

# Chemical accuracy for ligand-receptor binding Gibbs energies through multi-level SQM/QM calculations

Froze Jameel<sup>a</sup> and Matthias Stein<sup>\*a</sup>

<sup>a</sup> Max Planck Institute for Dynamics of Complex Technical Systems, Molecular Simulations and Design Group, Sandtorstrasse 1, 3910f Magdeburg, Germany.

\*Corresponding author: matthias.stein@mpi-magdeburg.mpg.de

## Abstract

Calculating the Gibbs energies of binding of ligand-receptor systems with a thermochemical accuracy of  $\pm 1$  kcal/mol is challenge to computational approaches. A semi-empirical quantum chemical exploration of the conformational space of host, ligand and complexes followed by a multi-level refinement of poses in terms of electronic energies and solvation is able to give Gibbs energies of binding of drug molecules to CB[8] and  $\beta$ -CD macrocyclic receptor with such an accuracy. The accurate treatment of a small number of poses outperforms system-specific force-matching and alchemical transfer model approaches without an extensive sampling and intergration.

## Introduction

The calculation of ligand-receptor Gibbs energies of binding with thermochemical accuracy is still a major challenge for state-of-the art computational approaches<sup>1, 2</sup> for example in drug discovery<sup>3</sup>. An accuracy of, at most, 1-2 kcal/mol can be achieved from quantitative modelling using force fields and extensive sampling techniques to compute relative Gibbs energies of binding.<sup>3</sup> In particular, charged guests were a challenge to the accuracy in SAMPL7.<sup>4</sup> Here, we show that this chemical accuracy can also be obtained using a combination of a fast tight-binding quantum chemical exploitation of the conformational space, plus a systematic and sequential refinement of describing solvation and interaction energies.<sup>5, 6</sup> The semi-empirical quantum chemical (SQM) GFN2 Hamiltonian allows an efficient exploration of the conformational space of complex molecular systems without the need for a re-parametrization of interaction terms even for non-standard binding situation, such as open-shell transition metal complexes.<sup>7, 8</sup> SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) are blind challenges to validate and improve computational methods as predictive tools in drug design.

For blind prediction of ligand-receptor Gibbs energies of binding, macrocyclic containers such as cucurbit[n]urils (CB[n])<sup>9, 10</sup> and cyclodextrins (CDs)<sup>11, 12</sup> with unreleased experimental data are sometimes chosen.

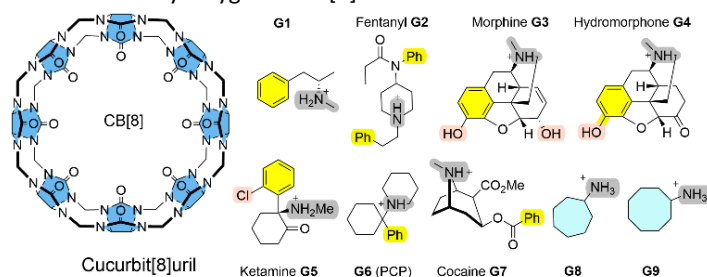
Here, we systematically refine the calculated Gibbs energies of binding from SQM calculations for the 'drugs of abuse' molecules to CB[8] of the SAMPL8 host-guest challenge. This challenge focused on binding of this host to nine drug molecules including morphine, hydromorphone, methamphetamine, cocaine, and others (see **Figure 1**).<sup>13</sup> It also included previously considered cycloheptanamine and cyclooctanamine (**G8** and **G9**). Experimental data were obtained from Isothermal Titration (ITC) and NMR spectroscopy.<sup>14</sup>

In addition, we address the calculation of Gibbs energies of binding of phenothiazine drug molecules to the  $\beta$ -cyclodextrin receptor, which was part of SAMPL9.

