Big Data Summer, 10 Sept 2019

Data-driven methods for soft matter

Max Planck Institute for Polymer Research, Mainz, Germany

Joseph F. Rudzinski

Max-Planck-Institut
für Polymerforschung

Max Planck Institute
for Polymer Research

Big Data Summer, 10 Sept 2019

Data-driven methods for soft matter

Some slides borrowed from Tristan Bereau

Research, Main

Content is highly biased!

Joseph F. Ru

Reviews on machine learning for soft matter

Main References:

Ferguson *J. Phys.: Condens. Matter* **(2018)** “Machine learning and data science in soft materials engineering”

Bereau *Handbook of Materials Modeling: Methods: Theory and Modeling* **(2018)** “Data-Driven Methods in Multiscale Modeling of Soft Matter”

Jackson *et al. Curr. Opin. Chem. Eng.* **(2019)** “Recent advances in machine learning towards multiscale soft materials design”

Klus *et al. Int. J. Nonlinear Sci.* **(2018)** “Data-Driven Model Reduction and Transfer Operator Approximation”

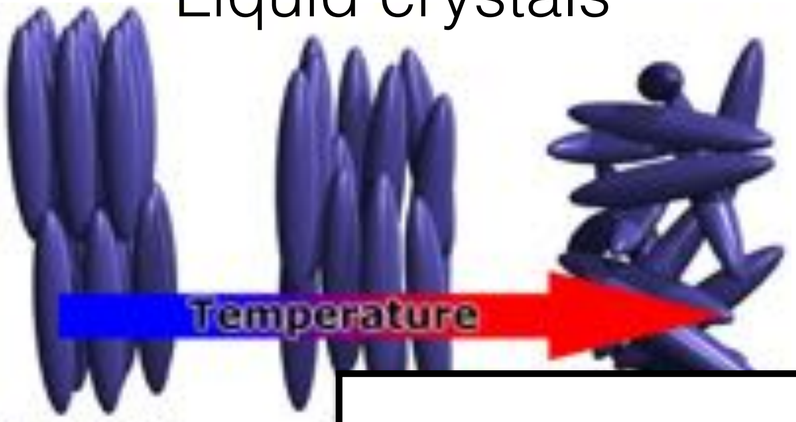
Noe *ArXiv* **(2019)** “Machine Learning for Molecular Dynamics on Long Timescales”

Kingma, Welling *ArXiv* **(2017)** “An Introduction to Variational Autoencoders”

Self assembly of soft-matter materials

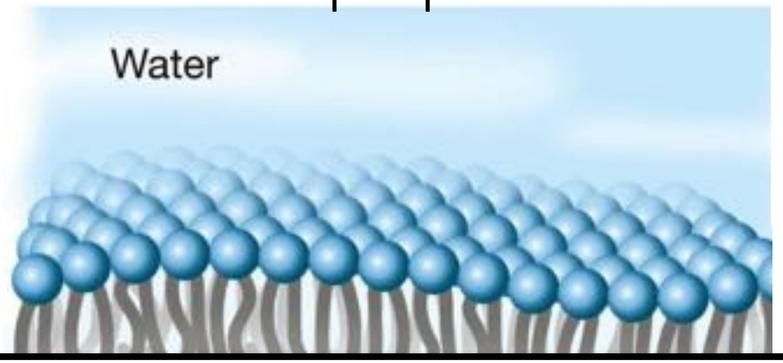
Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Liquid crystals

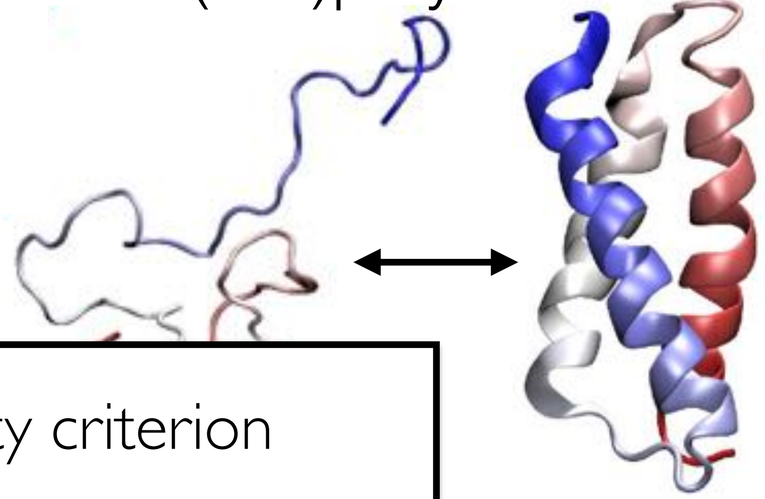


photonicswiki.org

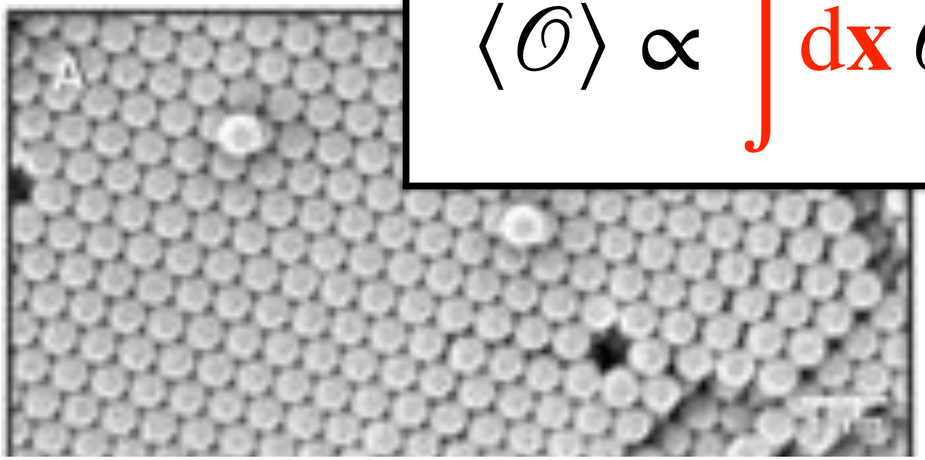
Amphiphiles



(Bio)polymers



Colloid



Stein, Schroden, *Curr Opin Solid State Mater Sci* (2001)

Configurational average

$$\langle \mathcal{O} \rangle \propto \int \mathbf{dx}\, \mathcal{O}(\mathbf{x})\, \mathrm{e}^{-\beta E(\mathbf{x})}$$

Stability criterion

$$F = E - TS$$

Weak interactions:

$$E \sim k_{\mathrm{B}}T$$

Thermal fluctuations lead to spontaneous self-assembly, mesoscale structures

Multiscale simulations

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

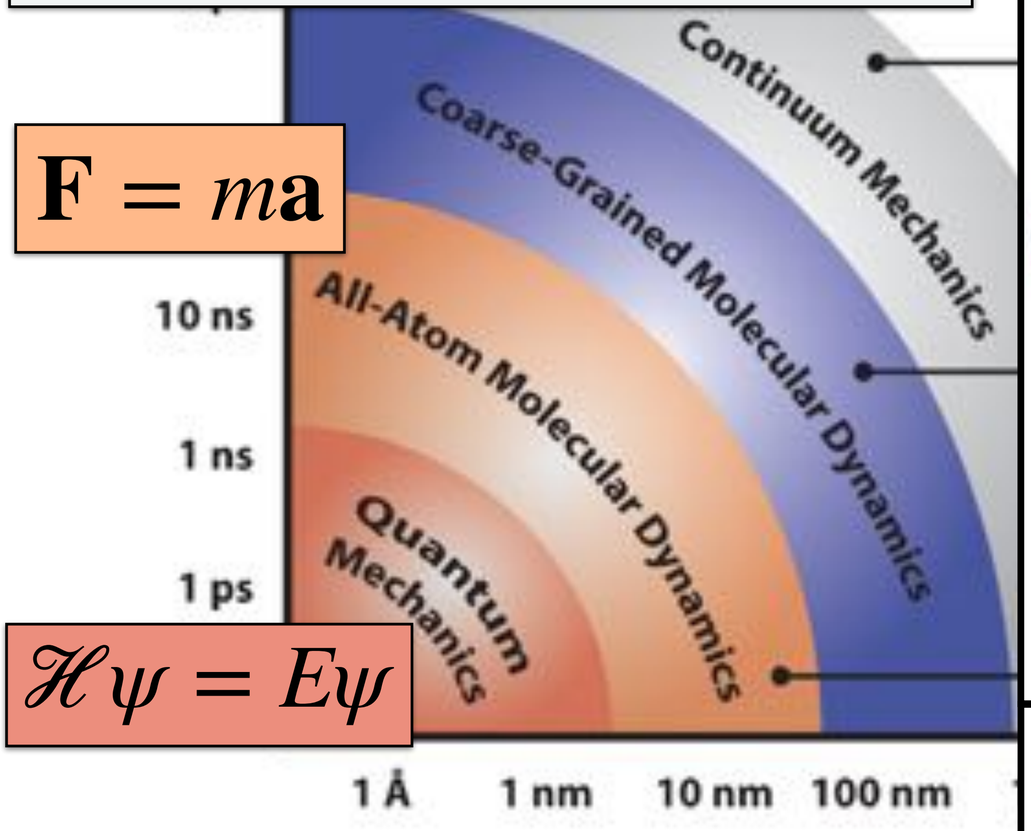
time

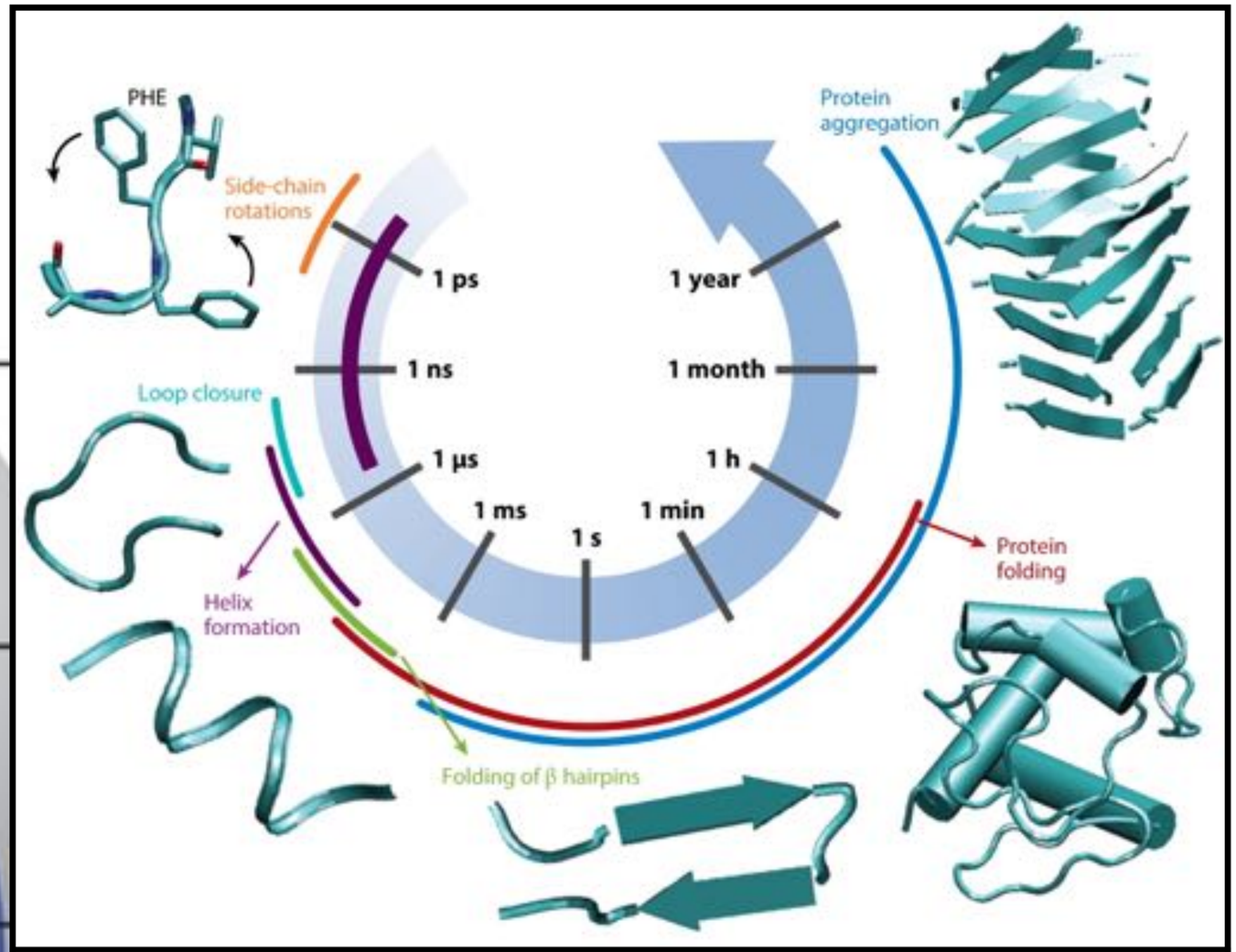
$$E_{\text{bend}} = \int_{\text{membrane}} \mathrm{d}A \left\{ \frac{1}{2}\kappa K^2 - \bar{\kappa} K_{\text{G}} \right\}$$

Helfrich, *Z. Naturforsch. C* **28** (1973)

$$\mathbf{F} = m\mathbf{a}$$

$$\mathscr{H}\psi = E\psi$$





Coupled hierarchy of kinetic processes leading to structure formation

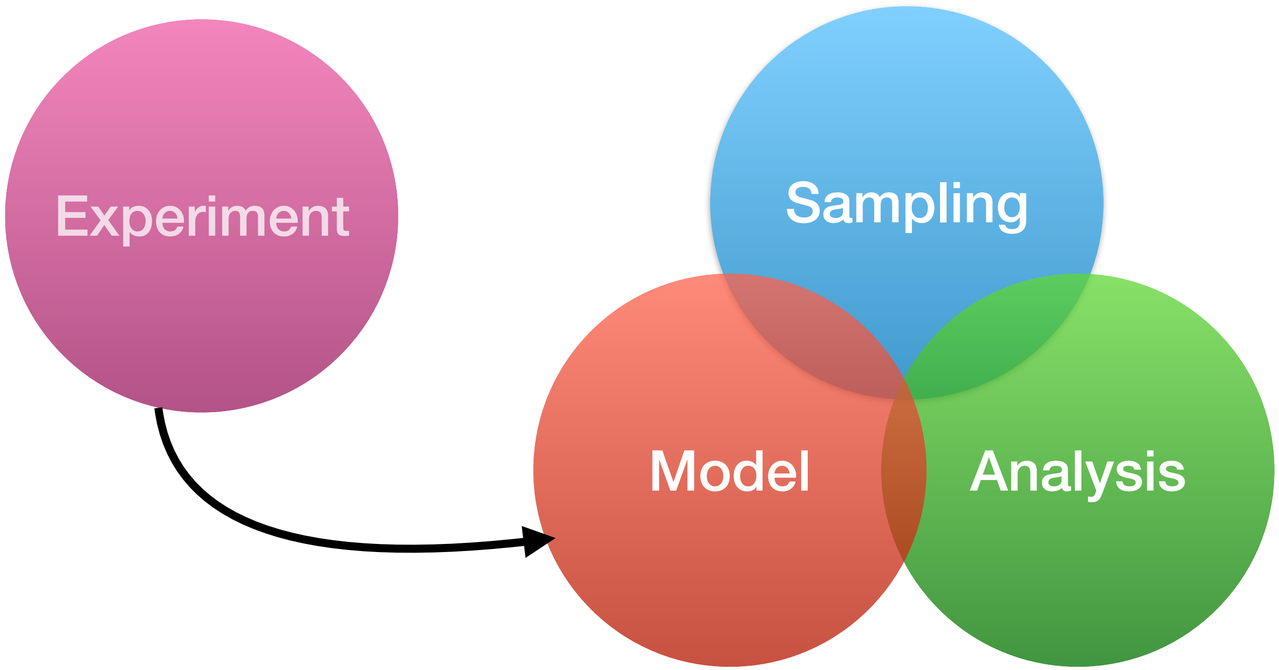
Bradley and Radhakrishnan, *Polymers* **5** (2013)

Fundamental challenges for soft matter modeling

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

bayesian reweighting
force field refinement

enhanced sampling
high throughput studies



force field development
coarse-grained modeling

kinetic modeling
(dimensionality reduction / clustering)

Bayesian inference

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Prediction { Prior beliefs; Sampled data

Embed data-driven techniques in physics-based models

M model

D data

Bayes' theorem

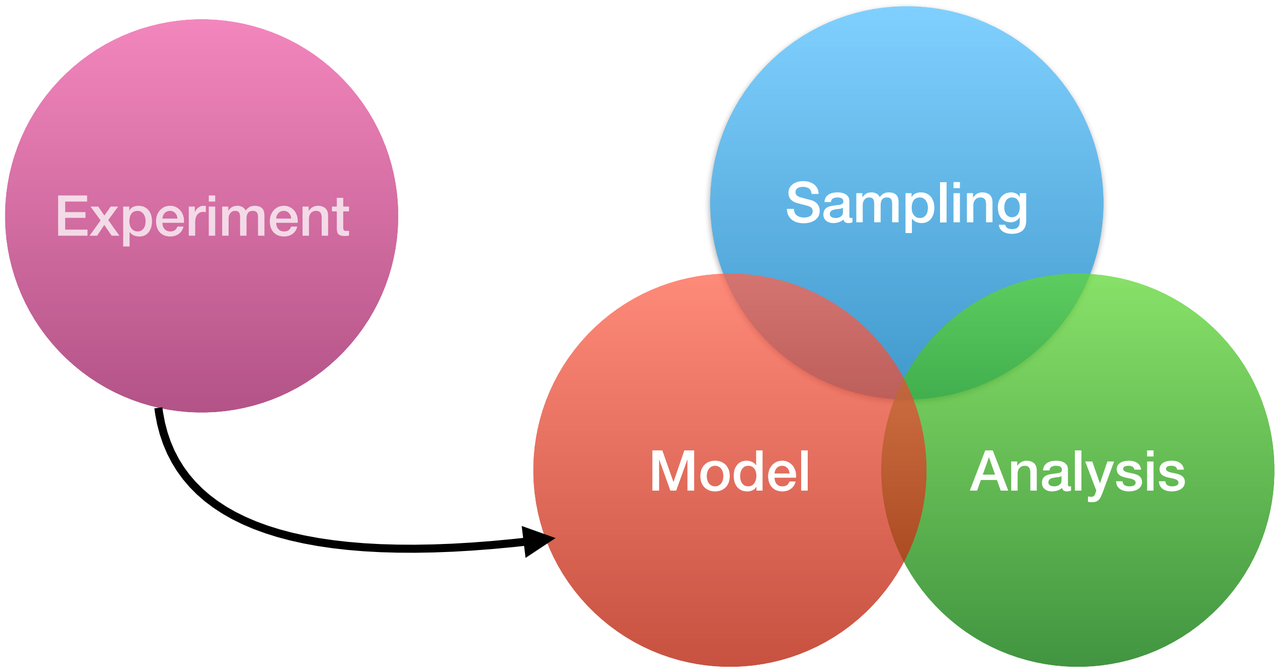
$$\underset{\text{posterior}}{p(M \mid D)} \propto \underset{\text{likelihood}}{p(D \mid M)}\ \underset{\text{prior}}{p(M)}$$

Fundamental challenges for soft matter modeling

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

bayesian reweighting
force field refinement

enhanced sampling
high throughput studies

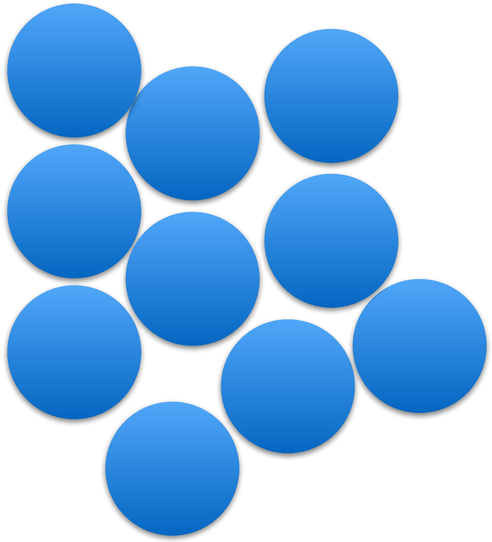


force field development
coarse-grained modeling

kinetic modeling
(dimensionality reduction / clustering)

Molecular dynamics

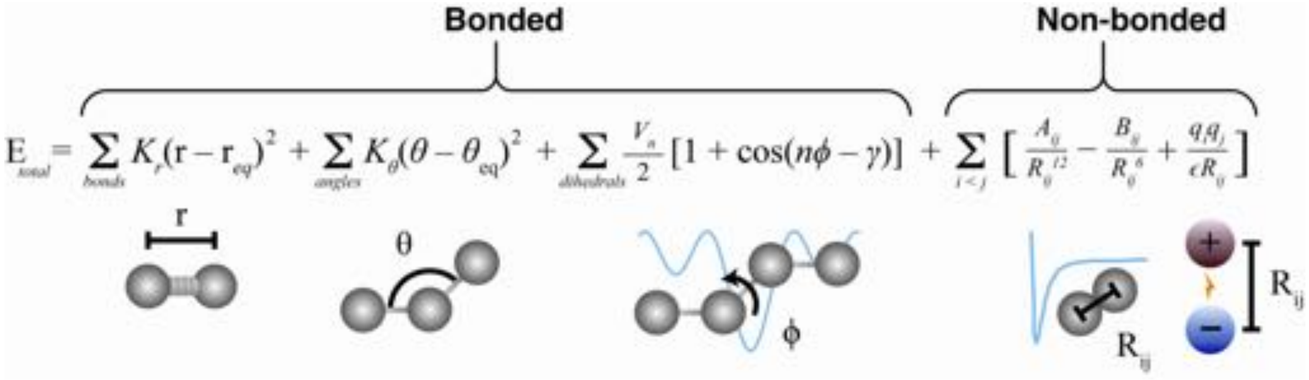
Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research



Numerically integrate particle positions

$$\mathbf{F} = m\mathbf{a}$$

Specify interparticle forces: “force field”



Durrant & McCammon, *BMC Biol* **9** (2011)

integration time step

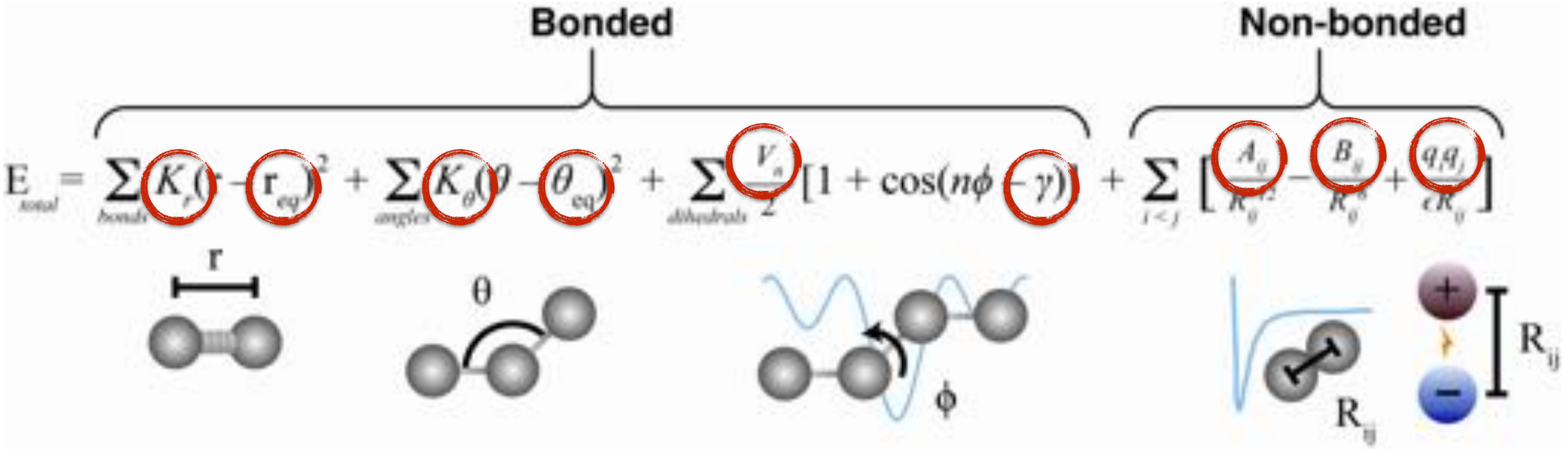
Timescales of interest

Emergent complexity

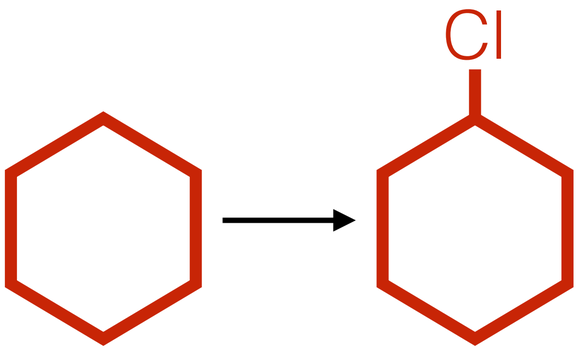
fs ps ns μs ms s

The craftsmanship of force-field parametrization

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research



Every new molecule requires parameter optimization



Durrant & McCammon, *BMC Biol* **9** (2011)

