



# Kinetic Model for the Coupling between Allosteric Transitions in GroEL and Substrate Protein Folding and Aggregation

Riina Tehver<sup>1</sup> and D. Thirumalai<sup>1,2\*</sup>

<sup>1</sup>*Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

<sup>2</sup>*Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA*

*Received 29 October 2007;  
received in revised form  
16 January 2008;  
accepted 20 January 2008  
Available online  
31 January 2008*

The bacterial chaperonin GroEL and the co-chaperonin GroES assist in the folding of a number of structurally unrelated substrate proteins (SPs). In the absence of chaperonins, SP folds by the kinetic partitioning mechanism (KPM), according to which a fraction of unfolded molecules reaches the native state directly, while the remaining fraction gets trapped in a potentially aggregation-prone misfolded state. During the catalytic reaction cycle, GroEL undergoes a series of allosteric transitions ( $T \leftrightarrow R \rightarrow R'' \rightarrow T$ ) triggered by SP capture, ATP binding and hydrolysis, and GroES binding. We developed a general kinetic model that takes into account the coupling between the rates of the allosteric transitions and the folding and aggregation of the SP. Our model, in which the GroEL allosteric rates and SP-dependent folding and aggregation rates are independently varied without prior assumption, quantitatively fits the GroEL concentration-dependent data on the yield of native ribulose biphosphate carboxylase/oxygenase (Rubisco) as a function of time. The extracted kinetic parameters for the GroEL reaction cycle are consistent with the available values from independent experiments. In addition, we also obtained physically reasonable parameters for the kinetic steps in the reaction cycle that are difficult to measure. If experimental values for GroEL allosteric rates are used, the time-dependent changes in native-state yield at eight GroEL concentrations can be quantitatively fit using only three SP-dependent parameters. The model predicts that the differences in the efficiencies (as measured by yields of the native state) of GroEL, single-ring mutant (SR1), and variants of SR1, in the rescue of mitochondrial malate dehydrogenase, citrate synthase, and Rubisco, are related to the large variations in the allosteric transition rates. We also show that GroEL/S mutants that efficiently fold one SP at the expense of all others are due to a decrease in the rate of a key step in the reaction cycle, which implies that wild-type GroEL has evolved as a compromise between generality and specificity. We predict that, under maximum loading conditions and saturating ATP concentration, the efficiency of GroEL (using parameters for Rubisco) depends predominantly on the rate of  $R \rightarrow R''$  transition, while the equilibrium constant of the  $T \leftrightarrow R$  has a small effect only. Both under sub- and superstoichiometric GroEL concentrations, enhanced efficiency is achieved by rapid turnover of the reaction cycle, which is in accord with the predictions of the iterative annealing mechanism. The effects are most dramatic at substoichiometric

\*Corresponding author. Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, Gaithersburg, MD 20742, USA. E-mail address: [thirum@glue.umd.edu](mailto:thirum@glue.umd.edu).

Abbreviations used: SP, substrate protein; KPM, kinetic partitioning mechanism; Rubisco, ribulose biphosphate carboxylase/oxygenase; IAM, iterative annealing mechanism; WT, wild type; NBA, native basin of attraction; mtMDH, mitochondrial malate dehydrogenase; CS, citrate synthase; GFP, green fluorescent protein.

conditions (most relevant for *in vivo* situations) when SP aggregation can outcompete capture of SP by chaperonins.

© 2008 Elsevier Ltd. All rights reserved.

**Keywords:** chaperonin GroEL; allosteric transitions; kinetic partitioning mechanism; substrate protein folding and aggregation; GroEL reaction cycle and protein folding

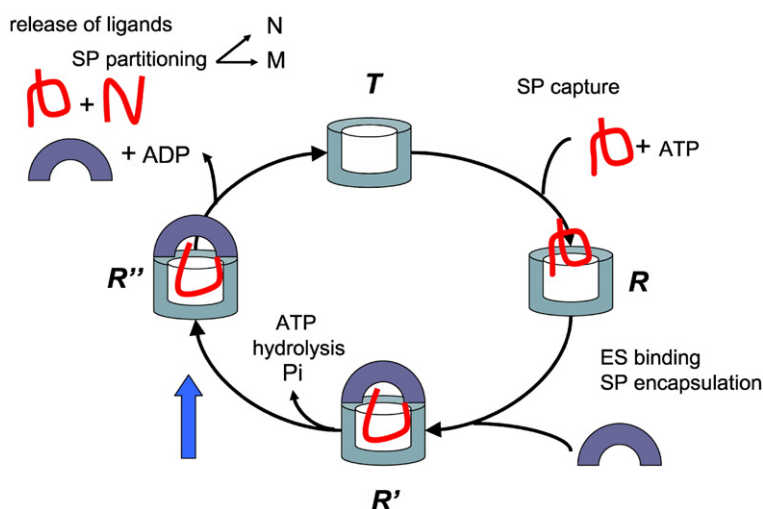
Edited by M. Levitt

## Introduction

Molecular chaperones are ATP-consuming machines that facilitate the folding of several structurally unrelated substrate proteins (SPs).<sup>1–5</sup> The best studied, both structurally<sup>6</sup> and mechanistically, of such machines are the *Escherichia coli* chaperonin GroEL and the co-chaperonin GroES.<sup>7–9</sup> GroEL consists of two heptameric rings that are stacked back to back. The coordinated interaction between the two rings in concert with SP and GroES binding and ATP hydrolysis allows for the GroEL machinery to rescue SPs that might otherwise aggregate. Although under maximal loading conditions (presence of SP, ATP, and GroES) it is necessary to consider the dual role of both rings, it is sufficient to use the single ring as the functional unit to describe the conceptual framework underpinning GroEL function.<sup>5,10</sup> Comparison of the structures<sup>11</sup> in the states GroEL visits during the catalytic cycle shown in Fig. 1 shows the spectacular conformational changes that it undergoes in response to interactions with SP, GroES, and ATP binding and hydrolysis. It follows from Fig. 1 that it is important to account for the coupling between the complex allosteric transitions that the GroEL particle undergoes and the fate of the SP during the reaction cycle in order to describe the workings of the GroEL–GroES machine.<sup>4,12–16</sup> Studies that artificially simplify the processes in the GroEL cycle (Fig. 1), such as using SPs that do not require GroES or mutants from which GroES does not dissociate on biologically relevant timescales,

while insightful, cannot accurately represent the mechanism of rescue of stringent SPs under non-permissive conditions. Moreover, genetic experiments show that various constituents cannot substitute for GroEL *in vivo*, which further attests to their limitations to serve as surrogate for the wild-type (WT) GroEL.<sup>17,18</sup>

Sometime ago, we developed the iterative annealing mechanism (IAM) to provide a quantitative description of the GroEL-assisted folding of SP.<sup>5,19</sup> The IAM incorporates the consequences of the structural changes in GroEL during the  $T \leftrightarrow R \rightarrow R''$  transitions (Fig. 1) and results from biophysical studies and theoretical description of monomeric folding. The IAM is based on the premise that GroEL plays an active role in SP folding. The large conformational changes that GroEL undergoes during the catalytic cycle and the experimental observations that, under operating conditions with an imposed load, the SP is ejected from the cavity,<sup>20–22</sup> folded or not, with each round of ATP hydrolysis, attest to the active role of GroEL.<sup>21,23</sup> It therefore follows that these large conformational changes enable the SP to fold under conditions in which spontaneous folding is unlikely. Even in the single-ring SR1 mutant, from which GroES dissociates with a lifetime of about 300 min,<sup>24</sup> the cavity volume has doubled as a result of GroES binding and ATP hydrolysis. In the process, the microenvironment that the SP feels changes as the GroEL  $T \leftrightarrow R \rightarrow R''$  transition takes place. A consequence of these observations is that, under conditions that place a load on the chaperonin



**Fig. 1.** Schematic sketch of the catalytic cycle of GroEL. SP is captured by GroEL in the T state. Binding of ATP triggers the equilibrium  $T \leftrightarrow R$  transition. Subsequently, GroES binding encapsulates the SP (assuming it can fit into the expanded cavity), leading to the  $R \rightarrow R''$  transition. Hydrolysis of ATP results in the  $R' \rightarrow R''$  transition. The volume of the cavity doubles from 85,000 Å<sup>3</sup> in the T state to about 170,000 Å<sup>3</sup> in the R'' (and presumably in the R') state. Folding of SP in the largely hydrophilic cavity in the R' or R'' state can occur with some probability. Interruption, shown by the continuous blue arrow, signals

release of SP, ADP, and GroES. Kinetic partitioning of SP, which is shown to occur for clarity in the bulk, could occur in the cell. In other words, folding to the native state or transition to the ensemble of misfolded structures can take place upon SP encapsulation.

machinery, GroEL is not merely a passive Anfinsen cage but participates actively as a stochastic machine that gives SPs multiple chances to fold by minimizing the possibility of aggregation.<sup>19,20,25</sup>

The goal of this work was to provide a kinetic framework that incorporates the coupling between allosteric transitions in GroEL and folding and aggregation of the SPs. Similar kinetic models have been successfully used to analyze experimental data on molecular motors<sup>26,27</sup> that also undergo a complex catalytic cycle in response to ATP binding and force. It is necessary to describe the various timescales whose interplay ultimately determines the efficiency of GroEL function in order to derive the kinetic model. It is well established, both theoretically<sup>28</sup> and experimentally,<sup>29,30</sup> that in the absence of GroEL the SPs fold by the kinetic partitioning mechanism (KPM).<sup>31</sup> According to KPM, a fraction of initially unfolded molecules reaches the native state rapidly, while the remaining fraction collapses into manifold misfolded conformations. The rates of transitions from the kinetically trapped misfolded structures to the native state, which might require partial unfolding of compact structures, are relatively slow. The KPM predicts that the fraction of molecules that do not fold at time  $t$  is:<sup>28</sup>

$$P_u(t) = \Phi e^{-k_F t} + \sum_{i=1}^m a_i e^{-k_i t} \quad (1)$$

where  $\sum_{i=1}^m a_i = 1 - \Phi$ ,  $k_F$  is the folding rate for the molecules that reach the native basin of attraction (NBA) without being kinetically trapped and  $k_i$  is the transition time from the  $i^{\text{th}}$  misfolded conformation to the NBA. Typically, the rate of collapse from the unfolded state to the misfolded ensemble is relatively rapid compared with  $k_i$ , while  $k_i/k_F \ll 1$ . In general, there may be an ensemble of misfolded structures. However, in practice, only a few (one or two) can be experimentally resolved. Here, without loss of generality, we will assume that escape from the collection of misfolded structures can be described by a single average rate,  $k_S$ . Thus, Eq. (1) simplifies to:

$$P_u(t) = \Phi e^{-k_F t} + (1 - \Phi) e^{-k_S t} \quad (2)$$

The KPM, which has also been used to describe RNA folding,<sup>31</sup> has been experimentally validated.<sup>29,30</sup>

The first step in the reaction cycle (Fig. 1) is the capture of the misfolded SP by GroEL in the T state. In the second stage, ATP binds to the seven sites in the equatorial domain that triggers movements (localized largely in the apical and intermediate domains) resulting in the transition to the R state. Recent NMR<sup>32</sup> and theoretical studies<sup>33</sup> showed that the SP is disordered and more expanded upon interacting with GroEL than it is in the bulk, which further reinforces the active role played by the chaperonins in the rescue of SPs. Two rates,  $k_{T \rightarrow R}$  and  $k_{R \rightarrow T}$ , describe the equilibrium of the  $T \rightarrow R$  transition.<sup>34,35,36</sup> In the third stage, GroES binds. The first of the irreversible steps in the cycle upon GroES binding involves ATP hydrolysis that drives the

allosteric  $R \rightarrow R''$  transition at a rate  $k_{R \rightarrow R''}$ . Currently, the structure of the  $R'$  [GroEL-(ATP)<sub>7</sub>-GroES] state is not known, although it is suspected that it is similar to that of the  $R''$  state. In addition, transition rates to or from  $R'$  have not been measured. It is likely that the irreversible steps in the reaction cycle after GroES binds are too rapid to resolve the  $k_{R \rightarrow R'}$  or  $k_{R' \rightarrow R''}$  transition. For these reasons, we did not consider the  $R'$  state, the inclusion of which would only increase the number of parameters without providing any increase in the accuracy of the predictions. The release of the SP (folded or not), ADP, and GroES requires signal from the lower (*trans*) ring (indicated by the blue arrow in Fig. 1).<sup>4,6,37</sup> Subsequently, the  $R''$  state relaxes back to the SP acceptor T state on a timescale  $k_{R'' \rightarrow T}$ . Thus, events in a single GroEL cycle in concert with changes in SP are minimally described by a number of rates, namely,  $k_F$ ,  $\{k_i\}$ ,  $k_{T \rightarrow R}$ ,  $k_{R \rightarrow T}$ , and  $k_{R \rightarrow R''}$ . In the proposed kinetic model, communication between rings, which occurs with negative cooperativity, is taken into account using the rate constant  $k_{R'' \rightarrow T}$  (see the blue arrow in Fig. 1). Inclusion of SP aggregation and release of SP from GroEL brings additional rate constants into assisted folding (see below).

