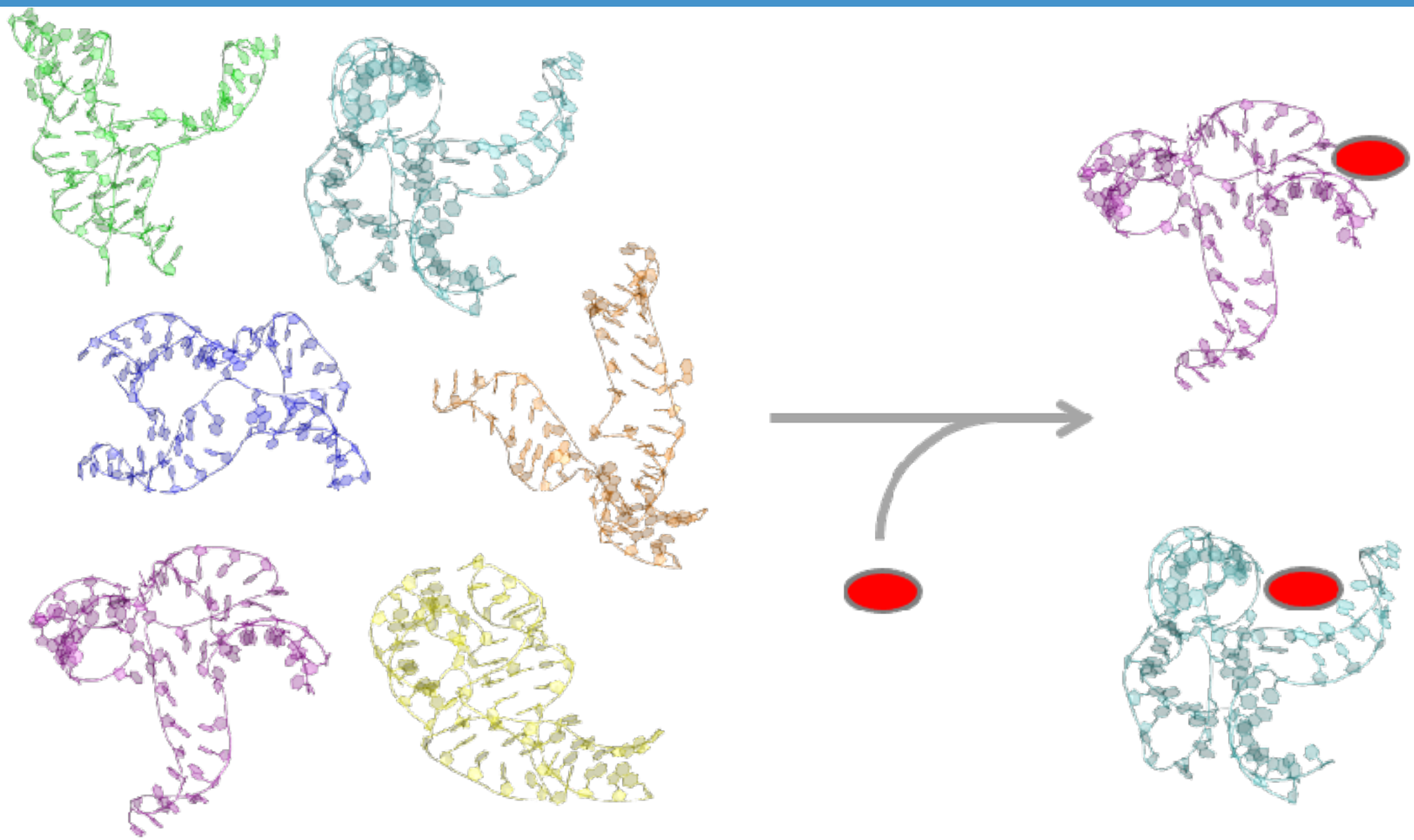
Dissecting RNA dynamics combining molecular simulations and solution experiments

Giovanni Bussi
Molecular and Statistical Biophysics
SISSA, Trieste, Italy

bussi@sissa.it
http://people.sissa.it/~bussi
twitter.com/bussilab

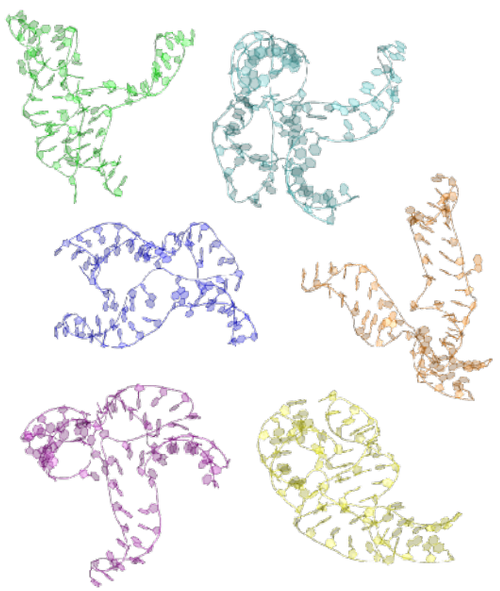
RNA structural dynamics

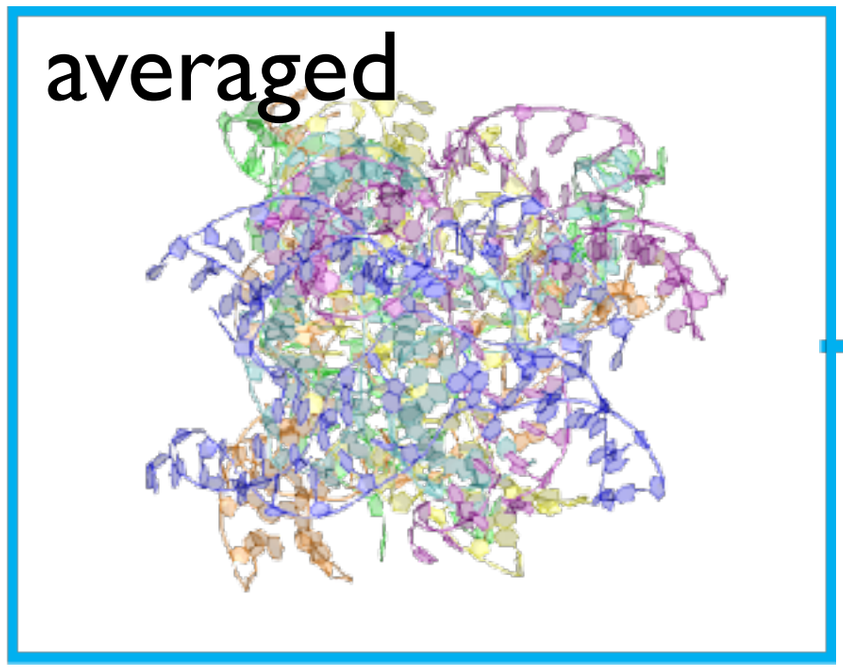


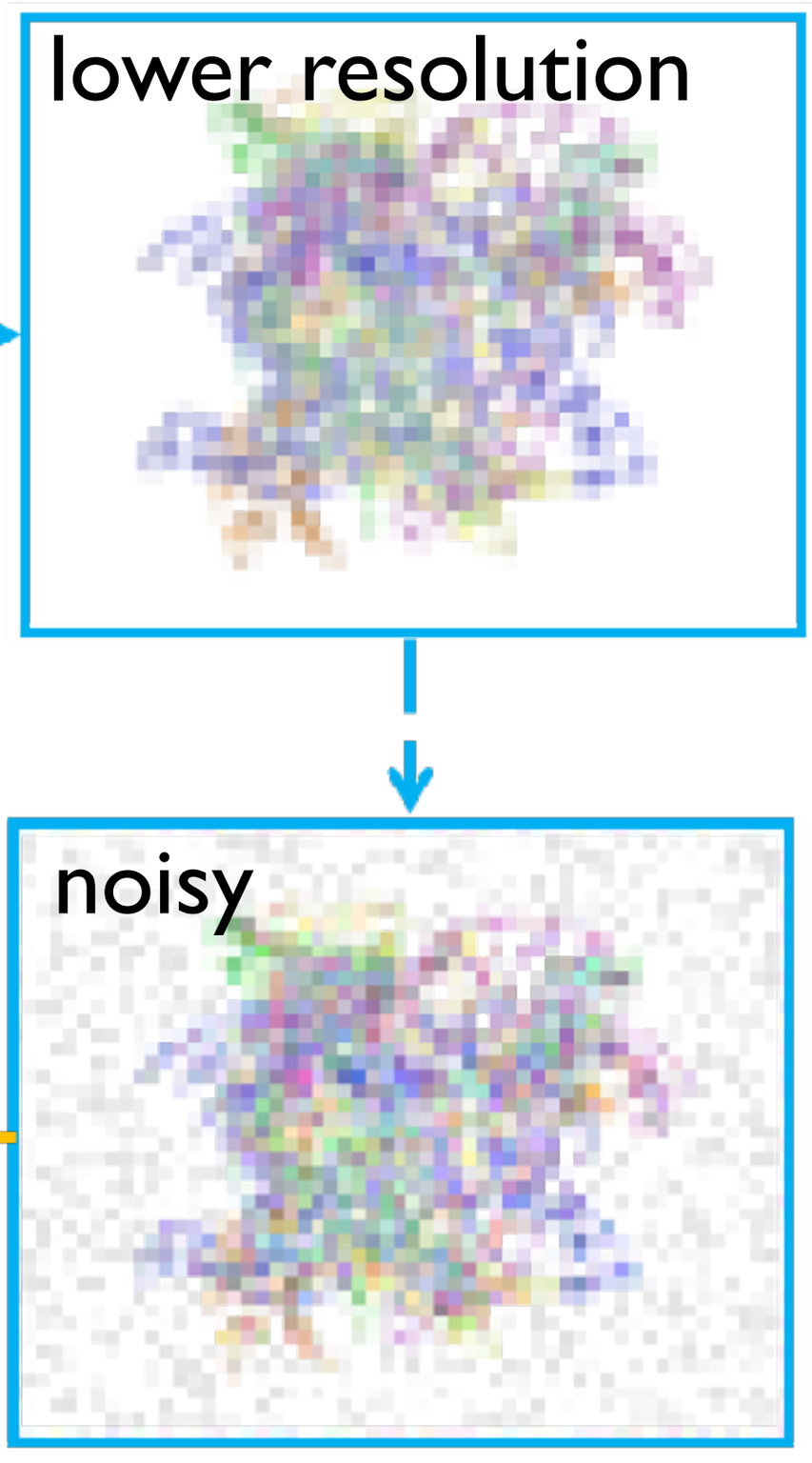
Conformational selection or induced fit
Crucial when interacting with proteins, ligands, ions, etc.

Bernetti and Bussi, COSB (2023)

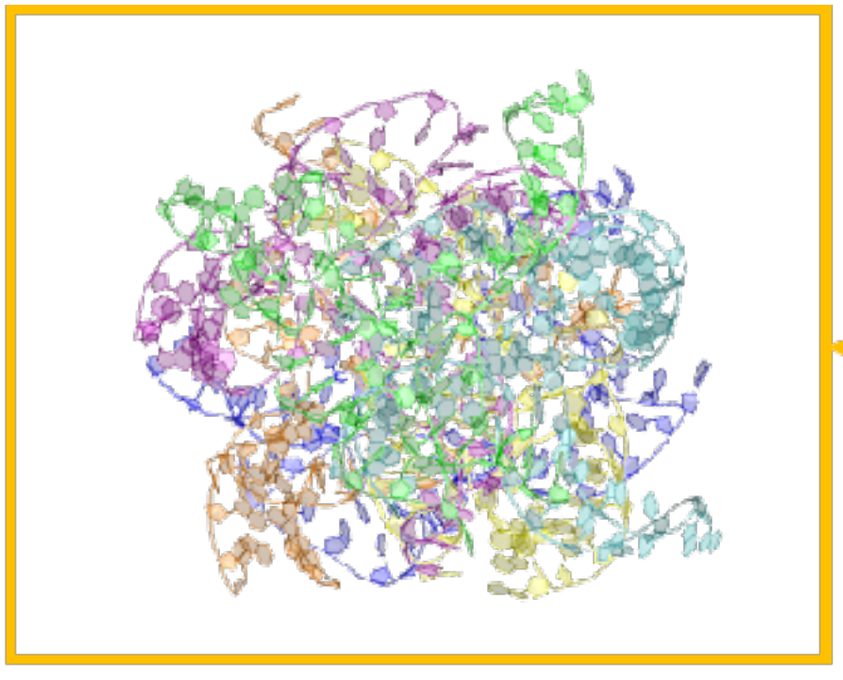
Dynamics from experiments

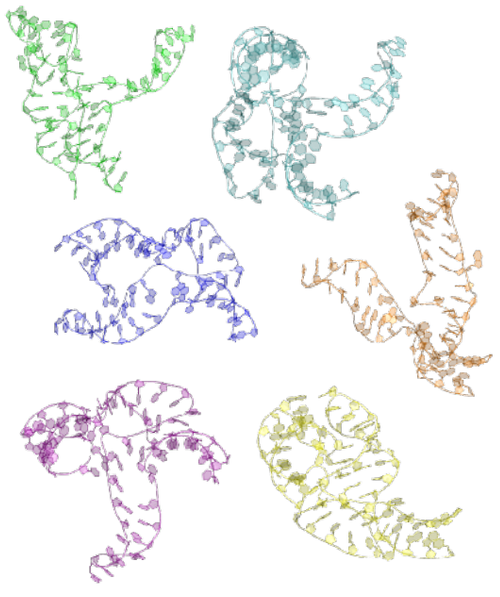






?





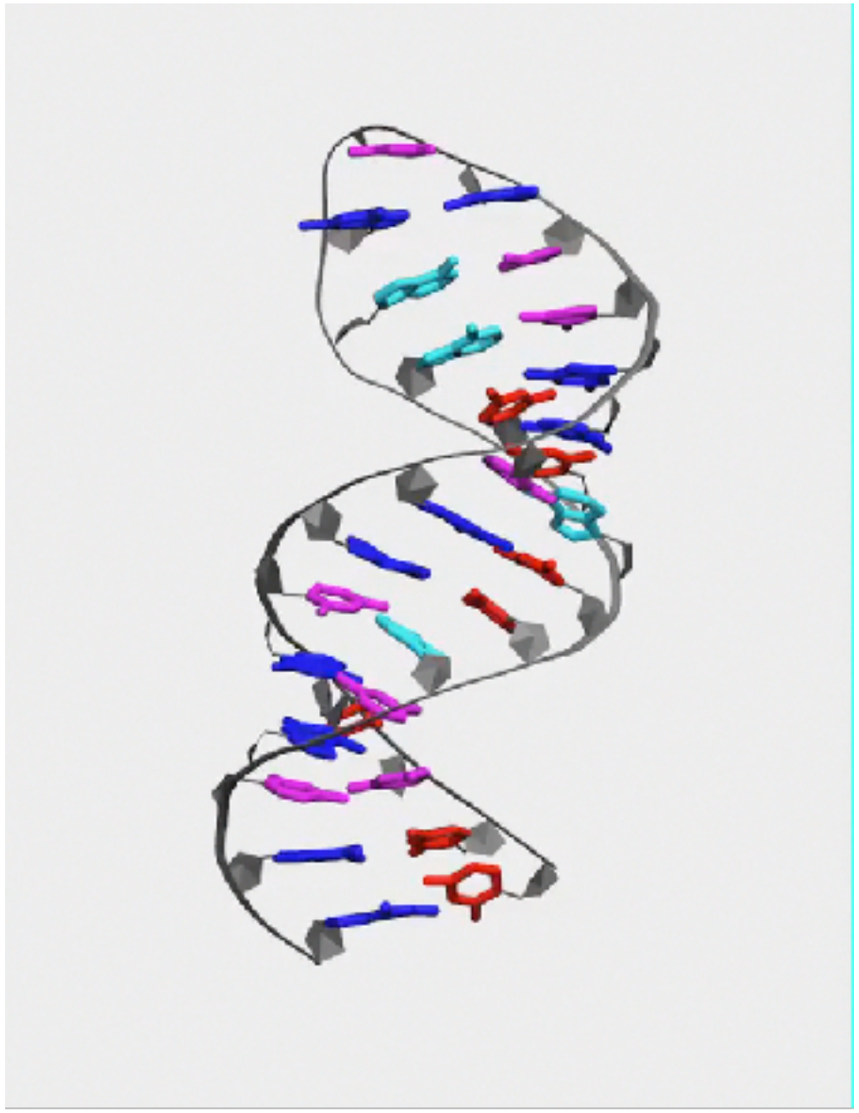
Bernetti, and Bussi, COSB (2023)

Molecular dynamics

$$E_{\text{total}} = \sum_{\text{bonds}} k_{\text{b}} \left(\ell - \ell_0\right)^2 + \sum_{\text{angles}} k_{\text{a}} \left(\theta - \theta_0\right)^2$$

$$+ \sum_{\text{torsions}} \frac{1}{2} V_{\text{n}} [1 + \cos(n\omega - \gamma)]^2$$

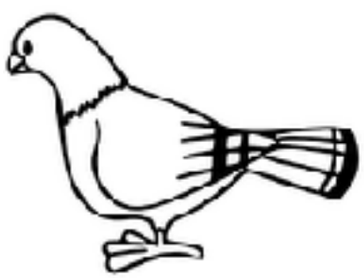
$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \varepsilon_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}$$



Empirical force field*:

- Chemically motivated interactions
- Atomistic details
- Explicit water and ions
- No polarization
- No chemical reactivity

