American Chemical Society


# SOMD SAMPL5 Predictions

Stefano Bosisio
University of Edinburgh
D3R Workshop
11th March 2016

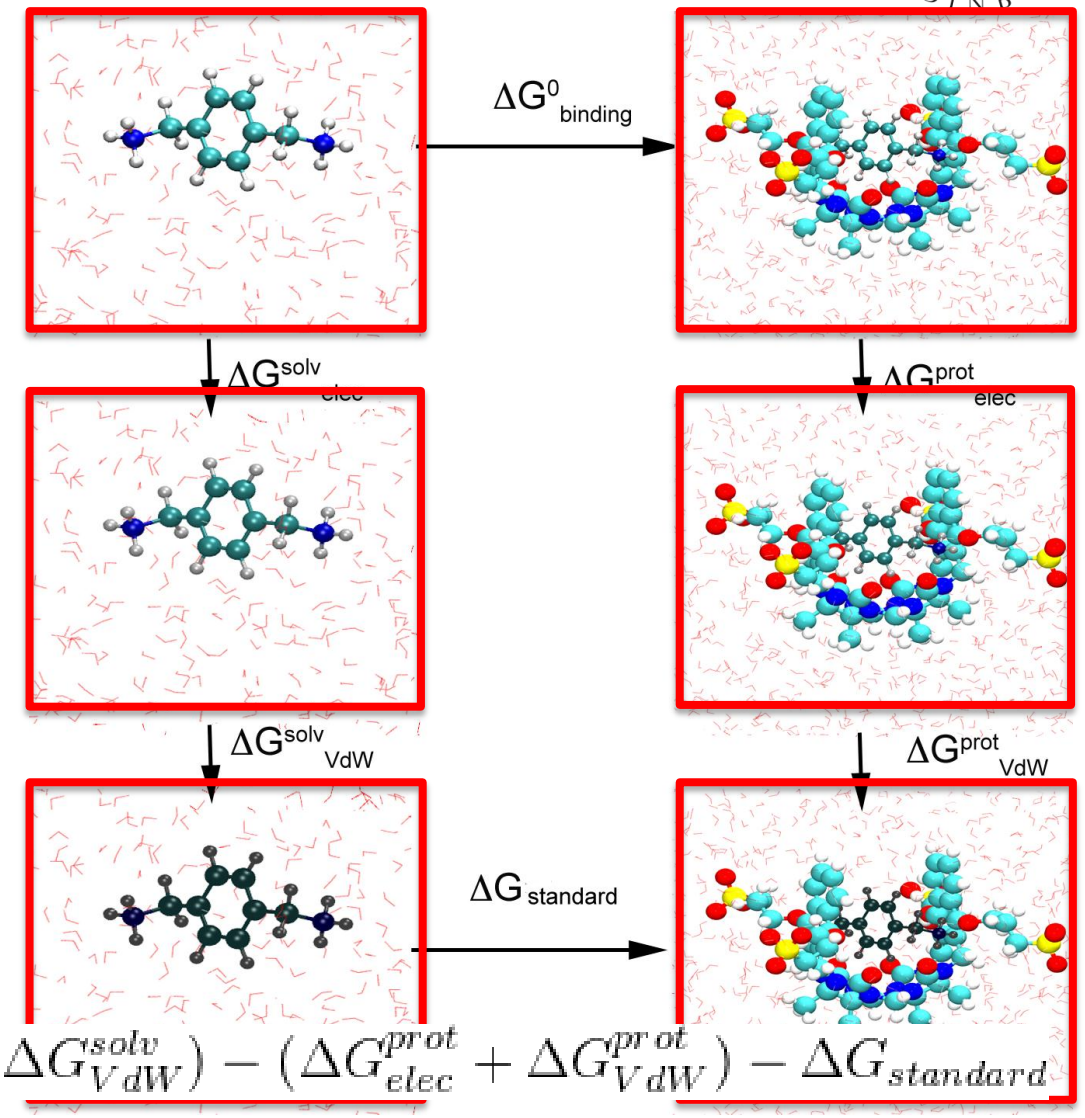
# OVERVIEW

- Host-Guest:
  - Methods: how we calculate binding free energy?
  - OAH-O1 as a reference study
  - Comparison of results for: OAH, OAMe,CBC

- Distribution Coefficient:
  - Minnesota Database as a reference study
  - Comparison with experimental results
  - Comparison with free energy estimations
  - Correction

# METHODS
## *Absolute free energy*

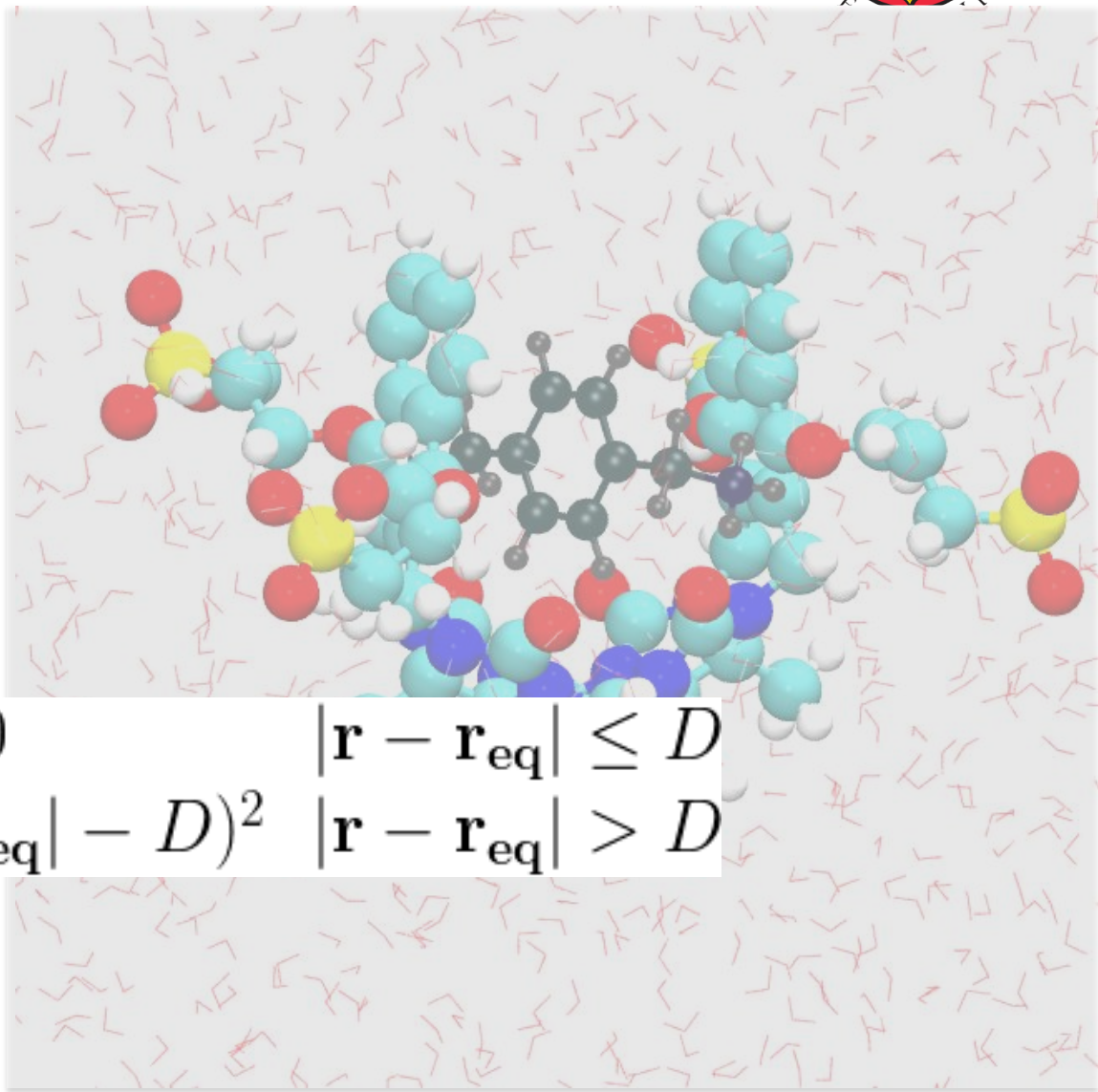
- Absolute free energy
- Double decoupling
- Single topology
- GAFF




$$\Delta G^{\circ}_{binding} = (\Delta G^{solv}_{elec} + \Delta G^{solv}_{VdW}) - (\Delta G^{prot}_{elec} + \Delta G^{prot}_{VdW}) - \Delta G_{standard}$$

# METHODS
## *Absolute free energy*

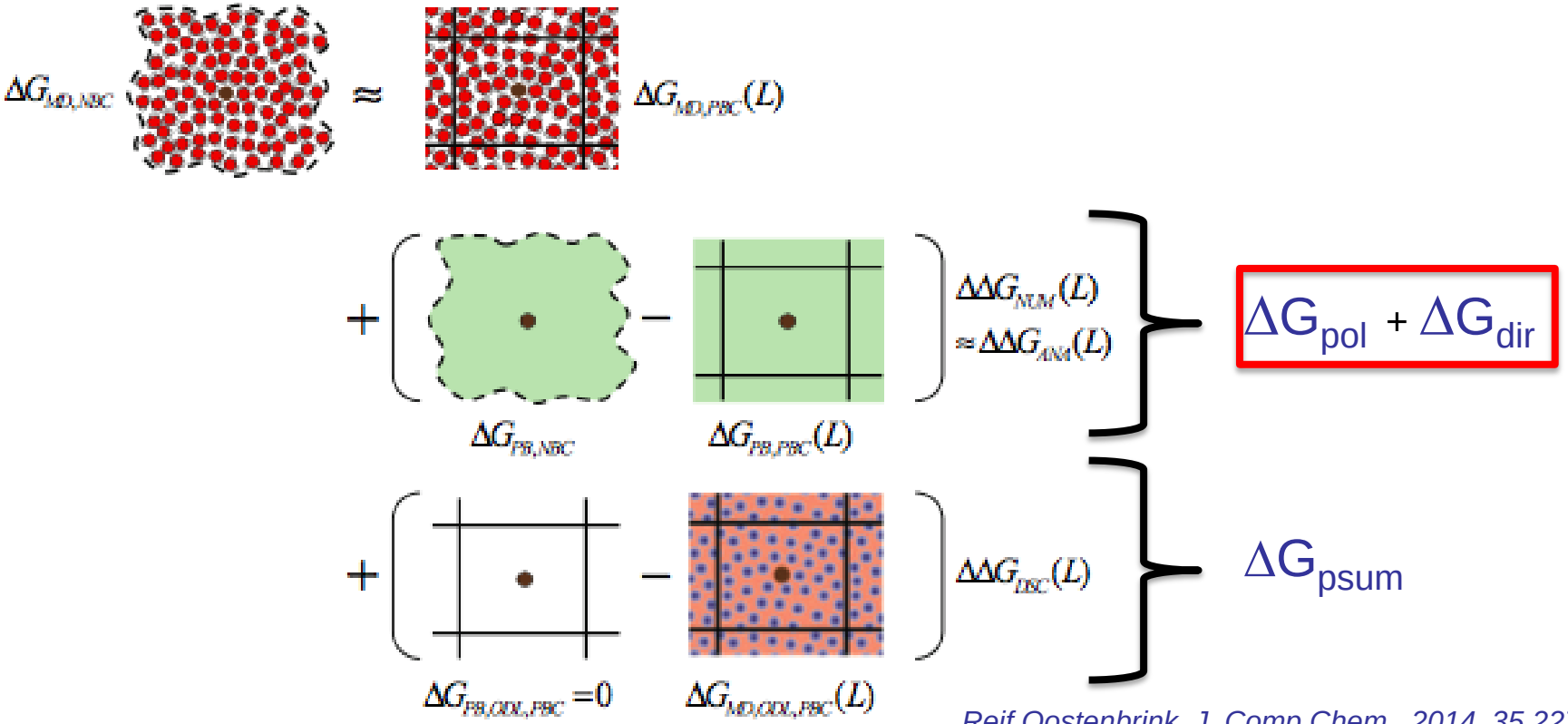
- Absolute free energy
- Double decoupling
- Single topology
- GAFF
- Flat bottom restraint:



$$U(\mathbf{r}) = \begin{cases} 0 & |\mathbf{r} - \mathbf{r_{eq}}| \leq D \\ K(|\mathbf{r} - \mathbf{r_{eq}}| - D)^2 & |\mathbf{r} - \mathbf{r_{eq}}| > D \end{cases}$$

# METHODS
## *Charging free energy*

*Rocklin et al. J. Chem. Phys, 2013, 139 184103*

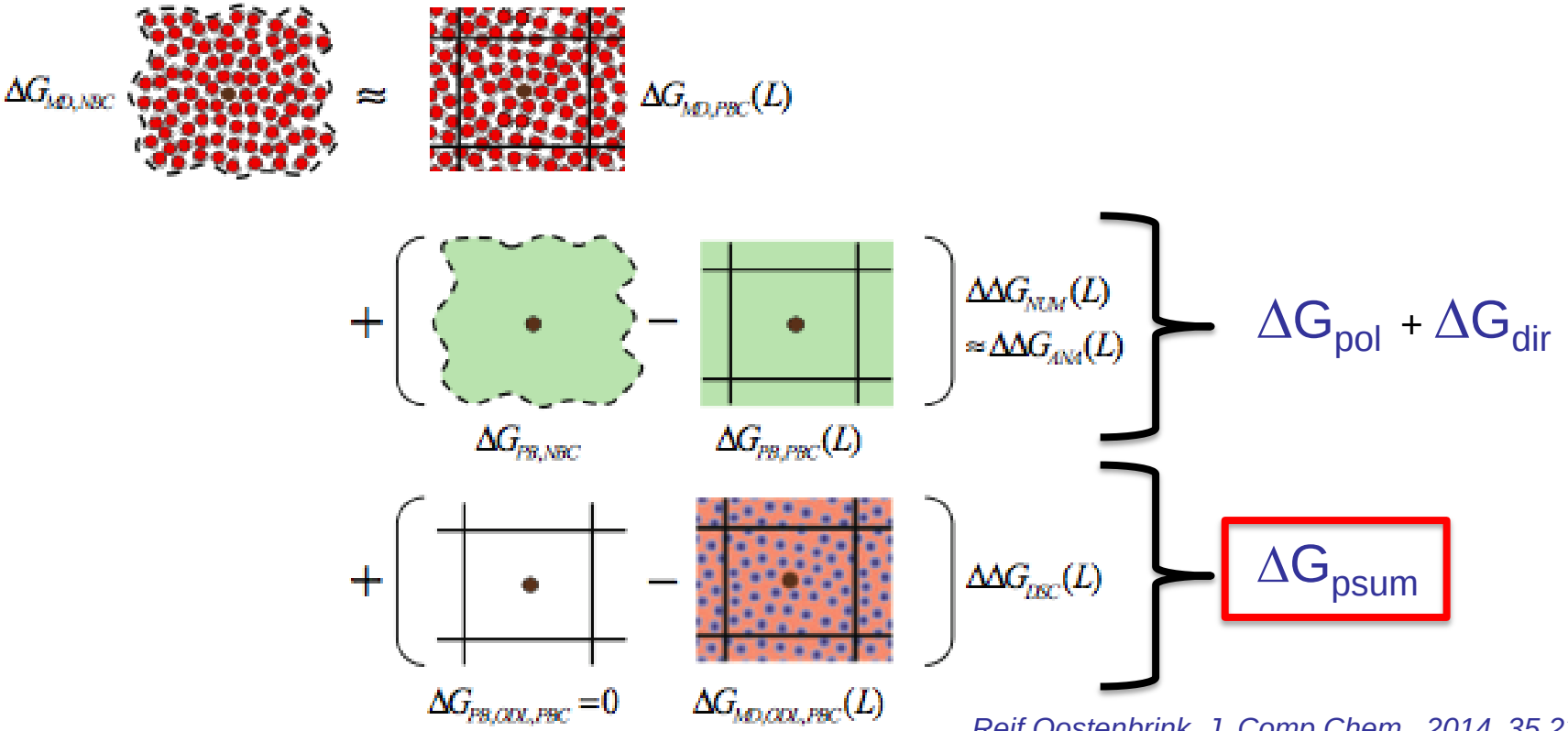



*Reif,Oostenbrink, J. Comp Chem., 2014, 35,227-243*

Reaction field Atom Based : APBS PB non periodic conditions
Custom Code given by P. Hunenberger