Transferable model for intermolecular interactions

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

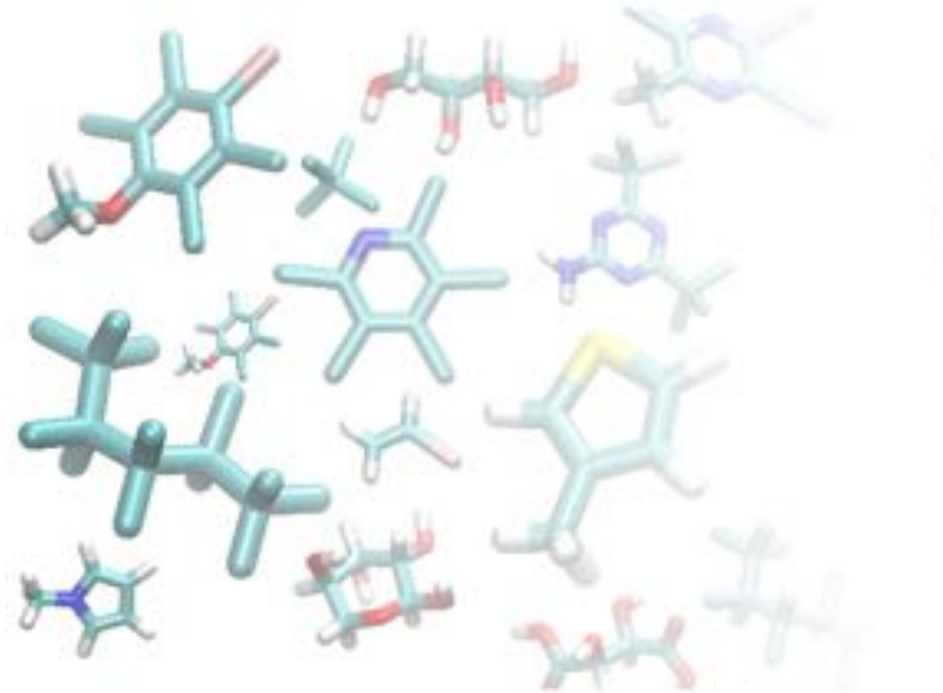
Physics-based models

- Encode laws, symmetries
- Little chemical information

Data-driven models

- Need to *learn* laws, symmetries
- Interpolate across chemistry

Any small molecule made of H, C, O, N
neutral compounds



Kernel methods are vintage

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Kernel

- needs a representation
- linear algebra
- can be efficient with small data

Deep learning

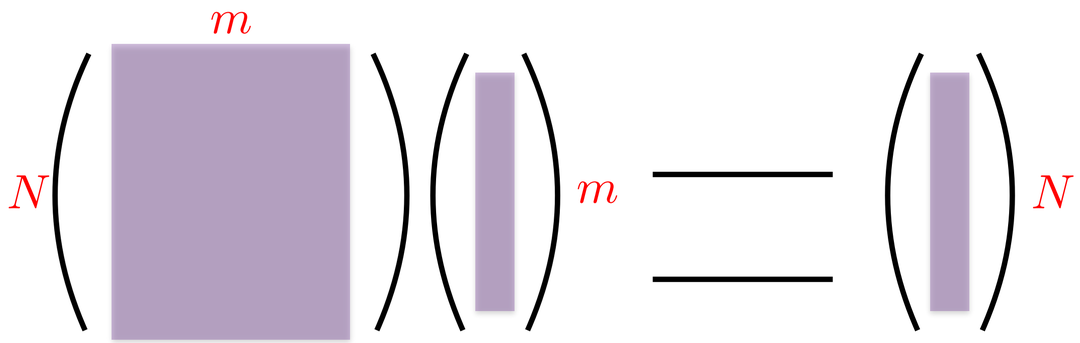
- learns the representation
- complex mathematical structure
- data hungry

Linear-ridge vs kernel-ridge regression

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Linear-ridge regression

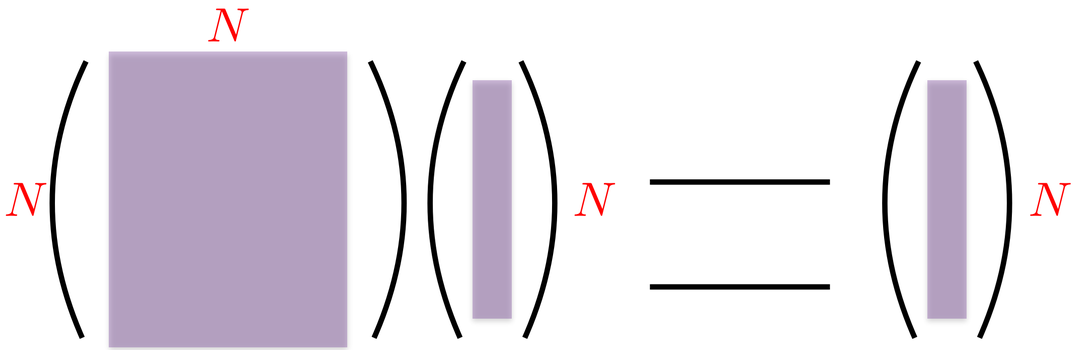
$$\mathbf{A}x = b$$



in general: $m \ll N$

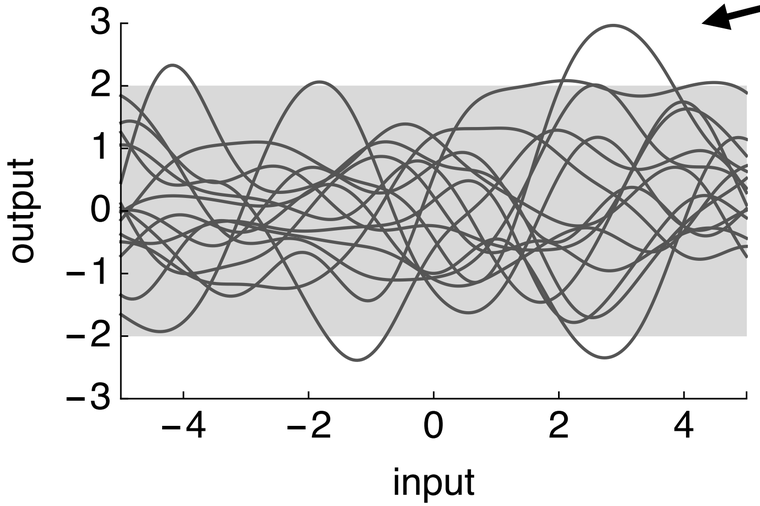
kernel-ridge regression (ML)

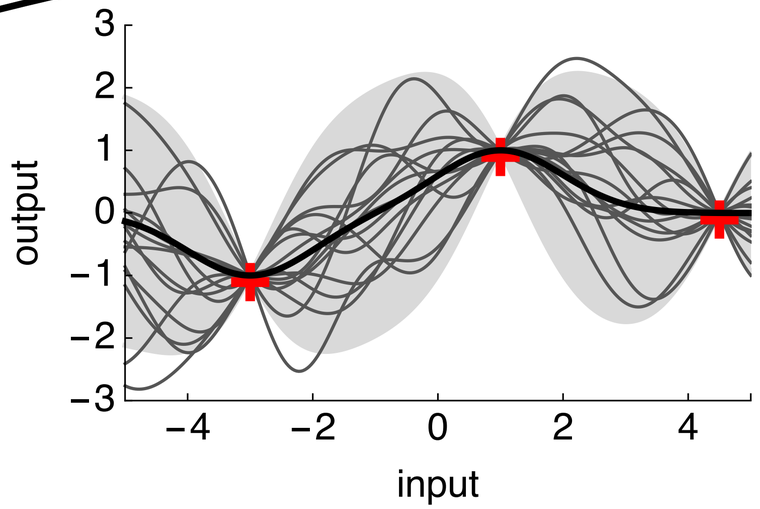
$$\mathbf{K}\alpha = p$$



$$\begin{aligned} K_{ij} &= K_{ij}(\mathbf{x}_i, \mathbf{x}_j) \\ &= K_{ij}(|\mathbf{x}_i - \mathbf{x}_j|) \\ &= \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|}{\sigma}\right) \end{aligned}$$

Bayesian inference





Bayes' formula

$$p(f, f^* \mid y) = \frac{p(y \mid f)\; p(f, f^*)}{p(y)}$$

likelihood

prior

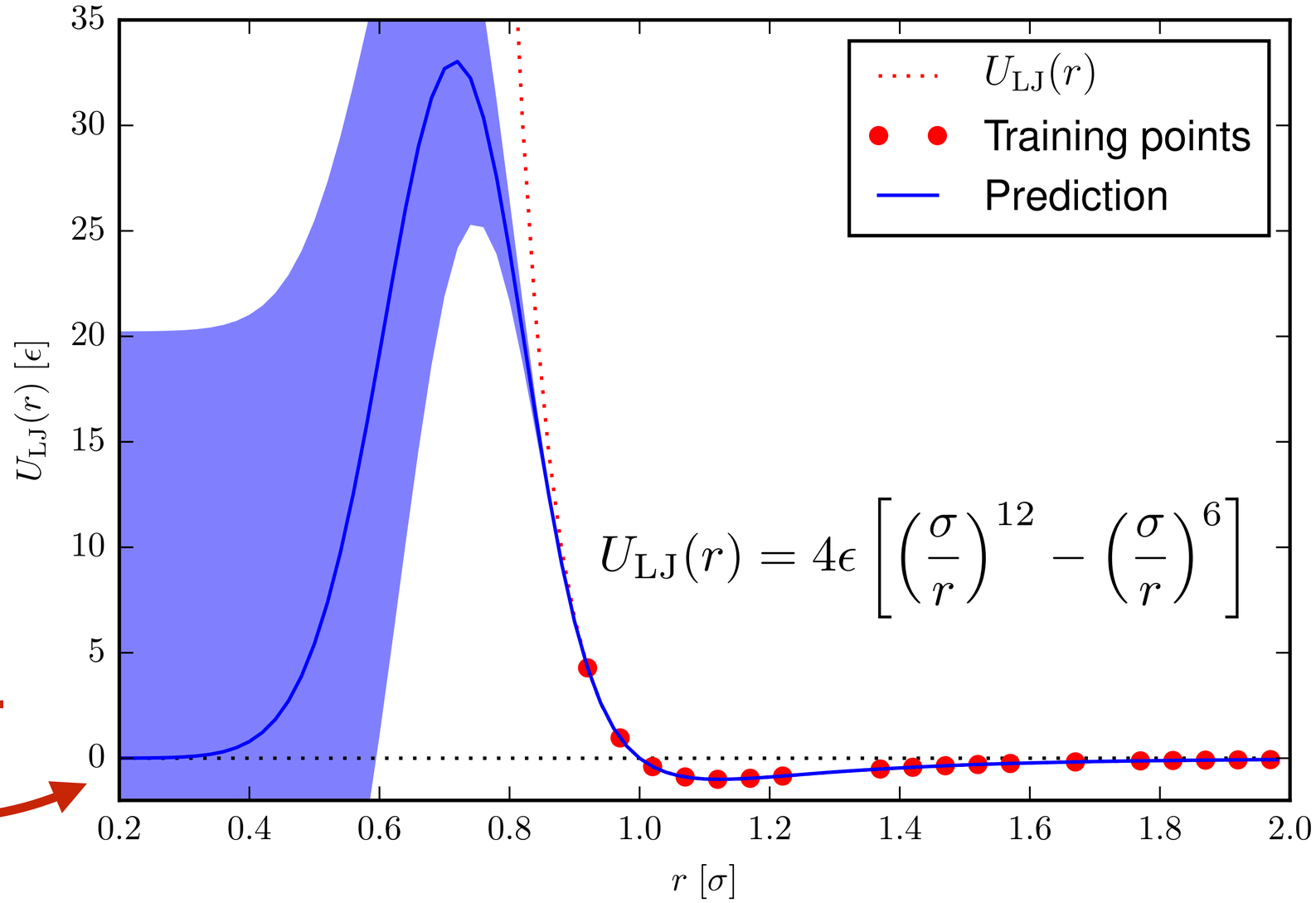
posterior

normalization

Rasmussen, *Advanced lectures on machine learning*. Springer, 63-71 (2004)

Extrapolation in machine learning

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research



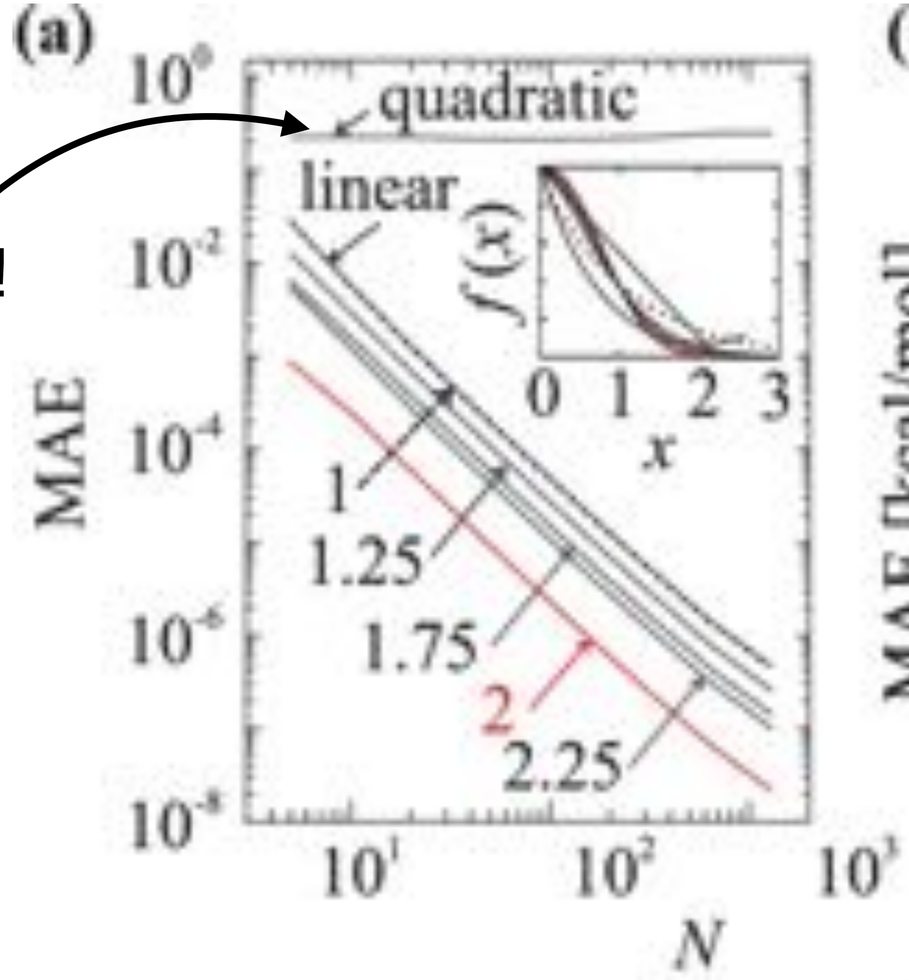
Uniform prior

Optimizing the representation links to the physics

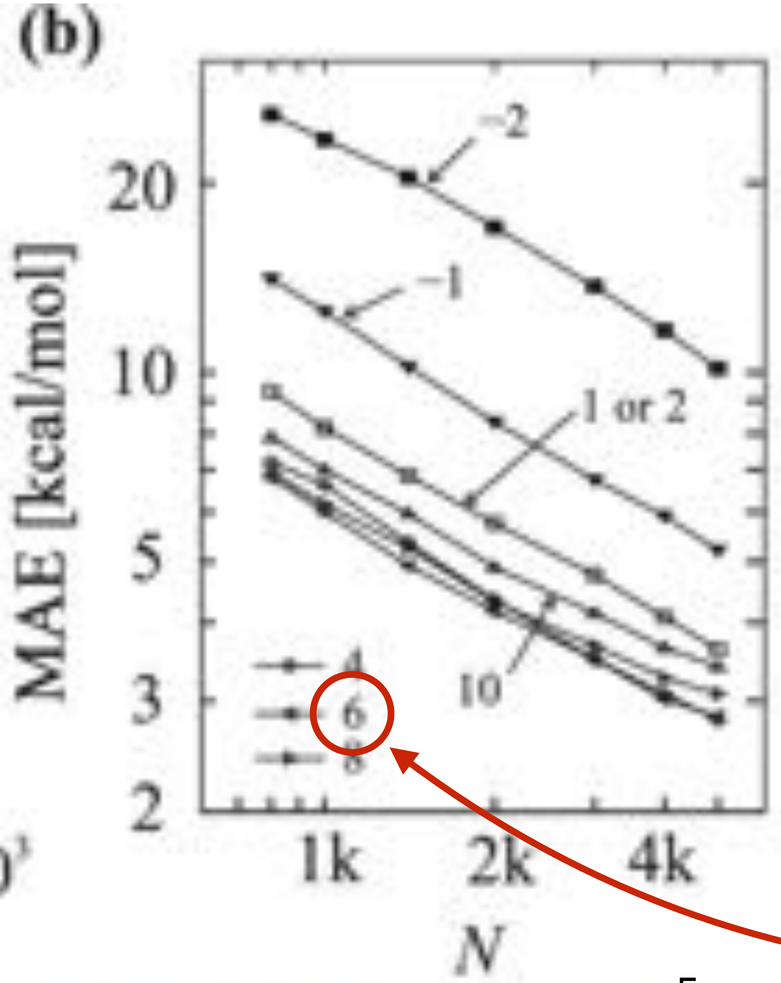
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Learning a Gaussian function

Non-unique representation!



Learning atomization energies



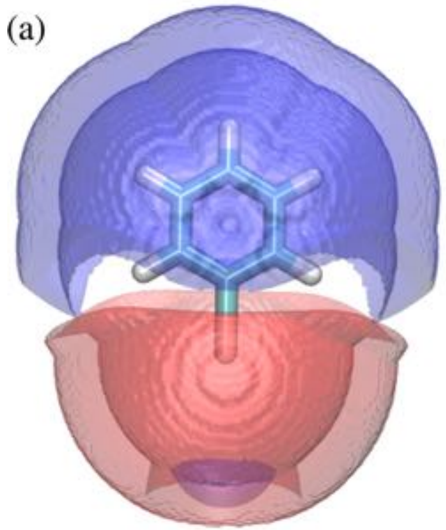
$$U_{\mathrm{LJ}}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

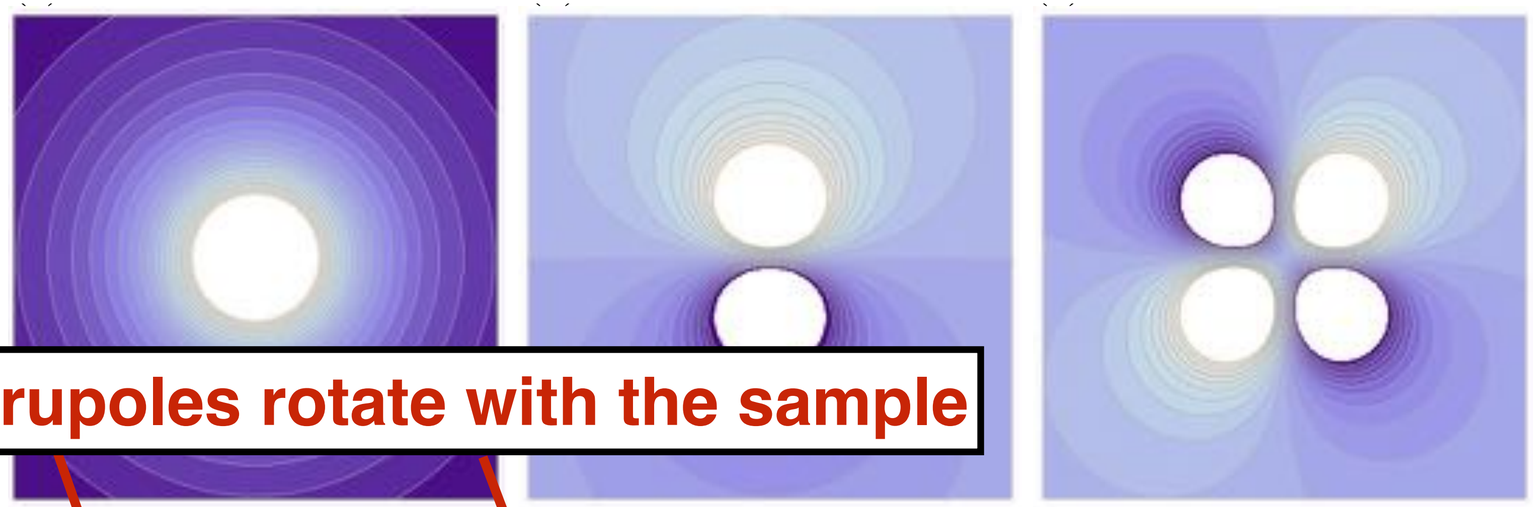
Huang and von Lilienfeld, *J Chem Phys* **145** (2016)

Static multipole electrostatics

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

$$V_{\mathrm{Coulomb}}(r) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r}$$





dipoles, quadrupoles rotate with the sample

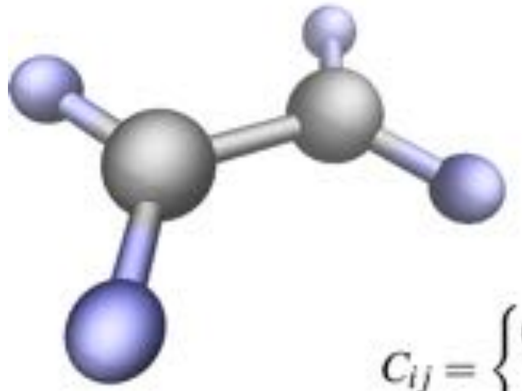
$$4\pi\varepsilon_0\Phi(\mathbf{r}) = \frac{q}{R} + \frac{\mu_\alpha R_\alpha}{R^3} + \frac{1}{3}\frac{\Theta_{\alpha\beta}\,3R_\alpha R_\beta - R^2\delta_{\alpha\beta}}{R^5} + \ldots$$

Stone, *The Theory of Intermolecular Forces*
Bereau and Meuwly, *Many-Body Effects and Electrostatics in Biomolecules*

Representation: the Coulomb matrix

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Symmetries of the representation should emulate symmetries of the system that keep the total energy constant



$$C_{ij} = \begin{cases} 0.5\, Z_i^{2.4} & \forall i = j \\ \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} & \forall i \neq j. \end{cases}$$

$\longrightarrow$ $\mathbf{C} =$

	H	H	C	C	H	H
H	0.5	0.3	2.9	1.5	0.2	0.2
H	0.3	0.5	2.9	1.5	0.2	0.2
C	2.9	2.9	36.9	14.3	1.5	1.5
C	1.5	1.5	14.3	36.9	2.9	2.9
H	0.2	0.2	1.5	2.9	0.5	0.3
H	0.2	0.2	1.5	2.9	0.3	0.5

1. Translation
2. Rotations
3. Mirror reflection

~ Coulomb's law $E = \dfrac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}$

Problems:

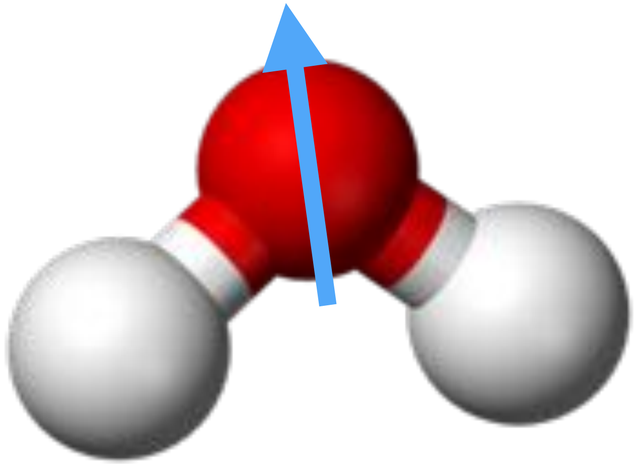
1. **Dimensionality from # atoms**
2. **Ordering of the atoms**

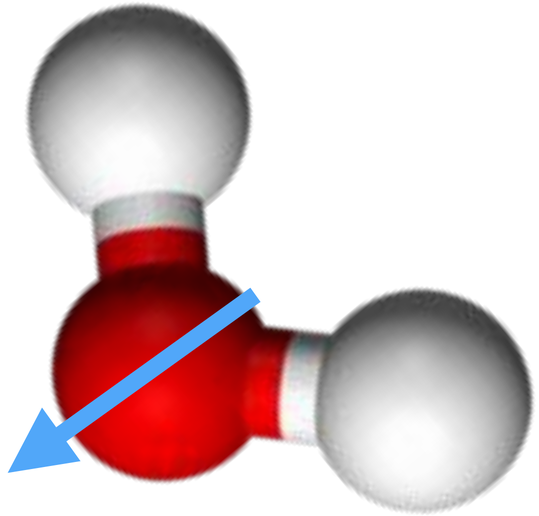
Hansen et al., *J Chem Theory Comput*, **9** (2013)

Covariant kernels

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Encode rotational properties of the target property in the **kernel**





"Build kernel so as to encode the rotational properties of the target property"

Glielmo, Sollich, De Vita, *Phys Rev B* **95** (2017)

Covariant kernels

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Encode rotational properties of the target property in the **kernel**

$$\hat{\mathbf{f}}(\mathcal{S}\rho \mid \mathcal{D}) = \mathbf{S}\hat{\mathbf{f}}(\rho \mid \mathcal{D})$$

Force prediction

Descriptor

Training data

Transformation (rotation/inversion)

Glielmo, Sollich, De Vita, *Phys Rev B* **95** (2017)

Covariant kernels

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Encode rotational properties of the target property in the **kernel**

$$\mathbf{K}(\mathcal{S}\rho, \mathcal{S}'\rho') = \mathbf{S}\mathbf{K}(\rho, \rho')\mathbf{S}'^{\mathrm{T}}$$

Kernel

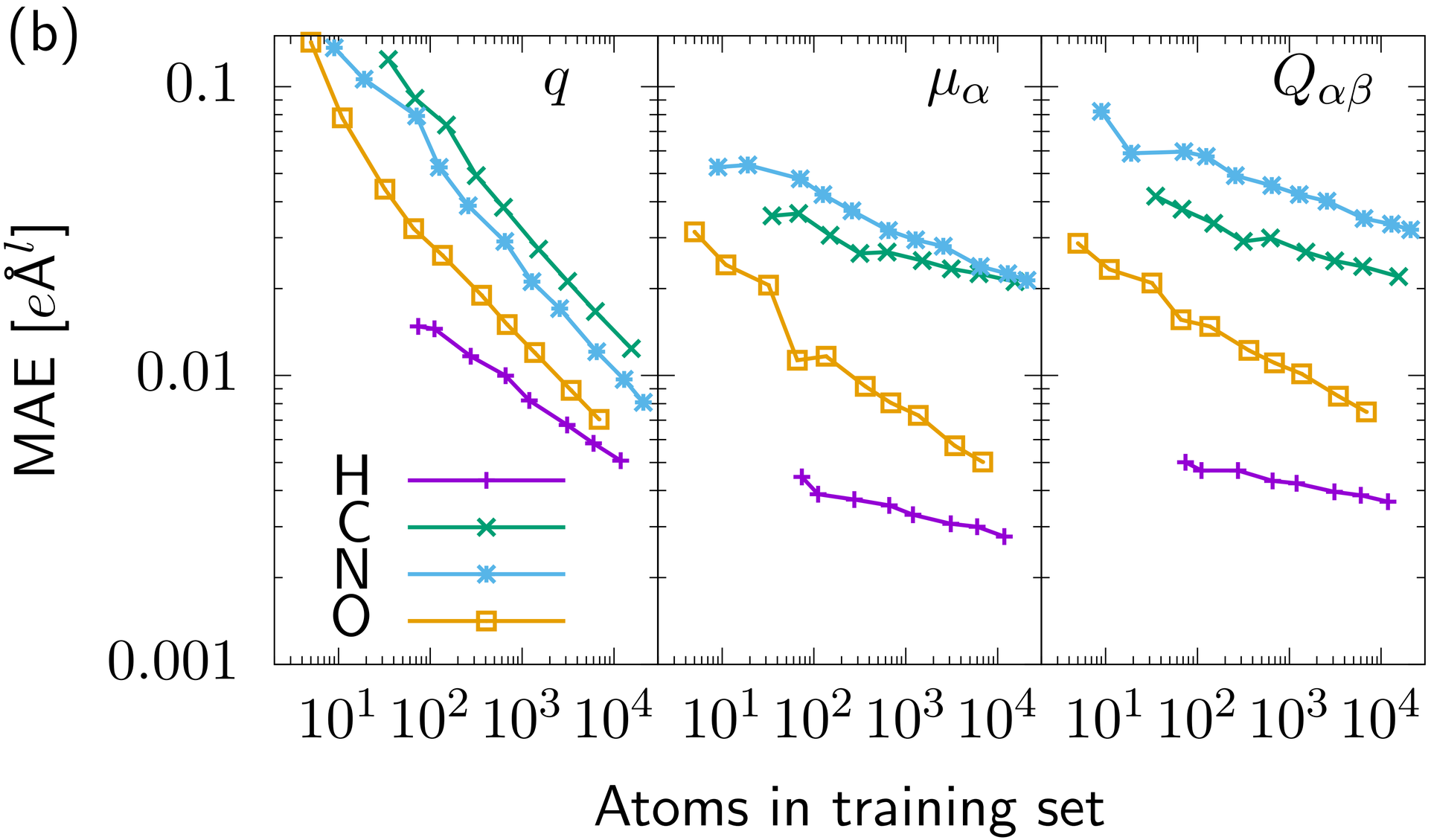
Transformations (rotation/inversion)

Configurations

1. Derivative of energy kernel
2. Integrate over all relevant orientations
3. Local axis system

Glielmo, Sollich, De Vita, *Phys Rev B* **95** (2017)

Multipoles: Learning curves



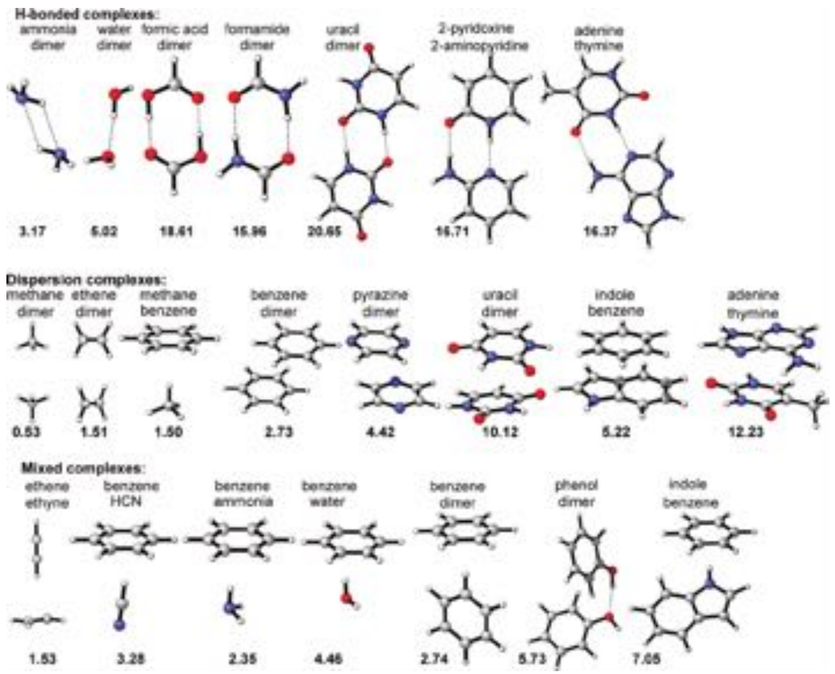
Easier to learn H,O than C,N

Bereau, DiStasio Jr., Tkatchenko, von Lilienfeld, *JCP* **148**, 241706 (2018); JCP Editor's Choice 2018

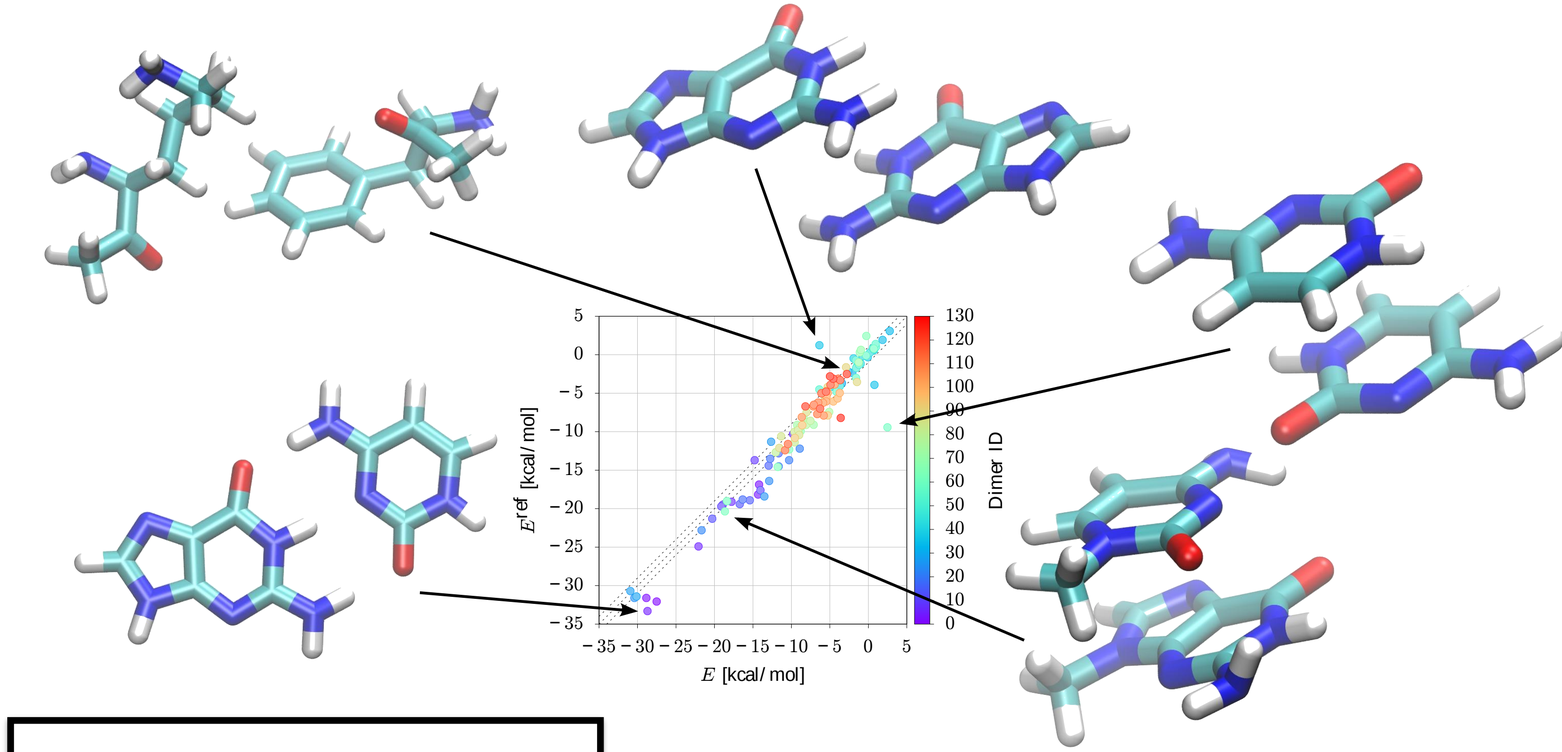
Intermolecular energy across conformations and composition

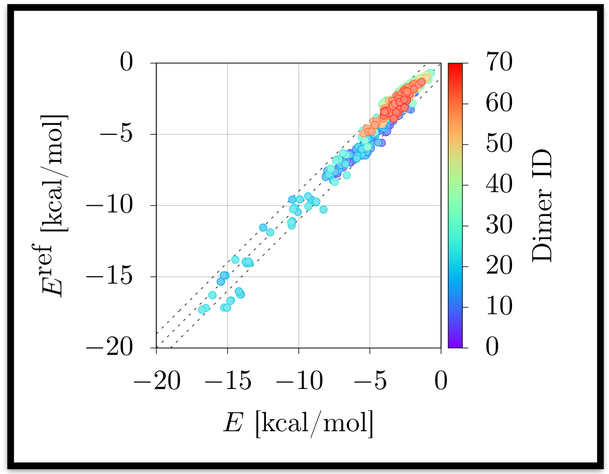
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

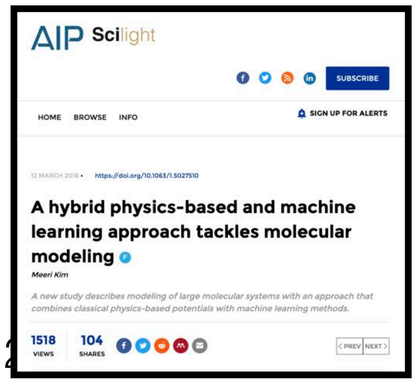
7 global parameters



Jurecka, Sponer, Cerny, Hobza, *PCCP* (2006)
Paton and Goodman, *J Chem Inf Model* (2009)







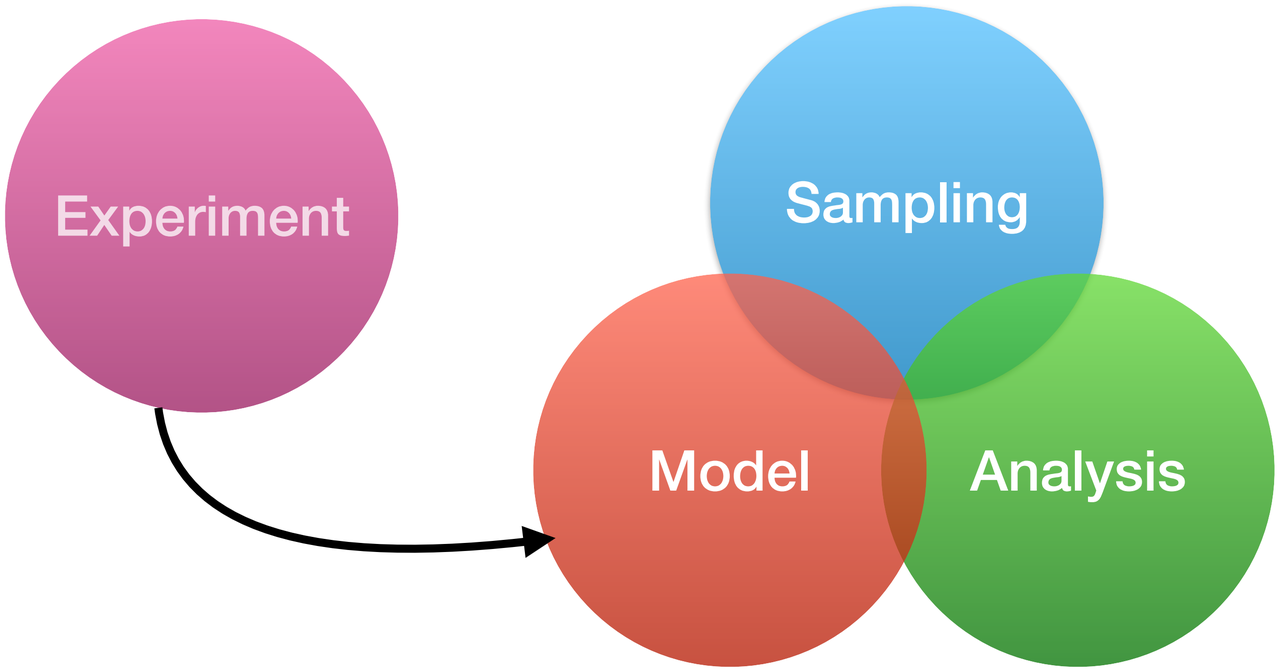
Bereau, DiStasio Jr., Tkatchenko, von Lilienfeld, *JCP* **148**, 241706 (2018); JCP Editor's Choice 2

Fundamental challenges for soft matter modeling

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

bayesian reweighting
force field refinement

enhanced sampling
high throughput studies



force field development
coarse-grained modeling

kinetic modeling
(dimensionality reduction / clustering)

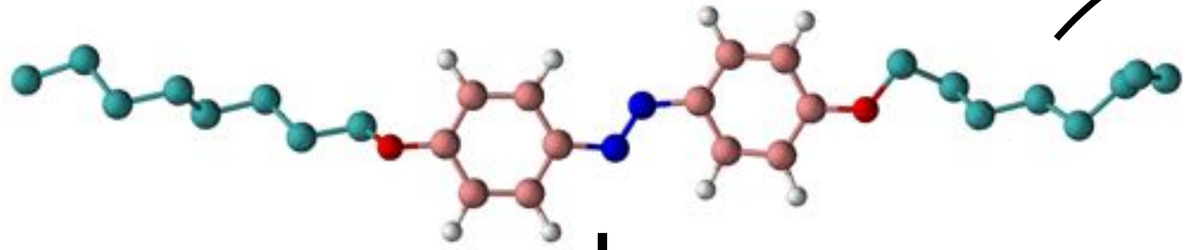
Coarse-graining as an inverse problem

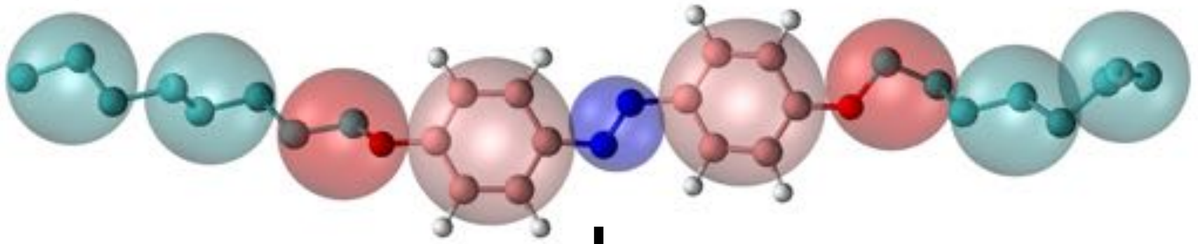
Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

high-resolution

CG mapping

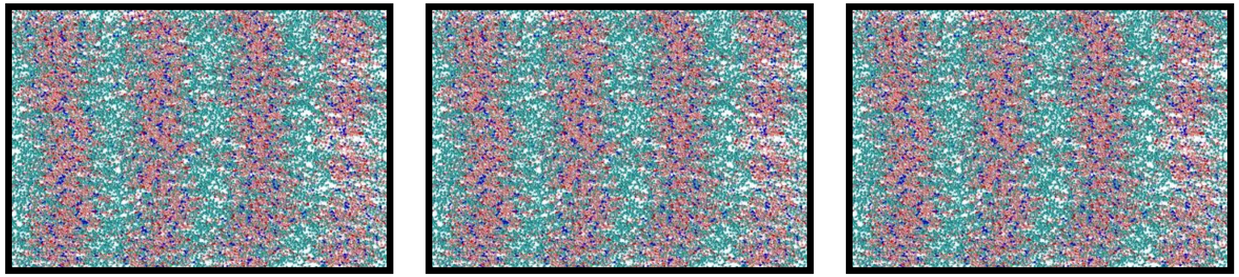
coarse-grained

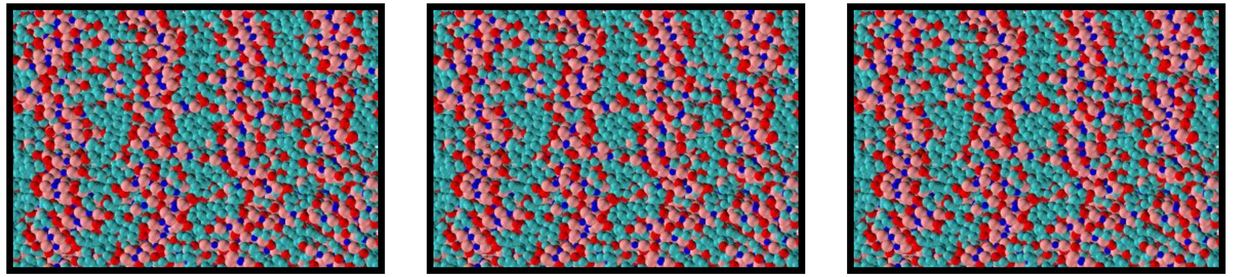




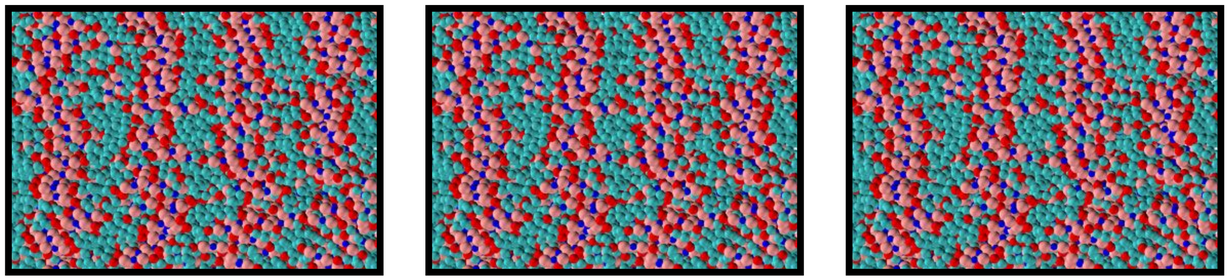
simulate with u(r)

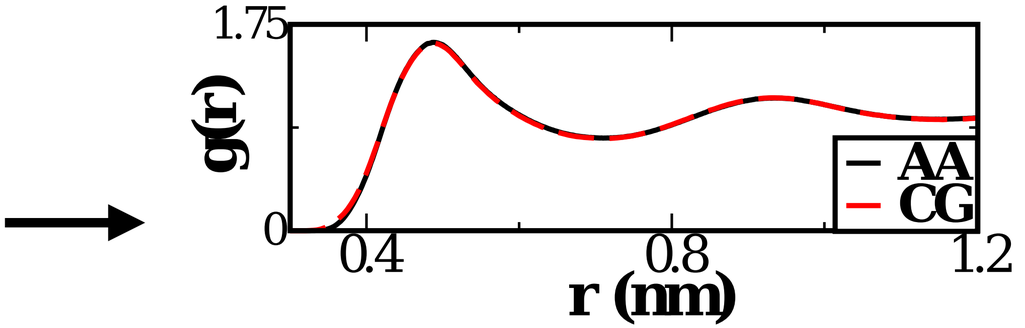
simulate with U(R)





