

## Thermal fluctuations assist mechanical signal propagation in coiled-coil proteins

Judit Clopés<sup>1</sup>, Jaeoh Shin<sup>1,2</sup>, Marcus Jahnel<sup>3,4</sup>, Stephan W. Grill<sup>1,3,4,5</sup>, and Vasily Zaburdaev<sup>1,6,7\*</sup>

<sup>1</sup>Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany

<sup>2</sup>Department of Chemistry, Rice University, Houston 77005 TX, USA

<sup>3</sup>Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany

<sup>4</sup>Biotechnology Center, Technical University Dresden, Tatzberg 47/49, 01307 Dresden, Germany

<sup>5</sup>Cluster of Excellence Physics of Life, TU Dresden, Dresden, Germany

<sup>6</sup>Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany and

<sup>7</sup>Max-Planck-Zentrum für Physik und Medizin, Erlangen, Germany

Recently it has been shown that the long coiled-coil membrane tether protein Early Endosome Antigen 1 (EEA1) switches from a rigid to a flexible conformation upon binding of a signaling protein to its free end. This flexibility switch represents a novel motor-like activity, allowing EEA1 to generate a force that moves vesicles closer to the membrane they will fuse with. To elucidate how binding of a single signaling protein can globally change the stiffness of a 200 nm long chain, we propose a simplified description of the coiled-coil as a one-dimensional Frenkel-Kontorova chain. Using numerical simulations, we find that an initial perturbation of the chain can propagate along its whole length in the presence of thermal fluctuations, changing the configuration of the entire molecule and thereby affecting its stiffness. Our work sheds light onto intramolecular communication and force generation in long coiled-coil proteins.

Coiled-coils (CC) are structural protein motifs composed of two or more intertwined  $\alpha$ -helices. They play essential roles in multiple cellular processes, including gene regulation, muscle contraction, and cell signaling [1]. Early Endosome Antigen 1 (EEA1) is an extended fibrous protein with a large dimeric CC domain that guides vesicles coated with the small GTPase Rab5 to early endosomes [2]. It has recently been shown that unbound EEA1 is rather stiff and becomes more flexible when active Rab5 selectively binds to its N-terminus [3]. The entropic force of the flexible chain brings the tethered vesicle closer to the endosome membrane to facilitate the downstream vesicle fusion. Next, GTP hydrolysis breaks the interaction between EEA1 and Rab5, suggesting reversibility of this process. Similar behavior has been observed for other long CC tethering proteins, such as the golgin family [4–7]. However, the underlying mechanism of the conformational switches is not fully understood. In particular, how does a local perturbation induced by a ligand binding at one end of the CC propagate through the whole structure of the protein to change its global flexibility?

Structurally, the center part of EEA1 folds as a long canonical homodimeric CC of approximately 200 nm (1275 amino acids) on almost its total length [8]. The sequence of  $\alpha$ -helices forming ideal left-handed CCs shows a periodic structural motif of *hpphpp* known as heptad repeat, where *h* and *p* are the hydrophobic and polar residues, respectively [9]. In an aqueous environment, the CC fold is formed by the assembly of  $\alpha$ -helices driven mainly by hydrophobic interactions established between the *h* residues of different strands. The regularity in the  $\alpha$ -helix sequence is interrupted by short discontinuity regions distributed along the structure (see Fig. 1,

(a)). While the nature of the fold in these regions is not known, the recent findings have related the flexibility change with the presence of such discontinuities; the mutant EEA1 that lacks these regions does not show the stiffness switch [3]. Therefore, we hypothesize that the CC regions could be involved in the energy transmission of the Rab5-EEA1 interaction through the whole structure whereas the discontinuities could be responsible for inducing the conformational change. To ensure energy transmission, the ideal CC regions in EEA1 (without discontinuities) must have a structure capable of propagating the perturbation with a sufficient signal amplitude to the discontinuity regions without destroying the dimeric CC fold.

The main interest of this work lies in understanding the feasibility of this energy transmission mechanism from the physical point of view. For that, we applied an overdamped one-dimensional Frenkel-Kontorova (FK) model [10] to a novel range of problems of signal transduction. We analyzed the dynamics of dislocations in FK chains for the length and time scales characteristic to the EEA1 protein, and we find that thermal fluctuations are necessary for the signal to propagate through the whole structure. Moreover, we provide the range of parameters when such signal propagation is feasible within the experimentally measured time. Finally, we briefly discuss how the small perturbation can trigger the change of the global properties of EEA1 in the framework of our model.

