THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 12, pp. 9259–9268, March 16, 2012

Two Distinct States of *Escherichia coli* Cells That Overexpress Recombinant Heterogeneous β-Galactosidase*[S]

Received for publication, November 25, 2011, and in revised form, January 31, 2012 Published, JBC Papers in Press, February 2, 2012, DOI 10.1074/jbc.M111.327668

Yun Zhao[‡§1], Wei He[‡§1], Wei-Feng Liu[¶], Chun-Chun Liu[‡§], Li-Kui Feng[¶], Lei Sun[‖], Yong-Bin Yan[¶2], and Hai-Ying Hang[‡3]

From the [‡]Key Laboratory of Protein and Peptide Pharmaceuticals and the [‖]Center for Biological Imaging, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, the [§]Graduate School of the Chinese Academy of Sciences, Beijing 100039, China, and the [¶]State Key Laboratory of Biomembrane and Membrane Biotechnology, School of Life Sciences, Tsinghua University, Beijing 100084, China

Background: Overexpressed heterogeneous protein often exists as soluble and insoluble forms.
Results: *Escherichia coli* cells overexpressing heterogeneous β-galactosidase are mainly composed of two subpopulations with soluble protein and inclusion bodies, respectively.
Conclusion: The ratio of soluble and insoluble protein is determined by the ratio of the two cell subpopulations.
Significance: Our data shed new light on the formation of the equilibrium between soluble and insoluble forms.

The mechanism by which inclusion bodies form is still not well understood, partly because the dynamic processes of the inclusion body formation and its solubilization have hardly been investigated at an individual cell level, and so the important detailed information has not been acquired for the mechanism. In this study, we investigated the *in vivo* folding and aggregation of *Aspergillus phoenicis* β-D-galactosidase fused to a red fluorescence protein in individual *Escherichia coli* cells. The folding status and expression level of the recombinant β-D-galactosidase at an individual cell level was analyzed by flow cytometry in combination with transmission electron microscopy and Western blotting. We found that individual *E. coli* cells fell into two distinct states, one containing only inclusion bodies accompanied with low galactosidase activity and the other containing the recombinant soluble galactosidase accompanied with high galactosidase activity. The majority of the *E. coli* cells in the later state possessed no inclusion bodies. The two states of the cells were shifted to a cell state with high enzyme activity by culturing the cells in isopropyl 1-thio-β-D-galactopyranoside-free medium after an initial protein expression induction in isopropyl 1-thio-β-D-galactopyranoside-containing medium. This shift of the cell population status took place without the level change of the β-D-galactosidase protein in individual cells, indicating that the factor(s) besides the crowdedness of the recombinant protein play a major role in the cell state transition. These results shed new light on the mechanism of inclusion body formation and will facilitate the development of new strategies in improving recombinant protein quality.

Recombinant protein production plays a significant role in both basic biological research and industrial protein engineering (1). Numerous heterogeneous protein expression systems have been developed to produce target proteins, and among them, the enterobacterium *Escherichia coli* is the most widely used host cells for overexpressing recombinant proteins because of its advantages of rapid growth, well studied genetic features, and easy manipulation (1–3). However, the application of *E. coli* in heterogeneous protein expression is significantly limited by the frequently encountered problem that overexpressed recombinant proteins fail to reach their correct conformation. This may be caused by the lack of necessary post-translational modifications, of folding modulators, or of the control of the folding rates of large proteins in the crowded *E. coli* cytoplasm (2, 4, 5). The misfolded proteins usually undergo proteolytic degradation or, more frequently, are deposited into insoluble aggregates called inclusion bodies (IBs)[4] (2). It was found that IBs are not the dead end of recombinant proteins, and IBs seem to be dynamic structures formed by unbalanced equilibrium between soluble proteins and IBs in *E. coli* (6). Although considerable progress has been achieved in developing new strategies (1, 3, 7–10), the lack of a thorough understanding of the formation of IBs in *E. coli* makes it often impossible to predict whether a recombinant protein will be soluble, partially, or completely deposited into IBs (11).

Protein folding and misfolding are complex processes both *in vivo* and *in vitro* (12, 13). The biophysical and biochemical properties of IBs in *E. coli*, as well as in other host cells, have been extensively studied in the past 20 years. IBs are porous ovoids or cylinders with diameter around 1 μm (14, 15) and are

* This work was supported by Grant 2011CBA00906 (to H.-Y. H.) from the National Key Basic Research Program of China (973) and Grant 2011YQ03013404 (to H.-Y. H.) from the National Special Projects for Key Scientific Instruments and Equipment Development of Ministry of Science and Technology, and Grant 30970559 (to Y.-B. Y.) from the National Natural Science Foundation of China.

[S] This article contains supplemental Tables S1 and S2 and Figs. S1–S5.

