

# SCIENTIFIC REPORTS



OPEN

## Cochaperones enable Hsp70 to use ATP energy to stabilize native proteins out of the folding equilibrium

Huafeng Xu 

The heat shock protein 70 (Hsp70) chaperones, vital to the proper folding of proteins inside cells, consume ATP and require cochaperones in assisting protein folding. It is unclear whether Hsp70 can utilize the free energy from ATP hydrolysis to fold a protein into a native state that is thermodynamically unstable in the chaperone-free equilibrium. Here I present a model of Hsp70-mediated protein folding, which predicts that Hsp70, as a result of differential stimulation of ATP hydrolysis by its Hsp40 cochaperone, dissociates faster from a substrate in fold-competent conformations than from one in misfolding-prone conformations, thus elevating the native concentration above and suppressing the misfolded concentration below their respective equilibrium values. Previous models would not make or imply these predictions, which are experimentally testable. My model quantitatively reproduces experimental refolding kinetics, predicts how modulations of the Hsp70/Hsp40 chaperone system affect protein folding, and suggests new approaches to regulating cellular protein quality.

The discovery of chaperones and their roles in assisting protein folding amended the long-held view that proteins spontaneously fold into their native structures<sup>1–3</sup>. Large, multi-domain proteins may take many hours to fold, or fail to fold properly altogether on their own<sup>2,4</sup>. ATP-consuming chaperones—including Hsp70s—provide critical assistance in the *in vivo* folding and the biological functions of broad sets of substrate proteins<sup>3</sup>. Extensive experimental studies have firmly established that within the same length of time, more denatured substrate proteins refold in the presence of the Hsp70 chaperones than in their absence<sup>5,6</sup>. Despite tremendous progress in the mechanistic studies of the Hsp70 chaperones<sup>2,7,8</sup>, including the development of theoretical models<sup>9–12</sup>, it remains unclear why ATP consumption is indispensable to these chaperones; many enzymes catalyze chemical reactions without consuming free energy. Recently it was demonstrated that chaperones such as GroEL and Hsp70 depend on continuous ATP hydrolysis to maintain a protein in a native state that is thermodynamically unstable<sup>13</sup>, but it is unknown how Hsp70 can utilize the ATP free energy to alter the folding equilibrium. In addition, Hsp70s require Hsp40—also known as J proteins<sup>14</sup>—cochaperones in assisting protein folding. It is yet unexplained why cochaperones are absolutely necessary.

The Hsp70 chaperones, such as the bacterial DnaK, consist of an N-terminal nucleotide binding domain (NBD) and a C-terminal substrate binding domain (SBD). The Hsp70 SBD adopts an open conformation when its NBD is ATP-bound (I call the Hsp70 to be in the ATP-state), which allows the substrate to bind and unbind at high rates, whereas when the NBD is ADP-bound (ADP-state), the SBD changes to a closed conformation, rendering both binding and unbinding orders-of-magnitude slower<sup>6,15,16</sup>. Hsp70s have low basal ATP hydrolysis activities<sup>17</sup>. The Hsp40 cochaperones, such as the bacterial DnaJ, can drastically stimulate the ATPase activity of Hsp70 using their N-terminal J domain (JD)<sup>18,19</sup>, shared by all Hsp40s (hence the name J proteins)<sup>14</sup>. Hsp40s also have a C-terminal domain (CTD) that can bind to denatured proteins<sup>20,21</sup>. Both Hsp70 and Hsp40 recognize exposed hydrophobic sites<sup>7,22,23</sup>. As a result, they can distinguish different protein conformations using the corresponding difference in the exposed hydrophobic sites. For example, Hsp70 has been shown to bind to both unfolded and partially folded, near native protein structures, but not to native structures<sup>24,25</sup>. Hsp40 and Hsp70 may simultaneously bind to different segments of the same substrate molecule, and the consequent spatial

Unaffiliated, Forest Hills, NY, 11375, USA. Correspondence and requests for materials should be addressed to H.X. (email: [huafeng@gmail.com](mailto:huafeng@gmail.com))

proximity then facilitates the J domain binding to Hsp70 and accelerating its ATP hydrolysis<sup>26–28</sup>. Following ATP hydrolysis, the chaperone returns from the ADP-state to the ATP-state through nucleotide exchange, which is often catalyzed by nucleotide exchange factors (NEF) such as the bacterial GrpE<sup>29,30</sup>.

It is unclear whether Hsp70 can use the free energy from ATP hydrolysis to drive its substrate protein toward the native state,  $N$ , and away from the misfolded state,  $M$ , such that  $f_N/f_M > f_{N,eq}/f_{M,eq}$ , where  $f_S$  is the fraction of the substrate in state  $S$  at the steady state of Hsp70-mediated folding, and  $f_{S,eq}$  is the corresponding fraction at the folding equilibrium in the absence of the chaperone. Previous models<sup>10,31</sup> mostly considered the chaperone as an unfoldase/holdase—which need not consume free energy—that pulls the substrate out of the misfolded state and holds it in an unfolded state. It was proposed that the free energy from ATP hydrolysis was used to achieve ultra-affinity in substrate binding<sup>9,32</sup>. As an unfoldase/holdase, Hsp70 would also pull the substrate out of the native state into the unfolded state; unless Hsp70 has a higher affinity for the native substrate than for the misfolded substrate, which contradicts experimental observations, these models would predict  $f_N/f_M \leq f_{N,eq}/f_{M,eq}$ .

Here I propose a model of Hsp70-mediated protein folding, in which Hsp70 and Hsp40 together constitute a molecular machine that uses the free energy from ATP hydrolysis to actively drive a protein toward its native state, so that  $f_N/f_M > f_{N,eq}/f_{M,eq}$ . It suggests that without Hsp40, Hsp70 alone cannot change the ratio  $f_N/f_M$  from the equilibrium value  $f_{N,eq}/f_{M,eq}$ . My model thus answers the question why Hsp70 requires both the Hsp40 cochaperones and ATP consumption in assisting protein folding. My model explains the puzzling non-monotonic dependency of folding efficiency on the chaperone and cochaperone concentrations. It makes quantitative predictions on how protein folding is affected by modulations of the chaperone environment, including changes in the ATPase activity or the nucleotide exchange rate of Hsp70. These predictions may be readily tested by experiments, and inform rational approaches to manipulating chaperone-mediated protein folding.

## Results

My model is based on two assumptions supported by experimental observations. The first assumption is that a substrate protein can adopt two additional conformational states besides the misfolded,  $M$ , and the native,  $N$ , states: the unfolded and misfolding-prone state,  $U$ , and the fold-competent state,  $F$ . A protein in the  $F$  state is unfolded but poised to fold into the native state (Fig. 1a,b). Such intermediate states of folding have been observed experimentally<sup>4</sup>. Conformational transitions can occur between  $M$  and  $U$ , between  $U$  and  $F$ , and between  $F$  and  $N$  (Fig. 1b). The second assumption is that Hsp40's affinity for a substrate protein is higher if the substrate is in the misfolding-prone conformation than if it is in the fold-competent conformation. Experimental observations suggest that the misfolding-prone conformation is less compact and exposes more hydrophobic sites—thus providing more accessible sites for Hsp40 binding—than the fold-competent conformation<sup>4</sup>. Consequently, a substrate in the  $U$  state is more likely to be Hsp40-bound than one in the  $F$  state.

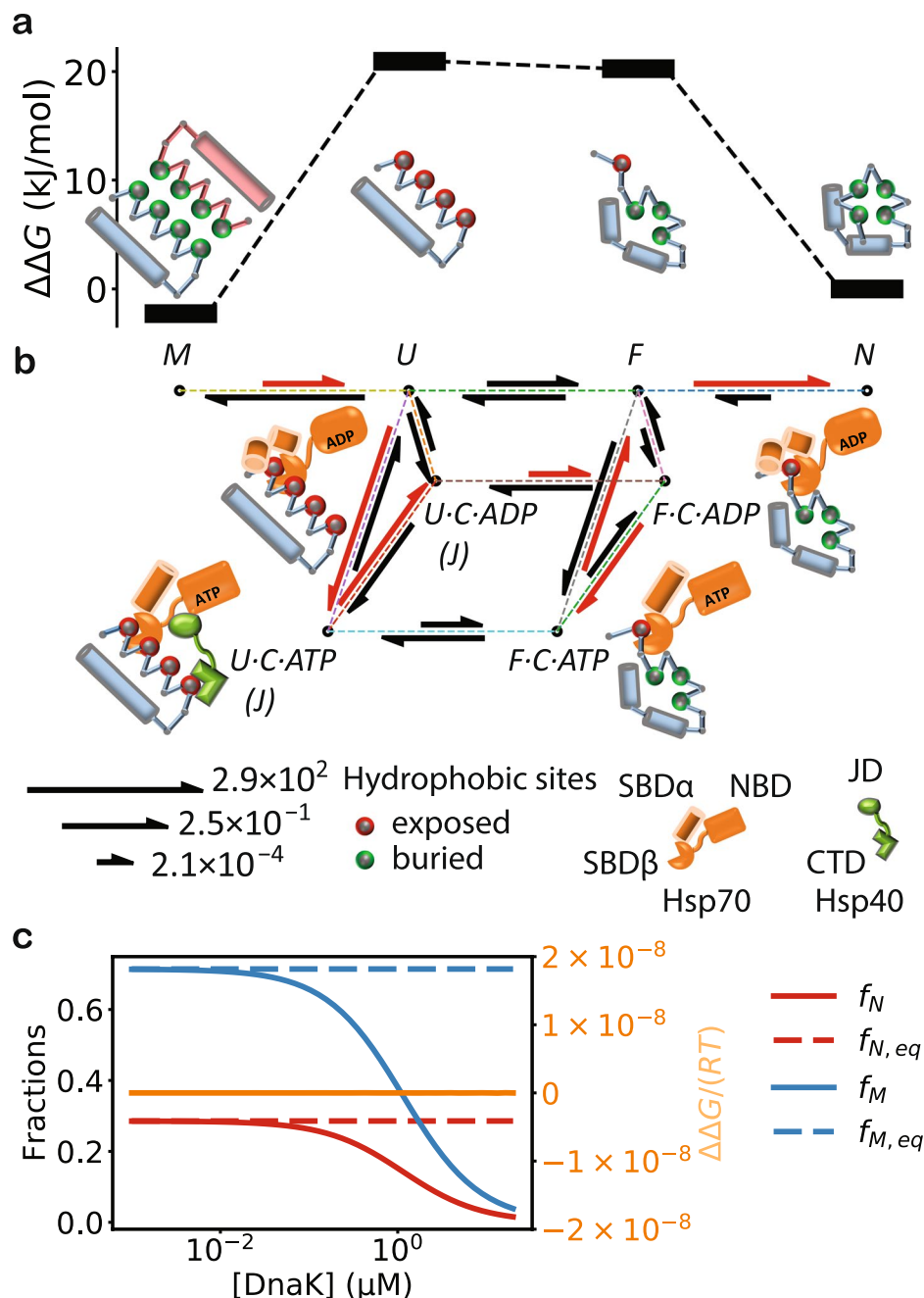
The key idea of my model is a mechanism by which Hsp70 actively drives a substrate toward the native state and away from the misfolded state. Based on the above assumptions, an Hsp70 molecule bound to a substrate molecule in the  $U$  state will on average have substantially higher ATP hydrolysis rate—because of the higher probability of *cis* stimulation by an Hsp40 molecule bound to the same substrate molecule—than if it is bound to a substrate molecule in the  $F$  state. If the nucleotide exchange rate is between these two hydrolysis rates, an Hsp70 bound to a substrate in the  $U$  state will be driven toward the ADP-state, where it slowly dissociates from the substrate, while an Hsp70 bound to a substrate in the  $F$  state will be driven toward the ATP-state, where it rapidly dissociates from the substrate. This enables Hsp70, when bound to substrate molecules, to act like a Maxwell's demon<sup>33</sup>; it quickly releases the fold-competent molecules so that they can fold, but it retains the misfolding-prone molecules to prevent misfolding. By this mechanism, Hsp70 drives the folding along the reaction path of  $M \rightarrow U \rightarrow U \cdot C \cdot ATP \rightarrow U \cdot C \cdot ADP \rightarrow F \cdot C \cdot ADP \rightarrow F \cdot C \cdot ATP \rightarrow F \rightarrow N$ , where  $S \cdot C \cdot X$  represents the complex between a substrate in conformation  $S$  and the chaperone  $C$  bound to nucleotide  $X = ATP, ADP$  (Fig. 1b). One ATP molecule is consumed in this reaction path and the free energy is used to compel the substrate into the native state.

