Multiple Parallel-pathway Folding of Proline-free Staphylococcal Nuclease

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · NOVEMBER 2003

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2003.07.002 · Source: PubMed

4 AUTHORS, INCLUDING:

Masaru Tanokura
The University of Tokyo
463 PUBLICATIONS **7,114** CITATIONS
SEE PROFILE

Kunihiro Kuwajima
The University of Tokyo
157 PUBLICATIONS **7,598** CITATIONS
SEE PROFILE

doi:10.1016/j.jmb.2003.07.002 *J. Mol. Biol.* (2003) **332**, 1143–1153

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Multiple Parallel-pathway Folding of Proline-free Staphylococcal Nuclease

Kiyoto Kamagata[1], Yoriko Sawano[2], Masaru Tanokura[2] and Kunihiro Kuwajima[1]*

[1]*Department of Physics Graduate School of Science University of Tokyo 7-3-1 Hongo, Bunkyo-ku Tokyo 113-0033 Japan*

[2]*Department of Applied Biological Chemistry Graduate School of Agricultural and Life Sciences University of Tokyo 1-1-1 Yayoi, Bunkyo-ku Tokyo 113-8657 Japan*

**Corresponding author*

When a protein exhibits complex kinetics of refolding, we often ascribe the complexity to slow isomerization events in the denatured protein, such as *cis/trans* isomerization of peptidyl prolyl bonds. Does the complex folding kinetics arise only from this well-known reason? Here, we have investigated the refolding of a proline-free variant of staphylococcal nuclease by stopped-flow, double-jump techniques, to examine the folding reactions without the slow prolyl isomerizations. As a result, the protein folds into the native state along at least two accessible parallel pathways, starting from a macroscopically single denatured-state ensemble. The presence of intermediates on the individual folding pathways has revealed the existence of multiple parallel pathways, and is characterized by multi-exponential folding kinetics with a lag phase. Therefore, a "single" amino acid sequence can fold along the multiple parallel pathways. This observation in staphylococcal nuclease suggests that the multiple folding may be more general than we have expected, because the multiple parallel-pathway folding cannot be excluded from proteins that show simpler kinetics.

Keywords: protein folding; double-jump; pathway; intermediate; isomerization

Introduction

It has been expected that protein folding may become much simpler if any slow isomerizations, like peptidyl prolyl *cis/trans* isomerization, in the denatured state are eliminated, and the complex kinetics of folding caused by kinetic partitioning, which is reported in lysozyme and dihydrofolate reductase (DHFR), might be rather exceptional.[1–4] However, the direct folding of proteins without the slow prolyl isomerizations has not been studied in detail. Probably, the only exception is a proline-free variant of tendamistat, which undergoes slow isomerizations of non-prolyl peptide bonds in the denatured state and exhibits complex folding kinetics.[5] Is the folding of proteins without any slow isomerizations in the denatured state much simpler or is it still complex?

To address the issue, we have studied the kinetic folding of proline-free variant (SNase (pro-)) of staphylococcal nuclease, a 149 residue, $\alpha + \beta$ protein,[6] which has no prosthetic groups nor cysteine residues, and hence is a typical model protein in protein-folding studies.[7–15] The folding reaction of this protein can be monitored by the intrinsic fluorescence of a sole tryptophan residue (Trp140) near the C terminus.

SNase (pro-) has been used for investigating the effect of prolyl *cis/trans* isomerization on the folding kinetics.[11–14] Systematic kinetic studies of SNase (pro-) were first reported by Walkenhorst *et al.*[11] and subsequently by Maki *et al.*[13,14] Although the folding kinetics of SNase (pro-) is simpler than that of wild-type SNase, SNase (pro-) still exhibits complex folding, in which at least three kinetic

Abbreviations used: DHFR, dihydrofolate reductase; SNase, staphylococcal nuclease; SNase (pro-), proline-free variant of staphylococcal nuclease; t_a, aging time which is the time between the first and second mixing in double-jump experiments; D, acid-denatured state; I, intermediate state; N, native state; GdnHCl, guanidinium chloride; EGTA, [ethylenebis](oxyethlene-nitrilo)tetraacetic acid; MALDI-TOF-MASS, matrix-assisted laser desorption/ionization time of flight mass spectrometry; m/z, mass-to-charge ratio.

E-mail address of the corresponding author: kuwajima@phys.s.u-tokyo.ac.jp

phases (fast, middle, and slow phases) of refolding from the fully denatured state were observed.[11,13,14] Walkenhorst *et al.* attributed the complex folding kinetics for SNase (pro-) to a structurally heterogeneous unfolded state that might be caused by *cis–trans* isomerization of a non-prolyl peptide bond.[11] However, their assignment was tentative and not tested by a stopped-flow double-jump assay that could monitor the slow isomerization in the unfolded state. Therefore, it still remains to be clarified why the proline-free protein exhibits multi-phasic kinetics of refolding. Is it because of slow *cis/trans* isomerizations of non-prolyl peptide bonds as observed in tendamistat,[5] or due to the presence of heterogeneous populations of on-pathway folding intermediates through which multiple parallel folding should take place?

To address the question, we have further investigated the folding of SNase (pro-) (P11A, P31A, P42A, P47T, P56A, P117G) in which all of the six proline residues are replaced with the other residues (Ala, Gly, or Thr). The results of stopped-flow double-jump experiments (interrupted refolding and interrupted unfolding) have shown that the fast phase is a lag phase in which the on-pathway folding intermediates are accumulated and that the subsequent multi-phasic kinetics arises from the presence of multiple parallel folding pathways. Amino acid sequence analysis and mass spectrometric analysis have shown chemical homogeneity of SNase (pro-) used, indicating that a single amino acid sequence can fold along multiple pathways. As far as we know, this is the first case where the multiple folding pathways of the proline-free variant have been revealed by the stopped-flow double-jump techniques.