# METHODS
## *Charging free energy*

*Rocklin et al. J. Chem. Phys, 2013, 139 184103*

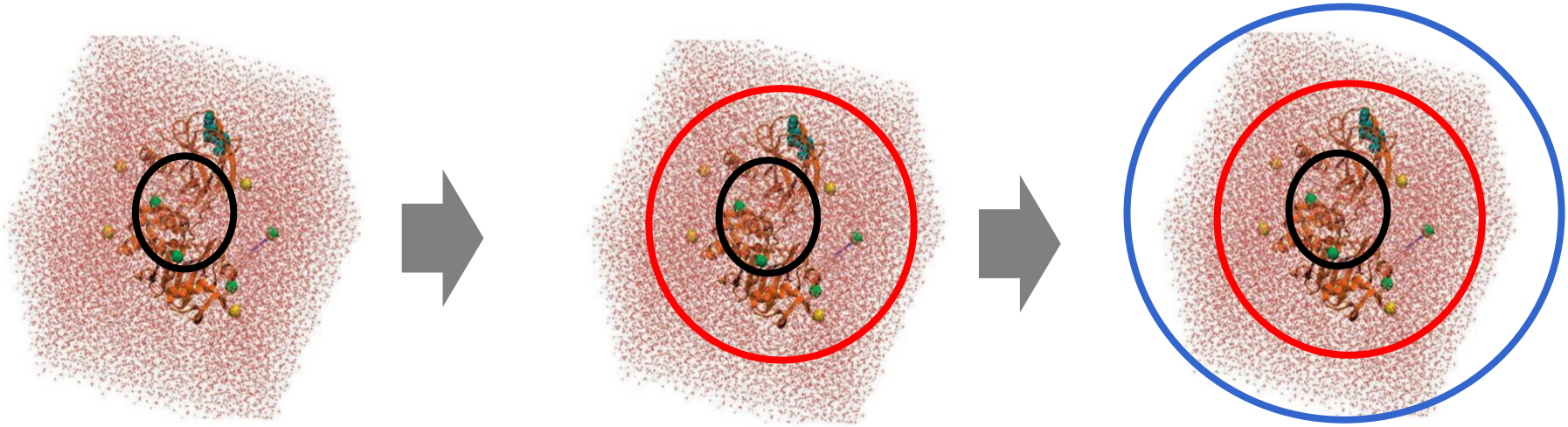



*Reif,Oostenbrink, J. Comp Chem., 2014, 35,227-243*

$$\Delta G_{psum} = -N_A(6\epsilon_0)^{-1}\left(\frac{2(\epsilon_{RF}-1)}{2\epsilon_{RF}+1}\right)\gamma_w Q_G\langle N_w(R_c)\rangle\left[\frac{4}{3}\pi R_c^3\right]^{-1}$$

# METHODS
## *Long-range dispersion correction*

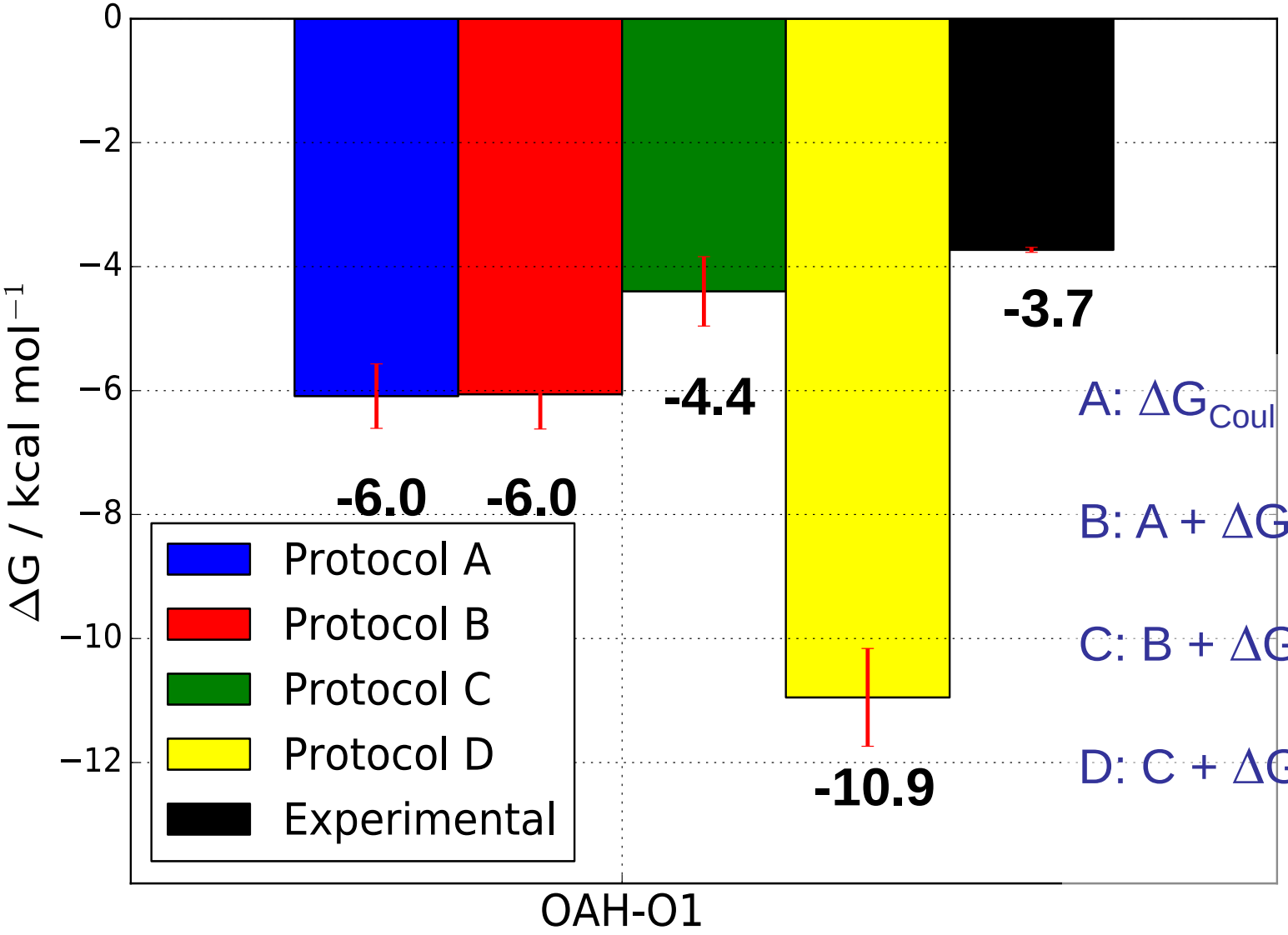
*Shirts et al. J. Phys. Chem. B, Vol. 111, No. 45, 2007*

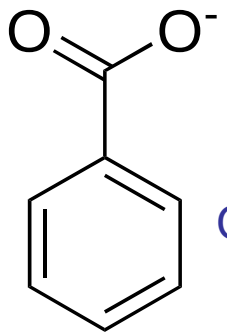


$U_{LJ,sim}$ $U_{LJ,LRC\text{-}num}$ $U_{LJ,LRC\text{-}ana}$

$$\Delta G_{LJLRC} = -\beta^{-1} \ln \langle exp(-(U_{LJ,long}(\mathbf{r}) - U_{LJ,sim}(\mathbf{r}))) \rangle_{sim}$$

