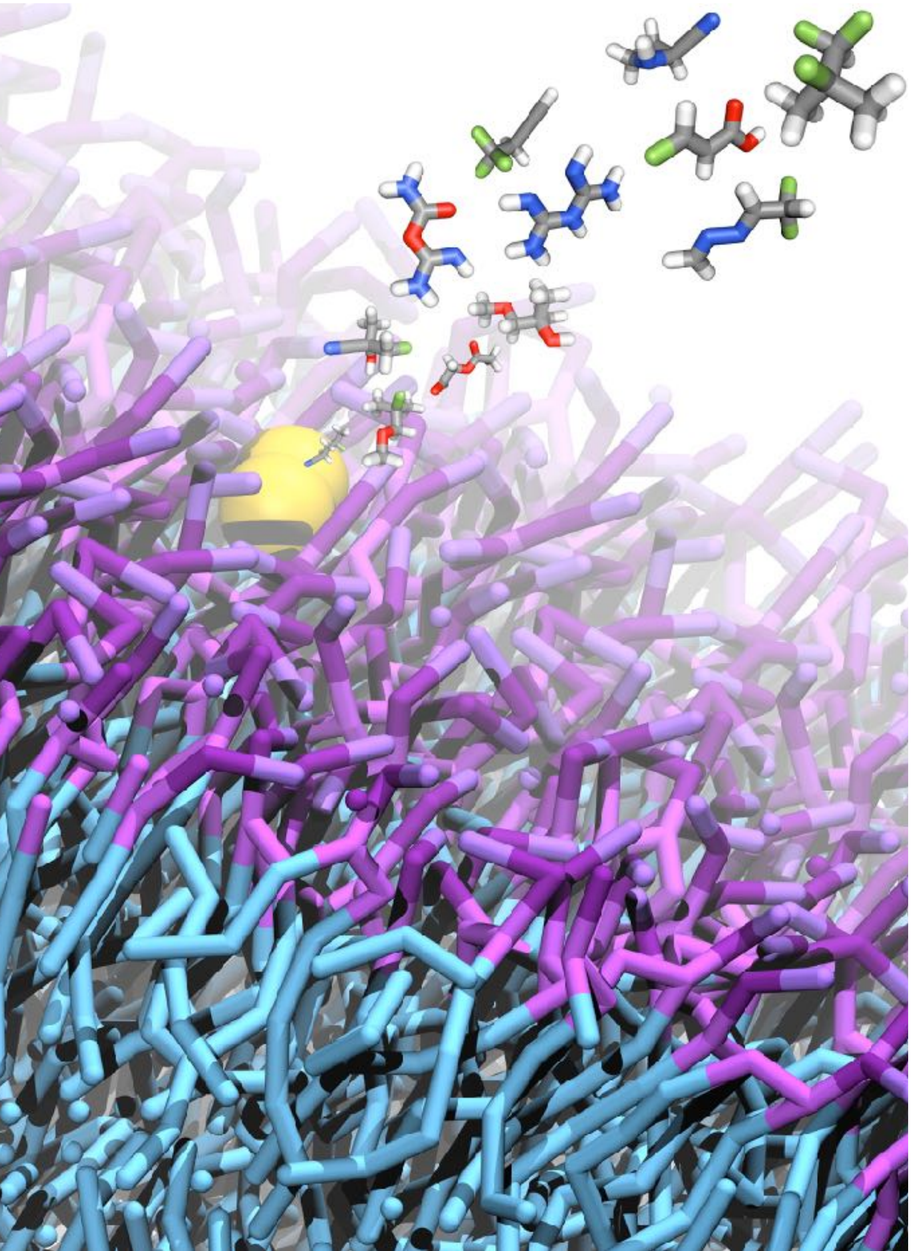
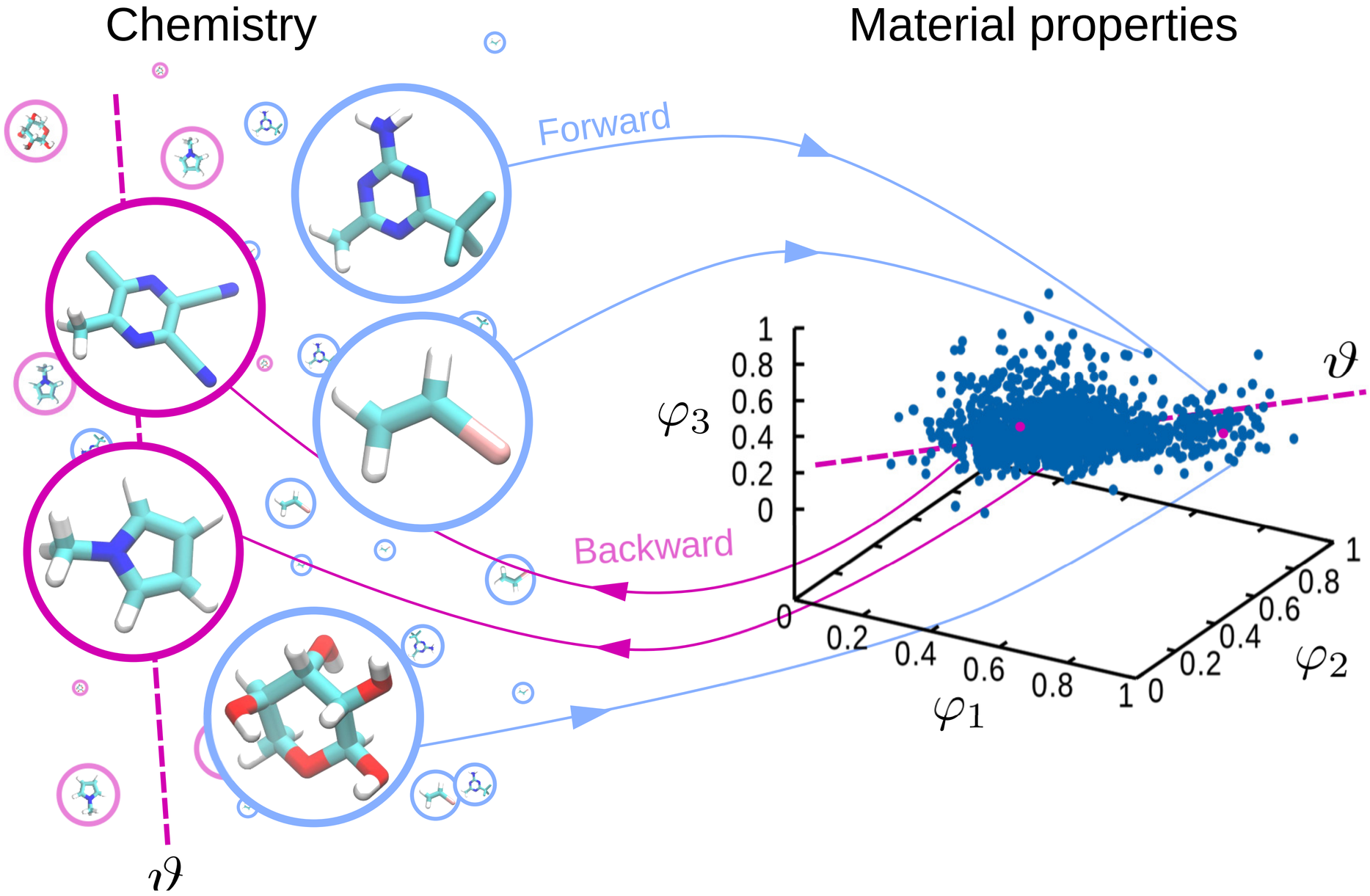


MAX PLANCK INSTITUTE
FOR POLYMER RESEARCH

Machine learning for molecular simulations: priors and predictive constraints

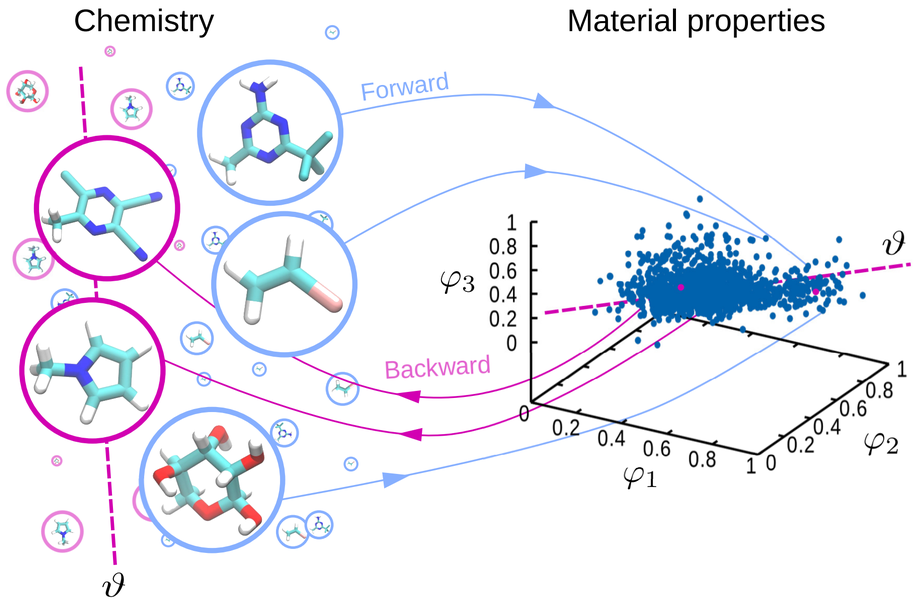
Tristan Bereau
Max Planck Institute for Polymer Research, Mainz, Germany

Rational design from structure-property relationships

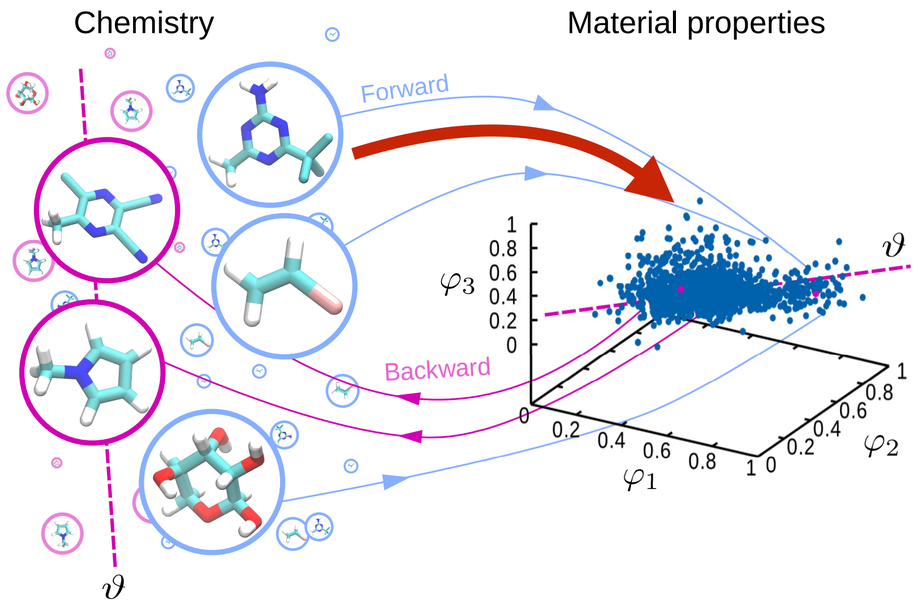


Pyzer-Knapp *et al.*, *Ann Rev Mat Res* **45** (2015); Jain *et al.*, *Nat Rev Mat* **1** (2016); Bereau *et al.*, *APL Materials* **4** (2016)

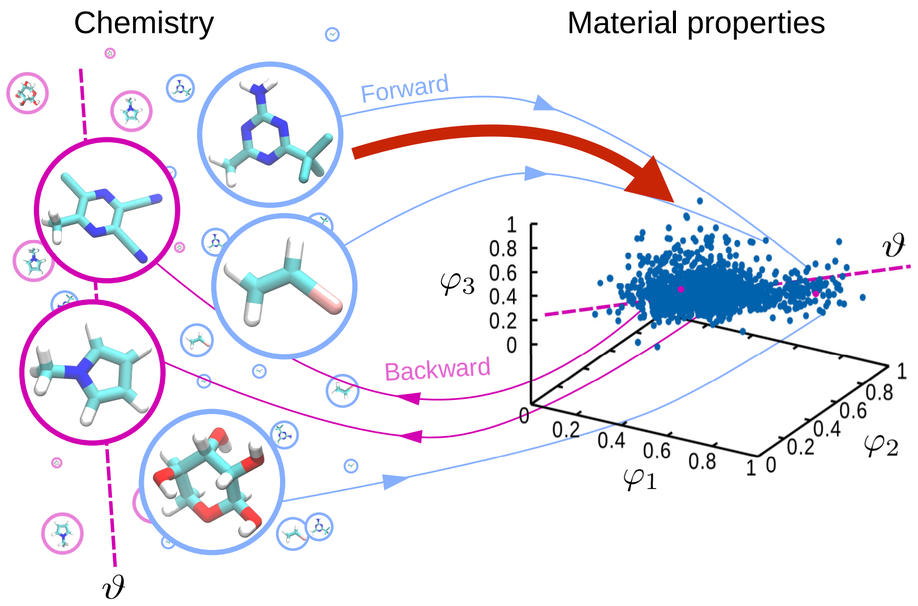
High-throughput screening



High-throughput screening

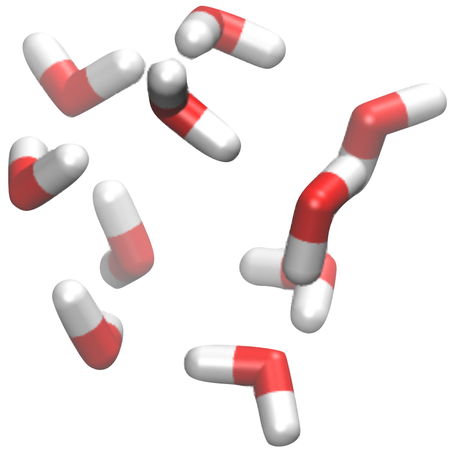


High-throughput screening



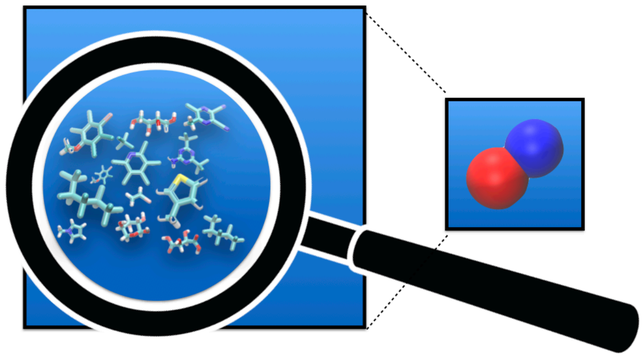
Outline

Outline



Intermolecular interactions across chemical space

High-throughput thermodynamics

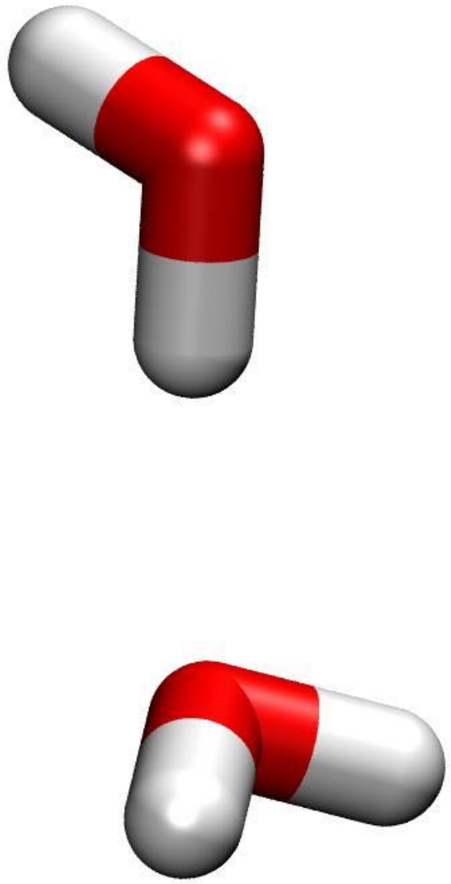


Predicting the intermolecular energy landscape

Predicting the intermolecular energy landscape

$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

Predicting the intermolecular energy landscape



Nuclear charges

$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

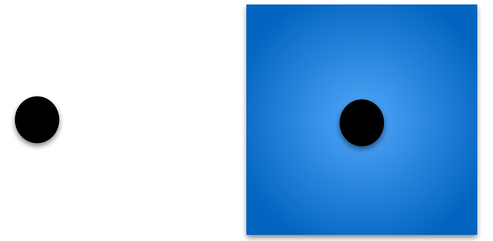
Coordinates

Potential energy

Transferability challenges

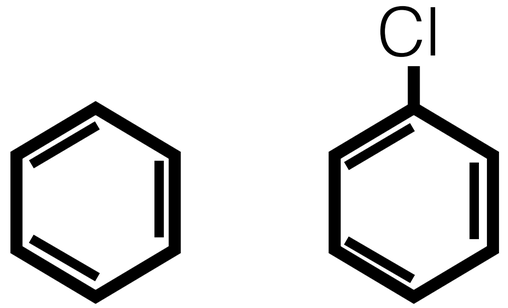
$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

Phase (e.g., gas vs. condensed)



Conformational (e.g., *cis* vs. *trans*)

Compositional

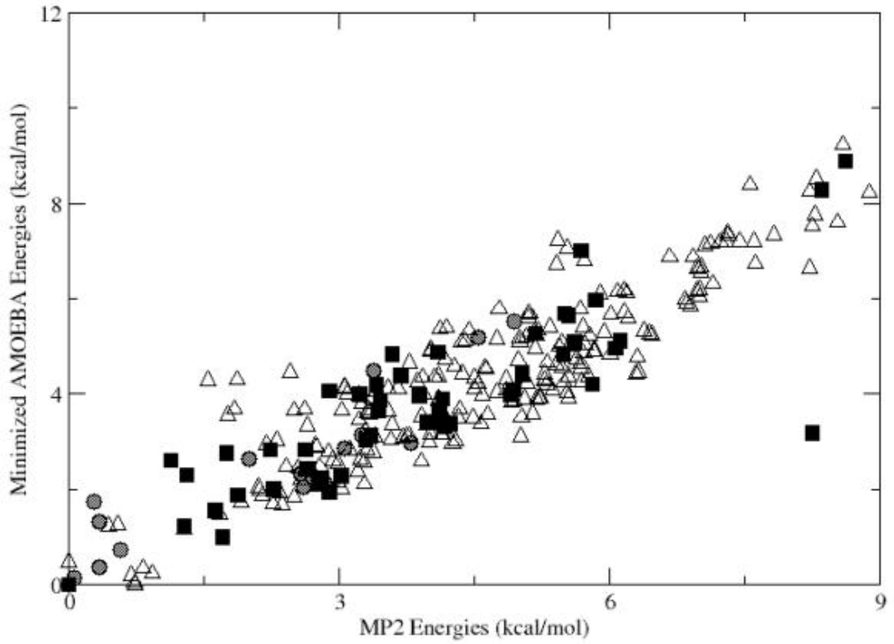


AMOEBA polarizable force field

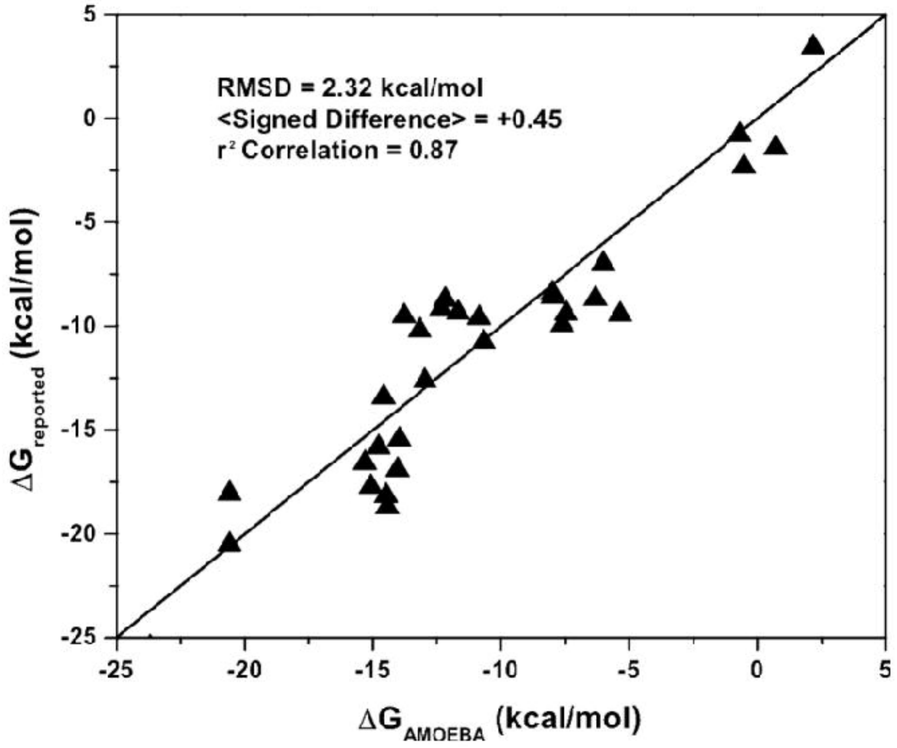
Max-Planck-Institut für Polymerforschung
Max Planck Institute for Polymer Research

$$U = U_{bond} + U_{angle} + U_{b\theta} + U_{oop} + U_{torsion} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{ind}$$

Water-sulfate anion clusters



Small molecule solvation free energies



Phase and conformational transferability

Ponder et al., J Phys Chem B (2010)

Strategies for energy landscape prediction

$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

1. Physics-based (force field)

2. Data-driven (machine learning)

Strategies for energy landscape prediction

$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

1. Physics-based (force field)

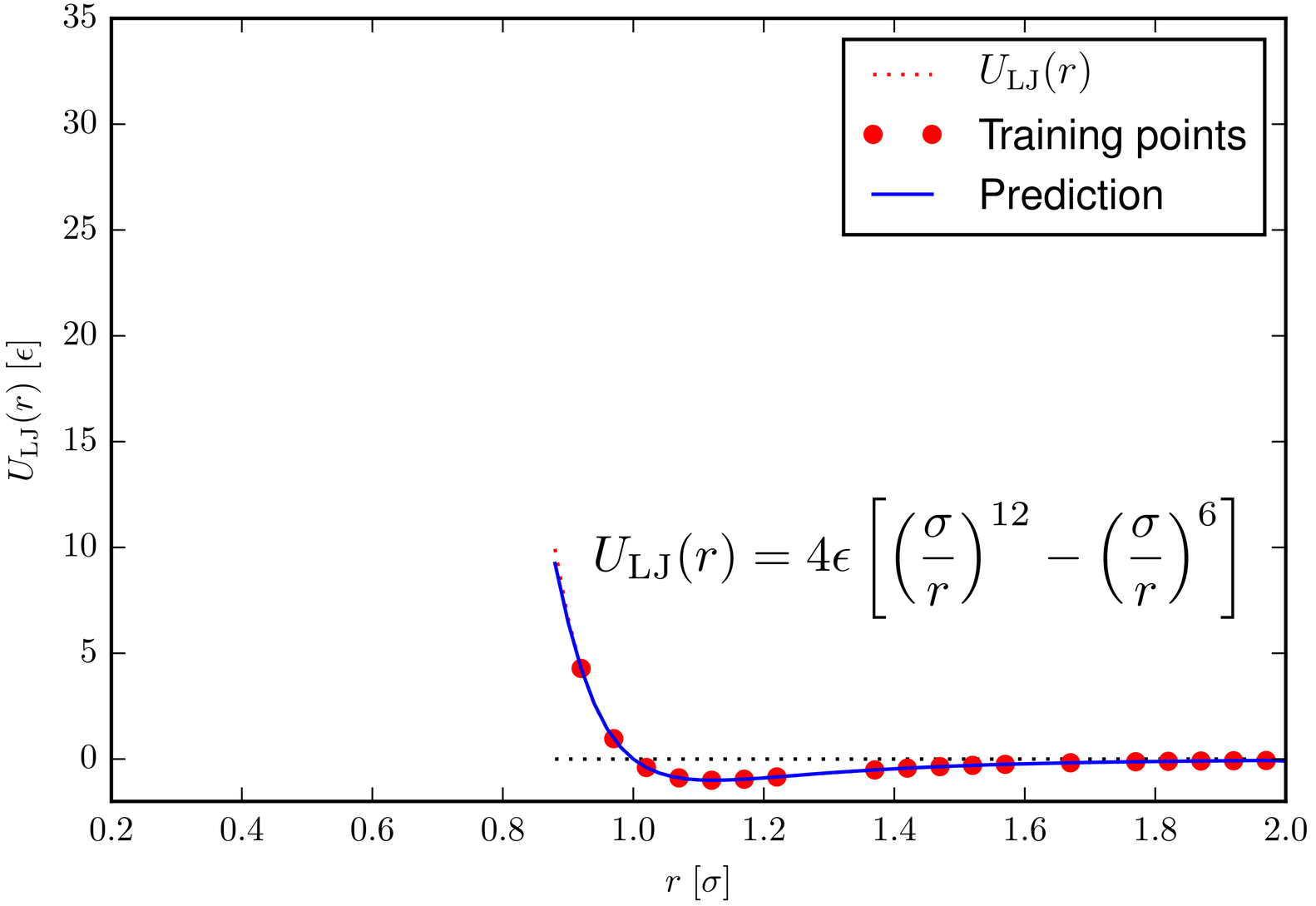
More parameters

Transferability

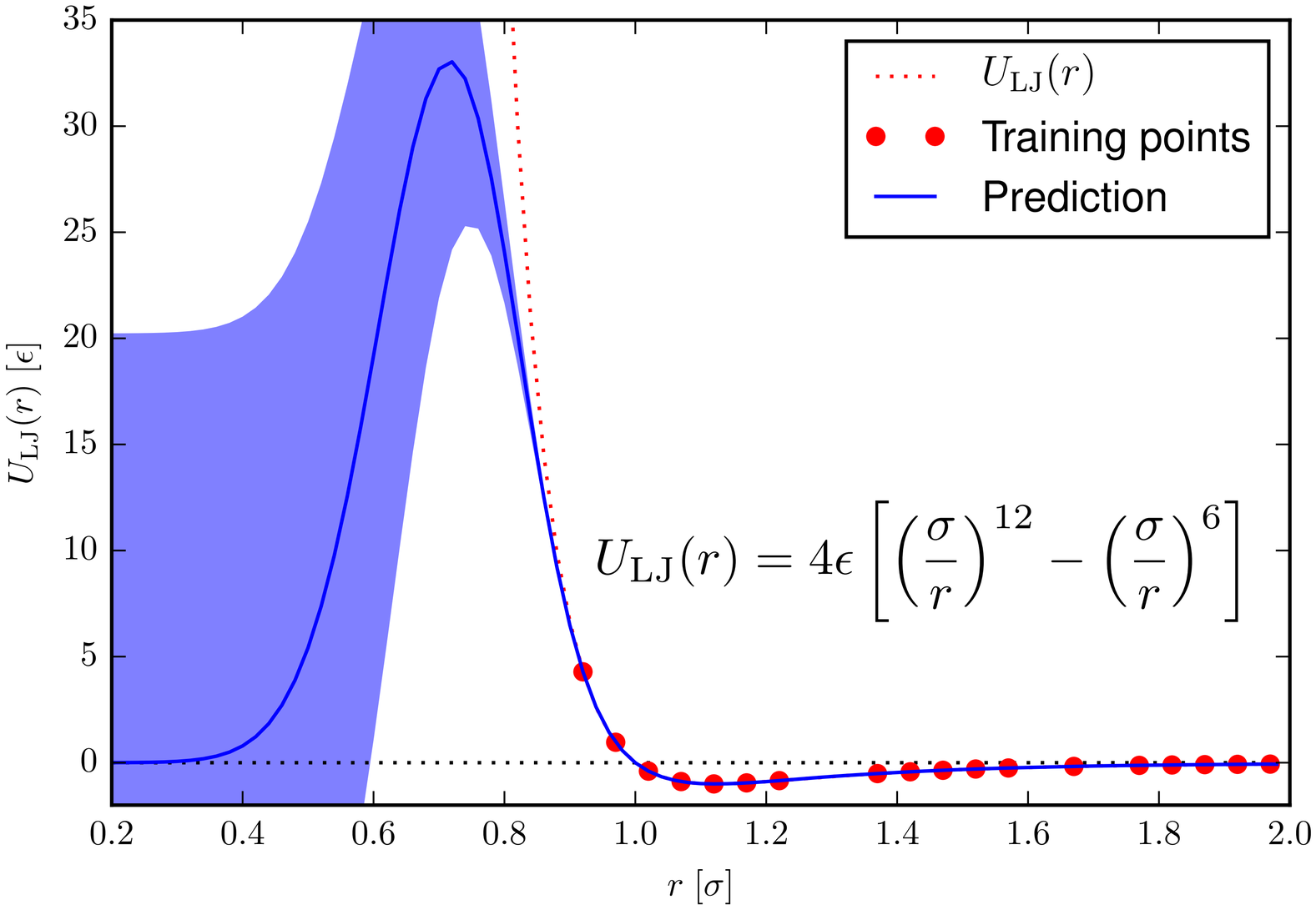
2. Data-driven (machine learning)

Extrapolation

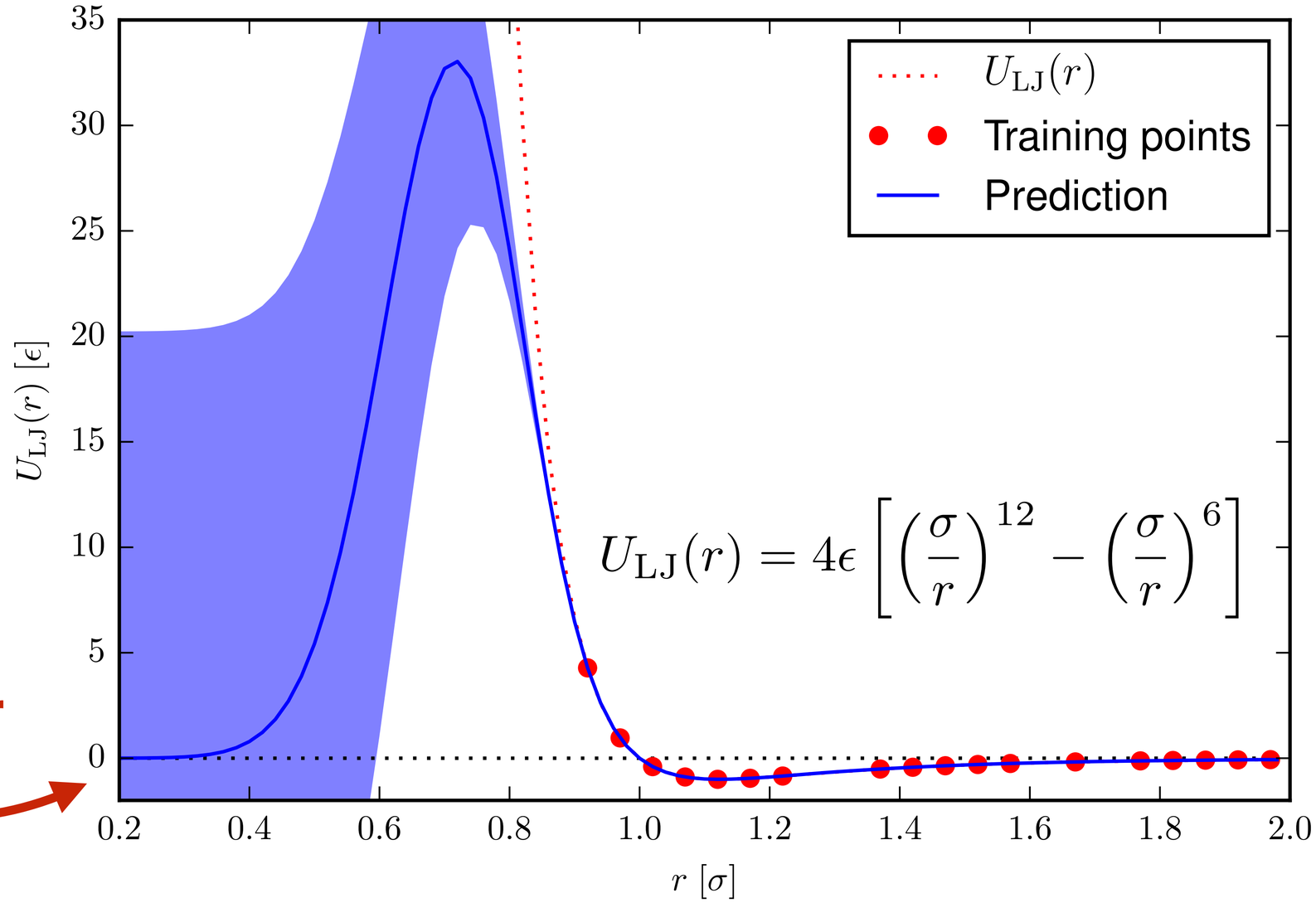
Extrapolation in machine learning



Extrapolation in machine learning

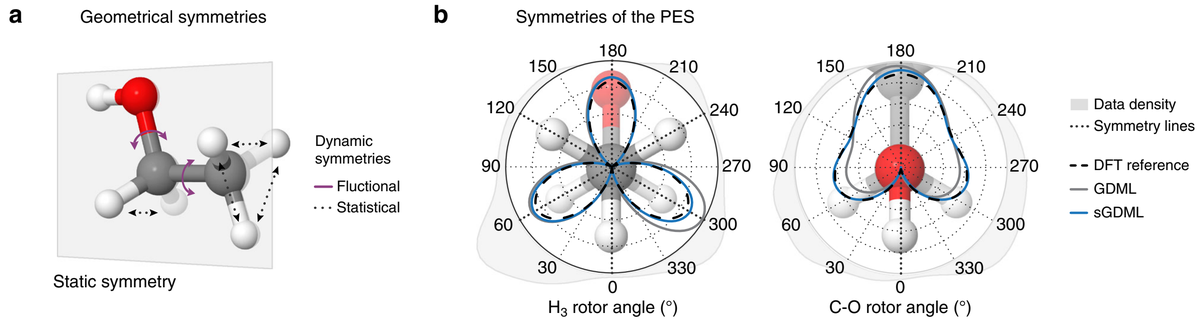


Extrapolation in machine learning



Uniform prior

Machine learning: encoding physics



$$\hat{\mathbf{f}}_{\mathbf{F}}(\mathbf{x}) = \sum_{i}^{M} \sum_{l}^{3N} \sum_{q}^{S} (\mathbf{P}_q \alpha_i)_l \frac{\partial}{\partial x_l} \nabla \kappa\left(\mathbf{x}, \mathbf{P}_q \mathbf{x}_i\right)$$

local symmetries

energy conservation

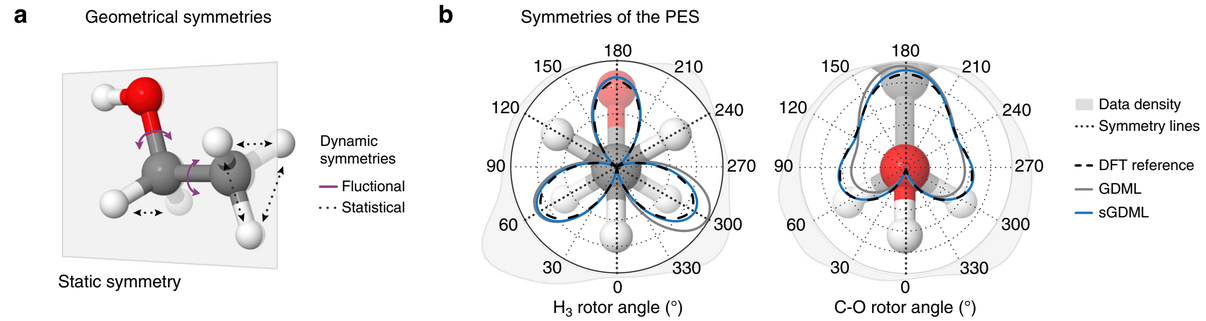
Chmiela, Sauceda, Müller, Tkatchenko, *Nat. Comm.* **9**:3887 (2018)

"Transform the configuration, and the prediction transforms with it"

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

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

Machine learning: encoding physics



$$\hat{\mathbf{f}}_{\mathbf{F}}(\mathbf{x}) = \sum_{i}^{M} \sum_{l}^{3N} \sum_{q}^{S} (\mathbf{P}_q \alpha_i)_l \frac{\partial}{\partial x_l} \nabla \kappa\left(\mathbf{x}, \mathbf{P}_q \mathbf{x}_i\right)$$

local symmetries

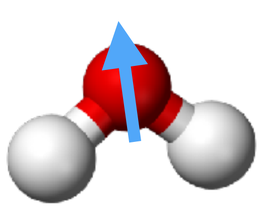
energy conservation

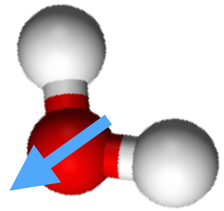
018)

Wishlist

- long-range interactions
- chemical-space transferability

Size of the interpolation-space?





