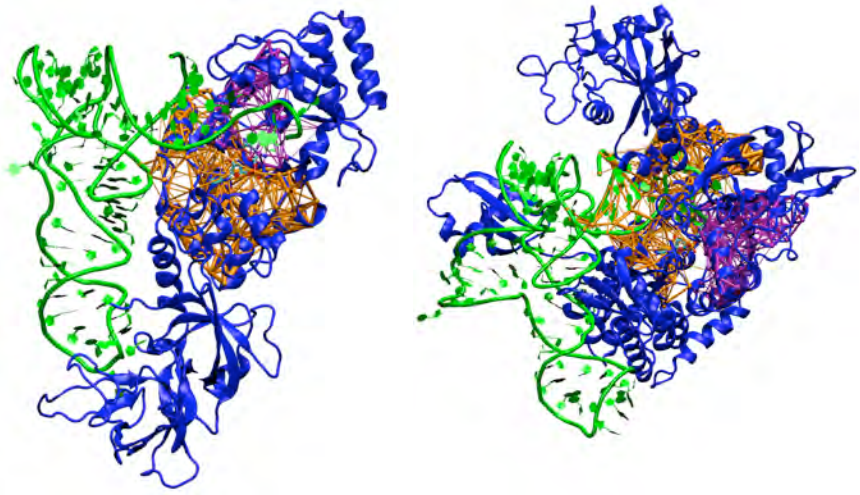


Similar Ligand Separation in GlnRS and LeuRS



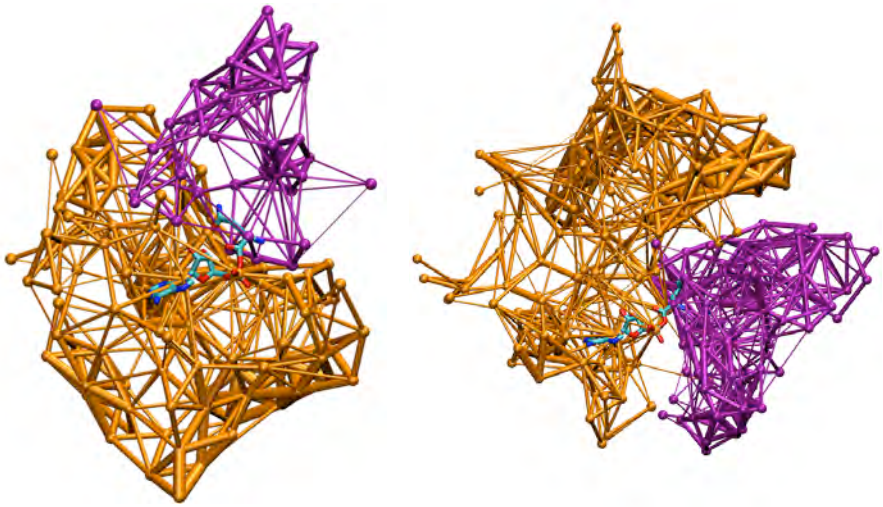
GlnRS LeuRS

Similar Ligand Separation in GlnRS and LeuRS



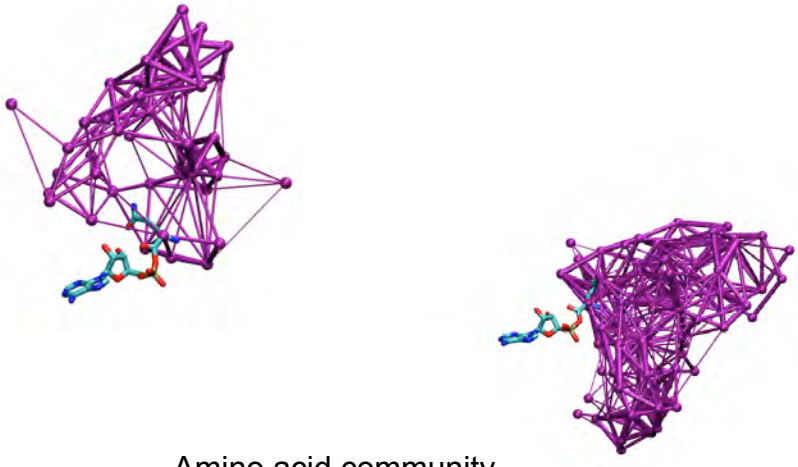
GlnRS LeuRS

Similar Ligand Separation in GlnRS and LeuRS



GlnRS LeuRS

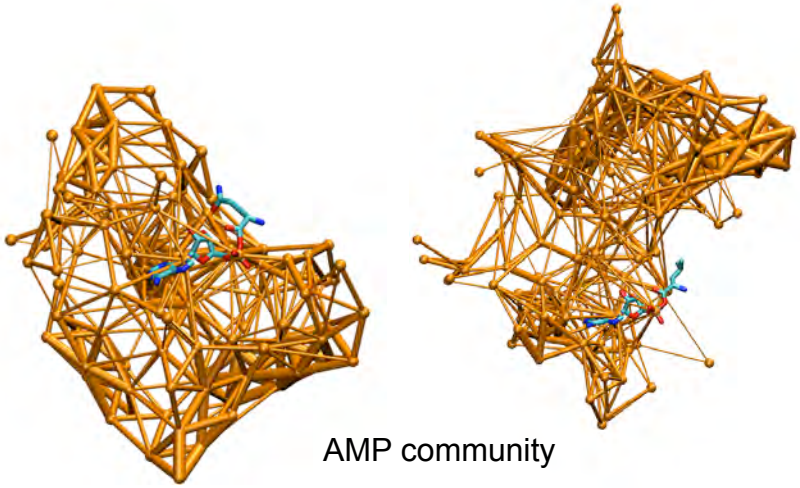
Similar Ligand Separation in GlnRS and LeuRS



Amino acid community

GlnRS LeuRS

Similar Ligand Separation in GlnRS and LeuRS



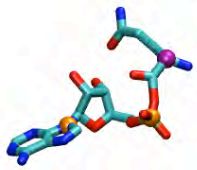
AMP community

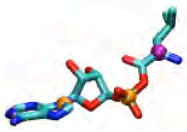
GlnRS LeuRS

Similar Ligand Separation in GlnRS and LeuRS

GlnRS LeuRS

Similar Ligand Separation in GlnRS and LeuRS

