Exploring RNA flexibility through all atoms and coarse-grained modeling

Samuela Pasquali

Laboratoire CiTCoM - Cibles Thérapeutiques et Conception de Médicaments
Faculté de Pharmacie, Université de Paris

Université de Paris

RNA zoo

20-25 ~30 40 60 80 100 300

microRNA
siRNA

piRNA

tRNA
riboswitches

ribozymes

snoRNA

mRNA

tRNA

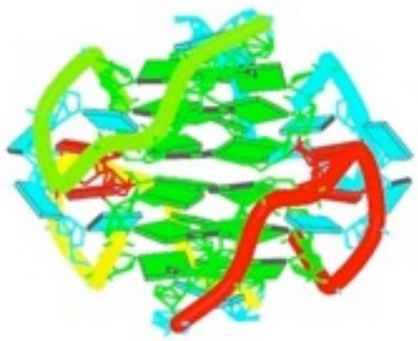
telomerase

viral fragment

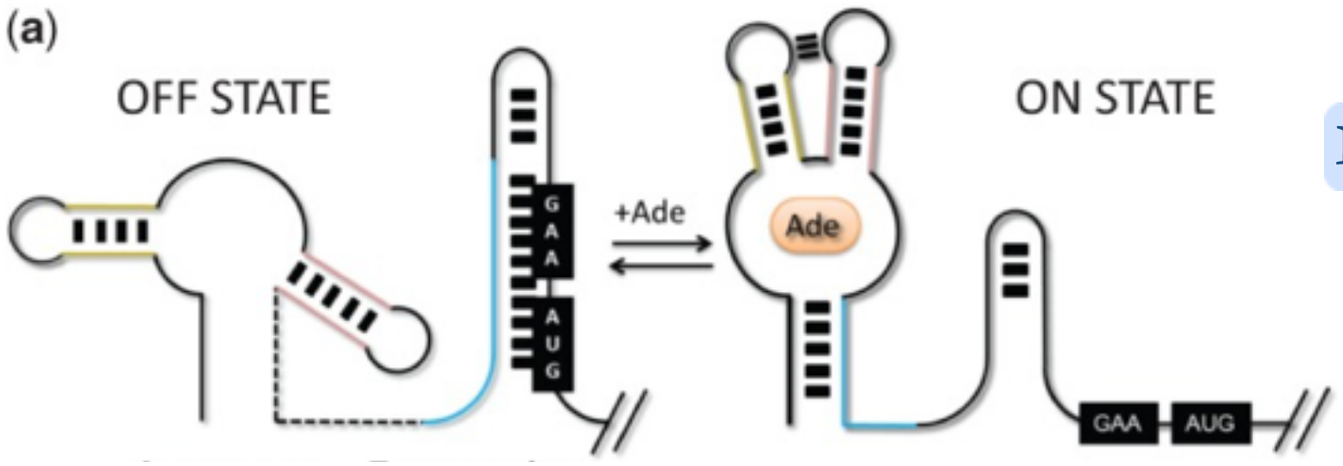
riboswitch

ribozyme

triple helix



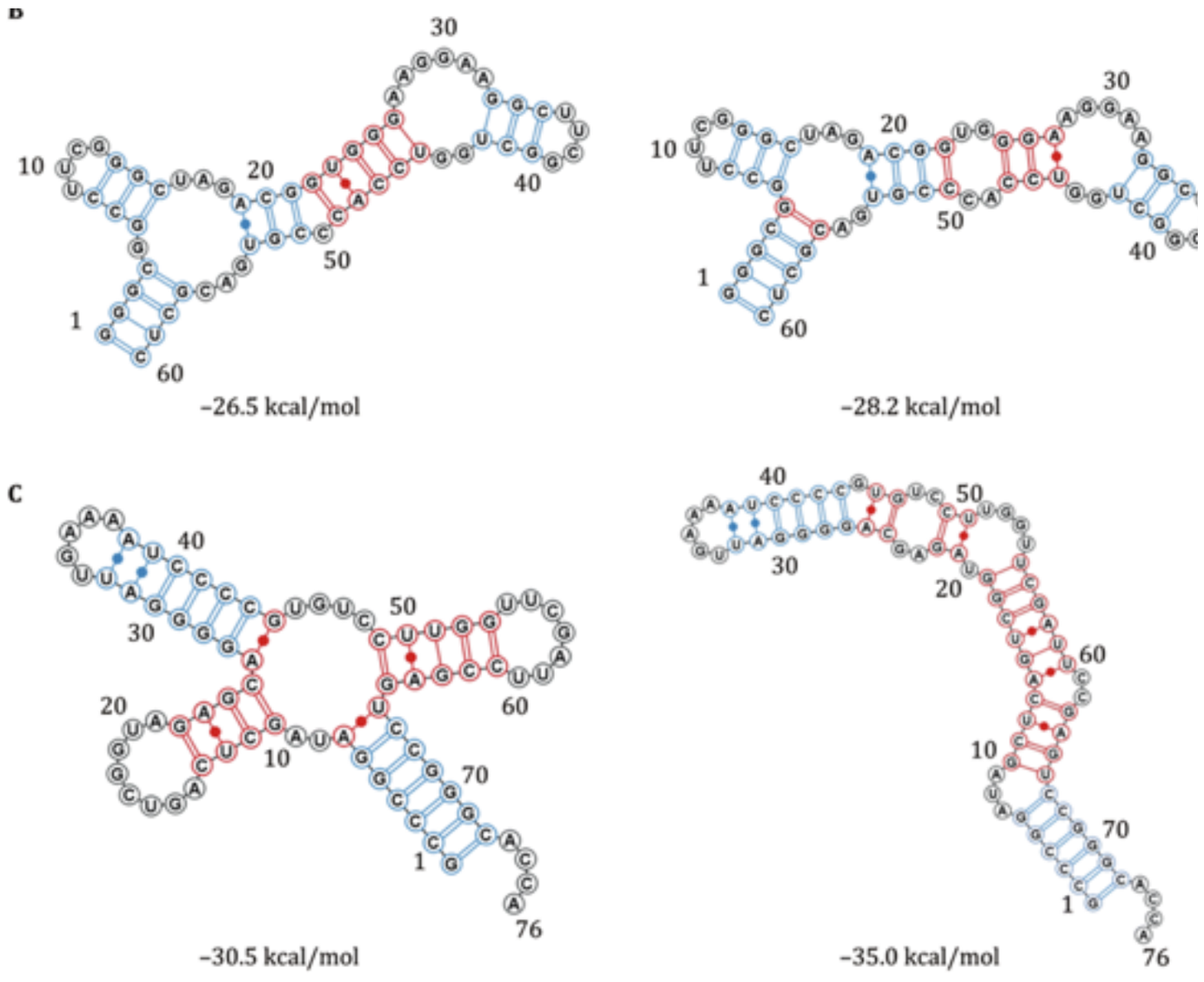
RNA flexibility



taken from Dalgarno, NAR (2013)

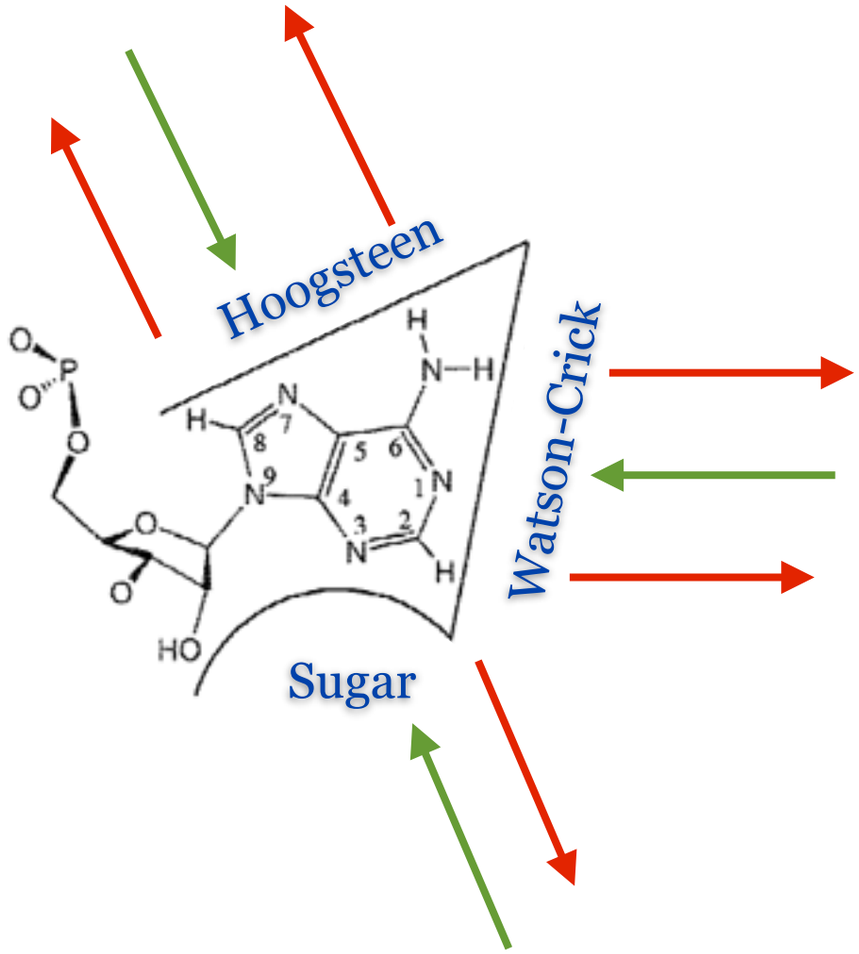
Large rearrangements

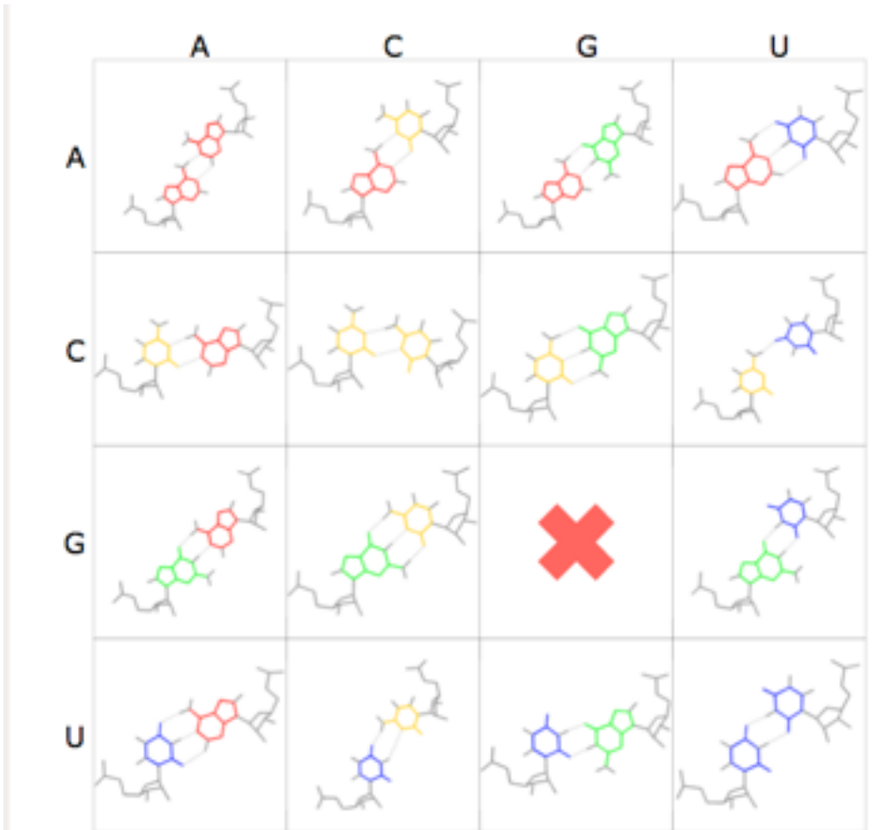
Small 2D changes result in very different topologies

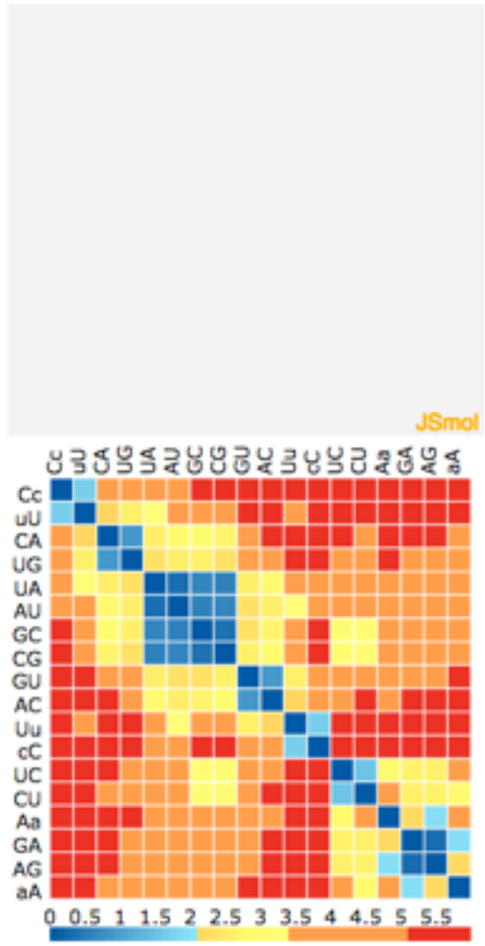


taken from Zeng, Biophysics Reports (2018)