[1] Both authors contributed equally to this work.
[2] To whom correspondence may be addressed: School of Life Sciences, Tsinghua University, Beijing, 100084, China. Tel.: 86-10-6278-3477; Fax: 86-10-6277-1597; E-mail: ybyan@tsinghua.edu.cn.
[3] To whom correspondence may be addressed: Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. Tel.: 86-10-6488-8473; Fax: 86-10-6488-8473; E-mail: hh91@sun5.ibp.ac.cn.

[4] The abbreviations used are: IB, inclusion body; IPTG, isopropyl 1-thio-β-D-galactopyranoside; RFP, red fluorescence protein.

composed of a high level (up to >90%) of target proteins (16, 17). The association behavior of misfolded proteins in IBs (18, 19) is similar to the amyloid-like fibrils found in many human diseases (12, 17, 20–22). That is, partially folded proteins with hydrophobic exposure specifically stick together via the formation of intermolecular β-sheet structures by a nucleation-dependent mechanism (17). Thus, aggregates are formed by the growth of protein polymers at a single or limited numbers of seeds, which resulted in a low copy number of IBs in the *E. coli* cell (20). The extent of the formation of IBs has been found to be affected by many factors, such as the amino acid sequence of the recombinant protein, the induction temperature, the culture condition, the rate of protein production and folding, and the availability of essential chaperones (6, 23–26).

However, it is unclear yet how an individual cell responses to the misfolding problem caused by recombinant protein overproduction. Single-cell studies provide an approach to monitor the inner cell information and its kinetic behaviors. Flow cytometry is capable of making rapid, multiparametric, and quantitative analyses at individual cell level (27–30). In this study, β-galactosidase from *Aspergillus phoenicis* was fused to a red fluorescence protein (mRFP1), and flow cytometry was applied to monitor the kinetics of the fusion protein expression and the β-galactosidase activity simultaneously at a single cell level. Flow cytometry was also used to sort the cells with different β-galactosidase activities that later were studied by the other methods for their folding statuses. To our surprise, the results indicate that IBs in a single *E. coli* cell followed a "only IB-bearing" or "soluble protein-dominated" distribution, and this two-state distribution was modulated into a single soluble protein-dominated state by shifting cultured cells to IPTG-free medium following an initial IPTG induction. This cell state transition was not accompanied by a level change of the overexpressed β-galactosidase protein in individual cells, suggesting that factors beyond the crowdedness of the overexpressed protein were involved in the process.

EXPERIMENTAL PROCEDURES

Strains and Vectors—All of the recombinant plasmids were constructed following the standard DNA cloning procedure (31) using *E. coli* strain Top10 (Invitrogen). The gene encoding *A. phoenicis* β-D-galactosidase was cloned from vector APCD6 (a gift provided by Dr. Zhiyang Dong in Institute of Microbiology of Chinese Academy of Sciences) by PCR using *Pfu* polymerase and the following primers: Gal-NheI-F, ATACGCGCTAGCATGAAGCTTTCCTCCGCTTG, and Gal-SacI-R, ATACGCGAGCTCGGTACCGTATGCACCCTTC-CGCTTCTT. The PCR fragment was digested with NheI and SacI and ligated into the expression vector pET28a(+) (Novagen), resulting in the pET28-galactosidase vector encoding β-D-galactosidase sequence. The gene for *RFP* (*mRFP1*) was PCR-amplified from the vector pmRFP-N1 using primers: mRFP-KpnI-5′, ATACGCGGTACCGCCTCCTCCGAGGAC GTCAT, and mRFP-NotI-3′, ATACGCGCGGCCGCTTAA-GATCTGGCGCCGGTGG. The PCR fragment was digested with KpnI and NotI and then ligated into pET28-galactosidase with a two-amino acid linker to avoid possible conformational interference between the adjacent domains. The final vector pET28-galactosidase-mRFP1 encoded the sequence of β-D-galactosidase fused by a His_6 tag sequence at the N terminus and the mRFP1 sequence at the C terminus.

Expression of the fusion protein was done in the *E. coli* strain Rosetta(DE3)pLysS (Novagen). A fresh single colony of the Rosetta(DE3)pLysS cells harboring the expression plasmids was picked and grown overnight at 37 °C in 2 ml of LB medium with the addition of 50 ng/ml kanamycin. The cultures were then inoculated in fresh LB medium (1:100) with 50 ng/ml kanamycin in 5-ml culture volume and grown at 200 rpm to reach an A_{600} value of 0.5–0.6 or designated values. Recombinant protein production was induced by the addition of given amounts of IPTG, followed by incubation for the given hours at 37 °C. The bacterial cells were harvested by centrifugation at 8000 × *g* for 5 min, washed twice with 20 mM PBS buffer (pH 7.0), and then suspended in the PBS buffer for further analysis.

