Article

Thermodynamic Protein Destabilization by GFP Tagging: A Case of Interdomain Allostery

Miri Sokolovski,[1] Arnab Bhattacherjee,[1] Naama Kessler,[1] Yaakov Levy,[1,*] and Amnon Horovitz[1,*]
[1]Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

ABSTRACT The Engrailed Homeodomain (EnHD) transcription factor of *Drosophila melanogaster* was fused to the enhanced green fluorescent protein (eGFP) either at its C- or N-terminus via three- or ten-residue flexible linkers. Here, we show that EnHD undergoes destabilization upon fusing it to eGFP regardless of the linker length used and whether the tethering is to its N- or C-terminus. The destabilization is reflected in melting points that are lower by up to 9°C. Thermodynamic analysis and coarse-grained molecular dynamic simulations indicate that this destabilization is due to eGFP-promoted entropic stabilization of the denatured state ensemble of EnHD. Our results provide, therefore, an example for destabilizing interdomain allostery. They are also important given the widespread use of eGFP tagging in cell biology, as they indicate that such tagging can cause unintended protein destabilization and concomitant effects.

INTRODUCTION

It has been estimated that multidomain proteins comprise ~65% and 40% of the proteomes of eukaryotes and prokaryotes, respectively (1). Multidomain proteins are more aggregation-prone because of the high effective protein concentration near each domain (2). Aggregation of multidomain proteins can occur as a result of kinetic partitioning between folding and aggregation during, for example, translation or transport across membranes. Alternatively, unfolding and aggregation can also occur as a result of thermodynamic destabilization that is triggered by a change in environmental conditions (e.g., pH, temperature, or denaturants) and/or by interactions between neighboring domains. Such domain-domain interactions can be stabilizing or destabilizing and specific or nonspecific. The C-terminal domain of γB-crystallin, for example, is stabilized by ~4 kcal mol^{-1} because of specific interactions with the folded state of the N-terminal domain (3), whereas spectrin domains are stabilized by nonnative interactions with the I27 titin domain when it is fused to their C-termini (4). An example for a destabilizing interdomain interaction is the utrophin tandem calponin-homology domain where the C-terminal domain appears to be thermodynamically and kinetically more stable than the full-length protein (5). An even more striking example for a destabilizing interaction is the thermodynamic tug-of-war described for an engineered chimera of barnase and GCN4 that causes one domain to be unfolded when its neighboring domain is folded (6). Computational work has indicated that protein destabilization can also occur because of tethering of a domain such as ubiquitin to an internal position in the protein (7).

Given that protein aggregation is harmful to cells and associated with many diseases (8), it is not surprising that various mechanisms have evolved for reducing the risk of aggregation. Such mechanisms include selection for 1) neighboring domains with low sequence identity (9) and 2) the N-terminal domains of two-domain proteins to be shorter than their neighboring C-terminal counterparts (10). It has also been suggested (11) that the allosteric mechanism of the eukaryotic chaperonin TRiC/CCT has evolved to be sequential, and not concerted as in the case of the prokaryotic GroEL, to facilitate domain-by-domain release and folding of its substrates and, thus, increase the folding efficiency of multidomain proteins. Such safeguards against aggregation are, however, unlikely to be present in engineered multidomain proteins that have not been subjected to natural selection. Because of the very extensive use of multidomain proteins such as fluorescent fusion proteins in biological research (12), we decided to test whether tagging with the enhanced green fluorescent protein (eGFP) can affect the tagged protein's stability. We chose to fuse eGFP with the Engrailed Homeodomain (EnHD) transcription factor of *Drosophila melanogaster* (Fig. 1). EnHD is a 61-residue, three-helix bundle protein flanked by substantially disordered tails at both the C- and N-termini. It folds on a microsecond timescale via a folding intermediate and its folding and stability properties have been studied extensively by both experiment and computation (13–15). EnHD was chosen for this study because it has a melting temperature that is lower than and distinct from that of eGFP and it undergoes heat-induced folding transitions that are reversible owing at least, in part, to its fast folding kinetics and residual structure in the denatured state

Submitted February 3, 2015, and accepted for publication April 20, 2015.
*Correspondence: amnon.horovitz@weizmann.ac.il or koby.levy@weizmann.ac.il
Editor: Nathan Baker.

