SCIENTIFIC REPORTS
nature research

OPEN

Delineating elastic properties of kinesin linker and their sensitivity to point mutations

Michał Świątek[1,3*] & Ewa Gudowska-Nowak[2]

We analyze free energy estimators from simulation trials mimicking single-molecule pulling experiments on a *neck linker* of a kinesin motor. For that purpose, we have performed a version of steered molecular dynamics (SMD) calculations. The sample trajectories have been analyzed to derive distribution of work done on the system. In order to induce stretching of the linker, we have applied a constant pulling force to the molecule and allowed for a subsequent relaxation of its structure. The use of fluctuation relations (FR) relevant to non-equilibrium systems subject to thermal fluctuations allows us to assess the difference in free energy between stretched and relaxed conformations. To further understand effects of potential mutations on elastic properties of the linker, we have performed similar *in silico* studies on a structure formed of a polyalanine sequence (Ala-only) and on three other structures, created by substituting selected types of amino acid residues in the linker's sequence with alanine (Ala) ones. The results of SMD simulations indicate a crucial role played by the Asparagine (Asn) and Lysine (Lys) residues in controlling stretching and relaxation properties of the linker domain of the motor.

The motor proteins' ability to generate movement creates a situation, where it is possible to treat different parts of one molecule as linked but functionally separate objects, that possess an ability to move at different times and speeds[1]. That makes the field of motor proteins a desirable testing ground for application of various theoretical models, that aim to filter the inherent complexity of biological systems[2–5]. Among the motors, members of the kinesin protein superfamily are consistently used throughout the years in research focusing on molecular motors' mechanical properties, both *in vitro*[2] and *in silico*[2,4–6].

A conformational change of a region within kinesin has been shown to be associated with its movement along microtubule[7]. This region has been labeled as *neck linker*. The label refers to a concise (less than 20 amino acids length) amino acid sequence in a single kinesin head that acts as a bridge between α-6 helix in coiled-coil dimerization domain and α-7 helix in the core motor domain, respectively[8]. While its exact length as well as the placement within sequence of its N- and C- termini are not set in stone[9], some of its residue patterns are present across all kinesin families, while *neck linker* sequences within a single family are very similar[8]. All this suggest that parts of *neck linker* region have been conserved, which makes *neck linker* a non-random, specific sequence. An on-going accumulation of experimental data evidence suggests that a transition of the *neck linker* from a disordered (random coil) state to an ordered (β-sheet) conformation is a key factor in determining a mechanism of force-generation that is a crucial element of molecular motors' ability to move along microtubules[7,8]. Attempts at substituting *neck linker* with a peptide of a different sequence, as well as extending it, has been made and resulted in impairment of kinesin functionality[10,11]. Strain through the *neck linker* ensures processive runs of the motor[1] and can be estimated by analyzing elastic properties of border regions between heads of the kinesin molecule. Molecular dynamics methods are a widespread tool usually employed to uncover possible system behaviour on atomic level. It has been used to describe kinesin domains' response to ATP binding and hydrolysis[12,13] and for comparison of its dynamics in different ATPase states[13,14]. Assessment of forces necessary to maintain *neck linker* at different lengths has also been made[8]. In this study, we ignore the complex interactions between *neck linker* and

[1]Department of Pharmaceutical Biophysics, Jagiellonian University Medical College, ul. Medyczna 9, 30-688, Kraków, Poland. [2]Jagiellonian University, Marian Smoluchowski Institute of Physics and Mark Kac Center for Complex Systems Research, ul. Prof. S.Łojasiewicza 11, Kraków, 30-348, Poland. [3]Jagiellonian University, Marian Smoluchowski Institute of Physics, ul. Prof. S.Łojasiewicza 11, Kraków, 30-348, Poland. *email: mgj.swiatek@uj.edu.pl

other elements of kinesin motor, focusing instead solely on *neck linker's* sequence, its mechanical properties and the effects of point mutations.

Biopolymer chains are often interpreted in terms of numerous approximations, all having roots in theoretical assumptions that form a basis of the Freely Jointed Chain (FJC) model. Among these, the worm-like-chain (WLC) model[15] seems to be the most relevant when it comes to describing a bending process of a semi-rigid biostructure, even though it has received some criticism[16,17]. Stretching of a peptide requires an application of a certain force and the relation between that force F and the stable extension x of the chain can be formulated as

$$F = \frac{k_B T}{L}\left[\frac{1}{4}\left(1 - \frac{x}{L_c}\right)^{-2} + \frac{x}{L} - \frac{1}{4}\right] \tag{1}$$

where k_B is Boltzmann constant, T stands for temperature, L_c is the contour length of the polymer and L its persistence length. In a previous work, we have already presented a preliminary venture into the matter at hand, showcasing a difference in stretching process between a *neck linker* and an Ala-only polypeptide[18]. Here, employing the methods of Molecular Dynamics and Normal Mode Analysis (NMA), we intend to deliver a more comprehensive description of a possible relation between *neck linker's* amino acid sequence and the specificity of its function. The paper is organised as follows: After a brief *Introduction*, the section *Material and Methods* presents basic methodology of the domain analysis and describes the setup of Molecular Dynamics (MD) simulations. A number of theoretical considerations are discussed, pertaining to thermodynamic description of an amino acid chain, ability to determine its elasticity via the force-extension relations and significance of non-equilibrium dynamics as used in our simulations. All these are placed in subsections of their own. Next, in *Results and Analysis*, data collected in series of simulations are presented and examined. The last section, *Conclusions*, contains our closing remarks in which we summarise findings and highlight points of interest for future research in this field.

Methods

Domain analysis of kinesin heads. In the initial part of our studies we have identified dynamic domains in the structure of *Kinesin Heavy Chain* (taken from PDB Databank(id:3kin)[19]) and analyzed deformations (low-frequency domain motions) which have been obtained with a simplified mechanical model proposed by Hinsen[20]. The method is essentially dependent on a fact that the low frequency normal modes describing motion of domains in proteins are influenced by anharmonic effects and in realistic environments become strongly overdamped, thus independent of the applied force field details. The positioning of heads in 3kin structure is such that they exhibit a rotational symmetry of 120° round an axis closely aligned with coiled coil's axis[19]. In that conformation, both heads couldn't attach themselves to microtubule lattice at the same time[19]. Considering the fact that the 3kin structure is not accompanied by a bound tail domain that limits free movement[21] and the fact that kinesin is known to rotate round stalk axis[22,23], we have no reason to doubt its validity.

The structure has been analyzed with *DomainFinder* application[24]. Firstly, an approximate NMA has been performed. Secondly, several choices had to be made regarding parameters of the domain analysis. More normal modes included in the domain analysis can result in a more detailed picture. However, a gradual addition of modes of ever rising average deformation energy inevitably leads to the ones responsible for rigid body motions having a decreased influence. It is also necessary to consider the number of modes selected for further analysis in conjunction with a choosing of a deformation energy threshold, meant to filter out regions of insufficient rigidity. Hinsen proposes values between 400 and 500 $kJ * mol^{-1}$ as a starting point[24]. If selected normal modes are to be meaningfully included in analysis, their average deformation energy should not exceed the threshold. The average deformation energy of the 16th mode equals 393.03 $kJ * mol^{-1}$. The average deformation energy of the 17th mode equals 416.98 $kJ * mol^{-1}$. The average energies of previous modes are lower while the average energies of further modes are higher. Taking a look at deformation energy distributions for particular normal mode ensembles (Fig. 1), it can be seen that as the number of modes grows, the fraction of low energy regions decreases. In case of 17 (and even more so 18) modes, the distribution starts to flatten. The percentage of regions, whose deformation energy is below 400 $kJ * mol^{-1}$ goes under 90% for 17 modes. All in all, the aforementioned observations have been deemed a good enough indicator to decide on 16 modes being stored for domain and deformation analysis with the energy threshold being set at 400 $kJ * mol^{-1}$. This choice has led to the acquisition of images presented in Fig. 2, in which domains are associated with internally stable regions of protein and off-domain regions are relatively fluid.

Here, in the upper panel, blue color represents regions of deformation energy well below threshold, while light blue and light red parts symbolize regions slightly below and slightly above the threshold, respectively. A crucial parameter differentiating between rigid regions with uniform motions and intermediate regions whose internal deformation yields systematic contributions to the overall motion between boundaries of the domain is the domain coarseness factor c. In brief, the coarseness parameter c establishes the threshold of similarity between residues' rigid body motions that must be exceeded if residues in question are to be considered a part of the same dynamical domain. As the coarseness parameter c decreases, the domain forming rule becomes more restrictive[25]. The coarseness parameter must be greater than 1 (see eq. (7)). We have settled on the default value of 10. The resulting dynamical domain decomposition can be seen in the lower panel in Fig. 2.

