

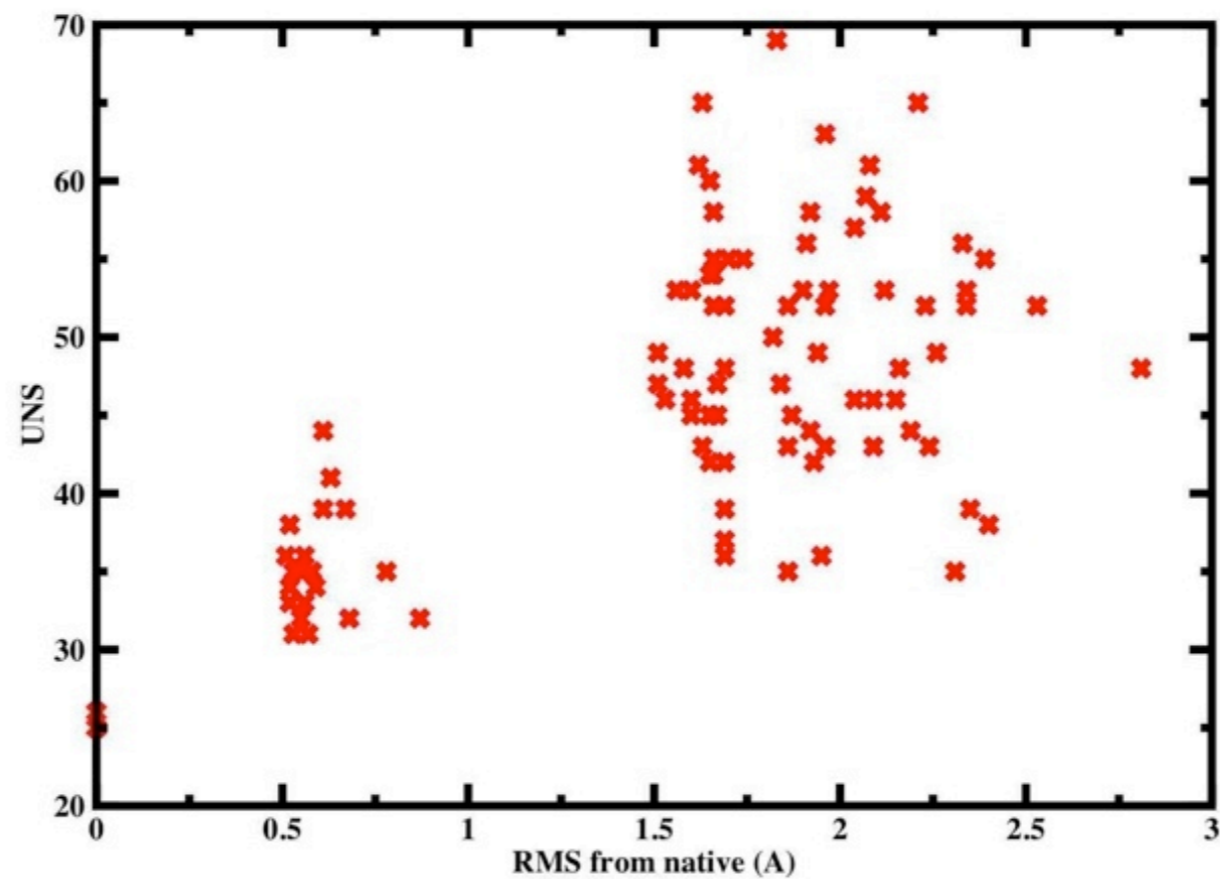
# The SHO model for implicit solvation

R. Whitney Smith  
John Karanicolas



## Phil's methods

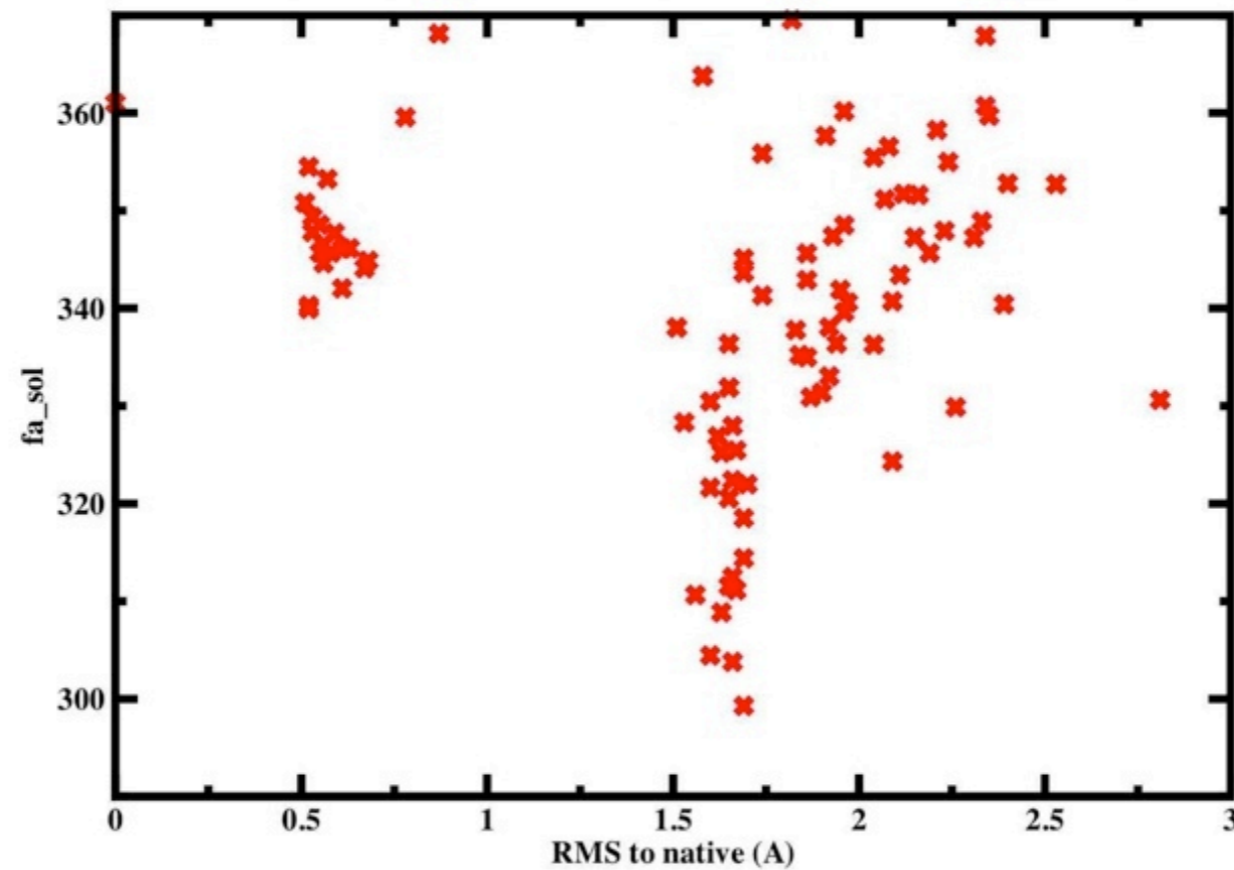
- Identification of UNS in monomeric proteins:



*RosettaCon 2005*

## Not recapitulation of fa\_sol

- Hopefully a forcefield term down the road...

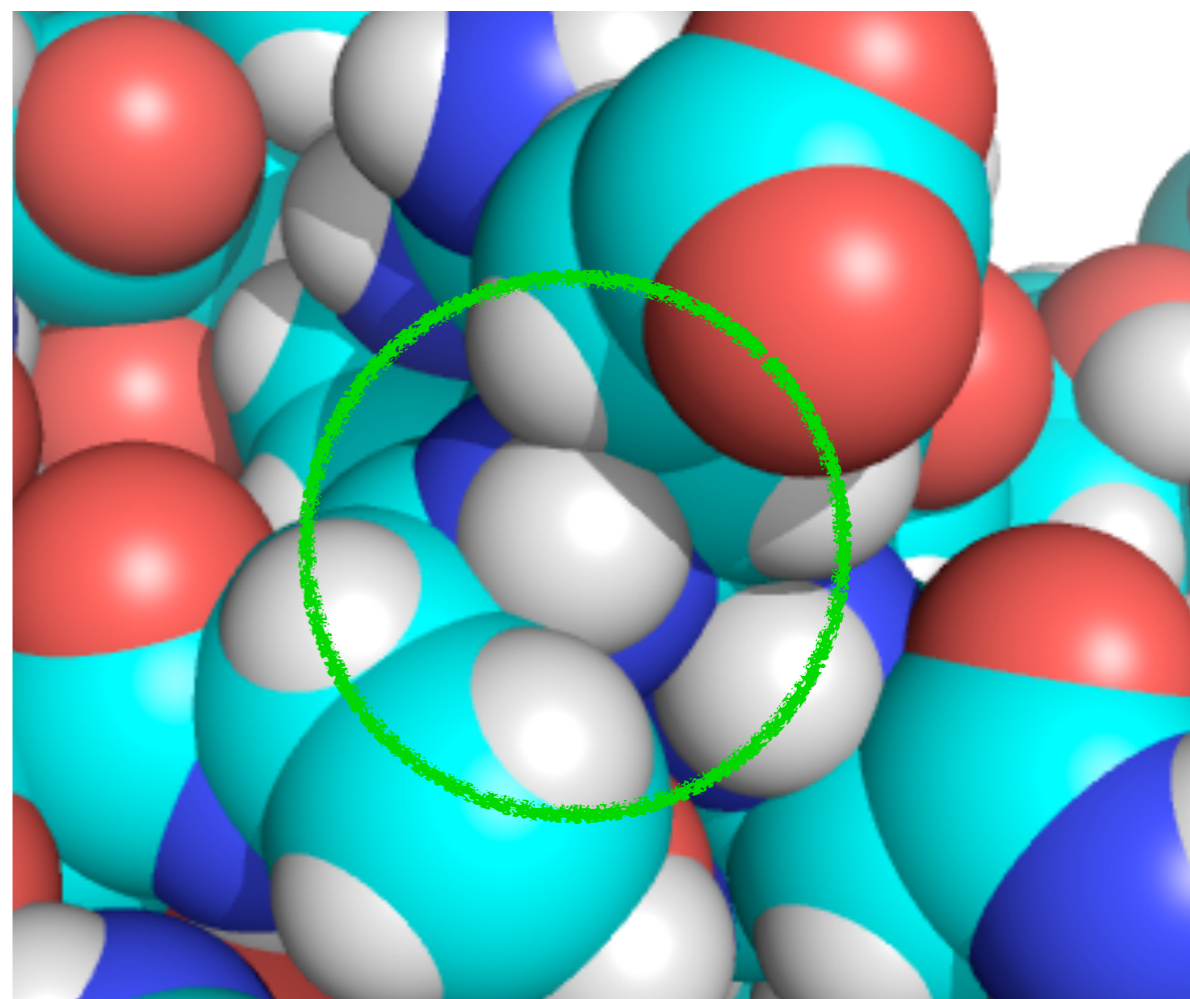
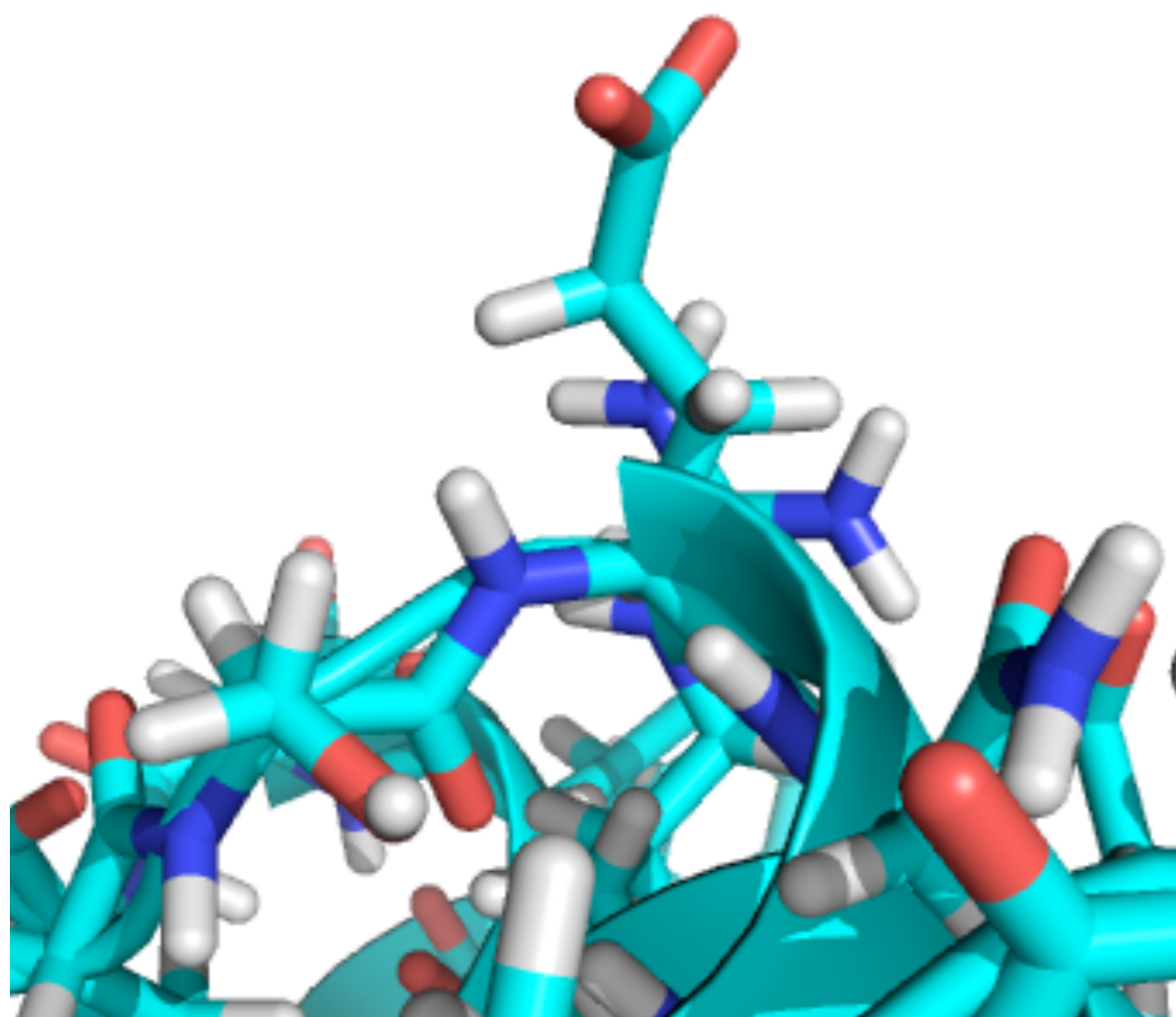


