Monitoring cotranslational protein folding in mammalian cells at codon resolution

Yan Han[a], Alexandre David[b], Botao Liu[c], Javier G. Magadán[b], Jack R. Bennink[b], Jonathan W. Yewdell[b], and Shu-Bing Qian[a,c,1]

[a]Division of Nutritional Sciences, Cornell University, Ithaca, NY 14853; [b]Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; and [c]Graduate Field of Genetics and Development, Cornell University, Ithaca, NY 14853

Edited by Arthur L. Horwich, Yale University School of Medicine, New Haven, CT, and approved June 20, 2012 (received for review May 14, 2012)

How the ribosome-bound nascent chain folds to assume its functional tertiary structure remains a central puzzle in biology. In contrast to refolding of a denatured protein, cotranslational folding is complicated by the vectorial nature of nascent chains, the frequent ribosome pausing, and the cellular crowdedness. Here, we present a strategy called folding-associated cotranslational sequencing that enables monitoring of the folding competency of nascent chains during elongation at codon resolution. By using an engineered multidomain fusion protein, we demonstrate an efficient cotranslational folding immediately after the emergence of the full domain sequence. We also apply folding-associated cotranslational sequencing to track cotranslational folding of hemagglutinin in influenza A virus-infected cells. In contrast to sequential formation of distinct epitopes, the receptor binding domain of hemagglutinin follows a global folding route by displaying two epitopes simultaneously when the full sequence is available. Our results provide direct evidence of domain-wise global folding that occurs cotranslationally in mammalian cells.

deep sequencing | ribosome profiling | protein quality

It is currently believed that protein folding generally begins during translation on the ribosome (1, 2). In mammalian cells, the rate of protein synthesis is approximately five residues per second, whereas folding is typically occurring on the microsecond scale (3, 4). Thus, many details of cotranslational folding pathway remain elusive. For example, what types of the structures and/or intermediates are formed in the nascent chain during cotranslational folding? How early in translation are these structures formed? In contrast to the in vitro refolding of full-length polypeptides, the cotranslational folding of emerging polypeptides is influenced by their sequential exposure from the ribosome exit tunnel to the cytosol (5). Cotranslational folding is further complicated by frequent ribosome pausing (6), as well as interactions with cellular binding partners (7).

Traditional methods of detecting cotranslational folding rely on monitoring of the enzymatic activity of model proteins synthesized in vitro (1, 2). These assays are impractical when applied to cells under physiological conditions. A few in vivo experiments supporting cotranslational folding were based on pulse-chase metabolic labeling coupled with folding-dependent cleavage analysis (8). A limitation of this approach is low resolution. Fluorescence-based techniques, such as FRET, allow detection of cotranslational folding and interactions of nascent chains with high resolution (9). However, FRET measurements require incorporation of modified amino acids and are limited to cell-free systems. None of these methods can be used for simultaneous monitoring of cotranslational folding of nascent chains with varied lengths in vivo. We developed an approach called folding-associated cotranslational sequencing (FactSeq) that overcomes many of these deficiencies. By harnessing the power of the ribosome profiling technique (10), FactSeq allows us to dissect at what point during translation the nascent chain acquires a specific conformation.

Results and Discussion

During translation elongation, the positions of ribosomes on a given mRNA, and hence the length of the synthesized polypeptide chain, can be determined by deep sequencing of the ribosome-protected mRNA fragments (RPFs) (10, 11). FactSeq is based on enriching ribosomes bearing nascent chains with recognizable structural features, followed by sequencing of the associated RPFs. Direct comparison of RPF distribution before and after ribosome enrichment provides sequence-specific structural information associated with the nascent chain. To pilot this technique, we generated a HEK293 cell line stably expressing the multidomain fusion protein Flag-FRB-GFP, in which the well-characterized FKBP12-rapamycin binding domain (FRB) was fused to the NH_2 terminus of GFP (Fig. 1). After collecting the ribosome fractions from the whole-cell lysates by using sucrose gradient sedimentation, we converted polysomes to single ribosomes by RNase I treatment to digest mRNAs not protected by the ribosome. We first enriched ribosomes bearing the NH_2-terminal Flag-tagged nascent chain by immunoprecipitation (IP) using anti-Flag mAb-coated beads. After extracting Flag tag-associated RPFs as well as total RPFs from the same sample, we constructed a cDNA library suitable for Illumina high-throughput sequencing (Fig. 1).

The sequencing results of RPFs obtained with or without Flag IP were of similar quality (Fig. S1). As expected, the majority of RPFs were approximately 30 nt in length. The 5′ end positions of RPF showed a strong 3-nt periodicity, confirming that the RPF accurately captures the ribosome movement along mRNAs. By using RPFs derived from the entire transcriptome, we built a ribosome density map on the Flag-FRB-GFP transcript (Fig. 2*A*). Consistent with the nonuniform rates of translation elongation, the transcript was punctuated with multiple ribosome pausing sites with a skewed number of reads located at the start codon region. Notably, the linker region between FRB and GFP showed the least RPF reads, possibly because of our selection of commonly used codons in creating the construct. Over four independent RPF deep-sequencing replicates, the ribosome distribution pattern on individual transcripts was highly reproducible (Fig. S2).