CG mapping

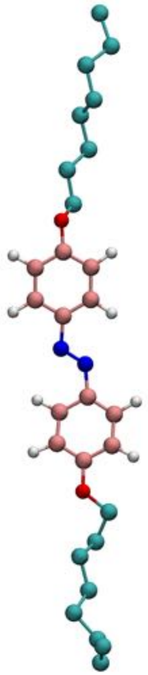


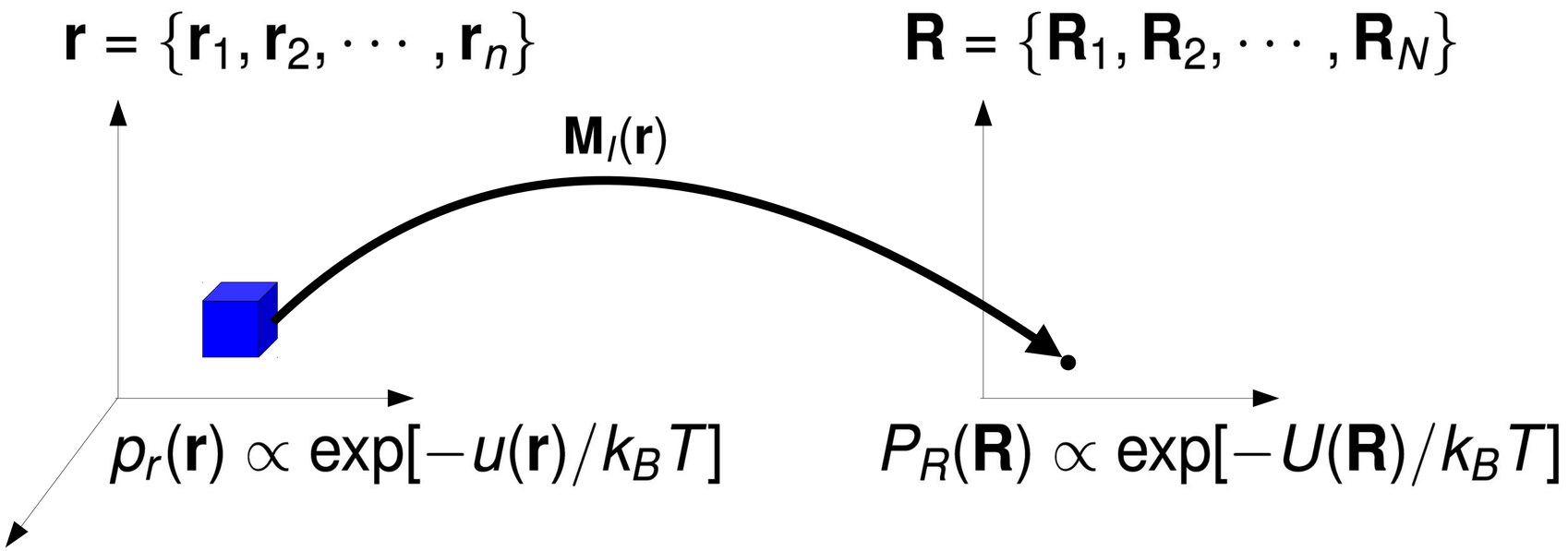


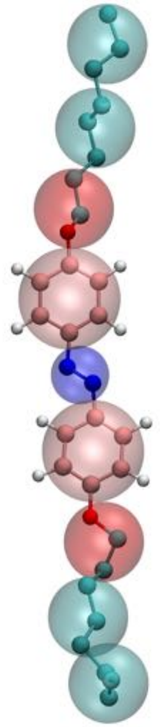
reproduce Boltzmann distribution at CG level

Coarse-graining as an inverse problem

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research







$$p_R(\mathbf{R}) = \int d\mathbf{r}\; p_r(\mathbf{r})\delta(\mathbf{M}(\mathbf{r}) - \mathbf{R})$$

Mapped high-resolution configurational probability

$$p_R(\mathbf{R}) = P_R(\mathbf{R})$$

Consistency criterion

$$U^0(\mathbf{R}) = -k_B T \ln p_R(\mathbf{R}) + \text{const}$$

Many-body potential of mean force

Kirkwood *J Chem Phys* (**1935**); Noid et al. *J Chem Phys* (**2008**)

Coarse-graining as an inverse problem

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

$$U^0(\mathbf{R}) = -k_B T \ln p_R(\mathbf{R}) + \text{const}$$

Many-body potential of mean force

$$U^0(\mathbf{R}) \approx U(\mathbf{R})$$

Approximate CG potential

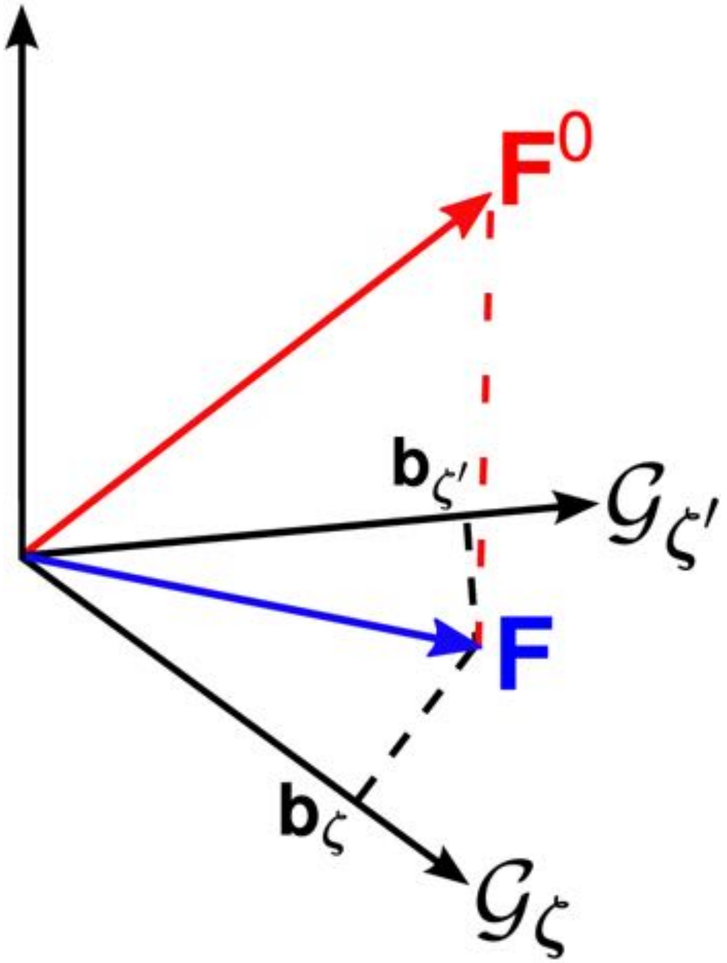
$$= \sum_{\text{bonds}(R_b)} U_b(R_b) + \sum_{\text{angles}(\theta)} U_\theta(\theta) + \sum_{\text{dihedrals}(\psi)} U_\psi(\psi) + \sum_{\text{pairs}(i,j)} U_p(R_{ij})$$

Kirkwood *J Chem Phys* (**1935**); Noid et al. *J Chem Phys* (**2008**)

Force-matching

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Geometric interpretation



Many-body mean force

$$F^0(\mathbf{R}) = -\nabla U^0(\mathbf{R})$$

Force field basis representation

$$\mathbf{F}_I(\mathbf{R}) = \sum_{\zeta} \sum_{d} \phi_{\zeta d} \mathcal{G}_{I;\zeta d}(\mathbf{R})$$

interaction type

basis function index

parameter

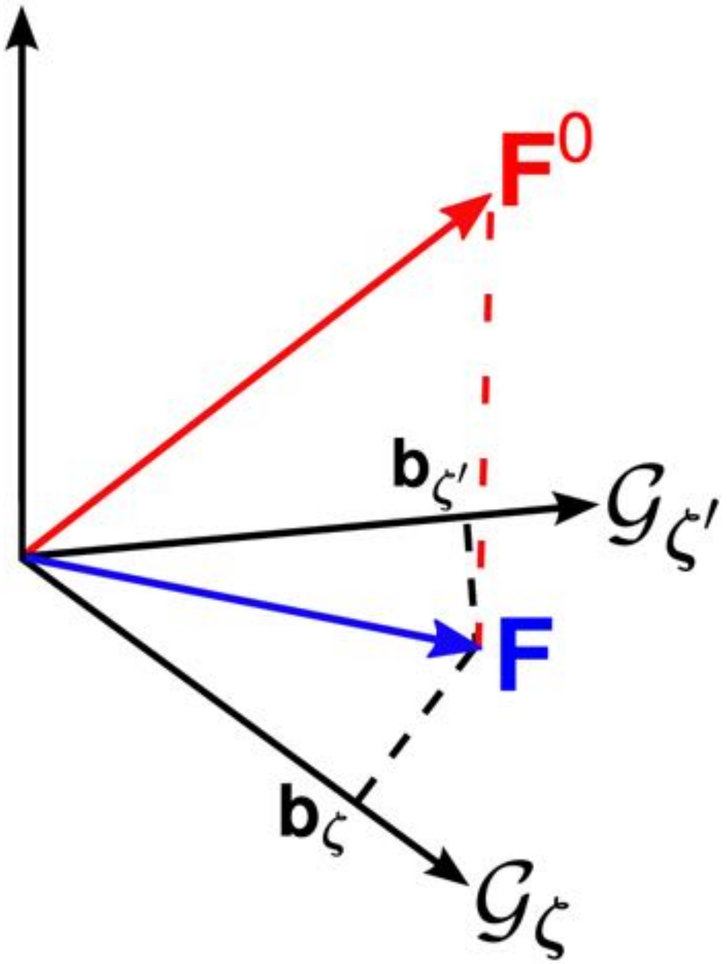
basis function

Izvekov, Voth *J Chem Phys* (**2005**); Noid et al. *J Chem Phys* (**2008**); Rudzinski, Noid *J Phys Chem B* (**2012**)

Force-matching

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Geometric interpretation



Many-body mean force

$$F^0(\mathbf{R}) = -\nabla U^0(\mathbf{R})$$

Variational Functional

$$\chi^2[U] = \frac{1}{3N}\left\langle \sum_I |\mathbf{f}_I(\mathbf{r}) - \mathbf{F}_I(\mathbf{M}(\mathbf{r})|U)|^2 \right\rangle$$
$$= \chi^2[U^0] + ||\mathbf{F}^0 - \mathbf{F}[U]||^2,$$

Izvekov, Voth *J Chem Phys* (**2005**); Noid et al. *J Chem Phys* (**2008**); Rudzinski, Noid *J Phys Chem B* (**2012**)

Relative Entropy Method

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Variational Functional

$$S_{\mathrm{rel}}[U] = k_B \int d\mathbf{r}\, p_r(\mathbf{r}) \ln\left[\frac{p_r(\mathbf{r})}{P_r(\mathbf{r}|U)}\right]$$

$$S_{\mathrm{rel}}[U] = k_B \int d\mathbf{r}\; p_r(\mathbf{r}) \ln\left[\frac{V^n}{V^N}\frac{p_r(\mathbf{r})}{P_R(\mathbf{M}(\mathbf{r})|U)}\right] + S_{\mathrm{map}}$$

Entropy lost during coarse-graining

$$S_{\mathrm{rel}}[U] = k_B \int d\mathbf{R}\, p_R(\mathbf{R}) \ln\left[\frac{p_R(\mathbf{R})}{P_R(\mathbf{R}|U)}\right]$$

Minimization procedure

$$\frac{\delta}{\delta U_\zeta(x)} S_{\mathrm{rel}}[U] = \Delta P_\zeta(x|U)/T$$

$$\Delta P_\zeta(x|U) = p_\zeta(x) - P_\zeta(x|U)$$

Shell *J. Chem. Phys.* **(2008)**; Rudzinski and Noid *J. Chem. Phys.* **(2011)**

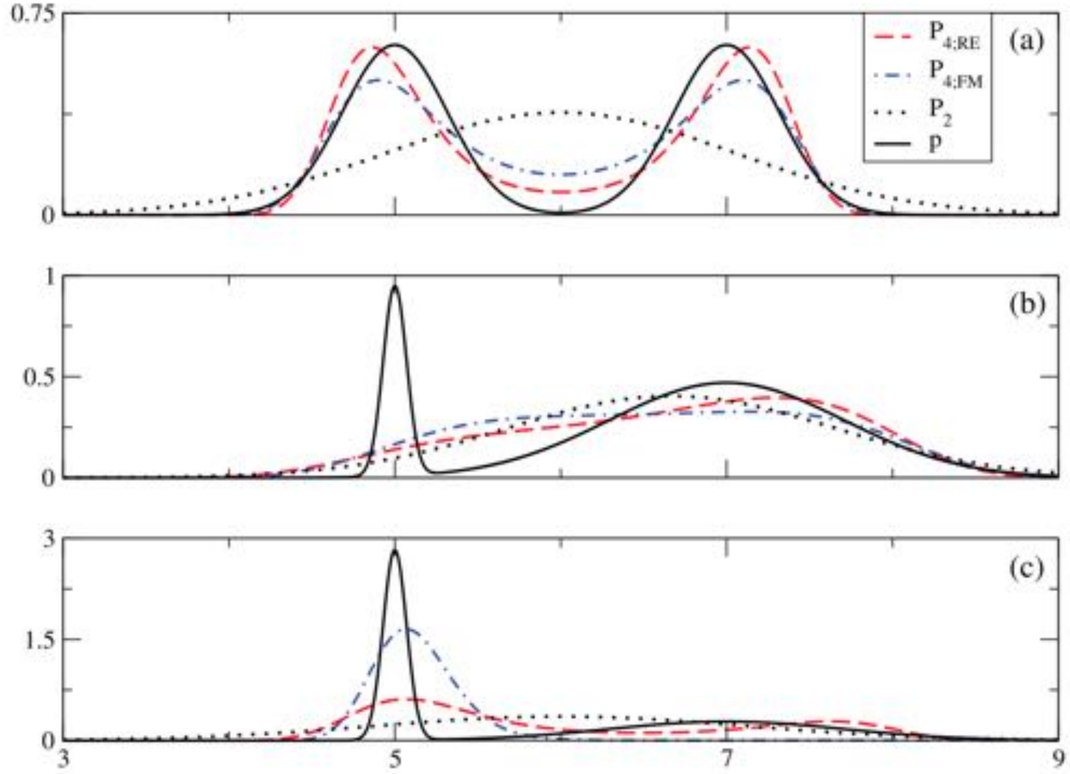
Relative Entropy Method

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

$$\Phi(\mathbf{R}|U) = \ln\left[\frac{p_R(\mathbf{R})}{P_R(\mathbf{R}|U)}\right]$$ **Information Functional**

$$S_{\mathrm{rel}}[U] = k_B \int d\mathbf{R} p_R(\mathbf{R}) \Phi(\mathbf{R}|U)$$

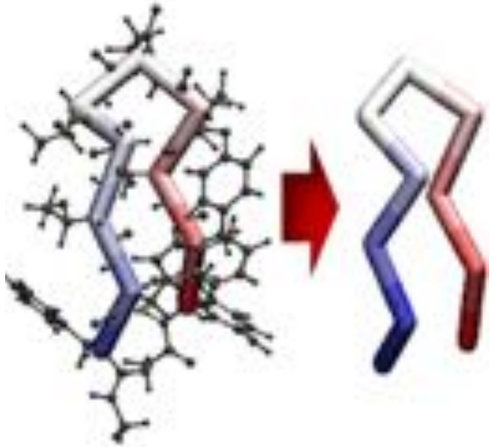
$$\chi^2[U] = \frac{(k_B T)^2}{3N} \int d\mathbf{R} p_R(\mathbf{R}) \, |\nabla \Phi(\mathbf{R}|U)|^2 + \chi^2[U^0]$$

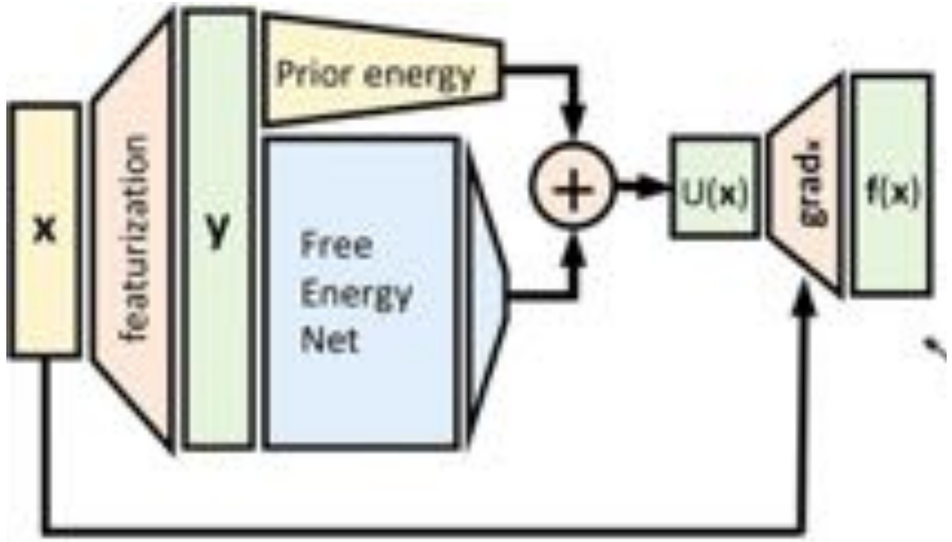


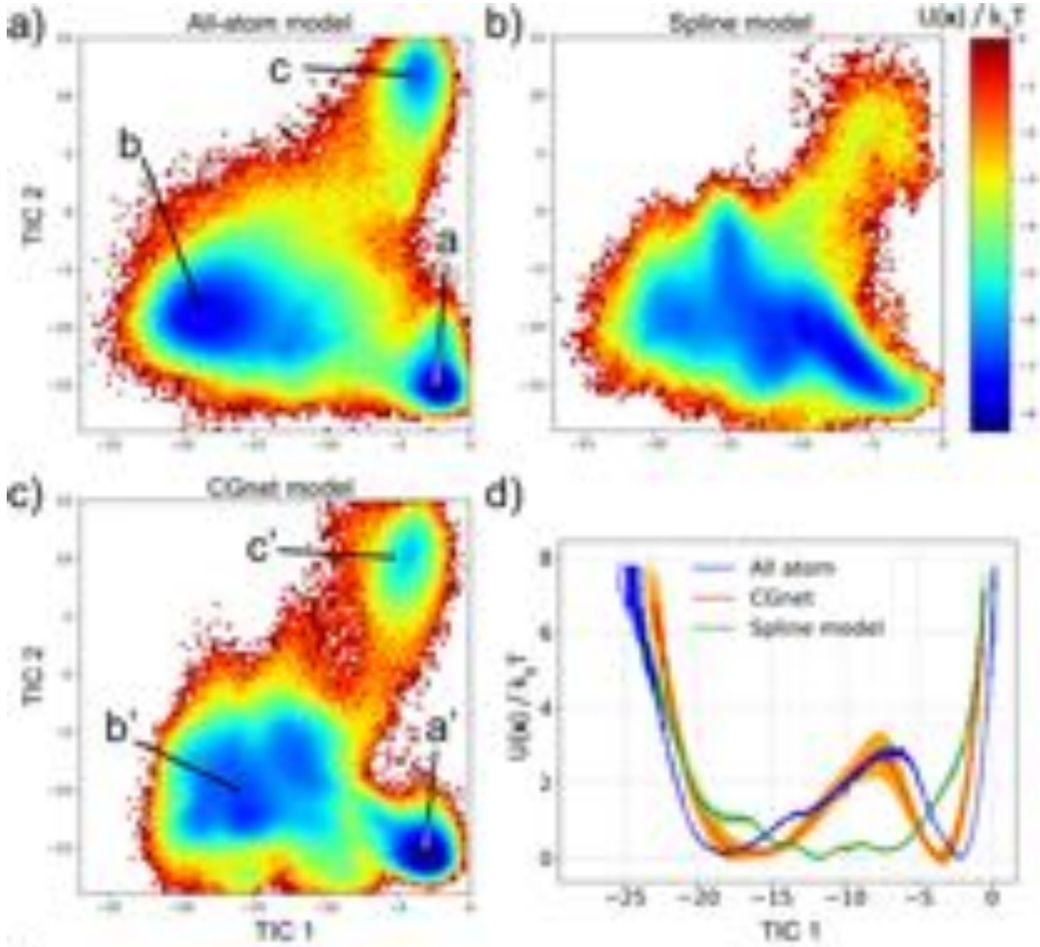
Shell *J. Chem. Phys.* **(2008)**; Rudzinski and Noid *J. Chem. Phys.* **(2011)**

CGnet: a neural network based force matching

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research







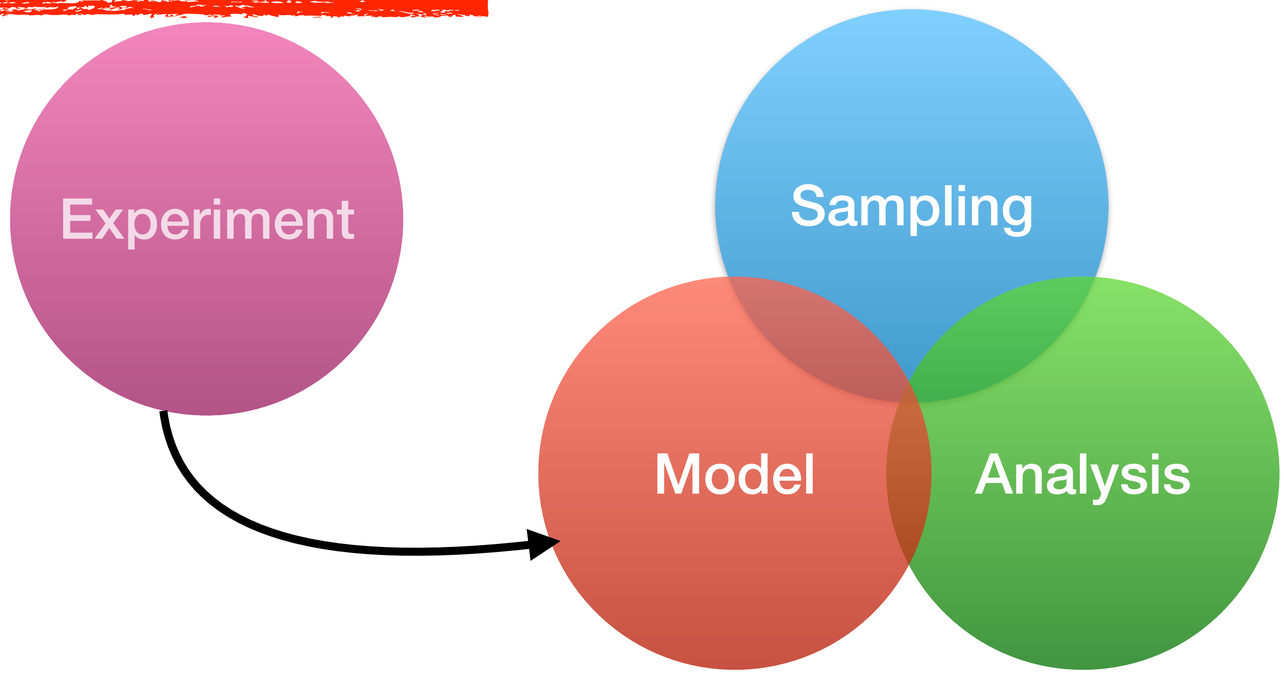
Wang et. al. *ACS Central Science* **(2019)** “Machine Learning of Coarse-Grained Molecular Dynamics Force Fields”

Fundamental challenges for soft matter modeling

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

bayesian reweighting
force field refinement

enhanced sampling
high throughput studies



force field development
coarse-grained modeling

kinetic modeling
(dimensionality reduction / clustering)

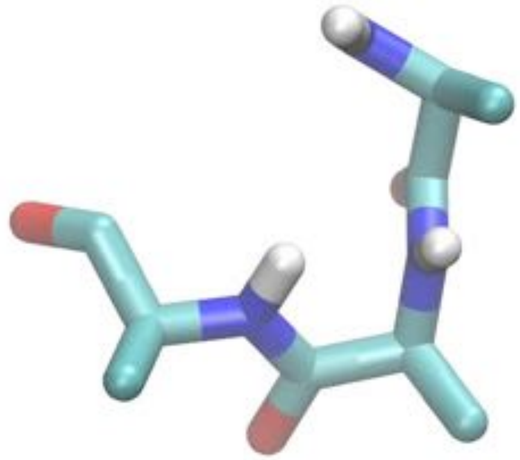
Characterizing the hierarchy of kinetic processes from simulation trajectories