Non-canonical pairs



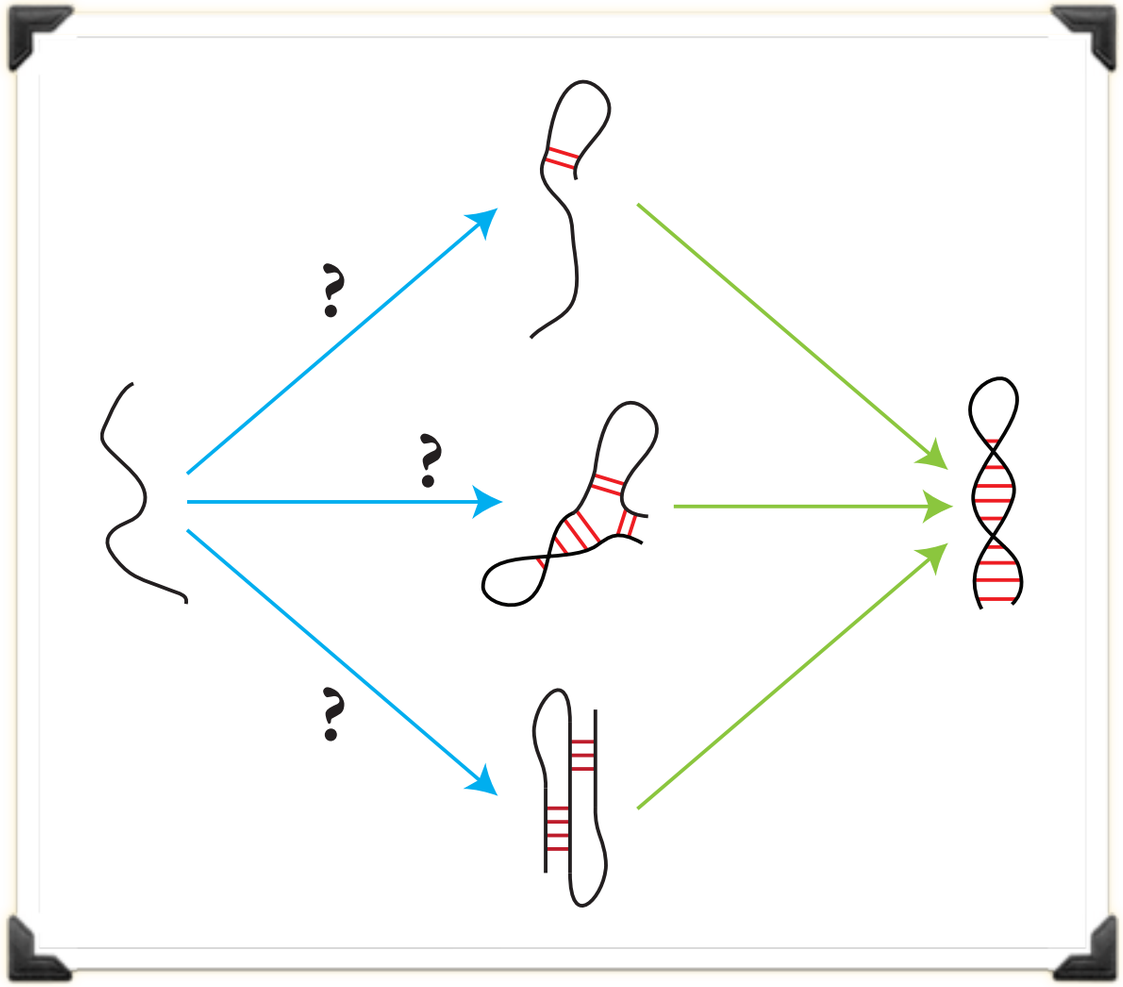


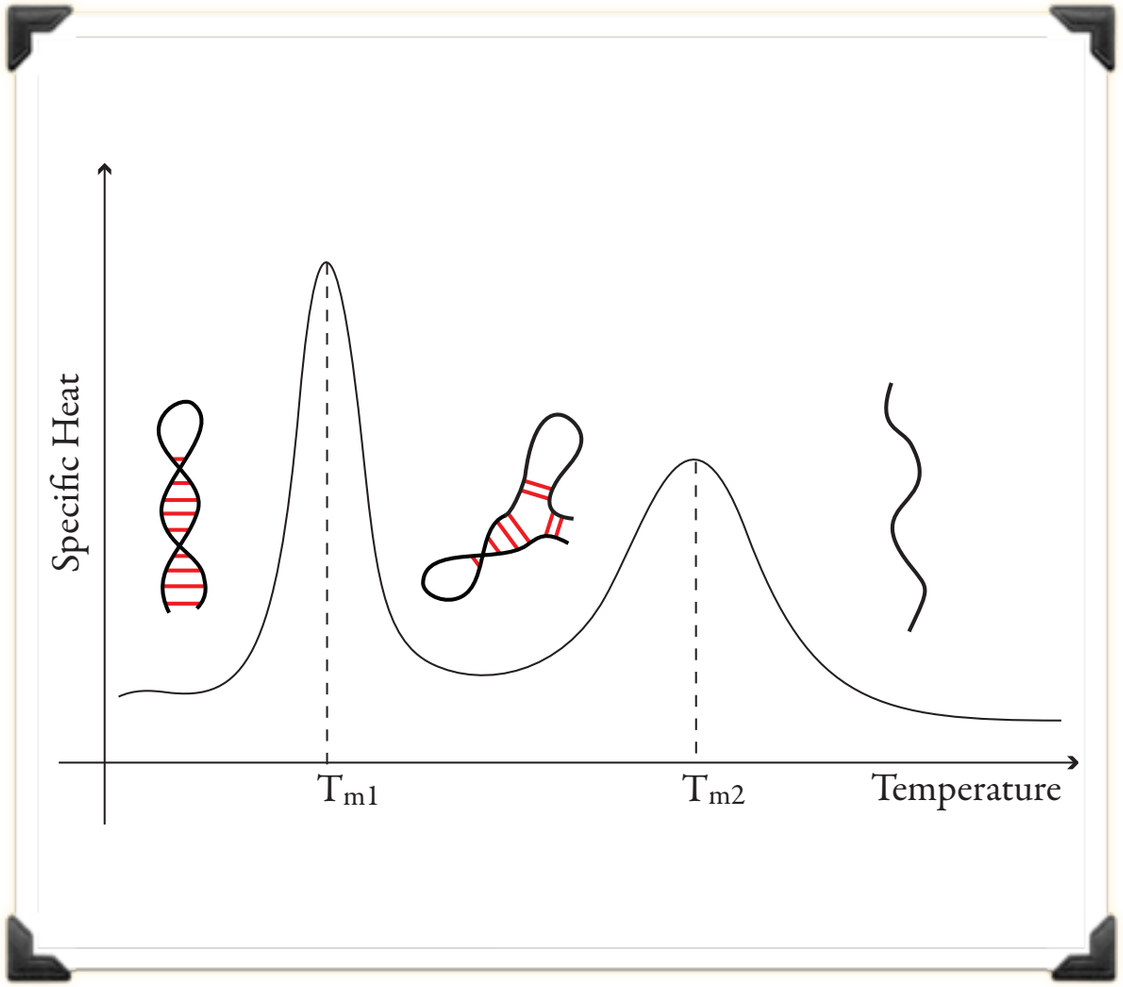


~150 different non-canonical pairs accounted for in the NDB

Ab initio physical modeling

- predict equilibrium structures,
- folding intermediates,
- thermodynamics,
- kinetic barriers,
- …

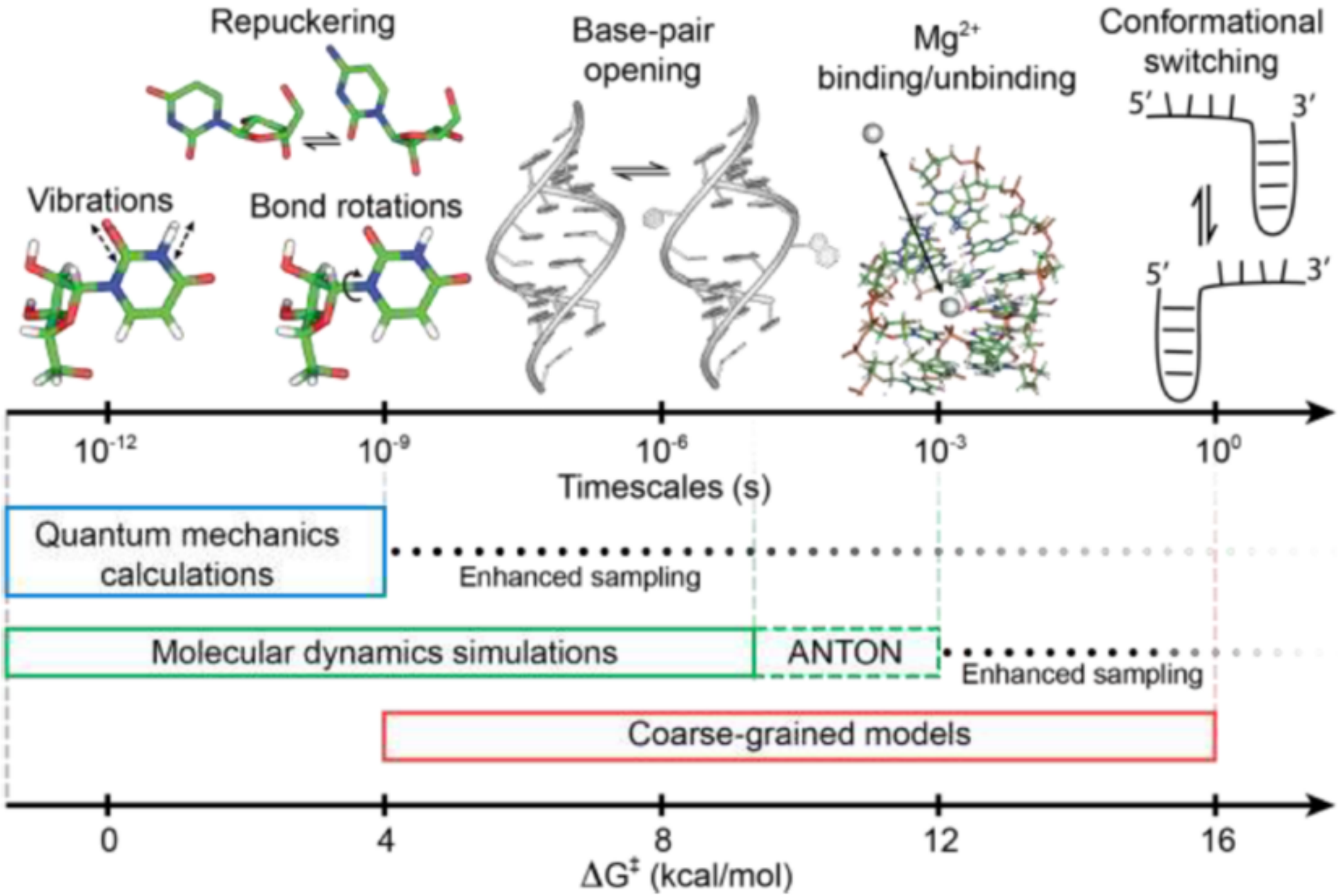




Behavior depends on environment

Temperature, ligands, ions, pH,…

RNA time scales and simulations



Prediction of structures, dynamical and thermodynamical behavior in 3D

Physical modeling methods : 3D, potential energy functions

Molecular dynamics

F=ma
energy model & force field

serial	temporal evolution dynamics	*limited sampling time*
replica	enhanced sampling thermodynamics (replicas in T)	*cpu*

Monte Carlo simulations

stochastic moves
energy model

simple	structure prediction	*limited sampling time*
basin-hopping	energy landscape kinetics	*cpu*

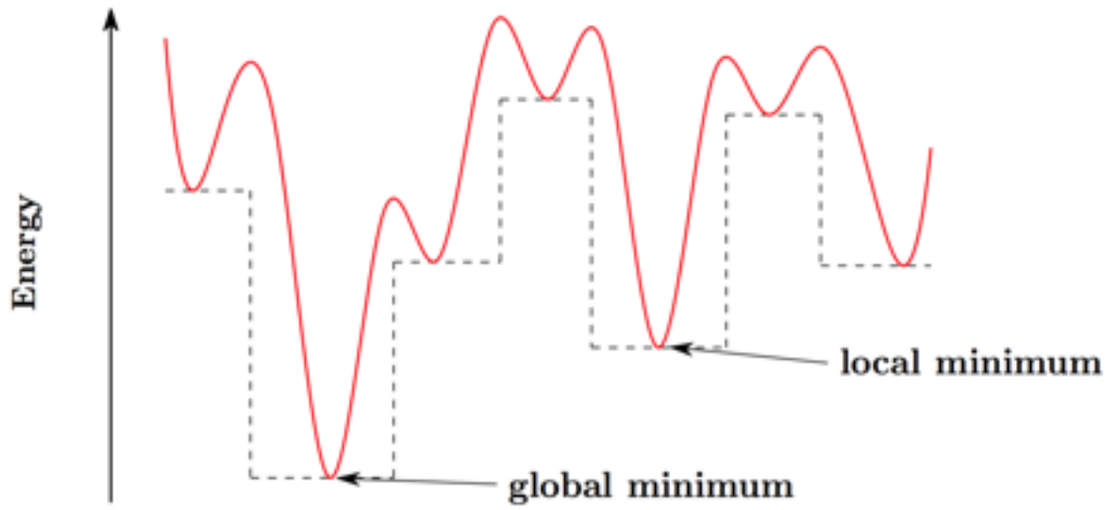
Normal mode analysis

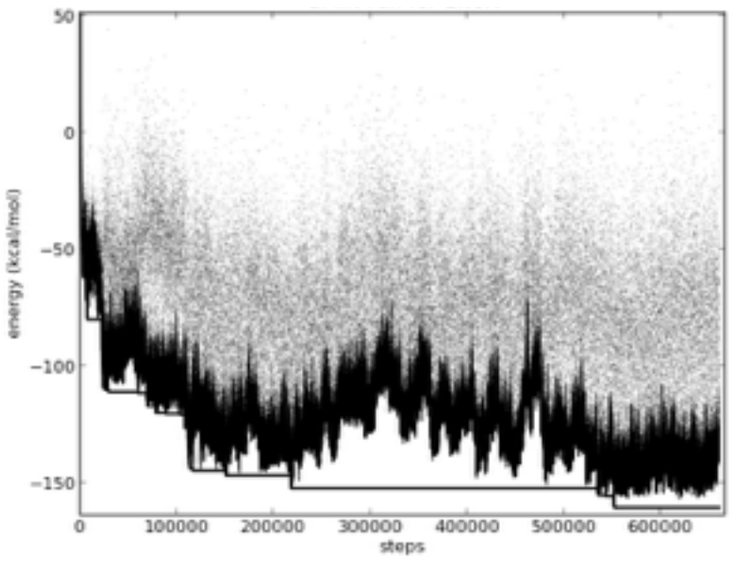
energy model
vibrational modes

cartesian coordinates	small structural modifications	*limited applicability*
internal coordinates	large structural modifications	*harder to implement*

