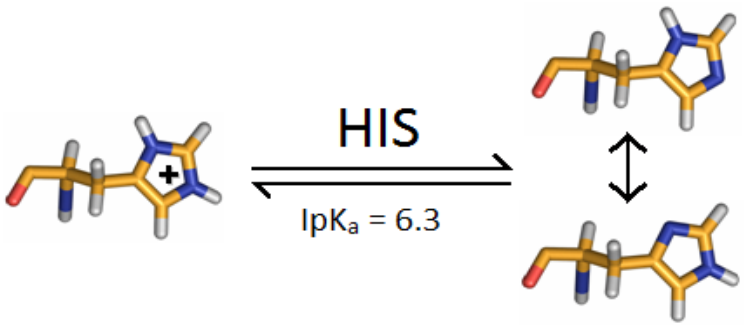
Towards fast and accurate calculation of protein pK_a values

Krishna Praneeth Kilambi & Jeffrey J. Gray

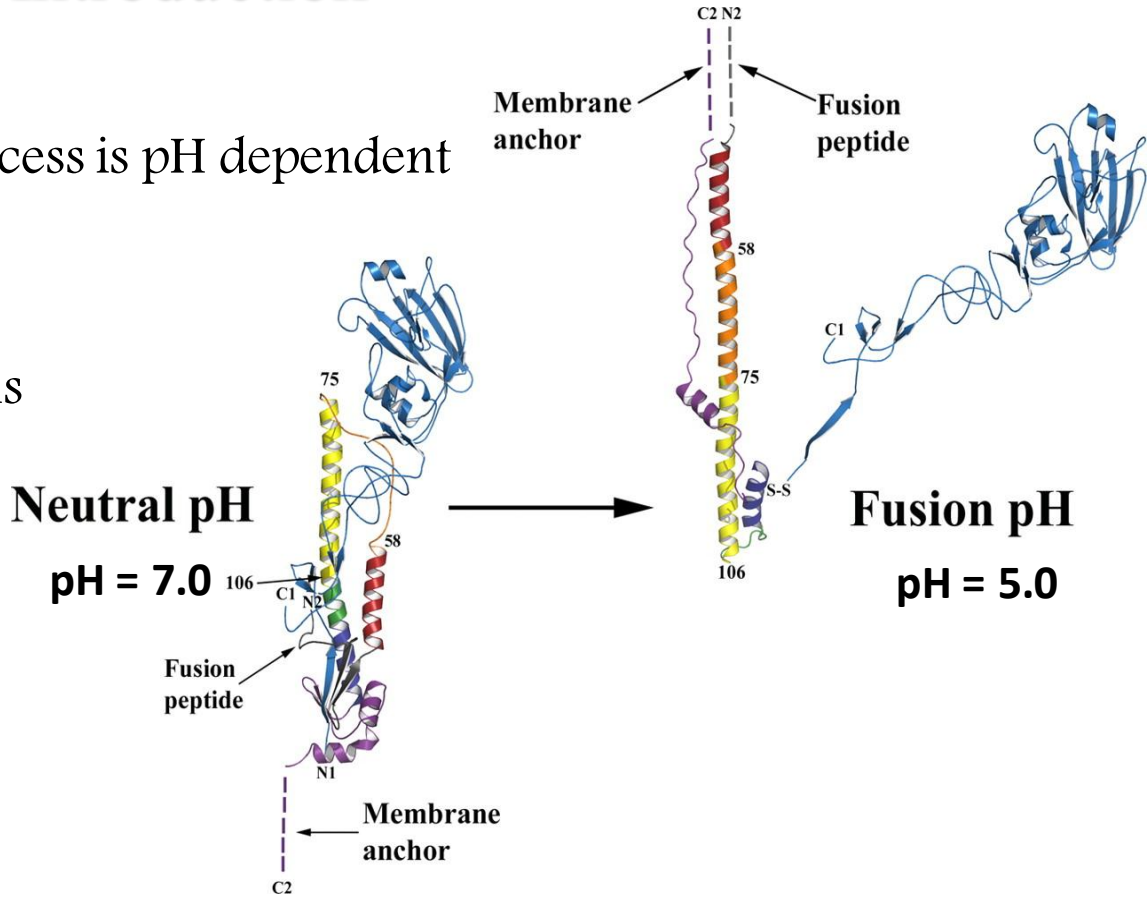
Johns Hopkins University

RosettaCon 2011



Introduction

- Virtually every biological process is pH dependent
  - protein folding
  - enzyme catalysis
  - protein-protein interactions
  - pathological conditions



- Influences protonation states of residues in proteins

Russell *et al.* 2008. Proc. Natl. Acad. Sci. USA. 105: 17736 -17741.

pK_a Values For Residues In Proteins

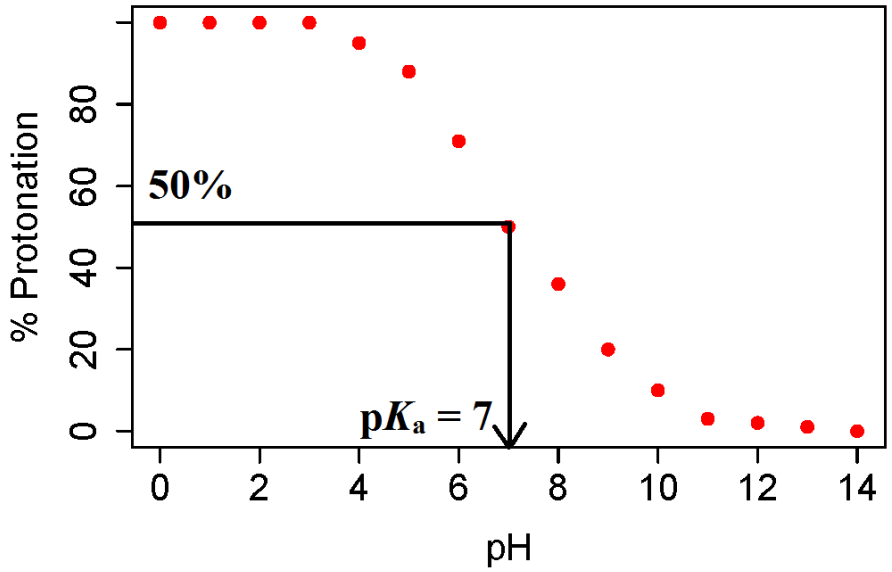
- Acid dissociation constant (pK_a) determines protonation state

$$[HA] \rightleftharpoons [H^+] + [A^-]$$

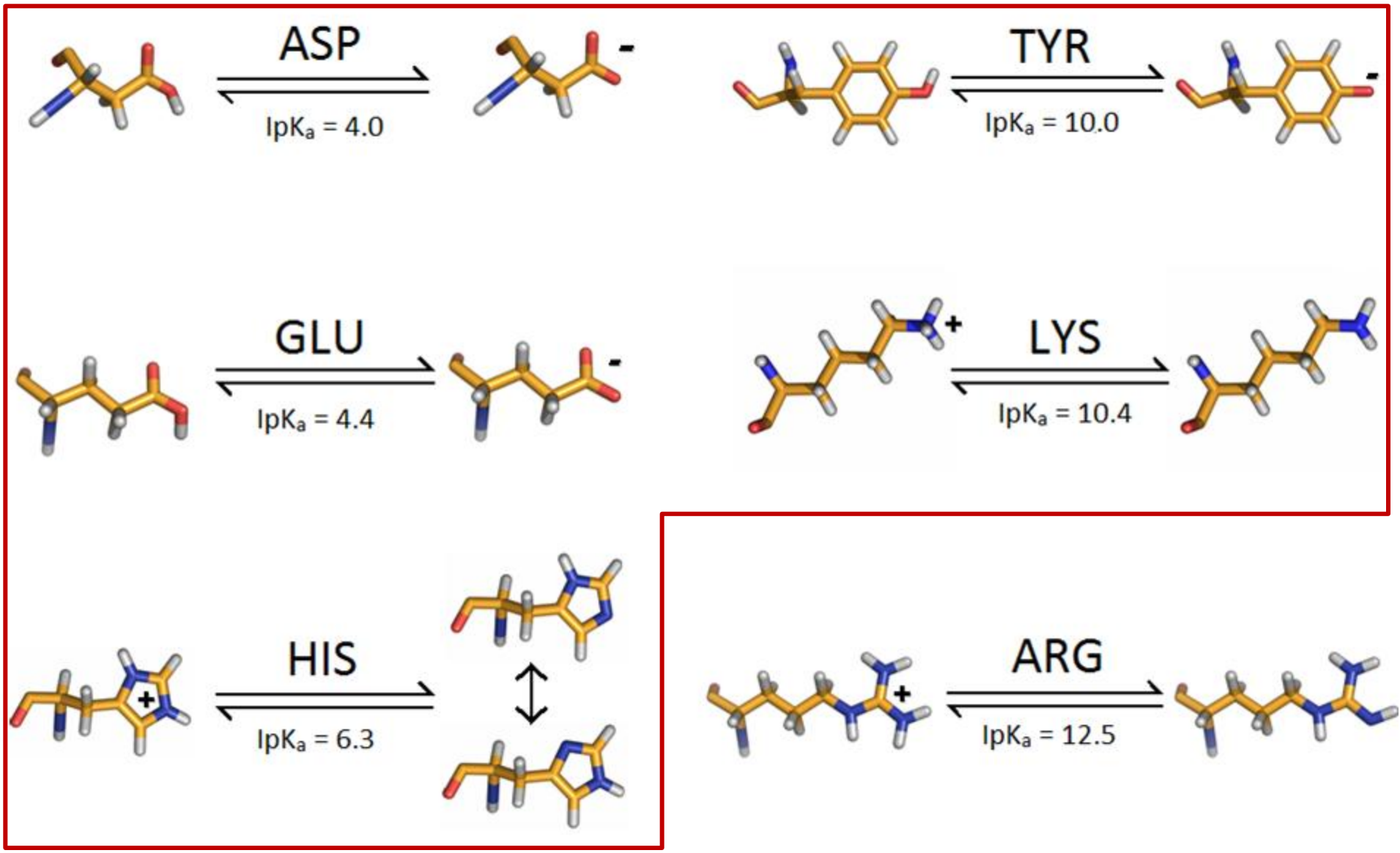
- Henderson–Hasselbalch equation

$$pH = pK_a + \log \frac{[A^-]}{[HA]}$$

- pH at 50% protonation = pK_a



- Residue pK_a in protein ≠ Intrinsic pK_a in solution

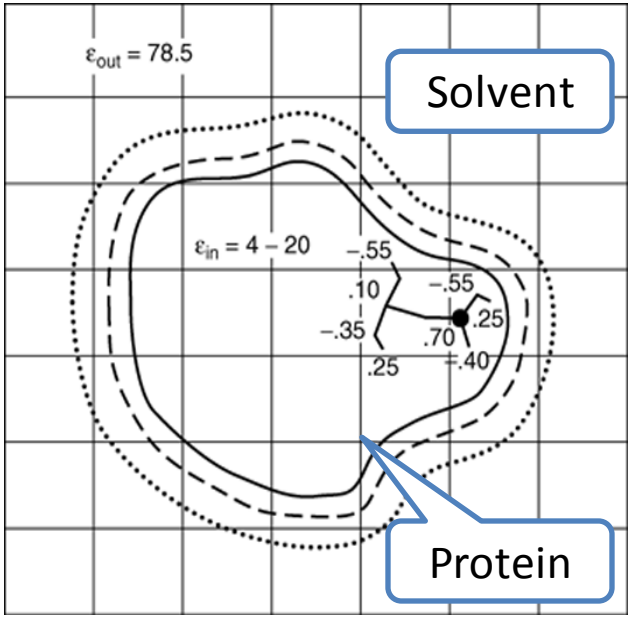

ASP
pK_a = 4.0
TYR
pK_a = 10.0
GLU
pK_a = 4.4
LYS
pK_a = 10.4
HIS
pK_a = 6.3
ARG
pK_a = 12.5

Current pK_a Prediction Methods

- Current methods
  1. continuum electrostatic models (Gunner, Harbury, McCammon etc.)
  2. all-atom MD simulations (Case etc.)
  3. empirical approaches (Jensen, Hellinga etc.)

