Catalysis of Protein Folding by Chaperones Accelerates Evolutionary Dynamics in Adapting Cell Populations

Murat Çetinbaş, Eugene I. Shakhnovich*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

Although molecular chaperones are essential components of protein homeostatic machinery, their mechanism of action and impact on adaptation and evolutionary dynamics remain controversial. Here we developed a physics-based ab initio multi-scale model of a living cell for population dynamics simulations to elucidate the effect of chaperones on adaptive evolution. The 6-loci genomes of model cells encode model proteins, whose folding and interactions in cellular milieu can be evaluated exactly from their genome sequences. A genotype-phenotype relationship that is based on a simple yet non-trivially postulated protein-protein interaction (PPI) network determines the cell division rate. Model proteins can exist in native and molten globule states and participate in functional and all possible promiscuous non-functional PPIs. We find that an active chaperone mechanism, whereby chaperones directly catalyze protein folding, has a significant impact on the cellular fitness and the rate of evolutionary dynamics, while passive chaperones, which just maintain misfolded proteins in soluble complexes have a negligible effect on the fitness. We find that by partially releasing the constraint on protein stability, active chaperones promote a deeper exploration of sequence space to strengthen functional PPIs, and diminish the non-functional PPIs. A key experimentally testable prediction emerging from our analysis is that down-regulation of chaperones that catalyze protein folding significantly slows down the adaptation dynamics.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: shakhnovich@chemistry.harvard.edu

Introduction

Evolutionary selection of protein sequences is a complex task whereby several traits such as translation efficiency, structural integrity (i.e. folding stability and kinetics), molecular function, as well as interactions with other proteins in the cellular milieu should be simultaneously optimized. Imposing simultaneous and often contradictive (pleiotropic) constraints on protein sequence evolution severely limits the repertoire of possible solutions in sequence space and thus slows down the evolutionary dynamics. It is widely accepted that strong selective pressure against protein misfolding plays a key role in determining the rate of protein evolution and sustainable mutational loads [1–5]. However, other constraints such as the need to avoid protein sequestration to non-functional protein-protein interactions (NF-PPIs) in the cytoplasm are also emerging as important determinants of the rates and outcomes of evolutionary dynamics of proteins [6–10].

From *de novo* folding of nascent polypeptides to refolding of mature misfolded proteins, chaperones or heat-shock proteins assist in maintaining the necessary abundance of folded proteins, compensating for the selective costs of erroneous protein synthesis, misfolding, and sequestration of proteins in NF-PPIs. In three domains of life, chaperones are essential components of protein homeostatic machinery. Chaperonins, like GroEL, effectively catalyze the folding process by increasing the rate at which misfolded proteins are converted into their folded conformations [11–13]; this process can lead to diminished aggregation and NF-PPIs due to the limited presence of aggregation-prone misfolded

species in the cytoplasm. Lindquist and others posited that chaperones may act as phenotypic capacitors by buffering the fitness effects of deleterious mutations [14], leading to a greater genetic diversity and speeding up adaptive evolution [15,16]. A recent in vivo study from our lab [12] also showed that the chaperone action in dynamic cellular milieu can be pleiotropic, i.e. it extends beyond the immediate effect of protein folding by reducing the participation of destabilized proteins in NF-PPIs and affecting their accessibility to ATP-dependent proteases .

Apparently, chaperones play a key role in sculpting the fitness landscape of organisms. However, understanding the evolutionary implications of this fact requires a multi-scale modeling that realistically represents the mechanism of chaperone action and reaches across the necessary length and time scales. Recently, we developed a multi-scale evolutionary model for population dynamics simulations [7], where the fitness (rate of division) of each cell is derived explicitly from its genomic sequence by using the physical principles of protein folding and interactions. The model provided insights into the co-evolution of molecular properties of proteins, their abundances in the cytoplasm, and their functional and NF-PPIs. Here we significantly extend this ab initio model to explicitly account for chaperone activity in the cytoplasm of model cells. The model elucidates not only the immediate pleiotropic effect of chaperone action on cellular fitness but also its long-term evolutionary consequences. We find that the chaperone activity provides a significant acceleration of adaptive evolution by minimizing the detrimental effect of protein misfolding and therefore opens new paths in sequence space for

Author Summary

Molecular chaperones or heat-shock proteins are essential components of protein homeostatic machinery in all three domains of life, whose role is not only to prevent protein aggregation but also catalyze the protein folding process by decreasing the energetic barrier for folding. Importantly, chaperones have often been implicated as phenotypic capacitors since they buffer the deleterious effects of mutations, promote genetic diversity, and thus speed up adaptive evolution. Here we explore computationally the consequences of chaperone activity in cytoplasm via long-time evolutionary dynamics simulations. We use a 6-loci multi scale model of cell populations, where the fitness of each cell is determined from its genome, based on statistical mechanical principles of protein folding and protein-protein interactions. We find that by catalyzing protein folding chaperones buffer the deleterious effect of mutations on folding stability and thus open up a sequence space for efficient and simultaneous optimization of multiple molecular traits determining the cellular fitness. As a result, chaperones dramatically accelerate adaptation dynamics.

Figure 1. Pictorial depictions of molecular interactions, chaperone interaction surface, and free energy-reaction coordinates diagram. (A) A schematic representation of molecular interactions in the model cell. The folded (red cubes) and MG state (blue cubes) proteins in the cytosol of model cell are allowed to interact with each other to form functional (red solid lines) and non-functional (black dashed lines) interactions, which include homodimeric self-interactions (black dashed loops). Black solid lines represent the PPI network of chaperone (green square). (B) Chaperone interaction surface. A single face of cube, consisting of nine amino acid residues is used to model the interaction between chaperone and unfolded proteins. (C) Reaction (rxn) coordinate vs. free energy diagram for protein folding with and without chaperones, highlighting the catalytic activity of chaperones.
doi:10.1371/journal.pcbi.1003269.g001

efficient and simultaneous optimization of multiple molecular traits, determining the fitness of model cells.

Results

Our ab initio 6-loci model cells contain explicit genomes that encode six essential, birth rate controlling, proteins that are modeled as 27-mer lattice proteins as introduced in [7]. The advantage of this coarse-grained protein model is that a crucial conformational subset, consisting of all maximally compact conformations, can be enumerated [17], making the calculations of binding and folding stabilities exact within a selected representative conformational ensemble. At the initial stage of the simulations, each protein in the model is assigned a conformation, which is deemed folded and thus functional, and each protein complex in the functional PPI network is assumed to be functional only in one specific docking mode out of 144 possible ones [7]. The model of Ref. [7] considered NF-PPIs only between folded proteins. Here we also take into account the misfolded compact Molten Globule (MG) states of proteins [18] by modeling the ensemble of unfolded states as maximally compact yet non-native conformations (see *Methods*). As shown in Fig. 1A, we allow all proteins in their folded and MG states to interact with each other in the cytoplasm of model cells to form functional and non-functional protein complexes. Experimental studies show that GroEL and several other chaperones do not interact strongly with proteins in their native state, see e.g. [19–22]. Therefore, here we only consider interactions between the model chaperone and proteins in their MG state. As shown in Fig. 1B, the interaction surface of the chaperone is modeled as a 2D (3×3) lattice fragment, consisting of nine amino acid residues that are found in the apical domain of the chaperonin GroEL and that have been shown to be essential for substrate binding [23].

We assume that functional protein complexes constitute the same prototypical PPI network as in [7]: the first protein is active in monomeric form, the second and third proteins are functional as a heterodimer, and finally, the fourth, fifth and sixth proteins form a "date triangle" where they function in various combinations of pairwise complexes between them (Fig. 1A). We then postulate, as in [7], that the division rate of an individual cell is a product of the functional concentrations of proteins for the postulated prototypical PPI network:

$$b = b_0 \frac{G_1 G_{23} (G_{45} G_{46} G_{56})^{1/3}}{1 + \alpha \left(\sum_{i=1}^{6} C_i - C_T \right)^2}. \tag{1}$$

Here b_0 is a parameter used to scale the rate and thus the time, C_T is the postulated "optimal" total concentration of proteins, which reflects the assumption that protein synthesis comes at a cost, C_i are the total concentrations of individual proteins, and α is a control parameter that defines a fitness penalty for deviation from the optimal *total* concentration of all proteins. Overall, the role of the denominator in Eq. **[1]** is to penalize the deviations from the optimal protein levels and to avoid a fitness gain by a mere overexpression of proteins. Hence, the cell division rate in our model is determined by a fitness function, which stems from an intuitive physical-biological assumption that a subset of gene products acts in concert to promote healthy cell divisions.

