3D RNA simulations

Samuela Pasquali

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

samuela.pasquali@parisdescartes.fr

UNIVERSITÉ PARIS DESCARTES

Intricate structures

20-25 ~30 40 60 80 100 300

microRNA
siRNA

piRNA

tRNA
riboswitches

ribozymes

snoRNA

mRNA

tRNA

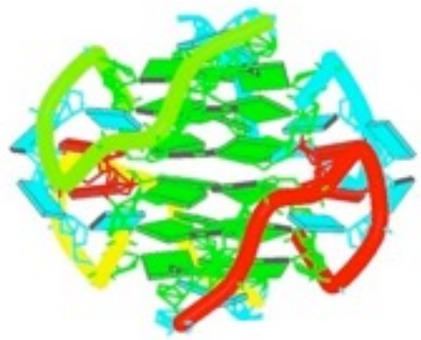
telomerase

viral fragment

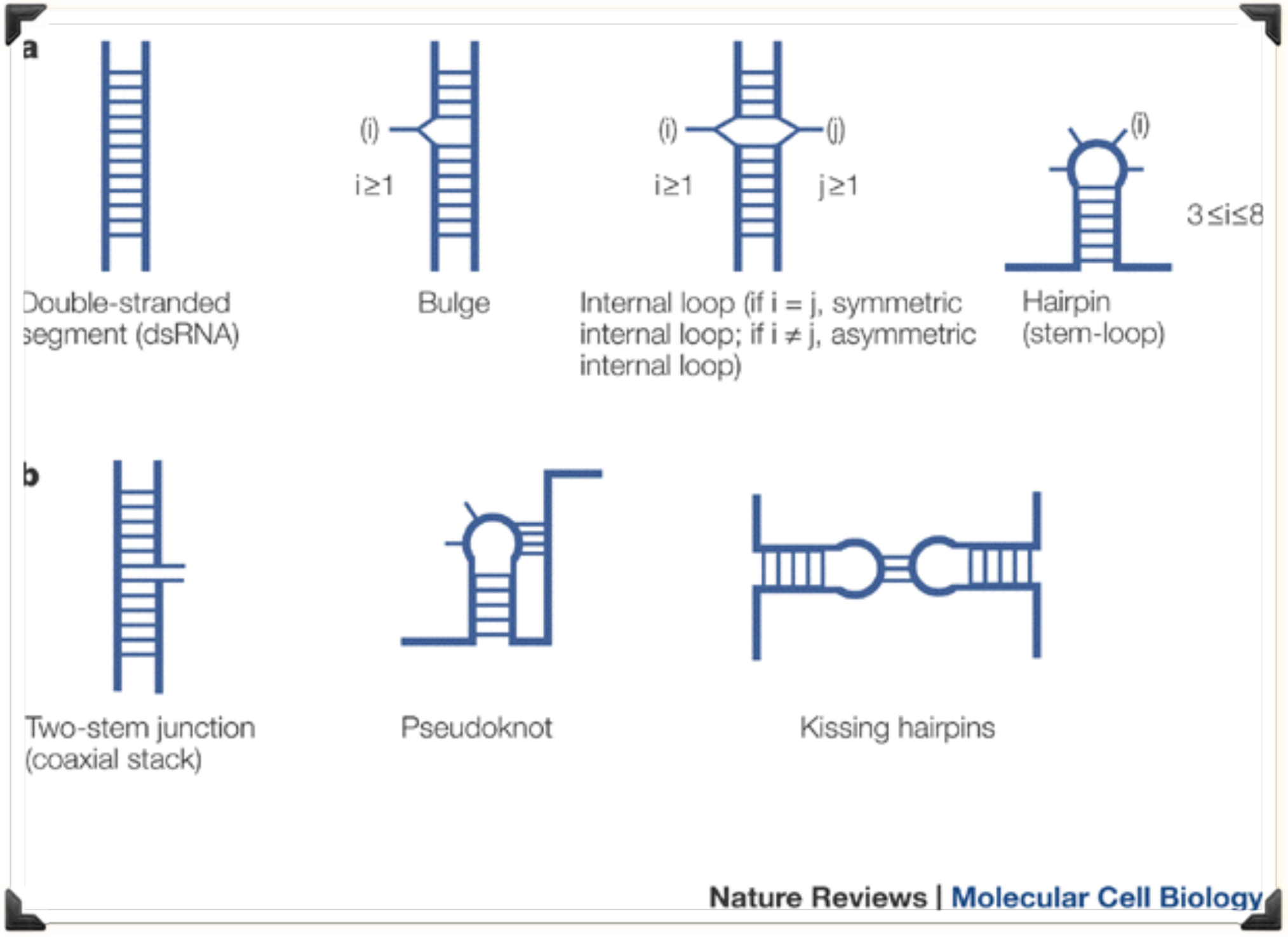
riboswitch

ribozyme

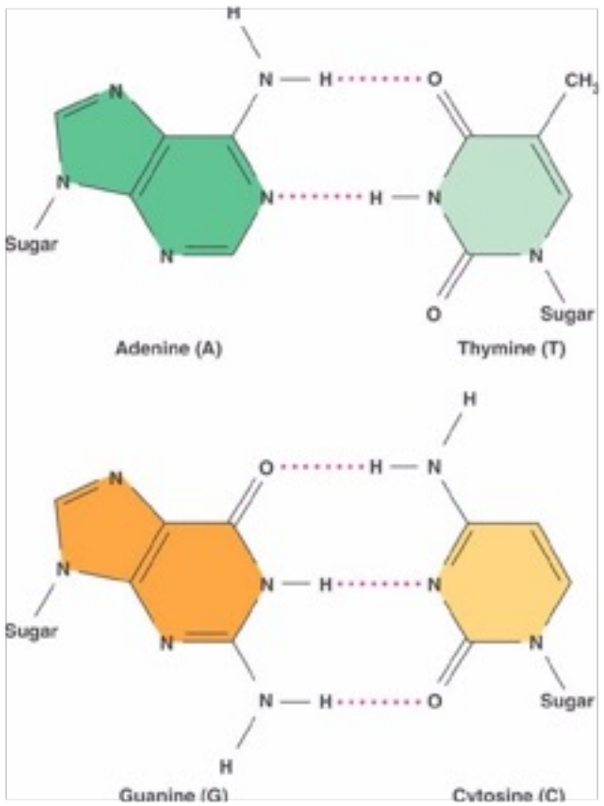
triple helix

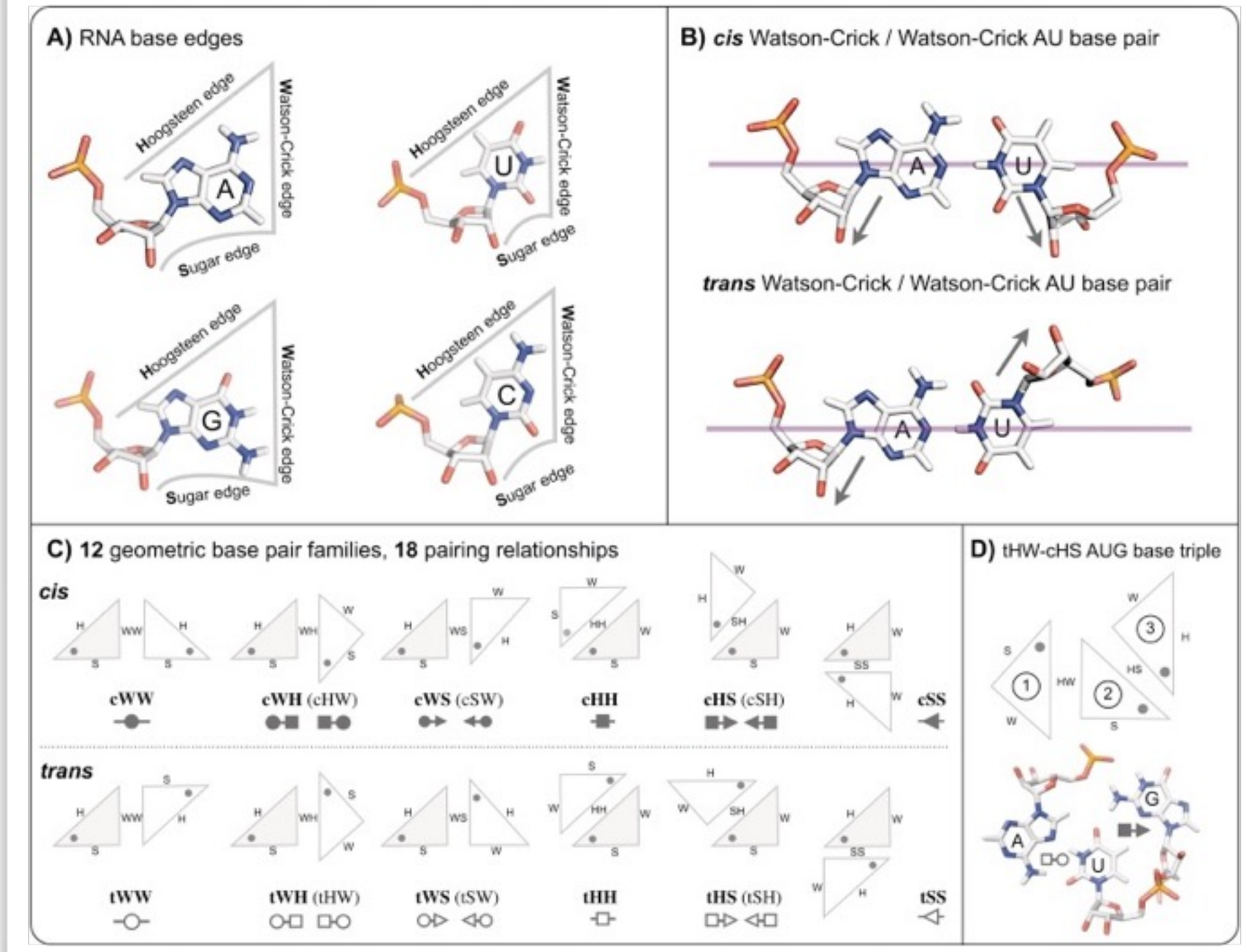


Secondary structures classification



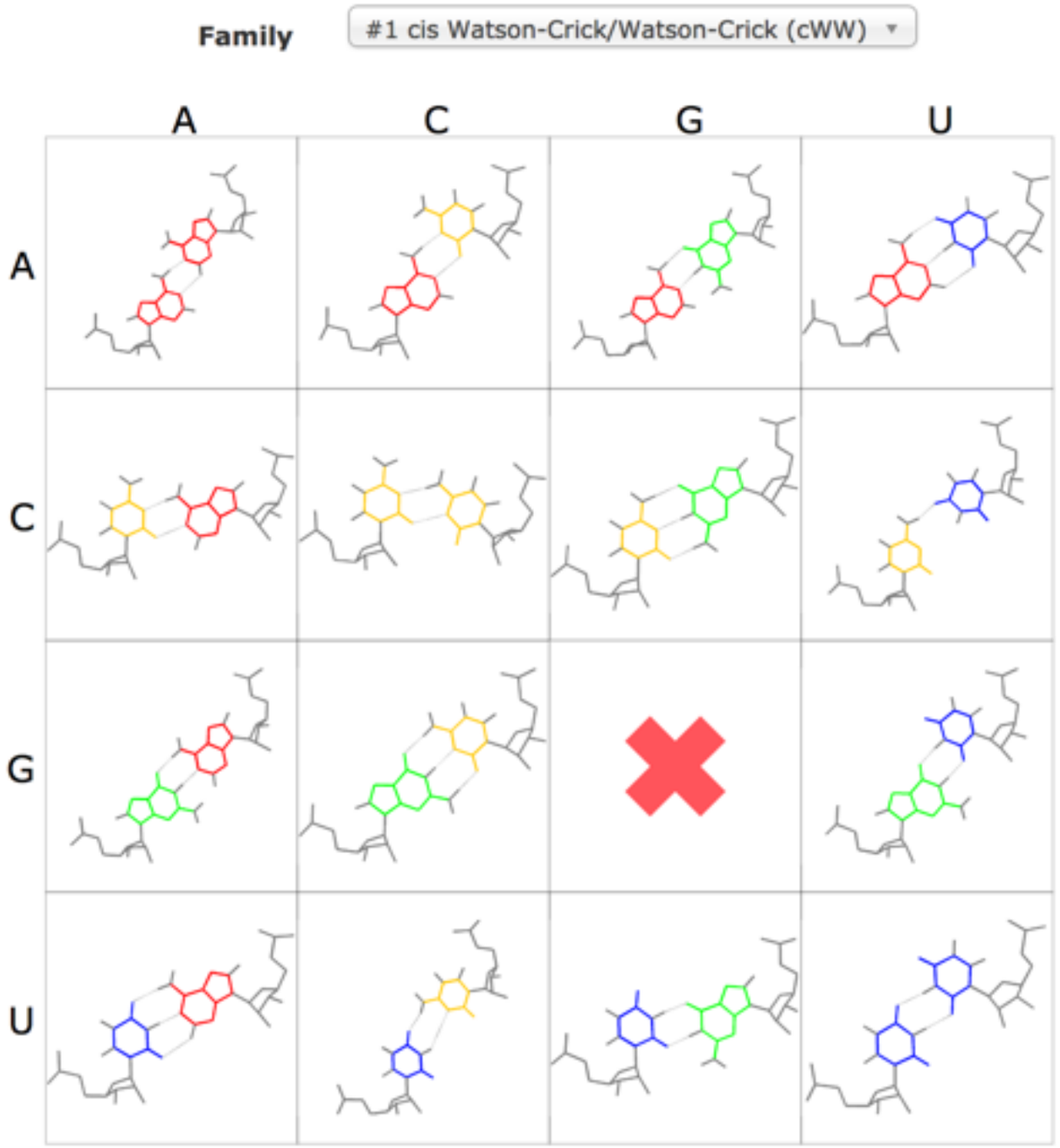
Canonical and non-canonical base pairing

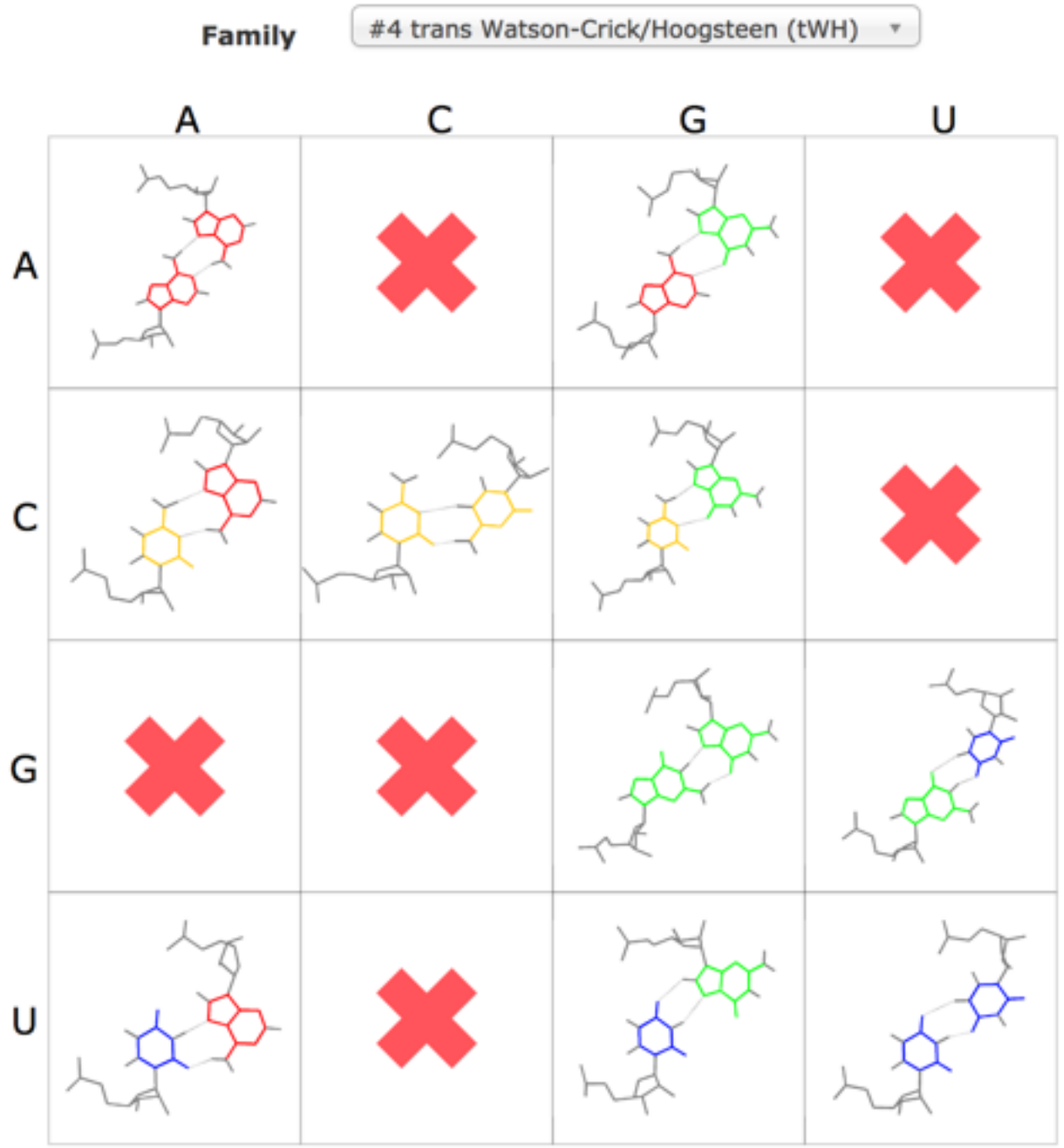




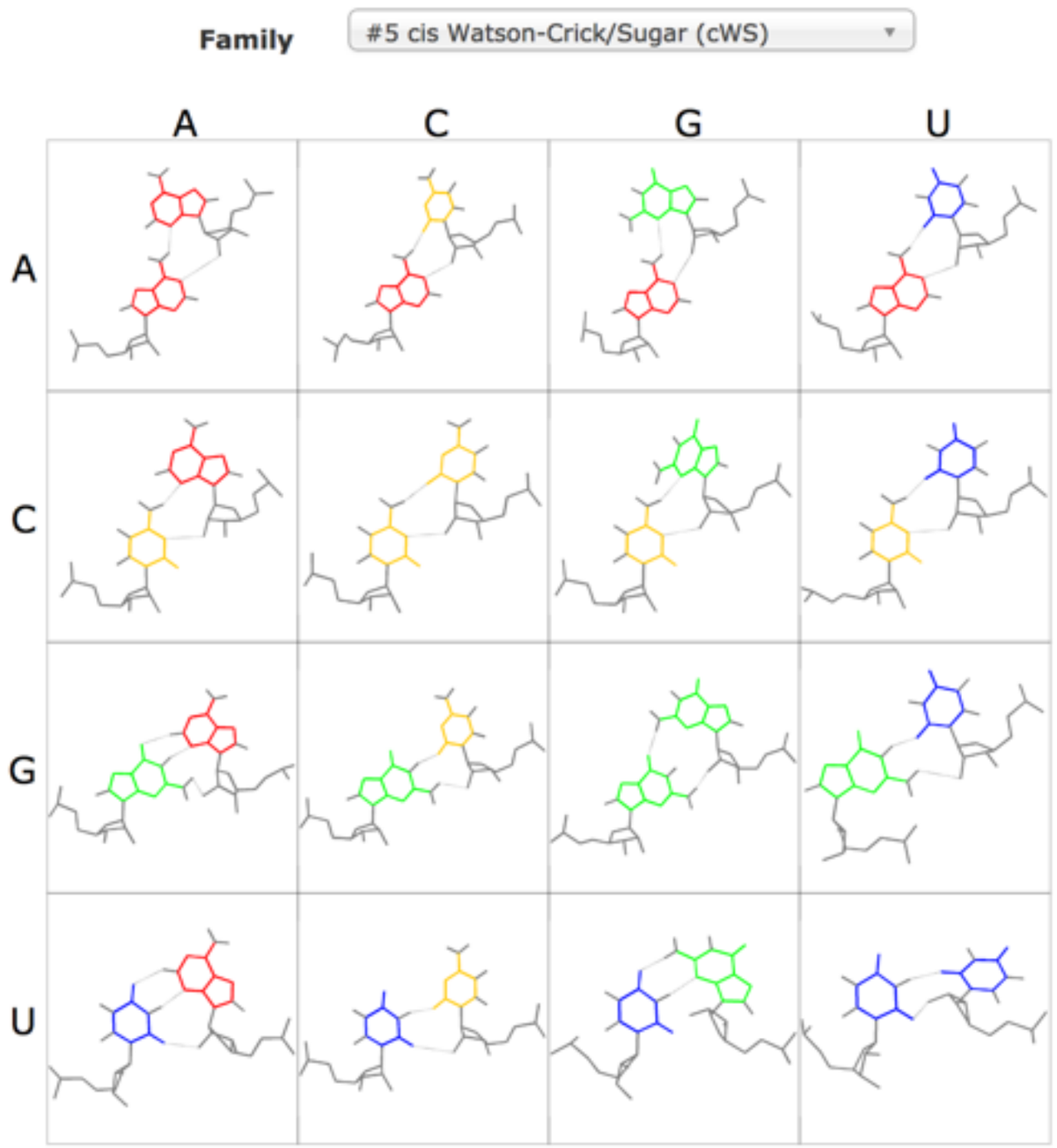
288 theoretically possible pairs ⇢ 145 found experimentally (NDB)

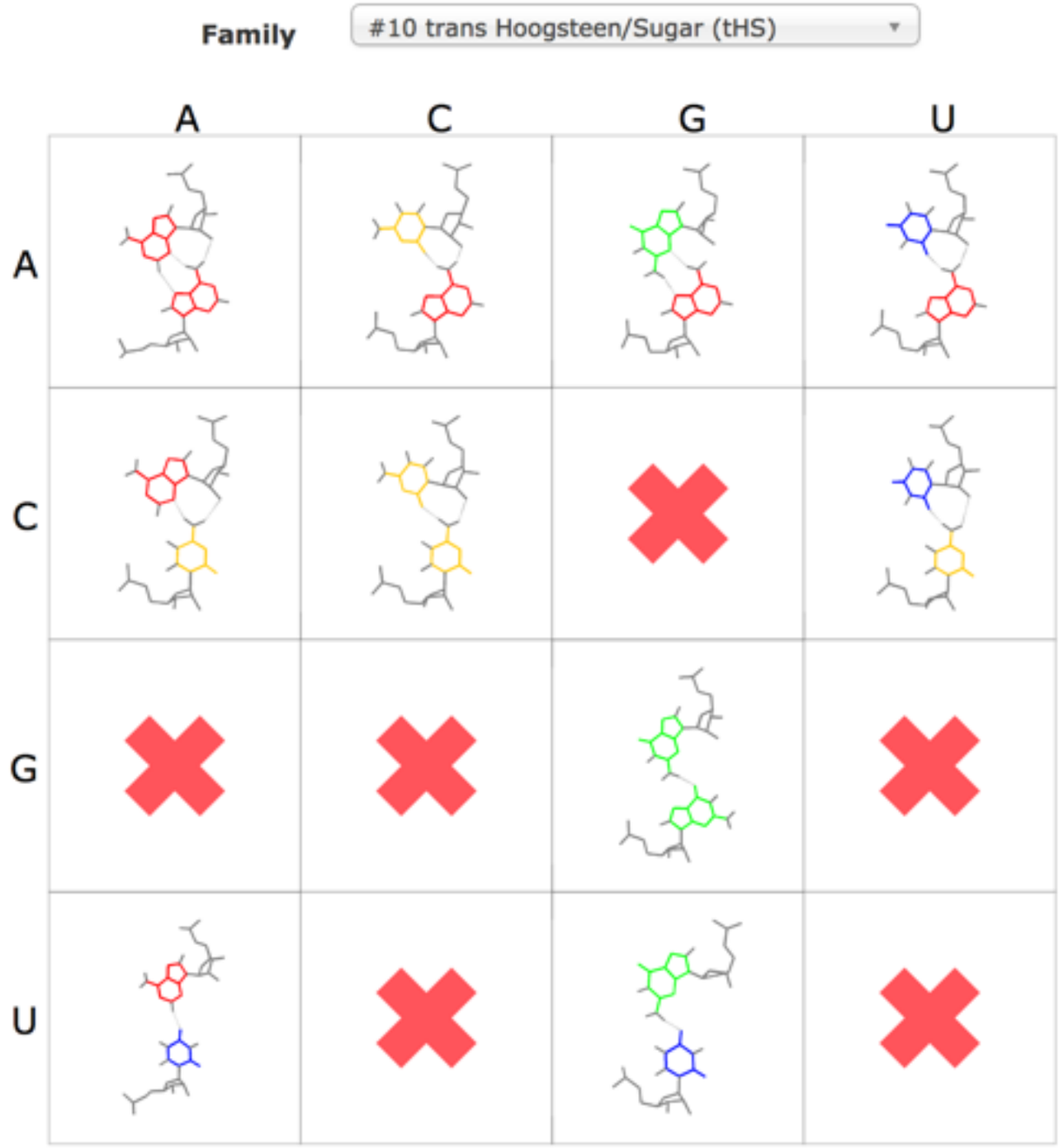
non-canonical base pairs





non-canonical base pairs





Exercice 1
Look at 3D structures

http://web.x3dna-dssr.org/

Look at PDB codes: 1Y26, 1L2X, 2K96

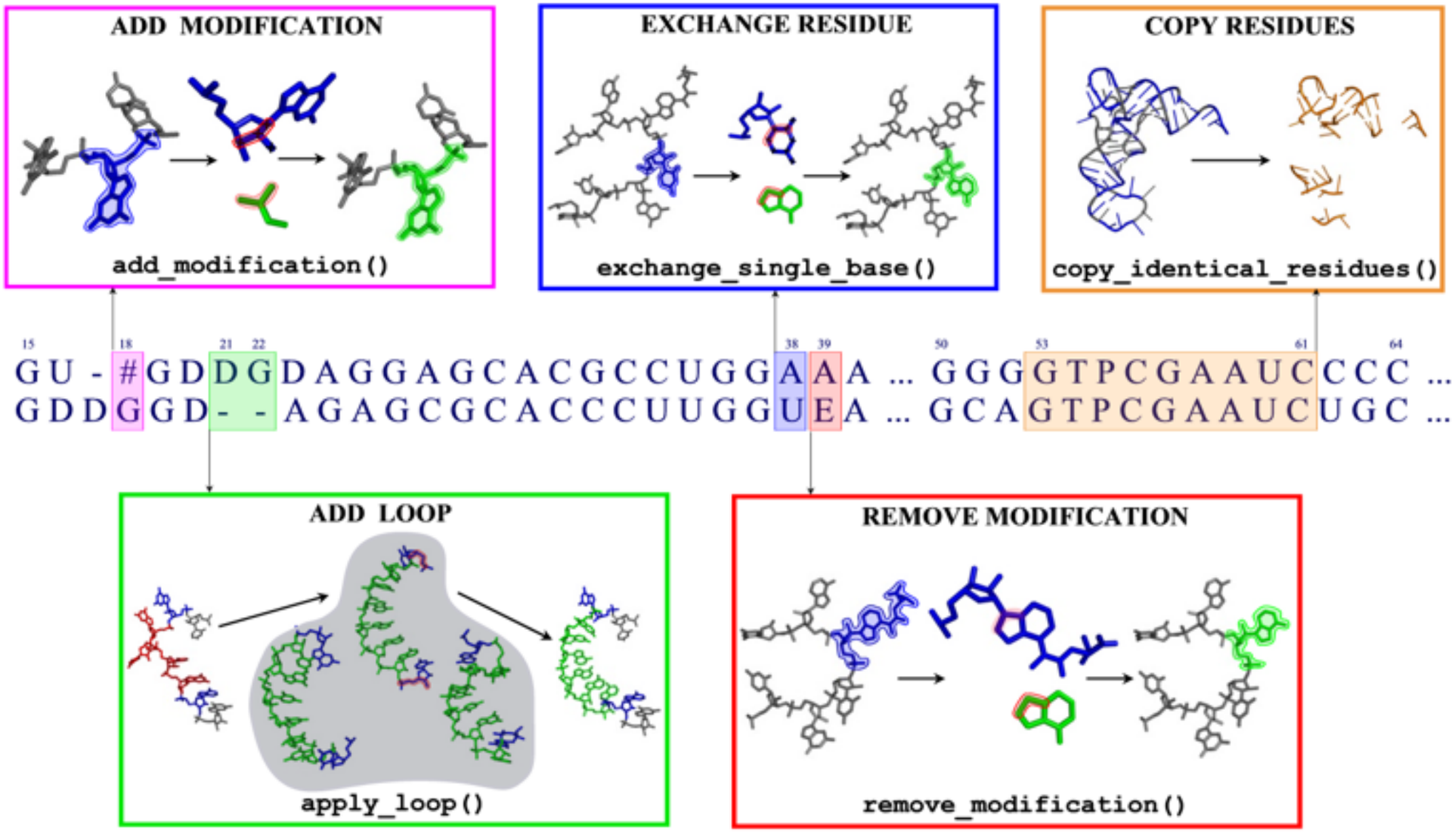
- For each system detect canonical pairs and non-canonical pairs.
- Are there multiple pairs (triplets or quadruplets)?
- How many stems are in the structure?
- What are the tertiary contacts?
- What is the secondary structure (topology) classification?

