Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary Material for *Bond dissociation energies of X−H bonds in proteins*

Wojtek Treyde,[ab] Kai Riedmiller,[a] and Frauke Gräter*[abc]

[a] *Heidelberg Institute for Theoretical Studies, Heidelberg, Germany; E-mail: frauke.graeter@h-its.org*

[b] *Max Planck School Matter-to-Life (MtL), Heidelberg, Germany;*

[c] *Interdisciplinary Center for Scientific Computing, Heidelberg University, Heidelberg, Germany.*

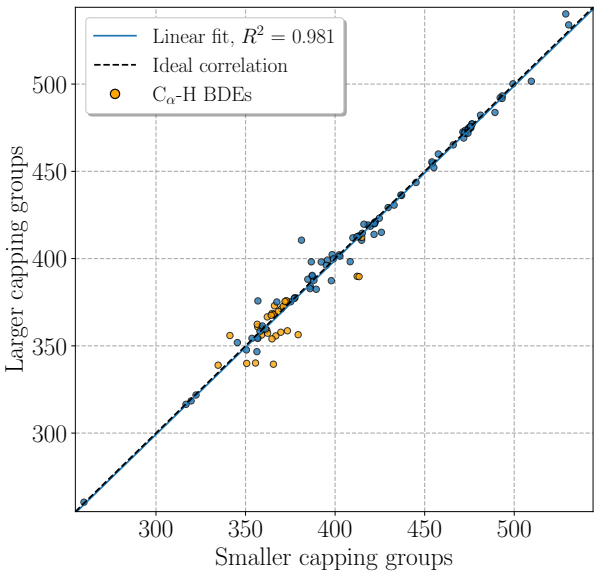

Fig. SI.1 BMK/6-31+G(2df,p) BDEs for dipeptides with larger capping groups (next two atoms along the protein backbone) versus BDEs for dipeptides with smaller capping groups (next atom along the protein backbone), in kJ/mol. Slope and intercept of the linear fit are 0.999 and -0.5 kJ/mol, respectively.

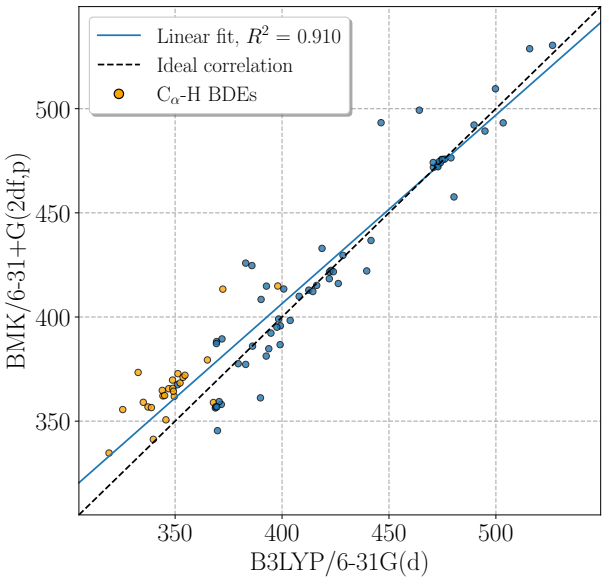

Fig. SI.3 BMK/6-31+G(2df,p) BDEs calculated using smaller capping groups versus B3LYP/6-31G(d) BDEs calculated by Moore *et al.*,[1] in kJ/mol. Slope and intercept of the linear fit are 0.905 and 44.3 kJ/mol, respectively.

