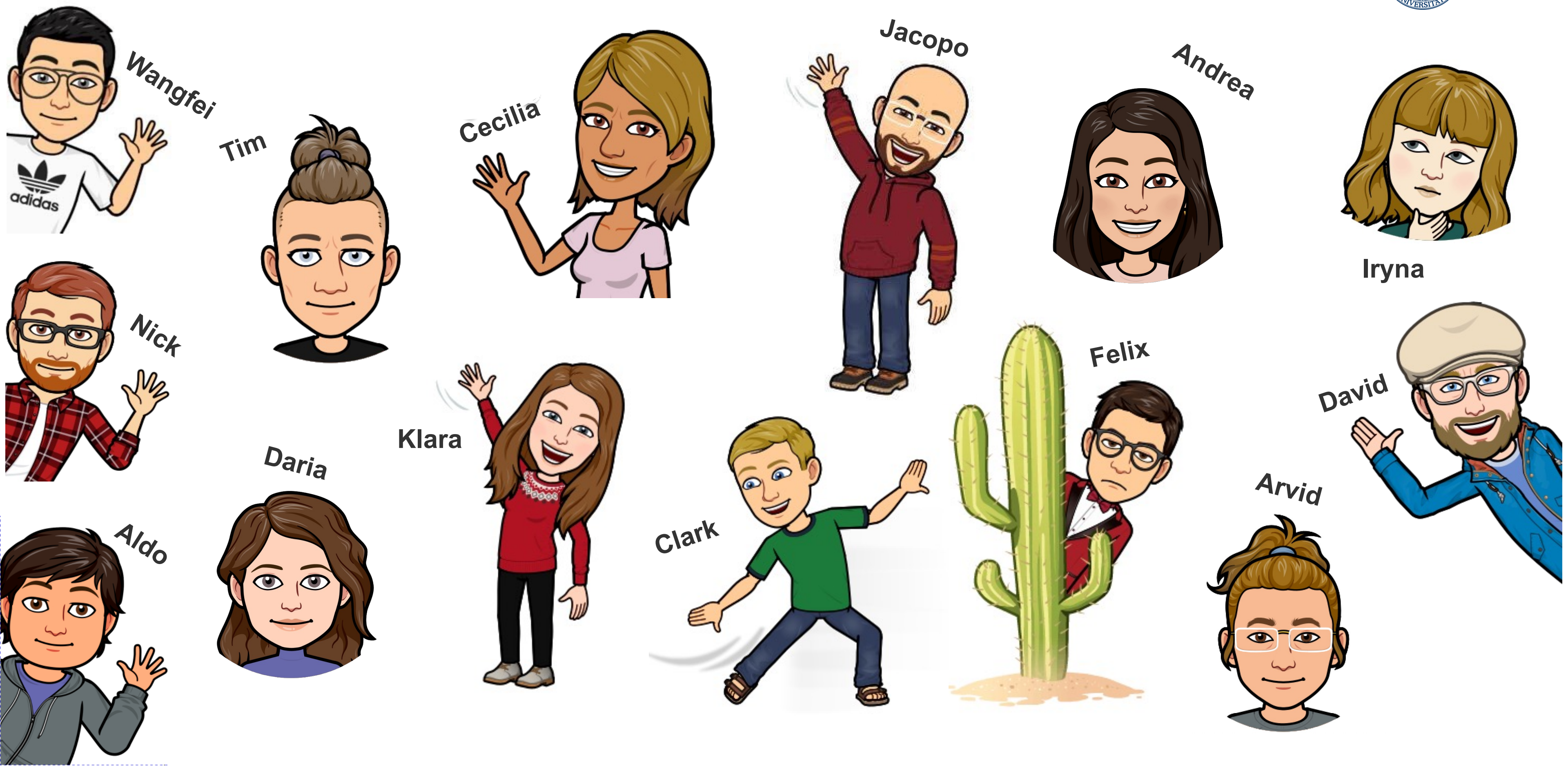
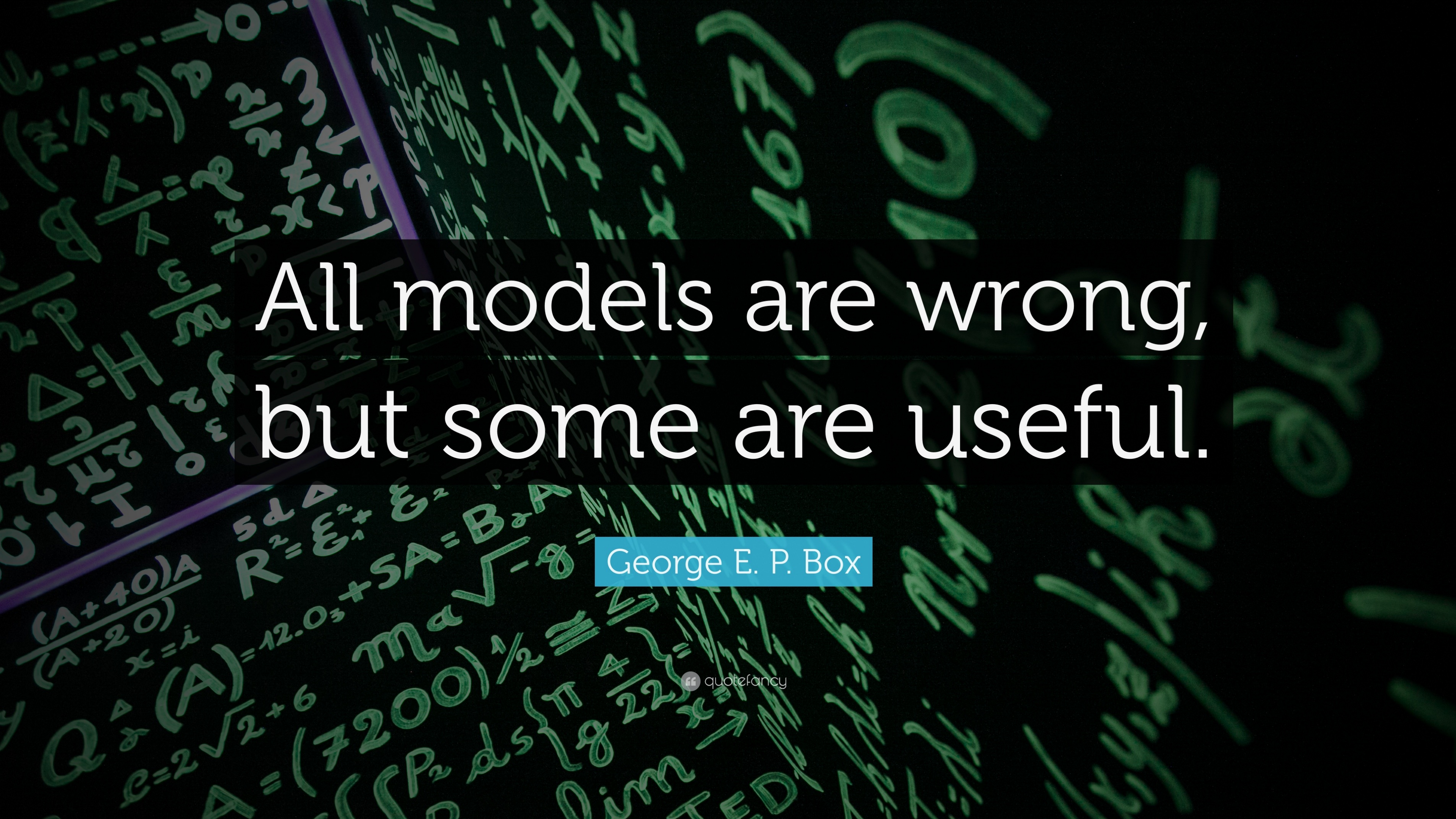
Coarse-graining for classical and quantum systems

Cecilia Clementi
Department of Physics
Freie Universität Berlin
cecilia.clementi@fu-berlin.de

Clementi Group - Theoretical and Computational Biophysics

Freie Universität Berlin

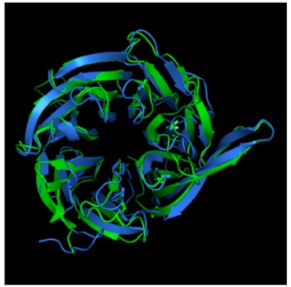



All models are wrong,
but some are useful.
George E. P. Box

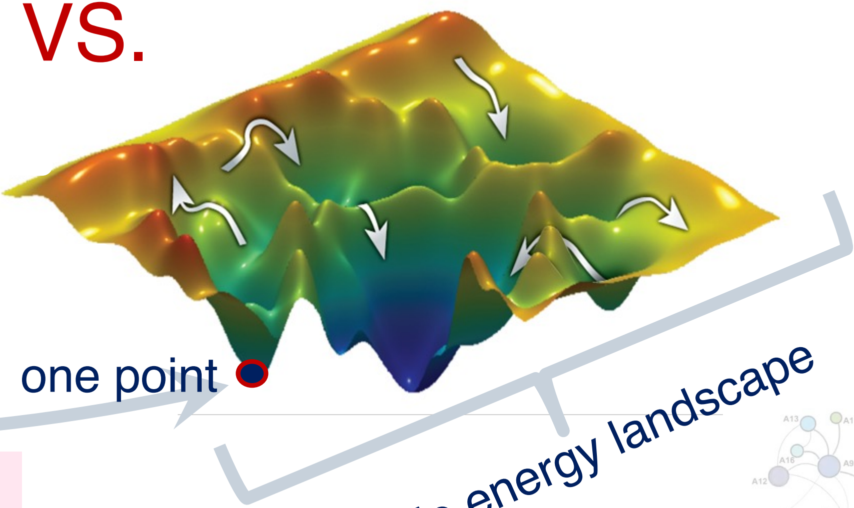
Ultimate goal

AlphaFold2@DeepMind
predicts
static structures
from sequences

Structures:
Ground truth (green)
Predicted (blue)



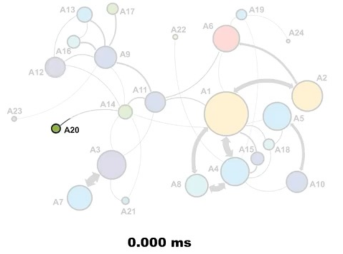
VS.

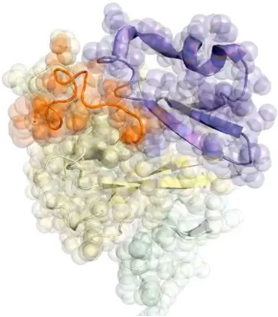


one point

the whole energy landscape

AlphaFold: a solution to a 50-year-old grand challenge in biology



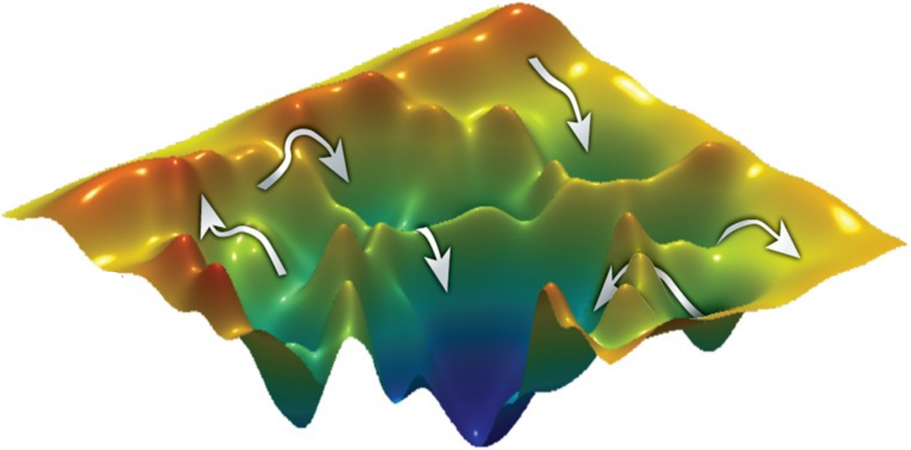


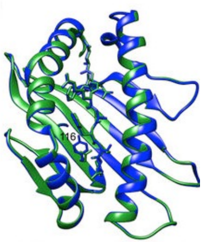
Biological function requires dynamics

Ultimate goal

Freie Universität Berlin

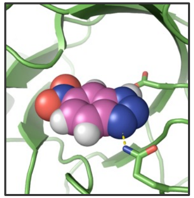
the whole energy landscape



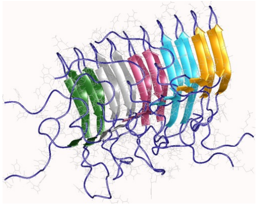


Vaccines

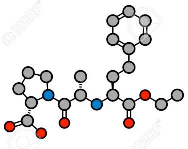
Antibodies



Enzymes



Amyloids



Drugs

Main challenge: wide range of spatiotemporal scales

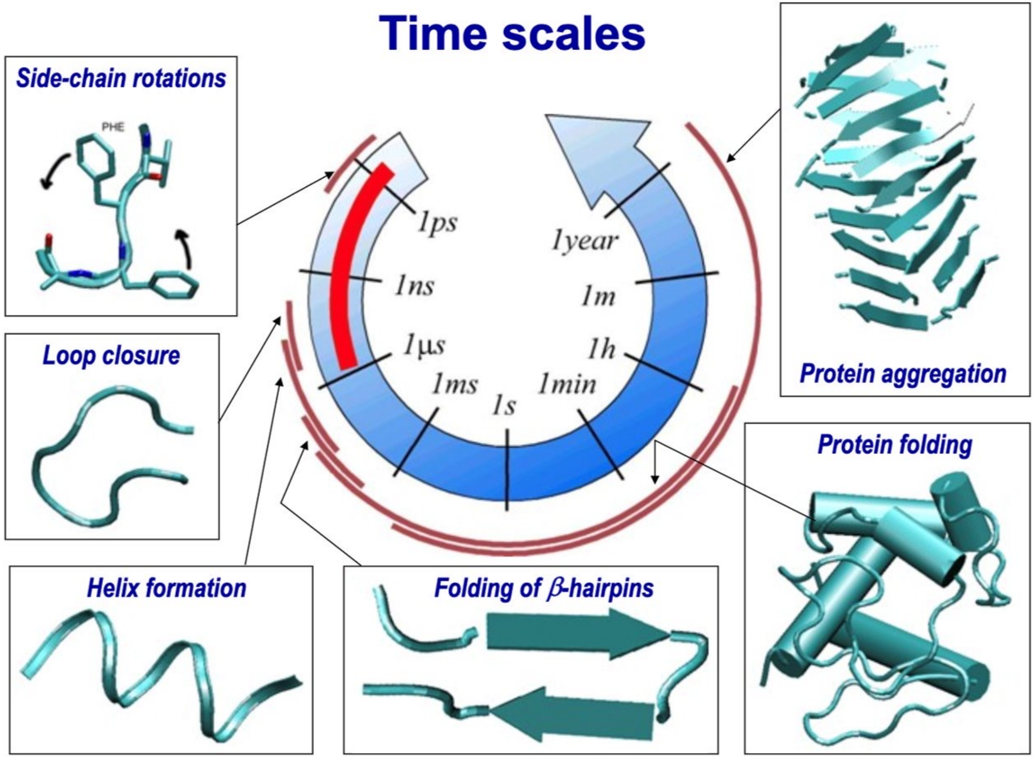
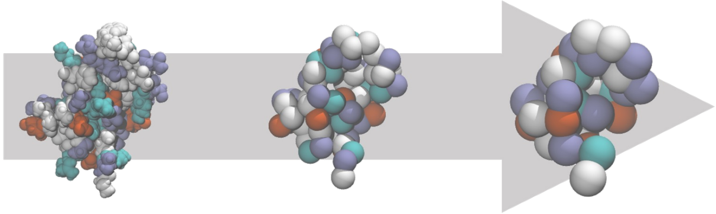


Figure from J.E. Shea

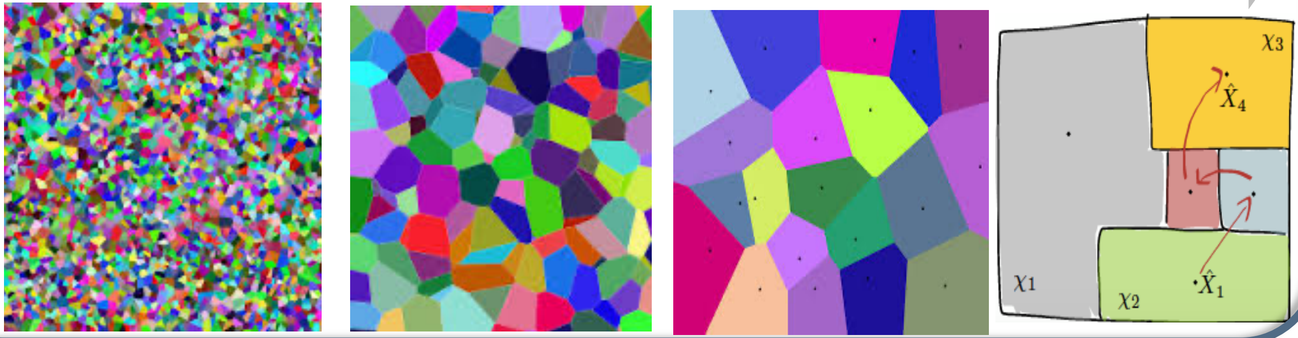
Coarse-graining idea:

Can we design a **minimalist** model speeding up the simulations significantly while preserving the relevant physics?

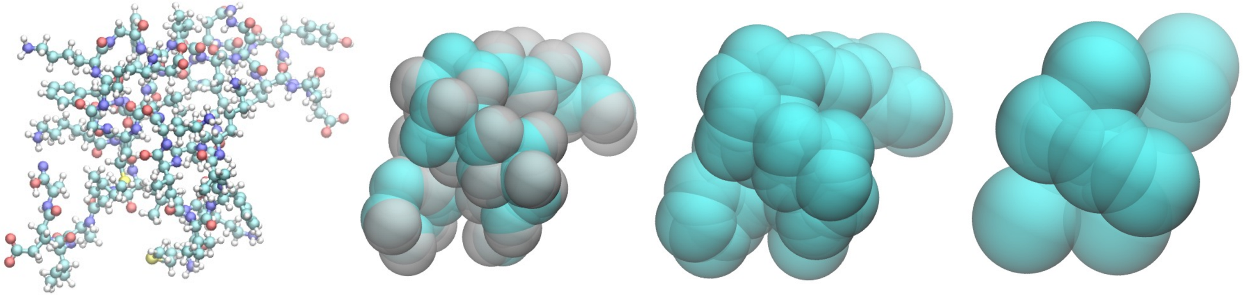


Coarse-Graining

$\mathbb{R}^{3N}$ coarse-graining in conformation space



$\mathbb{R}^{3}$ coarse-graining in structural space



Coarse-graining in conformation space

Available online at www.sciencedirect.com

ScienceDirect

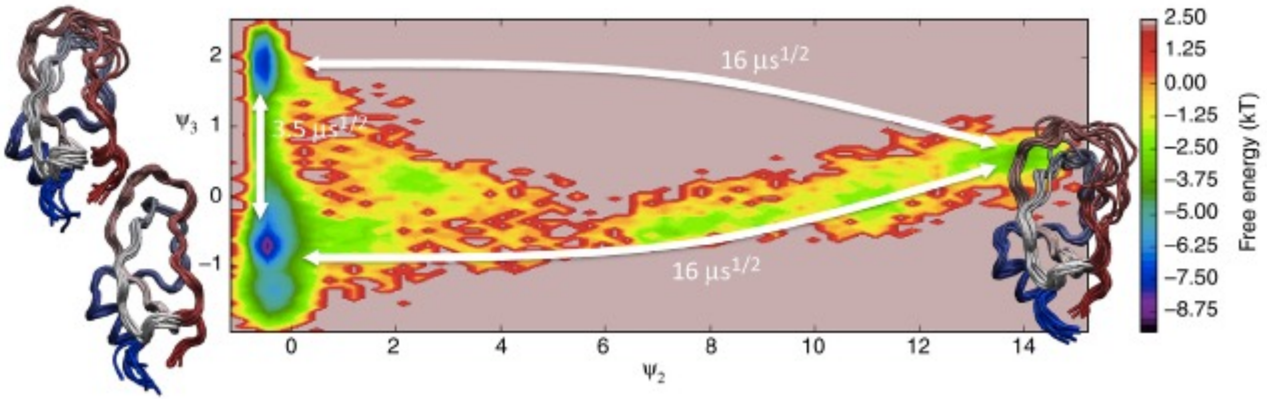
ELSEVIER

Current Opinion in Structural Biology

Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods

Frank Noé[1] and Cecilia Clementi[2]

