Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/12/eaax0024/DC1

Supplementary Materials for

Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions

Martin Stöhr and Alexandre Tkatchenko*

*Corresponding author. Email: alexandre.tkatchenko@uni.lu

Published 13 December 2019, *Sci. Adv.* **5**, eaax0024 (2019)
DOI: 10.1126/sciadv.aax0024

This PDF file includes:

Section S1. Computational details and vdW models

1.1 Density-Functional Tight-Binding calculations

Electronic structure calculations have been carried out on the Density-Functional Tight-Binding (DFTB) level of theory using a locally modified MPI-version of DFTB+ (*59,60*). For the study of the Fip35 Hpin1 WW-domain and the Nle/Nle mutant of villin HP35 we employed the second-order (self-consistent charges) DFTB method with recent mio-1-1 parameters (*38*). For the calculations on the Chignolin variant cln025 third-order DFTB (*61*) (DFTB3) with recent 3ob parameters (*62–64*) has been used. Effective atomic polarizabilities within the DFTB3 framework, as further used in our study, have been tested to comply with the results from the second-order approach. In all calculations we employed a self-consistency criterion of 10^{-5} elementary charges and Γ-point sampling.

1.2 van der Waals Dispersion Models

In our study on many-body dispersion effects on protein energetics and the collectivity of van der Waals (vdW) interactions in solvated biosystems, we have employed a combined approach of DFTB and the Many-Body Dispersion (MBD) formalism (*10,42,43*). The results from MBD have been compared with common pairwise approaches to vdW dispersion interactions, as they are commonly employed in state-of-the-art simulation techniques. Pairwise models are represented by the electronic structure-based vdW(TS) (*39*) and Grimme's D2 (*40*) and D3 (*41,65*) method. The central starting point of MBD and vdW(TS) is the definition of effective atomic dipole polarizabilities, $\alpha_{\text{eff}}^{(I)}$, according to the chemical environment. As presented in Reference (*37*), these can be derived from DFTB as,

$$\alpha_{\text{eff}}^{(I)} = \frac{\alpha_{\text{free}}^{(I)}}{Z_I} \cdot \sum_{i \in I} \mathbf{P}_{ii} \tag{SI-1}$$

where $\mathbf{P}$ is the Mulliken population matrix as obtained from DFTB. In vdW(TS), the obtained effective polarizabilities then define effective C_6 interaction coefficients, which then enter pairwise additive $C_6 \cdot R_{IJ}^{-6}$ potentials. The same functional form is used in Grimme's D2 and D3 schemes. Here, one relies on fixed (D2) or geometry-dependent (D3) interaction coefficients.

vdW(TS) calculations have been carried out using a standalone calculator based on '*semp_disp_corr.F90*', originally part of CASTEP (*66*), within the Atomic Simulation Environment (*67*). Calculations involving D2 or D3 have been performed using the DFTD3 module (*68*). For D3 calculations, Becke-Johnson damping as proposed in Reference (*65*) has been used.

In MBD, $\alpha_{\text{eff}}^{(I)}$ is first subject to self-consistent, electrodynamic screening (SCS) to account for the presence of the surrounding fluctuating atomic dipoles and obtain effective, screened atomic polarizabilities $\tilde{\alpha}_{\text{eff}}^{(I)}$. This is achieved by inverting the Dyson-like equation for the dynamic polarizabilities,

$$\tilde{\alpha}_{\text{eff}}^{(I)}(i\omega) = \sum_J \mathbf{B}_{IJ} \quad , \qquad \text{with} \quad \mathbf{B} = \left[\mathbf{A}^{-1}(i\omega) + \mathbf{T}_{\text{gg}}\right]^{-1} \tag{SI-2}$$

where $\mathbf{A}(i\omega)$ is the diagonal matrix of the dynamic polarizabilities $\alpha_{\text{eff}}^{(I)}(i\omega)$ at imaginary frequency $i\omega$ and $\mathbf{T}_{\text{gg}}$ is the effective short-range dipole potential constructed from the Coulomb interaction of two overlapping Gaussian charge densities. In the MBD formalism we model the interaction of intrinsic

electronic fluctuations in form of a set of dipole-coupled Quantum Harmonic Oscillators (QHOs). In accordance with this, the dynamic polarizability and effective excitation frequency, ω_I, are defined as

$$\alpha_{\text{eff}}^{(I)}(i\omega) = \frac{\alpha_{\text{eff}}^{(I)}}{1+\left(\frac{\omega}{\omega_I}\right)^2} \qquad \text{and} \qquad \omega_I = \frac{4}{3}\frac{\alpha_{\text{eff}}^{(I)}}{\alpha_{\text{free}}^{(I)}}\frac{C_{6,\text{free}}^{(II)}}{\alpha_{\text{free}}^{(I)}} \tag{SI-3}$$

respectively. The effective, screened polarizabilities then define the MBD Hamiltonian,

$$\mathcal{H}_{\text{MBD}} = \mathcal{T}_\zeta + \frac{1}{2}\zeta^T \mathcal{V}\,\zeta \quad , \qquad \text{with} \quad \mathcal{V}_{IJ}^{(k,l)} = \omega_I\omega_J\left[\delta_{IJ} + (1-\delta_{IJ})\sqrt{\tilde{\alpha}_I\tilde{\alpha}_J}\,\mathcal{D}_{IJ}^{(k,l)}\right] \tag{SI-4}$$

In eq. [SI-4], $\mathcal{T}$ is the kinetic energy, $\mathcal{V}$ the potential energy matrix, and ζ the direct sum of mass-weighted displacements of the individual QHOs. $\mathcal{V}_{IJ}$ is defined by the characteristic excitation frequencies, screened polarizabilities, and a damped dipole coupling tensor (Fermi-damping) $\mathcal{D}_{IJ}$ (*10,42,43*). All MBD and vdW(TS)@SCS calculations have been performed using a self-written implementation.

