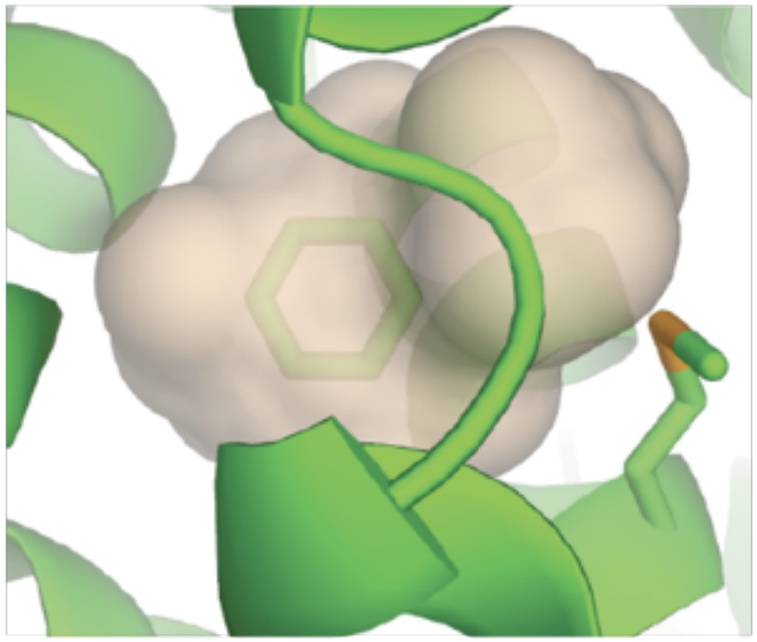
Examining how protein sampling impacts the convergence of relative binding free energy calculations

David Mobley

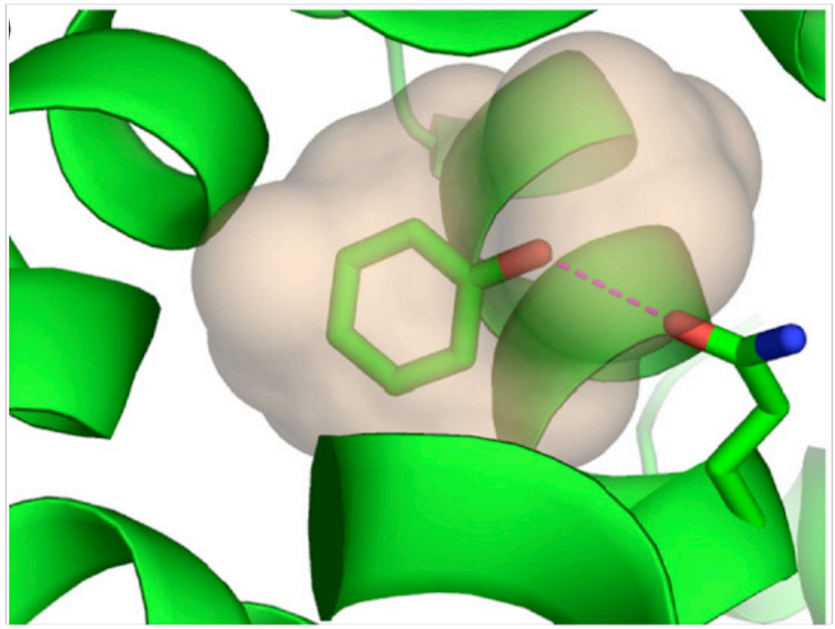
Nathan M. Lim, Lingle Wang, Robert Abel

We've been using a progression of model binding sites for free energy method development



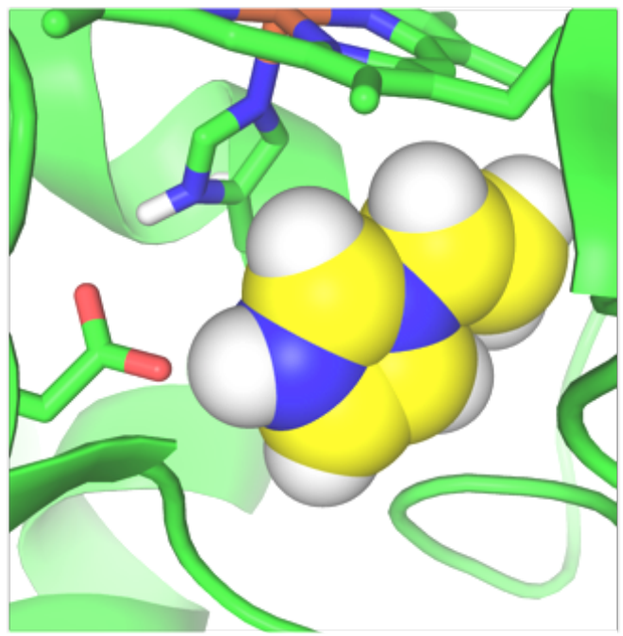
Lysozyme L99A

- Simple
- Nonpolar
- Dry



Lysozyme L99A/M102Q

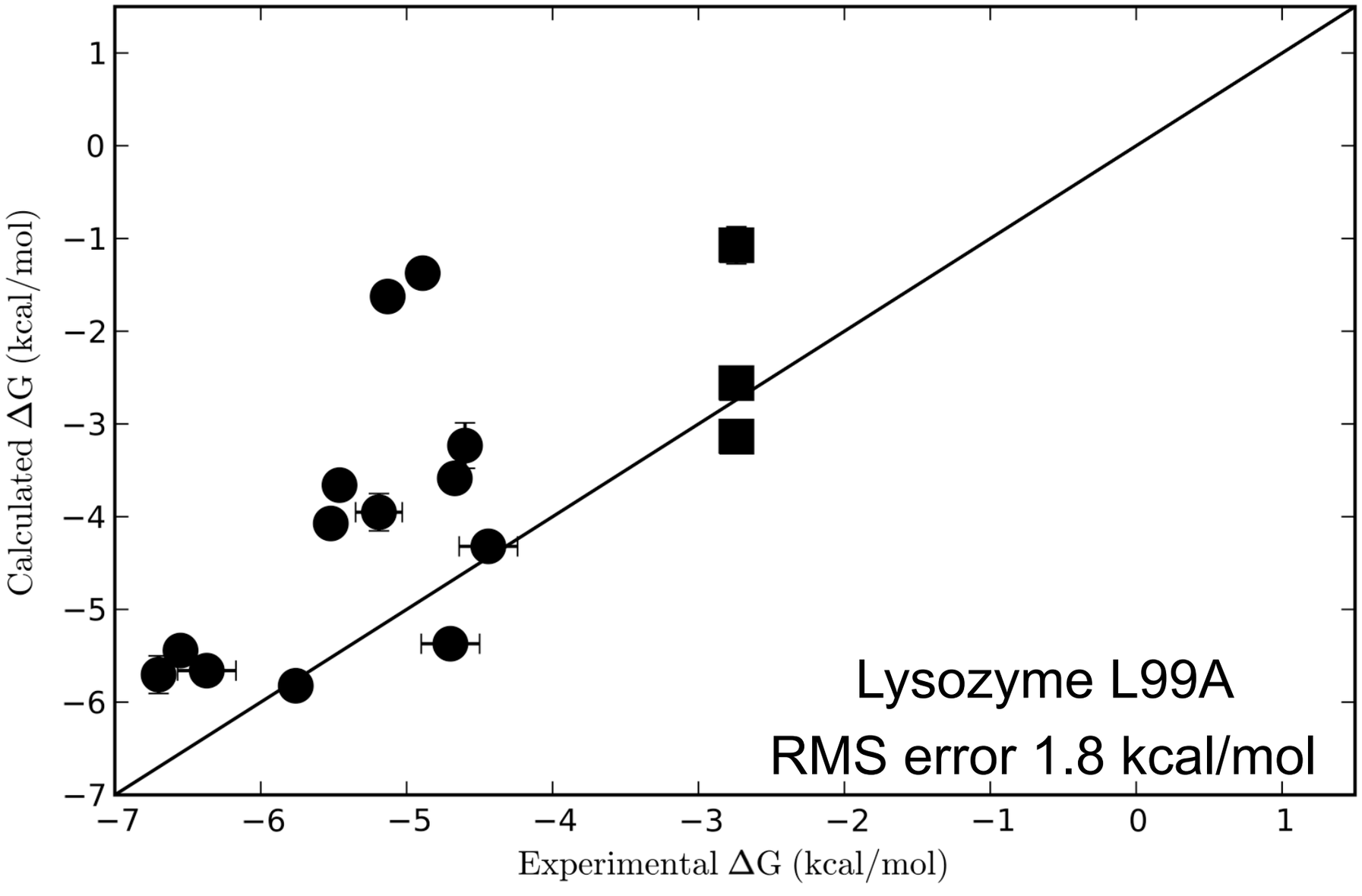
- Simple
- Polar
- Dry
- Additional stable binding modes



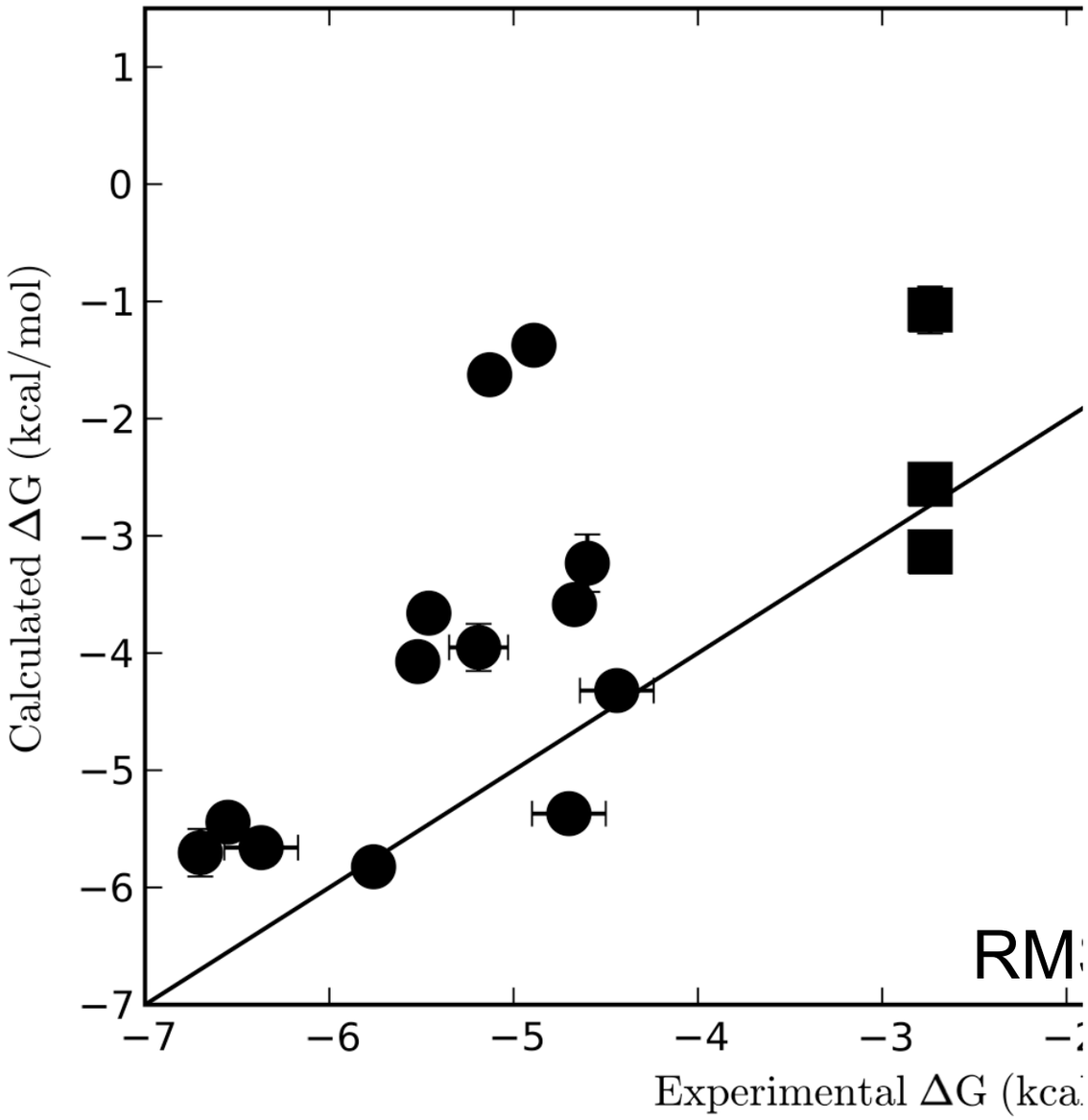
Cytochr. C Peroxidase

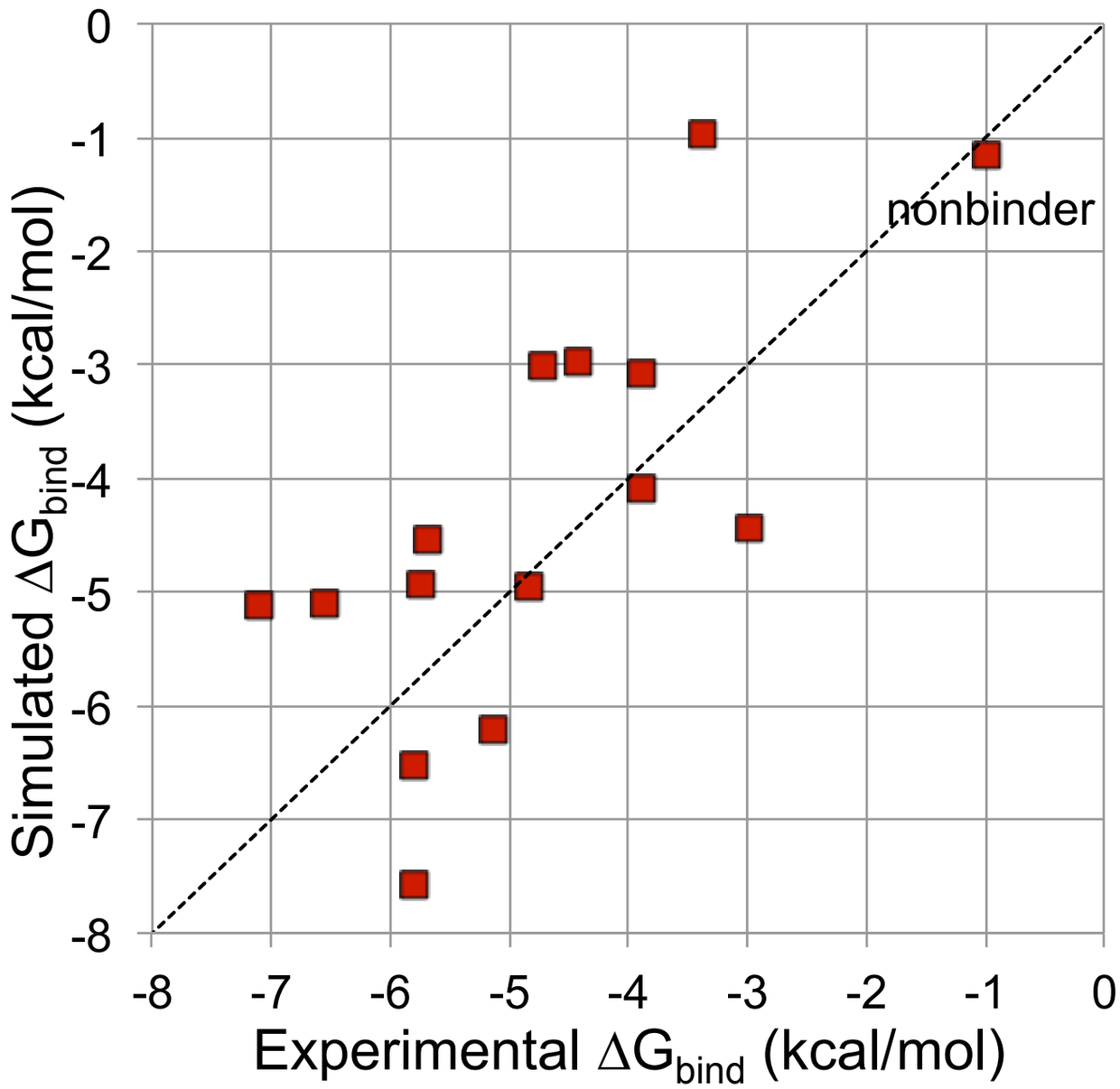
- Simple (?)
- Polar, Charged
- Wet
- Additional stable binding modes
- Force field issues?

Absolute free energy calculations on these sites are tractable and have taught us a great deal



Absolute free energy calculations on these sites are tractable and have taught us a great deal

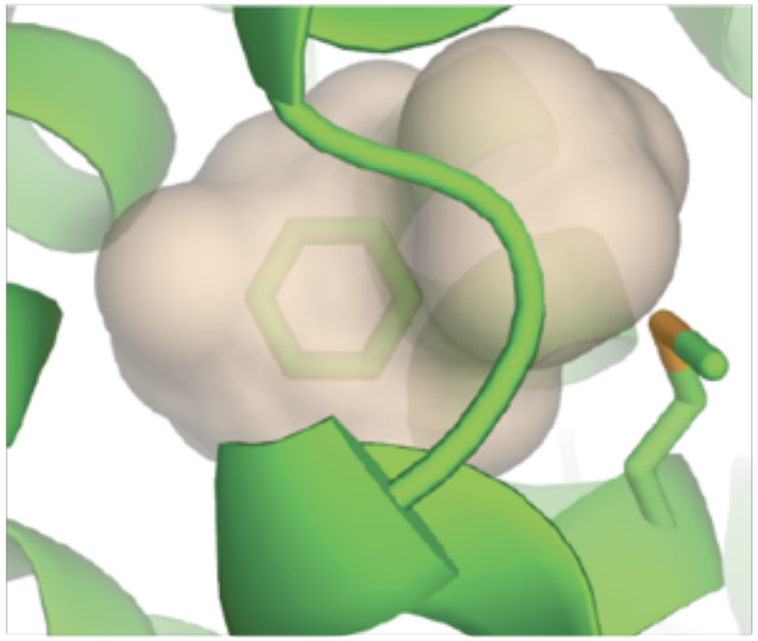


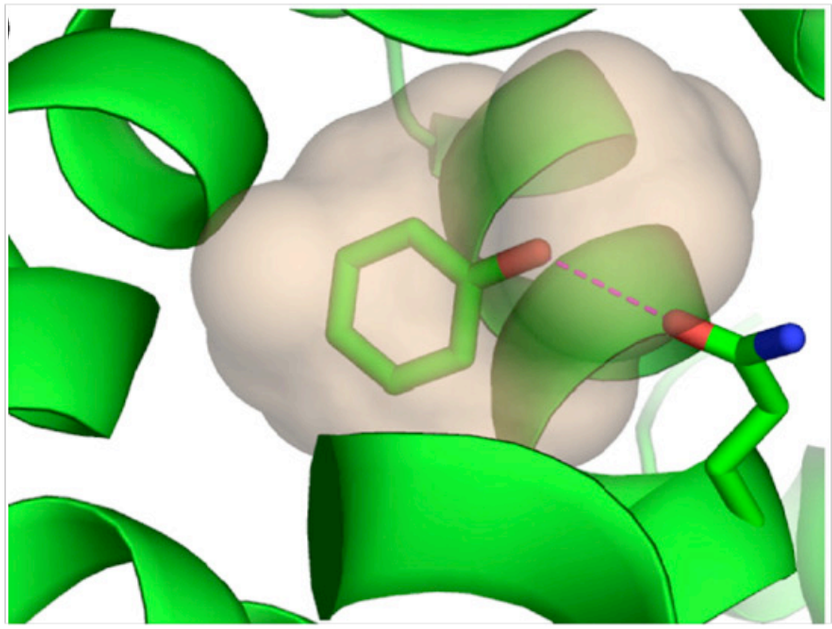


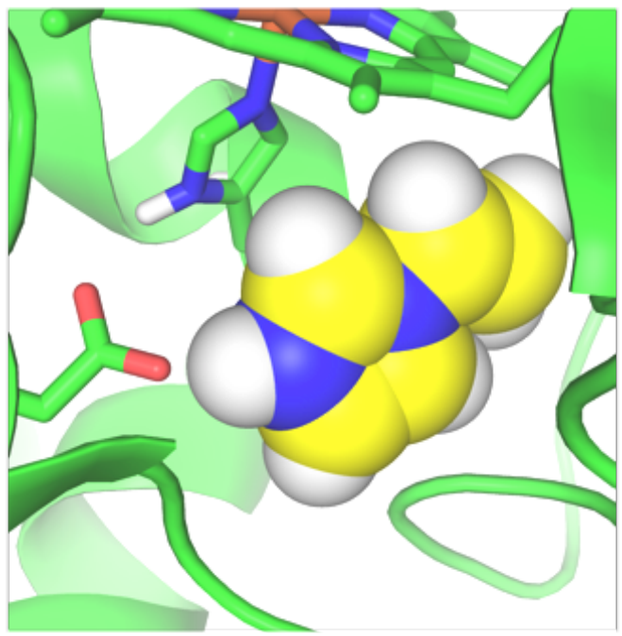
Cytochr. C Peroxidase

RMS error 1.4 kcal/mol

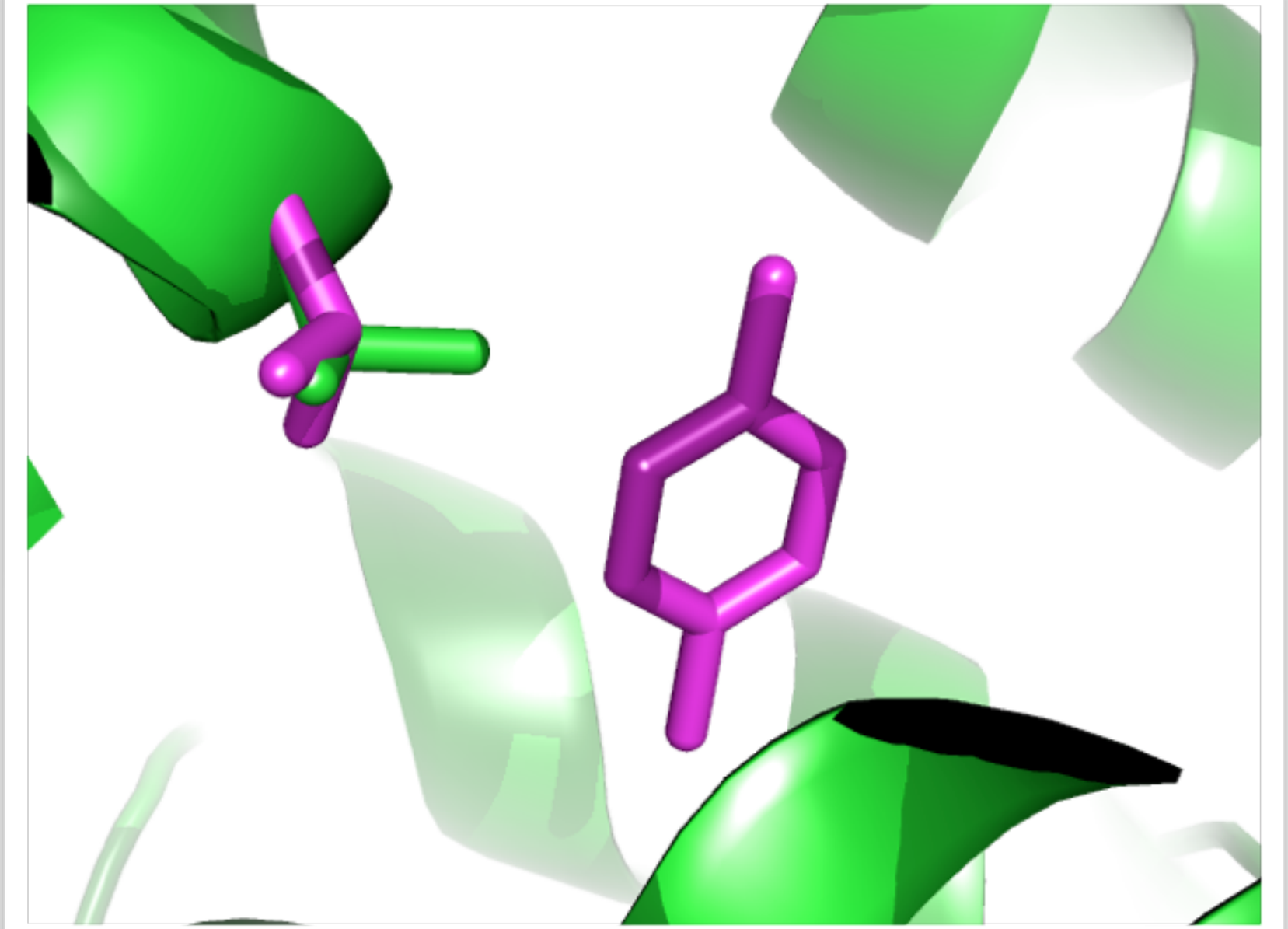
These model sites are simple, rigid, boring, easy - time to move to biological sites, right?