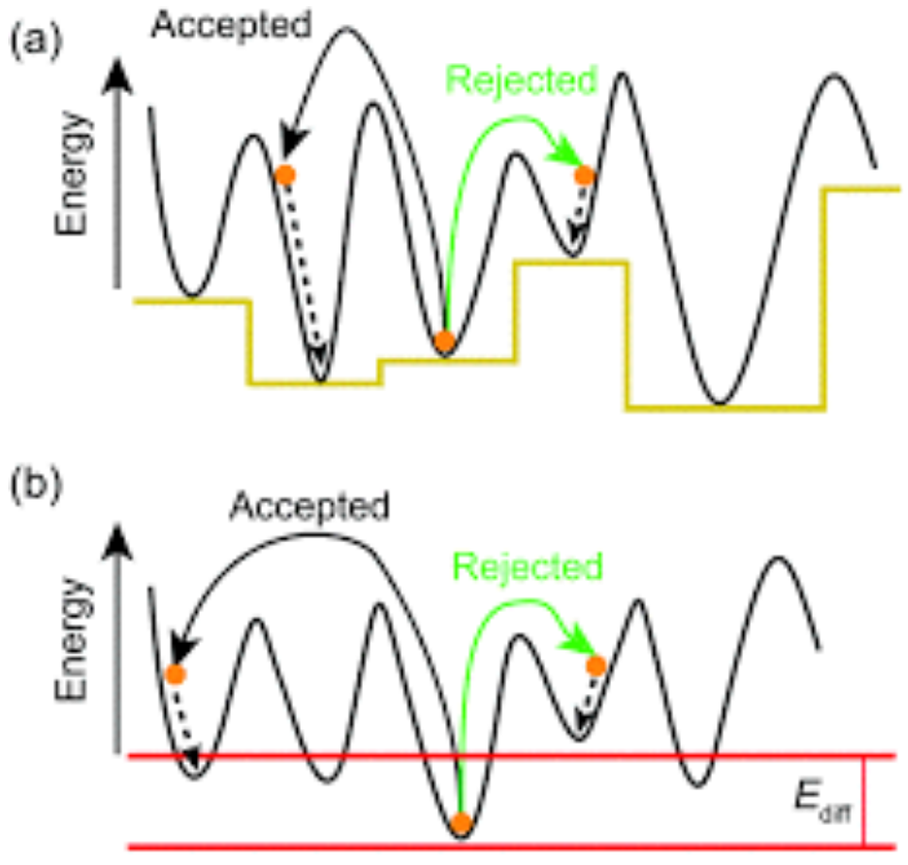
Basin-Hopping Monte Carlo

Monte Carlo
- create new configuration (no matter how),
- evaluate new energy,
- accept or refuse given a Metropolis condition





Basin-hopping MC



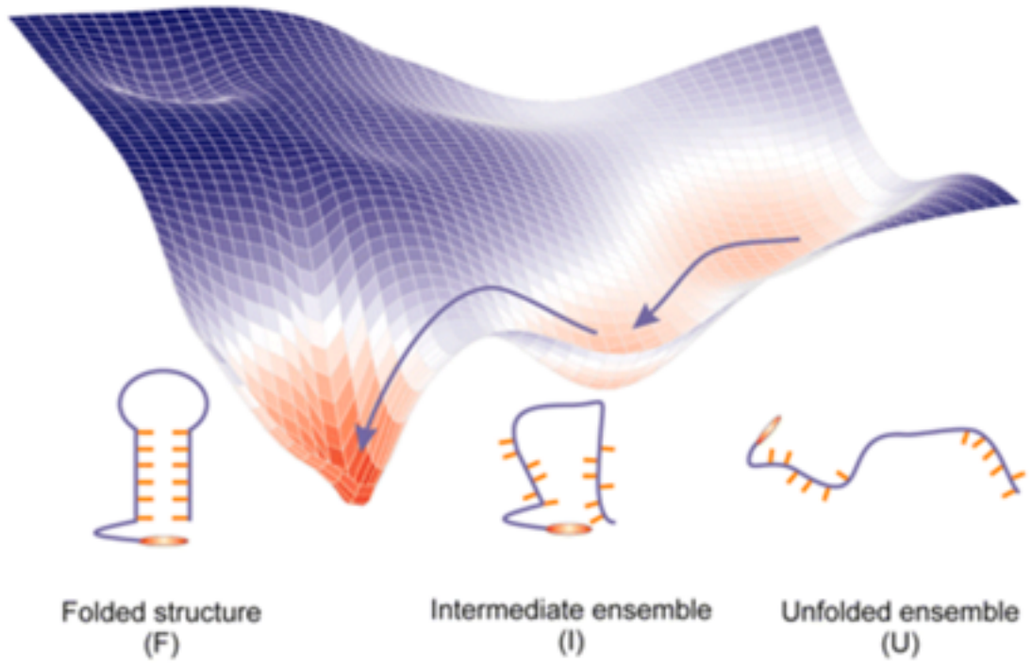
depth ⇢ stability

width ⇢ accessibility

Z. Li and H. A. Scheraga, J. Mol. Struct., 179, 333 (1988).

D.J. Wales and J.P.K. Doye, J. Phys. Chem. A, 101, 5111 (1997)

Discrete path sampling



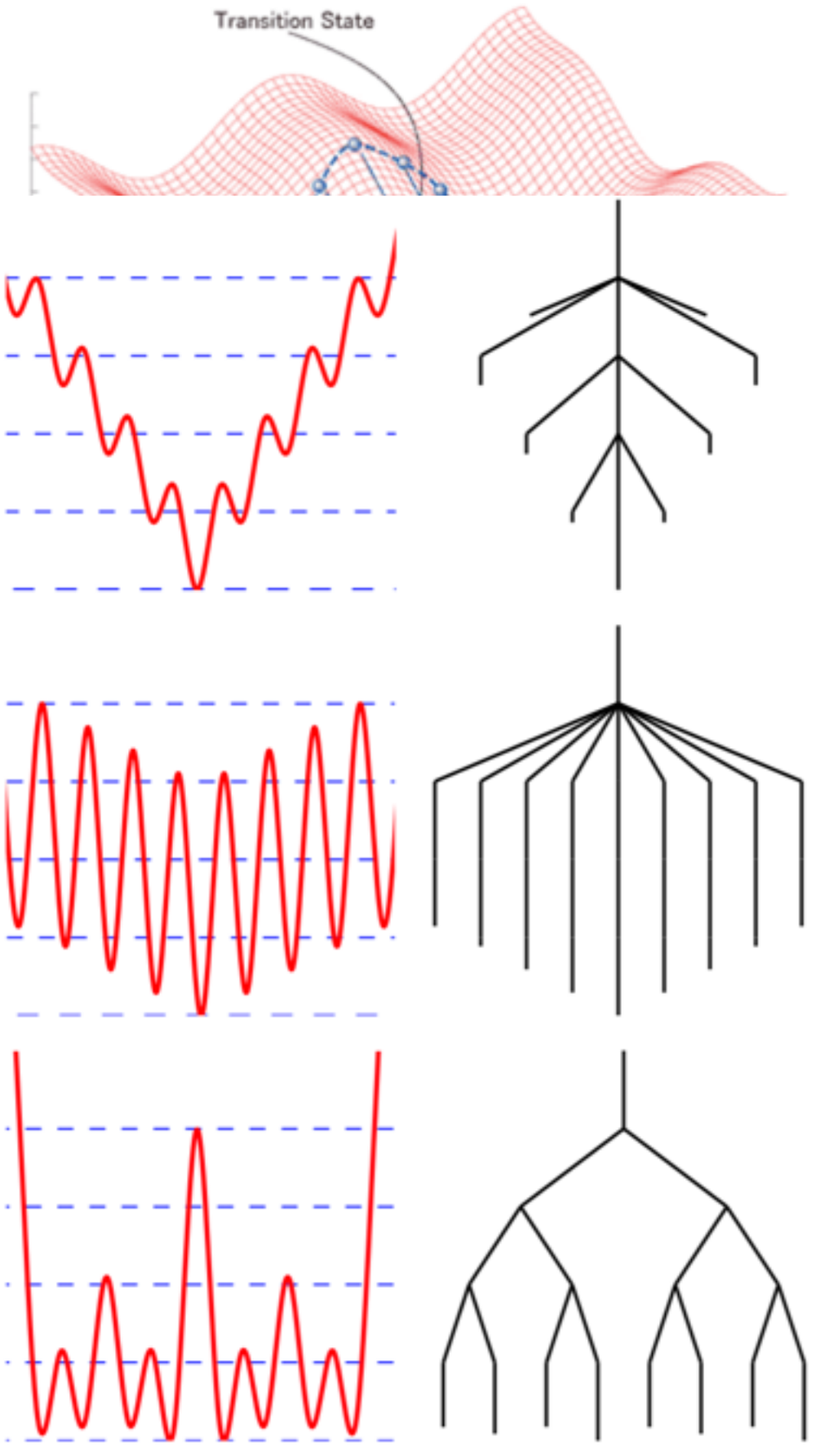
Ma, PNAS (2006)

1. Generate local minima.

2. Identify stationary points : minima and transition states.

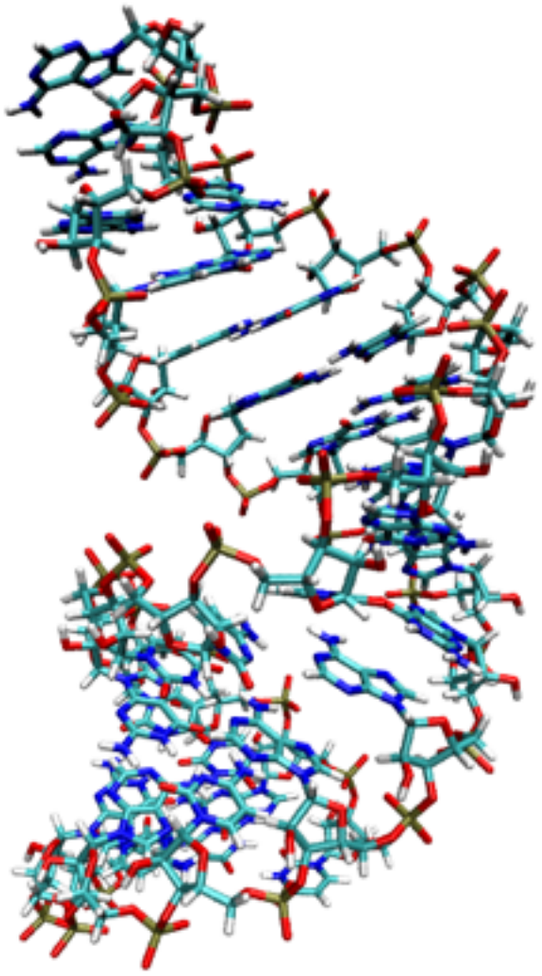
3. Determine lowest energy paths and connect all basins.