My model predicts that Hsp70 need Hsp40 in order to alter the folding equilibrium. For Hsp70 to drive substrate folding toward the native state, the above mechanism does not require that more Hsp70 molecules bind to a substrate in the  $U$  state than to a substrate in the  $F$  state, which is true and reflected in previous models; instead, it requires that an individual Hsp70 molecule, when bound to a substrate, dissociates slower if the substrate is in the  $U$  state than if it is in the  $F$  state. Hsp40 thus plays a critical role because a substrate-bound Hsp70 distinguishes, probabilistically, between the  $U$  and  $F$  states of the substrate by sensing whether an Hsp40 is also bound to the same substrate. Defining the excess free energy of folding as

$$\Delta\Delta G \equiv RT(\ln(f_N/f_M) - \ln(f_{N,eq}/f_{M,eq})),$$

where  $R$  is the gas constant and  $T$  the temperature, it can be shown algebraically (see *Methods*) and numerically (Fig. 1c) that without cochaperones,  $\Delta\Delta G = 0$ . This prediction is consistent with the results from the single-molecule experiment of DnaK-mediated refolding<sup>34</sup>, where DnaK alone in the presence of ATP was unable to alter the ratio of the misfolded and folded fractions.

In order to simplify calculations using my model, I assume that a protein's hydrophobic binding sites for Hsp40 and Hsp70—which can be exposed in the  $U$  and  $F$  states—become entirely buried upon folding ( $F \rightarrow N$ ) and misfolding ( $U \rightarrow M$ ), thus a protein in the  $M$  and  $N$  states has zero exposed hydrophobic binding sites, and it does not bind to either Hsp70 or Hsp40 (Fig. 1a). The hydrophobic burial in the misfolded state may be due to intramolecular contacts between incorrectly folded domains within a monomeric protein, or due to intermolecular contacts between different protein molecules as a result of oligomerization or mild, reversible aggregation. A protein may also aggregate irreversibly, and will not refold even in the presence of the chaperone system<sup>5</sup>. A protein in such an irreversibly aggregated state may have a prohibitive kinetic barrier to return to the  $M$  or  $U$



**Figure 1.** My model of Hsp70/Hsp40/NEF-mediated protein folding. (a) Conformational states in protein folding, and their relative free energies in the absence of chaperones. I assume that there are four conformational states: Misfolded, Unfolded and misfolding-prone, Fold-competent, and Native. The pink and blue chains in the  $M$  state may correspond to different molecules (reversibly aggregated) or to different domains in the same molecule. The high free energy barriers associated with the intermediate states  $U$  and  $F$  slow down the refolding of misfolded proteins. The exemplary free energies are derived from the kinetic parameters fit to the refolding experiments of luciferase at 25 °C (Table 2). (b) The transitions between the microscopic states in the chaperone-mediated folding pathway.  $S \cdot C \cdot X$  represents the complex between the substrate in the conformational state  $S$  ( $=U, F$ ) and the Hsp70 chaperone (denoted as  $C$ ) bound to nucleotide  $X$  ( $=ATP, ADP$ ). The transitions between  $S$  and  $S \cdot C \cdot X$  correspond to the chaperone binding to and unbinding from the substrate. The transition of  $S \cdot C \cdot ATP$  to  $S \cdot C \cdot ADP$  corresponds to ATP hydrolysis, and its reverse, nucleotide exchange. Hsp70 binding stabilizes the substrate in the intermediate states, thus catalyzing the folding reaction. Hsp40 ( $J$ ) can form a ternary complex with the substrate and Hsp70—thus stimulating ATP hydrolysis—if the substrate is in the  $U$  state, but not if the substrate is in the  $F$  state. Differential ATP hydrolysis by Hsp70 bound to the substrate in the  $U$  and  $F$  states drives the refolding through the pathway highlighted in red. The lengths of the reaction arrows are linear with respect to the logarithms of the exemplary rate constants (in 1/s) for the DnaK/DnaJ/GrpE-mediated refolding of luciferase at 25 °C (Tables 1 and 2). (c) Without cochaperones, Hsp70 cannot alter the balance between folding and misfolding. DnaK binding to the intermediate states decreases both the native

(red) and misfolded (blue) populations, but the ratio between the two remains unchanged from its equilibrium value:  $\Delta\Delta G = 0$  (orange, right y-axis). Here, I have taken  $k_{a,C-ATP}^{(U)}/k_{a,C-ATP}^{(F)} = k_{a,C-ADP}^{(U)}/k_{a,C-ADP}^{(F)} = 100$  to show that differential binding of Hsp70 to the substrate in different conformational states does not alter the folding/misfolding balance.

states. It is thus unamenable to Hsp70-assisted refolding and not included in my model. The quantitative details of my model are given in *Methods*.

I applied my model to the analysis of DnaK/DnaJ/GrpE-mediated refolding of luciferase<sup>5</sup>. Most of the relevant kinetic parameters at 25 °C for this bacterial Hsp70 system have been carefully determined experimentally<sup>35</sup> (Table 1). My model quantitatively reproduces the experimentally observed refolding kinetics under various conditions, capturing the slow spontaneous refolding and denaturation of luciferase, the acceleration of refolding with chaperone assistance, and the necessity of GrpE (Fig. 2a). The intermediate conformations  $U$  and  $F$  in my model in the case of luciferase may correspond to the experimentally identified intermediate conformations  $I_2$  and  $I_1$  of luciferase<sup>4</sup>: the free energy difference between  $N$  and  $F$  at 25 °C, according to the fitted parameters, is 20 kJ/mol, close to the experimental value of 15 kJ/mol between  $N$  and  $I_1$ , measured at 10 °C. Consistent with previous experimental observations<sup>31</sup>, my model suggests that the Hsp70-mediated refolding proceeds in two steps: (1) rapid unfolding of the misfolded substrate, stabilized by the ADP-bound DnaK, followed by (2) slow conversion to the native state (Fig. 3a).

Next, I used my model to analyze how the refolding kinetics and yield change with respect to the DnaK concentration, which have been experimentally studied for LucDHis6, a variant of luciferase<sup>31</sup>. The refolding data fits well with the kinetic equation of a first-order reaction,  $N(t) = N(\infty)(1 - \exp(-kt))$ , and the rate constant  $k$  is approximately unchanged across different DnaK concentrations<sup>8,36</sup>. DnaK thus affects the refolding kinetics mostly by altering the refolding yield  $N(\infty)$ . My model captures the refolding kinetics and the refolding yield of LucDHis6 in the experimental range of DnaK concentrations (Fig. 2c); although the experiments were performed at the temperature of 22 °C, the predictions of my model using the kinetic parameters derived for 25 °C are nevertheless in quantitative agreement with the experimental data. At the steady state, the reactive flux along the ATP-driven cycle  $U \rightarrow U \cdot C \cdot ATP \rightarrow U \cdot C \cdot ADP \rightarrow F \cdot C \cdot ADP \rightarrow F \cdot C \cdot ATP \rightarrow F (\rightarrow U)$  (Fig. 3b) keeps the protein folding out of equilibrium, elevating the native population above and suppressing the misfolded population below their respective equilibrium values (Fig. 2c). Notably, the refolding yield peaks around  $[DnaK] = 1 \mu M$ , and it decreases at higher DnaK concentrations. This non-monotonic dependence on the DnaK concentration was also reported for the wildtype luciferase while this work was under review<sup>36</sup>. According to my model, the excess free energy at the steady state always increases with increasing DnaK concentrations, but the native population reaches a maximum and then decreases (Fig. 2c), because at high DnaK concentrations, the substrate is trapped in the DnaK-bound state and thus prevented from folding into the native state.

I used my model to estimate the ATP consumption in the DnaK/DnaJ/GrpE-mediated folding, which has been experimentally measured for LucDHis6<sup>31</sup> (Fig. 4). My model estimates that, in the initial minutes of refolding, approximately 150 ATP molecules are consumed to refold one LucDHis6 (Fig. 4a), which is reasonably close to the experimental result of ~50 ATP molecules consumed per refolded LucDHis6 when the stoichiometry of DnaK:LucDHis6 is 1:1, significantly higher than the experimental number of ~5 when LucDHis6 is in excess of DnaK, and significantly lower than the estimates of >1000 for many other substrates in other experiments<sup>31,37-39</sup>. The discrepancy between the model and the experimental results may be attributable to the approximations in my model and the inaccuracies in the input kinetic parameters. ATP hydrolysis continues at the steady state and the free energy is utilized to promote the native state and suppress the misfolded state (Fig. 4b,c). As  $[DnaK]$  exceeds 1  $\mu M$ , the ATP consumption rate increases rapidly without commensurate increase in the excess free energy. My analysis thus suggests that DnaK may be most free energy efficient at maintaining protein folding out-of-equilibrium when its concentration is in the sub-micromolar range, a prediction that may be tested experimentally.

