Synonymous mutations can alter protein dimerization through localized interface misfolding involving self-entanglements

Lan Pham Dang[ǂ,1,2], Daniel Allen Nissley[ǂ,†,3], Ian Sitarik[ǂ,3], Quyen Vu Van[4], Yang Jiang[3], Philip To[5], Yingzi Xia[5], Stephen D. Fried[5,6], Mai Suan Li[1,4], Edward P. O'Brien[*3,7,8]

[1] Institute for Computational Sciences and Technology, Ho Chi Minh City, Vietnam
[2] Faculty of Physics and Engineering Physics, VNUHCM-University of Science, 227, Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam
[3] Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA
[4] Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland
[5] Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA
[6] Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218, USA
[7] Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA
[8] Institute for Computational and Data Sciences, Pennsylvania State University, University Park, PA 16802, USA

[ǂ] These authors contributed equally to this research project
[†]Current Affiliation: Department of Statistics, University of Oxford, Oxford OX1 3LB, UK
* to whom correspondence should be addressed: epo2@psu.edu

ABSTRACT

Synonymous mutations in messenger RNAs (mRNAs) can reduce protein-protein binding significantly without changing the protein's amino acid sequence. Here, we use coarse-grain simulations of protein synthesis, post-translational dynamics, and dimerization to understand how synonymous mutations can influence the dimerization of two *E. coli* homodimers, oligoribonuclease and ribonuclease T. We synthesize each protein from its wildtype, fastest- and slowest-translating synonymous mRNAs *in silico* and calculate the ensemble-averaged interaction energy between the resulting dimers. We find synonymous mutations alter oligoribonuclease's dimer properties. Relative to wildtype, the dimer interaction energy becomes 4% and 10% stronger, respectively, when translated from its fastest- and slowest-translating mRNAs. Ribonuclease T dimerization, however, is insensitive to synonymous mutations. The structural and kinetic origin of these changes are misfolded states containing non-covalent lasso-entanglements, many of which structurally perturb the dimer interface, and whose probability of occurrence depends on translation speed. These entangled states are kinetic traps that persist for long time scales. Entanglements cause altered dimerization energies for oligoribonuclease, as there is a large association (odds ratio: 52) between the co-occurrence of non-native self-entanglements and weak-binding dimer conformations. Simulated at all-atom resolution, these entangled structures persist for long timescales, indicating the conclusions are independent of model resolution. Finally, we show that regions of the protein we predict to have changes in entanglement are also structurally perturbed during refolding, as detected by limited-proteolysis mass spectrometry. Thus, non-native changes in entanglement at dimer interfaces is a mechanism through which oligomer structure and stability can be altered.

INTRODUCTION

Oligomerization, the process of assembling multiple macromolecules into dimers and higher-order oligomers, is necessary for a majority of proteins to function[1]. These functional oligomeric assemblies require the correct type, number, conformational state, and orientation of each constituent protein monomer[2,3]. For example, the monomers composing the active tetrameric forms of β-galactosidase[4] and hemoglobin[5] do not function efficiently on their own. An analysis of 452 human enzymes found roughly one-third (141) to be monomeric, one-third to be homodimers (125), and the remaining third to be heterodimers or higher order oligomers[6]. Just as the native structures of proteins represent their minimum free energy structure at equilibrium, thermodynamics is also thought to dictate the structural ensemble of oligomeric complexes. From this thermodynamic perspective, the initial conditions and history associated with a system have no long-term effect on its behavior, meaning that the influence of translation-elongation kinetics should be irrelevant to the structures a dimer adopts.

Contrary to this prediction, experiments have revealed that changes to the speed of protein translation can perturb post-translational oligomerization and protein function over biologically long timescales, indicating a role of kinetics and changes in co-translational processes. For example, when the sub-optimal codon usage in the *frq* gene encoding the FRQ circadian clock protein in *N. crassa* is “optimized” by replacing rare codons with common synonymous codons that tend to be translated faster, it binds 60% less to the WC-2 protein even after controlling for soluble expression level changes. This decrease in affinity effectively abolishes *N. crassa*’s circadian rhythm measured over the course of multiple days[7]. Thus, synonymous mutations can change the structure and function of protein complexes and cause phenotypic changes in organisms.

Recent studies[8,9] have suggested a mechanism by which synonymous mutations can alter monomeric protein enzyme structure and function, and how these changes can persist in the presence of the proteostasis machinery – such as chaperones and the proteasome – that evolved to fix or remove misfolded proteins. These studies indicate that long-lived misfolded states are self-entangled, leading to reduced structure and function. Many of these entangled structures resemble the native state and thus can evade chaperones, avoid aggregation, and fail to be degraded, allowing them to remain soluble but less functional on timescales ranging from seconds to months or longer. The partitioning of nascent proteins into such soluble but self-entangled conformations has the potential to explain how changes to translation kinetics are able to disrupt oligomer formation for long time periods.

Here, we use coarse-grain and all-atom molecular dynamics simulations to understand the structural origin of altered dimerization when synonymous mutations are introduced into a protein’s mRNA template. Because FRQ is an intrinsically disordered protein[10] whose binding interface and structure are unknown, we instead study the dimerization of two globular, cytosolic *E. coli* homodimers - oligoribonuclease and ribonuclease T - after synthesis from their wildtype, fastest-translating synonymous variant, and slowest-translating synonymous variant mRNA sequences. For ribonuclease T, which folds relatively quickly, the speed of translation has no discernible influence on its ability to dimerize. Oligoribonuclease’s dimerization, however, does depend on the mRNA variant from which it is synthesized. We find a molecular origin of this phenomenon, show the results are robust to changes in model resolution, explain why the mechanism we identify is likely to be widespread across the proteome, and find Limited Proteolysis Mass Spectrometry data are consistent with the computationally observed entangled states.

METHODS

Construction of coarse-grain protein and ribosome representations. We employ a previously published Gō-based coarse-grain methodology in which each amino acid is represented by a single interaction site[9,11–13]. Briefly, the potential energy of a configuration in this model is computed by the equation

$$E = \sum_i k_\mathrm{b}(r_i - r_0)^2 + \sum_i \sum_j^4 k_{\varphi,ij}\left[1 + \cos\left(j\varphi_i - \delta_{ij}\right)\right] + \sum_i -\frac{1}{\gamma}\ln\left\{\exp\left[-\gamma\left(k_\alpha(\theta_i - \theta_\alpha)^2 + \varepsilon_\alpha\right)\right] + \exp\left[-\gamma k_\beta\left(\theta_i - \theta_\beta\right)^2\right]\right\} + \sum_{ij} \frac{q_i q_j e^2}{4\pi\varepsilon_0\varepsilon_\mathrm{r} r_{ij}}\exp\left[-\frac{r_{ij}}{l_\mathrm{D}}\right] + \sum_{ij\in\{\mathrm{NC}\}} \varepsilon_{ij}^{\mathrm{NC}}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{ij\notin\{\mathrm{NC}\}} \varepsilon_{ij}^{\mathrm{NN}}\left[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \qquad [1]$$

Eq. 1 calculates the total potential energy (E) of a given conformation as a sum, respectively, of the contributions from C_α–C_α virtual bonds, dihedral angles, bond angles, electrostatic interactions, Lennard-Jones-like native interactions, and repulsive non-native interactions. Bonds are treated using a harmonic potential and dihedral terms are computed as previously described[14]. The bond angle energy is computed using a double-well potential that can adopt angles representative of either α or β structures[15]. Electrostatics are treated using Debye−Hückel theory with a Debye screening length, l_D, of 10 Å and a dielectric of 78.5. Lysine and arginine residues are assigned a charge of $q = +1e$, glutamate and aspartate $q = -1e$, and all other residue types $q = 0$[12]. Native and non-native interactions are computed using the 12-10-6 potential of Karanicolas and Brooks[14]. The minimum potential energy of a native contact is calculated as $\epsilon_{ij}^{\mathrm{NC}} = n_{ij}\,\varepsilon_\mathrm{HB} + \eta\varepsilon_{ij}$ where $\varepsilon_\mathrm{HB} = 0.75$ kcal/mol and ε_{ij} represent the energy contributions arising from hydrogen bonding and van der Waals interactions between residues i and j, respectively, and n_{ij} is the number of hydrogen bonds between residues i and j. η is a scaling factor that multiplicatively increases the values of ε_{ij}, which are initially set based on the Betancourt-Thirumalai pairwise potential[16]. The collision diameter for a native contact between residues i and j, denoted σ_{ij}, is set equal to the distance between the C_α atoms of the corresponding residues in the native-state crystal structure divided by $2^{\frac{1}{6}}$. Values of η were determined based on a previously published training set to reproduce realistic protein domain stabilities (see below)[13]. Interactions between all pairs of residues not in the native contact list are computed with $\varepsilon_{ij}^{\mathrm{NN}} = 1.32 \,\mathrm{x}\, 10^{-4}$ kcal/mol and collision diameters calculated as reported previously[17].

Synthesis simulations were conducted using a previously described protocol[9,11] with a cutout of the ribosome exit tunnel and surface. Briefly, ribosomal RNA is represented with one bead each for the ribose, phosphate, and pyrimidine nucleobases and two beads for purine nucleobases; ribosomal proteins are coarse-grained at C_α-resolution as described for other proteins above[11,12]. The peptidyl-transferase center is placed at the origin of the CHARMM coordinate system, with the positive x-axis pointing down the ribosome exit tunnel towards its opening into the cytosol. All coarse-grain simulations were carried out with a Langevin thermostat set to 310 K, a 15 fs integration time step, and a friction coefficient of 0.050 ps^{-1}.

Parameterization of intra- and inter-monomer protein interactions. We set realistic intra- and inter-monomer energy scales in the coarse-grain models of PDB IDs 1YTA and 2IS3, which represent the homodimeric structures of the *E. coli* proteins oligoribonuclease and ribonuclease T, respectively, by multiplying the $\varepsilon_{ij}^{\mathrm{NC}}$ term in Eq. 1 by a scaling factor η separately for intra- and inter-monomer native Lennard-Jones interactions. Missing heavy atoms and residues in the ribonuclease T structure were reconstructed based on default

CHARMM topology and parameters and then locally minimized as previously described[11] before construction of its coarse-grain representation. A coarse-grain monomer is considered to be reasonably stable if its fraction of native contacts, Q, is greater than the average $\langle Q_{\text{kin}} \rangle = 0.69$ determined from a training set[13] for at least 98% of the simulation frames in each of three 1-μs simulations initiated from the native state reference structure. The minimum value of η that results in a stable model based on this criterion is selected for each monomer and interface. A value of $\eta_{\text{intra}} = 1.359$ was selected by this procedure for all native Lennard-Jones interactions in both oligoribonuclease and ribonuclease T; inter-monomer contacts were scaled by $\eta_{\text{inter}} = 1.507$ and 1.235 for oligoribonuclease and ribonuclease T, respectively.

Construction of mRNA translation schedules for coarse-grain simulations. Wildtype mRNA sequences for oligoribonuclease and ribonuclease T were obtained from NCBI assembly eschColi_K12 using the University of California Santa Cruz microbe table browser (http://microbes.ucsc.edu/). Codon translation rates are taken from the Fluitt–Viljoen[18] model for *E. coli,* rescaled to produce an overall average elongation rate of 20 aa/s, and then further adjusted to account for the accelerated timescale of dynamic processes in our coarse-grain model[13]. When rescaled in this way, the translation times from the Fluitt-Viljoen model have a mean of 12.6 ns or 840,000 integration time steps each 15 fs in duration (see Supplementary Table 7 of ref. 11). Predicted fastest- and slowest-translating synonymous variant mRNAs were generated for oligoribonuclease and ribonuclease T by replacing each codon in their wildtype sequences with the codon predicted by the Fluitt-Viljoen model to be fastest- or slowest-translating synonymous codon, respectively. The average *in silico* translation times for the codons within the wildtype, fast-translating mutant, and slow-translating mutant sequences of oligoribonuclease are 10.6, 7.0, and 20.8 ns, respectively. Average translation times for the ribonuclease T wildtype, fast-translating mutant, and slow-translating mutant mRNAs are 12.7, 7.2, and 22.4 ns, respectively.

Coarse-grain simulations of monomer synthesis, ejection, and post-translational dynamics. One hundred statistically independent continuous synthesis simulations were performed as previously described for each monomer of each protein and for each mRNA sequence (*i.e.*, 200 trajectories per protein, 100 for Monomer A and 100 for Monomer B, for each of the wildtype, fast-translating mutant, and slow-translating mutant mRNAs)[9,11]. In these simulations, a coarse-grain cutout of the ribosome exit tunnel and 50S surface consisting of 3,800 interaction sites is explicitly represented (Figures 1a-c). The dwell time at a particular nascent chain length k was randomly selected from an exponential distribution with a mean equal to the average decoding time of the $k + 1$ codon (*i.e.*, the time to decode the codon in the ribosomal A-site). After synthesis is completed, the harmonic restraint on the C-terminal bead representing the covalent bond between the nascent protein and the P-site tRNA is removed, allowing ejection of the nascent protein from the exit tunnel (Figure 1c, panel 2). Simulations of nascent protein ejection are run until the C-terminal residue reached an x-coordinate of 100 Å or greater in the internal CHARMM coordinate system, indicating the nascent protein had left the ribosome exit tunnel. After ejection, the ribosome representation was deleted and 5 μs of post-translational dynamics simulated for each trajectory (Figure 1c, panel 3).

Computing the average interface interaction energy between monomers. Two hundred pairs of monomer structures were randomly selected with replacement from the 100 final conformations of Monomer A and 100 final conformations of Monomer B obtained after 5 μs of post-translational dynamics for a given protein and mRNA variant (Figure 1c, Panel 4). To generate dimer structures, random monomer pairs were first aligned to the crystal structure coordinates based on interface residue locations only. Steric clashes were then resolved by an iterative procedure to identify the lowest-energy dimer structure. In this procedure, the interaction energy between the two monomers is first calculated in CHARMM. If the energy is positive (*i.e.*, repulsive), the Monomer B structure is translated 0.5 Å away from the Monomer A structure along the vector connecting their interface centers of mass. This procedure is terminated when the interaction energy is found to be less than or equal to zero. The resulting conformation is used as the starting configuration for annealing simulations (Figure 1c, Panel 5). During annealing, the dimer structure is cooled from 310 to 0 K in 5 K increments. At each temperature, 150 ps of Langevin dynamics is simulated to allow for structural rearrangement of the interface. A harmonic root mean square deviation restraint with force constant 5 kcal/[mol Å^2] is applied to each monomer to maintain their initial conformations. Five hundred independent annealing simulations were run for each pair of randomly selected monomers and the structure with the lowest interface interaction energy after annealing selected. The average interaction energy between monomers generated for a particular protein and mRNA reported in Figure 2 is computed as the mean of these 200 lowest-energy values found from the sets of 500 annealing simulations for each of the 200 random dimer structures. It should be noted that this strategy assumes dimerization occurs exclusively in solution and not during synthesis by the ribosome (known as post-post assembly), which is reasonable given that they are both RNase H folds, which was not identified by Ref. 1 as a protein class that assembles co-translationally.