*a. The Frenkel-Kontorova model applied to coiled-coil structures.* The hydrophobic interactions between strands in an ideal CC fold can be described in terms of a one-dimensional FK chain [10, 12] (see Fig. 1, (b) and (c)). The original FK model was introduced in the late 1930s to describe defect dynamics in solids. In the context of biological systems, it has been used to model phenomena such as friction [13, 14], protein folding [15], chromatin dynamics [16], and DNA breathing

\* vasily.zaburdaev@fau.de

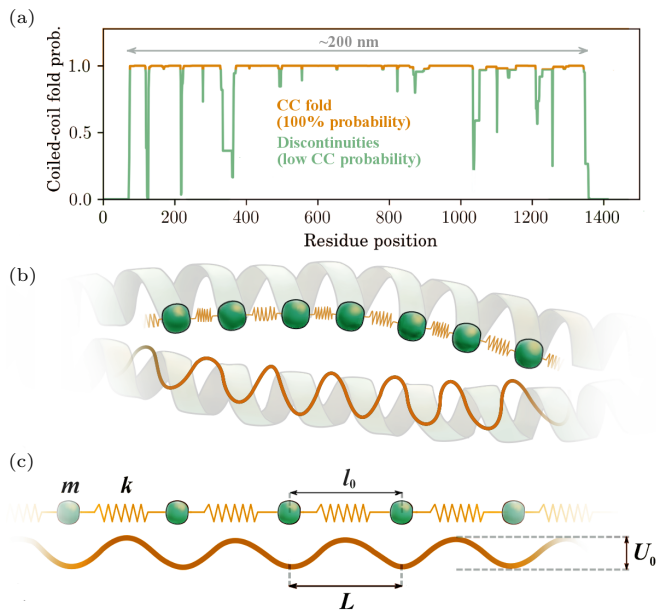


FIG. 1. Modeling hydrophobic interactions between strands in a canonical coiled-coil structure (a) Coiled-coil fold prediction for EEA1 in humans. Prediction plot computed with PCOILS bioinformatics toolkit [11] using a window of 28 amino acids. (b) Sketch of the application of the Frenkel-Kontorova model to map the hydrophobic contacts stabilizing the two  $\alpha$ -helical strands forming a homodimeric canonical coiled-coil protein. (c) Frenkel-Kontorova model represented by a chain of beads connected by harmonic springs in the periodic potential.  $U_0$  and  $L$  are the strength and the period of the hydrophobic interaction potential respectively,  $l_0$  is the half turn separating two hydrophobic residues and the equilibrium length of the harmonic spring connecting two beads with the spring constant  $k$ . Note that  $l_0$  coincides with  $L$  as both  $\alpha$ -helices in EEA1 are identical in the amino acid sequence.

[17]. When applied to CCs, this model is coarse-grained to the hydrophobic residue level, where each heptad repeat is described by two beads connected by a spring. The periodic potential with amplitude  $U_0$  and period  $L$  characterizes the strength of the hydrophobic interaction between residues of opposite strands. Since the two  $\alpha$ -helices forming the dimeric CC in EEA1 have an identical amino acid sequence and align in parallel [8], the effect of the hydrophobic interactions in the FK chain was chosen to be commensurate where the equilibrium length of the springs  $l_0$  coincides with the potential period  $l_0 = L$ . The elastic constant  $k$  of the spring connecting two neighboring beads is related to the contributions from both the rigidity at small stretches and the hydrogen bonds established between the 3rd and 4th nearest-neighbors in  $\alpha$ -helices (half heptad repeat). We have modeled the canonical coiled-coil regions in EEA1 by means of an overdamped FK chain. We first show that thermal fluctuations are necessary to ensure the propagation of an initial perturbation on one extreme through the whole

length of the chain. Then, we study the chain parameters which allow for the signal transmission and relate these results to the EEA1 protein.

