Different mechanistic requirements for prokaryotic and eukaryotic chaperonins: a lattice study

Etai Jacob¹, Amnon Horovitz² and Ron Unger¹,*

¹The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, 52900, Israel and ²Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel

ABSTRACT

Motivation: The folding of many proteins in vivo and in vitro is assisted by molecular chaperones. A well-characterized molecular chaperone system is the chaperonin GroEL/GroES from Escherichia coli which has a homolog found in the eukaryotic cytosol called CCT. All chaperonins have a ring structure with a cavity in which the substrate protein folds. An interesting difference between prokaryotic and eukaryotic chaperonins is in the nature of the ATP-mediated conformational changes that their ring structures undergo during their reaction cycle. Prokaryotic chaperonins are known to exhibit a highly cooperative concerted change of their cavity surface while in eukaryotic chaperonins the change is sequential. Approximately 70% of proteins in eukaryotic cells are multi-domain whereas in prokaryotes single-domain proteins are more common. Thus, it was suggested that the different modes of action of prokaryotic and eukaryotic chaperonins can be explained by the need of eukaryotic chaperonins to facilitate folding of multi-domain proteins.

Results: Using a 2D square lattice model, we generated two large populations of single-domain and double-domain substrate proteins. Chaperonins were modeled as static structures with a cavity wall with which the substrate protein interacts. We simulated both concerted and sequential changes of the cavity surfaces and demonstrated that folding of single-domain proteins benefits from concerted but not sequential changes whereas double-domain proteins benefit also from sequential changes. Thus, our results support the suggestion that the different modes of allosteric switching of prokaryotic and eukaryotic chaperonin rings have functional implications as it enables eukaryotic chaperonins to better assist multi-domain protein folding.

1 INTRODUCTION

Polypeptide chains carry all the information required to fold to their native functional 3D structure and do not require any additional molecules to direct the folding process (Anfinsen, 1973). However, in vivo, inside the complex milieu of a living cell, protein folding is facilitated by a class of proteins called ‘molecular chaperones’ whose role is especially essential under stress conditions such as heat shock. One major family of chaperones, called chaperonins, comprises ATP-dependent proteins that facilitate folding by binding the assisted protein (i.e. the substrate) in a cavity formed at each end of their double ring structure. A recent study in Escherichia coli (Kerner et al., 2005) suggested that close to a hundred proteins in this organism require the chaperonin system (called GroEL/GroES, see below) in order to fold properly.

The prokaryotic GroEL/GroES complex in E. coli is the best characterized chaperonin system. GroEL consists of two rings each formed by seven identical protein subunits. GroES is a single-ring heptamer that binds to GroEL in the presence of ATP and serves as the cap of the cavity formed by each ring structure (see Fig. 1). Each GroEL subunit can rotate and thus turn a different surface towards the inner cavity.

CCT (also called TCP-1 ring complex) is an eukaryotic chaperonin that is composed of eight similar (but not identical) subunits arranged in a fixed order around the ring (Liou and Willison, 1997). CCT does not have a GroES-like cap; instead it contains a ‘built-in’ lid that closes in an ATP-dependent manner to encapsulate its substrates, a process that is required for the folding process. How exactly chaperonins facilitate folding is still unclear. The passive ‘Anfinsen cage’ model suggests that the main effect of chaperonins is to supply each folding molecule a safe environment that protects it from aggregation with other folding molecules or protease digestion. However, the fact that chaperonins undergo coordinated ATP-dependent allosteric transitions during the process, suggests that they play a more active role. Thus, they may provide an environment that is able to guide the substrate towards structures with desired characteristics, for example, towards structures that have their polar residues on the surface.

Active involvement of chaperonins in the folding of the substrate proteins may involve two alternative mechanisms: (i) iterative annealing (see for example, Todd et al., 1996) where the protein substrate can bind several times to the chaperonin during its folding process and is, thus, offered multiple chances to reach the native state and (ii) confinement (or caging) (Baumketner et al., 2003).