## Results and discussion

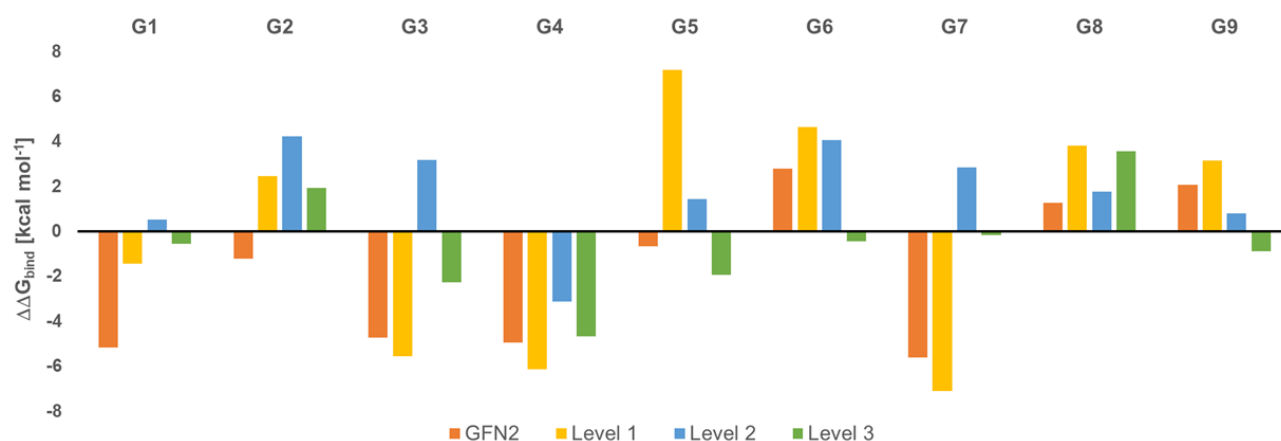
### Cucurbit[8]uril and drugs of abuse Gibbs energies of binding

Curcurbituril[8] is a host with only a few conformers accessible thus enabling a fast sampling of host conformations.<sup>15</sup> However, with certain force fields, cucurbituril hosts have been observed to collapse during simulations. CB[8] and guest molecules form 1:1 complexes with experimental Gibbs energies of binding in a range between -7.05 (for **G1**) and -14.07 kcal/mol (for **G6**).<sup>14</sup> High affinity measurements have attributed this to an interaction between the guest's ammonium group interacting and the carbonyl oxygen of CB[8].



**Figure 1.** CB[8] host and guest structures **G1** – **G9**.

**Figure 2** shows the deviation of calculated Gibbs energies of binding of CB[8] to ligands **G1**–**G9** from experiment. The conformer-rotamer ensemble from our SAMPL8 GFN2-xTB/MetaMD/GBSA work<sup>15</sup> was refined using a systematic



**Figure 2.** Deviation of calculated Gibbs energies of binding from experiment ( $\Delta\Delta G_{\text{bind}}$ , in kcal/mol) for **G1** to **G9** ligands to the CB[8] receptor.

improvement of description of electronic energies and solvation. We are using a three level approach with increasing refinement thresholds to reduce the number of structures to be considered at the next level (see Methods and ESI for more details). The original GFN2 CRE had a mean absolute deviation (MAD) of 3.16 kcal/mol from experiment which is already close to the top-ranked force matching approach with a MAD of 2.03 kcal/mol in SAMPL8.<sup>13</sup>

Since calculations at ‘Level 0’ are mere approximate energies and not Gibbs energies of binding to remove high-lying complexes, they cannot be directly compared with experiment and are not discussed further. A negative  $\Delta\Delta G_{\text{bind}}$  indicates an overbinding in **Figure 2**. GFN2 systematically overestimates the Gibbs energies of binding (only for cyclic amines **G8** and **G9** an underbinding by 1-2 kcal/mol is seen). The composite mGGA method r2scan-3c at Level 1 gives significantly better Gibbs energies of binding even for single-point calculations. It was originally found to outperform hybrid functionals in terms of conformational energies and non-covalent interactions at a significantly lower computational cost.<sup>16</sup> However, from **Figure 2** it becomes apparent that the structural re-optimization of GFN2 poses significantly reduces the deviation from experiment (MAD decreases from 4.6 kcal/mol at Level 1 to 2.45 kcal/mol at Level 2). The hybrid meta-GGA PW6B95 at Level 3 gives an additional reduction of MAD by  $\sim 1$  kcal/mol (see **Table 1**) and a final MSE of -0.6 kcal/mol.

