Dynamics of Protein Folding and Cofactor Binding Monitored by Single-Molecule Force Spectroscopy

Yi Cao and Hongbin Li*
Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT Many proteins in living cells require cofactors to carry out their biological functions. To reach their functional states, these proteins need to fold into their unique three-dimensional structures in the presence of their cofactors. Two processes, folding of the protein and binding of cofactors, intermingle with each other, making the direct elucidation of the folding mechanism of proteins in the presence of cofactors challenging. Here we use single-molecule atomic force microscopy to directly monitor the folding and cofactor binding dynamics of an engineered metal-binding protein G6-53 at the single-molecule level. Using the mechanical stability of different conformers of G6-53 as sensitive probes, we directly identified different G6-53 conformers (unfolded, apo- and Ni$^{2+}$-bound) populated along the folding pathway of G6-53 in the presence of its cofactor Ni$^{2+}$. By carrying out single-molecule atomic force microscopy refolding experiments, we monitored kinetic evolution processes of these different conformers. Our results suggested that the majority of G6-53 folds through a binding-after-folding mechanism, whereas a small fraction follows a binding-before-folding pathway. Our study opens an avenue to utilizing force spectroscopy techniques to probe the folding dynamics of proteins in the presence of cofactors at the single-molecule level, and we anticipated that this method can be used to study a wide variety of proteins requiring cofactors for their function.

INTRODUCTION

A large fraction of proteins in living cells (>30%) require cofactors, such as metal ions and small organic molecules, to carry out their biological functions, which range from electron transfer, metal ion transport and storage, to enzymatic catalysis (1–4). To reach their functional states, these proteins need to fold into their unique three-dimensional structures in the presence of cofactors (5,6). Therefore, the two processes, folding and binding of cofactors, intermingle with each other, raising questions about the interplay between these two processes and whether the binding of cofactors occurs before or after the protein has acquired its three-dimensional structure. Addressing this question is not only of fundamental importance for understanding the roles of cofactors in the functions of these proteins, but may also offer new insights into understanding of the protein-folding problem.

Pioneering work has demonstrated two possible pathways for the folding of proteins in the presence of cofactors (5,6). As shown in Fig. 1 A, in pathway I (binding-before-folding), cofactors can interact with and bind unfolded polypeptide in a specific manner to form an intermediate complex, which in turn significantly reduces the conformational entropy of the protein. Then, the cofactor-unfolded peptide complex serves as a nucleus for subsequent folding. Examples of proteins folding in this pathway include azurin (7–9) and Fe-S cluster proteins (10,11). In pathway II (binding after folding), unfolded polypeptides fold independently of cofactors to form apo-proteins (proteins without cofactors). Then, the cofactors bind the apo-protein to give rise to the holo-form (proteins with cofactors bound). Escherichia coli ribonuclease H1 (12) and staphylococcal nuclease A (13), which bind Mg$^{2+}$ and Ca$^{2+}$, respectively, are two examples that follow this pathway.

Despite the major progress in the study of protein folding in the presence of cofactors, it remains challenging to directly observe different protein conformers (unfolded, apo-, and holo-forms) along the folding pathway using ensemble measurements (5,6). Most of the experiments rely on ensemble kinetics data to deduce folding and cofactor binding mechanisms. Therefore, using single-molecule methodologies to directly probe different conformers during the folding process of proteins in the presence of cofactors is of great interest and importance. Over the last decade, development of single-molecule atomic force microscopy (AFM) techniques has made it possible to probe the mechanical unfolding/folding kinetics of proteins along a well-defined reaction coordinate at the single-molecule level (14–32). However, the application of single-molecule AFM techniques to elucidating the folding mechanism of proteins in the presence of cofactors remains to be demonstrated. Using an engineered metal binding protein G6-53 as a model system, here we employed single-molecule AFM to directly monitor the kinetics of protein folding and its cofactor binding of G6-53 in the presence of relatively high Ni$^{2+}$ concentrations. Our results revealed that single-molecule AFM can readily monitor the evolution of different protein conformers during the folding of proteins in the presence of its cofactor, and allows one to discern the folding...
mechanism of proteins in the presence of cofactors at the single-molecule level.

