bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Energy-dependent protein folding: modeling how a protein-folding machine may work

Harutyun K. Sahakyan[1] , Karen B. Nazaryan[1], Arcady R. Mushegian[2,3], and Irina N. Sorokina[4,*]

[1] Laboratory of Computational Modeling of Biological Processes, Institute of Molecular Biology, Yerevan, Armenia; [2] Division of Molecular and Cellular Biosciences, National Science Foundation, Alexandria, USA; [3] Clare Hall College, University of Cambridge, Cambridge, United Kingdom; [4] Strenic LLC, McLean, USA

* corresponding author.

Email: stenicbio@gmail.com

ORCID IDs: 0000-0003-3750-8118 (HKS); 0000-0002-5231-3928 (KBN); 0000-0002-6809-9225 (ARM); 0000-0003-1531-1832 (INS)

Keywords

Energy-dependent protein folding; co-translational protein folding; nascent peptide rotation; peptide backbone manipulation; peptide backbone rotation; protein-folding machine

Author Contributions

INS and HKS designed research. HKS performed research. HKS, KBN, ARM and INS analyzed the data and wrote the paper.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Proteins fold robustly and reproducibly in vivo, but many cannot fold in vitro in isolation from cellular components. The pathways to proteins' native conformations, either in vitro or in vivo, remain largely unknown. It is possible that the slow progress in recapitulating protein folding pathways in silico is due not to a lack of computational power but to fundamental deficiencies in our understanding of folding as it occurs in nature. We consider the possibility that protein folding in living cells may not be driven solely by the decrease in Gibbs free energy. We propose that protein folding in vivo should be modeled as an active energy-dependent process and that the mechanism of action of such a protein-folding machine might include direct manipulation of the peptide backbone. Considering the rotating motion of the tRNA 3'-end in the peptidyl transferase center of the ribosome, we hypothesized that this motion might introduce rotation to the nascent peptide and influence the peptide's folding pathway. We tested whether the formation of protein native conformations could be facilitated by the application of mechanical force to rotate one part of the folding polypeptide while simultaneously restricting rotation of another part of the peptide. We ran molecular dynamics simulations augmented by the application of torsion to the peptide backbones. The directional rotation of the C-terminal amino acid, with the simultaneous limitation of the N-terminal amino acid movements, indeed facilitated the formation of native structures in five diverse alpha-helical peptides. These simulations demonstrate the feasibility of a protein-folding machine.

Significance Statement

We show that in silico formation of the native helical structure in proteins can be substantially facilitated by simulating the application of mechanical rotation to the peptide backbone. When the backbone is manipulated by simultaneous directional rotation of one terminus and movement restriction of the other terminus, the peptide no longer behaves as a freely jointed chain, and its folding behavior changes dramatically. These results open a previously unexplored avenue in modeling and understanding the mechanism of protein folding that, in our model, is an intrinsically non-equilibrium process.

Main Text

Introduction

Once they are synthesized in a living cell, the majority of proteins rapidly attain their distinctive biologically active three-dimensional structures, called native conformations. These conformations are robustly achieved in vivo via a folding process that involves interactions of the folding chain with molecular chaperones and other maturation factors. The folding process often cannot be reproduced in vitro, in the absence of chaperones and other cellular components (1-4). However, some small proteins fold spontaneously in vitro in the absence of any other macromolecules (5).

What exactly happens during the folding of a linear polypeptide chain into a native conformation either in vivo or in vitro remains largely unknown. Despite decades of intense laboratory research, theory development and computer simulations, we still cannot recapitulate complete folding trajectories in silico, except for those of a few relatively short polypeptides (6). Knowledge of the intermediates in the folding pathways and the mechanisms that enable them is essential for determining the points of intervention at which folding and misfolding processes can be altered.

The painfully slow progress in our ability to fold in silico all but the shortest polypeptides could be due to the sheer complexity of the system: the number of possible conformations of a polypeptide chain, and the number of interactions between the atoms of all amino acid residues within the polypeptide itself and with the surrounding solvent, are so astronomically high that the existing computational power is not yet sufficient, and might never become sufficient, to capture the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

folding trajectories for longer proteins (7). It is also possible, however, that there are fundamental deficiencies in our understanding of folding as it occurs in nature, and progress in recapitulating protein folding pathways requires a more realistic physical model of folding than the one we have been relying upon.

The current dominant model of protein folding was prompted by early observations that some small proteins are able to fold in vitro into their native conformations spontaneously, in isolation from other proteins or cellular components (reviewed in (5)). These observations gave rise to the thermodynamic hypothesis of protein folding (5, 8), which in turn led to the development of the physical model that describes protein folding as a thermodynamically favorable, unassisted process. In a more recent, refined form, this model includes the description of a rugged funnel-shaped energy landscape, in which the various unfolded, unstructured conformations occupy the high-free-energy brim of the funnel (9-12). As the polypeptide chains fold, they sample conformations with progressively decreasing Gibbs free energy until they reach the native conformation, which is presumed to occupy the global thermodynamic minimum at the bottom of the funnel. The sampling of conformations during the folding process is assumed to occur via random thermal motions (13). The driving force of protein folding is assumed to be the decrease in free energy to the global minimum.

