Folding of a large protein at high structural resolution

Benjamin T. Waltersa,b, Leland Maynea, James R. Hinshawc, Tobin R. Sosnickd,e, and S. Walter Englandera,f

Kinetic folding of the large two-domain maltose binding protein (MBP; 370 residues) was studied at high structural resolution by an advanced hydrogen-exchange pulse-labeling mass-spectrometry method (HX MS). Dilution into folding conditions initiates a fast molecular collapse into a polyglobular conformation (<20 ms), determined by various methods including small angle X-ray scattering. The compaction produces a structurally heterogeneous state with widespread low-level HX protection and spectroscopic signals that match the equilibrium melting postransition-state baseline. In a much slower step (7-s time constant), all of the MBP molecules, although initially heterogeneously structured, form the same distinct helix plus sheet folding intermediate with the same time constant. The intermediate is composed of segments that are distant in the MBP sequence but adjacent in the native protein where they close the longest residue-to-residue contact. Segments that are most HX protected in the early molecular collapse do not contribute to the initial intermediate, whereas the segments that do participate are among the less protected. The 7-s intermediate persists through the rest of the folding process. It contains the sites of three previously reported destabilizing mutations that greatly slow folding. These results indicate that the intermediate is an obligatory step on the MBP folding pathway. MBP then folds to the native state on a longer time scale (~100 s), suggestively in more than one step, the first of which forms structure adjacent to the 7-s intermediate. These results add a large protein to the list of proteins known to fold through distinct native-like intermediates in distinct pathways.

SAXS | HDX | protein collapse | denatured state ensemble

Fifty years after Anfinsen’s seminal demonstration that an unfolded protein can refold spontaneously when placed under native conditions, major questions concerning the folding process remain unanswered (1, 2). What is the unfolded state like, its degree of compaction, the reality and character of residual structure before folding begins, and its possible role in guiding the folding process (3–7)? Analogous questions relate to folding intermediates and the folding pathway itself. Do proteins fold through many alternative independent pathways as earlier theoretical investigations have suggested (8–12), or do they fold through necessary intermediates and the folding pathway itself (13), as a growing list of experimental observations indicate (14, 15)? To answer these questions, it will be necessary to define experimentally the intermediate forms that proteins move through on their way to the native state. The problem has been that these transient states are beyond the reach of the usual high-resolution crystallographic and NMR structural methods. Most experimental folding studies have therefore relied on low-resolution optical methods that can follow folding in real time but rarely provide the structural information necessary to resolve the basic mechanistic questions.

Recent work has demonstrated an advanced hydrogen-exchange pulse-labeling mass-spectrometry technology (HX MS) that is able to detect and characterize local structure, even when it is only transiently present during the course of kinetic folding (15, 16). The HX pulse-labeling approach provides a snapshot of main chain amide sites that are protected against HX labeling by H bonds present at the time of the labeling pulse (17, 18). HX MS measurements can determine the position, stability, and dynamic behavior of native and nonnative H-bonded structure and whether it persists or dissipates in subsequent folding. In a recent application, the method was able to describe the structure and time-dependent formation of three sequential native-like folding intermediates in the 155-residue ribonuclease H protein (15).

Protein folding studies, whether theoretical or experimental, have been limited to relatively small proteins, with few exceptions. However, biological proteomes and the considerations they raise are dominated by large proteins (19). Here we extend the powerful HX MS technology to the two-domain, 370-residue, maltose binding protein (MBP). MBP is synthesized in the Escherichia coli cytoplasm and transported to the periplasm where it serves as a soluble receptor for the high-affinity capture and import of maltose and maltodextrins (20). The protein folds in vivo after deletion of a signal sequence (21); we study here the mature protein with the signal sequence deleted.

When unfolded MBP is placed into native conditions, we find that it rapidly adopts a dynamic collapsed state, which can lead to aggregation in vitro when the concentration is >1 μM and to inclusion body formation in vivo (22). Folding to the native state occurs much more slowly even in the absence of aggregation, with all molecules moving through one or more intermediate states to the native state. The HX MS experiment provides incisive information on the nature of the initially collapsed state, the slow formation and identity of at least one on-pathway native-like intermediate, and the even slower emergence of native structure.

Results

Equilibrium Stability and Unfolding. Previous reports have described the global unfolding of wild-type MBP with and without bound substrate analog and various mutations when driven by denaturant or temperature. In our hands denaturation follows a distorted sigmoidal curve with midpoint of the dominant unfolding phase at 0.6 M guanidinium chloride (GdmCl) and substantial posttransition baseline curvature (Fig. 1). The melting curve can be fit by the six-parameter, two-state, Santoro–Bolen equation (23) leading to equilibrium stability variously reported up to 14.5 kcal/mol. However, the prominent posttransition MBP baseline is not well fit by the two parameters allotted by Santoro–Bolen, making extrapolation and stability analysis problematic.

