Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2020

Supporting Information:

Modeling of spin-spin distance distributions for nitroxide labeled biomacromolecules

Sebastian Spicher,[†] Dinar Abdullin,[‡] Stefan Grimme,[*,†] and Olav Schiemann[*,‡]

†*Mulliken Center for Theoretical Chemistry, Institute of Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany*

‡*Institute of Physical and Theoretical Chemistry, University of Bonn, Wegelerstr. 12, 53115 Bonn, Germany*

E-mail: grimme@thch.uni-bonn.de; schiemann@pc.uni-bonn.de

Computational details

Input structures were generated with `mtsslWizard`.[S1] Structures were converted from pdb to xyz format with the `maestro`[S2] software package. All CREST/MD computations were performed with the `crest`[S3] and `xtb`[S4] stand alone programs. The default convergence criteria (10^{-7} E_h for energies and 10^{-5} E_h/Bohr for gradients) were used throughout. MMM calculations were performed with version *2020.2* employing the R1A-298K-UFF-216-r1-CASD rotamer library for the R1 side chain. MtsslWizard calculations were performed with the server version[S5] with clashes settings tight. For the azurin mutants, the distribution of the Cu(II) spin density was taken into account with each respective method. Trajectory evaluation was performed with the program `travis`.[S6] Structure visualization was done in `pymol`.[S7] MD simulations with GFN-FF were carried out for 1 ns at the respective freezing temperature of the solvents, employing the implicit GBSA(H_2O) solvation model. A time step of 2 fs at an increased hydrogen mass of 4 amu and equilibration phase of 200 ps was chosen.

For the mutants of azurin (1952 atoms), 100 ps of the MD simulation took on average 7.3 hours on 4 Intel© Xeon E5-2660 v4 @ 2.00 GHz CPUs. On the same machine, the 100 ps took on average 16.4 hours for the mutants of T4L (2683 atoms), again on 4 CPUs.

Statistical error measures

In this work, the following statistical measures were used. p is an arbitrary property.

$$\delta p = p_{calc.} - p_{ref.} \tag{1}$$

The error measures are defined by:

- Mean absolute deviation (MAD):

$$MAD = \frac{1}{N}\sum_{i}^{N}|\delta p_i| \tag{2}$$

- Standard deviation (SD):

$$SD = \sqrt{\frac{\sum_i^N |\delta p_i - MD|^2}{N-1}} \tag{3}$$

- relative MAD (*rel*MAD):

$$relMAD = \frac{1}{N}\sum_{i}^{N}\frac{|\delta p_i|}{p_{ref}}. \tag{4}$$

For the calculation of the distance distributions fur azurin, the distribution of spin densities is taken into account. ρ_{Cu} = 0.35, ρ_S = 0.60 and for nitroxide, the center of the N-O bond is taken ($\rho_N = \rho_O$ = 0.50). The spin density weighted distances are calculated according to:

$$\frac{1}{r^3} = \frac{1}{\rho_{Cu}+\rho_S}\left(\rho_{Cu}\,\frac{1}{r^3_{NO-Cu}} + \rho_S\,\frac{1}{r^3_{NO-S}}\right), \tag{5}$$

$$r = \left(\frac{0.95}{\frac{0.35}{r^3_{NO-Cu}} + \frac{0.60}{r^3_{NO-S}}}\right)^{\frac{1}{3}}. \tag{6}$$

The most probable distance r_p is the distance with the highest intensity/probability within the distance distribution. The mean distance $\bar{r}$ is calculated as follows,

$$\bar{r} = \frac{\sum_i r\,P(r_i)}{\sum_r P(r_i)} \tag{7}$$

where $I(r)$ is the respective probability associated with each distance.