**EXPERIMENTAL SECTION**

**Protein engineering**

The (G6-53)₈ gene was engineered using standard molecular biology techniques as previously reported (33). The polyprotein was expressed in DH5α strain, purified by Co²⁺ affinity chromatography, and eluted in phosphate-buffered saline with 300 mM NaCl and 150 mM imidazole. A quantity of Biophysical Journal 101(8) 2009–2017

The (G6-53)₈ gene was engineered using standard molecular biology techniques as previously reported (33). The polyprotein was expressed in DH5α strain, purified by Co²⁺ affinity chromatography, and eluted in phosphate-buffered saline with 300 mM NaCl and 150 mM imidazole. A quantity of Tris-HCl buffer, and the desired amount of Ni²⁺ solution was added to wash the proteins extensively using Tris-HCl buffer, and the desired amount of Ni²⁺ bound fraction, Nbound. The number of domains that were still in their unfolded form, Nunfolded, was calculated using Nunfolded = Ntotal − Nbound.

It is worth noting that although the contour-length increment ΔLc in single-molecule AFM experiments can be accurately measured and related to the folded protein structure with single amino-acid resolution, the apparent contour length of the polyprotein—that is, the initial contour length measured for the first unfolding force peak for the polyprotein—can deviate from the real contour length of the polyprotein being stretched. Because the polyprotein molecule is attached to the surface through nonspecific interactions, there are three factors that may contribute to the uncertainty of the measured apparent initial contour length. First, the molecule may not be stretched perpendicular to the surface but at an angle. This makes the measured contour length shorter than the real one (34,35).

Second, the polyprotein molecule is typically attached to a soft protein layer on the substrate instead of a hard surface. The thickness and extensibility of this protein layer make the measured apparent initial contour length longer than the real value of the stretched polyprotein (the sum of the length of folded domains and the length of unstructured linker sequences). Third, the protein is attached to the surface via a soft protein thin layer, it is possible that one or two domains in the polyprotein that are close to the substrate surface may have denatured/unfolded due to their interactions with the surface. In addition, the protein domain being picked up by the AFM tip may suffer a similar problem and thus remains unfolded. These unfolded domains, which make the probability of observing full-length polyprotein very low, make the apparent initial contour length of the polyprotein longer than the real value. For these reasons, the measured ambient contour length of the polyprotein can exhibit large variations—sometimes much larger than the calculated length based on the observed number of unfolding force peaks.

**Data analysis**

The kinetic data were fitted using the following model,

\[ U \xrightarrow{k_f} F \xrightarrow{k_b} F(M), \]

where \( U \) is the unfolded form, \( F \) is the folded apo-form protein, \( F(M) \) is the folded holo-form protein, \( k_f \) is the folding rate for apo-protein, \( k_b \) is the binding rate for metal ion to apo-protein, and \( k_{ub} \) is the unbinding rate for metal ion from holo-form protein.

\[ \frac{d[U]}{dt} = -k_f[U], \quad \text{(1)} \]

\[ \frac{d[F]}{dt} = k_f[U] - k_b[M][F] + k_{ub}[F(M)], \quad \text{(2)} \]

\[ \frac{d[F(M)]}{dt} = k_b[M][F] - k_{ub}[F(M)]. \quad \text{(3)} \]
Here, $k_b$ and $k_{ab}$ relate to each other through the dissociation constant $K_d$ (which was determined as 98 $\mu$M) (36): $K_d = k_{ab}/k_b$. Solving the above equations, we obtained:

$$[U](t) = e^{-k_b t}, \quad (4)$$

$$[F](t) = \frac{k_{ab} - k_f}{k_f - k_b[M] - k_{ab}} e^{-k_f t} + \frac{k_f k_b[M]}{(k_f - k_b[M] - k_{ab})(k_b[M] + k_{ab})} e^{-(k_b[M] + k_{ab}) t} + \frac{k_b}{k_b[M] + k_{ab}}. \quad (5)$$

$$[F(M)](t) = \frac{k_b[M]}{k_f - k_b[M] - k_{ab}} e^{-k_f t} - \frac{k_f k_b[M]}{(k_f - k_b[M] - k_{ab})(k_b[M] + k_{ab})} e^{-(k_b[M] + k_{ab}) t} + \frac{k_b}{k_b[M] + k_{ab}}. \quad (6)$$

Equations 4–6 were used to fit the experimental data obtained in single-molecule force spectroscopy experiments.