"Transform the configuration, and the prediction transforms with it"

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

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

Intermolecular interactions across chemical space

Intermolecular interactions across chemical space

Nuclear charges

$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

Coordinates

Potential energy

Intermolecular interactions across chemical space

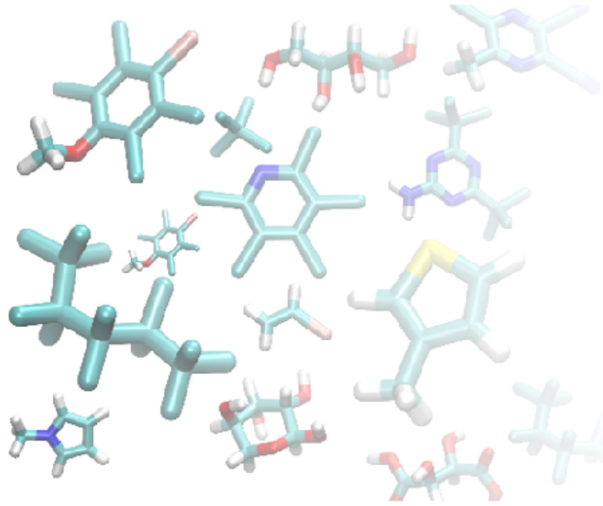
Nuclear charges

$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

Coordinates

Potential energy

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



Intermolecular interactions across chemical space

Nuclear charges

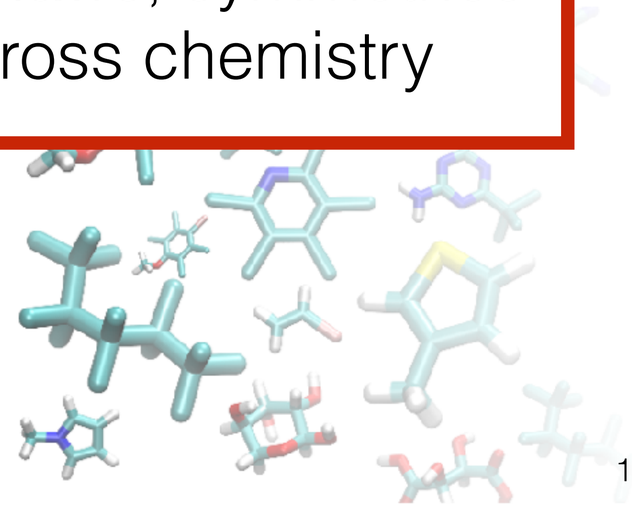
$$\{Z_i, \mathbf{r}_i\} \mapsto U(\{Z_i, \mathbf{r}_i\})$$

Coordinates

Potential energy

Physics-based models	**Machine learning**
• Encode laws, symmetries • Little chemical information	• Need to *learn* laws, symmetries • Interpolate across chemistry

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



Physics-based aspect

Long-ranged

- Static electrostatics
- Many-body dispersion
- Polarization

Short-ranged

- Charge penetration
- Repulsion
- (Charge transfer)

Van Vleet, Misquitta, Stone, and Schmidt, *J. Chem. Theory Comput.* **12** (2016); Vandenbrande, et al., *J. Chem. Theory Comput.* **13** (2017); Grimme, *J. Chem. Theory Comput.* **10** (2014); Verstraelen, et al., *J. Chem. Theory Comput.* **12** (2016); Metz et al, *J. Chem. Theory Comput* **12** (2016)

Physics-based aspect

Long-ranged

Perturbation theory

- Static electrostatics
- Many-body dispersion
- Polarization

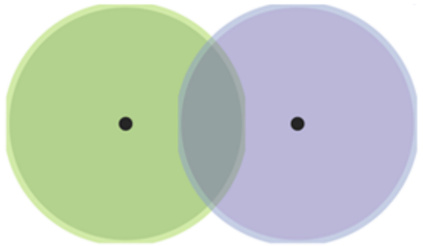
$$V_{\mathrm{Coulomb}}(r) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r}$$

Short-ranged

Overlap models

- Charge penetration
- Repulsion
- (Charge transfer)

$$S_{ij} = \int \mathrm{d}^3\mathbf{r}\rho_i(\mathbf{r})\rho_j(\mathbf{r})$$



Van Vleet, Misquitta, Stone, and Schmidt, *J. Chem. Theory Comput.* **12** (2016); Vandenbrande, et al., *J. Chem. Theory Comput.* **13** (2017); Grimme, *J. Chem. Theory Comput.* **10** (2014); Verstraelen, et al., *J. Chem. Theory Comput.* **12** (2016); Metz et al, *J. Chem. Theory Comput* **12** (2016)

Physics-based aspect

Long-ranged

- Static electrostatics
- Many-body dispersion
- Polarization

Short-ranged

- Charge penetration
- Repulsion
- (Charge transfer)

Use ML to predict atoms-in-molecules properties

- Multipole moments
- Hirshfeld ratios
- Atomic density widths/populations

Van Vleet, Misquitta, Stone, and Schmidt, *J. Chem. Theory Comput.* **12** (2016); Vandenbrande, et al., *J. Chem. Theory Comput.* **13** (2017); Grimme, *J. Chem. Theory Comput.* **10** (2014); Verstraelen, et al., *J. Chem. Theory Comput.* **12** (2016); Metz et al, *J. Chem. Theory Comput* **12** (2016)

Physics-based aspect

Long-ranged

- Static electrostatics
- Many-body dispersion
- Polarization

Multipole moments

Hirshfeld ratios

Short-ranged

- Charge penetration
- Repulsion
- (Charge transfer)

Atomic density widths/populations

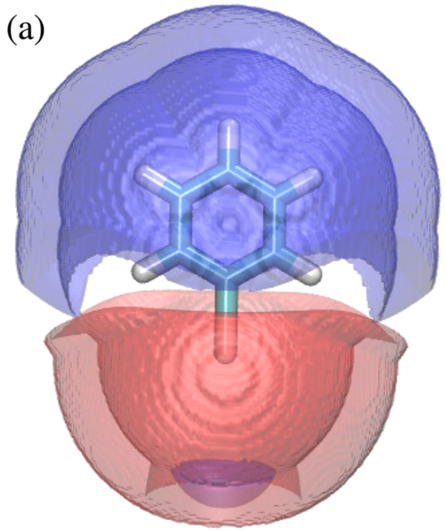
Use ML to predict atoms-in-molecules properties

- Multipole moments
- Hirshfeld ratios
- Atomic density widths/populations

Van Vleet, Misquitta, Stone, and Schmidt, *J. Chem. Theory Comput.* **12** (2016); Vandenbrande, et al., *J. Chem. Theory Comput.* **13** (2017); Grimme, *J. Chem. Theory Comput.* **10** (2014); Verstraelen, et al., *J. Chem. Theory Comput.* **12** (2016); Metz et al, *J. Chem. Theory Comput* **12** (2016)

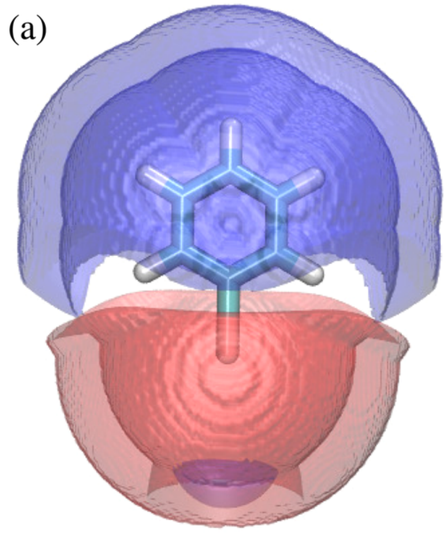
Static multipole electrostatics

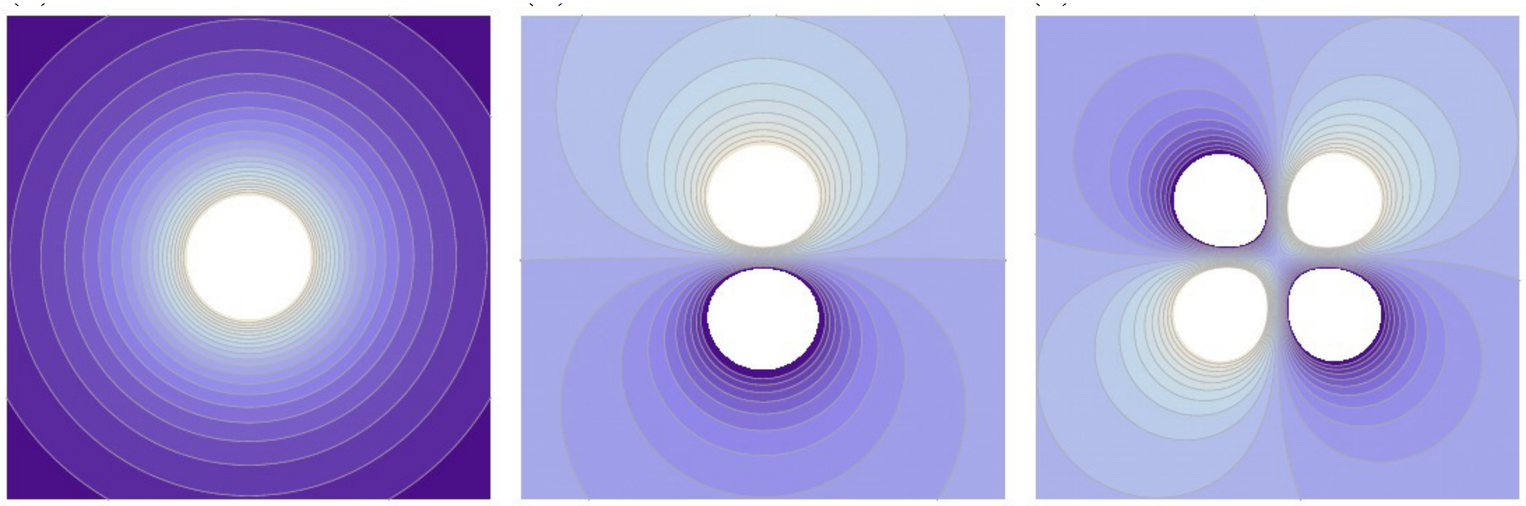
$$V_{\mathrm{Coulomb}}(r) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r}$$



Static multipole electrostatics

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

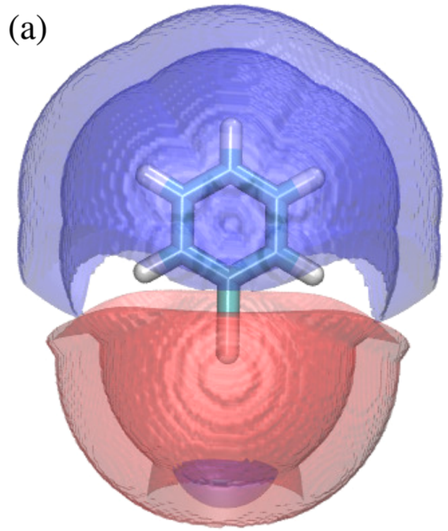


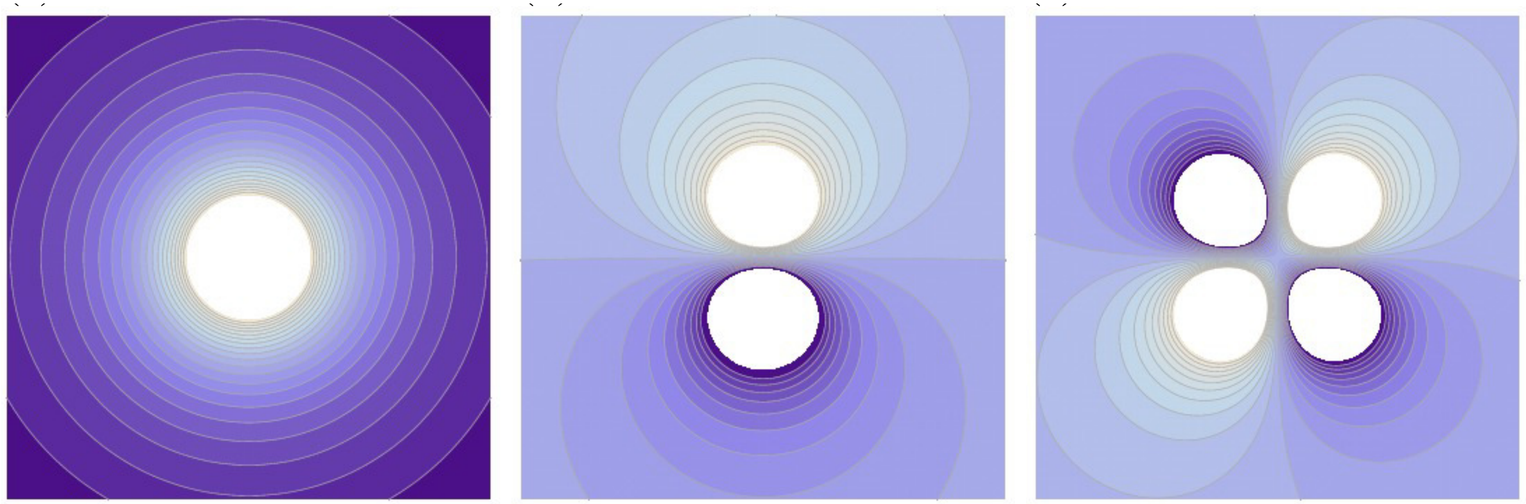


$$4\pi\varepsilon_0\Phi(\mathbf{r}) = \frac{q}{R} + \frac{\mu_\alpha R_\alpha}{R^3} + \frac{1}{3}\Theta_{\alpha\beta}\frac{3R_\alpha R_\beta - R^2\delta_{\alpha\beta}}{R^5} + \ldots$$

Static multipole electrostatics

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

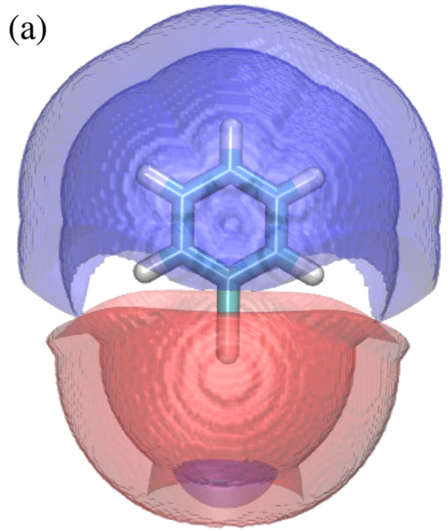


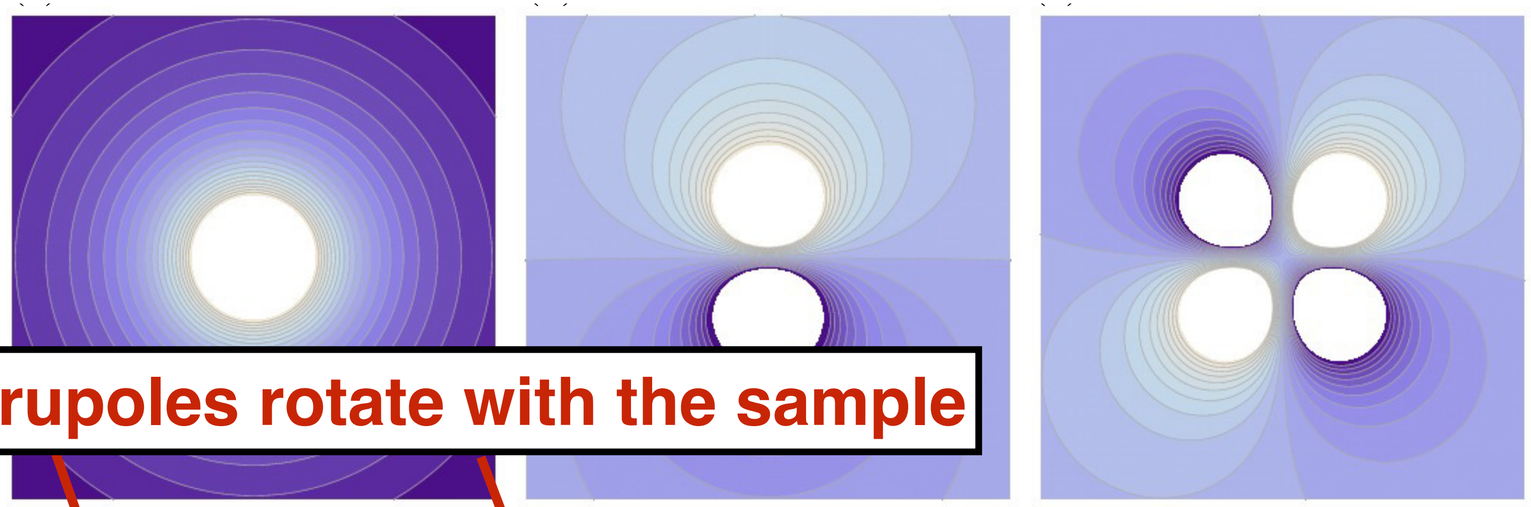


$$4\pi\varepsilon_0\Phi(\mathbf{r}) = \frac{q}{R} + \frac{\mu_\alpha R_\alpha}{R^3} + \frac{1}{3}\frac{\Theta_{\alpha\beta}\,3R_\alpha R_\beta - R^2\delta_{\alpha\beta}}{R^5} + \ldots$$

Static multipole electrostatics

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





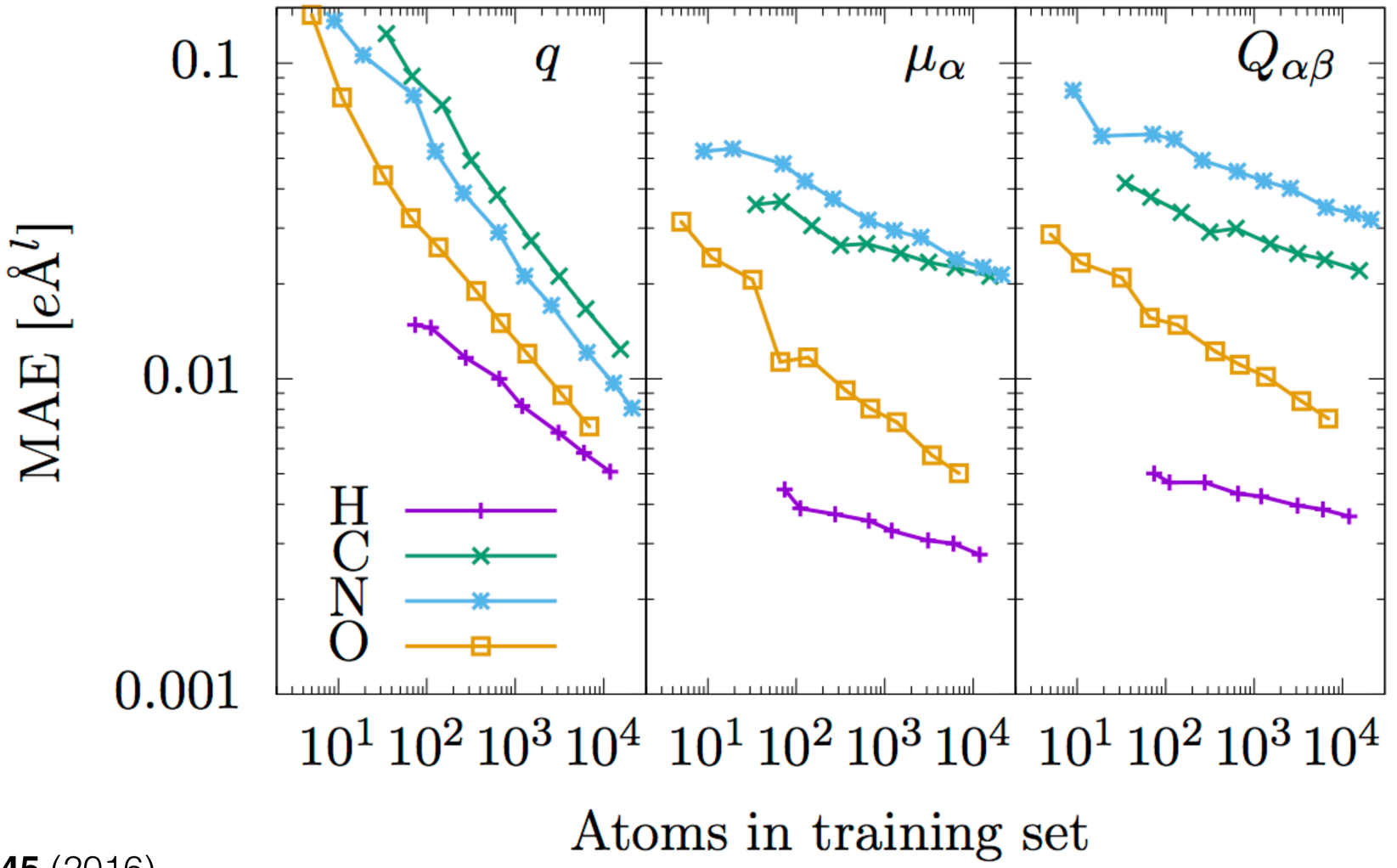
dipoles, quadrupoles rotate with the sample

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

Multipoles: Learning curves

kernel-ridge regression

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



Descriptor: aSLATM

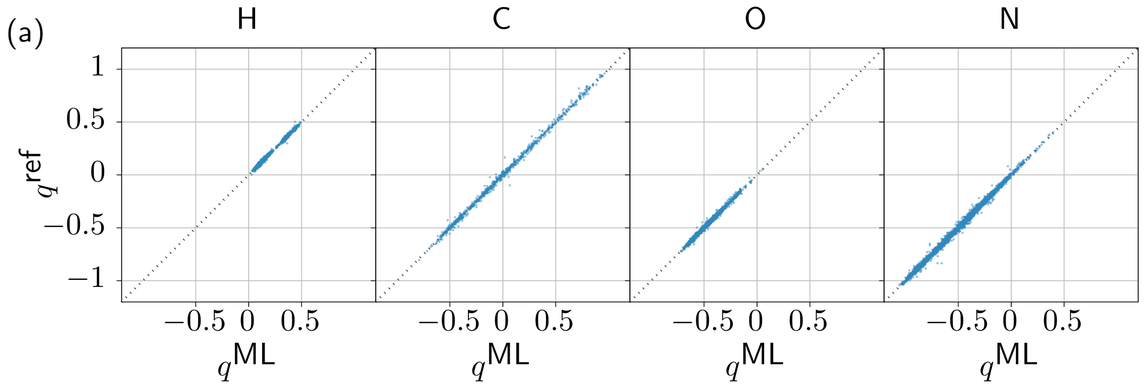
$$\mathbf{M}^I = [Z_I, \{\rho_I^{IJ}(R)\}, \{\rho_I^{IJK}(\theta)\}]$$

Huang and von Lilienfeld, *J. Chem. Phys.* **145** (2016)
Huang and von Lilienfeld, arXiv:1707.04146, 2017

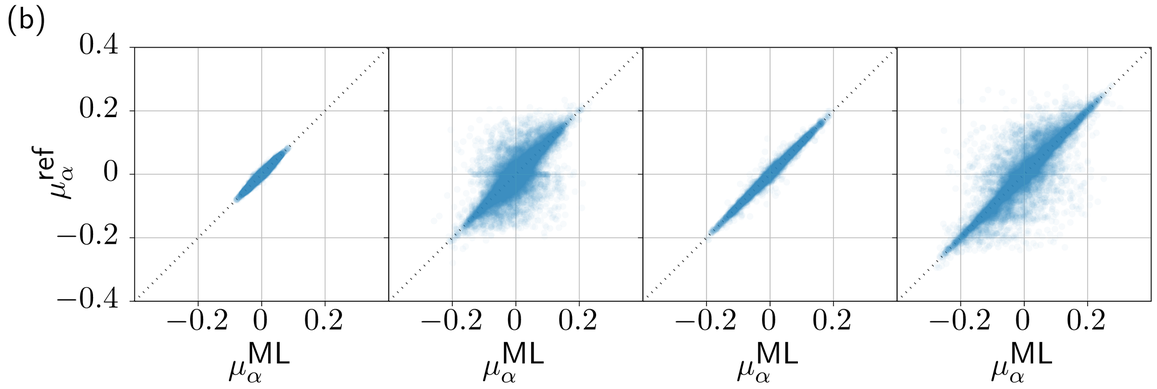
Easier to learn H,O than C,N

Multipoles: Correlation curves

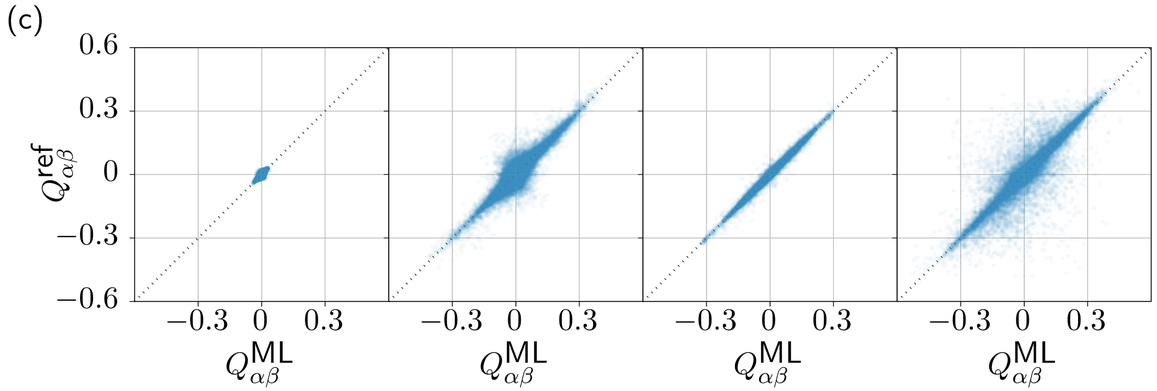
monopole



dipole



quadrupole



C,N have more complex valencies

Many-body dispersion

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

Pairwise

London dispersion:

$$E_{AB} = -\frac{C_{6AB}}{R^6}$$

Casimir-Polder:

$$C_{6AB} = \frac{3}{\pi}\int_0^\infty \mathrm{d}\omega \alpha_p(i\omega)\alpha_q(i\omega)$$

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

Many-body dispersion

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

Pairwise

London dispersion:

$$E_{AB} = -\frac{C_{6AB}}{R^6}$$

Casimir-Polder:

$$C_{6AB} = \frac{3}{\pi}\int_0^\infty \mathrm{d}\omega \alpha_p(i\omega)\alpha_q(i\omega)$$

Need atomic polarizabilities as input parameters

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

Many-body dispersion

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

Pairwise

London dispersion:

$$E_{AB} = -\frac{C_{6AB}}{R^6}$$

Casimir-Polder:

$$C_{6AB} = \frac{3}{\pi}\int_0^\infty \mathrm{d}\omega \alpha_p(i\omega)\alpha_q(i\omega)$$

Need atomic polarizabilities as input parameters

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

Many-body dispersion

Coupled fluctuating dipole model:

- Quantum harmonic oscillators

Donchev, J Chem Phys (2006)

Atomic polarizabilities

- Atom-in-molecule: Hirshfeld ratios

Tkatchenko et al., Phys Rev Lett (2012)

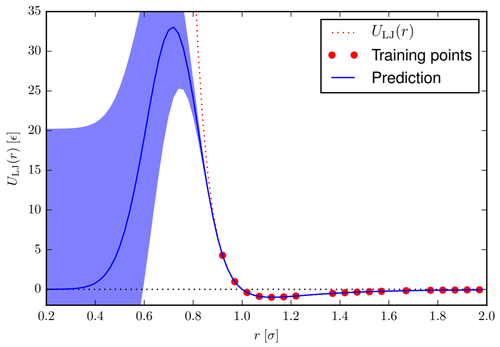
Many-body dispersion

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

Pairwise

London dispersion:

$$E_{AB} = -\frac{C_{6AB}}{R^6}$$



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

Casimir-Polder:

$$C_{6AB} = \frac{3}{\pi}\int_0^\infty \mathrm{d}\omega\alpha_p(i\omega)\alpha_q(i\omega)$$

Need atomic polarizabilities as input parameters

Many-body dispersion

Coupled fluctuating dipole model:

- Quantum harmonic oscillators

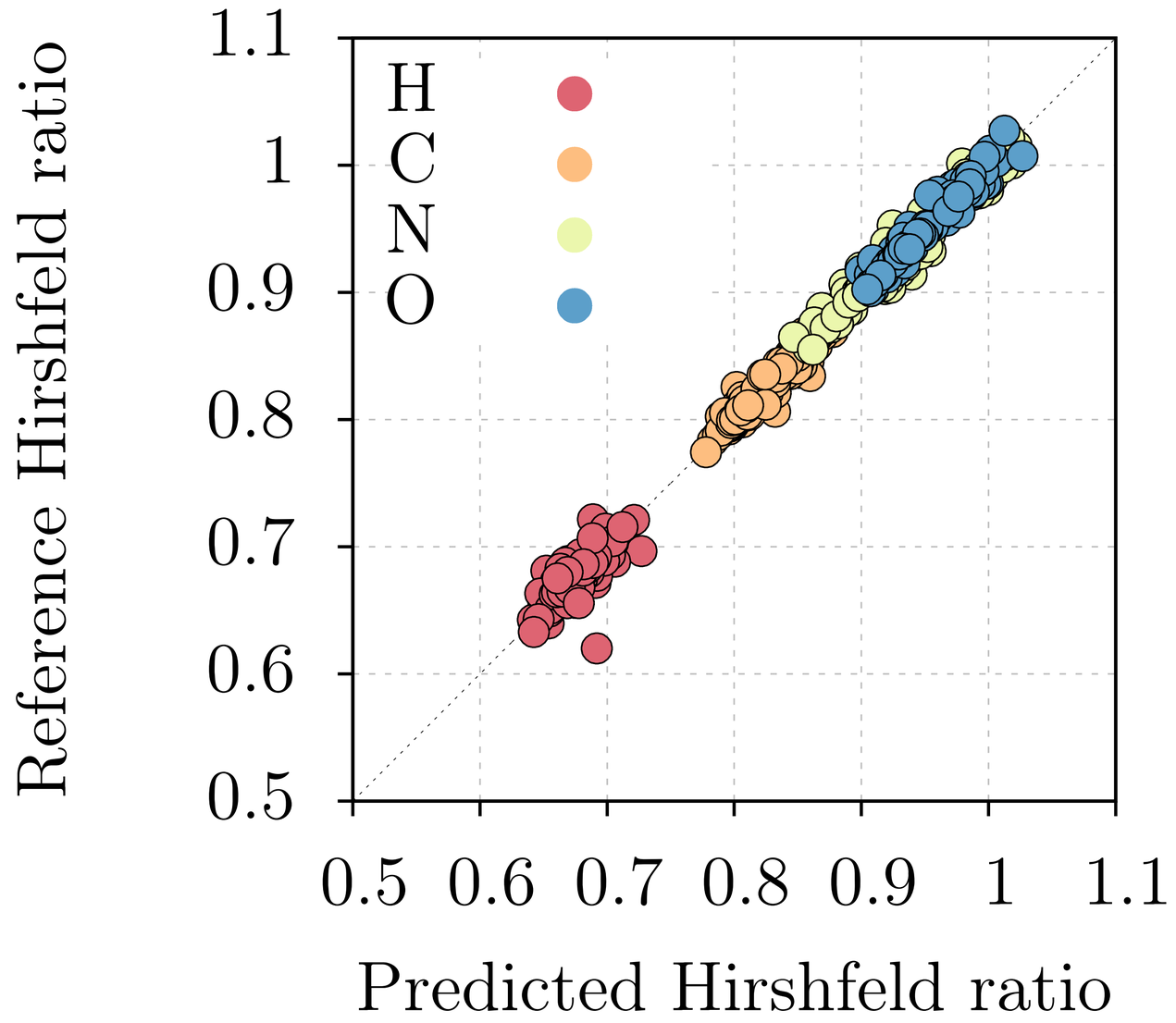
Donchev, J Chem Phys (2006)

Atomic polarizabilities

- Atom-in-molecule: Hirshfeld ratios

Tkatchenko et al., Phys Rev Lett (2012)

Learning Hirshfeld ratios



$$\alpha_p^0 \approx \frac{V_p^{\text{eff}}}{V_p^{\text{free}}}\alpha_p^{\text{free}} = \frac{\int \mathrm{d}\mathbf{r} r^3 w_p(\mathbf{r}) n(\mathbf{r})}{\int \mathrm{d}\mathbf{r} r^3 n_p^{\text{free}}}\alpha_p^{\text{free}}$$

Training: 12.3k atoms in 1k molecules

MAE: 0.006

$$C_{ij} = \begin{cases} \frac{1}{2} Z_i^{2.4} & \forall \ i = j \\ \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} & \forall \ i \neq j \end{cases}$$

Polarizability

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

Thole model

N point dipole polarizabilities placed in a homogeneous field **F**

Induced dipole moment at point *p*:

$$\mu_p = \alpha_p \left(\mathbf{F}_p - \sum_{q \neq p}^{N} T_{pq} \mu_q \right)$$

T_{pq} : dipole field tensor

$\mathbf{F}_p$ given by static electrostatic interactions at site *p*

Thole, *Chem Phys*, **59**, 341 (1981)

Polarizability

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

Thole model

N point dipole polarizabilities placed in a homogeneous field **F**

Induced dipole moment at point *p*:

$$\mu_p = \alpha_p \left(\mathbf{F}_p - \sum_{q \neq p}^{N} T_{pq} \mu_q \right)$$

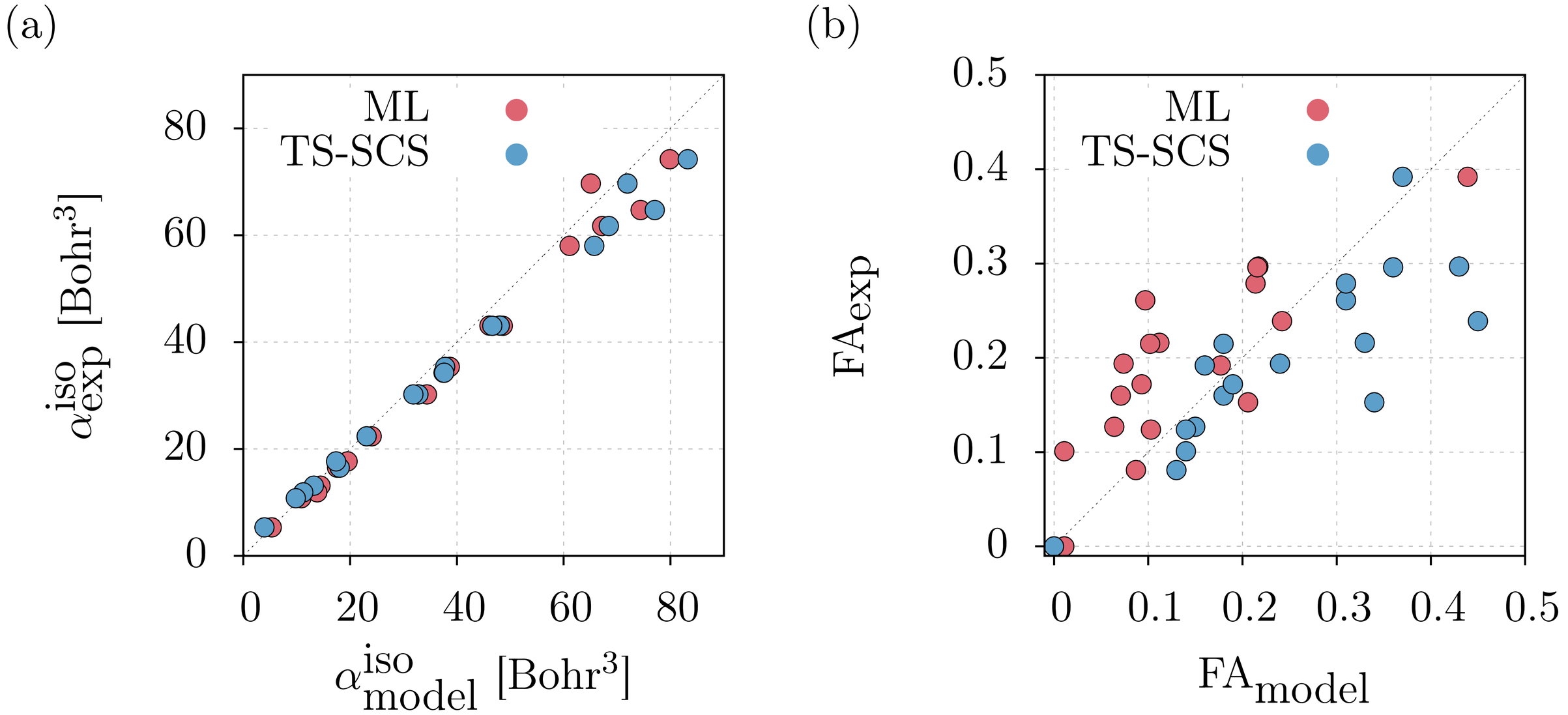
T_{pq} : dipole field tensor

$\mathbf{F}_p$ given by static electrostatic interactions at site *p*

Need atomic polarizabilities and multipoles as input parameters

Thole, *Chem Phys*, **59**, 341 (1981)