*RosettaCon 2005*

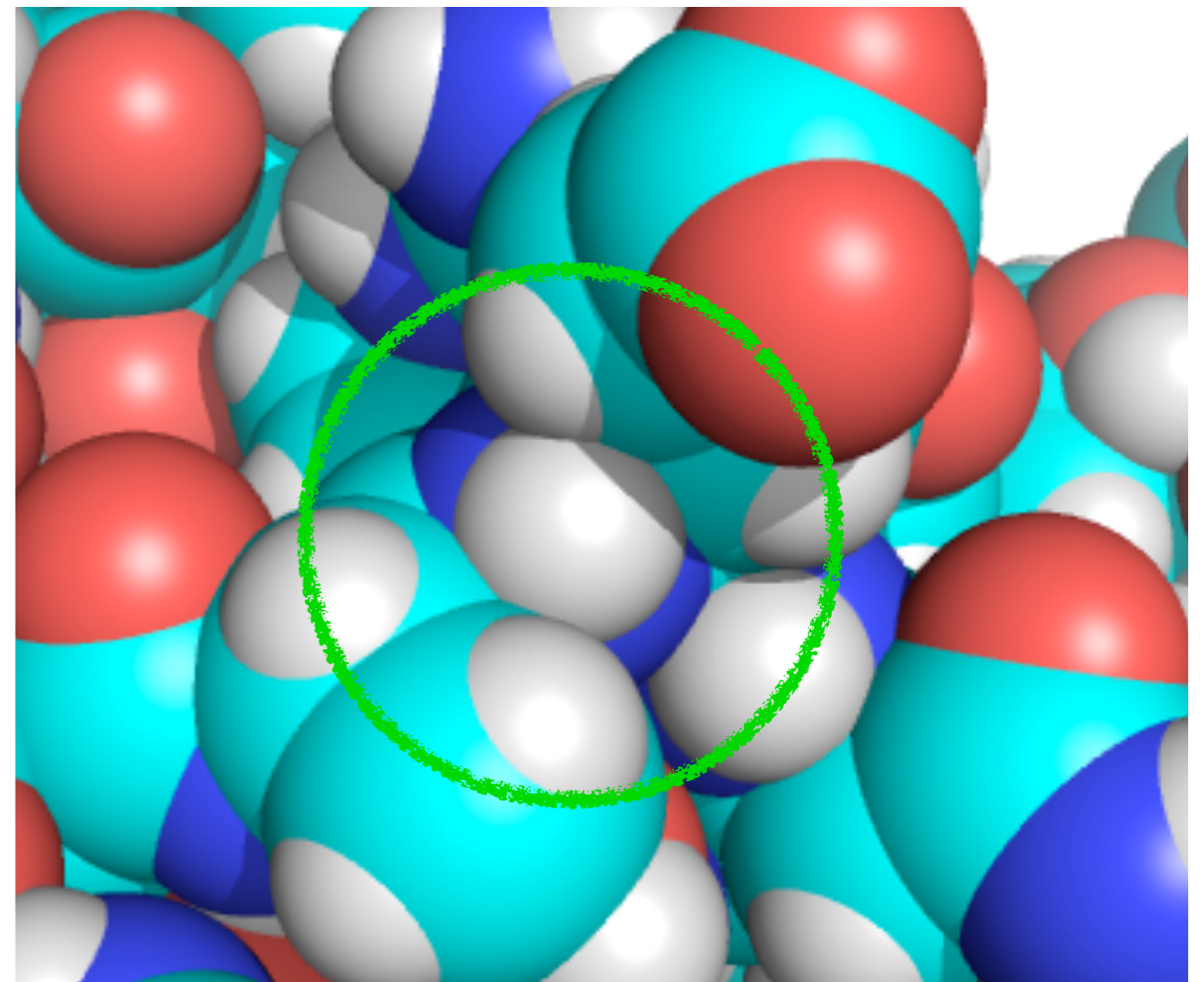
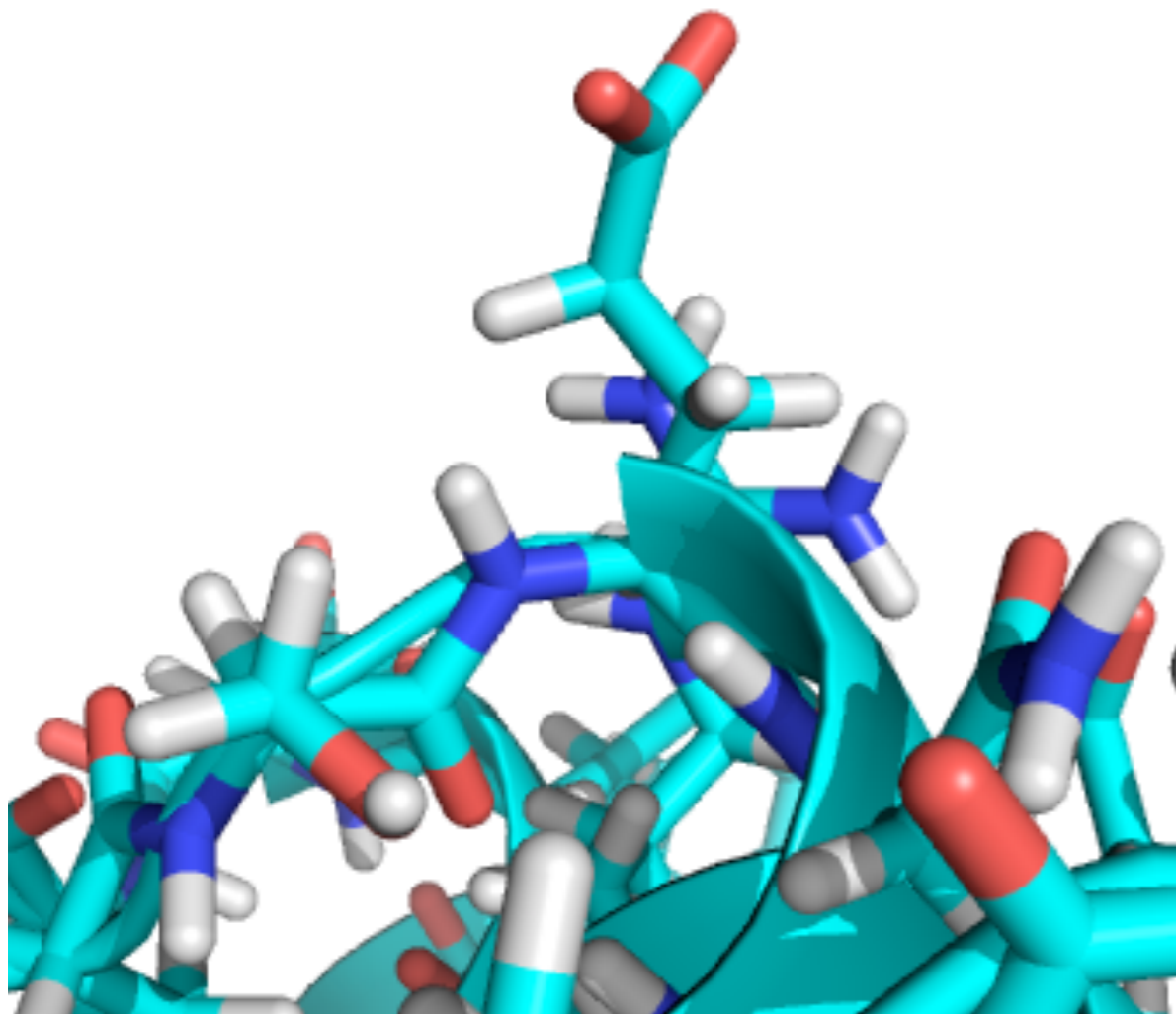
# Frequency of UNS

- Cost of an “UNS” backbone polar is 5-6 kcal/mol (Fleming and Rose, 2005)
- The few observed in crystal structures are artifacts (Fleming and Rose, 2005)
- Why did we identify ~25 in the native protein??

# Unintuitive “UNS” assignment

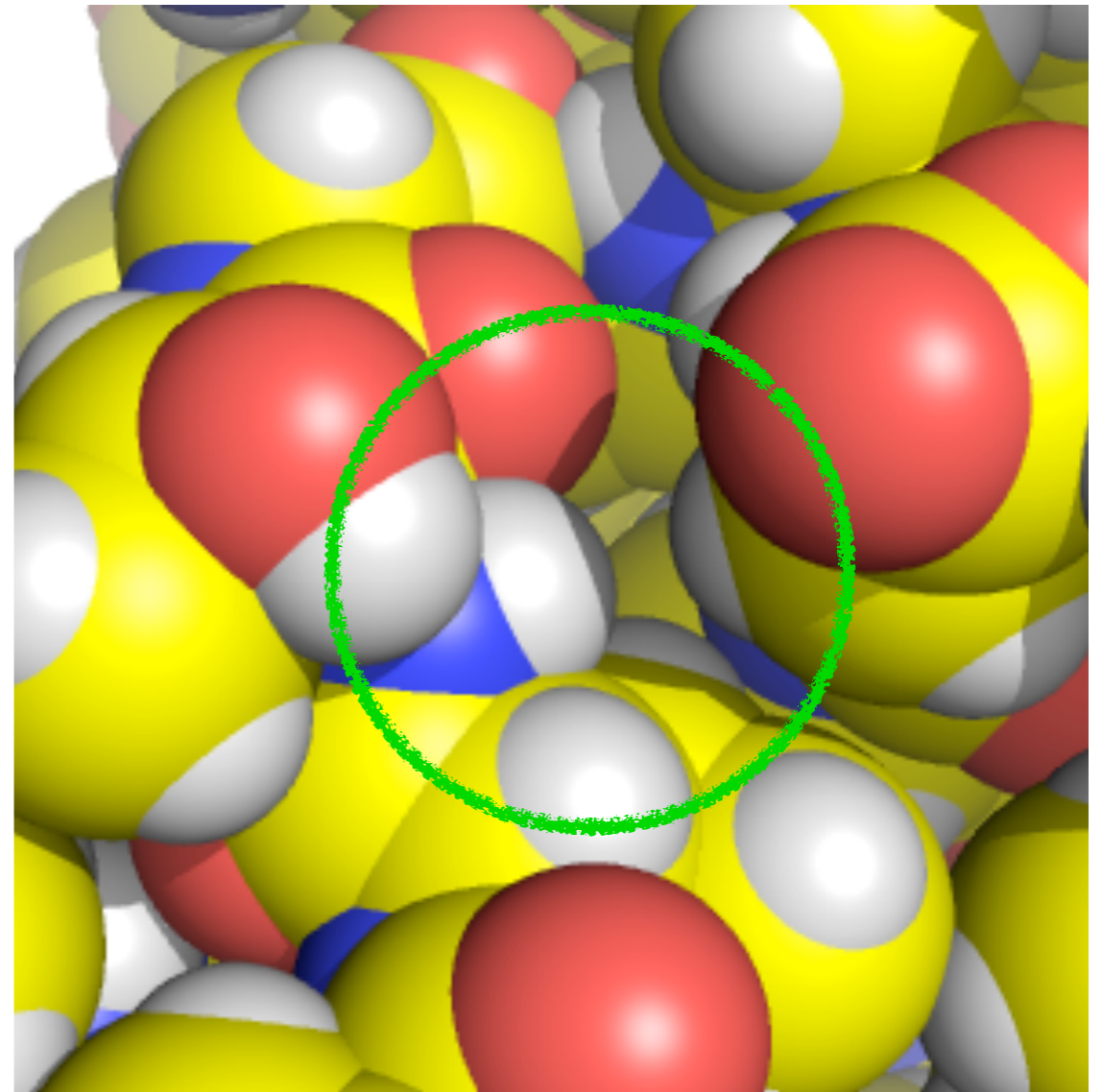
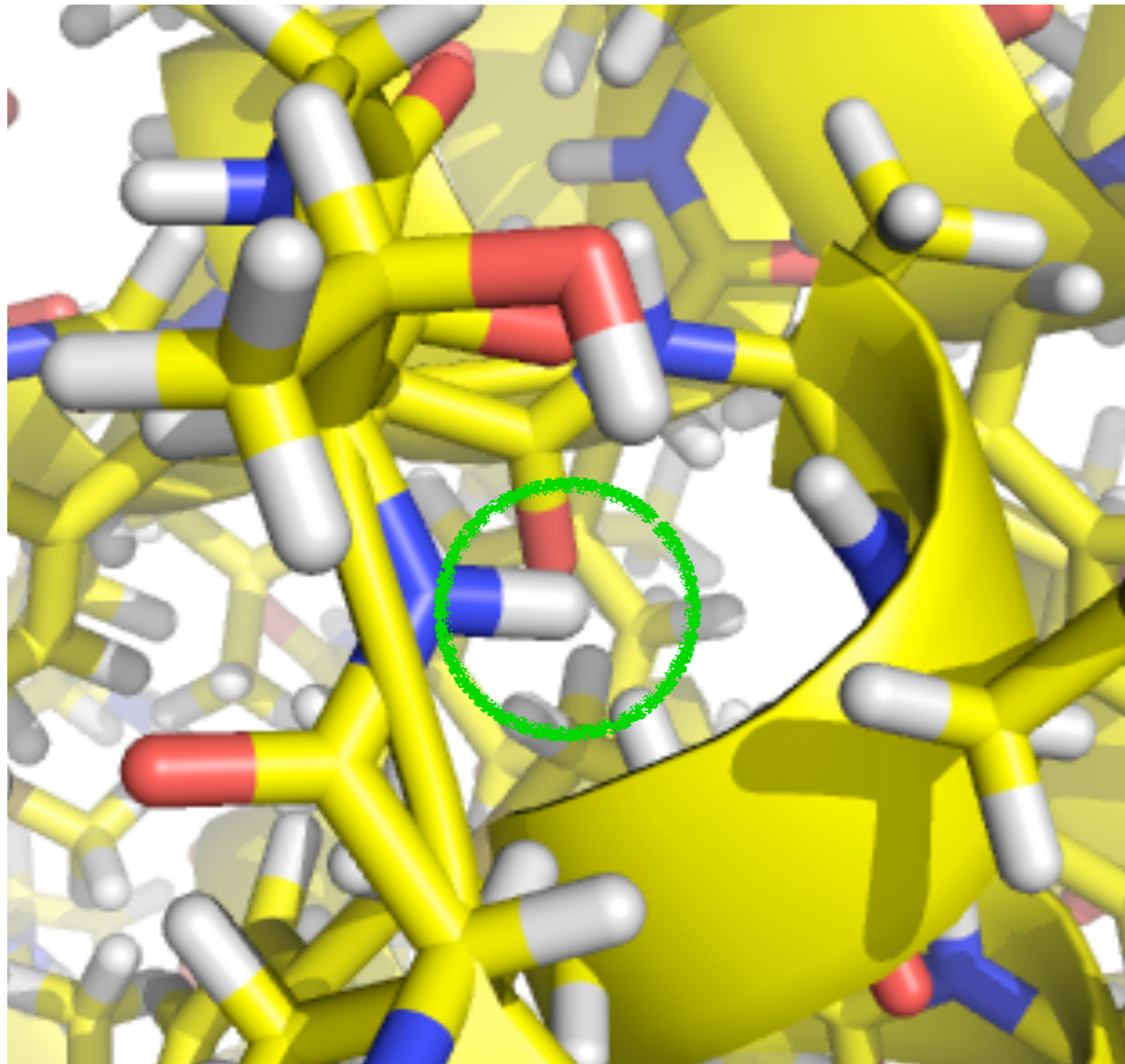