1.3 Molecular Dynamics Simulations with updated Force Field

To study the effect of the over-compaction of unfolded protein states as obtained from traditional molecular mechanics approaches and to rule out potential artifacts resulting from this, we have performed an additional sampling of the chignolin variant "cln025" in explicit water. To this end, we have employed the recently developed *a99SB-disp* force field (*34*). This approach has been designed to and shown to avoid spurious over-compaction of unfolded and disordered protein states and therefore provide a more balanced description of *intra*-protein and protein-water vdW interactions. The simulations have been performed using GROMACS (*69,70*) together with the force field definition provided by D. E. Shaw Research. We have performed two molecular dynamics simulations with equivalent set-up: one starting from an unfolded and one from a folded conformation of the previous molecular dynamics simulations (*46*). The systems were set-up according to standard protocol employing the corresponding topology and definitions of a99SB-disp.

Starting from the gas phase conformations of both initial structures, the protein has been resolvated, neutralized by sodium ions, relaxed with restrained positions down to a maximal force component of 100 kJ/mol/nm, and equilibrated for 3 ps in NVT (297 K) and for 8 ps in NPT (300 K, 1 bar) ensemble. Molecular dynamics simulations have subsequently been run for 50 and 100 ns for the folded and unfolded state, respectively, using the same NPT settings. The sampling has then been taken every 2 ns from the last 24 and 60 ns, respectively, where all thermodynamic properties have been assured to be well equilibrated. The final analysis of vdW energetics has been carried out according to the above protocol.

Section S2. Summary of gas-phase energetics for Fip35-WW

As can be seen from fig. S1, all considered vdW models show comparable behavior along the folding trajectory. The main differences between the approaches can be found for the relative stability of native states with respect to unfolded conformations. Hereby, the pairwise models overestimate the stability as seen from fig. S1 (bottom). The average over-stabilization is 3 kcal/mol in Grimme's D3 and vdW(TS) with screened C_6 interaction coefficients, "vdW(TS)@SCS". For D2 and conventional vdW(TS) we find an overestimation of the relative stability by 4 and 6 kcal/mol, respectively.

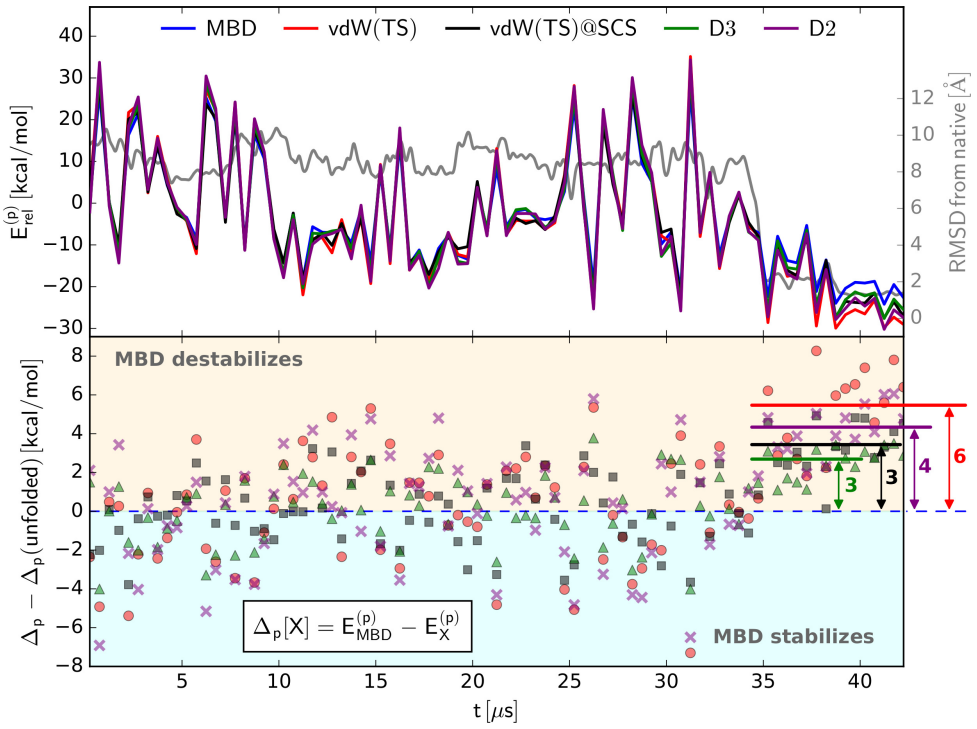

Fig. S1. vdW interaction energy of Fip35-WW in gas phase. Top: Dispersion interaction energy as obtained by all considered models. Bottom: Beyond-pairwise contributions to intraprotein interaction.