In what follows, we define the functional concentrations of monomer and dimers in Eq. **[1]** as

$$G_1 = [F_1] \quad \text{and} \quad G_{ij} = [F_i \cdot F_j] \times P_{\text{int}}^{ij}, \tag{2}$$

where P_{int}^{ij} is the Boltzmann probability that proteins i and j interact with each other in a specific docking conformation (see *Methods*), $[F_1]$ is the concentration of the monomeric protein product of gene 1 in its native folded form, $[F_i \cdot F_j]$ is the concentration of the binary complex formed by the folded states of proteins i and j.

We employ a simplified two-step kinetic model to describe the catalytic activity of active chaperones, as illustrated in Fig. 1C (see *Methods* for the technical aspects of the formulation). In this active model, the chaperone acts as a catalyst to accelerate the rate of protein folding. As a control, we also consider a passive model of chaperone action, whereby the role of chaperone is simply to bind and release proteins in their MG states. It is noteworthy that, in contrast to a conventional catalyst, which decreases the activation barrier for both forward and backward reactions, an active chaperone increases the rate of conversion of misfolded proteins into their folded form without increasing the rate of reverse reaction of unfolding. Such "one-way" catalysis, which requires consumption of ATP, increases the concentration of folded species, which is equivalent, under steady state conditions, to an effective increase of thermodynamic stability of a protein as outlined in [12]. We model binding of MG proteins to the chaperone with a pre-equilibrium assumption since the association/dissociation of chaperone with an MG protein is a fast process as compared to subsequent kinetic steps in which the actual protein folding occurs. It has been shown that these later kinetic steps, which lead to folding, are rate limiting as they almost always require ATP hydrolysis [24].

Examples of active chaperones with catalytic folding activity include the chaperonins, GroEL in prokaryotes [24] and TRiC in eukaryotes [25]. While the applicability of our model is not limited to the GroEL-like chaperonins, the catalytic activity of this class of chaperones has been well established, see e.g. [13,24]. Therefore, our model directly applies to this class of chaperones, which forms a good experimental system to test our predictions. Henceforth, unless otherwise indicated, we refer to the chaperones with catalytic activity simply as chaperones.

Chaperones dramatically speed up evolutionary dynamics

We explored the effect of chaperones on evolutionary dynamics by running long time evolutionary simulations (200,000 generations) of model cell populations. Our simulations start from monoclonal populations of model cells, whose sequences have been designed by using the method reported in [26] to provide high stabilities $P_{nat} > 0.8$ for all 6 proteins in their folded states without regard for their functional and NF-PPIs (see details in *Methods*). In our model, the acceleration of protein folding rate due to chaperone action is determined by the parameter x, which is the ratio of the rate at which a folded protein is released by the chaperone to the rate at which spontaneous protein folding occurs (defined in *Methods*).

To determine the effects of chaperone buffering on adaptive evolution, we tested two models – an active and a passive model. In the active model, the chaperone acts as a catalyst and accelerates protein folding. However, in the passive model, the chaperone assumes a simple role by merely binding and releasing proteins in their MG states. While for the passive model we set $x=0$, for the active model we assume a modest $x=0.1$ throughout this work, consistent with the estimates of the dynamic model given in [12]. In both cases we keep the chaperone concentration fixed at $[Ch]=0.1$. To highlight the role of chaperones we always, in parallel, run control simulations for cells without chaperones, i.e. setting $[Ch]=0$.

To determine broadly the effect of chaperones on adaptation dynamics we ran evolutionary simulations at three different temperatures, i.e. $T=0.85$ (low), $T=1.05$ (medium), and $T=1.25$ (high). Throughout this work, all temperatures are in units calibrated to Miyazawa-Jernigan (MJ) potentials [27].

The effect of chaperones on the evolution of fitness is presented in Fig. 2 as fitness ratio, i.e. the ratio of birth rate in the model with chaperone to that without chaperone. Fig. 2A shows the time evolution of $b_{[Ch]=0.1}/b_{[Ch]=0}$ for the active model $x=0.1$. The chaperones provide dramatic fitness benefit during the adaptive evolution, especially at early stages. The effect of chaperones is more pronounced at higher temperatures, where proteins in the MG state are more prevalent. While the fitness ratio reaches its peak of 100 at intermediate adaptation times for $T=0.85$, it peaks at 250 for $T=1.05$ and dramatically over 1000 for $T=1.25$. After the initial fast adaptation period, the relative fitness effect of chaperones abates. Up to this point, however, the cells already have gained a considerable fitness advantage, and in the long time limit, we see gradually declining fitness ratios as the organisms become more and more fit. Nevertheless, the evolutionary dynamics with chaperones always leads to a higher long-time fitness than the evolutionary dynamics without chaperones, although the final fitness ratio is not as dramatic as those observed at intermediate evolutionary times.

Fig. 2B shows the time evolution of $b_{[Ch]=0.1}/b_{[Ch]=0}$ for the passive chaperone model $x=0$. It is clear that the chaperones in the passive model do not provide a noticeable fitness gain but rather a small fitness loss (due to the sequestration of proteins by chaperones) at the initial stage of adaptation for all three temperatures. In light of these results, we conclude that the active folding of proteins by chaperones is necessary to provide a fitness benefit to cells in the evolutionary dynamics. In the following, unless otherwise indicated, we present the data only for the active model at the low temperature $T=0.85$ as representative of our general results.
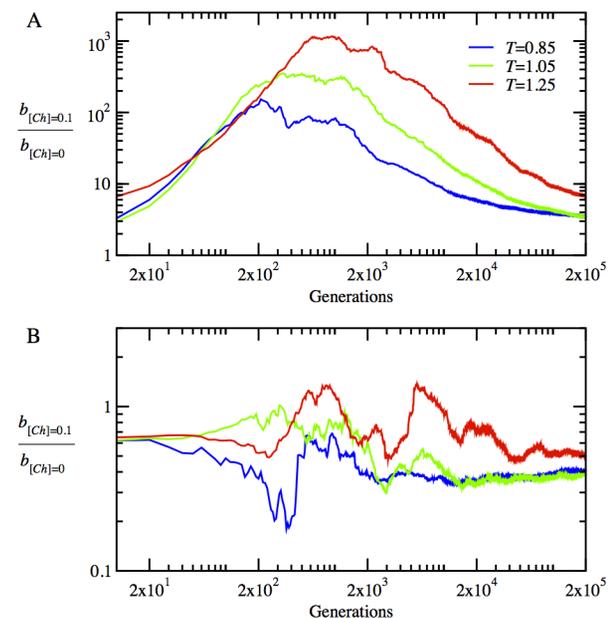

Figure 2. The time evolution of the fitness ratios $b_{[Ch]=0.1}/b_{[Ch]=0}$ **(i.e. the ratio of birth rates with chaperones and without chaperones) are presented for the active** $x=0.1$ **in (A) and passive** $x=0$ **model in (B) for three different temperatures.** The fitness ratios and evolutionary time are in log scale to convey the events clearly across all time scales. All data here and in the subsequent figures are ensemble averages over 100 independent stochastic trajectories.
doi:10.1371/journal.pcbi.1003269.g002

Chaperones promote epistasis in the evolution of molecular properties of proteins

Now, we turn to a detailed account of the evolutionary dynamics of the physicochemical properties of proteins, i.e. their stabilities P_{nat} and functional interaction probabilities P_{int} for the functional heterodimers and date triangles (see *Methods* for the definitions of these quantities). We present the time evolution of P_{nat} for the monomer in Fig. 3A, for the heterodimer proteins in Fig. 3B, and for the date triangle proteins in Fig. 3C. The chaperones provide a noticeable increase in stability for the monomeric proteins, as seen in Fig. 3A. Interestingly, at the initial stage of adaptation within 500 generations, the monomer loses its stability considerably by accumulating destabilizing mutations in the presence of chaperones. However, subsequent mutations bring about a rapid turnaround, resulting in a very stable monomer, which persists throughout the rest of the evolutionary dynamics. The non-monotonic dependence of stability of the monomeric protein on evolutionary time is an indication of a chaperone-enhanced epistatic behavior. The chaperone buffering relaxes significantly the stability constraint and allows the accumulation of more mutations in the locus encoding natively monomeric protein. This effect is mainly responsible for the initial sharp drop in the stability of the monomer. The resulting enhanced genetic diversity provides a path to a faster optimization of collective properties of all proteins in the cytoplasm such as NF-PPIs, as we show below.