Calculating potential of mean force of dimerization. Representative coarse-grain dimer structures were selected as the annealed structure (see previous section) with interaction energy closest to the median value within the set of 200 annealed structures generated for a given synonymous mRNA. Using this procedure, dimer structures were selected from the WT and slow mRNA ensembles for both oligoribonuclease and ribonuclease T. Initial structures for Hamiltonian (umbrella sampling) Replica Exchange (HREX)[19,20] were then generated by translating one monomer away from the other along the vector connecting the centers of mass of residues at the dimer interface to create structures spaced every 0.5 Å. HREX was simulated at 310 K, using center-of-mass harmonic restraints ranging from 6.5 Å to 100 Å. A total of 10,000 exchanges (~750 ns total simulation time) were attempted between nearest neighbor umbrellas and used to construct potentials of mean force as a function of the interface center-of-mass distance using the histogram free formulation of WHAM equations[21,22] and the following equation[23]:

$$F(\delta_c) = -k_{\mathrm{B}} T \ln[P(\delta_c)] \qquad [2]$$

where $P(\delta_c)$ is the probability of finding a distance between the monomer center of masses δ_c.

Folding times after release from the ribosome. A given trajectory of monomeric oligoribonuclease or ribonuclease T is considered to fold after its release from the ribosome when its Q first reaches ≥0.69 and remains ≥0.69 for at least 750 ps[13]. These cutoffs were set based on a training set of 18 proteins[13]. Based on this definition of folding, we computed the survival probability of the unfolded state of each protein as a function of time, denoted $S_{\mathrm{U}}(t)$. The resulting time series were then fit to the double-exponential equation $S_{\mathrm{U}}(t) = f_1 \exp(-k_1 t) + f_2 \exp(-k_2 t)$ with $f_1 + f_2 \equiv 1$. This fit equation corresponds to a kinetic scheme in which the unfolded and misfolded states proceed to the folded state by parallel folding pathways and there is no inter-transition between unfolded and misfolded states. The folding times of the two kinetic phases are computed as $\tau_1 = 1/k_1$ and $\tau_2 = 1/k_2$, with the larger of these two times determining the overall timescale of the folding process. The results of this fitting procedure for oligoribonuclease and ribonuclease T are summarized in Figure S1 and the resulting fit parameters are listed in Table S1.

Identifying changes in entanglement and the residues involved in entanglements. To identify non-covalent lasso entanglements we utilize Gaussian linking numbers[24], which describe the linking between two closed loops in three-dimensional space. This procedure is a modified version of a protocol previously used to detect entanglements in coarse-grain protein structures[25]. The first loop is composed of the peptide backbone connecting residues i and j that form a native contact. Outside this loop is an N-terminal segment, composed of residues 5 through $i-4$, and a C-terminal segment composed of residues $j+4$ through $N-5$ for oligoribonuclease. Similar terminal ranges were used for ribonuclease T, but with a 15-residue terminal offset instead of 5 to address transient virtual entanglements caused by the long flexible tails of ribonuclease T. We characterize the entanglement of each tail with the loop formed by the native contacts with two partial linking numbers denoted g_{N} and g_{C}. We use the approximation to the partial Gaussian double integration method introduced by Baiesi and co-workers[26] to calculate these partial linking numbers for a closed (loop) and open curve (termini). For a given structure of an N-residue protein, with a native contact present at residues (i,j), the coordinates $\boldsymbol{R}_l$ and the gradient $\mathrm{d}\boldsymbol{R}_l$ of the point l on the curves were calculated as

$$\begin{cases} \boldsymbol{R}_l = \frac{1}{2}(\boldsymbol{r}_l + \boldsymbol{r}_{l+1}) \\ \mathrm{d}\boldsymbol{R}_l = \boldsymbol{r}_{l+1} - \boldsymbol{r}_l \end{cases}, \qquad [3]$$

where $\boldsymbol{r}_l$ is the coordinates of the C$_\alpha$ atom in residue l. The linking numbers $g_{\mathrm{N}}(i,j)$ and $g_{\mathrm{C}}(i,j)$ were calculated as

$$\begin{cases} g_{\mathrm{N}}(i,j) = \dfrac{1}{4\pi}\displaystyle\sum_{m=6}^{i-5}\sum_{n=i}^{j-1}\frac{\boldsymbol{R}_m - \boldsymbol{R}_n}{|\boldsymbol{R}_m - \boldsymbol{R}_n|^3}\cdot(\mathrm{d}\boldsymbol{R}_m \times \mathrm{d}\boldsymbol{R}_n) \\ g_{\mathrm{C}}(i,j) = \dfrac{1}{4\pi}\displaystyle\sum_{m=i}^{j-1}\sum_{n=j+4}^{N-6}\frac{\boldsymbol{R}_m - \boldsymbol{R}_n}{|\boldsymbol{R}_m - \boldsymbol{R}_n|^3}\cdot(\mathrm{d}\boldsymbol{R}_m \times \mathrm{d}\boldsymbol{R}_n) \end{cases}, \qquad [4]$$

where we exclude the first 5 residues of the N-terminal curve, last 5 residues of the C-terminal curve (for ribonuclease T this cut-off is increased to 15), and 4 residues before and after the native contact to eliminate the error introduced by both the high flexibility and contiguity of the termini and trivial entanglements in local structure. The total linking number for a native contact (i,j) is therefore estimated as

$$g(i,j) = \mathrm{round}\big(g_{\mathrm{N}}(i,j)\big) + \mathrm{round}\big(g_{\mathrm{C}}(i,j)\big), \qquad [5]$$

Comparing the absolute value of the total linking number for a native contact (i,j) to that of a reference state allows us to detect a gain or loss of linking between the backbone trace loop and the terminal open curves as well as any switches in chirality. Therefore, there are six change in linking cases we should consider when using this approach to quantify entanglement (see Supplementary Table 9 of ref. 9).

The N- and C-terminal threading locations, $g_{\mathrm{N|C}}(i,j,r)$, of the most complex non-native entanglement is identified by first finding the native contact (i,j) where the total linking number is the equal to the global maximum of the set of all total linking numbers for the protein, $g(i,j) = MAX[g(i,j)]$, and at which there was a change of entanglement detected. We then employed the loop-piercing method in the python Topoloy[27] package to identify the residues along the threading tail that pierce the loop plane. Identifying the set of residues involved in the change of entanglement allows us to discern if the location disrupts the interface by examining the intersection of the set of interface residues with this set. An entanglement is considered to occur at the interface if any entangled residues are also identified to be interface residues, where interface residues are defined as those residues in Monomer A within 4.5 Å of Monomer B or vice versa.

Clustering of dimeric entangled structures. Considering an ensemble of dimeric structures that contains at least one intra-monomer change in entanglement (no inter-monomer entanglements were observed in the simulations) we separate them into clusters by examining the intersection between the sets of entangled residues in the two structures. Structures were merged in a leader algorithm[28] style where the leader challenge is as follows:

(1) Consider a leader superset of entangled residues $L = \{A, B\}$ and a subordinate superset $s = \{a, b\}$ where the sets A, B, a, b are the sets of residues in either monomer A(a) or B(b) that are involved in the entanglement.
(2) Calculate the intersections $I_{Aa} = A \cap a$, $I_{Ab} = A \cap b$, $I_{Bb} = B \cap b$, and $I_{Ba} = B \cap a$
(3) If $|I_{Aa}|\ \&\ |I_{Bb}| > 0 \text{ or } |I_{Ab}|\ \&\ |I_{Ba}| > 0$, the subordinate passes the challenge and becomes part of the leader group. Otherwise, the challenge is failed and the search continues. If a subordinate fails the challenge of every current leader it become a new leader.

The three most-populated (D1, D2, and D3) and the two lowest-energy (D4 and D5) dimeric entangled states of oligoribonuclease were selected for back-mapping to atomistic resolution as described below. Three all-atom monomeric systems were also generated by selecting three entangled monomers from within these five dimeric starting structures (see Supplementary Data File 1).

Back-mapping of coarse-grain monomer and dimer structures to all-atom resolution. Coarse-grain interaction sites representing the side-chain center-of-mass were rebuilt near their corresponding C_α beads in the selected dimer structures based on the native-state all-atom conformation[11]. Energy minimization was then performed using a two-bead C_α-C_β coarse-grain force field[29] generated from the original all-atom PDB with all C_α positions restrained. The backbone and sidechain atoms were then rebuilt with PD2[30] and Pulchra[31], respectively. The final all-atom structure was obtained after a further energy minimization *in vacuo* with all C_α positions restrained using OpenMM[32]. Representative starting coarse-grain and ending all-atom structures for the oligoribonuclease monomer and dimer are provided in Figure 1d and e, respectively.

All-atom simulations of entangled structures. The back-mapped protein was placed in a rectangular box with a minimum distance of 1 nm between the edge of the protein and the periodic boundary wall in all dimensions. The system was solvated in TIP3P[33] water and neutralized by Na^+ and Cl^- counter-ions before adding 0.15-M sodium chloride to mimic the salt concentration inside the cell[34]. We next minimized and equilibrated the system. First, 1 ns of dynamics was carried out in the NVT ensemble, followed by 1 ns of dynamics in the NPT ensemble with harmonic position restraint potential (spring constant k = 1000 kJ/[mol x nm^2]) applied to all heavy atoms of protein to relax the environment with the temperature and pressure held at 310 K and 1 atm, respectively. To allow the protein to reach equilibrium in the all-atom model and maintain the coarse-grain structure, we performed a second NPT simulation for 1 ns with harmonic restraints (k = 1000 kJ/[mol x nm^2]) applied to all C_α atoms. Finally, we ran production simulations for 500 ns for five dimer and three monomer structures with three statistically independent trajectories with different initial velocities generated from the Maxwell distribution for each starting structure. Simulations for one randomly selected dimer and monomer structure were extended to 1 μs with no qualitative change in results.

Simulations were performed with GROMACS 2018[35] using the AMBER99SB-ildn forcefield[36]. The AMBER99sb-ildn forcefield is widely used in all-atom protein simulations as it has improved side chain torsion parameters of four residues (Ile, Leu, Asp, Asn) to yield a better agreement with NMR data[36]. The particle mesh Ewald method[37] was used to calculate the long-range electrostatic interactions beyond 1 nm. Van der Waals interactions were calculated with a cut-off distance of 1 nm. The Nose-Hoover thermostat[38,39] and Parrinello-Rahman barostat[40] were employed to maintain the temperature and pressure at 310 K and 1 atm, respectively. The LINCS[41] algorithm was used to constrain all bonds involving hydrogen atoms. An integration time step of 2 fs was used for all simulations.

Calculating odds ratios and significance. The 200 post-annealing dimer structures generated for each protein and mRNA were labelled as strong or weak binding based on whether they had an interaction energy less than or equal to a threshold value selected from the cumulative distribution function of the interaction energy of the ensemble of annealed wildtype dimer structures. This threshold was initially set to the value at which the cumulative distribution function of interaction energy equals 5% and then increased to 95% in 10% increments. The number of structures containing non-native changes in entanglement was counted for each of the thresholds in increments of +5% from 5% to 95%. A contingency table at each threshold was then generated where the two events are: (1) strong or weak binding and (2) the presence or absence of non-native entanglements. Odds ratios and p-values for the contingency tables constructed in this way (Supplementary Data File 2) were computed in Python3 using the fisher_exact function in the SciPy (v1.10.1) stats module with the one-tailed hypothesis test.

Clustering of oligoribonuclease metastable states from post-translational simulations. To analyse the last 100 ns of the post-translational structural distribution resulting from 200 independent CG simulations, 400 k-means[42,43] clusters (micro-states) were made in the space spanned by the two order parameters G and Q. The clusters were aggregated into a small number of metastable states using the PCCA+ algorithm[44]. The optimal number of metastable states was chosen based on the existence of a gap in the eigenvalue spectrum of the transition probability matrix[45]. A probability surface $-log(P(G,Q))$ of the space can be used to show the population shifts of these metastable states under different translation schemes. Five representative structures of each metastable state were randomly sampled from all microstates according to the probability distribution of the microstates within the given metastable state. All the clustering and MSM building were performed by using the PyEmma package[46].

Calculation of relative change in solvent-accessible surface area for metastable states. Eq. 16 from ref. 9 was applied to compute the average change in solvent-accessible surface area over the sets of residues [84-94] and [166-175] for oligoribonuclease or [204-215] and [2] for ribonuclease T relative to the native state. The mean solvent-accessible surface area for the native state was computed as the average over ten 1-μs simulations initiated from the native state conformation. The resulting values and confidence intervals are provided in Tables S3 and S4.

Refoldability Experiments. Methods are described in greater detail in refs. 48 & 51. *E. coli* K12 cells (NEB) were grown in 2 sets of 3 × 50 mL (biological triplicates) MOPS EZ rich media from saturated overnight cultures with a starting OD_{600} of 0.05. As described in ref. 9, one set was supplemented with 0.5 mM [$^{13}C_6$]L-Arginine and 0.4 mM [$^{13}C_6$]L-Lysine and the other with 0.5 mM L-Arginine and 0.4 mM L-Lysine. Cells were cultured at 37°C with agitation (220 rpm) to a final OD_{600} of 0.8. Each heavy/light pair was pooled together; cells were collected by centrifugation at 4000 *g* for 15 mins at 4˚C, supernatants were removed, and cell pellets were stored at -20˚C until further use.

Frozen cell pellets were resuspended in a lysis buffer consisting of 900 µL of Tris pH 8.2 (20 mM Tris pH 8.2, 100 mM NaCl, 2 mM MgCl$_2$ and supplemented with DNase I to a final concentration (f.c.) of 0.1 mg mL^{-1}). Resuspended cells were cryogenically pulverized with a freezer mill (SPEX Sample Prep). Lysates were then clarified at 16000 *g* for 15 min at 4 °C to remove insoluble cell debris. To deplete ribosome particles, clarified lysates were ultracentrifuged at 33,300 rpm at 4 °C for 90 min using a SW55 Ti rotor. Protein concentrations of clarified lysates were determined using the bicinchoninic acid assay (Rapid Gold BCA Assay, Pierce) and diluted to 3.3 mg mL^{-1} using lysis buffer.

To prepare native samples, 3.5 µL of normalized lysates were diluted with 96.5 µL of Tris native dilution buffer (20 mM Tris pH 8.2, 100 mM NaCl, 10.288 mM MgCl$_2$, 10.36 mM KCl, 2.07 mM ATP, 1.04 mM DTT, 62 mM GdmCl) to a final protein concentration of 0.115 mg mL^{-1}. Native samples were then equilibrated by incubating for 90 min at room temperature. To prepare unfolded samples, 600 µL of normalized lysates, 100 mg of solid GdmCl, and 2.4 µL of a freshly prepared 700 mM DTT stock solution were combined, and solvent was removed using a vacufuge plus to a final volume of 170 µL. Unfolded lysates were incubated overnight at room temperature. To refold, 99 µL of refolding dilution buffer (19.5 mM Tris pH 8.2, 97.5 mM NaCl, 10.03 mM MgCl$_2$, 10.1 mM KCl, 2.02 mM ATP and .909 mM DTT) were rapidly added to 1 µL of unfolded extract. Refolded samples were then incubated at room temperature for 1 min, 5 min or 2 h.