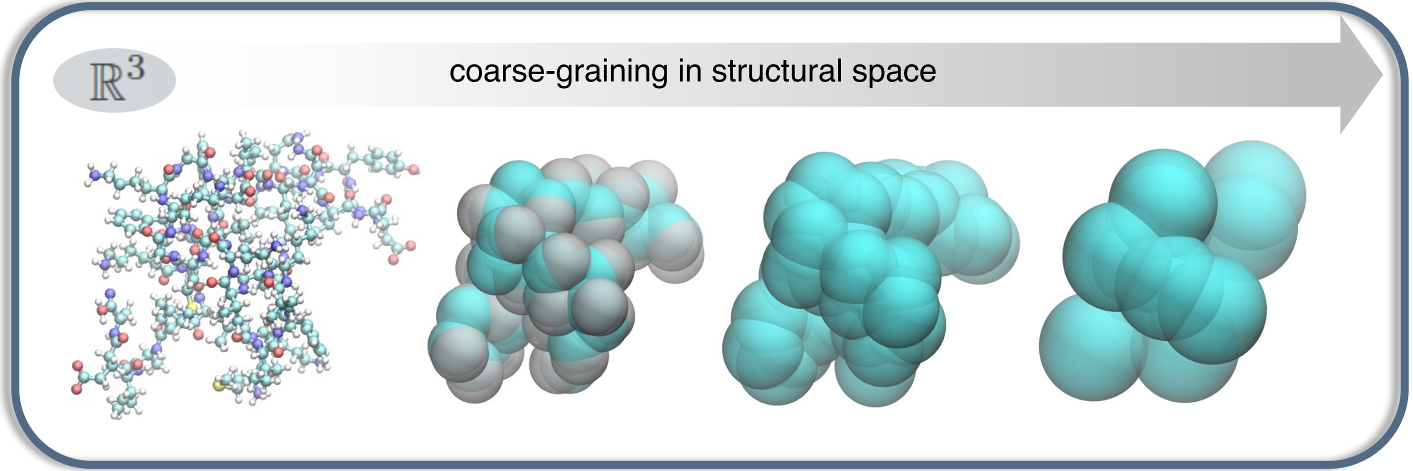
CrossMark



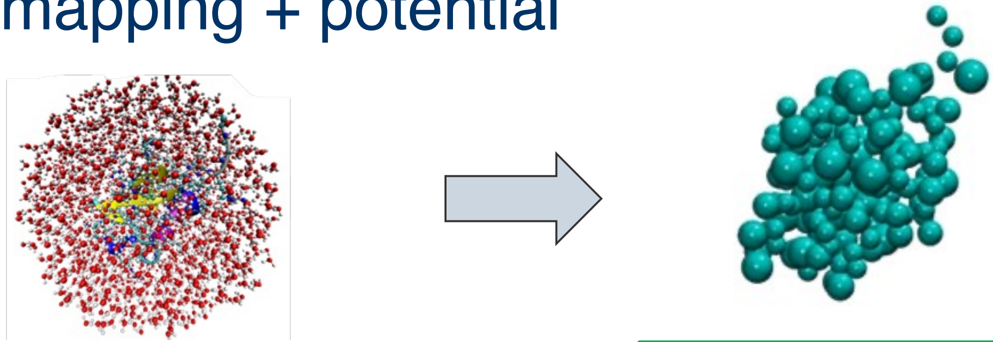
www.sciencedirect.com

Current Opinion in Structural Biology 2017, 43:141–147

Coarse-Graining

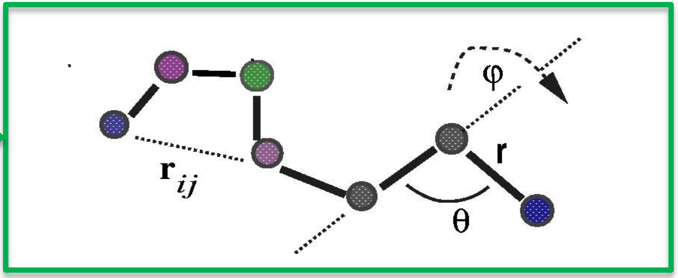


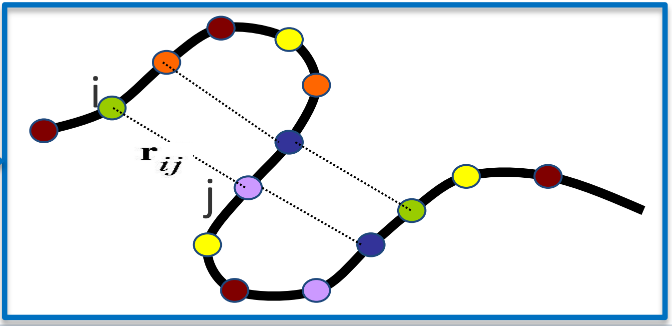
CG model = mapping + potential



$$\mathbf{H} = \sum_{\text{bonds}} K_r (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_d (1 + \cos(n(\phi - \phi_0))) +$$

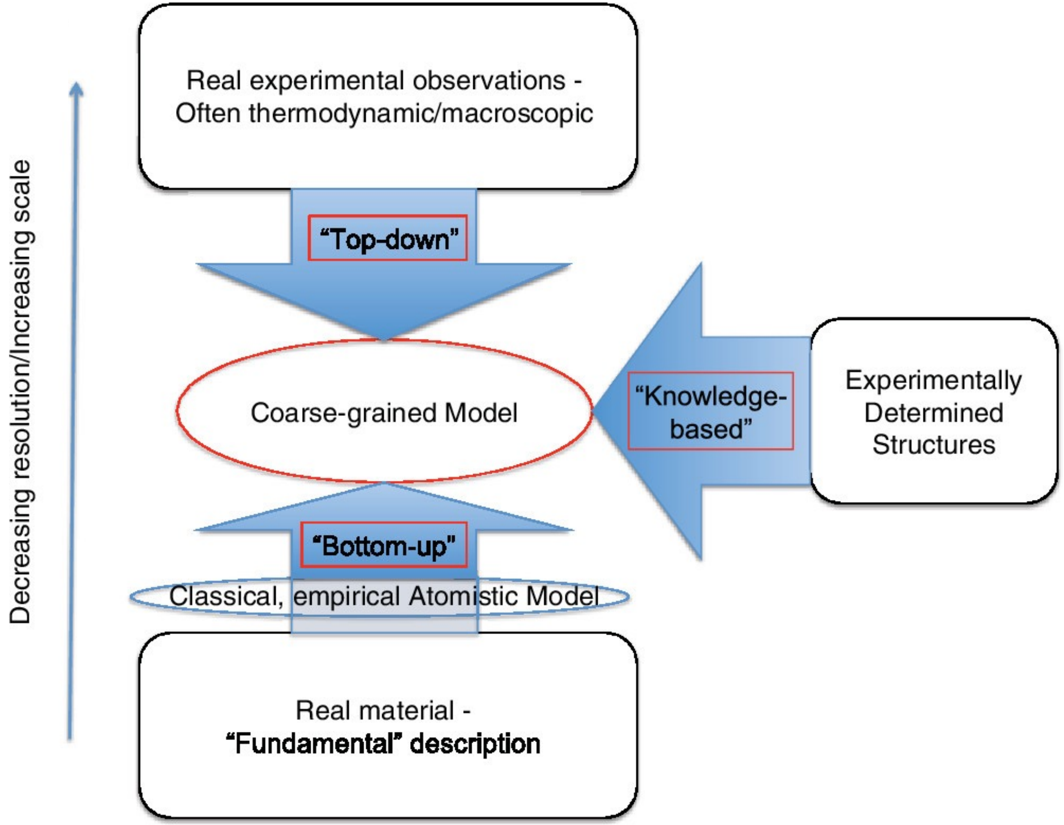
$$\sum_{i,j} \epsilon_{ij} \left[5 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right]$$





C. Clementi, *et al*, *J. Mol. Biol.* 298, 937-953 (2000)

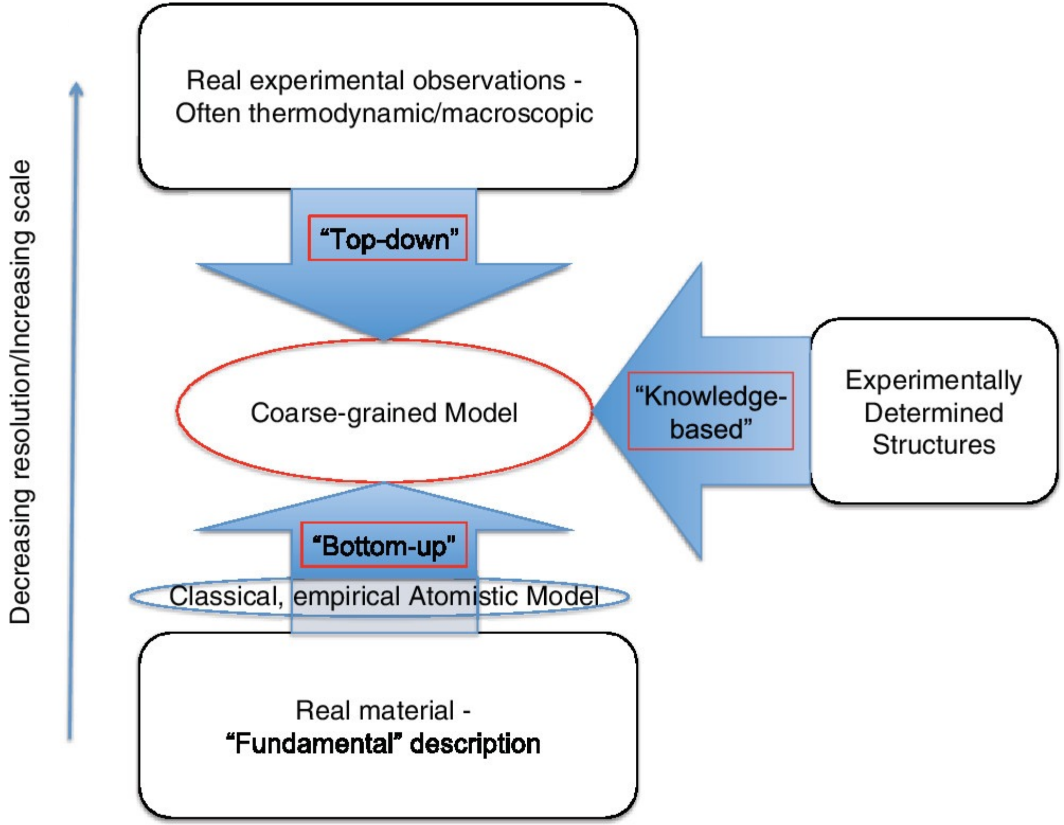
CG models: which properties to preserve?



1. Learning CG models from all-atom simulations
2. Incorporating nuclear quantum effects into all-atom simulation
3. Learning CG models from experimental data

W.G. Noid, *J. Chem. Phys.* 139:9, 090901 (2013)

CG models: which properties to preserve?



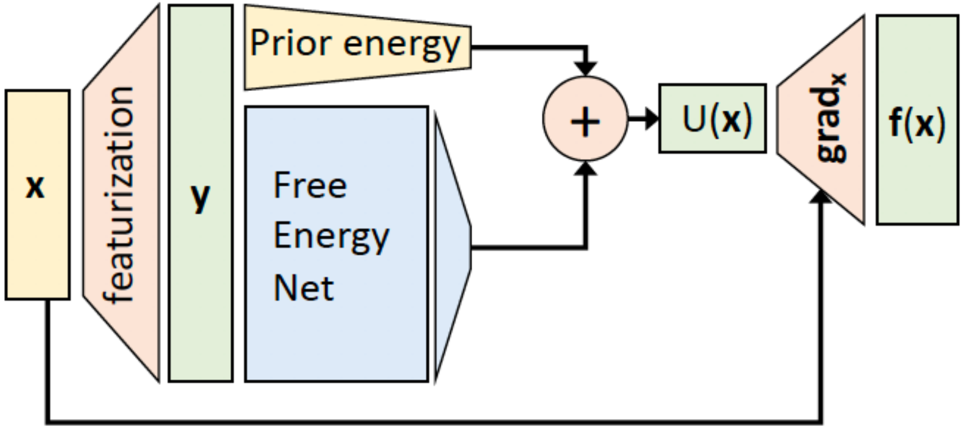
1. **Learning CG models from all-atom simulations**
2. Incorporating nuclear quantum effects into all-atom simulation
3. Learning CG models from experimental data

W.G. Noid, *J. Chem. Phys.* 139:9, 090901 (2013)

Force-matching

Enforcing the thermodynamic consistency is equivalent to minimize the force matching error:

$$\chi^2(\boldsymbol{\theta}) = \left\langle \|\xi(\mathbf{F}(\mathbf{r})) + \nabla U(\xi(\mathbf{r}); \boldsymbol{\theta})\|^2 \right\rangle_{\mathbf{r}}$$



S. Izvekov and G. A. Voth *J. Phys. Chem. B* 109:2469–2473 (2005)
W. G. Noid, et al, J. Chem. Phys. 128, 244114 (2008)

J. Wang, S. Olsson, C. Wehmeyer, A. Perez, N.E. Charron, G. de Fabritiis, F. Noé, C. Clementi, *ACS Cent. Sci*, 5, 755-767 (2019)

Force-matching

Enforcing the thermodynamic consistency is equivalent to minimizing the **force matching error**:

Instantaneous all-atom force **Model parameters**

$$\chi^2(\theta) = \left\langle \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) + \nabla U(\xi_r(\boldsymbol{r}); \theta) \right\|^2 \right\rangle_{\boldsymbol{r}}$$

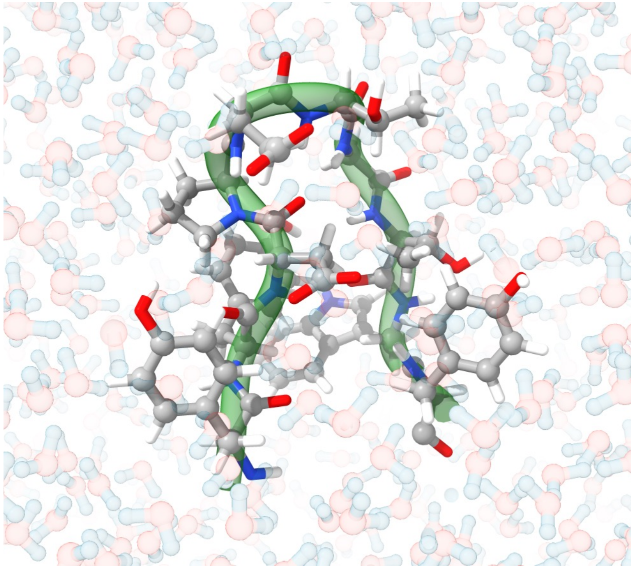
CG force mapping **CG configurational mapping**

S. Izvekov and G. A. Voth *J. Phys. Chem. B* 109:2469–2473 (2005)
W. G. Noid, et al, J. Chem. Phys. 128, 244114 (2008)

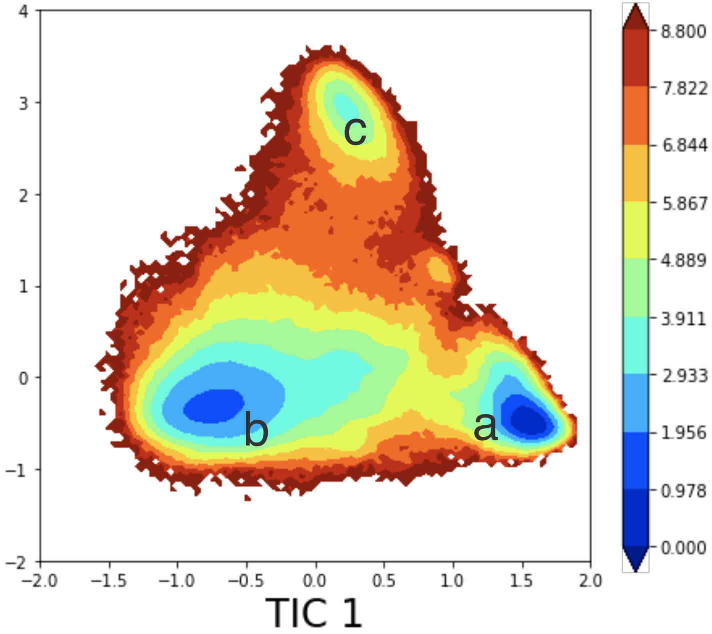
J. Wang, S. Olsson, C. Wehmeyer, A. Perez, N.E. Charron, G. de Fabritiis, F. Noé, C. Clementi, *ACS Cent. Sci*, 5, 755-767 (2019)

Test system: Chignolin folding/unfolding/misfolding

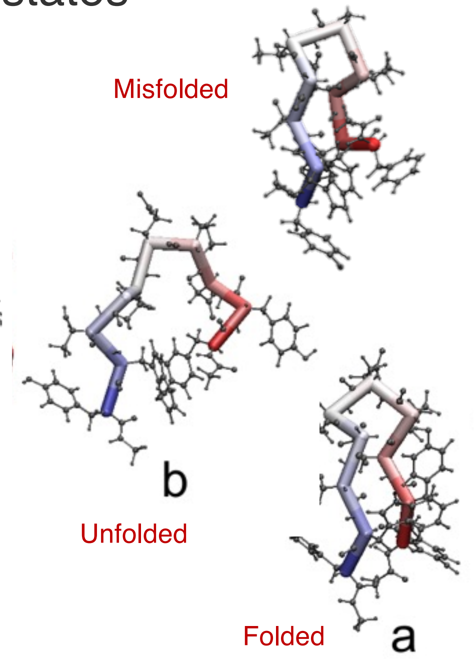
atomistic simulation in explicit solvent



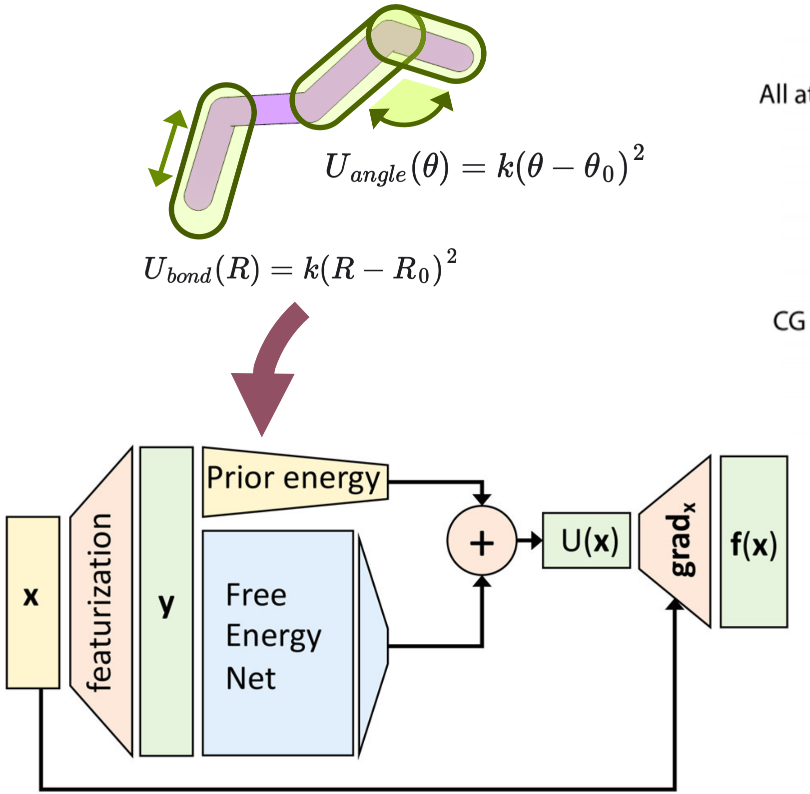
reference free energy

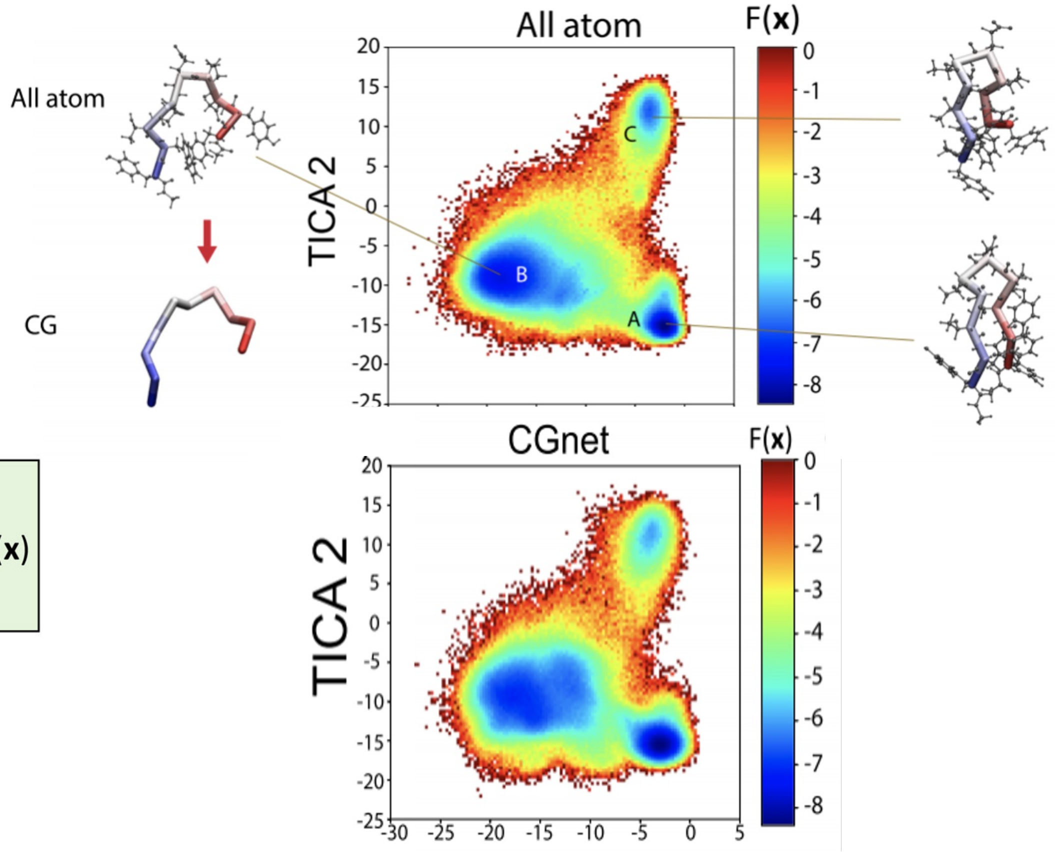


3 metastable states



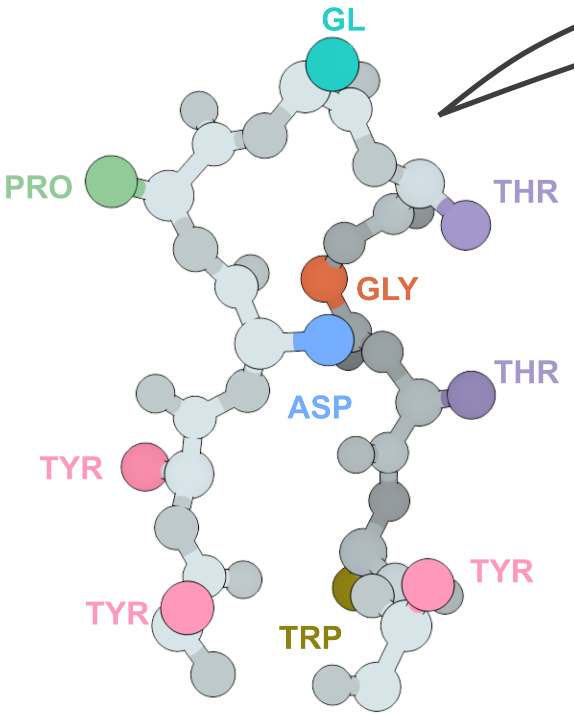
CGNet: Feedforward, Fully-Connected

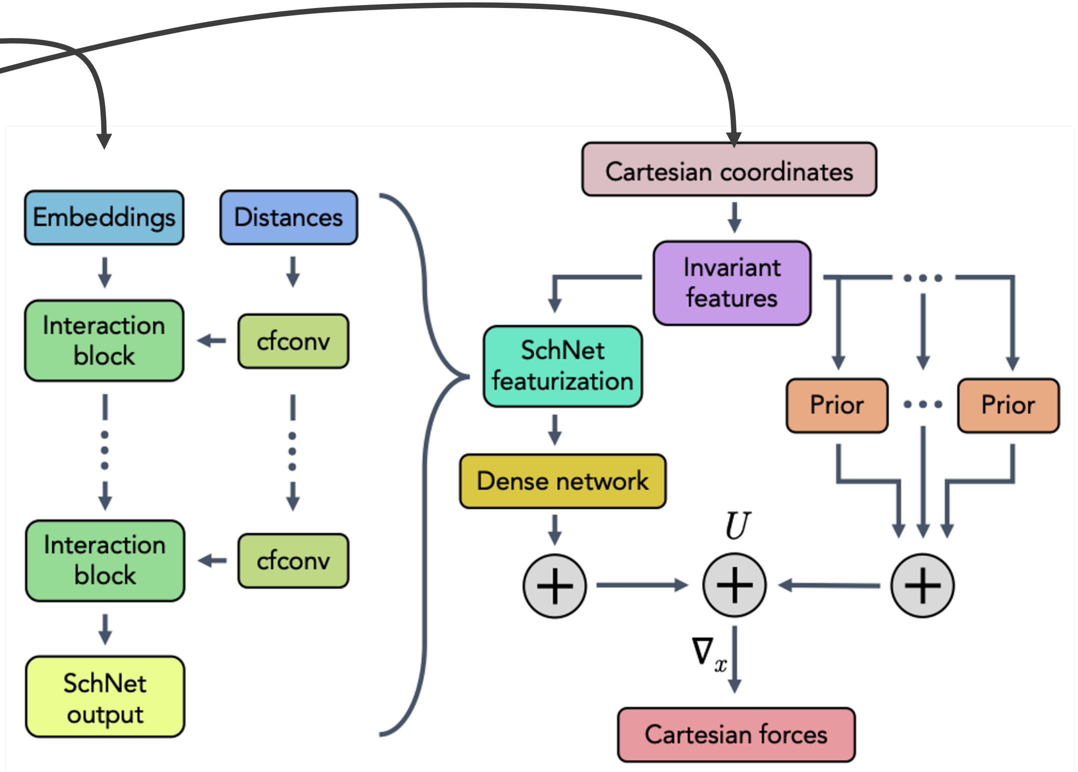




J. Wang, S. Olsson, C. Wehmeyer, A. Perez, N.E. Charron, G. de Fabritiis, F. Noé, C. Clementi, *ACS Cent. Sci*, 5, 755-767 (2019)