A deformation energy definition used by *DomainFinder* relates to the interaction energy between two particles i and j of the elastic network and reads

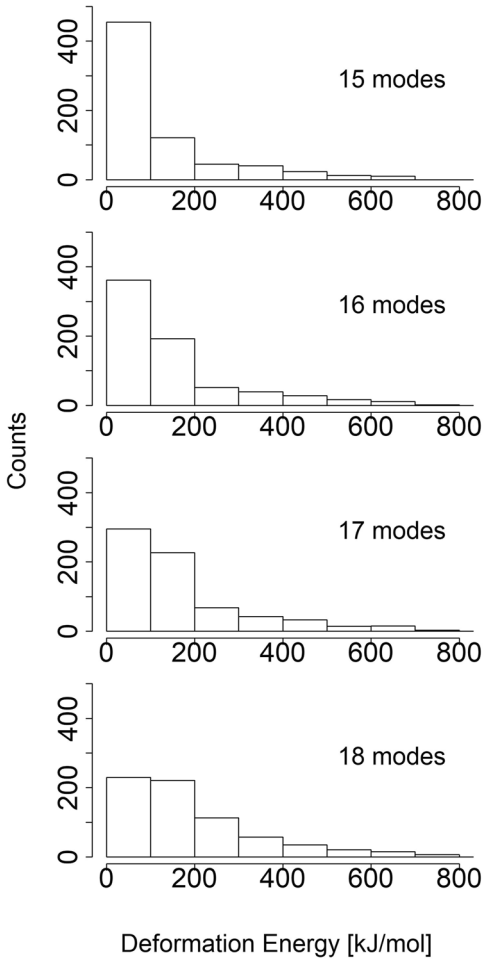

Figure 1. A comparison of deformation energy distributions in the kinesin chain. The deformation energy values of each distribution have been obtained by utilising different number of low-frequency normal modes. The deformation energy threshold that differentiates between rigid and flexible region has been set at 400 $kJ * mol^{-1}$.

$$E_i = \frac{1}{2}\sum_{j=1}^{N} k\left(R_{ij}^{(0)}\right)\frac{\left|\left(d_i - d_j\right)R_{ij}^{(0)}\right|^2}{\left|R_{ij}^{(0)}\right|^2} \tag{2}$$

where i, j denote C_α atoms, $d_{i,j}$ are corresponding infinitesimal displacements from original positions (displacement of the atom in the mode to be analyzed), R_{ij}^0 are distances between pairs of atoms i, j in a submitted structure and $k\left(R_{ij}^0\right)$ is an effective harmonic force constant, that attenuates with a spatial distance according to the relation:

$$k\left(R_{ij}^{(0)}\right) = C\exp\left(-\frac{\left|R_{ij}^{(0)}\right|^2}{r_0^2}\right) \tag{3}$$

in order to maintain a force field of short range (r_0), that excludes interactions between potential domains. Parameter C has been chosen arbitrarily, as 47.400 kJ $\times$ mol^{-1} nm^{-2} at temperature 300K, to ensure compatibility with the *Amber 94* forcefield. Amplitudes of $d_{i,j}$ are defined with an equation

$$\sum_{i=1}^{N} |d_i|^2 = fN, \tag{4}$$

f being a scaling factor of value 1 nm^2. Displacements d_i have been used to define rotation (Ω) and translation (T) vectors:

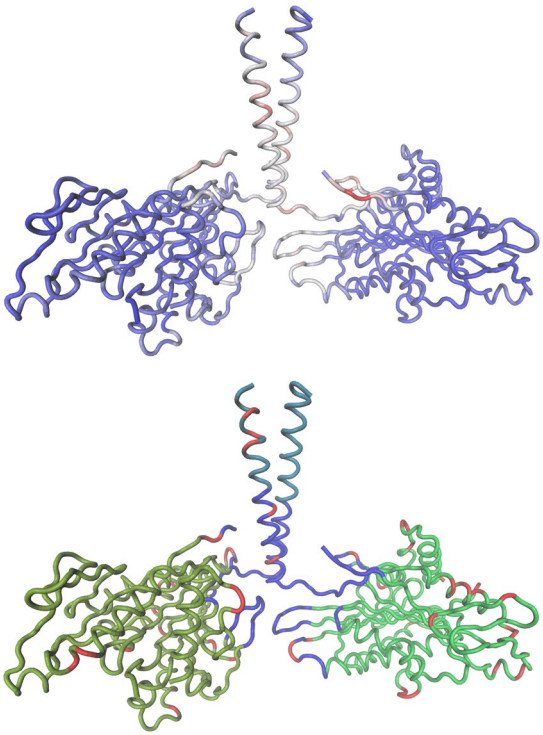

Figure 2. Distribution of deformation energies in the kinesin structure (upper panel) and derived domains (lower panel) based on 16 low-frequency modes. The deformation energy threshold has been set at 400 $kJ * mol^{-1}$, domain coarseness c at 10. Upper panel: as color scale goes from blue to red, deformation energy values go from low to high. Lower panel: each cluster of vibrational energies can be identified by its color. The domain incorporating *neck linker* region is blue, the heads are olive and green while the stalk fragment is cyan. The images has been created using *VMD 1.9.3*[63] and *Tachyon*[64].

$$d_i = T + \Omega \times R_i \tag{5}$$

where R_i stands for the i atom position. When displacement vectors do not describe pure rigid-body motion, linear least-squares fit is used to determine values of T and Ω. The structure is further divided into cubic compartments of side length 1.2 nm, all ignored unless containing at least 3 atoms and having average deformation energy below pre-defined threshold. Those cubes have their rotation and translation vectors calculated. A following definition of similarity is used to identify clusters of cubes having similar mobility.

$$\mathscr{S}_{ij} = 3\frac{\left|\Omega_i + \Omega_j\right|}{\left|\Omega_i - \Omega_j\right|} + \frac{\left|T_i + T_j\right|}{\left|T_i - T_j\right|} \tag{6}$$

The rotation vectors are empirically more precise in sorting out domains, thus are given greater weight. A cluster is finally created, by using the criterion

$$\mathscr{S}_{ik} > \frac{\mathscr{S}_{ij}^{max}}{c} \tag{7}$$

where c is pre-selected domain coarseness parameter. All cubes contributing to this relation, are then considered to compose one cluster.

Structure and relaxation of extended linker: mechanical and thermodynamic considerations. By definition, the partition function for simulations run at a constant volume condition is given by

$$Z = \mathscr{N} \int d\mathbf{r} e^{-\beta E(\mathbf{r})} \tag{8}$$

where $\beta = (k_B T)^{-1}$. Here T is absolute temperature and E stands for the energy of a given configuration state, with an average energy of the system given by

$$\langle E \rangle = \int d\mathbf{r} E(\mathbf{r}) \rho(\mathbf{r}), \tag{9}$$

in which $\rho(\mathbf{r})$ is the equilibrium probability density $\rho(\mathbf{r}) = e^{-\beta E(\mathbf{r})} \times \left[\int d\mathbf{r} e^{-\beta E(\mathbf{r})}\right]^{-1}$. Accordingly, the Gibbs free energy of the system is given by

$$G = \langle E \rangle + pV - T\left(k_B \log \mathcal{N} - k_B \int d\mathbf{r} \rho(\mathbf{r}) \log \rho(\mathbf{r})\right). \tag{10}$$

If the system energy is partitioned over many local minima (energy wells), the configurational integral Eq. (8) can be represented[26,27] in the form of a sum $Z = \sum_i Z_i$ with $Z_i = \mathcal{N} \int_i d\mathbf{r} e^{-\beta E(\mathbf{r})}$ and integral evaluated over the $i-th$ energy well, in which the probability density $\rho_i(\mathbf{r})$ can be expressed as

$$\rho_i(\mathbf{r}) = \frac{\rho(\mathbf{r})}{p_i}, \quad \mathbf{r} \in \Omega_i \tag{11}$$

with $p_i = \frac{Z_i}{Z}$[28]. The average energy of the system can be then rephrased as $\langle E \rangle = \sum_i p_i \langle E \rangle_i$. All in all, the weighted average Eq. (11), or the differences ΔG, pertinent to two (initial/final) states can be then accessed in a straightforward way by a histogram method counting the number of times the molecule "visited" given configurational state in course of MD simulations[26].