Well, the simple L99A site does have a slow conformational change

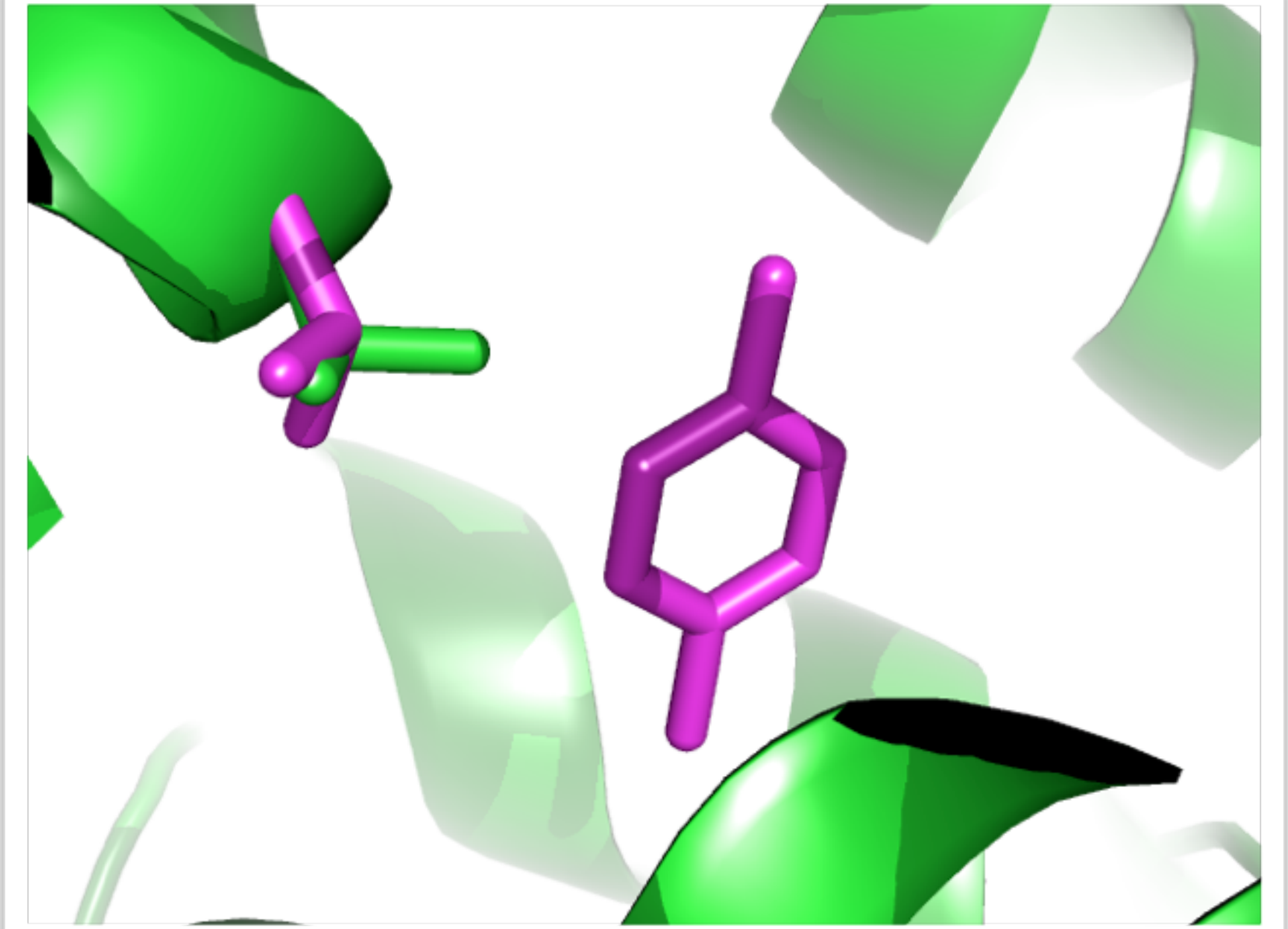


From apo structure: ΔG=-3.0+/-0.1 kcal/mol

From holo structure: ΔG=-7.3+/-0.1 kcal/mol

Experiment: -4.6 kcal/mol

It’s necessary to include conformational change

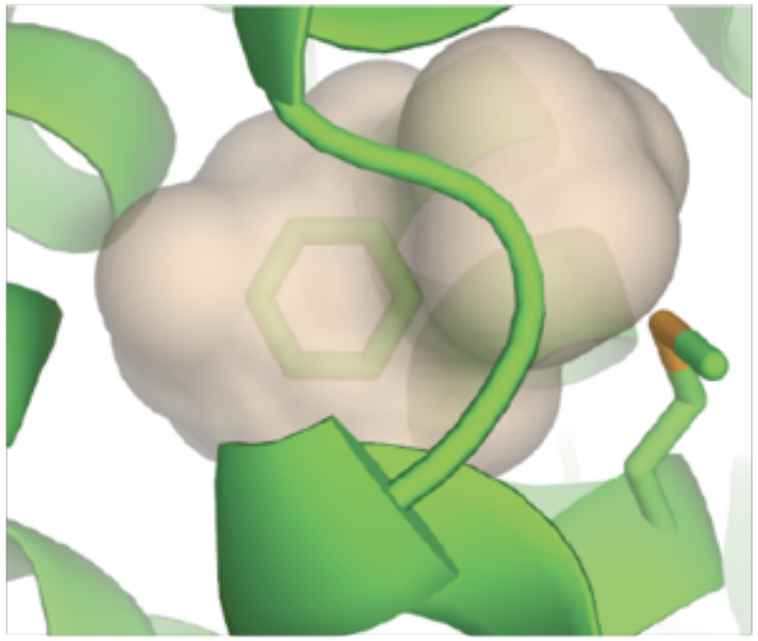


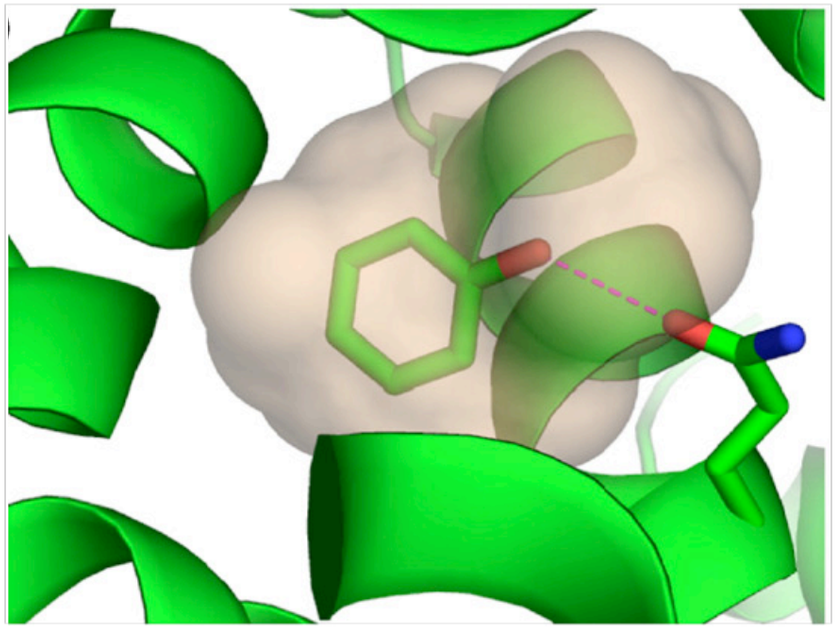
From apo structure: ΔG=-3.5+/-0.2 kcal/mol

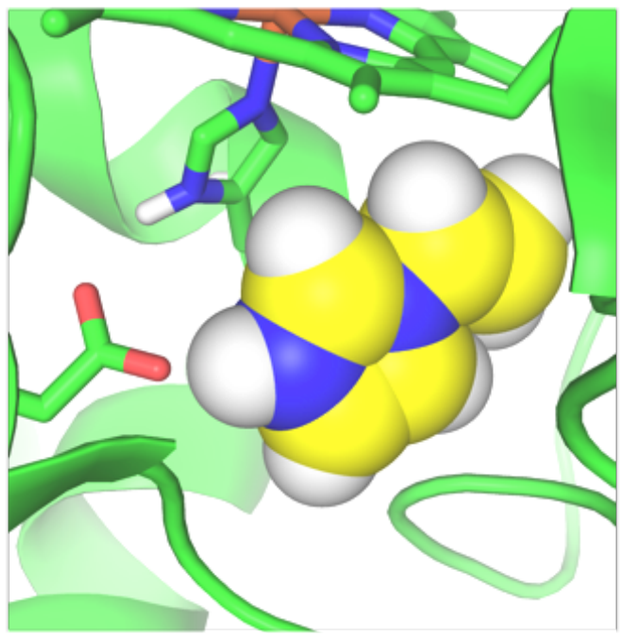
From holo structure: ΔG=-3.4+/-0.2 kcal/mol

Experiment: -4.6 kcal/mol

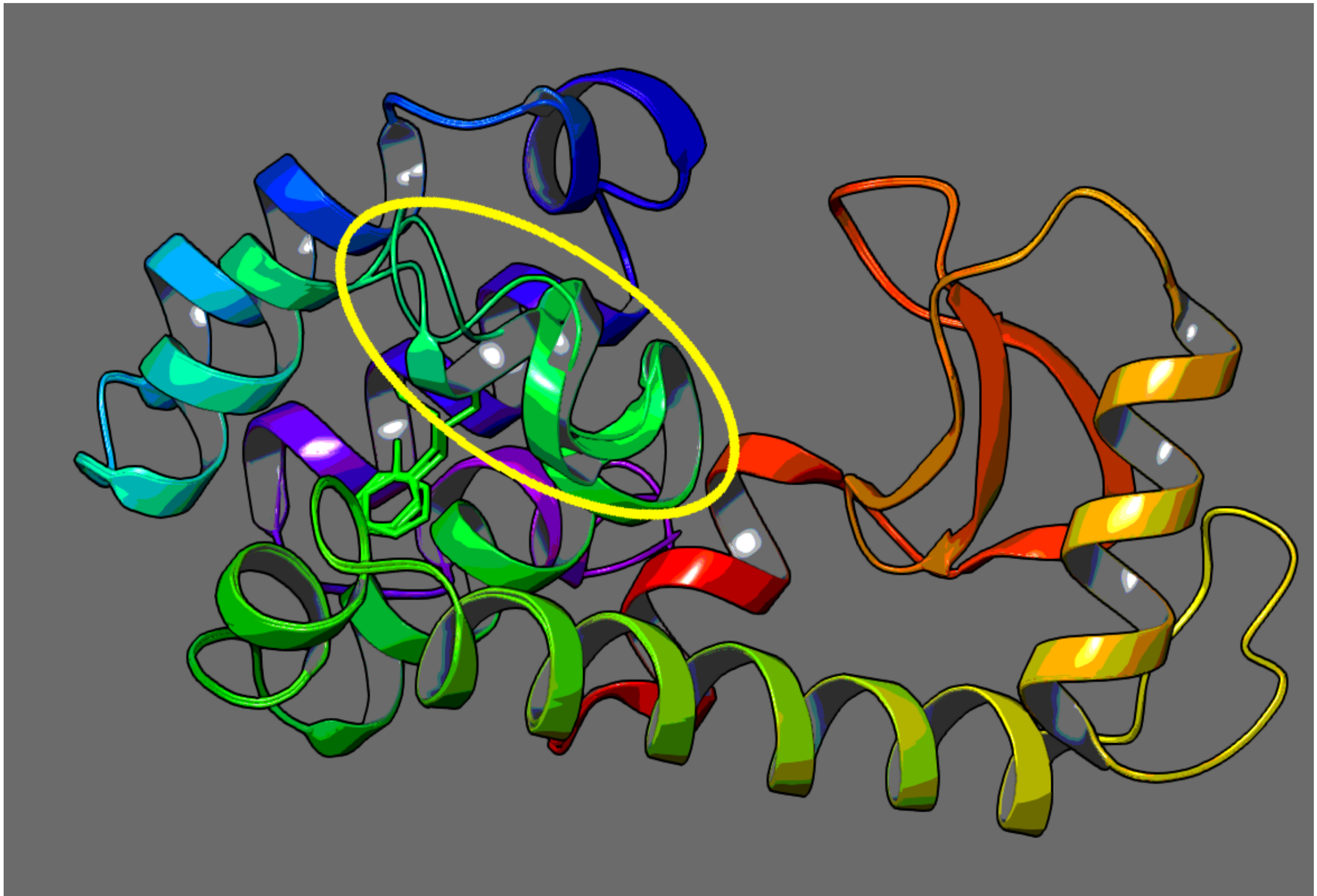
But that’s too small a conformational change. Real binding sites undergo larger conformational changes. Time to move to biological sites, right?



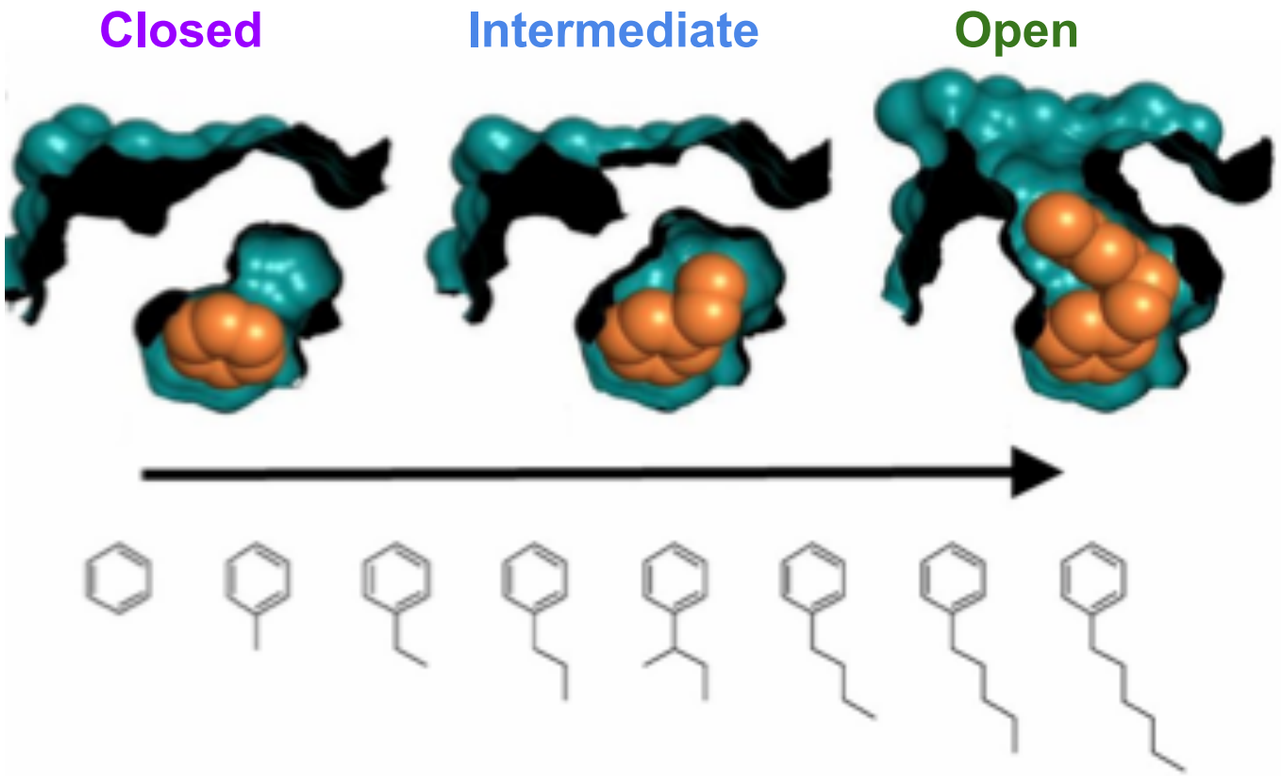


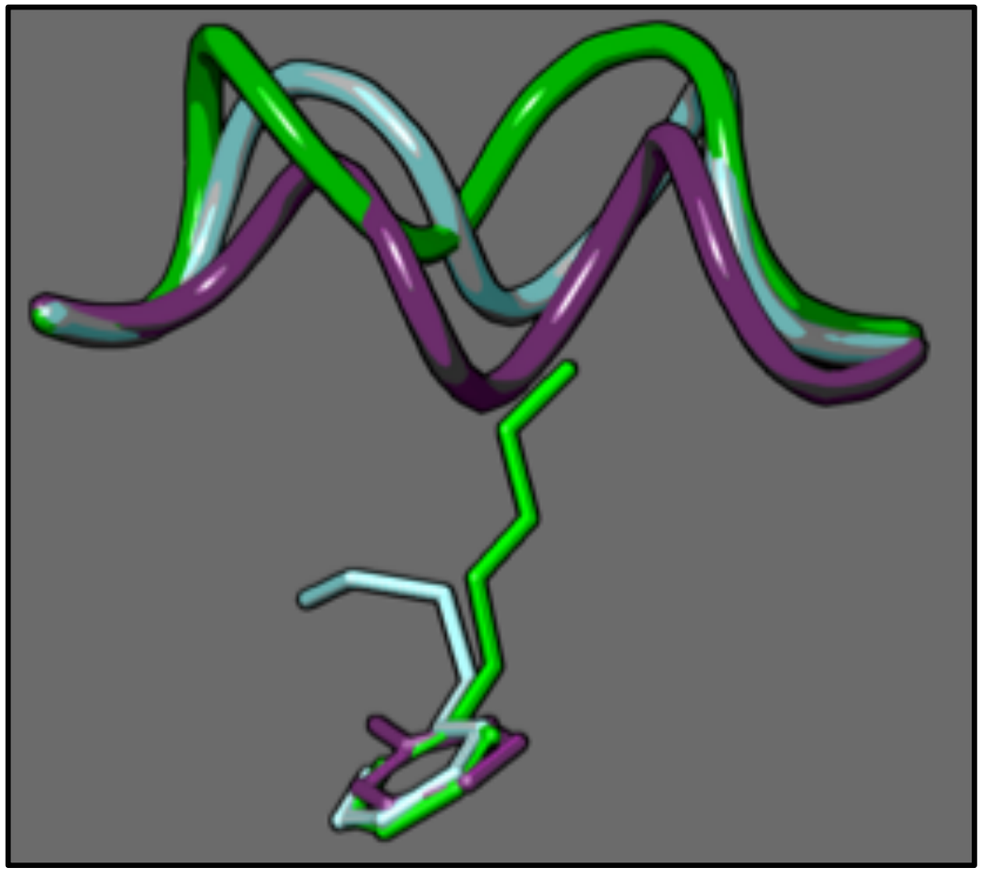


Well, L99A can undergo larger conformational changes too

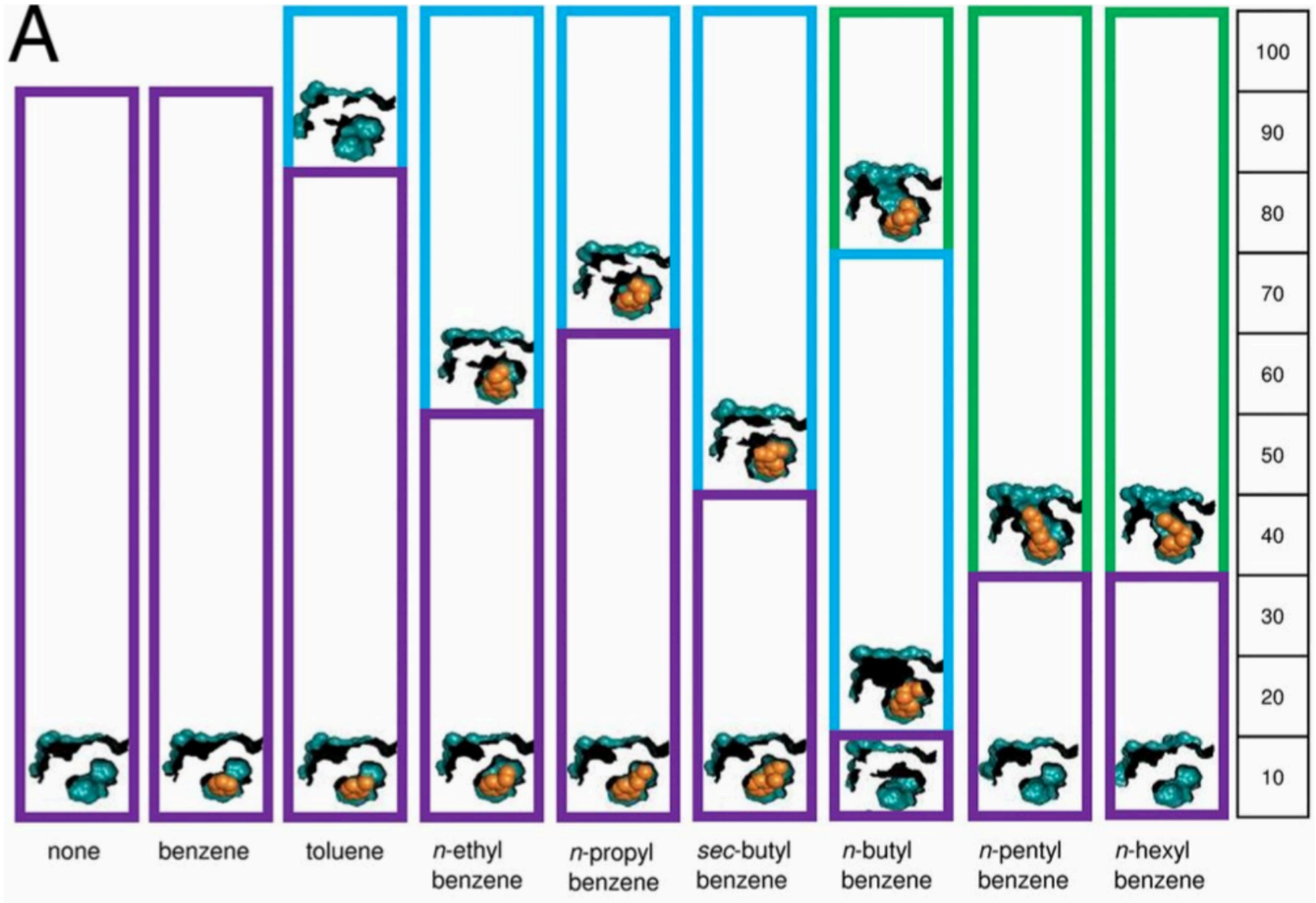


A series of related ligands gradually induce protein conformational changes

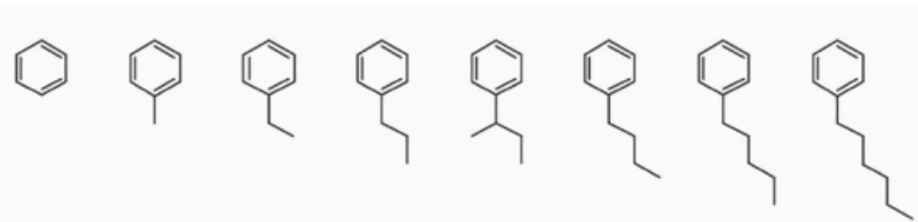




Crystallography resolves partial occupancies of different protein structures

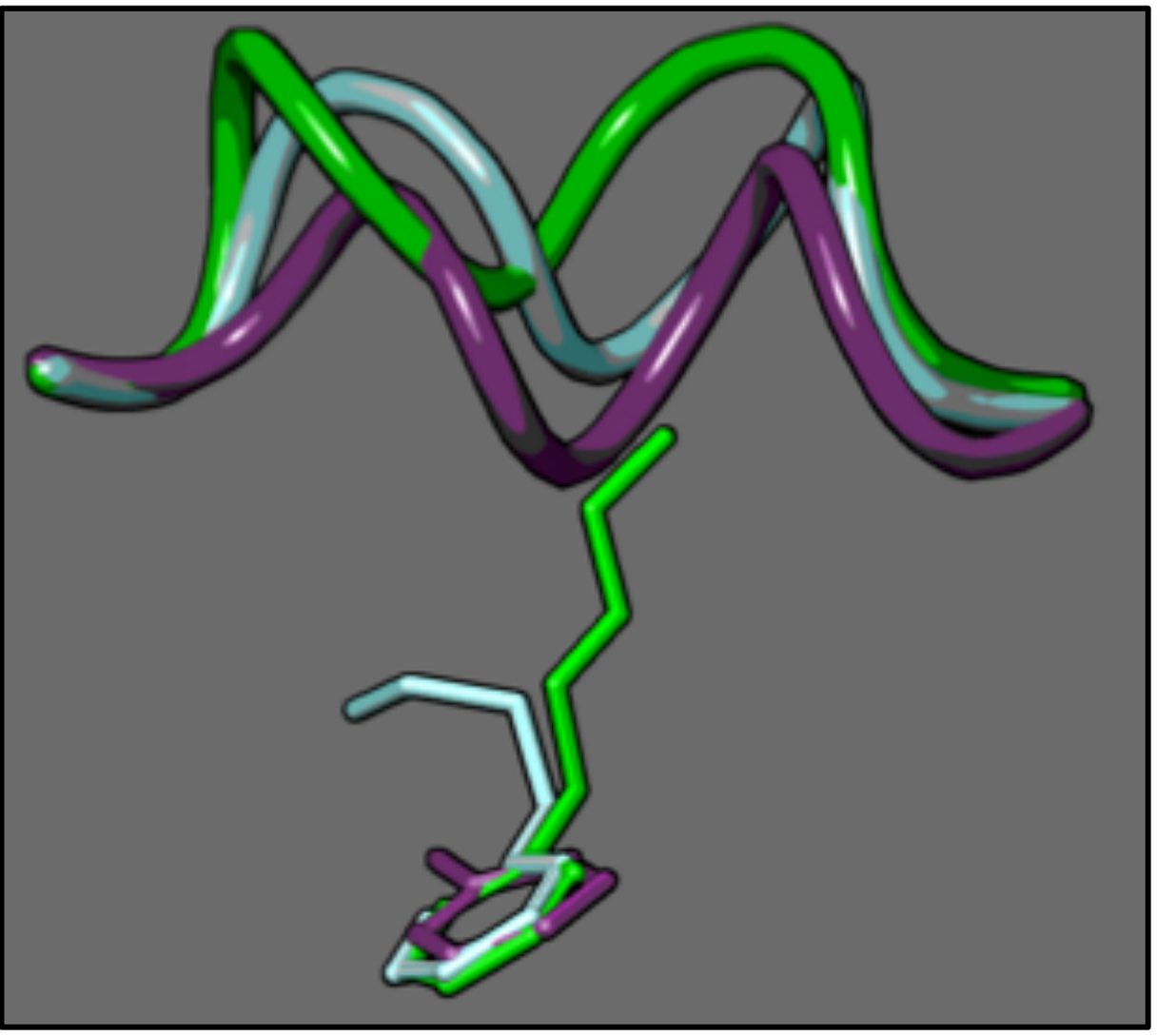


	Loop Occupancy		
Ligand	**C**	**I**	**O**
Benzene	**0.9**	-	-
Toluene	**0.8**	0.2	-
Ethyl	**0.5**	**0.5**	-
n-propyl	**0.6**	0.4	-
sec-butyl	0.4	**0.6**	-
n-butyl	0.1	**0.6**	0.3
n-pentyl	0.3		**0.7**
n-hexyl	0.3		**0.7**