100 µL of the native or refolded lysates was added to Proteinase K (enzyme:substrate ratio of 1:100 w/w ratio[47]), incubated for 1 min at room temperature, and quenched by boiling in a mineral oil bath at 110°C for 5. Boiled samples were transferred to tubes containing 76 mg urea. To prepare samples for mass spectrometry, dithiothreitol was added to a final concentration of 10 mM and samples were incubated at 37°C for 30 minutes. Iodoacetamide was added to a final concentration of 40 mM and samples were incubated at room temperature in the dark for 45 minutes. LysC was added to a 1:100 enzyme:substrate (w/w) ratio and samples were incubated at 37˚C for 2 h, urea was diluted to 2 M using 100 mM ammonium bicarbonate pH 8, then trypsin was added to a 1:50 enzyme:substrate (w/w) ratio and incubated overnight at 25°C.

Peptides were acidified, desalted with Sep-Pak C18 1 cc Vac Cartridges, dried down, and resuspend in 0.1% formic acid, as described in[48]. LC-MS/MS acquisition was conducted on a Thermo Ultimate3000 UHPLC system with an Acclaim Pepmap RSLC C18 column (75 µm × 25 cm, 2 µm, 100 Å) in line with a Thermo Q-Exctive HF-X Orbitrap, identically to as described in[48].

Proteome Discoverer (PD) Software Suite (v2.4, Thermo Fisher) and the Minora Algorithm were used to analyze mass spectra and perform Label Free Quantification (LFQ) of detected peptides. Default settings for all analysis nodes were used except where specified. The data were searched against Escherichia coli (UP000000625, Uniprot) reference proteome database. For peptide identification, the PD MSFragger node was used, using a semi-tryptic search allowing up to 2 missed cleavages[49]. A precursor mass tolerance of 10 ppm was used for the MS1 level, and a fragment ion tolerance was set to 0.02 Da at the MS2 level. Additionally, a maximum charge state for theoretical fragments was set at 2. Oxidation of methionine and acetylation of the N-terminus were allowed as dynamic modifications while carbamidomethylation on cysteines was set as a static modification. Heavy isotope labeling (^{13}C$_6$) of Arginine and Lysine were allowed as dynamic modifications. The Philosopher PD node was used for FDR validation. Raw normalized extracted ion intensity data for the identified peptides were exported from the .pdResult file using a three-level hierarchy (protein > peptide group > consensus feature). These data were further processed utilizing custom

Python analyzer scripts (available on GitHub, and described in depth previously in refs. 48 & 51).

RESULTS

Synonymous mutations alter oligoribonuclease's post-translational structural ensemble and ability to dimerize. To test if oligoribonuclease's ability to dimerize is perturbed by synonymous mutations we simulated its synthesis from mRNAs corresponding to its wildtype coding sequence, a slow-translating synonymous mRNA sequence composed of non-optimal codons, and a fast-translating synonymous mRNA sequence composed of optimal codons (Figure 2a, b). After synthesis, we simulated the release of oligoribonuclease from the ribosome followed by 5 μs of post-translational dynamics (the equivalent of approximately 20 s in real time[13]). We find that oligoribonuclease exhibits structural differences in its post-translational folding dynamics dependent on whether it was translated from the wildtype, fast-translating, or slow-translating mRNA sequences (Figure 2c). The slow-translating mRNA produces a monomer structural ensemble with a higher average fraction of native contacts ($Q = 0.89$, 95% CI: [0.87, 0.90], computed from bootstrapping 10^6 times, where Q only considers intra-monomer contacts) relative to the wild-type mRNA ($Q = 0.86$, 95% CI: [0.85, 0.88], 10^6 bootstraps) 5 μs after ribosome release. The ensemble of minimum energy dimeric structures from temperature annealing simulations (Figure 1c), reveals that this increase in native structure results in a more favorable dimer interface interaction energy of -64.1 kcal/mol (95% CI [-65.7, -62.4], 10^6 bootstraps) for the slow mRNA products compared to dimers produced from the wildtype mRNA of -58.3 kcal/mol (95% CI [-60.3, -55.9], 10^6 bootstraps). The difference between these interaction energies is significant ($p = 2 \text{ x } 10^{-5}$, permutation test 10^6 iterations). The average dimer interaction energy of -60.5 kcal/mol (95% CI: [-62.5, -58.4], 10^6 bootstraps) for the fast-translating mRNA is also different in comparison to the slow-translating mRNA (Figure 2d). These results demonstrate that oligoribonuclease's post-translational dimerization affinity is affected by changes in translation-elongation speed.

Dimerization of ribonuclease T is not influenced by synonymous mutations. Ribonuclease T's post-translational structural ensemble (Figure 2e-g) and dimerization interaction energy are not dependent on the translation schedule of the mRNA that encodes it (Figure 2h). No statistically significant differences are found between the average interface interaction energies of the dimer ensembles generated from the wildtype, fast-translating, or slow-translating mRNAs.

Oligoribonuclease, but not ribonuclease T, frequently populates self-entangled states that involve interface residues. Recent computational studies predict that misfolded proteins often contain entanglements that form long-lived, native-like kinetic traps[8,9]. In entangled protein structures, a segment of residues forms a loop (closed by a native contact) through which another segment of residues threads (Figure 3a)[26]. Mathematically, entanglements within a protein structure can be detected as a change in the Gauss linking number, $g(i,j)$, of the native contacts relative to the folded state (see Methods and Eq. (3)-(5)). This metric of protein structure is topologically invariant[50] and describes how segments of the protein are intertwined together in space (see Figure 3a of ref. 9).

We identified various entanglements in structures of oligoribonuclease that cause disruptions relative to the native state (Methods). Representative structures of two frequently occurring entanglements, in which residues 96-102 or residues 125-129 thread through a loop, are displayed in Figures 3b-e, and schematic representations are shown in Figures 3g and h. Entanglement can occur in an isolated monomer (Figure 3b), in one monomer that forms part of a dimeric complex (Figure 3c), or in both monomers within a dimer (Figures 3d,e). Each of these entangled structures is very similar to the native structure; for example, the entangled dimer structure shown in Figures 3d and 3e has ≤3-Å C_α RMSD from the native state (Figure 3i). The high similarity of these entangled conformations to the native state suggests that they may remain soluble and evade proteostasis quality controls[9]. Despite being well-folded overall, entanglements can structurally perturb the dimer interface. In the case of the entanglement of residues 125-129 in Monomer A, misplacement of a loop segment disrupts the formation of a β-sheet at the dimer interface (Figure 3j). This suggests that changes in dimer interaction energy could be caused by entanglements perturbing the dimer binding interface.

Synonymous mutations alter the population of entangled oligoribonuclease structures. To quantify the influence of synonymous mutations on the likelihood of entanglement in oligoribuclease and ribonuclease T we computed the fraction of monomers (out of 200 independent trajectories) and dimers (out of 200 annealed structures) that exhibit a non-native change in entanglement for both oligoribonuclease and ribonuclease T from their wild-type, fast- and slow-translating mRNAs (Figure 4). We find that ribonuclease T exhibits relatively little entanglement (fraction entangled < 0.15) in both its monomeric and dimeric forms regardless of the translation-rate schedule used during its synthesis. Statistically significant differences are present depending on the translation schedule used (see, for example, Figure 4c); however, the magnitude of these population differences is small (less than 10.5%). This suggests why ribonuclease T's dimer interaction energy is insensitive to synonymous mutations – any corresponding population changes in misfolded states are modest and have little effect on this protein's ability to dimerize.

In contrast, for oligoribonuclease, the population of conformations that display a non-native change in entanglement is larger in magnitude and more sensitive to changes in translation speed. For the wildtype, fast, and slow translation schedules, the fractions of dimer conformations with an overall entanglement are, respectively 0.62 (95% CI: [0.54, 0.68]), 0.51 (95% CI: [0.43, 0.57]), and 0.37 (95% CI: [0.29, 0.43], 10^6 bootstraps) (Figure 4c). This trend is anticorrelated with the average dimer interaction energies (Figure 2d) where values of -58.3, -60.5, and -64.1 kcal/mol, respectively, are found for the wildtype, fast, and slow mRNA templates. These results suggest that changes in the population of entangled states with perturbed interfaces, arising from changes in translation speed, cause the binding affinity between monomers to be altered.

Misfolded entangled states often involve the dimerization interface of oligoribonuclease. Next, we asked how frequent it was for misfolded states to have the entanglement located at the dimer interface (see Methods and Figure 4). We find ribonuclease T has relatively low levels of entanglement at the dimer interface of misfolded structures, with probabilities of less than 0.15 (Figures 4b and d). In comparison, oligoribonuclease displays more frequent interface entanglements in both monomer and dimer misfolded structures (Figures 4b and d). Specifically, misfolded dimer structures have probabilities of entanglements located at the interface of 0.49 (95% CI: [0.41, 0.55]), 0.38 (95% CI: [0.31, 0.45]), and 0.26 (95% CI: [0.21, 0.32], 10^6 bootstraps) for the wildtype, fast-translating, and slow-translating mRNAs, respectively (Figure 4d). Thus, at least for oligoribonuclease, non-native entanglements are relatively frequent at the dimer interface.

Entanglements greatly reduce the likelihood of strong dimer interactions. To test whether entanglements are associated with decreased average binding energy between monomers, we create a two-by-two contingency table categorizing annealed dimer structures as strongly or weakly bound and as having an entanglement present or not. A contingency table allows us to compute the conditional probability of these two events co-occurring, the odds ratio (effect size) that entanglement and weak binding are associated, and Fisher's Exact Test tells us whether the association is significant. Statistically significant odds ratios other than 1 would establish an association between these two phenomena.

To classify structures based on their binding energy, we define dimer complexes with interface interaction energy less than or equal to the X^{th} percentile value from the wildtype interaction energy distribution to be strong binding and all others to be weak binding. Then, as a test of robustness, we systematically vary this threshold, *X*, in increments of +5% from 5% to 95% and compute the odds ratio and p-value for each splitting of the data (Figure 5, Supplementary Data File 2).

For oligoribonuclease we find the odds ratio, where it is defined (plotted portions of Figures 5a and b), is significantly greater than 1 for a range of thresholds. For example, at a 50% threshold, the odds ratio is 51.9 and is statistically significant (p-value = 1.4 x 10^{-9}, Fisher's Exact test; see Supplementary Data Table 2) for the slow-translating mRNA variant. This effect size and the consistent significance of these odds ratios demonstrates there is a very strong association between the presence of entanglements and the occurrence of dimers with weak dimerization energies.

Conversely, ribonuclease T's odds ratio is never statistically different than 1.0 (Figure 5b and d). This indicates, as inferred earlier, that the modest population of entangled structures for ribonuclease T has no association with strong or weak dimerization occurring.

Equilibrium potentials of mean force of dimerization. A potential concern of the aforementioned results is that they reflect only potential energy changes upon dimerization, not free energy changes. Equilibrium binding-and-unbinding simulations are computationally expensive even when employing enhanced sampling and coarse-grained protein models. Therefore, we could only compute the potential of mean force for dimerization for a small number of monomeric conformations. Specifically, we selected a representative coarse-grain dimer structure with interaction energy closest to the median within the set of 200 annealed structures generated for a given synonymous mRNA. We ran Hamiltonian replica-exchange sampling to compute the potentials of mean force (Eq. 2) for the process of dimerization. For oligoribonuclease we find that the slow variant has a greater free-energy of dimerization than the wildtype variant, indicating slow synthesis produces more stable dimers than the wildtype sequence (Figure S5a). For ribonuclease T we find no difference in the free energy of dimerization between the wildtype and slow variants indicating invariance of the binding strength to synonymous variants (Figure S5b). Thus, these results, taken together with our earlier results reporting changes in the interaction potential energy and the statistically significant enrichment of entanglements at the interface of oligoribonuclease are consistently showing an association between binding strength and near native changes in self-entanglement.

Entangled states are long-lived kinetic traps. Entangled states that are long lived can have long-term impacts on protein structure and function. Therefore, we quantified the lifetime of these states in our simulations. While all monomers of ribonuclease T fold by 0.8 μs after release from the ribosome, some oligoribonuclease molecules fail to fold during the 5-μs post-translational simulations regardless of the mRNA sequence used. When synthesized from its wildtype, fast-translating, and slow-translating variants, respectively, 13% (95% CI [8%, 17%]), 14% (95% CI [9%, 19%]) and 10% (95% CI [6%, 14%], 10^6 bootstraps) of its monomers do not fold correctly (see Figure S1 and Methods). We used a kinetic curve-fitting procedure to estimate that these misfolded populations of oligoribonuclease require between 6 and 14 μs to fold (the equivalent of approximately 30-60 s of real time[13]). This indicates that these entangled states are kinetically trapped.

Entanglements persist in all-atom molecular dynamics for up to one microsecond. To test whether our results generated using a C$_\alpha$ coarse-grain representation are resolution dependent we back-mapped representative entangled conformations of the oligoribonuclease dimer and monomer to atomistic resolution and simulated their aqueous dynamics for 500 ns. We ran three statistically independent trajectories for five different dimer starting structures and three different monomer structures selected to represent the entangled conformations most frequently populated and lowest in energy (see Methods and Supplementary Data File 1). In each case, the entanglement present in the coarse-grain model persists at all-atom resolution for the duration of the 500-ns simulation. In addition to these 500-ns simulations, we also extended the simulations for one randomly selected dimer and monomer structures to 1 μs and find the entangled states persist. This is evidenced by the time series of $\langle G \rangle$ for four representative entanglements, one in a monomer and three in dimers, displayed in Figure S2. Thus, the entangled structures we observe in our coarse-grained simulations can also be populated and persist in all-atom models.

Entangled states are consistent with mass-spec signatures of altered structure. To experimentally test our predictions we utilized previously published limited proteolysis-mass spectrometry (LiP-MS)[9,48,51] data obtained from proteome-wide refolding studies conducted on *E. coli* extracts. In LiP-MS, comparison of proteolysis profiles is used to assess structural differences between native proteins (obtained directly from cell extracts) and those that have been chemically denatured and then refolded through a dilution jump. Proteolytic fragments that exhibit large changes in their populations between the refolded and native samples are indicative of regions of the protein that have altered protease susceptibility upon refolding. Entanglements can alter protease accessibility. Thus, LiP-MS data provides a means to test the computationally predicted changes in entanglement. In these data, we focus on peptides from oligoribonuclease and ribonuclease T. We consider only peptides that exhibit at least a 2 fold change in number between the refolded (R) and native (N) samples (i.e., $\left|\log_2\left(\frac{R}{N}\right)\right| \geq 1$) and are statistically different between the refolded and native sample ($-\log_{10} p \geq 2$, *i.e.* p<0.01)[48,51]. Here, we limit our analysis to long-lived misfolded states by considering only statistically significant changes detected 120 min after the dilution jump (see Supplementary Data File 4 of ref. 9). At that time point oligoribonuclease has two peptides that show significant changes, consisting of residues 84-94 and 166-175, while ribonuclease T exhibits no significant peptides (out of five peptides detected from that protein; see Table S2). The same results are found when the experiments are carried out in a cytosolic-like medium[48], and are moreover qualitatively consistent with our prediction that oligoribonuclease is more likely to populate entangled states than ribonuclease T (Figure 4).