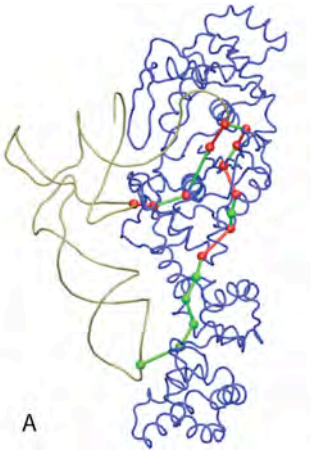




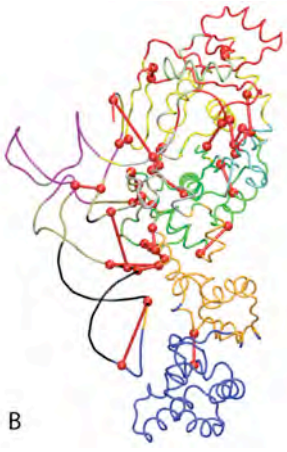
GlnRS

LeuRS

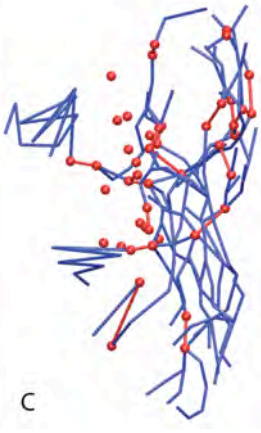
Dynamical Networks, Conservation, and Betweeness



Optimal signal pathways: U13, U35 to A76



Critical(conserved) nodes connecting communities



Betweeness routes - highest density pair optimal paths

Changing Networks: tRNA Migration from GluRS to EF-Tu

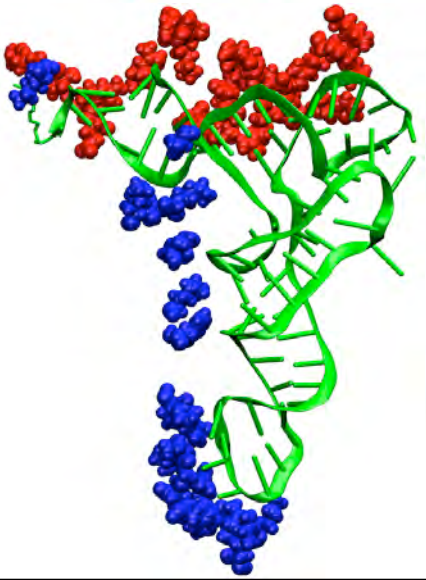
>2-fold increase in k_{cat} for CysRS:tRNA:EF-Tu

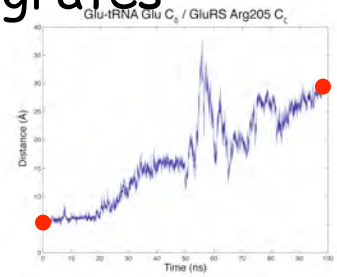
Zhang C, Perona J, Kang R, Francklyn C, Hou Y. *JMB* 2006

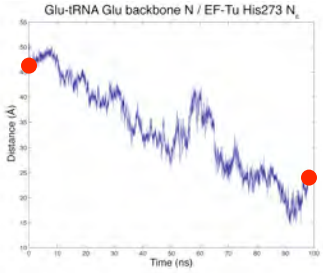
8-fold increase in k_{cat} for LeuRS:tRNA:EF-Tu