MD simulations' setup. A structure of a motor domain, belonging to a kinesin-like protein *KIF3B* (a Kinesin-2 family's member)[29], has been obtained from the PDB Databank(id:3b6u)[30]. A sequence of 19 amino acids, 17 of which are considered to be a functional part known as a *neck linker*, has then been extracted and optimized. In order to do so, *Steepest Descent* and *Conjugate Gradients* algorithms have been employed. The *neck linker* chain was subsequently placed in a box of water molecules (a Simple Point Charge model), and the whole system was optimized again. Next, a simulation of Molecular Dynamics (MD) was scheduled, with a goal of achieving a state of at least near equilibrium, producing a $6ns$ long trajectory (with a time step duration $\Delta t = 2 \times 10^{-3} ps$). A distribution of *end-to-end* distances was then created and used to determine the mean *end-to-end* distance of the equilibrated linker chain. Finally, a structure has been selected with an *end-to-end* distance sufficiently close to the mean, while belonging to a time frame from near the end of the simulation.

That selected *neck linker* structure has been taken out of the box of water molecules and placed in the implicit solvent. A short simulation, with positions of first and last C_α atoms fixed, served as a short equilibration routine. After that, the system was employed as a starting point of 10^4 simulation runs, where a constant force of 1300 $kJ * mol^{-1} * nm^{-1}$ (approximately 2160 pN) was applied between a mass centre of the 1st residue and a mass centre of the 19th residue. Results of several optical trap experiments[11,31,32] give a force value necessary to move kinesin motor as below 10 pN, while simulations centering on *neck linker* arrived at force values reaching and surpassing 200 pN[8,16]. We however desired to test our structure against a stronger external influence that causes a rapid response from our system and, with that in mind, the aforementioned force value has been chosen, after some preliminary MD runs. That kind of set up allows for examination of *neck linker's* spring-like behaviour. MD simulations that test mechanical properties of peptides have shown that upper limit of forces far exceeds 1300 $kJ * mol^{-1} * nm^{-1}$, which is the maximum force used by us[33,34]. In fact, the study researching deca-alanine peptide, a very similar one to a polyalanine sequence that we employ as a reference (details below), gives the value of force at stretching limit as around 3600 pN[33]. The simulations of non-equilibrium dynamics produced a set of 1 ps trajectories. The final states of these trajectories became starting points of another 10^4 simulations, each lasting 1 ps ($\Delta t = 2 \times 10^{-3}\ ps$), where the constant stretching force had been turned off, resulting in system relaxation. In all simulations, the *Berendsen* thermostat has been used to ensure stable temperature conditions (T = 300 K), while the levels of pressure have been controlled with *Parrinello-Rahman* barostat (p = 100 kPa). Since our interest has been focused on investigation of linker's specific elasticity, analogous steps have been taken with regards to an 18 amino acid long *Ala-only* peptide, with the equilibration simulation being 3 ns long. The alanine residue is the simplest possible, not possessing any side chain and, because of that, a polyalanine sequence has been deemed the best model for observing a protein polymer behaviour limited only to the protein backbone. Additionally, the above steps of modelling have been repeated for three "intermediate" structures between the original *neck linker* and the *Ala-only* polymer. Namely, by selecting some of amino acid residues (either Asparagine, or Proline, or Lysine) and substituting them with alanine residues, "mutant" versions of the neck-linker have been created. The equilibrium simulations of those three altered linker structures lasted $5ns$, $6ns$ and $3ns$ for *no-Asparagine*, *no-Lysine* and *no-Proline* chains, respectively. Finally, the whole process involving all 5 different sequences has been repeated for different stretching force values (130, 400, 700 and 1000 $kJ * mol^{-1} * nm^{-1}$). The 4.5.5 version[35] of the GROMACS package[36,37] and the *OPLS-aa* force field[38,39] have been employed to perform all MD simulations.

Steered molecular dynamics of kinesin linker structure. Molecular interactions and mechanical properties of individual molecules can be nowadays probed by use of combined techniques, like Atomic Force Microscopy (AFM) and optical tweezers[40]. In such experiments single molecules are held and stretched, and from the measurements of a cantilever spring restoring force in the AFM instrument, the information about elasticity (effective spring constants) and intensity of rupture forces can be derived[41,42]. Analogous to these experimental setups, steered molecular dynamics (SMD) simulations permit similar investigations to be performed *in silico*. In brief, the procedure of SMD applies external steering forces in molecular dynamics simulations to investigate processes of e.g. protein unfolding or binding/unbinding of substrates separated by some energy barriers. Practical designs of such simulations are based on relating free energy difference in nonequilibrium steady states achieved in course of manipulation with the work done through the process. The thermostated system is hold at the beginning of the action at equilibrium of a given temperature T. By changing an externally controlled parameter λ (a

particular "transformation coordinate"), the work W done on the system may be estimated from the external energy required to change the reaction coordinate from an initial value λ_0 to a final λ_f. The process is repeated many times so that the statistics of work performed is collected with free energy difference between the steady states related by Jarzynski equality[42,43]

$$\langle e^{-\beta W}\rangle = \int dW p(W) e^{-\beta W} = e^{-\beta \Delta G},\ \beta = (k_B T)^{-1} \tag{12}$$

with average taken over repeated realizations of the process and $p(W)$ being the relevant probability density function (PDF) for work distribution. The above work fluctuations equality holds under very general condition: originally proven for free energy differences between equilibrium states, it can be also derived for stochastic Markovian dynamics for prescribed protocols provided by changes of $\lambda(t)$ between stationary (nonequilibrium) states[44,45]. In the experiments with pulling force, the center of mass of the pulled molecule is attached to a spring with an elasticity constant k, so that the pulling force is $F = k(\lambda(t) - x)$ with the control parameter $\lambda(t) = x_0 + vt$. The thermodynamic work definition for the stretch dynamics becomes then

$$W = \int_{\lambda_0}^{\lambda_f} dt \dot{\lambda} \frac{\partial E(x;\ \lambda)}{\partial \lambda} \equiv -v \int_0^t dt' F(t') \tag{13}$$

where t stands for the time after which the new (stationary) state is reached and E is a sum of energies of the mean force and the external potential of the stretching force[42].

The implications of non-equilibrium dynamics described above have been taken into consideration as we set out to gauge the free energy difference between stretched and relaxed amino acid chains.

Results and Discussion

Judging from Fig. 2, the two kinesin heads are more rigid than the region that includes *neck linker* and the beginning of the two intertwined helices (coiled coil). Notably, rigidity rises again at the end of the helices, which is expected, since as the helices form the stalk, their intertwined structure should be quite stable. The distribution of deformation energy is not symmetrical in two kinesin heavy chains. According to the PDB file[19], in one of the chains the 371st and the 372nd residues are missing. These are residues that conclude the chain and their absence must cause a discrepancy between chains at the C-terminus. Additionally, almost all corresponding helices in the chains are slightly shifted in relation to each other. In one particular instance, a helix from one chain (Lys283 - Gly293) is split in two in the other chain (Lys283 - Ile287, Gln289 - Gly293). Sheets are affected to lesser degree. These observations carry on to the domain analysis. The domain of the highest similarity is the olive kinesin head, next is the green head, then the cyan fragment and finally the blue region that incorporates *neck linker*. As we rise the c parameter (up to the value of 54, where all of the kinesin is covered by one, all-encompassing domain), the green domain gradually gets incorporated into the olive one, while the blue *neck linker* domain gets incorporated into the cyan stalk domain. To sum up, a decomposition of the kinesin structure into two heads of relatively greater similarity with different, softer motions observed in the region of the *neck linker*, seems prevalent even as parameters change. *Neck linker*, the domain formed of 14–18 amino acids, is widely considered a key structure in the underlying kinesin's force-generating mechanism and has been examined in a series of experimental[46] and theoretical[8] studies. In particular, it has been proposed that a conformationally flexible unstructured state of the linker changes to a structured and docked one upon ATP binding, providing essential conformational change in the motor, responsible for subsequent stepping[46]. On one hand, the linker spring has to be then flexible enough to allow for diffusive search of the motor head of the next binding side. On the other, when both heads of the motor are simultaneously bound to the microtubule track, the *neck linker* has to be sufficiently stiff to ensure that mechanical forces between both head domains enable mechanical coupling. Accordingly, mechanical models and molecular dynamics simulations of this peptide structure are important contributions to understanding its elastic properties and ability to control kinesin's motion. It has been previously suggested by Zheng and Doniach that NMA does not describe kinesin dynamical properties as closely as in case of myosins[47]. Zheng and Doniach employed an approximated NMA model proposed by Tirion[48]. We employ a different approximation, where a distance dependence of the pair force constant is modelled exponentially, as described by Hinsen[20]. The decomposition into three domains, that we have observed and which follows expectations, opens up a possibility that Hinsen approximation is a better fit for kinesins.