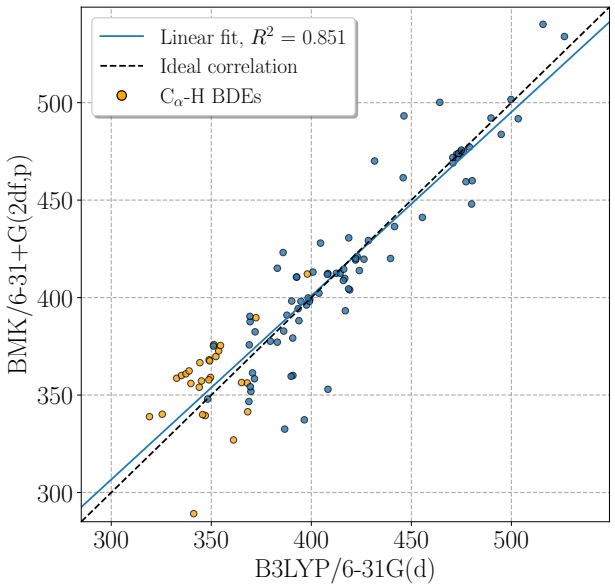

Fig. SI.2 BMK/6-31+G(2df,p) BDEs versus B3LYP/6-31G(d) BDEs calculated by Moore *et al.*,[1] in kJ/mol. Slope and intercept of the linear fit are 0.943 and 23.9 kJ/mol, respectively.

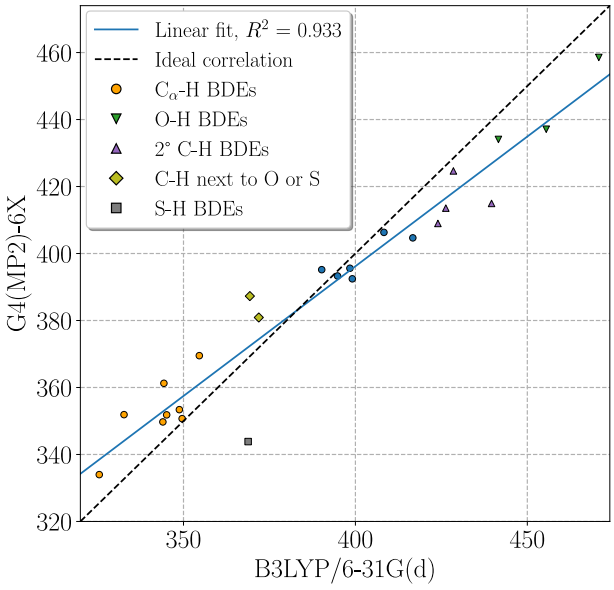

Fig. SI.4 G4(MP2)-6X BDEs versus B3LYP/6-31G(d) BDEs calculated by Moore *et al.*,[1] in kJ/mol. Slope and intercept of the linear fit are 0.775 and 86.4 kJ/mol, respectively.

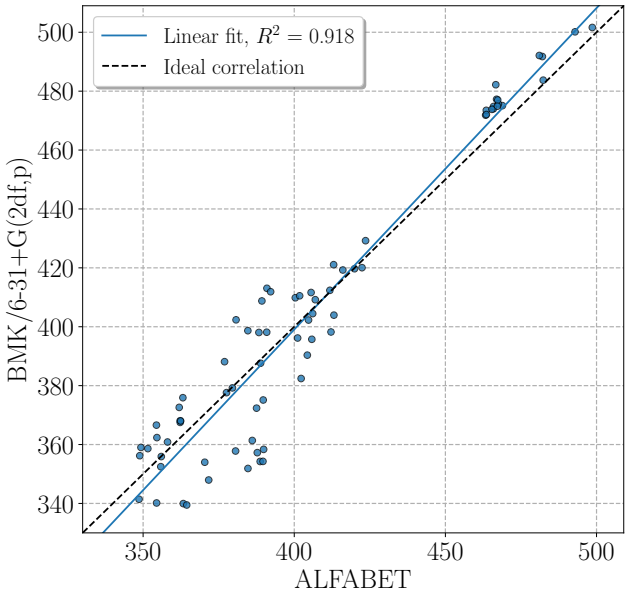

Fig. SI.5 BMK/6-31+G(2df,p) BDEs versus BDEs predicted by ALFABET,[2] in kJ/mol. Slope and intercept of the linear fit are 1.091 and -37.5 kJ/mol, respectively.