The equation of motion for the  $i$ -th bead in an overdamped FK chain in terms of Langevin dynamics is

$$\gamma \frac{dx_i(t)}{dt} = -\partial_{x_i} U_i(t) + \xi_i(t), \quad (1)$$

where  $x_i(t)$  is the position of the bead,  $\gamma$  is the friction introduced by both the medium and the interaction between opposite strands, and  $\xi_i(t)$  is a Gaussian white noise describing thermal fluctuations, satisfying  $\langle \xi_i(t) \rangle = 0$  and  $\langle \xi_i(t) \xi_j(t') \rangle = 2\gamma k_B T \delta(t - t') \delta_{i,j}$ . The potential energy  $U_i(t)$  is given as

$$U_i(t) = \frac{-U_0}{2} \cos\left(\frac{2\pi x_i}{l_0}\right) + \frac{k}{2} \left( (x_{i+1} - x_i - l_0)^2 + (x_{i-1} - x_i - l_0)^2 \right), \quad (2)$$

In the absence of thermal noise, the compression or expansion (kink and antikink) introduced between a pair of particles in an overdamped commensurate FK chain relax to a finite length characterized by  $\lambda = l_0^2 \sqrt{k/U_0}$  [12, 13]. This length  $\lambda$ , known as kink width, determines the region of the chain affected by a compression or an expansion of length  $l_0$ .

We introduce the Rab5-EEA1 interaction as a register shift leading to a relative sliding of half a heptad repeat in one strand with respect to the opposite. This mechanism could introduce long distance communication along the CC fold at a low energetic cost, since it does not involve the exposure of hydrophobic residues to the aqueous environment. A similar sliding mechanism has been suggested to be responsible for the allosteric communication along transmembrane proteins [21]. In our model, the initial perturbation from Rab5 binding is introduced by generating a compression (kink) of length  $l_0$  between the leftmost pair of beads of the FK chain and keeping the position of the first bead fixed. We then analyzed the dynamics of the potential energy distribution per bead on the FK chain  $U_i(t)$ . The signal propagation analysis was performed by averaging single kink trajectories for kinks introduced in one of the extremes and in the center of the chain. The signals were post-processed using the Savitzky-Golay filter [18]. We then analyzed the propagation front  $x_{fr}$  (see Fig. 2 (c)), defined as the length at which the filtered signal amplitude reaches the energy offset, which is the mean energy of the beads introduced solely by temperature in a resting chain.

*b. Characterization of the signal transmission* The energy transport along the FK chain is a result of two processes: the relaxation of the contraction and a diffusive propagation of the kink due to thermal effects. In the absence of thermal fluctuations, or at low temperature, the contraction applied to the extreme relaxes to a width

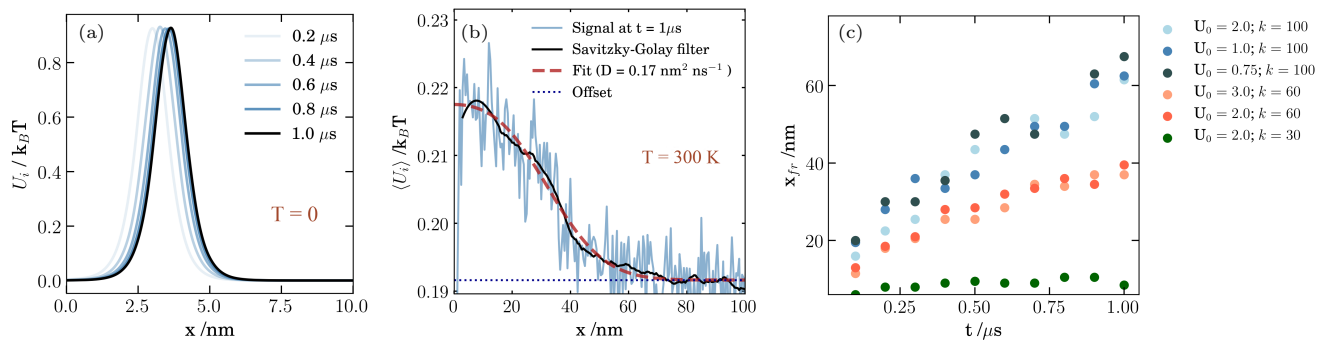


FIG. 2. Signal propagation analysis. (a) Evolution of the potential energy distribution profile along a Frenkel-Kontorova chain at zero temperature with  $U_0=1 k_B T$  and  $k=100 k_B T \text{ nm}^{-2}$ . The initial contraction relaxes and does not propagate. (b) Potential energy distribution profiles averaged over 5000 trajectories for identical chains as in (a), but now affected by thermal fluctuations. In this case, the compression is active and propagates diffusively. The averaged signal has been filtered by applying the Savitzky-Golay algorithm with a window size of 51 data points [18]. The fit curve (maroon dashed line) corresponds to the solution to the diffusion equation with a reflecting boundary at  $x = 0$  with the diffusion constant  $D$  extracted from the fit. (c) Dynamics of the position of the potential energy front  $x_{fr}$  for different  $U_0$  and  $k$  pairs. The front position  $x_{fr}$  is the position where the filtered signal reaches the offset. The units of  $U_0$  and  $k$  are  $k_B T$  and  $k_B T \text{ nm}^{-2}$  in all plots, respectively.

