bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Non-equilibrium protein folding and activation by ATP-driven chaperones

Huafeng Xu*
Roivant Sciences, New York, New York 10036, United States

Recent experimental studies suggest that ATP-driven molecular chaperones can stabilize protein substrates in their native structures out of thermal equilibrium. The mechanism of such non-equilibrium protein folding is an open question. Based on available structural and biochemical evidence, I propose here a unifying principle that underlies the conversion of chemical energy from ATP hydrolysis to the conformational free energy associated with protein folding and activation. I demonstrate that non-equilibrium folding requires the chaperones to break at least one of four symmetry conditions. The Hsp70 and Hsp90 chaperones each breaks a different subset of these symmetries and thus they use different mechanisms for non-equilibrium protein folding. I derive an upper bound on the non-equilibrium elevation of the native concentration, which implies that non-equilibrium folding only occurs in slow-folding proteins that adopt an unstable intermediate conformation in binding to ATP-driven chaperones. Contrary to the long-held view of Anfinsen's hypothesis that proteins fold to their conformational free energy minima, my results predict that some proteins may fold into thermodynamically unstable native structures with the assistance of ATP-driven chaperones, and that the native structures of some chaperone-dependent proteins may be shaped by their chaperone-mediated folding pathways.

I. INTRODUCTION

A commonly accepted view on protein folding is Anfinsen's thermodynamic hypothesis [1]: the native structure of a protein is uniquely determined by its amino acid sequence and it is the conformation of the lowest free energy. According to this view, a free energy gap separates the native structure and the denatured conformations, and protein folding is accompanied by a negative free energy change [2]. A protein, left to its own device and given sufficient time, will fold spontaneously to its native structure.

We now know that many proteins depend on the assistance of molecular chaperones for folding into their functional structures inside cells [3–5]. ATP-driven chaperones such as GroEL/GroES [6–8], Hsp70 [9, 10], and Hsp90 [11–15] represent an important class of chaperones that consume chemical energy in their functions. Biochemical and structural studies have established that these chaperones undergo a cycle powered by ATP hydrolysis through open and closed conformations [10, 16–20]. These chaperones can rescue their protein substrates from misfolded or aggregated structures and accelerate their refolding to their native structures [21–25]. This role of ATP-driven chaperones does not contradict Anfinsen's thermodynamic hypothesis: proteins still fold into the most thermodynamically stable structures, but the chaperones enable them to do so within a physiologically reasonable time [26].

Recent experimental studies suggest that ATP-driven chaperones may play a thermodynamic role besides the kinetic one: they may stabilize proteins in their native structures out of thermal equilibrium, converting the chemical energy of ATP hydrolysis into the conformational free energy of their substrates [24, 27]. Coincidental to these experimental studies, theoretical models were published around the same time that predicted such non-equilibrium stabilization [27–29]. In addition to quantitatively recapitulating the experimentally observed acceleration in folding kinetics, these models suggest that ATP-driven chaperones can maintain their protein substrates in their native structures at higher concentrations than thermodynamically permitted in the chaperone-free equilibrium. They explain why ATP hydrolysis is indispensable to the cellular functions of these chaperones, and in the case of Hsp70 [28] and Hsp90 [29], the critical roles of their respective cochaperones.

Here, I define non-equilibrium protein folding to be the phenomenon in which the native fraction of a protein is elevated by an energy-consuming process above its value in thermal equilibrium. Let $f_N = [N]/P_0$ be the steady state fraction of the protein substrate in its native structure in the presence of ATP-driven chaperones, and $f_{N,eq} = [N]_{eq}/P_0$ be the native fraction in the chaperone-free equilibrium, where P_0 is the total protein concentration and $[N]$ is the concentration of the protein in its native structure. Non-equilibrium protein folding occurs if $f_N > f_{N,eq}$, which of course requires energy consumption. I will introduce the gain factor of non-equilibrium folding

$$g \equiv \frac{f_N}{f_{N,eq}} = \frac{[N]}{[N]_{eq}} \tag{1}$$

* huafeng.xu@roivant.com

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

which measures the extent of out-of-equilibrium stabilization of the native structure. A protein that primarily occupies the non-native structures in equilibrium (*i.e.* $f_{N,eq} < 0.5$) but its native structure in the presence of ATP-driven chaperones (*i.e.* $f_N > 0.5$) would contest Anfinsen's hypothesis.

Note that the native fraction in my definition of non-equilibrium folding includes only the free (*i.e.* not chaperone-bound) native protein, because chaperones primarily bind to proteins that are at least partially unfolded [9, 30]. There are, however, examples in which chaperone-bound proteins retain some native activity. For instance, glucocorticoid receptor (GR) can bind to its ligand when it is in complex with Hsp90 [31]. In this case, however, GR may still need to dissociate from Hsp90 to function as an active transcription factor. Thus, in this work I will only consider non-equilibrium folding to a free, native protein.

Many mechanistic models have been proposed for chaperone-mediated protein folding [25, 26, 30, 32–34]. One prevalent hypothesis regards the chaperones as unfoldases or holdases [35], in that their primary function is to rescue a misfolded or aggregated protein substrate and to hold it in an unfolded state. Upon release from the chaperones, the protein molecule has a certain probability of folding into its native structure [36]. Models based on this hypothesis provide an explanation of how ATP-driven chaperones accelerate the folding of the substrates to their inherently stable native structures, but they do not provide an explicit mechanism for the chaperones to transfer the chemical energy from ATP hydrolysis into the folding free energy of the substrate protein. It has been proposed that the ATP energy is used by the chaperones to achieve ultra-affinity in substrate binding [34].

It is often unclear whether a model will imply non-equilibrium protein folding (*i.e.* $g > 1$), when microscopic reversibility [29, 37] is rigorously enforced. Based on thermodynamic principles, I have previously established one requirement of non-equilibrium protein folding: the substrate protein must undergo a conformational change when it is bound to the chaperone [29]. Supported by biochemical and structural evidence [38–40], this is a key assumption in my models of chaperone-mediated protein folding that couple the conformational dynamics of the protein substrate with the ATP-driven, open-close cycle of the chaperones (Fig. 1).

In this work, I introduce an additional requirement that an ATP-driven chaperone must satisfy to perform non-equilibrium protein folding. Specifically, I demonstrate mathematically that an ATP-driven chaperone must break at least one of the four kinetic symmetry conditions (Conditions 1–4 in Sec. III A) to use the energy from ATP hydrolysis for out-of-equilibrium stabilization of substrate proteins in their native structures. As discussed below, Hsp70, Hsp90, and GroEL/GroES each breaks a different subset of the symmetry conditions, thus they use different mechanisms to perform non-equilibrium folding. Despite the difference in their mechanistic details, I present a unifying principle by which symmetry breaking translates into non-equilibrium folding to the native structures.

In addition, I derive an upper bound on the extent to which an ATP-driven chaperone can elevate the native fraction of a substrate above its chaperone-free equilibrium value (Eqn. 53). My results suggest that for substantial non-equilibrium protein folding (*i.e.* $g \gg 1$) to occur, the chaperone—with the possible exception of chaperonins such as GroEL/GroES—must bind to an unstable intermediate conformation of the substrate, and the substrate protein must fold slowly on its own.

Whether Anfinsen's hypothesis holds true for an individual protein can be experimentally tested by comparing the protein's activity in the presence and in the absence of functional ATP-driven chaperones; I have previously proposed new experiments that may provide such tests on Hsp70- and Hsp90-mediated folding [28, 29]. In this work, I propose a potential proteomics-level experiment that may help identify proteins that depend on ATP-driven chaperones for maintenance of their native structures.

My models of non-equilibrium protein folding imply that the native structures of some proteins may be shaped by the chaperone-mediated folding pathways. They raise the possibility of discovering natural proteins—and engineering novel proteins—that adopt different conformations in the presence and absence of the chaperones.

A. Assumptions and notations

To facilitate the exposition, I summarize the assumptions and notations in my model as follows.

- The substrate protein can convert among a set of conformations $\mathbb{S}$, both when it is free in solution and when it is bound to the chaperone. I will use M to denote the misfolded/aggregated conformation and N the native conformation. In addition, I will consider two classes of intermediate conformations: the unfolded and misfold-tending (or aggregation-tending) conformation U, and the non-native but native-tending conformation F. To avoid a proliferation of symbols and to underscore the mechanistic commonality shared by protein folding and activation, in the discussion of kinase activation, I will use M to denote the inactive conformation, N the active conformation, U the inactive-tending conformation, and F the active-tending conformation.

- The chaperone can transition among a set of states $\mathbb{I}$, each state i characterized by its conformational state (*e.g.* open or closed) and the numbers and the types (ATP vs ADP) of bound nucleotides.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

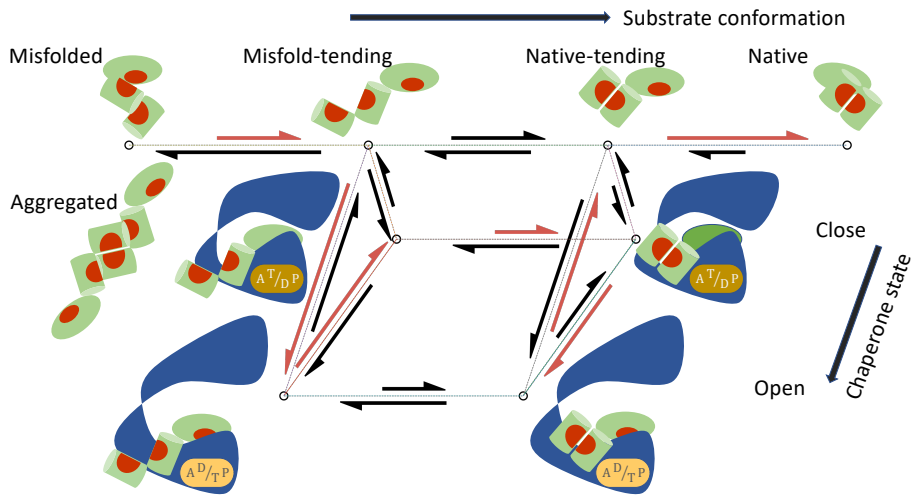

FIG. 1: A mechanistic model of chaperone-mediated non-equilibrium protein folding that couples the state cycle of the chaperone and the conformational dynamics of its substrate. The chaperone undergoes a cycle of open and closed conformations, driven by ATP hydrolysis and nucleotide exchange. The protein substrate can transition among four classes of conformations: Misfolded (M), Misfold-tending (U), Native-tending (F), and Native (N). The lengths of the reaction arrows signify the corresponding reaction rates. The red arrows indicate the predominant folding pathway.

My model includes the following reactions:

- The substrate in conformation S binds to the chaperone in state i with the association rate constant $k_{a,Si}$ and the dissociation rate constant $k_{d,Si}$:

$$S + H_i \underset{k_{d,Si}}{\overset{k_{a,Si}}{\rightleftharpoons}} SH_i \tag{2}$$

- The free substrate in solution converts between conformation S and conformation S':

$$S \underset{k_{S'\to S}}{\overset{k_{S\to S'}}{\rightleftharpoons}} S' \tag{3}$$

The corresponding conformational equilibrium constant is

$$K_{SS'} = \frac{k_{S\to S'}}{k_{S'\to S}} \tag{4}$$

- The substrate bound to the chaperone in state i converts between conformation S and conformation S':

$$SH_i \underset{k_{S'\to S,i}}{\overset{k_{S\to S',i}}{\rightleftharpoons}} S'H_i \tag{5}$$

- The chaperone transitions between state i and state j when it is bound to a substrate in conformation S:

$$SH_i \underset{k_{S,j\to i}}{\overset{k_{S,i\to j}}{\rightleftharpoons}} SH_j \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

II. METHODS AND MATERIALS

A. Proof that symmetry breaking is required for non-equilibrium protein folding

I will show that under the symmetry conditions (Conditions 1–4 in Sec. III A), the steady state concentrations of the substrate satisfies, for any pair of conformations S and S',

$$k_{S\to S'}[S] = k_{S'\to S}[S'] \iff \frac{[S']}{[S]} = \frac{k_{S\to S'}}{k_{S'\to S}} = K_{SS'} \tag{7}$$

where $[S]$ (or $[S']$) is the concentration of the free substrate in conformation S (or S'). Thus the steady state ratio $[S']/[S]$ is unchanged from that in the chaperone-free equilibrium $[S']_{eq}/[S]_{eq} = K_{SS'}$ for any pair of conformations S and S', including $[N]/[M] = [N]_{eq}/[M]_{eq}$, and the chaperone is unable to increase the native concentration of the substrate above that in the equilibrium.