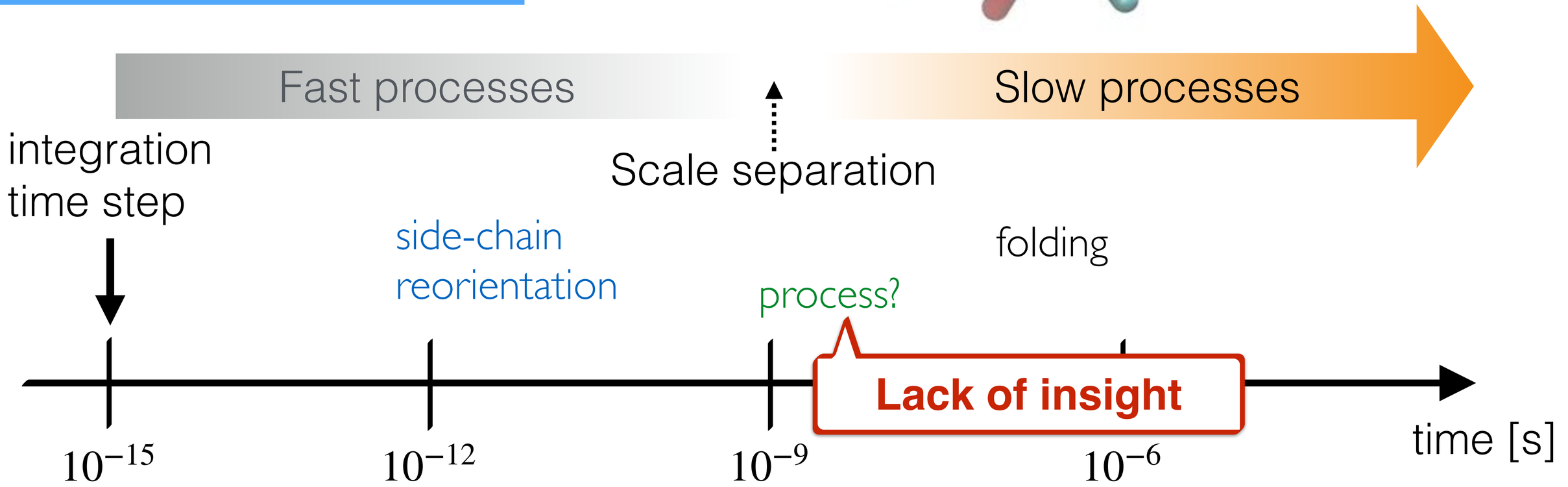
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Markov state models

Eigenvalue decomposition of the Master equation

Work from Pande, Noé, Chodera, Schütte, Dill, Swope, Bowman, Keller, and others





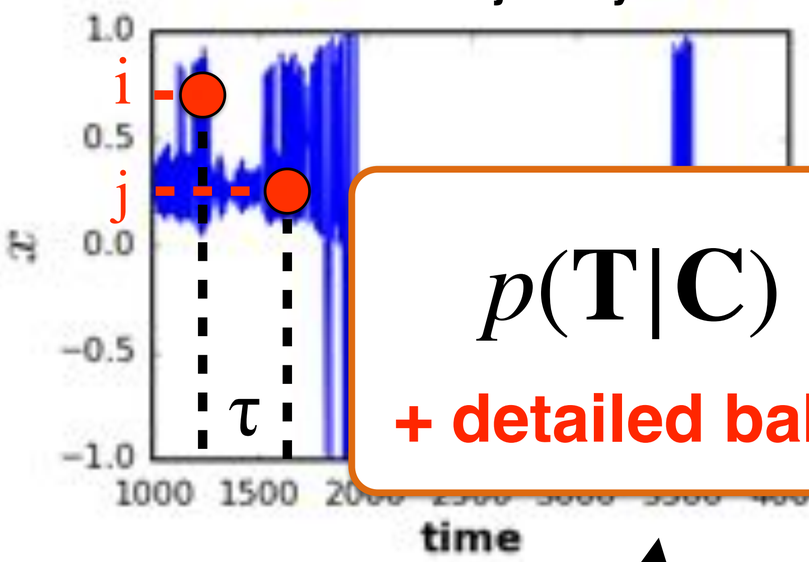
Markov State M

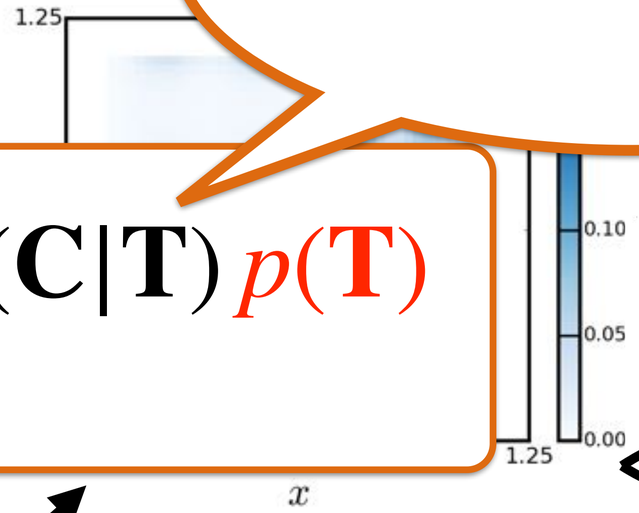
$$p(\mathbf{C}|\mathbf{T}) = \prod_{ij} T_{ij}^{C_{ij}}$$

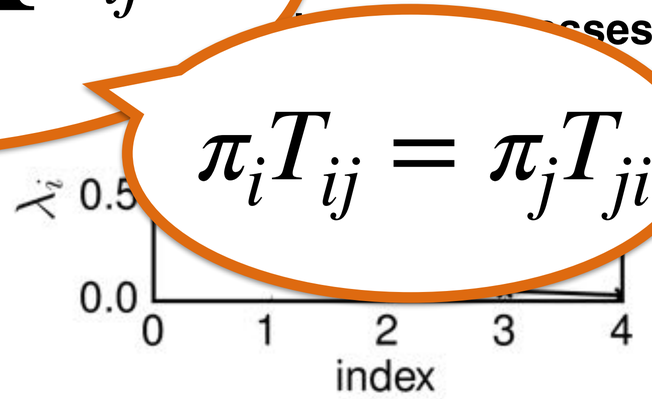
$$\pi_i T_{ij} = \pi_j T_{ji}$$

$$p(\mathbf{T}|\mathbf{C}) \propto p(\mathbf{C}|\mathbf{T})\, p(\mathbf{T})$$

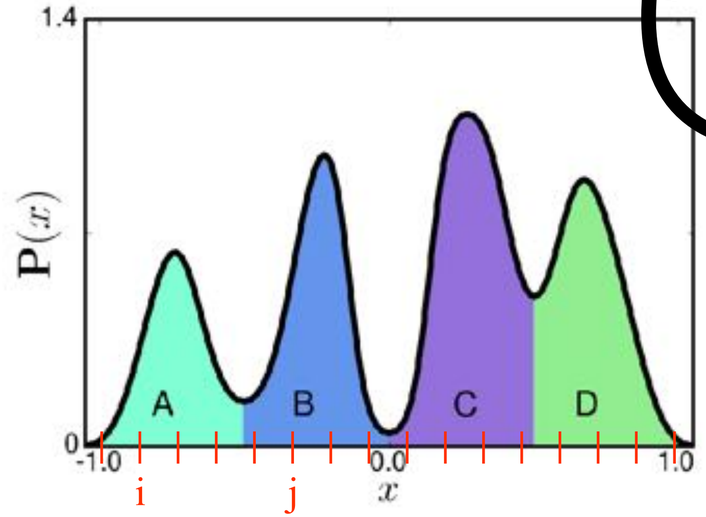
+ detailed balance







diagonalize

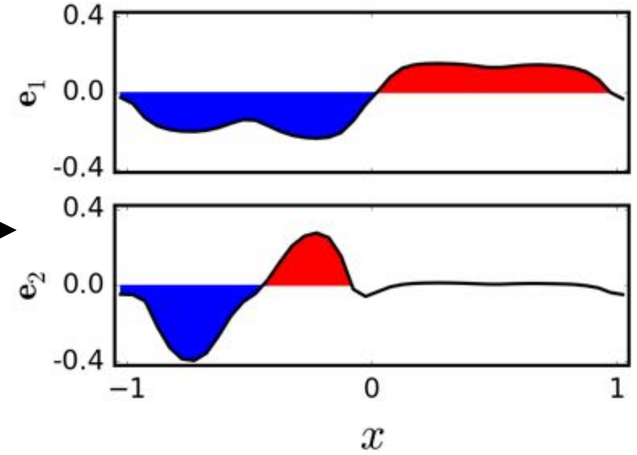


microstates

maximum likelihood optimization

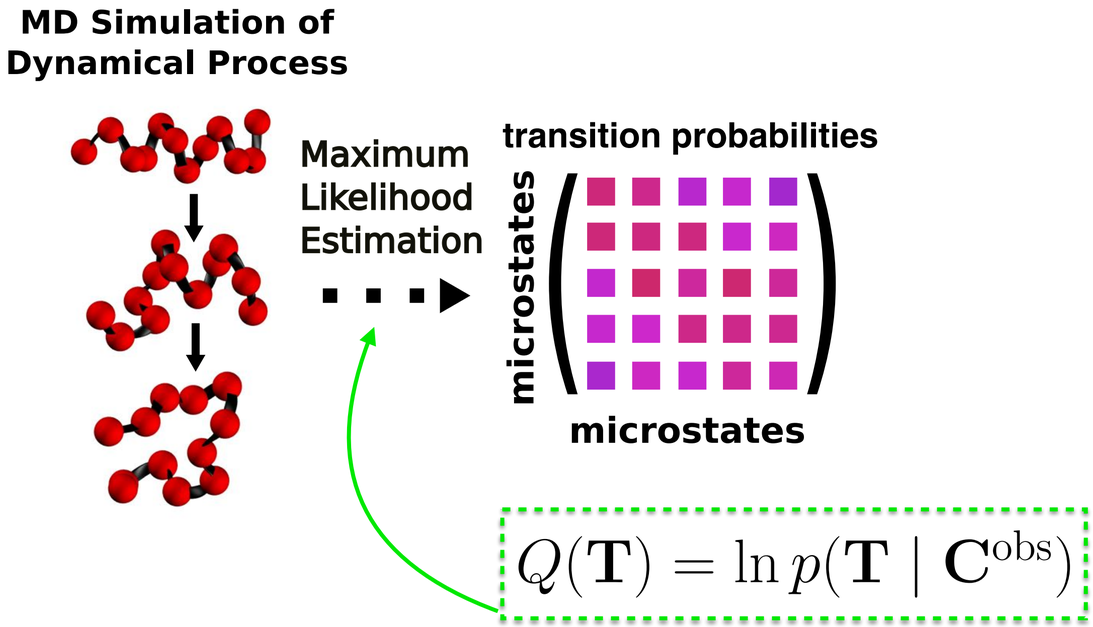
MSMs link microtrajectories with long timescale kinetic processes

eigenvectors (probability flux for kinetic processes)



Bowman *et al.* "An introduction to Markov state models…" *Springer* (2014)

Biased MSMs for dynamical reweighting



Bayesian approach

model data

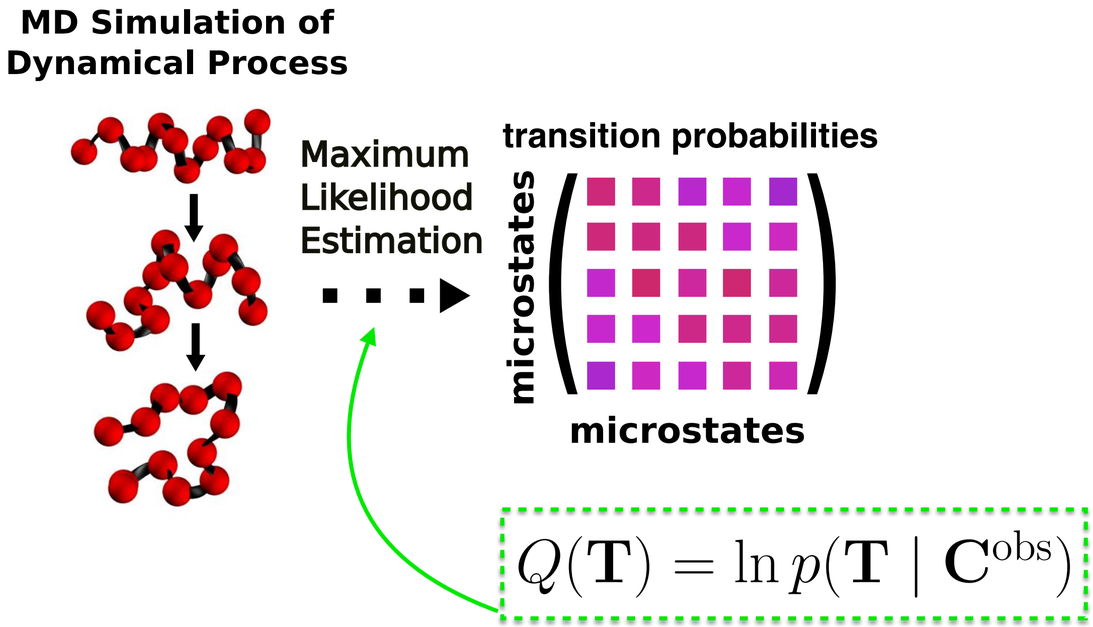
$$p(\mathbf{T} \mid \mathbf{C}^{\mathrm{obs}}) \propto p(\mathbf{C}^{\mathrm{obs}} \mid \mathbf{T})\, p(\mathbf{T})$$

Simulation data

Bowman *et al.* "An introduction to Markov state models…" *Springer* (2014)
Rudzinski, Kremer, Bereau *J Chem Phys* (2016)

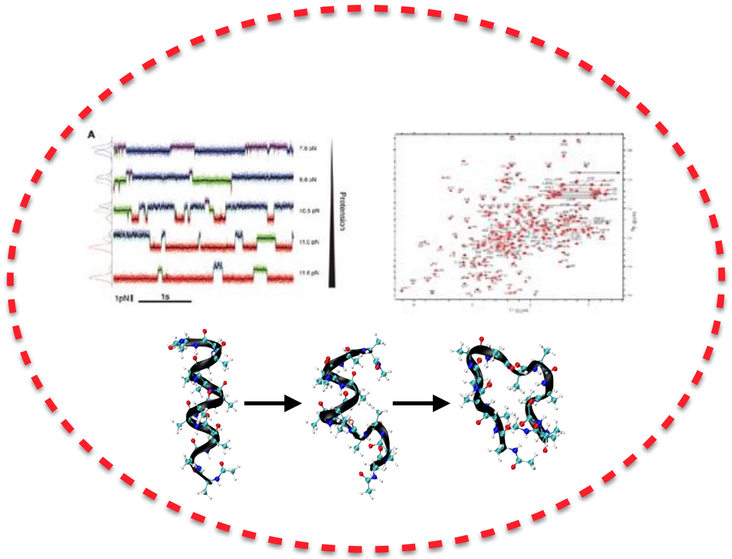
Biased MSMs for dynamical reweighting

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research



Bayesian approach

Best combination $p(\mathbf{T} \mid \mathbf{C}^{\mathrm{obs}})$ $\propto$ Simulation data $p(\mathbf{C}^{\mathrm{obs}} \mid \mathbf{T})$ Exp. data $p(\mathbf{T})$

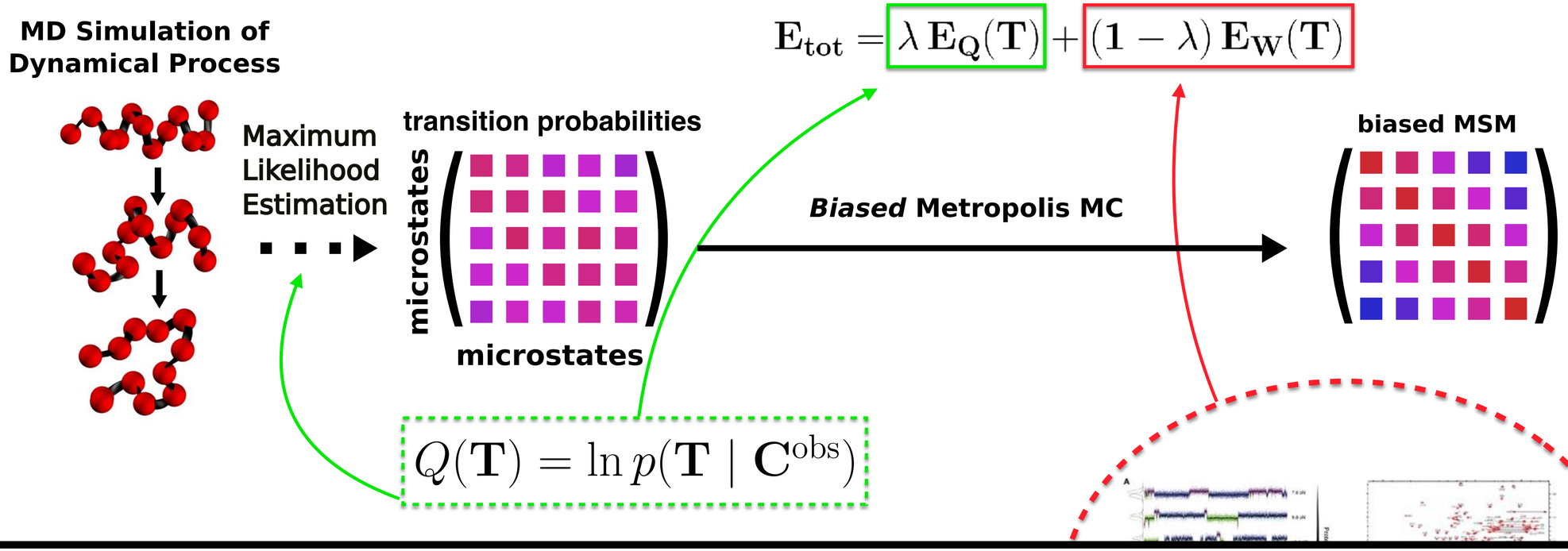


Reference Data

Rudzinski, Kremer, Bereau *J Chem Phys* (2016)

Biased MSMs for dynamical reweighting

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research



See Also:

Olsson *et al. PNAS* "Combining experimental and simulation data of molecular processes via augmented Markov models" (2017)

Dixit and Dill, *J. Chem. Theory Comput.* "Caliber Corrected Markov Modeling (C2M2): Correcting Equilibrium Markov Models" (2018)

Rudzinski, Kremer, Bereau *J Chem Phys* (2016)

Transfer Operators

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

probability density

$$p_{t+\tau}(x) = \mathscr{P}_\tau \, p_t(x)$$

Perron-Frobenius operator - “propagator”

$$\mathscr{K}_\tau \, f_t(x) = \int_{\mathbb{X}} dy \; p_\tau(y \,|\, x) f_t(y) = \mathbb{E}[f_t(\mathbf{X}_{t+\tau}) \mid \mathbf{X}_t = x]$$

Koopman operator

Transfer Operators

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Equilibrium

$$\mathscr{P}_\tau \pi = \pi$$ ← **Stationary distribution**

$$\pi(x)\, p_\tau(y \mid x) = \pi(y)\, p_\tau(x \mid y)$$ **Detailed balance**

Spectral properties

$$\mathscr{A}_\tau \varphi_l = \lambda_l(\tau)\, \varphi_l$$

$$1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots$$ **Eigenvalues**

$$\lambda_l(\tau) = \exp(-\kappa_l \tau)$$ **Characteristic timescales**

Klus *et al. Int. J. Nonlinear Sci.* **(2018)** “Data-Driven Model Reduction and Transfer Operator Approximation”

Transfer Operators

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Equilibrium

$$\mathscr{P}_\tau \pi = \pi$$ ← **Stationary distribution**

$$\pi(x)\, p_\tau(y \mid x) = \pi(y)\, p_\tau(x \mid y)$$ **Detailed balance**

Spectral properties

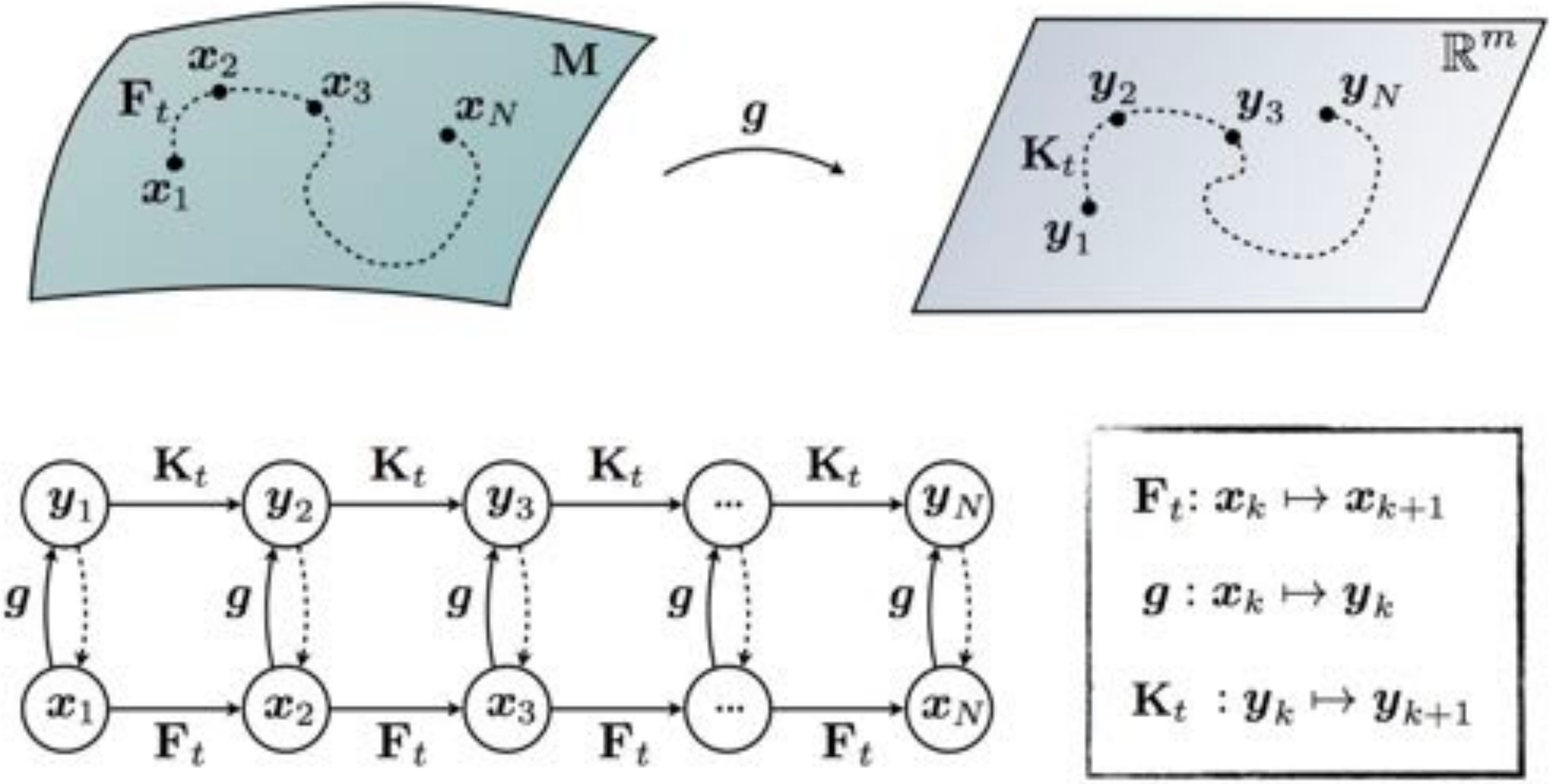
Separation of timescales?!

$$\mathscr{K}_\tau f = \sum_{l=1}^{\infty} \lambda_l(\tau) \langle f, \varphi_l \rangle_\pi \varphi_l \qquad f = \sum_{l=1}^{\infty} \langle f, \varphi_l \rangle_\pi \varphi_l$$

Eigenvector representation

Klus *et al. Int. J. Nonlinear Sci.* **(2018)** "Data-Driven Model Reduction and Transfer Operator Approximation"

Koopman Theory



S. Brunton *et al. PLoS ONE* **(2016)** "Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control"

Data driven approaches to transfer operators

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Variational Approach to Conformational dynamics (VAC)

$$\Psi_X = [\psi(x_1)\ \psi(x_2)\ \cdots\ \psi(x_m)] \text{ and } \Psi_Y = [\psi(y_1)\ \psi(y_2)\ \cdots\ \psi(y_m)]$$