Table SI.1 Theoretical and experimental BDEs of reference bonds used in the isodesmic reaction method in kJ/mol. Experimental values are taken from Luo[3]

Bond	BMK/6-31+G(2df,p)	G4(MP2)-6X	Experimental
1	422.7	423.0	420.5 ± 1.3
2	410.1	412.2	410.5 ± 2.9
3	399.3	403.2	400.4 ± 2.9
4	469.4	/[a]	472.2 ± 2.2
5	374.1	383.4	357.3 ± 6.3
6	356.5	372.2	362.8 ± 2.9
7	423.4	437.6	440.2 ± 3.0
8	391.3	396.0	401.2 ± 4.2
9	427.2	442.1	442.3 ± 2.8
10	386.2	392.0	396.5
11	462.6	471.2	468.6 ± 12.6
12	401.3	397.3	392.9 ± 8.4
13	359.5	366.5	365.7 ± 2.1
14	477.7	471.3	454.0
15	383.6	388.6	392.9 ± 8.4
16	410.7	415.1	425.1 ± 8.4
17	378.5	384.2	377.0 ± 8.4
18	384.5	391.5	395.8 ± 8.4
19	399.6	396.7	405.8 ± 8.4
20	442.1	442.0	445.6
21	316.3	328.5	/

[a] Convergence failure in CCSD(T) calculation.

Table SI.2 Comparison of $C_\alpha-H$ BDEs of glycine and alanine dipeptides with literature results from ref.[4], if not stated otherwise. The authors use Boltzmann-averaged enthalpies and methane as a reference system

Dipeptide	BMK/6-31+G(2df,p)[a]	G4(MP2)-6X[a]	G3(MP2)-RAD[4]	Other
Gly (**0**)	319.3	330.6	343.5	340.3[b], 340.2[d], 343.3[f], 341.5[g]„337.9[h], 336.9[i]
Ala (**0**)	316.3	328.5	336.8	332.9[b], 330.3[i]
Gly (**1**)	361.9	354.5	363.4	361.7[b], 360.9[d], 360.0[e]
Ala (**1**)	362.1	357.1	369.7	368.1[b], 364.1[e]
Gly (**2**)	359.0	350.3	365.2	363.8[b], 349.3[j], 351.0[k]
Ala (**2**)	366.6	360.8	373.8	372.3[c], 354.2[j], 357.0[k]

[a]This work. [b]G3B3. [c]IMOMO(G3B3, G3(MP2)-RAD). [d]G4-5H. [e]G3B3, *trans* conformation. [f]G3X(MP2)-RAD. [g]G2(MP2). [h]W1RO. [i]G3MP2, average of results from isodesmic reactions using a large number of reference compounds, lowest enthalpy conformations, taken from ref.[5] [j]DSD-PBE-P86/aug'-cc-pVTZ+d, [k]G4(MP2)-6X, computed directly using the lowest energy conformations from ref.[6], taken from ref.[7]

Table SI.3 Comparison of BDEs at the BMK/6-31+G(2df,p) level of theory with literature results

AA	Pos.	BMK	G4(MP2)-6X	Other
Asn	α	340.2	334	374.9[a], 375.2[b]
Asp	α	358.6	351.9	367.5[a], 366.9[b]
Cys	α	354.0	349.7	355.3[a], 355.9[b], 374.5[c], 373.0[d], 374.6[c,x], 375.2[e,x]
Cys	γ	346.7	343.8	367.5[c], 367.4[d], 369.5[e] 366.8[f], 367.9[g], 367.3[h]
Gln	α	341.4	/	354.7[a], 354.2 [b]
Glu	α	356.2	/	382.2[a], 382.1[b]
His (N_1-H)	α	339.5	/	384.4[a], 383.5[b]
Ile	α	375.9	/	378.8[a], 378.7[b]
Leu	α	355.9	/	376.4[a], 376.6[b]
Lys	α	360.9	/	380.1[a], 379.5[b]
Met	α	362.4	/	382.1[a], 382.0[b]
Phe	α	368.1	/	360.3[a], 370.2[c], 370.7[d], 371.0[c,x], 372.9[e,x]
Pro	α	389.7	/	381.9[a], 382.6[b], 393.0[c], 391.5[d], 396.7[e,x]
Ser	α	357.8	353.4	392.3[a], 393.8[b]
Thr	α	357.2	351.8	366.7[a], 368.7[b]
Trp	α	372.6	/	364.3[a]
Tyr	α	367.5	/	360.2[a], 369.9[c], 368.7[d] 371.2[c,x], 372.7[e,x]
Tyr	3	359.6	/	346.1[c,x], 365.6[d], 349.5[e,x]
Val	α	375.4	369.5	381.2[a], 381.9[b]

[a]DSD-PBE-P86/aug'-cc-pVTZ+d, [b]G4(MP2)-6X, computed directly using the lowest energy conformations from ref.[6], taken from ref.[7] [c]G3(MP2)-RAD, [d]IMOMO(G3B3,G3(MP2)-RAD), [e]ROMP2/6-311+G(3df,2p)//UB3LYP/6-31G(d), [f]G3B3, taken from ref.[8] [g]G3(MP2)//B3LYP, [h]G3, computed directly, taken from ref.[9] Zipse and colleagues[8,10] use SH_2, H_2O, and CH_4 as reference compounds and Boltzmann averages. [x]Best conformer only.