Let $[H_i]$ be the concentration of the chaperone in state i, and $[SH_i]$ the concentration of the substrate in conformation S bound to the chaperone in state i. The steady state condition for the reactions in Eqn. 2, 3, 5, and 6 is

$$\begin{aligned} 0 = \frac{d[S]}{dt} &= \sum_i \left(k_{d,Si}[SH_i] - k_{a,Si}[H_i][S]\right) + \sum_{S'} \left(k_{S'\to S}[S'] - k_{S\to S'}[S]\right) \\ 0 = \frac{d[SH_i]}{dt} &= k_{a,Si}[H_i][S] - k_{d,Si}[SH_i] + \sum_{j\neq i} \left(k_{S,j\to i}[SH_j] - k_{S,i\to j}[SH_i]\right) \\ &\quad + \sum_{S'} \left(k_{S'\to S,i}[S'H_i] - k_{S\to S',i}[SH_i]\right) \end{aligned} \tag{8}$$

According to Condition 1, the ratio $k_{a,S'i}/k_{a,Si}$ does not depend on i. I denote this ratio as

$$\frac{k_{a,S'i}}{k_{a,Si}} = \gamma_{SS'} K_{SS'}^{-1} \tag{9}$$

where $\gamma_{SS'}$ is a number that does not depend on i.

Consider a hypothetical, restricted system in which the substrate bound to the chaperone cannot change conformations (*i.e.*, setting $k_{S\to S',i} = 0$ for all i and all pairs of S and S' in Eqn. 8). Because the reaction $S \rightleftharpoons S'$ is not part of any energy consuming cycle in this restricted system, $[S']/[S] = K_{SS'}$ [29]. Let $\{[S]\} \bigcup \{[SH_i] | i \in \mathbb{I}\}$ be the steady state concentrations of the substrate in conformation S in this restricted system, I will show that

$$[S'] = K_{SS'}[S] \tag{10}$$

$$[S'H_i] = \gamma_{SS'}[SH_i] \tag{11}$$

are the steady state concentrations of the substrate in conformation S', and that $[SH_i]$ and $[S'H_i]$ satisfy

$$k_{S\to S',i}[SH_i] - k_{S'\to S,i}[S'H_i] = 0 \ \forall i \in \mathbb{I} \tag{12}$$

for the original $k_{S\to S',i} > 0$ and $k_{S'\to S,i} > 0$. Thus the steady state concentrations of the restricted system are also the solutions to the original steady state condition in Eqn. 8, and Eqn. 7 holds (it is equivalent to Eqn. 10).

To prove Eqn. 12, consider first an open state i. Thermodynamic cycle closure in the following reaction cycle (which does not consume chemical energy because the chaperone does not change state),

$$\begin{aligned} S &\underset{k_{S'\to S}}{\overset{k_{S\to S'}}{\rightleftharpoons}} S' \\ S' + H_i &\underset{k_{d,S'i}}{\overset{k_{a,S'i}}{\rightleftharpoons}} S'H_i \\ S'H_i &\underset{k_{S\to S',i}}{\overset{k_{S'\to S,i}}{\rightleftharpoons}} SH_i \\ SH_i &\underset{k_{a,Si}}{\overset{k_{d,Si}}{\rightleftharpoons}} S + H_i \end{aligned} \tag{13}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

implies that

$$\frac{k_{S'\to S,i}}{k_{S\to S',i}}\frac{k_{a,S'i}}{k_{d,S'i}}\frac{k_{d,Si}}{k_{a,Si}}K_{SS'} = 1 \tag{14}$$

$$\begin{aligned}
\iff\ & k_{S\to S',i} = k_{S'\to S,i}K_{SS'}\frac{k_{a,S'i}}{k_{a,Si}}\frac{k_{d,Si}}{k_{d,S'i}} = k_{S'\to S,i}\gamma_{SS'}\ (\because \text{Eqn.9 and 37}) \\
\implies\ & k_{S'\to S,i}[S'H_i] - k_{S\to S',i}[SH_i] \\
=\ & [SH_i]\left(k_{S'\to S,i}\gamma_{SS'} - k_{S\to S',i}\right)\ (\because \text{Eqn. 11}) \\
=\ & 0
\end{aligned} \tag{15}$$

If i is a closed state such that $k_{a,Si} = k_{a,S'i} = k_{d,Si} = k_{d,S'i} = 0$, Eqn. 14 no longer holds. According to Condition 4, however, the chaperone can reversibly transition between i and an open state j without the consumption of chemical energy, and according to Condition 3, the transition rates between i and j do not depend on the conformational state of the bound substrate, *i.e.*

$$\frac{k_{S',i\to j}}{k_{S,i\to j}} = \frac{k_{S',j\to i}}{k_{S,j\to i}} = 1 \tag{16}$$

Thus thermodynamic cycle closure in the following reversible reaction cycle

$$\begin{aligned}
S &\underset{k_{S'\to S}}{\overset{k_{S\to S'}}{\rightleftharpoons}} S' \\
S' + H_j &\underset{k_{d,S'j}}{\overset{k_{a,S'j}}{\rightleftharpoons}} S'H_j \\
S'H_j &\underset{k_{S',i\to j}}{\overset{k_{S',j\to i}}{\rightleftharpoons}} S'H_i \\
S'H_i &\underset{k_{S\to S',i}}{\overset{k_{S'\to S,i}}{\rightleftharpoons}} SH_i \\
SH_i &\underset{k_{S,j\to i}}{\overset{k_{S,i\to j}}{\rightleftharpoons}} SH_j \\
SH_j &\underset{k_{a,Sj}}{\overset{k_{d,Sj}}{\rightleftharpoons}} S + H_j
\end{aligned} \tag{17}$$

implies

$$\frac{k_{S'\to S,i}}{k_{S\to S',i}}\frac{k_{S,i\to j}}{k_{S,j\to i}}\frac{k_{S',j\to i}}{k_{S',i\to j}}\frac{k_{a,S'j}}{k_{d,S'j}}\frac{k_{d,Sj}}{k_{a,Sj}}K_{SS'} = 1 \tag{18}$$

$$\begin{aligned}
\implies\ & \frac{k_{S'\to S,i}}{k_{S\to S',i}}\frac{k_{a,S'j}}{k_{a,Sj}}\frac{k_{d,Sj}}{k_{d,S'j}}K_{SS'} = 1\ (\because \text{Eqn. 16}) \\
\implies\ & \frac{k_{S'\to S,i}}{k_{S\to S',i}}\gamma_{SS'} = 1\ (\because \text{Eqn. 9 and 37}) \\
\implies\ & k_{S'\to S,i}[S'H_i] - k_{S\to S',i}[SH_i] \\
=\ & [SH_i]\left(k_{S'\to S,i}\gamma_{SS'} - k_{S\to S',i}\right)\ (\because \text{Eqn. 11}) \\
=\ & 0
\end{aligned} \tag{19}$$

Thus Eqn. 12 is true for both open and closed states.

To prove that $\{[S']\}\bigcup\{[S'H_i]|i\in\mathbb{I}\}$ in Eqn. 10 and 11 satisfy the steady state condition Eqn. 8 (swapping S' and S), I only need to show that, for the reactions in Eqn. 2 and 6, the flux in each reaction involving the substrate in conformation S' is $\gamma_{SS'}$ times the flux of the corresponding reaction involving the substrate in conformation S, because $\{[S]\}\bigcup\{[SH_i]|i\in\mathbb{I}\}$ satisfy Eqn. 8 and the reactions in Eqn. 3 and 5 have zero flux (Eqn. 10 and 12).

Let $J_{S,ij} = k_{S,i\to j}[SH_i] - k_{S,j\to i}[SH_j]$ be the reactive flux of the state transition for the chaperone bound to a substrate in conformation S (Eqn. 6) and $J^a_{Si} = k_{a,Si}[H_i][S] - k_{d,Si}[SH_i]$ be the reactive flux of the substrate in conformation S binding to the chaperone in state i (Eqn. 2). The corresponding reactive fluxes for the substrate in conformation S' are

$$\begin{aligned}
J_{S',ij} &= k_{S',i\to j}[S'H_i] - k_{S',j\to i}[S'H_j] \\
&= \gamma_{SS'}\left(k_{S,i\to j}[SH_i] - k_{S,j\to i}[SH_j]\right)\ (\because \text{Eqn. 11 and Condition 3}) \\
&= \gamma_{SS'}J_{S,ij}
\end{aligned} \tag{20}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

$U \underset{k_{M\to U}}{\overset{k_{U\to M}}{\rightleftharpoons}} M$	Misfolding and aggregation; $K_M = k_{U\to M}/k_{M\to U}$
$U \underset{k_{F\to U}}{\overset{k_{U\to F}}{\rightleftharpoons}} F$	Transition between intermediate conformations; $K_F = k_{U\to F}/k_{F\to U}$
$F \underset{k_{N\to F}}{\overset{k_{F\to N}}{\rightleftharpoons}} N$	Folding to native structure; $K_N = k_{F\to N}/k_{N\to F}$
$S + O \underset{k_{d,S}}{\overset{k_{a,S}}{\rightleftharpoons}} SO$	Substrate in $S = U, F$ conformations binding to the open chaperone
$SO \underset{k_{S,C\to O}}{\overset{k_{S,O\to C}}{\rightleftharpoons}} SC$	Transition of chaperone between open and close states
$UH \underset{k_{F\to U,H}}{\overset{k_{U\to F,H}}{\rightleftharpoons}} FH$	Conversion of protein bound to chaperone in $H = C, O$ states

TABLE I: The reactions in chaperone-mediated protein folding. These reactions are depicted in Fig. 1. ATP hydrolysis and nucleotide exchange occur and inject chemical energy in the chaperone cycle.

and

$$\begin{aligned} J^a_{S'i} &= k_{a,S'i}[H_i][S'] - k_{d,S'i}[S'H_i] \\ &= k_{a,S'i}[H_i]K_{SS'}[S] - k_{d,Si}\gamma_{SS'}[SH_i] \ (\because \text{ Eqn. 10, 11, and 37 }) \\ &= \gamma_{SS'}\left(k_{a,Si}[H_i][S] - k_{d,Si}[SH_i]\right) \ (\because \text{Eqn. 9}) \\ &= \gamma_{SS'} J^a_{Si} \end{aligned} \tag{21}$$

Q.E.D.