Molecular polarizabilities

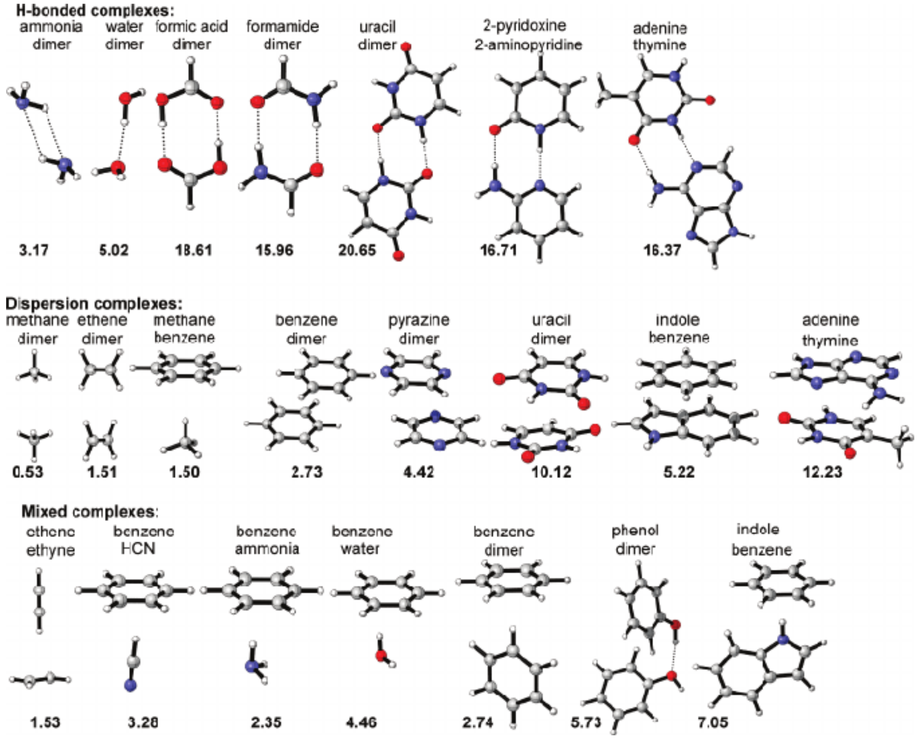


Tkatchenko and Scheffler, Phys Rev Lett (2009)

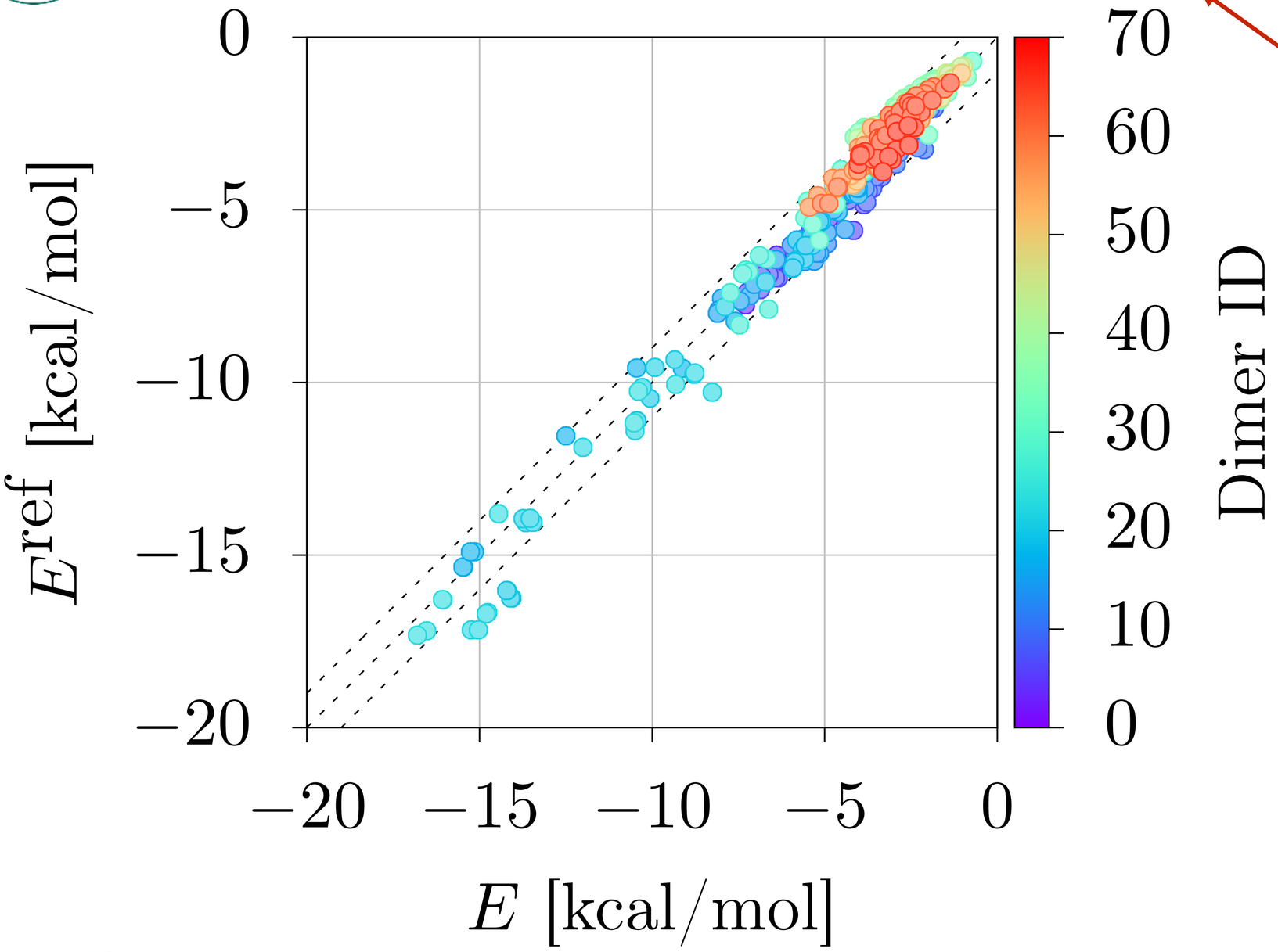
Few global parameters

7 parameters to cover chemical space of CHON intermolecular energies

Polarization	1
Many-body dispersion	2
Repulsion	C; H; O; N



Validation: S66a8

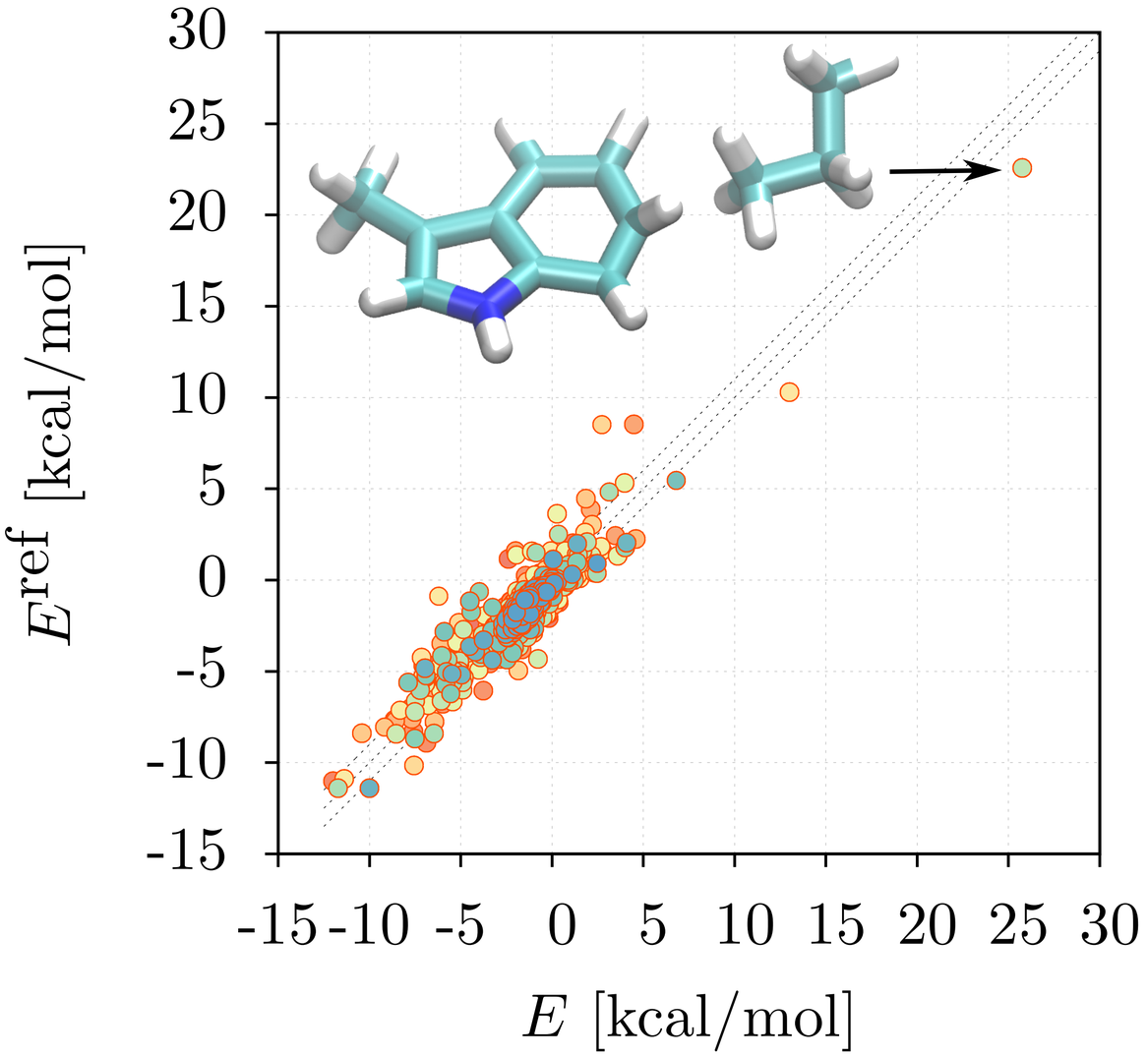


Angular-displaced nonequilibrium geometries

Error: **0.5 kcal/mol**

Rezác, Riley, and Hobza, *Journal of Chemical Theory and Computation*, **7** (2011).

Validation: SSI



2,200+ amino-acid pairs

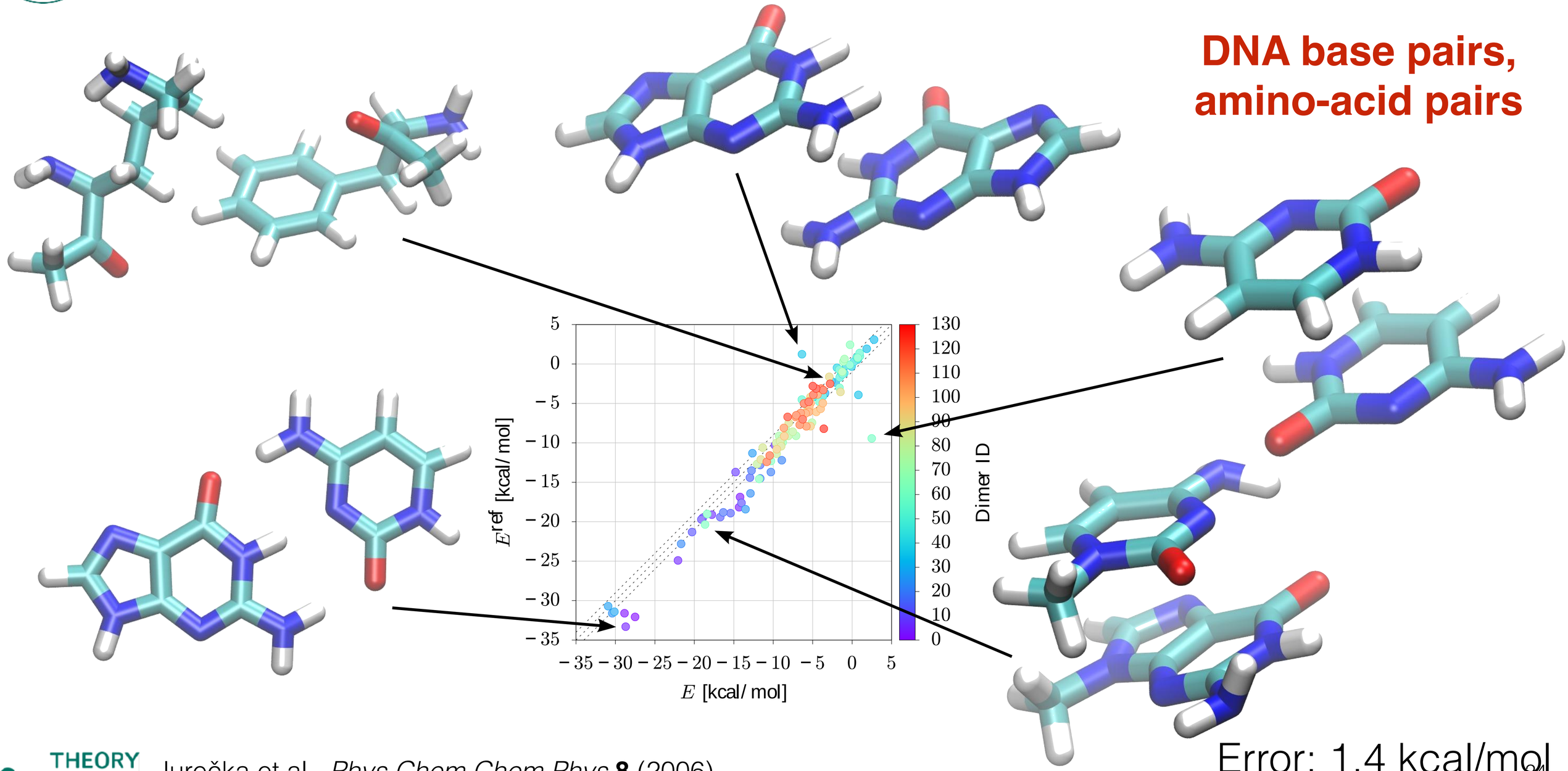
Ref: CCSD(T)/CBS

Error: **0.4 kcal/mol**

Burns *et al.*, *J Chem Phys*, **147** (2017).

Validation: JSCH-2005

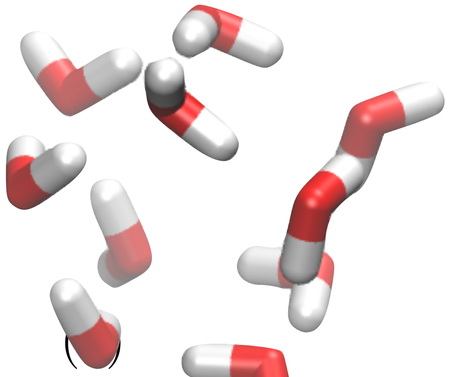
DNA base pairs, amino-acid pairs

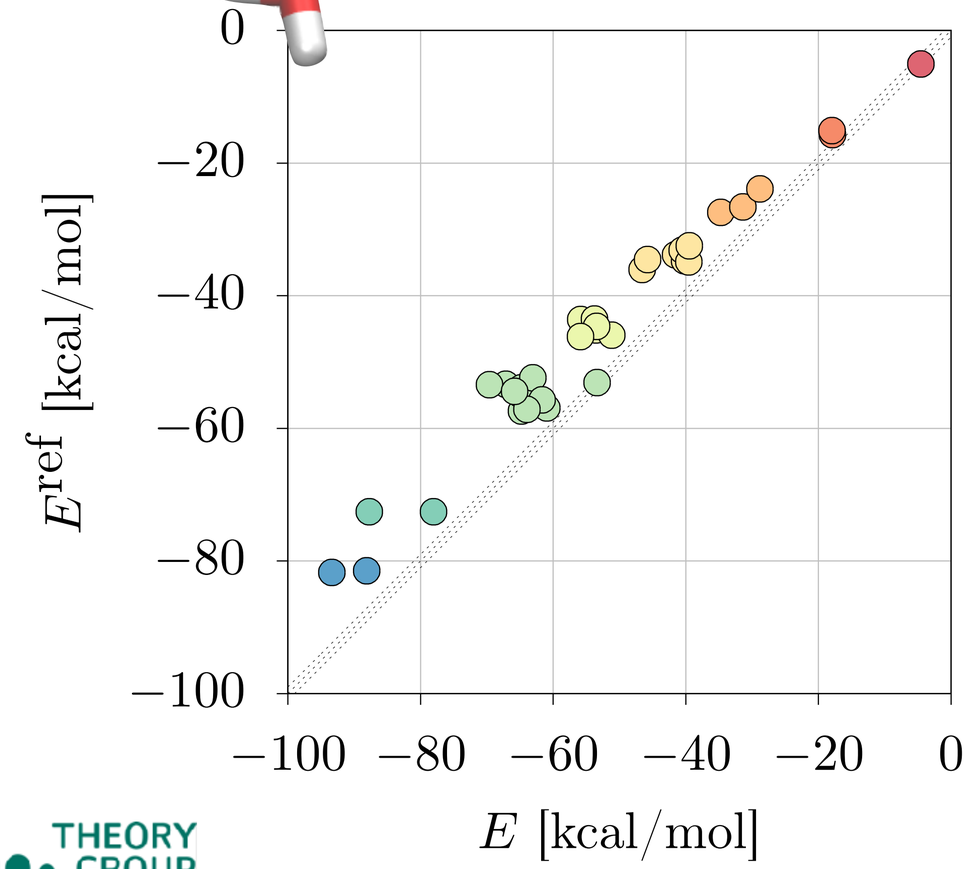


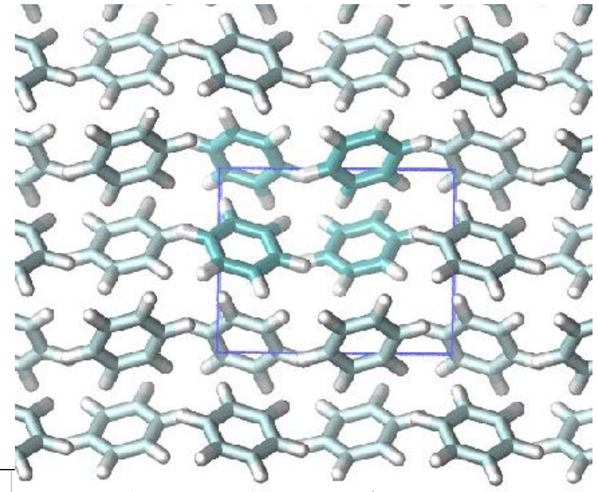
Error: 1.4 kcal/mol

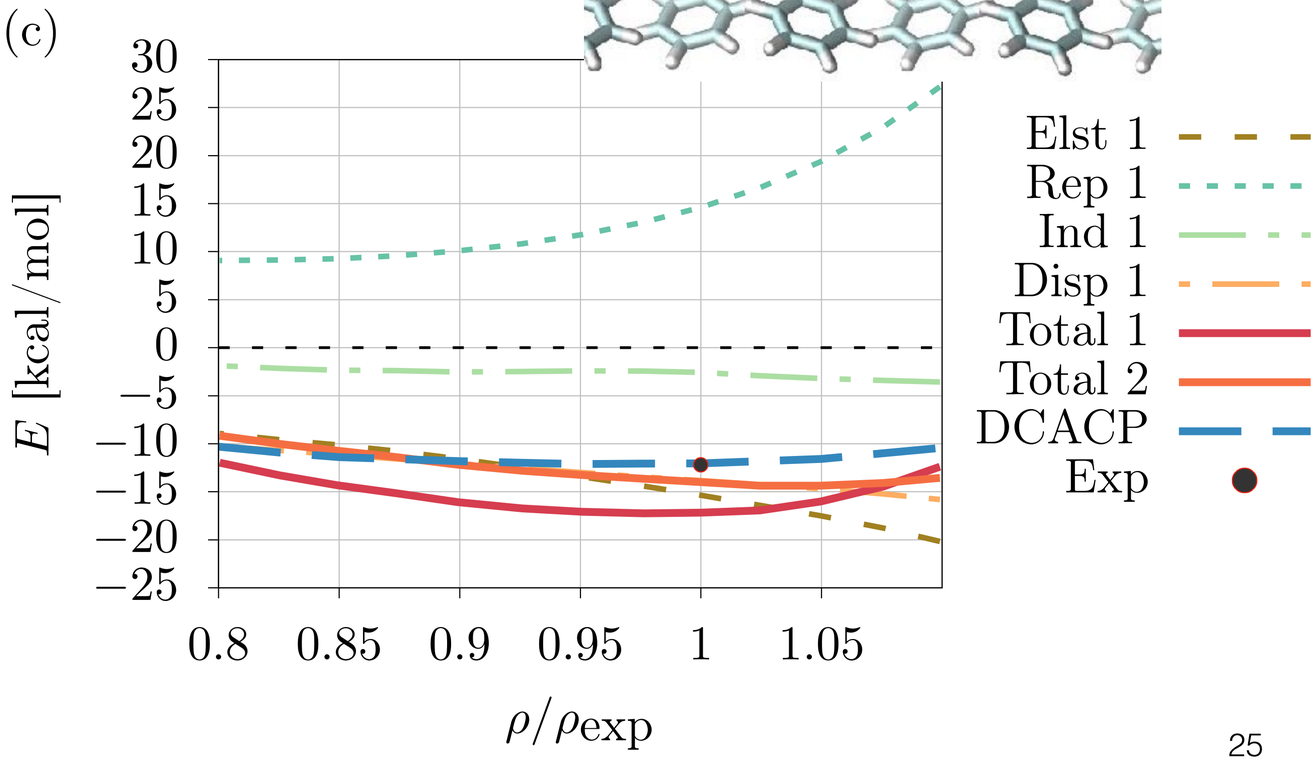
Jurečka et al., *Phys Chem Chem Phys* **8** (2006)

Toward the condensed phase

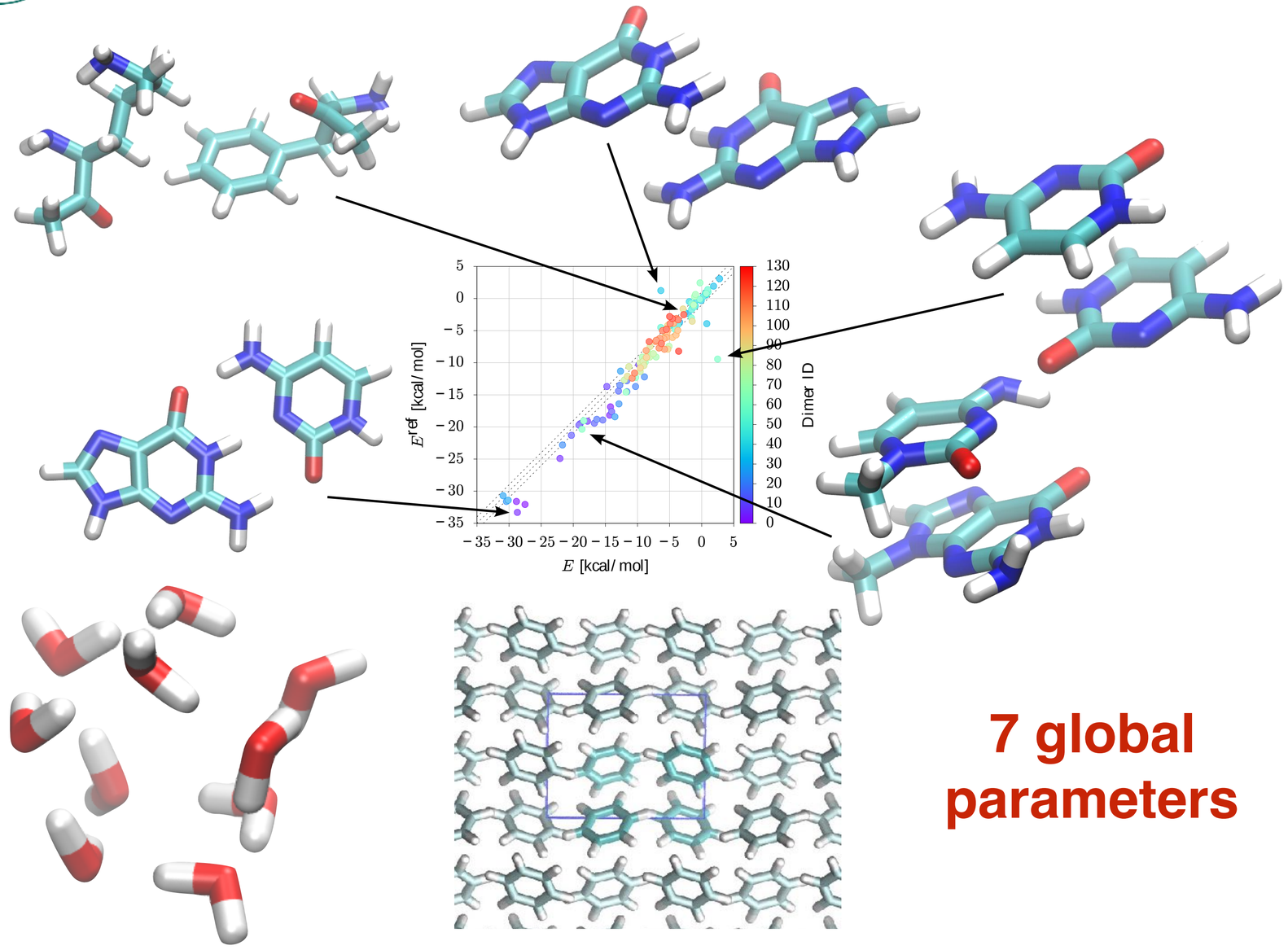


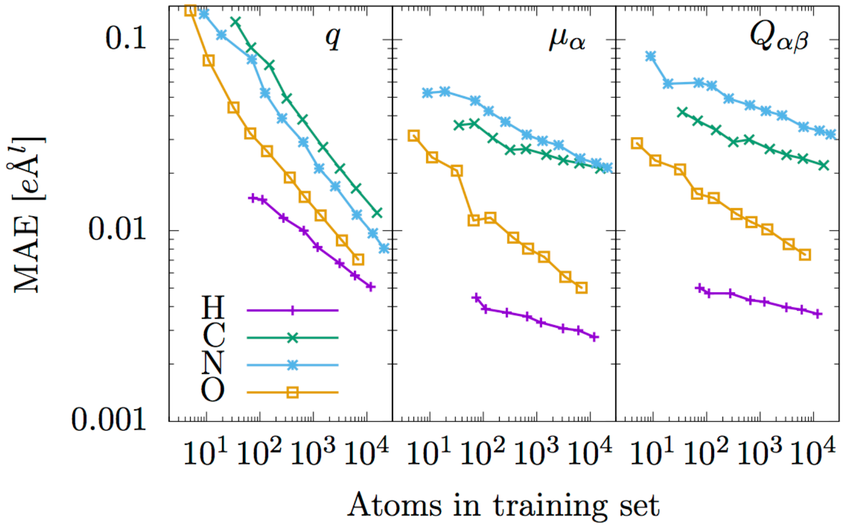






Intermolecular interactions across chemical space





7 global parameters

Lessons learned

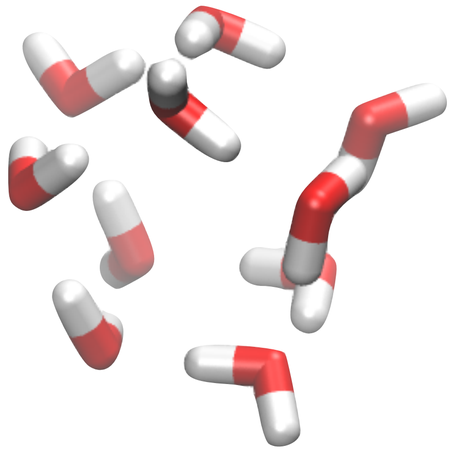
ML force fields can be systematically improved

- physical effects (perturbation theory at short range)
- chemical groups
- provide data + ML model

Bereau, DiStasio Jr., Tkatchenko, von Lilienfeld, *JCP* **148**, 241706 (2018)

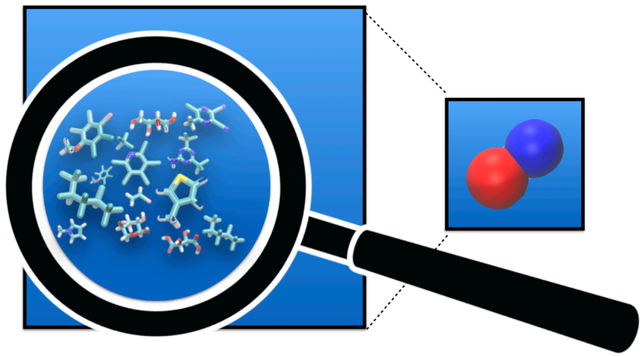
Outline

Outline

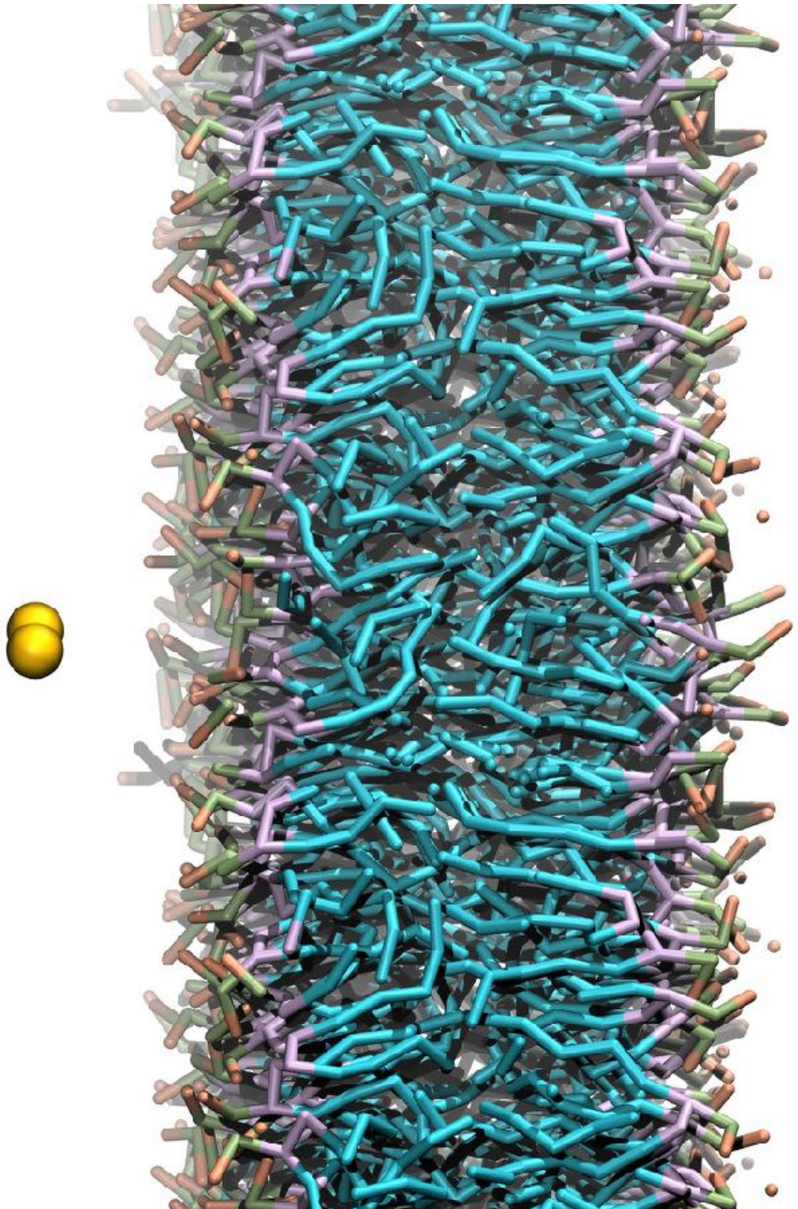


Intermolecular interactions across chemical space

High-throughput thermodynamics

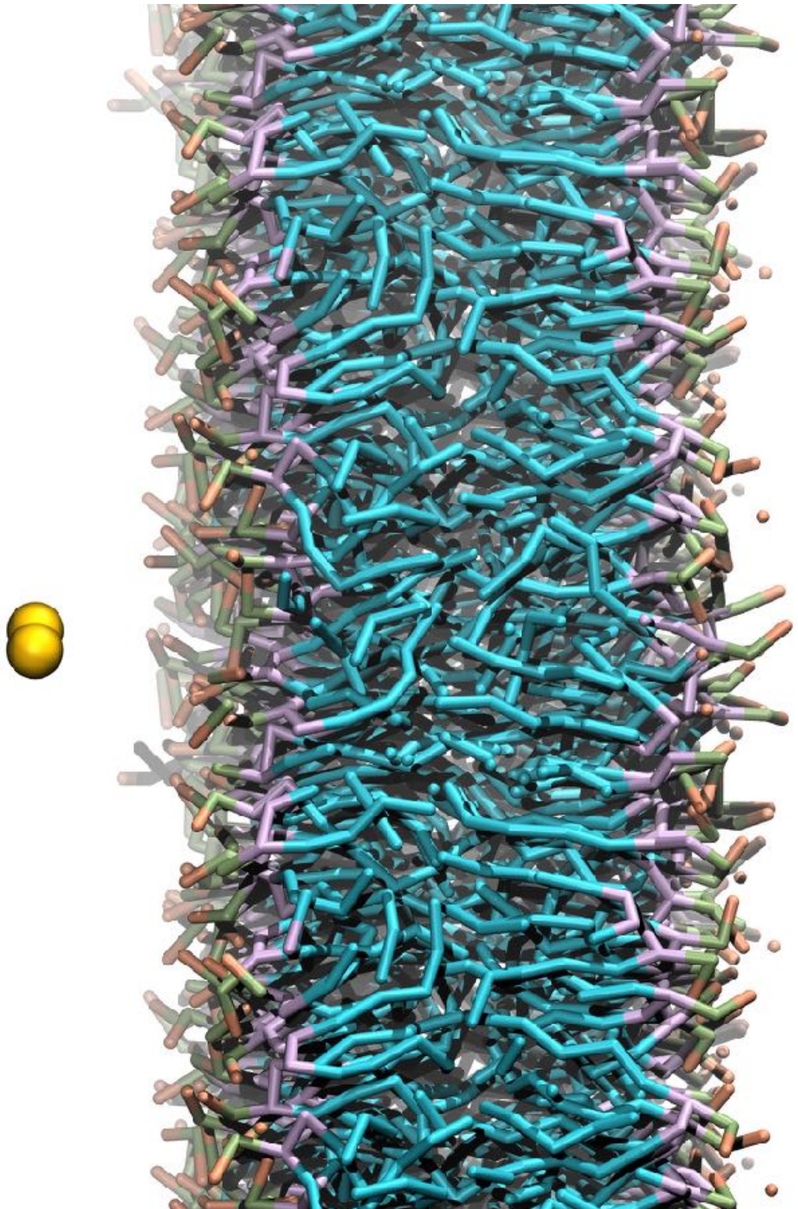


Drug permeability



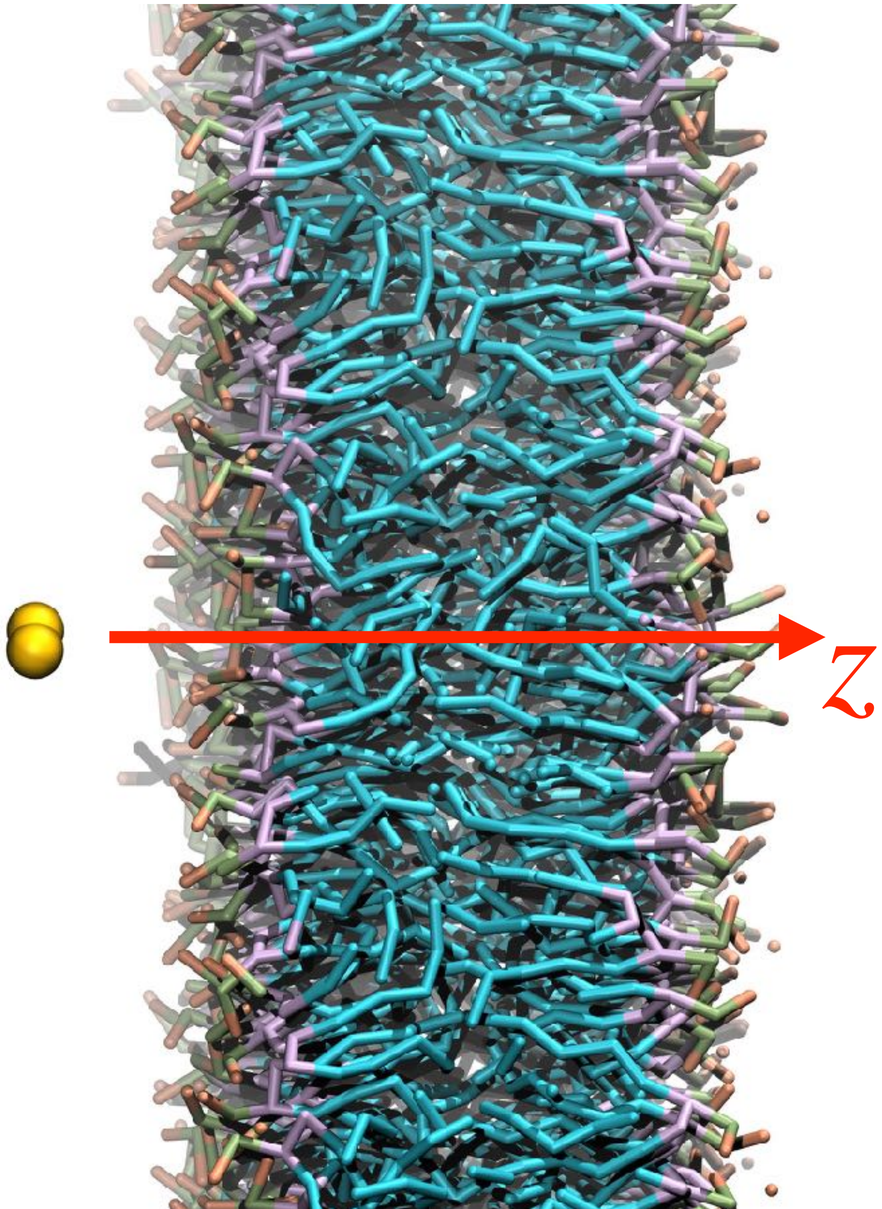
Drug permeability

Flux of drug permeation across a lipid membrane



Drug permeability

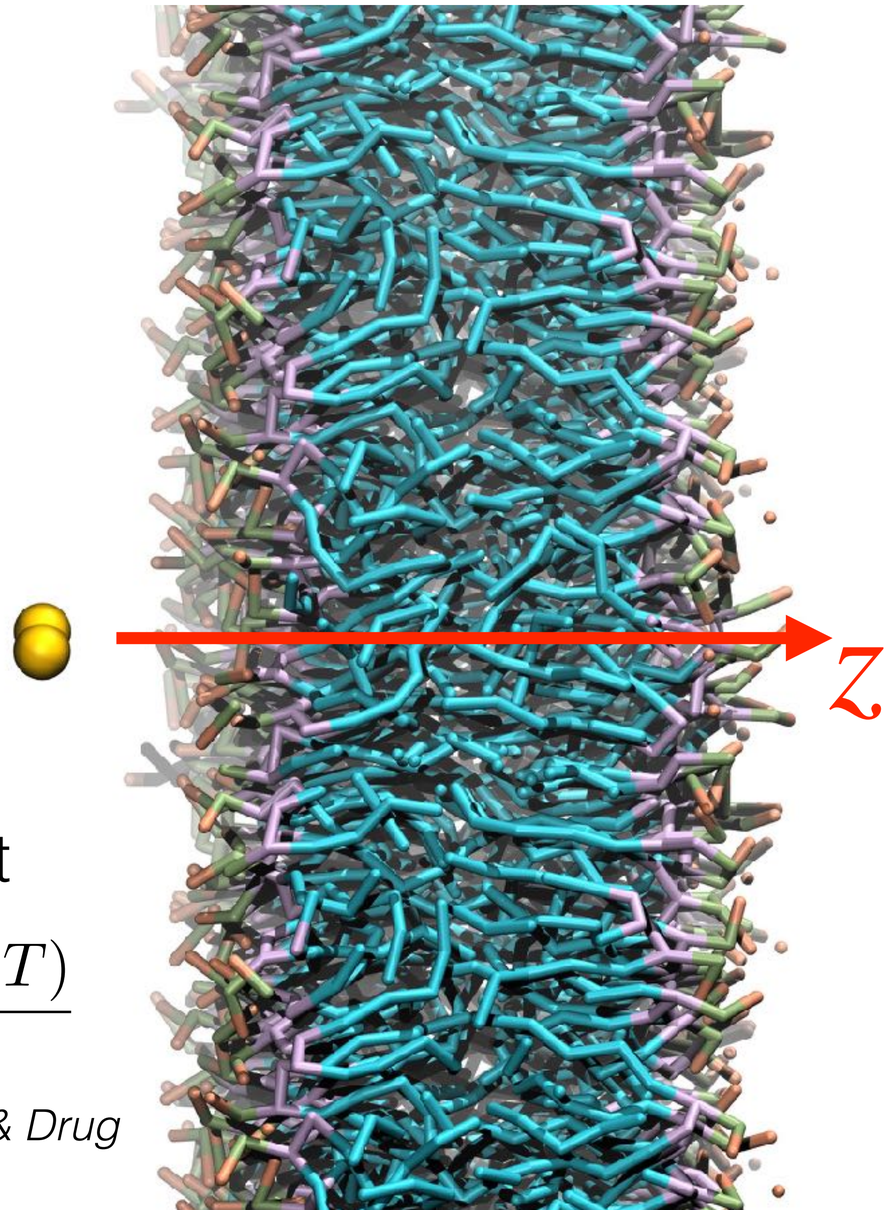
Flux of drug permeation across a lipid membrane