0006-3495/15/09/1157/6
http://dx.doi.org/10.1016/j.bpj.2015.04.032

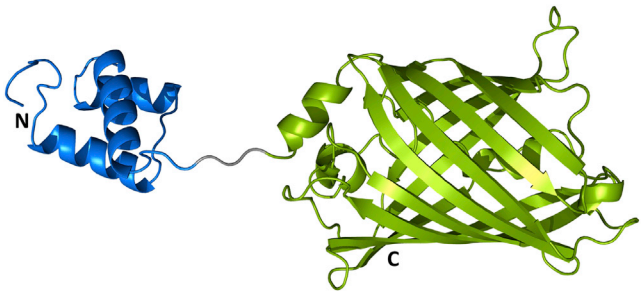

FIGURE 1 Model of the structure of EnHD-Gly_3-eGFP. The C-terminus of EnHD (*blue*, PDB ID: 2JWT) is fused to the N-terminus of eGFP (*green*, PDB ID: 4EUL) via a linker of three glycine residues (*gray*). To see this figure in color, go online.

(13,14). Here, we show that EnHD undergoes significant destabilization upon fusing it to eGFP either at its C- or N-terminus. Coarse-grained molecular dynamics (MD) simulations indicate that this destabilization is due to eGFP-promoted entropic stabilization of EnHD's denatured state ensemble.

MATERIALS AND METHODS

Molecular biology

All the mutagenesis, insertions, and gene cloning were carried out using restriction-free cloning (16,17) and confirmed by DNA sequencing of the entire genes. The mutation Ala-206→Lys in eGFP that stabilizes its monomeric state (18) was introduced into the gene coding for this protein in pET-22b using the forward and backward primers: 5′-CCTGAGCACCCA GTCCAAACTGAGCAAAGACCCCAA-3′ and 5′-GATCAGCCATGTTC GCGGCTTACTTATACAGCTCGTCCATGC-3′, respectively. A His_6-tag was added to the C-terminus of EnHD using the gene for this protein in pET-27b and the forward and backward primers: 5′-GCGTTCTCGAG CGAGCAGTTGGCCCGCCTGAAGCGCGAATTCAACGAGAATCGCT ACC-3′ and 5′-CTCGAGTGCGGCCGCTTAATGGTGATGGTGATGGT GCGTCGACTTCTTGATCTTGG-3′, respectively. The EnHD-Gly_3-eGFP-His_6 chimera was constructed by amplifying the gene for eGFP in pET-22b using the forward and backward primers: 5′-GCGCGCCAAGATCAAGA AGTCGACGGGCGGTGGCATGGTTAGCAAGGGCGAGGAGCTGTTC ACC-3′ and 5′- GCCGCTTAATGGTGATGGTGATGGTGCTTATACAG CTCGTCCATGCCGAGAG-3′, respectively. The purified polymerase chain reaction fragment was then used to generate the gene for EnHD-Gly_3-eGFP-His_6 in pET-27b. The eGFP-Gly_3-EnHD-His_6 chimera was constructed by amplifying the gene for eGFP in pET-22b using the forward and backward primers: 5′-GTTTAACTTTAAGAAGGAGATATACATA TGGTTAGCAAGGGCGAGGAGCTGTTCAC-3′ and 5′-CGCGGTGCGT GGACGCTTCTCCATGCCACCGCCCTTATACAGCTCGTCCATGCCG AG-3′, respectively. The eGFP-$(Gly\text{-}Gly\text{-}Ser)_3$-Gly-EnHD-$His_6$ chimera was generated from the gene for eGFP-Gly_3-EnHD-His_6 in pET-27b using the forward and backward primers: 5′-CGGCATGGACGAGCTGTATAA GGGCGGTAGCGGCGGTAGCGGCGGTAGCGGCATGGAGAAGCGT CCACGCACCGCG-3′ and 5′-CTCGAGTGCGGCCGCTTAATGGTGA TGGTGATGGTGCGTCGACTTCTTGATCTTGG-3′, respectively. The EnHD-$(Gly\text{-}Gly\text{-}Ser)_3$-Gly-eGFP-$His_6$ chimera was generated from the gene for EnHD-Gly_3-eGFP-His_6 in pET-27b using the forward and backward primers: 5′-GCGTTCTCGAGCGAGCAGTTGGCCCGCCTGAAG CGCGAATTCAACGAGAATCGCTACC-3′ and 5′-GAACAGCTCCTCG CCCTTGCTAACCATGCCGCTACCGCCGCTACCGCCGCTACCGCCC GTCGACTTCTTGATCTTGGC-3′, respectively.

