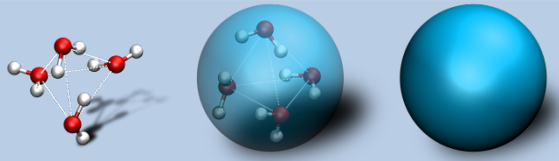


Developing an Implicit Solvation Machine Learning Model for Molecular Simulations of Ionic Media

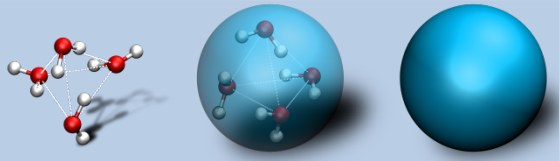
Matej Praprotnik

Laboratory for Molecular Modeling
National Institute of Chemistry

ML4MS, IJS, Ljubljana, May 17, 2024

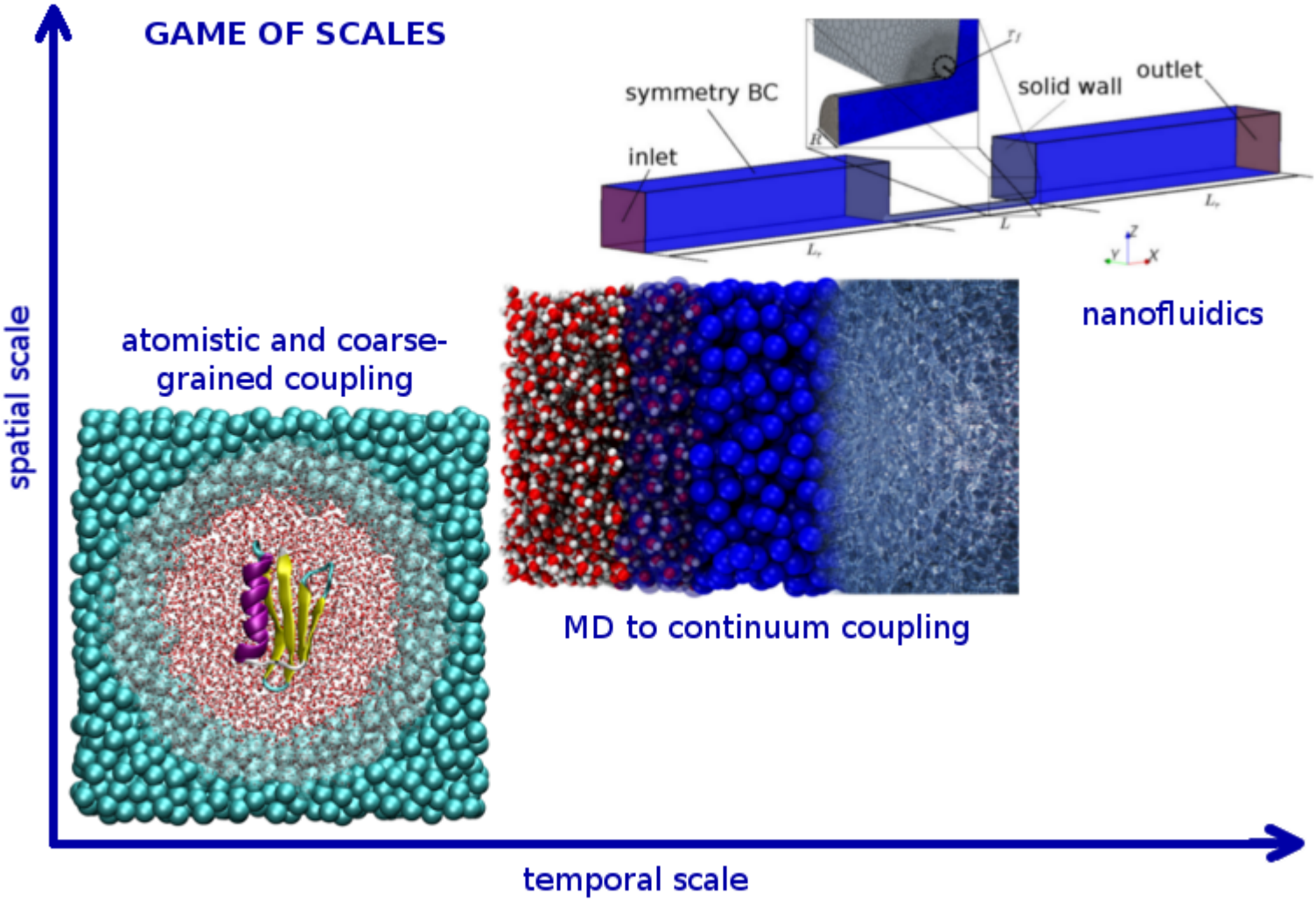
NATIONAL INSTITUTE OF CHEMISTRY

Outline

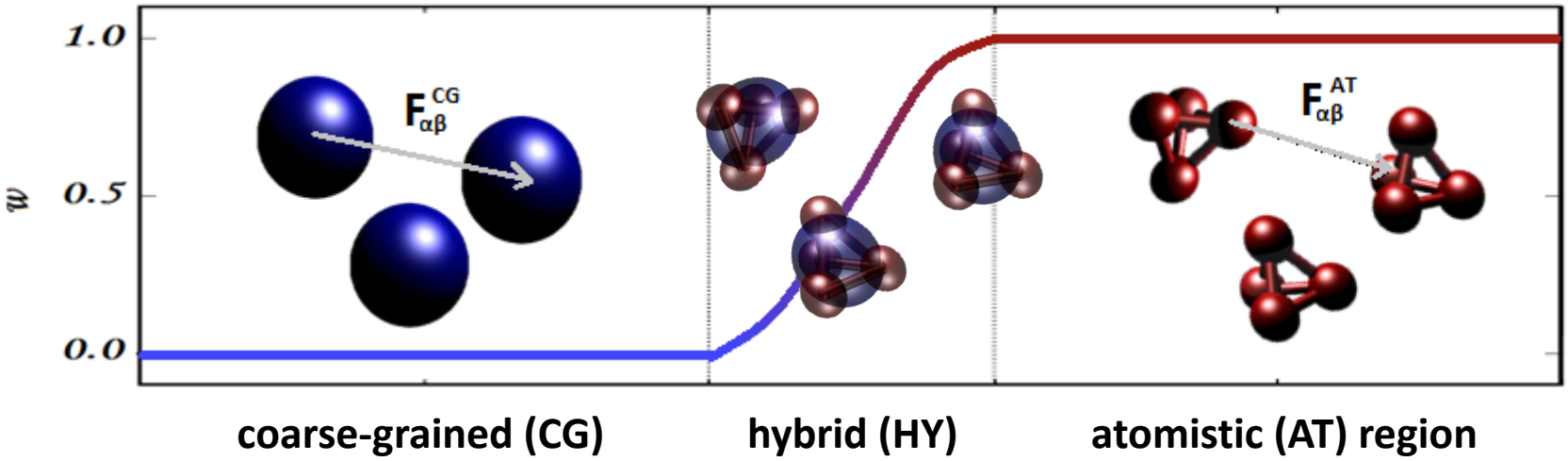


- Multiscale simulations of biomolecular systems:
  1. DNA molecule in salt solution
  2. Rotational dynamics of a protein under shear flow
  3. ultrasound propagation through liquid water
- Machine learning applications for biomolecular systems:
  1. Deep Implicit Solvation (DIS) model for sodium chloride solutions
  2. Learning of Effective Dynamics (LED) of alanine dipeptide

Multiscale modeling & simulation



Adaptive Resolution Scheme (AdResS)



- force between particle α and β:

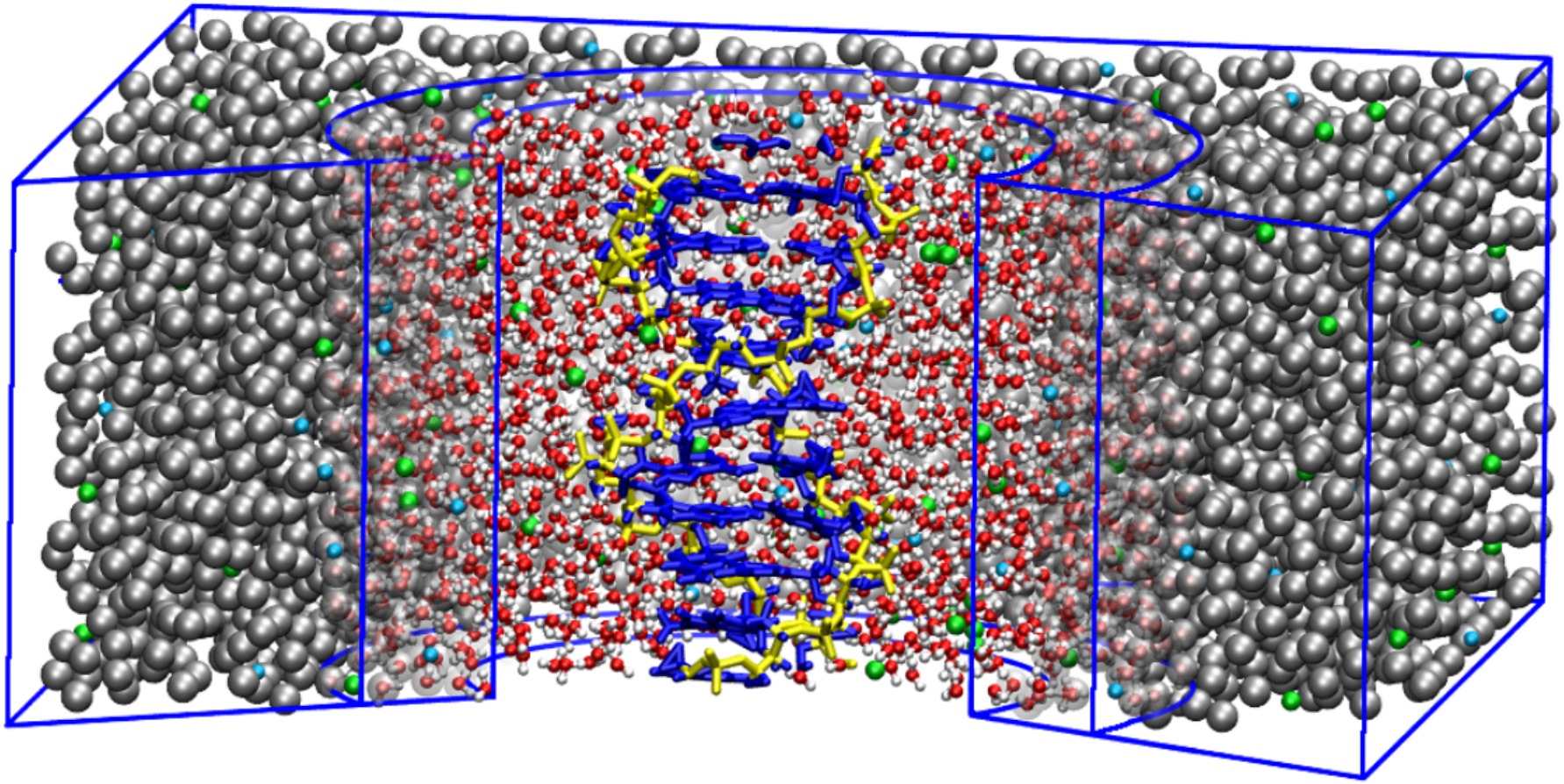
$$F_{\alpha}^{AdResS} = \sum_{\beta \neq \alpha} w\left(\left|R_{\alpha} - R\right|\right) w\left(\left|R_{\beta} - R\right|\right) F_{\alpha\beta}^{AT} + \sum_{\beta \neq \alpha} \left[1 - w\left(\left|R_{\alpha} - R\right|\right) w\left(\left|R_{\beta} - R\right|\right)\right] F_{\alpha\beta}^{CG} - F_{\alpha}^{TD}\left(\left|R_{\alpha} - R\right|\right)$$