Single-cell β-D-Galactosidase Assay—The staining reagent C_{12}-fluorescein-di-β-D-galactopyranoside was added to the harvested cell cultures to a final concentration of 20 μM. The reaction was carried out at 37 °C in a water bath for 1 h. After the substrate loading, phenylathylthio β-D-galactoside, a reversible inhibitor of β-D-galactosidase, was added to the reaction solutions to quench the staining reaction. Cell suspension of 100 μl was harvested at 4000 rpm for 5min at 4 °C, washed twice in PBS buffer, and then the samples were diluted 1:100 to perform the flow cytometry analysis.

Flow Cytometry Analysis—Bacterial cells were analyzed on a FACSCalibur flow cytometer or sorted with a FACSVantage Diva flow cytometer. The samples were illuminated with a water-cooled argon ion laser at 488 nm. Analytical flow cytometry histograms were recorded and analyzed using standard procedures provided by the manufacturer.

Transmission Electron Microscopy Analyses—The cell samples were collected using the same procedures as those for the flow cytometry analysis. The cells were fixed in 2.5% glutaraldehyde for 24 h and post-fixed in 2% OsO_4 for 2 h. After dehydration by gradient concentrations of ethanol, the samples were embedded in Epon812 and polymerized at 60 °C. Ultrathin sections (50 to 70 nm) were double stained by uranyl acetate and lead citrate. The analysis was performed on a transmission electron microscope (FEI Tecnai 20) at 120 kV.

Western Blotting—The harvested cells were disrupted by sonication in an ice bath. Fractionation of the supernatant and precipitation proteins was performed by centrifugation of the sonicated cells for 10 min at 10,000 rpm in a microcentrifuge. Soluble and insoluble proteins with high and low activity were resolved on 10% SDS-PAGE gel and then were transferred to a polyvinylidene difluoride membrane. The membrane was probed consecutively with the primary and peroxidase-conjugated secondary antibodies, and the signal was detected using SuperSignal West Pico Chemiluminescence Substrate system (product number 34077; Pierce). The primary and secondary antibodies used in this study are mouse anti-His tag antibody (Sigma) and peroxidase-conjugated anti-mouse IgG (w4021; Sigma).

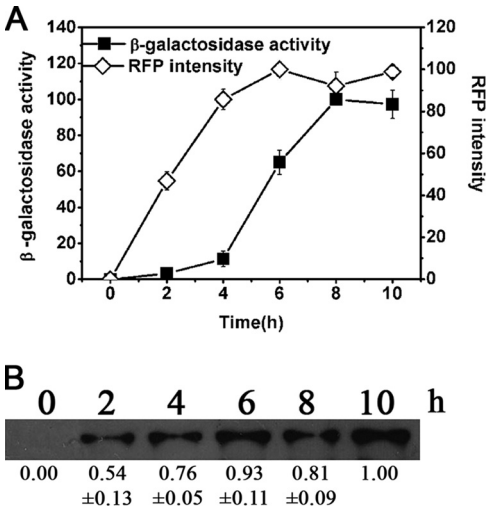

FIGURE 1. **The expression level and fluorescence intensity of mRFP in the fusion protein.** *A*, the β-galactosidase activity (green fluorescence intensity) and RFP intensity of the recombinant cells for designated times. *B*, Western blotting analysis of the recombinant protein at the designated times. Cell lysates from equal numbers of cells (1×10^8 cells/sample), after being induced for 2, 4, 6, 8, and 10 h, were loaded for Western blotting. The levels of the total recombinant protein were monitored by anti-His antibody. Both flow cytometric and Western blotting results were the averages of three repeated experiments.

RESULTS

Recombinant β-Galactosidase from A. phoenicis Used as Model to Study Protein Folding States in Individual Cells—β-Galactosidase, a hydrolase that catalyzes the hydrolysis of β-galactosides into monosaccharides, is frequently used as a reporter enzyme in *E. coli*. The activity of β-galactosidase *in vivo* can be detected by monitoring the hydrolytic extent of the fluorogenic substrate C_{12}-fluorescein-di-β-D-galactopyranoside, which can produce a green fluorescent product that is retained inside the cell (32). In this study, β-galactosidase from *A. phoenicis* was used as a model system to detect the folding status of heterogeneous recombinant proteins overexpressed in *E. coli*. A red fluorescence protein (mRFP1) was fused to the C terminus of β-galactosidase to monitor the *in vivo* level of gene expression (33). Flow cytometry was used to simultaneously monitor β-galactosidase activity and red fluorescence intensity of mRFP1 of the β-galactosidase-mRFP fusion protein in individual *E. coli* cells. In this manner, we intended to establish a quantitative relationship between β-galactosidase activity and its folding status at the same expression levels in individual cells.