determined by  $\lambda$ , and does not propagate further (see Fig. 2, (a)). At intermediate temperatures ( $U_i \gtrsim k_B T$ ), a single kink diffuses while maintaining a constant width  $\lambda$ . Higher temperatures ( $U_i < k_B T$ ) melt the coiled-coil structure by spontaneously generating kink-antikink pairs along the chain. The relaxation of the kink at zero temperature occurs much faster than the characteristic diffusion time. Therefore, both phenomena can be decoupled [22], and the diffusive regime is recovered at large time scales (see Fig. 2, (b)).

In the absence of thermal fluctuations, the chain remains on its new energy minimum configuration without any motion once the elastic energy accumulated by the register shift  $\delta U = \frac{k}{2} l_0^2$  is released. Thermal fluctuations enable the chain monomers to cross the potential barrier yielding further propagation of the energy. The kink jumps are, therefore, more frequent as the potential depth  $U_0$  decreases, so the overall signal propagation is faster for decreasing  $U_0$ . Furthermore, the propagation is enhanced in stiffer chains, i.e., by increasing the elastic constant  $k$ , since the energy introduced in the initial perturbation is greater (see Fig. 2 (c)). This phenomenon is reminiscent of the motion of a Brownian particle or a polymer in a tilted washboard potential [23]. The register shift in our model plays a similar role of tilting the potential.

The potential energy distribution at chain positions away from the boundary  $x > \lambda$  shows a diffusive evolution at large times (see Fig. 2 (b)). The fit (maroon dashed line) results from the solution of the diffusion equation with a reflecting boundary at  $x = 0$  and with the initial condition as the Dirac delta function at  $x_{pk}$ .

$$P(x \geq 0, t) = \frac{A_0}{\sqrt{4\pi Dt}} \left\{ \exp \left[ \frac{-(x - x_{pk})^2}{4Dt} \right] + \exp \left[ \frac{-(x + x_{pk})^2}{4Dt} \right] \right\}. \quad (3)$$

where the diffusion constant  $D$  is extracted from the fit and  $A_0$  is the normalization constant. In our situation,  $x_{pk}$  is expected to be very close to reflecting boundary, at the distance of the order of the width of the kink. For simplicity, we do not specify this value and just use it as an additional fit parameter.

*c. Relation to the case of the EEA1 coiled-coil protein*  
The parameter set used to reproduce the physiological conditions at which CC proteins are functional is listed in Table I. The equilibrium distance between two beads corresponds to the length of a turn in canonical CCs,  $l_0$ , which has an approximate value of 0.5 nm [24, 25]. The range of the elastic constant  $k$  corresponds to the stretching of coiled-coil regions of myosin in the limit of small pulling forces, where the elasticity is entropic, until the first hydrogen bond in the  $\alpha$ -helix (half heptad repeat) is broken at roughly  $65 k_B T \text{ nm}^{-2}$  [26], which is consistent with the maximum longitudinal bending stiffness of the tropomyosin coiled-coil used in [27]. For short peptide chains, the elastic constant  $k$  is reduced to values around  $40 k_B T \text{ nm}^{-2}$  [28]. The friction coefficient  $\gamma$  is estimated using the Stokes law  $\gamma = 6\pi\eta R$  with the dynamic viscosity of water at 300 K,  $\eta_w = 0.85 \text{ pN ns nm}^{-2}$  [29], and assuming the amino acid radius  $R$  to be of the order of  $l_0/2$ . We estimate a maximum strength of the hydrophobic interactions of  $4 k_B T$ , that corresponds to the hydration potential of isoleucine [30], although accurate measurements of the solvation free energies of biomolecules are difficult to access experimentally [31, 32]. Since the CC

structure is not melting, but sliding, this value for  $U_0$  might be overestimating the hydrophobic interaction established between the residues of opposite strands. Similarly, the energy required for transferring buried residues from the interior to the protein surface is much lower than its full solvation [30]. Since the Rab5-EEA1 interaction and the consequent signal propagation occur without GTP hydrolysis [3], the activation energy  $E_A$  must be less than  $20 k_B T$ .