The evolutionary dynamics of stability for the heterodimer and date triangle proteins show quite a different trend, as seen in Figs 3B and 3C. Initially, both the heterodimer and date triangle proteins lose their stability, but later on, the stability of date triangle proteins is slowly restored. However, in striking contrast to the monomer, the stability of heterodimer and date triangle proteins in the presence of chaperones shows a downward trend with evolutionary time, as compared to that of the chaperone-free cytoplasm of model cells.

The evolution of strengths of functional interactions, reflected in the parameter P_{int} for the heterodimer and date triangle proteins, is given in Fig. 3D and 3E, respectively. The chaperones significantly increase P_{int} for both the heterodimer and date triangle complexes. P_{int} for the heterodimer increases rapidly within first 1000 generations, in the presence of chaperones. The rate of increase of P_{int} for the date triangle is slower than that for the heterodimer; nevertheless, the chaperones provide a significant increase in P_{int} for the date triangle complexes as well. Hence, our results show that the chaperones shift the balance between the strengths of functional interaction and stability of proteins in favor of the former at the expense of the latter. Indeed, it is more advantageous for faster adaptation that the heterodimer and date triangle proteins primarily develop strong interaction surfaces to contribute to the fitness. High stability of proteins establishes later on once the strong functional interaction between them is ensured.
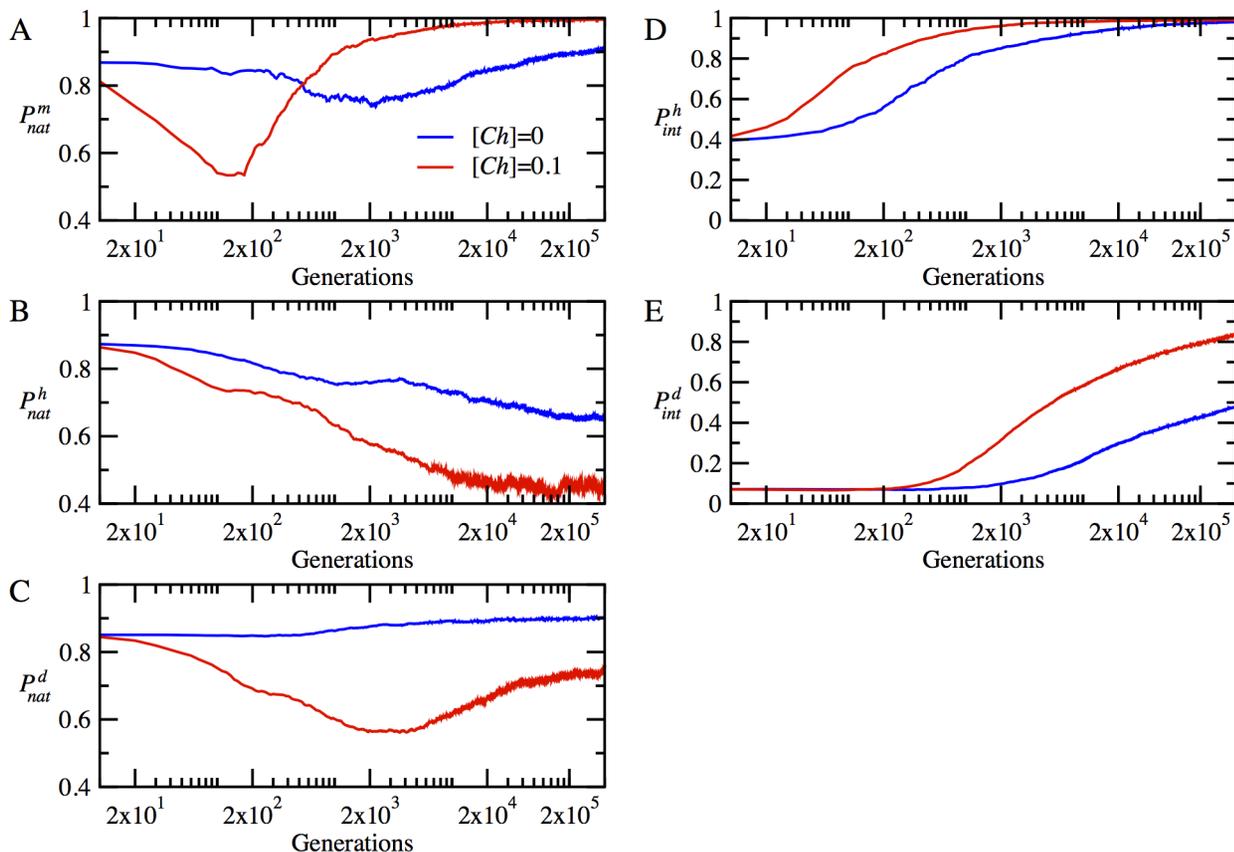

Figure 3. The time evolution of mean protein stabilities and mean interaction probabilities of functional dimers in the absence and presence of chaperones, i.e. for $[Ch]=0$ (blue lines) and $[Ch]=0.1$ (red lines), respectively, at temperature $T=0.85$. (A) The time evolution of stability $P_{nat}^m = P_{nat}^1$ for monomeric proteins. (B) The time evolution of mean stability, $P_{nat}^h = (P_{nat}^2 + P_{nat}^3)/2$ for heterodimer proteins. (C) The time evolution of mean stability, $P_{nat}^d = (P_{nat}^4 + P_{nat}^5 + P_{nat}^6)/3$ for date triangle proteins. (D) The time evolution of interaction probability, $P_{int}^h = P_{int}^{23}$ for the heterodimer complexes. (E) The time evolution of mean interaction probability, $P_{int}^d = (P_{int}^{45} + P_{int}^{46} + P_{int}^{56})/3$ for the date triangle complexes.
doi:10.1371/journal.pcbi.1003269.g003

Chaperones dramatically diminish the loss of proteins to NF-PPIs

While the effect of chaperones on protein stabilities and interactions is significant, it cannot fully account for the huge overall fitness increase, which transiently reaches up to a factor of 100 at the low temperature $T = 0.85$ (see Fig. 2A). Therefore, there must be another factor affecting fitness, where the effect of chaperones appears even more pronounced. To that end, we turn to the analysis of NF-PPIs, which affect fitness through modulation of concentrations of proteins in their functional form. We find that at the early stages of adaptation the chaperones dramatically decrease the concentrations of protein complexes engaged in NF-PPIs, releasing more proteins to become functional. This can be seen in Fig. 4, where we plot the time evolution of the fraction of protein material wasted in NF-PPIs in the absence and presence of chaperones. Specifically, we present the time evolution of $\Gamma_m = \Phi_1/C_1$ for the monomeric protein and $\Gamma_h = (\Phi_2 + \Phi_3)/(C_2 + C_3)$ for the heterodimers and $\Gamma_d = (\Phi_4 + \Phi_5 + \Phi_6)/(C_4 + C_5 + C_6)$ for the date triangles, where C_i is the total concentration protein i and Φ_i is the total concentration of protein i involved in NF-PPIs. Fig. 4 shows that the vast majority of proteins are lost to NF-PPIs at the beginning of evolutionary runs, where the sequences are optimized for stability only without regard for functional PPIs. Apparently, at the very early stage of adaptation, cell resources are mostly wasted unproductively to NF-PPIs. Both Fig. 4A and 4B show that the chaperones give rise to a rapid increase in the functional concentrations of monomeric and heterodimer proteins within the first 5,000 generations. As shown in Fig. 4C, the rate of decrease of NF-PPIs for the date triangle proteins is slower than that observed for the monomer and heterodimer proteins; nevertheless, with chaperones, it still occurs at the early stage of adaptation within 10,000 generations.

