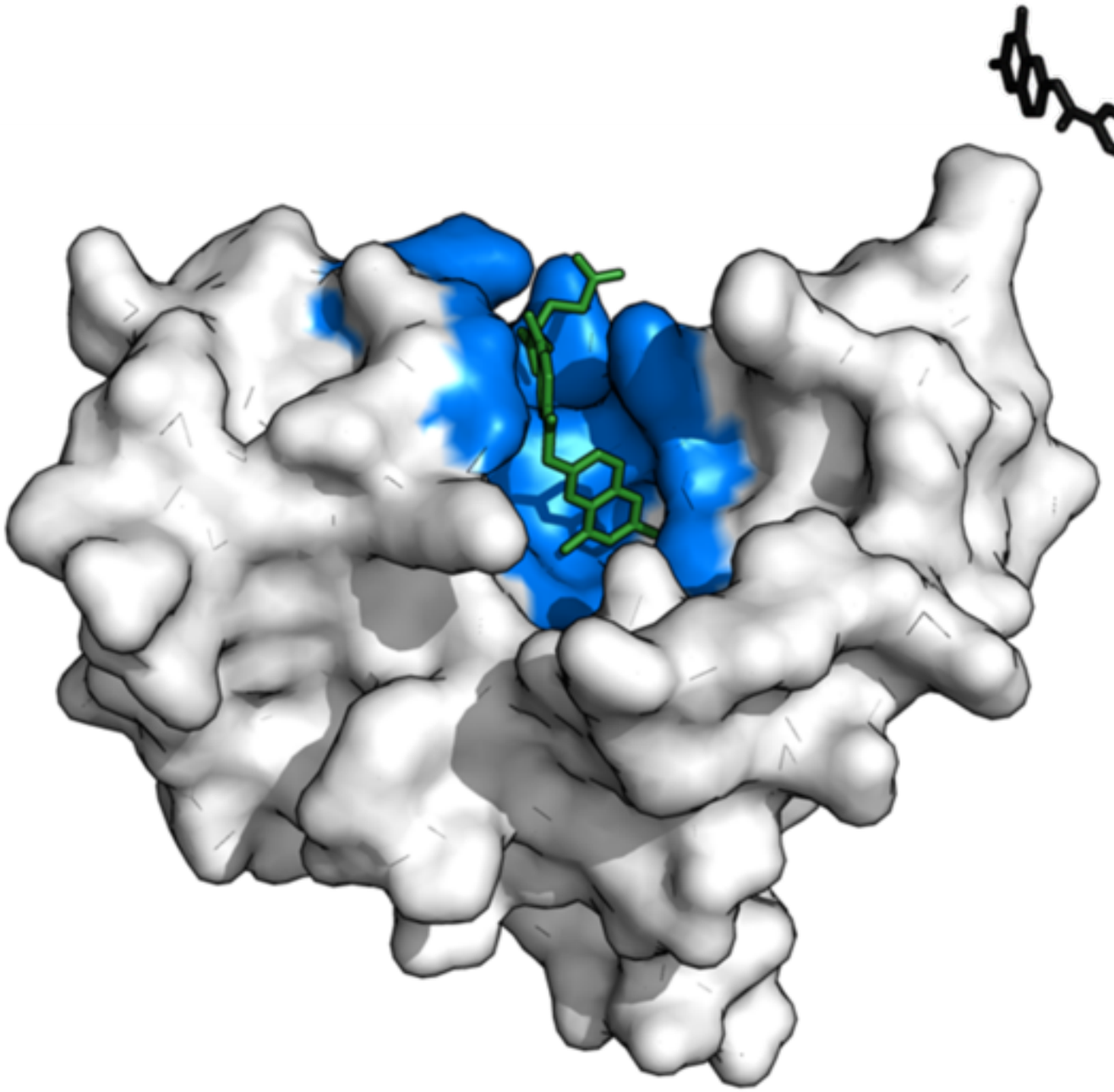
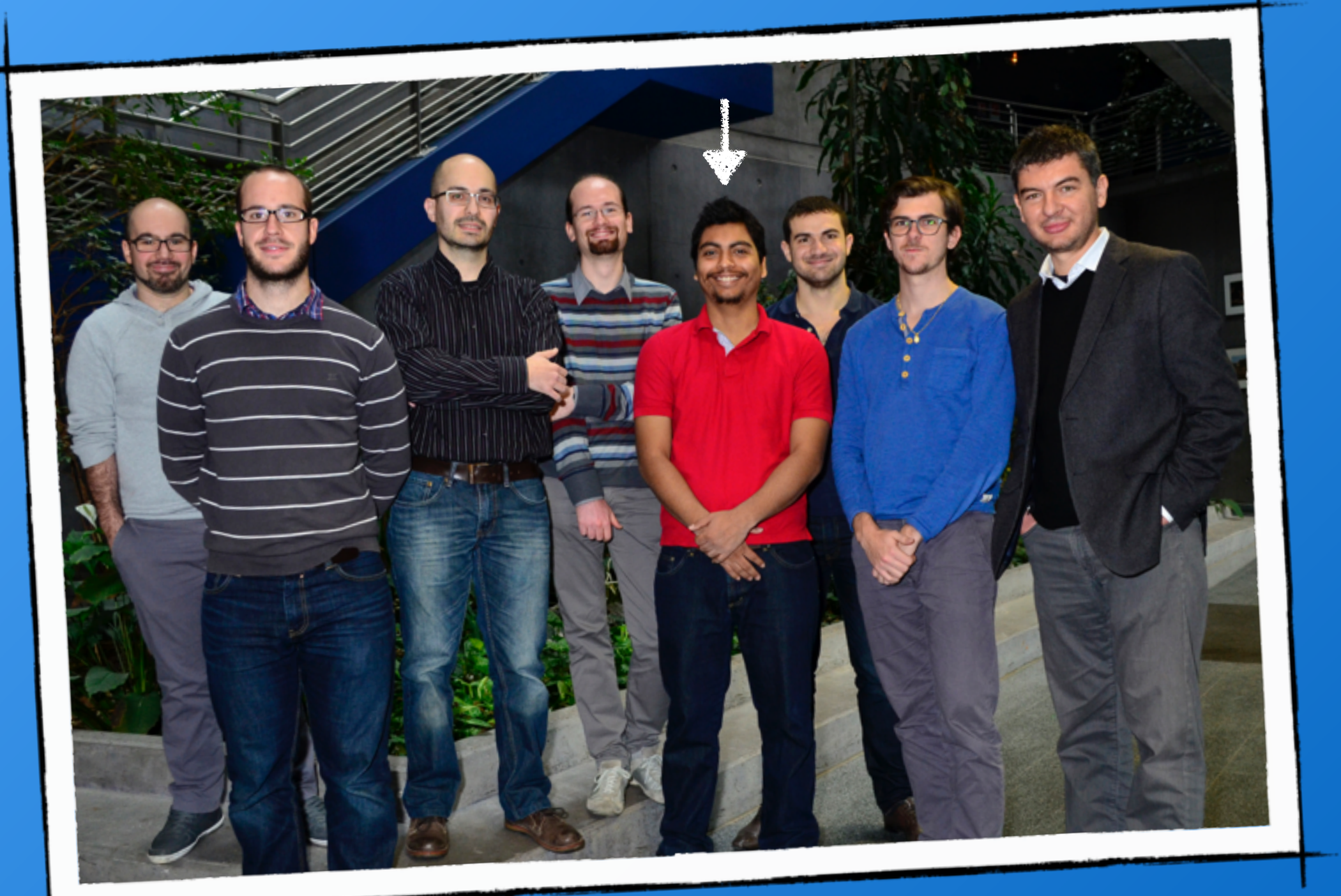
Computational Approaches to Protein-Ligand Binding



Marco Cecchini

Laboratoire d'Ingénierie des Fonctions Moléculaires
ISIS (UMR7006), Faculté de Chimie
Université de Strasbourg

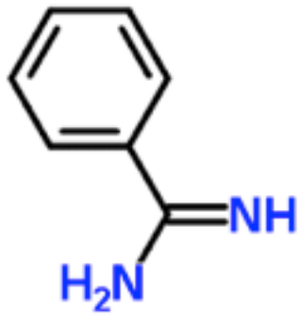
Lab d'Ingénierie des Fonctions Moléculaires



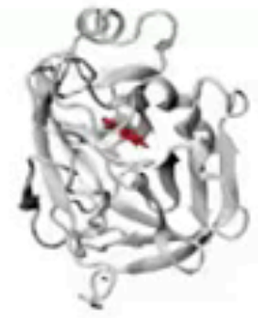
Joel Montalvo-Acosta

Outline

- Protein-ligand binding
- Computational approaches to ligand-binding affinity
- Provide a classification of methods based on stat mech
- Results on a host-guest system



Protein-Ligand Binding



Buch, Giorgino & De Fabritiis, *PNAS* (2011)