My model suggests that Hsp70 can keep a protein folded even if it thermodynamically tends to misfold and aggregate (Here I only consider reversible aggregation). The chaperone is thus able to play a critical role in maintaining protein conformations, not just in the folding of nascent chains<sup>40</sup>. Higher DnaK concentrations are required to suppress aggregation at increasing substrate concentrations (Fig. 5a) or at decreasing substrate stabilities (Fig. 5b). This may explain how cells that overexpress DnaK can tolerate higher numbers of mutations in the chaperone's substrates<sup>41</sup>. Because the excess free energy plateaus at high chaperone concentrations (Fig. 2c), my results imply a limit on the chaperones' capacity to prevent aggregation, in that there exists a threshold of aggregation tendency (Fig. 5a,b, the black arrows) above which the chaperone can no longer maintain high levels of native concentrations and prevent misfolding/aggregation at the same time.

My model suggests that Hsp70 only drives the folding of proteins with sufficiently slow conversion between  $U$  and  $F$  states (Fig. 5c-e), implying that Hsp70 substrates tend to be slow refolding proteins (Fig. 5d). If the conversion between  $U$  and  $F$  is too fast, the chaperone diminishes, rather than increases, the native fraction in comparison to the chaperone-free equilibrium. As the conversion slows, the chaperone drives the steady state native fraction higher, but at the price of longer refolding time (Fig. 5e), a trade-off reminiscent of that between speed and specificity in the kinetic proofreading mechanism<sup>42,43</sup>, where the expenditure of free energy (such as from ATP or GTP consumptions) is utilized to increase the specificity of chemical reactions.

My model explains the observation that folding is less efficient at both low and high DnaJ concentrations<sup>17</sup> (Fig. 6a). At low DnaJ concentrations, ATP hydrolysis is slow, and nucleotide exchange drives DnaK toward the ATP-state, in which it dissociates from the substrate rapidly and thus unable to prevent aggregation. At high DnaJ

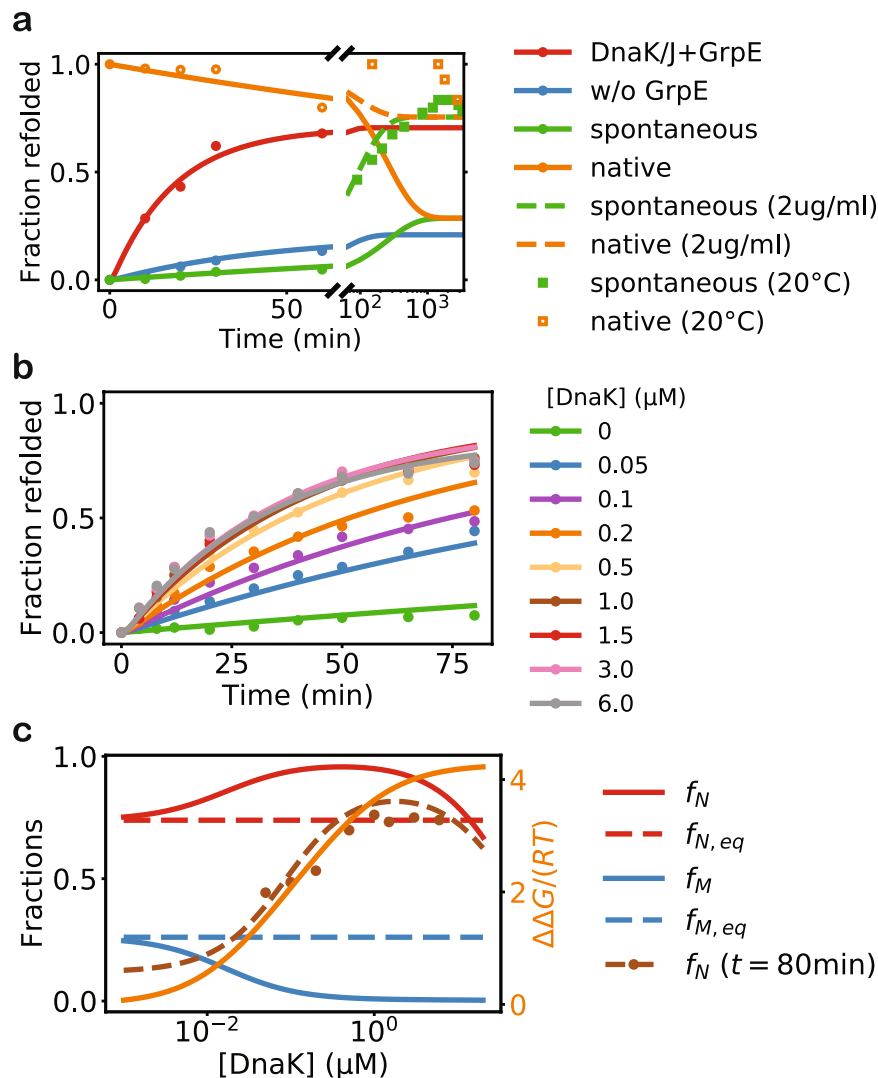
Reaction	Parameter	Rate	Reference
$S + C \cdot ATP \rightleftharpoons S \cdot C \cdot ATP$	$k_{a,C \cdot ATP}^{(0)}$	$1.28 \times 10^6 M^{-1} \cdot s^{-1}$	6
	$k_{d,C \cdot ATP}$	$2.31 s^{-1}$	6
$S + C \cdot ADP \rightleftharpoons S \cdot C \cdot ADP$	$k_{a,C \cdot ADP}^{(0)}$	$1.17 \times 10^4 M^{-1} \cdot s^{-1}$	6
	$k_{d,C \cdot ADP}$	$9.1 \times 10^{-4} s^{-1}$	6
$J + C \cdot ATP \rightleftharpoons J \cdot C \cdot ATP$	$k_{a,J \cdot C}$	$4.88 \times 10^4 M^{-1} \cdot s^{-1}$	17a
	$k_{d,J \cdot C}$	$1.6 \times 10^{-3} s^{-1}$	51
$S \cdot C \cdot ATP \rightarrow S \cdot C \cdot ADP + Pi$	$k_h^{(0)}$	$0.01 s^{-1}$	17b
$J \cdot S \cdot C \cdot ATP \rightarrow J \cdot S \cdot C \cdot ADP + Pi$	$k_h^{(J)}$	$1.8 s^{-1 c}$	17
$C \cdot ADP \rightarrow C + ADP$	$k_{d,ADP}$	$0.006 s^{-1}$	52
$E + C \cdot ADP \rightleftharpoons E \cdot C \cdot ADP$	$k_{a,E \cdot C}$	$7.85 \times 10^6 M^{-1} \cdot s^{-1}$	29d
	$k_{d,E \cdot C}$	$30 s^{-1}$	29
$C_{closed} \cdot ATP \rightarrow C_{open} \cdot ATP$	$k_C^{(0)}$	$1.1 \times 10^{23} s^{-1}$	50e
	$E_a/R$	17320 K	50

**Table 1.** The experimental kinetic parameters for the DnaK/DnaJ/GrpE chaperone system. All the parameters are determined at the temperature of 25 °C, except for the rate of conformational change of DnaK from closed to open conformations, which was measured at 15 °C, 25 °C, and 35 °C, yielding the parameters  $k_C^{(0)}$  and  $E_a$  from a fitting of the Arrhenius equation to the measured rates at different temperatures. <sup>a</sup>The association rate constant of DnaJ binding to DnaK was determined from the DnaJ concentration dependence of the DnaK ATP hydrolysis rates, by fitting the hydrolysis rate  $k_h([J]) = K_h(1 + K_h)^{-1}k_h^{(J,w/o S)} + (1 + K_h)^{-1}k_h^{(0,w/o S)}$  where  $K_h \equiv k_{a,J \cdot C} \cdot [J]/(k_{d,J \cdot C} + k_h^{(J,w/o S)})$ , to the experimental data with respect to  $k_{a,J \cdot C}$ ,  $k_h^{(J,w/o S)}$ , and  $k_h^{(0,w/o S)}$ , yielding the  $k_{a,J \cdot C}$  value in the table,  $k_h^{(J,w/o S)} = 0.76 s^{-1}$ , and  $k_h^{(0,w/o S)} = 6 \times 10^{-4} s^{-1}$ . I note that the hydrolysis rates are measured without a protein substrate (indicated by the superscript w/o S), which can further accelerate the hydrolysis. Surface plasmon resonance measurement of DnaJ binding to DnaK yielded  $k_{a,J \cdot C} = 2.3 \times 10^4 M^{-1} \cdot s^{-1}$ <sup>51</sup>, in good agreement with our fit. <sup>b</sup>Substrate-free DnaK has a basal ATP hydrolysis rate of  $0.001 s^{-1}$ , but the substrate can further accelerate ATP hydrolysis by up to 9-fold<sup>17,53</sup>. I take the ATP hydrolysis rate of substrate-bound DnaK to be 10-fold higher than the basal rate. The predictions of my model are insensitive to this parameter. <sup>c</sup>This is the experimental rate for the temperature  $T = 25$  °C. For the higher temperature  $T = 30$  °C, I used an arbitrary but reasonable 5.5-fold higher value of  $10 s^{-1}$  because I could not find any reported experimental value for this temperature. <sup>d</sup>The kinetic rates for GrpE binding to DnaK were not determined, but the steady state ratio in the two-step dissociation reaction,  $C \cdot ADP + E \rightleftharpoons E \cdot C \cdot ADP \rightleftharpoons E \cdot C + ADP$ ,  $K'_D = (k_{d,E \cdot C} + k_{d,ADP})/k_{a,E \cdot C} = 20 \mu M$  was determined. I chose an arbitrary diffusion limited association rate for GrpE binding to DnaK in our calculations. <sup>e</sup>I note that  $k_C^{(0)}$  and  $E_a$  appear too large to be physically meaningful; they should instead be taken simply as numerical parameters that yield an excellent fit of the Arrhenius equation to the experimental data.

