Correcting Molecular Mechanics Binding Free Energies with Machine Learning Potentials

Harry Moore[1]; **Finlay Clark**[2]; Dávid Péter Kovács[1]; Mateusz Bieniek[3]; Graeme Robb[4]; Julien Michel[2]; Gábor Csányi[1]; Daniel J. Cole[3]

[1]Engineering Laboratory, University of Cambridge, Cambridge CB2 1PZ, United Kingdom
[2]EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom
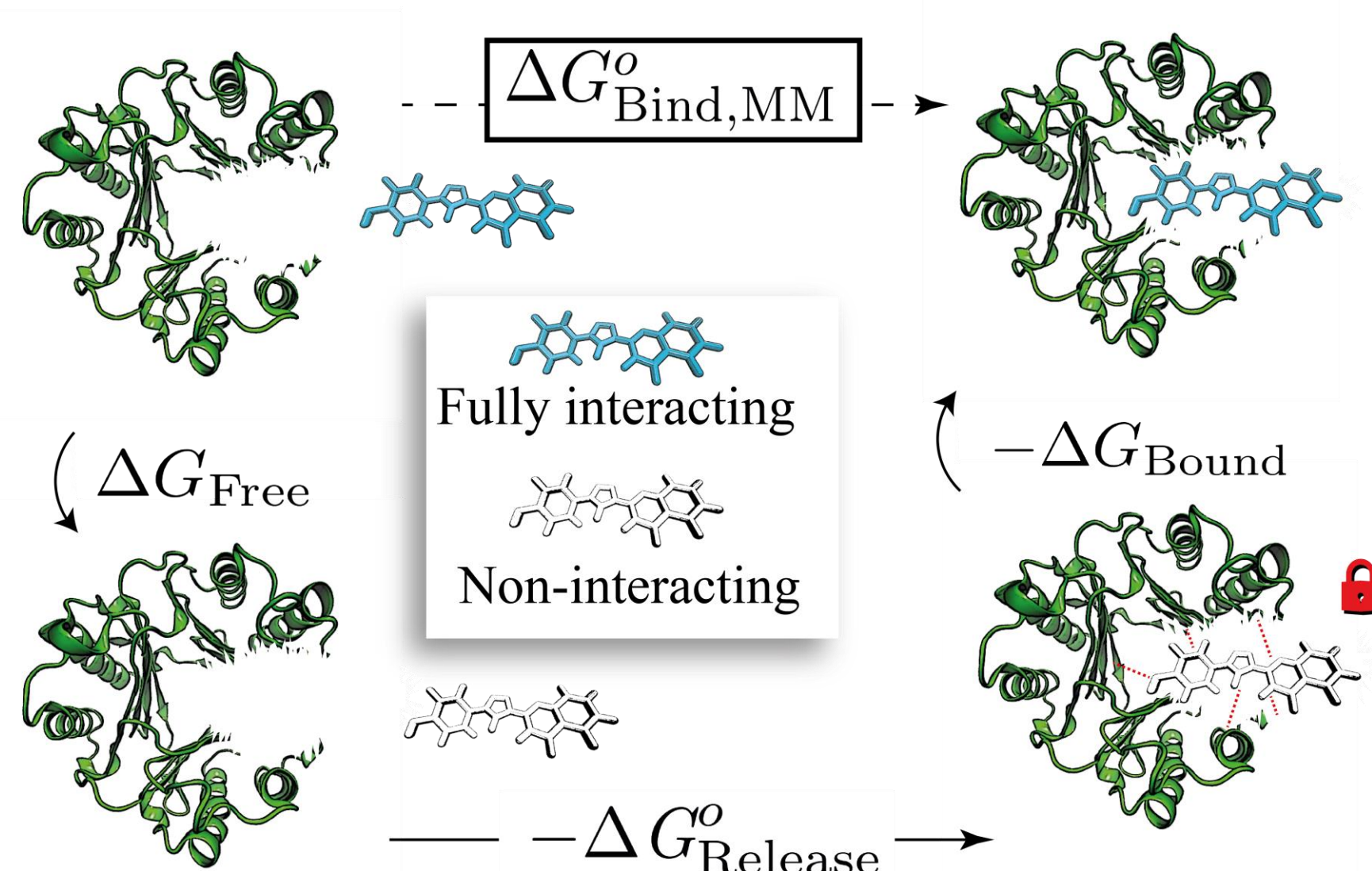[3]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
[4]Oncology R&D, AstraZeneca, Cambridge CB4 0WG, United Kingdom

