Simulations of Chaperone-Assisted Folding†

C. D. Sfatos, A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich*

Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Received August 25, 1995; Revised Manuscript Received November 1, 1995⊗

ABSTRACT: We investigated a chaperone mechanism of protein folding using a 36-mer model on a cubic lattice. The mechanism simulates folding, which proceeds with repetitive cycles of binding, unfolding, and releasing of misfolded metastable states. We measured the yield enhancement due to this mechanism for sequences selected by evolutionary design and showed that the binding and releasing mechanism is efficient for the yield enhancement of folding for sequences that are poorly designed, *i.e.*, where selection is not adequately strong. From this it follows that the chaperone mechanism can be considered as the evolutionary alternative to compensate for poor sequence design. On the other hand, random sequences show a decrease in yield and no effect on the total mean first passage time when the proposed chaperone mechanism is implemented, thus implying that sequence optimization is a necessary condition for the efficiency of the proposed mechanism. We qualitatively reproduced experimental results for folding in the presence of GroEL/GroES, fit our results with the aid of a double-exponential model of folding kinetics, and characterized the conditions under which this mechanism of chaperone action affects folding.

The classic experiments of Anfinsen (1973) revealed that denatured proteins can refold to the native state *in vitro* and, consequently, verified that all of the information required for folding into the native state is encoded in the primary structure of a protein. Recently it was observed that some proteins with large molecular weights are not able to fold *in vitro* in the absence of a class of proteins that appears to catalyze the folding process and that is present in the *in vivo* folding. Instead, aggregates of misfolded intermediates appear. These proteins have been characterized as molecular chaperones and facilitate folding in a way that always requires the hydrolysis of ATP (Ellis *et al*., 1993; Gething & Sambrook, 1992; Hendrick & Hartl, 1993; Boshkareva *et al*., 1988).

One important and extensively studied family of chaperones is the chaperonin ring family. A typical representative of this family is GroEL, which is known to bind non-native conformations of different proteins in its central cavity (Saibil *et al*., 1991) and release them through a mechanism that increases the yield in native form. One model for the mechanism of action of GroEL/GroES proposes (Jackson *et al*., 1993) that metastable misfolded intermediates bind to GroEL through their exposed hydrophobic surface (Tandon & Horowich, 1989; Mendoza *et al*., 1991; Landry *et al*., 1992). They then undergo specific conformational changes and are released in some conformation that is committed to fold to the native state. There is strong evidence that GroEL has one conformation with high affinity for binding the protein substrate and then, with hydrolysis of ATP, it undergoes a structural change to another conformation that has a low affinity for the substrate (Jackson *et al*., 1993; Todd *et al*., 1994). However, it is not clear whether the released protein conformation resembles that of the native or the unfolded state.

On the other hand, recent experiments, *e.g*., on monomeric rhodanese (Weissman *et al*., 1994) and dimeric rubisco (Todd *et al*., 1994), accumulate evidence supporting a mechanism of multiple rounds of binding and releasing of non-native forms. On the basis of evidence that protein folding proceeds through independent slow and fast phases (Udgaonkar & Baldwin, 1990; Radford *et al*., 1992; Bycroft *et al*., 1990), it is proposed (Todd *et al*., 1994; Weissman *et al*., 1994) that, through many rounds of binding and releasing, the protein can fold through a fraction of fast folding events undergoing kinetic proofreading (Hopfield, 1974; Gulukota & Wolynes, 1994). Long-lived misfolded intermediates can undergo structural changes that will remove them from the trap and allow them to refold, so that with a finite probability they proceed through a fast phase.

However, different behavior is observed among proteins whose folding is affected by the presence of chaperones. These are proteins such as rhodanese (Martin *et al*., 1991), for which the yield in native form under spontaneous refolding conditions is insignificant but reaches 75% when folding is catalyzed by chaperonins. Other proteins such as dehydrofolate reductase (DHFR) (Martin *et al*., 1991; Viitanen *et al*., 1991), which fold spontaneously with high efficiency, improve their yield very little by binding to GroEL. Although it is known that protein substrates bind directly on GroEL, the presence of another protein, GroES, is often essential for refolding.

In the present study, we differentiate between proteins by the extent of optimization performed through an evolutionary sequence design process. We find that a certain degree of evolutionary selection is a necessary condition for the efficiency of the binding and releasing model. We demonstrate that this chaperone mechanism is important for the sequences that are poorly designed, so that chaperones can act as a remedy to the cases where further evolutionary selection is not possible.