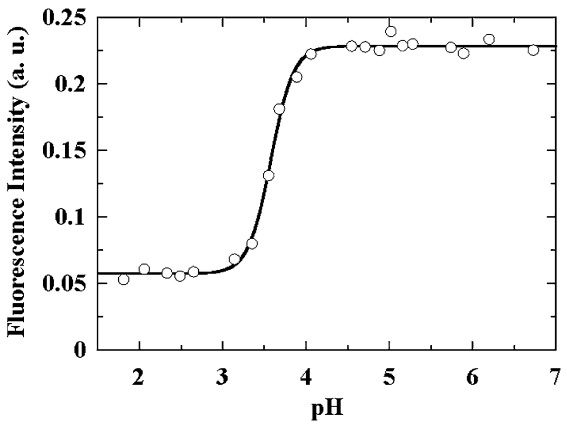


Figure 1. The acid-induced transition curve of SNase (pro-) at 20 °C. The transition curve was monitored by a change in intrinsic tryptophan fluorescence. The protein concentration was 0.01 mg/ml. The transition curve is analyzed by a two-state approximation in which only the native and the acid denatured states are assumed to be populated in the transition zone. The observed value of fluorescence, F, of SNase (pro-) at various values of pH are thus given by:

$$F = F_D \frac{K_0 \times 10^{-\Delta\nu pH}}{1 + K_0 \times 10^{-\Delta\nu pH}} + F_N \frac{1}{1 + K_0 \times 10^{-\Delta\nu pH}} \quad (7)$$

where F_N, F_D, K_0, and $\Delta\nu$ represent the fluorescence values in the native and the denatured states, the apparent equilibrium constant for the transition at pH = 0, and the number of bound protons which induces the denaturation transition, respectively. F_N and F_D are assumed to be independent of pH. The transition curve was analyzed using equation (7) by a non-linear least-squares method (KareidaGraph). As a result, the minimum number of bound protons $\Delta\nu$ is 3.3. The line represents the best-fit curve obtained by equation (7).

Results

Complex refolding and simple unfolding

Figure 1 shows the acid-induced denaturation curve of SNase (pro-) measured by the tryptophan fluorescence at 20 °C. The protein is in the native state above pH 4 and in the acid-denatured state below pH 3. Kinetic refolding and unfolding were induced by stopped-flow pH-jump techniques, and the reactions were monitored directly by the tryptophan fluorescence of SNase (pro-).

Typical kinetic refolding and unfolding curves of the protein are shown in Figure 2(a) and (b), respectively. The refolding kinetics (from pH 2.0 to pH 6.0) is well represented by a sum of four exponentials (equation (6)). The four phases are: (1) fast (the rate constant of 134(±21) s^{-1}, and the relative fluorescence amplitude of −11.3(±1.0)%), (2) middle (17.2(±0.5) s^{-1}, 57.3(±0.9)%), (3) slow (4.7(±0.1) s^{-1}, 35.3(±1.1)%), and (4) very slow phases (0.82(±0.04) s^{-1}, 4.5(±0.2)%). The major phases are the middle and slow phases, and the sum of their fluorescence amplitudes amount to about 93% of the total observable fluorescence change. The fast phase has a negative fluorescence amplitude of about −11%, indicating that this phase does not represent the direct folding into the native state. As will be shown later by a double-jump experiment, the fast phase is a "lag" phase caused by the accumulation of on-pathway folding intermediates. This has been suggested in previous studies,[11,13] and clearly established in this study by the double-jump assay (see below). Both the rate constant and amplitudes for the fast and the major (middle and slow) phases are coincident with the corresponding values previously reported for the kinetic refolding of the protein from urea-induced and guanidinium chloride (GdnHCl)-induced denatured states, when we take account of the dependence of the parameter values on the concentration of the denaturants.[8,11–13] Therefore, the refolding kinetics does not depend on the initial unfolding condition, but is determined solely by the final folding condition.

The unfolding kinetics (from pH 6.0 to pH 2.0) is well represented by a single exponential with an apparent rate constant of 7.69(±0.02) s^{-1} (equation (6)). Therefore, there is no kinetically observable unfolding intermediate. In both the refolding and unfolding, the sum of the observed amplitudes is

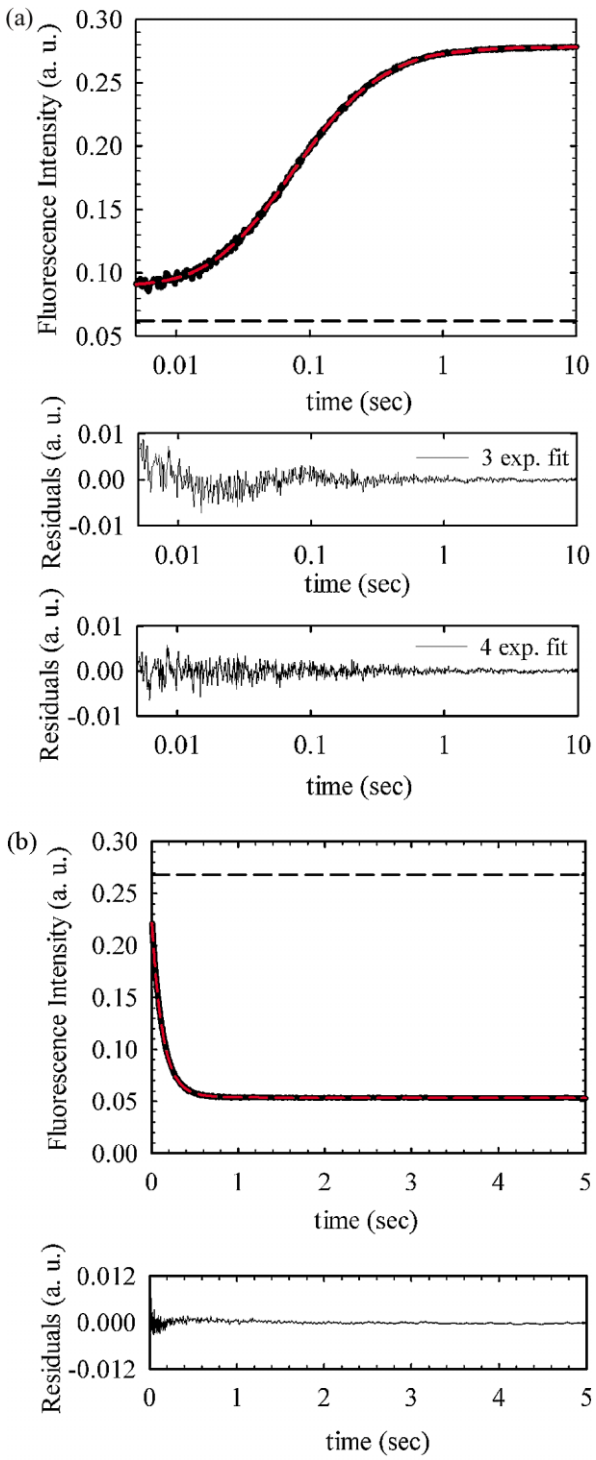


Figure 2. Direct refolding and unfolding kinetics of SNase (pro-). Both reactions were monitored by the change in intrinsic tryptophan fluorescence. The final protein concentration was 0.01 mg/ml. (a) Refolding was initiated by pH-jump from pH 2.0 to pH 6.0 and was monitored at 20 °C. The broken red line represents a quad-exponential fit of the data. The broken line represents the fluorescence intensity of the initial condition (pH 2.0). (b) Unfolding was initiated by pH-jump from pH 6.0 to pH 2.0 and was monitored at 20 °C. The broken red line represents a single-exponential fit of the data. The broken line represents the fluorescence intensity of the initial condition (pH 6.0).

less than the equilibrium difference between the native and denatured states, i.e. 86% and 79%, for refolding and unfolding, respectively. These might be interpreted in terms of the burst-phase intermediates for both the refolding and unfolding, but it is also likely that significant pH dependence of the fluorescence intensity of the native and/or denatured state(s) may exist and be interpreted into the missing amplitudes in the kinetic refolding and/or unfolding.

