Supplementary Information for

Sorting Drug Conformers in Enzyme Active Sites: The XTB Way

Bun Chan,*[,a,b] William Dawson,[b] and Takahito Nakajima[b]

[a] Graduate School of Engineering, Nagasaki University, Bunkyo 1-14, Nagasaki 852-8521, Japan

[b] RIKEN Center for Computational Science, 7-1-26, Minatojima-minami-machi, Chuo-ku, Kobe, 650-0047, Japan

Table of Contents

Comparison of Statistical Indicators

In the main text, we generally use the mean absolute deviation (MAD) for the assessment of the quantum chemistry methods. Alternative indicators may offer a complement to the MAD. One matter regarding the choice is how errors might be perceived as a function of the reference value. The use of MAD may obscure seemingly large percentage deviations for small reference values. This may be clarified by normalizing the errors on a percentage basis. However, such %MADs may mask large deviations for small percentages.

For the MEI196 set, the %MAD correlates well with the MAD (Fig. S1). Indeed, the distributions for the percentage deviations resemble those for the actual deviations (Fig. S2). We have further examined the correlations of the MADs and %MADs for the subsets and find them to be generally good; the subset with the largest spread in energy (the "xtra_enzyme_models" set) has an R^2 of 0.979. This is relevant because, in the application to enzyme binding, the sum of many interactions of diverse nature often give rise to a wide range of binding energies. Overall, within the context of the present study and for the purpose of ranking the methods, the two metrics appears to be functionally equivalent.

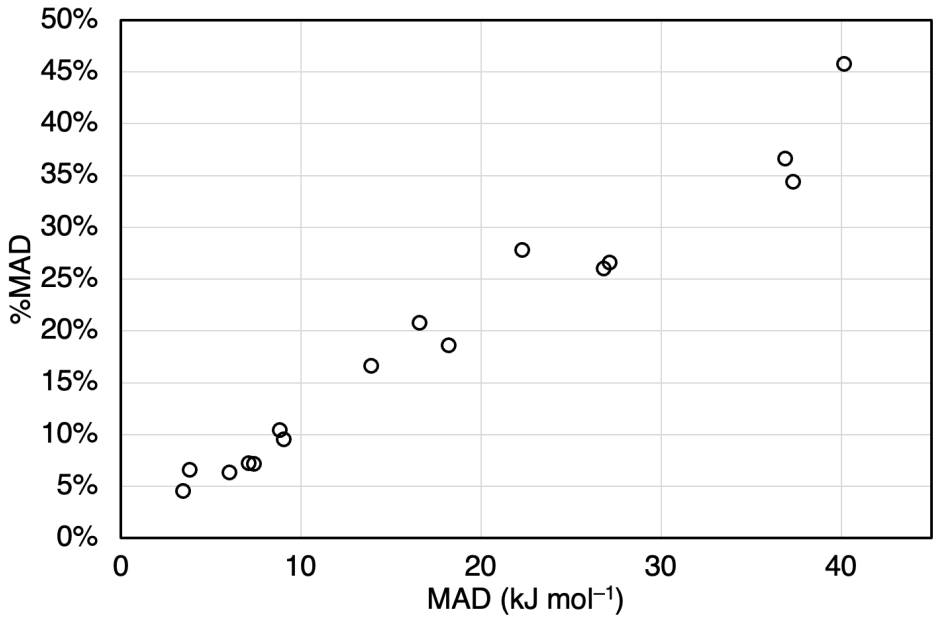

Fig. S1. The correlation between the MAD and the %MAD, i.e., MAD with each data point normalized to its reference value on a percentage basis, for the MEI196 set for all methods used in the present study (including those in Table S1 as well as Table 2 of the main text).