MODEL

Interest in the theoretical study of the effect of chaperones has already been demonstrated (Gulukota & Wolynes, 1994; Thirumalai, 1994). Here we describe a numerical experiment

† This work was supported by the Packard Foundation.
⊗ Abstract published in *Advance ACS Abstracts,* December 15, 1995.

0006-2960/96/0435-334$12.00/0

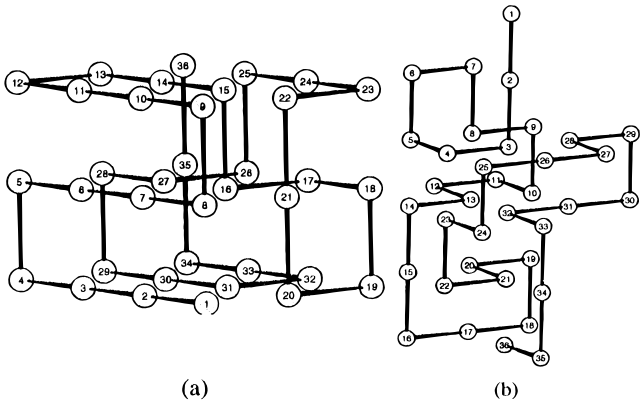

FIGURE 1: (a) Maximally compact structure with 40 contacts for a 36-mer on a cubic lattice. Folding of a well-designed (NKTVVGEPWHCLLFPRRDKNQMSYLTGIAGEDSAAI) and a poorly designed sequence (TNLWALNGRDFAVLKTQKPRDESECGIIMVHYSPAG) to this native structure has been tested in this work. (b) Less compact structure with only 28 out of 40 possible contacts used as the native structure for the random sequence (LLNKECIPHEVAQKWTGSRFTLDINYVDGSPRMAAG).

on a lattice model of proteins by adopting the model with rounds of binding and releasing (Chan & Dill, 1995; Todd et al., 1995). We model a protein sequence by a self-avoiding walk of 36 residues on a cubic lattice (Abkevich *et al*., 1994a). The amino acids on the sequence interact through contact potentials, and the total energy of each conformation on the lattice is given by

$$E(\{\sigma_i\}) = [B_0 + \sum_{i<j}^{N} U(\sigma_i\sigma_j)]\Delta_{ij} \qquad (1)$$

The identity of each amino acid in the sequence is represented by a parameter σ_i, and the interaction energy matrix $U(\sigma_i\sigma_j)$ between amino acids is taken from Miyazawa and Jernigan (1985). The function Δ_{ij} represents two residues i and j in contact on a cubic lattice as, for example, residues 18 and 23 in Figure 1a, where 18 and 24 are not in contact. We also omit the contribution of interactions between consecutive monomers along the chain, because they contribute only by a constant factor identical for all conformations. The overall attraction factor B_0 only contributes toward the compactness and stability of the ground state.

We designed the selected sequences on the basis of the following plan: we select a maximally compact structure shown in Figure 1a. Once a structure is randomly selected and labeled as "native", a design algorithm (Shakhnovich & Gutin, 1993), as outlined in the Appendix, is used to select sequences that deliver low energy to this "native" structure compared to unfolded or misfolded structures. Folding is started from a random coil conformation and proceeds through a Monte Carlo set of moves (Abkevich *et al*., 1994b).

The goal of selection through sequence design is to make native structures more stable thermodynamically. Extensive studies of this kind of designed sequences for the 36-mer on the cubic lattice showed that design accelerates folding. These studies show direct correlation between the energy of the "native" state for a particular designed sequence and the mean folding time [Figure 9b in Abkevich *et al*. (1994b); Sali *et al*., 1994; Socci & Onuchic, 1994): the lower the energy the faster the folding time. On the basis of this correlation, we characterize sequences with greatly enhanced stability of the "native" state as "well-designed", where sequences with small enhancement are called "poorly designed". The criteria and optimization scores for the sequences used as representatives throughout this work are given in the Appendix. Sequences that are chosen without any optimization and labeled as "random" correspond to higher energies of the "native" conformation and have much slower folding times.