Accumulation of the native state during refolding

To correlate the fluorescence-detected refolding kinetics with the formation of the native state, we performed a double-jump stopped-flow experiment (interrupted refolding). Denatured SNase (pro-) at pH 2.0 was first mixed with a refolding buffer (pH 6.2), which caused the denatured protein to refold into the native state at pH 6.0, and the protein was aged under the refolding condition for a certain time (aging time t_a). After various aging times, the second mixing with an unfolding solution (pH 1.1) prevented the protein from further folding and unfolded the already folded protein; the final pH was at pH 1.6.

All the unfolding reactions induced by the second mixing with different aging times were well represented by a single exponential with the rate constant (29.6–32.4 s^{-1}) that is equivalent to that of the direct unfolding at pH 1.6 in a single mixing. Therefore, only the unfolding from the native state was observed, and the unfolding from transient intermediates was not observed. It is likely that the unfolding from the intermediates is too rapid to measure and occurs within the dead-time (4 ms) of the stopped-flow apparatus. It is also likely that the fluorescence intensity of the transient intermediates is close to that of denatured state ensemble, so that the fluorescence change for unfolding from the intermediates, if any, is insufficient to be detected.

Figure 3 thus shows the accumulation of the native state as a function of the aging time during the refolding induced by the first mixing. The kinetics of the native-state accumulation is well represented by a sum of two exponentials with rate constants (and amplitudes) of 12.9(±2.4) s^{-1} (65.5(±8.2)%) and 2.5(±0.6) s^{-1} (34.4(±9.4)%), which are consistent with those of the major (middle and slow) phases observed in the direct refolding reaction. This indicates that the fluorescence changes of both the middle and slow phases represent the accumulation of the native state ($I_F \rightarrow N$, $I_M \rightarrow N$). Thus, a sequential reaction scheme, $D \overset{Fast}{\rightarrow} I_1 \overset{Middle}{\rightarrow} I_2 \overset{Slow}{\rightarrow} N$, does not produce these results, because the middle phase is limited by the slow phase with an accumulation of I_2, and does not accumulate the native state in this scheme.

The percentage of the accumulation of the native

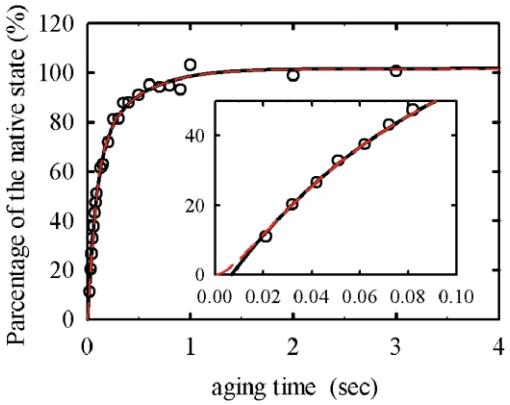
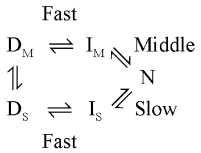


Figure 3. The time-course for the accumulation of the native state measured by the interrupted refolding experiments at 20 °C. Open circles (○) represent the accumulation of the native state (pH 6.0). The formation of the native state is better described by the sum of two exponentials, and the continuous line represents a double-exponential fit of the data. The inset shows the early time region. The percentage of the native state cannot be extrapolated to zero at aging time 0, indicating the presence of a lag phase. The kinetics of the native-state formation is also well described by the sum of three exponentials, in which the rate constant for the fast (lag) phase is assumed to be identical with that observed in the direct-refolding kinetics (broken red line).

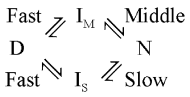
state is not extrapolated to zero at aging time 0 (Figure 3). This indicates the presence of a lag phase, so that the denatured state does not fold directly into the native state. The lag phase seems to be equivalent to the fast phase observed in the direct refolding reaction, because the kinetics of the native-state formation is well represented by a sum of three exponentials with the rate constants of (and amplitudes) 134 s^{-1} ($-6.8(\pm5.1)$%), 13.0($\pm$2.6) s^{-1} (69.9($\pm$8.7)%) and 2.5($\pm$0.6) s^{-1} (37.0($\pm$10.0)%) (134 s^{-1}, which is the rate constant of the fast phase obtained in the direct refolding reaction, was fixed during the fitting), as well as a sum of two exponentials described above.

Three possible folding schemes which explain both the results of the direct refolding and the interrupted refolding experiments of SNase (pro-) are Schemes 1–3 as shown below. In Scheme 1, there are at least two heterogeneous populations, D_M and D_S, in the acid denatured-state ensemble, and I_M and I_S are the on-pathway intermediates formed from D_M and D_S, respectively, in the fast phase, and the native state is formed through I_M and I_S along the middle and slow pathways. In this scheme, the interconversions between D_M and D_S and between I_M and I_S are much slower than the refolding reactions, indicating large free-energy barriers between D_M and D_S and between I_M and I_S. In Schemes 2 and 3, however, there are no energetically separated heterogeneous populations in the denatured-state ensemble. Instead, there is kinetic partitioning between the middle and slow pathways during the fast phase. Thus, the free-

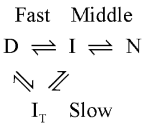


Scheme 1.

energy barrier between I_M and I_S should be produced during the fast phase. The difference between Schemes 2 and 3 is whether I_S directly folds into N (Scheme 2) or folds through I_M into N (Scheme 3). Assignment of the three (fast, middle, and slow) phases to individual steps in Schemes 1–3 is only applicable in strongly native conditions, where these steps are kinetically well separated.



Scheme 2.



Scheme 3.

Fast Interconversion within the denatured-state ensemble

To distinguish between the different schemes shown above, we performed another stopped-flow double-jump experiment (interrupted unfolding). Native SNase (pro-) (pH 5.4) was first mixed with an unfolding solution (pH 1.1), which caused the native protein to unfold at pH 1.8, and the protein was aged under the unfolding condition for an aging time t_a. After various aging times, the second mixing with a refolding buffer (pH 7.5) prevented the protein from further unfolding and refolded the already unfolded protein into the native state; the final pH was at pH 5.9.

