Compaction during Protein Folding Studied by Real-Time NMR Diffusion Experiments

Jochen Balbach*,†

Laboratorium für Biochemie, Universität Bayreuth D-95440 Bayreuth, Germany

Received December 30, 1999
Revised Manuscript Received April 19, 2000

During a protein folding reaction, the unstructured polypeptide chains find the highly ordered native state within a very short time, typically in the milliseconds to minutes range. Folding is always accompanied by a compaction of the protein chain. Recently, there have been several reports on measuring chain compaction directly by time-resolved experiments using small-angle X-ray scattering[1−4] (SAXS) and dynamic light scattering[5] (DLS).

Here I describe NMR diffusion experiments to determine the hydrodynamic radii (R_h) of the unfolded, partially folded and native states of bovine α-lactalbumin (BLA) at equilibrium and novel time-resolved NMR diffusion experiments to study R_h of a transiently populated collapsed state of the molten globule type. This state is as compact as the pH 2 state at equilibrium, which shows that the main chain compaction occurs before the rate-limiting step of folding.

It is still unclear whether there are universal rules for the succession of events in the complex process of protein folding, but every folding reaction involves significant compaction when the extended unfolded chain collapses to the compact native state.[6] In fact, for several proteins collapsed forms could be found both transiently and at equilibrium, which differed from the native and unfolded forms. Common features of these molten globule states (MG) include substantial secondary structure, the absence of nativelike tertiary structure, a globular compactness, and noncooperative unfolding transitions.[7,8]

Equilibrium molten globule states are often populated at low pH, and their radii of gyration (R_g) and Stokes radii are 10−30% larger than those of the corresponding native states.[9−13] In contrast, unfolded proteins in concentrated urea or guanidine hydrochloride (GdnHCl) solutions are substantially expanded, and their radii are increased between 60 and 300%, depending on the molecular weight of the protein.[1,12,14] Recently translational diffusion, determined with NMR techniques, has been used in combination with pulsed field gradients, to determine the size of folded and unfolded proteins at equilibrium.[14−16]

Figure 1 shows the NMR intensity of holo-BLA, of unfolded BLA in 6 M GdnDCl, and of dioxan in 6 M GdnDCl as a function of the gradient strength in a PFG-SLED experiment using constant diffusion periods. Equation 1[14,15] can be fitted to these NMR intensities $I(g)$ at the respective gradient strength g. The diffusion

$$I(g) = A \exp(-dg^2) \qquad (1)$$

coefficient D is proportional to d, and A represents the amplitude of the Gaussian diffusion curve.

Table 1 summarizes the equilibrium diffusion coefficients of six different states of BLA under various conditions relative to that of dioxan. Release of the Ca^{2+} cofactor or low pH conditions increase the hydrodynamic radius only marginally. A compact molten globule state of α-lactalbumin at equilibrium was also found by SAXS,[3,9,12] DLS,[13,17] and ^{17}O relaxation studies.[18] Unfolding of the oxidized peptide chain in 6 M GdnDCl, however, increases R_h by 30%, and reduction and carboxymethylation results in an increase of 68%.[19] This supports a recent ^{17}O relaxation dispersion study about internal and external hydration of unfolded BLA, which showed a persistent residual structure which disappears upon reduction.[18] The increase of R_g of α-lactalbumin upon unfolding varies between 60[3] and 90%[12] using SAXS.[20]

The diffusion coefficient of the kinetic MG of BLA was then determined by time-resolved PFG-SLED NMR experiments. To initiate refolding, 4.9 mM unfolded BLA in 6 M GdnDCl was diluted 10-fold within the NMR spectrometer by using a rapid mixing device as described previously.[21,22] At 5 °C, the kinetic MG forms rapidly (in the dead-time of the experiment), and in the absence of Ca^{2+} it converts to folded apo-BLA with a rate constant of about 0.15 min^{-1}. Forty-two PFG-SLED experiments could be recorded during this folding reaction with 10 different gradient strengths per experiment. Assuming a two-state model for the reaction from the MG to the native state, the total intensities of all protein resonances $I(g,t)$ in eq 2 depend on the known gradient strength g at every time point and upon the relative populations of the MG state and of the native state N, as determined by the reaction rate constant k. The differences in the relaxation rates between the MG and the native state are

* Correspondence should be addressed to Jochen Balbach, Biochemie III, Universität Bayreuth, D-95440 Bayreuth, Germany. Telephone: ++49 921 553545. Fax: ++49 921 553661. E-mail:jochen.balbach@uni-bayreuth.de.
† Laboratorium für Biochemie.

