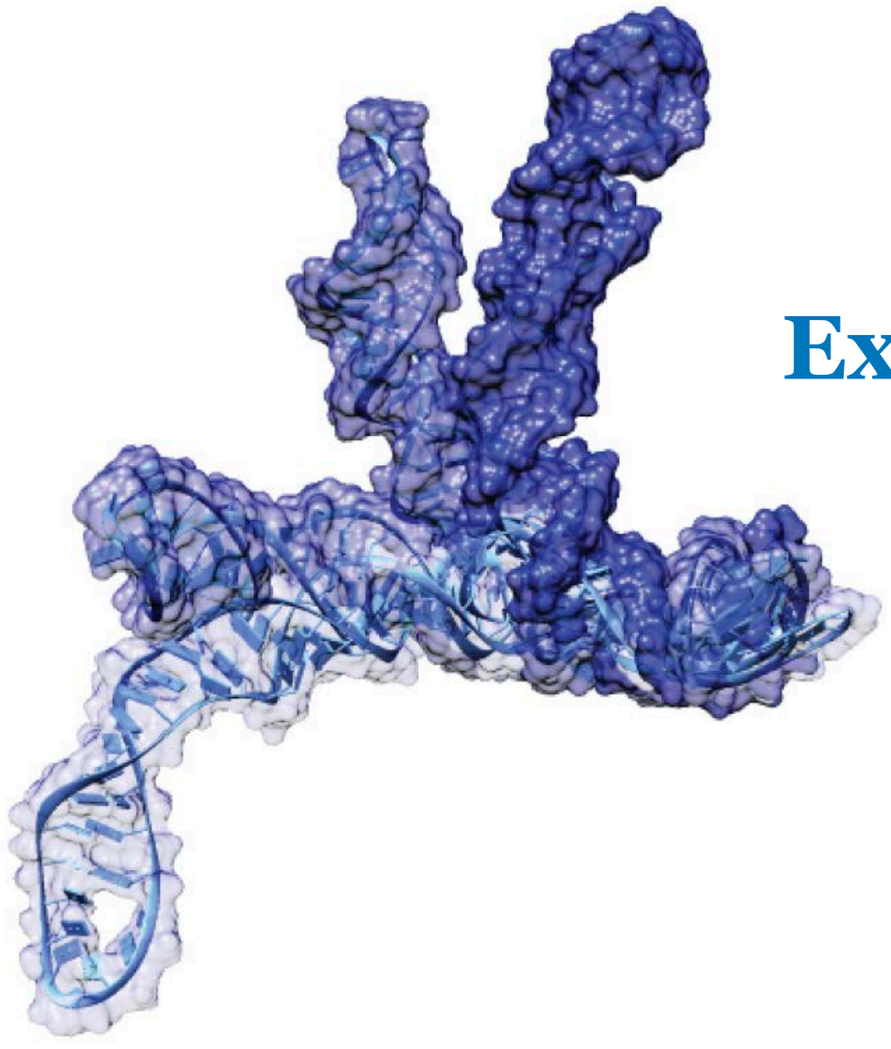
Université Paris Cité



Exploring RNA energy landscapes

Samuela Pasquali

Physics Professor

Laboratoire Biologie Fonctionnelle et Adaptative,

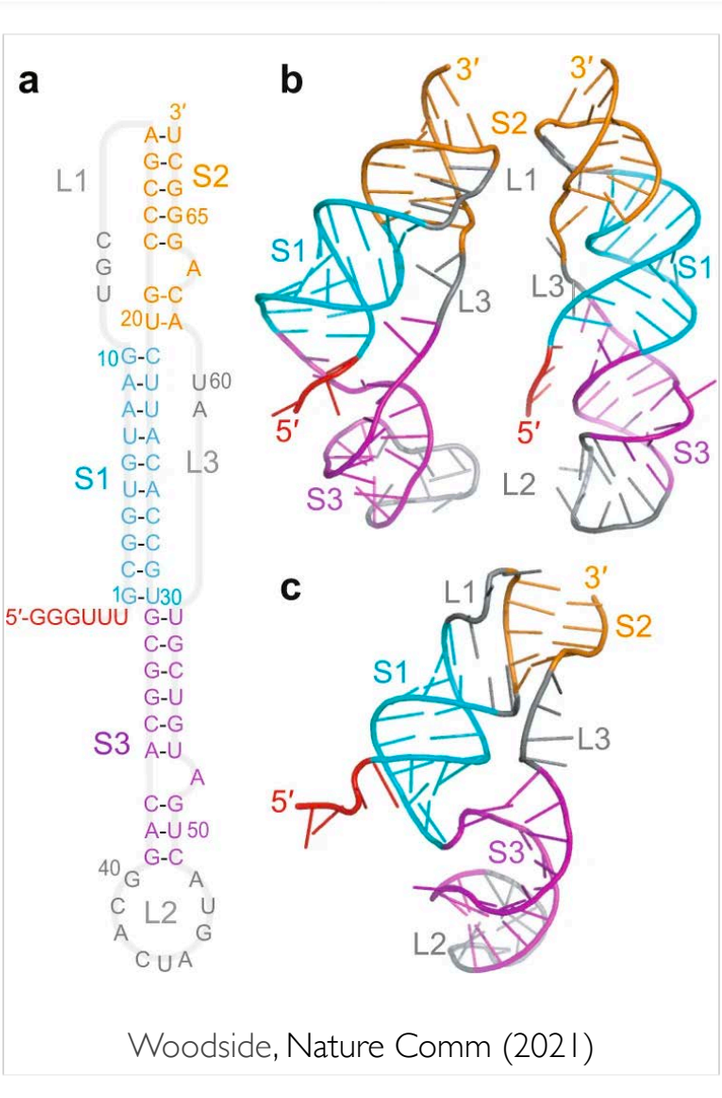
Molecular modeling and drug design team

School of Pharmacy and School of Science

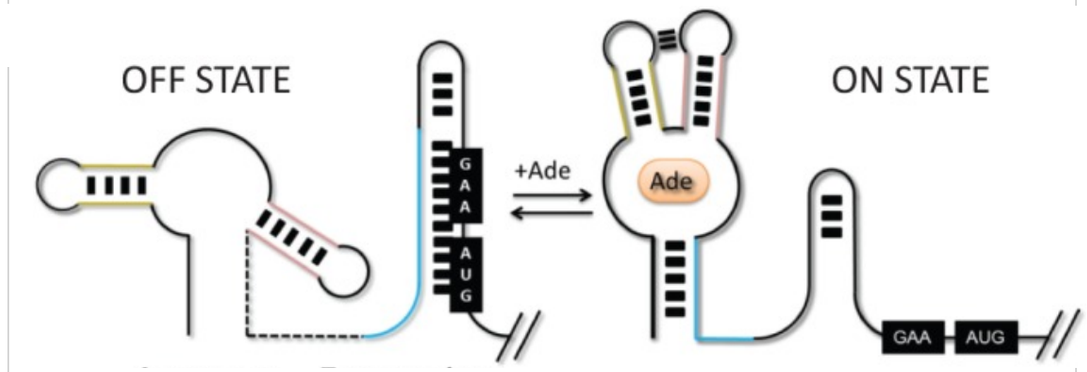
Structural Polymorphism

A given sequence can adopt multiple alternative structures, not just flexibility

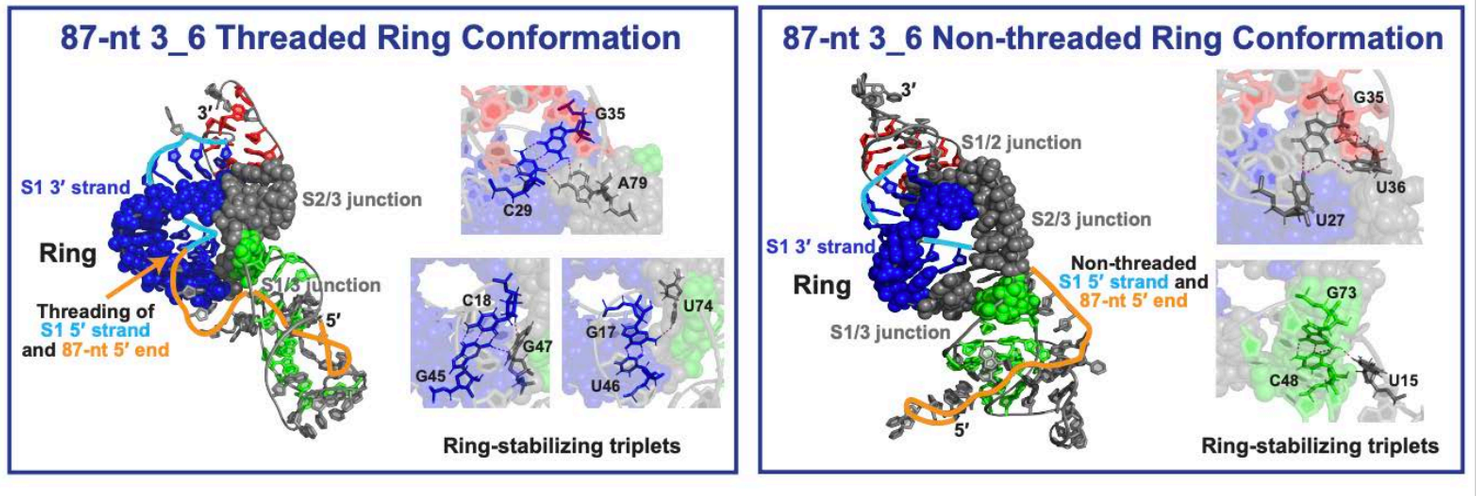
Large scale rearrangements



Woodside, Nature Comm (2021)

