More Accurate Predictions of Protein-Ligand Affinities in Pharmaceutical Research ?

Michael Brunsteiner & Nicolas Foloppe

What we want ...

- assist lead optimization
- predict rel. affinities in congeneric series
- performance: 10s of compounds in days
- accuracy: one order/magnitude (K_d)

What we want ...

- assist lead optimization
- predict rel. affinities in congeneric series
- performance: 10s of compounds in days
- accuracy: one order/magnitude (K_d)

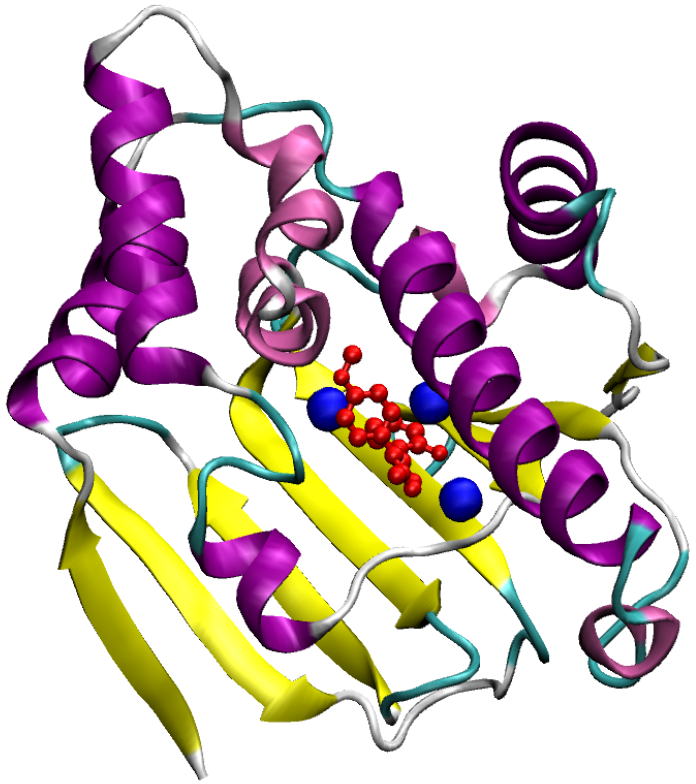
$RL_i \rightleftharpoons R + L_i \qquad K_d = \frac{[R][L_i]}{[RL_i]}$

$\Delta(\Delta G^\circ) = -RT \ln(K_d^1 / K_d^2) \approx 1.4$ kcal/mol

What we did ...

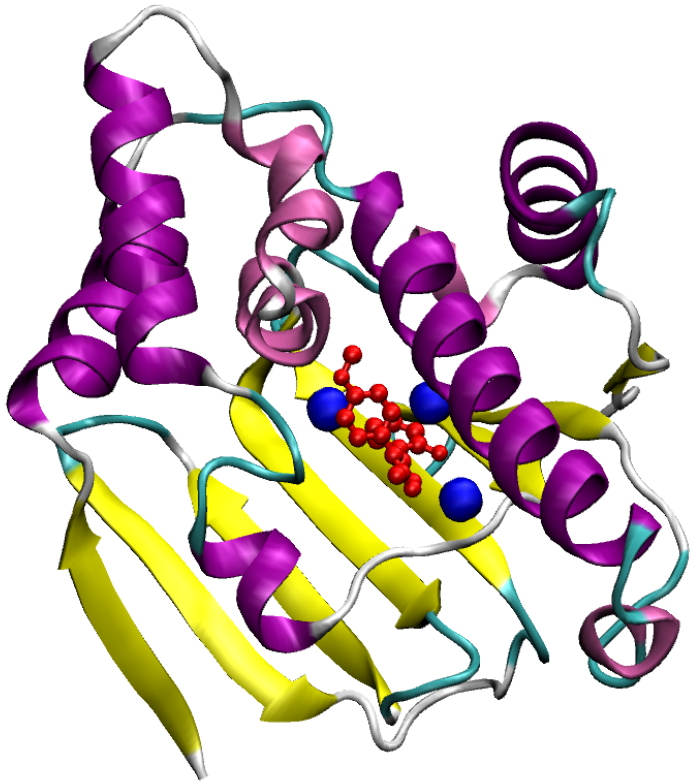
- Test free energy methodologies (MM/PBSA, LIE, TI)
- Use a large pharmaceutically relevant test set
- Good assay data + crystallographic support

The Test Case — Hsp90



- emerging oncology target
- structural data for ALL protein ligand complexes (Res.$<$2.5 Å)
- accurate binding data ($\pm$0.34 kcal/mol)

The Test Case — Hsp90



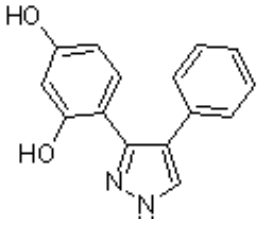
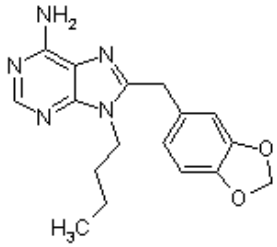
- emerging oncology target
- structural data for ALL protein ligand complexes (Res.$<$2.5 Å)
- accurate binding data ($\pm$0.34 kcal/mol)

real life example ...