Expression and purification of EnHD

Escherichia coli Rosetta (DE3) cells harboring the pET-27b plasmid that contains the gene for EnHD were grown in Luria broth medium containing 50 μg/ml kanamycin at 37°C until an $O.D._{600nm}$ of 0.6 was reached and protein expression was then induced by adding 1 mM isopropylthio-β-galactoside (IPTG). The cells were then further grown overnight at 16°C and harvested. The cell pellets were stored at −80°C until further use. Purification of EnHD was carried out by resuspending cells in 30 mM Tris buffer (pH 8.45) containing 100 mM NaCl, 10 mM imidazole, and 10 mM β-mercaptoethanol (buffer A) to which was added the cOmplete, EDTA-free, protease inhibitor cocktail (Roche Applied Science, Penzberg, Germany). The cells were then disrupted by sonication and the lysate was clarified by centrifugation. The supernatant was incubated for 1 h at 4°C with Ni-NTA resin that was equilibrated with buffer A. The resin was then transferred to a column and washed with buffer A. Elution of EnHD was carried out by washing the column with buffer A containing 200 mM imidazole. The eluted protein was dialyzed overnight at 4°C against 50 mM Tris buffer (pH 7.5) containing 100 mM NaCl and loaded onto a C-18 high-performance liquid chromatography column in the presence of 0.1% trifluoroacetic acid and 5% (v/v) acetonitrile. The protein was eluted at 45–50% acetonitrile using a 114 ml gradient of 5–70% acetonitrile. Fractions that contain purified EnHD were lyophilized and stored at −80°C until further use. The Leu-16→Ala EnHD mutant was purified like wt EnHD except that the cell lysis and the Ni-NTA affinity chromatography steps were carried out in the presence of 8 M urea.

Expression and purification of eGFP

eGFP was purified from *E. coli* Rosetta (DE3) cells harboring the pET-22b plasmid that contains the gene for this protein. Cells were grown in 2× tryptone yeast extract medium containing 100 μg/ml ampicillin at 37°C until an $O.D._{600nm}$ of 0.6 was reached and protein expression was then induced by adding 0.5 mM IPTG. The cells were then grown for another 4 h at 37°C and harvested. The cell pellets were stored at −80°C until further use. Purification of eGFP was carried out by resuspending the cells in 50 mM Tris buffer (pH 7.5) containing 100 mM NaCl, 10 mM imidazole, 0.1 mM DTT (buffer B), and the cOmplete, EDTA-free, inhibitor cocktail (Roche Applied Science). The cells were then disrupted using a French press and sonication and the lysate was clarified by centrifugation. The supernatant was loaded onto a 5 mL HisTrap HP column (Amersham Pharmacia, Uppsala, Sweden) and eGFP was eluted with a 10–200 mM imidazole gradient in buffer B. The eluted protein was transferred to 50 mM Tris buffer (pH 7.5) containing 10 mM NaCl and 0.1 mM DTT (buffer C) using a PD-10 desalting column (GE Healthcare, Uppsala, Sweden) and then loaded onto a 5 mL HiTrap Q FF column (Amersham Pharmacia). eGFP was eluted from the Q FF column with a 0.01–1 M NaCl gradient in buffer C. The eluted protein was then loaded onto a Superdex 75 column (Amersham Pharmacia) equilibrated with 10 mM phosphate buffer (pH 7) containing 0.1 mM DTT. Fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis and those containing eGFP were combined, divided into aliquots that were snap-frozen in liquid nitrogen, and stored at −80°C.

Expression and purification of chimeras of EnHD and eGFP

E. coli Rosetta (DE3) cells harboring the pET-27b plasmid that contains the gene for the appropriate chimera were grown in Luria broth medium containing 50 μg/ml kanamycin at 37°C until an $O.D._{600nm}$ of 0.6 was reached. 1 mM IPTG was then added to induce protein expression and the cells were grown for another 4 h at 37°C before harvesting. The pellets were stored at −80°C until further use and then resuspended in buffer A and disrupted by sonication. The lysate was clarified by centrifugation. The pellets were

then resuspended in buffer A containing 8 M urea and centrifuged to remove cell debris. The supernatant was loaded onto a 5 mL HisTrap HP column (Amersham Pharmacia) and the chimera was eluted with a 10–250 mM imidazole gradient in buffer A containing 8 M urea. The eluted protein (~20 ml) was added dropwise slowly to 700 ml of 50 mM Tris-HCl buffer (pH 7.5) containing 2.5 M NaCl and 0.1 mM DTT (buffer D). The protein was then loaded onto a butyl sepharose column and eluted from the column with a 2.5–0.05 M NaCl gradient in buffer D. The eluted protein was then transferred to 10 mM phosphate buffer (pH 7) containing 100 mM NaCl and 0.1 mM DTT using a PD-10 desalting column (GE Healthcare). Aliquots of the protein were snap-frozen in liquid nitrogen and stored at −80°C.