# Unintuitive “UNS” assignment

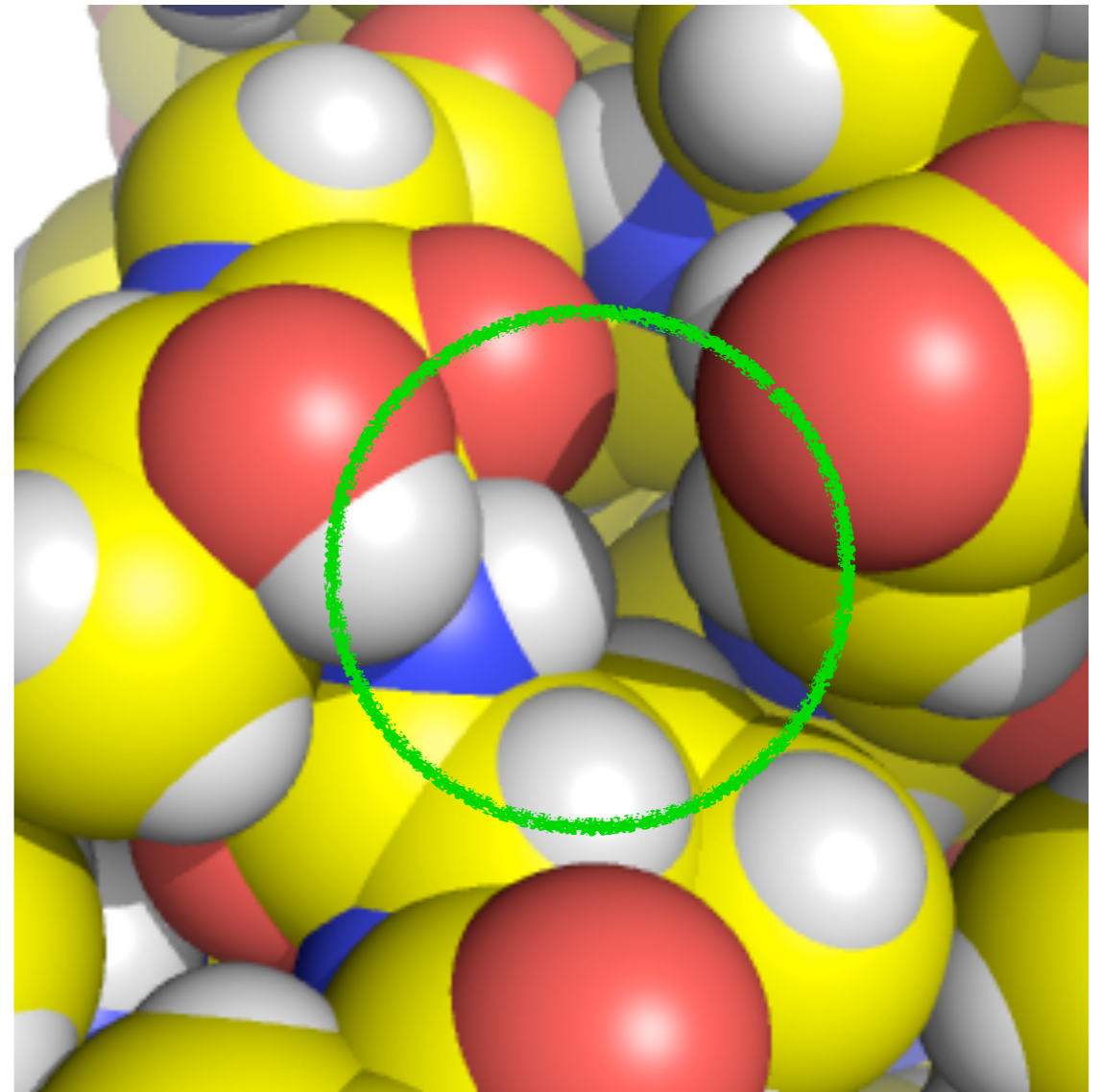
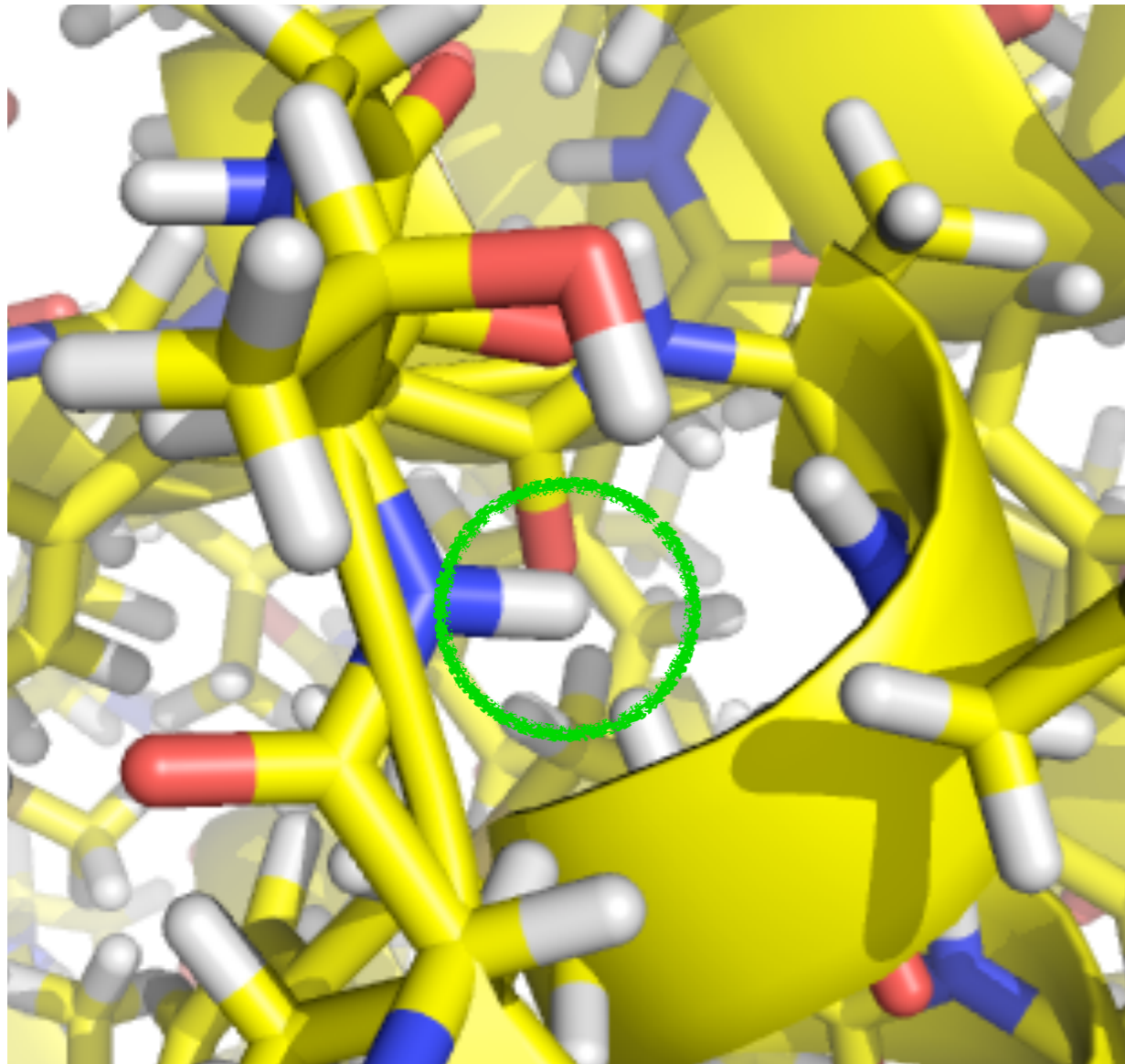


*Solution: smaller probe size?*

# Unintuitive “exposed” assignment



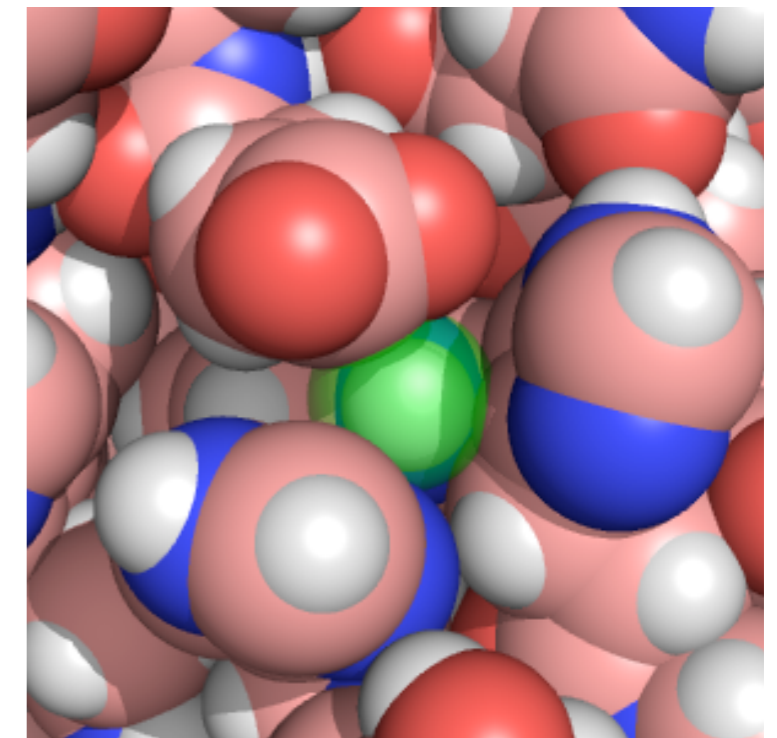
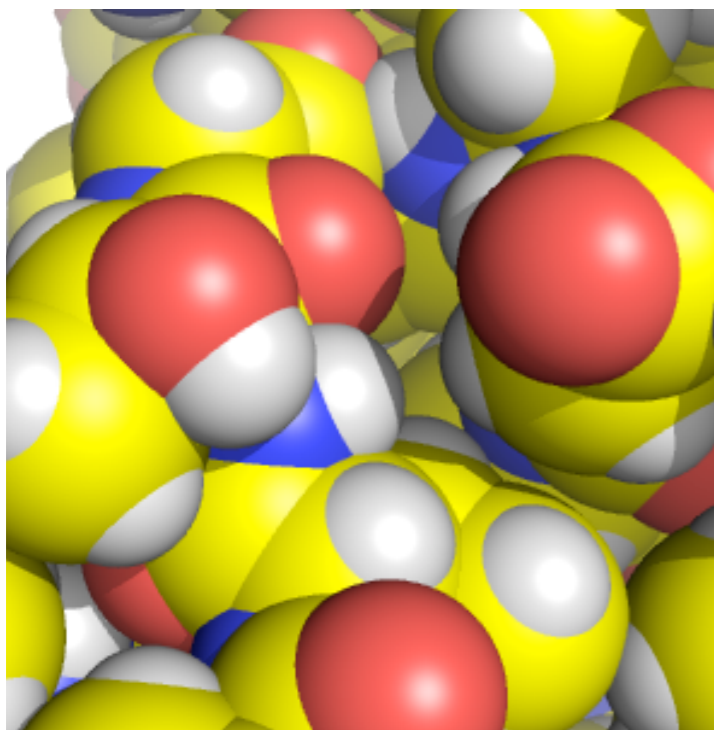
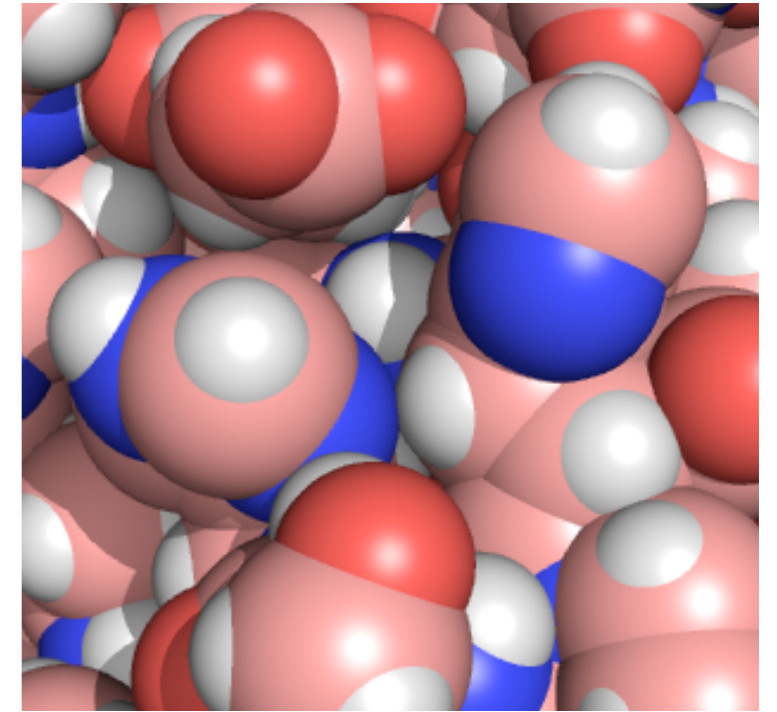
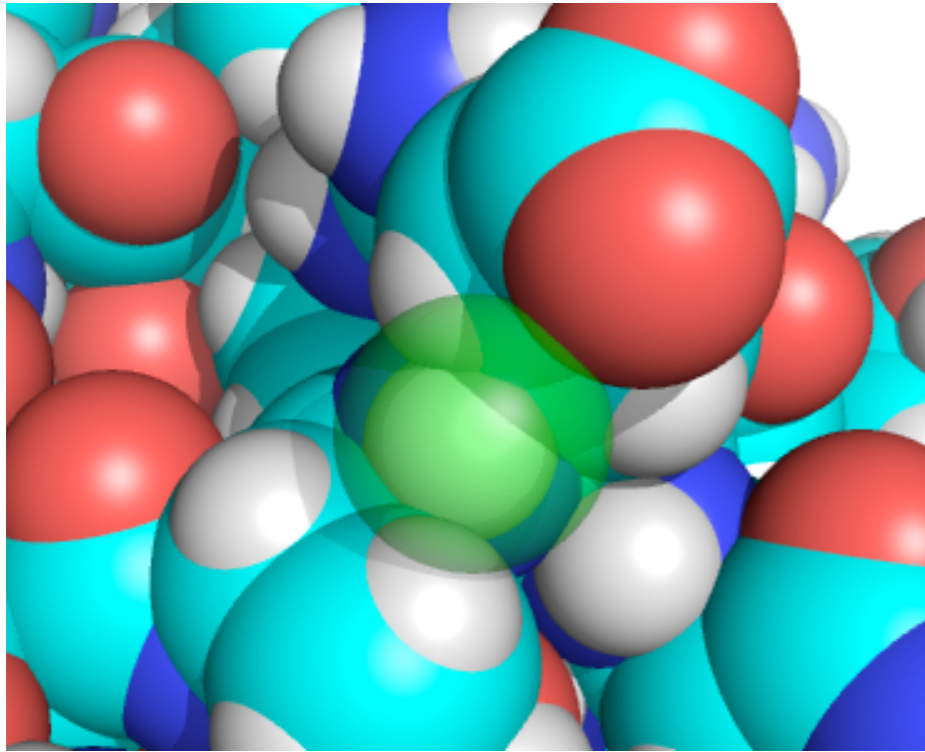
# Unintuitive “exposed” assignment



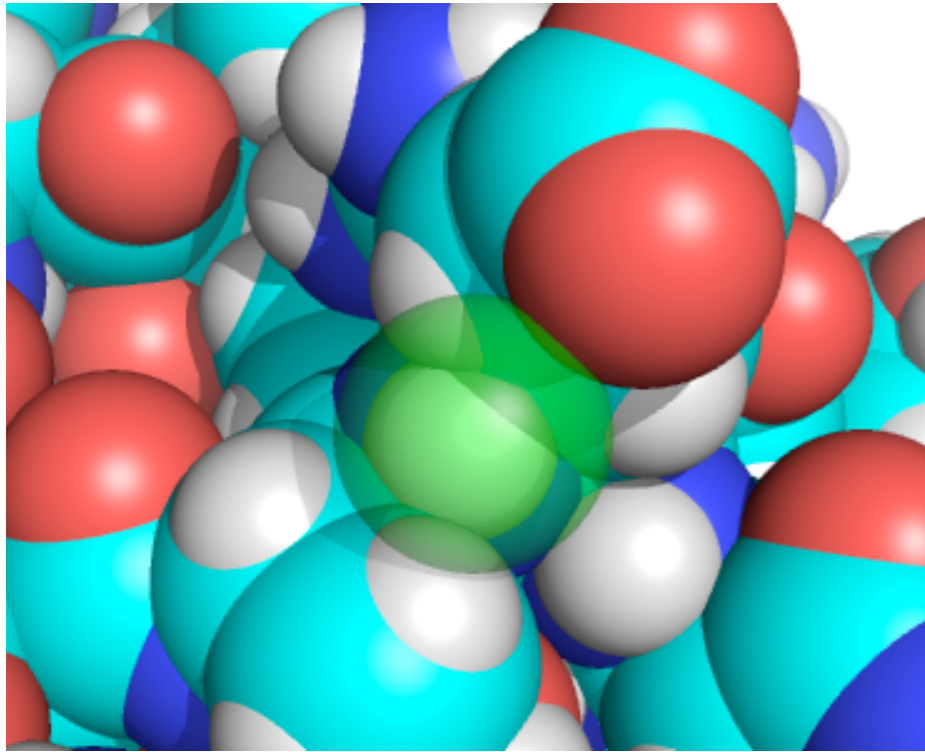
*Solution: try explicitly building (“rotameric”) waters*



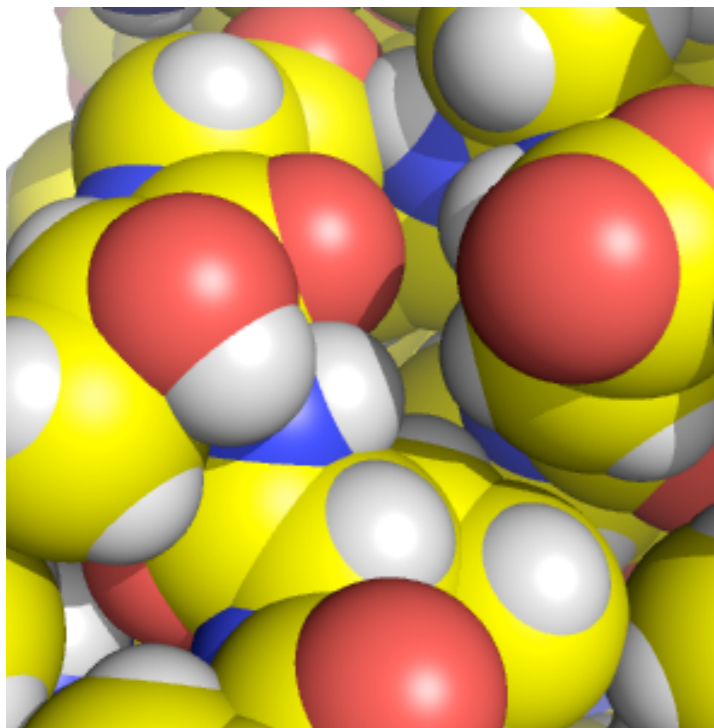
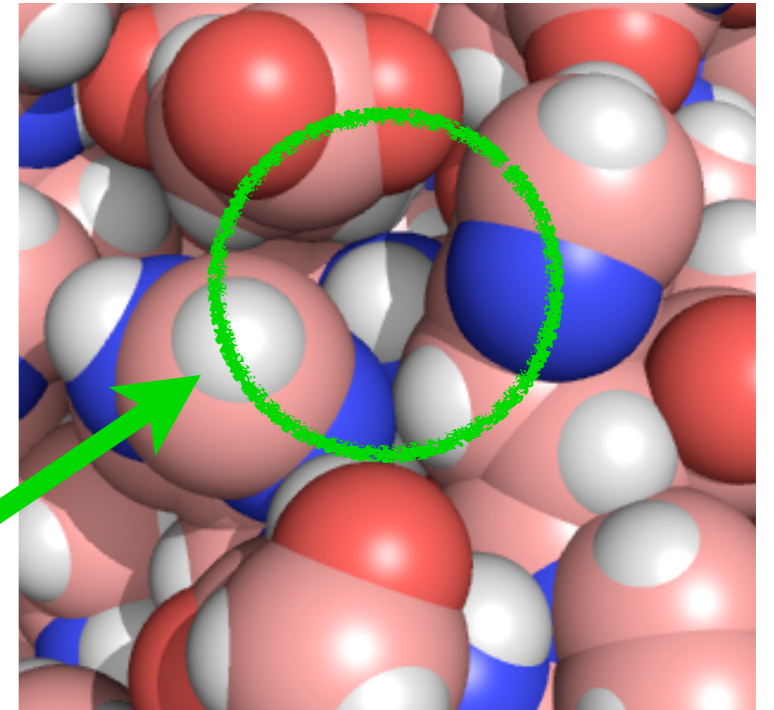
# Results match intuition