Approx ~50-500 ns/day



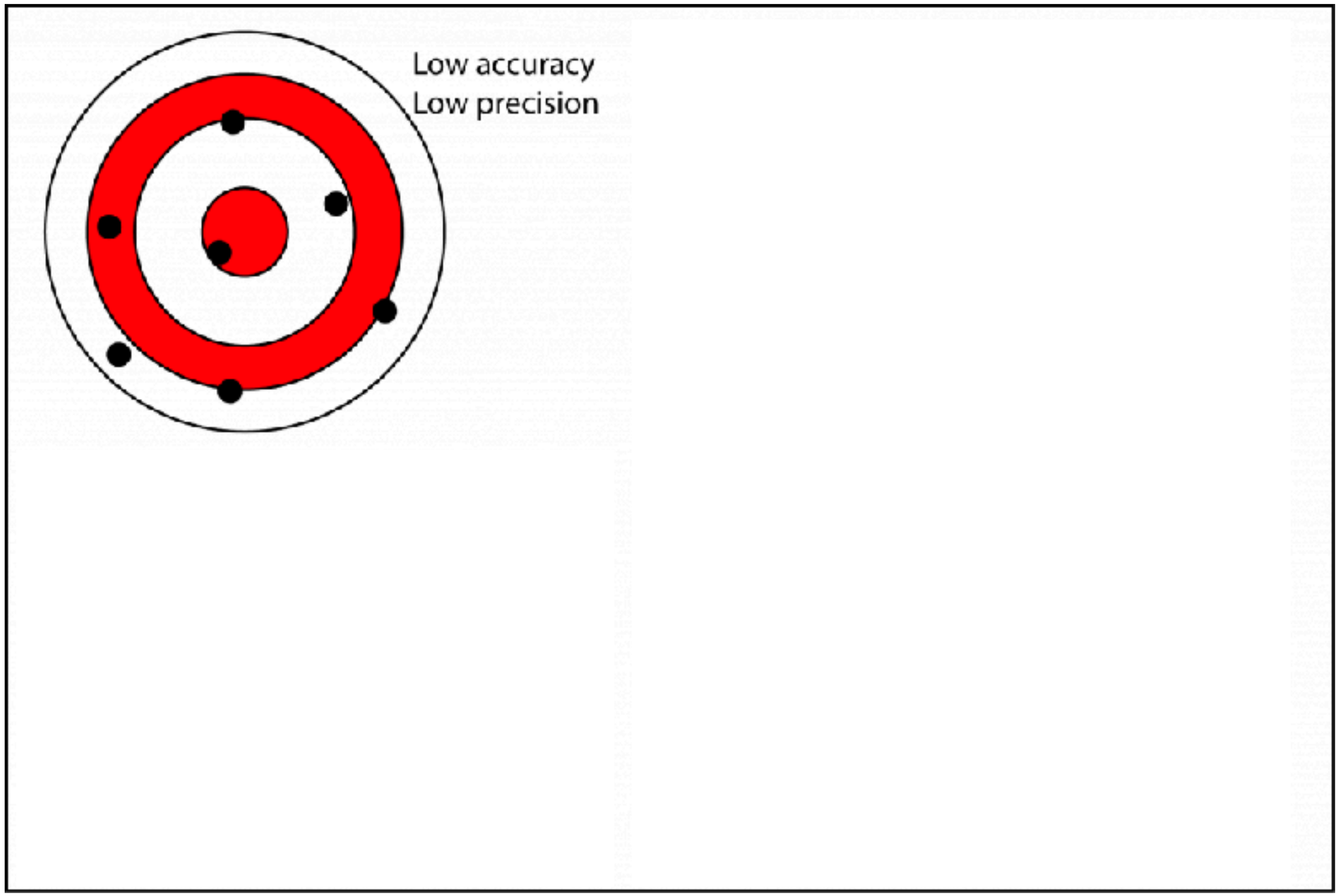
*AMBER (ff99+parmbsc0+ChiOL3+TIP3P or OPC)

gromacs.org + plumed.org

Accuracy and precision

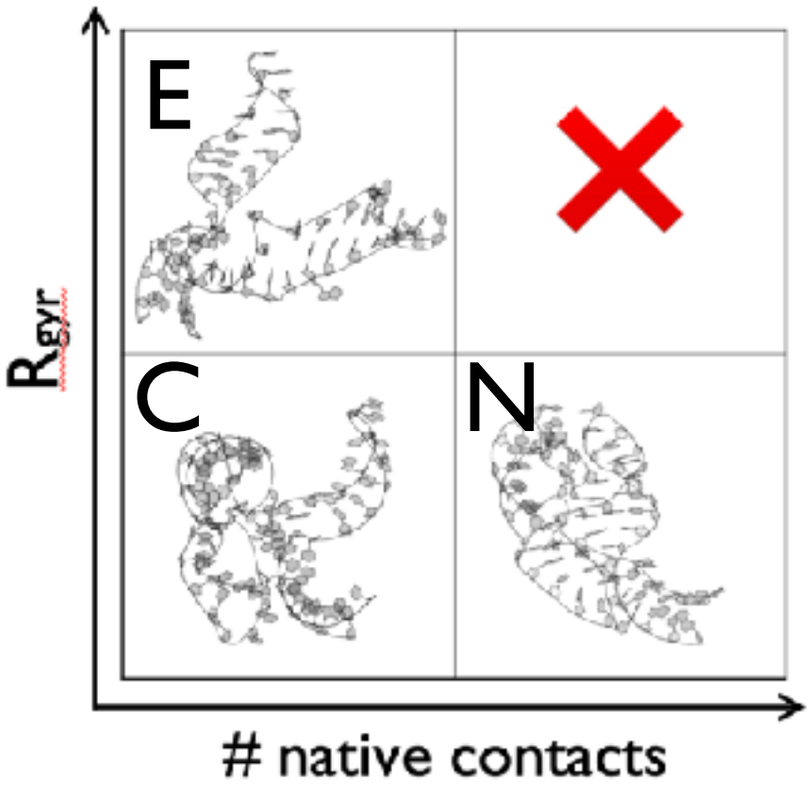
Enhanced sampling*

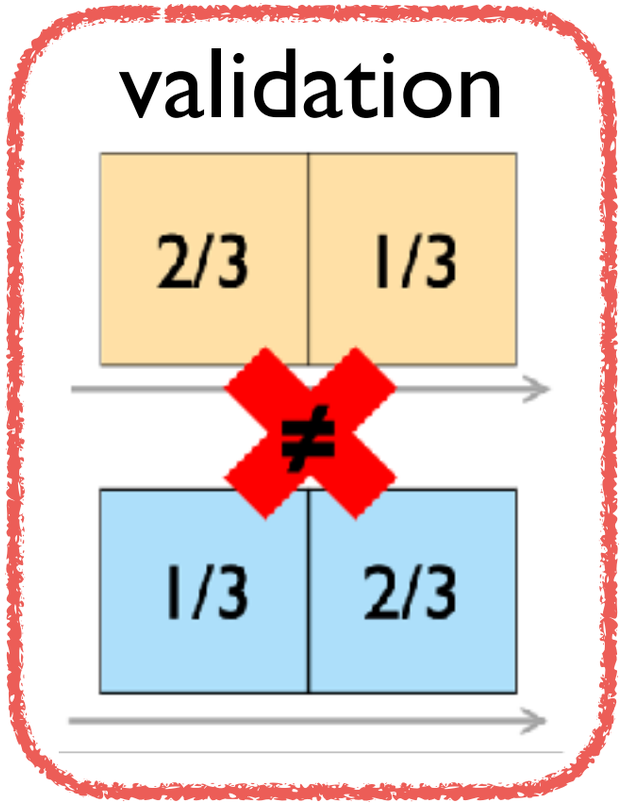
Combining with experimental data

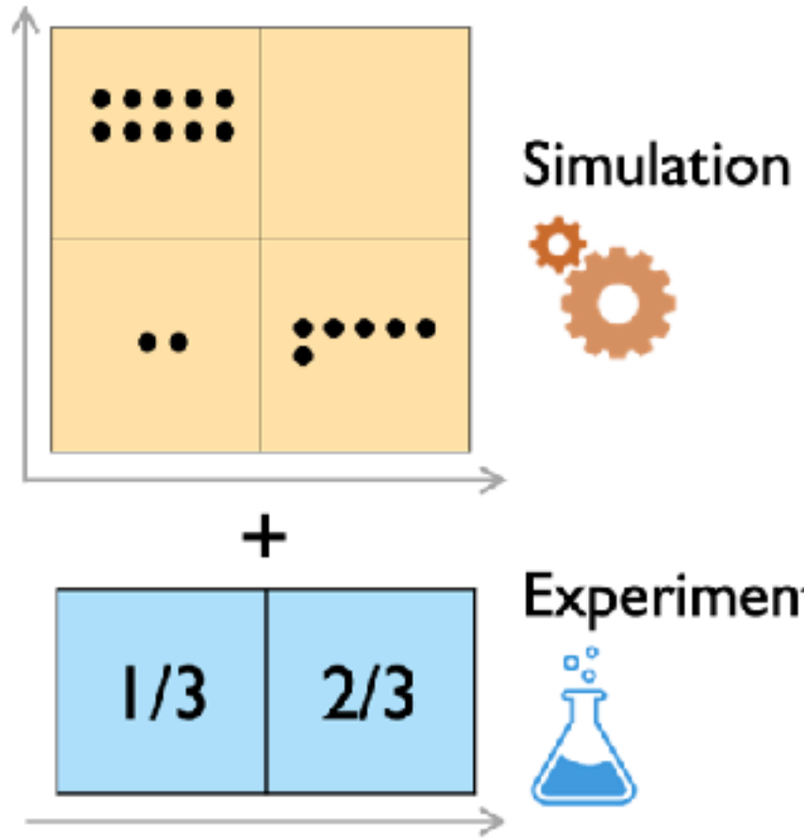


*Vast array of methods, Henin et al Living J. Comp. Mol. Sci. 2022
See also plumed.org/masterclass

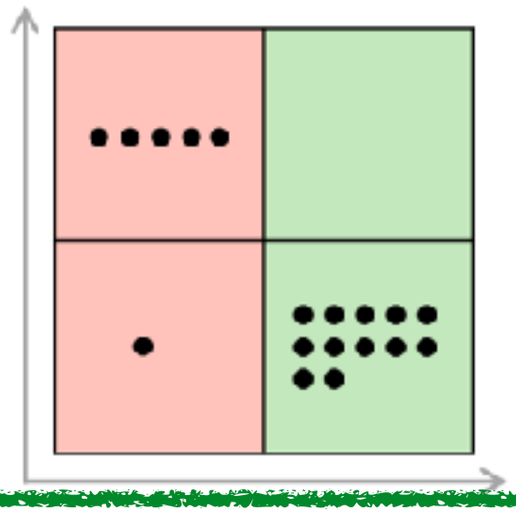
Combining simulations and experiment



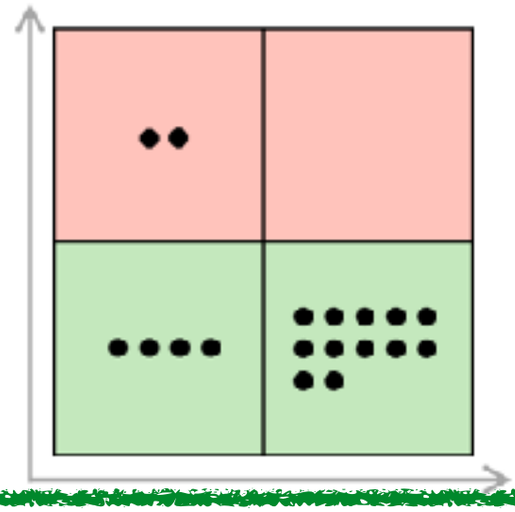




ensemble refinement with MaxEnt

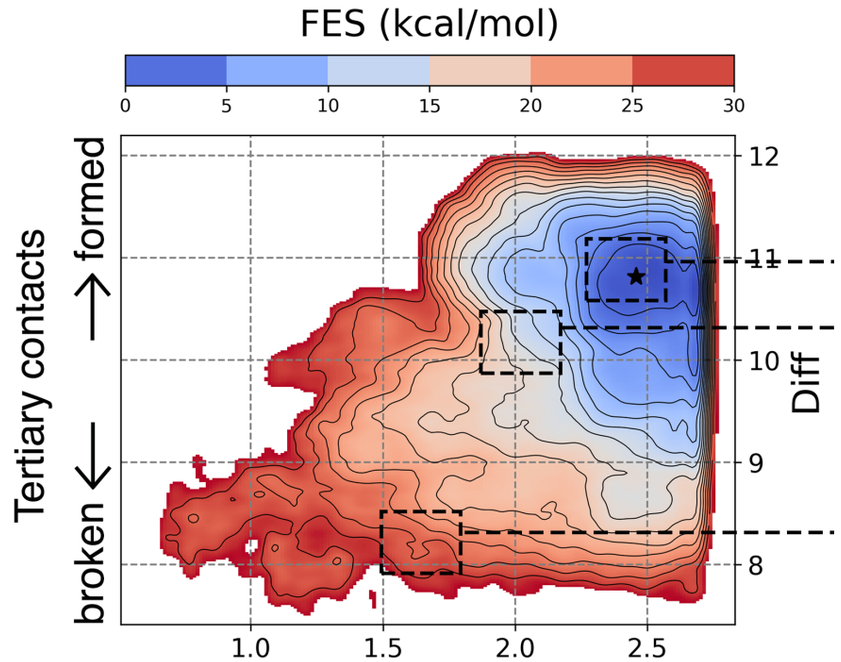


transferable force-field fitting

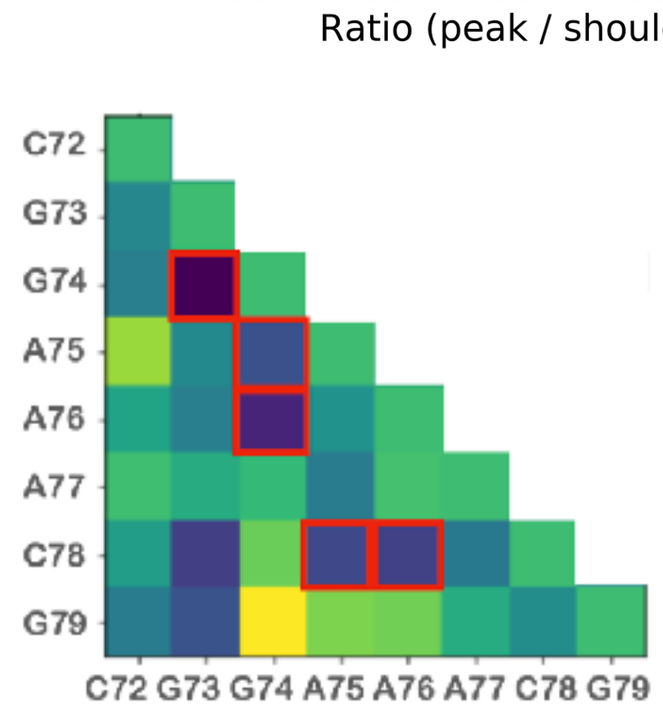


Agenda

GAC-RNA ensembles from MD and SAXS data#

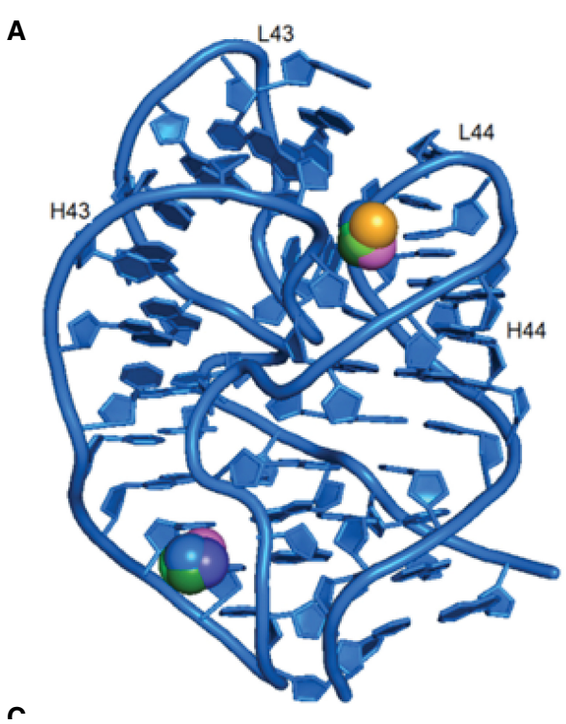


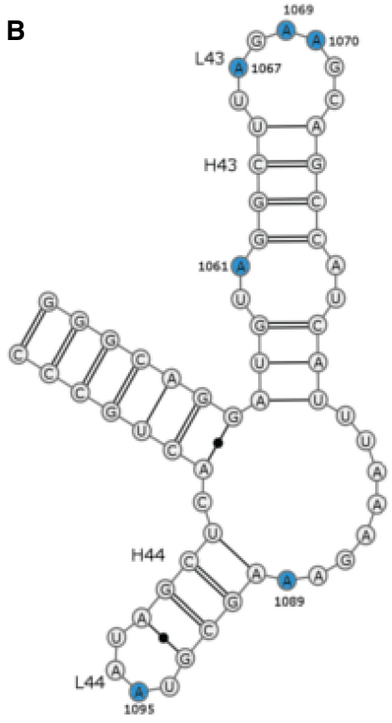
Cooperative effects in chemical probing experiments*

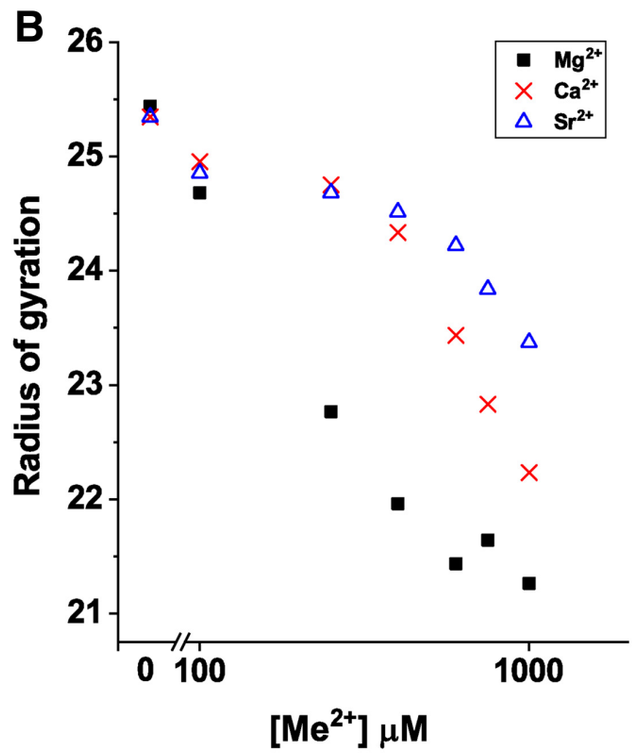


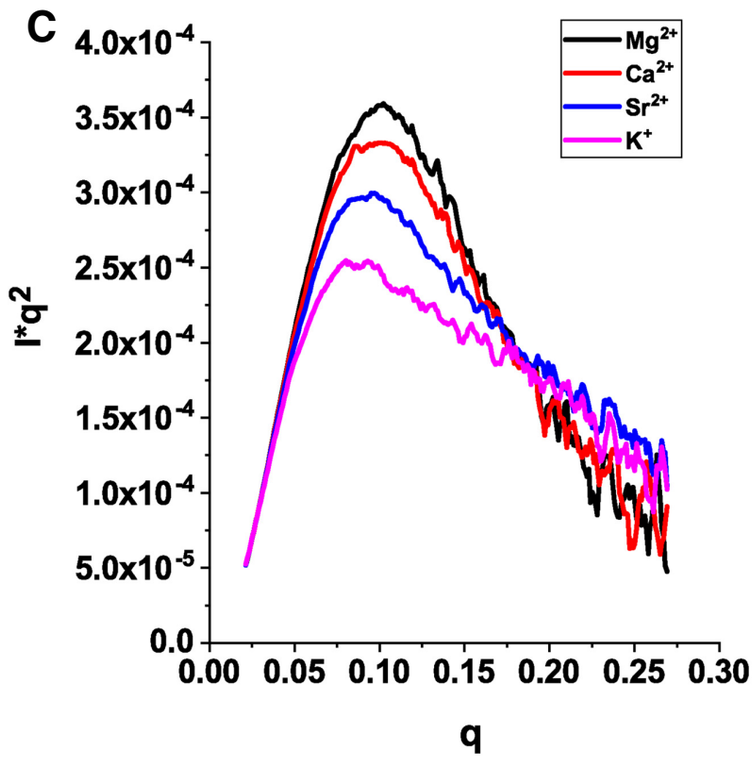
#Bernetti, Hall, and Bussi, NAR (2021) + Bernetti and Bussi EPJB (2021)
*Calonaci et al arXiv 2022