$w(r)$… position dependent weighting function

- above force coupling scheme obeys Newton‘s third law

Praprotnik, Delle Site, Kremer; Annu. Rev. Phys. Chem. (2008)

Atomistic DNA in multiscale salt solution



Zavadlav, Podgornik, Praprotnik; *J. Chem. Theory Comput.* (2015)

DNA molecule in multiscale salt solution

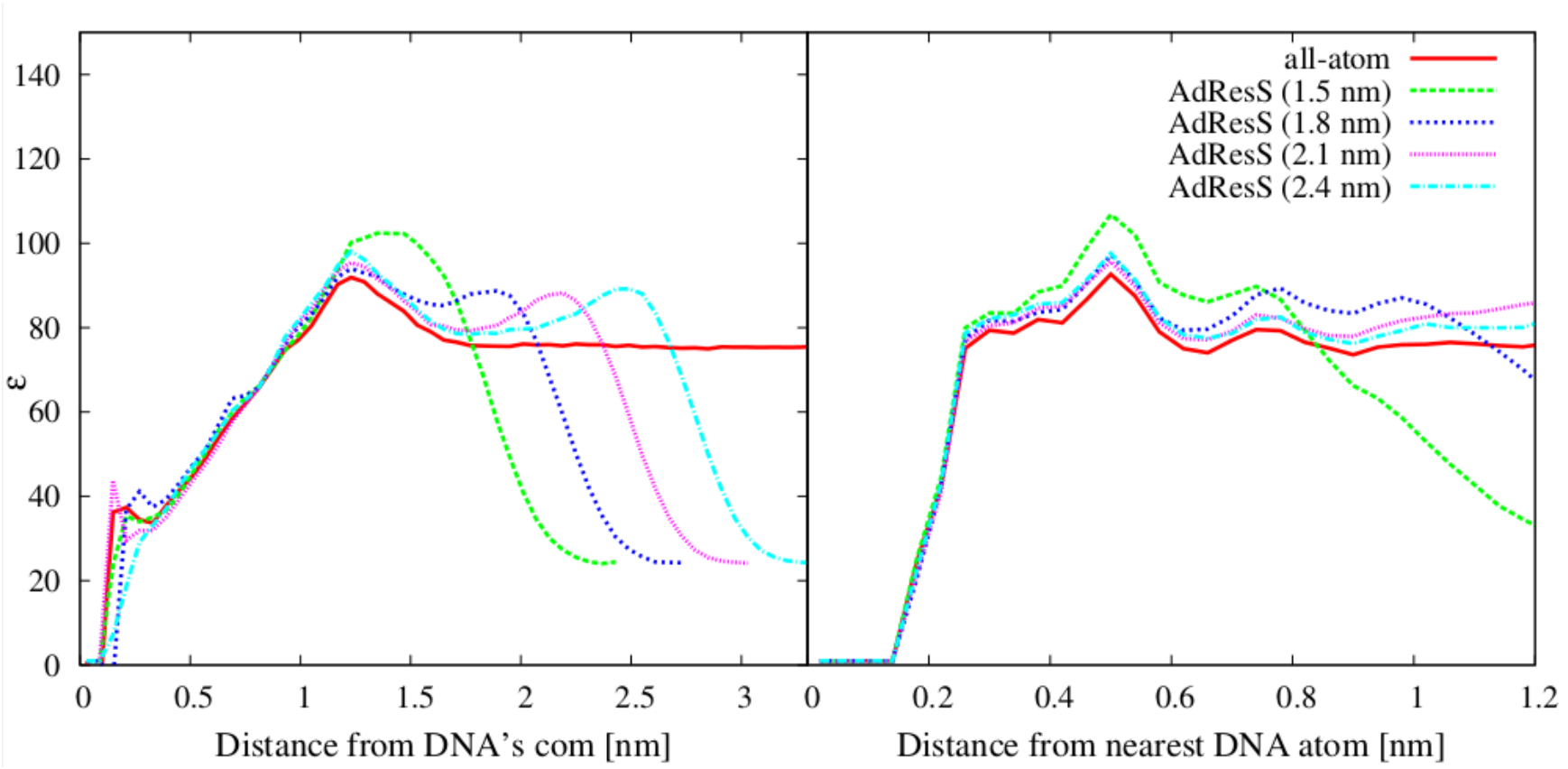
- dielectric constant of DNA molecule

Group	ε (1.5 nm)	ε (1.8 nm)	ε (2.1 nm)	ε (2.4 nm)	ε (∞)
Phosphate	18.3	21.1	17.3	15.7	17.0
Sugar	2.7	2.8	2.7	2.4	2.6
Base	2.1	2.1	2.0	2.0	2.1
DNA	5.6	5.9	5.4	4.5	5.0

Zavadlav, Podgornik, Praprotnik; *JCTC* (2015)

DNA molecule in multiscale salt solution

- dielectric constant of water



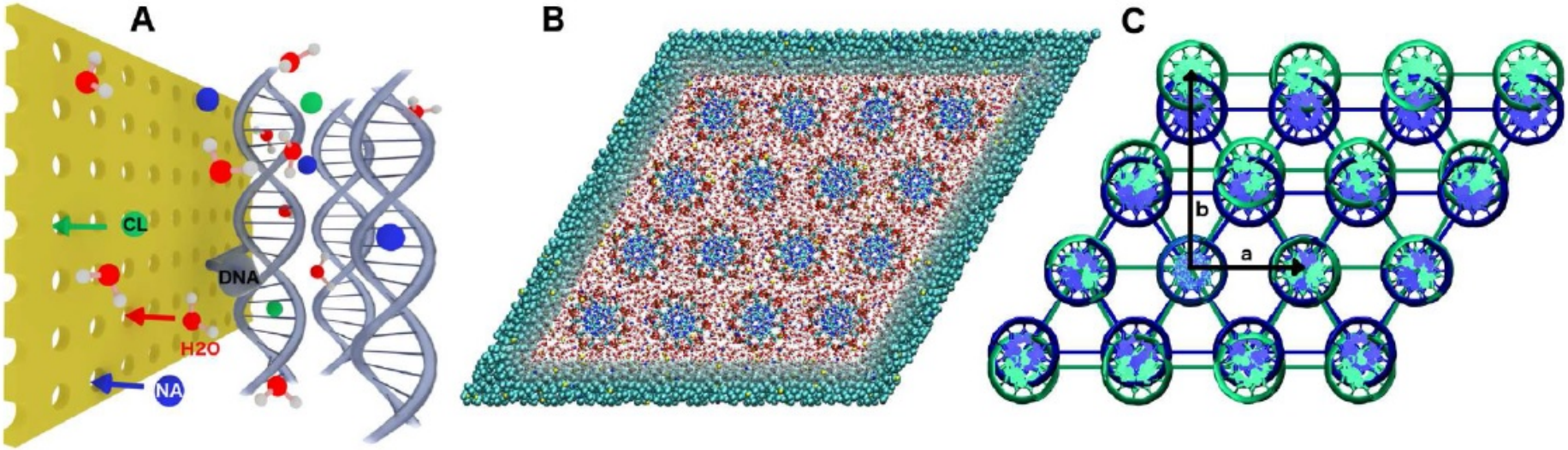
Zavadlav, Podgornik, Praprotnik; *JCTC* (2015)

Columnar phases of DNA arrays

Isotropic	Cholesteric	Hexagonal 2D progressive 3D longitudinal ordering	Orthorhombic
C(mg/ml) ≈	160 (*) 380	670	1055
	mean interhelices distance a_m 49 Å 32 Å	intermolecular distance a_H 31.5 Å 29 Å 23.7 Å	lattice parameters a = 24.09 Å a = 20.77 Å b = 39.33 Å b = 29.72 Å
		helix pitch P 34.6 Å ..... 30.2 Å	