B. Derivation of the upper bound of the native concentration at the steady state of non-equilibrium folding

To derive the upper bound in Eqn. 48, consider the reactions in Table I. These are simplifications of the reactions in Eqn. 2, 3, 5, and 6: only a substrate in intermediate conformations $S = U, F$ can bind to the chaperone (see Sec. III B 1), and only two chaperone states, open (O) and closed (C), are considered. The result holds as long as the substrate binds to all chaperone open states with the same association and dissociation rate constants, *i.e.*

$$\begin{aligned} k_{a,Si} &= k_{a,S} \\ k_{d,Si} &= k_{d,S} \ \forall \text{ open state } i \end{aligned} \tag{22}$$

Let

$$J_{FU} = k_{F\to U}[F] - k_{U\to F}[U] \tag{23}$$

be the reactive flux from F to U. At the steady state, there is no net flux into or out of any molecular species, implying

$$\begin{aligned} J_{FU} &= k_{a,U}[U][O] - k_{d,U}[UO] \\ &= k_{d,F}[FO] - k_{a,F}[F][O] \end{aligned} \tag{24}$$

Because no external chemical energy is consumed in the reaction cycle of

$$U + O \rightleftharpoons UO \rightleftharpoons FO \rightleftharpoons F + O \rightleftharpoons U + O \tag{25}$$

we have

$$\frac{k_{F\to U,O}}{k_{U\to F,O}} \cdot \frac{k_{a,F}}{k_{d,F}} \cdot \frac{k_{d,U}}{k_{a,U}} \cdot \frac{k_{U\to F}}{k_{F\to U}} = 1 \tag{26}$$

Thus

$$\begin{aligned} \frac{k_{F\to U,O}[FO]}{k_{U\to F,O}[UO]} &= \frac{k_{F\to U,O}}{k_{U\to F,O}} \cdot \frac{k_{a,F}}{k_{d,F}} \cdot \frac{k_{d,U}}{k_{a,U}} \cdot \frac{k_{U\to F}}{k_{F\to U}} \\ &\quad \cdot \frac{k_{d,F}[FO]}{k_{a,F}[F][O]} \cdot \frac{k_{a,U}[U][O]}{k_{d,U}[UO]} \cdot \frac{k_{F\to U}[F]}{k_{U\to F}[U]} \\ &= \frac{k_{d,F}[FO]}{k_{d,F}[F][O]} \cdot \frac{k_{a,U}[U][O]}{k_{d,U}[UO]} \cdot \frac{k_{F\to U}[F]}{k_{U\to F}[U]} \end{aligned} \tag{27}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

If J_{FU} in Eqn. 23 and 24 is positive, all three ratios on the right hand side of Eqn. 27 are greater than 1; if $J_{FU} < 0$, they are all smaller than 1. Thus the reactive flux

$$J_{UF,O} = k_{U\to F,O}[UO] - k_{F\to U,O}[FO] \tag{28}$$

must be of the opposite sign of J_{FU}.

If the chaperone drives the substrate toward the native structure, we have $[F]/[U] > K_F = k_{U\to F}/k_{F\to U}$, implying $J_{FU} > 0$ and $J_{UF,O} < 0$. Because the flux from conformation F to U in free substrates must balance the total flux from conformation U to F in chaperone-bound substrates, the steady state reactive flux of the reaction $UC \rightleftharpoons FC$

$$J_{UF,C} = k_{U\to F,C}[UC] - k_{F\to U,C}[FC] \tag{29}$$

satisfies

$$J_{FU} = J_{UF,C} + J_{UF,O} < J_{UF,C} \tag{30}$$

Thus

$$\begin{aligned} & k_{F\to U}[F] - k_{U\to F}[U] < k_{U\to F,C}[UC] - k_{F\to U,C}[FC] \\ \implies & k_{F\to U}[F] < k_{U\to F}\left([U] + \frac{k_{U\to F,C}}{k_{U\to F}}[UC]\right) \equiv k_{U\to F}([U] + \alpha[UC]) \\ \implies & [F] < K_F \cdot \max(1,\alpha) \cdot ([U] + [UC]) \end{aligned} \tag{31}$$

We also have, per Eqn. 23 and 24,

$$\begin{aligned} & J_{FU} = k_{F\to U}[F] - k_{U\to F}[U] = k_{a,U}[O][U] - k_{d,U}[UO] < k_{a,U}[O][U] \\ \implies & (k_{U\to F} + k_{a,U}[O])[U] > k_{F\to U}[F] \end{aligned} \tag{32}$$

At the steady state, there is no net flux in $M \rightleftharpoons U$ or in $F \rightleftharpoons N$, thus

$$\begin{aligned} [M] &= K_M[U] \\ [N] &= K_N[F] \end{aligned} \tag{33}$$

Because

$$[M] + [U] + [UC] + [F] + [N] < P_0 \tag{34}$$

we have

$$\begin{aligned} P_0 &> (K_F^{-1}\max(1,\alpha)^{-1} + 1)[F] + [M] + [N] \quad (\because \text{Eq. 31}) \\ &= (K_F^{-1}\max(1,\alpha)^{-1} + 1)[F] + K_M[U] + K_N[F] \\ &> (K_F^{-1}\max(1,\alpha)^{-1} + 1)[F] + K_M\frac{k_{F\to U}}{k_{U\to F} + k_{a,U}[O]}[F] + k_N[F] \quad (\because \text{Eq. 32}) \end{aligned} \tag{35}$$

Thus

$$[F] < \left(K_M K_F^{-1}\left(1 + \frac{k_{a,U}[O]}{k_{U\to F}}\right)^{-1} + K_F^{-1}\max(1,\alpha)^{-1} + 1 + K_N\right)^{-1} P_0 \tag{36}$$

and plugging in Eqn. 33 yields the upper bound in Eqn. 48.

III. RESULTS

A. Non-equilibrium folding requires kinetic symmetry breaking

I present a set of four symmetry conditions, if all satisfied, forbids an ATP-driven chaperone from elevating the native concentration $[N]$ of its substrate above the chaperone-free equilibrium concentration $[N]_{eq}$. A chaperone must break at least one of these symmetry conditions to be able to convert chemical energy into non-equilibrium stabilization of the native structure of the substrate. As I discuss below, different chaperones break different symmetry conditions, corresponding to different mechanisms of non-equilibrium protein folding and activation. The symmetry conditions are as follows:

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1. The ratio of association rate constants $k_{a,S'i}/k_{a,Si}$ does not depend on the chaperone state i for all pair of substrate conformations S and S' and for all open state i.

2. The dissociation rate constant $k_{d,Si}$ does not depend on the substrate conformation S, *i.e.*

$$k_{d,Si} = k_{d,i} \tag{37}$$

   for all open state i and for all conformation S.

3. The transition rates between chaperone states is independent of the conformation of the bound substrate, *i.e.* $k_{S,i\to j}$ does not depend on S for all pair (i, j).

4. For every closed state i of the chaperone, there is an open state j, such that the chaperone can reversibly transition between states j and i without consuming chemical energy.

In Sec. II A of Materials and Methods, I prove that if these four symmetry conditions are all satisfied, the ratio between the concentrations of the free substrate in any two conformations—say, S and S'—at the chaperone-mediated steady state is unchanged from that in the chaperone-free equilibrium, *i.e.* $[S']/[S] = [S']_{eq}/[S]_{eq}$, which implies $[N]/[M] = [N]_{eq}/[M]_{eq}$. Because chaperone-binding reduces the total concentration of the free substrate, the native concentration of the free substrate will be lower in the presence of chaperones than in the absence of chaperones, *i.e.* $g < 1$. (As noted in the Introduction, I only consider non-equilibrium folding to a free native protein.)

The above results regarding symmetry conditions hold for an arbitrary number of substrate conformations. For simplicity, I will assume only four representative conformations in the substrate, $\mathbb{S} = \{M \equiv \text{misfolded}, U \equiv \text{misfold-tending}, F \equiv \text{native-tending}, N \equiv \text{native}\}$, in the following discussion.

1. *Requisites for breaking the binding and unbinding symmetries (Conditions 1 and 2)*

The binding symmetry, Condition 1, is trivially satisfied if there is only one open chaperone state to which the substrate binds, or if the substrate binding rate does not depend on the chaperone state, *i.e.* $k_{a,Si} = k_{a,S}$. Note that the substrate in different conformations S may bind to the chaperone at different rates $k_{a,S}$, *e.g.* the substrate in an unfolded structure may bind to the chaperone faster than the substrate in a near-native structure, which is a common assumption in models of chaperone-mediated folding [32], but this conformation-selective binding alone does not permit non-equilibrium folding (defined by $g > 1$).

Condition 1 is approximately satisfied if the substrate in different conformations and the chaperone in different states bind using the same interface. In this case, the association rate constant is approximately

$$k_{a,Si} = p_S \times f_i \times k_a \tag{38}$$

where p_S is the probability that the binding surface on the substrate becomes accessible in conformation S, f_i is the probability that the binding surface on the chaperone is accessible in state i, and k_a is the intrinsic binding rate between the two binding surfaces once exposed (Eqn. 38 assumes that the conformational fluctuations exposing and occluding the binding surfaces are fast compared to the overall binding). The ratio

$$\frac{k_{a,S'i}}{k_{a,Si}} = \frac{p_{S'}}{p_S} \tag{39}$$

thus satisfies Condition 1.

Condition 1 is violated if the substrate binds to different binding surfaces on the chaperone depending on both the substrate conformation and the chaperone (open) state. This requires that the chaperone possesses multiple open states in which different binding surfaces are exposed. There has not been experimental demonstration of any ATP-driven chaperone breaking this symmetry condition.

The unbinding symmetry, Condition 2, is approximately satisfied if the chaperone binds to the substrate in different conformations using the same binding interface. The symmetry is broken if the substrate in different conformations form different protein-protein interactions with the chaperone.

In one limit of such binding interface change, the substrate in the misfold-tending conformation U with a slow dissociation rate $k_{d,U}$ may bind to the open chaperone and, after the chaperone closes, change to the native-tending conformation F in which its chaperone-binding surface is lost, so that when the chaperone opens again after the ATP-driven cycle the substrate unbinds rapidly with a fast dissociation rate $k_{d,F} \gg k_{d,U}$. This may happen in chaperones that can retain a substrate without a contact interface while allowing the bound substrate to change

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

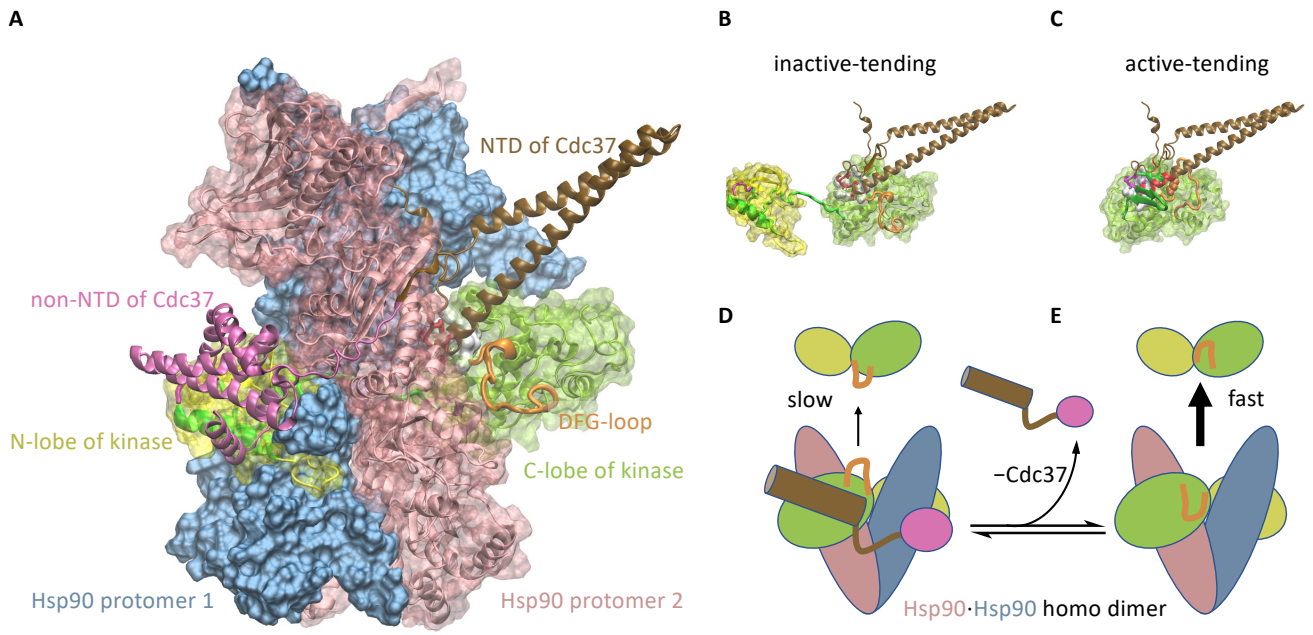

FIG. 2: Cochaperone Cdc37 enables a client kinase in different conformations to unbind from Hsp90 at different rates. **A**. The Hsp90-Cdk4-Cdc37 complex structure (PDB: 5FWM). The closed Hsp90 homo dimer clamps a partially unfolded Cdk4 kinase, and Cdc37 simultaneously binds to Cdk4 and Hsp90. **B**. Cdc37 can bind to the kinase in the inactive-tending conformation. **C**. Steric clashes prevent Cdc37 from binding to the kinase in the active-tending conformation, due to its DFG-loop configuration and other conformational features. **D**. Cdc37 helps to retain an inactive-tending kinase molecule inside the open Hsp90, resulting in slow unbinding of the kinase from the Hsp90. **E**. Without Cdc37, an active-tending kinase molecule unbinds rapidly from the open Hsp90.

conformation from U to F. Hsp90 and GroEL/ES are two such examples: Hsp90 clamps its client kinase between its closed homo dimer with a central hole that may accommodate substantial conformational changes in the client [29, 40], and GroEL/ES holds the substrate in its cavity inside which the substrate may fold [41]. These two chaperones may break Condition 2 by this mechanism and thus perform non-equilibrium protein folding.