Reaction	Parameter	Luciferase		
		25 °C	30 °C	22 °C
$M \rightleftharpoons U$	$k_{M \rightarrow U} (s^{-1})^a$	0.02	0.02	0.02
	$k_{U \rightarrow M} (M^{-1} \cdot s^{-1})$	$9.4 \times 10^8$	$3.4 \times 10^9$	$6.4 \times 10^7$
$U \rightleftharpoons F$	$k_{U \rightarrow F} (s^{-1})$	0.23	0.88	0.26
	$k_{F \rightarrow U} (s^{-1})$	0.18	0.13	0.042
$F \rightleftharpoons N$	$k_{F \rightarrow N} (s^{-1})^b$	5	5	5
	$k_{N \rightarrow F} (10^{-3} \cdot s^{-1})$	1.4	20	1.1
$S + C \cdot X \rightleftharpoons S \cdot C \cdot X$	$n_C^{(U)}$	9.1	6.2	1.9
	$n_C^{(U)}/n_C^{(F)c}$	20	20	20
$S + J \rightleftharpoons S \cdot J$	$K_{A,J}^{(U)} (10^6 \cdot M^{-1})$	13	22	54
$(J + C) \cdot S \rightleftharpoons (J \cdot C) \cdot S$	$[J]_{eff} (M)^d$	$5 \times 10^{-3}$	$5 \times 10^{-3}$	$5 \times 10^{-3}$

**Table 2.** The model parameters fit to the refolding experiments. <sup>a</sup>The predictions of my model depend on the equilibrium constant  $K_{M \rightleftharpoons U} = k_{M \rightarrow U}/k_{U \rightarrow M}$ , but are degenerate with respect to the individual rates. I thus fix an arbitrary but plausible value for  $k_{M \rightarrow U}$  and fit  $k_{U \rightarrow M}$  to the experimental refolding data. <sup>b</sup>For the similar reason as above, I fix  $k_{F \rightarrow N}$  and fit  $k_{N \rightarrow F}$  to the refolding data. <sup>c</sup>There should be more accessible DnaK binding sites in the U state than in the F state. Here, I arbitrarily set the ratio between the two. Although the values of the other fitting parameters will change accordingly, the quality of the fit and the predictions of my model are insensitive to this ratio (at least for values between 1 and 100). <sup>d</sup>I assume that the effective distance, L, between the Hsp70 molecule and the J domain of the Hsp40 molecule bound to the same substrate molecule is  $L = 4.3$  nm, the effective concentration is then  $[J]_{eff} = N_A^{-1}/(L^3 4\pi/3) = 5 \times 10^{-3} M$ . The quality of the fit and the predictions of my model are insensitive to this parameter.

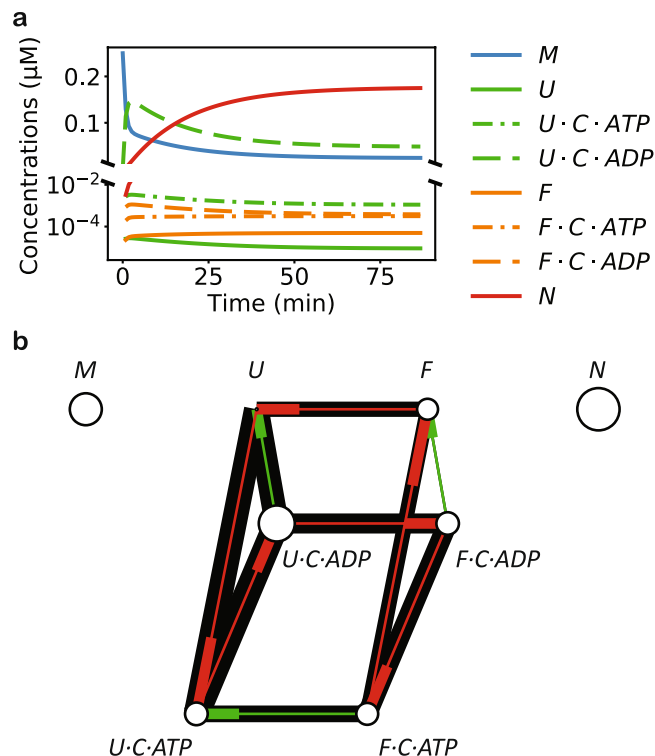




**Figure 2.** My model is in good agreement with previous experimental studies of DnaK/DnaJ/GrpE-mediated refolding. **(a)** The refolding of denatured luciferase under various conditions. The predictions of my model are shown in lines, whereas the experimental data are shown as filled circles<sup>5</sup>. The dashed lines show the spontaneous refolding and denaturation at the much lower substrate concentration of  $0.032 \mu\text{M}$  ( $2 \mu\text{g/ml}$ ), as in the corresponding experiments (empty circles and squares)<sup>4</sup>. The experiments of spontaneous refolding and denaturation were performed at  $20^\circ\text{C}$ , lower than the temperature of  $25^\circ\text{C}$  at which most of the kinetic parameters were obtained. In modeling the refolding, I assume that initially all the protein is in the misfolded state, *i.e.*,  $f_M(t=0) = 1$ ; in modeling the denaturation, I assume that initially all the protein is in the native state, *i.e.*,  $f_N(t=0) = 1$ . **(b)** The refolding of a luciferase mutant, LucDHis6<sup>31</sup>, in the presence of DnaK at various concentrations. **(c)** The native fraction ( $f_N = [N]/[S]$ , red) and the misfolded fraction ( $f_M = [M]/[S]$ , blue) of LucDHis6 at the steady state of DnaK-mediated refolding at various DnaK concentrations. The corresponding fractions in the chaperone-free folding equilibrium,  $f_{N,eq}$  and  $f_{M,eq}$  are shown as dashed lines. The unitless excess free energy  $\Delta\Delta G/(RT)$  is shown in orange (right y-axis). The fractions after 80 min of refolding, starting from misfolded LucDHis6, are shown in brown. My model is in good agreement with the experimental data (filled circles), and it suggests that the refolding is still incomplete even after 80 min. The fitting parameters in my model are given in Table 2, and the conditions of the experiments considered in this paper are summarized in Table 3.

concentrations, a large fraction of the substrate in the  $U$  state is bound to DnaJ. These DnaJ-bound substrate molecules are trapped in the  $U$  state, unable to progress toward the  $F$  state, resulting in diminished folding.

My model also explains the observation that folding decreases at both low and high GrpE concentrations<sup>29</sup> (Fig. 6b). For the chaperone to effectively assist folding, nucleotide exchange should be much slower than ATP hydrolysis when the chaperone binds to a substrate in the  $U$  state, but much faster than ATP hydrolysis when it binds to a substrate in the  $F$  state, so that the chaperone is driven toward the ADP-state in the former case, and toward the ATP-state in the latter case (Fig. 1b). At low GrpE concentrations, nucleotide exchange is slow, leaving DnaK bound to the substrate in the  $F$  state predominantly in the ADP-state—as reflected by the low population of  $F \cdot C \cdot \text{ATP}$  (Fig. 6b), slowing its dissociation from the substrate and thus preventing the latter from folding to



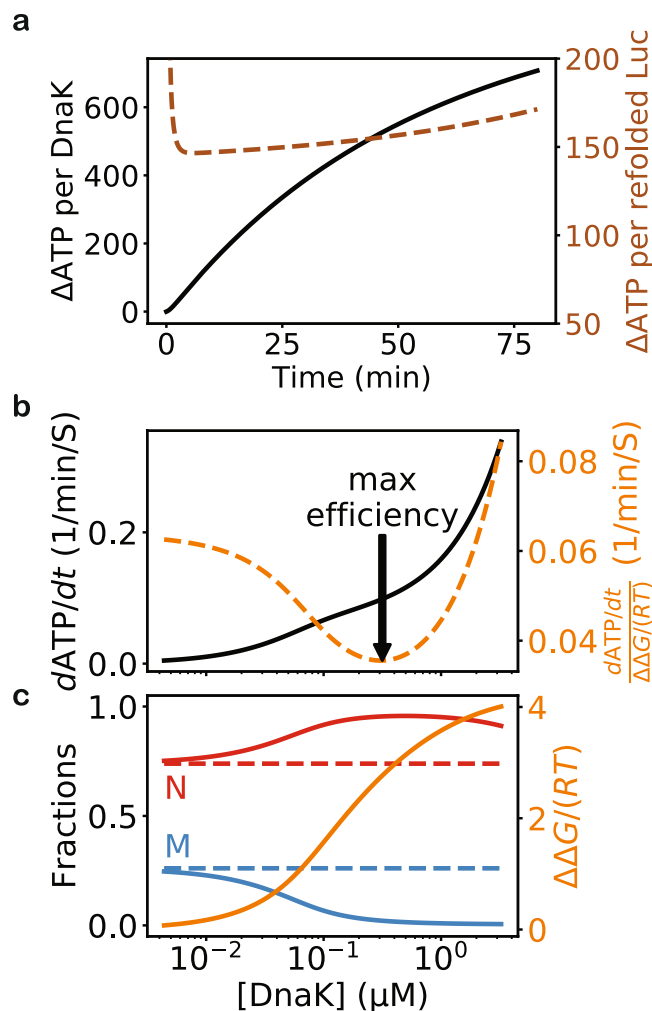
**Figure 3.** The mechanism by which the Hsp70 chaperone system accelerates refolding and maintains the folding out of equilibrium. **(a)** The model prediction of the concentrations of different molecular species in DnaK/DnaJ/GrpE-mediated refolding of denatured luciferase. The substrate is primarily held by ADP-bound chaperone in the  $U$  state before slowly moving into the native state. **(b)** The reactive flux at the steady state of LucDHis6 maintained by the DnaK/DnaJ/GrpE chaperone system. The thickness of each line is linear with respect to the logarithm of the absolute value of the reactive flux, and the thicker end of the thin, center line indicates the destination of the flux. The size of the circle at each node is linear with respect to the logarithm of the steady state fraction of the corresponding molecular species. The ATP-driven reaction cycle is highlighted in red.