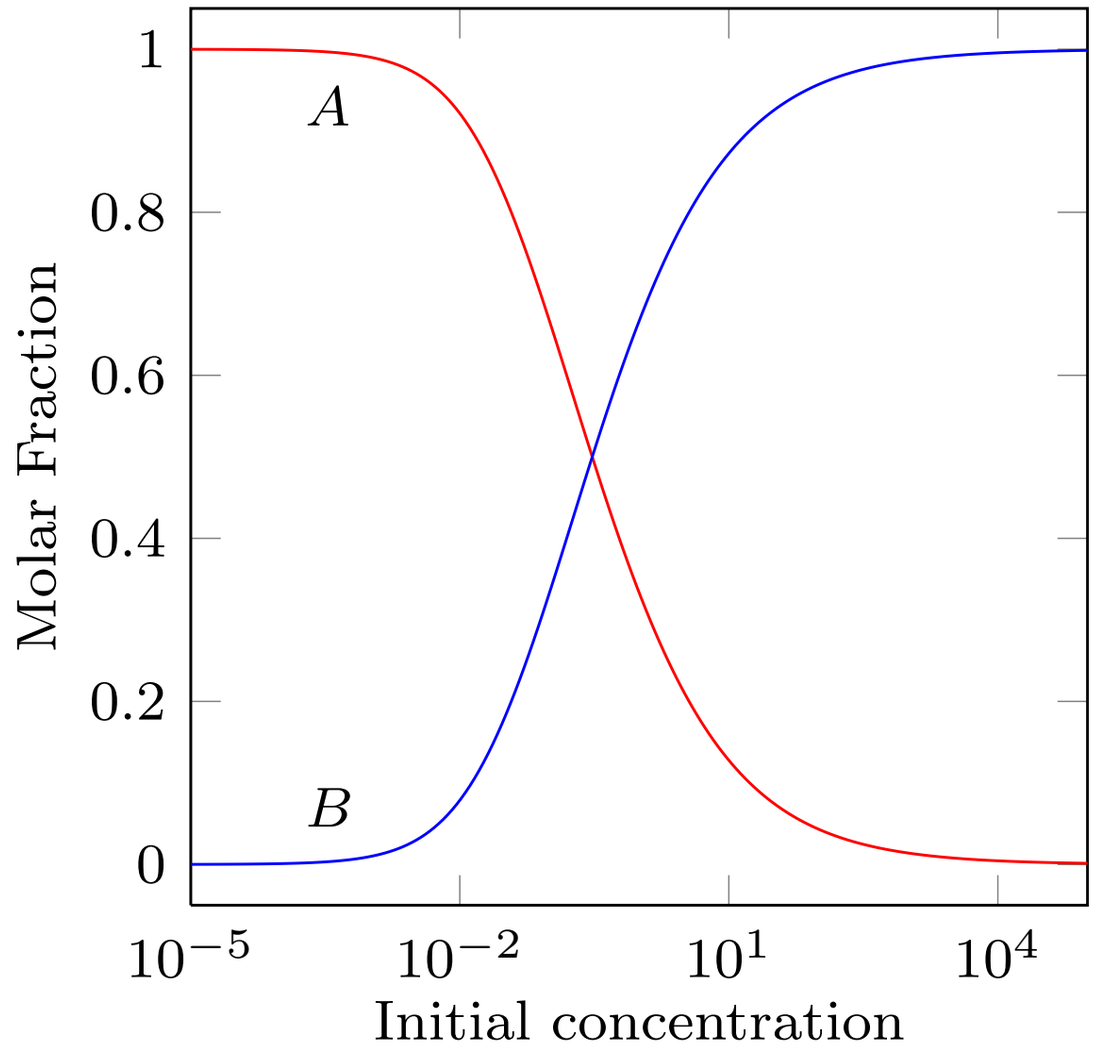
Ligand Binding Affinity

$$P + L \rightleftharpoons PL$$

$$K_d = 1/K_{\text{eq}} = \frac{[P][L]}{[PL]}$$

when $[L] = K_d$

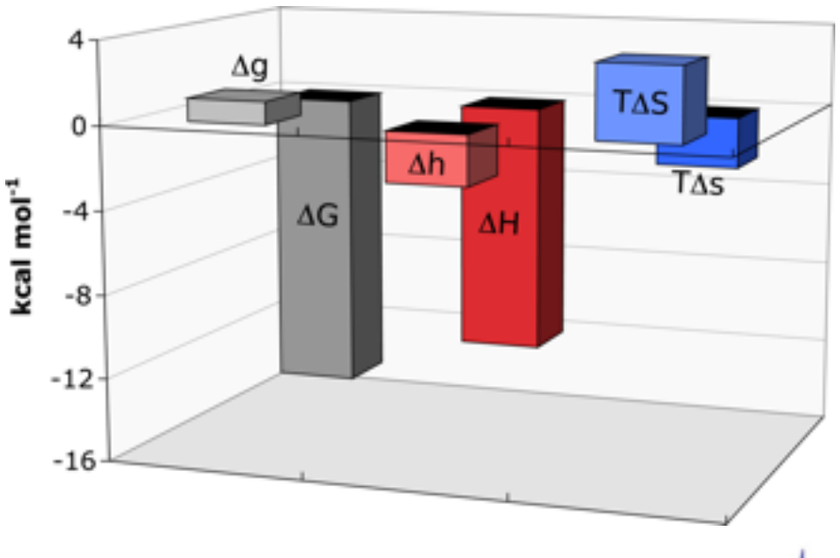
$$\frac{[PL]}{[PL] + [P]} = \frac{1}{2}$$

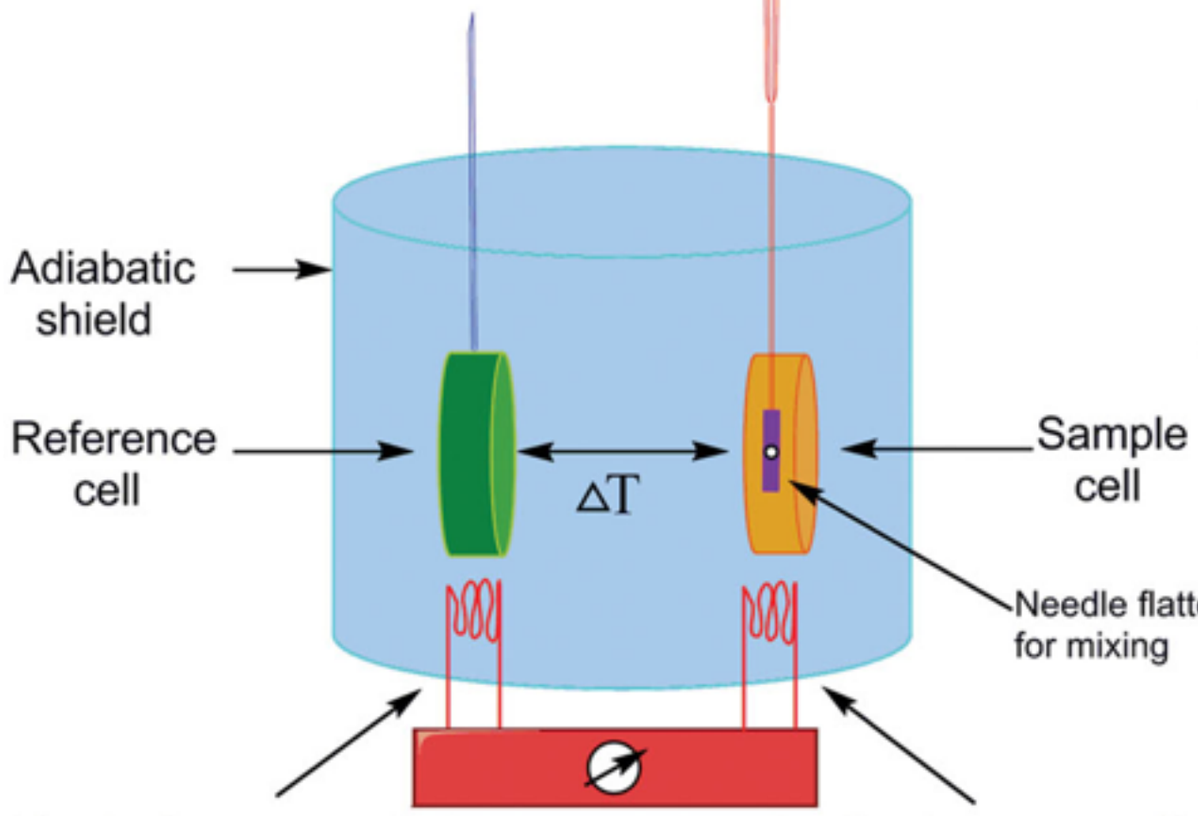


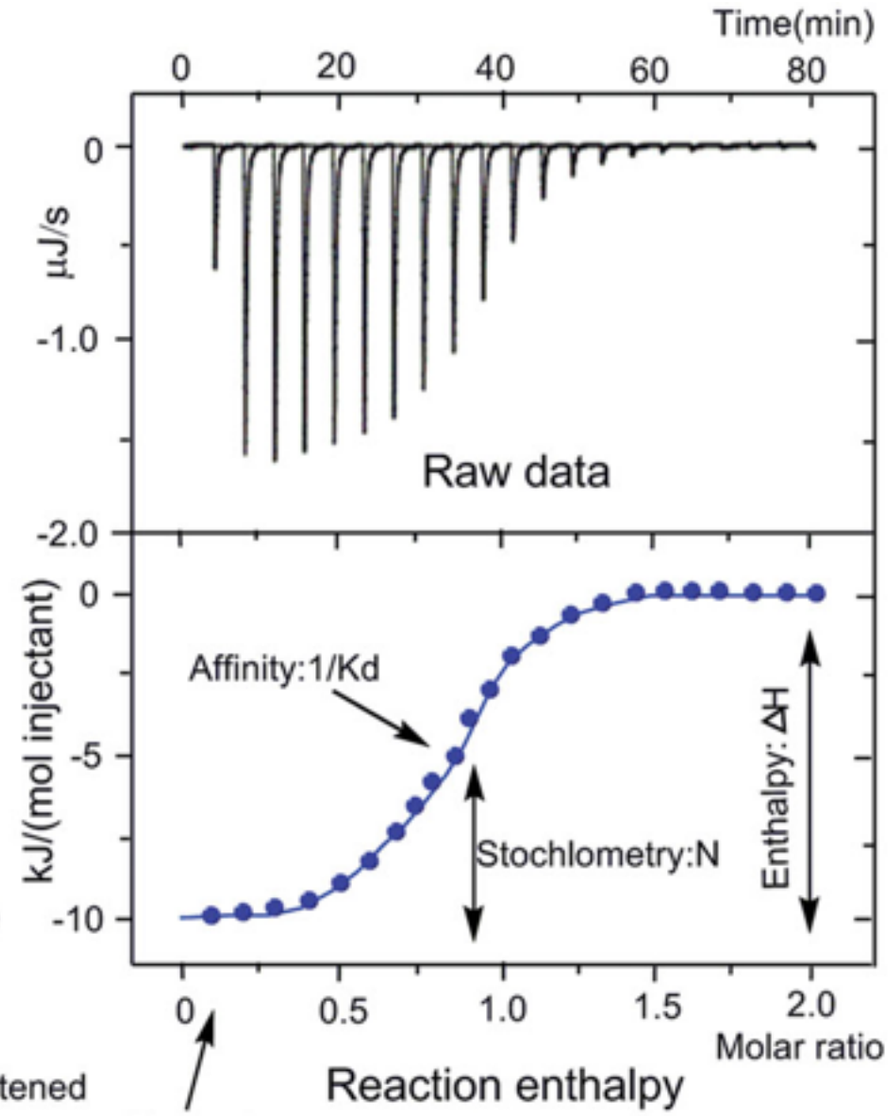
ligand potency (Kd)

$$\mu M < nM < pM$$

Isothermal Titration Calorimetry







Canonical Approach

Chemical thermodynamics

$$\exp\left(-\frac{\Delta\mu_b^\circ}{kT}\right) = \frac{C_{PL}\,(C^\circ)}{C_P\,C_L} = K_{eq}C^\circ$$

Classical Statistical Mechanics

$$F(N,V,T) = -kT\ln Q$$

$$\mu(V,T) = \left(\frac{\partial F}{\partial N}\right)_{V,T} = -kT\left(\frac{\partial \ln Q}{\partial N}\right)_{V,T}$$

$$K_{eq} \Rightarrow \Delta\mu_b^\circ \Rightarrow Q_L, Q_P, Q_{PL}$$

Chemical Potentials

In the limit of RRHO/BO & in vacuum

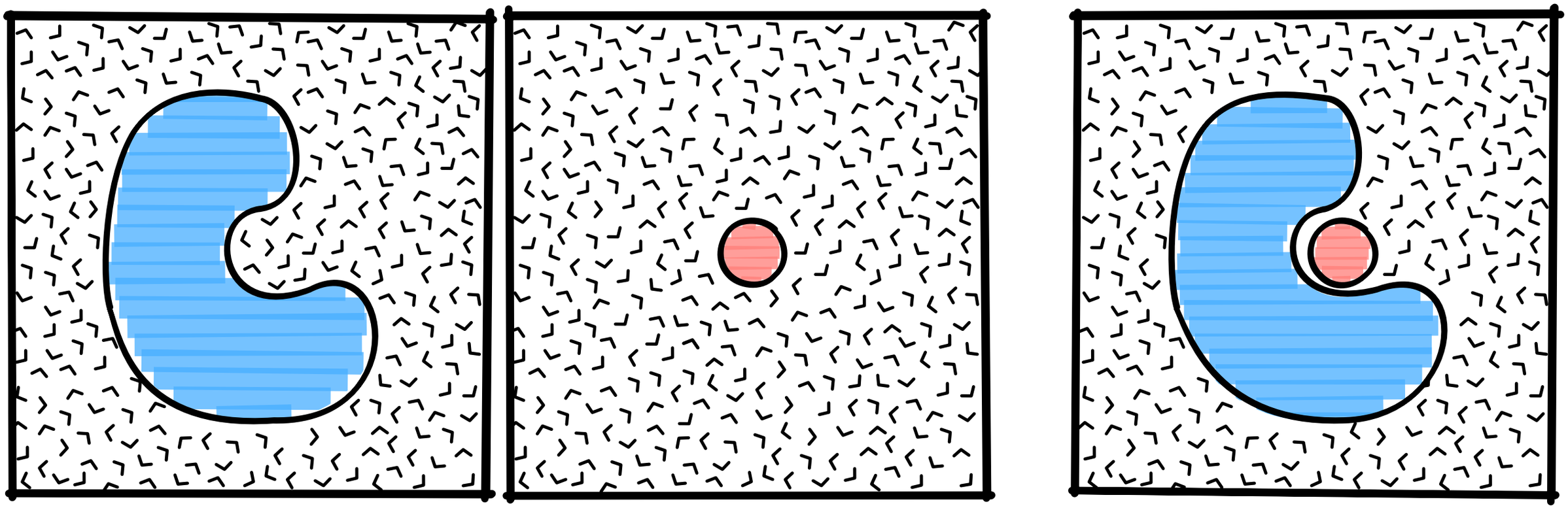
$$\mu_i(V,T) = \mu_{i,v}(V,T) + W_{bulk}(\mathbf{X}_0)$$

$$\begin{aligned}
\mu_{i,v}(V,T) \quad = \quad & -kT \ln \left[\left(\frac{2\pi mkT}{h^2} \right)^{\frac{3}{2}} \frac{V}{N} \right] + \\
& -kT \ln \left[\frac{\sqrt{\pi}}{\sigma} \left(\frac{8\pi^2 kT}{h^2} \right)^{\frac{3}{2}} \sqrt{I_X I_Y I_Z} \right] + \\
& -kT \sum_{j=1}^{3n-6} \ln \left[\left(\frac{kT}{h\nu_j} \right) \right] + \\
& -D_e
\end{aligned}$$

