Title:

Force Response of Polypeptide Chains from Water-Explicit MD Simulations

Author(s):
Richard Schwarzl, Susanne Liese, Florian N. Brünig, Fabio Laudisio, and Roland R. Netz

Citation:

"Macromolecules 2020, 53, 12, 4618–4629 ; https://doi.org/10.1021/acs.macromol.0c00138"

Force response of polypeptide chains from water-explicit MD simulations

Richard Schwarzl,[†] Susanne Liese,[†,‡] Florian N. Brünig,[†] Fabio Laudisio,[†] and Roland R. Netz[*,†]

[†]*Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany*

[‡]*Department of Mathematics, University of Oslo, 0851 Oslo, Norway*

E-mail: rnetz@physik.fu-berlin.de

Abstract

Using molecular dynamics simulations in explicit water, the force-extension relations for the five homopeptides polyglycine, polyalanine, polyasparagine, polyglutamic acid and polylysine are investigated. From simulations in the low force regime the Kuhn length is determined, from simulations in the high-force regime the equilibrium contour length and the linear and non-linear stretching moduli, which agree well with quantum-chemical density-functional theory calculations, are determined. All these parameters vary considerably between the different polypeptides. The augmented inhomogeneous partially freely rotating chain (iPFRC) model, that accounts for side-chain interactions and restricted dihedral rotation, is demonstrated to describe the simulated force-extension relations very well. We present a quantitative comparison between published experimental single-molecule force-extension curves for different polypeptides with simulation and model predictions. The thermodynamic stretching properties of polypeptides are investigated by decomposition of the stretching free energy into energetic and entropic contributions.

Introduction

Proteins consist of amino acids that are linked by peptide bonds. Numerous biological processes depend on the mechanical response of proteins[1–6] and the understanding of such processes requires as a first step the accurate description of the stretching response of single peptide chains. The elastic response of single protein molecules has been thoroughly investigated experimentally by atomic force microscopy (AFM).[7–13] The experimentally measured force-stretching relations of unfolded peptide chains are typically compared with the freely jointed chain (FJC)[14–16] model or the worm-like chain (WLC)[17,18] model. The WLC model has been shown to describe relatively stiff biological polymers such as double-stranded DNA[19] very well, and the persistence length that results from a fit to the force-extension relationship is related to the DNA bending stiffness and thus has physical meaning. For flexible polymers, which at the atomic scale consist of chemical bonds with a preferred bond length and preferred bond angle, neither the FJC nor the WLC model is microscopically correct. In fact, for the stretching of titin immunoglobulin domains, a multitude of different theoretical polymer models have been used that account for various molecular aspects such as fixed or extensible bond lengths, fixed bond angles, dihedral potential effects as well as specific interactions between polymer backbone and solvent.[9,20–22]

While experimental force-extension data of polypeptides is routinely fitted with the freely-jointed chain (FJC) model, which is equivalent to the Langevin model for the alignment of a dipole in an external field, the bond length that results from such a fit has no physical meaning and in particular does not agree with the actual chemical peptide bond length. As a step towards a more accurate description of experimental force-extension data of polypeptides, the stretching response of a freely rotating chain (FRC) model, which includes a fixed bond angle, was favorably compared with experimental data.[23,24] The disadvantage of the FRC model is that the force-extension relation is not available in closed form but must be calculated numerically by transfer matrix techniques. The resulting stretching behavior is quite complex and exhibits three distinct scaling regimes, a low-force linear regime, an intermedi-

ate regime that is similar to the response of a WLC chain, and finally a high force regime that can be mapped onto the FJC model.[23] But a peptide chain is also not accurately described by the FRC model, since the peptide backbone bond lengths and angles are not all the same and, more importantly, strong dihedral potentials exist that are very different for different bonds along the peptide backbone. In fact, one of the bonds, the so-called peptide bond, is almost unable to rotate. Motivated by this, the more general inhomogeneous partially freely rotating chain (iPFRC) model[25] was introduced, which, in addition to the inhomogeneous bond lengths and bond angles along the peptide backbone, accounts for the stiff peptide bond and fixes its dihedral angle to 180°. All these models are ideal and neglect interactions between amino acids as well as hydration effects, which for PEG have been demonstrated to dramatically influence the stretching behavior.[26] Particularly important for peptides are the interactions between side chains of neighboring amino acids, which are difficult to treat in theoretical polymer models.

In this paper, we investigate the force-extension relation of five different homopeptides consisting of 13 amino acids by means of molecular dynamics (MD) simulations in explicit water. In analogy to AFM experiments, we apply a constant force to the outermost C_α atoms and determine the average end-to-end distance between those C_α atoms, as schematically shown in Figure 1a and as done earlier for PEG and PNIPAM.[26,27] We choose polyglycine (GLY), which consists of only the peptide backbone and hence represents the structurally simplest peptide, and in addition polyalanine (ALA), polyasparagine (ASN), polyglutamic acid (GLU) and polylysine (LYS). These peptides span a wide range from hydrophobic to hydrophilic and charged polymers.

In order to separate effects due to interactions between neighboring amino acids from effects due to interactions between non-neighboring amino acids, we compare stretching simulations of tripeptides with stretching simulations of peptides consisting of 13 amino acids, both in water. From this comparison we conclude that interactions between side chains of neighboring amino acids are strong and modify significantly the unstretched peptide contour

length and the backbone extensibility. By comparison of tripeptide simulations in water at room temperature and in vacuum at zero temperature, we demonstrate that hydration effects do not modify the peptide contour length and the peptide contour extensibility, which we explain by a detailed hydrogen bonding analysis. The comparison with density-functional quantum-chemistry calculations of stretched tripeptides shows that the MD peptide force fields are well parameterized in terms of the backbone contour length and its extensibility. We then extract the peptide Kuhn length from the low-force stretching response of 13-mers. An analysis of dihedral angle distributions in terms of Ramachandran plots demonstrates that indeed the peptide bond is not rotating, which confirms the basic assumption of the iPFRC model. By comparison of the iPFRC model with the MD simulations at large stretching force, we determine the single undetermined parameter in the iPFRC model and show that it provides an excellent description of MD stretching curves over the entire force range if the contour length and contour extensibility are correctly taken into account. Finally, we compare the force-extension relations from MD simulations and the iPFRC model with experimental AFM results obtained previously[24] for two different heteropolymers, titin and spider silk (GVGVP), and for polylysine. The agreement is overall good but shows some discrepancy in the 200 pN force range, which might be due to peptide bond isomerization effects. We also determine the energetic and entropic contributions to the stretching response and find that water bridges, where one water forms two hydrogens bonds with the polypeptide, play only a marginal role in thermodynamically stabilizing homopeptide conformations against stretching. We conclude that the iPFRC model is an accurate model for the description of peptide stretching, provided that effects due to restricted backbone dihedral rotation and side-chain interactions are accounted for by suitably choosing the single fit parameter of the iPFRC model.

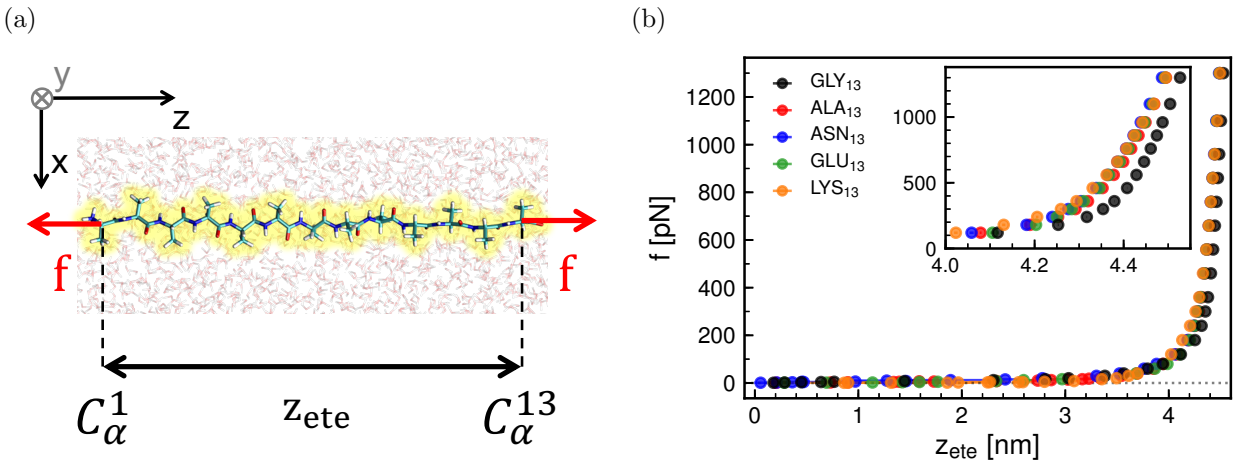

Figure 1: (a) Snapshot of an MD simulation of ALA_{13} at a stretching force of $f = 1300\,\mathrm{pN}$. (b) Simulation results for the force-extension relation of five different polypeptides. The inset highlights the high-stretching regime.

Results

In Figure 1a we show an MD simulation snapshot of an alanine chain with 13 residues, which we abbreviate by ALA_{13}, solvated in water at a stretching force of $f = 1300\,\mathrm{pN}$. Constant stretching forces f, parallel to the z-axis, act in opposite directions on the first and last C_α atoms. For each force, we calculate the time-averaged end-to-end distance z_{ete} in z-direction between these two C_α atoms. In Figure 1b we present force-extension results for all five polypeptides investigated in this study. All simulations are performed without added salt, in the supplementary information, salt effects are shown to be negligible in the the force range considered by us. Deviations between different peptides are revealed in the inset, where we show the large stretching regime. Of the five polypeptides, polyglycine shows the largest end-to-end distance z_{ete} for a given force, as we will explain based on an analysis of the contour lengths of the different peptides later on.