Section S3. vdW energetics in detail

3.1 Fip35-WW

Upon embedding in an aqueous environment, we do not observe systematic differences in the relative vdW energetics (*cf.* fig. S2). Also in the case of the pristine solvent, the pairwise models do not show consistent deviations from MBD. Interestingly, for the geometry-based D2 and D3, we additionally find a subtle relative over-stabilization of the pristine solvent corresponding to native state conformations. These add up with minor, yet non-systematic, discrepancies in the total vdW energy and ultimately lead to a overall underestimation of the relative solvation energy of native conformations by 12 kcal/mol for D2 and 7 kcal/mol for D3 (see main manuscript). The latter represents a surprising shortcoming when compared to the electronic structure-based vdW(TS) as D3 outperforms all other pairwise models for gas-phase vdW energies (*cf.* fig. S1). The spurious description of the pure solvent in the geometry-motivated models can most likely be ascribed to an inaccurate description of vdW interactions involving hydrogen atoms located near the cavity and thus mainly affects the pristine solvent. We conclude that, for an accurate description of edge effects in water, it is essential to include electronic-structure effects. In terms of the description of protein-water vdW interactions, ultimately, the pairwise approaches systematically lack a relative stabilization of 5 to 12 kcal/mol due to neglect of beyond-pairwise interaction terms.

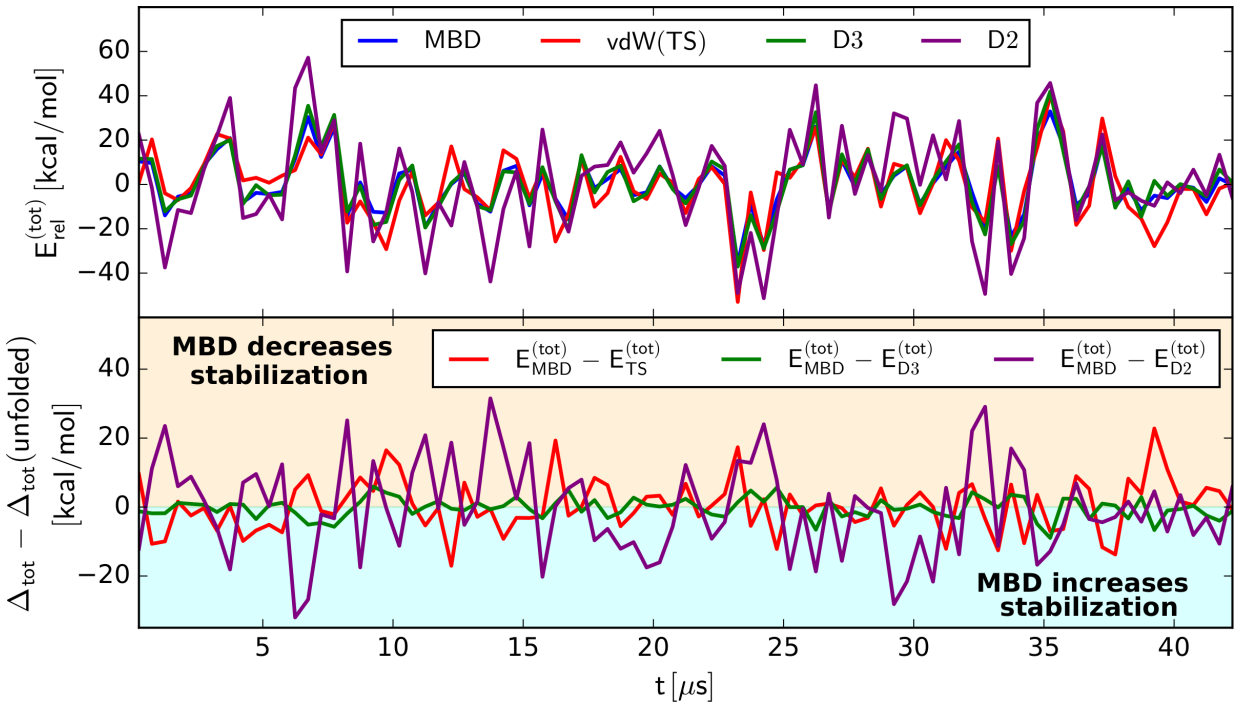

Fig. S2. vdW energetics of Fip35-WW in solvation. Top: Total relative vdW dispersion energy as obtained using MBD and the pairwise approaches vdW(TS), D2, and D3. Bottom: Beyond-pairwise effects on the relative vdW interaction energy of Fip35-WW in aqueous solvation.

3.2 Chignolin variant "cln025"

For the "cln025" variant of the *de novo* protein Chignolin, we observe a similar behavior for the gas-phase energetics. The dispersion interaction energy is maximized when going to the native state and the pairwise vdW model vdW(TS) overestimates the relative stability of the native state of cln025 in the absence of water in comparison to MBD by 5 kcal/mol (see fig. S4). Upon embedding in water, we still find a slight destabilization of native states via many-body dispersion effects (2 kcal/mol), while for the pure solvent we did not observe a statistically relevant change in the relative vdW energy along the folding trajectory. The discrepancy in the relative vdW solvation energy of cln025 between pairwise and many-body treatment, as depicted in fig S3, is thus governed by the neglect of many-body dispersion effects on *intra*-protein vdW energetics during folding and a slight destabilization of native states of cln025 in solvation via many-body effects. The increase of the relative stabilization through beyond-pairwise protein-water vdW interactions amounts to 2 kcal/mol.