	A <sup>4</sup>	B <sup>5</sup>	C <sup>17</sup>	D <sup>29</sup>	E <sup>31</sup>
[DnaK] ( $\mu\text{M}$ )	0	1.25	0.8	0.8	0 to 6
[DnaJ] ( $\mu\text{M}$ )	0	0.25	0 to 0.8	0.16	0.3
[GrpE] ( $\mu\text{M}$ )	0	1.25	0.4	0 to 4	0.8
[S] ( $\mu\text{M}$ )	0.0324	0.25	0.08	0.08	3
[ATP] (mM)	0	1	5	N/A	5
[ADP] (mM)	0	0	0	N/A	0
T ( $^{\circ}\text{C}$ )	20, 30	25	30	30	22

**Table 3.** The conditions of the refolding experiments (the corresponding references are in superscripts). A: Spontaneous refolding of guanidinium chloride (GdmCl)-denatured luciferase. B–D: DnaK/DnaJ/GrpE-mediated refolding of GdmCl-denatured luciferase. E: DnaK/DnaJ/GrpE-mediated refolding of LucDHis6 denatured by 4 freeze-thaw cycles. LucDHis6 is an engineered luciferase variant where the last 62 COOH-terminal residues are replaced by SKLSYEQDGLHAGSPAALHHHHHHH-COOH.

the native state. At high GrpE concentrations, nucleotide exchange is fast, and DnaK is driven into the ATP-state and does not stay bound to the substrate in the  $U$  state long enough—as reflected by the decreasing population of  $U \cdot C \cdot \text{ADP}$  (Fig. 6b)—to prevent the substrate from aggregation. To maximize substrate folding, higher nucleotide exchange rate should accompany higher stimulated ATP hydrolysis rate (Fig. 6c–e).

My model predicts that Hsp70 chaperones with higher Hsp40-stimulated ATP hydrolysis rates can drive substrate folding to higher native fractions (Fig. 6e), at the cost of higher free energy expenditure (Fig. 6f). This result explains a previous experimental observation that a small molecule that enhances ATP hydrolysis by Hsp40-bound Hsp70 can induce higher yields of substrate folding<sup>44</sup>. Modulation of the ATP hydrolysis or the nucleotide exchange rates by small molecules may represent a therapeutic opportunity in the treatment of misfolding- or proteostasis-related diseases<sup>45</sup>.

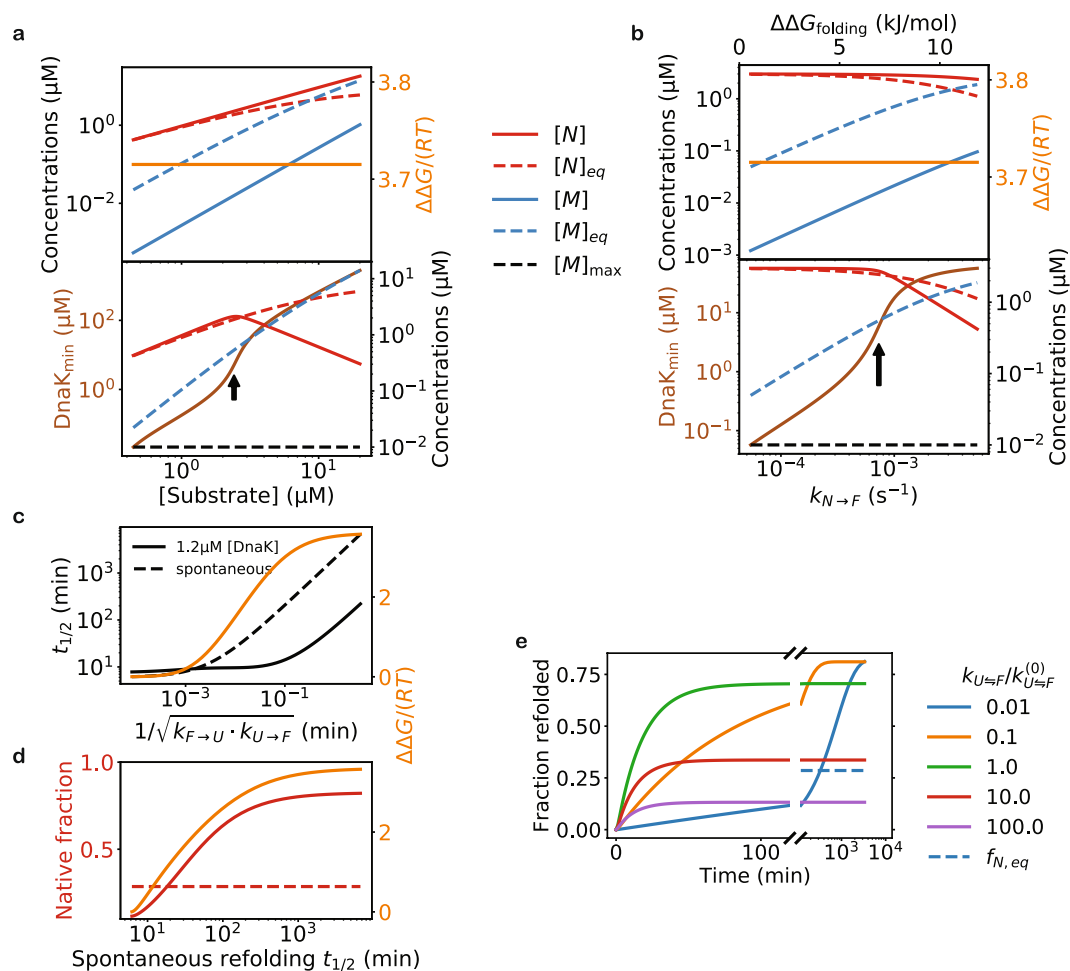


**Figure 4.** Free energy consumption in DnaK/DnaJ/GrpE-mediated folding of LucDHis6. (a) ATP consumption in the course of refolding of denatured LucDHis6. The instantaneous rate of ATP consumption is given by  $d\text{ATP}/dt = \sum_{S \in \{U, F\}} k_h^{(S)}(J) \cdot [S \cdot C \cdot \text{ATP}]$ , and the number of hydrolyzed ATP per molecule of refolded substrate can be estimated by dividing the cumulative consumption,  $\Delta\text{ATP} = \int_0^\tau d\text{ATP}$ , by the number of refolded substrate molecules after time  $\tau$ . Here, the DnaK concentration is  $0.5 \mu\text{M}$ , and the other kinetic parameters are given in Tables 1 and 2. The black curve shows the number of ATP molecules hydrolyzed per DnaK molecule after the given time. The brown curve shows the number of ATP molecules consumed per one molecule of refolded LucDHis6 (right y-axis) up to the given time, which increases to infinity at the steady state because no additional LucDHis6 is refolded, yet ATP hydrolysis continues. (b,c) ATP hydrolysis at the steady state. The ATP consumption rate per substrate, at various DnaK concentrations, is shown as the black curve in panel b, and the corresponding native (red) and misfolded (blue) fractions at the steady state are shown as solid lines in panel c. The native fraction is above and the misfolded fraction is below their respective equilibrium values (dashed flat lines). The excess free energy  $\Delta\Delta G/(RT)$  is shown in orange (right y-axis) in panel c. I can measure the chaperones' free energy efficiency in maintaining the non-equilibrium by the ratio of the ATP consumption rate to the excess free energy at the steady state (orange, right y-axis, in panel b). The arrow indicates the DnaK concentration at which the chaperones utilize the least amount of ATP per unit of excess free energy.

## Discussion

Key to my model is the assumption that Hsp40 has different affinities for a substrate in different conformations, favoring the misfolding-prone conformation over the fold-competent conformation. This is supported by a number of experimental observations. Hsp40 binds to exposed hydrophobic sites using a Zn finger-like domain within its CTD<sup>20</sup>. It has been inferred from experiments that the binding of Hsp40 CTD to substrate is the strongest for unfolded peptides, weaker for partially unfolded proteins, and the weakest for native proteins, and that Hsp40 is able to distinguish the substrate conformations<sup>22</sup>. Hsp40 has been experimentally shown to bind to a number of denatured proteins, including denatured luciferase<sup>20–22</sup>, but it binds to few native proteins, notably to  $\sigma^{32}$ , which adopts a loosely folded and highly flexible conformation<sup>46</sup>, and to RepE, the binding to which depends on the G/F-rich domain (outside CTD) of Hsp40 instead of the Zn finger-like domain<sup>22</sup>. These results suggest that Hsp40 preferentially binds to loose conformations with many exposed hydrophobic sites, which, together with

