SOMD- lessons learned from SAMPL6

Michail Papadourakis and Stefano Bosisio

Julien Michel's Group

Michail Papadourakis

Stefano Bosisio

22/02/2018

Outline

1. Protocol description: how we calculate the binding free energy?

2. Comparison of results of datasets

3. SAMPLing results

Software for the simulations

- SOMD: alchemical free energy calculation engine

Sire
MC & Free Energy
Woods, Christopher. “Sire 2016.3.1.” data.bris, January 17, 2017

OpenMM
MD & GPUs
Eastman, Peter, et al. “OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics.” *PLOS Computational Biology* 13, no. 7 (July 26, 2017): e1005659.

Sire/OpenMM (SOMD)
http://www.siremol.org

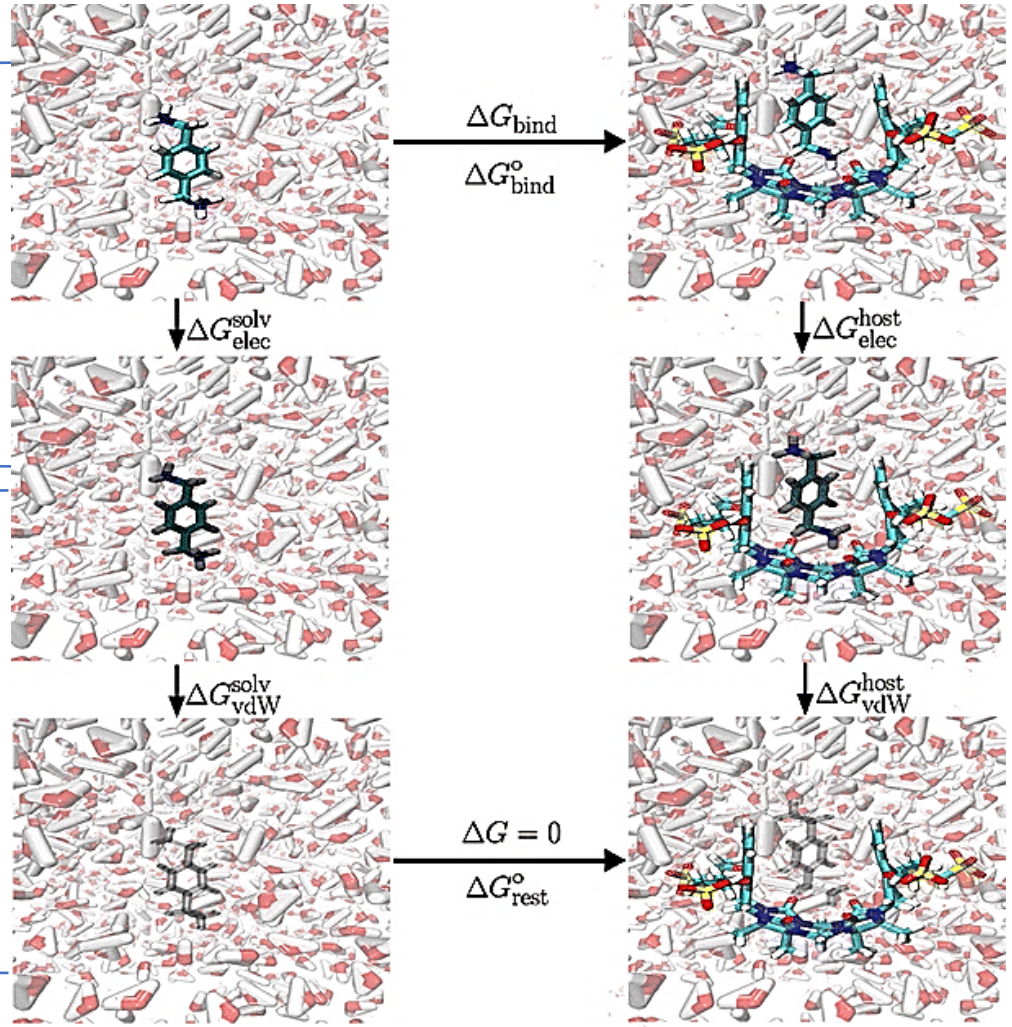
Gaetano Calabro

Antonia Mey

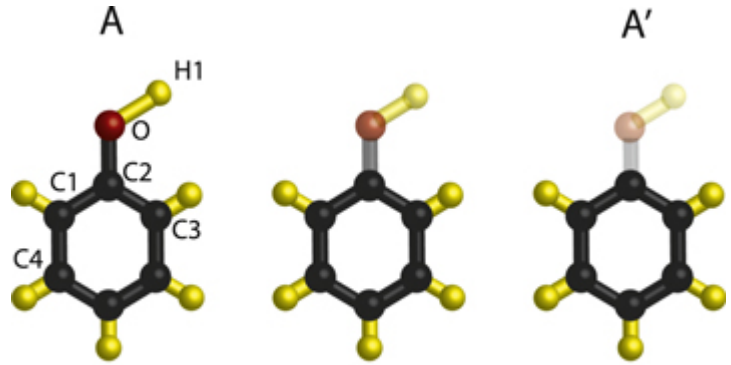
Model A: binding free energy

Switching off charges

Switching off van der Waals terms



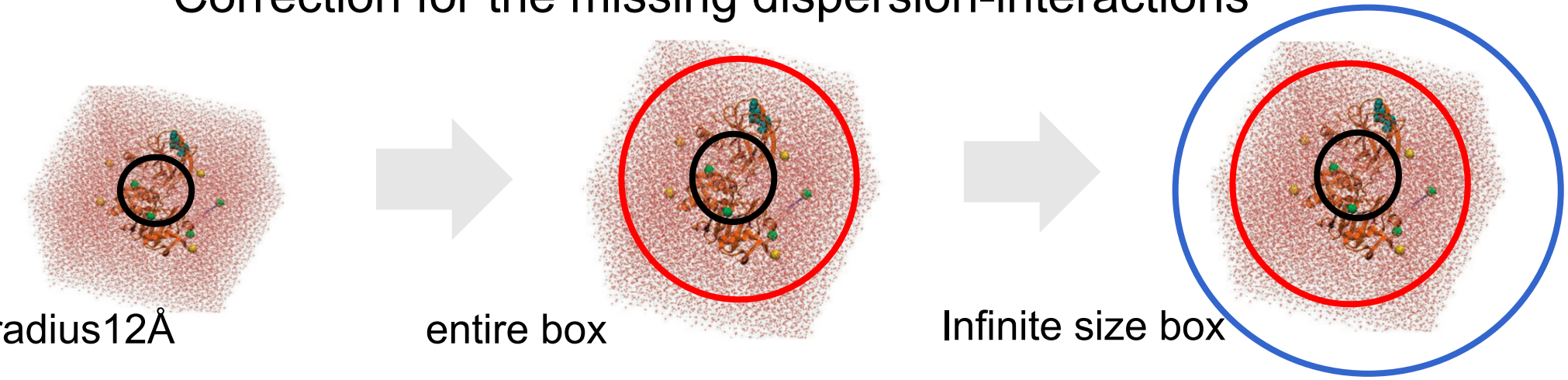
Alchemical intermediate states(0 < λ < 1)



$$\Delta G_{bind} = \left(\Delta G^{solv}_{elec} + \Delta G^{solv}_{vdW}\right) - \left(\Delta G^{host}_{elec} + \Delta G^{host}_{vdW}\right)$$