Significance

This study characterizes the initially collapsed but not yet folded denatured state ensemble of the large two-domain maltose binding protein (MBP), its contribution to the folding process, and an obligatory on-pathway native-like folding intermediate and probable further intermediates. The results add MBP to the growing list of proteins that fold through a distinct pathway with obligatory intermediates.


The authors declare no conflict of interest.

1To whom correspondence should be addressed. E-mail: engl@mail.med.upenn.edu.
The interpretation of equilibrium melting baselines has long been unclear. The results in Fig. 1A compare fluorescence and circular dichroism (CD) signals of the posttransition unfolded state with the signals produced in the initial burst phase refolding step described below. Similar burst phase signals have been thought to represent either a rapidly formed folding intermediate or, more simply, the still unfolded state as it exists under native conditions (24). Results described below characterize this condition for MBP and find it to represent a dynamic heterogeneously compacted condition that has little effect on the subsequent folding process except perhaps in a negative sense.

Whole Protein Folding Data. MBP unfolded in 2 M GdmCl was diluted into native conditions (0.2 M GdmCl, pH 9, 20 °C, 0.5 mM protein) and kinetic folding was observed by a variety of optical methods (Fig. 1B). Within the 4- to 20-ms dead time of the various spectroscopic observations, MBP exhibits a burst phase increase in tryptophan fluorescence and ANS binding and the formation of 40% of its native CD corresponding to 20% helical content, followed by much slower folding to the native state. The CD result suggests the development of significant helical content, which can be observed independently by HX protection (below). Fig. 1C shows that the amplitude of the burst phase (CD and fluorescence) decreases slowly and linearly with increasing denaturant rather than sigmoidally. This result is against a specific barrier-crossing process to some defined intermediate structure.

Fig. 1D shows $P(r)$, the distribution of pairwise atom-to-atom distances obtained by small angle X-ray scattering (SAXS) after 0.7 s of folding. The pairwise distribution function displays a form that is more compact than the unfolded state but more extended than the globular native protein, indicated by the tailing to longer interatom distances. The averaged parameter, radius of gyration ($R_g$), after 0.7 s of folding is 36 Å (0.2 M GdmCl) compared with 22 Å for native and 73 Å for unfolded MBP in 2 M GdmCl. Computed shape reconstructions (25) repeatedly produce envelopes like those in Fig. 1D, Inset, which we refer to as polyglobular rather than globular structure.

Fig. 1E shows the charge state distribution (CSD) produced by injecting MBP by electrospray ionization (ESI) into a mass spectrometer within ~50 msec of initiating folding. The spectrum is shifted from the high charge state pattern characteristic of un-folded MBP in 2 M GdmCl. Computed shape reconstructions (25) constrained the still unfolded state as it exists under native conditions (26, 27). A small population fraction with CSD similar to that of the unfolded protein is also seen but it can be noted that the populations measured in this way are greatly biased toward exaggerating the more unfolded component (26). All of these results point to a fast molecular compaction into an extended polyglobular condition.

The fast molecular collapse is followed by unusually slow folding, with final native state acquisition on a ~100-s time scale. Double jump experiments (Fig. 1B) designed to maintain proline residues in their native isomeric configuration (unfolded for 3 s, immediately refolded at 0.5 μM MBP) show modestly faster folding (approximately two times) indicating that the slow folding is affected but not determined by misisomerized proline barriers. CD and fluorescence experiments show essentially no dependence of the folding rate on solvent viscosity (glycerol; Fig. 1F). This observation suggests that polypeptide reconfiguration during the conformational searching that ultimately organizes the native structure is not limited by diffusional searching of the polypeptide chain through free solvent but rather by the difficulty of conformational reorganization within and between condensed polyglobular regions. This behavior studied in smaller proteins is commonly attributed to so-called internal friction (6, 28). Significant differences should be appreciated. The present time scale is seven orders of magnitude slower than is observed in small molecules, in part due to the degree of chain collapse, and also because the folding event requires a specific nearly simultaneous multipoint interaction rather than a general two-point interaction as for example in a FRET experiment.

In summary, upon dilution from unfolding denaturant, MBP experiences a fast molecular collapse into an ensemble of compact polyglobular forms and then folds much more slowly in a way that is limited by the difficulty of chain reconfiguration. However, these widely used methods only monitor whole molecule behavior. They provide little detailed information about structure in the compact unfolded state or the folding mechanism that produces the native state.