In order to quantify the stretching response of different peptides, we first need to determine the contour length in the absence of an external force, which need not be the same for the different peptides. For this we probe the stretching response of a tripeptide, which is the smallest subunit of a polypeptide for which the separation vector between the outer

C_α atoms in the stretched state is parallel to the end-to-end vector of the polypeptide.[24] This is so because one amino acid residue consists of an odd number of backbone atoms, namely three. Figures 2a and 2b show as solid lines the potential energy landscapes of all tripeptides in vacuum at zero temperature as a function of the separation between the outer C_α atoms, denoted by $2a$, which are obtained from ground state optimization of the MD force field (see Figure 2c for a snapshot and illustration of the geometry), where a denotes the contour length per amino acid residue.

The results demonstrate the existence of different competing ground states that exhibit sharp transitions as different ground state energies cross. The dotted lines in Figures 2a and 2b correspond to the free energy landscapes obtained from MD umbrella simulations of tripeptides in water at room temperature T=300 K. For large stretching the differences between the corresponding solid and broken lines are very small, which means that hydration as well as conformational fluctuation effects are not important in the stretched state. For small values of $2a$, i.e. in the unstretched and compressed state, the differences between the vacuum $T = 0$ K results and the hydrated finite-T results are significant and the sharp transitions between different groundstates that are visible in the $T = 0$ K results are expectedly washed out at room temperature.

In Figures 2e and 2f we show the stretching forces of tripeptides, which are obtained from the numerical derivative of the tripeptide energy and free energy landscapes. We define the equilibrium monomer contour length a_0 by the largest monomer length at which the force vanishes and base our results, which are denoted by vertical colored arrows, on the forces obtained within the vacuum calculations (solid lines). The resulting equilibrium monomer lengths are 0.74 nm for glycine, 0.72 nm for alanine and 0.70 nm for asparagine, glutamic acid and lysine. The fact that the glycine equilibrium monomer length is significantly larger than for the other investigated peptides is explained by the fact polyglycine forms a perfectly planar structure in the ground state, while the other polypeptides exhibit non-planar ground states due to repulsive interactions between side chains and the backbone. This is illustrated

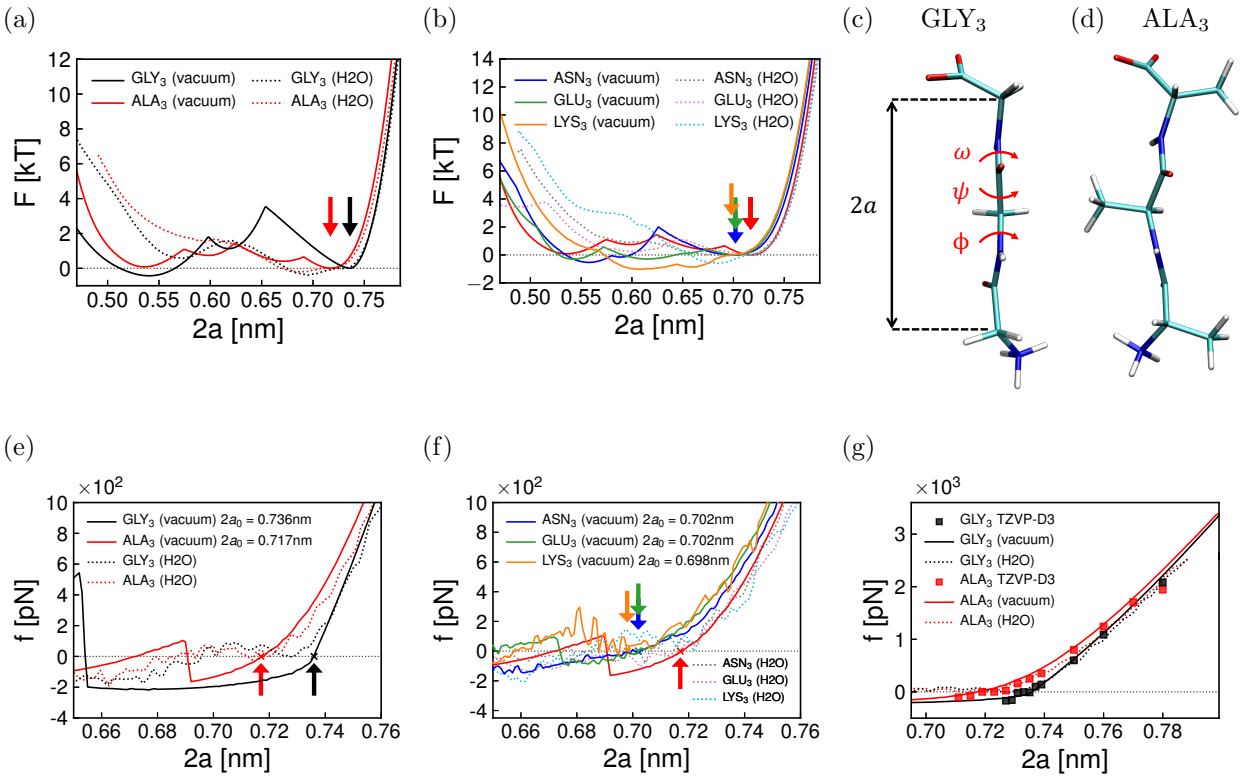

Figure 2: (a+b) Comparison of energy landscapes in vacuum at $T = 0\,\mathrm{K}$ (solid lines) and free energy landscapes in water at $T = 300\,\mathrm{K}$ (dotted lines) from MD simulations as function of the separation between the terminal C_α atoms, $2a$, for (a) GLY_3, ALA_3 and (b) ASN_3, GLU_3, LYS_3 (ALA_3 in red is shown for reference). Ground state conformations for a strongly stretched state ($2a = 0.77\,\mathrm{nm}$) of (c) GLY_3 and (d) ALA_3 obtained from the MD force field in vacuum at $T = 0\,\mathrm{K}$. The structures reveal that the backbone of ALA_3 is non planar, which is due to repulsive side-chain backbone interactions. GLY_3 on the other hand adopts a planar state at the same separation. (e+f) Stretching forces of the different tripeptides, obtained by numerical derivative of the energy and free energy landscapes in (a) and (b). The largest separation $2a_0$ where the stretching force in vacuum vanishes is used to define the equilibrium monomer lengths a_0, which are denoted by vertical colored arrows. Note that the range of separations shown is decreased compared to (a) and (b) in order to focus on the equilibrium configuration with separation $2a_0$. (g): Comparison of stretching forces from quantum chemistry DFT in vacuum at $T = 0\,\mathrm{K}$ (data points), MD force-fields in vacuum at $T = 0\,\mathrm{K}$ (solid lines), MD force-fields in water at $T = 300\,\mathrm{K}$ (dotted lines) for the stretching force of alanine and glycine tripeptides.

in Figure 2c and d, where we show MD snapshots in vacuum and at $T = 0\,\mathrm{K}$ of GLY_3 and ALA_3 at the same fixed separation $2a = 0.77\,\mathrm{nm}$. This fixed separation corresponds to a significant stretching force in the range of 2 nN. We see that the backbone of alanine is non planar even at such a strong stretching force.

In Figure 2g we compare DFT results in vacuum at $T = 0\,\mathrm{K}$ with MD force-field results

in vacuum at $T = 0\,\mathrm{K}$ and in water at $T = 300\,\mathrm{K}$ for the stretching force for glycine and alanine tripeptides. DFT and MD results agree very well in the relevant force regime between 0 pN and 1500 pN, which reflects that amber backbone force-field parameters have been optimized using gas-phase QM calculations.[28] The DFT results confirm the increased resistance to stretching of alanine compared to glycine. An analysis of the MD interaction energies in vacuum for $\mathrm{ALA_3}$ and $\mathrm{GLY_3}$ (cf. supporting information) shows that the smaller contour length of alanine is caused by repulsive Lennard-Jones interactions between the side-chain and the backbone. This means that backbone rotational degrees of freedom are restricted by the presence of side-chains in alanine, the same mechanism acts for the other polypeptides as well.

In order to further understand the structural differences between polyglycine and the other polypeptides, we investigate backbone dihedral angles of polypeptides in water. The nomenclature for the different backbone dihedral angles is defined in Figure 2c, ϕ denotes the $\mathrm{C{-}C_\alpha{-}N{-}C}$ dihedral angle and ψ denotes the $\mathrm{N{-}C{-}C_\alpha{-}N}$ dihedral angle. We calculate the dihedral angles ϕ and ψ for zero stretching force and for a large stretching force of 1300 pN by averaging over all 12 equivalent dihedral angles in the polypeptide. The resulting Ramachandran plots for zero force for $\mathrm{GLY_{13}}$ and for $\mathrm{ALA_{13}}$ are shown in Figure 3a and Figure 3b. We see that for glycine the distribution of the dihedral ϕ is symmetric and rather broadly distributed, whereas for alanine the distribution is asymmetric and restricted to a smaller region of the conformational space, which demonstrates the restricted backbone rotation due to side-chain steric hindrance. The symmetry breaking persists to large stretching forces of 1300 pN in Figure 3c and Figure 3d. These results show that even at large stretching forces alanine is not planar, while glycine is perfectly planar at such high forces. That this effect is not specific to alanine but holds also for all other investigated polypeptides, except glycine, can be seen in the plots Figures 3e to 3g. The intercept of the broken lines in Figure 3 indicates the planar state, where $\phi = 180°$ and $\psi = 180°$.

