Development of coarse-grained models for nucleic acids (and aromatic systems)

Samuela Pasquali & Elisa Frezza

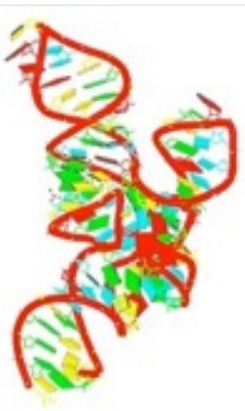
Laboratoire de Cristallographie et RMN Biologiques
Faculté de Pharmacie, Université Paris Descartes

UNIVERSITÉ PARIS DESCARTES

Nucleic acids complex structural architectures

nt: 20-25 | ~30 | 40 | 60 | 80 | 100 | 300 | > 1000

- microRNA, siRNA: 20-25 nt
- piRNA: ~30 nt
- tRNA, riboswitches: 80 nt
- ribosomal RNA: > 1000 nt
- ribozymes: 40 – 100 nt
- snoRNA: 60 – 300 nt
- mRNA: 100 – > 1000 nt



tRNA

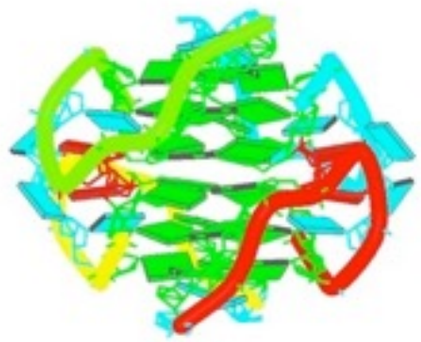
telomerase

viral fragment

riboswitch

ribozyme

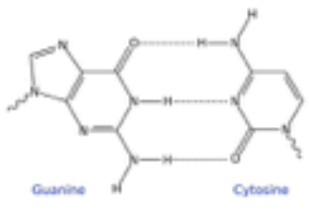
triple helix

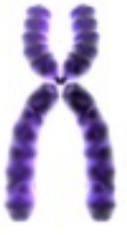


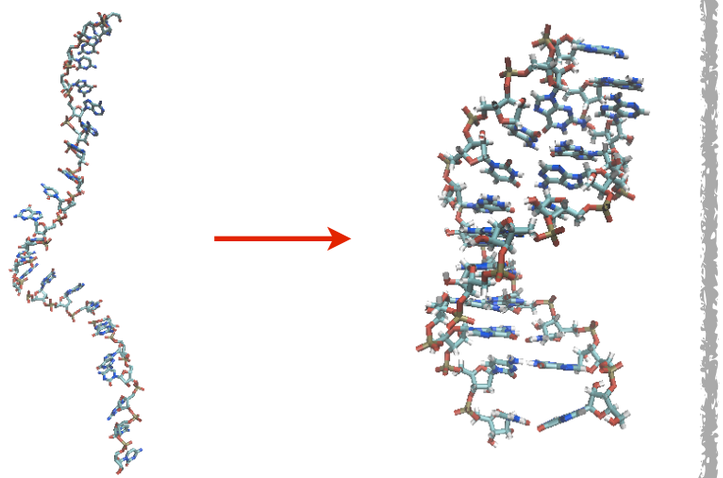
Physical description

Prediction of the dynamical and thermodynamical behavior in 3D

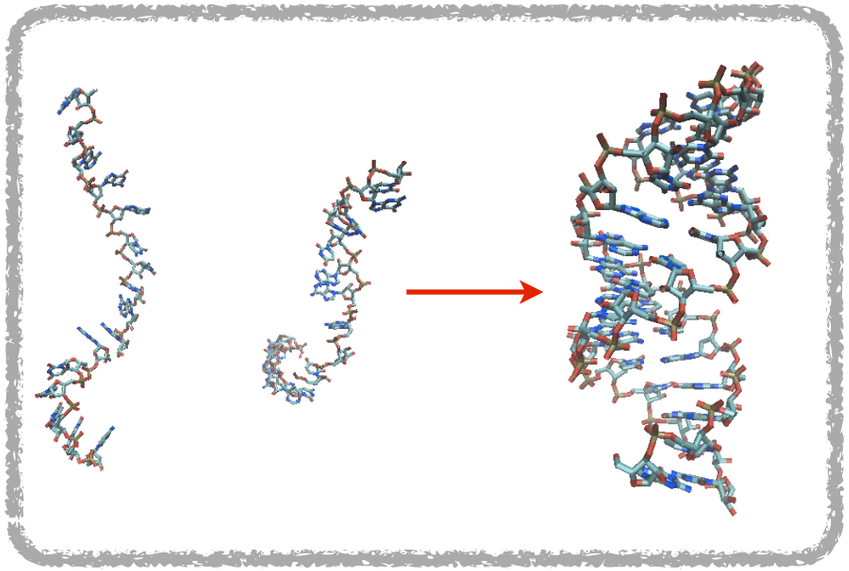
Å nm **10 nm** 100 nm μm







Folding



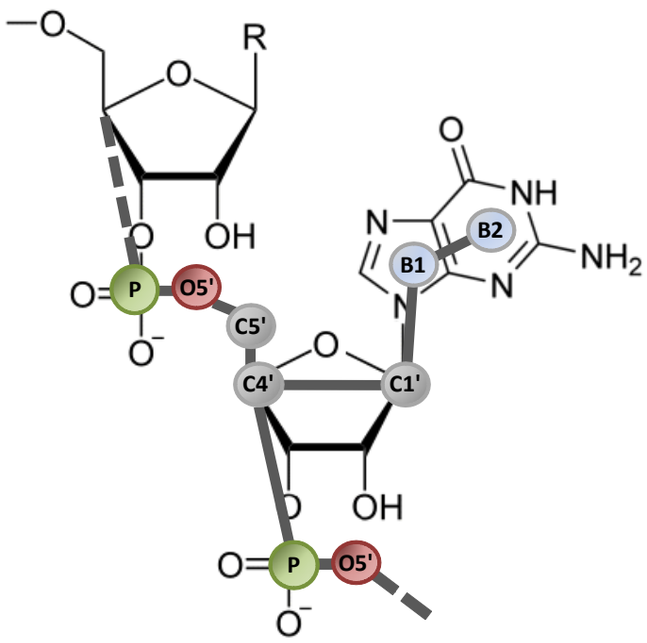
Assembly

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

model

Coarse-grained RNA modeling

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



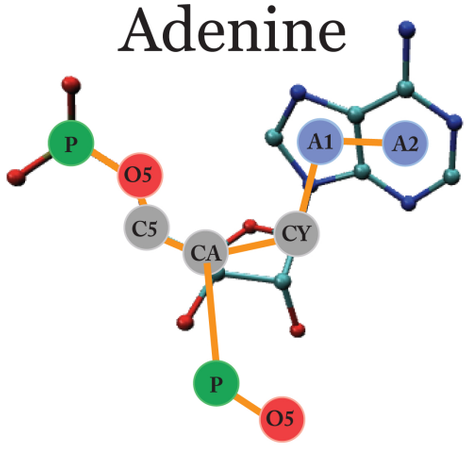
HiRE-RNA

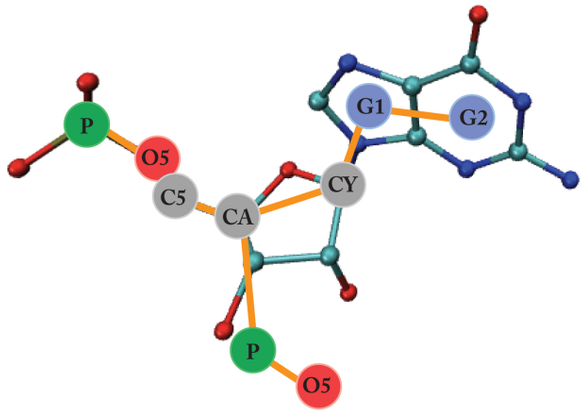
High Resolution Energy Model for RNA and DNA

Flexible, unconstrained

MD, REMD, ST, MC

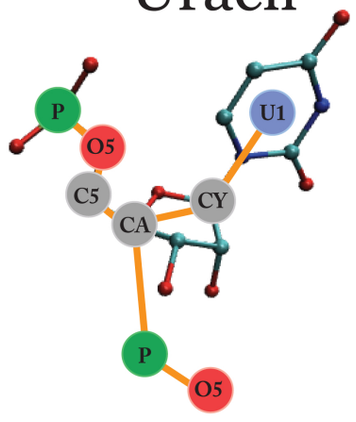
Adenine

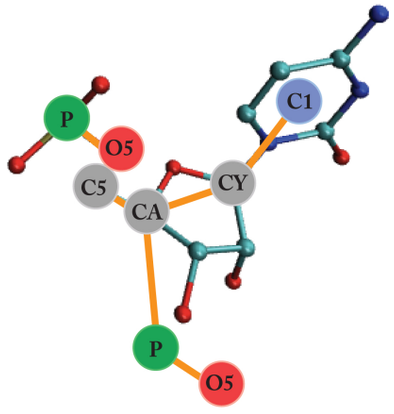




Guanine

Uracil





Cytosine

model

… Why is RNA not a protein …

At large distances the dominant effect should be the ELECTROSTATIC repulsion, with Van der Waals forces being subdominant

At long-range is LJ-like potential appropriate/necessary?