- water mediating ligand/receptor interactions
- flexible protein (3 known conformations)

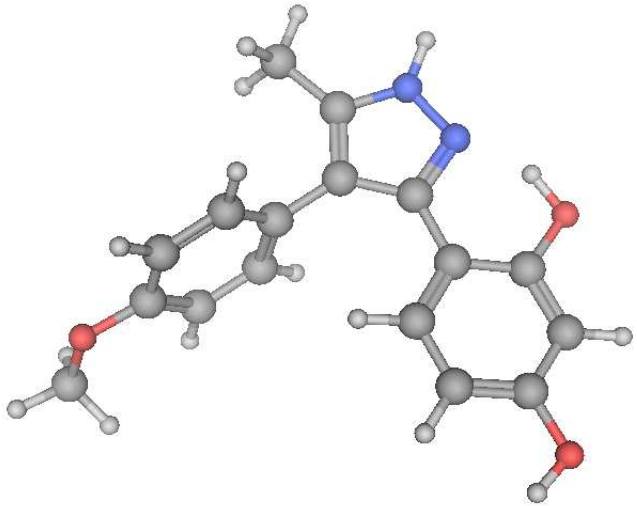
Ligands: 3 Congeneric Series

curated ligand test sets:

ID	chemistry	# of cmpds. (charged)	affinity range [$\log_{10}(IC_{50})$]	Conf.	example
A	Resorcinol	32 (16)	5.5	C	
B	PU3	17 (0)	3.2	H	
C	—	28 (7)	3.0	H	

Ligand Preperation

- tautomers: assigned based on structures
- partial charges: ESP, HF 6-31G*
- FF parameters: Momany-Rone FF, CHARMm



Methods Considered

physics based scoring

- Mol. Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA)
- Linear Interation Energies (LIE)

Methods Considered

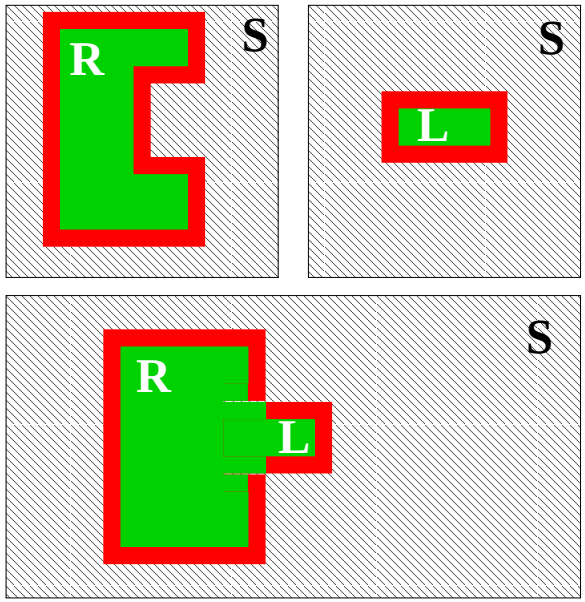
physics based scoring

- Mol. Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA)
- Linear Interation Energies (LIE)

rigorous (preliminary)

- Thermodynamic Integration (TI),

MM/PBSA



$$\begin{aligned}\Delta G &\approx \Delta U_{vdW}^{R-L} + \Delta U_{el}^{R-L} \\ &+ G_{solv}^{R-L} - G_{solv}^{L} - G_{solv}^{R} \\ &[+ T\Delta S_{\nu}]\end{aligned}$$

- direct interaction + desolvation + entropy
- continuum solvent: $G_{solv} \approx G_{PB} + \gamma$ SA
- entropy: harmonic approximation ... difficult.

MM/PBSA, The Protocol

MM/PBSA, The Protocol

MM force field:

protein: **CHARMM22** or **Momany-Rone**

Minimization:

protein non-H atoms: **fixed** or **relaxed**

PB calculation:

CHARMm PBEQ, conservative param., $\epsilon_{\mathrm{slte}} =$ **1 - 3**

Born radii: **LJ-radii** vs **PARSE** vs **Nina *et al.***

non-polar solvation

$\gamma =$ **0.0 - 0.05 kcal/mol/Å**^{2}

EM (SPE) vs MD sampling

MM/PBSA, Compound Series A

single point energies -

calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3

MM/PBSA, Compound Series A

single point energies -

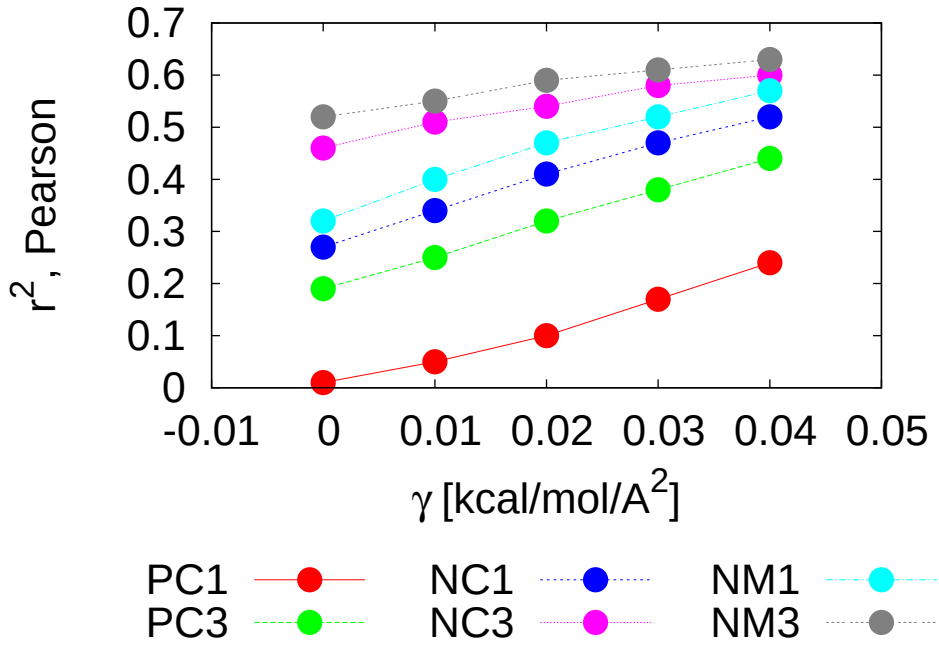
calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3



MM/PBSA, Compound Series A

single point energies -

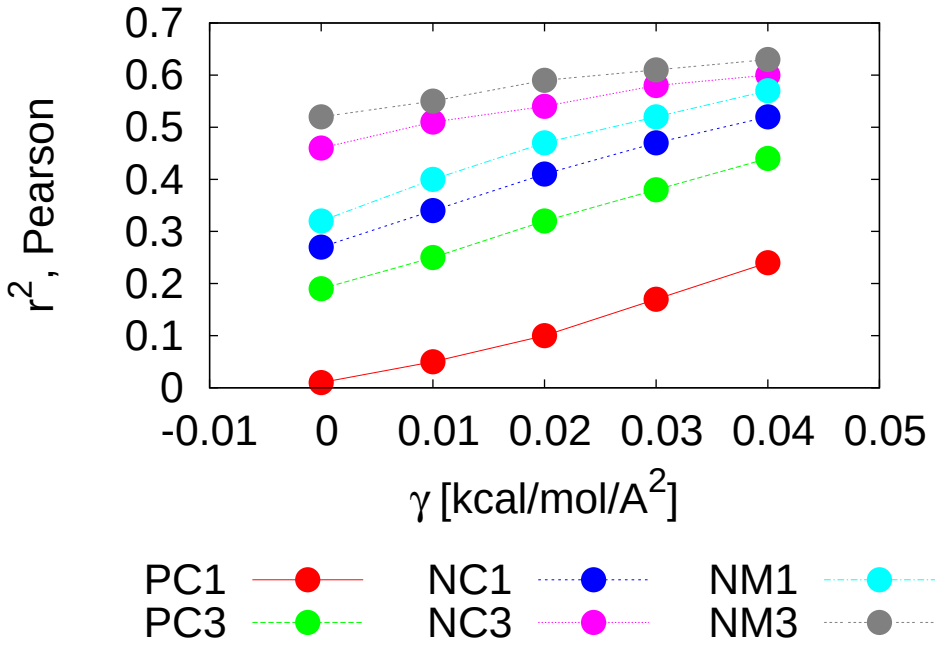
calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3



- MM/PBSA: best $r^2 = 0.63$ (Nina *et al* radii, Momany-FF, $\epsilon = 3$)
- better correlation with increasing $\gamma \leftarrow r^2(\Delta SA) = 0.73$!
- $r^2(\epsilon = 3) > r^2(\epsilon = 1)$
- Born radii from Nina *et al.* perform better than PARSE radii