- Advantages / Disadvantages
  - LPBE models overestimate charge interactions
  - empirical approaches are much faster

- **Rosetta toolset**
  - robust score function
  - varying levels of conformational flexibility
  - relatively computationally inexpensive

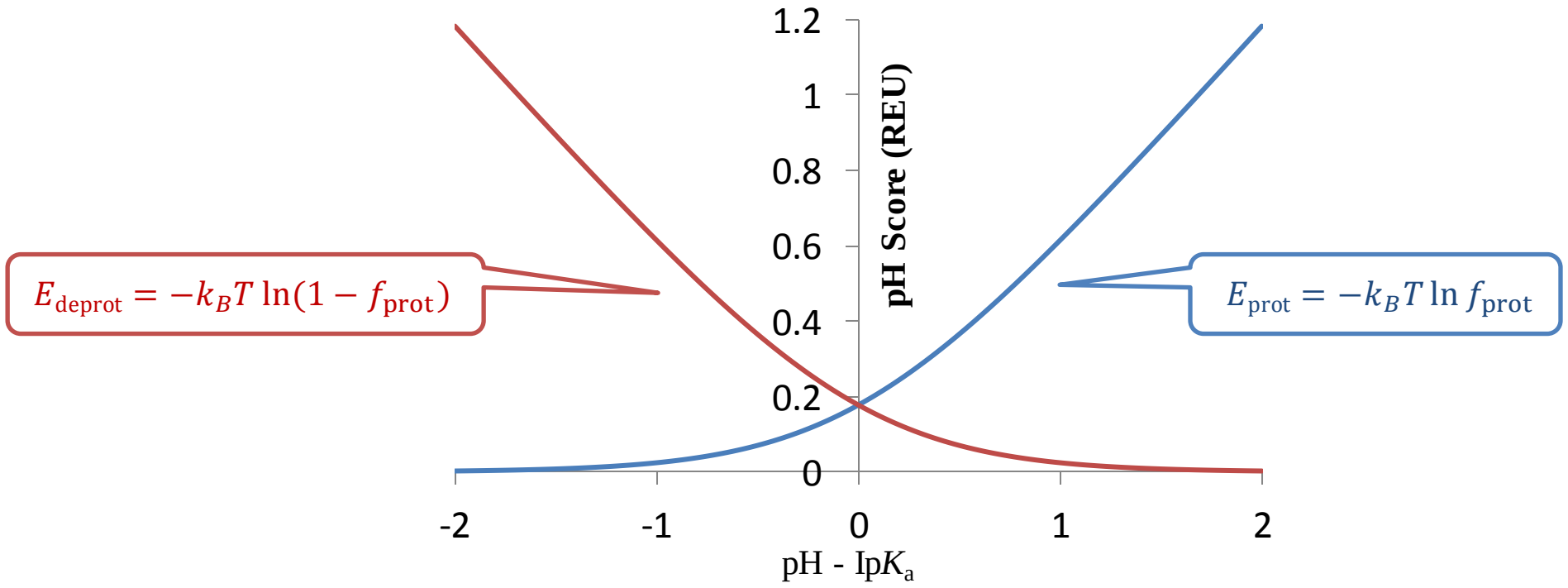


Protein-water system for FDPB calculations

Fitch, C.A., and B. García-Moreno E. 2002. Current Protocols in Bioinformatics

‘pH Score’ Captures Residue Protonation State Changes

- pH score
- Protonation probability $f_{\text{prot}} = \frac{1}{10^{\text{pH}-\text{Ip}K_a+1}}$ (Onufriev et al. 2001)
- $E_{\text{new}} = E_{\text{standard}} + E_{\text{pH}}$



Rotamer Frequency

Starting Structure
(Highly Acidic Environment pH = 1)

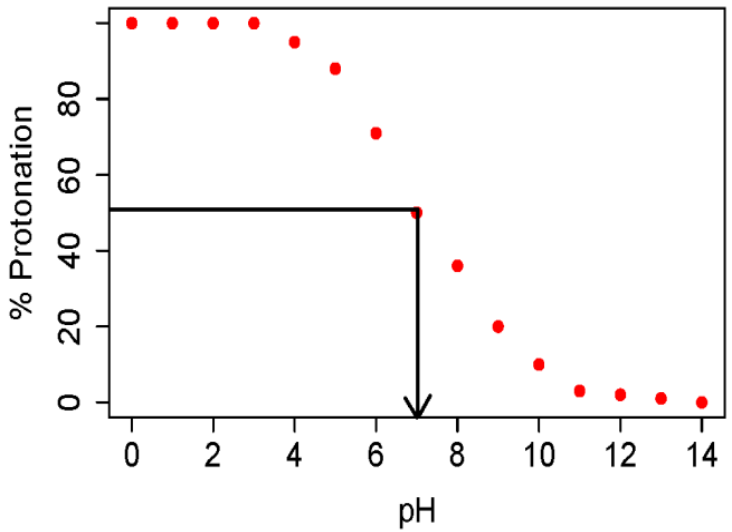
Sample Rotamer Library For Target Residue
(Includes Prot. And Deprot. Variants)

Collect Rotamer Protonation Statistics

Calculate pK_a From Residue Titration Plot

x 14

pH = pH + ΔpH
Titrate Until pH = 14



Starting Structure
(Highly Acidic Environment pH = 1)

Sample Rotamer Library For Target Residue
(Includes Prot. And Deprot. Variants)

Select Lowest Energy Rotamer

Did protonation state change?