GTPase center (GAC)









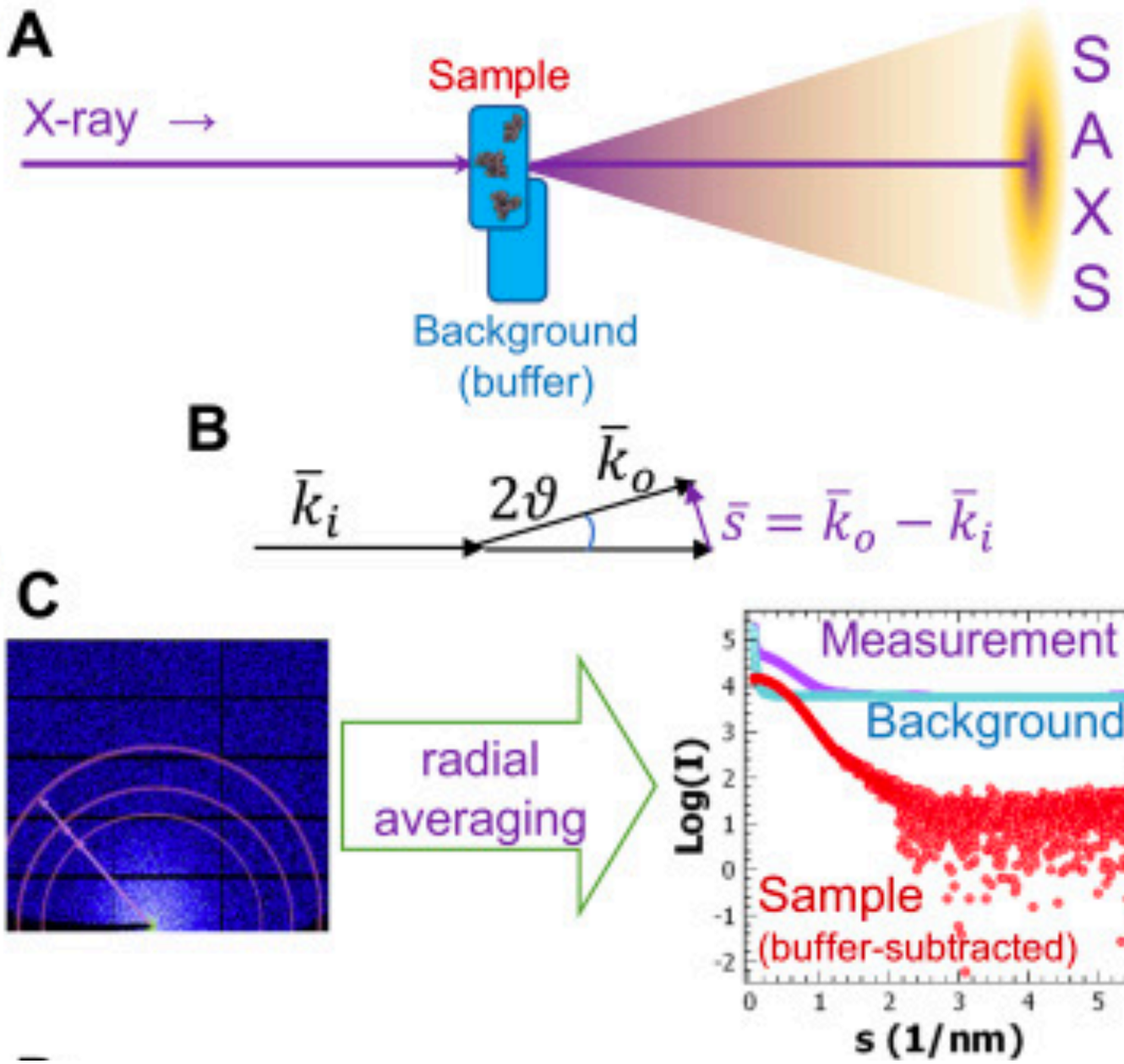
Domain folding is regulated by ions

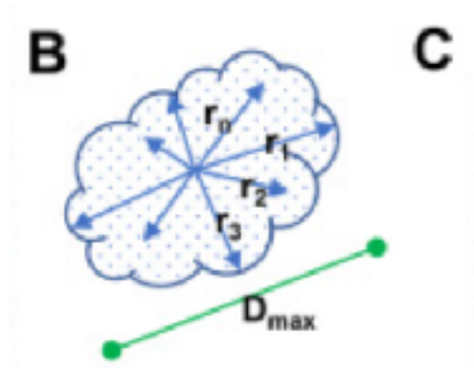
SAXS data shows compaction $K^{+} \rightarrow Sr^{2+} \rightarrow Ca^{2+} \rightarrow Mg^{2+}$

Data from Welty et al RNA (2018)

See also Welty et al JMB (2020) - RNA-only crystal structure

Small-angle X-ray Scattering





$$R_g^2 = \frac{\int_V \Delta\rho(\bar{r})\bar{r}^2 d\bar{r}}{\int \Delta\rho(\bar{r}) d\bar{r}}$$

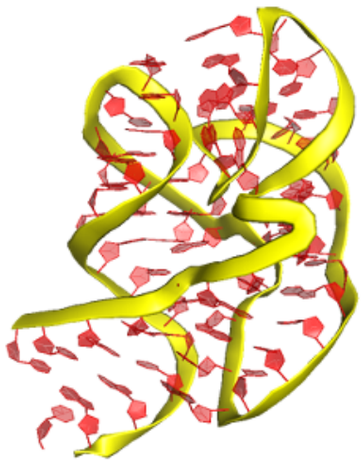
"Gross" information about shape

Require good molecular modeling

Da Vela and Svergun, COSB (2020)

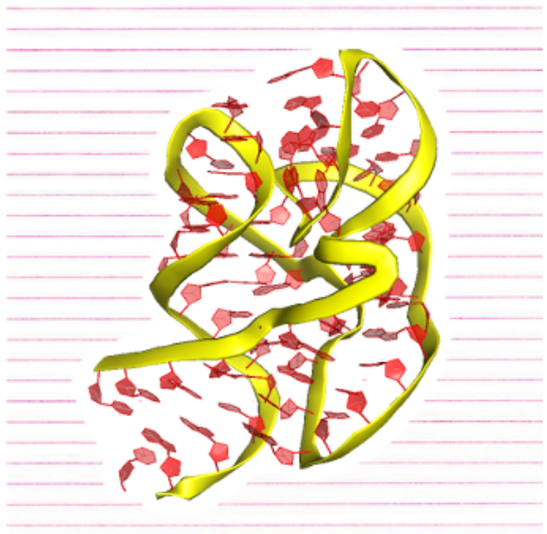
Implicit vs explicit solvent SAXS

Solute only



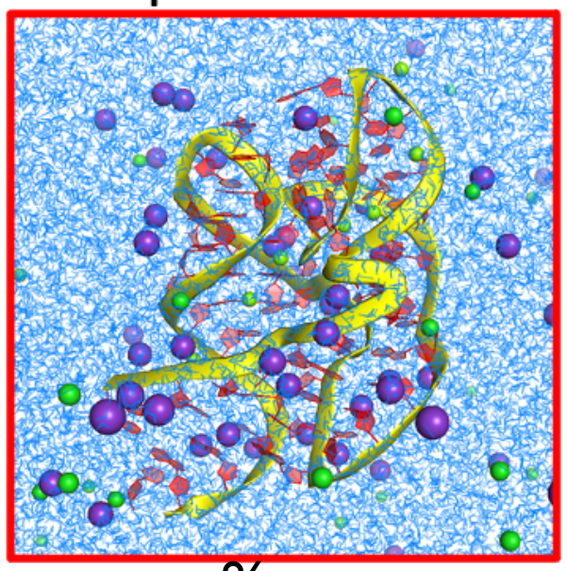
• PLUMED *

Implicit solvent



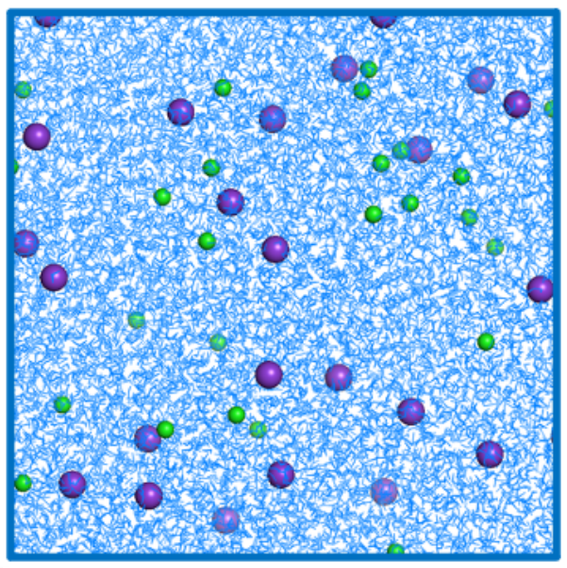
• CRYSOL &

explicit solvent



I_S

• WAXSIS % • CAPRIQORN #



I_B

Buffer subtraction: I_S - I_B

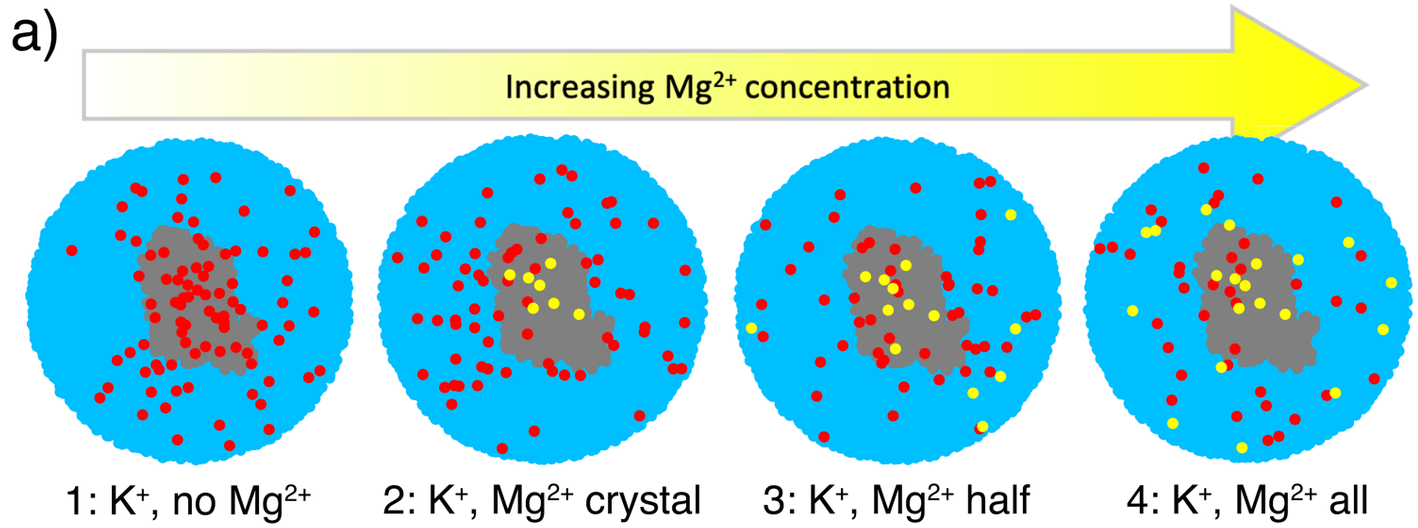
(closer to experiments)

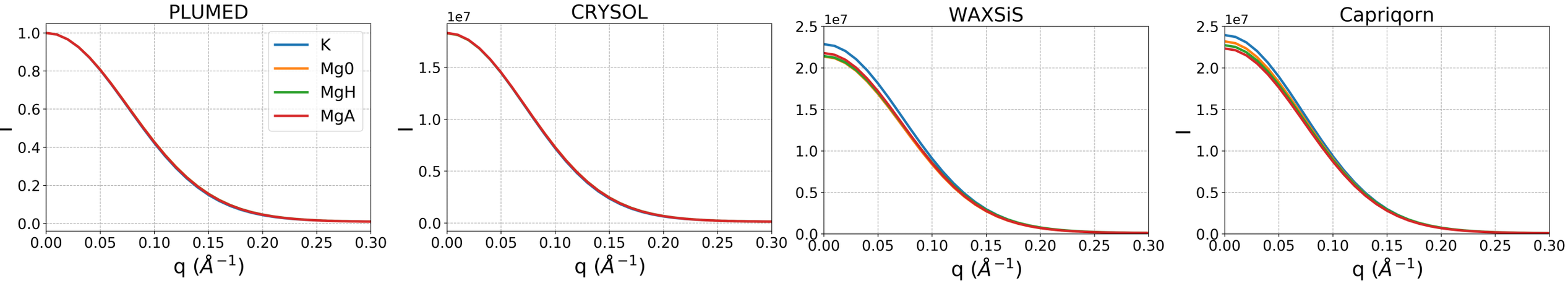
*Bonomi & Camilloni, Bioinf (2017)
&Svergun et al, JAC (1995)
%Knight & Hub, NAR (2015)
#Köfinger & Hummer, PRE (2013)