In order to select a proper starting structure for stretching simulations, a construction of an initial, well equilibrated ensemble is required. It is often assumed that, for a small system with properly functioning temperature and pressure coupling, a run that does not exceed 100*ps* is enough to achieve a state of equilibrium[8,49]. However, it has been shown that such assumption does not have to be necessarily true[50,51]. Taking into consideration possible difficulties in achieving equilibrium, we have decided to measure fluctuations of the *end-to-end* distance, aside from a routine check of parameters (e.g. Root Mean Square Displacement (RMSD) or average potential energy), typically used as indicative measures for an equilibrated system.

Figure 3 displays *end-to-end* distance distributions in different time windows of the simulation with Gaussian curves fitted to the data. For a chain made up of orientationally uncorrelated (free-jointed) links with a length of each segment randomly distributed, the *end-to-end* stretch distance is expected to follow statistics of the Gaussian law. Although we have observed that the length distributions of the *neck linker* as well as the *Ala-only* polypeptide do fit Gaussian curves in certain time regimes, the long-run equilibrium simulations of the segments clearly indicate deviations of the *end-to-end* distances from Gaussianity, see Fig. 3. This observation stays in line with the assumed GROMACS force field, which apart from the harmonic approximations on bonds and angles, contains also long range interactions:

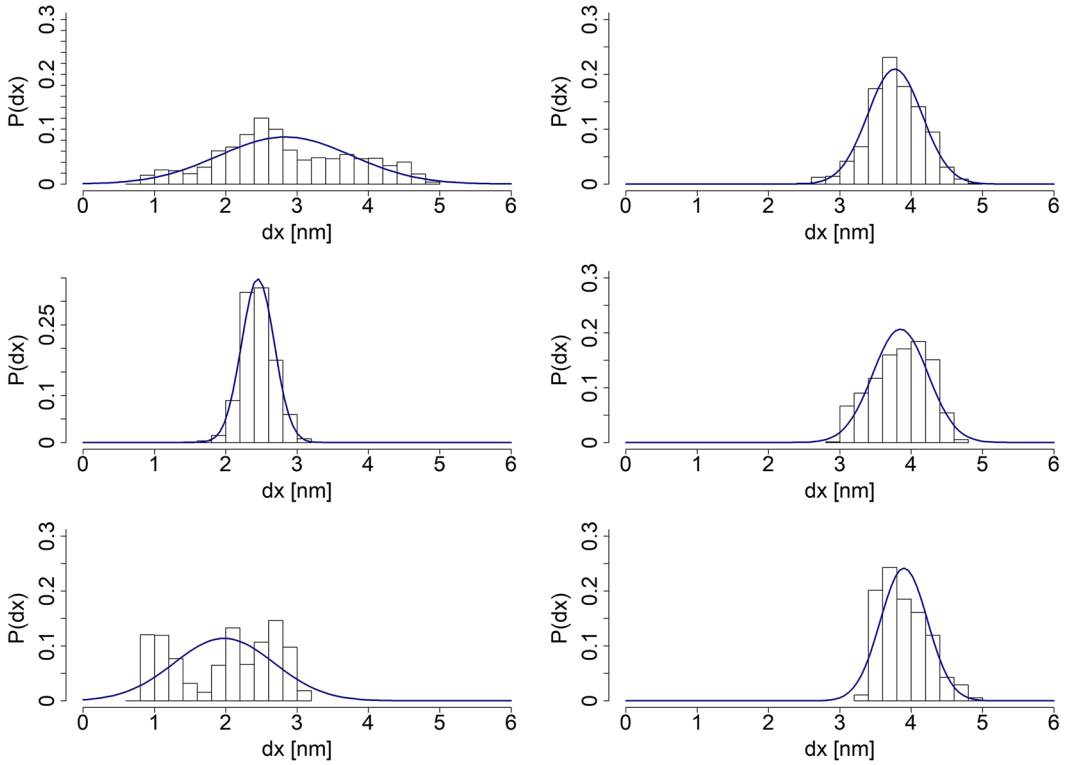

Figure 3. Left column includes (from top to bottom): the *neck linker's end-to-end* distance distributions of the whole equilibrium simulation, 10^4 recorded steps preceding the last 10^4 and, in the final row, the last 10^4 recorded steps. Right column includes the *Ala-only* peptide's *end-to-end* distance distributions arranged in an analogous way. Gaussian curves have been fitted to guide an eye, based on the mean and σ^2 of a given empirical distribution derived from the SMD simulations.

$$E = \sum_{bonds} \frac{k_i}{2}(x_i - x_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2}(1 + cos(n\omega - \gamma)) + \sum_{i}^{N} \sum_{j=i+1}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right). \quad (14)$$

Here x_i is a symbol of an i^{th} bond length, θ_i stands for an i^{th} angle, while $x_{i,0}$ and $\theta_{i,0}$ are their respective reference values. V_n is a parameter that gives information about rotation barriers of a torsion angle ω, while k_i refers to an i^{th} force constant. ϵ_{ij} is a minimal value of Van der Waals potential between atoms with indices i and j, r_{ij} is a distance between these atoms, σij is a distance between them when the Van der Waals potential value equals 0. The symbols q_i and q_j refer to charges of i^{th} and j^{th} atom respectively, and ϵ_0 stands for the dielectric constant.

Simulation results indicate that when the chain's structure attains a local minimum of the potential, the long range interactions do not play a significant role. Effectively, their influence on variations of the potential energy wanes temporarily. As a result, the chain is able to explore a narrow conformational subspace, behaving similarly to the Gaussian chain, before being pulled out of the energy well by thermal fluctuations. If the chain never leaves the vicinity of that particular energy well, an overall distribution of *end-to-end* distances approaches a normal distribution for sufficiently long simulation runs.

The mean *end-to-end* distance of the *Ala-only* chain over the whole simulation equals 3.77 ± 0.38 *nm*, hence such an equilibrated sequence has been chosen to be a starting point of the stretching process. In contrast, distribution of *end-to-end* distances of the *neck linker* structure is not as well fitted to a Gaussian curve. In order to make sure that a chosen structure is sufficiently close to the potential energy minimum, we have selected a model one with the *end-to-end* distance of 2.27 ± 0.56 *nm*, belonging to the class of conformations attained between 4.4 and 5.2 *ns* of simulation runs. Results of the pulling experiments performed on chosen *neck linker* and *Ala-only* chains are displayed in Fig. 4 and clearly indicate that both structures stretch at (almost) a constant rate for the majority of the process. At the same time, in accordance with findings reported in our prior studies[18], we observe that *Ala-only* chain's linear response drops much sooner than that of the *neck linker* structure.

While the *end-to-end* distance is a useful parameter in AFM experiments and those mimicking them *in silico*, it does not give detailed information regarding inner dynamics of the examined structure. In order to gather additional information that could hint at inner dynamic characteristics and elasticity of the analyzed biopolymer chains, we have measured pair distances between the 4th residue of the simulated chains and a selected residue of

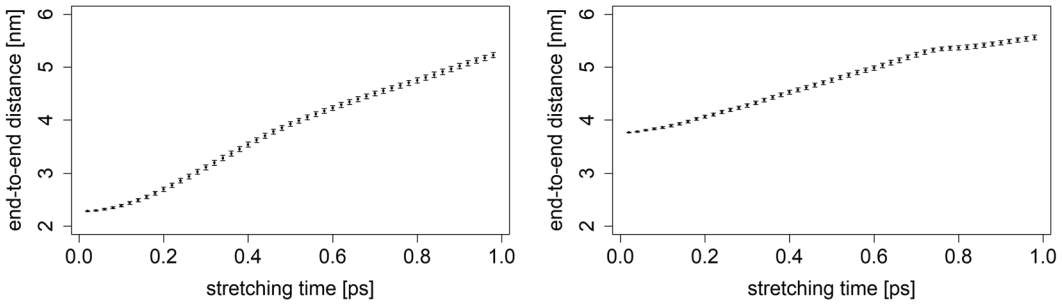

Figure 4. *End-to-end* distances as functions of time, averaged over 10^4 MD runs, each of duration $t = 1$ *ps*. The left plot depicts a change in the *neck linker's end-to-end* distance. The right plot displays analogous findings for the *Ala-only* chain.