STACKING is the hydrophobic behavior of bases and it is short-ranged

At short-range LJ-like potential between bases are inadequate

Hydrogen bonding occurs in the base PLANE

Local geometries have to be taken into account

Bases can form hydrogen bonds on 3 different SIDES

Non-canonical base pairs and multiple pairings have to be included

model

HiRE-RNA, version 3

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

harmonic
statistical parameters

Planarity
Non-canonical pairs

Base orientation

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

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

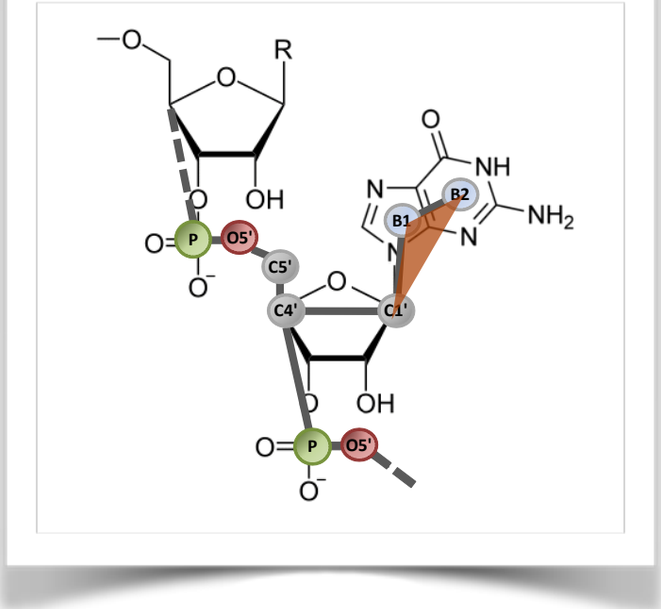
genetic algorithm parameter optimization
NDB - topology based

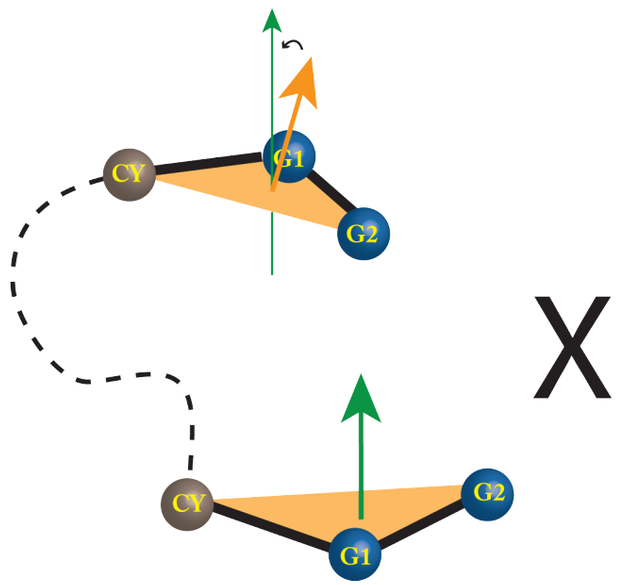
T. Cragnolini, Y. Laurin, P. Derreumaux, S. Pasquali, JCTC (2015)