For the purpose of investigating the chaperone mechanism, however, the most striking feature of designed sequences is that the distribution of folding times is nonexponential (Abkevich *et al*., 1994a; Guo & Thirumalai, 1995). This implies that the folding of designed sequences proceeds through independent multiple phases in which a fraction of the runs rapidly fold to the native state, where others fold at very long times through metastable traps. In that sense, the folding kinetics of these sequences resembles the folding kinetics of studied proteins (Udgaonkar & Baldwin, 1990; Radford *et al*., 1992; Bycroft *et al*., 1990). The appearance of nonexponentiality observed for designed sequences is related to this slow/fast phase competition. It must be noted that these two phases are due to independent pathways, and every run (molecule) can follow either one. This kind of interplay between the slow and fast pathways of folding should be differentiated from the effect of proline isomerization and the observed nonexponentiality due to this effect. In that case, molecules that contain *cis*-proline cannot follow the fast folding pathway unless they isomerize. Therefore, the proposed chaperone mechanism can have no effect on such nonexponentiality.

The numerical experiment performed here involving the chaperone binding and releasing mechanism allows a 36-mer to attempt folding up to a certain window time, t_w. If the molecule has not reached the native state before t_w, it is unfolded and given another try from a different random starting conformation. This t_w corresponds to the typical time that a protein molecule can spend free in solution between consecutive bindings on a chaperone.

The moment at which we reach this t_w without having reached the native state corresponds, in terms of the real biological problem, to the moment when an unfolded or misfolded protein molecule encounters a chaperone molecule and binds to it. According to the model of Todd *et al*. (1994) and Weissman *et al*. (1994), the protein molecule undergoes a conformational change, so that it is released in a less folded state and with high probability is removed from a trap. Then the molecule has another chance to fold through a fast folding pathway, if such a pathway exists, instead of remaining in the trap for a very long time and eventually aggregating.

In our model, the equivalent plan of action for the implemented chaperone mechanism resides in the fact that, when a conformation other than native reaches t_w, the molecule is unfolded and the folding attempt continues from a different randomly chosen unfolded conformation. This instantaneous unfolding corresponds to the structural changes that are proposed in the model of Todd *et al*. (1994) and Weissman *et al*. (1994). In the real system, such encounters of the misfolded protein with the chaperone molecule are more frequent as the concentration of the chaperone increases. Equivalently, in our numerical model the frequence of these encounters increases as the window time t_w decreases. This analogy allows qualitative comparisons between our results and the corresponding experiments with real chaperones.