Circular dichroism (CD) measurements

All CD measurements were preformed on a Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK). The temperature was controlled using a TC125 temperature controller (Quantum Northwest, Liberty Lake, Washington) and monitored with a temperature probe in the sample. All the measurements were made using a quartz suprasil cell of 1 mm path length (Hellma, Forest Hills, New York) and the protein concentration was 12 μM in 10 mM phosphate buffer (pH 7) containing 0.1 M NaCl and 0.0125% Tween. Changes in the far ultraviolet (UV) CD signal at 222 nm were monitored as a function of increasing temperature from 10 to 94°C (steps of 2°C) with an equilibration time of 5 min at each temperature before taking a reading. The melts of EnHD in the chimera or by itself were fully reversible until 65°C and 85°C, respectively. Wavelength scans were performed at 25°C and 65°C. The signal was monitored at a wavelength range from 200 to 240 nm in steps of 1 nm and the average recording time was 3 s. The CD spectra shown are averages of three consecutive scans that were corrected by subtracting the spectra of buffer alone.

Data fitting

Plots of the CD signal at 222 nm, Y, as a function of the temperature, T, were fitted to Eq. 1:

$$Y = \frac{Y_N + aT + (Y_U + bT)e^{\frac{\Delta H_m}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)+\frac{\Delta Cp}{R}\left(\frac{T_m}{T}+\ln\left(\frac{T}{T_m}\right)-1\right)}}{1+e^{\frac{\Delta H_m}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)+\frac{\Delta Cp}{R}\left(\frac{T_m}{T}+\ln\left(\frac{T}{T_m}\right)-1\right)}}, \quad (1)$$

where a and b are the slopes and Y_N and Y_U are the intercepts of the pre- and posttransition baselines, respectively, R is the gas constant, T_m is the melting temperature, ΔH_m is the denaturation enthalpy at T_m and ΔCp the change in heat capacity. The slopes, a and b, were obtained by linear fitting of the pre- and posttransition baselines and then fixed when fitting each phase separately using Eq. 1.

Computational methods

The solution structure of EnHD (Protein Data Bank (PDB) Id: 2JWT) was used to study its folding thermodynamics when it is alone or fused at its C- or N-terminus to the 239 residues long eGFP (PDB Id: 4EUL) via a flexible linker containing 3 glycine residues. The three and seven residues at the N- and C-terminus of eGFP that cannot be seen in its crystal structure were added by modeling them as disordered segments. A coarse-grained model was used in which the backbone atoms of each residue are represented by a single bead at the position of the C_α atom and the side chain is represented by another bead at the position of the C_β atom that is bonded to the backbone bead. The beads representing the side chains of the negatively (Asp and Glu) and positively (Arg and Lys) charged amino acids were assigned negative and positive unit charges, respectively. The conformational energy of the protein was estimated by using a native topology-based model (19) that ensures a funnel-like energy landscape by excluding nonnative interactions and represents native contact interactions using the Lennard-Jones potential. Electrostatic interactions between charged residues of the proteins (except for pairs that form native contacts) were also included and described using the Debye-Hückel potential (20). The native-state topologies of EnHD and eGFP were assumed to be the same when these proteins are tethered to each other or alone. Similar structure-based models have been used previously to successfully capture the essential details of folding of multidomain or conjugated proteins (7,21).

The molecular motion of the isolated and tethered systems were simulated by Langevin dynamics with a friction coefficient $\gamma = 0.01$. It should be noted that for a system with both fast and slow degrees of freedom (associated with the flexible linker and rigid protein domain, respectively), a Langevin thermostat is useful to avoid inhomogeneous distribution of the thermal energy (22). Using this setup, the temperature was varied from $k_BT = 1.3$ to $k_BT = 1.5$ with increments of 0.01 k_BT and at each temperature 10 independent fixed temperature simulations were performed for each system studied. These simulations were used to estimate statistical errors. Each simulation included at least 10^8 elementary MD steps and included numerous folding/unfolding events of EnHD. It should be noted that eGFP remains completely folded throughout the whole temperature range studied here because of its higher stability (due to a larger number of native contacts). The simulated trajectories yielded distributions of the enthalpies and number of native contacts, Q, in EnHD that were then analyzed using the weighted histogram analysis method (WHAM) (23). The WHAM analyses yielded plots of the specific heat capacity, C_V, as a function of temperature and free energy as a function of Q (24). A conformation was assigned as folded if its number of native contacts was larger than the number of contacts at the highest point of the transition barrier, Q_{TS}, and as unfolded if otherwise. The folding temperature, which is defined as the temperature at the peak of the specific heat capacity curve, provides a measure of the relative stability of EnHD in the presence and absence of tethered eGFP.

