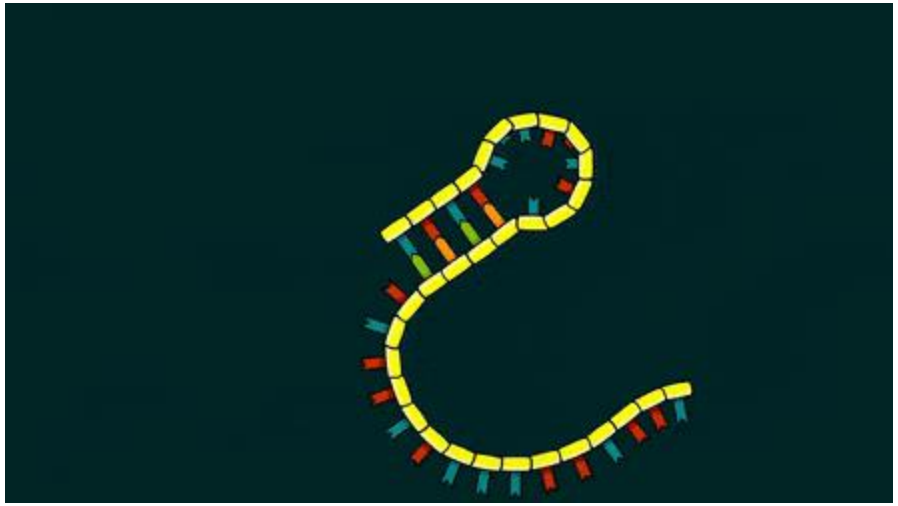
Predicting RNA Folding

Physics of Early Evolution

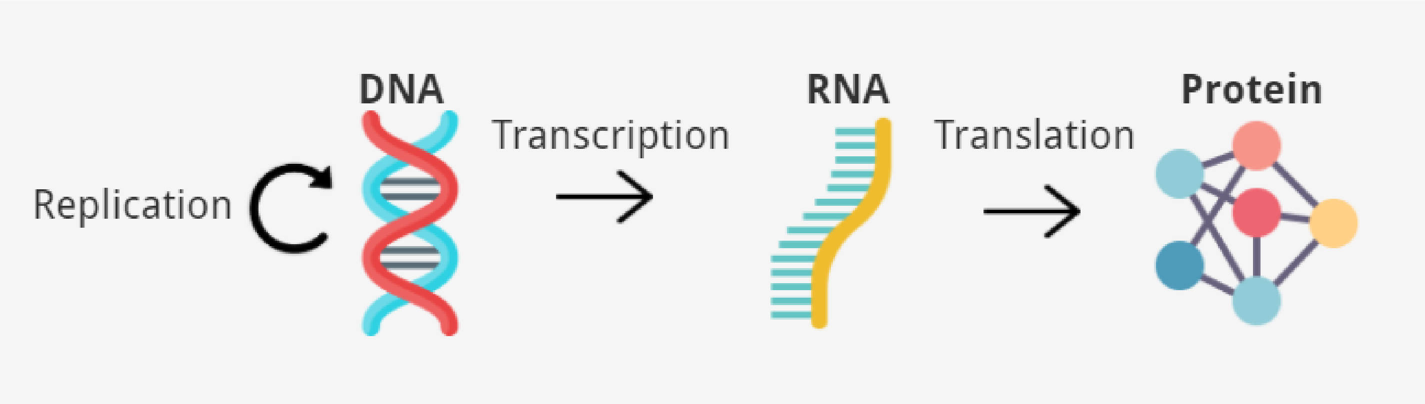
Girnar Goyal



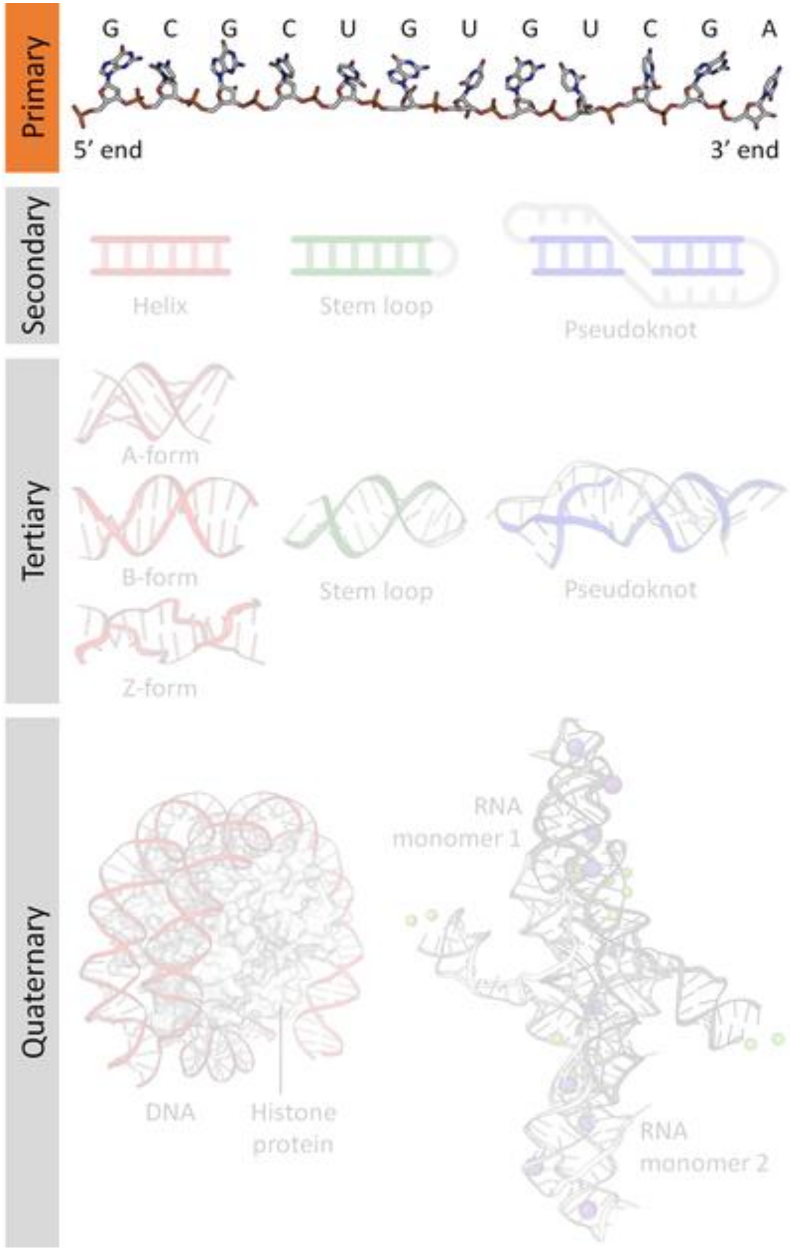
Contents

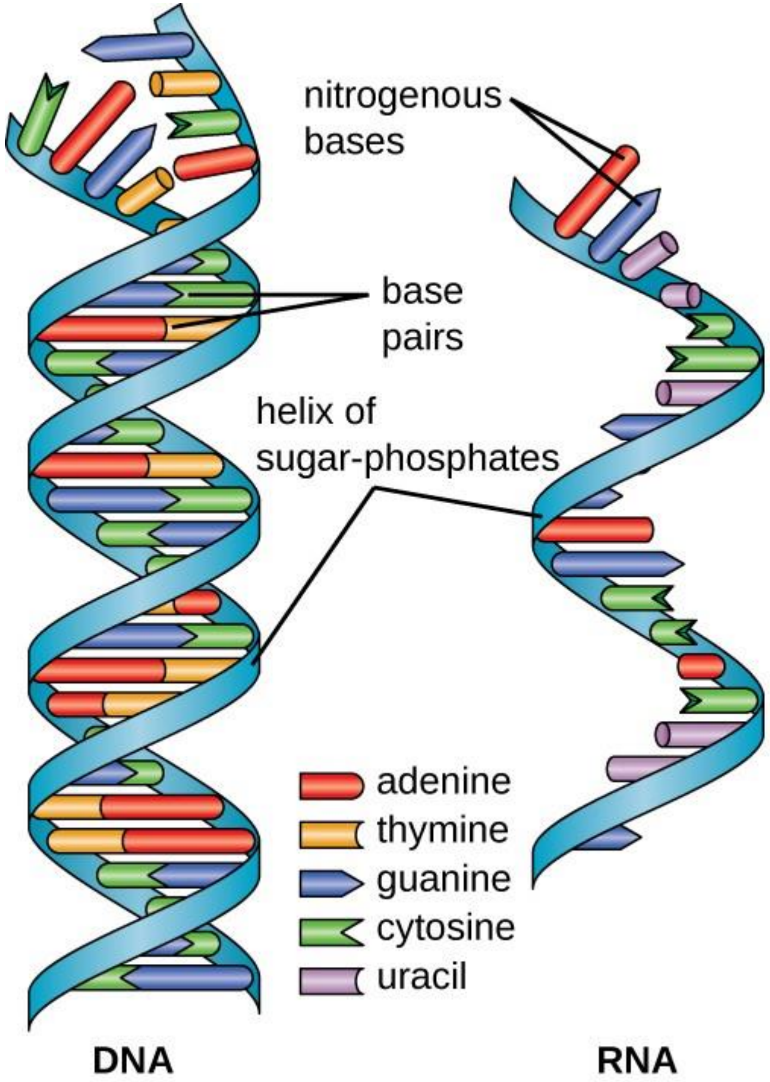
- RNA – Structure and Basics
- RNA World
- Early Observations in RNA Folding
- Middle RNA folding epoch – Folding Models
- Late RNA Folding Epoch – The Reconstitution Model