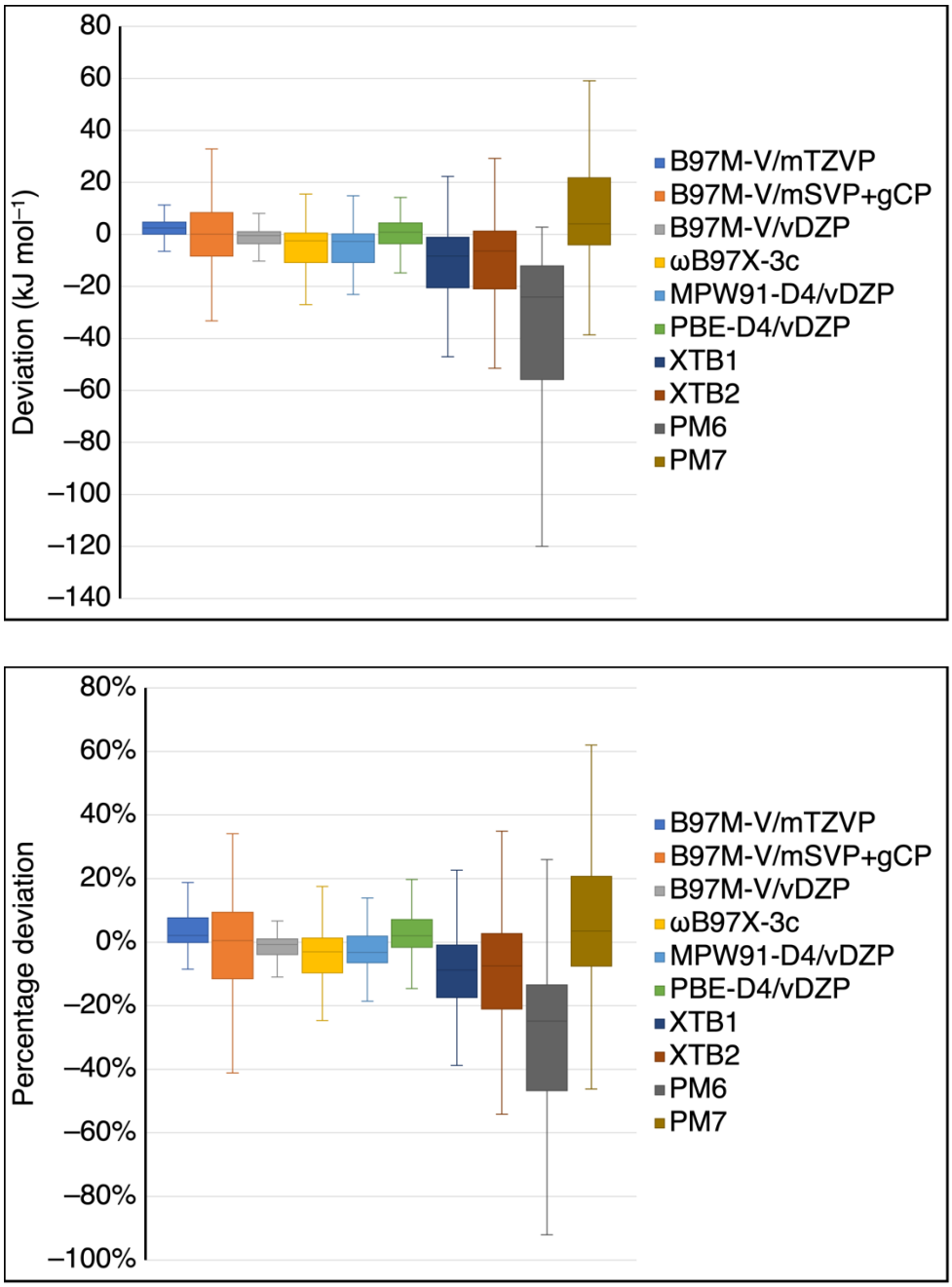

Fig. S2. Box plots that correspond to (top) Fig. 1 of the main text, and (bottom) the same figure but with percentage deviations.

Outliers for the MEI196 Set

In Fig. 1 of the main text, outliers for the various methods for the MEI196 set are omitted for the sake of clarity. Herein we provide the full box plot with the outliers included (Fig. S3). In general, the information conveyed by the outliers is consistent with that from the box plot in the absence of the outliers. For instance, among the low-cost DFT protocols with double-ζ basis sets, we can see that the B97M-V/mSVP+gCP method performs the worst, while B97M-V/vDZP performs the best. Among the semi-empirical-type methods, XTB1 edges out the other methods.

The outliers are identified in the Excel spreadsheet of the Supplementary Information. In general, they correspond to different systems for different methods. Altogether, they cover many subsets of the MEI196 set. Some exceptions are, for example, the solute–solvent-cluster-type systems such as the h2o_in_h2o and xtra_frames_c6h6 sets, from which no outliers originate. This is perhaps not surprising as these systems comprise relatively weak interactions between a neutral solute and its neutral environment and are typically associated with relatively small interaction energies. Their corresponding deviations are, in general, proportionally small in accordance with the correlation between the actual and percentage deviations (Fig. S1).