It is well established (Horovitz and Willison, 2005) that chaperonin rings can be in either a T (tense) or R (relaxed) state. In the T state of GroEL, the subunits exhibit a hydrophobic surface towards the cavity; this is an acceptor state for non-folded proteins, which have many exposed hydrophobic residues. In the R state of GroEL, chaperonins display polar residues towards the cavity, thereby enabling proteins to be released from the cavity surface and to continue folding either within the cavity volume or in bulk solution. This switch is mediated by ATP binding, since in the T state, the subunits have a low affinity for ATP and in the R state the subunits have high affinity for ATP (Yifrach and Horovitz,
1995). Increasing ATP concentration, therefore, leads to a cooperative change of all subunits from the T to the R state.

One major difference between the prokaryotic and eukaryotic chaperonins is the coordination between the surface change of the subunits. While in the GroEL/GroES system the change is concerted (Ma et al., 2000; Yifrach and Horovitz, 1998), i.e. all subunits switch together, it was recently shown (Rivenzon-Segal et al., 2005) that in CCT the change is sequential, i.e. the subunits switch conformation one after the other.

For relatively short, single-domain proteins, a concerted switch of the entire system is necessary since switching one subunit (i.e. one surface) to the release state is not effective if other surfaces are still in the T state and remain attached to other parts of the protein. However, it was suggested that a sequential change might be beneficial to eukaryotic proteins that tend to be larger and multi-domain as it may enable one domain of these larger proteins to detach from the cavity surface and fold while the other domain(s) is still attached to the surface. In a recent study (Kipnis et al., 2007), it was shown that a GroEL mutant that is defective in its ability to perform the concerted switch (Danziger et al., 2003) and thus behaves more like the CCT sequential chaperonin, can release a protein in a domain-by-domain manner.

In this study, we used a simple lattice model of the chaperonin–protein substrate system to explore the implications of the concerted versus sequential conformational switching. Are longer, multi-domain proteins more likely to benefit from a sequential mechanism of chaperonin transitions? We show here that our simulations are compatible with this hypothesis and, thus, support the idea that the different switching mechanism of prokaryotic versus eukaryotic chaperonins is related to the requirement of eukaryotic cells to fold multi-domain proteins.

2 METHODS

Studying the fundamental questions underlying the phenomenon of protein folding has been facilitated by the introduction of simple folding models. Simple, or even abstract, models of protein folding, while ignoring many of the small details of this process, are very useful for elucidating general principles regarding protein folding. For example, the importance of hydrophobicity in folding (Dill et al., 1995; Gutin et al. 1995); the relevance of energy landscapes (Bryngelson and Wolynes, 1987); the effectiveness of genetic algorithms in computational studies of folding (Unger and Moult, 1993); the concept of an energy gap (Sali et al., 1994) and the relative importance of short versus long-range interactions (Abkevich et al., 1995; Unger and Moult, 1996) have all been established by simple model studies. It is clear that such studies do not provide proof for the existence of folding-related phenomena but conclusions from them can certainly be used to promote and critically assess ideas about protein folding mechanisms.

In the first lattice model study of chaperone-assisted folding, Chan and Dill (1996) found that the folding yield depends on the amino acid sequence of the substrate, the chaperonin size and the binding and ejection rates from the chaperonin. In another study (Betancourt and Thirumalai, 1999) it was found, using a small number of simple protein substrates, that rapid cycling of the level of hydrophobicity of the surface of the chaperonin cavity can significantly reduce folding times and increase the folding yield under non-permissive (i.e. high temperature) conditions.

In order to enable the simulations described here, we have developed a computational engine that can be used to simulate many types of interactions between a folding protein and a chaperone on a lattice. Using this engine, it is possible to monitor the effects on the behavior of the system of changes in parameters such as the chaperonin cavity’s shape, size, surface composition, the way the surface changes,

![Fig.1. The GroEL–GroES complex of E. coli. Top- (top) and side-views (bottom) of the PDB structure (1AON) of the GroEL–GroES complex (Xu et al., 1997). This heat shock protein is composed of two attached rings of GroEL subunits (in blue and green) and a GroES cap (red) that binds at one end.](image-url)
the strength of interactions between amino acids (either between residues of the protein substrate or between substrate and chaperonin residues) and mechanisms of protein binding and release from the chaperonin. Furthermore, the simulation engine supports the distributed grid computational platform, and thus all simulations could be executed on the European GRID platform (EGEE).