4. Build “disconnectivity graphs”: representation of the energy it takes to convert any two given structures.

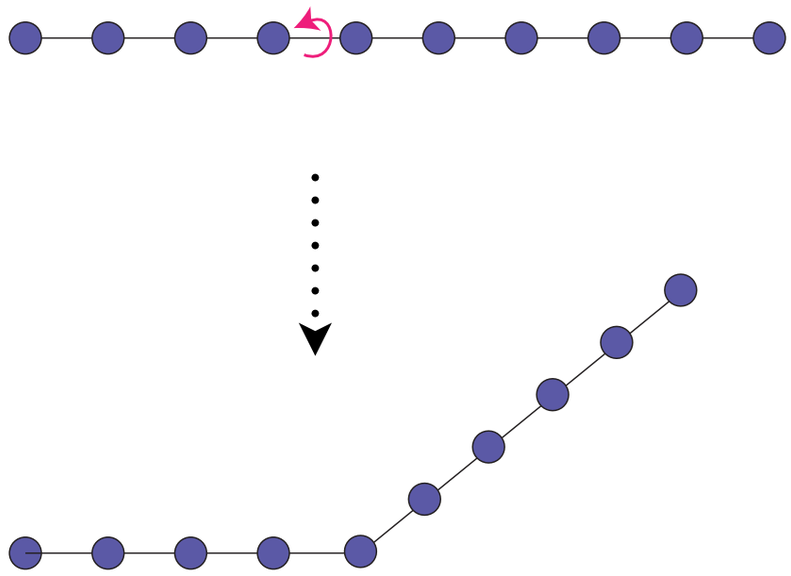


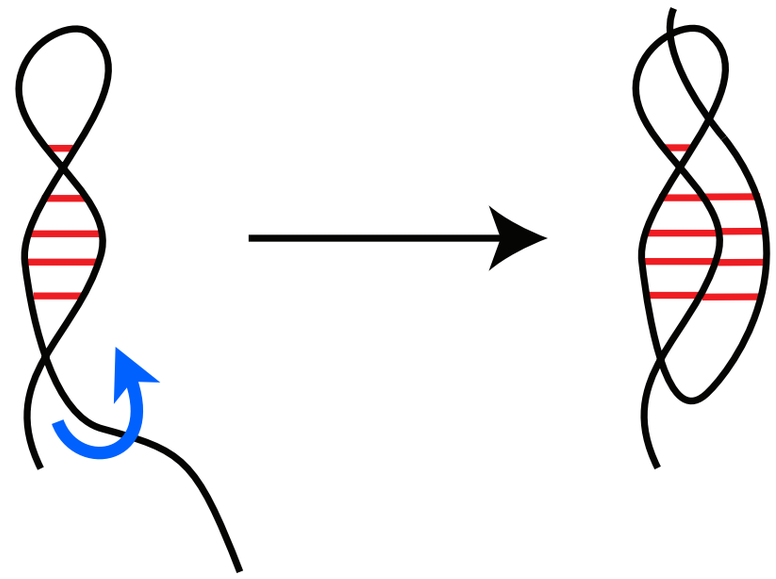
RNA Physical model

Atomistic representation, implicit solvent



local and global moves, torsions





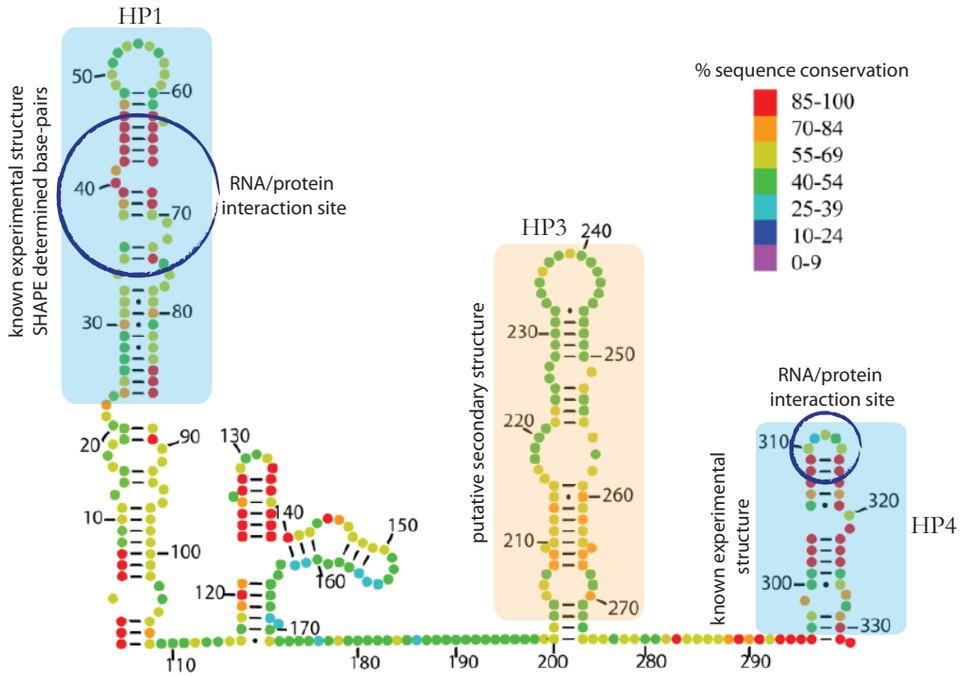
Amber Force Field

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

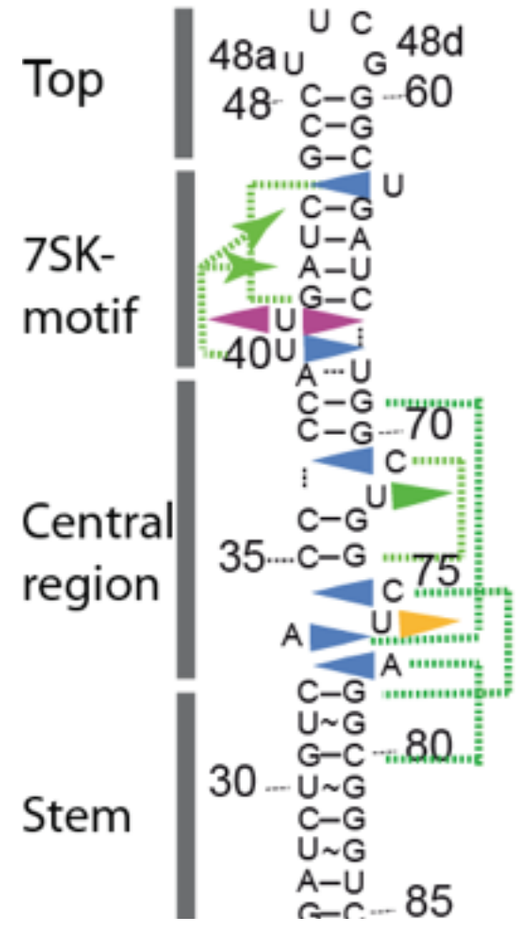
$$+ \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

7SK non-coding RNA (RNP)

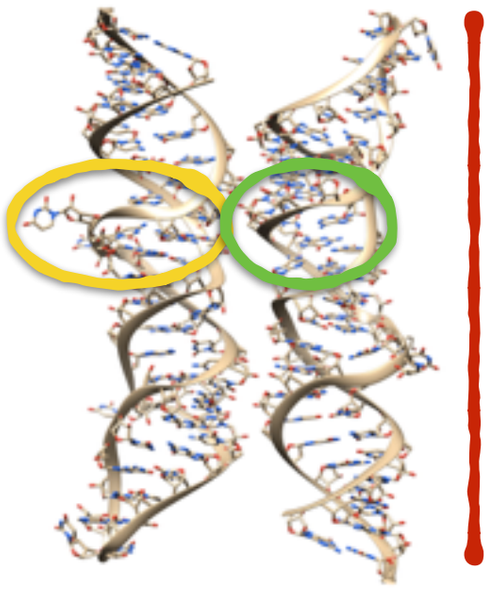
i.e. computational scientist's dreamland!



NMR
X-Ray
SHAPE
SAXS



X-ray structure is a dimer



Dock-Bregeon, Pasquali,

NMR
structure

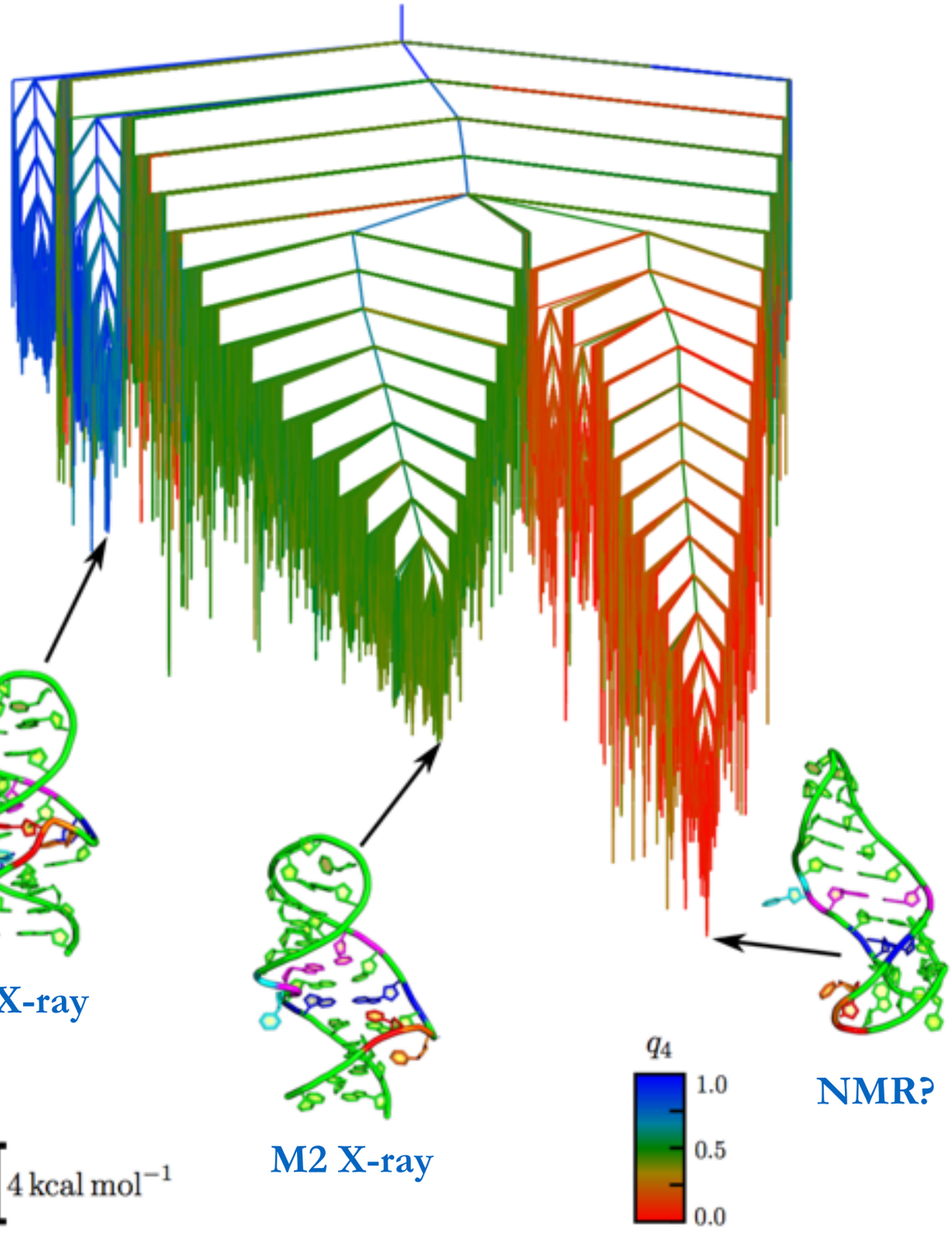
Dock-Bregeon, Lesbars,

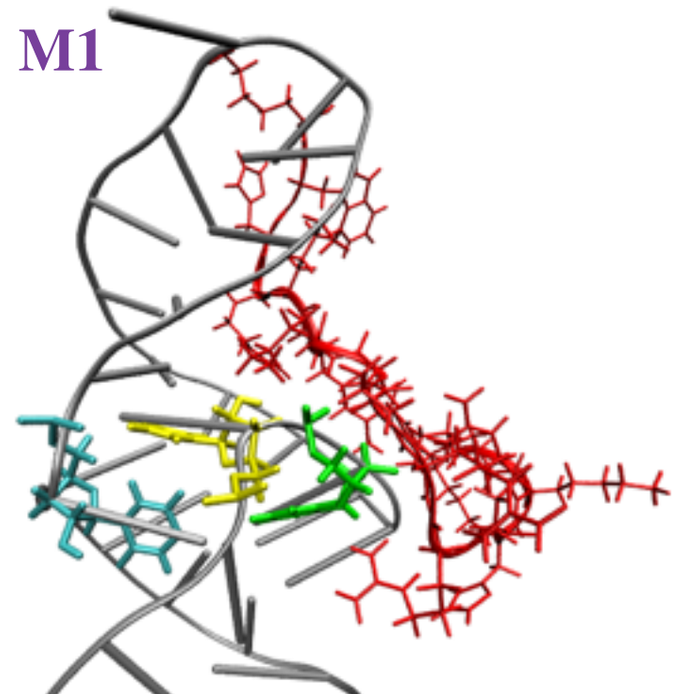
M1 2 triplets : **U40-C45-G64**, **U41-A43-U66**
M2 one triplet: **U40-A43-U66**, **U41 out**

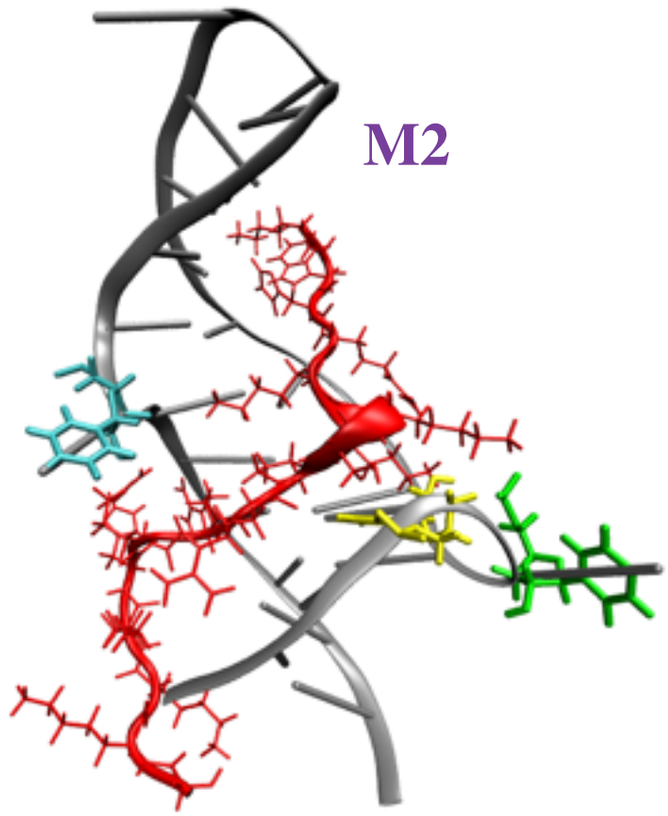
HP1 native disconnectivity graph

Generate a database of structures (~10^4-10^5 local minima)