**RESULTS**

**Probing the folding kinetics of G6-53 in the presence of Ni$^{2+}$ using single-molecule AFM**

The key to discerning the folding mechanism of proteins in the presence of cofactors is to identify different protein conformers along the folding pathway. During the folding process of a protein in the presence of its cofactor, four different conformers can exist: unfolded polypeptide ($U$), unfolded polypeptide and cofactor complex ($U(M)$), folded apo-form ($F$), and folded holo-form ($F(M)$). It is challenging to directly identify these different conformers in an unambiguous way to discern the folding mechanism. Single-molecule AFM studies have revealed that mechanical stability is an intrinsic property of proteins and depends on protein conformations and the binding of ligands (23,37–43). In these studies, the effects of cofactor binding on the mechanical properties of proteins in equilibrium are well documented. Here we propose to use mechanical stability as a sensitive probe to directly identify different conformers along the folding pathway and probe the kinetics of cofactor binding using single-molecule AFM techniques.

To demonstrate the feasibility of using single-molecule AFM to probe the folding dynamics of proteins in the presence of its cofactors, we used an engineered metal binding protein G6-53 as a model system (Fig. 1B). G6-53 is an engineered bi-histidine mutant (I6H, T53H) of the B1 IgG binding domain of protein G from *Streptococcus* (GB1) and binds divalent metal ions (such as Ni$^{2+}$) with high affinity (36,40). Although Ni$^{2+}$ ions are not cofactors for G6-53 in the strict sense, the binding of Ni$^{2+}$ to G6-53 can, in principle, capture the essential features found in the binding of cofactors to proteins. Thus, Ni$^{2+}$ serves the role of a “cofactor” for G6-53 in this folding study.

In our previous single-molecule AFM studies on polyprotein (G6-53)$_n$, we showed that stretching polyprotein (G6-53)$_n$ results in characteristic sawtoothlike force-extension curves, where sawtooth peaks correspond to the unfolding of individual G6-53 domains in a sequential fashion (36,40). The unfolding force, representing the mechanical stability of G6-53, depends on the specific states of G6-53. Apo-G6-53 ($F$ form), which refers to G6-53 without Ni$^{2+}$, unfolds at ~110 pN, whereas holo-G6-53 ($F(M)$ form), which refers to G6-53 with Ni$^{2+}$ bound, unfolds at a much elevated force of ~240 pN at a pulling velocity of 400 nm/s (36,40). In addition, unfolded G6-53 behaves as random coils in the absence or presence of Ni$^{2+}$, and the stretching of unfolded G6-53 results in featureless force-extension relationships typical of entropic springs. Thus, $U$ and $U(M)$ forms of G6-53 do not show any measurable mechanical stability within the resolution of our AFM (~20 pN). Such distinct mechanical stability provides sensitive probes allowing us to readily distinguish different conformers of G6-53 during the folding process of G6-53 in the presence of its cofactor Ni$^{2+}$. Therefore, by monitoring the concentration change of different conformers during folding, we should be able to directly discern the folding mechanism of G6-53 in the presence of Ni$^{2+}$.

**Direct observation of the evolution of different conformers during the folding of G6-53**

Based on the well-established double-pulse protocol (15,20,33,44–48), we used single-molecule AFM to monitor the folding kinetics of G6-53 in the presence of Ni$^{2+}$ (Fig. 2). In the first pulse, we stretched the polyprotein (G6-53)$_n$ between the cantilever tip and the substrate to unfold all the G6-53 domains in the polyprotein chain, allowing us to count the total number of G6-53 domains being stretched ($N_{total}$) and bring all the domains to the unfolded state. Then, the unfolded polypeptide chain was quickly relaxed to its original length to trigger the folding reaction. After waiting for a certain amount of time ($\Delta t$), the polypeptide was then stretched again in the second pulse, and the number of unfolding force peaks allowed one to count the number of G6-53 domains that had folded during the waiting time $\Delta t$. From the unfolding force of these unfolding events, one can directly determine the identity of the folded domains being $F$ or $F(M)$. G6-53 domains that failed to fold during time $\Delta t$ correspond to $U$ or $U(M)$.

