Supplementary Material

Establishing the accuracy of density functional approaches for the description of noncovalent interactions in biomolecules

Minho Kim,[1] Tim Gould,[2] Dario Rocca,[1] and Sébastien Lebègue[1, a)]

[1)] *Université de Lorraine and CNRS, LPCT, UMR 7019, Vandoeuvre-lès-Nancy 54506, France*

[2)] *Queensland Micro- and Nanotechnology Centre, Griffith University, Nathan, Queensland 4111, Australia*

[a)]Electronic mail: sebastien.lebegue@univ-lorraine.fr

I. APPENDIX

A. Definition of conformation energy

In the set of conformation energy of UpU23, PCONF21 and macrocyclic inhibitors in MPCONF196, we present three ways to represent conformation energy, that is, relative energy with respect to (i) the *average conformation energy* of all conformations at each level of theory, (ii) the energy of *the most stable conformer* determined by CCSD(T), and (iii) the *thermal average energy* of all conformations at each level of theory.

First, the conformation energy is defined relative to the *average energy of all conformations*, $\overline{E} = \frac{1}{N}\sum_{i<N} E_i$, for the tested method:

$$\Delta E_i = E_i - \overline{E} \; . \tag{1}$$

This measures the ability of the method to predict *all* conformers, including higher energy ones

Another way to define the conformation energy is relative to the energy of *the most stable conformer*, as determined by CCSD(T) calculations:

$$\Delta E_i = E_i - E_{0'} \; , \tag{2}$$

where the most stable conformation, $0'$, is determined from the CCSD(T) calculation (i.e. we do not change the reference state if an approximation mis-predicts the lowest energy conformer). According to the definition, we have $N-1$ relative energies for N conformers.

Since biomolecules are in the "heat bath" of a living thing, biological system will involve a statistical ensemble of low-lying conformers, not just the lowest energy one. Therefore, we introduce the *thermal average of conformation energy*, which can be defined as a weighted sum over all conformations of a molecule, $\langle E \rangle = Q^{-1}\sum_{j=1}^{N} E_j e^{-E_j/k_BT}$, where $Q = \sum_{j=1}^{N} e^{-E_j/k_BT}$ is the partition function, k_B is the Boltzmann constant and T is the temperature set to be 300 K (giving $k_BT = 0.596$ kcal.mol). The conformation energy relative to the thermal average is then defined as:

$$\Delta E_i = E_i - \langle E \rangle \; . \tag{3}$$

Since the thermal average is heavily weighted by the most stable conformation based on each method, the accuracy strongly depends on the accurate prediction of the most stable

conformation by each method. But is also influenced by its ability to predict the energy of low-lying conformers with similar energies.