We find therefore that, while the chaperones interact directly with proteins to affect its molecular properties, their greatest impact on cellular fitness occurs indirectly through the optimization of a collective property of all proteins in the cytoplasm of model cells, namely, their NF-PPIs.

Chaperones speed up evolution by promoting neutrality and polymorphism

Our results indicate that the chaperones significantly accelerate the rate of adaptive evolution. Customarily, a well-known parameter $\omega = dN/dS$, where dN and dS are the non-synonymous and synonymous substitution rates, respectively, represents a quantitative measure of evolutionary rate. A straightforward approach to calculate dN and dS at any time step in simulation is to compare the genome of the dominant clone in the population to the initial starting genome. However, we find that this approach is problematic for our model in the long time limit when multiple substitutions at a single site become frequent. Here, we employ a slightly different approach. Following Wilke [28], we define the evolutionary rate as $\omega = (cM_a/\bar{N}_a)/(cM_s/\bar{N}_s)$ where cM_a and cM_s are the cumulative non-synonymous and synonymous substitution counts summed over short time intervals of 100 generations, and $\bar{N}_a$ and $\bar{N}_s$ are arithmetic means of weights for non-synonymous and synonymous sites, which account for different degeneracies of codons in the genetic code, calculated over time frames of 100 generations, see *Methods* for details.

We summarize our results, averaged over multiple evolutionary runs, in Fig. 5A to 5C for ω to highlight the type and magnitude of selection acting on different proteins at different stages of adaptation. In Figs., from 5D to 5F, we present the cumulative weighted non-synonymous substitutions $c\bar{M}_a = cM_a/\bar{N}_a$ for different types of
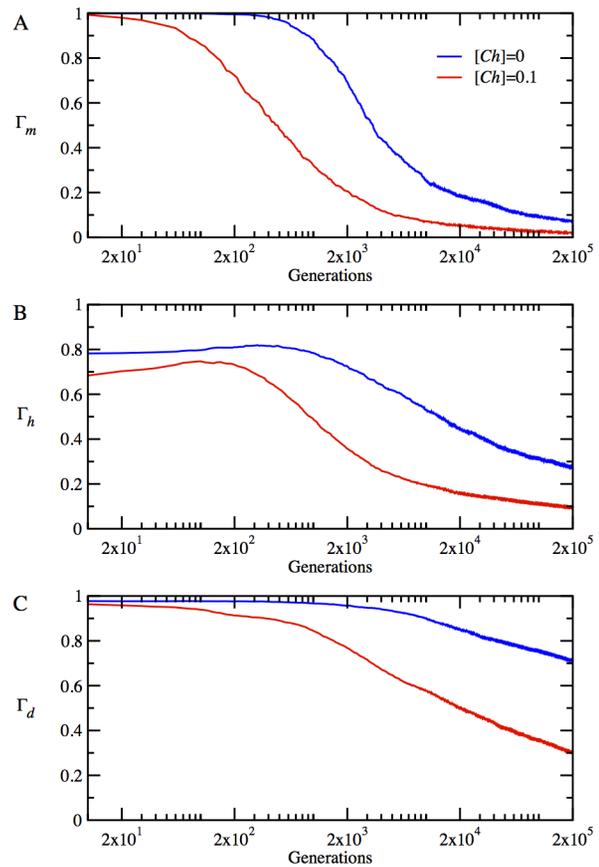

Figure 4. The time evolution of the mean value of the fractions of proteins involved in NFP-PPIs to their total concentrations for $[Ch] = 0$ (blue lines) and $[Ch] = 0.1$ (red lines), at temperature $T = 0.85$. (A) NF-PPI for functional monomer $\Gamma_m = \Phi_1/C_1$ (B) average NF-PPI for heterodimers $\Gamma_h = (\Phi_2 + \Phi_3)/(C_2 + C_3)$ and (C) NF-PPI for date triangles $\Gamma_d = (\Phi_4 + \Phi_5 + \Phi_6)/(C_4 + C_5 + C_6)$ where $\Phi_1 = C_1 - [F_1] - [U_1] - [U_1 \cdot Ch]$, $\Phi_2 = C_2 - G_{23} - [F_2] - [U_2] - [U_2 \cdot Ch]$, $\Phi_3 = C_3 - G_{23} - [F_3] - [U_3] - [U_3 \cdot Ch]$, $\Phi_4 = C_4 - G_{45} - G_{46} - [F_4] - [U_4] - [U_4 \cdot Ch]$, $\Phi_5 = C_5 - G_{45} - G_{56} - [F_5] - [U_5] - [U_5 \cdot Ch]$, $\Phi_6 = C_6 - G_{46} - G_{56} - [F_6] - [U_6] - [U_6 \cdot Ch]$.
doi:10.1371/journal.pcbi.1003269.g004

proteins in our system. Further, in Fig. S1, we provide the synonymous substitution rates $c\bar{M}_s = cM_s/\bar{N}_s$. The evolutionary dynamics of ω, $c\bar{M}_a$ and $c\bar{M}_s$ for individual trajectories are also given in Fig. S2.

We found that at the very early stage of adaptation, after 500 generation, the chaperones induce a strong positive selection pressure on the monomer, which lasts, in average for about 10000 generations, after which the monomer falls under purifying selection. However, without the chaperones, the monomer evolves under positive selection only for a short time between 1500 to 6000 generations. On the other hand, without the chaperones, the net selection on both the heterodimer and date triangle genes is negative, apparently due to the dominance of the stability constraint. In the presence of chaperone buffering, however, these loci evolve under positive selection for about 8000 to 10000 generations before they revert back to purifying selection. An important generic effect apparent in the time evolution of ω is that the chaperone buffering relaxes the negative selection pressure on all proteins and promotes the fixation of a greater number of beneficial mutations. Therefore, after the initial stage of fast adaptation, when all genes evolve under positive selection, we
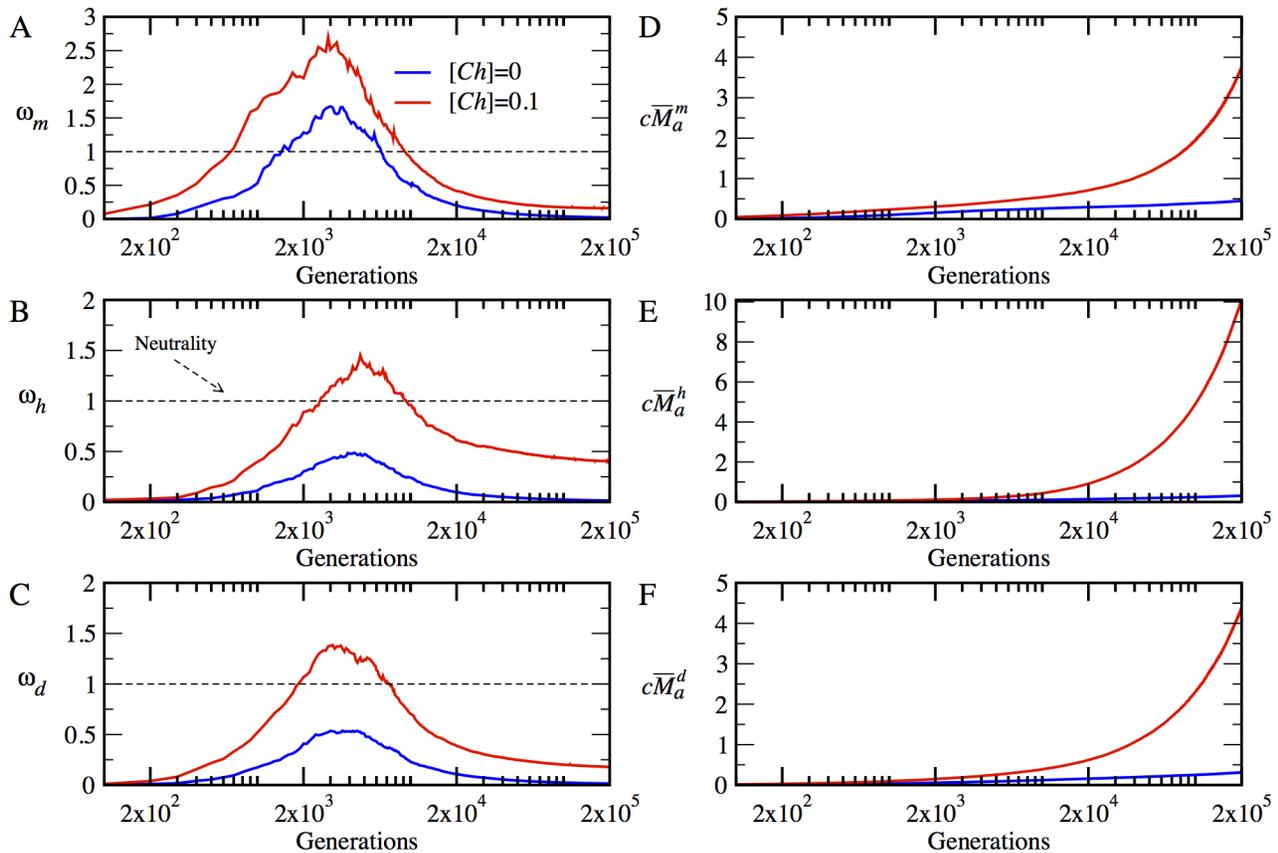
Figure 5. The time evolution of mean $\omega = (c\bar{M}_a/\bar{N}_a)/(c\bar{M}_s/\bar{N}_s)$ and $c\bar{M}_a$ in the absence and presence of chaperones, i.e. for $[Ch]=0$ (blue lines) and $[Ch]=0.1$ (red lines), respectively. The dashed line at $\omega=1$ represents the neutral evolution. The time evolution of ω is plotted in (A) for the monomer $\omega_m = \omega_1$, in (B) for the heterodimer $\omega_h = (\omega_2 + \omega_3)/2$, and in (C), for the date triangle $\omega_d = (\omega_4 + \omega_5 + \omega_6)/3$. The time evolution of $c\bar{M}_a$ is plotted in (D) for the monomer $c\bar{M}_a^m = c\bar{M}_a^1$, in (E) for the heterodimer $c\bar{M}_a^h = (c\bar{M}_a^2 + c\bar{M}_a^3)/2$, and in (F) for the date triangle $c\bar{M}_a^d = (c\bar{M}_a^4 + c\bar{M}_a^5 + c\bar{M}_a^6)/3$.
doi:10.1371/journal.pcbi.1003269.g005