In summary, the current general physical model of protein folding describes a process that occurs in a closed system in the absence of external sources of energy. It assumes that folding starts from a random, unstructured conformation and proceeds unassisted, with no apparent requirement for the folding chain to interact with other proteins or macromolecular cellular components. This model describes an extremely artificial process that is only likely to occur in vitro and has little resemblance to what takes place during the folding of all proteins in the living cell.

In nature, folding of the majority of proteins occurs in the environment of a living cell, which is an open system with a constant flow of energy and shifting chemical composition. In a cell, a polypeptide starts folding while it is still being synthesized on a ribosome, where it occupies a tight space that allows it to adopt only a limited set of conformations. The nascent peptide emerges into a crowded, viscous environment outside of the ribosomal tunnel and interacts with multiple proteins, including chaperones, and with other cellular components, at all stages of folding. In the course of peptide synthesis and co-translational folding, a large amount of energy is released by GTP hydrolysis. This energy is not required for the formation of peptide bonds (14), but may be spent, at least partially, on various motions and adjustments of the ribosomal components, directly affecting the folding environment of the nascent peptide (15−17). It is difficult to escape the conclusion that protein folding in vivo must be described by a physical model that takes into account the interactions of a folding polypeptide chain with its complex dynamic cellular environment.

We have recently proposed that a more realistic physical model of protein folding might be built on the assumption that protein folding in vivo is an active, energy-dependent process. In this alternative model, proteins that are not able to fold spontaneously must rely on additional external forces to achieve native conformations (18). We hypothesized that the mechanism of action of such a protein folding machine might include direct mechanical manipulation of the peptide backbone by the concerted actions of the ribosome and chaperone complexes (19,20). During translation in the peptidyl transferase center of the ribosome, the 3' terminus of the tRNA in the A-site swings by nearly 180 degrees in every elongation cycle (21,22). We hypothesized that this motion might lead to the rotation of the C-terminus of the nascent peptide. Simultaneously, the movements of the N-terminal regions of the nascent peptides may be restricted, first, by occlusions in the ribosome exit tunnel and then by steric capture mediated by the ribosome-associated "nascent chain welcoming committee", such as the trigger factor in bacteria and the nascent polypeptide-associated complex in archaea and eukaryotes (20). As a result the folding polypeptide may experience transient strained conformations with elevated free energy (18).

As the first step in exploring the feasibility of a protein folding machine capable of facilitating the attainment of native structure by mechanical manipulation of the peptide backbone, we performed

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

molecular dynamics simulations augmented by application of torsion to the peptide backbones. During the simulations, the C-termini of various polypeptides were mechanically rotated either clockwise or counterclockwise, while the motions of their N-termini were restricted. We compared the trajectories of both types of simulations with the folding of the same peptides without the application of torque. In our experiments, directional rotation of the C-terminal amino acids with simultaneous limitation of the movements of the N-termini indeed facilitated the formation of native structures in five diverse alpha-helical peptides.

Results

We performed atomistic molecular dynamics simulations to study peptide folding under conditions when, throughout the simulation, an external mechanical torque was applied to the C-terminal amino acid of a peptide and the motions of the N-terminal amino acid were restrained (Figure 1). We compared the folding trajectories of the peptide to which a mechanical force was applied to rotate the C-terminal amino acid in one of the two possible directions – either clockwise as in Figure 1, or counterclockwise – with the trajectories for the same peptide which was allowed to fold without any motion restriction or application of any mechanical force (referred to as "unassisted folding" below). As an additional control, we ran a fourth type of simulation, where motion restraints were applied to both ends of each peptide but the torque was omitted. The details of the simulations are described in the Methods section. Each of the four types of simulations were repeated three times, giving 12 simulations for each peptide.

The experiments were run on five peptides that are known to adopt alpha-helical conformations in their folded form (Table 1). Two of these, P1 and P2, have been designed de novo, and the other three, P3-P5, are parts of naturally occurring proteins. The folding of the peptides was monitored by calculating the root mean square deviation (RMSD) distance of the peptide backbone from the native structure of the same fragment determined by X-ray crystallography (peptides P2-P5), or computed ab initio (peptide P1). The results of the simulations for each peptide when folded unassisted in the standard force field, and when an external torque force was added to the field, are presented in Table 2 and Figure 2. All folding trajectories are available at https://zenodo.org/record/4008419 .

Within our simulation lengths, we observed the completion of unassisted folding into the native-like alpha-helical structure only in some runs for one peptide, P4, which represents the third helix and preceding loop in the villin headpiece domain HP35. Other peptides remained essentially unfolded throughout the 500-1500-nanosecond runs. The peptides also failed to fold when their ends were restricted in mobility but torque was not applied (Table 2). In contrast, when the external torsion force was applied to the C-termini of the peptides in the clockwise direction, as described in Methods and shown in Figure 1, peptides P1-P4 all folded into alpha-helical structures and were brought within 0.2 nm RMSD from their native structures in every run, typically within the first 100-200 ns of simulation. These peptides stayed in the native or nearly-native conformations for the remainder of the experiments. Peptide P5 was a special case; similarly to P1-P4, it adopted a compact conformation early in the experiments, but remained only partially folded for the duration of all runs (Figure 2).