Cochaperones that simultaneously bind to the chaperone and to the misfold-tending, but not the native-tending, conformation of the substrate may help break Condition 2. When the substrate in the misfold-tending conformation is bound to the cochaperone, the substrate-cochaperone complex together has an extended chaperone-binding surface with contributions from both the substrate and the cochaperone, which decreases the substrate's dissociation rate from the chaperone. Binding to and unbinding from the cochaperone, a substrate in the misfold-tending conformation has, in effect, a slower dissociation rate than the substrate in the native-tending conformation. One case in point may be that of Cdc37-assisted kinase activation by Hsp90, as discussed in the following.

2. *Cdc37 enables Hsp90 to differentiate between the active-tending and inactive-tending conformations of a client kinase*

Cdc37 is a cochaperone that specializes in assisting Hsp90 to activate client kinases [30, 42]. Experimental evidence suggests that Cdc37 binds to a locally unfolded conformation of the client kinase [43], and that Cdc37 can simultaneously bind to a client kinase and Hsp90 [40, 44, 45]. Based on the cryo-EM structure of the Hsp90-kinase-Cdc37 complex [40] (Fig. 2**A**), I have previously proposed a simple mechanism for Cdc37 to distinguish between the inactive-tending (U) and active-tending (F) kinase conformations, binding to the former with higher affinity than to the latter: in the inactive-tending conformation, the disordered DFG-loop of the kinase does not interfere with Cdc37

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

binding, whereas in the active-tending conformation, the DFG-loop may be ordered into a configuration that results in steric clashes with Cdc37 [29] (Fig. 2**B**,**C**). Thus, Cdc37 can help Hsp90 retain an inactive-tending client more than an active-tending client, and the effective rate of dissociation from Hsp90 is higher for a client in the active-tending conformation than for a client in the inactive-tending conformation (Fig. 2**D**,**E**), breaking symmetry Condition 2.

This mechanism implies the following reaction path of Hsp90-mediated kinase activation:

$$\begin{aligned} &U \overset{+\mathrm{Cdc37}}{\rightleftharpoons} U \cdot \mathrm{Cdc37} \overset{+\mathrm{Hsp90_{open}}}{\rightleftharpoons} \mathrm{Hsp90_{open}} \cdot U \cdot \mathrm{Cdc37} \rightleftharpoons \mathrm{Hsp90_{closed}} \cdot U \cdot \mathrm{Cdc37} \\ &\overset{-\mathrm{Cdc37}}{\rightleftharpoons} \mathrm{Hsp90_{closed}} \cdot U \rightleftharpoons \mathrm{Hsp90_{closed}} \cdot F \overset{\mathrm{ATP \to ADP+Pi}}{\rightleftharpoons} \mathrm{Hsp90_{open}} \cdot F \overset{-\mathrm{Hsp90_{open}}}{\rightleftharpoons} F \end{aligned} \tag{40}$$

Clearly, this mechanism requires that Cdc37 can dissociate from the Hsp90-kinase complex after Hsp90 closes. This requirement is indeed consistent with the observed structure of the Hsp90-kinase-Cdc37 complex: the closed Hsp90 clamps the client kinase between its N- and C-lobes to prevent the kinase from unbinding, but Cdc37 wraps around the exterior of Hsp90 so that it can disengage from the closed Hsp90 (Fig. 2**A**).

3. Cochaperone Hsp40 enables differential ATP hydrolysis by Hsp70 bound to a substrate in different conformations

Hsp70-mediated protein folding is an example of breaking symmetry Condition 3. The Hsp70 chaperones, such as the bacterial DnaK, adopts an open conformation when its nucleotide binding domain (NBD) is occupied by ATP. Upon ATP hydrolysis, Hsp70 changes to a closed conformation [46, 47] (Fig. 3**A**). By itself, Hsp70 has low basal ATP hydrolysis rate, but the J domain from the Hsp40 cochaperones—also known as J proteins—can stimulate Hsp70 and drastically increase its ATP hydrolysis rate [48, 49].

Both Hsp40 and Hsp70 bind to exposed hydrophobic sites on a substrate protein [50, 51] (Fig. 3**A**-**C**). Consequently, a substrate with multiple exposed hydrophobic sites may simultaneously bind to an Hsp70 and an Hsp40. This induces the proximity between the chaperone and the cochaperone, resulting in accelerated ATP hydrolysis in Hsp70 and its transition to the closed state. Because a substrate in the misfold-tending conformation often exposes more hydrophobic sites than a substrate in the native-tending conformation [52], an Hsp70 bound to the former is more likely to be stimulated by a nearby Hsp40 bound to the same substrate molecule than an Hsp70 bound to the latter. By recruiting Hsp40 to accelerate the ATP hydrolysis in Hsp70, a substrate in the misfold-tending conformation induces a higher rate of transition by Hsp70 from the open state to the closed state than a substrate in the native-tending conformation, *i.e.* $k_{U,\mathrm{open}\to\mathrm{closed}} > k_{F,\mathrm{open}\to\mathrm{closed}}$, breaking symmetry Condition 3 (Fig. 3**D**,**E**).

As a result of this symmetry breaking, an Hsp70 bound to a substrate in the misfold-tending conformation is more likely to be closed than one bound to a substrate in the native-tending conformation. Thus, a substrate is on average more quickly released from the Hsp70 if it is in the native-tending conformation than if it is in the misfold-tending conformation. This difference biases the substrate toward the native conformation [28].

4. Hsp70 and Hsp90 perform non-equilibrium folding by preferentially releasing substrate proteins in native-tending conformations

The cochaperone Cdc37 helps break symmetry Condition 2 in Hsp90-mediated kinase activation. The cochaperone Hsp40 helps break symmetry Condition 3 in Hsp70-mediated protein folding. Despite breaking different symmetries, Hsp70 and Hsp90 share the same kinetic consequence: both chaperones release a bound substrate in the native-tending (F) conformation faster than a bound substrate in the misfold-tending (U) conformation.

To see how this kinetic asymmetry promotes the native concentration, consider first a system in which the symmetry conditions are satisfied (Fig. 4**A**). A substrate in the U conformation binds to the chaperone faster than a substrate in the F conformation. As a result, the reactive flux through the ATP-driven cycle of a chaperone bound to a substrate in the U conformation is higher than that through the cycle of a chaperone bound to a substrate in the F conformation. But kinetic symmetry ensures that, at the steady state, the flux of U binding to the chaperone is the same as the flux of U unbinding from the chaperone, the same holds true for F binding to and unbinding from the chaperone, and there is no net flux between U and F. Under the symmetry conditions, there are two parallel, independent chaperone cycles with respective reactive fluxes:

$$J_{S+\mathrm{Hsp}} = (S \to S \cdot \mathrm{Hsp} \to S \cdot \{\text{states of Hsp} \cdots\} \to S \cdot \mathrm{Hsp} \to S) \quad \text{for } S = U, F, \tag{41}$$

and

$$J_{U+\mathrm{Hsp}} > J_{F+\mathrm{Hsp}} \tag{42}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

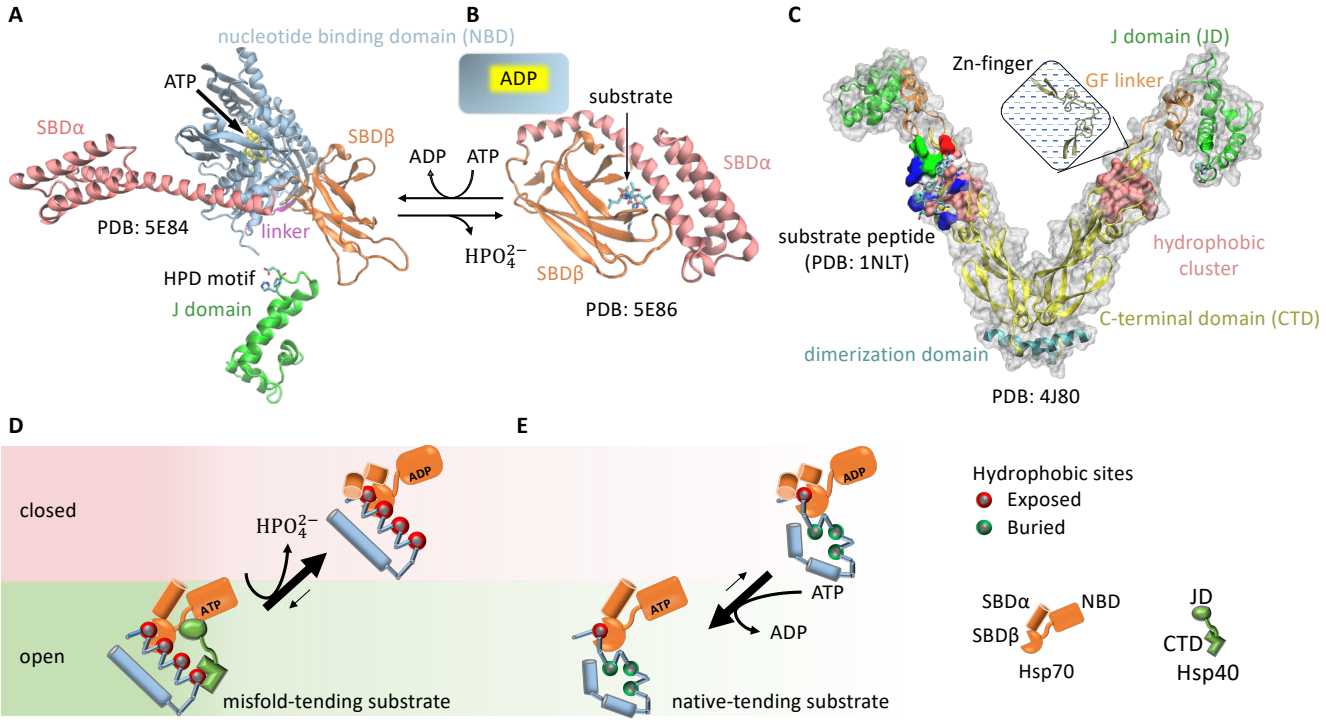

FIG. 3: Cochaperone Hsp40 enables Hsp70 to change the balance between its open and closed states in response to the conformation of a bound substrate. **A**. The ATP-bound, open state of Hsp70, which allows rapid binding and unbinding of the substrate. **B**. The ADP-bound, closed state of Hsp70, with slow binding and unbinding of the substrate. SBD: substrate binding domain. **C**. The structure of the Hsp40 cochaperone, including CTD that can bind to exposed hydrophobic sites on a substrate and the J domain that can stimulate the ATP hydrolysis of Hsp70. **D**. An Hsp70 bound to a misfold-tending substrate molecule with many exposed hydrophobic sites is likely to be in proximity to an Hsp40 bound to the same substrate molecule, thus the Hsp70 will be stimulated in ATP hydrolysis, which drives the Hsp70 to its ADP-bound, closed state. **E**. An Hsp70 bound to a native-tending substrate molecule with few exposed hydrophobic sites is unlikely to have a nearby Hsp40 and thus unlikely to be stimulated in ATP hydrolysis, and nucleotide exchange drives the Hsp70 toward its ATP-bound, open state.