MM/PBSA, Compound Series B

single point energies

calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3

MM/PBSA, Compound Series B

single point energies

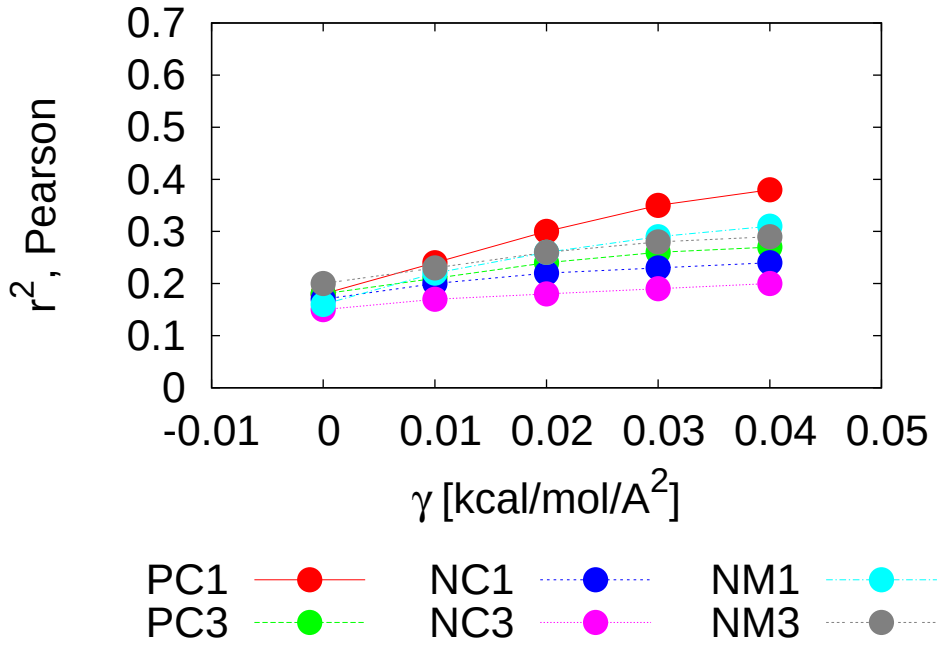
calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3



MM/PBSA, Compound Series B

single point energies

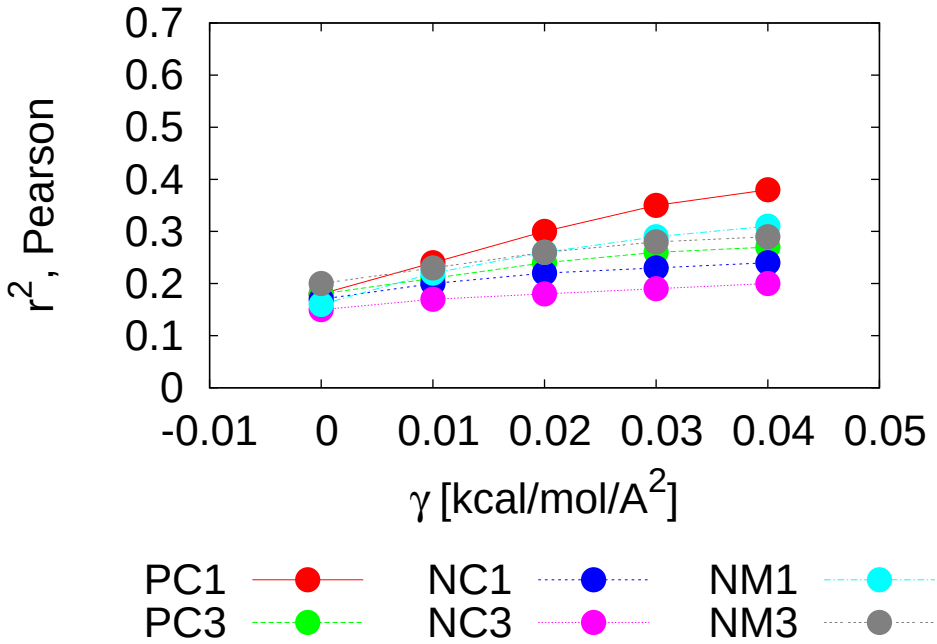
calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}: 1 vs 3



- best: r^2=0.38 (Parse radii, CHARMM22-FF, $\epsilon = 1$)
- better correlation with increasing γ, $r^2(\text{SA}) = 0.22$ (C), 0.34 (M)
- $r^2(\epsilon = 1) > r^2(\epsilon = 3)$
- PARSE radii perform better