RESULTS AND DISCUSSION

Four different chimeras were generated by fusing eGFP to either the N- or C-terminus of EnHD via a short linker comprising three glycine residues (Gly_3) or a longer 10-residue linker ($(Gly\text{-}Gly\text{-}Ser)_3\text{-}Gly$) (Fig. 1). CD spectra measured at 25°C show that EnHD and eGFP have the expected predominantly α-helical or β-strand rich structures, respectively. The percentages of α-helix content of EnHD and eGFP calculated (25) from their molar ellipticities at 208 nm, θ_{208nm}, are 55.8% and 17%, respectively. The chimeras contain both types of secondary structure elements (Fig. 2 and Fig. S1 in the Supporting Material) and their CD spectra can be generated from the sum of spectra of EnHD and eGFP alone (Fig. 2). Accordingly, the α-helix content of 23.5% calculated for EnHD-Gly_3-eGFP from θ_{208nm} is in agreement with the value of 25.6% that is calculated from the α-helix contents of EnHD and eGFP alone. CD spectra measured at 65°C show that the secondary structure of eGFP is maintained, whereas that of EnHD is lost in large part (Fig. 2) as reflected in respective α-helix contents of 17.9% and 34% calculated from θ_{208nm}. The CD spectra of the chimeras at 65°C show that their α-helical structure is

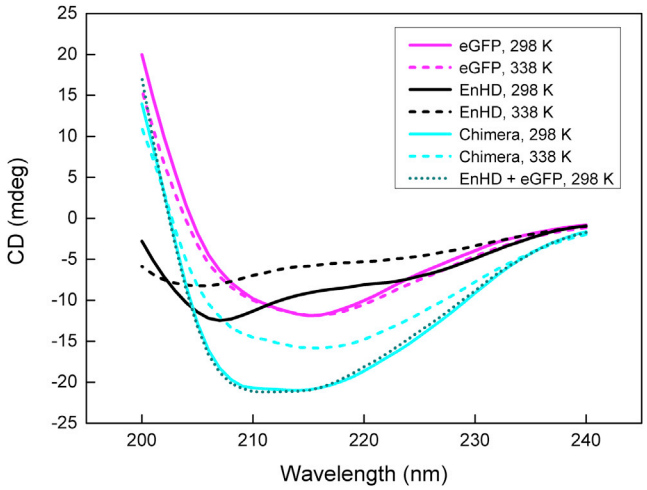

FIGURE 2 Far-UV CD spectra of EnHD, eGFP, and the EnHD-Gly_3-eGFP chimera. The far-UV CD spectra of EnHD (*black*), eGFP (*magenta*), and the EnHD-Gly_3-eGFP chimera (*cyan*) were recorded at 25°C (*solid lines*) and 65°C (*dashed lines*). The calculated sum of the spectra of EnHD and eGFP at 25°C is also shown (*dotted cyan*). See Materials and Methods for further details. To see this figure in color, go online.

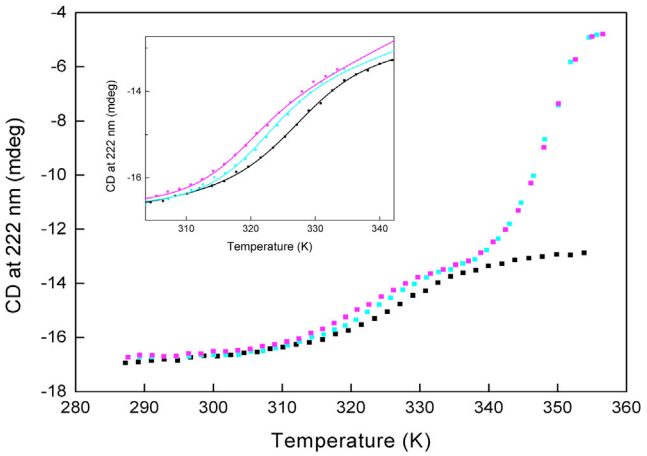

FIGURE 3 Thermal denaturation curves of EnHD and various chimeras. The CD signal at 222 nm of EnHD (*black*), eGFP-Gly_3-EnHD (*magenta*), and EnHD-Gly_3-eGFP (*cyan*) is plotted as a function of temperature. The inset shows the data at 305–345 K. The solid lines show the fits of the data to Eq. 1. See Materials and Methods for further details. To see this figure in color, go online.

partly lost but their β-strand-rich structure is maintained (Figs. 2 and S1). The α-helix content of 18.5% calculated for EnHD-Gly_3-eGFP at 65°C from θ_{208nm} agrees very well with the value of 18.2% that is calculated from the α-helix contents of EnHD and eGFP alone at 65°C. Taken together, these spectra show that heating the chimeras from 25 to 65°C results in an unfolding transition of their EnHD part, whereas the structure of their eGFP part remains essentially intact.