But there is no net flux between U and F:

$$J_{UF} = k_{U\to F}[U] - k_{F\to U}[F] = 0 \tag{43}$$

And the ratio between F and U is unchanged from the chaperone-free equilibrium:

$$\frac{[F]}{[U]} = \frac{k_{U\to F}}{k_{F\to U}} = K_{UF} = \frac{[F]_{eq}}{[U]_{eq}} \tag{44}$$

Symmetry breaking disrupts the independence between this pair of chaperone cycles. The release of a substrate in the U conformation from the chaperone is inhibited: in the case of Hsp90, Cdc37 helps the chaperone retain the bound client kinase; in the case of Hsp70, Hsp40-stimulated ATP hydrolysis and closure in Hsp70 diminishes the reactive flux to re-open the chaperone (Fig. 4**B**). This forces part of the reactive flux in $J_{U+\mathrm{HSP}}$ after the binding of the substrate to be diverted into the reactive flux of conformation conversion in the bound substrate:

$$J_{U\cdot\mathrm{Hsp}\to F\cdot\mathrm{Hsp}} > 0 \tag{45}$$

This in turn leads to a corresponding increase in the reactive flux of the chaperone's release of the substrate in the F conformation, which increases $[F]$ such that

$$\frac{[F]}{[U]} > \frac{[F]_{eq}}{[U]_{eq}} \tag{46}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

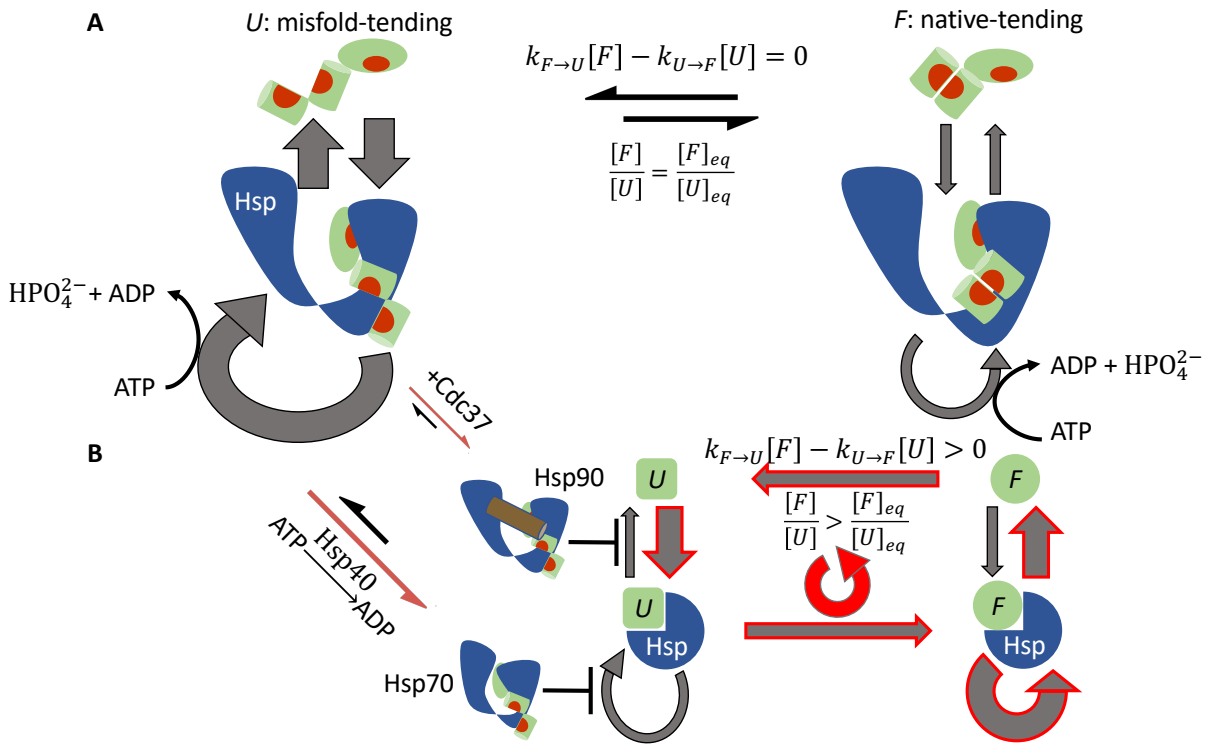

FIG. 4: Reactive flux in chaperone-mediated non-equilibrium protein folding. **A**. Under the symmetry conditions, there are two independent ATP-driven chaperone cycles: one with a higher reactive flux for a substrate in the misfold-tending (U) conformation (left) and one with a lower reactive flux for a substrate in the native-tending (F) conformation (right). There is no net flux between the substrate's two conformations, and the ratio $[F]/[U]$ is the same as its chaperone-free equilibrium value. **B**. Cochaperones break the kinetic symmetry. The release of a substrate in the U conformation from the chaperone is inhibited: Cdc37 assists Hsp90 to retain the substrate and Hsp40 stimulates ATP hydrolysis and closure of Hsp70. This restricts the reactive flux to release a substrate in the U conformation, forcing a partial diversion of the flux to the conformation conversion from $U \cdot \text{Hsp}$ to $F \cdot \text{Hsp}$ and resulting in a net reactive flux of $U \to U \cdot \text{Hsp} \to F \cdot \text{Hsp} \to F \to U$ (red cycle), which elevates the ratio $[F]/[U]$ above its chaperone-free equilibrium value.

Thus symmetry breaking biases the substrate toward the native-tending conformation and elevates the native concentration.

5. *The potential role of sequential hydrolyses of multiple ATPs in the chaperone cycle*

Breaking symmetry Condition 4 permits a net reactive flux along the following path that promotes the native-tending conformation F over the misfold-tending conformation U:

$$U \xrightarrow{+\text{Hsp}_{\text{open}}} U \cdot \text{Hsp}_{\text{open|closed}} \xrightarrow{\text{ATP}\to\text{ADP+Pi}} U \cdot \text{Hsp}_{\text{closed}} \to F \cdot \text{Hsp}_{\text{closed}}$$
$$\xrightarrow{\text{ATP}\to\text{ADP+Pi}} F \cdot \text{Hsp}_{\text{open}} \xrightarrow{-\text{Hsp}_{\text{open}}} F \tag{47}$$

A non-zero net flux of $U \cdot \text{Hsp}_{\text{closed}} \to F \cdot \text{Hsp}_{\text{closed}}$ does not violate thermodynamic cycle closure in this case because the reaction cycle in Eqn. 17 is no longer reversible—ATP hydrolysis occurs and chemical energy is consumed in that cycle—and thus Eqn. 18 and 19 no longer hold.

To break symmetry Condition 4, at least one closed state of the chaperone must be separated from all the open states by ATP hydrolysis. This requires at least two ATP to be hydrolyzed sequentially—not synchronously—per chaperone cycle, and the substrate has to change conformation between two ATP hydrolyses. Examples include Hsp90

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

that hydrolyzes two ATP molecules sequentially in its cycle [53] and the group II chaperonins in eukaryotes, such as TRiC/CCT, that hydrolyzes up to eight ATPs sequentially [54, 55]. The role of such sequential ATP hydrolyses—and the consequent symmetry breaking of Condition 4—in non-equilibrium protein folding is an open question.

B. An upper bound of non-equilibrium protein folding and its implications

Having established the symmetry breaking requirements for non-equilibrium folding, I now derive an upper bound on the folding capacity of an ATP-driven chaperone. The key result is

$$[N] < \frac{K_N}{K_M K_F^{-1}\left(1+\frac{k_{a,U}[O]}{k_{U\to F}}\right)^{-1} + K_F^{-1}\max(1,\alpha)^{-1} + 1 + K_N} P_0 \tag{48}$$

where $[O]$ is the concentration of free chaperone in the open state, the equilibrium constants K_N, K_M, K_F, the kinetic rate constants $k_{U\to F}$ and $k_{a,U}$, and their corresponding reactions are summarized in Table I, and

$$\alpha \equiv \frac{k_{U\to F,C}}{k_{U\to F}} \tag{49}$$

is an acceleration factor to indicate any potential rate change in conformation conversion when the substrate is bound to the closed chaperone. The proof of Eqn. 48 is given in Sec. II B of Methods and Materials.

Eqn. 48 gives a general upper bound applicable to any ATP-driven chaperone. The folding capacity of a specific type of chaperone need to be calculated by detailed models [28, 29], but it cannot exceed that given by Eqn. 48. This result allows an analysis of the common key factors in non-equilibrium folding without considering the mechanistic details of specific chaperones.

Introducing a combined equilibrium constant for the reaction $U \overset{\tilde{K}_F}{\rightleftharpoons} (F+N)$

$$\tilde{K}_F \equiv \frac{[F]_{eq}+[N]_{eq}}{[U]_{eq}} = (1+K_N)K_F \tag{50}$$

The upper bound in Eqn. 48 can be written as

$$[N] < \frac{1}{K_M \tilde{K}_F^{-1}\left(1+\frac{k_{a,U}[O]}{k_{U\to F}}\right)^{-1} + \tilde{K}_F^{-1}\max(1,\alpha)^{-1} + 1} \cdot \frac{K_N}{1+K_N} P_0 \tag{51}$$

Compare this to the native concentration in the chaperone-free equilibrium

$$\begin{aligned}[N]_{eq} &= \frac{K_N}{K_M K_F^{-1} + K_F^{-1} + 1 + K_N} P_0 \\ &= \frac{1}{K_M \tilde{K}_F^{-1} + \tilde{K}_F^{-1} + 1} \cdot \frac{K_N}{1+K_N} P_0\end{aligned} \tag{52}$$

The non-equilibrium gain factor is thus bounded by

$$g = \frac{[N]}{[N]_{eq}} < \frac{K_M \tilde{K}_F^{-1} + \tilde{K}_F^{-1} + 1}{K_M \tilde{K}_F^{-1}\left(1+\frac{k_{a,U}[O]}{k_{U\to F}}\right)^{-1} + \tilde{K}_F^{-1}\max(1,\alpha)^{-1} + 1} \tag{53}$$

1. *Chaperones bind to unstable intermediate conformations of substrates to drive non-equilibrium folding*

An implication of Eqn. 53 is that ATP-driven chaperones must bind to an intermediate unfolded conformation (U) of the substrate, not to the misfolded conformation (M) itself, to perform non-equilibrium folding, unless the conformation conversion of a substrate is accelerated when bound to the chaperone (*i.e.* $k_{U\to F,C} > k_{U\to F}$ hence $\alpha > 1$). This can be demonstrated by contradiction. If the substrate does not have an intermediate misfold-tending conformation and the chaperone directly binds to the misfolded conformation, *i.e.* M and U are the same, Eqn. 48 reduces to (by setting $K_M = 0$)

$$[N] < \left(\tilde{K}_F^{-1}\max(1,\alpha)^{-1} + 1\right)^{-1} \cdot \frac{K_N}{1+K_N} P_0 \tag{54}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and the upper bound of the non-equilibrium gain factor becomes

$$g < \frac{\tilde{K}_F^{-1} + 1}{\tilde{K}_F^{-1} \max(1, \alpha)^{-1} + 1} \tag{55}$$

In the absence of a mechanism for the substrate to accelerate its conformation conversion when it is bound to the chaperone ($\alpha \leq 1$), $g \leq 1$, the chaperone cannot elevate the native concentration.