MM/PBSA, Compound Series C

single point energies

calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3

MM/PBSA, Compound Series C

single point energies

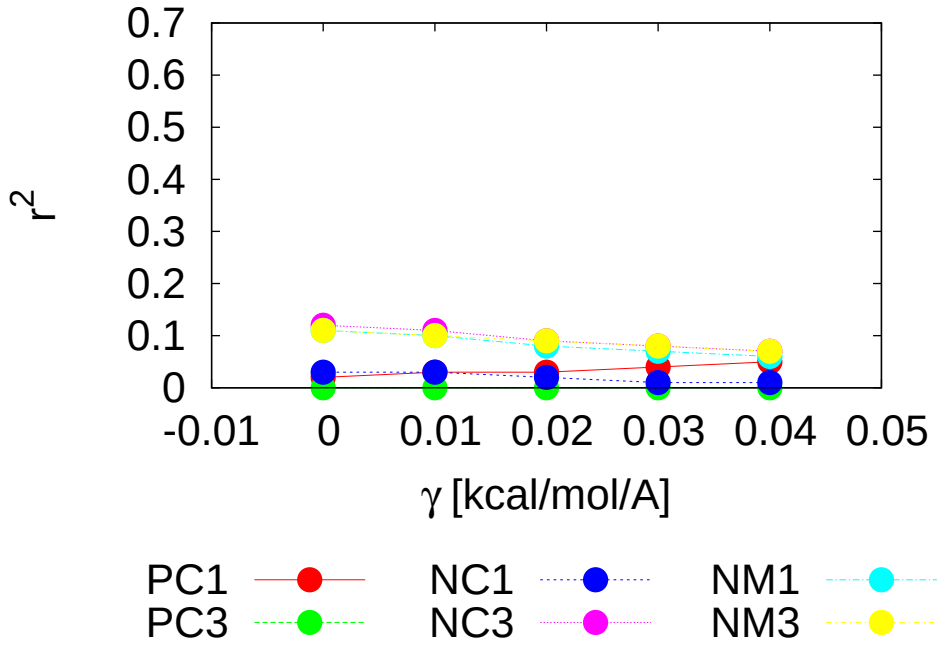
calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3



MM/PBSA, Compound Series C

single point energies

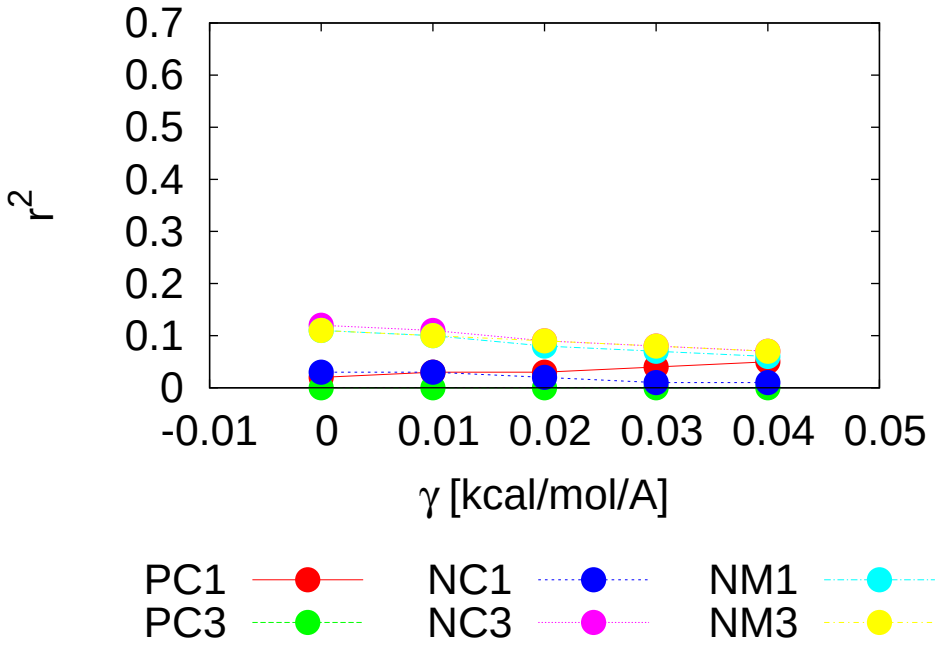
calculated vs experiment

radii: PARSE (P) vs Nina *et al* (N)

protein FF:

CHARMM22 (C) vs Momany (M)

ϵ_{solute}**:** 1 vs 3



- best: r^2=0.12 (N-radii, C-FF, $\epsilon = 3$)
- virtually NO corrlation

Does Sampling Help ?

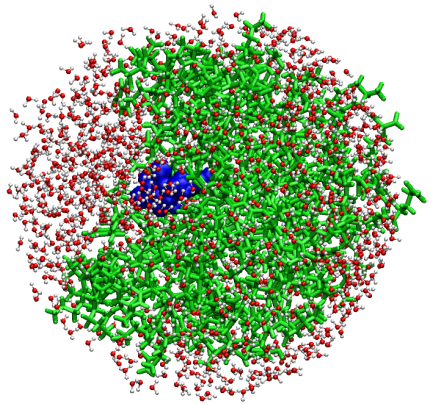
Setup:

- water droplet centered on ligand
- spherical boundary potential
- MD, 300 K, 5 nano-seconds (max)
- FF: CHARMM22, radii: Nina *et al.*
- $\gamma = 0.033$ kcal/mol/Å^2

Does Sampling Help ?