Drug permeability

Flux of drug permeation across a lipid membrane

Fokker-Planck; Smoluchoswki



Permeability coefficient

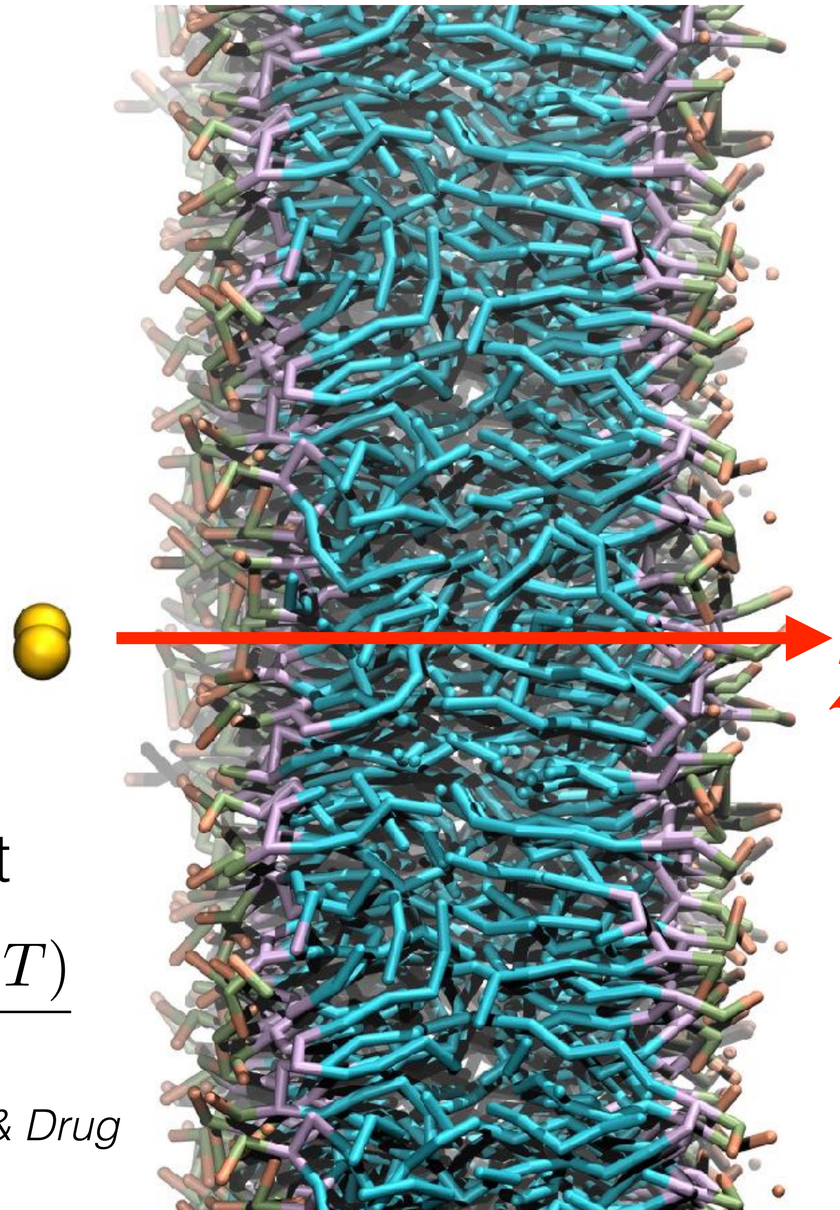
$$P^{-1} = \int \mathrm{d}z \frac{\exp\left(\Delta G(z)/k_{\mathrm{B}}T\right)}{D_z(z)}$$

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

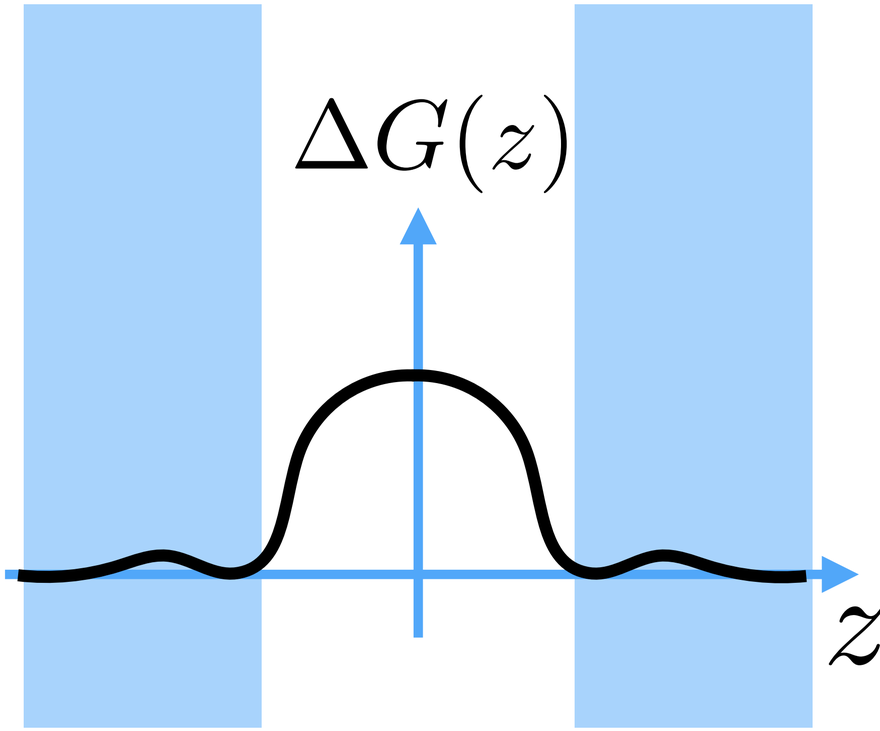
Drug permeability

Flux of drug permeation across a lipid membrane

Fokker-Planck; Smoluchoswki



Potential of mean force



Permeability coefficient

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

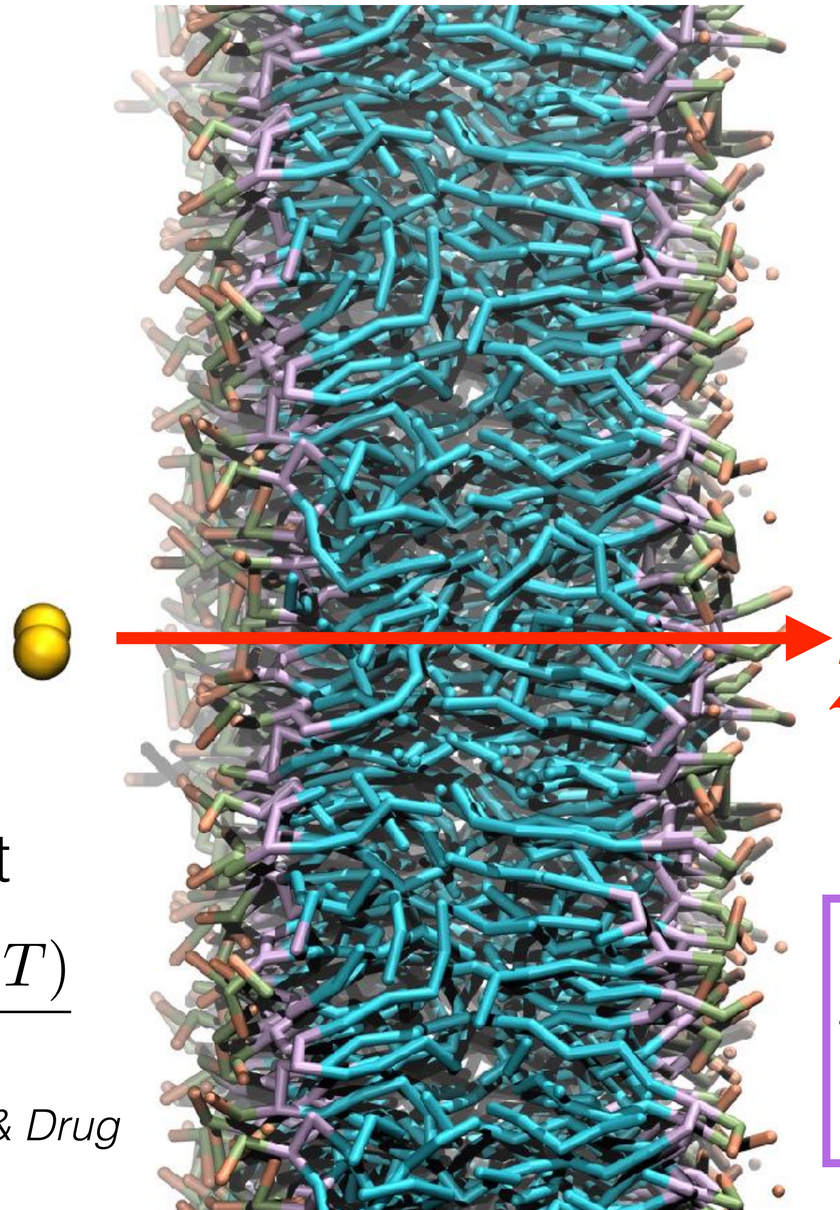
Swift & Amaro, *Chem Biol & Drug Design* **81** (2013)

THEORY GROUP

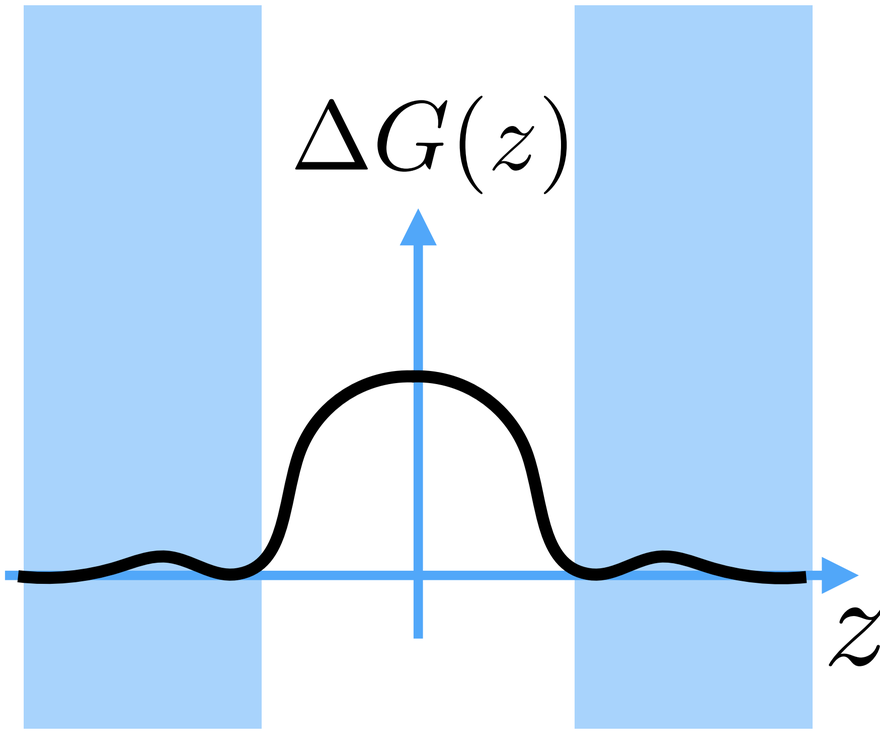
Drug permeability

Flux of drug permeation across a lipid membrane

Fokker-Planck; Smoluchoswki



Potential of mean force



Permeability coefficient

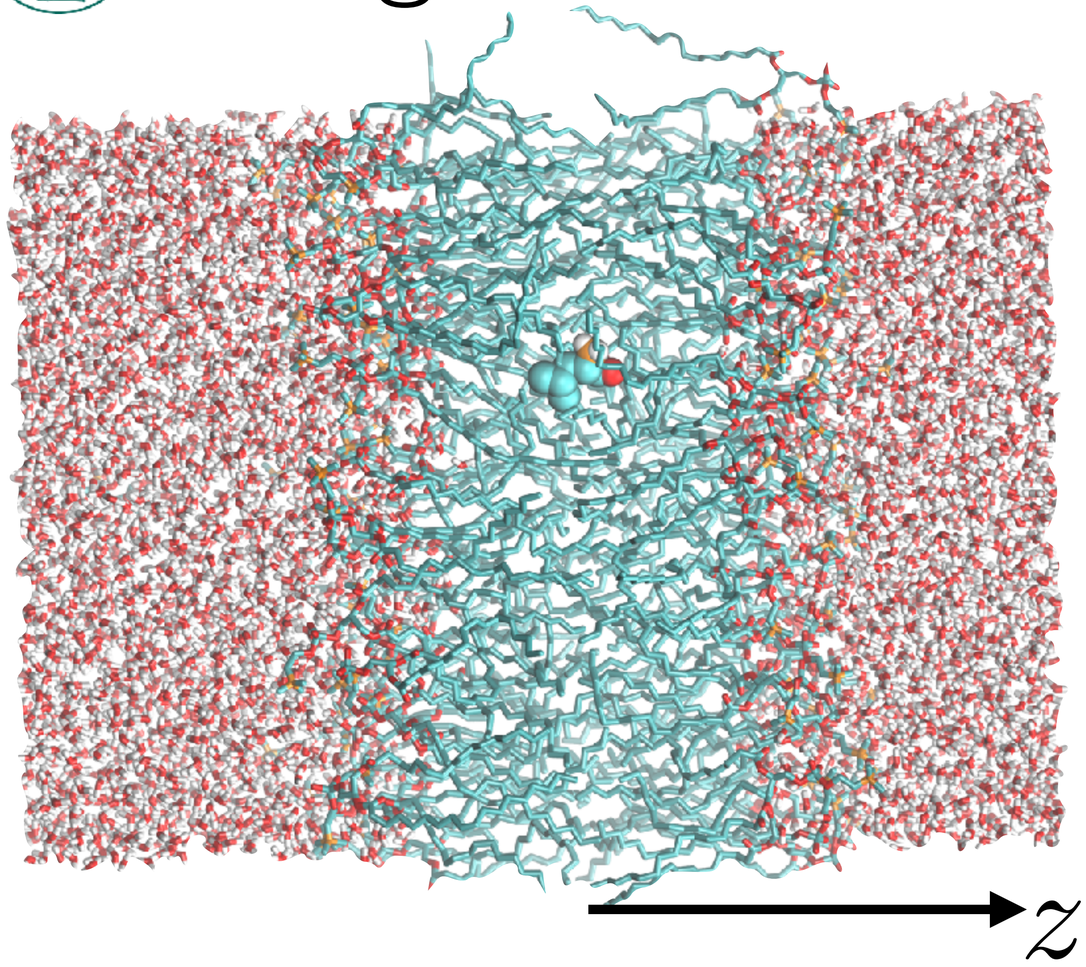
$$P^{-1} = \int \mathrm{d}z \frac{\exp\left(\Delta G(z)/k_\mathrm{B}T\right)}{D_z(z)}$$

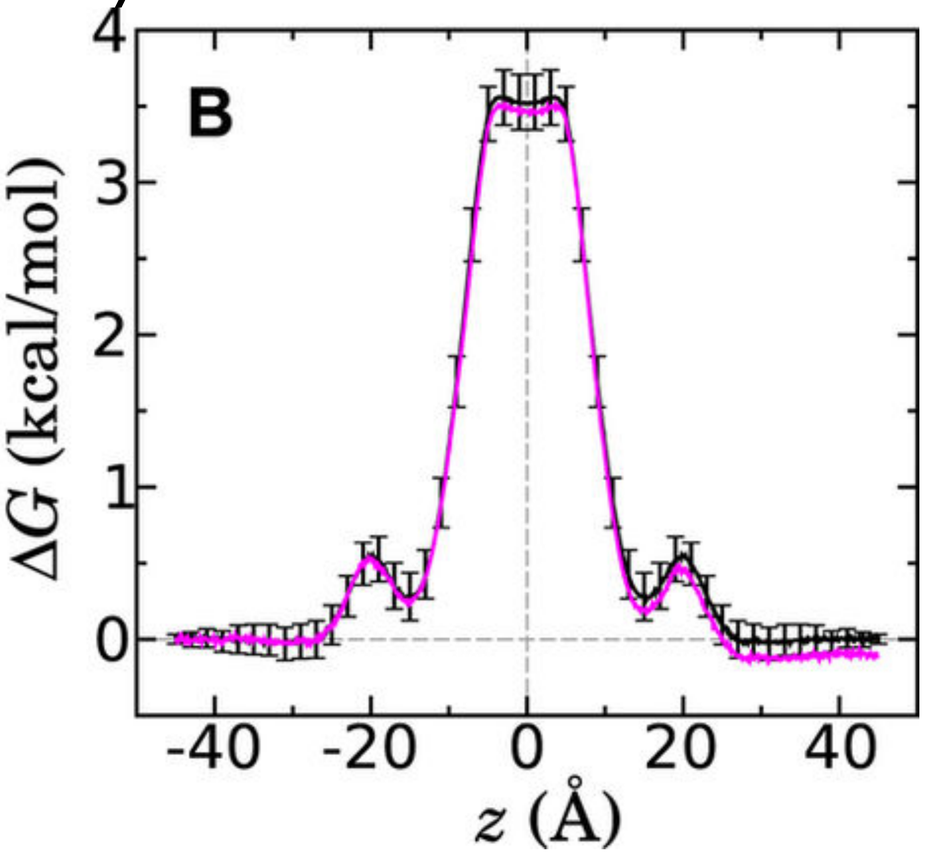
Swift & Amaro, *Chem Biol & Drug Design* **81** (2013)

Sample the potential of mean force from computer simulations

THEORY GROUP

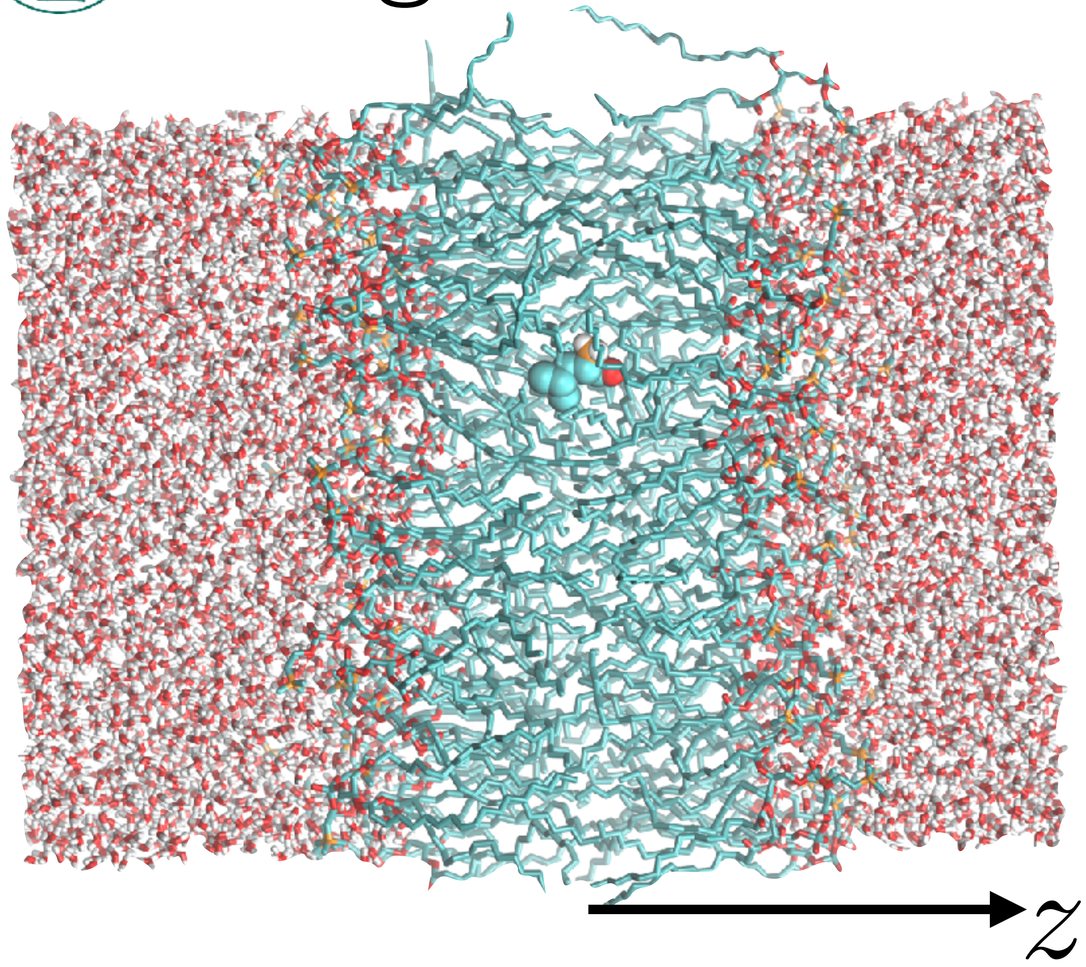
Drug-membrane thermodynamics *in silico*

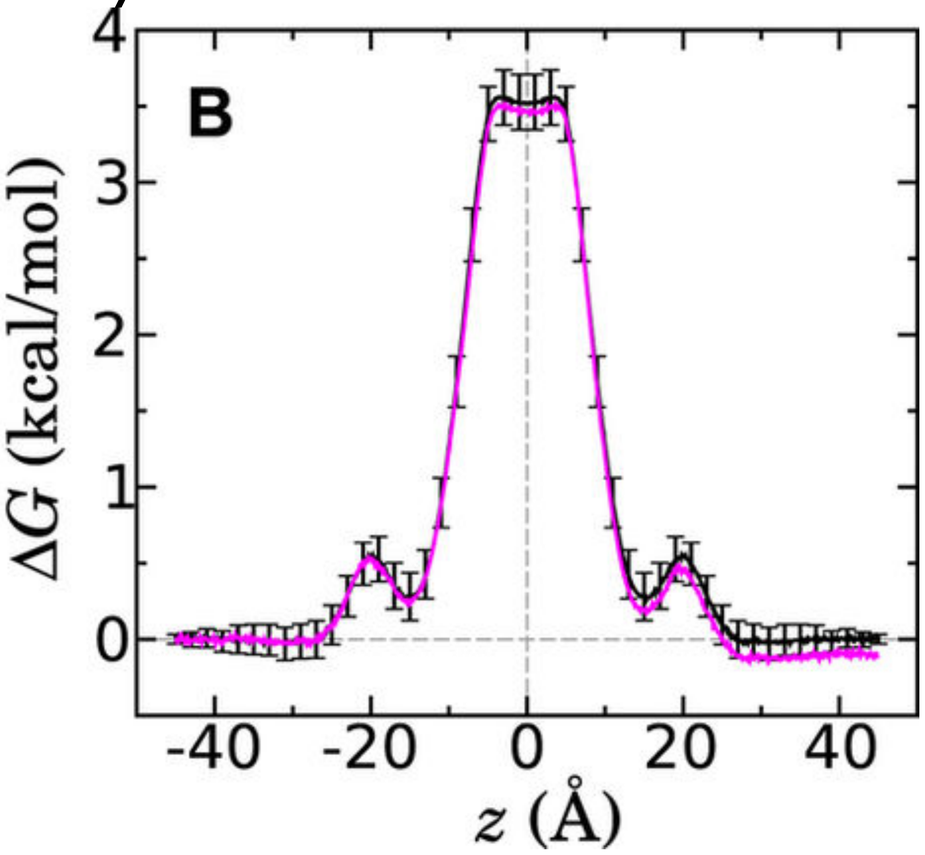




Chipot and Comer, *Scientific Reports* (2016)

Drug-membrane thermodynamics *in silico*

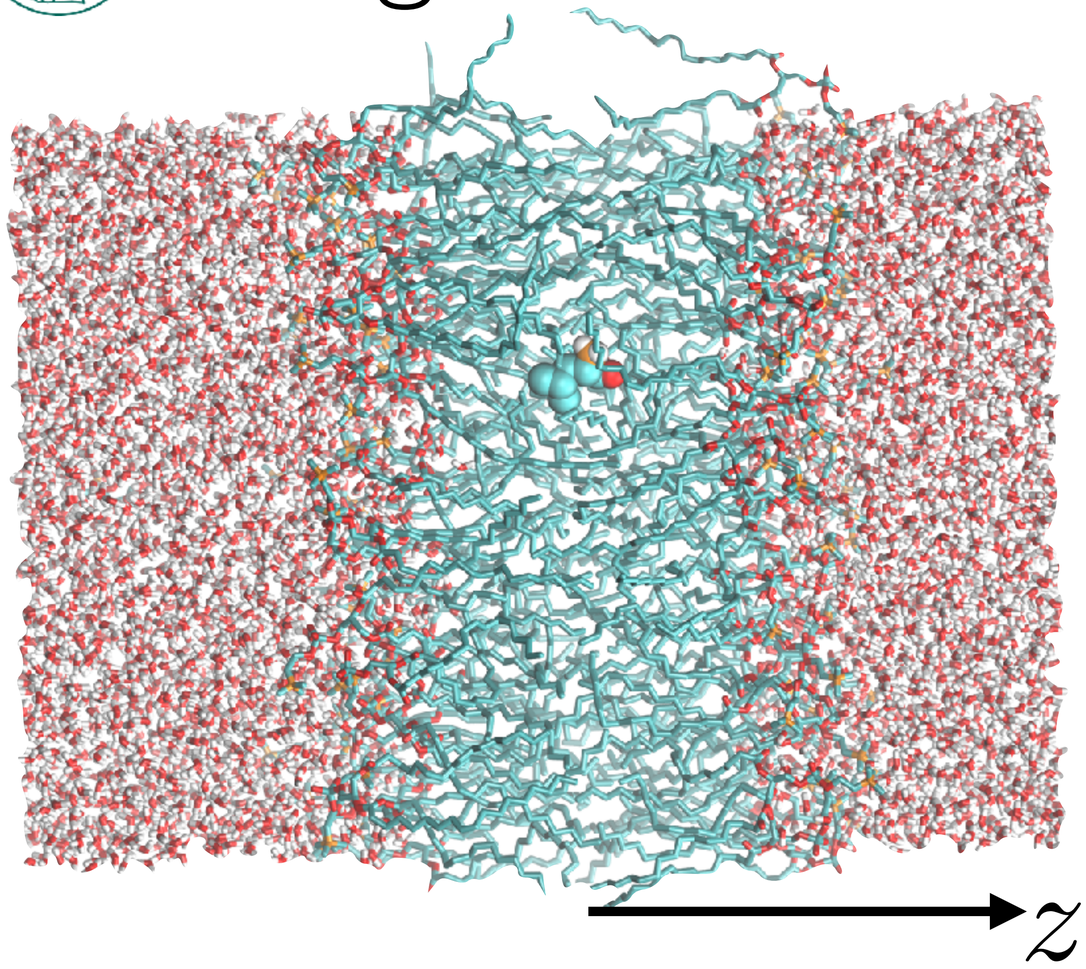




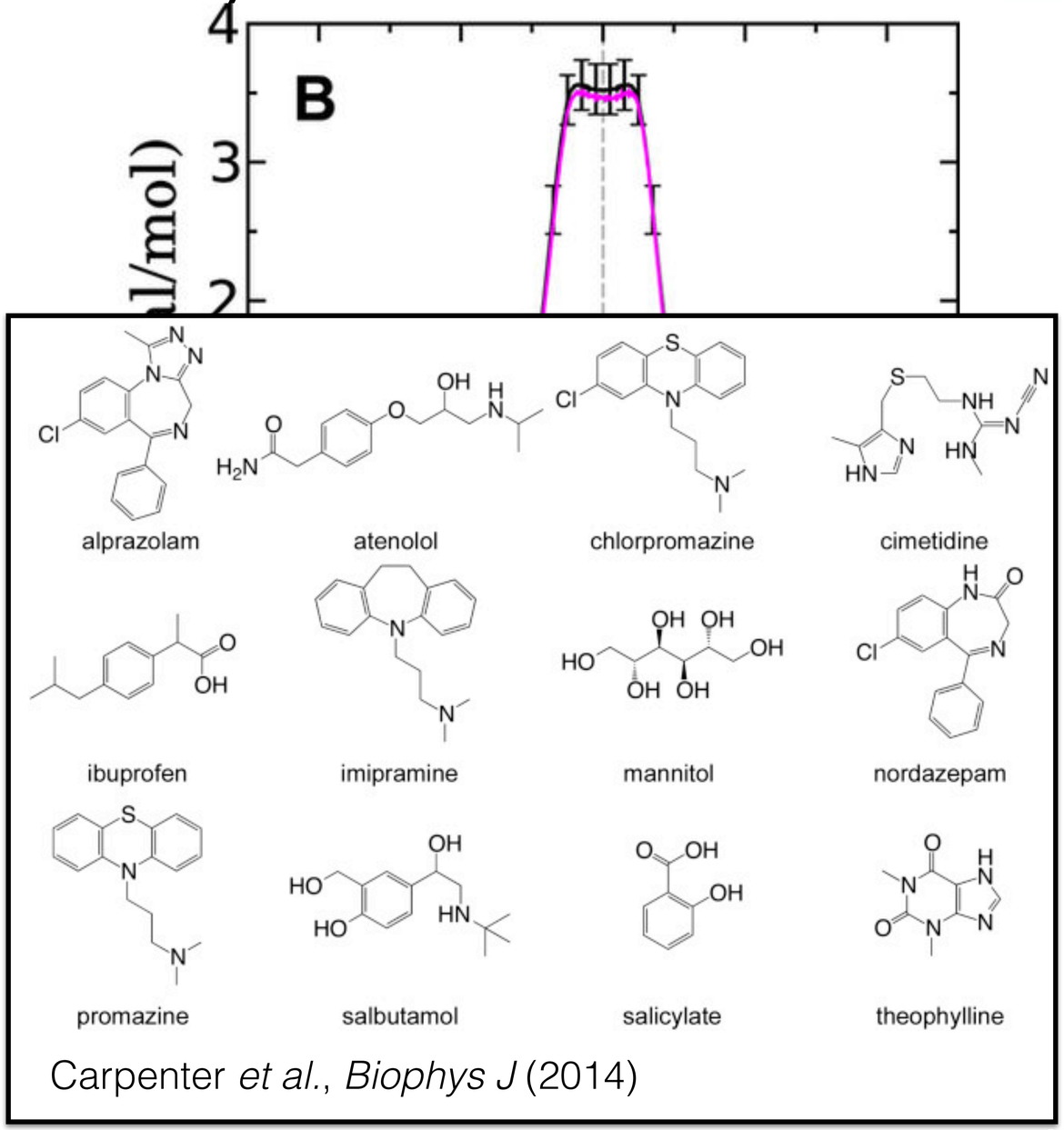
Chipot and Comer, *Scientific Reports* (2016)

- **Manual force-field parametrization**
- **Sampling: 100,000+ CPU-hours**

Drug-membrane thermodynamics *in silico*

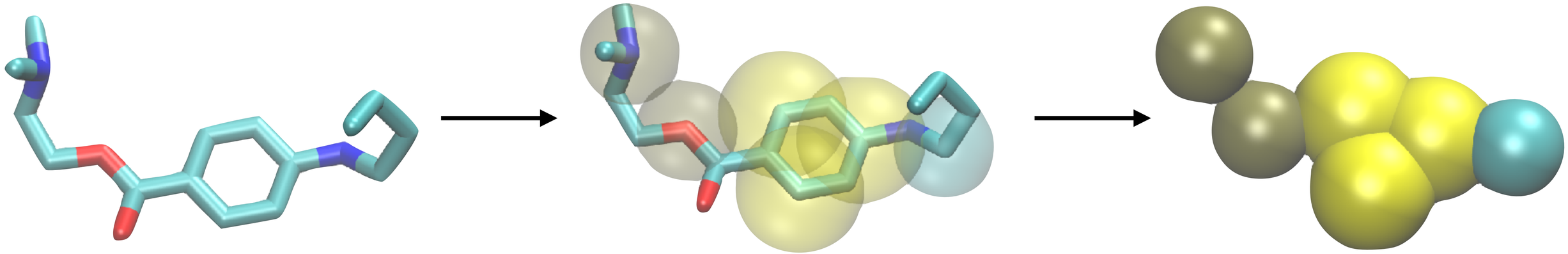


- **Manual force-field parametrization**
- **Sampling: 100,000+ CPU-hours**

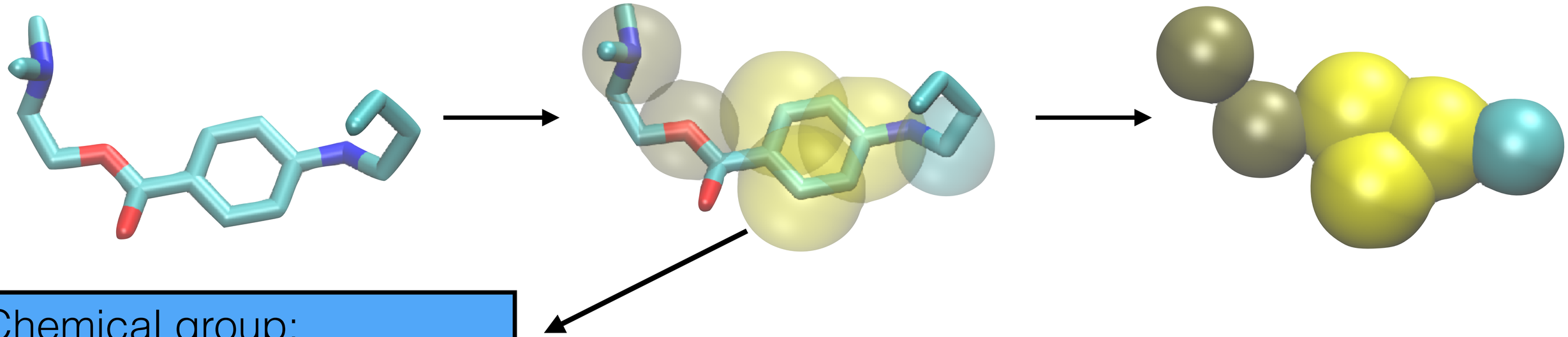


Carpenter *et al.*, *Biophys J* (2014)

