cloned into the Sph I site of pSL301 (Invitrogen). Colonies were screened for inserts by PCR with T7 and T3 sequences located outside the cloning site as primers.
10. Selected clones were manually sequenced as described [G. Del Sal, G. Manfioletti, C. Schneider, *Biotechniques* **7**, 514 (1989)] with 5′-GACGTCGAC-CTGAGGTAATTATAACC-3′ as primer. Sequence files were analyzed by means of the SAGE software group (*7*), which identifies the anchoring enzyme site with the proper spacing and extracts the two intervening tags and records them in a database. The 1000 tags were derived from 413 unique ditags and 87 repeated ditags. The latter were counted only once to eliminate potential PCR bias of the quantitation, as described in the text.
11. J. H. Han, L. Rall, W. J. Rutter, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 110 (1986); J. Takeda, H. Yano, S. Eng, Y. Zeng, G. I. Bell, *Hum. Mol. Genet.* **2**, 1793 (1993).
12. Pancreatic mRNA from the same preparation was used for SAGE and to construct a cDNA library in the ZAP Express vector. The ZAP Express cDNA Synthesis kit (Stratagene) was used according to the manufacturer's protocol. Analysis of 15 randomly selected clones indicated that 100% contained cDNA inserts. Plates containing 250 to 500 plaques were hybridized as described [J. M. Ruppert *et al.*, *Mol. Cell. Biol.* **8**, 3104 (1988)]. The cDNA probes for trypsinogen 1, trypsinogen 2, procarboxypeptidase A1, chymotrypsinogen, and elastase IIIB were derived by RT-PCR from pancreas RNA. The sequences of primers are available from the authors upon request. The trypsinogen 1 and 2 probes were 93% identical and hybridized to the same plaques under the conditions used. Likewise, the elastase IIIB probe was >95% identical to protease E.
13. Plates containing 250 to 2000 plaques were hybridized to oligonucleotide probes with the same conditions previously described for standard probes except that the hybridization temperature was reduced to room temperature (*12*). Washes were performed in 6× standard saline citrate–0.1% SDS for 30 min at room temperature. The probes consisted of 13-bp oligonucleotides that were labeled with [γ^{32}-P]ATP through use of T4 polynucleotide kinase.
14. An ABI 377 sequencer can produce a 451-bp read for 36 templates in a 3-hour run [(451 bp/11 bp per tag) × 36 = 1476 tags].
15. P. L. Deininger *et al.*, *J. Mol. Biol.* **151**, 17 (1981).
16. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lippman, *J. Mol. Biol.* **215**, 403 (1990).
17. Supported by NIH grants CA57345, CA35494, GM07309. B.V. is an American Cancer Society Research Professor and an Investigator of the Howard Hughes Medical Institute. We thank S. Kern, B. D. Nelkin, and members of our laboratories for their critical review and valuable discussions.

14 June 1995; accepted 11 August 1995

TECHNICAL COMMENT

The Radius of Gyration of an Apomyoglobin Folding Intermediate

Apomyoglobin (apoMb) forms a stable compact partially folded state under acidic conditions (*1*). This "molten globule" intermediate is slightly expanded relative to the native form of the protein, with a radius of gyration (R_g) of 23 (± 2) Å versus 19 (± 1) Å (*2*), and shows stable secondary structure (*3*) in the A, G, and H helices (Fig. 1).

We demonstrated recently, with the use of stopped-flow circular dichroism and pulse-labeling hydrogen exchange measurements, that the earliest detectable intermediate (formed within 6 ms) in the apoMb kinetic refolding pathway closely resembles the equilibrium molten globule state populated under acid conditions (*4*). A key question remained as to how compact this kinetic intermediate is compared to the equilibrium and native states. The cooperative unfolding of the kinetic intermediate and the significant protection from amide proton exchange (as compared to corresponding isolated peptides in solution) led us to propose that the kinetic intermediate is also compact (*4*, *5*). Such a proposal could best be verified by direct determination of the size of the protein as it folds, but measurements of this nature were not feasible at the time.