# OAH-O1
## *Reference simulation*




O1

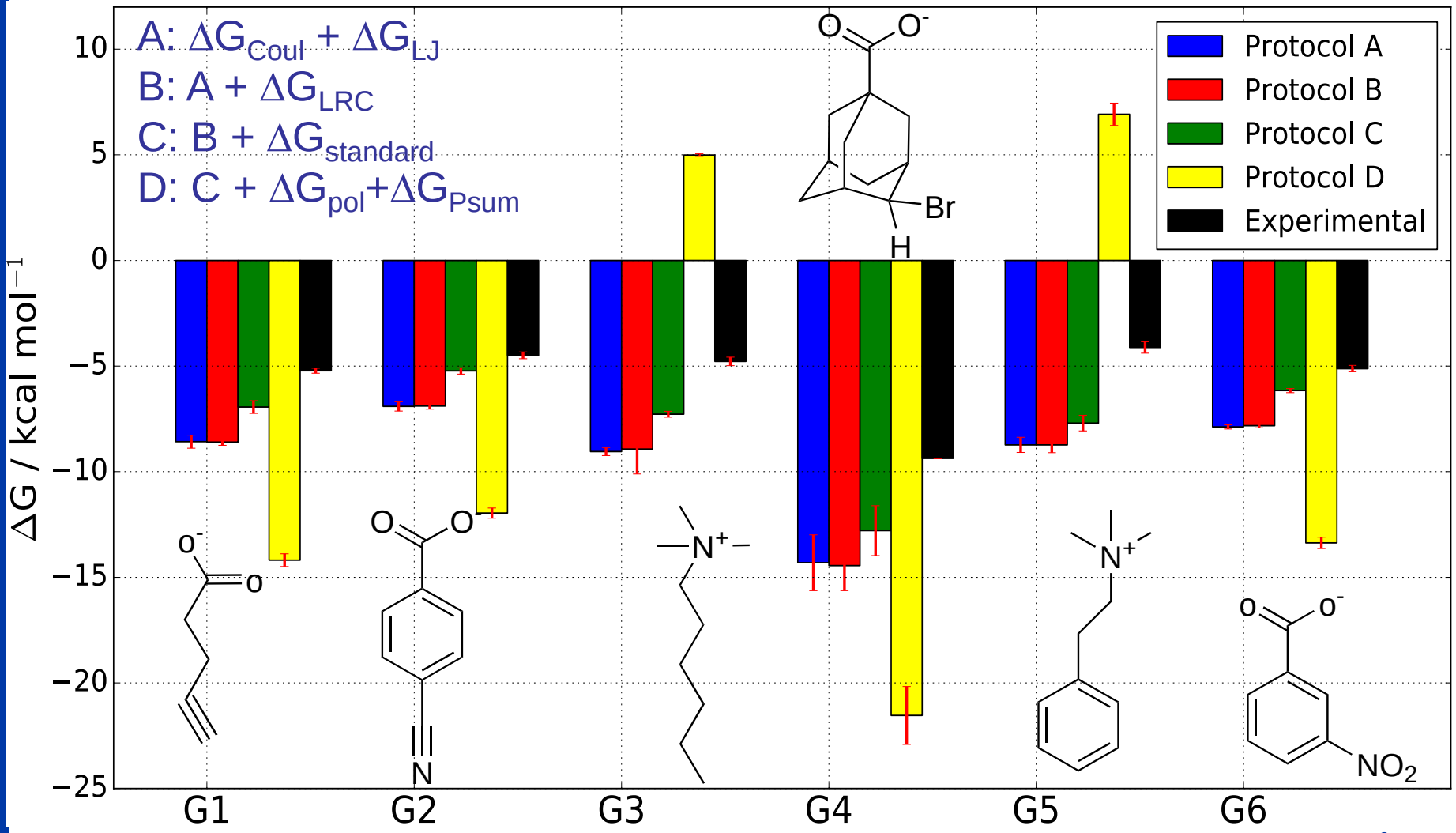
A: $\Delta G_{Coul} + \Delta G_{LJ}$

B: A + $\Delta G_{LRC}$

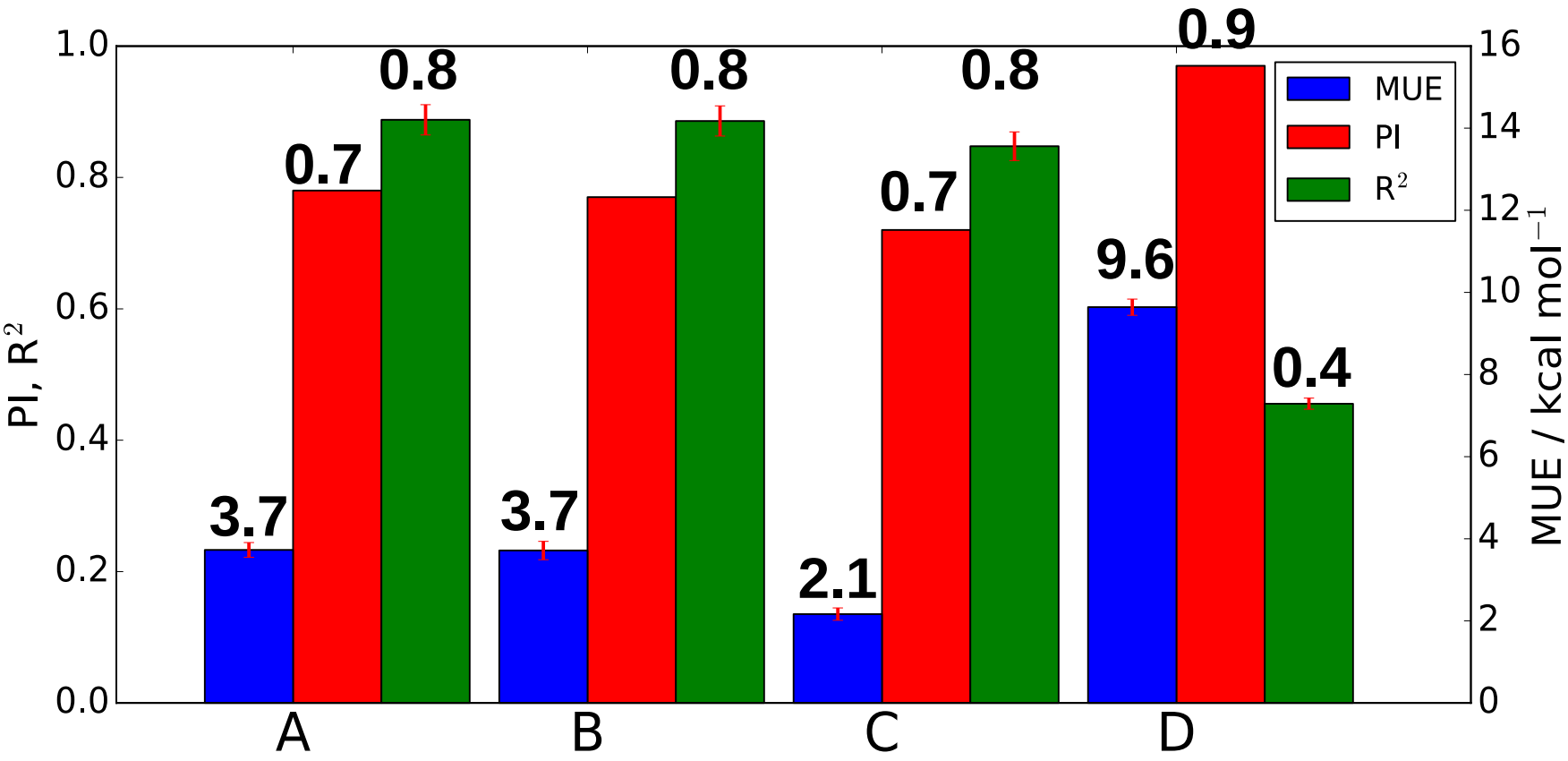
C: B + $\Delta G_{standard}$

D: C + $\Delta G_{pol} + \Delta G_{Psum}$

# OAH DATASET
## *Prediction vs experiment*

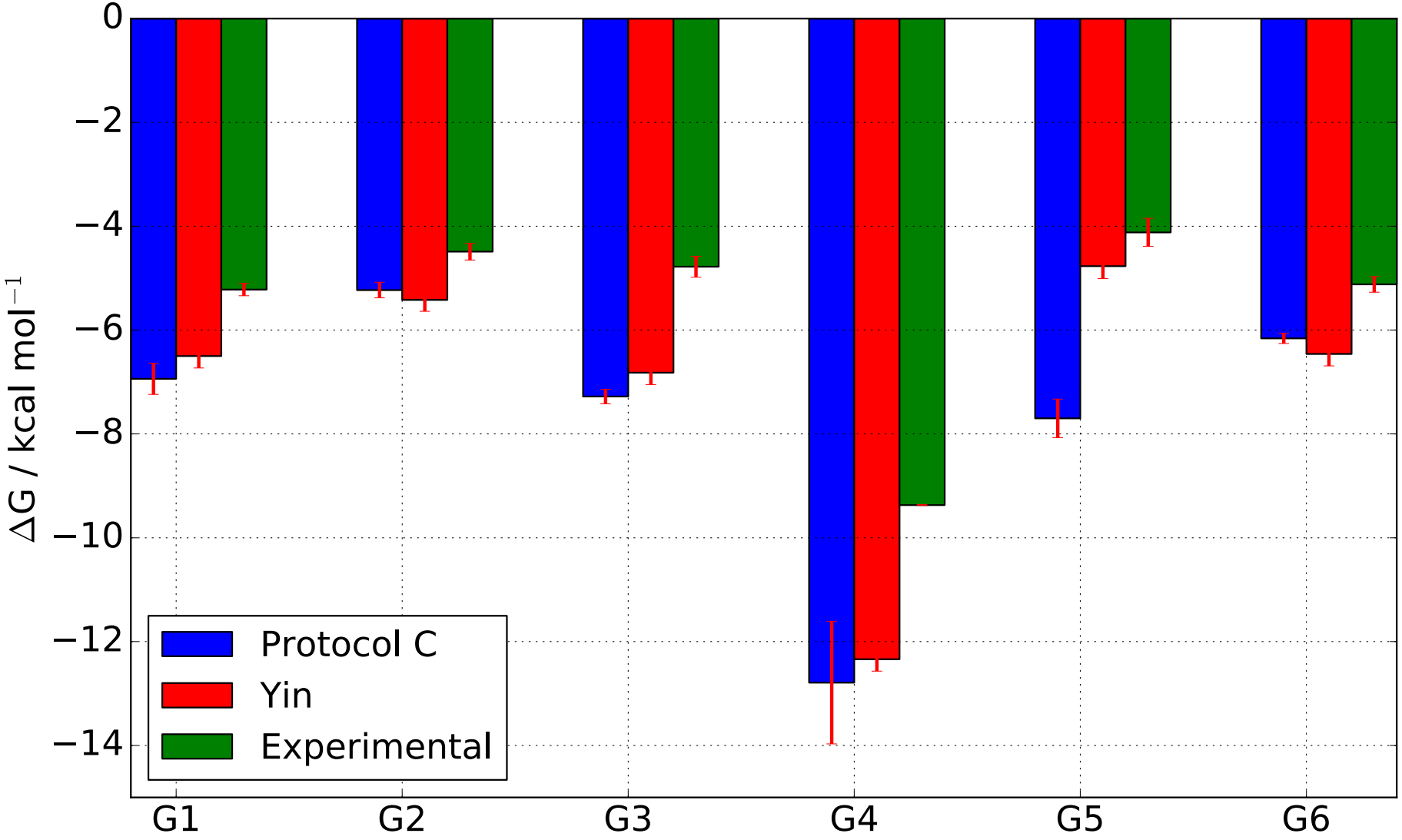


# OAH DATASET
## *Statistical analysis*

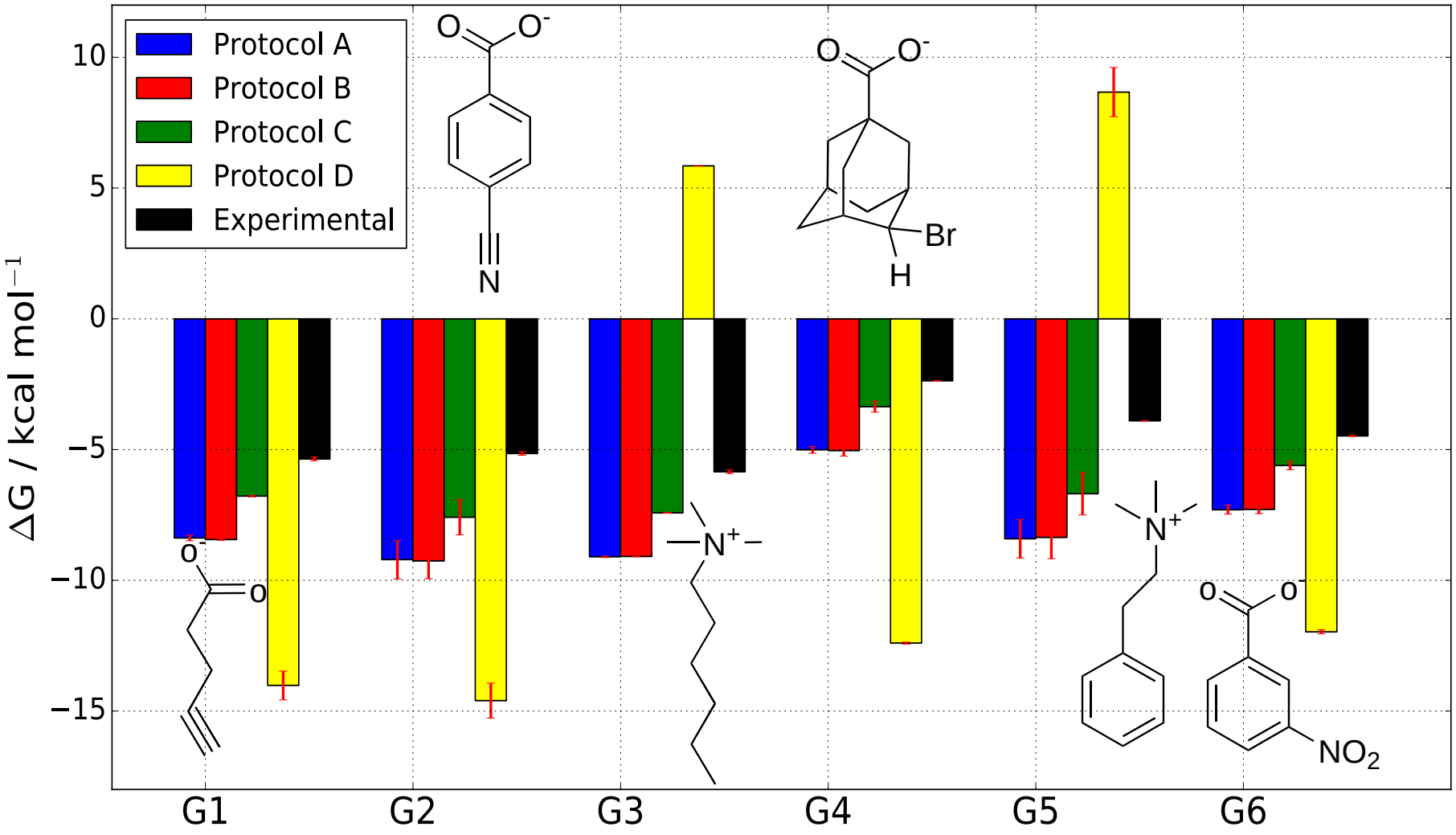



| | A | B | C | D |
|---|---|---|---|---|
| AUE: | 3.7 ± 0.4 | 3.7 ± 0.4 | 2.1 ± 0.5 | 9.6 ± 0.6 |
| R: | 0.9 ± 0.5 | 0.9 ± 0.5 | 0.9 ± 0.6 | 0.7 ± 0.2 |
| Max Err: | 2.4 ± 0.3 | 2.4 ± 0.4 | 0.6 ± 0.4 | 11 ± 2 |

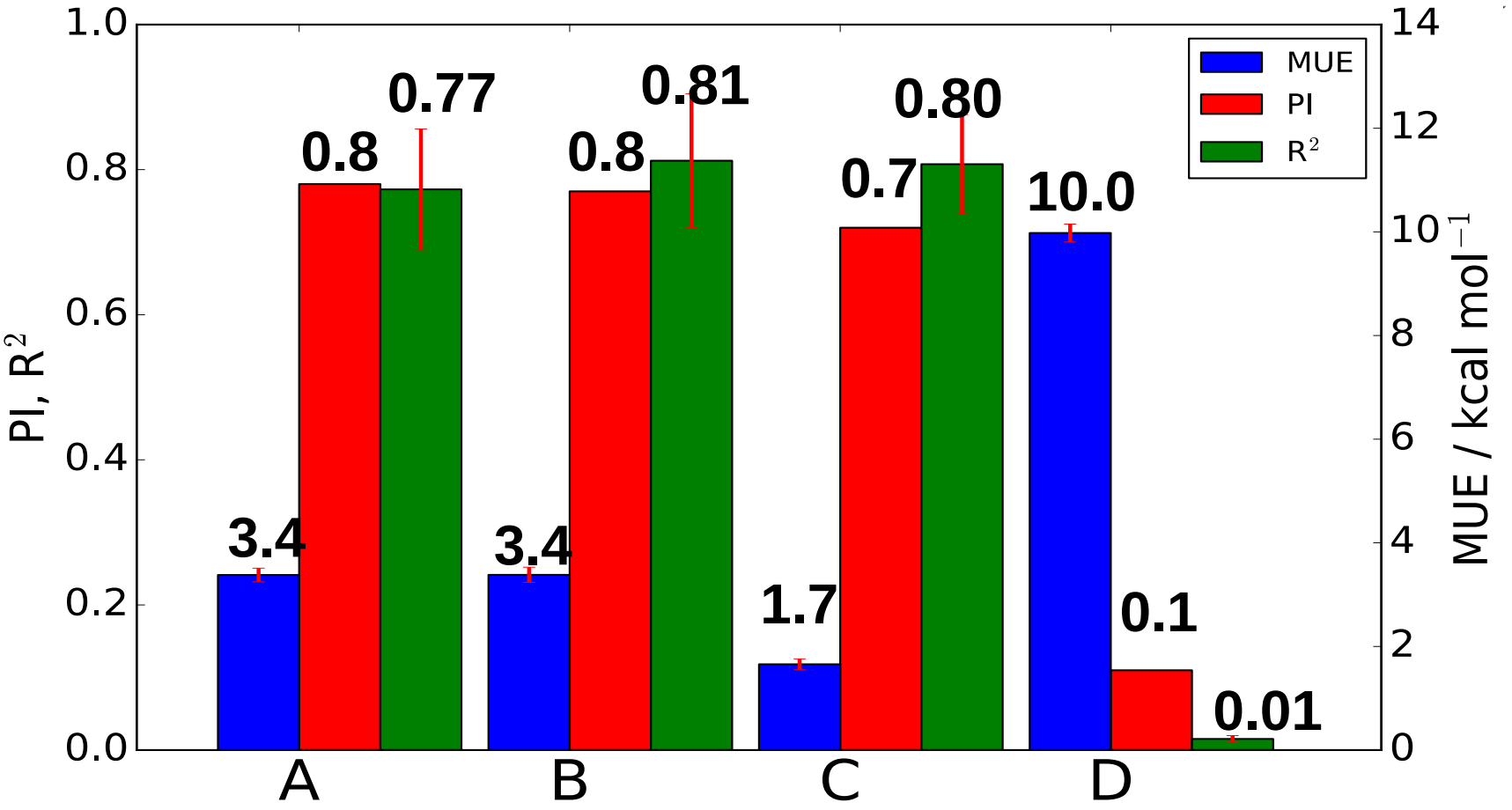
# OAH DATASET
## *Comparison with Yin's APR simulations*



# OAME DATASET
## *Prediction and experiment*

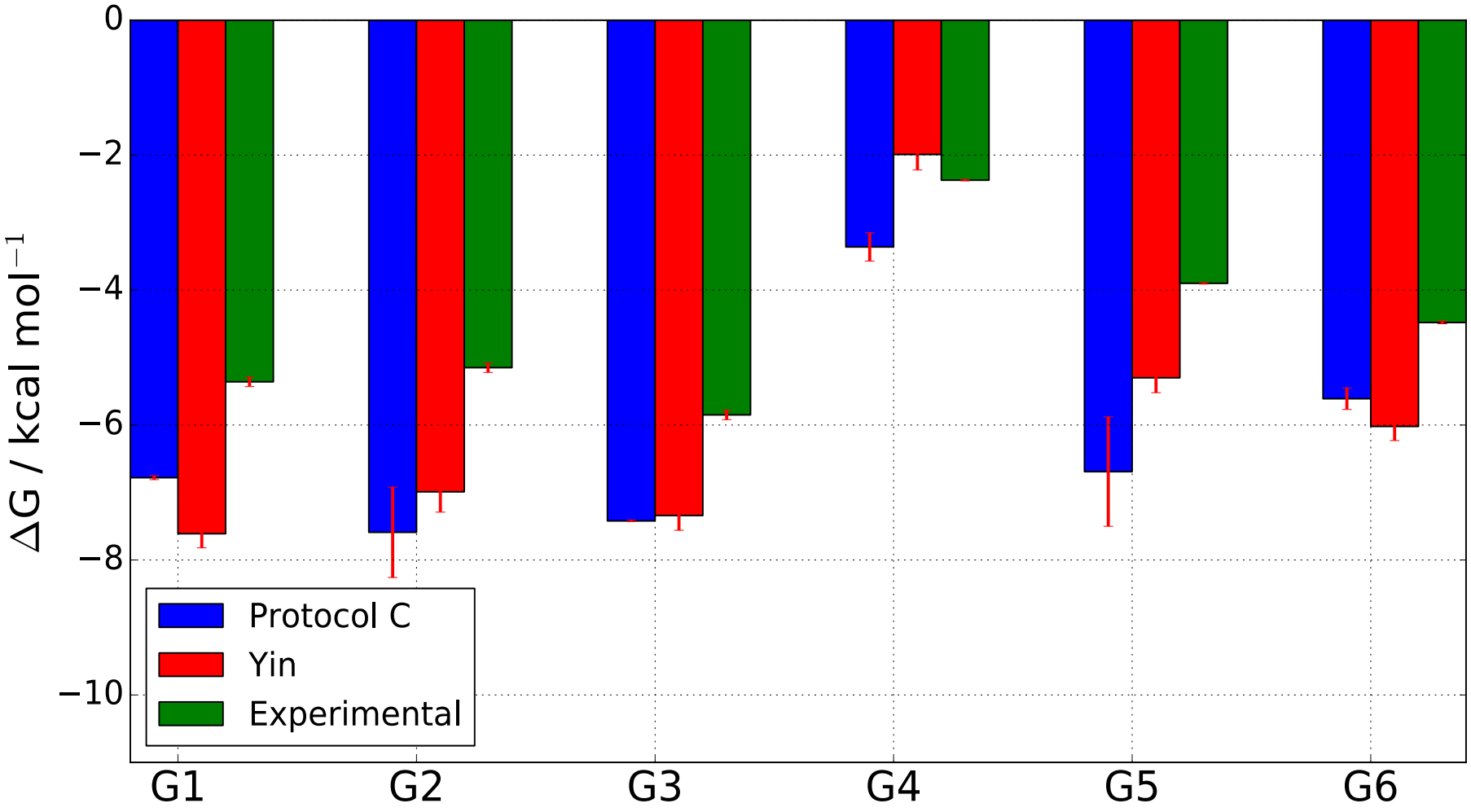


# OAME DATASET
## *Statistical analysis*

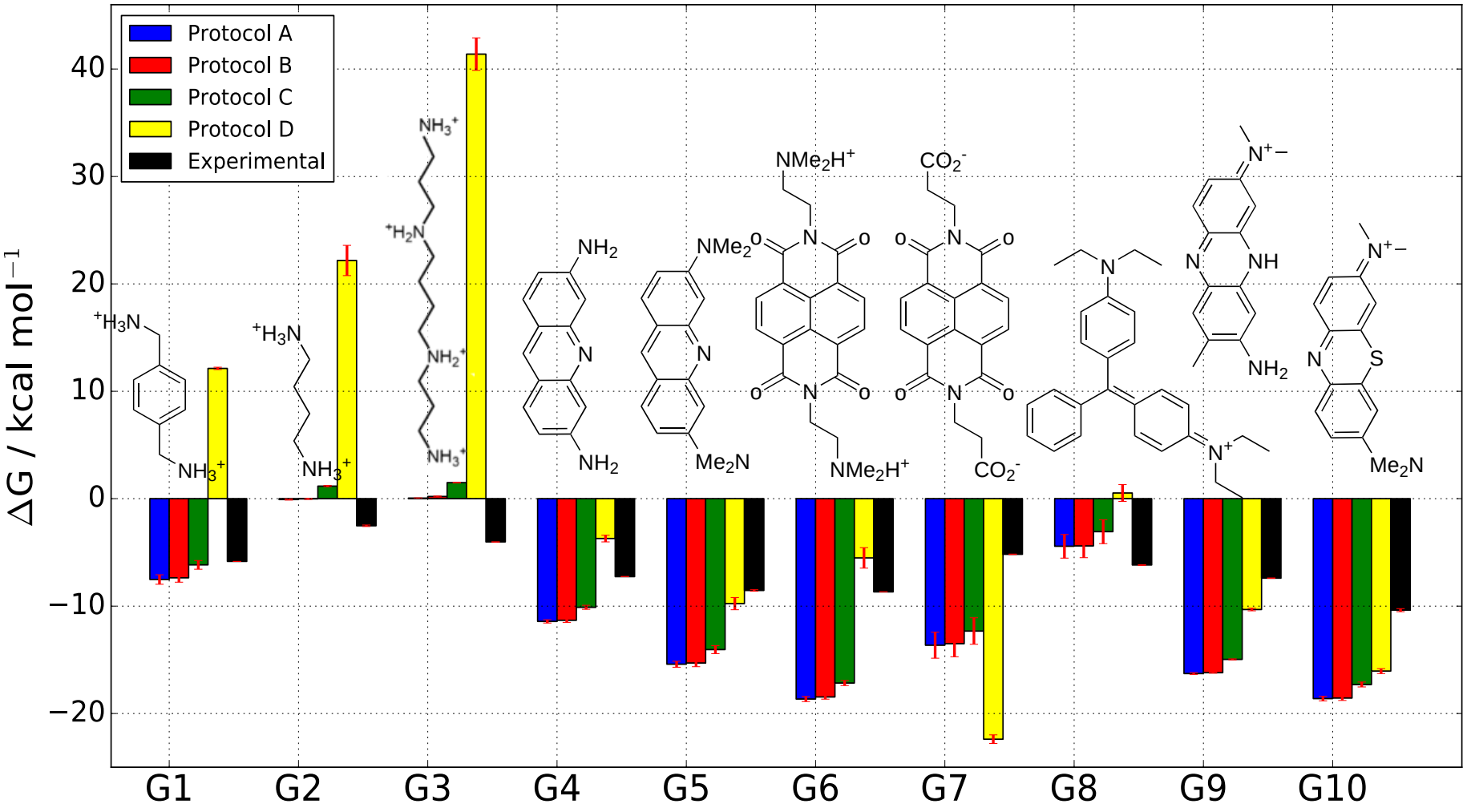



| | A | B | C | D |
|---|---|---|---|---|
| AUE: | 3.4 ± 0.3 | 3.4 ± 0.3 | 1.7 ± 0.3 | 10.0 ± 0.7 |
| R: | 0.9 ± 0.3 | 0.9 ± 0.3 | 0.9 ± 0.2 | -0.1 ± 0.5 |
| Max Err: | 2.6 ± 0.4 | 2.7 ± 0.4 | 0.1 ± 0.4 | 12.6 ± 0.5 |