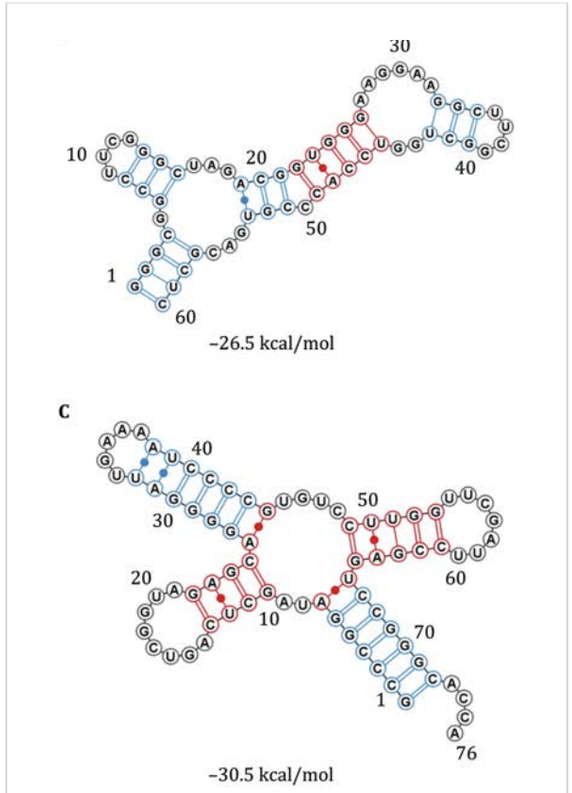


Dalgarno, NAR (2013)



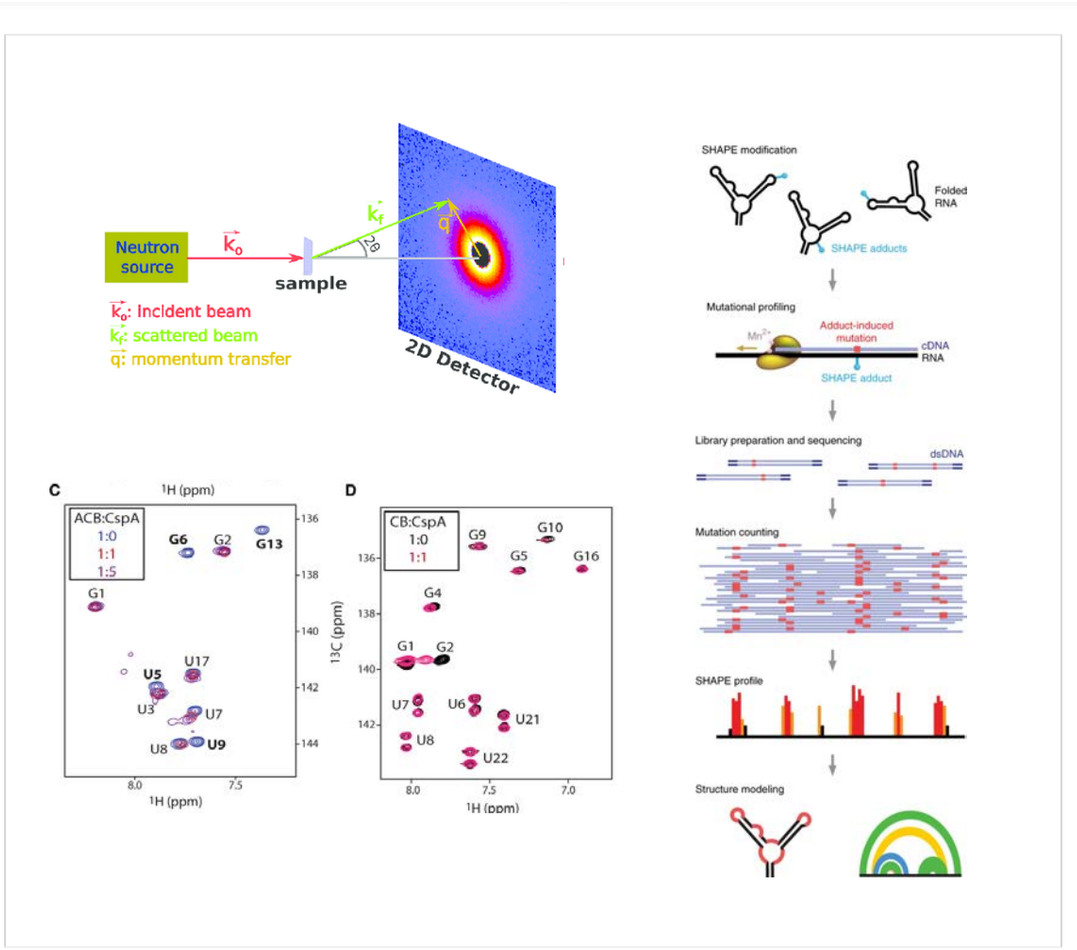
Schlick, Nature Comm (2022)

Role of modeling



Ensembles of 2D structures
canonical BP
~ pseudoknots

gap to be filled by physical modeling



Ensembles of 3D dynamical structures
sensitive to their environment
More or less indirect hints at the 3D structure

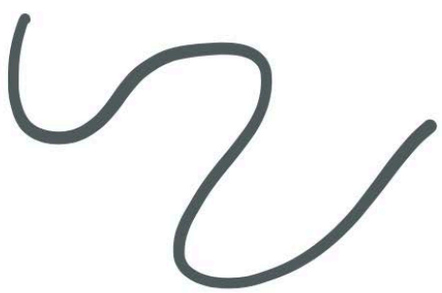
3D structures

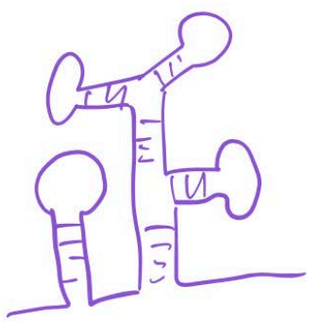
dynamics, thermodynamics, influence of the environment