As the NH_2-terminal Flag tag is present at the start of the nascent chain, the Flag mAb-associated RPFs should capture nearly all the ribosome footprints during elongation. Alignment of RPF reads on Flag-FRB-GFP transcript before and after Flag IP revealed a nearly identical pattern of ribosome density except the first 50 codon region (Fig. 2*A*, *Lower*). As ~40 aa of the

Author contributions: Y.H. and S.-B.Q. designed research; Y.H., A.D., B.L., J.G.M., and S.-B.Q. performed research; J.R.B. and J.W.Y. contributed new reagents/analytic tools; Y.H., B.L., J.W.Y., and S.-B.Q. analyzed data; and S.-B.Q. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: sq38@cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1208138109/-/DCSupplemental.

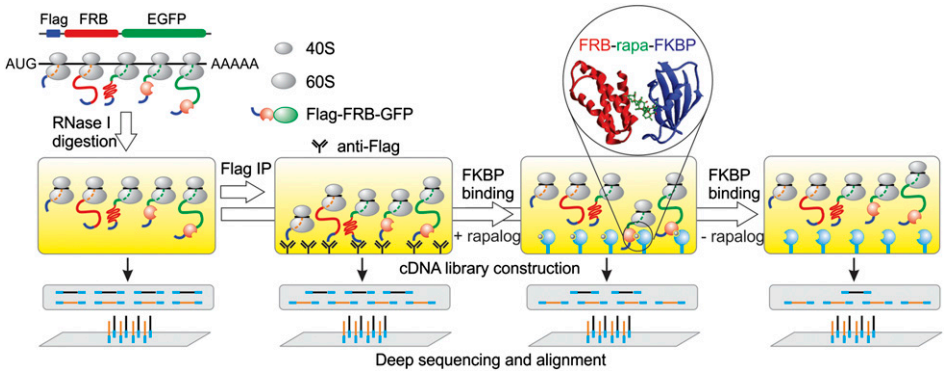

Fig. 1. Schematic for FactSeq approach. The polysomes from HEK293/Flag-FRB-GFP are converted into monosome by RNase I treatment, followed by IP using anti-Flag or recombinant FKBP in the presence or absence of rapalog. The RPFs are extracted and mixed with spike-in control before cDNA library construction. The deep sequencing results of RPFs are analyzed by transcriptome mapping. (*Inset*) Circle depicts structure of FKBP (blue, PDB 1A7X) and FRB (red, PDB 1AUE) dimerization in the presence of rapamycin (green).

growing polypeptide chain are buried within the ribosome exit tunnel, this corresponds well to the minimal length of 10 aa required for antibody binding (i.e., the full length of Flag tag). Notably, there was little reduction of Flag IP-associated RPF reads relative to total RPFs along the remaining part of the transcript, indicating that the Flag tag remains intact during the synthesis of Flag-FRB-GFP. This argues that cotranslational degradation is minimal in this multidomain fusion protein. Thus, the FactSeq approach allows tracking of the behavior of nascent chains with high accuracy and sensitivity.

We next extended the FactSeq approach to evaluate cotranslational folding of the FRB domain before the complete synthesis of GFP. To probe the folding status of FRB domain, we took advantage of its binding partner FKBP (FK506 binding protein). The dimerization of FKBP and FRB relies on their 3D structures and the presence of rapamycin or rapalog (12, 13) (Fig. 1). As expected, recombinant FKBP synthesized in *Escherichia coli* specifically pull down the Flag-FRB-GFP fusion protein in a rapalog-dependent manner (Fig. S3). Thus, FKBP-rapalog can be used as a bait to probe the folding status of FRB before the full-length fusion protein is released from the ribosome. Consistent with the high specificity of rapalog-mediated FRB–FKBP interaction, very few RPF reads were recovered in the absence of rapalog (Fig. 2*B*). By contrast, adding rapalog selectively restored a significant number of RPFs starting at the 150 codon position of the Flag-FRB-GFP transcript. Given 10 aa from the Flag tag and 40 aa buried in the ribosome tunnel, the appearance of RPF reads after 150 codon position corresponds to the minimal length of 100 aa at which the FRB domain starts to create the rapalog binding site and associate with FKBP. Thus, FRB domain is able to fold when the corresponding amino acid sequence immediately emerges from the ribosome.