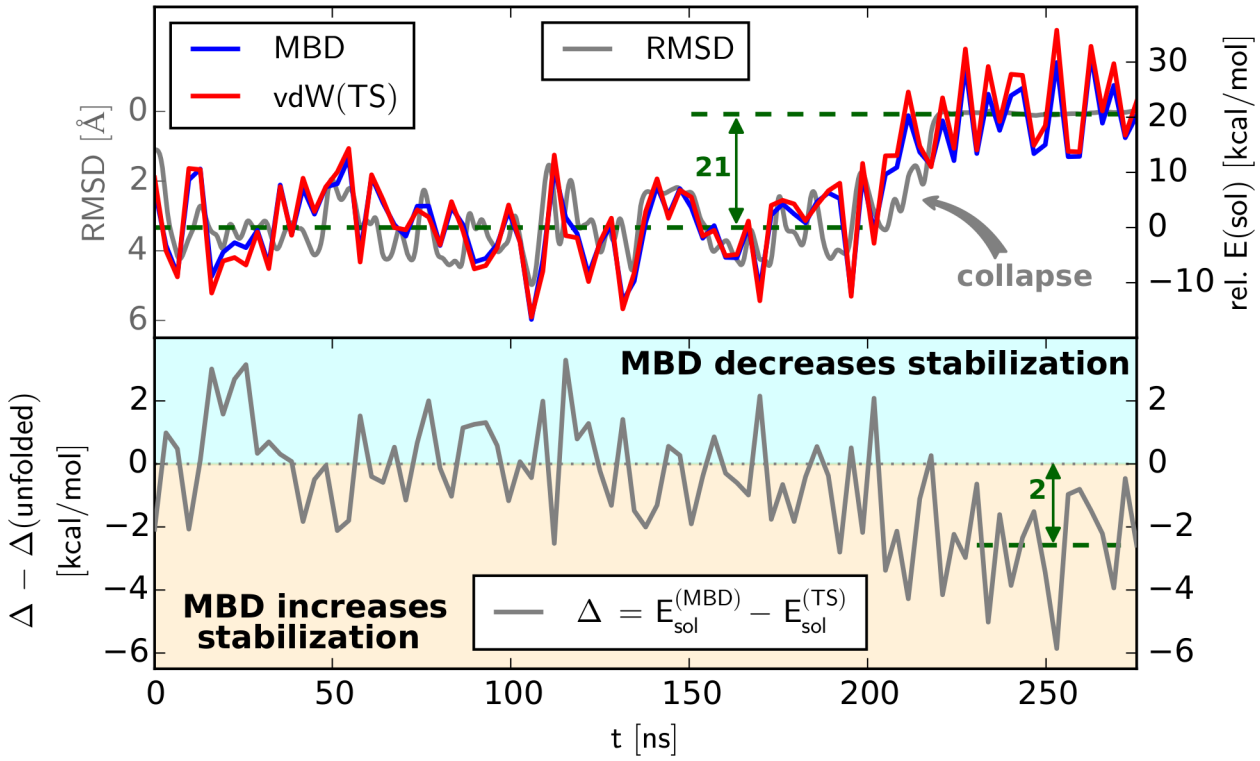

Fig. S3. Relative vdW solvation energy of cln025. Top: Protein-water vdW interaction as obtained with MBD and the pairwise vdW(TS) approach. Bottom: Beyond-pairwise contributions as given by the difference between many-body and pairwise treatment.