Figure 4 shows the rate constants and relative amplitudes of each phase for the interrupted-unfolding reactions. Even when the aging time t_a is 26 ms, four phases (the fast, middle, slow, and very slow phases) are already seen. Both the rate constant and the relative amplitude of each phase for the interrupted-unfolding reaction are constant within a range of experimental error and independent of aging time. Therefore, even if there are different populations in the unfolded-state ensemble, the conformational transition between the different populations must occur within 26 ms

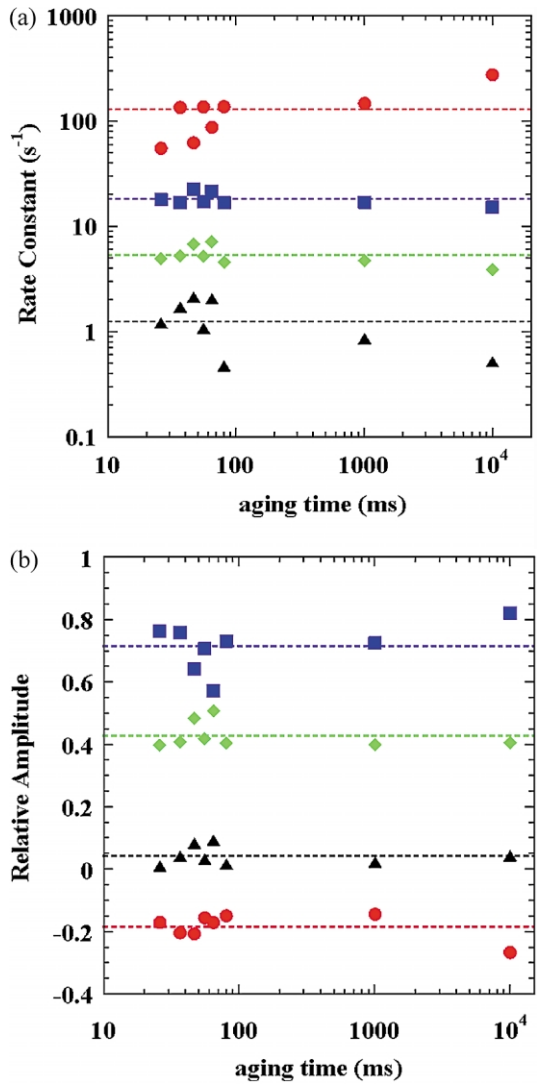


Figure 4. Interrupted-unfolding reactions for various aging times (20 °C, final pH 5.9). These reactions are represented by the sum of four exponentials. Filled circles (●), filled squares (■), filled diamonds (◆) and filled triangles (▲) indicate fast, middle, slow and very slow phases, respectively. (a) Apparent rate constants of each phase as a function of aging time. (b) Relative amplitudes of each phase as a function of aging time.

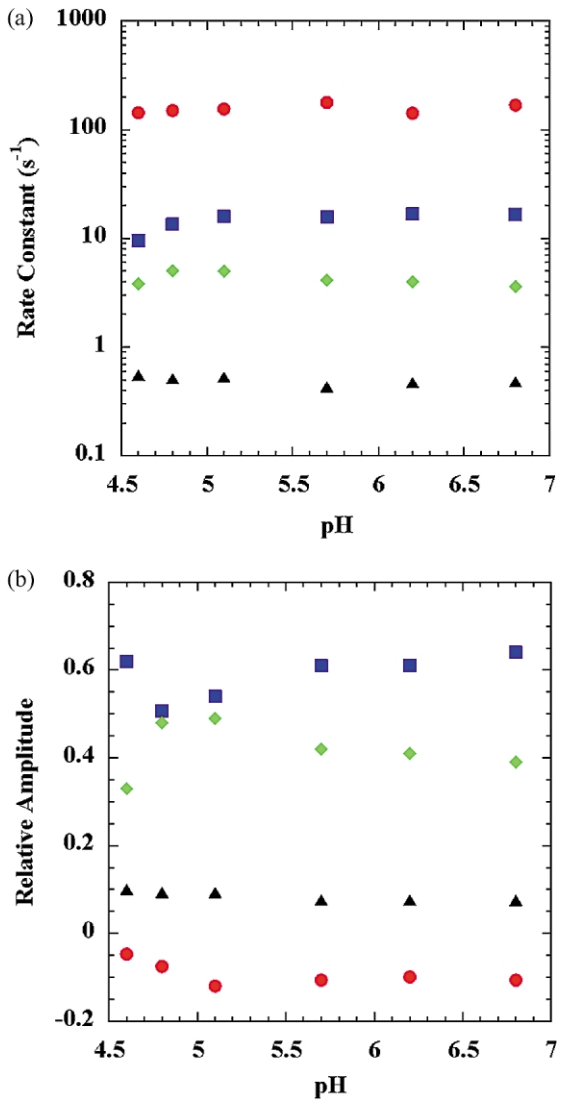


Figure 5. pH-dependence of direct refolding reactions (20 °C). These reactions are represented by the sum of four exponentials. Filled circles (●), filled squares (■), filled diamonds (◆) and filled triangles (▲) indicate fast, middle, slow and very slow phases, respectively. (a) Apparent rate constants of each phase as a function of final pH. (b) Relative amplitudes of each phase as a function of final pH.

and, hence be much faster than the subsequent folding processes (the middle and slow phases).

The above results thus clearly show that the slow conversion between the different populations of the unfolded state does not explain the present data. Therefore, Scheme 1 is excluded. Although the non-prolyl peptide *cis*/*trans* isomerization in the denatured state has been shown to produce complex refolding kinetics in tendamistat[5] and was once suggested in SNase (pro-)[11,15], the present results exclude this possibility, at least as an interpretation of the major phases of refolding. In fact, the non-prolyl *cis*/*trans* isomerizations about peptide bonds in proteins and model peptides occur with a time constant of 400 ms to 3.3 s,[5,16,17] and hence they are much slower than the conformational transitions, which must occur within 26 ms, within the denatured-state ensemble of SNase (pro-).

pH-dependence of relative fractions of SNase (pro-) molecules which fold through multiple pathways

Figure 5 shows the pH dependence of the apparent rate constant and relative amplitude of each phase for direct folding reactions of SNase (pro-) from pH 1.6 to pH 4.6–6.8. Under these conditions (pH 4.6–6.8), the protein is fully in the native state (see Figure 1). However, the relative amplitudes of the major two (middle and slow) phases depend

on the final pH below pH 5.5. This dependence indicates that the relative fractions of the middle and slow-folding species are not determined by the initial unfolding condition, but determined by the final folding condition, because the relative fluorescence amplitudes of the major two (middle and slow) phases correspond to the fraction of the native-state formation (interrupted-refolding experiments). The result thus again suggests that Scheme 1 is less likely than Schemes 2 and 3.