CGSchNet: A Message-passing, Graph Network

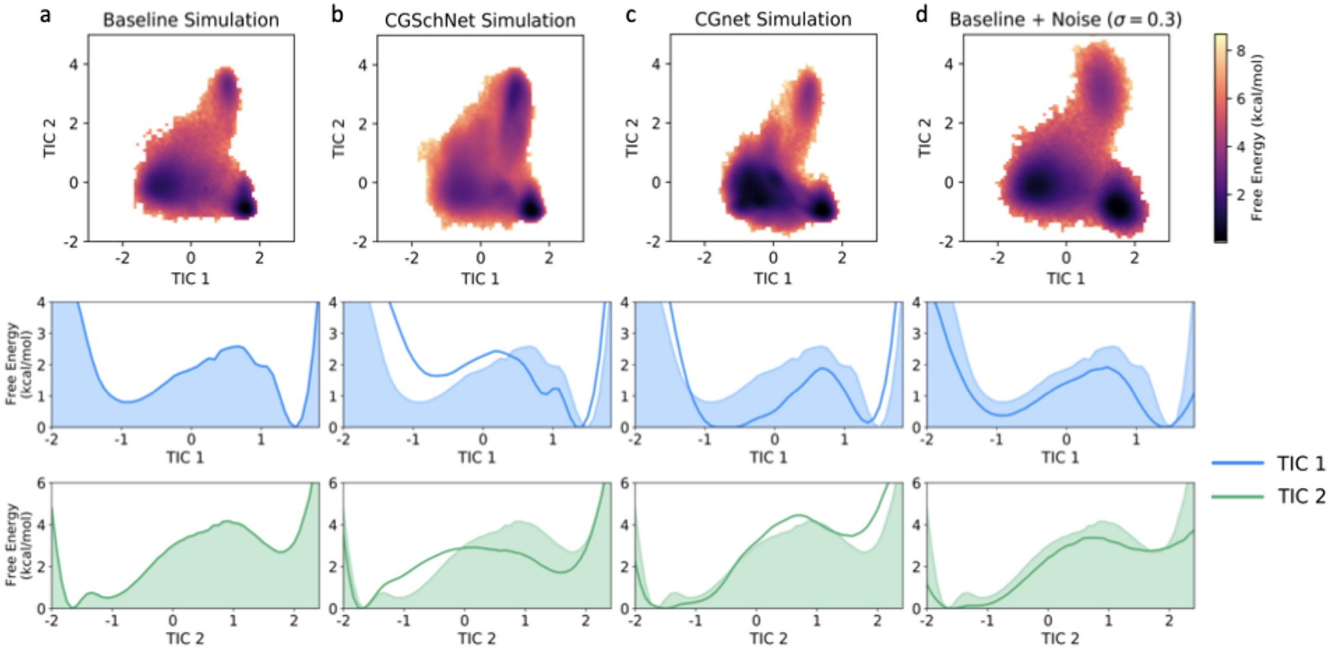


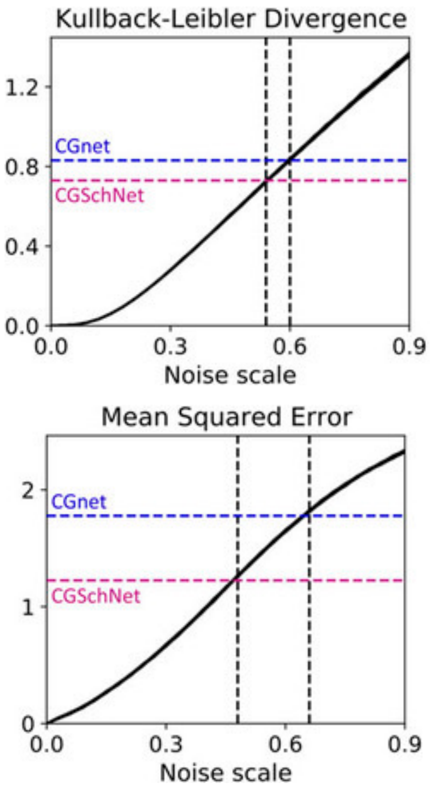


Husic, B. E., Charron, N. E., Lemm, D., Wang, J., Pérez, A., Majewski, M., Krämer, A., Chen, Y., Olsson, S., de Fabritiis, G., Noé, F., & Clementi, C.
Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys. 153*(19), 194101 (2020).

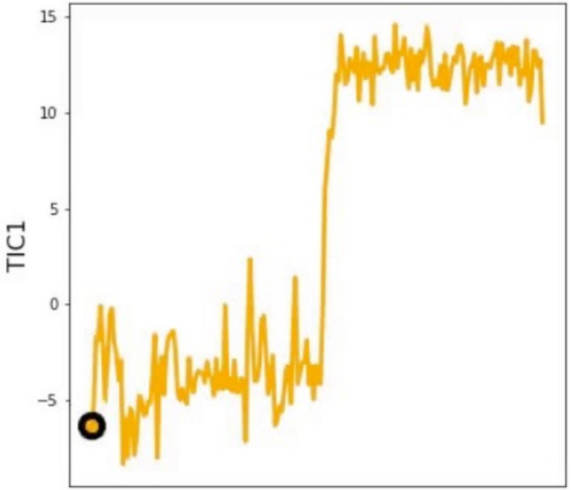
Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A., & Müller, K.-R.
SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys*, *148*(24), 241722 (2018).

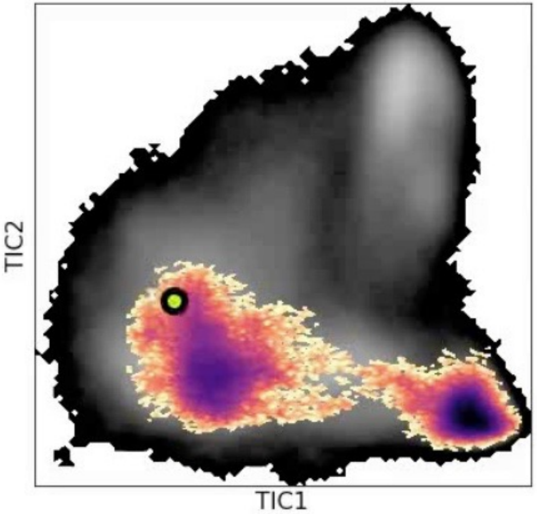
CGSchNet: A Message-passing, Graph Network

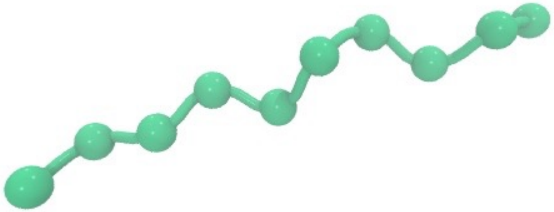




Husic, B. E., Charron, N. E., Lemm, D., Wang, J., Pérez, A., Majewski, M., Krämer, A., Chen, Y., Olsson, S., de Fabritiis, G., Noé, F., & Clementi, C. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys. 153*(19), 194101 (2020).


15
10
5
0
−5
TIC1

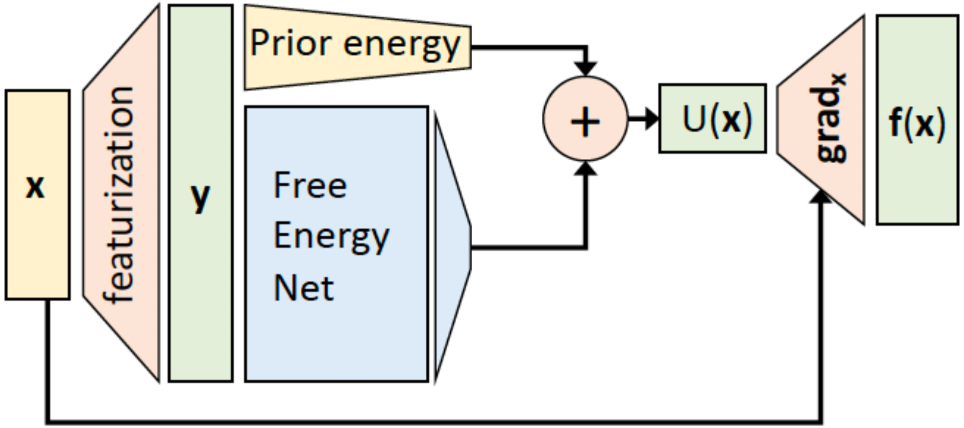

TIC2
TIC1



Force-matching

Enforcing the thermodynamic consistency is equivalent to minimize the force matching error:

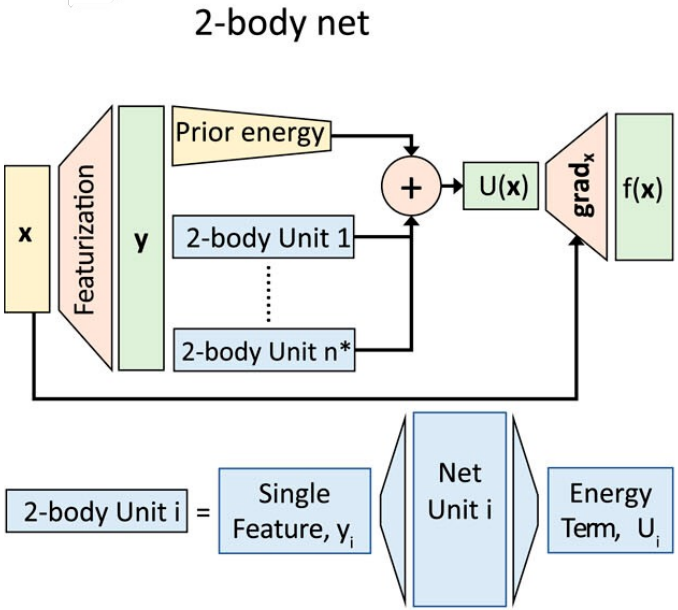
$$\chi^2(\boldsymbol{\theta}) = \left\langle \left\| \xi(\mathbf{F}(\mathbf{r})) + \nabla U(\xi(\mathbf{r}); \boldsymbol{\theta}) \right\|^2 \right\rangle_{\mathbf{r}}$$

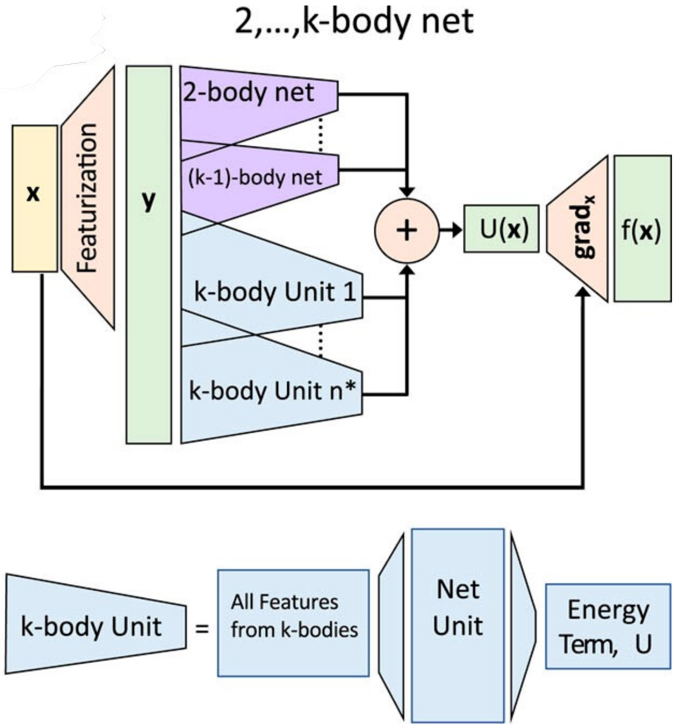


S. Izvekov and G. A. Voth *J. Phys. Chem. B* 109:2469–2473 (2005)
W. G. Noid, et al, J. Chem. Phys. 128, 244114 (2008)

J. Wang, S. Olsson, C. Wehmeyer, A. Perez, N.E. Charron, G. de Fabritiis, F. Noé, C. Clementi, *ACS Cent. Sci*, 5, 755-767 (2019)

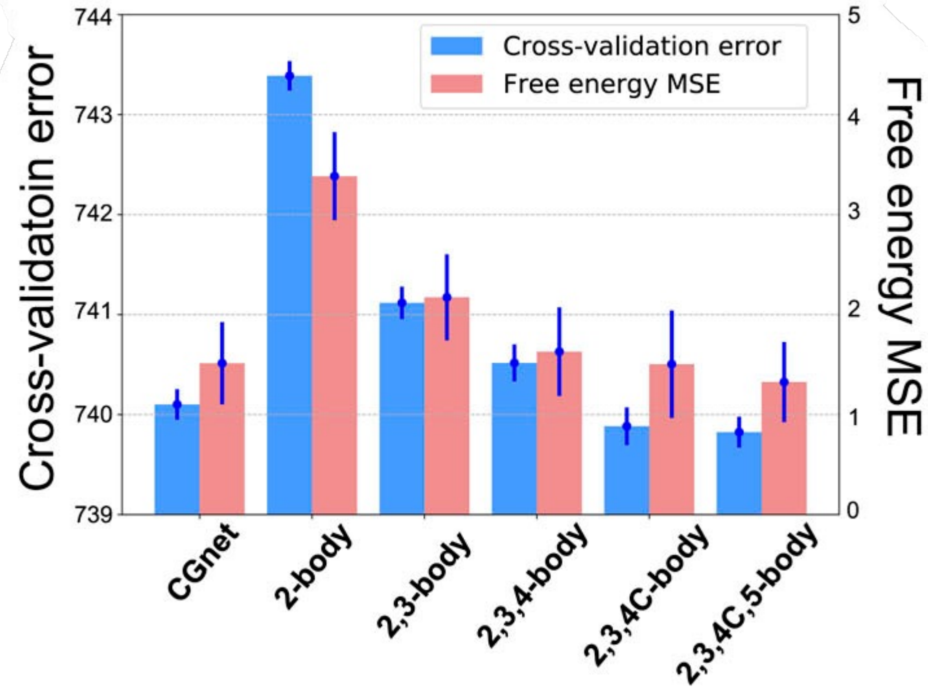
Multibody potentials: how many bodies?

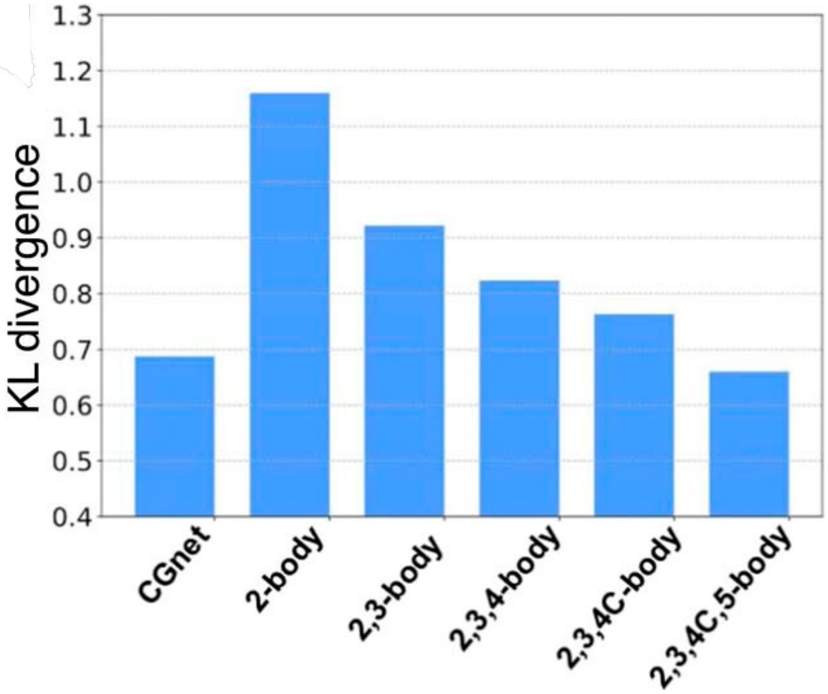




J. Wang, N.Charron, B. Husic, S. Olsson, F. Noé, C. Clementi, J. Chem. Phys., 154 (16), 164113 (2021)

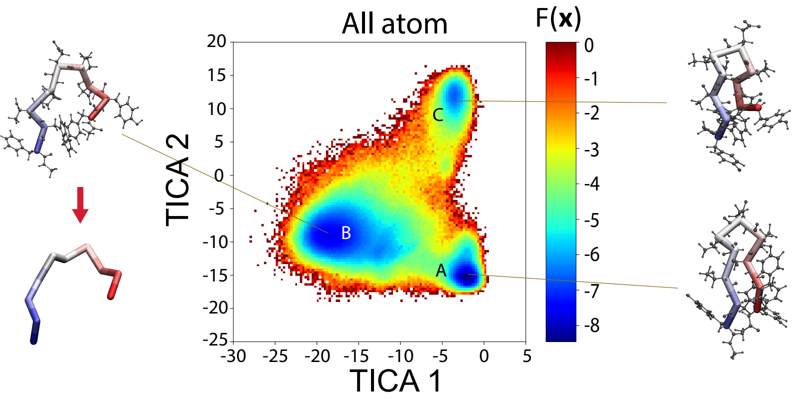
Multibody potentials: how many bodies?

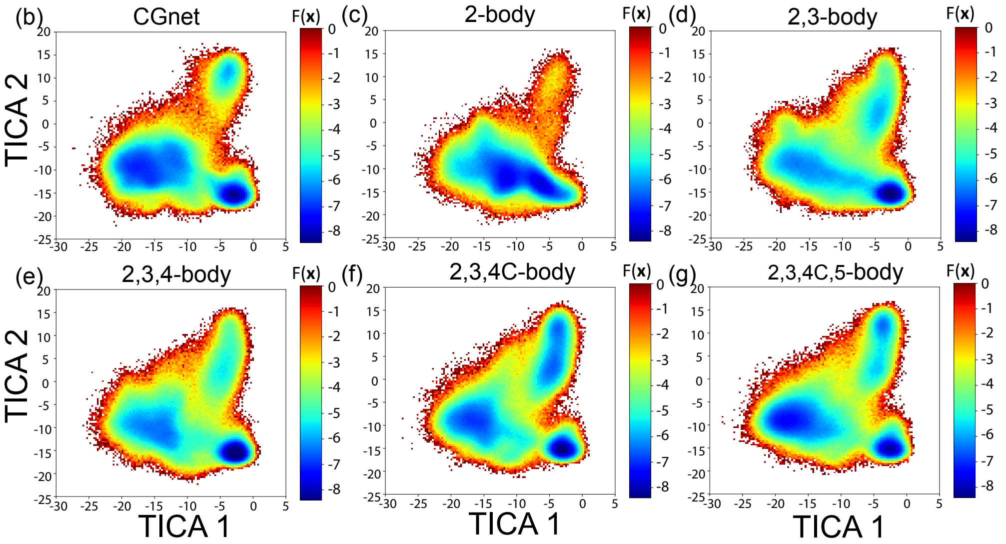




J. Wang, N.Charron, B. Husic, S. Olsson, F. Noé, C. Clementi, J. Chem. Phys., 154 (16), 164113 (2021)

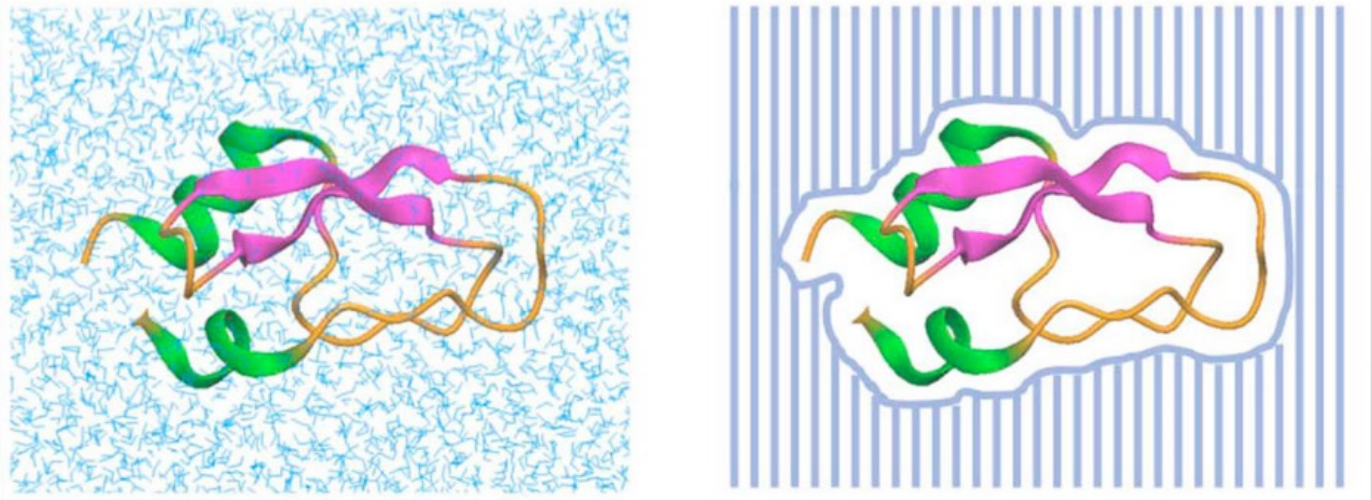
Multibody potentials: how many bodies?