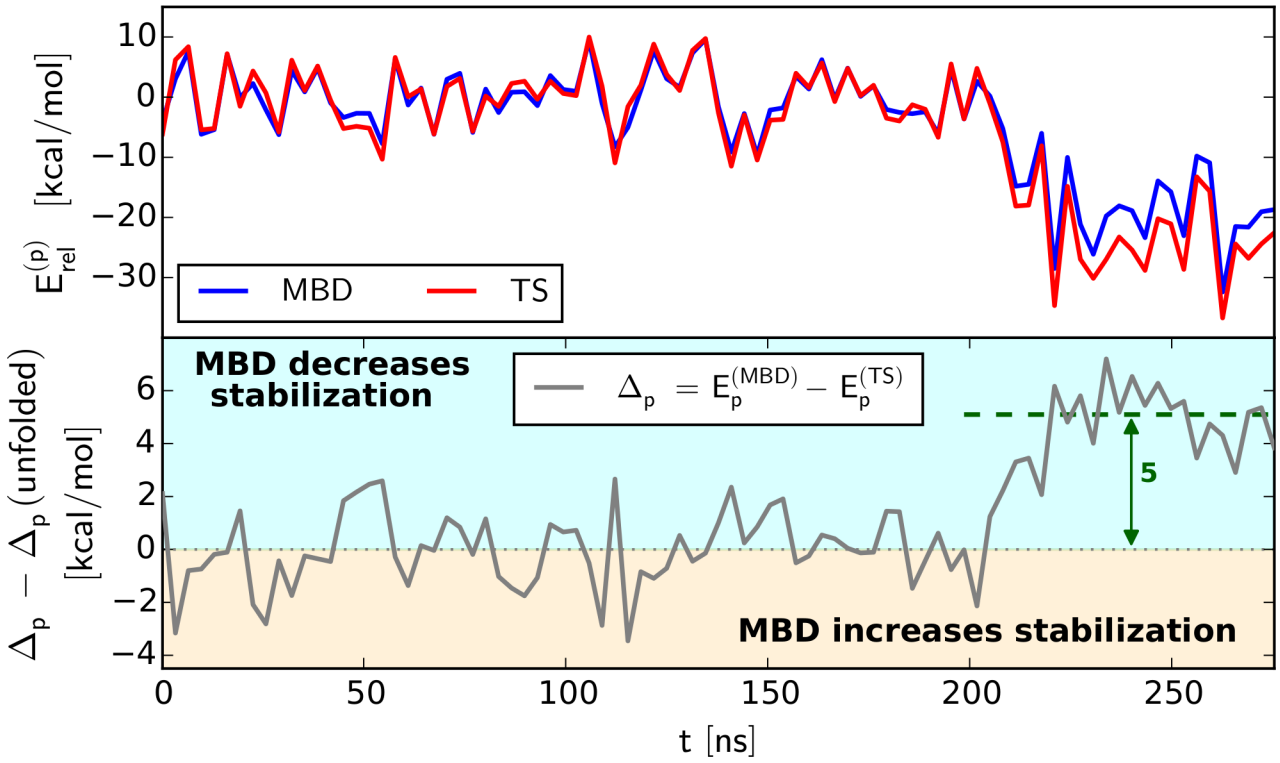

Fig. S4. Relative vdW energetics of cln025 in absence of solvent. Top: MBD and vdW(TS) intraprotein interaction along folding trajectory. Bottom: Beyond-pairwise contributions to relative vdW energetics.

3.3 Villin headpiece 35 Nle/Nle mutant

For HP35-NleNle, many-body dispersion effects destabilize the native state in gas-phase by 4 kcal/mol in comparison to the pairwise vdW(TS). For solvated HP35-NleNle and the pure solvent we did not observe a considerable consistent many-body effect on the energetics. The overestimation of the vdW solvation energy of HP35-NleNle by about 4 kcal/mol when comparing vdW(TS) to a full many-body treatment thus again roots from an inaccurate description of gas-phase energetics.

Section S4. Correlation and rescaling of vdW solvation energies

As can be seen from Fig. S1 and the main manuscript, the overestimation of vdW energies increases with the absolute vdW interaction energy. Correspondingly, a simple rescaling of the pairwise approaches considerably improves the agreement with the many-body treatment. Fig. S5 shows the correlation between such optimally rescaled vdW solvation energies and MBD. The obtained rescaling factors show that relying on electronic-structure based C_6 interaction coefficients as done within vdW(TS) provides the best estimate for vdW solvation energies. This can mainly be attributed to the description of the pure solvent as the geometry-based D2 and D3 methods outperform vdW(TS) for gas-phase energetics (*vide supra*). Despite the overall improvement, the deviation between the optimally rescaled pairwise approaches and MBD still regularly exceeds 4 kcal/mol. Furthermore, the optimal rescaling factors are highly system- and method-dependent and can only be obtained as an *a posteriori* correction.

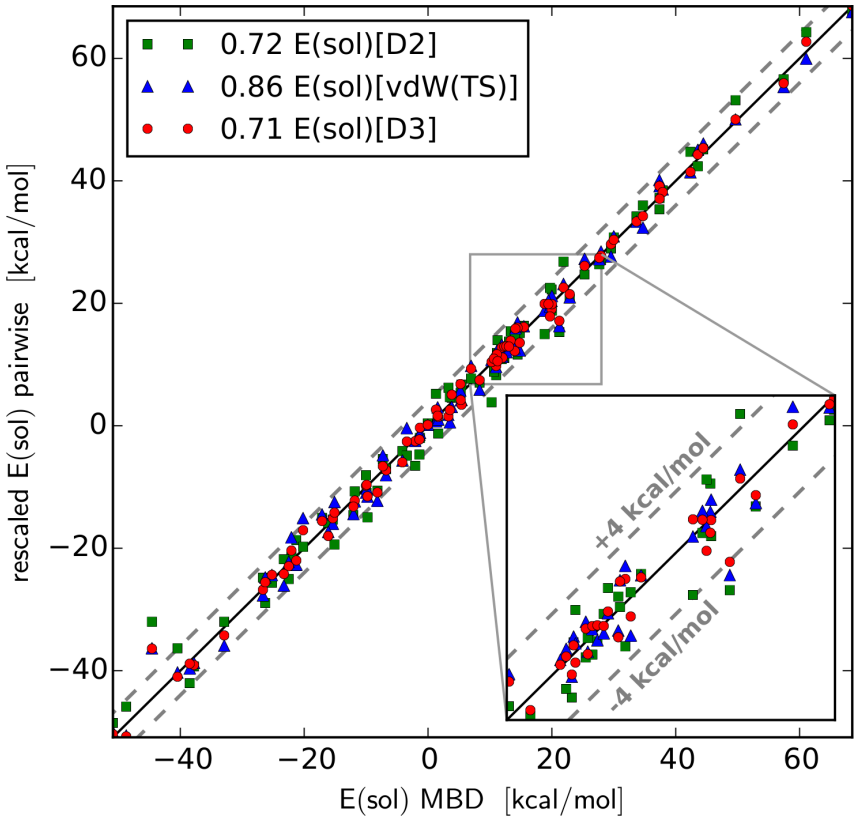

Fig. S5. Correlation of rescaled relative vdW solvation energies as obtained from pairwise models in comparison to the results obtained from many-body treatment.