**Table 1.** Analysis of error of calculated Gibbs energies of binding in kcal/mol for SAMPL8 entries.

	MSE <sup>a</sup>	SEM <sup>b</sup>	MAD <sup>c</sup>
<b>GFN2</b>	-1.79	1.13	3.16
<b>Level 0</b>	-16.84	1.71	16.84
<b>Level 1</b>	+0.11	1.78	4.61
<b>Level 2</b>	+1.75	0.76	2.45
<b>Level 3</b>	-0.60	0.79	1.82

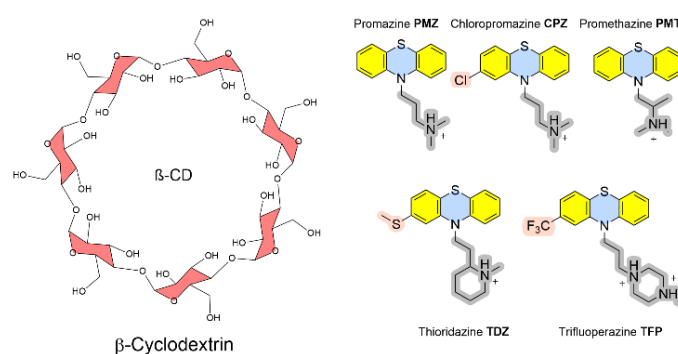
<sup>a</sup>mean signed error, <sup>b</sup>standard error of mean, <sup>c</sup> mean absolute deviation.

For the CB[8] host, the top-performer was using classical bonding and non-bonding parameters obtained via QM force-matching (FM) methods, QM-derived atomic charges and bonded parameters to yield a final MAD of 2.03 kcal/mol.<sup>13, 17</sup>

Derivation of QM charges and the molecular force matching was computationally expensive and required 10,000s of DFT force calculations. As an alternative approach with much less extensive sampling and a systematic refinement of QM interaction energies, a MSE of -0.60 kcal/mol and a MAD of 1.82 kcal/mol clearly outperforms the FM and MD ansatz.

#### Gibbs energies of binding of $\beta$ -cyclodextrin and drug molecules

The recent SAMPL competition included the prediction of host-guest Gibbs energies of binding between  $\beta$ -cyclodextrin and five phenothiazine-based antipsychotic drugs (see **Figure 3**).<sup>18</sup> In  $\beta$ -cyclodextrin, seven glucose subunits are  $\alpha$ -1,4 linked to give a cone-shaped host with  $\sim 6$  Å diameter, a hydrophobic interior and a slightly hydrophilic exterior surface. Cyclodextrins bind a range of guest molecules in aqueous solution by both hydrophobic and polar interactions and confer solubility, stability, and bioavailability to drug molecules and are thus used as drug carriers.<sup>11, 12</sup>



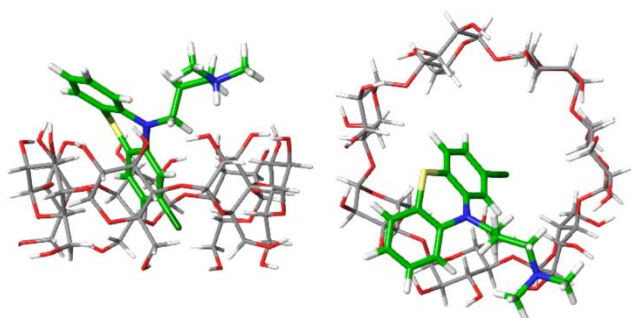
**Figure 3.** Structures of the  $\beta$ -cyclodextrin host and phenothiazine-derived guest molecules

Gibbs energies of binding were obtained from ITC and NMR to characterize the non-covalent interactions and revealed the formation of 1:1 complexes.<sup>19</sup> Gibbs energies of binding are in a

very narrow range between -4.5 and -5.7 kcal/mol and thus represent a challenge to current computational methods.