(1) Arai, M.; Ikura, T.; Semisotnov, G. V.; Kihara, H.; Amemiya, Y.; Kuwajima, K. *J. Mol. Biol.* **1998**, *275*, 149−162.
(2) Eliezer, D.; Jennings, P. A.; Wright, P. E.; Doniach, S.; Hodgson, K. O.; Tsuruta, H. *Science* **1995**, *270*, 487−488.
(3) Semisotnov, G. V.; Kihara, H.; Kotova, N. V.; Kimura, K.; Amemiya, Y.; Wakabayashi, K.; Serdyuk, I. N.; Timchenko, A. A.; Chiba, K.; Nikaido, K.; Ikura, T.; Kuwajima, K. *J. Mol. Biol.* **1996**, *262*, 559−574.
(4) Segel, D. J.; Bachmann, A.; Hofrichter, J.; Hodgson, K. O.; Doniach, S.; Kiefhaber, T. *J. Mol. Biol.* **1999**, *288*, 489−499.
(5) Nöppert, A.; Gast, K.; Zirwer, D.; Damaschun, G. *Folding Des.* **1998**, *3*, 213−221.
(6) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 868−893.
(7) Kuwajima, K. *Proteins: Struct., Funct., Genet.* **1989**, *6*, 87−103.
(8) Schulman, B. A.; Kim, P. S.; Dobson, C. M.; Redfield, C. *Nat. Struct. Biol.* **1997**, *4*, 630−634.
(9) Dolgikh, D. A.; Abaturov, L. V.; Bolotina, I. A.; Brazhnikov, E. V.; Bychkova, V. E.; Gilmanshin, R. I.; YuO, L.; Semisotnov, G. V.; Tiktopulo, E. I.; Ptitsyn, O. B. *Eur. Biophys. J.* **1985**, *13*, 109−121.
(10) Flanagan, J. M.; Kataoka, M.; Fujisawa, T.; Engelman, D. M. *Biochemistry* **1993**, *32*, 10359−10370.
(11) Kataoka, M.; Nishii, I.; Fujisawa, T.; Ueki, T.; Tokunaga, F.; Goto, Y. *J. Mol. Biol.* **1995**, *249*, 215−228.
(12) Kataoka, M.; Kuwajima, K.; Tokunaga, F.; Goto, Y. *Protein Sci.* **1997**, *6*, 422−430.
(13) Gast, K.; Damaschun, H.; Misselwitz, R.; Müller-Frohne, M.; Zirwer, D.; Damaschun, G. *Eur. Biophys. J.* **1994**, *23*, 297−305.
(14) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16424−16431.
(15) Jones, J. A.; Wilkins, D. K.; Smith, L. J.; Dobson, C. M. *J. Biomol. NMR* **1997**, *10*, 199−203.
(16) The radius of a spherical body in a continuous fluid is inversely proportional to the diffusion coefficient D (Stokes−Einstein equation). D can be determined by pulsed field gradient spin−echo experiments.[27] For nonspherical molecules such as proteins, the spherical radius has to be substituted by the effective hydrodynamic radius R_h, and a NMR pulse sequence (PFG-SLED) with a "longitudinal echo" must be applied, because of the rapid T_2 relaxation of large molecules during the diffusion period.[15,28] As proposed by Jones et al., we used dioxan as an internal standard to calibrate the system.[15]
(17) Gast, K.; Zirwer, D.; Müller-Frohne, M.; Damaschun, G. *Protein Sci.* **1998**, *7*, 2004−2011.
(18) Denisov, V. P.; Jonsson, B. H.; Halle, B. *Nat. Struct. Biol.* **1999**, *6*, 253−260.
(19) Wilkins, D. K.; Ph.D. Thesis, University of Oxford, 1998.
(20) It should be noted that R_g represents an ensemble average, whereas R_h a time average. Furthermore, R_h is sensitive to solvent molecules bound to the surface, diffusion of an extended non-globular chain follows not strictly the Stokes−Einstein equation, and an exact determination of R_g at high concentrations of denaturants is very difficult. Therefore, a direct comparison of R_h and R_g of the unfolded state is very limited.[14]
(21) Balbach, J.; Forge, V.; van Nuland, N. A. J.; Winder, S. L.; Hore, P. J.; Dobson, C. M. *Nat. Struct. Biol.* **1995**, *2*, 865−870.
(22) van Nuland, N. A. J.; Forge, V.; Balbach, J.; Dobson, C. M. *Acc. Chem. Res.* **1998**, *31*, 773−780.