The Central Dogma

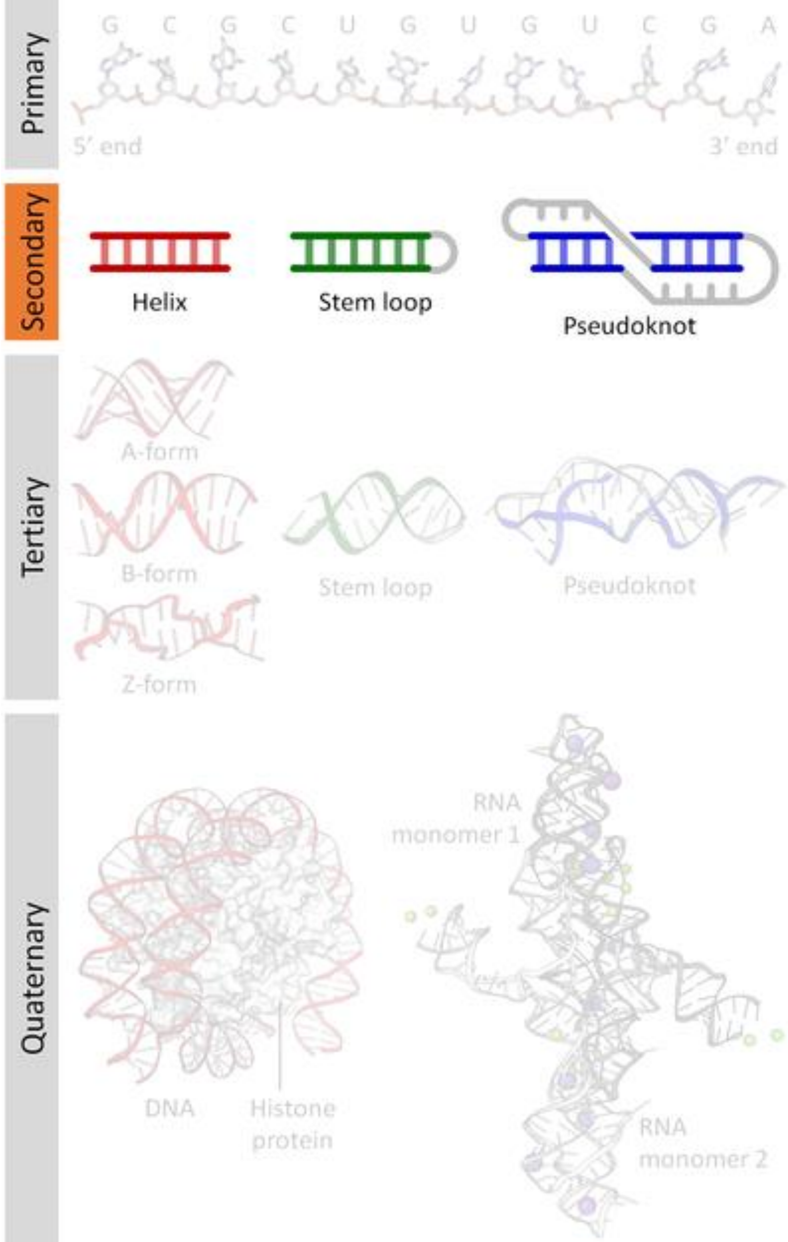


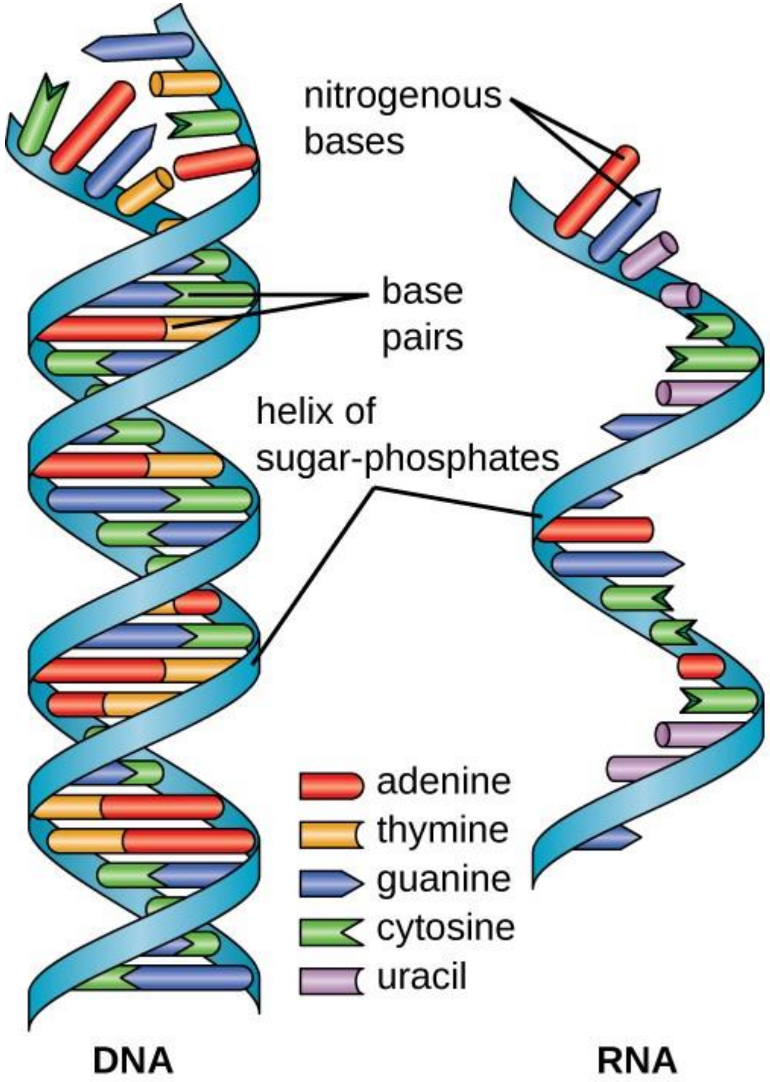
RNA Structure



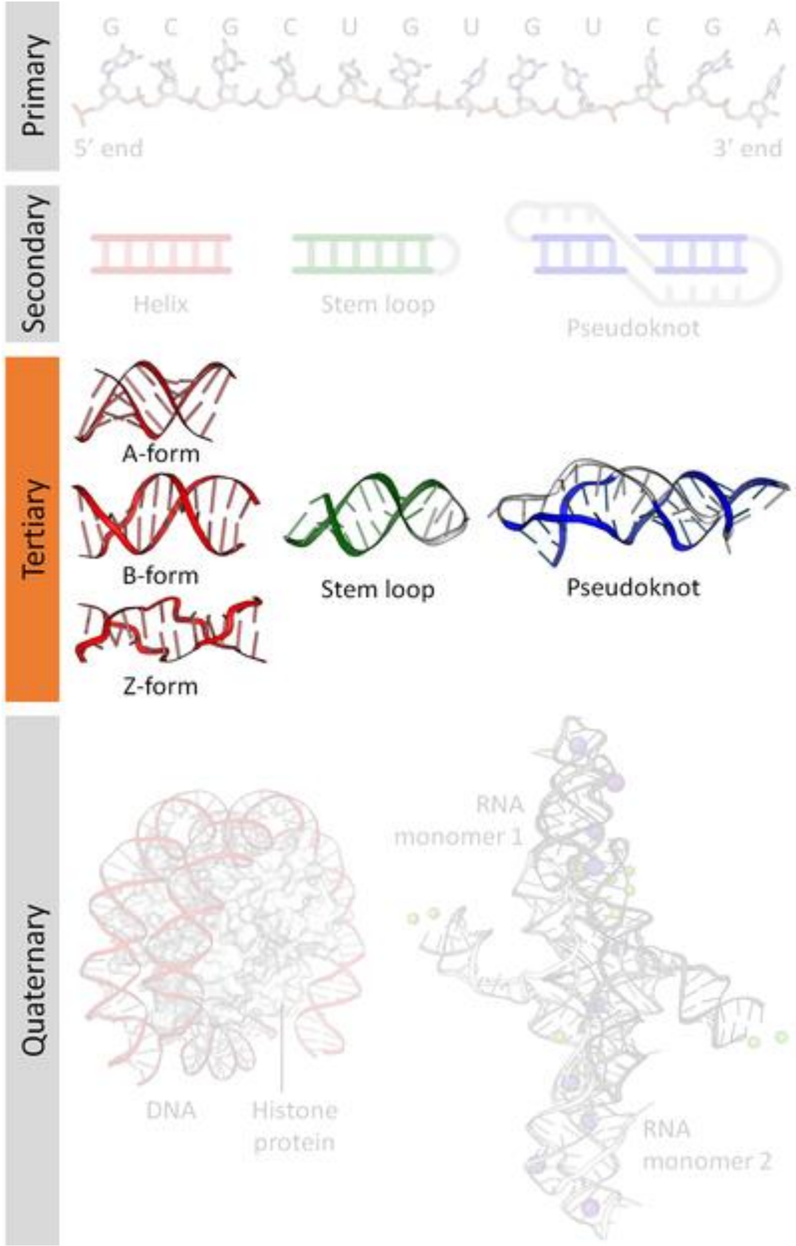


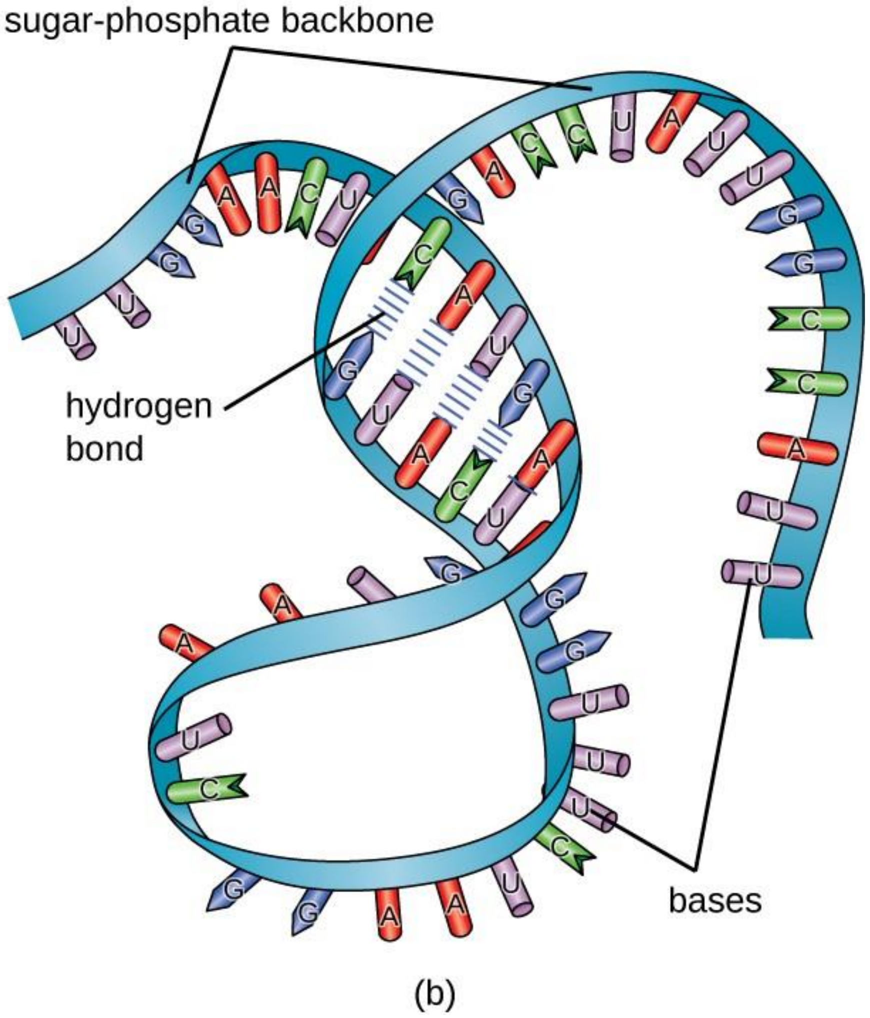
RNA Structure



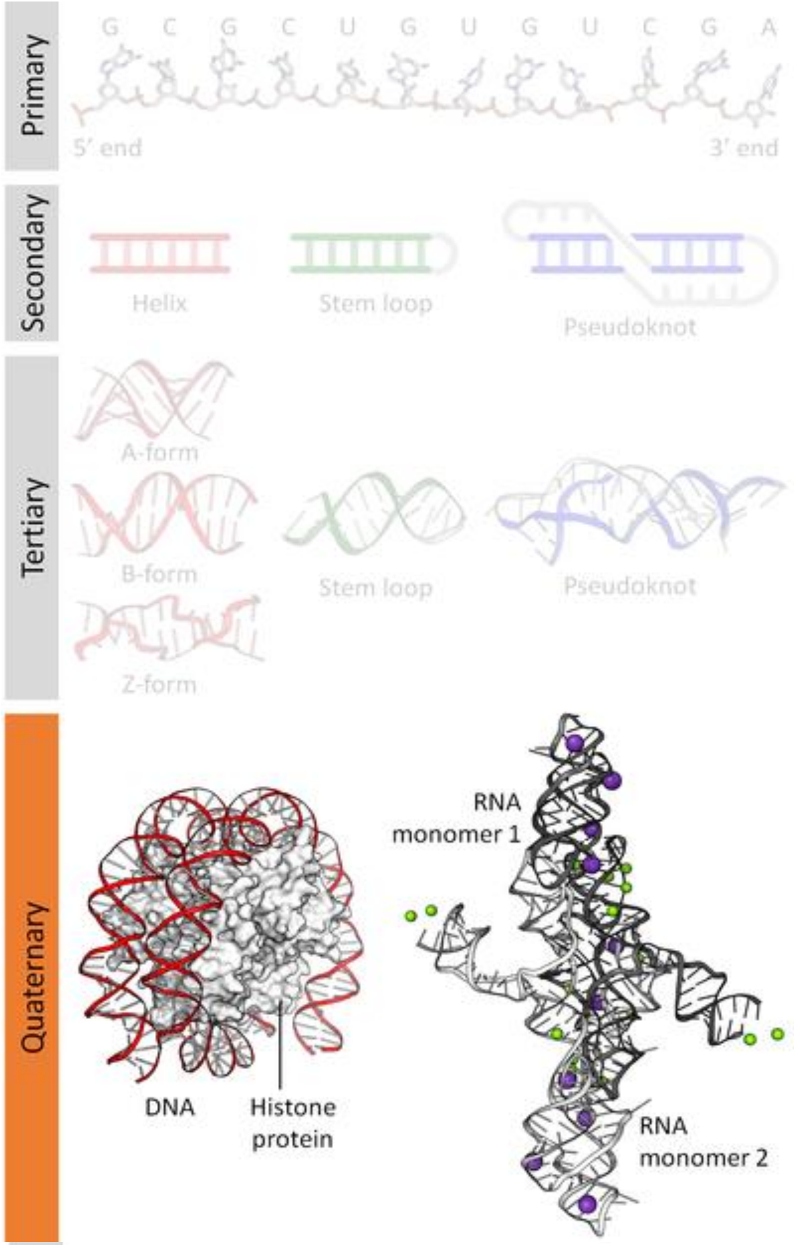


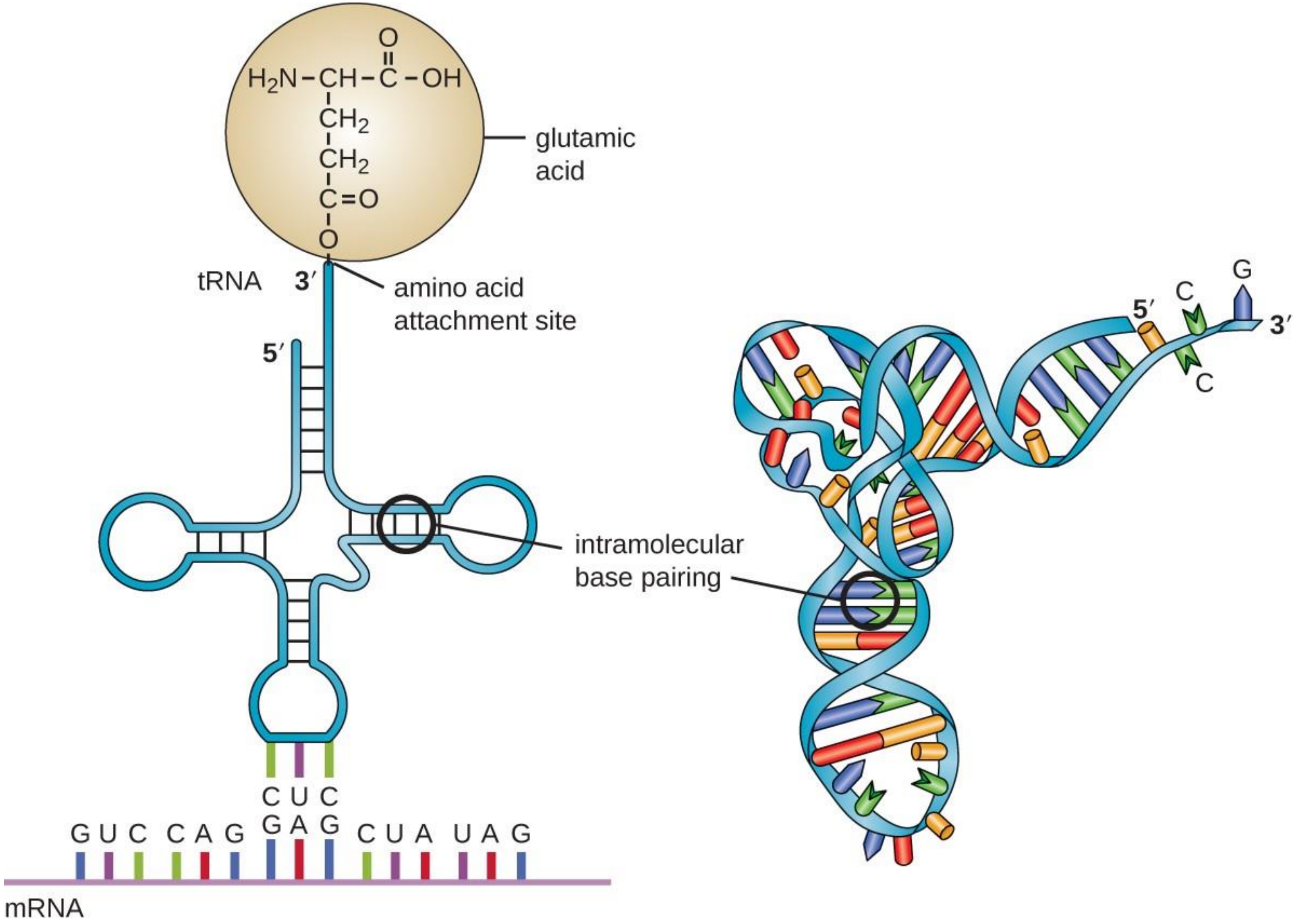
RNA Structure





RNA Structure





tRNA 2-dimensional structure

tRNA 3-dimensional structure

RNA World Hypothesis

> 4 billion years

1965	**Secondary structure Model**: Purification and sequence yeast tRNA, (1968 Nobel Prize)
1980	**Chemical Probes** to interrogate secondary and tertiary structure.
1988	Probing of RNA structures in Living cells using DMS
1989	**Hydroxyl Radical Footprinting:** Interactions that define exterior regions of RNA tertiary folds. Latham and Cech
2004	Computational RNA modeling with chemical probing data
2017	RT Stops and RT mutate methods. High throughput RNA structure probing

- Chicken-egg Problem (Protein-DNA).
- RNA enzymes (ribozymes) can catalyze chemical reactions.
- Replication with specificity. (A-U, G-C)
- Ability to form stable duplexes.
- Complex structure of tRNA resembled, a folded protein thus indicating the possibility of RNAs as catalysts.
- Ribosome’s catalytic core is made of RNA leads us to consider that this RNA core is possibly a vestige from an RNA world.

The Journey so far..

> 4 billion years

1965 **Secondary structure model**: Purification and sequence yeast tRNA, (**1968 Nobel Prize**)

1980 **Chemical Probes** to interrogate secondary and tertiary structure.