For all five peptides, folding was observed when the rotation force was applied to the C-terminal amino acid in the clockwise direction (Figure 1). In contrast, the torque applied to the C-terminus counterclockwise with the same force constant did not facilitate folding of P1-P3 and P5, and may have inhibited folding of P4 (Figure 2).

Discussion

To test the idea that inclusion of external forces can improve modeling of protein folding pathways in silico, we performed molecular dynamics simulations in which a standard force field was augmented by the application of external mechanical forces to the polypeptide backbone. We compared these simulations to control runs without any additional external forces. The directional rotation of the C-terminal amino acid with simultaneous restriction of the movements of the N-

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

terminal amino acid facilitated the formation of native structures in five diverse alpha-helical peptides, confirming that such constraints can have significant consequences for folding dynamics. Strikingly, application of mechanical force accelerated the folding of P4, a fragment of an on-pathway folding intermediate of the well-studied villin headpiece domain HP35, which is one of the fastest-folding protein domains known (6,23,24). The order-of-magnitude increase in the rate of P4 folding that was achieved in our experiments seems to suggest that the postulated "physical limit of folding" of HP35 as a whole (23,25) could be overcome by a protein-folding machine. The other four peptides in our experiments likewise attained their alpha-helical structure in the presence of the rotating force, but did not reach their native conformations when allowed to fold unassisted, even though we ran the control unassisted simulations for ~10 times longer than the simulations that included the application of the external force (Table 2). Some of those peptides might take a very long time to reach their native conformations without application of an external force, whereas others might never fold unassisted, if their unfolded states are more stable than the folded conformations.

These results are in line with our protein folding machine hypothesis (18). They also support a hypothetical mechanism through which the machine would directly alter the conformations of proteins by applying mechanical force to the peptide backbone (19,20). The feasibility of such a mechanism, however, is dependent on whether the torsion applied at one point of a peptide would propagate through the rest of the peptide chain and affect the movements of the distal parts of the peptide. The peptide backbone is often viewed as a freely jointed chain, due to the 360-degrees rotation ability around the phi- and psi-bonds within each amino acid (26). If the peptides in our simulations were to behave as freely jointed chains, the rotation of a single amino acid at the end of the peptide would not have any appreciable effect on the motions of the rest of the peptide. However, if a mechanical torque were applied to a peptide while it was being folded in a viscous crowded environment (e.g., co-translationally in a living cell), we predicted that the free rotation of the phi- and psi-bonds in the peptide backbone would be hindered enough that escape from the forbidden sections of the Ramachandran plots would become difficult for many residues, and as a result, the entire peptide backbone may experience transient strained conformations. Although our simulation could not account for all the details of the protein folding environment in vivo, we were able to devise a set of conditions under which the peptide indeed did not behave as a freely jointed chain. When a force was applied to a single amino acid residue, and the motion of just one other residue at least 15 amino acids apart was restricted, the folding trajectory of the entire peptide was affected dramatically, leading to the rapid attainment of the native helical conformation. Some of the steric hindrances that make this rapid folding possible involve amino acid side chains, and therefore the effect might be sequence-specific. For example, glycine residues are more likely to experience the full 360-degree rotation around the phi- and psi-bonds, relieving the strain in the main chain; this might explain why P5, a peptide with an internal glycine, first acquired and then partially lost its folded conformation in our experiments (Figure 2).

It remains unclear whether our simulation captures the main features of the folding process as it occurs in nature. For example, one of the parameters that differs between our simulations and real co-translational protein folding process is their characteristic times. The rotation of the backbone in our system occurs at the submicrosecond time scale, whereas the addition of amino acids to the nascent peptide is much slower, on the order of subseconds (27-29). Molecular dynamics simulations have been known to model, at a fast scale, the essential parts of the molecular processes that are much slower when observed with bulk kinetics or single-molecule methods (16,24), but the effect of the rotation rate on the peptide folding trajectory remains to be investigated.

The key feature of the hypothetical mechanism of co-translational protein folding that we simulated is the directional rotation of the peptide backbone. As discussed above, the 3' terminus of the tRNA in the A-site of the ribosome peptidyl transferase center turns by nearly 180 degrees in every translation elongation cycle. Only a 45-degree swing is necessary to achieve the proper stereochemistry of the peptide bond formation (30); the function of the remaining portion of the turn is unknown, and we have hypothesized that it may be needed to facilitate co-translational folding (19,20). It is notable, however, that the tRNA appears to turn in the counterclockwise

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

direction when looking from the C-terminus of the nascent peptide (21,22). In contrast, folding of all peptides into the right-handed alpha-helices in our experiments took place only with clockwise rotation of the C-termini (Figures 1 and 2). It remains to be determined what, exactly, happens to the nascent peptide in the peptidyl transferase center and in the ribosome exit tunnel. The nascent peptide might be rotated counterclockwise (in the direction of the tRNA swing), or clockwise (as a result of a gear-like interaction with the tunnel walls), or might not be rotated at all but rearranged in a more complex way, being subject to pushing and pulling forces as well as interactions with the exit tunnel walls and other components of the ribosomal complex.