If the folding of G6-53 in the presence of Ni$^{2+}$ follows pathway I (binding-before-folding), we should only observe...
second pulse. The dissociation constant of Ni\(^{2+}\) from
G6-53 is \(\sim 100 \mu M\) (36). Hence, \(\geq 90\%\) of G6-53 should
be in the Ni\(^{2+}\)-bound holo-form \((F(M)\) conformer) in the
presence of 2 mM Ni\(^{2+}\) after equilibrium (36). Indeed, in
the top force-extension curves of each pair, almost all
G6-53 domains unfolded at high forces \(\sim 240\) pN, indicating
that they are all Ni\(^{2+}\)-bound holo-G6-53. However, the
second pulse force-extension curves in the same pair
showed different unfolding patterns, depending on the wait-
ing time \(\Delta t\). For example, at a short waiting time of 0.005 s,
two out of six G6-53 domains folded into apo-form
\((F\) conformer) giving rise to unfolding events occurring at
\(\sim 110\) pN (as indicated by black dots), whereas the rest
four G6-53 domains remained unfolded. When the waiting
time increased to 0.1 s, four G6-53 domains managed to
fold giving rise to four unfolding events, in which three
occurred at \(\sim 110\) pN and one occurred at \(\sim 240\) pN, suggest-
ing that three G6-53 domains folded into the apo-form
\((F\) conformer) and one folded into Ni\(^{2+}\)-bound holo-form
\((F(M)\) conformer).

Upon further prolonging the waiting time, the number of
G6-53 domains that remained unfolded decreased and the
number of G6-53 domains that folded to the holo-form
increased. Eventually, after 2 s, all the G6-53 domains folded
into the holo-form (Fig. 2 D). By counting the number of
unfolding events occurring at \(\sim 110\) pN and \(\sim 240\) pN in
such experiments, the number of apo- and holo-forms of
G6-53, \(N_F\) and \(N_{FM}\), can be readily and unambiguously
determined during the folding process of G6-53 in the
presence of Ni\(^{2+}\). Accordingly, the number of unfolded G6-53
domains can be counted by subtracting the number of
unfolding events of G6-53 in the second trace from the total
number of G6-53 in the first trace, \(N_{\text{unfolded}} = N_{\text{total}} - N_F -
N_{FM}\). It is of note that here \(N_{\text{unfolded}}\) is the sum of \(N_U\) and
\(N_{UM}\), as we cannot distinguish these two conformers based
on their mechanical stability due to the limited force resolu-
tion of the AFM.

From Fig. 2, it is evident that apo-form G6-53 \((F\) con-
former) populated during the folding of G6-53 in the pres-
ence of 2 mM Ni\(^{2+}\). Because \(F\) conformer only populates
in pathway II (binding-after-folding), our results suggested
that the folding of G6-53 in the presence of Ni\(^{2+}\) follows
the binding-after-folding mechanism. To further confirm
this mechanism, we have monitored the evolution of
different conformers as a function of waiting time \(\Delta t\)
(Fig. 3 A).

Kinetic data corroborate that the major folding
mechanism for G6-53 in the presence of Ni\(^{2+}\)
is binding-after-folding

The evolution of different conformers along the folding
pathway will provide a direct test for the proposed folding
mechanism. Fig. 3 shows the evolution of the concentra-
tion of different conformers during the folding of G6-53 in the
order of Ni\(^{2+}\) to F is a second-order reaction and its reaction rate depends on both [F] and [Ni\(^{2+}\)]. Clearly, the concentration of all three conformers can be adequately fitted using the model derived from the pathway II mechanism, corroborating that pathway II is the major folding mechanism for G6-53 in the presence of Ni\(^{2+}\). From the fitting, we obtained the folding rate constant at zero force for apo-G6-53 of 10.5 ± 3.6 s\(^{-1}\), the binding rate constant of Ni\(^{2+}\) to apo-G6-53 of 1.84 ± 0.6 s\(^{-1}\)·mM\(^{-1}\), and the dissociation rate constant of Ni\(^{2+}\) from holo-G6-53 of 0.18 s\(^{-1}\) in the presence of 2 mM Ni\(^{2+}\).

Because the binding of Ni\(^{2+}\) to apo-G6-53 is a second-order reaction, the reaction rate for the binding step should be dependent upon the concentration of Ni\(^{2+}\). To further confirm the validity of the binding-after-folding mechanism, we performed single-molecule kinetic measurements in the presence of different concentrations of Ni\(^{2+}\) (Fig. 3, B and C). Indeed, the lower the Ni\(^{2+}\) concentration, the slower the binding reaction. However, the reaction rate of the folding step (U to F) remains similar to that in 2 mM Ni\(^{2+}\). As a consequence, the apo-form of G6-53 is more populated during the folding process. The maximum percentage of apo-G6-53 increases from ~60% to ~70%.