**HX Pulse Labeling and Fragment Separation Mass Spectrometry.** To study the fast and slow stages of MBP folding at higher structural resolution, we used a quench-flow HX pulse-labeling experiment with analysis by fragment-separation mass spectrometry (Fig. 2A) (15). MBP was unfolded and spontaneously deuterated by H-to-D exchange in D$_2$O, then diluted into folding conditions still in D$_2$O to avoid confounding D-to-H exchange during the long prefolding time. At a series of time points during kinetic folding, the deuterated protein was exposed to a short D-to-H labeling pulse (12–43 ms, pH 9, 20 °C, where average D-to-H exchange lifetime for unprotected amides is ~1 ms). Amides in

![Fig. 1](image-url)
already formed H-bonded structure tend to be protected and remain deuterated, whereas unprotected amides become protonated. The H-labeled protein was immediately quenched into a slow HX condition (pH 2.5, 0 °C), injected into an online analysis system, cleaved into small peptide fragments by acid proteolysis (pepsin), the peptide fragments were quickly separated by HPLC and mass spectrometry (LTQ Orbitrap XL) (29), and then identified and analyzed for carried D label (30).

The 225 unique peptide fragments obtained, often with multiple charge states, are indicated in Fig. 2B. Each peptide fragment monitors the folding behavior of the protein segment that it represents. The pulse-labeling analysis provides a snapshot of the HX protection of amides in each protein segment at the time of the labeling pulse. We associate HX protection with H-bond formation (31, 32). Analysis of the peptide-bound deuterium tracks the position and time dependence of amide H-bond formation throughout the protein. The degree of resistance of D label to exchange in the H-labeling pulse measures the stability and dynamic behavior of the structure that is formed.

The Rapidly Collapsed State: Structure, Stability, Dynamics. In the initial molecular collapse, sites in all of the peptides throughout MBP become mildly protected. HX pulse-labeling results at the 0.5-s time point (Fig. 3 A and B) show that when the D-to-H labeling pulse is increased from 12 to 43 ms (~10–40 HX lifetimes), the number of deuterons that resist exchange decreases, indicating relatively unstable H bonding and rapid fluctuational dynamics.

Most peptides appear as unimodal isotopic envelopes with low-level HX protection. Some peptides exhibit broadened envelopes, indicating protein-to-protein heterogeneity in HX protection at those positions. Fig. 3A plots the increment in centroid mass (sum of retained deuterons) as a function of pulse time for a sampling of peptides distributed through the sequence. To fit the broadened spectra, a minimum of two differently labeled fractions is required, one more protected (heavy) and one less protected (light). For these, the fraction “heavy” is plotted against pulse time in Fig. 3B. An example of each type of spectrum is in Fig. 3 C and D.

To display HX protection through the protein, the normalized centroid increment for the 12-ms pulse data are plotted as a black line through the protein sequence in Fig. 3E. Average D-occupancy levels of ~30% corresponds to HX protection factors of ~10–20, indicating stability against exchange in the range of 1 kcal/mol and H-bond reformation times, once HX-competent opening occurs, faster than ~1 msec. Among the broadened spectra, four regions (90–112, 155–180, 235–275, and 340–346) show a large fractional heavy population. For the blue bars, the gray bars in Fig. 3E show the heavy fraction at the 12-ms pulse. Two regions, residues 90–112 and 340–346, show especially high fractional protection.

An Obligatory Intermediate. From the compacted, heterogeneous, dynamic, polyglobular protein, a specific intermediate structure emerges on a longer time scale. The blue curves in Figs. 3 and 4 represent 15 peptides (residues 9–43, 60–62, and 260–278) that experience a concerted bimodal transition to a form that is highly protected against the HX labeling pulse (example in Fig. 4E). All become well protected with exponential time constant of 7 s. The number of protected sites is equal to the sites protected in the native protein. These peptides represent protein segments that are sequentially distant but are adjacent in the native structure (Fig. 4 B–D) where they come together to close the longest site-to-site loop in the native protein, analogous to some previous results (33). They form the two helices and two of the four β-strands pictured in Fig. 4C. Two intervening strands do not clearly show a bimodal transition (four peptides) because too few deuterons are incorporated, but they do protect several sites during the same ~7-s time period, matching the number of sites expected by H bonding in the native protein. These results point to the concerted formation of the native-like substructure shown in Fig. 4 C and D (in the MBP N domain). MS envelopes for these peptides directly show that these segments undergo the same concerted structural transition in 100% of the protein population (Fig. 4E). The time-dependent folding results show that the structure, once formed, is retained through final folding (e.g., Fig. 4F).