We next analyze the stretching response of all tripeptides quantitatively. In Figure 4

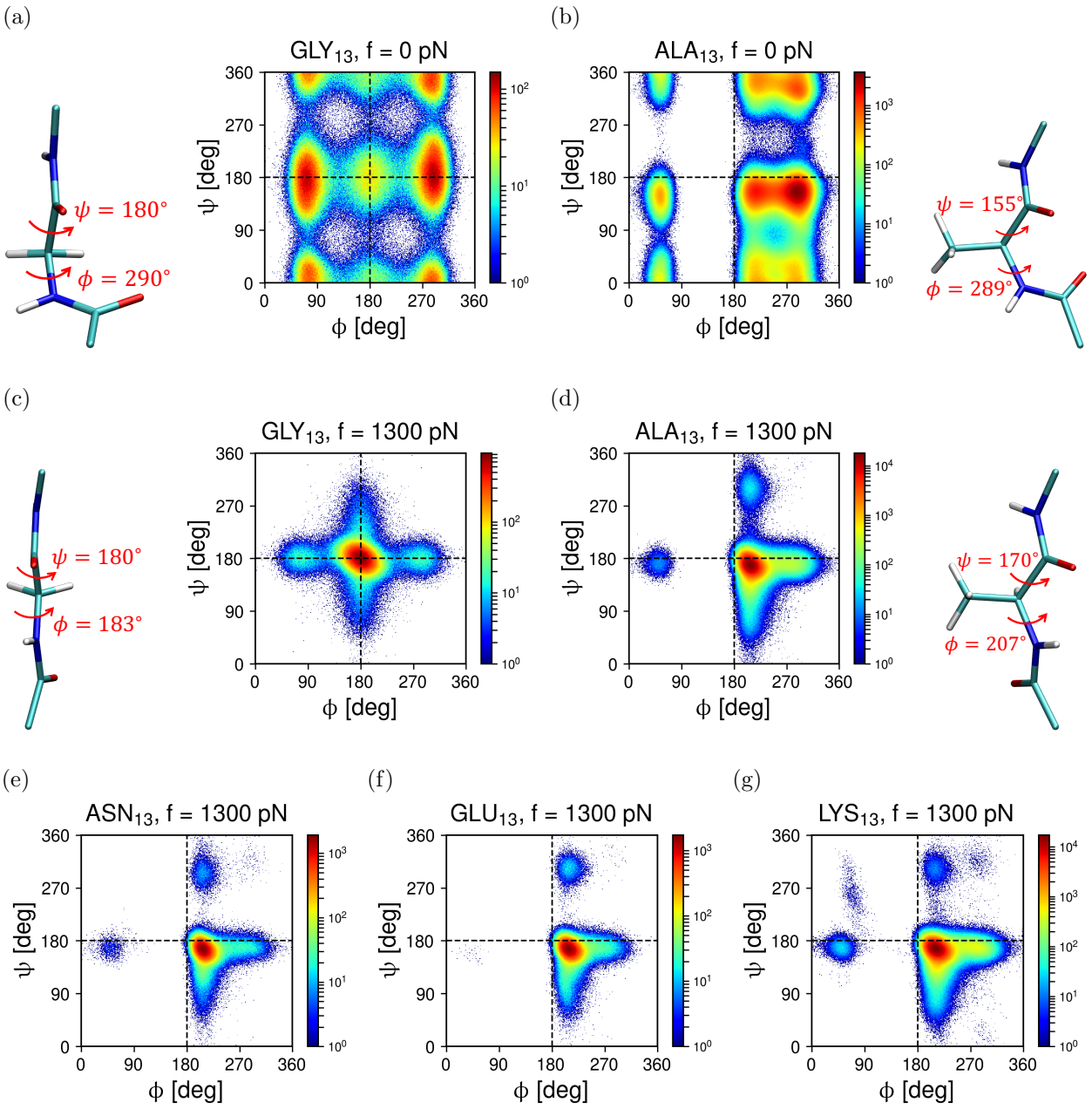

Figure 3: (a - d): Ramachandran plots for GLY_{13} and ALA_{13} at 0 pN and 1300 pN from MD simulations in water. All ϕ and ψ angles of the polypeptides are averaged. Side-chain interactions in polyalanine shift the most probable (red) region away from the symmetric planar state $\phi = 180°$, $\psi = 180°$ (intercept of the broken lines). The results for polyglycine on the other hand indicate a planar conformation. The results are in close agreement with previously published Ramachandran plots of GLY_3 and ALA_3.[28–31] MD simulations snapshots in water show the most probable conformations of a part of the polypeptides. (e - g): Ramachandran plots of ASN_{13}, GLU_{13} and LYS_{13} at a stretching force of 1300 pN show that all peptides except glycine adopt non planar most probable states.

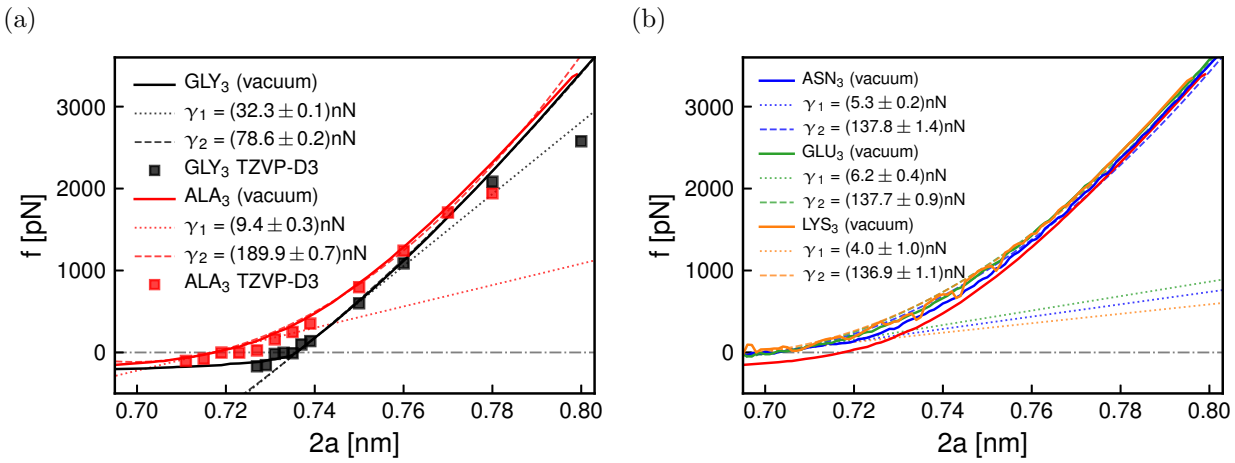

Figure 4: (a) Stretching force for GLY_3 and ALA_3 tripeptides in vacuum and at $T = 0\,\mathrm{K}$ as a function of the extension $2a$. Solid lines denote results from MD force fields and squares denote results from DFT. The first order (dotted lines) and second-order (broken lines) elastic responses defined in eq 5, which are fitted to the MD results, illustrate the relative importance of the stretching moduli γ_1 and γ_2. (b) Results for ASN_3, GLU_3 and LYS_3. Results for ALA_3 in red are included for reference.

we again show our results for the stretching force of tripeptides in vacuum and at $T = 0\,\mathrm{K}$ obtained from DFT and MD force fields as a function of the end-to-end distance $2a$. We fit the stretching force from the MD zero-temperature calculations by a second-order polynomial function, similar to a previous analysis,[24] given by

$$f = \gamma_1 \left(\frac{a}{a_0} - 1 \right) + \gamma_2 \left(\frac{a}{a_0} - 1 \right)^2 , \tag{1}$$

where we use the equilibrium monomer lengths a_0 as determined in Figures 2e and 2f. The linear stretching modulus coefficient γ_1 is determined by a local fit for monomer lengths a above the equilibrium monomer length a_0 in the range $0 \leq (a/a_0 - 1) < 0.015$. For the non-linear coefficient γ_2 the full range of data is used for fitting. The fits to the MD results are presented in Figure 4 as dotted lines (including only the first order) and as dashed lines (including first and second-order). It is seen that non-linear stretching effects are rather small for glycine while non-linear effects are important for alanine. This can be explained by the fact that alanine is rather soft for small stretching due to its non-planar ground state

and hardens for larger stretching, similar to results observed previously for the stretching of RNA molecules.[24] We conclude that the non-linear model eq 1 fits the $T = 0\,\mathrm{K}$ stretching response of tripeptides very well over the entire force range.

The parameters extracted so far describe the zero-temperature stretching response of polypeptides and thus neglect the effects of conformational fluctuations. Finite-temperature fluctuation effects in fact become important for stretching forces smaller than roughly 1 nN.[24] For low forces the stretching response of polymers is characterized by a linear relation between the end-to-end extension $\mathrm{z_{ete}}$ (normalized by the equilibrium unstretched contour length L_0) and the applied stretching force f according to

$$\frac{\mathrm{z_{ete}}}{\mathrm{L_0}} = \frac{f a_{\mathrm{Kuhn}}}{3 k_{\mathrm{B}} T}, \tag{2}$$

which defines the Kuhn length a_{Kuhn}.[32–34] This relation holds if the low-stretching condition $f a_{\mathrm{Kuhn}} < k_{\mathrm{B}} T$ is satisfied and if monomer-monomer interactions are neglected.[35,36] Since the polypeptides used in our simulations are rather short, the transition from the linear stretching response $\mathrm{z_{ete}} \sim f$ to the non-linear Pincus stretching response $\mathrm{z_{ete}} \sim f^{2/3}$, which