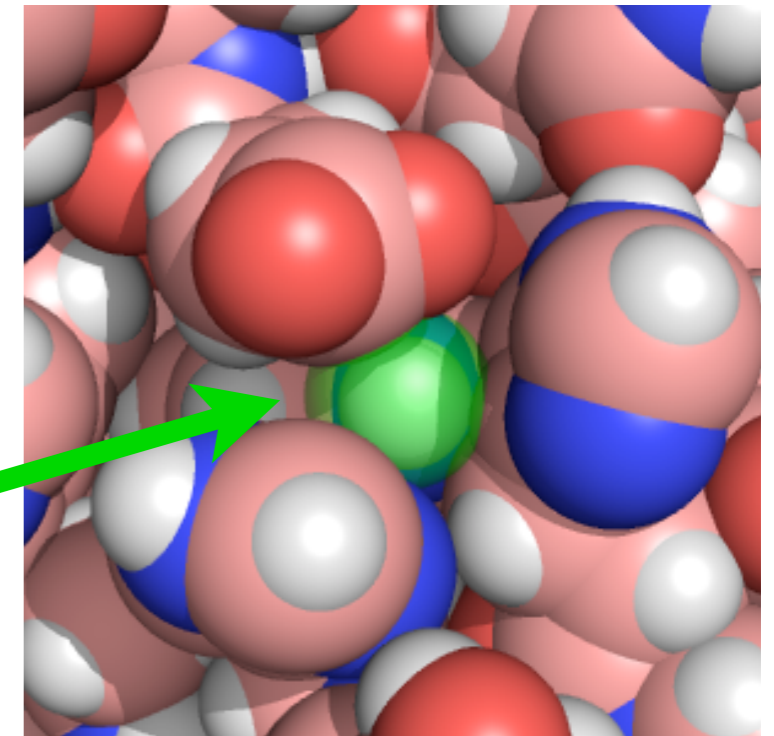
# Results match intuition



*Not  
accessible to  
1.0 Å probe*



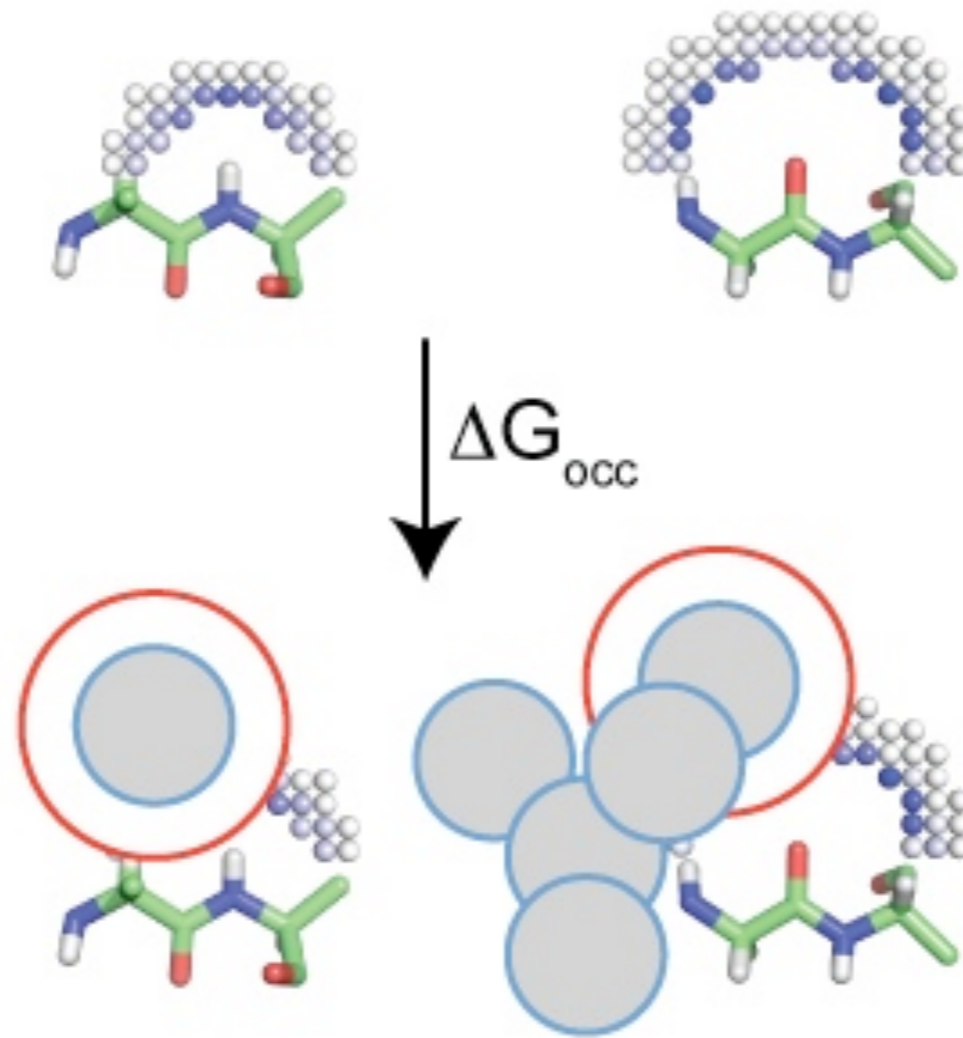
*But  
potentially  
solvated!*



# Shortcomings

- Description of partially buried polar groups not robust
- “Rotamer approximation” - what about non-ideal water positions?
- Prefer non-binary for incorporation into the energy function (in place of the polar part of EEF1)

# Moving off-rotamer



**“SHO”** : *solvent hydrogen-bond occlusion*

# The SHO approach

$$P(\text{grid point } j \text{ is occupied by solvent}) = \frac{\exp\left(-\frac{E_{hb}^j}{kT}\right)}{\sum_i^{\text{whole grid}} \exp\left(-\frac{E_{hb}^i}{kT}\right) + \exp\left(-\frac{E_{bulk}}{kT}\right)}$$

# The SHO approach

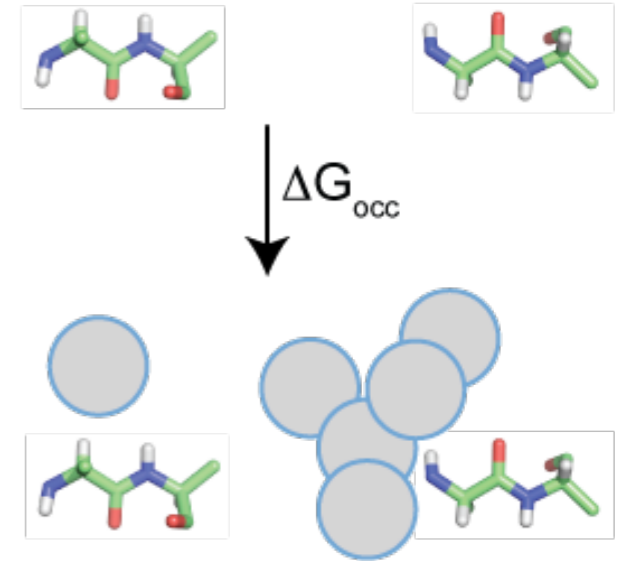
$$\begin{aligned} E_{SHO} &= -kT \ln[P(\text{no solvent at any occluded points})] \\ &= -kT \ln[1 - P(\text{solvent at one or more occluded points})] \end{aligned}$$

# The SHO approach

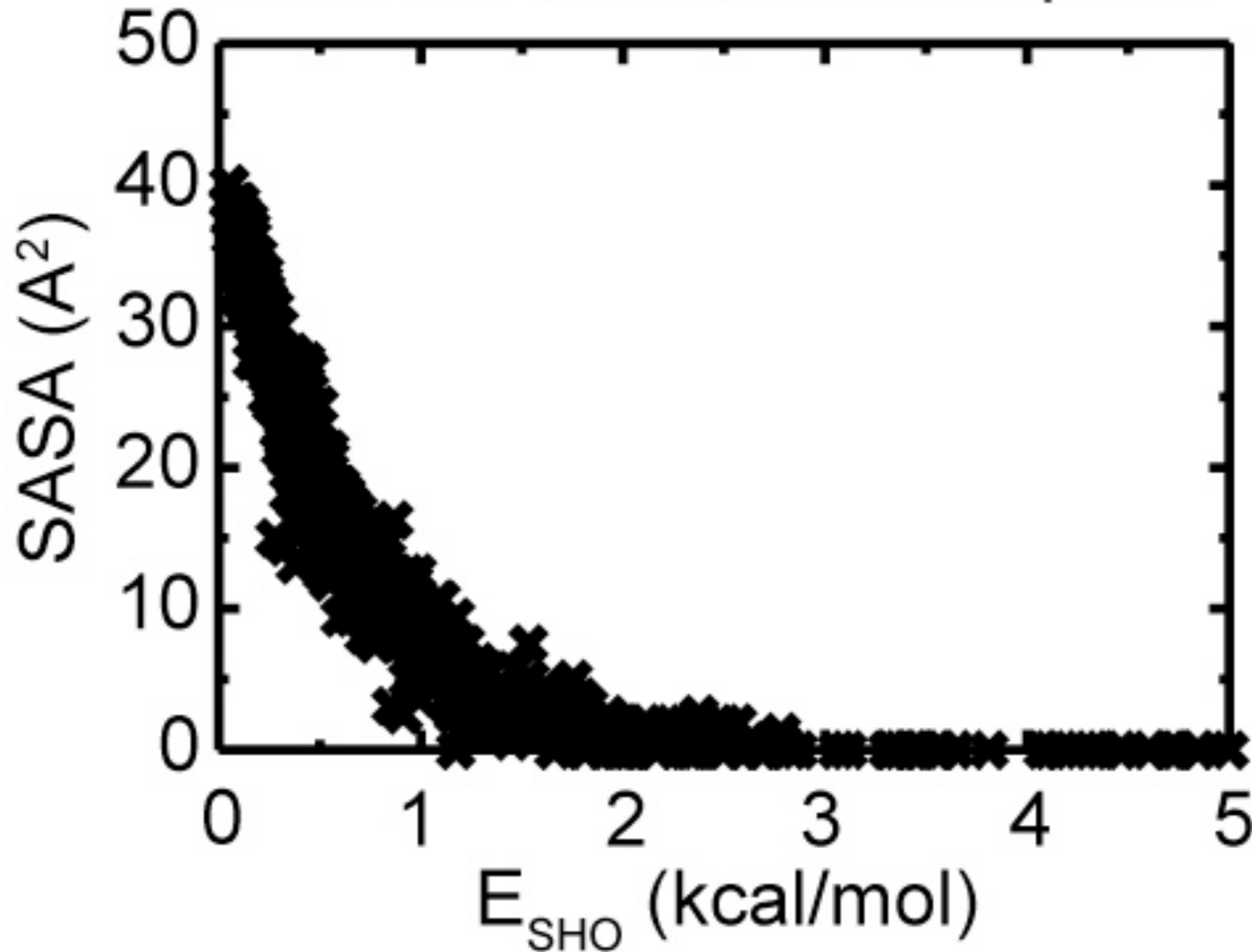
$$E_{SHO} = -kT \ln \left( 1 - \frac{\sum_j^{\text{occluded grid points}} \exp\left(-\frac{E_{hb}^j}{kT}\right)}{\sum_i^{\text{whole grid}} \exp\left(-\frac{E_{hb}^i}{kT}\right) + \exp\left(-\frac{E_{bulk}}{kT}\right)} \right)$$

Note:  $E_{SHO}$  ranges from 0 to 5 kcal/mol (sole adj. param)