NO

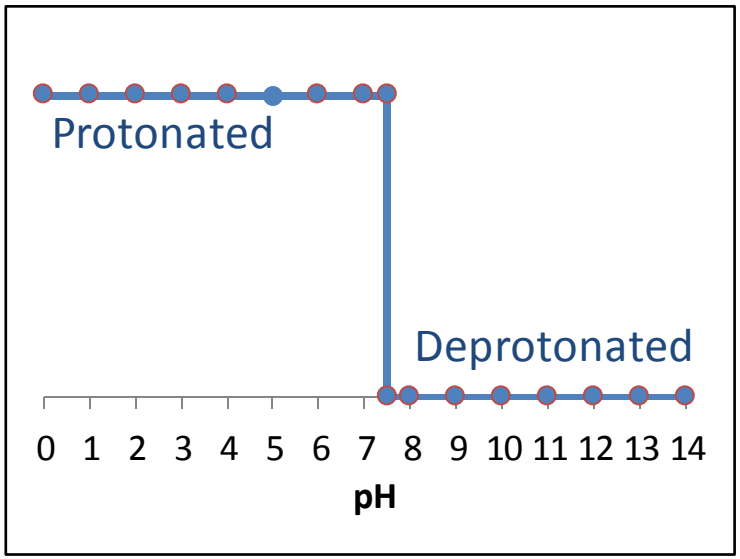
pH = pH + ΔpH

YES

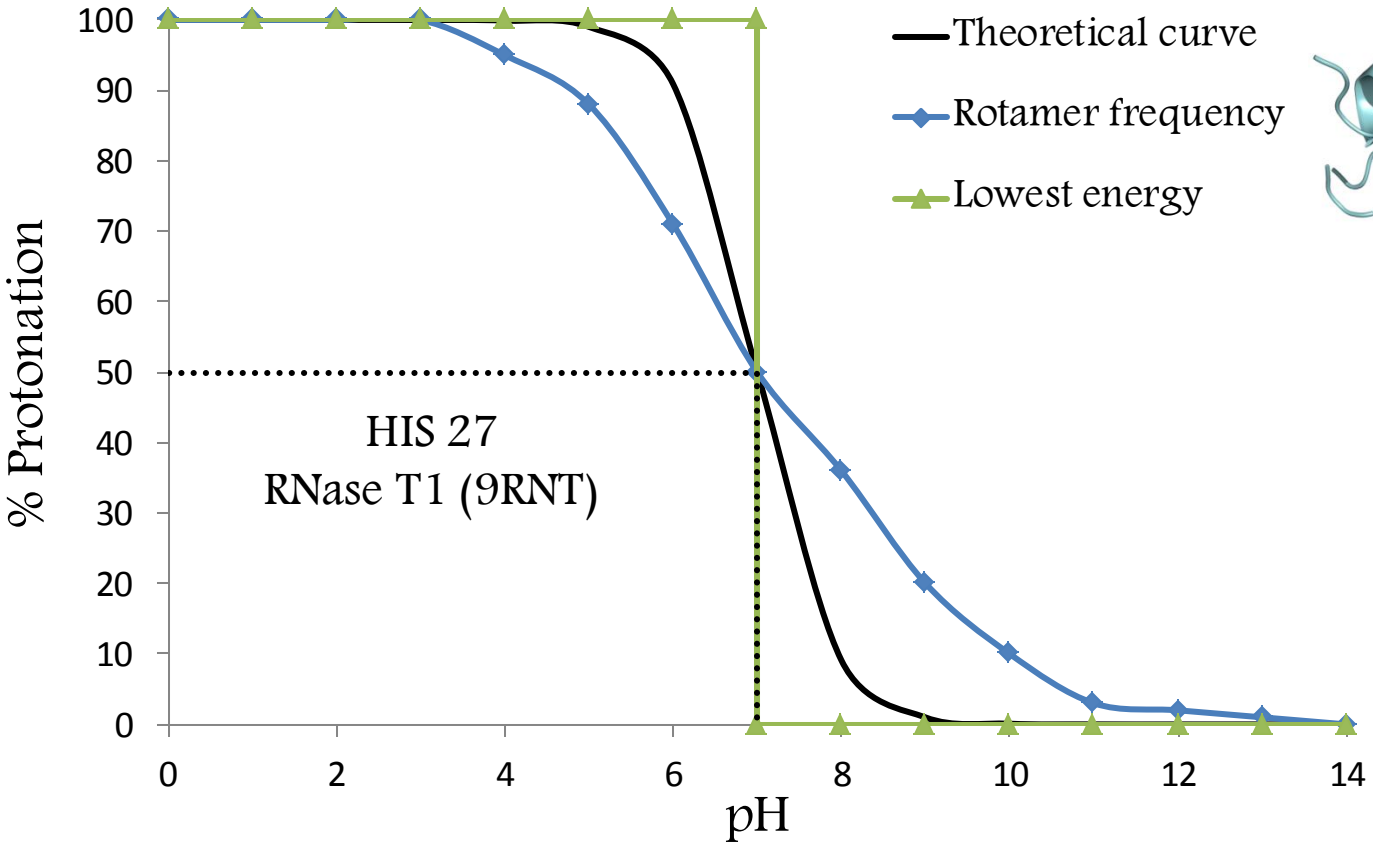
Output pK_a

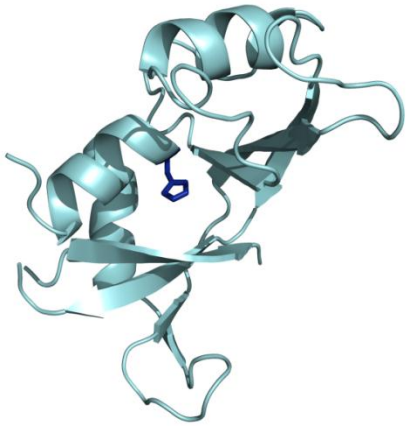
Lowest Energy

Optimize Rosetta Score Function For pK_a prediction

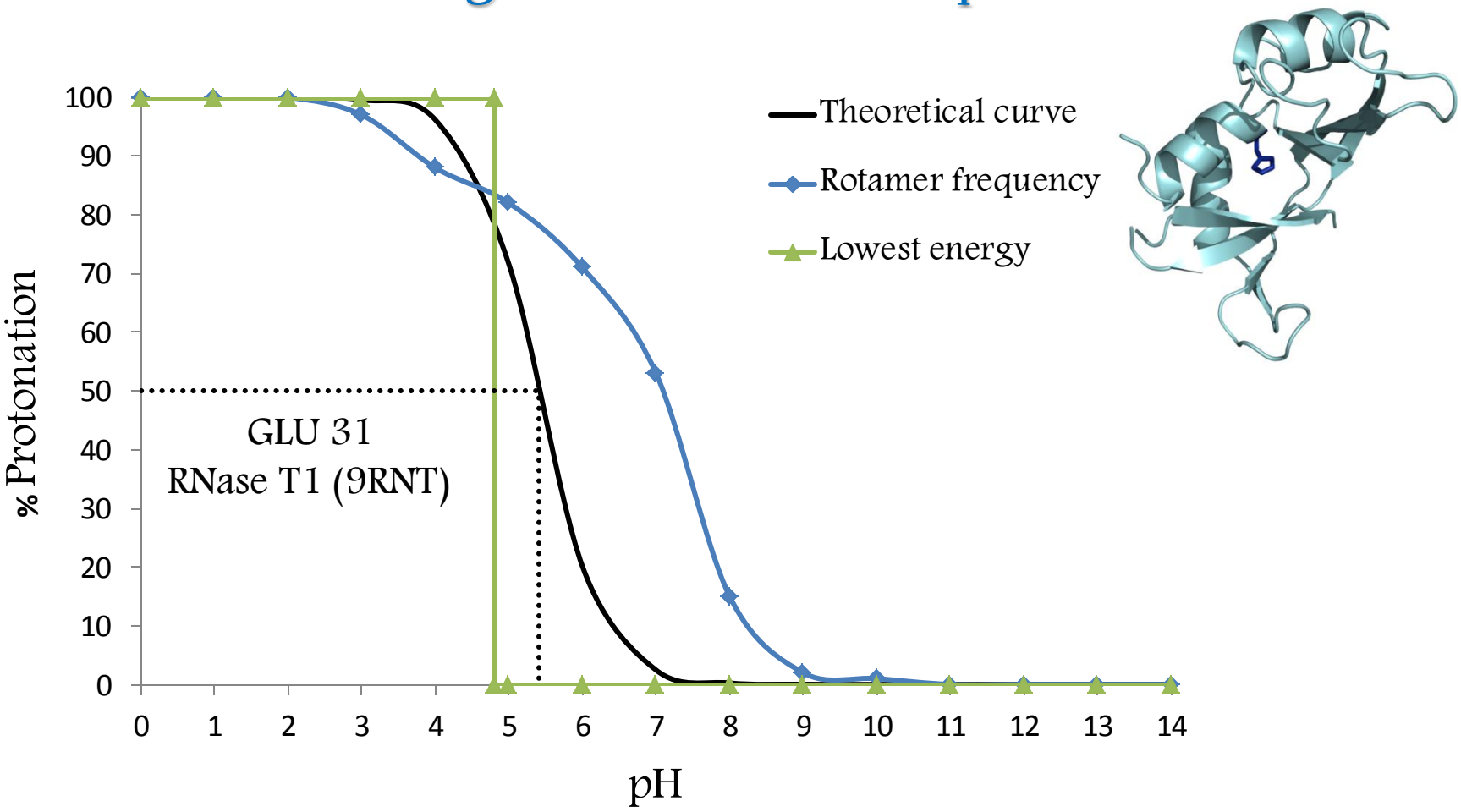


Lowest-Energy Conformation Is Better Than Boltzmann-Weighted Rotamer Frequencies





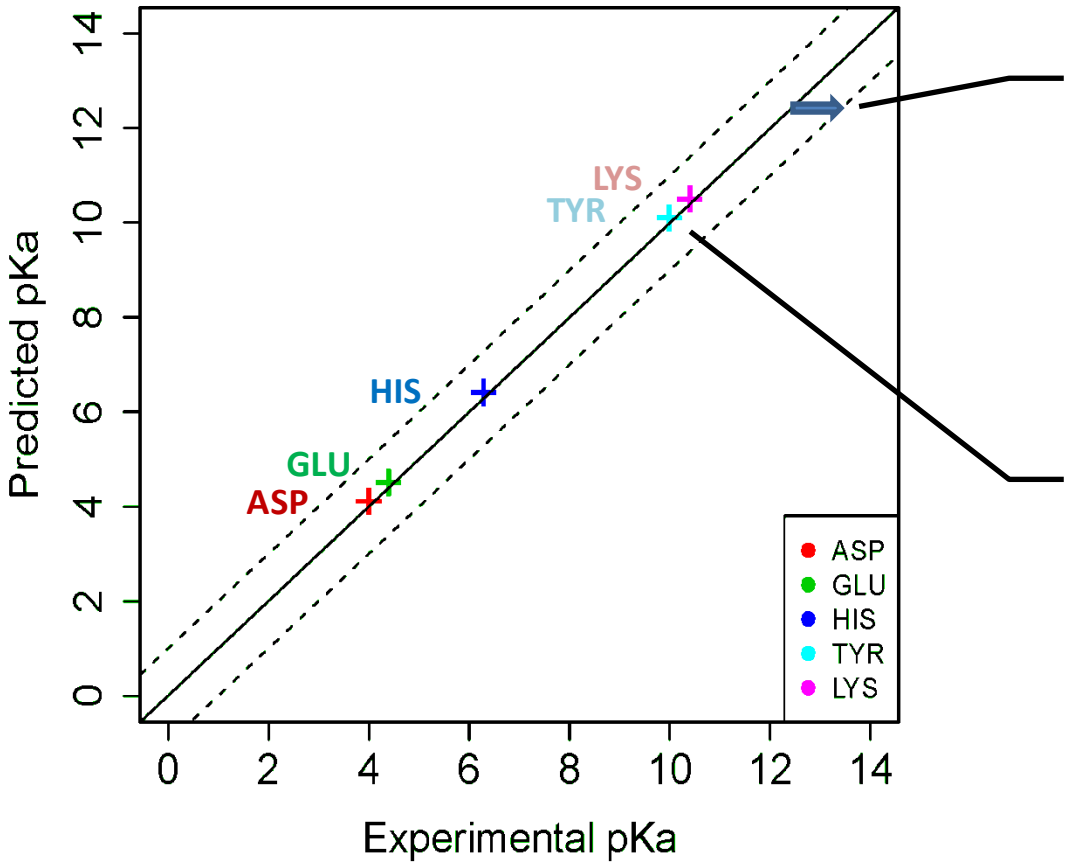
Lowest-Energy Conformation Is Better Than Boltzmann-Weighted Rotamer Frequencies



- Consistent with previous studies of side-chain entropy (Kortemme *et al.* 2002, Hu *et al.* 2006)

pK_a Prediction vs. Experiment

$$E = E_{\text{std}} + E_{\text{pH}}$$

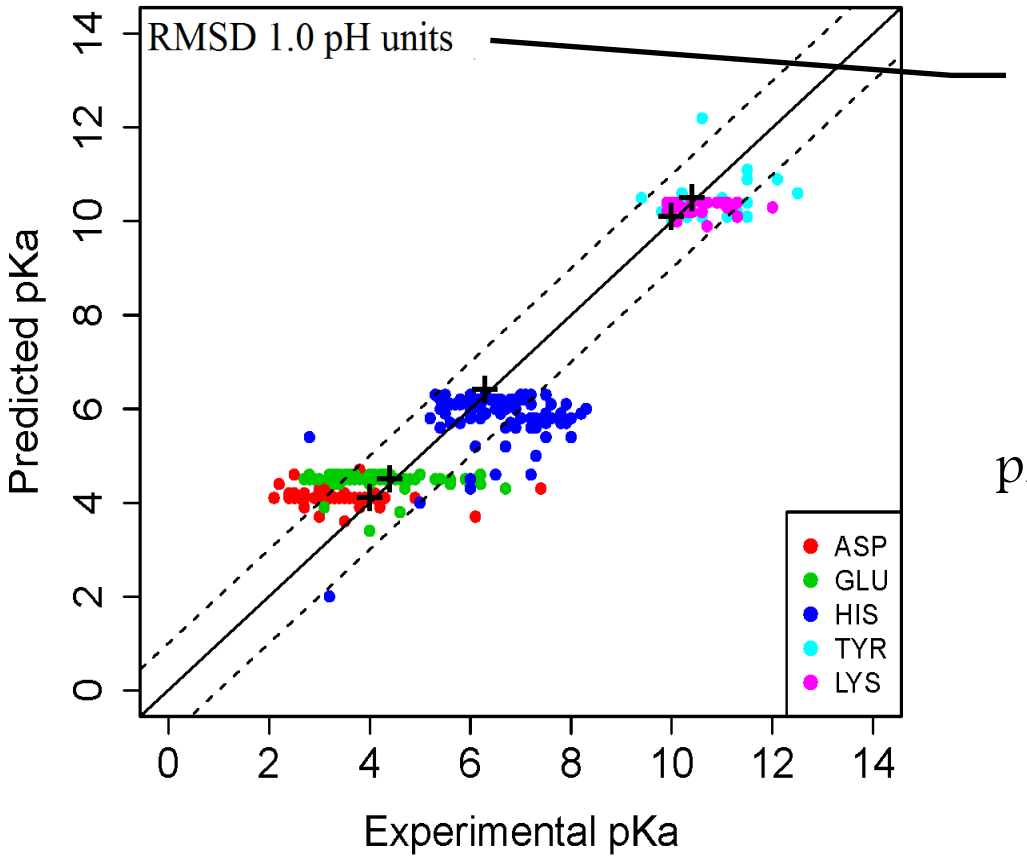


Error range 1 pH unit

Intrinsic pK_a of residue

Rosetta Standard Energy Function Predicts No pK_a Shifts

$$E = E_{std} + E_{pH}$$



RMSD of predicted pK_as from experimental pK_as

pK_a prediction methods (RMSD <1)

- MCCE2 RMSD 0.90 pH units
- PROPKA3 RMSD 0.79 pH units
- Null RMSD 0.92 pH units

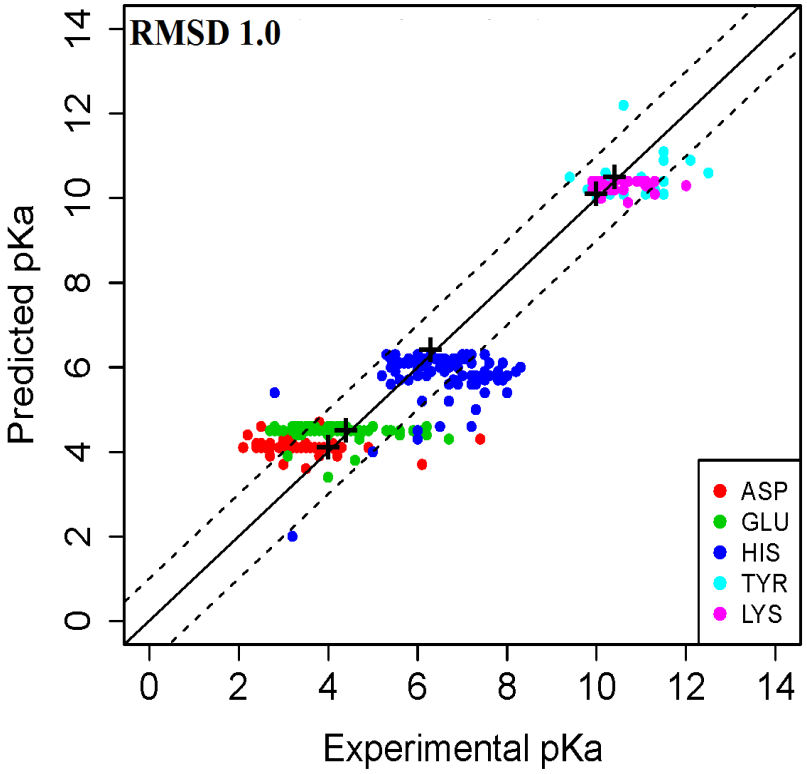
Song *et al.* 2009, Olsson *et al.* 2011