The change in the CD signal as a function of temperature was measured for EnHD, eGFP, and the various chimeras (Figs. 3 and S2). The plots obtained for EnHD and eGFP are monophasic, whereas those for the chimeras are biphasic (Figs. 3 and S2). The melts of EnHD by itself (26) and in the chimeras (Fig. S3) are reversible under our experimental conditions. The data could, therefore, be fitted to Eq. 1 yielding estimates for the values of the melting temperatures, T_m, changes in heat capacity, ΔCp, and denaturation enthalpies, ΔH_m (Table 1). It may be seen that the value of T_m for eGFP is unaffected by tethering it to EnHD. By contrast, the value of T_m for EnHD becomes lower when it is tethered to eGFP regardless of the linker length used and whether the tethering is to EnHD's N- or C-terminus. The largest destabilizing effect is observed when EnHD is fused at its N-terminus to eGFP via the shorter Gly_3 linker. The destabilizing effects of tethering to eGFP are even more pronounced, but with the same rank order, in the case of the Leu-16→Ala EnHD mutant (Fig. S4). These data, however, were not fitted to Eq. 1 because the pre- and posttransition baselines are absent or poorly defined.

A similar rank order of stabilities, i.e., EnHD > EnHD-Gly_3-eGFP > eGFP-Gly_3-EnHD, was found upon WHAM analysis (23,24) of coarse-grained Langevin dynamic simulations of these proteins (Fig. 4 *a*). The agreement between experiment and computation is also reflected in the similar relative changes in T_m,

TABLE 1 Summary of the values of the thermodynamic parameters determined from the temperature-induced melts of EnHD, eGFP, and their chimeras

	EnHD			eGFP		
Protein	T_m (K)	ΔCp (kcal mol^{-1} K^{-1})	ΔH (kcal mol^{-1})	T_m (K)	ΔCp (kcal mol^{-1} K^{-1})	ΔH (kcal mol^{-1})
EnHD	325.0 ± 0.5	0.69 ± 0.05	38.9 ± 5.2	–	–	–
eGFP	–	–	–	349.4 ± 1.0	2.6 ± 0.2	104.8 ± 9.7
EnHD-G_3-eGFP	320.5 ± 0.7	1.4 ± 0.1	43.9 ± 3.1	349.3 ± 0.6	6.5 ± 1.6	121.7 ± 11.8
eGFP-G_3-EnHD	316.6 ± 0.2	1.5 ± 0.1	40.1 ± 3.8	347.7 ± 0.7	5.9 ± 1.6	120.7 ± 14.1
EnHD-$(GGS)_3$G-eGFP	320.4 ± 0.8	1.4 ± 0.1	44.8 ± 3.9	349.3 ± 1.4	5.0 ± 0.6	123.9 ± 13.3
eGFP-$(GGS)_3$G-EnHD	319.2 ± 0.7	1.2 ± 0.2	45.7 ± 2.7	347.8 ± 0.2	5.0 ± 0.2	90.3 ± 3.2

All the temperature-induced melts were carried out in triplicate and the average values ± SD of each parameter are reported. The single-letter notation for amino acids is used.