Some of the phenothiazines (CPZ, TDZ, TFP) are substituted at the phenothiazine entity to give an asymmetric guest molecule. Increasing hydrophobicity at position 4 (-Cl and -CF<sub>3</sub> groups), varying alkyl chain lengths, branching and different terminal tertiary amines were found to be critical determinants for their biological activity.<sup>20</sup>

The conformational search yielded highly ranked poses in agreement with NMR studies. Part of the phenothiazine moiety is located at the secondary face of the host and part of the phenothiazine moiety penetrates deep into the host's cavity. Only the relatively bulky side-chains of TDZ and TFP were sufficiently locked to generate definite nuclear Overhauser effect (NOE) signals to suggest an interaction with the hydroxyl group. For all phenothiazine ligands, binding poses in agreement with structural interpretation from NMR were obtained (see **Figure 4**). They all reveal a bifurcated ammonium-hydroxyl interaction and an incorporation of the largely hydrophobic phenothiazine ring into the hydrophobic  $\beta$ -CD binding cavity. In medicinal chemistry, introduction of a chlorine atom is frequently used to increase the lipophilicity of drug compounds.<sup>19</sup> Promazine (PMZ) and chlorpromazine (CPZ) differ only in the replacement of a hydrogen by a chlorine, but chlorpromazine binds only slightly tighter than promazine (0.5 kcal/mol in experiment vs. 1.3 kcal/mol in calculations). This suggests that introducing the chlorine substituent is not significantly stabilizing ligand-receptor binding but may be beneficial for the drug's bioavailability.



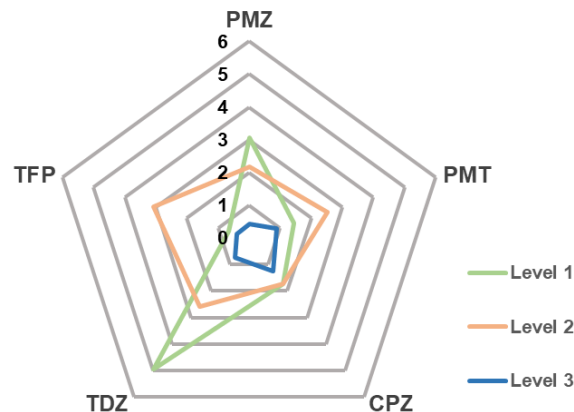
**Figure 4.** Top-ranked binding pose of CPZ binding to  $\beta$ -CD. Left: side view, right: top view.

**Table 2** gives the analysis of errors of phenothiazine binding to  $\beta$ -CD.

**Table 2.** Analysis of deviation of calculated (S)QM Gibbs energies of binding in kcal/mol for phenothiazine drug molecules to  $\beta$ -CD.

	MSE <sup>a</sup>	SEM <sup>b</sup>	MAD <sup>c</sup>
GFN2	-18.4	2.3	18.4
Level 0	+4.8	1.5	4.8
Level 1	+2.1	1.9	2.4
Level 2	+2.4	0.4	2.4
Level 3	+0.4	0.7	0.7

<sup>a</sup>mean signed error, <sup>b</sup>standard error of mean, <sup>c</sup>mean absolute deviation.



**Figure 5.** MAD from experiment in kcal/mol of calculated Gibbs energies of binding of phenothiazine drug molecules to  $\beta$ -CD.

As for the CB[8] receptor binding challenge, GFN2 generated structures are very plausible but the deviation in energies from experiment is even larger. Calculations at Level 0 and Level 1 reduce the MSE and MAD significantly (see **Table 2**). Structural optimizations at Level 2, however, do not lead to systematically lower errors. High-level double-hybrid DFT calculation generate very accurate Gibbs energies of binding with chemical accuracy (MSE +0.4 kcal/mol, MAD 0.7 kcal/mol). **Figure 5** shows that the MAD from experimental data can systematically be reduced from Levels 1 to 3.