Computational results

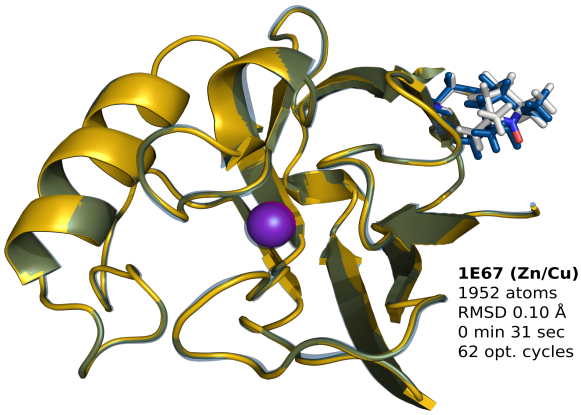

Figure S1: RMSD minimized structure overlay between the optimized GFN-FF geometries for zinc-azurin (yellow) and the copper-azurin analogue (transparent blue). The CSD identifier are given as well as heavy atom RMSD values, total computation wall-times, and the required number of geometry optimization cycles.

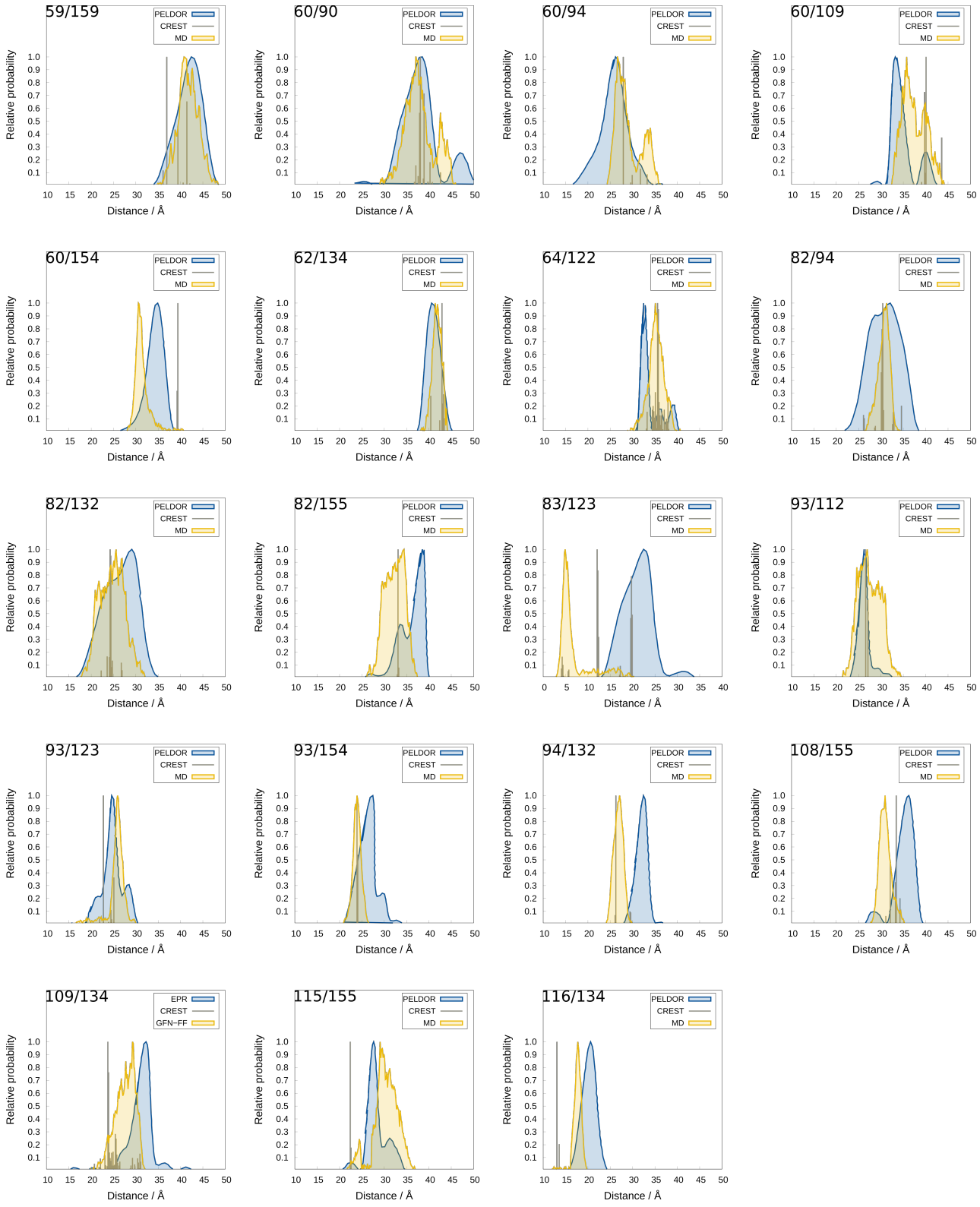

Figure S2: Spin label distance distribution functions for the 19 T4L mutants. The results from the CREST conformations without any MD averaging are shown in gray, the distribution obtained by the GFN-FF MD simulation in yellow and experimental EPR data are shown in blue.

Molecular dynamics settings

MD simulations with GFN-FF were carried out for 1 ns at the freezing point of the respective solvent employing the implicit GBSA(H_2O) solvation model. A time step of 2 fs (at