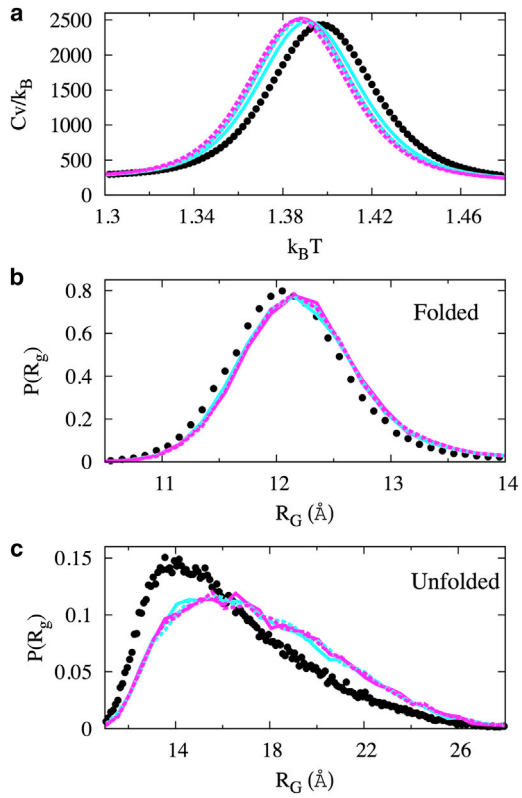

FIGURE 4 Characterization of tethered EnHD using coarse-grained MD simulations. (*a*) Plots of the specific heat capacity divided by the Boltzmann constant versus temperature are shown for EnHD alone (*black circles*), EnHD-Gly_3-eGFP (*solid cyan line*), and eGFP-Gly_3-EnHD (*solid magenta line*) in which eGFP is folded and the electrostatics of both proteins are considered. Also shown are plots for EnHD-Gly_3-eGFP (*dotted cyan*) and eGFP-Gly_3-EnHD (*dotted magenta*) with electrostatics only in EnHD. The peaks of the plots correspond to the transition folding temperature (T_m) at which the folded and unfolded state populations of EnHD are roughly the same (i.e., $\Delta G \sim 0$). The values of the specific heat capacity were calculated using $C_v = (\langle E^2\rangle - \langle E\rangle^2)/k_BT^2$ where $1.3 < k_BT < 1.5$ and the energy, E, is taken to be temperature-independent (30). (*b*) Probability distribution of the radius of gyration (R_g) of the folded state of EnHD alone and in the chimeras (color code as in *a*). (*c*) Probability distribution of R_g of the unfolded state of EnHD alone and in the chimeras (color code as in *a*). The R_g values were determined at the respective transition folding temperatures of the EnHD variants. Residues in eGFP were not included in the R_g calculations. The shift of the peak toward higher R_g values in the case of the tethered variants of EnHD reflects their expanded conformations due to tethering. To see this figure in color, go online.

$$\frac{T_m(\text{EnHD}) - T_m(\text{eGFP} - \text{Gly}_3 - \text{EnHD})}{T_m(\text{EnHD}) - T_m(\text{EnHD} - \text{Gly}_3 - \text{eGFP})},$$

which are 0.55 and 0.60, respectively. Interestingly, a similar rank order was observed when the simulations were done in the absence of electrostatic interactions in eGFP, thereby indicating that the observed destabilization of EnHD is not due to electrostatic interactions between the two proteins. The free energy of unfolding at the temperature of maximum stability, $\Delta G(T^*)$, can be calculated from the value of T_m and number of residues in the protein, N, to be 2.5 kcal mol^{-1} by using: $\Delta G(T^*) = (0.0069 \times N/T_m) \times (T_m\text{-}282.6)^2$ kcal mol^{-1} (27). This calculated value is identical to the experimentally determined value (14) and it follows, therefore, that EnHD when it is in the eGFP-Gly_3-EnHD chimera is destabilized by ~0.9 kcal mol^{-1}.

The values of ΔCp for EnHD either alone or fused to eGFP are found to be in very good agreement with the value of 0.93 kcal mol^{-1} K^{-1} that can be predicted (28) from the number of residues in this protein, thus providing further indication to that in Fig. 2 that the unfolding transition of EnHD in the various chimeras occurs while the eGFP part remains essentially folded. The values of ΔCp for the chimeras are found, however, to be larger than that for EnHD alone, thereby indicating that EnHD in the chimeras becomes more unfolded than EnHD alone since it is known that the value of ΔCp scales with the change in accessible surface area (28). The simulations of EnHD alone and in the chimeras (Fig. 4, *b* and *c*), support this conclusion because they show that the radius of gyration of the unfolded state of EnHD increases when it is fused to eGFP, whereas that of the folded state does not change. This conclusion is also supported by contact map analysis that shows that interactions between residues that are distant in sequence are weaker in the denatured state of the chimeras than in EnHD alone, as expected given their less compact structures (Fig. S5).

In summary, it is shown here that EnHD undergoes thermodynamic destabilization upon its tethering to eGFP. This destabilization is due to an increase in the entropy of the denatured state ensemble. The extent of destabilization is found to depend, to some extent, on whether the tethering is to the N- or C-terminus and on the linker length. In principle, such effects can also be strongly dependent on the linker sequences (29) and on various types of nonspecific interactions at the interface between the tethered domains as reported before (3,4). Given the widespread use of tagging with GFP in cell biology, more studies should be carried out to determine how common GFP fusion-promoted protein destabilization is.

SUPPORTING MATERIAL

Five figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00450-6.

AUTHOR CONTRIBUTIONS

M.S. and N.K. carried out all the experimental work. A.B. carried out the computational analyses. M.S., A.B., Y.L., and A.H. analyzed the data. Y.L. and A.H. designed the research and wrote the article.

ACKNOWLEDGMENTS

This work was supported by the Minerva Foundation with funding from the Federal German Ministry for Education and Research (to A.H.) and the Kimmelman Center for Macromolecular Assemblies. A.H. is an incumbent

of the Carl and Dorothy Bennett Professorial Chair in Biochemistry and Y.L. is an incumbent of the Morton and Gladys Pickman Professorial Chair in Structural Biology.