10.1021/ja994514d CCC: $19.00
Published on Web 05/31/2000

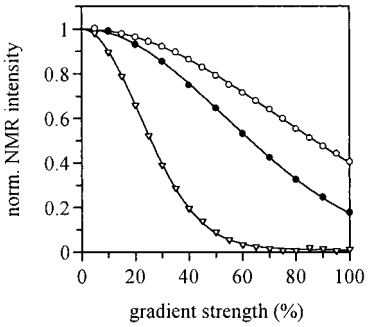

Figure 1. Normalized NMR intensity at different gradient strengths in PFG-SLED NMR experiments of (●) holo-BLA, (○) BLA in 6 M GdnDCl, and (▽) dioxan in 6 M GdnDCl. Solid lines represent a nonlinear, least-squares fit of eq 1 to the data.[25] The relative diffusion constants derived from these fits are given in Table 1.

Table 1. Relative Diffusion Constants of Different States of BLA at 5 °C

BLA state	d_{Diox}/d_{BLA} (%)[a]	BLA state	d_{Diox}/d_{BLA} (%)[a]
holo	9.21 ± 0.09 (100)	MG (kin.)[c]	10.77 ± 0.11 (108)
apo[b]	9.55 ± 0.10 (104)	unfolded (ox.)[d]	12.0 ± 0.12 (130)
MG (pH 2)	9.85 ± 0.10 (107)	unfolded (red.)[e]	15.5 ± 0.16 (168)

[a] Percentages are calculated relative to holo-BLA. [b] The apo form was generated by adding 2 mM EDTA to the NMR sample to complex Ca^{2+}. [c] Kinetic MG at 0.49 mM from Table 2. [d] Unfolded state in 6 M GdnDCl with oxidized disulfide bonds. [e] Unfolded state in 0 M GdnDCl with reduced and carboxymethylated cysteine side chains.[19] The protein concentration for all equilibrium experiments was 0.49 mM.

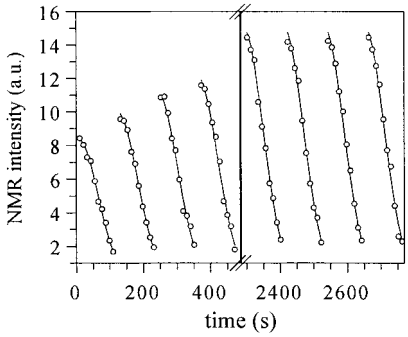

Figure 2. Two sections of the NMR intensity in a time-resolved 600 MHz PFG-SLED NMR experiment during the refolding of BLA from 6 M GdnDCl at 5 °C. Each open symbol represents the integral over the resonances between 0.92 and 1.86 ppm in the respective 1D diffusion spectrum, where the MG and the native state have identical integrals in the 1D NMR spectrum. Solid lines represent the nonlinear, least-squares fit of eq 2 to the data.[26]

reflected in the respective amplitudes A_{MG} and A_N.

$$I(g,t) = A_{MG}\exp(-d_{MG}g^2)\exp(tk) + A_N\exp(-d_Ng^2)(1-\exp(tk)) \quad (2)$$

Figure 2 illustrates the regression (solid line) of eq 2 to the experimental data (open circles) of the real-time PFG-SLED experiments. The optimized parameters are summarized in Table 2. The gradient strength was increased linearly every 10.9 s from 10 to 100% by steps of 10% periodically during the entire refolding reaction. Using eq 2 instead of fitting each set of 10 different gradient strengths to eq 2 separately, increases the time resolution by a factor 10.[23]

These real-time PFG-SLED experiments show that the transiently populated kinetic molten globule at typical protein concentrations for real-time NMR spectroscopy with BLA resembles the equilibrium MG and the native state in its hydrodynamic radius (Tables 1 and 2). This indicates that, starting from unfolded, but not fully extended, oxidized BLA in 6 M GdnDCl, the major compaction occurs concomitantly with the population of the kinetic MG state before the rate-limiting folding step. This compact state with an almost native amount of secondary structure[7,24] allows an efficient search for native side-chain contacts during the final folding step. These results confirm earlier, more indirect, findings by amide proton exchange and NOE transfer experiments[24] of the kinetic MG and a recent study using stopped-flow DLS.[17]