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

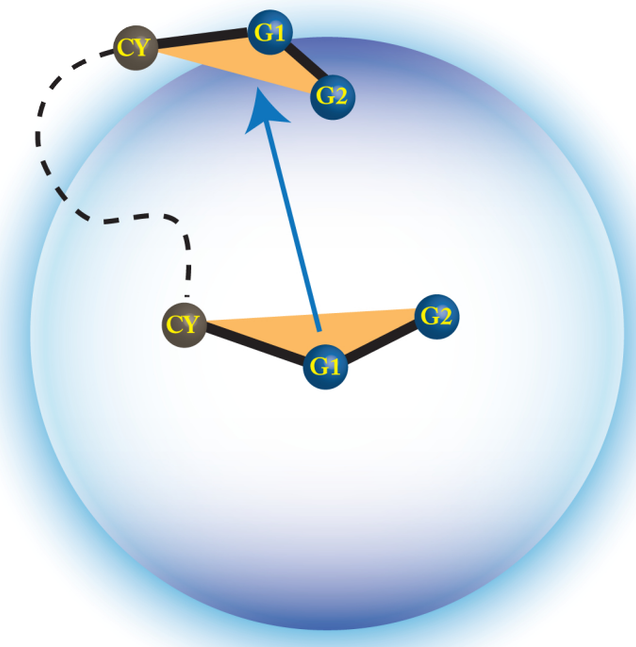
model

Stacking





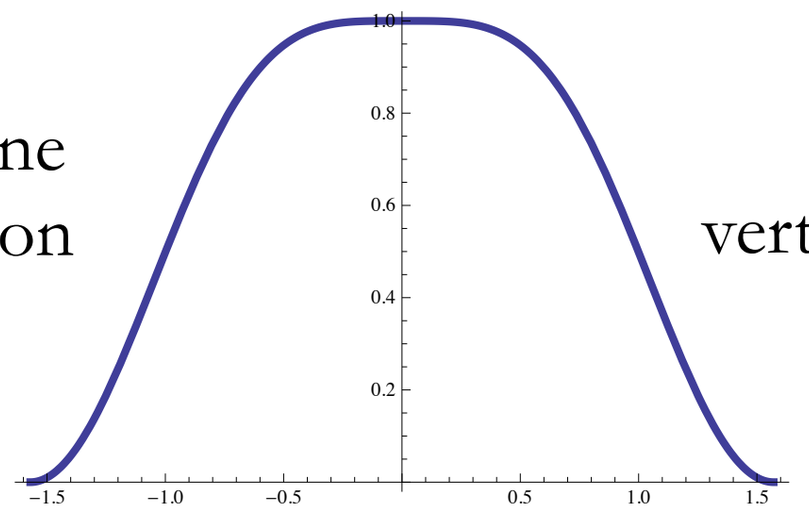
X



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

plane distance

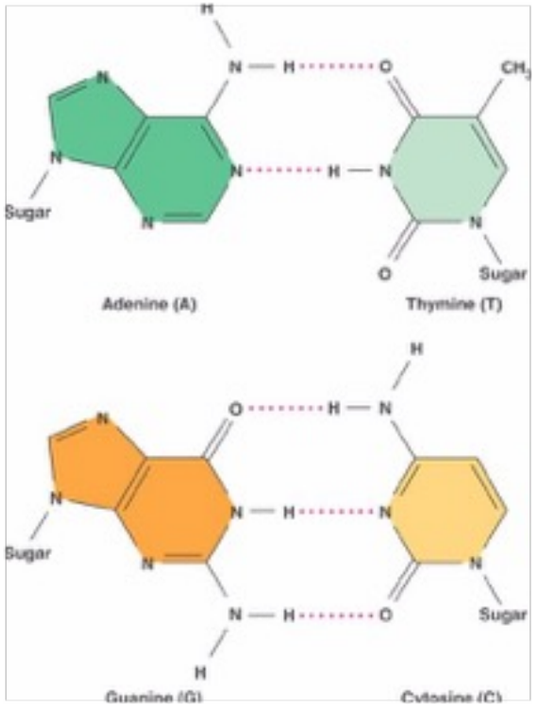
same plane orientation

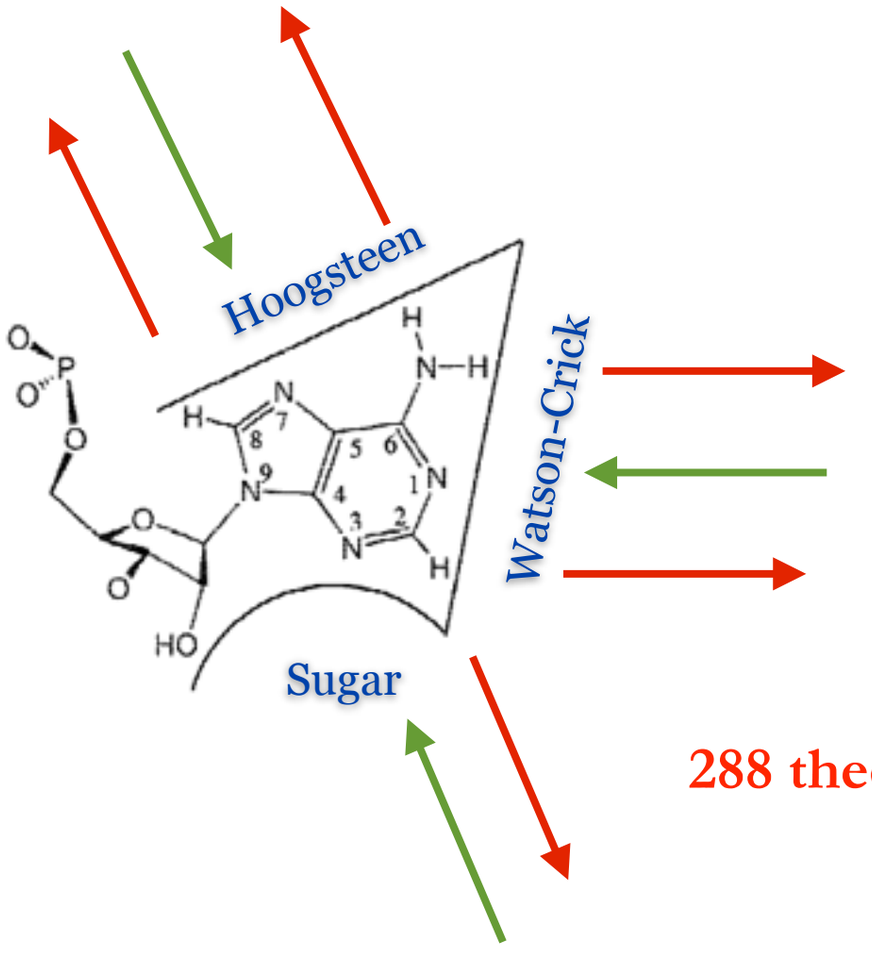


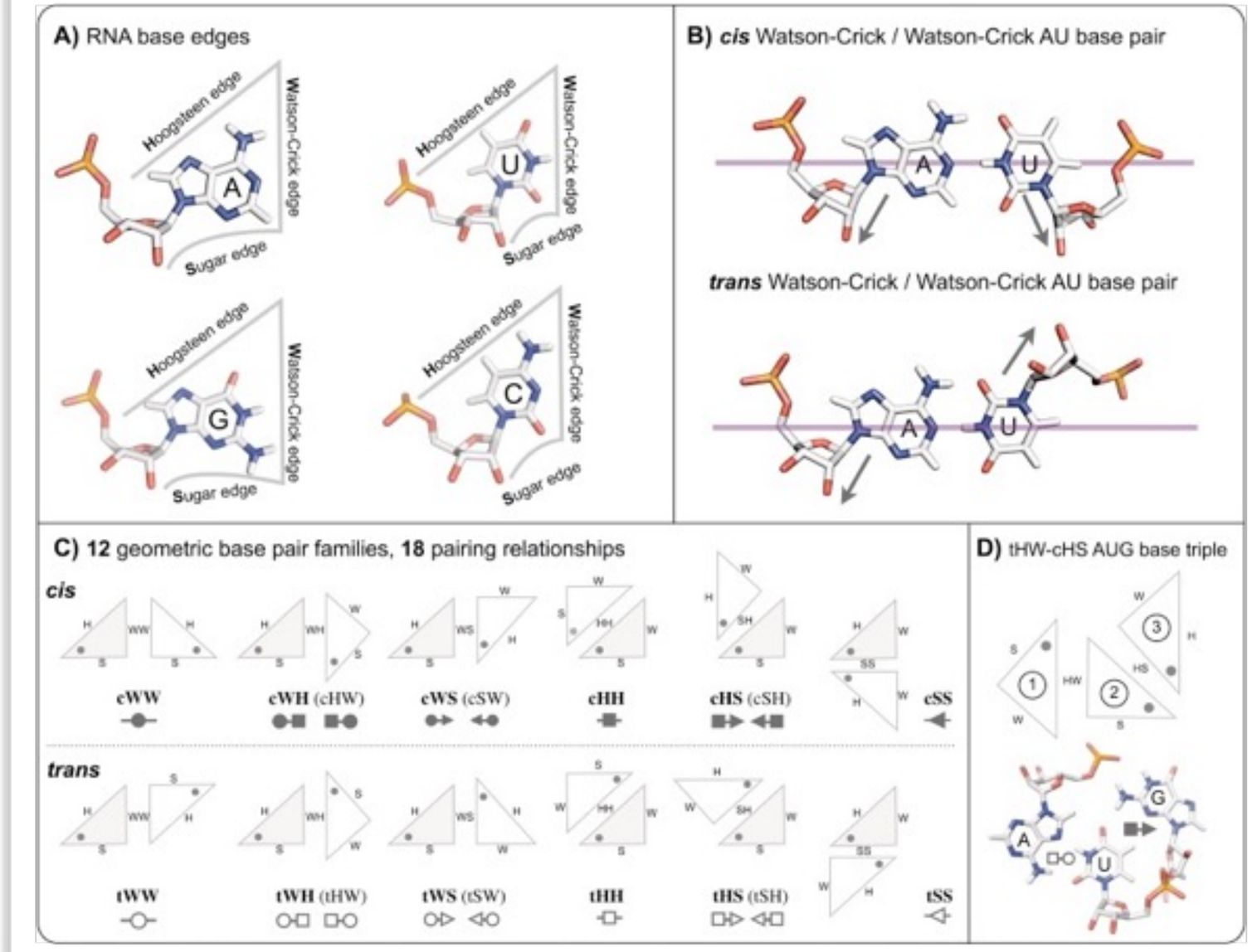
vertical position

Base pairing

canonical and non-canonical





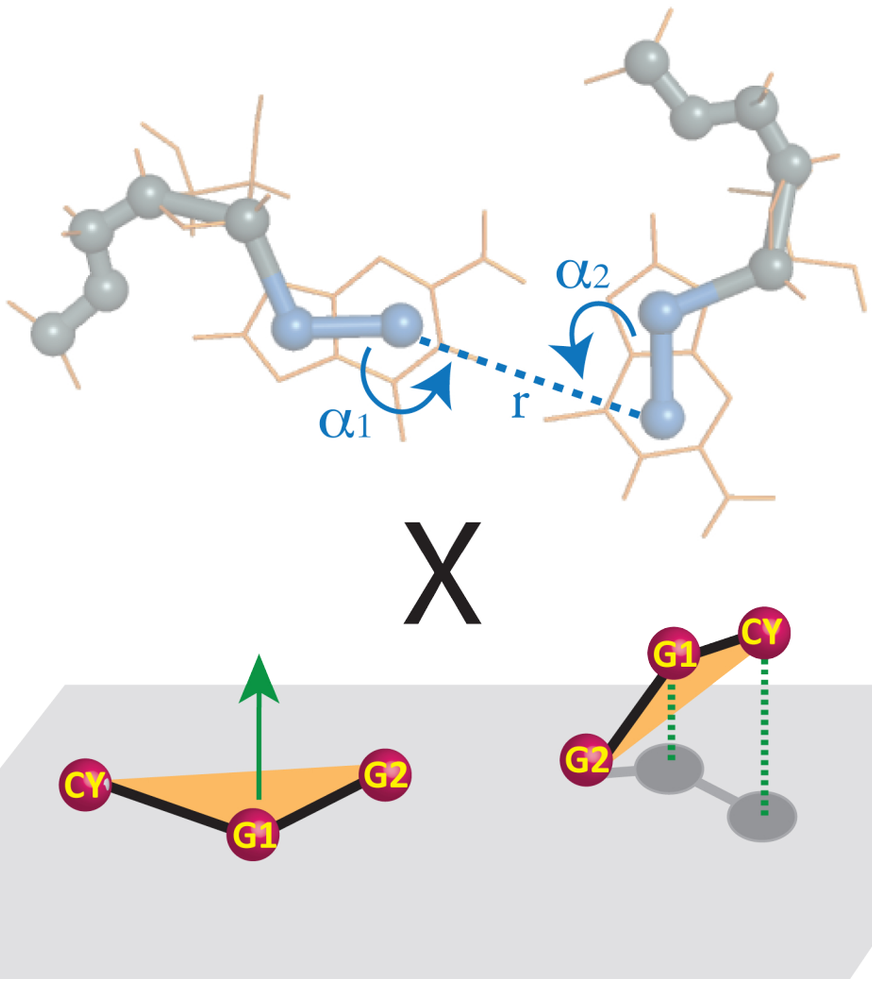


288 theoretically possible pairs ⇢ 145 found experimentally (NDB)

model

Base Pairing

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

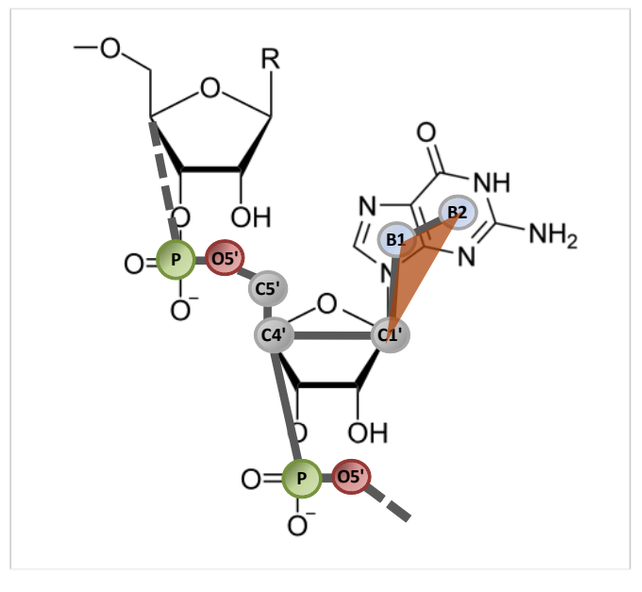


$$E_{hb} = \varepsilon_{hb} e^{-(r-\rho)^2/\xi}\, \nu(\alpha_1)\nu(\alpha_2)$$

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

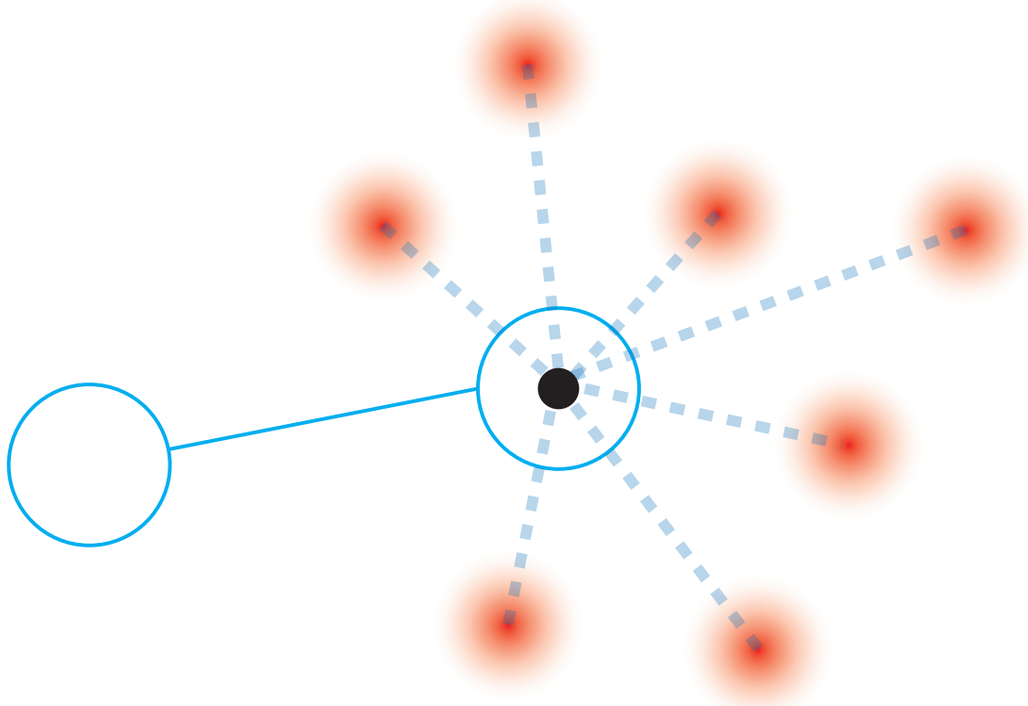
$$\alpha = \begin{cases} +\alpha, & \text{if } \cos(\tau) \ge 0; \\ -\alpha, & \text{otherwise} \end{cases}$$