Fast, approximate

Slow, accurate

GAC: little dependence on ions

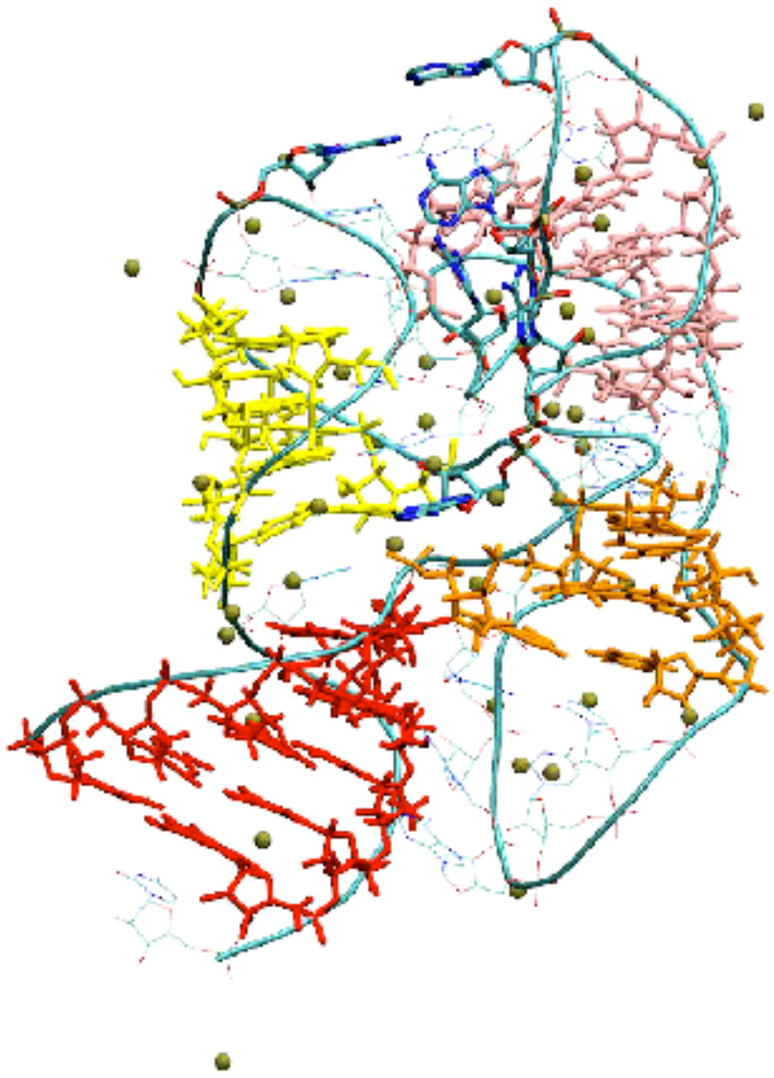


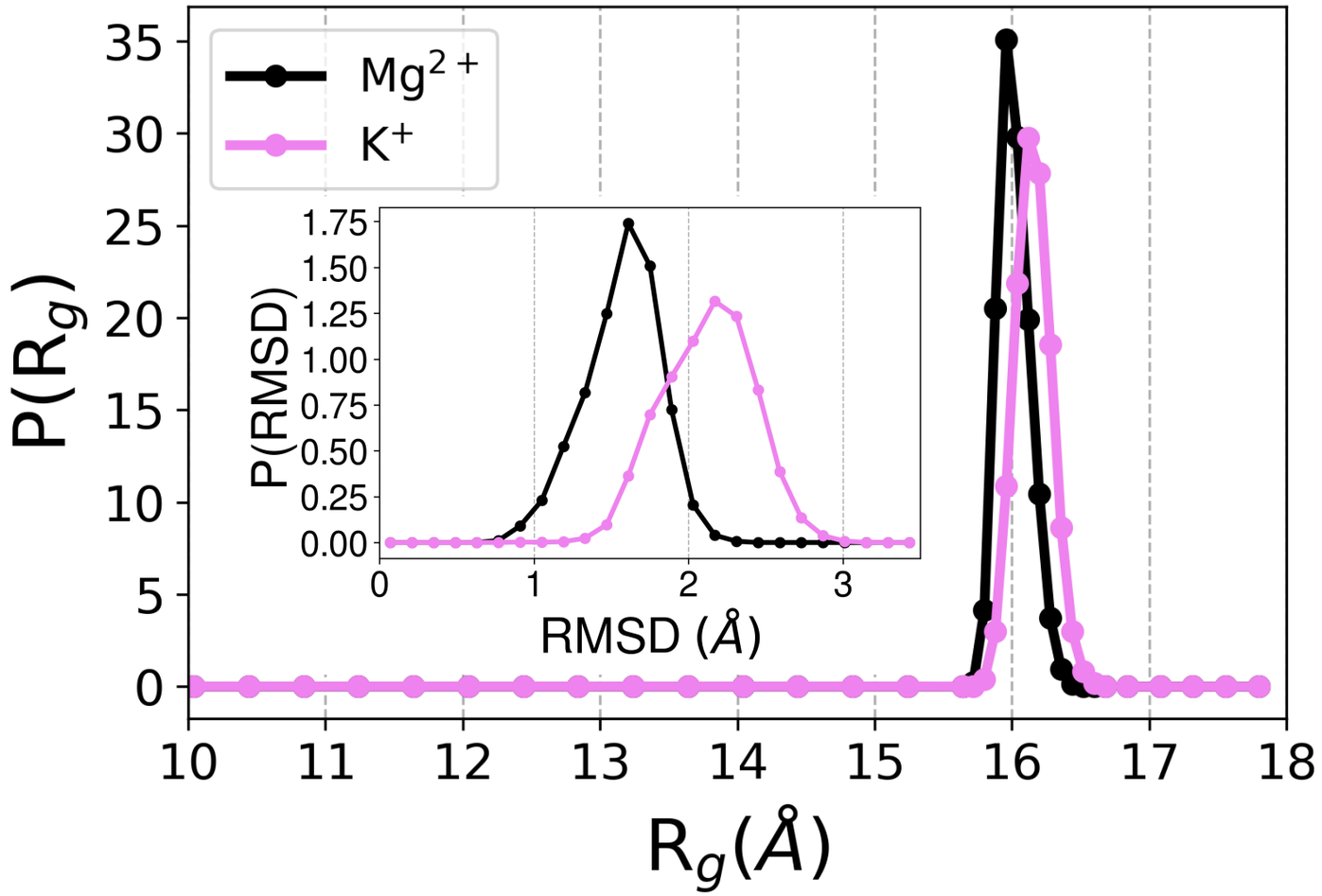


AMBER14 FF + OPC + Joung-Cheatam and Allner et al ions, RNA restrained to native

Bernetti and Bussi, EPJB (2021)

“Long” time scales: nothing happens

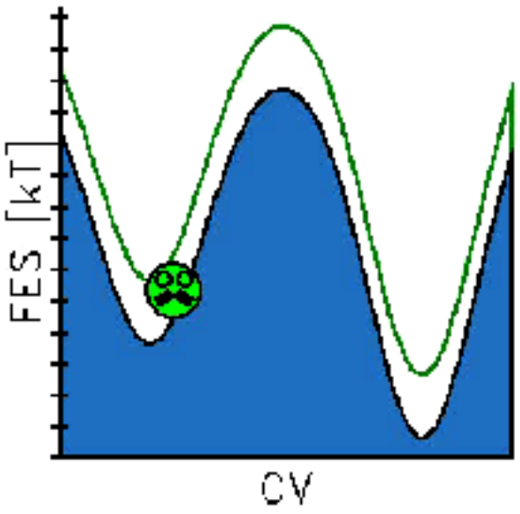




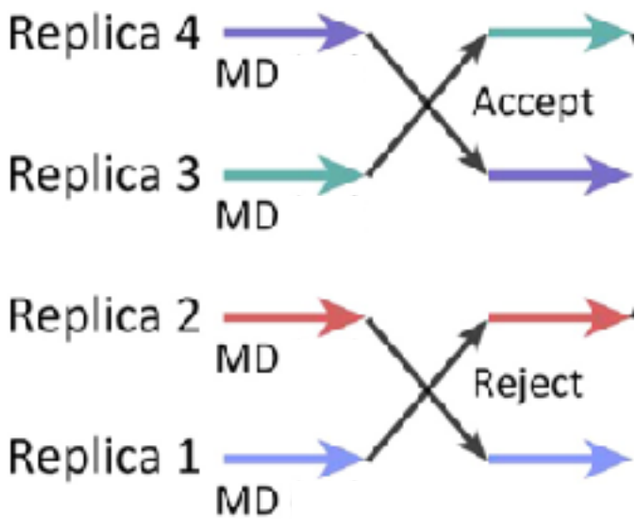
1 μs-long free MD with K^+ vs Mg^{2+}
Virtually identical trajectories

Bernetti, Hall, and Bussi, NAR (2021)

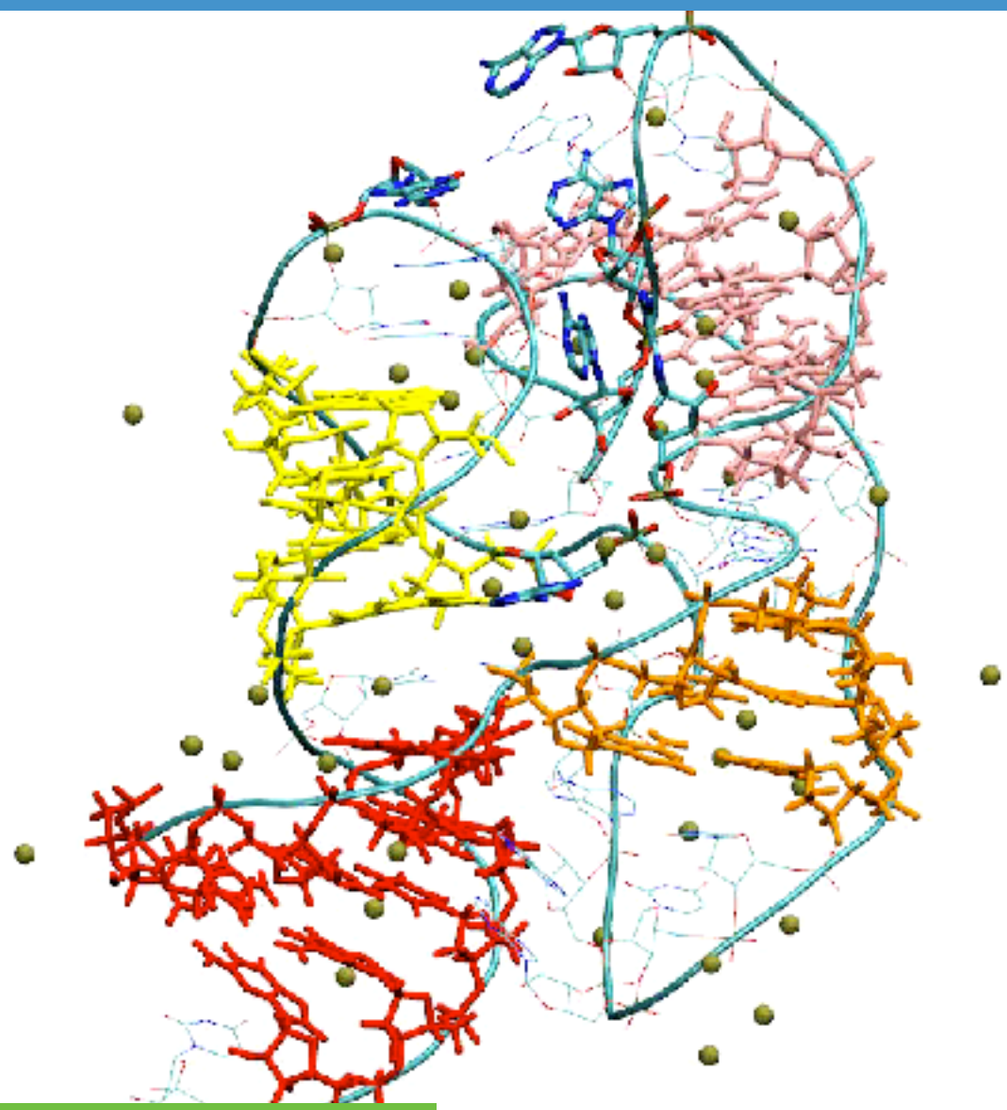
Metadynamics + replica exchange

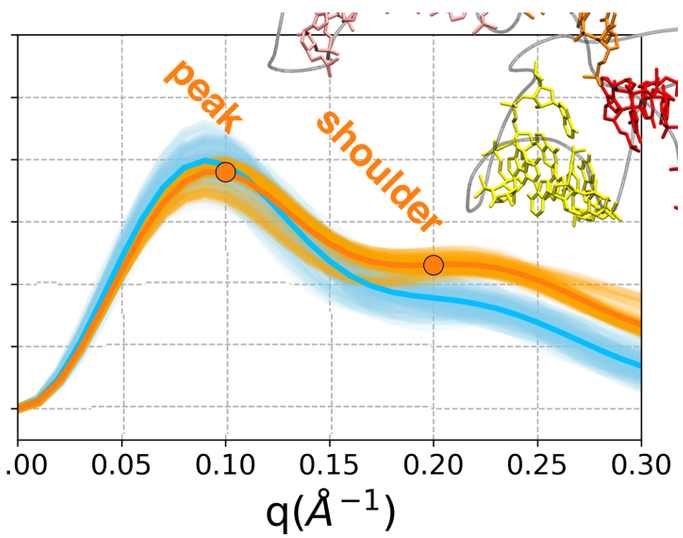


Metadynamics:
“Ratio” peak/shoulder
% non-2D contacts



"Solute tempering"
heating non-2D
structure
32 replicas



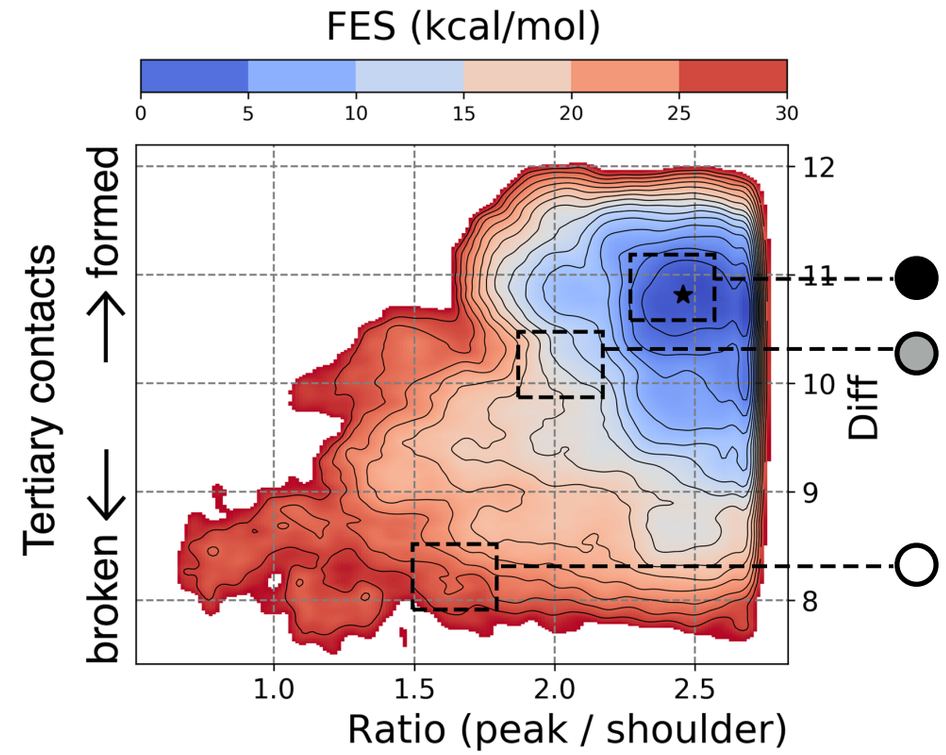


SAXS intensities computed on-the-fly
using a Martini-bead representation*
No exp. data used at this stage.
*Paissoni et al, JAC (2019)

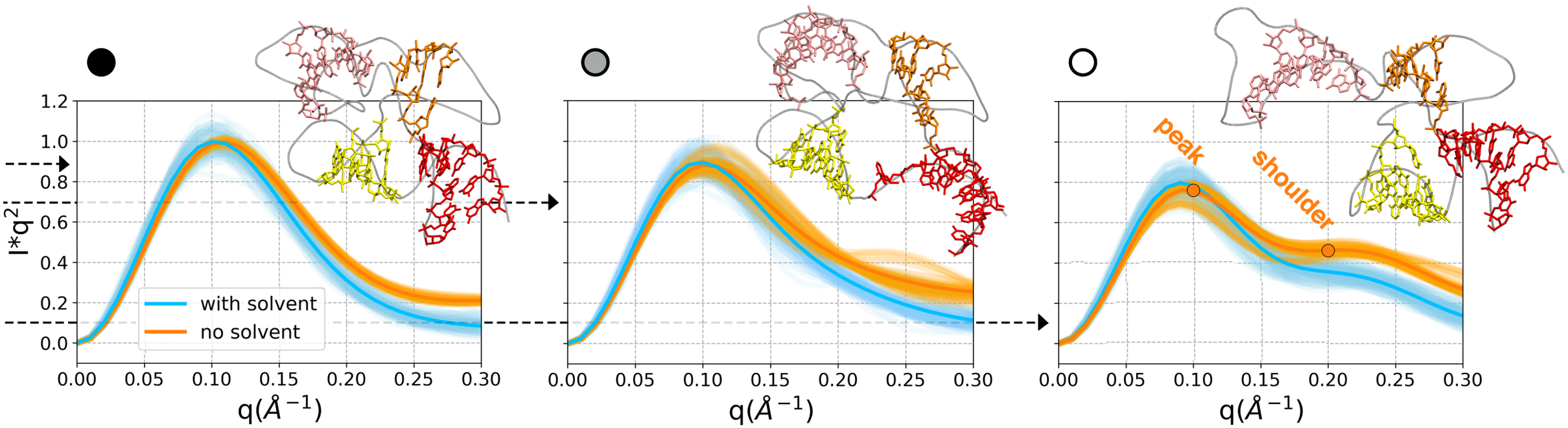
No Mg^{2+}!

Bussi and Laio, Nat Rev Phys (2020)
Bussi Mol Phys (2014)

Extracting structures in different regions



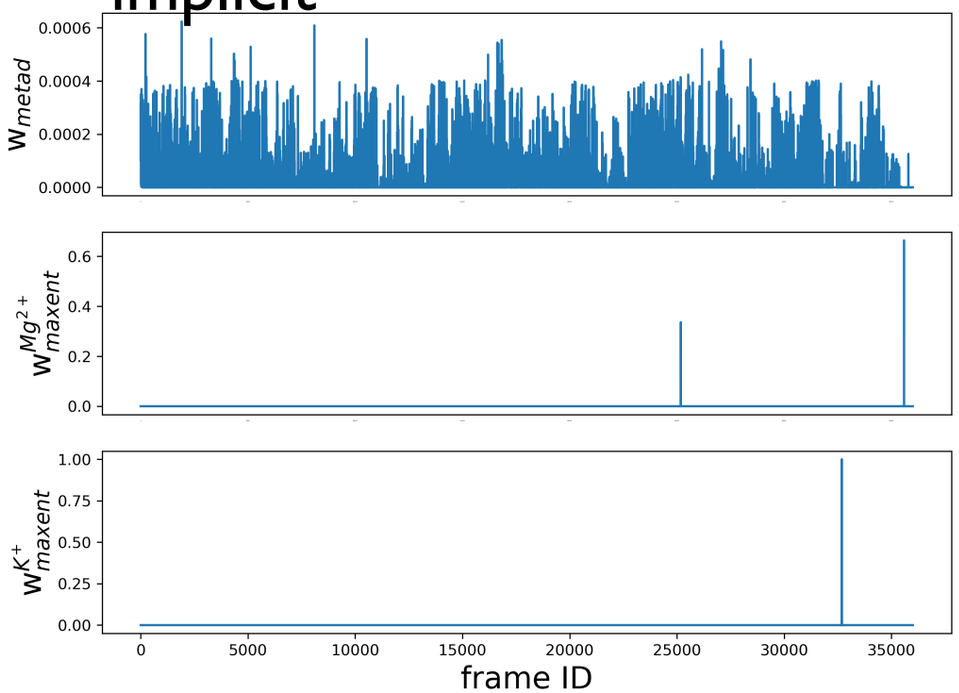
Implicit (PLUMED/MARTINI) SAXS useful to enhance sampling, but reports spectra different from explicit solvent at q ~ 0.2 A^{-1}



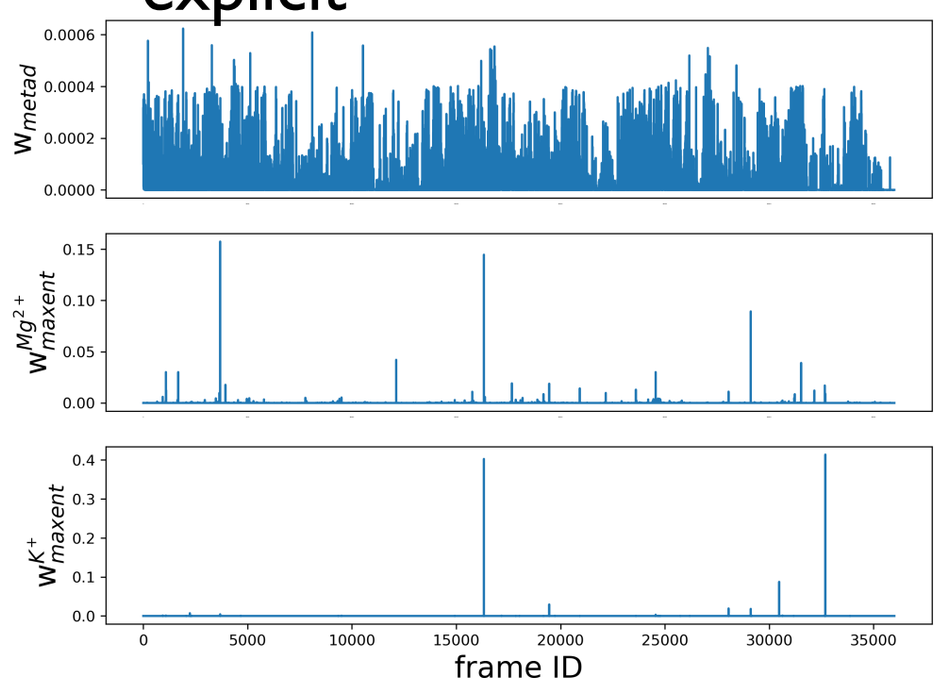
Bernetti, Hall, and Bussi, NAR (2021)

