A Kinetic Model for Chaperonin Assisted Folding of Proteins
Henri Orland, D. Thirumalai

To cite this version:

HAL Id: jpa-00247343
https://hal.archives-ouvertes.fr/jpa-00247343
Submitted on 1 Jan 1997

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A Kinetic Model for Chaperonin Assisted Folding of Proteins

Henri Orland (1,*) and D. Thirumalai (2)

(1) CEA, Service de Physique Théorique, CE-Saclay, 91191 Gif-sur-Yvette Cedex, France
(2) Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742-2341, USA

(Received 17 July 1996, received in final form 22 November 1996, accepted 26 November 1996)

PACS.82.20.-w – Chemical kinetics
PACS.87.15.-v – Molecular biophysics
PACS.87.15.Da – Physical chemistry of solutions of biomolecules; condensed states

Abstract. — A master equation formalism is used to account for the possible kinetic action of chaperonin assisted folding of proteins. The model is based on the assumption that chaperonins rescue misfolded protein structures and stochastically provide them pathways to reach the native state by a mechanism involving ATP hydrolysis. We assume that the misfolded structures are characterized by distinct free energies and that these structures are connected to the native conformation by a suitable transition probabilities. The chaperonins do not recognize the native state. For this model it is shown that in the presence of chaperonins the native state is populated exponentially. The exponential population of the native state, which is in accord with experiments, is independent of the distribution of the free energies of the misfolded structures as well as the transition probabilities connecting them to the native state.

1. Introduction

The classic experiments of Anfinsen [1] showed that protein folding, namely the acquisition of a unique structure with a well defined topology from the primary sequence of amino acids, is a self-assembly process. Recent studies [2-5], however, have shown that in many instances a group of chaperones, belonging to the family of heat shock proteins, is implicated in in vivo folding of proteins. It is suspected [6] that the molecular chaperones do not convey additional information (thus biasing the polypeptide chain to the native conformation) to enable the folding of the nascent polypeptide chains. Thus the in vivo folding of proteins should be viewed as an assisted self-assembly process. The chaperonins involved in the folding of proteins in E. Coli are GroEL and GroES (and perhaps others) [7].

A major challenge is to understand the mechanism of the action of chaperonins in the folding of proteins in cells. In very general terms one could classify the molecular chaperones into two groups, namely, intramolecular chaperones and intermolecular chaperones. In the former category are those classes of proteins that serve as signal sequences and whose presence seems necessary in the folding of α-lytic proteases [8], and in the folding of bovine pancreatic trypsin inhibitor [9]. To our knowledge no mechanism for the kinetic action of these intramolecular

(*) Author for correspondence (e-mail: orland@spht.saclay.cea.fr)
chaperonins has been proposed so far. In this article we consider only the possible role of intermolecular chaperones, namely, the chaperonin family.

The kinetic action of chaperonins has not been completely understood. The overall geometry of GroEL is that of a double toroid with seven subunits in each ring. The approximate length of the roughly cylindrical structure is 100 Å while the diameter of the hole is about 45 Å [10]. The geometry of GroEL was used by Saibil et al. [11, 12] to suggest that the partially folded protein gets trapped in the hole of GroEL. Assembly of proteins is thought to take place in the sequestered environment offered by GroEL. In this scenario, the chaperonins play a passive role and their presence is meant to ensure a completely protected environment, i.e. they basically provide an Anfinsen cage for folding. Fersht et al. [13] have suggested that folding could take place while the protein is still bound to the GroEL. These authors suggested this mechanism by studying the folding of barnase in the presence of GroEL. Their experiments did not require ATP hydrolysis which, on fairly general grounds, has been observed to be necessary in the assisted self-assembly of proteins. This suggests that under the conditions used in the experiments of Fersht and coworkers, the folding of barnase would not have required GroEL, i.e. there are permissive folding conditions. In other words the in vitro folding of barnase is almost spontaneous involving, perhaps, discrete intermediates.