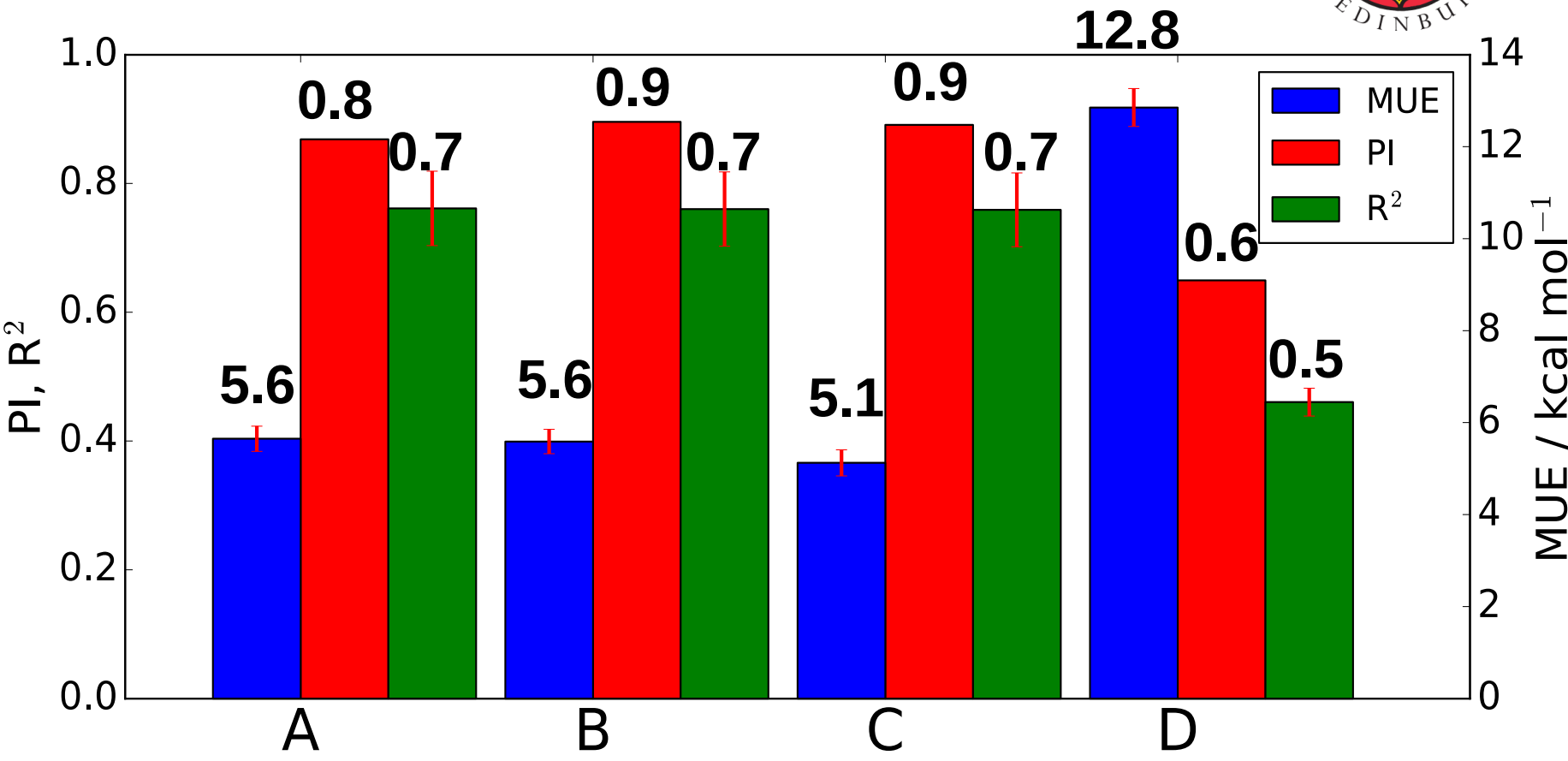
# OAME DATASET
## *Comparison with Yin's APR simulations*



# CBC DATASET
## *Prediction and experiment*

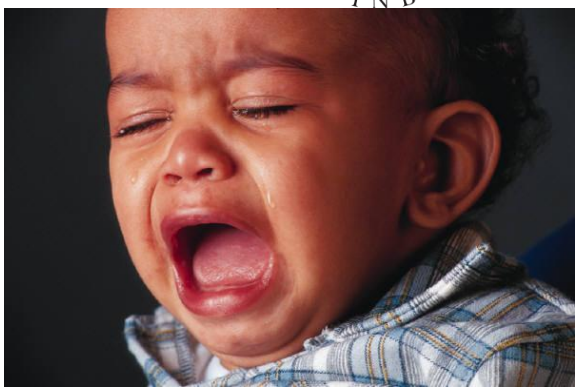


# CBC DATASET
## *Statistical Analysis*




| | A | B | C | D |
|---|---|---|---|---|
| AUE: | 5.7 ± 0.9 | 6.0 ± 0.1 | 5.1 ± 0.8 | 13.0 ± 4.0 |
| R: | 0.9 ± 0.1 | 0.9 ± 0.1 | 0.9 ± 0.1 | 0.7 ± 0.3 |
| Max Err: | 4.1 ± 0.6 | 4.2 ± 0.6 | 5.5 ± 0.6 | 45.0 ± 6.0 |

# HOST-GUEST
## *What we have learned*



- Charging corrections: CHECK THE CODE
- Long range corrections small contribution
- All protocols, except D, present a good $R^2$
- Reasonable reproducibility with other methods

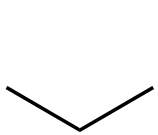
# DISTRIBUTION COEFFICIENT
## *Minnesota database*

- Absolute free energy of solvation
  $\Delta G_{HYD}$ and $\Delta G_{CYC}$
- GAFF/AM1BCC
- Protocols:
  A: $\Delta G_{Coul} + \Delta G_{LJ}$

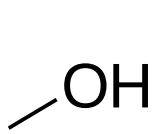
  B: A + $\Delta G_{LJ\text{-}LRC}$

  C: A + $\Delta G_{charge}$

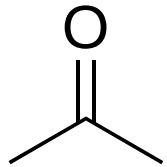
  D: B + C



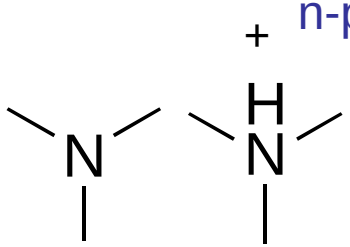
n-propane



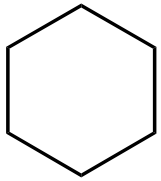

methanol



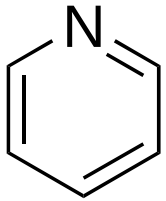

acetone



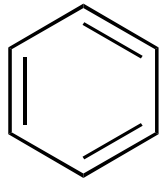

Trimethylamine



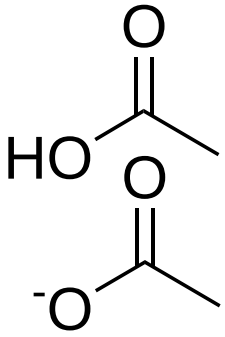
Cyclohexane



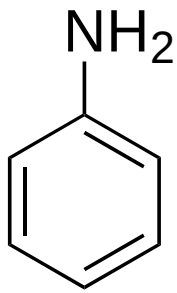

Pyridine



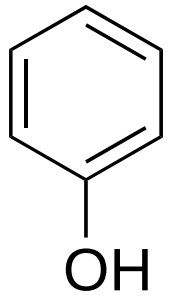
Benzene



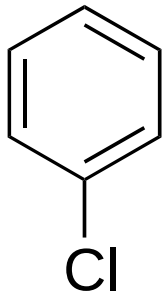

Acetic Acid




Aniline

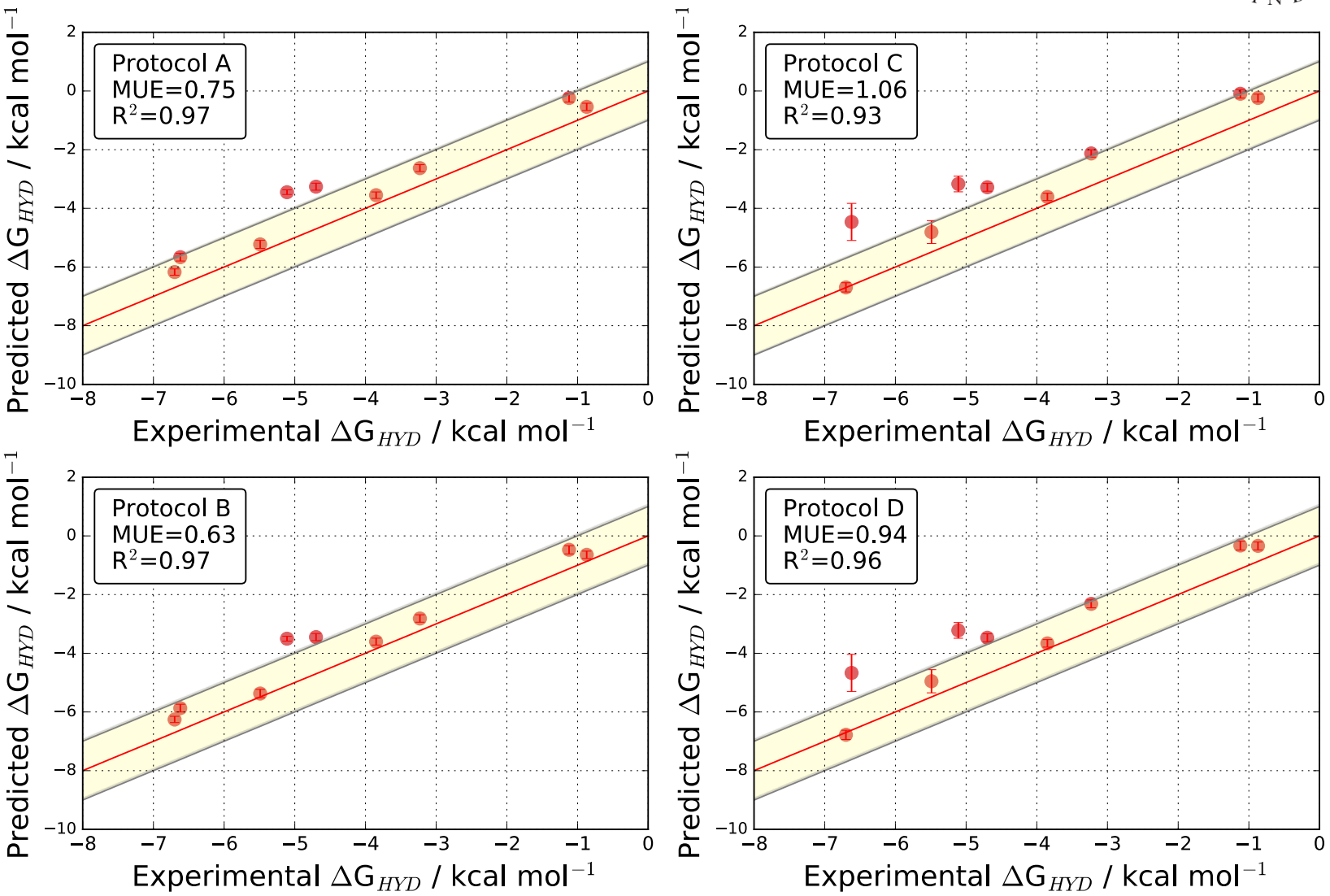



Phenol




Cl-Benzene

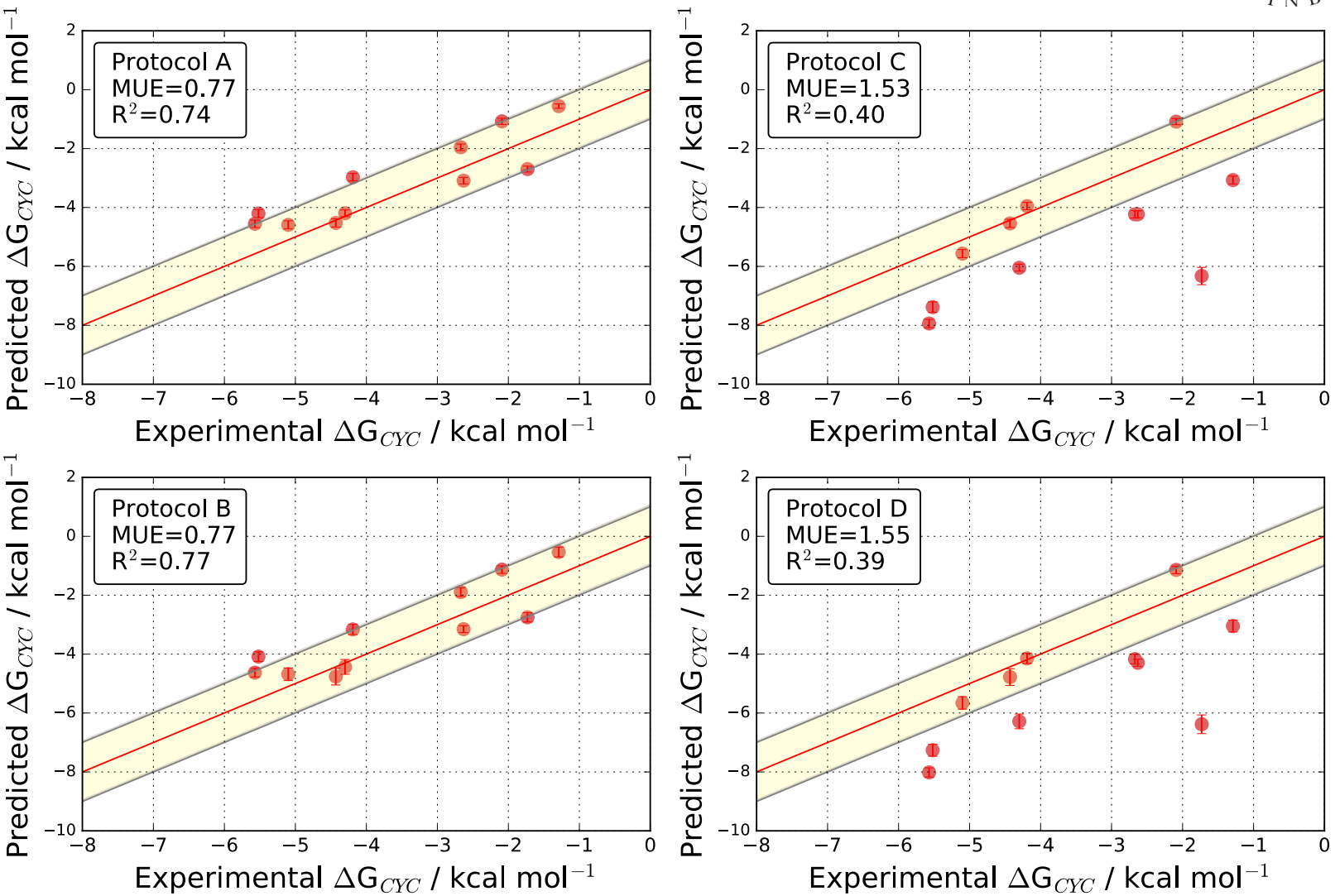
# DISTRIBUTION COEFFICIENT
## *Minnesota database*



# DISTRIBUTION COEFFICIENT
## *Minnesota database*

| Molecule | $\Delta G_{HYD}$ EXP | $\Delta G_{HYD}$ A | $\Delta G_{HYD}$ B | $\Delta G_{HYD}$ C | $\Delta G_{HYD}$ D |
|---|---|---|---|---|---|
| benzene | -0.9 | -0.7 | -0.9 | -0.5 | -0.6 |
| cyclohexane | 1.2 | 1.2 | 1.1 | 1.2 | 1.1 |
| trimethylamine_positive | -61.4 | -24.7 | -24.8 | -61.9 | -62.0 |
| acetic_acid_neutral | -6.7 | -5.8 | -6.0 | -5.4 | -5.6 |
| trimethylamine_neutral | -3.2 | -2.9 | -3.0 | -2.6 | -2.7 |
| chlorobenzene | -1.1 | -0.4 | -0.6 | -0.3 | -0.5 |
| methanol | -5.1 | -3.3 | -3.4 | -3.1 | -3.2 |
| n-propane | 2.0 | 2.4 | 2.4 | 2.5 | 2.4 |
| pyridine | -4.7 | -2.9 | -3.1 | -2.8 | -3.0 |
| phenol | -6.6 | -5.2 | -5.4 | -4.0 | -4.1 |
| acetic_acid_negative | -77.6 | -74.5 | -74.6 | -81.0 | -81.1 |
| acetone | -3.9 | -3.4 | -3.6 | -3.4 | -3.6 |
| aniline | -5.5 | -5.2 | -5.4 | -4.4 | -4.6 |
| MUE_all | | 3.70 | 3.58 | 1.20 | 1.11 |
| MUE_neutral | | 0.75 | 0.63 | 1.06 | 0.94 |