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

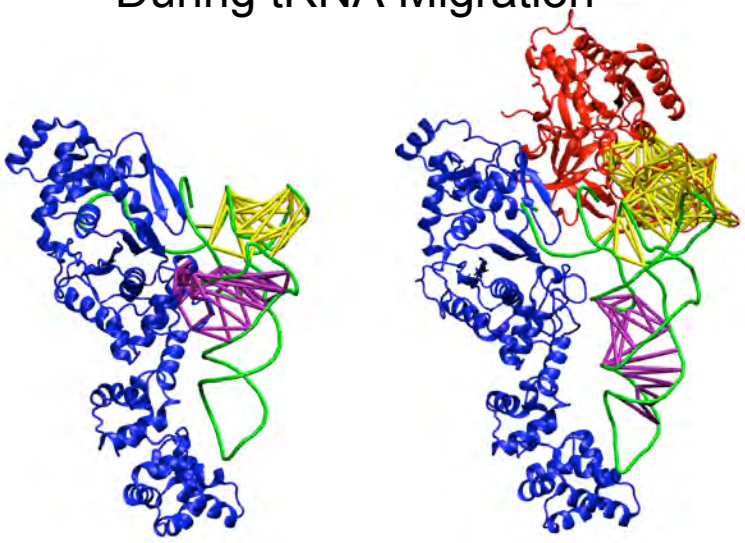
Change in Protein:RNA Contacts as tRNA Migrates





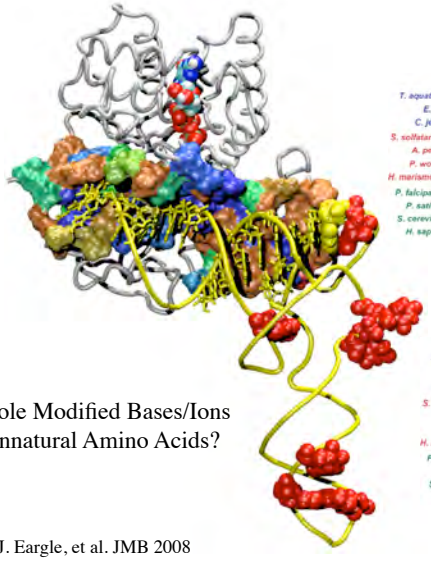


Change in Protein:RNA Contacts During tRNA Migration

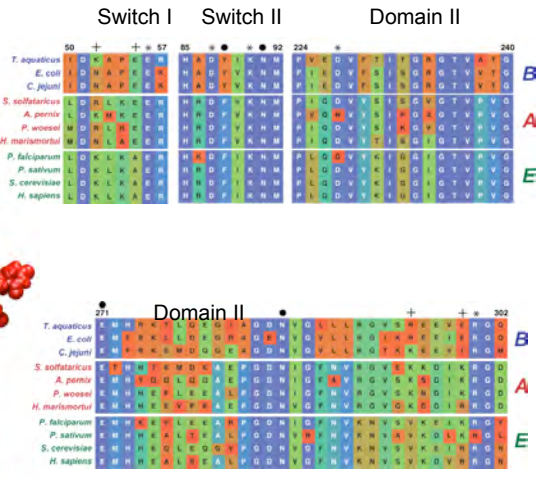


purple - D arm community yellow - T arm community

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



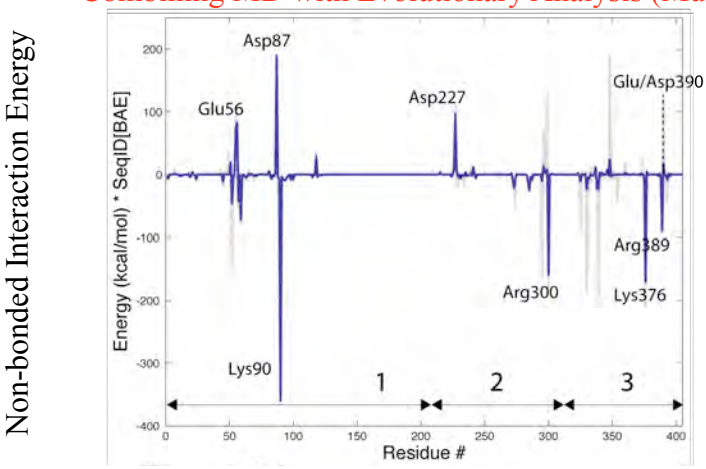
Role Modified Bases/Ions
Unnatural Amino Acids?