- system of 16 DNA molecules
- hexagonal/orthorhombic
- Na^+/Spd^{3+}

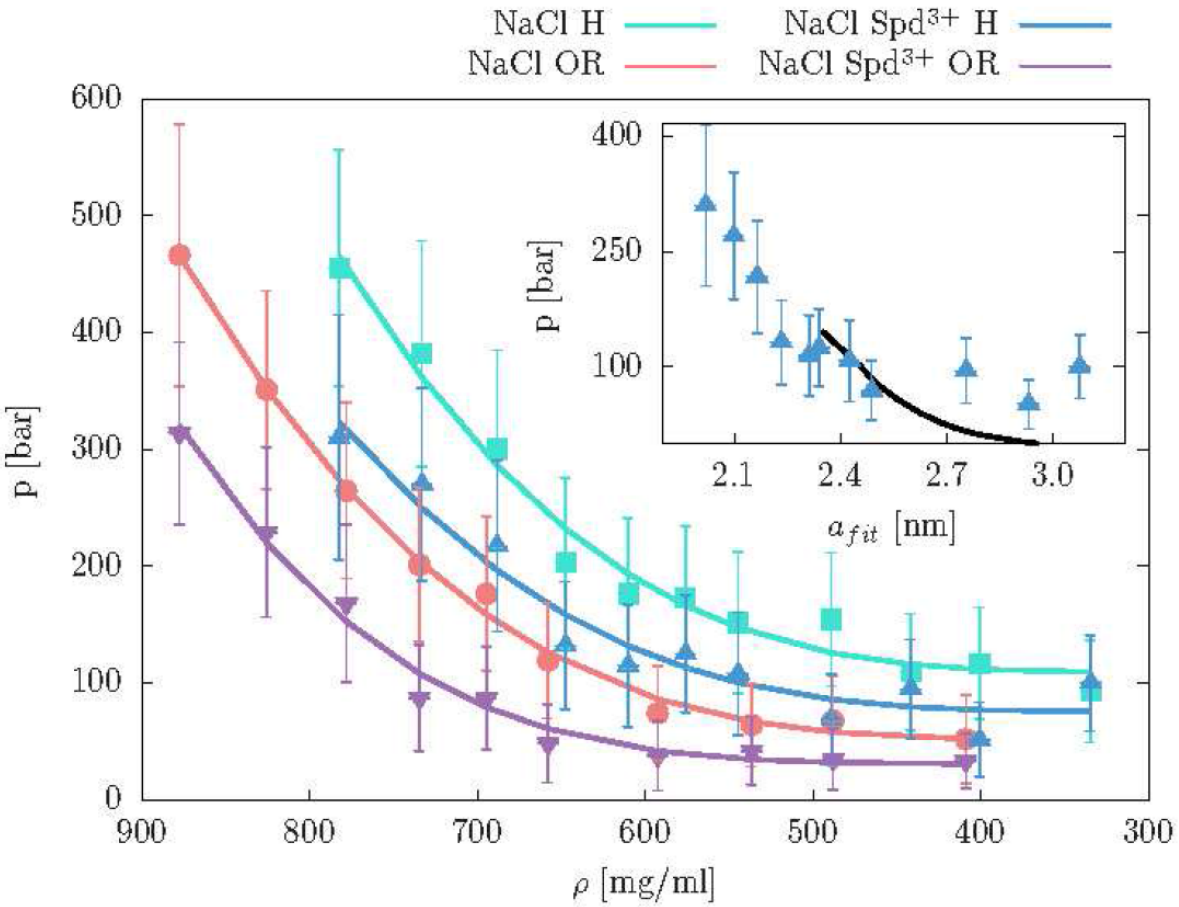
Durand, Doucet, Livolant, *J. Phys. II France* (1992)



Zavadlav, Podgornik, Praprotnik; *Sci. Rep.* (2017)

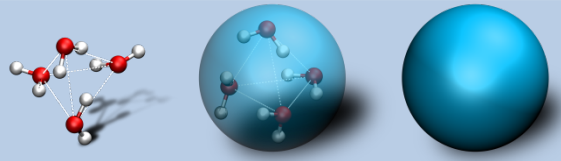
Columnar phases of DNA arrays

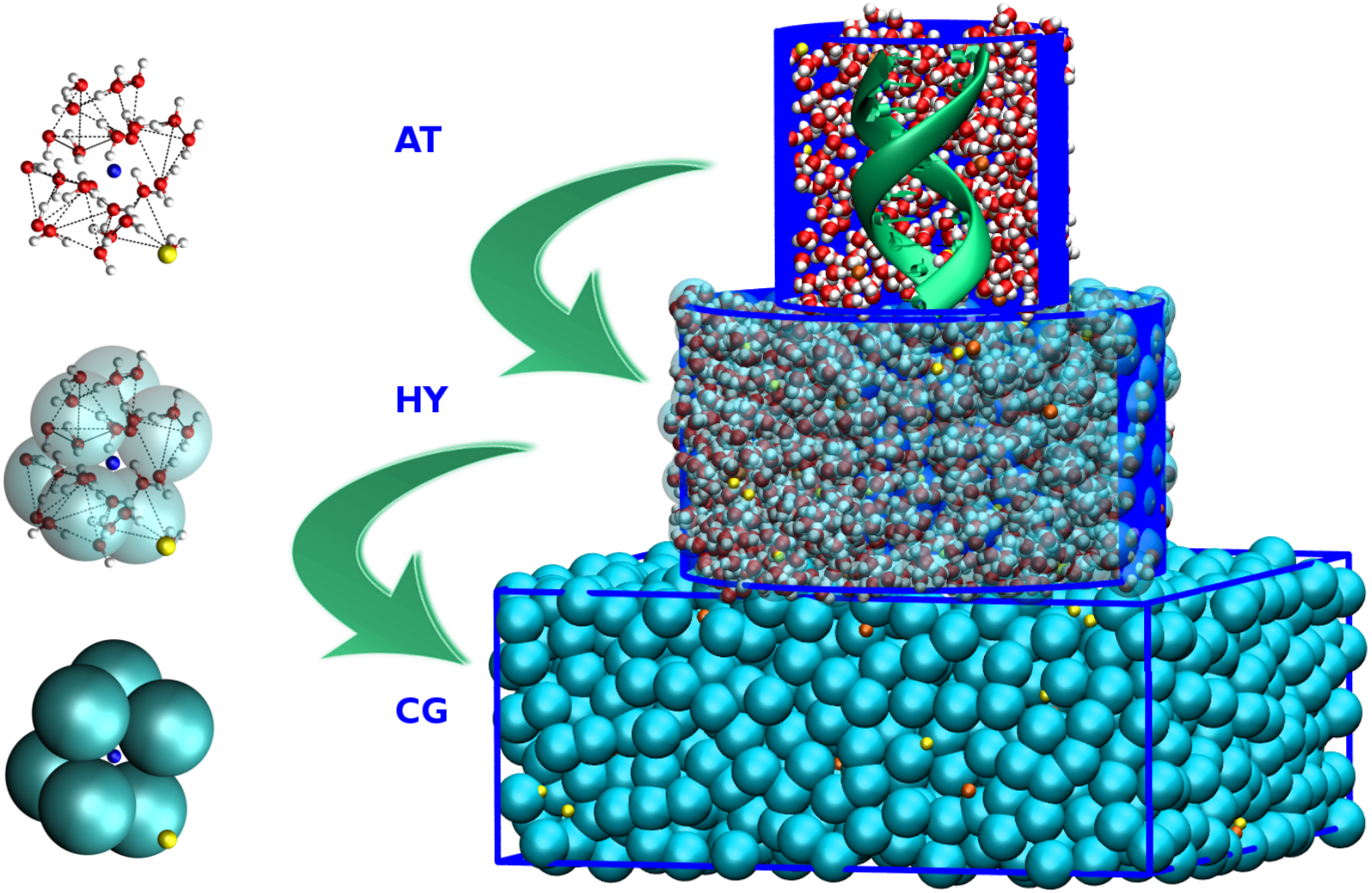
- osmotic pressure vs. DNA density



Zavadlav, Podgornik, Praprotnik; *Sci. Rep.* (2017)

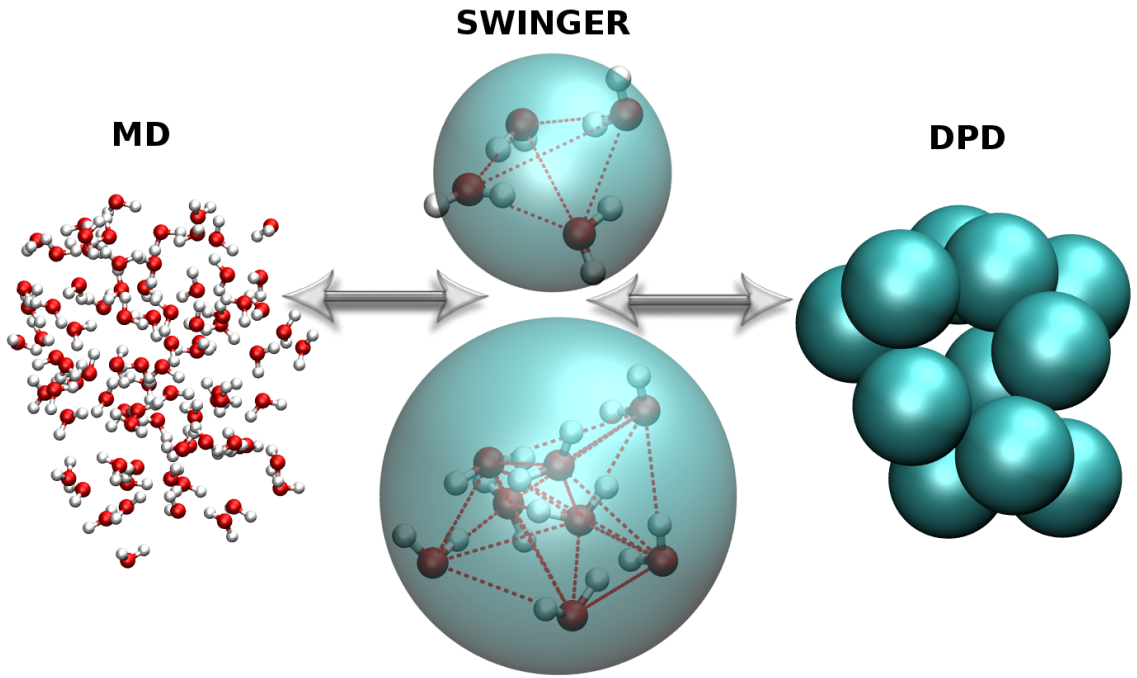
DNA molecule in bundled-SPC/MARTINI salt solution