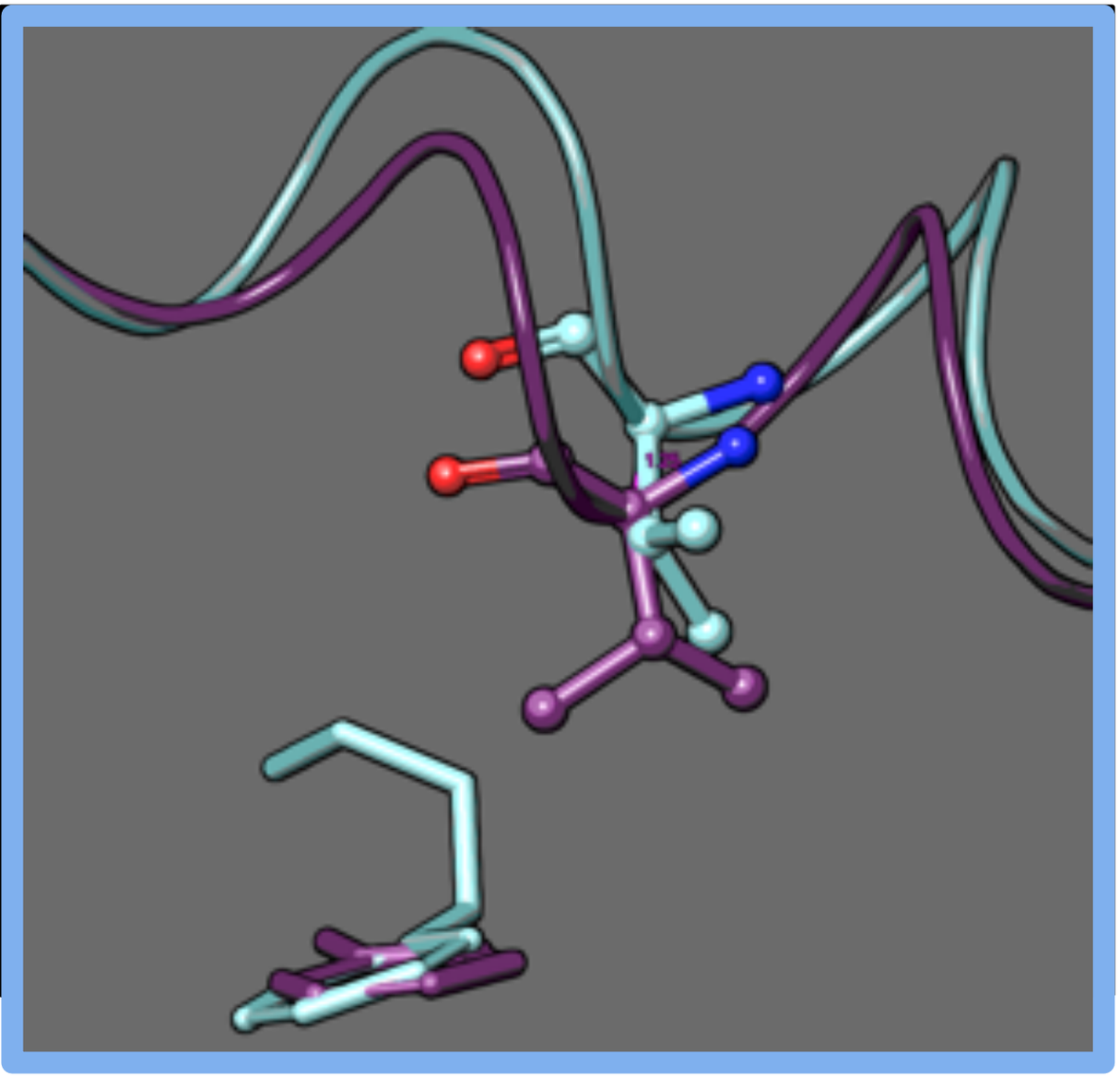
Merski, M., Fischer, M., Balius, T. E., Eidam, O., & Shoichet, B. K. (2015). Homologous ligands accommodated by discrete conformations of a buried cavity. Proceedings of the National Academy of Sciences, 112(16), 5039-5044.

Larger ligands seem to be better, but induce more conformational change



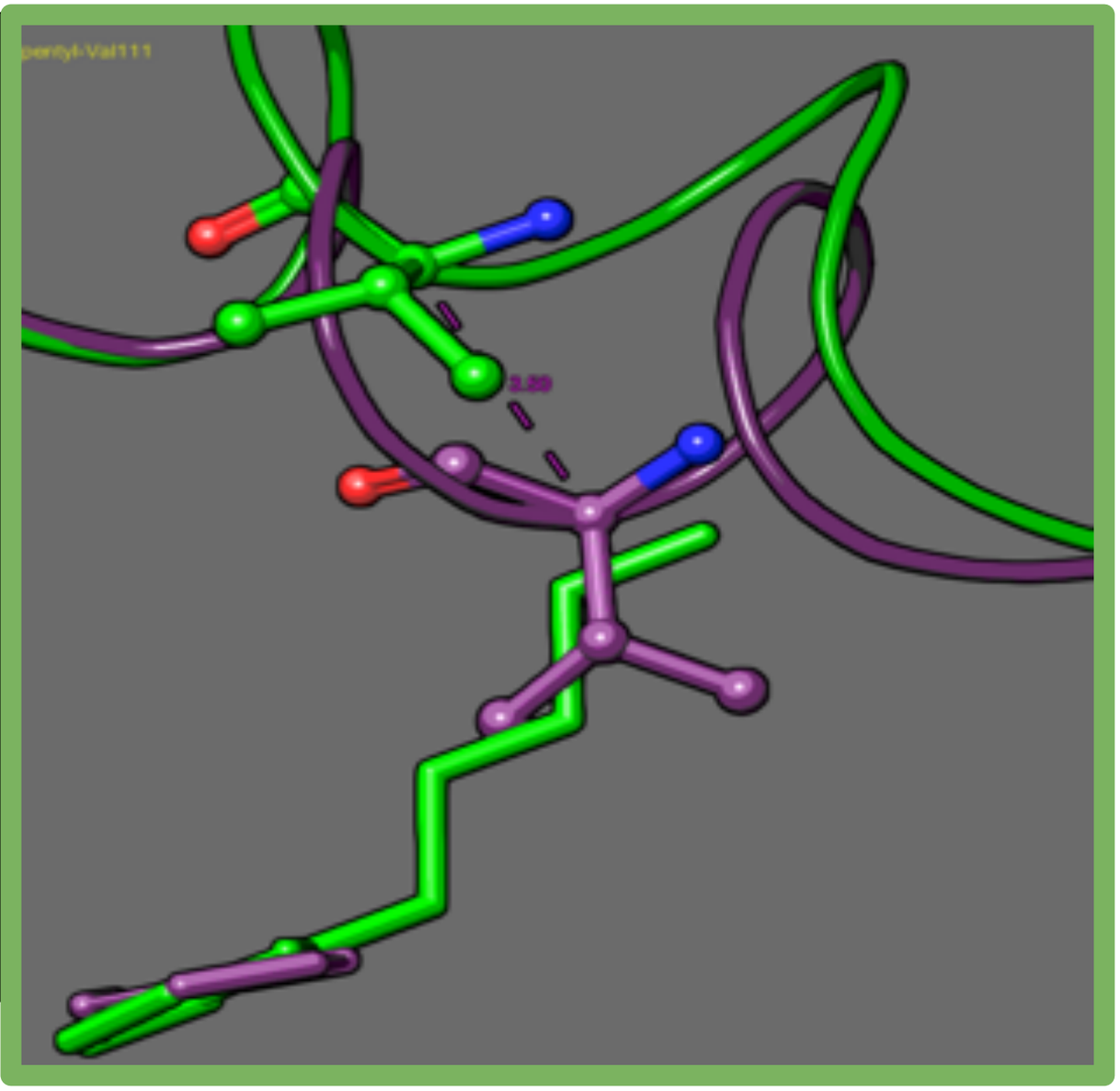
Ligand	ΔGexp (kcal/mol)
Benzene	**-5.19**
Toluene	**-5.52**
Ethylbenzene	**-5.76**
n-propylbenzene	**-6.55**
sec-butylbenzene	N/A
n-butylbenzene	**-6.70**
n-pentylbenzene	N/A
n-hexylbenzene	N/A

Larger ligands seem to be better, but induce more conformational change

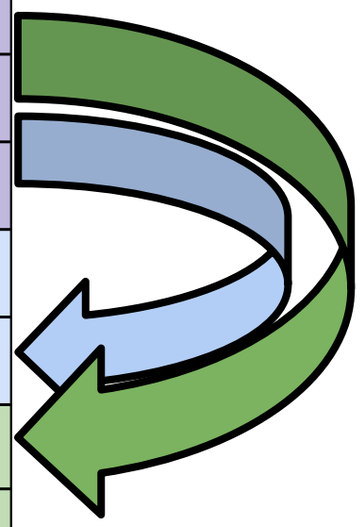


Ligand	ΔGexp (kcal/mol)
Benzene	**-5.19**
Toluene	**-5.52**
Ethylbenzene	**-5.76**
n-propylbenzene	**-6.55**
sec-butylbenzene	N/A
n-butylbenzene	**-6.70**
n-pentylbenzene	N/A
n-hexylbenzene	N/A

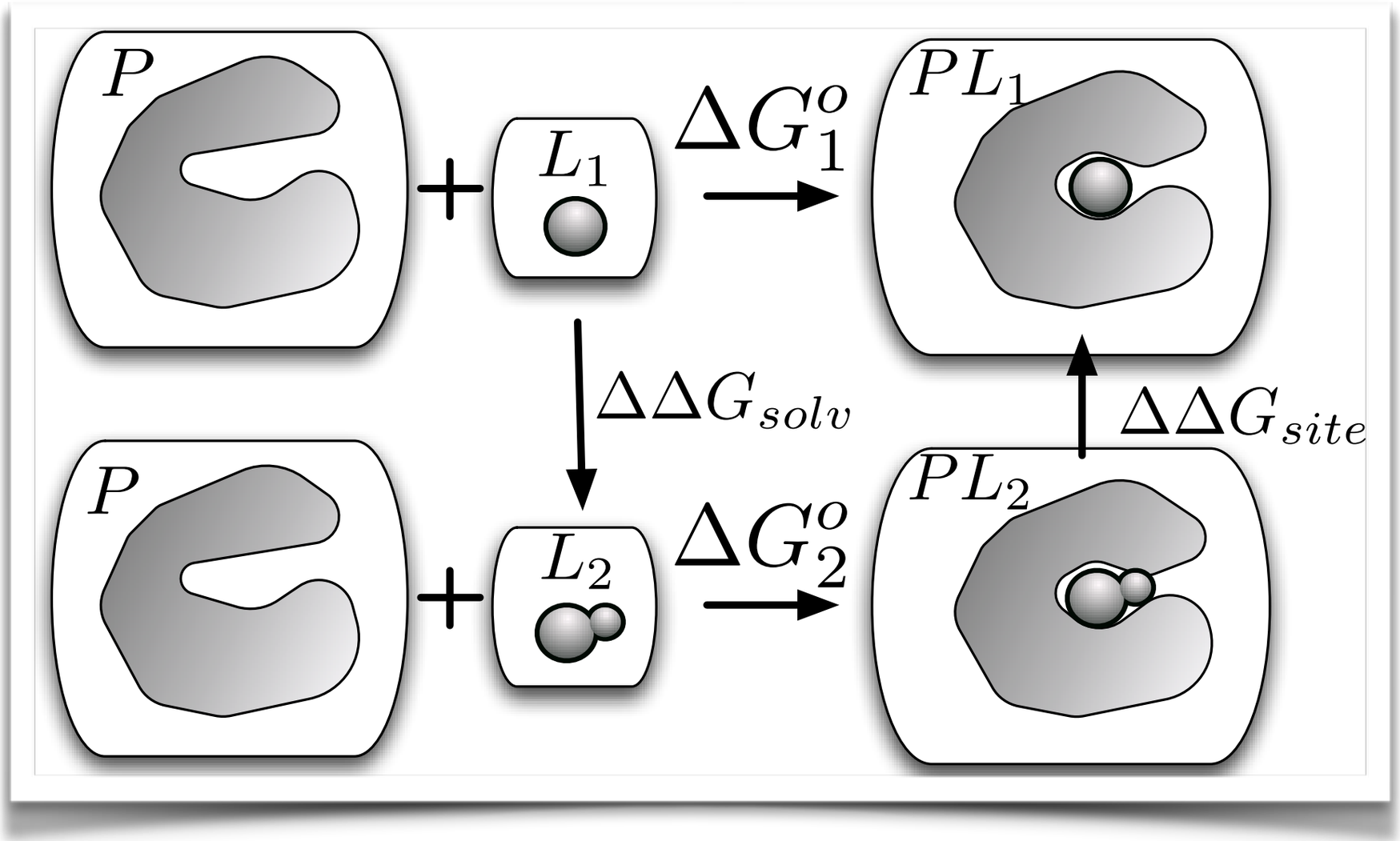
Larger ligands seem to be better, but induce more conformational change



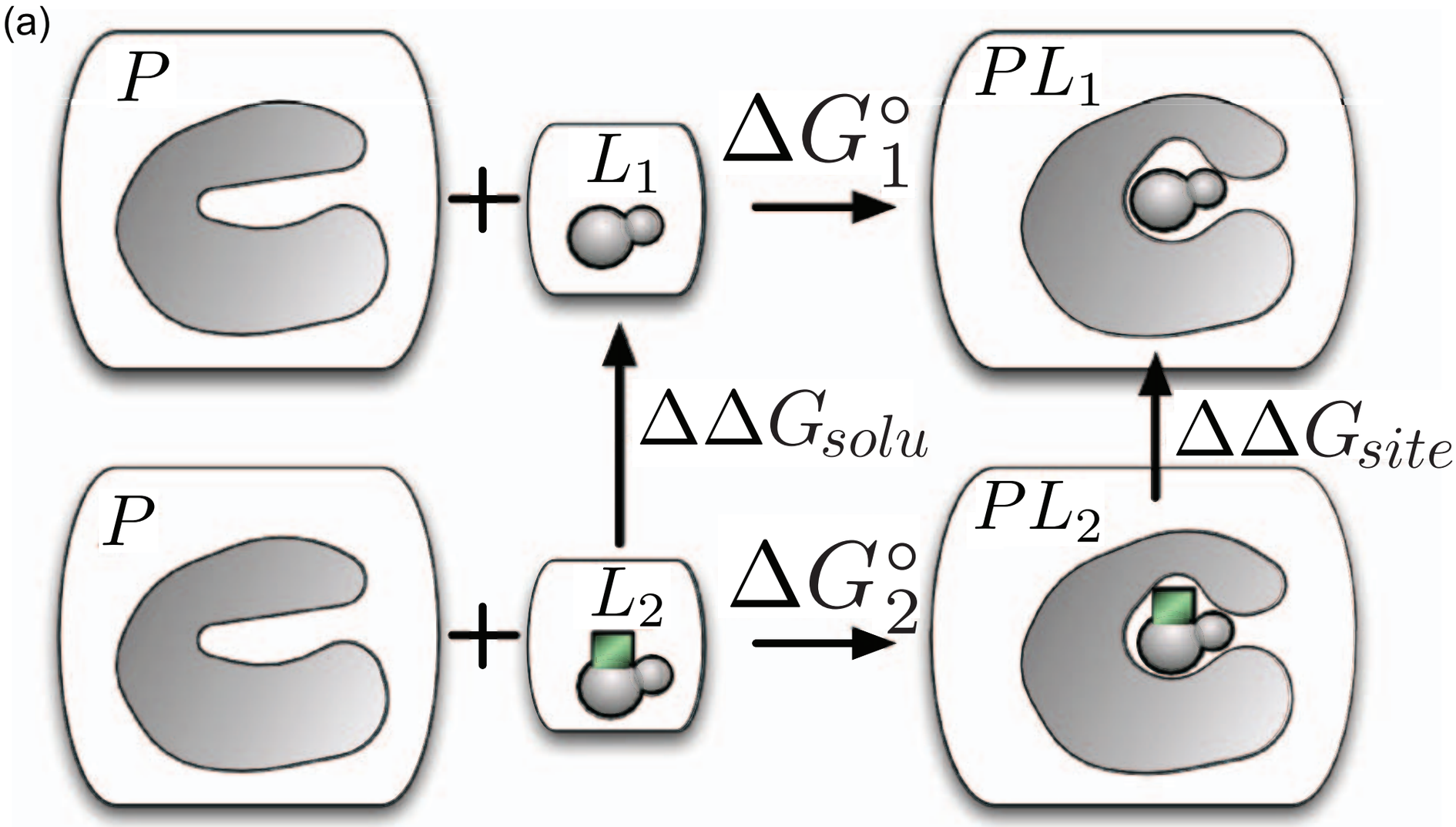
Ligand	ΔGexp (kcal/mol)
Benzene	**-5.19**
Toluene	**-5.52**
Ethylbenzene	**-5.76**
n-propylbenzene	**-6.55**
sec-butylbenzene	N/A
n-butylbenzene	**-6.70**
n-pentylbenzene	N/A
n-hexylbenzene	N/A



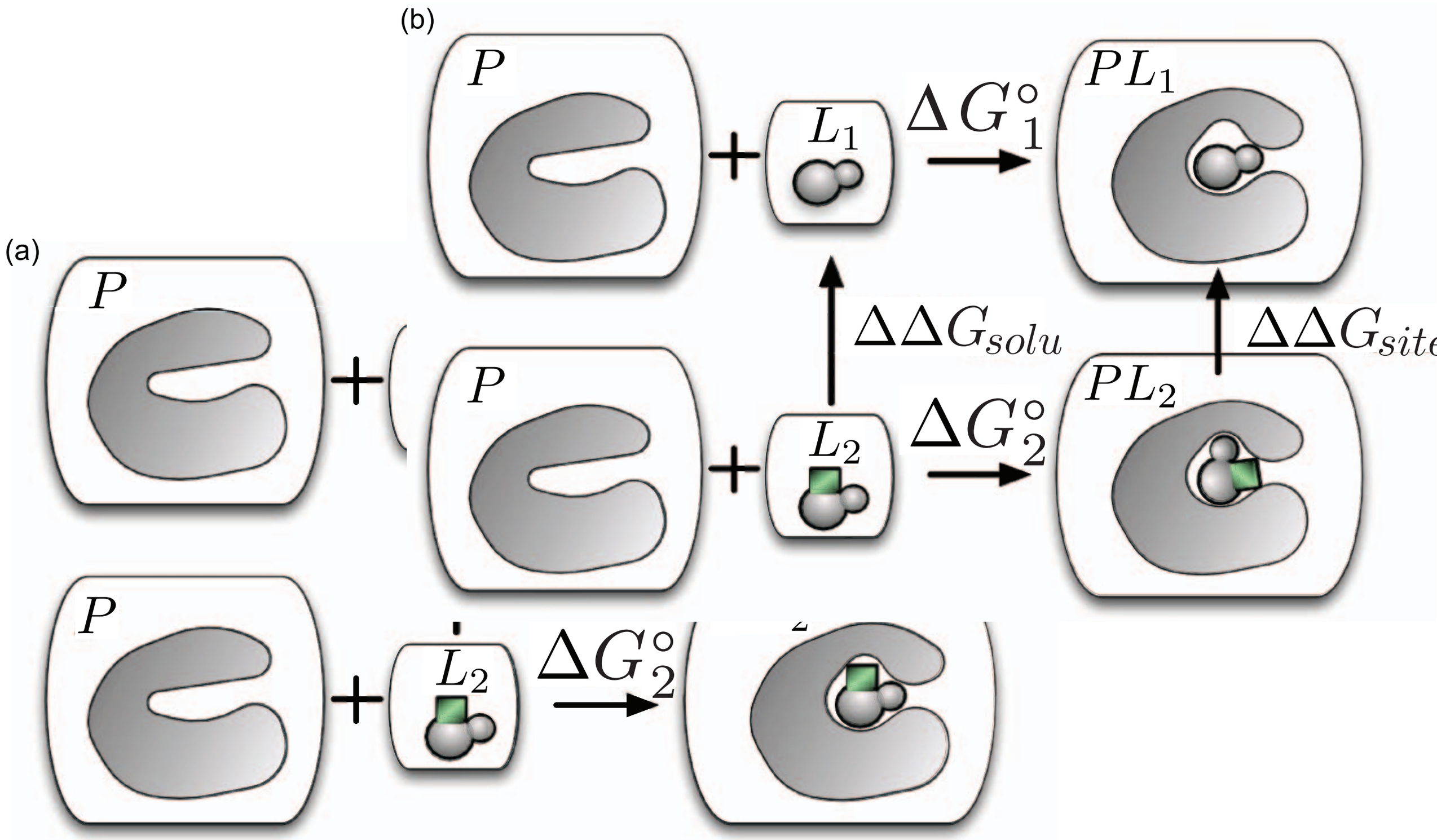
We use relative free energy calculations to explore this series with protein conformational change



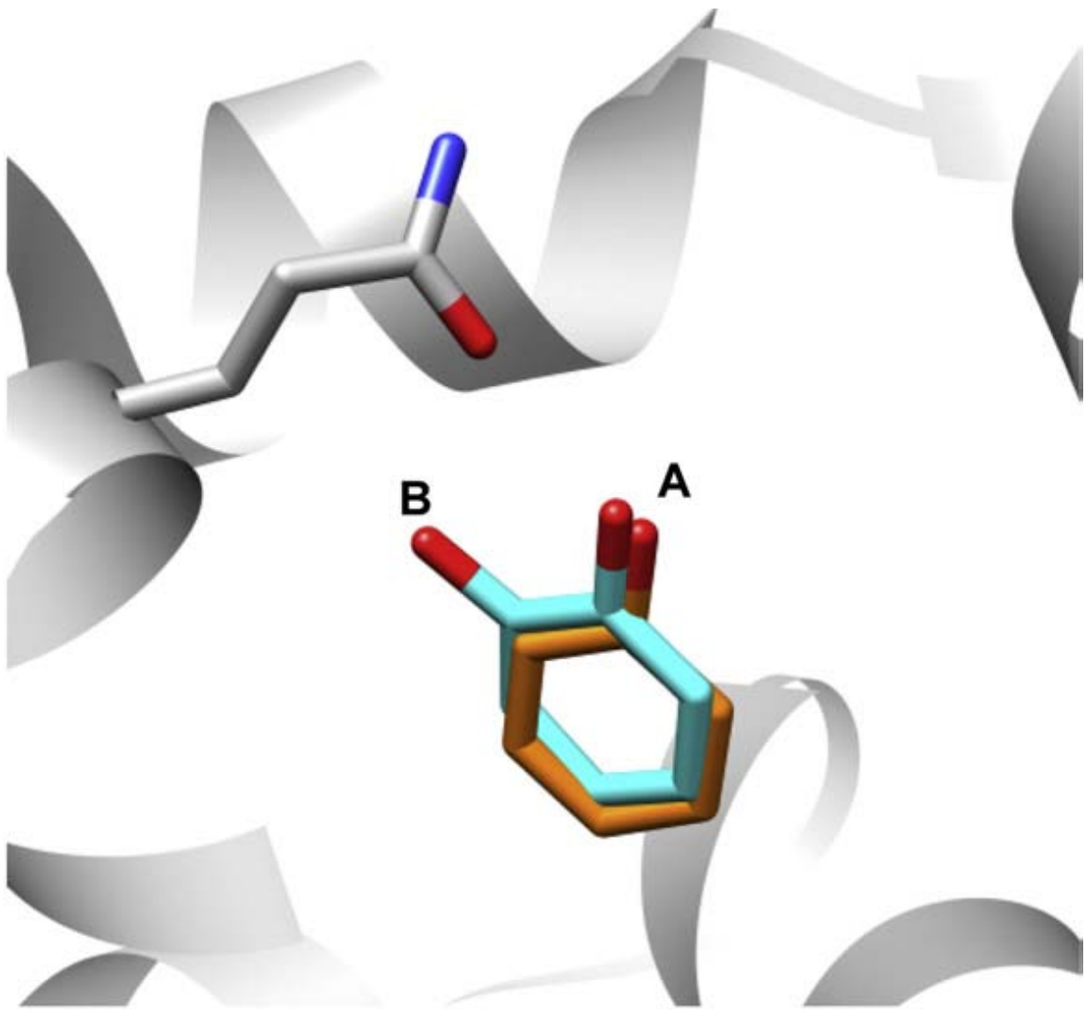
Potential sampling problems are “binding mode” sampling problems, if “binding mode” definition includes protein conformation