Setup:

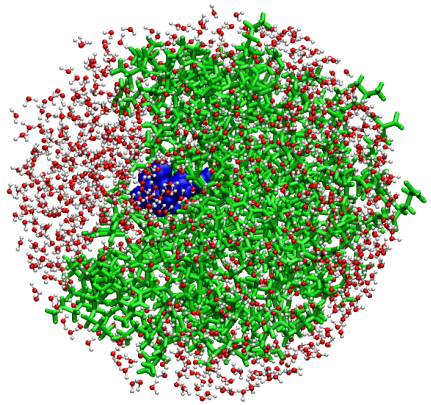
- water droplet centered on ligand
- spherical boundary potential
- MD, 300 K, 5 nano-seconds (max)
- FF: CHARMM22, radii: Nina *et al.*
- $\gamma = 0.033$ kcal/mol/Å^2



Does Sampling Help ?

Setup:

- water droplet centered on ligand
- spherical boundary potential
- MD, 300 K, 5 nano-seconds (max)
- FF: CHARMM22, radii: Nina *et al.*
- $\gamma = 0.033$ kcal/mol/Å^2

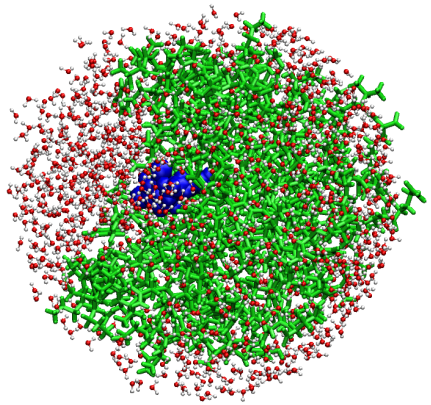


	A	B	C
MM/PBSA	0.52	0.24	0.01
MD/PBSA	**0.58**	**0.25**	**0.07**

Does Sampling Help ?

Setup:

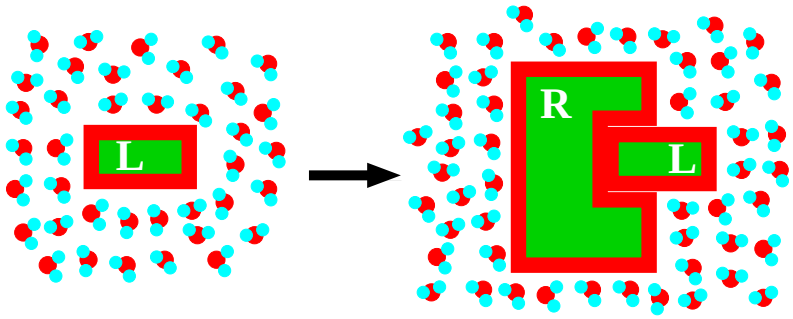
- water droplet centered on ligand
- spherical boundary potential
- MD, 300 K, 5 nano-seconds (max)
- FF: CHARMM22, radii: Nina *et al.*
- $\gamma = 0.033$ kcal/mol/Å^2



	A	B	C
MM/PBSA	0.52	0.24	0.01
MD/PBSA	**0.58**	**0.25**	**0.07**

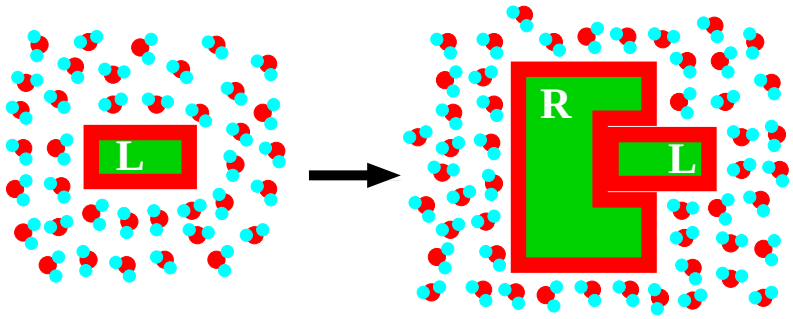
Sampling improves results marginally.

Linear Interaction Energies



$$\Delta G \approx \alpha \Delta U^{L}_{vdW} + \beta \Delta U^{L}_{el} \; (+\gamma)$$

Linear Interaction Energies



$$\Delta G \approx \alpha \Delta U^{L}_{vdW} + \beta \Delta U^{L}_{el} \; (+\gamma)$$

- explicit solvent
- ligand focussed
- empirical factors (weights)