We also studied the folding of G6-53 in the presence of a lower Ni\(^{2+}\) concentration (0.2 mM). At this concentration, we found that the binding reaction rate is significantly slowed down due to the decrease of [Ni\(^{2+}\)] whereas the folding reaction remains largely unchanged, again corroborating that pathway II (binding after folding) is the major folding mechanism for G6-53. It is of note that 0.2 mM Ni\(^{2+}\) is not a saturating concentration for the binding of Ni\(^{2+}\) to G6-53, thus only ~60% of G6-53 will be in the holo-form at equilibrium.

### Binding of Ni\(^{2+}\) to G6-53 slows down the folding kinetics of G6-53

It has been shown that Ni\(^{2+}\) can bind to the unfolded histidine-bearing proteins with lower affinity (47,48), thus the conformer U(M) must exist in the unfolded ensemble of G6-53 in the presence of Ni\(^{2+}\). Therefore, despite that the kinetic data for the folding of G6-53 indicate that binding-after-folding is the major mechanism, we cannot completely rule out the possibility of a minor binding-before-folding pathway due to the inability to distinguish conformer U from U(M) using single-molecule AFM. We again used single-molecule AFM to investigate the effect of relative high concentration of Ni\(^{2+}\) on the folding kinetics of G6-53. Using the same double pulse protocol, we measured the folding kinetics of G6-53 at zero force in the absence of Ni\(^{2+}\) (Fig. 4). We found that the folding rate constant of G6-53 is ~315 ± 32 s\(^{-1}\), which is slightly slower than that of wild-type GB1 (33). This folding rate constant (k\(_{U,F}\)) is much faster (~30-fold) than the folding rate of apo-G6-53 in the presence of high concentration of
Ni\textsuperscript{2+} (\(k_f\) for the \(U\rightarrow F\) transition in the binding-after-folding pathway) (\(10.5 \pm 3.6\) s\(^{-1}\)), suggesting that the presence of Ni\textsuperscript{2+} has a significant influence on the folding dynamics of the apo-form G6-53. Notably, the presence of Ni\textsuperscript{2+} does not slow down the folding of wild-type GB1, corroborating that the reduction of folding rate of G6-53 in the presence of Ni\textsuperscript{2+} is indeed due to the binding of Ni\textsuperscript{2+} to histidine residues.

However, in the majority of folding trajectories of G6-53 in the presence of Ni\textsuperscript{2+}, the folding of the apo-form (conformer \(F\)) of G6-53 is predominantly the obligatory step for the folding of holo-G6-53 (\(F(M)\)), suggesting that the formation of \(U(M)\) does not lead to the folding of \(F(M)\) and the pathway \(U(M)\rightarrow F(M)\) in the binding-before-folding mechanism is not favorable, possibly due to the unfavorable loop formation (entropic barrier) imposed by the binding of Ni\textsuperscript{2+} to the two histidine residues in G6-53. Thus, the dissociation of \(U(M)\) to \(U\) may be required for the folding of conformers \(F\) and \(F(M)\).

**Observation of a minor binding-before-folding pathway in single-molecule AFM experiments**

Although the majority of \(U(M)\) does not lead to the folding of holo-G6-53 (\(F(M)\)), misfolding events of G6-53 provided valuable information about the pathway of binding-before-folding.

In force spectroscopy experiments, polyprotein engineering is a general approach to provide a fingerprint for identifying single-molecule stretching events. If the metal ions bind to the two histidines from the same domain, if the metal ions bind to the histidines from neighboring G6-53 domains, such a misaligned intermediate structure may lead to the formation of a misfolded dimeric G6-53, in which two neighboring G6-53 domains fold together to form a superfold, possibly through a domain-swapping mechanism. Such misfolding events have been reported for different polyproteins (49,50). Because a misaligned Ni\textsuperscript{2+}-unfolded peptide complex facilitates the formation of such misfolded superfolds, the observation of such misfolding events will provide a glimpse of the binding-before-folding pathway.