1988 Probing of RNA structures in Living cells using DMS

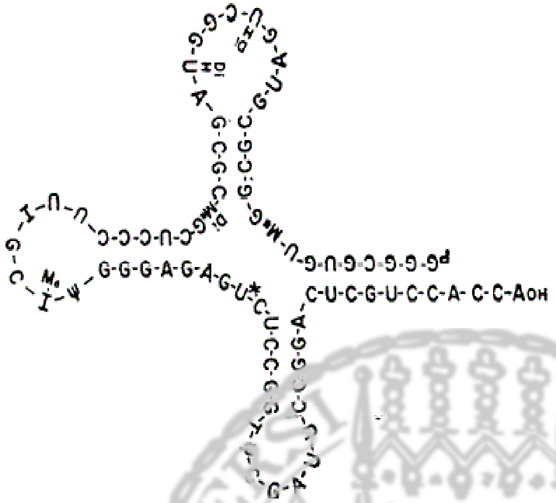
1989 **Hydroxyl Radical Footprinting:** Interactions that define exterior regions of RNA tertiary folds. Latham and Cech

2004 Computational RNA modeling with chemical probing data

2017 RT Stops and RT mutate methods. High throughput RNA structure probing

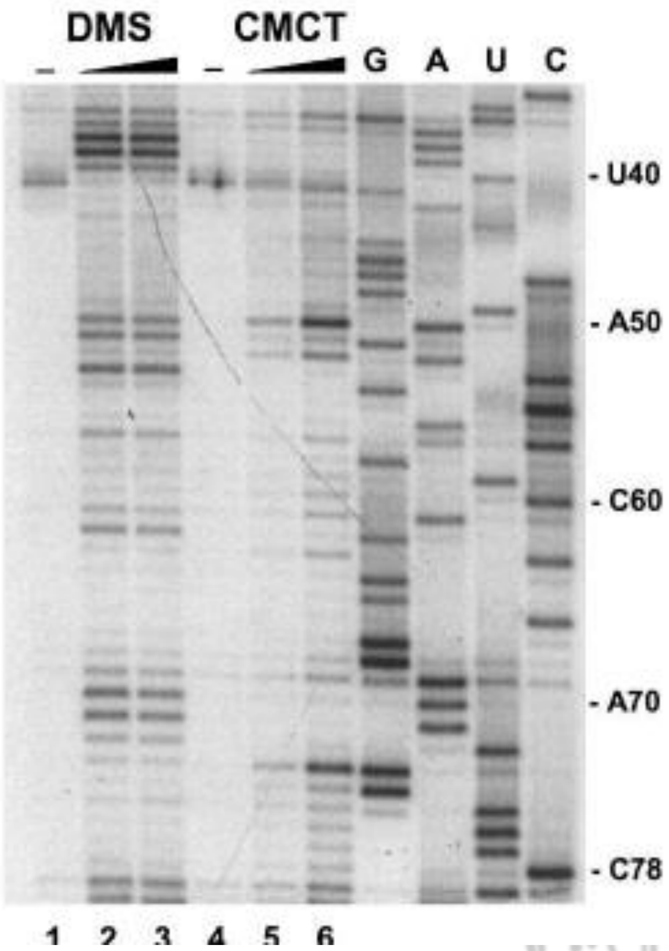
"for their interpretation of the genetic code and its function in protein synthesis."

Robert W. Holley

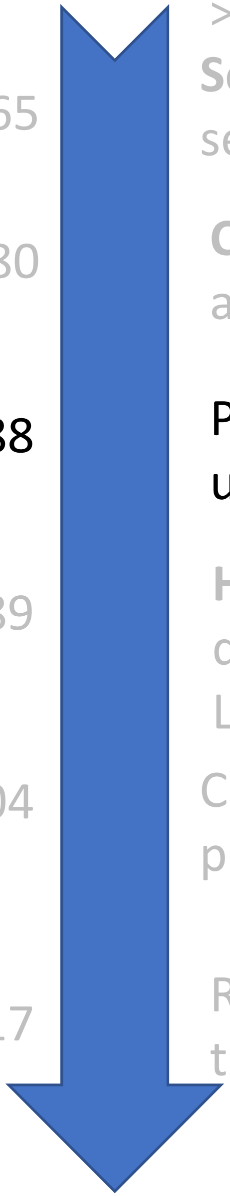


The Journey so far..

	> 4 billion years
1965	**Secondary structure Model**: Purification and sequence yeast tRNA, (1968 Nobel Prize)
1980	**Chemical Probes** to interrogate secondary and tertiary structure.
1988	Probing of RNA structures in Living cells using DMS
1989	**Hydroxyl Radical Footprinting:** Interactions that define exterior regions of RNA tertiary folds. Latham and Cech
2004	Computational RNA modeling with chemical probing data
2017	RT Stops and RT mutate methods. High throughput RNA structure probing



The Journey so far..



> 4 billion years

1965 **Secondary structure Model**: Purification and sequence yeast tRNA, (1968 Nobel Prize)

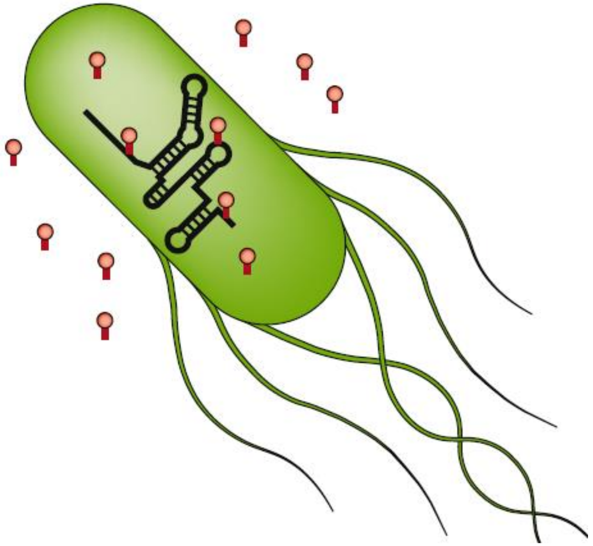
1980 **Chemical Probes** to interrogate secondary and tertiary structure.

1988 Probing of RNA structures in Living cells using DMS

1989 **Hydroxyl Radical Footprinting:** Interactions that define exterior regions of RNA tertiary folds. Latham and Cech

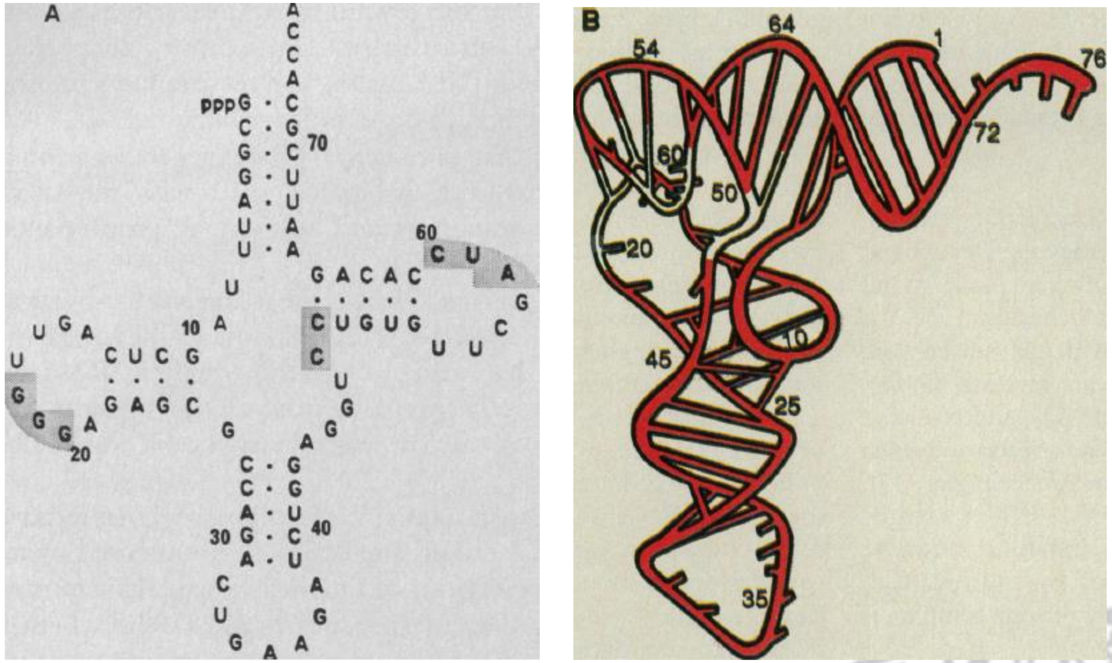
2004 Computational RNA modeling with chemical probing data

2017 RT Stops and RT mutate methods. High throughput RNA structure probing



The Journey so far..

	> 4 billion years
1965	**Secondary structure Model**: Purification and sequence yeast tRNA, (1968 Nobel Prize)
1980	**Chemical Probes** to interrogate secondary and tertiary structure.
1988	Probing of RNA structures in Living cells using DMS
1989	Hydroxyl Radical Foot printing
2004	Computational RNA modeling with chemical probing data
2017	RT Stops and RT mutate methods. High throughput RNA structure probing



Latham and Cech (1989)

The Journey so far..

	> 4 billion years
1965	**Secondary structure Model**: Purification and sequence yeast tRNA, (1968 Nobel Prize)
1980	**Chemical Probes** to interrogate secondary and tertiary structure.
1988	Probing of RNA structures in Living cells using DMS
1989	**Hydroxyl Radical Footprinting:** Interactions that define exterior regions of RNA tertiary folds. Latham and Cech
2004	Computational RNA modeling with chemical probing data
2017	RT Stops and RT mutate methods. High throughput RNA structure probing

A dynamics programming algorithm for predicting RNA secondary structure accommodating folding constraints and free energy increments.

The Journey so far..

> 4 billion years

1965 **Secondary structure Model**: Purification and sequence yeast tRNA, (1968 Nobel Prize)

1980 **Chemical Probes** to interrogate secondary and tertiary structure.

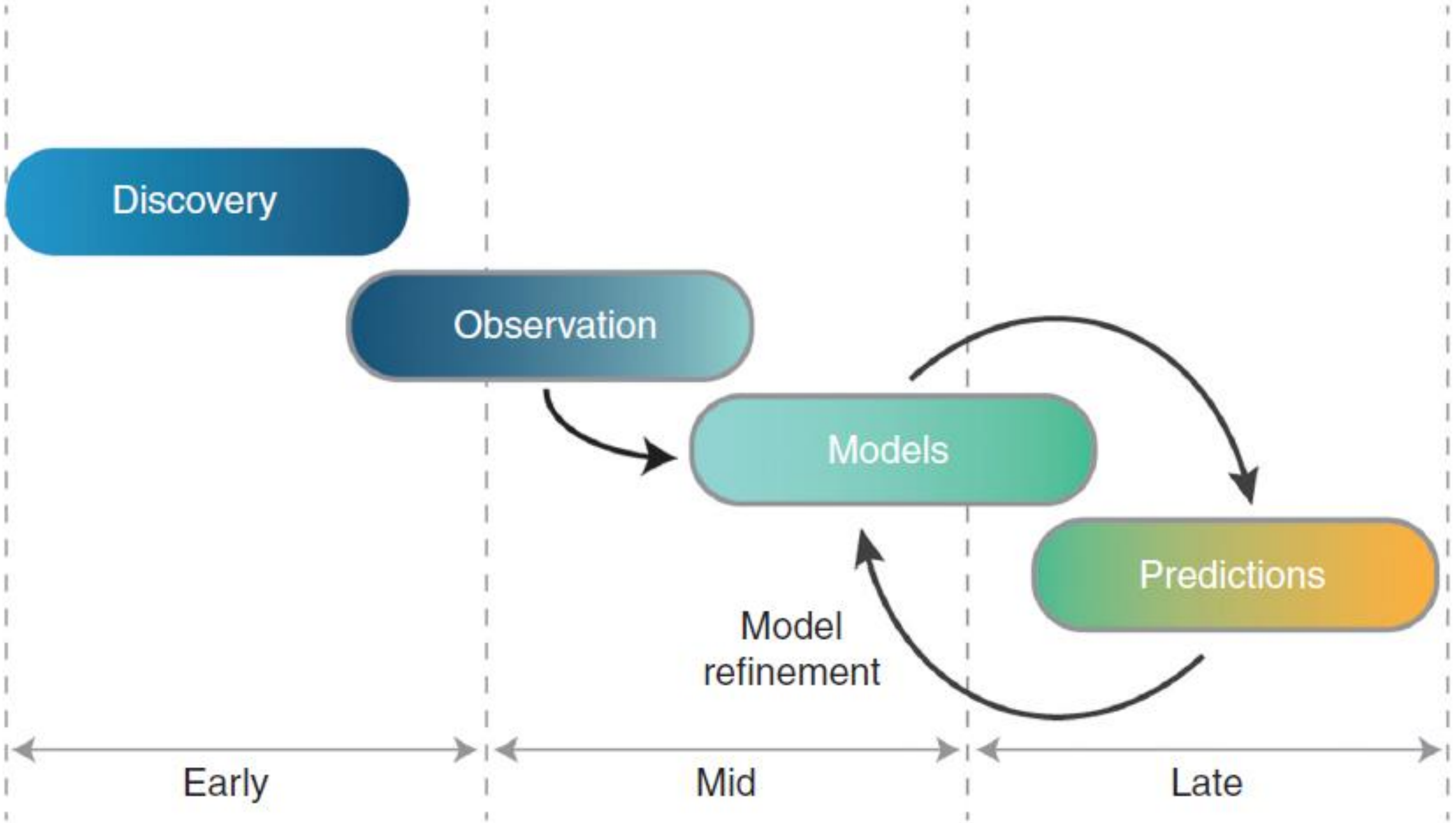
1988 Probing of RNA structures in Living cells using DMS

1989 **Hydroxyl Radical Footprinting:** Interactions that define exterior regions of RNA tertiary folds. Latham and Cech

2004 Computational RNA modeling with chemical probing data

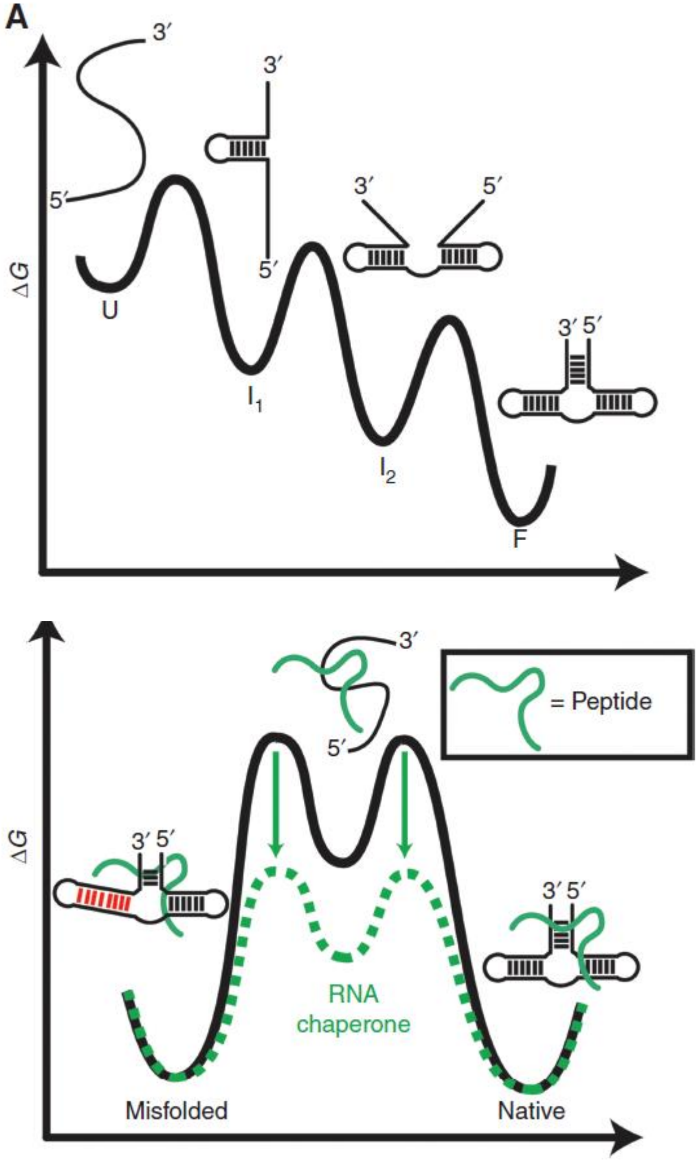
2017 RT Stops and RT mutate methods. High throughput RNA structure probing

The Journey so far..