observe that the chaperones bring all genes closer to neutral regime in the adapted populations.

Next, we evaluated the effect of chaperones on the polymorphism in evolving populations of model cells. To that end we determined the average sequence entropy for each protein locus in our model. This quantity is determined from the alignment of gene sequences between all model organisms within the population (see *Methods*). These results are presented in Fig. 6.

Overall, we find that chaperones greatly enhance polymorphism in evolving populations. For all protein types, the sequence entropy rapidly increases within a few hundred generations with chaperone. For the monomer and date triangle proteins, the entropy stays approximately at the same level for the duration of an evolutionary run after the initial fast increase. For the heterodimer proteins, however, the entropy gradually increases reaching a level, which is almost two times higher than that for the monomer and date triangle proteins. A greater degree of polymorphism observed for the heterodimer proteins helps these proteins evolve faster than other loci in the model, as we note below.

The enhanced neutrality due to chaperone buffering also increases the rate of protein evolution considerably. Indeed, as seen in Figs., from 5D to 5F, the chaperones increase the net number of non-synonymous mutations for all loci. Initially, the monomer still evolves with the chaperones faster than the heterodimer and date triangle genes. However, the rate of

evolution of heterodimer is the fastest as a result of more phenotypic diversity of this gene in population as indicated by the entropy plot (see Fig. 6B). Apparently, the evolutionary rates of the heterodimer and date triangle loci are slower that that of the monomer throughout the evolutionary dynamics with or without chaperones. Finally, Fig S1 shows that the rate of synonymous substitutions is approximately the same for all protein types, as could be expected. However, we also see that the rate of synonymous substitutions is slightly faster with chaperones as compared to that of chaperone-free evolution. Such slightly faster evolution of synonymous substitutions might be due to hitchhiking of neutral mutations with beneficial ones that should more pronounced with chaperone evolution.

Discussion

Our ab intio cell model, while much simpler than real biological systems, captures the essence of biological complexity that stems from the fact that the main effects, epistatic effects, and pleiotropic effects on different parameters often act in antagonistic directions. The pleiotropic concept of optimization of antagonistic traits in evolutionary biology, which gives rise to a complex fitness landscape, has its analog in the concept of frustration in physics, where competing interactions lead to a complex energy landscape with many suboptimal minima equal to or close to global minimum [29,30]. In our model, the molecular traits, whose
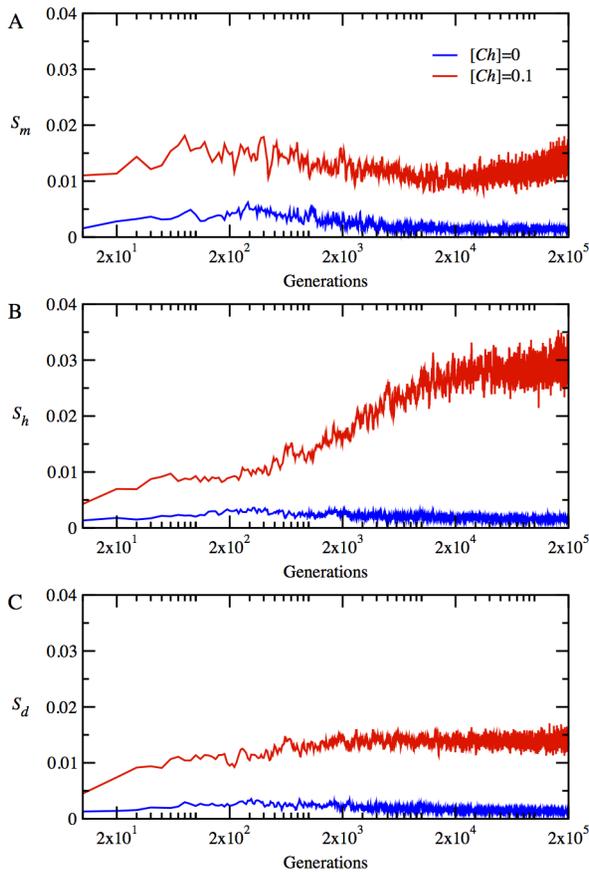
Figure 6. The time evolution of mean sequence entropy is plotted in the absence and presence of chaperones, i.e. for $[Ch]=0$ (blue lines) and $[Ch]=0.1$ (red lines), respectively, at temperature $T=0.85$. The time evolution of mean sequence entropy is given in (A) for the monomer $S_m = S_1$, in (B) for the heterodimer $S_h = (S_2 + S_3)/2$, and in (C) for date triangle proteins $S_d = (S_4 + S_5 + S_6)/3$.
doi:10.1371/journal.pcbi.1003269.g006

optimization might be antagonistic, include protein stability, abundances, functional, and NF-PPIs. An earlier study showed how antagonistic constraints result in a peculiar co-evolution of protein abundances and functional PPIs [7]. Here we introduced a new essential component of the cellular milieu – the chaperone activity, which enhances the conversion of proteins from the MG state to their native conformation. The chaperone action in our model partially relaxes an essential constraint on protein sequences to maintain high stability of proteins. The resulting chaperone buffering dramatically affects evolutionary dynamics by opening up sequence space, to provide a dramatic acceleration of the adaption process. We find that only the active chaperone model has a strong effect on evolutionary dynamics, while the passive chaperone model, where an MG protein is bound to the chaperone to prevent its sequestration to NF-PPIs has no effect on fitness (Fig. 2B). However, an important caveat here is that our model does not consider an irreversible aggregation and other elements of protein quality control such as proteolytic activity. Hence, the passive model might also be efficient when all the kinetic aspects of protein quality control are taken into account.