# DISTRIBUTION COEFFICIENT
## *Minnesota database*



# DISTRIBUTION COEFFICIENT
## *Protocols setup*

A: $\Delta G_{Coul} + \Delta G_{LJ}$

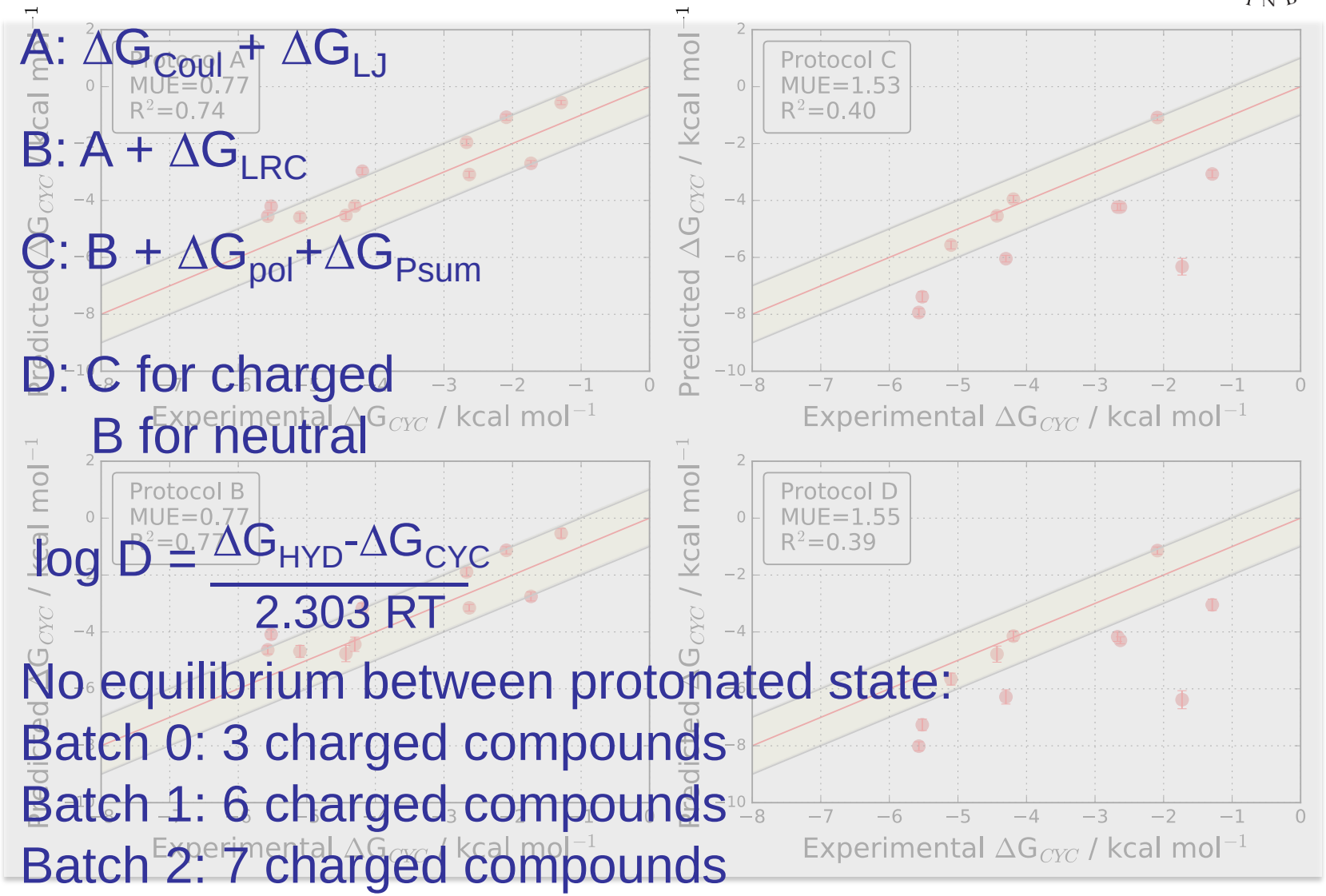
B: A + $\Delta G_{LRC}$

C: B + $\Delta G_{pol} + \Delta G_{Psum}$

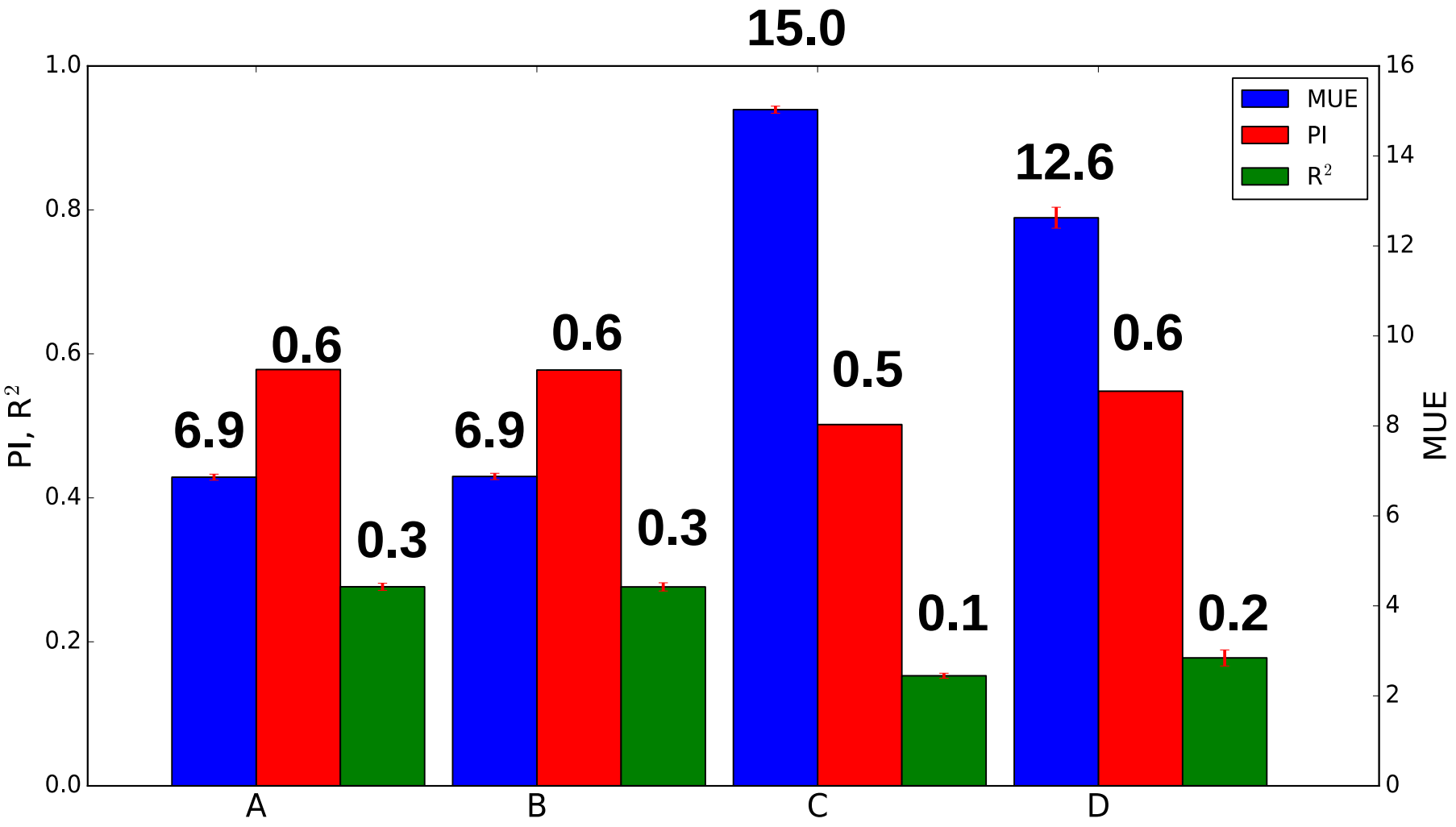
D: C for charged
B for neutral

$$\log D = \frac{\Delta G_{HYD} - \Delta G_{CYC}}{2.303\ RT}$$

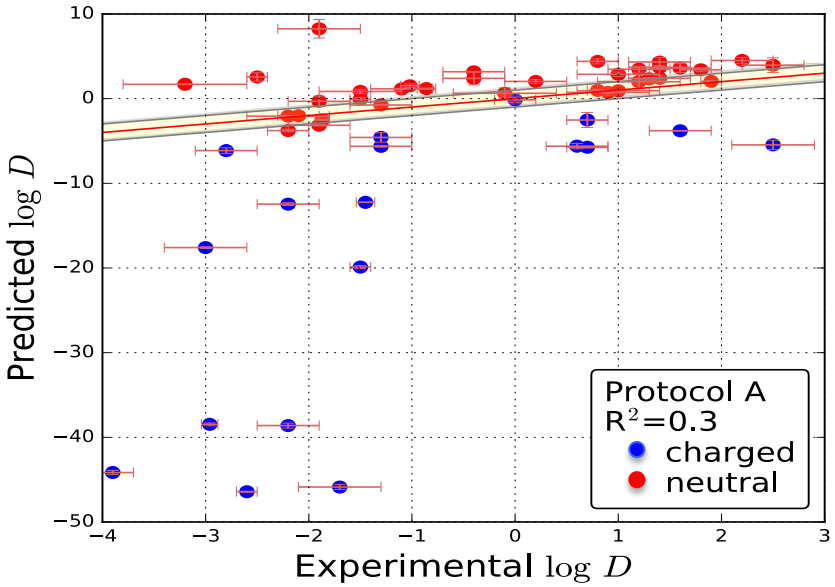
No equilibrium between protonated state:
Batch 0: 3 charged compounds
Batch 1: 6 charged compounds
Batch 2: 7 charged compounds