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



model

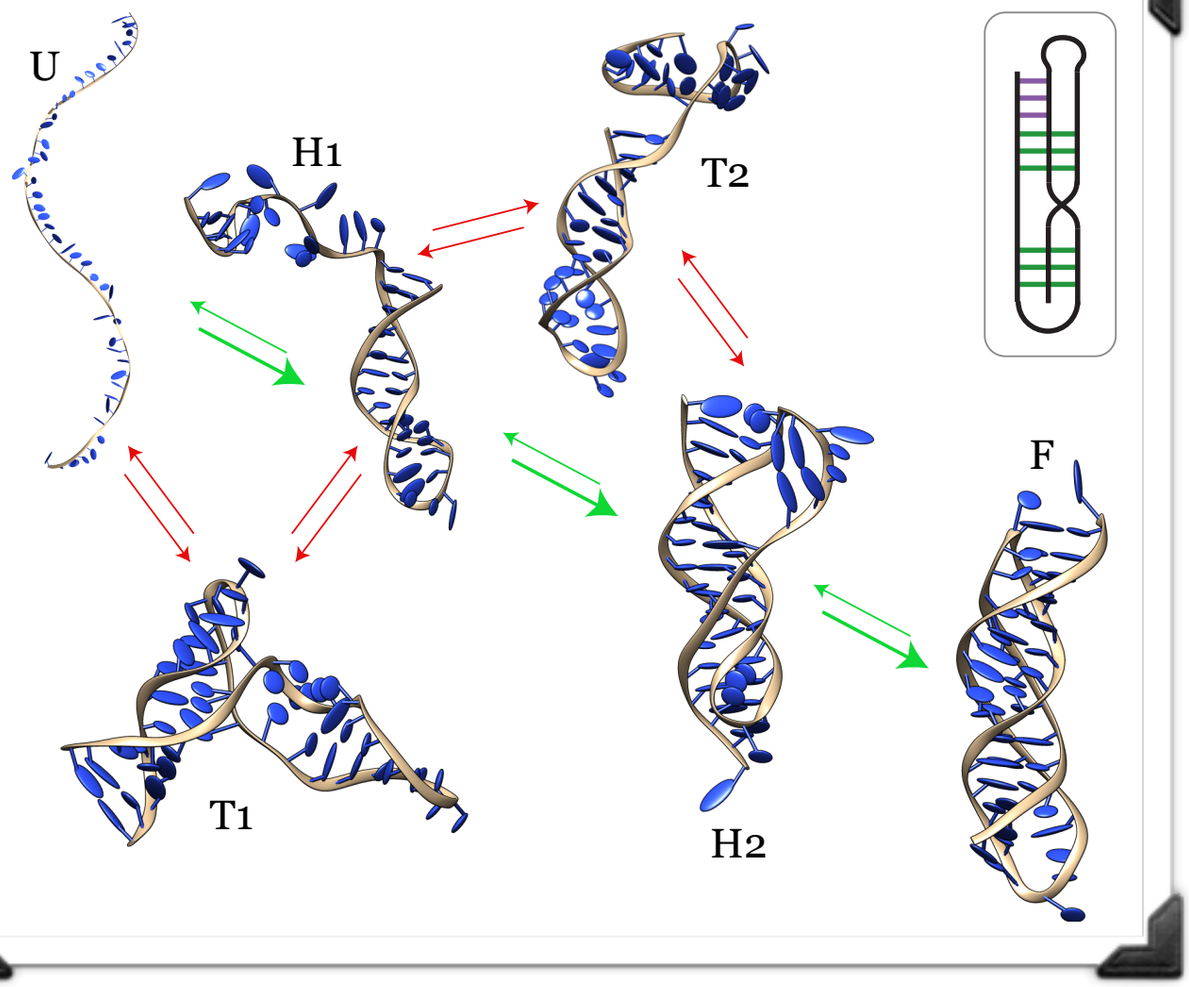
Non-canonical pairings



A A *trans HH* (2)	A A *trans HS* (1)	A A *trans WcWc* (2)
A C *cis WcWc* (1)	A C *cis WcWc* (1)	A G *cis WcS* (1)
A G *cis WcWc* (2)	A G *trans HS* (2)	A G *trans WcWc* (1)
A U *cis WcWc* (2)	A U *trans WcH* (2)	C C *trans WcWc* (2)
C U *cis WcWc* (1)	C U *trans WcH* (1)	G C *cis WcWc* (3)
G C *trans HH* (1)	G G *cis WcH* (1)	G G *cis WcH* (2)
G U *cis WcWc* (2)	G U *trans WcWc* (1)	U U *cis WcWc* (2)
U U *trans WcH* (1)		

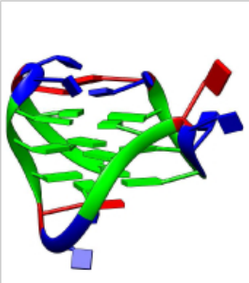
Wc: Watson-Crick
H: Hoogsteen
S: Sugar

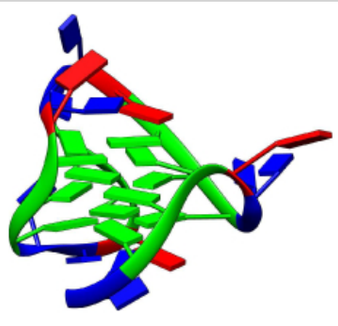
HiRE-RNA, version 3

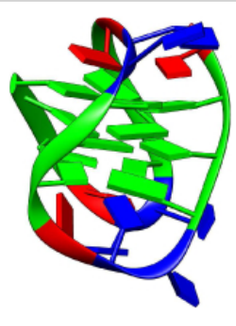


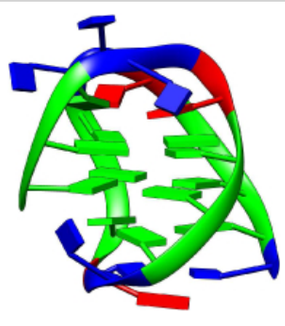
← Triple helix folding

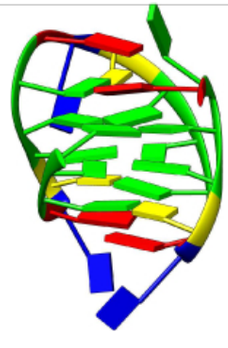
G-quadruplexes unfolding

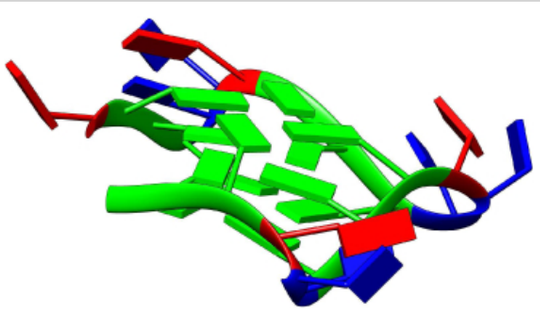

(1)$3+1_1$
2JSM


(2)$3+1_2$
2JSL


(3)basket_1
143D


(4)basket_2
2KF8


(5)chair
2KM3


(6)propeller
1KF1

Exp

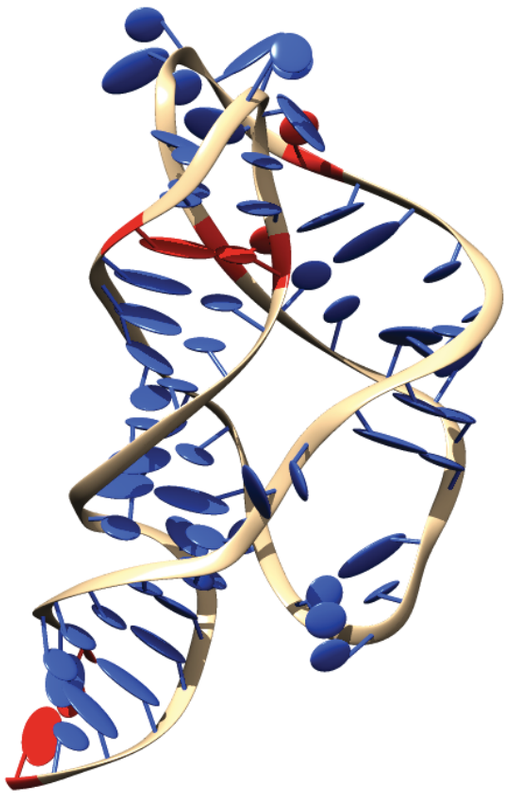
Inclusion of experimental data

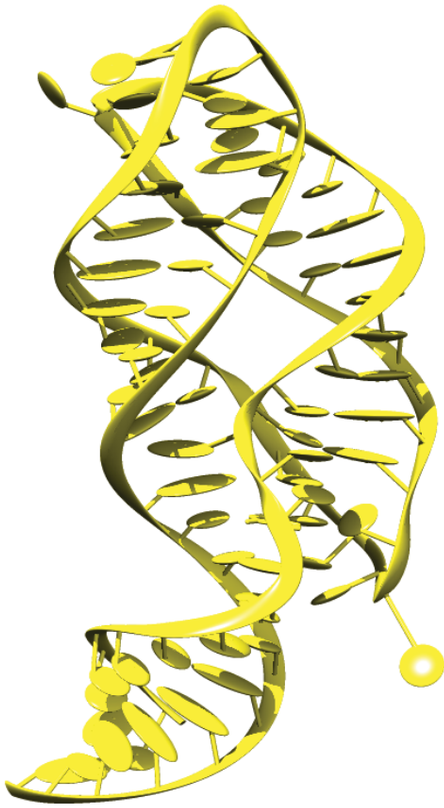
Low-resolution techniques : SAXS, Cryo-EM → **Biased simulations**
Interactive simulations