Model C: Long Range Correction + Standard State

Correction for the missing dispersion-interactions



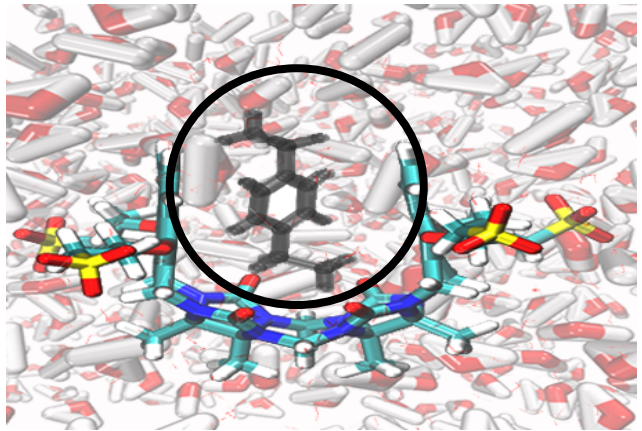
Cutoff radius12Å

entire box

Infinite size box

Free energy cost for imposing a restraint – Standard State correction

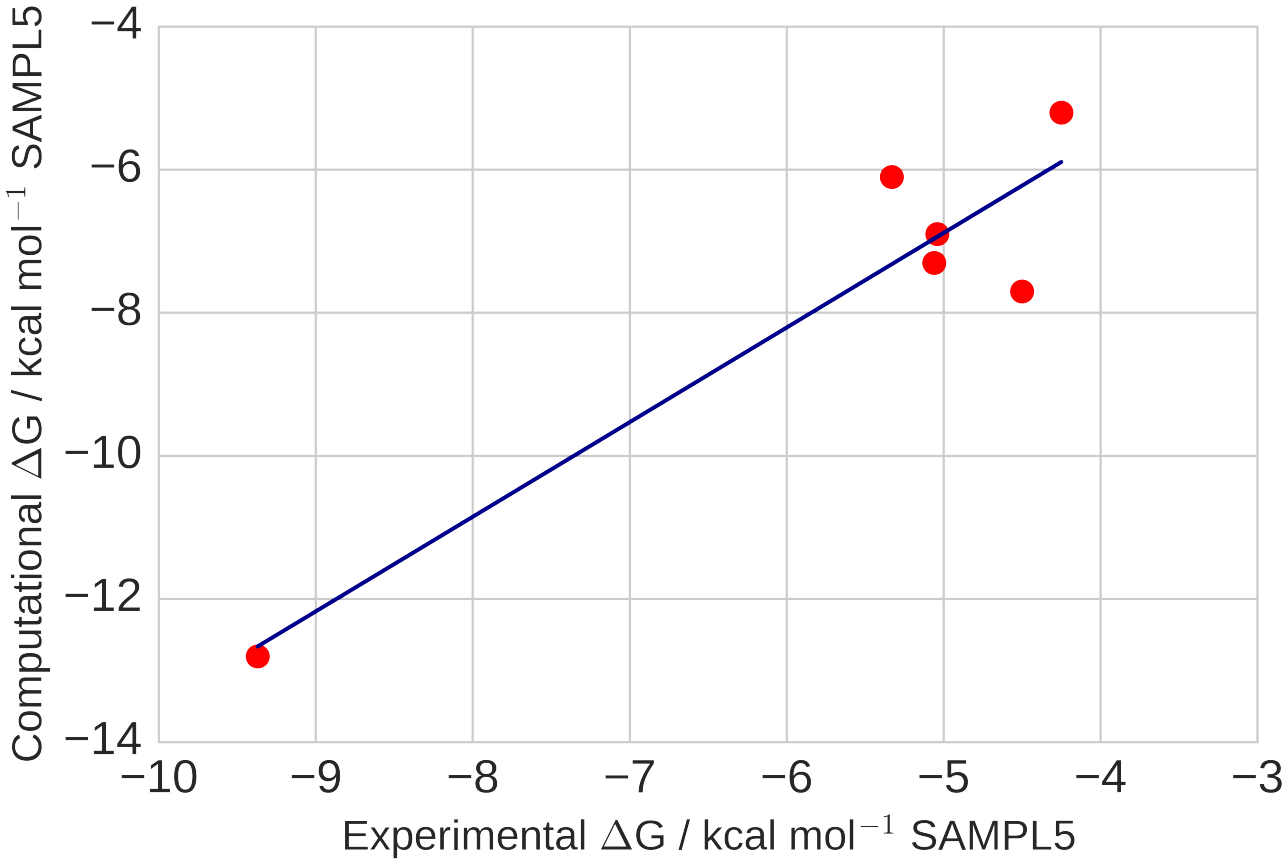
$A + B \rightleftarrows AB$



Model D: Model C + Empirical Estimation

Linear regression from the comparison of the experimental results $\Delta G^{\circ}_{experimental}$ and our computational results $\Delta G^{\circ}_{computational}$ in SAMPL5.

- $\Delta G^{\circ}_{computational\ SAMPL5}$ = a $\Delta G^{\circ}_{experimental\ SAMPL5}$ + b
- $\Delta G^{\circ}_{mod\ C}$ = a $\Delta G^{\circ}_{mod\ D}$ + b
- $\Delta G^{o}_{mod\ D} = \frac{\Delta G^{\circ}_{modC} - b}{a}$



Modelling the buffer

❖ Three models submitted for the calculation of the free energy of binding:

1. **Model A** = Binding Free Energy
2. **Model C** = Model A + Lennard Jones correction term + Standard state correction term
3. **Model D** = Model C + Empirical estimation of binding free energy