J. Eargle, et al. JMB 2008

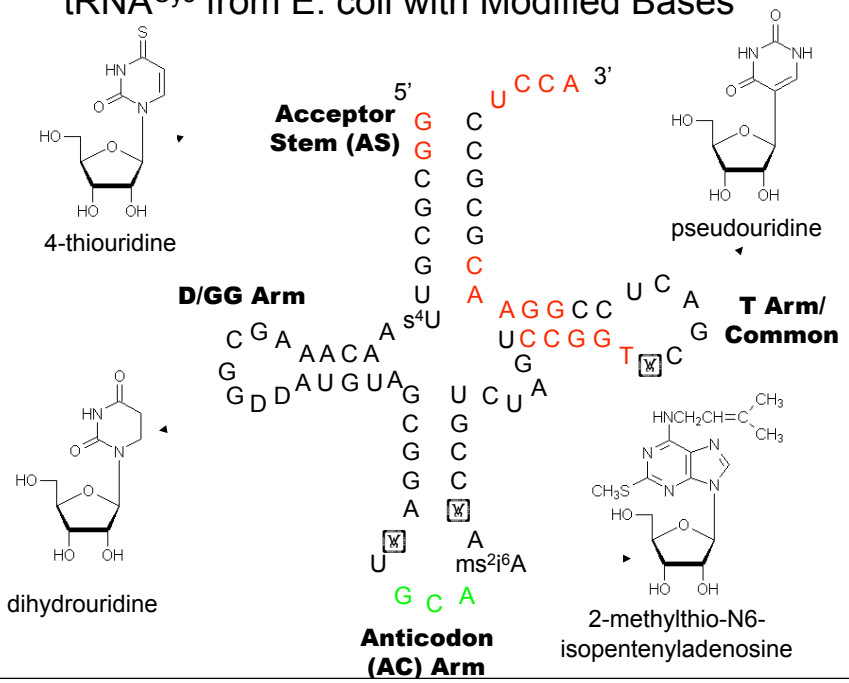
Dynamical Recognition EF-Tu/tRNA (E.coli)

Combining MD with Evolutionary Analysis (MultiSeq)



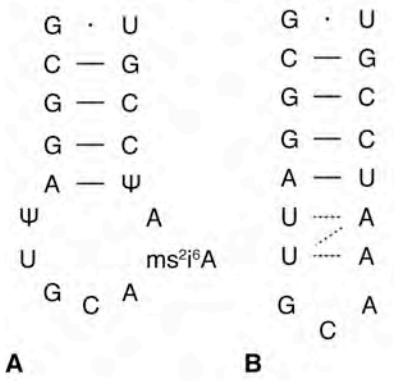
J. Eargle, A. Sethi, A. Black, L. Trabuco & Z. Luthey-Schulten. JMB "Dynamics of Recognition in EF-Tu/tRNA Complex" (2008)

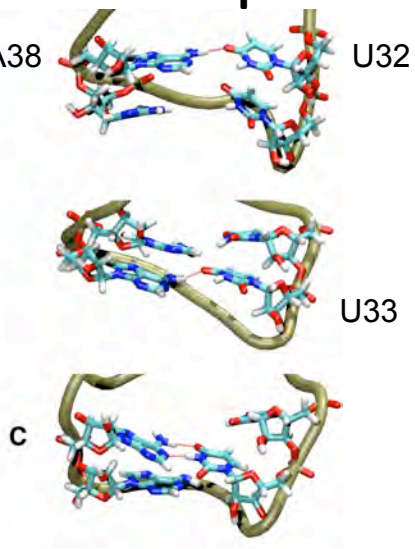
tRNA^{Cys} from E. coli with Modified Bases



Effects of Modified Bases on Dynamics of AC Stem/Loop

MD simulations of unmodified tRNA:Cys

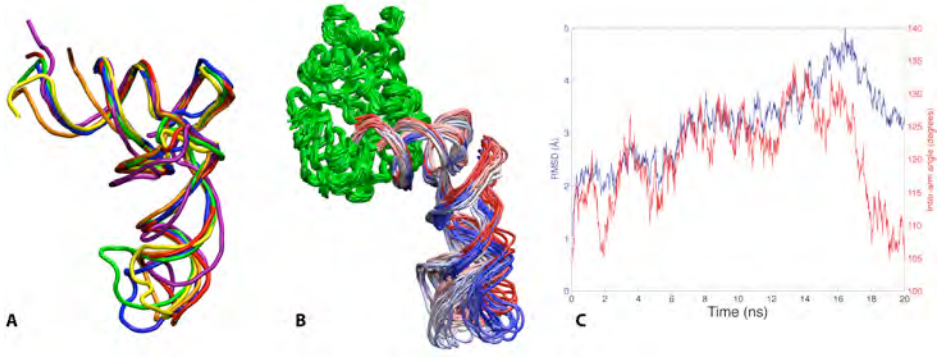




Similar NMR structures for unmodified tRNA-Phe
Nikonowicz, *JMB* 2002, *personal communication*