# What are we capturing?

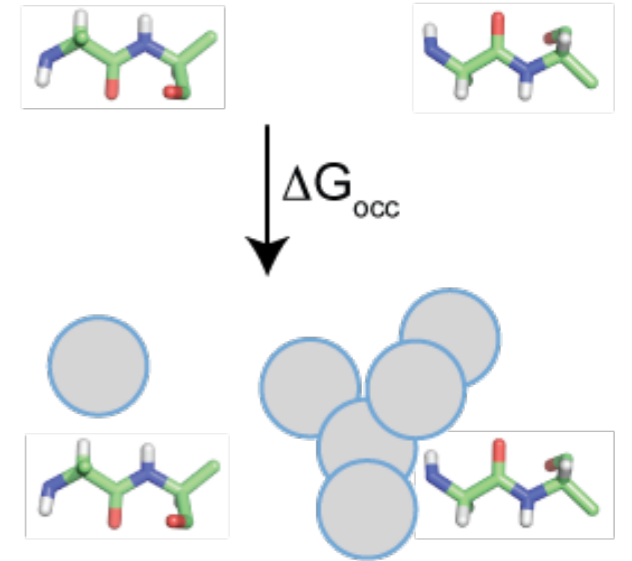


Asn/Gln sidechain acceptors

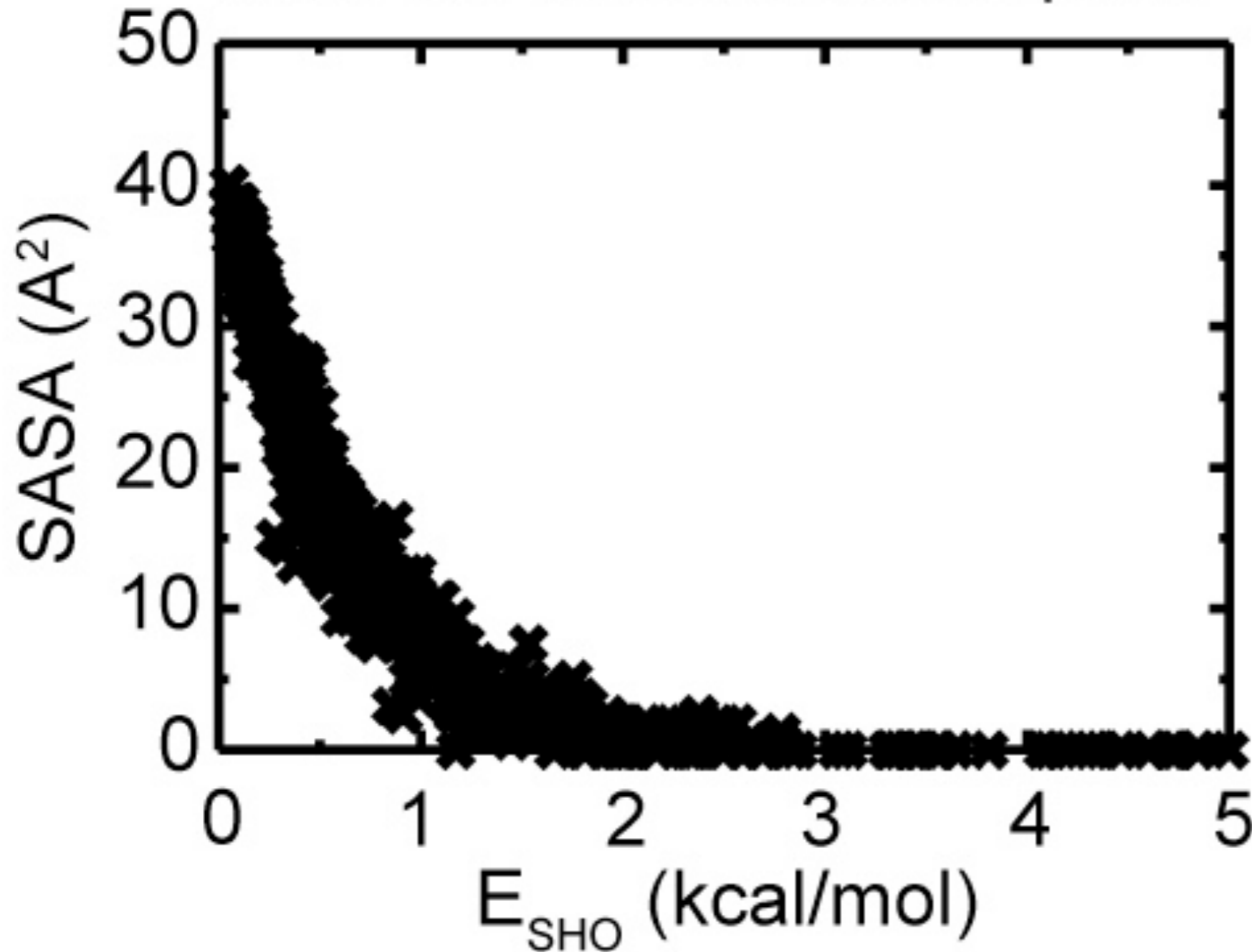




# What are we capturing?

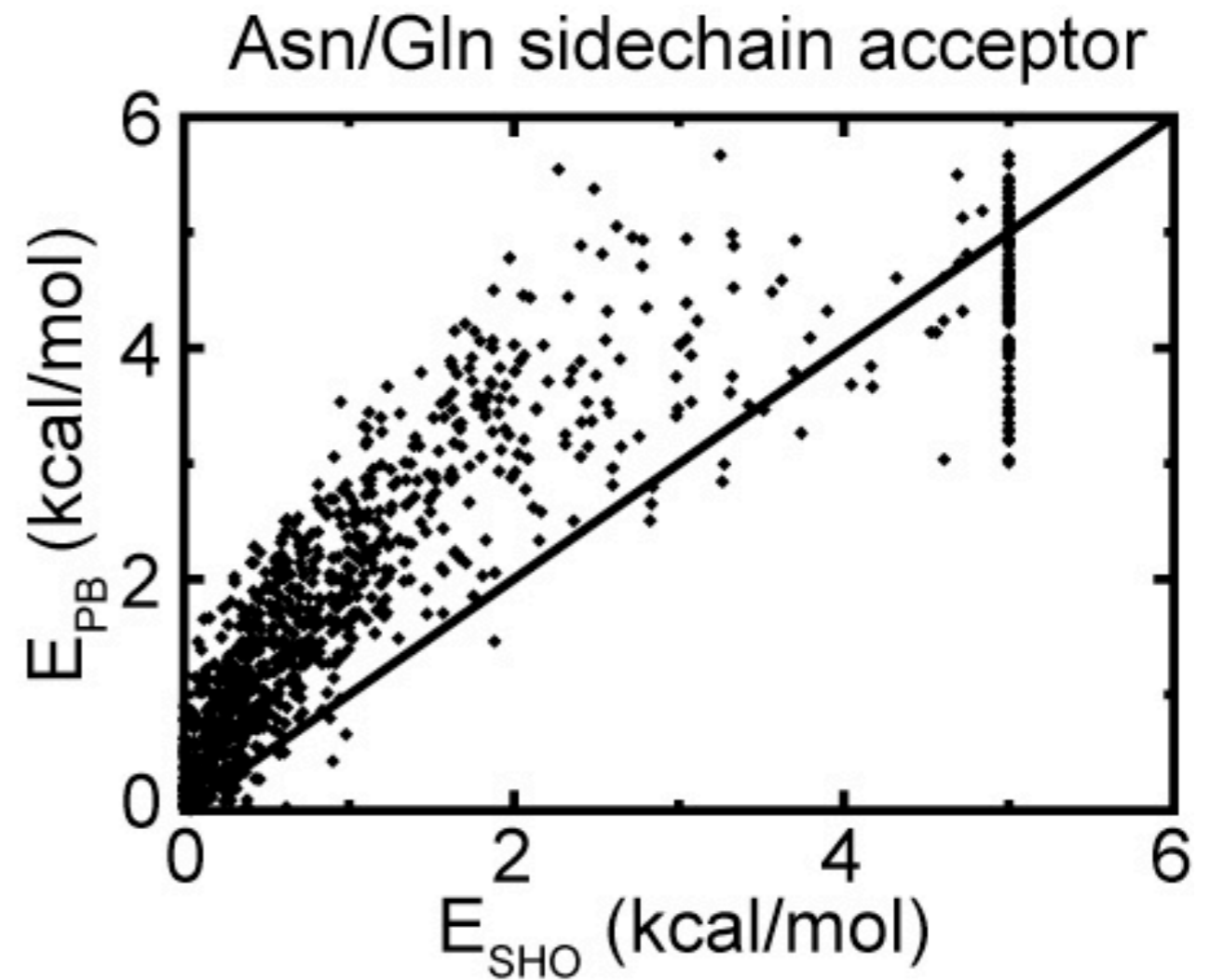
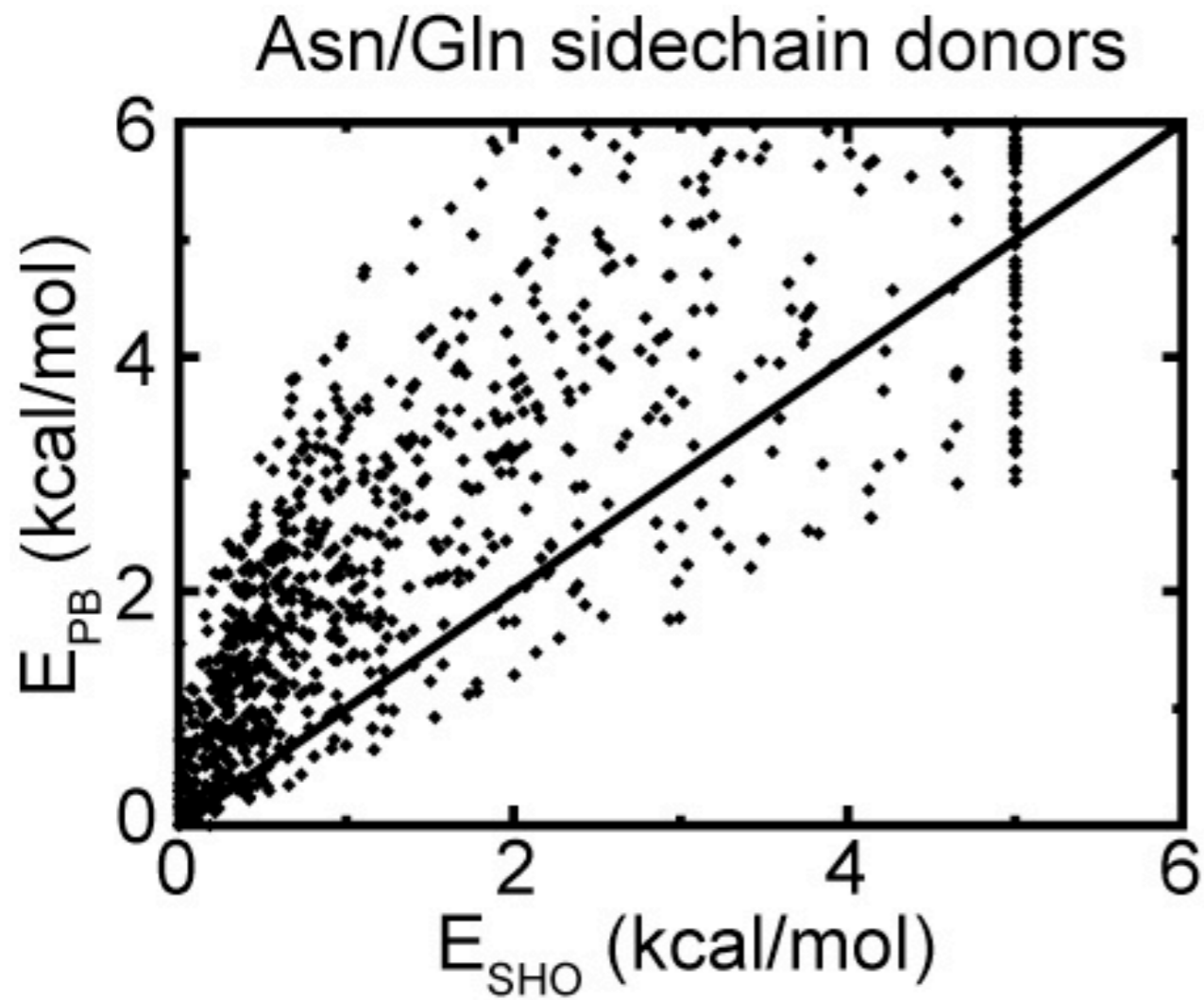
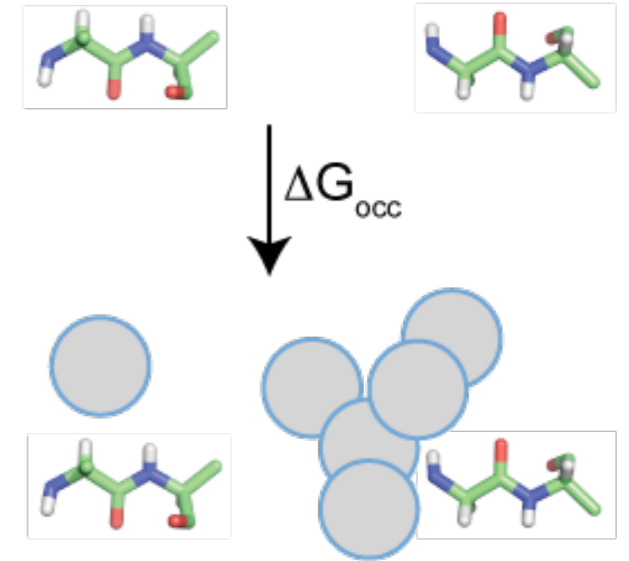


Asn/Gln sidechain acceptors

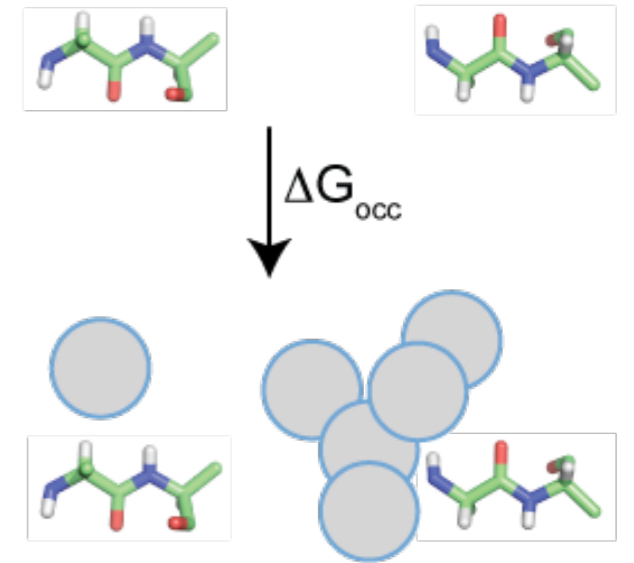


*Recall: many spurious UNS in native proteins (1.4  $\text{\AA}$  probe)*

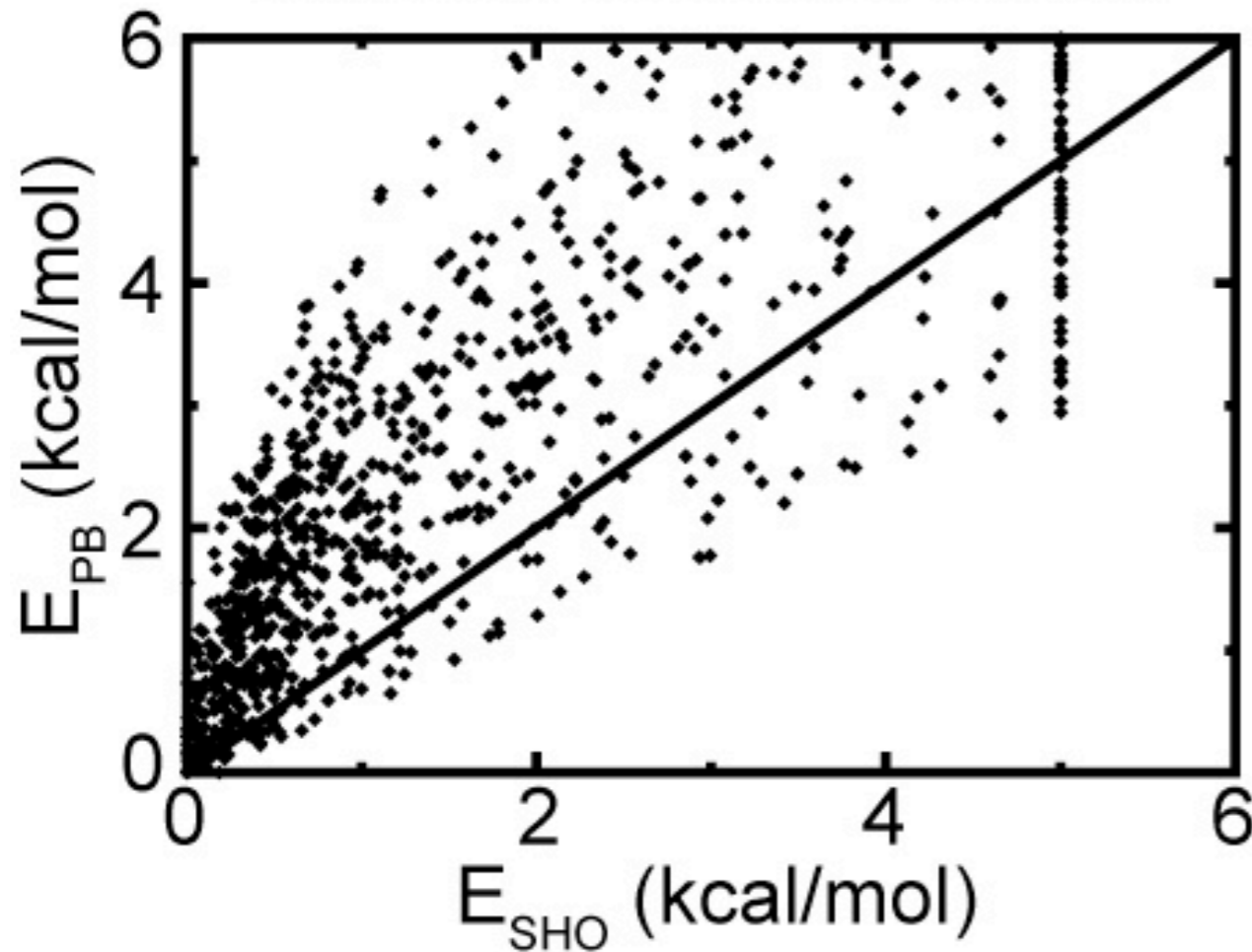
# What are we capturing?



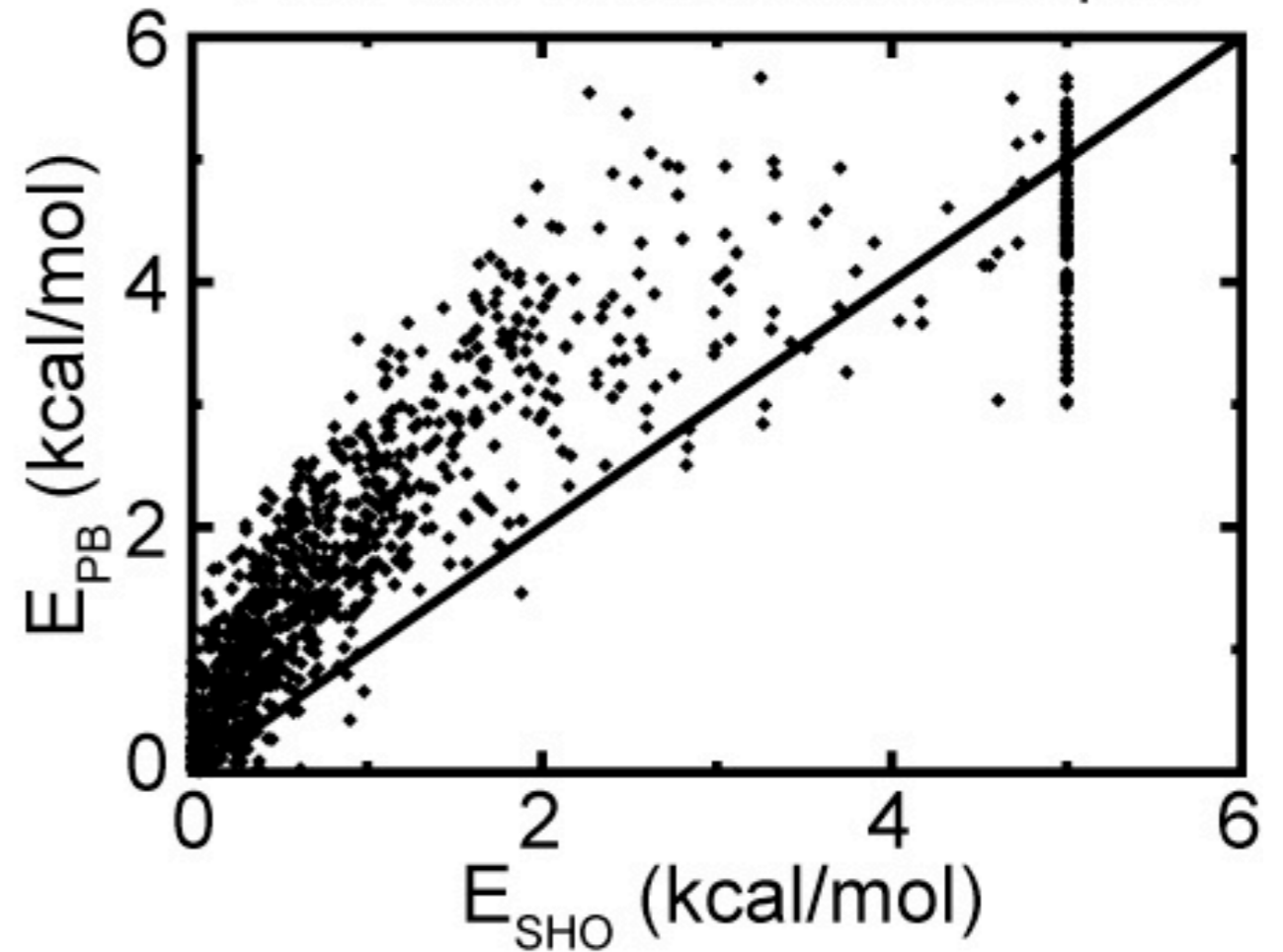
# What are we capturing?