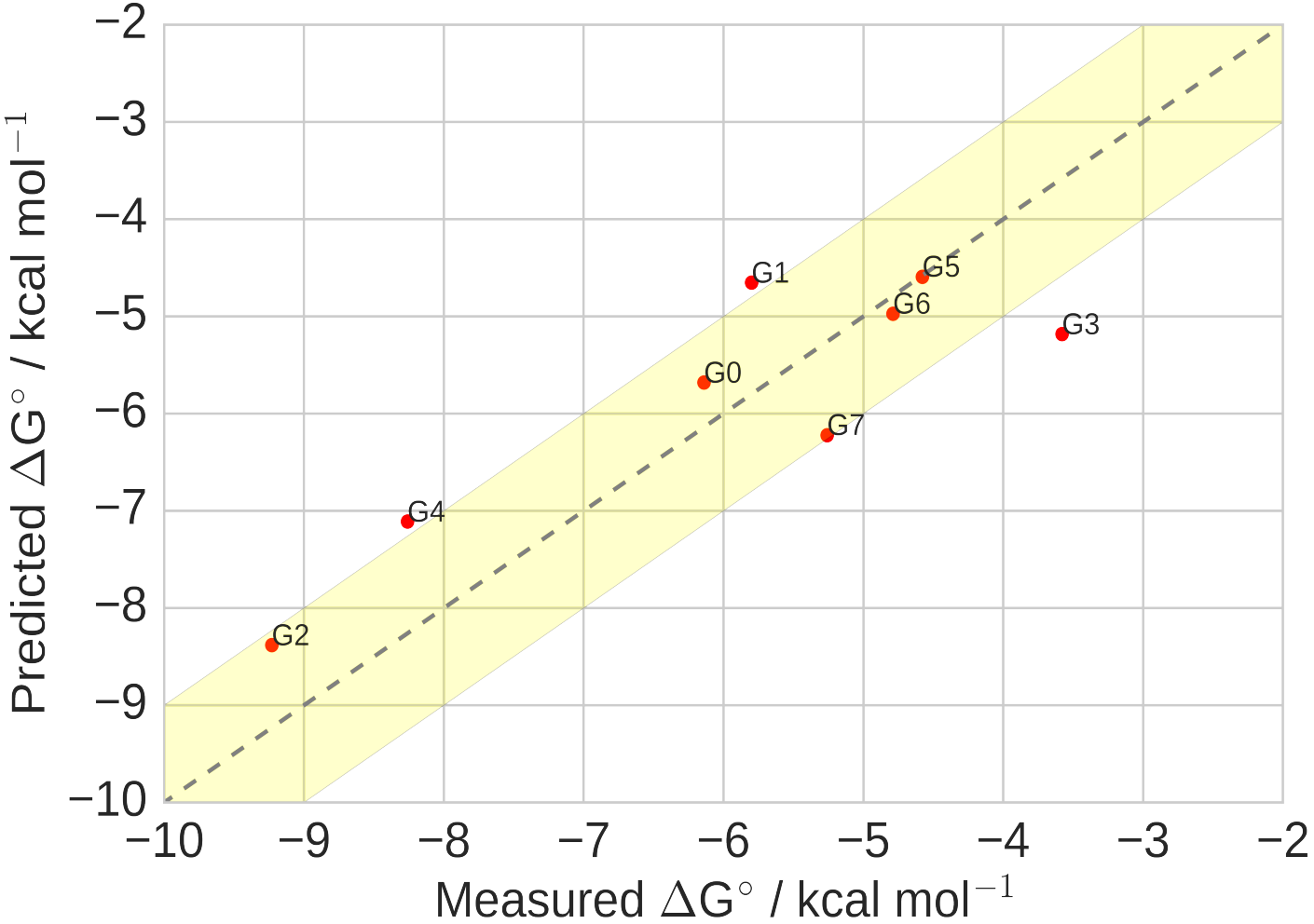
❖ Two sets of simulation:

I. **No buffer**: Ions to neutralize the overall charge of the box

II. **Buffer**: NaCl to model the sodium phosphate buffer by replicating the ionic strength

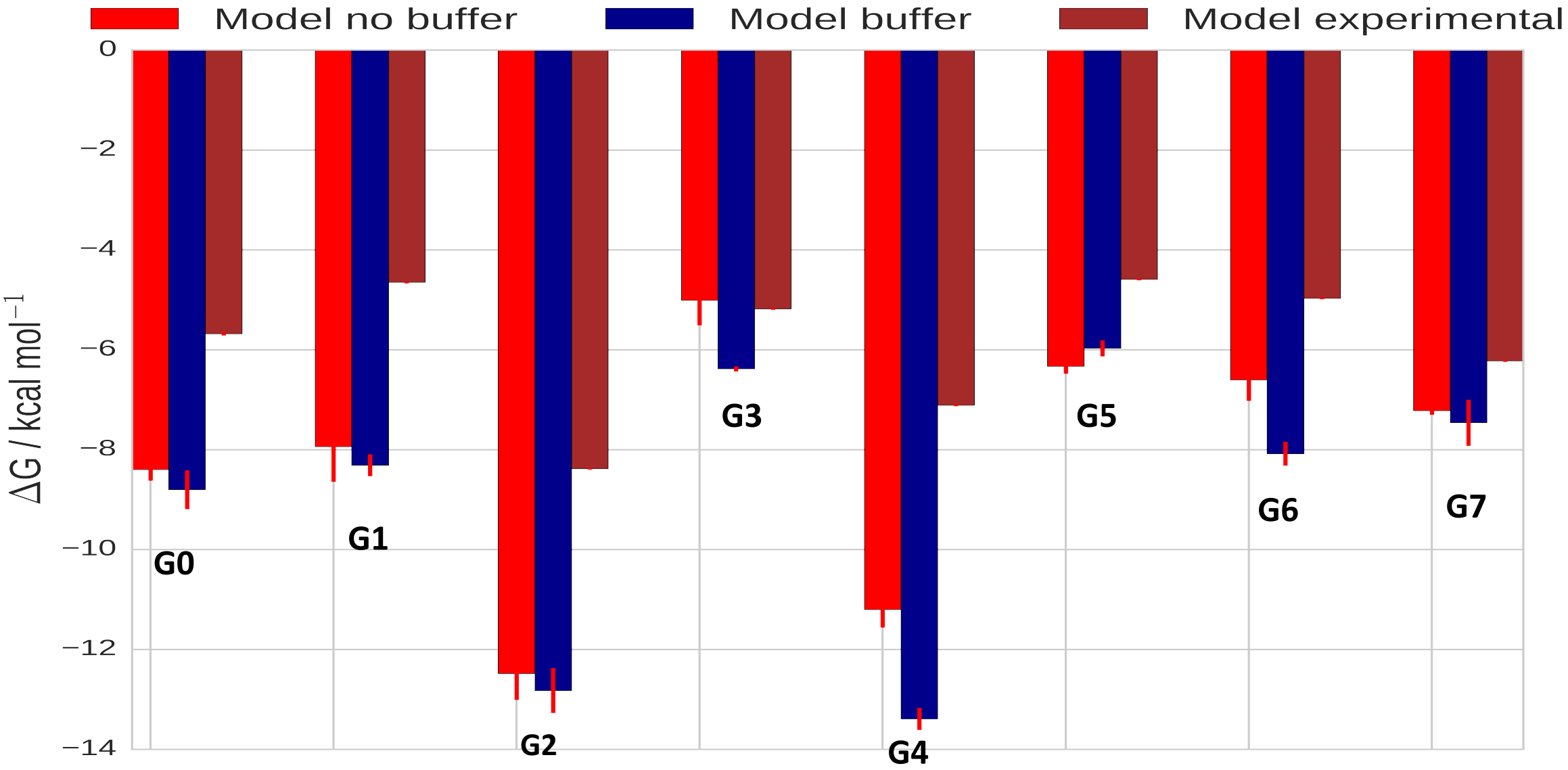
✓ Mobley, David L., and Michael K. Gilson. "Predicting Binding Free Energies: Frontiers and Benchmarks." *Annual Review of Biophysics* 46 (May 22, 2017): 531–58.