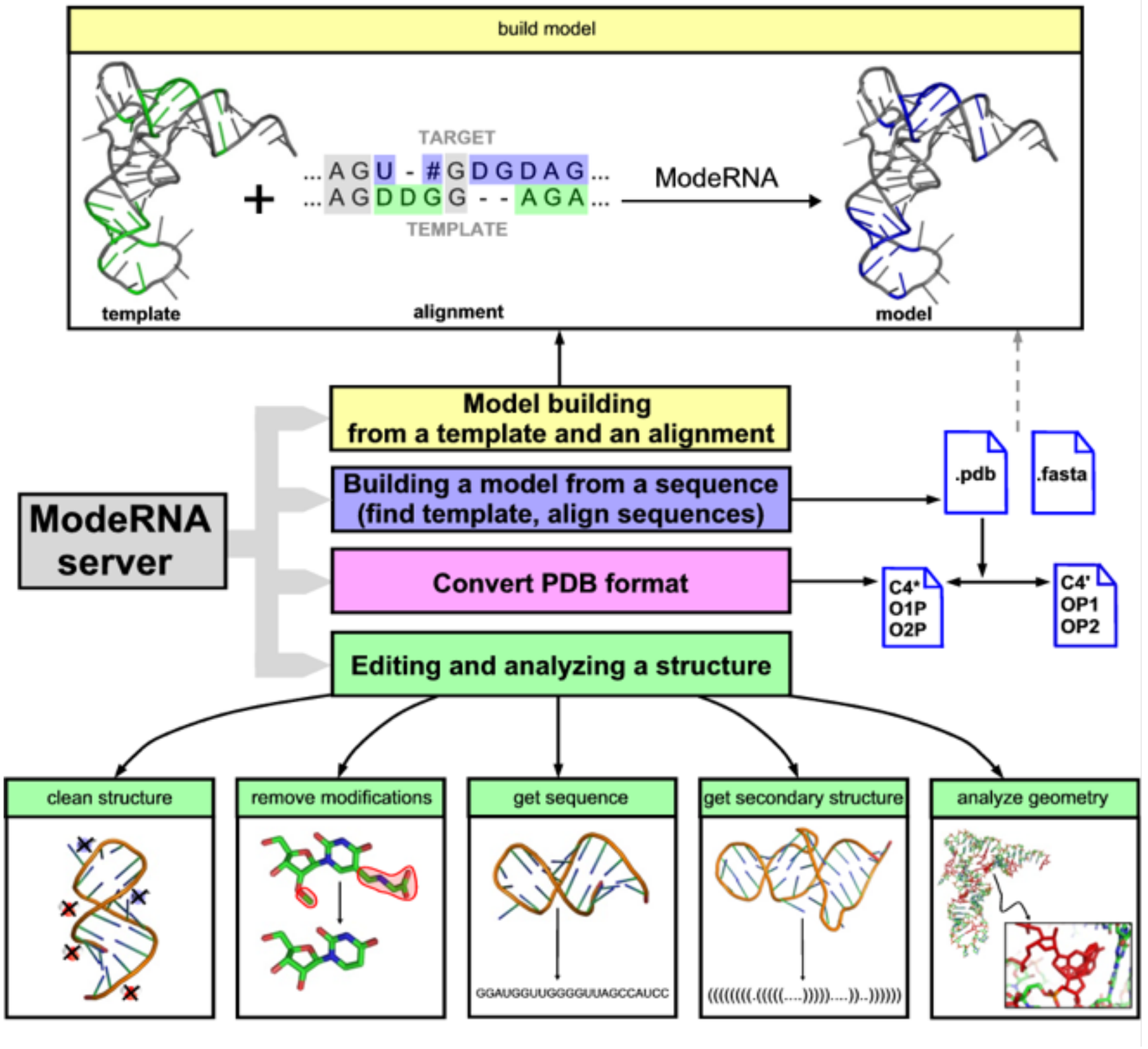
3D structure predictions

Homology modeling

Structures inferred by comparison with resolved structures of similar sequences

Reliable sequence alignment
Extensive structural database




build model
TARGET
... A G U - # G D G D A G ...
... A G D D G G - - A G A ...
TEMPLATE
ModeRNA
template
alignment
model
Model building
from a template and an alignment
Building a model from a sequence
(find template, align sequences)
Convert PDB format
Editing and analyzing a structure
ModeRNA
server
.pdb
.fasta
C4*
O1P
O2P
C4'
OP1
OP2
clean structure
remove modifications
get sequence
get secondary structure
analyze geometry
GGAUGGUUGGGGUUAGCCAUCC
(((((((((.(((((....)))))....)))..)))))))

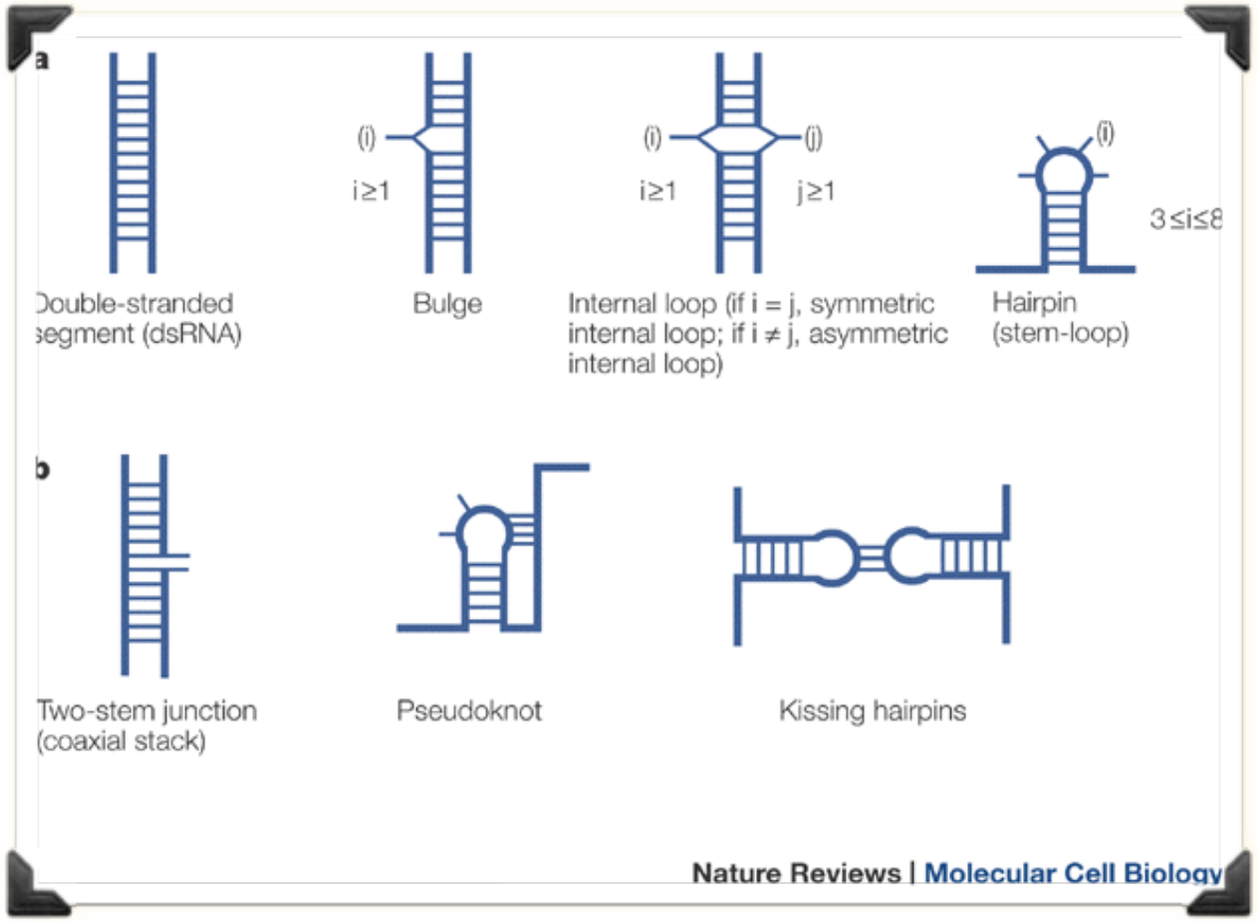
Exercice 2
Use online homology model server

http://iimcb.genesilico.pl/modernaserver/

- Launch test case 1QF6_B_tRNA.pdb
- Download 1QF6_B_tRNA.pdb on your machine and analyze it using DSSR.
- Download results from homology modeling and analyze them using DSSR.
- Detect the main similarities and differences between the two structures by comparing the DSSR results.
- Compare the 3D structures using a visualization software (Pymol, Chimera, VMD)

If you don't have homologues...

Hierarchical folding hypothesis



Most of the energy of folding goes in the formation of double helices

+

Double helical regions form first

↓

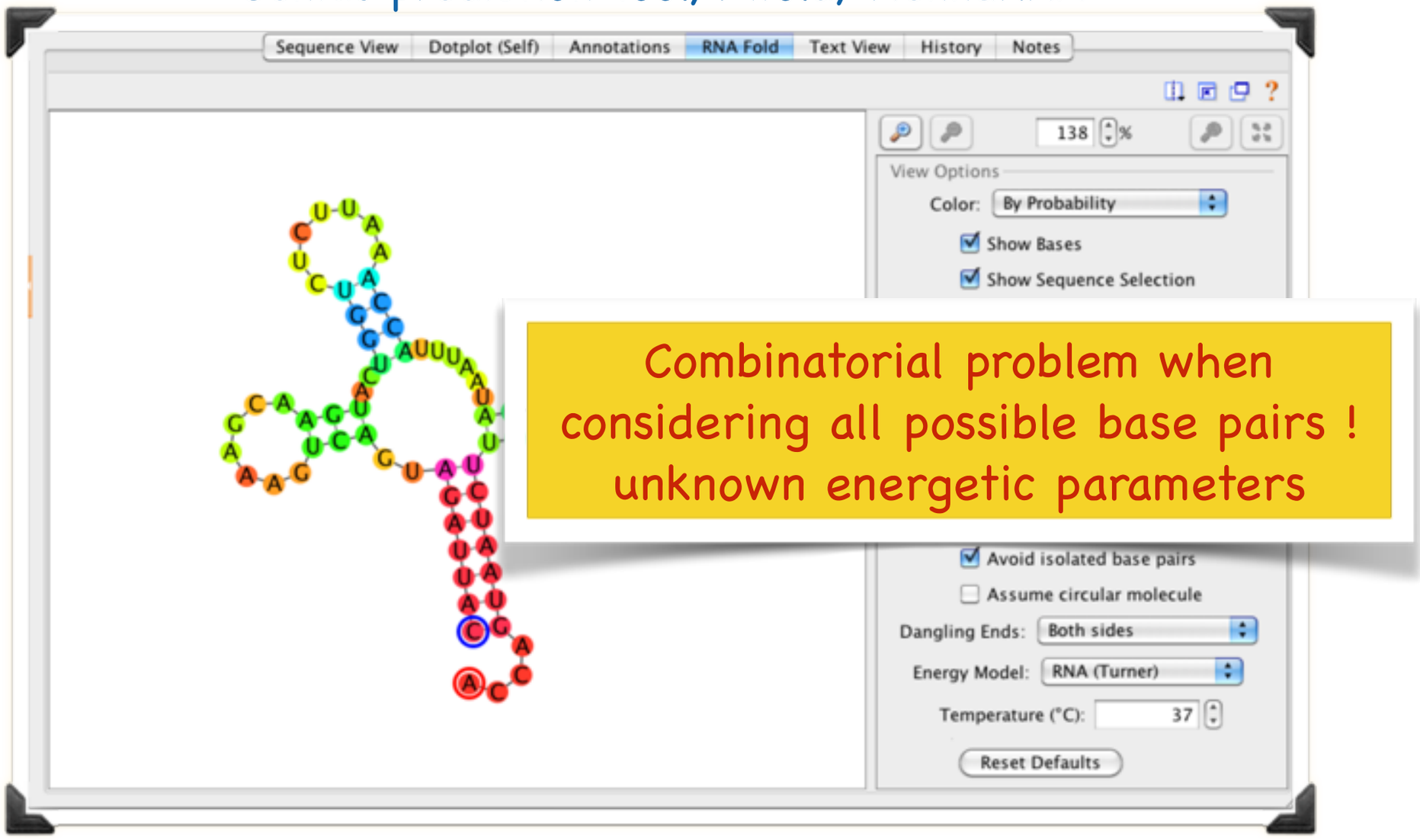
3D (non-helical) pairs can be considered as a "perturbation" to the secondary structure

↓

Find the 2D structure first, then move in space to find 3D contacts

Secondary structure predictions

Genius prediction tool, Mfold, ViennaRNA



Combinatorial problem when considering all possible base pairs !
unknown energetic parameters

- Based on a thermodynamic model considering the free energy of pair formation and stacking with the following pair : nearest-neighbor (Turner's) model
- Very fast

- Consider only canonical paris
- Does not include pseudoknots

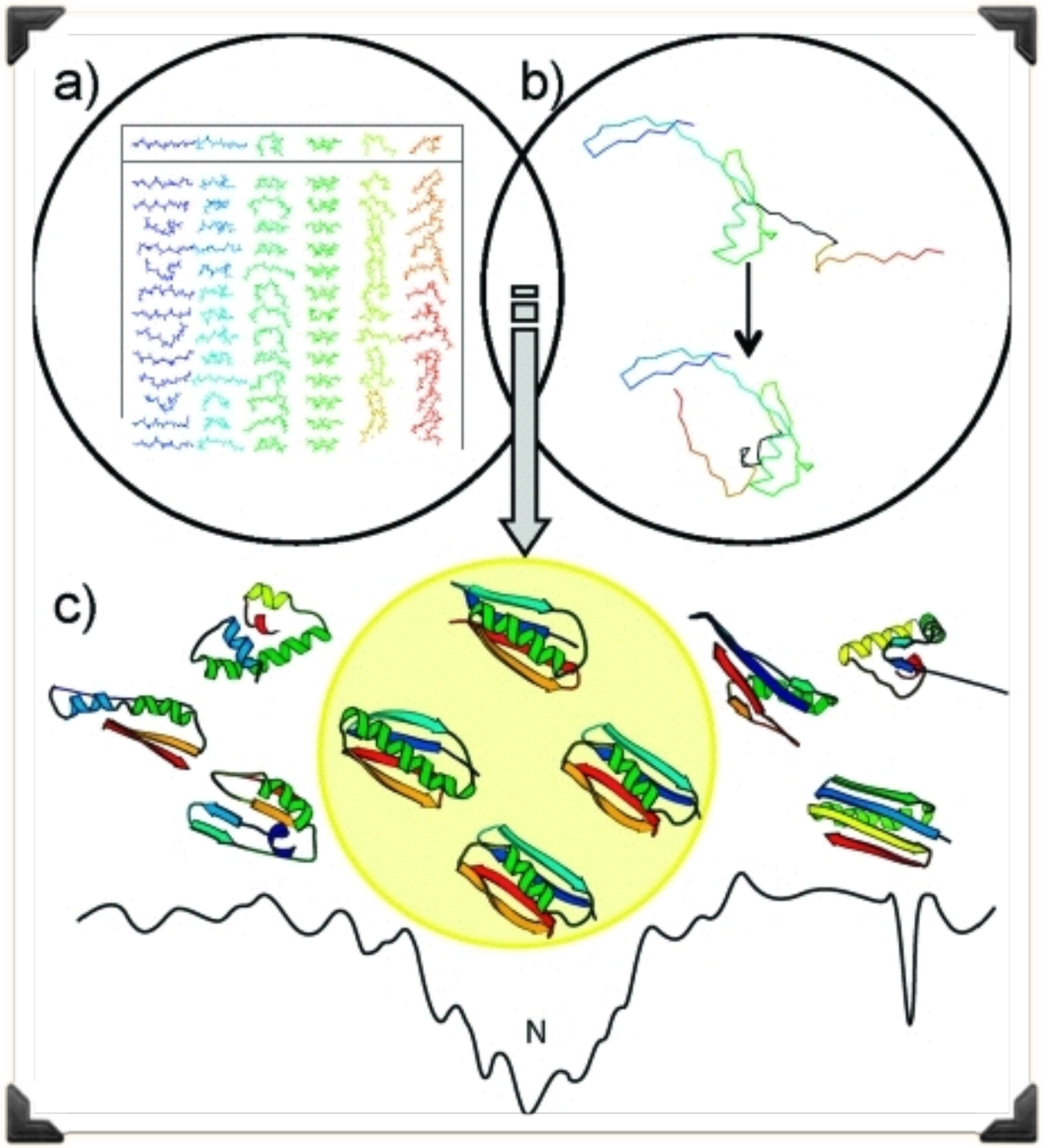
3D structure predictions

Hybrids methods

Fragment reconstruction, 3D scaffolds, ...

Good at predicting local structures

Strongly rely on prediction of secondary structures first



Analyze sequence as short fragments (4-5 nt)

Predict 2D structure and assign each fragment to a specific 2D

Search a structural database for fragments of the same sequence and same 2D environment.

Assemble all fragments.

Relax (minimize) structure according to a given force field.

Exercice 3
Use online fragment reconstruction server

http://rnacomposer.ibch.poznan.pl/

- Launch test case
  #HIV-2 DIS RNA hairpin
  `>example1`
  `GCUCCUAGAAAGGCGCGGGCCGAGGUACCAAGGCAGCGUGUGGAGC`
  `((((((.....(((((..(((.....)))..)))))..))))))`

- Download the results on your machine and analyze it using DSSR.

- Launch a second test in which you modify the secondary structure given to the server (changes highlighted in red)
  #HIV-2 DIS RNA hairpin
  `>example1`
  `GCUCCUAGAAAGGCGCGGGCCGAGGUACCAAGGCAGCGUGUGGAGC`
  `((((((.....(((((.....)))..)))))..))))))`

- Download the results on your machine and analyze it using DSSR.

- What are the differences between the two 3D structures?
  (you can use a web server such as this: http://rna.ucf.edu/WebSTAR3D or your favorite visualization software)

Exercice 4
Use online scaffold modeling server

http://rna.physics.missouri.edu/vfold3D/index.html

- Launch test case

```
>example1
GCUCCUAGAAAGGCGCGGGCCGAGGUACCAAGGCAGCGUGUGGAGC
((((((.....(((((..(((.....)))..)))))..))))))
```

- Download the results on your machine and analyze it using DSSR.
- Compare the two prediction for this same system (RNAcomposer and Vfold)

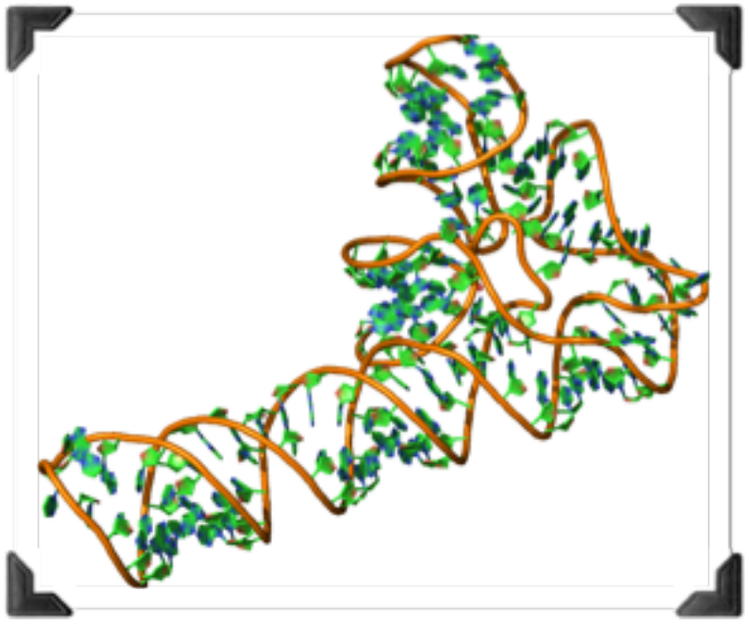
- Launch a second test in which you modify the secondary structure given to the server (changes highlighted in red:

```
>example2
GCUCCUAGAAAGGCGCGGGCCGAGGUACCAAGGCAGCGUGUGGAGC
((((((.....(((((.....)))..)))))..))))))
```

- Download the results on your machine and analyze it using DSSR.