Regardless of whether the peptide torsion mechanism operates during co-translational folding on the ribosome in vivo, we demonstrate that it is possible to facilitate protein folding under conditions when an external mechanical force is applied to the peptide backbone. Importantly, we show that the peptide does not always behave as a freely jointed chain, opening the possibility that in vivo the peptide backbone can be manipulated into conformations that cannot be reached without assistance because they are either thermodynamically unstable or kinetically inaccessible. The results of our simulations thus demonstrate the feasibility of a protein folding machine. Some recently published results, including studies of the role of the exit tunnel in nascent chain folding (31-35) and of direct coupling between ATP hydrolysis and protein refolding by the chaperones of the HSP70 family (36-38), may be also interpreted as evidence of protein folding in vivo being an active process.

The notion of an active, energy-dependent protein folding mechanisms in vivo is better compatible with the current understanding of evolution than the generally accepted, standard thermodynamic hypothesis of protein folding. Although it is accepted that the ability of proteins to attain their native conformations must have evolved by natural selection of sequences that fold quickly and correctly ("evolution solved the protein folding problem" (39)), models of unassisted folding sidestep the fact that ribosomes and translation factors are among the oldest molecular machines shared by all extant cellular life (40), and were present during much of the evolution of proteins and of their folding pathways. The evolutionary optimization of the tempo and mode of protein folding, for at least 3.5 billion years of biological evolution, has taken place not in dilute solutions of isolated proteins, but in a dynamic environment of living cells with their constant flow of matter and energy. Thus, the ability of any present-day protein to fold in isolation and without assistance is likely to be either an incidental or derived property, not shared by most other proteins. Realistic computational modeling of protein folding must therefore take into account the presence of a multitude of external forces. Further studies should attempt to more closely recreate the conditions of protein folding in vivo.

Materials and Methods

The initial stretched structures of peptides with four additional alanine residues, two at each end, were generated using ICM software (41). These alanines were attached as handles to which the rotation or restraint could be applied directly without affecting the sequence whose folding was investigated, and were not considered in the RMSD calculations. We aligned a peptide along the X-axis and solvated it in a dodecahedron box in the case of the simulations of unassisted folding and triclinic box in all other cases, with minimum distance of 1.5 nm between a peptide and the simulations box. Potassium and sodium ions were added to neutralize the charges in the system. The system was then minimized with the steepest descent algorithm, equilibrated for 100 ps in the NVT ensemble using V-rescale thermostat (42) for temperature coupling, and continued in the NPT ensembles for 1 ns using V-rescale thermostat and Berendsen barostat (43). After the equilibration, we kept temperature and pressure constant at 300 K and 1 bar respectively, using Nose−Hoover thermostat (44, 45) and isotropic Parinello−Rahman barostat (46). The properties of simulation boxes for each peptide are summarized in Additional File 1.

For all simulations, we used the ff14SB force field (47) with the TIP3P water model (48) and ion parameters modified by Joung and Cheatham (49). Electrostatic interactions were calculated using particle-mesh Ewald (PME) summation (50) with a Fourier grid spacing of 0.135 nm. For

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

non-bonded Coulomb and Lennard-Jones interactions, 1 nm cutoff was used. We constrained the hydrogen bonds with the LINCS algorithm (51) and used a 2-fs integration time step.

To exert an external mechanical torque to the C-termini of the peptides, we adopted the enforced rotation method, originally designed to study rearrangements during the rotation of a folded protein within the F_1-ATPase assembly (52), implemented in the GROMACS molecular dynamics package. To this end, we restrained the positions of the O and N atoms of the C-terminal alanine to keep it aligned with the X-axis, about which the rotation was applied. The restraints with a force constant of 10000 kJ/mol*nm^2 were applied only for the YZ-plane, so the C-terminal amino acid could move along the X-axis. In addition, we restrained the O atom of the C-terminal amino acid in the X direction with a force constant of 5 kJ/mol*nm^2 and N and Cα atoms of the N-terminal alanine with a force constant 10000 kJ/mol*nm^2 in all directions. The C-terminal amino acid was rotated using a flexible axis approach (Vflex2) with a reference rotation rate of 60 degrees/ps and a force constant of 1500 kJ/mol*nm^2.

The GROMACS package version 2020.2 (53) was used for all simulations and trajectory analyses. The simulations were carried out on CUDA-enabled GPUs with Turing architecture, running Ubuntu Linux. For visualization of protein structures and trajectories, the programs ICM-Pro 3.9 (41) and VMD 1.9.3 (44) were used.