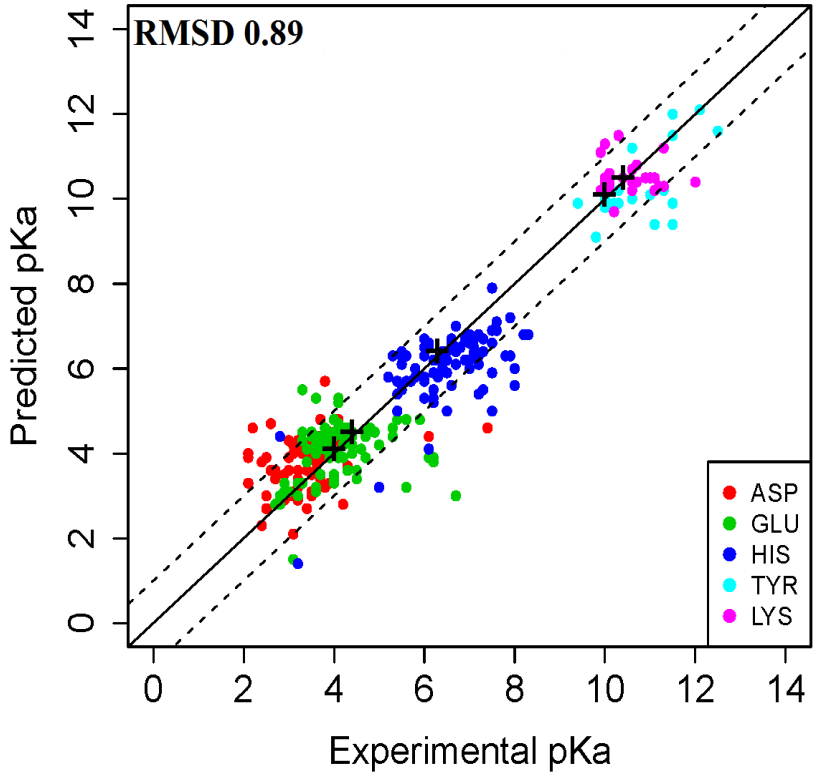
Using Explicit Coulomb Potential Improves pK_a Predictions

(hack_elec E $\propto \frac{q_1 q_2}{\varepsilon r}$; $\varepsilon \propto r$)

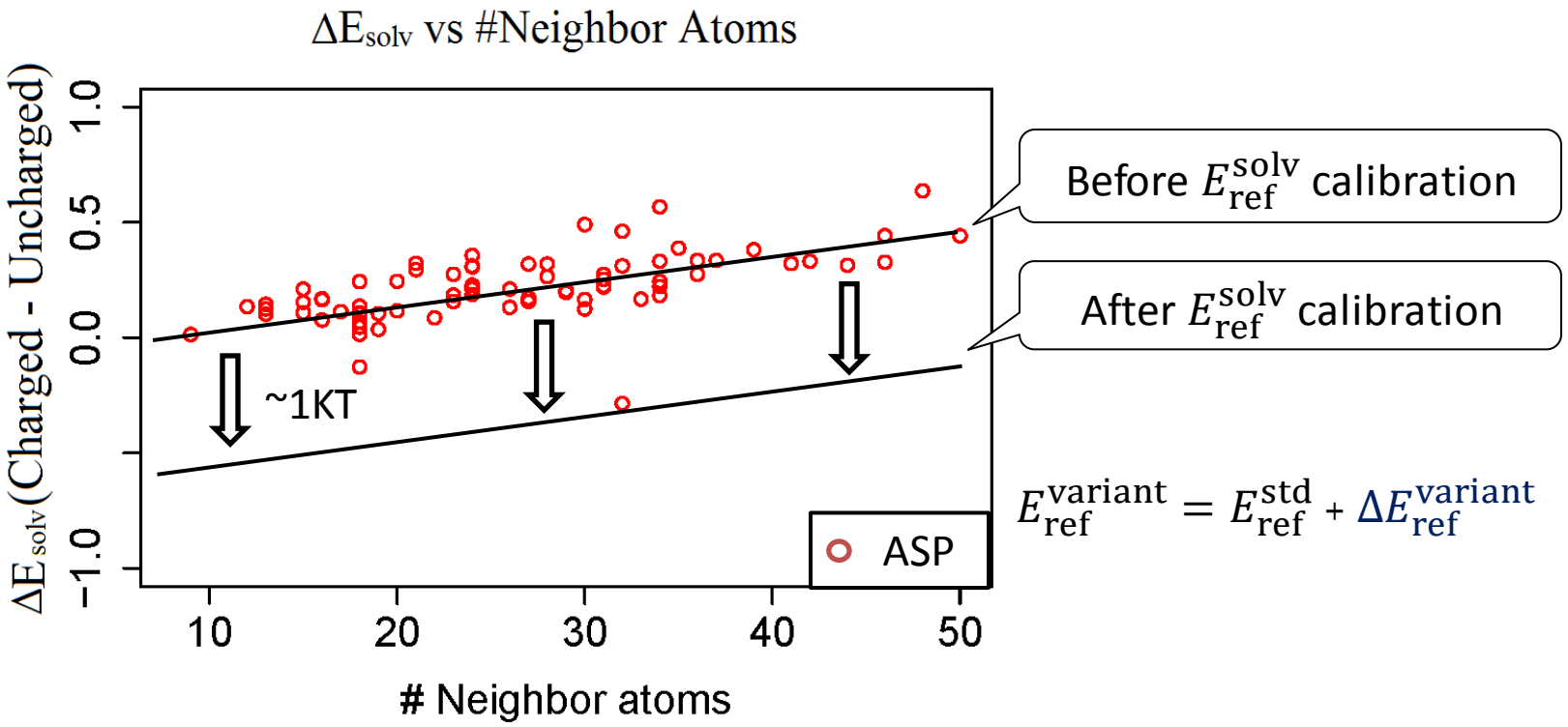
$$E = E_{\text{std}} + E_{\text{pH}}$$

$$E = E_{\text{std}} + E_{\text{pH}} + E_{\text{elec}}$$





Recalibrating LK Solvation Reference Energies For New Residue Protonation Variants

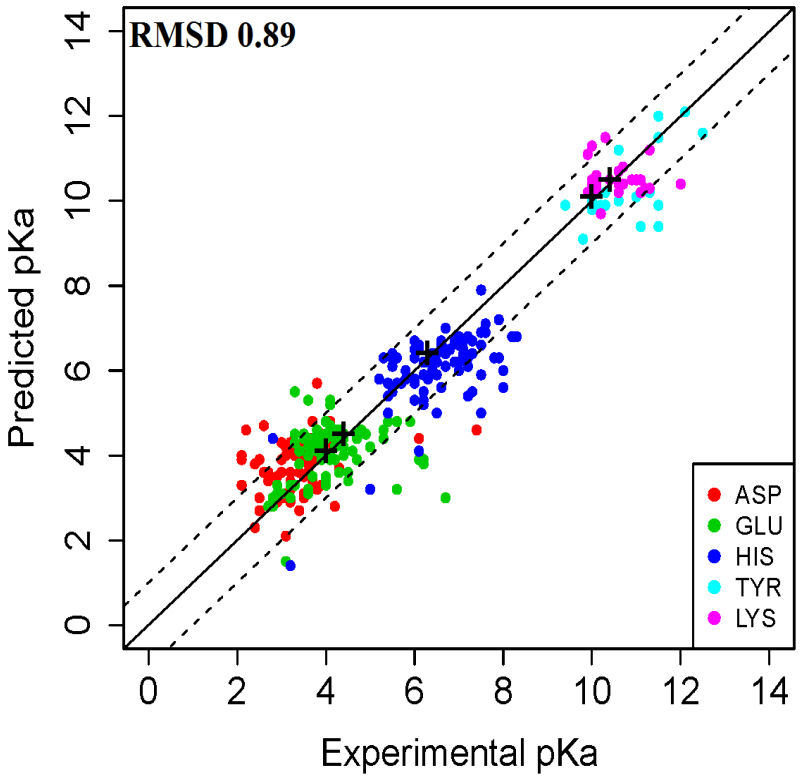


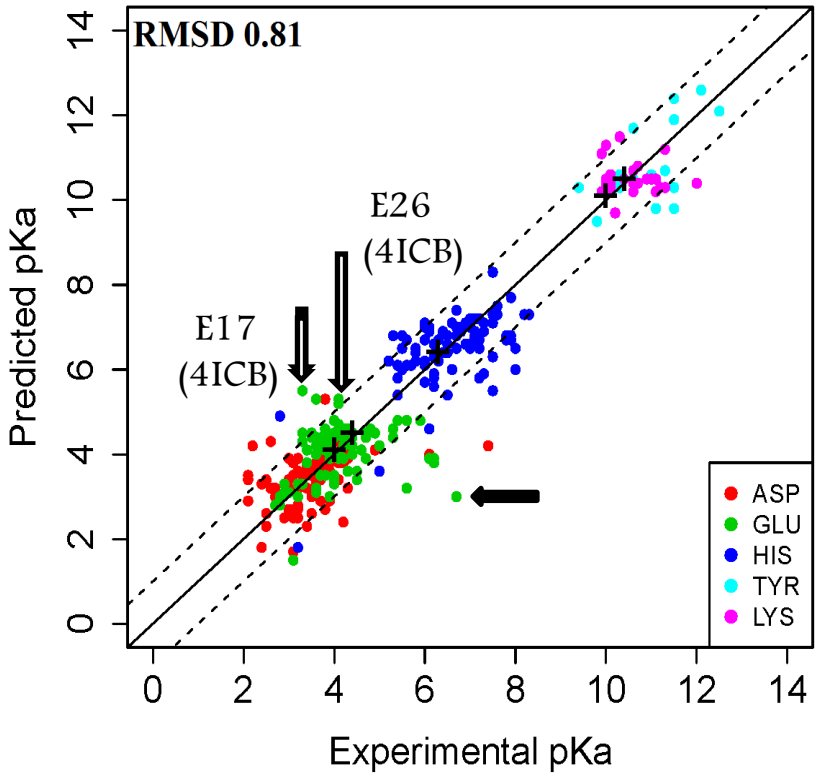
- $E_{charged}^{solv} - E_{uncharged}^{solv} \sim 0$ when exposed to solvent
- $\Delta E_{residue}^{solv} = \sum_{atoms}(\Delta E_{i}^{ref} - \sum_{j \neq i} f_i(r_{ij}) V_j)$ (Lazaridis *et al.* 1999)

Rosetta pH Energy (RPH) Function For pK_a Prediction

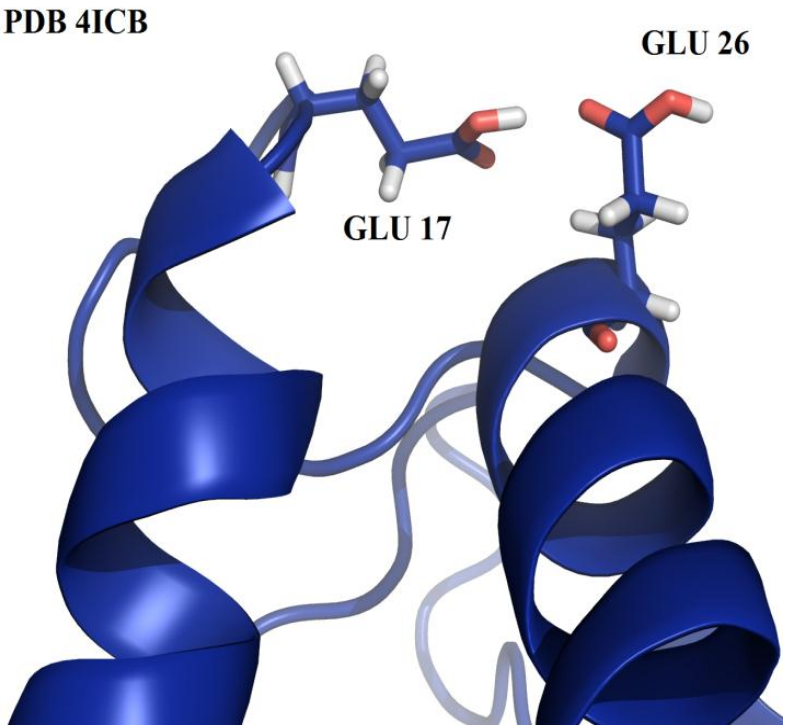
$$E = E_{std} + E_{pH} + E_{elec}$$

$$E = E_{std} + E_{pH} + E_{elec} + E_{ref}$$



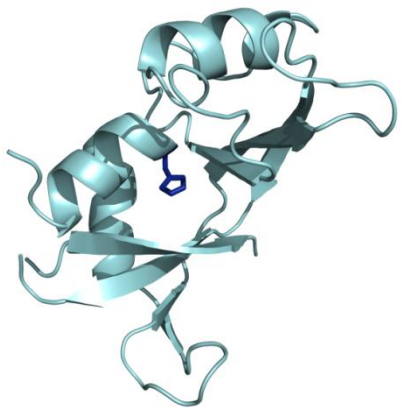


Errors ?