- Compare the two prediction for this same system (RNAcomposer and Vfold) (http://rna.ucf.edu/WebSTAR3D)

Limitations



VFold : 3.7 Å RMSD
FARFAR : 4.5 Å RMSD

VFold : 22.5 Å RMSD
FARFAR : 11.6 Å RMSD

2D helps, but it's not the end of the story

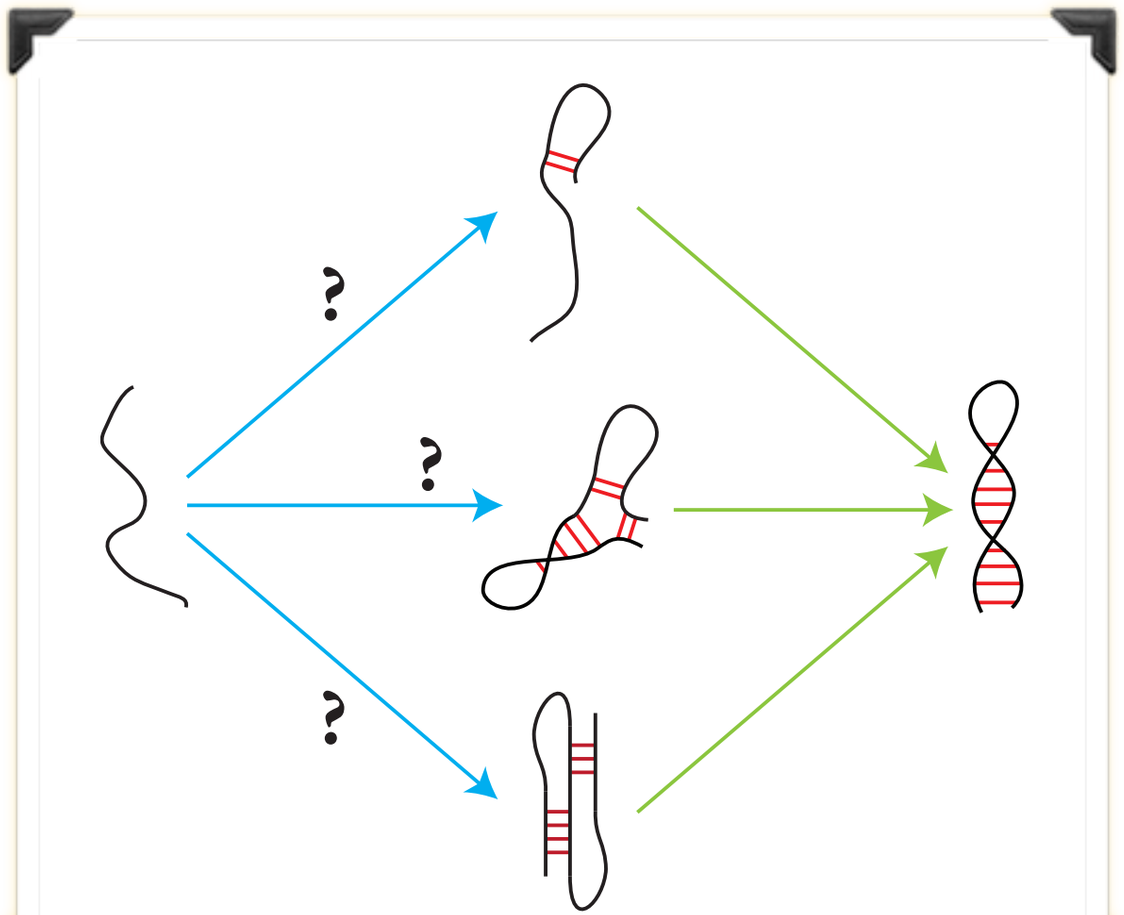
- canonical base pairing occurs in helices, but does not regulate long range interactions that ultimately determine the 3D structure (often non-canonical)
- pseudoknots are hard to account for.
- some molecules undergo BP shifts through their life that alter the 3D structure completely
- interactions that hold the molecule together are often non-canonical pairings

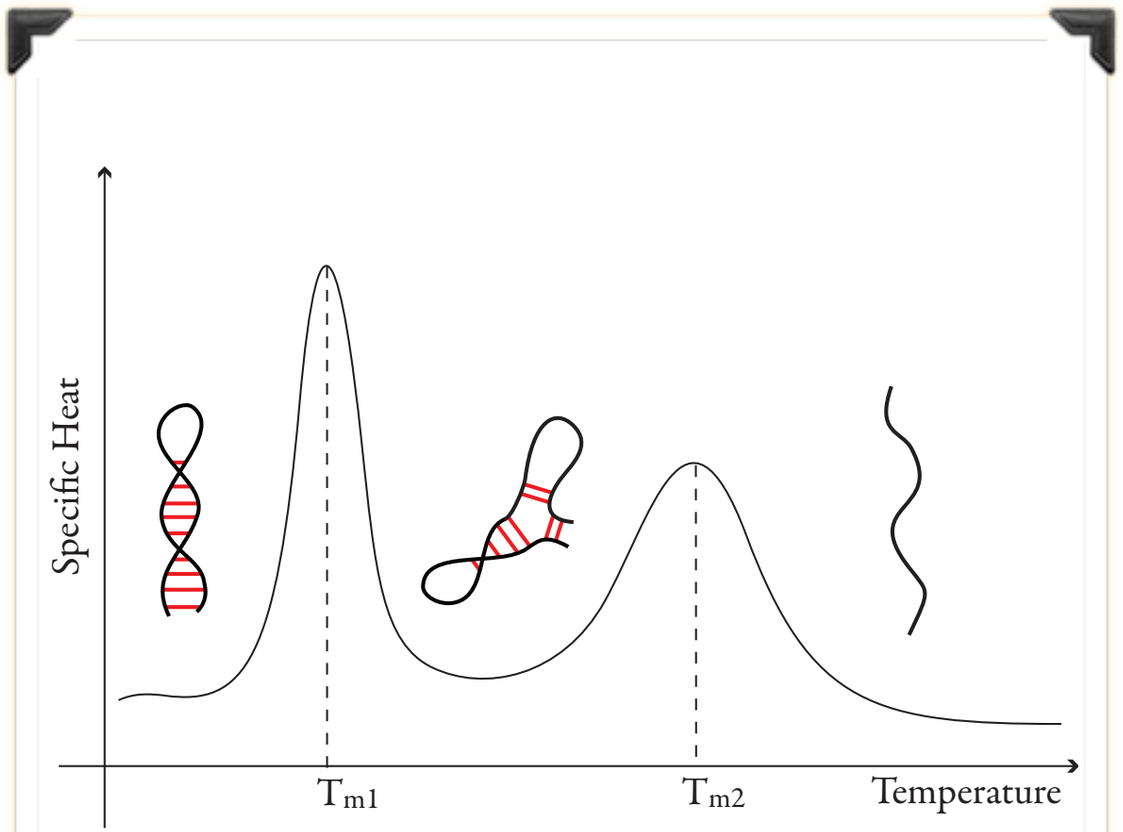
Physics-based simulations

Ab initio modeling

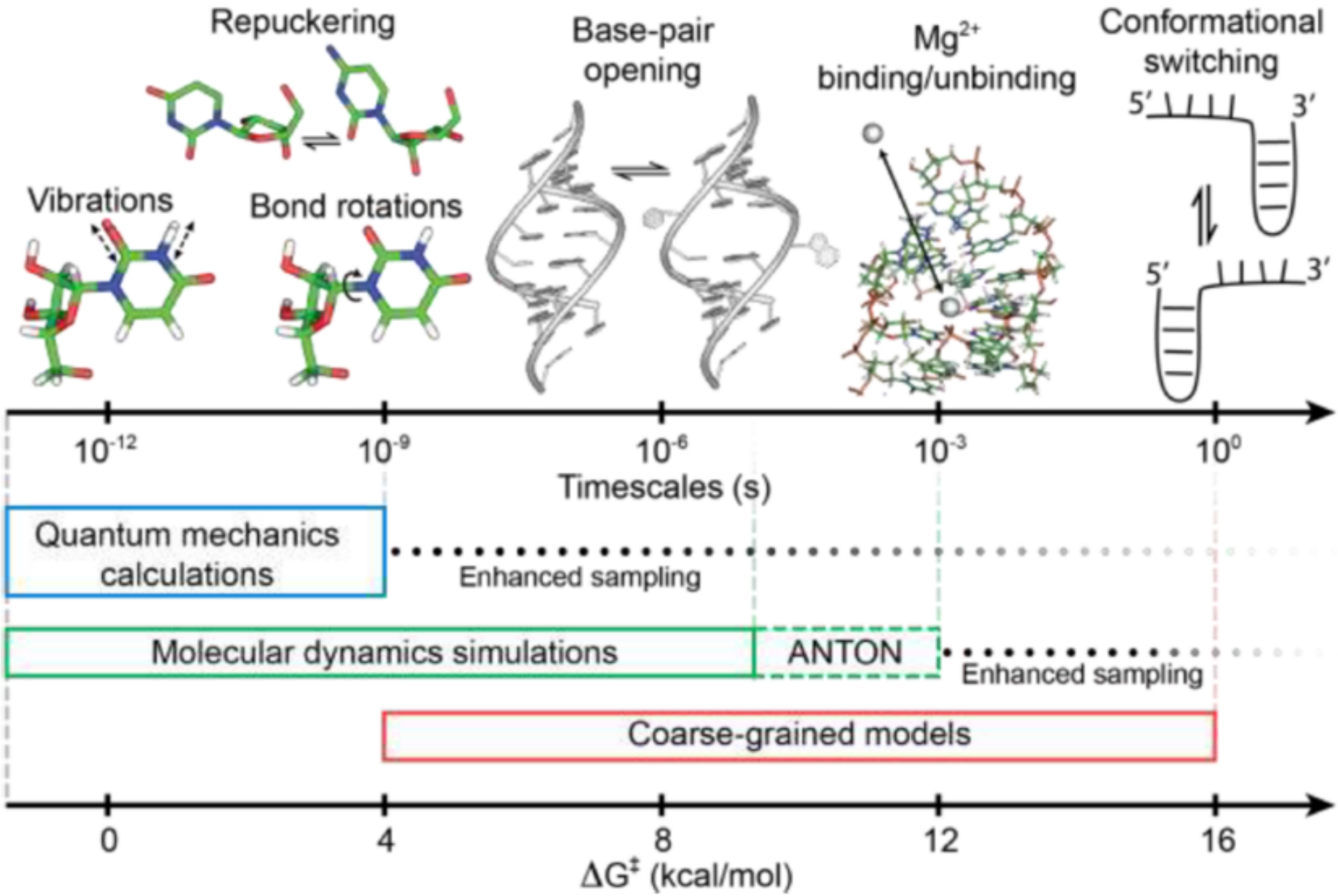
Physical models of the system, aiming at predicting equilibrium structures, folding intermediates, thermodynamics, kinetic barrier,...

Quantum : base-base interactions
Atomistic : stability of known structures
Coarse-grained : folding of a given sequence





RNA time scales and simulations



Simulation 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*

Simulation methods

Classical molecular dynamics - force based

particles subject to force-field and friction (T dependent)

Generate conformation S_i

Compute all potential energies for S_i

Compute all forces : $F = -\frac{dV}{dx}$

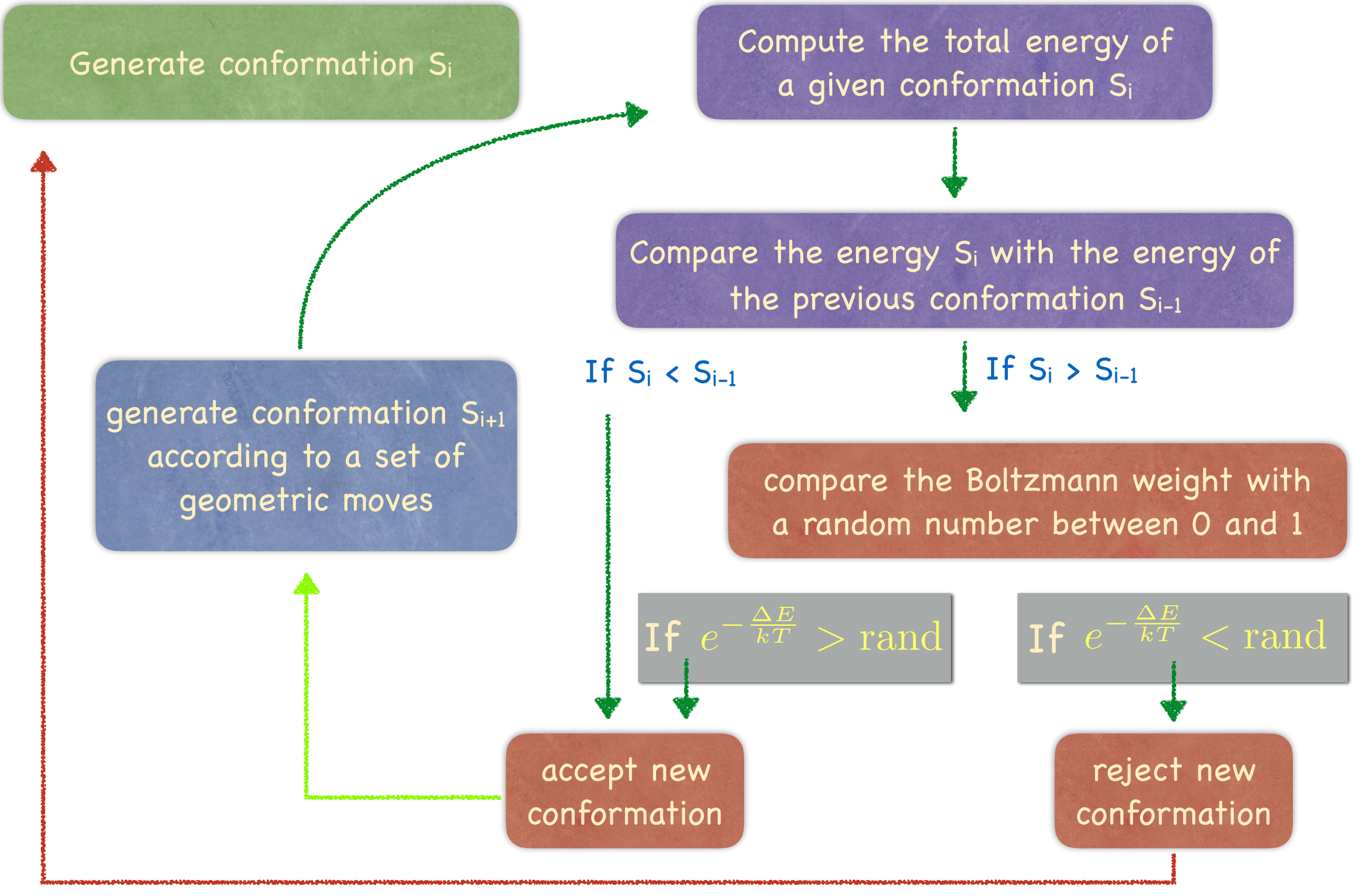
Accelerations from Netwon's equation F=ma

Integrate to find x(t) and v(t): trajectory

Move according to the trajectory for a time τ

Simulation methods

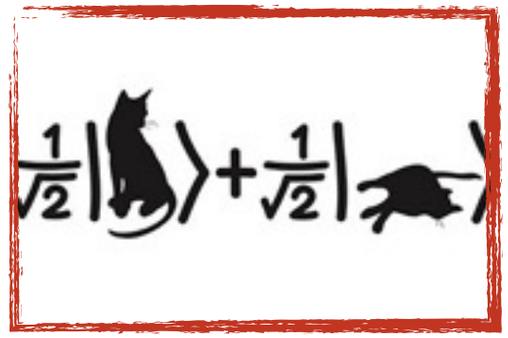
Monte Carlo - energy based

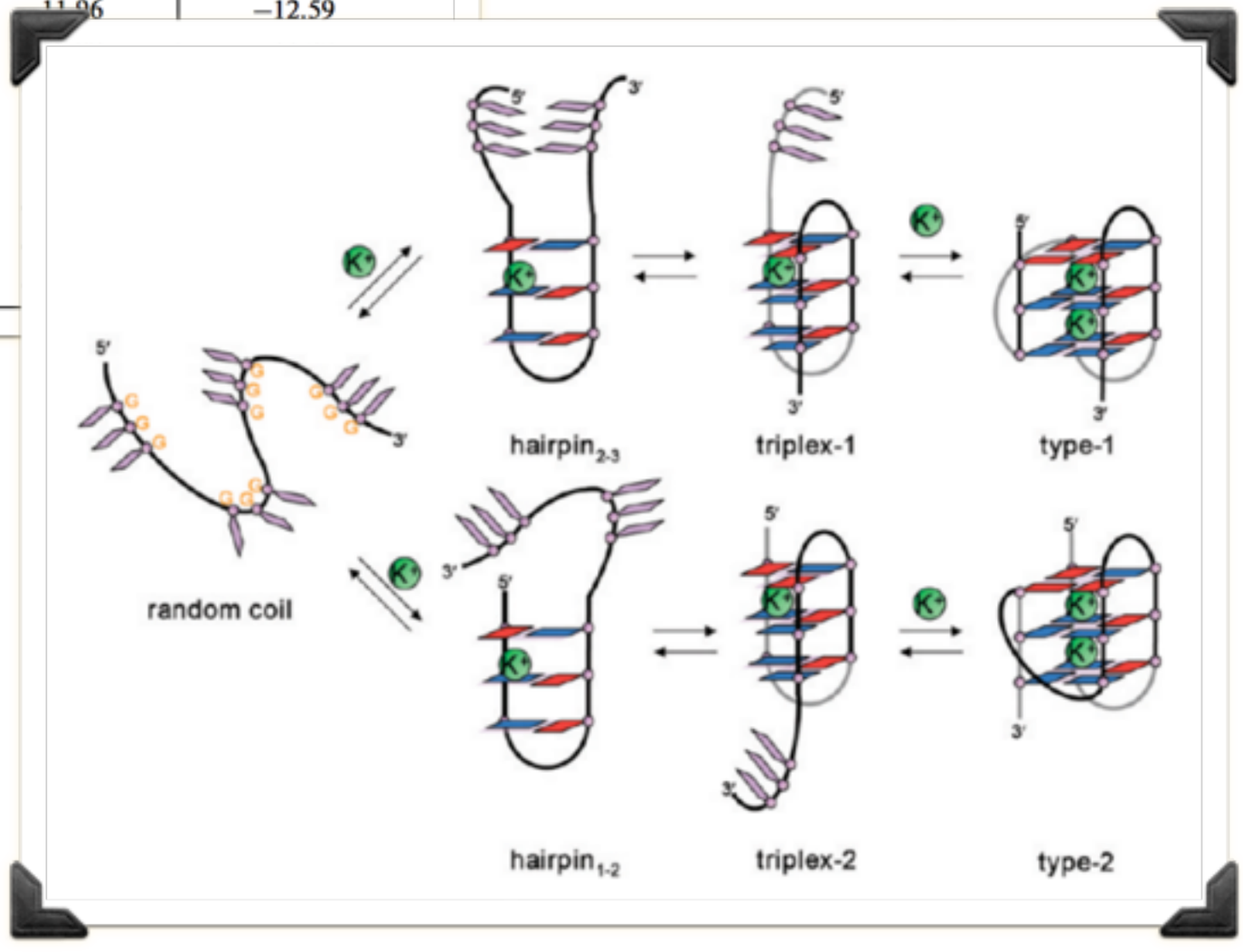


Quantum mechanics calculations

Table 1. Energies of water–base interaction (kcal/mol) calculated by different methods in local minima

Minima	Centers of hydrogen bonds	QM	PM	AMBER BCC	AMBER RESP
		m^9Ade (−10.6[a])			
1	N7, H62	−9.34	−10.84	−11.23	−11.08
2	N1, H61	−8.39	−8.42	−9.40	−9.53
3	N3	−6.58	−6.97	−7.48	−8.08
		m^1Thy (−10.4[a])			
1	O4	−4.62	−7.07	−8.59	−7.70
2	O4, H3	−7.72	−8.85	−9.17	−8.97
3	O2, H3	−7.39	−8.07	−8.90	−8.99
4	O2	−5.49	−6.89	−9.31	−8.55
		m^9Gua (−14.0[b])			
1	N7, O6	−7.69	−12.41	−9.99	−10.44
2	O6, H1	−10.42	−11.94	−11.96	−12.59
3	H1, H21	−9.02	−12.20		
4	N3, H22	−7.93	−9.86		
		m^1Cyt(−11.4[a])			
1	H42	−5.42	−7.79		
2	N3, H41	−9.68	−11.22		
3	N3, O2	−6.81	−11.38		
4	O2	−6.07	−8.67		

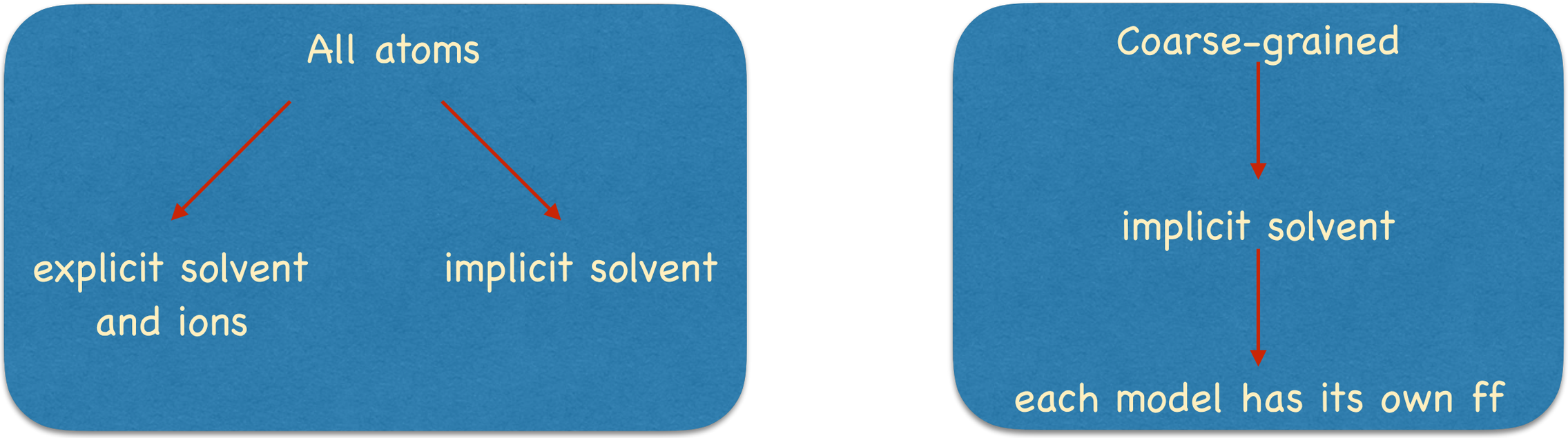




Atomistic Force Fields

to compute both energies and forces for all kinds of simulations

Force fields depends on the representation used for the particles
Typically referred to classical mechanics



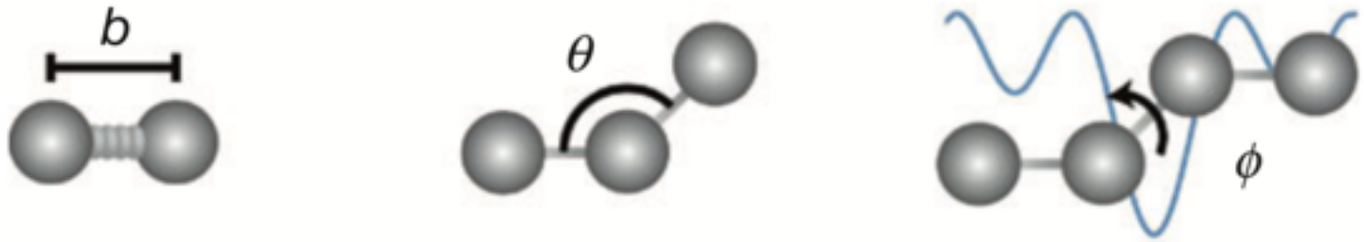
CHARMM AMBER ... later

Vangaveti, Ranganathan, Chen, “Advances in RNA molecular dynamics: a simulator's guide to RNA force fields”, Wiley Interdiscip Rev RNA. 2017

Šponer J, Bussi G, Krepl M, et al. RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. Chem Rev. 2018

Atomistic force fields

$$V(R)_{bonded} = \sum_{bonded} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{torsions} \sum_{n} k_\phi \cos(n\phi - \gamma)$$



$$V(R)_{nonbonded} = \sum_{van\ der\ waals} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \right) + \sum_{electrostatic} \left(\frac{q_i q_j}{\varepsilon_D r_{ij}} \right)$$

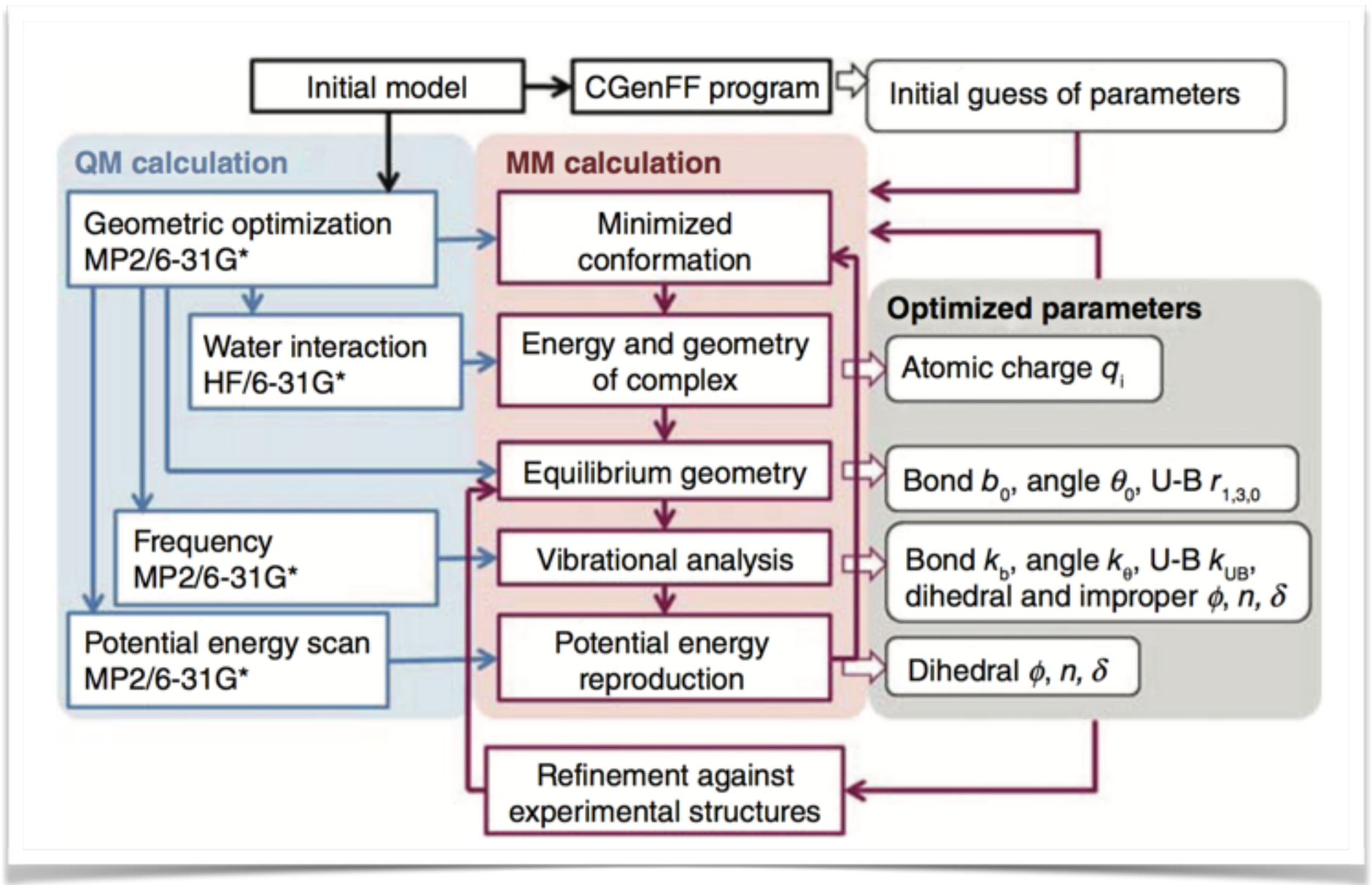
Calibration against QM calculations
May include in the parametrization physical properties of pure model compounds (heat of vaporization, free energy of salvation, ...)
data on conformational equilibria (NMR) and high resolution X-ray structural data.

! While protein folding is governed by hydrophobic effect and side chains packing, RNA folding is given by aromatic stacking, base-pair formation and RNA-ions interactions.

CHARMM

Parametrization philosophy:

- Complete scan of the QM potential energy surface (not just lowest-energy)
- Additional calculations at each H donor/acceptor site with the addition of a TIP3P water molecule -> indirect encoding of solvation properties
- Iterative optimization for best global agreement with input criteria (QM calculations + selected experimental structures and data such as various energies)



CHARMM + and -

Because of the highly delicate and iterative process it is hard to include new compounds such as ligands, modified nucleotides or chemical labels.

Recent publication of the parameters for the 100 naturally occurring modified RNAs

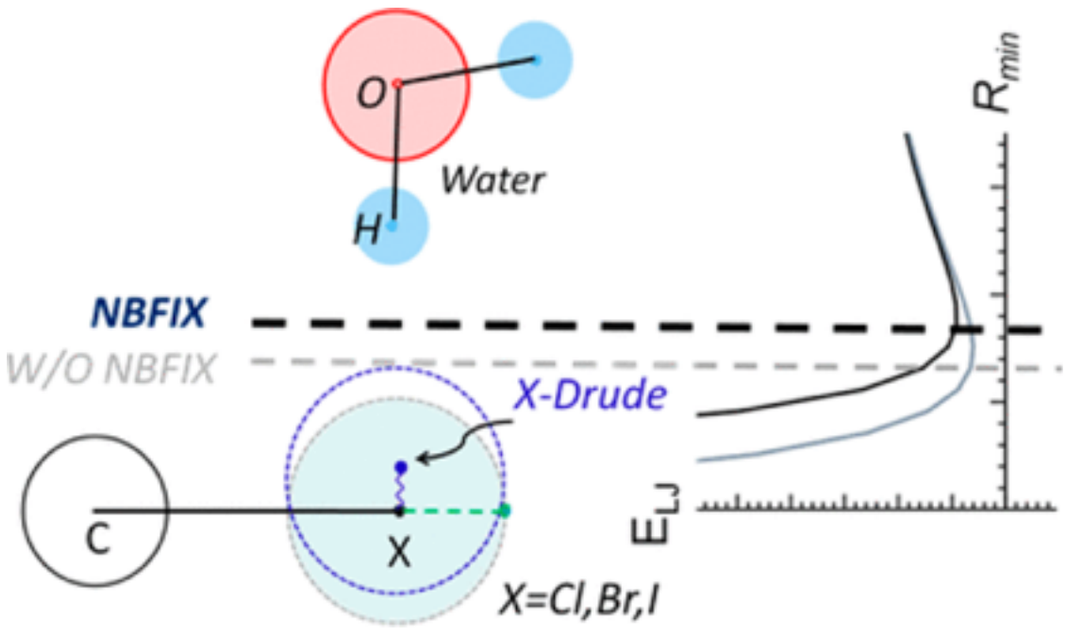
Transferability of the parameters from gas phase to solution phase is assumed (and not explicitly calibrated).