Automated coarse-graining: Martini

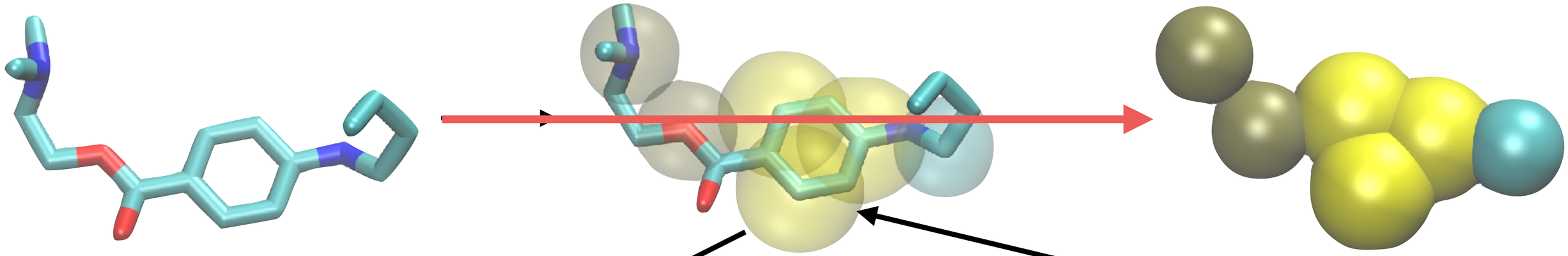


Automated coarse-graining: Martini

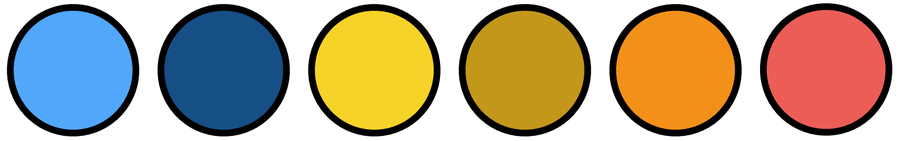


Chemical group:
- net charge
- hydrogen bond
- water/octanol partitioning

Automated coarse-graining: Martini



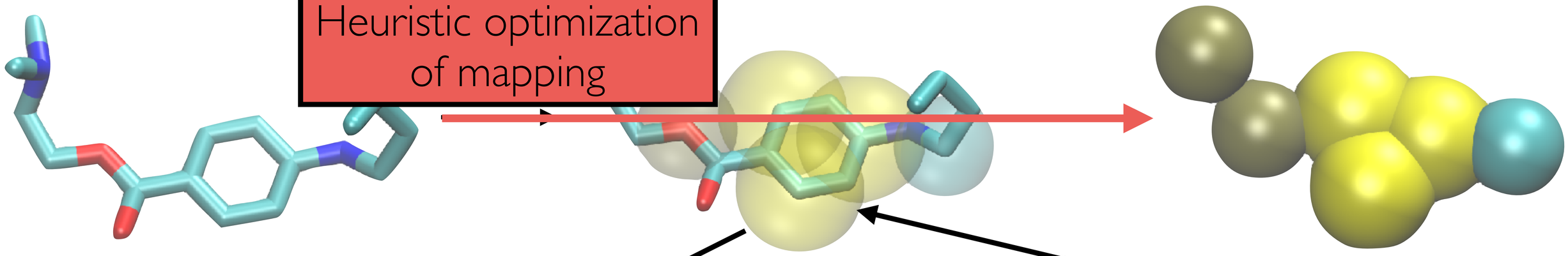
Chemical group:
- net charge
- hydrogen bond
- water/octanol partitioning



18 bead types: *chemical fragments*

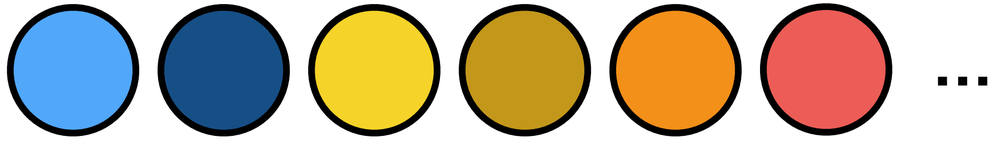
Marrink, Tieleman, Chem Soc Rev **42** (2013)
Periole, Marrink, Biomolecular Simulations (2013)

Automated coarse-graining: Martini



Heuristic optimization of mapping

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



18 bead types: *chemical fragments*

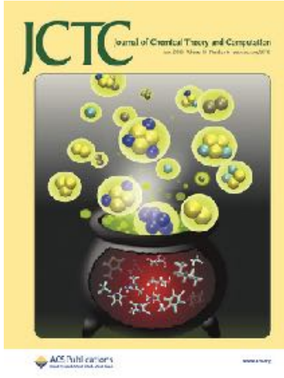
Marrink, Tieleman, Chem Soc Rev **42** (2013)
Periole, Marrink, Biomolecular Simulations (2013)

neural network trained on experimental data

Tetko et al., *J Chem Inf Comput Sci* **41** (2001)

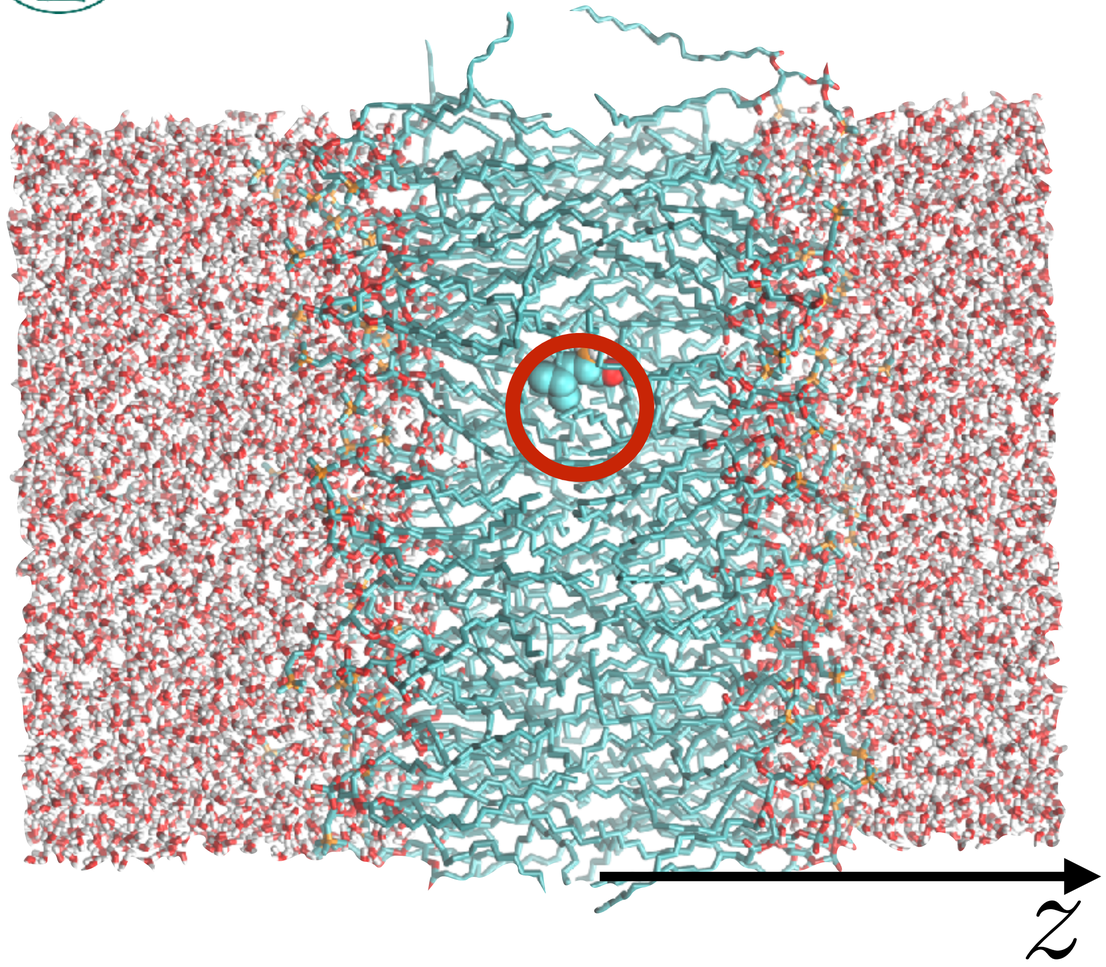
Automated parametrization for small molecules

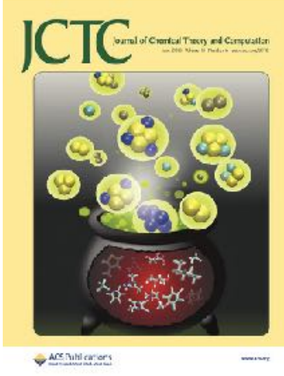
Bereau & Kremer, *J Chem Theory Comput* **11** (2015)



THEORY GROUP

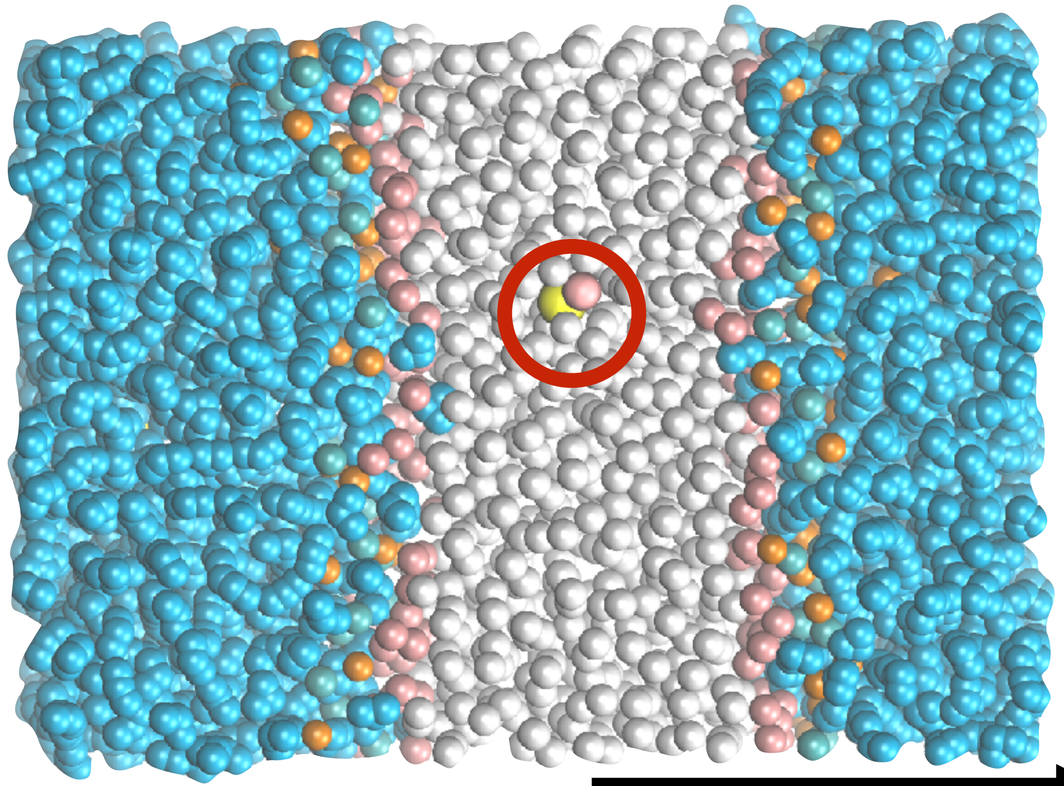
Solute insertion in the membrane

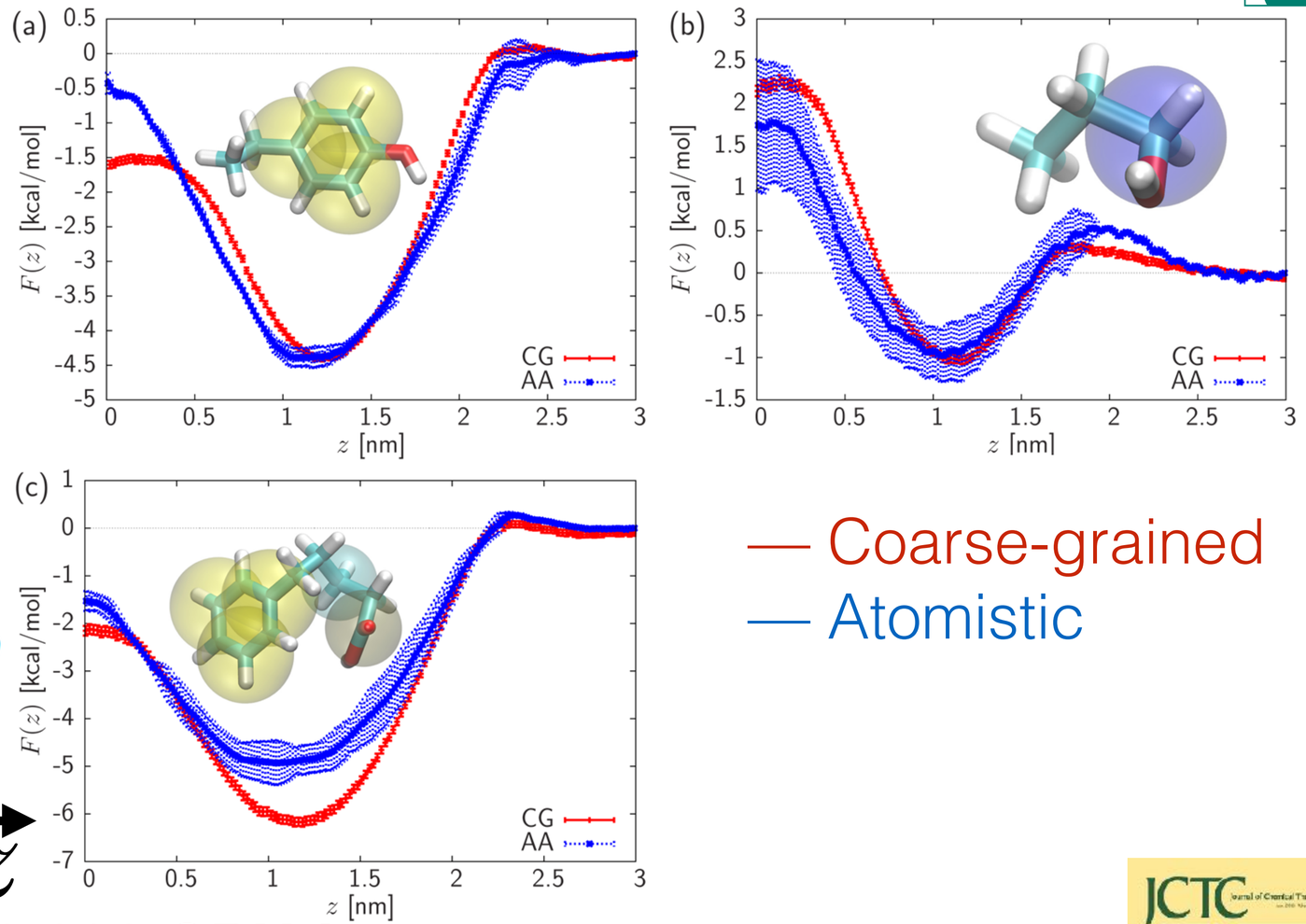




Jakobtorweihen et al. *J Chem Phys* **141** (2014); Bereau & Kremer, *J Chem Theory Comput*, **11** (2015)

Solute insertion in the membrane

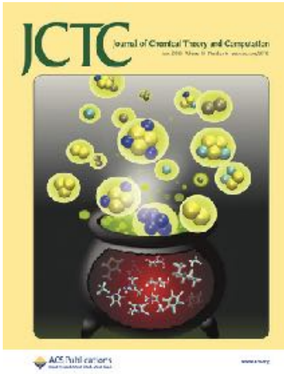




— Coarse-grained
— Atomistic

Atomistic simulations: >100,000 CPU-hours

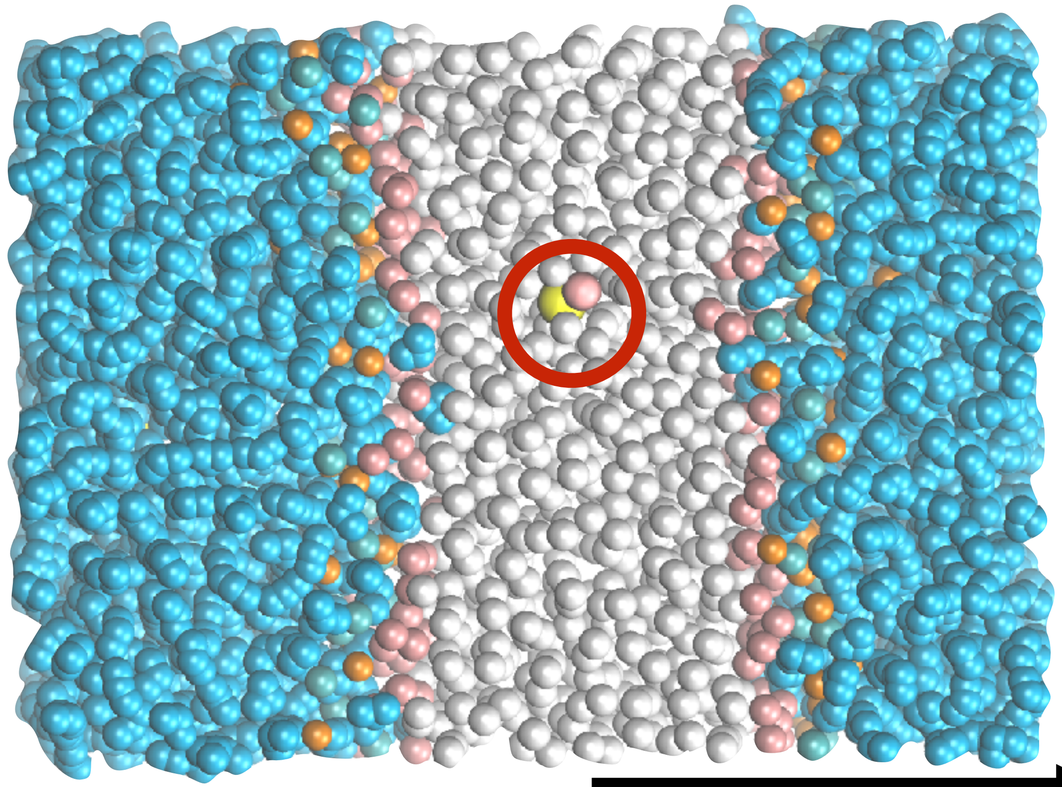
Coarse-grained simulations: ~100 CPU-hours

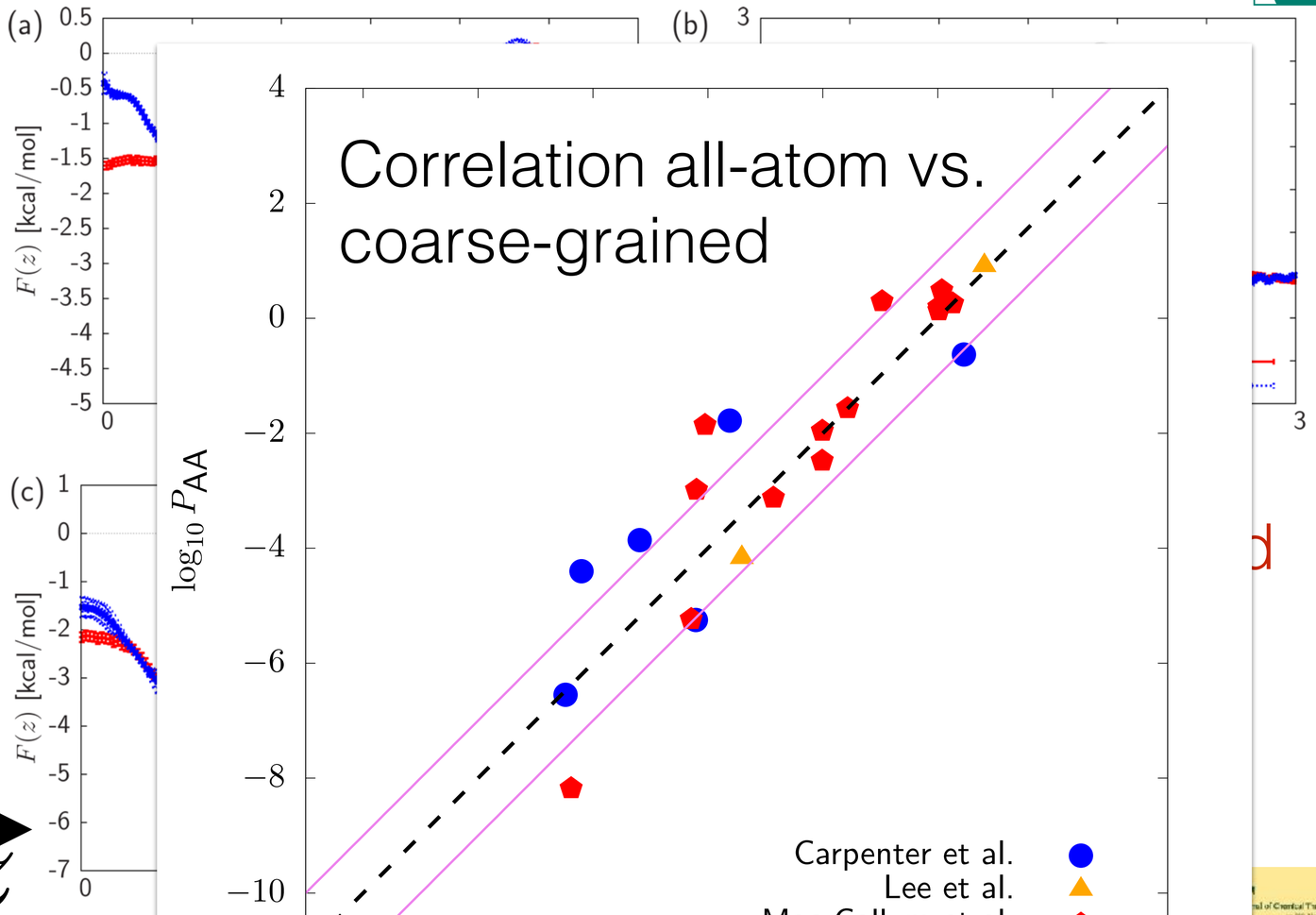


THEORY GROUP

Jakobtorweihen et al. *J Chem Phys* **141** (2014); Bereau & Kremer, *J Chem Theory Comput*, **11** (2015)

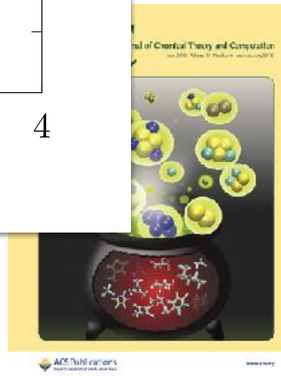
Solute insertion in the membrane





Atomistic simulations: >100,000

Coarse-grained simulations: ~100



Jakobtorweihen et al. *J Chem Phys* **141** (2014); Bereau & Kremer, *J Chem Theory Comput*, **11** (2015)

Construct reduced chemical space

We don’t focus on specific compounds, instead **explore chemical diversity**

Molecular weight:

1 ... (x14)

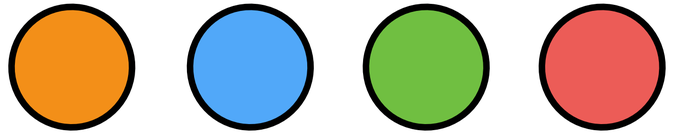
2 ... (x105)

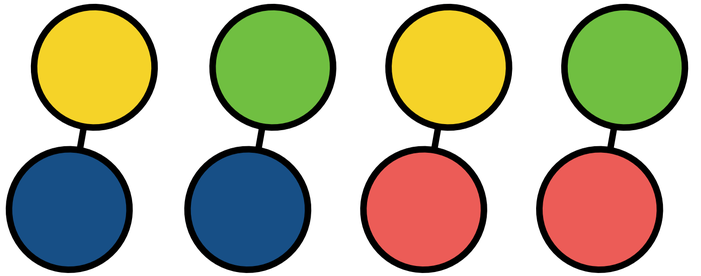
~ 30-160 Da

Construct reduced chemical space

We don’t focus on specific compounds, instead **explore chemical diversity**

Molecular weight:

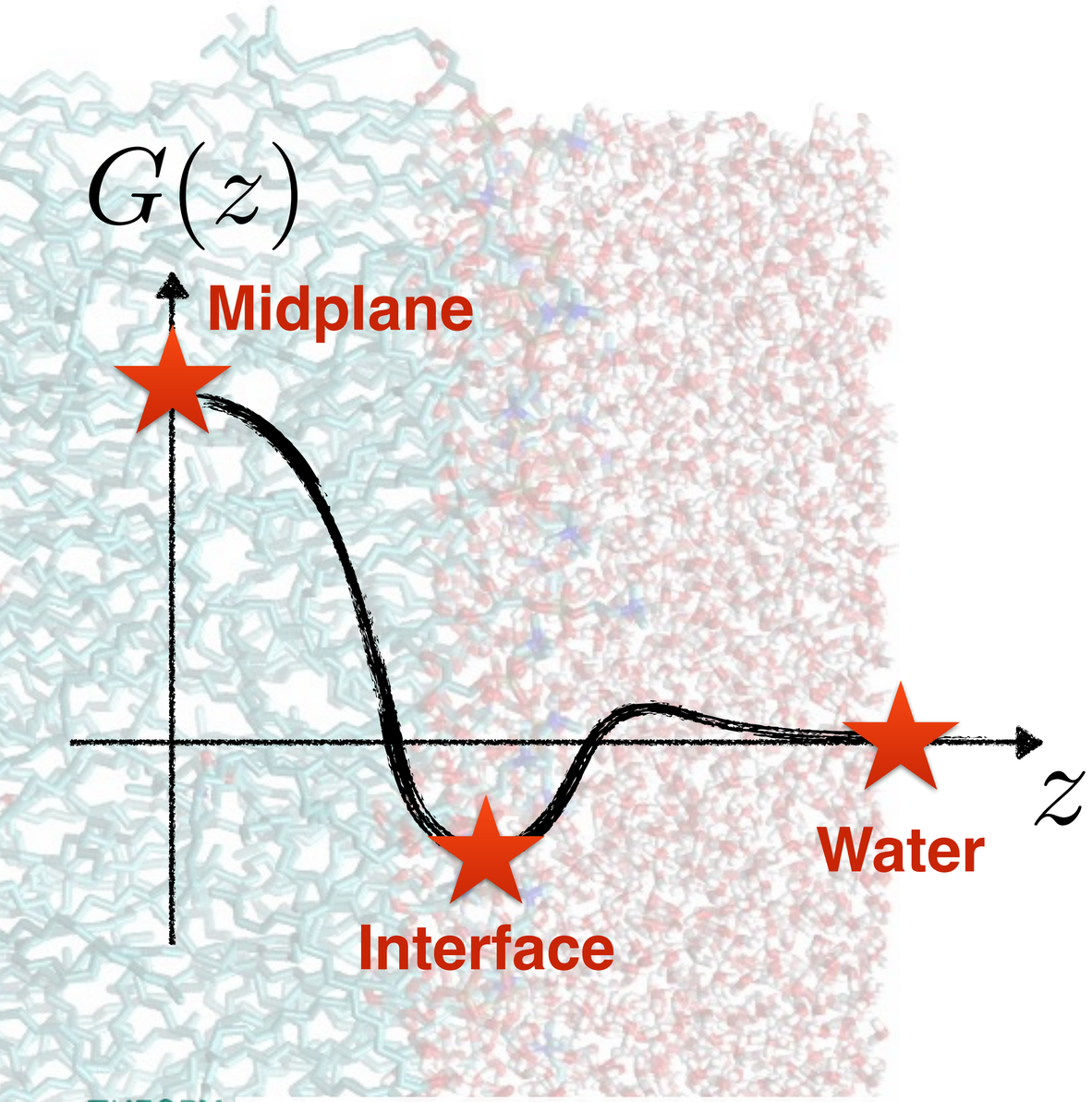
1  ... (x14)

2  ... (x105)

~ 30-160 Da

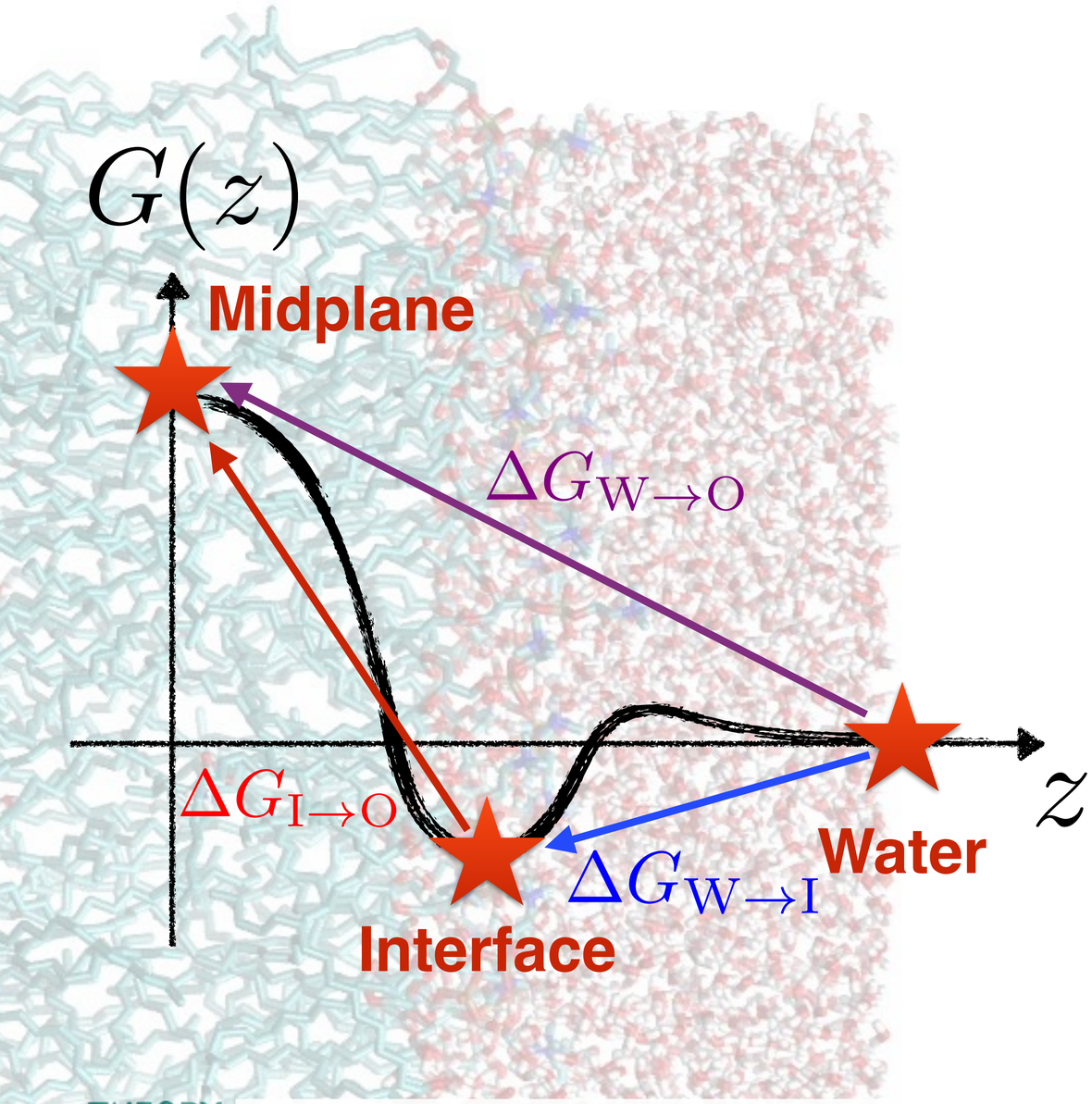
CG: combinatorial explosion is reduced

Identifying simple thermodynamic relations



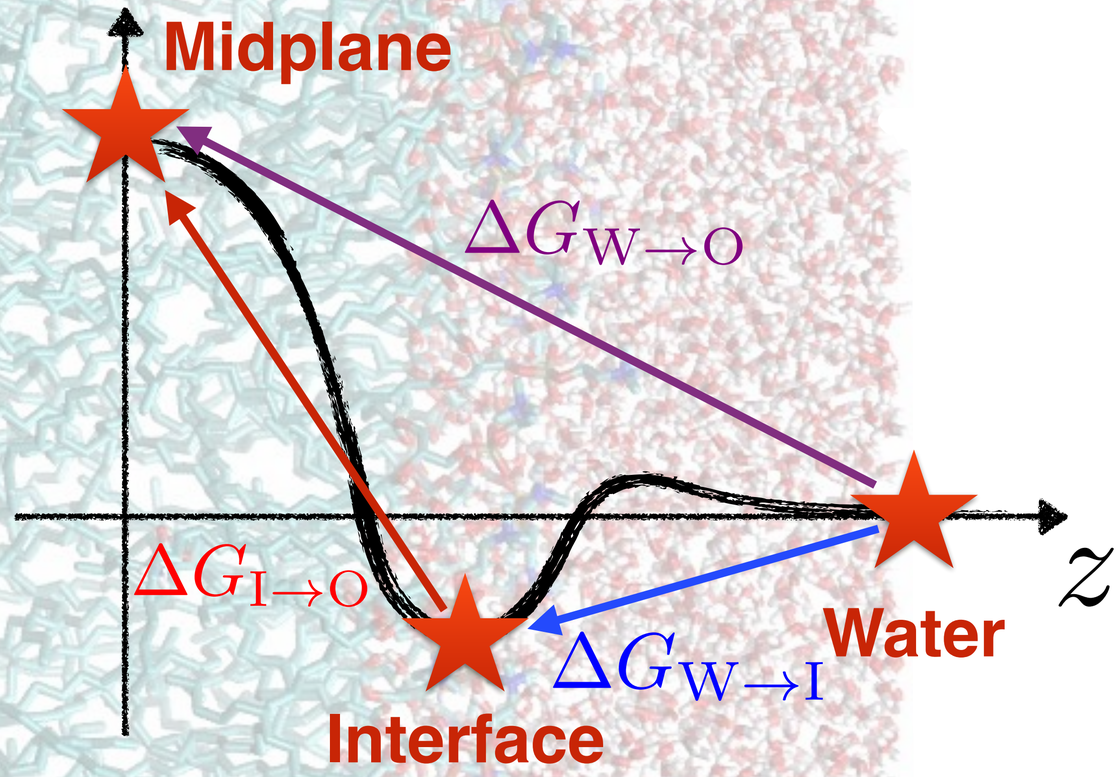
Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

Identifying simple thermodynamic relations



Identifying simple thermodynamic relations

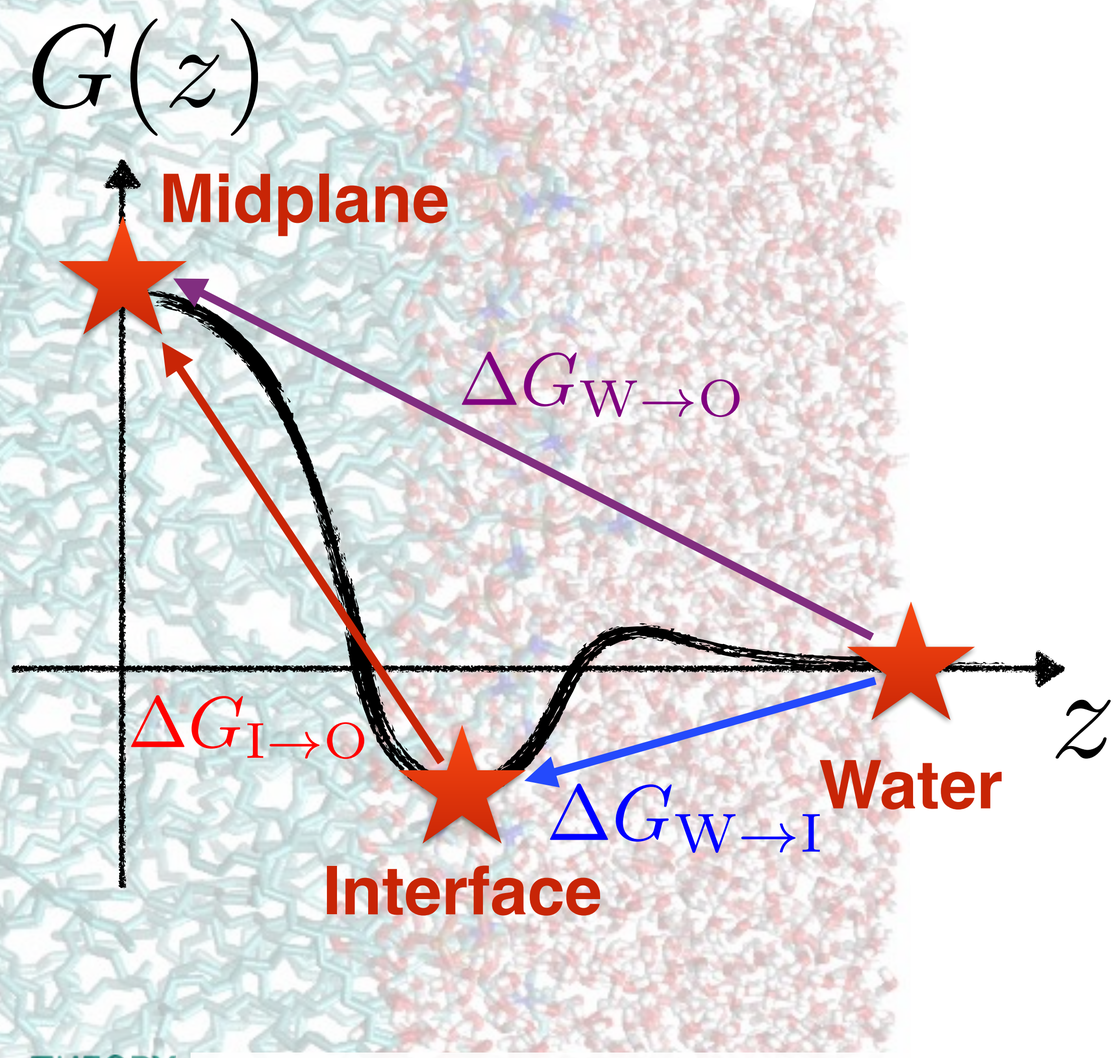
$$\Delta G_{\mathrm{W}\to\mathrm{I}} + \Delta G_{\mathrm{I}\to\mathrm{O}} - \Delta G_{\mathrm{W}\to\mathrm{O}} = 0$$