an increased hydrogen mass of 4 amu) and equilibration phase of 200 ps was chosen. In the following, the effects of different MD simulation lengths and temperatures are investigated. Regarding the simulation time, the limiting factor is the computation time. GFN-FF is a physically motivated partially polarizable force-field with many sophisticated energy terms that lead to a scaling which is roughly a factor of 10 slower than specialized protein FFs. For 1 ns an overall wall time of four to seven days (dependent on the system size) is reached on 4 Intel© Xeon E5-2660 v4 @ 2.00 GHz CPUs.

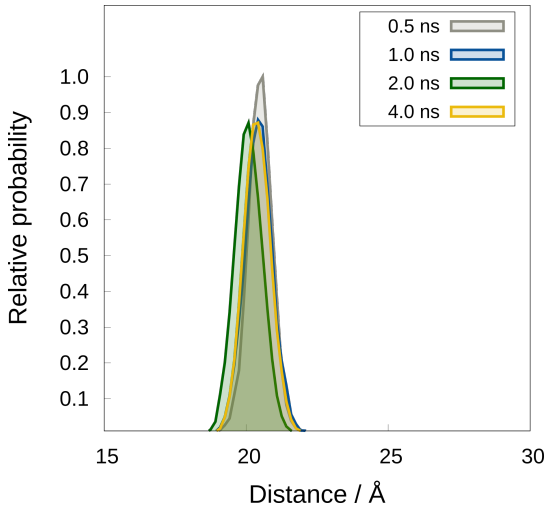

Figure S3: Effect of different MD settings on the radial distribution function of azurin mutant T21R1. For longer MD simulation times the intensities decrease while the distribution width becomes slightly larger. The mean distances remain roughly the same.