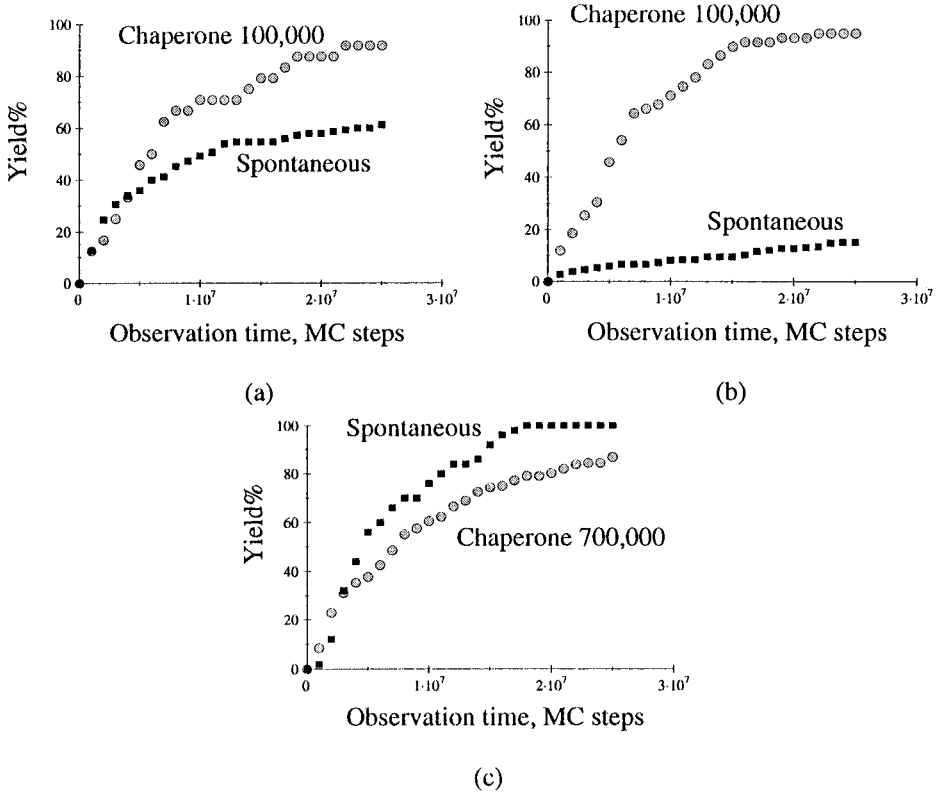

FIGURE 2: (a) Reaction profile for the well-designed sequence (see Figure 1a) for spontaneous folding and chaperone folding with a window (binding period) $t_w = 100\,000$ MC steps at temperature $T = 0.15$ and overall attraction $B_0 = 0.05$. (b) Reaction profile for the poorly designed sequence (see Figure 1a) for spontaneous folding and chaperone folding with a window $t_w = 100\,000$ MC steps at temperature $T = 0.15$ and overall attraction $B_0 = 0.05$. (c) Reaction profile for the random sequence (see Figure 1b) for spontaneous folding and chaperone folding with a window $t_w = 700\,000$ MC steps at temperature $T = 0.15$ and overall attraction $B_0 = 0.0$.

The final folding time is evaluated from one successful folding into the native conformation to the next. For every given observation time t_{ob}, we examine the percentage of those runs that fall within t_{ob} and calculate the yield.

RESULTS

The reaction profile of yield against observation time in Monte Carlo (MC) steps is given in Figure 2a,b for two sequences folding into the same "native" structure (Figure 1a) at the same temperature $T = 0.15$ and overall attraction $B_0 = 0.05$. Results are presented for folding with the chaperone mechanism with the same $t_w = 10^5$ MC steps and also for spontaneous folding for each sequence. The well-designed sequence (Figure 2a), which has very low energy for the native conformation relative to misfolded conformations, folds very efficiently in spontaneous folding conditions, and the chaperone binding and releasing mechanism improves the yield very little. It should also be noted that at early times the chaperone yield is lower, as in the case of DHFR (Martin *et al.*, 1991). On the contrary, the poorly designed sequence (Figure 2b), with higher relative energy of the native conformation, in which spontaneous folding provides a very low yield within a certain t_{ob}, dramatically improves in yield due to the kinetic proofreading of the chaperone mechanism. The yield for the poorly designed sequence as a function of the window size at a fixed observation time $t_{ob} = 25 \times 10^6$ MC steps is given in Figure 3. We see that a very wide range of t_w values produces a large enhancement in yield. The average time of the successful chaperone foldings is longer than the average time of the spontaneous foldings within the same observation time (results not shown).

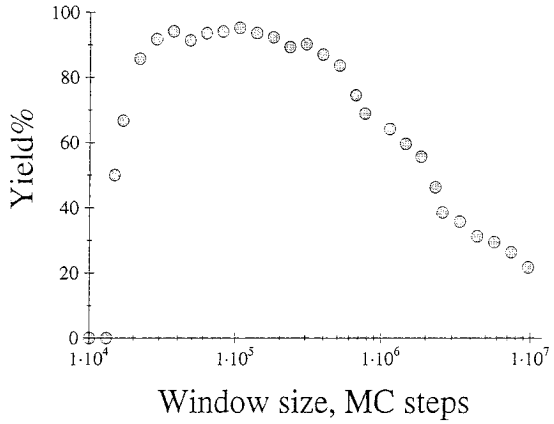

FIGURE 3: Yield as a function of the window t_w (binding period) for the poorly designed sequence at an observation time $t_{ob} = 25 \times 10^6$ MC steps at temperature $T = 0.15$ and $B_0 = 0.05$. As the window size increases to 25×10^6 MC steps, the yields decrease to the spontaneous one (Figure 2b).

It is also interesting to examine the effect of chaperones on random sequences. We took a random sequence and found the structure with the lowest energy, which we labeled as "native", with 28 instead of 40 native contacts shown in Figure 1b. At all temperatures examined, we found that chaperones have a small but strictly retarding effect, as shown in Figure 2c.

The discrimination in the behavior of different kinds of sequences under the influence of the chaperone mechanism is a crucial result of this work. Previously, a similar mechanism of kinetic proofreading applied to a numerical spin model resulted in an indiscriminate increase in yield, even for random sequences (Gulukota & Wolynes, 1994).