Zavadlav, Podgornik, Melo, Marrink, Praprotnik; *EPJST* (2016)

MD/DPD water



MD:

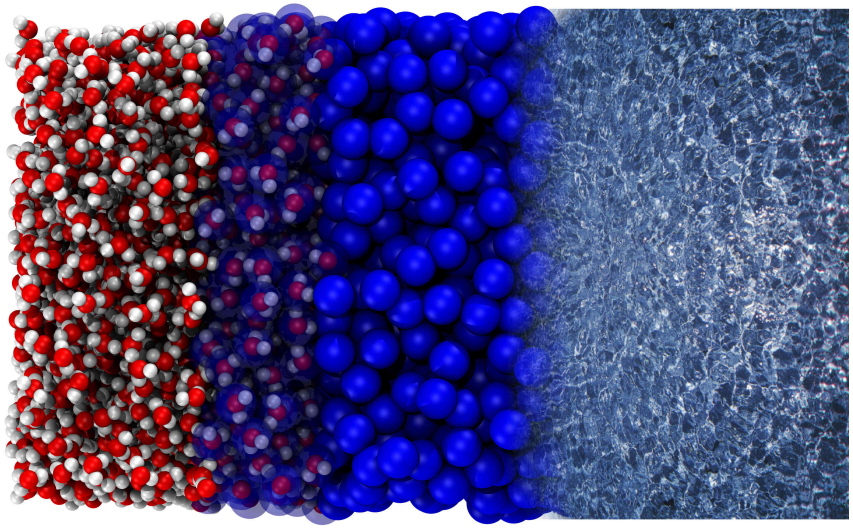
$$\mathbf{F}_{ij}^{MD,C}(\mathbf{r}_{ij}) = -\frac{\partial U^{MD}}{\partial \boldsymbol{r}_{ij}}$$

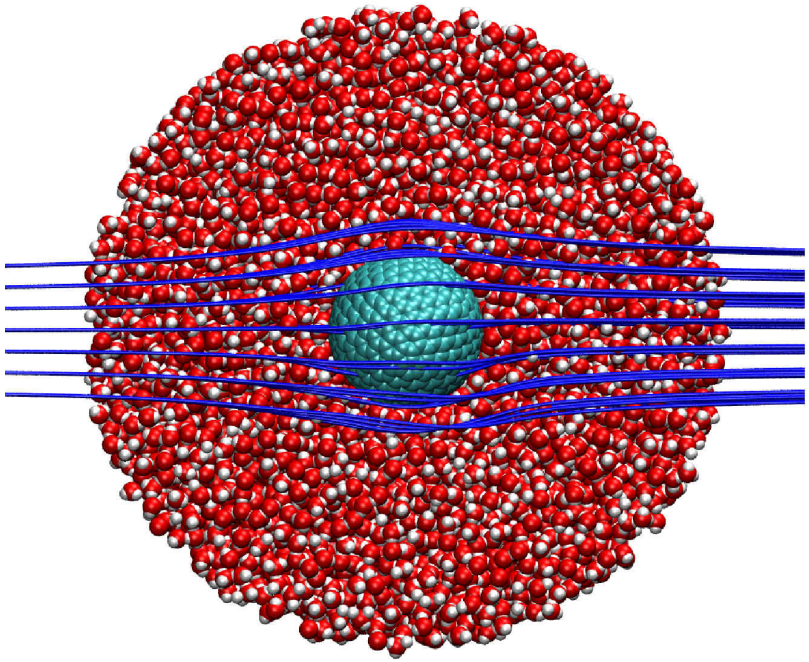
DPD:

$$\mathbf{F}_{\alpha\beta}^{DPD,C}(\mathbf{R}_{\alpha\beta}) = a_{\alpha\beta}(1 - R_{\alpha\beta}/R_c)\hat{\mathbf{R}}_{\alpha\beta}$$
$$\mathbf{F}_{\alpha\beta}^{DPD,R}(\mathbf{R}_{\alpha\beta}) = \sqrt{2\gamma_{\alpha\beta}k_BT}(1 - R_{\alpha\beta}/R_c)\zeta_{ij}\hat{\mathbf{R}}_{\alpha\beta}$$
$$\mathbf{F}_{\alpha\beta}^{DPD,D}(\mathbf{R}_{\alpha\beta}) = -\gamma_{\alpha\beta}(1 - R_{ij}/R_c)^2(\hat{\mathbf{R}}_{\alpha\beta} \cdot \mathbf{V}_{\alpha\beta})\hat{\mathbf{R}}_{\alpha\beta}$$