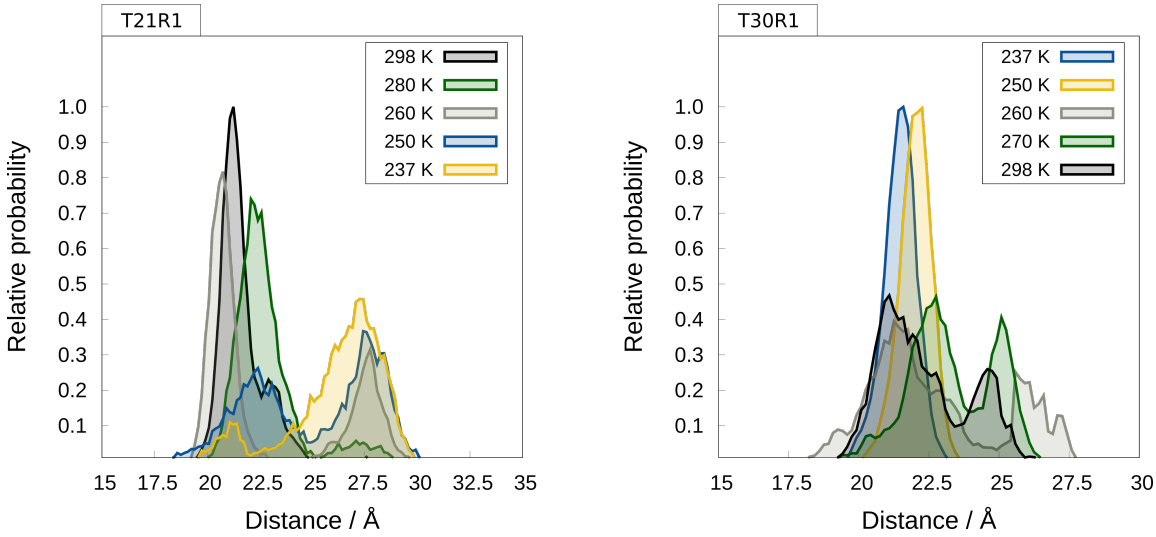

Figure S4: Effect of different MD temperatures on the radial distribution function of azurin mutant T21R1 and T30R1. For T21R1 the starting conformation is taken from the MTSL conformation of larger distance. For T30R1 the shorter conformation was the starting point of the MD. More information is in the manuscript.

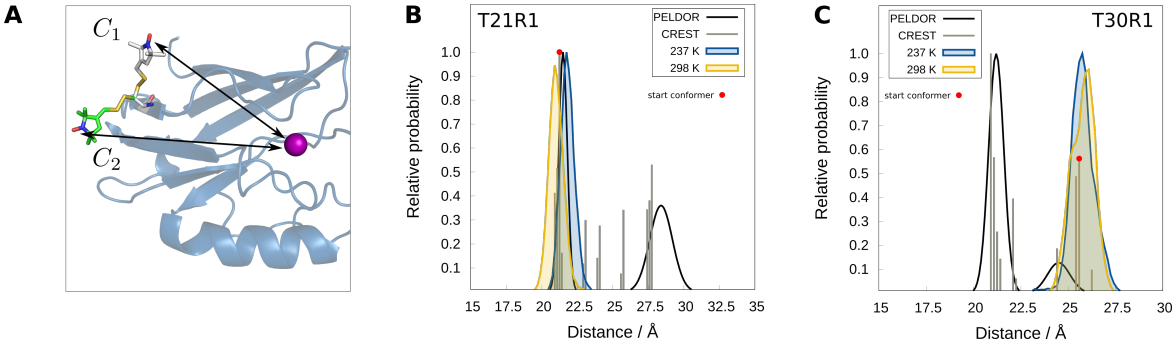

Figure S5: Investigation of temperature effects in the MD simulation. In **A**, two conformations with shorter (C_1) and longer (C_2) inter-spin distances of T21R1 are illustrated. Computations are performed at the freezing point of the solvent (237 K) and at room temperature (298 K), respectively, for the azurin mutants T21R1 (**B**) and T30R1 (**C**). The starting conformation for the MD simulation is indicated by a red dot.

CREST conformations

From the constrained CREST calculations for the R1 side chain, conformer ensembles with sizes from 20 (for azurin mutants) to 200 (for T4L mutants) were obtained. For each conformer the distance between the respective spin center was calculated. The distance distribution was obtained as the sum of all the distances from the conformers as illustrated for S100R1 in Figure S6. The Boltzmann population of each conformer determines the intensity. Relative probabilities were calculated by dividing the intensity of each conformer by the maximum value of the energetically lowest conformer. If multiple conformers existed that showed a very similar distance between the spin center (<1 Å), they were considered as one R1 conformer cluster. For S100R1, 34 conformers exited, which lead to three R1 conformer clusters as shown in Figure S6. The division into conformer clusters was done manually. Since only one R1 conformer cluster had a relative probability of more than 0.5, one GFN-FF MD simulation was carried out for the conformer with the highest Boltzmann population within the respective R1 conformer cluster. In blue the conformer ensemble without Boltzmann weighting is shown. Here, a critical aspect of the applied procedure should be addressed. All conformational energies were obtained at the GFN-FF level of theory for a constrained system. Even though this was tested in Ref. S8 for small to medium sized systems, the extrapolation to large biomolecules can not be assumed.