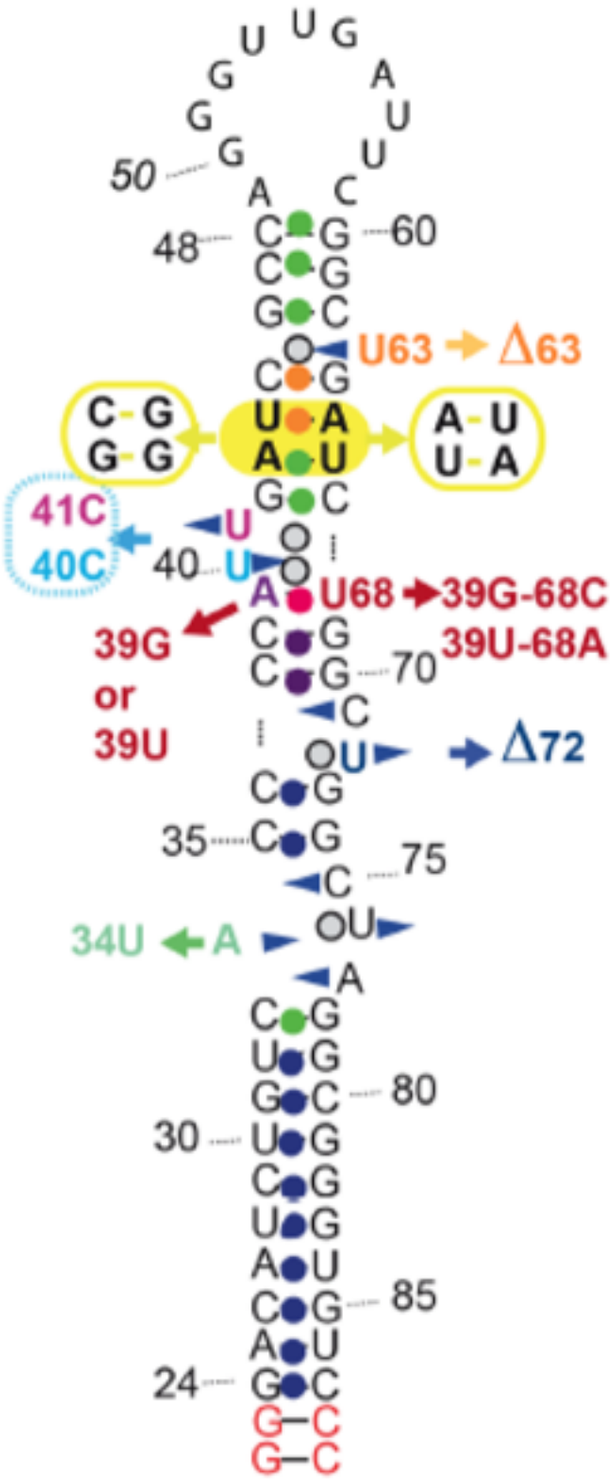
Connect them through path sampling

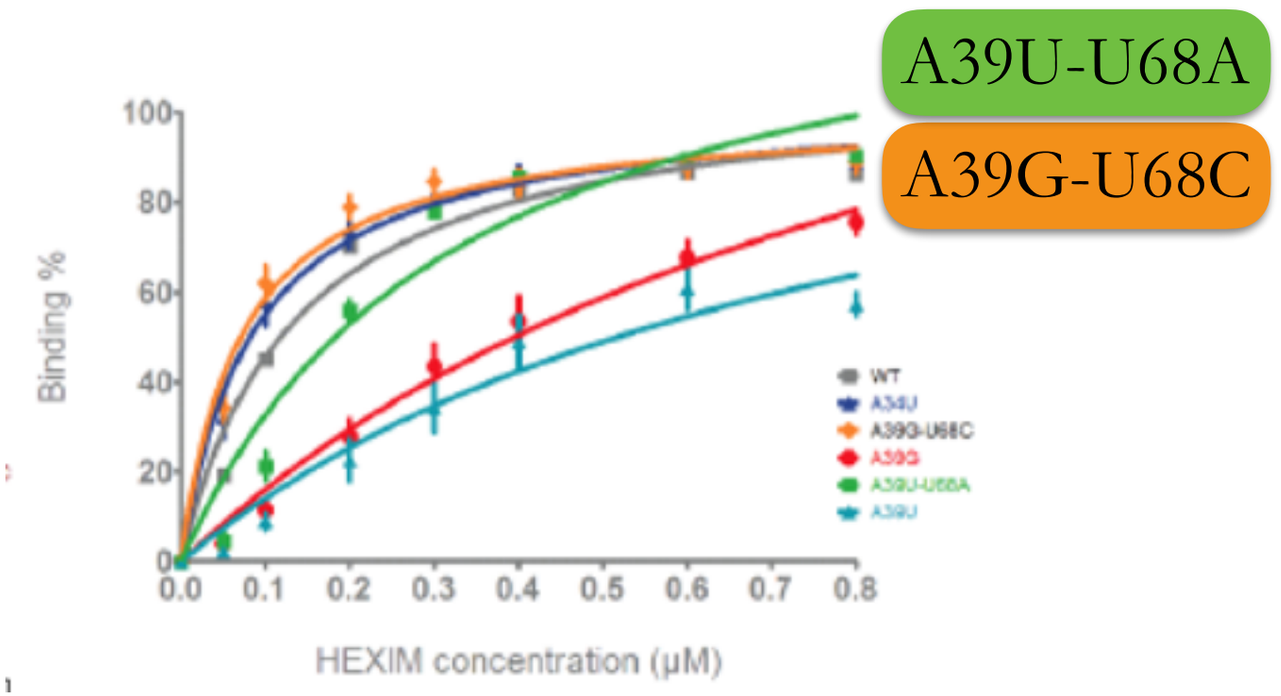


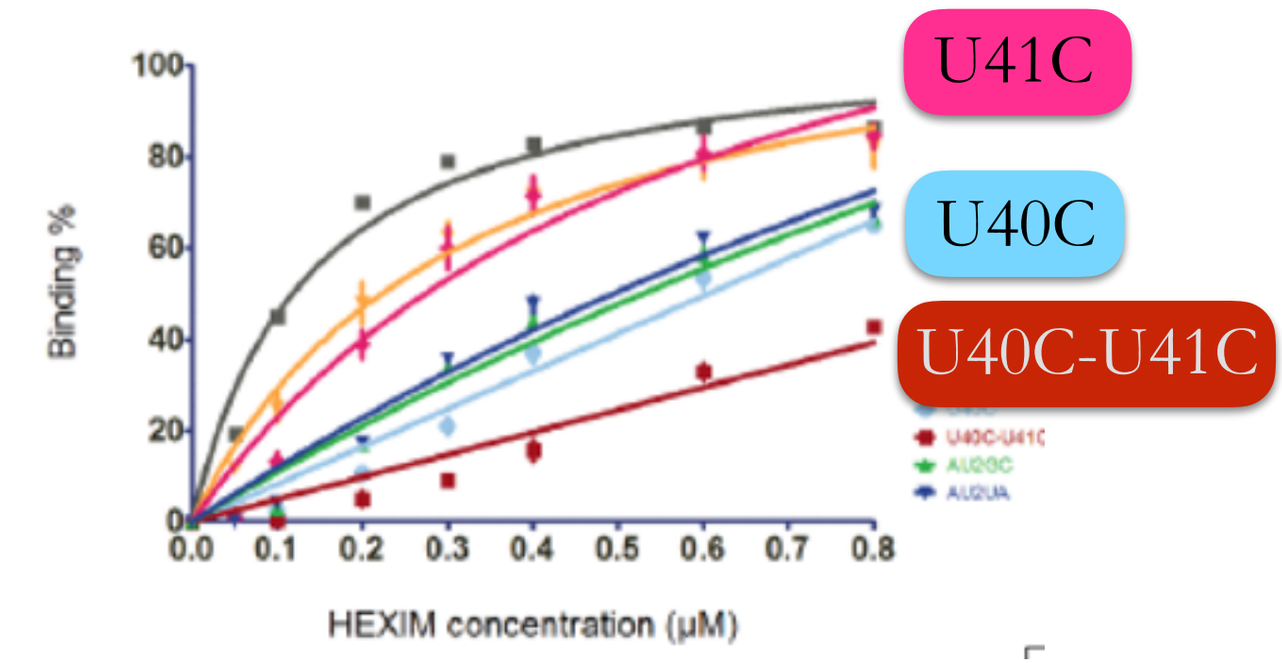




HP1 mutants

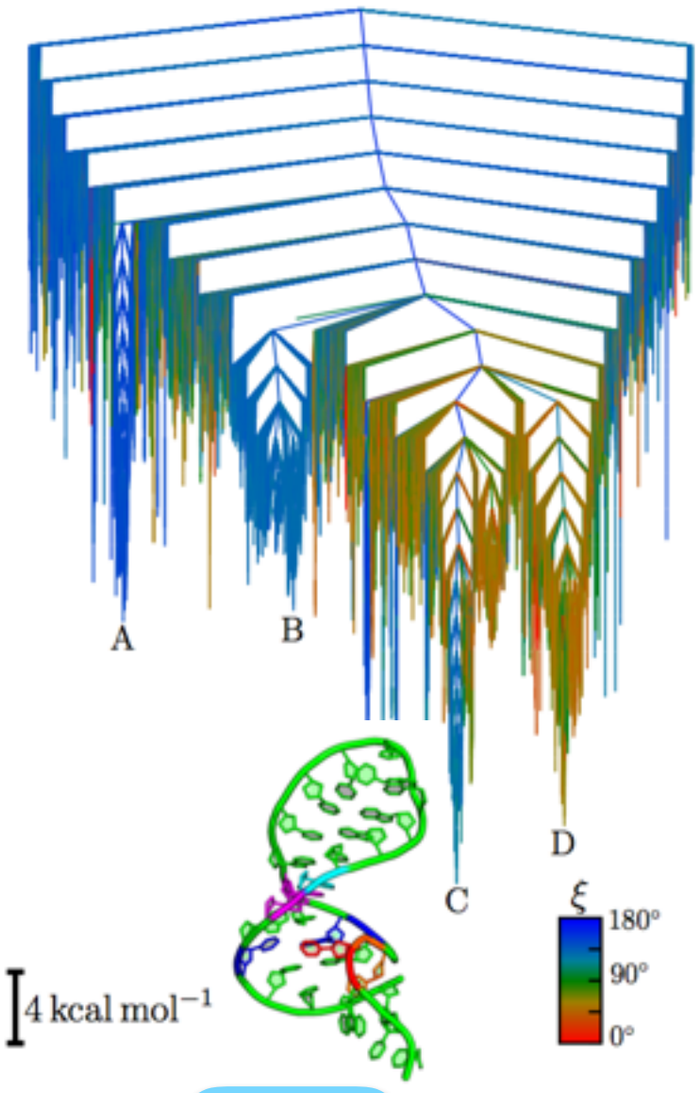

U63 → Δ63
41C
40C
U68 → 39G-68C
39G
or
39U
39U-68A
Δ72
34U
50
48
60
70
75
80
30
85
24
35


A39U-U68A
A39G-U68C
Binding %
HEXIM concentration (μM)
WT
A39U
A39G-U68C
A39G
A39U-U68A
A39U


U41C
U40C
U40C-U41C
Binding %
HEXIM concentration (μM)

HP1 mutants simulations

Poor binding



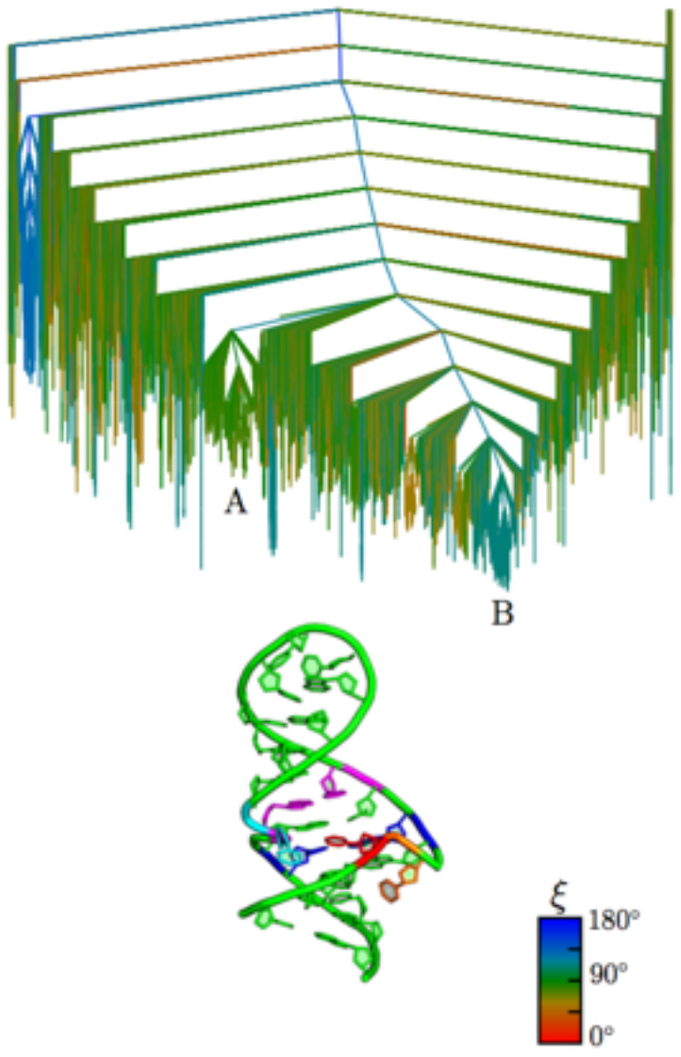
U40C

Very few M2 structures

C40 inside, no T2

U63 blocks the groove

very poor binding



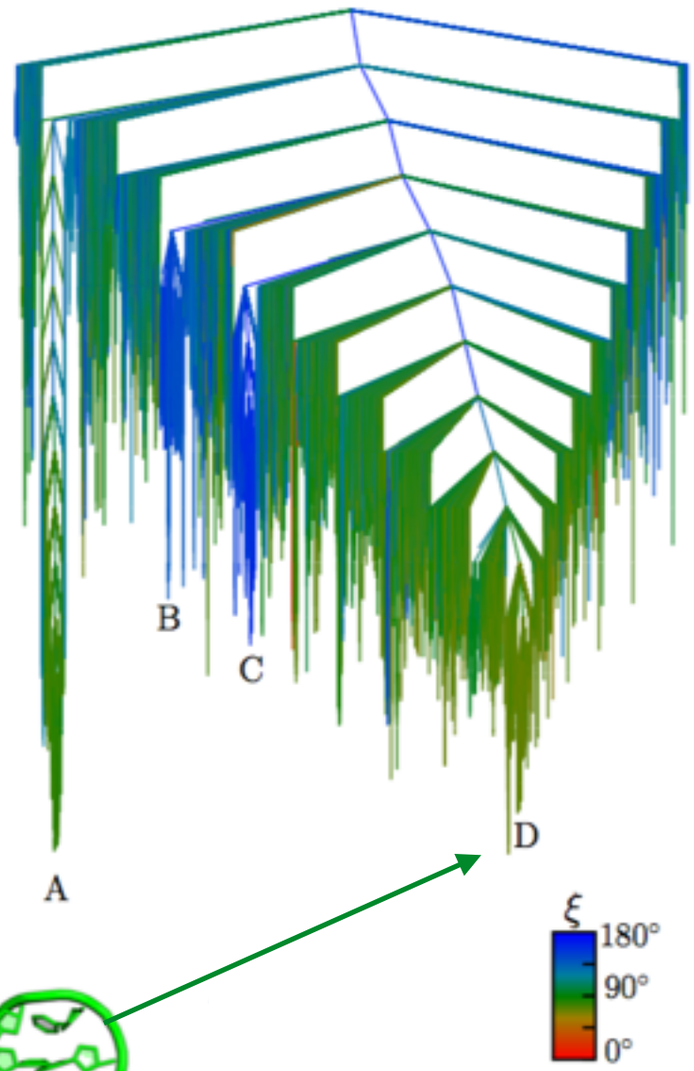
U40C-U41C

Mainly M2 structures

Both C40 and C41 inside, no T2

U63 blocks the groove

high binding



A39U-U68A

Low energy M2 structures

U40 kept in place by T2

U63 on the outside

Take home message 1

The good

- Basin-hopping samples widely the conformational space and allows to explore the energy landscape. Much more effective than other usual 3D methods.
- Path sampling allows to define “families” of conformations that can be linked to experimental data.
- Kinetic and thermodynamical properties can be extracted (theoretical framework and algorithms)

The bad

- Only as good as the force-fields
- Expensive computations even in implicit solvent (but still less expensive than more common methods).
- Need to have insights to define good order parameters for interpreting experimental data (but not for generating the EL)

If you want to use expensive 3D physical simulations to make sense of structural experimental data path sampling is the way to go

Physical modeling methods : 3D, potential energy functions

Molecular dynamics

energy model & force field

serial	temporal evolution dynamics	*limited sampling time*
replica	enhanced sampling thermodynamics (replicas in T)	*cpu*

Monte Carlo simulations

energy model

simple	structure prediction	*limited sampling time*
basin-hopping	energy landscape kinetics	*cpu*