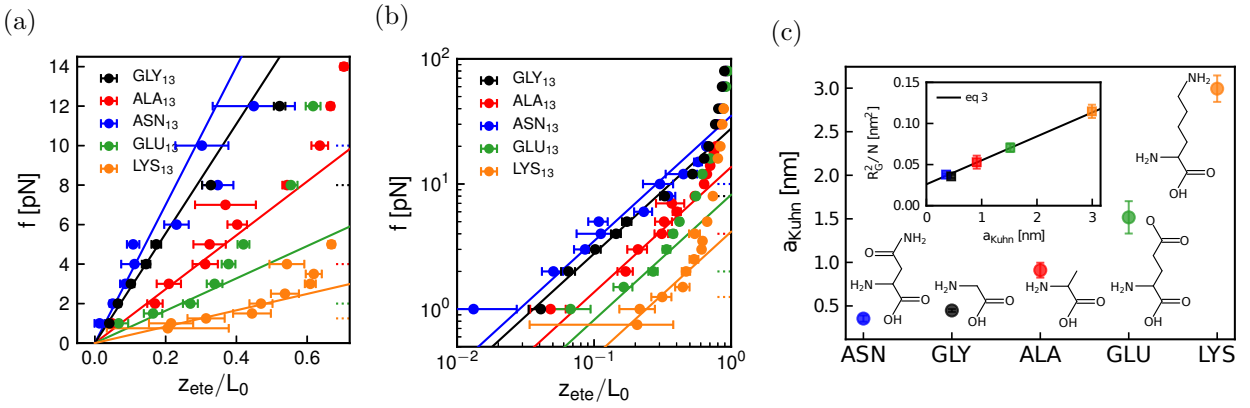

Figure 5: (a): Low-force MD simulation results. Solid lines depict linear fits according to eq 2 and yield estimates for the Kuhn length a_{Kuhn} for the different polypeptides. (b): Same results as in (a) on a double-logarithmic scale. (c): Results for the Kuhn length are shown in ascending order, corresponding chemical peptide structures are included next to each data point. The inset shows the correlation between the squared radius of gyration divided by $N = 12$ at zero stretching force and the Kuhn length according to eq 3.

is expected to occur in the presence of repulsive monomer-monomer interactions at very low stretching forces, cannot be observed.[37–39] In the supporting information we demonstrate that the addition of salt into our simulations does not change the stretching response, we conclude that electrostatic repulsive interactions between charged side chains are not relevant in the force regime we are investigating. Figure 5a shows the MD force-extension data in the small force regime. We extract a_{Kuhn} from the MD simulation results by linear fits in the force range $0 \leq f < f_{\mathrm{max}}$, indicated by solid straight lines. The upper bound f_{max} is determined by the aforementioned condition $f_{\mathrm{max}} a_{\mathrm{Kuhn}} = k_{\mathrm{B}}T$ and is indicated by dotted horizontal lines in Figure 5a. The contour length L_0 at zero force in eq 2 is given by $12 \times a_0$, where the equilibrium monomer lengths a_0 are taken from Figures 2e and 2f. The resulting Kuhn lengths are shown in Figure 5c for all different peptides. Except for asparagine, a larger side-chain size leads to a larger Kuhn length. The inset of Figure 5c shows the squared radius of gyration R_{G} at zero force as a function of the Kuhn length. The relation between the squared radius of gyration R_{G}^2 and the Kuhn length a_{Kuhn} is described by the heuristic relation

$$R_{\mathrm{G}}^2/N = \alpha a_{\mathrm{Kuhn}} a_0 + \beta, \tag{3}$$

where $\alpha = 0.04$ and $\beta = 0.025\,\mathrm{nm}^2$ are the best-fit values. The presence of a finite offset β is unexpected but obviously needed to describe the data.

We have so far determined the most important polymer parameters of the different polypeptides investigated in this study, namely the equilibrium unstretched monomer length a_0, from which the contour length $L_0 = Na_0$ in the absence of a stretching force can be calculated, the Kuhn length a_{Kuhn}, and the non-linear elastic coefficients that characterize the stretching response of the contour length. Notably, all these parameters are rather different for different polypeptides. The Kuhn length describes the low-force stretching response, while the elastic parameters describe the high-force stretching response of polymers. We now set out to describe the polypeptide stretching response at intermediate forces. Based on the stretching response of a partially restricted freely-rotating chain model,[23] a simple heuris-

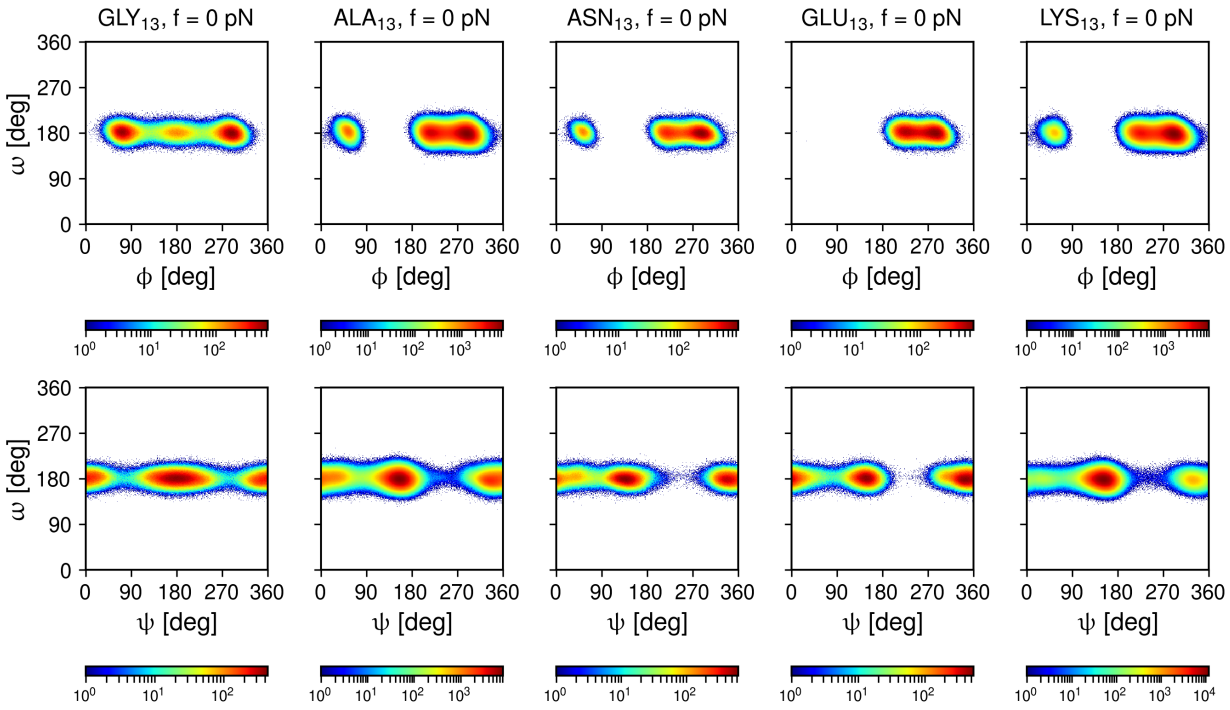

Figure 6: Ramachandran plots of the dihedral angles for zero stretching force demonstrate that the dihedral angle ω that characterizes rotations around the peptide bond is restricted and fixed to 180°. The other two dihedral angles, ϕ and ψ, exhibit rather broad distributions the shapes of which depend sensitively on the peptide type.

tic formula for the force-extension relation of polypeptides, which omits the intermediate worm-like-chain stretching regime, was given as [25]

$$f_{\mathrm{iPFRC}} = \frac{k_{\mathrm{B}}T\,\mathrm{z_{ete}}}{\mathrm{L}(f)}\left(\frac{3}{a_{\mathrm{Kuhn}}} + \frac{1}{c\;a_0}\frac{\mathrm{z_{ete}}/\mathrm{L}(f)}{1-\mathrm{z_{ete}}/\mathrm{L}(f)}\right). \tag{4}$$

Note that this expression is constructed such as to reproduce both the low-force and the high-force limits of the stretching response of a rotating chain model. We here generalize the iPFRC model by allowing the contour length $L(f)$ to be force-dependent, which is motivated by the results shown in Figures 4a and 4b. As a main feature, the inhomogeneous partially freely rotating chain (iPFRC) model accounts for the fact that the dihedral angle ω, which describes the peptide bond, defined by $\mathrm{C}_\alpha\mathrm{-N-C-C}_\alpha$, is fixed at 180°, see Figure 2c for a graphical definition. In the original iPFRC model, the other two dihedral angles, ϕ and ψ (as defined in Figure 2c), are free to rotate, which clearly is an approximation. To test

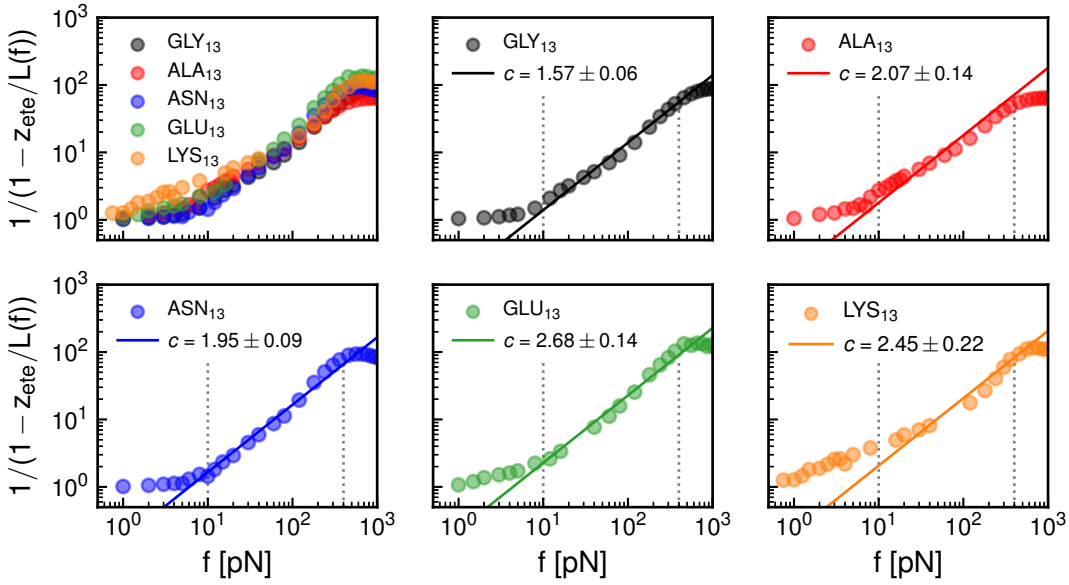

Figure 7: The parameter c, which determines the large-force stretching response, is determined by a least-square fit of the simulation results according to eq 6 (solid lines). For the contour length $L(f)$ the force-dependent expression eq 5 is used. The vertical dashed lines indicate the force range that is used for the fit.

whether the dihedral angle ω is restricted in our simulations, we consider Ramachandran plots of all different combinations of dihedral angles. Two-dimensional probability histograms of dihedral angles from simulations at zero stretching force are shown in Figure 6. The results support the assumption that the dihedral angle characterizing the peptide bond ω is restricted to 180°. Furthermore, we see that the other two dihedral angles, ϕ and ψ, are restricted in their rotation to different degrees, depending on the peptide type. This suggests that the parameter c in eq 4, which was determined from transfer-matrix calculations where two of the three dihedral angles in the peptide backbone were completely free to rotate, must be adjusted.