Reweighting (implicit vs explicit)

implicit



explicit

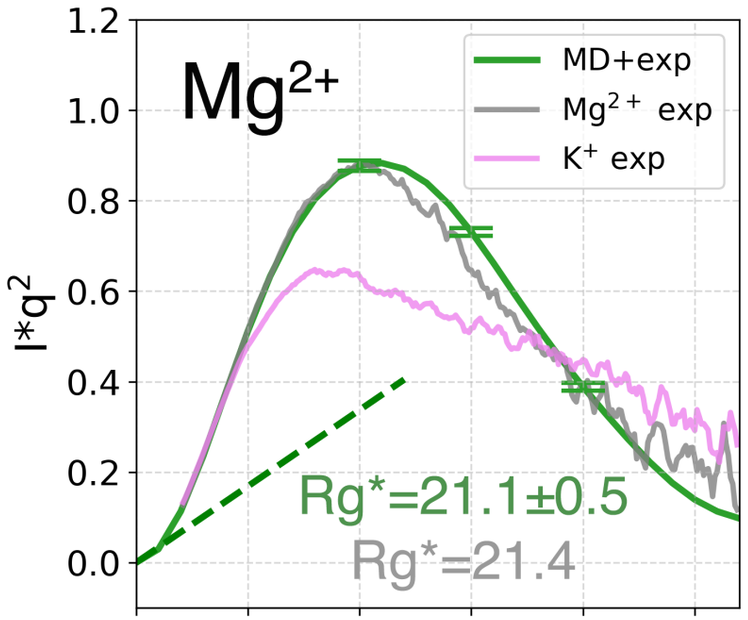


	Ratio_{exp}	$R^*_{g,exp}$	Ratio_{solv}	$R^*_{g,solv}$	Ratio_{solute}	$R^*_{g,solute}$
Mg^{2+}	2.25	21.35	2.26	21.12	1.62	21.17
K^+	1.43	25.17	1.44	24.92	0.67	22.54

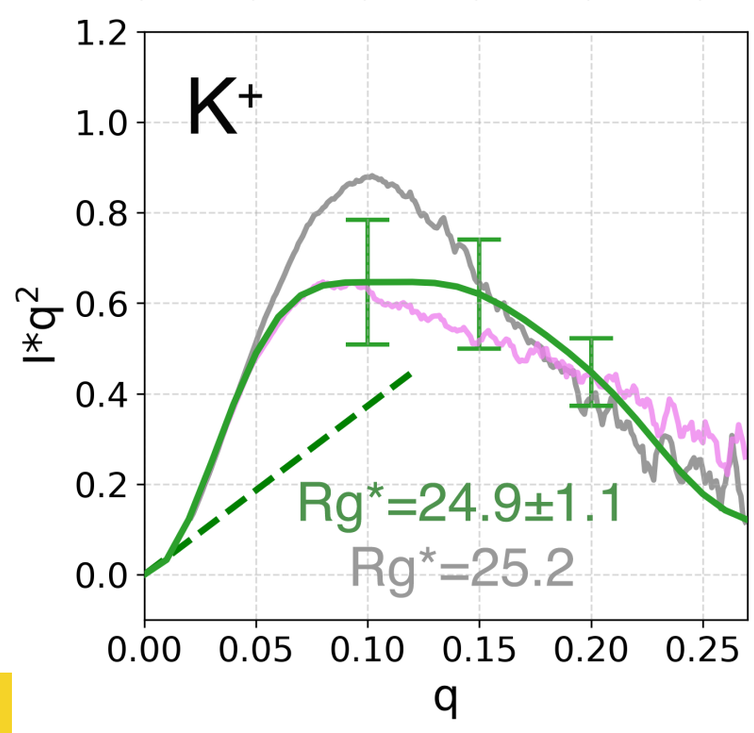
Reweighting to match experiments with implicit solvent SAXS does not work
(no way to reproduce experimental spectra)

SAXS spectra from reweighting

Exp. spectrum with Mg^{2+}
~1% extended



Exp. spectrum with K+
~42% extended
(Few extended structures in MD, high statistical error)



Reminder: simulation had no Mg^{2+}!

Bernetti, Hall, and Bussi, NAR (2021)

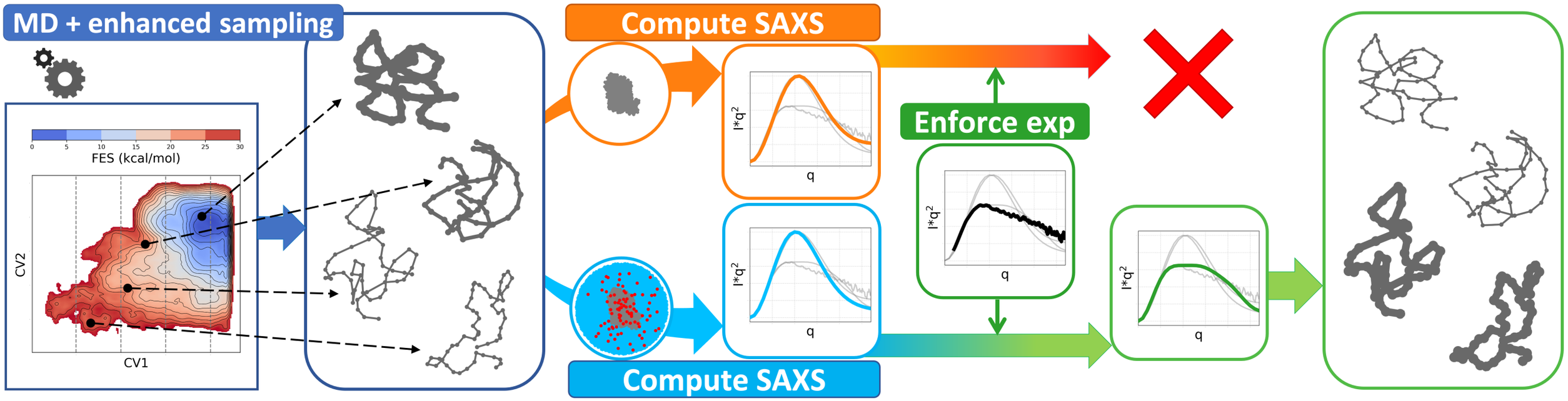
Partial summary

Enhanced sampling (heterogeneity) + MaxEnt (match experiments)

(Fast) implicit solvent SAXS, rough estimates and enhanced sampling
(Slow) explicit solvent SAXS, match experiments

Little impact of ions on SAXS → run with K^+, match K^+ and Mg^{2+}

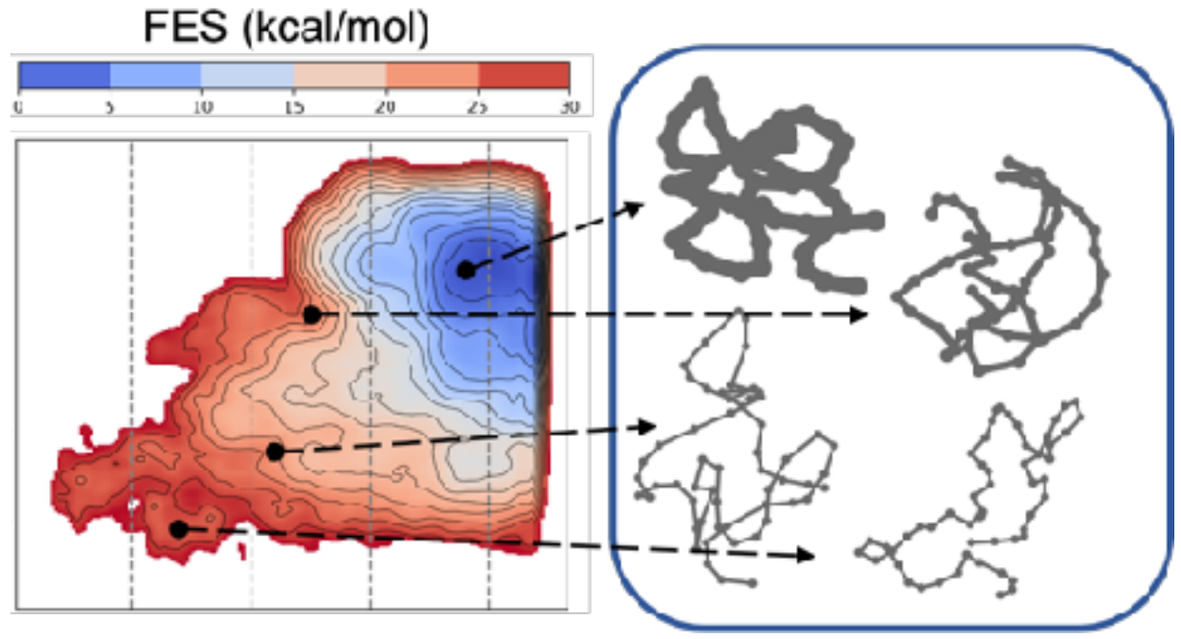
Mg^{2+} → K^+ results in shift in extended population (1% vs 42%)



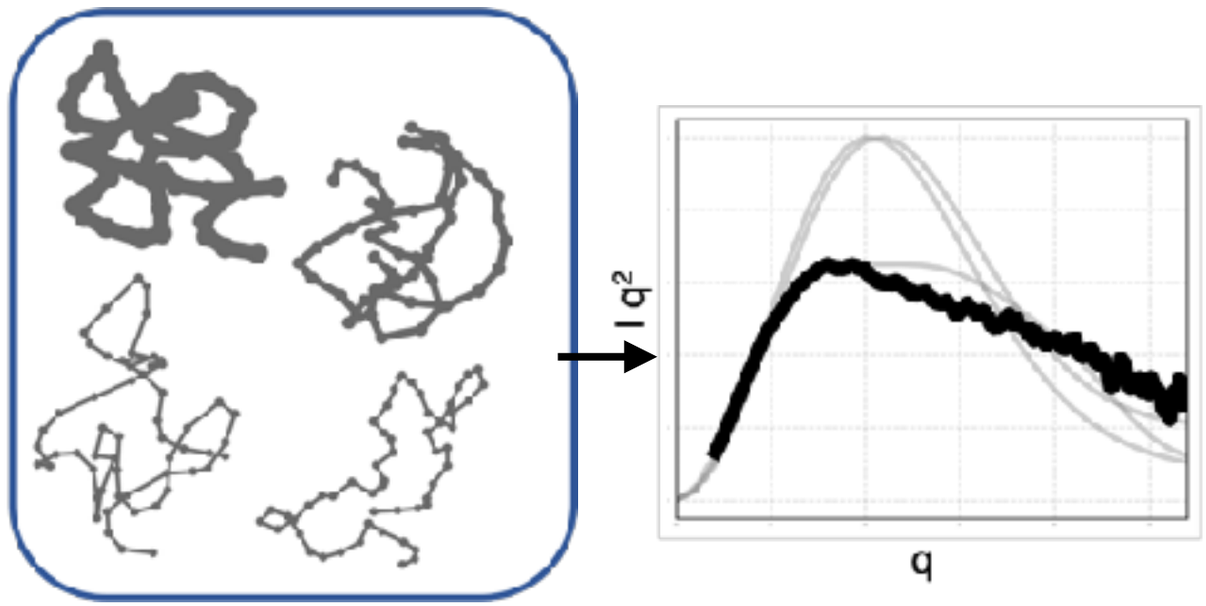
Bernetti and Bussi, EPJB (2021)
Bernetti, Hall, and Bussi, NAR (2021)

Crucial ingredients

A method to generate heterogeneous ensembles (MD with enhanced sampling, etc.)

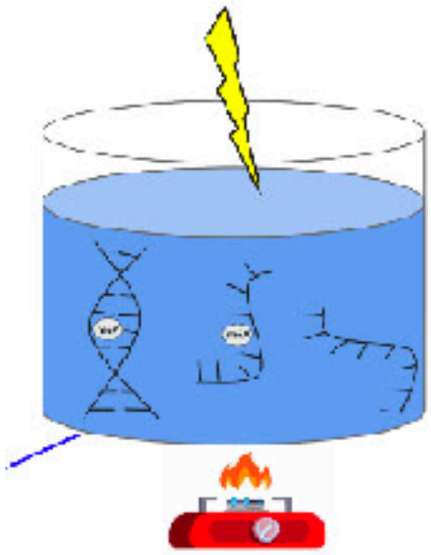


A good “forward model” to back-calculate experiments from ensembles



Other "forward models" we are using

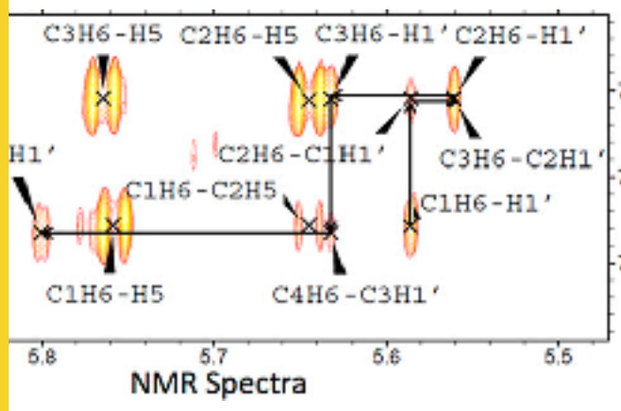
Force field optimization for m6A using melting experiments (V. Piomponi)

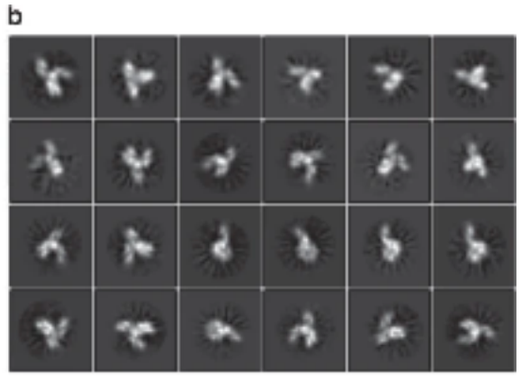


General RNA force field optimization using NMR and thermodynamic data (T. Froehlking, I. Gilardoni, collaboration with J. S

Inosine duplex ensem
(V. Piomponi, in collab
Sattler, Munich)

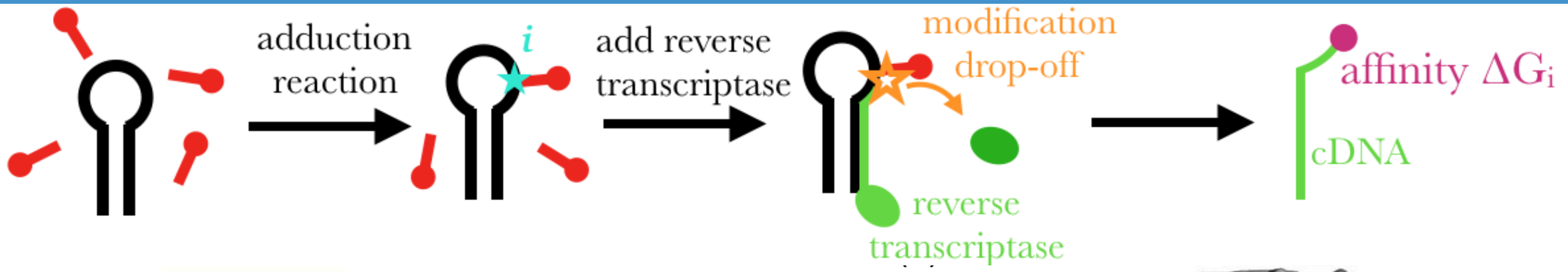
What about chemical probing data?