The red fluorescence protein mRFP1 is a very fast folder and matured 10 times faster than its parent version, DsRFP (34). To test the suitability of mRFP1 for monitoring the expression levels of the β-galactosidase-mRFP1 fusion protein, we carried out the following experiments. The protein production was induced by 0.5 mM IPTG at 37 °C for the designated times. The mRFP1 fluorescence increased very fast and almost reached the plateau at 4 h, and the red fluorescence dynamic changes largely paralleled to the total expressed β-galactosidase-mRFP1 levels measured by Western blotting (Fig. 1). In contrast, β-galactosidase activity was still at a very low level at 4 h and increased at later times (Fig. 1*A*). These results suggest that the mRFP1 in the fusion protein folds fast largely independent of the β-galactosidase folding status and that the mRFP1 fluorescence is suitable for monitoring the fusion protein levels in individual cells. In the experiments of the following sections, the mRFP1 fluorescence was shown to change little in both soluble and insoluble forms of β-galactosidase-mRFP1 fusion protein (see Fig. 3), as well as in both inclusion bodies and the other cellular parts (see Figs. 5*B* and 7, *A* and *C*), further confirming the suitability of mRFP1 as an indicator of β-galactosidase-mRFP1 levels.

Recombinant β-Galactosidase Existed as Only One of Two Dominant Distinct Folding Statuses in Each Individual Cell—After the induction for 2 h, three dominant populations (R1, R2, and R3) appeared on a flow cytometric plot of side scattering light *versus* forward scattering light (Fig. 2*A*). The dot signals in the lower part of the plot were from the background and appeared when running cell-free PBS through the cytometer (supplemental Fig. S1). Side scattering light and forward scattering light reflect the intracellular granularity and cell size, respectively (28). Most IPTG-induced cells fell into the R1 region, and these cells were large and highly granular. The R3 region contains a small number of cells (<0.47% of the total cell population between 0 and 10 h after adding IPTG) with large cell sizes but little granularity (Fig. 2, *A* and *B*). The cells in the R2 region were small with little granularity (Fig. 2*A*). These cells had very low level of RFP fluorescence intensities, suggesting that these cells expressed the recombinant protein at very low levels or did not express the recombinant protein (Fig. 2*F*). In contrast, the cells in the R1 and R3 regions shared significantly higher red fluorescence intensities, suggesting that a larger cell size reflects high protein production in *E. coli*, whereas a small size reflects no or very low protein production in *E. coli.*

The activity of the recombinant β-galactosidase in the three populations was investigated by monitoring the green fluorescence by flow cytometry (Fig. 2, *C* and *E*). The green fluorescence intensity in R2 region was very low, and the peak value was largely overlapped with that of the negative control cells, indicating that the β-galactosidase activity of the fusion protein was very low in these cells (Fig. 2*D*). A single peak with very high β-galactosidase activity was found in the R3 region cells (Fig. 2, *C* and *E*), suggesting that the recombinant proteins were well folded to its active form. Interestingly, two distinct peaks were identified in the R1 region after culturing in IPTG containing medium for 6 h or longer times according to their different green fluorescence intensities: a low intensity peak and a high intensity peak (Fig. 2*C*). This result suggests that there were two dominant states of *E. coli* cells in the R1 region: one containing the recombinant β-galactosidase-mRFP1 with high activity and the other bearing the enzyme with very low activity. The difference in the β-galactosidase activity of these two states was not due to the disparity of transcription or translation of the recombinant proteins because there was only a single peak of the red fluorescence level in the R1 region cells (Fig. 2*B*, *bottom row*). Similar two-activity cell populations also appeared when the β-galactosidase-mRFP1 form was replaced by the β-galactosidase (supplemental Fig. S2); thus, the distribution reflects the genuine folded statuses of the β-galactosidase but not only those of the β-galactosidase-mRFP1 fusion protein. Because the cells in R2 region had almost no expression of the recombinant protein (Fig. 2*F*) and the R3 region contained very few cells relative to those of the total population (Fig. 2*D*), we will not