Previous studies have found several destabilizing MBP mutations that greatly slow folding and increase the tendency toward aggregation (V8G, Y283D, G32/I33 to D32/P33) (34–37). All are contained within the apparent 7-s intermediate structure, represented by stars in Fig. 4C. A similar double mutation at a similar position in the C domain was similarly destabilizing but had no effect on folding rate (35). These results support the conclusion that the 7-s structure formation represents an obligatory, on-pathway, native-like folding intermediate. Interestingly, this first formed intermediate is placed on the solvent-exposed protein surface, not buried in the protein core (Fig. 4D).

Given that the pulse time was 43 ms (~40 HX lifetimes) and the D occupancy is maintained at the native level, the protection of the 7-s intermediate structure against pulse labeling is >100, corresponding to ~3 kcal/mol of stability, assuming EX2 (38) behavior, and/or a dynamic unfolding rate of less than 150 ms.

Slow Native Transition. All of the other peptides transition to a protected state on a ~100-s time scale (Fig. 4A), matching the spectroscopically monitored folding kinetics in Fig. 1B. These segments exhibit a spread of folding rates much broader than is seen for the 7-s class (Fig. 4A; compare also Fig. 4 E–G), suggesting that the spread is significant and that different protein segments may fold sequentially. However, the large number of peptides with halflives between 60 and 120 s fold too closely in time to resolve them definitively into clearly separate groupings. It is interesting that peptides that occupy the earliest part of the spread (green in Fig. 4; residues 78–89) represent MBP segments that are adjacent to the 7-s intermediate structure in the native protein. This is also expected from the sequential stabilization mechanism (14) which posits that folding pathways tend to be sequential because already formed structure tends to guide and stabilize the formation of interacting structure. The slowest folding peptides shown in red are remote, in the C-terminal domain of native MBP (Fig. 4; residues 180–195 and 315–370), also as expected for a defined sequential pathway mechanism.

![Fig. 2.](https://example.com/fig2.png)
The folding history of the peptides that develop high fractional HX protection in the early collapsed phase (Fig. 3) is indicated by the black curves in Fig. 4A. They do not contribute to the 7-s intermediate. The segments that do produce the 7-s intermediate (blue curves in Figs. 3 and 4) are among the least protected segments in the early condensed protein. Thus, measured protection in the early condensed state does not correlate with subsequent structure formation.

With respect to the question of multidomain folding (19), these results indicate that (parts of) the MBP N domain fold first. Reported single molecule forced unfolding experiments unfold the N domain first but this is because the pulling force is applied at the molecular N and C termini, both of which emanate from the N domain (39).

**Discussion**

This work used standard optical methods, SAXS, and a developing HX MS pulse-labeling method to study the folding of the large, two-domain, 370-residue maltose binding protein. Upon mixing into folding conditions, the unfolded MBP polypeptide quickly condenses. The results characterize the initially condensed state, the subsequent formation of an obligatory on-pathway intermediate, and the even slower folding to the native state.

**Protein Condensation.** A quantity of work has focused on the character of the unfolded state and its possible role in guiding subsequent folding (3, 4, 6, 40–42). Is the denatured state ensemble (DSE) compact at the start of the folding process? Is significant prefolding structure present? If so, does it help to guide or hinder the folding process?

The methodology used here provides some answers. Earlier SAXS studies found that the DSE for small two-state folding proteins under native conditions has the same extended Rg as in high denaturant (43), pointing to the absence of distinct structure (2). However, when unfolded MBP is mixed into folding conditions, the 350-residue polypeptide chain rapidly condenses to a polyglobular form. It seems likely that large proteins like MBP with many more possible hydrophobic interactions will bias more toward an initially collapsed condition (44, 45). Hydrophobic interactions that drive condensation bring together sites that allow ANS-to-protein binding (Fig. 1A) and can similarly promote protein-to-protein aggregation. The same situation promotes the binding of exposed hydrophobic sites of condensed but still unfolded large proteins like MBP to the hydrophobic sites of GroEL and other

---

**Fig. 3.** HX protection in the fast collapse phase after 0.5 s of folding. (A) The response to the length of the labeling pulse is plotted in terms of fractional increase in centroid mass. (B) Peptides with broad perhaps bimodal spectra are plotted as the protected population fraction versus the pulse length. (C and D) Envelopes of typical peptides (black curves in A and B) illustrating unimodal and bimodal spectra. The blue curves in A and B represent peptides that form the 7-s intermediate (already ∼7% folded at 0.5 s). (E) Comparison of some peptides (from A and B, color coded) with helical propensity, hydrophobicity, and native conformation (Lower bar).