In the absence of Ni\textsuperscript{2+}, G6-53 folds with high fidelity just like wild-type GB1 and we did not observe any misfolding events out of >500 refolding curves (33). However, in the presence of Ni\textsuperscript{2+}, we have detected the formation of such a misfolded G6-53 superfold. Two examples are shown in Fig. 5: during the refolding of (G6-53)_8, two neighboring G6-53 domains can misfold, leading to the observation of a skip in the force-extension curves. For example, there are ten G6-53 unfolding events with an identical contour-length increment \(\Delta L_c\) of ~18 nm in curve 1 (Fig. 5). However, in the second force-extension curve, only nine unfolding events were observed. One of the unfolding events has a \(\Delta L_c\) of ~40 nm, giving rise to the appearance

**FIGURE 4** Folding kinetics of G6-53 in the absence of Ni\textsuperscript{2+}. The folding of G6-53 in the absence of Ni\textsuperscript{2+} ions was measured using standard double-pulse protocol. (Shaded line) Exponential fitting to the experimental data using first-order kinetics. The folding rate constant of G6-53 in the absence of Ni\textsuperscript{2+} is estimated to be 315 ± 32 s\(^{-1}\).

**FIGURE 5** Small fraction of G6-53 follows the binding-before-folding mechanism. In the presence of Ni\textsuperscript{2+}, a small number of G6-53 misfold into dimeric superfolds, giving rise to the appearance of skips in the force-extension curves (highlighted in dark gray). The contour-length increment \(\Delta L_c\) of such a misfolded G6-53 is ~42.5 nm, slightly bigger than \(2 \times \Delta L_c\) (~38.3 nm) of G6-53. The superposition of the two curves clearly showed the difference in \(\Delta L_c\)-misfold and \(2 \times \Delta L_c\). Such a misfolded superfold likely resulted from the binding of Ni\textsuperscript{2+} to the two histidines from two neighboring G6-53 domains. (Top panel) Schematics of the misfolding of two neighboring G6-53 domains into a misfolded superfold. It is of note that polyprotein (G6-53)_8 has two cysteine residues at its C-terminus. The oxidation of cysteine residues can lead to the formation of dimeric (G6-53)_8. (Curves 1 and 2) Stretching of such a dimeric (G6-53)_8 molecule.
of a skip. It is of note that, consistent with the picture that two G6-53 domains coalesced into one superfold, $\Delta Lc(skip)$ is slightly bigger than $2 \times \Delta Lc$ of G6-53 (see superposition of curves 1 and 2 as well as curves 3 and 4). The difference between $\Delta Lc(skip)$ and $2\Delta Lc$ is $\sim 4$ nm. For such a superfold misfolding scenario, the contour length of the unfolded and fully extended superfold equals to the sum of the contour length of two unfolded and fully extended G6-53 domains ($2 \times Lc(unfolded)$) and the length of the linker region ($L_{\text{linker}}$) between these two domains; and the folded length ($Lc(folded)$) remains the same as G6-53. Therefore, $\Delta Lc(skip)$ should be $2 \times Lc(unfolded)+ L_{\text{linker}}- Lc(folded)$, which equals to $2 \times \Delta Lc+ L_{\text{linker}}- Lc(folded)$. Clearly, the difference between $\Delta Lc(skip)$ and $2\Delta Lc$ is the sum of the length of the linker sequence between the two G6-53 domains and the length of the force-bearing motif of one G6-53 domain (49).

Such misfolding events are rare and occur in $\sim 1\%$ of the force-extension curves of G6-53 in the presence of Ni$^{2+}$ (23 out of 2246). In fact, after mechanical unfolding of the skips, the two G6-53 domains were found to always resume their normal folding behavior and in the subsequent force-extension curves, the skip disappeared. It is worth noting that Ni$^{2+}$ cannot bind nonspecifically to other residues in G6-53 except for histidine. In a control experiment, in the presence of 4 mM Ni$^{2+}$, wild-type GB1 does not show any misfolding events during $>1000$ refolding cycles. Hence, the observation of such misfolded superfold indicates that Ni$^{2+}$ does bind unfolded G6-53 ($U$ form), leading to the formation of the unfolded G6-53/Ni$^{2+}$ complex ($U(M)$) form. This result indicates that the formation of misfolded superfold follows the binding-before-folding pathway.

**DISCUSSION**

Cofactors play important functional and structural roles for many proteins. Metalloproteins are a typical class of proteins requiring cofactors. Some of these proteins require cofactors for proper folding, whereas others can only bind cofactors after they acquire their native structures and the binding sites form. Misfolding of such proteins will disrupt their biological functions and lead to different diseases. For example, the homeostasis of metal ions directly affects the folding and aggregation of $\beta\beta$ protein that is associated with Alzheimer’s (51). Thus, understanding the folding mechanism and the intricate/delicate interplay between folding and cofactor binding is of great importance from the perspective of fundamental understanding of protein folding as well as the biomedical relevance of such proteins.