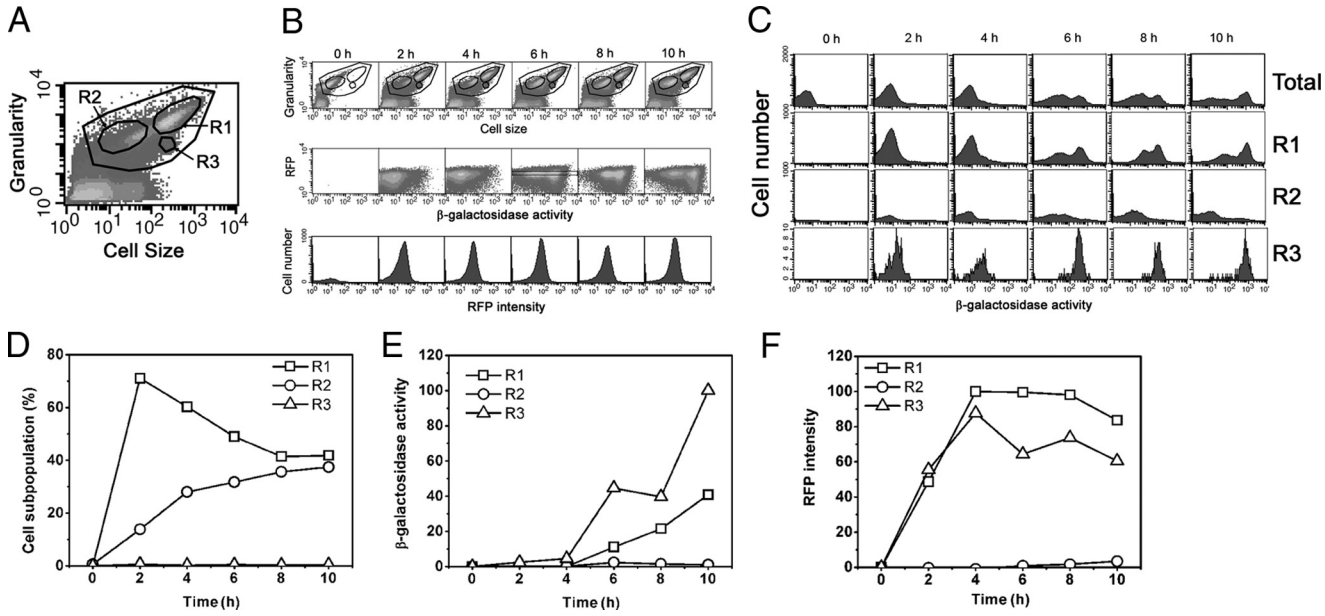

FIGURE 2. **The *in vivo* β-galactosidase activities of the recombinant fusion protein detected by flow cytometry.** *A*, three population cells induced by IPTG. The recombinant cells were induced by 0.5 mM IPTG at 37 °C for 2 h. The cell size (reflected by forward scattering light) and granularity (reflected by side scattering light) of individual *E. coli* cells were analyzed, and three cell populations (R1, R2, and R3) were identified. The three cell populations existed between 2 and 10 h (see *B*). *B*, R1 region cells. The recombinant cells were induced by 0.5 mM IPTG at 37 °C for designated times. *Top row*, three dominant cell populations (R1, R2, and R3) in dot plots were characterized. *Middle row*, RFP fluorescence intensities *versus* β-galactosidase activities in R1 region were shown in dot plots. *Bottom row*, RFP fluorescence intensities in R1 regions were demonstrated in histogram plots. C, β-galactosidase activities of the total and three populations demonstrated in flow cytometric histogram plots. *D*, the percentages of R1, R2, and R3 cell populations during the cell culturing in IPTG-containing medium. The culturing conditions were the same as in *B*. The percentages of each population were calculated on the base that the three populations were 100% at any time. *E*, the β-galactosidase activities of the three cell populations. *F*, RFP intensities of the three cell populations.

further discuss these two cell populations but focus on the cells in the R1 region.

To further confirm the above finding, Western blotting analysis of the recombinant protein was performed for the two cell subpopulations with respectively low and high levels of β-galactosidase activity in R1 region. The two bacterial subpopulations were sorted by flow cytometry based on different green fluorescence levels but within the same intensity scale of red fluorescence to ensure the same expression levels of the recombinant proteins (Fig. 3*A*). After ultrasonication and centrifugation, the supernatants and pellets were analyzed by SDS-PAGE, followed by Western blotting (probed with His tag antibody). The results in Fig. 3*B* clearly indicate that the high β-galactosidase-activity protein was fully soluble, whereas the majority of the low β-galactosidase-activity protein was detected in the precipitated part and only a very small fraction in the soluble fraction. The above data together suggest that there are two subpopulations of expressing β-galactosidase-mRFP1 recombinant protein with distinct folded statuses of the protein. Because a fresh single colony of the bacteria was picked for each of the experiments in this study, the distinct statuses of the cells were not derived from a starting mixed population.