To make a detailed structural comparison between our predicted misfolded states and these LiP-MS data we identify metastable states in the 2-D Log-probability surface defined by the order parameters Q and G (Figures 6 a-d, see Methods). Both proteins contain several metastable states with a high fraction of native contacts involving a change in entanglement, indicating these states are well folded but with structural perturbations introduced by entanglements (Figure 6). Two of the metastable states identified for oligoribonuclease (0 and 1 in Figures 6a and 6b) display increases in the solvent-accessible surface area of residues 84-94 relative to the native state reference simulations and together make up 10-16% of the conformations in the overall structural ensemble (Table S3, Figure 6e) produced by translation of the wildtype, fast, and slow mRNAs. Similarly, for residues 166-175, two metastable states (1 and 5) show a statistically significant increase in the solvent-accessible surface area and make up 65-80% of the structural ensembles across wildtype, fast, and slow (Figure 6f, Table S3). Thus, metastable states within our entangled state ensemble are consistent with the experimentally observed increase in protease accessibility of both proteolytic peptide fragments.

Ribonuclease T, on the other hand, only displays a single misfolded metastable state across all three variants with appreciable population (State 7, see Table S4). Of the five sites identified for ribonuclease T, LiP-MS found no significant change in protease susceptibility after refolding; in agreement, our simulations show small changes in solvent-accessible surface area relative to the native state at these sites (Table S4), with all values close to zero. This observation is consistent with our prediction that this protein would refold efficiently, leading to a similar protease accessibility between the native and refolded samples.

DISCUSSION

Our results provide a structural explanation for how changes in translation speed induced by synonymous mutations can alter the ability of soluble proteins to dimerize over long time scales. For the homodimer oligoribonuclease, synonymous mutations change the proportion of protein molecules that partition into soluble, misfolded, self-entangled conformations. These entangled conformations weaken the ensemble-averaged binding energy between the monomers over long time scales. In comparison to oligoribonuclease, ribonuclease T is largely insensitive to synonymous mutations that alter translation speed, with far fewer states with entangled dimer interfaces (≤15%) populated and those that are entangled exhibiting little population dependence on translation speed. We find that the key entangled states identified for oligoribonuclease persist at all-atom resolution for long timescales. On the other hand, ribonuclease T's dimer binding energy does not change with the introduction of synonymous mutations. Finally, we used structure-based Markov State Models to compare metastable misfolded conformations with LiP-MS analysis of proteome-wide refolding, finding consistency with experimental data that indicate structural perturbation at residues 84-94 and 166-175 for oligoribonuclease.

A commonly held assumption in the nascent protein folding field is that slower translation will result in more co-translational protein folding[52–56]. Therefore, one would predict that any changes in dimer interaction energy would follow the trend that the slow-translating mRNA will result in the strongest binding, followed by the wild-type and fast-variant mRNAs of oligoribonuclease. This is not what we observe – we find, respectively, that slow, fast, and wild-type mRNA variants result in increasingly weaker dimer affinities. This result is explained by both kinetic and simulation models showing the influence of translation kinetics on co-translational protein folding is, for some proteins, non-monotonic. Faster translation can result in an increased yield of correctly folded protein by translating quickly through protein segments that are prone to misfolding[57,58]. Is it surprising that the wildtype sequence produces the most entangled oligoribonuclease in our simulations? The evolutionary principle of parsimony, that evolution does not further optimize features that are already 'sufficient', suggests that the wildtype sequence is "good enough" under normal environmental conditions. That is, it produces enough folded, functional protein that its codon sequence need not be optimized to generate 100% folded, functional protein. A classic example of this phenomenon is the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) protein, the dysfunction of which causes cystic fibrosis[59]. A large proportion of wildtype CFTR is tagged for degradation by the endoplasmic reticulum-associated degradative pathway[60,61], indicating that the majority of the protein produced from even the wildtype sequence is recognizably defective. Further, in epithelial cell lines the restoration of functional CFTR to only 25% of cells leads to the same function as wildtype[62]. Thus, not all wildtype mRNA sequences have been maximally optimized for protein folding efficiency, with many mRNAs leading to enough functional protein so as not to be problematic for the cell.

Oligomer assembly can begin early in the life of a protein, with some nascent chains co-translationally dimerizing between adjacent ribosomes[1]. It is unknown how many different proteins engage in such co-translational assembly, though it appears to be preferred by homodimers in specific folds, particularly coiled coils[1]. Both ribonuclease T and oligoribonuclease have a RNase-H-like fold, which was not found by Bertolini and co-workers to be a high-confidence co-co assembling candidate. Therefore, in this study we chose to consider their dimerization after their release from the ribosome only. Additionally, the motivating experiments on FRQ assessed only post-translational dimerization. Based on our results, we speculate that co-translational interface interaction energies are likely to follow similar mechanisms as we have identified in this post-translational study. Investigating how synonymous mutations influence co-translational dimerization is an interesting avenue of future research for systems that are likely to dimerize co-translationally.

In summary, our results indicate that for some proteins synonymous mutations can modulate the amount of nascent protein that misfolds into non-native self-entangled conformations with reduced dimer interface interaction energies. For oligoribonuclease, slowing down or speeding up translation relative to its wildtype translation schedule leads to a reduction in entanglement, especially at the interface, leading to more stable dimers on average. Ribonuclease T, however, folds quickly and is less prone to misfolding, and is therefore largely unaffected by synonymous mutations. Finally, LiP-MS experiments and Markov state modeling of our post-translational data demonstrate that oligoribonuclease has specific patterns of misfolding that may correlate with entanglements. Taken in combination with a recent large-scale study of entanglement in the *E. coli* proteome[9] and an in-depth analysis of the influence of entanglement on enzymatic activity[8], our results support an emerging view that near-native, entangled misfolded states are likely to be a common phenomenon that influences a wide range of protein functions.

ACKNOWLEDGEMENTS

L.P.D. acknowledges the Department of Science and Technology at Ho Chi Minh City for their support (Grant 07/2020/HĐ-KHCNTT). S.D.F. acknowledges support from the NIH Director’s New Innovator Award (DP2GM140926) and from the National Science Foundation Division of Molecular and Cellular Biology (MCB-2045844). M.S.L. acknowledges support by Narodowe Centrum Nauki in Poland (Grant 2019/35/B/ST4/02086) and by PLGrid Infrastructure and the supercomputer centre TASK in Gdansk, Poland. E.P.O. acknowledges funding from the National Institutes of Health (R35-GM124818) and the National Science Foundation (MCB-1553291). This work used the Extreme Science and Engineering Discovery Environment (XSEDE[63]), which is supported by National Science Foundation grant number ACI-1548562.

DATA AVAILABILITY STATEMENT

Coarse-grain and all-atom molecular dynamics simulation data is available upon request. Codes used for back mapping to all-atom resolution and detection of self-entanglements are available at: https://github.com/orgs/obrien-lab/

REFERENCES

(1) Bertolini, M.; Fenzl, K.; Kats, I.; Wruck, F.; Tippmann, F.; Schmitt, J.; Auburger, J. J.; Tans, S.; Bukau, B.; Kramer, G. Interactions between Nascent Proteins Translated by Adjacent Ribosomes Drive Homomer Assembly. *Science (80-.).* **2021**, *371* (6524). https://doi.org/10.1126/science.abc7151.

(2) Goodsell, D. S.; Olson, A. J. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105–153.

(3) Levy, E. D.; Pereira-Leal, J. B.; Chothia, C.; Teichmann, S. A. 3D Complex: A

Structural Classification of Protein Complexes. *PLoS Comput. Biol.* **2006**, *2* (11), 1395–1406. https://doi.org/10.1371/journal.pcbi.0020155.

(4) Juers, D. H.; Matthews, B. W.; Huber, R. E. LacZ β-Galactosidase: Structure and Function of an Enzyme of Historical and Molecular Biological Importance. *Protein Sci.* **2012**, *21* (12), 1792–1807. https://doi.org/10.1002/pro.2165.

(5) Schechter, A. N. Hemoglobin Research and the Origins of Molecular Medicine ASH 50th Anniversary Review Hemoglobin Research and the Origins of Molecular Medicine. *Blood* **2008**, *112* (10), 3927–3938. https://doi.org/10.1182/blood-BLOOD.

(6) Marianayagam, N. J.; Sunde, M.; Matthews, J. M. The Power of Two: Protein Dimerization in Biology. *Trends Biochem. Sci.* **2004**, *29* (11), 618–625. https://doi.org/10.1016/j.tibs.2004.09.006.

(7) Zhou, M.; Guo, J.; Cha, J.; Chae, M.; Chen, S.; Barral, J. M.; Sachs, M. S.; Liu, Y. Non-Optimal Codon Usage Affects Expression, Structure and Function of Clock Protein FRQ. *Nature* **2013**, *495* (7439), 111–115. https://doi.org/10.1038/nature11833.

(8) Jiang, Y.; Neti, S. S.; Sitarik, I.; Pradhan, P.; To, P.; Xia, Y.; Fried, S. D.; Booker, S. J.; O'Brien, E. P. How Synonymous Mutations Alter Enzyme Structure and Function over Long Timescales. *Nat. Chem.* **2022**. https://doi.org/10.1038/s41557-022-01091-z.

(9) Nissley, D. A.; Jiang, Y.; Trovato, F.; Sitarik, I.; Narayan, K. B.; To, P.; Xia, Y.; Fried, S. D.; O'Brien, E. P. Universal Protein Misfolding Intermediates Can Bypass the Proteostasis Network and Remain Soluble and Less Functional. *Nat. Commun.* **2022**, *13* (1). https://doi.org/10.1038/s41467-022-30548-5.

(10) Wright, P. E.; Dyson, H. J. Intrinsically Disordered Proteins in Cellular Signaling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (1), 18–29. https://doi.org/10.1038/nrm3920.

(11) Nissley, D. A.; Vu, Q. V; Trovato, F.; Ahmed, N.; Jiang, Y.; Li, M. S.; O'Brien, E. P. Electrostatic Interactions Govern Extreme Nascent Protein Ejection Times from Ribosomes and Can Delay Ribosome Recycling. *J. Am. Chem. Soc.* **2020**. https://doi.org/10.1021/jacs.9b12264.

(12) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Trigger Factor Slows Co-Translational Folding through Kinetic Trapping While Sterically Protecting the Nascent Chain from Aberrant Cytosolic Interactions. *J. Am. Chem. Soc.* **2012**. https://doi.org/10.1021/ja302305u.

(13) Leininger, S. E.; Trovato, F.; Nissley, D. A.; O'Brien, E. P. Domain Topology, Stability, and Translation Speed Determine Mechanical Force Generation on the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2019**. https://doi.org/10.1073/pnas.1813003116.

(14) Karanicolas, J.; Brooks, C. The Origins of Asymmetry in the Folding Transition States of Protein L and Protein G. *Protein Sci.* **2002**, *11*, 2351–2361. https://doi.org/10.2807/1560-7917.ES2014.19.46.20966.

(15) Best, R. B.; Chen, Y. G.; Hummer, G. Slow Protein Conformational Dynamics from Multiple Experimental Structures: The Helix/Sheet Transition of Arc Repressor. *Structure* **2005**. https://doi.org/10.1016/j.str.2005.08.009.

(16) Betancourt, M. R.; Thirumalai, D. Pair Potentials for Protein Folding: Choice of Reference States and Sensitivity of Predicted Native States to Variations in the Interaction Schemes. *Protein Sci.* **1999**, *8* (2), 361–369. https://doi.org/10.1110/ps.8.2.361.

(17) Karanicolas, J.; Brooks, C. L. Improved Gō-like Models Demonstrate the Robustness of Protein Folding Mechanisms towards Non-Native Interactions. *J. Mol. Biol.* **2003**, *334* (2), 309–325. https://doi.org/10.1016/j.jmb.2003.09.047.

(18) Fluitt, A.; Pienaar, E.; Viljoen, H. Ribosome Kinetics and Aa-TRNA Competition Determine Rate and Fidelity of Peptide Synthesis. *Comput. Biol. Chem.* **2007**. https://doi.org/10.1016/j.compbiolchem.2007.07.003.

(19) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**. https://doi.org/10.1016/S0009-2614(99)01123-9.

(20) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J. Chem. Phys.* **2000**, *113* (15), 6042. https://doi.org/10.1063/1.1308516.

(21) Nguyen, H. L.; Lan, P. D.; Thai, N. Q.; Nissley, D. A.; O'Brien, E. P.; Li, M. S. Does SARS-CoV-2 Bind to Human ACE2 More Strongly than Does SARS-CoV? *J. Phys. Chem. B* **2020**, *124* (34), 7336–7347. https://doi.org/10.1021/acs.jpcb.0c04511.

(22) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. https://doi.org/10.1002/JCC.540130812.

(23) O'Brien, E. P.; Okamoto, Y.; Straub, J. E.; Brooks, B. R.; Thirumalai, D. Thermodynamic Perspective on the Dock-Lock Growth Mechanism of Amyloid Fibrils. *J. Phys. Chem. B* **2009**. https://doi.org/10.1021/jp9050098.

(24) Kauffman, L.; Balachandran, A. P. Knots and Physics . *Phys. Today* **1992**. https://doi.org/10.1063/1.2809632.

(25) Vu, Q. V; Jiang, Y.; Sitarik, I.; Li, M. S.; OBrien, E. P. A New Class of Protein Misfolding Is Observed in All-Atom Folding Simulations. *bioRxiv* **2022**. https://doi.org/10.1101/2022.07.19.500586.

(26) Baiesi, M.; Orlandini, E.; Seno, F.; Trovato, A. Exploring the Correlation between the Folding Rates of Proteins and the Entanglement of Their Native States. *J. Phys. A Math. Theor.* **2017**. https://doi.org/10.1088/1751-8121/aa97e7.

(27) Dabrowski-Tumanski, P.; Rubach, P.; Niemyska, W.; Gren, B. A.; Sulkowska, J. I. Topoly: Python Package to Analyze Topology of Polymers. *Brief. Bioinform.* **2021**, *22* (3), 1–8. https://doi.org/10.1093/bib/bbaa196.

(28) Jain, A. K.; Murty, M. N.; Flynn, P. J. Data Clustering: A Review. In *ACM Computing Surveys*; 1999. https://doi.org/10.1145/331499.331504.

(29) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of Denaturants and Osmolytes on Proteins Are Accurately Predicted by the Molecular Transfer Model. *Proc. Natl. Acad. Sci.* **2008**. https://doi.org/10.4404/hystrix-28.2-12255.

(30) Moore, B. L.; Kelley, L. A.; Barber, J.; Murray, J. W.; MacDonald, J. T. High-Quality Protein Backbone Reconstruction from Alpha Carbons Using Gaussian Mixture Models. *J. Comput. Chem.* **2013**. https://doi.org/10.1002/jcc.23330.

(31) Rotkiewicz, P.; Skolnick, J. Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations. *J. Comput. Chem.* **2008**. https://doi.org/10.1002/jcc.20906.