McQuarrie, *Statistical Mechanics*, New York: Harper & Row, 1976

1. Absolute Chemical Potentials

$$P + L \rightleftharpoons PL$$



$$\mu_i^{\circ}(T) = \mu_{i,v}^{\circ}(T) + W_{bulk}(\mathbf{X}_0)$$

Grand Canonical Approach

$$\mu_i^\circ(T) = -kT \ln\left(\frac{\mathcal{Q}_i(p,T)}{V}\right)$$

$$\exp\left(-\frac{\Delta\mu_b^\circ}{kT}\right) = \frac{(\mathcal{Q}_{PL}/V_{PL})}{(\mathcal{Q}_P/V_P)(\mathcal{Q}_L/V_L)} = K_{eq}$$

@ infinite dilution

$$\mathcal{Q}_i(p,T) \approx \frac{Q_{N,1}}{Q_{N,0}}$$

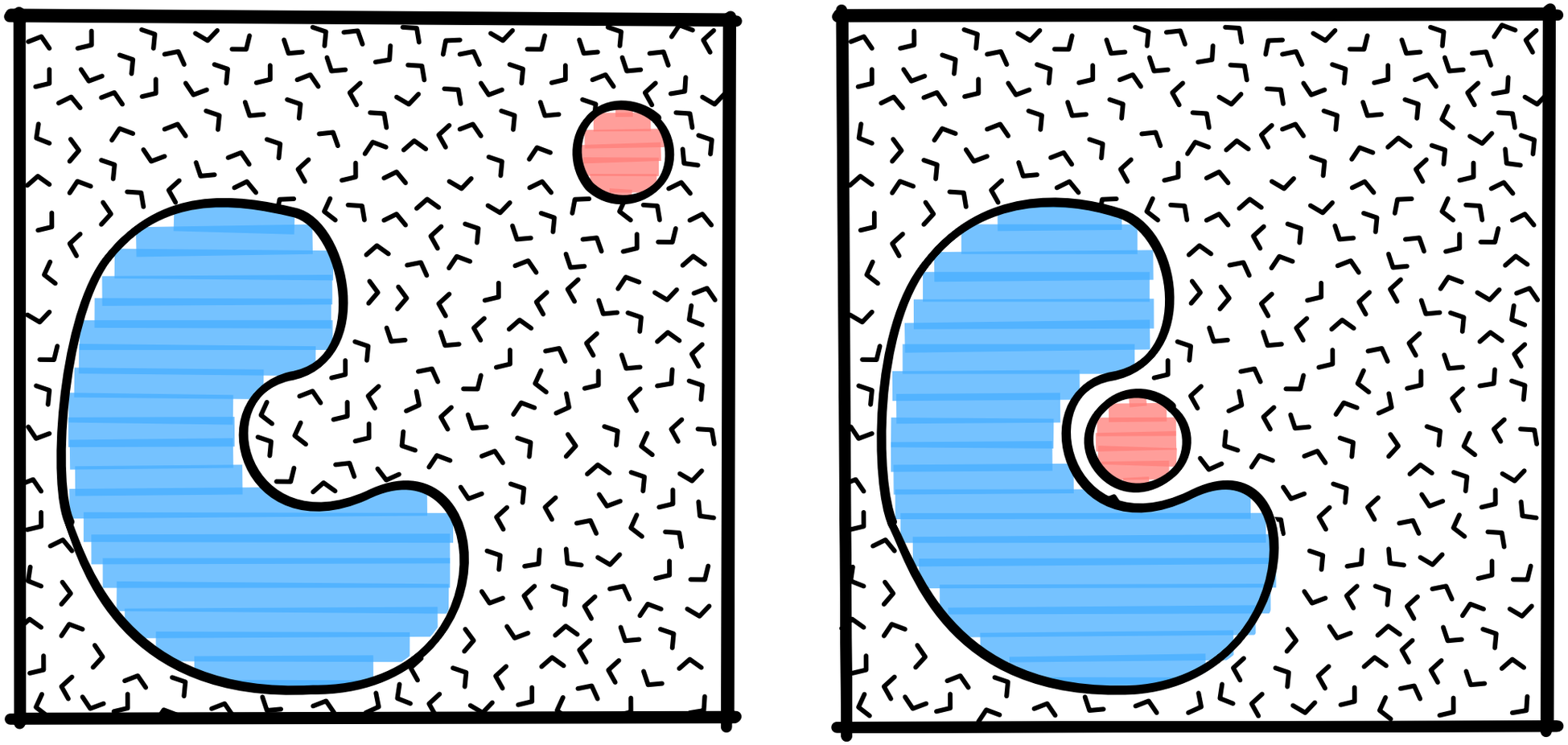
& small ligand

$$K_{eq} = \frac{Q_{N,LP}/Q_{N,P}}{(Q_{N,L}/Q_{N,0})/V_L}$$

Hill, *Cooperativity Theory in Biochemistry*, Springer Verlag 1985

2. Ligand Partitioning

$$L_{\text{free}} \rightleftharpoons L_{\text{bound}}$$

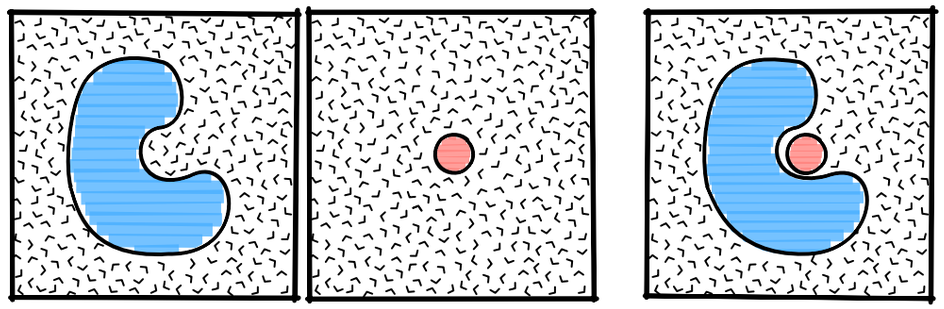


$$K_{eq} = \frac{Q_L(P)}{(Q_L/V_L)}$$

Rigorous Approach

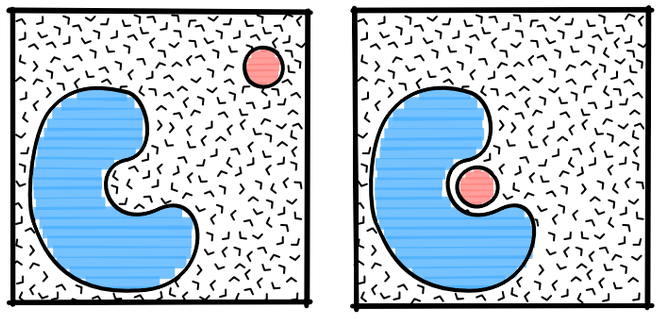
1. Absolute Chemical Potentials

$$\mu_i(V,T) = \mu_{i,v}(V,T) + W_{bulk}(\mathbf{X}_0)$$



2. Ligand Partitioning

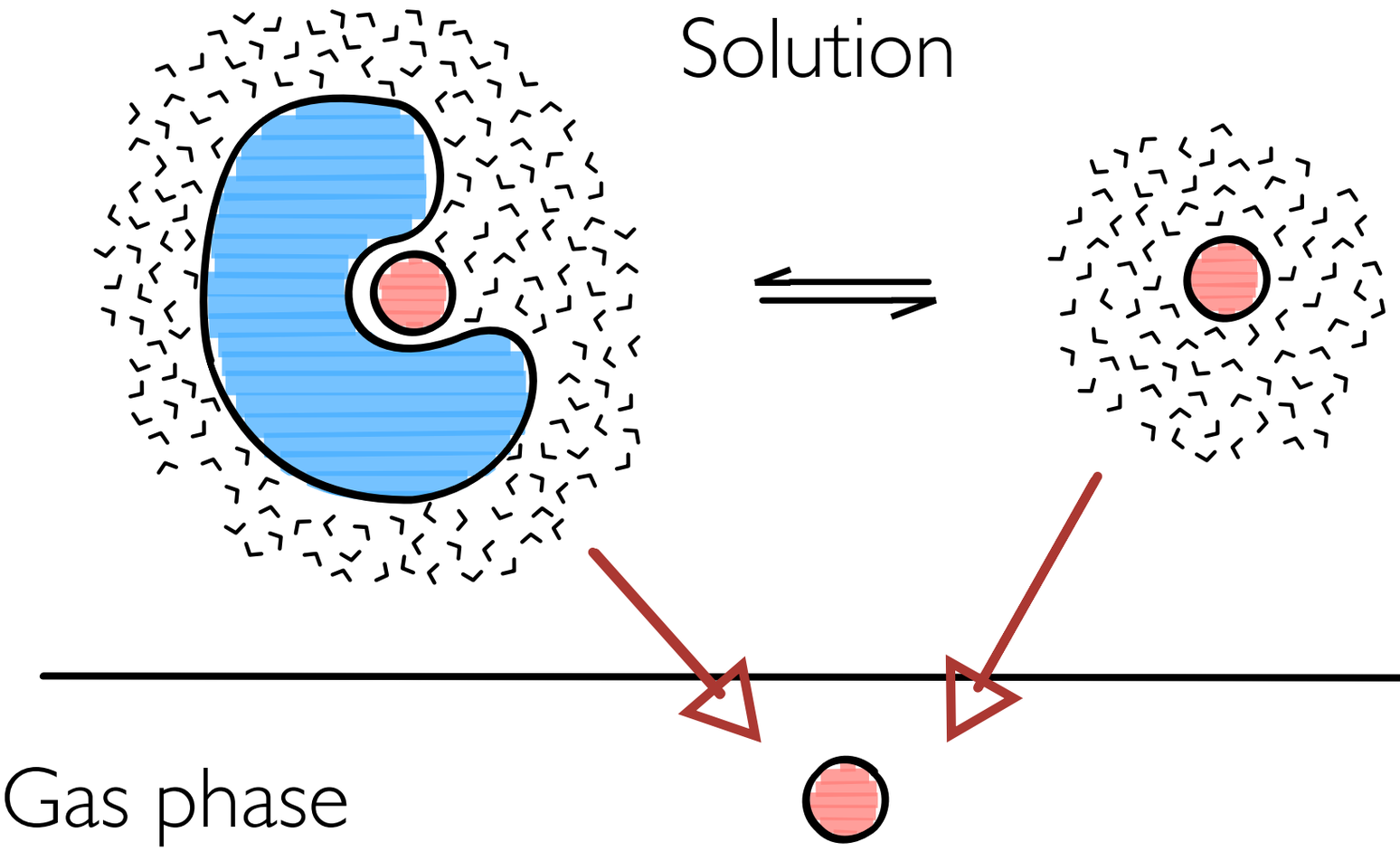
$$K_{eq} = \frac{\mathcal{Q}_L(P)}{(\mathcal{Q}_L/V_L)}$$



$$K_{eq} = \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left(-\beta U\right)}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L - \mathbf{r}^*\right) \int d\mathbf{X} \exp\left(-\beta U\right)}$$