Effects of IPTG Concentration on Ratio of Two Cell Populations with Distinct β-Galactosidase Activities—IPTG concentration in medium is commonly known as an important factor influencing IB formation through inducing different rates of protein synthesis. Therefore, in this study IPTG concentration was tested for its effect on the two-state distribution of the cells that express recombinant β-galactosidase (Fig. 4). After induction for 5.5 h at 37 °C, no significant change was observed for the intensity of the red fluorescence from mRFP1 when IPTG concentration was increased from 0.1 mM to 2 mM (Fig. 4*C*), whereas the percentages of the high activity cells were different at different IPTG concentrations, and its maximum value of ~27% appeared at IPTG concentration = 0.5 mM (Fig. 4, *A* and *B*). The above experimental results suggest that the distribution of the low and high activity cells is an important intermediate parameter through which changing IPTG concentration alters the ratio of IB to soluble recombinant protein in the whole cell population in R1 region.

Effects of Conditioned LB Medium on Ratio of Two Cell Populations with Distinct β-Galactosidase Activities—The above experimental data indicate that IPTG concentration is a major factor that influences the distribution of the cells bearing the recombinant protein with different folded statuses. We asked whether IPTG exerts its role in making the slightly active recombinant enzyme through inducing too much protein in a short time. After being induced by 0.5 mM IPTG at 37 °C for 4 h, the cells were harvested by centrifugation and washed twice with fresh LB medium. The cells were recultured in prewarmed conditioned LB medium, which was collected from a bacterial culture without IPTG for 4 h at 37 °C by centrifuging and removing *E. coli*. The cell samples were then collected at 2-h intervals. As shown in Fig. 5*B*, removing IPTG did not significantly change the RFP intensity, indicating similar expression levels of recombinant protein. However, the percentage of high activity cell population was increased remarkably by culturing in IPIG-free conditioned medium (Fig. 5, *A* and *C*). The preconditioned medium was used to keep everything else the same except IPTG for the experiments. We also replaced the old

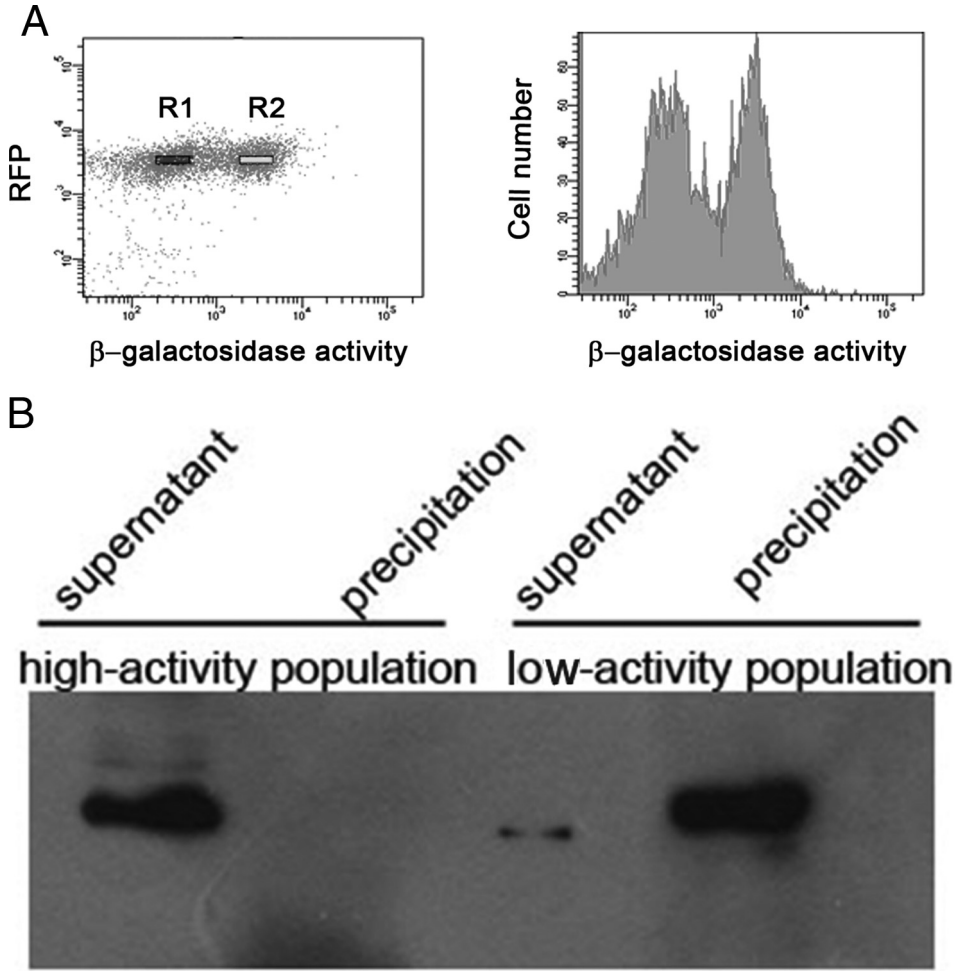

FIGURE 3. **Western blotting analysis of the cells with high or low β-galactosidase activity in the R1 population.** *A*, 5×10^6 cells induced 5.5 h with IPTG were first gated by flow cytometry according to the equal intensity of red fluorescence to ensure an identical expression level of recombinant proteins and then were sorted into two populations according to the intensity of green fluorescence. *B*, these two sorted populations, the high activity population and low activity population, were then collected and treated by ultrasonication followed by centrifugation. The supernatants and pellets of centrifugation were analyzed by Western blotting. Lysates from equal numbers of both cell population were loaded for Western blotting analysis.