-> only TIP3P water molecules can be used (as they are the ones included in the original parametrization for solvation)

Use of high-level QM calculations give better description of relative energies between competing minima, hydrogen bonding energies and charge dipole moments than HF calculations that are commonly used to calibrate ff (less demanding).

Dispersion-mediated interactions such as base stacking still require higher level of electron correlation to be accurately described.

-> Introduction of some degree of polarizability can help (Drude model).



AMBER

(Assisted Model Building and Energy Refinement)

AMBER ff94 : nucleic acids LJ parameters codeveloped together with other small organic compounds. Specific RNA torsions from fitting a QM (low-level) profile for the phosphate group and for the ribose sugar.

Transitions between A and B-form helices.

AMBER ff99 : refit of all nucleotide torsions against detailed QM PES.

AMBER ff99-BSC0 : α/γ torsions

-> eliminates backbone distorsions of B-form duplexes.

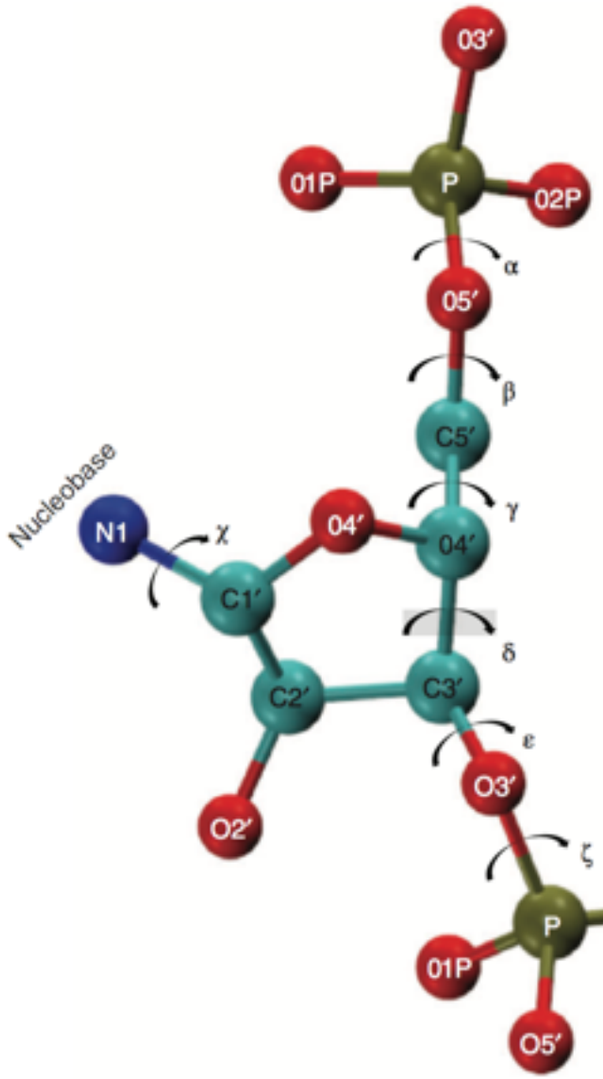
AMBER ff99-BSC1 : ε/ζ torsions (sugar puckering)

-> correctly reproduces twist and roll in DNA

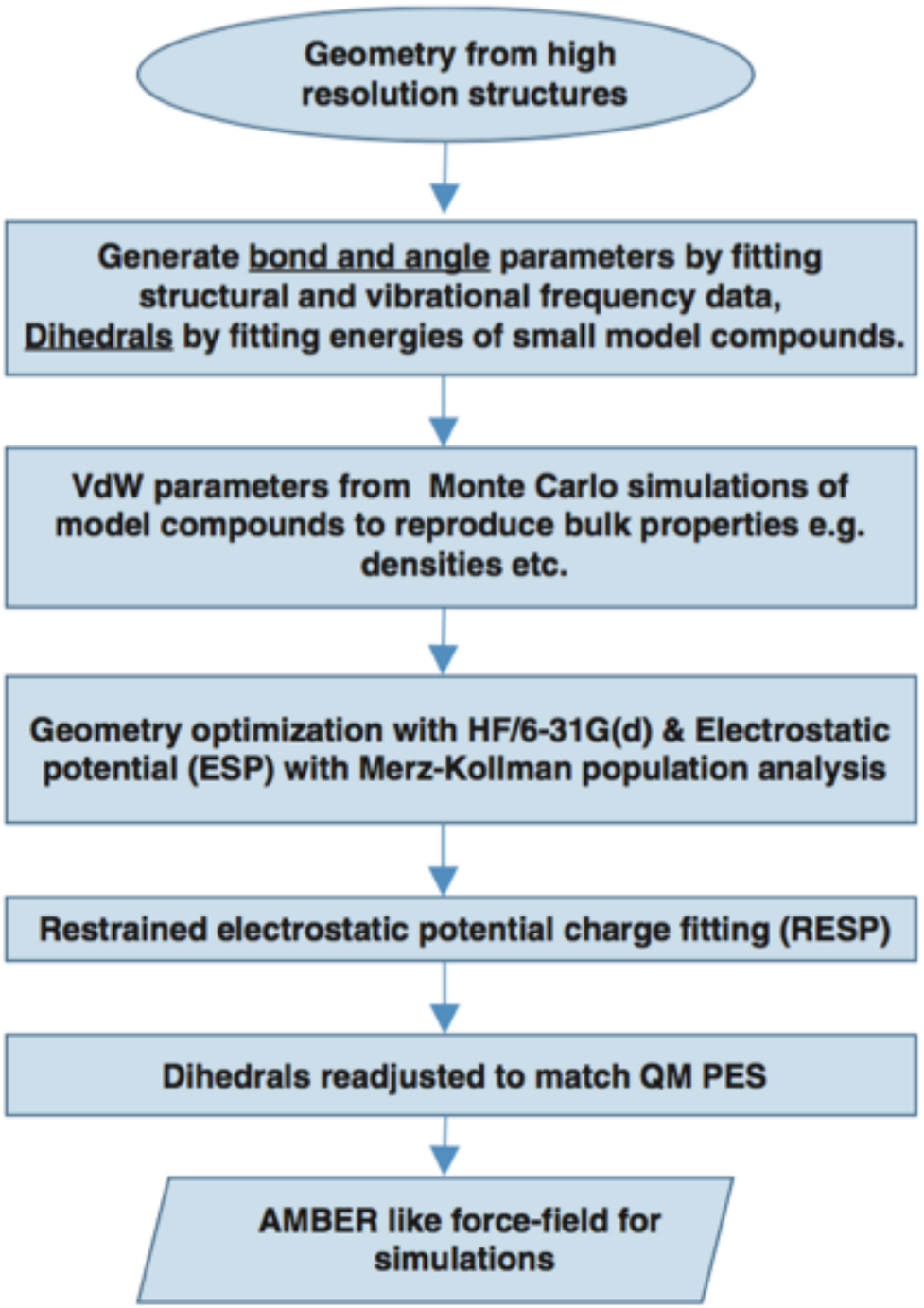
AMBER ff99-χOL : refit of glycosidic torsion χ

-> better anti/syn ratios, more realistic major groove widths.

→ AMBER ff16 for RNA



AMBER



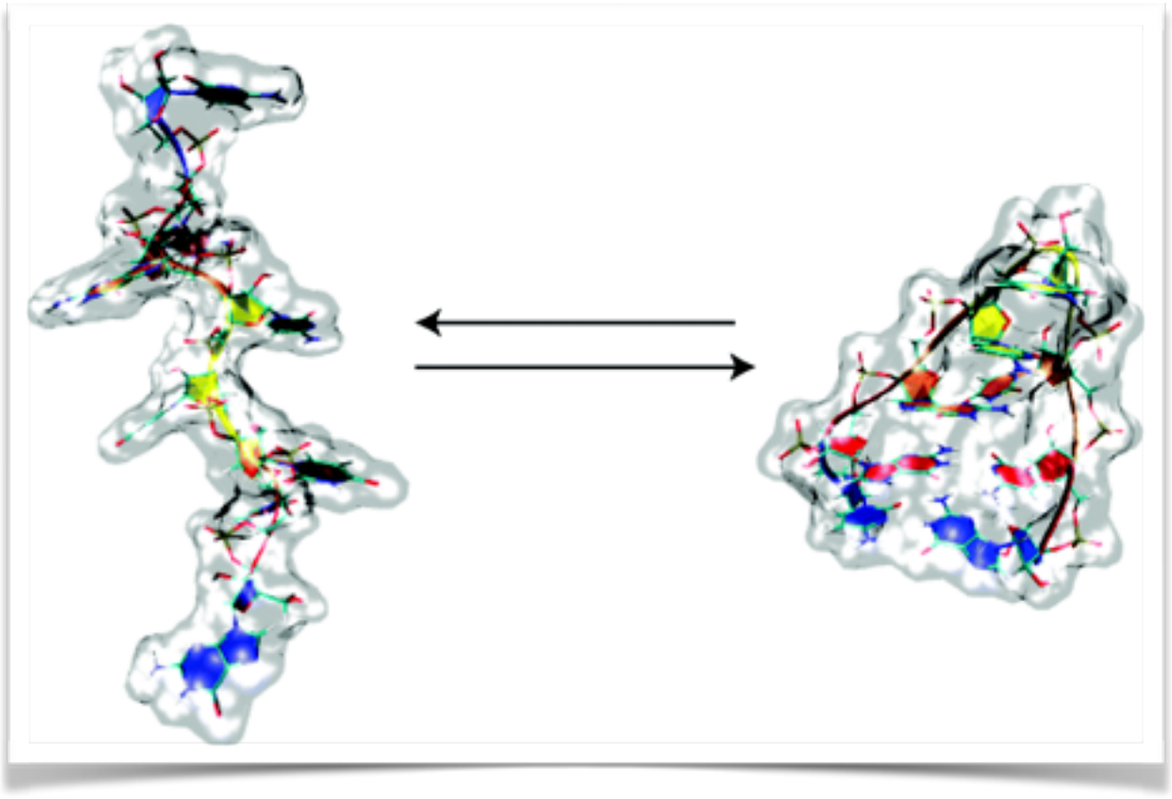
Parametrization protocol is linear, water-model independent, and assumes transferability of parameters developed for small organic molecules

the modular and non-iterative nature of the parametrization scheme allows for improvements to be "easily" introduced and tested (that's why there are a zillions variants...)

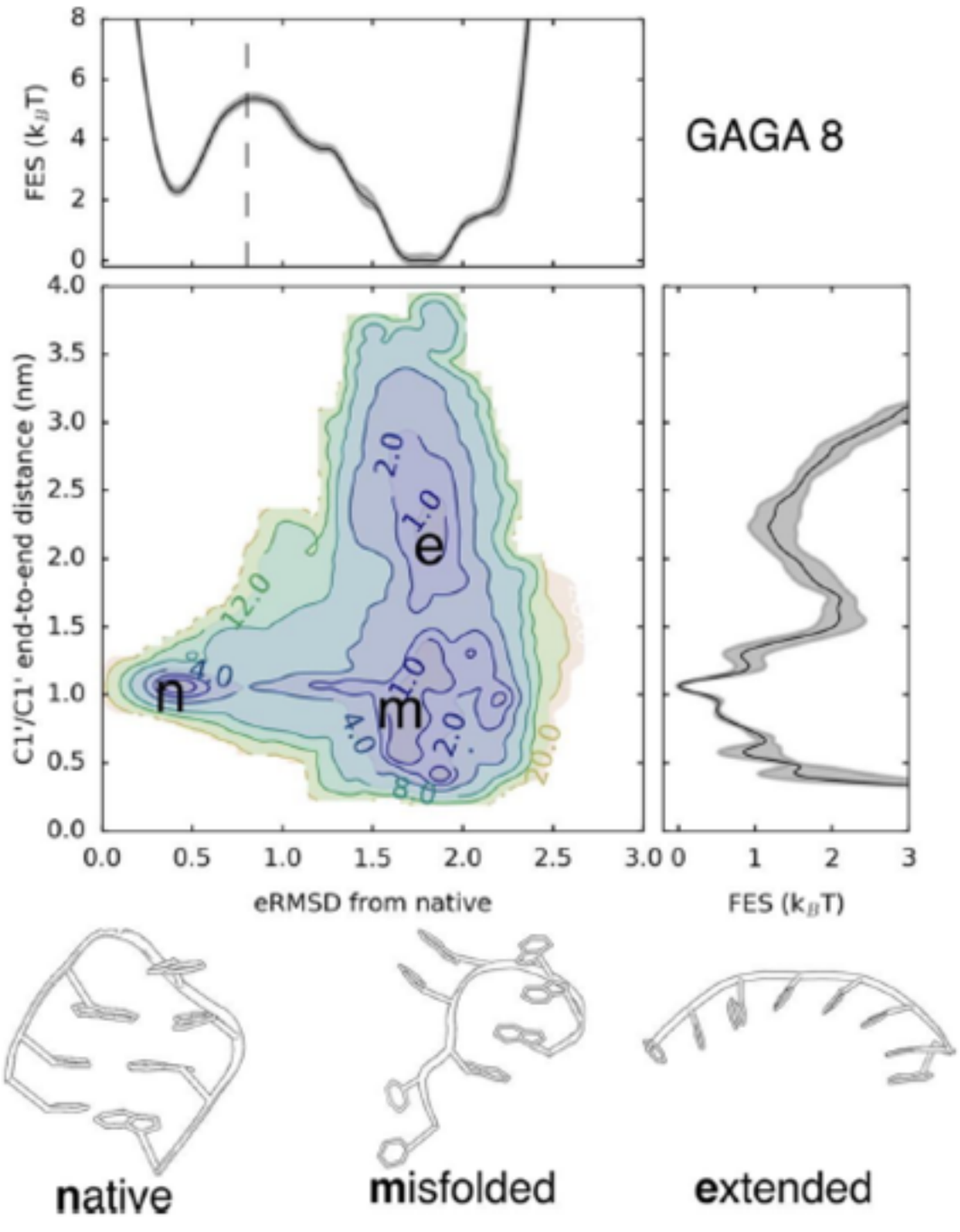
! Even though all water models can be used with AMBER (unlike CHARMM), simulation results are dependent on the model chosen.

AMBER

example : tetraloop folding

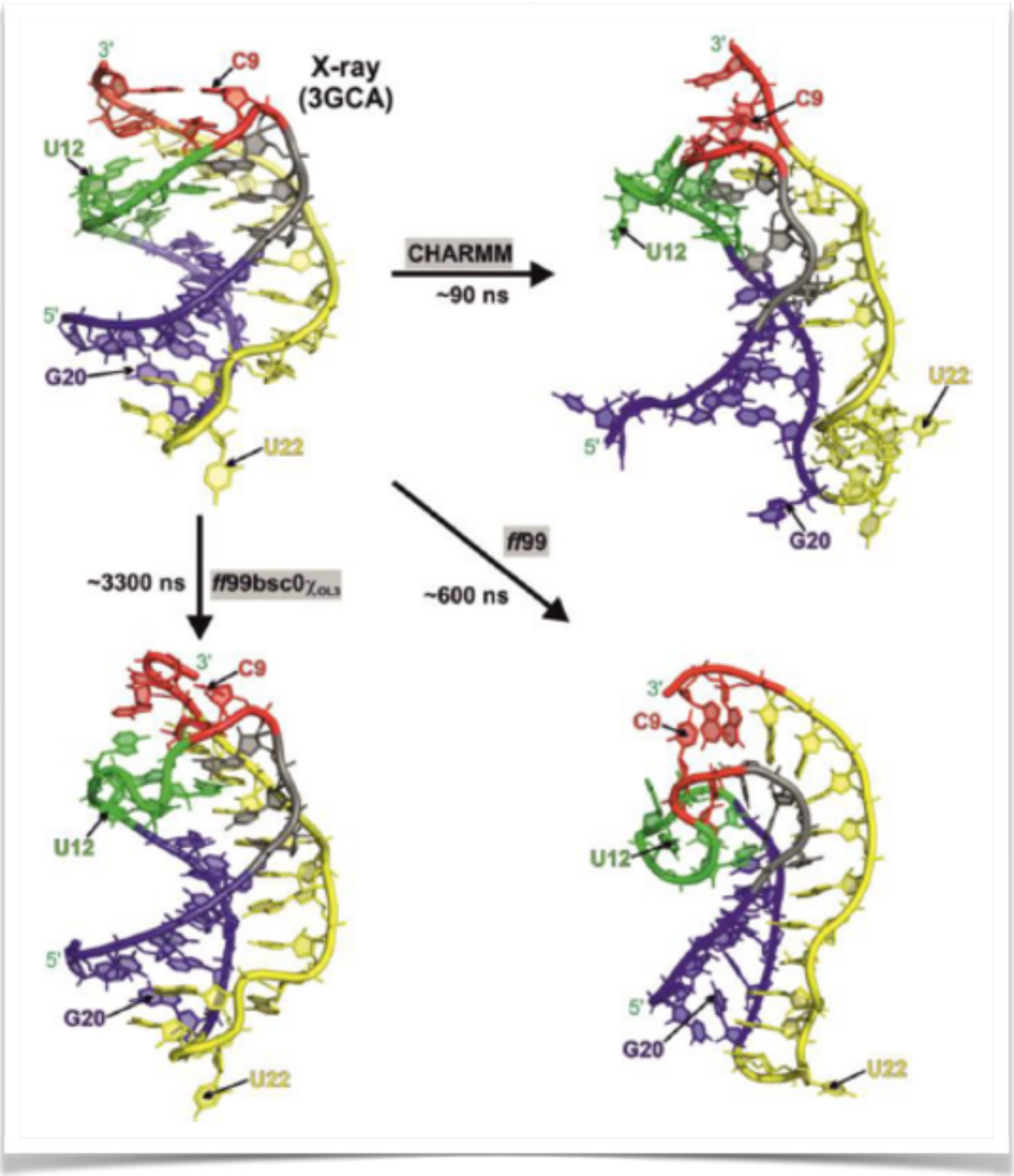


Garcia 2008: REMD of 226 ns per replica, using 52 replicas.
-> unbiased folding of all replicas starting from extended conformations.



Bussi 2016: metadynamics + REMD
-> native structure despite that it is not the most stable structure according to the ff

atomistic ff performance



- only 8-12 nt RNAs have been folded from random initial conformations (no matter the ff) and using enhanced sampling methods

- longer RNAs can be simulated from an experimental structure as long as large structural changes are not expected and/or being investigated.

- Due to the current difficulties in describing base-pairing energetics, atomistic simulations cannot be used to melt duplexes or observe conformational equilibria involving base-pairing rearrangements.

Current aa ff standing

Current RNA atomistic force fields are still inadequate

Differences in the ff arise from different optimization of torsions, but torsions have the LEAST physical basis of all the potentials in the ff.

-> torsions are 'the garbage' where we hide all the flaws in the ff

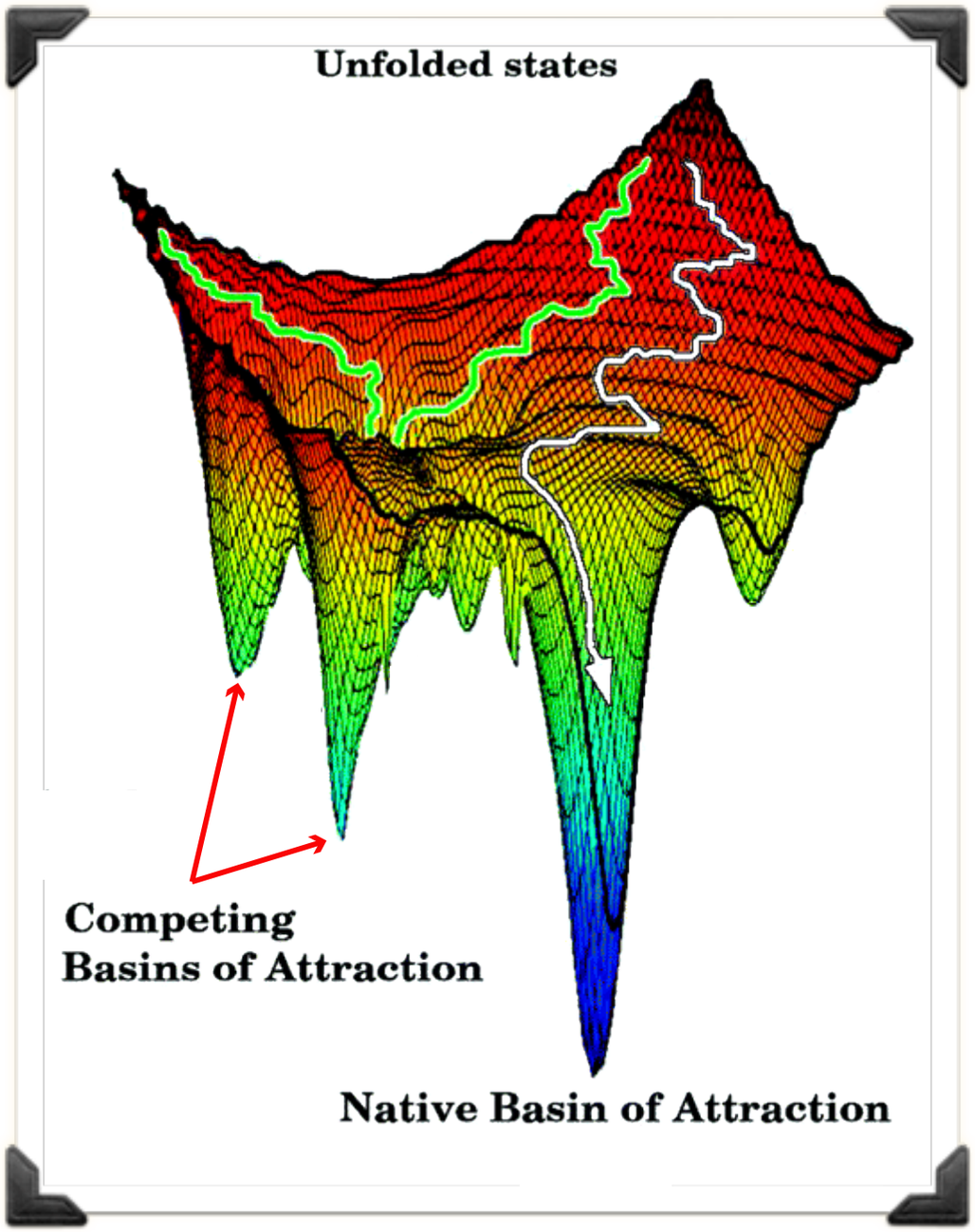
To assess the ff accuracy one simulates a variety of experimentally solved crystal structures for as long as is feasible to make sure that the molecule does not drift away.

-> RNA molecules are very flexible and can adopt various alternative conformations. Maintaining one single structure is not enough to ensure the correctness of the ff.

When QM is used for parametrization a transferability is assumed from properties in gas phase to properties in solution (not so obvious!)

When ff contain calibrations from solution-phase experiments using empirical approaches, these perform best under those same conditions for calibration (transferability issues)

MD limitations



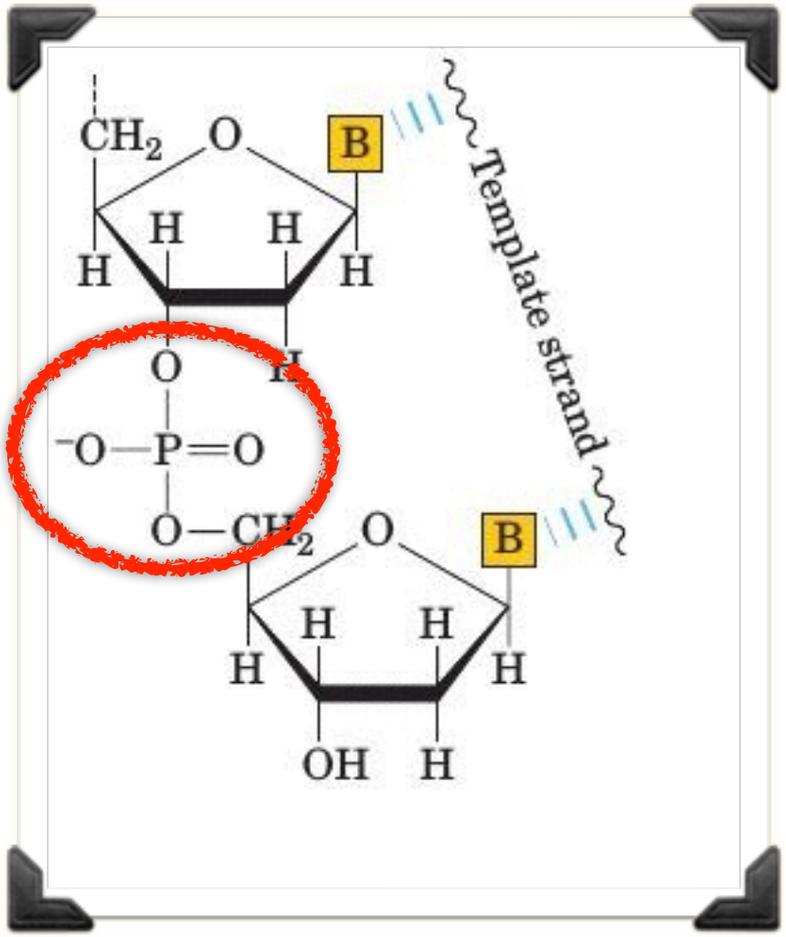
- At low temperatures the molecule gets trapped in local minima.
- At high temperature the molecule does not fold.