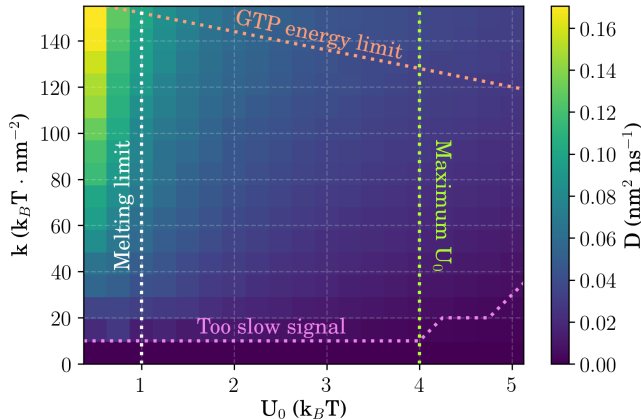


FIG. 3. Diffusion coefficient of a kink generated in the center of a Frenkel-Kontorova chain for different pairs of  $k$  and  $U_0$ , each extracted from the energy average of 2000 kink trajectories. The upper limit (orange) corresponds to the activation energy  $E_A$  required to overcome the potential energy barrier for energies below that released by GTP hydrolysis ( $E_A = 20 k_B T$ ). The magenta line limits the cases where the diffusion constant is too low to be able to transmit the signal within the maximum allowed time,  $t_c = 0.1s$ . Hydrophobic energy differences  $U_0$  below  $1 k_B T$  lead to unstable chains (white). Finally, the maximum energy released between a pair of hydrophobic residues  $U_0$  is about  $4 k_B T$  [30] (shown in light green).

The ranges of  $k$  and  $U_0$  that would allow the signal to propagate until the opposite end of the EEA1 protein are shown in Fig. 3. The diffusion constant  $D$  has been obtained by fitting the solution to diffusion equation (Gaussian distribution) in an unbounded domain to the average of 2000 kink trajectories generated at the center of the chain. The diffusion constant  $D$  for kinks introduced both at the center and at the extreme of the chain are equivalent in the long time limit.

The sum of the energy contributions required to overcome the potential barrier  $U_0$ , and to compress the first monomer to a length  $l_0$ , must not surpass the energy released by the hydrolysis of a GTP molecule ( $\approx 20 k_B T$ ),

$$\frac{1}{2} k l_0^2 + U_0 \leq 20 k_B T. \quad (4)$$

This provides an upper limit for the values of the elastic constant  $k$  and the potential well  $U_0$  (shown in orange). The time required to complete a full EEA1-Rab5

cycle is around  $t_c = 0.1 s$  [3]. This delimits the minimum required signal speed according to  $t_c \geq \langle N^2 \rangle / 2D$ , where  $N$  is the length of the chain. From the results shown in Fig. 3 and the physiological parameters in Table I, we predict the ranges for the hydrophobic interaction energy  $U_0 \in (1.0, 4.0) k_B T$  and an elasticity ranging from  $k \in [10, 150] k_B T \text{ nm}^{-2}$  to  $k \in [10, 125] k_B T \text{ nm}^{-2}$  depending on  $U_0$ , allowing for signal propagation in EEA1 at physiological conditions. Hydrophobic contacts  $U_0$  with less than  $1 k_B T$  lead to unstable chains due to the spontaneous generation of kink-antikink pairs already observed in the results (white dotted line).

TABLE I. Choice of biologically relevant parameters for the Frenkel-Kontorova model applied to describe ideal coiled-coil structures.