II. SUPPLEMENTARY FIGURES

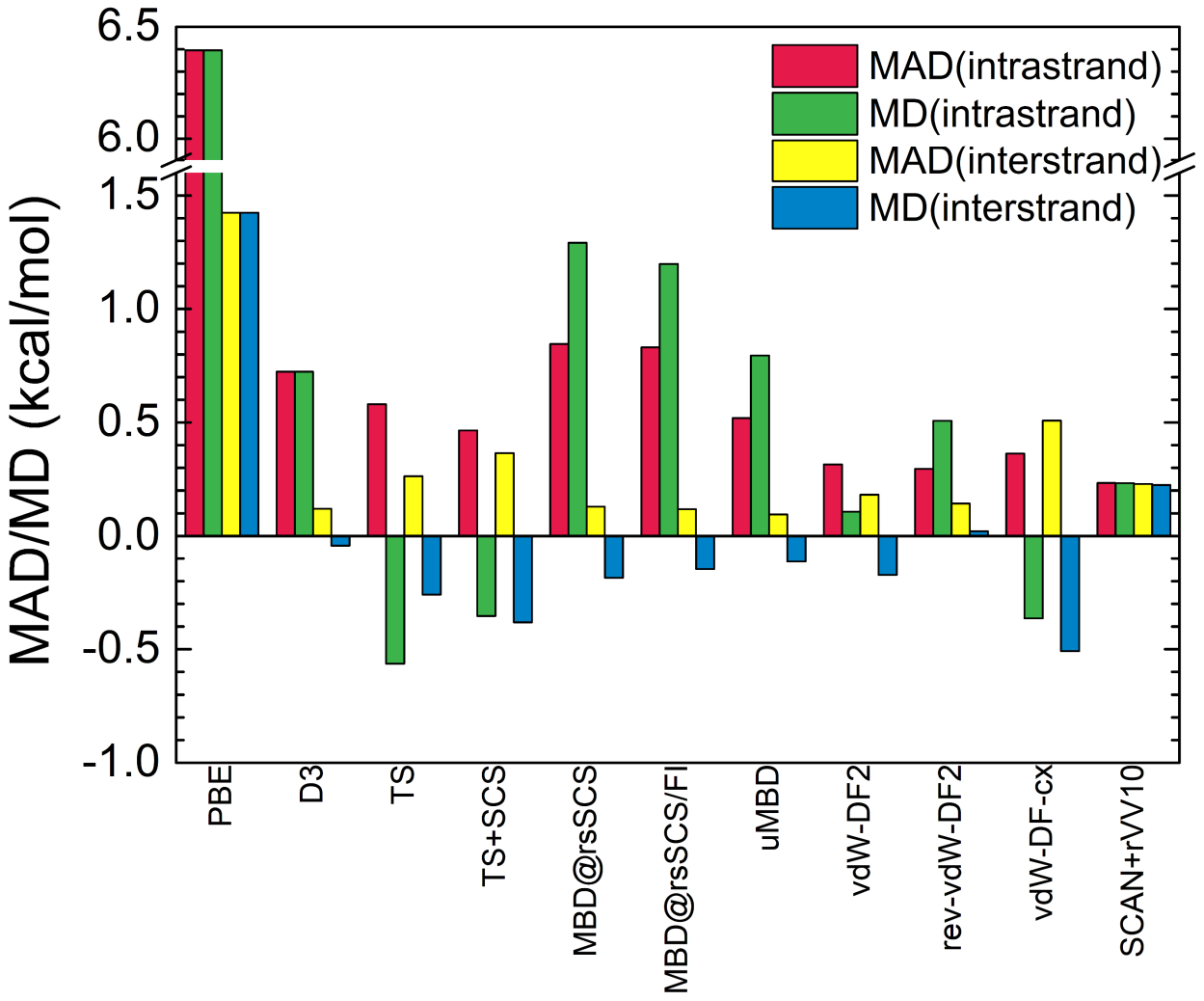

FIG. S1. Mean absolute deviations (MADs) and mean deviations (MDs) of PBE and the selected vdW-corrected DFT methods and nonlocal vdW density functionals for predicting the binding energies of the inter-base pairs, $\Delta E_{stack,ij}$

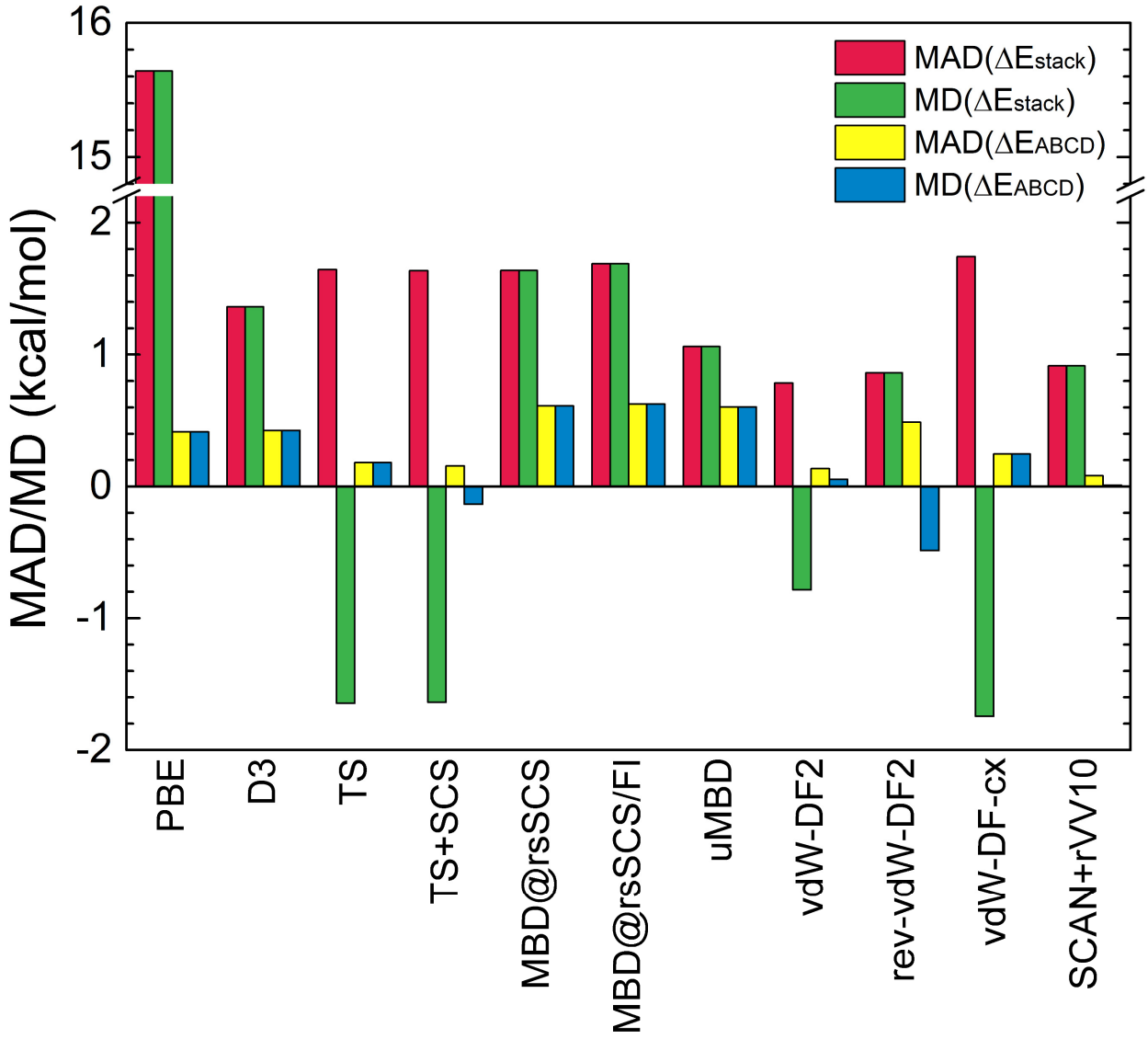

FIG. S2. Mean absolute deviations (MADs) and mean deviations (MDs) of PBE and the selected vdW-corrected DFT methods and nonlocal vdW density functionals for predicting the pairwise contribution, ΔE_{stack}, and the 4-body nonadditive contribution, ΔE_{ABCD}, of the stacking energies of the B-DNA base-pair steps, $\Delta E_{4stack} = \Delta E_{stack} + \Delta E_{ABCD}$.