Experimental details

Gas-phase BDEs and BDFEs at 298.15 K were computed using Gaussian09.[11] Frequency analyses were carried out to verify the structures were local minima. Spin contamination was found to be minimal for all systems (at maximum $\langle \hat{S}^2 \rangle = 0.7509$ after correction). All side chain X−H bonds in all canonical amino acids, hydroxyproline, hydroxylysine, DOPA, as well as X−H bonds in the C and N termini, the backbone, and the acetylated and N-methyl amidated N and C termini were considered. To model the local environment in a protein, C termini were extended by an N-methyl amino group and N termini by an acetyl group. The model peptides were built using Avogadro[12] and GaussView.[11]

To minimize intramolecular hydrogen bonding and obtain more reliable BDEs, for amino acids with polar side chains, side chain dihedral angles were set to $\chi_1 \approx 60^\circ$ and $\chi_i \approx 180^\circ$, $i > 1$, before geometry optimization, and the maximum size for the initial optimization step was set to 0.1 a_0. If the geometry optimization still led to hydrogen bonding, the maximum size for the optimization step was set to 0.05 a_0 and then to 0.01 a_0, and updating the step size during optimization was suppressed.

Because of the risk of hydrogen bonding, conformer sampling was not performed. For dipeptide radicals, however, starting structure resulting from the deletion of each chemically equivalent H atom were considered. If the geometry optimization converged to a saddle point, the structure was displaced along the vibration with the imaginary frequency in both directions and the geometry optimization was restarted. Most systems were fully optimized, however, to exclude big conformational changes, optimizations were stopped in some cases after only the forces converged. If more than one chemically equivalent structure converged to a local minimum, the lower energy conformer or the conformer that did not exhibit hydrogen bonding between the side chain and the backbone was selected.

SMILES strings for ALFABET[2] predictions were obtained from optimized reactant PDB files using RDKit.[13] ALFABET predictions for which the reactant was outside the model's domain of validity were discarded, resulting in 74 BDEs that were compared to the corresponding BMK values.

References

1 B. N. Moore and R. R. Julian, *Physical Chemistry Chemical Physics*, 2012, **14**, 3148–3154.

2 P. C. St John, Y. Guan, Y. Kim, S. Kim and R. S. Paton, *Nature communications*, 2020, **11**, 1–12.

3 Y.-R. Luo, *Comprehensive Handbook of Chemical Bond Energies*, CRC Press, 2007.

4 J. Hioe, M. Mosch, D. M. Smith and H. Zipse, *RSC Adv.*, 2013, **3**, 12403–12408.

5 K. M. Uddin, D. J. Henry, R. A. Poirier and P. L. Warburton, *Theoretical Chemistry Accounts*, 2016, **135**, 224.

6 Y. Yuan, M. J. L. Mills, P. L. A. Popelier and F. Jensen, *The Journal of Physical Chemistry A*, 2014, **118**, 7876–7891.

7 B. Chan, A. Karton, C. J. Easton and L. Radom, *Journal of Chemical Theory and Computation*, 2016, **12**, 1606–1613.

8 J. Hioe and H. Zipse, *Faraday Discussions*, 2010, **145**, 301–313.

9 M. V. Roux, C. Foces-Foces, R. Notario, Ribeiro da Silva, Manuel A. V., Ribeiro da Silva, Maria das Dores M. C., Santos, Ana Filipa L. O. M. and E. Juaristi, *The Journal of Physical Chemistry B*, 2010, **114**, 10530–10540.

10 J. Hioe, G. Savasci, H. Brand and H. Zipse, *Chemistry – A European Journal*, 2011, **17**, 3781–3789.

11 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, Montgomery, J. A., Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, 2013.

12 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Journal of Cheminformatics*, 2012, **4**, 17.

13 *RDKit: Open-source cheminformatics*.