Potential sampling problems are “binding mode” sampling problems, if “binding mode” definition includes protein conformation



Even the lysozyme polar site has issues with multiple binding modes being poorly sampled



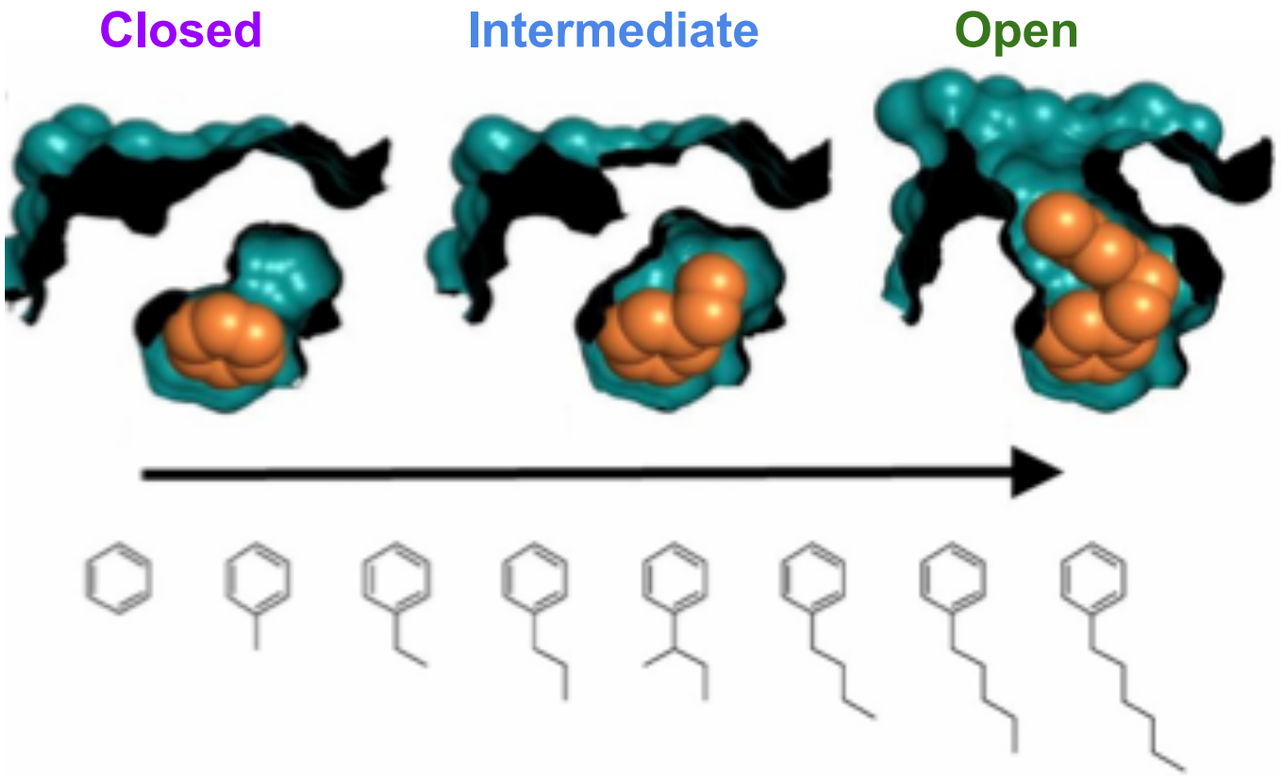
Orange: Phenol
Cyan: Catechol

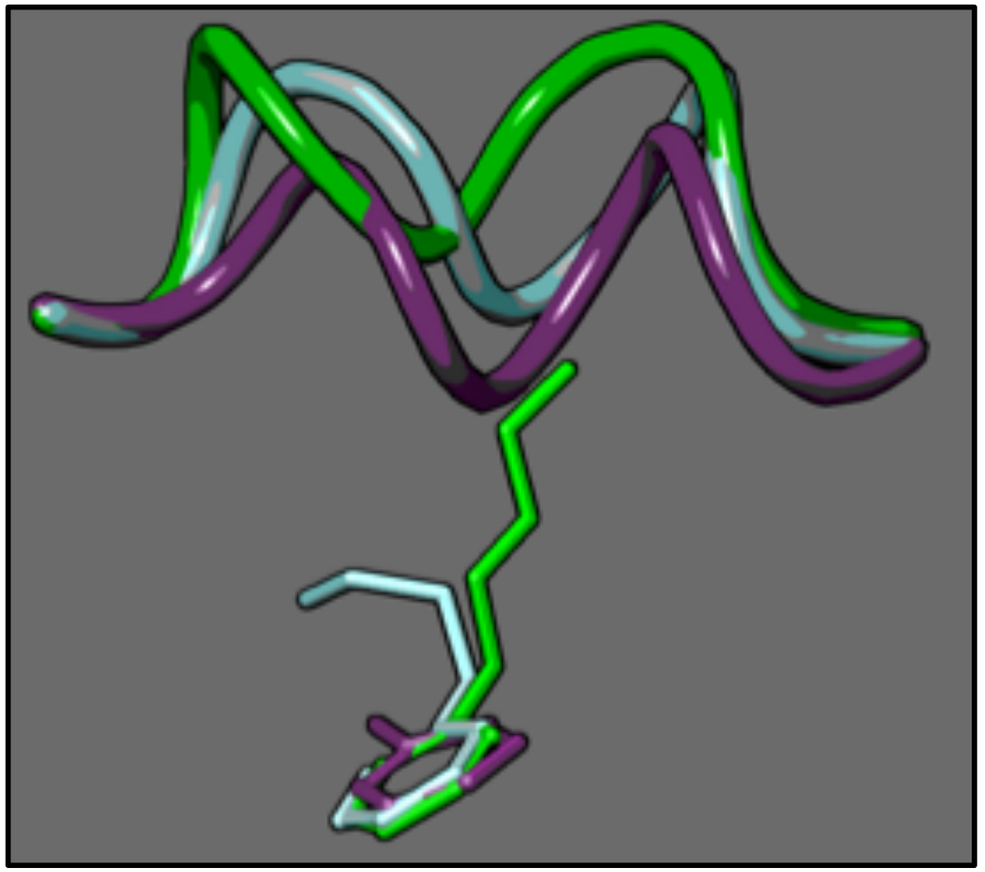
Relative calculations for derivatives like 2-methoxyphenol have four likely binding modes

Interconversion is slow

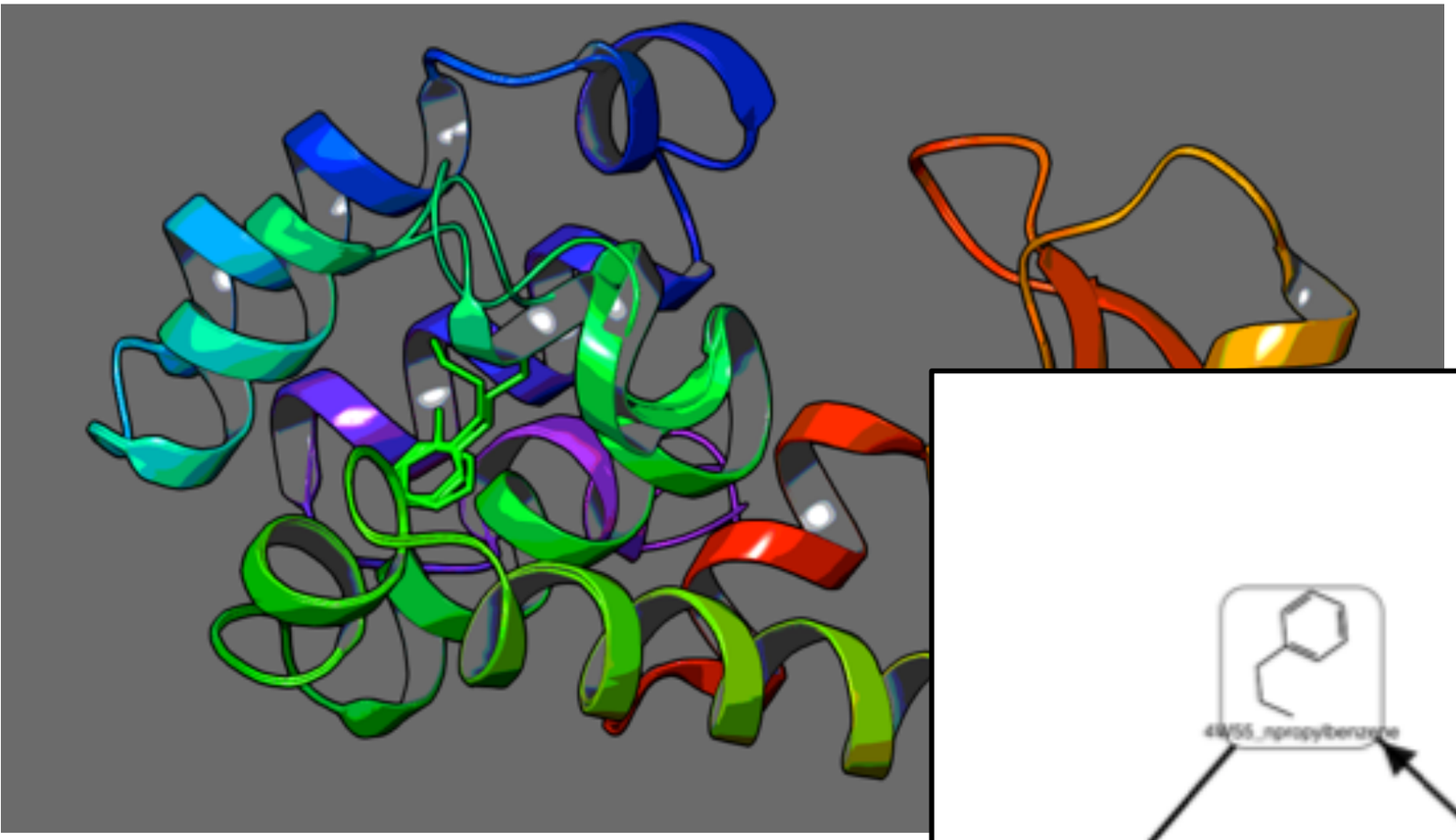
Calculated relative binding free energies differ by up to 4 kcal/mol depending on starting structure

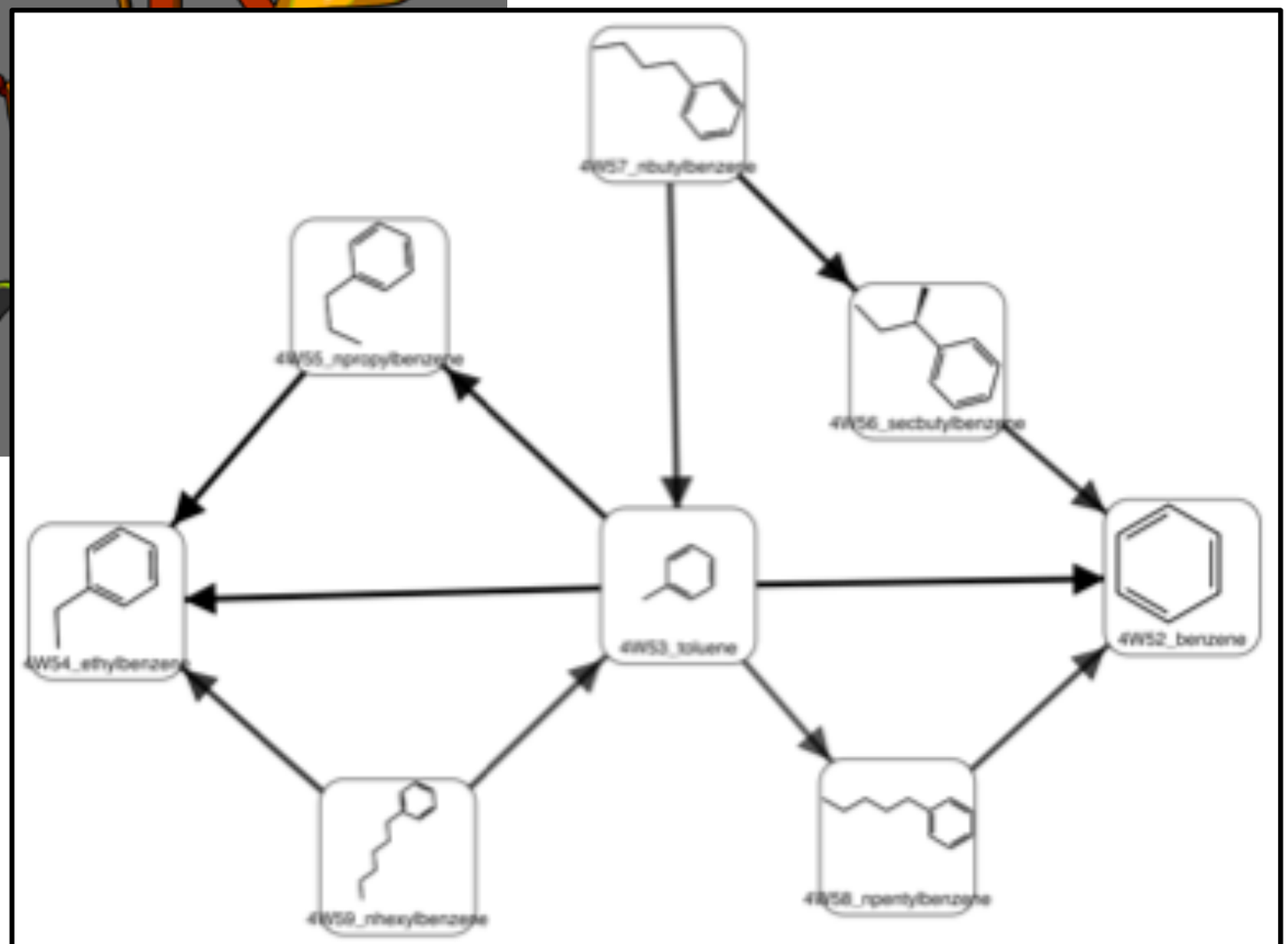
Let’s get back to protein conformational changes and see how they affect relative calculations



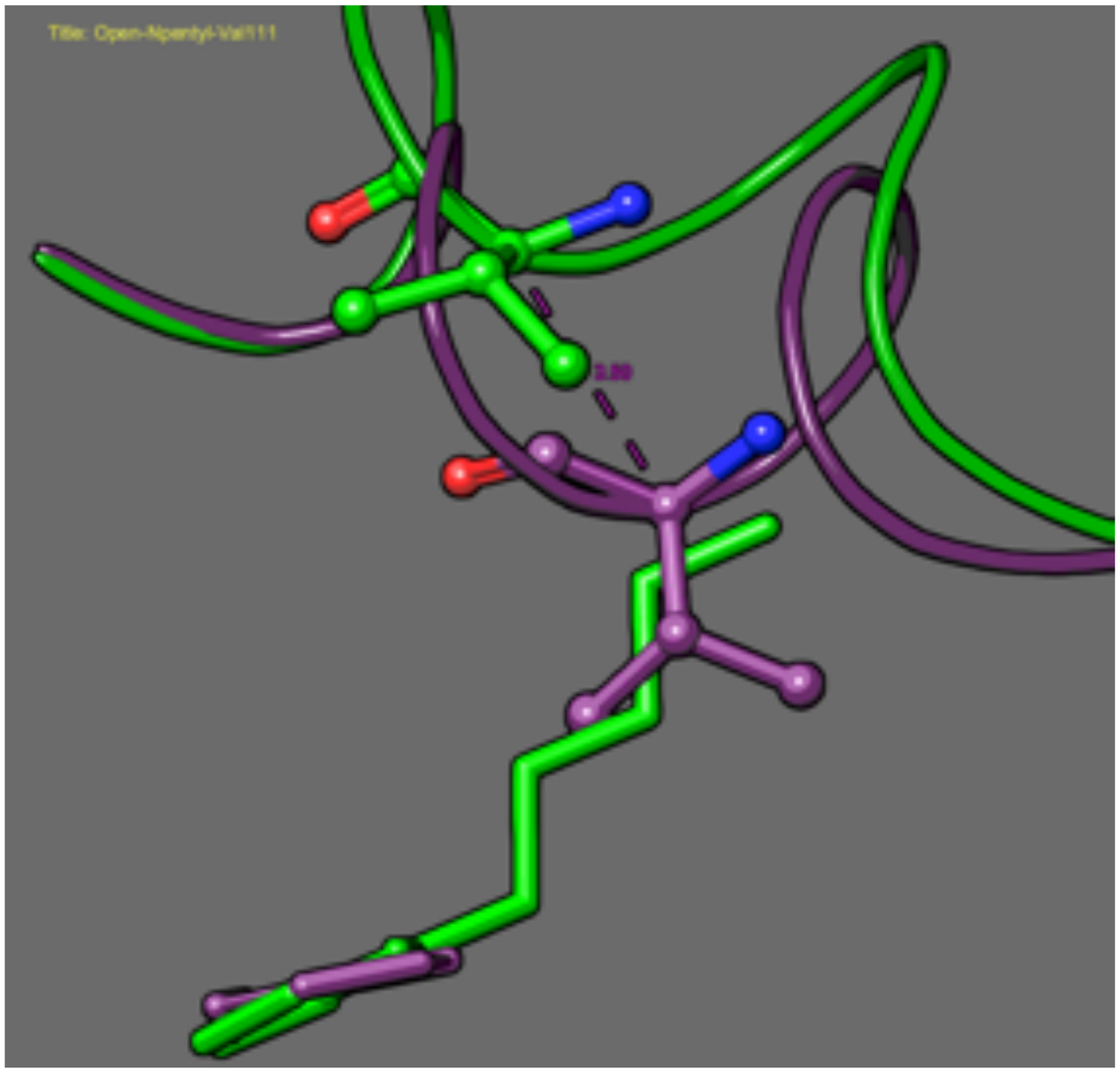


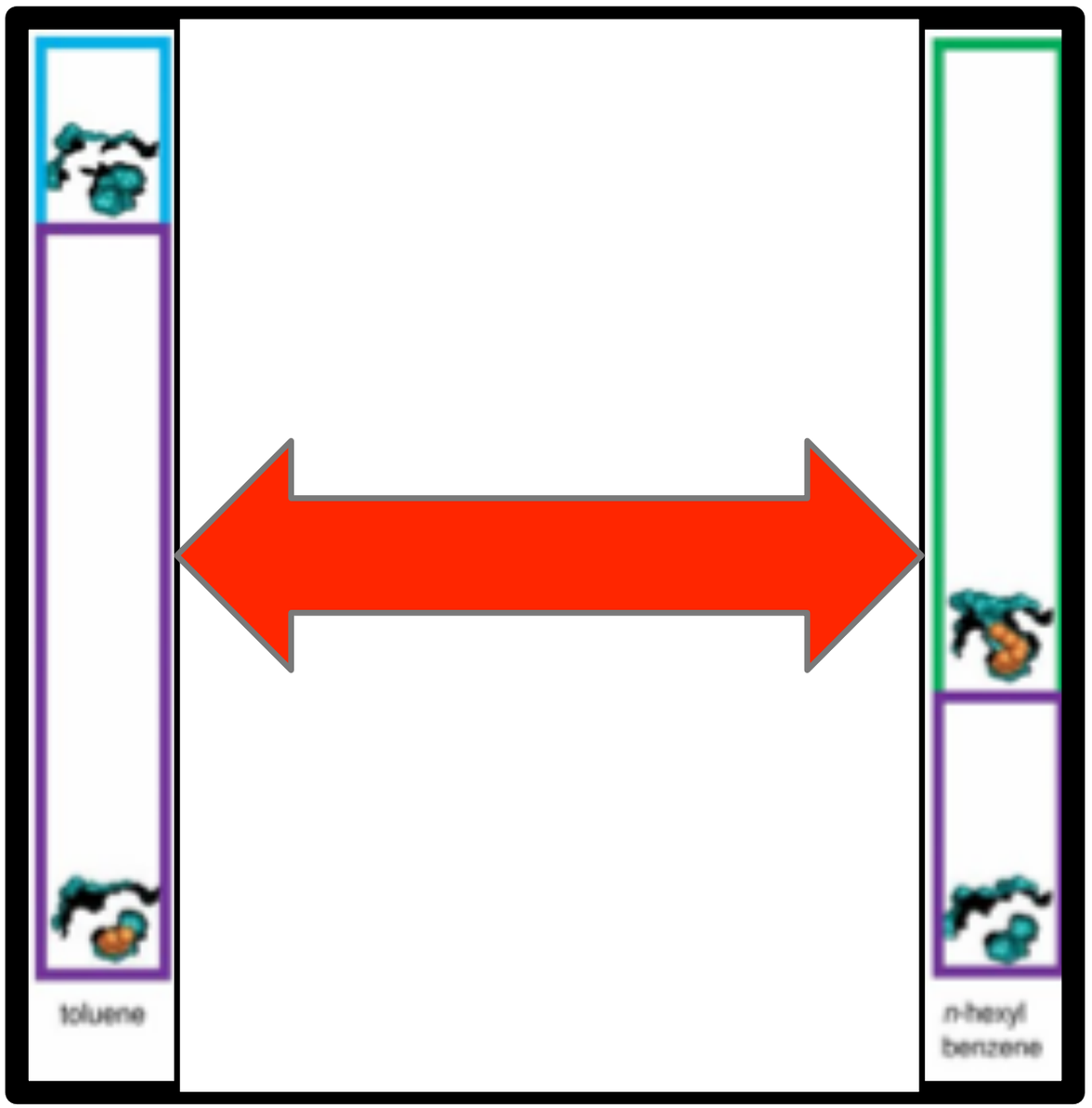
This is a good test system to see how protein conformational change affects relative free energies





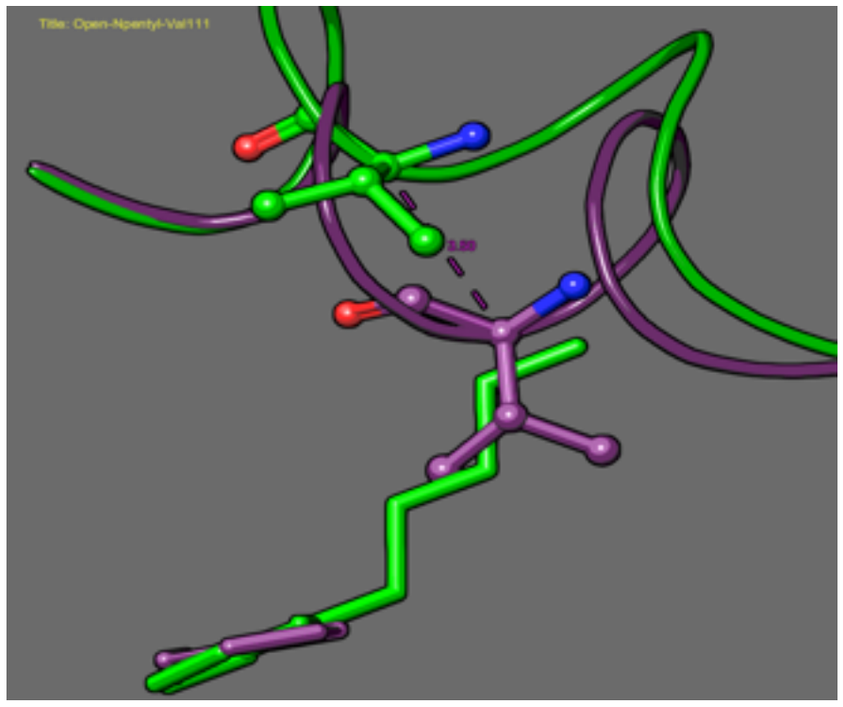
We start simulations from the protein **closed** or **open** conformation and check for consistency





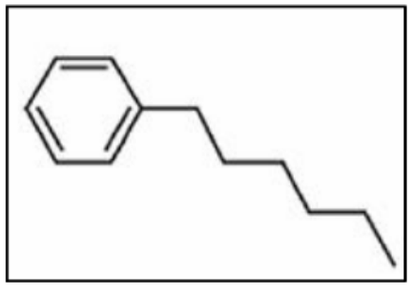
In many cases we find significant dependence on the starting structure for “large” perturbations

Closed vs Open ddG				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-hexyl	4.1	-0.6	**4.7**
toluene	n-hexyl	2.9	-1.6	**4.5**
ethyl	n-hexyl	3.6	-0.8	**4.4**
n-propyl	n-hexyl	5.9	0.1	**5.8**