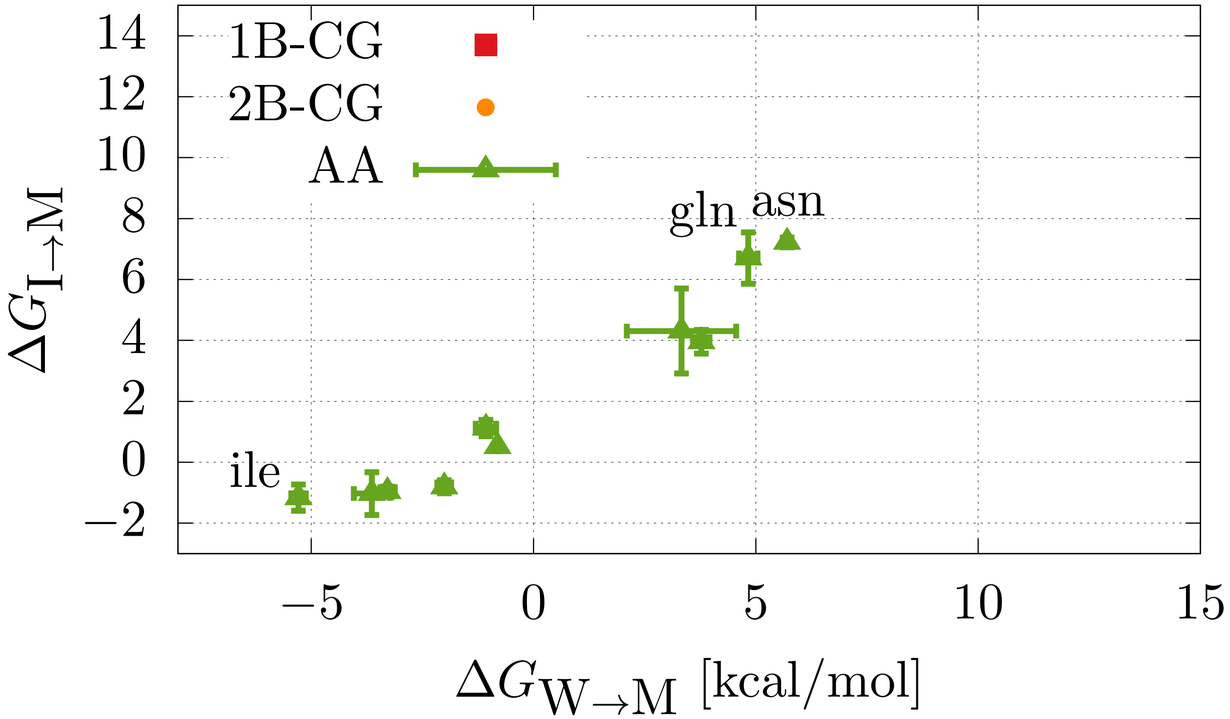


Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

Identifying simple thermodynamic relations

$$\Delta G_{\mathrm{W}\to\mathrm{I}} + \Delta G_{\mathrm{I}\to\mathrm{O}} - \Delta G_{\mathrm{W}\to\mathrm{O}} = 0$$

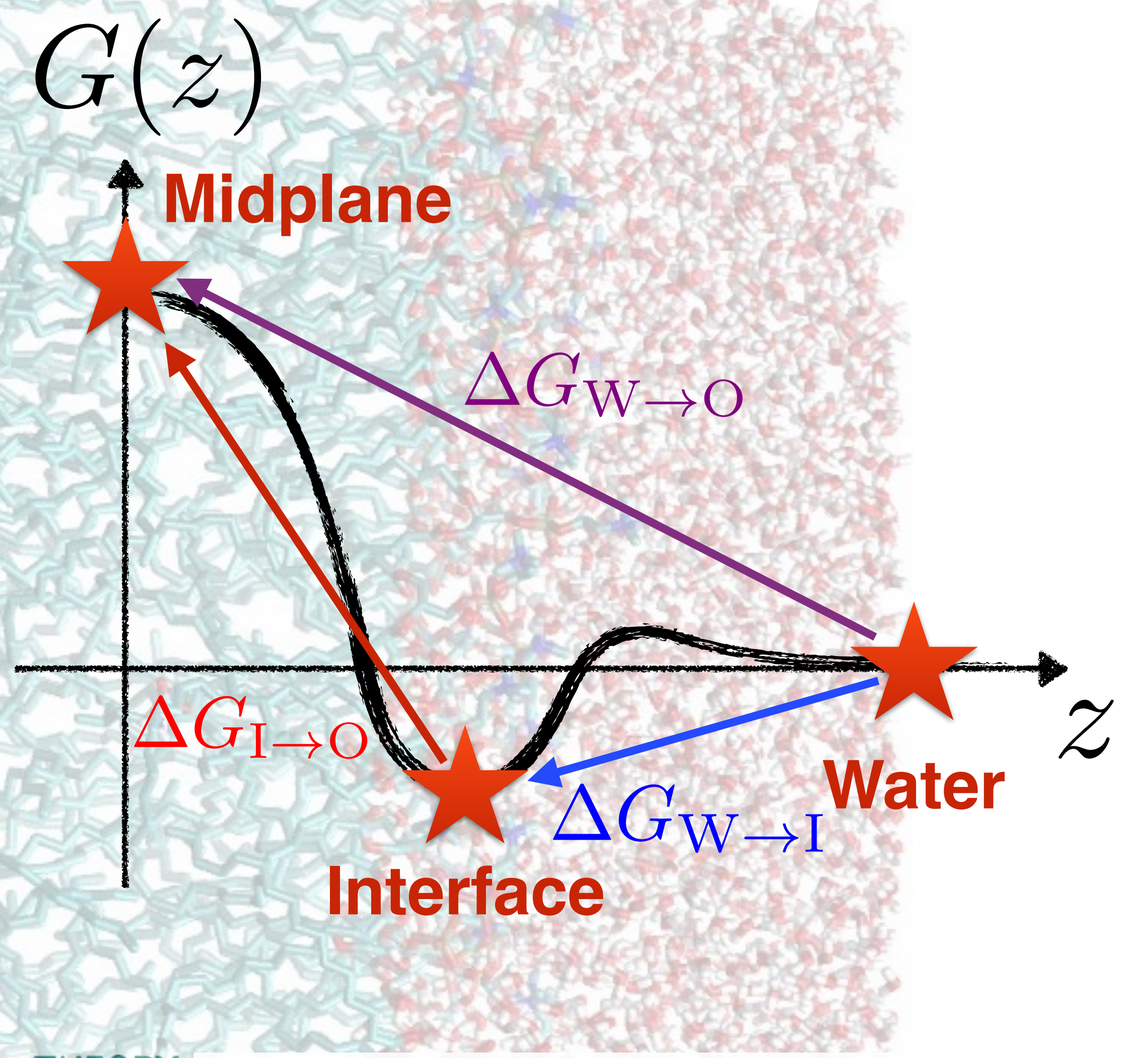


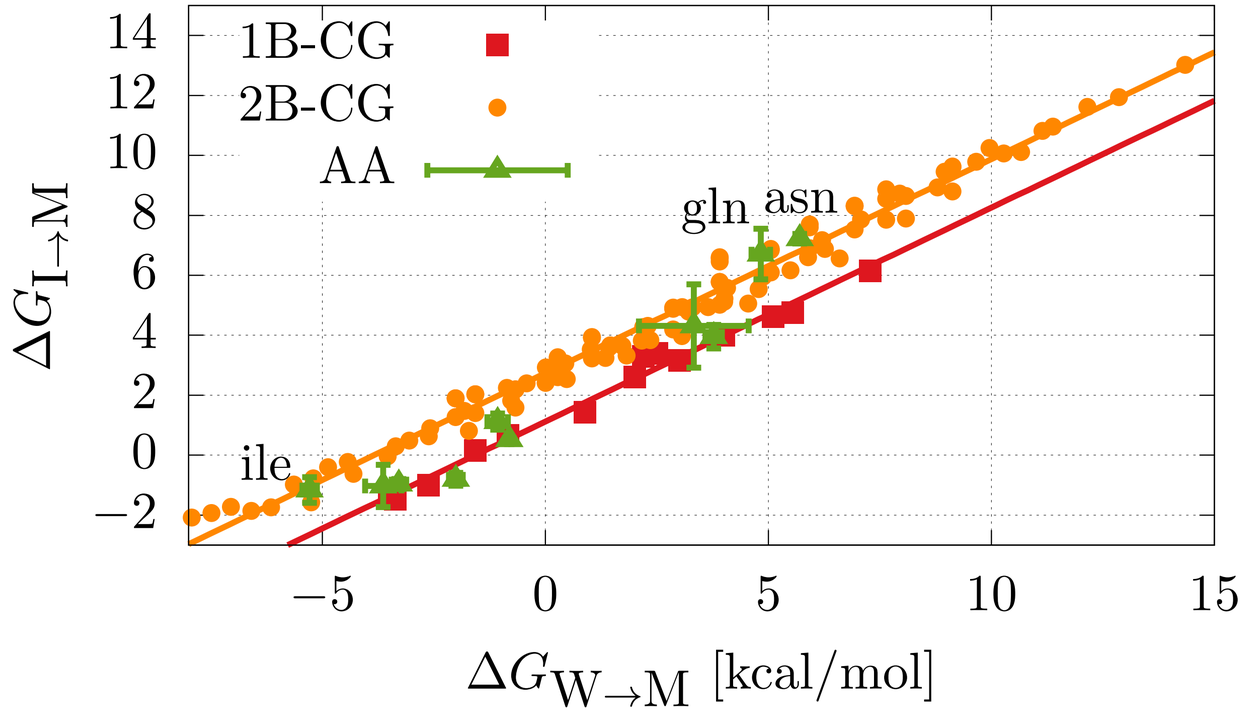


Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

Identifying simple thermodynamic relations

$$\Delta G_{\mathrm{W}\to\mathrm{I}} + \Delta G_{\mathrm{I}\to\mathrm{O}} - \Delta G_{\mathrm{W}\to\mathrm{O}} = 0$$

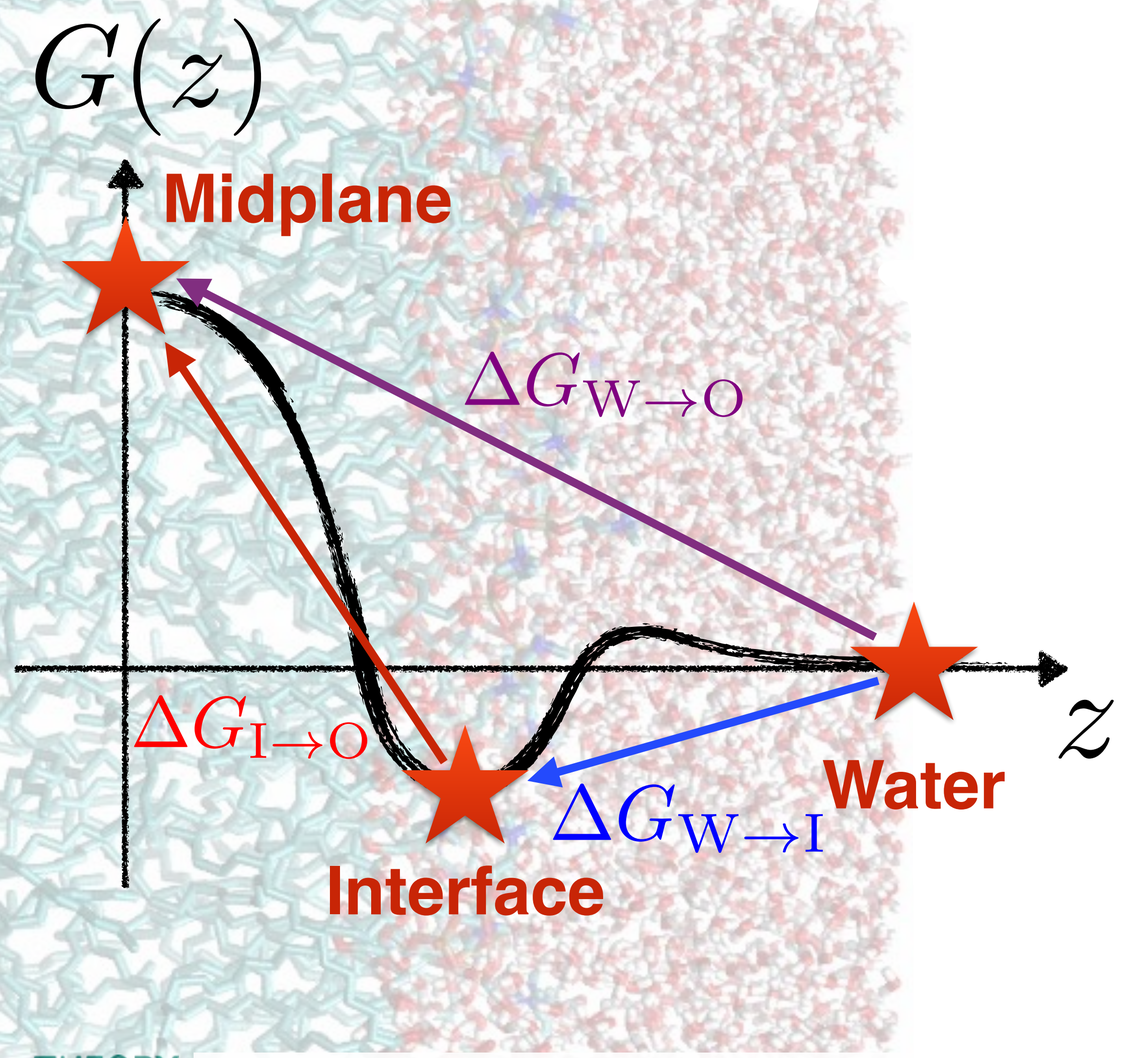


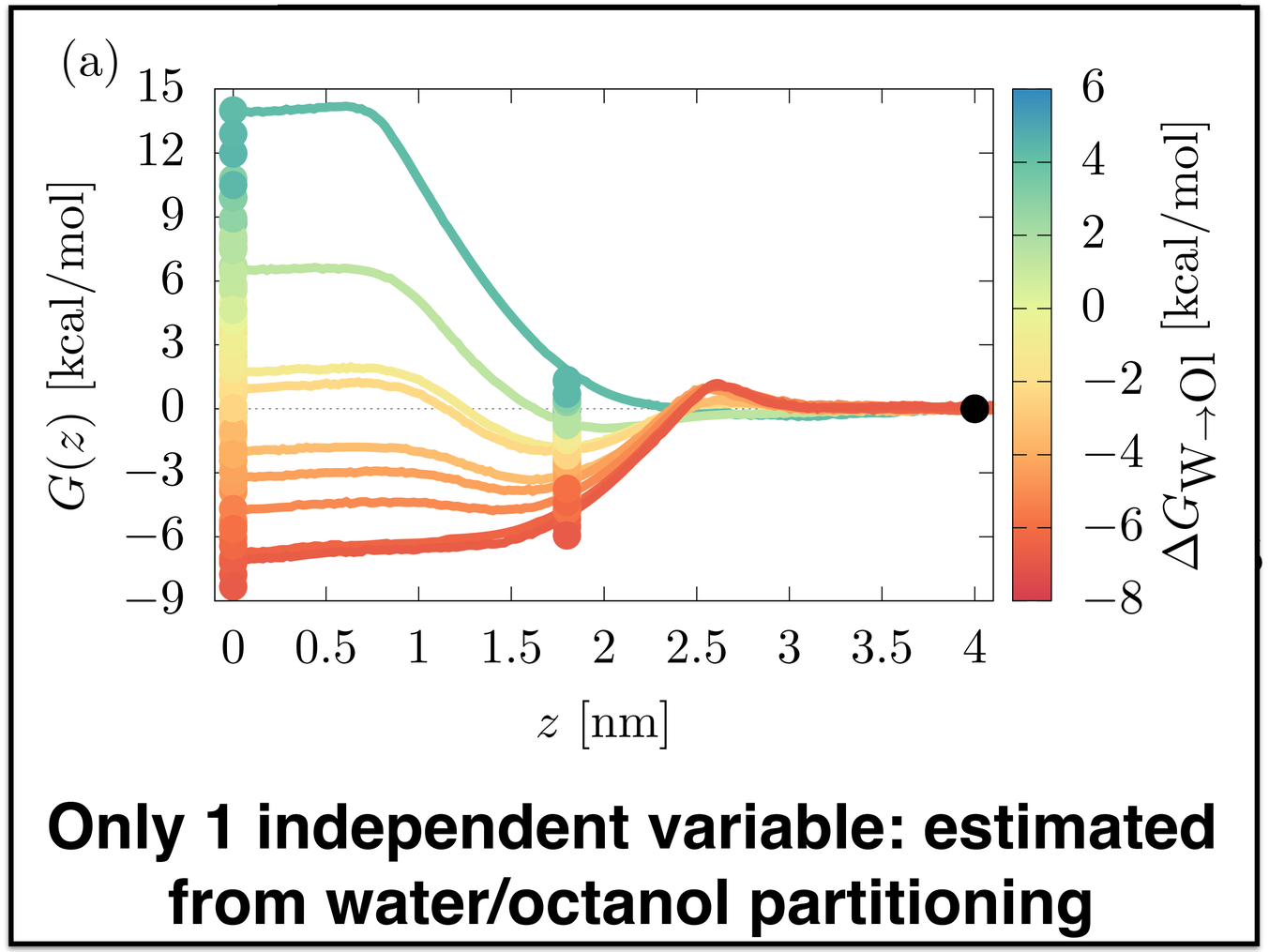


Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

Identifying simple thermodynamic relations

$$\Delta G_{\mathrm{W}\to\mathrm{I}} + \Delta G_{\mathrm{I}\to\mathrm{O}} - \Delta G_{\mathrm{W}\to\mathrm{O}} = 0$$

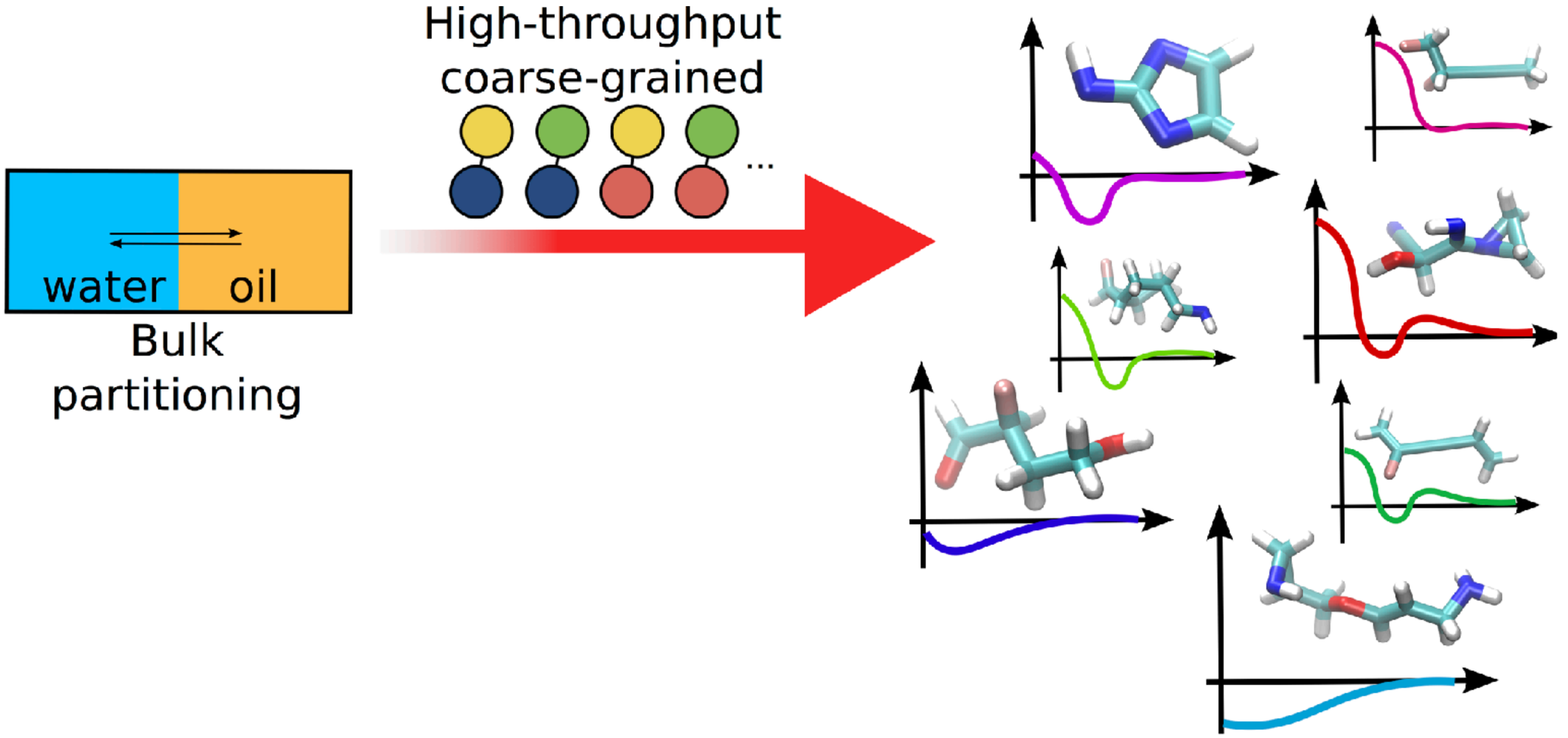




Only 1 independent variable: estimated from water/octanol partitioning

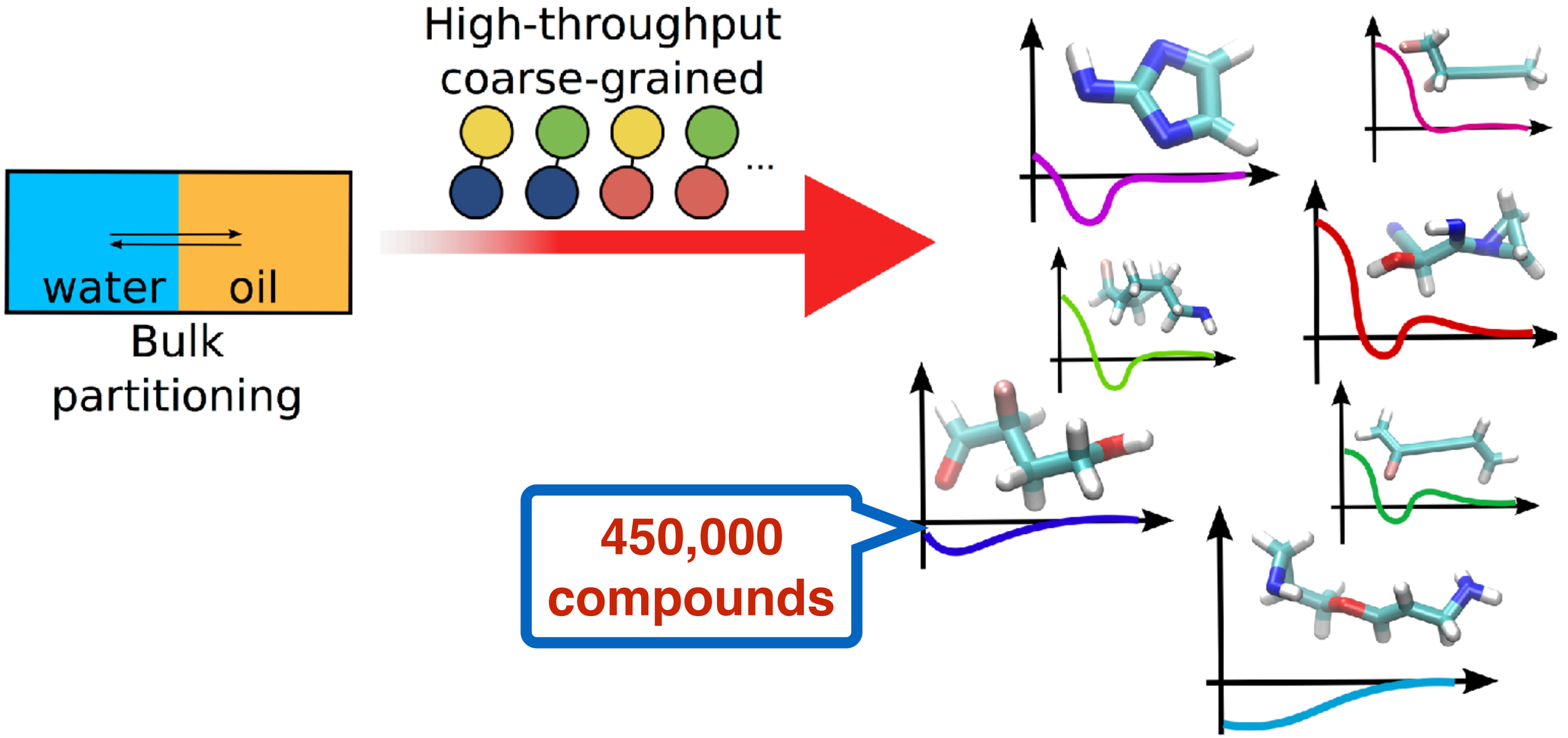
Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

PMF prediction from experimental value



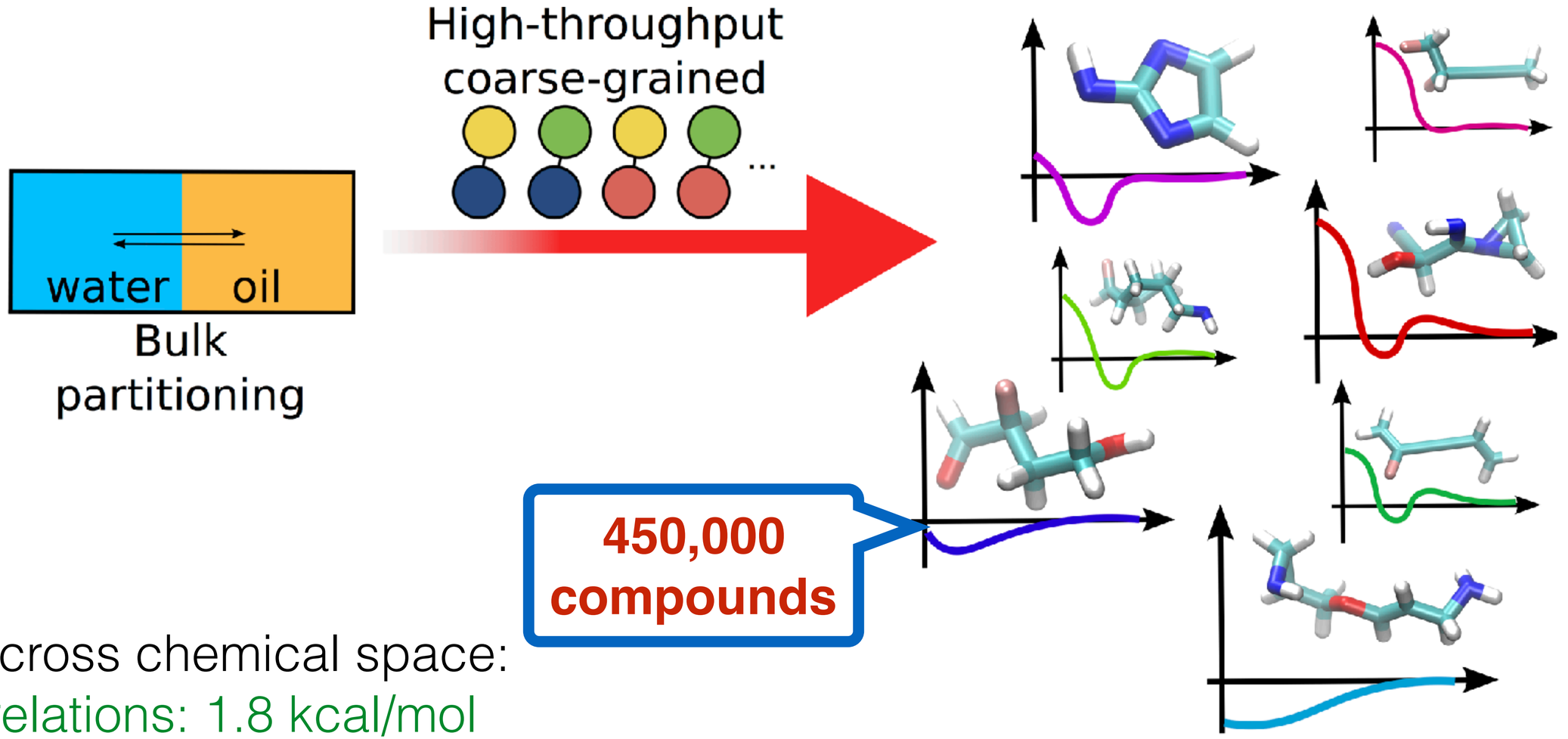
Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)

PMF prediction from experimental value



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

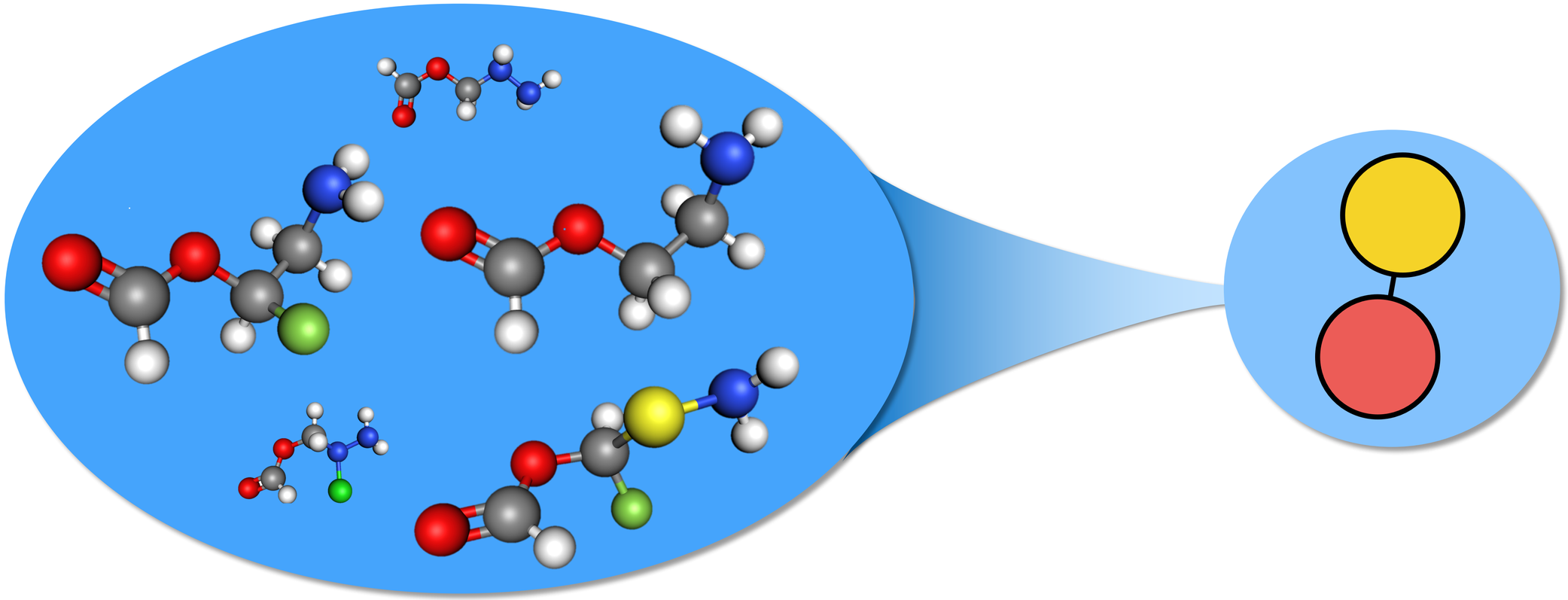
PMF prediction from experimental value



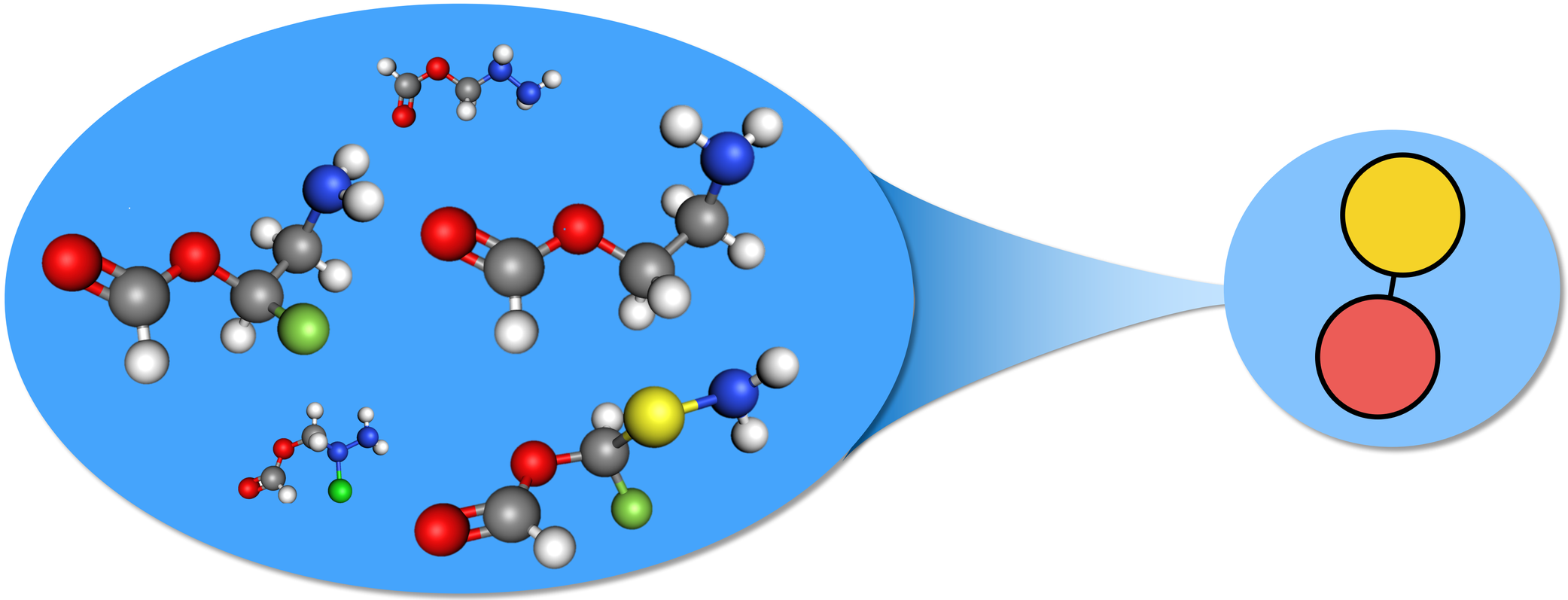
Error across chemical space:
linear relations: 1.8 kcal/mol
coarse-grained: 1.4 kcal/mol