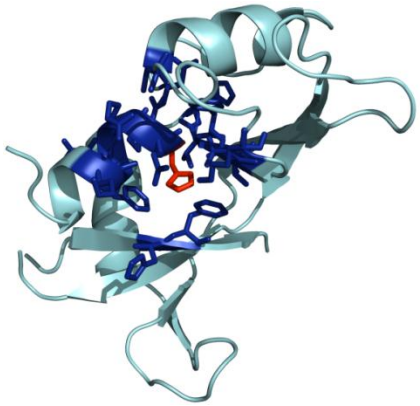


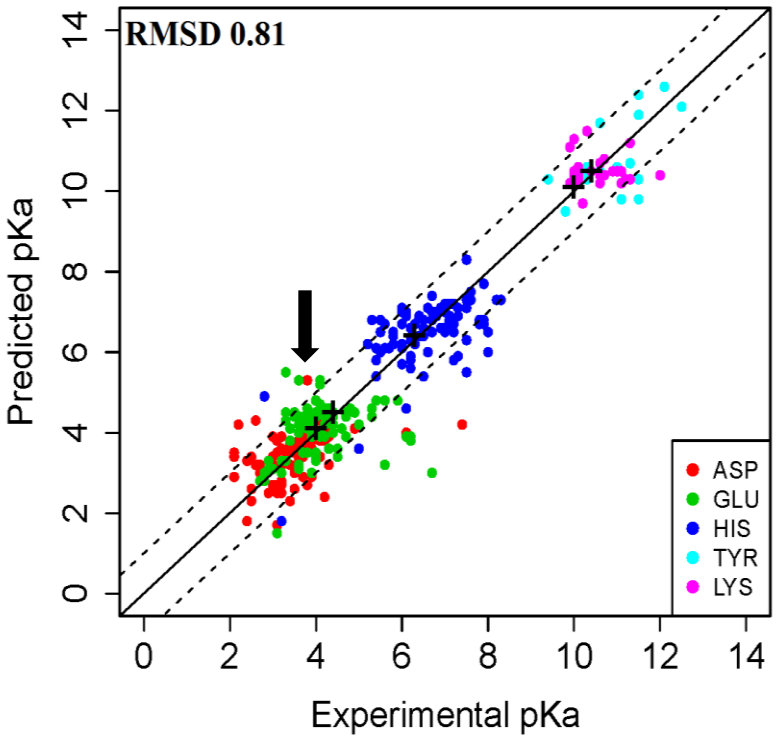
Glu-Glu interaction in Calbindin D9K

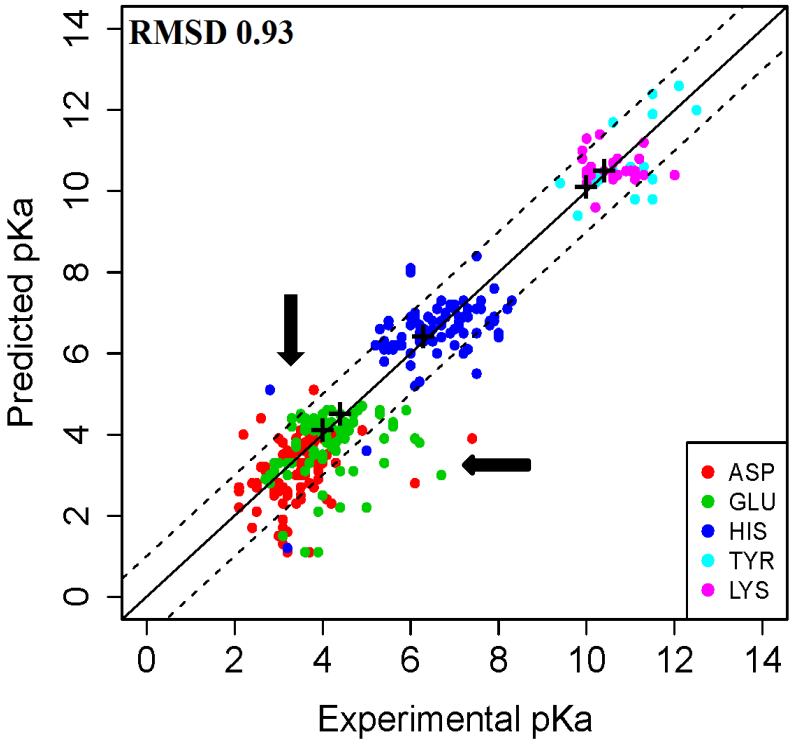
Packing Neighbors (6Å): Some Improvements, But Negates pK_a Shifts



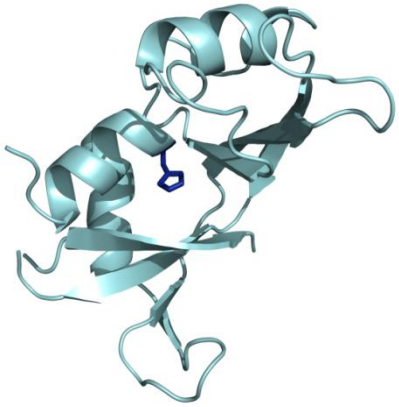
Repack < 6Å



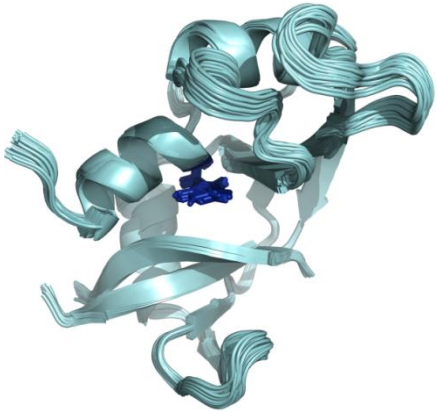


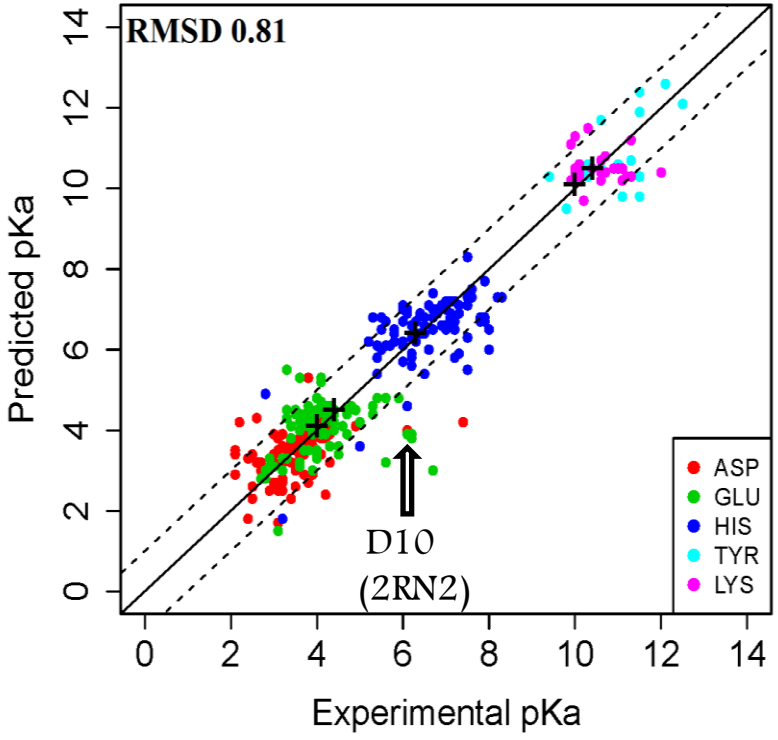


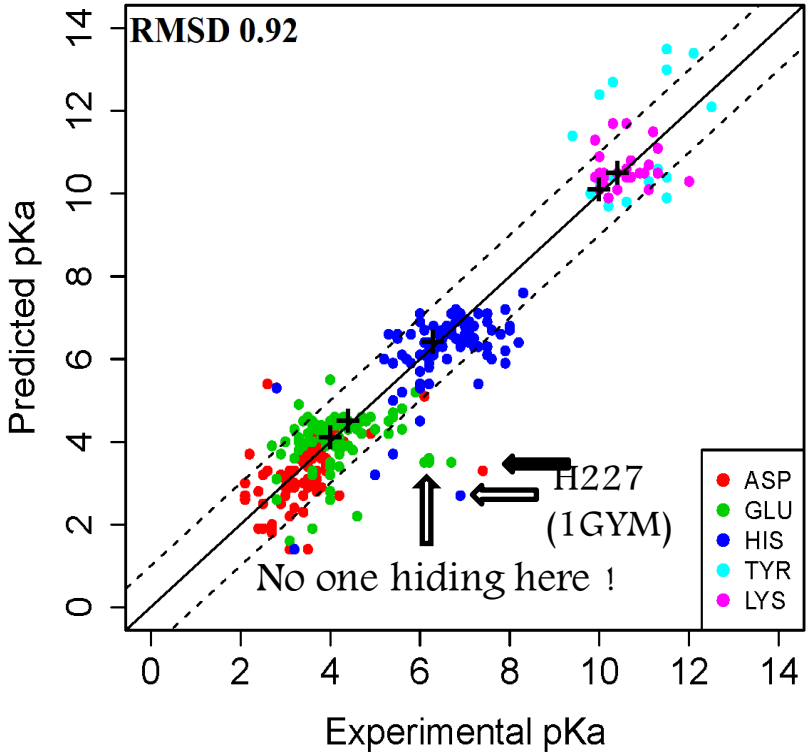
Structural Ensemble : Large Range Of Predictions → Highly Sensitive To Backbone Conformation