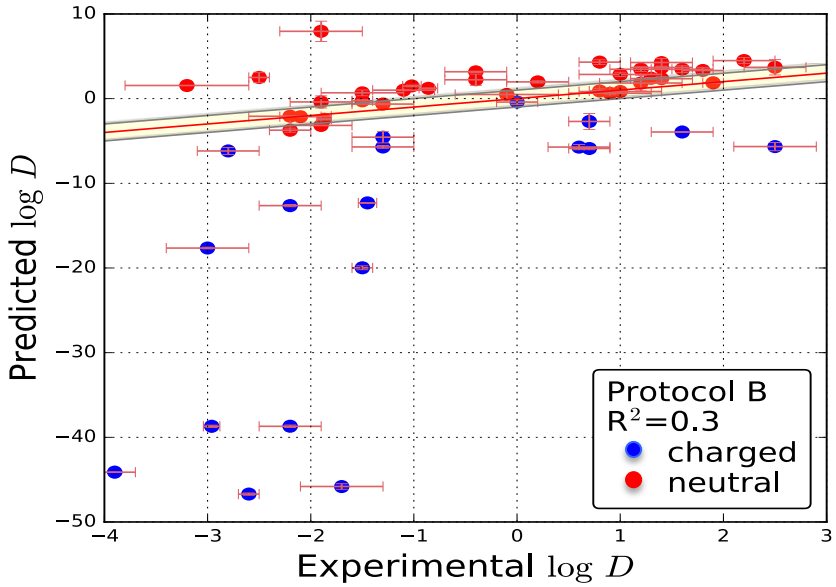


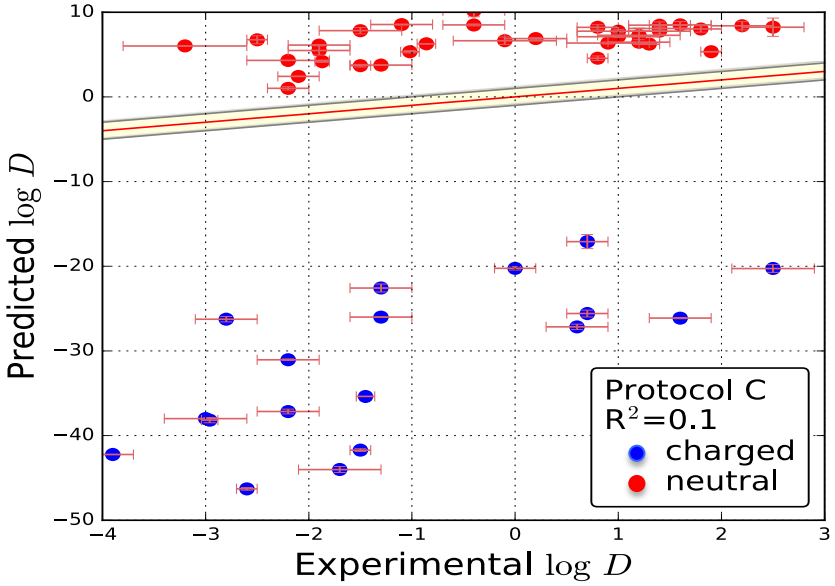
# ALL DATASET
## *MUE – PI – $R^2$*

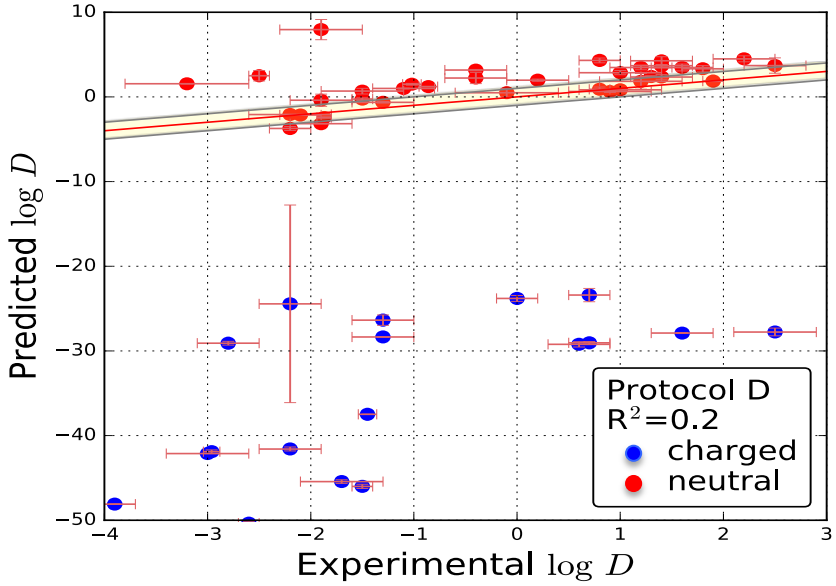


# ALL COMPOUNDS

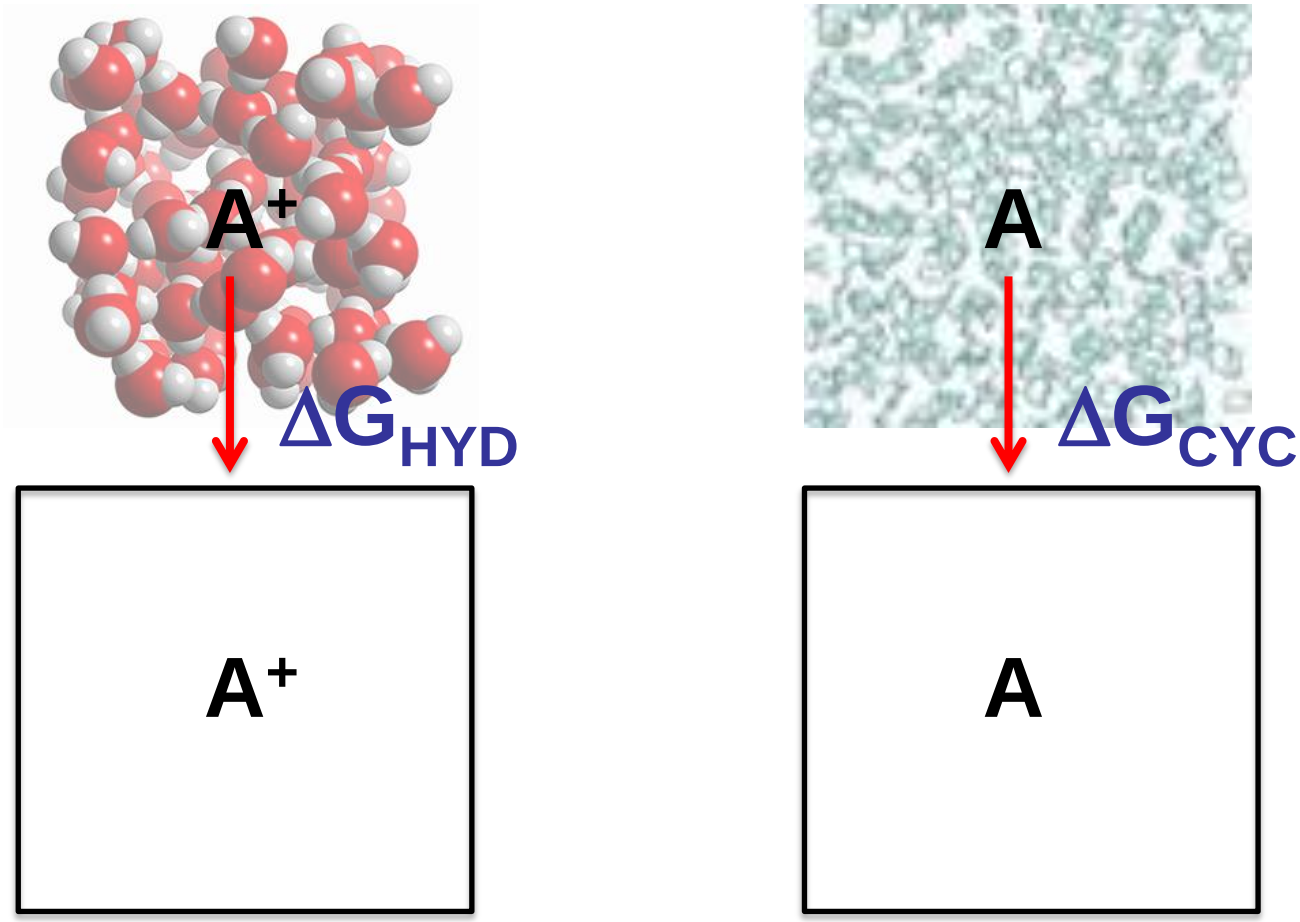


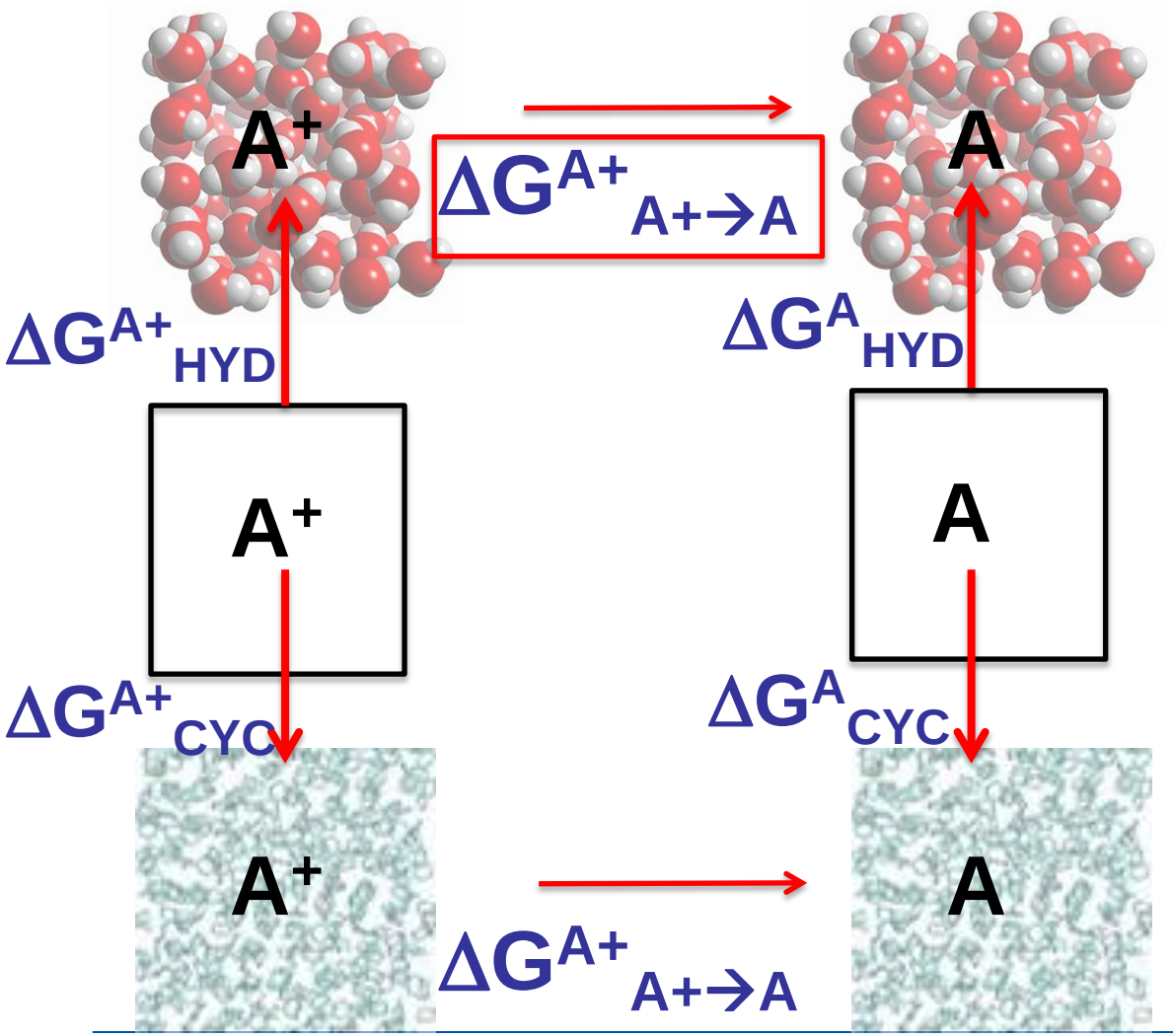



# A FIRST LESSON
# One-specie equilibrium model



# CORRECTION
# Multi-species equilibrium model

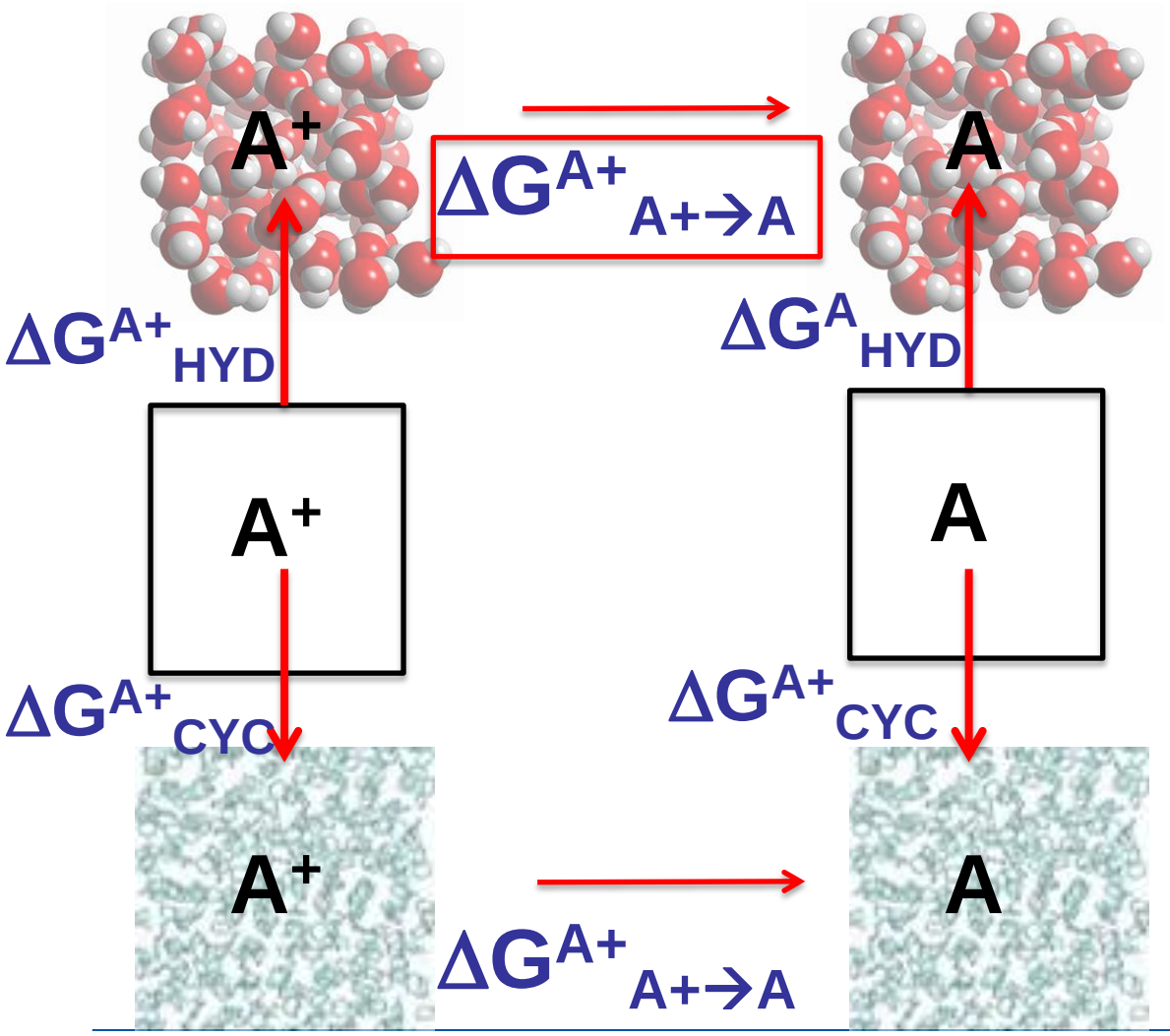



chemicalize.org

$[A^{w}_{tot}] = [A^{w+}] + [A^{w}]$

1M = 1-x x

# CORRECTION
# Multi-species equilibrium model



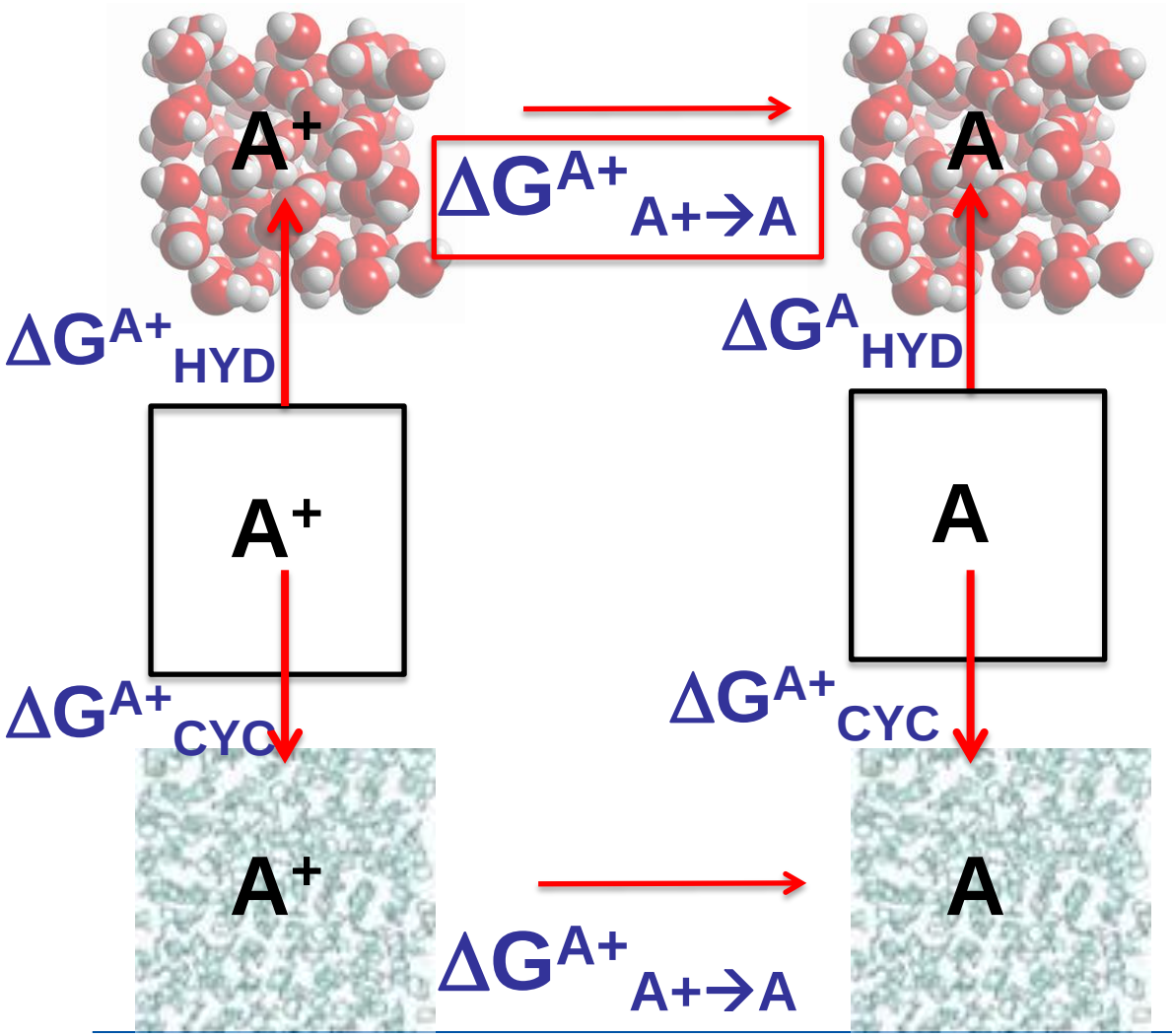

chemicalize.org

$[A^{w}_{tot}] = [A^{w+}] + [A^{w}]$

1M = 1-x x

$[A^{c}_{tot}] = [A^{c+}] + [A^{c}]$

# CORRECTION
# Multi-species equilibrium model




chemicalize.org

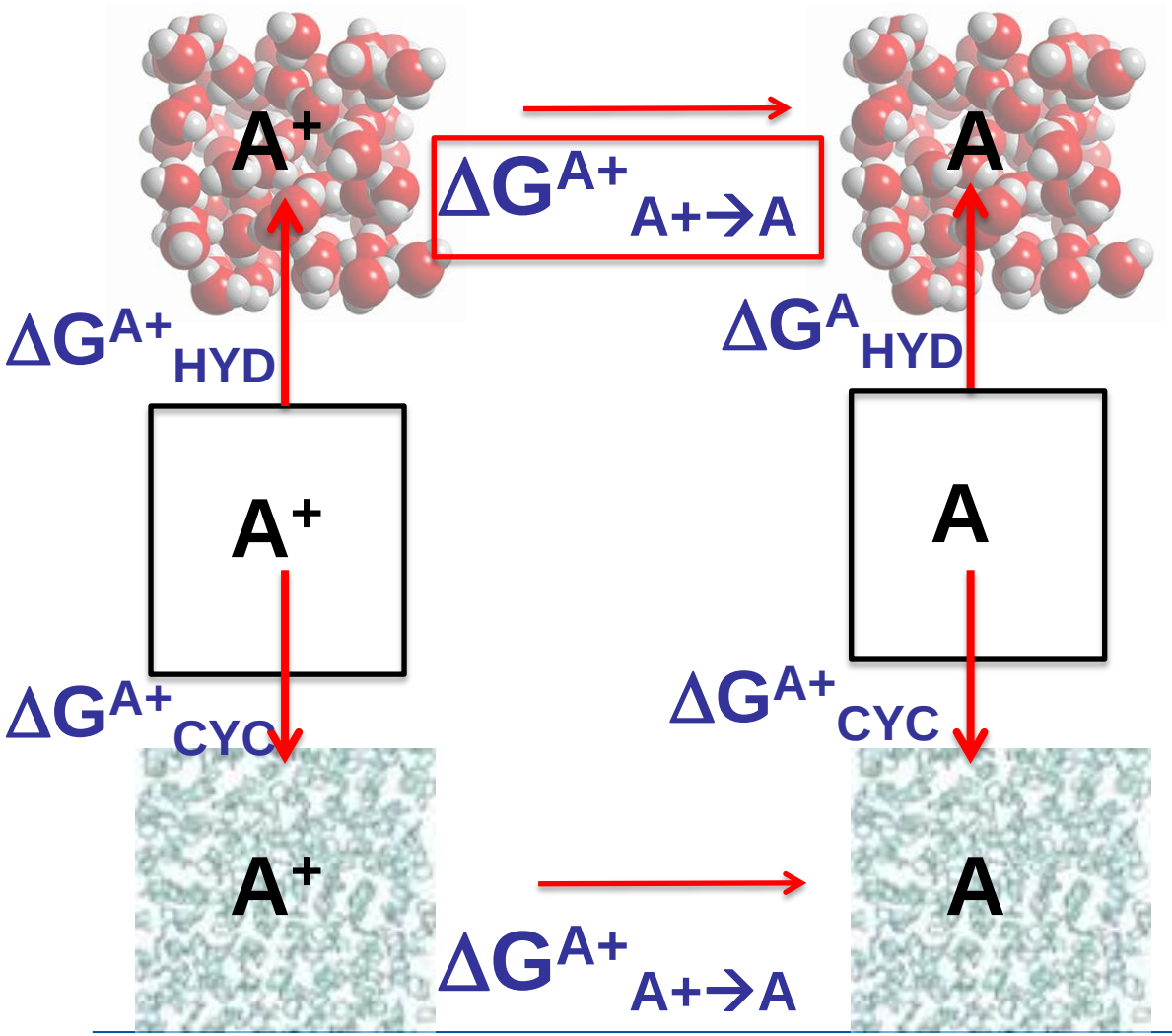
$[A^{w}_{tot}] = [A^{w+}] + [A^{w}]$

1M = 1-x x

$[A^{c}_{tot}] = [A^{c+}] + [A^{c}]$

$= \varepsilon$ y