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

Flexibility in tRNA structures observed in crystallography and simulations



X-ray 5 tRNAs

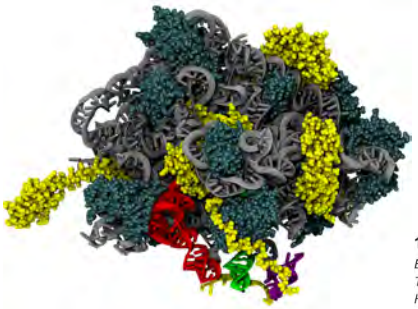
MD EF-Tu:tRNA:cys

MD: RMSD and Interarm Angle in tRNA:cys

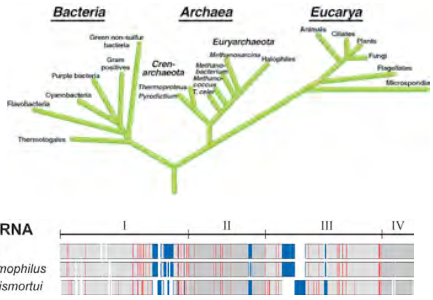
JMB 2008 & FEBS Let. 2010

Molecular Signatures in Ribosome Evolution

Archaeal



Universal Phylogenetic Tree

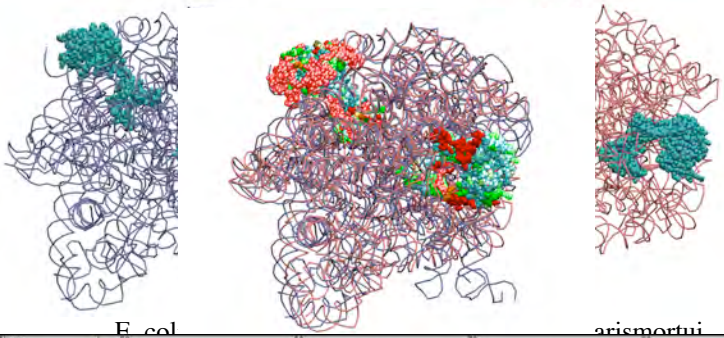


Dynamical function of ribosomal signatures: idiosyncrasies in ribosomal RNA and/or proteins characteristic of the domains of life

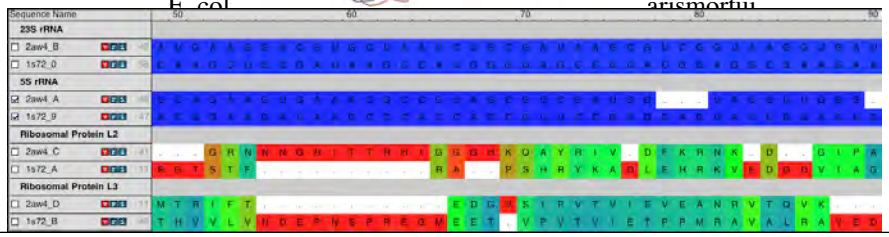
Roberts, Sethi, Montoya, Woese, Luthey-Schulten (2008) ***PNAS*** 105:13953
Chen, Eargle, Sarkar, Gruebele, Luthey-Schulten (2010) ***Biophysical Journal***

Structural Overlaps with STAMP

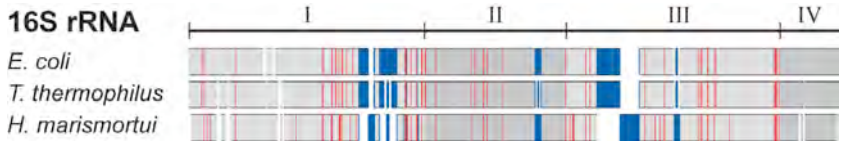
Comparing ribosomal large subunit with r-proteins L2 and L3
180,000 atoms in 4 rRNAs and 58 proteins



E. col arismortui

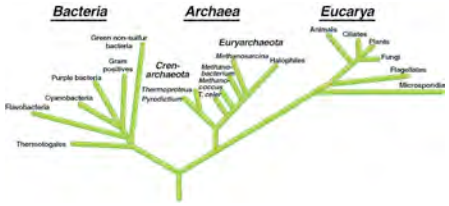


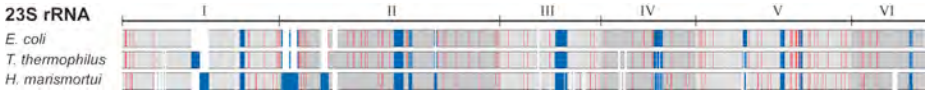
Sequence and **Structure** Signatures in Ribosomal RNA



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

Signature analysis

- Signatures are sequence and/or structural features that are characteristic of a domain of life*.
- Identify the sequence signatures in the 16S rRNA and proteins using MultiSeq
- MultiSeq has a coloring based on signatures, but use Scripts to collect quantitative data about the signatures: position, composition.