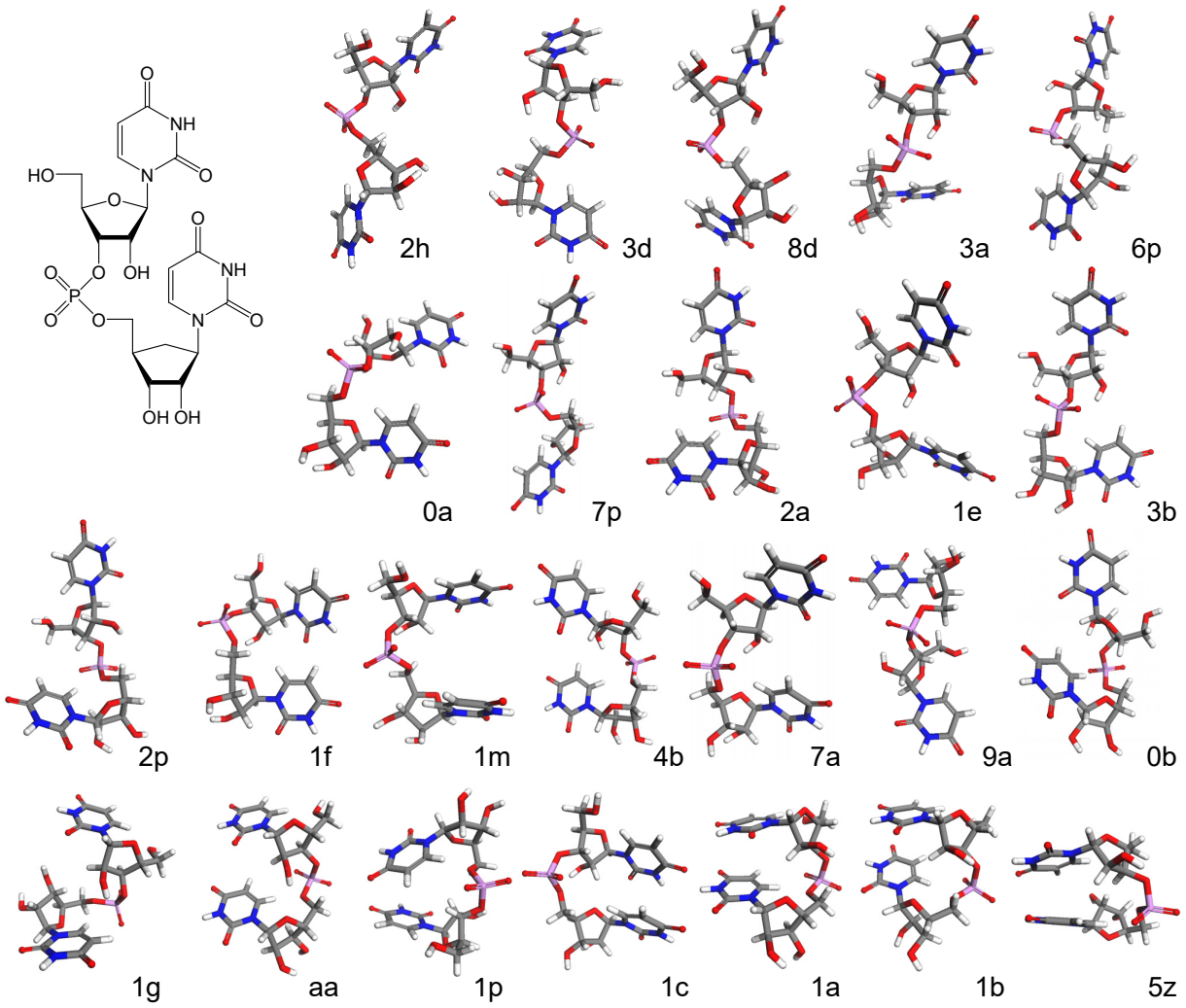

FIG. S3. Structures of 24 uracil dinucleotide conformers in the UpU23 set in ascending order of pairwise TS dispersion correction energy. Atoms are colored by white for H, grey for C, red for O, blue for N, and pink for P.

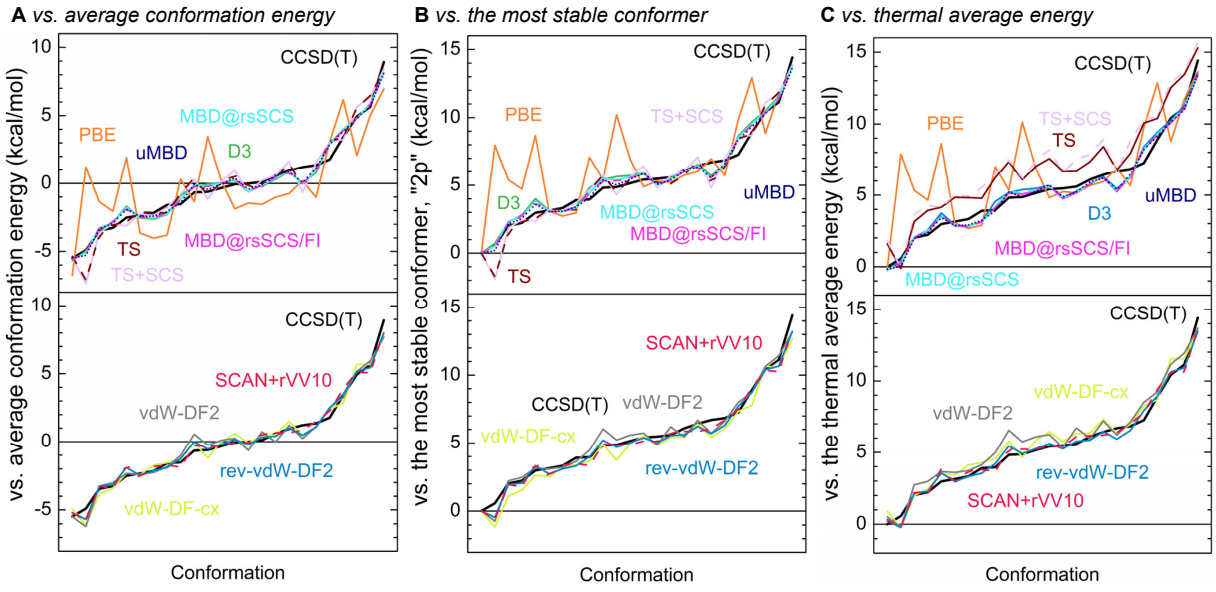

FIG. S4. Individual conformation energy of the uracil dinucleotide conformers in the UpU23 set in ascending order of CCSD(T) conformation energy. Conformation energy is defined as a relative energy to **A** the *average conformation energy*, **B** the energy of *the most stable conformer*, and **C** the *thermal average energy*.