### 2.1 Lattice model of proteins

We used the same lattice model platform as in our study of the behavior of termini residues of proteins (Jacob and Unger, 2007). Here we summarize this model in brief. We use a 2D square lattice with an alphabet of four amino acids: hydrophobic (H), neutral polar (P), positively charged (+) and negatively charged (-). Interactions are considered only between residues in neighboring lattice points (diagonal points are not considered neighboring). Interactions between consecutive residues in the sequence are not considered since they are always present and are independent of the conformation (see Fig. 2 for an example of a model structure). Note that we chose to use a 2D lattice rather than a 3D lattice, as for feasible sizes of proteins, a 2D lattice offers a more realistic ratio of surface to core residues. For example, a 25-residue sequence compactly folded to a perfect 5 \times 5 square on a 2D grid will have a ratio of 16/9 surface to core residues. A similar 27-long sequence will have a ratio of 26/1 on a 3 \times 3 \times 3 cubic lattice. The energy of sequence \( S \) of length \( N \) in conformation \( C \) is given by:

\[
E(S) = \sum_{j=1}^{N} P_{ij} \delta(|r_i - r_j|)
\]

where

\[
\delta(x) = \begin{cases} 
1 & \text{if } x = 1 \\
0 & \text{otherwise}
\end{cases}
\]

and \( P_{ij} \) represents the contact interaction energies which are defined as shown in Table 1.

### 2.2 Simulation technique

Folding dynamics are simulated using the Monte Carlo (MC) method with the Metropolis criterion (Metropolis et al., 1953). Starting from a random conformation, the following steps are iterated: From a conformation \( S_i \) with energy \( E_i \), accept a random move to conformation \( S_j \) if \( E_j < E_i \), or if the following non-deterministic test is met:

\[
\text{Rnd} < \exp \left( \frac{E_i - E_j}{C_k T_f} \right)
\]

where Rnd is a random number between 0 and 1, and \( C_k = 1 \). A value of 0.5 for \( T_f \) was used for selecting the proteins and \( T = 0.75 \) was used for the folding simulations (see Methods section). If the move is not accepted, the former conformation \( S_i \) is retained. Two types of moves are considered: a tail move, which is a random left or right turn of the first or last residue of the chain, and an internal move which is performed as follows: (a) Two residues with a sequence separation up to \( L \) residues are randomly selected. Then, (b) the trajectory between the two residues is replaced by another trajectory, taken from a predefined library of trajectories of the same length and the same relative translocation between their end points (see Fig. 3). Only trajectories that do not collide with another part of the chain are considered. Here we used move size up to \( L = 7 \). Note that this notion of local moves is a generalization of standard local moves, for example, corner flips (also known as kink jumps) and crankshafts for \( L = 3 \) (see review in Skolnick and Kolinski, 1991).

### 2.3 Model of chaperonins

In our model, chaperonins are modeled as proteins with static conformations (octagonal or square) whose sequence is composed of only H and P residues. The same table of interactions (Table 1) was used to evaluate the interactions between protein (substrate) residues and chaperonin residues. Each chaperonin has a cavity that can contain a semi or fully compact collapsed protein. In accordance with current

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>P</th>
<th>+</th>
<th>–</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>0</td>
<td>0</td>
<td>-0.125</td>
<td>-0.125</td>
</tr>
<tr>
<td>+</td>
<td>0</td>
<td>-0.125</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>–</td>
<td>0</td>
<td>-0.125</td>
<td>-1</td>
<td>+1</td>
</tr>
</tbody>
</table>

These contact interaction values conceptually represent the average strength of interactions in empirical mean force potentials (Miyazawa and Jernigan, 1993), where HH and (+)(–) interactions are considered most favorable, HP and H(+)(–) interactions are neutral, P(+)(–) interactions are weakly favorable, (+)(+) or (–)(–) are repulsive and PP interactions are neutral.
Two different mechanisms of the way substrate proteins interact with chaperonins were simulated:

(a) Binding and release—Substrate protein commences its folding process in an open environment, where it can make any movement without colliding with a lattice boundary. After a predefined duration, the protein binds to the chaperonin which confines it in its cavity for an additional predefined duration [e.g. 100,000 Monte Carlo Steps (MCs)]. After this time, the protein is released from the chaperonin cavity back to the open environment. This cycle of binding and release may be repeated several times during the folding process.