Results of OA – Model D buffer

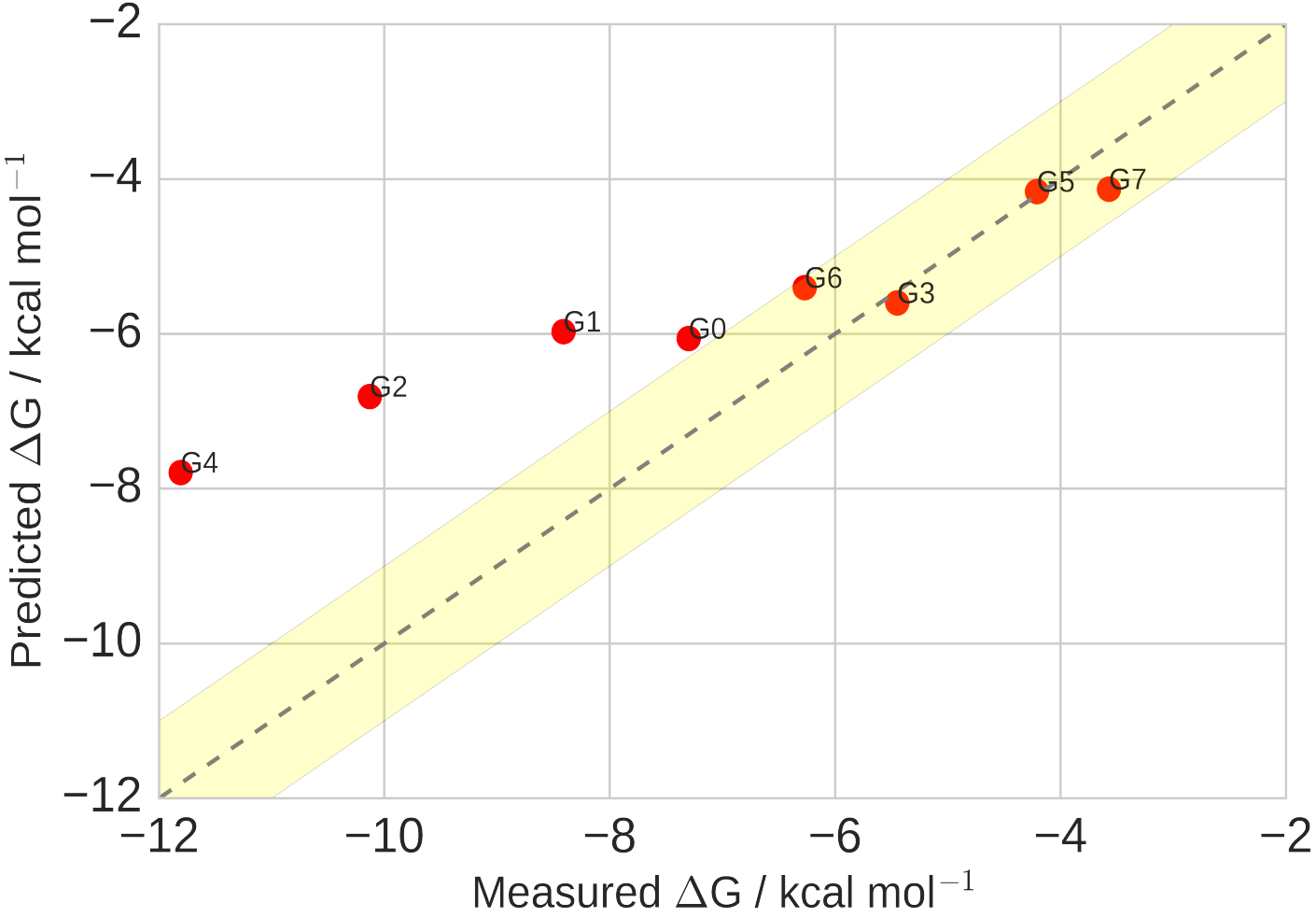


- Model D, **R^2** : 0.62 < 0.74< 0.84
  MUE : 0.65< 0.82 < 1.00 kcal/mol
  Kendall tau : 0.43 < 0.54 < 0.64
- Model C, **R^2** : 0.62 < 0.74 < 0.84
  MUE: 2.15 < 2.37< 2.60 kcal/mol
  Kendall tau : 0.43 < 0.54< 0.64
- Model A , **R^2** : 0.63< 0.75 < 0.85
  MUE: 4.16< 4.41 < 4.66 kcal/mol
  Kendall tau : 0.43 < 0.55< 0.64
- **Buffer** did not improve the results