To my knowledge, accelerated folding of protein substrates when bound to a chaperone has only been reported for the GroEL/ES chaperonins [41, 56–60]. In general, steric hindrance from the chaperone is more likely to impede rather than to accelerate conformation conversions in a bound substrate; this impedance was observed for the rhodanese protein trapped in GroEL/GroES by a single-molecule experiment [61]. For Hsp90 and Hsp70, there has not been any experimental demonstration that a substrate exhibits faster conformation conversions when bound to the chaperone than when free in the solution. This suggests that chaperones, with the potential exceptions of chaperonins, must bind to intermediate unfolded conformations of the substrate proteins to drive non-equilibrium protein folding.

Assuming $\alpha \leq 1$, the upper bound on the non-equilibrium gain factor becomes

$$g < g_{max} = \frac{K_M \left(\tilde{K}_F + 1\right)^{-1} + 1}{K_M \left(\tilde{K}_F + 1\right)^{-1} \left(1 + \frac{k_{a,U}[O]}{k_{U\to F}}\right)^{-1} + 1} \tag{56}$$

For the gain factor to substantially exceed 1, the following must be true:

$$K_M \left(\tilde{K}_F + 1\right)^{-1} \gg 1 \implies K_M \gg 1 \tag{57}$$

Eqn. 57 implies that the intermediate conformation U to which the chaperone binds must be intrinsically unstable and it will predominantly convert to the misfolded conformation M in the absence of the chaperone. This result is intuitive: if the chaperone binds to a dominant conformation of the substrate, it will trap a substantial fraction of the substrate and hinder its folding to the native structure. As a result, the chaperone will be unable to elevate the native concentration. The difficulty to observe the chaperone-binding conformations in biophysical experiments [62] attests to their transiency.

2. *Chaperones stabilize the native structures of slow-folding proteins*

Non-equilibrium folding also requires, as implied by Eqn. 56 and $g \gg 1$,

$$\frac{k_{a,U}[O]}{k_{U\to F}} \gg 1 \iff k_{U\to F} \ll k_{a,U}[O] \tag{58}$$

Taken together, Eqn. 57 and 58 suggest that chaperones stabilize the native structures of slow-folding proteins. Assuming the binding rate constant to be on the order of $k_{a,U} \sim 10^6$ /M/s, the spontaneous (*i.e.* without chaperones) refolding rate of the protein, which is approximately $K_M^{-1} k_{U\to F}$, should be much slower than 1 /s to admit of effective non-equilibrium folding by chaperones at a concentration of $[O] \sim 1$ μM.

3. *ATP-driven chaperones buffer destabilizing mutations*

About 18% of protein molecules in the cell harbor at least one missense mutation due to errors in translation [63]. In addition, proteins incur mutations due to germline and somatic gene polymorphism [64]. Given that about 30-40% of random substitutions disrupt protein functions [65, 66], most probably by loss-of-folding [67, 68], it is likely that many cellular protein molecules have compromised thermal stability and the native structures of some will not be the free energy minima. ATP-driven chaperones may buffer such destabilizing mutations [69, 70] and maintain the native concentrations of these proteins by non-equilibrium folding [28].

The missense mutations may alter one or more of the transition rates and the equilibrium constants in protein folding dynamics: *e.g.* it may decrease the thermal stability of the protein by increasing K_M, decreasing $k_{U\to F}$ or increasing $k_{F\to U}$ (hence decreasing $K_F = k_{U\to F}/k_{F\to U}$), or decreasing K_N. Assuming $\alpha \leq 1$ as discussed above, the

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

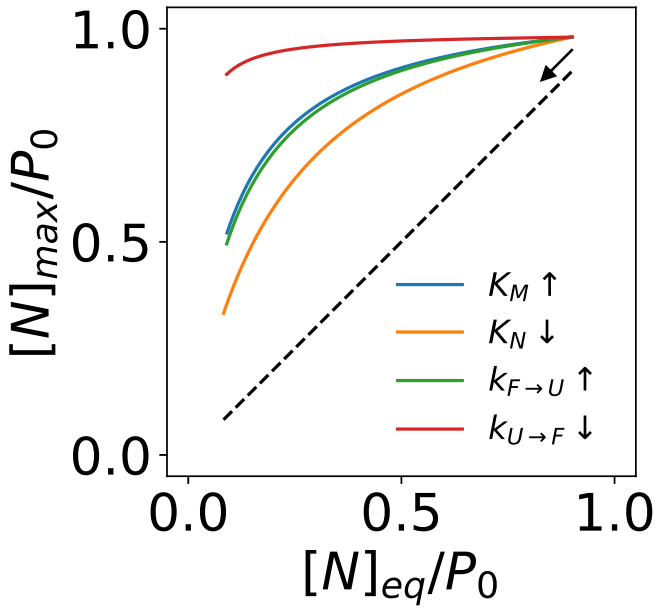

FIG. 5: The capacity of ATP-driven chaperones to maintain elevated native fractions in response to destabilizing mutations in a protein substrate. The kinetic parameters of the wild-type protein are $K_M = 10^2$, $K_F = 10$, $K_N = 10^2$, $k_{U\to F} = 0.1s^{-1}$, and $k_a = 10^6 M^{-1} \cdot s^{-1}$; the concentration of the open chaperone is set to $[O] = 1\mu$M.

maximum native concentration mediated by a chaperone is

$$\begin{aligned}[N]_{max} &= g_{max}[N]_{eq} \\ &= \frac{1}{K_M(\tilde{K}_F+1)^{-1}\left(1+\frac{k_a[O]}{k_{U\to F}}\right)^{-1}+1}\frac{1}{1+\tilde{K}_F^{-1}}\frac{K_N}{1+K_N}P_0 \end{aligned} \tag{59}$$

Eqn. 59 suggests that the capacity of ATP-driven chaperones to buffer a destabilizing mutation depends on both the wild-type substrate's folding kinetics and how the mutation alters the kinetic parameters (Fig. 5). For instance, chaperones may be more effective in buffering mutations that slows down the transition from the misfold-tending conformation (U) to the native-tending conformation (F)—*i.e.* decreasing $k_{U\to F}$ by *e.g.* stabilizing the U conformation—than mutations that destabilize the native state by decreasing K_N. Such differential buffering may play a role in selecting tolerated genetic variations and shaping their consequences in human disease [71].

IV. DISCUSSION

Breaking the symmetry conditions 1– 4 is necessary but on its own insufficient for non-equilibrium folding. $g > 1$ often requires both substantial deviation from the symmetry conditions and other enabling kinetic conditions, as exemplified by Eqn. 53. Detailed mechanistic models [28, 29] are needed to quantitatively predict the extent of non-equilibrium folding. Nonetheless, these symmetry conditions can help assess whether a proposed mechanism of chaperone function will imply non-equilibrium folding.

Given both the theoretical models and the experimental evidence suggesting that ATP-driven chaperones can stabilize the native or active structures of substrate proteins out of the thermal equilibrium, Anfinsen's hypothesis does not *need* to be true for protein folding in cells. ATP-driven chaperones may not only kinetically accelerate the folding of proteins to thermodynamically stable native structures, but also actively fold some proteins to native structures that are thermodynamically unstable.

Most proteins have evolved to be marginally stable [72, 73]. If ATP-driven chaperones can indeed buffer destabilizing mutations and maintain the native structures and functions of unstable mutants, as discussed in Sec. III B 3, it is then plausible that the native structures of some proteins may have become thermodyanmically unstable as a consequence of this chaperone-buffered evolution. They may not stay folded on their own, but depend on the energy-consuming chaperones to maintain their native structures.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

How many proteins in a cell take exception to Anfinsen's hypothesis and depend on non-equilibrium folding by a particular ATP-driven chaperone? Emerging proteomics techniques may help answer this question. For example, cell lysates may be subject to proteolytic digestion [74] and the resulting products analyzed by mass spectrometry (MS), identifying proteins with permissible digestion sites, which approximately reflect their state of folding [75]. This proteolysis-MS assay may be repeated for lysates incubated with chaperone inhibitors [76, 77] or chaperone agonists [78]. Proteins more susceptible to proteolysis in the presence of chaperone inhibitors—and less susceptible in the presence of chaperone agonists—are candidates that may depend on the chaperone for non-equilibrium folding to their native structures. The lysates should be incubated in the presence of protein synthesis inhibitors so that the analysis can isolate the chaperone's effects on *maintaining* the native structures of its substrates from its effects on the folding of their nascent chains; the former demonstrates non-equilibrium stabilization of thermodynamically unfavorable native structures while the latter may be attributable to kinetic acceleration of protein folding. This analysis may be more applicable to GroEL/ES and Hsp70 than to Hsp90, because the latter mediates the late-stage folding and activation of its substrates [31, 79], which may not be associated with significant changes in the protein disorder detectable by the proteolysis-MS assay.

A. Implications for protein native structures and their folding pathways

My model of non-equilibrium protein folding and activation suggests the tantalizing possibility that ATP-driven chaperones may play a role in shaping the native structures of some proteins. Consistent with a previous experimental demonstration that chaperones alter the folding pathway of a substrate protein [80], my model implies that an ATP-driven chaperone may bias a substrate protein to fold along pathways that expose few cochaperone binding sites during folding, with consequences on the resulting structures.

Consider two conformations M and N of a substrate protein, where M is the free energy minimum but associated with a folding pathway inhibited by the chaperone, and N has a higher free energy but can be reached through a folding pathway uninhibited by the chaperone (Fig. 6). The protein will fold into its free energy minimum conformation M in the absence of the chaperone, but if the chaperone-induced pathway bias is sufficiently strong, it will fold into the alternative conformation N in the presence of the chaperone. Note that M can be an ensemble of rapidly inter-converting conformations, such as in intrinsically disordered proteins (IDP) or intrinsically disordered protein regions [81–85]. Can ATP-driven chaperones fold some IDPs into well-ordered structures?

It has been proposed that some proteins may fold into native structures that are more kinetically accessible than conformations of the lowest free energy [86]. Indeed, experimental observations have been reported that synonymous codon substitutions result in conformational changes in the translated proteins, due to kinetic changes in the co-translational folding of the nascent chain on the ribosome [87–90]. These results are consistent with the idea that the native structures of some proteins may be determined by kinetics rather than thermodynamics. One implication is that the solution to the structure prediction problem for such proteins in cell may depend on the solution to the protein folding problem, and in-cell protein folding may be an active, energy-dependent process [91]. Predicting the cellular conformation of these proteins—in the presence of ATP-driven chaperones—may require the search for folding pathways that limit the exposures of cochaperone-binding, *e.g.* hydrophobic, sites.

V. CONCLUSIONS

In this work, I have proposed a theoretical framework to analyze non-equilibrium protein folding by ATP-driven chaperones. The symmetry breaking conditions may help determine whether a chaperone—by a proposed mechanism of action—can convert the energy from ATP hydrolysis to the out-of-equilibrium stabilization of the native structures of its substrate proteins. I have discussed how Hsp70 and Hsp90 may have broken different symmetries and how the symmetry breaking enables them to perform non-equilibrium protein folding and activation. My models predict that some proteins may fold to native structures that do not correspond to the free energy minima, and that their native structures may be shaped by the chaperone-mediated folding pathways. These predictions may be tested by experiments, some of which I have suggested above.