Normal mode analysis

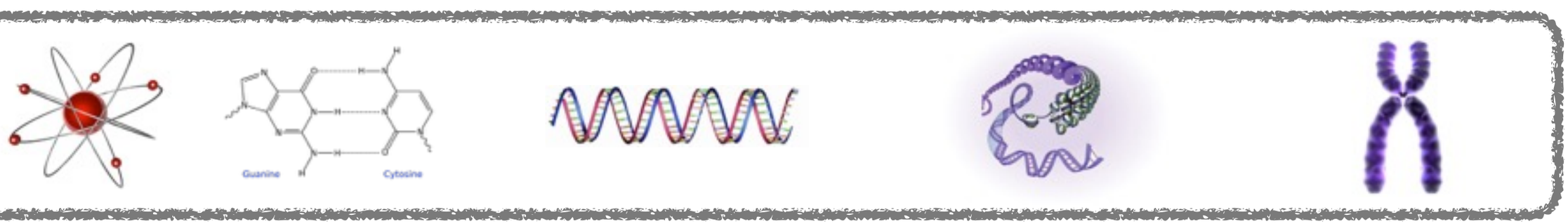
energy model

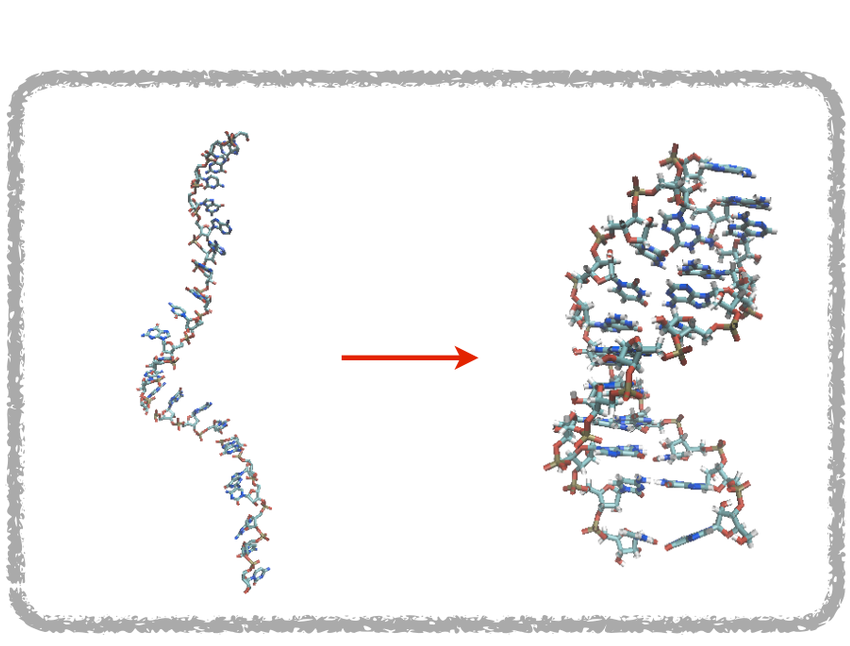
cartesian coordinates	small structural modifications	*limited applicability*
internal coordinates	large structural modifications	*harder to implement*

The scales problem

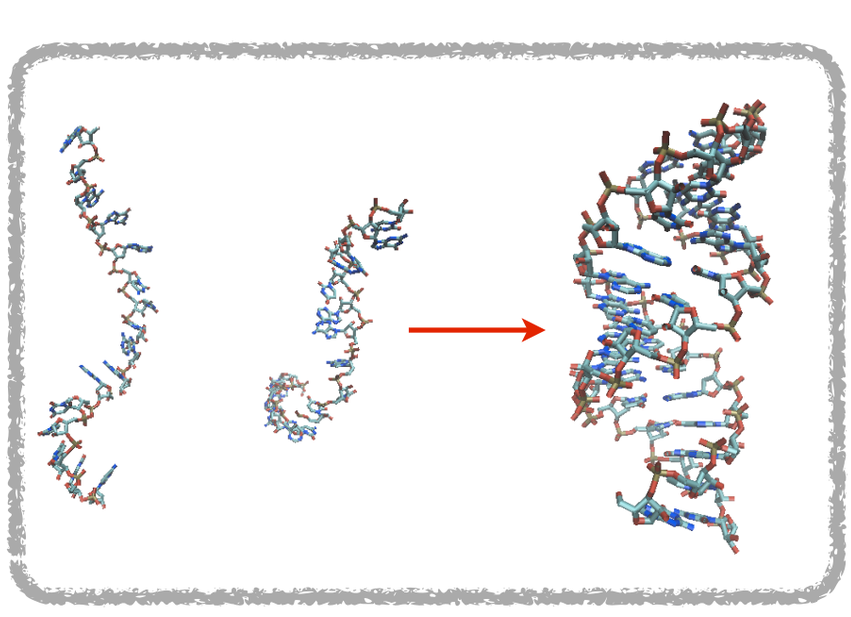
(DNA spans 10^9 orders of magnitude in length!)

Å nm **10 nm** 100 nm μm





Folding



Assembly

Too large for an atomistic description
Too small for a mesoscoptic description

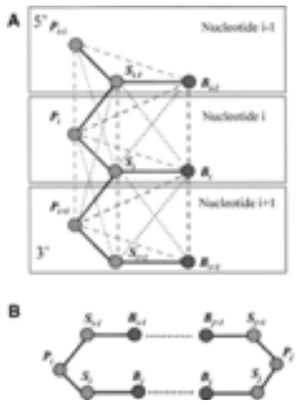
Coarse-grained RNA modeling

simplified models to represent the meaningful degrees of freedom of the system and the process of interest

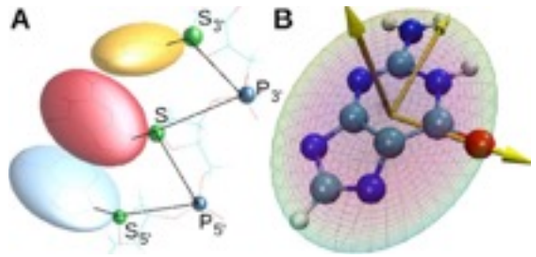
Coarse-graining, implicit solvent, implicit ions*

Jonikas, 2009

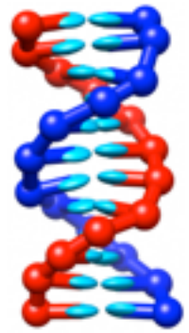
Dokholyan, 2008

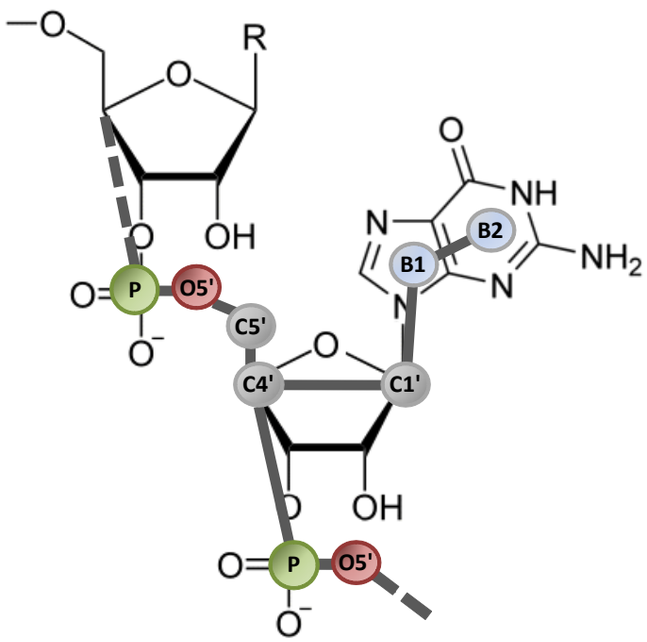


Plotkin, 2008



Ouldridge, 2011





HiRE-RNA

High Resolution Energy Model for RNA and DNA

Flexible, unconstrained

MD, REMD, ST, GMIN

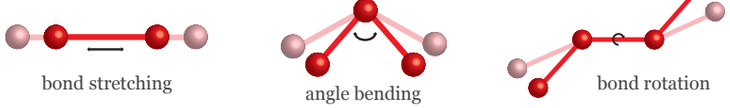
S. Pasquali, P. Derreumaux, JPCB (2010)

HiRE-RNA, version 3

Revisiting "traditional" atomic-like potentials

(non-isotropic interactions)

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



harmonic
statistical parameters

Planarity
Non-canonical pairs

Base orientation

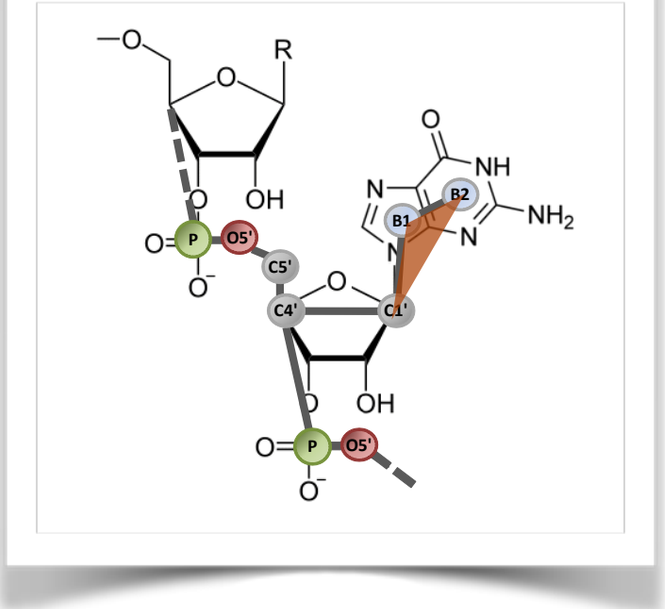
$$E_{ex\ vol} = \varepsilon_{ex}\ e^{-\kappa(r-r_v)}$$

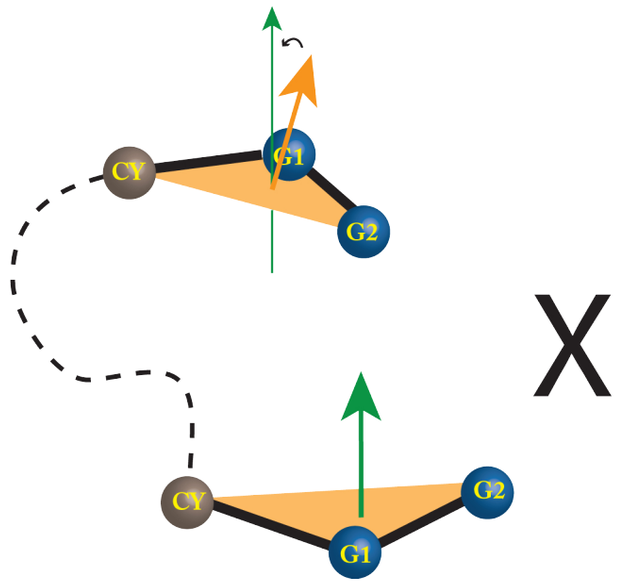
$$E_{el} = \varepsilon_{el}\ \frac{q^2}{4\pi\epsilon_0\epsilon_r r} e^{-r/\ell}$$

T. Cragnolini, Y. Laurin, P. Derreumaux, S. Pasquali, JCTC (2015)

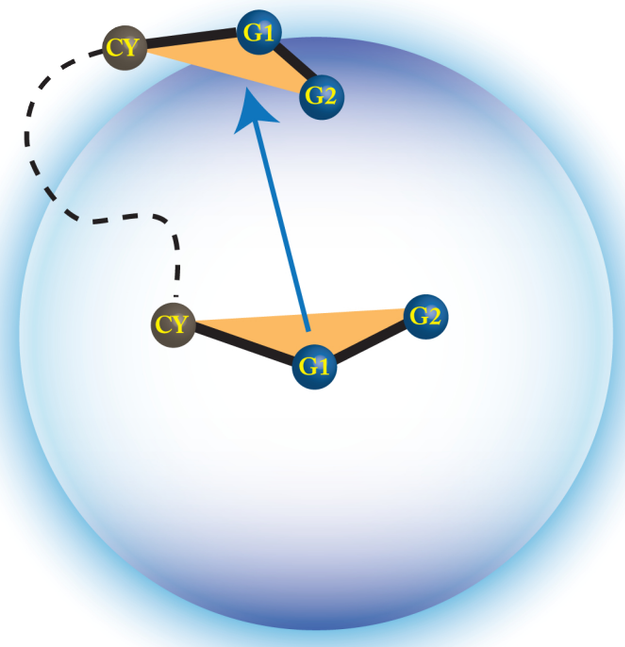
T. Cragnolini, P. Derreumaux, S. Pasquali, J. Physics: Condensed Matter (2015)

Stacking





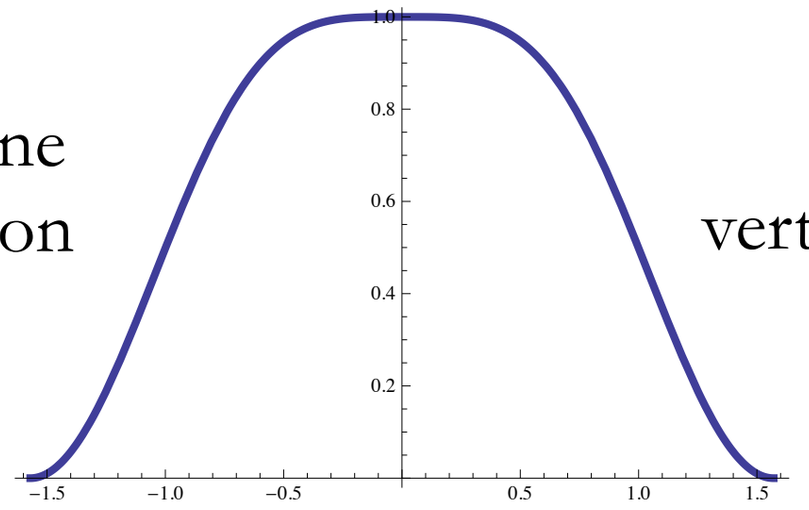
X



$$E_{st} = \varepsilon_{st}\, e^{-\frac{(r-r_{st})^2}{\sigma}} \; (\vec{n_i}\cdot\vec{n_j})^2 \left(1-|\vec{n_i}\times\vec{r}|^4\right)\left(1-|\vec{n_j}\times\vec{r}|^4\right)$$