Chemical homogeneity of SNase (pro-)

Because all the above arguments are based on the assumption that SNase (pro-) used in the present study is chemically homogeneous, this assumption was tested by N-terminal amino acid sequence analysis and by mass spectrometric analysis. A potential reason why the protein might be chemically heterogeneous is found in incomplete processing of the N-terminal methionine residue for a protein over-expressed in *Escherichia coli* (*E. coli*), and this has been reported for other proteins.[18] The N-terminal methionine residue could change the energy landscape of protein folding/unfolding.[19,20] Thus, the N-terminal amino acid sequence of the first ten residues of SNase (pro-) was analyzed, and our purified recombinant protein has been found not to contain any extra N-terminal methionine residue (data not shown).

A mass spectrometric profile of our SNase (pro-) sample is shown in Figure 6. There is only a single major peak, which indicates a mass of 16,668.9 ± 1.9 that is identical with the mass (16,670.8) of SNase (pro-). In conclusion, our purified SNase (pro-) is chemically homogeneous.

Discussion

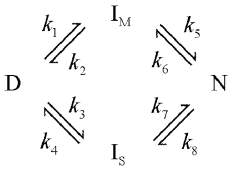
Multiple parallel-pathway folding of SNase (pro-)

The present results clearly demonstrate that SNase (pro-) folds into the native state along at least two accessible parallel pathways, and the presence of the aging time independence of a very slow phase of refolding in the interrupted unfolding experiments further suggests that there may be at least three parallel folding pathways. Because SNase (pro-) is chemically homogeneous in terms of the amino acid sequence analysis and the mass spectrometric analysis, the single homogeneous amino acid sequence can fold along the multiple parallel pathways. Because the slow-folding species that fold along the slow and very slow pathways appear rapidly, within 26 ms in the interrupted unfolding experiments (Figure 4), the kinetic partitioning is the only mechanism that explains the present data, and the simplest models are given by Schemes 2 and 3. The kinetic partitioning models are further supported by the dependence of the relative amplitudes of the major two phases of refolding (Figure 5). It is difficult for such dependence to be explained by Scheme 1, in which the amplitudes are determined by initial populations of different denatured species (D_M and D_S). Similar dependence of the relative amplitudes on the final folding condition has also been observed in the refolding study of SNase (pro-) from the urea-induced unfolded state, in which the amplitude of the middle phase increases while that of the slow phase decreases as urea concentration increases under native folding conditions.[14]

Between the two schemes, however, Scheme 2 is more likely than Scheme 3. Previous studies on the refolding kinetics of SNase have shown that the dependence of the rate constant for denaturant concentration (urea and GdnHCl) is very similar between the middle and slow phases, indicating that the conformational changes in terms of accessibility to the denaturant during the middle and slow processes are also similar to each other.[11,14] Therefore, Scheme 3, in which the slow phase is not parallel with the middle phase, is less likely than Scheme 2.

Numerical simulation analysis

Although the qualitative kinetic arguments presented above can distinguish among alternative mechanisms depicted by Schemes 1–3, further validation of the kinetic mechanism of this complexity may require quantitative simulation analysis. Therefore, we have solved the kinetic differential equations for the reaction scheme (Scheme 4):



Scheme 4.

where k_i represents an ith microscopic rate constant. The differential equations are given by:

$$\begin{aligned}
\frac{df_D}{dt} &= -(k_1 + k_3)f_D + k_2 f_{I_M} + k_4 f_{I_S} \\
\frac{df_{I_M}}{dt} &= k_1 f_D - (k_2 + k_5)f_{I_M} + k_6 f_N \\
\frac{df_{I_S}}{dt} &= k_3 f_D - (k_4 + k_7)f_{I_S} + k_8 f_N \\
\frac{df_N}{dt} &= k_5 f_{I_M} + k_7 f_{I_S} - (k_6 + k_8)f_N
\end{aligned} \tag{1}$$

where f_D, f_{I_M}, f_{I_S}, and f_N are the fractions of D, I_M, I_S,

and N, respectively. The observed rate constants, λ_1, λ_2, and λ_3, for the three (fast, middle, and slow) phases, respectively, are given by the solutions of the cubic equation:

$$\lambda^3 - a\lambda^2 + b\lambda - c = 0 \quad (2)$$

where:

$$a = \sum_{i=1}^{8} k_i,$$

$$\begin{aligned} b = {} & k_1k_4 + k_1k_5 + k_1k_6 + k_1k_7 + k_1k_8 + k_2k_3 + k_2k_4 \\ & + k_2k_6 + k_2k_7 + k_2k_8 + k_3k_5 + k_3k_6 + k_3k_7 \\ & + k_3k_8 + k_4k_5 + k_4k_6 + k_4k_8 + k_5k_7 + k_5k_8 \\ & + k_6k_7 \end{aligned} \quad (3)$$

$$\begin{aligned} c = {} & k_1(k_4k_5 + k_4k_6 + k_4k_8 + k_5k_7 + k_5k_8 + k_6k_7) \\ & + k_2(k_3k_6 + k_3k_7 + k_3k_8 + k_4k_6 + k_4k_8 + k_6k_7) \\ & + k_3(k_5k_7 + k_5k_8 + k_6k_7) + k_4k_5k_8 \end{aligned}$$

In addition, the following relation should be satisfied because of requirement of thermodynamic equilibrium:

$$k_1k_4k_5k_8 = k_2k_3k_6k_7 \quad (4)$$

The amplitudes, A_1, A_2, and A_3, for the three phases of the observed kinetic equation:

$$A(t) = A(\infty) + A_1\, e^{-\lambda_1 t} + A_2\, e^{-\lambda_2 t} + A_3\, e^{-\lambda_3 t} \quad (5)$$

are obtained from the initial conditions ($f_D = 1$, $f_{I_M} = 0$, $f_{I_S} = 0$, and $f_N = 0$ for folding), and the fluorescence values of the individual states. We assume that the fluorescence values of the intermediates (I_M and I_S) are equivalent to that of the denatured state (D), and hence the fluorescence change is only observed when the native state (N) is formed during the refolding.

The kinetic refolding curve shown in Figure 2(a) is well reproduced by a numerical simulation analysis with equations (2)–(5) using the parameter values, $k_1 = 82.3\ s^{-1}$, $k_3 = 56.1\ s^{-1}$, $k_5 = 17.2\ s^{-1}$, and $k_7 = 4.7\ s^{-1}$, and by assuming that the rate constants for the reversed processes, k_2, k_4, k_6, and k_8, are negligibly small; the very slow phase is not discussed here because of the small relative amplitude.

We have also investigated, by the numerical simulation analysis, whether or not Scheme 4 is consistent with the previously observed multiphasic kinetic behavior of SNase (pro-) in the presence of the denaturant, GdnHCl.[11] This is particularly important, because the GdnHCl dependence of the apparent rate constants (λ_2 and λ_3) and the corresponding amplitudes of the middle and slow phases has been interpreted in terms of Scheme 1,[11] in which slow *cis*/*trans* isomerization of a non-prolyl peptide bond is the mechanism that produces the multi-phasic kinetics.