Basis function representation

$$C_0 = \frac{1}{m-1}\sum_{k=1}^{m} \psi(x_k)\,\psi(x_k)^\top = \frac{1}{m-1}\Psi_X\,\Psi_X^\top$$

$$C_\tau = \frac{1}{m-1}\sum_{k=1}^{m} \psi(x_k)\,\psi(y_k)^\top = \frac{1}{m-1}\Psi_X\,\Psi_Y^\top$$

Covariance

$$M_{\mathrm{VAC}} = C_0^+\,C_\tau$$

$$M_{\mathrm{VAC}}\,\xi_l = \lambda_l\,\xi_l$$

$$\varphi_l(x) = \xi_l^*\,\psi(x)$$

Eigenvalue problem

Klus *et al. Int. J. Nonlinear Sci.* **(2018)** "Data-Driven Model Reduction and Transfer Operator Approximation"

Data driven approaches to transfer operators

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Variational Approach to Conformational dynamics (VAC)

$$\sum_{l=1}^{M} \lambda_l = \sup \sum_{l=1}^{M} \langle \mathscr{K}_\tau v_l, v_l \rangle_\pi$$

Rayleigh trace is maximized by first M eigenvectors of the Koopman operator

$$\langle v_l, v_{l'} \rangle_\pi = \delta_{ll'}$$

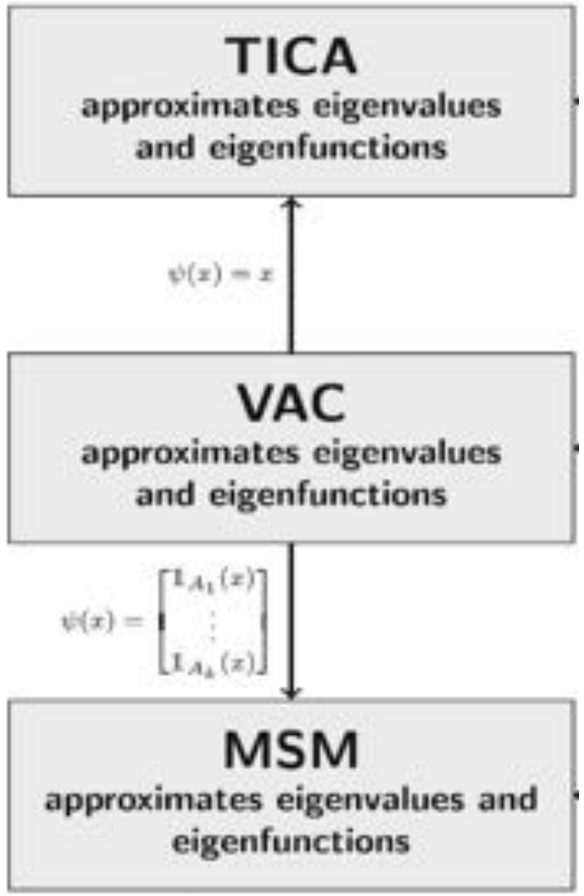
$$M_{\mathrm{VAC}}\, \xi_l = \lambda_l\, \xi_l$$

Solution to VAC eigenvalue problem maximizes Rayleigh trace

Klus *et al. Int. J. Nonlinear Sci.* **(2018)** "Data-Driven Model Reduction and Transfer Operator Approximation"

Data driven approaches to transfer operators

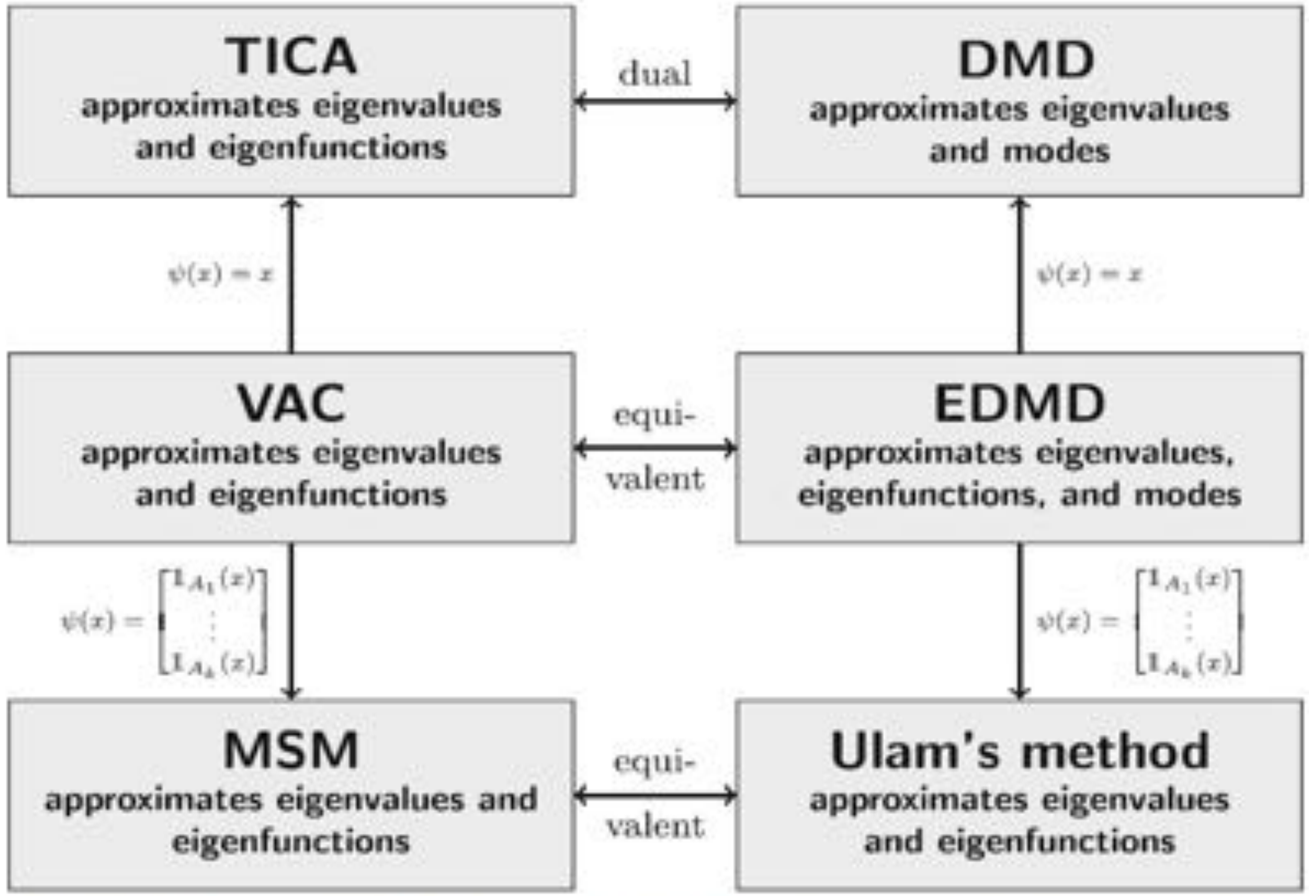
Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research



Klus *et al. Int. J. Nonlinear Sci.* **(2018)** "Data-Driven Model Reduction and Transfer Operator Approximation"

Data driven approaches to transfer operators

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research



Klus *et al. Int. J. Nonlinear Sci.* **(2018)** "Data-Driven Model Reduction and Transfer Operator Approximation"

Variational Autoencoders

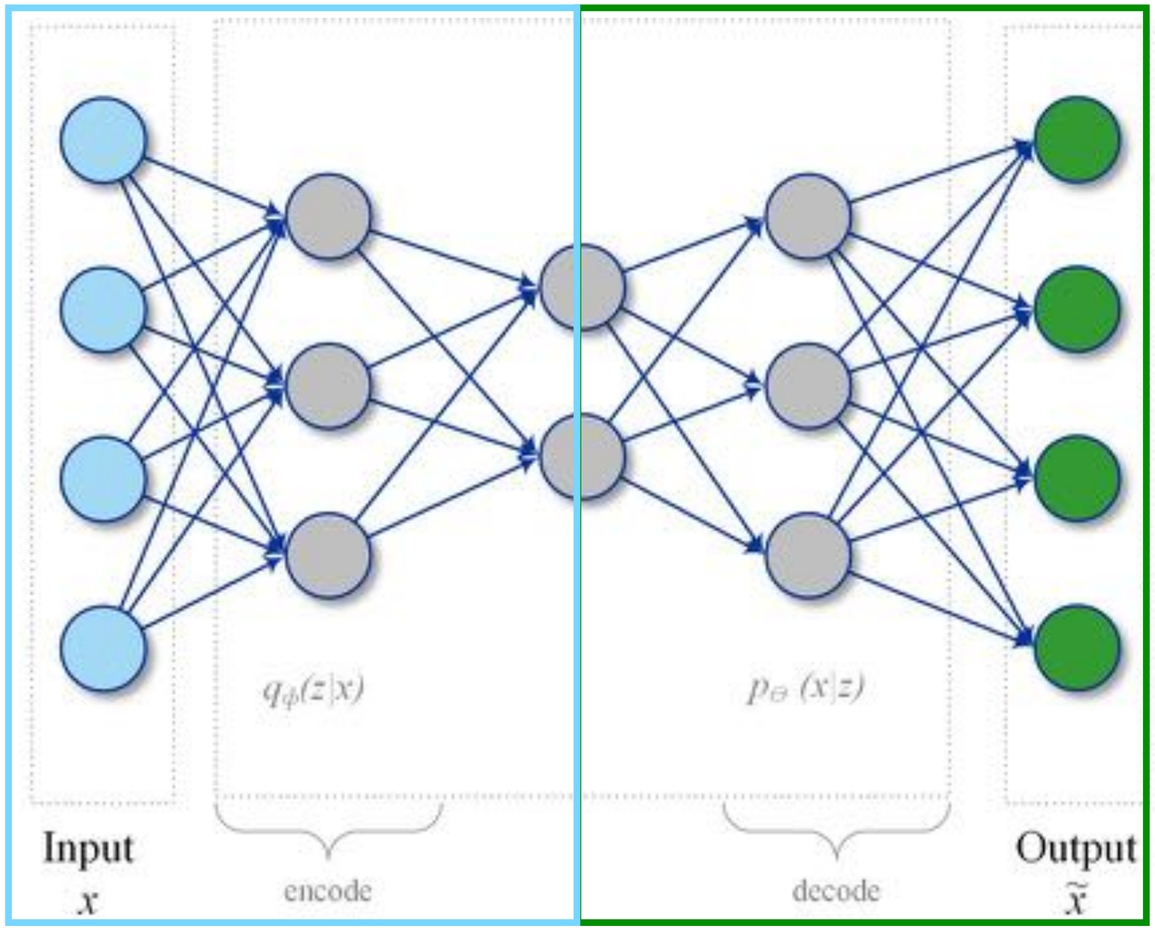
$$p_\theta(z|x) \approx q_\phi(z|x) \qquad p_\theta(x|z)$$

$$p_\theta(z|x) = \frac{p_\theta(x,z)}{p_\theta(x)}$$

$$p_\theta(x) = \int \mathrm{d}z\, p_\theta(x,z)$$

marginal likelihood

intractable



$$p_\theta(x|z) = \frac{p_\theta(x,z)}{p_\theta(z)}$$

prior

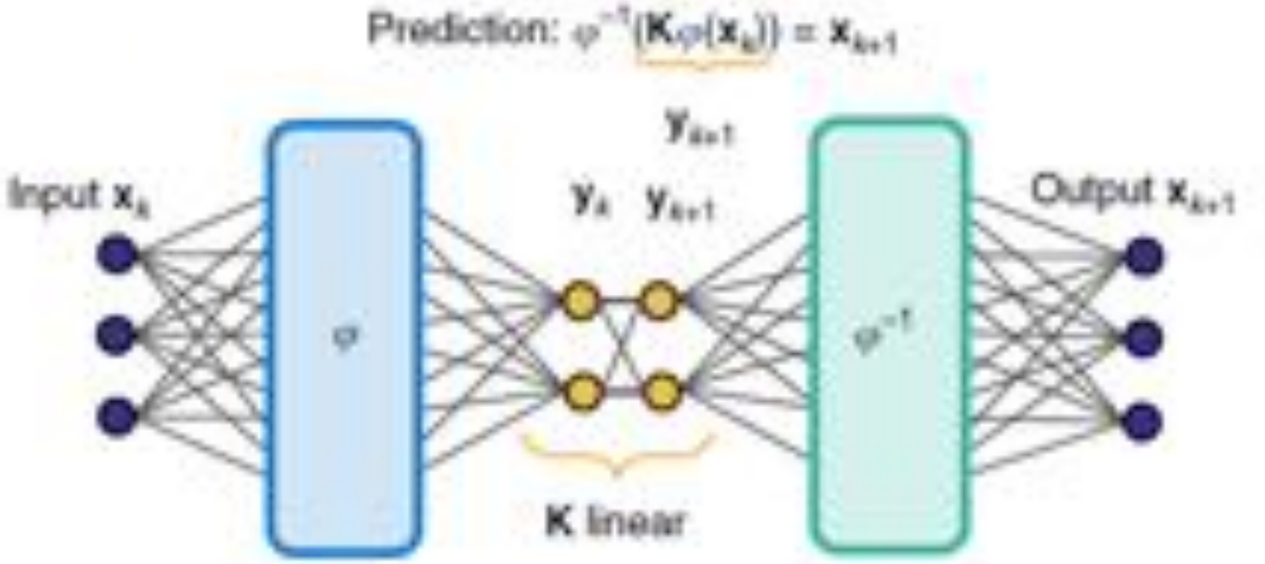
Mohammadi *et al. IEEE Commun. Surv. Tutor* **(2017)** "Deep Learning for IoT Big Data and Streaming Analytics: A Survey"

Kingma, Welling *ArXiv* **(2017)** "An Introduction to Variational Autoencoders"

Data driven approach to Koopman Theory

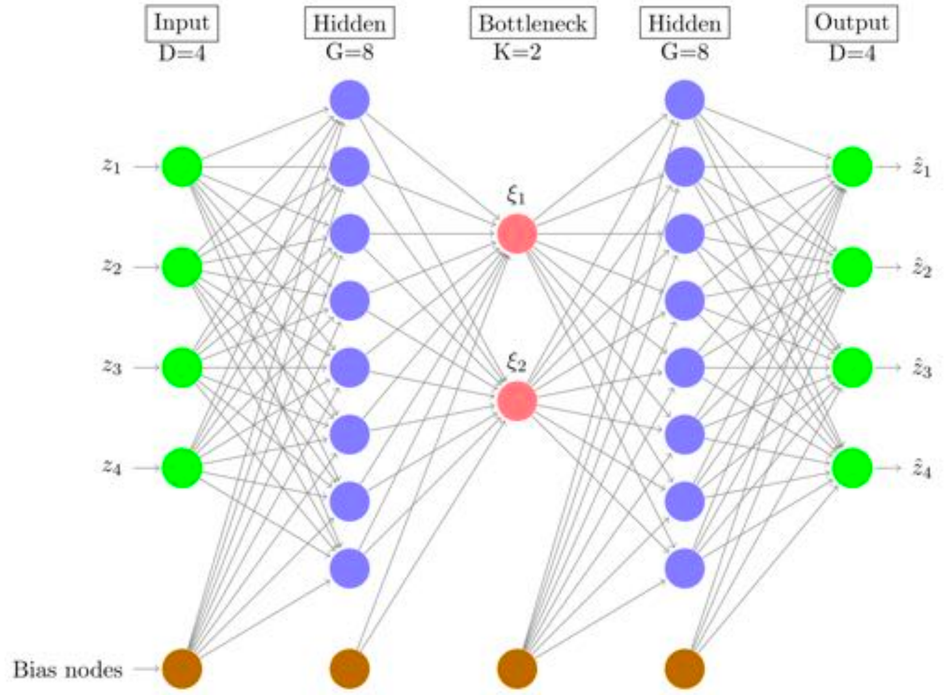
Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Data driven approach to Koopman Theory



Lusch *et al. Nat Comm* **(2018)** "Deep learning for universal linear embeddings of nonlinear dynamics"

Data driven approach to collective variable discovery

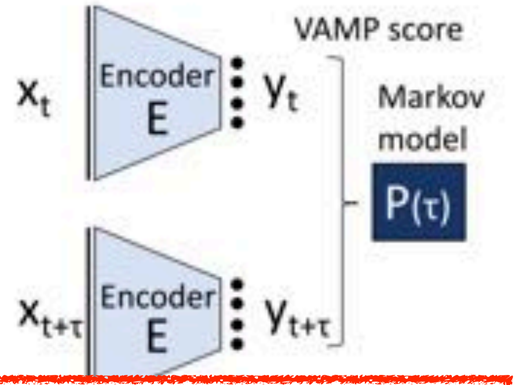


Chen, Ferguson *J Comp Chem* **(2018)** "Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration"

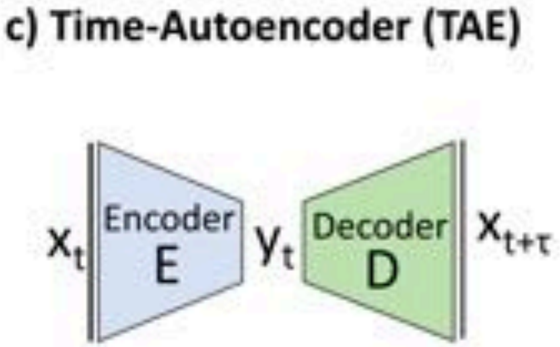
VAEs for MD analysis

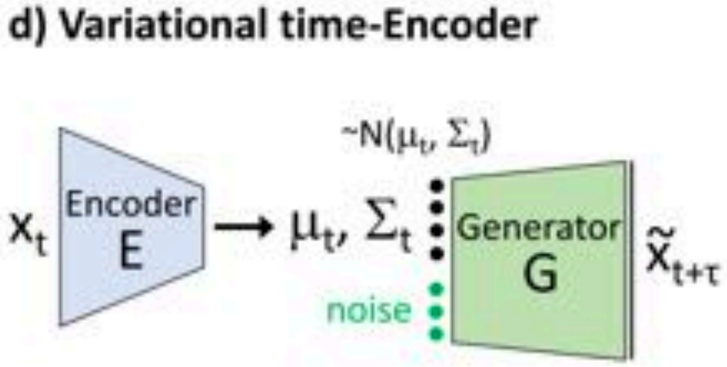
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Find latent space representation that characterises dynamics

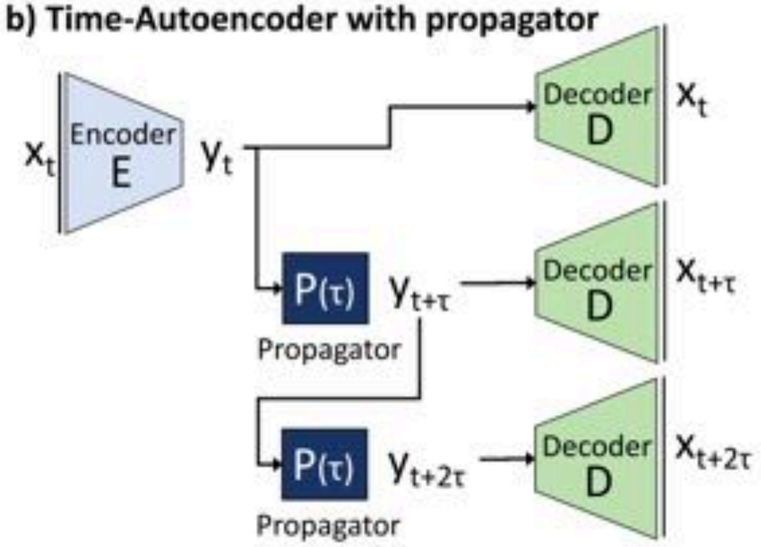


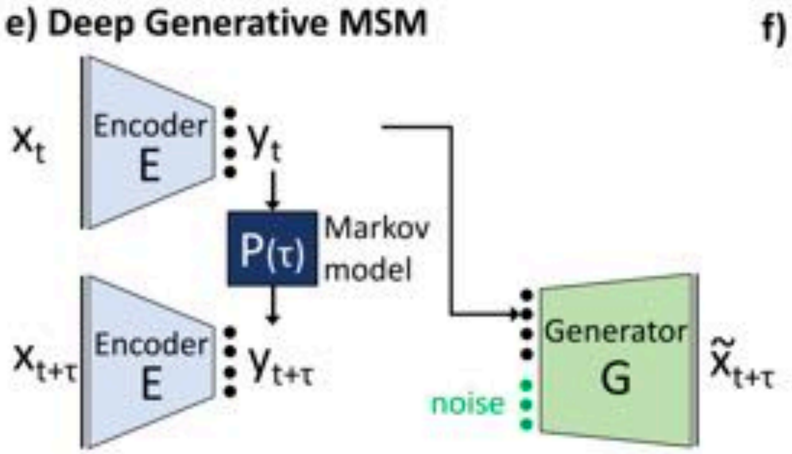
Learn propagator directly

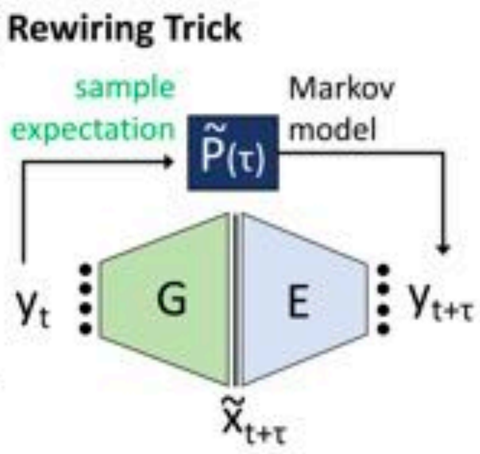




Focus on generation of samples







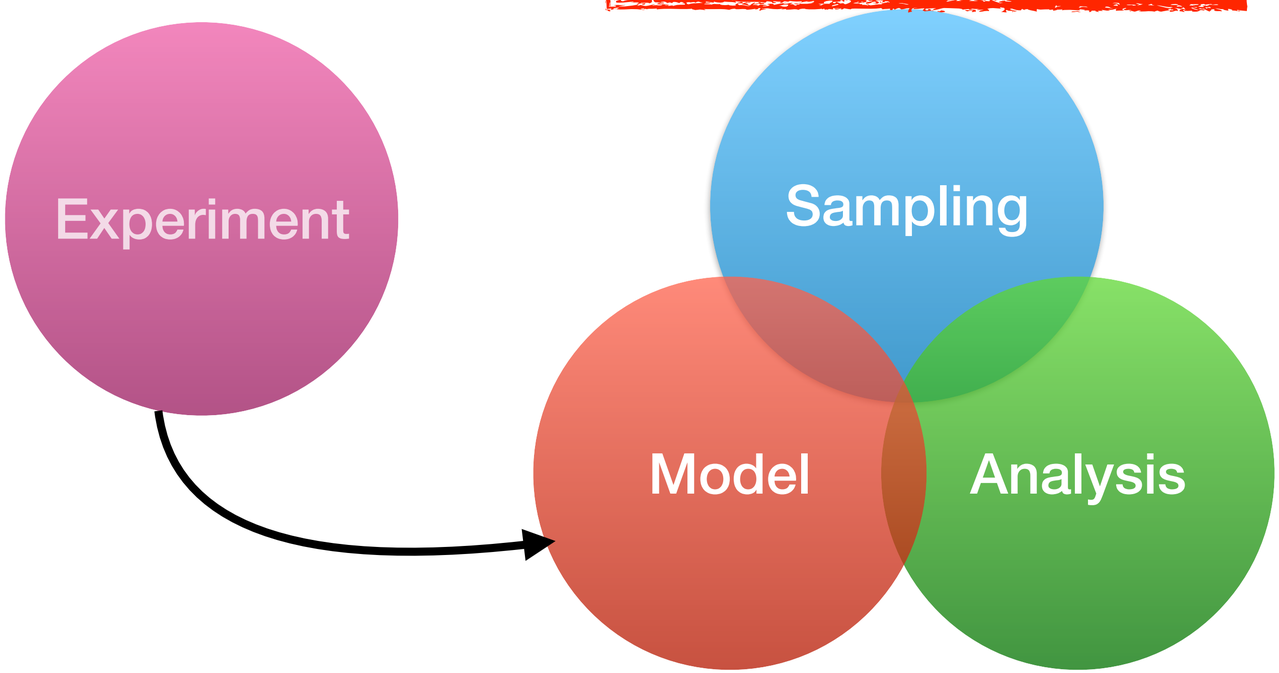
Noe *ArXiv* **(2019)** "Machine Learning for Molecular Dynamics on Long Timescales"