High-resolution techniques : biochemistry, NRM, X-ray → **Constraints**

Single-molecule experiments : FRET, optical tweezers → **External forces**
Constraints

Contraintes d'appariement de bases





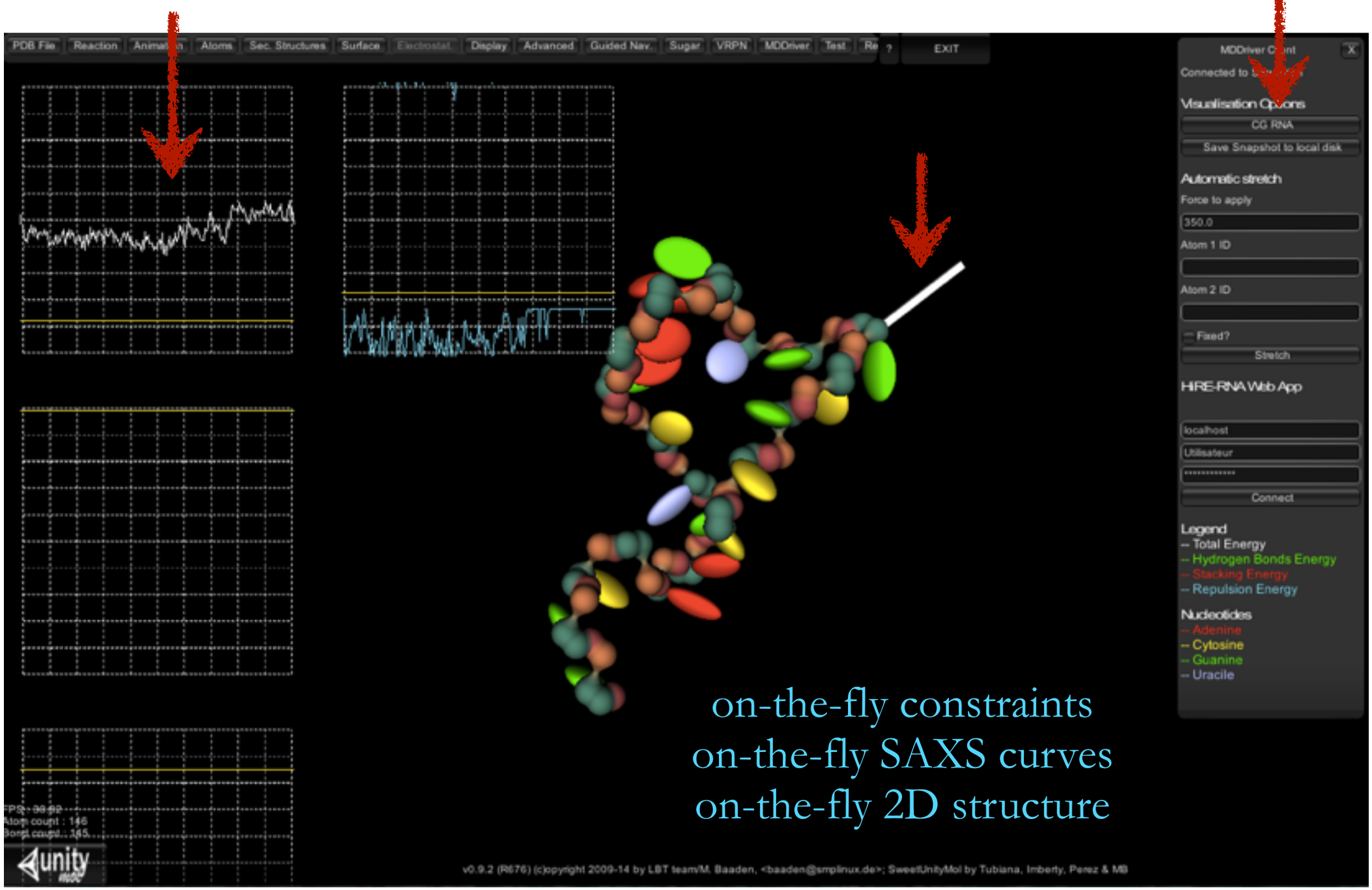
3 contraintes d'appariement de bases
~7Å rmsd