$$K_{eq} = \frac{\mathcal{Q}_L(P)}{(\mathcal{Q}_L/V_L)}$$

Double Decoupling



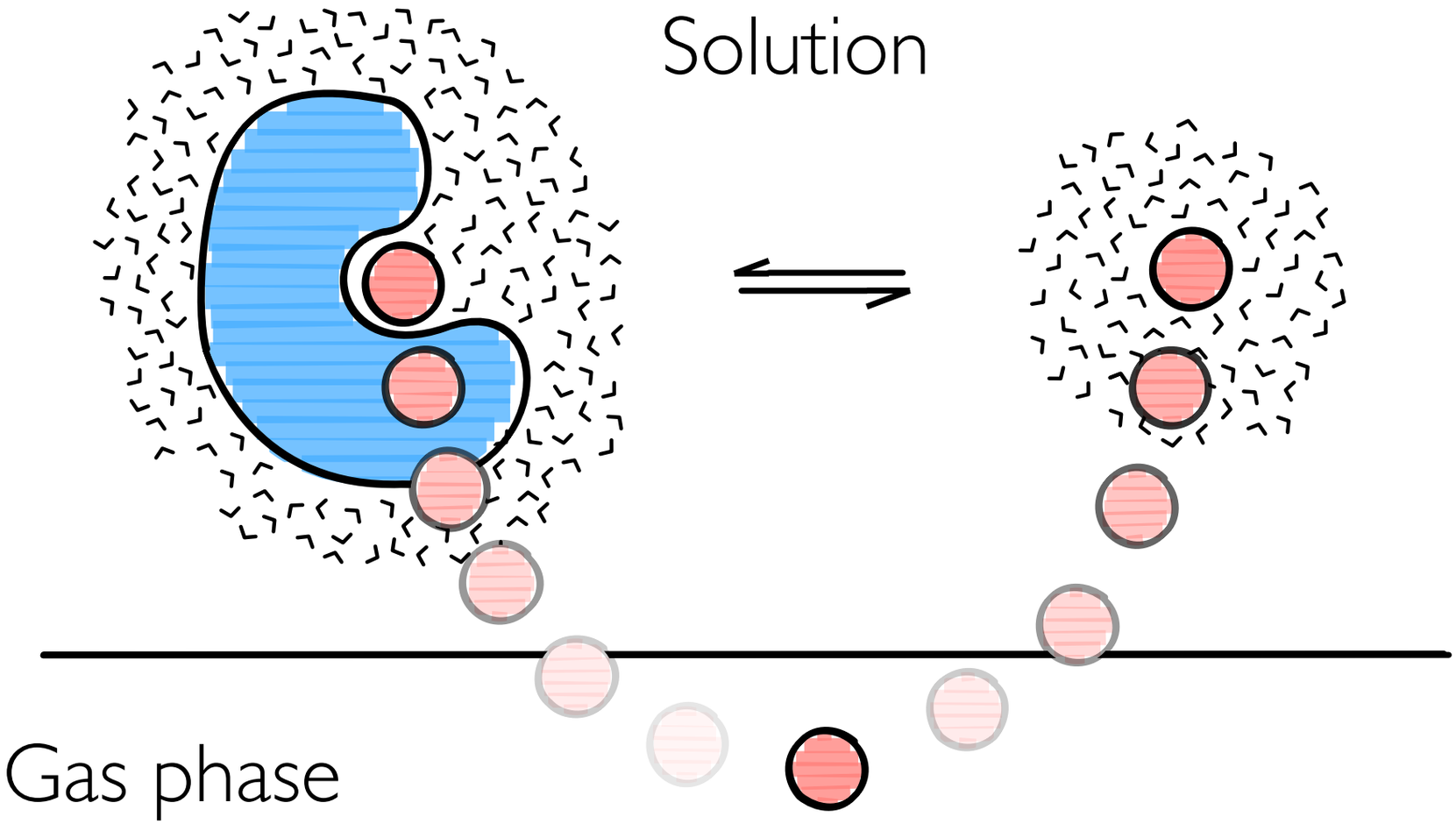
$$K_{eq} = \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left(-\beta U_1\right)}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L - \mathbf{r}^*\right) \int d\mathbf{X} \exp\left(-\beta U_0\right)} \times \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L - \mathbf{r}^*\right) \int d\mathbf{X} \exp\left(-\beta U_0\right)}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L - \mathbf{r}^*\right) \int d\mathbf{X} \exp\left(-\beta U_1\right)}$$

Jorgensen et al, *J Chem Phys* (1988)

Gilson et al, *Biophys J* (1997)

$$K_{eq} = \frac{\mathcal{Q}_L(P)}{(\mathcal{Q}_L/V_L)}$$

Alchemical Route



FEP

$$\Delta F = -kT \ln \left\langle e^{-\beta(U_1 - U_0)} \right\rangle_{U_0}$$

…with Restraints

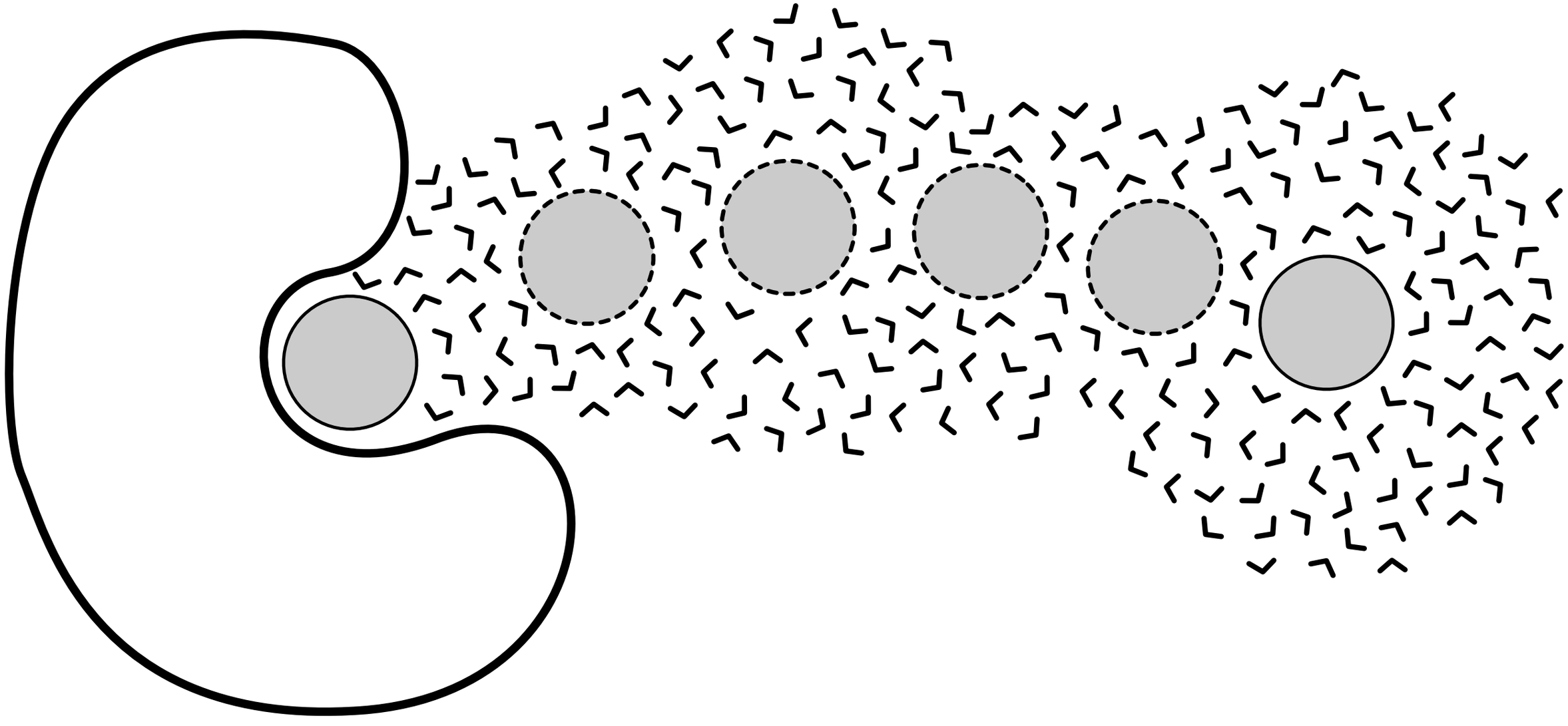
Uc, Uo, Up
conformational, orientational & positional restraints

2 alchemical transformations

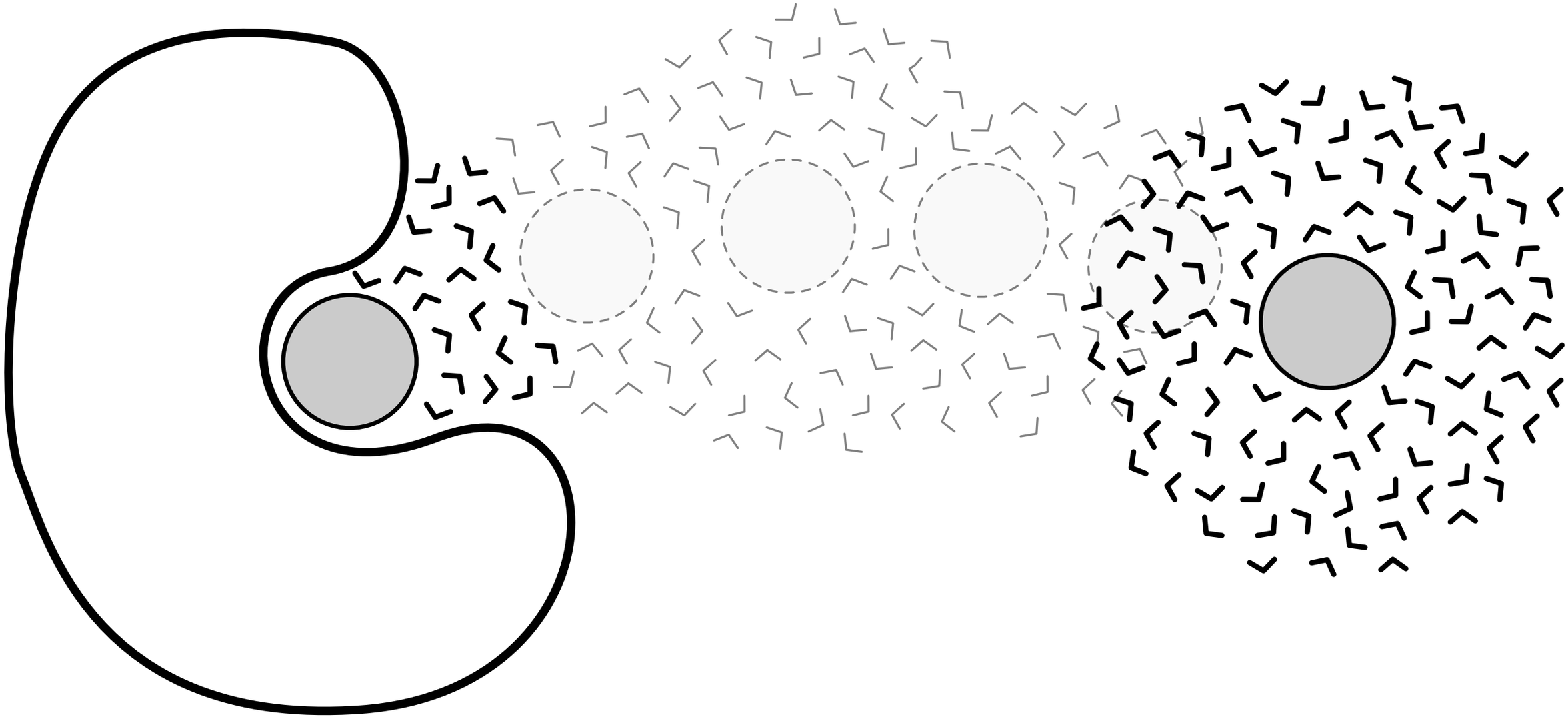
$$
\begin{aligned}
K_{eq} = & \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left(-\beta U_1\right)}{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]} \times \\
& \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]}{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o\right)\right]} \times \\
& \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o\right)\right]}{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o+u_p\right)\right]} \times \\
& \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o+u_p\right)\right]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o+u_p\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o+u_p\right)\right]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o\right)\right]]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c\right)\right]]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_1\right)\right]}
\end{aligned}
$$

Gumbart, Chipot & Roux, *J Chem Theory Comput* (2013)