The reaction cycle in Fig. 1 shows that, even when inter-protein interactions between SPs are neglected, there are a number of distinct timescales that have to be taken into account for a quantitative description of assisted folding. Here, we propose a general kinetic model that accounts for the coupling between SP folding, GroEL allostery, and a pathway leading to SP aggregation. The proposed kinetic model is used to analyze experimental data on chaperonin-assisted reconstitution of mitochondrial malate dehydrogenase (mtMDH), citrate synthase (CS), and the stringent substrate ribulose biphosphate carboxylase/oxygenase (Rubisco). The excellent agreement between the kinetic model and experiments not only validates the kinetic model but also allows us to explore other scenarios for efficient GroEL function. Because of the presence of a large number of timescales that are needed to describe the kinetics of the reaction cycle (Fig. 1), several scenarios for yield enhancement can emerge. We show that changes in the rate constants in the catalytic cycle, which depend on intrinsic folding properties of monomeric proteins, the concentrations of SPs, and the SP- and ATP-dependent GroEL allosteric rates, can dramatically change the efficiency of the function of this remarkable machine. Finally, we show that optimization of the yield of a particular protein at the expense of decrease in efficiency for other SPs can be achieved by tuning the  $k_{R \rightarrow R''}$  rate (Fig. 1).

## Results

### Steady-state solutions of the kinetic equations

In order to reveal the interplay between the allosteric transition rates and SP folding in a transparent

manner, we first obtained analytical expressions for the concentrations of the folded SPs and aggregated species in the steady-state limit. The applicability of the results obtained in the steady-state limit depends on the actual timescales. The cell doubling time of *E. coli*, under minimal glucose conditions, is on the order of 30 min, which is the most relevant biological timescale. The results obtained in the steady-state limit provide some insights into the more complex kinetic model. If the protein folding and chaperonin reaction rates are significantly faster than 30 min, then the relatively simple approach is meaningful. Below we analyze the steady-state results for two specific situations.

### Superstoichiometric chaperonin concentration

If the chaperonin concentration exceeds that of the SP, we can approximate  $[EL_T] \approx [EL_0]$  and linearize the protein binding reaction [Eq. (11)]  $\lambda_B[EL_T] \approx \lambda_B[EL_0] = k_B$ . In this limit, the native states and the aggregates are produced at rates that are independent of the chaperonin allostery-related timescales  $k_{T-R}^{-1}$ ,  $k_{R-T}^{-1}$ , and  $k_{R-R''}^{-1}$ . The steady-state rates of production of aggregates ( $v_{MA}$ ) and folded SPs ( $v_N$ ) are as follows:

$$\begin{aligned} v_{MA} &= \lambda_A \left[ \sqrt{\left( \frac{k_S}{2\lambda_A} + \frac{w}{2\lambda_A} \right)^2 + \frac{(1-\Phi)v_0}{\lambda_A} - \frac{k_S}{2\lambda_A} - \frac{w}{2\lambda_A}} \right]^2 \\ v_N &= v_0 - v_{MA} \end{aligned} \quad (3)$$

where we have assumed that nascent proteins are introduced to the cell at the rate  $v_0$ , and the parameter  $w$  is:

$$w = \frac{k_B}{1 + k_M/k_N} \quad (4)$$

Spontaneous folding without chaperonins corresponds to  $w=0$ .

The solutions have a few properties that are useful to point out. First, the production rate of the native proteins depends strongly on the effective SP binding rate  $k_B$ . For large  $k_B$  (compared with  $\lambda_A$ ),  $w$  [Eq. (4)] becomes large and  $v_{MA} \rightarrow 0$ ,  $v_N \rightarrow v_0$ , according to Eq. (3) (i.e., all proteins fold successfully). Thus, any mutation in the SP or GroEL that increases the binding rate would increase the native-state yield. Second, mutations that directly affect the protein release rates without affecting the folding rate in the cavity (i.e., that would affect  $k_N$  and  $k_M$  by the same factor) would have no effect on the native-state yield. In other words, the yield and aggregation depend on the ratio  $k_M/k_N$ , which is the partitioning of the ejected SPs only. Third, chaperonins cannot aid folding in the steady-state limit if folded proteins are not ejected from the cavity ( $w \rightarrow 0$  as  $k_N \rightarrow 0$ ). Finally, chaperonin effectiveness can be enhanced by increasing the ratio  $k_N/k_M$ . In the limit,  $k_N \gg k_M$ ,  $w \rightarrow k_B$  and the effectiveness are determined solely by the binding rate.

### Substoichiometric chaperonin concentration

The steady-state solutions for a low chaperonin concentration yield a lengthy expression for the reaction rates. However, if the second-order aggregation reaction rate can be replaced with an effective first-order rate  $k_A \approx \lambda_A[M]$ , simple expressions for the production of native states and aggregates can be found. In these expressions, the allosteric timescales appear in the following form:

$$A = \frac{k_N + k_M}{k_{R-R''}} \left( \frac{k_{R-T} + k_{R-R''}}{k_{T-R}} + 1 \right) + 1 \quad (5)$$

One can show that the native-state production rate increases as  $A$  decreases. It follows from Eq. (5), which is derived in Appendix A, that  $A$  can be decreased if the rates  $k_{T-R}$  and  $k_{R-R''}$  increase or upon the decrease of  $k_{R-T}$ . Thus, if the concentration of GroEL is less than that of the SP, then faster turn-around of the allosteric cycle leads to a bigger yield, as was predicted based on lattice simulations and the IAM. Because of the paucity of the number of GroEL/S particles in *E. coli*, it is likely that the substoichiometric condition is most appropriate for the *in vivo* situation. In this situation, even the steady-state solutions clearly show that rapid GroEL turnover leads to yield enhancement of the folded SP. Thus, in general, speeding of the  $k_{T-R}$  and  $k_{R-R''}$  should lead to increased efficiency (measured by the amount of native SP yield in a given time) of chaperonins. These predictions are borne out by solving the time-dependent kinetic equations (see Methods).

### The allosteric kinetic model accounts for [EL]-dependent folding kinetics of Rubisco

Although the steady-state solutions give insights into the variables in the reaction cycle that control the native-state yield, it is necessary to solve the coupled kinetic equations in order to compare the predictions with experiments. To explore the full effect of GroEL dynamics, from now on, we use the coupled time-dependent equations [Eq. (12)]. *In vitro* experiments can be used to measure the time-dependent yield of the native state of SPs as a function of GroEL, GroES, and ATP concentrations. In order to validate the kinetic model, we first analyze the measured yield of Rubisco as a function of time at various GroEL concentrations. In the process, we extract kinetic rates in the reaction cycle that are very difficult to measure.

Since the differential equations (see Methods) cannot be solved analytically, we integrate the kinetic equations [Eq. (12)] numerically. The model is used to extract parameters using experimental data on Rubisco,<sup>19</sup> whose folding was probed at various GroEL concentrations under non-permissive conditions. It is important to keep in mind, however, that variations in the experimental conditions, such as the temperature and the concentrations of co-chaperonins, nucleotides, and cations, could affect the specific rates significantly. Other SPs will have

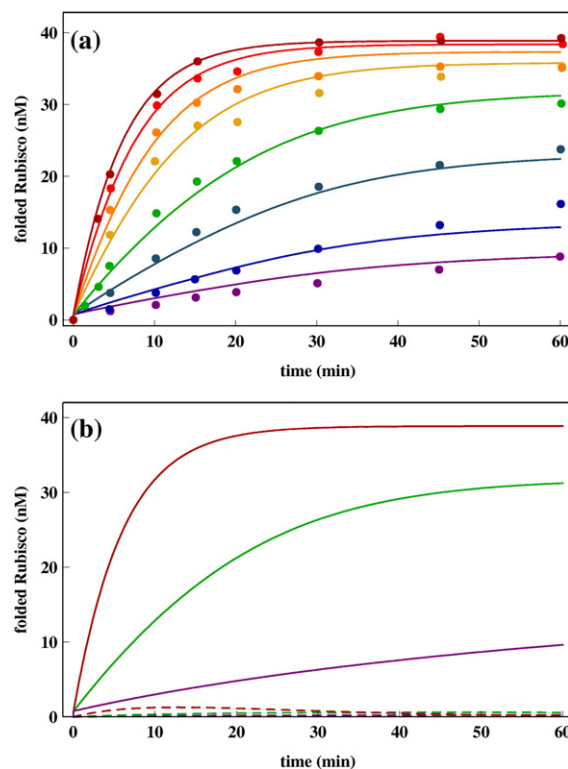


different intrinsic folding timescales and can also affect the GroEL allosteric rates (see below). Nevertheless, many of the qualitative results obtained here also provide predictions or illustrations of trends and dependencies of the GroEL efficiency on allosteric transitions or spontaneous folding rates of SP.

The data at eight values of GroEL concentration in Fig. 5 of Ref. 19 allow us to demonstrate that our model can reproduce the observed time-dependent yield of the native state of Rubisco and its dependence on the concentration of GroEL. In the process, we can extract the nine kinetic parameters [Eq. (12)]. Fortunately, the number of measurements reported in Ref. 19 is large enough that an unbiased fit to the equations can be carried out. In the first test of Eq. (12), we varied, without prejudice, the nine parameters to fit all eight curves at different [EL] concentrations. Further tests of the robustness of our model are described below. In assessing the validity of the model, it is important to not only check the quality of the fits but also ensure that the values of the parameters are physically reasonable. Our kinetic model fits the experimental data extremely well (Fig. 2). We should emphasize that all the data points were globally fit *using only one set* of parameters, describing the kinetic reaction cycle of GroEL and intrinsic Rubisco rates. By fitting the experimental data using our kinetic model, we obtained physically reasonable reaction rates for Rubisco (see Table 1). The intrinsic timescales associated with Rubisco are in accord with theoretical expectations. For example, the timescale for the fast process  $\tau_F \approx k_F^{-1}$  is in the range of time expected for a 491-residue protein.<sup>38</sup> The value  $k_S$  is consistent with measurements that monitored spontaneous folding of Rubisco in the presence of chloride ion.<sup>39</sup> The ratio  $k_F/k_S \approx 10^4 \gg 1$  is in accord with the predictions of KPM. Recall that  $k_S$  is an average rate for the transition from the ensemble of misfolded structures to the native state. It is likely a small fraction of misfolded states that can fold more rapidly (i.e.,  $1 \ll k_F/k_S < 10^4$ ). That, under experimental conditions, Rubisco folding requires the chaperonin machinery implies that the aggregation must compete with transition from misfolded structures, especially considering that the partitioning factor  $\Phi = 0.02$  is small. Indeed, the pseudo-first-order process for aggregation  $k_A \approx \lambda_A[\text{SP}] > k_S$  when  $[\text{SP}]_0 \approx 40$  nM, as in the experiments.<sup>19</sup> Efficient function of GroEL also requires that the binary rate constant for [ELSP] formation exceeds the low-order aggregation rate. Our fits show that  $\lambda_B = 0.6 \text{ nM}^{-1} \text{ min}^{-1} \gg \lambda_A$ , and hence GroEL can effectively rescue stringent substrates.