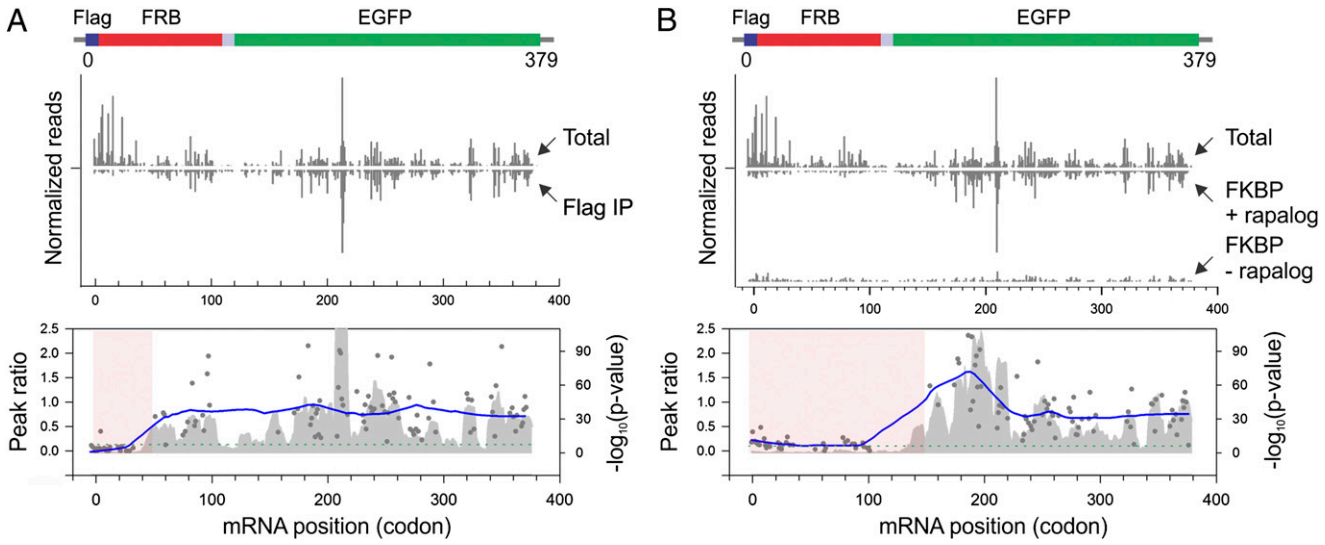

Fig. 2. Monitoring cotranslational behavior of Flag-FRB-GFP polypeptide in mammalian cells. (*A*) Comparison of RPF distribution on Flag-FRB-GFP transcript before and after anti-Flag IP. Both the total and Flag IP-associated RPFs are aligned based on the sequence position of Flag-FRB-GFP. *Lower*: Pattern analysis using single codon peak ratio (dot plot) and significance (*P* value) of RPF density vs. background in a 10-codon sliding window (field plot). The line plot represents the LOESS-smoothed trend line for single codon peak ratio (sampling proportion, 0.2). The colored areas represent regions of nascent chains that inaccessible to anti-Flag antibody. Cutoff line was set at $P = 0.001$ (green dashed line). (*B*) Comparison of RPF distribution on Flag-FRB-GFP transcript before and after FKBP affinity purification. (*Upper*) Alignment of RPFs associated with total and FKBP binding in the presence or absence of rapalog. *Lower*: Pattern analysis using single codon peak ratio (dot plot) and significance (*P* value) of RPF density vs. background in a 10-codon sliding window (field plot). The line plot represents the LOESS-smoothed trend line for single codon peak ratio (sampling proportion, 0.2). The colored areas represent regions of nascent chains inaccessible to FKBP. Cutoff line was set at $P = 0.001$ (green dashed line).