Modeling

this is not a cow

this is not a cow either





Real life is messy
Pragmatic and greedy approach

Use multiple models at different resolutions

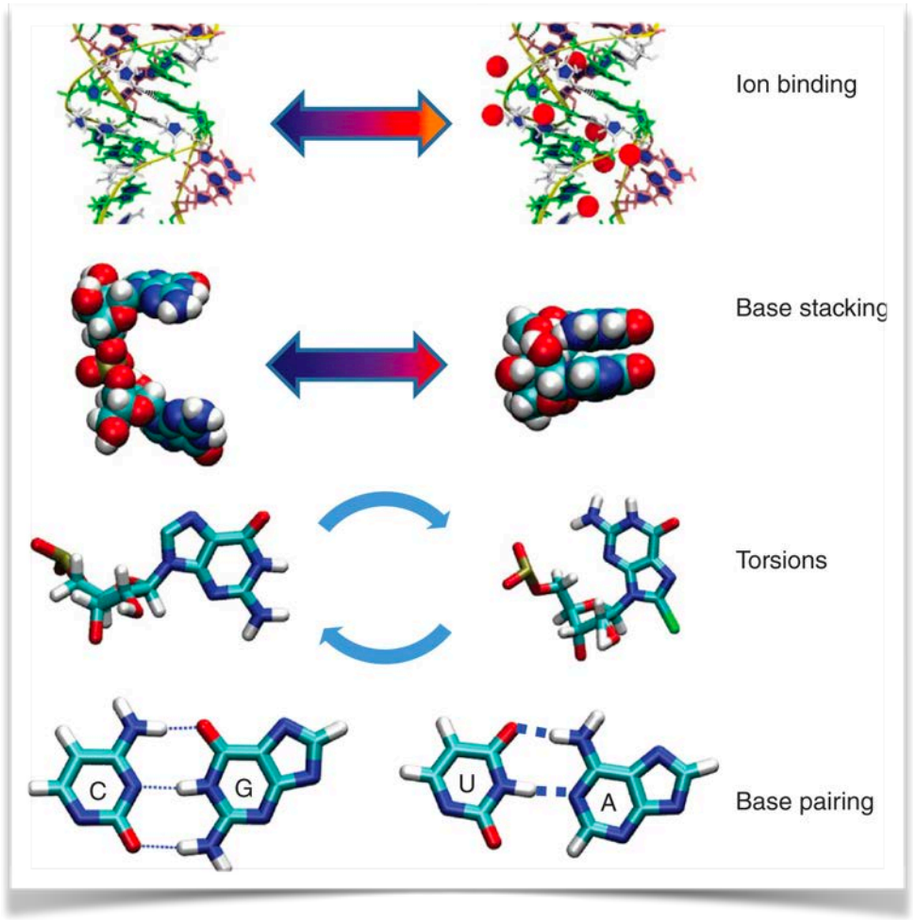
Use different sampling methods

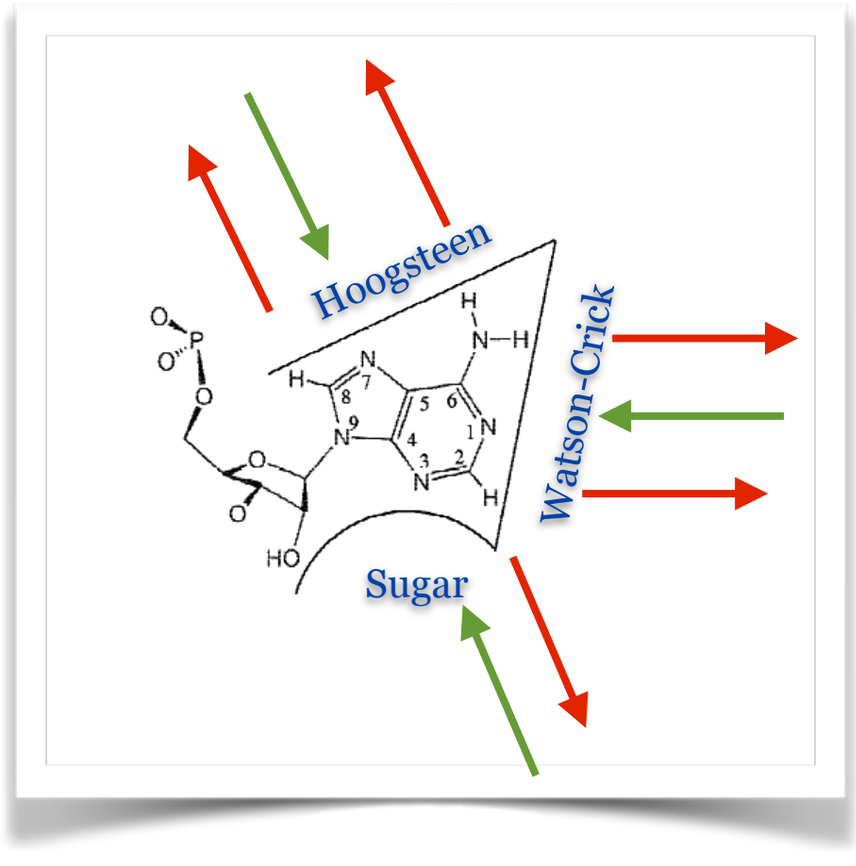
Mix and match

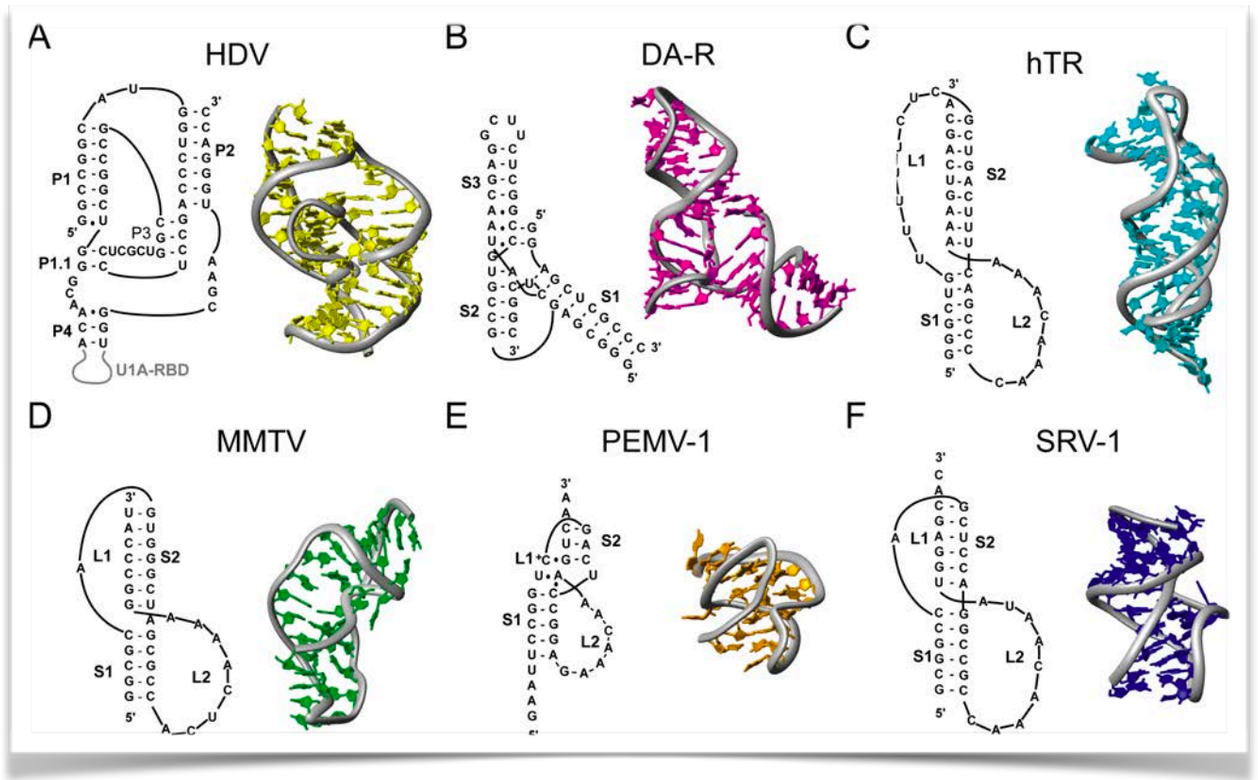
Talk to experimentalists!

Make the model as close as possible to experiments

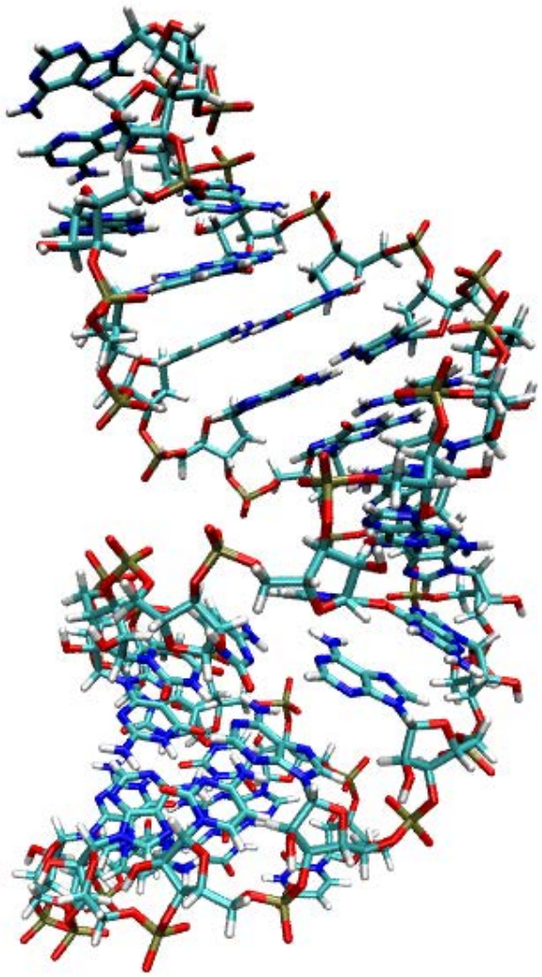
RNA structural features







Physical modeling (how dirty do you want to get?)

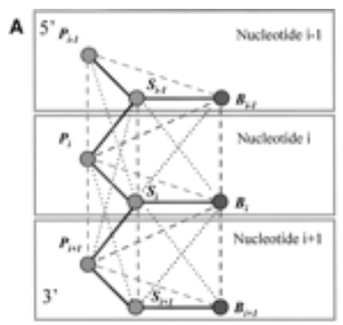


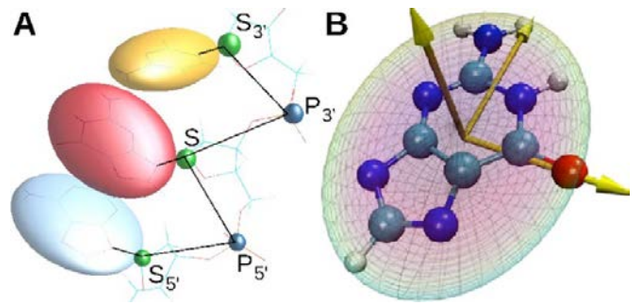
Atomistic description (explicit or implicit solvent)

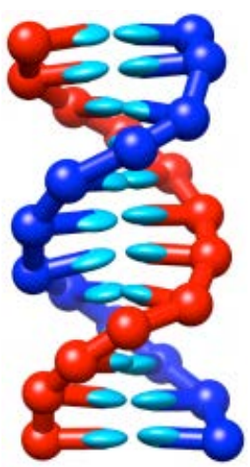
$$V(r^N) = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2$$