Section S5. Total electronic energy of solvation

In contrast to the vdW solvation energy, the total electronic energy of solvation does not provide a clear-cut distinction between folded and unfolded states (*cf.* fig. S6 and S7). This is not surprising as the, in the end decisive, free energy of a solvated molecule has a large entropic component and it is known that the hydrophobic effect mainly arises from entropic contributions (*5*). In the case of cln025, we do observe a slight shift in the total electronic energy of solvation when comparing folded and unfolded states, see fig. S7. Considering the absolute spread of the solvation energy, however, this shift is less clear as for its

mere vdW component (fig. S3) and coincides with the spread of total solvation energies of the unfolded conformations.

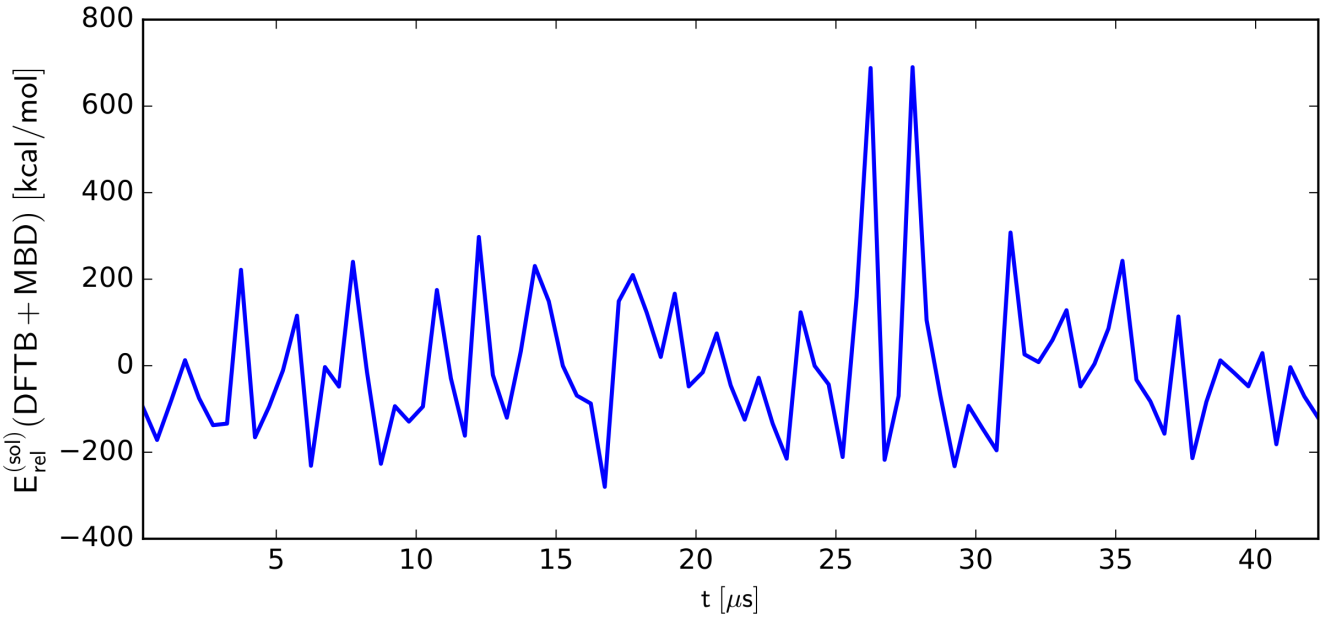

Fig. S6. Total electronic energy of solvation of the Fip35-WW as obtained with DFTB in conjunction with the MBD model.