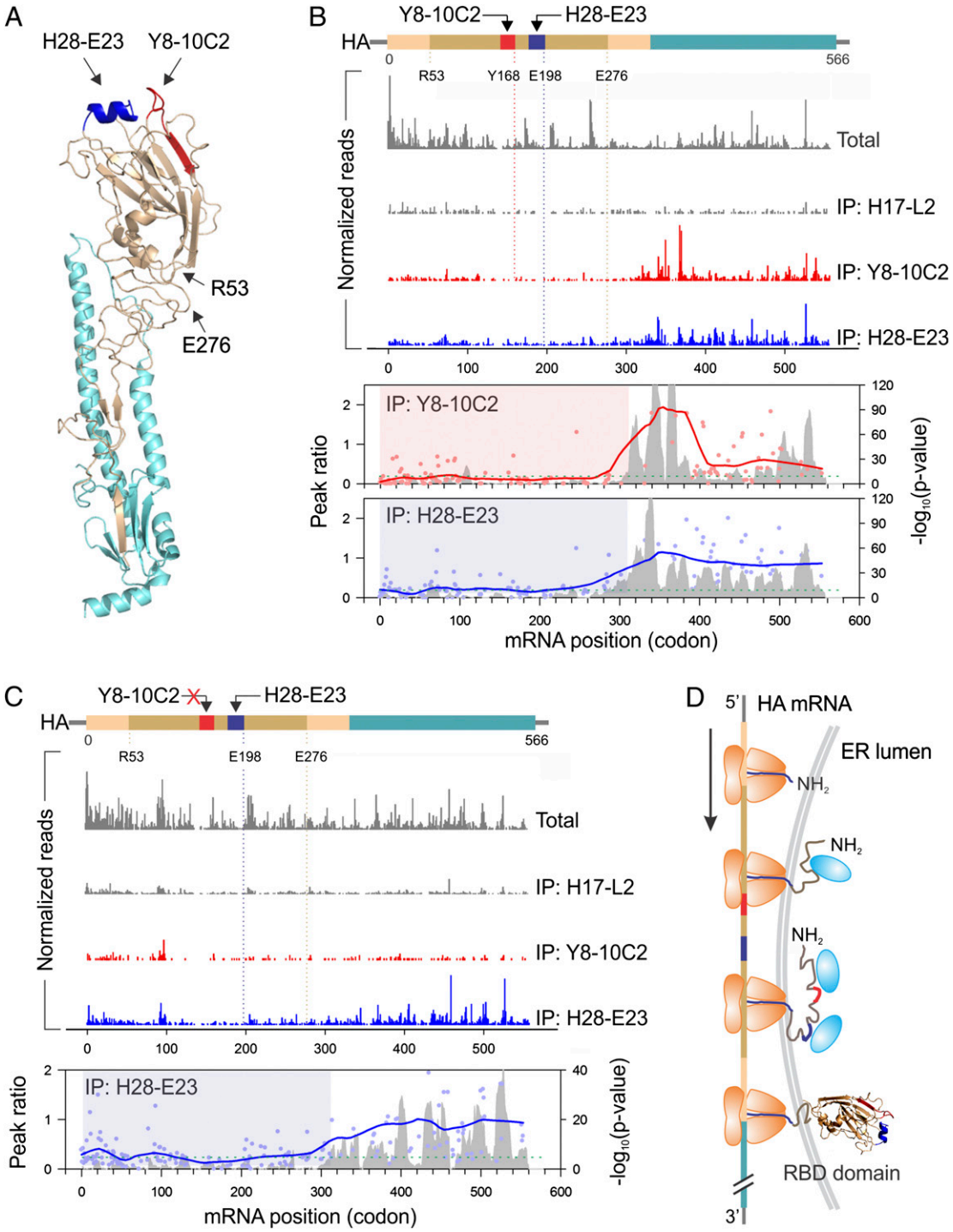

Fig. 3. Monitoring cotranslational folding of HA in influenza A-infected cells. (*A*) Epitope sites on the RBD domain of HA. The structure of monomeric HA is shown as a ball-and-stick model with HA1 in light brown and HA2 in light blue (PDB 1RU7). The Sa site recognized by Y8-10C2 is shown in red, whereas the Sb site by H28-E23 is dark blue. The RBD domain is identified from R53 to E276. (*B*) Comparison of RPF distribution on HA transcript derived from PR8 H1N1 influenza A before and after IP. Polysome fractions were prepared from influenza A-infected HeLa cells. IP was performed by using a panel of antibodies, followed by deep sequence of RPFs. *Lower*: Pattern analysis after IP with Y8-10C2 (red) and H28-E23 (blue) using single codon peak ratio (dot plot) and significance (*P* value) of RPF density vs. background in a 10-codon sliding window (field plot). The line plot represents the LOESS-smoothed trend line for single codon peak ratio (sampling proportion, 0.2). The colored areas represent regions of nascent chains inaccessible to antibodies. Cutoff line was set at $P = 0.001$ (green dashed line). (*C*) Comparison of RPF distribution on HA transcript derived from CV1 mutant that escapes the Y8-10C2 recognition. (*Lower*) Pattern analysis after IP with H28-E23 (blue) using single codon peak ratio (dot plot) and significance (*P* value) of RPF density vs. background in a 10-codon sliding window (field plot). The line plot represents the LOESS smoothed trend line for single codon peak ratio (sampling proportion, 0.2). Cutoff line was set at $P = 0.001$ (green dashed line). (*D*) A model of domain-wise global folding of the HA nascent chain attached to the ribosome. Rapid cotranslational folding occurs only after the full RBD domain sequence is available. The blue ovals represent possible binding partners of nascent chains, such as molecular chaperones.

Intriguingly, there was a significant reduction of FKBP-associated RPF reads after 200 codon position of the Flag-FRB-GFP transcript relative to total RPFs ($P = 6.256 \times 10^{-5}$; Fig. 2*B*, *Lower*). We cannot attribute this to cotranslational degradation, because the NH_2-terminal Flag tag was continuously present (Fig. 2*A*). Rather, the appearance of GFP polypeptide could partially prevent folded FRB from interacting with FKBP, possibly by associating molecular chaperones or other binding partners (5).