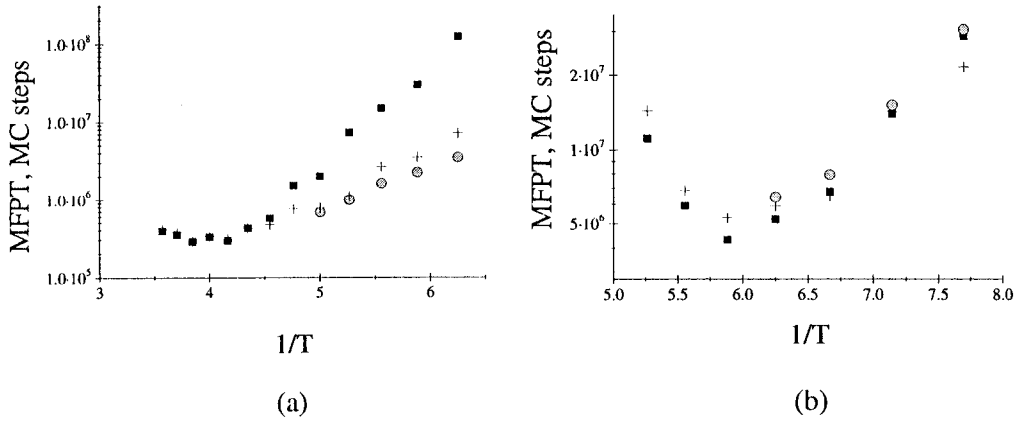

FIGURE 4: (a) Mean first passage time (MFPT) (■) against $1/T$ for the well-designed sequence at $B_0 = 0.05$. The Arrhenius behavior starts at about $T = 0.23$. The MFPT for chaperone folding (●) shows significant acceleration because of the acceleration of rare but very strong traps. The chaperone window is $t_w = 5 \times 10^5$ MC steps. At this window size, MFPT acceleration is optimum (see Figure 5). The deviation from single-exponential behavior is indicated by the deviation of $t_{1/2}/\ln 2$ from MFPT for the spontaneous folding (+). (b) Same results for the random sequence (Figure 1b). We see that the chaperone mechanism does not affect the MFPT because of the single-exponential behavior. The results are shown for $t_w = 1.5 \times 10^6$ MC steps, which corresponds to the fastest of the results observed. The positive deviation is reminiscent of the effect of lag time.

Our results can be explained by proposing the following simple scheme: As mentioned earlier, designed sequences follow nonexponential kinetics with independent fast and slow phases of folding. Well- and poorly designed sequences may have almost equally rapid fast phases of folding, but well-designed sequences have a much larger amplitude of the fast phase and therefore fold efficiently in the absence of chaperones. This corresponds to the fact that traps are relatively rare and spontaneous folding works very efficiently in well-designed sequences, although a few folding attempts that go through the slow phase can result in folding times that are orders of magnitude longer than the average time. We see here how, in the case of poorly designed sequences, the proposed chaperone mechanism acts as a repair mechanism to compensate for the weak evolutionary design.

In order to support this scheme, we studied the behavior of the mean first passage time (MFPT) with temperature for the kinds of sequences examined here. We allowed all runs to reach the native state and then calculated the average folding time. The results for the well-designed sequence are shown in Figure 4a. We see that below some temperature T_f the system exhibits Arrhenius behavior, *i.e.*, the logarithm of MFPT becomes a linear function of the inverse temperature. This implies that below this temperature the energy of the barrier does not depend on temperature.

From these data, we can demonstrate that designed sequences follow nonexponential behavior for temperatures slightly below T_f. For this purpose, we calculate the median time $t_{1/2}$, the time required for successful folding of 50% of the attempted spontaneous runs. For a single-exponential time distribution, $t_{1/2}$ divided by ln 2 must be equal to the MFPT. Deviation from this rule signals more than one exponential phase. This deviation from single-exponential kinetics is shown in Figure 4a and is a property of designed sequences only (V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, unpublished results). The random sequence tested earlier follows, on the basis of this criterion, single-exponential kinetics as shown in Figure 4b. The multiexponential behavior is a result of fast and slow phases of folding, and to the lowest approximation we can consider them as one fast and one slow exponential phase. There is also a small lag time, possibly related to a fast compactization stage, so that very short times are not observed.

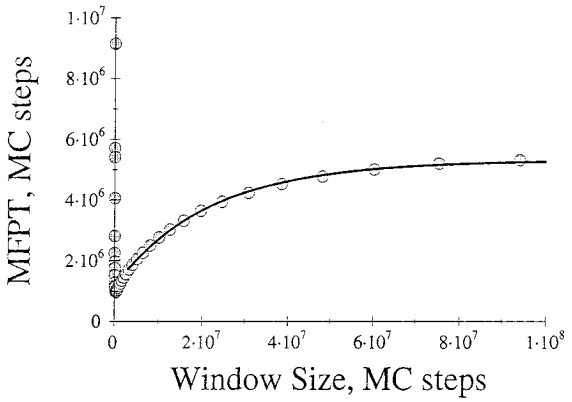

FIGURE 5: Chaperone MFPT against window size for the well-designed sequence at $T = 0.19$. For a very large window, chaperone MFPT relaxes to the spontaneous time (Figure 4a) at this temperature. For very small times it increases drastically due to a lag time period. The fit of eq 3 is shown for $f = 4.52$, $\tau_1 = 1.09 \times 10^6$, and $\tau_2 = 2.437 \times 10^7$.