Configuration space not fully explored

Enhanced sampling solves part of the problem

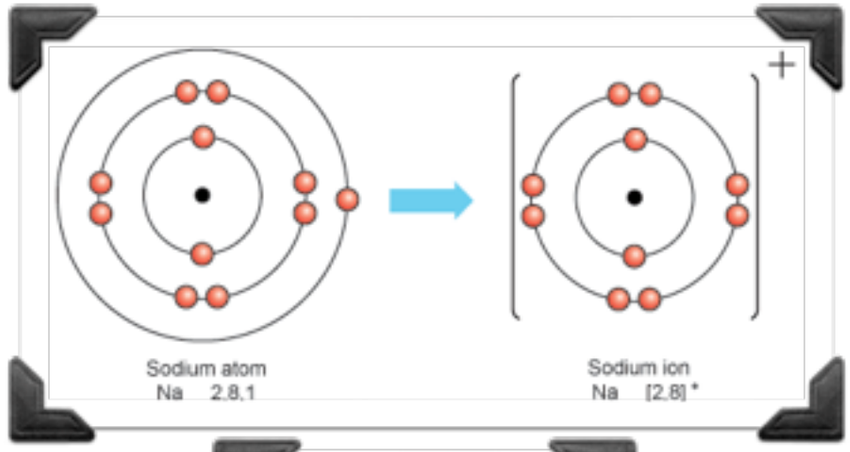
Ions - nucleic acids weak spot

At physiological conditions nucleic acids carry one -1 charge on each phosphate group

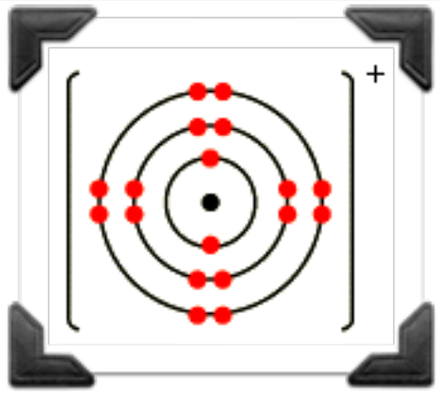


RNA and DNA are surrounded by positive metal ions!

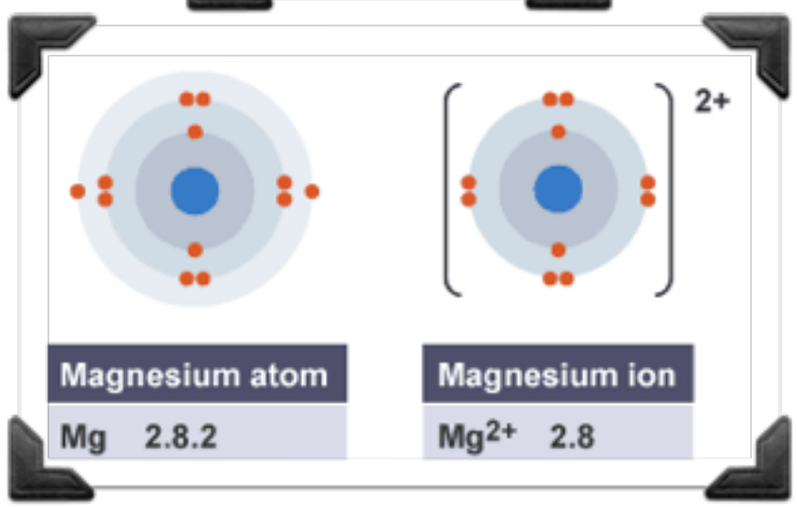
Na^+



K^+

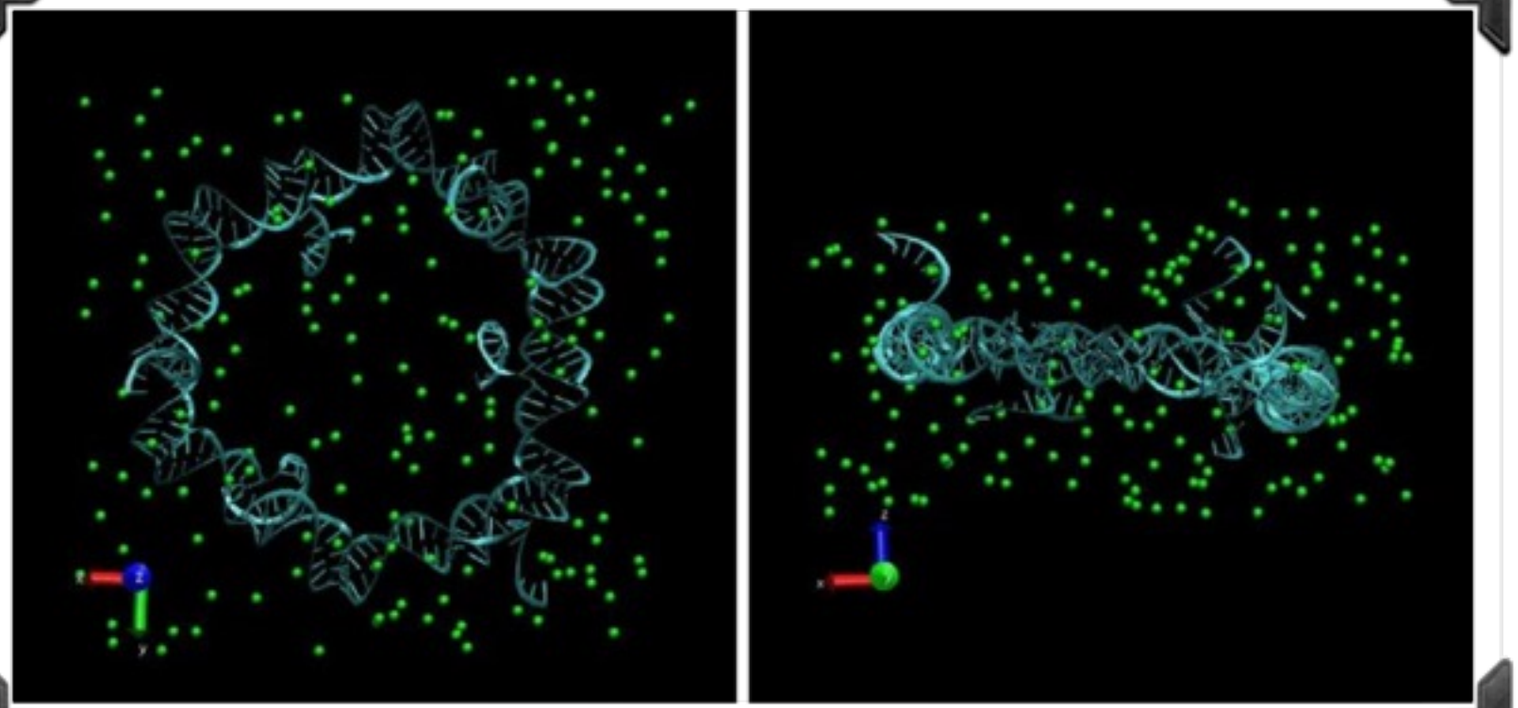


Mg^{+2}

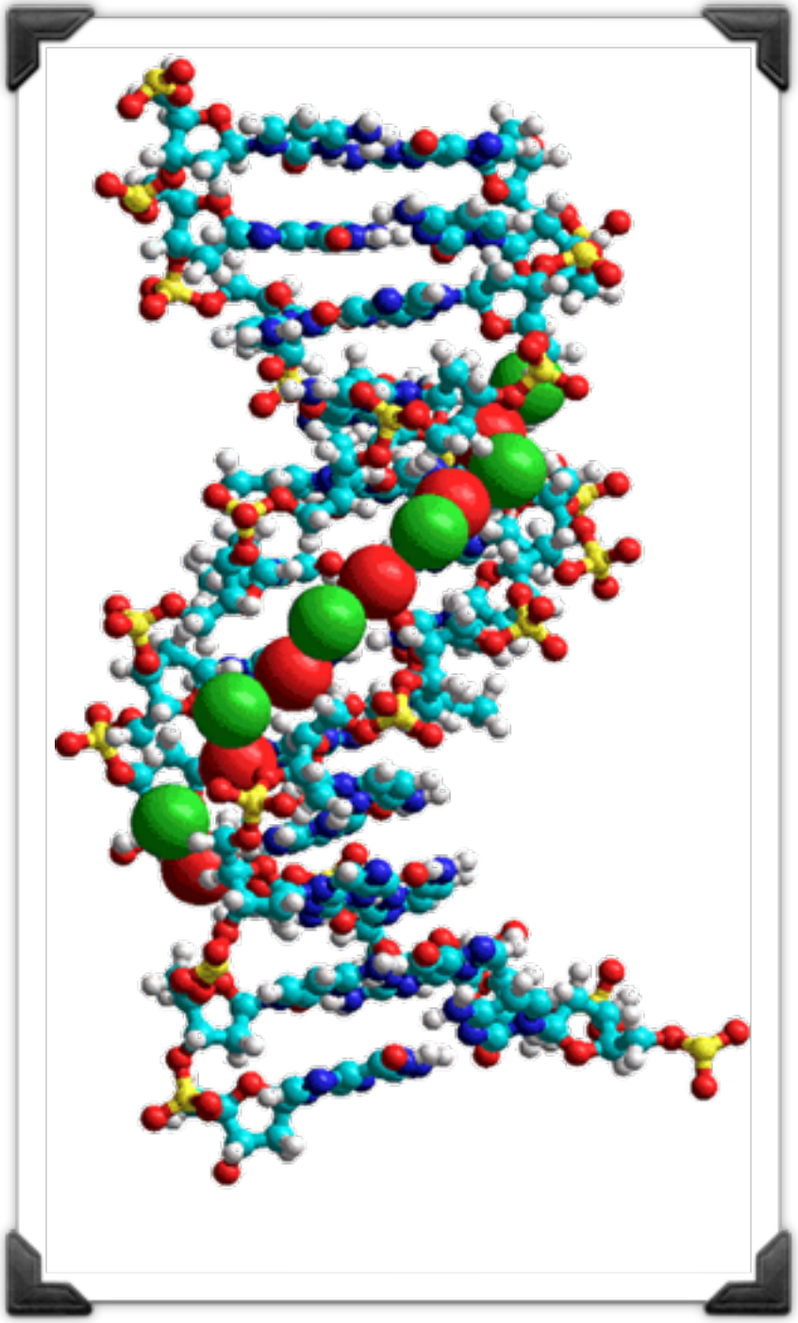


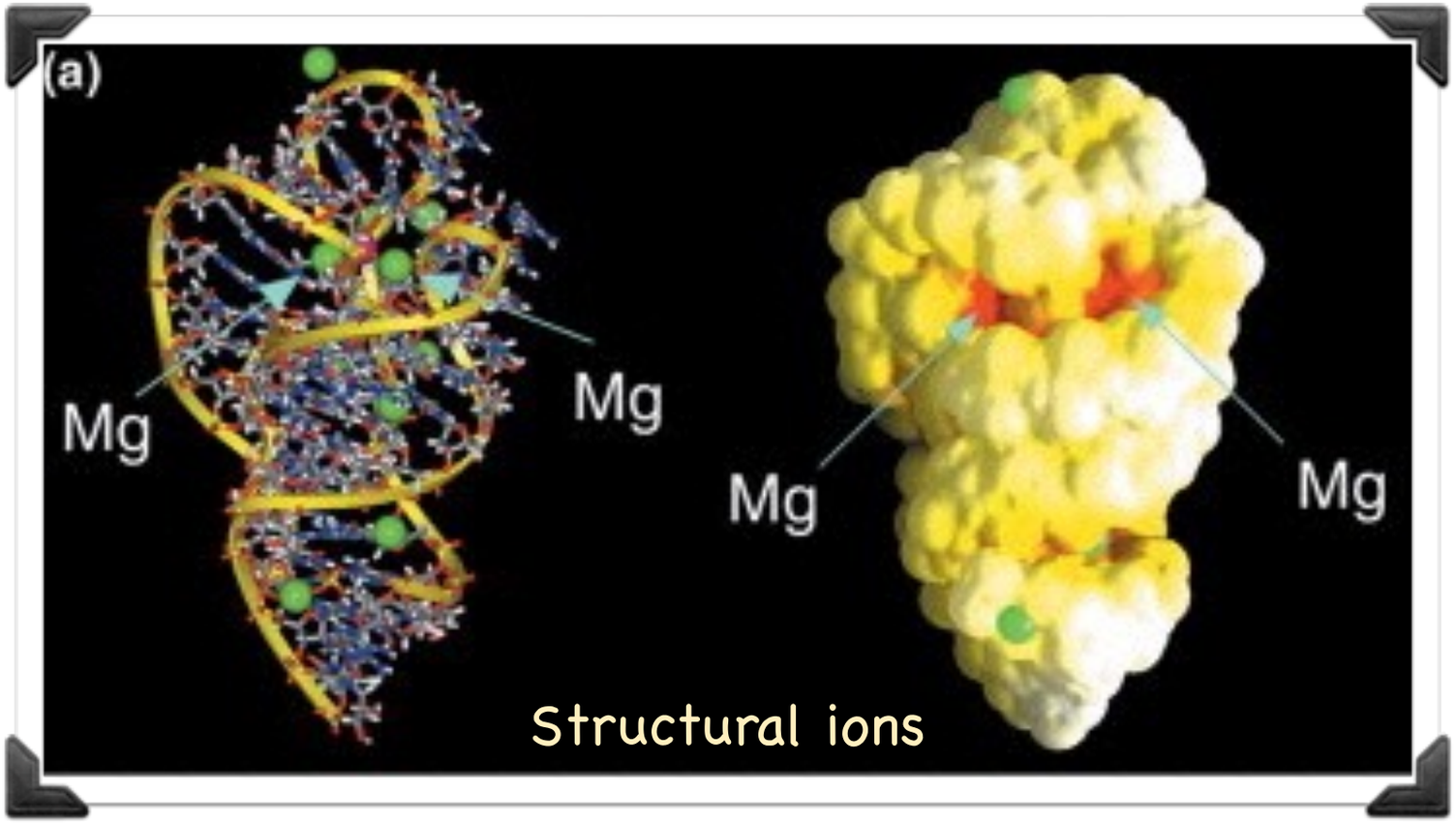
Ions

Counter-ions (bulk)

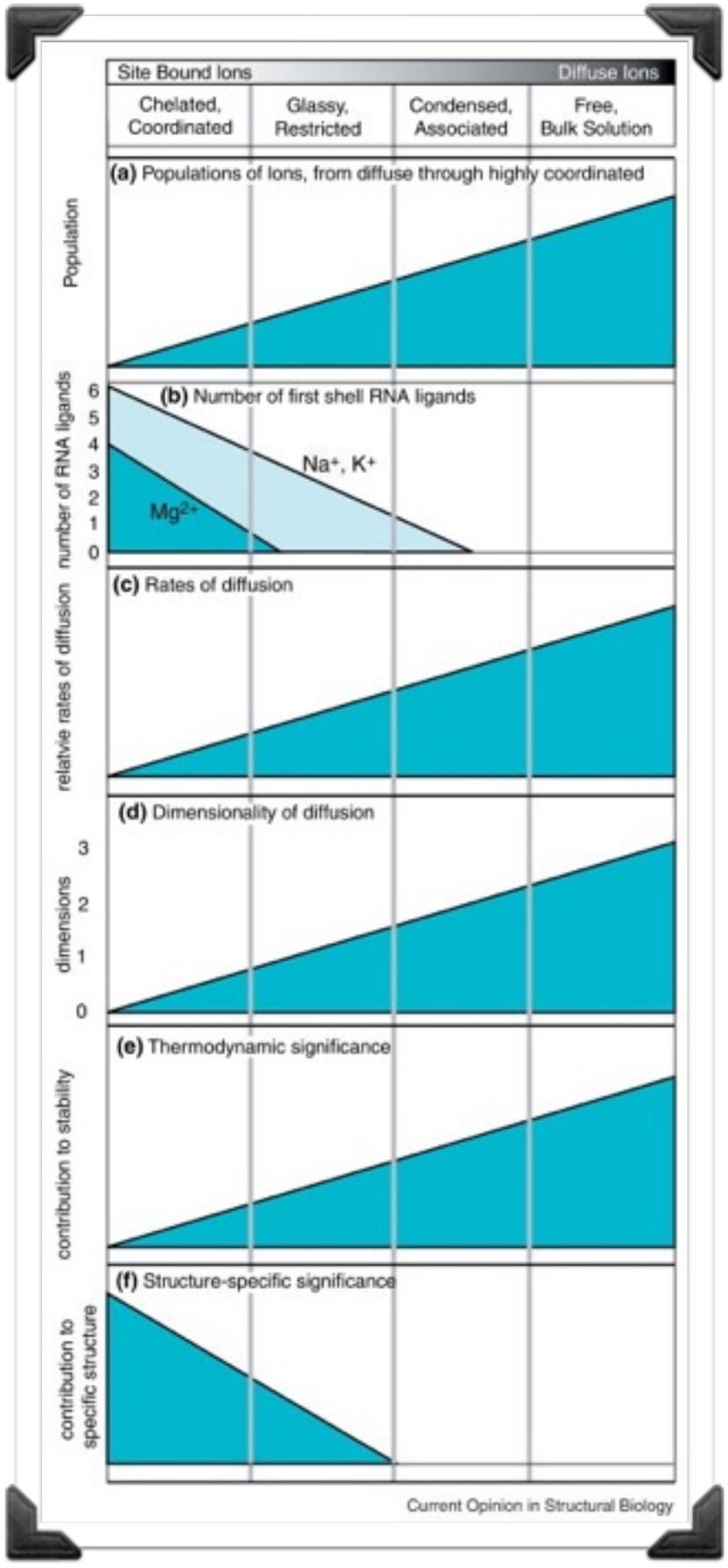


Surface ions (ionic cloud)





Ion classification



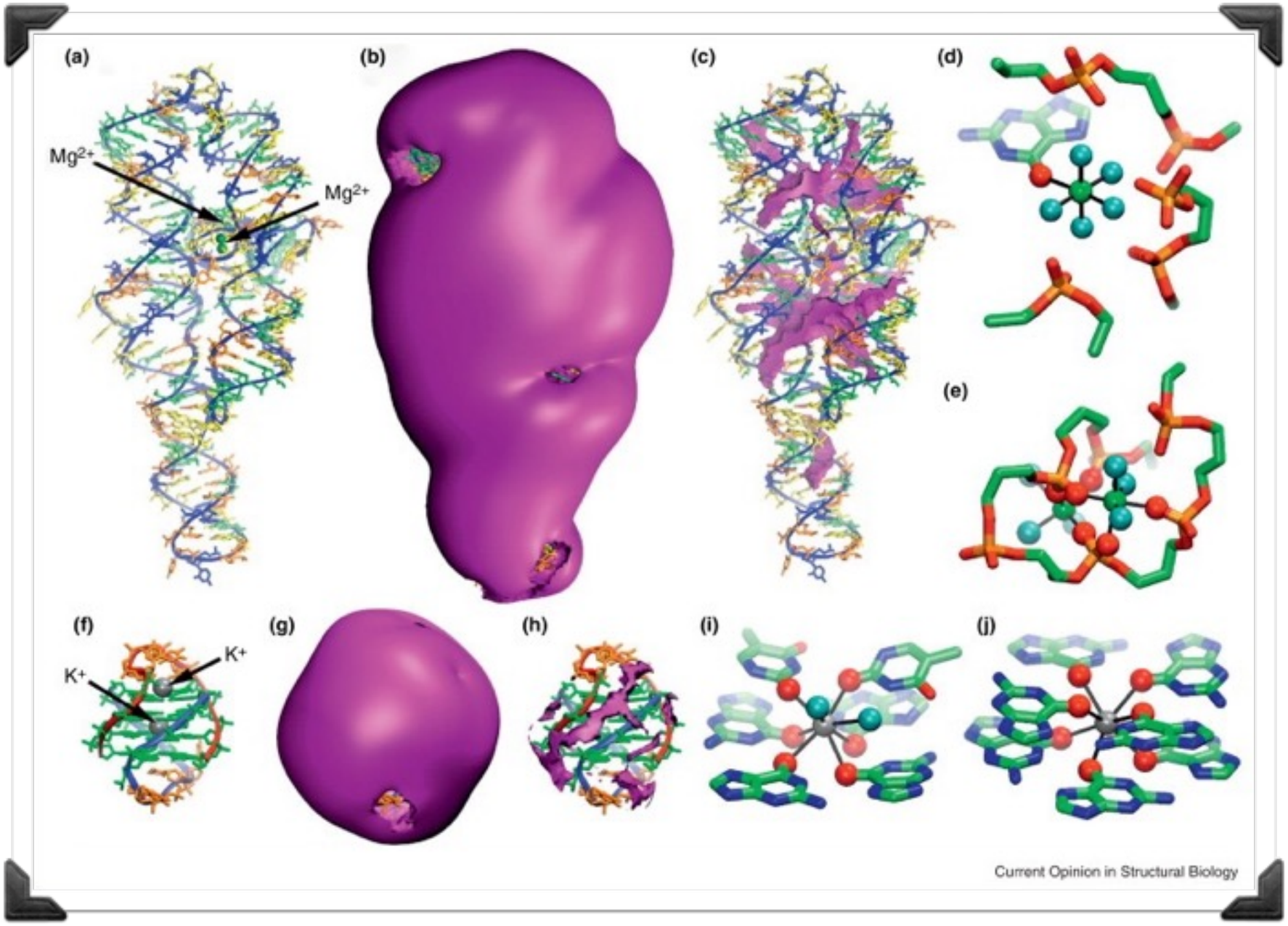
Chelated: directly linked to RNA through coordination, immobile (ion clamps)

Glassy: closely associated with DNA and RNA, restricted mobility (i.e. ions in grooves)

Condensed: mobile ions in close distance from DNA or RNA, they explore the whole NA surface

Bulk: completely free ions, they explore the whole space, they provide the neutrality of the solution

Ion classification



a) structure of P4-P6 domain of group 1 intron RNA

b) envelope of condensed cations

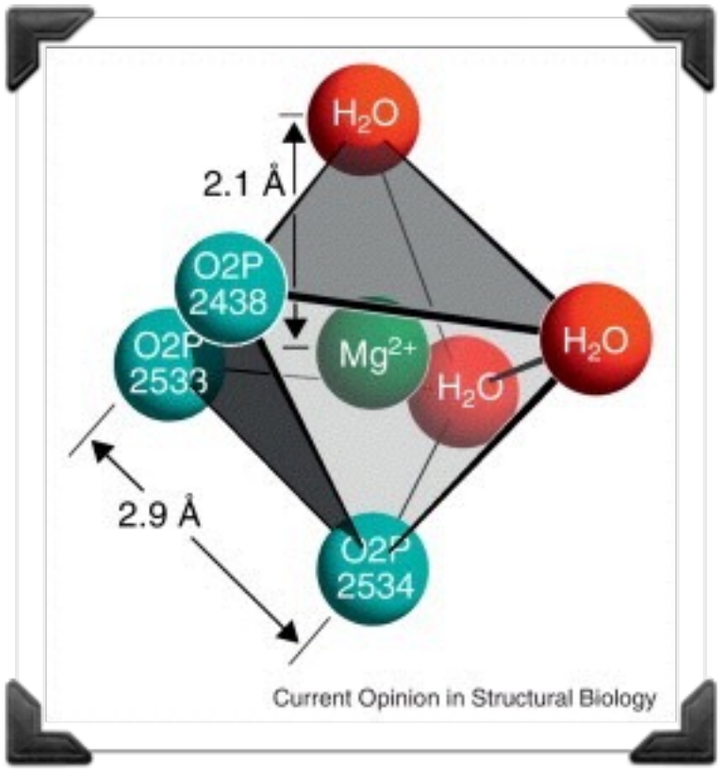
c) glassy cations in the grooves

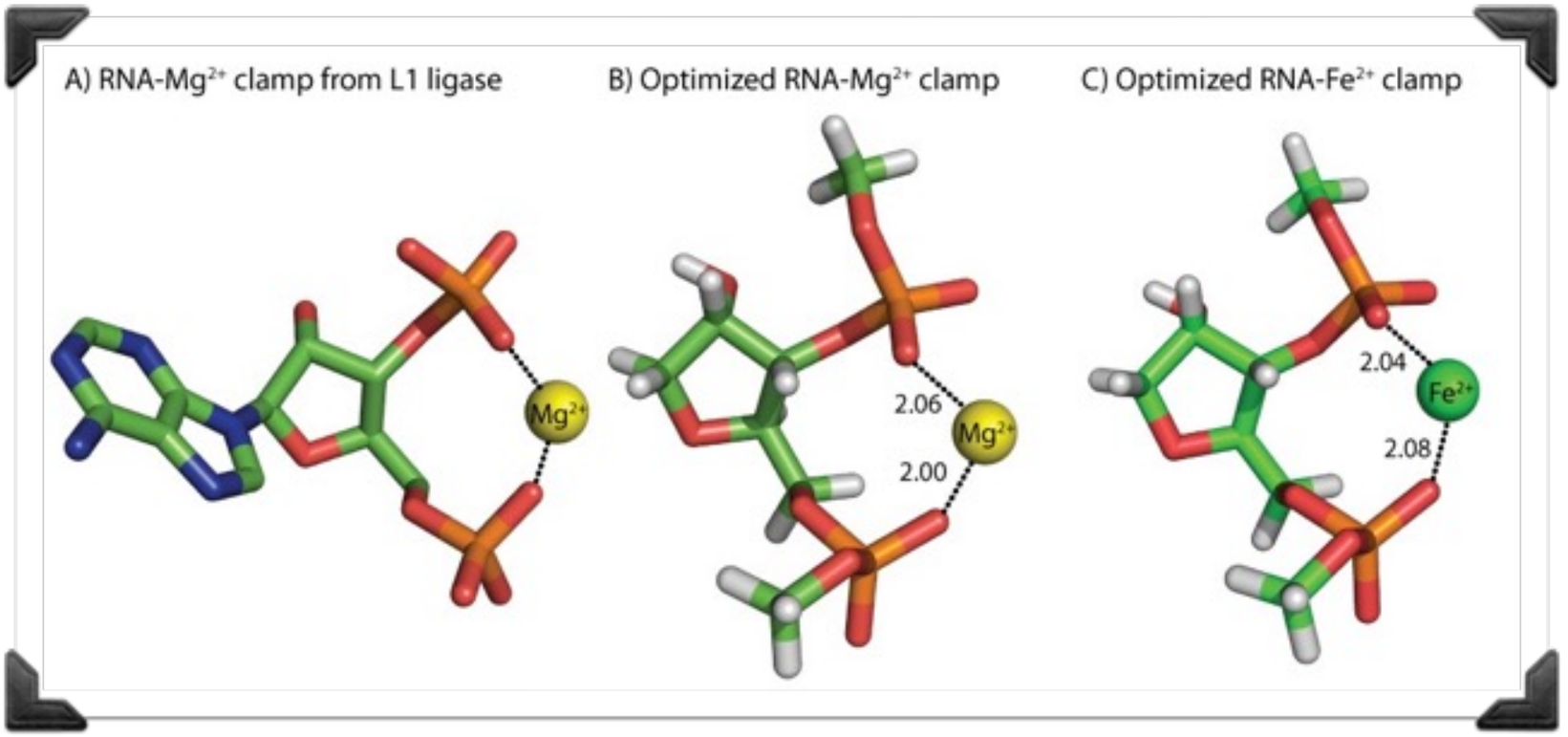
d) coordination of glassy Mg^{2+} ions

e) highly coordinated structural Mg^{2+}

Chelated ions

Structural ions are an integral part of the structure.
They coordinate with surface water molecule and with phosphate groups
to promote and stabilize RNA architectures





electrostatic as well as non-Coulombic interactions

polarization
charge-transfer
correlation

Modeling ions

Neutralizing the negative charge of the nucleic acids backbone

explicit ion simulations | implicit representation

Monovalent ions : Na^+, K^+

Coulomb + LJ | Debye-Huckle

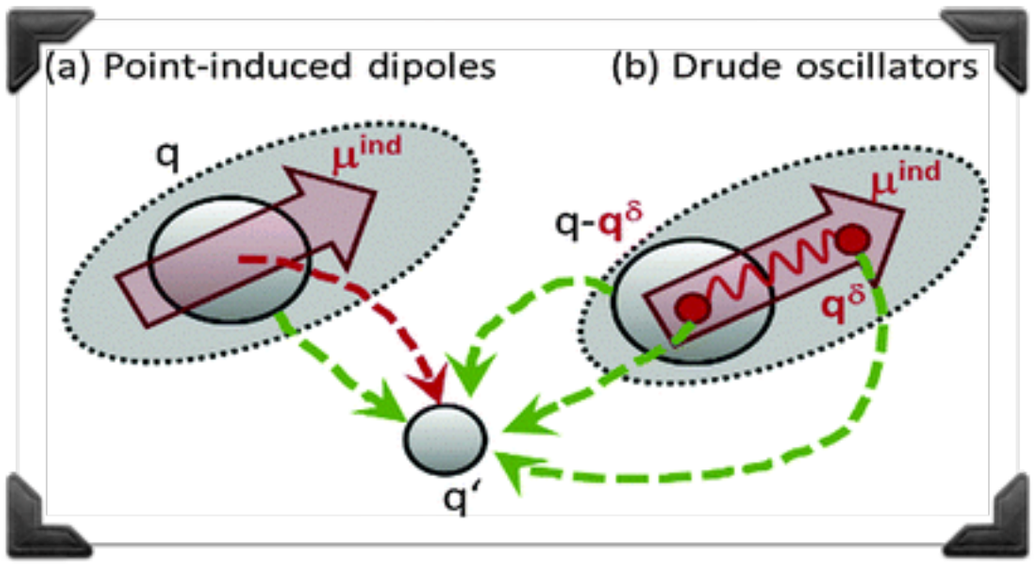
Divalent Mg^{2+}

explicit water, explicit ion simulations :
very slow water dynamics
slow ionic diffusion -> poor statistics

implicit representation:
how to account for coordination?