Table 2. Relative Diffusion Constant, Amplitude of the Gaussian Diffusion Curve, and Refolding Rate of the Kinetic Molten Globule of BLA at 5 °C

concentration (mM) ion	d_{native}/d_{MG} (%)[a]	A_{MG}/A_{native} (%)[a]	folding rate (h^{-1})
0.31	107 ± 11	55.9 ± 3.4	9.0 ± 0.8
0.49	108 ± 5	55.7 ± 1.2	7.9 ± 0.4

[a] Values for d_{native}, d_{MG}, A_{native}, and A_{MG} are obtained by fitting eq 2 to the data as illustrated in Figure 2.

The radius of a transient protein folding state was pioneered by Eliezer et al. for apo-myoglobin using stopped-flow SAXS.[2] This collapsed state is also almost as compact as the native state and resembles the equilibrium MG in its R_g value. The burst phase intermediate of β-lactoglobulin, which has non-native α-helical structure elements, is only 10% larger than the corresponding native state.[1] These examples illustrate that compaction of the protein chain occurs early in refolding, even though the side chain packing has not finished, as indicated by the absence of a near UV−CD spectrum. On the other hand, stopped-flow SAXS experiments with bovine carbonic anhydrase (BCAB), yeast phosphoglycerate kinase (PGK),[3] and hen egg white lysozyme,[4] as well as DLS experiments[5] with PGK and ribonuclease A showed that substantially extended intermediates and both, reversible and irreversible oligomerization can occur during un- and refolding reactions and that the major compaction can also occur during the rate-limiting folding reaction.

The results presented here indicate that time-resolved NMR diffusion experiments can be used to follow changes in the hydrodynamic radius of a folding protein chain in real time. A standard NMR spectrometer can be used in contrast to time-resolved SAXS protein folding experiments, for example, which require synchroton radiation.

Acknowledgment. I thank D. K. Wilkins, J. A. Jones, K. Gast, and F. X. Schmid for very helpful discussions and C.M. Dobson for his enthusiastic support and spectrometer time at the Oxford Center of Molecular Science. My research is supported by the Fonds der Chemischen Industrie, Gesellschaft Deutscher Chemiker and Deutscher Akademischer Austauschdienst. This paper is dedicated to Professor Horst Kessler, Technische Universität München, on the occasion of his 60th birthday.

JA994514D

(23) The rate constant k was either determined from the kinetic dataset by fitting a single exponential function to intensities at constant gradient strength or by recording an interleaved 1D spectrum after every gradient block. The present instrumentation requires about 100 s to record one set of gradient strengths and is mainly limited by the signal/noise ratio feasible with the PFG-SLED experiment. This time resolution is sufficient to study slow folding reactions of an increasing number of proteins accessible by real-time NMR methods (see references in review[22]). Stopped-flow SAXS studies allow data acquisition within the first 100 ms of the reaction, and continuous-flow SAXS between 14 and 100 ms.[4]

(24) Forge, V.; Wijesingha, R. T.; Balbach, J.; Brew, K.; Robinson, C. V.; Redfield, C.; Dobson, C. M. *J. Mol. Biol.* **1999**, *288*, 673−688.

(25) The proposed pulse sequence by Jones et al.[15] has been used at 600 MHz, 5 °C, with a 30 ms diffusion time. Each curve is an average of three separate experiments with typically 64 scans per gradient strength and a sweep width of 9615 Hz. The NMR intensity of the protein (0.49 mM) was obtained by integrating the spectrum between 0.65 and 0.75 ppm.

(26) A series of 42 PFG-SLED experiments such as in Figure 1 were performed during 5000 s with eight scans for each of the 10 different gradient strengths (10−100%, step 10%) per experiment. The first four sets up to 480 s and four sets between 2290 and 2770 s are presented.

(27) Stejskal, E. O.; Tanner, J. E. J. *Chem. Phys.* **1965**, *42*, 288−292.

(28) Gibbs, S. J.; Johnson, C. S., Jr. *J. Magn. Reson.* **1991**, *93*, 395−402.