**Fig. 4.** The folding history of different segments. (A) Blue and black segments are from Fig. 3B. Blue segments go on to form the 7-s intermediate, which persists through subsequent folding (E). Black curves track the early protected segments, which do not contribute to the 7-s intermediate. Segments that fold in the very slow broader grouping are graded from early (green), through gray, to late (red). (B) Contact map colored as for A. (C and D) Placement of the blue, green, and red segments in native MBP. (E–G) HX mass spectra comparing the time-dependent folding of some color-marked segments. The blue 7-s structure includes residues 9–43, 60–62, and 260–278. Residues 78–89 are shown in green; residues 180–195 and 315–370 are in red. Pulse time was 43 ms.
chaperones (46, 47), which shield them from aggregation during their vulnerable slow-folding time period (48).

Over time the entire MBP refolding population self-organizes, forming a distinctly structured, obligatory native-like intermediate, and then proceeds to assemble the native state. What is the relationship between structure in the initial collapsed state and subsequent folding? The general character of the collapsed state is indicated by CD and fluorescence results (Fig. L4), which correlate with the postmelting condition, and by HX results, which indicate widespread, diverse, low-level HX protection among the various sites within each protein molecule. Especially interesting are several unusually well-protected (H-bonded) segments (in Fig. 3E; 90–112, 155–180, 235–275, and 340–346). The detailed spectroscopy of heterogeneous HX labeling is ambiguous. Different scenarios can be considered, depending on HX options [EX1 (38), local fluctuations, unfolding] and the rate of conformational interconversion. Some other observations are indicative. These segments are differently protected from one molecule to another. They are not selected by helical propensity or by helical conformation in the native protein. Rather, there is a correlation with high hydrophobicity, especially for the segments with the highest protection (104–112 and 340–346), which will promote chain burial (32). Unlike the segments that come together to form the 7–5 intermediate, these segments do not interact in the native protein. The segments that do form the intermediate are among the less protected in the initial compaction.

These observations do not favor the possibility that early HX protection presages later structure formation. The compacted condition does contribute to the slow folding in a negative sense because it constrains chain reconfigurational searching. Smaller proteins that do not collapse fold much faster.

Nature of Protein Folding Pathways. In the early history of the protein folding field, it was assumed that proteins fold through discrete intermediates in discrete pathways, like other biochemical pathways. Early theoretical efforts to study folding mechanisms led to a view of a heterogeneous unfolded state, reminiscent of the initial condensed state studied here. It was inferred that proteins then fold through multiple pathways guided by a funnel-shaped energy landscape (8, 10–12, 41, 49–54). Despite a dearth of experimental verification, this view is still current, and experimental as well as theoretical folding results are often phrased in this language. Some spectrometry-based experiments have been taken to suggest more than one folding pathway (55–59), but it has been shown that this kind of data cannot distinguish alternative parallel pathways from a given pathway with alternative misfolding barriers (60, 61).

A quantity of more recent experimental work using HX and associated methods has found much more organized folding behavior. Many proteins have been found to form at least one specific structure, usually the native state. The present work used an advanced mass spectrometry-based analysis of the fragments generated from a trap column, allowing for the detection of fragments with a range of masses. The results indicate that proteins fold through at least two different pathways, which may or may not be identical. This supports the idea that the folding process is not strictly defined by a single, well-defined pathway, but rather involves multiple, parallel pathways.

Materials and Methods

The E. coli apo-MBP (Protein Data Bank id: 10MP) used here is the wild-type mature protein without its 26-residue leader sequence. Protein expression was performed in E. coli in distinct pathways. Recent progress in computer simulations of the course of native structure formation now finds similar folding pathways for a number of small proteins (79–81).

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health (NIH) Research Grants R01 GM031847 (to S.W.E.), R01 GM055694 (to T.R.S.), a Structural Biology Predoctoral Training Grant GM08275 (to B.W.T.), and a National Science Foundation Research Grant MCB1020649 (to S.W.E.). Use of the Advanced Photon Source, an Office of Science User Facility, operated for the Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the DOE under Contract No. DE-AC02-06CH11357. This project was supported by grants from the National Center for Research Resources (2P41RR008630-18) and the National Institute of General Medical Sciences (9 P41 GM103622-18), both under NIH.


This work adds to the growing list of experimental demonstrations that proteins tend to fold through distinct intermediates in direct pathways. Recent progress in computer simulations of the course of native structure formation now finds similar folding pathways for a number of small proteins (79–81).