For  $\beta$ -CD phenothiazine binding, the best SAMPL submission was using an alchemical transfer model (ATM)<sup>21, 22</sup> with a proprietary ligand force field to give a MSE of -0.9 kcal/mol and a MAD of 1.6 kcal/mol. In total, 64 individual alchemical Gibbs energy calculations and hundreds of nanoseconds of simulation needed to be performed. Our approach gives results that are outperforming the most accurate SAMPL submissions.

## Computational Details

Ligands were manually positioned inside the host entities. The complexes were minimized using GFN2. Ensembles of non-covalent binding poses were generated using CREST<sup>5</sup> and an ALPB continuum solvent representation for water. The CENSO workflow<sup>6</sup> with SPH thermostatical corrections and the modified rigid-rotor harmonic oscillator (mRRHO) was used.<sup>23</sup> Solvent contributions were incorporated at various levels such as ALPB, DCOSMO-RS and COSMO-RS (see ESI for details).

- Level 0: SP B97-D/def2-SV(P) + E<sub>solv</sub>(GFN2-ALPB)
- Level 1: SP r2scan-3c/def2-mTZVPP + E<sub>solv</sub>(GFN-ALPB) + G<sub>mRRHO</sub>(GFN2-ALPB-SPH)
- Level 2: OPT r2scan-3c/def2-mTZVPP/DCOSMO-RS + G<sub>solv</sub>(COSMO-RS) + G<sub>mRRHO</sub>(GFN2-ALPB-SPH)
- Level 3: SP PW6B95-D/def2-TZVPD + G<sub>solv</sub>(COSMO-RS) + G<sub>mRRHO</sub>(GFN2-ALPB-SPH)

Since calculations at 'level 0' are mere energies of binding, they cannot be directly compared with experiments and are not discussed in the main text.

## Conclusions

The quantum chemical refinement of a moderate number of SQM poses in combination with an increasing level of description of electronic energies and (implicit) solvation is able to provide very accurate Gibbs energies of binding of drug molecules to CB[8] and  $\beta$ -CD receptors. Machine learning (ML) approaches are now alternatives to obtain such accurate Gibbs energies of binding. For example, for the pillar[n]arene WP6<sup>24</sup> host-guest binding challenge,<sup>21</sup> ML results trained on extensive experimental datasets were superior.<sup>25</sup> For realistic host-guest complexes<sup>26</sup>, however, training data may be scarce. Here, a ML framework to control the error of DFT calculations may be more appropriate.<sup>27</sup> Our approach is able to provide such computational Gibbs energies of binding that are of comparable accuracy as experimental. It does not require a system-specific force matching or force field parametrization. At every step in this systematic workflow, it also allows the control of accuracy of results and possible ranges of errors

## Author Contributions

FJ: Investigation, Formal analysis, Writing – original draft, Data curation, Methodology; MS: Conceptualization, Data curation, Investigation, Validation, Writing - original draft, Writing - review & editing.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

All coordinates and calculated Gibbs energies are freely available from zenodo.org doi:10.5281/zenodo.10657701.

## Acknowledgements

We thank the Max Planck Society for Advancement of Science (MPG) for financial support. This work is part of the Research Initiative "SmartProSys: Intelligent Process Systems for the Sustainable Production of Chemicals" funded by the Ministry for Science, Energy, Climate Protection and the Environment of the State of Saxony-Anhalt. We appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley (UC Irvine).

## References

1. C. E. M. Schindler, H. Baumann, A. Blum, D. Böse, H.-P. Buchstaller, L. Burgdorf, D. Cappel, E. Chekler, P. Czodrowski, D. Dorsch, M. K. I. Eguida, B. Follows, T. Fuchß, U. Grädler, J. Gunera, T. Johnson, C. Jorand Lebrun, S. Karra, M. Klein, T. Knehans, L. Koetzner, M. Krier, M. Leiendecker, B. Leuthner, L. Li, I. Mochalkin, D. Musil, C. Neagu, F. Rippmann, K. Schiemann, R. Schulz, T. Steinbrecher, E.-M. Tanzer, A. Unzue Lopez, A. Viacava