#### Extracted kinetic parameters are in accord with experiments

We further calibrate our model by comparing the extracted parameters for GroEL allostery with rates that have been obtained by independent measurements. In particular, the transition  $R \rightarrow R''$  is believed to be the rate-limiting step in the GroEL



**Fig. 2.** (a) Native-state yield of Rubisco as a function of time at different chaperonin concentrations. The points are experimental data from Ref. 19. The lines are the fits to the data using the kinetic model. The set of parameters that provide a global fit to all eight data sets is given in Table 1. The initial concentration of Rubisco is 40 nM. The chaperonin concentrations for the curves from bottom to top are as follows:  $[\text{EL}_0] = 1$  nM,  $[\text{EL}_0] = 2$  nM,  $[\text{EL}_0] = 5$  nM,  $[\text{EL}_0] = 10$  nM,  $[\text{EL}_0] = 20$  nM,  $[\text{EL}_0] = 30$  nM,  $[\text{EL}_0] = 50$  nM, and  $[\text{EL}_0] = 100$  nM, corresponding exactly to Ref. 19. (b) Results of sensitivity analysis of the kinetic model [Eq. (12)] using Rubisco parameters (Appendix B). The time-dependent native-state yield  $[N](t)$ , continuous curves, for three chaperonin concentrations,  $[\text{EL}_0] = 1$  nM,  $[\text{EL}_0] = 10$  nM, and  $[\text{EL}_0] = 100$  nM, is compared with the robustness measure  $\Delta[N](t)/[N](t)$ , dashed curves [Eq. (23)]. The small values of  $\delta[N](t)/[N](t)$  for all  $t$  and at various [EL] concentrations suggest that the allosteric model is robust to changes in the kinetic parameters around the values that provide the best fit in (a).

allosteric cycle in the presence of saturating SP. Our fits yield  $k_{R-R'} = 12 \text{ min}^{-1}$  (Table 1) or  $k_{R-R'}^{-1} = 5$  s, which is consistent with the GroE hemicycle time. The load-dependent GroEL allosteric cycle time  $\tau_1$  is usually estimated to be 10–15 s.<sup>40</sup> The value of  $\tau_1$  can be as small as 6 s, depending on the nature of the ligand that is bound to the *trans* ring (J. Gresham, J. Grason, and G.H. Lorimer, unpublished data). The chaperonin binding rate has been measured to be in the range  $\lambda_B \approx 10^7\text{--}10^8 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>9</sup> The binding constant  $\lambda_B = 10^7 \text{ s}^{-1} \text{ M}^{-1}$  obtained from our fits (Table 1) falls within this range. The  $\text{TT} \rightarrow \text{TR}$  transitions have been characterized extensively by Yifrach and Horovitz,<sup>41</sup> who found that, for various GroEL mutants,  $k_{T-R} \sim 2000\text{--}9000 \text{ min}^{-1}$ . The value

**Table 1.** Folding and aggregation rates of Rubisco and allosteric transition rates of GroEL<sup>a</sup>

$k_F$ (min <sup>-1</sup> ) <sup>b</sup>	100
$k_S$ (min <sup>-1</sup> ) <sup>b</sup>	0.003
$\Phi$ <sup>b</sup>	0.02
$\lambda_A$ (nM <sup>-1</sup> min <sup>-1</sup> ) <sup>c</sup>	0.001
$\lambda_B$ (nM <sup>-1</sup> min <sup>-1</sup> ) <sup>d</sup>	0.6
$k_{R-R''}$ (min <sup>-1</sup> ) <sup>d</sup>	12
$k_{T-R}$ (min <sup>-1</sup> ) <sup>d</sup>	5000
$k_{R-T}$ (min <sup>-1</sup> ) <sup>d</sup>	1
$k_{R''-T}$ (min <sup>-1</sup> ) <sup>d</sup>	60
$\Phi^e$	0.02
$\chi^f$	0.37

<sup>a</sup> The theoretical fits and experimental data are compared in Fig. 2.

<sup>b</sup> Defined in Eq. (2).

<sup>c</sup> Extracted aggregation rate.

<sup>d</sup> Parameter characterizing the reaction cycle (Fig. 1).

<sup>e</sup> Probability of folding to the native state with  $k_N = \Phi k_{R''-T}$  and  $k_M = (1 - \Phi) k_{R''-T}$ .

$$^f \chi = \frac{1}{\sqrt{M}} \sqrt{\sum_{c,i}^M \frac{(N^{\text{exp}}(c, i) - N^{\text{model}}(c, t_i))^2}{N^{\text{model}}(c, t_i)}}, \text{ where } M \text{ is the}$$

number of experimental data points.

of  $k_{T-R} = 5000 \text{ min}^{-1}$  (Table 1) is consistent with their measurements. We also found that  $k_{R-T} \ll k_{R-R''}$  to favor the forward reaction from the R state in the presence of ATP and GroES. Incidentally, our fits are not sensitive to the precise values of  $k_{R-T}$  and  $k_{T-R}$  as long as they remain non-rate-limiting in the reaction cycle. The final ejection of ligands is associated with the release of SP (folded or not), ADP, and the transition from the R'' state back to the T state. We found that this transition timescale is on the order of 1 s, which is in accord with expectations.<sup>40</sup>

#### Independent estimate of $\Phi$ agrees with fits to experimental data

According to the IAM, in the limit  $k_F^{-1} \ll \tau_1 \ll k_S^{-1}$ , the native-state yield in the case of superstoichiometric chaperonin concentration can be approximated as:

$$[N](t) = 1 - (1 - \Phi)^n = 1 - \exp(-t/\tau_E) \quad (6)$$

where  $n = t/\tau_1$  is the number of GroEL cycles within the time  $t$ . Alternatively, the yield of the GroEL-assisted folding of the native SP can be fit using an exponential function and thus is characterized by an effective folding timescale  $\tau_E$ . Thus, in this limit, the kinetic partitioning  $\Phi$  can be estimated using:

$$\frac{\tau_E}{\tau_1} = -\frac{1}{\ln(1 - \Phi)} \quad (7)$$

For the data in Fig. 2 at the highest chaperonin concentration, we found  $\tau_E = 6.3 \text{ min}$ . The independent estimate of  $\Phi = 0.02$  is consistent with our fitted value. We should stress that the quality of our fits may not change as long as  $\Phi < 0.05$ , which was the value reported in our earlier study.<sup>19</sup>

#### Coupling to GroEL allostery is required for accurate predictions of Rubisco yield

We should point out that kinetic schemes that completely ignore the GroEL reaction cycle can also be constructed to fit the experimental data on Rubisco (see Methods). The solutions to such a simple kinetic scheme do not provide as good fits to the data as those in Fig. 2 and lead to unphysical values of the scales. More importantly, Eq. (14) and even simpler conceivable schemes cannot account for dramatic efficiency using GroEL mutants that clearly affect the allosteric transition rates.<sup>17,41</sup>

#### Average rate $k_S$ describes transitions from the misfolded state ensemble to the native state

The KPM predicts that there ought to be an ensemble of misfolded conformations such that escape from these kinetic traps should be specified by a collection of rates  $\{k_i\}$  [Eq. (1)]. In deriving the kinetic model, we replaced those by an average rate  $k_S \approx \frac{1}{m} \sum_{i=1}^m k_i$  Eqs. (1) and (2). Mathematically, including separate misfolded species as in Eq. (1) only complicates the solutions (the summation over discrete misfolded species has to be explicitly taken into account) without adding any new insight. More importantly, bulk experiments cannot resolve these states. Here, we further justify using an average  $k_S$  for the misfolded ensemble using experimental data on Rubisco.<sup>19</sup> Figure 4 of Ref. 19 shows the time-dependent Rubisco folding yields after the addition of GroEL at  $t=0$  and after delays of 15 and 30 min. Each data set can be fit using a single exponential, giving an effective folding timescale (data not shown). Since the effective timescale does not vary significantly between the sets, we can conclude that the population of GroEL-binding misfolded species does not change over time, and our assumption that we can use an average  $k_S$  for the misfolded state ensemble, while approximate, is fully justified.

#### Assessing the robustness of the kinetic scheme

The complexity of the reaction cycle (Fig. 1) and the intrinsic folding rates of the SP require, minimally, nine parameters (see Methods). If some of the parameters are constrained to known experimental values, then the number of parameters required to analyze the Rubisco data can be reduced. We reanalyzed the experimental data for assisted Rubisco folding by using known values of the allosteric rates ( $k_{T-R}$ ,  $k_{R-T}$ ,  $k_{R-R''}$ , and  $k_{R''-T}$ ) in the GroEL reaction cycle. In particular, we fixed the values of the ATP-dependent rates  $k_{R-T}$  and  $k_{T-R}$  as well as  $k_{R-R''}$  and  $k_{R''-T}$  at values in the middle of the experimental range. The partition factor,  $\Phi$ , which can be independently determined from experimental data, was set to 0.02. With these values fixed, there are only four Rubisco-dependent parameters, namely,  $k_S$ ,  $k_F$ ,  $\lambda_A$ , and the binding rate  $\lambda_B$ . The best fit obtained by varying these parameters is indis-

tinguishable from the results shown in Fig. 2 (data not shown). The consistency between the results obtained using the full kinetic model in which all parameters are varied to obtain the best fit and the ones obtained by constraining the allosteric rates and  $\Phi$  to experimental values gives additional credence to the proposed kinetic model. It should be emphasized that the proposed kinetic scheme is a general minimal model that not only can be used to analyze available data but also is useful in predicting experimental outcomes, especially in situations where the GroEL concentration is less than the SP concentration (see below).

We also performed sensitivity analysis (see Appendix B for details) to identify the kinetic parameters that most strongly influence the time-dependent changes in the native-state yield and assess their effect on the yield curves in Fig. 2. The changes in the native-state yield to variations in the kinetic parameters around the optimum values (those that produce the fits shown in Fig. 2) depend on GroEL concentration. For all values of  $[EL]_0/[SP]_0$ , the values of  $\delta[N]/\delta k_\alpha$  ( $\alpha = R \rightarrow T$ ,  $T \rightarrow R$ ,  $R \rightarrow R''$ , or  $R'' \rightarrow T$ ) are relatively small. The largest value over the time course ( $0 < t < 60$  min) corresponds to  $\delta[N]/\delta k_{R \rightarrow R''}$ , which lies in the range ( $0 < |\delta[N]/\delta k_{R \rightarrow R''}| < 0.8$ ) nM min. Thus, the native-state yield is not sensitive to small changes in the allosteric rates. The derivatives  $\delta[N]/\delta k_\alpha$  are larger for SP-dependent parameters, especially for the binary aggregation rate  $\lambda_A$ . However, the cumulative effect on the time-dependent changes in the native-state yield is very small (Fig. 2b). We have calculated the time-dependent changes in the native-state yield  $\Delta[N](t)$  [see Eq. (23) in Appendix B] due to variations in the kinetic parameters for three values of  $[EL]$ . At all values of  $[EL]$ ,  $\Delta[N](t)/[N](t) \ll 1$  for all  $t$ . Both the sensitivity analysis and the ability to reproduce the experimental results with merely four parameters for a large data set by using measured allosteric rates in the reaction cycle attest to the robustness of the kinetic model.

### Single-ring mutant (SR1) can reconstitute Rubisco with decreased efficiency

The single-ring mutant of GroEL, called SR1, has been used as a surrogate to probe assisted folding.<sup>18,24,42</sup> The inability to discharge the ligands implies that inter-ring negative cooperativity is suppressed in the GroEL allosteric cycle. The allosteric cycle of SR1 is similar to that of GroEL. Just like GroEL, SR1 can bind SPs and GroES in the presence of ATP and undergo the structural changes shown in Fig. 1. However, without the *trans* ring, and hence a signal from it, SR1 is missing a “timer” and cannot readily discharge the ligands. Several studies have shown that when SR1<sup>24,42</sup> and SR1 mutants<sup>18</sup> are present in excess of SP, they can aid in the folding of stringent substrates. However, the efficiency of SR1 is greatly compromised.<sup>18</sup> Nevertheless, SR1 and other mutants of SR1 have been useful in providing insights into the role confinement effects play in the

SP folding as a result of the expanded cavity in the  $R''$  state. Our theoretical model can be used to explore folding in SR1 and assess the extent to which yield enhancement occurs when the SP is sequestered in the central cavity for a long time. It should be borne in mind that even in SR1 the microenvironment of SP is altered from hydrophobic to hydrophilic as a consequence of  $T \rightarrow R \rightarrow R''$  allosteric transitions. In terms of IAM, cycling in SR1 corresponds to  $n=1$ , and  $k_{R'' \rightarrow T} \approx 0$  (Fig. 1).