medium with a freshly prepared LB medium devoid of IPTG, and the result was similar to that using the preconditioned medium (supplemental Fig. S3). The above results together suggest that the replacement of IPTG-containing medium with IPTG-free medium significantly enhances the percentage of the high activity cells. We added tetracycline to the medium (to inhibit protein translation) to confirm the above phenomenon that the cell state shifting was not due to the reduction of the recombinant protein level in individual cells. After induction for 4 h, the cells were either washed or left unwashed and then incubated for further 4 h. The data showed that adding 10 μg/ml tetracycline has no obvious effect on both the enzyme activity of the R1 cells and the percentage of the high activity cells in both washed and unwashed cells (Fig. 6). The "no effect" is not because tetracycline did not inhibit translation, because adding tetracycline to cells at 0 h completely blocked the translation of the recombinant protein (Fig. 6).

Effects of Conditioned LB Medium on Percentages of IB-free Cells Identified by Transmission Electron Microscope—The above experimental results show that IPTG removal shifts the distribution between the cells containing high and low activity recombinant β-galactosidases. Fig. 3 demonstrates that the majority of the β-galactosidase-mRFP1 in the low activity cells was insoluble, whereas the enzyme proteins in the high activity cells were soluble. We inferred that the majority of β-galactosidase-mRFP1 in the low activity cells is in the IB form and the β-galactosidase-mRFP1 in the high activity cells are in the soluble non-IB form. An investigation using transmission electron microscopy demonstrated that there were mainly two types of the IPTG-induced unwashed cells regarding their internal protein distribution (Fig. 7*A*). One type contained small dense patches, and the rest contained large areas that appeared dark. The dense patches are most likely to be IBs as reported previously (35, 36), whereas the dark areas probably contained no or little recombinant protein. Some of this type of cells also contained less dense materials besides small dense patches. The other type of cells contained many less dense patches in much larger areas of the cells. These two morphologies might correspond to the two-state distribution of enzyme activity. Consistent with this suggestion, the majority of the washed cells (IPTG-induced only for 4 h and incubated without IPTG for the rest of the time) contained less dense white materials in a much wider area and thin pale materials over the whole cell (Fig. 7*A*). The randomly picked "feature-free areas" in the unwashed and