Mechanistically, our main finding is that the chaperones act pleiotropically to affect fitness in a number of ways. Firstly, the relaxation of the stability constraint allows achieving stronger

functional PPIs at the expense of lower thermodynamic stability for the proteins participating in the functional PPIs (Fig. 3). Secondly, a more dramatic effect of the chaperone on cellular fitness stems from faster and greater decrease of the NF-PPIs in the course of the evolutionary dynamics (Fig. 4). The NF-PPIs are a collective property of all proteins in the cellular milieu. There is evidence that proteins in their MG state are largely responsible for NF-PPIs [31]. The active chaperone in our model converts the proteins in their MG states into their folded conformations, leading to a drop in NF-PPIs with an ensuing increase in fitness due to a diminished sequestration of functional proteins.

Recent experimental and theoretical studies with the chaperonin GroEL corroborate some of our findings. Tokuriki and Tawfik performed a series of random mutagenesis experiments on a number of non-endogenous enzymes expressed in *E. Coli* to investigate the impact of overexpression of GroEL on enzyme evolution [16]. They found that GroEL indeed helps folding of destabilized proteins and potentially facilitates the evolution of enzymes to gain new functions. The acceleration of adaptive evolution by GroEL is also found in a recent study [15] in which the evolutionary rates of GroEL clients and non-clients [32] were compared. It was found that the GroEL obligatory proteins evolve 35% faster than the proteins that fold spontaneously without the GroEL assistance [15]. The importance of GroEL for adaptive evolution is highlighted by the case of the endosymbiotic bacterium *Buchnera*, which often undergoes population bottlenecks through maternal transmission and thus quickly accumulates random mutations that destabilize proteins [33]. Remarkably, the expression level of GroEL in *Buchnera* is almost 8 times greater than that of *E. coli* under the normal conditions.

Quayle and Bullock define evolvability as the number of generations that it takes for a population to reach its phenotypic target that maximizes fitness [34]. Our study highlights the dual role of chaperones not only as a catalyst of protein folding but also as a catalyst on the fitness landscape, which lowers the genetic "barriers" between phenotypes and thus promotes evolvability. A key prediction emerging from our analysis is that the catalytic activity of chaperone gives rise to a dramatic acceleration of adaptive evolution. Hence, we predict that the depletion of active chaperones through down-regulation of their expression should directly affect the rate at which organisms adapt to new environments, which can be directly experimentally testable. This work is currently in progress in our lab.

Methods

Protein stability and interactions

Our proteins consist of 27 amino acid residues that fold into $3\times3\times3$ cubic lattice conformations [17]. We use the MJ potentials to model intra- and inter-molecular interactions [27]. While the 27-mer lattice model has 103,346 maximally compact conformations [17], we employ a uniform subset of randomly selected 10,000 conformations as our conformational ensemble to speed up calculations [7]. We calculate the Boltzmann probability of folding to a native state, i.e. P_{nat}^i for each protein $i=1,..,6$ as follows

$$P_{nat}^i = \frac{\exp[-E_1^i/T]}{\sum_{k=1}^{10000} \exp[-E_k^i/T]}, \qquad (3)$$

where E_1^i is the energy of the native conformation and T is the temperature in units corresponding to the calibration with MJ potentials.

We model the functional protein-protein and protein-chaperone interactions using a rigid docking scheme. The six faces of a cubic lattice provide six possible interaction surfaces and there are four rotational degrees of freedom to dock two interaction surfaces of two lattice proteins. Hence, in total, there exist $6 \times 6 \times 4 = 144$ docking modes for a binary protein complex. The Boltzmann probability of interaction P_{int}^{ij} between the dimer proteins i and j are calculated as

$$P_{int}^{ij} = \frac{\exp[-E_f^{ij}/T]}{\sum_{k=1}^{144} \exp[-E_k^{ij}/T]}, \qquad (4)$$

where E_f^{ij} is the interaction energy of the functional binding mode, E_k^{ij} are the interaction energies for 144 docking modes and T is the temperature. Because 27-mers are quite small (as compared to real protein sizes), in order to represent both interior and interaction surface of proteins, we employed two different lattice conformations to model our proteins: one for interior part that determines stability and one for interaction surface that determines PPI, as has been done in previous studies, see e.g. Ref. [7]. Hence, the lattice conformations that we used to represent protein surfaces in order to calculate P_{int}^{ij} are randomly chosen conformations but they are not the same lattice conformation that we used to represent the stability energetics of native folds. This approach provides a less tight coupling between interior and exterior of proteins that would be the case for small 27-mers representing therefore a more realistic description of protein geometry and energetics.

The model of active chaperone action

In the absence of chaperones, the folded state and the unfolded ensemble of states (which also includes compact MG states) for any protein "i" are at equilibrium, satisfying detailed balance with the corresponding folding k_f^i and unfolding rates k_u^i:

$$U_i \underset{k_u^i}{\overset{k_f^i}{\rightleftharpoons}} F_i. \qquad (5)$$

The active chaperone changes this picture dramatically. In general, the operation of chaperones requires input energy by ATP hydrolysis. The energy flux due to the ATP hydrolysis by chaperone causes the violation of detailed balance between the folded and unfolded forms of a protein. Therefore, following the findings in [19–22], we assume that the chaperon Ch interacts with a protein in its misfolded MG conformation U_i to form a pre-equilibrium dimer complex $[U_i \cdot Ch]$, from which the protein is released in its folded form F_i,

$$U_i + Ch \xrightarrow{K_{Ch}^i} [U_i \cdot Ch] \xrightarrow{k_{Ch}^i} F_i + Ch \qquad (6)$$
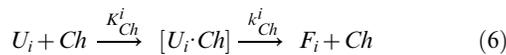
where K_{Ch}^i are the pre-equilibrium constants for the chaperone-unfolded protein complex, and k_{Ch}^i are the rate constants for the chaperone assisted folding. While the native state is uniquely defined by a single conformation in our model, the unfolded states constitute an ensemble of conformations, which we take into account as a representative ensemble via a mean field approximation (see below).

The steady state solution of Eqs. 5 and 6 leads to the following,

$$[F_i] = \frac{k_f^i}{k_u^i}[U_i] + \frac{k_{Ch}^i}{k_u^i}[U_i \cdot Ch]. \qquad (7)$$

By introducing the ratio of the rate constant for chaperone assisted folding to the rate constant for unassisted folding, i.e. $k_{Ch}^i = x k_f^i$ and also by making use of the following two equilibrium relations,

$$[U_i \cdot Ch] = K_{Ch}^i[U_i][Ch] \quad \text{and} \quad \frac{k_f^i}{k_u^i} = \frac{P_{nat}^i}{1 - P_{nat}^i}, \qquad (8)$$

we arrive at the equation,

$$[F_i] = \frac{P_{nat}^i}{1 - P_{nat}^i}[U_i] + x \frac{P_{nat}^i}{1 - P_{nat}^i}[U_i \cdot Ch], \qquad (9)$$

where $x = \frac{k_{Ch}^i}{k_f^i}$ is the ratio of the rate of release of native proteins from the chaperone complex to the rate of spontaneous folding.

Stochastic simulation algorithm

Our simulations start from initial sequences designed to be stable in their respective native conformations $P_{nat}^i > 0.8$; the PPIs of initial sequences were not optimized. We randomly assigned the functional docking modes for the heterodimers and date triangles. In order to keep protein folds fixed throughout the simulations, we discarded the cells encoding proteins whose assigned native folds were no longer the lowest energy configurations. A constant population size of $M = 1000$ is maintained throughout the simulations.

We use a variant of the Gillespie algorithm to generate stochastic evolutionary trajectories in our simulations. Using two uniformly distributed random numbers, $r_1, r_2 \in (0,1)$, the algorithm decides when and which cell undergoes a cell division. Given the normalized birth rates as $\tilde{b}_i = b_i / \beta_M$ where $\beta_M = \sum_{i=1}^{M} b_i$ for each cell i in a population of size M, we define the cumulative probabilities as $\beta_i = \beta_{i-1} + \tilde{b}_i$. Note that $\beta_1 = \tilde{b}_1$ and $\beta_M = \sum_{i=1}^{M} \tilde{b}_i$. The waiting time dt for the next cell division to occur at time $t+dt$ is determined by $dt = (1/\beta_M) \ln r_1$. The cell "i" divides when the second random number falls in the interval $\beta_{i-1} \le r_2 < \beta_i$.