Our simulation results have shown that it is possible to fit Scheme 4 to the previously observed kinetics in the presence of GdnHCl. We assume that every microscopic rate constant has the GdnHCl dependence given by $k_i = k^\circ_i \exp(m_i c)$, where k°_i is the k_i at 0 M GdnHCl, m_i is a constant that represents the GdnHCl dependence, and c is the molar concentration of GdnHCl. We have found that the following set of parameters gives a reasonable fit with the reported rate constants and amplitudes at different GdnHCl concentrations below 2 M; an additional process is known to contribute to the kinetics above 2 M GdnHCl and should not be considered here. These parameters are: $k_1 = 100\ s^{-1}$, $k_2 = 0.0251 \exp(6.45c)\ s^{-1}$, $k_3 = 15.8\ s^{-1}$, $k_4 = 0.0117 \exp(6.45c)\ s^{-1}$, $k_5 = 10.0 \exp(-3.91c)\ s^{-1}$, $k_6 = 0.0251 \exp(2.30c)\ s^{-1}$, $k_7 = 2.34 \exp(-3.68c)\ s^{-1}$, and $k_8 = 0.00200 \exp(2.53c)\ s^{-1}$. Interestingly, the above parameter set has reproduced a dent of the GdnHCl dependence of logarithmic λ_3 around 0.4 M GdnHCl, which has been detected experimentally,[11] but has not been reproduced in the previous analysis based on Scheme 1. However, the present simulation analysis did not reproduce roll-over behavior of the logarithmic λ_2 and λ_3 observed at very low GdnHCl concentrations. Therefore, if this roll-over is real, it might be required to assume additional intermediates between I_M (and I_S) and N.

Comparison to multiple parallel-pathway folding of other proteins

Here, we further discuss the multiple folding pathways of other proteins, but will not discuss multiple kinetics caused by some additional chemical reactions, such as heme misligations in heme proteins[21] or disulfide bond formation and rearrangements in disulfide-containing proteins,[22] because these reactions, though resulting in additional complexity in the folding kinetics, are clearly not related to SNase that does not have heme or disulfide. Nevertheless, it should be mentioned that the recent study on the cytochrome *c* folding by Goldbeck *et al.*[21] has indicated that the kinetic heterogeneity is also related to diffusional nature of early folding dynamics of this heme protein.

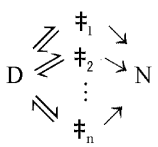
Multiple parallel folding pathways, except for those produced by slow prolyl isomerizations in the denatured state, have also been observed in the folding of other proteins, DHFR and hen egg lysozyme, but these have been considered to be rather exceptional. DHFR from *E. coli* folds through four parallel folding pathways containing on-pathway intermediates, but in this case, there are four different native states that may contribute to the multiple folding.[3] DHFR from human and from *Lactobacillus casei* also folds through similar multiple folding pathways.[4] Hen egg lysozyme, which consists of α and β-domains, folds through

two competing pathways under a specific condition (pH 5.2).[1,2] The fast pathway represents the simultaneous formation of the α and β-domains, while the slow pathway represents the structural formation of the α-domain in advance of formation of the β-domain. Furthermore, a third parallel pathway with another folding intermediate appears in addition to the fast and slow pathways under other conditions (pH 9.2 or ≥100 mM NaCl at pH 5.2).[23,24] The observation of the multiple parallel-pathway folding in SNase, which is a typical model protein in protein folding studies, suggests that the parallel-pathway folding may be more general than previously thought.

In a proline-free variant of tendamistat, non-prolyl *cis/trans* peptide-bond isomerizations in the denatured state are known to cause multi-phasic kinetics of refolding.[5] Because such peptide-bond isomerizations must be universal in denatured proteins, there is no reason why this is absent in SNase (pro-). However, it is very likely that the non-prolyl *cis/trans* isomerizations, if any, could not be detected experimentally in SNase (pro-). The expected amplitude of the non-prolyl-isomerization phase is only a few percent, and its rate constant (0.3–3 s^{-1}) is close to that of the slow phase (4.7 s^{-1}, 38%) in the SNase (pro-) folding. These facts would make the detection of the non-prolyl peptide isomerizations in SNase (pro-) extremely difficult, indicating that the present data should not be inconsistent with those previously reported for the non-prolyl variant of tendamistat.

The multiple parallel-pathway folding as a general mechanism of protein folding

Finally, it should be emphasized that the fast accumulation of the stable on-pathway intermediates (Scheme 4), separated by a large free-energy barrier from one another, is critical for experimental verification of the multiple parallel-pathway folding. The accumulation of the intermediates has resulted in the multi-exponential kinetics of refolding, and hence the presence of the multiple parallel pathways has been revealed experimentally by the use of the stopped-flow, double-jump techniques. When there are only two macroscopically stable species, namely, the native and denatured states, the refolding kinetics must be single-exponential even if the real folding takes place along the multiple parallel pathways as illustrated by Scheme 5:



Scheme 5.

where ‡$_i$ (i = 1,2,…,n) is an ith unstable intermediate (e.g. the transition state) that may exist on each of the multiple folding pathways. Therefore, it should be noticed that the single-exponential kinetics offers no guarantee that the folding occurs along a single pathway.[25] This, in turn, again suggests that the multiple parallel-pathway folding may be more general than we have expected. Further characterizations of the multiple parallel-pathway folding may be important for us to fully elucidate the molecular mechanisms of *in vitro* folding of globular proteins.

Materials and Methods

Chemicals

All chemicals used in this study were either specially prepared or guaranteed reagent-grade chemicals.

Expression and purification of SNase (pro-)

The expression and purification of SNase (pro-) were carried out as described.[26] The samples used for kinetic measurements were >93% pure, where the purity was estimated using the peak area of the profile obtained from a reversed-phase HPLC chromatogram detected by UV-absorbance at 215 nm using a TSKgel Octadecyl-4PW column (TOSOH) with a linear gradient elution of 30%–50% (v/v) acetonitrile in the presence of 0.1% (v/v) trifluoroacetic acid at a flow rate of 0.5 ml per minute. The refolding kinetics for the >99% purified sample, which was further purified by the reversed-phase HPLC, was identical with that for >93% purified sample (data not shown).

Sample preparation

The sample preparation for the equilibrium measurements was carried out as described.[26] SNase (pro-) solutions were adjusted to various pH (50 mM sodium acetate plus 50 mM NaCl (below pH 5.5), or 50 mM sodium cacodylate plus 50 mM NaCl and 1.0 mM [ethylenebis](oxyethlenenitrilo)tetraacetic acid (EGTA) (above pH 5.5)), and the final protein concentration was adjusted to 0.01 mg/ml. The protein concentration of SNase (pro-) was determined from the UV absorbance with an extinction coefficient $E^{1\%}_{1\,cm}$ of 9.3 at 280 nm.