Acknowledgments and Funding

We are grateful to many colleagues, most of all Dr. Yuri Wolf, for useful discussions. We thank Dr. Alexandra Mushegian for critical reading of this manuscript. INS is grateful to an anonymous angel investor for their support of Strenic, LLC. HKS and INS were funded by Strenic, LLC. HKS and KBN are supported by the National Academy of Sciences of the Republic of Armenia and Foundation Armenia (Switzerland). ARM is a Program Director at the National Science Foundation (NSF), the agency of the U.S. Government; his work on this project and stay at Clare Hall College, University of Cambridge, were supported by the NSF Long-Term Professional Development Program and Clare Hall Visiting Fellows Program, but the statements and opinions expressed here are made in a personal capacity and do not constitute an endorsement by the NSF or the Government of the United States.

References

1. Martin, J., Langer, T., Boteva, R., Schramel, A., Horwich, A. L., & Hartl, F. U. (1991). Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. *Nature*, *352*(6330), 36–42.

2. Chow, D., Srikakulam, R., Chen, Y., Winkelmann, D. A. (2002). Folding of the striated muscle myosin motor domain. *J. Biol. Chem*, *277*(39), 36799–36807.

3. Balchin, D., Miličić, G., Strauss, M., Hayer-Hartl, M., & Hartl, F. U. (2018). Pathway of Actin Folding Directed by the Eukaryotic Chaperonin TRiC. *Cell*, *174*(6), 1507–1521.e16.

4. Bauer, D., Meinhold, S., Jakob, R. P., Stigler, J., Merkel, U., Maier, T., Rief, M., & Žoldák, G. (2018). A folding nucleus and minimal ATP binding domain of Hsp70 identified by single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. USA, 115*(18), 4666–4671.

5. C. B. Anfinsen, Principles that govern the folding of protein chains. *Science* **181**, 223–30 (1973).

6. K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding proteins fold. *Science* **334**, 517–520 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

7. C. Levinthal, How to fold graciously. *Mössbauer Spectroscopy in Biological Systems Proceedings. Univ. of Illinois Bulletin* **67**, 22–24 (1969).

8. E. Haber, C. B. Anfinsen, Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* **237**, 1839–1844 (1962).

9. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).

10. J. N. Onuchic, Z. Luthey-Schulten, P. G. Wolynes, Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* **48**, 545–600 (1997).

11. K. A. Dill, S. B. Ozkan, M. S. Shell, W. T. R, The Protein Folding Problem. *Annual Rev Biophysics* **37**, 289–316 (2008).

12. G. C. Rollins, K. A. Dill, General Mechanism of Two-State Protein Folding Kinetics. *J Amer Chem Soc* **136**, 11420–7 (2014).

13. K. A. Dill, J. I. MacCallum, The protein-folding problem, 50 years on. *Science* **338**, 1042–6 (2012).

14. A. S. Spirin, *Ribosomes* (Kluwer Academic/Plenum Publishers;, 1999).

15. V. I. Lim, J. F. Curran, M. B. Garber, Ribosomal Elongation Cycle: Energetic, Kinetic and Stereochemical Aspects. *Journal of Molecular Biology* **351**, 470–480 (2005).

16. L. V. Bock, Blau, C., Schröder, G. F., Davydov, I. I., Fischer, N., Stark, H., Rodnina, M. V., Vaiana, A. C., & Grubmüller, H. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat Struct Mol Biol* **20**, 1390–1396 (2013).

17. C. Maracci, M. V. Rodnina, Translational GTPases. *Biopolymers* **105**, 463–75 (2016).

18. I. Sorokina, A. Mushegian, Modeling protein folding in vivo. *Biol. Direct* **13**, 13 (2018).

19. I. Sorokina, A. Mushegian, The role of the backbone torsion in protein folding. *Biol. Direct* **11**, 64 (2016).

20. I. Sorokina, A. Mushegian, Rotational restriction of nascent peptides as an essential element of co-translational protein folding: possible molecular players and structural consequences. *Biol. Direct* **12**, 14 (2017).

21. A. Bashan, I. Agmon, R. Zarivach, F. Schluenzen, J. Harms, R. Berisio, H. Bartels, F. Franceschi, T. Auerbach, H.A. Hansen, E. Kossoy, M. Kessler, M., A. Yonath, Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol Cell* **11**, 91–102 (2003).

22. I. Agmon, A. Bashan, R. Zarivach, A. Yonath, Symmetry at the active site of the ribosome: structural and functional implications. *Biol Chem* **386**, 833–44 (2005).

23. D. L. Ensign, P. M. Kasson, V. S. Pande, Heterogeneity Even at the Speed Limit of Folding: Large-scale Molecular Dynamics Study of a Fast-folding Variant of the Villin Headpiece. *Journal of Molecular Biology* **374**, 806–816 (2007).

24. E. Wang, P. Tao, J. Wang, Y. Xiao, A novel folding pathway of the villin headpiece subdomain HP35. *Phys. Chem. Chem. Phys.* **21**, 18219–18226 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. J. Kubelka, J. Hofrichter, W. A. Eaton, The protein folding “speed limit.” *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).

26. P. J. Fleming, G. D. Rose, “Conformational Properties of Unfolded Proteins” in *Protein Science Encyclopedia*, A. R. Fersht, Ed. (Wiley-VCH Verlag GmbH & Co. KGaA, 2008), p. sf06.