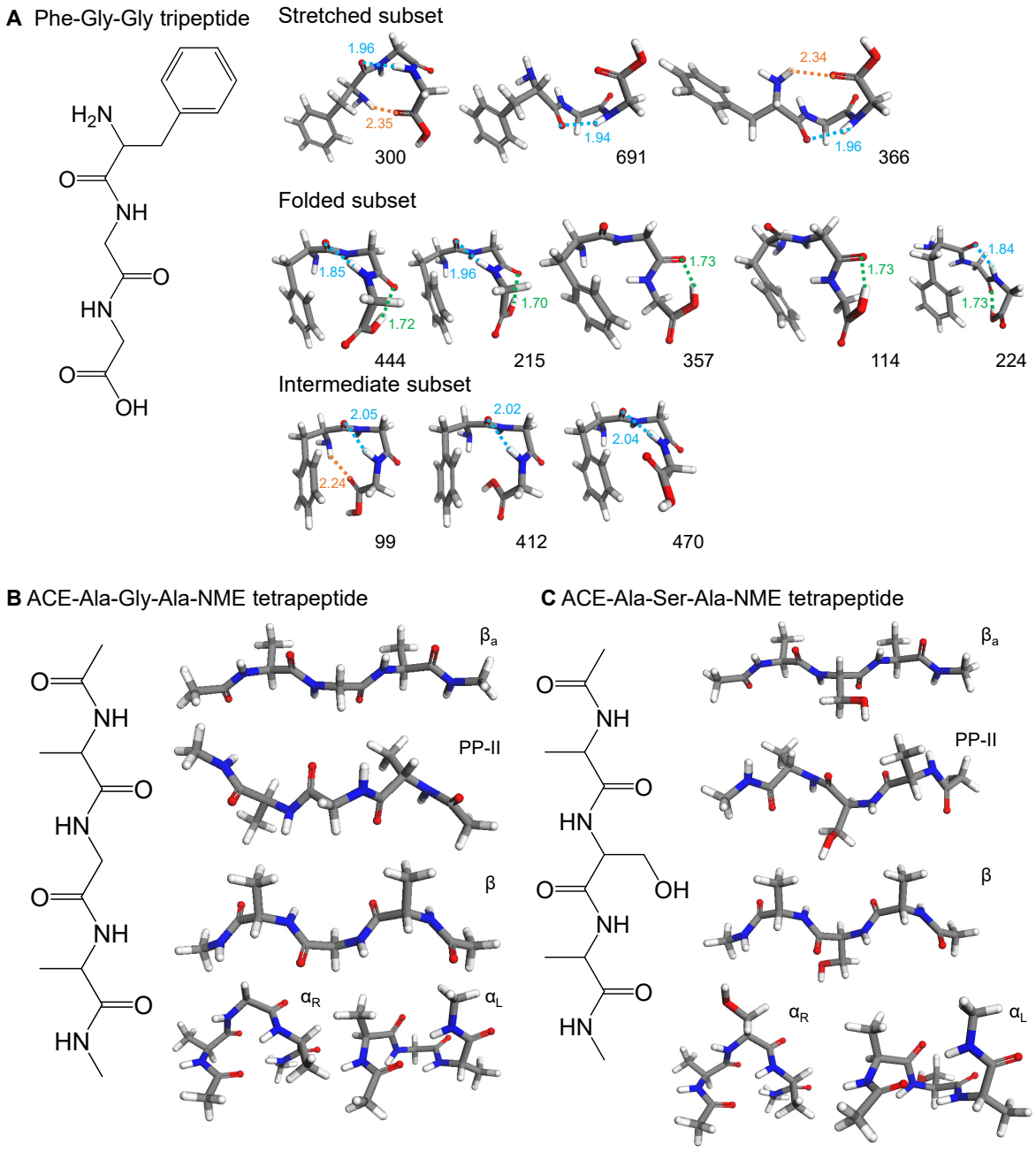

FIG. S5. Structures of **A** 11 conformers of Phe-Gly-Gly tripeptide, **B** 5 conformers of ACE-ALA-GLY-ALA-NME tetrapeptide, and **C** 5 conformers of ACE-ALA-SER-ALA-NME tetrapeptide. β_a is an antiparallel β sheet, PP-II is a polyproline-II helix, β is a parallel β sheet, α_R is an α helix in right-handed, and α_L is an α helix in left-handed, respectively. The conformers of Phe-Gly-Gly tripeptide are listed with the numbers indicating the distances of hydrogen bond, and the atoms are colored by white for H, grey for C, red for O, and blue for N, respectively.