(32) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**. https://doi.org/10.1371/journal.pcbi.1005659.

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**. https://doi.org/10.1063/1.445869.

(34) Ando, T.; Skolnick, J. Crowding and Hydrodynamic Interactions Likely Dominate in Vivo Macromolecular Motion. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (43), 18457–18462. https://doi.org/10.1073/pnas.1011354107.

(35) James, M.; Murtola, T.; Schulz, R.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(36) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958. https://doi.org/10.1002/prot.22711.

(37) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald

Sums in Large Systems. *J. Chem. Phys.* **1993**, *10089*. https://doi.org/10.1063/1.464397.

(38) Nosé, S.; Klein, M. L. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

(39) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics. *J. Chem. Phys.* **1984**, *81*, 511–519. https://doi.org/10.1063/1.447334.

(40) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.

(41) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.

(42) Steinhaus, H. Sur La Division Des Corps Matériels En Parties. *Bull. L'Académie Pol. des Sci.* **1956**, *4*, 801–804.

(43) MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA:*; 1967; pp 281–297. https://doi.org/10.1007/s11665-016-2173-6.

(44) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classif.* **2013**, *7* (2), 147–179.

(45) Buchete, N. V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112* (19), 6057–6069. https://doi.org/10.1021/jp0761665.

(46) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11* (11), 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743.

(47) Feng, Y.; De Franceschi, G.; Kahraman, A.; Soste, M.; Melnik, A.; Boersema, P. J.; De Laureto, P. P.; Nikolaev, Y.; Oliveira, A. P.; Picotti, P. Global Analysis of Protein Structural Changes in Complex Proteomes. *Nat. Biotechnol.* **2014**, *32* (10), 1036–1044. https://doi.org/10.1038/nbt.2999.

(48) To, P.; Xia, Y.; Lee, S. O.; Devlin, T.; Fleming, K. G.; Fried, S. D. A Proteome-Wide Map of Chaperone-Assisted Protein Refolding in a Cytosol-like Milieu. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (48). https://doi.org/10.1073/pnas.2210536119.

(49) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry-Based Proteomics. *Nat. Methods* **2017**, *14* (5), 513–520. https://doi.org/10.1038/nmeth.4256.

(50) Rolfson, D. *Knots and Links*; 1976.

(51) To, P.; Whitehead, B.; Tarbox, H. E.; Fried, S. D. Nonrefoldability Is Pervasive across the E. Coli Proteome. *J. Am. Chem. Soc.* **2021**, *143* (30), 11435–11448. https://doi.org/10.1021/jacs.1c03270.

(52) Komar, A. A.; Lesnik, T.; Reiss, C. Synonymous Codon Substitutions Affect Ribosome Traffic and Protein Folding during in Vitro Translation. *FEBS Lett.* **1999**. https://doi.org/10.1016/S0014-5793(99)01566-5.

(53) Siller, E.; DeZwaan, D. C.; Anderson, J. F.; Freeman, B. C.; Barral, J. M. Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency. *J. Mol. Biol.* **2010**, *396* (5), 1310–1318. https://doi.org/10.1016/j.jmb.2009.12.042.

(54) Spencer, P. S.; Siller, E.; Anderson, J. F.; Barral, J. M. Silent Substitutions Predictably Alter Translation Elongation Rates and Protein Folding Efficiencies. *J. Mol. Biol.* **2012**, *422* (3), 328–335. https://doi.org/10.1016/j.jmb.2012.06.010.

(55) Zhang, G.; Hubalewska, M.; Ignatova, Z. Transient Ribosomal Attenuation Coordinates Protein Synthesis and Co-Translational Folding. *Nat. Struct. Mol. Biol.* **2009**, *16* (3), 274–280. https://doi.org/10.1038/nsmb.1554.

(56) Nissley, D. A.; Sharma, A. K.; Ahmed, N.; Friedrich, U. A.; Kramer, G. G.; Bukau, B.; O'Brien, E. P. Accurate Prediction of Cellular Co-Translational Folding Indicates

Proteins Can Switch from Post- to Co-Translational Folding. *Nat. Commun.* **2016**. https://doi.org/10.1038/ncomms10341.

(57) O'brien, E. P.; Vendruscolo, M.; Dobson, C. M. Kinetic Modelling Indicates That Fast-Translating Codons Can Coordinate Cotranslational Protein Folding by Avoiding Misfolded Intermediates. *Nat. Commun.* **2014**, *5*, 2988. https://doi.org/10.1038/ncomms3988.

(58) Trovato, F.; O'Brien, E. P. Fast Protein Translation Can Promote Co- and Posttranslational Folding of Misfolding-Prone Proteins. *Biophys. J.* **2017**, *112* (9), 1807–1819. https://doi.org/10.1016/j.bpj.2017.04.006.

(59) Ratjen, F.; Bell, S. C.; Rowe, S. M.; Goss, C. H.; Quittner, A. L.; Bush, A. Cystic Fibrosis. *Nat. Rev. Dis. Prim.* **2015**, *1* (May), 15010. https://doi.org/10.1038/nrdp.2015.10.

(60) Ward, C. L.; Omura, S.; Kopito, R. R. Degradation of CFTR by the Ubiquitin-Proteasome Pathway. *Cell* **1995**, *83* (1), 121–127. https://doi.org/10.1016/0092-8674(95)90240-6.

(61) Varga, K.; Jurkuvenaite, A.; Wakefield, J.; Hong, J. S.; Guimbellot, J. S.; Venglarik, C. J.; Niraj, A.; Mazur, M.; Sorscher, E. J.; Collawn, J. F.; Bebok, Z. Efficient Intracellular Processing of the Endogenous Cystic Fibrosis Transmembrane Conductance Regulator in Epithelial Cell Lines. *J. Biol. Chem.* **2004**, *279* (21), 22578–22584. https://doi.org/10.1074/jbc.M401522200.

(62) Zhang, L.; Button, B.; Gabriel, S. E.; Burkett, S.; Yan, Y.; Skiadopoulos, M. H.; Dang, Y. L.; Vogel, L. N.; McKay, T.; Mengos, A.; Boucher, R. C.; Collins, P. L.; Pickles, R. J. CFTR Delivery to 25% of Surface Epithelial Cells Restores Normal Rates of Mucus Transport to Human Cystic Fibrosis Airway Epithelium. *PLoS Biol.* **2009**, *7* (7). https://doi.org/10.1371/journal.pbio.1000155.

(63) John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, N. W.-D. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16* (5), 62–74. https://doi.org/doi:10.1109/MCSE.2014.80.

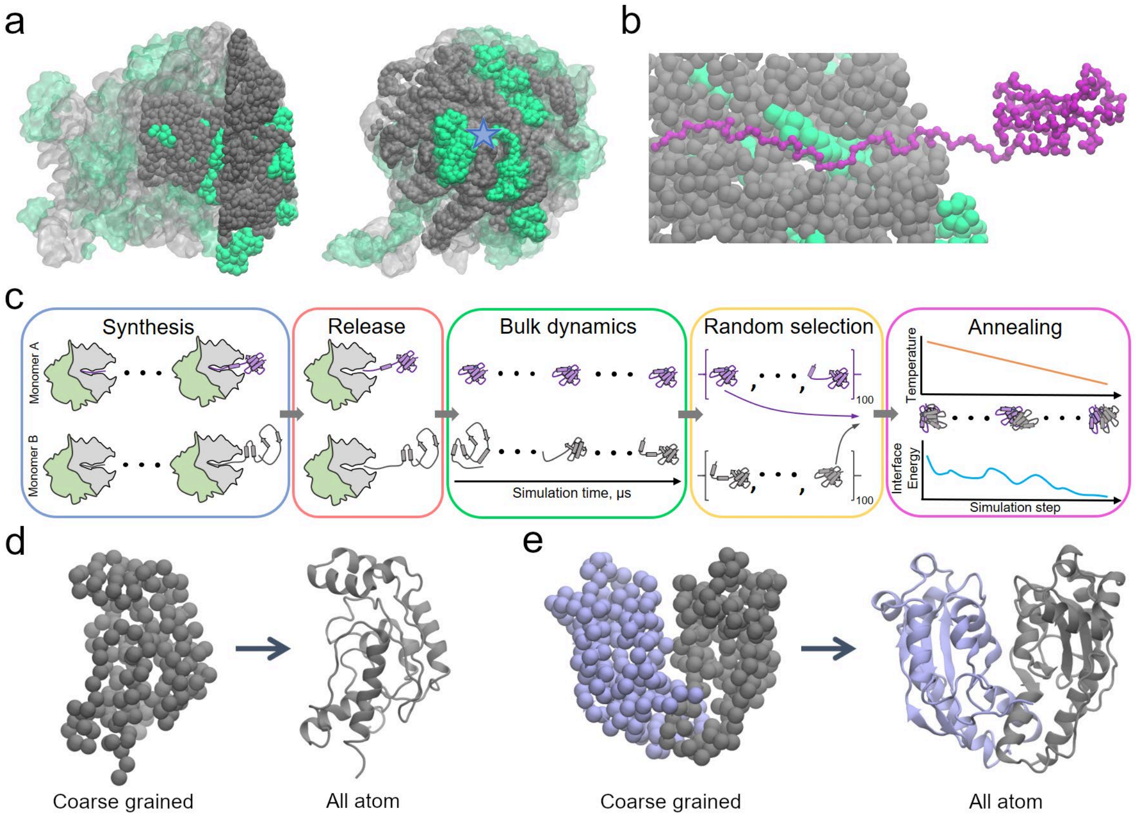

Figure 1. **Simulating protein dimerization and entanglement at multiple resolutions**. (a) Side (left) and top (right) views of the coarse-grained 50S *E. coli* ribosome cutout (filled spheres) used in our simulations superimposed over the entire all-atom 50S subunit (transparent) from PDB ID 3R8T. Ribosomal RNA and protein are displayed in grey and green, respectively. The approximate location of the ribosome exit tunnel is indicated by a blue star in the top view. (b) Side view of a 181-residue oligoribonuclease ribosome nascent chain complex just prior to its release from the ribosome. Note that one side of the exit tunnel was cut away for visualization only. (c) Schematic of simulation protocol. One hundred nascent protein conformations are generated for Monomer A (purple) and Monomer B (grey). Each monomer is synthesized one amino acid at a time using coarse-grain protein and ribosome models (represented here by the purple/grey lines and green/grey shapes, respectively). After synthesis, the monomer is released from the ribosome and its bulk dynamics then simulated for 5 μs. Random combinations of the final structures from bulk dynamics are then selected from the sets of 100 Monomer A and 100 Monomer B trajectories and their lowest-energy dimer configurations determined by temperature annealing. (d) Initial coarse grain and resulting all-atom structures of oligoribonuclease monomer before and after back-mapping. (e) Same as (d) but for a dimeric oligoribonuclease structure.

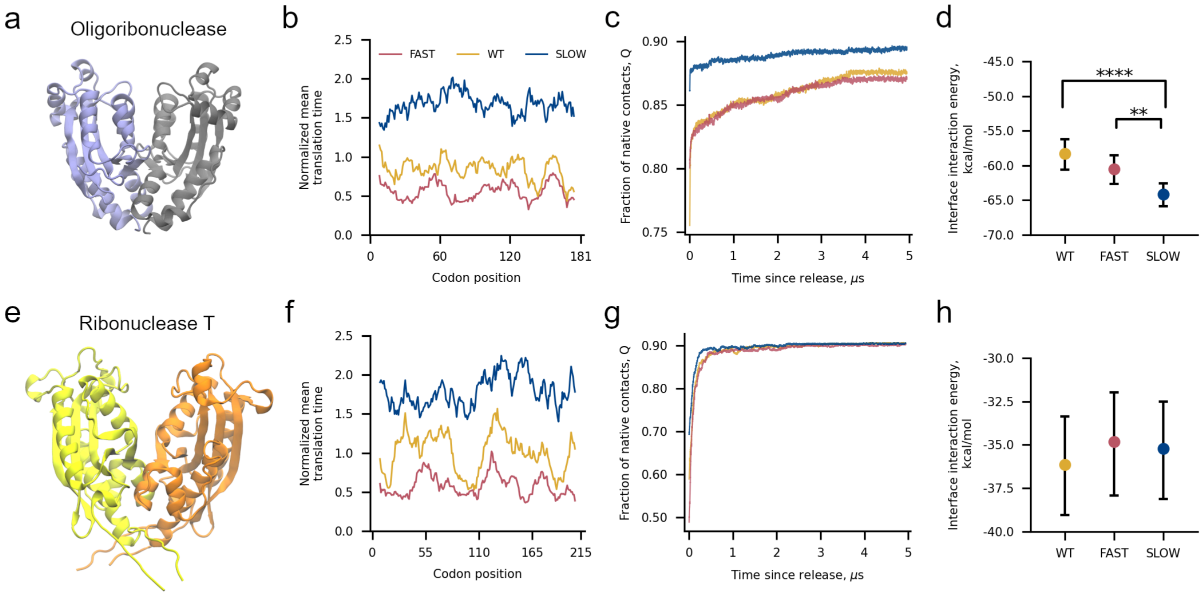

Figure 2. **Altering translation kinetics affects the binding affinity of the oligoribonuclease homodimer**. (a) 3D structure of oligoribonuclease from PDB ID 1YTA with Monomers A and B colored light purple and grey, respectively. (b) Mean translation time of codon positions, normalized by the average codon translation time across the 64 codons, within the fast-translating mutant (FAST, red), wildtype (WT, yellow), and slow-translating mutant (SLOW, blue) mRNA sequences used for oligoribonuclease simulations smoothed with a 15-codon moving average. (c) Moving average of fraction of intra-monomer native contacts as a function of time since oligoribonuclease's release from the ribosome computed over all 200 trajectories (100 Monomer A + 100 Monomer B) for each mRNA. Individual time series were first smoothed by taking the mode within a sliding 15-ns window and then averaged together across all 200 monomer trajectories. (d) Average interface interaction energy between Monomers A and B computed over 200 different random pairs of monomers after annealing as described in Methods and Figure 1c. Error bars are 95% confidence intervals computed from bootstrapping 10^6 times. Brackets and asterisks indicate statistical significance of comparisons between means determined from permutation tests with 10^6 samples. (e) 3D structure of ribonuclease T from PDB ID 2IS3 with monomers A and B colored yellow and orange, respectively. (f) Normalized mean translation time of codon positions in ribonuclease T mRNAs used in our simulations. (g) Fraction of native contacts versus time computed from 200 Ribonuclease T trajectories for each mRNA template used. (h) Same as (d) but for interactions between monomers of ribonuclease T. One, two, three, or four asterisks indicate $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, or $p \leq 0.0001$, respectively.