THE UNIVERSITY of EDINBURGH · Newcastle University · EPSRC Engineering and Physical Sciences Research Council · AstraZeneca

Alchemical binding free energy calculations are an important tool in drug discovery

- Alchemical relative binding free energy calculations are used routinely in the hit-to-lead and lead optimisation stages of drug discovery[1]
- Alchemical absolute binding free energy calculations show promise for virtual screening

$-\Delta G^o_{\text{Bind,MM}}$

Fully interacting
Non-interacting

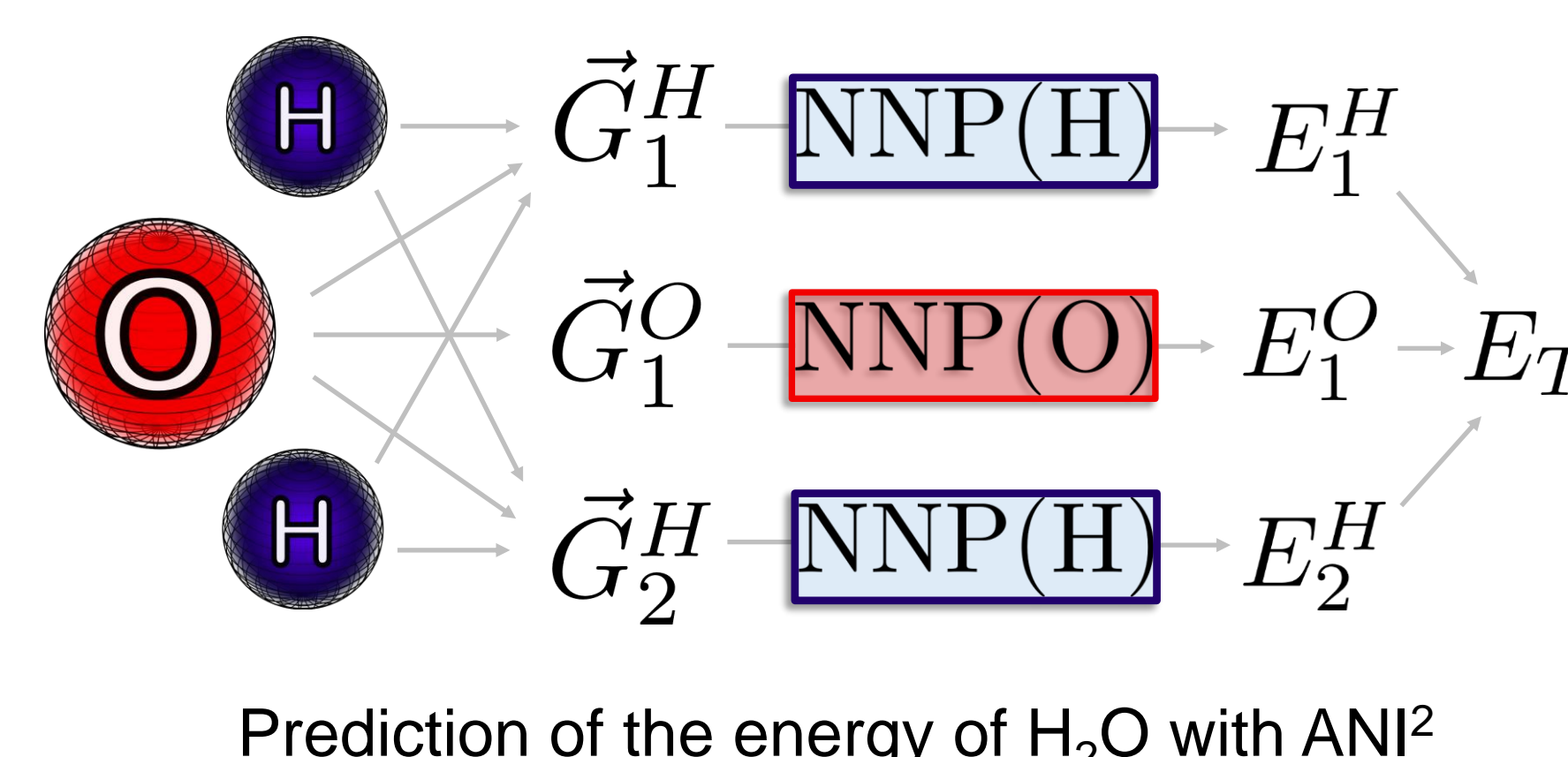ΔG_{Free} · $-\Delta G_{\text{Bound}}$ · $-\Delta G^o_{\text{Release}}$