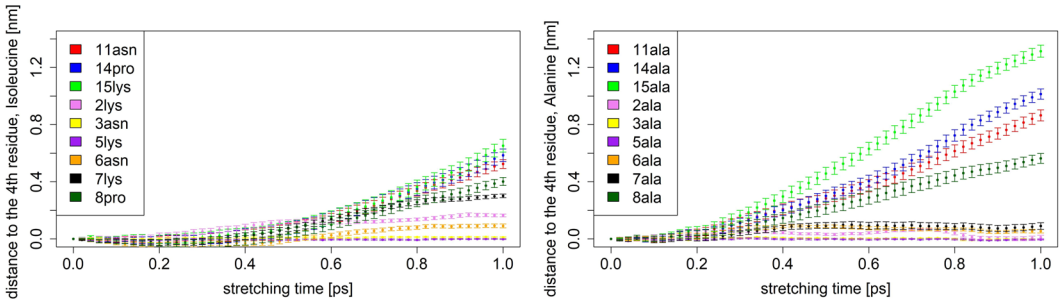

Figure 5. Distances (extensions) between residues as functions of time, averaged over 10^4 MD runs, each of duration $t = 1$ *ps*. The left plot depicts different rates of distance changes in the *neck linker's* chain between the 4th residue and selected residues, listed in the inset. The right plot displays analogous findings for the *Ala-only* chain.

interest. The 4th residue of our *neck linker* sequence is an Isoleucine amino acid which is one of the most prevalent elements at this position in a *neck linker* sequence across numerous kinesin families[8]. The second selected residue in the pair has been chosen as either neighboring Asparagine, Lysine or Proline. Asparagine and Lysine have been chosen for their significant propensity to be in contact with water (polar Asparagine and positively charged Lysine), while Proline - because of the presence of a Pyrrolidine, five-member ring in its side chain, being the only steric group of that kind in the whole *neck linker* chain. Effects of the simulation runs are displayed in Fig. 5 and document considerable difference in response to mechanical perturbations between the *neck linker* and the *Ala-only* chain.

The extensions between the 4th residue and its proximal contacts (the 2nd, 3rd, 5th and 6th residues in the chain) remain relatively unchanged throughout the stretching time, regardless of the type of the representative polymer chain. Stronger variations are observed for more distant pairs: for the *neck linker* structure significant changes in extension profiles between pairs of residues emerge in time windows longer than 0.6 *ps*. At the same time the *Ala-only* sequence shows pronounced variability in conformations by comparison to a much more rigid structure of the *neck linker*.

Positions of the 8th and 14th residues in the *neck linker* are taken by Proline which is known to reduce flexibility of the chain in the kink (cis) conformation. In fact, in former *in silico* studies examining mechanical properties of the *neck linker* domain from sequence analysis[8] it has been argued that the cis-trans isomerization of a conserved proline residue switching between straight and kink forms accounts for variations in resulting force-extension profiles and supports experimental observations[52] of the proline's isomerization influence on duration and effectiveness of biological processes dependent on protein folding.

In case of the *Ala-only* chain, displacement patterns between Isoleucine at 4th position and subsequent Alanine residues differ significantly from those observed for the *neck linker* chain stretched at constant pulling speed: In course of pulling experiment inner distances do not deviate much from their averages, whereas distances to external residues (at 8th, 11th, 14th and 15th positions) show pronounced extensibility. Altogether, while the final *end-to-end* distance of the *Ala-only* chain has been on average smaller than that of the *neck linker* (see Fig. 4), its inner extension distances reach greater lengths, to the point, where 15th residue of the polyalanine chain has almost twice the final value, when compared to the largest inner distances of the *neck linker*. This indicates that stretching of the *Ala-only* chain is far more complex, possibly with emergent inner dynamic domains facilitating extensions.

In order to further explore the specificity of *neck linker's* sequence and to determine how the presence of particular amino acid types affects that specificity, we have prepared 3 modified *neck linker* chains, where all Asparagine, all Lysine residues and the two Proline residues have been substituted with Alanine amino acids,

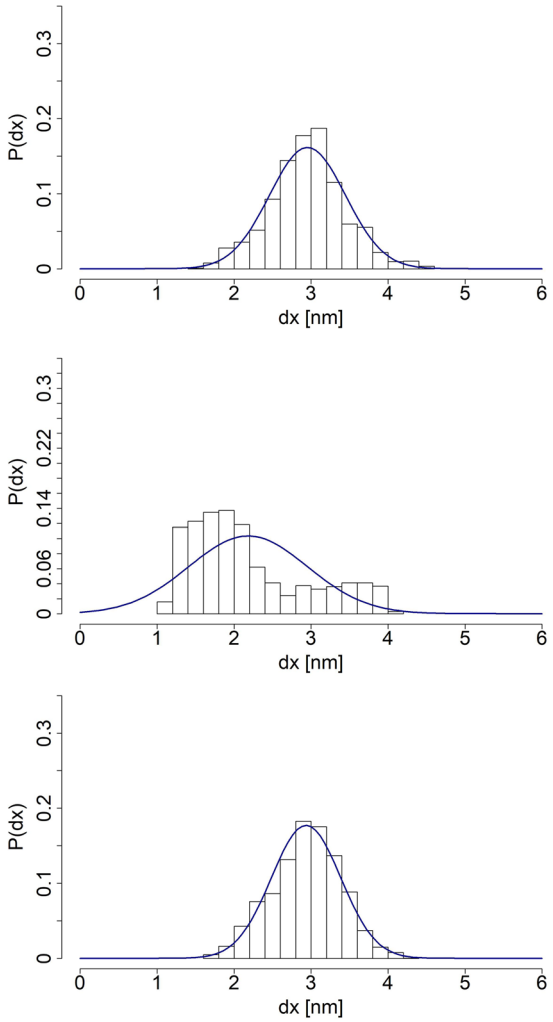

Figure 6. The *end-to-end* distance distributions of the *no-Asparagine* chain (top), the *no-Lysine* chain (middle) and the *no-Proline* chain (bottom), all pertaining to equilibrium simulations. Appropriate Gaussian curves have been fitted to the data, based on derived means and σ^2 of respective distributions.

respectively. Simulations' setup has been identical to the one employed in case of the unchanged *neck linker* and the *Ala-only* chain. The *no-Asparagine* and the *no-Proline* peptides seems to have easily achieved a local minimum of potential energy. On the other hand, the distribution of the *no-Lysine* chain *end-to-end* distances from the full simulation does not fit the Gaussian curve at all (Fig. 6). The possible reasons for such behaviour has been discussed above. Accordingly, in order to meet the requirement of beginning simulation runs with a mechanically equilibrated structure, as a starting conformation of the *no-Lysine* chain we have selected a structure from a time period, in which the *no-Lysine end-to-end* distances have been distributed normally (Fig. 7). All in all, *end-to-end* distances of equilibrated structures has been estimated to equal 2.95 ± 0.49 *nm*, 1.81 ± 0.21 *nm* and 2.94 ± 0.45 *nm* *no-Asparagine*, *no-Lysine* and *no-Proline* chain respectively.

It can be concluded from the Jarzynski's equality[43], as shown by Crooks[53], that a crossing point between work distributions of forward and reverse processes is equivalent to free energy difference ΔG between the resulting states and has been used in various studies of mechanical stability and folding/unfolding dynamics of biopolymers[54,55]. The approach allows us to determine ΔG between stretched and relaxed structures of the *neck linker* and *Ala-only* chains. It is difficult to ascertain *a priori* the possible range of work values in such an experiment, since it depends on several factors, such as the experiment's duration, the stretching force's value, the way that force is applied, etc. Nevertheless, we can infer from examples of similar experiments, like the one performed on DNA hairpins[40], that we should expect work values in range of hundreds of k_BT (where k_BT can be given as approximately 4.14 *pN* * *nm*). The stretching of protein barnase by way of optical tweezers gives work values of over 1000 k_BT[56] and the authors comment on different contributing factors. A study of short peptide's adsorption on amorphous SiO_2 shows that a friction emerging from pulling a peptide through water contributes to work in a linear way, quickly rising with dragging speed[57], while maximum work of pulling 18 residue-long pVEC peptide variants through the POPE bilayer has been estimated to be between 797 and 919 kcal/mol[58]. As it is, work values, we have arrived at, are within expectations for this type of simulation. As we can see in Fig. 8 (averages included

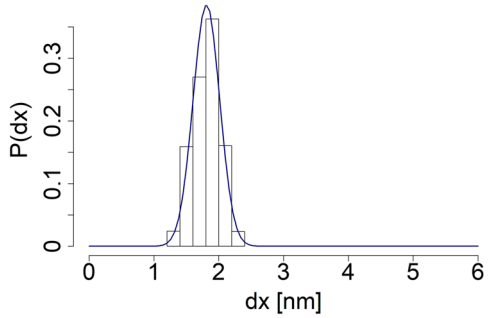

Figure 7. The *end-to-end* distance distribution of the last 10000 recorded steps taken from the *no-Lysine* peptide equilibrium simulation data. An appropriate Gaussian curve has been fitted to the data, based on its mean and σ^2.