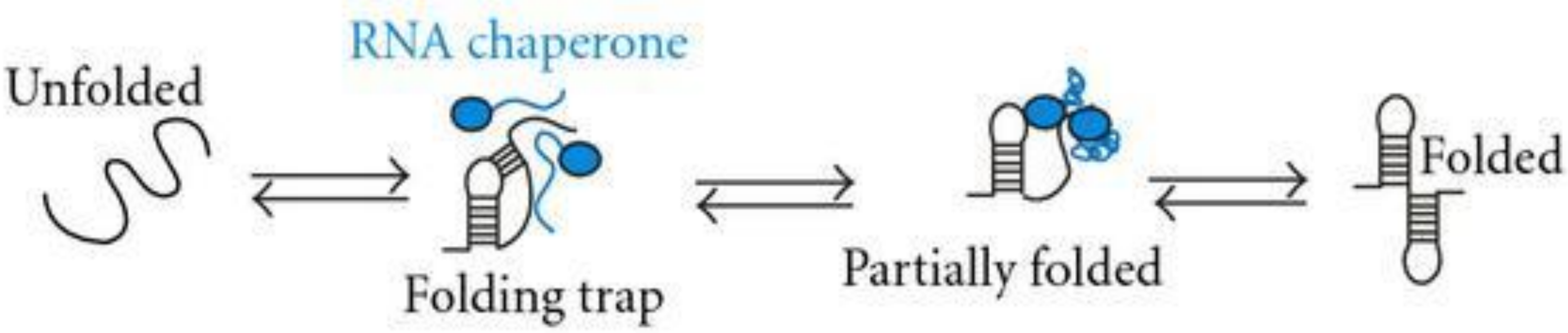
RNA Molecular Properties

- RNA secondary structure is highly stable.
- Ability to form stable structure with short oligomeric strands facilitates development of function under primitive conditions.
- Potential for Hydrogen bonding in residues.
- Stable secondary structure and stability => Misfolded state.
- “RNA Chaperons”, proteins that facilitate unfolding and refolding of DNA



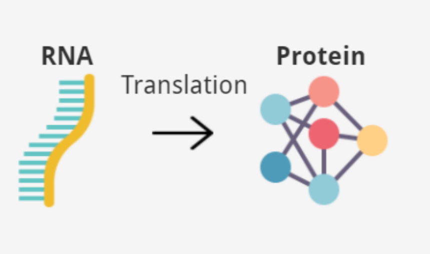
Early Observations

- RNA chaperons may have endowed RNA with increased functionality.
- Short, non specific peptides functioned as early RNA chaperones in an RNA World.
- Strong local interactions and RNA's ability to form multiple quasi-stable states may account for
    - Pre-mRNA splicing.
    - Telomere maintenance.
    - RNA interference.



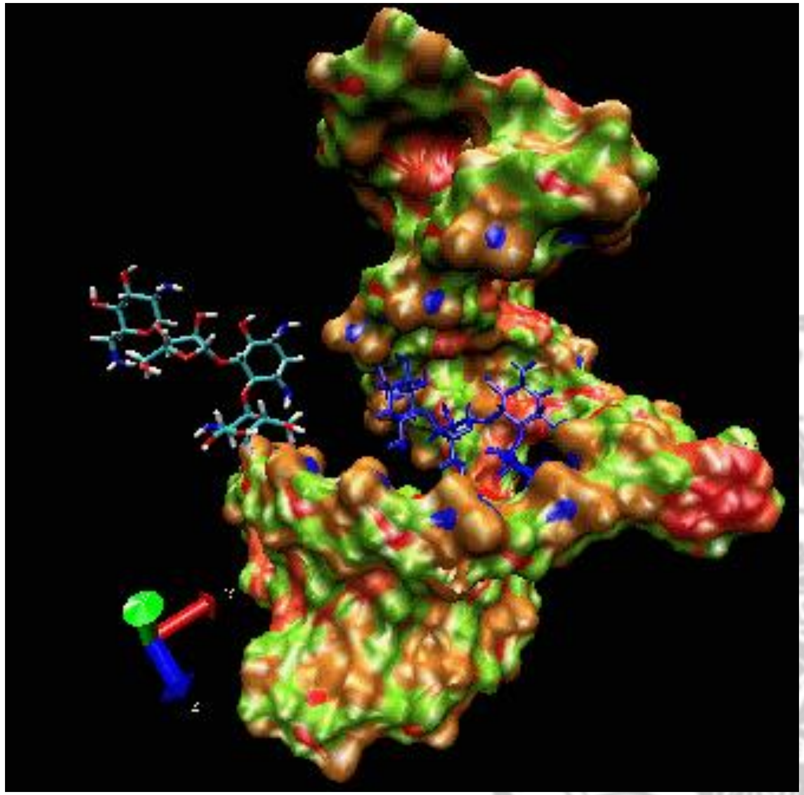
Early Observations

What might have caused proteins to ultimately take over?



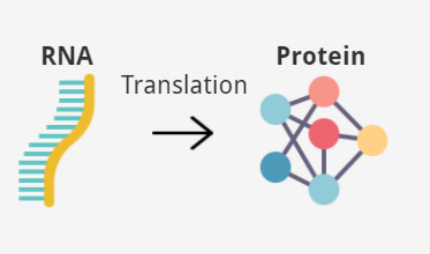
- **Favorable folding properties**, relative to RNA.

- Large number of more catalytically effective and **diverse side chains of proteins**.
- **Speed:** Rate constant of ligand binding for RNA << Rate constant of ligand binding for protein limit ~ diffusional limit.

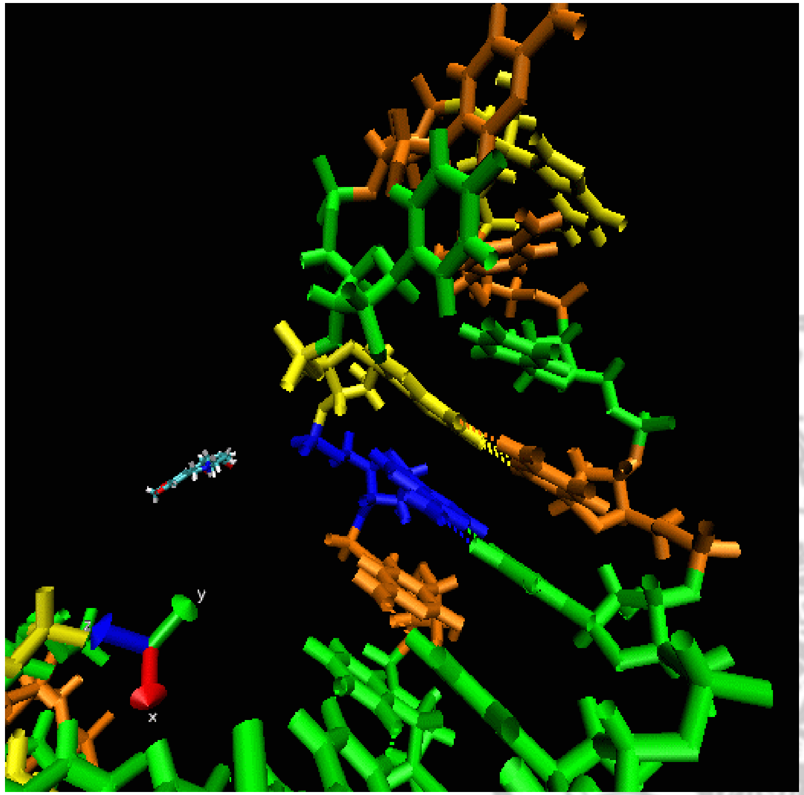


Early Observations

What might have caused proteins to ultimately take over?



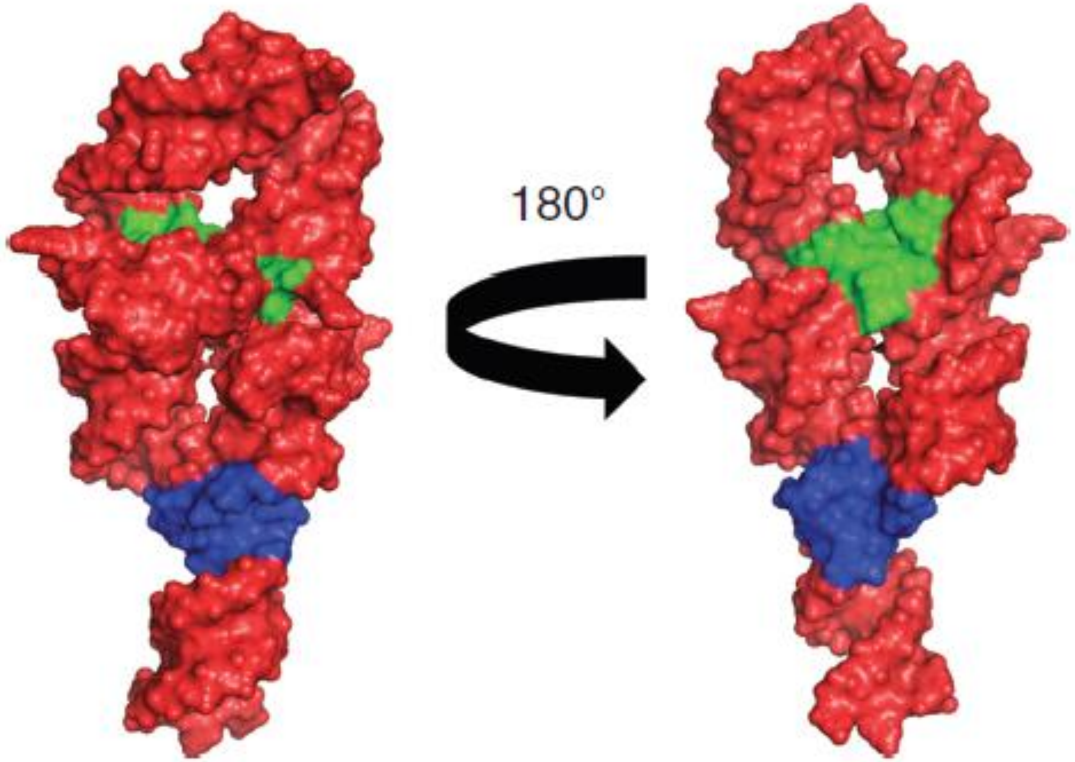
- **Favorable folding properties**, relative to RNA.
- Large number of more catalytically effective and **diverse side chains of proteins**.
- **Speed:** Rate constant of ligand binding for RNA << Rate constant of ligand binding for protein limit ~ diffusional limit.

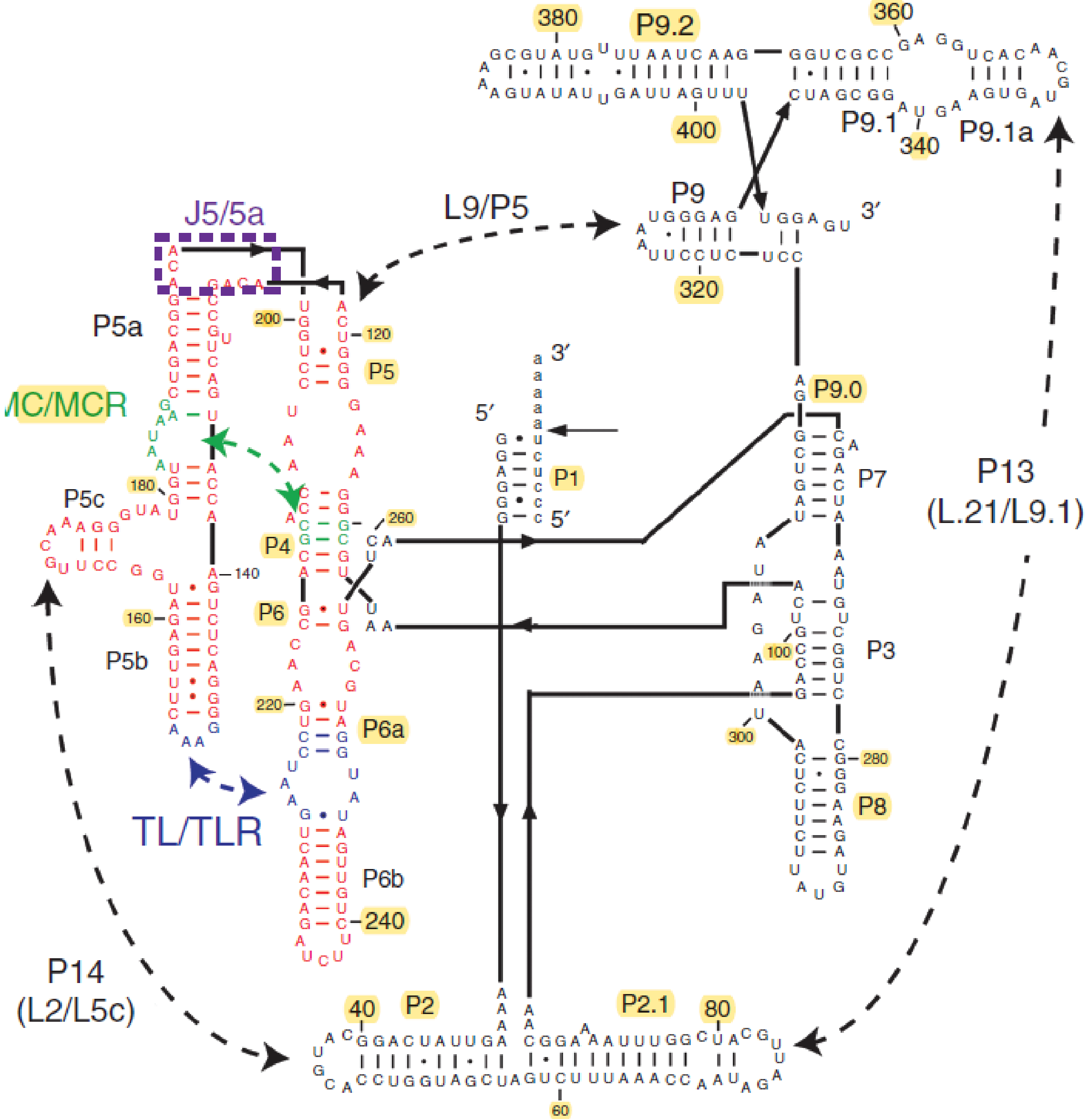


Middle Epoch – RNA Folding Models

- Integrated Kinetic, Thermodynamic and structural framework .
- Globular nature of ribozymes similar to globular proteins. For e.g, Tetrahymena group I intron
- Existence of multiple stable intermediates => Time-resolved hydroxyl radical cleavage.
- Evidence of kinetic traps during folding.