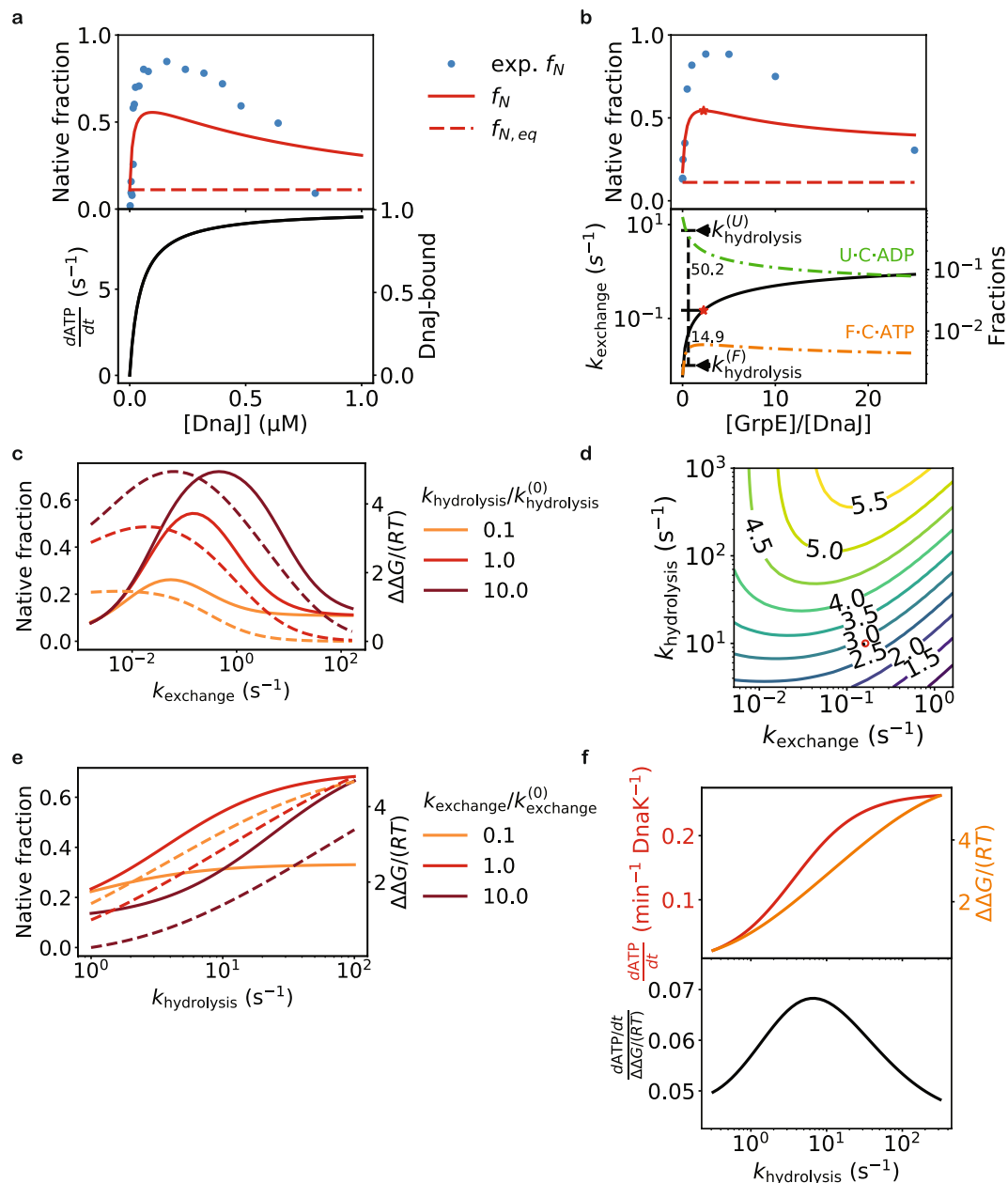




**Figure 5.** The capacity of the DnaK/DnaJ/GrpE chaperones to prevent reversible aggregation and to elevate native protein concentrations. **(a)** The chaperones can prevent aggregation at increasing substrate concentrations. The rate of aggregation is taken to be proportional to the substrate concentration (see *Methods*). Top: the native and the misfolded steady state concentrations at increasing total LucDHis6 concentrations, with the DnaK concentration fixed at  $1.2 \mu\text{M}$ . Bottom: the DnaK concentration required to maintain the misfolded protein concentration at or below  $[M]_{\text{max}} = 0.01 \mu\text{M}$ , as well as the steady state concentration of the native substrate at that DnaK concentration. **(b)** The chaperones can prevent aggregation at decreasing substrate stability. I vary the protein stability by changing the rate constant of conversion,  $k_{N \rightarrow F}$ , from the  $N$  state to the  $F$  state; the corresponding change in the folding free energy  $\Delta\Delta G_{\text{folding}}$  is indicated on the top axis. The native and the misfolded concentrations, as well as the DnaK concentration required to prevent aggregation, are shown as in panel a. **(c,d)** Hsp70 is more efficient at folding substrates with slower conversion between the  $U$  and the  $F$  states. Here, I take the kinetic parameters of luciferase folding at  $25^\circ\text{C}$ , and simultaneously scale the forward and reverse rates of the reaction  $U \rightleftharpoons F$  by the same factor, thus changing the kinetics without affecting the folding equilibrium. The times,  $t_{1/2}$ , for the refolding of the misfolded substrate to reach half of the native fraction at equilibrium (spontaneous refolding) or the steady state (mediated by DnaK/DnaJ/GrpE), as well as the excess free energy (orange, right y-axis), are plotted against the hypothetical rates of conversion in c. The native fractions (red, left y-axis) and the excess free energy (orange, right y-axis) at the steady state are plotted against  $t_{1/2}$  of spontaneous refolding in d; the equilibrium native fraction is shown as the red dotted line. **(e)** The time courses of Hsp70-mediated refolding of the misfolded substrate at different hypothetical rates of conversion between  $U$  and  $F$  (keeping  $k_{U \rightarrow F}/k_{F \rightarrow U}$  constant). Higher steady state native fractions are obtained at the price of longer refolding times.

the experimental observation that the fold-competent conformation is more compact with fewer exposed hydrophobic sites than the misfolding-prone conformation<sup>4</sup>, provides support for the assumption. Future experiments may directly test the assumption.

My model makes two distinct predictions that subject it to future experimental tests and possible falsification. First, it predicts that some thermodynamically unstable substrates depend on continuous Hsp70 assistance to maintain their native structures, and such a substrate in the steady state of Hsp70-mediated folding will gradually lose its native structure upon disruption of the chaperone system. Second, it predicts that an Hsp70 molecule



**Figure 6.** The rates of Hsp40-catalyzed ATP hydrolysis and NEF-catalyzed nucleotide exchange affect the efficiency of Hsp70-mediated folding. **(a)** Refolding of luciferase at various DnaJ concentrations. Top: the steady state native fractions predicted by my model, compared to the experimental data of refolding after 30 min at 30 °C<sup>17</sup>. Bottom: the ATP hydrolysis rate (left y-axis) and the fraction of substrate bound to DnaJ (right y-axis). **(b)** Refolding of luciferase at various GrpE concentrations. Top: the steady state native fractions predicted by my model, compared to the experimental data of refolding after 2 hours at 30 °C<sup>29</sup>. Bottom: the rate of nucleotide exchange at different GrpE concentrations (black line, left y-axis), and the populations of  $F \cdot C \cdot ATP$  and  $U \cdot C \cdot ADP$  (right y-axis). The optimal GrpE concentration is indicated by the red stars, and the in-plot numbers show the corresponding ratios of the nucleotide exchange rate to the ATP hydrolysis rates in the  $U$  and  $F$  states. In **a** and **b**, the rates of GrpE-catalyzed nucleotide exchange and DnaJ-catalyzed ATP hydrolysis are adjusted for the temperature of 30 °C (see *Methods* and *Table 1*). **(c)** Folding efficiency at different hypothetical rates of nucleotide exchange, for different values of the ATP hydrolysis rate in the  $U$  state. Native fractions (solid lines, left y-axis) are diminished at both low and high nucleotide exchange rates. At high rates of nucleotide exchange, the excess free energies (dashed lines, right y-axis) approach zero, indicating that Hsp70 can no longer drive protein folding. **(d)** The excess free energy as a function of the nucleotide exchange and the DnaJ-catalyzed ATP-hydrolysis rates. The rates used to model DnaK/DnaJ/GrpE-mediated folding at 30 °C are indicated by the red circle. **(e)** Folding efficiency increases with the DnaJ-catalyzed ATP hydrolysis rate, yielding higher native fractions (solid lines, left y-axis) and larger excessive free energies (dashed lines, right y-axis). **(f)** Higher ATP hydrolysis rate yields larger excess free energy (orange, right y-axis, top), at the price of higher rate of ATP consumption (red, left y-axis, top). The ratio of the two (bottom) changes only slightly.

bound to a substrate molecule will dissociate faster if the substrate is in the fold-competent conformation than if it is in the misfolding-prone conformation, and that this difference will disappear in the absence of Hsp40.

In support of the first prediction above, a recent experiment has demonstrated that luciferase at 37 °C can be kept active by the DnaK/DnaJ/GrpE chaperone system when there is sufficient ATP, but it rapidly loses its activity when ATP is depleted by the addition of apyrase<sup>13</sup>. The interpretation of this experiment, however, is complicated because apyrase also affects the luciferase activity assay. Here, based on my model, I propose an alternative experiment, in which Hsp70-mediated maintenance of luciferase activity is disrupted by inhibiting the simultaneous binding of Hsp40 to Hsp70 and to the substrate protein. For example, an isolated J-domain (e.g., DnaJ with its CTD deleted) can be used to compete against Hsp40 in binding to the Hsp70; alternatively, two D-peptides known to compete against substrate for binding to DnaJ, without binding to DnaK, can be used to inhibit DnaJ binding to the substrate<sup>47</sup>. When the J-domain or the D-peptide is added in excess to luciferase kept active by the DnaK/DnaJ/GrpE system, my model predicts that luciferase will lose its activity.

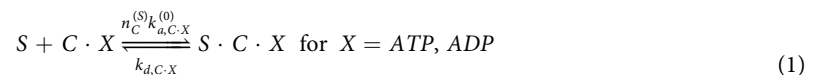
The second prediction of my model may be tested by kinetic experiments. Luciferase can be unfolded to different extent at different concentrations of the chemical denaturant guanidinium chloride (GdmCl): luciferase adopts a more compact unfolded structure with fewer exposed hydrophobic sites at lower concentrations of the denaturant than at higher concentrations of the denaturant<sup>4</sup>. My model predicts that, in the presence of Hsp40 and ATP, the dissociation rate of Hsp70 from luciferase denatured by low concentrations of GdmCl will be higher than from luciferase denatured at high concentrations of GdmCl, and that this difference in the dissociate rates will disappear in the absence of Hsp40.

Single molecule experiments<sup>25,34</sup> may provide a more stringent test of the second prediction of my model, if one can monitor both the residence time of Hsp70 on a substrate molecule and the probability that the same substrate molecule subsequently folds into the native structure. My model predicts that in the presence of Hsp40 and ATP, the folding probability will be higher if the residence time is shorter, but this correlation will vanish in the absence of Hsp40. Such experiments may be feasible if, for instance, separate fluorescence signals to detect Hsp70-substrate binding and substrate folding become available.

## Methods

**Model of Hsp70-mediated protein folding.** I denote Hsp70 as *C*, Hsp40 (J protein) as *J*, and the NEF as *E*. [*Y*] denotes the solution concentration of the molecular species *Y*. There are four types of reactions explicitly considered in my model (Fig. 1b):

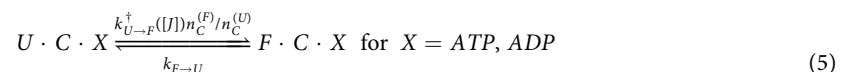
- 1) Hsp70 binding to the substrate.



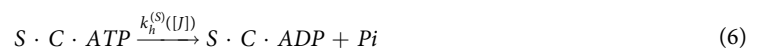
- 2) Conformational transitions of the substrate. An Hsp70-free substrate can adopt any of the four conformational states



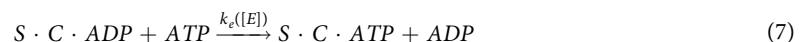
The chaperone-bound substrate can only be in and transition between the *U* and *F* states



- 3) ATP hydrolysis.



- 4) Nucleotide exchange.



The details of the kinetic rates of the above reactions are described below.