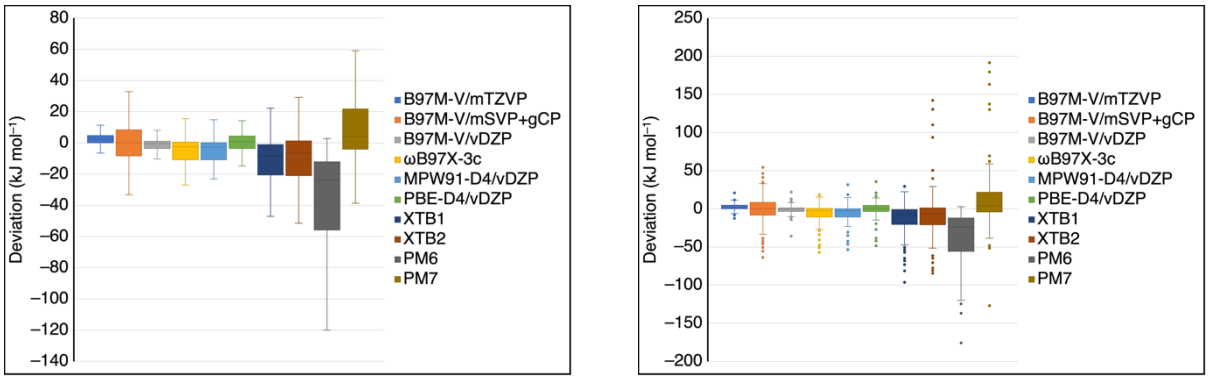

Fig. S3. Box plots that correspond to (left) Fig. 1 of the main text, and (right) the same figure but with the outliers included.

Additional Benchmark with the MEI196 Set

In the main text, we show the key results of our assessment of low-cost methods, with some subtle observations deserving further discussion. Thus, we have carried out calculations with additional methods for a more in-depth examination. The results are shown in Table S1, which also include selected results from the main text for comparison.

Table S1. Mean absolute deviation (MAD, kJ mol^{-1}) mean deviation (MD), standard deviation of the deviations (SD), and largest deviation (LD) from benchmark values for the MEI196 set of solute–solvent and drug–enzyme intermolecular interactions

method	MAD	MD	SD	LD
B97M-V/vDZP	3.5	–1.2	5.2	–35.9
*ω*B97M-V/vDZP	6.0	4.3	8.2	38.9
*ω*B97X-V/vDZP	7.4	5.3	9.9	46.4
*ω*B97X-D4/vDZP	8.8	–6.0	12.6	–57.1
XTB1	16.6	–14.0	21.4	–96.5
B97M-V/mSVP+gCP	13.9	–1.5	20.2	–63.8
PBE-D4/vDZP	7.1	–0.2	11.2	–48.4
PBE/6-31G(d)	27.1	11.1	34.3	–90.1
PBE-D4/6-31G(d)	40.2	40.1	32.0	142.7
PBE0/6-31G(d)	26.8	9.0	33.6	81.2
PBE0-D4/6-31G(d)	36.9	36.8	31.6	145.6

The B97M-V/vDZP method shows better performance for MEI196 than the arguably more advanced *ω*B97X-D4/vDZP method, with B97M-V formally being a non-hybrid DFT (rung-3) while the latter being a hybrid DFT (rung-4). While the overall better accuracy for *ω*B97X-D4 for a wider range of species and chemical properties does not warrant a better accuracy for a specific set of systems, it is nonetheless instructive to further clarify the differences.

The B97M-V and *ω*B97X-D4 methods differ in four ways, (1) they use different dispersion corrections “V” vs “D4”, (2) B97M-V is a meta-functional, signified by “M”, while *ω*B97X-D4 is not, (3) B97M-V is a non-hybrid method, while *ω*B97X-D4 is a range-separated “*ω*” hybrid “X” method, and (4) because of the various differences of 1–3, the internal parameters of the functionals are different as a result of statistical fitting in their formulations.

To unravel the effects of these differences, we examine two intermediate methods. The ωB97X-D4 and ωB97X-V methods show the difference in dispersion corrections, and we see that, in this case, the use of the V correction leads to somewhat better agreements with the reference. Further slight improvements can be seen when going from the non-meta ωB97X-V method to the meta ωB97M-V method. However, its agreement with the reference is still notably worse than that for B97M-V.

At this point, we cannot decouple the effect of hybrid vs non-hybrid from the effect of different internal parameters. However, other additional results [PBE (non-hybrid) vs PBE0 (hybrid), and equivalently PBE-D4 vs PBE0-D4] seem to suggest that hybrid methods do hold an advantage over non-hybrid equivalents. It thus appears that the parametrization of ωB97M-V may be less suitable for MEI196 than those of B97M-V.

Let us now turn our attention to the performance of the XTB1 method, which is the most accurate semi-empirical method in our assessment. While it does not match the performance of many DFT methods presented in the main text, it does come close to that for B97M-V/def2-mSVP+gCP. In this regard, it is of interest to compare with other widely used low-cost DFT protocols. We note that the 6-31G(d) has been and still is a popular choice for basis set, and we have further examined its use with the PBE and PBE0 methods with and without the D4 dispersion correction.

The results are disappointing, with the lowest MAD being 26.8 kJ mol^{-1} for PBE0. We also note that, for PBE-D4, the use of 6-31G(d) leads to a large deterioration in the accuracy (MAD = 40.2 kJ mol^{-1}) in comparison with the vDZP basis set (7.1 kJ mol^{-1}). This further illustrates the advantage of using a more advanced and higher-quality basis set as opposed to legacy ones, which may remain relevant for some applications but are increasingly becoming inferior.