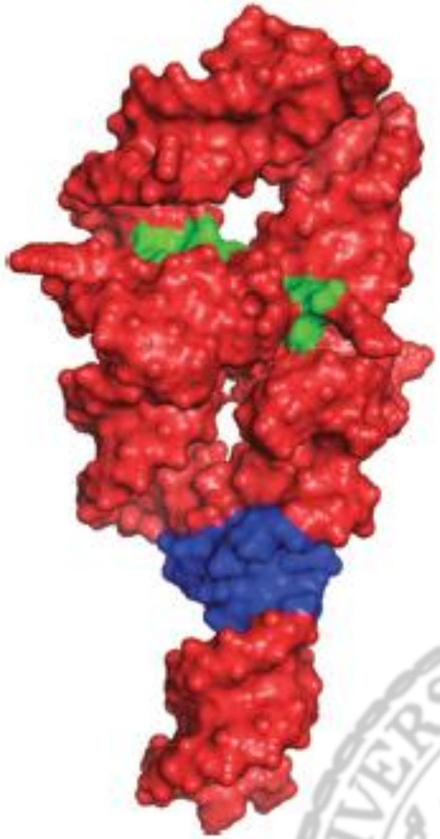
Tetrahymena Group I Intron



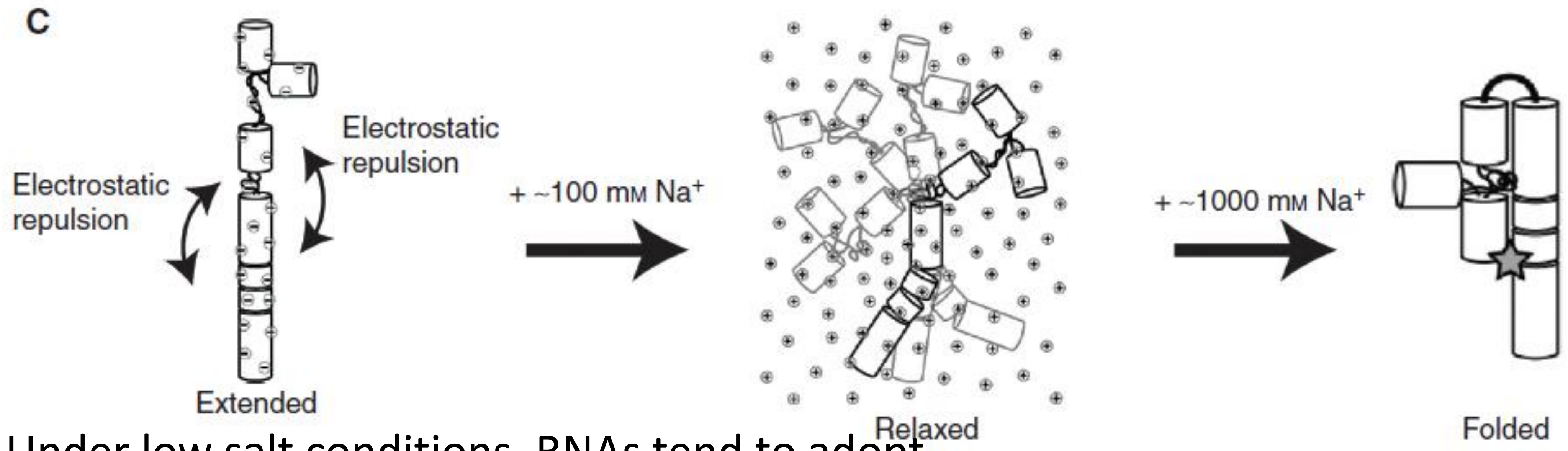


Tetrahymena Group I intron

- Discovery of the P4-P6 region of the ribozyme led to model in RNA folding studies.
- Nucleation around Mg^{2+} core was tested using a mutant of Tetrahymena ribozyme lacking the Mg^{2+} binding domain (P5abc.).
- Catalytic activity on addition of Mg^{2+} binding domain (P5abc) > Folding initiated simultaneous addition of Mg^{2+} and P5abc.
- This shows ribozyme can fold along multiple pathways. Mg^{2+} core formation is not an obligate early folding step.



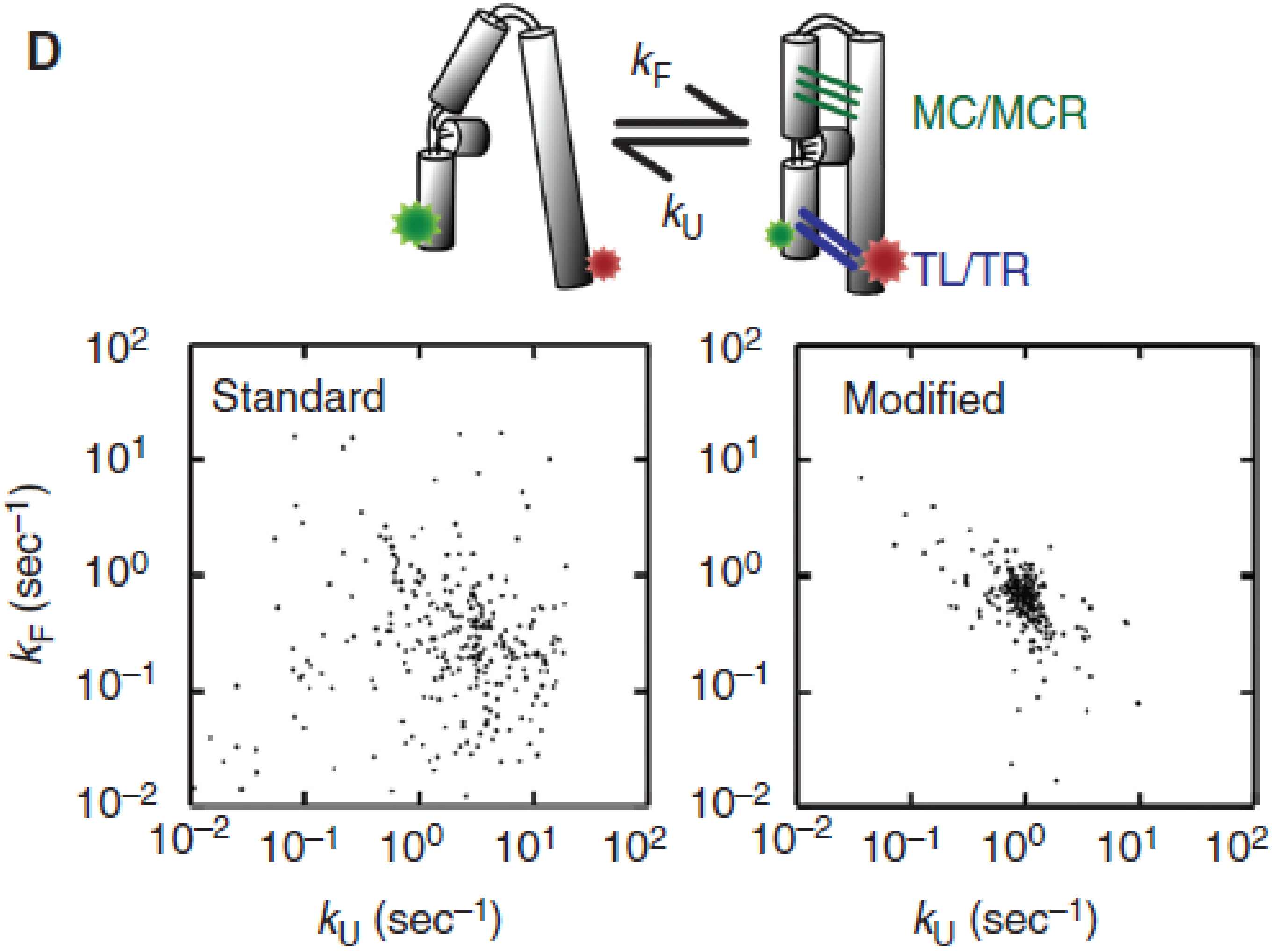
Tetrahymena Group I intron



- Under low salt conditions, RNAs tend to adopt extended states, attributable to electrostatic repulsion between negatively charged phosphates of the RNA backbone.

- Direct link between physical association of Mg^{2+} and thermodynamics of folding.

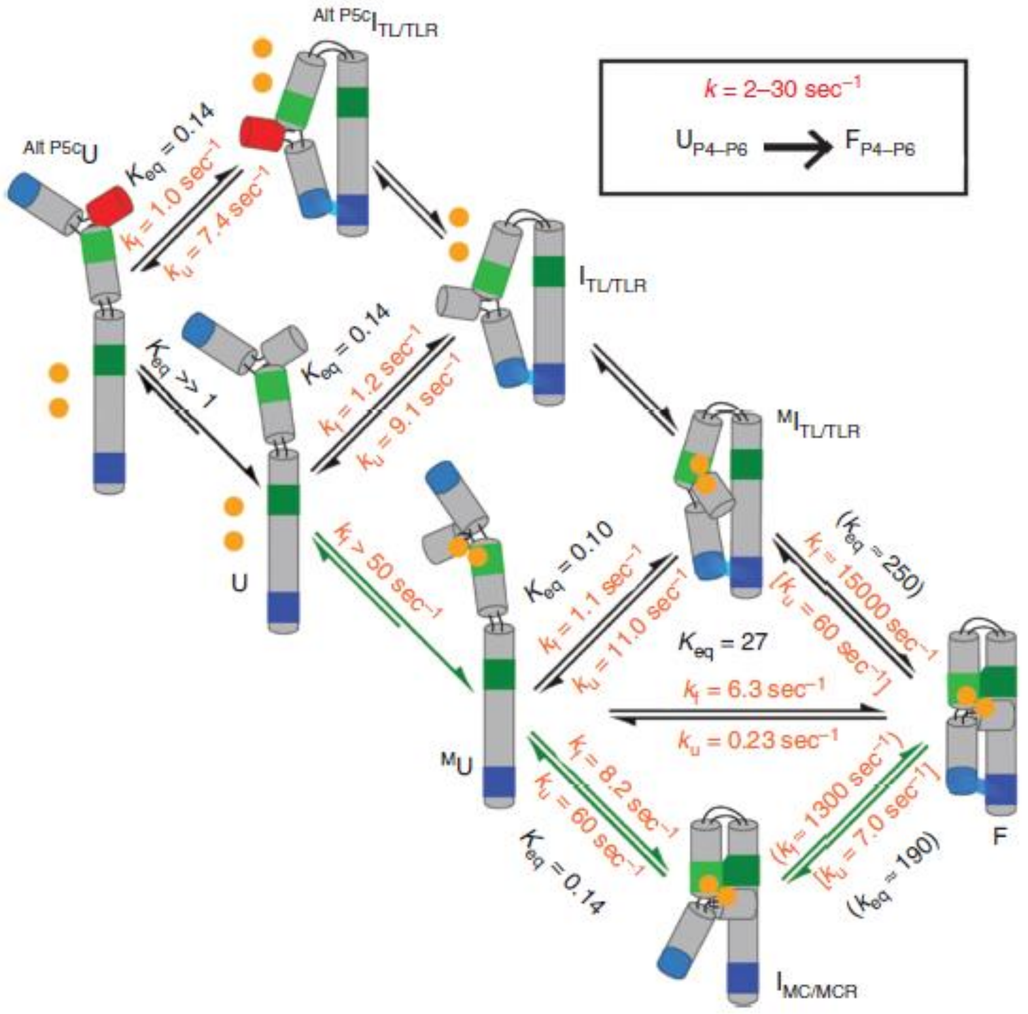
Tetrahymena Group I intron



- Heterogeneity greatly reduced in single molecule studies with a changed protocol. (compare left and right)

Tetrahymena Group I intron

- sMFRET approach to characterize folding of P4-P6 mutants.
- Native conformation of the P5abc domain dominates.
- Following Mg^{2+} binding, MC/MCR forms first followed by rapid formation of TL/TLR. [90% RNA molecules form MC/MCR first).
- 10% form more remote contact first.
- The pathways and choice of pathways are influenced by **collision frequencies and also by the properties of individual tertiary elements.**



Approaches - folding pathways

- **smFRET**: Different RNA molecules display different kinetic, thermodynamic or catalytic behaviour in smFRET experiments.

- **SAXS**: Information about overall compactness and shape.
  - In Tetrahymena ribozyme compaction occurs even before first tertiary interactions. => Mg^{2+} is involved in more than one folding steps.

- **Chemical Probing with DMS:** Local information about initial and final structural states.

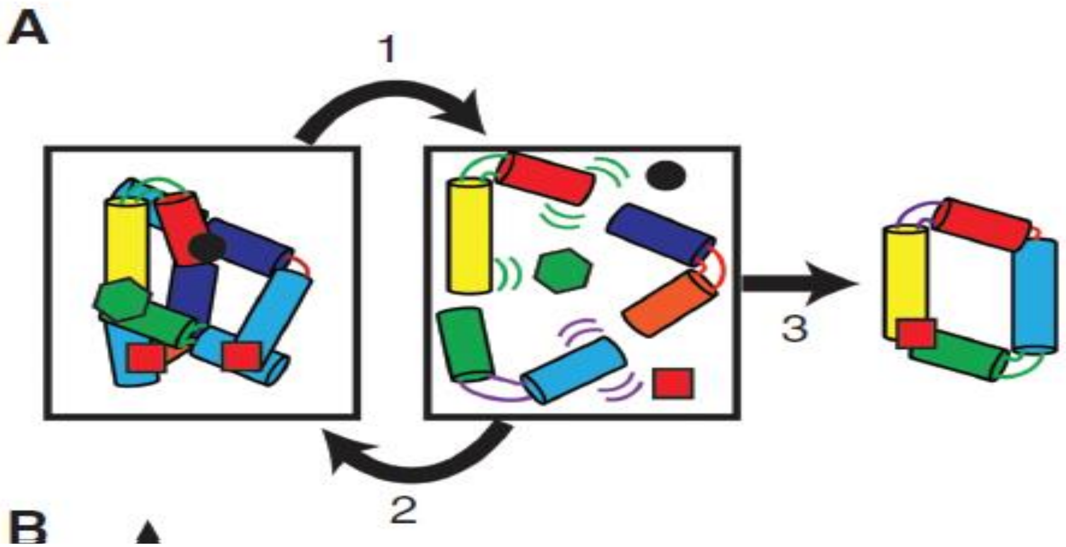
- **Mutations:** To stabilize or destabilize known interactions.