[1] C. B. Anfinsen, Principles that Govern the Folding of Protein Chains, Science **181**, 223 (1973).
[2] K. A. Dill, Dominant forces in protein folding, Biochemistry **29**, 7133 (1990).
[3] B. Bukau, J. Weissman, and A. Horwich, Molecular Chaperones and Protein Quality Control, Cell **125**, 443 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

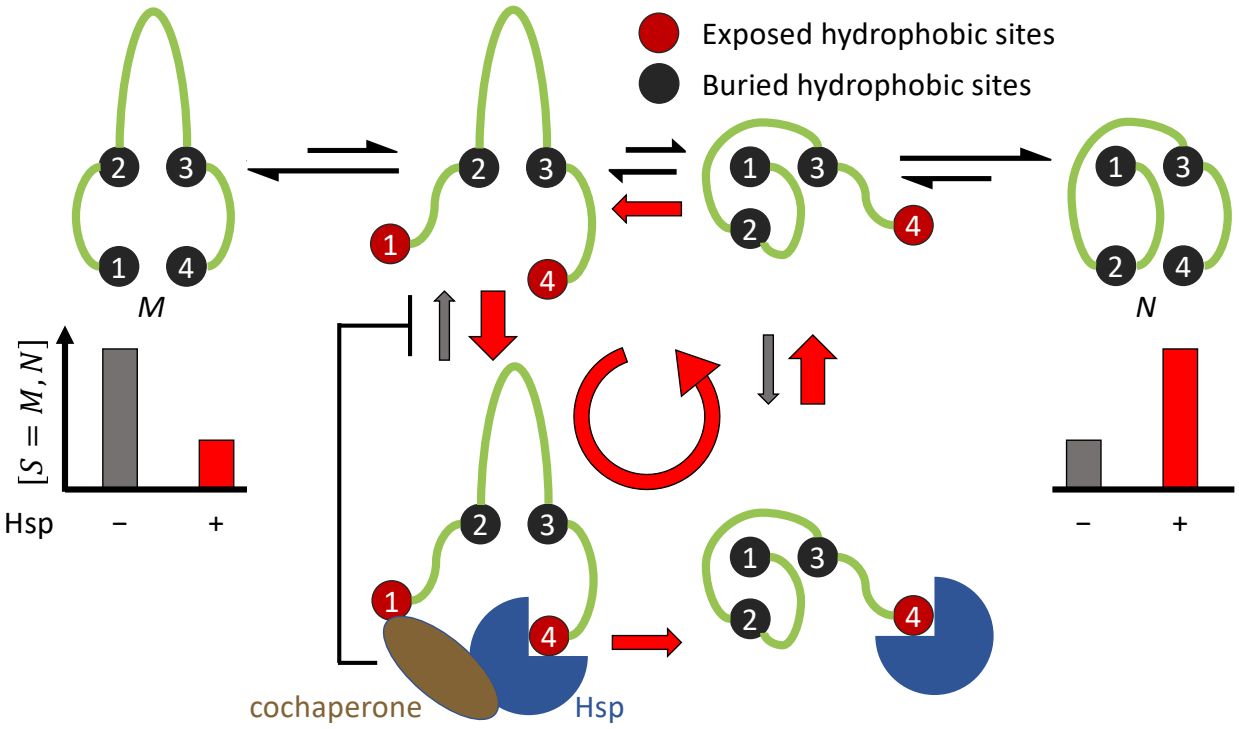

FIG. 6: An ATP-driven chaperone (Hsp) may favor the protein folding pathway that exposes few cochaperone binding sites and drive the protein to a different conformation (N) than its most stable conformation (M) in the absence of the chaperone.

[4] F. U. Hartl, A. Bracher, and M. Hayer-Hartl, Molecular chaperones in protein folding and proteostasis, Nature **475**, 324 (2011).

[5] D. Balchin, M. Hayer-Hartl, and U. F. Hartl, In vivo aspects of protein folding and quality control, Science **353**, aac4354 (2016).

[6] D. Thirumalai and G. H. Lorimer, CHAPERONIN-MEDIATED PROTEIN FOLDING, Annual Review of Biophysics and Biomolecular Structure **30**, 245 (2001).

[7] Z. Lin and H. S. Rye, GroEL-Mediated Protein Folding: Making the Impossible, Possible, Critical Reviews in Biochemistry and Molecular Biology **41**, 211 (2008).

[8] A. L. Horwich and W. A. Fenton, Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding, Quarterly Reviews of Biophysics **42**, 10.1017/S0033583509004764 (2009).

[9] M. P. Mayer, Hsp70 chaperone dynamics and molecular mechanism, Trends in Biochemical Sciences **38**, 507 (2013).

[10] M. P. Mayer and L. M. Gierasch, Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones, Journal of Biological Chemistry **294**, 2085 (2019).

[11] L. H. Pearl, Review: The HSP90 molecular chaperone—an enigmatic ATPase, Biopolymers **105**, 10.1002/bip.22835 (2016).

[12] M. Radli and S. G. Rüdiger, Dancing with the Diva: Hsp90-Client Interactions, Journal of Molecular Biology **430**, 10.1016/j.jmb.2018.05.026 (2018).

[13] F. H. Schopf, M. M. Biebl, and J. Buchner, The HSP90 chaperone machinery, Nature Reviews Molecular Cell Biology **18**, 345 (2017).

[14] G. Chiosis, C. A. Dickey, and J. L. Johnson, A global view of Hsp90 functions, Nature Structural & Molecular Biology **20**, 1 (2013).

[15] M. Taipale, D. F. Jarosz, and S. Lindquist, HSP90 at the hub of protein homeostasis: emerging mechanistic insights, Nature Reviews Molecular Cell Biology **11**, 515 (2010).

[16] N. A. Ranson, G. W. Farr, A. M. Roseman, B. Gowen, W. A. Fenton, A. L. Horwich, and H. R. Saibil, ATP-Bound States of GroEL Captured by Cryo-Electron Microscopy, Cell **107**, 869 (2001).

[17] K. A. Krukenberg, T. O. Street, L. A. Lavery, and D. A. Agard, Conformational dynamics of the molecular chaperone Hsp90, Quarterly Reviews of Biophysics **44**, 229 (2011).

[18] D. K. Clare, D. Vasishtan, S. Stagg, J. Quispe, G. W. Farr, M. Topf, A. L. Horwich, and H. R. Saibil, ATP-Triggered Conformational Changes Delineate Substrate-Binding and -Folding Mechanics of the GroEL Chaperonin, Cell **149**, 113

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(2012).

[19] R. Kityk, M. Vogel, R. Schlecht, B. Bukau, and M. P. Mayer, Pathways of allosteric regulation in Hsp70 chaperones, Nature Communications **6**, 8308 (2015).

[20] T. Alderson, J. Kim, and J. Markley, Dynamical Structures of Hsp70 and Hsp70-Hsp40 Complexes, Structure **24**, 1014 (2016).

[21] J. S. Weissman, Y. Kashi, W. A. Fenton, and A. L. Horwich, GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms, Cell **78**, 693 (1994).

[22] A. Szabo, T. Langer, H. Schröder, J. Flanagan, B. Bukau, and F. U. Hartl, The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE, Proc. Natl. Acad. Sci. **91**, 10.1073/pnas.91.22.10345 (1994), this shows that DnaK can be added subsequent to DnaJ and still fold the denatured luciferase, but incubation with DnaK without DnaJ leads to aggregation that cannot be rescued by subsequent addition of DnaJ. The order of addition of DnaK/DnaJ matters.

[23] S. M. Doyle, O. Genest, and S. Wickner, Protein rescue from aggregates by powerful molecular chaperone machines, Nature Reviews Molecular Cell Biology **14**, 617 (2013).

[24] R. Imamoglu, D. Balchin, M. Hayer-Hartl, and F. U. Hartl, Bacterial Hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein, Nature Communications **11**, 10.1038/s41467-019-14245-4 (2020).

[25] P. D. L. Rios and P. Goloubinoff, Hsp70 chaperones use ATP to remodel native protein oligomers and stable aggregates by entropic pulling, Nature Structural & Molecular Biology **23**, 766 (2016).

[26] S. Chakrabarti, C. Hyeon, X. Ye, G. H. Lorimer, and D. Thirumalai, Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium, Proceedings of the National Academy of Sciences **114**, E10919 (2017).

[27] P. Goloubinoff, A. S. Sassi, B. Fauvet, A. Barducci, and P. Rios, Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins, Nature Chemical Biology **14**, 10.1038/s41589-018-0013-8 (2018).

[28] H. Xu, Cochaperones enable Hsp70 to use ATP energy to stabilize native proteins out of the folding equilibrium, Scientific Reports **8**, 13213 (2018).

[29] H. Xu, ATP-driven Non-equilibrium Activation of Kinase Clients by the Molecular Chaperone Hsp90, Biophysical Journal 10.1016/j.bpj.2020.08.038 (2020).

[30] K. A. Verba and D. A. Agard, How Hsp90 and Cdc37 Lubricate Kinase Molecular Switches, Trends in Biochemical Sciences **42**, 799 (2017).

[31] E. Kirschke, D. Goswami, D. Southworth, P. Griffin, and D. Agard, Glucocorticoid Receptor Function Regulated by Coordinated Action of the Hsp90 and Hsp70 Chaperone Cycles, Cell **157**, 1685 (2014).

[32] M. J. Todd, G. H. Lorimer, and D. Thirumalai, Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism, Proceedings of the National Academy of Sciences **93**, 4030 (1996).

[33] B. Hu, M. P. Mayer, and M. Tomita, Modeling Hsp70-Mediated Protein Folding, Biophysical Journal **91**, 496 (2006).

[34] P. D. L. Rios and A. Barducci, Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption, eLife **3**, e02218 (2014).

[35] S. K. Sharma, P. Rios, P. Christen, A. Lustig, and P. Goloubinoff, The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase, Nature Chemical Biology **6**, 914 (2010).

[36] S. G. Burston, J. S. Weissman, G. W. Farr, W. A. Fenton, and A. L. Norwich, Release of both native and non-native proteins from a cis-only GroEL ternary complex, Nature **383**, 96 (1996).

[37] D. R. Astumian, Microscopic reversibility as the organizing principle of molecular machines, Nature Nanotechnology **7**, nnano.2012.188 (2012).

[38] A. Sekhar, A. Velyvis, G. Zoltsman, R. Rosenzweig, G. Bouvignies, and L. E. Kay, Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions, eLife **7**, e32764 (2018).

[39] S. J. Park, B. N. Borin, M. A. Martinez-Yamout, and H. J. Dyson, The client protein p53 adopts a molten globule–like state in the presence of Hsp90, Nature Structural & Molecular Biology **18**, 537 (2011).

[40] K. A. Verba, R. Y.-R. Wang, A. Arakawa, Y. Liu, M. Shirouzu, S. Yokoyama, and D. A. Agard, Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase, Science **352**, 1542 (2016).

[41] M. Hayer-Hartl, A. Bracher, and F. U. Hartl, The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding, Trends in Biochemical Sciences **41**, 62 (2016).

[42] A. J. Caplan, A. K. Mandal, and M. A. Theodoraki, Molecular chaperones and protein kinase quality control, Trends in Cell Biology **17**, 87 (2007).

[43] D. Keramisanou, A. Aboalroub, Z. Zhang, W. Liu, D. Marshall, A. Diviney, R. Larsen, R. Landgraf, and I. Gelis, Molecular Mechanism of Protein Kinase Recognition and Sorting by the Hsp90 Kinome-Specific Cochaperone Cdc37, Molecular Cell **62**, 260 (2016).