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

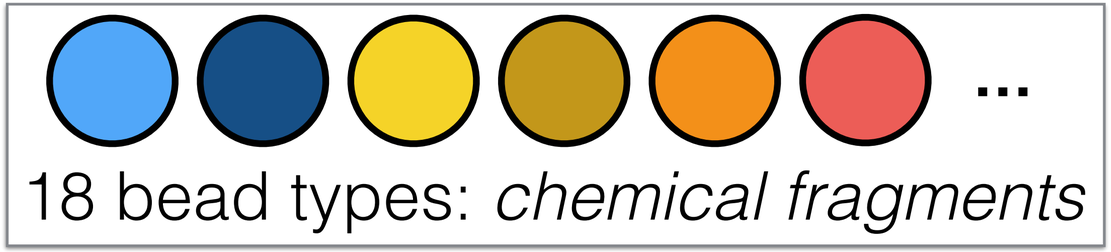
Coarse-graining reduces chemical space



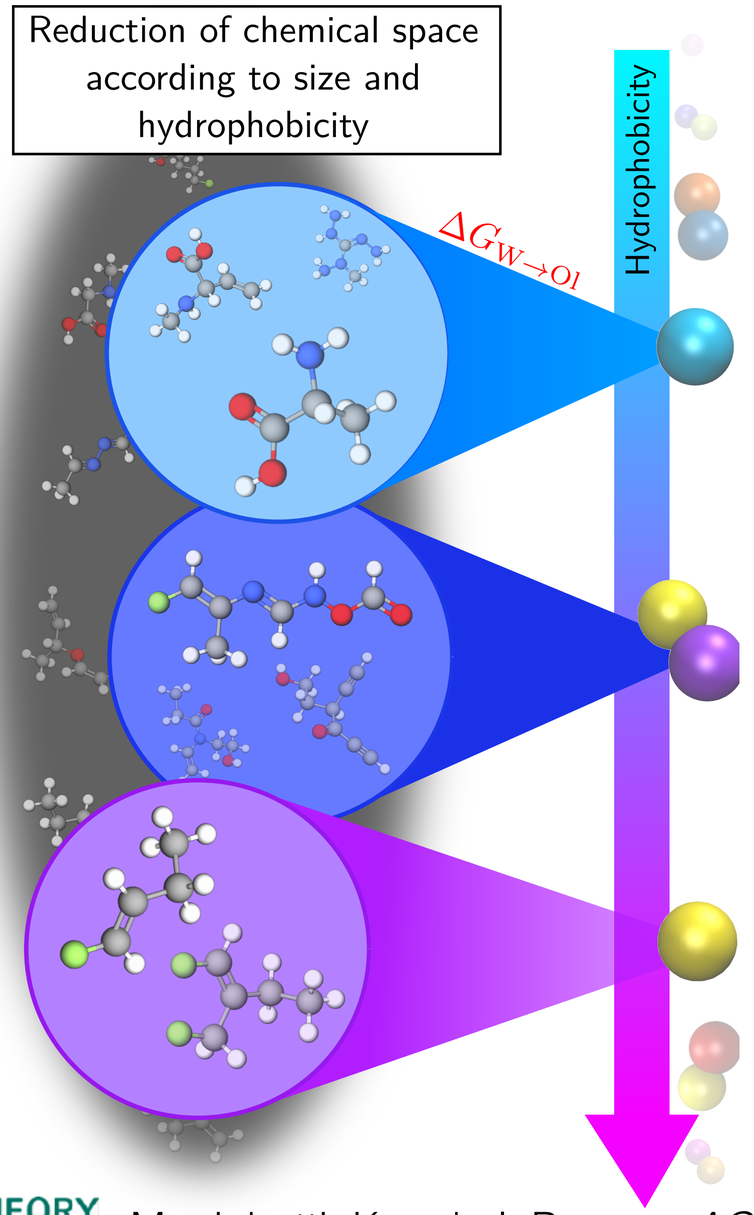
Coarse-graining reduces chemical space



Reduction due to limited number of bead types

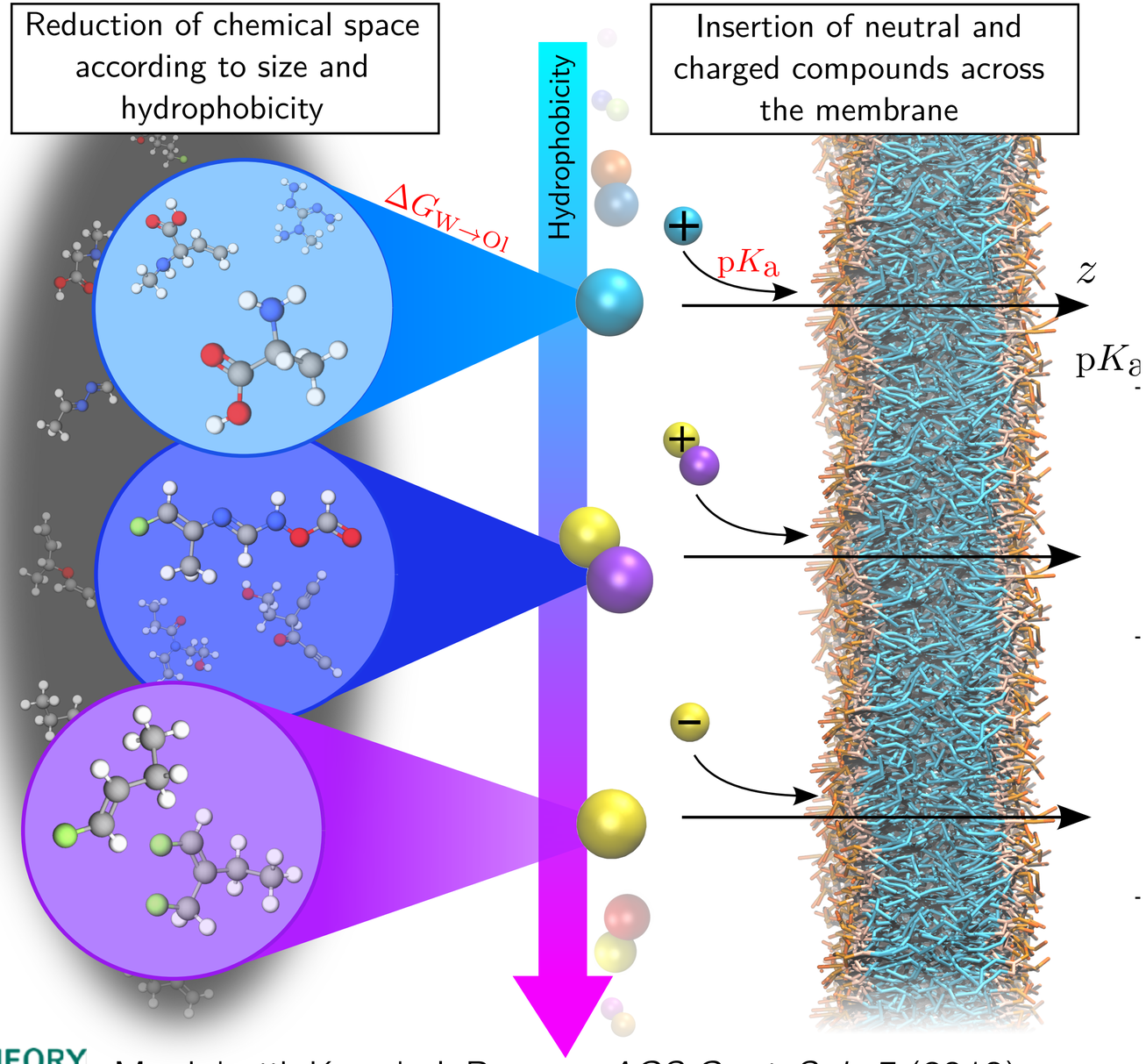


High-throughput coarse-graining scheme



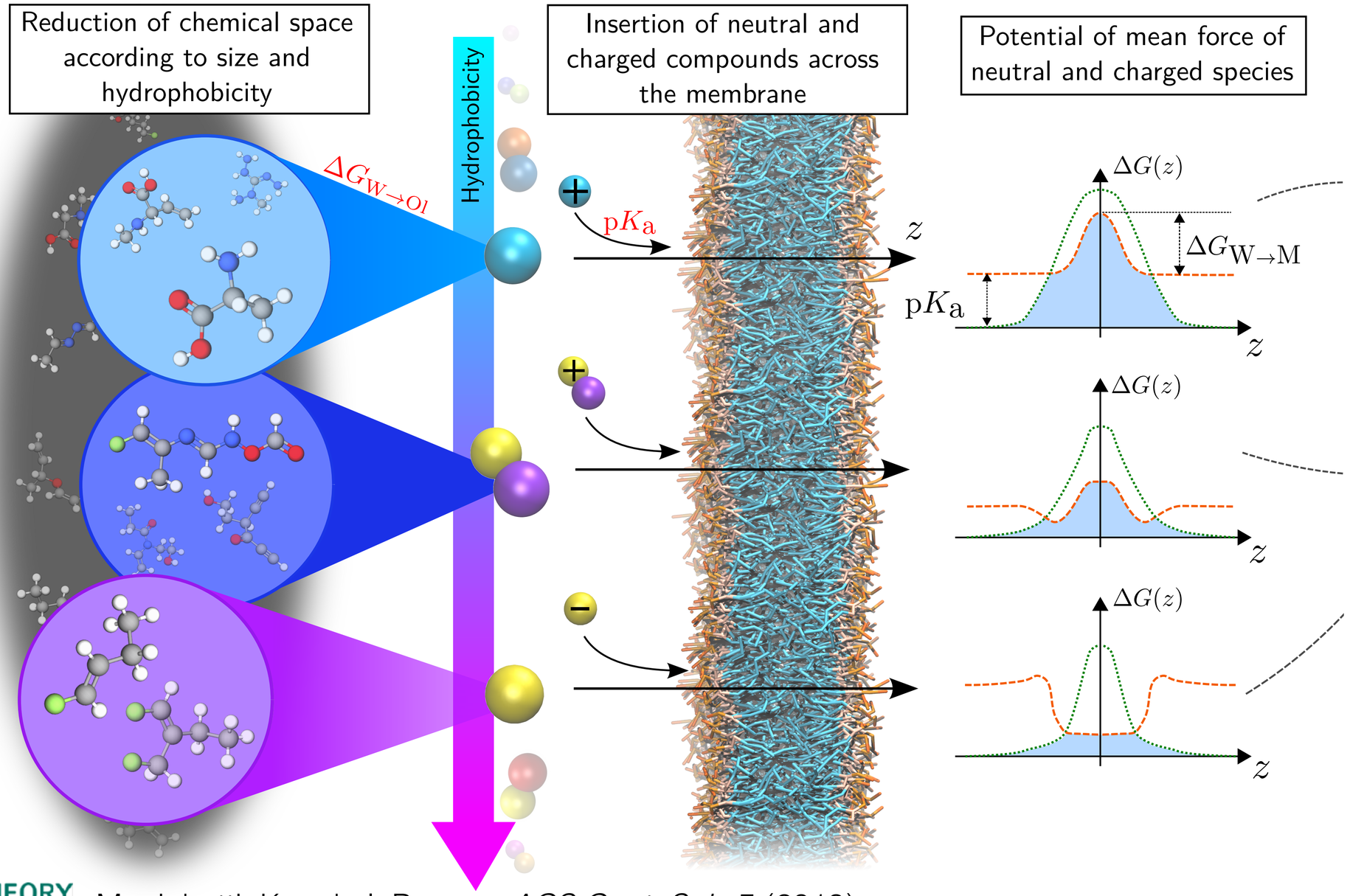
Menichetti, Kanekal, Bereau, *ACS Cent. Sci.* **5** (2019)

High-throughput coarse-graining scheme



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

High-throughput coarse-graining scheme



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

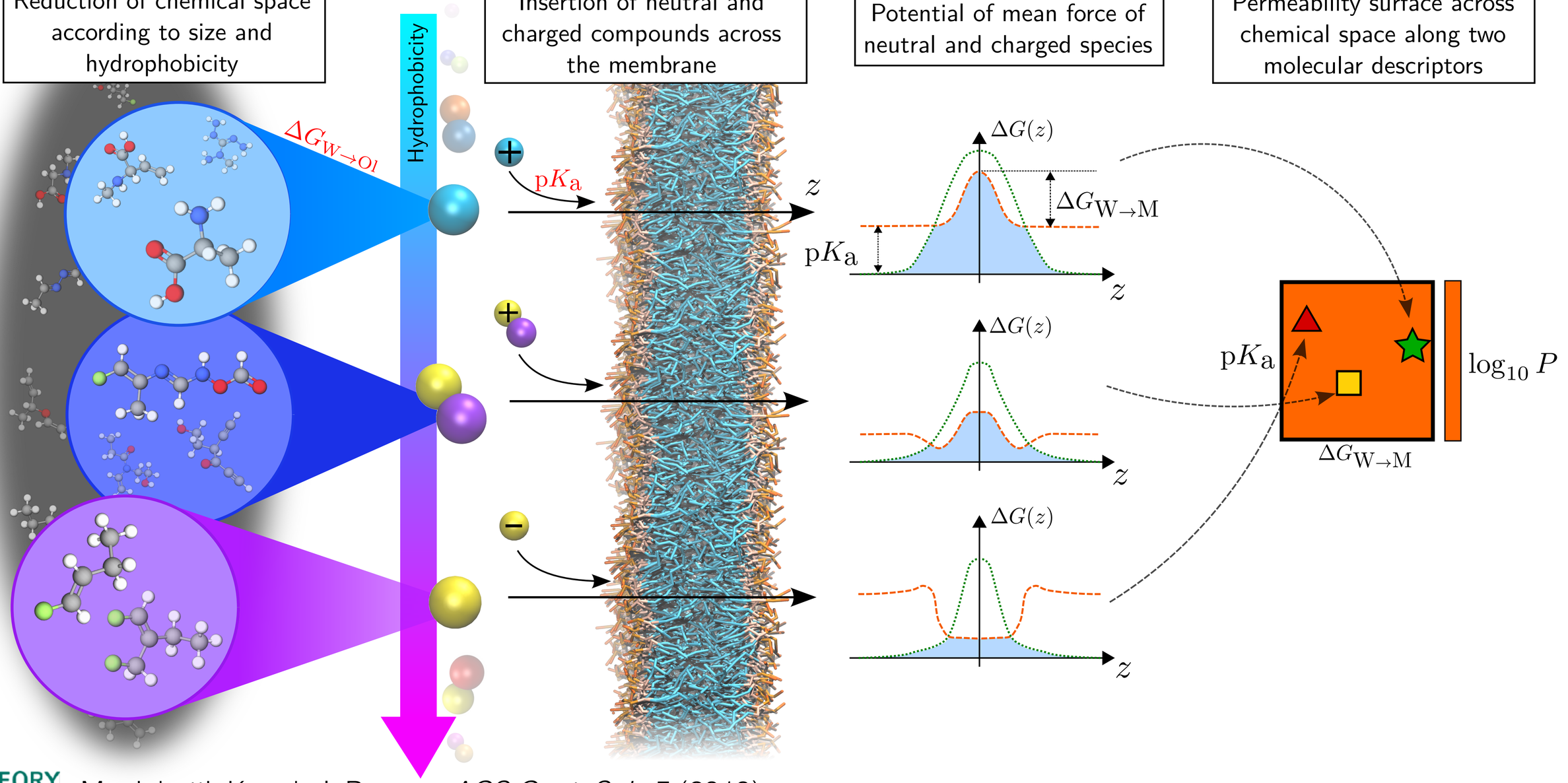
High-throughput coarse-graining scheme

Reduction of chemical space according to size and hydrophobicity

Insertion of neutral and charged compounds across the membrane

Potential of mean force of neutral and charged species

Permeability surface across chemical space along two molecular descriptors



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

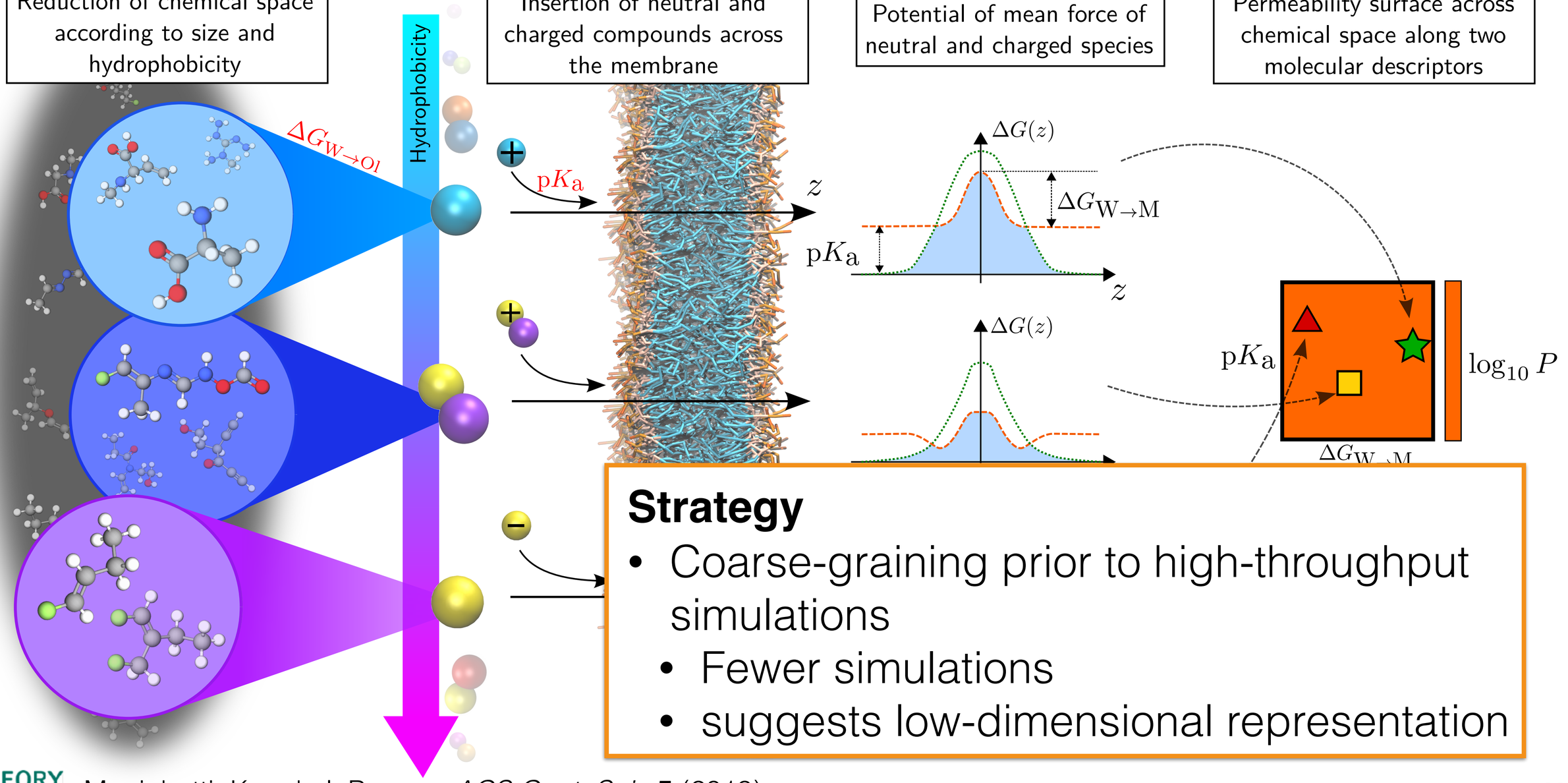
High-throughput coarse-graining scheme

Reduction of chemical space according to size and hydrophobicity

Insertion of neutral and charged compounds across the membrane

Potential of mean force of neutral and charged species

Permeability surface across chemical space along two molecular descriptors



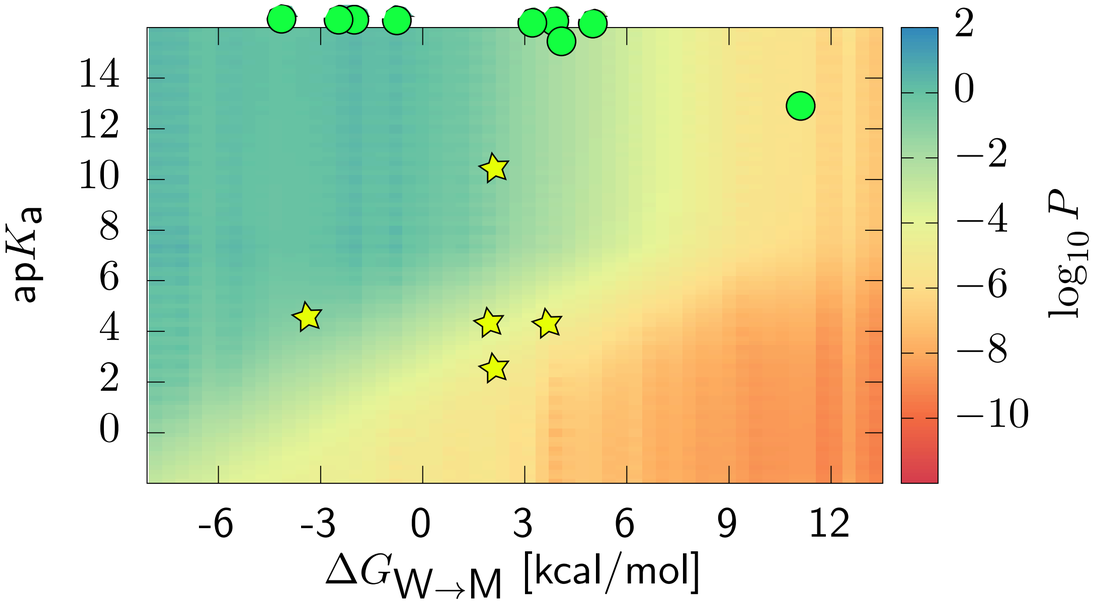
Strategy

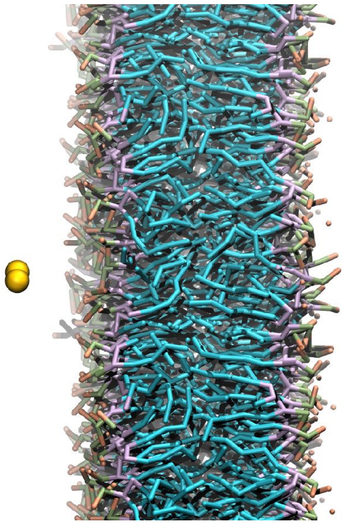
- Coarse-graining prior to high-throughput simulations
  - Fewer simulations
  - suggests low-dimensional representation

THEORY GROUP

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

Chemical-space coverage

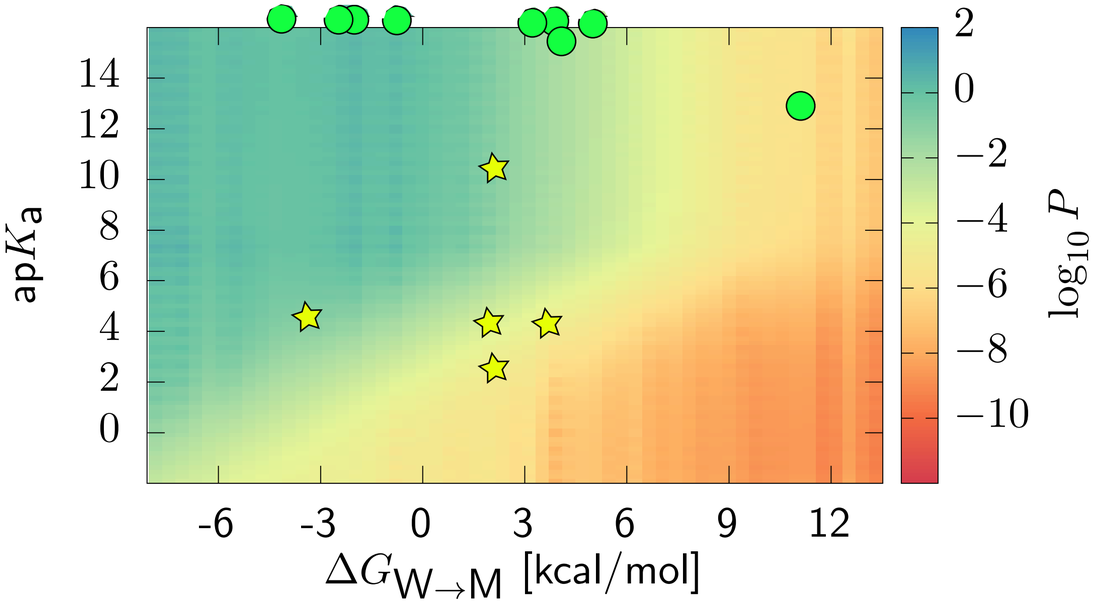


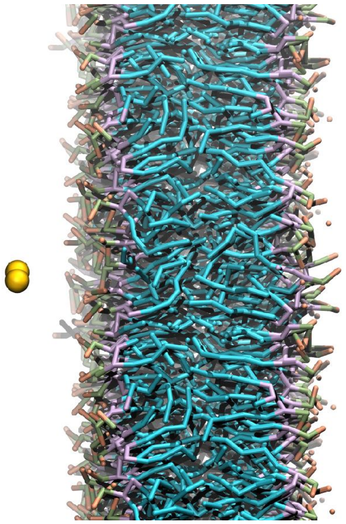


What type of chemistry covers the surface?

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

Chemical-space coverage



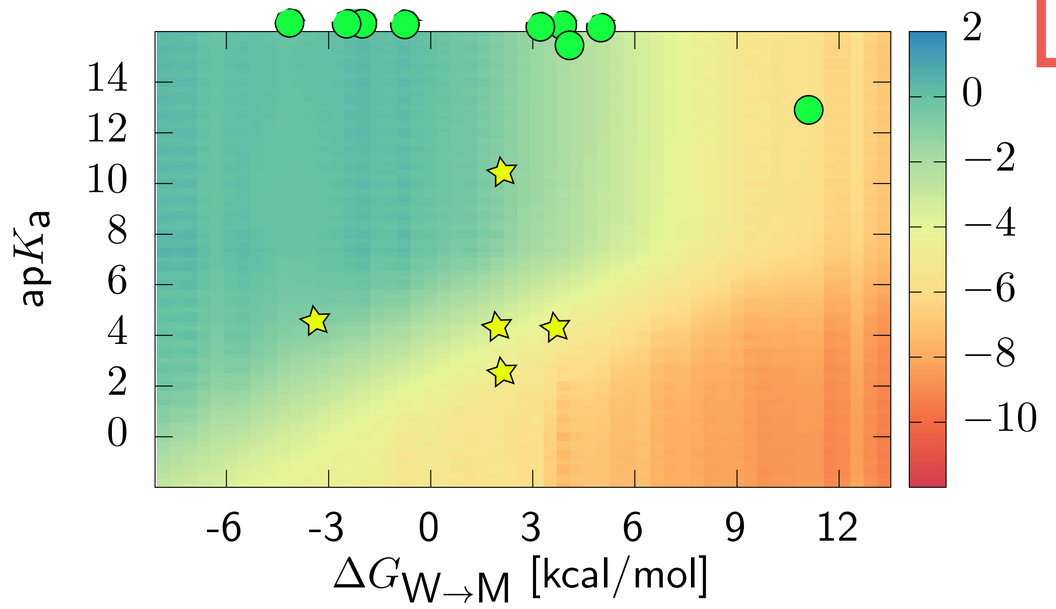


What type of chemistry covers the surface?

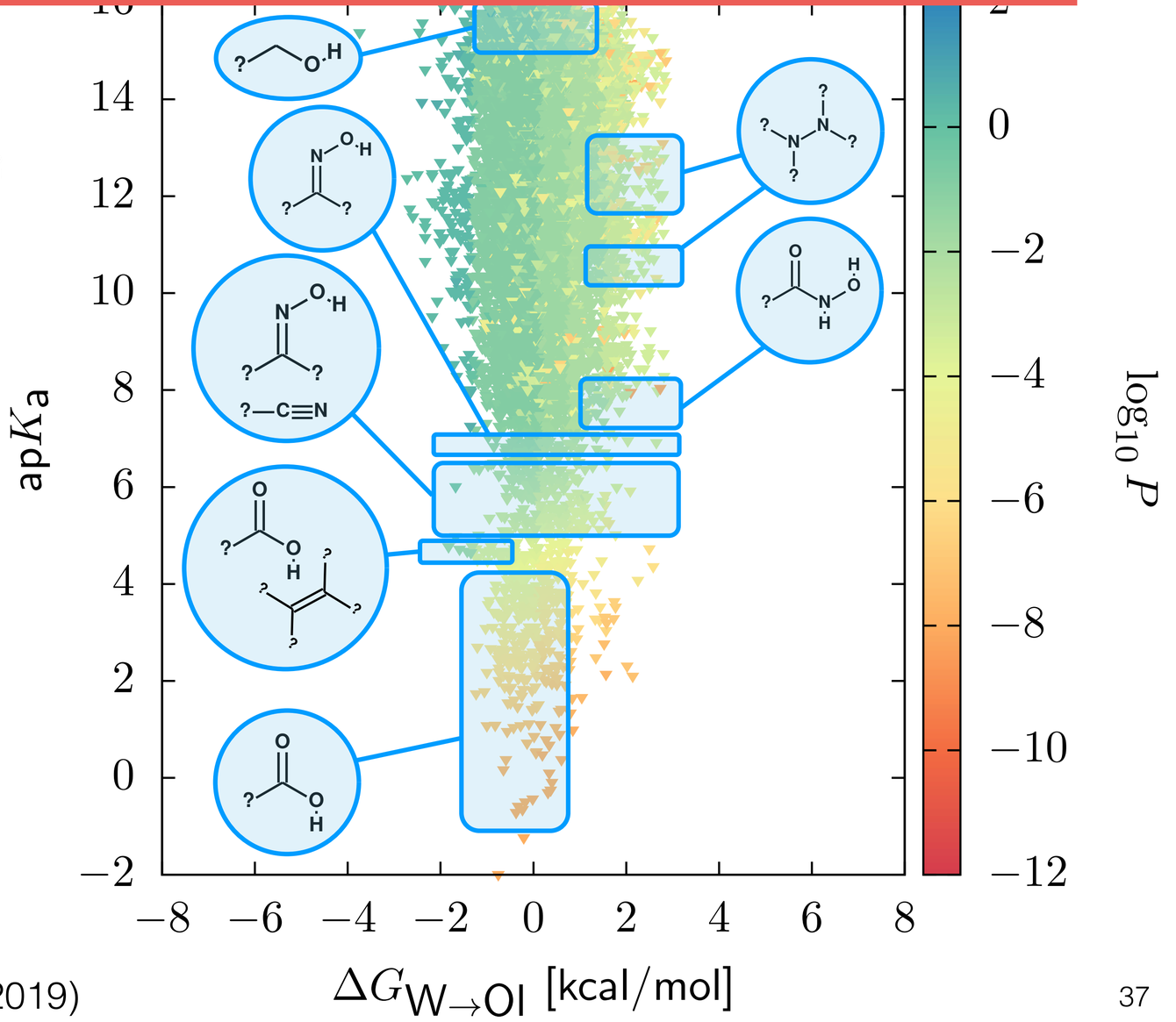
Analysis of **500,000+** compounds

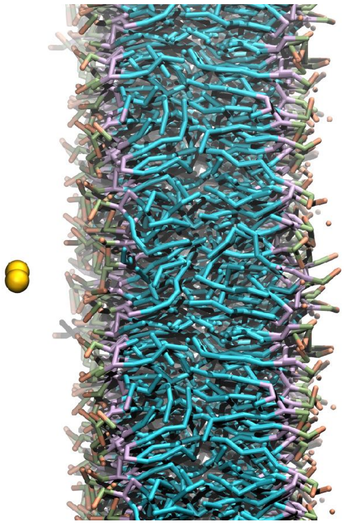
Menichetti, Kanekal, Bereau, *ACS Cent. Sci.* **5** (2019)

Chemical-space coverage



Impact of chemical group on permeability



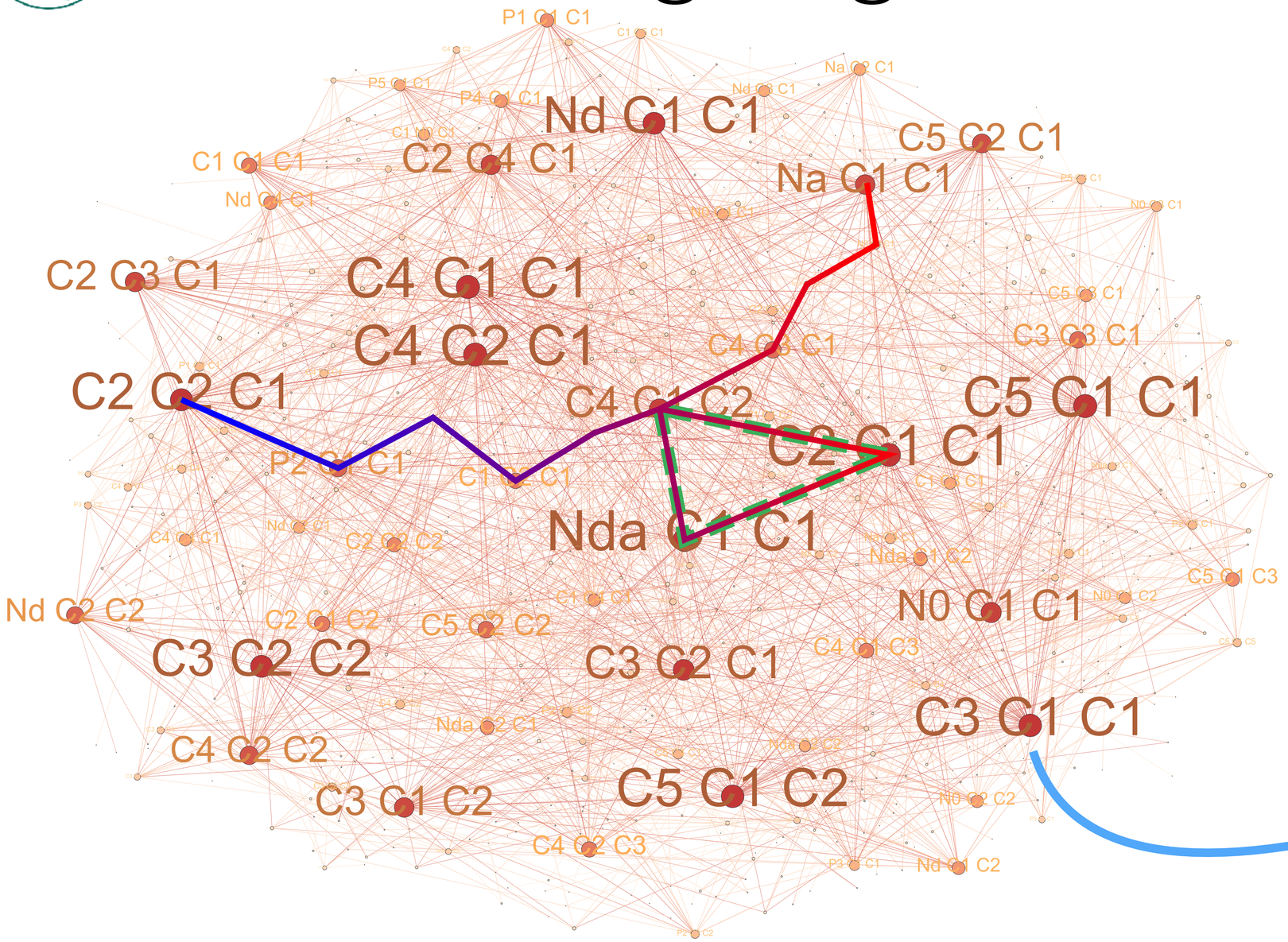


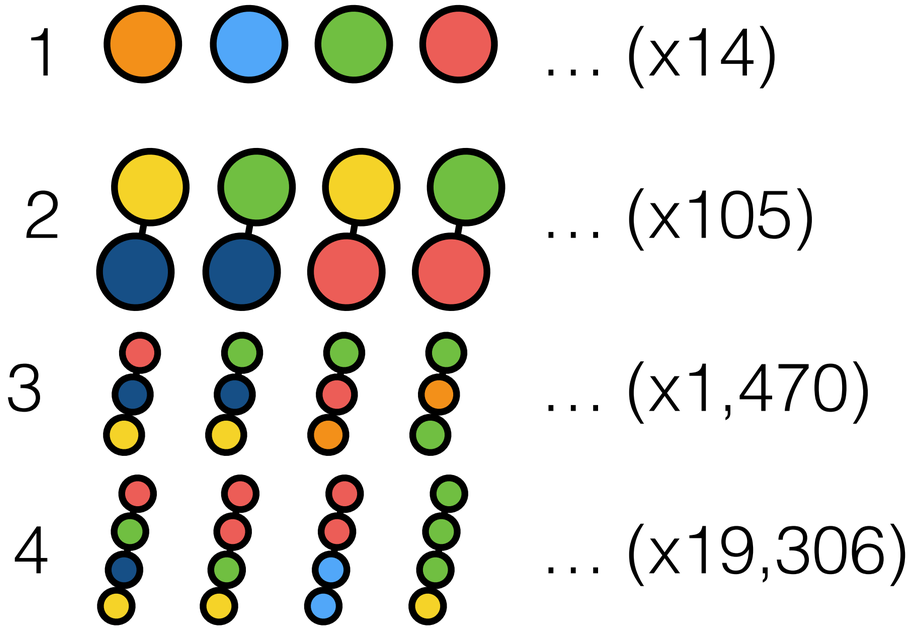
What type of chemistry covers the surface?