$$+ \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

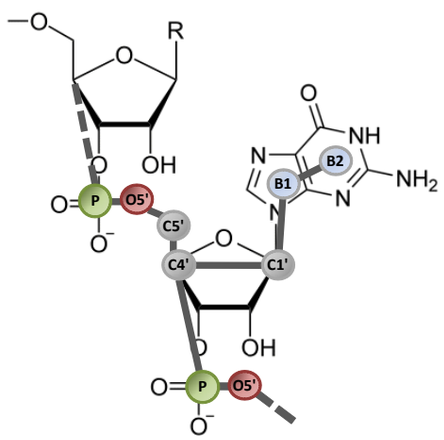
Coarse-graining





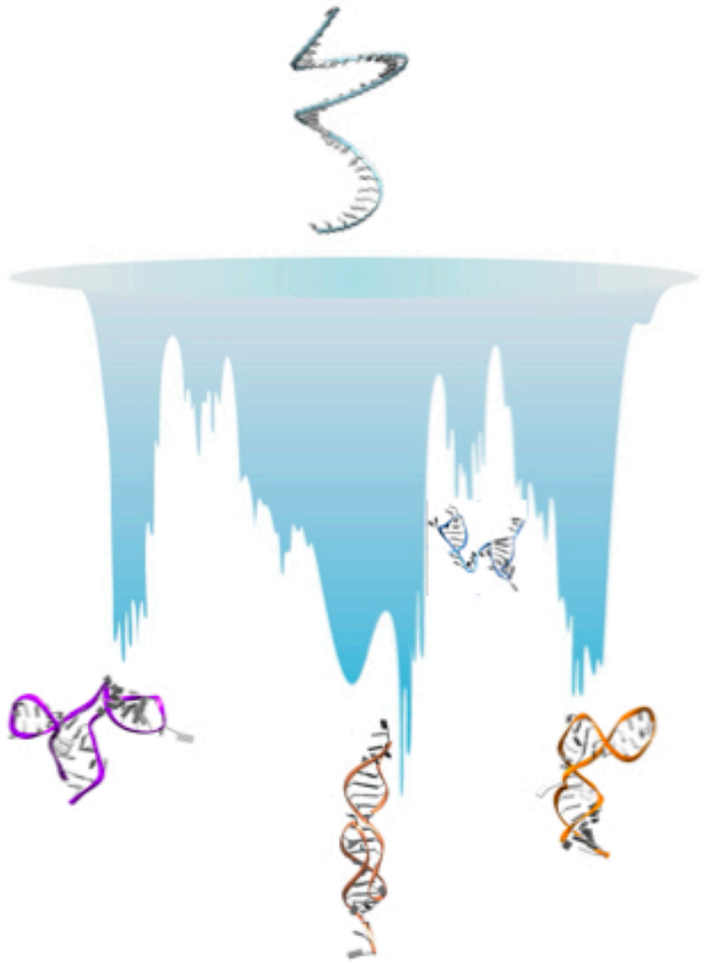


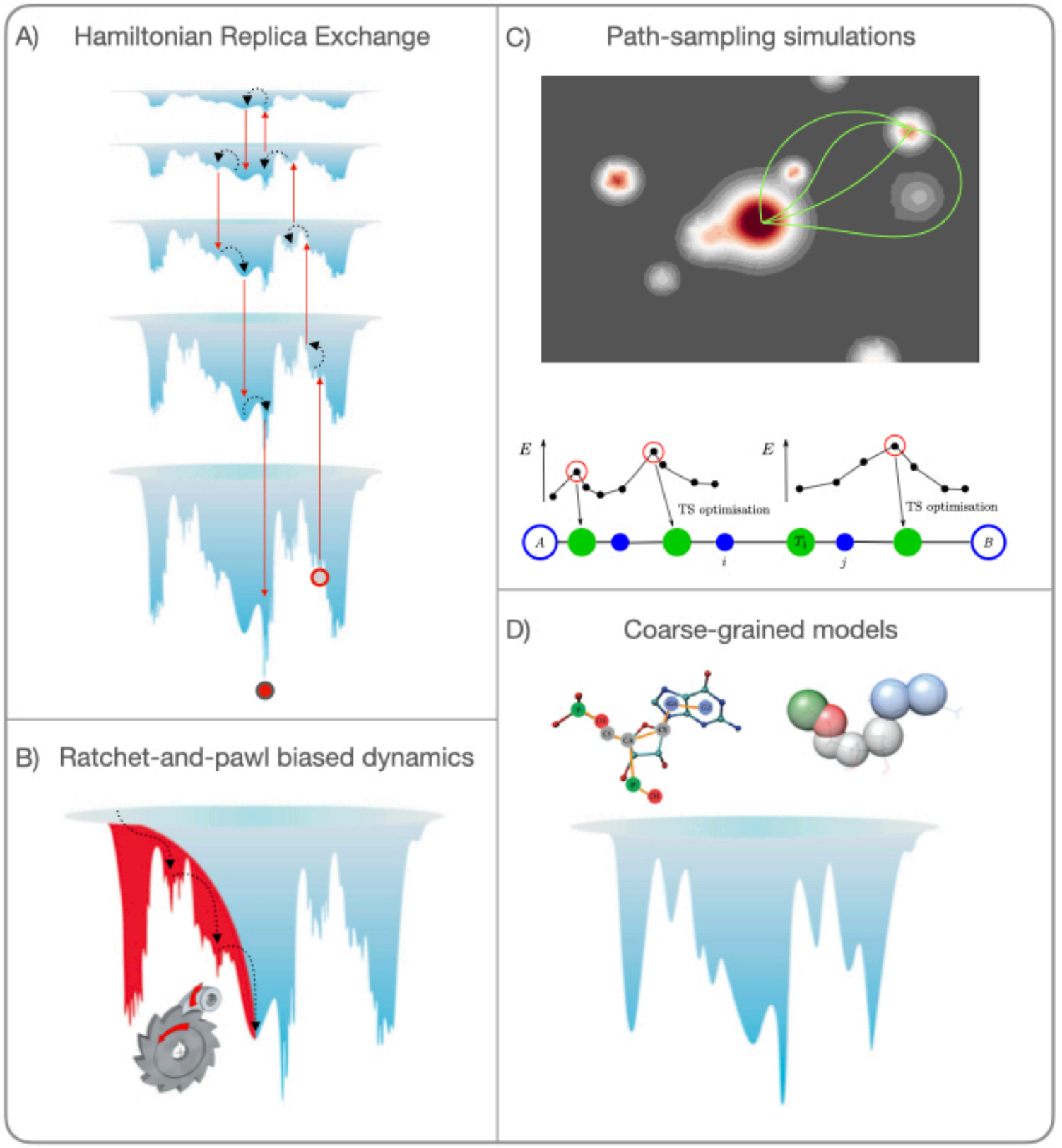
and more...



Use whatever works!

i.e. No method is perfect but if you combine more than one you might have a chance to get things right





Roeder, Stirnemann, Faccioli, Pasquali, QRB Discovery 2022

Ups and downs

H-REX

✓ Full Solvation
Only solute is affected by the energy rescaling
Use standard force fields

✗ Need multiple copies of the system (~30)
Can deal only with small systems (~few dozens nt)

DPS

✓ Sample widely the conformational space
Allows to define “families” of structures (basins)
Allow to obtain kinetic data

✗ Implicit solvent
Can deal only with small systems (~few dozens nt)
Hard time with entropy

rMD

✓ Full Solvation
Use standard force fields
Good statistics on folding trajectories

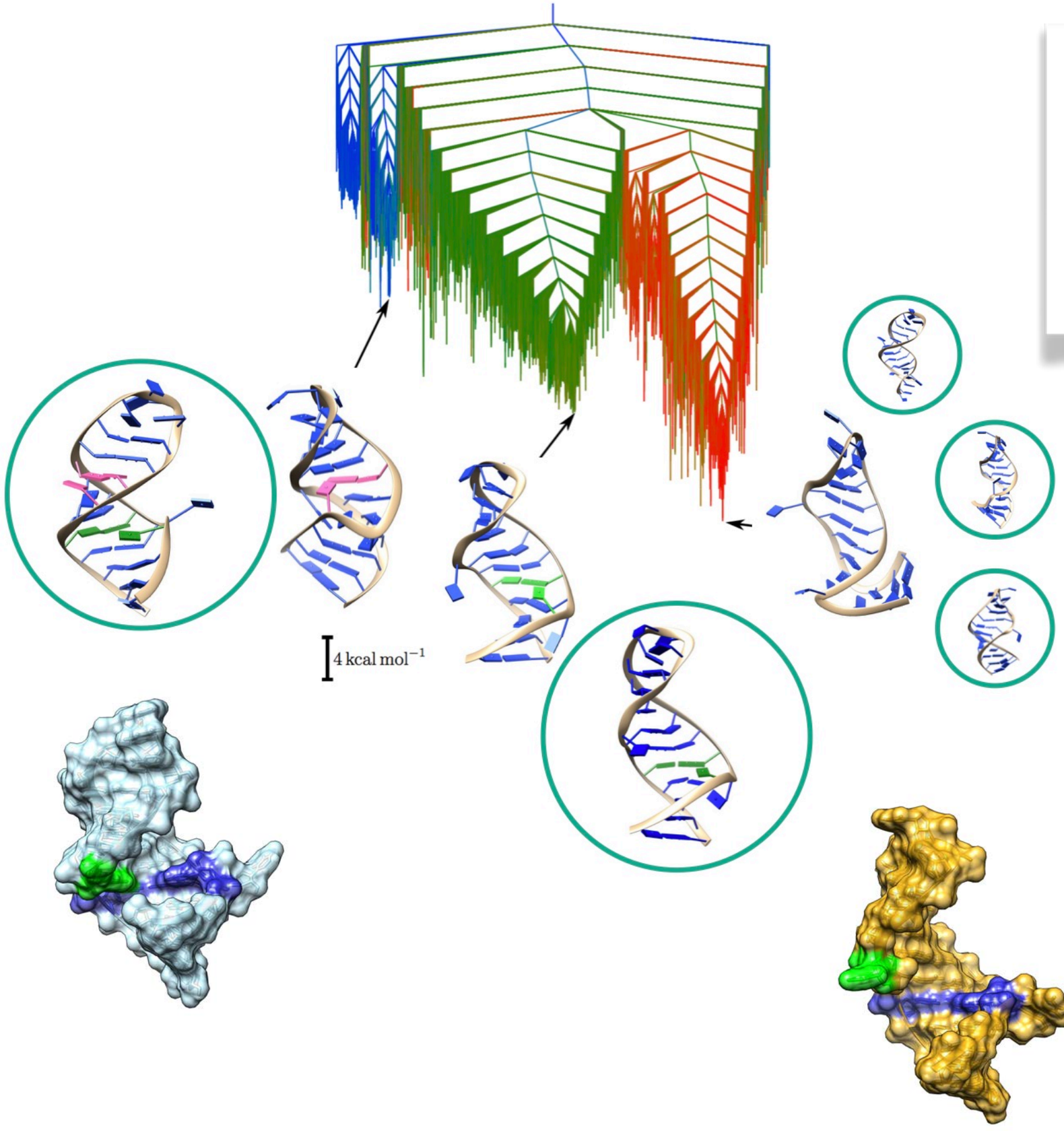
✗ Knowledge of native state required
Cannot study alternative foldings in a single MD

CG models

✓ Allow to study larger systems for longer time scales
Allow to highlight essential elements of the system description

✗ Loss of atomistic details
Implicit solvent
No generally recognized model

Combining strategies: DPS + REST2 REX



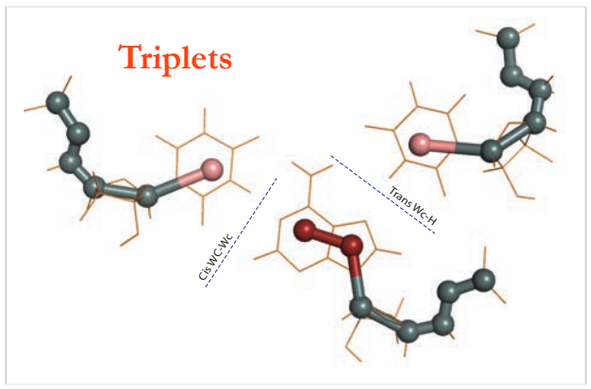
REST2 (24 H-replicas)

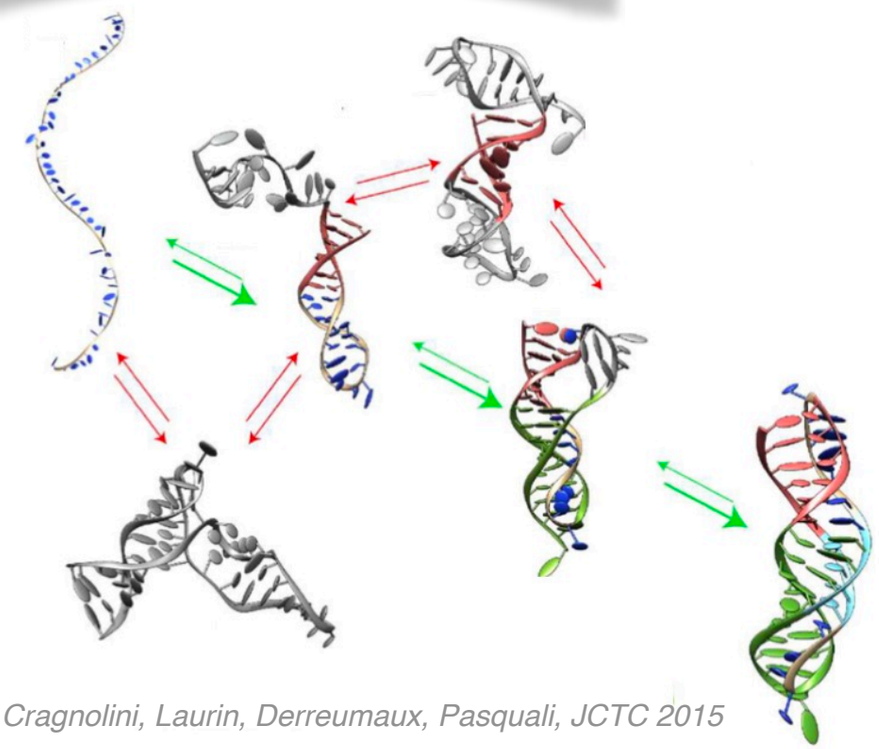
Explicit solvent: interactions with water and ions, solvation effects full entropy contribution.