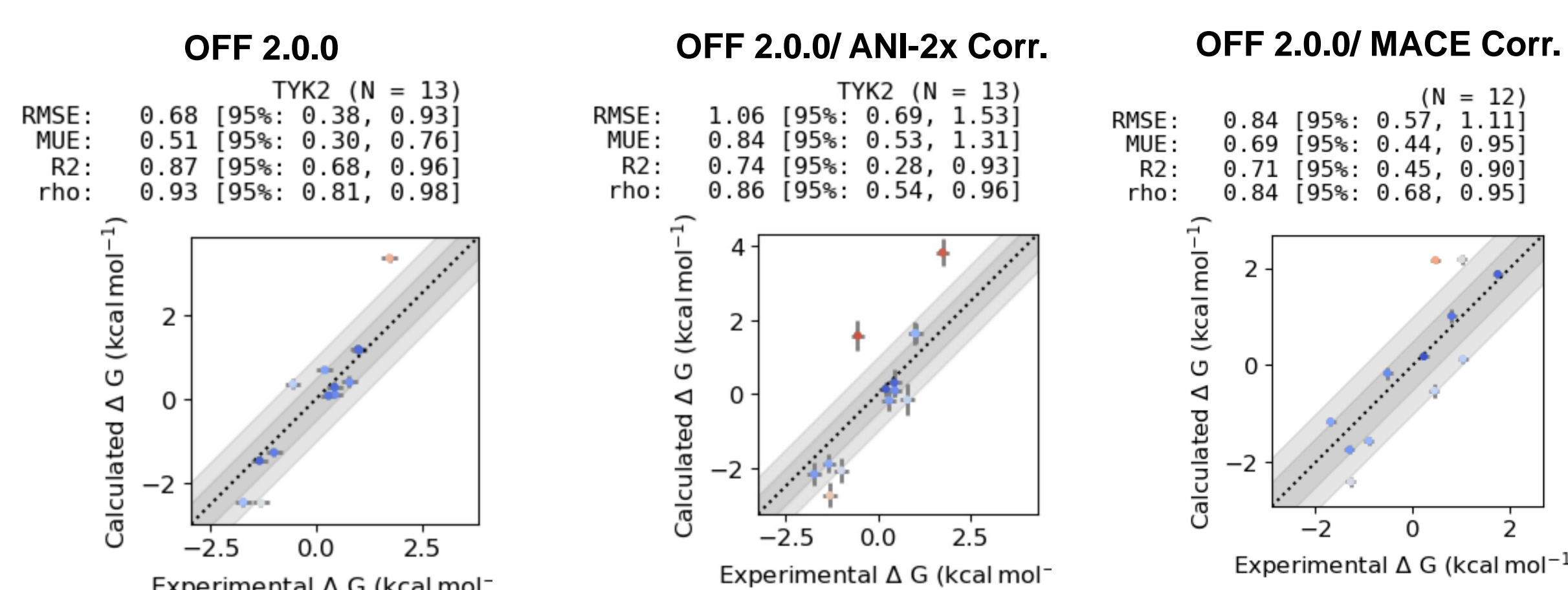## Molecular mechanics force fields produce an upper limit on accuracy of predictions