[44] S. Sreeramulu, H. R. Jonker, T. Langer, C. Richter, R. C. Lancaster, and H. Schwalbe, The Human Cdc37·Hsp90 Complex Studied by Heteronuclear NMR Spectroscopy, Journal of Biological Chemistry **284**, 3885 (2009).

[45] J. M. Eckl, M. J. Scherr, L. Freiburger, M. A. Daake, M. Sattler, and K. Richter, Hsp90·Cdc37 Complexes with Protein Kinases Form Cooperatively with Multiple Distinct Interaction Sites, Journal of Biological Chemistry **290**, 30843 (2015).

[46] M. P. Mayer, H. Schröder, S. Rüdiger, K. Paal, T. Laufen, and B. Bukau, Multistep mechanism of substrate binding determines chaperone activity of Hsp70, Nature Structural & Molecular Biology **7**, 586 (2000), kinetics measured for various DnaK mutants.

[47] J. Yang, M. Nune, Y. Zong, L. Zhou, and Q. Liu, Close and Allosteric Opening of the Polypeptide-Binding Site in a Human Hsp70 Chaperone BiP, Structure **23**, 2191 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[48] T. Laufen, M. P. Mayer, C. Beisel, D. Klostermeier, A. Mogk, J. Reinstein, and B. Bukau, Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones, Proceedings of the National Academy of Sciences **96**, 10.1073/pnas.96.10.5452 (1999).

[49] P. Wittung-Stafshede, J. Guidry, B. E. Horne, and S. J. Landry, The J-Domain of Hsp40 Couples ATP Hydrolysis to Substrate Capture in Hsp70†, Biochemistry **42**, 10.1021/bi027333o (2003), j domain binds to DnaK ATPase domain with KD=23uM. This is a constraints to be used in the model.

[50] S. Rüdiger, J. Schneider-Mergener, and B. Bukau, Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone, The EMBO Journal **20**, 1042 (2001).

[51] J. Perales-Calvo, A. Muga, and F. Moro, Role of DnaJ G/F-rich Domain in Conformational Recognition and Binding of Protein Substrates, Journal of Biological Chemistry **285**, 34231 (2010).

[52] R. Herbst, U. Schäfer, and R. Seckler, Equilibrium Intermediates in the Reversible Unfolding of Firefly (Photinus pyralis) Luciferase, Journal of Biological Chemistry **272**, 7099 (1997).

[53] D. Elnatan, M. Betegon, Y. Liu, T. Ramelot, M. A. Kennedy, and D. A. Agard, Symmetry broken and rebroken during the ATP hydrolysis cycle of the mitochondrial Hsp90 TRAP1, eLife **6**, e25235 (2017).

[54] T. Lopez, K. Dalton, and J. Frydman, The Mechanism and Function of Group II Chaperonins, Journal of Molecular Biology **427**, 2919 (2015).

[55] D. Gestaut, A. Limatola, L. Joachimiak, and J. Frydman, The ATP-powered gymnastics of TRiC/CCT: an asymmetric protein folding machine with a symmetric origin story, Current Opinion in Structural Biology **55**, 50 (2019).

[56] K. Chakraborty, M. Chatila, J. Sinha, Q. Shi, B. C. Poschner, M. Sikor, G. Jiang, D. C. Lamb, F. U. Hartl, and M. Hayer-Hartl, Chaperonin-Catalyzed Rescue of Kinetically Trapped States in Protein Folding, Cell **142**, 112 (2010).

[57] Y.-C. Tang, H.-C. Chang, A. Roeben, D. Wischnewski, N. Wischnewski, M. J. Kerner, F. U. Hartl, and M. Hayer-Hartl, Structural Features of the GroEL-GroES Nano-Cage Required for Rapid Folding of Encapsulated Protein, Cell **125**, 903 (2006).

[58] A. J. Gupta, S. Haldar, G. Miličić, F. U. Hartl, and M. Hayer-Hartl, Active Cage Mechanism of Chaperonin-Assisted Protein Folding Demonstrated at Single-Molecule Level, Journal of Molecular Biology **426**, 2739 (2014).

[59] A. Baumketner, A. Jewett, and J. Shea, Effects of Confinement in Chaperonin Assisted Protein Folding: Rate Enhancement by Decreasing the Roughness of the Folding Energy Landscape, Journal of Molecular Biology **332**, 701 (2003).

[60] J. L. England, D. Lucent, and V. S. Pande, A Role for Confined Water in Chaperonin Function, Journal of the American Chemical Society **130**, 11838 (2008).

[61] H. Hofmann, F. Hillger, S. H. Pfeil, A. Hoffmann, D. Streich, D. Haenni, D. Nettels, E. A. Lipman, and B. Schuler, Single-molecule spectroscopy of protein folding in a chaperonin cage, Proceedings of the National Academy of Sciences **107**, 11793 (2010).

[62] S. Rüdiger, S. M. Freund, D. B. Veprintsev, and A. R. Fersht, CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90, Proceedings of the National Academy of Sciences **99**, 11085 (2002).

[63] D. A. Drummond and C. O. Wilke, Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution, Cell **134**, 341 (2008).

[64] A. Cagan, A. Baez-Ortega, N. Brzozowska, F. Abascal, T. H. H. Coorens, M. A. Sanders, A. R. J. Lawson, L. M. R. Harvey, S. Bhosle, D. Jones, R. E. Alcantara, T. M. Butler, Y. Hooks, K. Roberts, E. Anderson, S. Lunn, E. Flach, S. Spiro, I. Januszczak, E. Wrigglesworth, H. Jenkins, T. Dallas, N. Masters, M. W. Perkins, R. Deaville, M. Druce, R. Bogeska, M. D. Milsom, B. Neumann, F. Gorman, F. Constantino-Casas, L. Peachey, D. Bochynska, E. S. J. Smith, M. Gerstung, P. J. Campbell, E. P. Murchison, M. R. Stratton, and I. Martincorena, Somatic mutation rates scale with lifespan across mammals, Nature **604**, 517 (2022).

[65] H. H. Guo, J. Choe, and L. A. Loeb, Protein tolerance to random amino acid change, Proceedings of the National Academy of Sciences **101**, 9205 (2004), https://www.pnas.org/doi/pdf/10.1073/pnas.0403255101.

[66] P. Markiewicz, L. Kleina, C. Cruz, S. Ehret, and J. Miller, Genetic studies of the lac repressor. xiv. analysis of 4000 altered escherichia coli lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence, Journal of Molecular Biology **240**, 421 (1994).

[67] A. A. Pakula and R. T. Sauer, Genetic analysis of protein stability and function, Annual Review of Genetics **23**, 289 (1989), pMID: 2694933, https://doi.org/10.1146/annurev.ge.23.120189.001445.

[68] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, Protein stability promotes evolvability, Proceedings of the National Academy of Sciences **103**, 5869 (2006), https://www.pnas.org/doi/pdf/10.1073/pnas.0510098103.

[69] C. Queitsch, T. A. Sangster, and S. Lindquist, Hsp90 as a capacitor of phenotypic variation, Nature **417**, 618 (2002).

[70] J. Aguilar-Rodríguez, B. Sabater-Muñoz, R. Montagud-Martínez, V. Berlanga, D. Alvarez-Ponce, A. Wagner, and M. A. Fares, The Molecular Chaperone DnaK Is a Source of Mutational Robustness, Genome Biology and Evolution **8**, 2979 (2016), https://academic.oup.com/gbe/article-pdf/8/9/2979/20593410/evw176.pdf.

[71] G. I. Karras, S. Yi, N. Sahni, M. Fischer, J. Xie, M. Vidal, A. D. D'Andrea, L. Whitesell, and S. Lindquist, HSP90 Shapes the Consequences of Human Genetic Variation, Cell **168**, 856 (2017).

[72] D. M. Taverna and R. A. Goldstein, Why are proteins marginally stable?, Proteins: Structure, Function, and Bioinformatics **46**, 105 (2002).

[73] K. Ghosh and K. Dill, Cellular Proteomes Have Broad Distributions of Protein Stability, Biophysical Journal **99**, 3996 (2010).

[74] C. Park and S. Marqusee, Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding, Nature Methods **2**, 207 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.21.492947; this version posted May 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[75] P. To, B. Whitehead, H. E. Tarbox, and S. D. Fried, Nonrefoldability is Pervasive Across the E. coli Proteome, Journal of the American Chemical Society **143**, 11435 (2021).

[76] J. I.-J. Leu, P. Zhang, M. E. Murphy, R. Marmorstein, and D. L. George, Structural Basis for the Inhibition of HSP70 and DnaK Chaperones by Small-Molecule Targeting of a C-Terminal Allosteric Pocket, ACS Chemical Biology **9**, 2508 (2014).

[77] W. Wen, W. Liu, Y. Shao, and L. Chen, VER-155008, a small molecule inhibitor of HSP70 with potent anti-cancer activity on lung cancer cell lines, Experimental Biology and Medicine **239**, 638 (2014).

[78] S. Wisén, E. B. Bertelsen, A. D. Thompson, S. Patury, P. Ung, L. Chang, C. G. Evans, G. M. Walter, P. Wipf, H. A. Carlson, J. L. Brodsky, E. R. Zuiderweg, and J. E. Gestwicki, Binding of a Small Molecule at a Protein–Protein Interface Regulates the Chaperone Activity of Hsp70–Hsp40, ACS Chemical Biology **5**, 611 (2010), the small molecule increases ATP hydrolysis turnover rate and thus increases the refolding efficiency.

[79] U. Jakob, H. Lilie, I. Meyer, and J. Buchner, Transient Interaction of Hsp90 with Early Unfolding Intermediates of Citrate Synthase IMPLICATIONS FOR HEAT SHOCK IN VIVO, Journal of Biological Chemistry **270**, 7288 (1995).

[80] A. Sekhar, R. Rosenzweig, G. Bouvignies, and L. E. Kay, Hsp70 biases the folding pathways of client proteins, Proceedings of the National Academy of Sciences **113**, E2794 (2016).

[81] C. K. Fisher and C. M. Stultz, Constructing ensembles for intrinsically disordered proteins, Current Opinion in Structural Biology **21**, 426 (2011).

[82] P. Tompa, Unstructural biology coming of age, Current Opinion in Structural Biology **21**, 419 (2011).

[83] C. J. Oldfield and A. K. Dunker, Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions, Annual Review of Biochemistry **83**, 553 (2014).

[84] S.-H. Chong, P. Chatterjee, and S. Ham, Computer Simulations of Intrinsically Disordered Proteins, Annual Review of Physical Chemistry **68**, 1 (2016).

[85] R. B. Best, Computational and theoretical advances in studies of intrinsically disordered proteins, Current Opinion in Structural Biology **42**, 147 (2017).

[86] D. Baker and D. A. Agard, Kinetics versus Thermodynamics in Protein Folding, Biochemistry **33**, 7505 (1994).

[87] C. Kimchi-Sarfaty, J. M. Oh, I.-W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar, and M. M. Gottesman, A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity, Science **315**, 525 (2007).

[88] S. J. Kim, J. S. Yoon, H. Shishido, Z. Yang, L. A. Rooney, J. M. Barral, and W. R. Skach, Translational tuning optimizes nascent protein folding in cells, Science **348**, 444 (2015).

[89] I. M. Walsh, M. A. Bowman, I. F. S. Santarriaga, A. Rodriguez, and P. L. Clark, Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness, Proceedings of the National Academy of Sciences **117**, 3528 (2020).

[90] A. A. Rosenberg, A. Marx, and A. M. Bronstein, Codon-specific Ramachandran plots show amino acid backbone conformation depends on identity of the translated codon, Nature Communications **13**, 2815 (2022).

[91] I. Sorokina, A. R. Mushegian, and E. V. Koonin, Is Protein Folding a Thermodynamically Unfavorable, Active, Energy-Dependent Process?, International Journal of Molecular Sciences **23**, 521 (2022).