J. Wang, N.Charron, B. Husic, S. Olsson, F. Noé, C. Clementi, J. Chem. Phys., 154 (16), 164113 (2021)

ISSchNet: Implicit Solvent CG Models

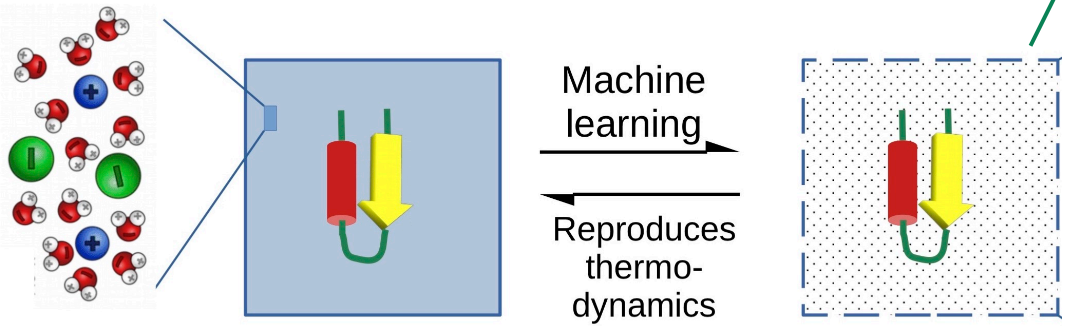


$$H = H_{\text{prot}} + H_{\text{prot-wat}} + H_{\text{wat}}$$

$$H = H_{\text{prot}} + \Delta G_{\text{solv}}$$

Chen, Y., Krämer, A., Charron, N. E., Husic, B. E., Clementi, C., & Noé, F.
Machine learning implicit solvation for molecular dynamics. *J. Chem. Phys*, *155*(8), 084101 (2021).

ISSchNet: Implicit Solvent CG Models



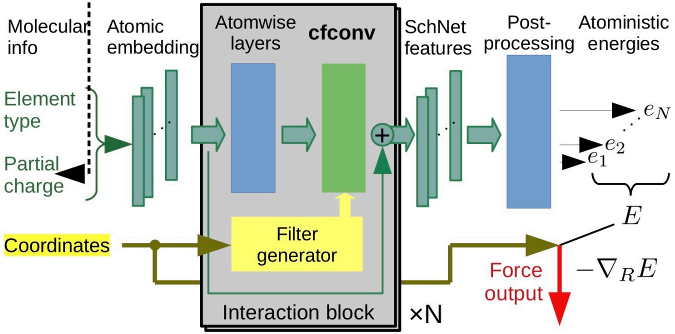
Machine learning

Reproduces thermo-dynamics

Methods

Neural network force field

- ISSNet = an extension of SchNet,[3] (graph neural network)
- Including electrostatic information by embedding atomic partial charges
- Implicit solvent MD with ISSNet



3. Schütt, et al., *J. Chem. Phys.* 148, 241722 (2018).

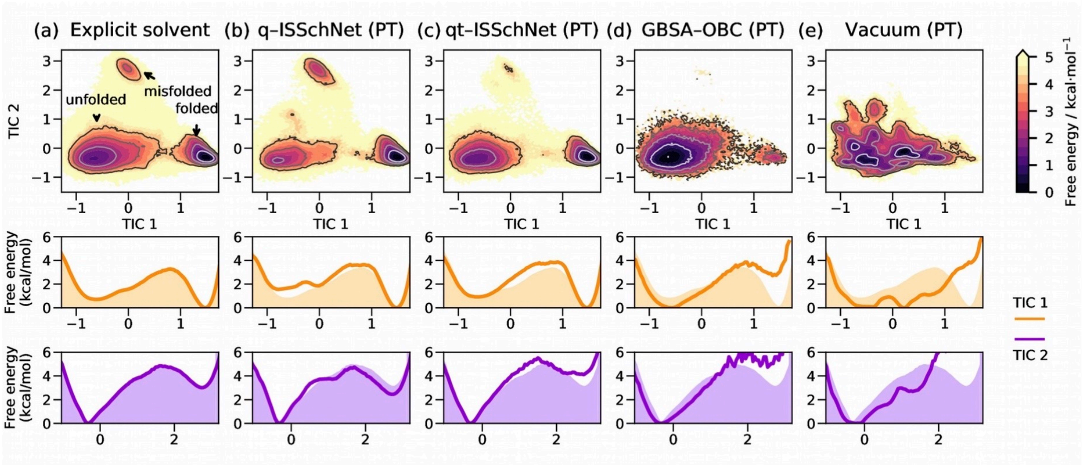
Chen, Y., Krämer, A., Charron, N. E., Husic, B. E., Clementi, C., & Noé, F.
Machine learning implicit solvation for molecular dynamics. *J. Chem. Phys*, *155*(8), 084101 (2021).

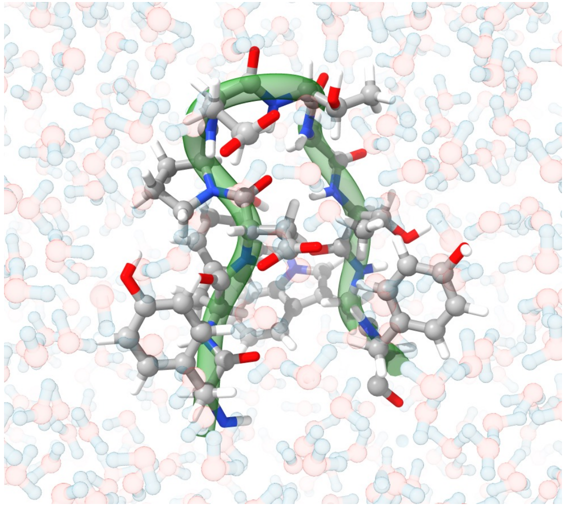
ISSchNet: Implicit Solvent CG Models

ISSNet model can correctly reproduce the folding temperature of miniprotein **chignolin**

- CLN025 variant: 175 solute atoms
- Comparing to GBSA model, trained ISSNet model
  - more accurately reproduces thermodyanmics (shown below), and
  - gives more reasonable folding temperature →

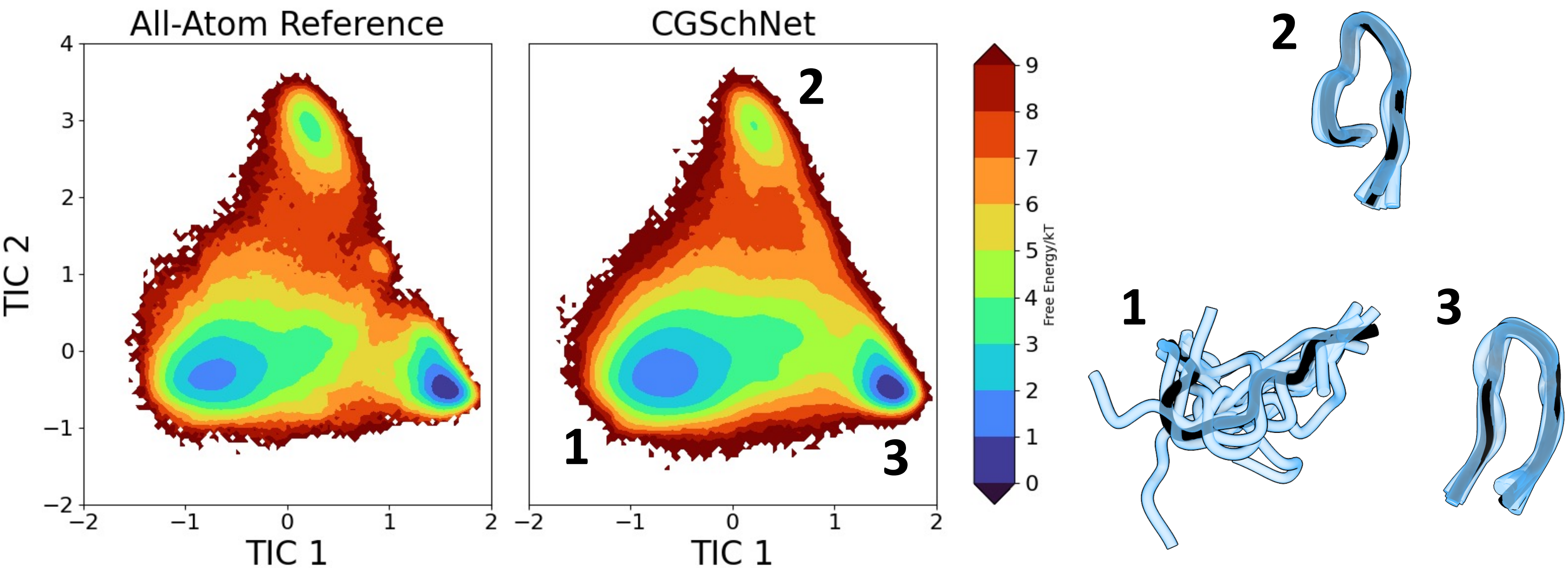
Solvation model	T_m / K
q-ISSNet	**~368.3**
GBSA–OBC2 (comparable with Anandakrishnan et al., Biophys. J., 2015)	~266.5
(Ref) Explicit solvent (Lindorff-Larsen et al., *Science*, 2011)	361.0–393.0
(Ref) Experimental (S. Honda et al., *J. Am. Chem. Soc.*, 2008)	~343





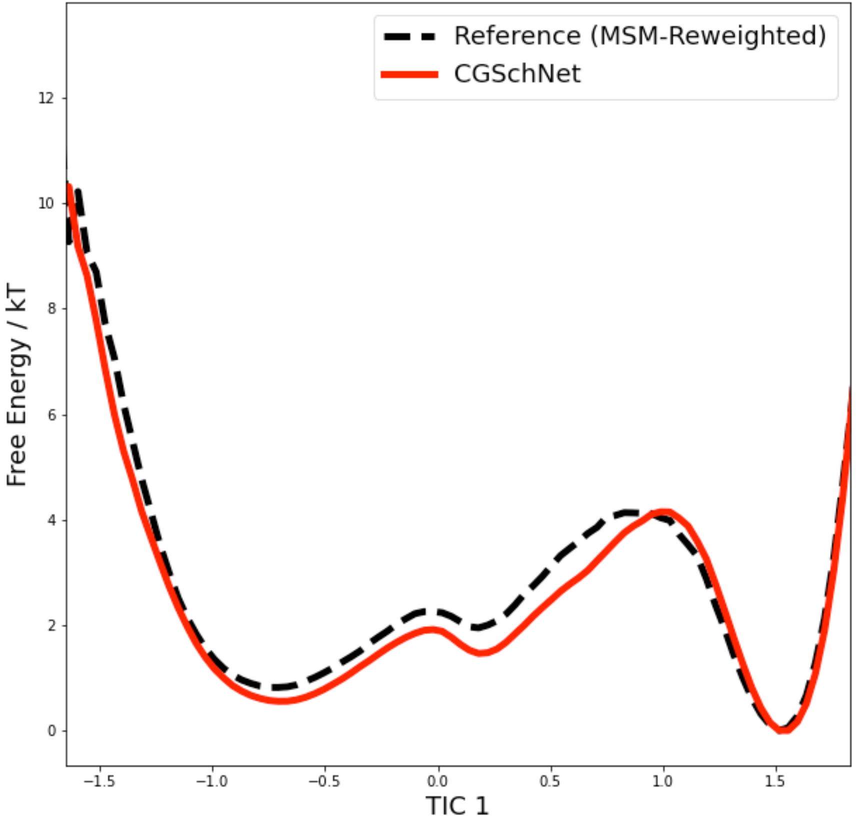
Chen, Y., Krämer, A., Charron, N. E., Husic, B. E., Clementi, C., & Noé, F.
Machine learning implicit solvation for molecular dynamics. *J. Chem. Phys*, *155*(8), 084101 (2021).

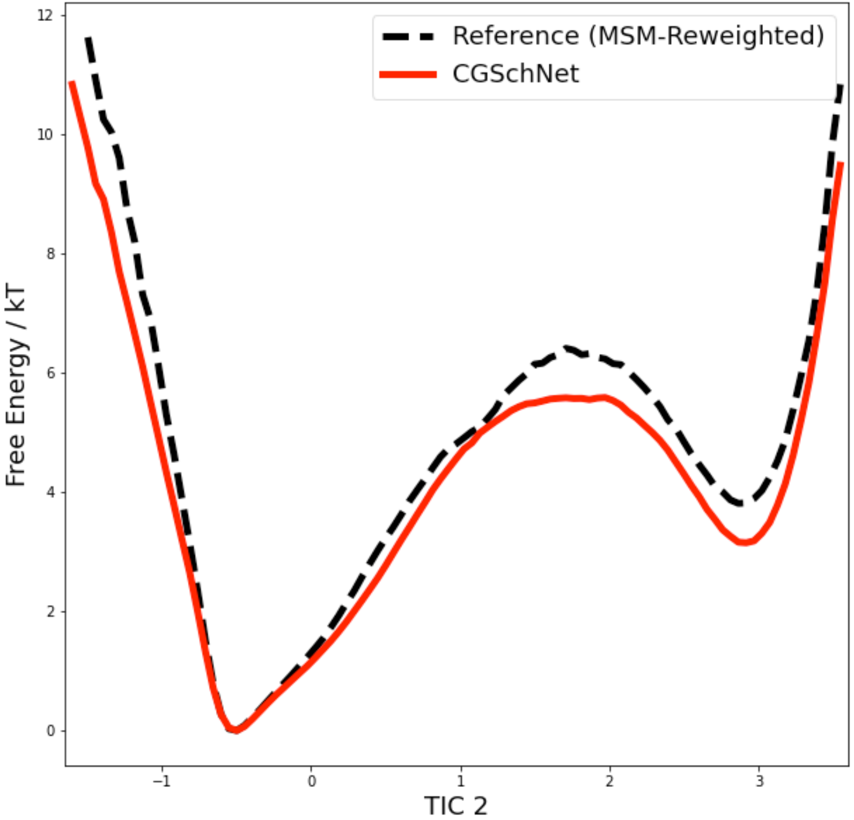
State of The Art MLCG Model



J. Wang, S. Olsson, C. Wehmeyer, A. Perez, N.E. Charron, G. de Fabritiis, F. Noé, C. Clementi, *ACS Cent. Sci*, 5, 755-767 (2019)
B. E. Husic, N. E. Charron, …, G. de Fabritiis, F. Noé, C. Clementi *J. Chem. Phys.* 153 (19), 194101 (2020)

State of The Art MLCG Model



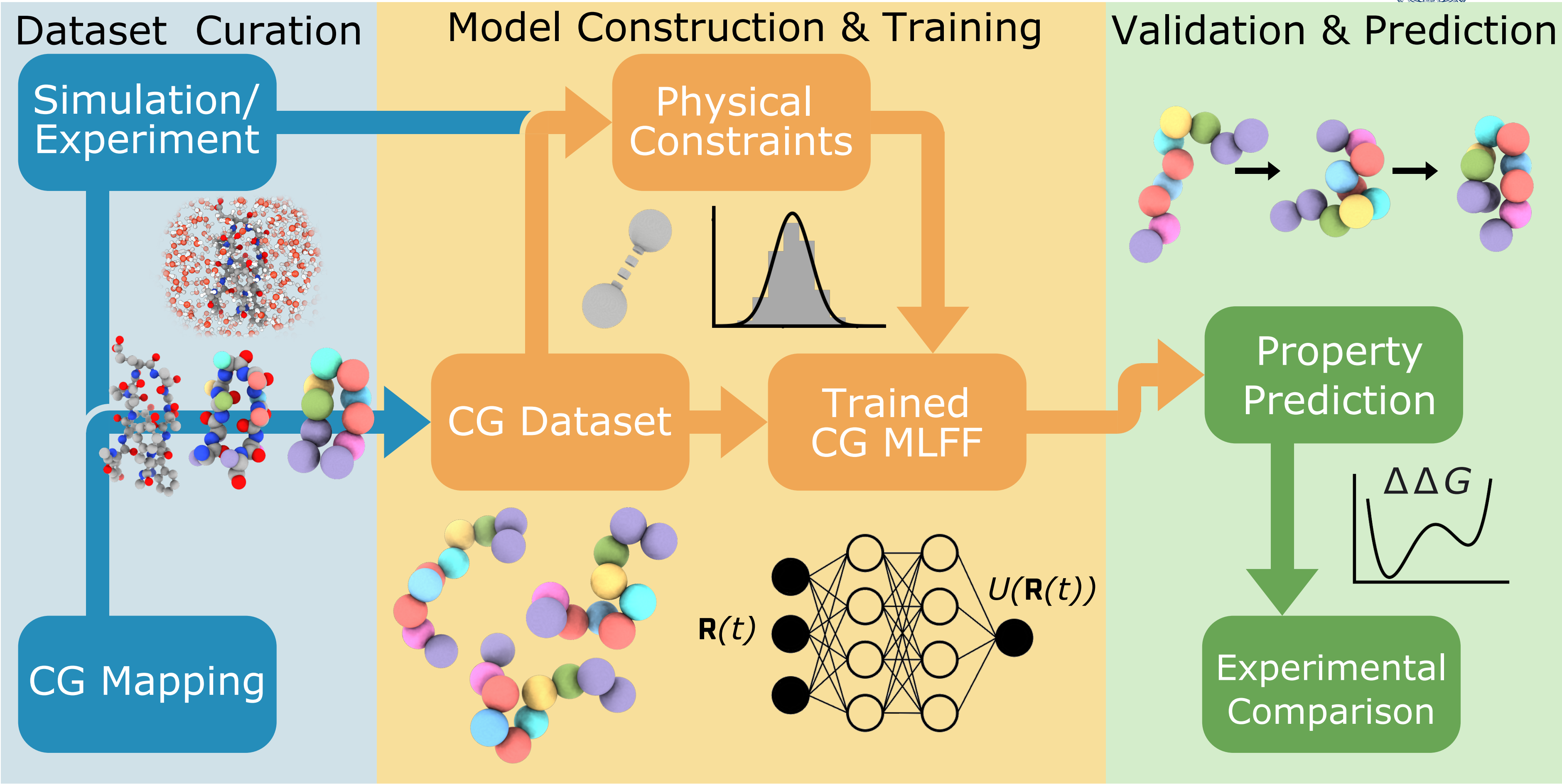


J. Wang, S. Olsson, C. Wehmeyer, A. Perez, N.E. Charron, G. de Fabritiis, F. Noé, C. Clementi, *ACS Cent. Sci*, 5, 755-767 (2019)

B. E. Husic, N. E. Charron, …, G. de Fabritiis, F. Noé, C. Clementi *J. Chem. Phys.* 153 (19), 194101 (2020)

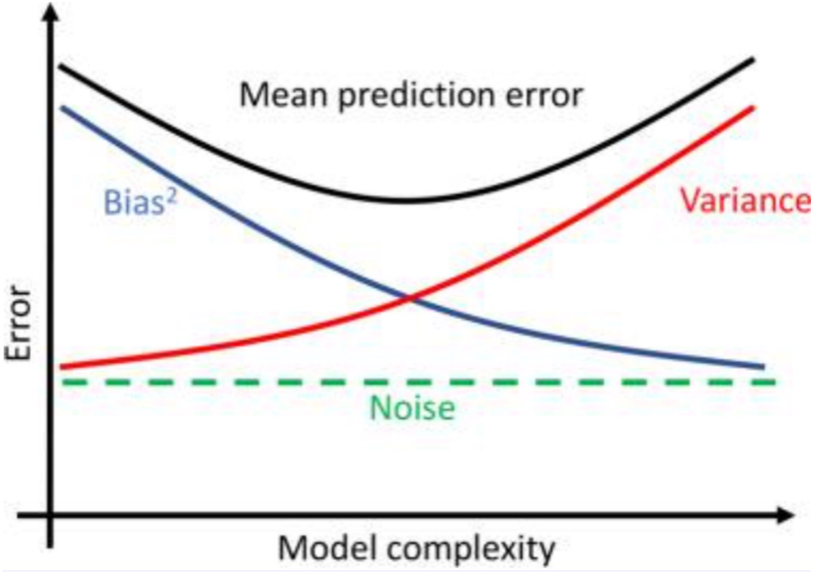
How do we achieve state of the art?

Freie Universität Berlin



Optimization: CG Force Noise

$$\chi^2(\theta) = \left\langle \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) + \nabla U(\xi_r(\boldsymbol{r}); \theta) \right\|^2 \right\rangle_{\boldsymbol{r}}$$



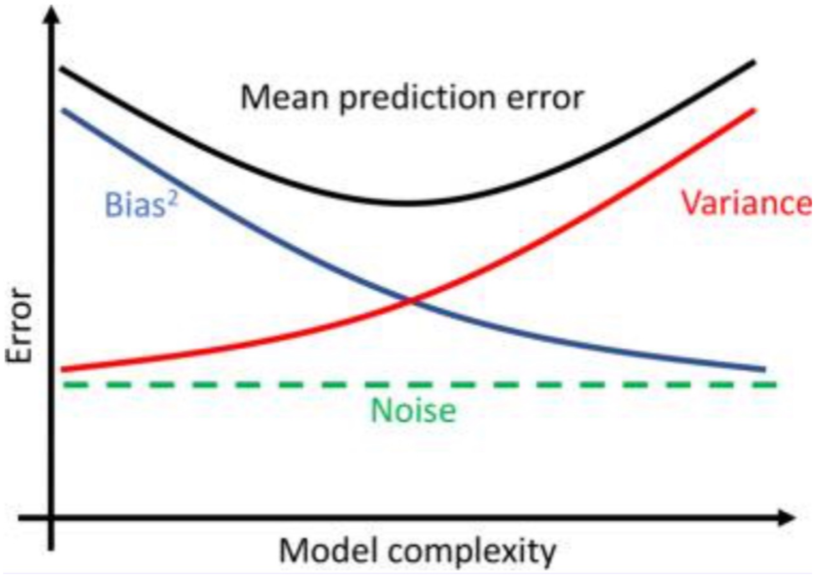
Wang, J., Olsson, S., Wehmeyer, C., Pérez, A., Charron, N. E., de Fabritiis, G., Noé, F., & Clementi, C. (2019). Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Central Science*.

Optimization: CG Force Noise

$$\chi^2(\theta) = \left\langle \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) + \nabla U(\xi_r(\boldsymbol{r}); \theta) \right\|^2 \right\rangle_{\boldsymbol{r}}$$

$$\downarrow$$

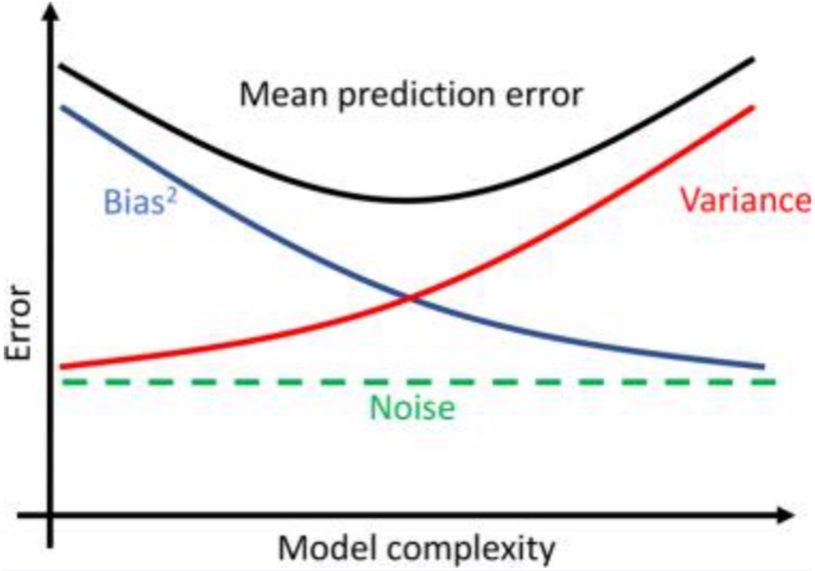
$$L(\theta) = \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) + \nabla U(\xi_r(\boldsymbol{r}); \theta) \right\|^2$$



Wang, J., Olsson, S., Wehmeyer, C., Pérez, A., Charron, N. E., de Fabritiis, G., Noé, F., & Clementi, C. (2019). Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Central Science*.