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

Signature analysis script

```
proc get_taxa_signatures {args} {

    initializeMultiSeqEnvironment

    # Load the alignment.
    set sequenceIDs [::SeqData::Fasta::loadSequences $alignmentFilename]

    # Find the reference species.
    set refSequenceIDs {}
    foreach species $referenceSpecies {
        set found 0
        foreach sequenceID $sequenceIDs {
            if {[string toupper [::SeqData::getScientificName $sequenceID]] == [string toupper $species]} {
                lappend refSequenceIDs $sequenceID
                set found 1
                break
            }
        }
    }

   # Group the sequence ids.
    array set groupData {names {}}
    foreach sequenceID $sequenceIDs {

        set group ""
        if {$rankName == "domain"} {
            set group [::SeqData::getDomainOfLife $sequenceID]
        } else {
            set group [::SeqData::getLineageRank $sequenceID $rankName]
        }
        if {![info exists groupData($group)]} {
            lappend groupData(names) $group
            set groupData($group) {}
        }
        lappend groupData($group) $sequenceID
    }
```

Signature analysis script (cont)

```
    # Calculate the signatures for the groups.
    set signatures [::SeqEdit::Metric::Signatures::calculateSignatures $sequenceGroups $groupConsensusCutoff \
                                $otherGroupMaxCutoff $otherGroupMaxGapFraction $maxConservedBlockDistance $minConse

    # Print the signatures.
    puts ""
    puts "Ordered Signatures"
    printSignatureHeader $groups $refSequenceIDs
    for {set position 0} {$position < [llength [lindex $signatures 0]]} {incr position} {

        # Get the signature for each of the groups.
        set isSignature 0
        for {set groupIndex 0} {$groupIndex < [llength $groups]} {incr groupIndex} {
            set groupSignature [lindex [lindex $signatures $groupIndex] $position]
            if {[string length $groupSignature] == 1} {
                set isSignature 1
            }
        }

        if {$isSignature} {
            printSignatureLine $position $groups $refSequenceIDs $signatures $sequenceGroups
        }
    }
}
```

Signature analysis script (cont)

```
proc getCompositionString {sequenceIDs position} {

    # Go through each group to count the elements.
    array set counts {values {}}
    foreach sequenceID $sequenceIDs {
        set element [::SeqData::getElement $sequenceID $position]
        if {![info exists counts($element)]} {
            lappend counts(values) $element
            set counts($element) 1
        } else {
            incr counts($element)
        }
    }

    set ret ""
    foreach element $counts(values) {
        set percentage [expr round(100.0*double($counts($element))/double([llength $sequenceIDs]))]
        if {$percentage > 5} {
            if {$ret != ""} {
                append ret "/"
            }
            append ret "${element}($percentage\%)"
        }
    }

    return $ret
}
```

Signature script output

Ordered Signatures

Pos	E._coli	T._therm	H._maris	Archaea	Bacteria
35	33	33	27	U(28%)/C(72%)	*A(96%)
42	39	39	33	*U(100%)	*G(95%)
51	47	47	42	*G(97%)	C(85%)/U(7%)/G(8%)
59	53	53	49	*C(100%)	A(83%)/G(14%)
222	113	106	87	*C(100%)	*G(95%)
661	248	243	227	*G(97%)	C(75%)/G(9%)/U(11%)
749	314	309	293	*G(100%)	*C(97%)
753	317	312	296	*G(97%)	*U(91%)/G(7%)
777	338	333	317	*G(97%)	A(89%)/N(9%)
779	339	334	318	*G(97%)	*C(98%)
780	340	335	319	*C(94%)	*U(92%)/A(6%)
790	349	344	328	G(86%)/N(11%)	*A(92%)
791	350	345	329	C(67%)/U(19%)/N(11%)	*G(98%)
800	358	353	337	*G(100%)	*U(91%)/C(7%)
802	359	354	338	*C(97%)	G(59%)/U(9%)/C(9%)/A(20%)
804	361	356	340	*C(94%)	G(84%)/A(14%)
809	365	360	344	*A(92%)/U(8%)	*U(97%)
811	367	362	346	C(83%)/G(14%)	*U(97%)
823	377	372	356	C(89%)/U(11%)	*G(92%)
833	386	381	365	*G(94%)	C(89%)/U(6%)
840	393	388	372	G(83%)/C(14%)	*A(97%)
851	403	398	382	*A(100%)	C(89%)
920	447	441	416	C(31%)/U(67%)	*G(92%)
990	487	471	424	G(89%)/A(8%)	*A(91%)
1083	508	491	446	*A(97%)	*U(92%)
1085	510	493	448	*G(97%)	*A(97%)
1087	512	495	450	*C(92%)/U(8%)	*U(97%)
1089	514	497	452	*G(92%)/N(6%)	C(89%)/U(8%)
1101	523	506	461	*C(94%)/A(6%)	*A(96%)
1119	537	520	475	*C(100%)	*G(91%)/A(8%)
1121	539	522	477	*G(92%)/A(6%)	A(89%)/G(10%)
1134	549	532	487	*U(94%)	*C(99%)