Exp

Interactive simulations: UnityMol + HiRE-RNA

Energetic monitoring:
total, electrostatic, stacking, base-pairing

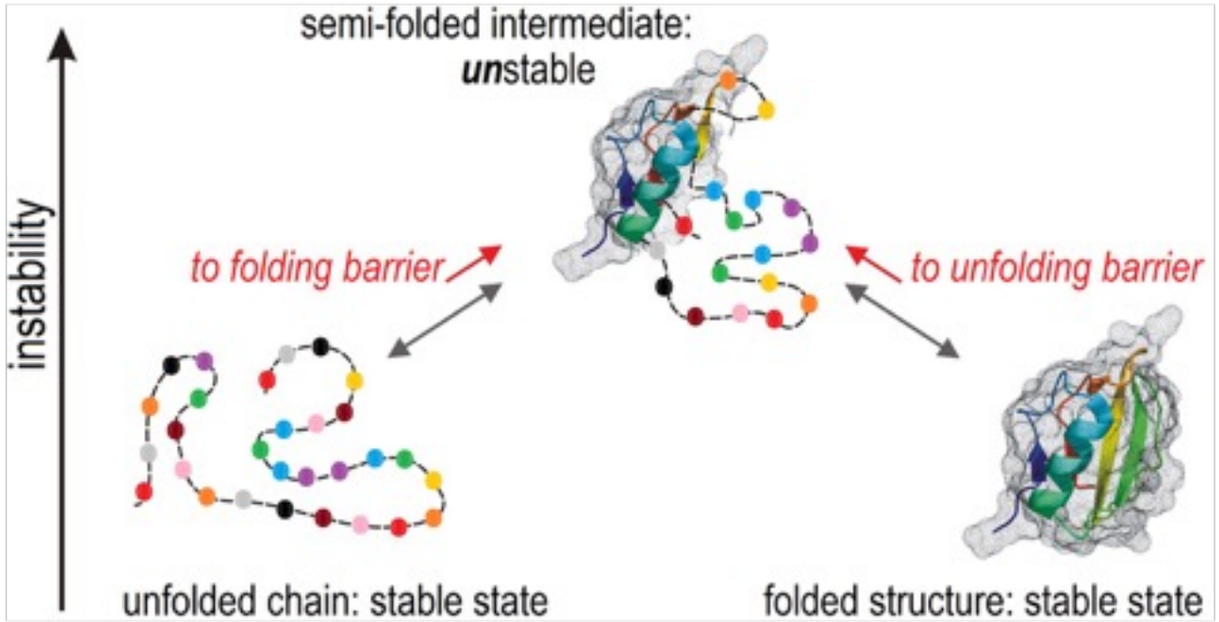
Simulation interface

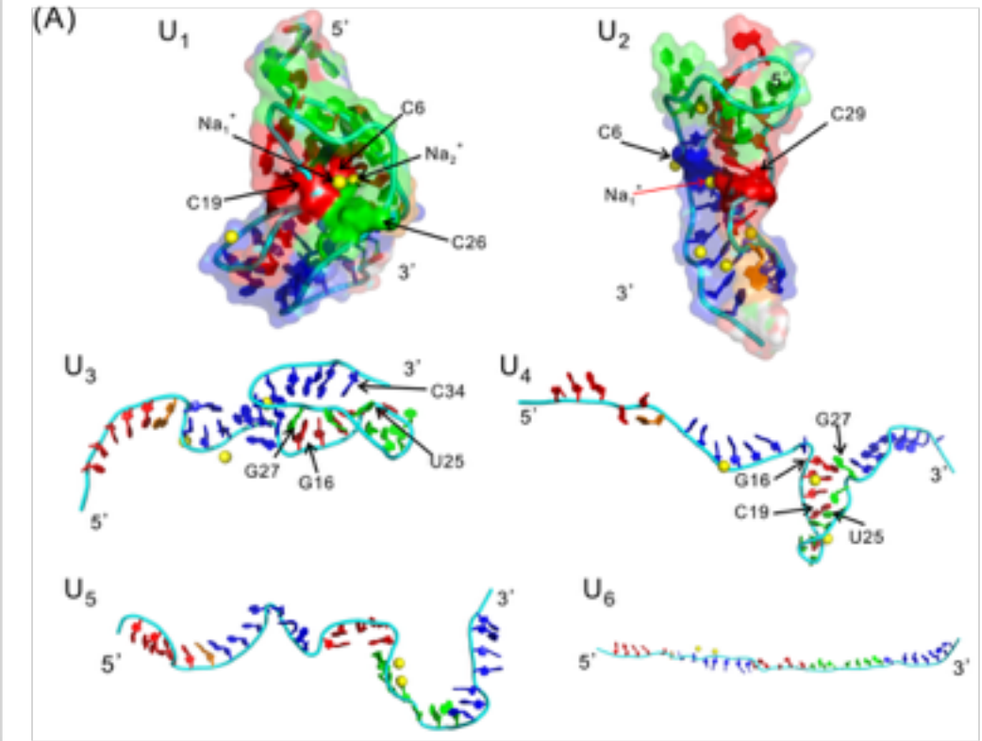


S. Doutreligne, P. Derreumaux, S. Pasquali, M. Baaden (2015)

S. Doutreligne, L. Mazzanti, A. Taly, P. Derreumaux, M. Baden, S. Pasquali (2017)

Behavior of biomolecules



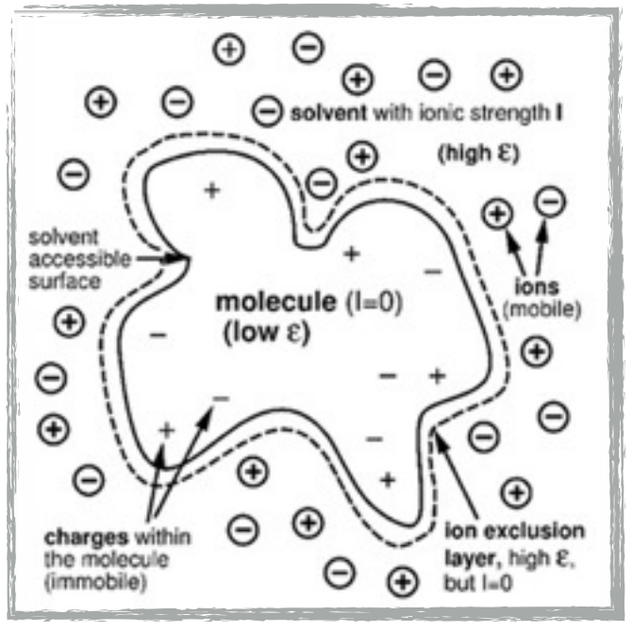


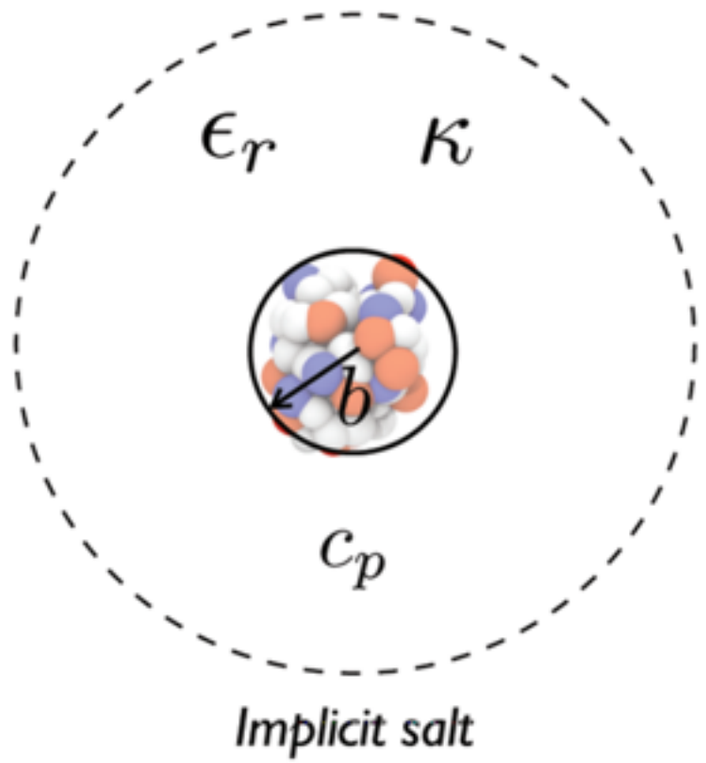
Behavior depends on environment

Temperature, ligands, ions, … pH

Coarse-grained models to study structural changes ⟷ Titration scheme to account for pH and salt

Tanford-Kirkwood model (1934, 1957)





Molecule represented as a sphere impenetrable to solvent.
Titratable group are independent (interact only through electrostatics)
⇢molecule's titration curve as superposition of titration curves of individual types of groups

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

↘ protonation (+)
deprotonation (-)