Rigorous Methods

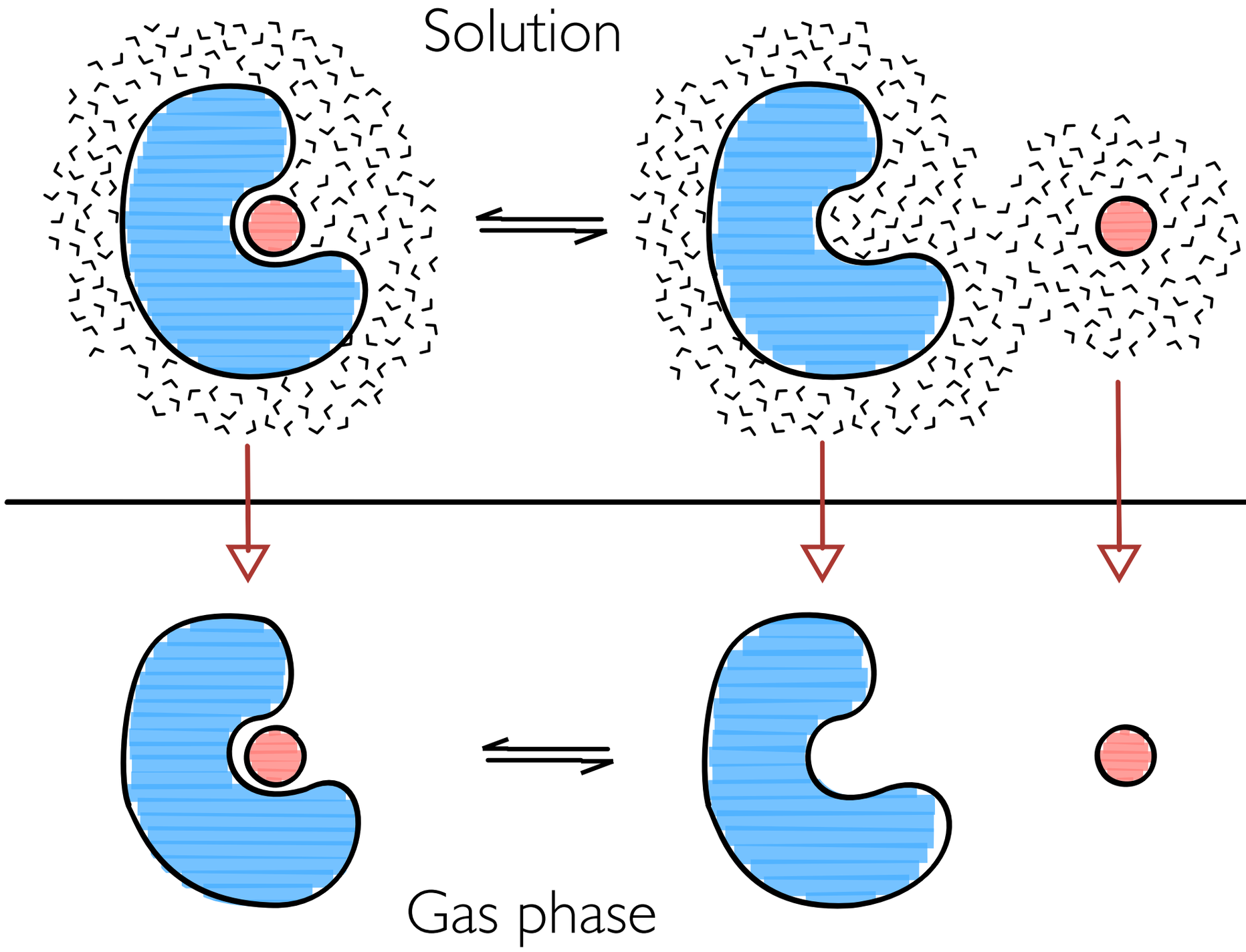


End-points Approach



$$\mu_i(V,T) = \mu_{i,v}(V,T) + W_{bulk}(\mathbf{X}_0)$$

MM/PBSA



Kollman *et al*, *Acc Chem Res* (2000)

$$\mu_i(V,T) = \mu_{i,v}(V,T) + W_{bulk}(\mathbf{X}_0)$$

MM/PBSA

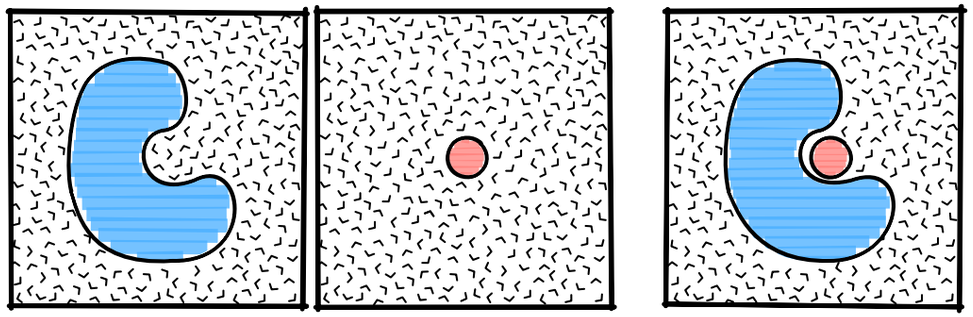
By separating out enthalpy vs entropy contributions

$$\mu_i(V,T) = (3n-3)\,kT - D_{e,v} - TS_i(V) + W_{solv}$$

MM/PBSA assumptions

1. $W_{solv} \approx G_{\mathrm{PBSA}}$ implicit solvent
2. $-D_{e,v} = \langle U \rangle - \sum^{\kappa} \frac{1}{2} kT$ Force Field

$$\mu_i(V,T) = \frac{3}{2} n_i kT + \langle U \rangle + \langle G_{\mathrm{PBSA}} \rangle - TS_i(V)$$

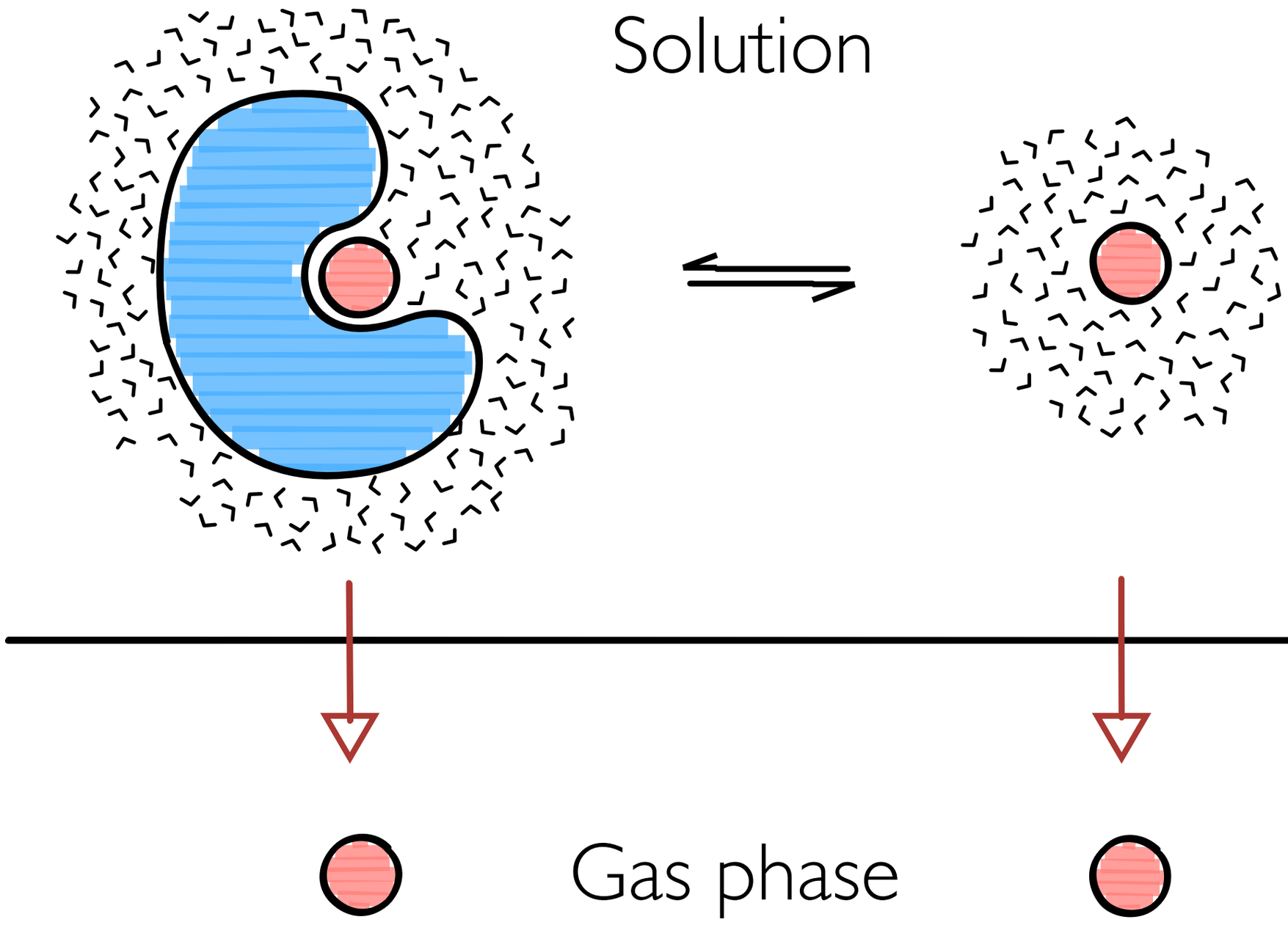


MM/PBSA (HOWTO)

$$\mu_i(V,T) = \tfrac{3}{2} n_i kT + \overline{E}_{bond} + \overline{E}_{elec} + \overline{E}_{vdW} + \overline{G}_{pol} + \overline{G}_{np} - TS_i(V)$$

- Run explicit-water MD for P, L & PL
- Extract sampling & compute ensemble averages of both the internal energy in a vacuum & the solvation free energy
- Evaluate the configurational entropy at standard state (1M) by statistical mechanics formulas & NMA

Linear Interaction Energy (LIE)



$$K_{eq} = \frac{\mathcal{Q}_L(P)}{(\mathcal{Q}_L/V_L)}$$

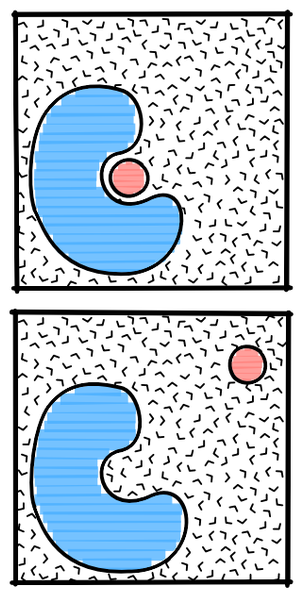
Åqvist, Medina & Samuelsson, *Protein Eng* (1994)

$$K_{eq} = \frac{\mathcal{Q}_L(P)}{(\mathcal{Q}_L/V_L)} \qquad q = q_{\mathrm{tr}}\, q_{\mathrm{rot}}\, q_{\mathrm{vib}}\, q_{\mathrm{elec}}$$

LIE

In the limit of RRHO & rigid ligands:

$$K_{eq} = \frac{q_{\mathrm{CM}}\, q_{\mathrm{rock}}\, e^{-\beta W_{site}}}{(q_{\mathrm{tr}}/V)\, q_{\mathrm{rot}}\, e^{-\beta W_{bulk}}}$$



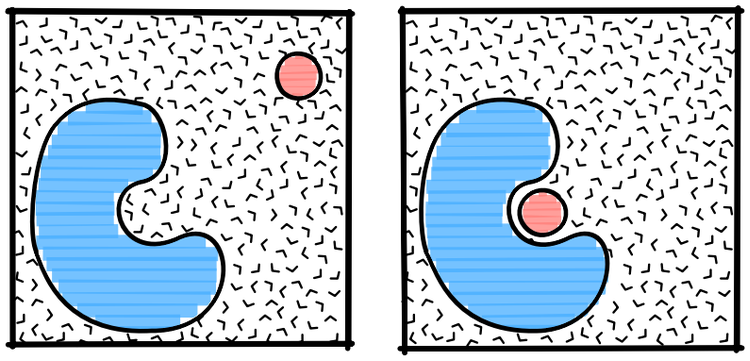
$$\Delta\mu_b^\circ = -kT \log \zeta + W_{site} - W_{bulk}$$