Slow convergence, expensive calculations, limited exploration of new conformations

Roeder, Stirnemann, Dock-Bregeon, Wales, Pasquali, NAR (2020)

Triple helix folding: CG-MD + rMD



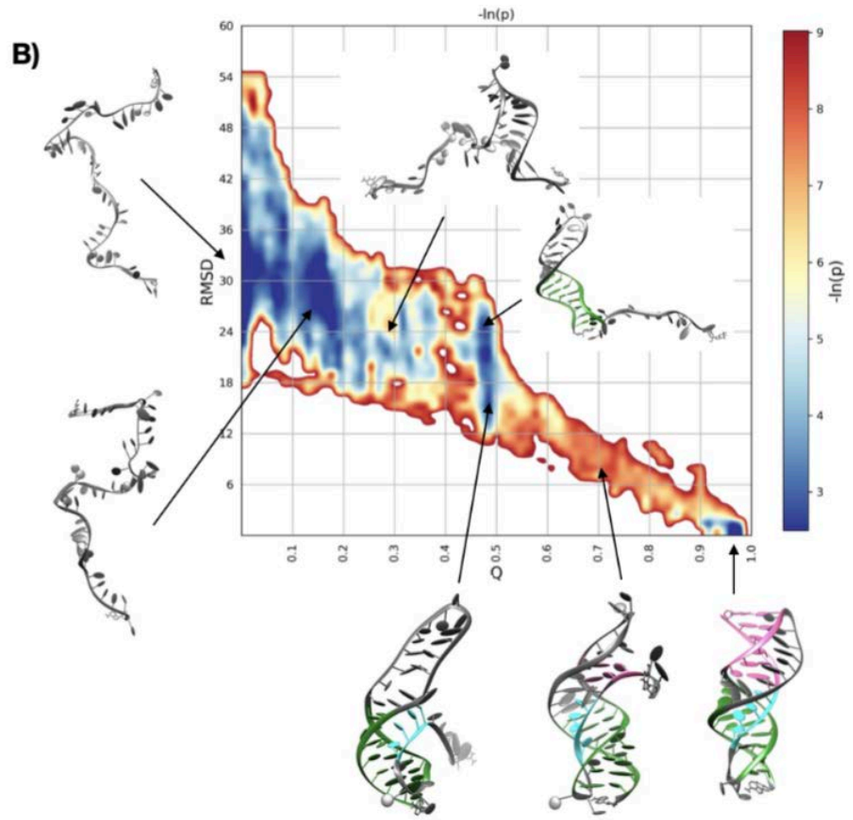


Cragnolini, Laurin, Derreumaux, Pasquali, JCTC 2015

rMD (100 short simulations)

Full solvation, fast folding.

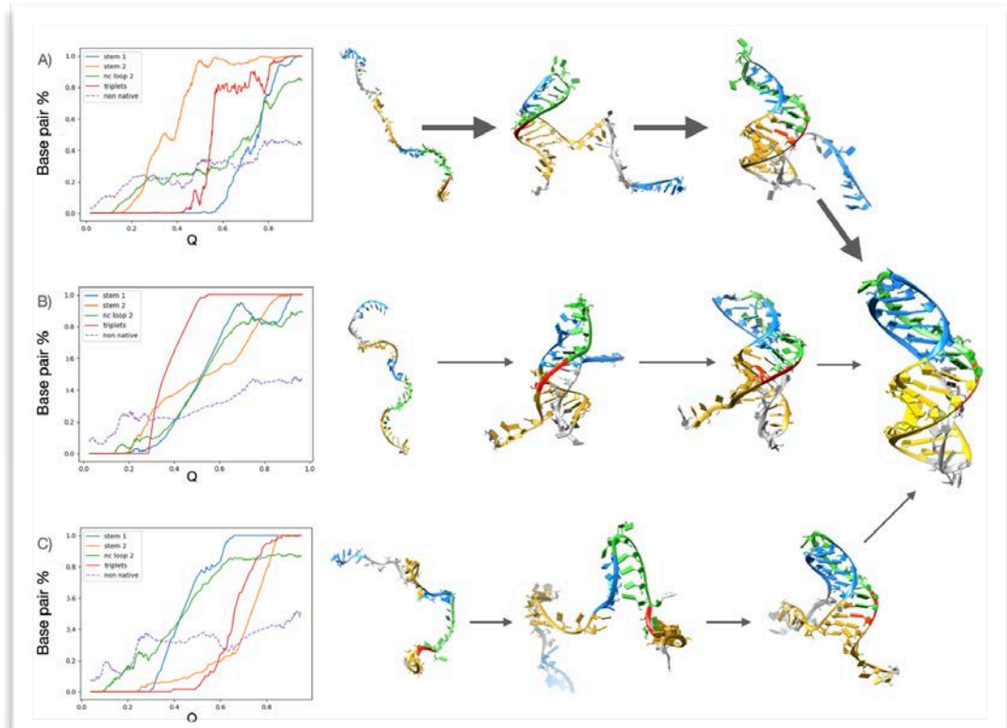
Strong bias, need of a good reaction coordinate.



HiRE-RNA (64 replicas REMD)

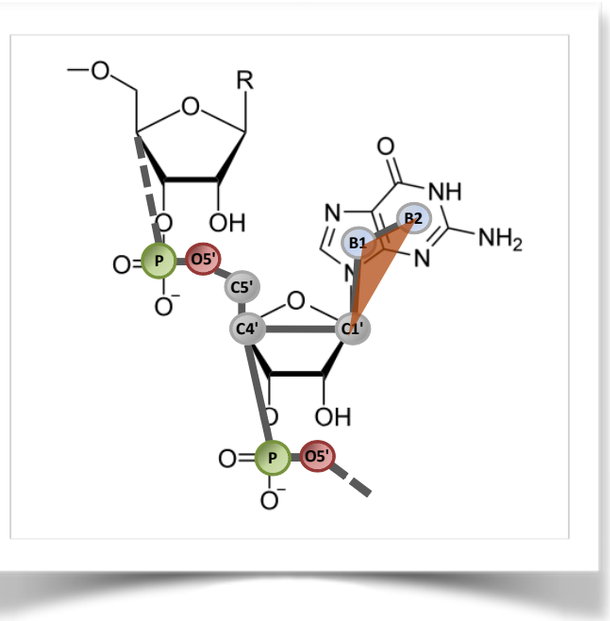
No bias, large exploration of conformational space

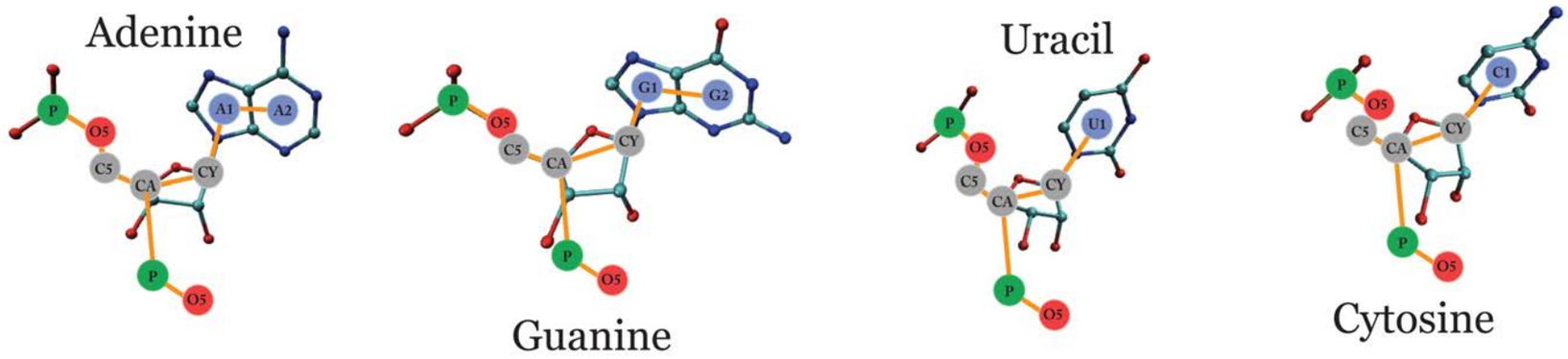
Approximate force field, no solvent nor ions



Lazzari, Micheletti, Pasquali, Faccioli, preprint 2022

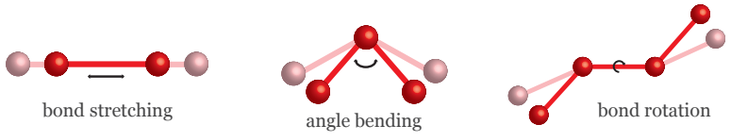
Coarse-grained RNA modeling: HiRE-RNA





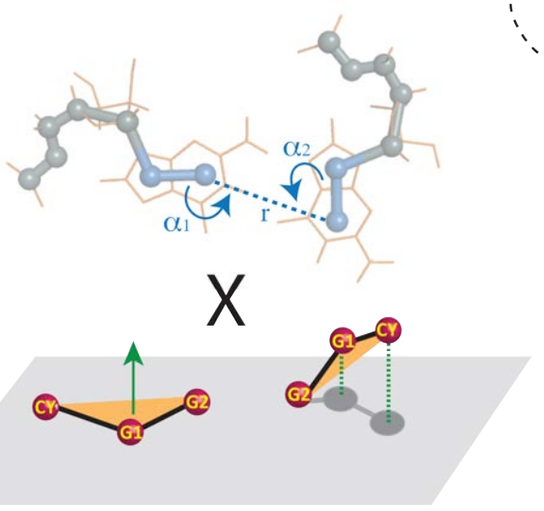
$$E = E_{\text{local}} + E_{\text{ex vol}} + E_{\text{BP}} + E_{\text{electrostatics}} + E_{\text{stacking}}$$