In terms of the kinetic model, the SR1 mutant undergoes the complete allosteric cycle [Eq. (11)], but the  $R'' \rightarrow T$  transition, characterized by  $k_{R'' \rightarrow T}$ , has been prolonged to approximately 300 min.<sup>24</sup> Thus, during the  $R \rightarrow R'' \rightarrow T$  transitions, the SP is encapsulated within the GroEL cavity, where it can fold in the restricted space.<sup>42</sup> Rates of SP folding in a hydrophilic cavity are typically greater than those in the bulk,<sup>43,44,45</sup> assuming that the interaction between SP and the walls of cavity is repulsive. Even in the most favorable case, the confinement-induced folding rate remains within 1 order of magnitude of the bulk rates. Because these changes are much smaller than the lifetime of the SR1 mutant, we assume that folding inside the cavity is not very different from folding in the exterior, except that aggregation is suppressed. The consequence of changes in SP folding in confined space is explored later. With this assumption, after time  $\tau_c = k_{R \rightarrow R''}^{-1}$ , we found that the fractions of misfolded and folded concentrations are as follows:

$$[M](k_{R \rightarrow R''}^{-1}) = \frac{(1 - \bar{\Phi})\bar{k}_F}{k_F - k_S} \left( e^{-\bar{k}_S k_{R \rightarrow R''}^{-1}} - e^{-\bar{k}_F k_{R \rightarrow R''}^{-1}} \right) \alpha k_M$$

$$[N](k_{R \rightarrow R''}^{-1}) = 1 - \frac{\bar{\Phi}\bar{k}_F - \bar{k}_S}{k_F - k_S} e^{-\bar{k}_F k_{R \rightarrow R''}^{-1}} - \frac{(1 - \bar{\Phi})\bar{k}_F}{k_F - k_S} e^{-\bar{k}_S k_{R \rightarrow R''}^{-1}} \alpha k_N \quad (8)$$

We have used the notation  $\bar{k}_i$ ,  $\bar{\Phi}$  to emphasize that the interior (in the cavity) and bulk rates may differ. We have also assumed that SP folding inside the cavity follows the KPM.

In almost all cases, 300 min is sufficient for SPs to fold inside the cavity, and thus mostly folded or assembly competent SPs would be ejected from it. For instance, for Rubisco, even under the assumption that there is no effect of confinement in the hydrophilic cavity ( $\bar{k}_i = k_i$ ,  $\bar{\Phi} = \Phi$ ), 300 min is sufficient for 61% of the SPs to fold using the parameters in Table 1. A modest enhancement of the slow folding timescale upon confinement would increase the partitioning to the folded states even more. Thus, our kinetic model predicts that SPs can fold in the central cavity as in the SR1 mutant. However, from Eq. (8), it follows that even in this case, the yield of the folded states depends not only on the intrinsic folding characteristics of the SP ( $k_F$ ,  $k_S$ , and  $\Phi$ , and their values upon encapsulation) but also on  $k_{R \rightarrow R''}$  (i.e., SP folding and GroEL allostery are coupled). More importantly, the efficiency of SR1 in producing Rubisco in the native state, even after 300 min,  $\sim 10$  times the cell doubling time, is far less



than that of the WT GroEL. The efficiency of SR1 mutants is linked to a decrease in  $k_{R''-T}$  (see below).

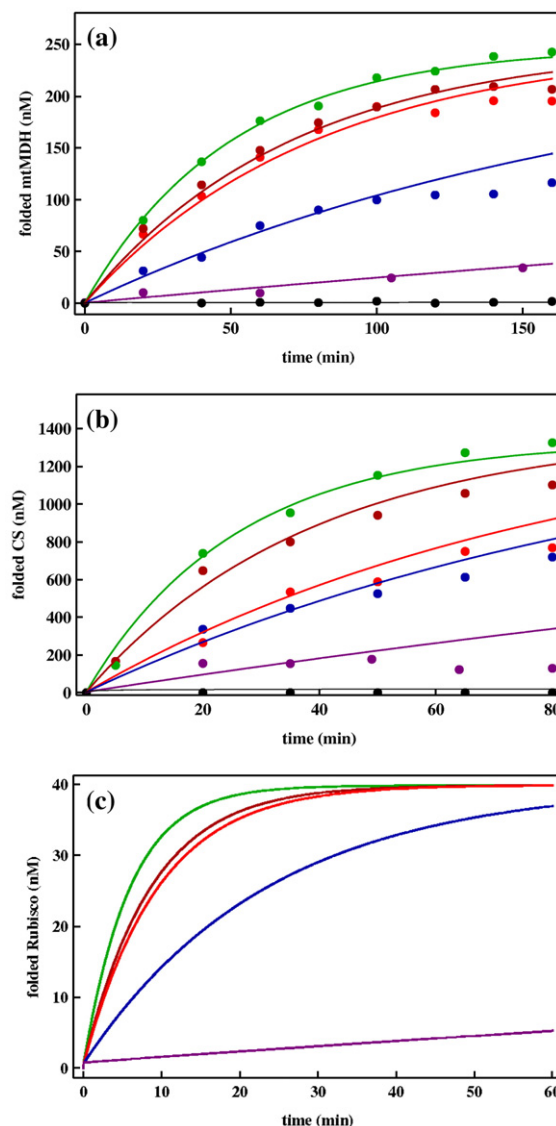
### Efficiency of SR1 and its variants in folding mtMDH and CS depends on intrinsic GroEL allosteric rates

In the quest to find single-ring constructs that can substitute for GroEL *in vivo*, Sun *et al.* used genetic experiments to screen for SR1 mutants from which GroES can disassociate.<sup>18</sup> Two of the three SR1 mutants, namely, SR-D115N and SR-A399T, were judged to be nearly as efficient as GroEL<sup>18</sup> in that they facilitated folding of mtMDH and CS almost as well as did GroEL. They could also function well *in vivo* at 37 °C. The third one, SR-T522I, while considerably more efficient than SR1, was unable to sustain growth of *E. coli*. The experiments were performed under non-permissive conditions such that spontaneous folding of mtMDH and CS was negligible (see Figs. 5 and 6 of Ref. 18). The time-dependent changes in the activity of mtMDH, which presumably mirrors the yield of the folded state, in the presence of GroEL, SR1, SR-D115N, SR-A399T, and SR-T522I are reported in Fig. 5 of Ref. 18.

#### mtMDH

We fitted the experimental data using our kinetic model, with the same GroEL parameters (varying only  $k_{R''-T}$  for different mutants; see below) as we obtained in the analysis of Rubisco (Table 1) and treating mtMDH folding within the single-ring cavity using Eq. (8). In particular, we analyzed the data in Fig. 5 of Ref. 18 in two steps. First, we used the data of spontaneous folding and GroEL-assisted folding from the figure to determine the intrinsic rates of mtMDH folding. While GroEL allosteric rates are dependent on the SP and on the specific experimental conditions, we used the rates  $\lambda_B$ ,  $k_{R-R'}$ ,  $k_{T-R}$ ,  $k_{R-T}$ , and  $k_{R''-T}$  from Table 1 for the WT GroEL. Thus, for mtMDH, the intrinsic folding rates are as follows:  $k_F=100 \text{ min}^{-1}$ ,  $k_S=0.0012 \text{ min}^{-1}$ ,  $\Phi=0.002$ , and  $\lambda_A=0.021 \text{ nM}^{-1} \text{ min}^{-1}$ . Consistent with these rates, the values for  $k_N$  and  $k_M$  [ $k_N=\Phi k_{R''-T}$  and  $k_M=(1-\Phi)k_{R''-T}$ ] for GroEL are as follows:  $\Phi=0.002$  and  $k_{R''-T}=60 \text{ min}^{-1}$ . For SR1, we used the same  $\lambda_B$ ,  $k_{R-R'}$ ,  $k_{T-R}$ , and  $k_{R-T}$  values; however,  $k_{R''-T}=1/300 \text{ min}^{-1}$ . The corresponding data and curves are shown in Fig. 3.

Second, we fitted the yield data for mtMDH for the three SR1 mutants (SR-D115N, SR-A399T, and SR-T522I), using only one free parameter each, namely, the  $R'' \rightarrow T$  transition rate,  $k_{R''-T}$ , which reflects the kinetic efficiency of inter-ring communication. For the mutants SR-T522I, SR-A399T, and SR-D115N, the fits yielded  $k_{R''-T}=2.5 \text{ min}^{-1}$ ,  $k_{R''-T}=11 \text{ min}^{-1}$ , and  $k_{R''-T}=14 \text{ min}^{-1}$ , respectively. The fits are shown in Fig. 3. While it is likely that other SR1 kinetic rates were also altered as a result of these mutations, we chose the  $k_{R''-T}$  transition rate because it is most closely related to the GroES dissociation rate (Table 5 of Ref. 18). The GroES  $K_d$  measurements for these



**Fig. 3.** Yield of SPs as a function of time. (a) The points are taken from the experimental measurements<sup>18</sup> for folding mtMDH. The lines are fits to the data using the kinetic model. The black line is for spontaneous folding. Assisted folding in the presence of GroES and SR1 (purple), SR-T522I (blue), SR-A399T (red), and SR-D115N (dark red) was used to assess the efficiencies of these three single-ring chaperonins relative to GroEL (green). (b) The same as (a) except that the folding of CS instead of mtMDH is analyzed. The only GroEL allosteric rate that is varied in obtaining the results in (a) and (b) is  $k_{R''-T}$ , while all others were taken from Table 1. (c) Predictions for the time-dependent yield of Rubisco for GroEL (green) and the single-ring mutants SR-T522I (blue), SR-A399T (red), SR-D115N (dark red), and SR1 (purple) based on chaperonin allosteric rates. The values of  $k_{R''-T}$  were taken by analyzing the mtMDH data in (a). In (a) and (b), superstoichiometric concentrations of chaperonins were used.

mutants (Table 5 of Ref. 18) indicate an increased dissociation rate for the mutants compared with SR1. A larger GroES dissociation rate would correspond to speeding of the  $R'' \rightarrow T$  transition rate



compared with SR1. Thus, our kinetic model, with physically reasonable parameters, accounts for folding in a number of single-ring mutants of GroEL. Increased efficiencies of GroEL, SR1 mutants, and SR1 correlate with decreasing values of  $k_{R''-T}$ ; that is, the values of  $k_{R''-T}$  can be arranged in the order GroEL > SR-D115N > SR-A399T > SR-T522I > SR1.

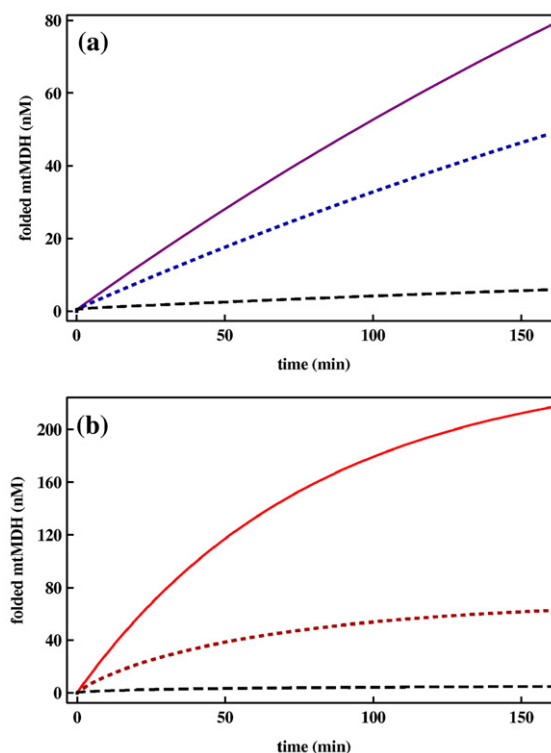
## CS

Sun *et al.* also reported the efficiency of the same GroEL mutants (SR-D115N, SR-A399T, and SR-T522I) in assisted folding of CS (see Fig. 6 of Ref. 18). We analyzed the CS data using exactly the same procedure described for mtMDH. As before, for  $k_{R-T}$ ,  $k_{T-R}$ ,  $k_{R-R''}$ ,  $k_{R''-T}$ , and the binding rate  $\lambda_B$ , we used the values from Table 1. We fitted the GroEL-assisted and spontaneous folding data and obtained the CS intrinsic folding rates ( $k_S = 0.01 \text{ min}^{-1}$ ,  $k_F = 100 \text{ min}^{-1}$ ,  $\Phi = 0.003$ , and  $\lambda_A = 0.01 \text{ nM}^{-1} \text{ min}^{-1}$ ). We then fitted the data set for each mutant by varying only one reaction rate,  $k_{R''-T}$ . For CS, the effective rates associated with the transition from the R'' state to the T state were as follows:  $10 \text{ min}^{-1}$  for SR-D115N,  $0.8 \text{ min}^{-1}$  for SR-A399T, and  $0.1 \text{ min}^{-1}$  for SR-T522I. For SR1, we used  $k_{R''-T} = 1/300 \text{ min}^{-1}$ , as before. The fits are shown in Fig. 3b. While the precise values of the allosteric rate constant depend on the SP, the trend in  $k_{R''-T}$  using mtMDH and CS, namely, GroEL > SR-D115N > SR-A399T > SR-T522I > SR1, is the same.