For direct unfolding and interrupted-unfolding experiments, the preparation method was essentially the same as that for the equilibrium preparation. The final protein concentration was adjusted to 0.02 mg/ml (direct unfolding measurement) or 0.04 mg/ml (interrupted unfolding measurements) and the buffer (pH 6.0 (direct unfolding measurement) or pH 5.4 (interrupted unfolding measurements), 100 mM sodium cacodylate) was used.

For direct refolding and interrupted-refolding experiments, lyophilized SNase (pro-) was dissolved into pH 2.0 solution (100 mM HCl/NaCl) and the solution was filtrated through a Millipore membrane filter (Millex-HV) with a pore size of 0.45 μm. The protein concentration of SNase (pro-) was determined as described above and was adjusted to 0.02 mg/ml (direct refolding measurements) or 0.04 mg/ml (double-jump measurements). For direct refolding experiments (from pH 1.6 to

various final pH values), the preparation method was essentially the same as described above.

Equilibrium and kinetic measurements

The pH-induced unfolding transition was studied by measuring the intrinsic tryptophan fluorescence of 21 equilibrium samples (pH 1.8–6.8). The equilibrium transition curve was measured in an Applied PhotoPhysics model SX.18MV stopped-flow fluorescence apparatus (20(±0.1) °C). The excitation wavelength was 295 nm. The emission of the fluorescence was detected using a 320 nm cut off filter to detect only the fluorescence of tryptophan. The optical path length was 2 mm.

All kinetic experiments were performed using the SX.18MV stopped-flow fluorescence apparatus (20(±0.1) °C). These reactions were also monitored by the change in intrinsic tryptophan fluorescence above 320 nm with excitation at 295 nm. The dead-time of the apparatus was approximately 4 ms using the chemical reactions at a mixing ratio of 1:1 or 1:10.6.[27,28]

Data fitting

Kinetic data were fitted by the non-linear least-squares method with an equation:

$$F(t) = F(\infty) + \sum_{i=1}^{n} \Delta F_i \exp(-\lambda_i t) \quad (6)$$

where $F(t)$ and $F(\infty)$ are the observed fluorescence at time t and infinite time, respectively. ΔF_i and λ_i are the fluorescence amplitude and the rate constant, respectively, of the ith phase. These analyses were performed using KaleidaGraph (Synergy Software) and SigmaPlot (SPSS Science).

Direct refolding and unfolding experiments

All direct refolding and unfolding experiments were performed using the SX.18MV stopped-flow fluorescence apparatus (20(±0.1) °C). In the case of direct refolding experiments, acid-induced denatured SNase (pro-) (pH 2.0, 100 mM HCl/NaCl, 0.02 mg/ml) was mixed with the refolding buffer (pH 6.2, 100 mM sodium cacodylate, 2 mM EGTA) at a ratio of 1:1, which initiated the refolding reaction (final pH 6.0, 0.01 mg/ml). For direct refolding experiments from pH 1.6 to various final pH values, acid-denatured SNase (pro-) (0.09 mg/ml) in 50 mM sodium acetate plus 50 mM NaCl at pH 1.6 was mixed with various refolding buffers (50 mM sodium cacodylate, 50 mM NaCl plus 1 mM EGTA at different pH values from 4.8 to 7.5) at a mixing ratio of 1:10.6 (protein solution: refolding buffer), which gave the final pH values from 4.6 to 6.8 and a final protein concentration of 0.007 mg/ml.

In direct-unfolding experiments, native SNase (pro-) (pH 6.0, 100 mM sodium cacodylate) was mixed with the unfolding solution (pH 1.2, 100 mM HCl/NaCl) at a ratio of 1:1, which initiated the unfolding reaction (final pH 2.0, 0.01 mg/ml). These refolding and unfolding reactions were fitted by equation (6).

Double-jump experiments (interrupted refolding and interrupted unfolding)

Double-jump experiments were performed using the SX.18MV stopped-flow fluorescence apparatus (20(±0.1) °C). For interrupted-refolding experiments, acid-denatured SNase (pro-) (pH 2.0, 100 mM HCl/NaCl, 0.04 mg/ml) was mixed with the refolding buffer (pH 6.2, 100 mM sodium cacodylate, 2 mM EGTA) at a ratio of 1:1, which initiated the refolding reaction (pH 6.0). After an aging time t_a, the second mixing with the unfolding solution (pH 1.1, 100 mM HCl/NaCl) at a ratio of 1:1 prevented the protein from further folding and unfolded the already folded protein at various aging times (final pH 1.6, 0.01 mg/ml). A series of the interrupted-refolding reactions with different aging times was thus measured as the unfolding reactions from the folded state. These unfolding reactions were fitted by equation (6). The relative difference between $F(0)$ and $F(\infty)$ was assumed to correspond to the accumulation of the native state, where $F(0)$ is the fluorescence value of the fitting curve extrapolated to the unfolding-reaction time 0 and $F(\infty)$ is the fluorescence value at equilibrium.

In interrupted-unfolding experiments, native SNase (pro-) (pH 5.4, 100 mM sodium cacodylate) was mixed with the unfolding solution (pH 1.1, 100 mM HCl/NaCl) at a ratio of 1:1, which made the native protein unfold (pH 1.8). After an aging time t_a, the second mixing with the refolding buffer (pH 7.5, 100 mM sodium cacodylate, 2 mM EGTA) at a ratio of 1:1 prevented the protein from further unfolding and folded the already unfolded protein into the native state (final pH 5.9). A series of the interrupted-unfolding reactions with different aging times was thus measured as the refolding reactions from the already unfolded state. These refolding reactions were fitted by equation (6).

Mass spectrometry and N-terminal sequence analysis

Mass spectra of our purified protein were measured by the MALDI-TOF-MASS with an Applied Biosystems model Voyager DE STR mass spectrometer (Figure 6). Sinapinic acid was used as the matrix. Insulin, myoglobin and myoglobin dimer were used as standard proteins to calibrate the mass axis (m/z). N-terminal sequencing of our purified protein was carried out

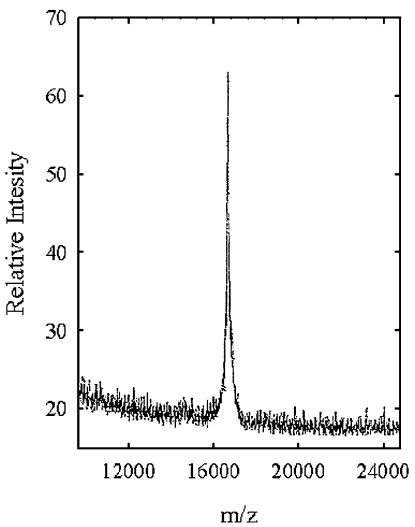


Figure 6. The mass spectrum of our purified recombinant SNase (pro-) measured by MALDI TOF-MASS.

using an Applied Biosystems model PROCISE-cLC automated sequencer. In this study we analyzed the first ten residues in the protein.