harmonic
statistical parameters

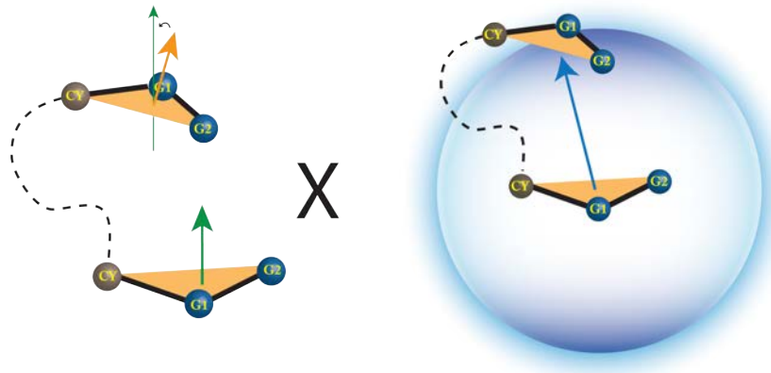


Planarity
Non-canonical pairs

$$E_{\text{BP}} = E_{\text{HB}} \times E_{\text{plane}}$$

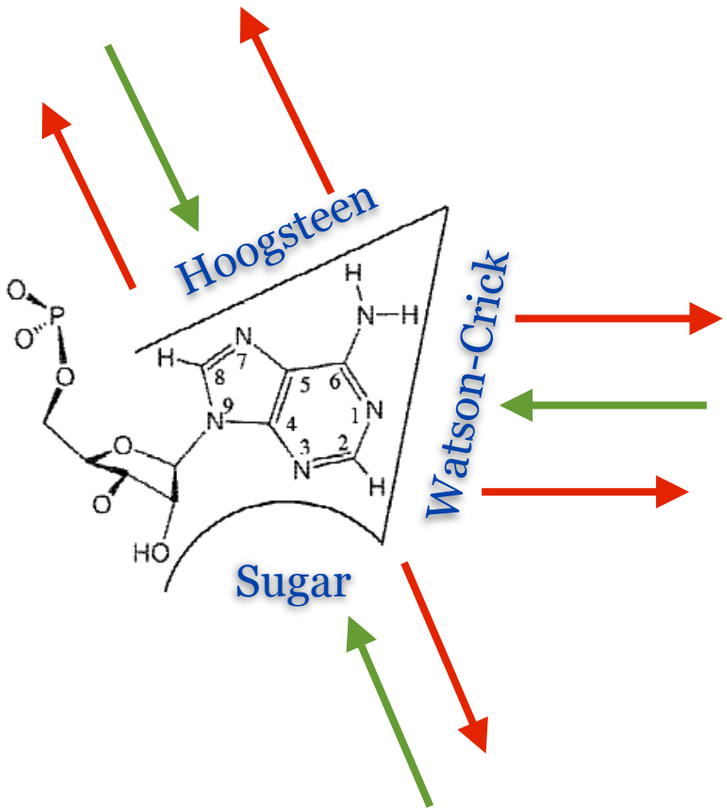


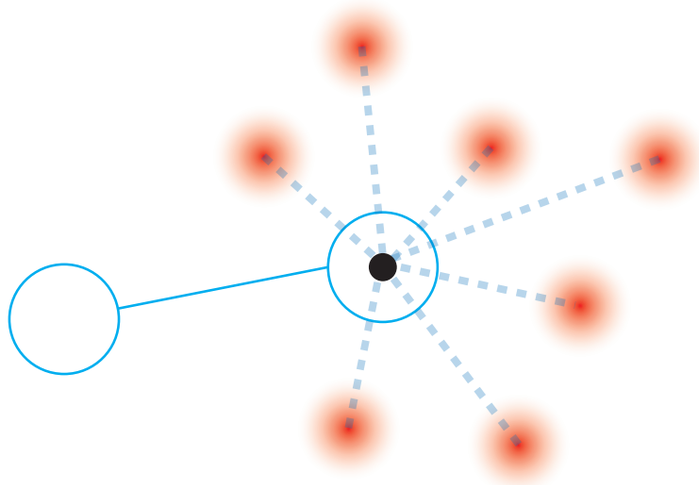
Base orientation

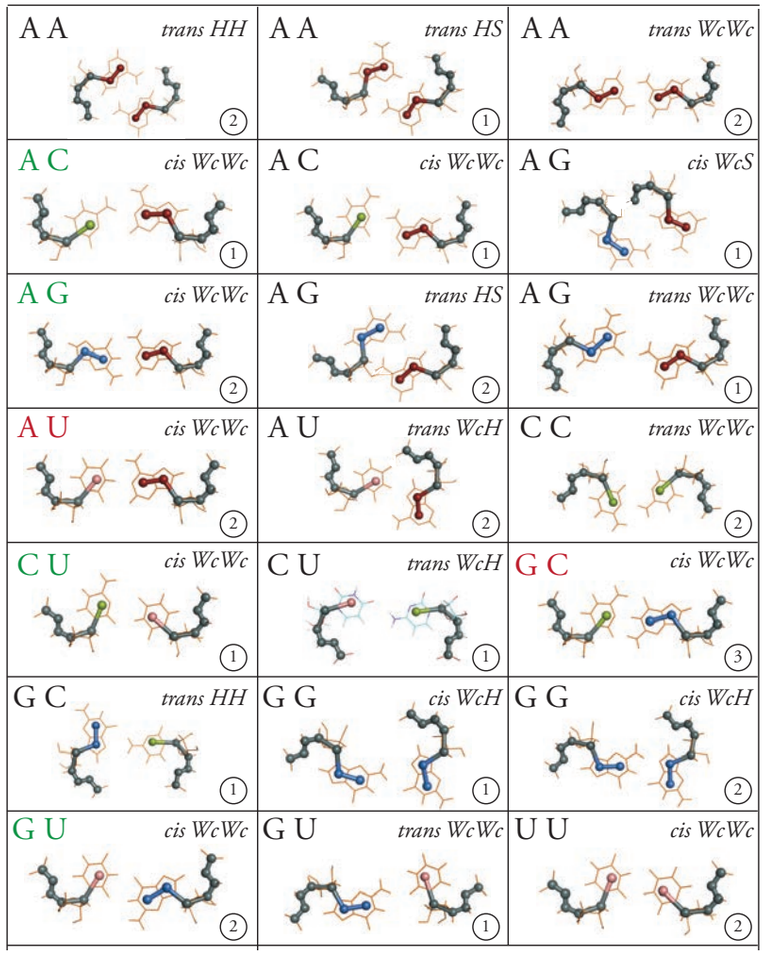


Pasquali, Derreumaux, JPCB 2010
Cragnolini, Derreumaux, Pasquali, JCTC 2015

Base pairing







~150 different experimentally accounted interactions 1, 2 or 3 HB per pair

Stacking

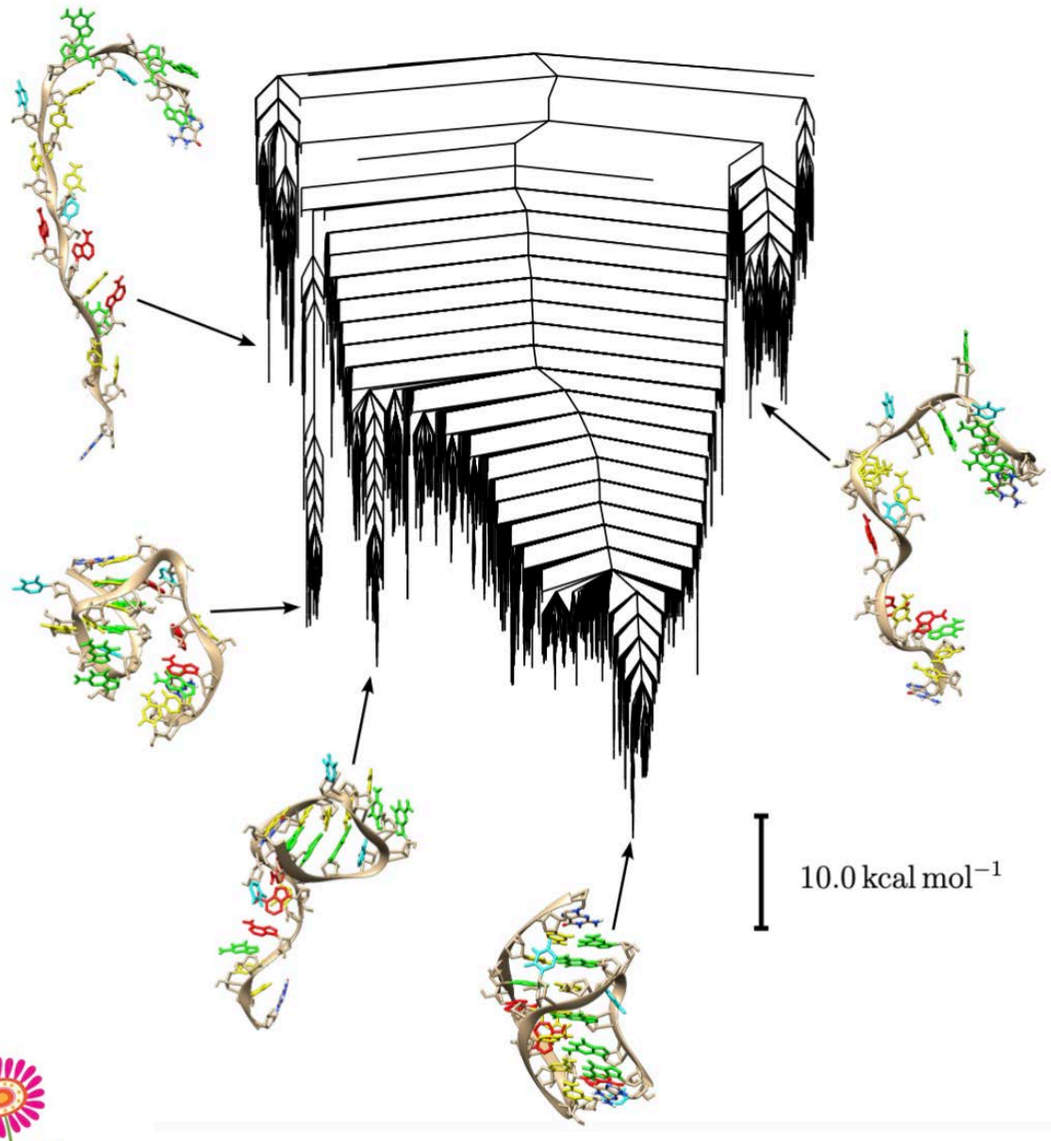
$$E_{STK} = -\varepsilon e^{-\sigma(r_{ij}-r_0)^2}\,(\overrightarrow{n_i}\cdot\overrightarrow{n_j})^2\,(e^{-\alpha(1-\cos(\theta-\theta_0)^2}+e^{-\alpha(1-\cos(\pi-\theta-\theta_0)^2})$$