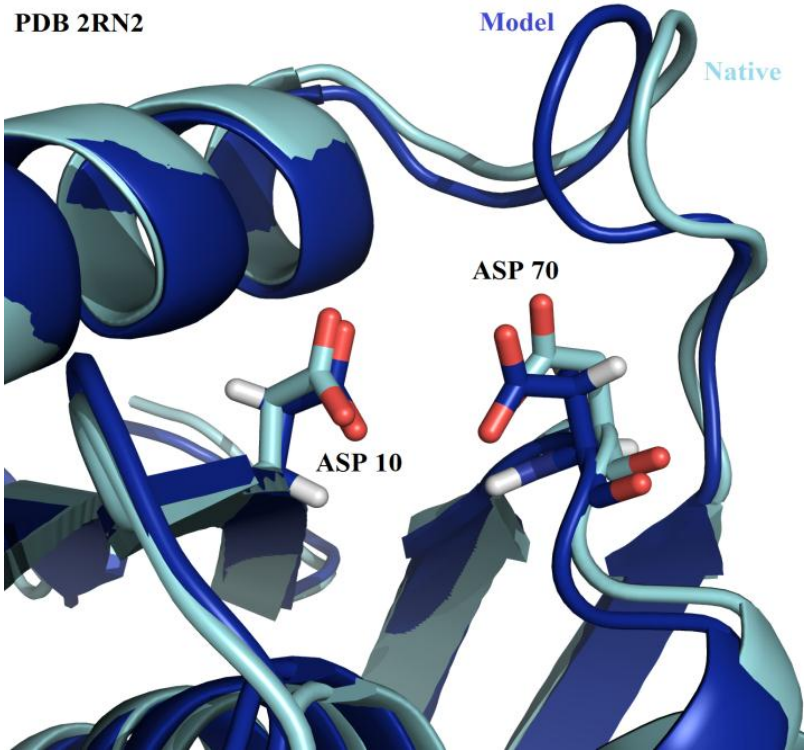


ClassicRelax

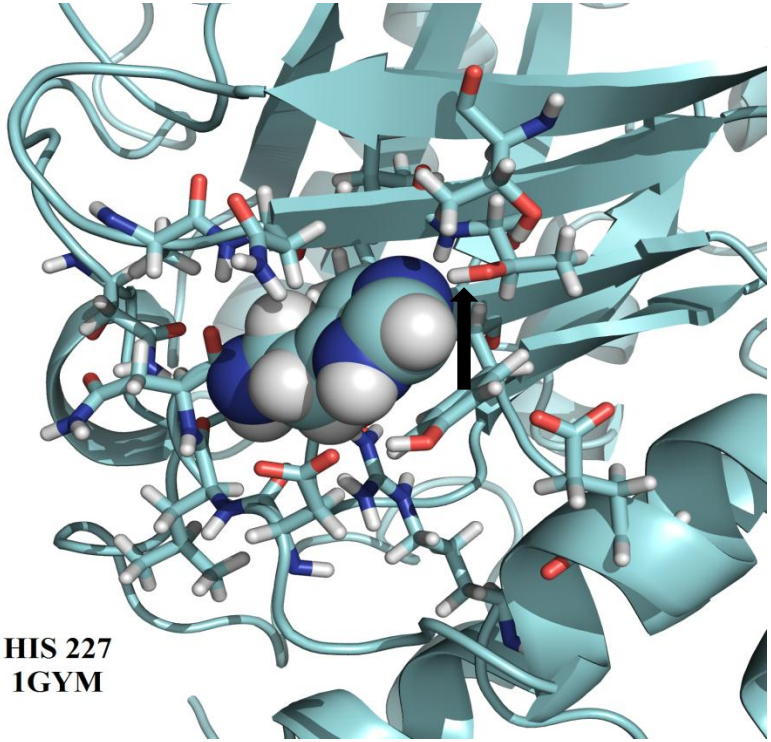








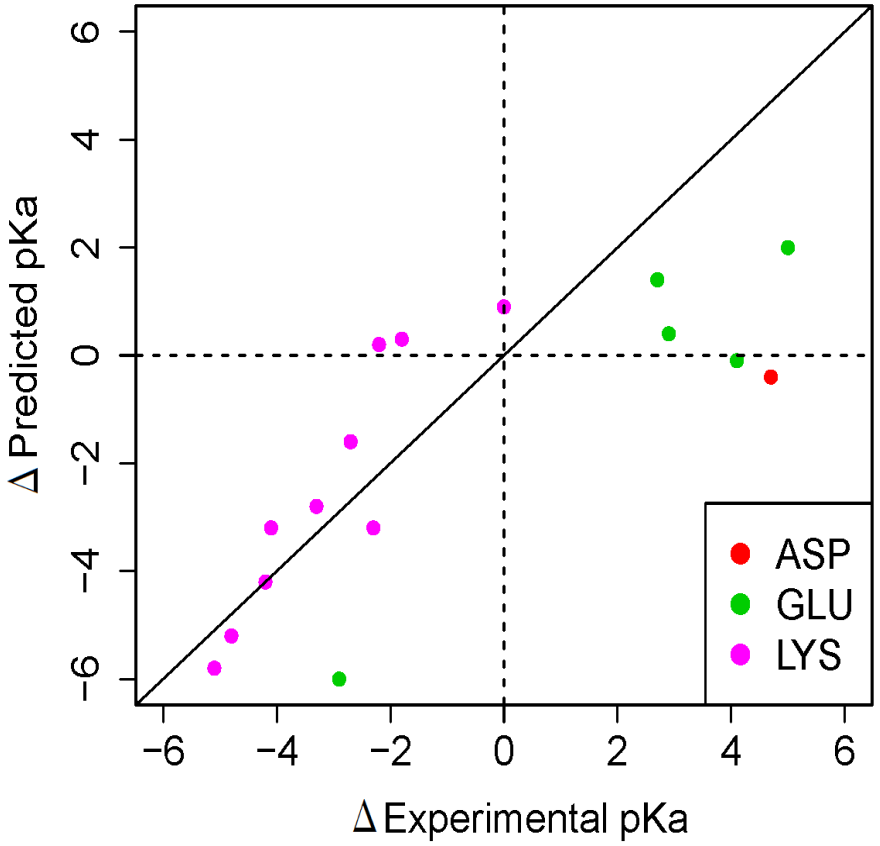
Asp–Asp interaction in RNase H captured with backbone flexibility

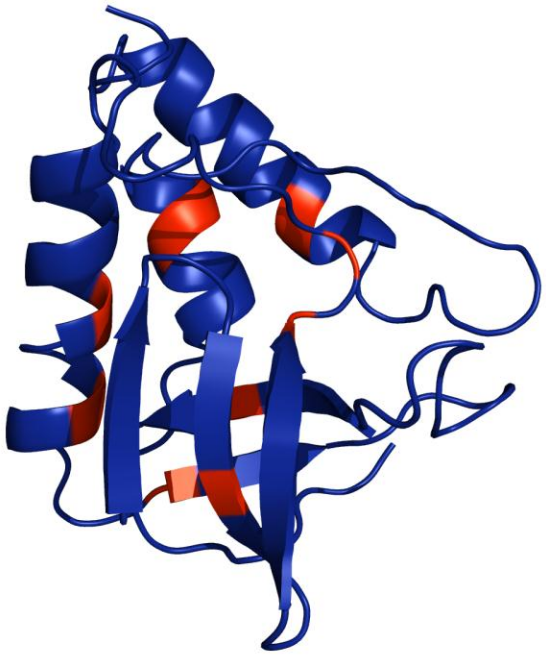


His in Phospholipase C is constrained with no room for a proton

Rosetta Accurately Predicts Some Extreme pK_a Shifts

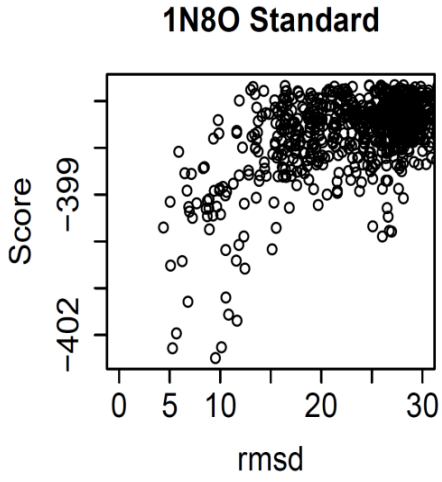
- New mutants of *Staphylococcal nuclease* with dramatic shifts up to 5 pH units (~ 7 kcal/mol)
  [Bertrand Garcia Moreno lab / JHU]

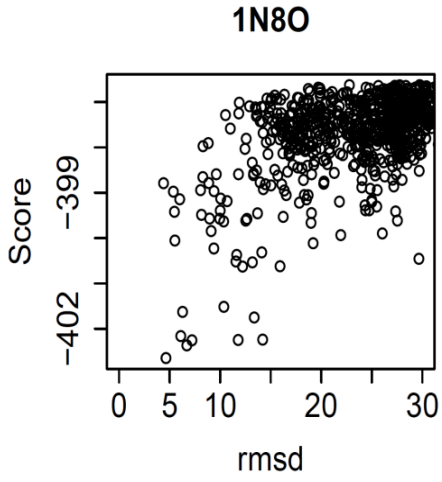


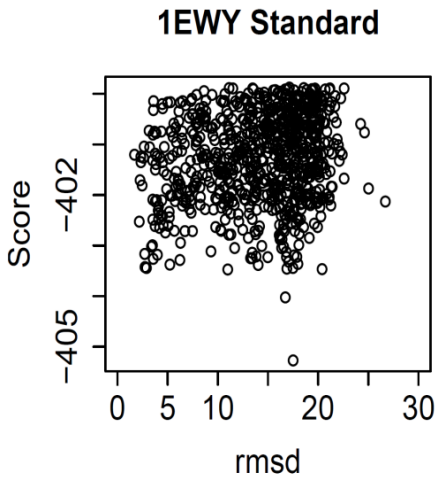


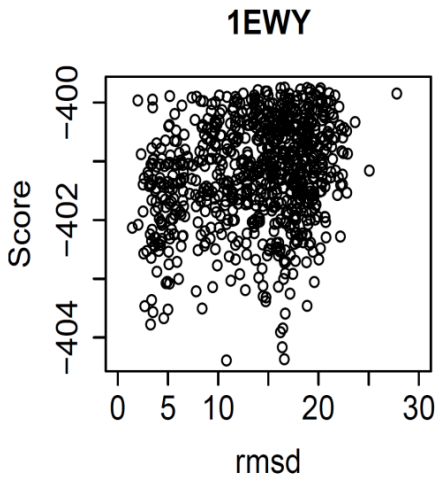
Application to protein-protein docking

- 30% of interface residues ionizable in Docking Benchmark set
- 90% of interfaces – at least one residue changes protonation state (Aguilar *et al.* 2010)
- Docking results with early score function show some improvements (work in progress)









Summary / Future Work

- Explicit electrostatic model is essential for accurate pK_a predictions
- RPH mode predicts pK_as to 0.81 RMSD, comparable to leading methods
- Higher conformational flexibility –
  1. Improves accuracy of some pK_a predictions
  2. Use of pH mode during ClassicRelax to be tested
- **pK_a prediction test can be used to benchmark score functions**
- Protonation states may improve docking, folding, and design

Acknowledgments

- Prof. Jeffrey J Gray
- Prof. Bertrand Garcia Moreno
- Ryan Harrison
- Lab members
- NIH R01 - GM 078221
- Rosetta Community

Thank You

Questions?