Folding with the chaperone mechanism of binding and releasing shows a dramatic decrease in the MFPT, up to 2 orders of magnitude, for a designed sequence with $t_w =$ 500 000 MC steps (Figure 4a), while it shows no effect at all for the random sequence (Figure 4b). This behavior of random sequences is also in contrast with the results in Gulukota and Wolynes (1994).

The plot of Figure 5 shows MFPT for folding of the well-designed sequence at temperature $T = 0.19$ as function of the chaperone window size t_w. We see that for a very large window size the MFPT relaxes to the spontaneous folding time. For very short window size, the existence of the lag time period decreases the number of successful foldings in this short period and increases MFPT sharply. For intermediate windows we see a minimum far below the spontaneous MFPT, which corresponds to the acceleration effect of the chaperone mechanism on MFPT. A short chaperone window time t_w corresponds to a high concentration of the chaperone and a large binding frequency. The results of Figures 3 and 5 are in agreement with the results for

reconstitution of native rubisco (Goloubinoff *et al*., 1989) for different concentrations of the chaperone.

DOUBLE-EXPONENTIAL MODEL

For any arbitrary time distribution of folding time $P(t)$, the average chaperone time, $\langle t_{ch} \rangle$, as a function of the window size is given by

$$\langle t_{ch} \rangle = (P_-/P_+)t_w + \langle t_+ \rangle \quad (2)$$

where P_- and P_+ are the probabilities that a given run will not or will be completed, respectively, within the given window time t_w, and $\langle t_+ \rangle$ is the average time for the runs that are completed within t_w. For a single-exponential distribution with a decay time τ, this formula gives $\langle t_{ch} \rangle = \tau$, *i.e*., chaperones do not have any effect on single-exponential kinetics. If we ignore the lag time effect, for large window size we can represent our nonexponential process with a double-exponential distribution. Then $\langle t_{ch} \rangle$ is given by

$$\langle t_{ch} \rangle = \frac{f\tau_1(1 - e^{-t_w/\tau_1}) + \tau_2(1 - e^{-t_w/\tau_2})}{f(1 - e^{-t_w/\tau_1}) + (1 - e^{-t_w/\tau_2})} \quad (3)$$

where τ_1 and τ_2 are the two time scales of the proposed double-exponential time distribution, and f is the ratio of the fast to the slow phase amplitudes.

By using this equation, we made a fit of the data of Figure 5 for large window size where the lag time does not affect the result and the double-exponential process should be adequate (Abkevich *et al*., 1994a). The fit gave $f \sim 4.5$, $\tau_1 \sim 1 \times 10^6$ MC steps, and $\tau_2 \sim 2.4 \times 10^7$ MC steps. This fit reveals the existence of a fast and a slow phase where the fast phase has a 4.5 times larger amplitude. This means that it is 4.5 times more likely that this well-designed sequence will fold through a fast rather than through a slow folding pathway. For the poorly designed sequence, we were able to estimate $f \sim 0.1$ by fitting the data for a short range of t_w after the optimal window; if we assume $\tau_1 \ll t_w \ll \tau_2$, we can estimate from eq 3 that $\langle t_{ch} \rangle \sim t_w/f$.

It is clear that the action of chaperones eliminates, to a large extent, the slow phase. As far as MFPT is concerned, well-designed sequences are relieved from folding through rare but strong traps that greatly contribute to the MFPT in spontaneous folding without chaperones.

The experimentally interesting parameter for proteins, however, is the yield at a given observation time because, due to aggregation, not all folding events are allowed to go to completion. When the system does not follow single-exponential kinetics, the yield depends on the ratio of the amplitude of the fast phase to the amplitude of the slow phase. The chaperone mechanism is more efficient in increasing the yield under conditions for which this ratio is small. Under these conditions, the spontaneous folding yield is very small as in the case of poorly designed sequences. Inversely, when this ratio is large, spontaneous folding is very efficient as in the case of well-designed sequences. When the slow phase is eliminated to a large extent, due to the action of chaperones, the increase in yield is apparently larger for the first case with the smaller ratio of fast to slow phase amplitude.