A proposal very different from the two described above has been made by Todd et al. [14] based on the energy landscape perspective, and the in vitro kinetic partition mechanism. This mechanism termed the iterative annealing mechanism (IAM) assumes that the major role of GroEL is to utilize the energy from ATP hydrolysis to rescue misfolded structures that are separated from the native conformation by barriers which are difficult to overcome on biological time scale. With this energy, the misfolded protein is released into solution and a fraction of such molecules partition to the native conformation whereas the remaining ones get trapped in a large number of misfolded conformations. Then, another round of ATP hydrolysis ensues and this kinetic partitioning is repeated until sufficient yield is obtained. This process of repeated release of misfolded and subsequent annealing is the reason for labeling this scenario the iterative annealing mechanism. In this mechanism, GroEL is thought to play an active role which when coupled with ATP hydrolysis serves to enhance the rate of folding as was conjectured earlier. It should be emphasized that the IAM does not imply that the polypeptide chain does not undergo large structural fluctuations that might lead to unfolding even after the native state is reached. The only assumption is that the GroEL does not bind to the native conformation of the protein. The IAM mechanism seems to have experimental support, although more work is required to confirm various aspects of this scenario. The IAM bears some resemblance to the kinetic proofreading mechanism familiar in the context of DNA replication. The contrast between IAM and that based on kinetic proofreading treatment and the random energy model [15] is presented elsewhere. We propose in this paper a master equation approach to predict some of the consequences of the IAM. Recently there have been lattice simulations of in vivo folding based on the IAM [16, 17].

2. The Model

It is useful to recall the properties of chaperonins which are crucial in their ability to rescue misfolded structures [14]. Briefly they are: (1) perhaps the most important requirement of the chaperonin system is that they enable folding to occur under non-permissive conditions i.e. when spontaneous folding under biologically relevant time scale is unlikely. In general this requires the presence of both GroEL and GroES; (2) it has been shown by Lorimer et al. [18] that chaperonins are promiscuous in their interaction with substrate proteins i.e. they recognize a large number of structurally unrelated proteins as long as they are not in the native
conformation; (3) upon ATP hydrolysis the substrate protein is released in a non specific stochastic manner and the released protein either reaches the native conformation or gets trapped in one of the many misfolded structures. These two properties are mimicked using a simple master equation and the kinetic consequences of the theory are derived below.

For simplicity, we model the protein as existing in several free energy states \( f_a \) \((a = 0, 1, 2, \ldots)\). We imagine that to a large extent, these states correspond to misfolded structures that have to cross large free energy barriers to get to the native conformation, which is chosen to be conformation \( a = 0 \). The master equation giving the probability of occupation of each of the states \( P_a(t) \) can be written as

\[
\frac{d}{dt} P_a(t) = \sum_b W_{ab} P_b(t) - \sum_b W_{ba} P_a(t) \tag{1}
\]

where as usual the transition probabilities \( W_{ab} \) (probability for the system to make a transition from state \( b \) to \( a \) per unit time) satisfy detailed balance, namely

\[
W_{ab} \exp(-\beta f_b) = W_{ba} \exp(-\beta f_a) \tag{2}
\]

with \( \beta = 1/kT \). For concreteness we choose

\[
W_{ab} = \tau_0^{-1} \exp\left(-\frac{\beta}{2}(f_a - f_b)\right) \tag{3}
\]

where \( \tau_0 \) is a time constant that sets the scale for an elementary event in the problem. In order to adapt this general formulation to the chaperonin problem, we make additional approximations for the transition probabilities involving the 0th state (referred to as the native state of the protein). It is known that GroEL does not recognize the native conformation and only binds to the misfolded structures [18]. Furthermore the IAM allows for the possibility of any misfolded structures to be connected with some probability (however small) to the native conformation. These important properties can be mimicked by setting \( W_{a0} = 0 \) for all states \( a \neq 0 \). This implies that on certain time scales, the native conformation acts as an absorbing sink and once it is reached, transitions to other states are unlikely. With these approximations, which are intended to be a minimal representation of chaperonin action according to the IAM, the equations to be analyzed become:

\[
\frac{d}{dt} P_0(t) = \sum_{b \neq 0} W_{0b} P_b(t) \tag{4}
\]

and

\[
\frac{d}{dt} P_a(t) = \tau_0^{-1} \left[ \sum_{b \neq 0} e^{-\frac{\beta}{2}(f_a - f_b)} P_b(t) - \sum_b e^{-\frac{\beta}{2}(f_b - f_a)} P_a(t) \right] \tag{5}
\]

for \( a \neq 0 \). It is easily seen that this set of equations conserves the sum of the \( P_a(t) \).

These equations can be readily solved to yield:

\[
P_0(t) = \tau_0^{-1} e^{-\beta f_0/2} \int_0^t dt' P_T(t') \tag{6}
\]

where

\[
P_T(t) = \sum_{b \neq 0} e^{\beta f_b/2} P_b(t) \tag{7}
\]
and
\[ P_b(t) = \tau_0^{-1} e^{-\beta f_b/2} \int_0^t dt' e^{-A_b\lambda(t-t')/\tau_0} P_T(t') + C_b e^{-A_b\lambda t/\tau_0} \]  \tag{8}

In equation (8), the various constants are given by
\[ \lambda = \sum_{\text{all } b} e^{-\beta f_b/2} \]
\[ A_b = e^{\beta f_b/2} \]  \tag{9}
\[ C_b = P_b(t=0) \]  \tag{10}
and for simplification, we have assumed that the initial population of the native protein is 0.

Using the Laplace transform of \( P_0(t) \) and analyzing the structure of its poles, it can be shown that \( P_0(t) \) decays exponentially. Indeed, denoting by \( \tilde{P}_0(p) \) the Laplace transform of \( P_0(t) \), after some algebra, we obtain:
\[ \tilde{P}_0(p) = \frac{1}{A_0 z} \left( 1 - \frac{1}{\lambda} \sum_{b \neq 0} \frac{1}{z + \lambda A_b} \right) \sum_{b \neq 0} \frac{\tau_0 C_b A_b}{z + \lambda A_b} \]  \tag{12}

where \( z = \tau_0 p \).

The function \( P_0(t) \) is given by the inverse Laplace transform:
\[ P_0(t) = \int_{C-i\infty}^{C+i\infty} \frac{dp}{2i\pi} e^{pt} \tilde{P}_0(p) \]  \tag{13}
where \( C \) is a constant larger than the largest singularity of the integrand.

This integral can be performed by the residue method. Denoting respectively by \( z_\alpha = \tau_0 p_\alpha \) and \( R_\alpha \) the poles and residues of \( \tilde{P}_0(p) \), we have:
\[ P_0(t) = \sum_\alpha R_\alpha e^{z_\alpha t/\tau_0} \]  \tag{14}

The large time behavior of \( P_0(t) \) is determined by the pole structure of \( \tilde{P}_0(p) \). As seen in equation (12), there is a pole at \( z = 0 \), with a residue which is easily seen to be equal to 1.

One might think that there are poles at \( z_b = -\lambda A_b, b \neq 0 \), but these poles are neutralized by the denominator of the premultiplying fraction, which also diverges at these poles. Therefore, \( \tilde{P}_0(p) \) is regular at \( z_b \). Note that due to the definition of \( \lambda \) and \( A_b \), equations (9, 10), we see that for any distribution of energies, \( z_b < -1 \).