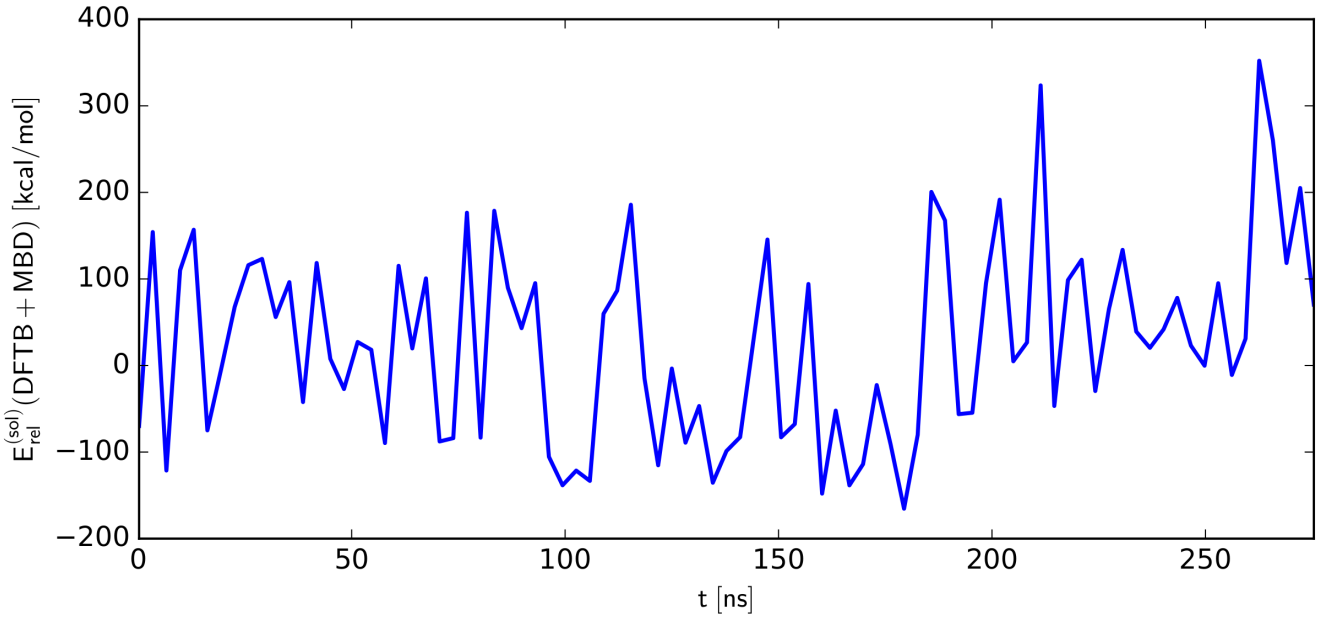

Fig. S7. Total electronic energy of solvation of the chignolin variant cln025 as obtained with DFTB in conjunction with the MBD model.

Section S6. Effect of overcompaction of unfolded states

As detailed in the manuscript, conventional molecular mechanics force fields in conjunction with traditional water models such as the standard TIP3P, predict unfolded protein states that are too compact. This has largely been assigned to an unbalanced description of *intra*-protein and protein-water vdW interactions. Our present findings thereby provide a fundamental, quantum-mechanical explanation for this argument: Due to many-atom effects beyond the traditional pairwise formulation of vdW energetics, we observe a system- and conformation-dependent change in the relative magnitude of the individual vdW contributions. These findings, however, have so far been based on structures obtained employing the above mentioned unbalanced description and thus featured unfolded states, which are too compact. As an ultimate confirmation of our results, we have performed our analysis based on a new sampling of the proteins conformational space using the recent *a99SB-disp* force field together with the TIP4P-D water model as developed by Robustelli *et al.*. This setup has been designed and shown to provide a superior description of unfolded and disordered protein states avoiding spurious over-compaction (*34*). To study potential effects of this over-compaction, we performed new molecular dynamics simulations of the chignolin variant "cln025" in explicit water starting from an unfolded and folded state, respectively. Subsequently, we evaluate the vdW energetics in the same manner as above, where the only difference is the usage of a more correct and representative sampling of unfolded states as obtained from the new molecular dynamics simulations.

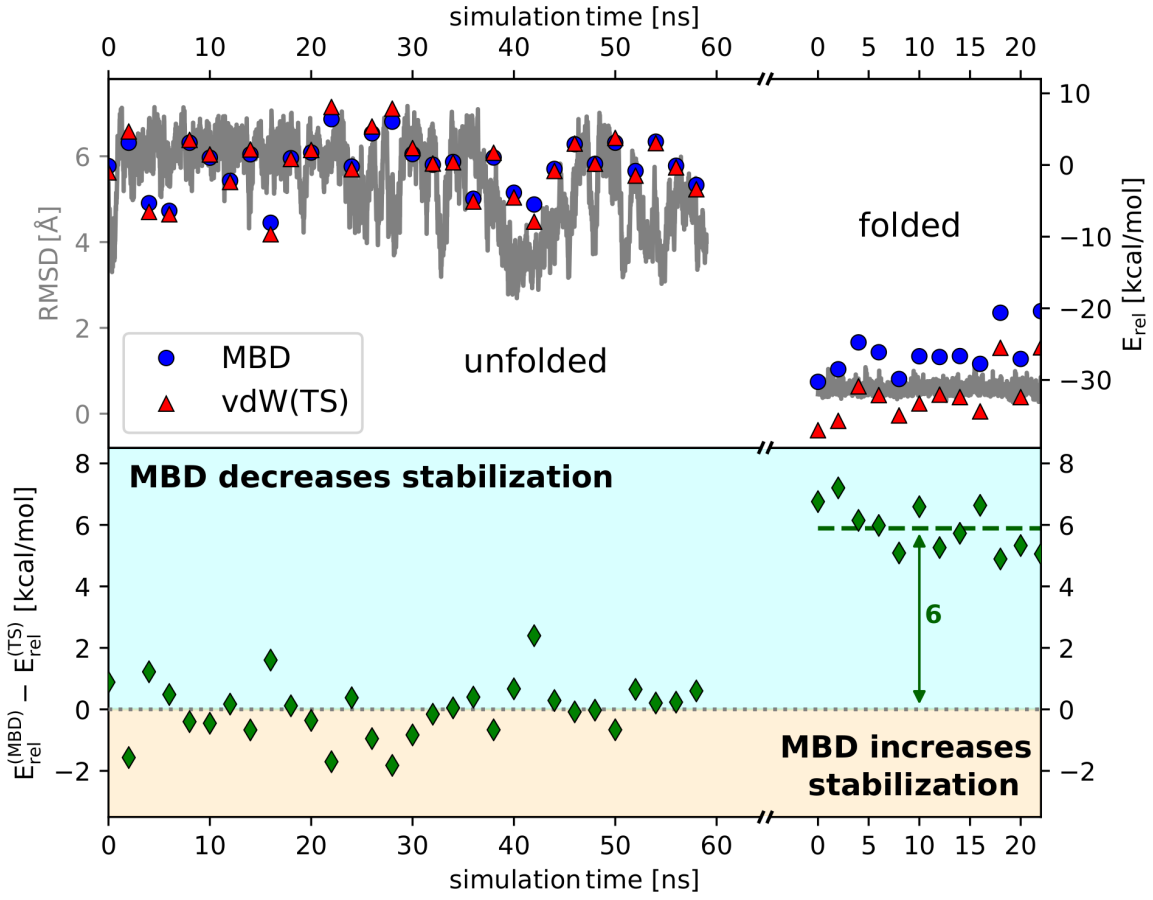

Fig. S8. Intraprotein vdW interaction energy of cln025 based on improved structural sampling. Top: relative intraprotein vdW interaction as obtained from MBD formalism and pairwise description in vdW(TS). Backbone RMSD (grey) taken with respect to native state. Bottom: Many-body contributions as defined by the difference between MBD and vdW(TS).