The experiments with SR1 are generally performed with superstoichiometric chaperonin concentrations. The panels in Fig. 4 show plots of predicted mtMDH yields with super- and substoichiometric concentrations of SR1 (Fig. 4a) and SR-A399T (Fig. 4b). When present in excess molar concentrations ( $[EL]_0/[SP]_0 > 1$ ), SR1 and SR-A399T can prevent aggregation as they isolate misfolded SPs (data not shown). However, experiments<sup>18</sup> and the theoretical analysis here show that fast cycling rates are required for efficient chaperonin-aided SP folding, especially in *in vivo* situations, where, typically,  $[EL]_0/[SP]_0 < 1$ .

## SR1 is least efficient in the folding of Rubisco

Normalized activity of Rubisco in the presence of SR1 has been previously reported.<sup>42</sup> However, to our knowledge, time-dependent increases in the absolute yield (or activity) for SR1-assisted folding of Rubisco do not exist in the literature. Using the allosteric transition rates for SR1 and SR1 mutants (SR-D115N, SR-A399T, and SR-T522I), we can predict the trend in the yield as a function of time for single ring-assisted folding of Rubisco using the kinetic model. With the initial concentrations of 100 nM for the chaperonins and 40 nM for unfolded Rubisco, we found that SR1 is the least efficient in folding this stringent SP (Fig. 3c). Numerically, we used the  $k_{R''-T}$  rates from mtMDH data fits. Just as with refolding mtMDH and CS, the SR-D115N can



**Fig. 4.** (a) Theoretical predictions for the expected yield of mtMDH at various concentrations of SR1. The concentrations of SR1 are 1500 nM (continuous purple line), 150 nM (dotted blue line), and 15 nM (dashed black line), where the initial unfolded mtMDH concentration is 300 nM. The allosteric transition rates for SR1 are obtained from the theoretical fits in Fig. 3a. (b) Same as (a) except for the expected yield of mtMDH at various concentrations of SR1 mutant SR-A399T instead of SR1. (The continuous red line corresponds to 1500 nM SR-A399T, the dotted red line corresponds to 150 nM SR-A399T, and the dashed black line corresponds to 15 nM SR-A399T.) The predictions are based on the predicted allosteric transition rate  $k_{R''-T} = 11 \text{ min}^{-1}$ , obtained from fits to Fig. 3a.

lead to yields that are similar to those obtained using GroEL. The yield of folded Rubisco at long times, in the presence of SR-A399T, is somewhat less than that obtained using GroEL or SR-D115N. These predictions also reinforce that easy and frequent discharge of the SP from the central cavity is the most efficient way of enhancing chaperonin-assisted yield in finite (less than cell doubling) time.

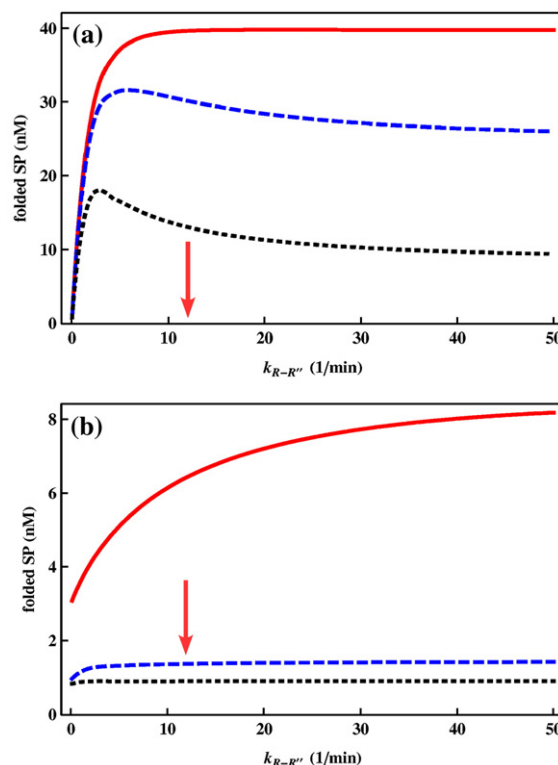
## Rate of $k_{R-R'}$ of substrate-optimized chaperonin is greater than WT GroEL

It is well established that the WT GroEL machinery can assist in the folding of a variety of proteins that do not share a sequence or structural similarity. The ability of GroEL to function as a general chaperone for a range of substrates might render it inefficient for a given SP. To probe the basis of the conflict between the generality and specificity for selected SPs, Wang *et al.*<sup>17</sup> performed rounds of *in vivo* and *in vitro* directed evolution to identify

GroEL/S mutants that can reconstitute a particular SP, namely, green fluorescent protein (GFP), more efficiently than WT GroEL (see Fig. 2B of Ref. 17). However, the enhanced efficiency of GFP comes at the expense of GroE<sub>3-1</sub> (the mutant that refolds GFP more efficiently than WT) not being able to fold other natural GroEL SPs as well. It was argued that efficient folding of GFP by GroE<sub>3-1</sub> is related to changes in the ATPase activity in GroE<sub>3-1</sub> compared with the WT and enhanced hydrophilic character of the GroE<sub>3-1</sub> cavity.<sup>17</sup> Thus, mutation-induced alterations in the kinetics of the GroEL reaction cycle can be used to increase the yield for a particular SP at the expense of others. The specific residues that were altered in the directed evolution experiments resulting in GroE<sub>3-1</sub> were located in the equatorial (D490G) and intermediate (A163V) domains near a hinge region that is critical to the chaperonin allosteric transitions. GroES-Y71H, the accompanying co-chaperonin, had a single mutated residue that putatively changed the polarity of the chaperonin cavity.

While it is difficult to ascertain the specific allosteric kinetic rates in the reaction cycle and the extent to which they were altered in GroE<sub>3-1</sub>, our model can explain how changes in the rates may affect the folding efficiency of a specific SP in a positive way while hindering the folding of others. For instance, based on the location of the GroEL mutated residues in GroE<sub>3-1</sub>, we propose that the allosteric transition-determining rates,  $k_{R-R'}$  in particular, are altered. In addition, we can presume that the associated mutation in GroES changed the folding rates and/or the kinetic partitioning of GFP inside the cavity. With the assumption that  $k_{R-R'}$  is altered in the GroE<sub>3-1</sub> chaperonin, we can explore, using the kinetic model, whether the optimal value of the rate for one SP (say, GFP) compromises efficiency for others. Thus, a given SP ( $k_F$  and  $k_S$  are fixed) chaperonin efficiency is determined by  $k_{R-R'}$  and  $\lambda_A$  (binary aggregation rate), assuming that all other parameters do not change dramatically with GroE<sub>3-1</sub>. The optimum value of  $k_{R-R'}$  should arise only if the SP, under the experimental conditions, is aggregation prone (see below).

Typically, enhancement in  $k_{R-R'}$  increases the yield of folded SPs. The continuous red line in Fig. 5 is an example of the dependence of the native-state yield after 30 min on an SP (we used the Rubisco parameters in Table 1) using superstoichiometric chaperonin concentrations (100 nM GroEL and 40 nM Rubisco). From this result, we surmise that optimizing the efficiency of the chaperonin cannot be achieved unless  $\lambda_A$  (or other rates) is also changed. For instance, the dashed curve in Fig. 5 shows the dependence of the native-state yield (also at 30 min) on  $k_{R-R'}$  for a highly aggregation-prone protein (we used  $\lambda_A = 0.1 \text{ nM}^{-1} \text{ min}^{-1}$ ) with superstoichiometric chaperonin concentrations. In this case, it is better to prolong the time the protein spends in the cavity in order to reduce the amount of misfolded species in the bulk (Fig. 5b). The presence of a maximum in the yield curve in Fig. 5, indicating an optimum time-



**Fig. 5.** Prediction of the effect of  $k_{R-R'}$  on the yield of folded SPs over cell doubling time (30 min). (a) The red line, obtained using Eqs. (8) and (12), is for the expected yield of Rubisco as  $k_{R-R'}$  is varied. All other parameters were taken from Table 1. Dashed blue curve shows yield for a protein whose aggregation rate  $\lambda_A = 0.1 \text{ nM}^{-1} \text{ min}^{-1}$ , which is 100-fold greater than that of Rubisco. Dotted black curve shows yield for a protein whose aggregation rate  $\lambda_A = 1 \text{ nM}^{-1} \text{ min}^{-1}$ , which is a thousand times that of Rubisco. The concentration of chaperonin  $[EL_0] = 100 \text{ nM}$ , which exceeds the initially unfolded concentration of SP (40 nM). (b) Same as (a) except that  $[EL_0] = 1 \text{ nM}$ . In this case, even rapid cycling does not significantly yield a rapidly aggregating protein. However, the paucity of chaperonins makes it best to increase the cycling rate. The arrows give the value of  $k_{R-R'}$  for WT GroEL (see Table 1).

scale for the chaperonin allosteric transition, can be understood as a competition between two effects. On one hand, a small  $k_{R-R'}$  would result in most SPs remaining captive within GroEL and not being ejected quickly enough for sufficient yield. On the other hand, if  $k_{R-R'}$  is large, then most SPs are released in a misfolded state and hence would aggregate because  $\lambda_A$  is large. Our calculation (dashed curves in Fig. 5) demonstrates the plausibility of an optimum value of  $k_{R-R'}$  and is consistent with the experiments of Wang *et al.*<sup>17</sup> These calculations show that optimized chaperonin activity depends not only on the timescales in the reaction cycle of GroEL but also on the aggregation and folding characteristics of the SP. It is important to emphasize that for proteins that are only moderately aggregation prone, the yield increases as the cycling rate increases. The experimental observation that the folding of other SPs was not as efficient with GroE<sub>3-1</sub> is also

consistent with this observation (continuous red curve in Fig. 5).

The results in Fig. 5a were obtained using the concentration of chaperonins in excess of the SP. We predict in Fig. 5b the yield for the same SPs at the same time in the case of substoichiometric chaperonin concentration (we used 1 nM chaperonin). The presence of an additional competing factor—the availability of chaperonins—favors faster chaperonin turnover; hence, we see that even in the case of a highly aggregation-prone SP, it is better to increase the chaperonin cycling rate. It is expected that the amount of chaperonins is rather limited in *in vivo* situations. Therefore, we predict that even for aggregation-prone SPs, the most efficient chaperonin function *in vivo* requires fast cycling.

## Discussion

The interplay between the ligand-dependent (ATP and SP) timescales describing the GroEL allosteric transition rates in the reaction cycle (Fig. 1) and those that determine the fate of SPs determines the efficiency of chaperonins. Our work shows that the kinetics of assisted folding of SPs, under non-permissive conditions, can be used to accurately describe a number of experiments, provided the coupling between allosteric transitions and SP folding and aggregation is taken into account. The dependence of Rubisco folding on the EL concentration can be quantitatively obtained only by including the coupling between the kinetic processes in the reaction cycle. More importantly, using our kinetic model, we can also account for SP folding assisted by SR1 and SR1 mutants and link the efficiency of these chaperonins to the kinetic rates in the reaction cycle. Finally, our model also explains naturally, in terms of the rates of the allosteric transitions that depend on the nature and concentrations of ligands and the tendency of SPs to aggregate, the observation that enhanced efficiency of GroEL mutants for a particular SP comes at the expense of their inability to rescue other SPs. Although the full kinetic model is complicated [Eq. (12)], the variations in the efficiency of the chaperonins are, to a large extent, determined by the rate-limiting step in the reaction cycle (it is typically  $k_{R-R'}$  under maximal loading conditions), the capture rate of the SP, and the concentration-dependent rate of SP aggregation. Most of the readily available yield data can be quantitatively explained primarily by varying these parameters. The availability of the complete theoretical model allows us to explore other possibilities that can be experimentally realized through directed evolution or targeted mutation of GroEL and GroES.