LIE's assumptions

1. $W_i^{pol} = \frac{1}{2}\left\langle U_{l/s}^{elec} \right\rangle$

2. $W_i^{np} \approx \alpha \left\langle U_{l/s}^{vdW} \right\rangle$

$$\Delta\mu_b^\circ = \tfrac{1}{2}\left[\langle U_{l/s}^{elec}\rangle_{site} - \langle U_{l/s}^{elec}\rangle_{bulk}\right] + \alpha\left[\langle U_{l/s}^{vdw}\rangle_{site} - \langle U_{l/s}^{vdw}\rangle_{bulk}\right] + \gamma$$



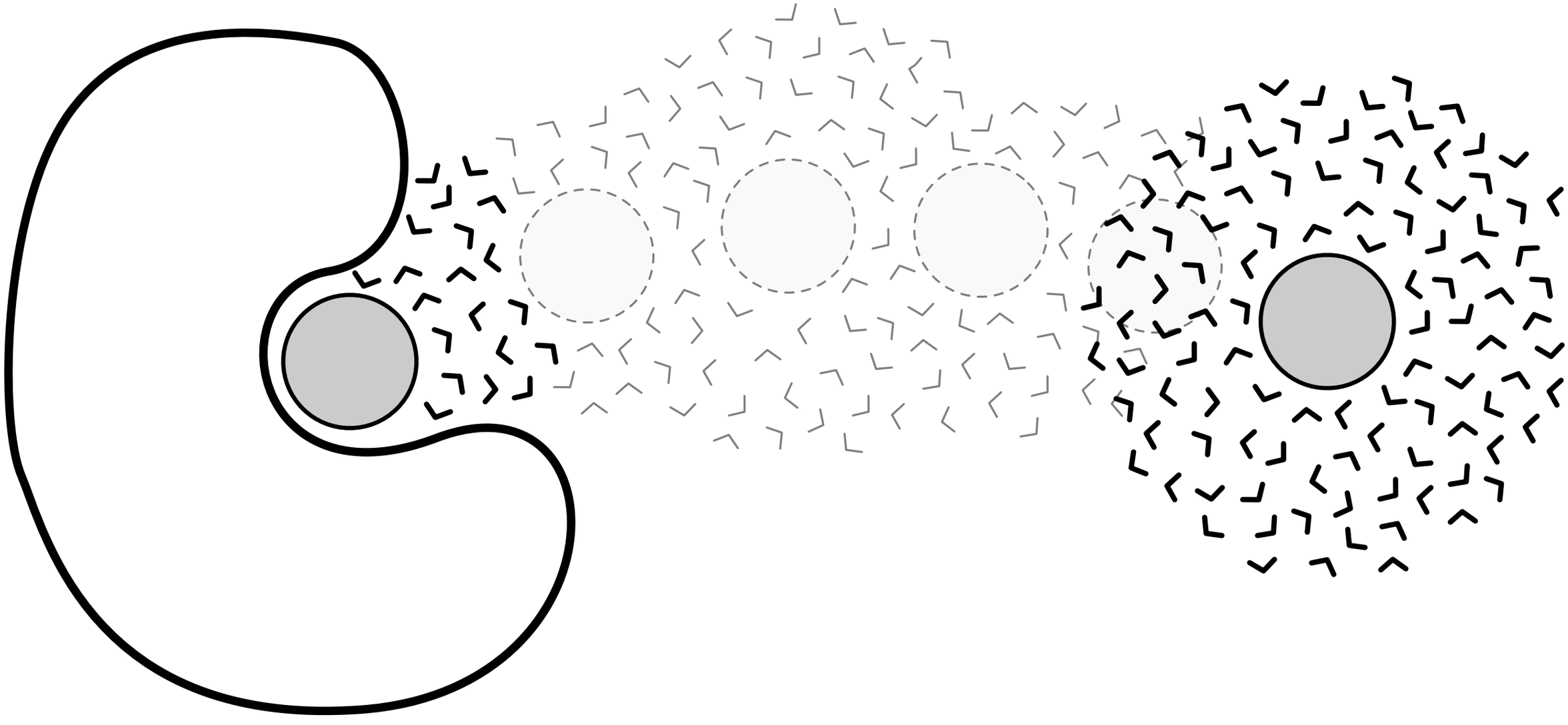
LIE (HOWTO)

$$\Delta\mu_b^\circ = \beta\left[\langle U_{l/s}^{elec}\rangle_{site} - \langle U_{l/s}^{elec}\rangle_{bulk}\right] + \alpha\left[\langle U_{l/s}^{vdw}\rangle_{site} - \langle U_{l/s}^{vdw}\rangle_{bulk}\right] + \gamma$$

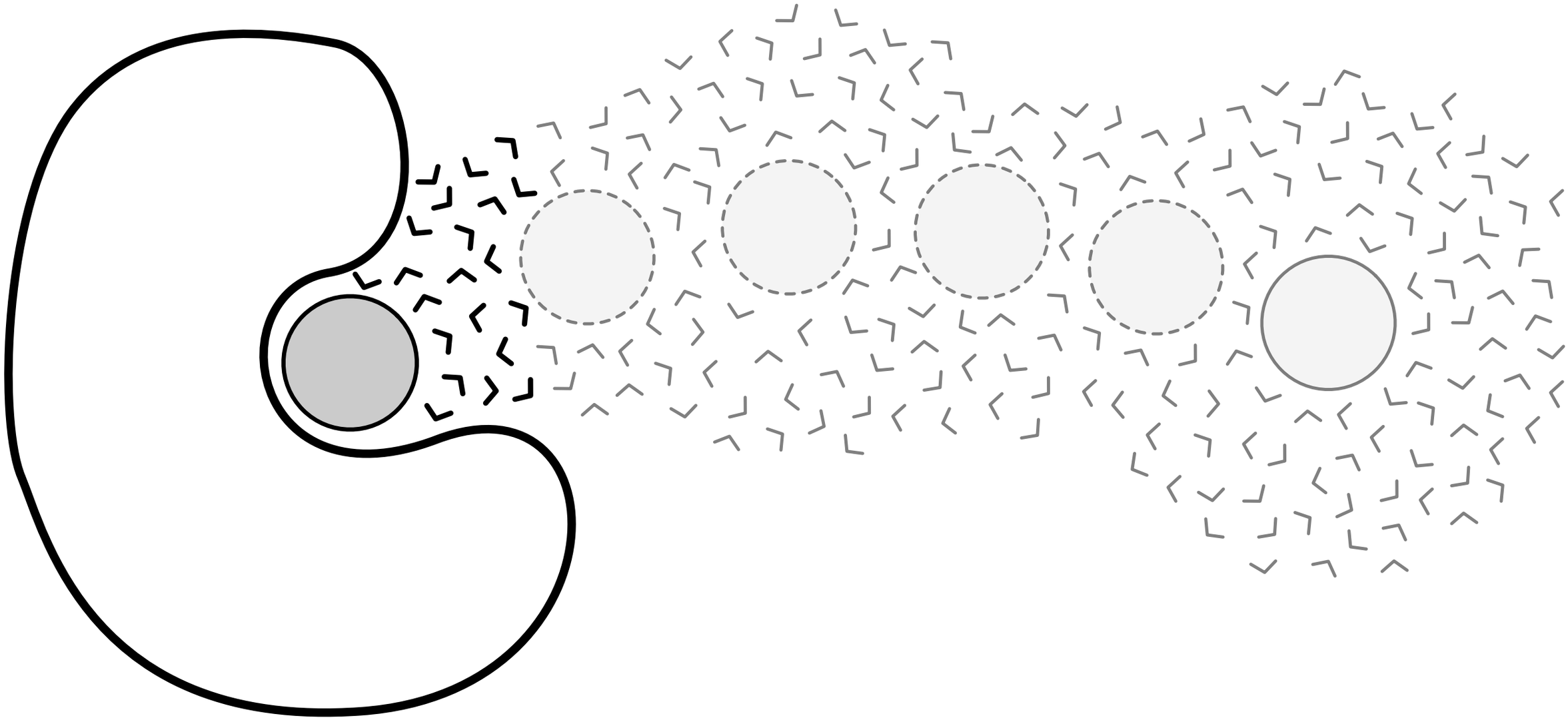
- Run two explicit-water MD for L & PL
- Compute ensemble averages of the electrostatic & the Van Der Waals interaction energy of the ligand with the surroundings
- Assign appropriate **β** for your ligand(s)
- Determine **α**,**γ** by fitting on experiments

Guitiérrez-de-Terán & Åqvist, *Computational Drug Discovery and Design* (2012)

End-points Approach



Empirical Approach

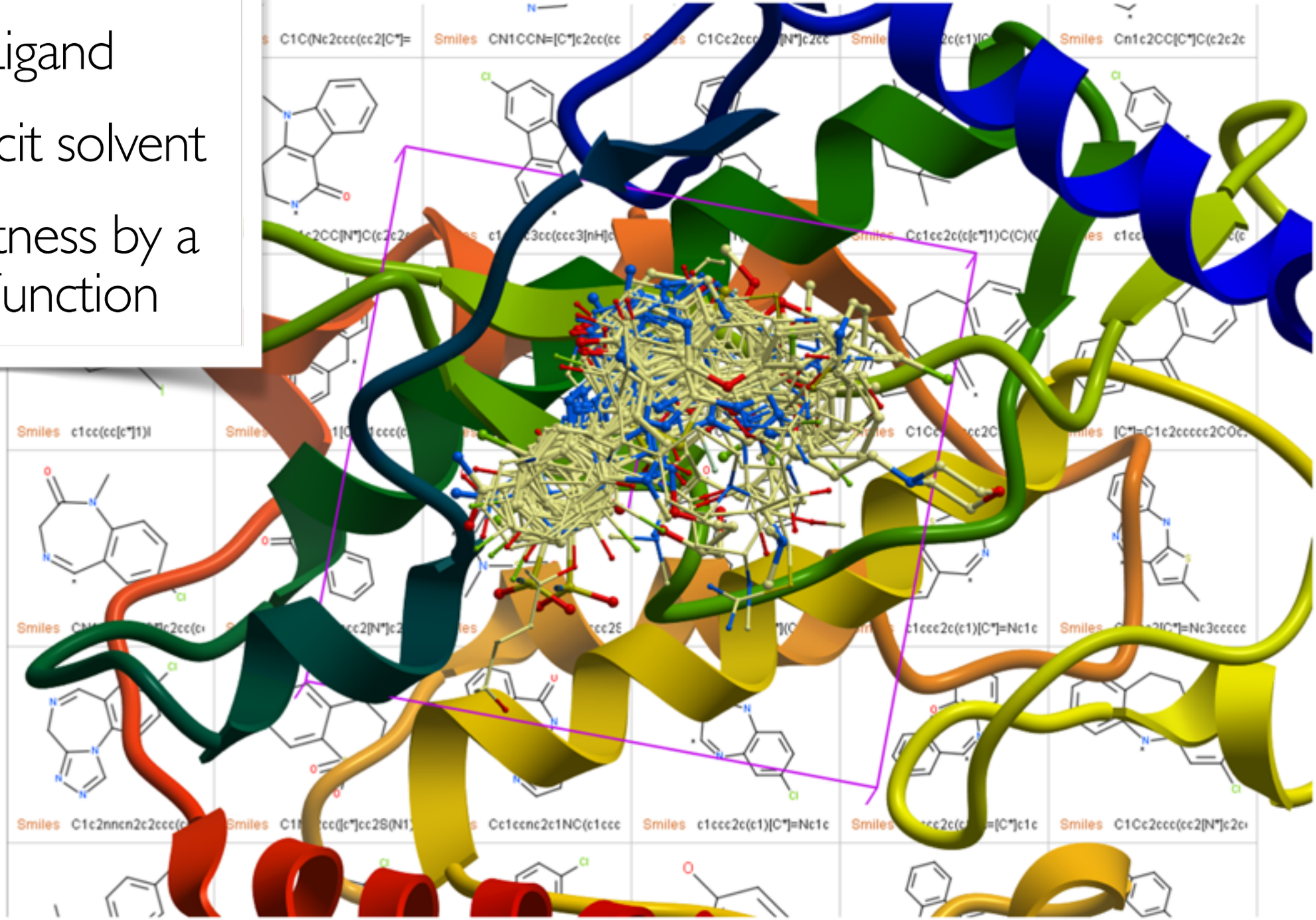


virtual High-Throughput Screening (vHTS)