(b) Caging—The protein is inside the cavity of the chaperonin during the entire simulation.

### 2.4 Chaperonin substrates

In eukaryotic cells, ~70% of proteins are multi-domain whereas in prokaryotes single-domain proteins are more common. We wanted to examine whether this fundamental difference between eukaryotic and prokaryotic cells may have had a selective effect on the mechanism of allosteric switching of their respective chaperonins. Hence, three types of substrate proteins interacting with chaperonins were studied: (i) proteins of 25 residues in length (these are single-domain proteins as proteins of that size cannot form two domains); (ii) single-domains of 55 residues and (iii) double-domains of 55 residues (total length). For each type, ~100 different sequences were tested (123 sequences of 25-residue single-domains, 117 sequences of 55-residue single-domains and 104 sequences of 55-residue double-domain proteins). All model sequences were generated by random selection of 25- or 55-residue sequences, drawn from a distribution of 45% H, 30% P, 12.5% (+) and 12.5% (−). This composition is based on the composition of amino acid groups in the PDB (http://us.expasy.org/sprot/relnotes/relstat.html, PFB release 49.1) that is: 44.3% neutral, 30.7% polar, 12% positively charged amino acids and 13% negatively charged amino acids. Oppositely charged amino acids (+/−) were assigned to both termini to reflect the fact that protein termini (amino and carboxyl groups) are charged.

#### 2.4.1 Sequences of single-domain substrates with 25 residues

The generation of the 25-residue sequences was based on a thermodynamic selection criterion followed by a kinetic selection.

**Thermodynamic selection**

We created sequences with the amino acid composition mentioned above. For each sequence, we considered the lowest energy conformation amongst all possible 9,646,215 conformations that fit into a compact 6 x 6 square. If there was more than one conformation with the same minimum, one was arbitrarily chosen as the native conformation. Conformations for which the simulation (to be described below) demonstrated that the minimal energy is not a compact structure (i.e. the simulations found a minimal energy conformation that could not be contained within a 6 x 6 square) were excluded retroactively from consideration. We encountered only very few (<1%) such cases. Using this procedure, 1084 sequences were analyzed.

There is a large variance in the spectrum of energy values of the conformational space of different proteins. A significant energy gap is important in order to ensure kinetic accessibility of the native structure as suggested by Sali et al. (1994). Thus, for each sequence, we measured the difference between the minimal energy (i.e. the native conformation) and the average energy of all conformations, and expressed these in units of standard deviations of the average energy. The larger the difference between these two numbers, the more pronounced the
For each sequence, 200 independent, long \((10^7\) MCS) simulations were performed. If in any one of these 200 simulations, the simulation found a non-double-domain conformation that had a significantly lower energy than that of the double-domain structure, then the sequence was excluded from further analysis. A total of 104 sequences were selected under these criteria. Figure 6B illustrates an example of a double-domain structure.

**Single-domains with sequences of 55 residues**

Since we cannot enumerate all possible conformations for sequences of length 55 in order to identify a sequence with a native conformation which is kinetically accessible, we selected sequences for which long \((10^7\) MCS) simulations converged to a similar structure within a distance of 0.9 RMSD (i.e. found the same structure as minimal) in more than 5% of 200 runs. A total of 117 out of 1000 randomly chosen sequences with the residue composition described above, satisfied this criterion and were included in the set of 55-residue single-domains.

### 2.5 Computational platform

The computational requirements of this project are enormous and could not have been met in feasible time with a single or even several clusters. Thus, we performed all our simulations on the European GRID infrastructure, through the Enabling Grids for e-sciencE (EGEE) project. Briefly, GRID computing provides the ability to distribute high throughput computing on an infrastructure that virtually links an enormous memory capacity and thousands of CPUs. The computational challenge in this study was to perform a large number of repeated simulations (e.g. 250 simulations for 350 different sequences with 10 different modes of interactions with the chaperonins). By using the GRID platform, it was possible to distribute all the executions between hundreds of CPUs that are clustered at different locations (e.g. different universities around Europe) and dramatically increase the number of CPUs that are simultaneously used (e.g. 1600 computation hours are performed in less than a single day).