Comparison buffer-nobuffer in OA host-guest systems
Model C

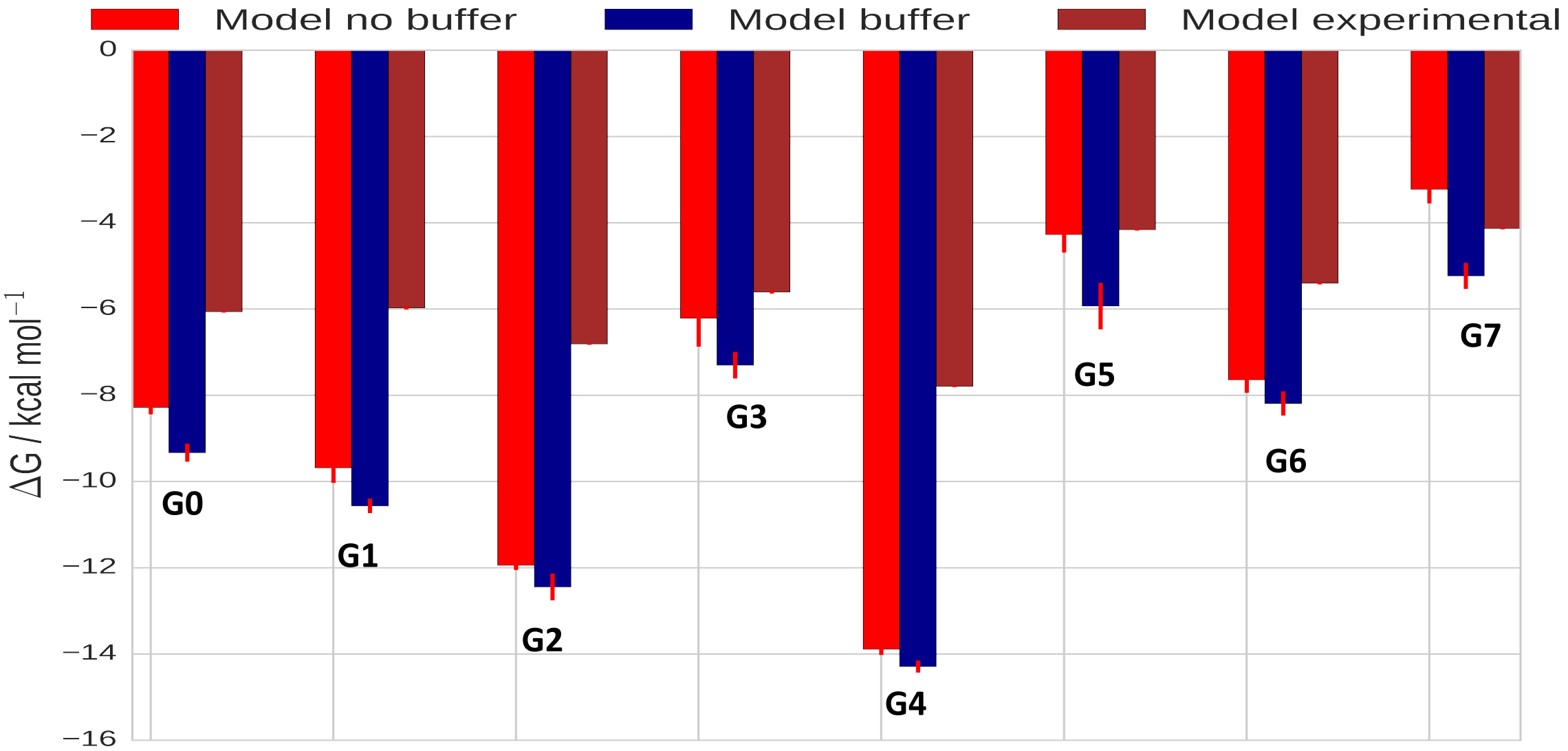


Results of TEMOA – Model D buffer

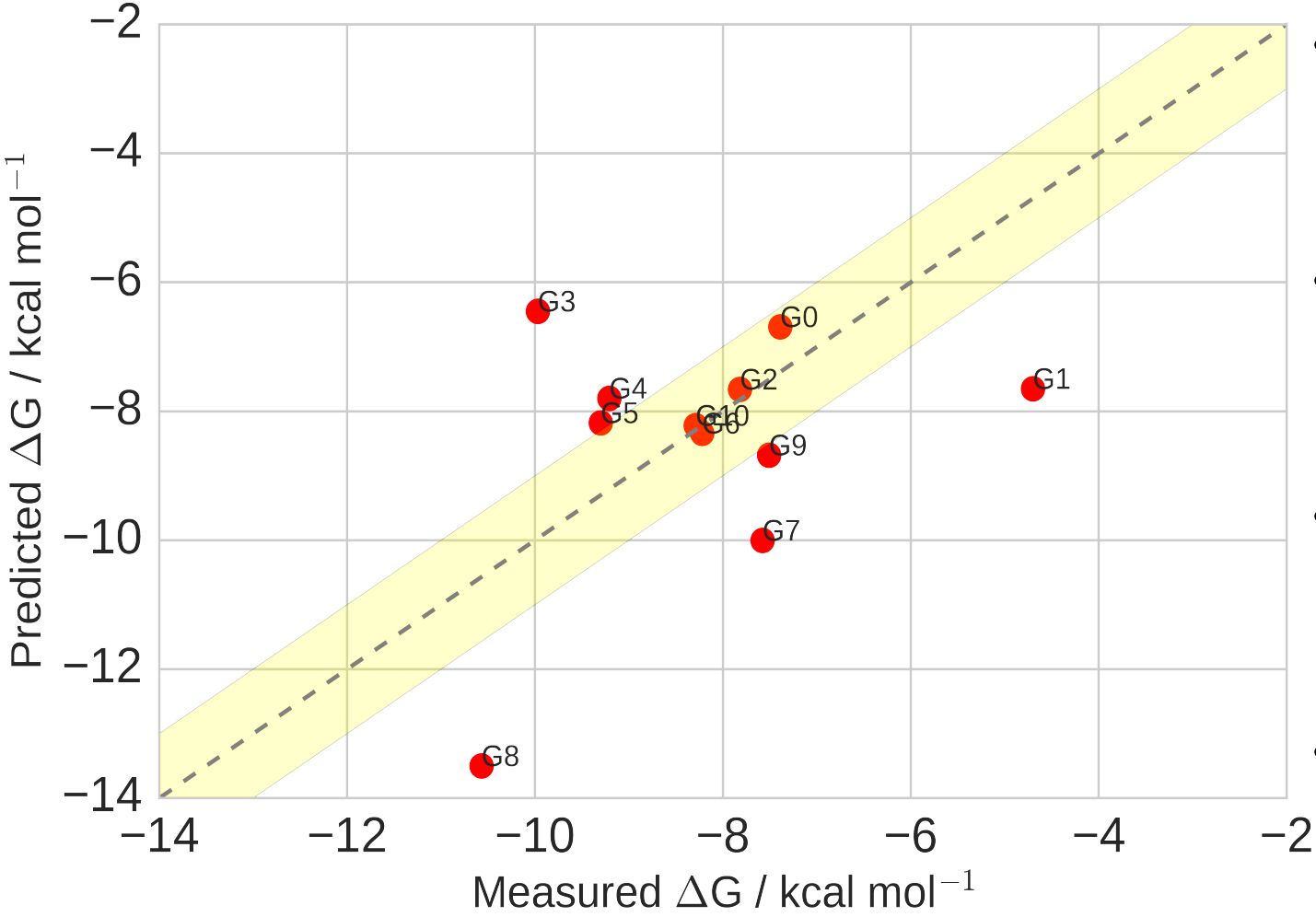


- Model D, **R^2** : 0.89< 0.93 < 0.96
  MUE : 1.51 < 1.64< 1.77 kcal/mol
  Kendall tau : 0.79 < 0.85 < 0.86
- Model C, **R^2** : 0.89 < 0.93 < 0.96
  MUE: 3.23 < 3.42 < 3.59 kcal/mol
  Kendall tau : 0.79 < 0.85 < 0.86
- Model A , **R^2** : 0.89 < 0.93< 0.96
  MUE: 5.48 < 5.66 < 5.85 kcal/mol
  Kendall tau : 0.79 < 0.84 < 0.86
- **Buffer** did not improve the results