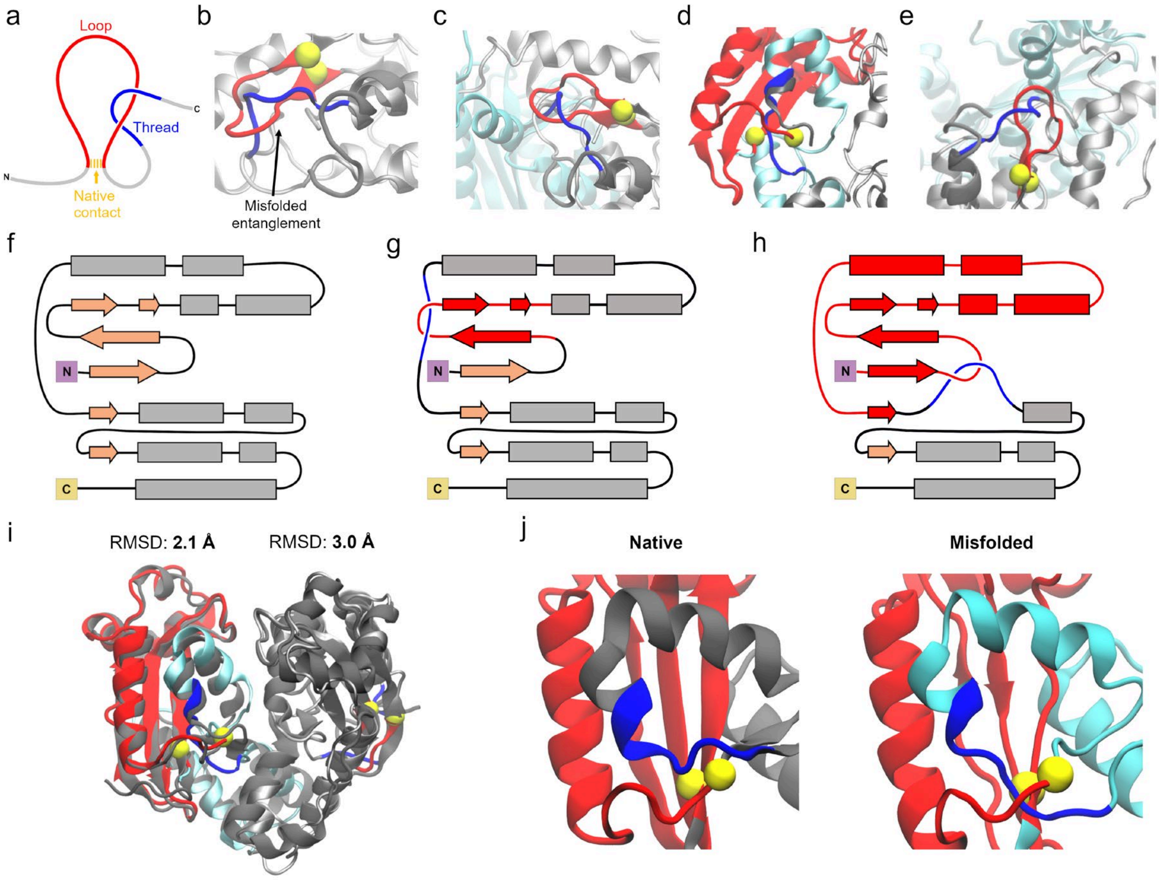

Figure 3. Entanglements in oligoribonuclease perturb its dimer interface. (a) Entanglements occur when a threading segment (blue) passes through a loop (red) formed by another segment of residues and closed by a native contact (yellow). (b) Structure of oligoribonuclease Monomer B in which residues 96-102 thread through the loop closed by the native contact between residues 31 and 41. Portions of the misfolded structure not involved in the entanglement are displayed in gray, while the location of the threading segment in the native state is shown in dark grey. This structure corresponds to M3 (see Supplementary Data File 1). (c) Same as (b) except in the context of a dimeric complex after annealing (structure D4), with Monomer A displayed in cyan. (d) Structure of oligoribonuclease dimer in which residues 125-129 of Monomer A thread through the loop closed by the contact between residues 9 and 103 (structure D5). (e) Structure of oligoribonuclease Monomer B with the same entanglement as (b) but in the context of a dimer in which Monomer A is also entangled (structure D5). (f) Secondary structure diagram of the native state of a monomer of oligoribonuclease. (g) Secondary structure diagram of the entanglement shown in (b), (c), and (e). (h) Same as (g) but for the entanglement shown in (d). (i) Alignment of the Monomer A and Monomer B structures shown in (d) and (e) to the native state structure indicates they are overall native-like with ≤3-Å C_α Root Mean Square Deviation (RMSD) from the crystal structure after backmapping. (j) Left: native state dimer interface of Monomer A. Right: interface view of the entangled structure of Monomer A from (d). Residues 125-129 of Monomer A thread through the loop from residues 9 to 103, disrupting the formation of a β-sheet that forms part of the dimer interface. Throughout all panels, loops, threads, and the native contact that closes the loop are colored red, blue, and yellow, respectively. Misfolded conformations of Monomers A and B are show in cyan and silver, respectively, and native conformations of the threading segment (as in (b)-(e)) or the overall structure (as in (i) and (j)) are shown in dark grey. Coarse-grain structures were back-mapped to all-atom resolution to generate visualizations.

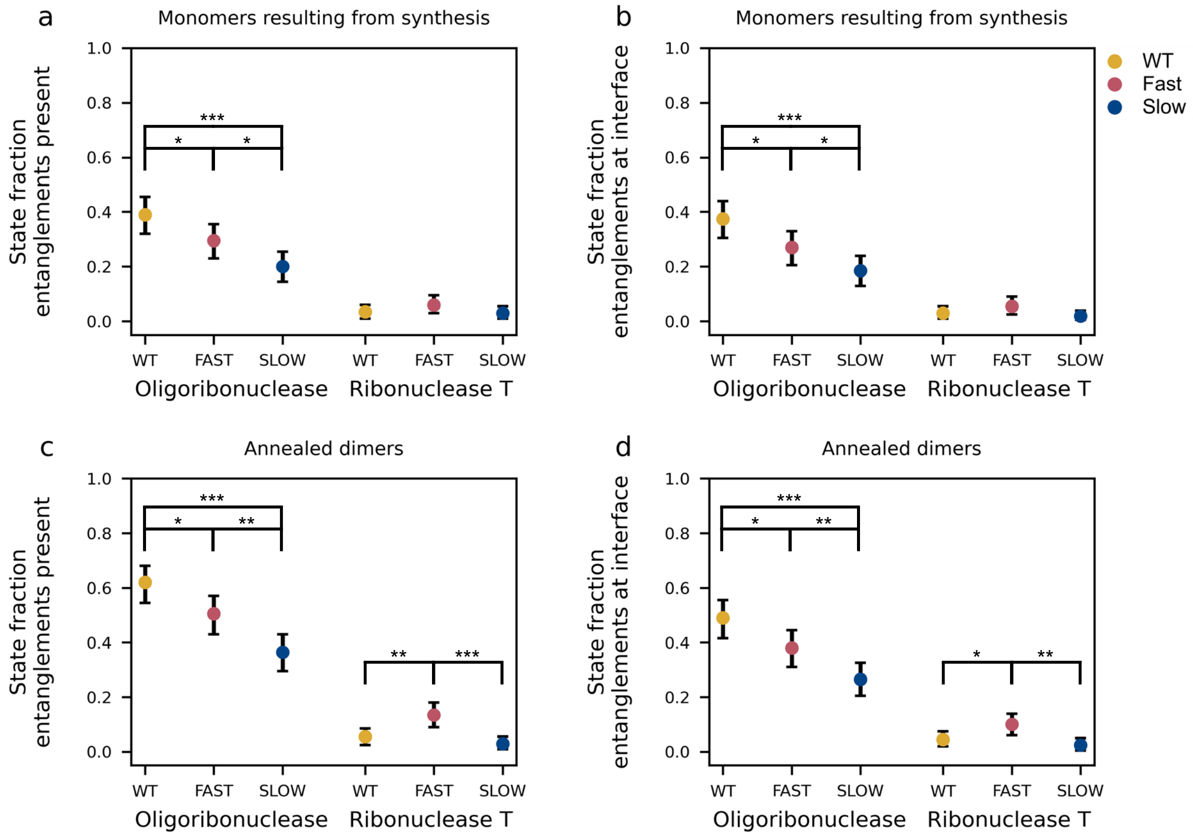

Figure 4. Changes in the population of self-entangled structures correlate with differences in dimer interaction energies. (a) Fraction of monomer structures of oligoribonuclease and ribonuclease T generated by coarse-grain synthesis, ejection, and post-translational dynamics simulations using the wildtype (WT), fast-translating mutant (FAST), and slow-translating (SLOW) mutant mRNAs that have a gain in entanglement relative to the native state somewhere in their structure. (b) Same as (a) but limited to the specific set of entanglements involving residues at the dimer interface. (c) Same as (a) but computed for the dimer structures generated by annealing random pairs of monomers. (d) Same as (c) but limited to the specific set of entanglements involving interface residues. All error bars are 95% confidence intervals computed from bootstrapping 10^6 times. Brackets and asterisks indicate the statistical significance of comparisons between means determined from permutation tests with 10^6 samples. One, two, or three, asterisks indicate $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$, respectively.

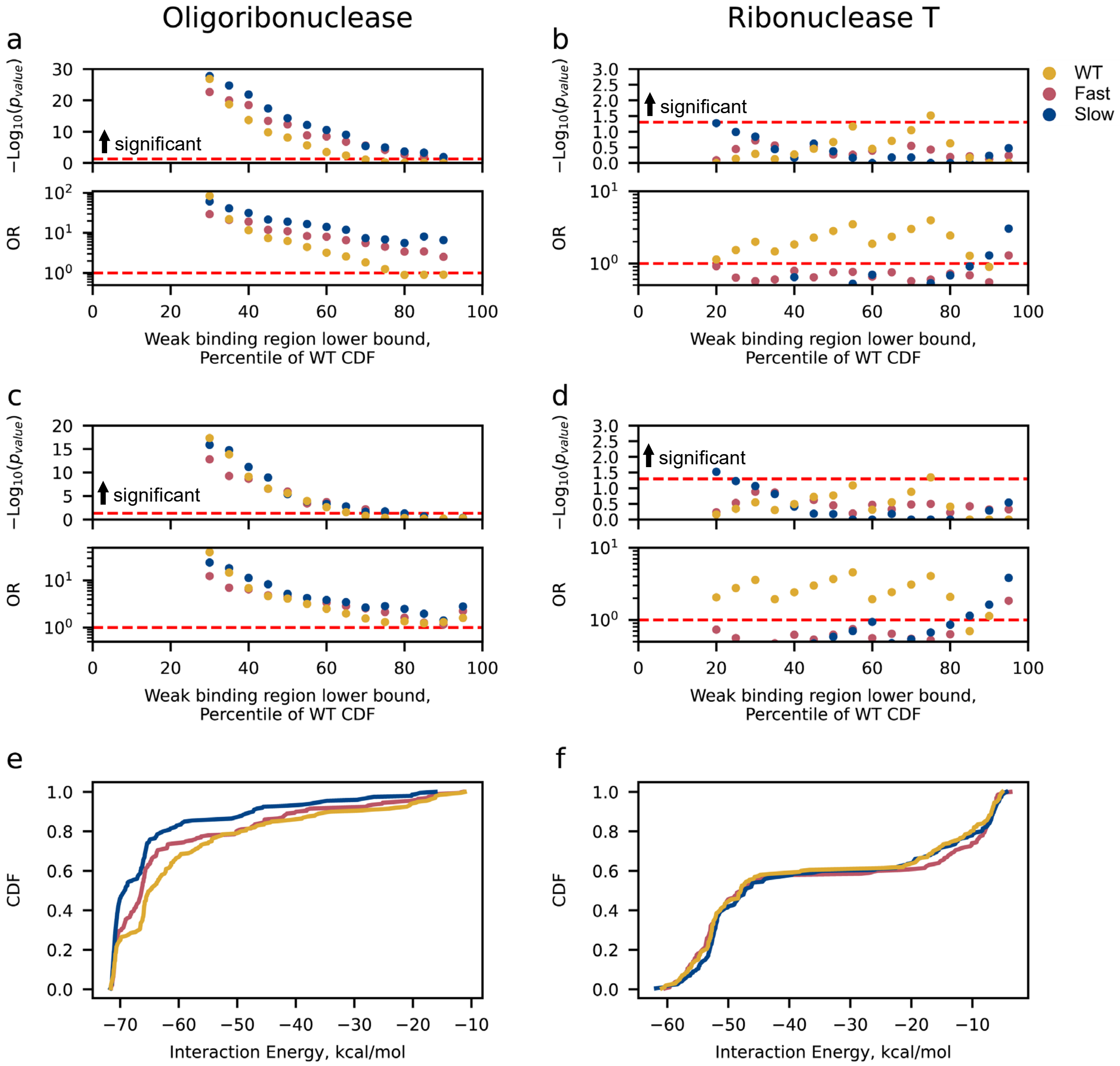

Figure 5. The presence of entanglements is strongly associated with weak dimer interaction energies. (a & b) $-\mathrm{Log}_{10}(p_{\mathrm{value}})$ and odds ratio resulting from Fisher's Exact Test applied to contingency tables for oligoribonuclease and ribonuclease T. The events were defined as (1) the presence of any change in entanglement of the annealed dimer relative to a reference native state and (2) the interaction energy falls within the weak binding region. To test the robustness of the results, the upper bound of the strong binding region was swept from the 5th percentile to the 95th percentile of the appropriate WT distribution of interaction energies. Regions where the odds ratio is not well defined near the extremes (*i.e.*, 0 entries in the contingency table) are not plotted, and the red dotted lines represent 1 and 0.05 on the odds ratio and p-value axes, respectively. (c & d) are the same as (a & b) but the first event is defined as the presence of entanglement at the interface (e & f). Cumulative distribution functions (CDFs) of the WT, Fast, and Slow ensemble interaction energies.