Newly developed improvements in time-resolved small angle x-ray scattering (SAXS) experiments allow direct measurement of the time-dependent change of R_g of a protein as it folds in the millisecond to second time frame (*6*, *7*). We initiated studies of the refolding of apoMb using this technique, under conditions similar to those employed in our previous work (*4*). SAXS data collected during the first 100 ms after initiation of the refolding reaction (*8*) are shown in Fig. 2.

Data collected from the fully refolded protein and unfolded protein are given for comparison (Fig. 2). The data obtained 100 ms after the initiation of folding are within experimental error of the data obtained for the refolded protein, and easily distinguishable from data obtained for the unfolded state. An R_g value of 23 (± 2) Å is obtained at 100 ms, only 1 Å greater than the 22 (± 1) Å value obtained for the refolded protein. By contrast, the unfolded state has an R_g of 34 (± 2) Å. The slightly higher than expected R_g value obtained for the refolded state may result from either experimental error (*9*) or a small degree of sample aggregation owing to radiation damage during exposure. It is possible that the R_g value obtained at 100 ms is similarly inflated, and it may therefore be considered an upper bound on the true R_g.

Our conclusion that the intermediate is compact is based on the small differences

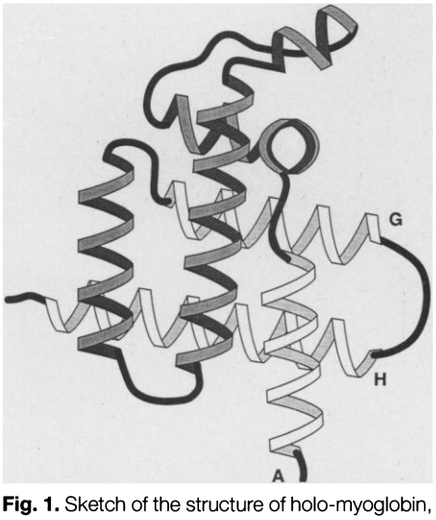

Fig. 1. Sketch of the structure of holo-myoglobin, illustrating the location of the A, G, and H helices, which are present in both the equilibrium and kinetic folding intermediates of the apoprotein.

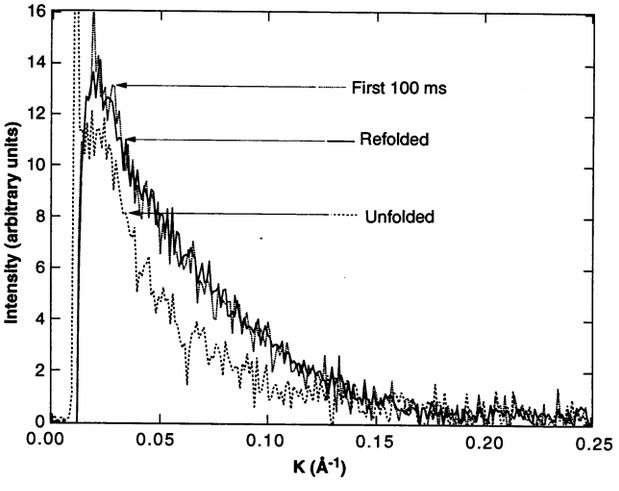

Fig. 2. SAXS data from sperm whale apomyoglobin after 100 ms of folding, after 4.2 s of folding, and in the unfolded state. Detected intensity is plotted as a function of *K*. Data from the unfolded state is scaled to match the folded state data at zero scattering angle. The data obtained from the fully folded protein and that obtained after 100 ms of folding are barely distinguishable from each other and are different from the data for the unfolded protein.

between both the raw SAXS data and the R_g values from the kinetic intermediate and from the fully refolded protein. It is unnecessary to invoke specific models to reach this conclusion. Indeed, the low resolution of SAXS data and the uncertainties inherent in time-resolved SAXS measurements make it both inappropriate and unwise to attempt to interpret the current data in terms of specific structural models.