Having successfully monitored the cotranslational folding of FRB domain, we next applied FactSeq to evaluate the cotranslational folding process of influenza A hemagglutinin (HA). Influenza presents a serious public health challenge, and HA is a prime candidate for vaccine design and drug development (14). HA is a type I transmembrane glycoprotein with multiple folding domains. The protein is cotranslationally glycosylated in the endoplasmic reticulum. After posttransaltional trimerization, the HA trimer then traffics to the cell surface. The crystal structure of HA reveals a globular ectodomain sitting atop an extended stalk (Fig. 3*A*). The HA mediates attachment to cells via its receptor binding site in the ectodomain [i.e., receptor-binding domain (RBD)], which also contains most of the epitopes for neutralizing antibody recognition (15). All known neutralizing anti-HA mAbs recognize conformational determinants, i.e., their epitopes are formed by folding of the primary sequences. For instance, the Y8-10C2 mAb recognizes an epitope on the HA comprising residues from E158 to Y168 [Puerto Rico/8/34 (PR8)], whereas the H28-E23 epitope is principally formed by residues from N187 to E198 (16) (Fig. 3*A*). As these mAbs are capable of binding HA monomers based on local folding of their respective domains, we used them to probe cotranslational formation of distinct epitopes during the biosynthesis of HA. As a negative control, we used the H17-L2 mAb, whose binding is dependent on trimerization, as its epitope is formed by residues on each side of the trimer interface (17).

We purified ribosomes 5 h after viral infection of HeLa cells with PR8 (Fig. S4). The ribosome fractions were then immunoprecipitated using each of the mAbs followed by deep sequencing of RPFs. Total RPF reads aligned to the full length of HA transcript exhibited typical pausing sites (Fig. 3*B*). Trimer-specific H17-L2 set the background level of reads. In contrast, both Y8-10C2 and H28-E23 recovered a large number of RPFs over background. Thus, FactSeq could be used to investigate the cotranslational folding of endoplasmic reticulum proteins. Intriguingly, Y8-10C2 exhibited an almost 110-codon delay in the appearance of RPFs relative to the emergence of its epitope ending at residue Y168 from the ribosome exit tunnel (Fig. 3*B*). H28-E23 also had a lag of approximately 70 codons after the residue E198 emerged from the ribosome. Notably, both mAbs initiated simultaneous binding with the emergence of the entire globular RBD domain, which ends at residue E276 (Fig. 3*A*). These results indicate that Y8-10C2 and H28-E23 epitopes are not formed sequentially. Although we could not exclude the possibility of mAb binding-induced nascent chain folding, our previous study suggests that it is unlikely for such an event to contribute to the specific RPFs revealed by FactSeq (15). First, refolding of denatured HA by antibody binding occurs over days, not within 1 h. Second, mAb binding-induced HA folding, if rapid, would lead to continuous enrichment of mAb-associated RPFs along with elongation. However, similar to cotranslational folding of the FRB domain, the HA exhibited reduced accessibility of Y8-10C2 after its initial epitope formation. The discontinuous antibody accessibility to the continuously elongated nascent chain could represent a previously unrecognized feature of cotranslational folding pathway. Thus, the coappearance of distinct epitopes after the full ectodomain is available is consistent with a model of global folding pathway that occurs rapidly when the RBD sequence has emerged from the ribosome (Fig. 3*D*).

To validate the cotranslational folding propensity of HA revealed by FactSeq, we performed [^{35}S]methionine pulse-chase of influenza A virus-infected cells coupled with IP. Pulse-chase was performed at 20 °C to slow down elongation and trimerization of HA polypeptides. Y8-10C2 recovers HA fragments in addition to the full-length polypeptide. In contrast, HA trimer-specific antibody H17-L2 recovers only full-length HA. Y8-10C2–binding fragments completely resolve during the chase into full-length HA, demonstrating a precursor–product relationship and indicative of the ability of HA fragments to generate the Y8-10C2 conformational epitope (Fig. S5*A*). Remarkably, the pattern of HA fragments matches well with that revealed by FactSeq (Fig. S5*B*). For instance, the smallest fragment from each assay was approximately 30 kDa. Despite relatively low resolution, pulse labeling analysis confirms the discontinuous nature of mAb binding during elongation. These data support the conclusion that the high resolution pattern offered by FactSeq represents the true behavior of nascent chain synthesis.

We next asked how mutation of an epitope affects the default folding pathway of the HA globular domain. To this end, we chose a PR8 escape mutant CV1, whose single K165E substitution reduces Y8-10C2 avidity more than 100-fold (18) (Fig. S6). Only background levels of RPF reads were recovered by Y8-10C2 from CV1, providing an important confirmation of the specificity of the FactSeq method (Fig. 3*C*). Interestingly, whereas CV1 maintained the similar codon lag as WT HA in the generation of H28-E23 epitope, the full extent of epitope formation was delayed, indicating that the single K165E substitution alters the kinetics of the cotranslational folding of HA (Fig. 3 *C* and *D*).