Predictive Models

- A quantitative model: Sequence => 3D structure => RNA dynamics and energetics.
- Structural, dynamic and thermodynamic models required to predict complex interactions and processes involving structured RNAs inside the cell.
- Blind trials
- Unknown sources of variations in complex problems, unconscious bias and publication bias provide limitations to make predictions before solvated structures.
- ROSETTA- structure prediction from crystallographic structures.

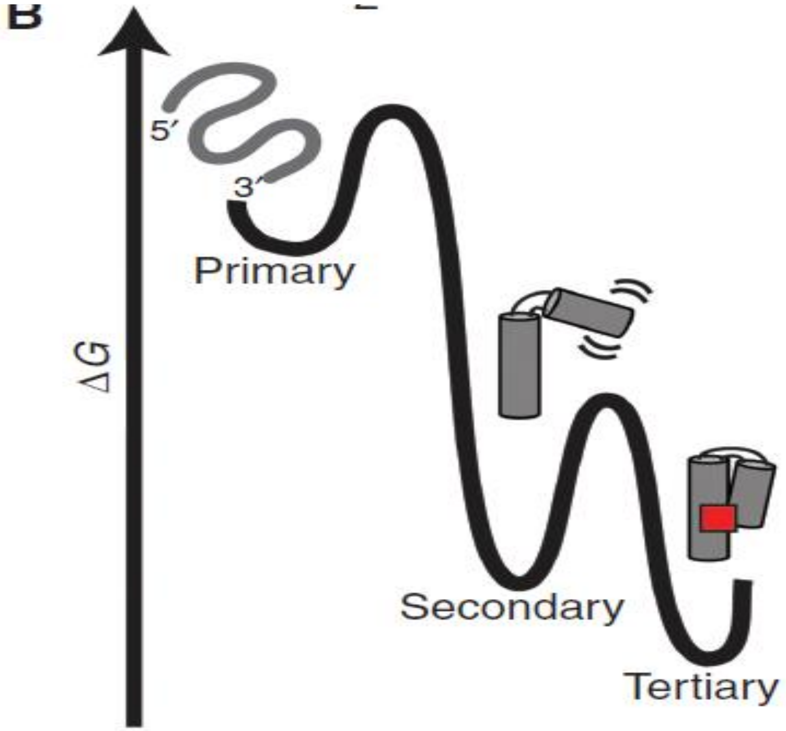
The Reconstitution Model

- Helix Junction Helix elements
- Tertiary contact motifs.
- Study in isolation.



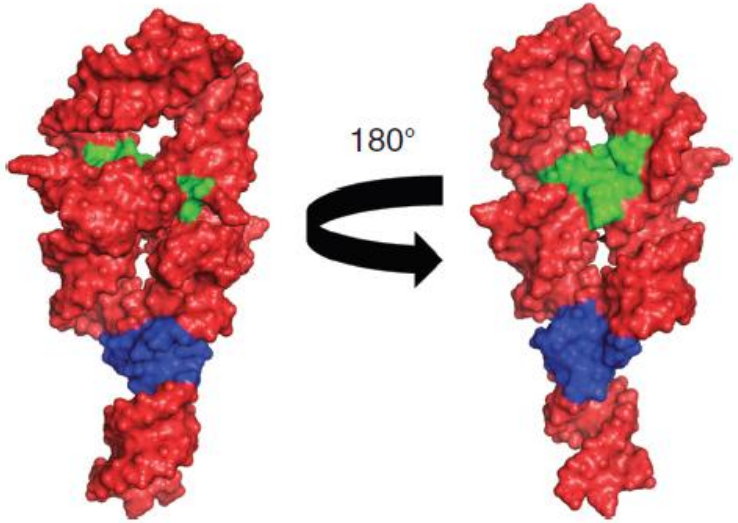
Features of Reconstitution Model

- RNA folding is hierarchical.
- Most RNA residues interact only with their primary or secondary structure neighbours.
- Existence of recurring tertiary motifs in RNA. E.g, TL/TLR in Tetrahymena



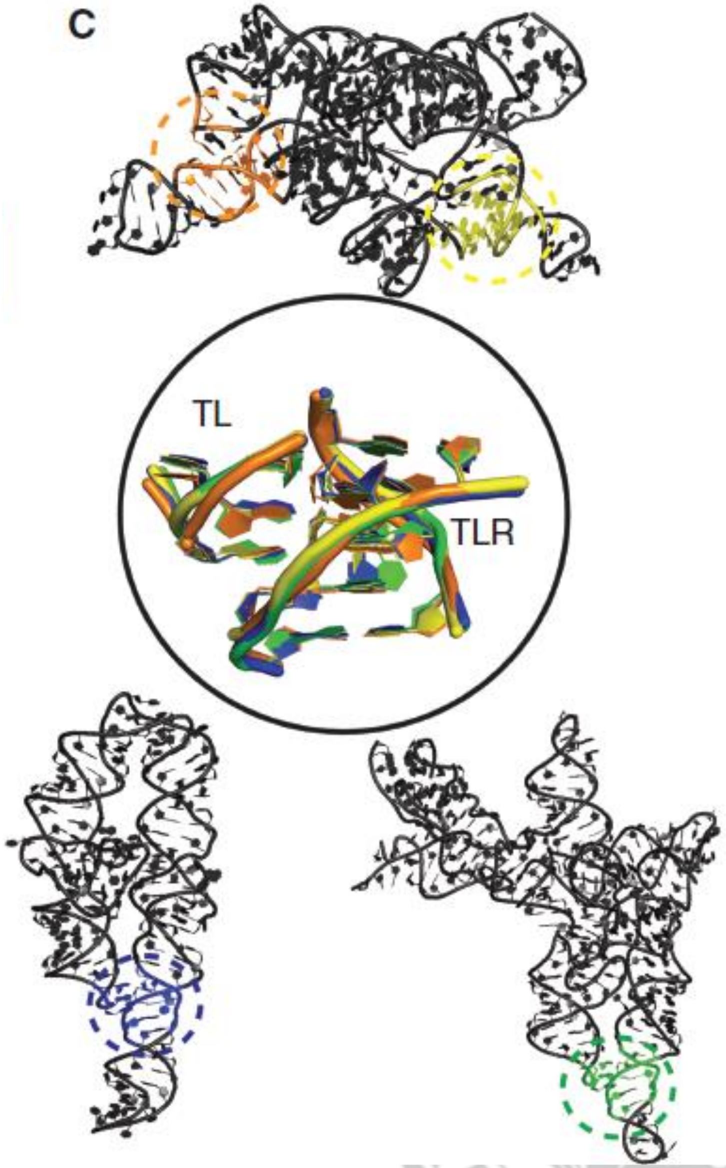
Features of Reconstitution Model

- RNA folding is hierarchical.
- Most RNA residues interact only with their primary or secondary structure neighbours.
- Existence of recurring tertiary motifs in RNA. E.g, TL/TLR in Tetrahymena



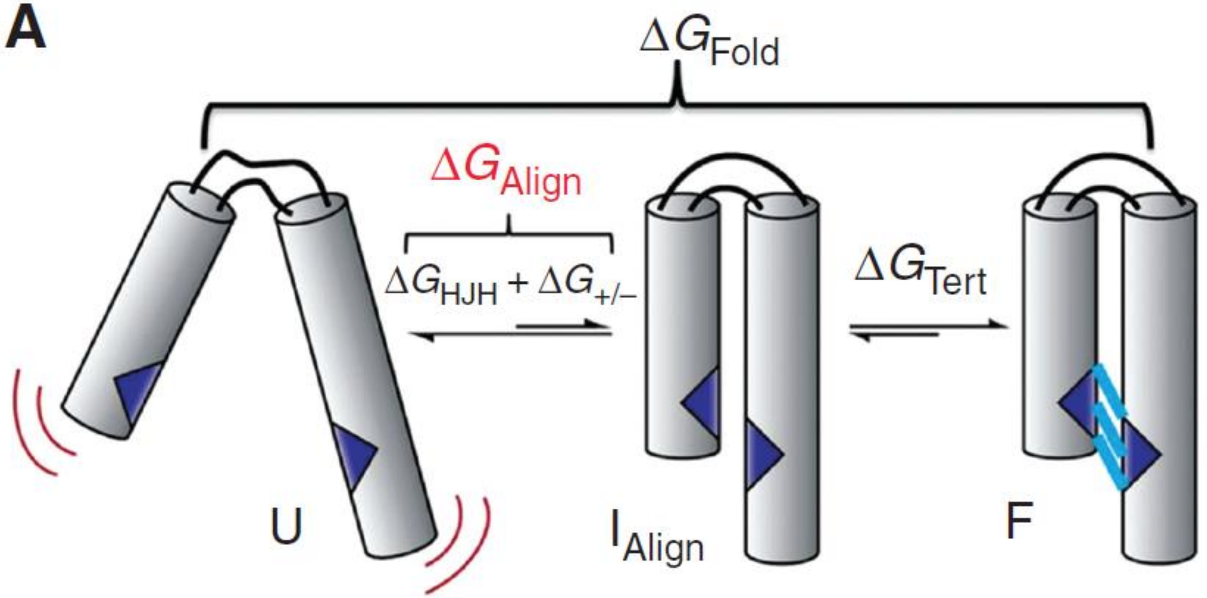
Features of Reconstitution Model

- RNA folding is hierarchical.
- Most RNA residues interact only with their primary or secondary structure neighbours.
- Existence of recurring tertiary motifs in RNA. E.g, TL/TLR in Tetrahymena



Features of Reconstitution Model

- Probability of forming a folded structure (F) depends on:
  1. Tertiary contact alignment ($\Delta G_{Align} = \Delta G_{HJH} + \Delta G_{+/-}$)
  2. Strength of the tertiary interactions. (ΔG_{Tert})
- Junction plays a crucial role in shaping conformational ensembles.
- Floppier Junctions can accommodate misalignments.

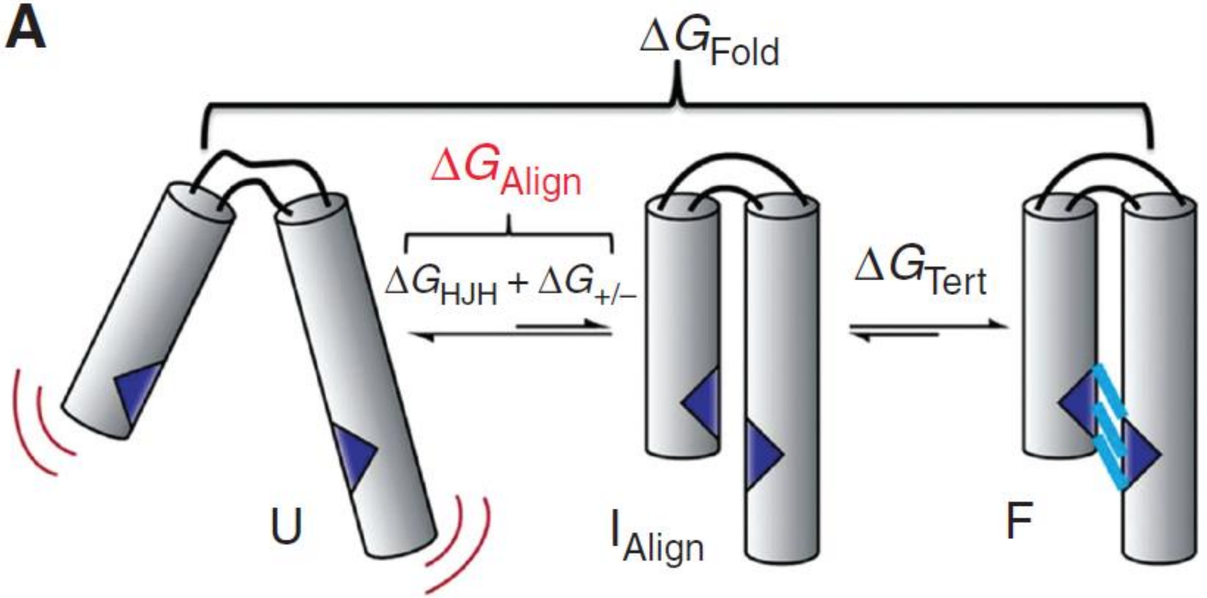


Unfolded RNA

folded RNA

Features of Reconstitution Model

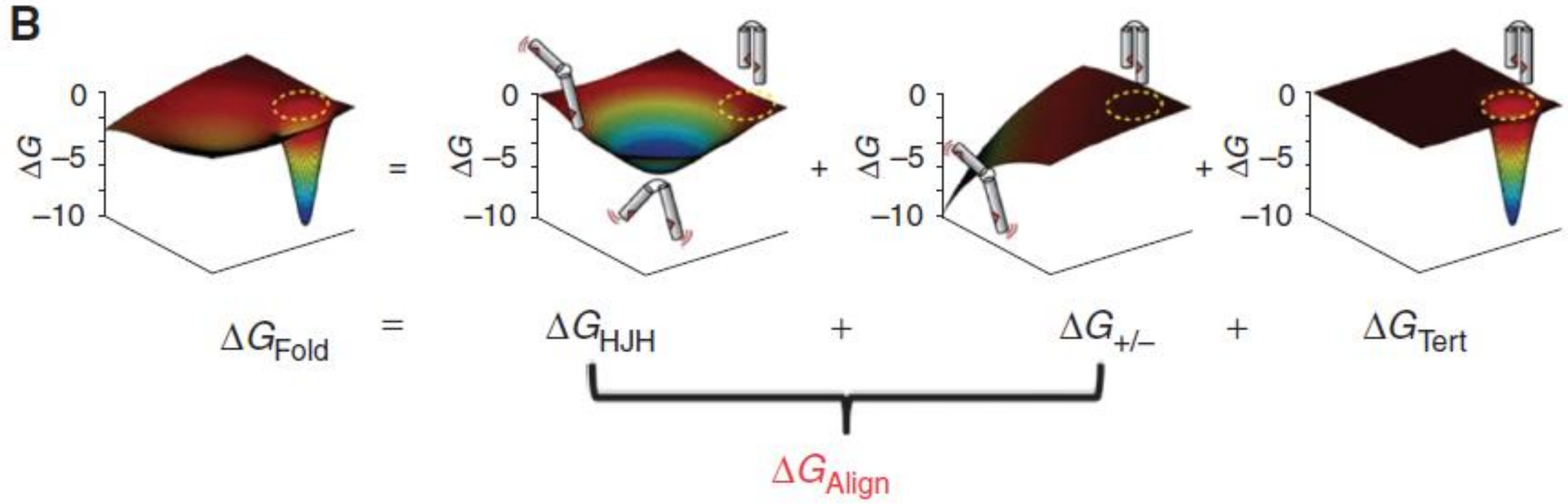
- More tertiary hydrogen bonds and stacking interactions, more favourable ΔG_{Tert} will be.
- Matching of the conformations preferences also matter.
- "Energetic Additivity" cannot be expected for RNA tertiary structure.
- "Ensemble Additivity"