The relation between the force-dependent contour length $\mathrm{L}(f)$ and the unstretched contour length L_0 follows from eq 1 as

$$\mathrm{L}(f) = \mathrm{L}_0 \left(1 + \frac{\sqrt{\gamma_1^2 + 4\gamma_2 f} - \gamma_1}{2\gamma_2} \right) \quad , \text{for } \frac{\sqrt{\gamma_1^2 + 4\gamma_2 f} - \gamma_1}{2\gamma_2} \geq 0. \tag{5}$$

The large force stretching response from eq 4 follows as

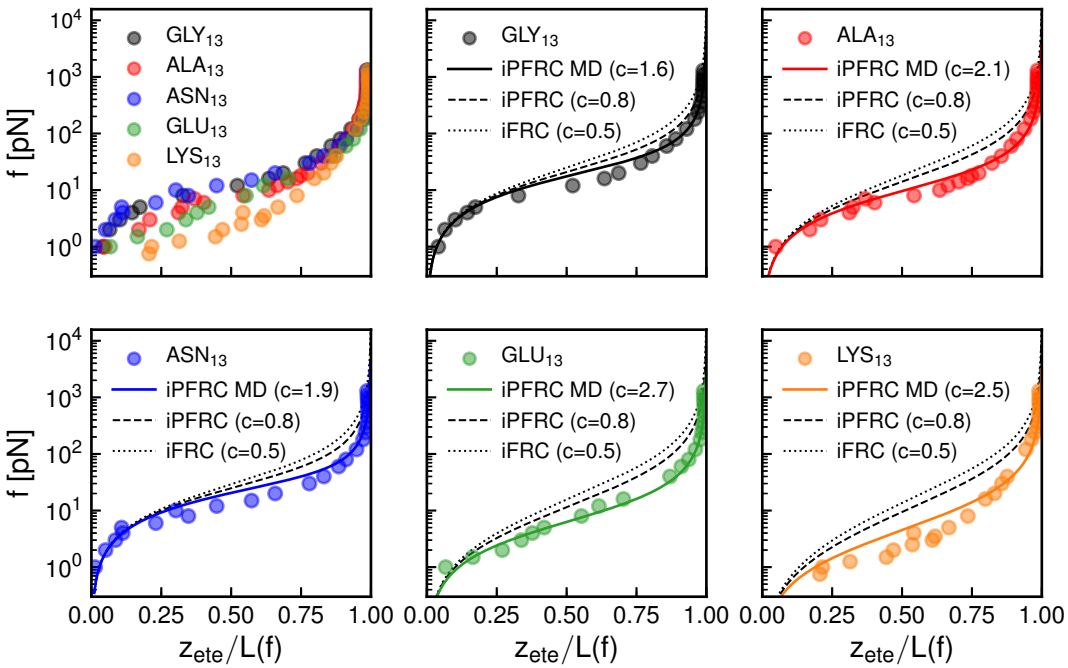

Figure 8: Comparison of the simulated force-stretching data to the stretching model eq 4 (lines). The force-dependent contour length is given by eq 5, for the solid lines the c parameter values are used as deduced from the MD simulations in Figure 7. We also show predictions using the previously determined parameter values for the iFRC and iPFRC models[25] (dotted and dashed lines).

$$\left(1 - \frac{\mathrm{z_{ete}}}{\mathrm{L}(f)}\right)^{-1} = \frac{ca_0}{k_\mathrm{B}T} \times f. \tag{6}$$

To determine the parameter c we plot $1/(1-\mathrm{z_{ete}}/\mathrm{L}(f))$ as a function of the applied stretching force f in Figure 7. We use a least-square fit algorithm based on eq 6 for forces between 10 pN and 400 pN, since for forces above 400 pN solvation effects, finite-temperature effects and higher-order elastic effects presumably modify the dependence of the contour length $L(f)$ on force compared to the elastic response we extracted from our MD simulations of tripeptides at zero temperature. The resulting values for c are given in Figure 7.

Now that we have determined all parameters of the iPFRC model from simulations, which are compiled in Table 1, we can describe the simulated force-stretching data over the entire force range without any adjustable parameters and thereby test the applicability of the heuristic force-extension relation eq 4. In Figure 8 we present the MD simulation

data by plotting the logarithm of the force versus the linear end-to-end distance divided by the force-dependent contour length $L(f)$. In Figure 8 a we compare data for all different

Table 1: Parameters obtained from simulations including the equilibrium monomer length a_0, the Kuhn length a_{Kuhn}, the linear and non-linear stretching moduli γ_1 and γ_2 and the parameter c of the iPFRC model.

	a_{Kuhn}[nm]	a_0[nm]	γ_1[nN]	γ_2[nN]	c
Glycine	0.45 ± 0.02	0.368	32.3 ± 0.1	78.6 ± 0.2	1.57 ± 0.06
Alanine	0.91 ± 0.09	0.359	9.4 ± 0.3	189.9 ± 0.7	2.07 ± 0.14
Asparagine	0.35 ± 0.04	0.351	5.3 ± 0.2	137.8 ± 1.4	1.95 ± 0.09
Glutamic acid	1.49 ± 0.18	0.351	6.2 ± 0.4	137.7 ± 0.9	2.68 ± 0.14
Lysine	2.92 ± 0.15	0.349	4.0 ± 1.0	136.9 ± 1.1	2.45 ± 0.22

polypeptides and in the other subfigures we compare data for individual polypeptides with eq 4. For comparison, we here also show results for previously determined values of c, as predicted from the inhomogeneous partially freely rotating chain model, $c_{\mathrm{iPFRC}} = 0.807$, and the inhomogeneous freely rotating chain model, $c_{\mathrm{iFRC}} = 0.515$.[25] Our results show that with the c values extracted by fits to the MD simulation data, which are generally higher than the previously estimated values c_{iPFRC} and c_{iFRC}, less force is needed in the intermediate regime to stretch the peptides. This reflects the fact that previous models have neglected side-chain interactions, which of course are included in the MD simulations.