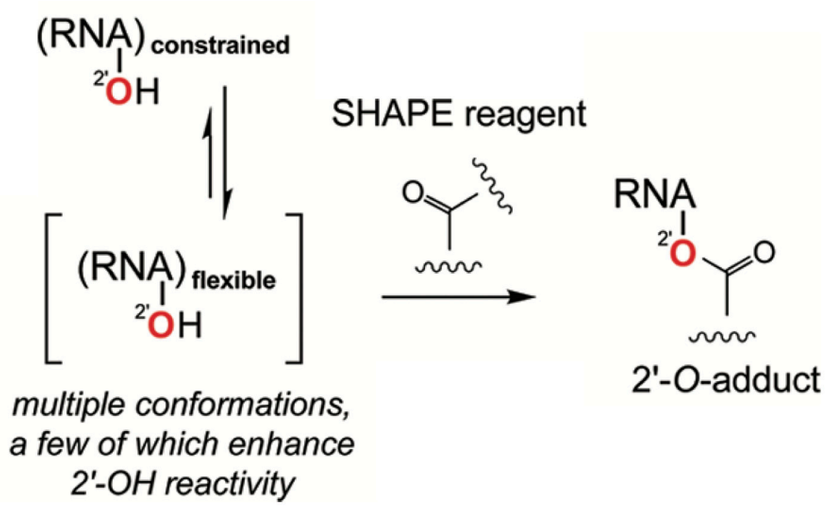


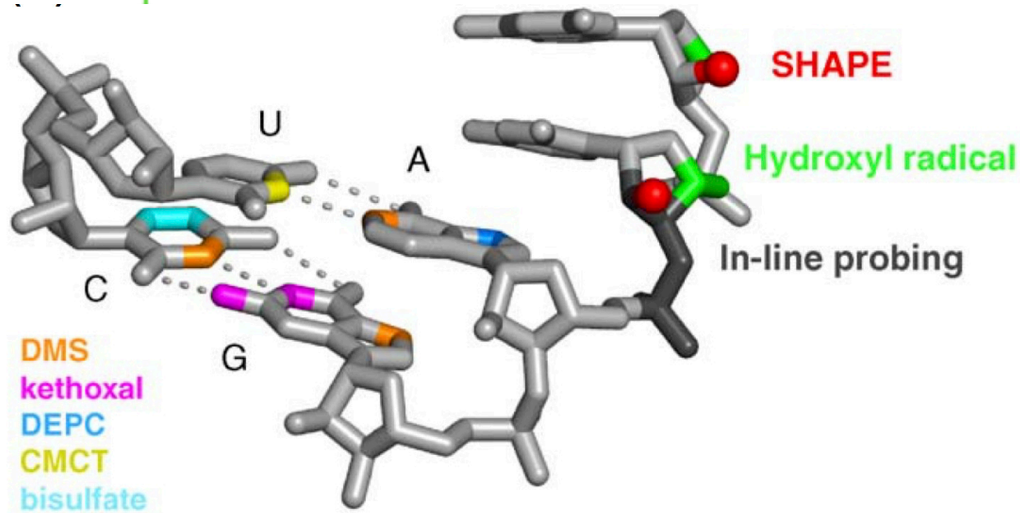


Ribozymes dynamics with cryo-EM (E. Posani, in collaboration with A. Magistrato, CNR; M. Bonomi, Pasteur; N. Toor, UCSD)

Chemical probing







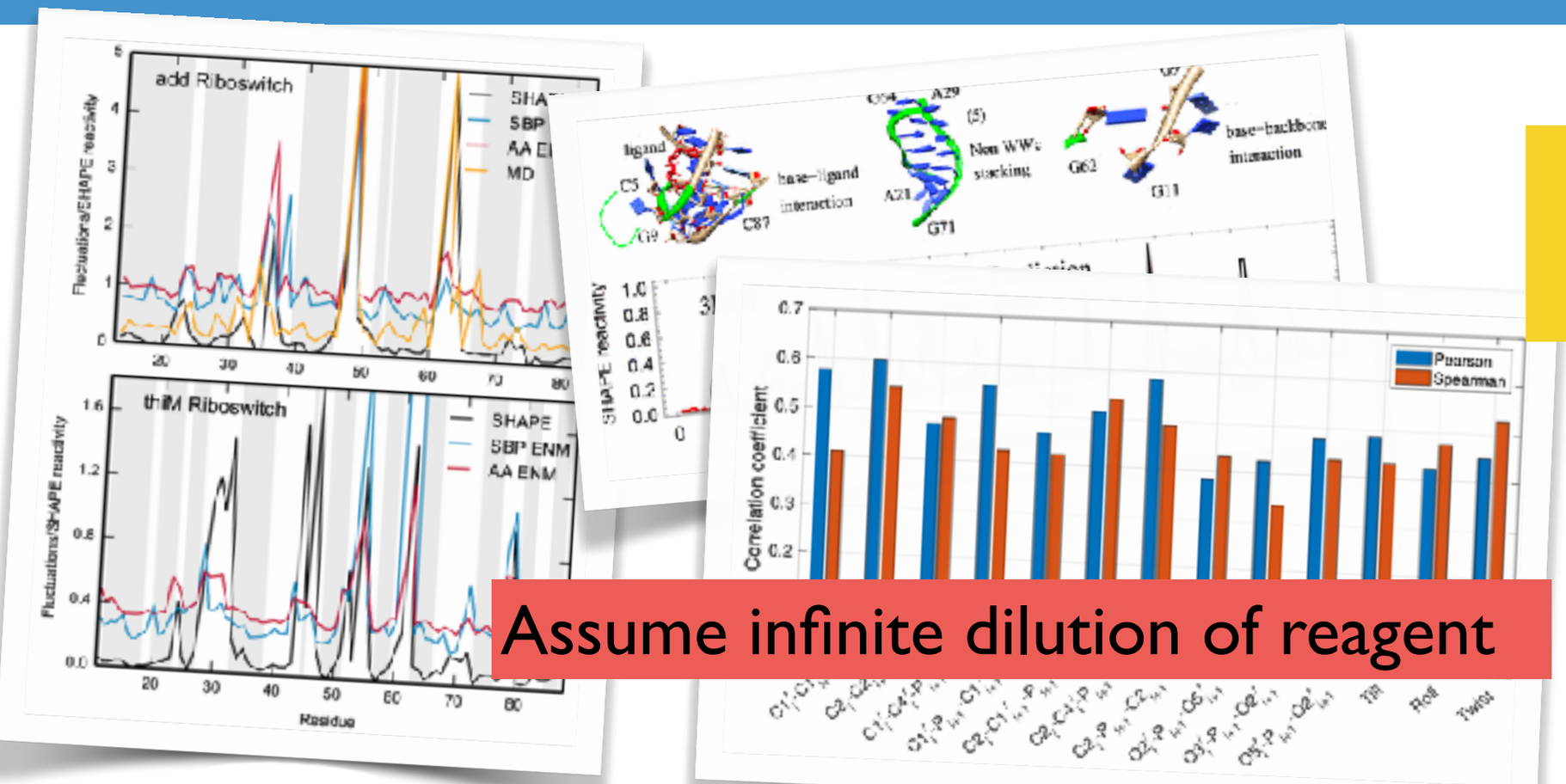
Different probes (SHAPE/DMS/CMCT/etc)

Concentration of cDNA (# reads) ~ proportional to probability to form an adduct at that position

No clear quantitative forward model from 2D/3D structure (Roee's and Redmond's talks)

Weeks COSB (2010); McGinnis et al JACS (2012)

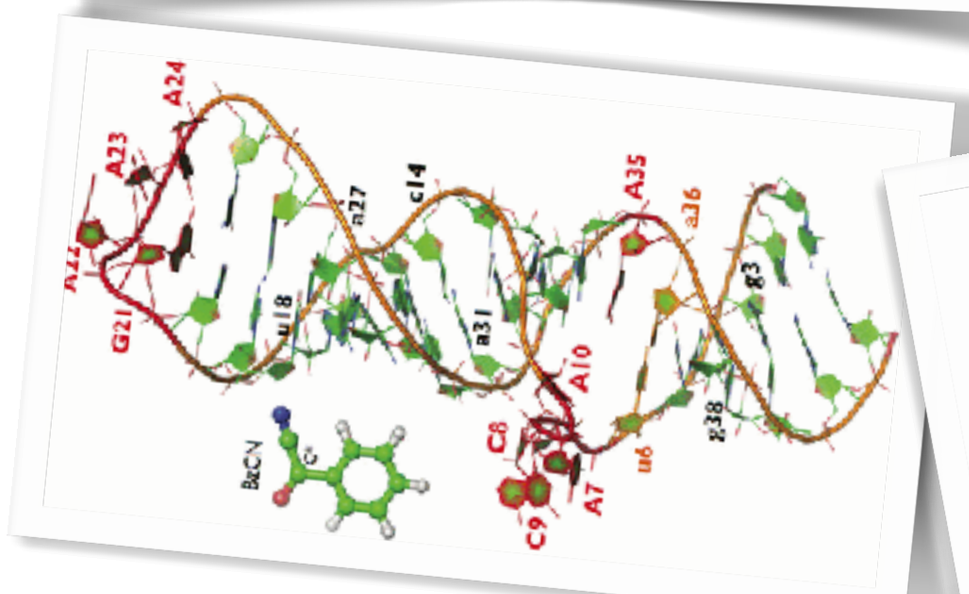
Relating SHAPE and structure/fluctuations

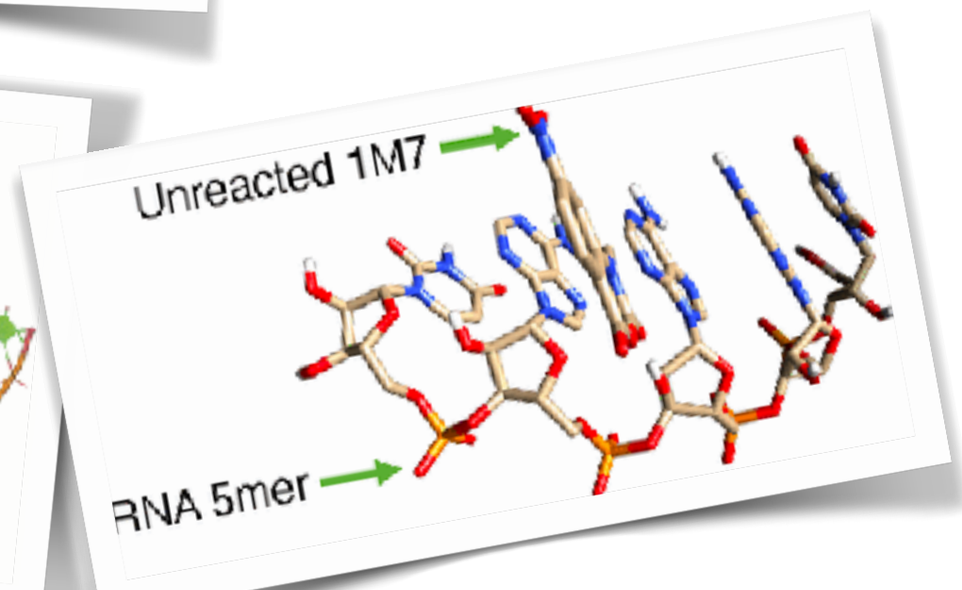


No reagent modeled

Assume infinite dilution of reagent

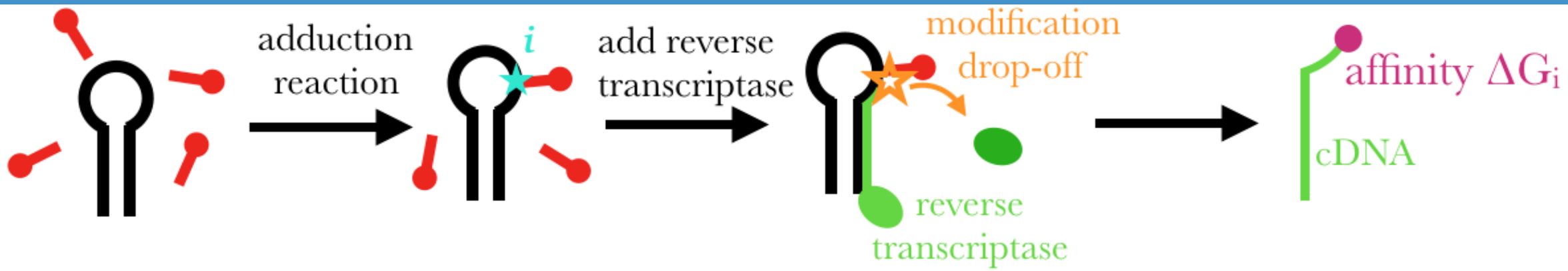
One reagent copy modeled

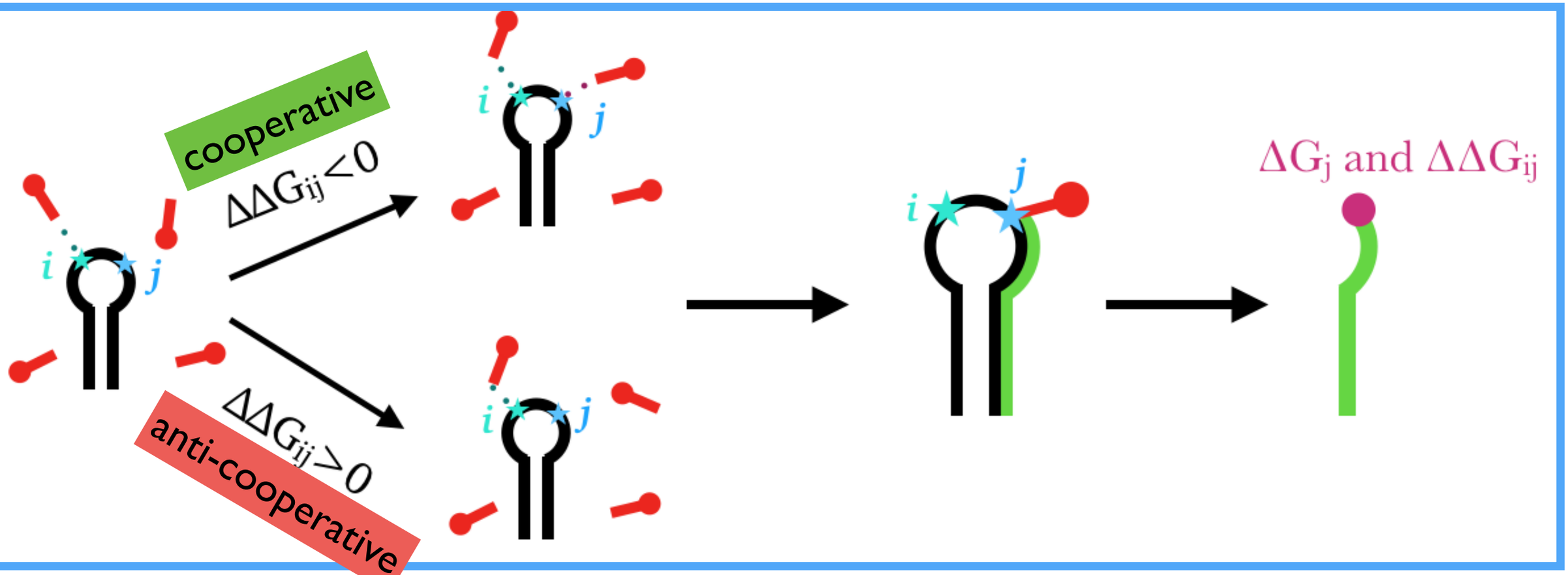




Pinamonti et al, NAR (2015); Hurst et al JPCB (2018); Frezza et al Methods (2019)
Mlynsky and Bussi, JPCL (2018); Hurst and Chen, JPCB (2021)

(Anti-)cooperative effects?





Are (anti-)cooperative effects relevant at the typical concentrations?

Physical vs Chemical binding

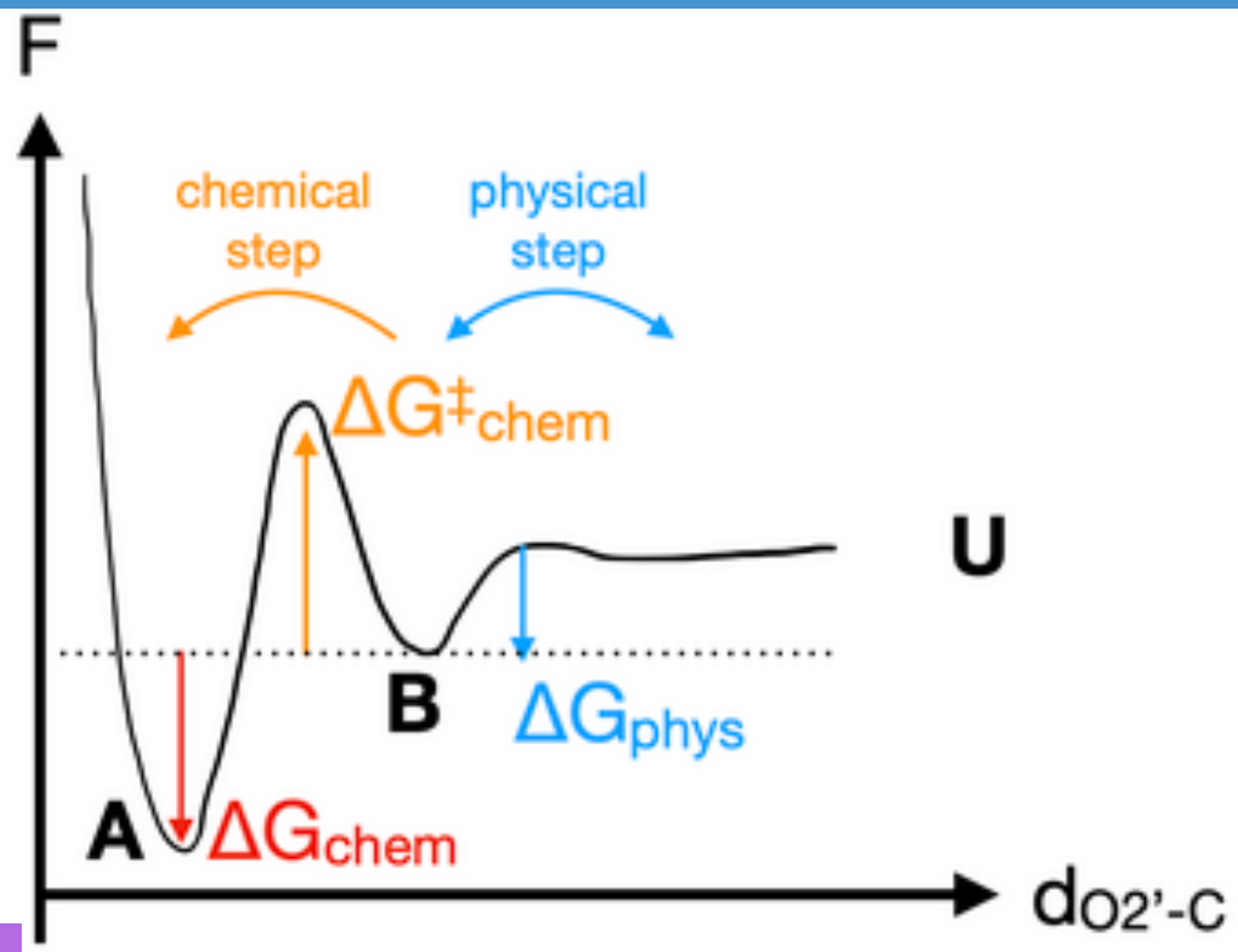
Non-equilibrium process:

- Rate-limiting irreversible chemical step
- SHAPE reagents hydrolyse water as well

Typical reagent concentrations: 10-100 mM

Usually, single-hit kinetics for short (≲100nt) RNAs:

- 1 adduction event (A) per molecule.
- How many "physical" binding events (B)?