Conclusion

It is widely believed that cotranslational folding is a universal feature of newly synthesized polypeptides (1, 2). However, monitoring this dynamic process is challenging, in particular inside mammalian cells. By harnessing the power of the ribosome profiling technique with the folding-sensitive affinity reagents, FactSeq provides a unique view of the folding competency of the nascent chain during its elongation. The acquisition of the functional FRB and the HA RBD immediately after their sequences emerge from the ribosome exit tunnel strongly favors a domain-wise global folding pathway. Despite the limitations of FactSeq in providing real-time kinetics of folding pathway, the snapshot taken by FactSeq consists of continuous frames of ribosomes with varied length of nascent chains. FactSeq also requires folding-sensitive affinity reagents to capture specific folding status of ribosome-attached nascent chains. With the increasing number of available conformation-specific antibodies or binding factors for various gene products (19), FactSeq is readily applicable to endogenous proteins thanks to the high throughput of deep sequencing that covers the entire transcriptome. In addition, taking advantage of the intrabodies that are designed to be expressed intracellularly (20, 21), FactSeq has the potential to capture cotranslational folding in live cells before cell lysis. Beyond cotranslating folding, the prototype of FactSeq can also be applied to other cotranslational events. For instance, by using an NH_2-terminal tag, the same concept can be adapted to investigate cotranslational degradation by comparing the loss, rather than the gain, of RPFs. Finally, FactSeq can also be expanded to study cotranslational chaperone interaction with nascent chains (7, 22). The applicability of FactSeq is not limited to studying cotranslational events. The basic principle can be used to design in vivo folding reporter to investigate cellular factors influencing cotranslational folding. We envision that this

approach will provide novel insights into protein triage decisions under physiological as well as pathological conditions.

Materials and Methods

Cells and Reagents. HEK293 cells stably expressing Flag-FRB-GFP were maintained in DMEM with 10% (vol/vol) FBS. Rapalog AP21967 was provided by Ariad. Anti-Flag and anti-HA antibodies were purchased from Sigma, and protein A/G beads from Santa Cruz. TRIzol reagent was purchased from Invitrogen.

Influenza A Infection. HeLa cells were infected with Influenza A/PR8 strain at a multiplicity of 20 pfu/cell in AIM medium, pH 6.6. After adsorption at 37 °C for 1 h, infected monolayers were overlaid with DMEM containing 7.5% FBS, and incubated for an additional 5 h.

Ribosome Profiling. Sucrose solutions were prepared in polysome gradient buffer [10 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 5 mM DTT, and 20 U/mL SUPERase_In (Ambion)]. Sucrose density gradients [15–45% (wt/vol)] were freshly made in SW41 ultracentrifuge tubes (Fisher) using a BioComp Gradient Master (BioComp) according to the manufacturer's instructions. HEK293/Flag-FRB-GFP cells were plated to four 10-cm dishes before ribosome profiling. Cells were first treated with cycloheximide (100 μg/mL) for 3 min at 37 °C to freeze the translating ribosomes, followed by ice-cold PBS solution wash. Cells were then harvested by ice-cold polysome lysis buffer [10 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 5 mM DTT, 20 U/mL SUPERase_In, and 2% (vol/vol) Triton X-100]. After centrifugation at 4 °C and 10,000 × *g* for 10 min, approximately 650 μL supernatant was loaded onto sucrose gradients, followed by centrifugation for 100 min at 38,000 rpm, 4 °C, in an SW41 rotor. Separated samples were fractionated at 0.375 mL/min by using a fractionation system (Isco) that continually monitored OD_{254} values. Fractions were collected into tubes at 1-min intervals.

Ribosome Purification. To convert the polysome into monosome, *E. coli* RNase I (Ambion) was added into the pooled polysome samples (750 U per 100 A260 units) and incubated at 4 °C for 1 h. Preclearance was conducted by incubating the ribosome samples with 30 μL protein A/G beads coated with 4% BSA for 1 h at room temperature. For IP using mAbs, 30 μL protein A/G beads were first incubated with 5 μg mAbs for 1 h at room temperature followed by blocking with 4% BSA for 1 h. The mAb-coated beads were then incubated with the precleared ribosome samples at 4 °C for 1 h, followed by washing with polysome lysis buffer for three times. For FKBP binding assay, 20 μg recombinant HA-FKBP proteins purified from *E. coli* (BL21) were first immobilized on protein A/G beads using anti-HA antibody. After blocking with 4% BSA for 1 h, the beads were then incubated with the precleared ribosome samples at 4 °C for 1 h in the absence or presence of 1 μM rapalog. After washing with polysome lysis buffer three times, total RNA extraction was performed by using TRIzol reagent.

cDNA Library Construction of Ribosome-Protected mRNA Fragments. Purified RNA samples were first mixed with 1 nM of synthetic 28-nt random RNA (5′-AUGUACACGGAGUCGACCCGCAACGCGA-3′) as the spike-in control. The mixed RNA samples were then dephosphorylated in a 15 μL reaction containing 1× T4 polynucleotide kinase buffer, 10 U SUPERase_In, and 20 U T4 polynucleotide kinase (NEB). Dephosphorylation was carried out for 1 h at 37 °C, and the enzyme was then heat-inactivated for 20 min at 65 °C. Dephosphorylated samples were mixed with 2× Novex TBE-Urea sample buffer (Invitrogen) and loaded on a Novex denaturing 15% polyacrylamide TBE-urea gel (Invitrogen). The gel was stained with SYBR Gold (Invitrogen) to visualize the RNA fragments. Gel bands containing RNA species corresponding to 28 nt were excised and physically disrupted by using centrifugation through the holes of the tube. RNA fragments were dissolved by soaking overnight in gel elution buffer (300 mM NaOAc, pH 5.5, 1 mM EDTA, 0.1 U/mL SUPERase_In). The gel debris was removed using a Spin-X column (Corning) and RNA was purified by using ethanol precipitation.