Folding and binding processes are intermingled during the folding of proteins in the presence of their cofactors. Both processes are analogous in that they both attempt to maximize the intra- or intermolecular interactions along a funnel-shaped energy landscape (52). To elucidate the folding mechanism of such proteins requires one to resolve all the conformers involved in the folding and their dynamic evolution. Here, using the distinct mechanical stability of different conformers of the engineered metal-binding protein G6-53 as sensitive probes, we employed single-molecule AFM to identify different conformers populated during the folding process and monitored their time-evolution course along their folding pathways, which are otherwise difficult to measure in ensemble measurements.

Our results demonstrated that the majority of G6-53 folds via a binding-after-folding mechanism, whereas a small portion (<1%) folds via a binding-before-folding mechanism. Extensive simulations and experiments suggested that for wild-type GB1, the second $\beta$-hairpin forms in the folding transition state and serves as nuclei for the folding of the rest part (53,54). However, the engineered metal-ion-binding site is located on the first and fourth $\beta$-strands. The structure of this region is only formed after GB1 is completely folded. Therefore, it is not surprising that the binding-after-folding is the dominant mechanism for the folding of G6-53 in the presence of Ni$^{2+}$. However, rare events following the binding-before-folding pathway do exist for G6-53. In addition, the presence of high concentration of Ni$^{2+}$ slows down the folding kinetics (U-F) of apo G6-53. These results suggest that two competing pathways are coexisting. However, the pathway $U-U(M)$ is largely not productive, as the conformer $F$ serves as an obligatory intermediate for the folding of $F(M)$ (holo G6-53). In contrast, the binding-after-binding pathway is much more efficient for reaching the holo-form of G6-53. Thus, the conformer $U(M)$ can be considered as an off-pathway intermediate state for the folding of $F(M)$.

Although G6-53 is an engineered simple metal-binding protein, our results revealed the complexity of its folding process in the presence of Ni$^{2+}$. It can be anticipated that for naturally occurring proteins requiring cofactors, such as metalloproteins, the folding mechanism can be even more complex. However, the utility of single-molecule AFM in studying the folding of proteins in the presence of cofactors, as demonstrated here, can be readily applied to investigating the folding mechanism of a wide range of proteins requiring the presence of cofactors. In particular, we anticipate that this new, to our knowledge, method will enable the elucidation of the folding mechanism of many biologically important metalloproteins, such as zinc-finger proteins and rubredoxins.

One limitation of the method here presented is the limited force resolution of single-molecule AFM ($>10$ pN). Due to this limitation, conformer $U(M)$ cannot be directly detected. Improving the force resolution of AFM will be key for the further application of single-molecule AFM in studying protein folding in the presence of cofactors. For example, if the force resolution is high enough, conformers $U$ and $U(M)$ can easily differentiated by their distinct mechanical behaviors: stretching $U$ will lead to a featureless wormlike
chain behavior, whereas stretching $U(M)$ will possibly lead to sawtoothlike force-extension curves where the sawtooth peaks correspond to the mechanical rupture of His-Ni$^{2+}$-His complex and should occur at very low forces (<20 pN).

Moreover, in the polyprotein (G6-53)$_k$, the His-Ni$^{2+}$-His complex formed between histidines of different G6-53 domains will lead to different contour-length increments, which can be easily differentiated from that formed between histidines of a single G6-53 domain. Such possibilities will make a force spectroscopy-based method very attractive. In the same vein, the application of optical tweezers has been used to monitor the folding of riboswitch aptamers with a subpiconewton resolution in force. Another possible limitation lies in the limited temporal resolution. For example, in the binding-after-folding mechanism, if the binding rate is much faster than the folding rate and the temporal resolution is not high enough, the apo-form will be difficult to capture—leading to incorrect interpretation of the folding mechanism. Therefore, improving the temporal resolution is another key for resolving all the conformers during the folding process.

We anticipate that with improved single-molecule techniques, force spectroscopy-based assays can be used widely to characterize the folding and binding mechanism of a wide variety of proteins that require cofactors to function.

We thank anonymous reviewers for the helpful suggestions.

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Canada Research Chairs Program, and Canada Foundation for Innovation.

REFERENCES