### Chaperonin concentration and binding rate of SP determine GroEL efficiency

Misfolded proteins either fold to the native state or aggregate. Chaperonins provide an additional pathway out of the misfolded states. Only if the flux

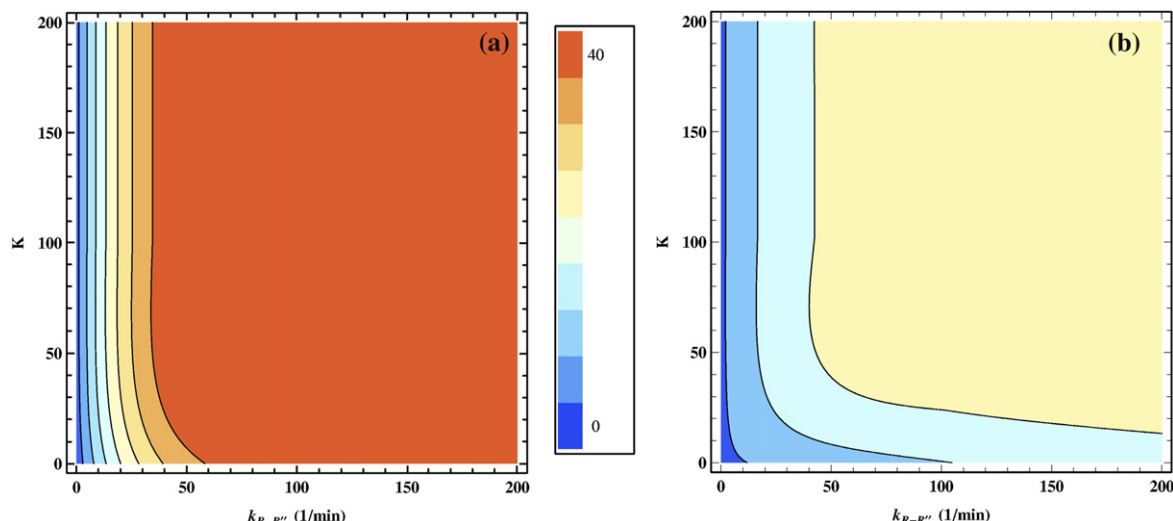
along that path is significant can chaperonins substantially reduce aggregation. Thus, quantitatively, chaperonins can measurably reduce aggregation only if  $\lambda_B C_0 > \lambda_A [X_0]$ . In other words, the product of chaperonin concentration and binding rate determines how effectively chaperonins can isolate the misfolded SPs.  $\lambda_B C_0 > \lambda_A [X_0]$  is a necessary kinetic criterion for efficient function of chaperonins.

Given a sufficient amount of chaperonins, the binding rate  $\lambda_B$  has a large effect on aggregation prevention illustrated by steady-state solutions to the kinetic equations. We expect that binding to the misfolded proteins, and hence isolating them from the aggregation-prone environment, is beneficial. In most cases, increasing  $\lambda_B$  or the concentration also increases the yield. In Fig. 2, we show the native-state yield increase as the chaperonin concentration increases. Theoretically, however, it is also possible to find a regime where increasing the chaperonin concentration or increasing  $\lambda_B$  may not necessarily increase the native-state yield. Such a possibility is realized if the chaperonin allosteric cycle is very slow, slower than the characteristic timescale  $k_S^{-1}$  associated with the protein escaping from the misfolded conformations spontaneously. The effect of chaperonin concentration on yield and aggregation is larger in the substoichiometric cases (i.e., if  $C_0 < [X_0]$ ). If  $C_0 > [X_0]$ , then GroEL–GroES can fold all SPs (Fig. 2).

### The T ↔ R equilibrium constant has only a moderate effect on SP folding yield in the presence of excess ATP

A consequence of IAM<sup>43</sup> is that it is generally beneficial to speed the chaperonin allosteric cycle, such that, when needed, chaperonin can perform multiple rounds of binding and release. This can result in a rapid enhancement in the yield of native-state SP [Eq. (6)]. In the context of the T → R transition, this means that, upon SP binding, driving the equilibrium toward the R state should enhance efficiency of SP folding. Yifrach and Horovitz showed (see Fig. 1 of Ref. 41) how the rate of refolding of mDHFR increases as the TT → TR conformational transition is sped. Using our kinetic model, we have analytically shown [Eq. (5)] in the steady-state limit that the native-state yield can be increased (and aggregation reduced) upon increasing  $k_{T-R}$  and reducing  $k_{R-T}$ . For WT GroEL, under non-permissive conditions in the presence of saturating ATP and GroES, the T → R transition does not appear to be the rate-determining step in the catalytic cycle. We have varied both the equilibrium constants  $K = k_{T-R}/k_{R-T}$  and  $k_{R-R'}$  to predict the changes in the yield of the folded SP (Fig. 6). Under superstoichiometric conditions, even large changes in  $K$  do not have a significant effect on the SP native-state yield (Fig. 6a). Even when  $C_0 < [X_0]$ , alterations in SP yield are relatively small, except perhaps when  $K$  and  $k_{R-R'}$  are greatly changed (Fig. 6b). It is unclear if this situation can be easily realized. Thus, we predict that changes in  $K$  do not significantly affect GroEL





**Fig. 6.** Theoretical predictions on yield of the native SP for Rubisco. (a) The extent of SP folding as a function of the equilibrium constant  $K$  in the  $T \leftrightarrow R$  transition and that in  $k_{R-R''}$  are varied. The GroEL concentration is 100 nM. (b) Same as (a) except that the concentration of GroEL is lowered to 2 nM. In both (a) and (b), the GroEL parameters, except those that are varied, were taken from Table 1. The scale reporting the yield of folded Rubisco, whose initial unfolded concentration is 40 nM, is given in the middle.

efficiency, at least for stringent SPs. Even for non-stringent substrates that do not require GroES, the SP folding rate changes by only a small factor even as  $K$  varies greatly. These conclusions are only valid when the ATP concentration is large such that the  $T \leftrightarrow R$  occurs readily. If the ATP concentration is very low or GroEL mutations slow the first steps in the reaction cycle, then it is likely that the efficiency of GroEL can also be greatly reduced.

### R $\rightarrow$ R'' transition rates and confinement-induced changes in SP folding dramatically alter native-state yield

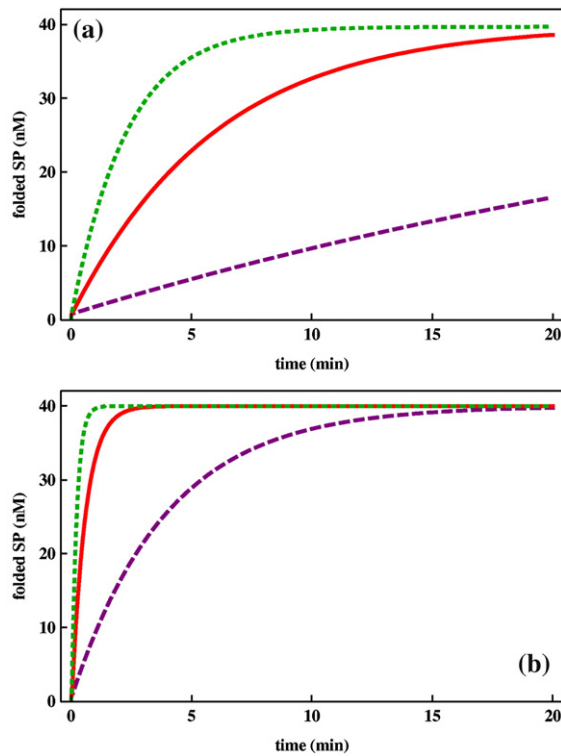
The efficiency of chaperonin function changes dramatically as  $k_{R-R''}$ , which includes GroES binding, SP encapsulation, and ATP hydrolysis, is varied. During the events that occur on the timescale  $k_{R-R''}^{-1}$ , the characteristics of SP folding can also be different from those in the bulk. In particular, as discussed in the context of SR1, it is possible that  $k_N$  and  $k_M$  are functions of  $k_{R-R''}$ . It is of interest to explore the efficiency of assisted folding as SP-dependent rates and  $k_{R-R''}$  are simultaneously altered. It was argued,<sup>5</sup> using the IAM, that an SP is pulled out of a misfolded state during the  $T \rightarrow R$  transition. The stretching occurs because ATP favors the R state, while SP, which has high affinity for the T state, resists the  $T \rightarrow R$  transition. During the  $R \rightarrow R''$  transition, SP is encapsulated in the central cavity, where it can refold on a timescale  $\sim k_{R-R''}^{-1}$ . At the end of the  $R \rightarrow R''$  transition, the concentrations of SP in various states to be released from the cavity are given by Eq. (8). In general, the folding rates  $k_S$  and  $k_F$  and kinetic partitioning factor  $\Phi$  inside the cavity can be different from the bulk values. Detailed computations show that rates of folding inside a cavity change at best by a factor of 10.<sup>46,47</sup> Using the kinetic model, we ex-

plored two possibilities (Fig. 7). First, assuming that the folding rates inside the cavity are the same as those outside, we plotted the native-state yield as a function of time for three  $k_{R-R''}$  rates in Fig. 7a. The yield increases as  $k_{R-R''}$  increases. Thus, even in this case where the rates are not altered by confinement, it is beneficial to speed the chaperonin allosteric cycle. For comparison, we also plotted, in Fig. 7b, the native-state yield in the case where the folding rates inside the cavity are different. It has been shown that confinement and the hydrophobicity of the cavity walls<sup>43</sup> affect folding rates, probably by favoring the native states. We used  $\bar{k}_F \approx k_F$ ,  $\bar{\Phi} \approx 10\Phi$ , and  $\bar{k}_S = 10 * k_S$ , so the population of the misfolded states is somewhat suppressed. We can see from Fig. 7b that chaperonins increase the native-state yield even more. Benefits of confinement-induced changes in SP and multiple iterations have a cooperative effect on SP folding. However, regardless of the changes in the intrinsic confinement-induced folding characteristic of the SP yield, enhancement is only possible if  $k_{R-R''}$  increases. This central result, obtained in experiments (Figs. 2 and 3a) and in the exploration of the complete kinetic model [Eq. (12)], shows clearly that when presented with SP, ATP, and GroES, efficient function of GroEL requires frequent cycling—a prediction that follows from the IAM.<sup>5,19</sup>

## Methods

### Allosteric kinetic model

In order to derive the kinetic model, we have to translate the steps in the catalytic cycle (Fig. 1) into a set of kinetic equations. In addition to the timescales that characterize the KPM for the SP ( $k_F$ ,  $\{k_i\}$ ) and the GroEL allostery ( $k_{T \rightarrow R}$ ,  $k_{R \rightarrow T}$ , and  $k_{R-R''}$ ), we have to include the rates of binding



**Fig. 7.** Effect of variations of  $k_{R-R''}$  on the native yield of Rubisco. The concentration of unfolded SP is 40 nM, and that of GroEL is 100 nM. (a) The folding characteristics inside the central cavity are the same as those in the bulk (for values, see Table 1). The values of  $k_{R-R''}$  are 1.2 min<sup>-1</sup> (dotted green line), 12 min<sup>-1</sup> (continuous red line), and 120 min<sup>-1</sup> (dashed purple line). (b) Same as in (a) except that the enhancement in confinement-induced folding of SP is taken into account. Specifically, we used the same value of  $k_F$  and  $k_S=10k_S$ , and  $\Phi=10\Phi$ . In both cases, frequent cycling, as predicted by IAM, greatly enhanced the yield of folded Rubisco. The combined effect of multiple iterations and favorable values of folding in a cavity results in more rapid production of the native state.

and release of SP (linked to the  $R'' \rightarrow T$  transition) for a complete description of the kinetic model. It is possible to solve the kinetic equations assuming that there are distinct structures in the misfolded ensemble. In this case, the kinetic processes associated with the misfolded ensemble would involve summations over the discrete structures, like in Eq. (1). Experimental resolution of the various misfolded ensembles is difficult if not impossible, and treating the discrete misfolded structures explicitly does not provide new insights into the GroEL function. As a result, we assume that the transition to the NBA from the ensemble of misfolded conformations is given by an average rate  $k_S$ . Because of the commonly made assumption that the  $R'$  state [GroEL-(ATP)<sub>7</sub>-GroES] and the  $R''$  state [GroEL-(ADP)<sub>7</sub>-GroES] are not that structurally different, we did not consider the  $R'$  state explicitly. Even with these physically motivated simplifications, there are 10 timescales that can vary depending on the SP and binding conditions (concentrations of SP, ADP, and GroEL). From general considerations, we can anticipate that the dynamic range of possibilities for assisted folding is large. As a consequence, a number of scenarios can emerge depending on the values of the rate constants.