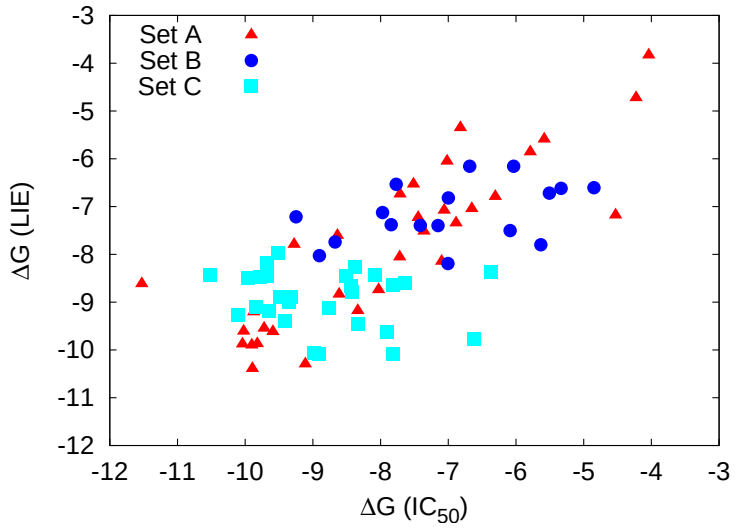
LIE - Absolute Energies

$\Delta G \propto \alpha\, \Delta U_{vdW} + \beta\, \Delta U_{el} + \gamma$ ..... fit α, β, γ

cmpd.	α	β	γ	r^2	MUE
A	0.16	0.00	-0.76	0.73	0.67
B	0.13	0.00	-1.42	0.17	0.97
C	0.14	0.00	-2.87	0.03	1.01

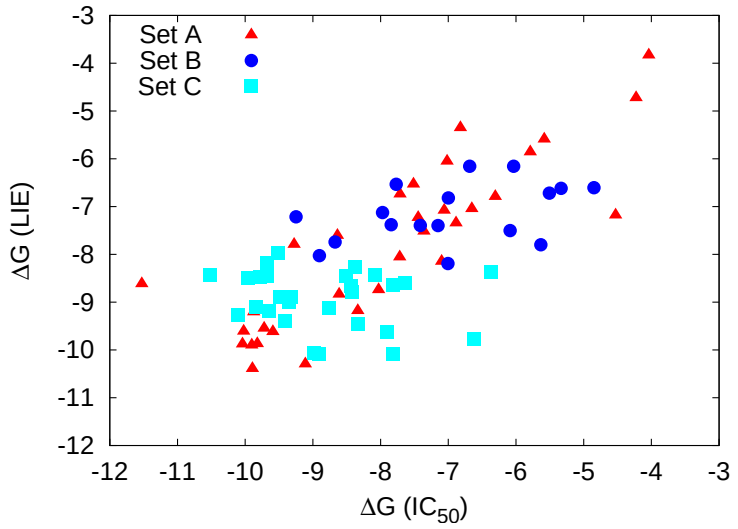
- trends similar to MM/PBSA, good r^2 for A
- electrostatics does not contribute to specificity

Resolution



	A	B	C
MM/PBSA	0.52	0.24	0.01
MD/PBSA	0.58	0.25	0.07
LIE ($\beta = 0$)	0.73	0.17	0.03

Resolution



	A	B	C
MM/PBSA	0.52	0.24	0.01
MD/PBSA	0.58	0.25	0.07
LIE ($\beta = 0$)	0.73	0.17	0.03
$\Delta\Delta G$	7.5	4.4	4.2
$\Delta \log_{10} IC_{50}$	5.5	3.2	3.0

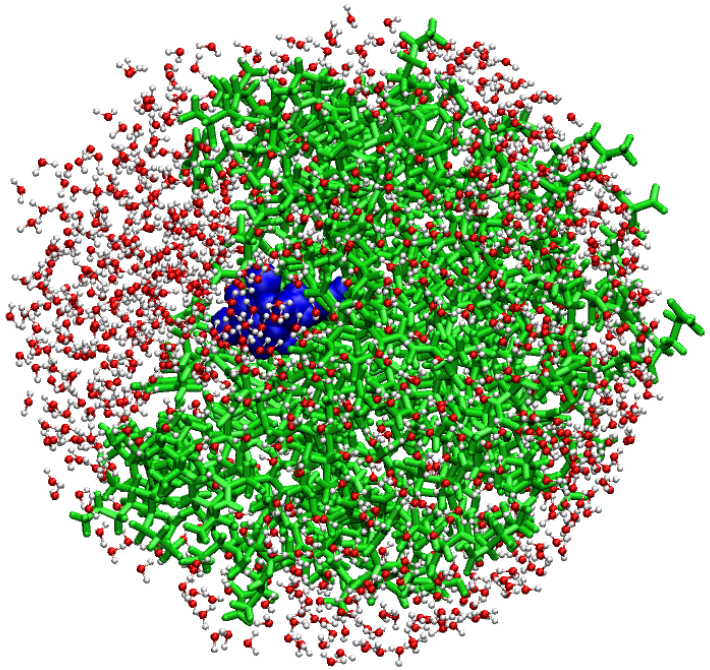
- Affinity differences smaller than 3 orders of magnitude can NOT be resolved in this case.
- Protein conformation may play a role.