Fundamental challenges for soft matter modeling

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

bayesian reweighting
force field refinement

enhanced sampling
high throughput studies

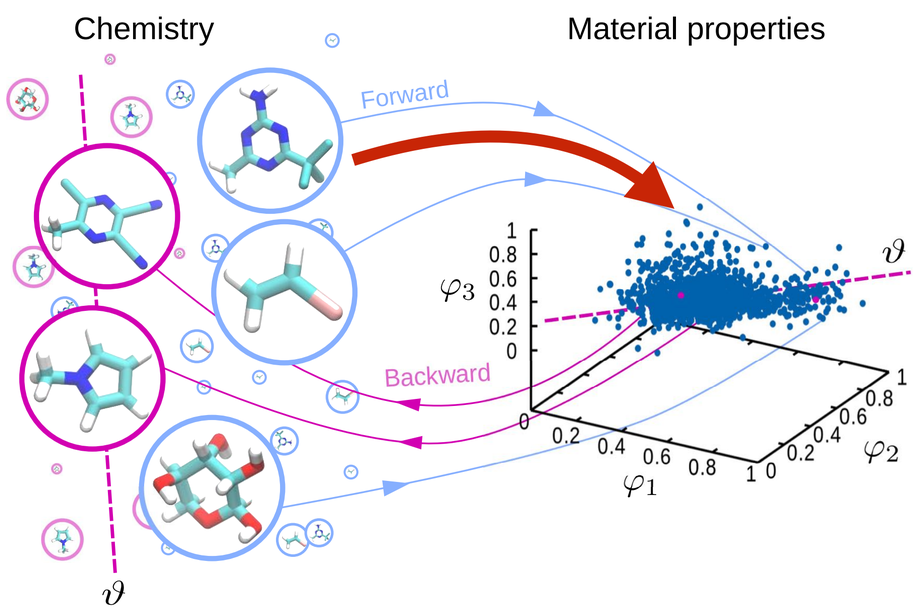


force field development
coarse-grained modeling

kinetic modeling
(dimensionality reduction / clustering)

High-throughput screening

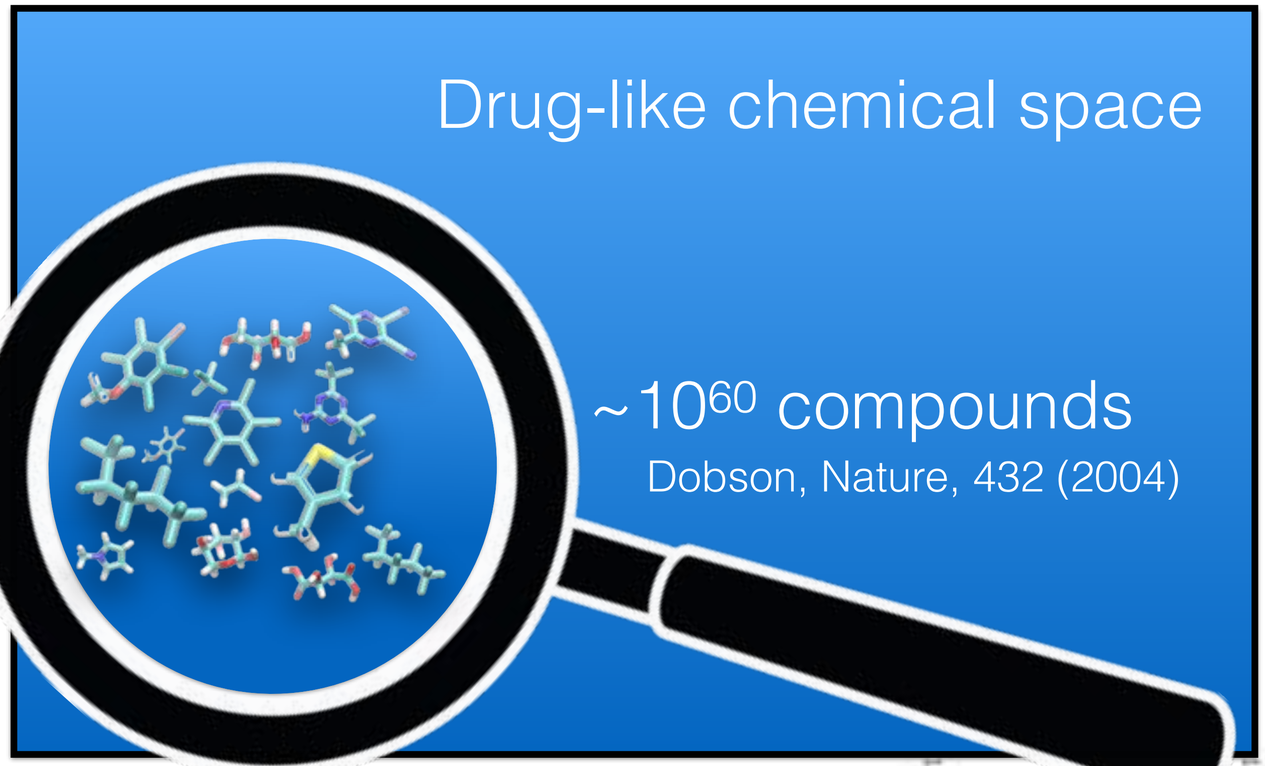
Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

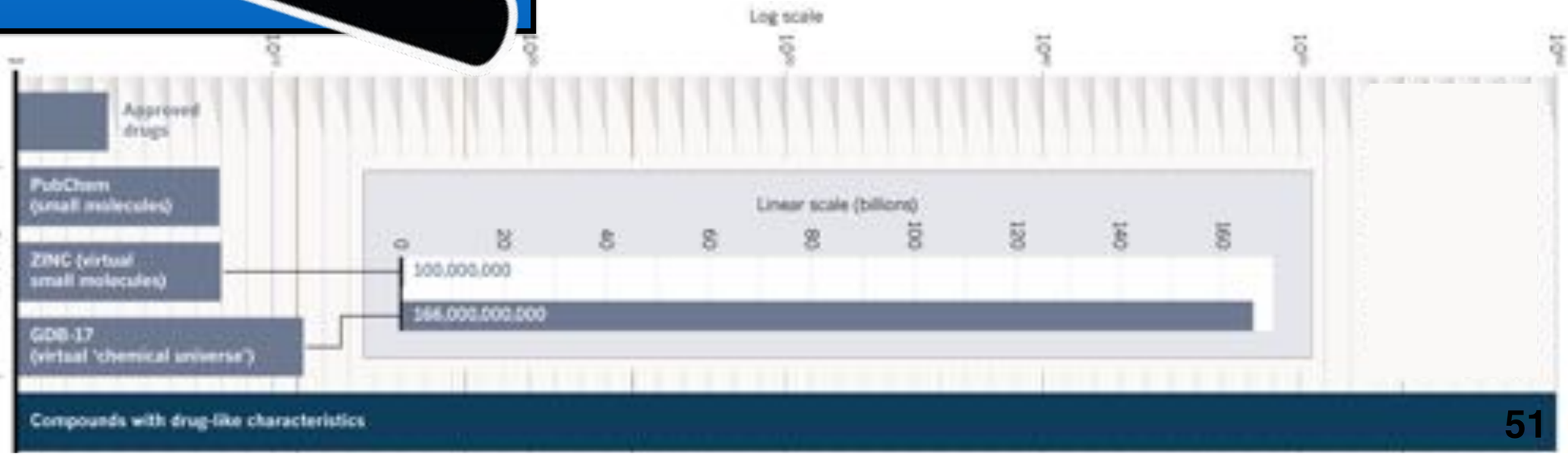


Systematic measurements, Automated sample preparation

Interpolation (chemical) space is large

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research





Mullard,
Nature, 549 (2017)

Drug permeability

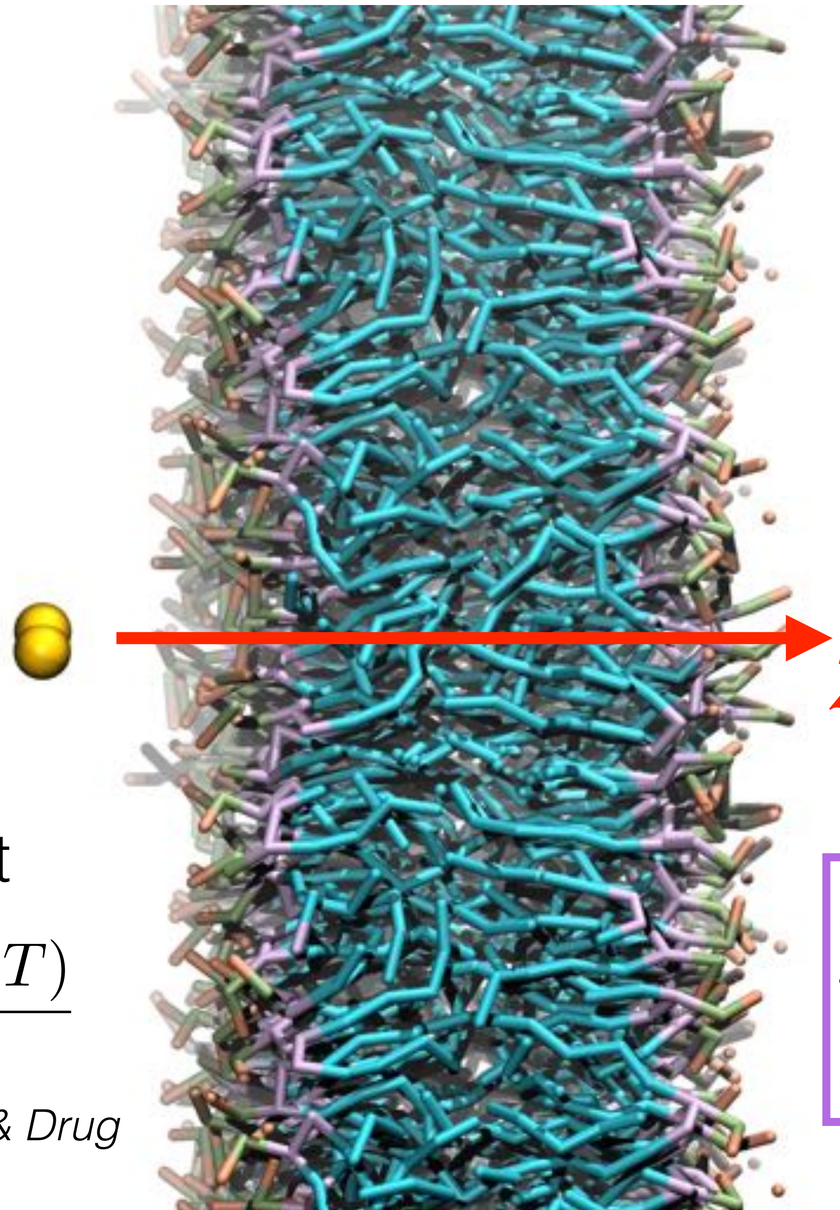
Flux of drug permeation across a lipid membrane

Fokker-Planck; Smoluchoswki

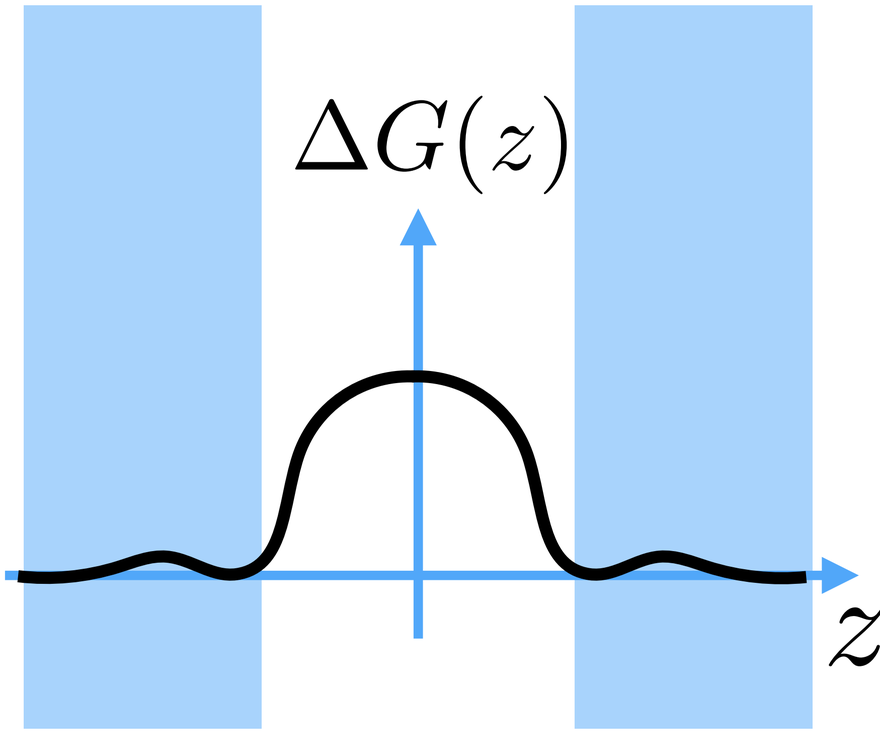
Permeability coefficient

$$P^{-1} = \int \mathrm{d}z \frac{\exp\left(\Delta G(z)/k_\mathrm{B}T\right)}{D_z(z)}$$

Swift & Amaro, *Chem Biol & Drug Design* **81** (2013)



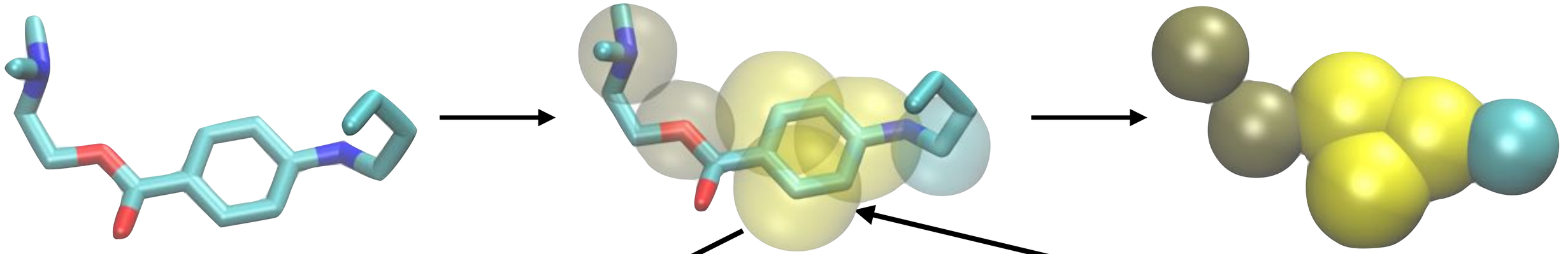
Potential of mean force



Sample the potential of mean force from computer simulations

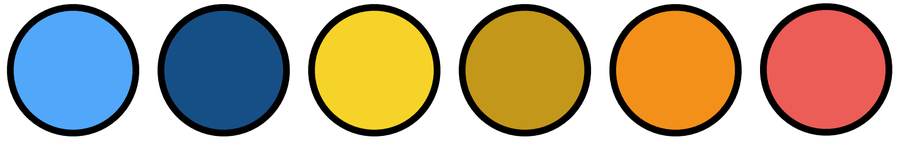
Coarse-graining the thermodynamics of partitioning

Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research



Chemical group:
- net charge
- hydrogen bond
- water/octanol partitioning

cheminformatics & machine learning

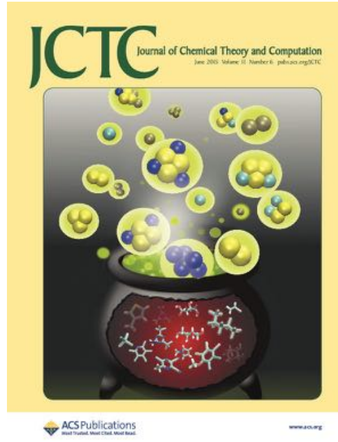


18 bead types: *chemical fragments*

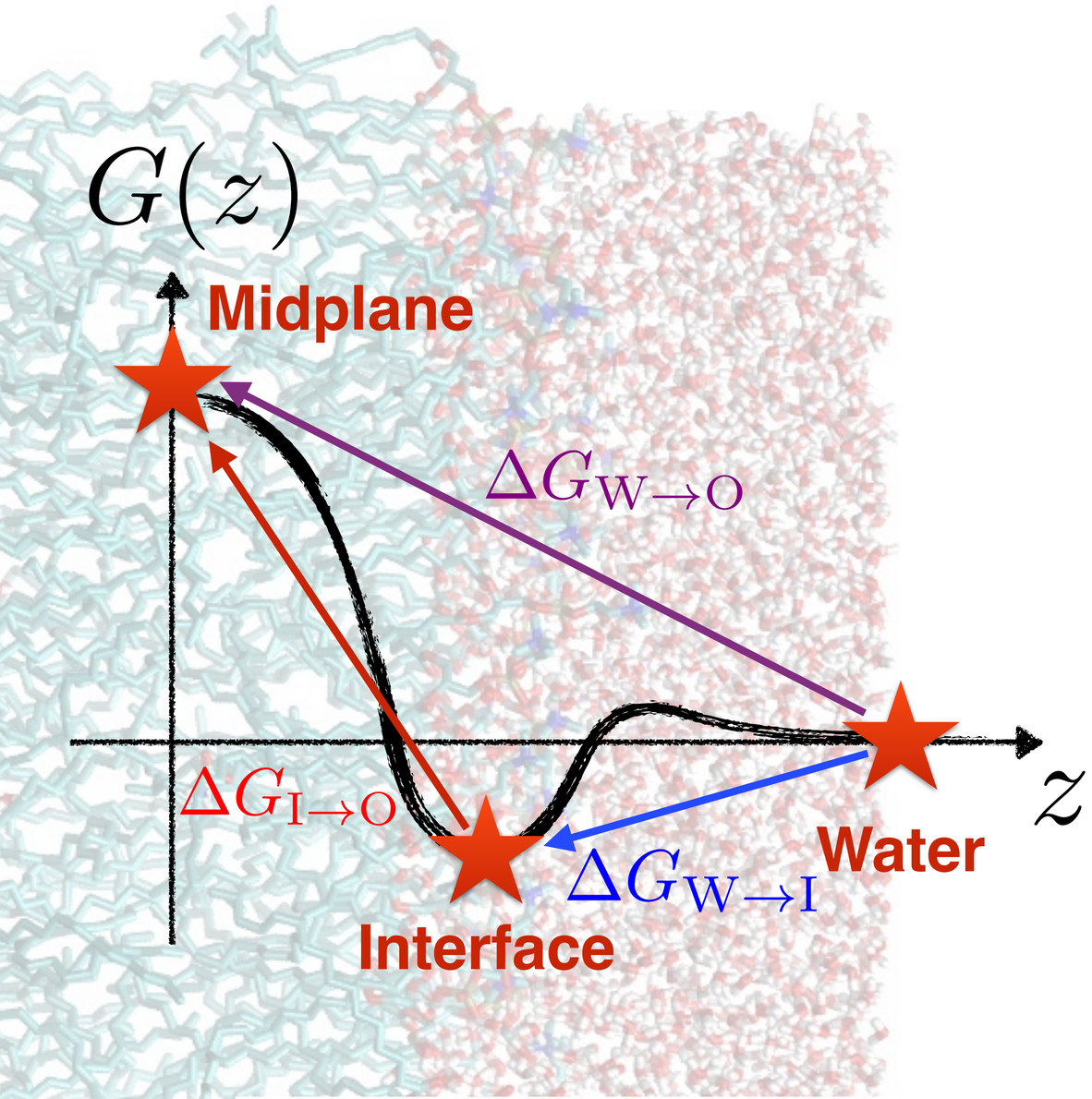
Marrink, Tieleman, Chem Soc Rev **42** (2013)
Periole, Marrink, Biomolecular Simulations (2013)

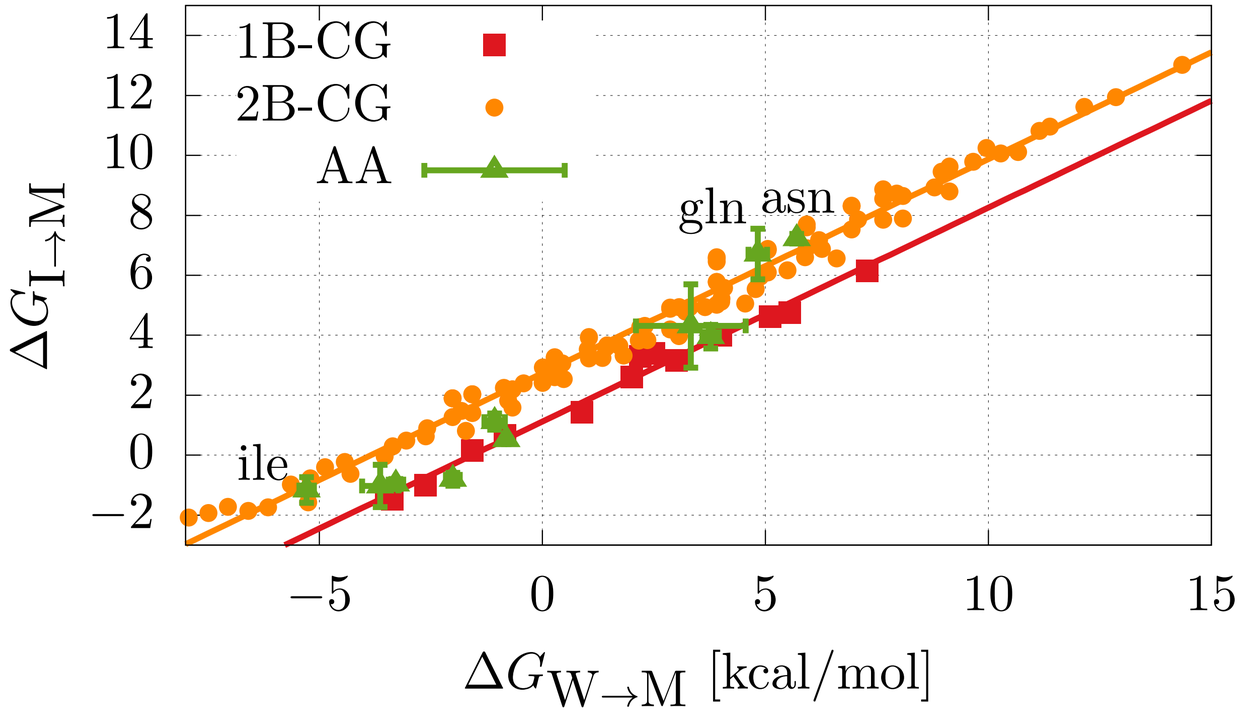
Automated parametrization for small molecules

Bereau & Kremer, *J Chem Theory Comput* **11** (2015)