In the absence of GroEL and under non-permissive conditions, a small fraction of unfolded molecules folds directly to the native state and the remaining molecules misfold into a long-lived metastable state (i.e., SP folding is described by the KPM).<sup>28</sup> The misfolded molecules can also aggregate. Using  $X$ ,  $N$ ,  $M$ , and  $M_A$  to denote the unfolded conformations, native conformations, ensemble of misfolded conformations, and the aggregates, respectively, the relevant reactions are as follows:

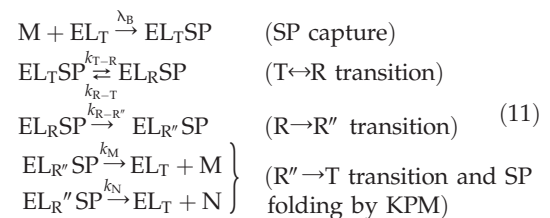


We have assumed that along the fast track ( $X \rightarrow N$ ), specific collapse and folding occur nearly simultaneously. In addition, the commitment to partitioning, which also results in  $M$ , occurs on timescale  $k_F$  (i.e., timescale for non-specific collapse  $\sim k_F$ ). Explicit simulations using coarse-grained models<sup>48</sup> and theoretical considerations lend support to these assumptions. For simplicity, we have also ignored the higher-order aggregation reactions, such as  $M_2 + M \rightarrow M_3$ . For our purposes, it suffices to include aggregation process as a pathway that prevents the formation of the native structure. The inclusion of higher-order aggregation reactions does not alter the conclusions relating to GroEL. First-order rate constants are denoted by  $k_\alpha$  ( $\alpha = S, F, \dots$ ); second-order rate constants, by  $\lambda_\alpha$ . The kinetic equations for the reaction scheme in Eq. (9) are as follows:

$$\begin{aligned} \frac{d[X]}{dt} &= -k_F[X] \\ \frac{d[M]}{dt} &= (1 - \Phi)k_F[X] - k_S[M] - \lambda_A[M]^2 \\ \frac{d[N]}{dt} &= k_S[M] + \Phi k_F[X] \\ \frac{d[M_A]}{dt} &= \lambda_A[M]^2 \end{aligned} \quad (10)$$

The case  $\lambda_A=0$ ,  $k_F \gg k_S$  describes SP folding by the KPM for which  $P_u(t)$  is given by Eq. (2). Throughout the article, we assume that a single rate constant ( $k_S$ ) describes the transition from the  $M$  state to the  $N$  state.

In the presence of GroEL, additional reactions that reflect the coupling between allosteric transitions and the fate of SP have to be taken into account. The minimal set of GroEL allosteric state-dependent reactions that account for the major steps in the reaction cycle in Fig. 1 is as follows:



where EL stands for GroEL, the subscript on EL indicates its allosteric state, and an SP in an unknown state is denoted by SP. Note:  $k_N$ , which is the rate for reaching the native state either in the cavity or upon release from the *cis* cavity, is assumed to be different from  $k_F$ .

In our model, EL can bind misfolded proteins [the first step in Eq. (11)]; it then undergoes  $T \rightarrow R \rightarrow R''$  allosteric transitions (the next two reactions). The SPs that are ejected

from the chaperonin cavity are either misfolded (M) or folded (N). We have assumed that while the chaperonin may unfold the SP during the initial steps of the allosteric cycle, the collapse from the unfolded state X into M or N is fast (commitment to partitioning occurs rapidly), and, thus, even if there are unfolded proteins ejected from the cavity, they will partition into M and N quickly.

The kinetic equations that describe assisted folding and include SP folding by KPM, GroEL allostery, and SP aggregation are as follows:

$$\begin{aligned}
 \frac{d[X]}{dt} &= -k_F[X] \\
 \frac{d[M]}{dt} &= (1 - \Phi)k_F[X] + k_M[EL_R^*SP] - k_S[M] \\
 &\quad - \lambda_B[EL_T][M] - \lambda_A[M]^2 \\
 \frac{d[EL_TSP]}{dt} &= \lambda_B[EL_T][M] + k_{R-T}[EL_RSP] - k_{T-R}[EL_TSP] \\
 \frac{d[EL_RSP]}{dt} &= k_{T-R}[EL_TSP] - k_{R-T}[EL_RSP] - k_{R-R^*}[EL_RSP] \\
 \frac{d[EL_R^*SP]}{dt} &= k_{R-R^*}[EL_RSP] - (k_N + k_M)[EL_R^*SP] \\
 \frac{d[N]}{dt} &= k_S[M] + \Phi k_F[X] + k_N[EL_R^*SP] \\
 \frac{d[M_A]}{dt} &= \lambda_A[M]^2
 \end{aligned} \tag{12}$$

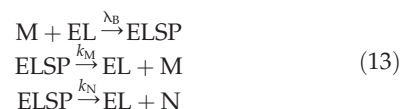
With the use of the conservation of the number of EL particles,  $[EL_0] = [EL_T] + [EL_TX] + [EL_RX] + [EL_R^*X]$ , where  $[EL_0]$  is the initial (and total) chaperonin concentration, the coupled non-linear differential equations in Eq. (12) need to be solved consistently to obtain time-dependent concentrations of the different molecular species. There are, in principle, 10 independent reaction rates needed to solve the seven coupled differential equations. The interplay between these rates creates a number of possibilities, several of which may not be always experimentally realizable or biologically relevant. Physical arguments show that  $\lambda_B/\lambda_A \gg 1$  to prevent aggregation. In other words, the rate of SP recognition must exceed the binary SP aggregation rates. Although chaperones belonging to the other heat shock protein family (e.g., HSP70) may break up aggregates, there is no evidence that GroEL can facilitate such process. Even with physically motivated values for the rate constants, there are a number of factors that determine GroEL efficiency.

### Kinetic model without allostery

Even though the kinetic model that takes into account coupling between the kinetic steps in the well-established reaction cycle and SP folding quantitatively fits the experimental data for Rubisco and mtMDH (Figs. 2 and 3), it is legitimate to wonder if a reduced description suffices. Before we consider models that ignore GroEL allostery, it is important to state the reasons for deriving a complete kinetic model for GroEL-assisted folding. The multi-timescale model agrees better with experimental data. An example of a model without chaperonin allostery (see below) does not provide as good a fit of Rubisco folding into the data in Ref. 19. Moreover, the rates in our kinetic model have a physical meaning. The steps in the GroEL allosteric cycle are distinct, and their rates can be separately measured and/or manipulated by mutations. Our extracted rates for the allosteric transitions agree with independent experiments. In addition to giving less accurate results for chaperonin concentration and time-

dependent yield profiles, a fit using a reduced model results in reaction rates that cannot be easily interpreted or used to make predictions. It is unclear how to predict the effects of mutations in the GroEL–GroES system using the model that ignores allostery. Experiments on various mutants of GroEL–GroES have linked their efficiencies to alterations in ATPase activity that are intrinsic to GroEL function. Nevertheless, we have considered a simplified reaction scheme to see if the experimental data can be fit without accounting for allosteric transitions.

In the absence of allosteric transitions, the chaperonin-related reactions reduce to:

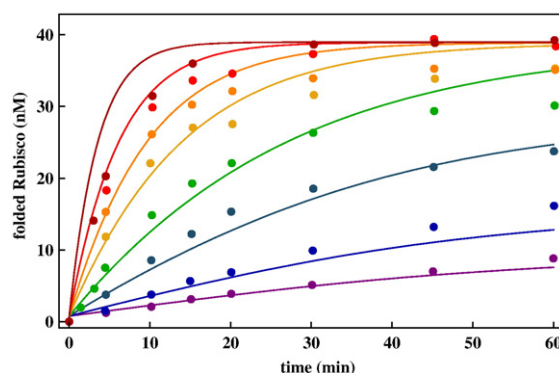


It is conceivable that kinetic schemes in which EL plays a passive role can be proposed. Because such schemes do not even account for GroEL concentration-dependent yield of SP, they do not warrant further discussion. The kinetic equations that describe assisted folding and include SP folding by KPM, GroEL without allostery, and SP aggregation, corresponding to the reactions in Eq. (13), are as follows:

$$\begin{aligned}
 \frac{d[X]}{dt} &= -k_F[X] \\
 \frac{d[M]}{dt} &= (1 - \Phi)k_F[X] + k_M[ELSP] - k_S[M] \\
 &\quad - \lambda_B[EL][M] - \lambda_A[M]^2 \\
 \frac{d[ELSP]}{dt} &= \lambda_B[EL][M] - (k_M + k_N)[ELSP] \\
 \frac{d[N]}{dt} &= k_S[M] + \Phi k_F[X] + k_N[ELSP] \\
 \frac{d[M_A]}{dt} &= \lambda_A[M]^2
 \end{aligned} \tag{14}$$

The solutions to Eq. (14) were used to fit the data for Rubisco. Although the quality of fits in Fig. 8 (especially at intermediate concentrations of EL) is not as good as that in Fig. 2, it appears that the reduced set of equations can approximately account for experimental observation. The parameters extracted from the fits are as follows:  $k_F=100 \text{ min}^{-1}$ ,  $k_S=0.0003 \text{ min}^{-1}$ ,  $\Phi=0.02$ ,  $\lambda_A=0.0003 \text{ nM}^{-1} \text{ min}^{-1}$ ,  $\lambda_B=6 \text{ nM}^{-1} \text{ min}^{-1}$ ,  $k_M=1400 \text{ min}^{-1}$ , and  $k_N=1 \text{ min}^{-1}$ .

There are notable differences between the extracted parameters using Eq. (14) and those that include allosteric



**Fig. 8.** Rubisco native-state yield at different chaperonin concentrations. Data points are from Ref. 19, and curves are fits based on a kinetic model without allostery [Eq. (14) in Methods].



transitions [Eq. (12)]. In particular, the value of  $k_S$  obtained for the reduced kinetic model is a factor smaller than that shown in Table 1. The much smaller value of  $k_S$  would result in the prediction that SR1 would not be able to fold Rubisco to any measurable degree, which is in apparent contradiction to the report by Brinker *et al.*<sup>42</sup> More importantly, the  $k_M$  value in the reduced model is far greater than all other timescales and is unphysical. Perhaps the greatest drawback with the kinetic scheme in Eq. (14) is that it cannot be used to predict or understand the effect of GroEL/S mutations that have a profound effect on the efficiency of chaperonin function. The mutations alter the rates in the reaction cycle of GroEL. Only by taking those changes into account can the action of WT GroEL and its variants be obtained (Figs. 3–7).

## Acknowledgements

We have benefited greatly from discussions with George H. Lorimer. R.T. is a Ruth A. Kirschstein Postdoctoral Fellow, and her work was supported by the National Institute of General Medical Sciences (1F32GM082009). D.T. is grateful to the National Science Foundation (CHE 05-14056), Air Force Office of Scientific Research, and the National Institutes of Health (1R01GM067851-01) for support.

## Appendix A. Derivation of Eq. (5) (steady-state substoichiometric limit)

If we assume that the second-order aggregation reaction rate can be replaced with an effective first-order rate  $k_A \approx \lambda_A [M]$ , Eq. (12) in the steady-state limit becomes:

$$\begin{aligned} 0 &= -k_F[X] + v_0 \\ 0 &= (1 - \Phi)k_F[X] + k_M[EL_{R''}SP] - k_S[M] \\ &\quad - \lambda_B[EL_T][M] - k_A[M] \\ 0 &= \lambda_B[EL_T][M] + k_{R-T}[EL_RSP] - k_{T-R}[EL_TSP] \\ 0 &= k_{T-R}[EL_TSP] - k_{R-T}[EL_RSP] - k_{R-R''}[EL_RSP] \\ 0 &= k_{R-R''}[EL_RSP] - (k_N + k_M)[EL_{R''}SP] \\ 0 &= k_S[M] + \Phi k_F[X] + k_N[EL_{R''}SP] - v_N \\ 0 &= \lambda_A[M]^2 - v_{M_A} \end{aligned}$$

where we have included  $v_0$ ,  $v_N$ , and  $v_{M_A}$  as the rate of introduction of nascent unfolded proteins, the rate of production of native proteins, and the rate of production of aggregated proteins, respectively. From the above equations, we obtain

$$[EL_TSP] = \frac{k_{R-T} + k_{R-R''}}{k_{T-R}} [EL_RSP] \text{ and} \quad (15)$$

$$[EL_RSP] = \frac{k_N + k_M}{k_{R-R''}} [EL_{R''}SP]. \quad (16)$$

Using these equations, we obtain

$$\lambda_B[EL_T][M] = (k_N + k_M)[EL_{R''}SP]. \quad (17)$$

By using the conservation of chaperonin concentration,  $[EL]_0 = [EL_T] + [EL_TSP] + [EL_RSP] + [EL_{R''}SP]$ , Eq. (17) can be rewritten as

$$\lambda_B([EL]_0 - [EL_TSP] - [EL_RSP] - [EL_{R''}SP])[M] = (k_N + k_M)[EL_{R''}SP] \quad (18)$$

By substituting Eqs. (15) and (16), Eq. (18) becomes

$$\lambda_B \left[ [EL]_0[M] - \lambda_B \left[ \frac{k_N + k_M}{k_{R-R''}} \left( \frac{k_{R-T} + k_{R-R''}}{k_{T-R}} + 1 \right) + 1 \right] [EL_{R''}SP][M] = (k_N + k_M)[EL_{R''}SP] \right] \quad (19)$$

or

$$\lambda_B[EL]_0[M] - \lambda_B A [EL_{R''}SP][M] = (k_N + k_M)[EL_{R''}SP] \quad (20)$$

where [see also Eq. (5)]

$$A = \frac{k_N + k_M}{k_{R-R''}} \left( \frac{k_{R-T} + k_{R-R''}}{k_{T-R}} + 1 \right) + 1. \quad (21)$$

Combining the first two steady-state equations along with Eqs. (17) and (20) yields the following quadratic equation for the concentration of the misfolded species:

$$\begin{aligned} [M]^2 + \left[ \frac{(k_M + k_N)(k_A + k_S) + \lambda_B[EL]_0 k_N}{\lambda_B(k_S + k_A)A} - \frac{(1 - \Phi)v_0}{k_S + k_A} \right] \\ [M] - \frac{(k_M + k_N)(1 - \Phi)v_0}{\lambda_B(k_A + k_S)A} = 0 \end{aligned} \quad (22)$$

Because  $\Phi$  and all the rates are positive, the only physical solution for  $[M]$  is a monotonically increasing function of  $A$ . The native-state yield decreases as  $[M]$  increases; thus, the native-state yield production increases as  $A$  decreases.