**Hsp70 binding to the substrate.** The association rate constant of Hsp70 binding to a substrate in state *S*,  $k_{a,C,X}^{(S)}$ , is determined by the number of accessible binding sites,  $n_C^{(S)}$ , in that conformation:  $k_{a,C,X}^{(S)} = n_C^{(S)} k_{a,C,X}^{(0)}$ , where  $k_{a,C,X}^{(0)}$  is the association rate constant of the chaperone binding to a fully accessible binding site, and  $X = ATP, ADP$ . The dissociation rate constant of Hsp70 from the substrate,  $k_{d,C,X}$ , does not depend on the sub-

strate conformation, but depends on whether nucleotide  $X = ATP$  or  $ADP$  is bound. Experimentally,  $k_{a,C-ATP}^{(0)} \gg k_{a,C-ADP}^{(0)}$  and  $k_{d,C-ATP} \gg k_{d,C-ADP}$ . I assume  $n_C^{(U)} > n_C^{(F)}$ .

**Conformational transitions of the substrate.** The transition rates between conformations  $S$  and  $S'$  are different between a chaperone-free substrate ( $k_{S \rightarrow S'}$ ) and a chaperone-bound substrate ( $k_{S \cdot C \cdot X \rightarrow S' \cdot C \cdot X}$ ) (Fig. 1b). The condition of thermodynamic cycle closure dictates that

$$\frac{k_{S \cdot C \cdot X \rightarrow S' \cdot C \cdot X}}{k_{S' \cdot C \cdot X \rightarrow S \cdot C \cdot X}} = \frac{k_{a,C \cdot X}^{(S')} k_{S \rightarrow S'}}{k_{a,C \cdot X}^{(S)} k_{S' \rightarrow S}} = \frac{n_C^{(S')} k_{S \rightarrow S'}}{n_C^{(S)} k_{S' \rightarrow S}}. \quad (8)$$

Because Hsp40 has different affinities for different substrate conformations, the transition rates between the conformations will depend on whether the substrate is bound to Hsp40. I treat the effects of Hsp40 on the reactions implicitly by making the affected rate constants dependent on the solution Hsp40 concentration  $[J]$  (see below).

For the transition  $U \cdot C \cdot X \rightleftharpoons F \cdot C \cdot X$ , I assume that the bound chaperone does not hinder the substrate to go from the  $F$  state to the  $U$  state, because a binding site available in the  $F$  state is most likely also available in the  $U$  state (based on the assumption  $n_C^{(U)} > n_C^{(F)}$ ). Thus I take  $k_{F \cdot C \cdot X \rightarrow U \cdot C \cdot X} = k_{F \rightarrow U}$ . It follows from thermodynamic cycle closure that the rate of the reverse transition—I use the superscript dagger to indicate that they are influenced by the presence of Hsp40—is

$$k_{U \cdot C \cdot X \rightarrow F \cdot C \cdot X}^\dagger([J]) = k_{F \rightarrow U}^\dagger([J]) n_C^{(F)} / n_C^{(U)} \quad (9)$$

I take the rate of (reversible) aggregation to be proportional to the substrate concentration:

$$k_{U \rightarrow M}^\dagger([U], [J]) = [U] \cdot k_{U \rightarrow M}^\dagger([J]) \quad (10)$$

The rates  $k_{U \rightarrow S}^\dagger([J])$  (for  $S = F, M$ ) depend on the affinities of Hsp40 for the substrate in different conformational states. For simplicity, I assume that Hsp40 only binds to the substrate in the  $U$  state, and consequently only the Hsp40-free substrate can change from conformation  $U$  to  $F$  or  $M$ . The corresponding transition rates are

$$k_{U \rightarrow S}^\dagger([J]) \approx (1 - p_j([J])) \cdot k_{U \rightarrow S} \approx \frac{1}{1 + [J] \cdot K_{A,J}^{(U)}} k_{U \rightarrow S}, \quad (11)$$

where  $k_{U \rightarrow S}$  is the rate of transition  $U \rightarrow S$  (for  $S = F, M$ ) for an Hsp40-free substrate,  $p_j$  is the probability that the substrate is Hsp40-bound, and  $K_{A,J}^{(S)}$  is the binding constant of Hsp40 for the conformational state  $S$ .

**Hsp40-substrate binding.** To keep my model simple, I do not explicitly consider the kinetics of binding and unbinding between Hsp40 and the substrate, and make the approximation that they are always at equilibrium. The key assumption of my model is that the fold-competent conformation  $F$  is much less accessible to Hsp40 than the misfolding-prone conformation  $U$ , i.e.,  $K_{A,J}^{(F)} \ll K_{A,J}^{(U)}$ . To reduce the number of unknown parameters, I take  $K_{A,J}^{(F)} \approx 0$ , i.e., the binding of Hsp40 to the fold-competent conformation is negligible, as in our derivation of the transition rates above. I also neglect subtleties such as that the  $J$  domain may bind with different affinities to Hsp70 in the ATP- and ADP-states<sup>48</sup>. The above approximations may contribute to quantitative differences between the predictions of my model and the experimental observations, particularly in predicting how folding changes with Hsp40 concentrations. The binding and unbinding of Hsp40 to the substrate and to Hsp70 can be explicitly included in my model at the cost of greater complexity and additional fitting parameters, but my simplified treatment above is adequate for the key results in this work.

**ATP hydrolysis.** The Hsp40-stimulated Hsp70 ATP hydrolysis rate,  $k_h^{(J)}$ , can be orders-of-magnitude higher than the unstimulated basal rate  $k_h^{(0)}$ . The ATP hydrolysis rate of Hsp70 bound to the substrate in the  $F$  state is simply  $k_h^{(F)}([J]) = k_h^{(0)}$ , following our approximation that no Hsp40 binds to the substrate in the  $F$  state. When Hsp70 is bound to the substrate in the  $U$  state, its average rate of ATP hydrolysis, given the solution Hsp40 concentration,  $[J]$ , can be approximated by

$$k_h^{(U)}([J]) \approx p_j([J]) \left( \frac{K_h}{1 + K_h} k_h^{(J)} + \frac{1}{1 + K_h} k_h^{(0)} \right) + (1 - p_j([J])) k_h^{(0)}, \quad (12)$$

where  $K_h \equiv k_{a,J \cdot C} \cdot [J]_{\text{eff}} / (k_{d,J \cdot C} + k_h^{(J)})$ , with  $k_{a,J \cdot C}$  and  $k_{d,J \cdot C}$  being the association and dissociation rates of  $J$  domain binding to Hsp70, and  $[J]_{\text{eff}}$  being the effective concentration of a substrate-bound Hsp40 molecule around an Hsp70 molecule bound to the same substrate molecule. The first term on the right hand side is the steady state rate of catalysis, weighted by the probability  $p_j$  that an Hsp40 is bound to the substrate and thus present to catalyze the hydrolysis. Here I assume a high ATP concentration such that ATP binding to Hsp70 is fast compared to other steps in ATP hydrolysis.

**NEF-catalyzed nucleotide exchange.** Because of the high concentration of ATP in cells and in the refolding experiments, I treat this reaction as irreversible. The reaction proceeds in three steps: (1) dissociation of ADP, (2) binding of ATP, and (3) conformational change of Hsp70 from the closed conformation in the ADP-state to the open conformation in the ATP-state. The conformational equilibrium between the open and closed conformations may be influenced by Hsp40 binding to Hsp70<sup>49</sup>, but this effect is not considered in my model for

simplicity and lack of experimental parameters. In the absence of the nucleotide exchange factor, the rate limiting step in the reaction is the dissociation of ADP, with the rate constant  $k_{d,ADP}$ , whereas when catalyzed by the NEF, the rate limiting step is the conformational change, with rate constant  $k_C$ <sup>29,50</sup>. The overall rate of reaction at a given NEF concentration,  $[E]$ , is then approximately

$$k_e([E]) = \frac{K_e}{1 + K_e} k_C + \frac{1}{1 + K_e} k_{d,ADP}, \quad (13)$$

where  $K_e \equiv k_{a,E-C} \cdot [E] / (k_{d,E-C} + k_C)$ , with  $k_{a,E-C}$  and  $k_{d,E-C}$  being the association and dissociation rates of NEF binding to Hsp70. The temperature dependence of  $k_C$  for DnaK has been determined to satisfy the Arrhenius equation:  $k_C(T) = k_C^{(0)} \exp(-E_a/(RT))$ , where  $R = 8.314$  J/mol/K is the gas constant.

**Solving the kinetic equations.** To simplify the calculations of refolding kinetics, I make the approximation that the solution concentrations of Hsp70, Hsp40, and NEF remain constant throughout the refolding process, which is true if they are in large excess of the substrate-bound chaperone, cochaperone, and NEF. Under this approximation, refolding kinetics is described by a set of linear ordinary differential equations, which are solved by the technique of eigenvalue decomposition of the rate matrix. This simplification allows quick and robust fitting of the folding kinetic parameters to the experimental refolding data. The steady state calculations do not use this approximation.

**Proof that without cochaperones Hsp70 cannot alter the population ratio between the native and the misfolded states.** Consider a hypothetical, single-conformation substrate  $s$  and a reference reaction cycle of chaperone binding ( $s + C \cdot X \rightleftharpoons s \cdot C \cdot X$  for  $X = ATP, ADP$ ), ATP hydrolysis ( $s \cdot C \cdot ATP \rightarrow s \cdot C \cdot ADP + Pi$ ), and nucleotide exchange ( $s \cdot C \cdot ADP + ATP \rightarrow s \cdot C \cdot ATP + ADP$ ), where the ATP-bound and ADP-bound chaperones bind to  $s$  with the rate constants  $k_{a,C-ATP}^{(0)}$  and  $k_{a,C-ADP}^{(0)}$ , respectively. Let  $q$ ,  $q_{C-ATP}$  and  $q_{C-ADP}$  be the fractions of the hypothetical substrate that are chaperone-free ( $s$ ), bound to an ATP-bound chaperone ( $s \cdot C \cdot ATP$ ), and bound to an ADP-bound chaperone ( $s \cdot C \cdot ADP$ ), respectively, at the steady state of this reaction cycle. For the real substrate in Hsp70-mediated folding, let  $f_S$  and  $f_{S,C \cdot X}$  ( $X = ATP, ADP$ ) be the fractions of the free and the  $C \cdot X$ -bound substrates in conformation  $S$ , and let  $f_{S,eq}$  be the fraction of the substrate in conformation  $S$  at the folding equilibrium in the absence of the chaperones. It can be verified that

$$f_S = Q^{-1} f_{S,eq} q \quad (14)$$

$$f_{S,C \cdot X} = Q^{-1} f_{S,eq} n_C^{(S)} q_{C \cdot X}, \text{ for } X = ATP, ADP \quad (15)$$

where

$$Q = \sum_S f_{S,eq} (q + n_C^{(S)} (q_{C-ATP} + q_{C-ADP})), \quad (16)$$

are the steady state solutions to the kinetic equations, given the condition of thermodynamic cycle closure in equation (8). Thus the ratio  $f_S / f_{S'} = f_{S,eq} / f_{S',eq}$  is not altered by the chaperone, despite the free energy expenditure of ATP hydrolysis. This holds true for all numbers of intermediate states, so long as the ATP hydrolysis and the nucleotide exchange rates of the chaperone do not depend on the conformational state of the bound substrate.