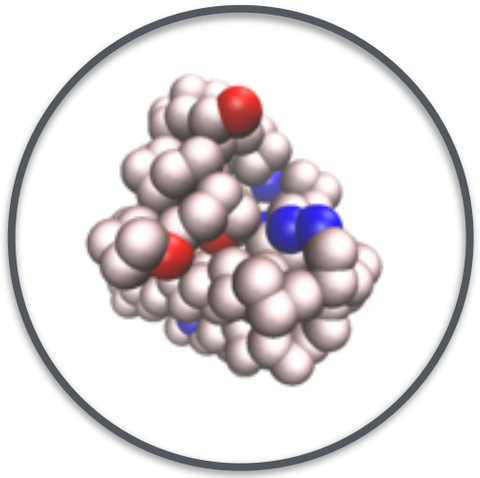
Fast Monte Carlo titration scheme

Texeira, Lund, Barroso da Silva, JCTC, 2010

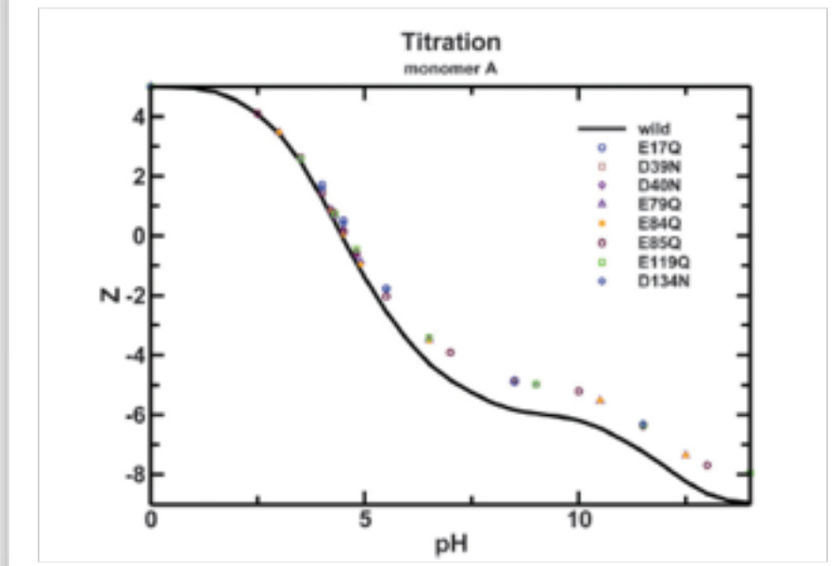
Barroso da Silva, MacKernan, JCTC 2017

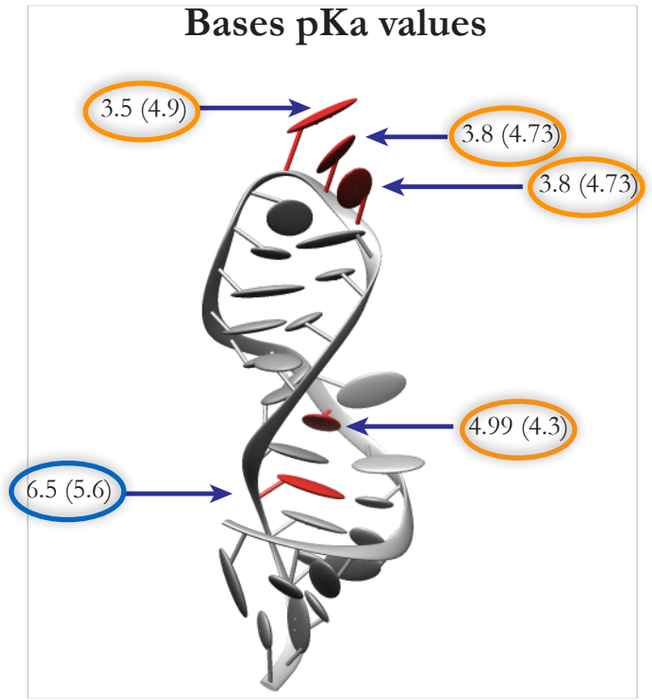
pH

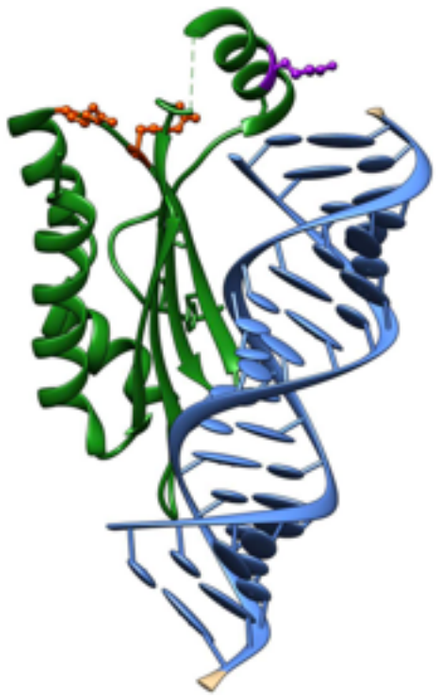
Fast MC titration

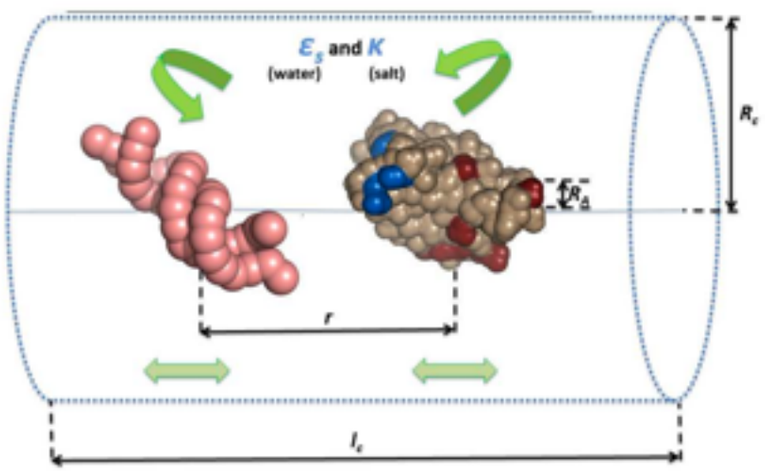


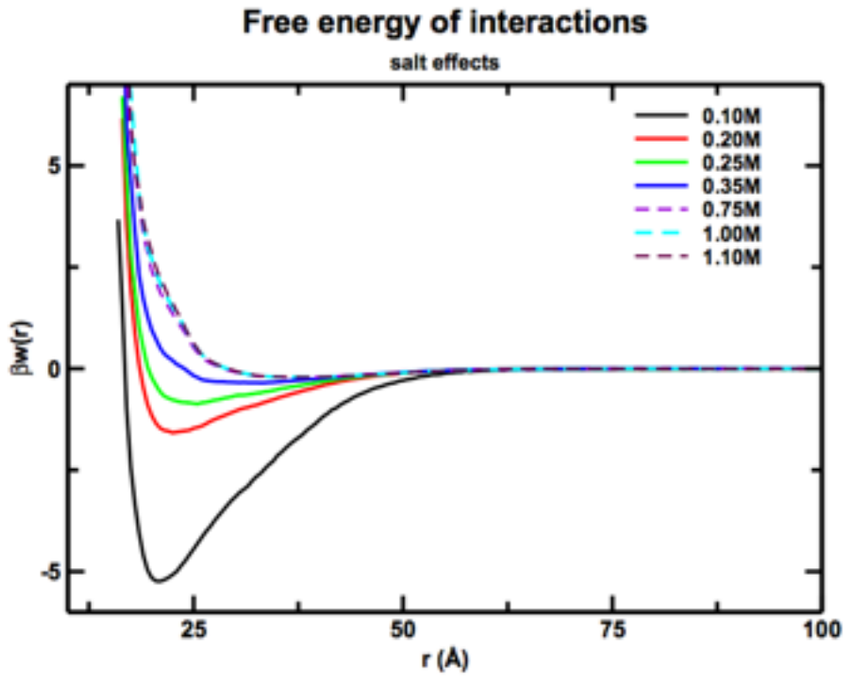
individual effective pK_a values









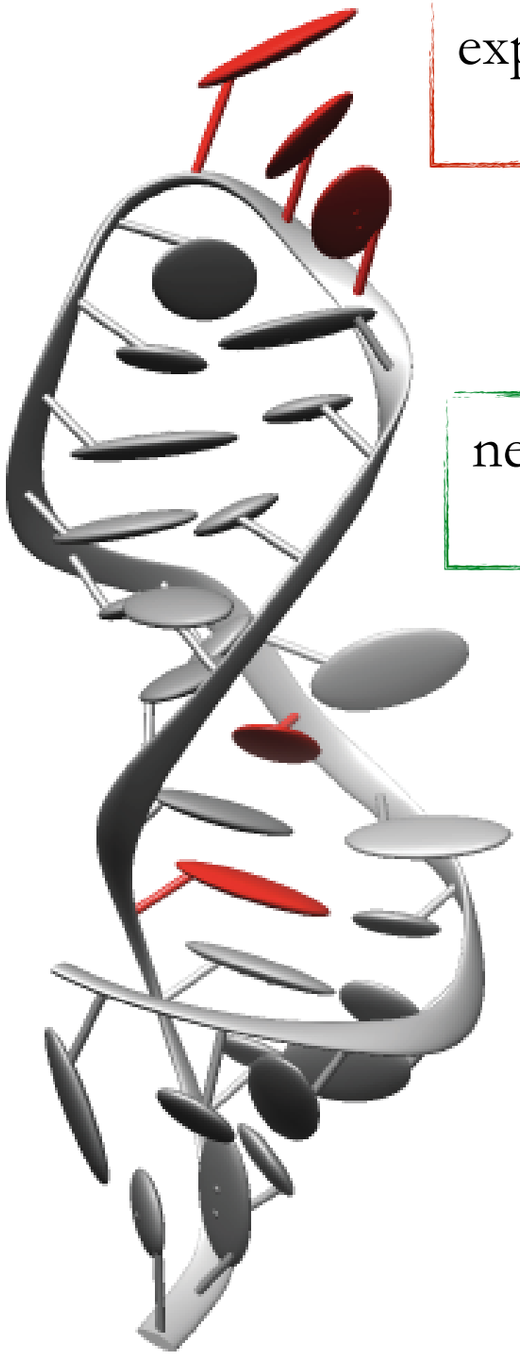


Barroso da Silva, Pasquali, Derreumaux, Dias, Soft Matter 2016

Barroso da Silva, Derreumaux, Pasquali, BBRC 2017

Barroso da Silva, Derreumaux, Pasquali, J Chem Phys 2017

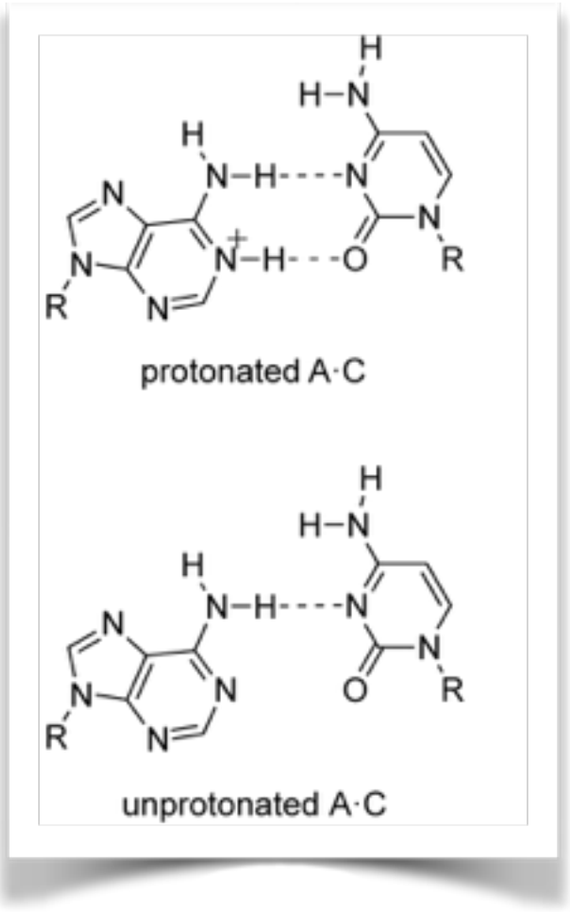
Base protonation is intertwined with base pairing!

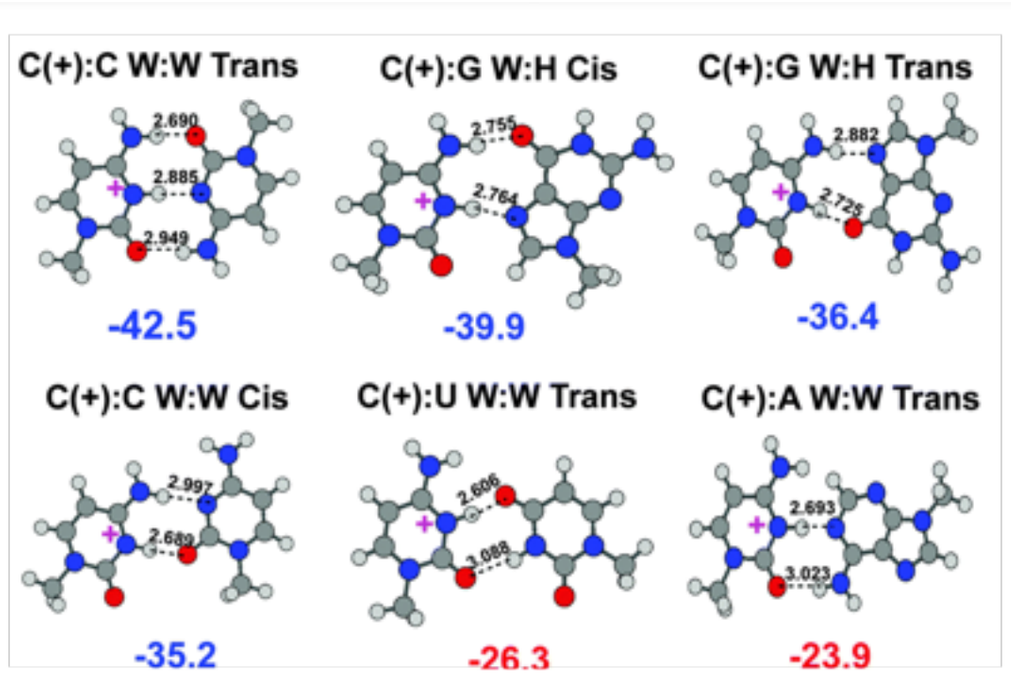


exposed ⇢ protonable
pK_a ~ isolated

neutral paired ⇢ protected
lower pKa

N^+ paired ⇢ protected
higher pKa





Idex Brazil project

HiRE-RNA v3 achievements

- ☑ **Correctly fold molecules of complex architectures, including triplets and quadruplets, giving access to folding pathways and metastable states.**
- ☑ **Investigate the importance of non-canonical paris in RNA folding**
- ☑ **Give access to the plurality of states of G-quadruplexes and study the possible interconversions between different conformations.**
- ☑ **Development of interactive simulation software for teaching and experimentalists (software presentation on Friday)**

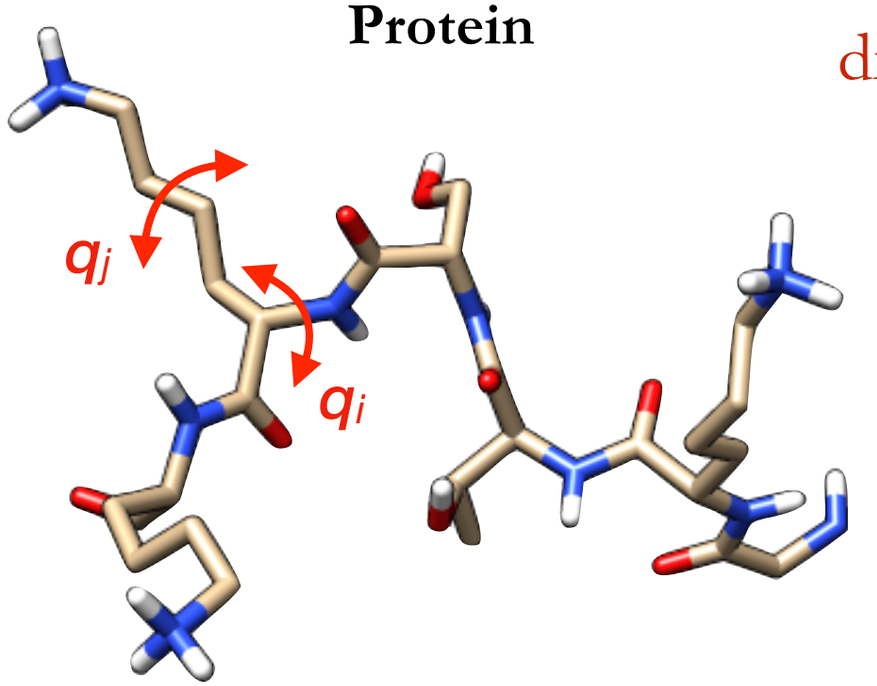
Future directions (to do list)