Taken as an upper bound, the 23 (± 2) Å R_g obtained at 100 ms illustrates that the first intermediate observed in the kinetic refolding reaction of apoMb is at least as compact as the equilibrium "molten globule" state of apoMb. The fact that this R_g value is only 1 Å greater than the value measured for the fully refolded protein, together with the great similarity of the actual SAXS data at 100 ms to the data from the refolded protein, suggests that the kinetic intermediate may be nearly as compact as the native state itself (*10*). SAXS data collected during the first 20 ms of folding indicate these same results, but with a lower signal to noise ratio (*11*). Thus the present experiments provide a direct measurement of the size of the early kinetic folding intermediate of apoMb.

David Eliezer
Patricia A. Jennings*
Peter E. Wright
Department of Molecular Biology,
The Scripps Research Institute,
La Jolla, CA 92037, USA

Sebastian Doniach
Department of Applied Physics,
Stanford University,
Stanford, CA 94305, USA
Keith O. Hodgson
Department of Chemistry and
Stanford Synchrotron Radiation Laboratory,
Stanford University
Hirotsugu Tsuruta
Stanford Synchrotron Radiation Laboratory

*Present address: Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92037, USA

REFERENCES AND NOTES

1. Y. V. Griko, P. L. Privalov, S. Y. Venyaminov, V. P. Kutyshenko, *J. Mol. Biol.* **202**, 127 (1988).
2. D. Eliezer *et al.*, unpublished data. Similar values have been reported by K. Gast *et al.* [*Eur. Biophys. J.* **23**, 297 (1994)] and for a trichloroacetate stabilized equilibrium intermediate by I. Nishii, M. Kataoka, F. Tokunaga, and Y. Goto [*Biochemistry* **33**, 4903 (1994)].
3. F. M. Hughson, P. E. Wright, R. L. Baldwin, *Science* **249**, 1544 (1990).
4. P. A. Jennings and P. E. Wright, *ibid.* **262**, 892 (1993).
5. P. A. Jennings, H. J. Dyson, P. E. Wright, in *Statistical Mechanics, Protein Structure, and Protein Substrate Interactions*, S. Doniach, Ed. (Plenum, New York, 1994).
6. D. Eliezer *et al.*, *Biophys. J.* **65**, 912 (1993).
7. Data were collected at the Stanford Synchrotron Radiation Laboratory with the use of a new high-flux multi-layer x-ray monochromator calibrated to a photon energy of 8980 electron volts. The bandwidth of radiation transmitted by this monochromator is 10-fold greater than standard Si(III) double-crystal monochromators, providing the additional flux necessary for studies at protein concentrations low enough to avoid dimerization of highly association-prone folding intermediates.
8. Refolding was triggered by rapid dilution of 10 mg/ml protein in 5.6 M urea to 1.4 mg/ml protein in 0.8 M urea. The dead time of the rapid mixer (Unisoku Inc., Osaka) is on the order of 10 ms. Kinetic data were accumulated from 1200 individual mixing events. Radii of gyration were extracted from the background-subtracted data using Guinier fits to the region $K = 0.034$ to $K = 0.063$ where K, the scattering vector amplitude, equals $4\pi\sin(\theta/2)/\lambda$ (θ is the scattering angle and λ is the x-ray photon wavelength).
9. The lowest possible protein concentration was used to prevent the possible oligomerization of kinetic folding intermediates. An unfortunate consequence of such a low concentration is that a much larger fraction of the detected x-ray photons are from background sources, leading to larger experimental errors.
10. In such a highly compact intermediate it seems likely that the polypeptide chain segment that forms helices B through F in native myoglobin would also have undergone some degree of collapse. The possibility that the kinetic intermediate may be more compact than the equilibrium intermediate is supported by our previous observation (*5*) that the apoMb equilibrium intermediate would be about 1 kcal/mol more stable under the conditions used for the kinetic studies than under the partially denaturing conditions in which it is typically studied.
11. The time-resolved circular dichroism and amide exchange data (*4*) indicate a lag phase between 5 and 350 ms into the folding reaction, where no change in mean residue ellipticity or amide proton protection is observed.
12. SSRL is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, and Office of Health and Environmental Research, and by the NIH Biomedical Research Technology Program. Additional funding was provided by the National Institutes of Health through grant numbers DK34909 (to P.E.W.) and RR01209 (to K.O.H.) and through postdoctoral award GM14541 (to P.A.J.).

2 May 1995; revised 6 August 1995; accepted 22 August 1995