Figure S8 shows the *intra*-protein vdW interaction energy based on this new trajectories (left: unfolded state sampling, right: folded state sampling) in the top graph. The RMSD of both samplings is taken with respect to the native conformation. The bottom graph again depicts the difference between the pairwise description in vdW(TS) and the many-body energetics obtained from MBD. All in all, fig. S8 confirms our previous findings of the pairwise formalism overestimating the internal stabilization in the pairwise description.

As can be seen from fig. S9, the improved sampling of unfolded conformations neither affects our conclusions for the vdW solvation energy: It still tracks with the (inverse) geometrical RMSD (top graph), and many-body effects increase the relative protein-water vdW interaction in the native state (bottom graph). In conclusion, the results obtained for the previous trajectories (as detailed above and in the main manuscript) can clearly be assigned to a failure of the pairwise approximation and do not simply represent an artifact of the incorrect sampling of unfolded states.

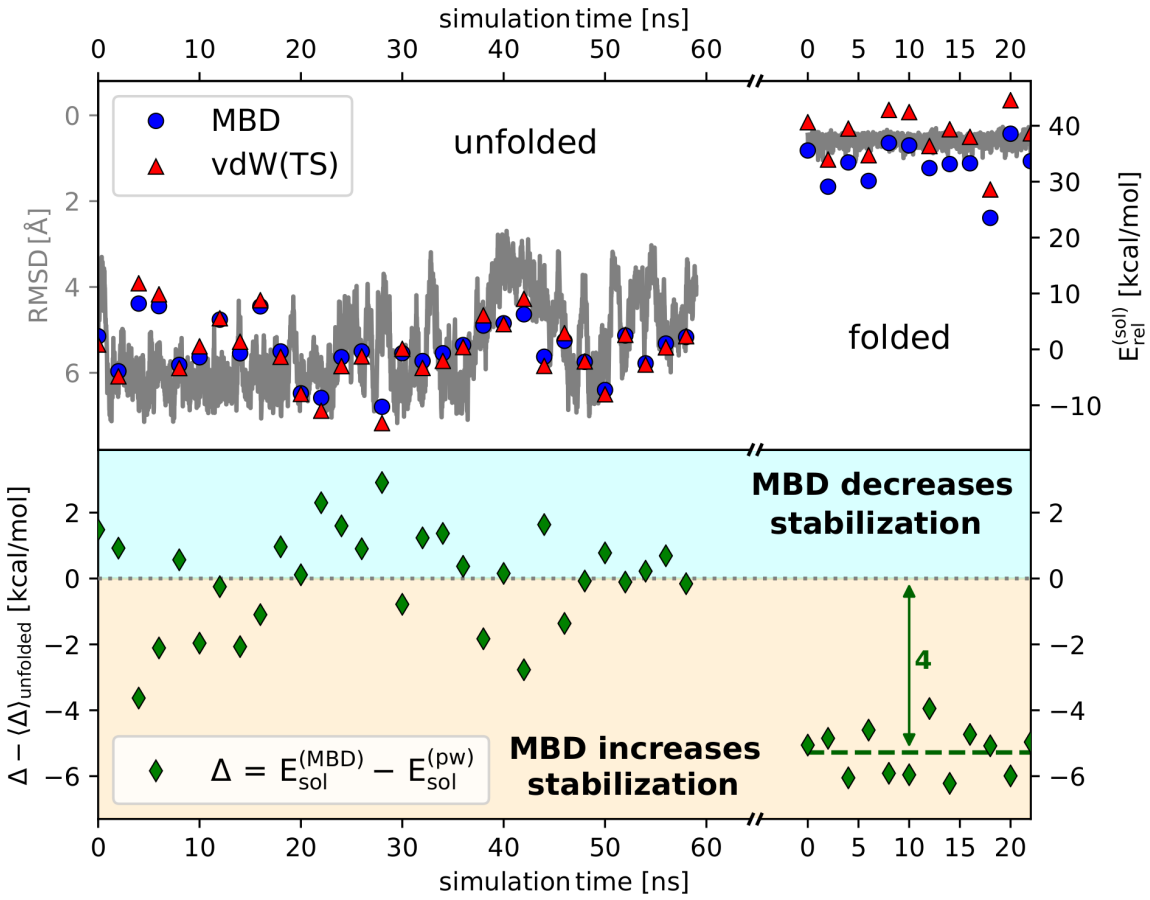

Fig. S9. Relative vdW solvation energy of cln025 based on improved structural sampling. Top: relative vdW solvation energy as obtained from MBD formalism and pairwise description in vdW(TS). Backbone RMSD (grey) taken with respect to native state. Bottom: Beyond-pairwise contributions as given by the difference between MBD and vdW(TS).