Optimization: CG Force Noise

Freie Universität Berlin



$$\chi^2(\theta) = \left\langle \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) + \nabla U(\xi_r(\boldsymbol{r}); \theta) \right\|^2 \right\rangle_{\boldsymbol{r}}$$

$$\downarrow$$

$$L(\theta) = \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) + \nabla U(\xi_r(\boldsymbol{r}); \theta) \right\|^2$$

$$\downarrow$$

$$\mathbb{E}[L(\theta)] = \text{Parameter Bias}^2 + \text{Parameter variance} + \text{CG Bias}$$

$$\boldsymbol{f} = \langle \xi_f \boldsymbol{F}(\boldsymbol{r}) \rangle_{r|R}$$

$$\text{Parameter Bias}^2 = \| \boldsymbol{f} - \mathbb{E}[-\nabla U(\xi_r \boldsymbol{r}; \theta)] \|^2$$

$$\text{Parameter Var} = \mathbb{E}\left[\| \mathbb{E}[-\nabla U(\xi_r \boldsymbol{r})] + \nabla U(\xi_r \boldsymbol{r}; \theta) \|^2 \right]$$

$$\text{CG Force Noise} = \left\langle \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) - \boldsymbol{f} \right\|^2 \right\rangle_{\boldsymbol{r}}$$

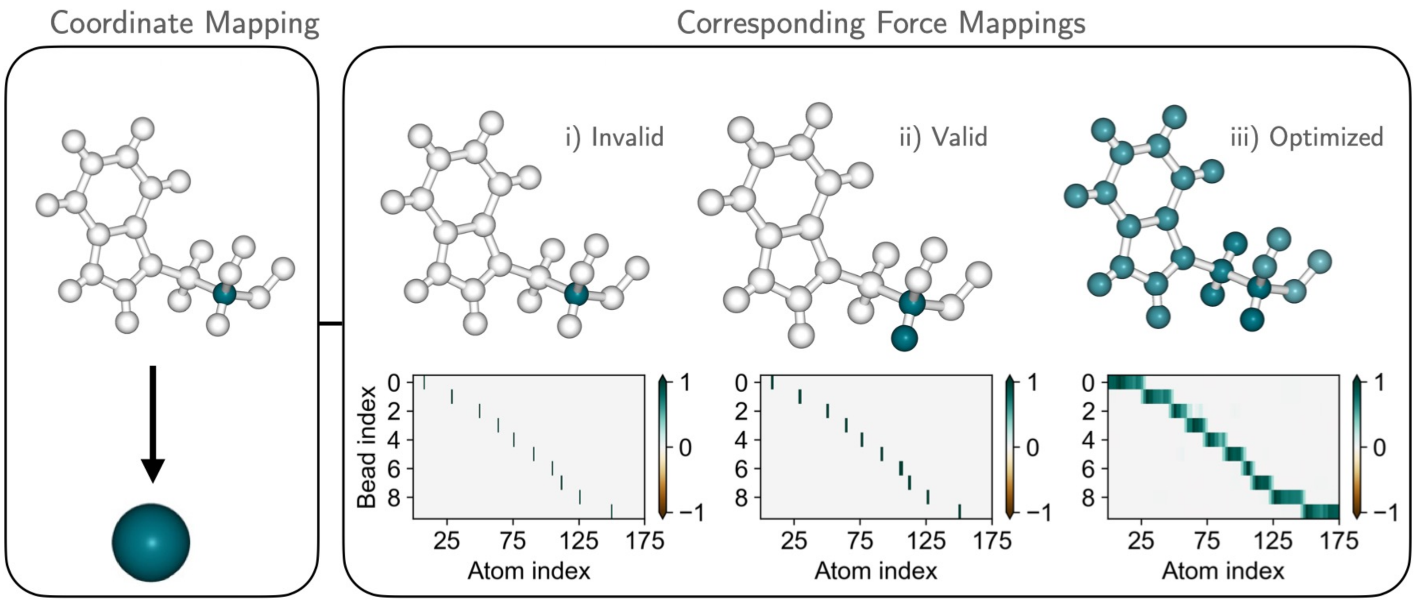
Wang, J., Olsson, S., Wehmeyer, C., Pérez, A., Charron, N. E., de Fabritiis, G., Noé, F., & Clementi, C. (2019). Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Central Science*.

Force Aggregation Strategies

Freie Universität Berlin

$$\xi_r \neq \xi_f$$

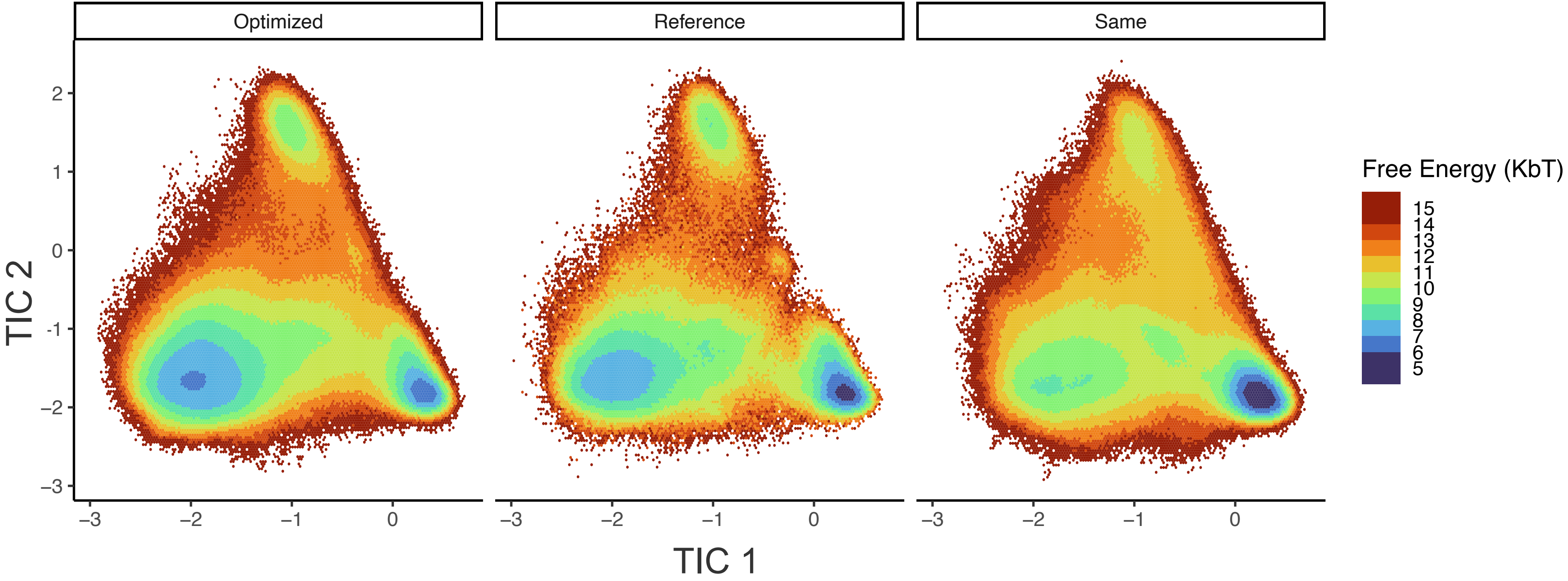
$$\text{CG Force Noise} = \left\langle \left\| \xi_f \boldsymbol{F}(\boldsymbol{r}) - \boldsymbol{f} \right\|^2 \right\rangle_{\boldsymbol{r}}$$



Krämer, A.,Durumeric, A.P., Charron, N.E., Chen, Y., Clementi, C. & Noé, F., (2023) *in preparation*
Statistically optimal force aggregation for coarse-grained molecular dynamics

Force Aggregation Strategies

$$\xi_r \neq \xi_f \qquad\qquad \xi_r = \xi_f$$

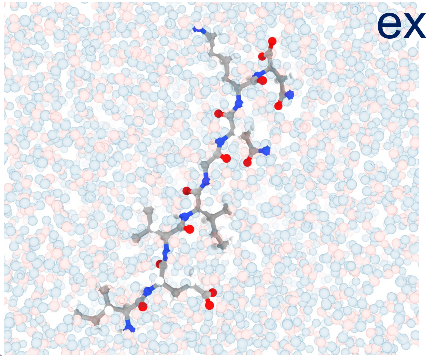


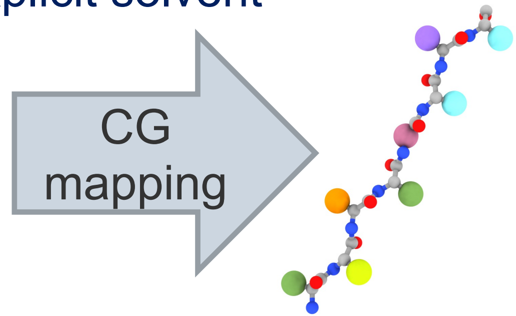
Krämer, A.,Durumeric, A.P., Charron, N.E., Chen, Y., Clementi, C. & Noé, F., (2023) *in preparation*
Statistically optimal force aggregation for coarse-grained molecular dynamics

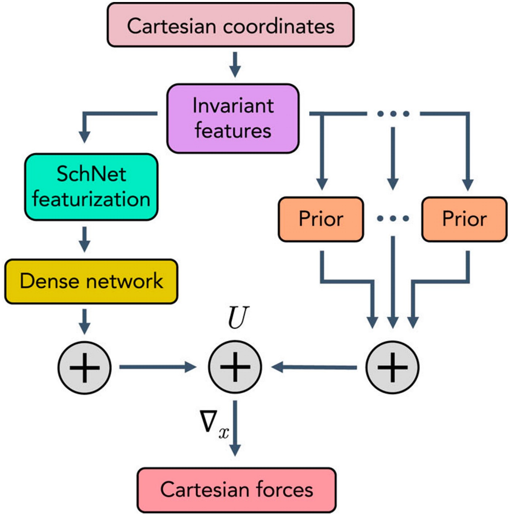
Developing transferable CG force fields

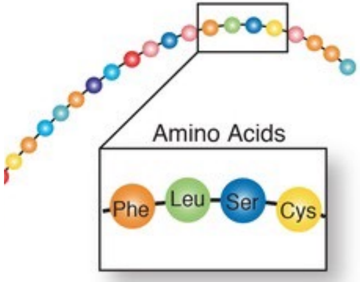
INPUT

1100 different octapeptide sequences in explicit solvent





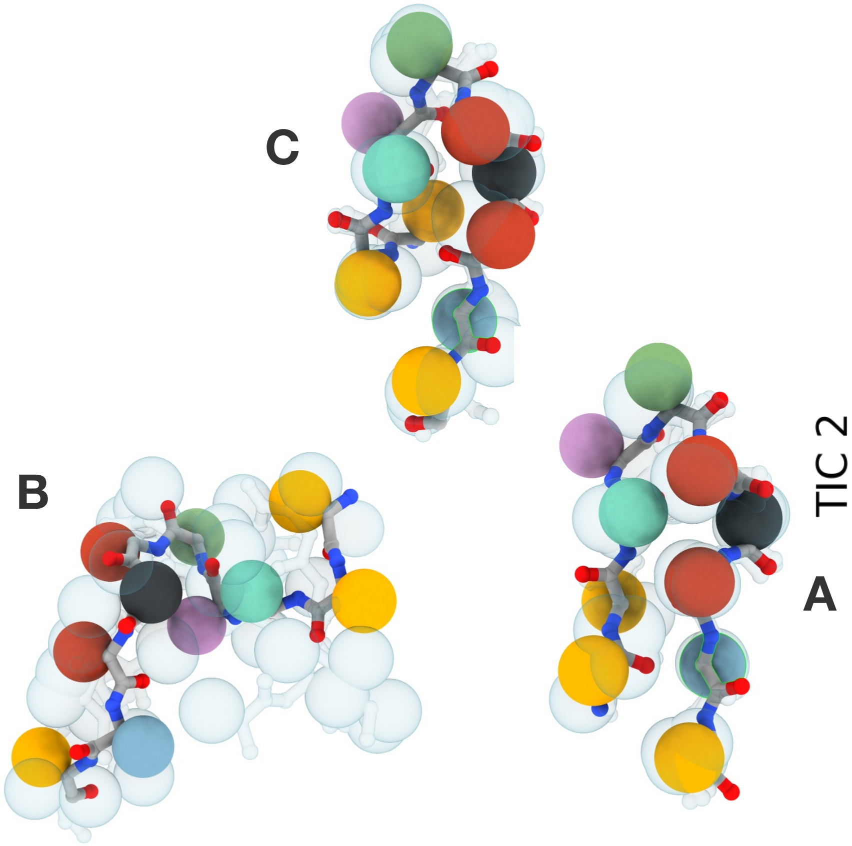


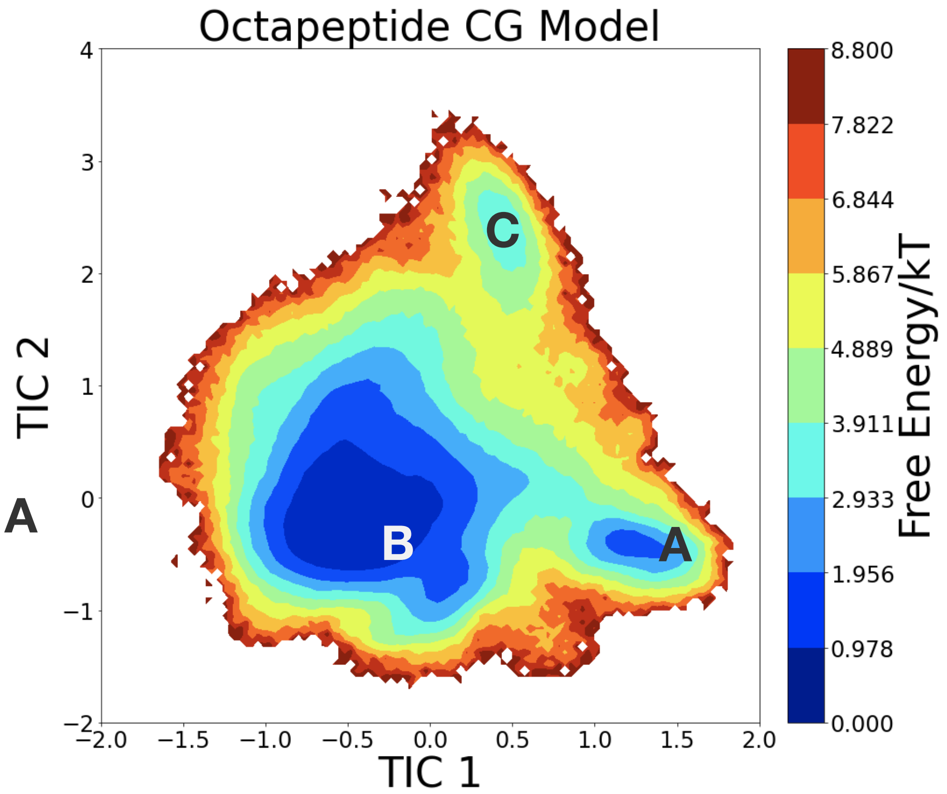


OUTPUT

Learn a "universal" CG force-field

Developing transferable CG force fields





Force-matching without forces

force matching

$$\text{MD} \xrightarrow{\text{simulate}} r, f \xrightarrow[\text{matching}]{\text{force}} V(r; \theta_{\text{pot}})$$

relative entropy method

density estimation

optimize

$$\text{MD} \xrightarrow{\text{simulate}} r \quad V(r; \theta_{\text{pot}}) \quad V(r; \theta_{\text{pot}})$$

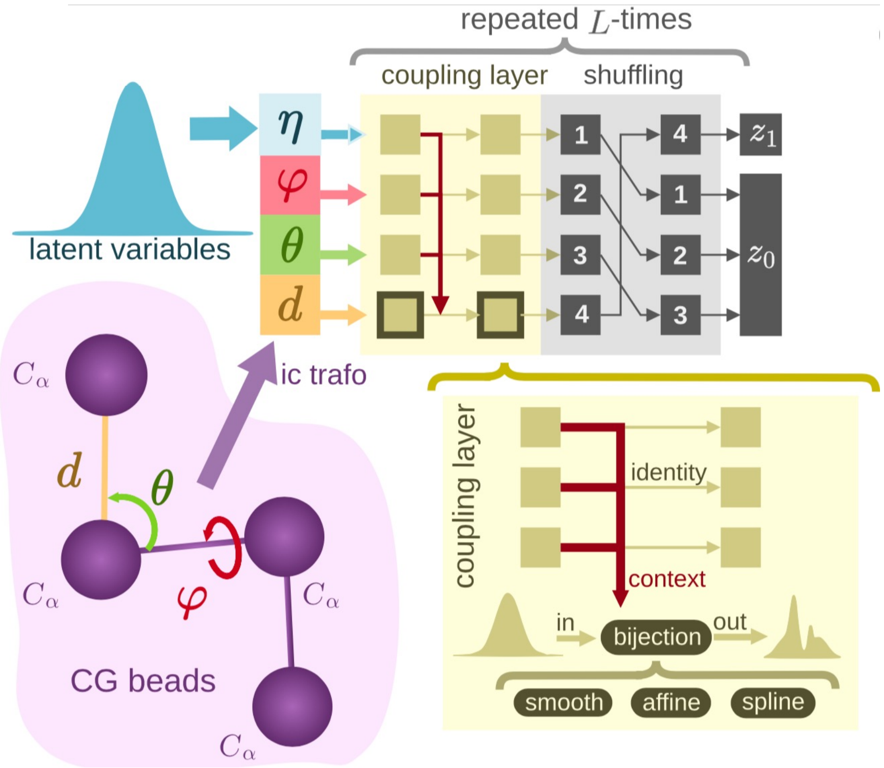
simulate

force matching without forces

$$\text{MD} \xrightarrow{\text{simulate}} \underbrace{r \xrightarrow[\text{estimation}]{\text{density}} \mathcal{V}(r; \theta_{\text{flow}})}_{\text{step 1}} \xrightarrow{\text{sample}} \underbrace{r, f \xrightarrow[\text{matching}]{\text{force}} V(r; \theta_{\text{pot}})}_{\text{step 2}}$$

J. Köhler, Y. Chen, A. Krämer, C. Clementi, F. Noé, arXiv:2203.11167 (2022)

Force-matching without forces



Flow density p corresponds to a (noisy) coarse-grained potential

$$\mathcal{V}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}) = -\log p(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}),$$

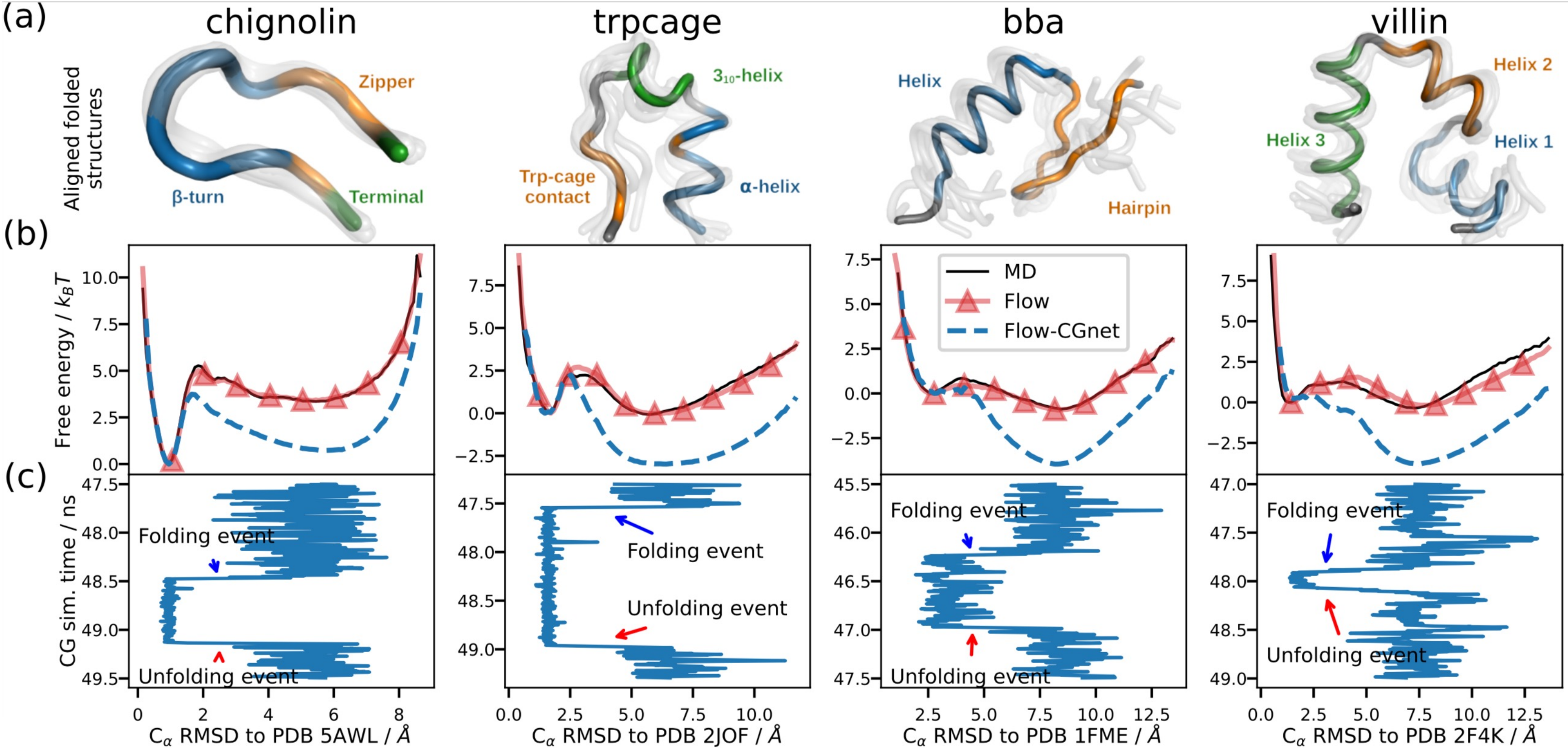
Noisy forces

$$\boldsymbol{f}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}) = \nabla_{\boldsymbol{r}} \mathcal{V}(\boldsymbol{r}, \boldsymbol{\eta}; \boldsymbol{\theta}),$$