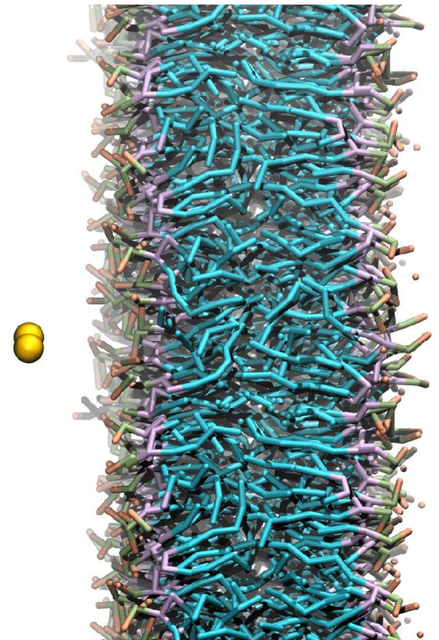
Analysis of **500,000+** compounds

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

Navigating chemical space





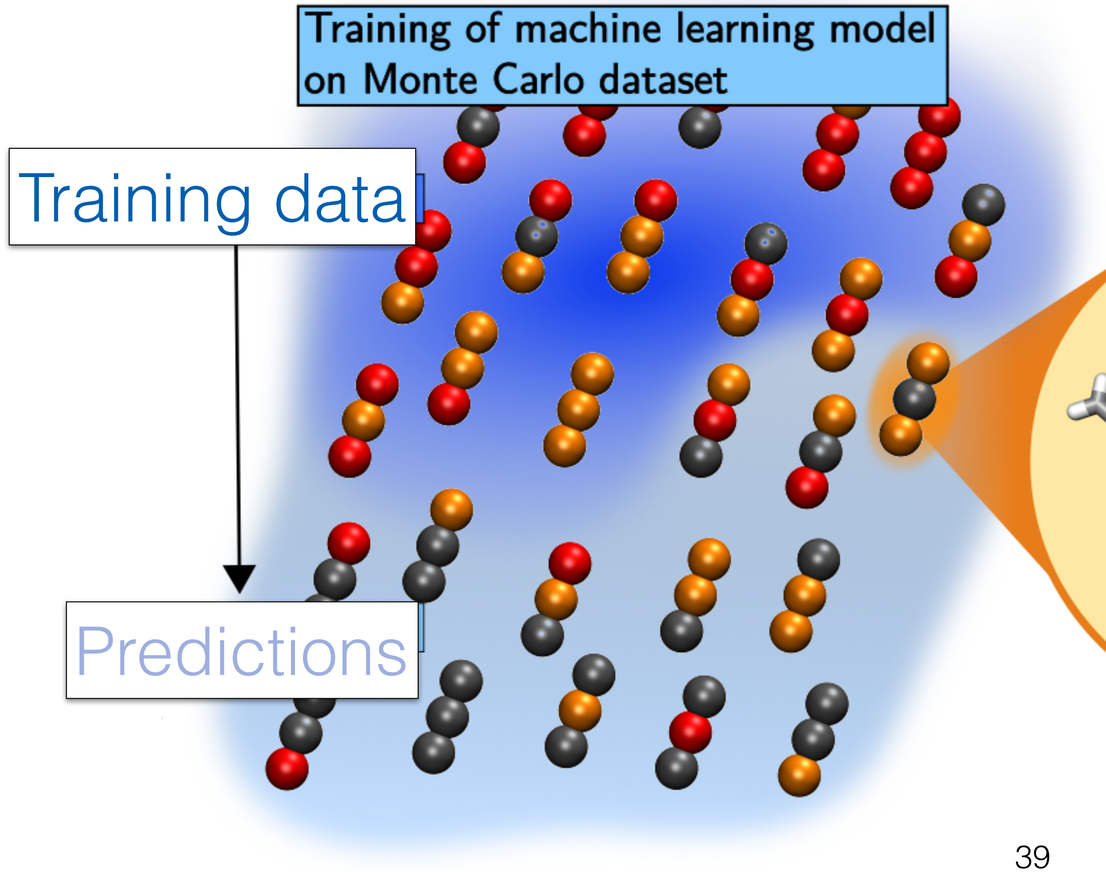


Hoffmann, Menichetti, Kanekal, Bereau, *Phys Rev E* **100** (2019)

Boosting the database with machine learning

Kernel-based machine learning

$$\Delta G(\mathbf{x}) = \sum_i \alpha_i K(\mathbf{x}_i^*, \mathbf{x})$$



Hoffmann, Menichetti, Kanekal, Bereau, *Phys Rev E* **100** (2019)

Boosting the database with machine learning

Kernel-based machine learning

$$\Delta G(\mathbf{x}) = \sum_i \alpha_i K(\mathbf{x}_i^*, \mathbf{x})$$

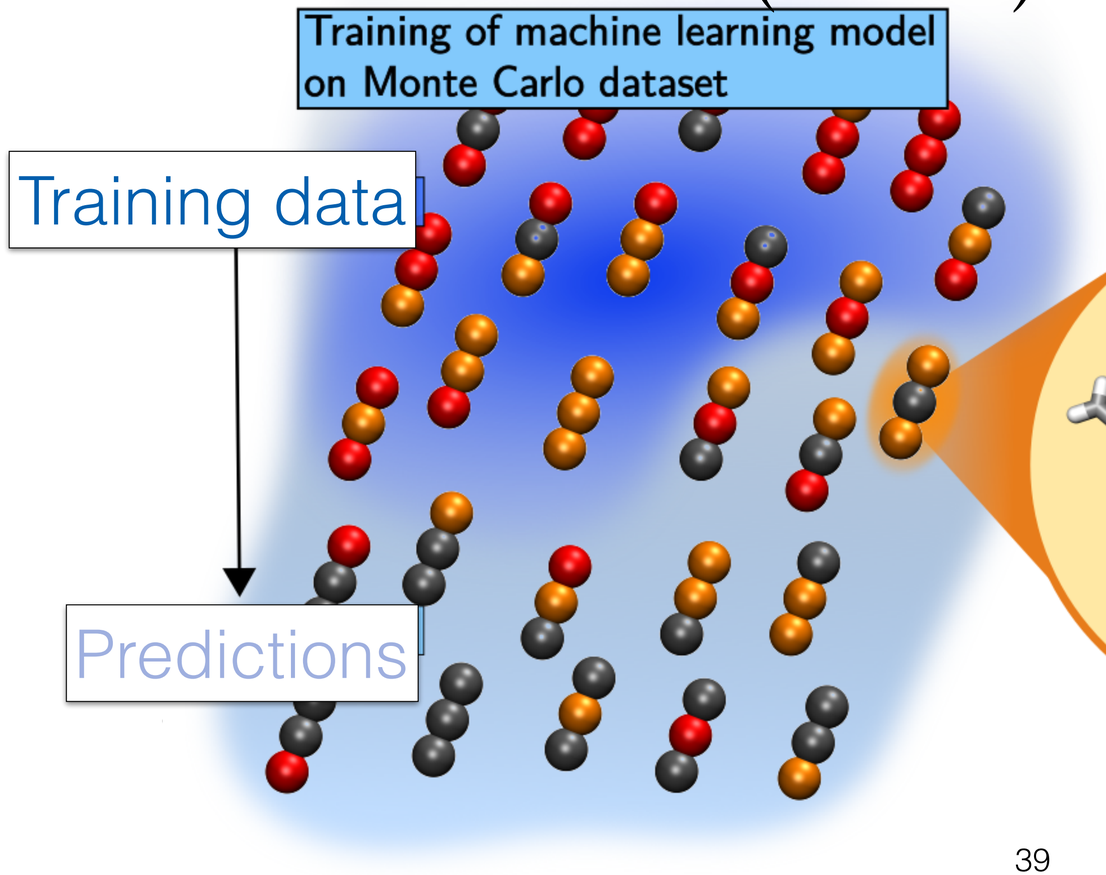
Predicted free energy

Training points

Query sample

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|_1}{\sigma}\right)$$

$$\mathbf{x} = \begin{pmatrix} \Delta G^{(1)}_{\mathrm{W}\rightarrow\mathrm{Ol}} \\ \Delta G^{(2)}_{\mathrm{W}\rightarrow\mathrm{Ol}} \\ \Delta G^{(3)}_{\mathrm{W}\rightarrow\mathrm{Ol}} \end{pmatrix}$$



Hoffmann, Menichetti, Kanekal, Bereau, *Phys Rev E* **100** (2019)

Boosting the database with machine learning

Kernel-based machine learning

$$\Delta G(\mathbf{x}) = \sum_i \alpha_i K(\mathbf{x}_i^*, \mathbf{x})$$

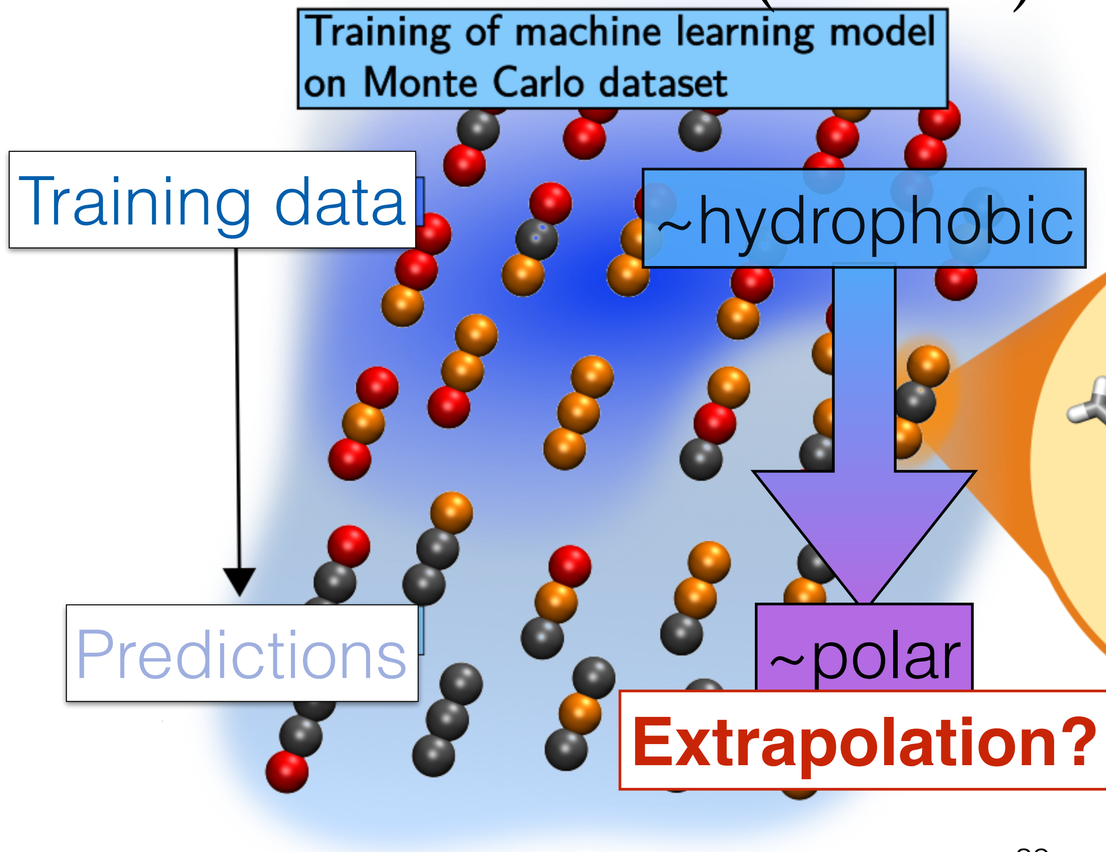
Predicted free energy

Training points

Query sample

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|_1}{\sigma}\right)$$

$$\mathbf{x} = \begin{pmatrix} \Delta G^{(1)}_{\mathrm{W}\to\mathrm{Ol}} \\ \Delta G^{(2)}_{\mathrm{W}\to\mathrm{Ol}} \\ \Delta G^{(3)}_{\mathrm{W}\to\mathrm{Ol}} \end{pmatrix}$$

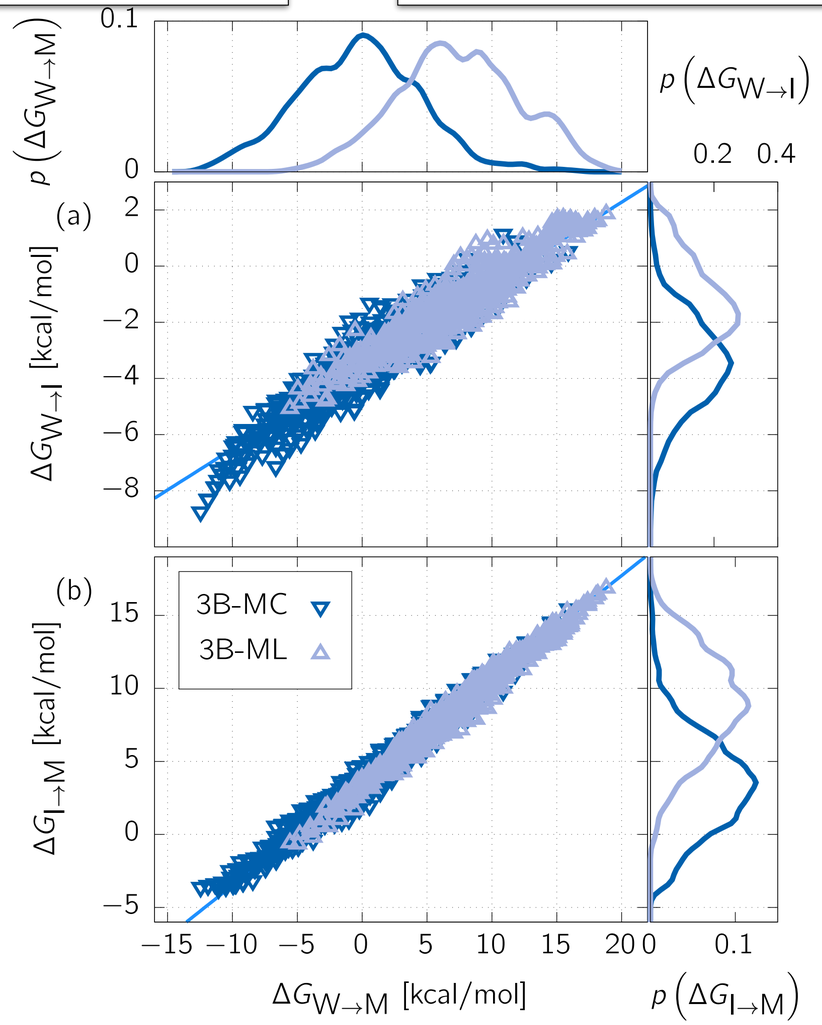


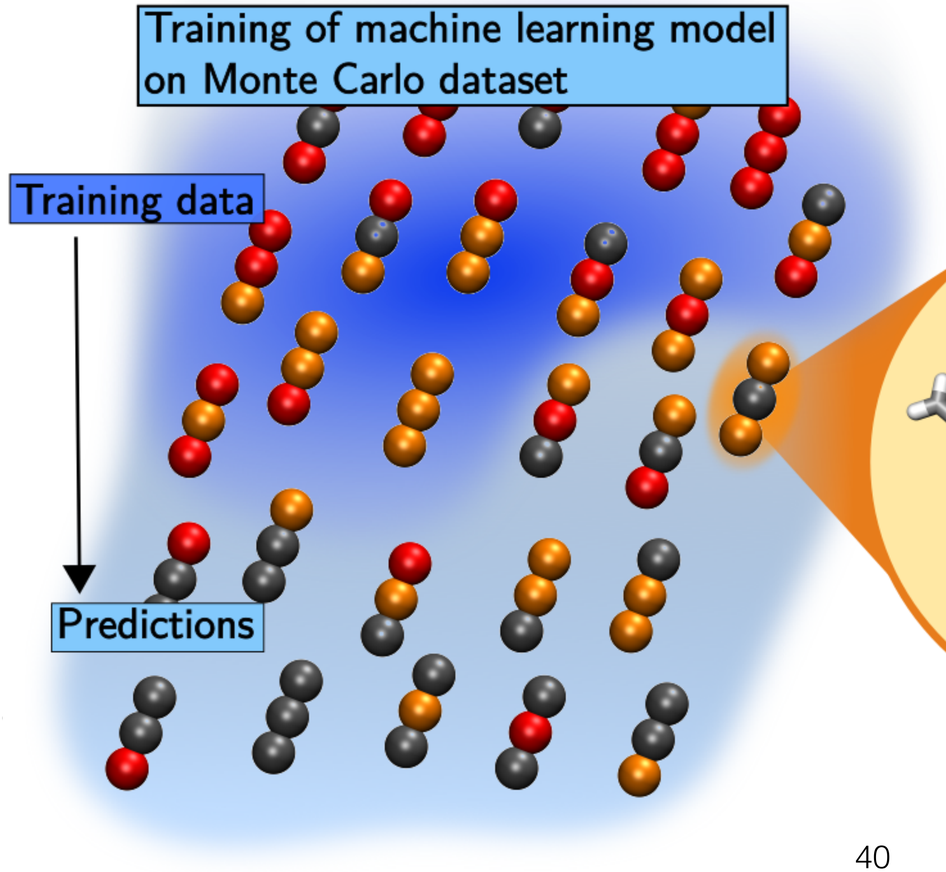
Hoffmann, Menichetti, Kanekal, Bereau, *Phys Rev E* **100** (2019)

Physics offers controlled exploration of chemical space

Existing data

Machine learning



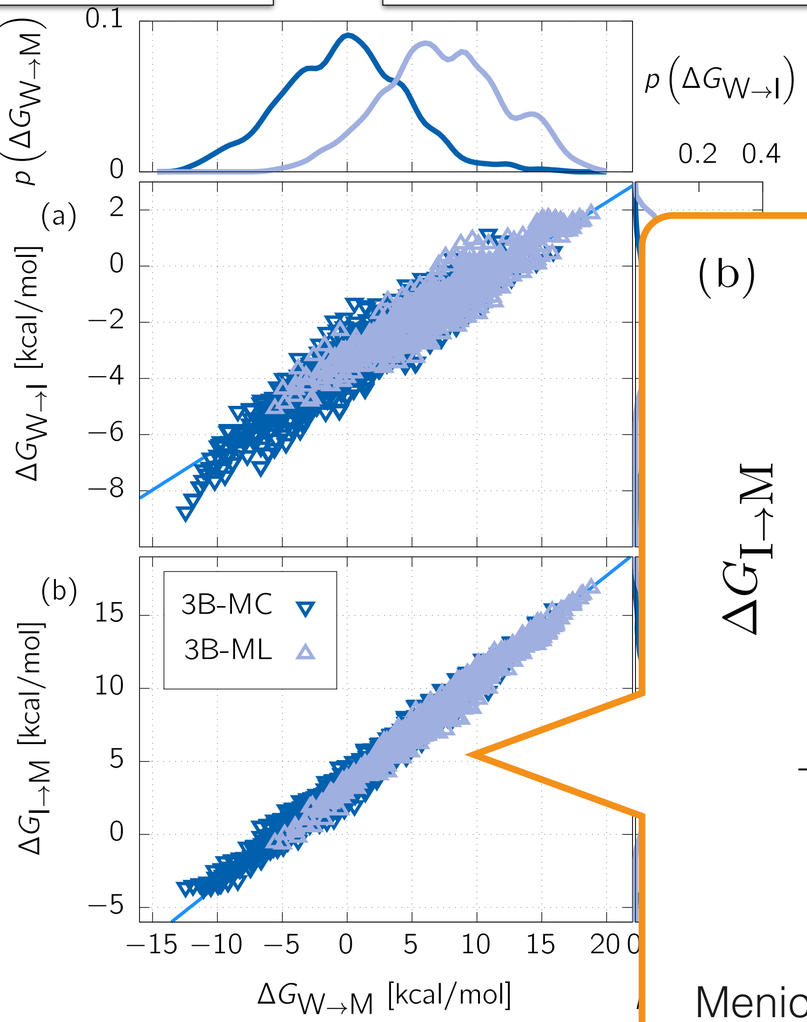


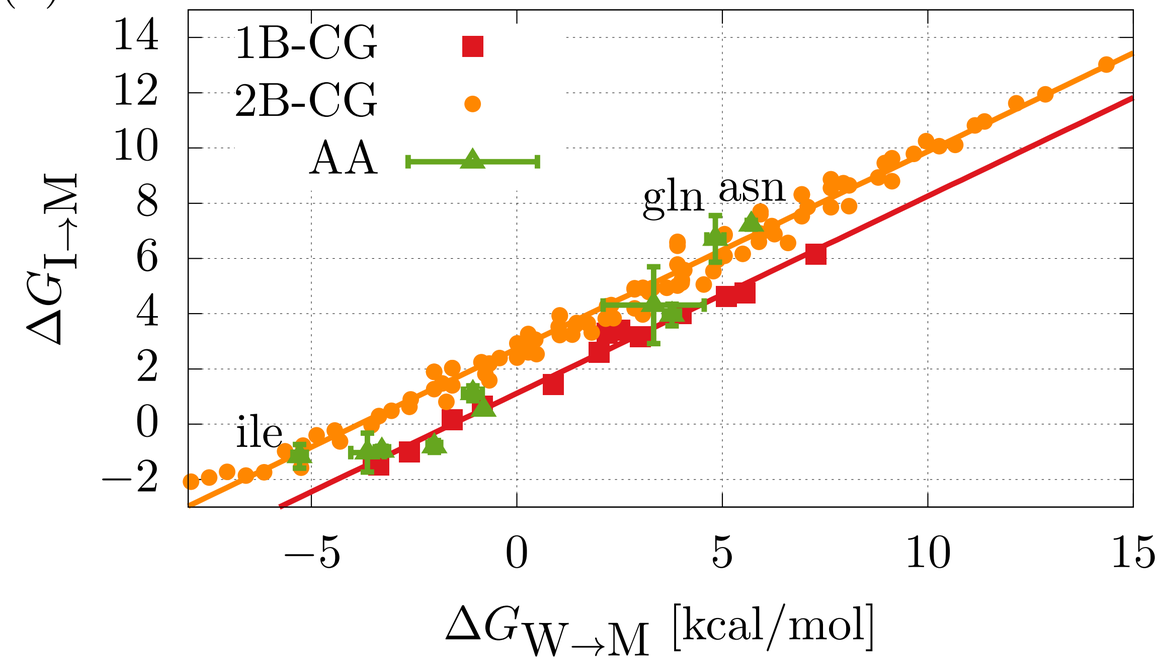
Hoffmann, Menichetti, Kanekal, Bereau, *Phys Rev E* **100** (2019)

Physics offers controlled exploration of chemical space

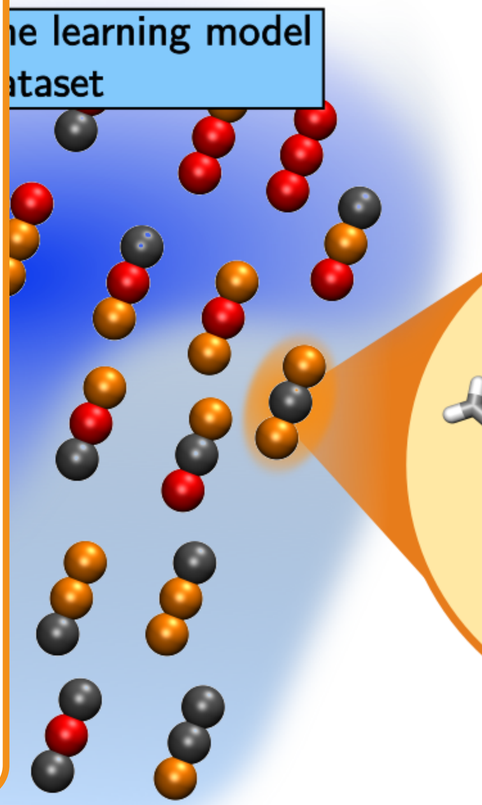
Existing data

Machine learning





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



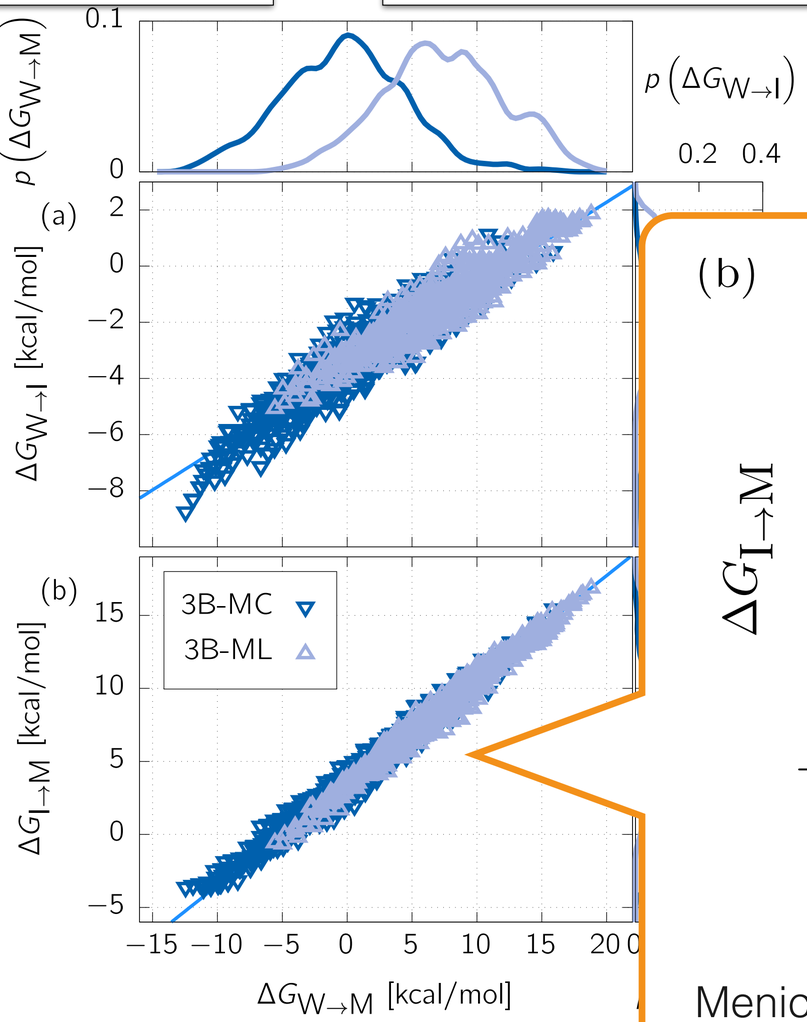
Hoffmann, Menichetti, Kanekal, Bereau, *Phys Rev E* **100** (2019)

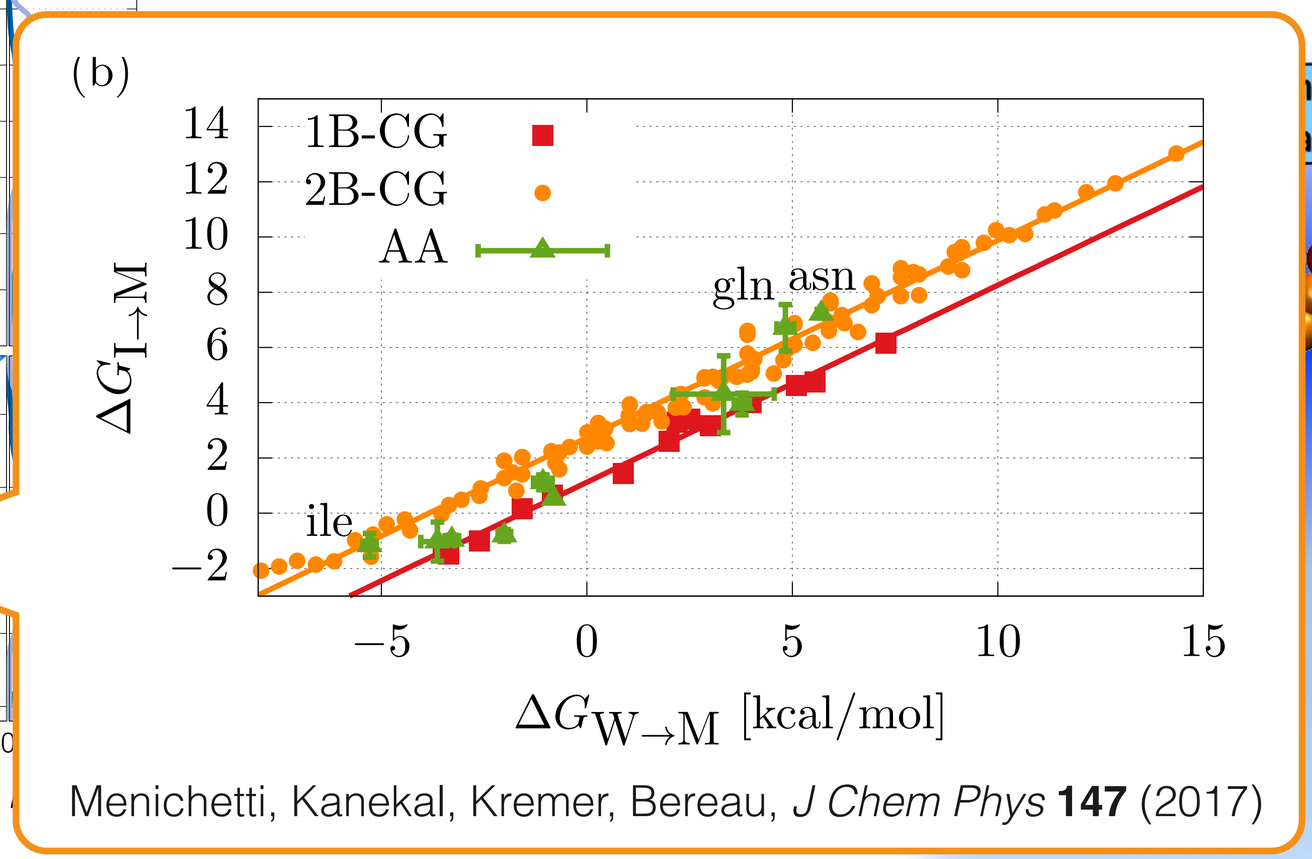
Physics offers controlled exploration of chemical space

Existing data

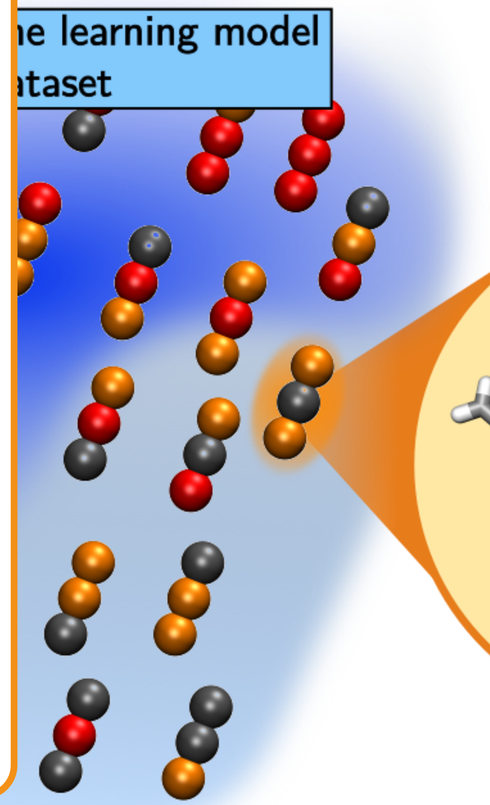
Machine learning

- Thermodynamics: global constraint
- CG representation: better ML transferability
- Boosted database: **1.36M** compounds





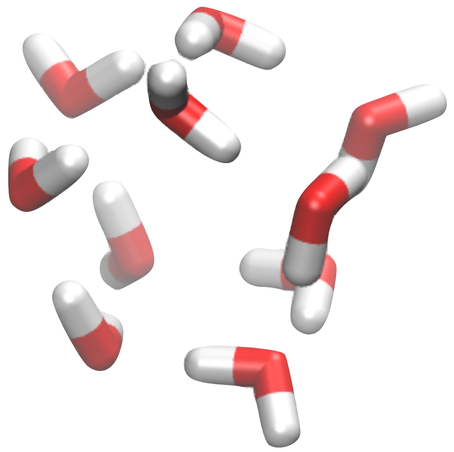
Menichetti, Kanekal, Kremer, Bereau, *J Chem Phys* **147** (2017)



Hoffmann, Menichetti, Kanekal, Bereau, *Phys Rev E* **100** (2019)

Conclusions

Conclusions

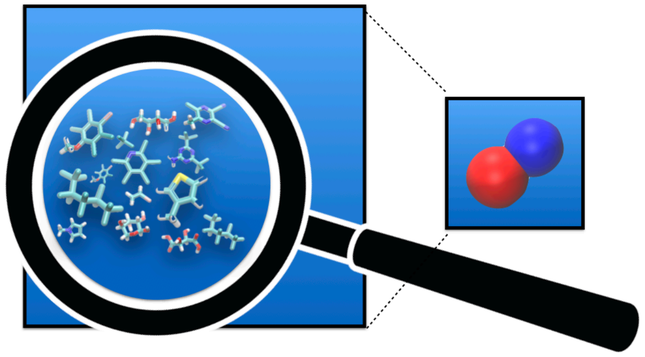


Force fields across chemical space

- Physics encode long-range interactions
- Symmetries reduce the interpolation space
- Learning from QM (atomistic) or physicochemical properties (CG)

High-throughput thermodynamics

- Coarse-graining reduces the size of chemical space: fewer simulations
- Structure-property relationships: CG suggests low-dimensional representation

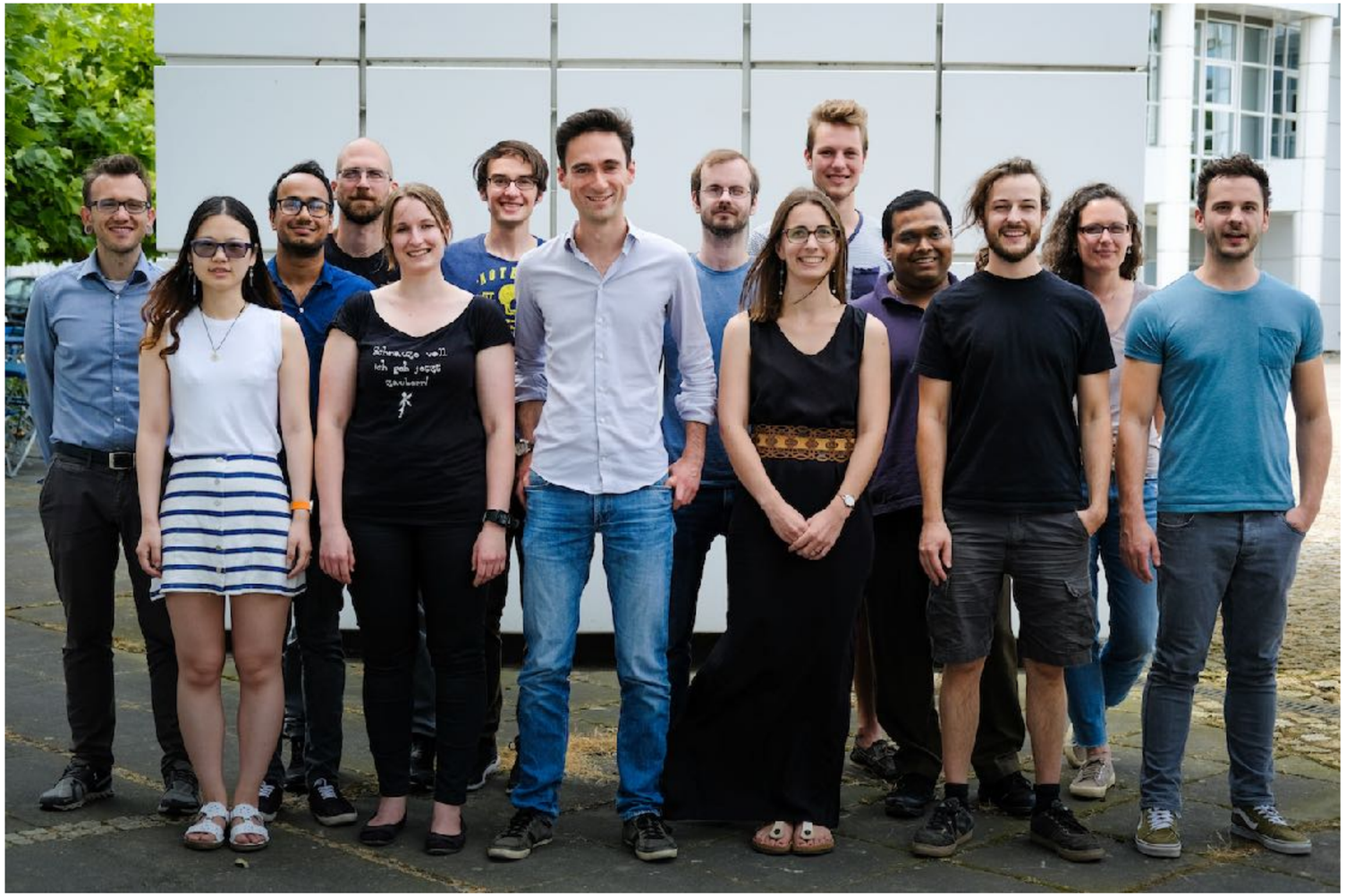


Acknowledgments

MAX PLANCK INSTITUTE
FOR POLYMER RESEARCH

Group members

- Marius Bause
- Yasemin Bozkurt Varolgunes
- Alessia Centi
- Arghya Dutta
- Martin Girard
- **Christian Hoffmann**
- **Kiran Kanekal**
- Chan Liu
- **Roberto Menichetti**
- Bernadette Mohr
- René Scheid
- Clemens Rauer
- Joseph F. Rudzinski
- Marc Stieffenhofer
- Timon Wittenstein
- Svenja Wörner



Collaborators

- Robert DiStasio Jr. (Cornell)
- Anatole von Lilienfeld (Basel)
- Alexandre Tkatchenko (Luxembourg)
- Dirk Schneider (JGU Mainz)
- Michael Wand (JGU Mainz)
- Kurt Kremer
- Denis Andrienko
- Sapun Parekh
- Tanja Weil

NIC

ipam

Carl Zeiss Stiftung

ChemAxon

BiG max MAX PLANCK RESEARCH NETWORK on big-data-driven materials science

TRR 146

Emmy Noether-Programm

DFG Deutsche Forschungsgemeinschaft