Zavadlav, Praprotnik; *J. Chem. Phys. (2017)*

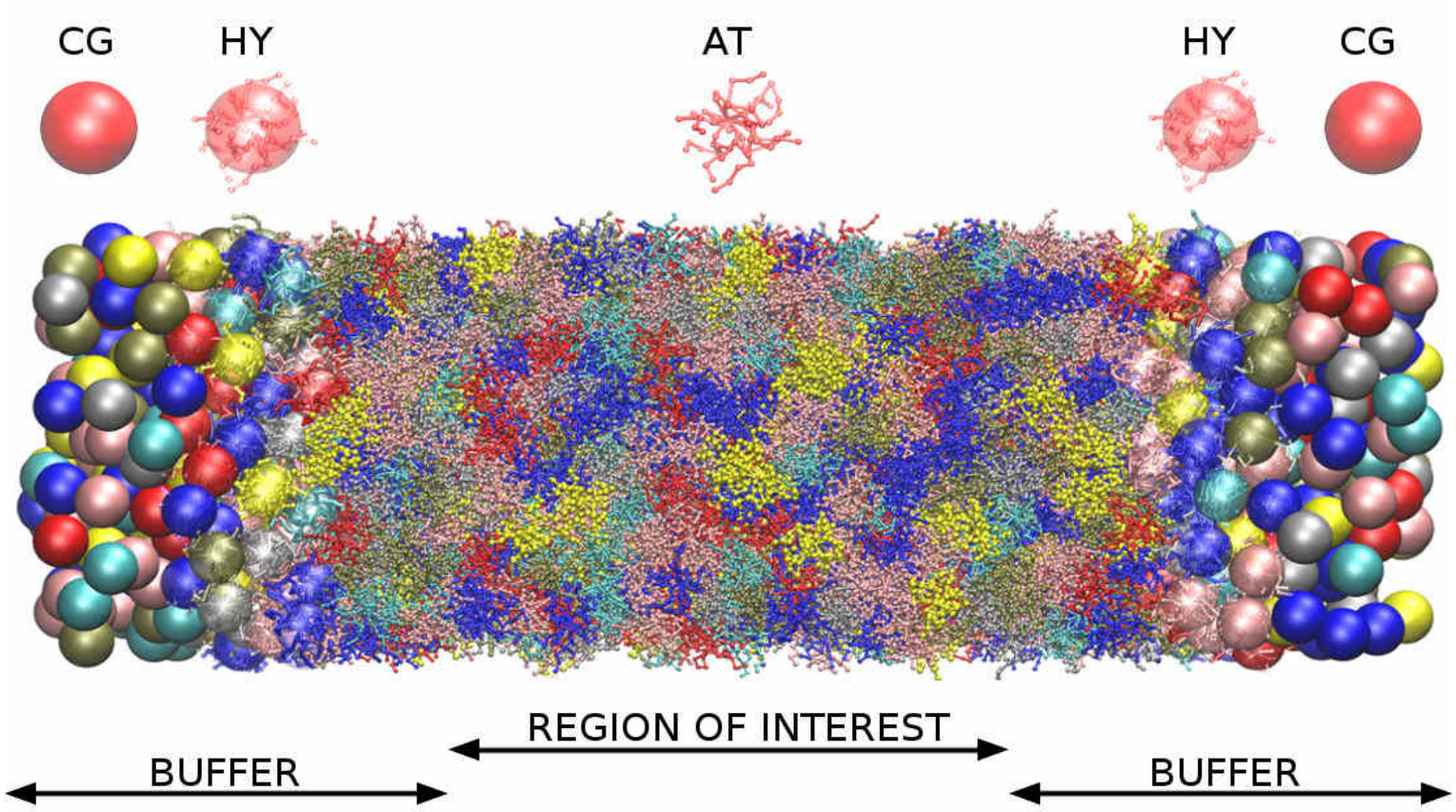
Coupling to CFD





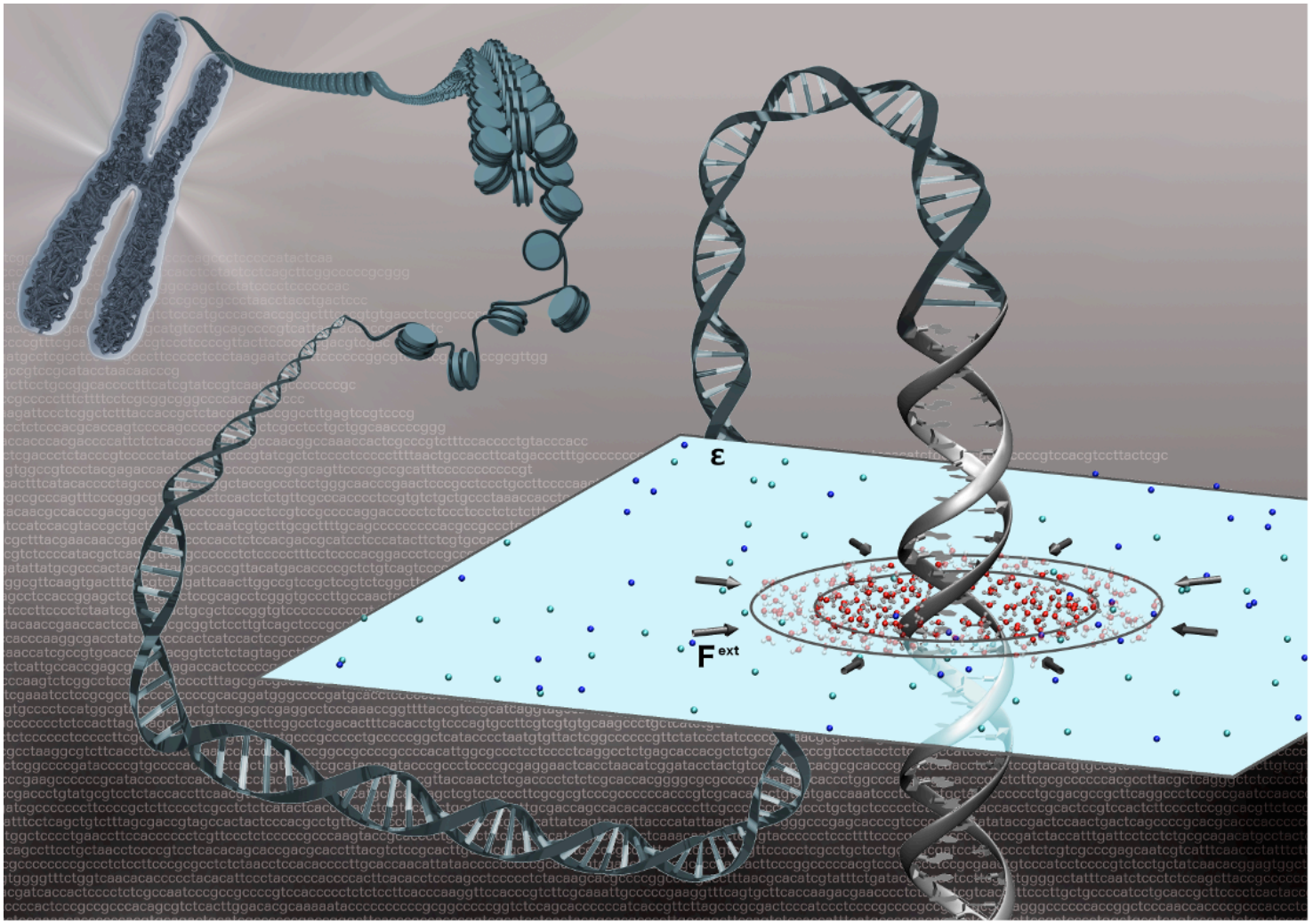
Delgado-Buscalioni, Sablić, Praprotnik; *Eur. Phys. J. Special Topics* (2015)
Walther et al.; *J. Comput. Phys.* (2012)

Open Boundary Molecular Dynamics



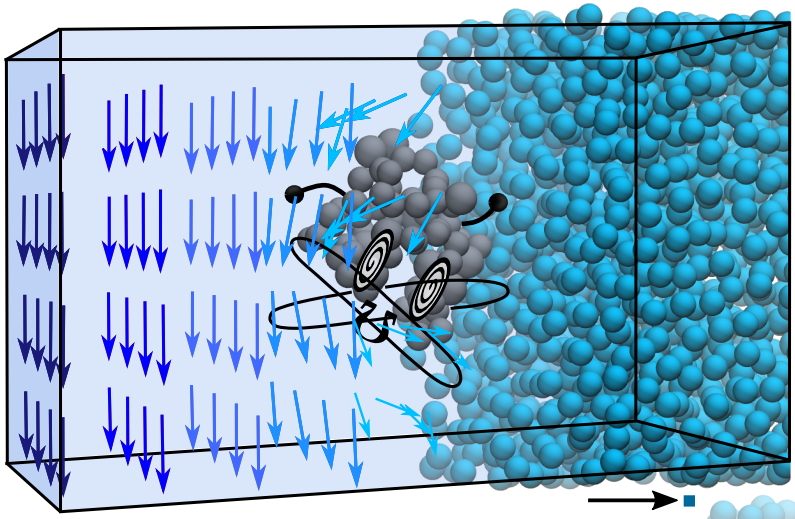
Delgado-Buscalioni, Sablić, Praprotnik; *Eur. Phys. J. Special Topics* (2015)
Sablić, Praprotnik, Delgado-Buscalioni; *Soft Matter* (2016)
Delle Site, Praprotnik; Phys. Rep. (2017)

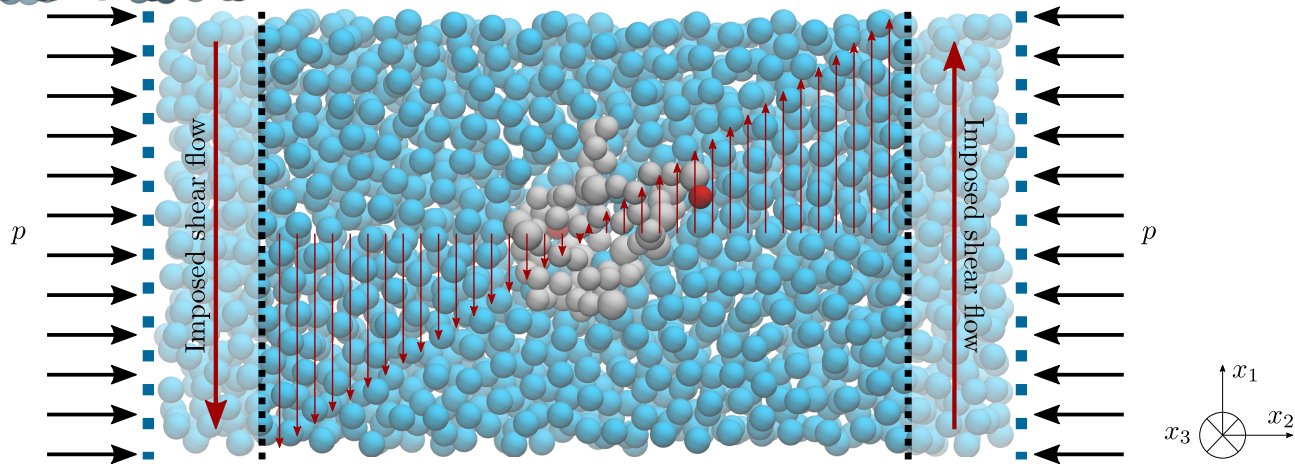
Open Boundary Molecular Dynamics



Zavadlav, Sablić, Podgornik, Praprotnik; *Biophys. J.* (2018)

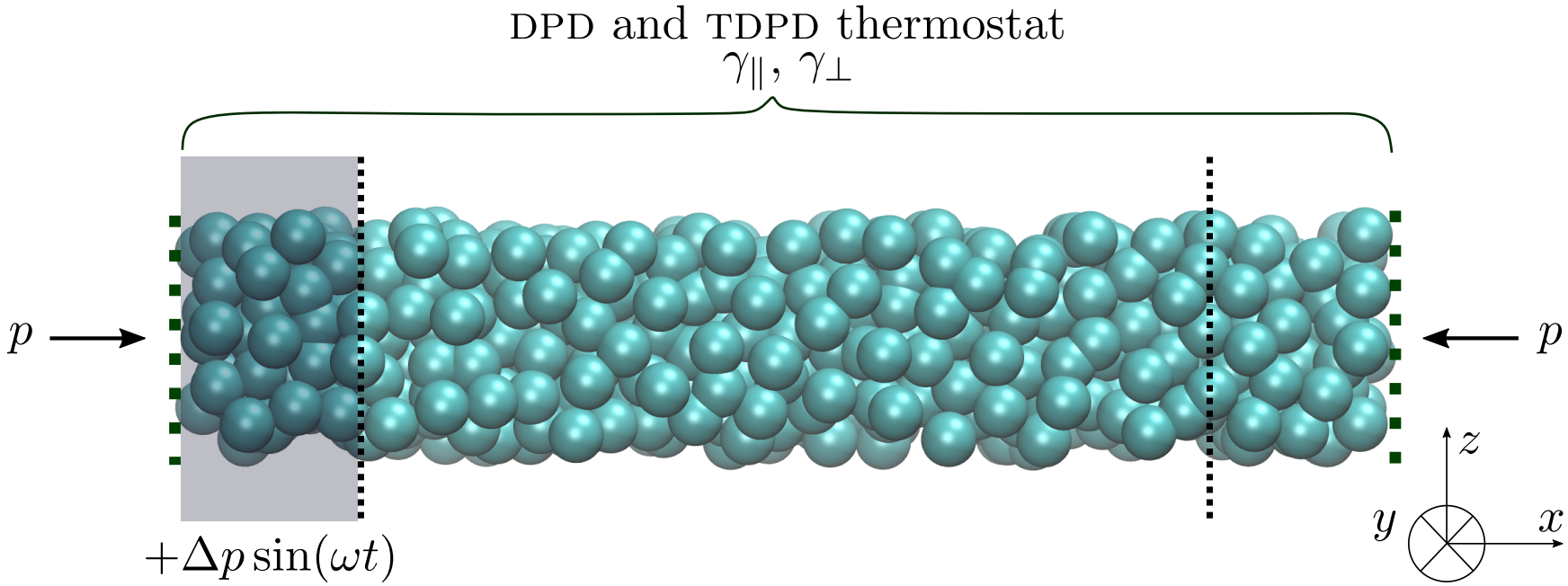
Rotational dynamics of a protein under shear flow





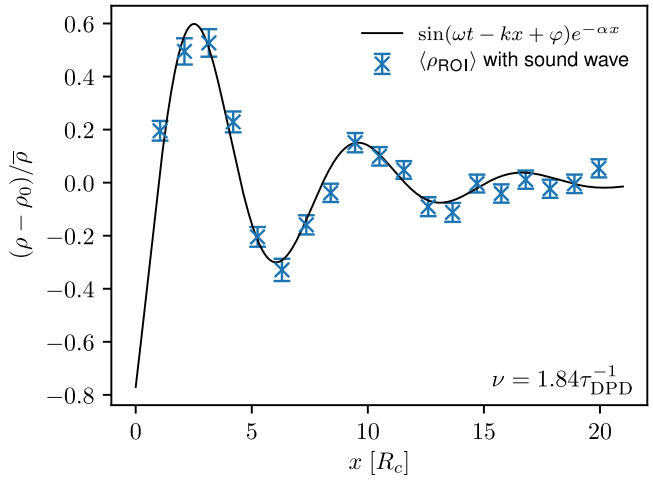
Papež, Merzel, Praprotnik; *J. Phys. Chem B* (2023)