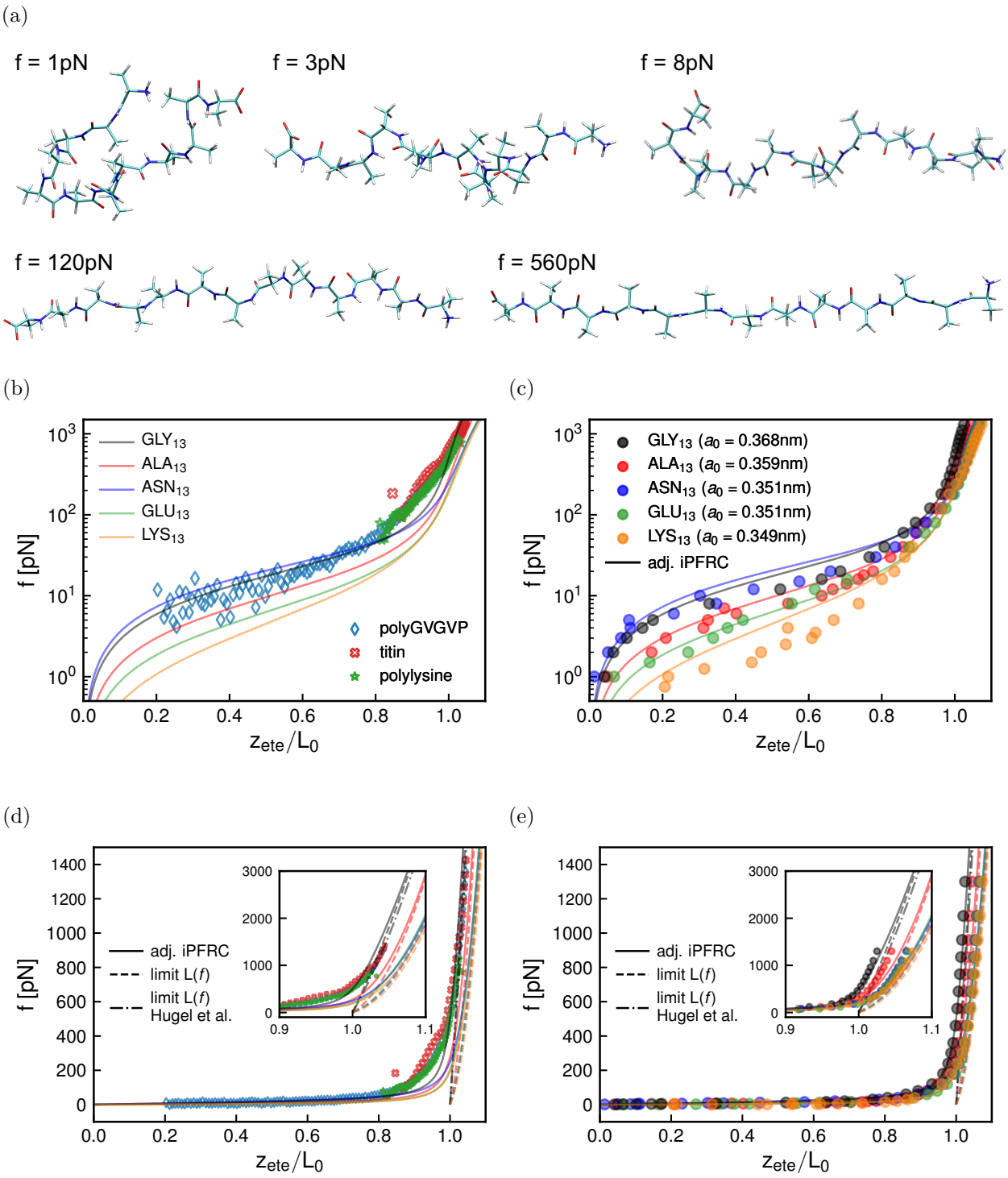

Figure 9: (a) Snapshots of ALA_{13} from MD simulations for several different applied forces. Comparison of the iPFRC model predictions (solid lines) with (b) experimental single polypeptide data from AFM stretching experiments[24] (open data points) and with (c) simulation data (filled data points) for the logarithm of the stretching force versus the end-to-end distance rescaled by the unstretched contour length L_0. (d) and (e): Same data as in (b) and (c) but on a linear force scale. Here in addition the force-dependent contour lengths are shown as determined in this study (dashed lines) and as determined by Hugel et al. for polyglycine[24] (dashed-dotted line).

In Figure 9b and c we compare iPFRC modeling results (solid lines) with previously published experimental AFM results for the stretching of different single polypeptides[24] and with our MD simulation results, where we show the logarithm of the force versus the end-to-end distance z_{ete} rescaled by the unstretched contour length L_0. The experimental contour length L_0 we take from the original paper:[24] There, a fit of the experimental data in the large force regime according to $z_{ete} = L(f) \times (1 - k_B T/(2bf))$ with $b = 0.154\,\text{nm}$ was done, where $L(f)$ is given by eq 5 with the parameters $a_0 = 0.365\,\text{nm}$, $\gamma_1 = 27.4\,\text{nN}$ and $\gamma_2 = 109.8\,\text{nN}$. These parameters were determined previously from DFT calculations of polyglycine.[24] For the rescaling of our MD simulation results we use the parameters for a_0, γ_1, γ_2 and c as derived from MD simulations in the present paper, which for polyglycine are very similar to the parameters obtained earlier.[24] The comparison does not employ additional fit parameters. The agreement between experimental and iPFRC results in Figure 9b is fair over the experimentally probed force range, which primarily validates the MD force-field that was used to obtain the iPFRC parameters. Note that only one of the three experimental data sets, polylysine, corresponds to a homo-polypeptide; polyGVGVP is a repeating pentameric sequence and titin consists of a titin domain repeat, so the comparison with our homopolymeric results can strictly be done only for polylysine. In fact, the three different experimental sequences agree with each other for forces above 30 pN, where data for all three sets is available. Unfortunately, for smaller forces, where our results would predict differences between different sequences to show up, data for only polyGVGVP is available, which is seen to agree quite well with our results for polyglycine and polyasparagine. Interestingly, deviations between the experimental and modeling results appear in the force regime around 200-400 pN. The origin of this discrepancy is unclear, we speculate that it might be related to kinetically frozen dihedral angle flips. Additional experiments that specifically probe aging and memory effects and that use different stretching rates would be highly desirable in order to understand these deviations.

In Figure 9d and Figure 9e we compare experimental and simulation data with iPFRC

results on a linear force scale. Here we in addition also show the force-dependent contour length as given by eq 5. The contour stretching response as determined from our simulation results is shown as broken lines, the contour stretching response for glycine using the parameters determined previously by Hugel et al.[24] is shown as a dashed-dotted line, these two estimates are seen to differ only slightly. This shows that the way we compare experimental data with modeling results is consistent.

In Figure 9a we depict MD simulation snapshots of ALA_{13} for a few different applied forces. Only for the smallest force $f = 1$ pN are monomer-monomer contacts seen, the snapshots thus indicate that for the force range we are interested in, which is substantially larger, interactions between amino acid residues can be neglected.

Finally, we investigate the entropic and energetic contributions to the stretching free energy. The stretching free energy is calculated by integrating the force-extension relation according to $\Delta F(\mathrm{z_{ete}}) = \int_0^{\mathrm{z_{ete}}} f(\mathrm{z'_{ete}})\,\mathrm{dz'_{ete}}$. The change in internal energy ΔU is directly calculated from the simulation trajectories. The change in entropy then follows as $-T\Delta S = \Delta F - \Delta U$. All contributions are divided by the number of monomeric units and thus refer to monomeric quantities. In Figure 10a we show the free energy and its decomposition into internal energy and entropy. We see that the stretching response is mostly entropic in nature and that for alanine and glutamic acid a noticeable attractive energetic contribution is present. All energy curves increase steeply for large force, which in the supporting information is demonstrated to be due to backbone bond stretching and backbone angle deformation. In Figure 10b we show the mean number of hydrogen bonds per residue between peptides and water, $\mathrm{n_{HB}/N}$, as well as the ratio between the mean number of water-bridges and the mean number of hydrogen-bonds between peptides and water, $2\mathrm{n_{WB}/n_{HB}}$. A water bridge is defined as a water molecule that forms two hydrogen bonds with the polypeptide.[26] An example of a water bridge is shown in the inset of Figure 10b. We see that the number of hydrogen bonds between peptide and water does not change with the stretching force, from which we conclude by inversion that interactions with water do only have a marginal impact

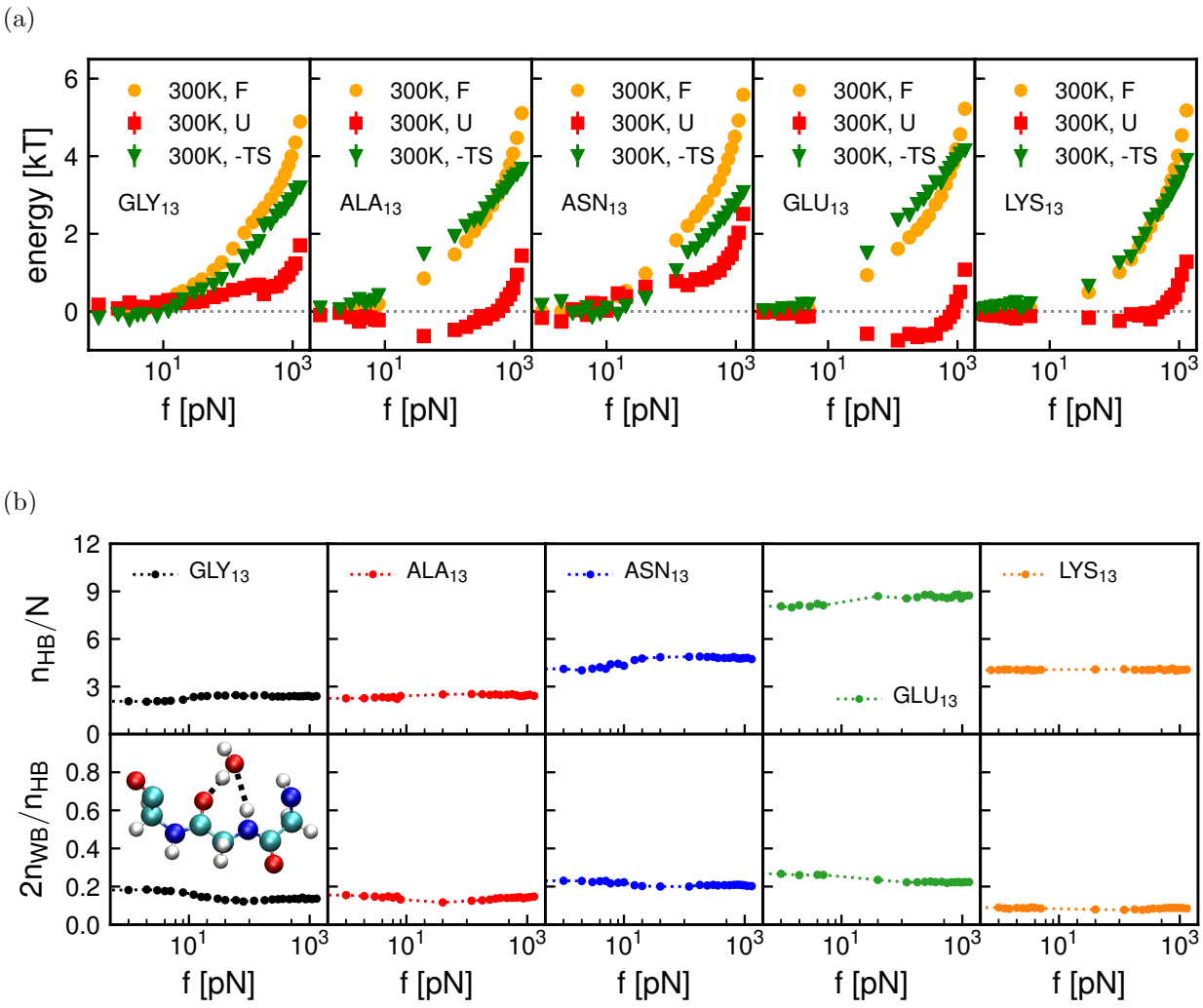

Figure 10: (a) The polypeptide stretching free energy is decomposed into internal energy and entropic contributions. The stretching free energy is calculated by the integral of the force-extension relation. (b) The mean number of hydrogen bonds per monomeric unit between the polypeptides and water, n_{HB}/N, is shown to change only weakly as a function of the stretching force. The force-dependent ratio between the mean number of water-bridges and the mean number of hydrogen bonds between the polypeptide and water, $2n_{WB}/n_{HB}$, depends strongly on peptide type but only weakly on stretching force.

on the stretching response of polypeptides. This is in line with our previous observation that the stretching response of tripeptides is very similar in vacuum and in water.