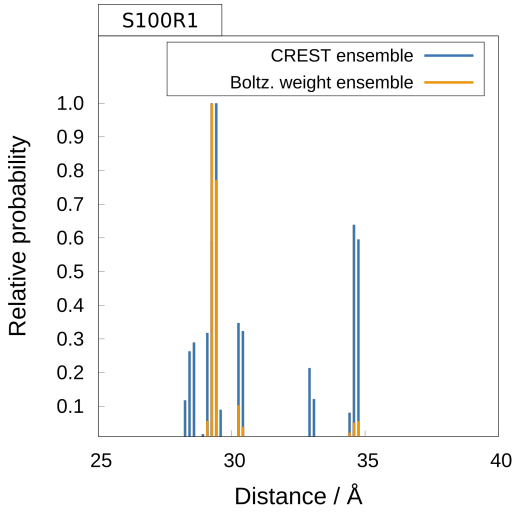

Figure S6: Conformer ensemble for azurin mutant S100R1. In blue, the sum of all distances obtained from the entire CREST ensemble. In yellow, the Boltzmann weighted sum of all distances.

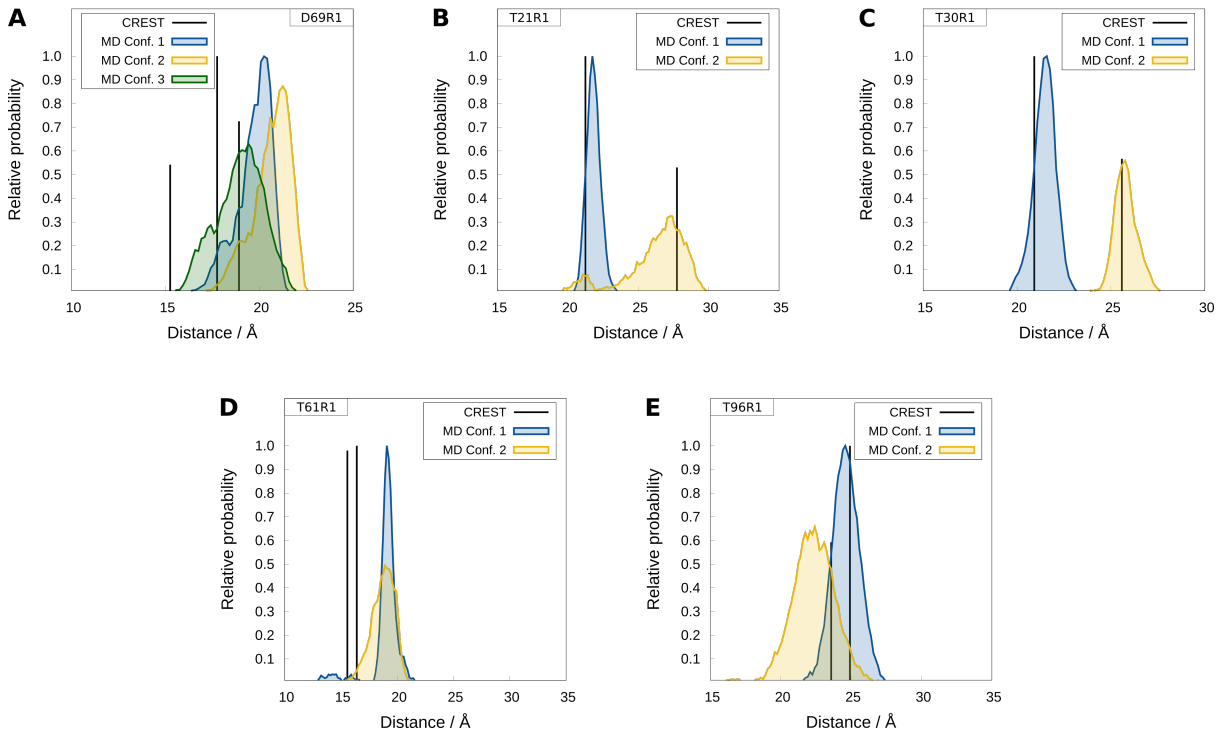

Figure S7: Contributions of the individual conformers to the radial distribution. For D69R1 MD simulations are performed for three conformers. T21R1, T30R1, and T62R1 show two conformations with a relative probability larger than 0.5. The MTSL conformers are named according to their relative probability, e.g Conf. 1 for the highest probability. Only conformers with a relative probability larger than 0.5 are shown.

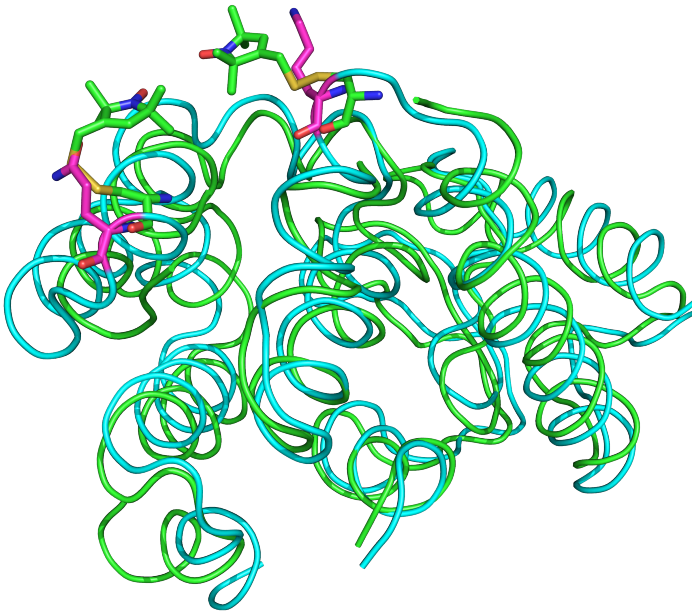

Figure S8: Overlay of an MD snapshots (green) with the initial X-ray structure (blue) for mutant 83/123 revealing non-covalent interactions between the two spin-labels. Even though the tertiary structure is similar on a global view, significant differences in the positions of individual loops and helices are visible. These conformational changes cause the short mean distance of only 5.8 Å.

References

(S1) Hagelueken, G.; Ward, R.; Naismith, J. H.; Schiemann, O. MtsslWizard: in silico spin-labeling and generation of distance distributions in PyMOL. *Appl. Magn. Reson.* **2012**, *42*, 377–391.

(S2) Schrödinger Release 2020-2: Maestro, Schrödinger, LLC, New York, NY, 2020.

(S3) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

(S4) "Semiempirical Extended Tight-Binding Program Package `xtb`", https://github.com/grimme-lab/xtb. Accessed: 2020-04-20.

(S5) "MtsslWizard Server Version", http://www.mtsslsuite.isb.ukbonn.de/. Accessed: 2020-10-08.

(S6) Brehm, M.; Kirchner, B. TRAVIS - A Free Analyzer and Visualizer for Monte Carlo and Molecular Dynamics Trajectories. *J. Chem. Inf. Model.* **2011**, *51*, 2007–2023, PMID: 21761915.

(S7) The PyMOL Molecular Graphics System, Version 2.2.0, SchrÃűdinger, LLC.

(S8) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem. Int. Ed.* **2020**, *59*, 15665–15673.