Parameter	Range	Reference
Turn length, $L, l_0$	0.5 nm	[24, 25]
Chain length, $N$	400 beads	[33]
Hydrophobic int., $U_0$	$< 4 k_B T$	[30]
Elastic constant, $k$	$1 - 65 k_B T \text{ nm}^{-2}$	[26–28]
Dyn. water viscosity, $\eta_w$	$0.85 \text{ pN ns nm}^{-2}$	[29]
Activation energy, $E_A$	$< 20 k_B T$	[34]

*d. Conclusions* How a local interaction affects the global property of a molecule is an important question in physical and biological systems. Recent experiments on the membrane tether EEA1 have shown that binding of the small GTPase Rab5 to the N-terminus of EEA1 significantly reduces the bending stiffness of this long coiled-coil protein. To elucidate this novel phenomenon, we proposed a simplified description of the CC as a finite length FK chain, modeling ligand binding as a register shift in the chain. Using numerical simulations, we found that the signal can reach the other end of the chain only in the presence of thermal fluctuations. Importantly, we determine the range of parameters that enable the intramolecular communication within the experimentally estimated time; those can be used to predict the elastic parameters and the hydrophobic interaction strength of the EEA1 CC.

Experimentally it has been found that the flexibility transition occurs only in the presence of discontinuities. Within our model, the discontinuities can be considered as defects in the FK potential, which enable a wider fluctuation of FK monomers. Therefore, we can interpret the experimental finding as follows: in the presence of discontinuities, the binding of Rab5 can trigger a significant distortion of EEA1, which is revealed as a reduced bending stiffness. In the mutant EEA1 (without discontinuities), however, the signal can propagate through the whole chain while the chain remains stable. Our results indicate that as EEA1 remains in the vicinity of the critical point between stable and unstable state, it can dramatically change its global flexibility upon being trig-

gered by the ligand. Including the discontinuities in our model is a future direction of research.

Planck Society.

## ACKNOWLEDGMENTS

We acknowledge helpful discussions with M. Zerial and D. Murray. This research was supported by the Max