Conclusions

We present extensive MD simulation results for the force-extension relation of five different homopeptides in explicit water and compare with the iPFRC model force-extension relation, with quantum-chemistry calculations as well as with previously published experimental AFM data. The parameters for the equilibrium monomer length a_0 and the linear and nonlinear contour length stretching moduli γ_0 and γ_1 are determined from tripeptide simulations in vacuum and in water. The Kuhn length a_{Kuhn} is extracted from the linear relation of stretching force and relative extension for 13-mers in the low-force regime. All these parameters depend strongly on the peptide type. The applicability of the inhomogeneous partially freely rotating chain (iPFRC) model is analyzed by detailed analysis of Ramachandran plots and it is shown that side-chain interactions in peptides modify the iPFRC parameter c that plays a prominent role in the high-force regime. It is shown that the iPFRC model describes the simulated force-extension very well over the entire force range without additional freely adjustable parameters. Agreement with published experimental data is good over the force range where experimental data is available, but shows significant deviations in the mid-force regime around 200-400 pN, which suggests further experimental studies in the future. We thus demonstrate that if all relevant parameters such as the the equilibrium monomer length, the Kuhn length, the iPFRC parameter c and the stretching moduli are determined, the global force-stretching relation of peptides can be quantitatively predicted. This will help to interpret polypeptide force-extension relations measured experimentally in different applications and scenarios. The stretching of the contour length is important in the high-force regime. The effect of amino-acid interactions is expected to show up in the low-force regime and can be included by incorporating existing scaling models.[35,36] A decomposition of the stretching free energy into energetic and entropic contributions shows that peptide stretching is mostly entropic, which is different from previous results found for PEG.[26,27] This result is rationalized by an analysis of the number of hydrogen bonds between water and peptide and of the number of water-bridges, which depend very little on the applied

force.

Methods

All MD simulations are performed using the Gromacs simulation package version 5.1.5, some simulations were additionally done with the version 2016.1.[40] The time step is set to 2 fs, the temperature is set to 300 K and the pressure is set to 1 bar with a water compressibility of $4.5 \times 10^{-5}\,\mathrm{bar}^{-1}$. For temperature coupling we use the v-rescale[41] and for pressure coupling the Parrinello-Rahman[42] algorithm with relaxation times of 0.1 ps and 0.5 ps, respectively. All simulations are performed using periodic boundary conditions. The cut-off for non-bonded interactions is set to 1.0 nm. The particle mesh Ewald method[43] is used to calculate the long-range electrostatic interactions. Except for the tripeptide calculations, all homopeptides investigated consist of 13 amino acids. The first and last C_α atoms along the backbone are the anchor points for constant forces that are applied in opposite directions in the z-direction. The polypeptides are simulated in a box with dimensions of 4.5 nm in x- and y- and 12.5 nm in z-direction. Production runs for low forces have a simulation time of up to 830 ns, for high forces the simulations times are at least 200 ns. The mean end-to-end distance z_{ete} in stretching-direction is defined by the time average over the distance in z-direction between the first and the last C_α atom of the backbone of the polypeptide as $z_{ete} = \langle z_{C_\alpha^1 - C_\alpha^{13}} \rangle$. We simulate polyalanine, polyasparagine, polyglutamic acid, polyglycine and polylysine. The caps of the polypeptides are chosen in the zwitterionic form. All polypeptides are simulated using the amber99sb force field.[44] The bond lengths are not constraint. In the initial configuration of the simulations the dihedral angles are all 180°, which for GLY_{13} corresponds to a length of $z_{C_\alpha^1 - C_\alpha^{13}} = 4.38\,\mathrm{nm}$. Initial configurations are created with Avogadro.[45] The exported pdb file is converted to Gromacs input files using pdb2gmx. For the polypeptides with charged side chains, such as polylysine and polyglutamic acid, all side chains are charged. The system is neutralized by inserting the same number of counter

ions Na^+ or Cl^-. Simulations with added salt are shown in the supplementary information. For water, the recommended force field tip3p is used.[46] To equilibrate the system before production runs, an energy minimization followed by a 1 ns NPT simulation is performed. For a given polypeptide, the same equilibrated configuration is used for initializing all production runs at different forces. We analyzed the auto-correlation function of the end-to-end separation to confirm that the relaxation time is much smaller than the simulation time. From the ratio of simulation times and relaxation times the number of uncorrelated configurations is estimated, which is then used for the calculation of the standard error of expectation values. For the hydrogen bond analysis, a donor-acceptor distance cutoff of 3.5 Å and an angle cutoff of 150° is used, which agrees with cutoffs used in the definition of strong hydrogen bonds.[47–49] The number of hydrogen bonds between water and the polypeptide is analyzed every 0.2 ns using the tools `MDAnalysis.analysis.hbonds.hbond_analysis`.[50,51] A water molecule that forms two hydrogen bonds with the polypeptide is denoted as a water bridge.

For the umbrella simulations, tripeptides are put in a $2.5 \times 2.5 \times 2.5\,\mathrm{nm}^3$ box with 503 water molecules. After energy minimization, the box is equilibrated for 1 ns in an NPT simulation. Afterwards, a harmonic potential with a force constant of $200\,000\,\mathrm{kJ/(mol\ nm^2)}$ is applied between the first and the last C_α atoms. The minimum of the harmonic potential is varied between 0.50 nm and 0.80 nm in steps of 0.01 nm. The number of windows is therefore given by 30. The simulations are analyzed using the WHAM algorithm.[52]

For the estimate of the force response of a monomer in vacuum, we perform energy minimizations in double precision, where a single stretched tripeptide is put in a $2.5 \times 2.5 \times 2.5\,\mathrm{nm}^3$ box without any water molecules. The distance $2a$ between the first and the last C_α atoms is constrained during each energy minimization. For each value of $2a$ we perform two indepedent minimizations using the steepest descent method and using the conjugate gradient method and choose the result that gives the lower energy, which helps to avoid the pitfall of being stuck in a local energy minimum. We then change the value of $2a$ in steps of 0.01 nm and use the result from the previous minimization as starting point. By

starting from various different initial configurations, we ensure to obtain the lowest energy structures for all values of $2a$. For steepest descent, the maximum step size is set to 0.01 nm and the maximum tolerance is set to 1 kJ/(mol nm). The same tolerance value is used for the conjugate gradient method. The dielectric constant for the energy minimization is set to the bulk value of tip3p water, which we found to be 106 at 300 K. This value matches reasonable well the reported value of 97 ± 7 by Höchtl et al.,[53] which was simulated with AMBER 4.1[54] in a box with 2.5 nm side length.

In order to support the results from classical simulations and exclude the possibility that differences between glycine and other peptides are force-field specific effects, density functional theory (DFT) calculations are performed. The tripeptide structures are simplified compared to the MD simulations by substituting the capping groups for neutral methyl groups ($CH_3-NH-CO-CR-NH-CO-CH_3$). We limit the computations to two systems, glycine and alanine. Energy minimizations of the structures in vacuum at fixed separations between the C atoms of the caps are performed for several values around $2a = 0.73$ nm, starting from optimized geometries of the corresponding classical simulations. For the DFT calculations, we use the CP2K 4.1 environment with a correlation-consistent polarized triple-zeta basis set augmented with diffuse functions, the BLYP exchange-correlation functional and D3 dispersion corrections.[55–57]

Associated Content

Supporting Information. Stretching response in salt solution, origin of different backbone stretching moduli. This material is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgement

The authors acknowledge the Max-Planck Water Initiative for funding and the North-German Supercomputing Alliance (HLRN) for providing HPC resources that have con-

tributed to the research results reported in this paper.

References

(1) Wang, K.; Forbes, J. G.; Jin, A. J. Single molecule measurements of titin elasticity. *Progress in Biophysics and Molecular Biology* **2001**, *77*, 1–44, DOI: https://doi.org/10.1016/S0079-6107(01)00009-8.

(2) Puchner, E. M.; Gaub, H. E. Force and function: probing proteins with AFM-based force spectroscopy. *Current Opinion in Structural Biology* **2009**, *19*, 605 – 614, DOI: https://doi.org/10.1016/j.sbi.2009.09.005.

(3) Geisler, M.; Xiao, S.; Puchner, E. M.; Gräter, F.; Hugel, T. Controlling the Structure of Proteins at Surfaces. *Journal of the American Chemical Society* **2010**, *132*, 17277–17281, DOI: 10.1021/ja107212z.

(4) Kienle, S.; Liese, S.; Schwierz, N.; Netz, R. R.; Hugel, T. The Effect of Temperature on Single-Polypeptide Adsorption. *ChemPhysChem* **2012**, *13*, 982–989, DOI: 10.1002/cphc.201100776.

(5) Wei, M.; Gao, Y.; Li, X.; Serpe, M. J. Stimuli-responsive polymers and their applications. *Polym. Chem.* **2017**, *8*, 127–143, DOI: 10.1039/C6PY01585A.

(6) Kuroyanagi, S.; Shimada, N.; Fujii, S.; Furuta, T.; Harada, A.; Sakurai, K.; Maruyama, A. Highly Ordered Polypeptide with UCST Phase Separation Behavior. *Journal of the American Chemical Society* **2019**, *141*, 1261–1268, DOI: 10.1021/jacs.8b10168.