- ☐ **HiRE-RNA v4, including ions and base-phosphate interactions**
- ☐ **Enhance sampling for rare events** (collaboration D. Wales)
- ☐ **Proteins/Nucleic acids systems** (collaboration LBT)
- ☐ **Strenghten coupling with experiments** (collaborations LCRB, LBT)
- ☐ **Couple Titration and HiRE-RNA** (collaboration F. Barroso da Silva)
- ☐ **Generalization to other aromatic systems** (collaboration B. Baumeier)

immediate future

Internal coordinates

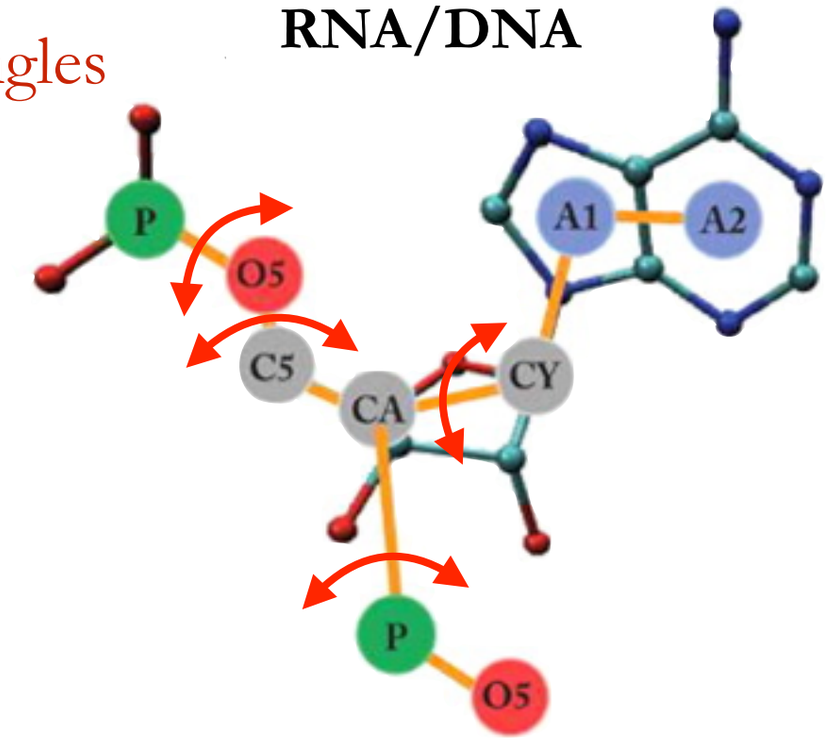
Protein



All-atom or CG representation

dihedral angles

RNA/DNA



HiRE-RNA representation

Energy minimisation

$$\frac{\partial E}{\partial q_i}$$

Internal normal mode analysis

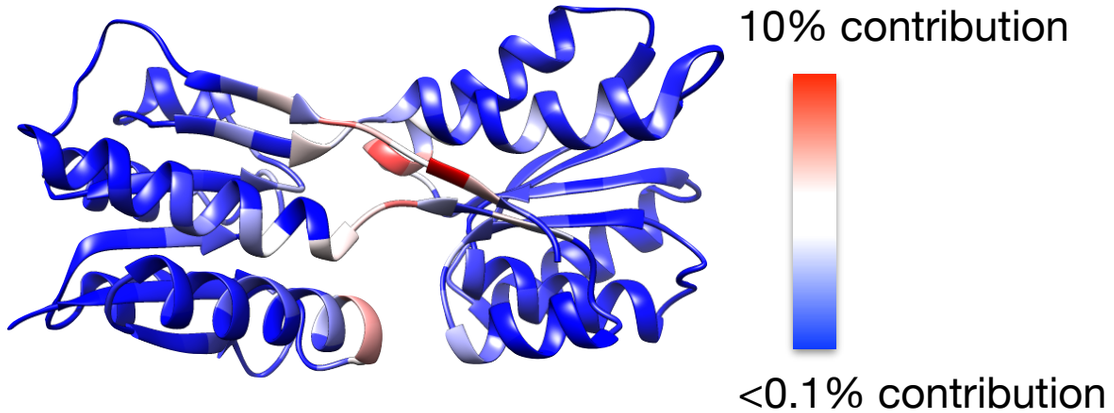
Technical caveat:
Conversion from internal to cartesian coordinate space (non linear)

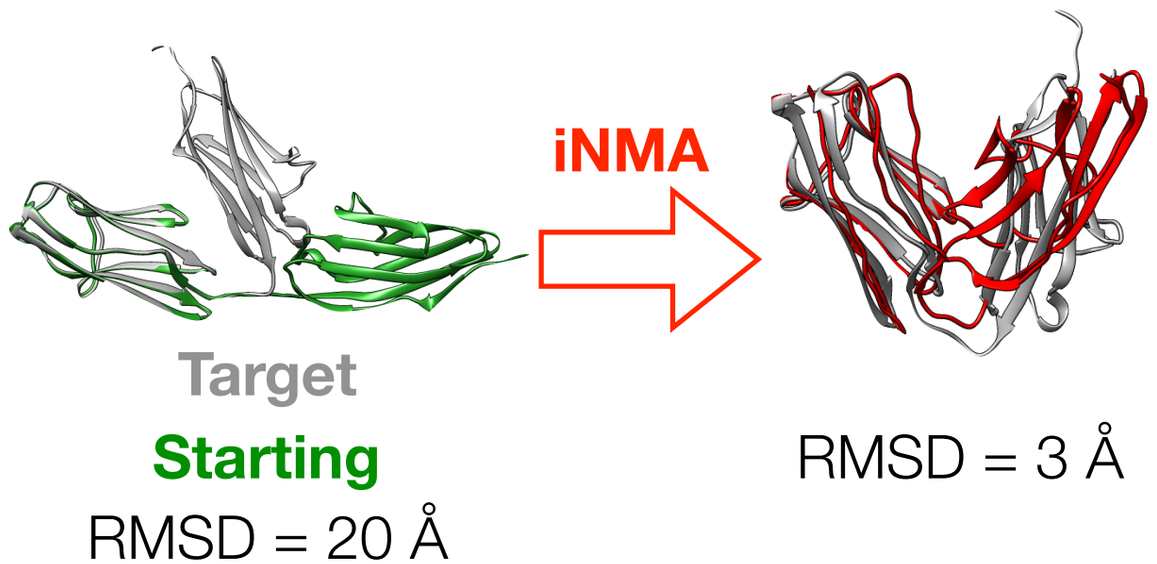
immediate future

Internal Normal Mode Analysis

Advantages

- Faster and more harmonic exploration
- Better sampling for large conformational changes
- Determination of torsions implied in the global movements
- Conformational changes better described by the lower frequency modes (<5)
- No deformation of the structure, but large conformational changes





Frezza and Lavery JCTC 2015

Frezza and Lavery in preparation

immediate future

Internal Normal Mode Analysis

Applications

- Sampling methods
- Prediction of candidate structures for docking experiments
- Prediction of RNA structure by combining SAXS data and MD
- Parametrization and optimisation of a coarse-grained force-field

Acknowledgements

Philippe Derreumaux
LBT, Paris 7

HiRE-RNA

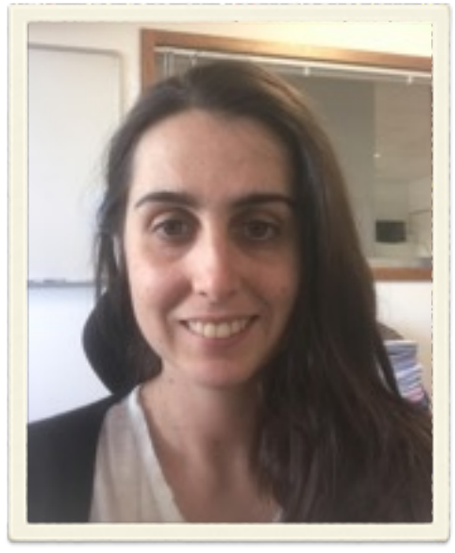
Marc Baaden
LBT, CNRS

UnityMol

Fernando LB Da Silva
University of Sao Paolo

Titration

Elisa Frezza
LCRB



Internal coordinates

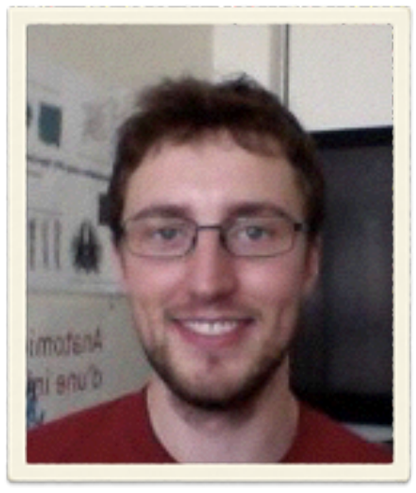
Tristan Cragnolini
Post-doc Cambridge

HiRE-RNA, v2 & v3

Liuba Mazzanti
Post-doc Cambridge

HiRE-RNA + SAXS

Sébastien Doutreligne
grad student



UnityMol

J. Sponer's group

D. Wales's group