### 3 RESULTS

We used a lattice model to investigate the effects of chaperonins on the folding of different substrate proteins. In particular, we wished to determine the effect of the chaperonin cavity surface, which interacts with the substrate, on the yield of successfully folded substrate proteins. The fundamental measure used in this study is the improvement in the percentage of ‘successful’ simulations (simulations that found the native structure) for a given protein substrate. All protein substrates were first subjected to a few hundred simulations in the absence of chaperonins. We considered a run to be successful if during the folding process of a pre-defined duration (e.g. a simulation of \(10^6\) MCS) the native minimal structure was found (or a structure within RMSD <0.5 for a 25-residue sequence and RMSD <0.9 for a 55-residue sequence). The percentage of successful simulations, out of the total few hundred simulations executed for each protein, was defined as the **folding yield** of a protein. The same number of simulations under the exact same conditions (e.g. temperature, interaction potential, etc.) was then executed for each protein in the presence of a chaperonin. The difference between the folding yield with and without the chaperonin is defined as the improvement in the folding yield of the protein. The ratio between the folding yields with versus without the chaperonin is defined as the **improvement factor**. For instance, for a protein with a folding yield
of 20% in the absence of a chaperonin, and folding yield of 30% with a chaperonin, the improvement factor is 1.5. All analyses presented here were tested by paired t-test and were found to be statistically significant.

We start by exploring several basic aspects of protein substrate-chaperonin interactions in the case of 25-residue single-domains and then continue with more sophisticated models of 55-residue single- and double-domains.

### 3.1 Analysis of basic aspects of protein substrate–chaperonin interactions in the case of 25-residue single-domains

The simulations for the 123 sequences of length 25 were more than 90% successful at an ideal simulation temperature of 0.5. As was observed before (Betancourt and Thirumalai, 1999) in lattice simulations (and also in several experimental studies), the effect of the chaperonin on the folding yield under ideal folding conditions is minimal and sometimes even adverse. Thus, the effect of chaperonins must be evaluated under conditions that are less than ideal for folding. Hence, we simulated folding of these sequences at a higher temperature of 0.75 (\(T_f = 0.75\)). The increase in temperature resulted, as expected, in a lower percentage of successful simulations. In the presence of a chaperonin, we observed a significant improvement in the folding yield when folding of the substrate was simulated inside a chaperonin that undergoes concerted surface changes and by invoking a binding-release mechanism (Fig. 7). In contrast, no significant improvement was obtained when substrate sequences were folded inside a chaperonin that undergoes sequential surface changes.

Another fundamental aspect we investigated was the effect of chaperonin cavity size on the folding yield of substrate proteins. We expected that the action of caging a protein would reduce its entropy and thereby improve the process of finding the structure with the minimal energy (which is usually also compact). Figure 8 illustrates the effect of the size of the chaperonin cavity on the folding yield of substrate proteins. As expected, smaller cavities of chaperonins result in better folding yield improvements. A square-shaped cavity was used and results for the binding and release mechanism are presented.

### 3.2 Analysis of protein–substrate chaperonin interactions on 55-residue single-domain substrates

In lattice simulations, longer sequences require longer MC runs to fold. Hence, the 117 sequences of 55 residues were simulated using \(10^7\) MCS runs.

In order to include in this analysis the standard error of the result of a series of simulations for a given sequence, we performed the following approximation. The standard error for each sequence is estimated from the binomial distribution.

\[
\sigma = \sqrt{\frac{\hat{p}(1 - \hat{p})}{N}}
\]

where \(\hat{p}\) is the estimated probability of a successful simulation (folding yield) out of the \(N\) independent simulations
chaperons undergo a sequential mode of surface changes (Rivenzon-Segal et al., 2005) and it was, therefore, suggested that these two phenomena are linked. Thus, we next investigated whether the effect of a chaperonin that undergoes sequential surface changes is stronger on double-domain versus single-domain structures. To this end, 104 sequences that are 55 residues long and have a homo-double domain native structure were simulated for $10^7$ MCS. The results in Figure 10 show that, in contrast with the results obtained for the 25- and 55-residue single-domain substrate sequences, the effect of a chaperonin with sequential surface behavior on the homo-double-domain substrates was as strong as that of the concerted chaperonin.