Comparison buffer-nobuffer in TEMOA host-guest systems-Model C

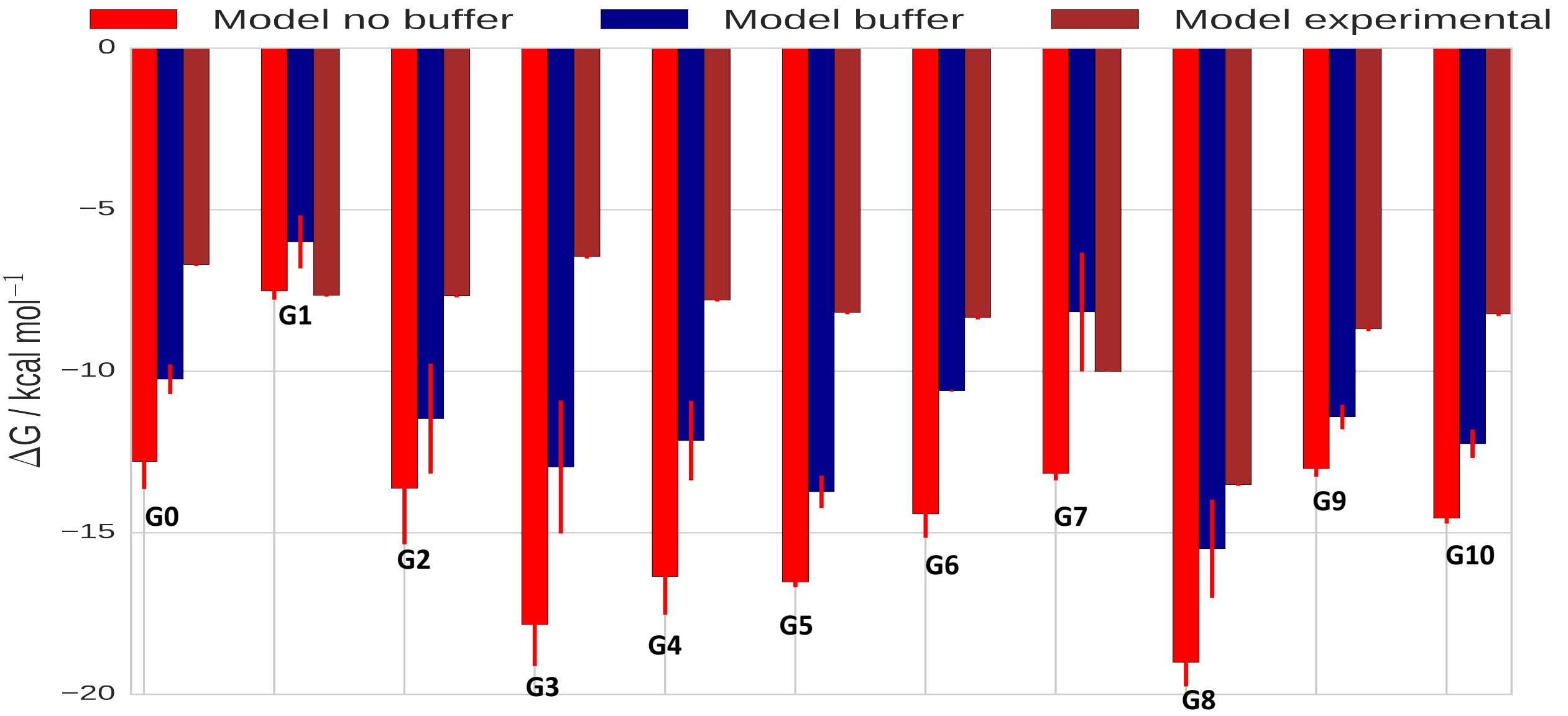


Results of CB8 – Model D no buffer



- Model D, **R^2** : 0.04< 0.13 < 0.23
  MUE : 1.40 < 1.58 < 1.76 kcal/mol
  Kendall tau : -0.09 < 0.05 < 0.20
- Model C **R^2** : 0.04< 0.13 < 0.23
  MUE: 5.57 < 6.00 < 6.43 kcal/mol
  Kendall tau : -0.09 < 0.05 < 0.20
- Model A , **R^2** : 0.04< 0.13 < 0.23
  MUE: 6.90< 7.33 < 7.75 kcal/mol
  Kendall tau : -0.09 < 0.05 < 0.20
- **Buffer** improved the results in Model C as it reduced the MUE by 3 kcal/mol

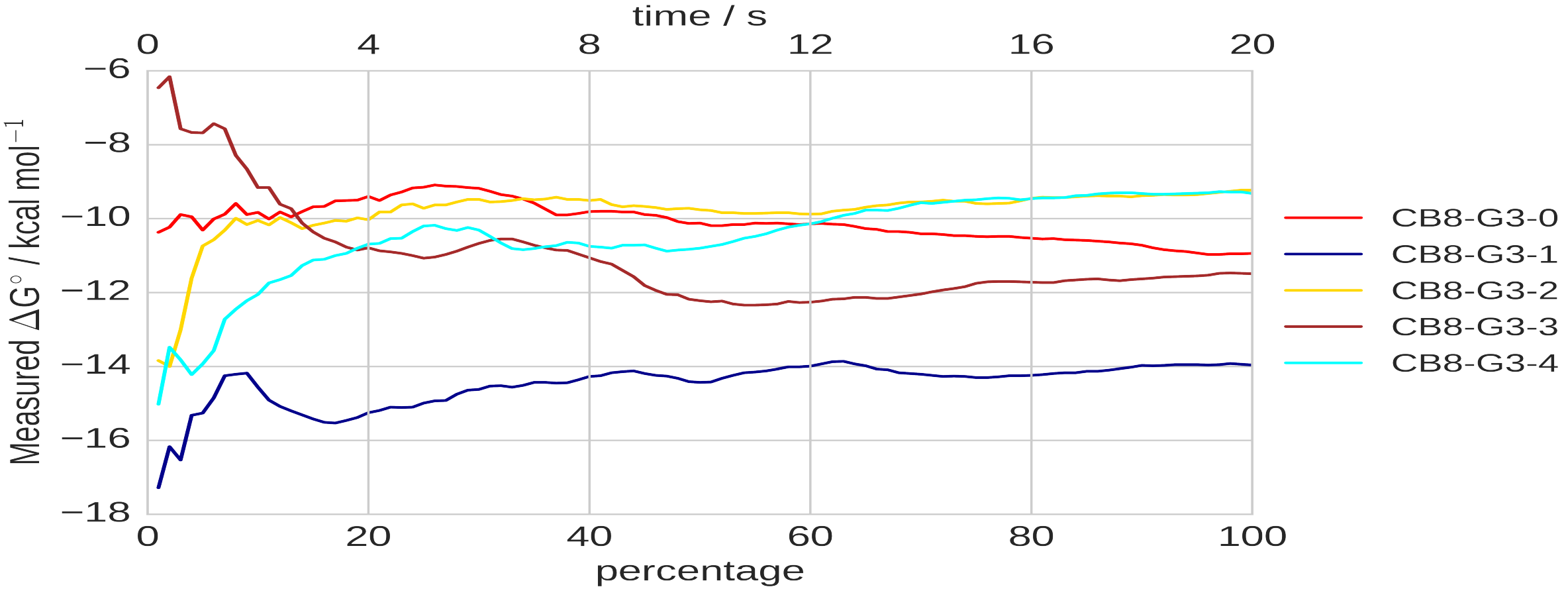
Comparison buffer-nobuffer in CB8 host-guest systems-Model C