Upon cell division, a mother cell gives birth to a daughter cell. A newborn cell replaces a randomly chosen cell in the population in order to maintain constant population size. Also, whenever a mutation changes the lowest energy protein fold or hits a stop codon, such cells are discarded from the population. Upon semi-conservative replication, both the mother and daughter cells are subject to either a mutational event with constant probability $m = 0.001$ per gene per replication (whereby a nucleotide is randomly changed) or the expression level of one randomly chosen protein in a cell can change with a constant rate $er = 0.01$ per cell division such that the concentration of a protein i in a daughter cell is derived from that of a mother cell as follows $[C_i^{new}] = [C_i^{old}](1 + \varepsilon)$ where ε is a Gaussian random number with zero mean and variance 0.1. At the beginning of our simulations, we set the concentrations of each protein and chaperone equally at $C_i = 0.1$.

A mean field approximation for equilibrium constants

Six proteins encoded in our cell model make four *functional* interactions in total. In addition, we allow all possible *non-functional* interactions between all proteins in their folded as well as MG conformations (see Fig. 1A). More specifically, we consider the equilibrium reactions, forming homo- as well as heterodimers between the folded proteins,

$$F_i + F_j \xrightleftharpoons{K_{F_iF_j}} [F_i \cdot F_j] \quad \text{where} \quad K_{F_iF_j} = \frac{[F_i \cdot F_j]}{[F_i][F_j]}, \quad (10)$$

between the folded and unfolded proteins,

$$F_i + U_j \xrightleftharpoons{K_{F_iU_j}} [F_i \cdot U_j] \quad \text{where} \quad K_{F_iU_j} = \frac{[F_i \cdot U_j]}{[F_i][U_j]}, \quad (11)$$
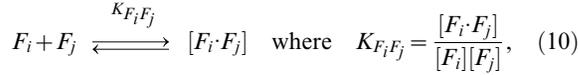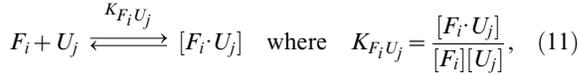
and between the unfolded proteins,

$$U_i + U_j \xrightleftharpoons{K_{U_iU_j}} [U_i \cdot U_j] \quad \text{where} \quad K_{U_iU_j} = \frac{[U_i \cdot U_j]}{[U_i][U_j]}. \quad (12)$$
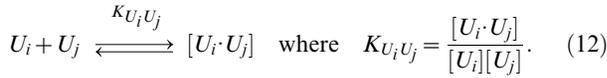
Next, we discuss how we calculate the equilibrium constants for protein-protein and protein-chaperone interactions. We use the index set $X \equiv \{U, F, Ch\}$ for different molecular species, the subscripts i and j for different proteins, and the superscripts n and m for different protein conformations. Given the two conformations n and m of lattice proteins i and j, the binding constant can be written as

$$Q_{X_i^n X_j^m} = \sum_{k=1}^{144} \exp[-E_{X_i^n X_j^m}^k / T] \quad (13)$$

where $E_{X_i^n X_j^m}^k$ are the interactions energies and T is the temperature.

In order to reduce computational cost in calculating the binding equilibrium constants, we make use of a mean field approximation in which we choose $N_u = 10$ representative MG conformations randomly out of 10,000 conformations and assume that each of these conformations is equally likely to occur in the MG ensemble. In what follows, we calculate the binding equilibrium constants for the dimers formed by the folded proteins exactly as

$$K_{F_iF_j} = Q_{F_iF_j}, \quad (14)$$

and the dimers formed by the folded and unfolded proteins as

$$K_{F_iU_j} = (1/N_u) \sum_{m=1}^{N_u} Q_{F_iU_j^m}. \quad (15)$$

The binding equilibrium constants for the heterodimers, i.e. $i \neq j$, formed by the unfolded proteins are calculated by

$$K_{U_iU_j} = (1/N_u^2) \sum_{n=1}^{N_u} \sum_{m=1}^{N_u} Q_{U_i^n U_j^m}, \quad (16)$$

and the homodimers formed by the unfolded proteins are given by

$$K_{U_iU_i} = \frac{2}{N_u(N_u+1)} \sum_{n=1}^{N_u} \sum_{m=n}^{N_u} Q_{U_i^n U_i^m}. \quad (17)$$

To model the protein-chaperone interactions, we use a 3×3 square lattice face to mimic an interaction surface for the chaperone (See Fig. 1B). For the protein-chaperone interactions, there are $1 \times 6 \times 4 = 24$ docking modes. Hence, the binding constant for an unfolded protein U_i with conformations n and the chaperone Ch is of the form,

$$Q_{U_i^n Ch} = \sum_{k=1}^{24} \exp[-E_{U_i^n Ch}^k / T]. \quad (18)$$

By using our mean field approximation, the pre-equilibrium constant for the association of chaperone with an unfolded protein is given by

$$K_{U_i Ch} = (1/N_u) \sum_{n=1}^{N_u} Q_{U_i^n Ch}. \quad (19)$$

The conservation of mass for each protein C_i and chaperone C_{Ch} in our system can be written as

$$[C_i] = [F_i] + [U_i] - [F_i \cdot U_i] + [U_i \cdot Ch]$$
$$+ \sum_{j=1}^{6} \left([F_i \cdot F_j] + [F_i \cdot U_j] + [U_i \cdot F_j] + [U_i \cdot U_j] \right) \quad (20)$$
$$[C_{Ch}] = [Ch] + \sum_{j=1}^{6} [U_i \cdot Ch].$$

An iterative method to solve the LMA equations

Due to the non-linear nature of the law of mass action (LMA) equations, a direct integration of these coupled equations may only be possible for small systems. However, by using iterative algorithms, the LMA equations can easily be solved even for large systems. Previous iterative algorithms were developed to solve the LMA equations involving only equilibrium reactions between different molecular species. The LMA equations in our system involve not only different molecular species but also equilibrium reactions between conformational isomers of the same molecular species and therefore may not be solved by the existing algorithms, see e.g. [35]. Here, we present a straightforward generalization of the existing iterative algorithms.

We start our iterative algorithm by initializing the concentrations of monomeric unfolded proteins and chaperone, i.e. $[U_i] = [C_i](1 - P_{nat}^i)$ and $[Ch] = [C_{Ch}]$. Our algorithm consists of iterations of three sets of equations to find equilibrium concentrations of all chemical species in our system. First, by substituting $[U_i]$ and $[C_{Ch}]$ into the right hand side of Eq. [9] we determine $[F_i]$. Second, by using the new value of $[F_i]$ along with $[U_i]$ and $[C_{Ch}]$ we calculate the two quantities:

$$\Sigma_i = [F_i] + [U_i] - K_{F_iU_i}[F_i][U_i] + K_{U_iCh}[U_i][Ch] \quad (21)$$
$$+ \sum_{j=1}^{6} \left(K_{F_iF_j}[F_i][F_j] + K_{F_iU_j}[F_i][U_j] + K_{U_iF_j}[U_i][F_j] + K_{U_iU_j}[U_i][U_j] \right)$$
$$\Sigma_{Ch} = [Ch] + \sum_{j=1}^{6} K_{U_jCh}[U_j][Ch]$$

Third, we find the new values of $[U_i]$ and $[C_{Ch}]$ by using,

$$[U_i]_{new} = ([C_i] \times [U_i]) / \Sigma_i \quad \text{and} \quad [Ch]_{new} = ([C_{Ch}] \times [Ch]) / \Sigma_{Ch}. \quad (22)$$

By updating the old values of $[U_i]$ and $[C_{Ch}]$ with the new values, i.e. $[U_i] = [U_i]_{new}$ and $[C_{Ch}] = [C_{Ch}]_{new}$, and substituting them back into Eq. [9], we continue our iterations until the error threshold

$\zeta < 10^{-6}$ is achieved where ζ is defined by

$$\zeta = \sum_{i=1}^{6} \left([C_i] - [\Sigma_i] \right)^2 + \left([C_{Ch}] - [\Sigma_{Ch}] \right)^2. \qquad (23)$$

Calculations of sequence entropy

In order to determine the degree of polymorphism in a population, we calculated the sequence entropy for each protein $k = 1, \ldots, 6$ by using the formula

$$S_k = -\frac{1}{27} \sum_{i=1}^{20} \sum_{j=1}^{27} p_{ij}^k \log p_{ij}^k, \qquad (24)$$

where p_{ij}^k is the frequency of amino acid type "i" in the jth position in the multiple alignment (among all cells in the population) of sequences of a protein "k".