Numerical simulation analysis

Numerical simulation analysis for equations (2)–(5) was performed using Microsoft EXCEL 2000 and a computer program coded by ourselves (Visual Basic version 6.0).

Acknowledgements

The authors express their appreciation to Dr M. Arai for his assistance in taking dead-time measurements of the stopped-flow apparatus, and to Mr R. Yoshida and Mr A. Kadooka for invaluable advice and discussions.

References

1. Kiefhaber, T. (1995). Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. USA*, **92**, 9029–9033.
2. Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **270**, 294–304.
3. Jennings, P. A., Finn, B. E., Jones, B. E. & Matthews, C. R. (1993). A reexamination of the folding mechanism of dihydrofolate reductase from *Escherichia coli*: verification and refinement of a four-channel model. *Biochemistry*, **32**, 3783–3789.
4. Wallace, L. A. & Matthews, C. R. (2002). Highly divergent dihydrofolate reductases conserve complex folding mechanisms. *J. Mol. Biol.* **315**, 193–211.
5. Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G. & Kiefhaber, T. (2001). Nonprolyl *cis* peptide bonds in unfolded proteins cause complex folding kinetics. *Nature Struct. Biol.* **8**, 452–458.
6. Hynes, T. R. & Fox, R. O. (1991). The crystal structure of staphylococcal nuclease refined at 1.7 Å resolution. *Proteins: Struct. Funct. Genet.* **10**, 92–105.
7. Schechter, A. N., Chen, R. F. & Anfinsen, C. B. (1970). Kinetics of folding of staphylococcal nuclease. *Science*, **167**, 886–887.
8. Sugawara, T., Kuwajima, K. & Sugai, S. (1991). Folding of staphylococcal nuclease A studied by equilibrium and kinetic circular dichroism spectra. *Biochemistry*, **30**, 2698–2706.
9. Kuwajima, K., Okayama, N., Yamamoto, K., Ishihara, T. & Sugai, S. (1991). The Pro117 to glycine mutation of staphylococcal nuclease simplifies the unfolding–folding kinetics. *FEBS Letters*, **290**, 135–138.
10. Jacobs, M. D. & Fox, R. O. (1994). Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, **91**, 449–453.
11. Walkenhorst, W. F., Green, S. M. & Roder, H. (1997). Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry*, **36**, 5795–5805.
12. Ikura, T., Tsurupa, G. P. & Kuwajima, K. (1997). Kinetic folding and cis/trans prolyl isomerization of staphylococcal nuclease. A study by stopped-flow absorption, stopped-flow circular dichroism and molecular dynamics simulations. *Biochemistry*, **36**, 6529–6538.
13. Maki, K., Ikura, T., Hayano, T., Takahashi, N. & Kuwajima, K. (1999). Effects of proline mutations on the folding of staphylococcal nuclease. *Biochemistry*, **38**, 2213–2223.
14. Maki, K., Ikura, T., Mohs, A. & Kuwajima, K. (1999). Equilibrium and kinetics of folding of staphylococcal nuclease and its proline mutants. In *Old and New Views of Protein Folding* (Kuwajima, K. & Arai, M., eds), pp. 271–278, Elsevier, Amsterdam.
15. Walkenhorst, W. F., Edwards, J. A., Markley, J. L. & Roder, H. (2002). Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. *Protein Sci.* **11**, 82–91.
16. Scherer, G., Krammer, M. L., Schultkkowski, M., Reimer, U. & Fischer, G. (1998). Barriers to rotation of secondary amide peptide bonds. *J. Am. Chem. Soc.* **120**, 5568–5574.
17. Odefey, C., Mayr, L. M. & Schmid, F. X. (1995). Nonprolyl *cis–trans* peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *J. Mol. Biol.* **245**, 69–78.
18. Ben-Bassat, A., Bauer, K., Chang, S. Y., Myambo, K., Boosman, A. & Chang, S. (1987). Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J. Bacteriol.* **169**, 751–757.
19. Ishikawa, N., Chiba, T., Chen, L. T., Shimizu, A., Ikeguchi, M. & Sugai, S. (1998). Remarkable destabilization of recombinant alpha-lactalbumin by an extraneous N-terminal methionyl residue. *Protein Eng.* **11**, 333–335.
20. Chaudhuri, T. K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T. P. *et al.* (1999). Effect of the extra N-terminal methionine residue on the stability and folding of recombinant alpha-lactalbumin expressed in *Escherichia coli*. *J. Mol. Biol.* **285**, 1179–1194.
21. Goldbeck, R. A., Thomas, Y. G., Chen, E., Esquerra, R. M. & Kliger, D. S. (1999). Multiple pathways on a protein-folding energy landscape: kinetic evidence. *Proc. Natl. Acad. Sci. USA*, **96**, 2782–2787.
22. van den Berg, B., Chung, E. W., Robinson, C. V., Mateo, P. L. & Dobson, C. M. (1999). The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. *EMBO J.* **18**, 4794–4803.
23. Bieri, O. & Kiefhaber, T. (2001). Origin of apparent fast and non-exponential kinetics of lysozyme folding measured in pulsed hydrogen exchange experiments. *J. Mol. Biol.* **310**, 919–935.
24. Bieri, O., Wildegger, G., Bachmann, A., Wagner, C. & Kiefhaber, T. (1999). A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding. *Biochemistry*, **38**, 12460–12470.
25. Fersht, A. R., Itzhaki, L. S., Elmasry, N. F., Matthews, J. M. & Otzen, D. E. (1994). Single *versus* parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl. Acad. Sci. USA*, **91**, 10426–10429.
26. Kamagata, K. & Kuwajima, K. (2003). Parallel folding pathway of proline-free staphylococcal nuclease studied by the stopped-flow double-jump method. *Spectrosc. Intl J.* **17**, 203–212.
27. Paul, C., Kirschner, K. & Haenisch, G. (1980).

Calibration of stopped-flow spectrophotometers using a two-step disulfide exchange reaction. *Anal. Biochem.* **101**, 442–448.

28. Brissette, P., Ballou, D. P. & Massey, V. (1989). Determination of the dead time of a stopped-flow fluorometer. *Anal. Biochem.* **181**, 234–238.

Edited by F. Schmid

(Received 21 March 2003; received in revised form 29 July 2003; accepted 29 July 2003)