In addition to a small value of the amplitude ratio f, the two decay times must be well separated. It was confirmed, through the $t_{1/2}$ test, that the chaperone folding times follow single-exponential kinetics. For $t_w \lesssim \tau_1$, we can estimate that for large f (good design) $\langle t_{ch} \rangle \sim \tau_1$, while for $f \ll 1$ (bad design) $\langle t_{ch} \rangle \sim \tau_1/f$. The chaperone mechanism will be effective as long as $\langle t_{ch} \rangle \gtrsim \tau_2$, *i.e*., as long as $f \gg \tau_1/\tau_2$. From this analysis it is seen that for $\tau_1/\tau_2 \ll f \ll 1$ the performance of the chaperone mechanism will be significantly more efficient than the spontaneous double-exponential process.

Furthermore, we can estimate a maximum performance in yield enhancement for $\tau_1/\tau_2 \ll f \ll 1$ as follows. For $\langle t_{ch} \rangle \sim \tau_1/f \ll t_{ob} \ll \tau_2$, the chaperone yield is almost 100%. The yield enhancement will be $r \sim 1/(f + t_{ob}/\tau_2)$. We can distinguish two limiting cases: for $(\tau_1/\tau_2)^{1/2} \ll f \ll 1$, the yield enhancement will be $r \sim 1/f$; for $(\tau_1/\tau_2) \ll f \ll (\tau_1/\tau_2)^{1/2}$, the enhancement is $r \sim \tau_2 f/\tau_1$. Between these two limiting cases there is an optimum at $f \sim (\tau_1/\tau_2)^{1/2}$. The corresponding yield enhancement is $r \sim (\tau_2/\tau_1)^{1/2}$.

CONCLUSIONS

In this work the efficiency of the binding and releasing mechanism of chaperone-assisted folding has been tested for sequences with different degrees of evolutionary optimization. It was demonstrated that poorly designed sequences are the ones that benefit the most in yield enhancement.

In the traditional sense, a well-designed sequence is one that delivers very low energy to the native conformation and provides numerous fast folding pathways. A poorly designed sequence uses the binding and releasing chaperone mechanism to compensate for the weak evolutionary sequence selection. Therefore, we can propose that the chaperone mechanism evolves as an alternative to allow for folding to occur in cases where the traditional sequence optimization is not a preferred or possible evolutionary route.

It is also shown that unoptimized random sequences cannot benefit from such a mechanism because of the lack of any distinct fast folding pathway. Evolutionary design is the factor that allows for multiexponential kinetics, with well-separated decay times between fast and slow phases. Therefore, evolutionary selection is a necessary condition for the efficiency of the binding and releasing mechanism.

It is reasonable to propose that the parameters that determine the enhancement in yield, such as the amplitude ratio f and the separation between the different decay times of the spontaneous multiexponential process, may also depend on the chain length. One may think that it is harder to design long sequences with large f, so that the assistance of chaperones is badly needed in the folding of long proteins. Such studies of the effect of the length are in progress.

It can be seen from Figure 3 that very small window times t_w result in a decrease in the yield. Small t_w means a higher frequency of restarting, which, as explained in the Model section, corresponds to a higher binding frequency, *i.e*., higher concentration of the chaperone in the real biological problem. The trend described in Figure 3 is in agreement with the results of Figure 3b of Goloubinoff *et al*. (1989), where the yield in native rubisco decreases with an increasing concentration of GroEL.

One must note that certain proteins with efficient spontaneous folding, such as (DHFR) (Martin *et al*., 1991), show little increase in yield and a retardation effect upon the action of chaperones like the one shown for well-designed sequences (Figure 2b), as well as no drastic effect of GroES.

Within the framework of the binding and releasing mechanism, GroES can be held responsible for altering the conformation of the misfolded intermediate so that it can be removed from the trap. For proteins like DHFR that fold efficiently in spontaneous conditions, GroES does not have any major effect because it assists only the very small fraction of folding attempts that proceed through the small-amplitude slow phase. In proteins like rhodanese, rubisco, and citric synthase (Buchner *et al*., 1991), GroES is necessary for a significant increase in yield in order to liberate the chain from frequent traps.

ACKNOWLEDGMENT

We are very grateful to O. B. Ptitsyn for suggesting that we simulate the kinetic proofreading mechanism (Todd *et al*., 1994; Weissman *et al*., 1994) on the lattice model.