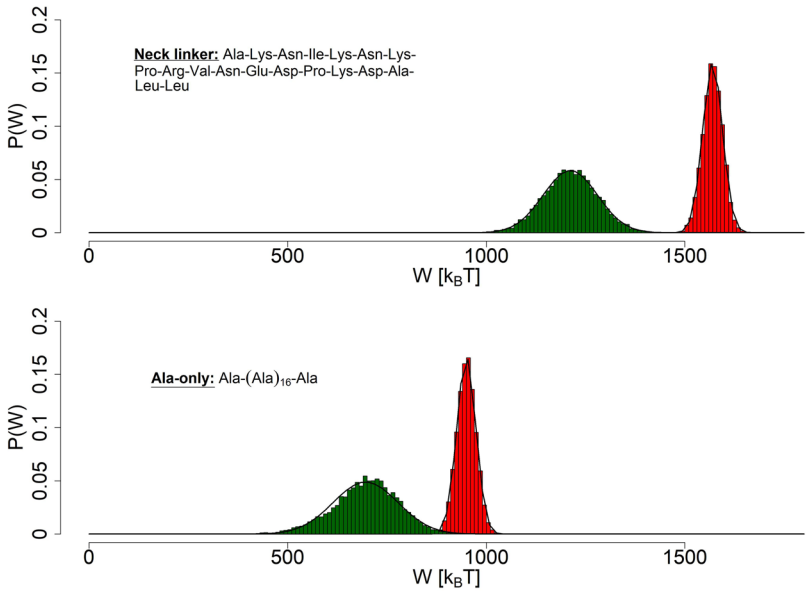

Figure 8. Reconstruction of probability density functions (PDFs) of work done on the *end-to-end* distance in pulling *in-silico* experiments of the modeled peptides (the stretching force value set at 1300 $kJ * mol^{-1} * nm^{-1}$). The Gaussian curves are fitted to the derived histograms of the *neck linker* (top) and *Ala-only* peptide (bottom). Crossing points of the curves are at 1462.4 k_BT and 882 k_BT, respectively. The ensembles of stretched structures are red, while the ensembles of relaxed structures are green colored.

Peptide type	$< W >_{stretch}$ $[k_BT]$	$< W >_{relax}$ $[k_BT]$	Δ G $[k_BT]$
neck linker	1569.9 ± 24.9	1211.6 ± 68.2	1462.4 ± 0.1
Ala-only	949.7 ± 23.9	696.0 ± 81.5	882 ± 0.1
no-Asparagine	733.7 ± 30.5	651.0 ± 63.3	690.5 ± 0.1
no-Lysine	985.6 ± 40.4	388.0 ± 66.7	752.8 ± 0.1
no-Proline	1056.8 ± 26.8	888.0 ± 63.9	997.6 ± 0.1

Table 1. Distribution means of stretched and relaxed chain ensembles of the original *neck linker*, the *Ala-only* peptide and 3 modified (mutant-like) structures, with the stretching force value set to 1300 $kJ * mol^{-1} * nm^{-1}$. ΔG values equivalent to crossing points between Gaussian curves of stretched and relaxed ensembles.

in Table 1) - for constant stretching force of 1300 $kJ*mol^{-1} * nm^{-1}$, work distributions derived from ensembles of stretched and relaxed *neck linker* structures are well separated and exhibit larger average values than those typical for the *Ala-only* chain. The free energy difference between two equilibrated states ΔG can be identified as $\Delta G = 1462.4\ k_BT$ for the *neck linker* and $\Delta G = 882\ k_BT$ for the *Ala-only* chain. These rough estimates of ΔG, used in conjunction with the Aarhenius definition of the rate constant, seem to imply that the *neck linker* not only stretches more effectively than the plain polyalanine peptide but also returns faster to the relaxed conformation. This conclusion is in line with its documented elasticity[8].

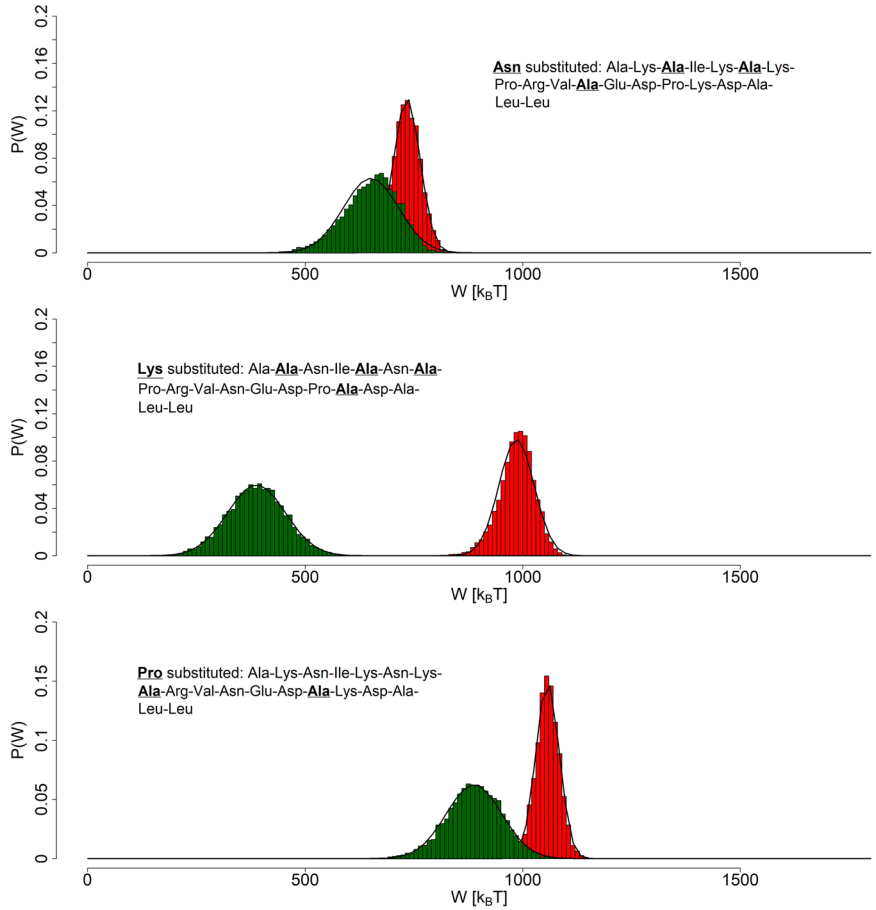

Figure 9. Reconstruction of probability density functions (PDFs) of work done on the *end-to-end* distance in pulling *in-silico* experiments on the modified linker chains (the stretching force value set at 1300 $kJ * mol^{-1} * nm^{-1}$). The Gaussian curves are fitted to the derived histograms of the *no-Asparagine* peptide (top), the *no-Lysine* peptide (middle) and the *no-Proline* one (bottom). The ensembles of stretched structures are red colored, while the ensembles of relaxed structures are green.

Distributions displayed in Fig. 9 (averages included in Table 1) suggest that, for constant stretching force of 1300 $kJ * mol^{-1} * nm^{-1}$, the elasticity of all 3 modified *neck linker* sequences suffered, compared to the original one (see Fig. 8). Just like in case of the polyalanine peptide, the value of work performed on them hardly crosses the point of 1000 k_BT, while that of the *neck linker* easily passes the 1500 k_BT mark. Additionally, the removal of Asparagine from the chain has influenced its ability to spontaneously retract most severely, both that ability and its stretching are less effective than that of the polyalanine peptide. The substitution of Proline seems to result in similar behaviour to the *Ala-only* chain. The *no-Lysine* chain's ability to retract seems to even surpass that of the original chain. All this may hint at Asparagine being a crucial part, when it comes to retracting during relaxation process. Asparagine side chain consists of sole amine group, which may have a stabilizing effect through its ability to partake in hydrogen bonding. Lysine side chain contains amine group as well, it is however preceded by a conventional chain of 4 methylene groups, a fact that probably keeps the amine group away from the peptide's backbone. Proline may be adding to the stabilising effects, as well as providing an extra push to the stretching ability with its conformational changes.