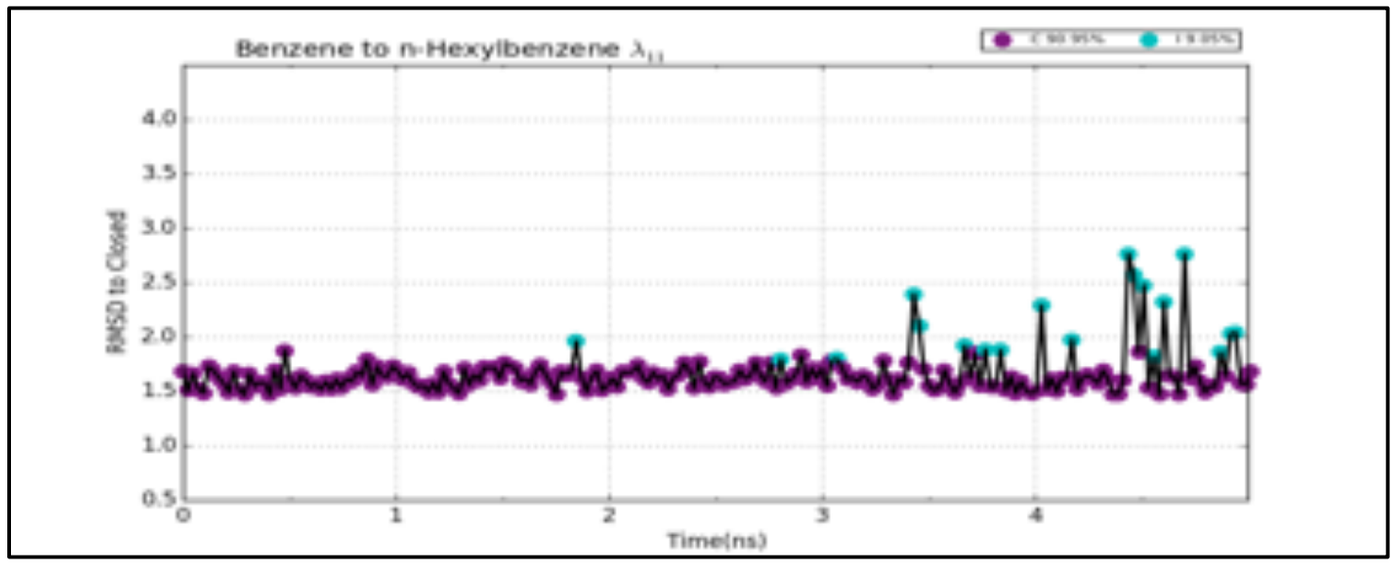


C-O RMSE: **4.9** kcal/mol

This is a problem of kinetic trapping or slow conformational change

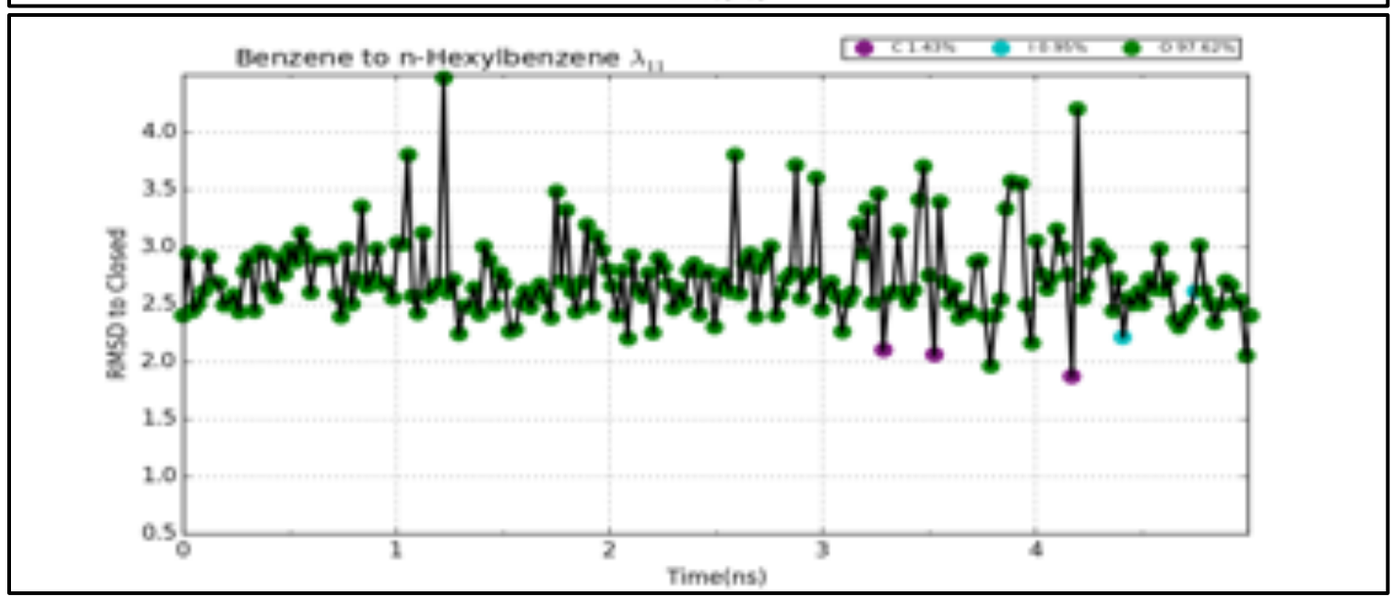


Exp: 30% occupancy
Start from Closed:
Trapped in **Closed**

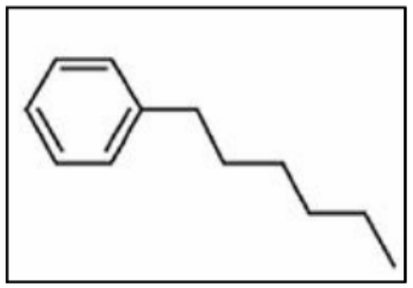


5ns Closed vs Open ddG predicted				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-hexyl	4.1	-0.6	**4.7**

Exp: 70% occupancy
Start from Open:
Trapped in **Open**



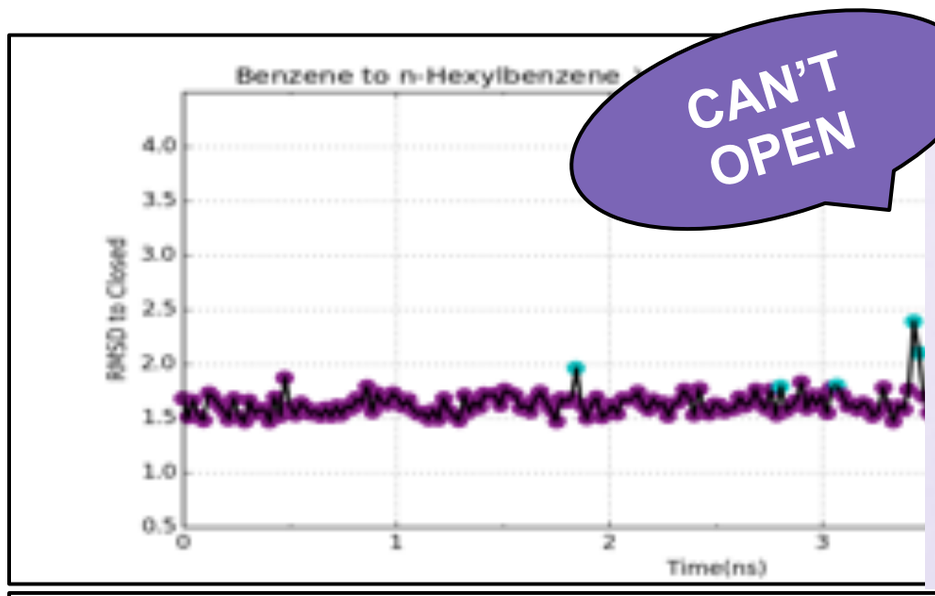
This is a problem of kinetic trapping or slow conformational change

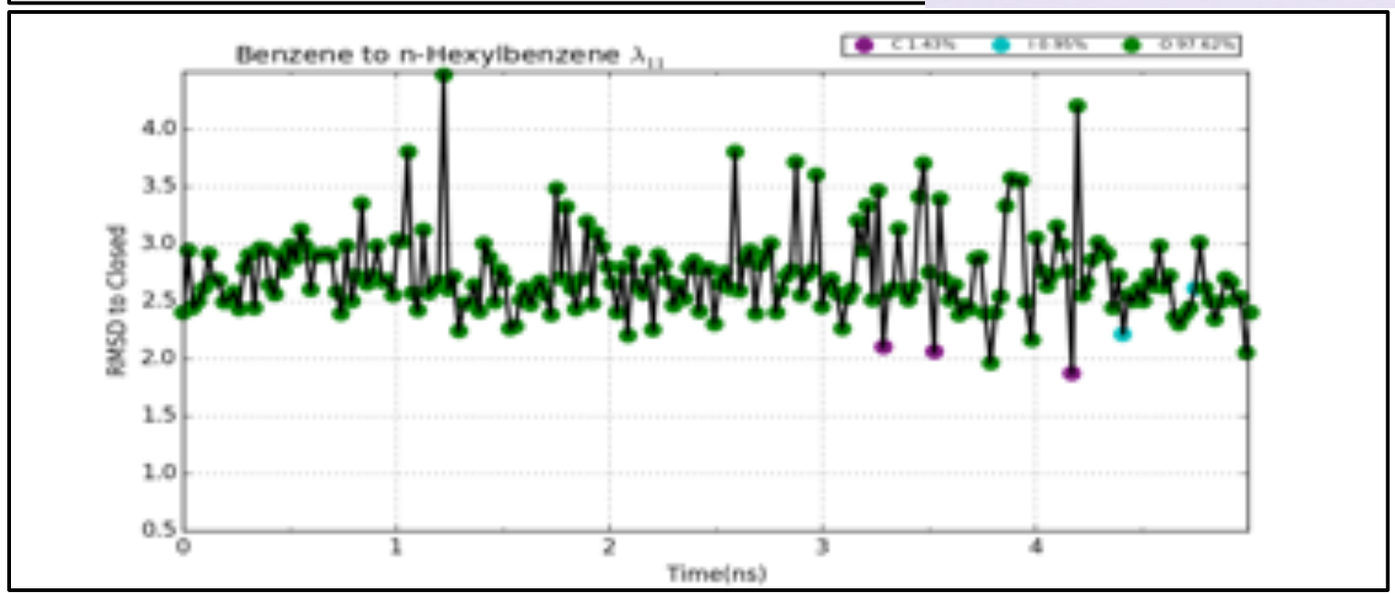


Exp: 30% occupancy
Start from Closed:
Trapped in **Closed**

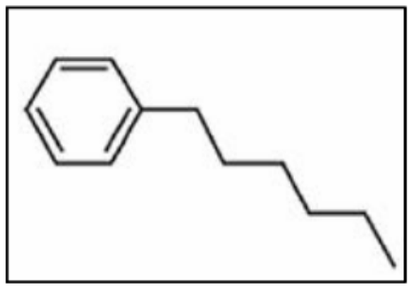
5ns Closed vs Open ddG predicted				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-hexyl	4.1	-0.6	**4.7**

Exp: 70% occupancy
Start from Open:
Trapped in **Open**





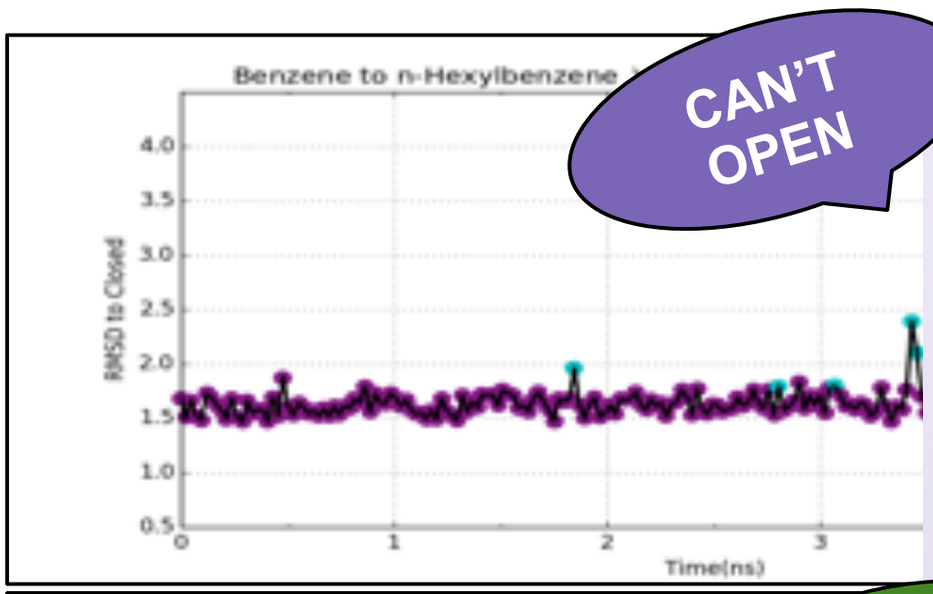
This is a problem of kinetic trapping or slow conformational change

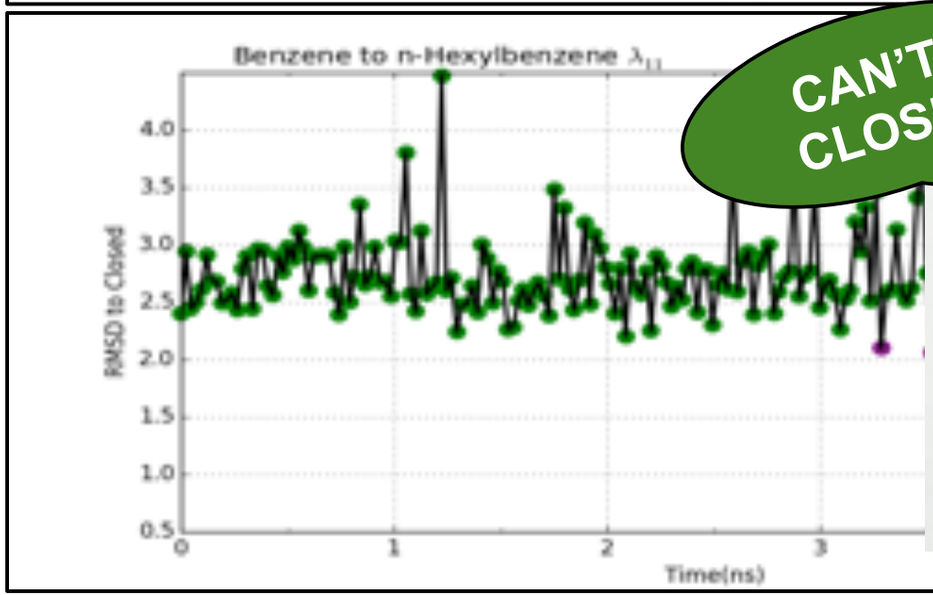


Exp: 30% occupancy
Start from Closed:
Trapped in **Closed**

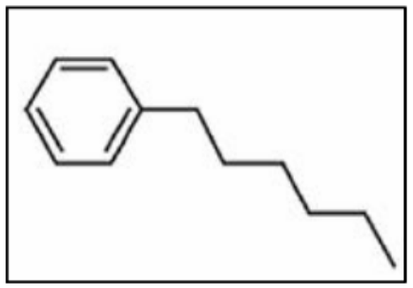
5ns Closed vs Open ddG predicted				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-hexyl	4.1	-0.6	**4.7**

Exp: 70% occupancy
Start from Open:
Trapped in **Open**





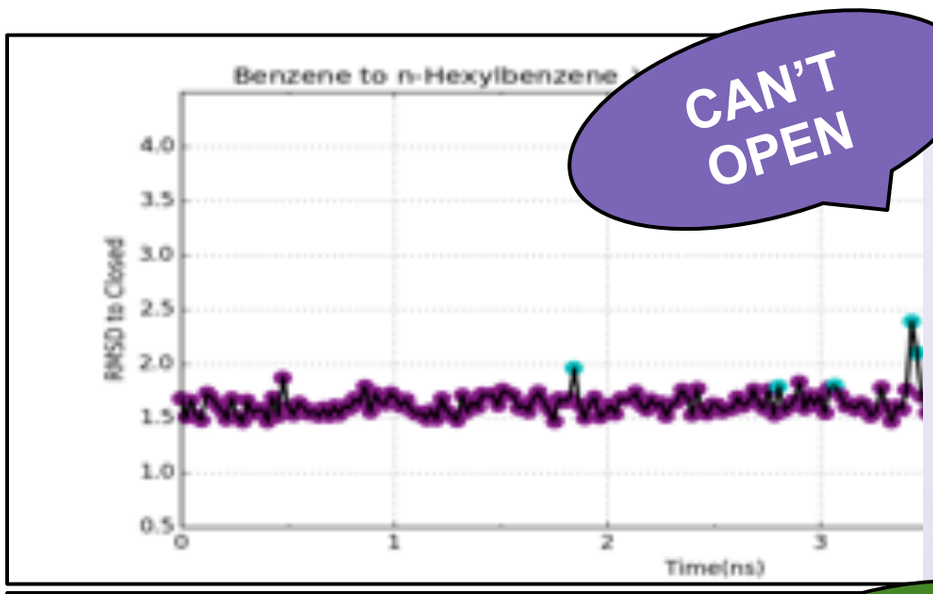
This is a problem of kinetic trapping or slow conformational change

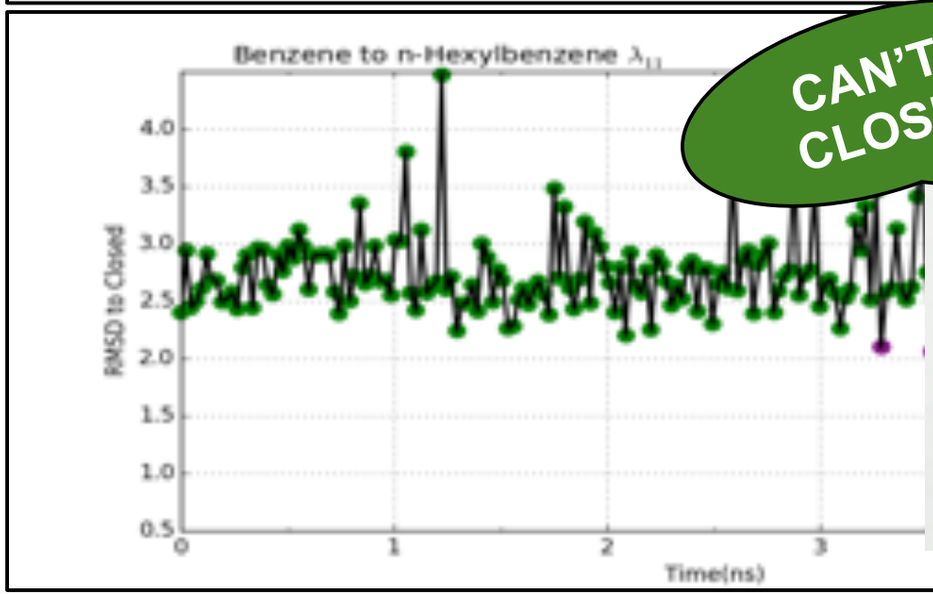


Exp: 30% occupancy
Start from Closed:
Trapped in **Closed**

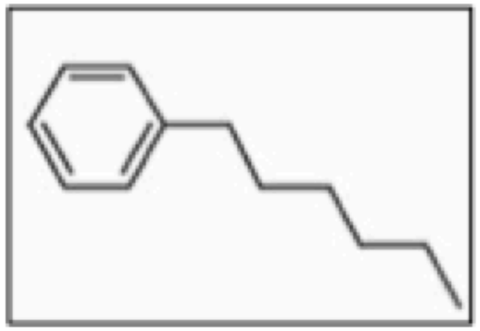
5ns Closed vs Open ddG predicted				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-hexyl	4.1	-0.6	**4.7**

Exp: 70% occupancy
Start from Open:
Trapped in **Open**

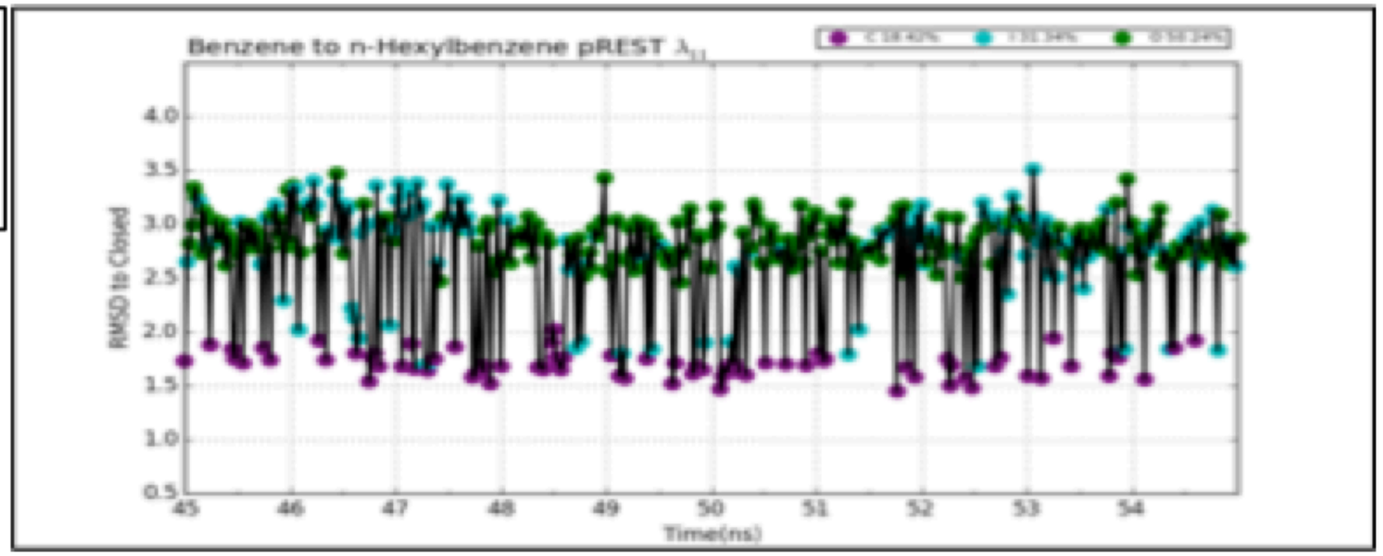




Simulating 10x longer and expanding the enhanced sampling region helps reduce discrepancy

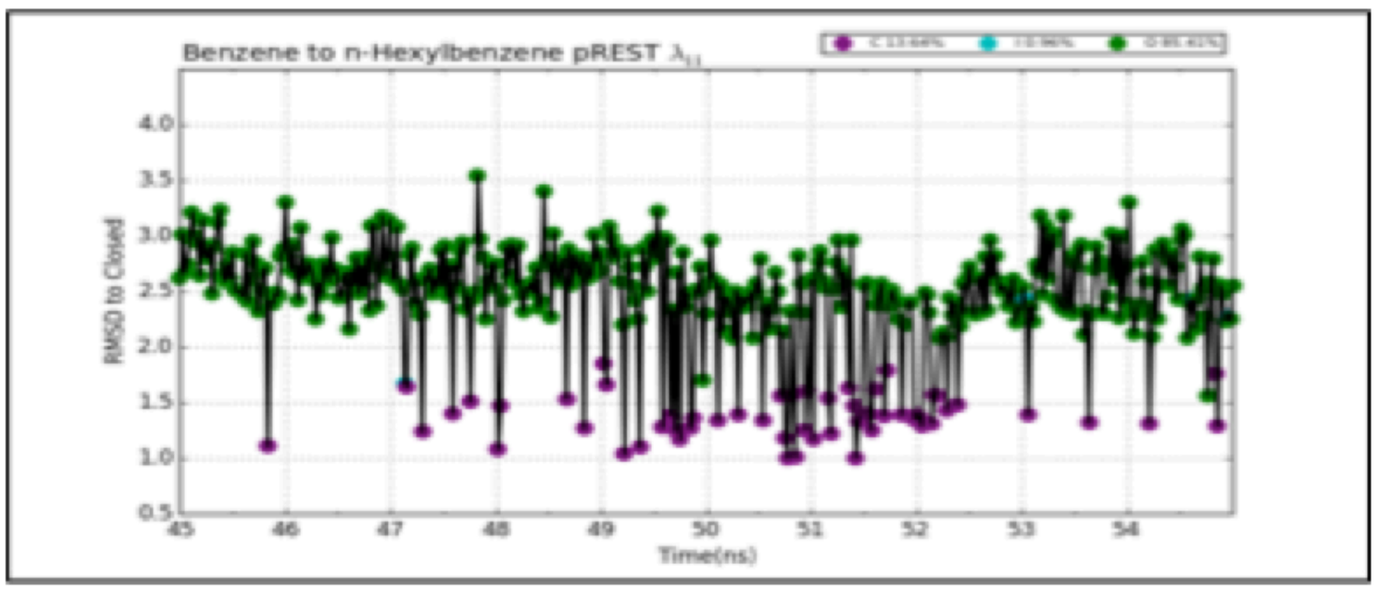


Exp: 30% occupancy
Start from Closed:
Some transitions

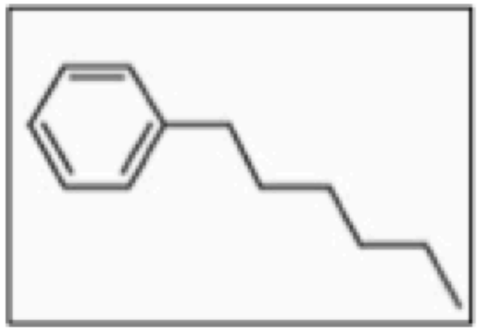


50-55ns Closed vs Open ddG predicted				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-hexyl	2.1	1.4	**0.7**