Molecular Docking

BUT to be fast…

- Bound State
- Rigid Receptor
- Flexible Ligand
- No/Implicit solvent
- Ligand Fitness by a Scoring Function



Force-Field Scoring

DOCK score:

$$\Delta\mu_b^\circ = \sum_i^{prot}\sum_j^{lig}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332.0\frac{q_i q_j}{\epsilon\left(r_{ij}\right) r_{ij}}\right)$$

FF's assumptions

$$\langle \ldots \rangle = (\ldots)$$ NO sampling

$$\left(U_{l/s}^{elec}\right)_{bulk} = \left(U_{l/s}^{vdw}\right)_{bulk} = 0$$ NO desolvation

$$\Delta\mu_b^\circ = \left[\left(U_{l/s}^{elec}\right)_{site} - 0\right] + \left[\left(U_{l/s}^{vdw}\right)_{site} - 0\right] = \sum_i^{prot}\sum_j^{lig}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\epsilon\left(r_{ij}\right) r_{ij}}\right)$$

Empirical Scoring

$$\Delta\mu_b^\circ = \sum_i W_i \Delta\mu_i$$

Böhm's score:

$$\begin{aligned}
\Delta\mu_b^\circ \quad = \quad & \Delta G_0 + \Delta G_{hb} \sum_{hbonds} f\left(\Delta R, \Delta\alpha\right) \quad + \\
& \Delta G_{io} \sum_{io\ int.} f\left(\Delta R, \Delta\alpha\right) \quad + \\
& \Delta G_{lipo} \sum_{lipo\ cont.} A_{lipo} \quad + \\
& \Delta G_{aro} \sum_{aro\ int.} f\left(\Delta R\right) + \Delta G_{rot} \times N_{rot}
\end{aligned}$$

FRESNO's score:

$$\Delta\mu_b^\circ = K + \alpha\left(HB\right) + \beta\left(LIPO\right) + \gamma\left(ROT\right) + \delta\left(BP\right) + \epsilon\left(DESOLV\right)$$

Summary

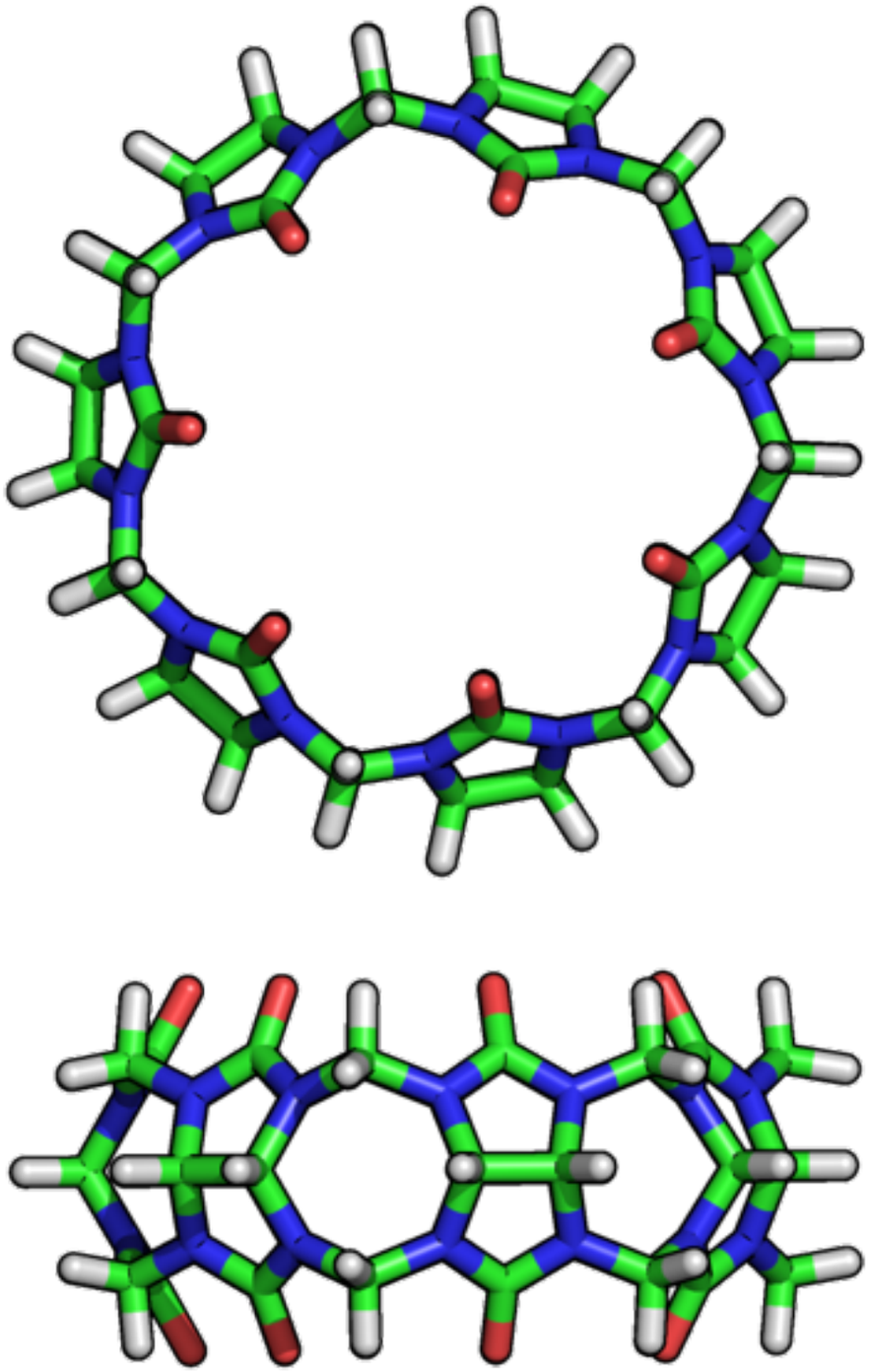
Class of Methods	*1.* Absolute Chemical Potentials	**2.** Ligand Partition Equilibrium	Focus	Context (No. compounds)
Rigorous ($week^{-1}$)		DDM **FEP/PMF** DAM	Full Reaction Path	*lead optimization* ($10\text{-}10^2$)
End-points (day^{-1})	QM/MM quasi-harmonic **MM/PBSA** MM/GBSA one-average	QMLIECE **LIE** LIE(α,β) LIECE	Bound & Unbound States	*hit-to-lead* ($10\text{-}10^3$)
Empirical (sec^{-1})		Dock AutoDock **FF** Böhm Fresno **ES**	Bound State	*hit identification* ($10^4\text{-}10^6$)

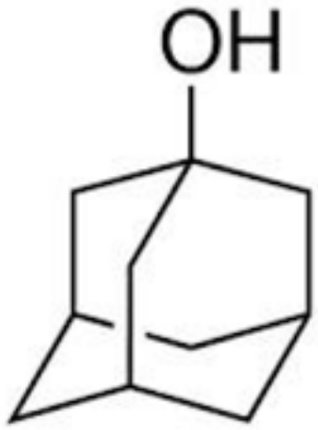
Accuracy ↑ ↓ *Efficiency*

Montalvo-Acosta & Cecchini, *Mol Inform* (2016)

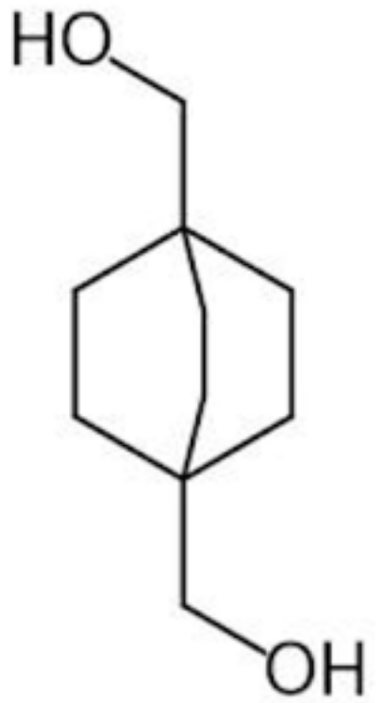
Host-Guest System

Cucurbit-[7]-uril (CB7)

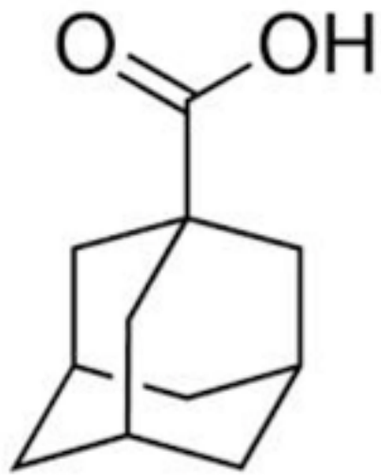




−14.1

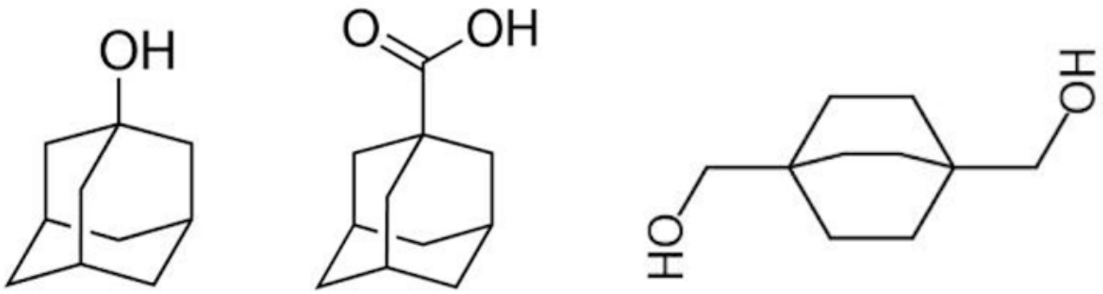


−13.4



−11.8

Moghaddam *et al*, *J Am Chem Soc* (2011)



Benchmark

FEP

$$
\begin{aligned}
K_{eq} = & \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left(-\beta U_1\right)}{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]} \times \\
& \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]}{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o\right)\right]} \times \\
& \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o\right)\right]}{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o+u_p\right)\right]} \times \\
& \frac{\int_{site} d\mathbf{L} \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c+u_o+u_p\right)\right]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o+u_p\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o+u_p\right)\right]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c+u_o\right)\right]]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_0+u_c\right)\right]]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]} \times \\
& \frac{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_1+u_c\right)\right]]}{\int_{bulk} d\mathbf{L}\, \delta\left(\mathbf{r}_L-\mathbf{r}^*\right) \int d\mathbf{X} \exp\left[-\beta\left(U_1\right)\right]}
\end{aligned}
$$

MM/PBSA

$$\mu_i(V,T) = \frac{3}{2} n_i kT + \langle U \rangle + \langle G_{\text{PBSA}} \rangle - TS_i(V)$$