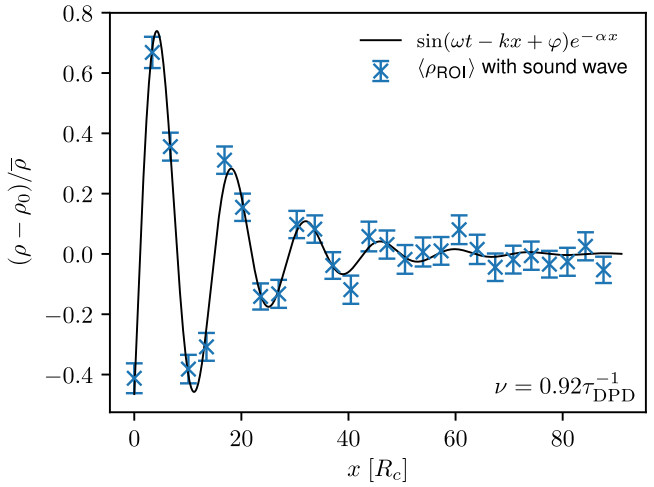
Virtual ultrasound machine

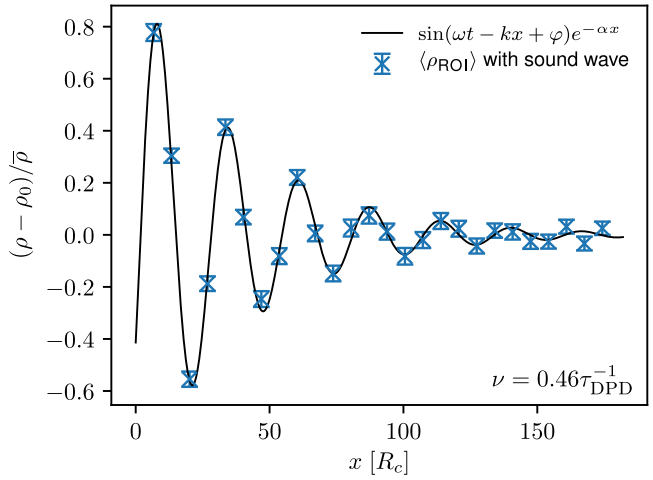


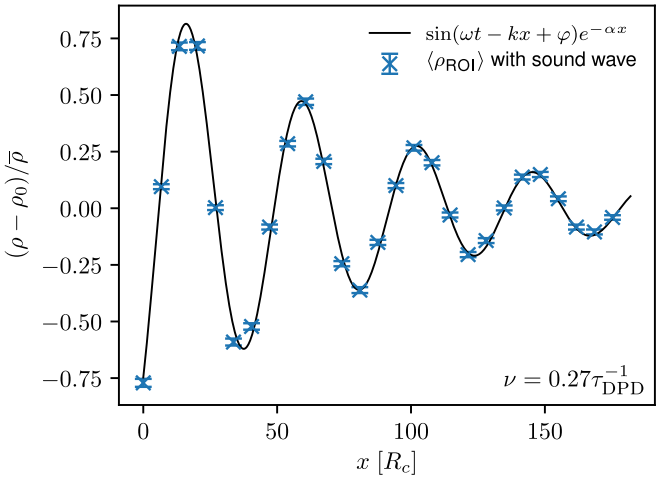
Papež, Praprotnik; J. Chem. Theory Comput. (2022)

Ultrasound propagation through liquid water



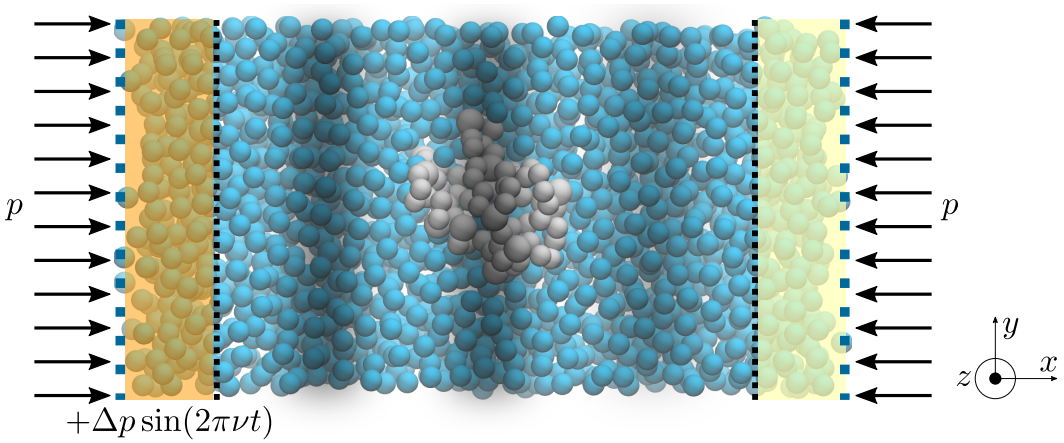






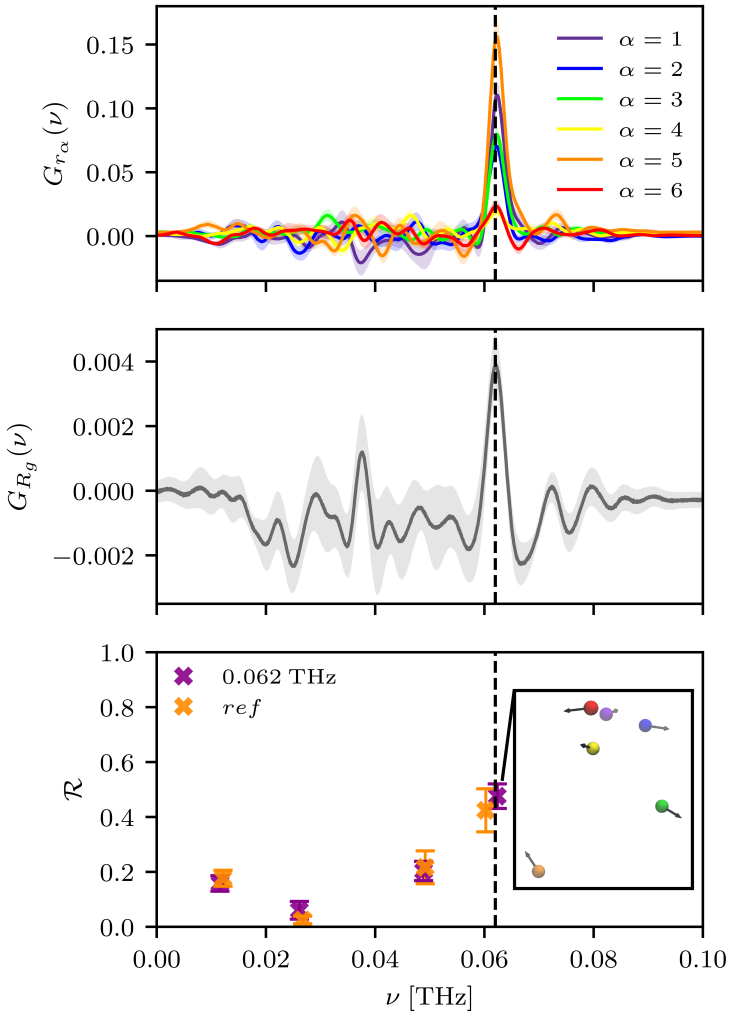
Papež, Praprotnik; J. Chem. Theory Comput. (2022)

Sub-THz acoustic excitation of protein motion



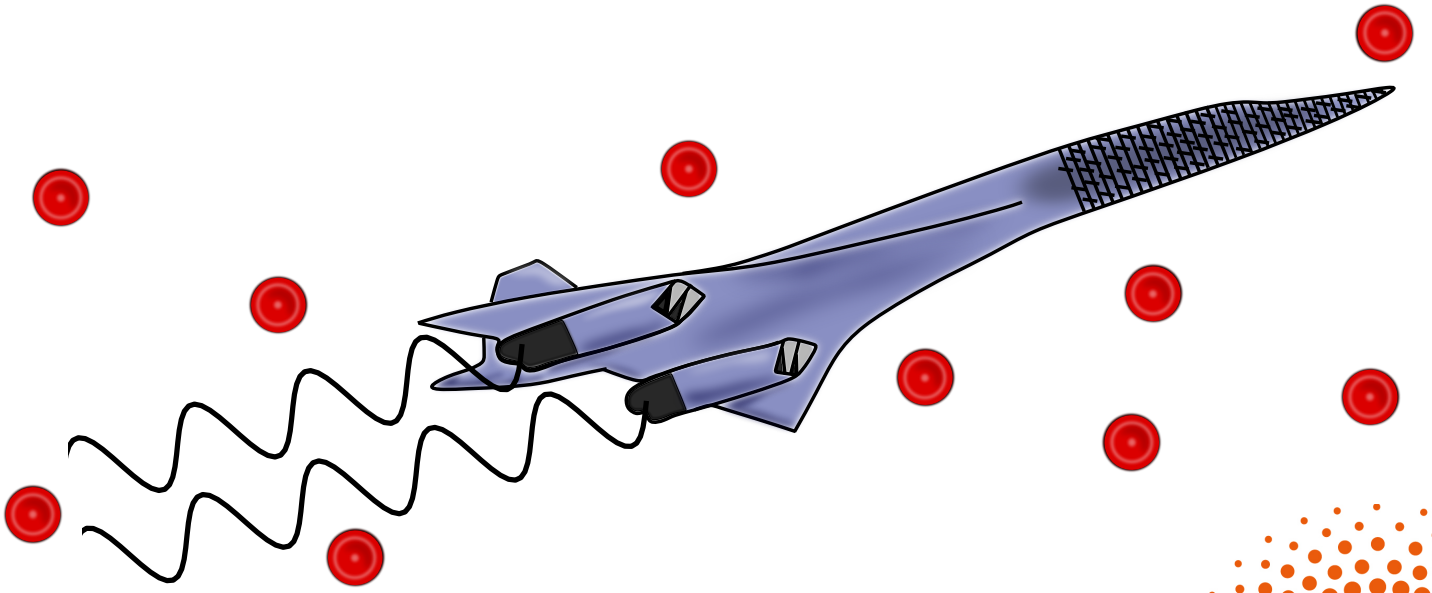
$$g_{R_g/r_\alpha}(\nu) = \frac{1}{2\pi}\int\int V_{g/\alpha}(\tau+t)V_{g/\alpha}(\tau)d\tau e^{i2\pi\nu t}$$

$$G_{R_g/r_\alpha}(\nu) = g_{R_g/r_\alpha}(\nu) - g^0_{R_g/r_\alpha}(\nu)$$



Papež, Merzel, Praprotnik; *J. Chem. Phys.* (2023)