- Molecular mechanics (MM) functional form is a severe approximation

$$V(\mathbf{r}^N) = \sum_{\text{bonds}} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$
$$+ \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right)$$

- Machine learning potentials (MLPs) such as ANI-2x[2] and MACE[3] offer substantial improvements in accuracy at a much lower cost than QM methods

$\vec{G}^H_1 \rightarrow$ NNP(H) $\rightarrow E^H_1$
$\vec{G}^O_1 \rightarrow$ NNP(O) $\rightarrow E^O_1 \rightarrow E_T$
$\vec{G}^H_2 \rightarrow$ NNP(H) $\rightarrow E^H_2$

Prediction of the energy of H_2O with ANI[2]

Current results show no improvement for relative calculations and are affected by a systematic offset

OFF 2.0.0
TYK2 (N = 13)
RMSE: 0.68 [95%: 0.38, 0.93]
MUE: 0.51 [95%: 0.30, 0.76]
R2: 0.87 [95%: 0.68, 0.96]
rho: 0.93 [95%: 0.81, 0.98]

OFF 2.0.0/ ANI-2x Corr.
TYK2 (N = 13)
RMSE: 1.06 [95%: 0.69, 1.53]
MUE: 0.84 [95%: 0.53, 1.31]
R2: 0.74 [95%: 0.28, 0.93]
rho: 0.86 [95%: 0.54, 0.96]

OFF 2.0.0/ MACE Corr.
(N = 12)
RMSE: 0.84 [95%: 0.57, 1.11]
MUE: 0.69 [95%: 0.44, 0.95]
R2: 0.71 [95%: 0.45, 0.90]
rho: 0.84 [95%: 0.68, 0.95]

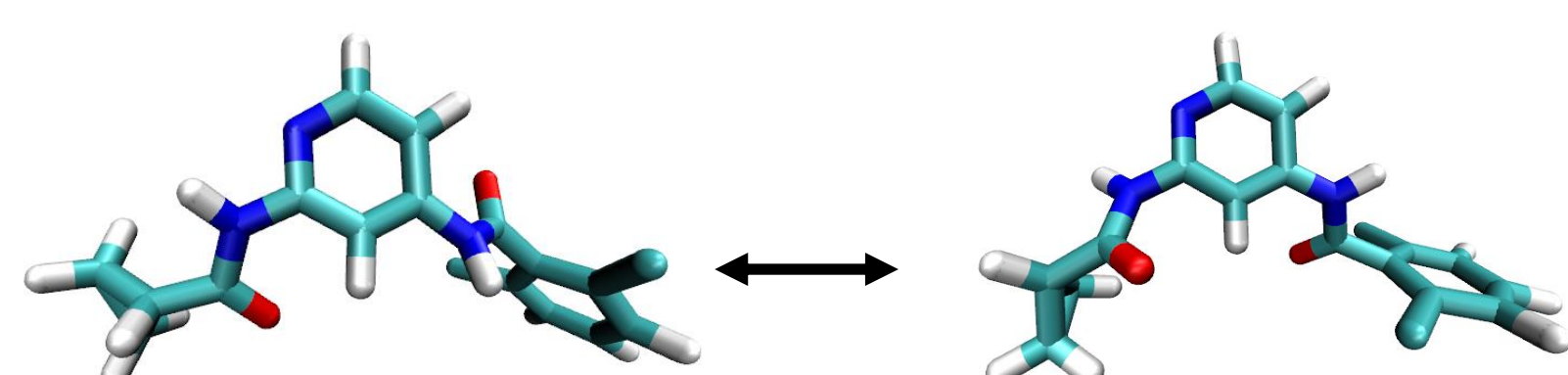(Plots: Calculated Δ G (kcal mol⁻¹) vs Experimental Δ G (kcal mol⁻¹))