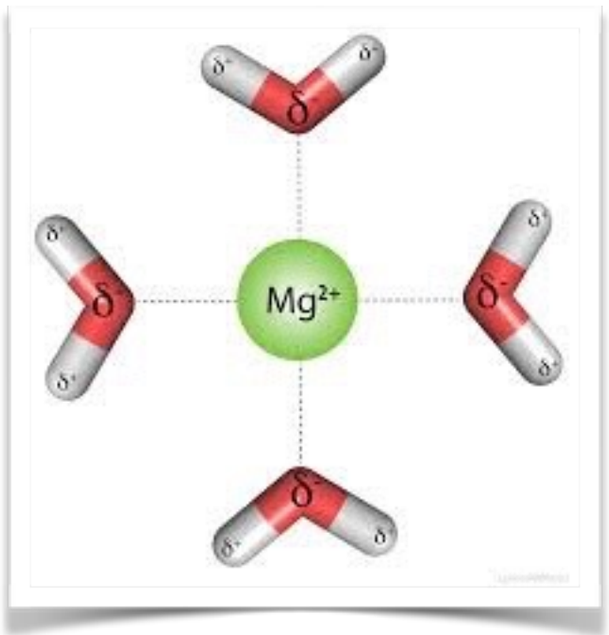
Polarization

Atomistic polarizable force-fields (Drude model)



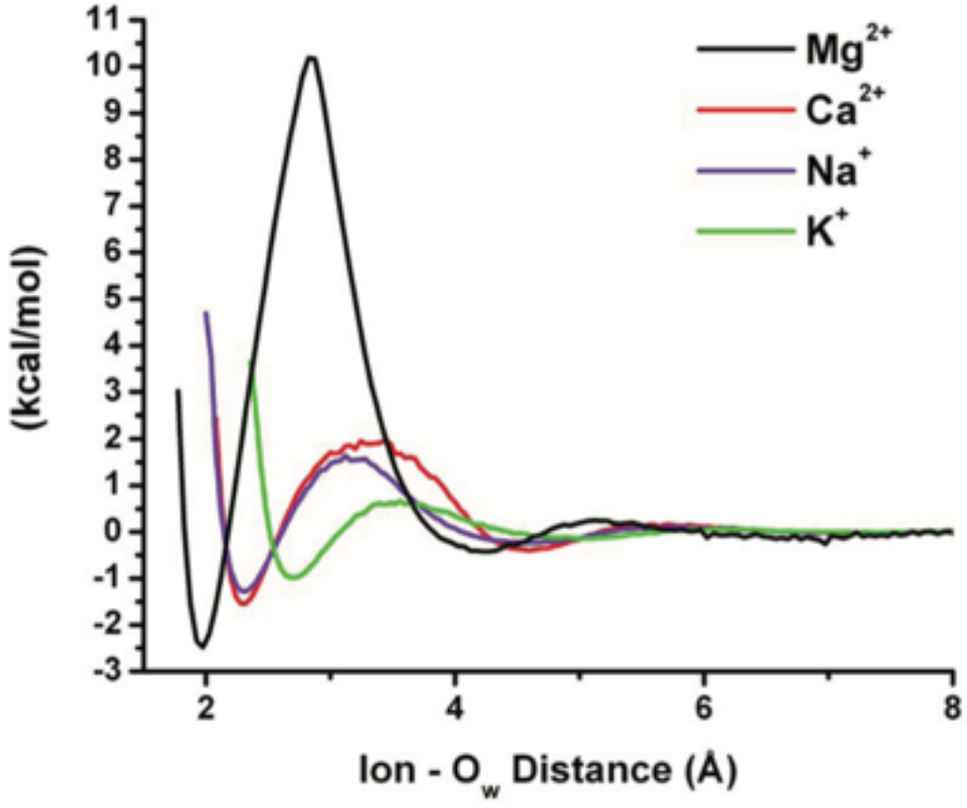
Why ions pose problems

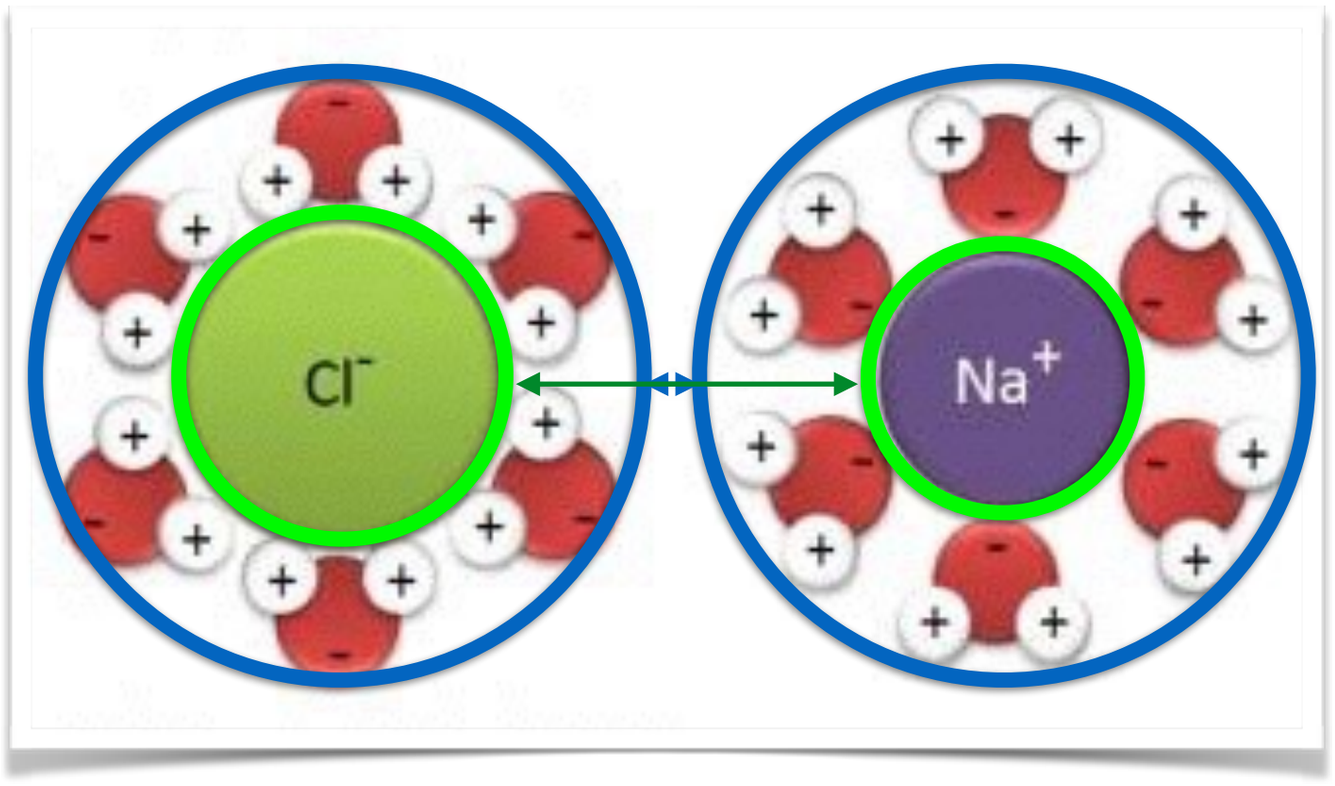
very delicate equilibrium with the environment



Ions (Mg+2 in particular) coordinate strongly with water.

Difficult to define LJ parameters for the interactions with ions and the other particles because the size changes if water is present or not.

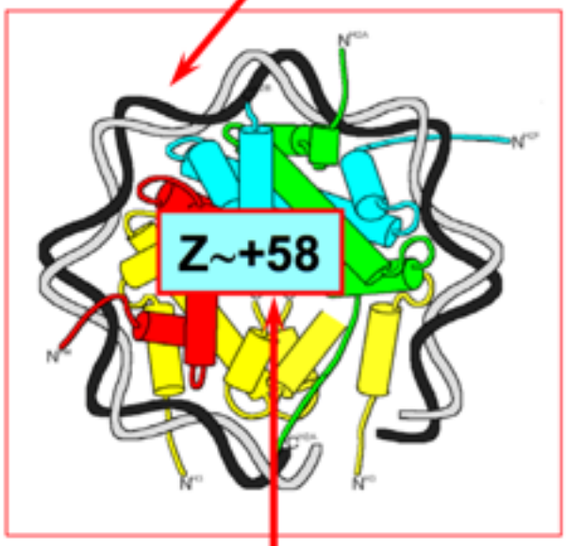


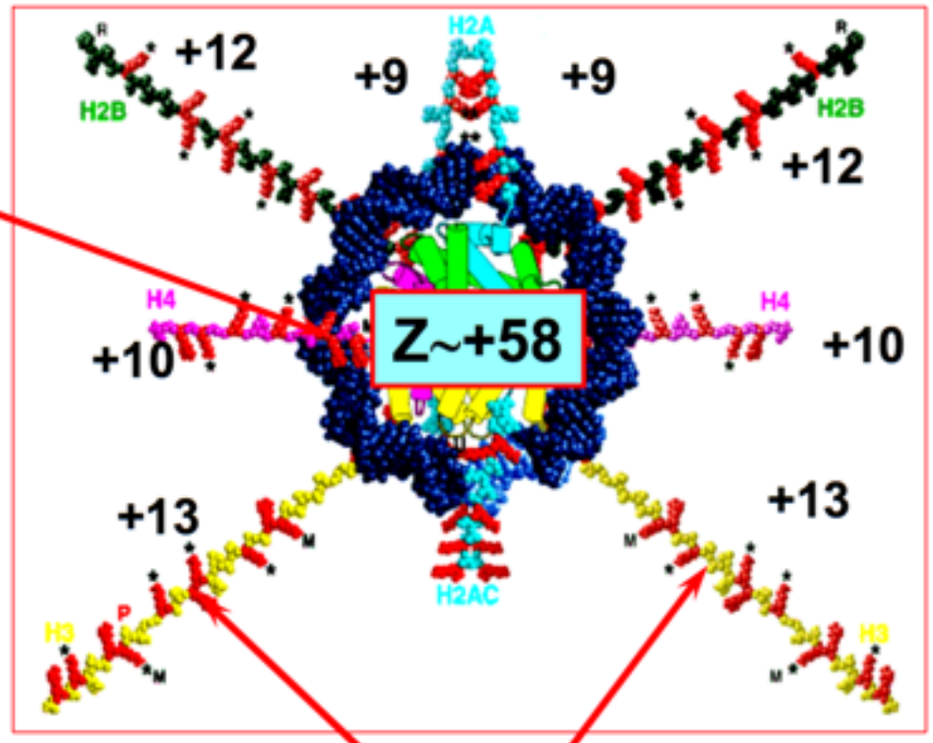


Example : histones

The Nucleosome Core Particle (NCP):

DNA, ~147 bp
Z = –294

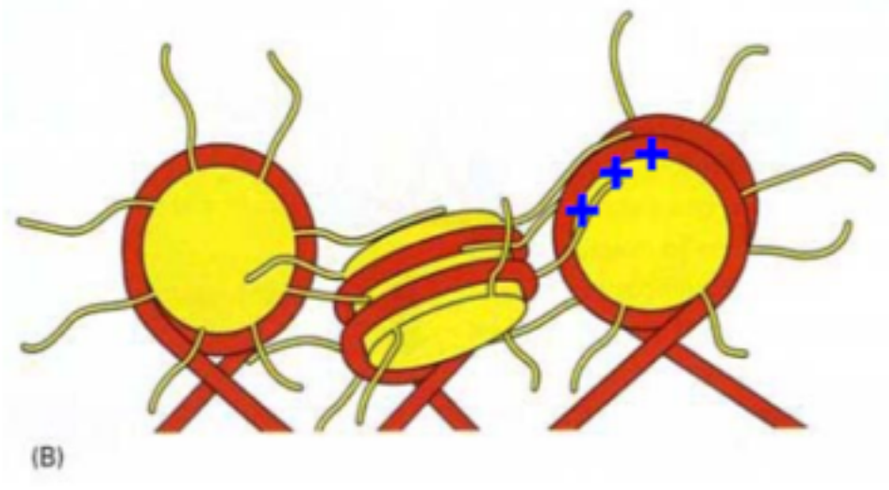




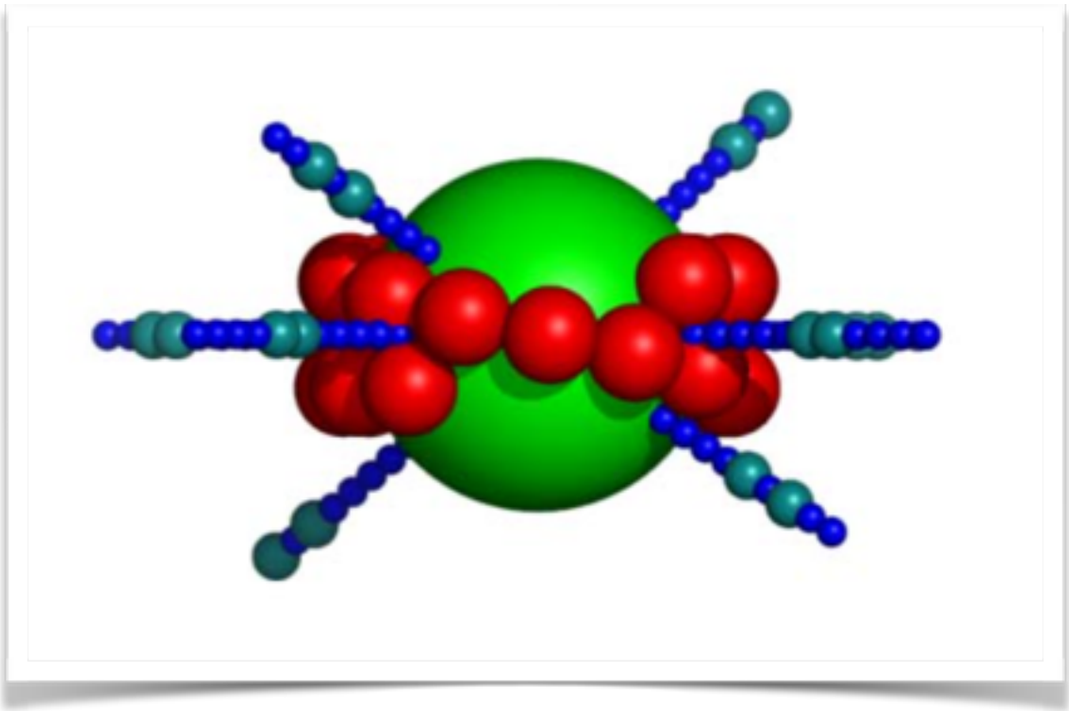
NCP Total charge: ~ -148

Histone protein octamer core:
2xH2A, 2xH2B, 2xH3, 2xH4

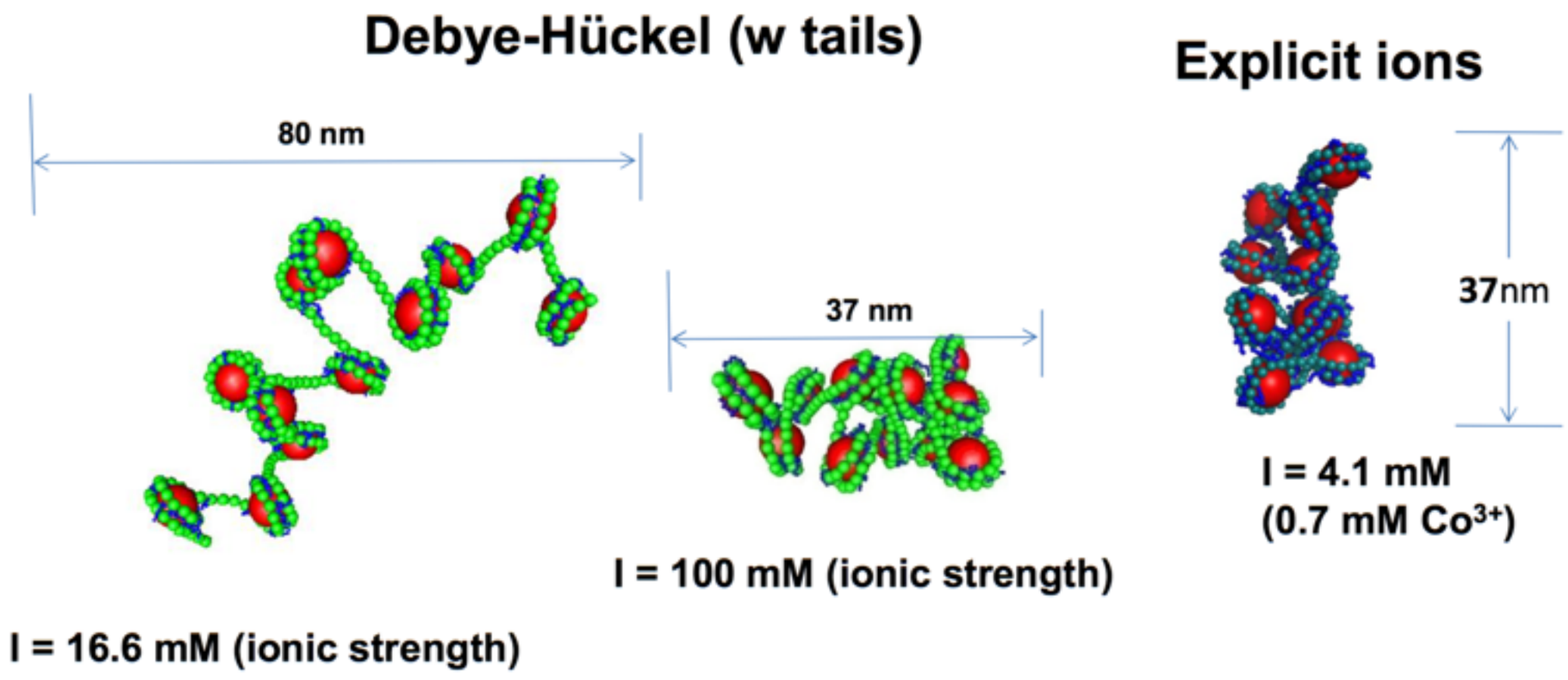
N-Histone tails Z= ~ +88



Example : histones



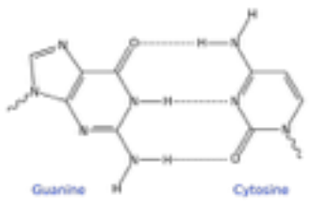
Nucleosome Core Particle (NCP) model
(L. Nordenskiold)

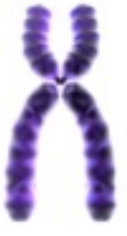


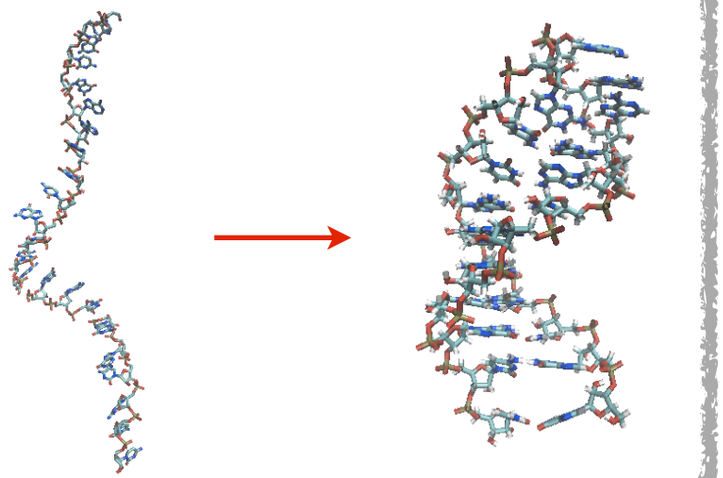
Folding problem

Prediction of structure, dynamical and thermodynamical behavior in 3D

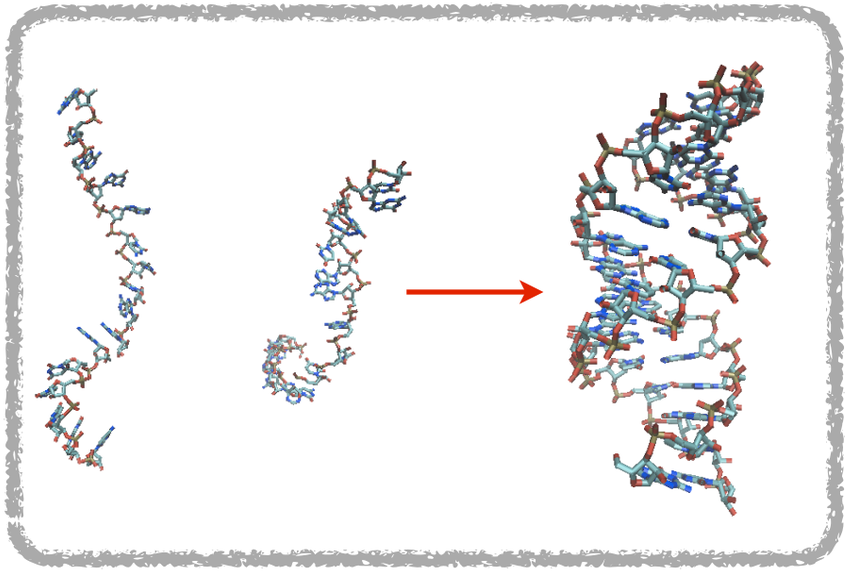
Å nm 10 nm 100 nm μm







Folding



Assembly

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

Accuracy and speed

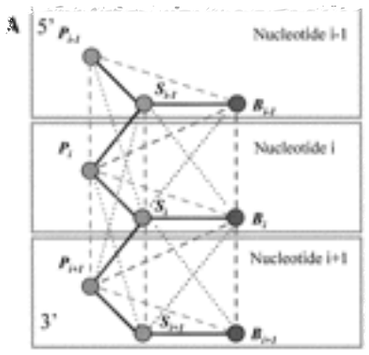
Speed → Accuracy

Jonikas, 2009

NAST

Biased potentials

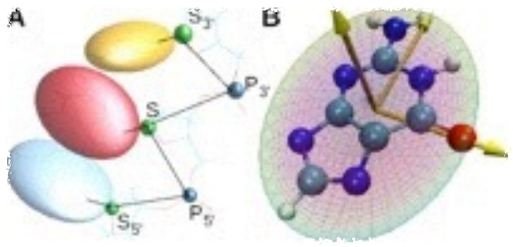
Dokholyan, 2008



iFoldRNA

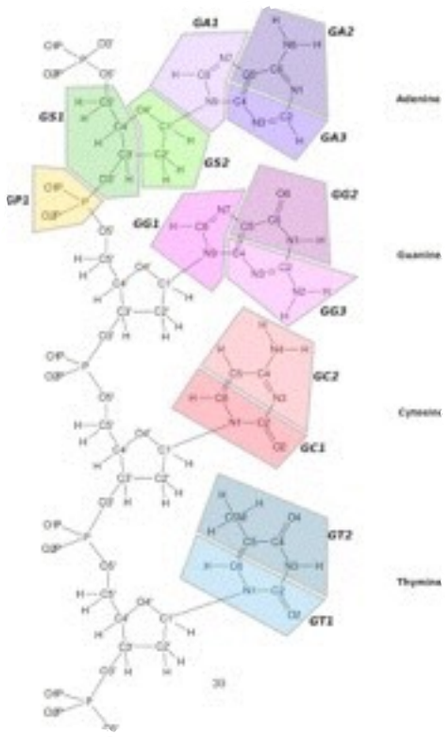
Geometric constraints

Plotkin, 2008



bottom-up potentials

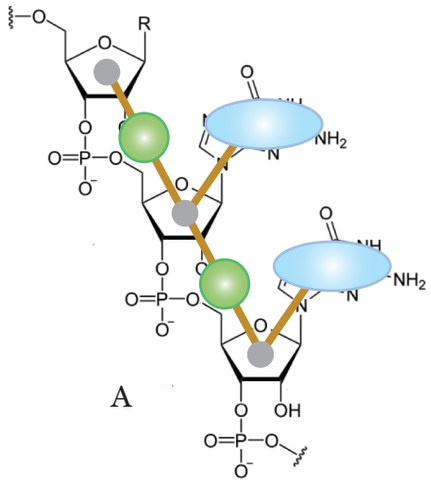
Poulin, 2008

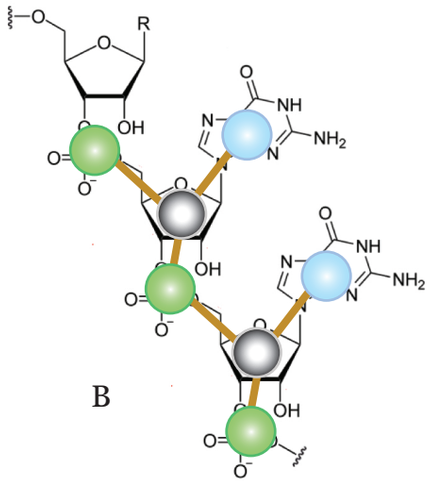


Rigid body

What model?