2. Follis, A. Wegener and D. Kuhn, *Journal of Chemical Information and Modeling*, 2020, **60**, 5457-5474.
3. D. L. Mobley and M. K. Gilson, *Annual Review of Biophysics*, 2017, **46**, 531-558.
4. Z. Cournia, B. Allen and W. Sherman, *Journal of Chemical Information and Modeling*, 2017, **57**, 2911-2937.
5. M. Amezcua, L. El Khoury and D. L. Mobley, *Journal of Computer-Aided Molecular Design*, 2021, **35**, 1-35.
6. S. Grimme, *Journal of Chemical Theory and Computation*, 2019, **15**, 2847-2862.
7. S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *The Journal of Physical Chemistry A*, 2021, **125**, 4039-4054.
8. J. Barrera, H. H. Haeri, J. Heinrich, M. Stein, D. Hinderberger and N. Kulak, *Dalton Transactions*, 2023, **52**, 3279-3286.
9. J. Heinrich, K. Bossak-Ahmad, M. Riisom, H. H. Haeri, T. R. Steel, V. Hergl, A. Langhans, C. Schattschneider, J. Barrera, S. M. F. Jamieson, M. Stein, D. Hinderberger, C. G. Hartinger, W. Bal and N. Kulak, *Chemistry – A European Journal*, 2021, **27**, 18093-18102.
10. J. Lagona, P. Mukhopadhyay, S. Chakrabarti and L. Isaacs, *Angewandte Chemie International Edition*, 2005, **44**, 4844-4870.
11. S. Liu, C. Ruspic, P. Mukhopadhyay, S. Chakrabarti, P. Y. Zavalij and L. Isaacs, *Journal of the American Chemical Society*, 2005, **127**, 15959-15967.
12. M. V. Rekharsky and Y. Inoue, *Chemical Reviews*, 1998, **98**, 1875-1918.
13. K. Uekama, F. Hirayama and T. Irie, *Chemical Reviews*, 1998, **98**, 2045-2076.
14. M. Amezcua, J. Setiadi, Y. Ge and D. L. Mobley, *Journal of Computer-Aided Molecular Design*, 2022, **36**, 707-734.
15. S. Murkli, J. Klemm, A. T. Brockett, M. Shuster, V. Briken, M. R. Roesch and L. Isaacs, *Chemistry – A European Journal*, 2021, **27**, 3098-3105.
16. E. Boz and M. Stein, *International Journal of Molecular Sciences*, 2021, **22**, 3078.
17. S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *The Journal of Chemical Physics*, 2021, **154**.
18. P. S. Hudson, F. Aviat, R. Meana-Pañeda, L. Warrensford, B. C. Pollard, S. Prasad, M. R. Jones, H. L. Woodcock and B. R. Brooks, *Journal of Computer-Aided Molecular Design*, 2022, **36**, 263-277.
19. M. J. Ohlow and B. Moosmann, *Drug Discovery Today*, 2011, **16**, 119-131.
20. B. Andrade, A. Chen and M. K. Gilson, *Physical Chemistry Chemical Physics*, 2024, **26**, 2035-2043.
21. J. M. Ford, W. C. Prozialeck and W. N. Hait, *Molecular Pharmacology*, 1989, **35**, 105.
22. M. Amezcua, J. Setiadi and D. L. Mobley, *ChemRxiv*, 2023, DOI: 10.26434/chemrxiv-2023-nv011
23. S. Khuttan, S. Azimi, J. Z. Wu, S. Dick, C. Wu, H. Xu and E. Gallicchio, *Physical Chemistry Chemical Physics*, 2023, **25**, 24364-24376.
24. S. Grimme, *Chemistry – A European Journal*, 2012, **18**, 9955-9964.
25. T. Ogoshi, T.-a. Yamagishi and Y. Nakamoto, *Chemical Reviews*, 2016, **116**, 7937-8002.
26. O. T. Unke, S. Chmiela, H. E. Saucedo, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Chemical Reviews*, 2021, **121**, 10142-10186.
27. R. Sure and S. Grimme, *Journal of Chemical Theory and Computation*, 2015, **11**, 3785-3801.

27. M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller and K. Burke, *Nature Communications*, 2020, **11**, 5223.