Purified RNA fragments were resuspended in 10 mM Tris (pH 8) and denatured briefly at 65 °C for 30 s Poly-(A) tailing reaction was performed in a 8 μL with 1 × poly-(A) polymerase buffer, 1 mM ATP, 0.75 U/μL SUPERase_In, and 3 U *E. coli* poly-(A) polymerase (NEB). Tailing was carried out for 45 min at 37 °C. For reverse transcription, the following oligos containing barcodes were synthesized:

MCA02, 5′-pCAGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTTTTTVN-3′; LGT03, 5′-pGTGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT TTTTTTTTTTTTTTTTTTVN-3′; YAG04, 5′-pAGGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT TTTTTTTTTTTTTTTTTTVN-3′; HTC05, 5′-pTCGATCGTCGGACTGTAGAACTCTCAAGCAGAAGACGGCATACGATT TTTTTTTTTTTTTTTTTTVN-3′.

In brief, the tailed RNA product was mixed with 0.5 mM dNTP and 2.5 mM synthesized primer and incubated at 65 °C for 5 min, followed by incubation on ice for 5 min. The reaction mix was then added with 20 mM Tris (pH 8.4), 50 mM KCl, 5 mM MgCl, 10 mM DTT, 40 U RNaseOUT, and 200 U SuperScript III (Invitrogen). RT reaction was performed according to the manufacturer's instructions. RNA was eliminated from cDNA by adding 1.8 μL 1 M NaOH and incubating at 98 °C for 20 min. The reaction was then neutralized with1.8 μL 1 M HCl. Reverse transcription products were separated on a 10% polyacrylamide TBE-urea gel as described earlier. The extended first-strand product band was expected to be approximately 100 nt, and the corresponding region was excised. The cDNA was recovered by using DNA gel elution buffer (300 mM NaCl, 1 mM EDTA).

First-strand cDNA was circularized in 20 μL of reaction containing 1× CircLigase buffer, 2.5 mM $MnCl_2$, 1 M Betaine, and 100 U CircLigase II (Epicentre). Circularization was performed at 60 °C for 1 h, and the reaction was heat-inactivated at 80 °C for 10 min. Circular single-strand DNA was relinearized with 20 mM Tris-acetate, 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, and 7.5 U APE 1 (NEB). The reaction was carried out at 37 °C for 1 h. The linearized single-strand DNA was separated on a Novex 10% polyacrylamide TBE-urea gel (Invitrogen) as described earlier. The expected 100-nt product bands were excised and recovered as described earlier.

Deep Sequencing. Single-stranded template was amplified by PCR by using the Phusion High-Fidelity enzyme (NEB) according to the manufacturer's instructions. The oligonucleotide primers qNTI200 (5′-CAAGCAGAAGACGGCATA-3′) and qNTI201 (5′-AATGATACGGCGACCACCG ACAGGTTCAGAGTTCTACAGTCCGACG-3′) were used to create DNA suitable for sequencing, i.e., DNA with Illumina cluster generation sequences on each end and a sequencing primer binding site. The PCR contains 1× HF buffer, 0.2 mM dNTP, 0.5 μM oligonucleotide primers, and 0.5 U Phusion polymerase. PCR was carried out with an initial 30 s denaturation at 98 °C, followed by 12 cycles of 10 s denaturation at 98 °C, 20 s annealing at 60 °C, and 10 s extension at 72 °C. PCR products were separated on a nondenaturing 8% polyacrylamide TBE gel as described earlier. Expected DNA at 120 bp was excised and recovered as described earlier. After quantification by Agilent BioAnalyzer DNA 1000 assay, equal amount of barcoded samples were pooled into one sample. Approximately 3–5 pM mixed DNA samples were used for cluster generation followed by sequencing by using sequencing primer 5′-CGACAGGTTCAGAGTTC TACAGTCCGACGATC-3′ (Illumina Genome Analyzer 2 or HiSeq).