plane distance

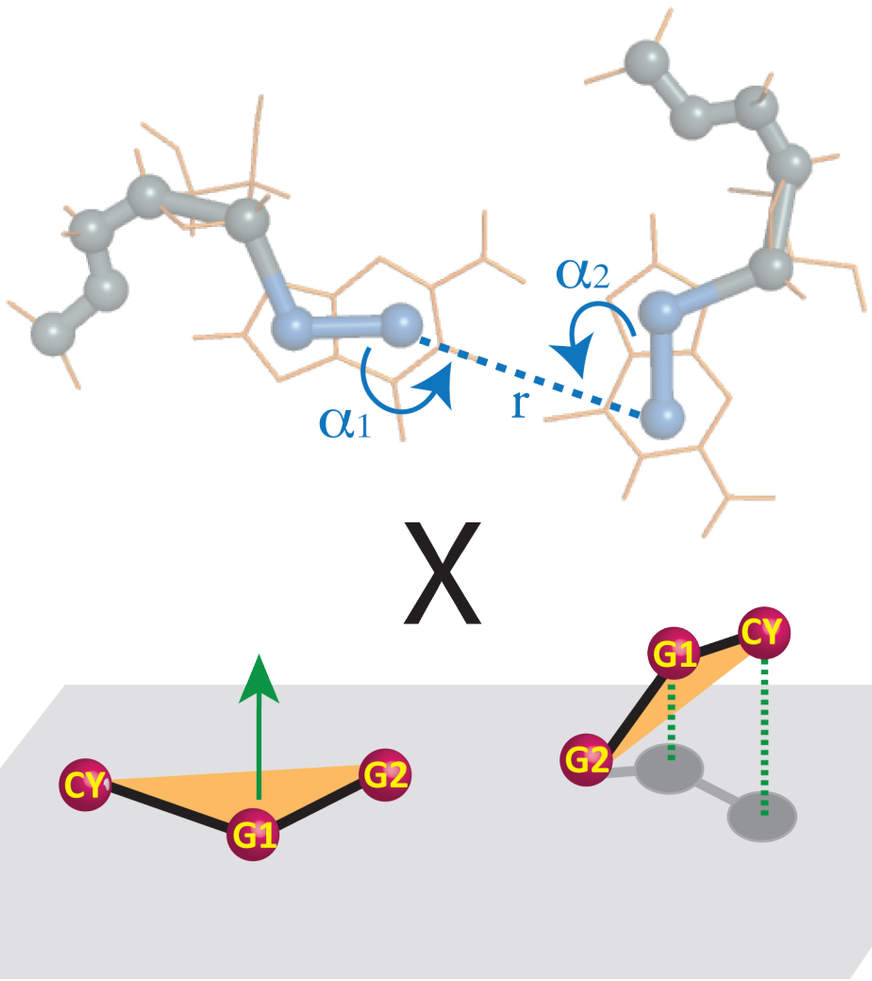
same plane orientation



vertical position

Base Pairing

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

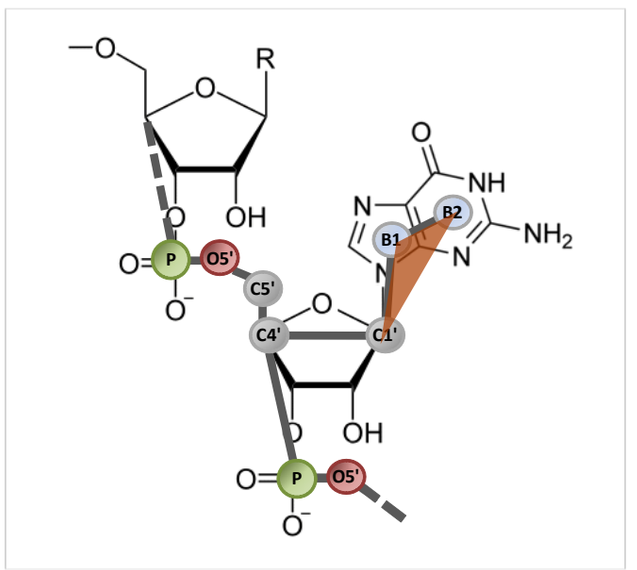


$$E_{hb} = \varepsilon_{hb} e^{-(r-\rho)^2/\xi}\, \nu(\alpha_1)\nu(\alpha_2)$$

$$\nu(\alpha) = \begin{cases} \cos^6(\alpha - \Omega), & \text{for } -90^\circ \le (\alpha - \Omega) \le 90^\circ; \\ 0, & \text{otherwise} \end{cases}$$

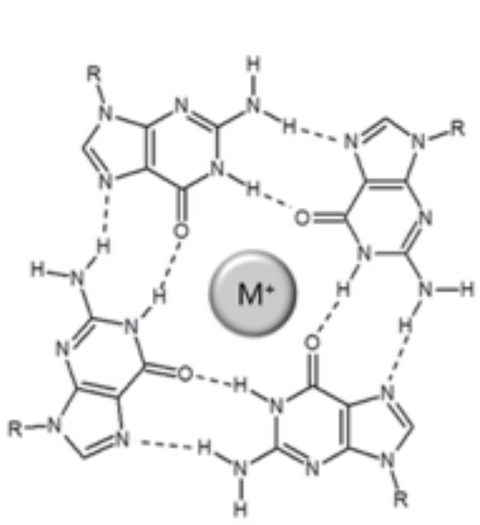
$$\alpha = \begin{cases} +\alpha, & \text{if } \cos(\tau) \ge 0; \\ -\alpha, & \text{otherwise} \end{cases}$$

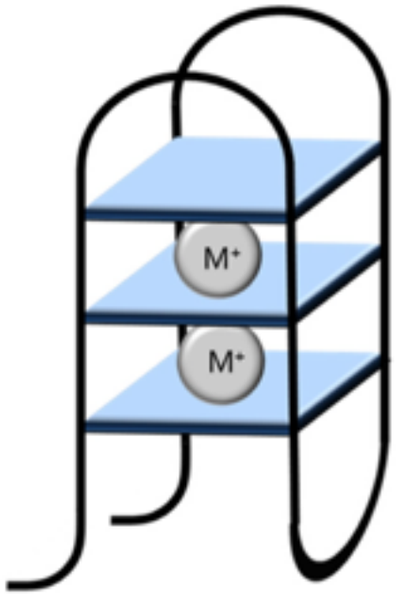
$$E_{plane} = \varepsilon_{pl} \sum_{k_j=1}^{3} \left(e^{-(d_{B_i}^{k_j}/\delta)^2} + e^{-(d_{B_j}^{k_j}/\delta)^2} \right)$$

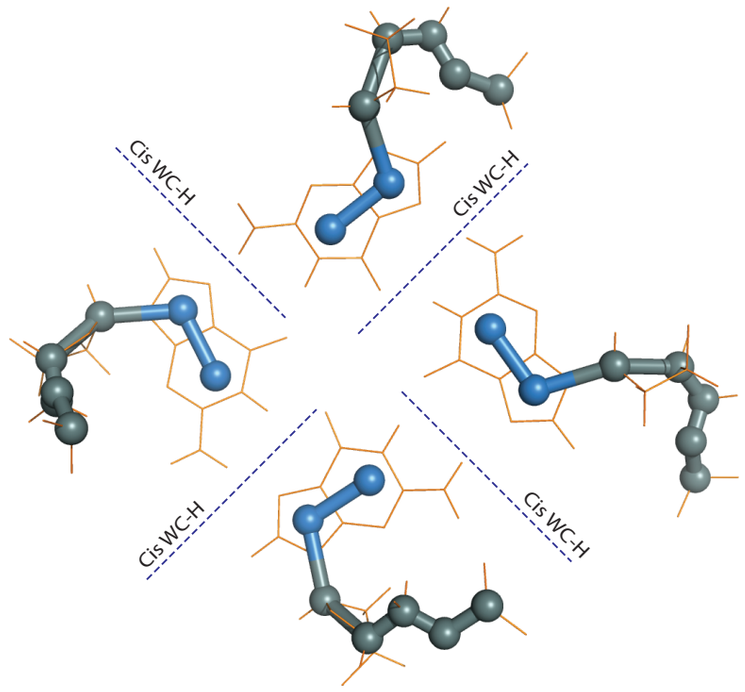


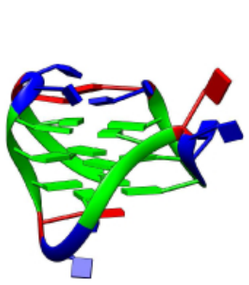
Toward multi-scale modeling: G-quadruplexes