- Relative binding free energy results for TYK2 (ff14SB/TIP3P/OFF 2.0.0) do not change significantly when corrected with ANI-2x or MACE
- Absolute binding free energy results for TYK2[5] would be significantly worsened by large (~ 4 kcal mol⁻¹) systematic offsets from the corrections towards weaker free energies of binding. These cancel out in relative calculations.
- Absolute binding free energy results for MIF / MIF180 also substantially worsened

Method	ΔG^o_{Bind} / kcal mol⁻¹
Experiment[6]	-8.98 ± 0.28
MM	-9.96 ± 0.34
MM/ML	-12.07 ± 0.62

ML
MM

$\Delta G^o_{\text{Bind,MM/ML}}$

Machine learning potentials can be used to correct the results of molecular-mechanics-based calculations

- Rufa et al.[4] calculated the **relative binding free energies** for a series of ligands to non-receptor tyrosine kinase TYK2 using the ff14SB/TIP3P/OFF 1.0.0 force fields.
- **Correcting the ligand intramolecular energies** with ANI-2x significantly reduced RMSE from 0.97 [95% CI: 0.68, 1.21] to 0.47 [95% CI: 0.31, 0.63] kcal mol⁻¹

$-\Delta G^o_{\text{Bind,MM/ML}}$

$\Delta G_{\text{MM/ML} \rightarrow \text{MM,Free}}$ · $-\Delta G_{\text{MM/ML} \rightarrow \text{MM,Bound}}$

Fully interacting, MLP intramolecular energies

ΔG_{Free} · $-\Delta G_{\text{Bound}}$ · $-\Delta G^o_{\text{Release}}$

- We attempt to correct binding free energies using ANI-2x or a transferrable MACE model, using a replica-exchange equilibrium sampling protocol implemented using OpenMM 8, OpenMM-ML, and OpenMMTools