# CORRECTION
# Multi-species equilibrium model




chemicalize.org

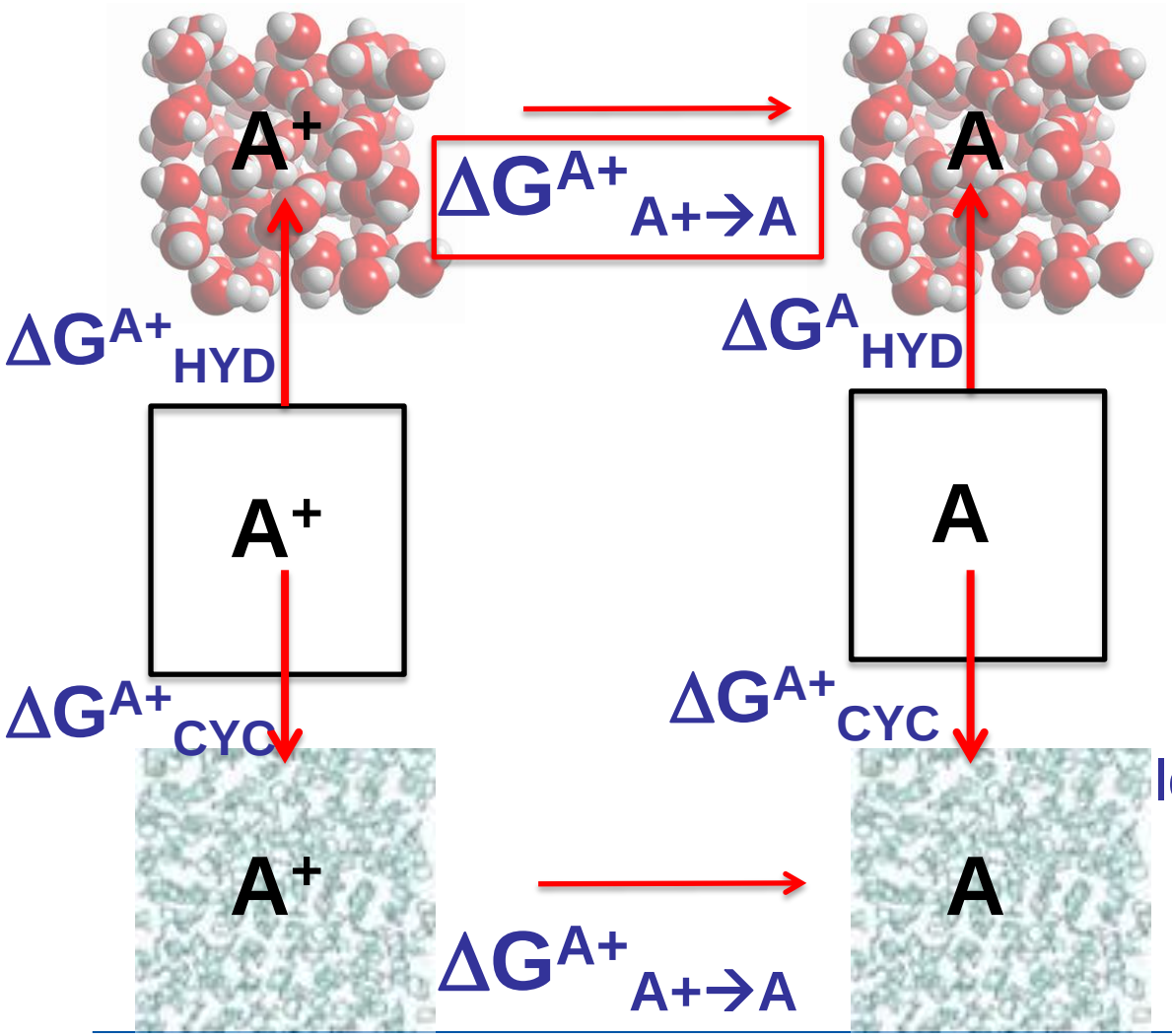
$[A^{w}_{tot}] = [A^{w+}] + [A^{w}]$

1M = 1-x-ε x-y

$[A^{c}_{tot}] = [A^{c+}] + [A^{c}]$

= ε y

# CORRECTION
# Multi-species equilibrium model




chemicalize.org

$$[A^{w}_{tot}] = [A^{w+}] + [A^{w}]$$

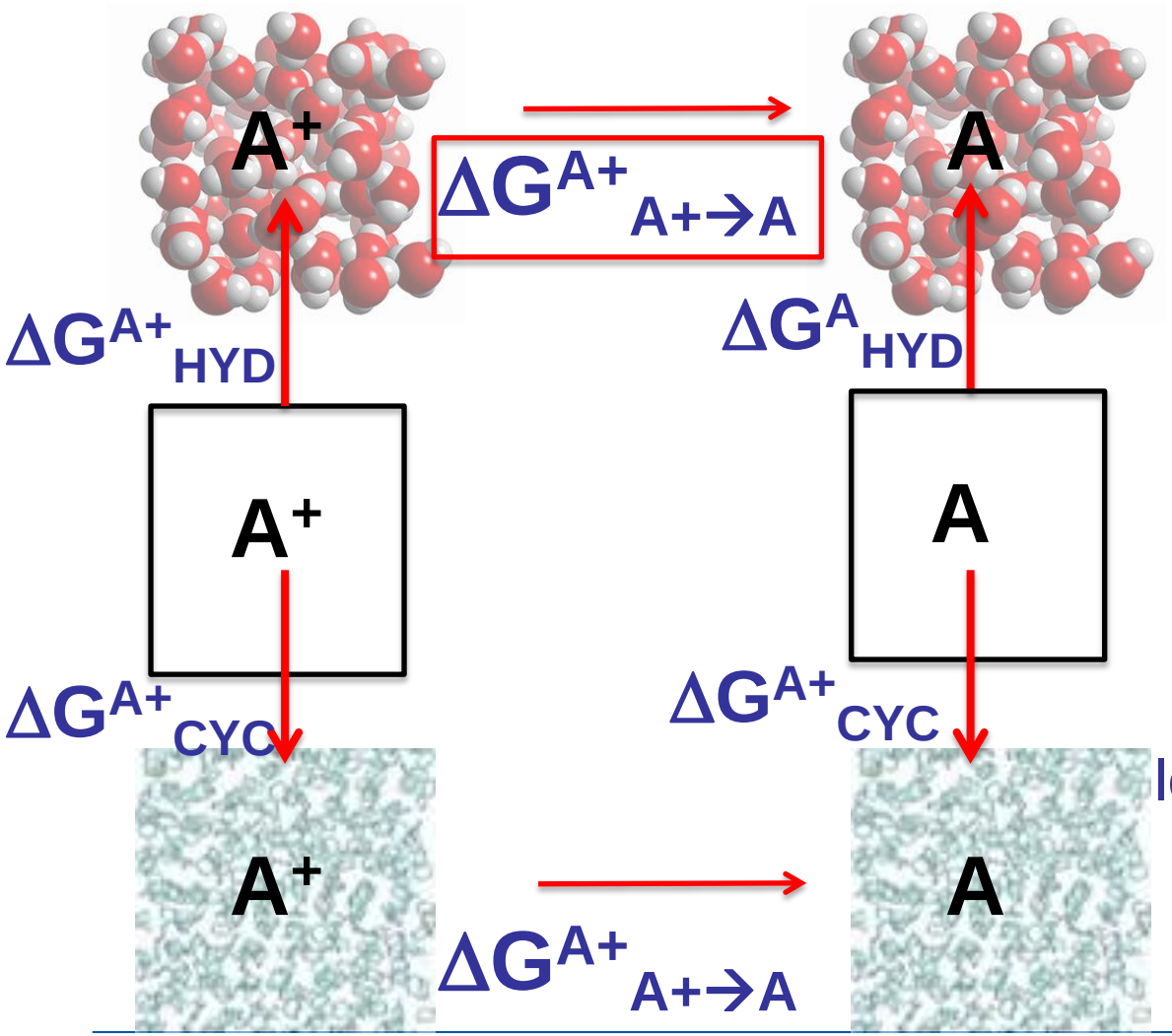
$$1M = 1\text{-}x\text{-}\varepsilon \quad x\text{-}y$$

$$[A^{c}_{tot}] = [A^{c+}] + [A^{c}]$$

$$= \varepsilon \qquad y$$

$$\log D = \log_{10} \frac{[A^{c+}] + [A^{c}]}{[A^{w+}] + [A^{w}]}$$

# CORRECTION
# Multi-species equilibrium model




chemicalize.org

$[A^{w}_{tot}] = [A^{w+}] + [A^{w}]$

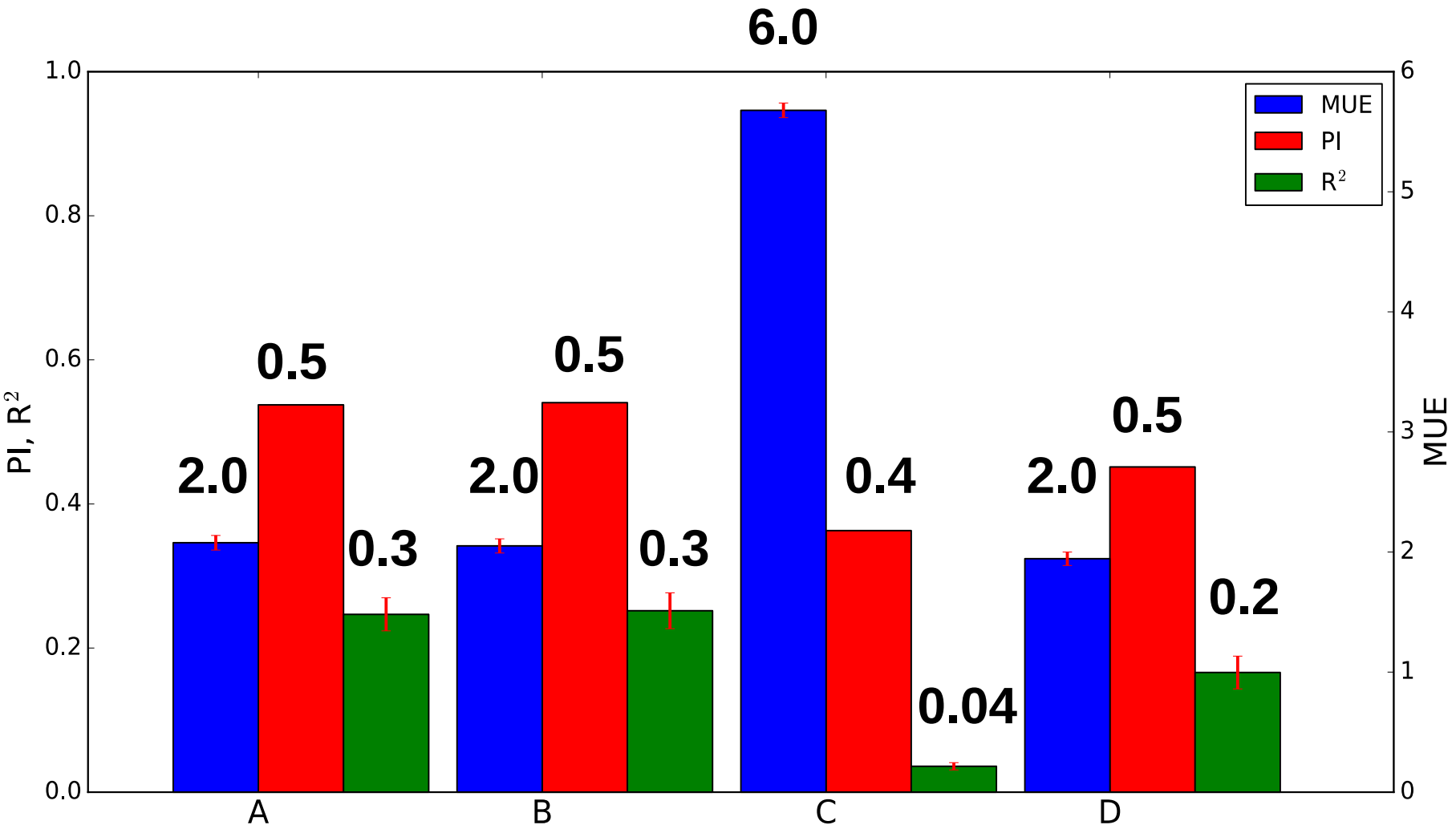
1M = 1-x-ε x-y

$[A^{c}_{tot}] = [A^{c+}] + [A^{c}]$

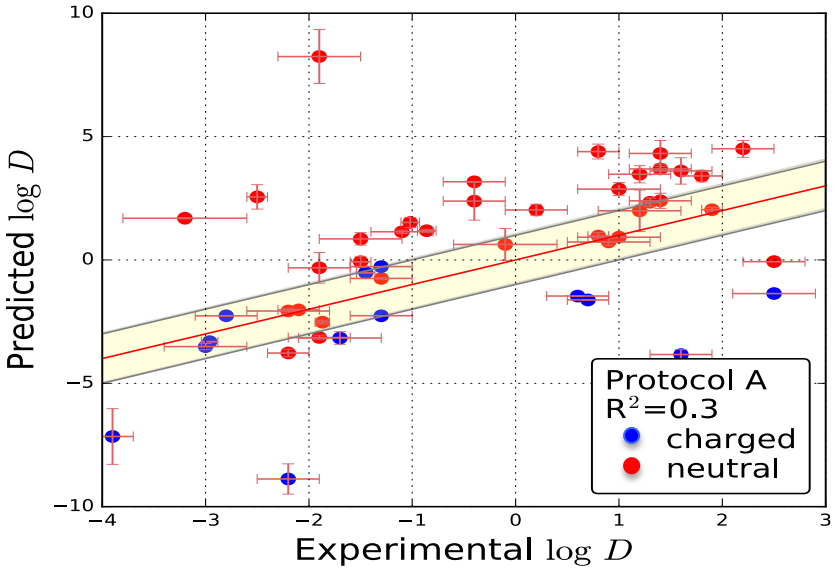
= ε y

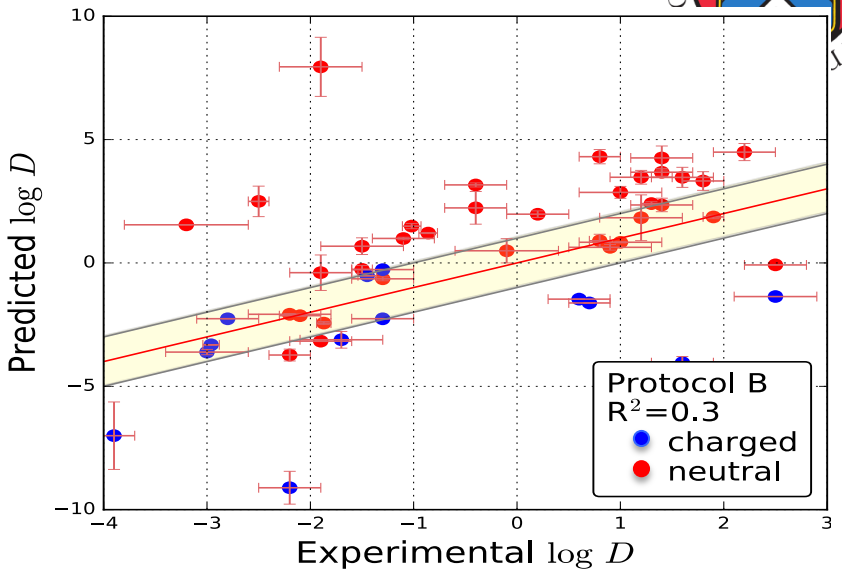
$$\log D = \log_{10} \frac{\varepsilon + y}{1-x-\varepsilon \quad x-y}$$