SUPPORTING CITATIONS

Reference (31) appears in the Supporting Material.

REFERENCES

1. Ekman, D., A. K. Björklund, …, A. Elofsson. 2005. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J. Mol. Biol.* 348:231–243.
2. Han, J. H., S. Batey, …, J. Clarke. 2007. The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.* 8:319–330.
3. Jaenicke, R. 1999. Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* 71:155–241.
4. Randles, L. G., S. Batey, …, J. Clarke. 2008. Distinguishing specific and nonspecific interdomain interactions in multidomain proteins. *Biophys. J.* 94:622–628.
5. Bandi, S., S. M. Singh, and K. M. Mallela. 2014. The C-terminal domain of the utrophin tandem calponin-homology domain appears to be thermodynamically and kinetically more stable than the full-length protein. *Biochemistry.* 53:2209–2211.
6. Radley, T. L., A. I. Markowska, …, S. N. Loh. 2003. Allosteric switching by mutually exclusive folding of protein domains. *J. Mol. Biol.* 332:529–536.
7. Hagai, T., and Y. Levy. 2010. Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. *Proc. Natl. Acad. Sci. USA.* 107:2001–2006.
8. Luheshi, L. M., and C. M. Dobson. 2009. Bridging the gap: from protein misfolding to protein misfolding diseases. *FEBS Lett.* 583:2581–2586.
9. Wright, C. F., S. A. Teichmann, …, C. M. Dobson. 2005. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature.* 438:878–881.
10. Jacob, E., R. Unger, and A. Horovitz. 2013. N-terminal domains in two-domain proteins are biased to be shorter and predicted to fold faster than their C-terminal counterparts. *Cell Reports.* 3:1051–1056.
11. Rivenzon-Segal, D., S. G. Wolf, …, A. Horovitz. 2005. Sequential ATP-induced allosteric transitions of the cytoplasmic chaperonin containing TCP-1 revealed by EM analysis. *Nat. Struct. Mol. Biol.* 12:233–237.
12. Tsien, R. Y. 2010. Nobel lecture: constructing and exploiting the fluorescent protein paintbox. *Integr. Biol. (Camb).* 2:77–93.
13. Mayor, U., J. G. Grossmann, …, A. R. Fersht. 2003. The denatured state of Engrailed Homeodomain under denaturing and native conditions. *J. Mol. Biol.* 333:977–991.
14. Mayor, U., N. R. Guydosh, …, A. R. Fersht. 2003. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature.* 421:863–867.
15. Jiang, F., and Y. D. Wu. 2014. Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *J. Am. Chem. Soc.* 136:9536–9539.
16. van den Ent, F., and J. Löwe. 2006. RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods.* 67:67–74.
17. Unger, T., Y. Jacobovitch, …, Y. Peleg. 2010. Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol.* 172:34–44.
18. Zacharias, D. A., J. D. Violin, …, R. Y. Tsien. 2002. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science.* 296:913–916.
19. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* 298:937–953.
20. Azia, A., and Y. Levy. 2009. Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *J. Mol. Biol.* 393:527–542.
21. Arviv, O., and Y. Levy. 2012. Folding of multidomain proteins: biophysical consequences of tethering even in apparently independent folding. *Proteins.* 80:2780–2798.
22. Mor, A., G. Ziv, and Y. Levy. 2008. Simulations of proteins with inhomogeneous degrees of freedom: The effect of thermostats. *J. Comput. Chem.* 29:1992–1998.
23. Kumar, S., D. Bouzida, …, J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–1021.
24. Cho, S. S., Y. Levy, and P. G. Wolynes. 2006. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. USA.* 103:586–591.
25. Greenfield, N. J., and G. D. Fasman. 1969. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry.* 8:4108–4116.
26. Mayor, U., C. M. Johnson, …, A. R. Fersht. 2000. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. USA.* 97:13518–13522.
27. Rees, D. C., and A. D. Robertson. 2001. Some thermodynamic implications for the thermostability of proteins. *Protein Sci.* 10:1187–1194.
28. Robertson, A. D., and K. P. Murphy. 1997. Protein structure and the energetics of protein stability. *Chem. Rev.* 97:1251–1268.
29. Ma, B., C. J. Tsai, …, R. Nussinov. 2011. Dynamic allostery: linkers are not merely flexible. *Structure.* 19:907–917.
30. Chan, H. S., Z. Zhang, …, Z. Liu. 2011. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.* 62:301–326.
31. Sobolev, V., A. Sorokine, …, M. Edelman. 1999. Automated analysis of interatomic contacts in proteins. *Bioinformatics.* 15:327–332.