—> Match these forces on average

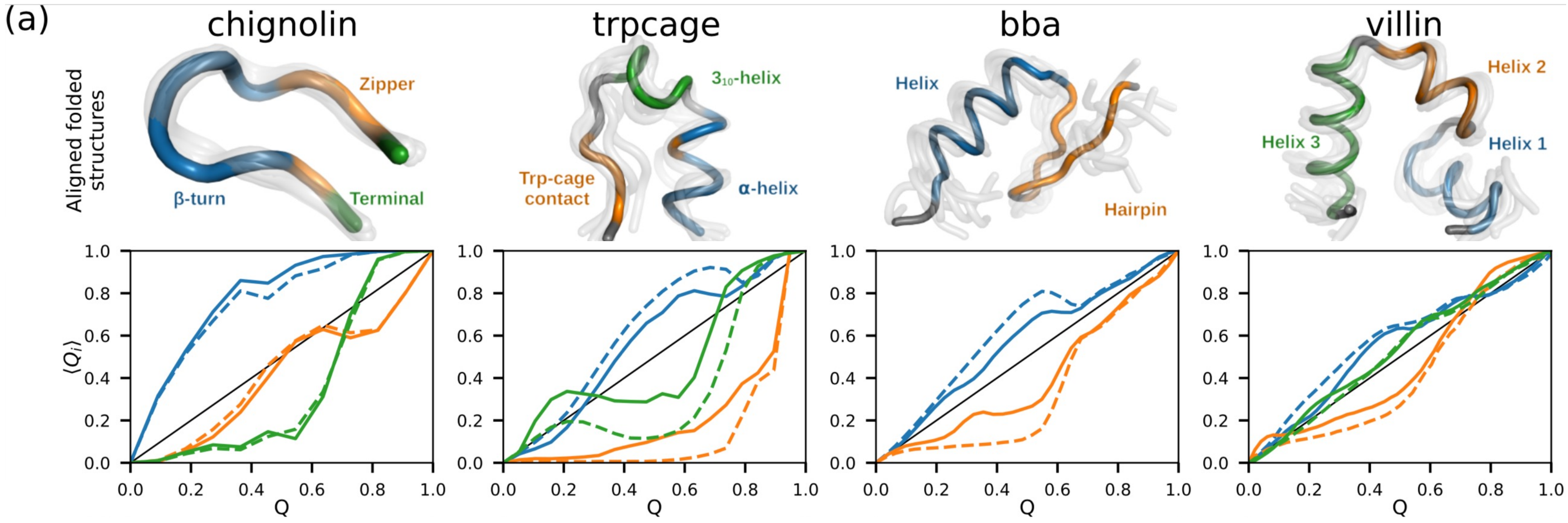
J. Köhler, Y. Chen, A. Krämer, C. Clementi, F. Noé, arXiv:2203.11167 (2022)

Force-matching without forces



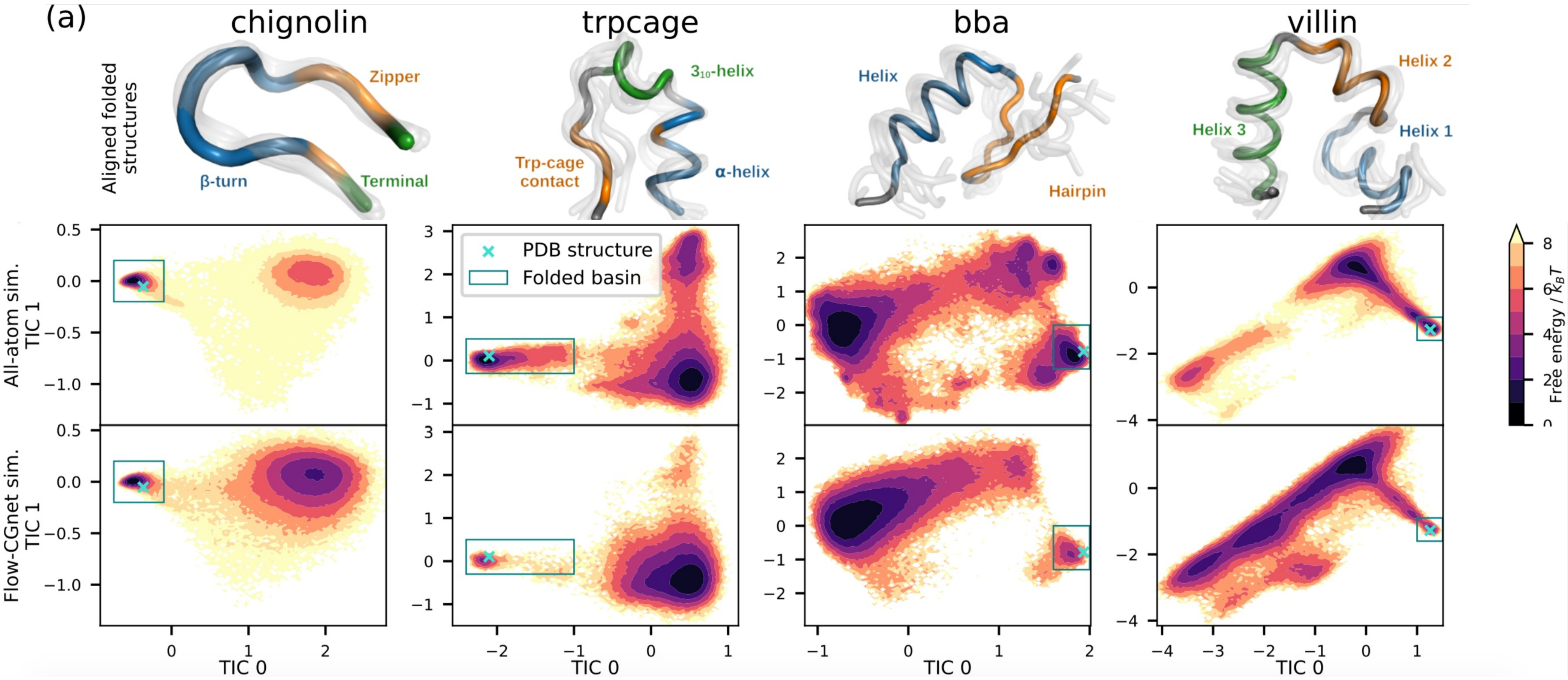
J. Köhler, Y. Chen, A. Krämer, C. Clementi, F. Noé, arXiv:2203.11167 (2022)

Force-matching without forces



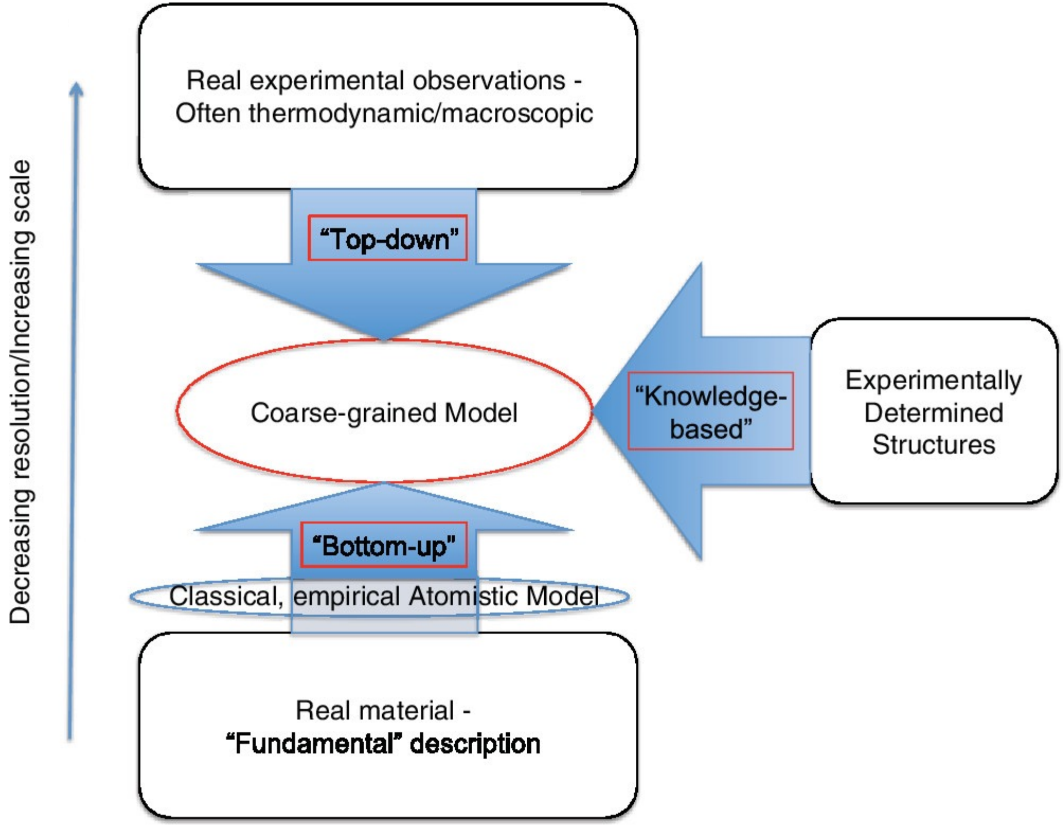
J. Köhler, Y. Chen, A. Krämer, C. Clementi, F. Noé, arXiv:2203.11167 (2022)

Force-matching without forces



J. Köhler, Y. Chen, A. Krämer, C. Clementi, F. Noé, arXiv:2203.11167 (2022)

CG models: which properties to preserve?



1. Learning CG models from all-atom simulations
2. **Incorporating nuclear quantum effects into all-atom simulation**
3. Learning CG models from experimental data

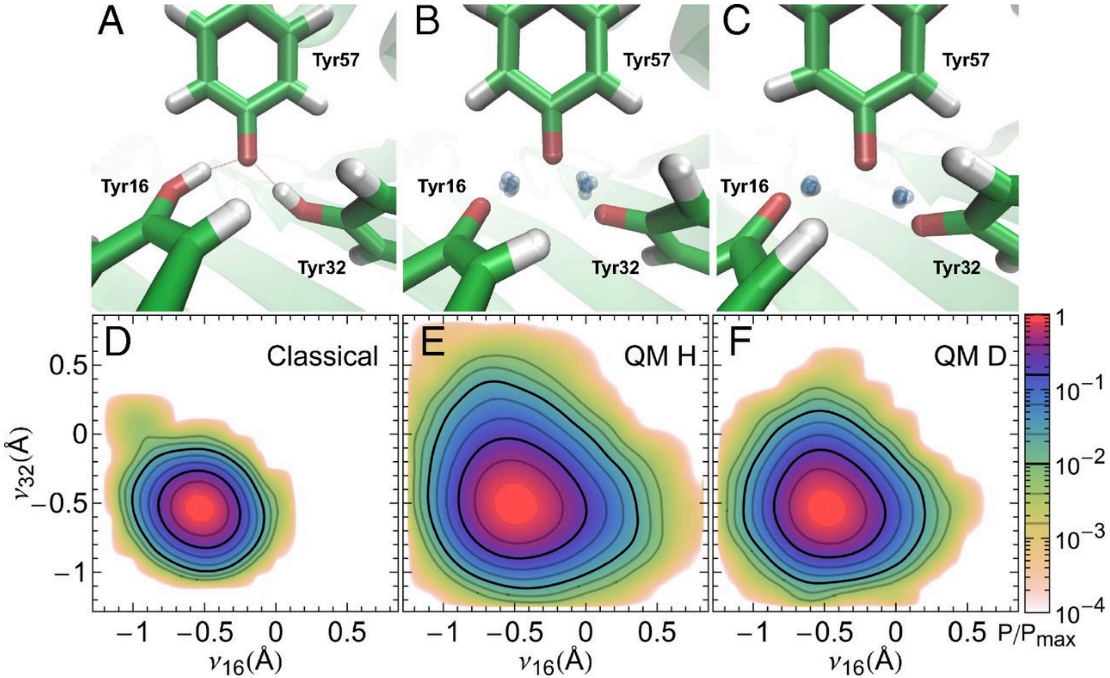
W.G. Noid, *J. Chem. Phys.* 139:9, 090901 (2013)

Coarse-graining nuclear quantum effects

Classical picture: nuclei move on a potential energy surface according to the laws of classical physics

Quantum picture: nuclei obey Schrödinger equation

Effects present only in quantum picture: tunnelling, isotope effects ...

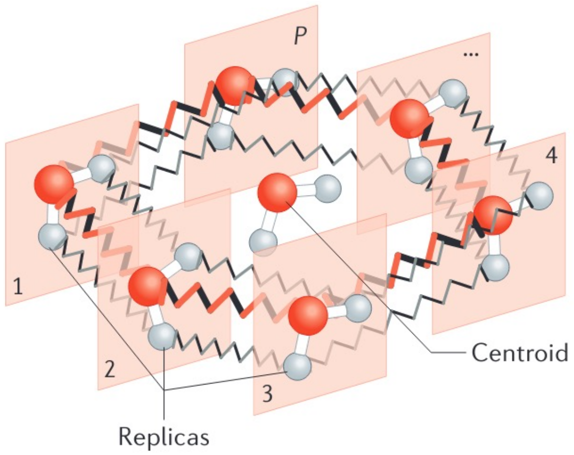


Delocalized proton in the active site of ketosteroid isomerase D40N

Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. Quantum Delocalization of Protons in the Hydrogen-Bond Network of an Enzyme Active Site. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (52), 18454–18459.

Quantum-classical isomorphism and path integral molecular dynamics

Freie Universität Berlin



Sample the quantum canonical distribution by moving to an extended phase space:

$$H(\mathbf{q},\mathbf{p}) = \sum_{k=1}^{P}\left[\frac{\mathbf{p_k}^2}{2m'} + \frac{1}{2}m\omega_P^2(\mathbf{q_{k+1}} - \mathbf{q_k})^2 + \frac{1}{P}U(\mathbf{q_k})\right]_{\mathbf{q_{P+1}}=\mathbf{q_1}}$$

$$\omega_P = \frac{\sqrt{P}}{\beta\hbar} \qquad P \to \infty$$

Can we coarse-grain the ring polymer back to a centroid?

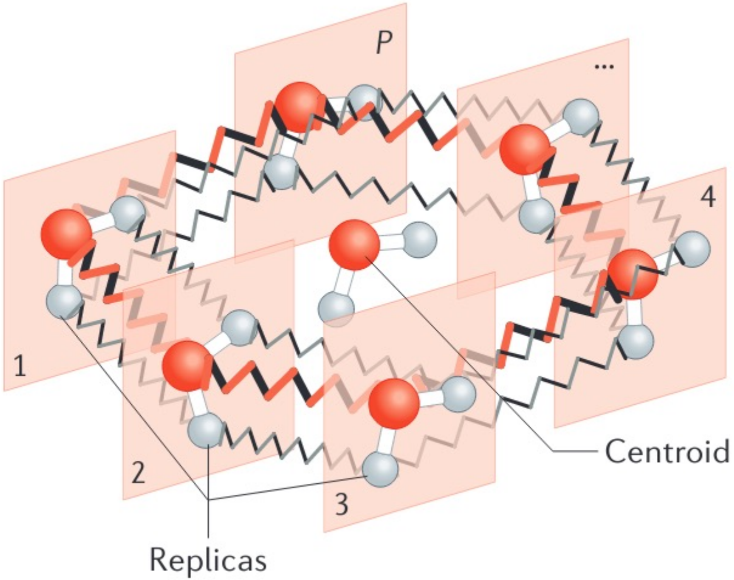
Markland, T. E.; Ceriotti, M. Nuclear Quantum Effects Enter the Mainstream. *Nat. Rev. Chem.* **2018**, *2* (3).

Quantum-classical isomorphism and path integral molecular dynamics

Freie Universität Berlin

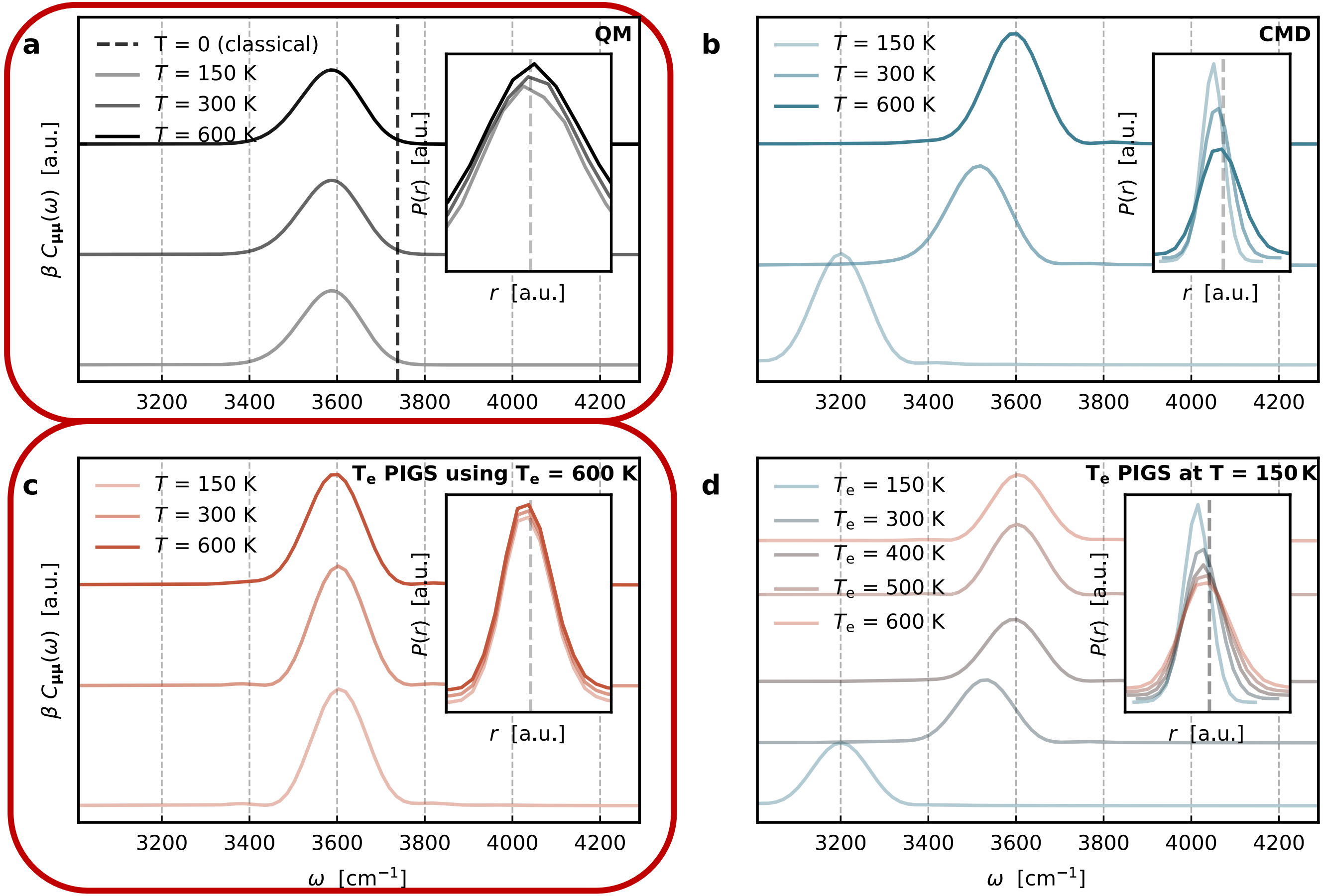
Can we coarse-grain the ring polymer back to a centroid?

centroid **P**ath **I**ntegral coarse-**G**rained **s**imulation
PIGs



F. Musil, I. Zaporozhets, F. Noe, C. Clementi, V. Kapil, *J. Chem. Phys.* 157, 181102 (2022)

Benchmark system: OH bond

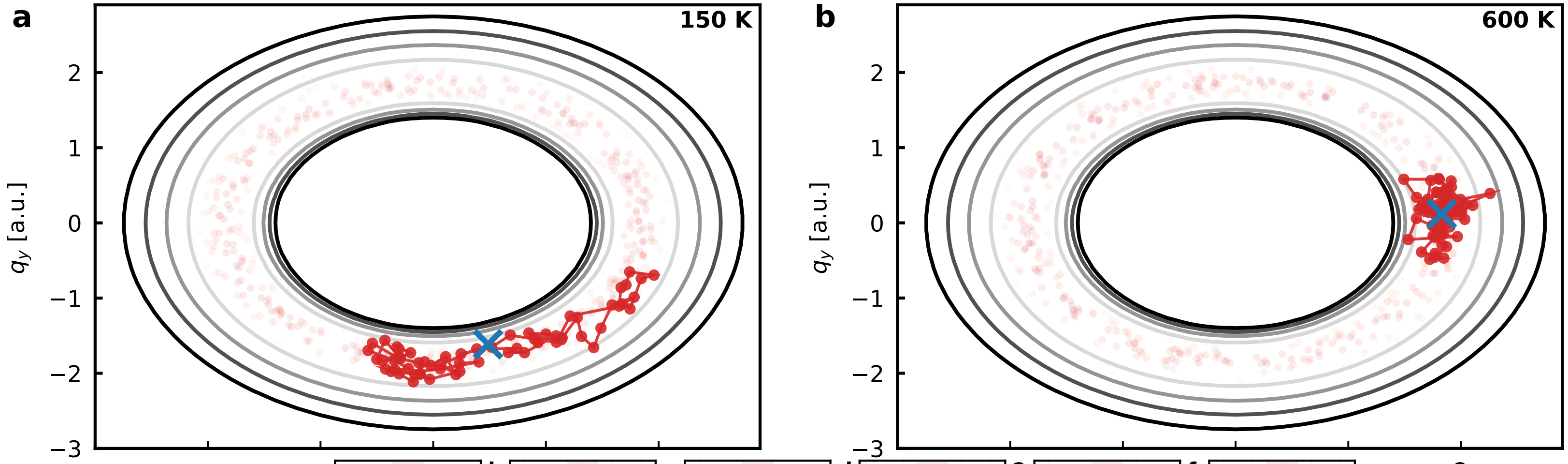


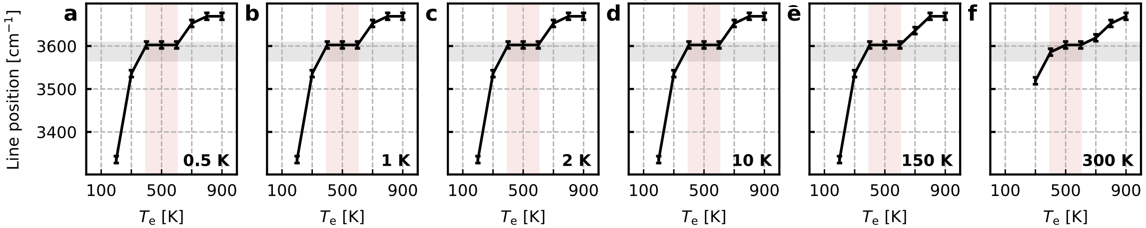
F. Musil, I. Zaporozhets, F. Noe, C. Clementi, V. Kapil, *J. Chem. Phys.* 157, 181102 (2022)