## Appendix B. Sensitivity analysis of the kinetic model [Eq. (12)]

We have assessed the effects of the changes in the kinetic parameters around the optimal values that produce the best fits to the Rubisco data (Fig. 2 and Table 1) by performing a local sensitivity analysis. We computed the time-dependent change in the native-state yield,  $\Delta[N](t)$ , in response to variations in the kinetic parameters. The change  $\Delta[N](t)$  is computed by using:

$$\Delta[N](t) = \sum_{\alpha} (\delta[N]/\delta p_{\alpha}) \Delta p_{\alpha} \quad (23)$$

where  $\Delta p_{\alpha}$  is the variation in the kinetic parameter  $p_{\alpha}$  [Eq. (12)] and  $(\delta[N]/\delta p_{\alpha})$  is the time-dependent derivative of the native-state yield with respect to  $p_{\alpha}$ . The magnitude of  $\Delta[N](t)$  is an estimate of the variation in the native-state yield in response to fluctuations in  $\Delta p_{\alpha}$ . The magnitude of  $(\delta[N]/\delta p_{\alpha})$  is an indication of the extent to which the parameter  $p_{\alpha}$

affects  $\Delta[N](t)$ . We computed  $\Delta[N](t)$  by varying each rate constant by  $\pm 5\%$  around the values given in Table 1. The derivatives, obtained using a simple finite difference scheme, depend on  $[EL]$ . Among the allosteric rates in the reaction cycle, the largest value corresponds to  $|\delta[N]/\delta k_{R \rightarrow T}|$  when  $[SP]_0 < [EL]_0$ . The maximum value ( $\approx 0.8$ ) is found at  $t \approx 10$  min. Similarly, the values of  $|\delta[N]/\delta \lambda_A|$  change from 40  $\text{nM}^2 \text{ min}$  at early times to 120  $\text{nM}^2 \text{ min}$  for  $\geq 20$  min. Even these seemingly large values produce negligible changes in  $\Delta[N](t)$  (Fig. 2b). All other derivatives are much smaller. Using all the values of  $\delta[N]/\delta p_{\alpha}$ , we have calculated  $\Delta[N](t)$  [using Eq. (23)]. The time-dependent changes in  $\Delta[N](t)$ , which are residuals in the best fit of the data using Eq. (12), are small. Thus, the kinetic scheme is robust to variations in the kinetic parameters.

## References

- Lin, Z. & Rye, H. S. (2006). GroEL-mediated protein folding: making the impossible, possible. *Crit. Rev. Biochem. Mol. Biol.* **41**, 211–239.
- Saibil, H. (2000). Molecular chaperones: containers and surfaces for folding, stabilizing or unfolding proteins. *Curr. Opin. Struct. Biol.* **10**, 251–258.
- Horwich, A. L., Farr, G. W. & Fenton, W. A. (2006). GroEL–GroES-mediated protein folding. *Chem. Rev.* **106**, 1917–1930.
- Horovitz, A. & Willison, K. R. (2005). Allosteric regulation of chaperonins. *Curr. Opin. Struct. Biol.* **15**, 646–651.
- Thirumalai, D. & Lorimer, G. H. (2001). Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269.
- Xu, Z. & Sigler, P. B. (1999). GroEL/GroES: structure and function of a two-stroke folding machine. *J. Struct. Biol.* **124**, 129–141.
- Viitanen, P. V., Gatenby, A. A. & Lorimer, G. H. (1992). Purified chaperonin 60 (GroEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Sci.* **1**, 363–369.
- Lorimer, G. H. (1996). A quantitative assessment of the role of the chaperonin proteins in protein folding *in vivo*. *FASEB J.* **10**, 5–9.
- Fenton, W. A. & Horwich, A. L. (1997). GroEL-mediated protein folding. *Protein Sci.* **6**, 743–760.
- Lorimer, G. H. (1997). Protein folding with a two-stroke motor. *Nature*, **388**, 720–721.
- Xu, Z., Horwich, A. L. & Sigler, P. B. (1997). The crystal structure of the asymmetric GroEL–GroES–(ADP)<sub>7</sub> chaperonin complex. *Nature*, **388**, 741–750.
- Kipnis, Y., Papo, N., Haran, G. & Horovitz, A. (2007). Concerted ATP-induced allosteric transitions in GroEL facilitate release of protein substrate domains in an all-or-none manner. *Proc. Natl Acad. Sci. USA*, **104**, 3119–3124.
- Amir, A. & Horovitz, A. (2004). Kinetic analysis of ATP-dependent inter-ring communication in GroEL. *J. Mol. Biol.* **338**, 979–988.
- Chennubhotla, C. & Bahar, I. (2006). Markov propagation of allosteric effects in biomolecular systems: application to GroEL–GroES. *Mol. Syst. Biol.* **2**, 36.
- Keskin, O., Bahar, I., Flatow, D., Covell, D. G. & Jernigan, R. L. (2002). Molecular mechanisms of chaperonin GroEL–GroES function. *Biochemistry*, **41**, 491–501.
- Hyeon, C., Lorimer, G. H. & Thirumalai, D. (2006). Dynamics of allosteric transitions in GroEL. *Proc. Natl Acad. Sci. USA*, **103**, 18939–18944.
- Wang, J. D., Herman, C., Tipton, K. A., Gross, C. A. & Weissman, J. S. (2002). Directed evolution of substrate-optimized GroEL/S chaperonins. *Cell*, **111**, 1027–1039.
- Sun, Z., Scott, D. J. & Lund, P. A. (2003). Isolation and characterisation of mutants of GroEL that are fully functional as single rings. *J. Mol. Biol.* **332**, 715–728.
- Todd, M. J., Lorimer, G. H. & Thirumalai, D. (1996). Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc. Natl Acad. Sci. USA*, **93**, 4030–4035.
- Todd, M. J., Viitanen, P. V. & Lorimer, G. H. (1994). Dynamics of the chaperonin ATPase cycle: implications for facilitated protein folding. *Science*, **256**, 659–666.
- Weissman, J. S., Kashi, Y., Fenton, W. A. & Horwich, A. L. (1994). GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell*, **78**, 693–702.
- Bukau, B., Weissman, J. S. & Horwich, A. (2006). Molecular chaperones and protein quality control. *Cell*, **125**, 443–451.
- Burston, S. G., Weissman, J. S., Farr, G. W., Fenton, W. A. & Horwich, A. L. (1996). Release of both native and non-native proteins from a *cis*-only GroEL ternary complex. *Nature*, **383**, 96–99.
- Weissman, J. S., Hohl, C. M., Kovalenko, O., Kashi, Y., Chen, S., Braig, K. *et al.* (1995). Mechanism of GroEL action: productive release of polypeptide from a sequestered position under GroES. *Cell*, **93**, 577–587.
- Fedorov, A. N. & Baldwin, T. O. (1997). GroE modulates kinetic partitioning of folding intermediates between alternative states to maximize the yield of biologically active protein. *J. Mol. Biol.* **268**, 712–723.
- Kolomeisky, A. B. & Fisher, M. E. (2007). Molecular motors: a theorist's perspective. *Annu. Rev. Phys. Chem.* **58**, 675–695.
- Wu, Y., Gao, Y. Q. & Karplus, M. (2007). A kinetic model of coordinated myosin V. *Biochemistry*, **46**, 6318–6330.
- Guo, Z. & Thirumalai, D. (1995). Kinetics of protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers*, **36**, 83–102.
- Kiefhaber, T. (1995). Kinetic traps in lysozyme folding. *Proc. Natl Acad. Sci. USA*, **92**, 9029–9033.
- Matagne, A., Radford, S. E. & Dobson, C. M. (1997). Fast and slow tracks in lysozyme folding: insight into the role of domains in the folding process. *J. Mol. Biol.* **267**, 1068–1074.
- Hyeon, C. & Thirumalai, D. (2005). RNA and protein folding: common themes and variations. *Biochemistry*, **44**, 4957–4970.
- Horst, R., Bertelsen, E. B., Fiaux, J., Wider, G., Horwich, A. L. & Wüthrich, K. (2005). Direct NMR observation of a substrate protein bound to the chaperonin GroEL. *Proc. Natl Acad. Sci. USA*, **102**, 12748–12753.
- Stan, G., Lorimer, G. H., Thirumalai, D. & Brooks, B. R. (2007). Coupling between allosteric transitions in GroEL and assisted folding of a substrate protein. *Proc. Natl Acad. Sci. USA*, **104**, 8803–8808.
- Yifrach, O. & Horovitz, A. (1995). Nested cooperativity in the ATPase activity of the oligomeric chaperonin GroEL. *Biochemistry*, **34**, 5303–5308.
- Inbar, E. & Horovitz, A. (1997). GroES promotes the T to R transition of the GroEL ring distal to GroES in the GroEL–GroES complex. *Biochemistry*, **36**, 12276–12281.
- Yifrach, O. & Horovitz, A. (1998). Mapping the transition state of the allosteric pathway of GroEL

- by protein engineering. *J. Am. Chem. Soc.* **120**, 13262–13263.
37. Ma, J., Sigler, P. B. & Karplus, M. (2000). A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* **302**, 303–313.
38. Li, M. S., Klimov, D. K. & Thirumalai, D. (2004). Thermal denaturation and folding rates of single domain proteins: size matters. *Polymer*, **45**, 573–579.
39. Viitanen, P. V., Lubben, T. H., Reed, J., Goloubinoff, P., O'Keefe, D. P. & Lorimer, G. H. (1990). Chaperonin-facilitated refolding of ribulosebiphosphate carboxylase and ATP hydrolysis by chaperonin 60 (groEL) are  $K^+$  dependent. *Biochemistry*, **29**, 5665–5671.
40. Rye, H. S., Roseman, A. M., Furtak, K., Fenton, W. A., Saibil, H. R. & Horwich, A. L. (1999). GroEL–GroES cycling: ATP and non-native polypeptide direct alternation of folding-active rings. *Cell*, **97**, 325–338.
41. Yifrach, O. & Horovitz, A. (2000). Coupling between protein folding and allostery in the GroE chaperonin system. *Proc. Natl Acad. Sci. USA*, **97**, 1521–1524.
42. Brinker, A., Pfeifer, G., Kerner, M. J., Naylor, D. J., Hartl, F. U. & Hayer-Hartl, M. (2001). Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*, **107**, 223–233.
43. Betancourt, M. R. & Thirumalai, D. (1999). Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *J. Mol. Biol.* **287**, 627–644.
44. Takagi, F., Koga, N. & Takada, S. (2003). How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc. Natl Acad. Sci. USA*, **100**, 11367–11372.
45. Jewett, A. I., Baumketner, A. & Shea, J.-E. (2004). Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an alternate fast folding pathway. *Proc. Natl Acad. Sci. USA*, **101**, 13192–13197.
46. Klimov, D. K., Newfield, D. & Thirumalai, D. (2002). Simulations of beta-hairpin folding confined to spherical pores using distributed computing. *Proc. Natl Acad. Sci. USA*, **99**, 8019–8024.
47. Cheung, M. S., Klimov, D. K. & Thirumalai, D. (2005). Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl Acad. Sci. USA*, **102**, 4753–4758.
48. Guo, Z. & Thirumalai, D. (1996). Kinetics and thermodynamics of folding of a *de novo* designed four-helix bundle. *J. Mol. Biol.* **263**, 323–343.