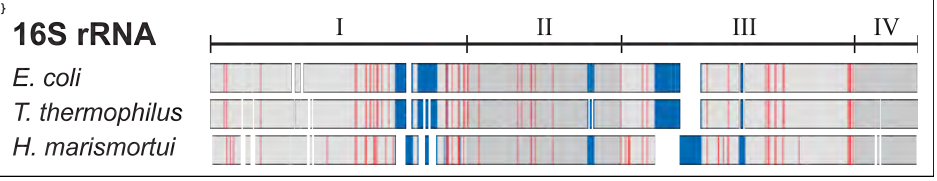
Draw signature figure

- Want to create a graphical representation of the signatures.
- Can use the TCL graphics object to draw figures.

```
proc draw_one_d_alignment {args} {

    # Load the alignment.
    set alignment [::SeqData::Fasta::loadSequences $alignmentFilename]
    set numberSequences [llength $alignment]
    set numberPositions [::SeqData::getSeqLength [lindex $alignment 0]]
    puts "Loaded $numberSequences sequences of length $numberPositions"

    # Create the drawing canvas.
    set canvasWidth [expr $draw(xBorder)+($numberPositions*$draw(positionWidth))+$draw(xBorder)]
    set canvasHeight [expr $draw(yBorder)+($numberSequences*$draw(sequenceHeight))+(($numberSequences-1)*\

                    $draw(sequenceSpacing))+$draw(yBorder)]
    set g [canvas .drawing -width $canvasWidth -height $canvasHeight]

    # Draw the alignment.
    for {set i 0} {$i < $numberSequences} {incr i} {
        set sequenceID [lindex $alignment $i]
        set y1 [expr $draw(yBorder)+($i*($draw(sequenceHeight)+$draw(sequenceSpacing)))]
        set y2 [expr $y1+$draw(sequenceHeight)]
        drawSequencePositions $g $draw(xBorder) $y1 $y2 $sequenceID 0 [expr $numberPositions-1] \
                                                                        $draw(positionWidth) $draw(sequenceColor) $dra
    }

    # Save the canvas.
    $g postscript -x 0 -y 0 -width $canvasWidth -height $canvasHeight -file $outputFilename

    cleanupMultiSeqEnvironment
}
```