ERC AdG 2019: MULTraSonicA

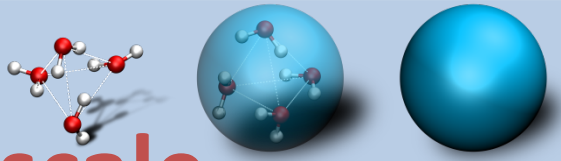


NATIONAL INSTITUTE OF CHEMISTRY

European Research Council

Established by the European Commission

EuroHPC JU CoE: MultiXscale



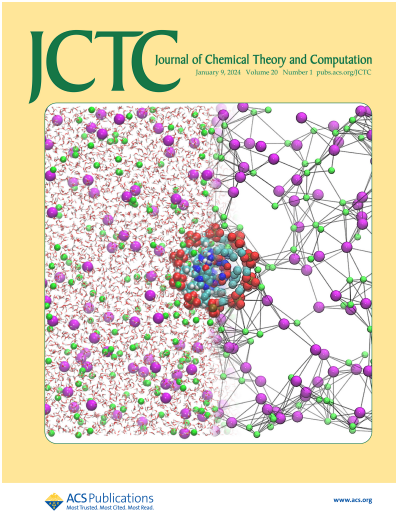
https://www.multixscale.eu

NATIONAL INSTITUTE OF CHEMISTRY

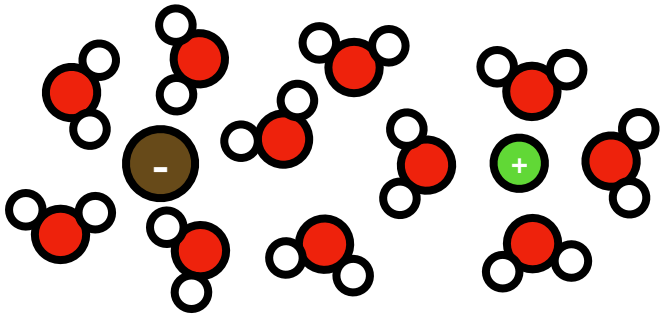
Deep Implicit Solvation (DIS) model

- Delta learning approach :

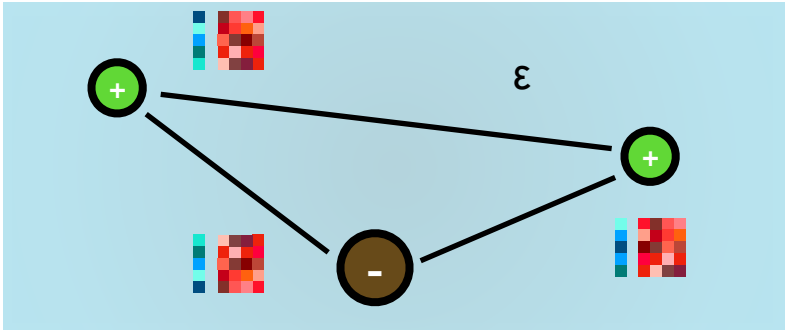
    1. Prior model: physics-based potential, captures the long-range interactions
    2. Machine Learning Potential : Equivariant neural potential (Allegro)



All-atom with explicit hydration model

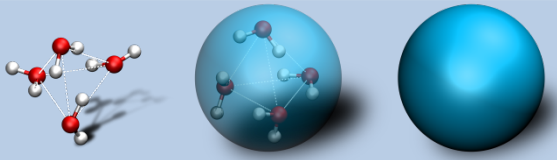


DIS model



Coste, Slejko, Zavadlav, Praprotnik; J. Chem. Theory Comput. (2024)

The DIS model



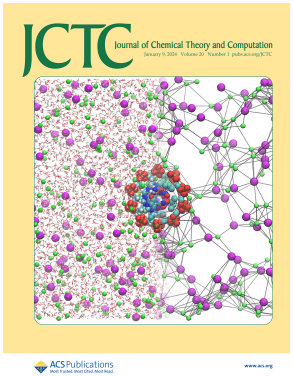
- Prior model: physics-based potential, captures the long-range interactions
- ML potential: equivariant potential (Allegro)
  - $E = \sum_i \sum_{j \in \mathcal{N}(i)} E_{ij}$
  - Tensor product is use to update the information

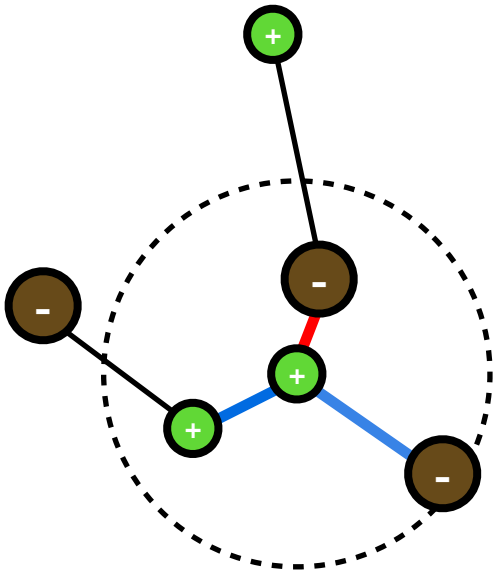
Pair feature Environment feature

$$\mathbf{V}^{ij,L} = \mathbf{V}^{ij,L-1} \otimes \left(\sum_{k \in \mathcal{N}(i)} w^{ik,L} \vec{Y}^{ik} \right)$$

- Delta learning approach :

$$L = \frac{1}{3N_{\text{data}}N_{\text{ions}}} \sum_{i=1}^{N_{\text{data}}} \sum_{j=1}^{N_{\text{ions}}} \sum_{k=1}^{3} \| F_{ijk}^{\text{ML}} - (F_{ijk}^{\text{all-atom}} - F_{ijk}^{\text{prior}}) \|$$

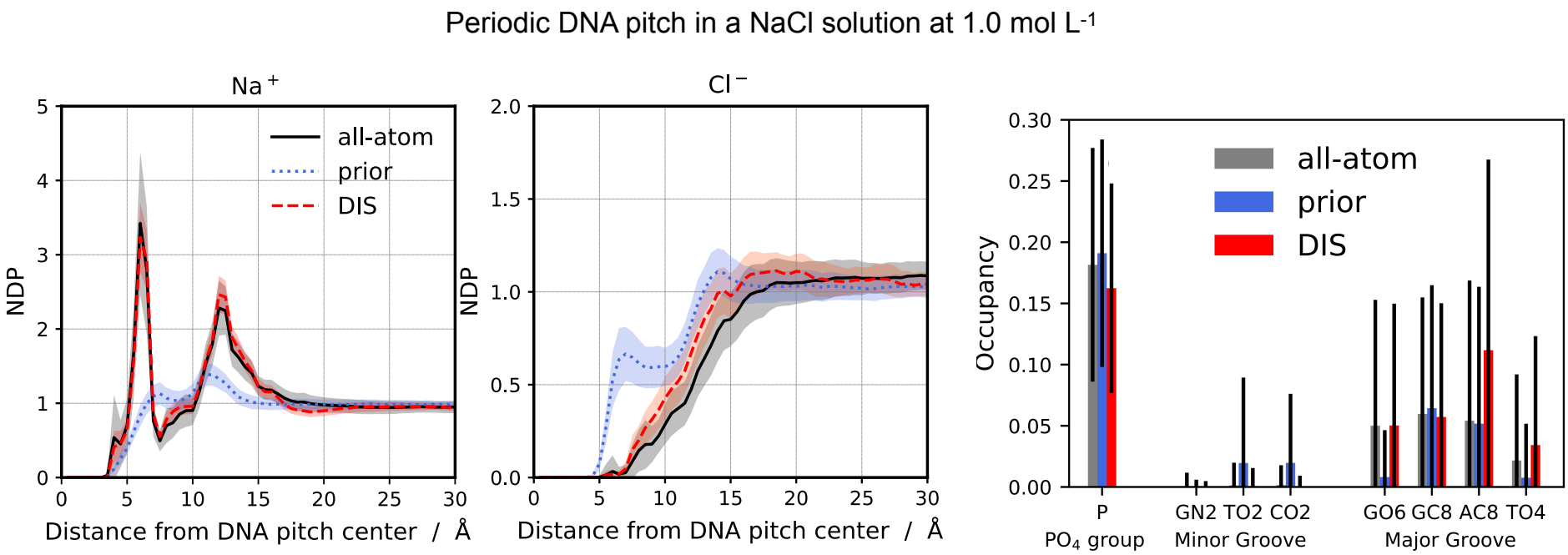




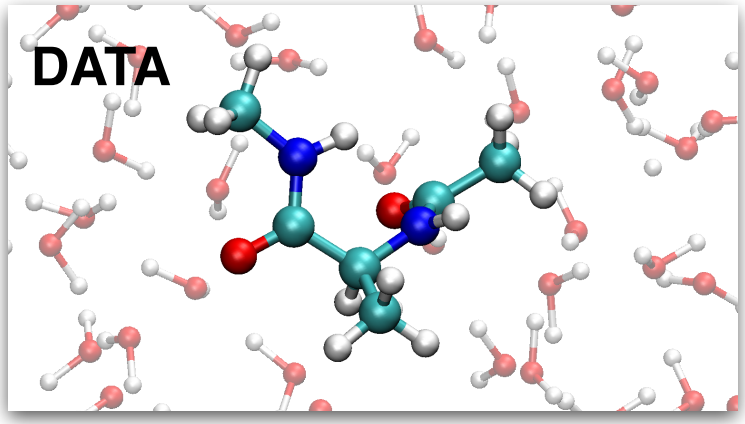
Coste, Sjleko, Zavadlav, Praprotnik; J. Chem. Theory Comput. (2024)
Musaelian, A., Batzner, S., Johansson, A., Sun L., Owen C. J., Kornbluth M., Kozinsky B.; Nat Commun (2023)