Quantum-classical isomorphism and path integral molecular dynamics

Freie Universität Berlin

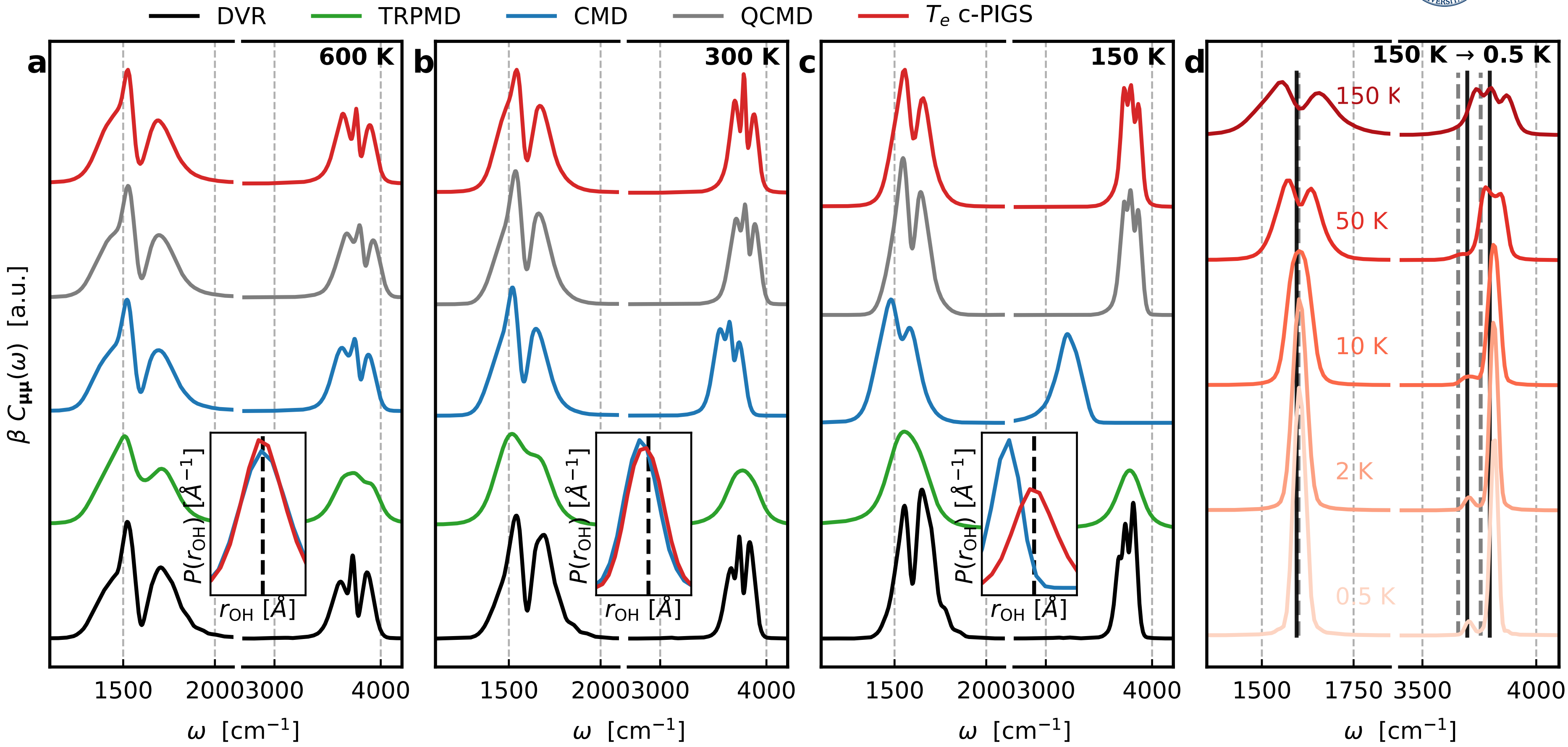
Temperature elevation approximation
T_e





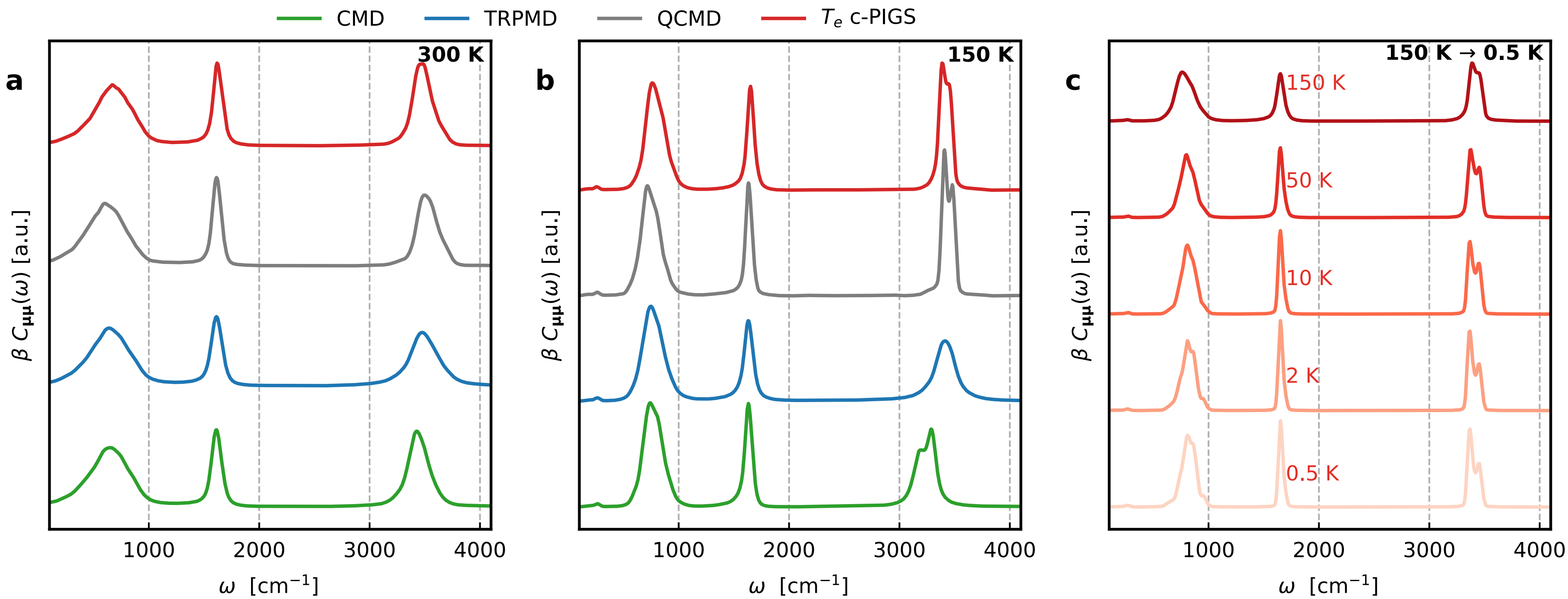
F. Musil, I. Zaporozhets, F. Noe, C. Clementi, V. Kapil, *J. Chem. Phys.* 157, 181102 (2022)

Water molecule



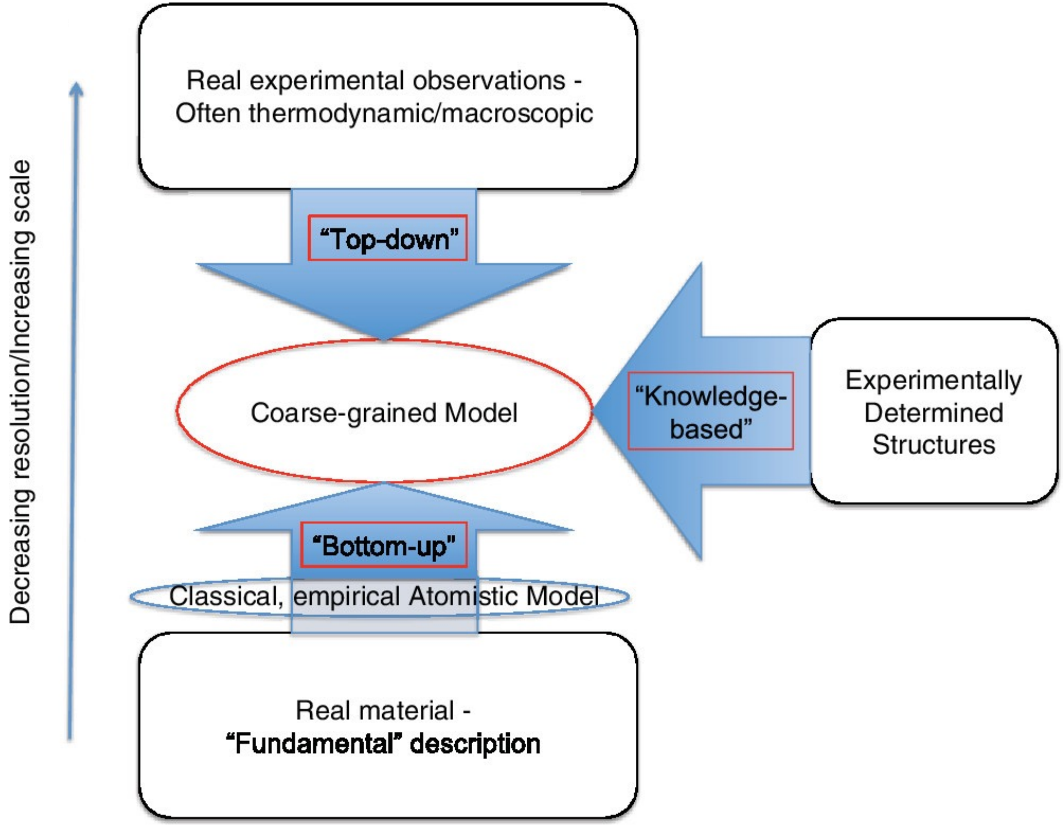
F. Musil, I. Zaporozhets, F. Noe, C. Clementi, V. Kapil, *J. Chem. Phys.* 157, 181102 (2022)

Bulk water



F. Musil, I. Zaporozhets, F. Noe, C. Clementi, V. Kapil, *J. Chem. Phys.* 157, 181102 (2022)

CG models: which properties to preserve?



1. Learning CG models from all-atom simulations
2. Incorporating nuclear quantum effects into all-atom simulation
3. **Learning CG models from experimental data**

W.G. Noid, *J. Chem. Phys.* 139:9, 090901 (2013)

Observable-driven Design of Effective Molecular Models (ODEM)

$$V = V_{bond}(r, k_b) + V_{angle}(r, k_a) + V_{dihedral}(r, k_d) + V_{nonbonded}(r, \epsilon)$$

Input:

Parameter guess $\{\epsilon_i^{(0)}\}$

Experimental observables $\{f_k\}$

Estimate:

Equilibrium distribution π

$$\pi^{(1)} = \pi^{(0)} \exp[-\beta \Delta V(r, \epsilon)]$$

Observable values

$$e_k = \sum_i \pi_i(\epsilon) e_{ki}$$

Maximize:

Model likelihood

$$Q = \prod_{k=1}^{N} exp\left[-\frac{(e_k(\epsilon) - f_k)^2}{2\Delta f_k^2}\right]$$

Output:

Set of optimal parameters $\{\epsilon_i^{(1)}\}$

J. Chen, J. Chen, G. Pinamonti, C. Clementi *J. Chem. Theory Comput.*, **2018**, *14* (7), pp 3849–3858

Freie Universität Berlin

Observable-driven Design of Effective Molecular Models (ODEM)

$$V = V_{bond}(r, k_b) + V_{angle}(r, k_a) + V_{dihedral}(r, k_d) + V_{nonbonded}(r, \epsilon)$$

Input:

Parameter guess $\{\epsilon_i^{(0)}\}$

Experimental observables $\{f_k\}$

→

Estimate:

Equilibrium distribution π

$$\pi^{(1)} = \pi^{(0)} \exp[-\beta \Delta V(r, \epsilon)]$$

Observable values

$$e_k = \sum_i \pi_i(\epsilon) e_{ki}$$

→

Maximize:

Model likelihood

$$Q = \prod_{k=1}^{N} exp\left[-\frac{(e_k(\epsilon) - f_k)^2}{2\Delta f_k^2}\right]$$

→

Output:

Set of optimal parameters $\{\epsilon_i^{(1)}\}$

J. Chen, J. Chen, G. Pinamonti, C. Clementi *J. Chem. Theory Comput.*, **2018**, *14* (7), pp 3849–3858

Observable-driven Design of Effective Molecular Models (ODEM)

$$V = V_{bond}(r, k_b) + V_{angle}(r, k_a) + V_{dihedral}(r, k_d) + V_{nonbonded}(r, \epsilon)$$

Input:

Parameter guess $\{\epsilon_i^{(0)}\}$

Experimental observables $\{f_k\}$

Estimate:

Equilibrium distribution π

$$\pi^{(1)} = \pi^{(0)} \exp[-\beta \Delta V(r, \epsilon)]$$

Observable values

$$e_k = \sum_i \pi_i(\epsilon) e_{ki}$$

Maximize:

Model likelihood

$$Q = \prod_{k=1}^{N} exp\left[-\frac{(e_k(\epsilon) - f_k)^2}{2\Delta f_k^2}\right]$$

Output:

Set of optimal parameters $\{\epsilon_i^{(1)}\}$

J. Chen, J. Chen, G. Pinamonti, C. Clementi *J. Chem. Theory Comput.*, **2018**, *14* (7), pp 3849–3858

Observable-driven Design of Effective Molecular Models (ODEM)

$$V = V_{bond}(r, k_b) + V_{angle}(r, k_a) + V_{dihedral}(r, k_d) + V_{nonbonded}(r, \epsilon)$$

Input:

Parameter guess $\{\epsilon_i^{(0)}\}$

Experimental observables $\{f_k\}$

Estimate:

Equilibrium distribution π

$$\pi^{(1)} = \pi^{(0)} \exp[-\beta \Delta V(r, \epsilon)]$$

Observable values

$$e_k = \sum_i \pi_i(\epsilon) e_{ki}$$

Maximize:

Model likelihood

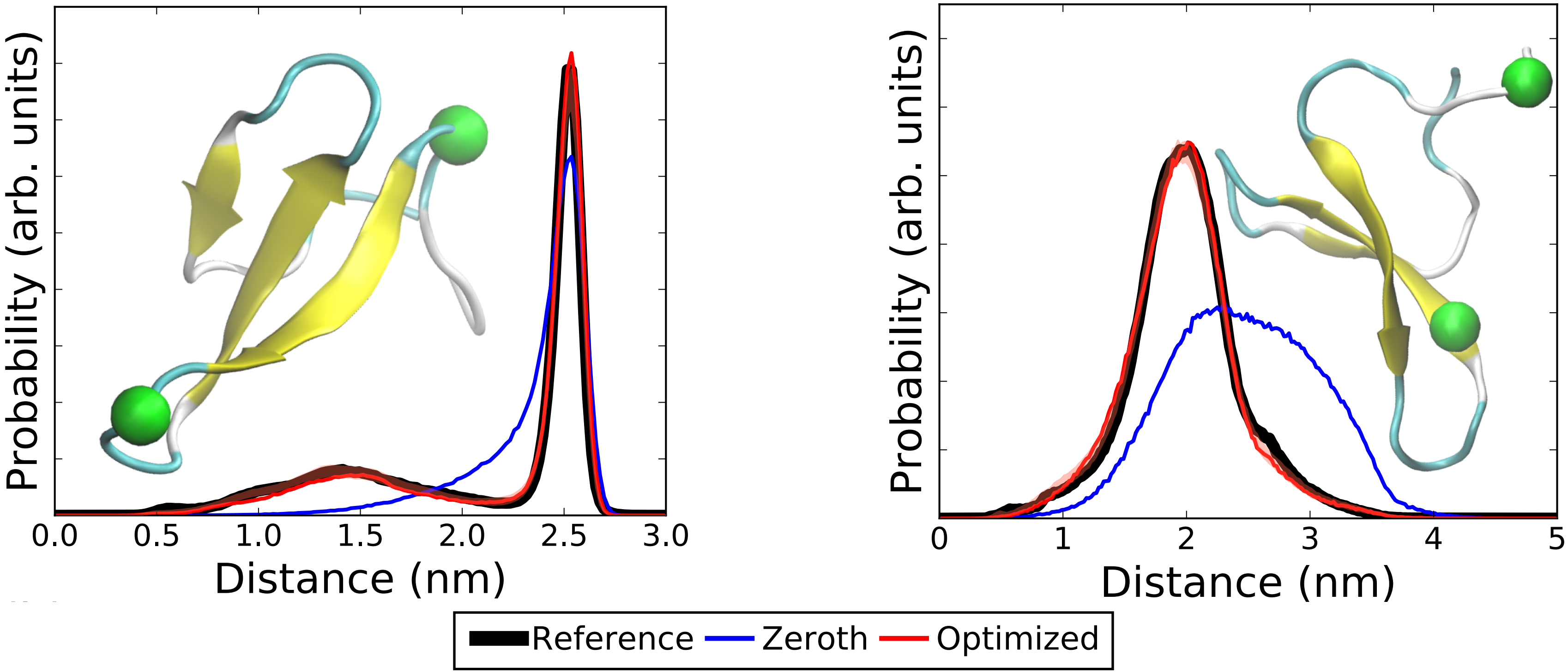
$$Q = \prod_{k=1}^{N} exp\left[-\frac{(e_k(\epsilon) - f_k)^2}{2\Delta f_k^2}\right]$$

Output:

Set of optimal parameters $\{\epsilon_i^{(1)}\}$

J. Chen, J. Chen, G. Pinamonti, C. Clementi *J. Chem. Theory Comput.*, **2018**, *14* (7), pp 3849–3858

CG model of FIP35 from "synthetic FRET"



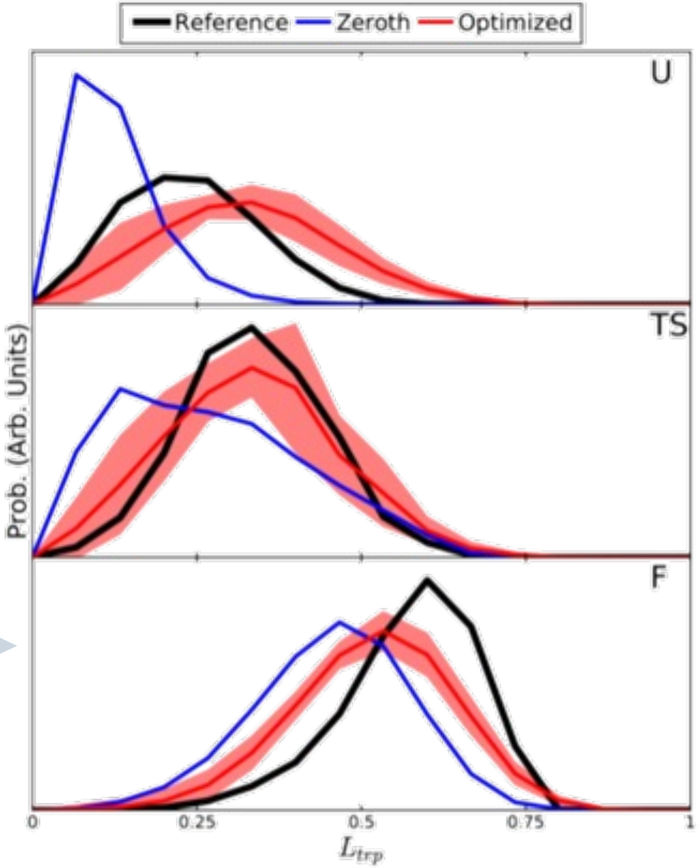
J. Chen, J. Chen, G. Pinamonti & C. Clementi, *JCTC* 14, 3849 (2018)

Model parametrized with FRET reproduces TRP fluorescence

Freie Universität Berlin

Model trained on synthetic FRET data can better reproduce folding mechanism and TRP fluorescence

Histograms for the simulated Trp fluorescence measurements for unfolded (U), transition (TS), and folded (F) states:

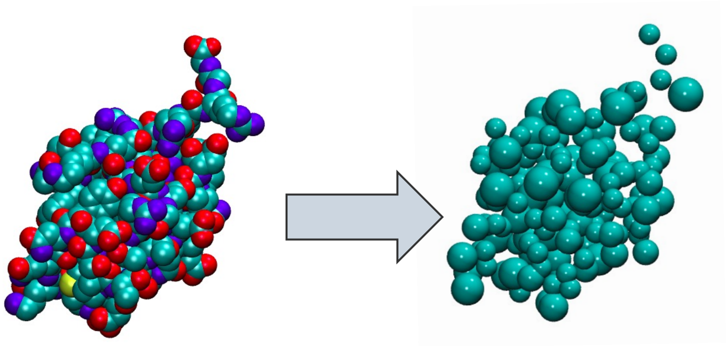


J. Chen, J. Chen, G. Pinamonti & C. Clementi, *JCTC* 14, 3849 (2018)

Model parametrization with NMR data

- **Test system:**

  Ubiquitin, structure-based CαCβ model.