Data Analysis. The deep sequencing data of ribosome footprints was processed and analyzed by using a collection of custom Perl scripts. The barcoded multiplex sequencing output files were separated into individual sample datasets according to the first 2-nt barcodes. Second, the 3′ polyA tails allowing one mismatch were identified and removed. After that, the high-quality reads of length ranging from 25 to 35 nt were retained whereas other reads were excluded from the downstream analysis. The sequences of the longest transcript isoform for each human gene were downloaded from the Ensembl database to construct a human transcriptome reference. In addition, the sequence of HA (NC_002017) from the National Center for Biotechnology Information database and the plasmid sequence of Flag-FRB-GFP were used as the reference. The trimmed reads were aligned to the corresponding reference transcripts by SOAP 2.0, allowing as many as two mismatches, and only unique mapping hits were retained. Last, the 5′ end positions of aligned reads were mapped into the coding frame and the number of reads was counted at each codon ranging from −20 codon 5′ UTR to the stop codon for the downstream analysis.

The reproducibility of the RPF distribution on individual transcripts was evaluated by Pearson correlation. The replicates were clustered by Cluster 3.0, and heat maps were produced by Treeview. To compare the RPF distribution on transcript before and after the affinity purification, the reads in the first 30-codon window were considered as the background because the polypeptides are still buried within the ribosome exit tunnel and cannot be accessed by binding partners. The significance (i.e., *P* value) of the RPF density vs. background within a 10-codon sliding window in the pull-down sample was calculated by Fisher exact test across the transcripts compared with the total sample. The first position at which the *P* value was less than 0.001 was considered as the folding start point. Based on the number of reads after the folding start point, the total and pull-down data were normalized to the same scale. To decrease the counting error, only the positions with reads above the mean reads density in the total sample were treated as comparable sites. Subsequently, the single codon peak ratio was calculated by dividing the normalized reads of pull-down sample to those of total

sample at the same codon. The trend line of the single codon peak ratio was determined by locally estimated scatterplot smoothing (LOESS)by using SigmaPlot 11.0 (Systat).

ACKNOWLEDGMENTS. The authors thank the Cornell University Life Sciences Core Laboratory Center for performing deep sequencing and Dr. Yuxin Mao for help in preparing the hemagglutinin structure shown in Fig. 3*A*. This work was supported by the National Institute of Allergy and Infectious Diseases Division of Intramural Research (A.D., J.R.B., and J.W.Y.), National Institutes of Health Grant 1DP2 OD006449-01 (to S.-B.Q.), Ellison Medical Foundation Grant AG-NS-0605-09 (to S.-B.Q.), and US Department of Defense Exploration-Hypothesis Development Award TS10078 (to S.-B.Q.).

1. Fedorov AN, Baldwin TO (1997) Cotranslational protein folding. *J Biol Chem* 272: 32715–32718.
2. Komar AA (2009) A pause for thought along the co-translational folding pathway. *Trends Biochem Sci* 34:16–24.
3. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147: 789–802.
4. Sosnick TR, Barrick D (2011) The folding of single domain proteins—have we reached a consensus? *Curr Opin Struct Biol* 21:12–24.
5. Kramer G, Boehringer D, Ban N, Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* 16:589–597.
6. Buchan JR, Stansfield I (2007) Halting a cellular production line: Responses to ribosomal pausing during translation. *Biol Cell* 99:475–487.
7. Frydman J (2001) Folding of newly translated proteins in vivo: The role of molecular chaperones. *Annu Rev Biochem* 70:603–647.
8. Nicola AV, Chen W, Helenius A (1999) Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat Cell Biol* 1:341–345.
9. Johnson AE (2005) The co-translational folding and interactions of nascent protein chains: A new approach using fluorescence resonance energy transfer. *FEBS Lett* 579: 916–920.
10. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324:218–223.
11. Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466:835–840.
12. Choi J, Chen J, Schreiber SL, Clardy J (1996) Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* 273:239–242.
13. Qian SB, et al. (2009) Engineering a ubiquitin ligase reveals conformational flexibility required for ubiquitin transfer. *J Biol Chem* 284:26797–26802.
14. Skehel JJ, Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* 69:531–569.
15. Yewdell JW (2010) Monoclonal antibodies specific for discontinuous epitopes direct refolding of influenza A virus hemagglutinin. *Mol Immunol* 47:1132–1136.
16. Yewdell JW, et al. (1993) Mutations in or near the fusion peptide of the influenza virus hemagglutinin affect an antigenic site in the globular region. *J Virol* 67:933–942.
17. Yewdell JW, Yellen A, Bächi T (1988) Monoclonal antibodies localize events in the folding, assembly, and intracellular transport of the influenza virus hemagglutinin glycoprotein. *Cell* 52:843–852.
18. Gerhard W, Yewdell J, Frankel ME, Webster R (1981) Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* 290:713–717.
19. Clark PL, King J (2001) A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *J Biol Chem* 276:25411–25420.
20. Lobato MN, Rabbitts TH (2003) Intracellular antibodies and challenges facing their use as therapeutic agents. *Trends Mol Med* 9:390–396.
21. Stocks MR (2004) Intrabodies: production and promise. *Drug Discov Today* 9:960–966.
22. Oh E, et al. (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* 147:1295–1308.