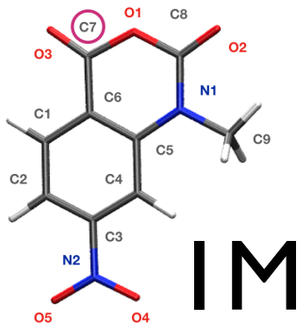


A: (chemical) adduct
B: (physical) bound
U: unbound

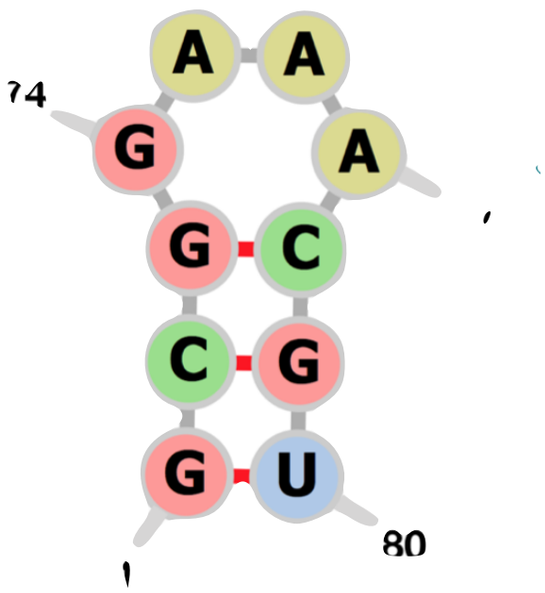
$K_{d,phys}$~0.2-6.4 M* $\Longrightarrow$ 0.1%-50% of "physical" sites (B) occupied

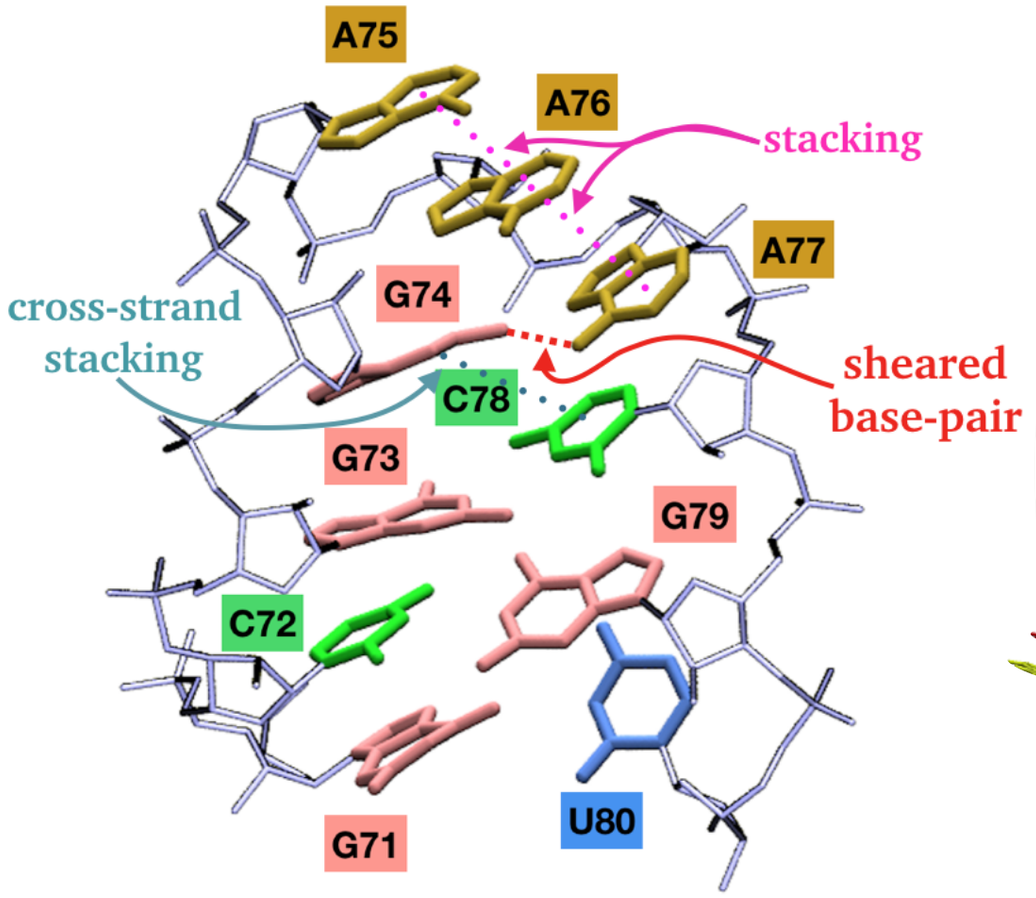
*estimated from MD simulations with NMIA, Mlynsky and Bussi, JPCL (2018)

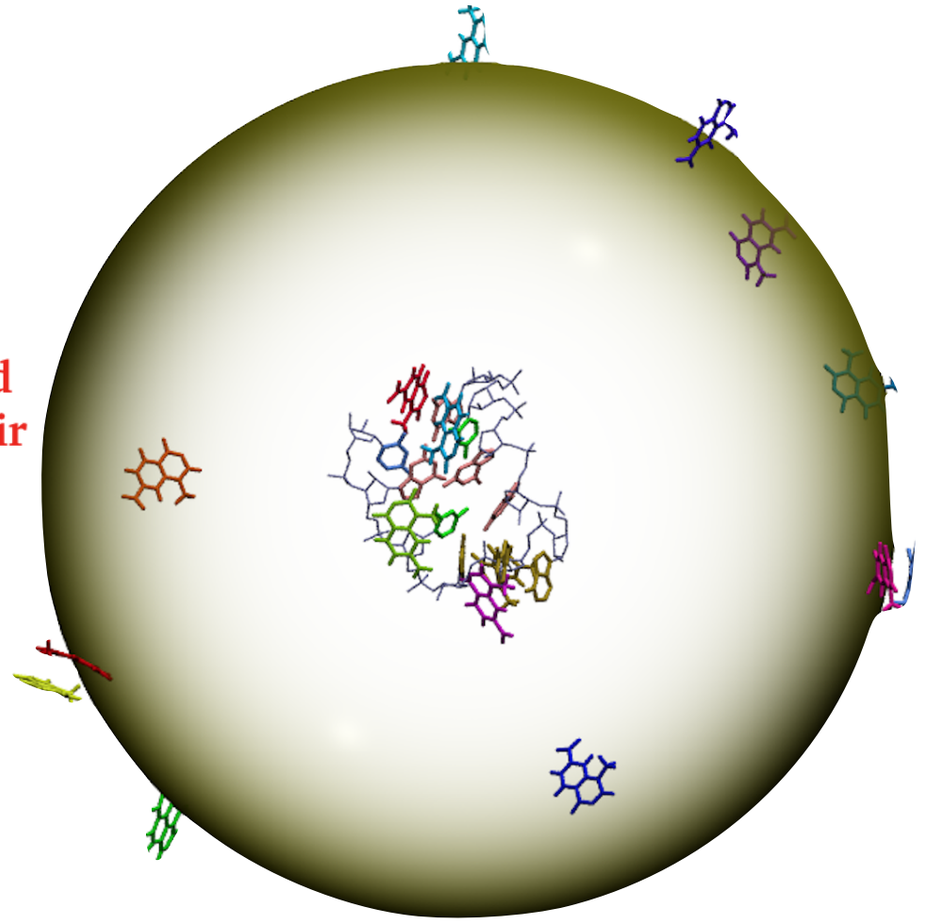
Prototype system: GAAA tetraloop



IM7







Amber FF + GROMACS, plain MD (no enhanced sampling)
Multiple (1,2…,19) copies of the reagent per simulation box
19 x 1 μs long simulations - multiple binding/unbinding events

Calonaci et al, arXiv (2022)

Grand-canonical reweighting

Probability to observe particles in A/B depending on $N=N_A+N_B$

$$P^N_{A/B}(N_{A/B}) \propto \Omega_{A/B}(N_{A/B})\Omega_{B/A}(N - N_{B/A})$$

Likelihood for the actual histograms t_{Nk}

$$P\left(t_{Nk}\right) \propto \prod_N \prod_k \left(c_N \Omega_A(k)\Omega_B(N-k)\right)^{t_{Nk}}$$

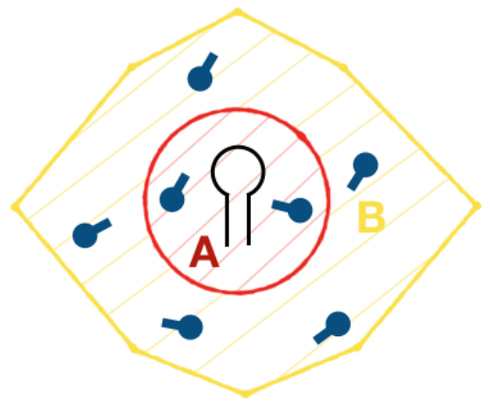
Maximizing P leads to (similar to WHAM*):

$$\Omega_A(k) = \frac{A_k}{\sum_N L_N c_N \Omega_B(N-k)}$$

$$\Omega_B(k) = \frac{B_k}{\sum_N L_N c_N \Omega_A(N-k)}$$

Grand-canonical averages:

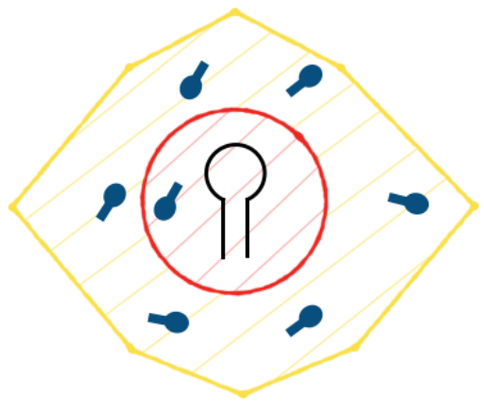
$$P^{GC}_{A/B}(N_{A/B}) \propto \Omega_{A/B}(N_{A/B}) e^{-\mu N_{A/B}/RT}$$



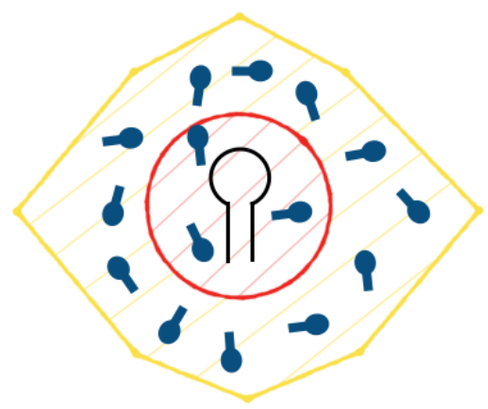
A = Binding region

B = Buffer region

Simulation with N = 4 reagents
Frame with N_A = 1 and N_B =3



Simulation with N = 7 regaents
Frame with N_A = 1 and N_B =6



Simulation with N = 15 reagents
Frame with N_A = 3 and N_B =12

Chemical potential μ estimated from concentration in buffer B.

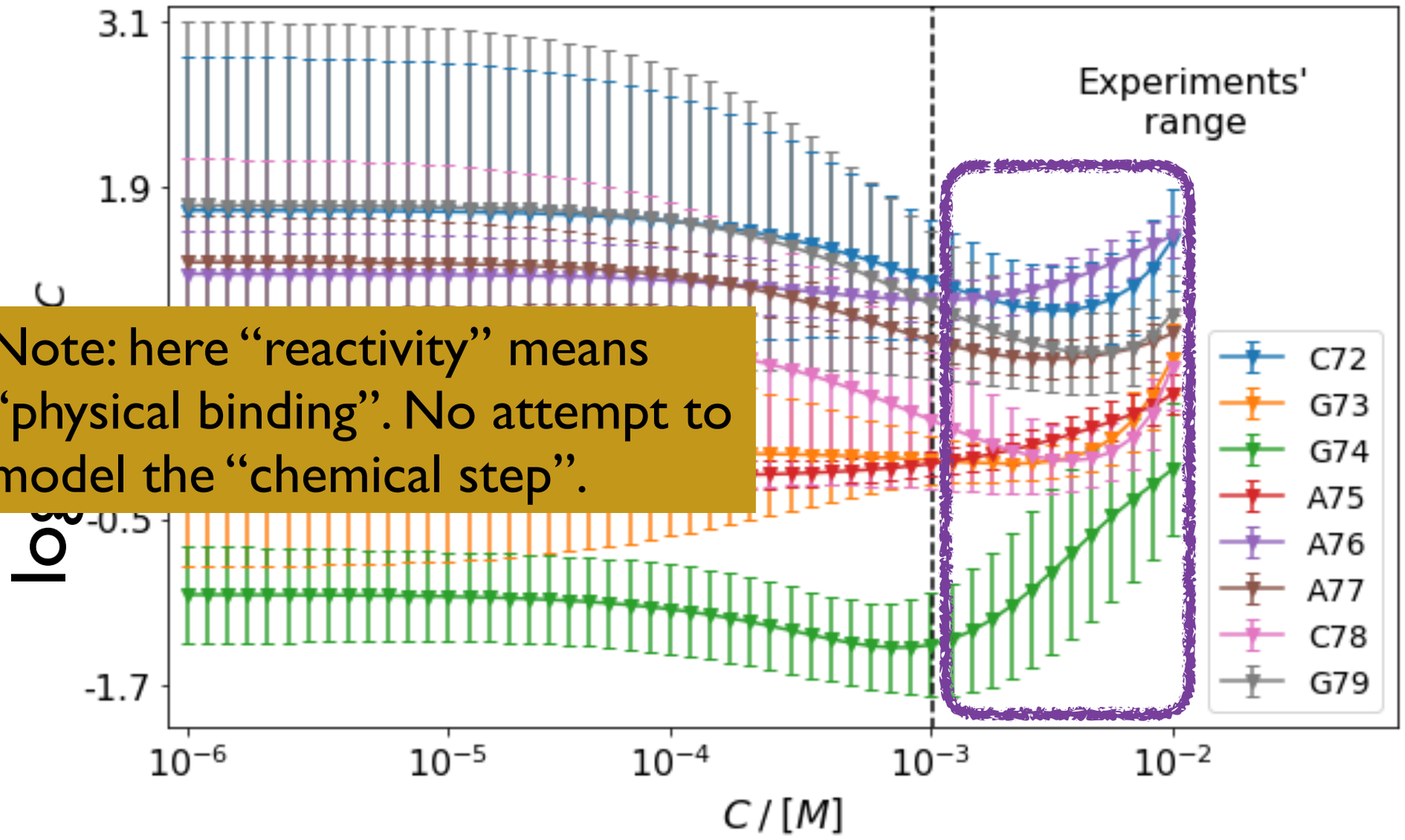
Result: smooth concentration-dependent averages!

*Kumar et al JCC (1992)

Calonaci et al, arXiv (2022)

Concentration-dependent binding

Non-linear behaviour!



Note: here "reactivity" means "physical binding". No attempt to model the "chemical step".

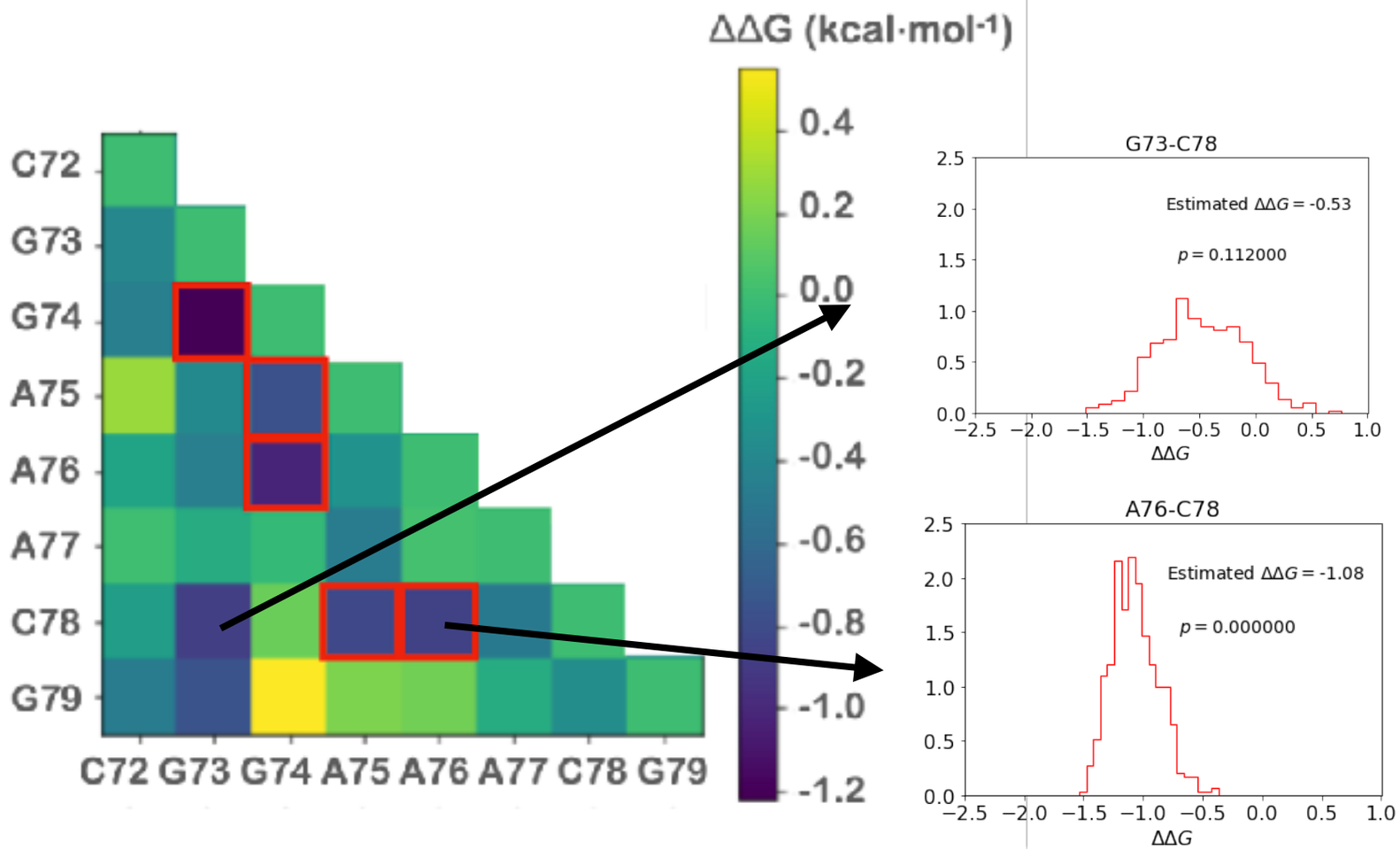
Physical occupation of an adduction site as a function of concentration

Calonaci et al, arXiv (2022)