SAMPLing results – CB8-G3

SAMPL6 Model C: $\Delta G^{\circ}_{Model\ C}$ = -12.96 kcal/mol

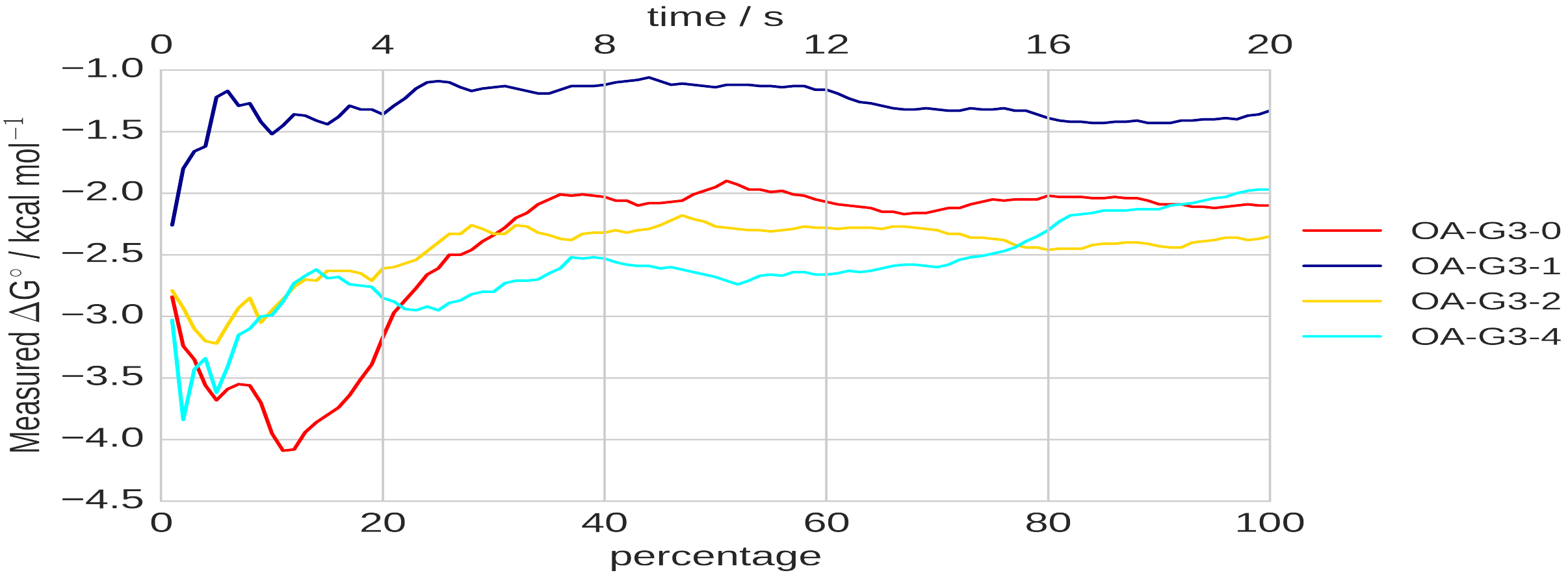
SAMPL6 experimental: $\Delta G^{\circ}_{experimental}$ = -6.45 kcal/mol



SAMPLing results – OA-G3

SAMPL6 Model C: $\Delta G^{\circ}_{Model\ C}$ = -6.38 kcal/mol

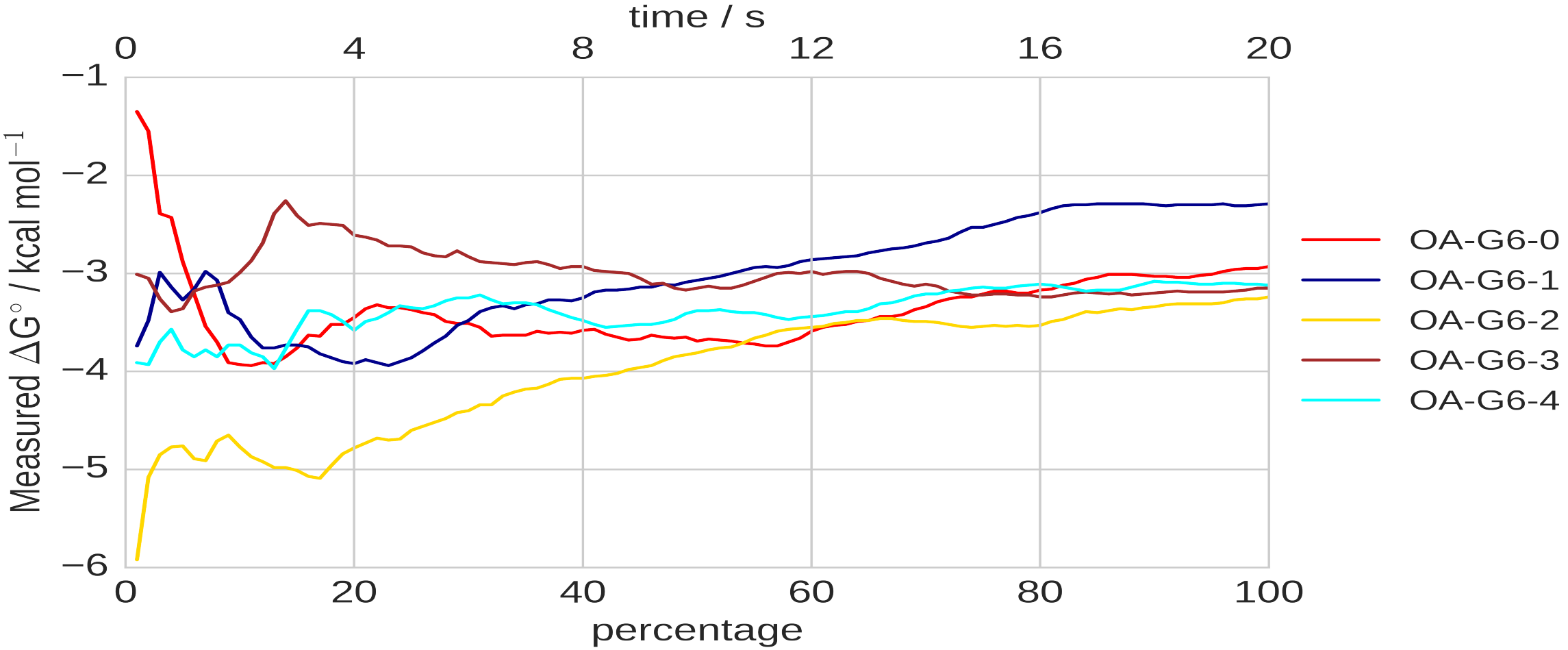
SAMPL6 experimental: $\Delta G^{\circ}_{experimental}$ = -5.18 kcal/mol



SAMPLing results – OA-G6

SAMPL6 Model C: $\Delta G^{\circ}_{Model\ C}$ = -8.08 kcal/mol

SAMPL6 experimental: $\Delta G^{\circ}_{experimental}$ = -4.97 kcal/mol



What lessons we have learned

1. Buffer did not improve the results for the OA and TEMOA host-guest systems but improved MUE in Model C for CB8 by 3 kcal/mol.