(7) Meyer, G.; Amer, N. M. Erratum: Novel optical approach to atomic force microscopy. *Applied Physics Letters* **1988**, *53*, 2400–2402, DOI: 10.1063/1.100425.

(8) Allen, M. J.; Bradbury, E. M.; Balhorn, R. AFM analysis of DNA-protamine complexes bound to mica. *Nucleic Acids Research* **1997**, *25*, 2221–2226, DOI: 10.1093/nar/25.11.2221.

(9) Rief, M. Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM. *Science* **1997**, *276*, 1109–1112, DOI: 10.1126/science.276.5315.1109.

(10) Hugel, T.; Seitz, M. The Study of Molecular Interactions by AFM Force Spectroscopy. *Macromolecular Rapid Communications* **2001**, *22*, 989–1016, DOI: 10.1002/1521-3927(20010901)22:13<989::AID-MARC989>3.0.CO;2-D.

(11) Lyubchenko, Y. L.; Shlyakhtenko, L. S. AFM for analysis of structure and dynamics of DNA and protein-DNA complexes. *Methods* **2009**, *47*, 206 – 213, DOI: https://doi.org/10.1016/j.ymeth.2008.09.002.

(12) Camunas-Soler, J.; Ribezzi-Crivellari, M.; Ritort, F. Elastic Properties of Nucleic Acids by Single-Molecule Force Spectroscopy. *Annual Review of Biophysics* **2016**, *45*, 65–84, DOI: 10.1146/annurev-biophys-062215-011158.

(13) Grebikova, L.; Radiom, M.; Maroni, P.; Schlüter, A. D.; Borkovec, M. Recording stretching response of single polymer chains adsorbed on solid substrates. *Polymer* **2016**, *102*, 350–362, DOI: https://doi.org/10.1016/j.polymer.2016.02.045.

(14) Fixman, M.; Kovac, J. Dynamics of stiff polymer chains. I. *The Journal of Chemical Physics* **1974**, *61*, 4939–4949.

(15) Gennes, P. D. Scaling theory of polymer adsorption. *Journal de Physique* **1976**, *37*, 1445–1452, DOI: 10.1051/jphys:0197600370120144500.

(16) Manca, F.; Giordano, S.; Palla, P. L.; Zucca, R.; Cleri, F.; Colombo, L. Elasticity of flexible and semiflexible polymers with extensible bonds in the Gibbs and Helmholtz ensembles. *The Journal of Chemical Physics* **2012**, *136*, 154906, DOI: 10.1063/1.4704607.

(17) Marko, J. F.; Siggia, E. D. Stretching DNA. *Macromolecules* **1995**, *28*, 8759–8770, DOI: 10.1021/ma00130a008.

(18) Odijk, T. Stiff Chains and Filaments under Tension. *Macromolecules* **1995**, *28*, 7016–7018, DOI: 10.1021/ma00124a044.

(19) Bouchiat, C.; Wang, M. D.; Allemand, J.-F.; Strick, T.; Block, S.; Croquette, V. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophysical journal* **1999**, *76*, 409–413.

(20) Tskhovrebova, L.; Trinick, J.; Sleep, J. A.; Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **1997**, *387*, 308–312, DOI: 10.1038/387308a0.

(21) Gräter, F.; Shen, J.; Jiang, H.; Gautel, M.; Grubmüller, H. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophysical journal* **2005**, *88*, 790–804, DOI: 10.1529/biophysj.104.052423.

(22) Grützner, A.; Garcia-Manyes, S.; Kötter, S.; Badilla, C. L.; Fernandez, J. M.; Linke, W. A. Modulation of titin-based stiffness by disulfide bonding in the cardiac titin N2-B unique sequence. *Biophysical journal* **2009**, *97*, 825–834, DOI: 10.1016/j.bpj.2009.05.037.

(23) Livadaru, L.; Netz, R. R.; Kreuzer, H. J. Stretching Response of Discrete Semiflexible Polymers. *Macromolecules* **2003**, *36*, 3732–3744, DOI: 10.1021/ma020751g.

(24) Hugel, T.; Rief, M.; Seitz, M.; Gaub, H. E.; Netz, R. R. Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-Initio Theory. *Physical Review Letters* **2005**, *94*, 048301, DOI: 10.1103/PhysRevLett.94.048301.

(25) Hanke, F.; Serr, A.; Kreuzer, H. J.; Netz, R. R. Stretching single polypeptides: The

effect of rotational constraints in the backbone. *EPL (Europhysics Letters)* **2010**, *92*, 53001, DOI: 10.1209/0295-5075/92/53001.

(26) Liese, S.; Gensler, M.; Krysiak, S.; Schwarzl, R.; Achazi, A.; Paulus, B.; Hugel, T.; Rabe, J. P.; Netz, R. R. Hydration Effects Turn a Highly Stretched Polymer from an Entropic into an Energetic Spring. *ACS Nano* **2017**, *11*, 702–712, DOI: 10.1021/acsnano.6b07071.

(27) Kolberg, A.; Wenzel, C.; Hackenstrass, K.; Schwarzl, R.; Rüttiger, C.; Hugel, T.; Gallei, M.; Netz, R. R.; Balzer, B. N. Opposing Temperature Dependence of the Stretching Response of Single PEG and PNiPAM Polymers. *Journal of the American Chemical Society* **2019**, *141*, 11603–11613, DOI: 10.1021/jacs.9b04383.

(28) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725, DOI: 10.1002/prot.21123.

(29) Ho, B. K.; Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMC structural biology* **2005**, *5*, 14, DOI: 10.1186/1472-6807-5-14.

(30) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2016**, *14*, 71–73, DOI: 10.1038/nmeth.4067.

(31) Balaji, G. A.; Nagendra, H. G.; Balaji, V. N.; Rao, S. N. Experimental conformational energy maps of proteins and peptides. *Proteins: Structure, Function and Bioinformatics* **2017**, *85*, 979–1001, DOI: 10.1002/prot.25266.

(32) de Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press, 1979.

(33) Netz, R. R.; Andelman, D. Neutral and charged polymers at interfaces. *Physics Reports* **2003**, *380*, 1–95, DOI: 10.1016/S0370-1573(03)00118-2.

(34) Dittmore, A.; McIntosh, D. B.; Halliday, S.; Saleh, O. A. Single-Molecule Elasticity Measurements of the Onset of Excluded Volume in Poly(Ethylene Glycol). *Phys. Rev. Lett.* **2011**, *107*, 148301, DOI: 10.1103/PhysRevLett.107.148301.

(35) Pincus, P. Excluded Volume Effects and Stretched Polymer Chains. *Macromolecules* **1976**, *9*, 386–388, DOI: 10.1021/ma60051a002.

(36) Saleh, O. A.; McIntosh, D. B.; Pincus, P.; Ribeck, N. Nonlinear Low-Force Elasticity of Single-Stranded DNA Molecules. *Phys. Rev. Lett.* **2009**, *102*, 068301, DOI: 10.1103/PhysRevLett.102.068301.

(37) Morrison, G.; Hyeon, C.; Toan, N. M.; Ha, B.-Y.; Thirumalai, D. Stretching Homopolymers. *Macromolecules* **2007**, *40*, 7343–7353, DOI: 10.1021/ma071117b.

(38) Saleh, O. A. Perspective: Single polymer mechanics across the force regimes. *The Journal of Chemical Physics* **2015**, *142*, 194902, DOI: 10.1063/1.4921348.

(39) Berezney, J. P.; Saleh, O. A. Electrostatic Effects on the Conformation and Elasticity of Hyaluronic Acid, a Moderately Flexible Polyelectrolyte. *Macromolecules* **2017**, *50*, 1085–1089, DOI: 10.1021/acs.macromol.6b02166.

(40) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718, DOI: 10.1002/jcc.20291.

(41) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*, 14101, DOI: 10.1063/1.2408420.

(42) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190, DOI: 10.1063/1.328693.

(43) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593, DOI: 10.1063/1.470117.

(44) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725, DOI: 10.1002/prot.21123.

(45) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **2012**, *4*, 17, DOI: 10.1186/1758-2946-4-17.

(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935, DOI: 10.1063/1.445869.

(47) Desiraju, G.; Steiner, T. *International Union of Crystallography Monographs on Crystallography*; Oxford University Press: Oxford, 2001; p 480, DOI: 10.1093/acprof:oso/9780198509707.001.0001.

(48) van der Spoel, D.; van Maaren, P. J.; Larsson, P.; Tîmneanu, N. Thermodynamics of Hydrogen Bonding in Hydrophilic and Hydrophobic Media. *The Journal of Physical Chemistry B* **2006**, *110*, 4393–4398, DOI: 10.1021/jp0572535.

(49) Wood, P. A.; Allen, F. H.; Pidcock, E. Hydrogen-bond directionality at the donor H atom-analysis of interaction energies and database statistics. *CrystEngComm* **2009**, *11*, 1563–1571, DOI: 10.1039/B902330E.

(50) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327, DOI: 10.1002/jcc.21787.

(51) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016; pp 98–105, DOI: 10.25080/Majora-629e541a-00e.

(52) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021, DOI: 10.1002/jcc.540130812.

(53) Höchtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. Rationalization of the dielectric properties of common three-site water models in terms of their force field parameters. *Journal of Chemical Physics* **1998**, *109*, 4927–4937, DOI: 10.1063/1.477104.

(54) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, *91*, 1–41, DOI: https://doi.org/10.1016/0010-4655(95)00041-D.

(55) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; Vandevondele, J. Cp2k: Atomistic simulations of condensed matter systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 15–25, DOI: 10.1002/wcms.1159.

(56) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms

revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992**, *96*, 6796–6806, DOI: 10.1063/1.462569.

(57) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010**, *132*, DOI: 10.1063/1.3382344.