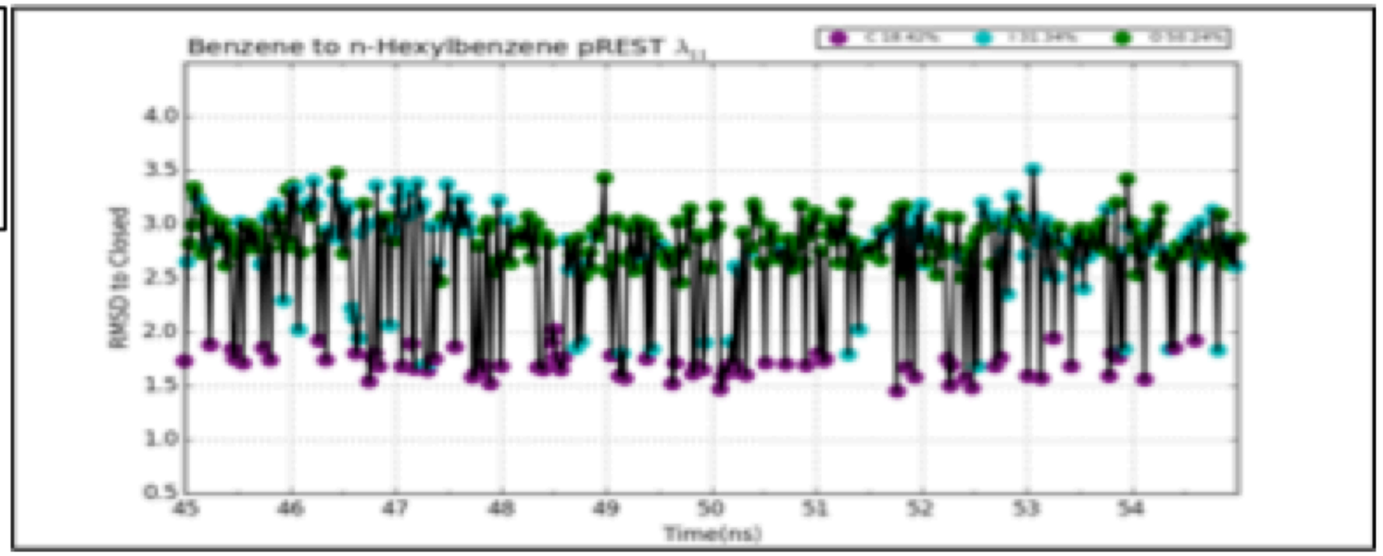
Exp: 70% occupancy
Start from Open:
Some transitions



Simulating 10x longer and expanding the enhanced sampling region helps reduce discrepancy

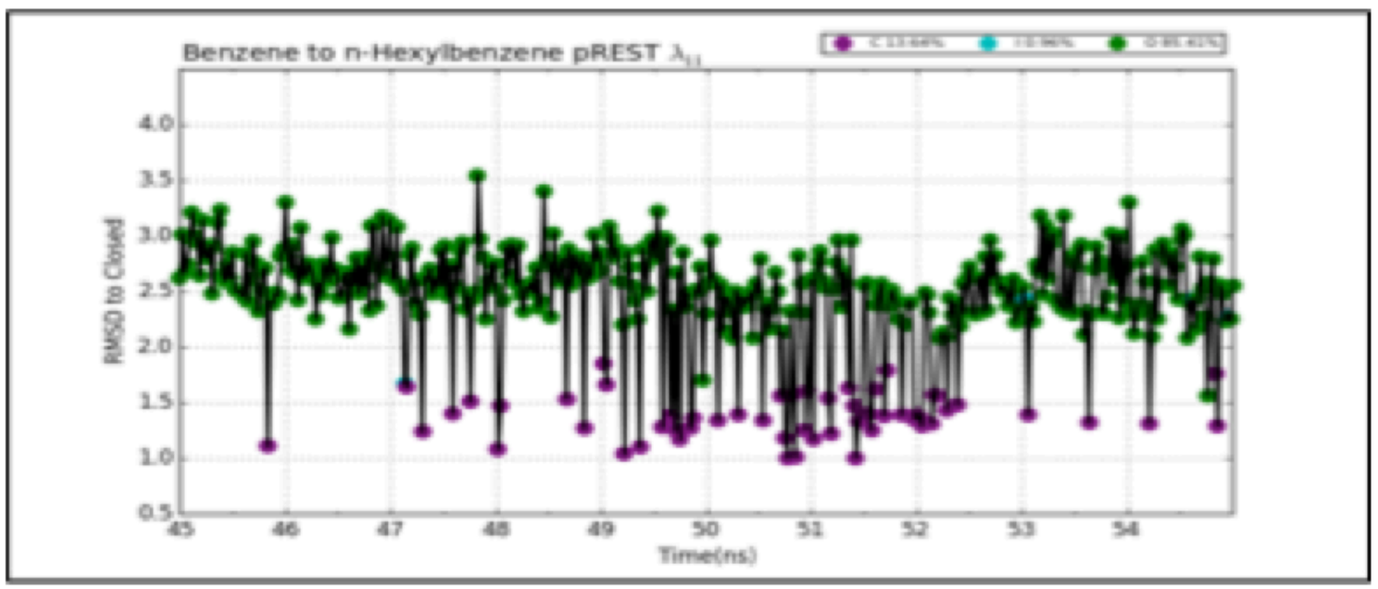


Exp: 30% occupancy
Start from Closed:
Some transitions



50-55ns Closed vs Open ddG predicted				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-hexyl	2.1	1.4	**0.7**

Exp: 70% occupancy
Start from Open:
Some transitions

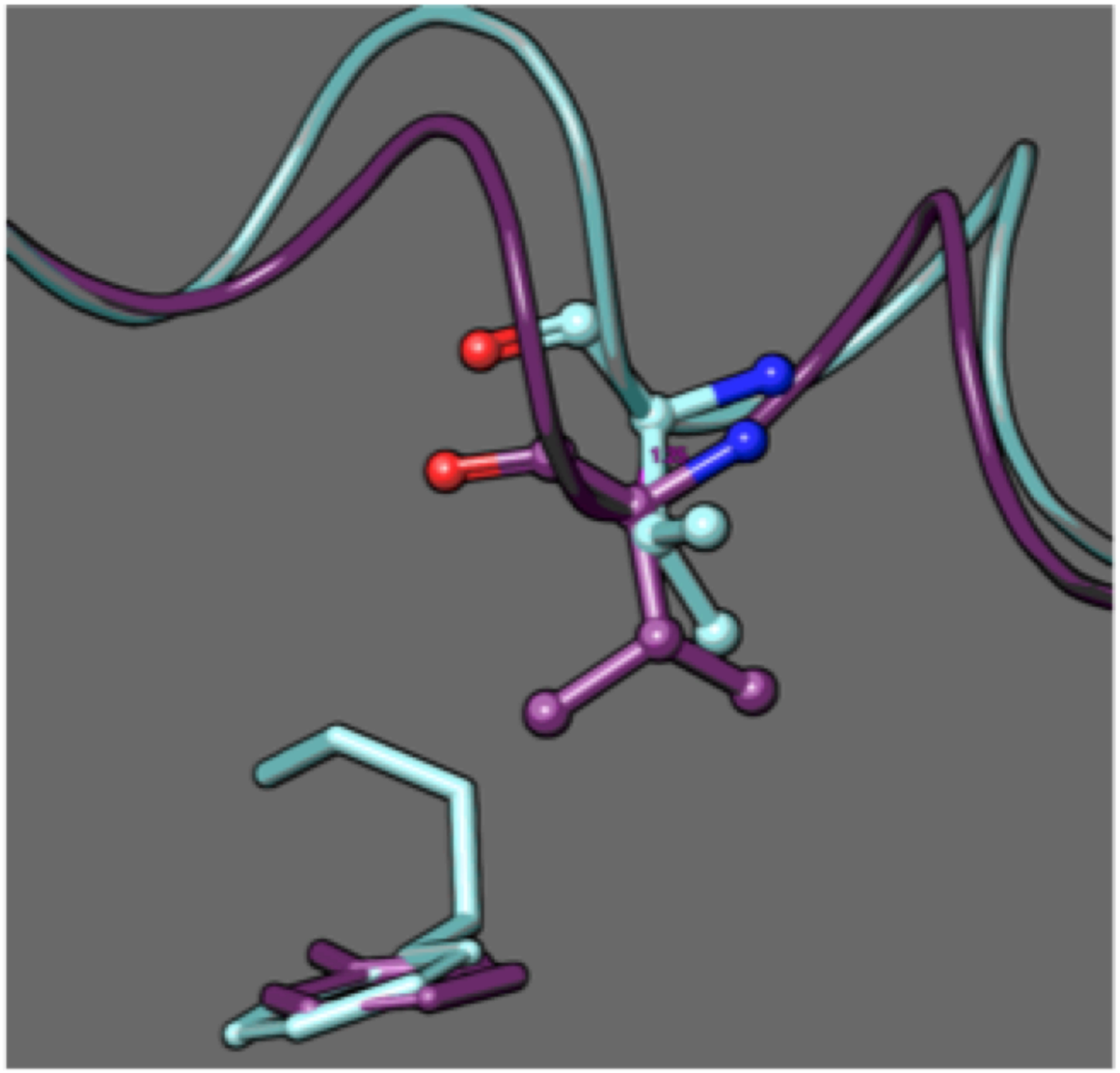


For smaller perturbations, dependence on the starting structure is almost eliminated

Time: 5ns FF: OPLS3
Method: FEP/(p)REST

Closed vs Open ddG calculated				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-butyl	-0.1	-0.7	**0.6**
toluene	n-butyl	+0.9	-0.4	**1.3**
ethyl	n-butyl	-0.2	-0.4	**0.2**
n-propyl	n-butyl	+1.0	+0.5	**0.5**

C-O RMSE: **0.8** kcal/mol

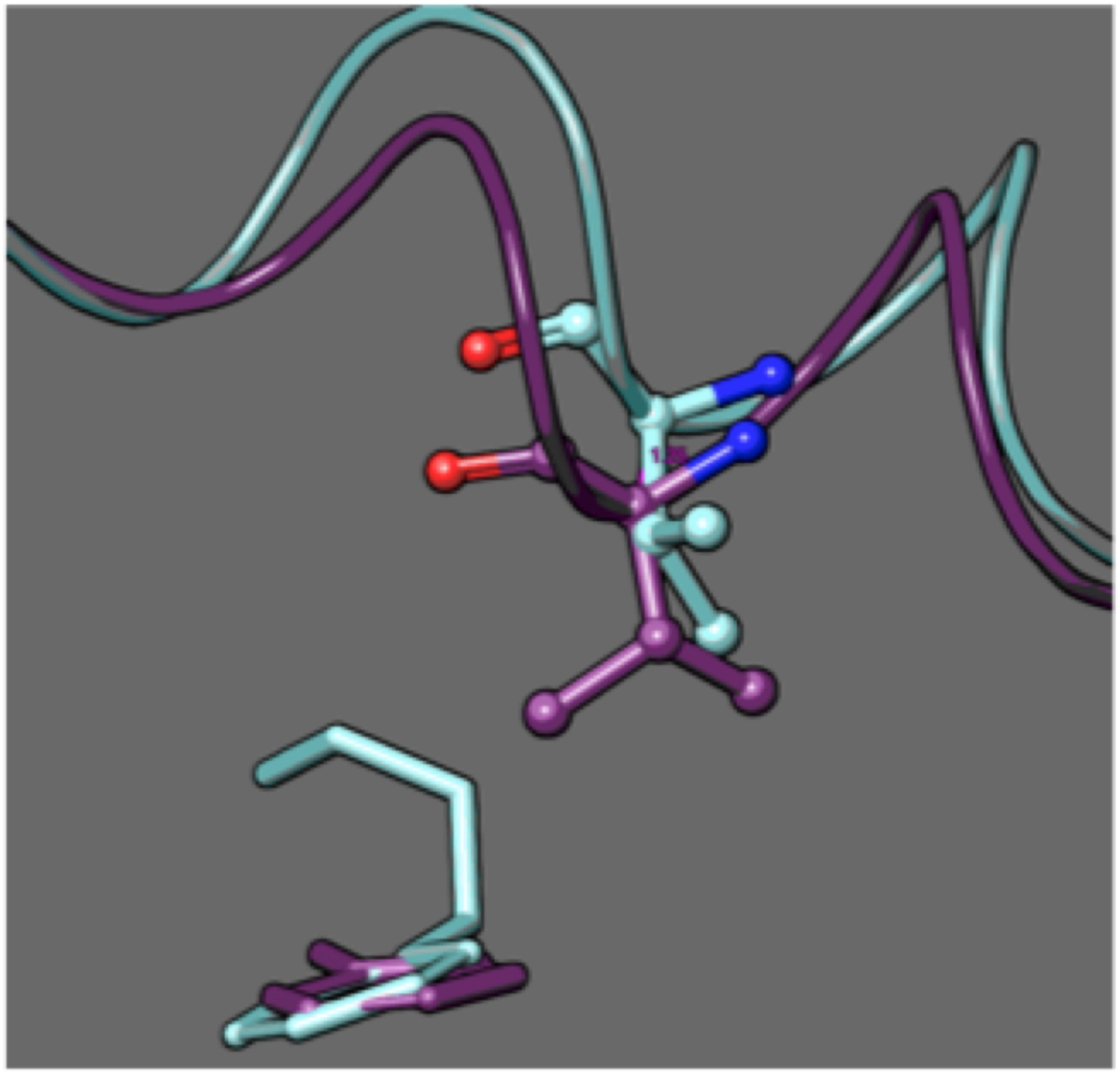


For smaller perturbations, dependence on the starting structure is almost eliminated

Time: 5ns FF: OPLS3
Method: FEP/(p)REST

Closed vs Open ddG calculated				
Lig 1	**Lig 2**	**Closed**	**Open**	***C-O Diff***
benzene	n-butyl	-0.1	-0.7	**0.6**
toluene	n-butyl	+0.9	-0.4	**1.3**
ethyl	n-butyl	-0.2	-0.4	**0.2**
n-propyl	n-butyl	+1.0	+0.5	**0.5**

C-O RMSE: **0.8** kcal/mol

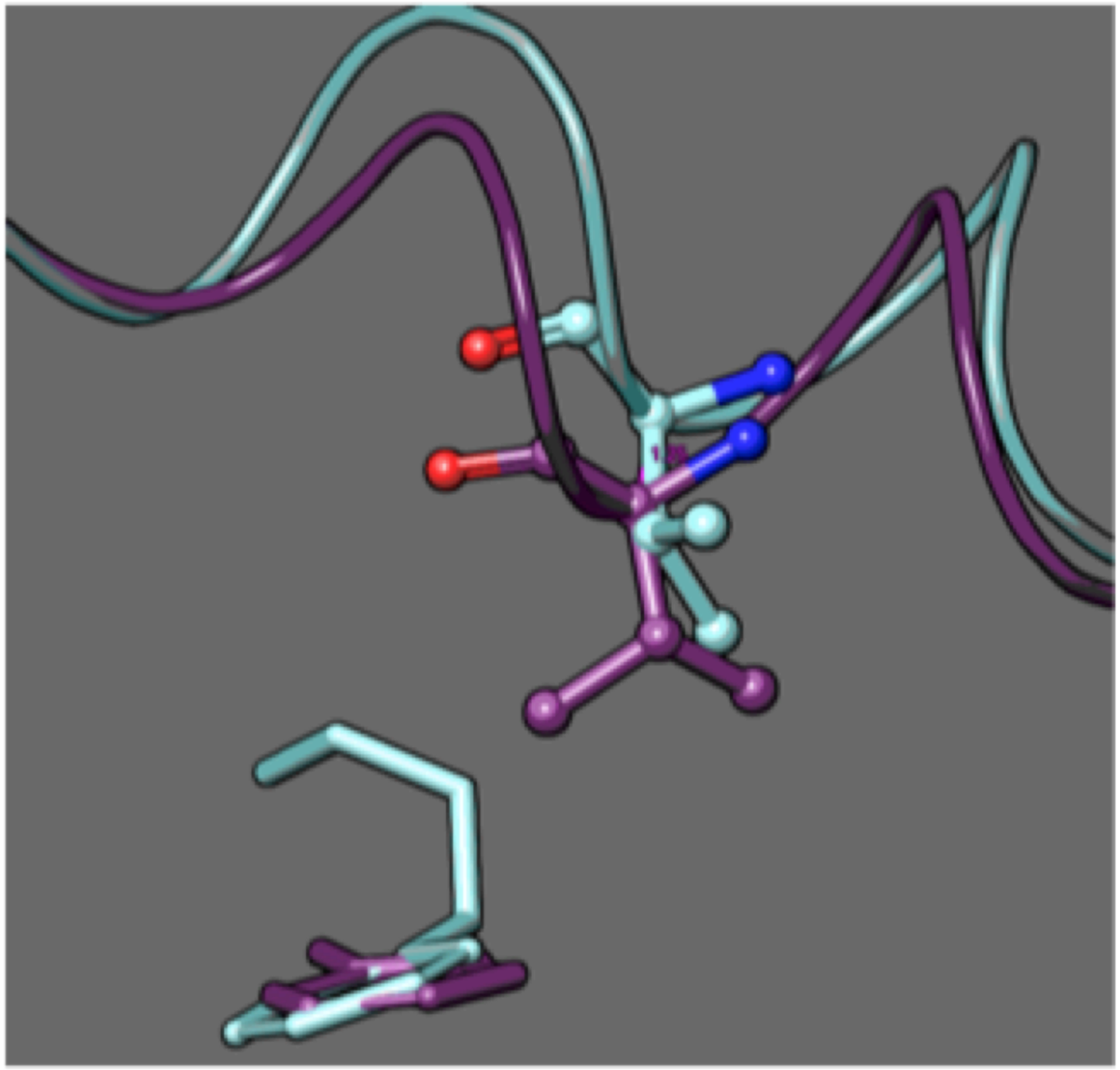


But the discrepancy is still in a consistent direction, increasing error relative to experiment

Time: 5ns FF: OPLS3

Method: FEP/(p)REST

Error from Exp ddG			
Lig 1	**Lig 2**	**Closed**	**Open**
benzene	n-butyl	1.4	0.8
toluene	n-butyl	2.1	0.8
ethyl	n-butyl	0.7	0.5
n-propyl	n-butyl	1.2	0.6
EXP RMSE:		1.4	**0.70**

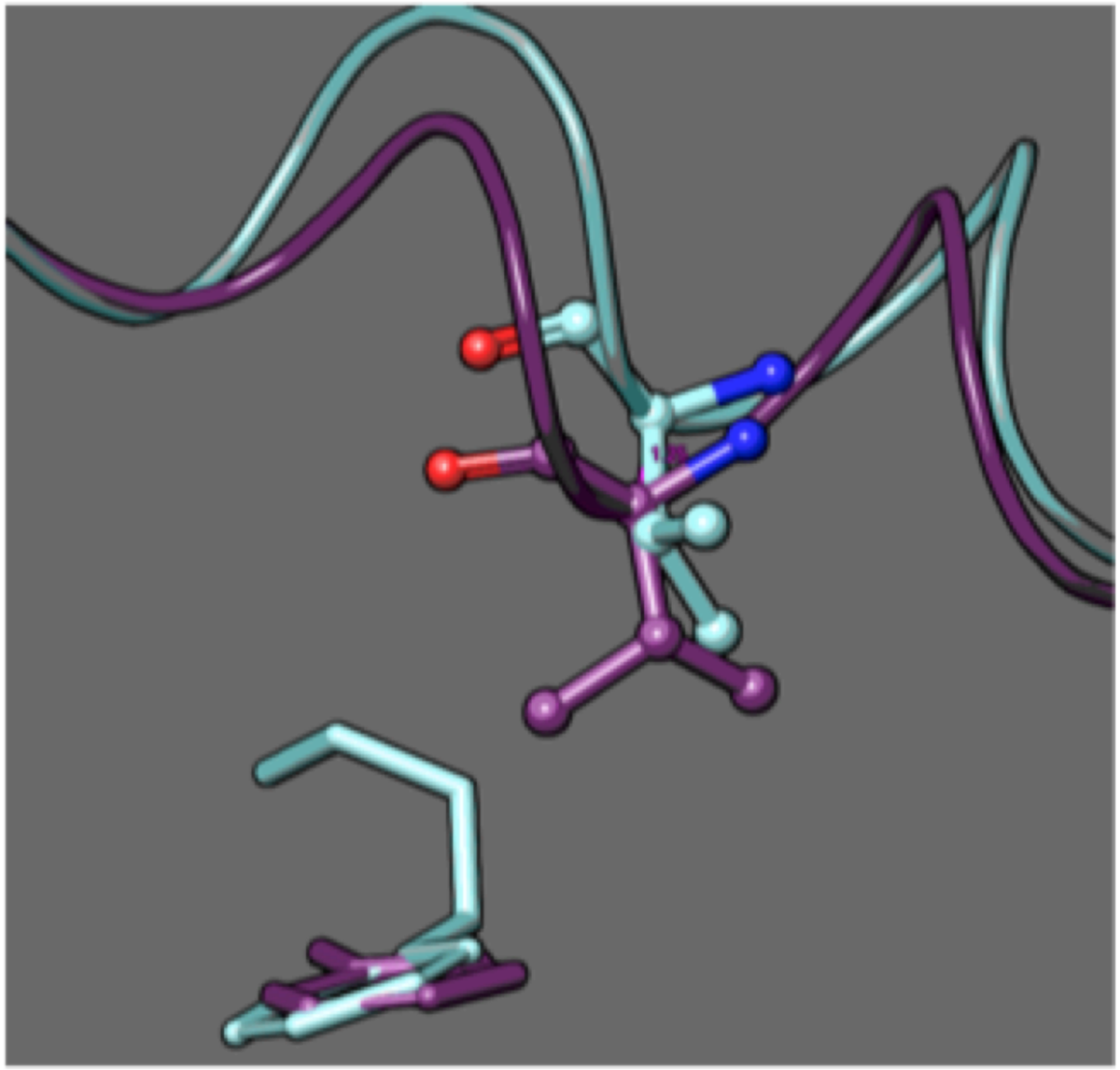


But the discrepancy is still in a consistent direction, increasing error relative to experiment

Time: 5ns FF: OPLS3

Method: FEP/(p)REST

Error from Exp ddG			
Lig 1	**Lig 2**	**Closed**	**Open**
benzene	n-butyl	1.4	0.8
toluene	n-butyl	2.1	0.8
ethyl	n-butyl	0.7	0.5
n-propyl	n-butyl	1.2	0.6
EXP RMSE:		1.4	**0.70**

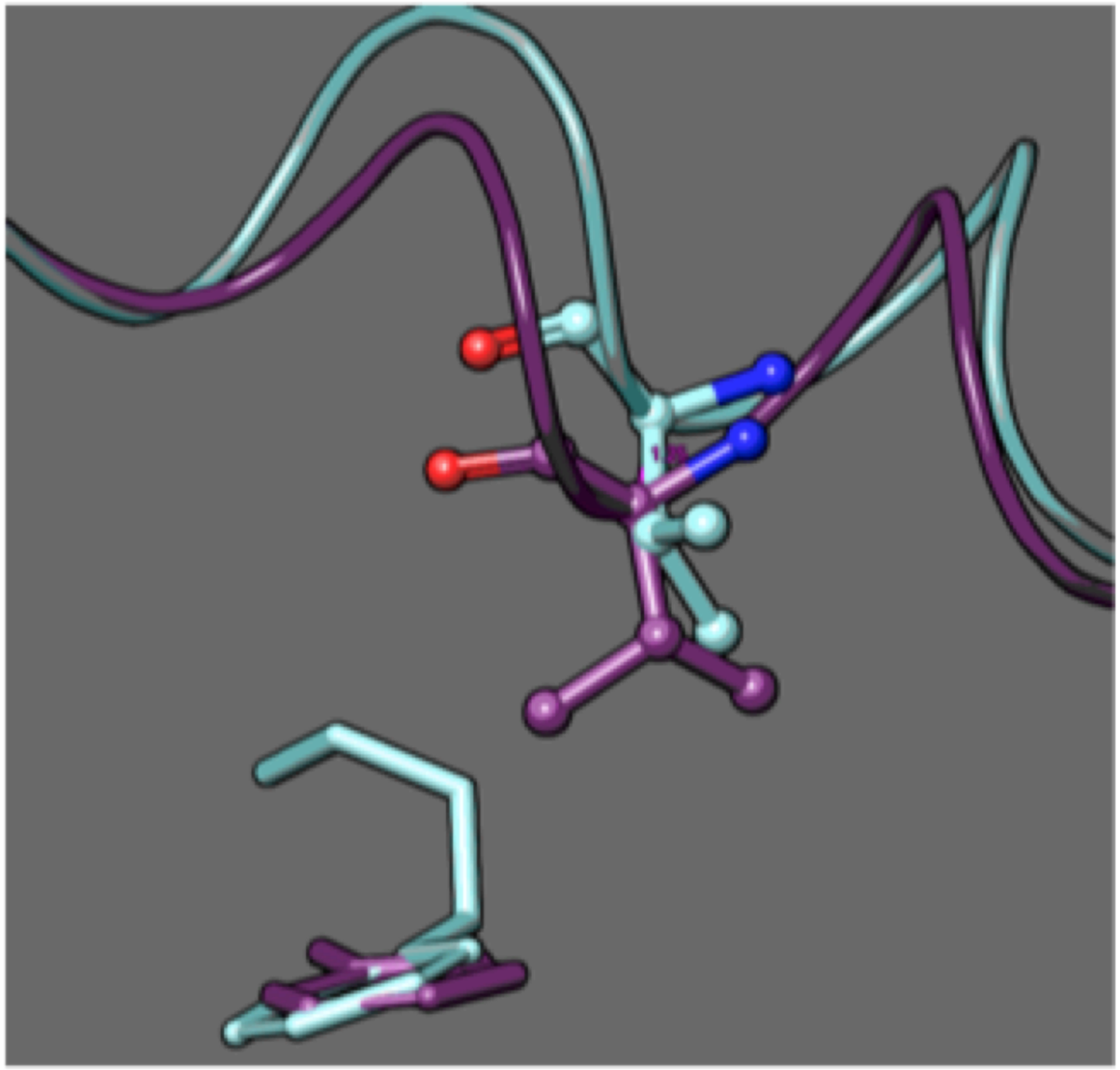


But the discrepancy is still in a consistent direction, increasing error relative to experiment