APPENDIX: THE DESIGN PROCEDURE

We selected a maximally compact structure shown in Figure 1a. Once a structure is randomly selected and labeled as "native", a design algorithm (Shakhnovich & Gutin, 1993) is used to select sequences that deliver low energy to this "native" structure compared to unfolded or misfolded structures. Selection is based upon maximization of the parameter

$$z = \frac{E_{av} - E_{nat}}{\sigma} \qquad \text{(A1)}$$

where σ is the standard deviation of energies of all contacts. The parameter is referred to as the "*z*-score" (Bowie *et al*., 1991). The *z*-score value for the well-designed sequence is 56.5 and that for the poorly designed sequence is 43.6, where the interaction energies are calculated from Miyazawa and Jernigan (1985).

Folding starts from a random coil conformation and proceeds through a Monte Carlo set of moves that contains corner flips, crankshaft moves, and tail moves (Abkevich *et al*., 1994b).

REFERENCES

Abkevich, V. I., Gutin, A. M., & Shakhnovich, E. I. (1994a) *J. Chem. Phys. 101*, 6052−6062.
Abkevich, V. I., Gutin, A. M., & Shakhnovich, E. I. (1994b) *Biochemistry 33*, 10026−10036.
Anfinsen, C. B. (1973) *Science 181*, 223−230.
Boshkareva, E. S., Lissin, N. M., & Girshovich, A. S. (1988) *Nature 336*, 254−257.
Bowie, J. U., Luthy, R., & Eisenberg, D. (1991) *Science 253*, 164−170.
Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., & Kiefhaber, T. (1991) *Biochemistry 30*, 1586−1591.
Bycroft, M., Matouschek, A., Kellis J., Serano, & Fersht, A. (1990) *Nature 346*, 488−490.
Chan, H. S., & Dill, K. A. (1995) *Proteins* (in press).
Ellis, R. J., Laskey, R. A., & Lorimer, G. H., Eds. (1993) *Molecular Chaperones*, Chapman and Hall, London.
Gething, M. J., & Sambrook, J. (1992) *Nature 355*, 33−45.
Goloubinoff, P., Christeller, J. T., Gatenby, A. A., & Lorimer, G. H. (1989) *Nature 342*, 884−889.
Gulukota, K., & Wolynes, P. G. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9292−9296.
Guo, Z. Y., & Thirumalai, D. (1995) *Biopolymers 35*, 137−140.
Hendrick, J. P., & Hartl, F.-U. (1993) *Annu. Rev. Biochem. 62*, 349−384.
Hopfield, J. J. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4135−4139.
Jackson, G. S., Staniforth, R. A., Halsall, D. J., Atkinson, T., Holbbrook, J. J., Clarke, A. R., Burston, S. G. (1993) *Biochemistry 32*, 2554−2563.
Landry, S. J., Jordan, R., McMaken, R., & Gierasch, L. M. (1992) *Nature 355*, 455−457.
Martin, J., Lange, T., Boteva, R., Schramel, A., Horwich, A. L., & Hartl, F.-U. (1991) *Nature 352*, 36−42.
Mendoza, J. A., Rogers, E., Lorimer, G. H., & Horowich, P. M. (1991) *J. Biol. Chem. 266*, 13044−13049.
Miyazawa, S., & Jernigan, R. (1985) *Macromolecules 18*, 534−552 (Table VI).
Radford, S., Dobson, C., & Evans, P. (1992) *Nature 358*, 302−307.
Saibil, H., Dong, Z., Wood, S., & auf der Mauer, A. (1991) *Nature 353*, 25−26.
Sali, A., Shakhnovich, E. I., & Karplus, M. (1994) *J. Mol. Biol. 235*, 1614−1636.
Shakhnovich, E. I., & Gutin, A. M. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7195−7199.
Socci, N. D., & Onuchic, J. N. (1994) *J. Chem. Phys. 101*, 1519−1528.
Tandon, S., & Horowich, P. M. (1989) *J. Biol. Chem. 264*, 9859−9866.
Thirumalai, D. (1994) *Statistical Mechanics, Protein Structure, and Protein-Substrate Interactions* (Doniach, S., Ed.) Plenum, New York.
Todd, M. J., Viitanen, P. V., & Lorimer, G. H. (1994) *Science 265*, 659−666.
Todd, M. J., Thirumalai, D., & Lorimer, G. H. (1995) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Udgaonkar, J. B., & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8197−8201.
Viitanen, P. V., Donaldson, G. K., Lorimer, G. H., Lubben, T. H., & Gatenby, A. A. (1991) *Biochemistry 30*, 9716−9723.
Weissman, J. S., Kashi, Y., Fenton, W. A., & Horwich, A. L. (1994) *Cell 78*, 693−702.

BI952033A