Identifying simple thermodynamic relations

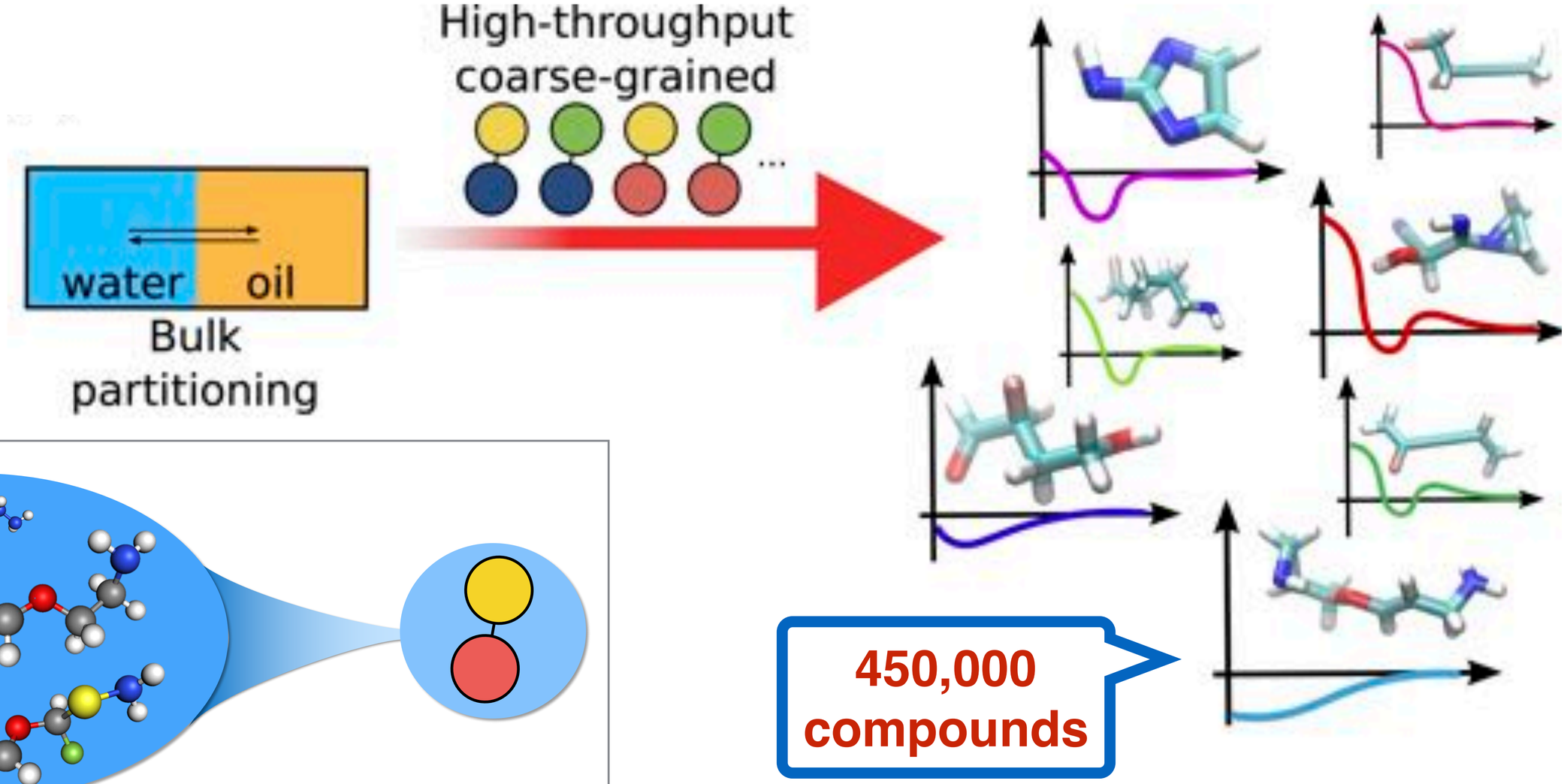




Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

Generating databases of drug-membrane PMFs

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

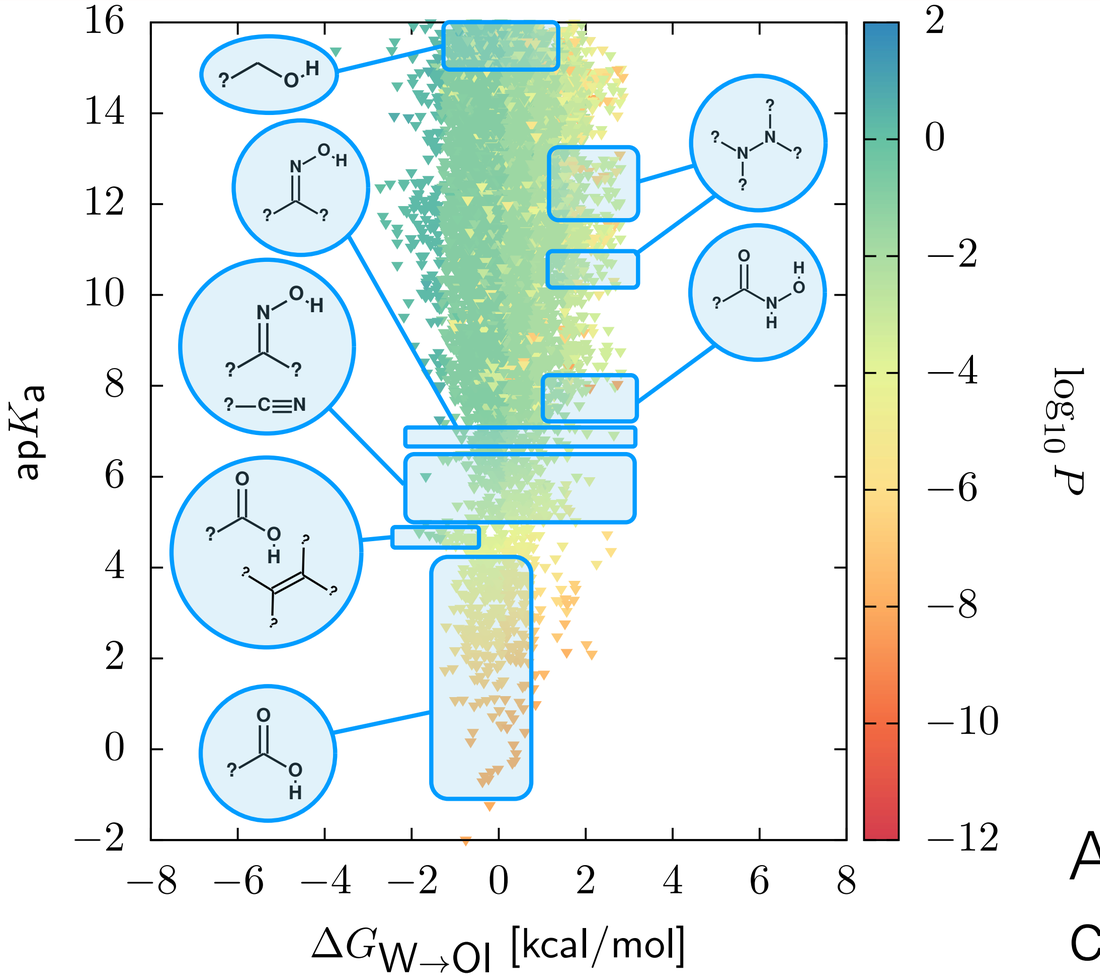


Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

Building a surface of permeabilities

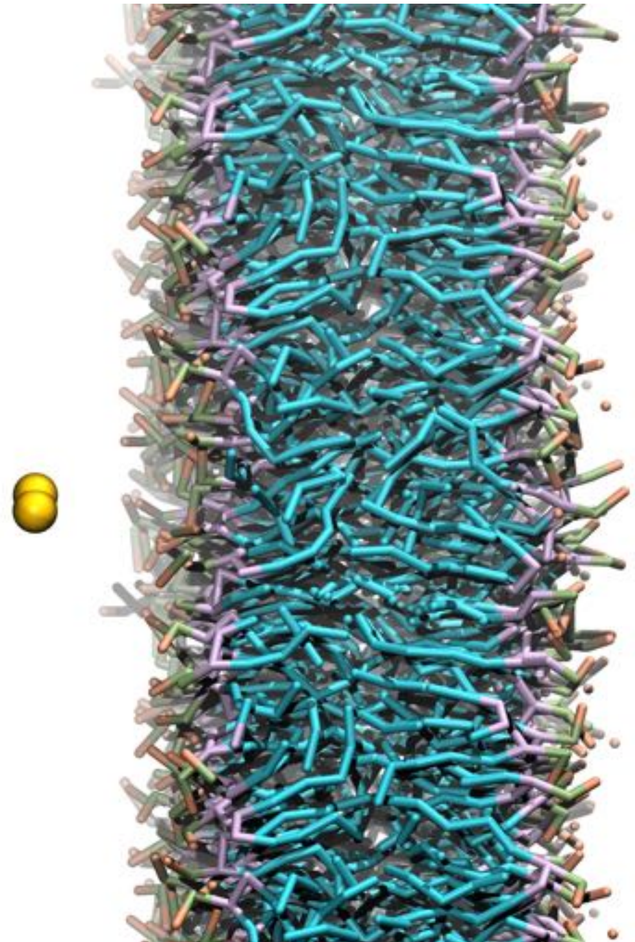
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Impact of chemical group on permeability



Permeability coefficient

$$P^{-1} = \int \mathrm{d}z \frac{\exp\left(\Delta G(z)/k_\mathrm{B}T\right)}{D_z(z)}$$

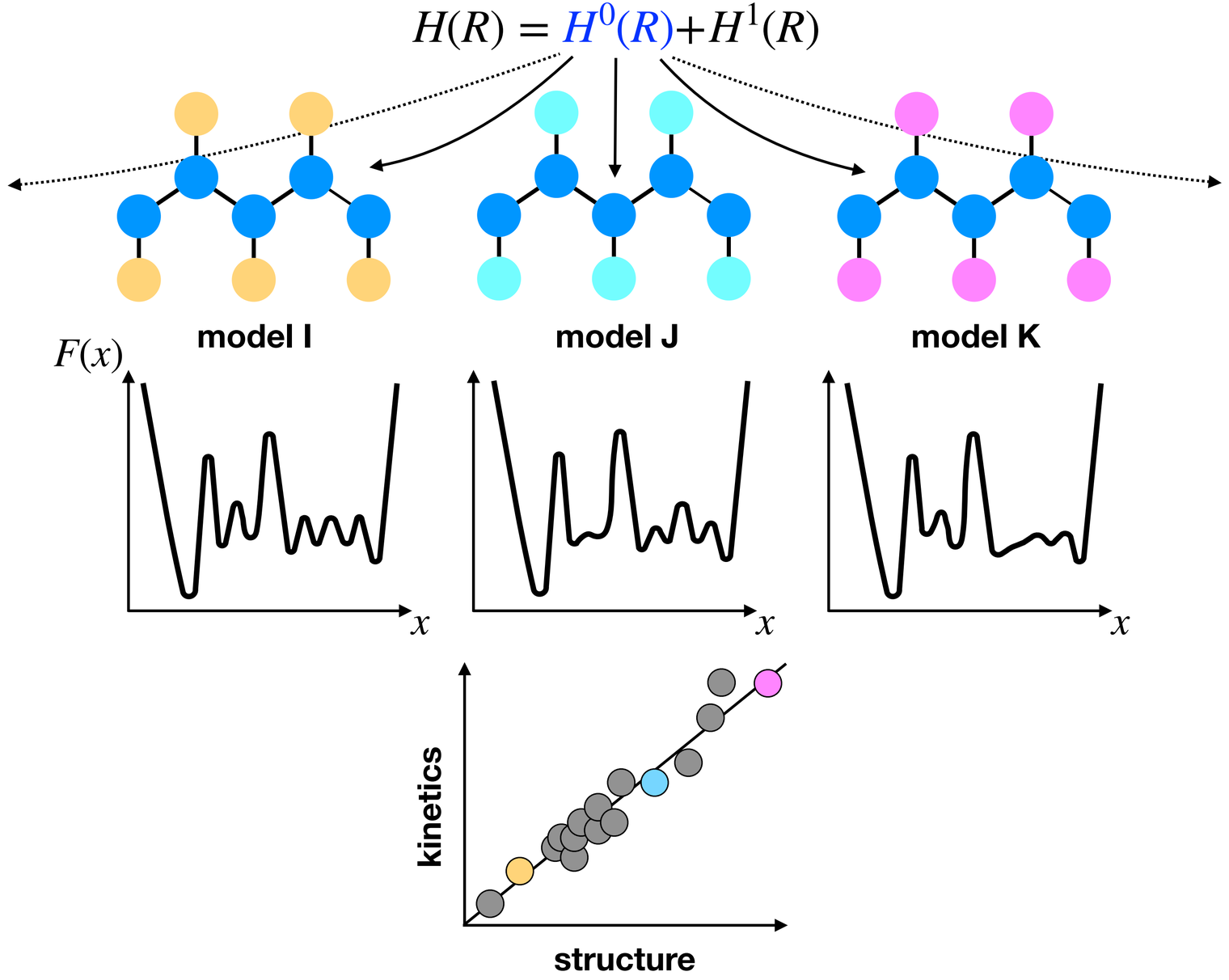


Analysis of **500,000+** compounds

Menichetti, Kanekal, Bereau, *ACS Cent. Sci.* **5** (2019)

Connecting structure and kinetics

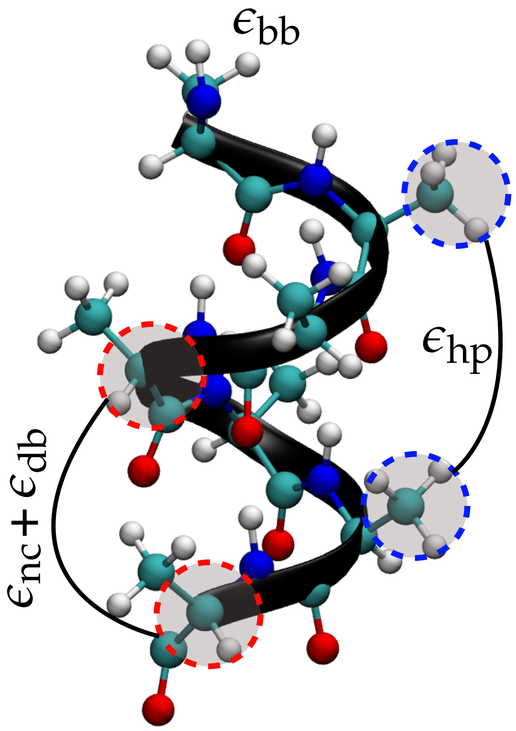
Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research



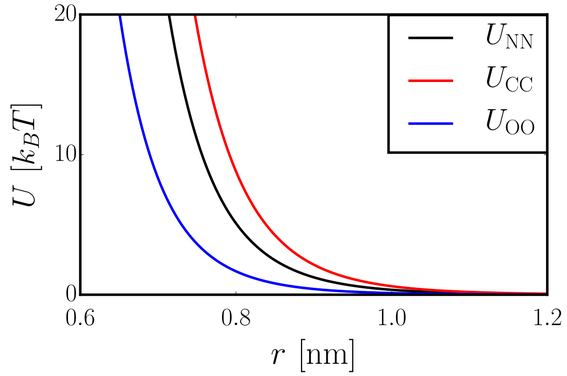
Investigation of structural-kinetic relationships for helix-coil transitions

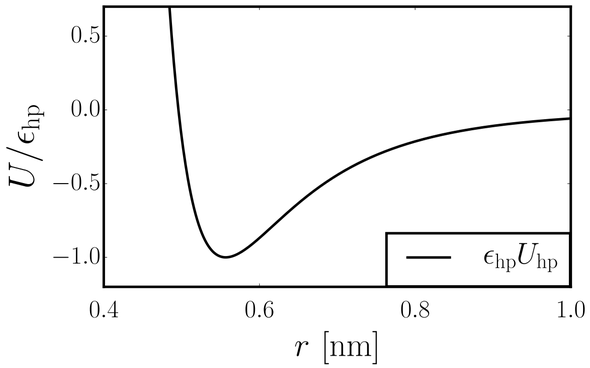
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Probing the role of steric interactions in the formation of free-energy barriers



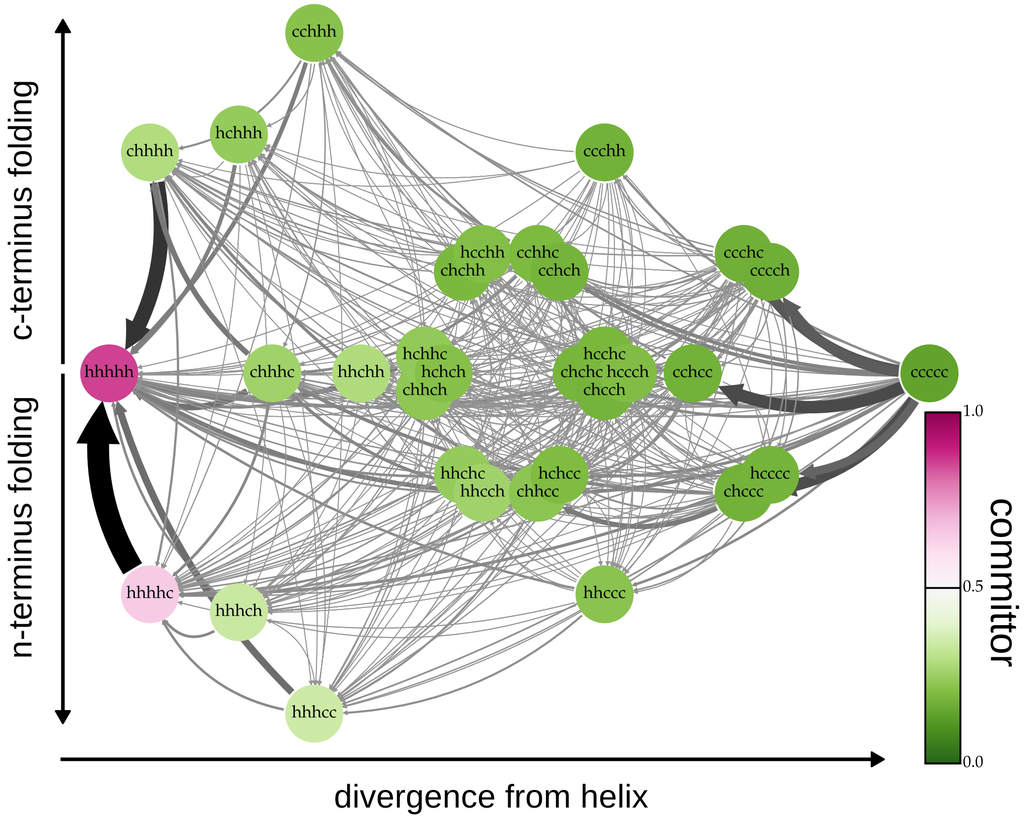
Detailed sterics





Simple energetics

Systematic characterization of kinetics through network description

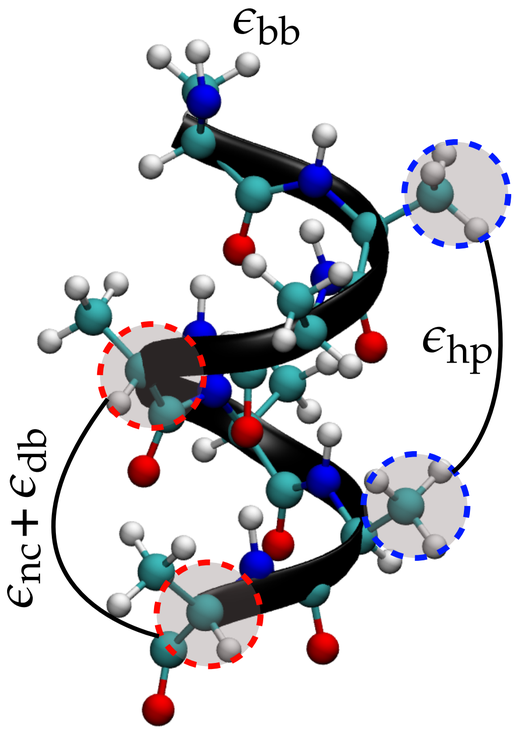


JFR, Bereau *Computation* (2018); JFR, Bereau *J Chem Phys* (2018)

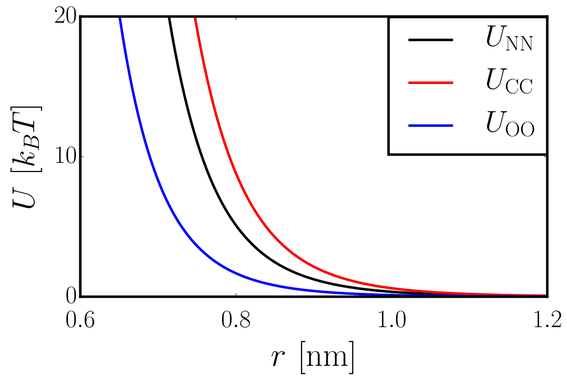
Investigation of structural-kinetic relationships for helix-coil transitions

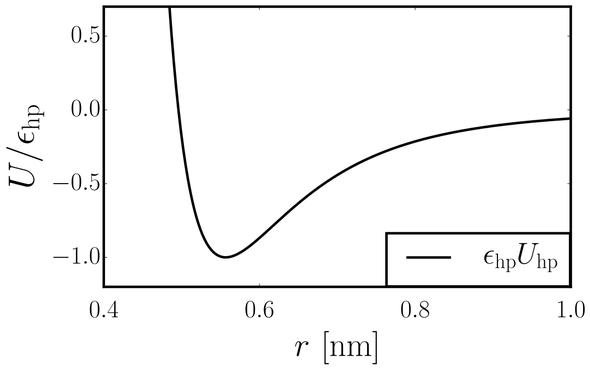
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

Probing the role of steric interactions in the formation of free-energy barriers



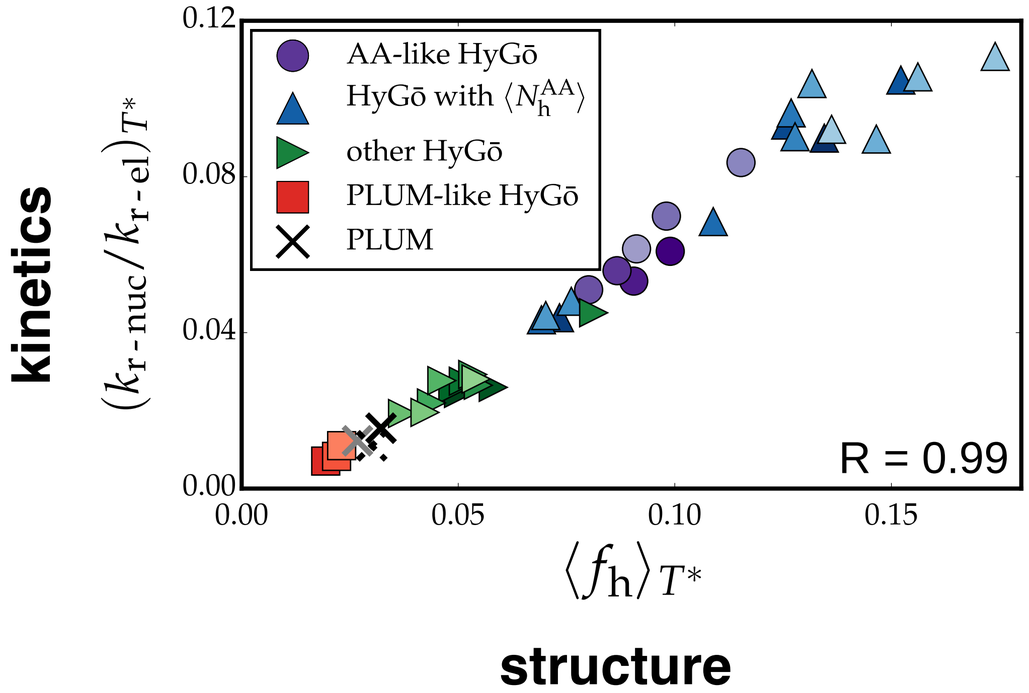
Detailed sterics





Simple energetics

Steric interactions alone determine a simple relationship between structure and kinetics



JFR, Bereau *Computation* (2018); JFR, Bereau *J Chem Phys* (2018)

Data-driven methods for soft matter

Max-Planck-Institut
für Polymerforschung
Max Planck Institute
for Polymer Research

Thank you for your attention!

Slides posted @ RudzinskiResearch.com

Main References:

Ferguson *J. Phys.: Condens. Matter* **(2018)** “Machine learning and data science in soft materials engineering”

Bereau *Handbook of Materials Modeling: Methods: Theory and Modeling* **(2018)** “Data-Driven Methods in Multiscale Modeling of Soft Matter”

Jackson *et al. Curr. Opin. Chem. Eng.* **(2019)** “Recent advances in machine learning towards multiscale soft materials design”

Klus *et al. Int. J. Nonlinear Sci.* **(2018)** “Data-Driven Model Reduction and Transfer Operator Approximation”

Noe *ArXiv* **(2019)** “Machine Learning for Molecular Dynamics on Long Timescales”

Kingma, Welling *ArXiv* **(2017)** “An Introduction to Variational Autoencoders”