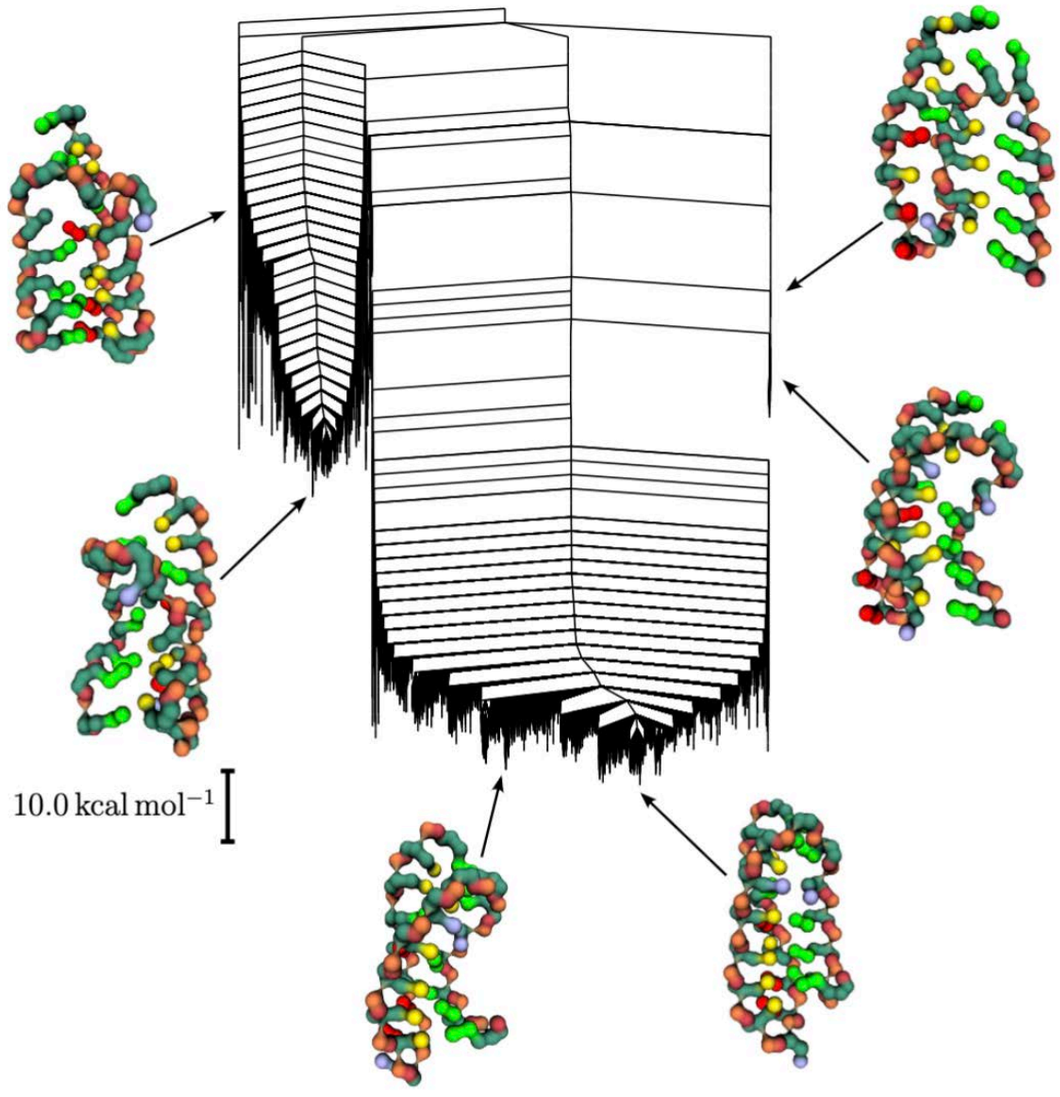
Semi-sequence dependent parameters : purine-purine, purine-pyrimidine, pyrimidine-pyrimidine

> 200 parameters

Force field optimization

Comparison of atomistic and CG energy landscapes ➜ optimize to match energy differences





Can we use EL comparisons also to assess the quality of atomistic force-fields?

Generalized landscapes

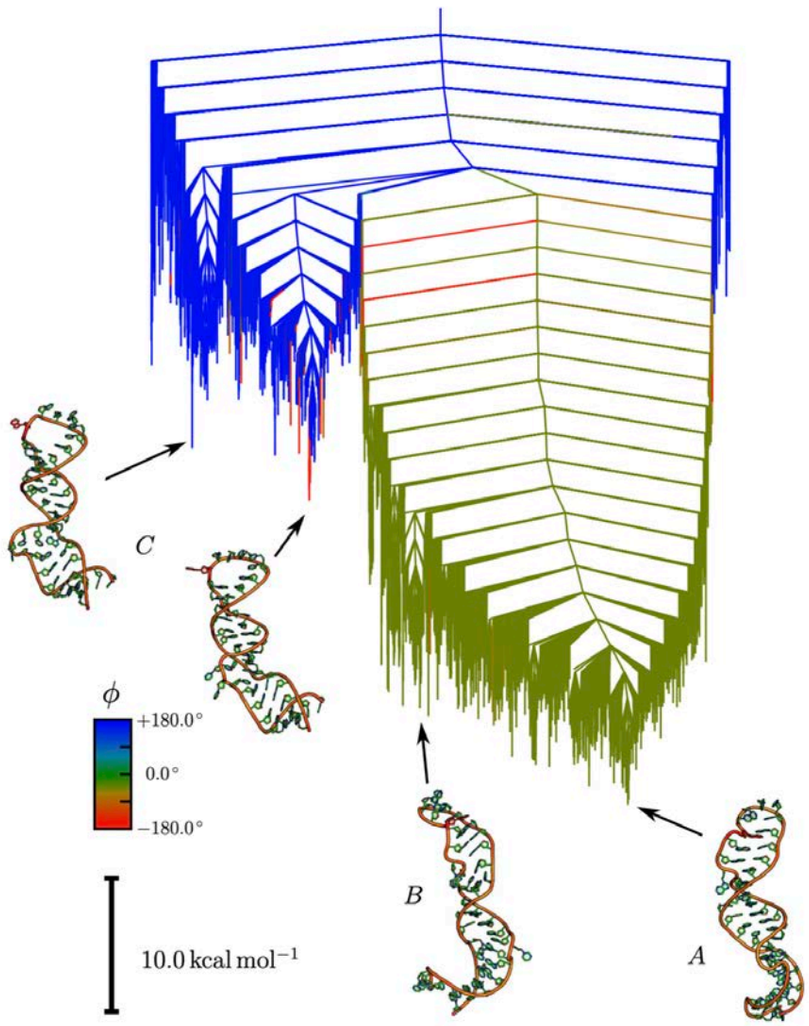
Discrete variables

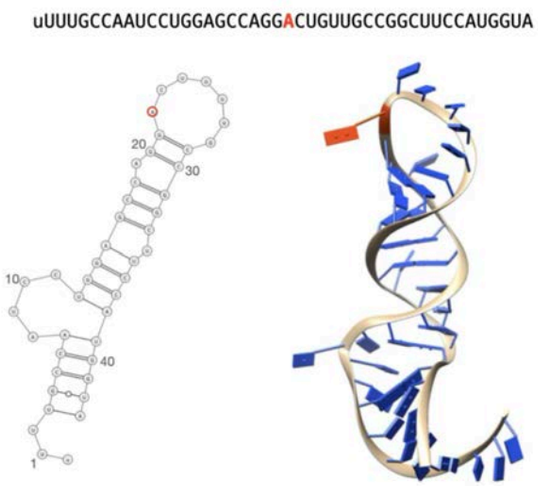
mutations, chemical modifications

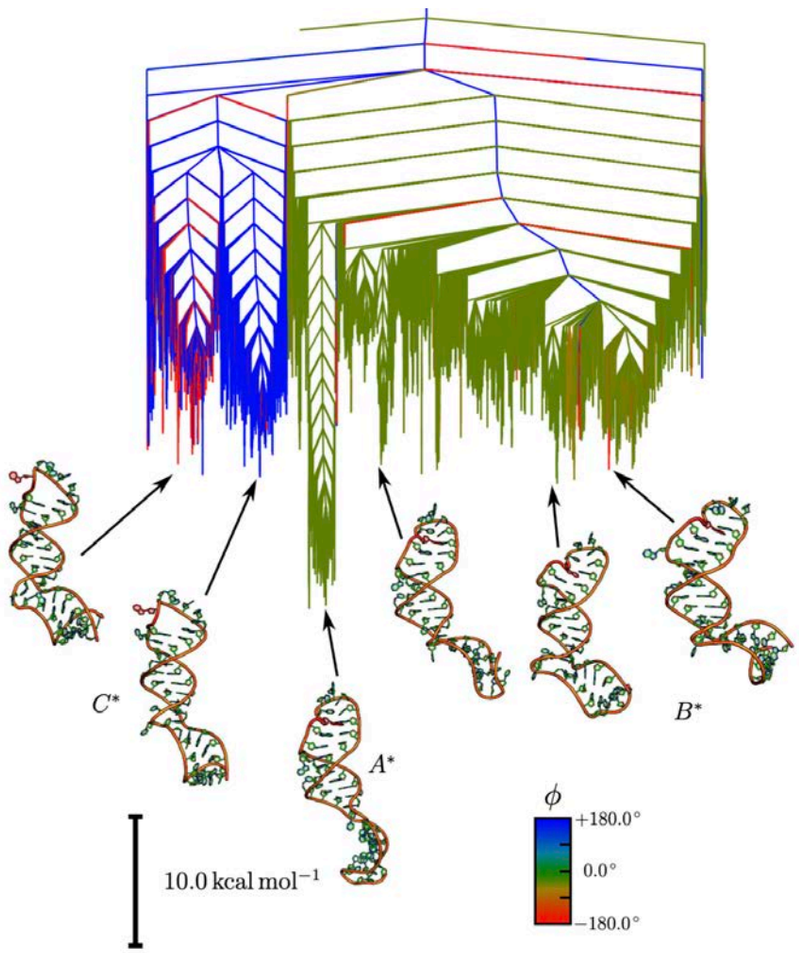
methylation

Continuous variables

Temperature, pH, ionic conditions...