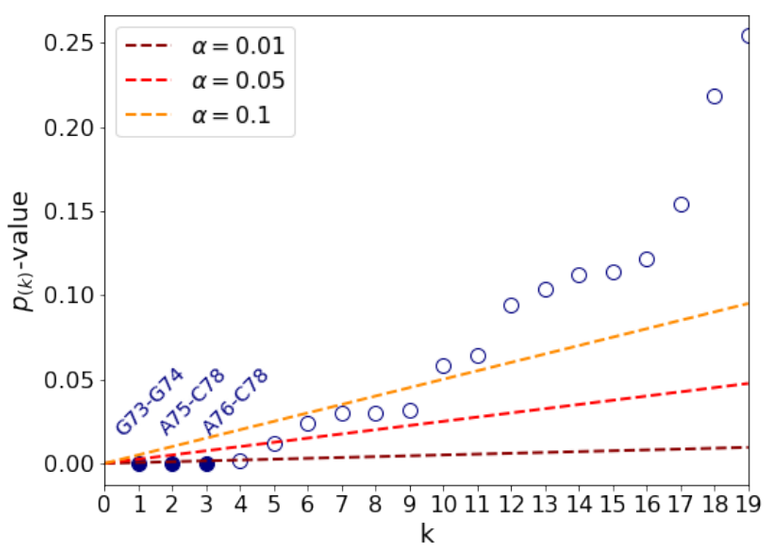
Cooperative binding

Does binding at *i* influence binding at *j*?

$$\Delta\Delta G_{ij} = -RT \log \frac{p_{ij}(1,1)p_{ij}(0,0)}{p_{ij}(1,0)p_{ij}(0,1)}$$

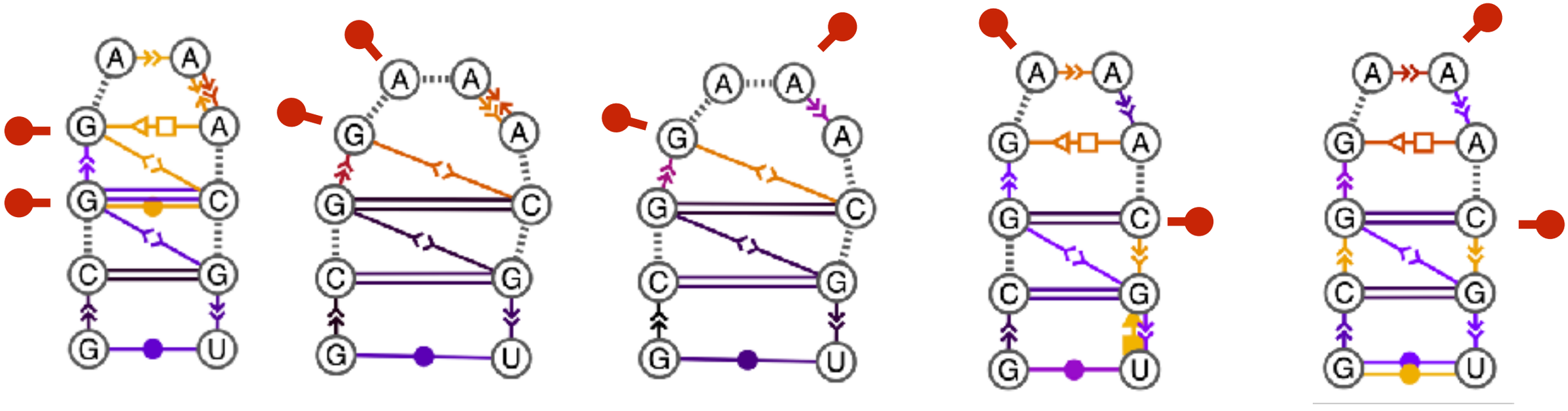


False-discovery rate (Benjamini–Hochberg test) to check for multiple hypothesis



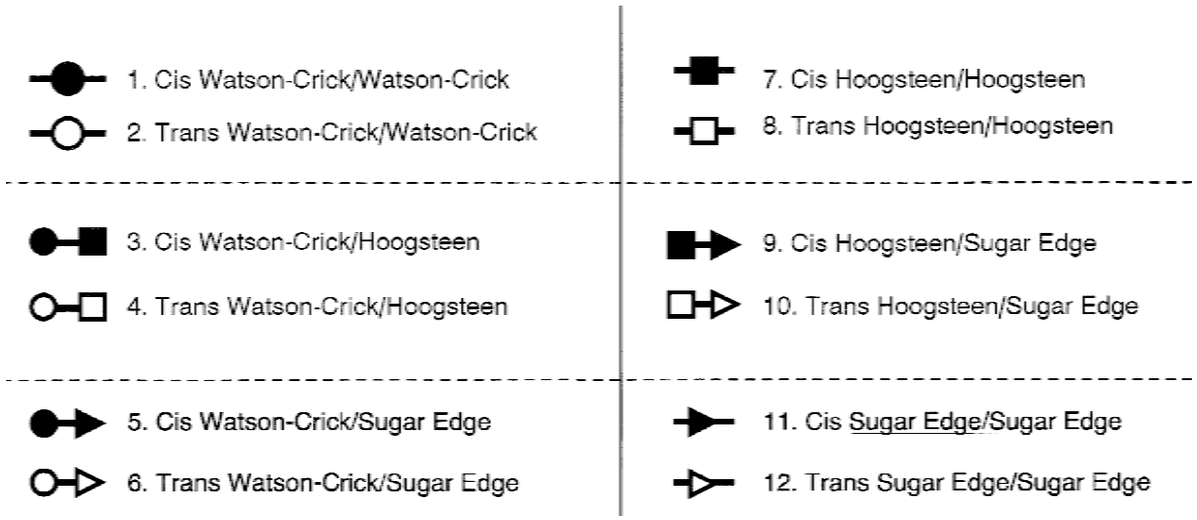
Bootstrap can tell us which pairs are non-zero "by chance"

Annotated 2D structures



Dynamical secondary structure* conditioned to double reagent binding

- Stacking between copies of reagent
- Loop reconformation



Calonaci et al, arXiv (2022)

Westhof-Leontis annotations made with BaRNAba, Bottaro et al RNA (2019)

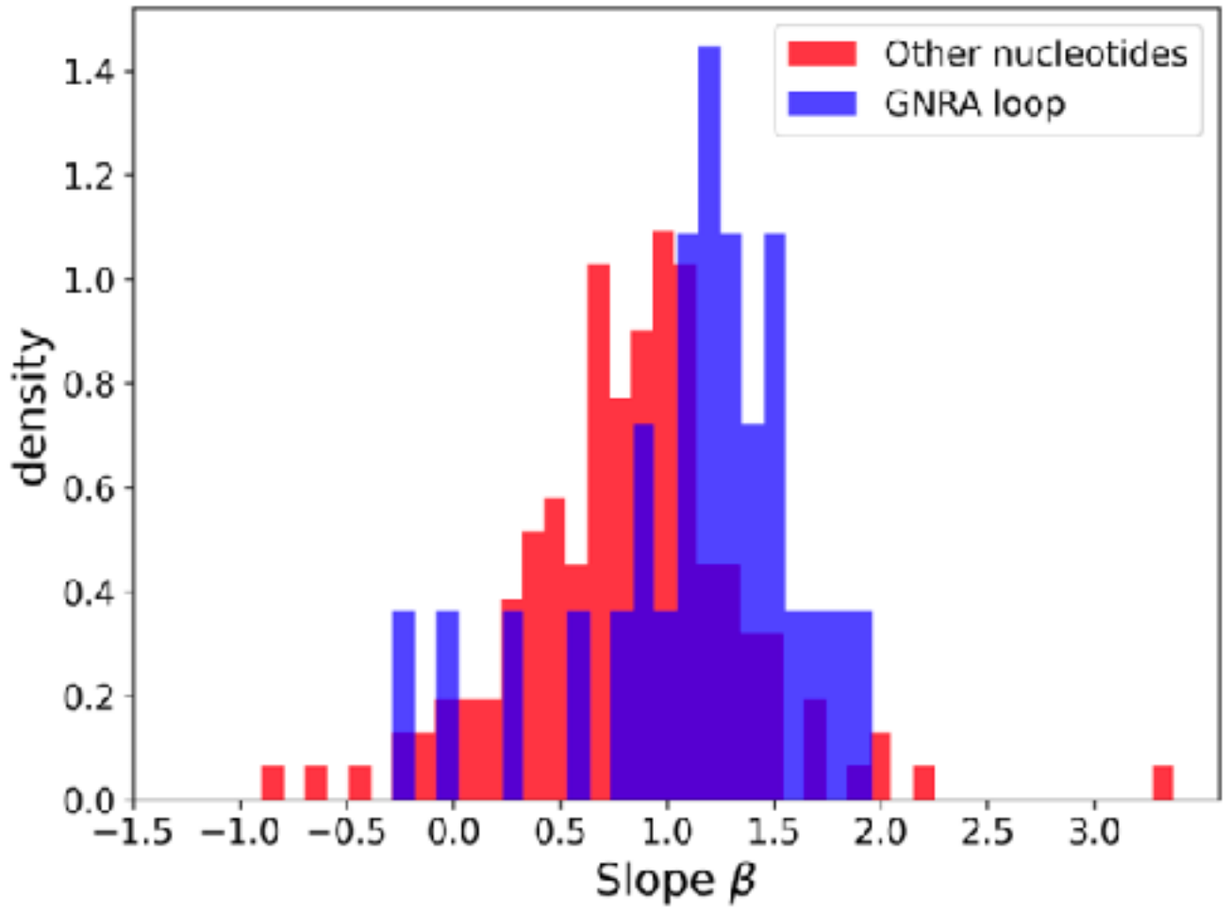
Experimental validation (qualitative)

$$\tilde{R}_i(C) = \frac{R_i(C)}{\sum_i R_i(C)} \times C$$

$$\log \tilde{R}(C) = \alpha + \beta \log (C/[M])$$

$$\log \frac{1}{1 + \left(\frac{K_A}{[L]}\right)^n} \approx \alpha + n \log[L]$$

Same as low [L] limit of Hill's equation



β>1 means superlinear (wrt average)

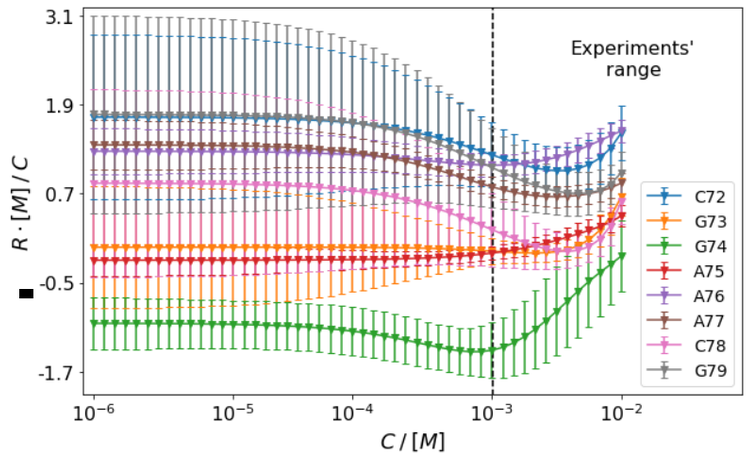
Reads are taken at 3 concentrations (32mM, 64mM, 125mM).

Normalisation is independent of concentration by construction to avoid biases (e.g. different number of cycles or other technical differences)

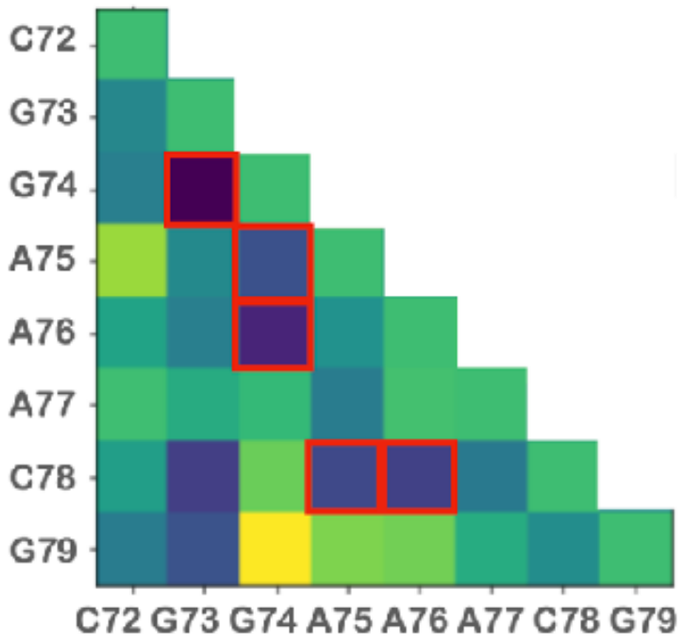
Partial summary

New method for grand-canonical averaging combining simulations at fixed number of particles

Smooth concentration-dependent curves and rigorous error analysis



At relevant reagent concentration, we predict non-linear effect. Non-linearity due to cooperative binding.

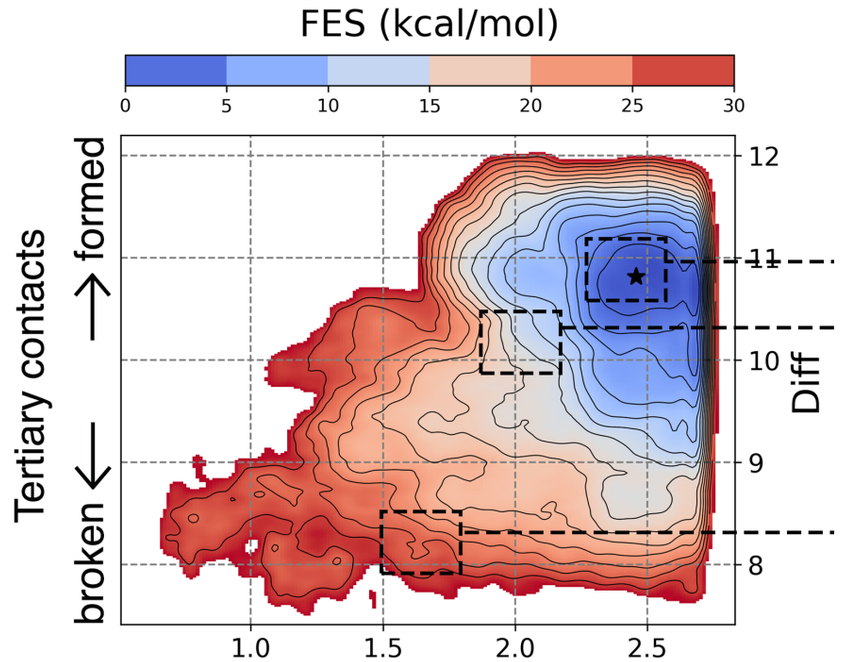


Perspective: combine experiments at different concentration to search for typical patterns

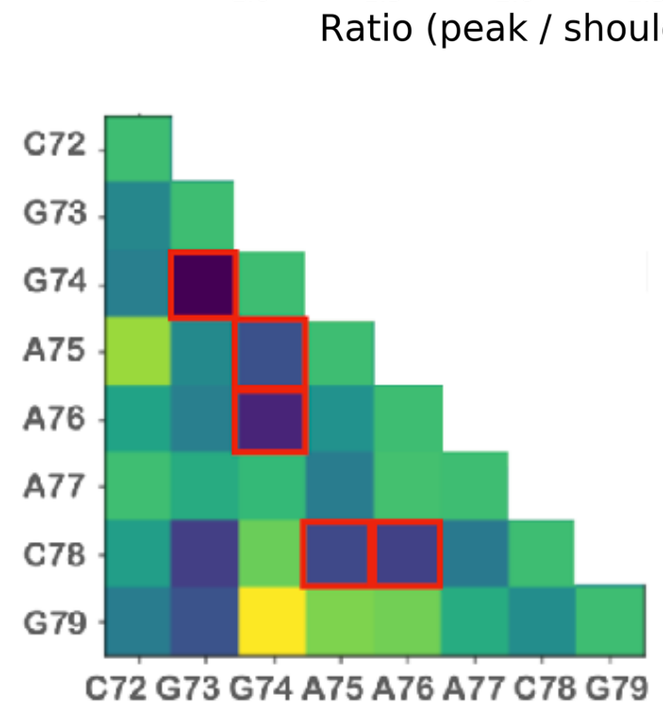
Calonaci et al, arXiv (2022)

Agenda

GAC-RNA ensembles from MD and SAXS data#



Cooperative effects in chemical probing experiments*



#Bernetti, Hall, and Bussi, NAR (2021) + Bernetti and Bussi EPJB (2021)
*Calonaci et al arXiv 2022

Acknowledgements

Tomas Silva
Olivier Languin-Cattoen
Valerio Piomponi
Ivan Gilardoni
Elisa Posani

Pavlina Pokorna
Zhengyue Zhang
(Brno, Czech Republic)

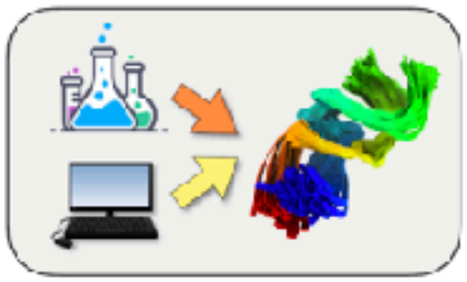
Mattia Bernetti
(IIT, Bologna, Italy)

Nicola Calonaci
(UniTS, Italy)

Kathleen Hall (St Louis, USA)

Alisha Jones & Michael Sattler
(Munich, Germany)

Cecam workshops



June 26, 2023 – June 28, 2023

RNA DYNAMICS FROM EXPERIMENTAL AND COMPUTATIONAL APPROACHES

Type
Flagship Workshop

Multinodal event
CECAM-FR-MOSER
CECAM-IT-SISSA-SNS

Location
CECAM-FR-MOSER

Organizers
Massimiliano Bonomi (Institut Pasteur - CNRS),
Giovanni Bussi (Scuola Internazionale Superiore di Studi Avanzati),
Paraskevi Gkeka (Sanofi),
Michael Sattler (Technical University of Munich)

July 3, 2023 – July 6, 2023

ENHANCED SAMPLING METHODS WITH PLUMED

Type
Flagship School

Location
CECAM-HQ-EPFL, Lausanne, Switzerland

Organizers
Massimiliano Bonomi (Institut Pasteur - CNRS),
Giovanni Bussi (Scuola Internazionale Superiore di Studi Avanzati),
Carlo Camilloni (University of Milano),
Gareth Tribello (Queen's University Belfast)

Openings

PLUMED

One 30-months developer position for PLUMED (Python/C++)
GitHub/plumed/opening-2023
Deadline Feb 23

MOLECULAR AND STATISTICAL BIOPHYSICS

4 fully funded PhD positions
Physics and Chemistry of Biological Systems
@SISSA
Deadline Mar 20

Faculty position at the tenure track / associate professor level, expression of interest
Deadline Mar 31

nature careers