- 
- [1] A. N. Lupas and J. Bassler, Coiled coils—a model system for the 21st century, *Trends in biochemical sciences* **42**, 130 (2017).
- [2] F.-T. Mu, J. M. Callaghan, O. Steele-Mortimer, H. Stenmark, R. G. Parton, P. L. Campbell, J. McCluskey, J.-P. Yeo, E. P. C. Tock, and B.-H. Toh, EEA1, an early endosome-associated protein, *J. Biol. Chem.* **270**, 13503 (1995).
- [3] D. H. Murray, M. Jahnel, J. Lauer, M. J. Avellaneda, N. Brouilly, A. Cezanne, H. Morales-Navarrete, E. D. Perini, C. Ferguson, A. N. Lupas, Y. Kalaidzidis, R. G. Parton, S. W. Grill, and M. Zerial, An endosomal tether undergoes an entropic collapse to bring vesicles together, *Nature* **537**, 107 (2016).
- [4] T. M. Witkos and M. Lowe, The golgin family of coiled-coil tethering proteins, *Front. Cell. Dev. Biol.* **3**, 1 (2016).
- [5] C. A. Laughton, B. F. Luisi, J. V. Pratap, and C. R. Calladine, A potential molecular switch in an alpha-helical coiled-coil, *Proteins* **70**, 25 (2007).
- [6] A. K. Gillingham and S. Munro, Long coiled-coil proteins and membrane traffic, *BBA-Mol. Cell Res.* **1641**, 7185 (2003).
- [7] A. K. Gillingham, At the ends of their tethers! how coiled-coil proteins capture vesicles at the golgi, *Biochem. Soc. Trans.* **46**, 43 (2018).
- [8] J. Callaghan, A. Simonsen, J.-M. Gaullier, B.-H. Toh, and H. Stenmark, The endosome fusion regulator Early-Endosomal Autoantigen 1 (EEA1) is a dimer, *Biochem. J.* **338**, 539 (1999).
- [9] C. Cohen and D. A. D. Parry,  $\alpha$ -helical coiled coils a widespread motif in proteins, *Trends Biochem. Sci.* **11**, 245 (1986).
- [10] Y. I. Frenkel and T. Kontorova, On the theory of plastic deformation and twinning, *Sov. Phys. USSR* **1**, 137 (1939).
- [11] V. Alva, S.-Z. Nam, J. Soeding, and A. N. Lupas, The MPI bioinformatics toolkit as an integrative platform for advanced protein sequence and structure analysis, *Nucleic Acids Res.* **44**, 410 (2016).
- [12] O. M. Braun and Y. S. Kivshar, *The Frenkel-Kontorova model*, Springer-Verlag (1987).
- [13] A. Ward, F. Hilitki, W. Schwenger, D. Welch, A. W. C. Lau, V. Vitelli, L. Mahadevan, and Z. Dogic, Solid friction between soft filaments, *Nat. Mater.* **14**, 583 (2015).
- [14] A. Vanossi, N. Manini, and E. Tosatti, Static and dynamic friction in sliding colloidal monolayers, *Proc. Natl. Acad. Sci. USA* **109**, 16429 (2012).
- [15] A. K. Sieradzan, A. Niemi, and X. Peng, Peierls-Nabarro barrier and protein loop propagation, *Phys. Rev. E* **90**, 062717 (2014).
- [16] I. M. Kulić and H. Schiessel, Chromatin dynamics: nucleosomes go mobile through twist defects, *Phys. Rev. Lett.* **91**, 148103 (2003).
- [17] A. Sulaiman, F. P. Zen, H. Alatas, and L. T. Handoko, Dynamics of DNA breathing in the Peyrard-Bishop model with damping and external force, *Physica D* **241**, 1640 (2012).
- [18] A. Savitzky and M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* **38**, 1627 (1964).
- [19] K. Hirose and L. A. Amos, *Handbook of dynein*, CRC Press Book (2012).
- [20] T. Kon, K. Imamula, A. J. Roberts, R. Ohkura, P. J. Knight, I. R. Gibbons, S. A. Burgess, and K. Sutoh, Helix sliding in the stalk coiled coil of dynein couples at-pase and microtubule binding, *Nat. Struct. Mol. Biol.* **16**, 325333 (2009).
- [21] E. E. Matthews, M. Zoonens, and D. M. Engelman, Dynamic helix interactions in transmembrane signaling, *Cell* **127**, 447 (2006).
- [22] N. R. Quintero, Á. Sánchez, and F. G. Mertens, Overdamped sine-Gordon kink in a thermal bath, *Phys. Rev. E* **60**, 222 (1999).
- [23] G. Costantini and F. Marchesoni, Threshold diffusion in a tilted washboard potential, *EPL (Europhysics Letters)* **48**, 491 (1999).
- [24] F. H. C. Crick, Is alpha-keratin a coiled-coil?, *Nature* **4334**, 882 (1952).
- [25] L. Truebestein and T. A. Leonard, Coiled-coils: the long and short of it, *Bioessays* **38**, 903 (2016).
- [26] D. D. Root, V. K. Yadavalli, J. G. Forbes, and K. Wang, Coiled-coil nanomechanics and uncoiling and unfolding of the superhelix and alpha-helices of myosin, *Biophys. J.* **90**, 2852 (2006).
- [27] O. N. Yogurtcu, C. Wolgemuth, and S. X. Sun, Mechanical response and conformational amplification in alpha-helical coiled coils, *Bioph. J.* **99**, 3895 (2010).
- [28] M. A. Lantz, S. P. Jarvis, H. Tokumoto, T. Martinsky, T. Kusumi, C. Nakamura, and J. Miyake, Stretching the  $\alpha$ -helix: a direct measure of the hydrogen-bond energy of a single-peptide molecule, *Chem. Phys. Lett.* **315**, 61 (1999).
- [29] IAPWS, *Iapws industrial formulation 2008 for the thermodynamic properties of water and steam*, International Steam Tables. Springer, Berlin, Heidelberg (2008).
- [30] R. Wolfenden, L. Andersson, P. M. Cullis, and C. C. B. Southgate, Affinities of amino acid side chains for solvent water, *Biochemistry* **20**, 849 (1981).
- [31] G. König, S. Bruckner, and S. Boresch, Absolute hydration free energies of blocked amino acids: Implications for protein solvation and stability, *Biophys. J.* **104**, 453 (2013).

- [32] A. Nicholls, D. L. Mobley, J. P. Guthrie, J. D. Chodera, C. I. Bayly, M. D. Cooper, and V. S. Pande, Predicting small-molecule solvation free energies: An informal blind test for computational chemistry, *J. Med. Chem.* **51**, 453 (2008).
- [33] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Böckmann, S. Ferro, E. Gasteiger, H. Huang, R. López, M. Magrane, M. J. Martín, D. A. Natale, C. O'Donovan, N. Redaschi, and L.-S. L. Yeh, Uniprot: the universal protein knowledgebase, *Nucleic Acids Res.* **45** (2017).
- [34] D. L. Nelson and M. M. Cox, *Principles of biochemistry*, Lehninger (7th ed.) (2017).