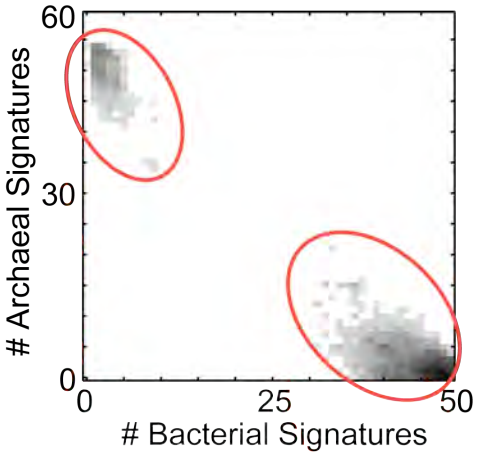
Draw signature figure (cont)

```
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
Distinct A & B Sequence Signatures

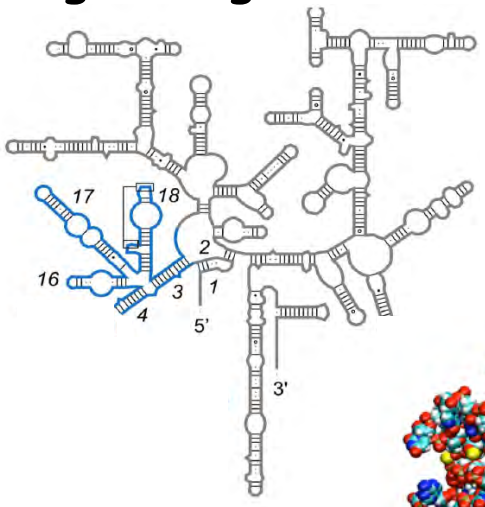
- Analysis of the ribosomal signatures in 90,000 new environmental samples shows that no "gray" area exists: a ribosome is either bacterial or archaeal in nature.
- Split across cluster: 10,000 sequence on each node

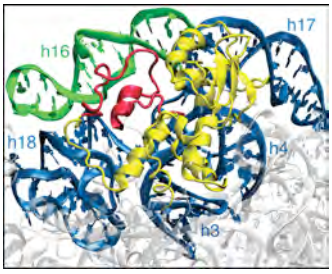


Roberts, et al. *PNAS* 2008

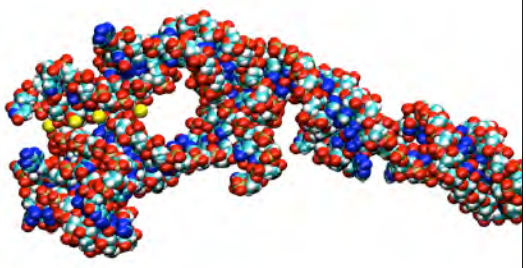
Data: "Greengenes", Lawrence Berkeley

Largest signature region in the SSU

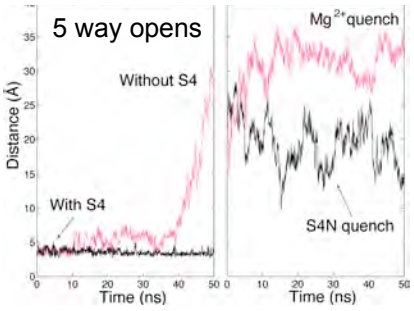


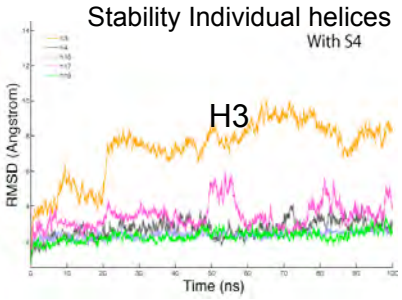


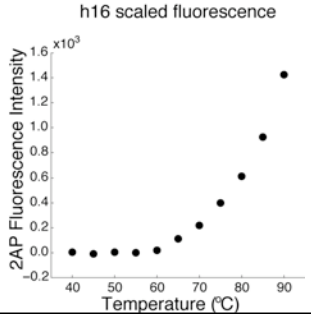
5-way helical junction rRNA with S4



Stability of 5-way junction helices with/without S4





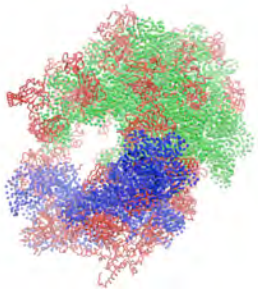


Model for Assembly ?

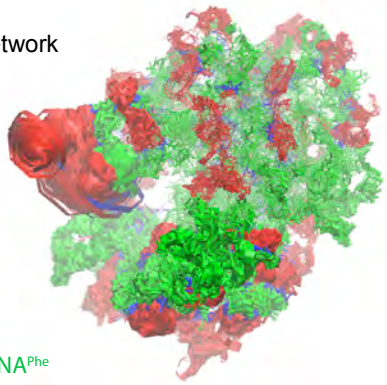
Center for the Physics of Living Cells

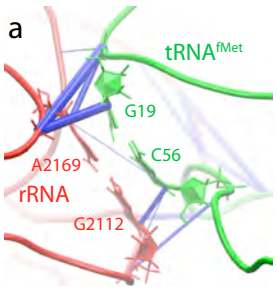
Collaboration with Gruebele, Ha, Woodson labs

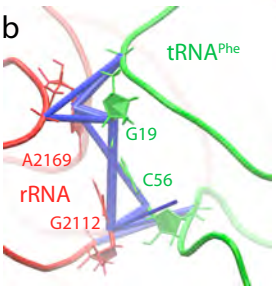
Signaling Networks in LSU &SSU



MD 140ns network







P/E site tRNA:rRNA network

L. Trabuco, E. Schreiner, J. Eargle et al. JMB 2010

New Tools in VMD/MultiSeq

Protein / RNA Sequence Data

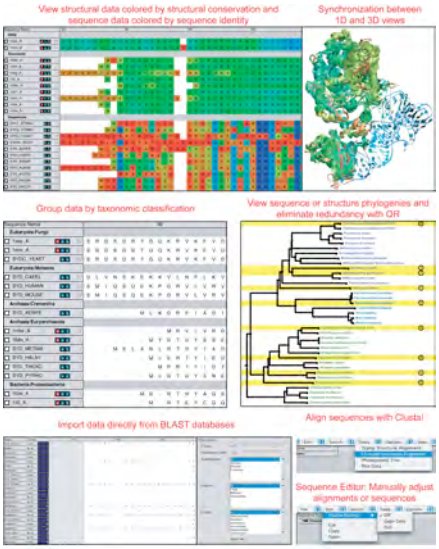
SwissProt DB (400K), Greengenes RNA (100K) Signatures, Zoom

Metadata Information, Clustal & Phylogenetic Trees

RAXml Trees, Genomic Content, Temperature DB

Blast & PsiBlast

Sequence Editor



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)

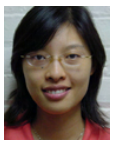
VMD/MultiSeq Tutorials

1. Evolution of Translation: AARS:tRNA
2. Evolution of Translation: EF-Tu:tRNA
3. Evolution of Translation: Ribosome
4. Dynamical Network Analysis
5. Participant's project

Part III Towards in silico cells: Simulating processes in entire bacterial cells

Acknowledgements

Collaborators
Carl Woese
Mike Heath



Martin Gruebele
Taekjip Ha
Sarah Woodson
Susan Martinis
Wen-Mei Hwu
John Stone
Julio Ortiz, W. Baumeister

Students/Postdocs
Alexis Black
Ke Chen
John Eargle
Li Li
Piyush Labhsetwar
Jonathan Montoya
Elijah Roberts
Anurag Sethi
Dan Wright
Mike Hallock






NIH, BI, NSF, DOE & Nvidia Cuda Center of Excellence

NIH Resource for Macromolecular Modeling and Bioinformatics, Beckman Institute, UIUC – John Stone, Kirby Vandivort