Time: 5ns FF: OPLS3

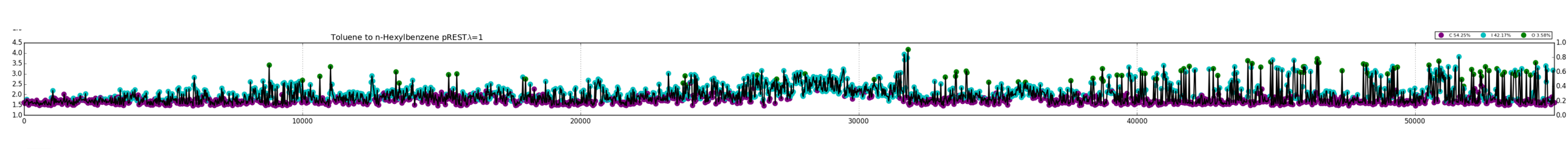
Method: FEP/(p)REST

Error from Exp ddG			
Lig 1	**Lig 2**	**Closed**	**Open**
benzene	n-butyl	1.4	0.8
toluene	n-butyl	2.1	0.8
ethyl	n-butyl	0.7	0.5
n-propyl	n-butyl	1.2	0.6
EXP RMSE:		1.4	**0.70**

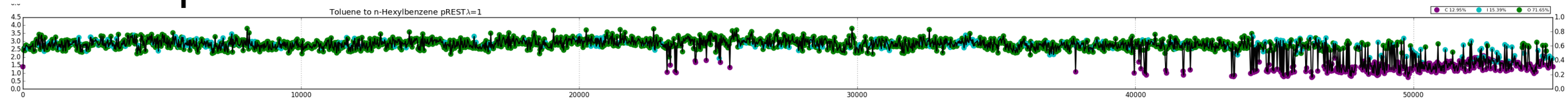


Note: For these “large” transformations such as toluene to n-hexylbenzene, it takes 50 ns per window to get anywhere close

From closed:

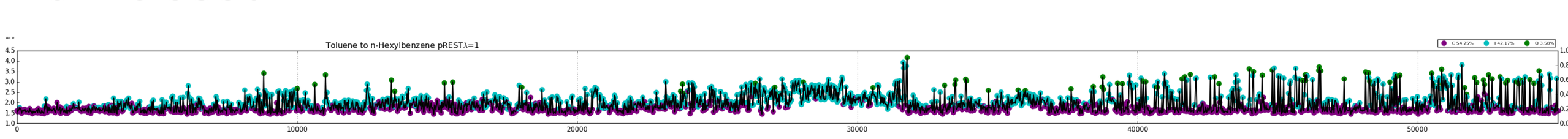


From open:

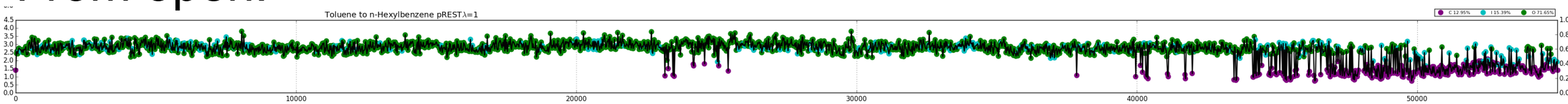


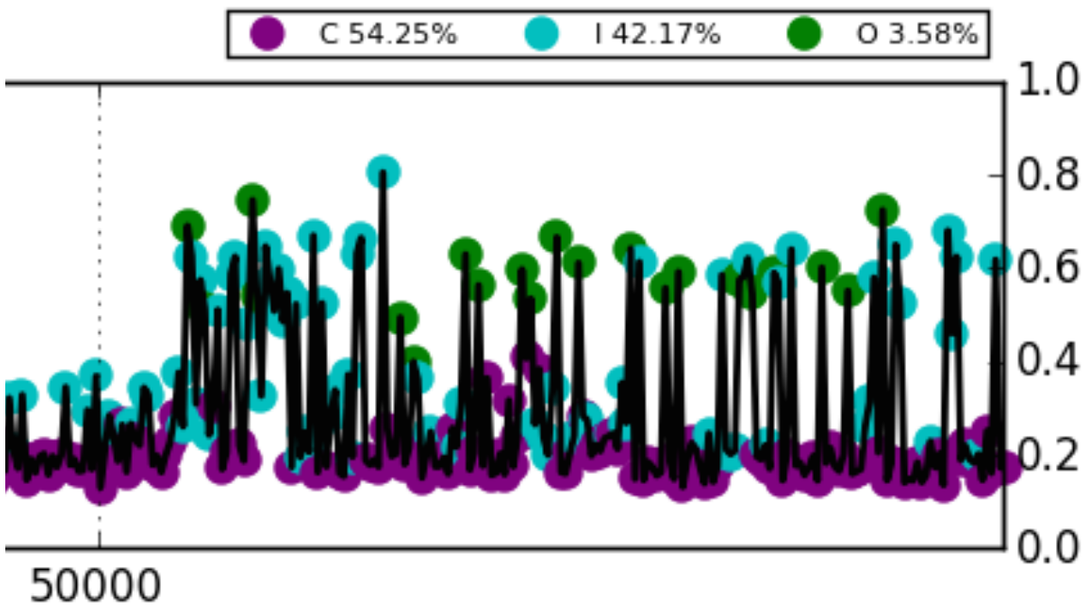
Note: For these “large” transformations such as toluene to n-hexylbenzene, it takes 50 ns per window to get anywhere close

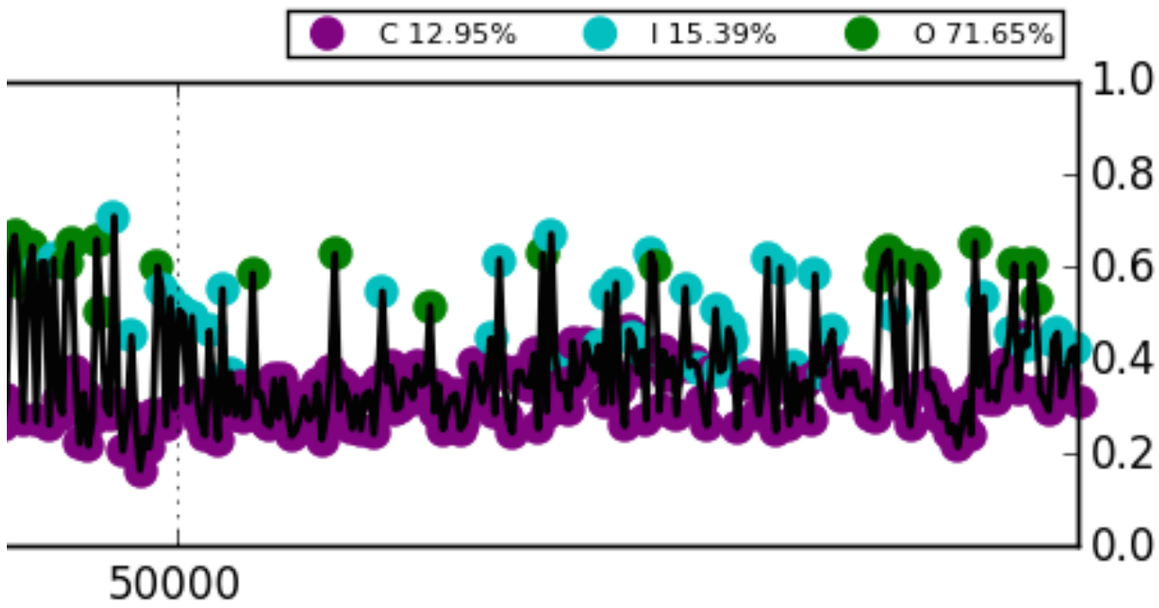
From closed:



From open:



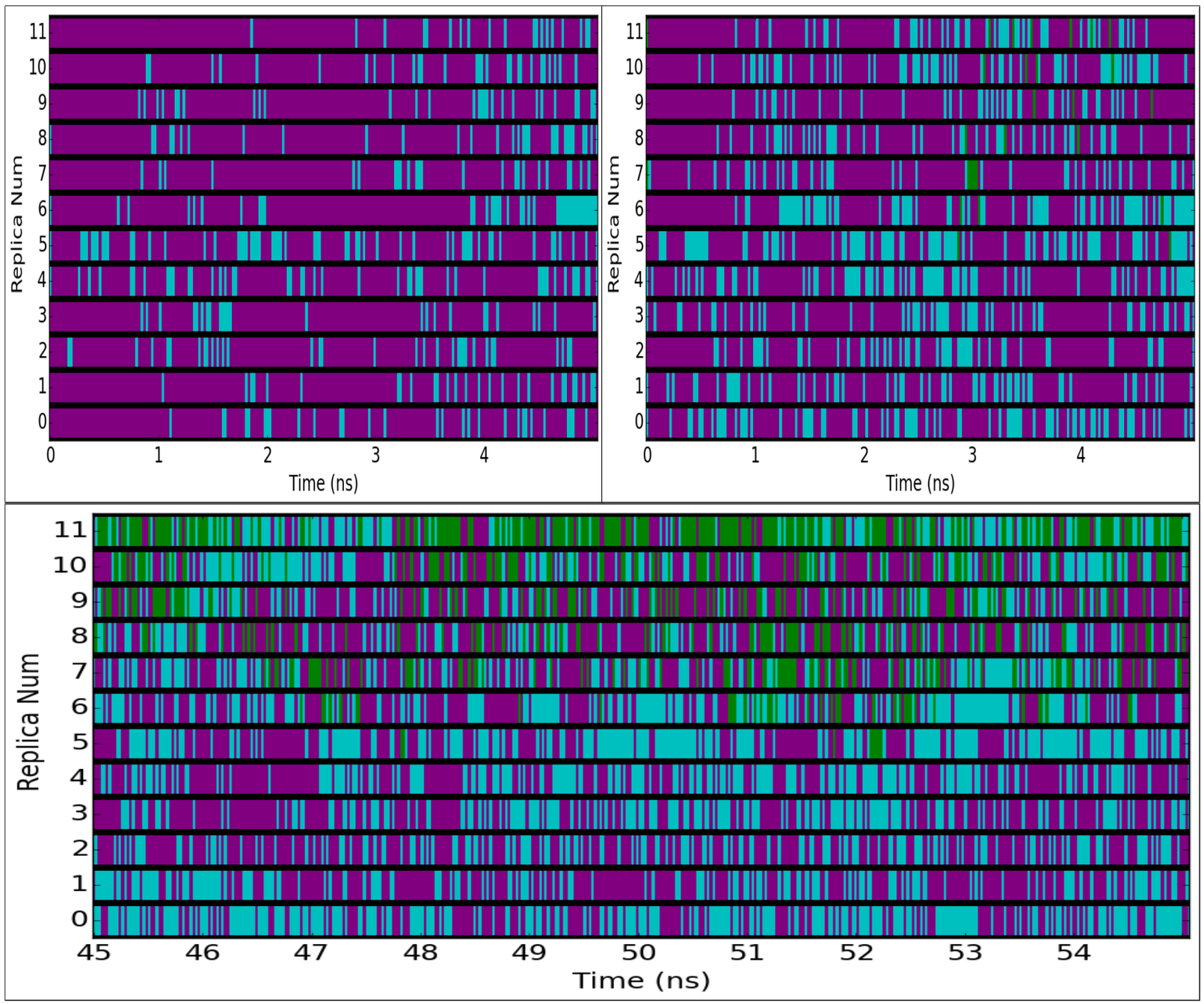




Really, I should be showing you mixing across all lambda windows

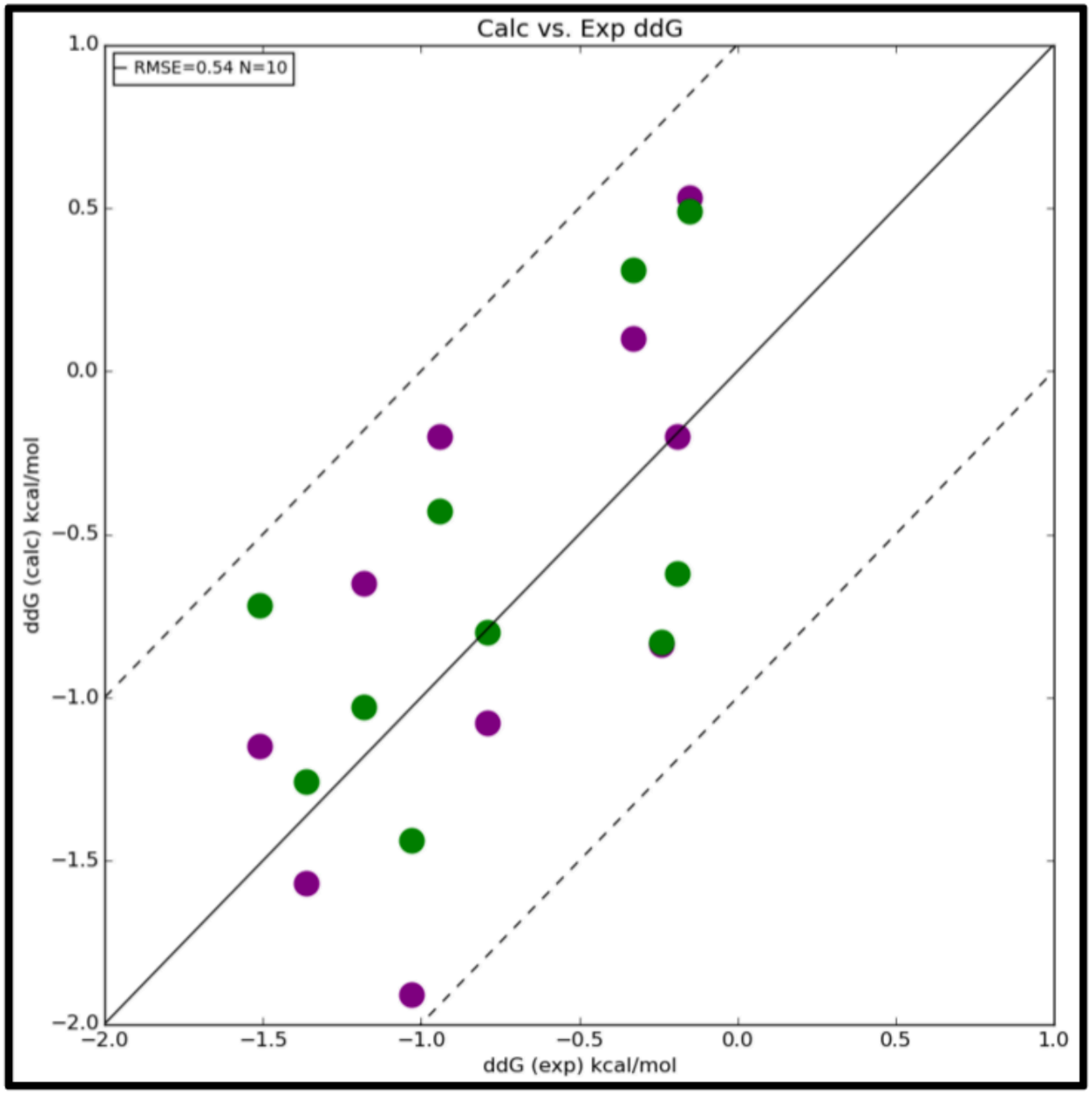
(a) Default

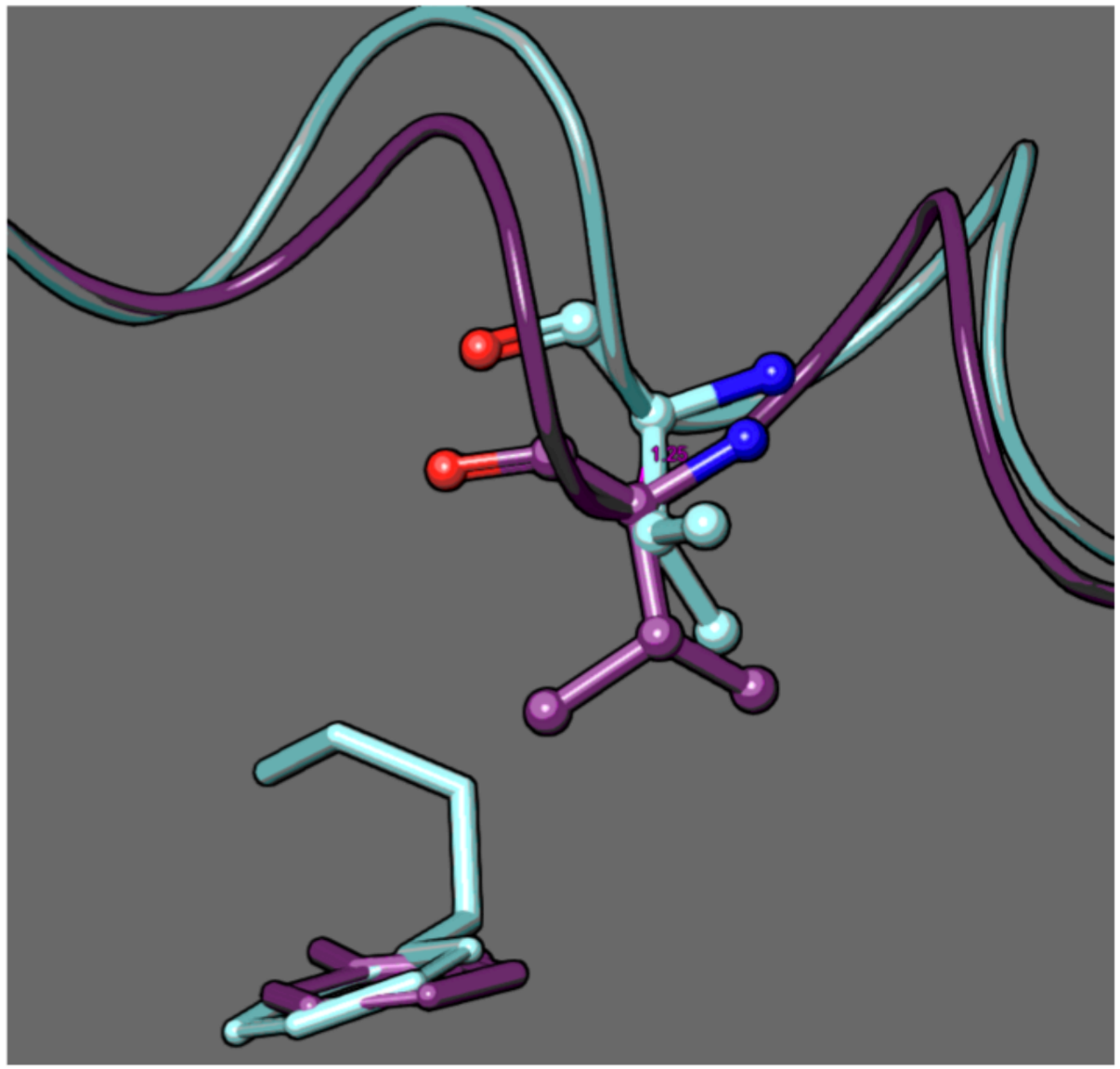
(b) pREST



(c) pREST extended

Overall, with longer simulations, we recapitulate experiment fairly well

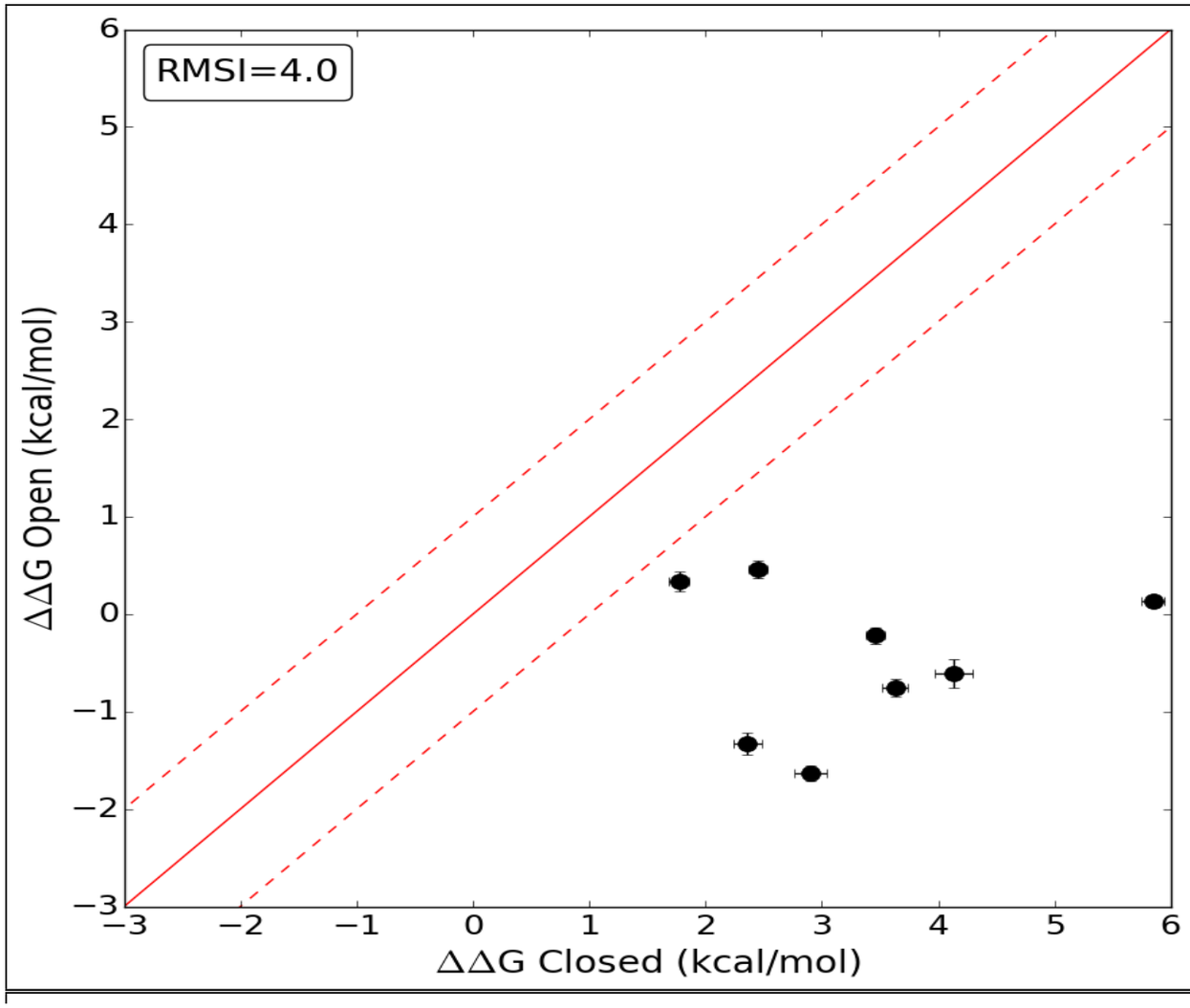


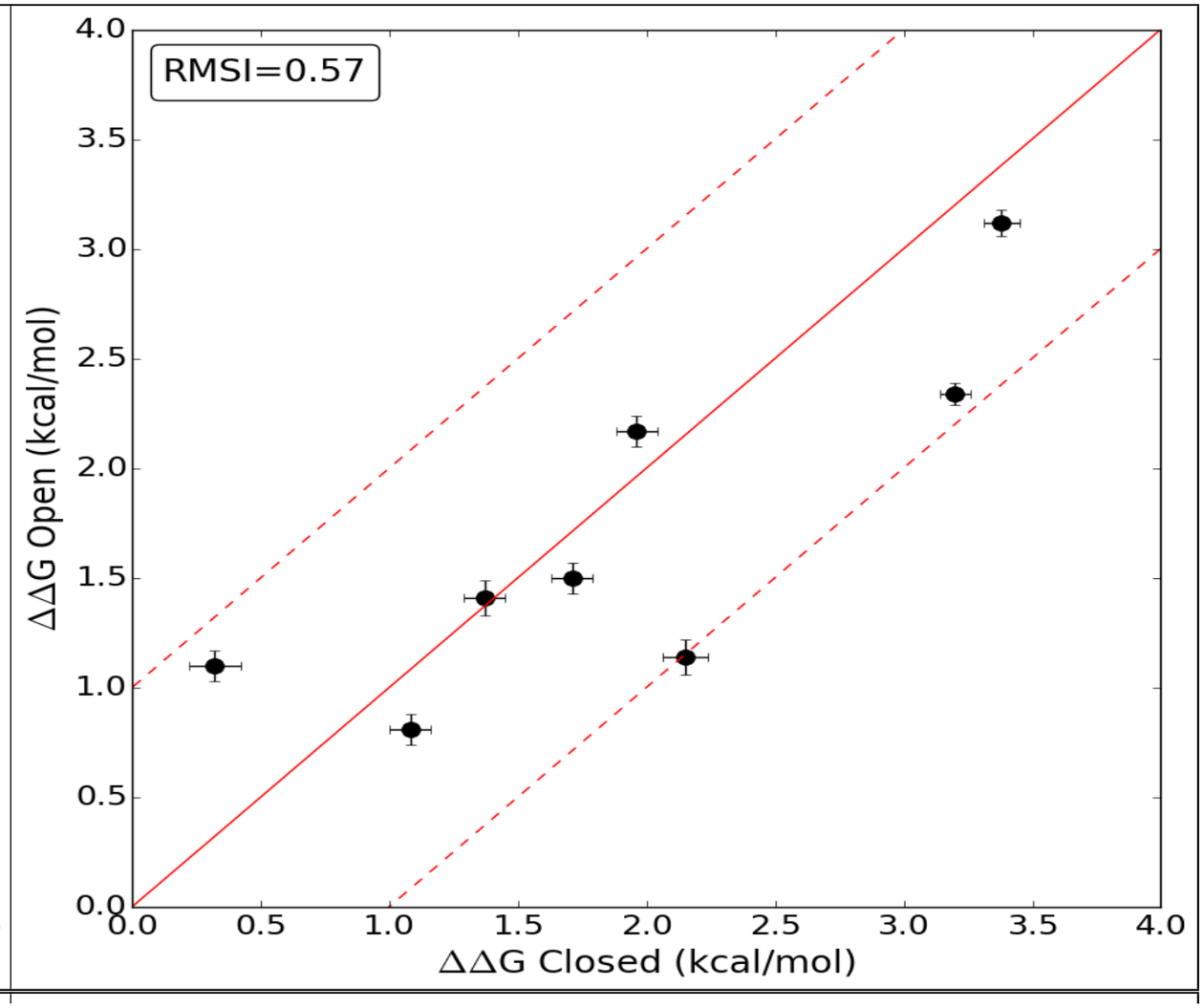


Longer simulations and pREST mostly remove the inconsistency

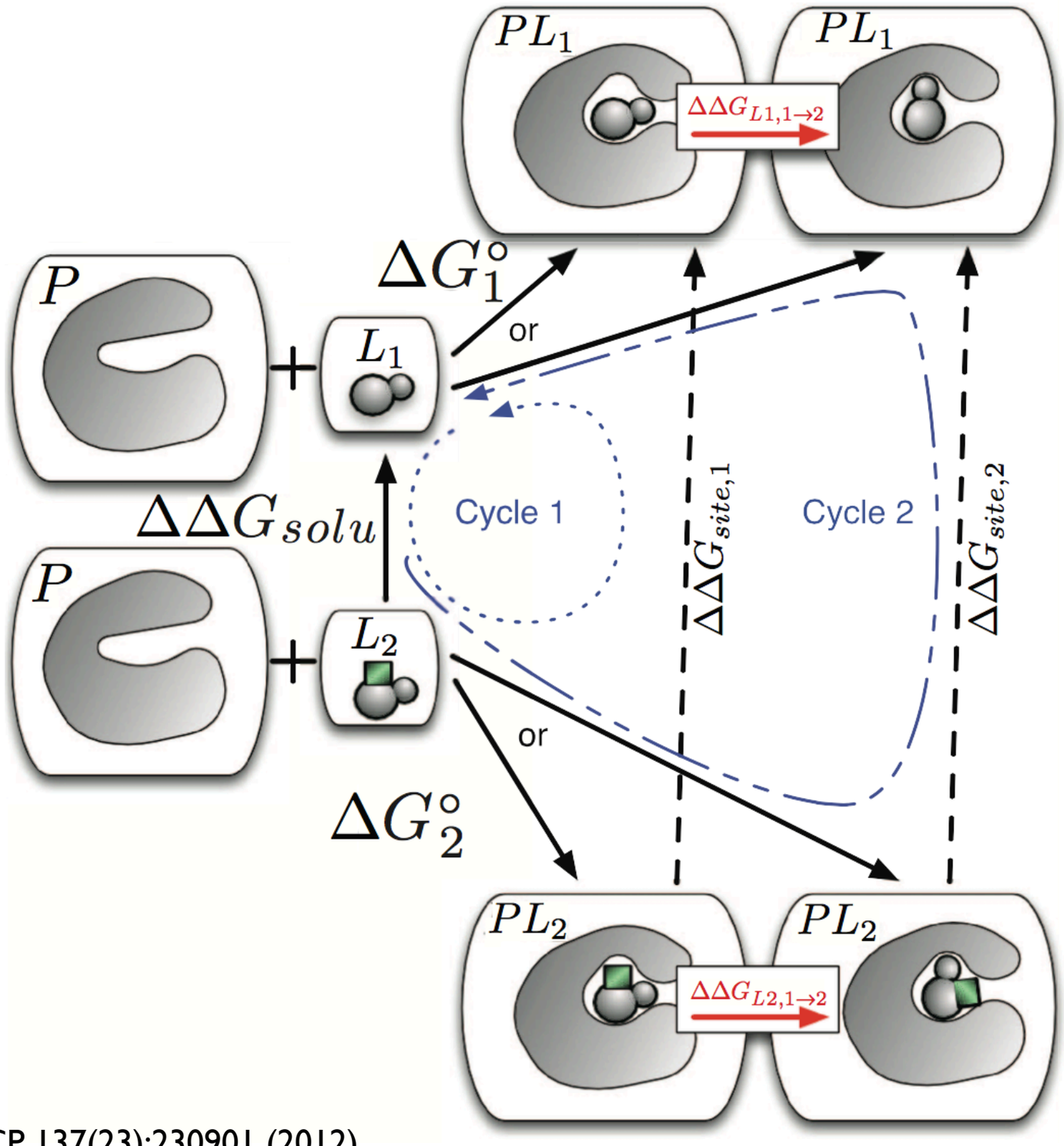
(a) Closed-Open: Default

(b) Closed-Open: pREST



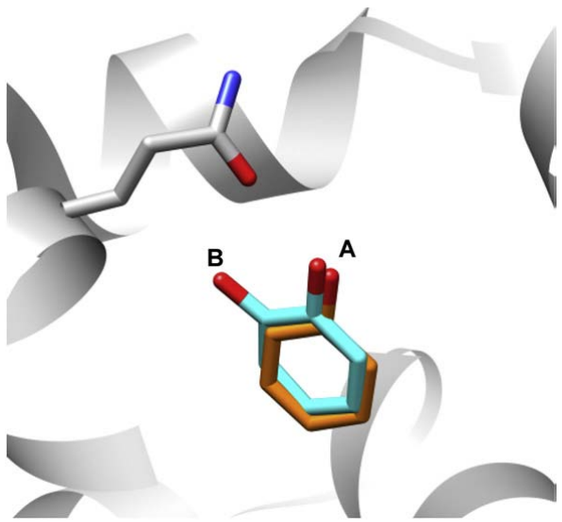


This means we need to sample a *lot* better in general, or use separation of states approaches



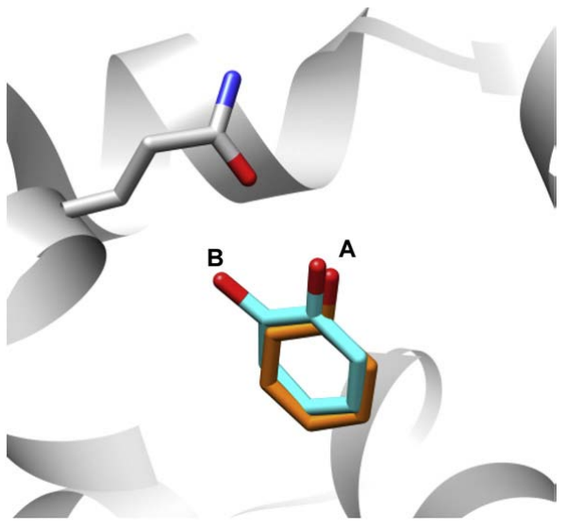
Klimovich and Mobley, JCP 137(23):230901 (2012)

What did we learn?

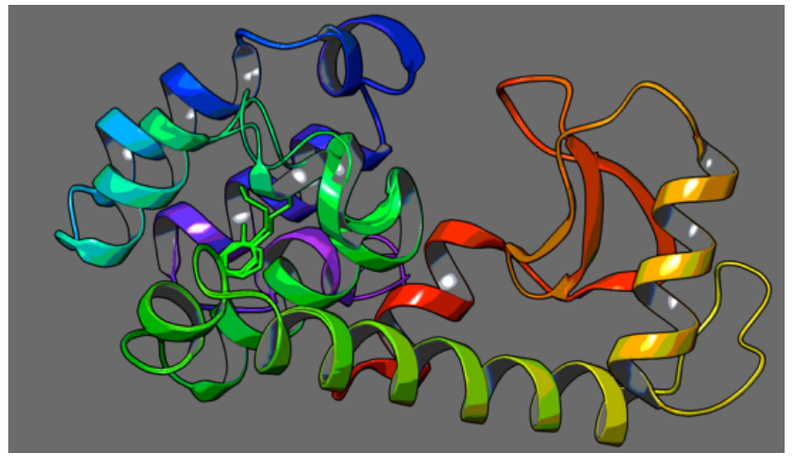


Ligand binding mode sampling is a general problem

What did we learn?

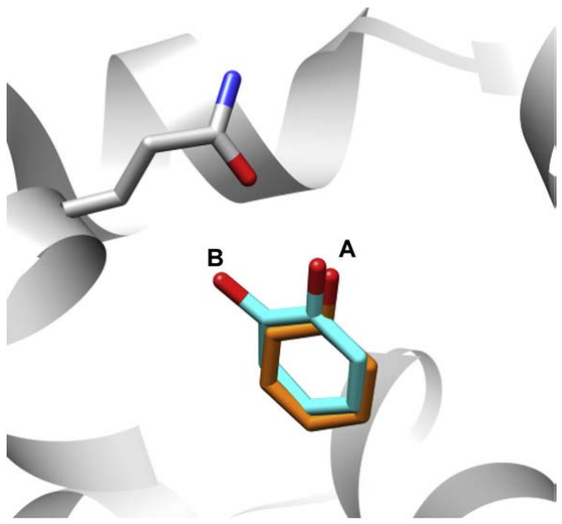


Ligand binding mode sampling is a general problem

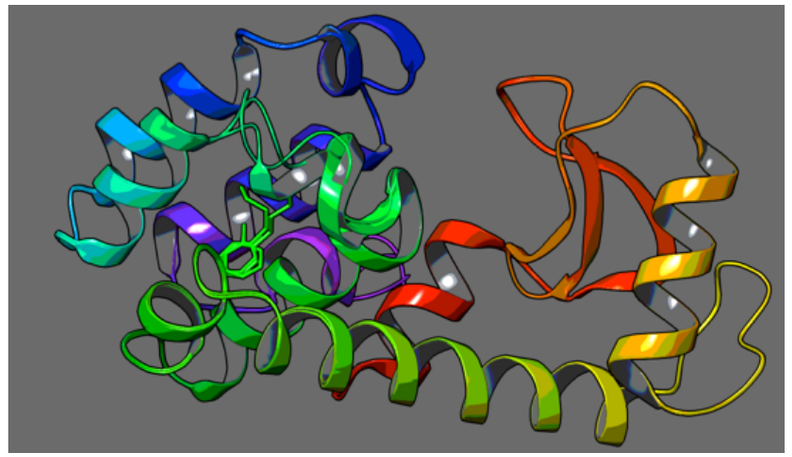


Sampling remains a major challenge and lysozyme provides a model system

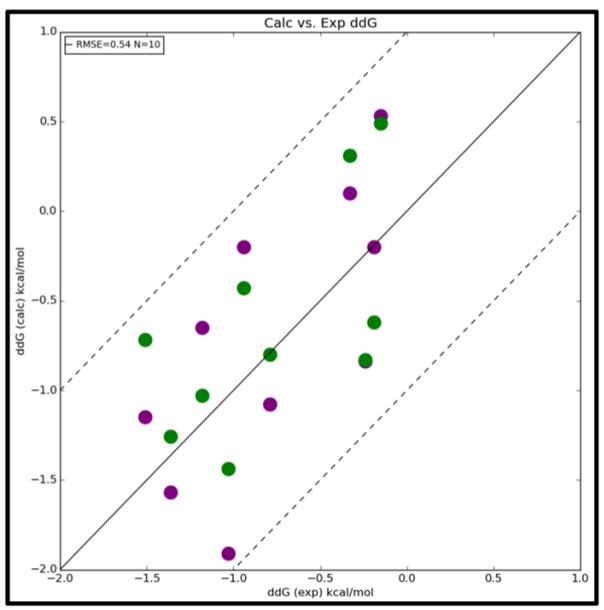
What did we learn?



Ligand binding mode sampling is a general problem

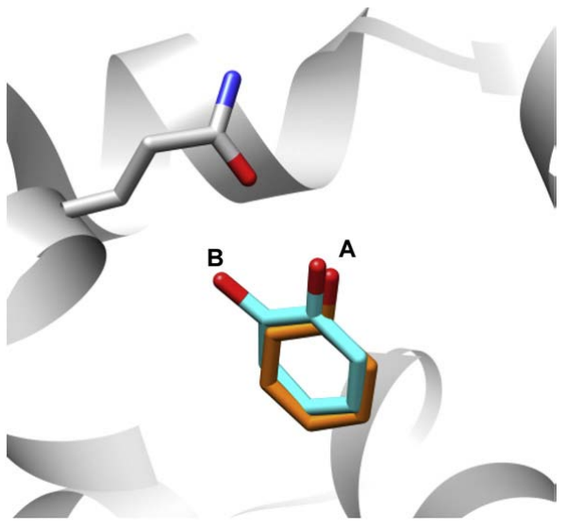


Sampling remains a major challenge and lysozyme provides a model system

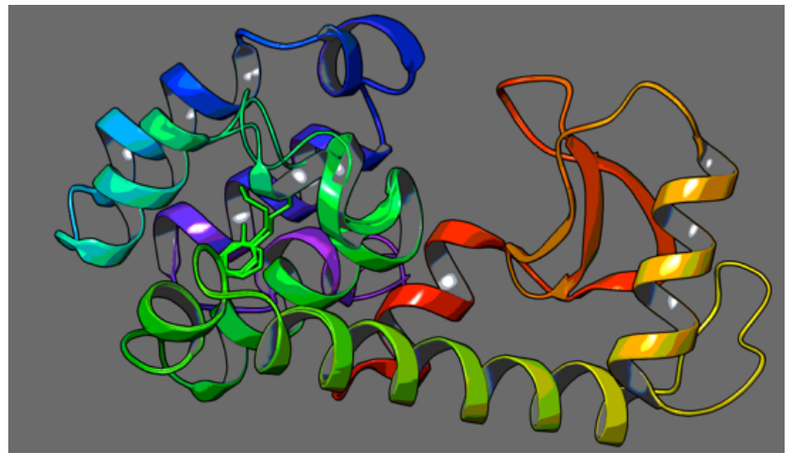


Results are reasonably good

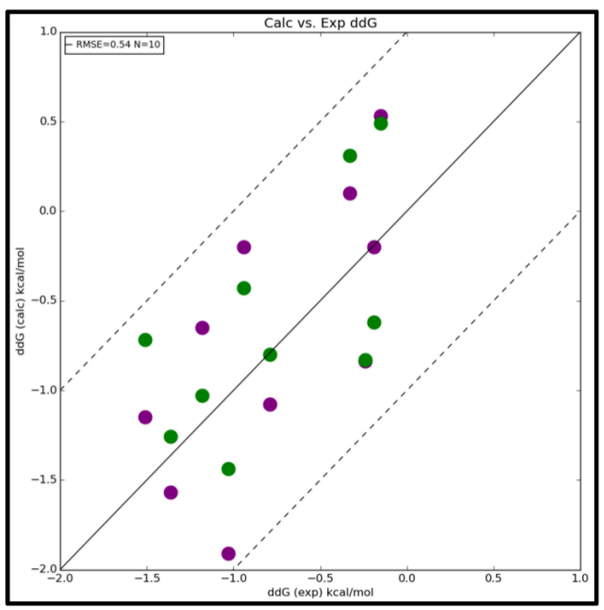
What did we learn?



Ligand binding mode sampling is a general problem

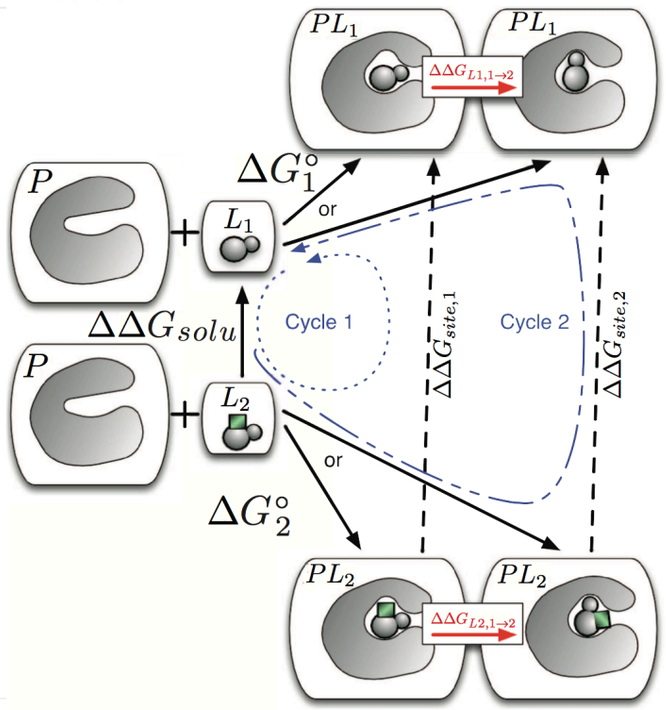


Sampling remains a major challenge and lysozyme provides a model system

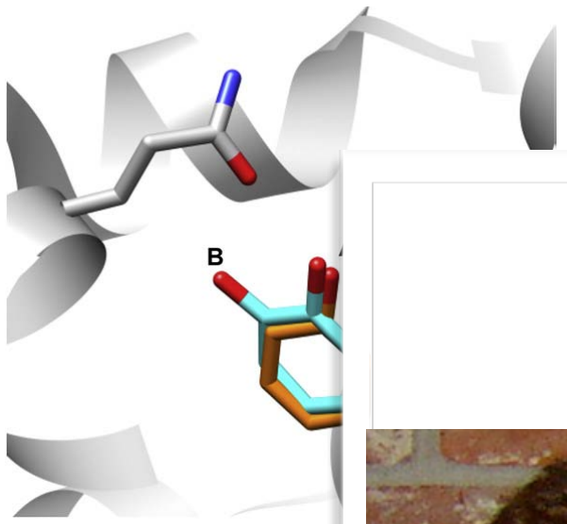


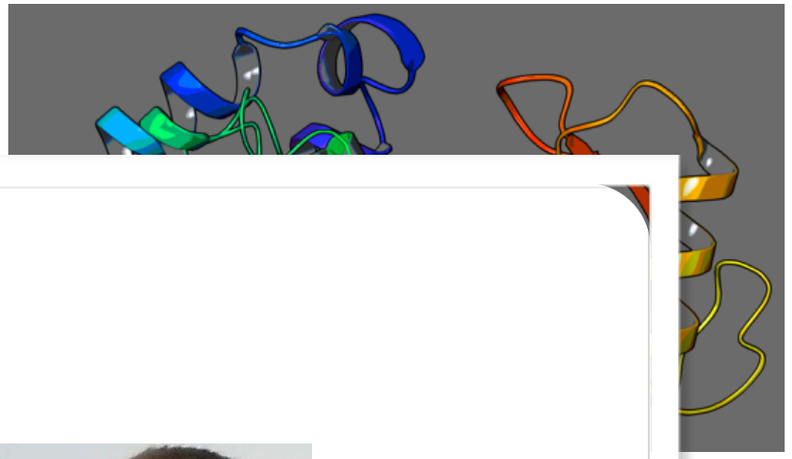
Results are reasonably good

Separation of states approaches may be an alternative



What did we learn?

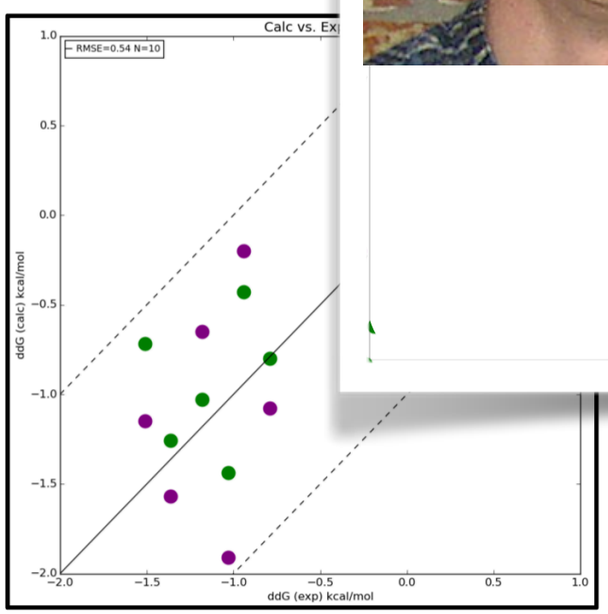




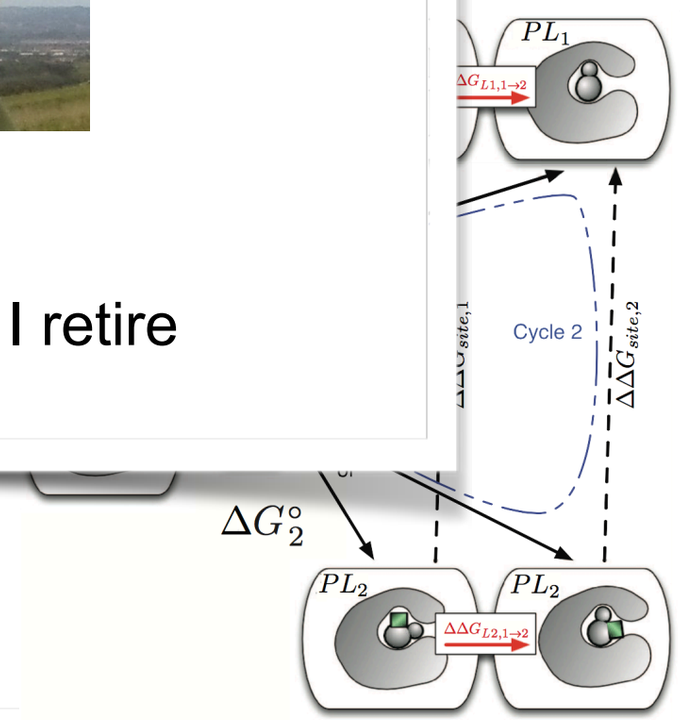
allenge
el system

???

I might still be studying lysozyme when I retire



alternative



What implications does this have for the “domain of applicability”? My view:

“Domain of applicability” means your model captures the relevant phenomena:

- Adequate description of forces
- Correct system (cofactors, counterions, construct, …)
- Sufficient sampling of relevant motions

What implications does this have for the “domain of applicability”? My view:

“Domain of applicability” means your model captures the relevant phenomena:

- Adequate description of forces
- Correct system (cofactors, counterions, construct, …)
- Sufficient sampling of relevant motions

Therefore, domain of applicability is ligand-dependent. Maybe this ligand:

- has unique FF problems
- requires new cofactor (or new waters, ions, …)
- induces new slow motions

Acknowledgments

My group:

Nathan Lim

Guilherme Matos

Camilla Zanette

Sam Gill

Hanh Nguyen

Caitlin Bannan

Gaetano Calabro

Kalli Burley

Vickie Lim

Kevin Hoang

Daisy Kyu

Collaborators:

J. Peter Guthrie

John Chodera

Michael Shirts

Andrew Paluch

Gabriel Rocklin

Ken Dill

Brian Shoichet, Sarah Boyce, Alan Graves (UCSF)

Chris Fennell

Chris Summa, J. Redmann (UNO)

Schrödinger, including **Teng Lin, Lingle Wang, Robert Abel**

OpenEye, especially Ant Nicholls and Christopher Bayly

Julien Michel, Hannes Loeffler

Dave Cerutti

Former group members:

Shuai Liu

Pavel Klimovich

Karisa Wymer

Chris Savoie

Anasuya Kolavennu

Kim Keyes

Vivian Jaber

Sreeja Parameswaran

Stefania Evoli

Kevin Hoang

Shannon Cao