Quadruplets

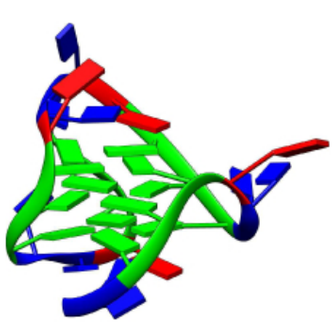




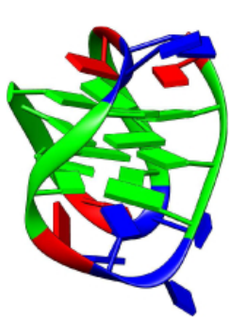




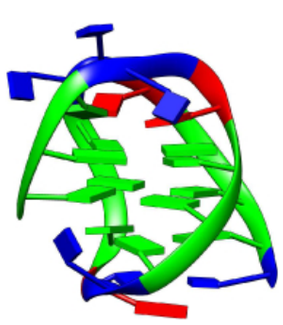
(1)$3+1_1$
2JSM



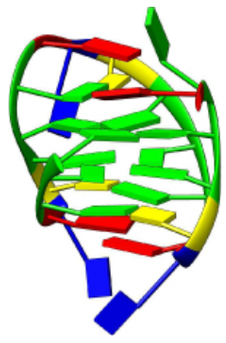
(2)$3+1_2$
2JSL



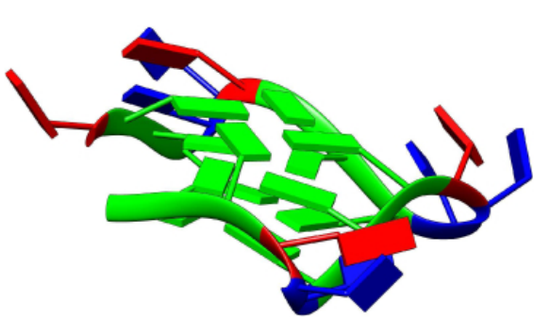
(3)basket_1
143D



(4)basket_2
2KF8



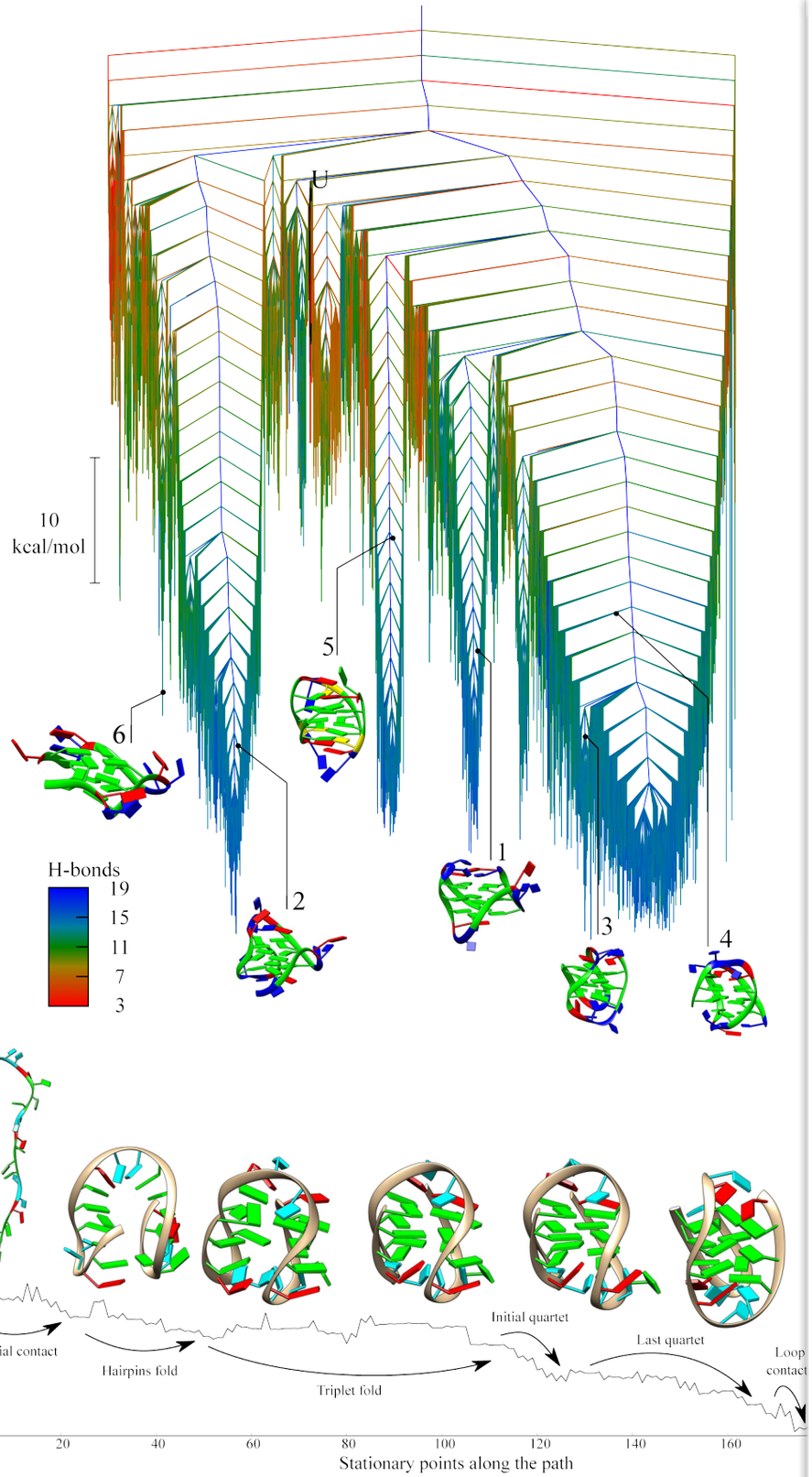
(5)chair
2KM3

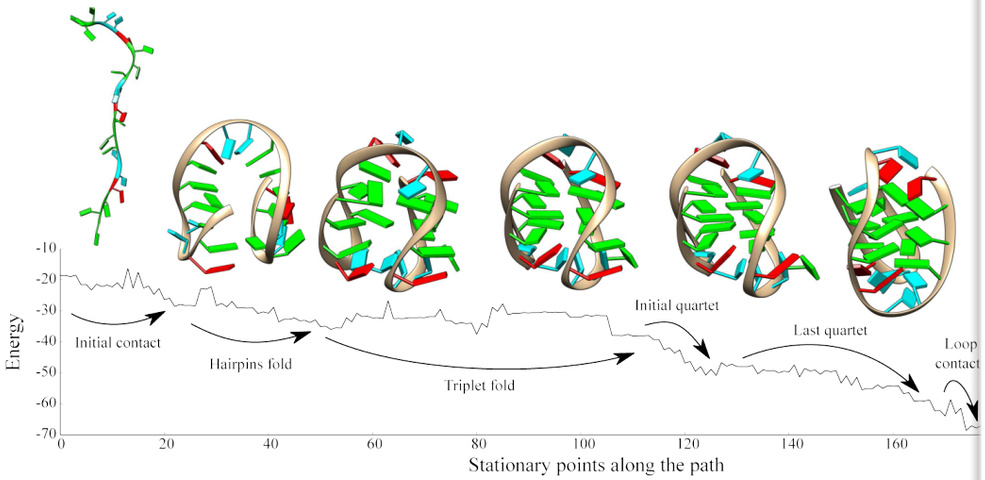


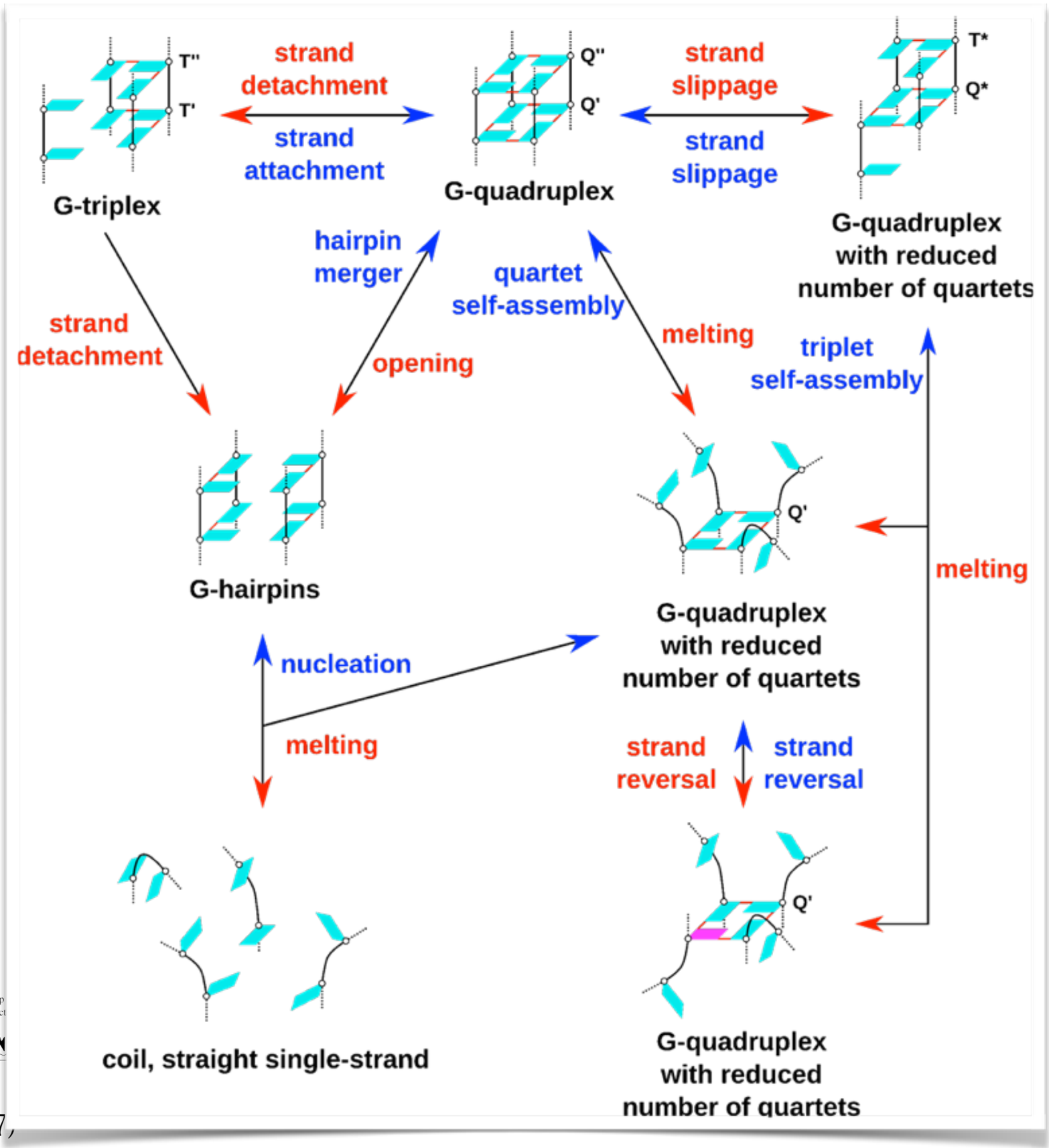
(6)propeller
1KF1

Why multiple folds ? why slow folding?

G-quadruplexes landscape

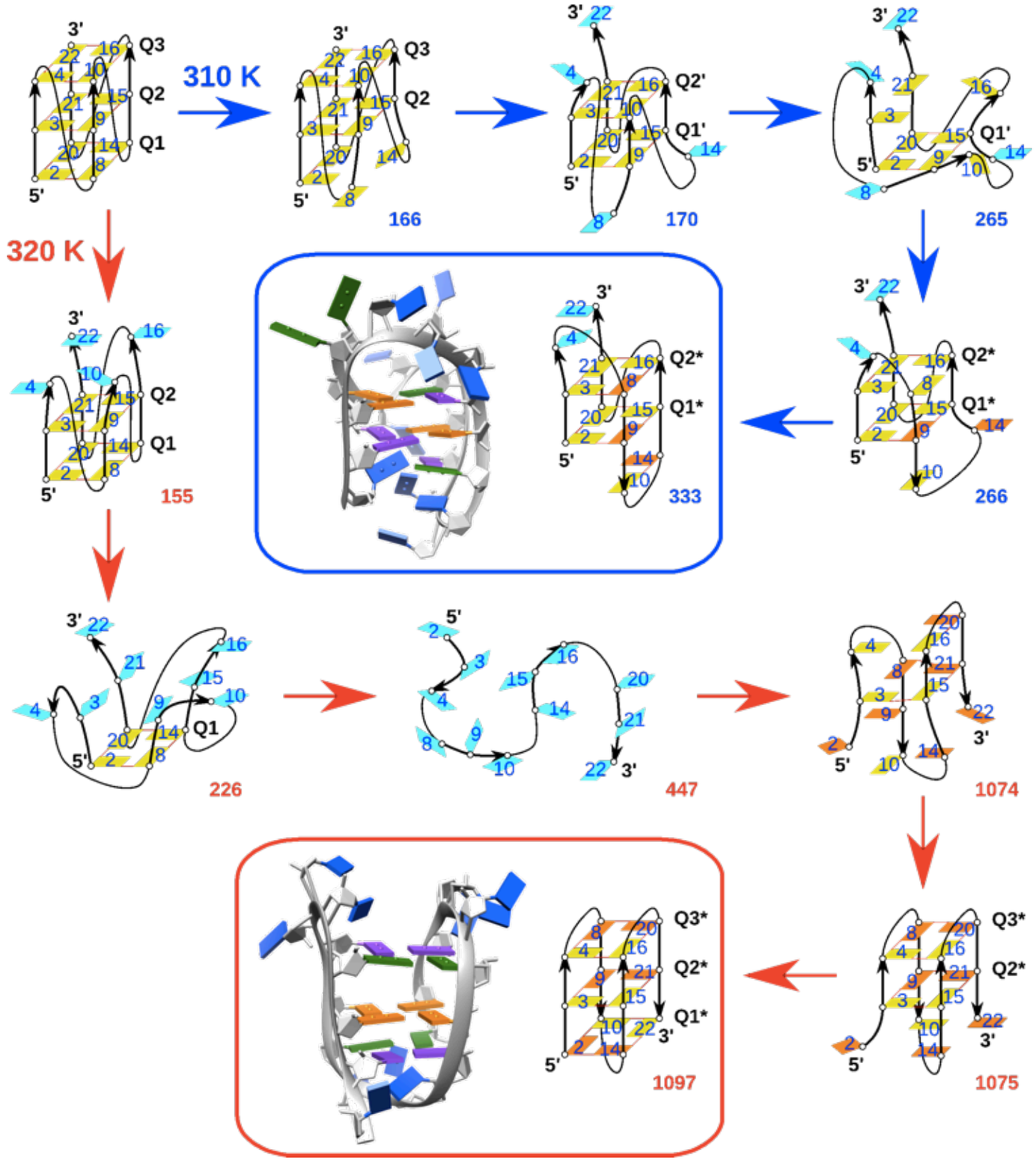


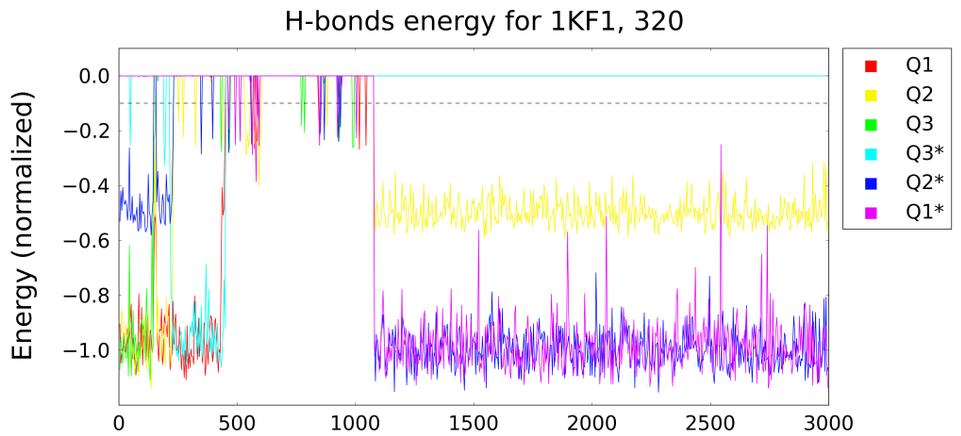


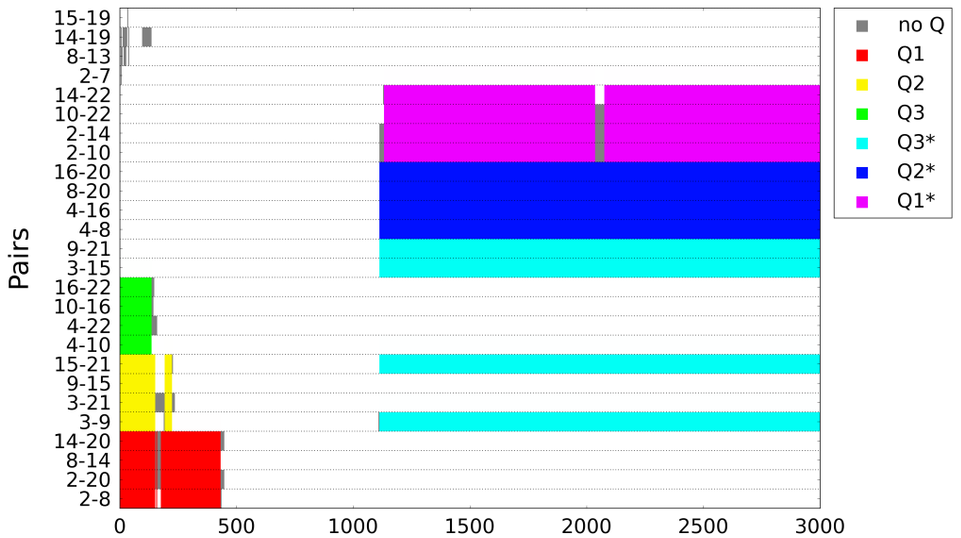


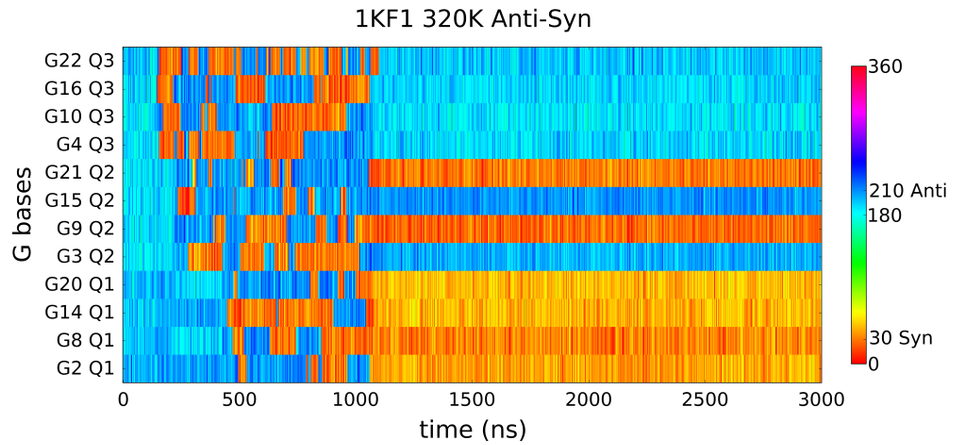
Cagnolini, xxx, Pasquali, Wales, J Chem Phys (2017)

Fixed temperature CG simulations









Take home message 2

The good

- Coarse-graining allows to access a wide conformational space that can not be explored with atomistic simulations.
- Kinetic and thermodynamical properties can be extracted.
- High-resolution allows to account for details of biological relevance and to go back and forth between coarse-grained and atomistic resolutions (multi-scale modeling).
- Experimental data (or data from other prediction tools such as 2D) can be incorporated in the simulations and the influence of environmental conditions can be tested.

The bad

- Only as good as the force-fields.
  Base-pairing parametrization main challenge when including non-canonical BP.
  Try to predict specific non-canonical motifs to get reasonable parameters?
- Expensive computations.
  Need for a big cluster and weeks of cpu time.
- Proper thermodynamic parametrization missing.
  Learn from OxRNA.
- Messy Code.
  But a LAMMPS version will soon be available.

RNA Energy Landscapes

unique secondary structure
one well defined minimum

small, regular structures
hairpins, duplexes

↓

Fast folding

base-pairing frustration
many alternative secondary structures
equivalent in energy

large, complex structures

↓

Extremely long lived metastable states

↓

Folding dynamics becomes relevant
to determine the attained structure

Acknowledgments

Laboratoire CiTCoM
Université de Paris

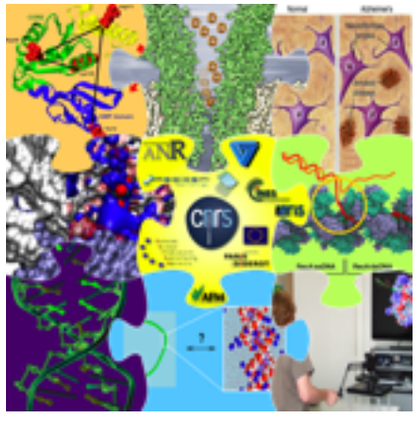
Elisa Frezza
Bruno Sargueil

Chemistry Department
University of Cambridge

David Wales
Tristan Cragnolini
Konstantin Roeder

Pharmacy Department
Universidade de São Paulo
RIBEIRÃO PRETO
Fernando Barroso da Silva

Laboratoire de Biochimie Théorique
Université de Paris



Philippe Derreumaux
Marc Baaden
Liuba Mazzanti

Institute of Biophysics
Czech Republic's Academy of Sciences
IBP

Jiri Sponer
Petr Stadlbauer