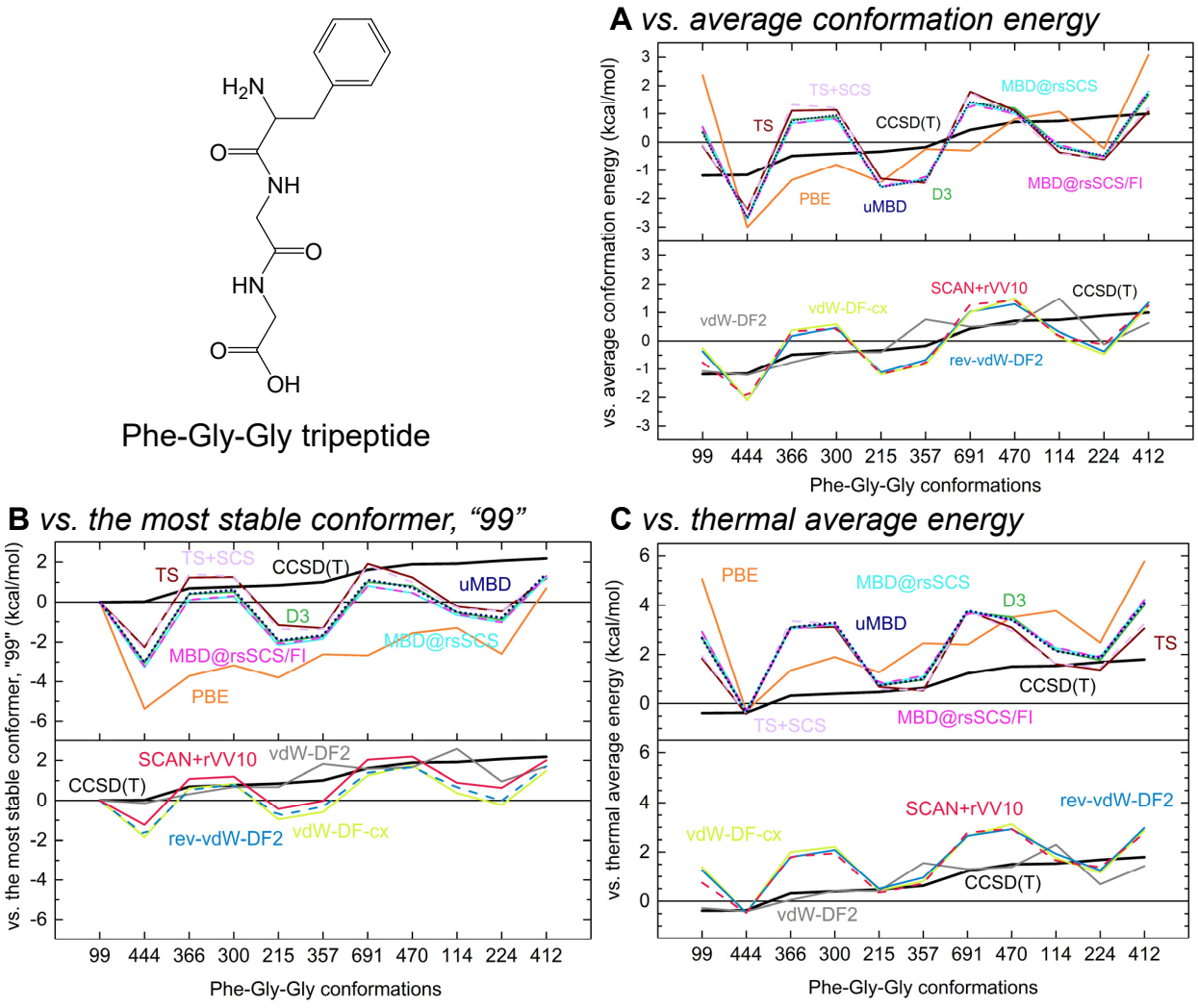

FIG. S6. Individual conformation energy of the Phe-Gly-Gly tripeptide conformers in the PCONF21 set in ascending order of CCSD(T) conformation energy. The conformation energy is defined as a relative energy to **A** the *average conformation energy*, **B** the energy of *the most stable conformer*, and **C** the *thermal average energy*.