The data gathered in remaining *in-silico* experiments (stretching force values: 130, 400, 700 and 1000 $kJ * mol^{-1} * nm^{-1}$) have been used together with the data for the stretching force value of 1300 $kJ * mol^{-1} * nm^{-1}$ showcased in previous figures. In Fig. 10 the relation between fractional extension and ΔG has been shown, with all chains and force values included. The quadratic curves fitted to the data points confirm harmonic spring-like behaviour of simulated chains and stay in line with analysis of single polymer dynamics presented elsewhere (see e.g. ref. [45]). The *neck linker* curve is characterised by a quadratic coefficient of a greater value than all other curves, save one. In fact, the *neck linker* curve and the *polyalanine* (the *Ala-only* peptide) curve seem to define a range of coefficient values that go from the least unique chain (the *Ala-only* peptide) to the one present in properly functioning biostructures (the *neck linker*). Predictably, the least unique chain is also the least effective spring, while the other one is more effective. The curves depicting the characteristics of *no-Asparagine* chain and *no-Proline* chain fall in between these two extremes. It could thus suggest that chains decrease in effectiveness towards the *polyalanine* with the removal of *Proline* and *Asparagine* residues. The lack of *Asparagine* in sequence seems to have affected the stiffness least profoundly, with stiffness coefficient of *no-Asparagine* chain keeping the closest to that of *neck linker*. Intriguingly, the *no-Lysine* curve is to the far left of the other curves, including *neck linker* curve.

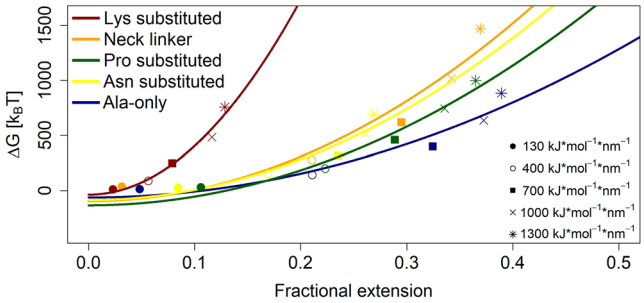

Figure 10. ΔG expressed as a function of the fractional extension for the *neck linker*, *Ala-only* peptide, *no-Asparagine* chain (*Asn substituted*), *no-Lysine* chain (*Lys substituted*) and *no-Proline* chain (*Pro substituted*). Fractional extension is defined as a ratio of the extension measured from the relaxed mean *end-to-end* distance to the rest of the path (*L* minus the relaxed mean *end-to-end* distance, where *L* is the contour length of the chain in question). The contour length per amino acid has been assumed to be equal to 4.0 Å[65]. Several different ΔG values have been acquired by repeating the *in silico* experiments while using different values of a constant stretching force equal to 130, 400, 700, 1000 and 1300 $kJ*mol^{-1}*nm^{-1}$. Parabolic curves have been fitted to the plots. Quadratic coefficients of fitted plots are: *Lys* substituted: 44.3 * $10^3 k_BT$; *neck linker*: 10.1 * $10^3 k_BT$; *Asn substituted*: 9.2 * $10^3 k_BT$; *Pro substituted*: 7.9 * $10^3 k_BT$; *Ala-only*: 5.4 * $10^3 k_BT$. They represent the effective stiffness of the stretched structure. The intercept coefficients are: −37.3, −95.8, −86.6, −134.4 and −63.1 k_BT, respectively.

Its quadratic coefficient value reflects this difference: it is close to 44.3 * $10^3 k_BT$ while the coefficients of 4 other curves have values ranging from 5.4 to 10.1 (*10^3 k_BT). Judging from this, we may argue that the substitution of *Lysine* residues actually improves the spring-like performance of the chain. It is possible to draw a conclusion from this particular result that the substitution of *Lysine* residues may improve *neck linker*'s performance within the context of kinesin motor's movement along microtubules. However, it is also possible to assume that there is an optimal range within which biologically viable springs operate and that such drastic surge of stiffness takes *no-Lysine* chain outside of this range. If it were to be so, then it is probable that this optimal biological range coincides with the range defined by *neck linker* and *polyalanine* chains, as shown in the Fig. 10. Indeed, further inquiries may prove enlightening as to whether *neck linker* chain corresponds to the optimal structure in the context of, first, kinesin protein, then whole group of molecular motors and, finally, in the context of all protein native structures.

Conclusions

Pulling experiments on single-molecules provide a quantitative characterization of unfolding and relaxation mechanisms of biomolecules. Despite many nanotechniques like atomic force microscopy, laser tweezers or fluorescence resonance energy transfer are available today and used in combined protocols, they may not reveal molecular mechanisms underlying modulation of protein's elasticity, especially under costly conditions of manipulating local mutations of investigated molecules. In order to overcome these difficulties, mechanical models of molecular dynamics can be used as guiding insight into consequences of local modifications of protein structure on its elasticity and response to external mechanical stress[16,26,27,59]. Computational all-atoms MD or coarse-grained MD simulations on single molecule pulling experiments are frequently a complementary tool in analysis of entropic elasticity of polymers or protein molecules and facilitate development and design of single-molecule force spectroscopy[8,40,45,46].

Notably, entropic forces have not been a frequent subject of discussion when it comes to nanomechanical devices, despite the fact that the entropy-functional devices may very well be easier to steer by means of external parameter manipulation (e.g. control of temperature, application of external fields etc.), with the chemical structure of the components remaining intact. It seems that gaining insight into the particulars of design of such devices could be beneficial to the engineering efforts that focus on artificial motors for nanoscale transport.

Here we have investigated force response in the structure of kinesin focusing on elastic properties of a spring connecting two separate domains (heads) of the motor protein. The fact that the motor is operating at molecular scale in the presence of viscous forces causes the inertia to be less pronounced and leads to the overdamped dynamics becoming a valid approximation[1,60–62].

In contrast, our numerical simulations follow the procedure in which full set of Newton's equations of motion are solved to propagate in time the coordinates of all atoms of the structure that is under consideration.

Presented study adds to the notion that the *neck linker* regions possess mechanical properties not found in an arbitrary amino acid chain. The *neck linker's* models modified with point mutations clearly exhibit different responses to stretching force in comparison to the intact, original structure. The lack of *Lysine* residues seems to alter the behaviour of the chain most significantly and in an unexpected opposite direction, in contrast to other mutations. At this point, it still remains unclear, how much those differences depend on amino acid type and how much on the number of residues being substituted or whether the placement within the sequence factors in. It is also impossible to assert, whether the amino acid types' impacts on *neck linker* properties are merely additive, or if the certain residues' combinations produce more nuanced interactions. Future studies, both experimental and

theoretical, may shed more light not only on mutation-dependent mechanical properties of *neck linker* but subsequently on effects of perturbing motor proteins by introduced point mutations and on adaptation of the resulting stepping mechanism to those changes.

Data availability

The file containing initial atom coordinates of the *neck linker* region, used in the beginning of the simulations' setup, is available at https://www.rcsb.org/structure/3b6u. The GROMACS package employed to execute simulations can be obtained from http://www.gromacs.org/Downloads. As for the datasets generated and analysed during the current study, they are available from the corresponding author on reasonable request.

Received: 20 September 2018; Accepted: 24 February 2020;
Published online: 16 March 2020