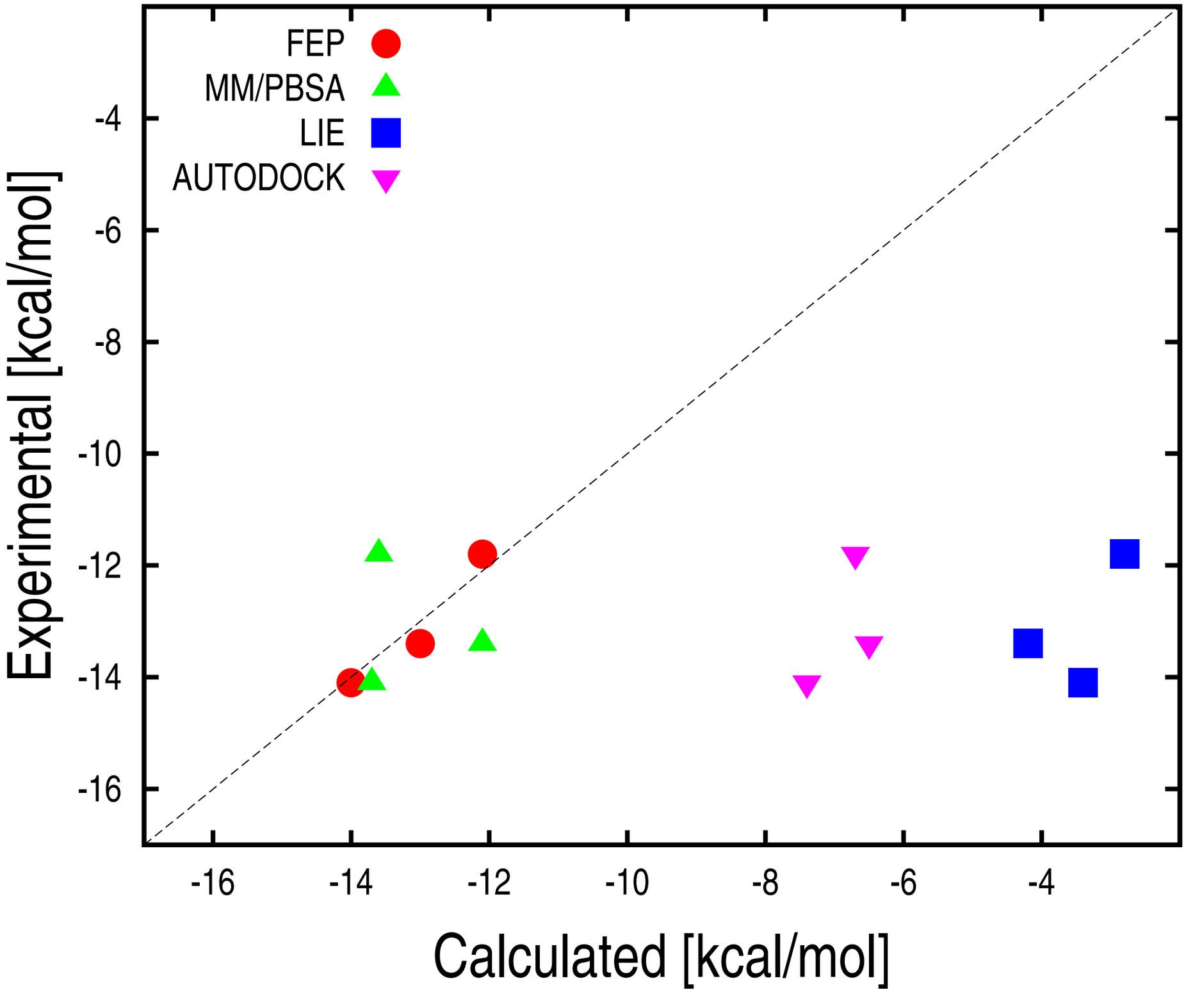
LIE

$$
\begin{aligned}
\Delta\mu_b^\circ = & \frac{1}{2}\left[\langle U_{l/s}^{elec}\rangle_{site} - \langle U_{l/s}^{elec}\rangle_{bulk}\right] + \\
& \alpha\left[\langle U_{l/s}^{vdw}\rangle_{site} - \langle U_{l/s}^{vdw}\rangle_{bulk}\right]
\end{aligned}
$$

AUTODOCK

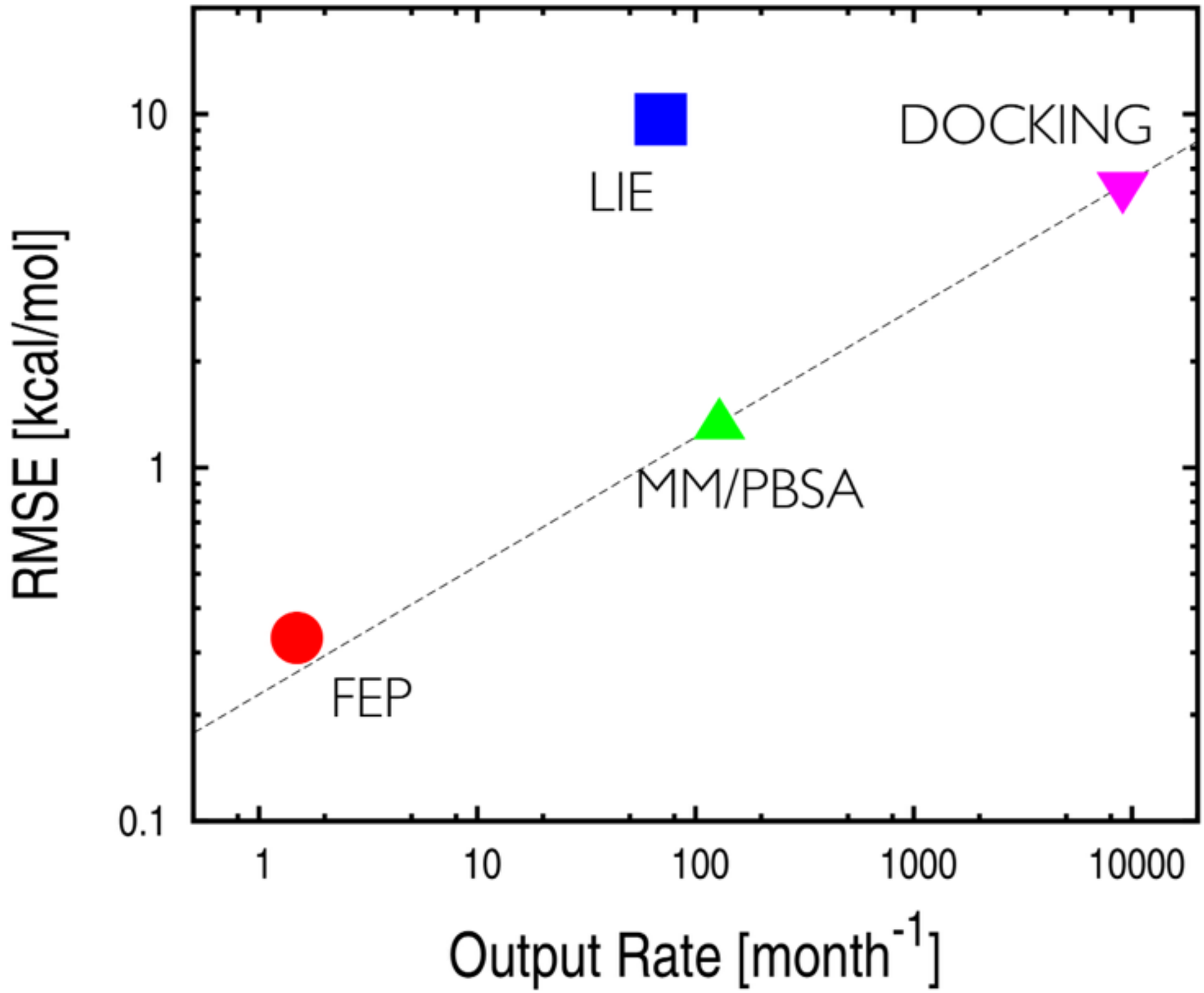
$$
\begin{aligned}
\Delta\mu_b^\circ = & W_{vdw} \sum_i^{prot} \sum_j^{lig} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \\
& W_{ele} \sum_i^{prot} \sum_j^{lig} \left(\frac{q_i q_j}{\epsilon\left(r_{ij}\right) r_{ij}}\right) + \\
& W_{hbond} \sum_i^{prot} \sum_j^{lig} \left(E\left(t\right) \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + \\
& W_{sol} \sum_i^{prot} \sum_j^{lig} \left(S_i V_j + S_j V_i\right) \exp\left(\frac{-r_{ij}^2}{2\sigma^2}\right)
\end{aligned}
$$

Binding Free Energy



Accuracy/Efficiency

$$y = 0.22\sqrt[3]{x}$$



Take Home Message

Our statistical mechanics interpretation of protein-ligand binding:

- provides a useful classification of existing methods to the binding constant
- highlights their inherent approximations and provides guidelines for future development
- has allowed to quantify their performances (accuracy/efficiency) on a model host-guest system

Montalvo-Acosta & Cecchini, *Mol Inform* (2016)

Acknowledgments

- Joel Montalvo-Acosta

- Nicolas Muzet (Sanofi)
- Michael Schaefer (Novartis)

EQUIP@MESO

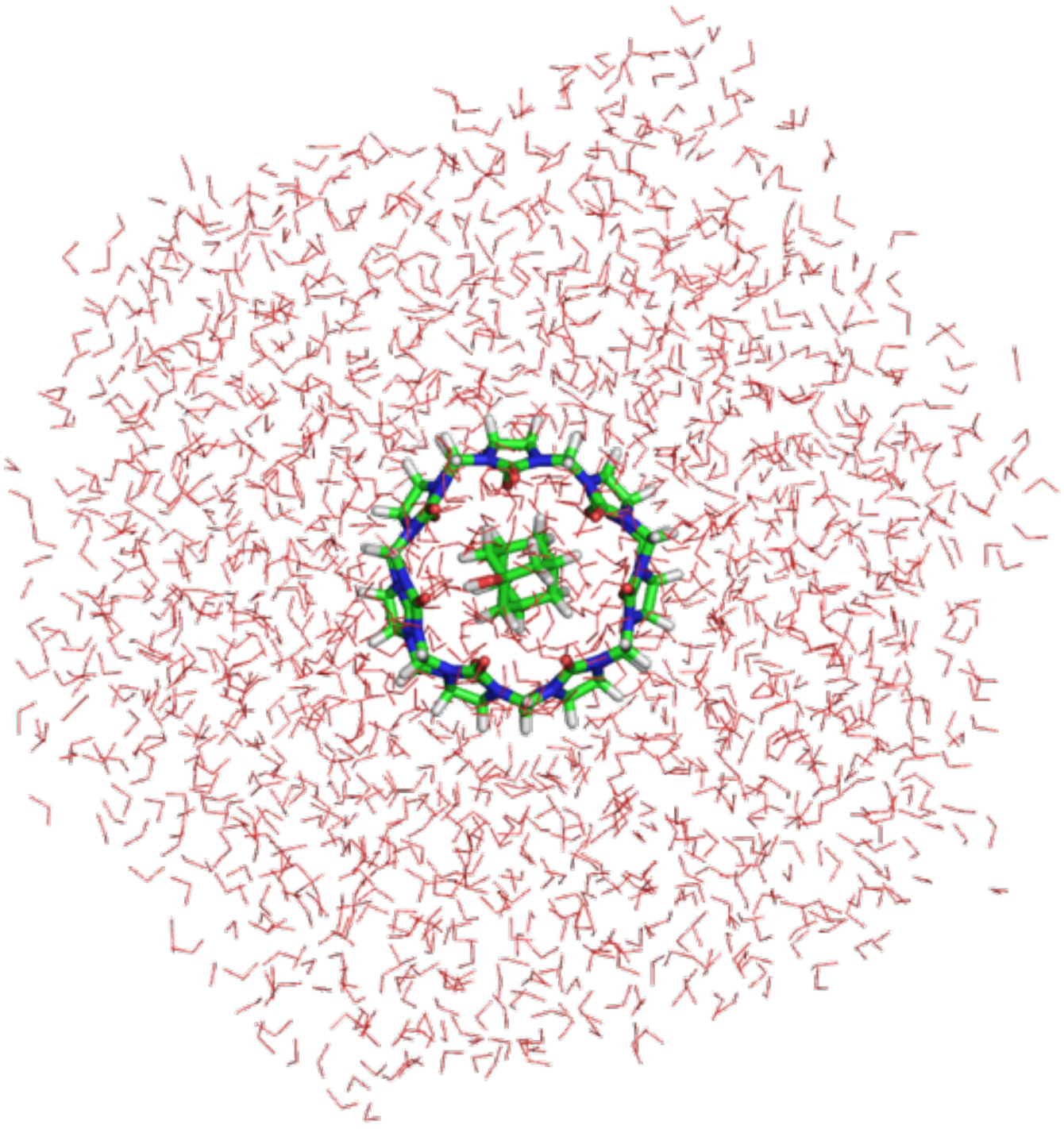
FRC
International Center for Frontier Research in Chemistry

INVESTISSEMENTS D'AVENIR

GENCI

PRACE

Simulation Setup



- Explicit-water MD simulations for three host-guest systems
- Initial coordinates for the complexes obtained from the Cambridge Structural Database
- Dodecahedron box with a layer of >14Å around the solute
- Force-Field parameters for the guests by CGenFF
- NPT @ 298K & 1atm
- Production runs of 20 ns with GROMACS 5.1 compiled with PLUMED plugin