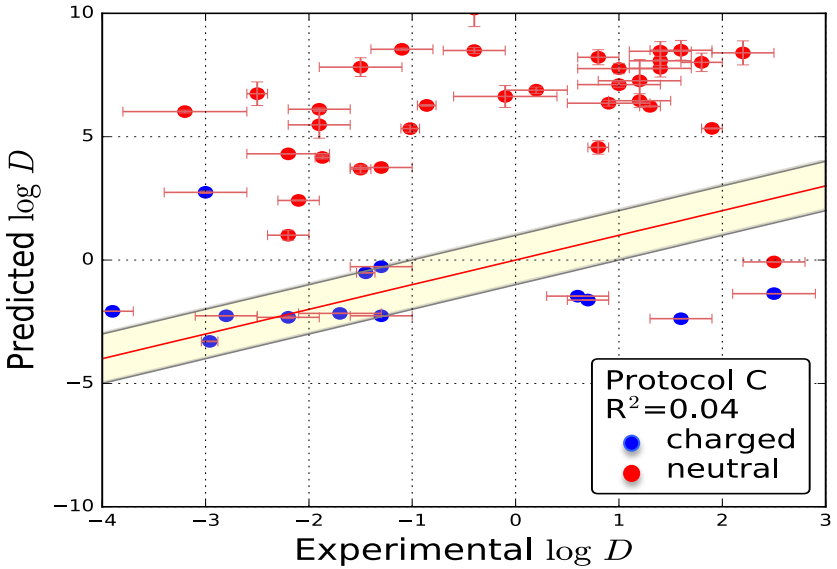
# CORRECTION
## Multispecies statistics

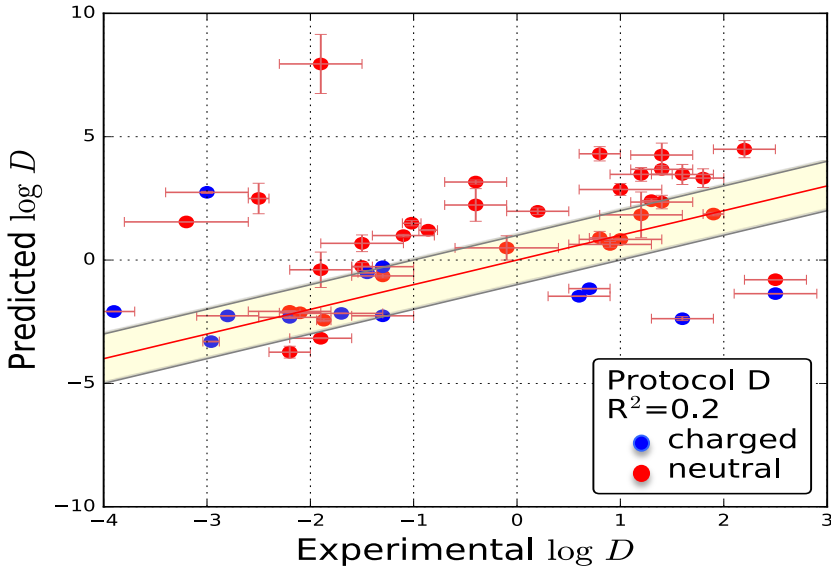


# CORRECTION

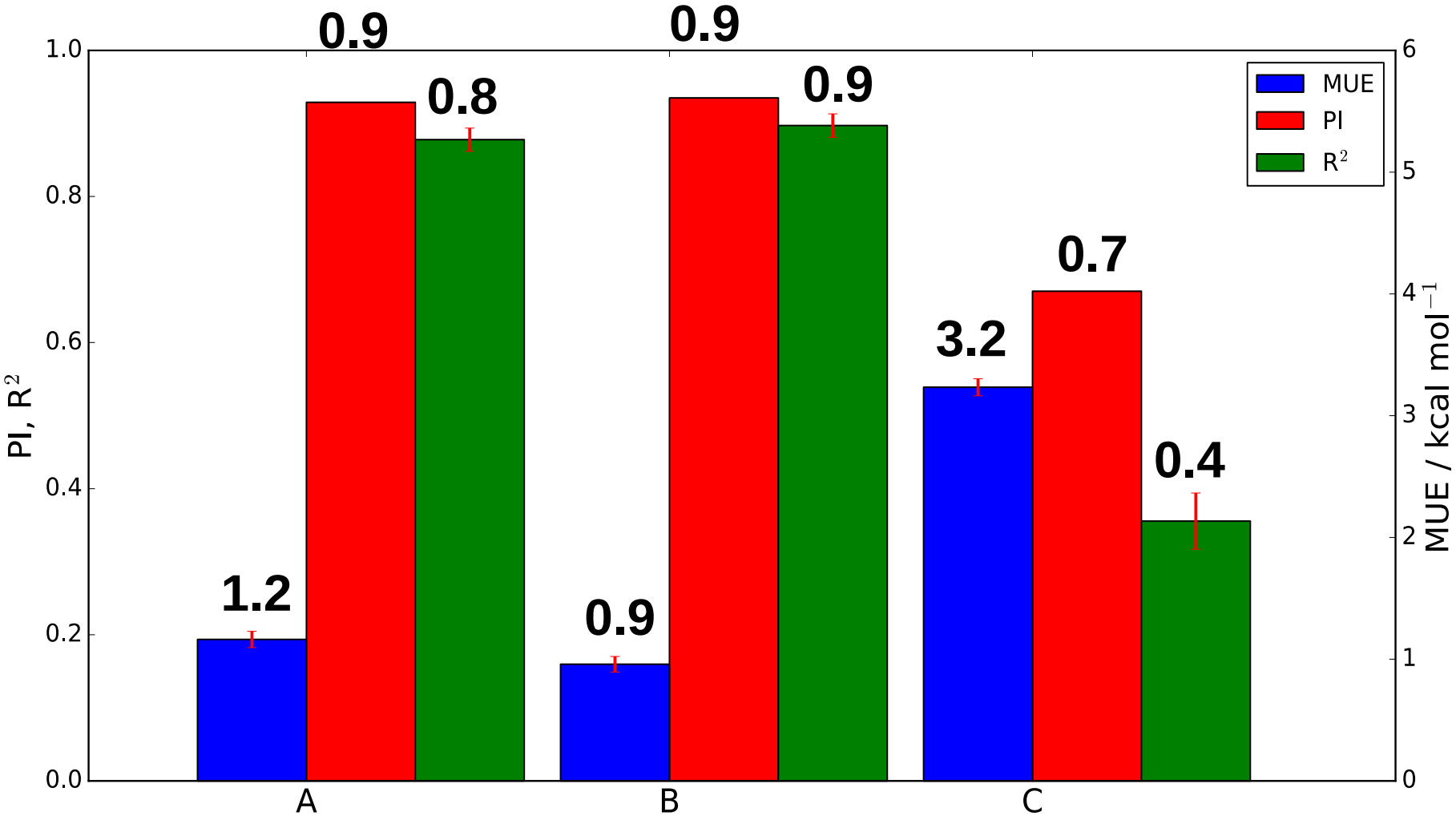

Predicted log $D$
Experimental log $D$
Protocol A
$R^2$=0.3
charged
neutral


Predicted log $D$
Experimental log $D$
Protocol B
$R^2$=0.3
charged
neutral

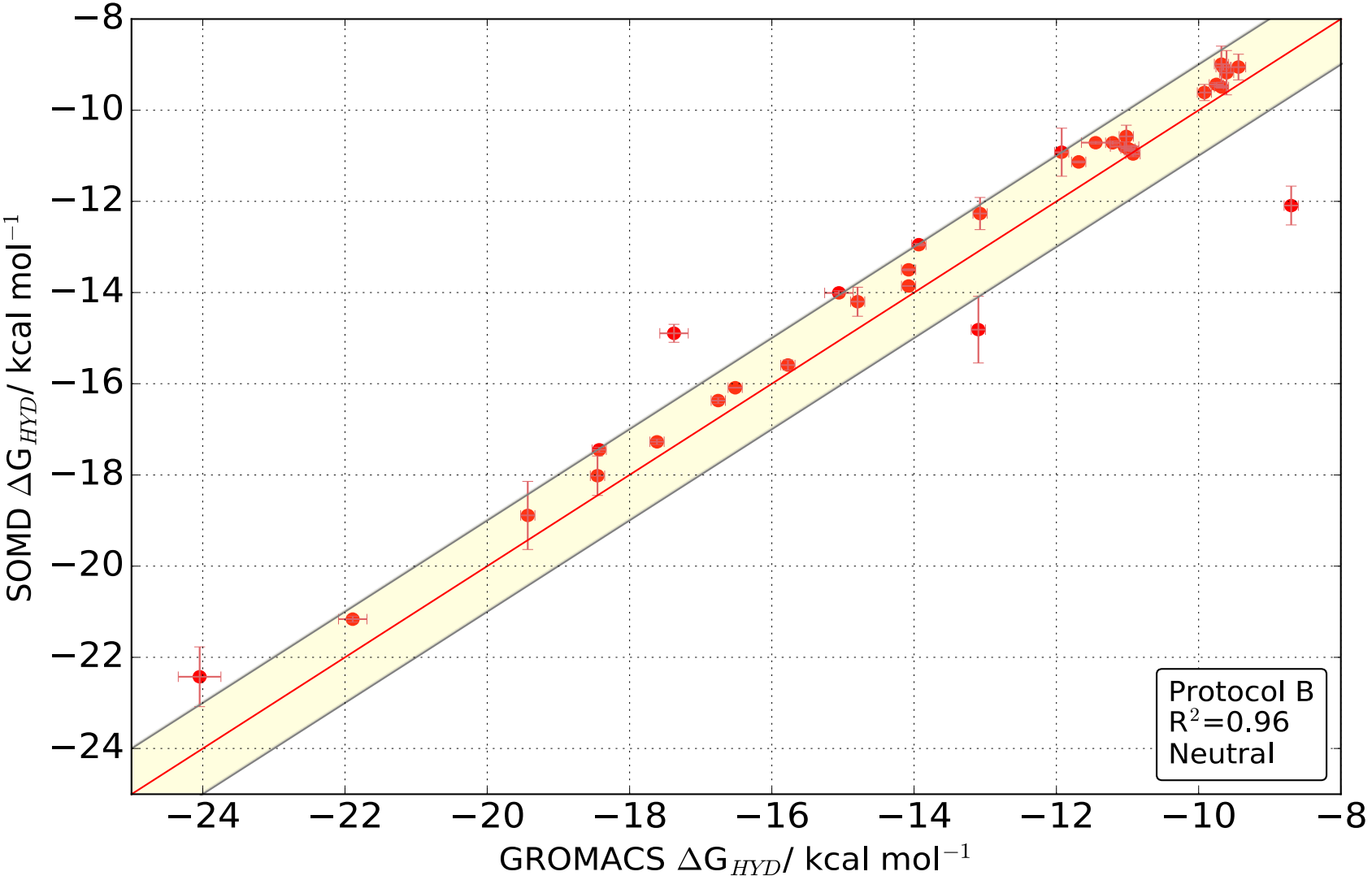

Predicted log $D$
Experimental log $D$
Protocol C
$R^2$=0.04
charged
neutral


Predicted log $D$
Experimental log $D$
Protocol D
$R^2$=0.2
charged
neutral

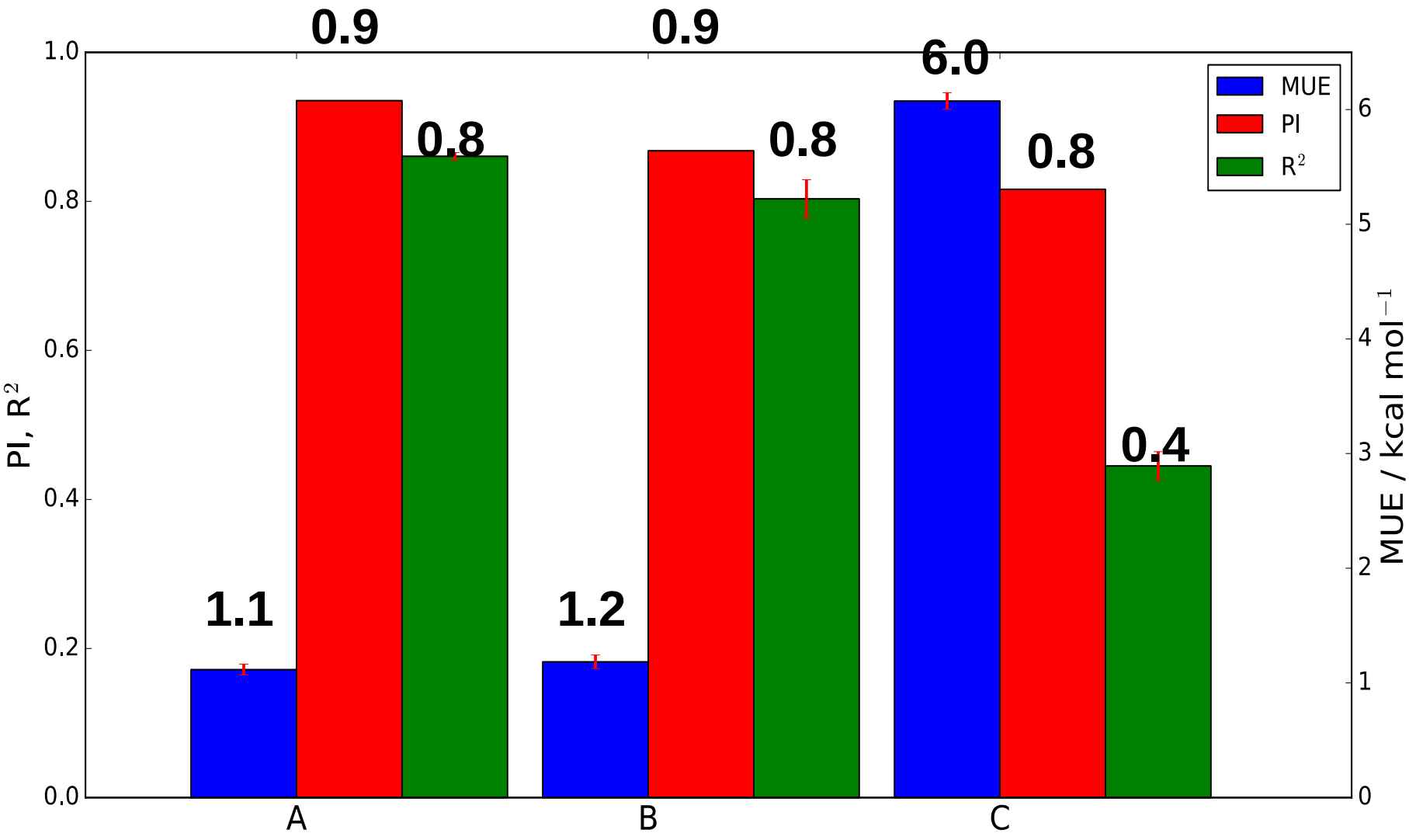
# ALL NEUTRAL?
## A Comparison with GROMACS $\Delta G_{HYD}$



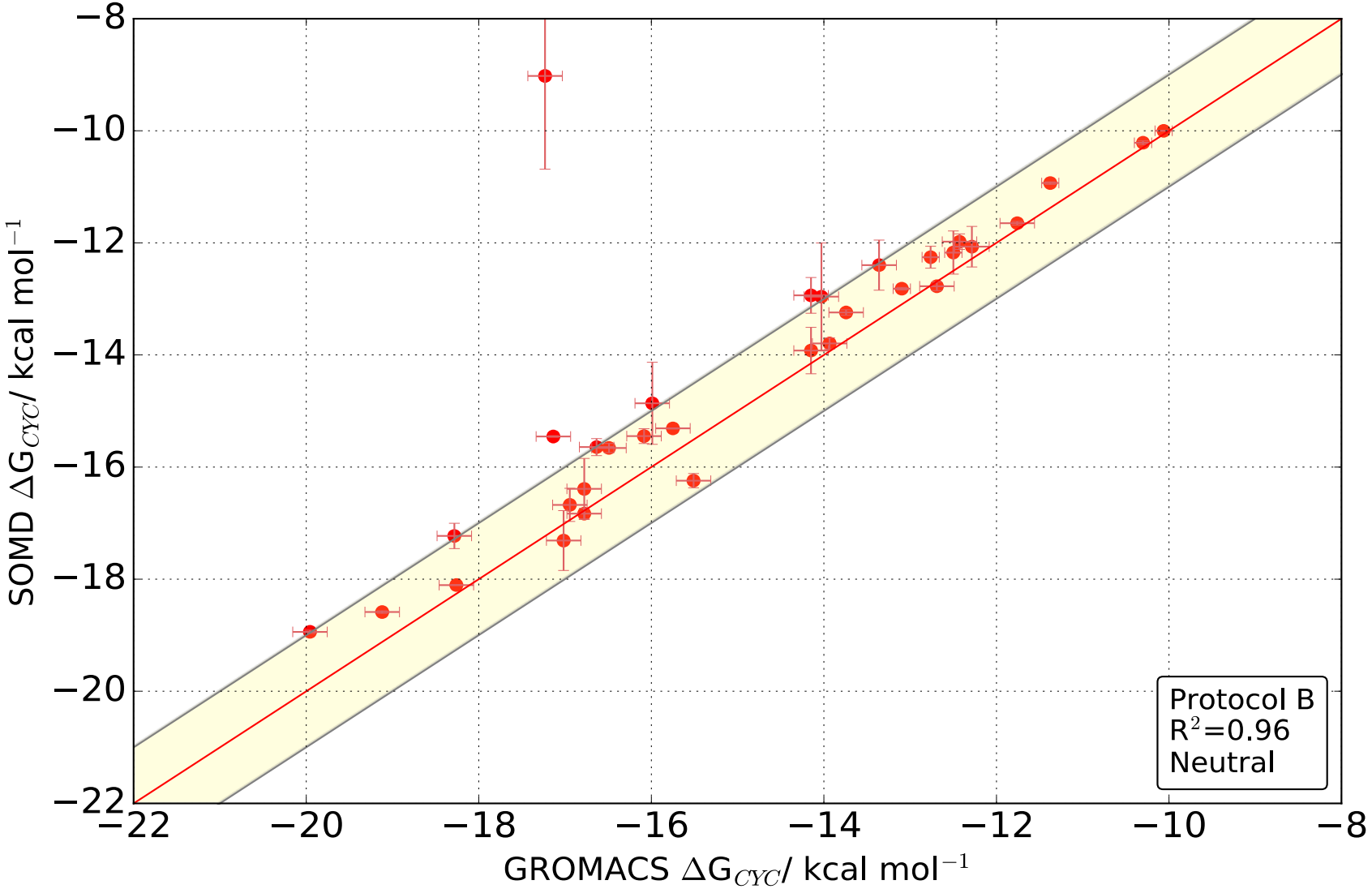
# ALL NEUTRAL?Hydration Free Energies A Comparison with GROMACS



# ALL NEUTRAL?
# A Comparison with GROMACS $\Delta G_{CYC}$



# All neutral Reproducibility-GROMACS



# CONCLUSIONS

HOST-GUEST:

- Good $R^2$ for all protocols, except D
- Very good agreement with other simulations

DISTRIBUTION COEFFICIENT:

- Protonation state must be considered
- Poor results unless neutral species are considered in water
- GAFF is too hydrophobic, always favorable to cyclohexane

# AKNOWLEDGMENT

**Julien Michel Group:**

Dr. Julien Michel

Dr. Antonia Mey
Dr. Jordi Juarez Jimenez

Dr. Alessio De Simone, Dr. Arun Gupta, Charis Georgiou, Pattama Wapeesittipan, Joan Clark Nicholas, Harris Ioannidis, Lisa Patrick

**D3R Workshop Organizers**

David Mobley, Michael Gilson, Kenneth Tomory


Thanks for your attention
QUESTIONS?