Calculations of ω, synonymous and non-synonymous substitution rates

In calculation of ω, we used the following formula $\omega = c\bar{M}_a / c\bar{M}_s$, where $c\bar{M}_a = cM_a / \bar{N}_a$ and $c\bar{M}_s = cM_s / \bar{N}_s$ are the normalized cumulative non-synonymous and synonymous substitutions rates, and $\bar{N}_a$ and $\bar{N}_s$ are the arithmetic mean of non-synonymous and synonymous substitutions reflecting the instant composition and degeneracies in the genetic code [28]. In order to calculate the quantities cM_a, cM_s, $\bar{N}_a$ and $\bar{N}_s$ we partitioned the overall simulation time into the time intervals of length 100 generations. Given the two DNA sequences, say, DNA-1 and DNA-2 that are 100 generations apart, we first count the number of synonymous M_s and non-synonymous M_a substitutions between them, and second determine the number of synonymous $N_s = d_4 + (1/3)d_2$ and non-synonymous $N_a = d_0 + (2/3)d_2$ sites at the frames in this time interval, where d_0 is non-degenerate, d_2 is 2-fold and d_4 4-fold degenerate sites, respectively [36]. By using the above quantities, we next calculate the cumulative sums cM_s and cM_a, over all time intervals and the arithmetic averages $\bar{N}_s$ and $\bar{N}_a$, and finally determine ω.

Supporting Information

Figure S1 The time evolution of $c\bar{M}_s$ is given in the absence and presence of chaperones, i.e. for $[Ch] = 0$ (blue lines) and $[Ch] = 0.1$ (red lines), respectively. The time evolution of $c\bar{M}_s$ is plotted in (A) for the monomer $c\bar{M}_s^m = c\bar{M}_s^1$, in (B) for the heterodimer $c\bar{M}_s^h = (c\bar{M}_s^2 + c\bar{M}_s^3)/2$, and in (C) for the date triangle $c\bar{M}_s^d = (c\bar{M}_s^4 + c\bar{M}_s^5 + c\bar{M}_s^6)/3$. All results are ensemble averages over 100 independent stochastic trajectories.
(TIFF)

Figure S2 The time evolution of ω, $c\bar{M}_a$ and $c\bar{M}_s$ for the locus encoding functional monomeric protein (#1) is plotted for 5 different individual stochastic trajectories. Each color marks an individual trajectory. The time evolution of ω, $c\bar{M}_a$ and $c\bar{M}_s$ is plotted in subfigures (A), (B), and (C), respectively, for chaperone-free evolution. The time evolution of ω, $c\bar{M}_a$ and $c\bar{M}_s$ is plotted in subfigures (D), (E), and (F), respectively, for evolution with chaperones.
(TIFF)

Author Contributions

Conceived and designed the experiments: MC EIS. Performed the experiments: MC. Analyzed the data: MC EIS. Contributed reagents/materials/analysis tools: MC EIS. Wrote the paper: MC EIS.

References

1. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134: 341–352.
2. Zeldovich KB, Chen PQ, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. Proceedings of the National Academy of Sciences of the United States of America 104: 16152–16157.
3. Wylie CS, Shakhnovich EI (2011) A biophysical protein folding model accounts for most mutational fitness effects in viruses. Proc Natl Acad Sci U S A 108: 9916–9921.
4. Lobkovsky AE, Wolf YI, Koonin EV (2010) Universal distribution of protein evolution rates as a consequence of protein folding physics. Proc Natl Acad Sci U S A 107: 2983–2988.
5. Serohijos AW, Rimas Z, Shakhnovich EI (2012) Protein biophysics explains why highly abundant proteins evolve slowly. Cell Rep 2: 249–256.
6. Zhang J, Maslov S, Shakhnovich EI (2008) Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. Mol Syst Biol 4: 210.
7. Heo M, Maslov S, Shakhnovich E (2011) Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. Proc Natl Acad Sci U S A 108: 4258–4263.
8. Johnson ME, Hummer G (2011) Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. Proc Natl Acad Sci U S A 108: 603–608.
9. Levy ED, De S, Teichmann SA (2012) Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proceedings of the National Academy of Sciences of the United States of America 109: 20461–20466.
10. Yang JR, Liao BY, Zhuang SM, Zhang J (2012) Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proceedings of the National Academy of Sciences of the United States of America 109: E831–840.
11. England JL, Lucent D, Pande VS (2008) A role for confined water in chaperonin function. Journal of the American Chemical Society 130: 11838–11839.
12. Bershtein S, Mu W, Serohijos AW, Zhou J, Shakhnovich EI (2013) Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. Mol Cell 49: 133–144.
13. Chakraborty K, Chatila M, Sinha J, Shi QY, Poschner BC, et al. (2010) Chaperonin-Catalyzed Rescue of Kinetically Trapped States in Protein Folding. Cell 142: 112–122.
14. Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. Nature 417: 618–624.
15. Bogumil D, Dagan T (2010) Chaperonin-dependent accelerated substitution rates in prokaryotes. Genome Biol Evol 2: 602–608.
16. Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459: 668–673.
17. Shakhnovich E, Gutin A (1990) Enumeration of All Compact Conformations of Copolymers with Random Sequence of Links. Journal of Chemical Physics 93: 5967–5971.
18. Ptitsyn OB (1995) Molten globule and protein folding. Adv Protein Chem 47: 83–229.
19. Viitanen PV, Donaldson GK, Lorimer GH, Lubben TH, Gatenby AA (1991) Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. Biochemistry 30: 9716–9723.
20. Fink AL (1999) Chaperone-mediated protein folding. Physiological Reviews 79: 425–449.
21. Tehver R, Thirumalai D (2008) Kinetic model for the coupling between allosteric transitions in GroEL and substrate protein folding and aggregation. Journal of Molecular Biology 377: 1279–1295.
22. Tsurupa GP, Ikura T, Makio T, Kuwajima K (1998) Refolding kinetics of staphylococcal nuclease and its mutants in the presence of the chaperonin GroEL. Journal of Molecular Biology 277: 733–745.
23. Fenton WA, Kashi Y, Furtak K, Horwich AL (1994) Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371: 614–619.
24. Horwich AL, Fenton WA (2009) Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. Q Rev Biophys 42: 83–116.

25. Frydman J, Nimmesgern E, Erdjument-Bromage H, Wall JS, Tempst P, et al. (1992) Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. EMBO J 11: 4767–4778.
26. Berezovsky IN, Zeldovich KB, Shakhnovich EI (2007) Positive and Negative Design in Stability and Thermal Adaptation of Natural Proteins. PLoS Comput Biol 3: e52.
27. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 256: 623–644.
28. Wilke CO (2004) Molecular clock in neutral protein evolution. BMC Genet 5: 25.
29. Campbell IA, Bernardi L (1995) Frustration, randomness, and the spin-glass transition. Phys Rev B Condens Matter 52: R9819-R9822.
30. Sfatos CD, Shakhnovich E (1997) Statistical Mechanics of Random Hetero-plymers. Physics Reports 288: 77–108.
31. Booth DR, Sunde M, Bellotti V, Robinson CV, Hutchinson WL, et al. (1997) Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. Nature 385: 787–793.
32. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122: 209–220.
33. Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U S A 93: 2873–2878.
34. Quayle AP, Bullock S (2006) Modelling the evolution of genetic regulatory networks. Journal of Theoretical Biology 238: 737–753.
35. Maslov S, Lspolatov I (2007) Propagation of large concentration changes in reversible protein-binding networks. Proceedings of the National Academy of Sciences of the United States of America 104: 13655–13660.
36. Hartl DL, Clark A.G. (2007) Principles of population genetics, 4th ed. Sunderland, MA: Sinauer.