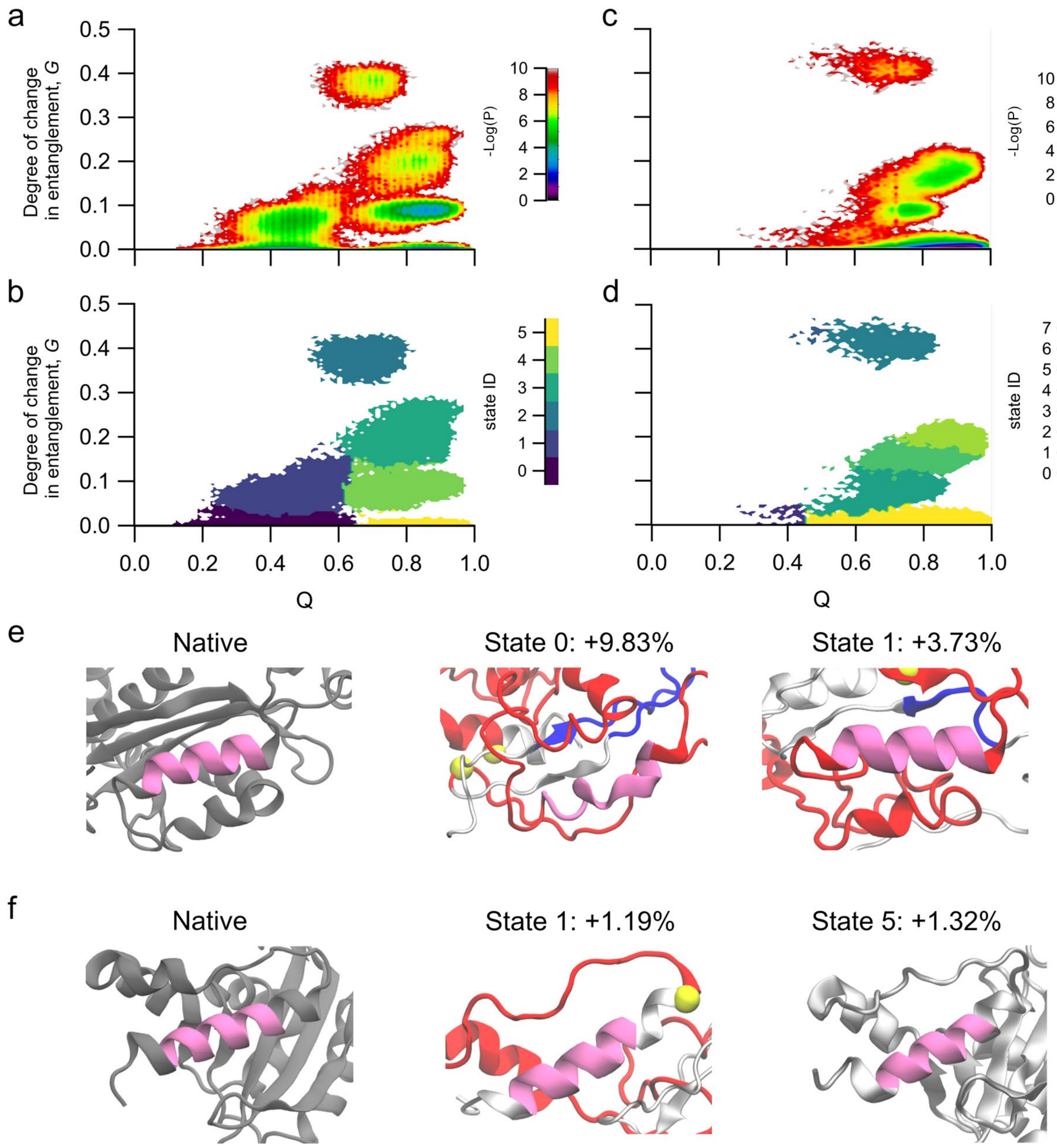

Figure 6. Predicted changes in solvent accessible surface area are consistent with LiP-MS refolding experiments. (a) and (c) are the $-\mathrm{Log}(P)$ surfaces spanning the fraction of native contacts, Q, and change in entanglement, G, for oligoribonuclease and ribonuclease T, respectfully, across WT, fast-translating, and slow-translating variants (see Figures S4 and S5 for results separated by variant). (b) and (d) are the resulting metastable states generated by Markov State models for the data in (a) and (c); the native-like states are 5 and 7, respectively. (e) Structures of oligoribonuclease in the native state and two entangled states with residues 84-94 highlighted in mauve. Percentages are the change in solvent-accessible surface area, relative to the average from the native state simulations, of residues 84 through 94 computed from the ensemble of structures arising from the wildtype translation schedule (see Table S5 for confidence intervals). For entangled conformations, the loop, thread, and contact closing the loop are shown in red, blue, and yellow, respectively (as in Figure 3). (f) Same as (e) but for residues 166-175. Note that State 5 does not contain an entanglement.

Supplementary Information

Synonymous mutations can alter protein dimerization through localized interface misfolding involving self-entanglements

Lan Pham Dang[ᵻ,1,2], Daniel Allen Nissley[ᵻ,3], Ian Sitarik[ᵻ,3], Quyen Vu Van[4], Yang Jiang[3], Philip To[5], Yingzi Xia[5], Stephen D. Fried[5,6], Mai Suan Li[1,4], Edward P. O'Brien[*3,7,8]

[1] Institute for Computational Sciences and Technology, Ho Chi Minh City, Vietnam
[2] Faculty of Physics and Engineering Physics, VNUHCM-University of Science, 227, Nguyen Van Cu Street, District 5, Ho Chi Minh City, Vietnam
[3] Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA
[4] Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland
[5] Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA
[6] Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218, USA
[7] Bioinformatics and Genomics Graduate Program, The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA
[8] Institute for Computational and Data Sciences, Pennsylvania State University, University Park, PA 16802, USA

[ᵻ] These authors contributed equally to this research project
* to whom correspondence should be addressed: epo2@psu.edu

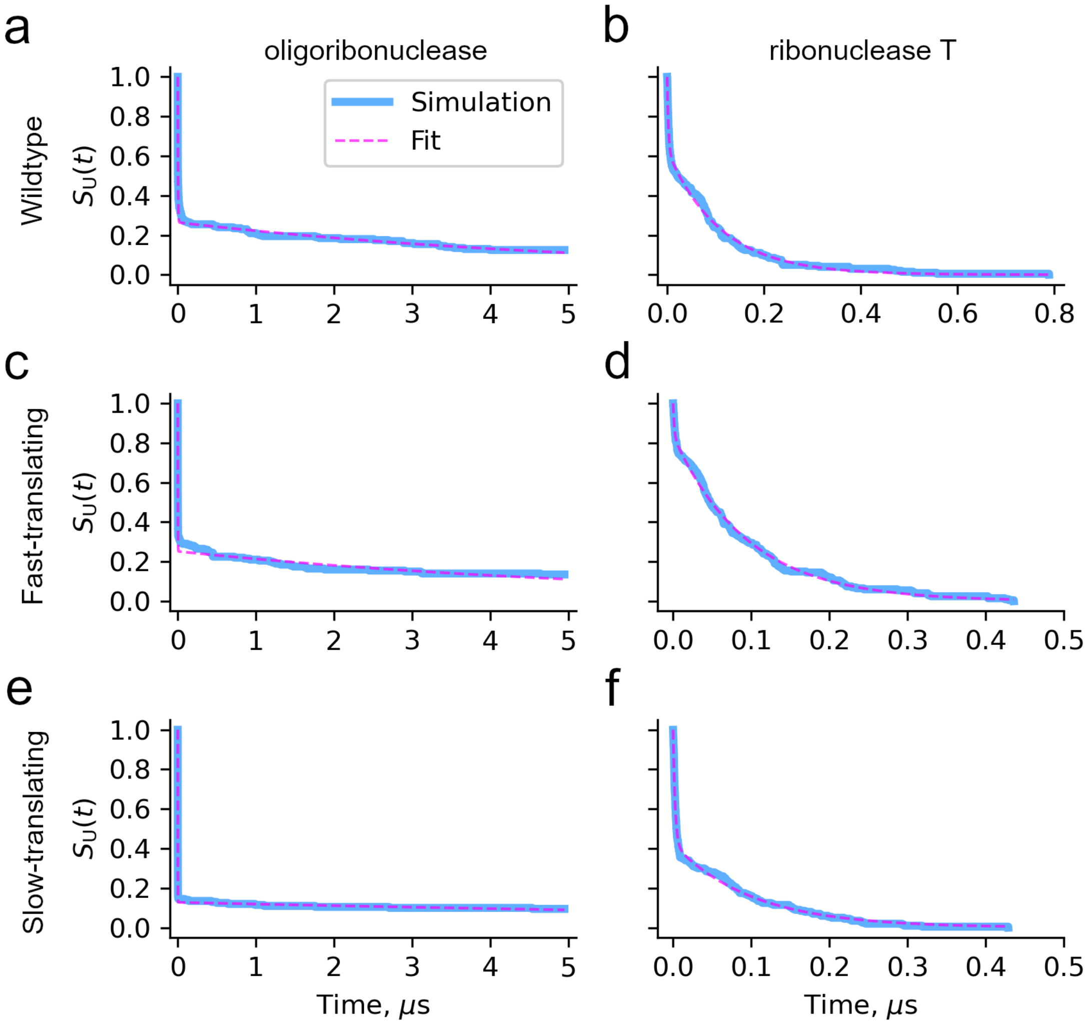

Figure S1. Calculation of post-translational folding timescales using kinetic curve fitting. (a) Survival probability of the unfolded state as a function of time since release from the ribosome (blue) and fit to the double-exponential equation $S_{\mathrm{U}}(t) = f_1 \exp(-k_1 t) + f_2 \exp(-k_2 t)$ (magenta dashed line, see Methods) for oligoribonuclease translated from its wildtype mRNA. (b) Same as (a) but for ribonuclease T wildtype mRNA simulations. (c) $S_{\mathrm{U}}(t)$ and fit for oligoribonuclease fast-translating mRNA simulations. (d) Same as (c) but for ribonuclease T fast-translating mRNA simulations. (e) $S_{\mathrm{U}}(t)$ and fit for oligoribonuclease slow-translating mRNA simulations. (f) Same as (e) but for ribonuclease T slow-translating mRNA simulations.

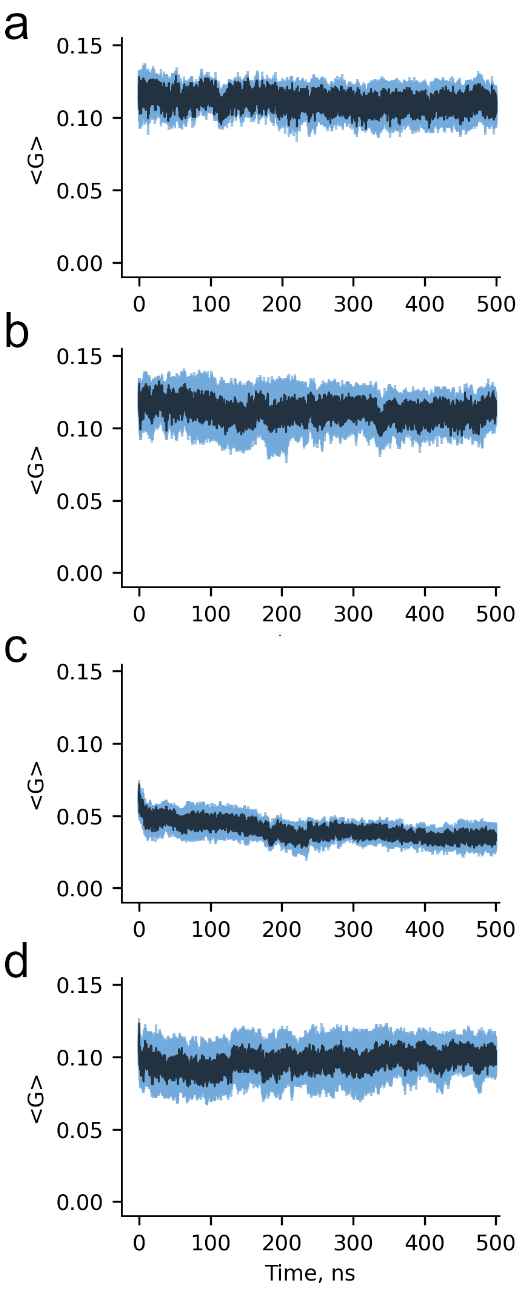

Figure S2. Entanglements in monomer and dimer structures of oligoribonuclease persist in aqueous all-atom simulations for 500 ns. (a) $\langle G \rangle$ (Eq. 5) as a function of simulation time computed over three all-atom trajectories each initiated from the same back-mapped entangled oligoribonuclease Monomer B structure (structure M3, see Supplementary Data File 1), in which residues 96-102 are entangled. The blue shaded region indicates the standard error of the average over the three trajectories. (b) Same as (a) except computed for dimer structure D4, in which residues 96-102 of Monomer B are entangled. The time series of $\langle G \rangle$ for Monomer A is not shown as it fluctuates around zero. (c) Same as (a) except for Monomer A of dimer structure D5, in which residues 125-129 are entangled. (d) Same as (a) except for Monomer B of dimer structure D5, in which residues 96-102 are entangled.

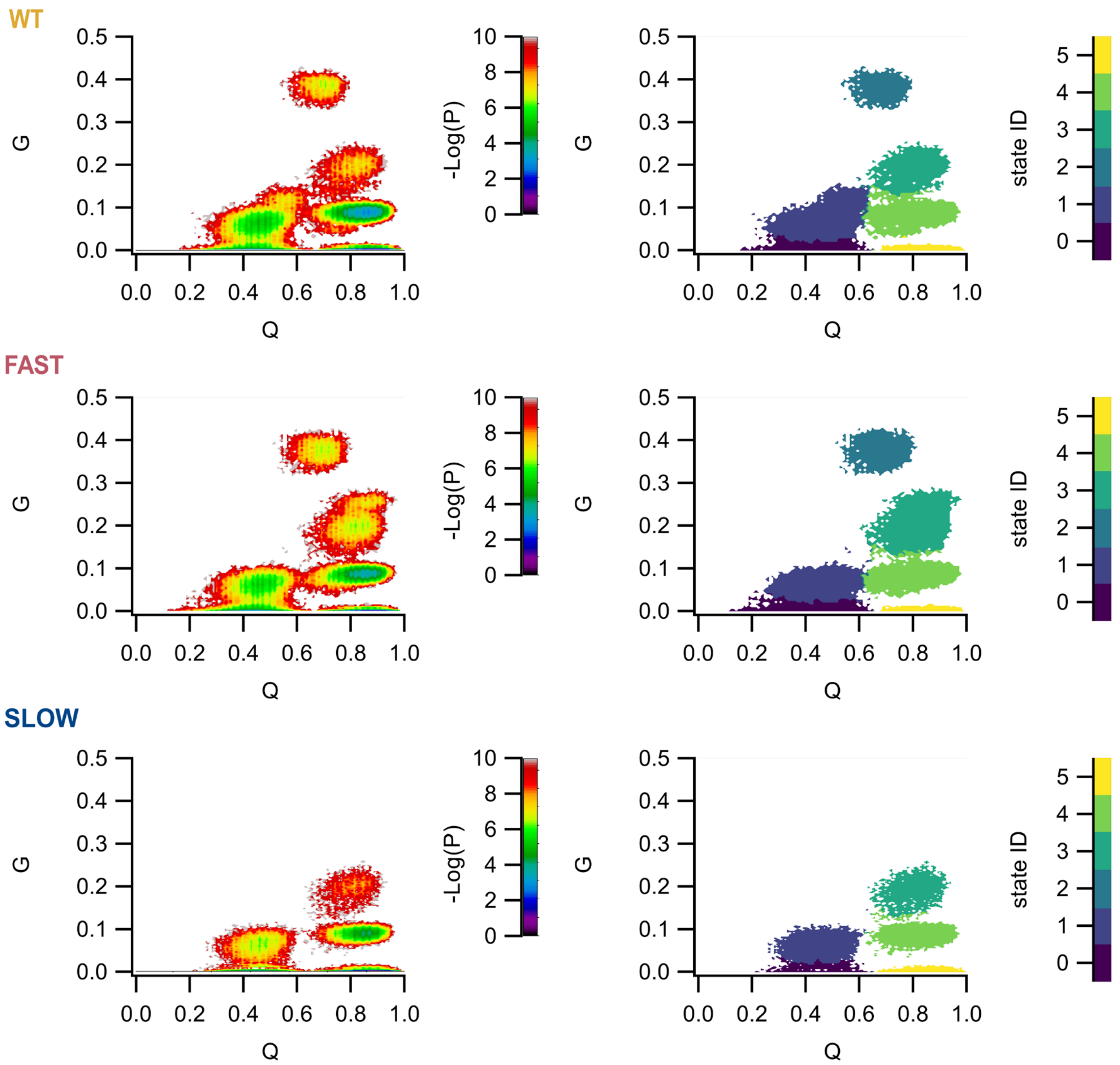

Figure S3. $-\mathbf{Log}(\boldsymbol{P})$ landscapes for oligoribonuclease synonymous variants. (Left) The $-\mathrm{Log}(P)$ landscapes for each variant for the last 100 ns of post-translational dynamics projected onto the fraction of native contacts (Q) and change in entanglement (G) order parameters. (Right) the same landscape as on the left but coloured to represent the resulting meta-stable states obtained after Markov state modeling (see Methods).