27. V. Talkad, E. Schneider, D. Kennell, Evidence for variable rates of ribosome movement in Escherichia coli. *Journal of Molecular Biology* **104**, 299–303 (1976).

28. S.-O. Olofsson, S. O., Boström, K., Carlsson, P., Borén, J., Wettesten, M., Bjursell, G., Wiklund, O., & Bondjers, G.Structure and biosynthesis of apolipoprotein B. *American Heart Journal* **113**, 446–452 (1987).

29. G.-W. Li, D. Burkhardt, C. Gross, J. S. Weissman, Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **157**, 624–635 (2014).

30. A. Gindulyte, A. Bashan, I. Agmon, L. Massa, A. Yonath, J. Karle,The transition state for formation of the peptide bond in the ribosome. *Proc Natl Acad Sci U S A* **103**, 13327–32 (2006).

31. O. B. Nilsson, R. Hedman, J. Marino, S. Wickles, L. Bischoff, M. Johansson, A. Müller-Lucks, F. Trovato, J.D. Puglisi, E.P. O'Brien, R. Beckmann, G. von Heijne, Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep* **12**, 1533–40 (2015).

32. J. A. Farías-Rico, S. K. Goetz, J. Marino, G. von Heijne, Mutational analysis of protein folding inside the ribosome exit tunnel. *FEBS Lett.* **591**, 155–163 (2017).

33. G. Kemp, R. Kudva, A. de la Rosa, G. von Heijne, Force-Profile Analysis of the Cotranslational Folding of HemK and Filamin Domains: Comparison of Biochemical and Biophysical Folding Assays. *J. Mol. Biol.* **431**, 1308–1314 (2019).

34. A. Bitran, W. M. Jacobs, X. Zhai, E. Shakhnovich, Cotranslational folding allows misfolding-prone proteins to circumvent deep kinetic traps. *Proc Natl Acad Sci USA*, 201913207 (2020).

35. P. Tao, E. Wang, Y. Xiao, Pathway regulation mechanism revealed by cotranslational folding of villin headpiece subdomain HP35. *Phys. Rev. E* **101**, 052403 (2020).

36. A. Barducci, P. De Los Rios, Non-equilibrium conformational dynamics in the function of molecular chaperones. *Curr. Opin. Struct. Biol.* **30**, 161–169 (2015).

37. P. Goloubinoff, A. Sassi, B. Fauvet, A. Barducci, P. De Los Rios, Molecular chaperones inject energy from ATP hydrolysis into the nonequilibrium stabilisation of native proteins. *Nature Chem Biol* **14**, 388–395 (2018).

38. S. Assenza, A.S. Sassi, R. Kellner, B. Schuler, P. De Los Rios, A. Barducci, A. Efficient conversion of chemical energy into mechanical work by Hsp70 chaperones. *Elife* **8** (2019).

39. P. G. Wolynes, Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie* **119**, 218–230 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. J. C. Bowman, A. S. Petrov, M. Frenkel-Pinter, P. I. Penev, L. D. Williams, Root of the Tree: The Significance, Evolution, and Origins of the Ribosome. *Chem. Rev.*, acs.chemrev.9b00742 (2020).

41. R. Abagyan, M. Totrov, D. Kuznetsov, ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506 (1994).

42. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101 (2007).

43. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684–3690 (1984).

44. S. Nosé, A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52**, 255–268 (1984).

45. W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).

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

47. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

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

49. I. S. Joung, T. E. Cheatham, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **112**, 9020–9041 (2008).

50. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089–10092 (1993).

51. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463–1472 (1997).

52. C. Kutzner, J. Czub, H. Grubmüller, Keep It Flexible: Driving Macromolecular Rotary Motions in Atomistic Simulations with GROMACS. *J. Chem. Theory Comput.* **7**, 1381–1393 (2011).

53. M. J. Abraham, *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

54. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures and Tables

Peptide	Peptide description	Sequence	Length, amino acids	PDB ID
P1	Peptide Fs (Folded short), designed de novo	AAAA(AAARA)$_3$	19	n/a
P2	First helix of the three-helix bundle, designed de novo	SWAEFKQRLAAIKTR	15	2A3D
P3	Fragment of the tetramerization domain of potassium channel Kv7.1	HLNLMVRIKELQRRLDQSL	19	6UZZ
P4	Loop and third helix of the villin headpiece fragment HP35	PLWLQQHLLKEKGLF	15	2F4K
P5	Fragment of the coiled-coil region of pyrin	KIQKQLEHLKKLRKSGEEQRS	21	4CD4

Table 1. Peptides used for the molecular dynamics simulations in this study.