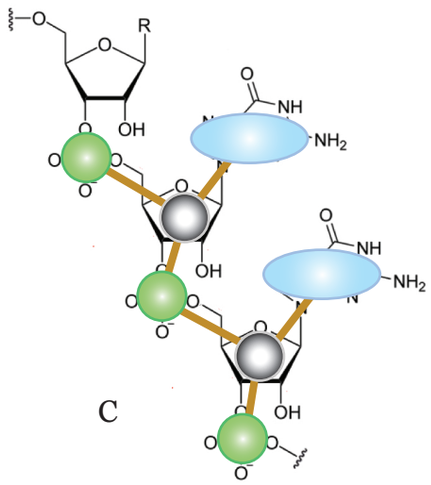
NARES-2P
two interacting particles +
virtual particle for geometry
dipolar bases

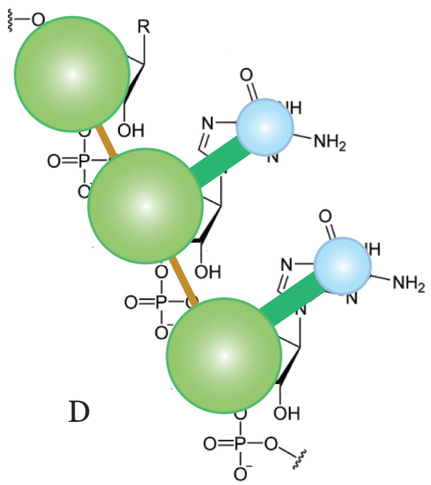




TIS
Three-interacting-particles
at the center of mass of each
group

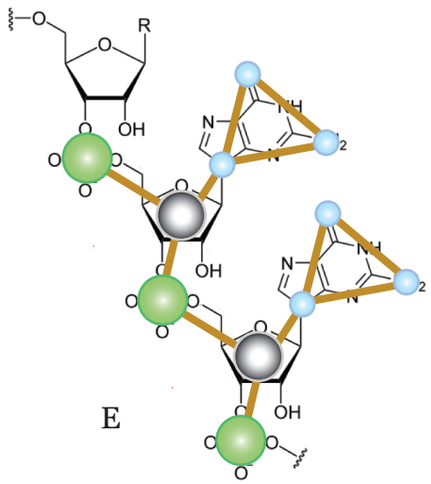
Plotkin
Three interacting particles
Elliptical base

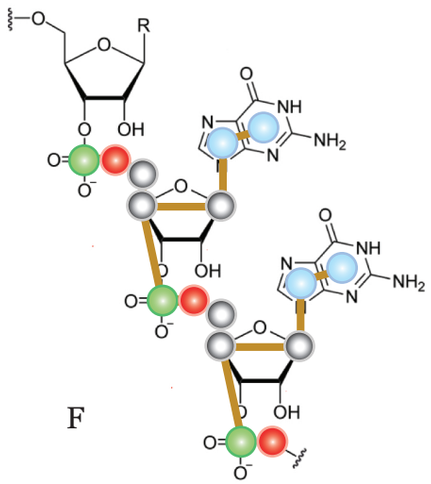




OxDNA/RNA
Rigid-body composed of 5
interaction centers

Xia
5 interacting particles, 3 of
which to describe the base



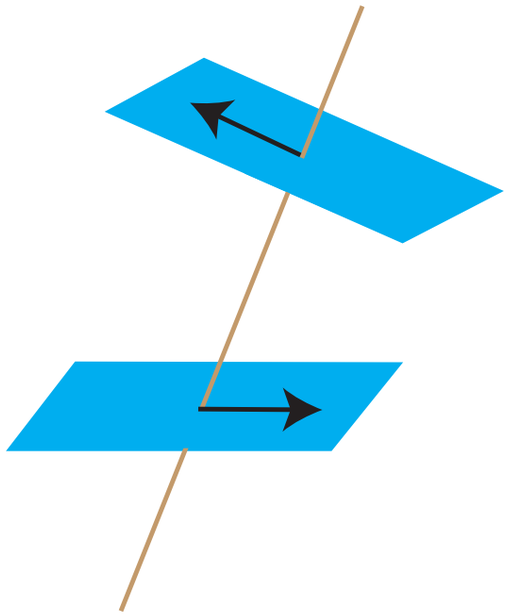


HiRE-RNA/DNA
6 or 7 interacting particles

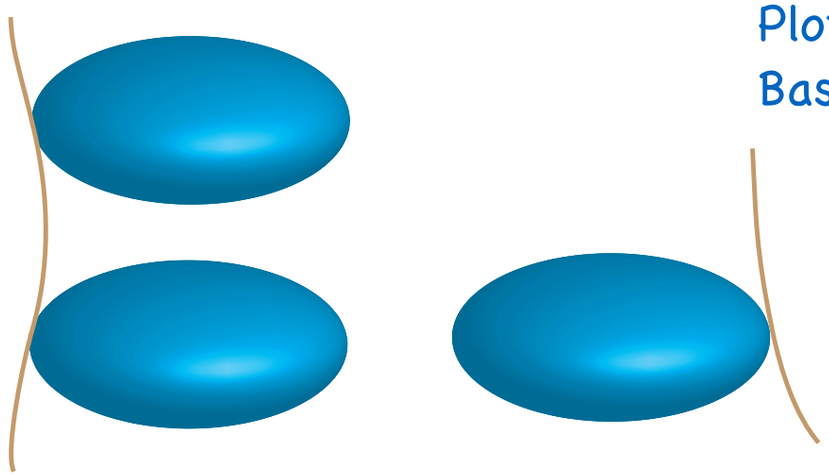
Flat or round?

A

NARES-2P
Bases as dipoles

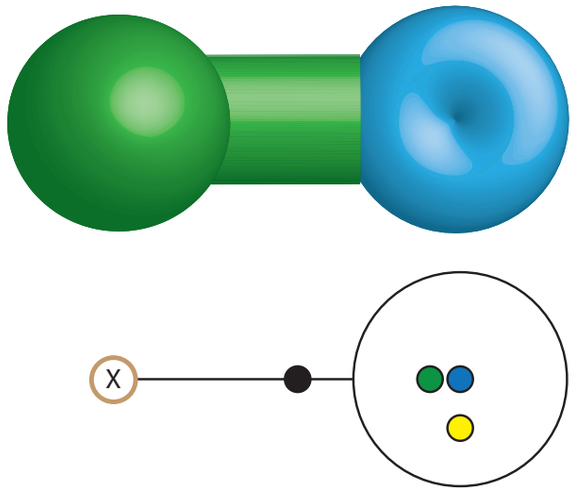


B



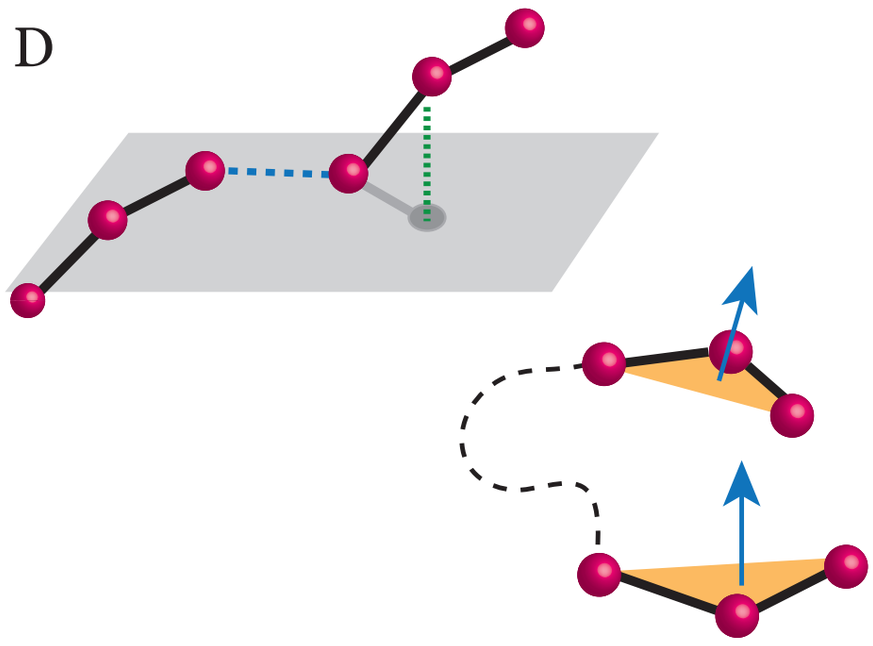
Plotkin
Bases as ellipsoids

C



OxDNA
of-centered interaction sites

D

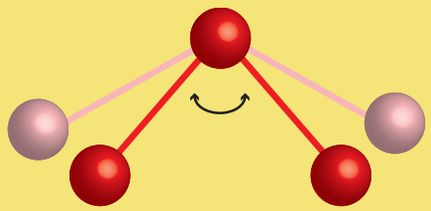


HiRE-RNA
base-plane from 3 particles

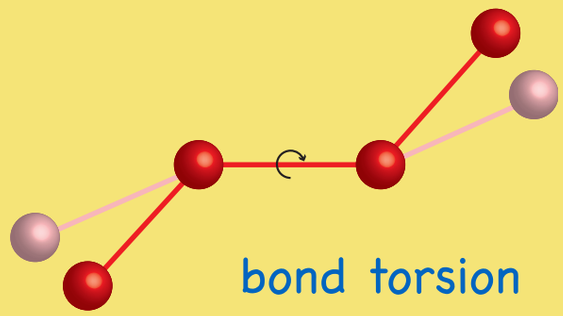
Force field?

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

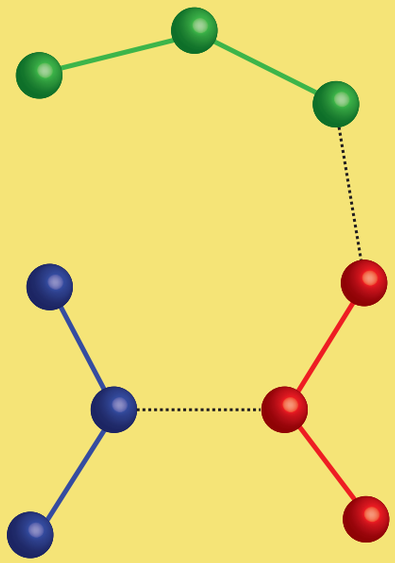
bond stretching



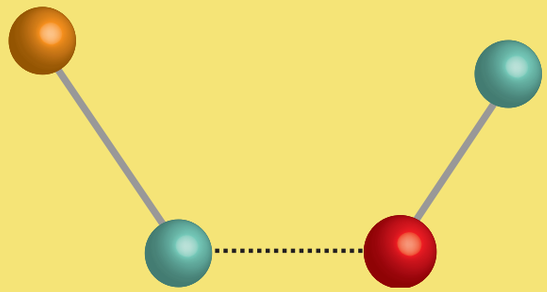
angle bending



bond torsion



non-bonded interaction
(Van der Waals)



hydrogen bond

electrostatic
interaction

geometric parameters
-> statistical analysis NDB

energetic parameters
-> experimental data, QM calculations, ?

Choosing functional forms

Breakable bonds : Morse potential

$$V(x) = \epsilon \left(1 - e^{-a(x-x_0)}\right)^2$$

Long-range : Lennard Jones potential

$$V(x) = \epsilon \left[\left(\frac{\sigma}{x}\right)^{12} - 2\left(\frac{\sigma}{x}\right)^{6}\right]$$

spherical particles

$$V(r, \theta, \phi) = V_1(r) \times V_2(\theta, \phi)$$

Pure Coulomb

$$V(x) = \frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon_r}\frac{1}{x}$$

Screened ionic potential: Debye-Huckel

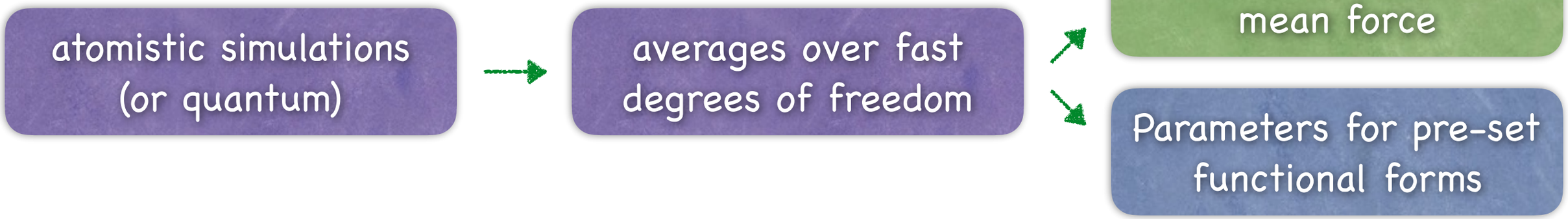
$$V(x) = \frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon_r}\frac{e^{-\frac{x}{\lambda}}}{x}$$

elliptical (anisotropic) particles: Guy-Berne potential

$$V(\vec{u}_i, \vec{u}_j, \vec{r}) = 4\epsilon(\vec{u}_i, \vec{u}_j, \vec{s}) \left\{ \left[\frac{\sigma_0}{4 - \sigma(\vec{u}_i, \vec{u}_j, \vec{s}) + \sigma_0}\right]^{12} - \left[\frac{\sigma_0}{4 - \sigma(\vec{u}_i, \vec{u}_j, \vec{s}) + \sigma_0}\right]^{6} \right\}$$

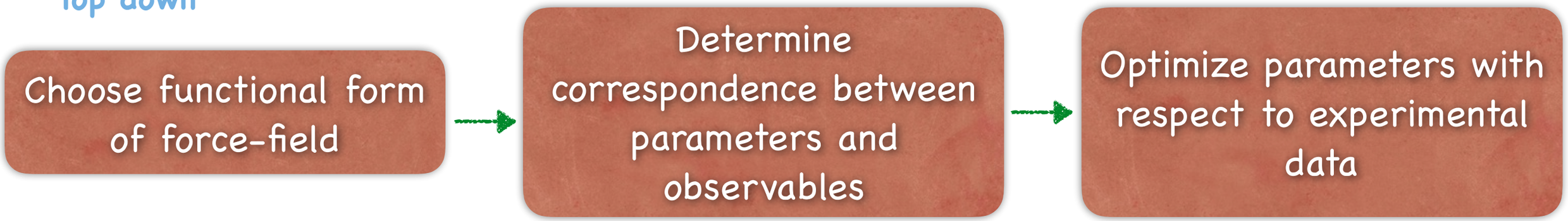
Up or down?

"Bottom up"



Rely on the validity of the underlying model.
Ok for proteins, but still problematic for nucleic acids.

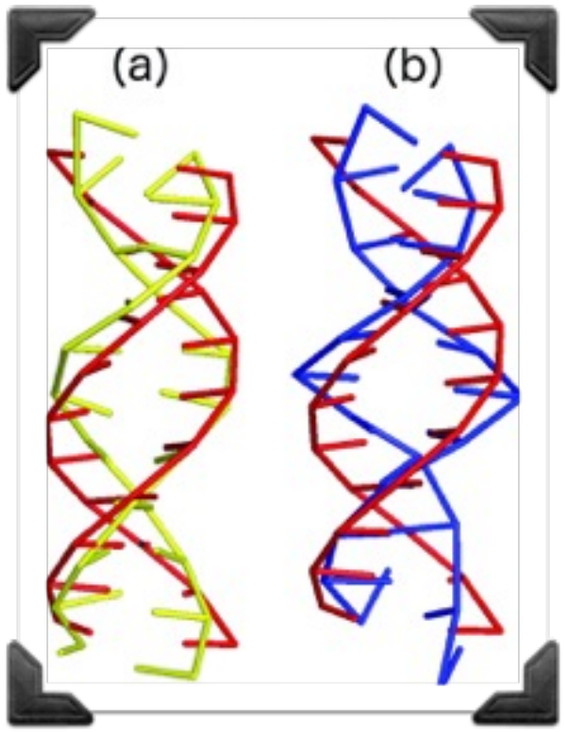
"Top down"

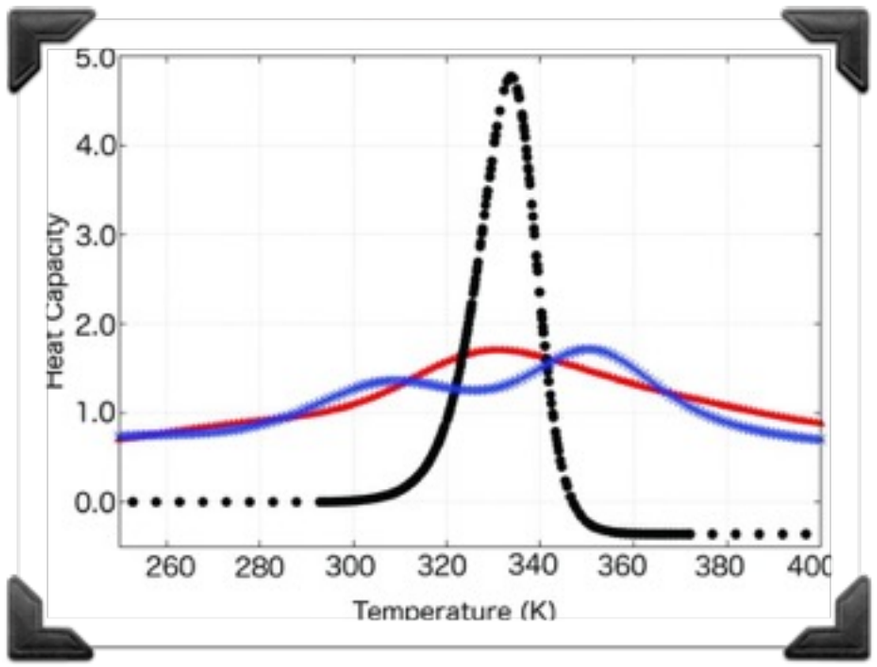


Difficulty of extracting suitable information from experimental data.

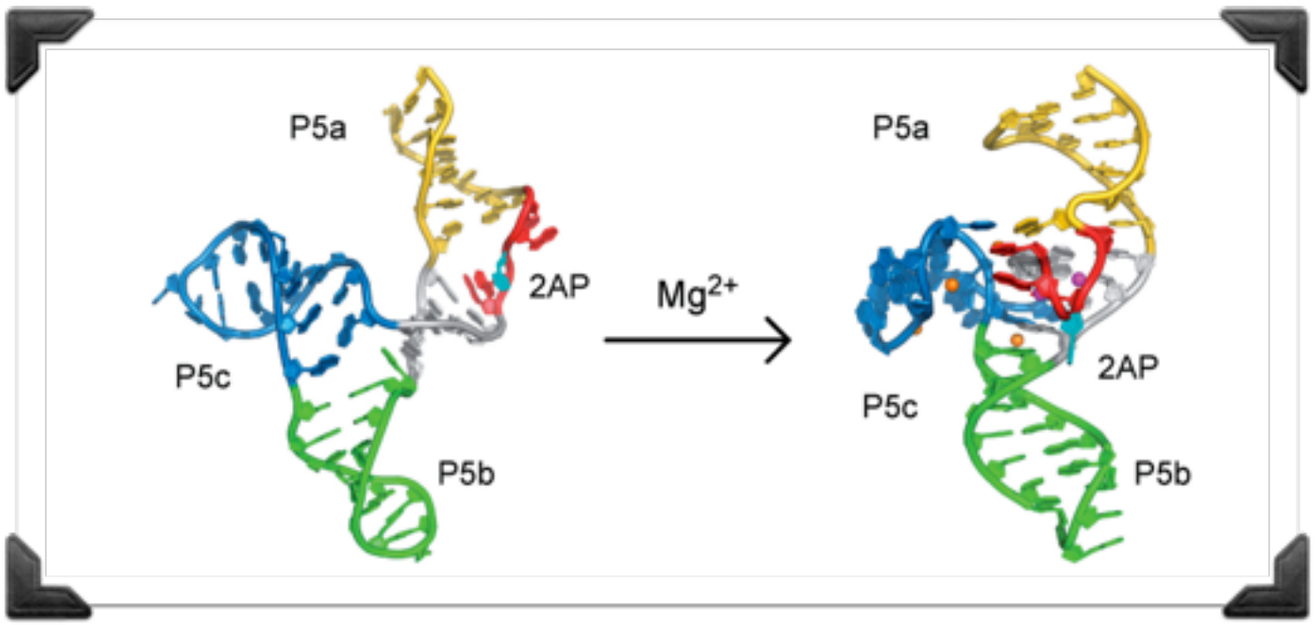
What for?

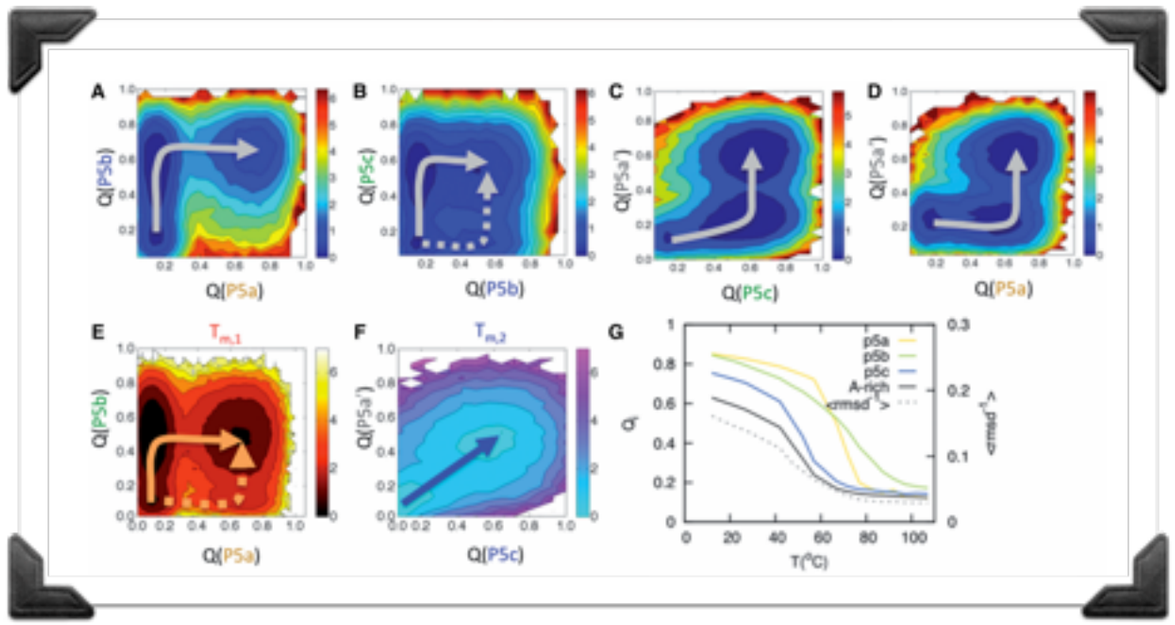
NARES-2P (Sheraga 2015) : DNA melting, proof of principle on dipoles





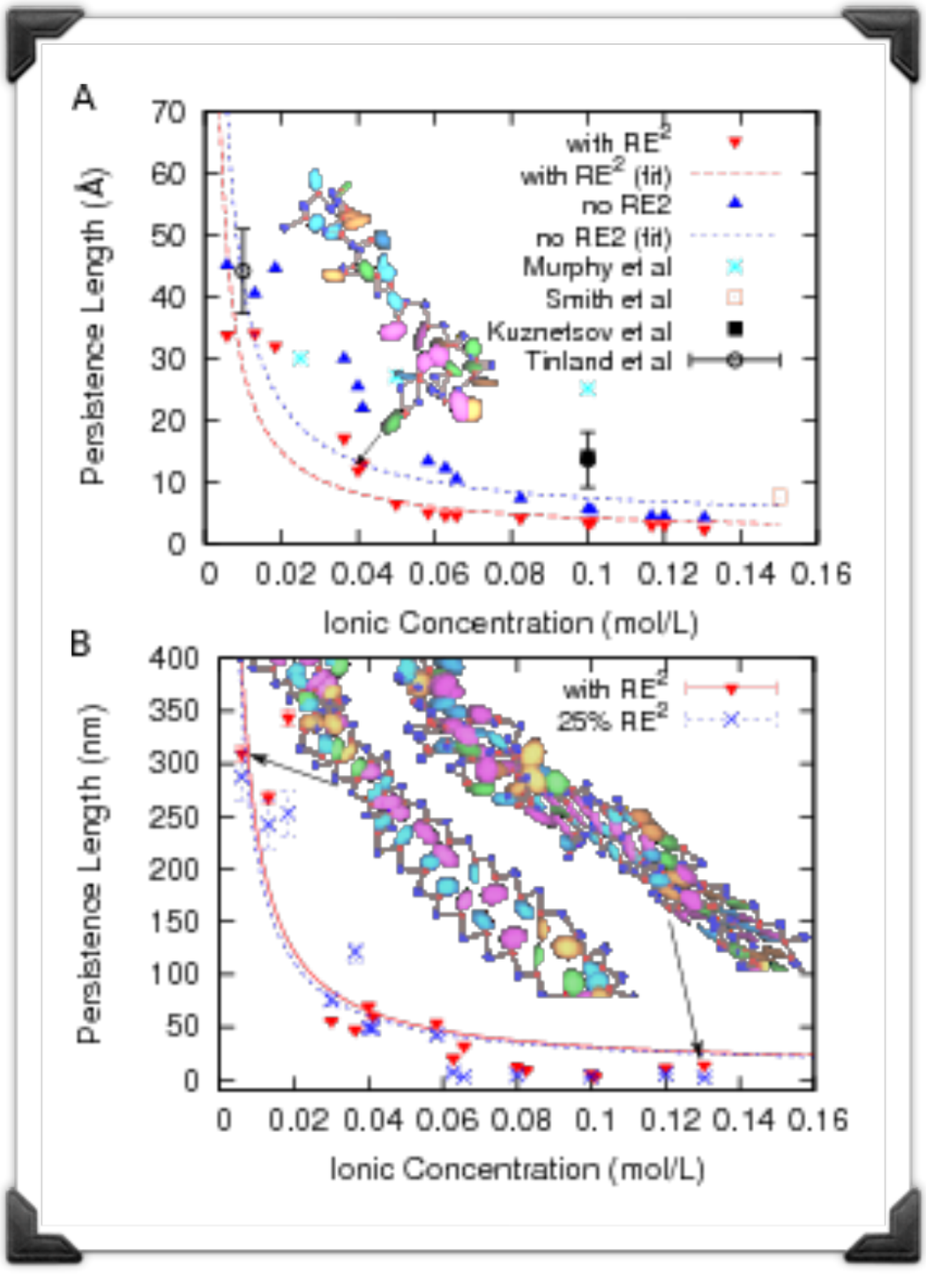
TIS (Thirumalai 2012) : Ribozyme structural transitions on folding





What for?