Differences in torsional sampling are not responsible for the systematic offset
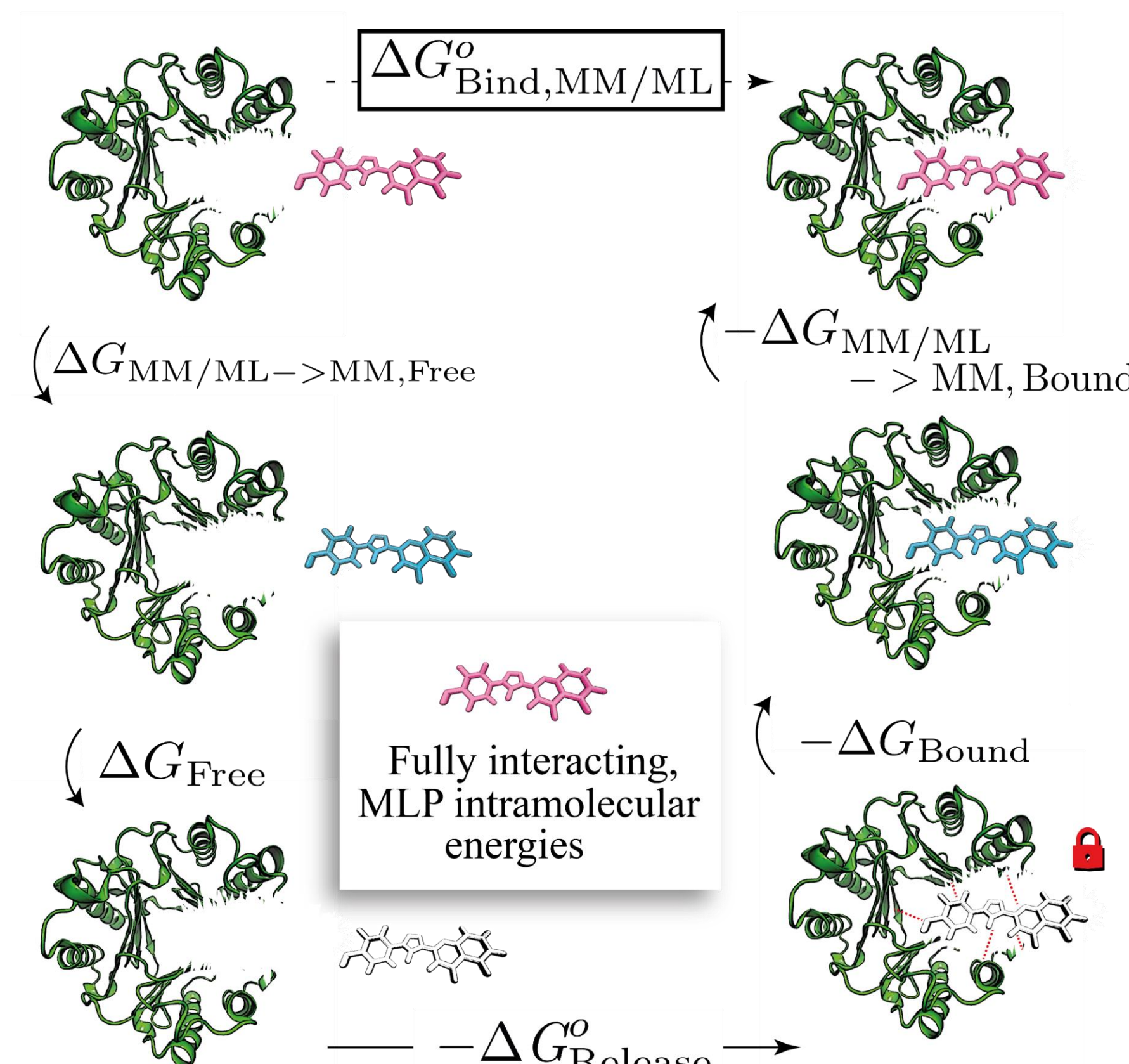
- Substantial differences in torsional distributions between OFF1 and ANI-2x were highlighted by Rufa et al.
- Rerunning a calculation using an MM force field with torsions re-fit to ANI-2x did not decrease the systematic offset observed for the TYK2 ligands
- Current scheme uses point charges and LJ terms from the MM force field to describe the intermolecular interactions throughout. In future, improved embedding schemes such as the electrostatic embedding scheme proposed by Zinojev[7] may reduce errors.

Conclusions

- Corrections with ANI-2x and MACE do not significantly change relative binding free energy results for TYK2
- Corrections for TYK2 show a large systematic offset towards weaker binding free energies which would dramatically worsen absolute binding free energy results
- Accuracy may be improved by moving beyond mechanical embedding[7]

References

1. Z. Cournia, B. Allen and W. Sherman, *J. Chem. Inf. Model.*, 2017, **57**, 2911–2937.
2. C. Devereux, J. S. Smith, K. K. Huddleston, K. Barros, R. Zubatyuk, O. Isayev and A. E. Roitberg, *J. Chem. Theory Comput.*, 2020, **16**, 4192–4202.
3. I. Batatia, D. P. Kovacs, G. N. C. Simm, C. Ortner, and G. Csanyi, "MACE: Higher order equivariant message passing neural networks for fast and accurate force fields," in *Advances in Neural Information Processing Systems*, ed. A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, 2022.
4. D. A. Rufa, H. E. Bruce Macdonald, J. Fass, M. Wieder, P. B. Grinaway, A. E. Roitberg, O. Isayev and J. D. Chodera, *bioRxiv*, 2020, DOI: 10.1101/2020.07.29.227959.
5. Y. Khalak, G. Tresadern, M. Aldeghi, H. M. Baumann, D. L. Mobley, B. L. de Groot and V. Gapsys, *Chem. Sci.*, 2021, **12**, 13958–13971.
6. J. A. Cisneros, M. J. Robertson, M. Valhondo and W. L. Jorgensen, *J. Am. Chem. Soc.*, 2016, **138**, 8630–8638.
7. K. Zinovjev, *J. Chem. Theory Comput.*, 2023, **19**, 1888–1897.