References

1. Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Associates, Sunderlans, MA, 2001).
2. Kolomeisky, A. B. Motor proteins and molecular motors: how to operate machines at the nanoscale. *J. Physics: Condens. Matter* **25**, 463101 (2013).
3. Kolomeisky, A. B. & Phillips Iii, H. Dynamic properties of motor proteins with two subunits. *J. Physics: Condens. Matter* **17**, S3887 (2005).
4. Teimouri, H., Kolomeisky, A. B. & Mehrabiani, K. Theoretical analysis of dynamic processes for interacting molecular motors. *J. Phys. A: Math. Theor.* **48**, 065001 (2015).
5. Hyeon, C. & Onuchic, J. N. A structural perspective on the dynamics of kinesin motors. *Biophys. J.* **101**, 2749–2759 (2011).
6. Zhang, Z. & Thirumalai, D. Dissecting the kinematics of the kinesin step. *Structure* **20**, 628–640 (2012).
7. Rice, S. *et al*. A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).
8. Hariharan, V. & Hancock, W. O. Insights into the mechanical properties of the kinesin neck linker domain from sequence analysis and molecular dynamics simulations. *Cell. Mol. Bioeng.* **2**, 177–189 (2009).
9. Phillips, R. K., Peter, L. G., Gilbert, S. P. & Rayment, I. Family-specific kinesin structures reveal neck-linker length based on initiation of the coiled-coil. *J. Biol. Chem.* **291**, 20372–20386 (2016).
10. Case, R. B., Rice, S., Hart, C. L., Ly, B. & Vale, R. D. Role of the kinesin neck linker and catalytic core in microtubule-based motility. *Curr. Biol.* **10**, 157–160 (2000).
11. Yildiz, A., Tomishige, M., Gennerich, A. & Vale, R. D. Intramolecular strain coordinates kinesin stepping behavior along microtubules. *Cell* **134**, 1030–1041 (2008).
12. Scarabelli, G. & Grant, B. J. Mapping the structural and dynamical features of kinesin motor domains. *PLoS Computational Biology* **9**, e1003329 (2013).
13. Hwang, W., Lang, M. J. & Karplus, M. Kinesin motility is driven by subdomain dynamics. *Elife* **6**, e28948 (2017).
14. Chakraborty, S. & Zheng, W. Decrypting the structural, dynamic, and energetic basis of a monomeric kinesin interacting with a tubulin dimer in three atpase states by all-atom molecular dynamics simulation. *Biochemistry* **54**, 859–869 (2015).
15. Bouchiat, C. *et al*. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.* **76**, 409–413 (1999).
16. Kutys, M., Fricks, J. & Hancock, W. O. Monte Carlo analysis of neck linker extension in kinesin molecular motors. *PLoS Comput. Biol.* **6**, e1000980 (2010).
17. Padinhateeri, R. & Menon, G. I. Stretching and bending fluctuations of short DNA molecules. *Biophys. J.* **104**, 463–471 (2013).
18. Lisowski, B., Świątek, M., Żabicki, M. & Gudowska-Nowak, E. Understanding operating principles and processivity of molecular motors. *Acta Phys. Pol. B* **43**, 1073 (2012).
19. Kozielski, F. *et al*. The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. *Cell* **91**, 985–994 (1997).
20. Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins* **33**, 417–429 (1998).
21. Kaan, H. Y. K., Hackney, D. D. & Kozielski, F. The structure of the kinesin-1 motor-tail complex reveals the mechanism of autoinhibition. *Science* **333**, 883–885 (2011).
22. Ramaiya, A., Roy, B., Bugiel, M. & Schäffer, E. Kinesin rotates unidirectionally and generates torque while walking on microtubules. *Proc. Natl. Acad. Sci*. **114**, 10894–10899 (2017).
23. Liu, H.-L., Pemble IV, C. W. & Endow, S. A. Neck-motor interactions trigger rotation of the kinesin stalk. *Sci. Reports* **2**, 236 (2012).
24. Hinsen, K. & Field, M. J. Analysis of domain motions in large proteins. *Proteins* **34**, 369–382 (1999).
25. Reuter, N., Hinsen, K. & Lacapère, J.-J. Transconformations of the serca1 ca-atpase: a normal mode study. *Biophys. J.* **85**, 2186–2197 (2003).
26. Meirovitch, H. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulations. *Curr. Opin. Struct. Biol.* **17**, 181–186 (2007).
27. Chong, S.-H. & Ham, S. Configurational entropy of protein. *Chem. Phys. Lett.* **504**, 225–229 (2011).
28. Hill, T. *An Introduction to Statistical Thermodynamics* (Dover Publications, New York, NY, 1986).
29. Lawrence, C. J. *et al*. A standardized kinesin nomenclature. *J. Cell Biol.* **167**, 19–22 (2004).
30. Berman, H. M. *et al*. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
31. Uemura, S. *et al*. Kinesin–microtubule binding depends on both nucleotide state and loading direction. *Proc. Natl. Acad. Sci.* **99**, 5977–5981 (2002).
32. Uemura, S. & Ishiwata, S. Loading direction regulates the affinity of adp for kinesin. *Nat. Struct. & Mol. Biol.* **10**, 308 (2003).
33. Hamdi, M., Sharma, G., Ferreira, A. & Mavroidis, C. Characterization of protein based spring-like elastic joints for biorobotic applications. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*., 1794–1799 (IEEE, 2006).
34. Kumari, J. L. J., Sudan, R. J. J. & Sudandiradoss, C. Evaluation of peptide designing strategy against subunit reassociation in mucin 1: A steered molecular dynamics approach. *PloS One* **12**, e0183041 (2017).
35. Pronk, S. *et al*. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
36. Berendsen, H. J., van der Spoel, D. & van Drunen, R. Gromacs: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
37. Lindahl, E., Hess, B. & Van Der Spoel, D. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model. Annual* **7**, 306–317 (2001).
38. Jorgensen, W. L. & Tirado-Rives, J. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988).
39. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).

40. Alemany, A. & Ritort, F. Fluctuation theorems in small systems: extending thermodynamics to the nanoscale. *Europhys. News* **41**, 27–30 (2010).
41. Hummer, G. & Szabo, A. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.* **85**, 5–15 (2003).
42. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics using Jarzynski's equality. *J. Chem. Phys.* **119**, 3559 (2003).
43. Jarzynski, C. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Ann. Rev. Cond. Matt. Phys* **2**, 329–351 (2011).
44. Dellago, C. & Hummer, G. Computing equilibrium free energies using non-equilibrium molecular dynamics. *Entropy* **16**, 41–61 (2014).
45. Latinwo, F. & Schroeder, C. M. Determining elasticity from single polymer dynamics. *Soft Matter* **10**, 2178–2187 (2014).
46. Block, S. Kinesin motor mechanics; binding, stepping, tracking, gating and limping. *Biophys. J* **92**, 2986–2995 (2007).
47. Zheng, W. & Doniach, S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl. Acad. Sci.* **100**, 13253–13258 (2003).
48. Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908 (1996).
49. Zhang, Z., Shi, Y. & Liu, H. Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys. J.* **84**, 3583–3593 (2003).
50. Genheden, S. & Ryde, U. Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys. Chem. Chem. Phys.* **14**, 8662–8677 (2012).
51. Smith, L. J., Daura, X. & van Gunsteren, W. F. Assessing equilibration and convergence in biomolecular simulations. *Proteins* **48**, 487–496 (2002).
52. Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. Prolyl cis-trans isomerization as a molecular timer. *Nat. Chem. Biol.* **3**, 619–629 (2007).
53. Crooks, G. E. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **61**, 2361–2366 (2000).
54. Alemany, A., Ribezzi-Crivellari, M. & Ritort, F. From free energy measurements to thermodynamic inference in nonequilibrium small systems. *New J. Phys.* **17**, 075009 (2015).
55. Fox, R. F. Using nonequilibrium measurements to determine macromolecule free-energy differences. *Proc. Natl. Acad. Sci.* **100**, 12537–12538 (2003).
56. Alemany, A., Rey-Serra, B., Frutos, S., Cecconi, C. & Ritort, F. Mechanical folding and unfolding of protein barnase at the single-molecule level. *Biophys. J.* **110**, 63–74 (2016).
57. Meißner, R. H., Wei, G. & Ciacchi, L. C. Estimation of the free energy of adsorption of a polypeptide on amorphous sio 2 from molecular dynamics simulations and force spectroscopy experiments. *Soft Matter* **11**, 6254–6265 (2015).
58. Akdag, I. O. & Ozkirimli, E. The uptake mechanism of the cell-penetrating pvec peptide. *J. Chem*. **2013** (2013).
59. Harris, N. C., Song, Y. & Kiang, C.-H. Experimental free energy surface reconstruction from single-molecule force microscopy using Jarzynski's equality. *Phys. Rev. Lett.* **99**, 068101 (2007).
60. Linke, H., Downton, M. & Zuckermann, M. Performane characteristics of Brownian motors. *Chaos* **15**, 026111 (2005).
61. Bier, M. Processive motor protein as an over damped Brownian stepper. *Phys. Rev. Lett.* **91**, 148104 (2003).
62. Żabicki, M., Ebeling, W. & Gudowska-Nowak, E. The thermodynamic cycle of an entropy-driven stepper motor walking hand-over-hand. *Chem. Phys.* **375**, 472–478 (2010).
63. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
64. Stone, J. *An Efficient Library for Parallel Ray Tracing and Animation*. Master's thesis, Computer Science Department, University of Missouri-Rolla (1998).
65. Ainavarapu, S. R. K. *et al*. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. *Biophys. J.* **92**, 225–233 (2007).

Acknowledgements

This project has been supported in part by the grant from National Science Center 2014/13/B/ST2/02014.

Author contributions

M.Ś. and E.G-N. conceived and designed the research. SMD simulations were conducted by M.Ś. and followed by statistical analysis and interpretation of the results performed together with E.G-N. Both authors (M.Ś. and E.G-N.) wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.Ś.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.