Roeder, Barker, Whitehouse, Pasquali, PLOS Computational Bio (2022)

Generalized landscapes

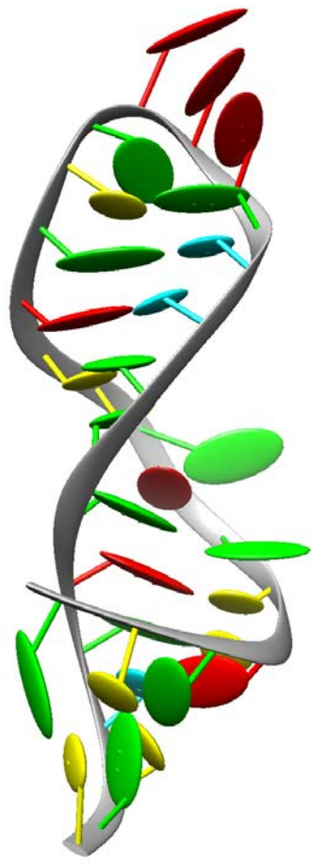
Discrete variables

mutations, chemical modifications

Continuous variables

Temperature, pH, ionic conditions...

pH



exposed ⇢ protonable
pK_a ~ isolated

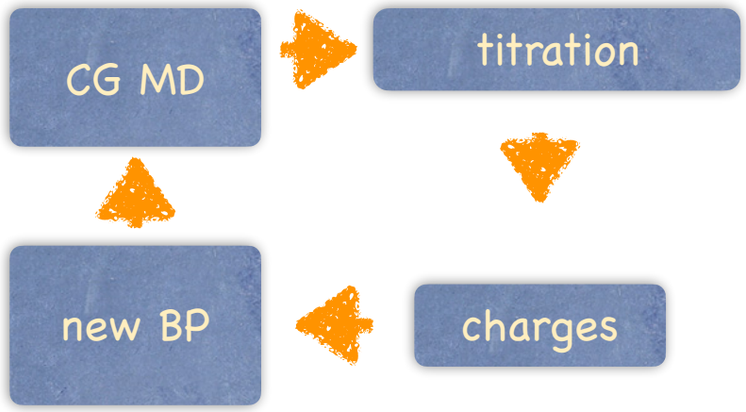
neutral paired ⇢ protected
lower pKa

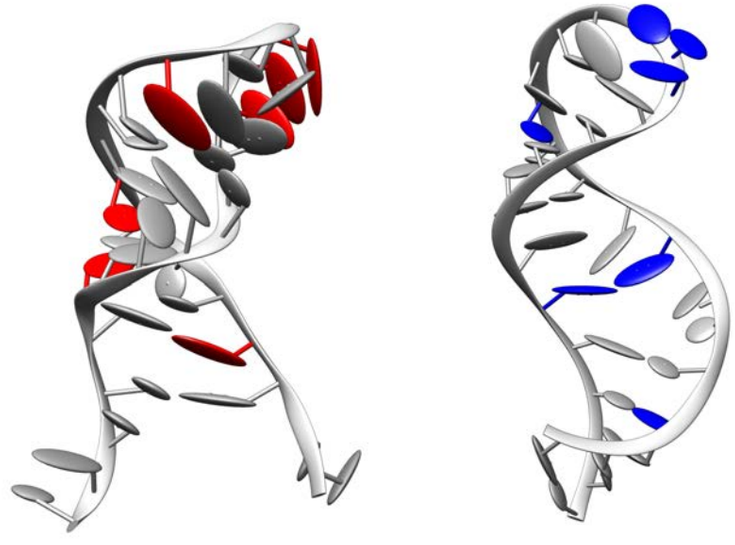
N^+ paired ⇢ protected
higher pKa

$$w_{TK} \approx \frac{e^2}{8\pi\epsilon_0\epsilon_r} \sum_{i>j}^{N_p} \left(\frac{z_i z_j}{r_{ij}} - \frac{Z_p^2 \kappa}{2(1+\kappa b)} \right) \pm (pH - pK_a)$$

Texeira, Lund, Barroso da Silva, JCTC, 2010

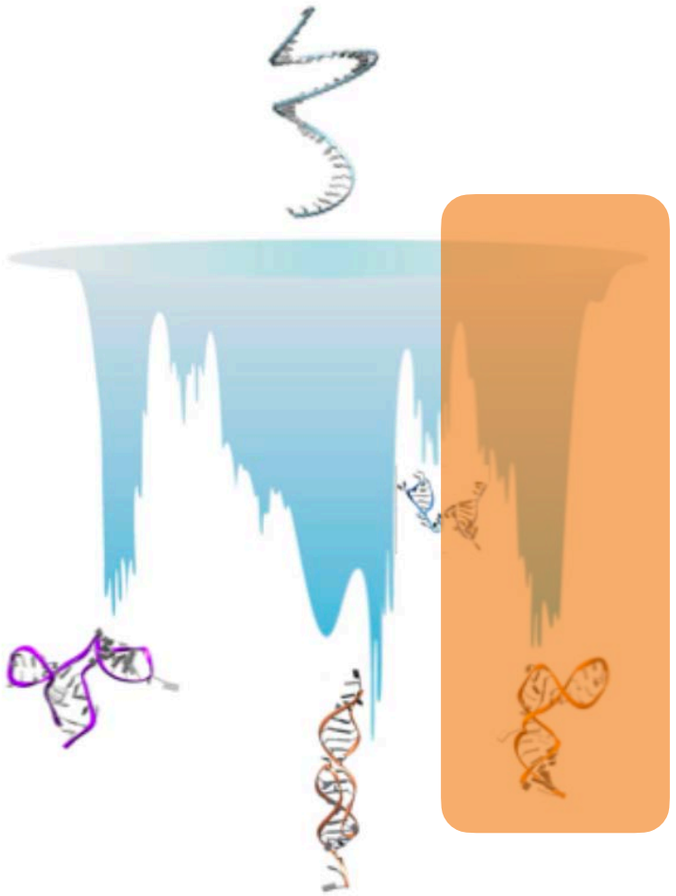
Fast MC protonation scheme





Pasquali, Frezza, Barroso da Silva, Interface Focus (2019)

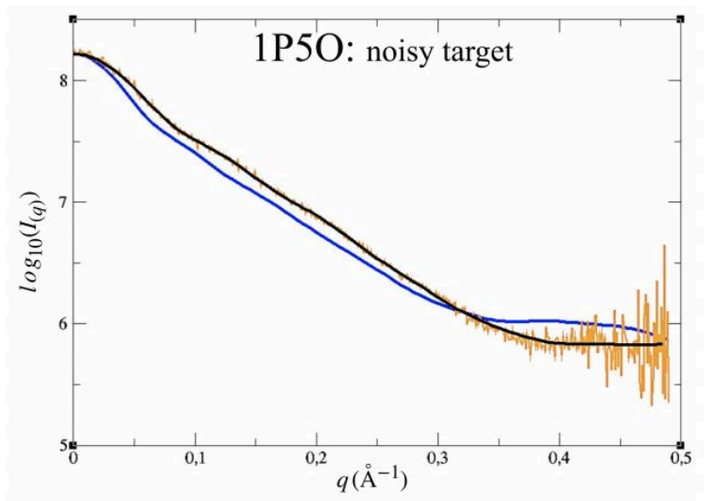
Including experimental data and experimental conditions

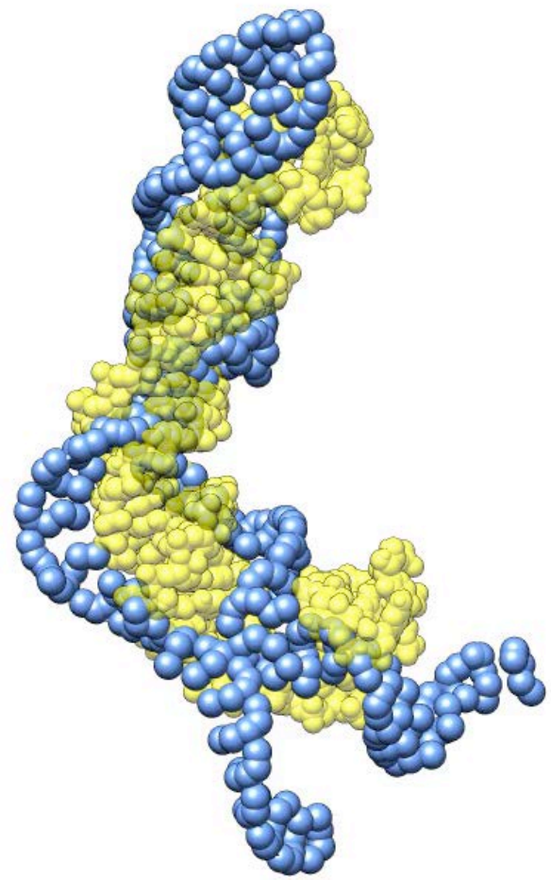


Experimental data restricts the conformational space to be explored

- What does the experiment actually measure?
- Is it possible to compute the outcome from hypothetical structures?
- Can we bias simulations based on this account?

SAXS





Mazzanti, Alferkh, Frezza, Pasquali JCTC (2021)

Thanks to...

Chemistry Department
University of Cambridge

David Wales
Tristan Cragnolini*

King's College London

Konstantin Roeder

Università di Trento

Pietro Faccioli

Institute of Biophysics
Czech Republic's Academy of Sciences

Jiri Sponer
Petr Stadlbauer

SISSA

Cristian Micheletti

Laboratoire de Biochimie Théorique
Université Paris Cité

Philippe Derreumaux
Marc Baaden
Guillaume Stirnemann

Laboratoire CiTCoM
Université Paris Cité

Elisa Frezza
Liuba Mazzanti*
Bruno Sargueil

University of Sao Paolo

Fernando Barroso da Silva