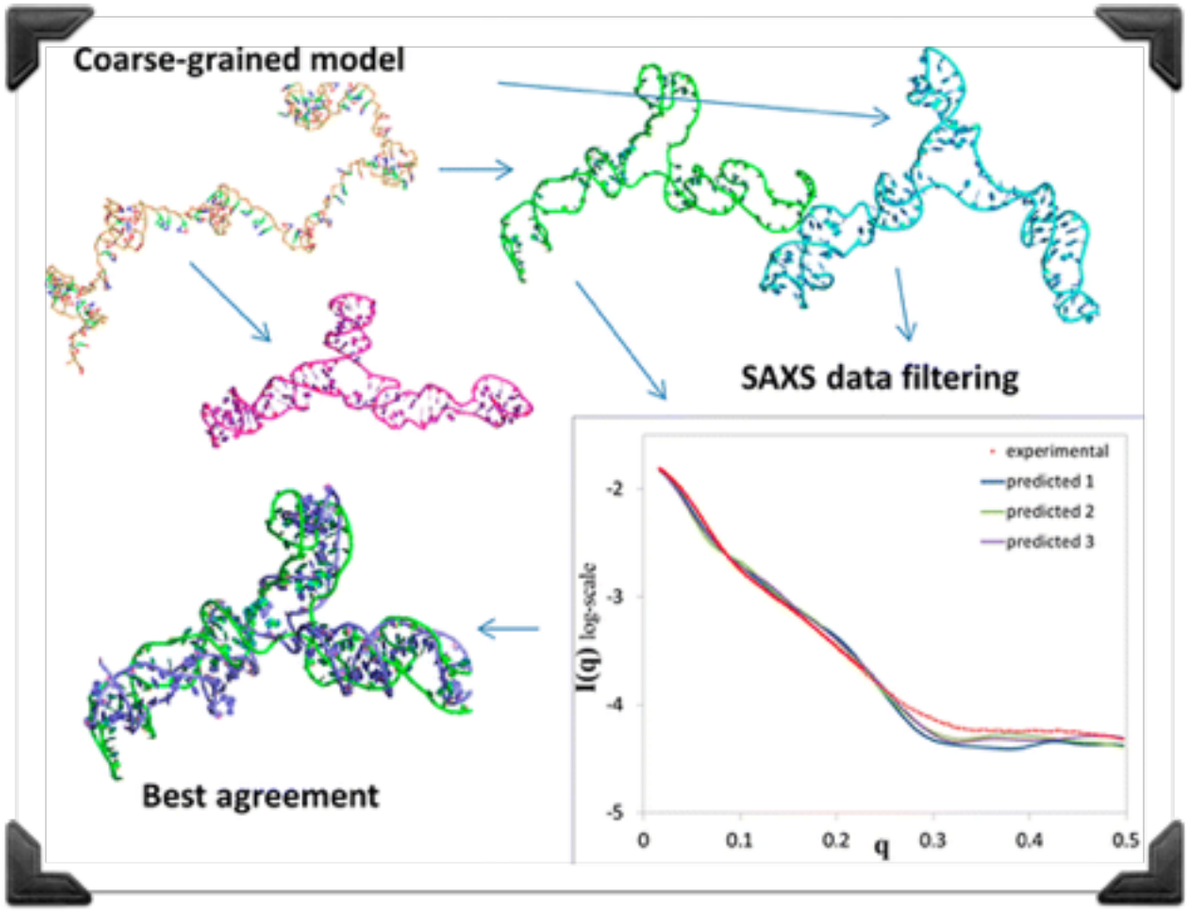
Plotkin (2010) : DNA persistence length



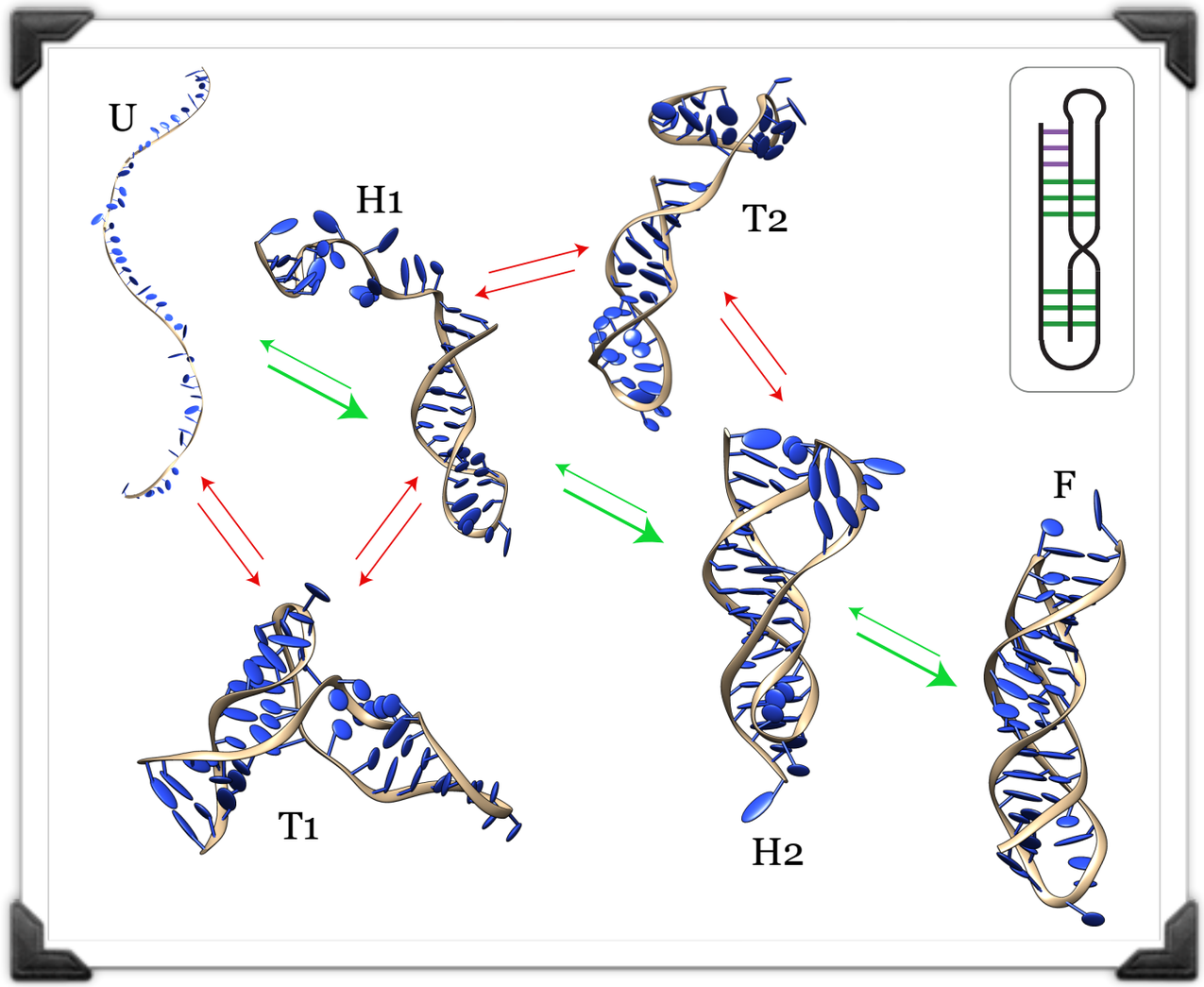
OxDNA (Ouldridge 2010) :
DNA origami and nanotechnology

What for?

Xia (2013) : RNA structure prediction, inclusion of experimental data to guide folding

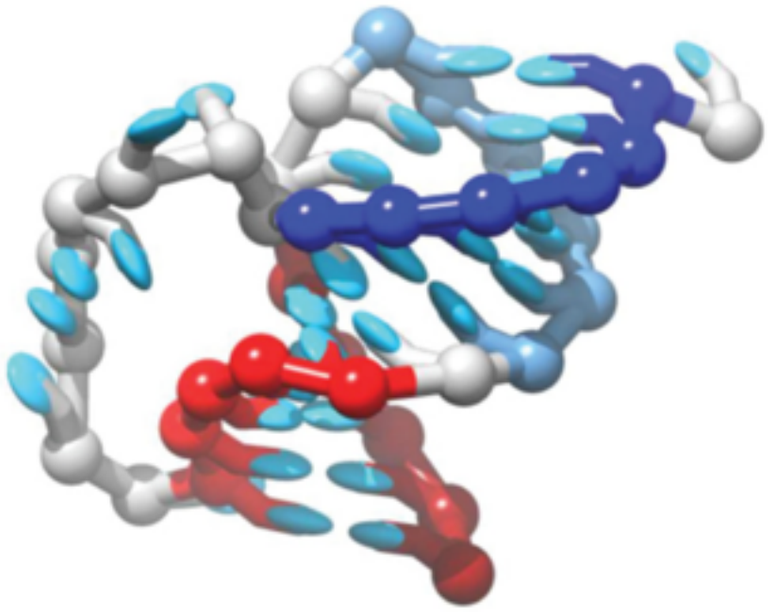


HiRE-RNA (Cragnolini 2015) : RNA folding predictions, folding pathways, dynamics

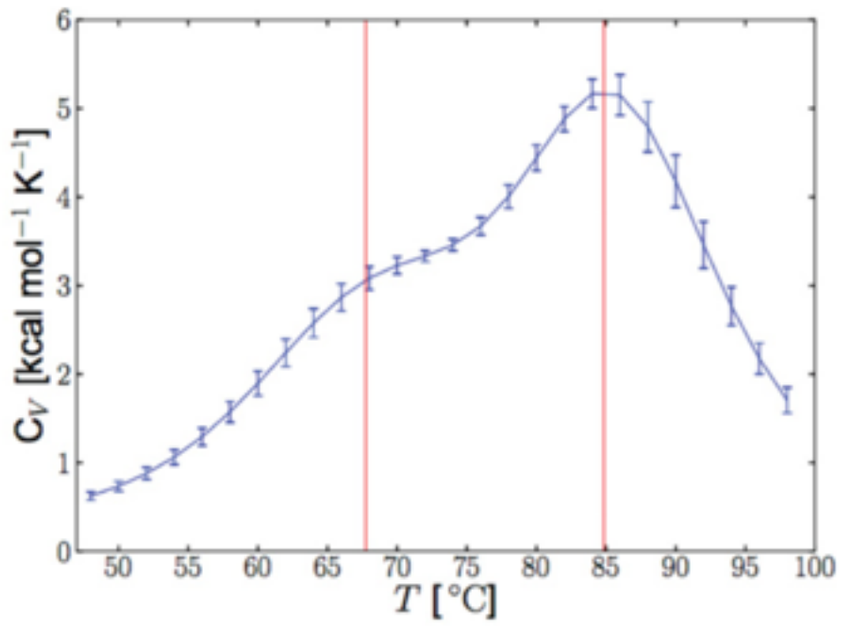


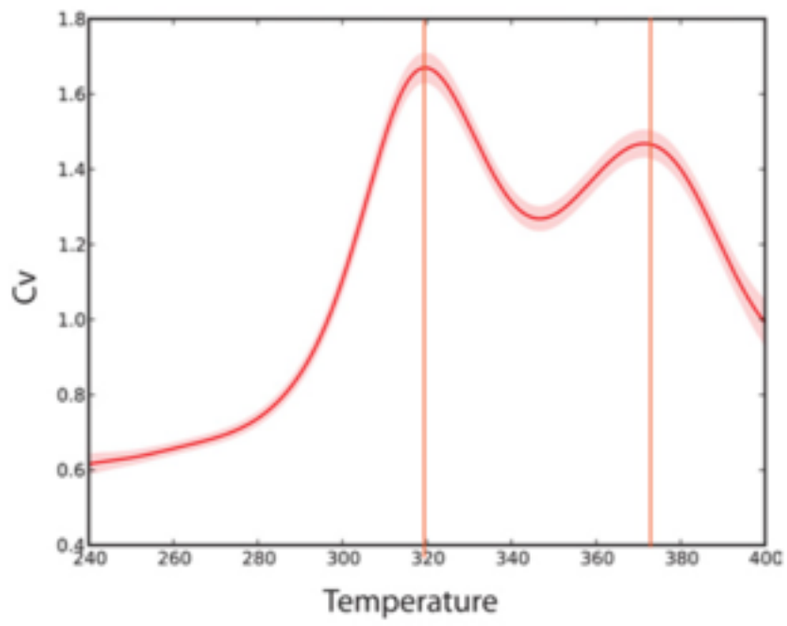
Pseudoknot comparison example

MMTV (OxRNA)

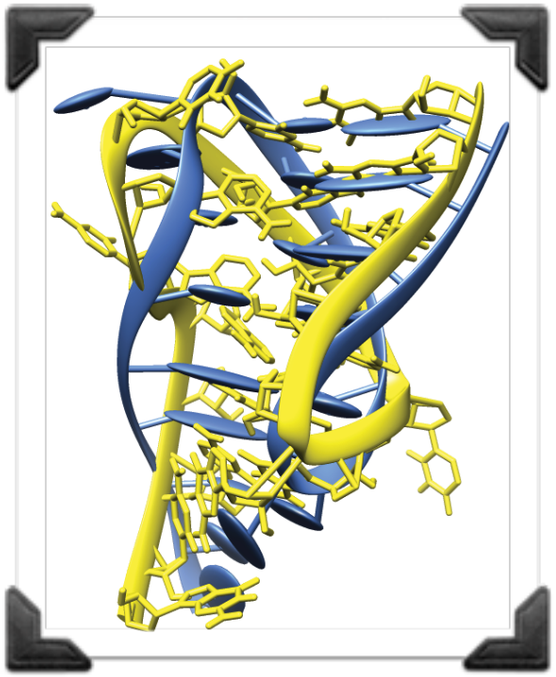


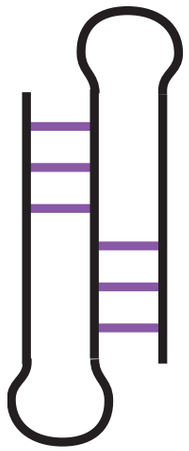
2G1W (HiRE-RNA)



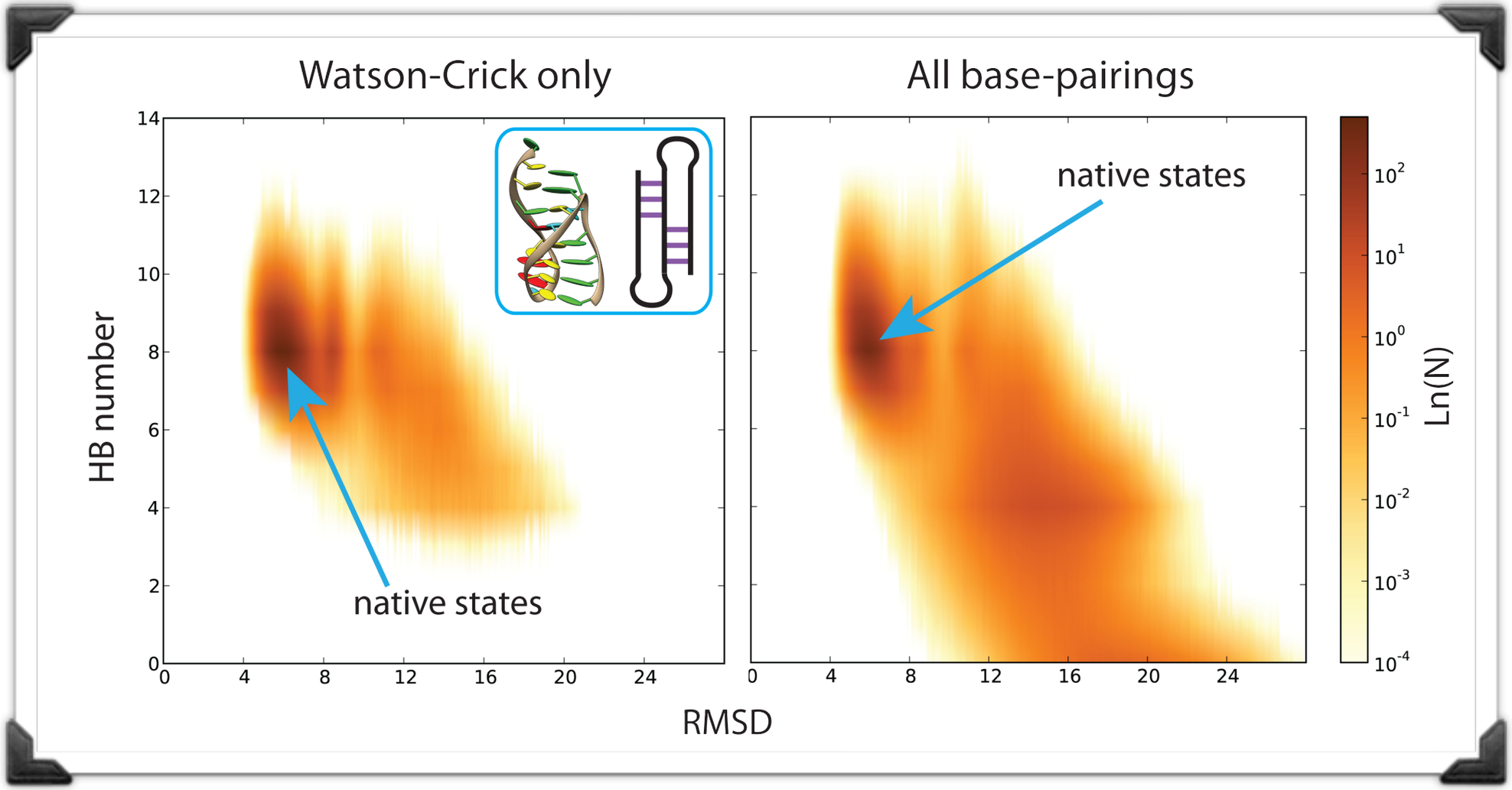


Model's choices do matter



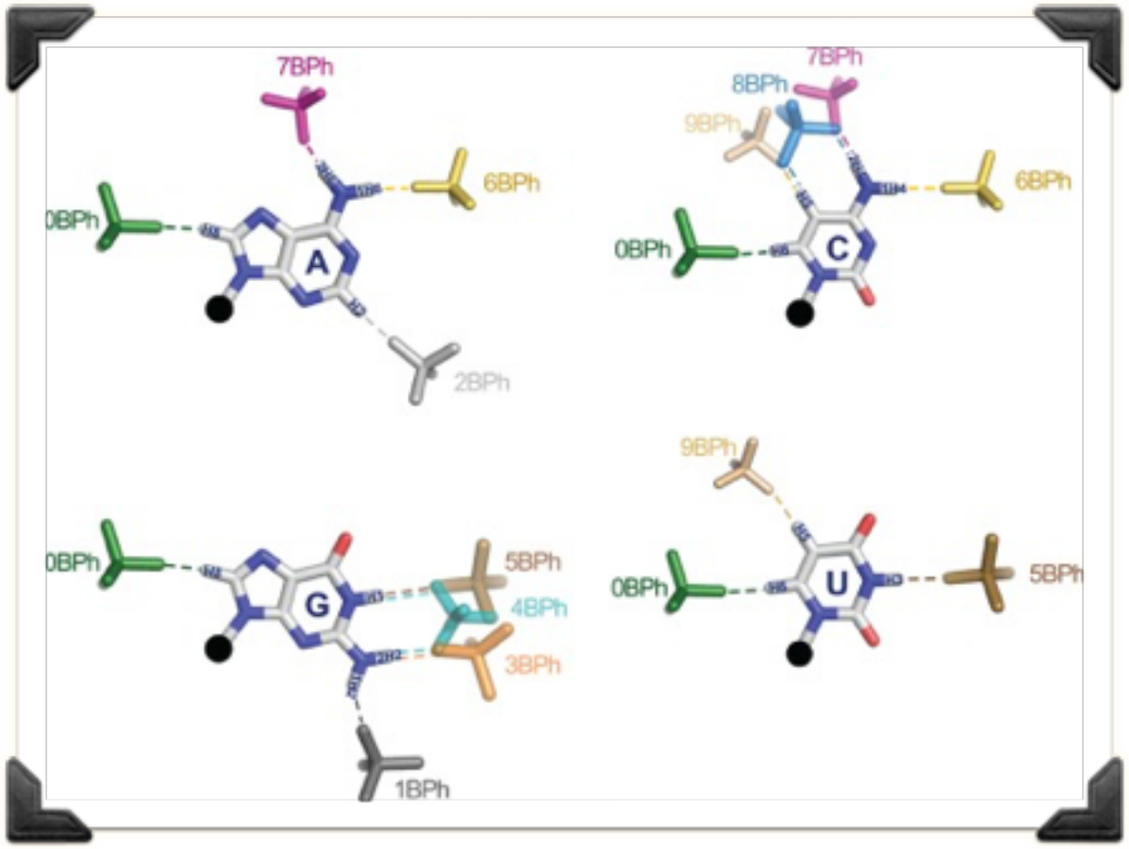


Same structural prediction.
Impact on dynamics, thermodynamics

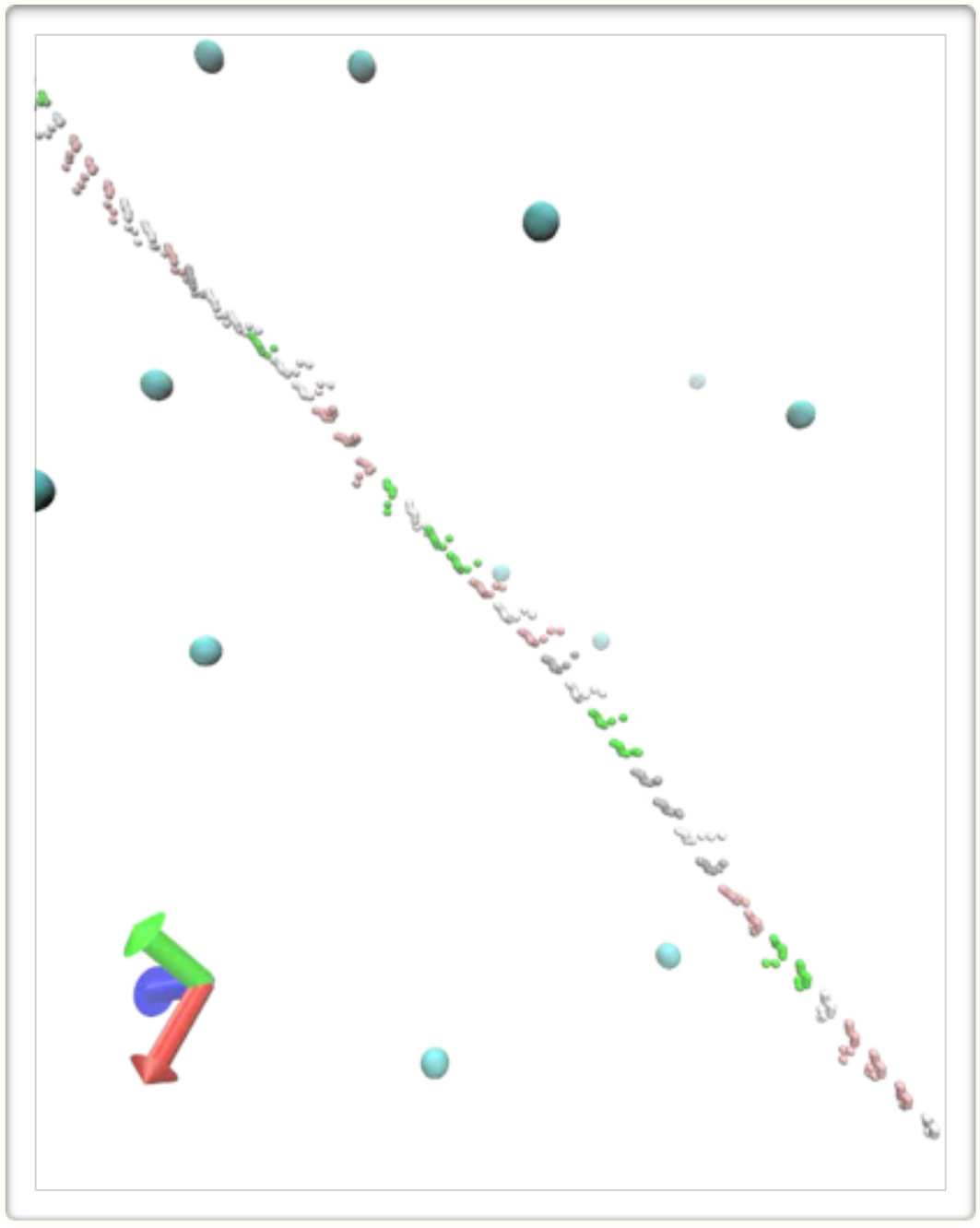


what's missing ?

Base-phosphate interactions



Ions



Exercice 5
Look at an atomistic simulation

You are going to use the program VMD to visually analyze the trajectory of the molecule 1Y26

- Properly visualize the molecule.
  - Using different representations highlight the backbone of the molecule and the bases.
  - Based on the analysis of the initial structure by DSSR highlight stems (with different colors for example).
  - and identify the kissing loop.

- Visualize the trajectory and highlight the movements of the molecule.
  - Compute RMDS.
  - Superpose movements.

- Assess the stability of base pairs by plotting the time evolution of the distance between atoms forming hydrogen bonds. Can you detect differences between canonical and non-canonical pairs?

Exercice 6
Look at a CG simulation

You are going to use the program VMD to visually analyze the trajectory of the molecule 1Y26

- Visualize the molecule and familiarize with the coarse-grained representation. Notice that the visualization of VMD is not optimized for the GC model.

- Visualize the trajectory and highlight the movements of the molecule.
    - Compute RMDS.
    - Superpose movements.

- Assess the stability of base pairs by plotting the time evolution of the distance between atoms forming hydrogen bonds.

- Draw a comparison between atomistic and coarse-grained simulations for the system under investigation.