Asn/Gln sidechain donors

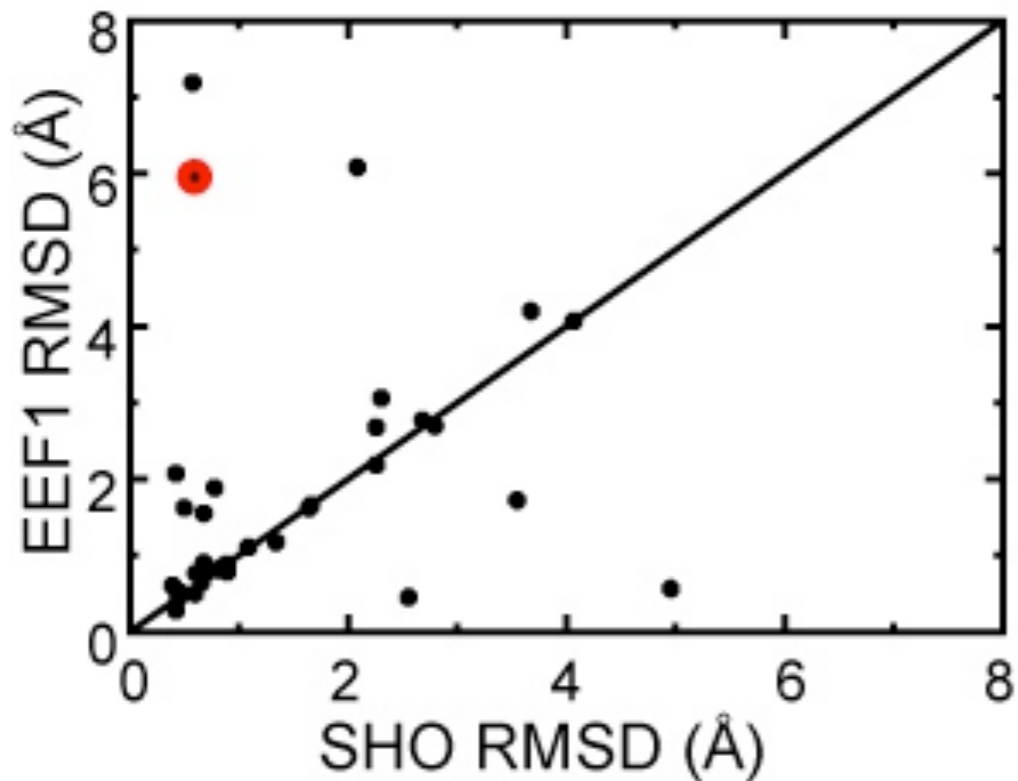


Asn/Gln sidechain acceptor

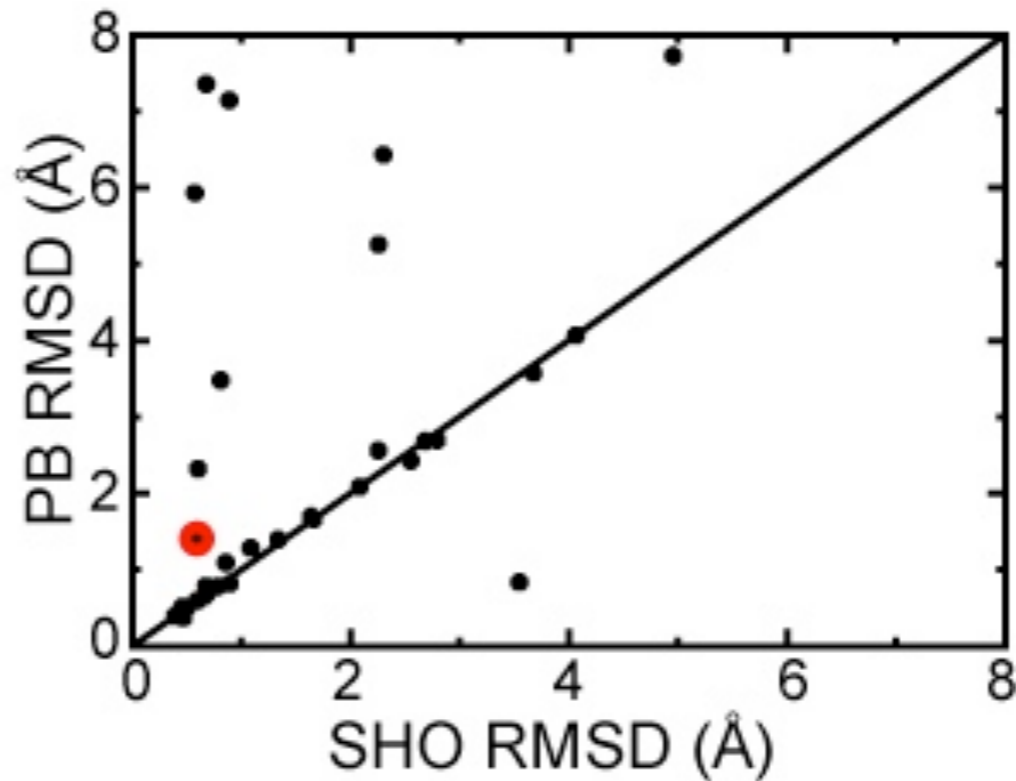


*Mean unsigned error is 0.9 (donors), 0.6 kcal/mol (acceptors)*

# KIC loop modeling

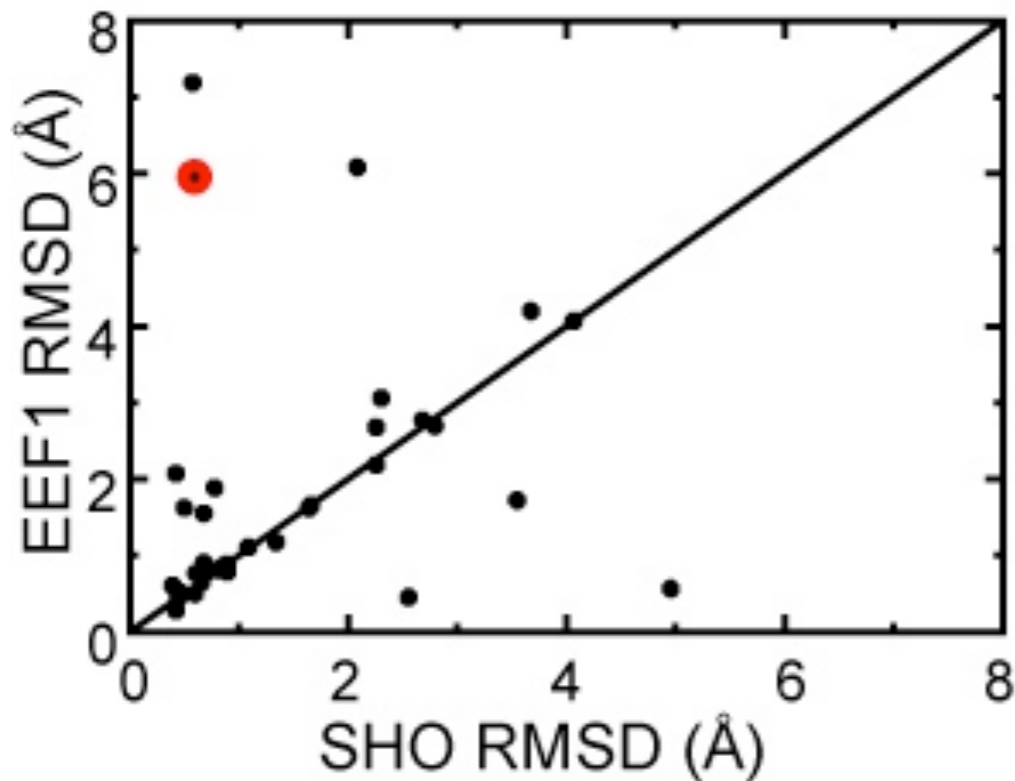


*SHO outperforms EEF1 for  
discrimination,  $p < 0.05$*

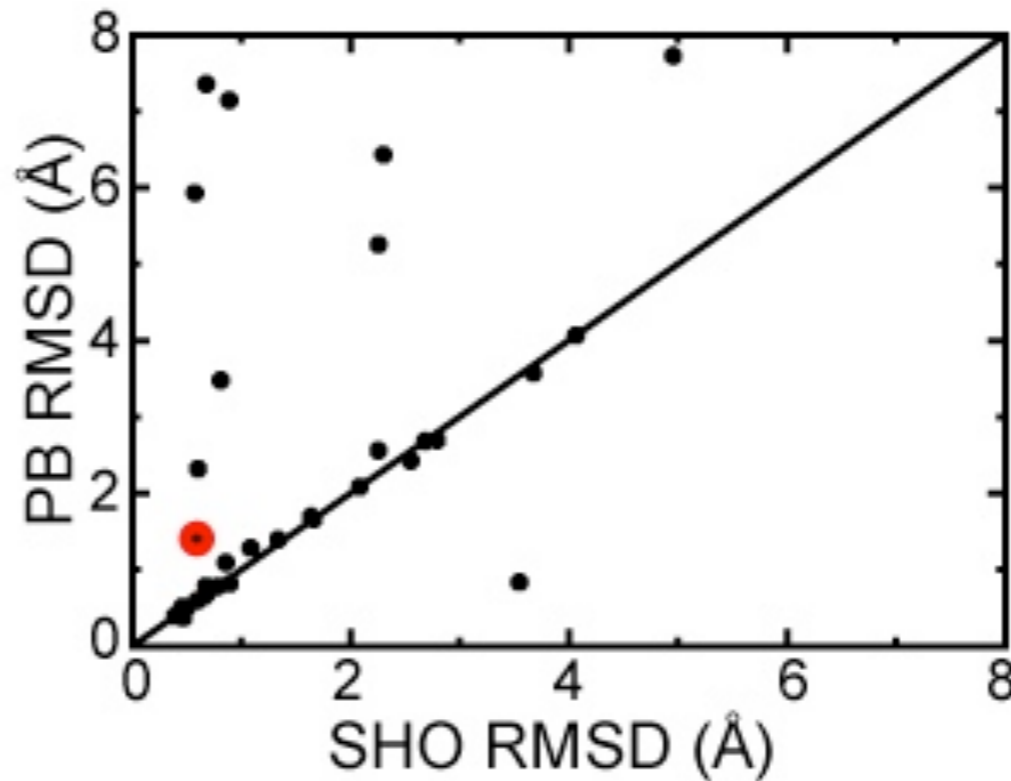


*SHO outperforms PB for  
discrimination,  $p < 0.02$*

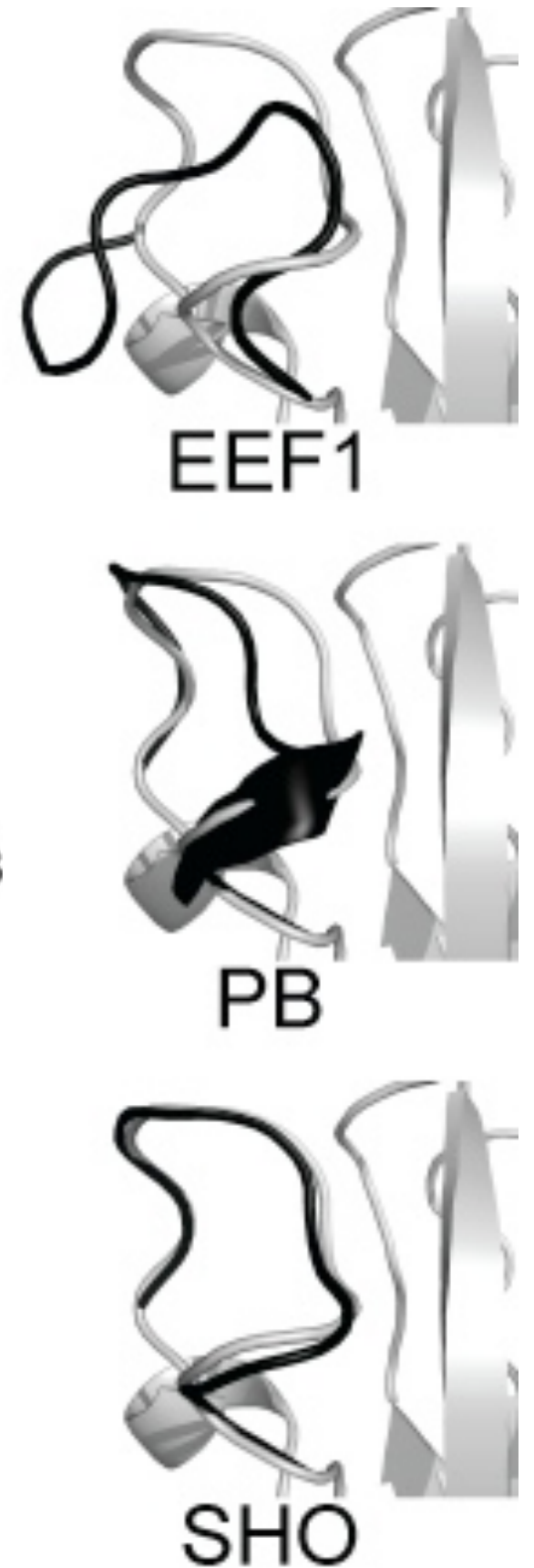
# KIC loop modeling



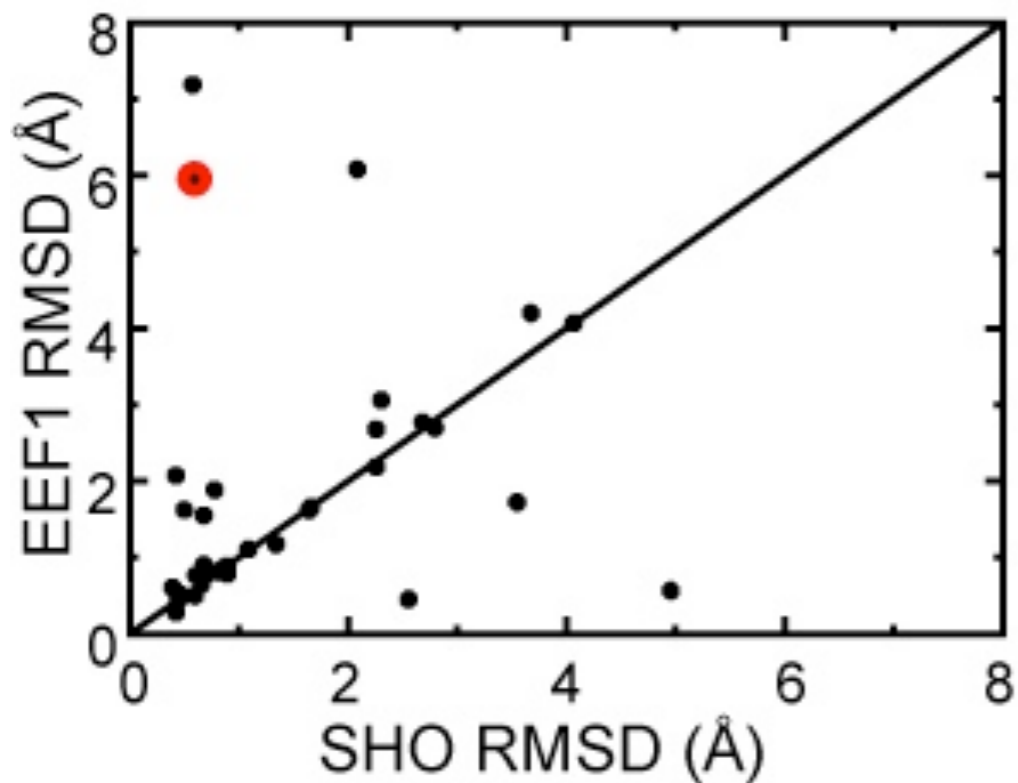
*SHO outperforms EEF1 for discrimination,  $p < 0.05$*