Finally, there is one series of poles given by the equation:
\[ \sum_{a \neq 0} \frac{1}{z + \lambda A_a} = 1. \]  \tag{15}

This equation can easily be solved graphically (see Fig. 1). To characterize this set of poles, they are interleaved with the \( \{z_b\} \) and are all smaller than \(-1\), except for one pole, which we denote by \( z_1 = -K_0 \), which is separated from the others and satisfies the inequality \(-1 < z_1 < 0 \). This result again holds for any distribution of energies. In particular, if
the spectrum of energies becomes continuous, all poles will form a continuous band below \(-1\), except for the pole at \(z = 0\) and the pole \(z_1\) which remains above \(-1\).

As a result of our analysis, we see that the large time behavior of \(P_0(t)\) is given by:

\[
P_0(t) \sim 1 - R_1 e^{-K_0 t}
\]

(16)

A simple estimate of the effective rate constant \(K_0\) can be obtained if we assume that it is close to \(\tau_0^{-1}\). We obtain:

\[
K_0 \tau_0 = 1 - \sum_{a \neq 0} \frac{1}{-1 + \lambda A_a} - 1
\]

(17)

The form for \(P_0(t)\) given in equation (16) asserts that no matter what the distribution of free energies of the misfolded structures are, the native state is populated exponentially, provided that on a certain time scale, transitions from the native conformation are unlikely. We have verified explicitly that similar results are obtained for other choices of \(W_{ab}\). In particular we have shown that equation (16) is obtained (with a different \(K_0\)) if \(W_{ab} = \tau_0^{-1} \exp(-\beta f_a)\).

Recently Zwanzig [19] has suggested a simple model for protein folding kinetics in which the reaction kinetics is described in terms of changes in an order parameter that measures the similarity of a given conformation to the native state. A master equation description of this model, in which the native conformation has the largest probability of occupation, also shows that the native conformation is populated in an exponential manner. Our results show that this is more general. In fact they are obtained using the rather weak assumptions that the misfolded structures are connected with some probability to the native conformation and that the transition out of the native state is unlikely on a certain time scale. For the chaperonin mediated folding this time scale is essentially set by \(K_0^{-1}\).
An estimate of the rate of production of the native state in the chaperonin induced folding is complicated due to several reasons. Consider a single turnover experiment. In such an experiment the GroEL is initially bound to the polypeptide chain which is in a misfolded conformation. After a round of ATP hydrolysis, which leads to the release of the polypeptide chain, GroEL goes to the off state for a period of time. In the off state GroEl does not bind to the misfolded states. After the ATP hydrolysis is complete and a fraction of the polypeptide chain molecules has partitioned to the native conformation (referred to as the partition factor in [14]) GroEL reaches the on state and once again binds to the molecules still in the misfolded conformation. The folding rate $K_0^{-1}$, which depends on the concentration of GroEL, is essentially set up by the transition time for the GroEL to reach the on state from the off state. In experiments it is of the order of a minute. The utility of equation (16) is that it can be used to estimate the fraction of molecules that reach the native conformation ($\propto R_1$) in one round of ATP hydrolysis. Using the experimental rate of population of the native conformation for RUBISCO [14] we estimate the partition factor to about 5% which is consistent with the experimental measurement [14].

3. Conclusions

In this paper, we have proposed a simple kinetic model for chaperonin facilitated folding of proteins. The crucial experimental observation [14] that GroEL does not recognize the native state has been used to construct a master equation in which the probability of escape from the native conformation is negligible. This simple assumption leads to the fact that the population of the native conformation occurs exponentially, in agreement with recent experiments. It is remarkable that our result appears to be independent of the free energies of the misfolded structures, i.e. the states other than the native one. Previous studies of the use of master equations for disordered systems have shown that one could obtain stretched exponential [20] or power law [21] behavior depending upon the exact form of the transition probability $W_{ab}$ used. The $W_{ab}$ used in this paper depend explicitly on the initial and final states. We have explicitly checked that our conclusion (cf. Eq. (16)) remains unchanged if the $W_{ab}$ of reference [20] is utilized. The physical reason for obtaining exponential relaxation is due to the fact that the native conformation acts essentially as a trap and that there is some path that connects all misfolded structures to the native state, i.e. $W_{0a} \neq 0$, for all $a$. In the case of chaperonin assisted folding, $W_{0a} \neq 0$ is ensured because according to the iterative annealing mechanism the barriers separating the misfolded structure and the native states are overcome by consuming the energy from ATP hydrolysis. Thus frozen states, which are the hallmark of glassy systems, are not essentially allowed and this translates into fast population of the lowest free energy minimum which in the case of proteins corresponds to the native conformation.