Results: The DIS model

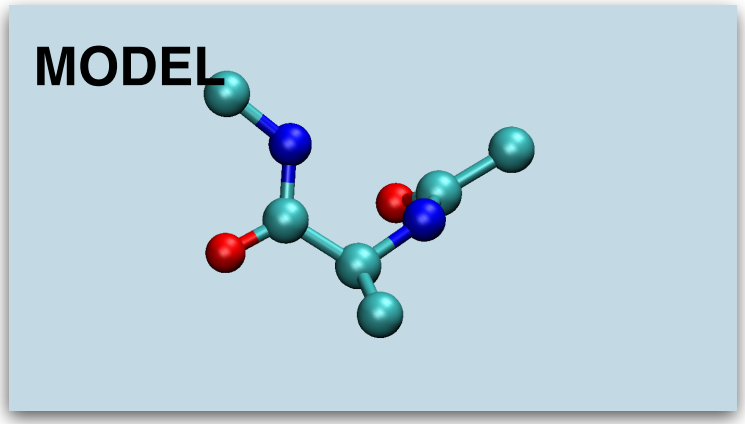
- DIS model captures the effective ion interactions in the vicinity and far away from a DNA molecule, while implicit solvent treatment renders the model inexpensive



Alanine Dipeptide



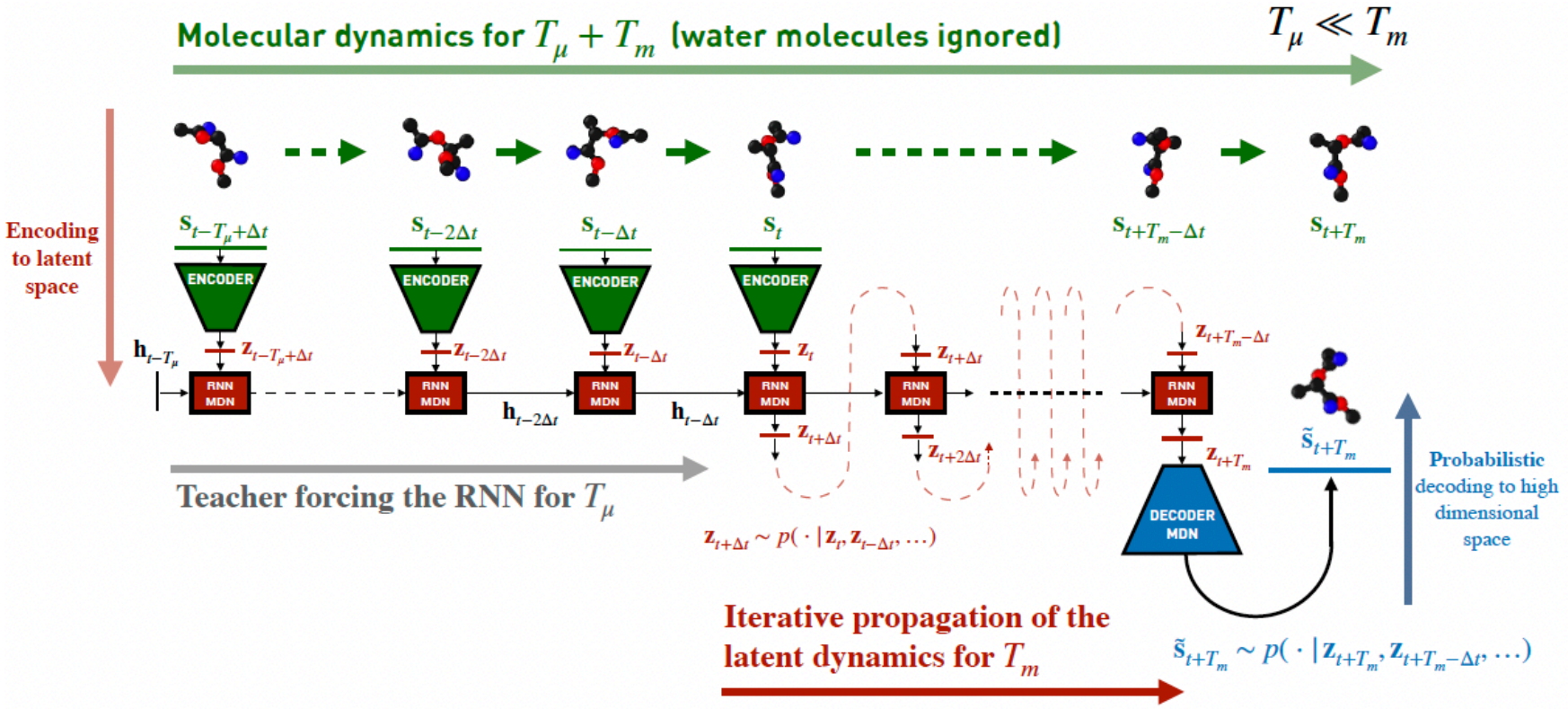
Machine



atomistic simulation
in explicit water

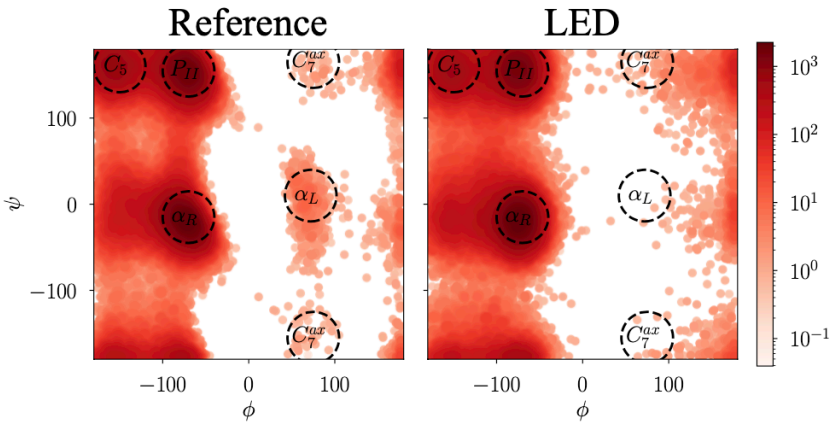
Coarse-grained representation:
no hydrogens, implicit water
internal coordinates: bonds, angles, dihedrals

ML as a surrogate model

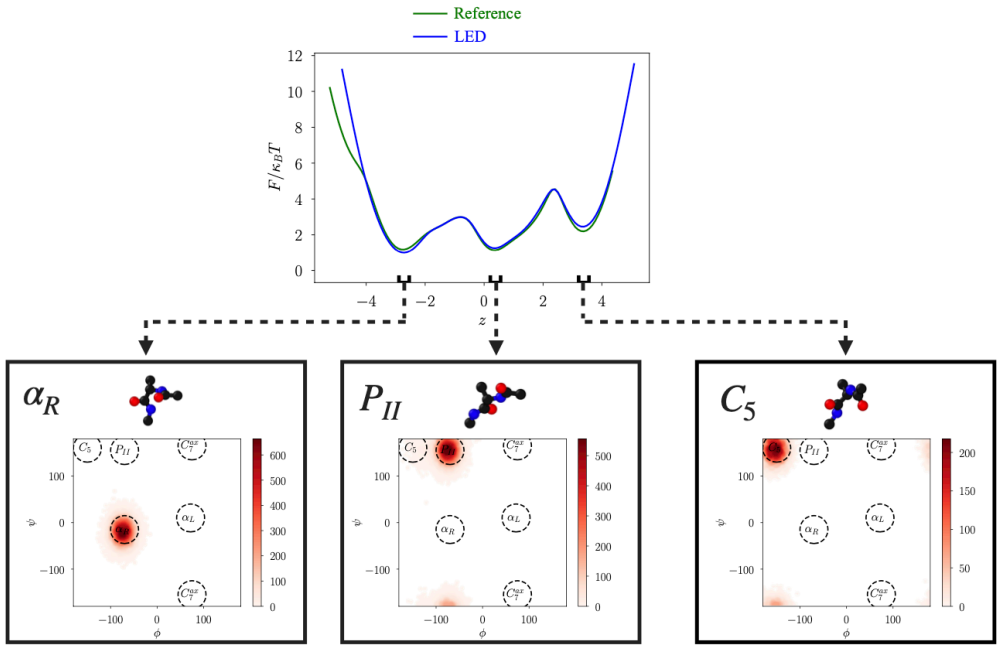


Vlachas, Zavadlav, Praprotnik, Koumoutsakos, *J. Chem. Theory Comput.* 18, 538-549, 2022

Learning Effective Dynamics (LED) of Alanine Dipeptide



MFPT [ns]	MSM – 10ps on MD data Reference	MSM – 10ps on LED – 0.1ps data MFPT	Error (%)
$T_{C_5 \to P_{II}}$	0.105	0.103	**2**
$T_{C_5 \to \alpha_R}$	0.104	0.082	**21**
$T_{P_{II} \to C_5}$	0.226	0.242	**7**
$T_{P_{II} \to \alpha_R}$	0.105	0.083	**21**
$T_{\alpha_R \to C_5}$	0.236	0.258	**9**
$T_{\alpha_R \to P_{II}}$	0.116	0.119	**2**



Vlachas, Zavadlav, Praprotnik, Koumoutsakos, *J. Chem. Theory Comput.* 18, 538-549, 2022

Acknowledgements

- **Amaury Coste**; National Institute of Chemistry, Slovenia
- **Ema Slejko**; National Institute of Chemistry, Slovenia
- **Petra Papež**; National Institute of Chemistry, Slovenia
- **Franci Merzel**; National Institute of Chemistry, Slovenia
- **Julija Zavadlav**; TUM, Germany
- **Petros Koumoutsakos**; Harvard, USA

- **Slovenian Research Agency** for funding