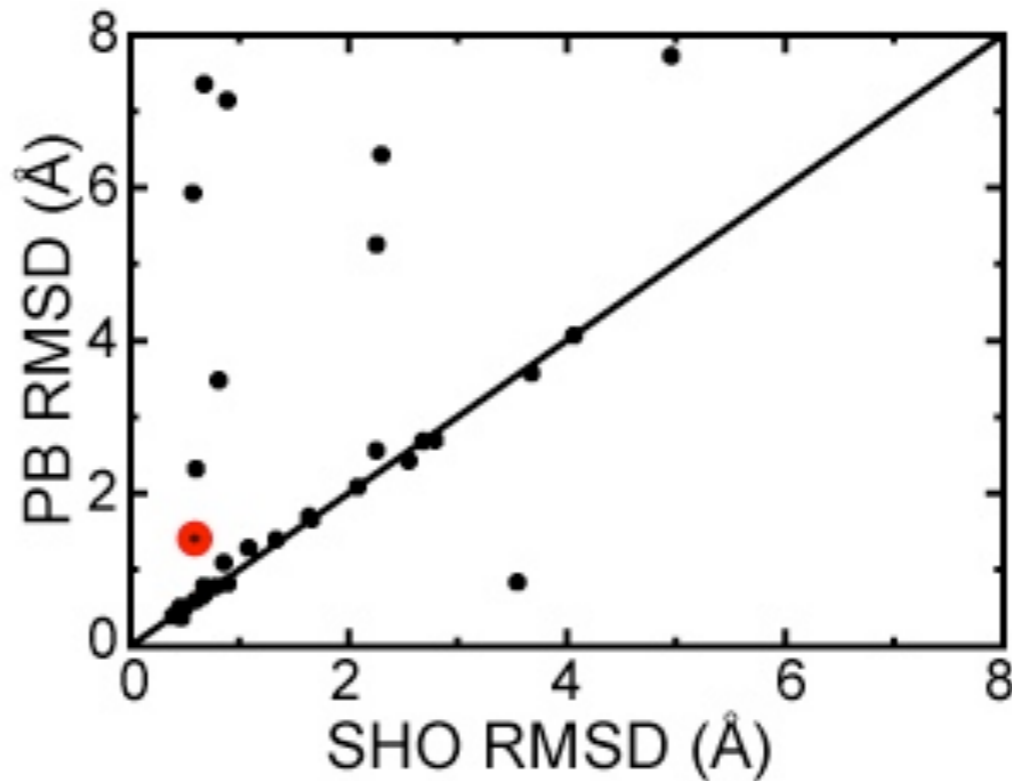
*SHO outperforms PB for discrimination,  $p < 0.02$*



# KIC loop modeling

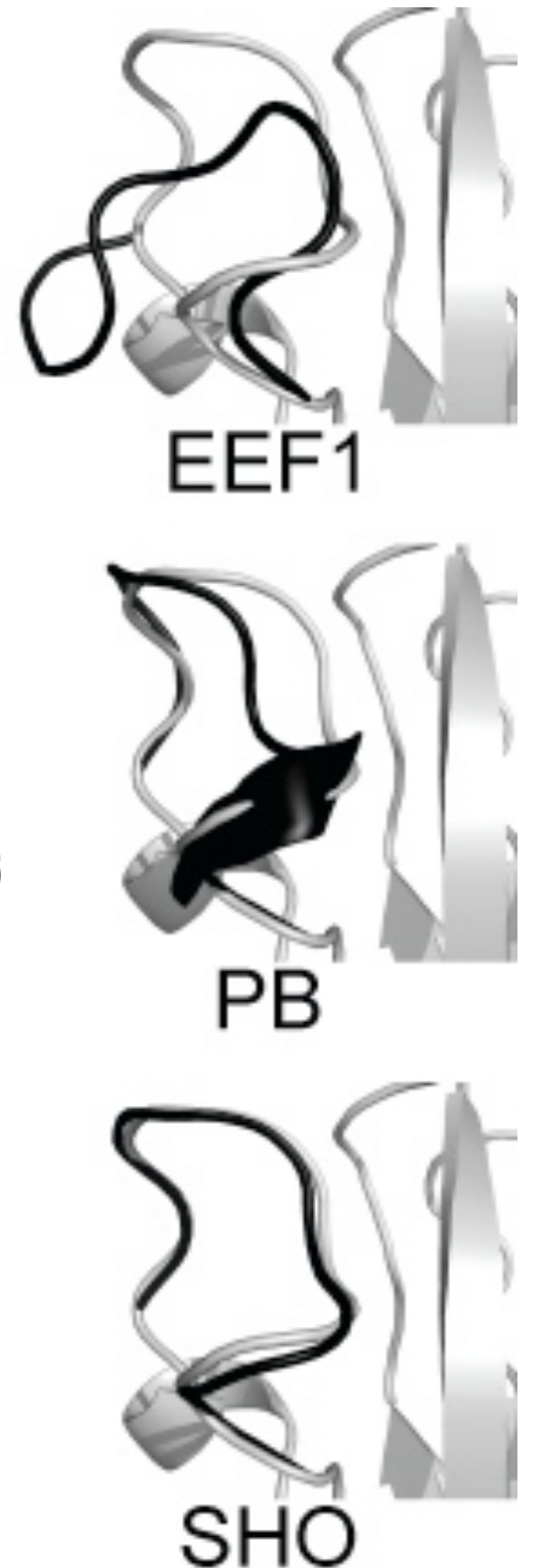


*SHO outperforms EEF1 for discrimination,  $p < 0.05$*



*SHO outperforms PB for discrimination,  $p < 0.02$*

*Average UNS in lowest-energy decoy:  
EEF1=1.8, PB=0.6, SHO=0.1*



# Loop modeling debriefing

- We haven't reweighted yet!

# Loop modeling debriefing

- We haven't reweighted yet!
- Most successes are very high-resolution examples (sub-angstrom predictions)



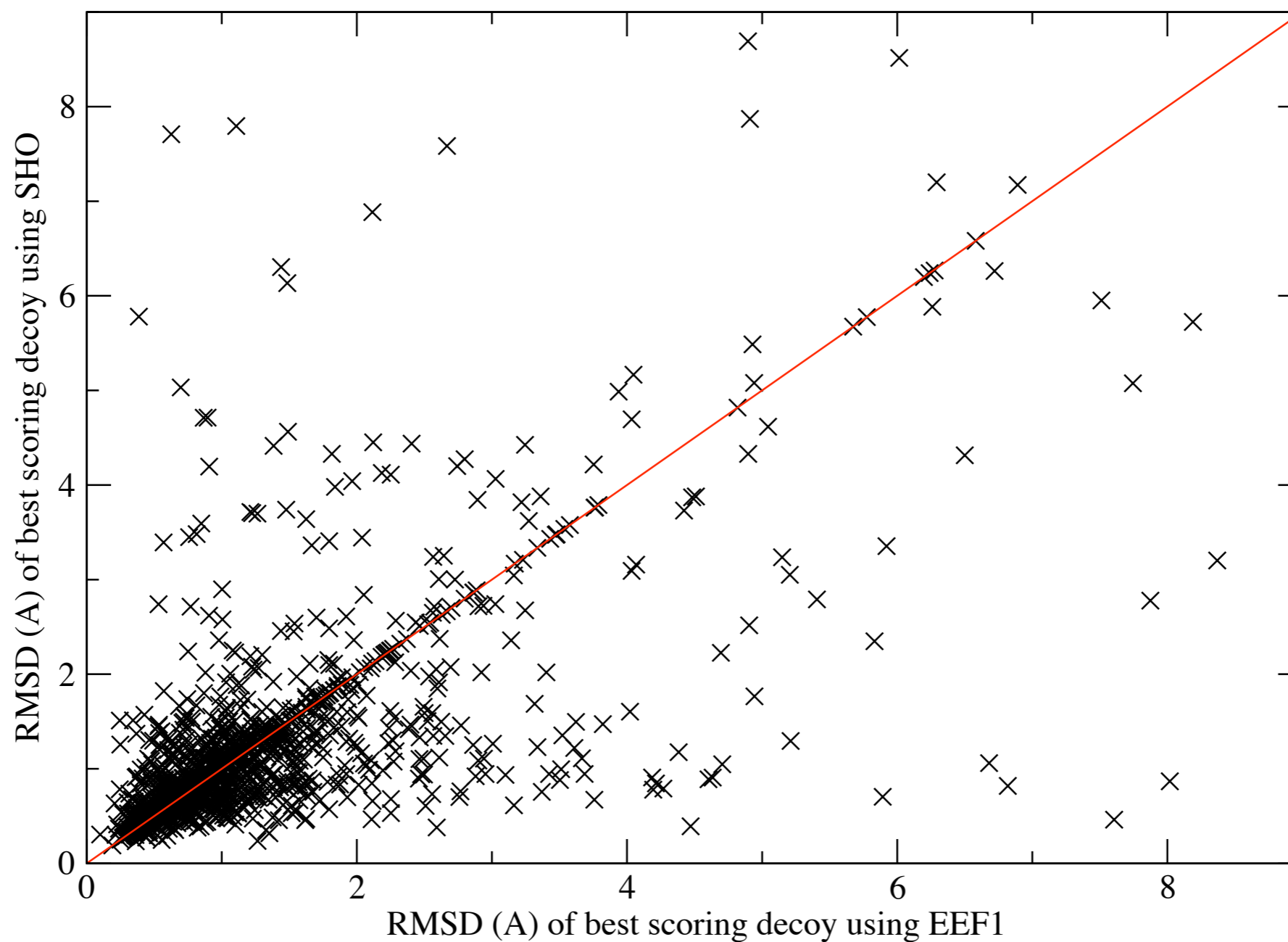
# Loop modeling debriefing

- We haven't reweighted yet!
- Most successes are very high-resolution examples (sub-angstrom predictions)
- Expect more rugged landscape than EEFI, SHO may not be useful if no sub-angstrom decoys are sampled

# Loop modeling debriefing

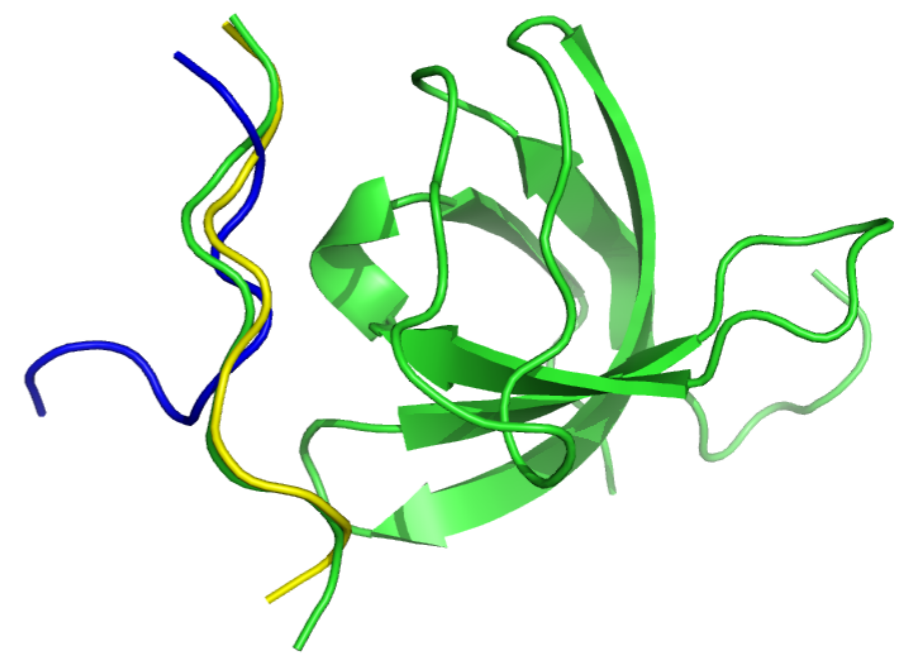
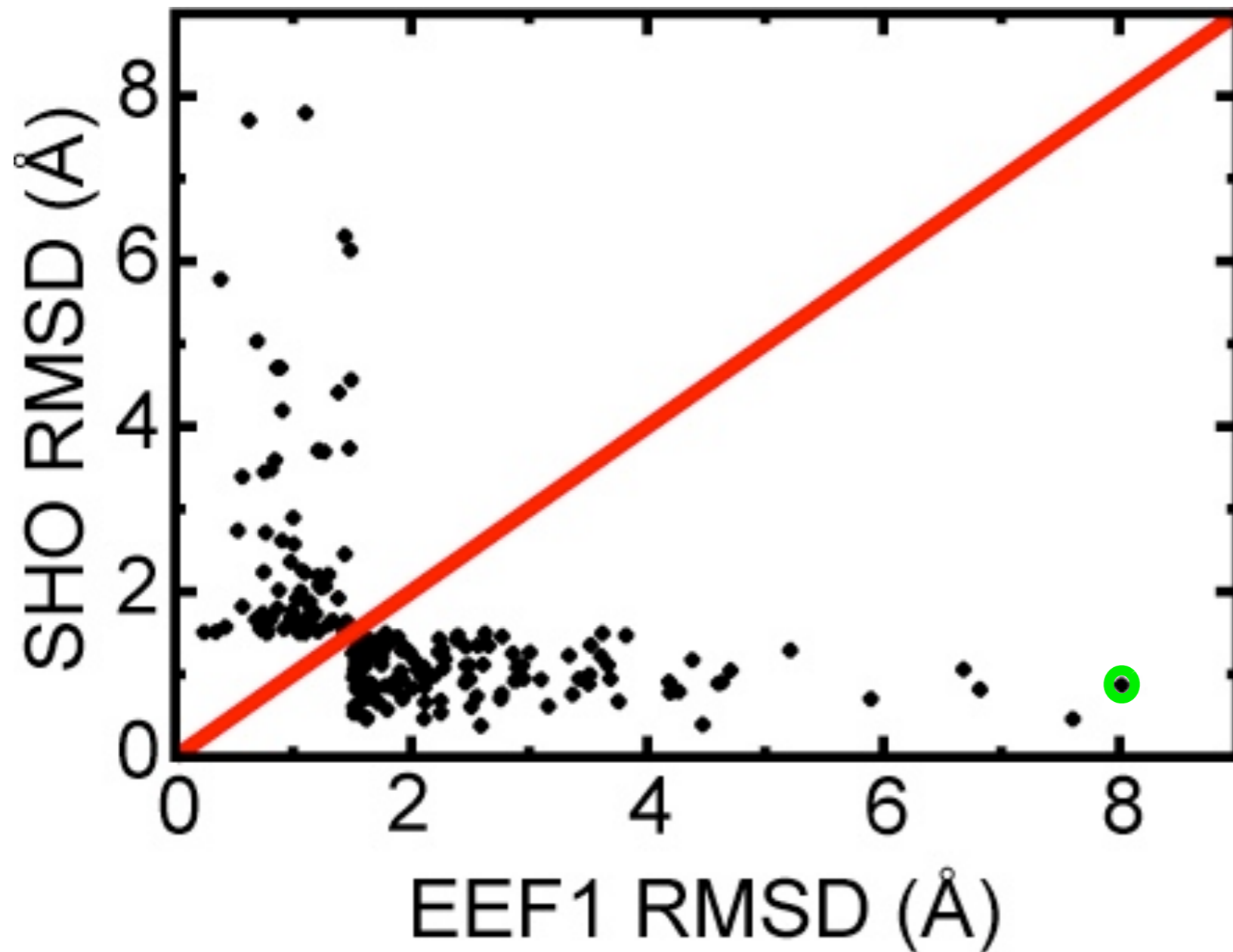
- We haven't reweighted yet!
- Most successes are very high-resolution examples (sub-angstrom predictions)
- Expect more rugged landscape than EEFI, SHO may not be useful if no sub-angstrom decoys are sampled
- Avoidance of UNS is a *very* stringent criterion - probably cuts down conformational space dramatically
  - Evidence from filtering of protein interface designs, etc.