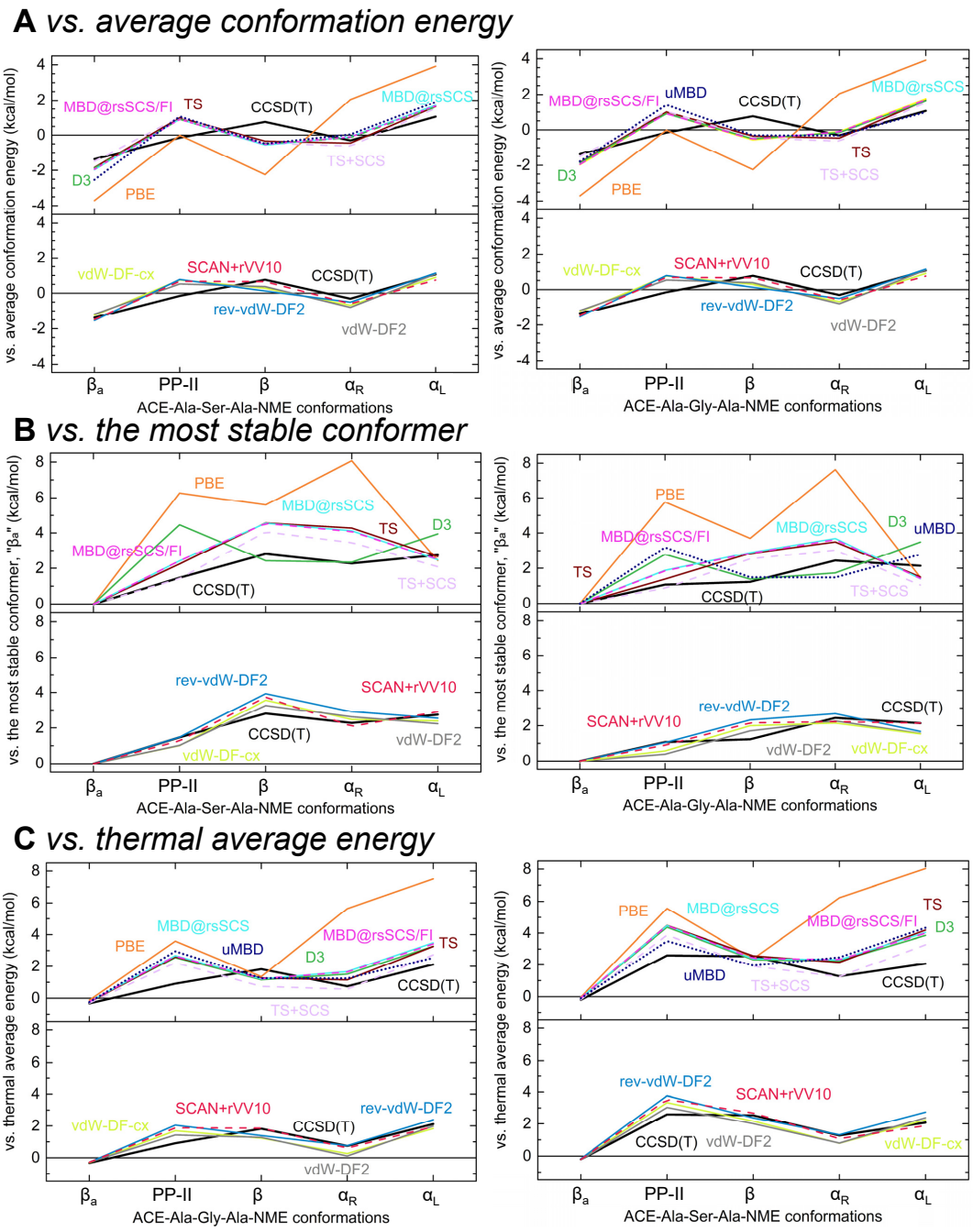

FIG. S7. Individual conformation energy of the (left) ACE-Ala-Gly-Ala-NME and (right) ACE-Ala-Ser-Ala-NME tetrapeptide conformers in the PCONF21 set. The conformation energy is defined as a relative energy to **A** the *average conformation energy*, **B** the energy of *the most stable conformer*, and **C** the *thermal average energy*.

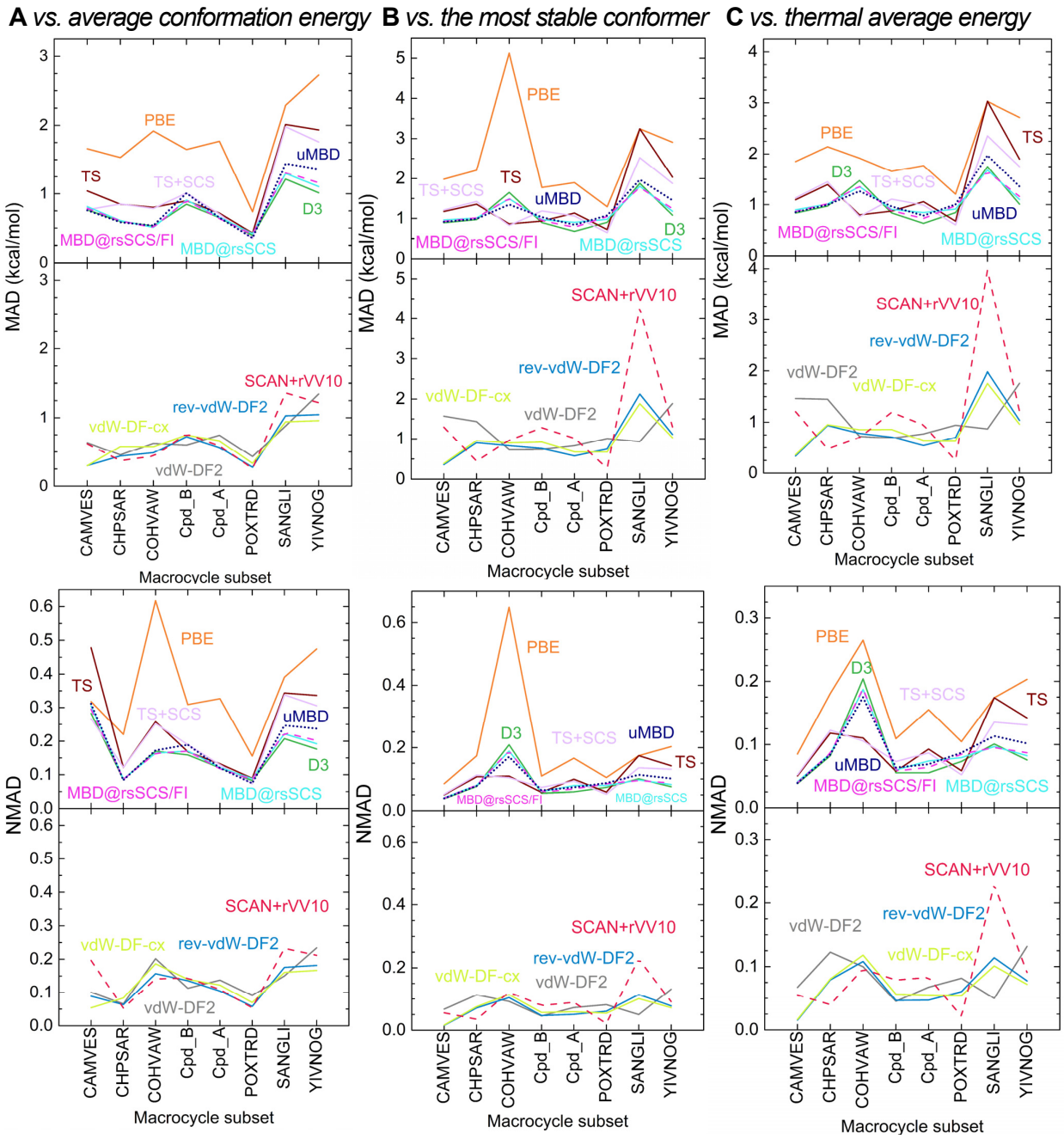

FIG. S8. (Top) Mean absolute deviations (MADs, in kcal/mol) and (bottom) normalized mean absolute deviation (NMAD) of the tested methods in each macrocyclic inhibitor of MPCONF196 set. The conformation energy is defined as a relative energy to **A** the *average conformation energy*, **B** the energy of *the most stable conformer*, and **C** the *thermal average energy*.