Thermodynamic Integration

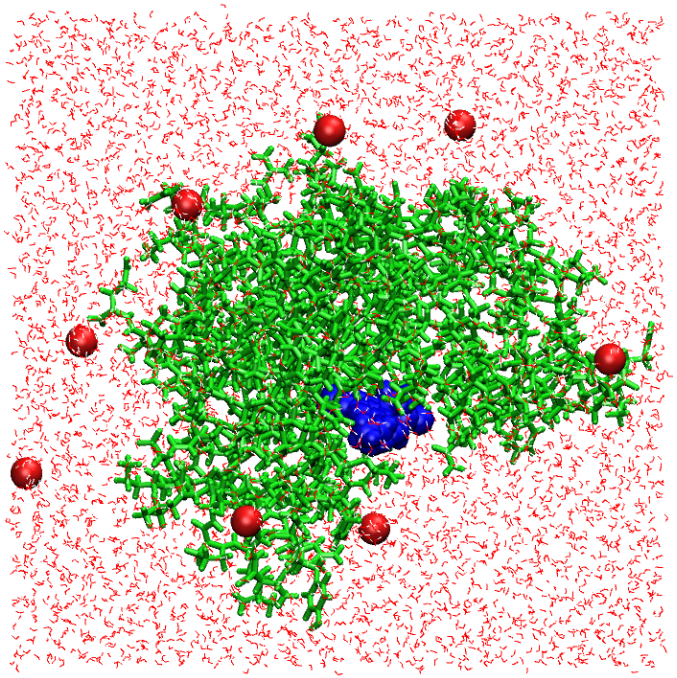
Thermodynamic Integration

- How much CPU-time does it require ?
- Protocol, long range interactions, system size, etc ?

spherical BC



full PBC + counter-ions



Thermodynamic Integration - Accuracy

2 compound pairs from
series C:
1 charged, 1 neutral

small change: S $\rightarrow$ O
IC_{50} $\searrow$ fact. 100

Thermodynamic Integration - Accuracy

2 compound pairs from series C:
1 charged, 1 neutral

small change: S $\longrightarrow$ O
IC_{50} $\searrow$ fact. 100

	$\Delta\Delta G$ [kcal/mol]	
	case 1 neutral	case 2 charged
xptl	2.3	2.3
SBC	2.3	4.5
PBC/ctof		4.1
PBC/PME	2.2 $\pm$ 0.03	2.9 $\pm$ 0.04

Thermodynamic Integration - Accuracy

2 compound pairs from series C:
1 charged, 1 neutral

small change: S $\longrightarrow$ O
$IC_{50} \searrow$ fact. 100

	$\Delta\Delta G$ [kcal/mol]	
	case 1 neutral	case 2 charged
xptl	2.3	2.3
SBC	2.3	4.5
PBC/ctof		4.1
PBC/PME	2.2 $\pm$ 0.03	2.9 $\pm$ 0.04

- error $<$ 1 kcal/mol in two cases
- full PBC + Ewald gives best result
- required CPU-time reasonable
- results converge fast with PBC/Ewald

Summary

- Hsp90 ligand binding specifity:

  dominated by vdW interactions.

Summary

- Hsp90 ligand binding specifity:

  dominated by vdW interactions.
- MM/PBSA: optimal protocol varies

  with the ligand series / protein conformation

Summary

- Hsp90 ligand binding specifity:
  dominated by vdW interactions.
- MM/PBSA: optimal protocol varies
  with the ligand series / protein conformation
- MM/PBSA or LIE reliably resolve only large differences
  in binding affinities in this case.

Summary

- Hsp90 ligand binding specifity:
  dominated by vdW interactions.
- MM/PBSA: optimal protocol varies
  with the ligand series / protein conformation
- MM/PBSA or LIE reliably resolve only large differences
  in binding affinities in this case.
- extensive sampling barely improves results.

Summary

- Hsp90 ligand binding specifity:
  dominated by vdW interactions.
- MM/PBSA: optimal protocol varies
  with the ligand series / protein conformation
- MM/PBSA or LIE reliably resolve only large differences
  in binding affinities in this case.
- extensive sampling barely improves results.
- TI/FEP: accuracy $\approx$ 1 kcal/mol (preliminary)

Summary

- Hsp90 ligand binding specifity:
  dominated by vdW interactions.
- MM/PBSA: optimal protocol varies
  with the ligand series / protein conformation
- MM/PBSA or LIE reliably resolve only large differences
  in binding affinities in this case.
- extensive sampling barely improves results.
- TI/FEP: accuracy $\approx$ 1 kcal/mol (preliminary)
- TI/FEP: required CPU time reasonable
  usage in lead optimization feasible.

Acknowledgements

Kate Grant (assays)
Lisa Fisher, David Robinson (X-ray)

Medicinal chemistry:
Paul Brough, Brian Dymock, Martin Drysdale

The modeling team:
Xavier Barril, Ijen Chen, Alba Macias,
Christine Richardson

and Rod Hubbard