# FlexPepDock



*Decoys from Raveh, London, Schueler-Furman*

# FlexPepDock



Native  
SHO (0.9 Å)  
EEF1 (8.0 Å)

*SHO better in 132 cases, EEF1 better in 69 cases  
SHO outperforms EEF1 for discrimination,  $p < 0.0001$*

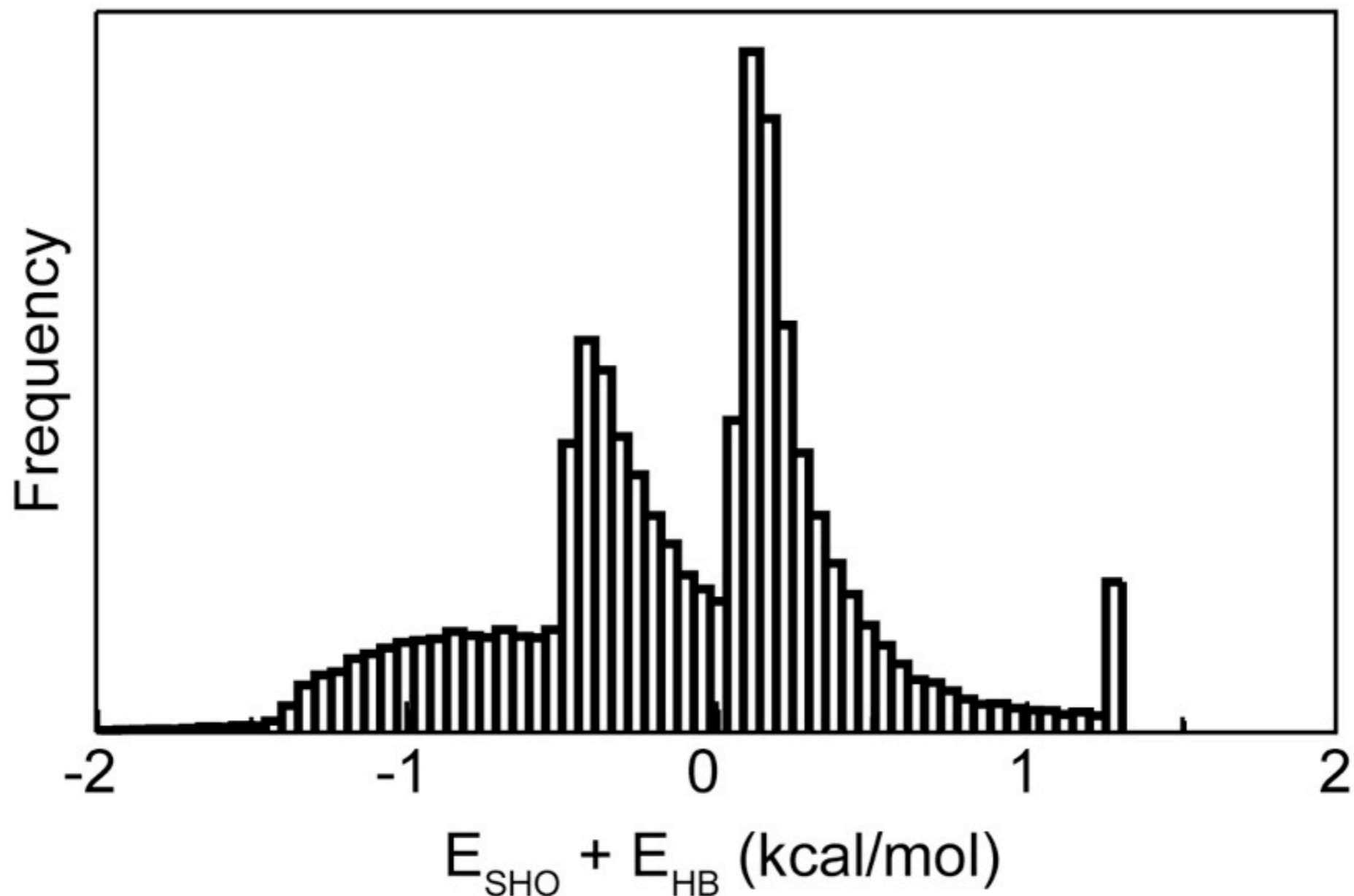
*Decoys from  
Raveh, London,  
Schueler-Furman*

# Watch for it!

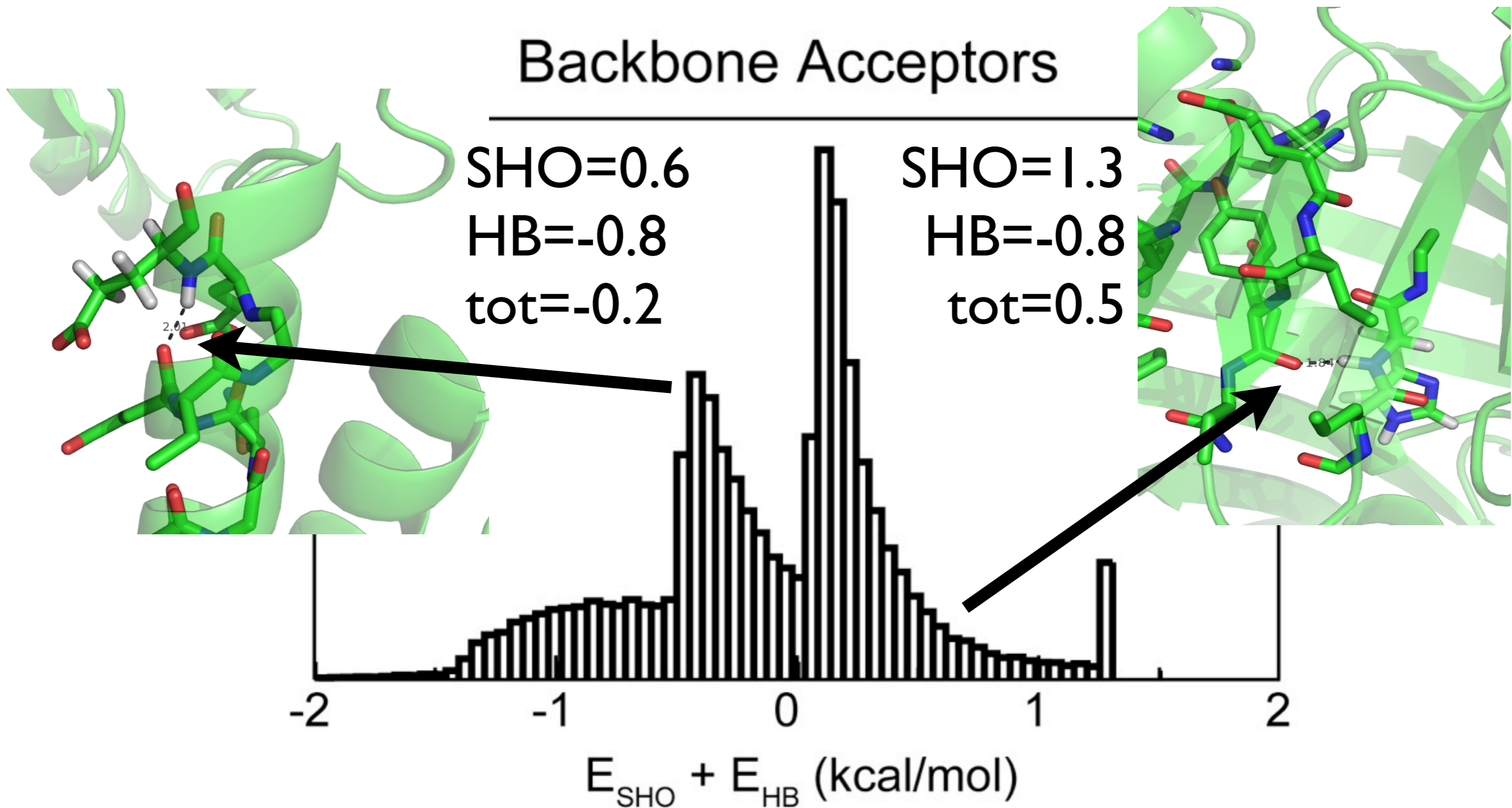
- We're hard at work on a fast, differentiable, pairwise-additive approximate version
- *Important caveat* - we've only done discrimination tests so far, haven't tried generating decoys yet

# Incomplete energetic trade-off between Hbonding and solvation

Backbone Acceptors

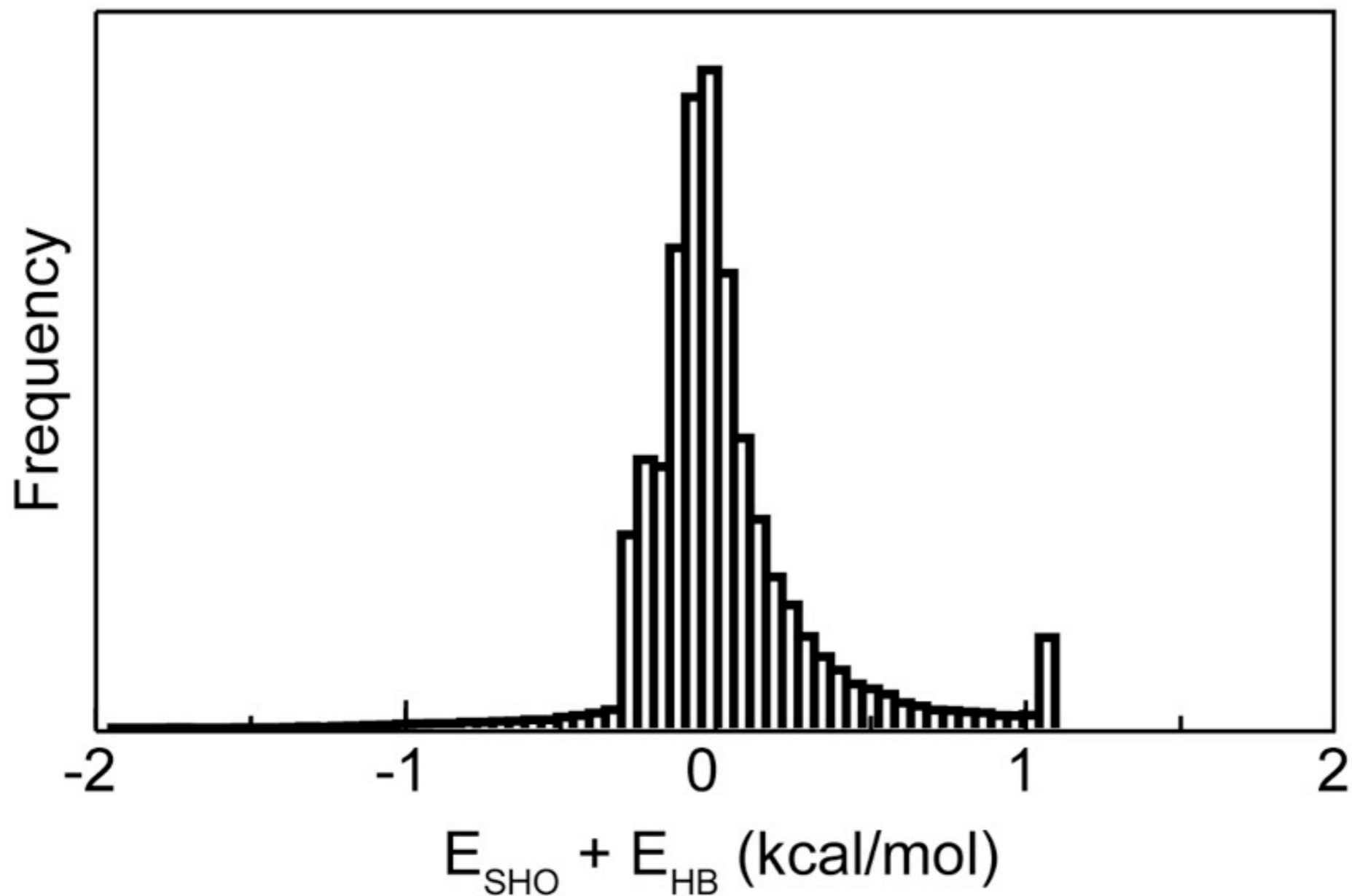


# Incomplete energetic trade-off between Hbonding and solvation



# Energetic trade-off requires environment dependent Hbonding

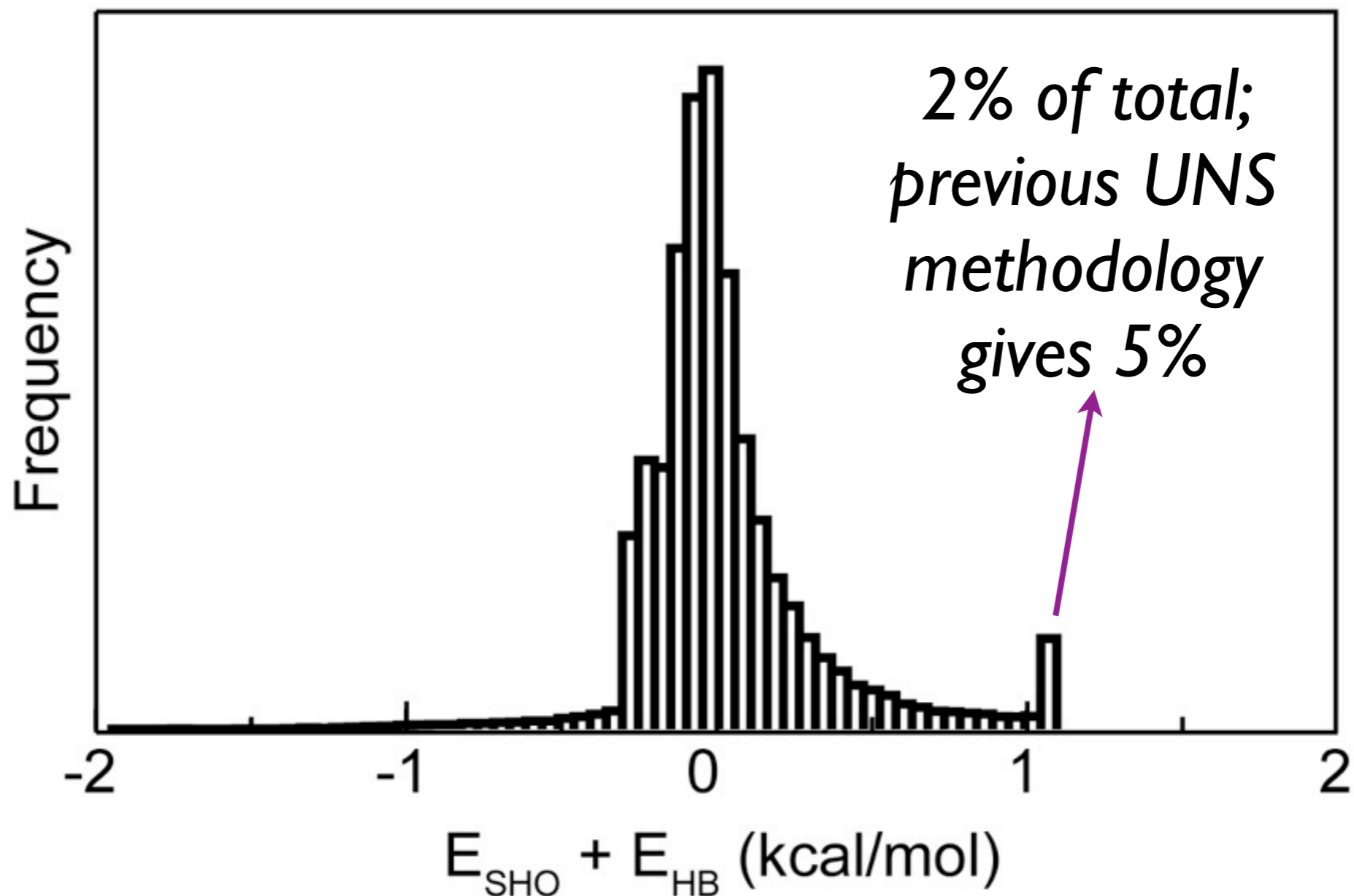
Backbone Acceptors





# Energetic trade-off requires environment dependent Hbonding

## Backbone Acceptors



# Recap

- Model is build by considering solute's potential for Hbonding to *discrete* solvent molecules
- As such, specifically penalizes occlusion that leads to UNS
- Seems to work well for discrimination, decoy generation is upcoming
- May also represent a better way of identifying UNS in decoys / designs

# Rosetta Acknowledgments

- Tanja Kortemme
- Rhiju Das
- Brian Kuhlman
- Dan Mandell
- ~~Barak Raveh~~ Nir London
- Ora Schueler-Furman
- Phil Bradley
- Colin Smith
- Sarel Fleishman
- Liz Kellogg
- David Baker
- Jeff Gray

# Acknowledgments

- Karanicolas lab:

Whitney Smith  
David Johnson  
Ragul Gowthaman  
Chet Egan  
Katelyn Deckert  
Yan Xia



KU startup funds

Alfred P. Sloan Research  
Fellowship

NIH COBRE PSF