Unfolded RNA

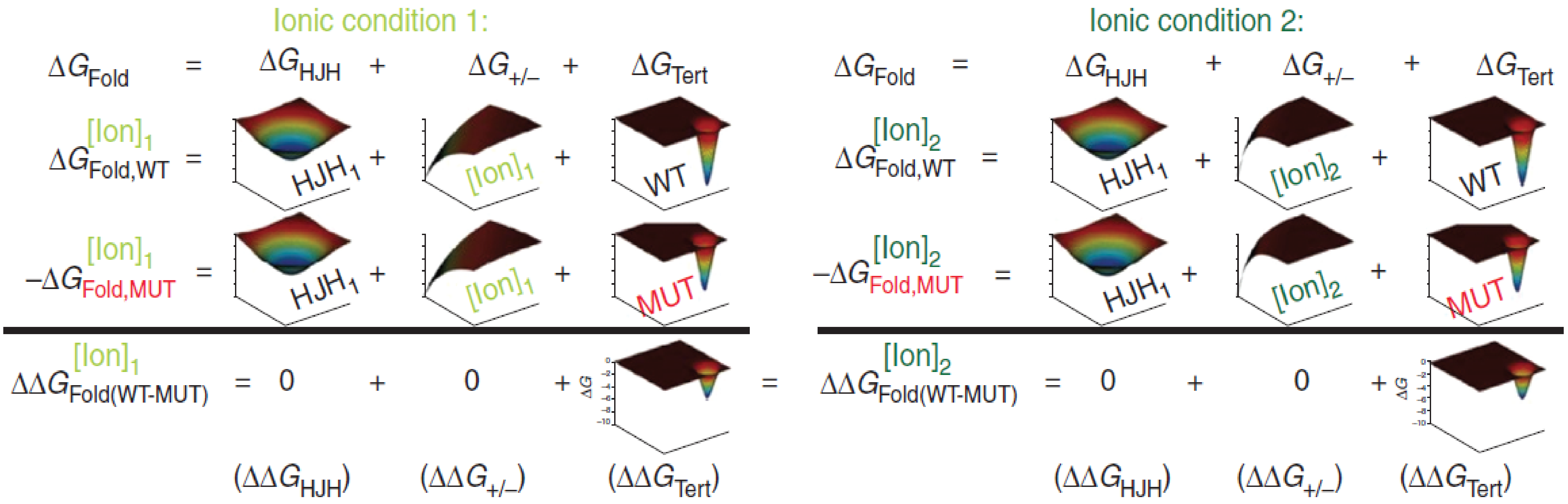
folded RNA

Features of Reconstitution Model



- Perturbation in one of the constituent energy landscapes should not alter the others.
- Overall effect of folding ($\Delta\Delta G_{fold}$) is the same at each ionic condition

Features of Reconstitution Model



Ionic condition 1:

$$\Delta G_{\mathrm{Fold}} = \Delta G_{\mathrm{HJH}} + \Delta G_{+/-} + \Delta G_{\mathrm{Tert}}$$

$$\Delta G^{[\mathrm{Ion}]_1}_{\mathrm{Fold,WT}} = \mathrm{HJH}_1 + [\mathrm{Ion}]_1 + \mathrm{WT}$$

$$-\Delta G^{[\mathrm{Ion}]_1}_{\mathrm{Fold,MUT}} = \mathrm{HJH}_1 + [\mathrm{Ion}]_1 + \mathrm{MUT}$$

$$\Delta\Delta G^{[\mathrm{Ion}]_1}_{\mathrm{Fold(WT\text{-}MUT)}} = 0 + 0 +$$

$(\Delta\Delta G_{\mathrm{HJH}})$ $(\Delta\Delta G_{+/-})$ $(\Delta\Delta G_{\mathrm{Tert}})$

=

Ionic condition 2:

$$\Delta G_{\mathrm{Fold}} = \Delta G_{\mathrm{HJH}} + \Delta G_{+/-} + \Delta G_{\mathrm{Tert}}$$

$$\Delta G^{[\mathrm{Ion}]_2}_{\mathrm{Fold,WT}} = \mathrm{HJH}_1 + [\mathrm{Ion}]_2 + \mathrm{WT}$$

$$-\Delta G^{[\mathrm{Ion}]_2}_{\mathrm{Fold,MUT}} = \mathrm{HJH}_1 + [\mathrm{Ion}]_2 + \mathrm{MUT}$$

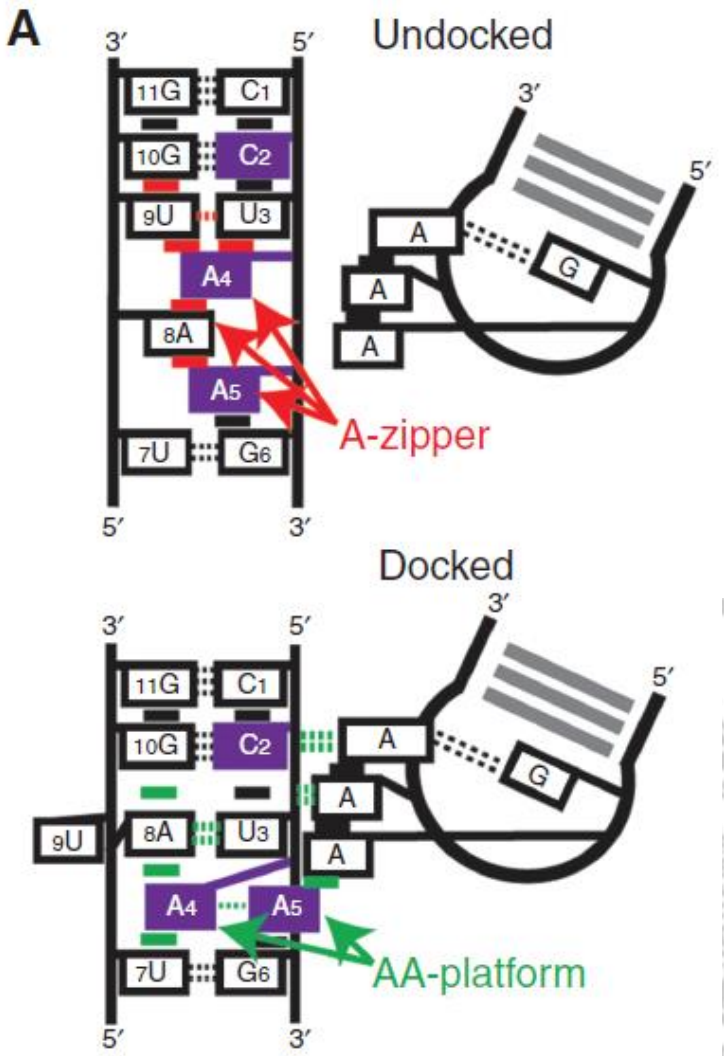
$$\Delta\Delta G^{[\mathrm{Ion}]_2}_{\mathrm{Fold(WT\text{-}MUT)}} = 0 + 0 +$$

$(\Delta\Delta G_{\mathrm{HJH}})$ $(\Delta\Delta G_{+/-})$ $(\Delta\Delta G_{\mathrm{Tert}})$

- Overall effect of folding ($\Delta\Delta G_{fold}$) is the same at each ionic condition

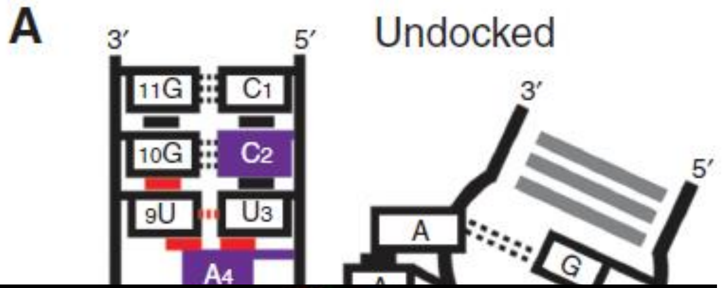
Features of Reconstitution Model

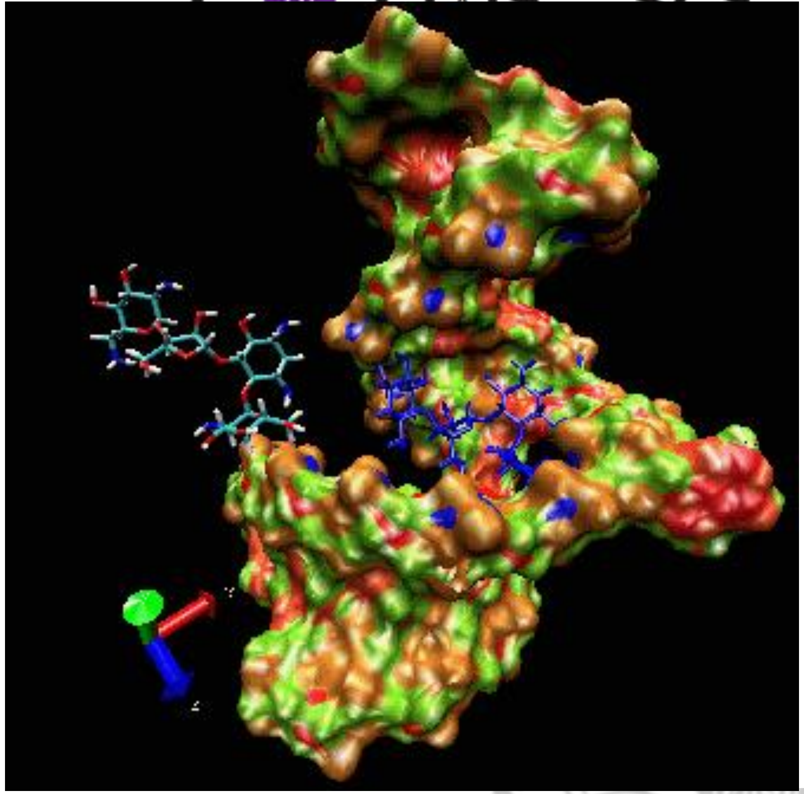
- Unfolding rate constants are local properties of the tertiary motif involved and thus are transferable across RNAs for a given set of conditions.
- Some TL/TLR variants involve additional electrostatic and metal ions effect ($\Delta G_{\text{Tert, +/-}}$)



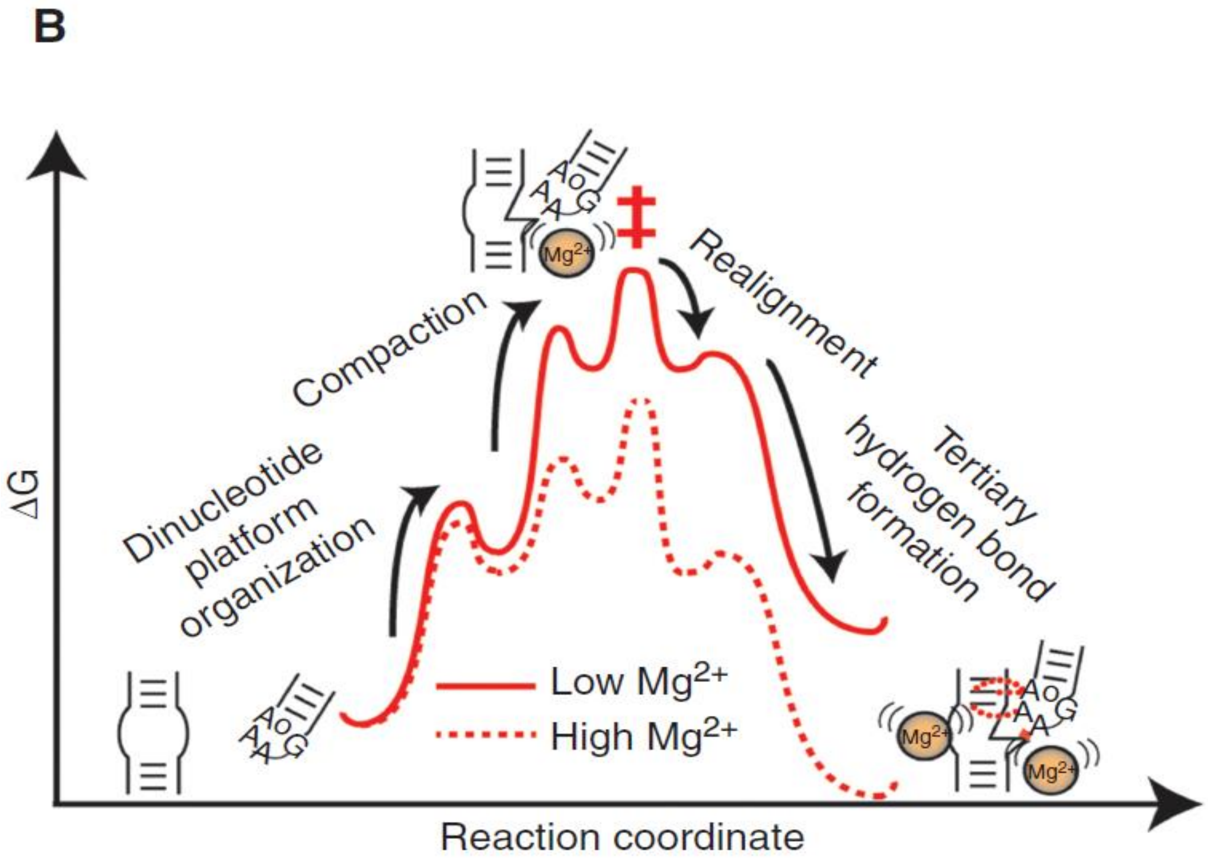
Features of Reconstitution Model

- Unfolding rate constants are local properties of the tertiary motif involved and thus are transferable across RNAs for a given set of conditions.
- Some TL/TLR variants involve additional electrostatic and metal ions effect ($\Delta G_{\text{Tert, +/-}}$)