Peptide	No rotation, no restraints	No rotation, restrained ends	Restrained ends, rotation clockwise	Restrained ends, rotation counterclockwise
P1	500/500 500/500 500/500	500/500 500/500 500/500	22/100 18/100 29/100	500/500 500/500 500/500
P2	1500/1500 1500/1500 1500/1500	1500/1500 1500/1500 1500/1500	160/200 140/200 105/200	1500/1500 1500/1500 1500/1500
P3	1500/1500 1500/1500 1500/1500	1500/1500 1500/1500 1500/1500	105/150 95/150 125/150	1500/1500 1500/1500 1500/1500
P4	1500/1500 545/1500 360/1500	250/1500 1500/1500 1500/1500	130/175 80/175 150/175	1500/1500 1500/1500 1500/1500
P5	1500/1500 1500/1500 1500/1500	1500/1500 1500/1500 1500/1500	230/1500 380/1500 170/1500	1500/1500 1500/1500 1500/1500

Table 2. Peptide folding rates (ns) in the molecular dynamics simulations. The rate was defined as the time spent before reaching the RMSD of 2 nm from the native conformation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.01.277582; this version posted September 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends.

Figure 1. Schematic representation of the energy-dependent peptide folding protocol employed in this study. The force vectors applied to the C- and N-termini of a peptide in the simulation box are shown by black arrows. All force values are in kJ/mol*nm^2. The purple curled arrow indicates the direction of the clockwise rotation of the peptides that resulted in the accelerated productive folding of all peptides to their helical conformations. The restrained groups are shown by green outline.

Figure 2. **Peptide folding in the force field augmented by the application of external rotation forces to the polypeptide backbone, compared to the controls run without additional external forces.** Left side: folding of peptides P1-P5 simulated as indicated in Table 2. In each pane, top three curves (purple, green, and light blue) indicate three independent runs for the same peptide in the standard force field without external torsion, and the bottom three curves (dark blue, orange, and yellow) indicate three runs in the presence of the clockwise (+) rotational force. Right side: In each pane, top three curves (purple, green, and light blue) indicate three independent runs for the same peptide as on the left in the presence of the counterclockwise (-) rotational force, and the bottom three curves indicate three runs in the presence of the clockwise rotational force, i.e., the same three runs as in the corresponding left pane.

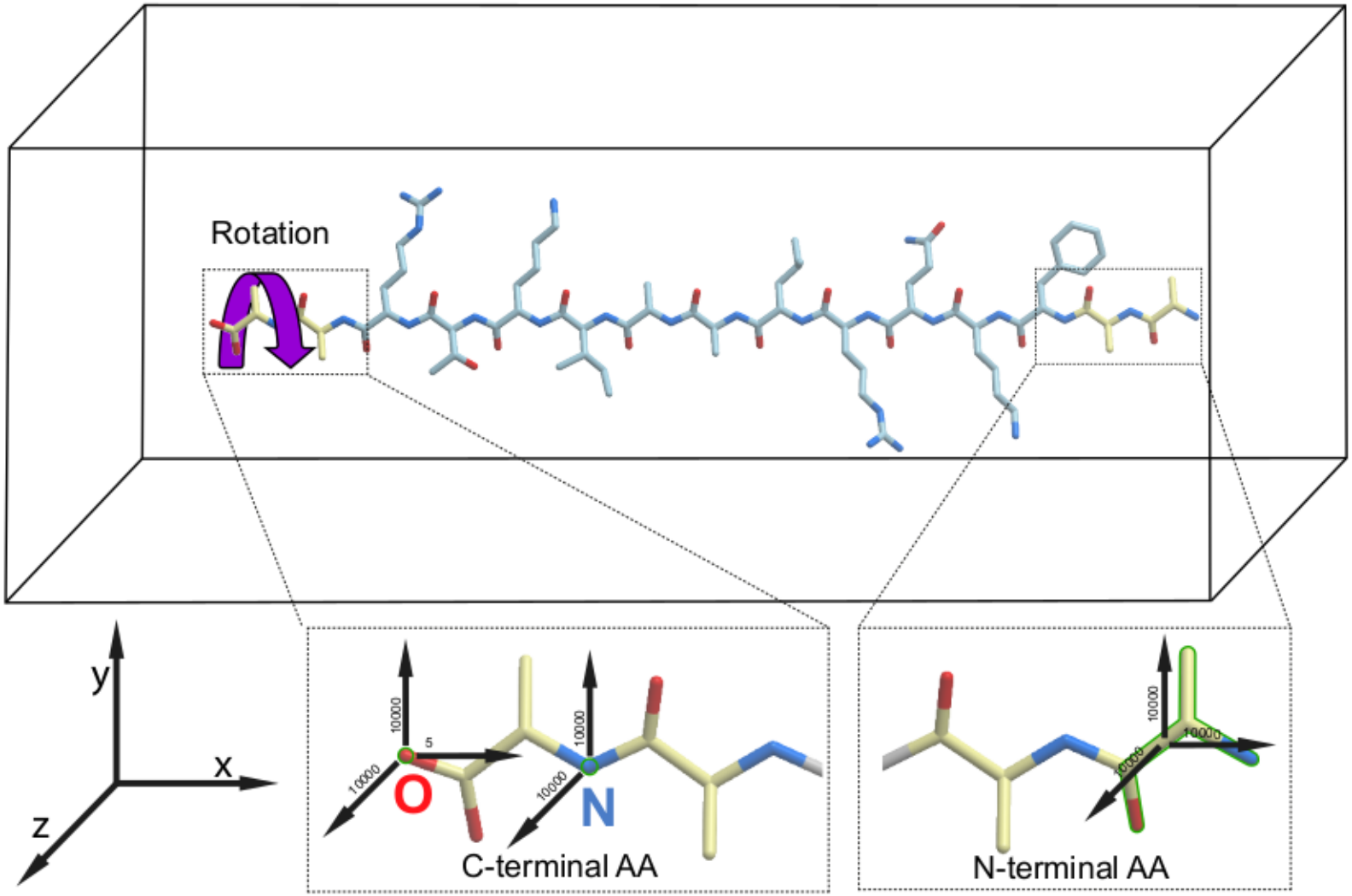
Rotation
y
x
z
10000
5
10000
10000
10000
O
N
C-terminal AA
10000
10000
10000
10000
N-terminal AA