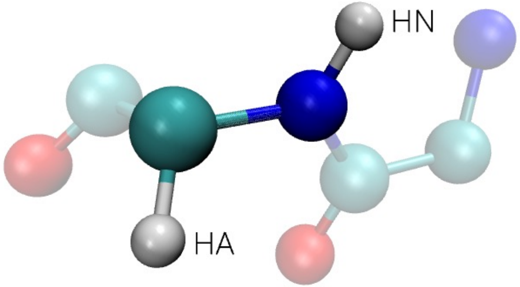


- **Adjustable parameters:**

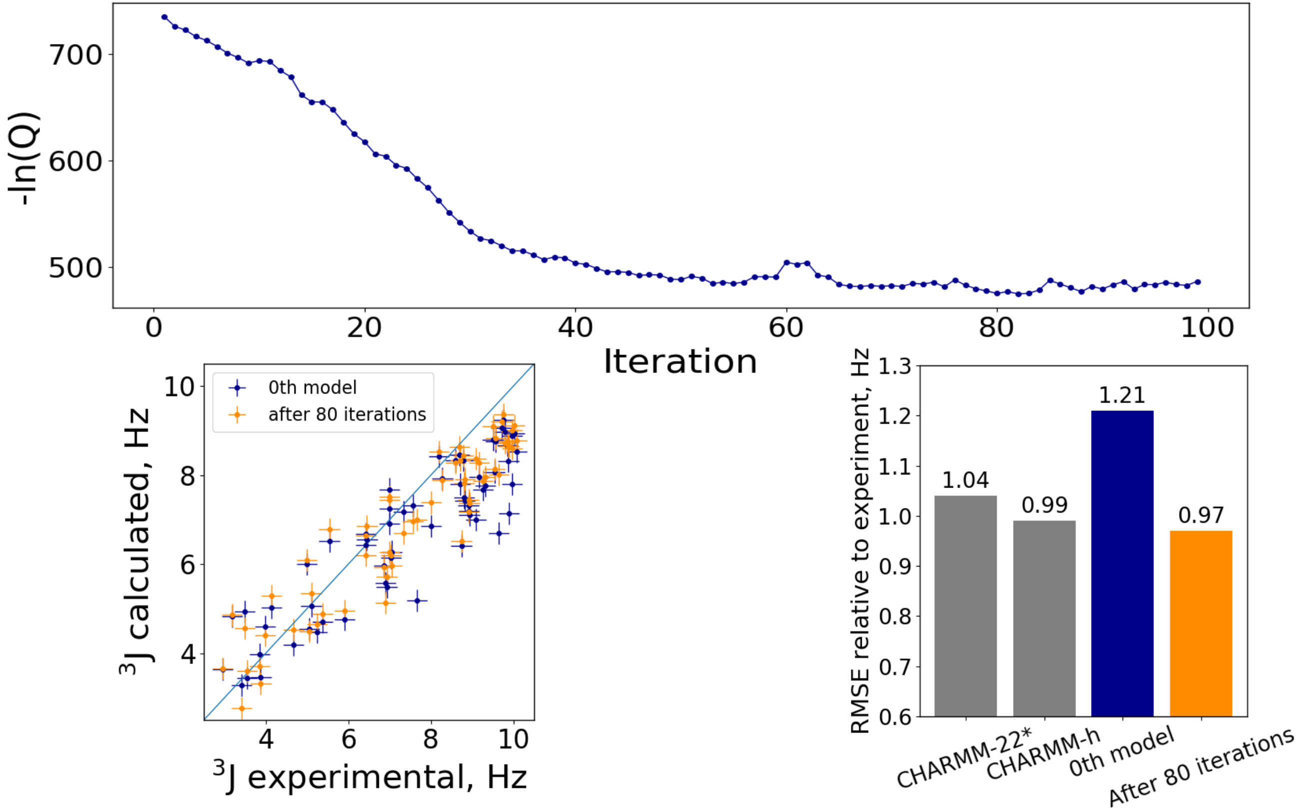
  Strength of native interactions (219 parameters)

- **Experimental observables**:

  Scalar spin-spin coupling constants (^{3}J (HA-HN), N = 63)



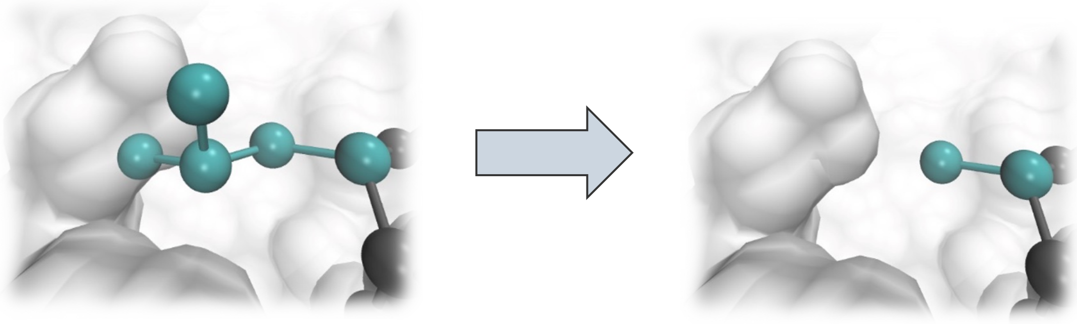
Model parametrization with NMR data



I. Zaporozhets & C. Clementi, unpublished (2022)

Coarse-grained modelling with mutational data

$$\Delta\Delta G = \Delta G_{folding}^{mutant} - \Delta G_{folding}^{wild-type}$$

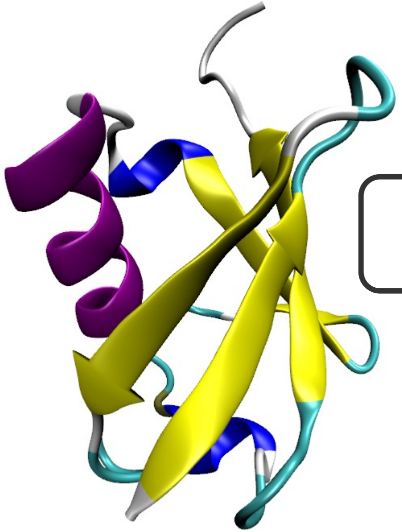


$V^0(\epsilon)$ $\quad +\delta V(\epsilon)$ $\quad V^{mutant}(\epsilon)$

$$\beta\Delta\Delta G = ln \frac{\langle exp(-\beta\delta V(\epsilon))\rangle^{unfolded}}{\langle exp(-\beta\delta V(\epsilon))\rangle^{folded}}$$

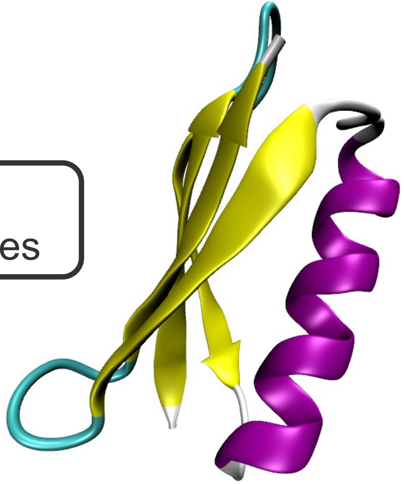
Matysiak, S.; Clementi, C. J. Mol. Biol. 2004, 343 (1), 235–248.

Coarse-grained modelling with mutational data



1UBQ
76 residues

1PGB
56 residues

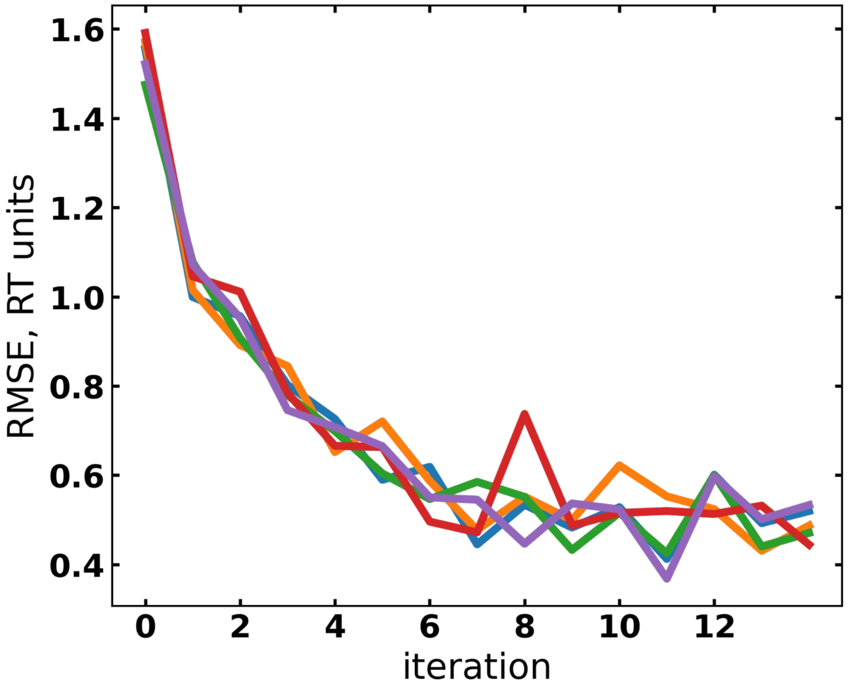


Adjustable parameters:

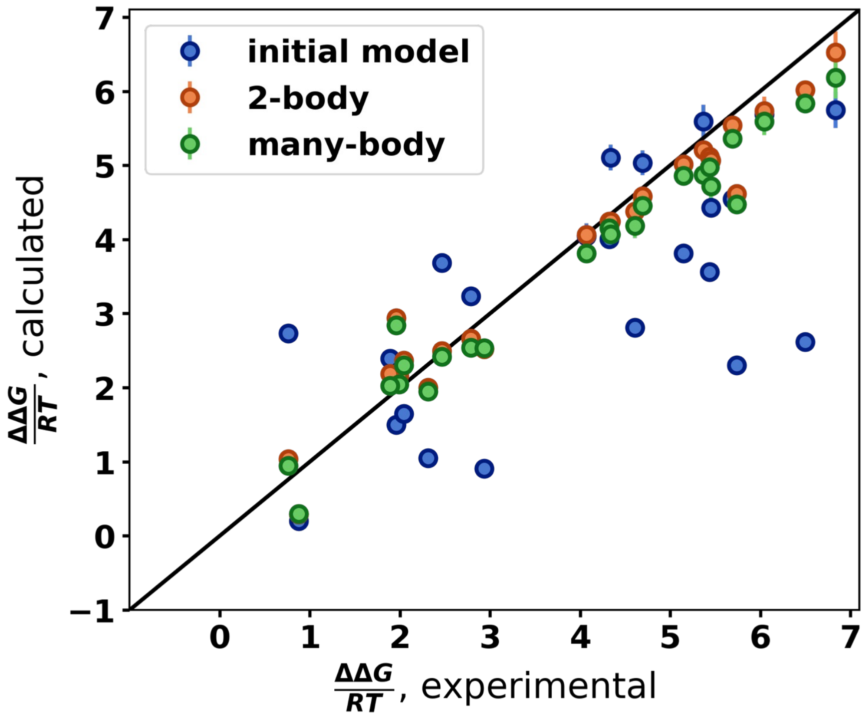
Strength of both native and nonnative interactions

Representation: Cα-model

Coarse-grained modelling with mutational data



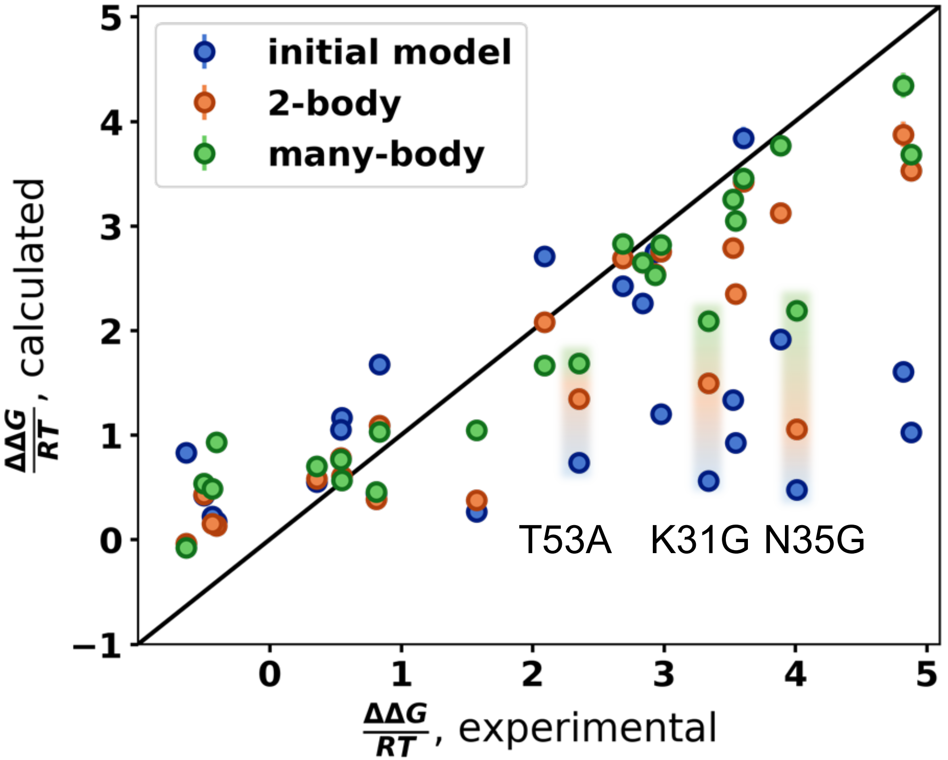
Root mean square error: change during the optimization

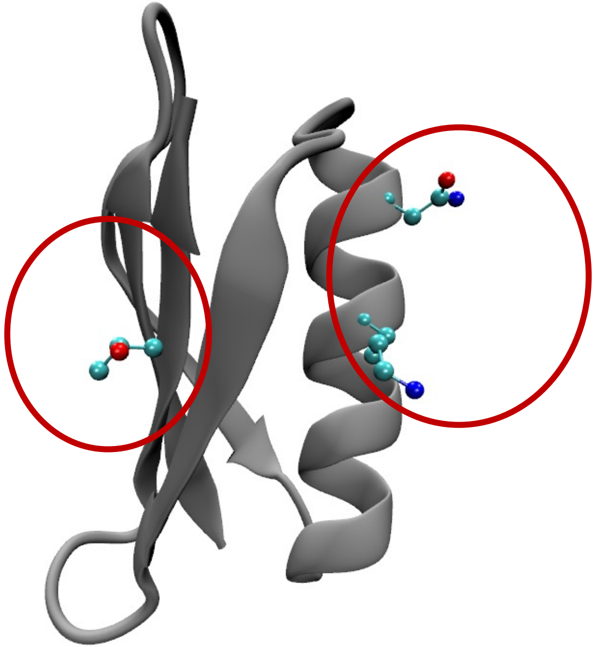


Correlation between calculated and experimental data

I. Zaporozhets & C. Clementi, submitted (2022)

Coarse-grained modelling with mutational data: results for protein G





I. Zaporozhets & C. Clementi, submitted (2022)

Protein G: Multibody terms are needed!

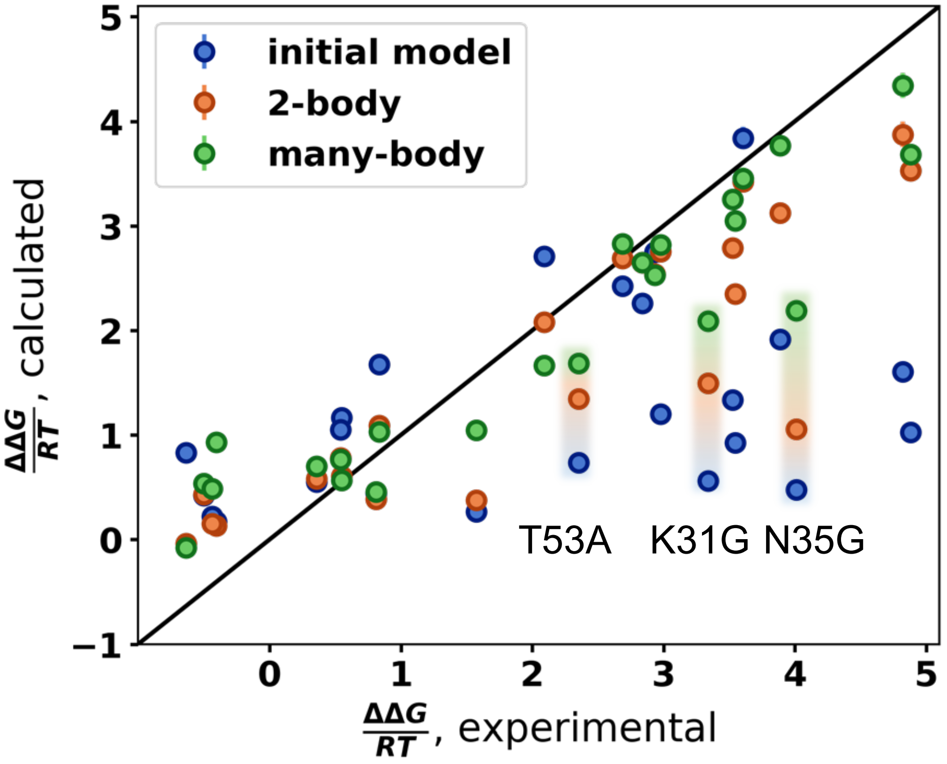
Freie Universität Berlin

$$V = V_{2-\mathrm{body}} + \frac{\lambda_{burial}}{2} \sum_{\mu=1}^{3} \sum_{i=1}^{N} \gamma_{burial}(a_i, \mu) \left[\tanh\left(\eta\left(\rho_i - \rho_{min}^{\mu}\right)\right) + \tanh\left(\eta\left(\rho_{max}^{\mu} - \rho_i\right)\right)\right]$$

$\eta, \lambda_{burial}, \rho_{max}, \rho_{min}$ - fixed parameters

$\gamma_{burial}(a_i, \mu)$ - learnable parameters

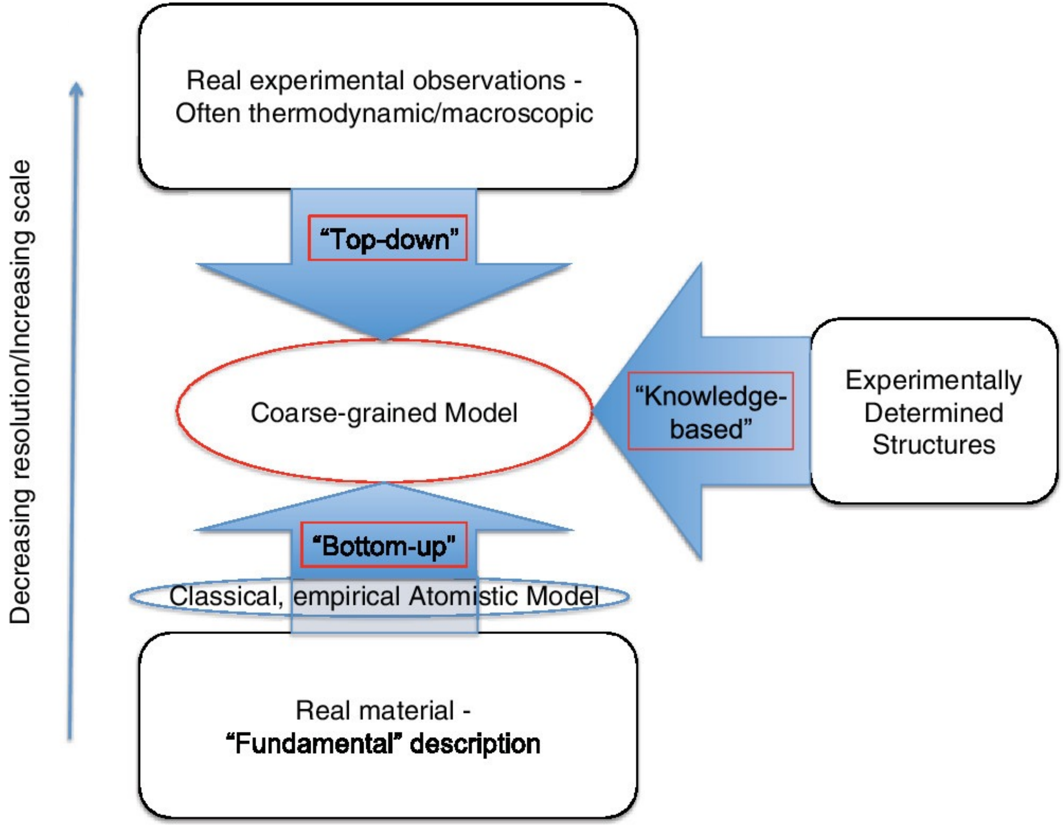
ρ - local density



Additional multibody terms produce better agreement with experimental data

I. Zaporozhets & C. Clementi, submitted (2022)

Future direction



W.G. Noid, *J. Chem. Phys.* 139:9, 090901 (2013)

Ultimate goal:
combine
top-down and bottom-up
approaches to construct
accurate and transferable yet
computationally inexpensive
coarse-grained force fields

Freie Universität Berlin

Cecilia Clementi's research group

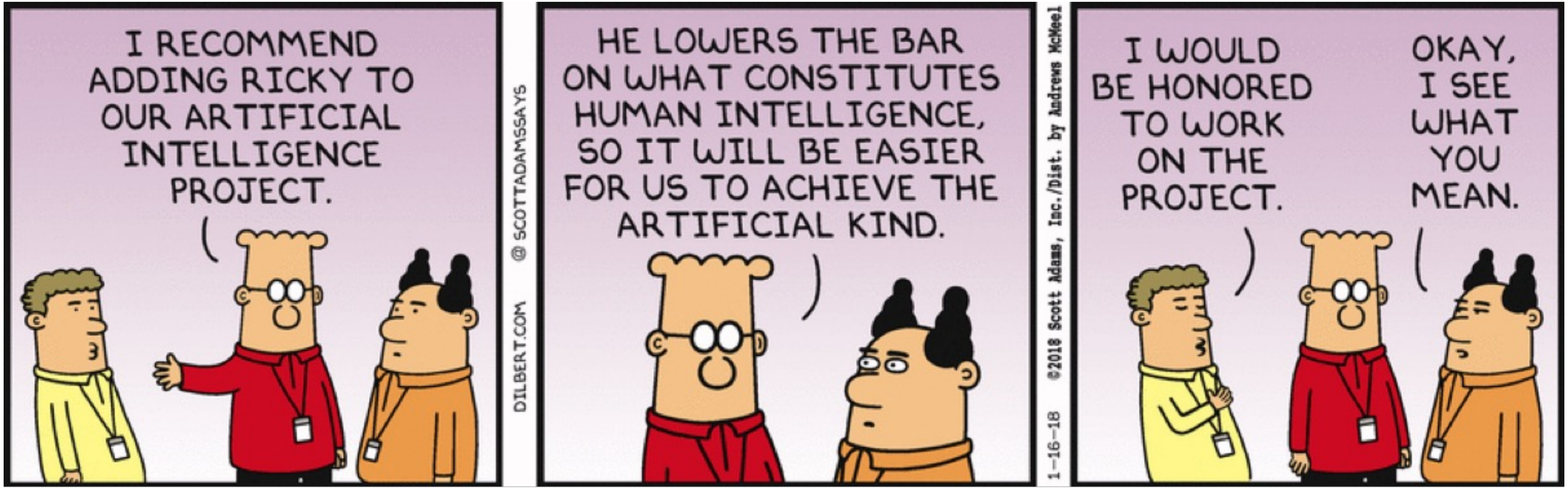
https://www.physik.fu-berlin.de/en/einrichtungen/ag/ag-clementi/

In collaboration with:
Frank Noé
Gianni de Fabritiis
Venkat Kapil

Clementi's group
Dr. Felix Musil
Dr. David Rosenberger
Dr. Clark Templeton
Wangfei Yang
Nick Charron
Iryna Zaporozhets
Tim Hempel
Andrea Guljas
Klara Bonneau
Aldo Sayeg
Jacopo Venturin
Daria Gusew
Arvid Krein

previous:
Dr. Giovanni Pinamonti
Dr. Fabio Trovato
Dr. Jiang Wang (Huaqiao U.)
Dr. Eugen Hruska (Emory)
Dr. Justin Chen (Google)
Dr. Feliks Nüske (U. Padeborn)
Dr. Lorenzo Boninsegna (UCLA)
Dr. Fernando Yrazu (Rice)
Dr. Jordane Preto (U Alberta)
Dr. Mary Rohrdanz (MD Anderson)
Dr. Wenwei Zheng (U Arizona)
Dr. Amarda Shehu (GMU)
Dr. Payel Das (IBM)
Dr. Silvina Matysiak (U. Maryland)
Dr. Brad Lambeth (Shell)

Thank you for your attention!



Questions?