Supplemental Data

	Number of χ angles	*Null Model*	*RPH Function*	*Neighbor Repack*	*Ensemble Average*
Asp	2	0.95	0.77	0.97	0.8
Glu	3	0.86	0.92	1.1	0.95
His	2	0.98	0.79	0.84	0.97
Tyr	2	1.1	0.76	0.76	1.3
Lys	4	0.56	0.67	0.66	0.68
All	-	0.92	0.81	0.93	0.92

Work with Docking benchmark set 3.0

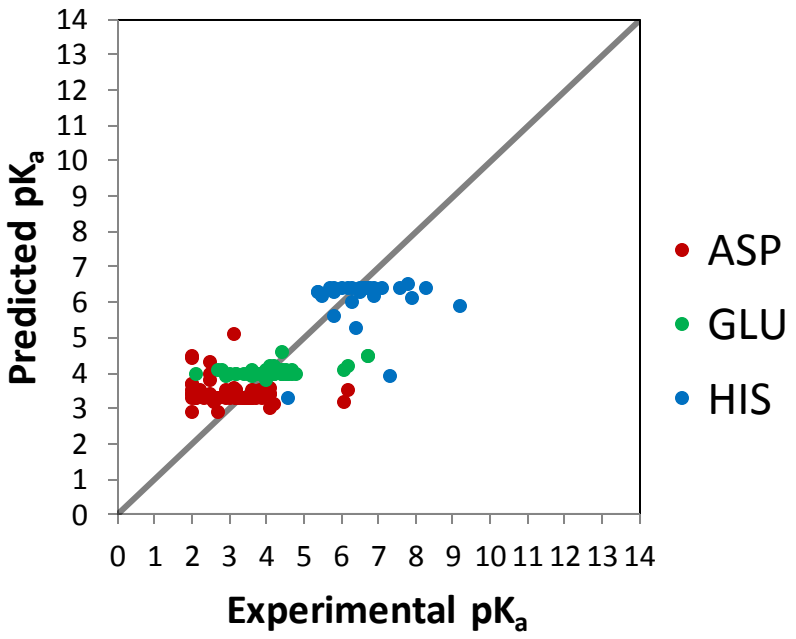
Total number of interface residues (within 5A at the interface)

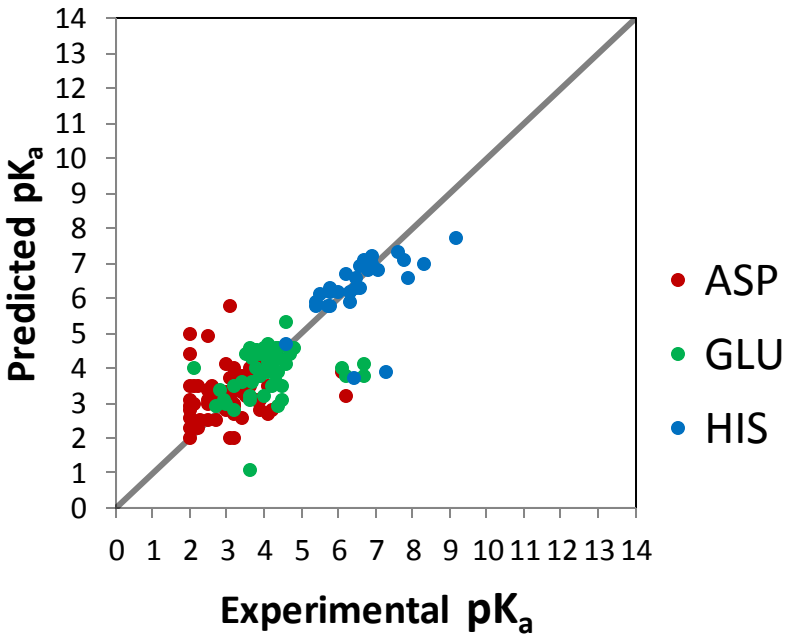
Type	All	Enz-inh	Ab-Ant	Others
ASP	6.4	6.9	6	6.2
GLU	6.8	5	5.6	8.2
HIS	2.9	4.2	2.2	2.5
TYR	6.5	6.6	11	5
LYS	6.2	4.6	4.7	7.5
Chargeable %	28.8	27.3	29.5	29.4
Non-chargeable %	71.2	72.7	70.5	70.6

pK_a shifts at interface

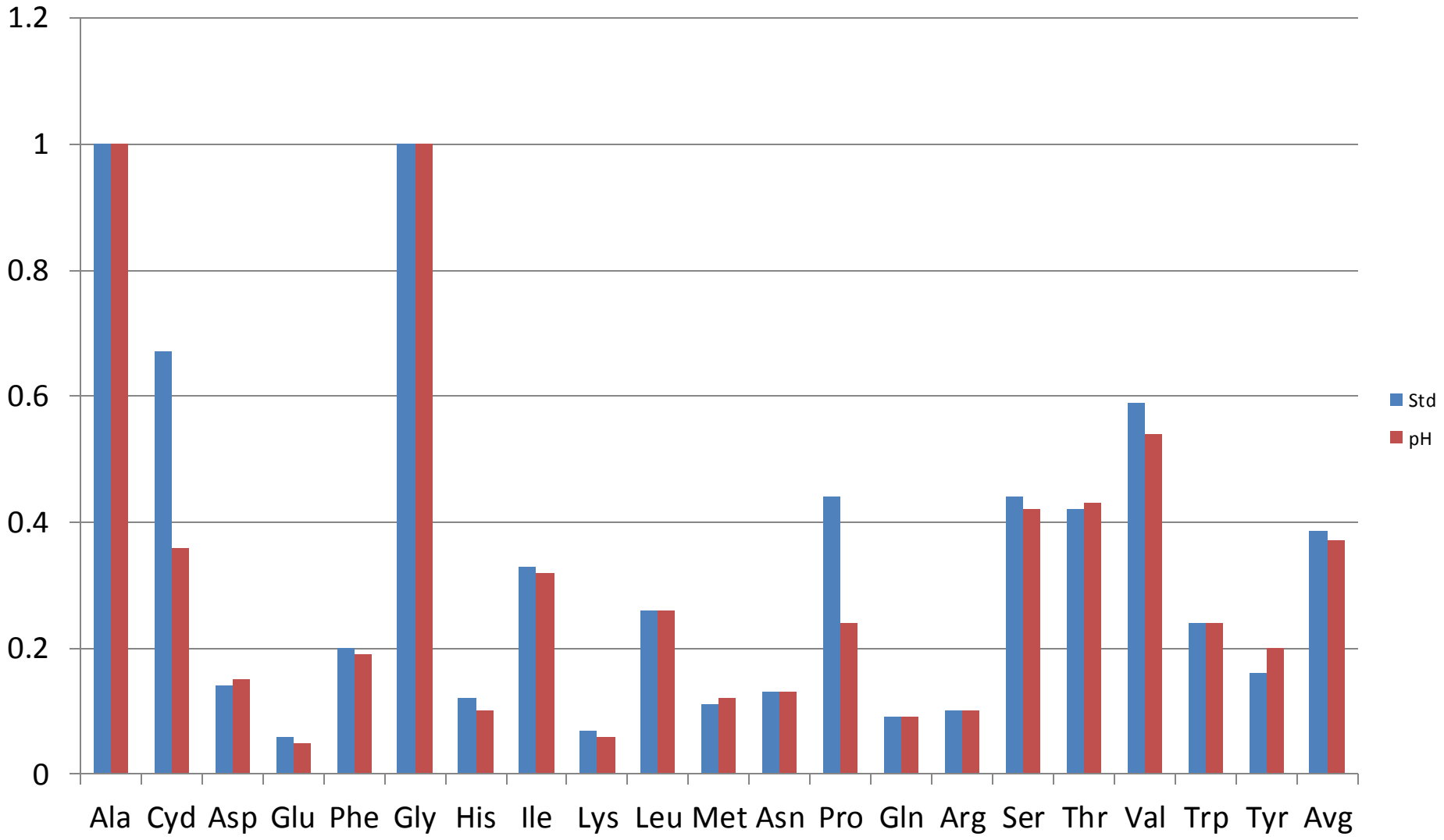
Residue	Total no	Neg_shift	Pos_shift	Avg_shift
ASP	360	233	90	0.72
GLU	381	222	132	0.67
HIS	132	71	55	0.89
TYR	328	143	152	0.69
LYS	330	90	211	0.72