Additional Assessment for Relative Binding Energies

In the main text, we have shown the good performance of XTB1 for the calculation of relative binding energies for conformers of drug molecules in their host. To further put this into perspective, we have also obtained the relative energies with alternative semi-empirical methods (Table S2 and Fig. S4). We can see that the MADs for XTB1 are smaller than those for other methods, for both absolute and relative binding energies. For the other methods, the MADs for absolute binding energies show a wide range. The corresponding MADs for relative binding energies, as expected, are generally much smaller. Among the four methods, XTB1 and XTB2 outperform the PM6 and PM7 methods considerably.

Table S2. Mean absolute deviation (kJ mol^{-1}) from B97M-V/vDZP values for absolute and relative binding energies for several sets of substrate–enzyme models, in which each system contains a set of different docking poses of the substrate for the determination of relative binding energies

species	absolute					relative				
	XTB1	XTB2	PM6	PM7	Vina	XTB1	XTB2	PM6	PM7	Vina
all	15.3	37.0	88.1	39.8	91.7	9.6	10.8	12.9	19.8	31.3
1YSG	15.2	28.4	90.8	62.0	86.5	8.9	12.7	18.7	23.9	32.3
4WMV	10.8	33.7	87.8	37.7	102.6	9.2	9.8	11.6	13.9	22.3
5MZP	21.9	36.4	84.8	29.8	83.7	7.4	8.4	16.1	9.9	61.1
7E2Y	19.6	43.1	88.1	38.0	92.3	10.0	6.2	8.7	8.6	29.3
8DE4	10.0	46.3	91.0	47.3	90.6	7.7	5.3	13.4	18.4	26.1
PX12	13.6	30.6	86.1	16.0	100.9	4.4	8.8	7.2	13.0	23.1

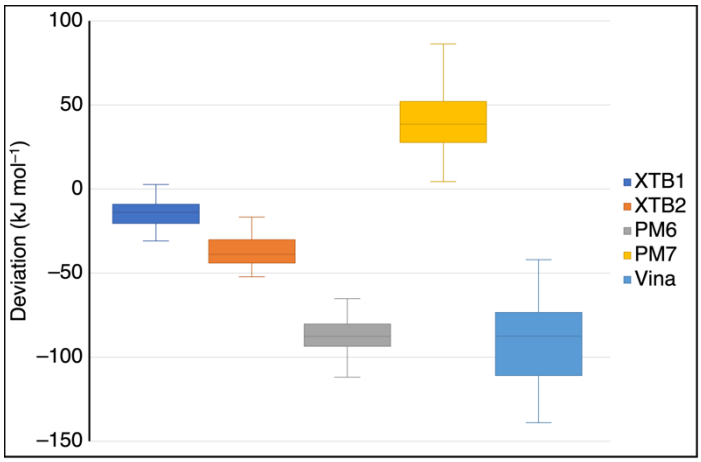

Fig. S4. Box plot for the full validation set of absolute drug–enzyme interactions.

Comparison of XTB1 and XTB2 for Thermochemical Properties

For the MEI196 set, the XTB1 method very slightly outperforms XTB2. This contrasts with the generally better accuracy of XTB2. As a sanity check, we have carried out addition comparison of the two methods for a more diverse range of thermochemical properties. We use mainly the systems that we have applied to the development of a high-level composite method (*J. Comput. Chem.* 2022, **43**, 1394). The chemical properties in the benchmark include atomization energy, reaction barrier, non-covalent interaction, artificial molecular reactions, and isodesmic-type reactions. We further complement this collection with the lattice energy of some prototypical ionic clusters (*J. Phys. Chem. A* 2023, **127**, 5652).

Admittedly, many of these systems are challenging for methods that are more sophisticated and computationally significantly more demanding than XTB, and it would be unrealistic to expect the low-cost XTB methods to provide an adequate description for many of these properties. The present assessment simply aims to compare the relative performance of the two methods. In this regard, we can see that XTB2 indeed show better agreements with the reference than XTB1. The rationale for the somewhat better performance of XTB1 in the particular case of the MEI set is unclear, and we deem the XTB2 method to remain the more robust method among the two in a more general sense.

Table S3. Mean absolute deviation (MAD, kJ mol^{-1}) for a representative range of thermochemical properties

test set	brief description	XTB1	XTB2
E0s	basic properties	140.2	100.2
P34s	heavy main-group systems	597.6	368.3
MB13	artificial molecules	249.8	231.0
BH28	barriers	46.3	39.8
plat	aliphatic isodesmic reactions	204.2	188.7
PAH	aromatic isodesmic reactions	10.5	10.1
MX35	ionic clusters	194.5	150.5