### 3.4 Comparison between the effects of a chaperonin with sequential surface changes on single-domain and double-domain substrates

In order to compare the distinct averaged effects on single-domain and double-domain substrate sets on a single scale, we needed to normalize the results. We define the improvement factor as the ratio of folding yield (the fraction of successful simulations) in the presence of a chaperonin versus the folding yield without a chaperonin. Figure 11 shows the normalized results. It is clear that, in the presence of a chaperonin with
Different mechanistic requirements for prokaryotic and eukaryotic chaperonins: a lattice study

Fig. 11. The effects of concerted and sequential chaperonins on 55-residues single-domain and double-domain substrates. The graph plots the improvement in folding yield as a function of MCS. The factor is normalized relative to the maximal yield of successful simulations of each group since each group has a different average folding yield. A structure of RMSD <0.9 from the native structure was considered a success for both groups. In addition, these results apply to the same octagonal chaperonins with side of size 7 using a binding and release mechanism. These data show that single-domain folding is better facilitated by a concerted surface change, while double-domain folding is better facilitated by a sequential surface change. All differences shown here are statistically significant according to paired t-test analysis.

sequential surface behavior, there is a significant improvement in folding yield of double-domain substrates but only a slight improvement in the case of single-domain substrates. In contrast, the effect of a chaperonin with concerted surface behavior is significantly larger in the case of single-domain substrates than double-domain substrates.

4 DISCUSSION

The prokaryotic chaperonin GroEL undergoes ATP-driven concerted conformational switching between a protein acceptor (T) state and a protein-release (R) state (Horovitz and Willison, 2005). A major difference between these states is that the surface of the chaperonin’s folding chamber is hydrophobic in the T state (thus favoring non-folded protein substrate binding) (Braig et al., 1994) and hydrophilic in the R state thus favoring protein substrate release (Ranson et al., 2001). In contrast, it has been shown that the eukaryotic chaperonin CCT undergoes ATP-driven conformational changes that are sequential (Rivenzon-Segal et al., 2005). In this work, we tested possible implications of these different allosteric mechanisms for the folding function of these chaperonins using 2D lattice simulations. These simulations involve many simplifications. Our model consists of only four types of residues (i.e. amino acids) for the substrate and only two types (P and H) for the chaperonin surface and is 2D rather than 3D. Also, the relative sizes of the substrate and the chaperonin cavity are not realistic.

However, intra-protein substrate and substrate–chaperonin interactions were treated in the same way and our simulations of assisted folding by the concerted and sequential chaperonins involved the same simplifications. In addition, our simulations were carried out on hundreds of different sequences and they reproduced expected effects of confinement and temperature on folding yields. Thus, we suggest that our results may have captured real properties of chaperonin-mediated folding in prokaryotic and eukaryotic systems.

We found that the folding yields of single-domain protein substrates (both 25- and 55-residue sequences) are greater when the chaperonin undergoes concerted but not sequential conformational changes. In contrast, the folding yields of double-domain proteins are greater in the presence of a chaperonin that undergoes either sequential or concerted conformational changes (Fig. 11). It is likely that single-domain protein folding is adversely affected by a sequential mechanism since non-local native interactions cannot be formed if not all parts of the protein are released from the chaperonin. In the case of double-domain proteins, there is a trade-off between this effect and prevention of non-native contacts between domains. Our results are consistent with findings that indicate that large multi-domain proteins are more common in eukaryotes compared with prokaryotes. Hence, they support the suggestion (Rivenzon-Segal et al., 2005) that the different allosteric mechanisms of prokaryotic and eukaryotic chaperonins can be explained by the need of eukaryotic chaperonins to facilitate folding of proteins with a multi-domain structure.

ACKNOWLEDGEMENTS

We thank Inna Myslyuk for assistance with the artwork and Ehud Schreiber for the useful comments. The computations in this research were performed on the EGEE grid and we thank Assaf Gottlieb for technical support. This work was supported by Israel Science Foundation grant (67/05) to A.H.

Conflict of Interest: none declared.

REFERENCES