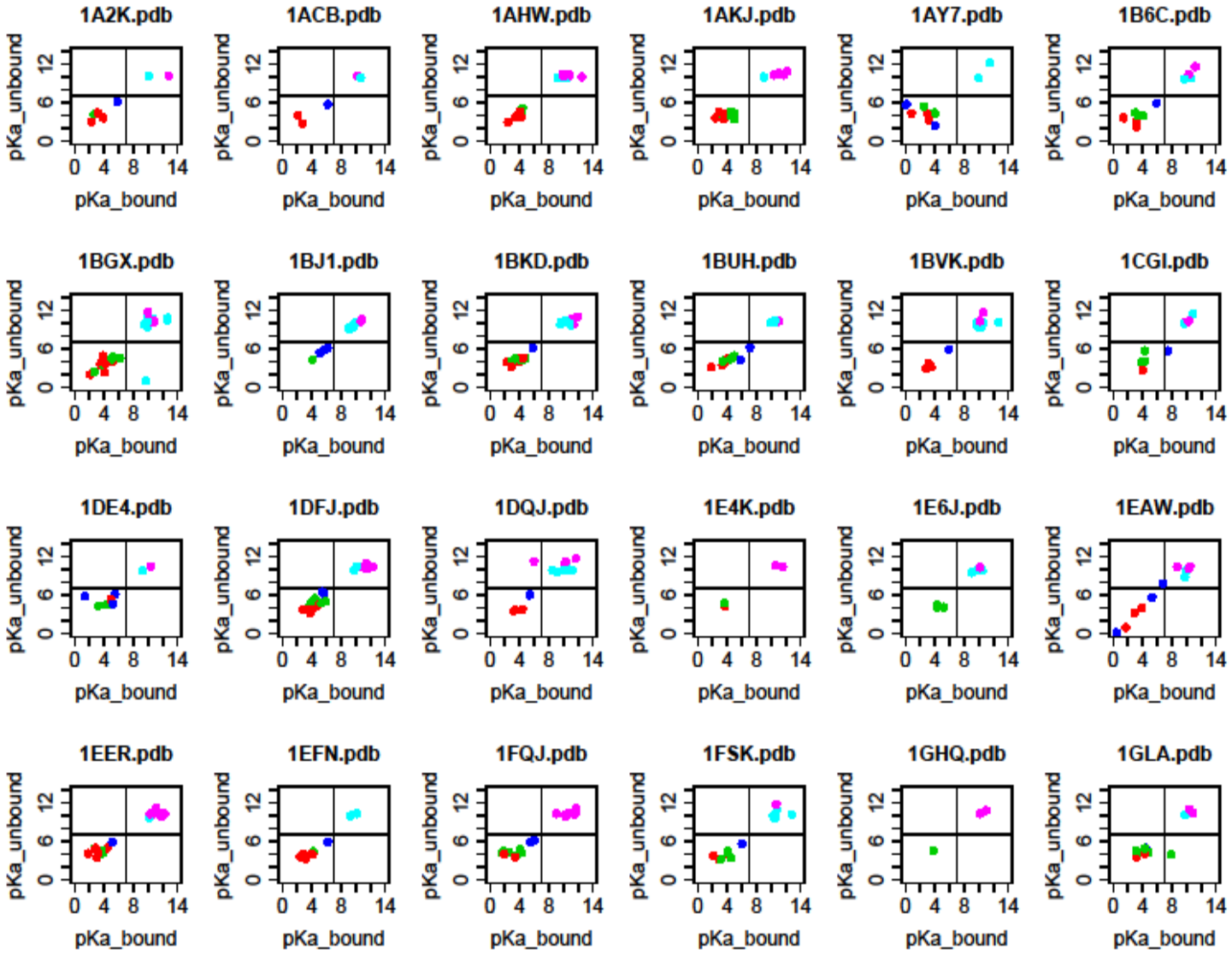
Electrostatic model

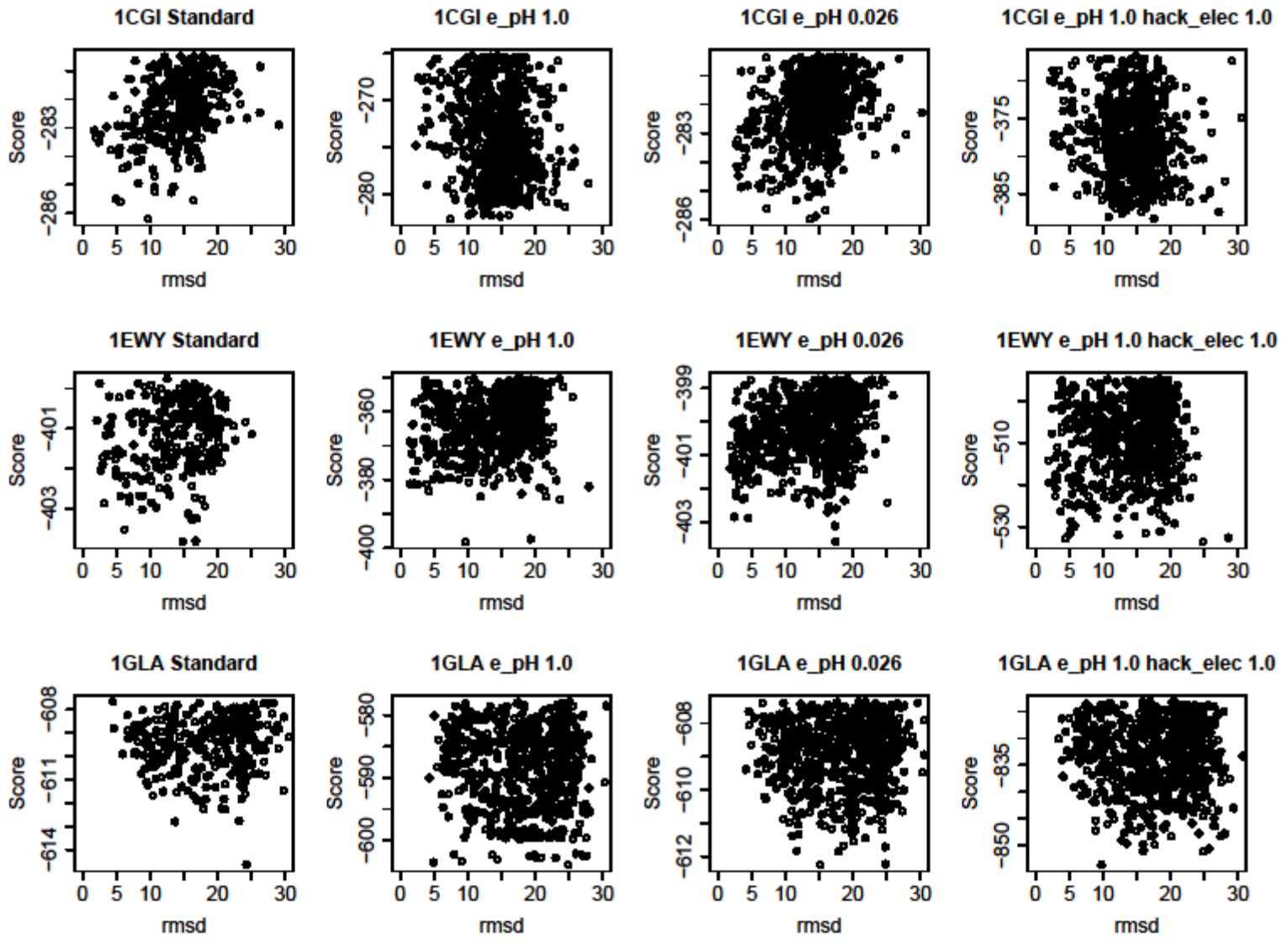


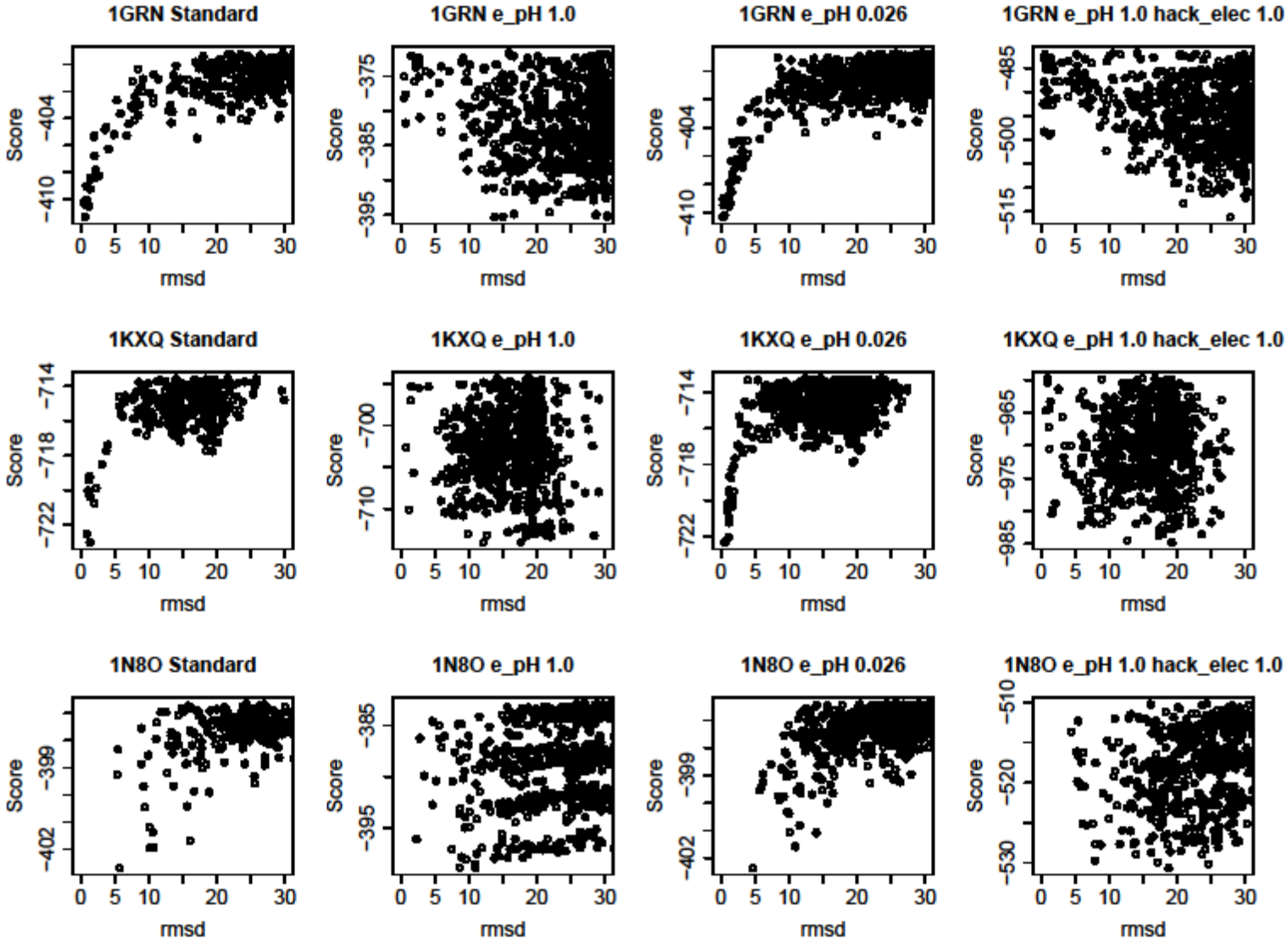


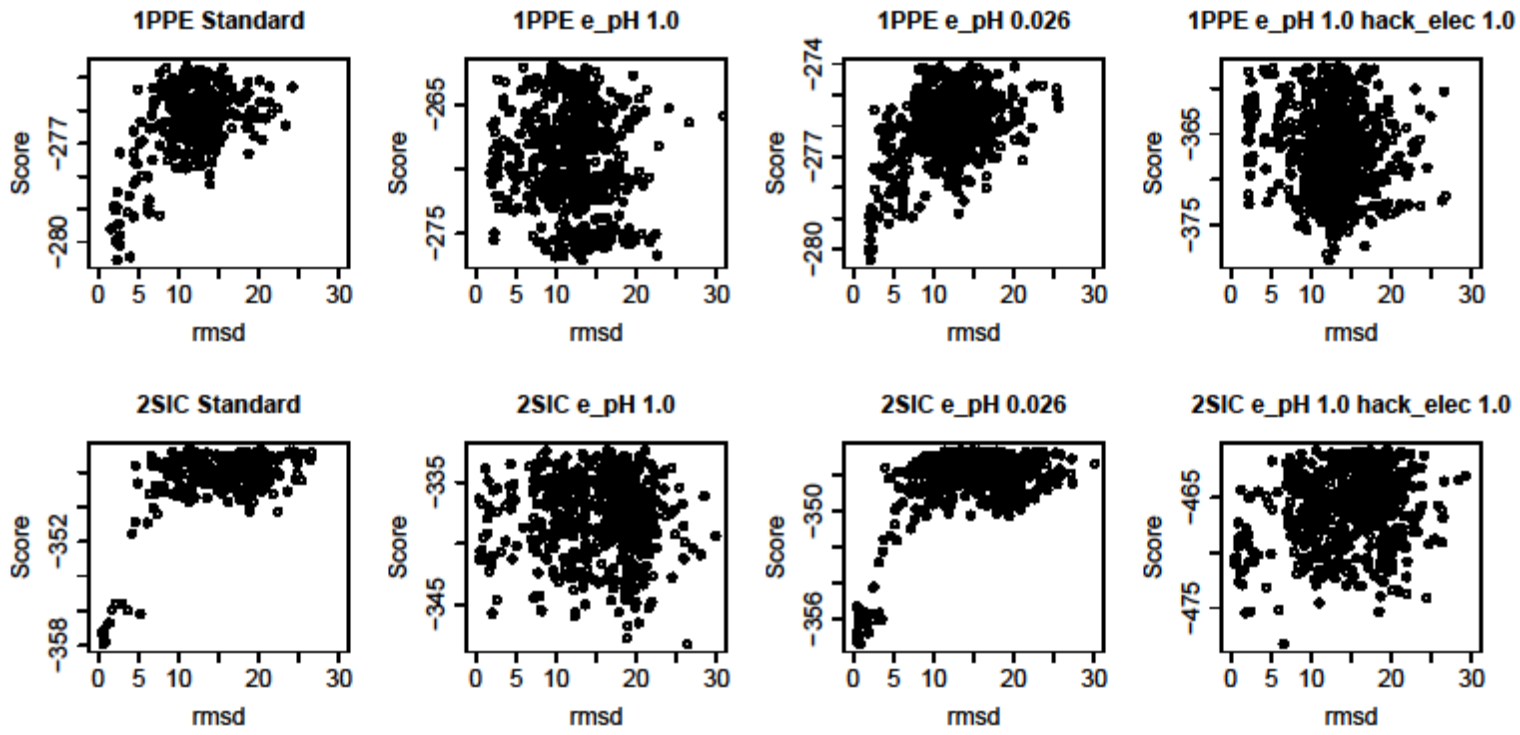
Rotamer Recovery




1A2K.pdb
1ACB.pdb
1AHW.pdb
1AKJ.pdb
1AY7.pdb
1B6C.pdb
1BGX.pdb
1BJ1.pdb
1BKD.pdb
1BUH.pdb
1BVK.pdb
1CGI.pdb
1DE4.pdb
1DFJ.pdb
1DQJ.pdb
1E4K.pdb
1E6J.pdb
1EAW.pdb
1EER.pdb
1EFN.pdb
1FQJ.pdb
1FSK.pdb
1GHQ.pdb
1GLA.pdb
pKa_unbound
pKa_bound
0
6
12
0
4
8
14


1CGI Standard
1CGI e_pH 1.0
1CGI e_pH 0.026
1CGI e_pH 1.0 hack_elec 1.0
1EWY Standard
1EWY e_pH 1.0
1EWY e_pH 0.026
1EWY e_pH 1.0 hack_elec 1.0
1GLA Standard
1GLA e_pH 1.0
1GLA e_pH 0.026
1GLA e_pH 1.0 hack_elec 1.0
Score
rmsd


1GRN Standard
1GRN e_pH 1.0
1GRN e_pH 0.026
1GRN e_pH 1.0 hack_elec 1.0
1KXQ Standard
1KXQ e_pH 1.0
1KXQ e_pH 0.026
1KXQ e_pH 1.0 hack_elec 1.0
1N8O Standard
1N8O e_pH 1.0
1N8O e_pH 0.026
1N8O e_pH 1.0 hack_elec 1.0
Score
rmsd


1PPE Standard
1PPE e_pH 1.0
1PPE e_pH 0.026
1PPE e_pH 1.0 hack_elec 1.0
2SIC Standard
2SIC e_pH 1.0
2SIC e_pH 0.026
2SIC e_pH 1.0 hack_elec 1.0
Score
rmsd

Starting Structure
(Highly Acidic Environment pH = 1)

Sample Rotamer Library For Target Residue

NO

pH = pH + ΔpH

Did protonation state change?

YES

Output pK_a

Select Lowest Energy Rotamer

Calibrate Rosetta Score Function For pK_a prediction

Target Residue

Neighbors (< 6Å)

Backbone

Conformational Flexibility