In describing the model in this article we have not explicitly addressed the role of GroES a peptide which like GroEL also has a seven fold symmetry and acts as a dome when bound to GroEL. Although it has been demonstrated that protein folding can occur in the absence of GroES [2-4] recent experiments [23] suggest that GroES does play an important role. The basic assumption of IAM, namely, the chaperonin machinery (GroEL, GroES, and others) enables misfolded structures an opportunity to reach the native state, still remains valid. However, in the presence of GroES the release of misfolded structures takes place only when the dome opens. This additional event, which may have biological significance, does not alter the mathematical model presented here. The only input in our treatment is that chaperonins provide the necessary free energy so that barriers connecting a given misfolded structure and the native conformation can be overcome. Thus the various misfolded structures are connected to the native state with some probability.
The theory proposed here is based on the iterative annealing mechanism (IAM) of chaperonin action. One of the assumptions of IAM is that the free energy released upon ATP hydrolysis is sufficient to release the substrate protein from GroEL so that the folding to the native conformation can occur via the kinetic partition mechanism. If the interaction between the protein and GroEL is too strong then it follows that the GroEL-protein complex would be too stable. If this were the case folding rate (if the protein folds at all) would be considerably slowed down. An estimate of the optimal interaction strength between the chaperonin system and the substrate protein can be obtained using scaling arguments and the fact that ATP hydrolysis occurs in the heptameric GroEL structure in a "quantized" [22] manner. These calculations show that the interaction energy per residue between the protein and the GroEL should satisfy the inequality [22]

$$\epsilon \leq \left(\frac{n\Delta}{Nk_BT}\right)^{2/5}$$

(18)

where $n$ is the stoichiometry of the moles of ATP involved in each iteration, $\Delta$ is the free energy released upon ATP hydrolysis, $N$ is the number of residues in the substrate protein, and $T$ is the temperature. If the above inequality is not satisfied we expect that folding in the presence of chaperonin system is either retarded or does not take place at all. In the GroEL system $n$ is seven corresponding to the heptameric ring with each acting as a center for ATP hydrolysis, $\Delta$ is roughly about 7 kcal/mole. This gives $\epsilon$ between (0.6-1.2) kcal/mole for $N = 150$ depending on the precise value used for $\Delta$. Thus optimal folding is expected to occur by the IAM mechanism if $\epsilon$ lies between (1 - 2) $k_BT$.

The assessment that the strength of interaction between GroEL and the substrate protein and the folding process could be correlated is consistent with recent experimental observations [24]. The above arguments also suggest that if the inequality in equation (18) is not satisfied then the folding of the substrate protein, if it occurs at all, would have to take place while it is still bound to GroEL. If this were the case the mechanism by which the folded protein would be released is not at all clear. These considerations suggest that in all likelihood the proteins that require the chaperonin system (i.e. when spontaneous folding does not take place) are bound to GroEL in such a way that equation (18) is satisfied and hence the native conformation would be reached by the iterative annealing mechanism. It is clear that further experiments of folding of a variety of proteins in the presence of chaperonin system as well as the energetics of interaction of the substrate proteins with GroEL are required to clarify these issues.

Acknowledgments

One of us (DT) is grateful to G.H. Lorimer for several useful discussions on various aspects of chaperonin mediated protein folding. This work was supported in part by a grant from the National Science Foundation and the Air Force Office of scientific research.

References