2. In general, Model C reduces the MUE but still overestimates the binding free energy and R^2 remains the same in comparison to Model A

3. Model D improved the prediction in terms of MUE but R^2 remains the same in comparison to Model C

4. In SAMPLing challenge – surprisingly no convergence in our results despite the length in our simulations and no agreement with our prediction in Model C of SAMPL6 and the experimental result

Julien Michel Group members

Julien Michel

Stefano Bosisio

Alessio De Simone

Charis Georgiou

Arun Gupta

Harris Ioannidis

Jordi Juarez-Jimenez

Salome Llabres

Cesar Mendoza Martinez

Antonia Mey

Joan Clark Nicholas

Lisa Patrick

Pattama Wapeesittipan

Marie Bluntzer

Supplementary info (1) - buffer

- Simulating the buffer in **OA** and **TEMOA** host-guest systems: We modelled the 10mM sodium phosphate buffer with NaCl by replicating the ionic strength. This results in 60mM of NaCl to mimic 10mM of Na3PO4, thus 4 Na and 4 Cl where added to each OA and TEMOA system. Initially buffer was added to the bound phase systems and the solvated guest files were created by substituting the host's heavy atoms with water molecules, to keep the same box size between both system files and the same NaCl concentration.
- Simulating the buffer in **CB8** host-guest systems: We modelled the 25mM sodium phosphate buffer with NaCl by replicating the ionic strength. This results in 150mM of NaCl to mimic 25mM of Na3PO4, thus 8 Na and 8 Cl where added to each CB8 system. Initially buffer was added to the bound phase systems and the solvated guest files were created by substituting the host's heavy atoms with water molecules, to keep the same box size between both system files and the same NaCl concentration.

Lennard Jones correction term

- The Lennard Jones dispersion correction term can be calculated from the Zwanzig relationship:

$$\Delta G^{X}_{LJLRC} = k_B T \ln < \exp\left[-\beta\left(U_{LJ,long(r)} - U_{LJ,sim(r)}\right)\right] >_X + U_{LJana}$$

- An analytical correction over an infinite size box, $U_{LJ,ana}$ is needed, which is given below:

$$U_{LJ,ana} = 8\pi\rho \sum_i \sum_{j>i} \left[\frac{\varepsilon_{ij}\sigma_{ij}^{12}}{9r_c^9} - \frac{\varepsilon_{ij}\sigma_{ij}^{6}}{3r_c^3}\right]$$

Definition of a standard state

A computational vial with definite volume

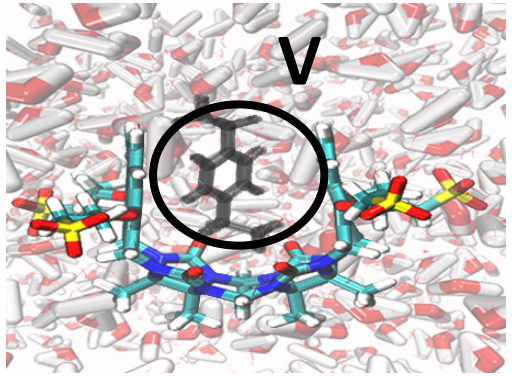
Experimentally: volume, pressure and concentration

$A + B \rightleftarrows AB$

At the equilibrium for the i-th species

$$\mu_{sol,i} = \mu^0_{sol,i} + RT\ ln\frac{\gamma_i C_i}{C^0} \Rightarrow \Delta G^0_{AB} = -RTln\left(\frac{\gamma_{AB}}{\gamma_A\gamma_B}\frac{C^0 C_{AB}}{C_A C_B}\right)$$

Computationally we already compute some terms but:



$$\Delta G_{std} = -RTln\frac{C}{C^0}$$

$$\Delta G_{std} = -RTln\frac{V}{1660\frac{\text{Å}^3}{mol}}$$

Overview of SAMPL6-OA results

OA-MODELA-NOBUFFER	OA-MODELC-NOBUFFER	OA-MODELD-NOBUFFER	OA-MODELA-BUFFER	OA-MODELC-BUFFER	OA-MODELD-BUFFER
Rsqrt: 0.63< 0.75 < 0.85	Rsqrt: 0.62 < 0.74 < 0.84	Rsqrt: 0.62 < 0.74< 0.84	Rsqrt: 0.67< 0.74< 0.80	Rsqrt: 0.64< 0.71 < 0.78	Rsqrt: 0.64 < 0.71 < 0.78
mue: 4.16< 4.41 < 4.66 kcal/mol	mue: 2.15 < 2.37< 2.60 kcal/mol	mue: 0.65< 0.82 < 1.00 kcal/mol	mue: 4.98< 5.15 < 5.31 kcal/mol	mue: 2.88 < 3.05 < 3.23 kcal/mol	mue: 0.95 < 1.09 < 1.22 kcal/mol
Kendall tau : 0.43 < 0.55< 0.64	Kendall tau : 0.43 < 0.54< 0.64	Kendall tau : 0.43 < 0.54 < 0.64	Kendall tau : 0.50 < 0.57 < 0.64	Kendall tau : 0.43 < 0.52 < 0.64	Kendall tau : 0.43 < 0.52 < 0.64

Overview of SAMPL6-TEMOA results

TEMOA-MODELA-NOBUFFER	TEMOA-MODELC-NOBUFFER	TEMOA-MODELD-NOBUFFER	TEMOA-MODELA-BUFFER	TEMOA-MODELC-BUFFER	TEMOA-MODELD-BUFFER
Rsqrt: 0.91 < 0.94 < 0.97	Rsqrt: 0.90 < 0.94 < 0.97	Rsqrt: 0.90 < 0.94< 0.97	Rsqrt: 0.89 < 0.93< 0.96	Rsqrt: 0.89 < 0.93 < 0.96	Rsqrt: 0.89< 0.93 < 0.96
mue: 4.47 < 4.66 < 4.84 kcal/mol	mue: 2.50< 2.67 < 2.86 kcal/mol	mue: 1.55< 1.73 < 1.91 kcal/mol	mue: 5.48 < 5.66 < 5.85 kcal/mol	mue: 3.234 < 3.42 < 3.59 kcal/mol	mue: 1.51 < 1.64< 1.77 kcal/mol
Kendall tau : 0.86 < 0.85 < 0.86	Kendall tau : 0.79 < 0.85 < 0.86	Kendall tau : 0.79 < 0.85 < 0.86	Kendall tau : 0.79 < 0.84 < 0.86	Kendall tau : 0.79 < 0.85 < 0.86	Kendall tau : 0.79 < 0.85 < 0.86

Overview of SAMPL6-CB8 results

CB8-MODELA-NOBUFFER	CB8-MODELC-NOBUFFER	CB8-MODELD-NOBUFFER	CB8-MODELA-BUFFER	CB8-MODELC-BUFFER	CB8-MODELD-BUFFER
Rsqrt: 0.04< 0.13 < 0.23	Rsqrt: 0.04 < 0.13 < 0.23	Rsqrt: 0.04 < 0.13 < 0.23	Rsqrt: 0.00 < 0.14< 0.35	Rsqrt: 0.00 < 0.13< 0.35	Rsqrt: 0.00 < 0.13< 0.34
mue: 6.90< 7.33 < 7.75 kcal/mol	mue: 5.57 < 6.00 < 6.43 kcal/mol	mue: 1.40 < 1.58 < 1.76 kcal/mol	mue: 4.12< 4.64 < 5.18 kcal/mol	mue: 2.97 < 3.52 < 4.09 kcal/mol	mue: 1.79 < 2.06< 2.33 kcal/mol
Kendall tau : -0.09 < 0.05 < 0.20	Kendall tau : -0.09 < 0.05 < 0.20	Kendall tau : -0.09 < 0.05 < 0.20	Kendall tau : -0.09 < 0.09< 0.31	Kendall tau : -0.09 < 0.10< 0.31	Kendall tau : -0.09 < 0.10< 0.31