## Data Availability

No datasets were generated or analyzed during the current study.

## References

- Bukau, B., Weissman, J. & Horwich, A. Molecular chaperones and protein quality control. *Cell* **125**, 443–451 (2006).
- Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
- Balchin, D., Hayer-Hartl, M. & Hartl, F. U. *In vivo* aspects of protein folding and quality control. *Science* **353**, aac4354, <https://doi.org/10.1126/science.aac4354> (2016).
- Herbst, R., Schäfer, U. & Seckler, R. Equilibrium intermediates in the reversible unfolding of firefly (*Photinus pyralis*) luciferase. *J. Biol. Chem.* **272**, 7099–7105 (1997).
- Szabo, A. *et al.* The ATP hydrolysis-dependent reaction cycle of the *Escherichia coli* Hsp70 system DnaK, DnaJ, and GrpE. *Proc. Natl. Acad. Sci.* **91**, 10345–10349 (1994).
- Mayer, M. P. *et al.* Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol.* **7**, 586–593 (2000).
- Mayer, M. P. Hsp70 chaperone dynamics and molecular mechanism. *Trends Biochem. Sci.* **38**, 507–514 (2013).
- Cleric, E. M., Tilitsky, J. M., Meng, W. & Gierasch, L. M. How Hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J. Mol. Biol.* **427**, 1575–1588 (2015).
- Barducci, A. & Rios, D. L. P. Non-equilibrium conformational dynamics in the function of molecular chaperones. *Curr. Opin. Struct. Biol.* **30**, 161–169 (2015).
- Goloubinoff, P. & Rios, P. D. L. The mechanism of Hsp70 chaperones: (entropic) pulling the models together. *Trends Biochem. Sci.* **32**, 372–380 (2007).
- Santra, M., Farrell, D. W. & Dill, K. A. Bacterial proteostasis balances energy and chaperone utilization efficiently. *Proc. Natl. Acad. Sci.* **114**, E2654–E2661 (2017).
- Chakrabarti, S., Hyeon, C., Ye, X., Lorimer, G. H. & Thirumalai, D. Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium. *Proc. Natl. Acad. Sci.* **114**, E10919–E10927 (2017).
- Goloubinoff, P., Sassi, A. S., Fauvet, B., Barducci, A. & Rios, P. D. L. Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins. *Nat. Chem. Biol.* **14**, 388–395 (2018).



14. Kampinga, H. H. & Craig, E. A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell Biol.* **11**, 579–592 (2010).
15. Yang, J., Nune, M., Zong, Y., Zhou, L. & Liu, Q. Close and allosteric opening of the polypeptide-binding site in a human Hsp70 chaperone BiP. *Structure* **23**, 2191–2203 (2015).
16. Kityk, R., Kopp, J., Sinning, I. & Mayer, M. P. Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. *Mol. Cell* **48**, 863–874 (2012).
17. Laufen, T. *et al.* Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. *Proc. Natl. Acad. Sci.* **96**, 5452–5457 (1999).
18. Alderson, T. R. R., Kim, J. H. H. & Markley, J. L. L. Dynamical structures of Hsp70 and Hsp70-Hsp40 complexes. *Structure* **24**, 1014–1030 (2016).
19. Wittung-Stafshede, P., Guidry, J., Horne, B. E. & Landry, S. J. The J-Domain of Hsp40 couples ATP hydrolysis to substrate capture in Hsp70. *Biochemistry* **42**, 4937–4944 (2003).
20. Szabo, A., Korszun, R., Hartl, F. U. & Flanagan, J. A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO J.* **15**, 408–417 (1996).
21. Lu, Z. & Cyr, D. M. The conserved carboxyl terminus and Zinc finger-like domain of the co-chaperone Ydj1 assist Hsp70 in protein folding. *J. Biol. Chem.* **273**, 5970–5978 (1998).
22. Perales-Calvo, J., Muga, A. & Moro, F. Role of DnaJ G/F-rich domain in conformational recognition and binding of protein substrates. *J. Biol. Chem.* **285**, 34231–34239 (2010).
23. Rüdiger, S., Schneider-Mergener, J. & Bukau, B. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *EMBO J.* **20**, 1042–1050 (2001).
24. Sekhar, A. *et al.* Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions. *eLife* **7**, e32764, <https://doi.org/10.7554/eLife.32764> (2018).
25. Mashaghi, A. *et al.* Alternative modes of client binding enable functional plasticity of Hsp70. *Nature* **539**, 448–451 (2016).
26. Ahmad, A. *et al.* Heat shock protein 70 kDa chaperone/DnaJ cochaperone complex employs an unusual dynamic interface. *Proc. Natl. Acad. Sci.* **108**, 18966–18971 (2011).
27. Han, W. & Christen, P. Mechanism of the targeting action of DnaJ in the DnaK molecular chaperone system. *J. Biol. Chem.* **278**, 19038–19043 (2003).
28. Johnson, J. L. & Craig, E. A. An essential role for the substrate-binding region of Hsp40s in *Saccharomyces cerevisiae*. *J. Cell Biol.* **152**, 851–856 (2001).
29. Packschies, L. *et al.* GrpE accelerates nucleotide exchange of the molecular chaperone DnaK with an associative displacement mechanism. *Biochemistry* **36**, 3417–3422 (1997).
30. Brehmer, D. *et al.* Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. *Nat. Struct. Biol.* **8**, 427–432 (2001).
31. Sharma, S. K., Rios, P. D. L., Christen, P., Lustig, A. & Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat. Chem. Biol.* **6**, 914–920 (2010).
32. Rios, P. D. L. & Barducci, A. Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption. *eLife* **3**, e02218, <https://doi.org/10.7554/eLife.02218> (2014).
33. Astumian, R. D. Microscopic reversibility as the organizing principle of molecular machines. *Nat. Nanotechnol.* **7**, 684–688 (2012).
34. Nunes, J. M., Mayer-Hartl, M., Hartl, F. U. & Müller, D. J. Action of the Hsp70 chaperone system observed with single proteins. *Nat. Commun.* **6**, 6307, <https://doi.org/10.1038/ncomms7307> (2015).
35. Hu, B., Mayer, M. P. & Tomita, M. Modeling Hsp70-mediated protein folding. *Biophys. J.* **91**, 496–507 (2006).
36. Luengo, T. M., Kityk, R., Mayer, M. P. & Rüdiger, S. G. D. Hsp90 breaks the deadlock of the Hsp70 chaperone system. *Mol. Cell* **70**, 545–552 (2018).
37. Diamant, S., Peres Ben-Zvi, A., Bukau, B. & Goloubinoff, P. Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. *J. Biol. Chem.* **275**, 21107–21113 (2000).
38. Diamant, S. & Goloubinoff, P. Temperature-controlled activity of DnaK-DnaJ-GrpE chaperones: Protein-folding arrest and recovery during and after heat shock depends on the substrate protein and the GrpE concentration. *Biochemistry* **37**, 9688–9694 (1998).
39. Goloubinoff, P. *et al.* Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl. Acad. Sci.* **96**, 13732–13737 (1999).
40. Calloni, G. *et al.* DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Rep.* **1**, 251–264 (2012).
41. Aguilar-Rodríguez, J. *et al.* The molecular chaperone DnaK is a source of mutational robustness. *Genome Biol. Evol.* **8**, 2979–2991 (2016).
42. Murugan, A., Huse, D. A. & Leibler, S. Speed, dissipation, and error in kinetic proofreading. *Proc. Natl. Acad. Sci.* **109**, 12034–12039 (2012).
43. Hopfield, J. J. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci.* **71**, 4135–4139 (1974).
44. Wisén, S. *et al.* Binding of a small molecule at a protein-protein interface regulates the chaperone activity of Hsp70-Hsp40. *ACS Chem. Biol.* **5**, 611–622 (2010).
45. Labbadia, J. & Morimoto, R. I. The biology of proteostasis in aging and disease. *Annu. Rev. Biochem.* **84**, 435–464 (2015).
46. Rist, W., Jørgensen, T. J. D., Roepstorff, P., Bukau, B. & Mayer, M. P. Mapping temperature-induced conformational changes in the *Escherichia coli* heat shock transcription factor  $\sigma$ 32 by amide hydrogen exchange. *J. Biol. Chem.* **278**, 51415–51421 (2003).
47. Bischofberger, P., Han, W., Feifel, B., Schönfeld, H.-J. & Christen, P. D-peptides as inhibitors of the DnaK/DnaJ/GrpE chaperone system. *J. Biol. Chem.* **278**, 19044–19047 (2003).
48. Kim, J. H., Alderson, T. R., Frederick, R. O. & Markley, J. L. Nucleotide-dependent interactions within a specialized Hsp70/Hsp40 complex involved in Fe-S cluster biogenesis. *J. Am. Chem. Soc.* **136**, 11586–11589 (2014).
49. Yang, J. *et al.* Conformational transitions of the polypeptide-binding pocket support an active substrate release from Hsp70s. *Nat. Commun.* **8**, s41467–017, <https://doi.org/10.1038/s41467-017-01310-z> (2017).
50. Slepnev, S. V. & Witt, S. N. Kinetics of the reactions of the *Escherichia coli* molecular chaperone DnaK with ATP: evidence that a three-step reaction precedes ATP hydrolysis. *Biochemistry* **37**, 1015–1024 (1998).
51. Suh, W.-C. *et al.* Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. *Proc. Natl. Acad. Sci.* **95**, 15223–15228 (1998).
52. Russell, R., Jordan, R. & McMacken, R. Kinetic characterization of the ATPase cycle of the DnaK molecular chaperone. *Biochemistry* **37**, 596–607 (1998).
53. McCarty, J. S., Buchberger, A., Reinstein, J. & Bukau, B. The role of ATP in the functional cycle of the DnaK chaperone system. *J. Mol. Biol.* **249**, 126–137 (1995).

## Author Contributions

H.X. designed and performed the research and wrote the paper.

## Additional Information

**Competing Interests:** The author declares no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018