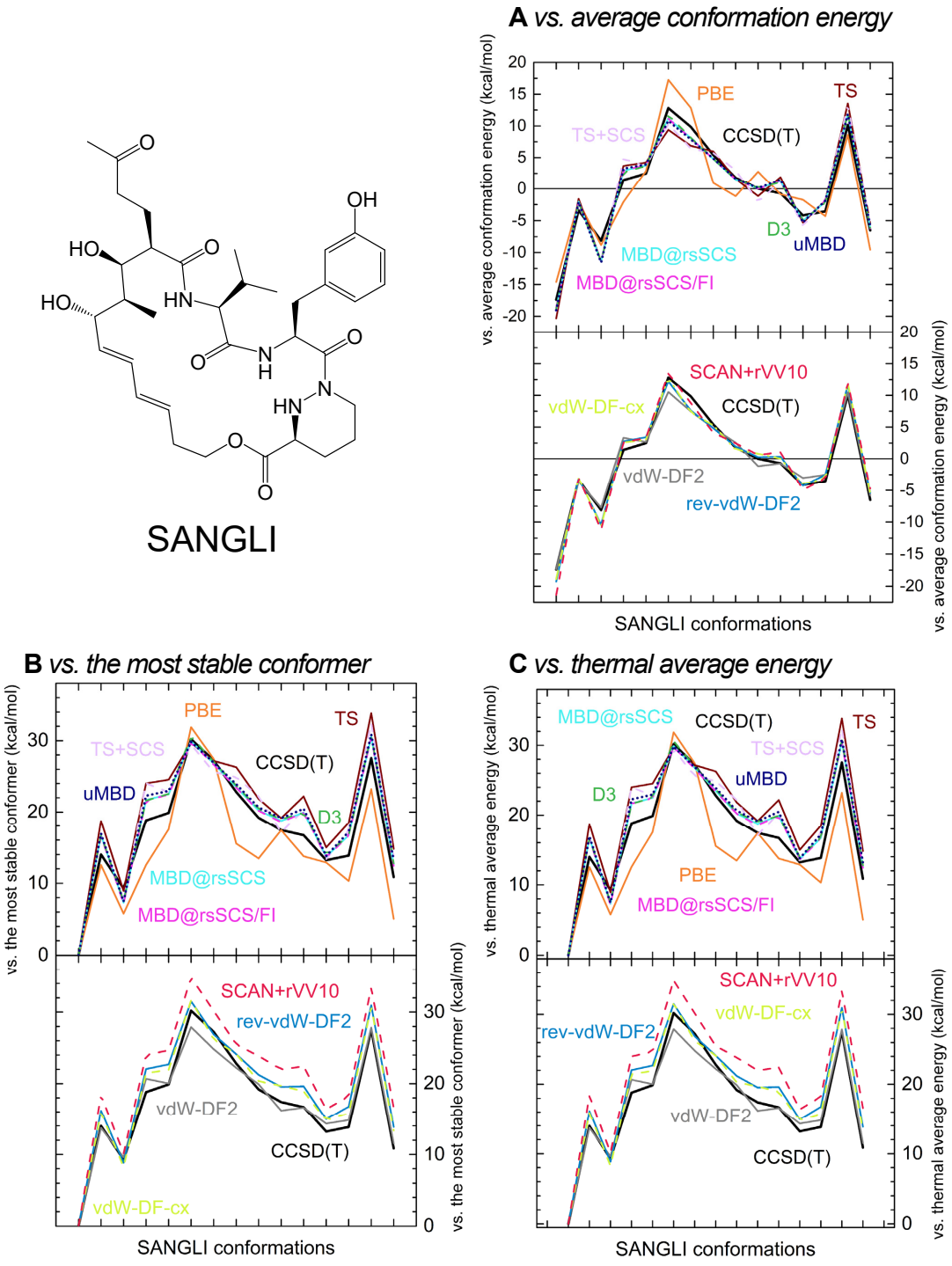

FIG. S9. Individual conformation energy of the SANGLI macrocyclic inhibitor in the MPCONF196 set. The conformation energy is defined as a relative energy to **A** the *average conformation energy*, **B** the energy of *the most stable conformer*, and **C** the *thermal average energy*.

TABLE S1. Mean absolute deviation (MAD, kcal/mol) and mean deviation (MD, kcal/mol) of the tested methods in DNA inter-base pair set. Hybrid B3LYP and meta-GGA functionals (with their dispersion correction) are also listed for comparison from Ref S1.

DNA inter-base pairs	Pair-stacking energy, $\Delta E_{stack,ij}$	
	MAD (kcal/mol)	MD (kcal/mol)
PBE and vdW-correction methods (this work)		
PBE	3.91	+3.91
D3(BJ)	0.42	+0.34
TS	0.42	-0.41
TS+SCS	0.42	-0.41
MBD@rsSCS	0.49	+0.41
MBD@rsSCS/FI	0.47	+0.42
uMBD	0.31	+0.27
Nonlocal vdW functionals (this work)		
vdW-DF2	0.25	-0.20
rev-vdW-DF2	0.22	+0.22
vdW-DF-cx	0.44	-0.44
SCAN+rVV10	0.23	+0.23
Hybrid and meta-GGA functionals[S1]		
B3LYP-D3(BJ)	0.16	0.00
M06-2X	1.02	+1.02
M06-2X-D3(0)	0.34	+0.34
M11	1.81	+1.81
M11-D3(BJ)	0.68	+0.68
ωB97M-V	0.18	-0.08
ωB97M-D3(BJ)	0.35	+0.35

TABLE S2. Mean absolute deviation (MAD, kcal/mol) and mean deviation (MD, kcal/mol) of the tested methods in UpU23 set. Hybrid B3LYP and meta-GGA functionals (with their dispersion correction) are also listed for comparison from Ref S2.