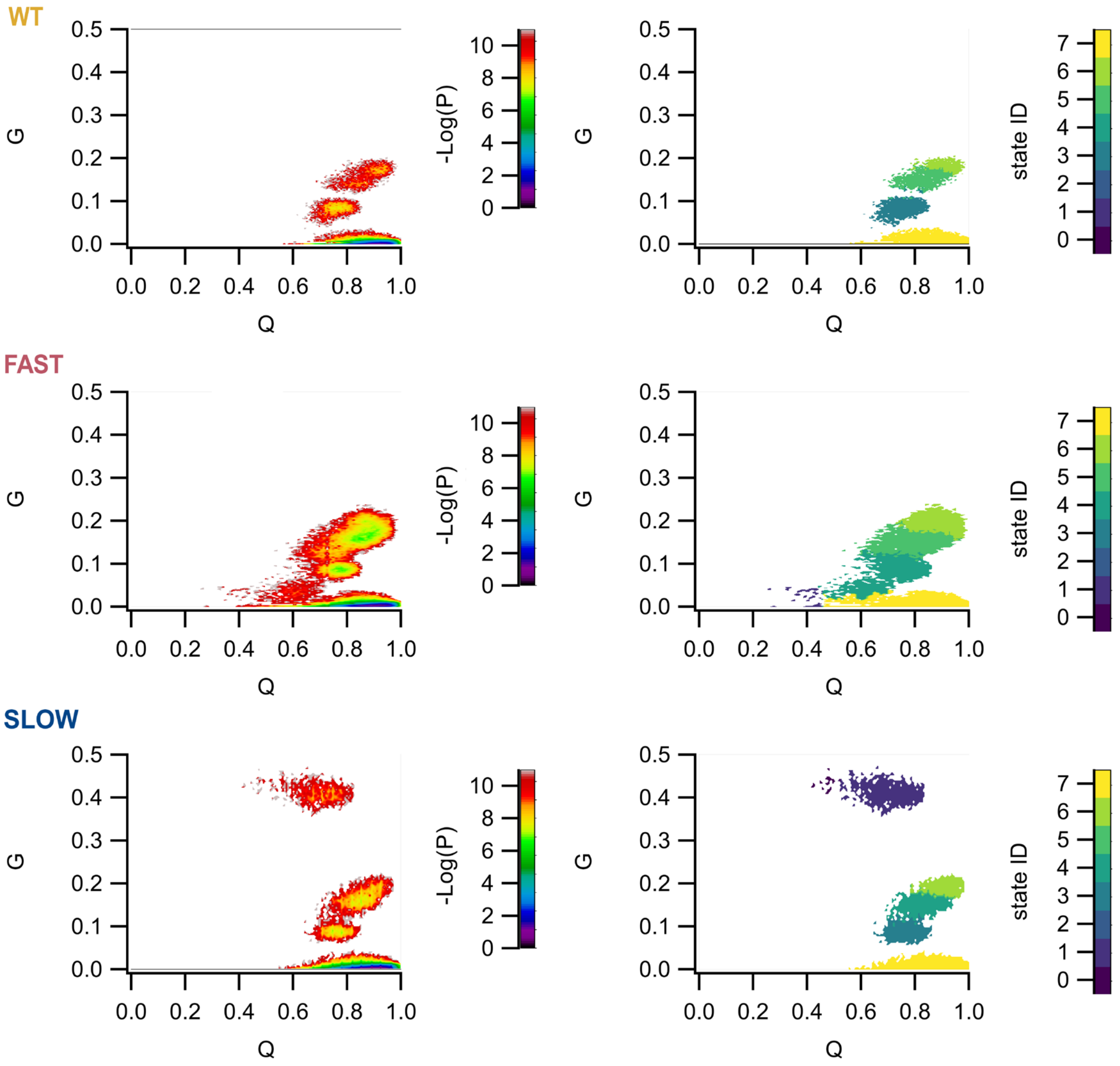

Figure S4. $-\mathbf{Log}(\boldsymbol{P})$ landscapes for ribonuclease T synonymous variants. (Left) The $-\mathrm{Log}(P)$ landscapes for each variant for the last 100 ns of post-translational dynamics projected onto the fraction of native contacts (Q) and change in entanglement (G) order parameters. (Right) the same landscape as on the left but coloured to represent the resulting meta-stable states obtained after Markov state modeling (see Methods).

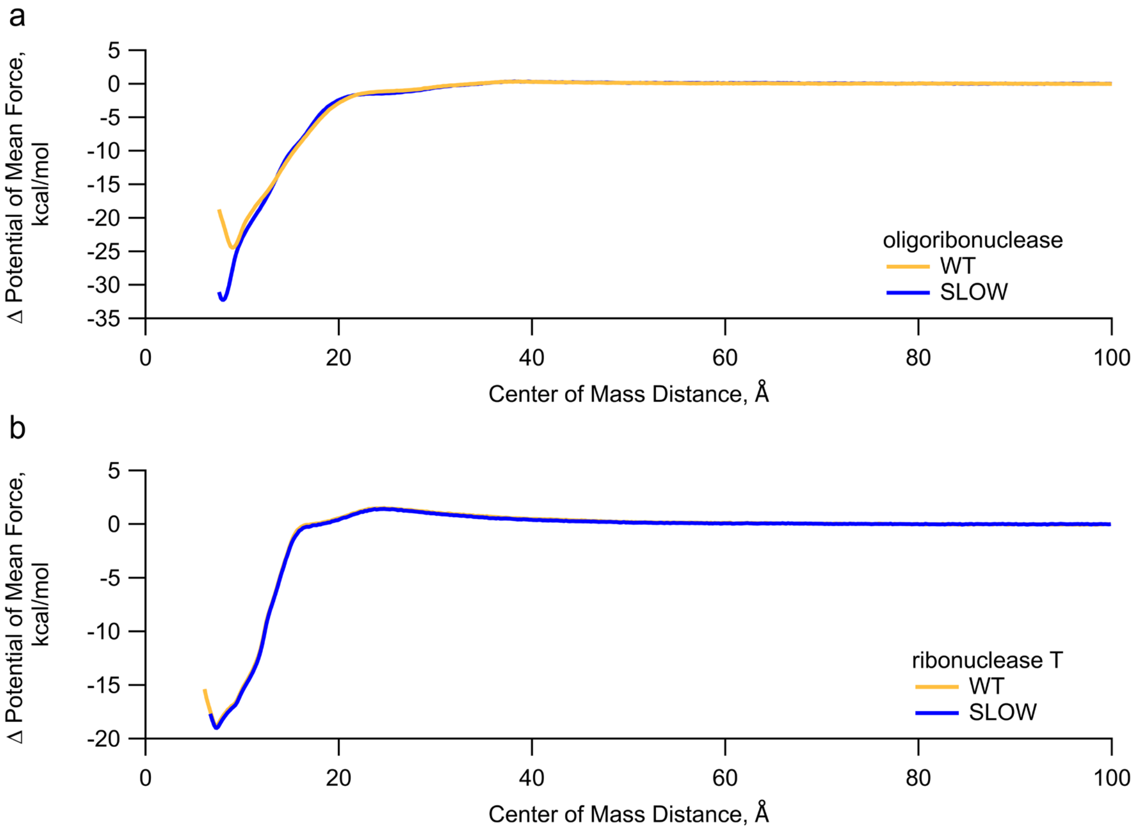

Figure S5. **$\Delta\mathrm{PMF}$ profiles for representative dimeric structures.** The $\Delta\mathrm{PMF}$ as a function of the center of mass (CoM) distance for the selected annealed dimers of WT and SLOW variants of oligoribonuclease (a) and (b) ribonuclease T. Where $\Delta\mathrm{PMF} = \mathrm{PMF} - \mathrm{PMF}(\mathrm{CoM} = 100\text{Å})$. Shaded regions on the curves represents standard errors obtained from 200 bootstrapped PMFs.

Table S1. Kinetic fitting parameters to the equation $S_{\mathrm{U}}(t) = f_1 \exp(-k_1 t) + f_2 \exp(-k_2 t)$ for oligoribonuclease and ribonuclease T post-translational folding time courses in Figure S1.

Protein	mRNA	f_1	k_1, μs^{-1}	τ_1, μs	f_2	k_2, μs^{-1}	τ_2, μs	Pearson R^2
oligoribonuclease	wildtype	0.26	1.74 x 10^{-1}	5.75	0.74	2.89 x 10^{2}	3.46 x 10^{-3}	0.95
	fast	0.25	1.65 x 10^{-1}	6.06	0.75	6.26 x 10^{2}	1.60 x 10^{-3}	0.86
	slow	0.13	7.10 x 10^{-2}	1.41 x 10^{1}	0.87	1.19 x 10^{3}	8.40 x 10^{-4}	0.86
ribonuclease T	wildtype	0.63	9.07	1.10 x 10^{-1}	0.37	3.85 x 10^{2}	2.60 x 10^{-3}	0.99
	fast	0.84	1.05 x 10^{1}	9.52 x 10^{-2}	0.16	6.56 x 10^{2}	1.52 x 10^{-3}	1.00
	slow	0.42	9.88	1.01 x 10^{-1}	0.58	3.91 x 10^{2}	2.56 x 10^{-3}	0.99

Table S2. Peptides detected by LiP-MS 120 min after dilution jump established refolding conditions for oligoribonuclease (Orn) and ribonuclease T (Rnt).

Protein Name	Gene name	Residue(s)	$\log_2 \frac{R}{N}$	$-\log_{10} p$
Oligoribonuclease	Orn	[119-130]	-10.3114	1.3226
		[166-175]*	-7.2116	2.4444
		[144-154]	-1.3235	1.2013
		[102-114]	-1.2497	0.6847
		[84-94]*	-1.007	3.5499
		[131-138]	-0.8993	2.0449
		[143-154]	-0.6540	1.7436
		[77-94]	0.4510	0.5612
Ribonuclease T	Rnt	[190-200]	-9.3280	1.2855
		[33-45]	-0.8579	0.6775
		[A40]	0.8072	1.5798
		[S2]	-0.1413	0.1037
		[204-215]	0.5255	1.1250

*peptides with a significant difference in population between refolded and native samples *and* a greater than 2 fold change between the refolded and native samples.

Table S3. Relative increase in solvent-accessible surface area of significant LiP-MS peptides for oligoribonuclease with 95% confidence intervals from 10^6 bootstraps listed in square brackets.

Synonymous Mutant	State ID	Proportion of conformations in state*	$\zeta_{[84-94]}$, %	$\zeta_{[166-175]}$, %
WT	0	0.072	9.83 [9.59, 10.1]	-6.98 [-7.22, -6.74]
	1	0.071	3.73 [3.62, 3.85]	1.19 [0.98, 1.40]
	2	0.010	-3.47 [-3.62, -3.32]	-0.69 [-1.03, -0.36]
	3	0.009	6.25 [6.04, 6.45]	-12.6 [-12.9, -12.2]
	4	0.261	-0.57 [-0.60, -0.55]	-0.33 [-0.40, -0.27]
	5 (Native)	0.577	-0.21 [-0.23, -0.19]	1.32 [1.28, 1.36]
Fast	0	0.103	10.01 [9.81, 10.2]	-6.24 [-6.44, -6.05]
	1	0.055	5.59 [5.42, 5.76]	2.20 [1.97, 2.42]
	2	0.015	-3.23 [-3.35, -3.11]	-2.75 [-3.03, -2.46]
	3	0.025	5.14 [5.01, 5.26]	-8.85 [-9.09, -8.62]
	4	0.161	-0.06 [-0.10, -0.01]	0.21 [0.13, 0.29]
	5 (Native)	0.641	-0.16 [-0.17, -0.14]	1.18 [1.14, 1.22]
Slow	0	0.056	9.13 [8.88, 9.38]	-5.02 [-5.27, -4.76]
	1	0.038	9.24 [8.96, 9.52]	1.95 [1.68, 2.22]
	2	0.000	-	-
	3	0.009	6.37 [6.17, 6.58]	-12.2 [-12.5, -11.8]
	4	0.130	-0.69 [-0.73, -0.66]	0.72 [0.63, 0.80]
	5 (Native)	0.766	-0.12 [-0.14, -0.11]	1.25 [1.22, 1.29]

*sparsely populated (probability <0.001) states are omitted from this analysis

Table S4. Relative increase in solvent-accessible surface area of non-significant LiP-MS peptides for ribonuclease T with 95% confidence intervals from 10^6 bootstraps.

Synonymous Mutant	**State ID**	**Proportion of conformations in state***	$\zeta_{[S2]}$, %	$\zeta_{[204-215]}$, %
WT	0	0.000	-	-
	1	0.000	-	-
	2	0.000	-	-
	3	0.000	-	-
	4	0.010	-0.59 [-1.25, 0.07]	-0.58 [-0.87, -0.28]
	5	0.006	-0.63 [-1.42, 0.16]	-1.49 [-1.87, -1.12]
	6	0.003	-1.08 [-2.31, 0.13]	-0.06 [-0.60, 0.48]
	7 (Native)	0.981	-0.44 [-0.51, -0.37]	-1.01 [-1.04, -0.98]
Fast	0	0.000	-	-
	1	0.000	-	-
	2	0.000	-	-
	3	0.000	-	-
	4	0.014	-0.09 [-0.62, 0.43]	-0.94 [-1.19, -0.69]
	5	0.019	-0.36 [-0.83, 0.10]	-1.07 [-1.29, -0.85]
	6	0.011	-0.56 [-1.16, 0.04]	-0.40 [-0.67, -0.13]
	7 (Native)	0.956	-0.42 [-0.48, -0.35]	-1.07 [-1.10, -1.04]
Slow	0	0.000	-	-
	1	0.000	-	-
	2	0.000	-	-
	3	0.005	0.44 [-0.49, 1.36]	-1.68 [-2.13, -1.23]
	4	0.005	-0.64 [-1.55, 0.26]	-0.71 [-1.13, -0.29]
	5	0.006	-0.31 [-1.11, 0.46]	-0.76 [-1.13, -0.38]
	6	0.003	-0.04 [-1.21, 1.11]	-1.05 [-1.61, -0.51]
	7 (Native)	0.981	-0.39 [-0.45, -0.32]	-1.00 [-1.03, -0.97]

*sparsely populated (probability <0.001) states are omitted from this analysis