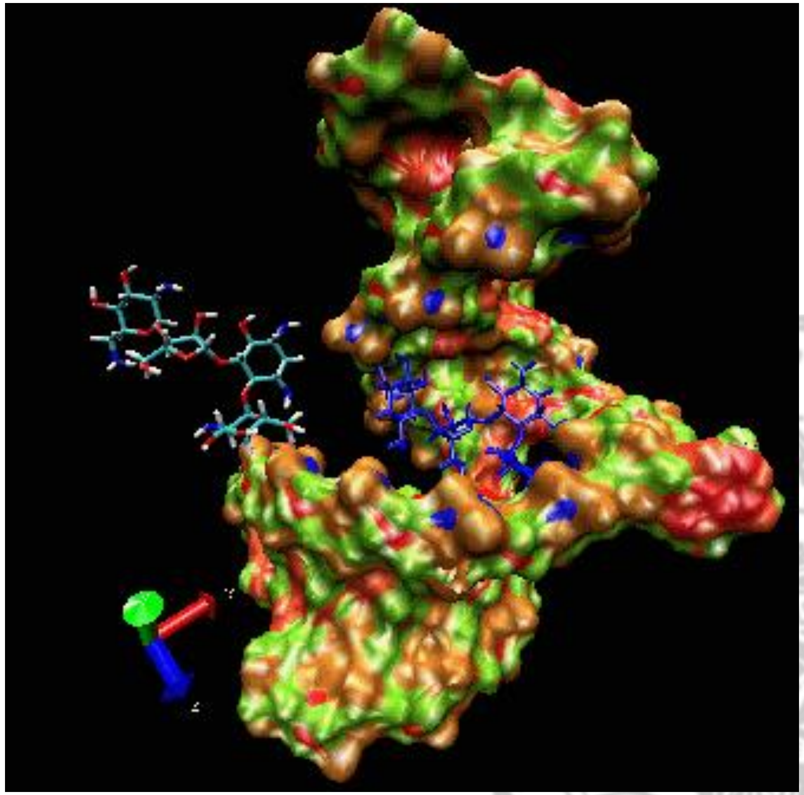




Features of Reconstitution Model

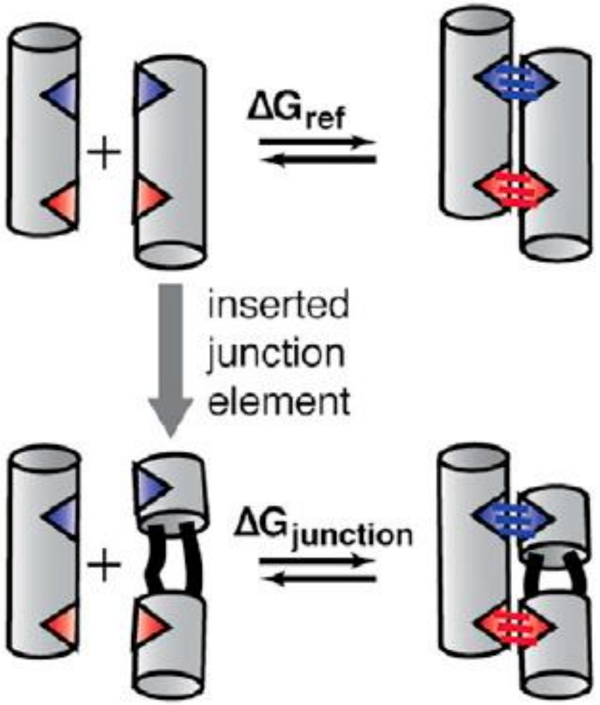


- Model for conformational changes and nature of transition state for TL/TLR.



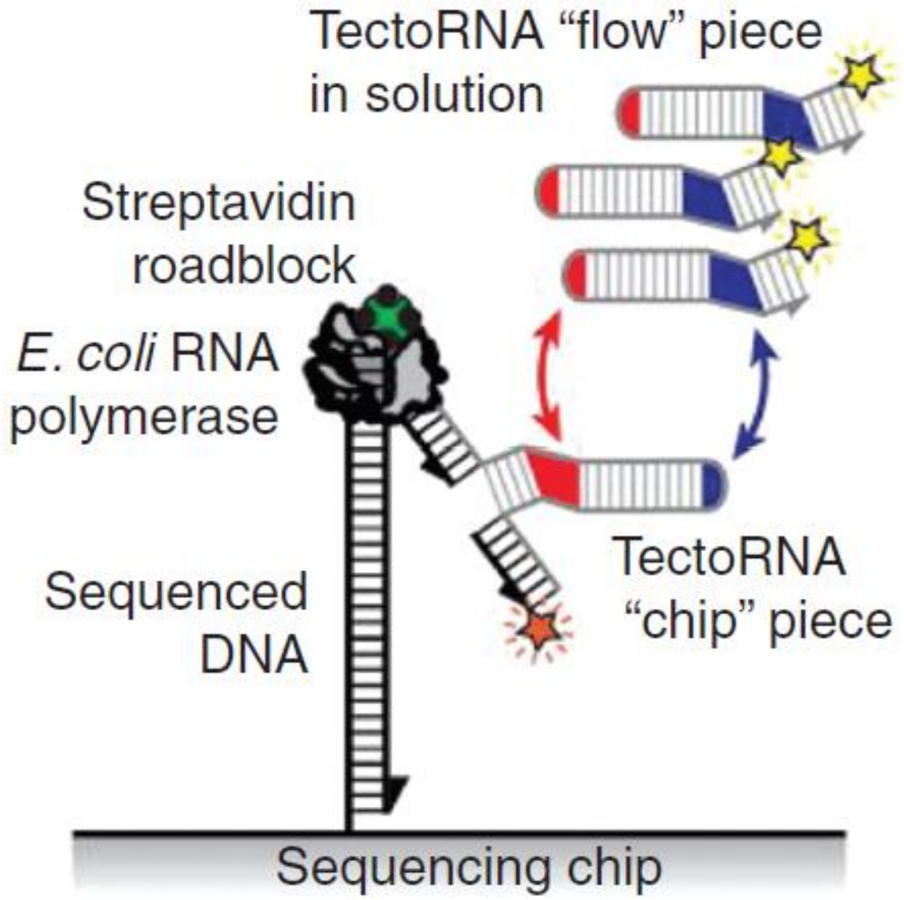
High throughput characterization

- Broad implementation of “The reconstitution model’
- Atomic-level data about junctions and tertiary contact elements not required.
- Instead, preferred orientations of the helices emerging from junctions to determine spatial probabilities of tertiary contact elements connected to the helices.



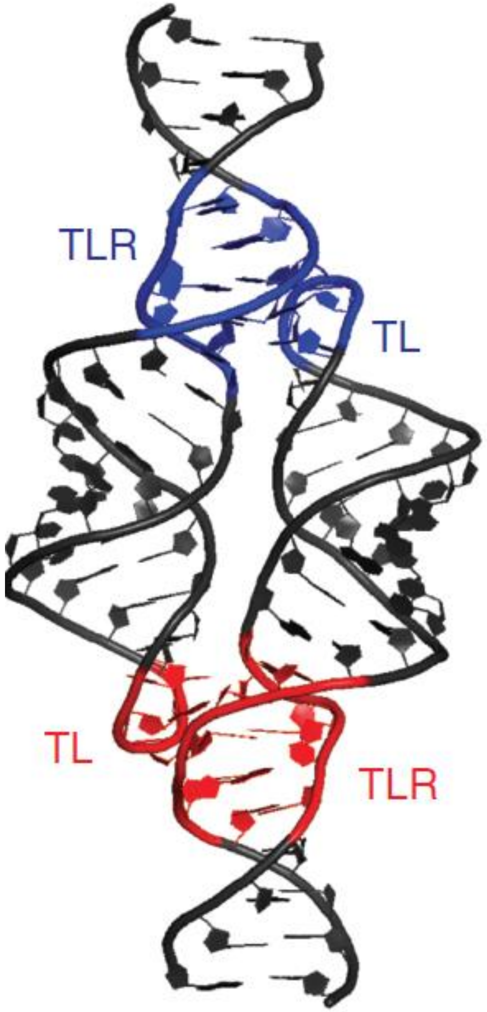
High throughput characterization

- An engineered dimer "TectoRNA", assembles via formation of two distinct tertiary contacts.
- Likelihood of forming both contacts depends on the alignments.
- Thermodynamic fingerprints reflect conformational preferences.



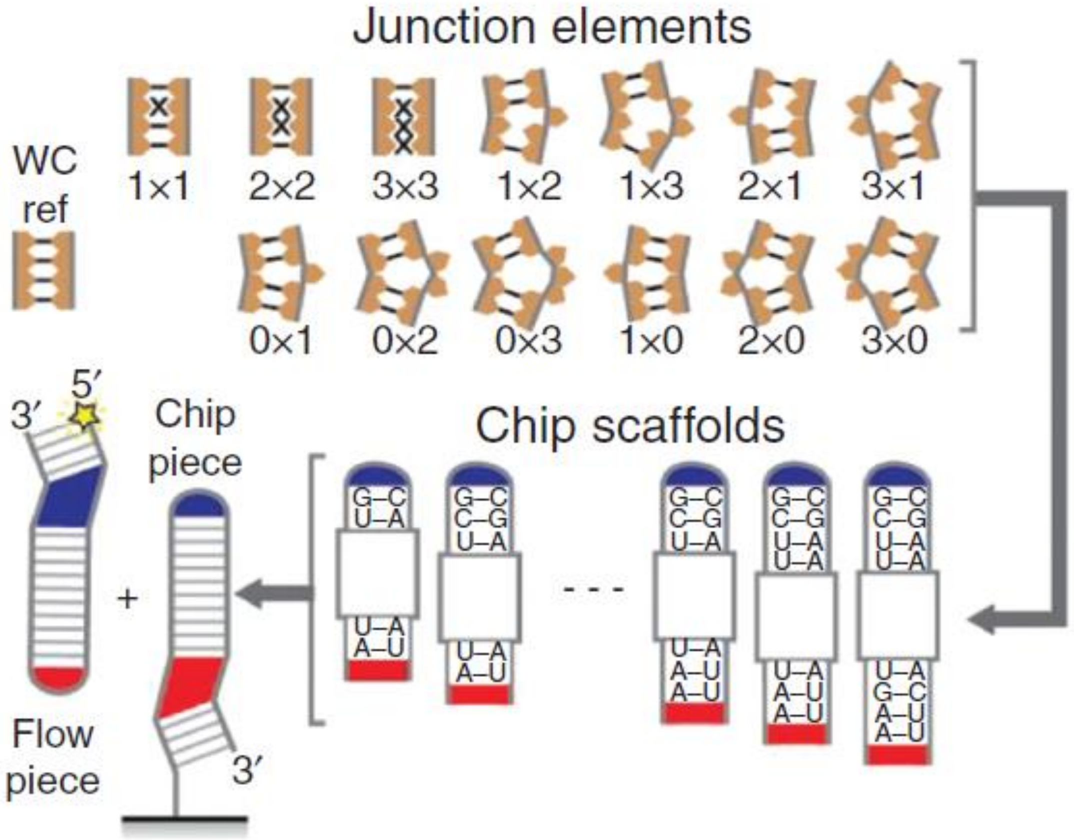
High throughput characterization

- An engineered dimer "TectoRNA", assembles via formation of two distinct tertiary contacts.
- Likelihood of forming both contacts depends on the alignments.
- Thermodynamic fingerprints reflect conformational preferences.



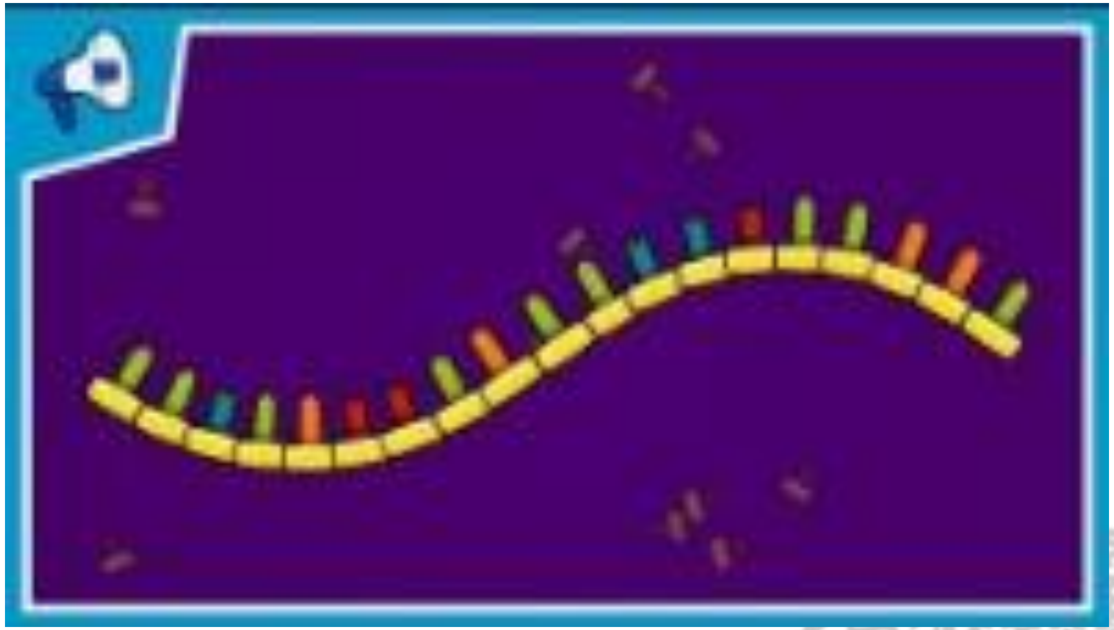
High throughput characterization

- TectoRNA system is used as a scaffold for insertion of a wealth of two-way junctions and components that could be systematically varied such as
  - Helix length and content
  - Junction type and sequence
  - Tertiary motif identity



Conclusion

- Basis of early life RNA -> Proteins
- ‘Energy Additivity’ in predicting thermodynamic stability of secondary structures.
- Predictive Phase in RNA folding and Dynamics.
- Strong evidence in favor of ‘The Reconstitution Model’
- Ability of Quantitative frameworks that ultimately predict cellular control and function.



https://youtu.be/K1xnYFCZ9Yg?t=203

References

[1]. Cech, T. R. et al. *In vitro* splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–496 (1981)

[2]. Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982)

[3]. Doudna, J. A. & Szostak, J. W. RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**, 519–522 (1989)

[4]. Doherty EA, Doudna JA. The P4-P6 domain directs higher order folding of the Tetrahymena ribozyme core. Biochemistry. 1997 Mar 18;36(11):3159-69. doi: 10.1021/bi962428+. PMID: 9115992.

[5]. Robert W. Holley – Facts. NobelPrize.org. Nobel Media AB 2021. Wed. 20 Jan 2021. https://www.nobelprize.org/prizes/medicine/1968/holley/facts/

[6]. Daniel Herschlag et al., The Story of RNA Folding, as Told in Epochs, Cold Spring Harb Perspect Biol 2018;10:a032433

References

[7]. Denny et al., High-Throughput Investigation of Diverse Junction Elements in RNA Tertiary Folding, Cell 174, 377–390, July 12, 2018

[8]. Mathemws et al., Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, www.pnas.orgcgidoi10.1073pnas.0401799101

Thanks!