UpU23	vs. average conformation energy		vs. the most stable conformer		vs. thermal average energy	
MAD (kcal/mol)	MAD	NMAD	MAD	NMAD	MAD	NMAD
PBE and vdW-correction methods (this work)						
PBE	2.03	0.84	1.92	0.34	1.84	0.34
D3(BJ)	0.46	0.19	0.50	0.09	0.47	0.09
TS	0.57	0.24	0.60	0.11	1.53	0.28
TS+SCS	0.61	0.25	0.63	0.11	1.85	0.34
MBD@rsSCS	0.44	0.18	0.45	0.84	2.03	0.84
MBD@rsSCS/FI	0.42	0.17	0.43	0.08	0.43	0.08
uMBD	0.35	0.15	0.37	0.06	0.39	0.07
Nonlocal vdW functionals (this work)						
vdW-DF2	0.47	0.19	0.49	0.09	0.70	0.13
rev-vdW-DF2	0.32	0.13	0.38	0.07	0.31	0.06
vdW-DF-cx	0.44	0.18	0.61	0.11	0.57	0.10
SCAN+rVV10	0.34	0.14	0.40	0.07	0.34	0.06
Hybrid and meta-GGA functionals[S2]						
B3LYP	2.48	1.03	2.49	0.44	2.37	0.43
B3LYP-D3(BJ)	0.45	0.19	0.61	0.11	0.78	0.14
M06	0.52	0.22	0.53	0.09	0.74	0.13
M06-D3	0.83	0.34	0.83	0.14	4.05	0.74
M06-2X	0.56	0.23	0.61	0.11	0.55	0.10
M06-2X-D3(0)	0.47	0.20	0.50	0.09	1.04	0.19
ωB97X-D3	0.49	0.20	0.77	0.14	1.00	0.18

TABLE S3. Mean absolute deviation (MAD, kcal/mol) and mean deviation (MD, kcal/mol) of the tested methods in PCONF21 set. Hybrid B3LYP and meta-GGA functionals (with their dispersion correction) are also listed for comparison from Ref S2.

UpU23	vs. average conformation energy		vs. the most stable conformer		vs. thermal average energy	
MAD (kcal/mol)	MAD	NMAD	MAD	NMAD	MAD	NMAD
PBE and vdW-correction methods (this work)						
PBE	1.62	2.17	3.66	2.63	2.32	1.91
D3(BJ)	0.94	1.27	1.43	1.03	1.23	1.01
TS	0.93	1.25	1.25	0.90	1.08	0.89
TS+SCS	0.92	1.24	1.18	0.85	0.98	0.80
MBD@rsSCS	0.96	1.29	1.57	1.13	1.30	1.07
MBD@rsSCS/FI	0.94	1.27	1.54	1.11	1.29	1.06
uMBD	0.89	1.20	1.32	0.95	1.17	0.96
Nonlocal vdW functionals (this work)						
vdW-DF2	0.35	0.47	0.43	0.31	0.35	0.29
rev-vdW-DF2	0.57	0.77	0.72	0.52	0.67	0.55
vdW-DF-cx	0.60	0.81	0.80	0.57	0.69	0.56
SCAN+rVV10	0.53	0.71	0.58	0.42	0.56	0.46
Hybrid and meta-GGA functionals[S2]						
B3LYP	2.09	2.08	3.81	2.74	2.49	2.05
B3LYP-D3(BJ)	0.51	0.68	0.53	0.38	0.46	0.38
M06	0.33	0.45	0.41	0.29	0.29	0.24
M06-D3	0.73	0.98	1.19	0.86	1.06	0.87
M06-2X	0.54	0.72	0.88	0.64	0.76	0.63
M06-2X-D3(0)	0.69	0.93	1.09	0.78	0.95	0.78
ωB97X-D3	0.34	0.45	0.31	0.22	0.28	0.23

TABLE S4. Mean absolute deviation (MAD, kcal/mol) and mean deviation (MD, kcal/mol) of the tested methods in inhibitors in MPCONF196 set. Hybrid B3LYP and meta-GGA functionals (with their dispersion correction) are also listed for comparison from Ref S3.

MPCONF196	vs. average conformation energy		vs. the most stable conformer		vs. thermal average energy	
MAD (kcal/mol)	MAD	NMAD	MAD	NMAD	MAD	NMAD
PBE and vdW-correction methods (this work)						
PBE	2.55	0.21	1.78	0.35	2.05	0.16
D3(BJ)	1.12	0.09	0.75	0.16	1.07	0.09
TS	1.43	0.10	1.08	0.24	1.36	0.10
TS+SCS	1.34	0.09	1.02	0.21	1.28	0.10
MBD@rsSCS	1.16	0.09	0.79	0.17	1.12	0.09
MBD@rsSCS/FI	1.15	0.09	0.79	0.17	1.11	0.09
uMBD	1.21	0.09	0.84	0.18	1.15	0.09
Nonlocal vdW functionals (this work)						
vdW-DF2	1.14	0.08	0.72	0.14	1.08	0.08
rev-vdW-DF2	0.93	0.07	0.61	0.12	0.90	0.07
vdW-DF-cx	0.93	0.07	0.64	0.12	0.90	0.07
SCAN+rVV10	1.34	0.09	0.70	0.14	1.26	0.09
Hybrid and meta-GGA functionals[S3]						
B3LYP	2.58	0.19	2.26	0.45	2.49	0.19
B3LYP-D3(BJ)	0.83	0.06	0.57	0.13	0.78	0.06
M06-2X	1.10	0.08	0.72	0.14	1.03	0.07
M06-2X-D3(0)	1.15	0.08	0.81	0.15	1.08	0.08
M06L	1.69	0.12	1.00	0.20	1.62	0.12
M06L-D3(0)	1.73	0.12	1.12	0.22	1.65	0.12
ωB97X-D3	0.92	0.06	0.64	0.14	0.88	0.06

REFERENCES

[S1]H. Kruse, P. Banas and J. Šponer, *Journal of chemical theory and computation*, 2018, **15**, 95–115.

[S2]L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Physical Chemistry Chemical Physics*, 2017, **19**, 32184–32215.

[S3]J. Rezac, D. Bím, O. Gutten and L. Rulisek, *Journal of chemical theory and computation*, 2018, **14**, 1254–1266.