Figure 1.

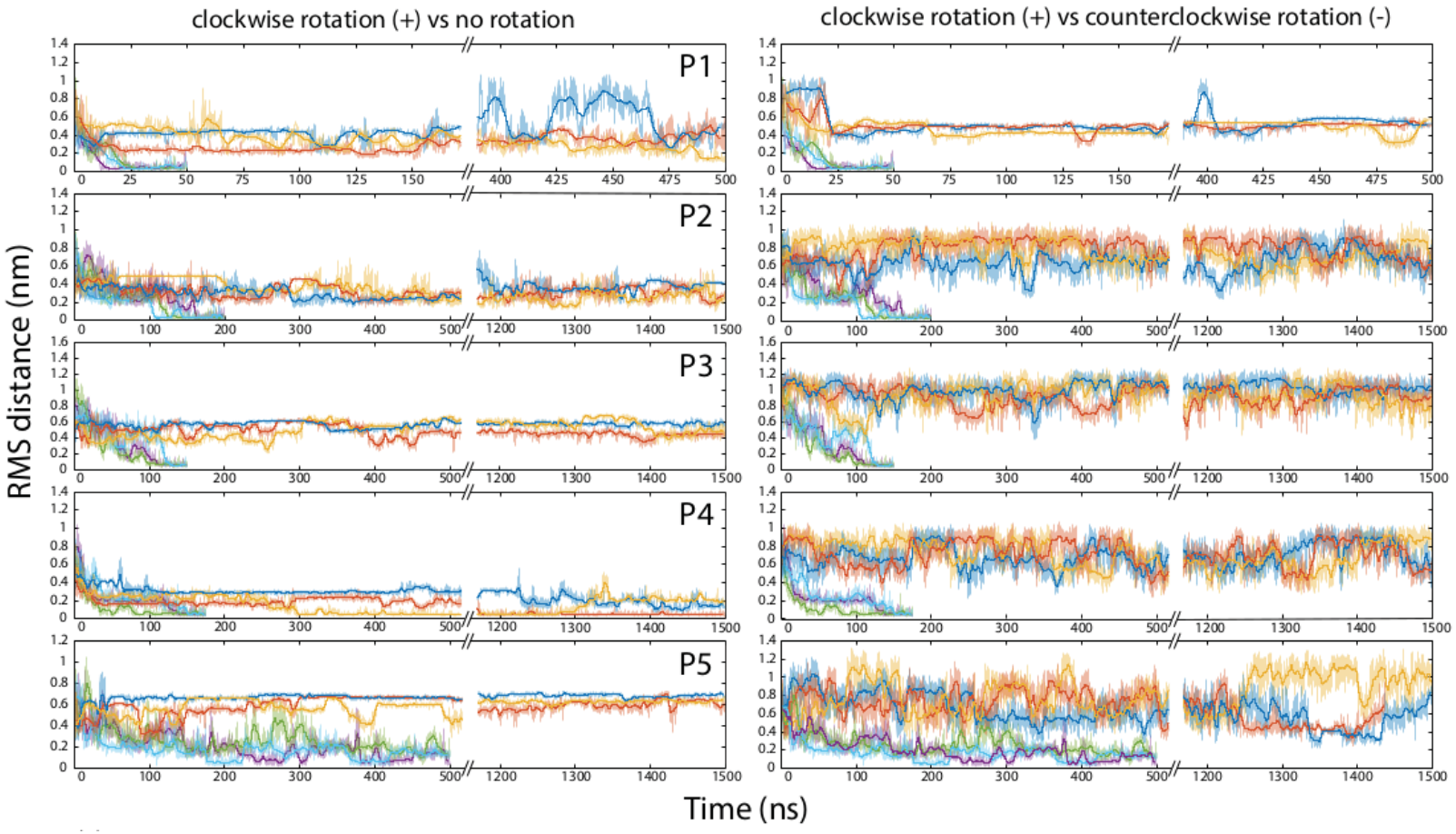
clockwise rotation (+) vs no rotation
clockwise rotation (+) vs counterclockwise rotation (-)
P1
P2
P3
P4
P5
RMS distance (nm)
Time (ns)

Figure 2.