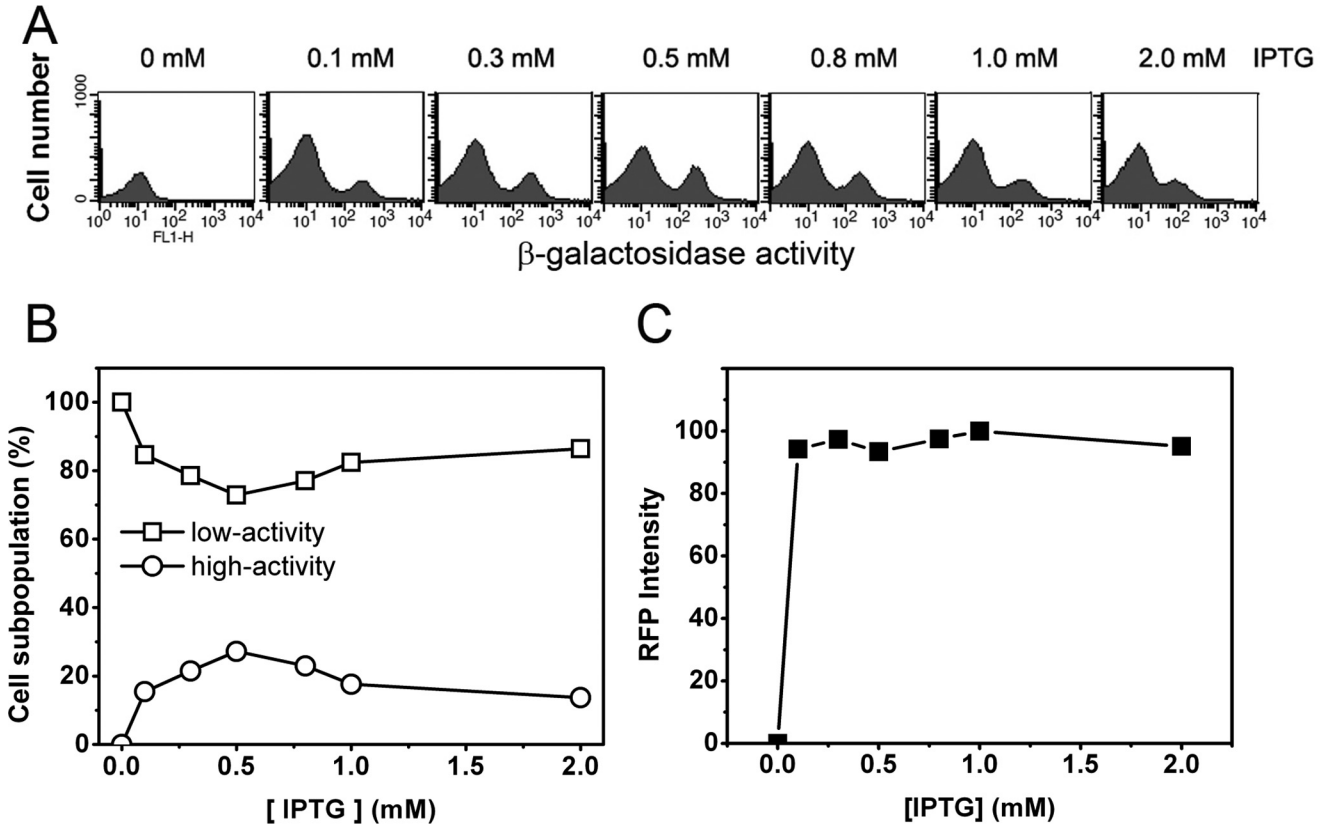

FIGURE 4. **Effects of IPTG concentrations on the percentages of two β-galactosidase activity populations in the R1 region.** *A*, the β-galactosidase activity distribution identified by FL1 (green fluorescence) using flow cytometry. *B*, the percentages of two β-galactosidase activity populations. *C*, the red fluorescence intensities of the R1 region cells. The cells were cultured for 5.5 h.

washed cells were amplified six times. The "featureless areas" in the washed cells obviously contained clearly observable materials, in sharp contrast to the randomly picked dark areas in unwashed cells (Fig. 7*B*). If IB-free cells are defined as the cells containing loose white materials and possessing no dark area, the number of inclusion body-free cells was much increased (37.5–91.2% of the total cells at 10 h) after the removal of IPTG (Fig. 7*C*, *white bars*). This result suggests that IPTG withdraw increases the percentage of the IB-free cells. It is clear that the low activity cells almost disappeared, and the high activity cells accounted for 98% of R1 cells after being washed and cultured in IPTG-free conditioned medium for 6 h (Fig. 5, *A* and *C*). Therefore, the cells containing very dense patches, as well as loose patches and thin white materials in a wide area (7.1% of R1 cells at 10 h), are very likely to be in the group with high enzyme activity. In summary, there were two cell states: one in which white thin materials existed in all the intracellular space and another in which there were no discernable materials in the majority of the intracellular space; the percentages of the two types of cells correlated respectively with the percentages of the high and low enzyme activity R1 cells.

DISCUSSION

In this study three dominant groups of an *E. coli* cell population overexpressing β-galactosidase-mRFP1 fusion protein were identified in side scattering light *versus* forward scattering light plots (Fig. 2, *A* and *B*). Red fluorescence was shown to be able to measure the fusion protein levels in individual cells (Fig. 1), and green fluorescence was shown to distinguish the individual cells bearing the highly active and soluble β-galactosidase from those containing the slightly active and insoluble β-galactosidase (Figs. 2 and 3). The average enzyme activities of the high activity enzyme-containing cells were 13.1–17.5-fold those of the low activity enzyme-containing cells 4–10 h after IPTG induction (supplemental Table S1). This capability of simultaneously measuring multiple parameters in individual *E. coli* cells by flow cytometry allowed us to identify a large-sized and highly granular cell population as the major cells overexpressing the recombinant protein and to find that these cells consisted of two subpopulations containing, respectively, the highly and slightly active β-galactosidase proteins. Further analyses using flow cytometric sorting and Western blotting indicates that the highly active enzyme-containing cells carried only the soluble form of the recombinant protein, whereas the slightly active enzyme-containing cells carried mainly the insoluble form of this protein (Fig. 3). Transmission electron microscopy analysis demonstrates that there were three types of cells that were induced by IPTG to overexpress the recombinant protein (Fig. 7*A*, *top row*). In the first type of cells, very dense patch(s) (IBs) were located in very small area(s), and darkness filled in the rest of the areas. In the second type of cells, loose white materials were located in very large areas or filled in the whole intracellular space. A small number of the second type of cells also contained IBs. In the third type of cells, the darkness was shown to occupy all of the area inside the cells. We

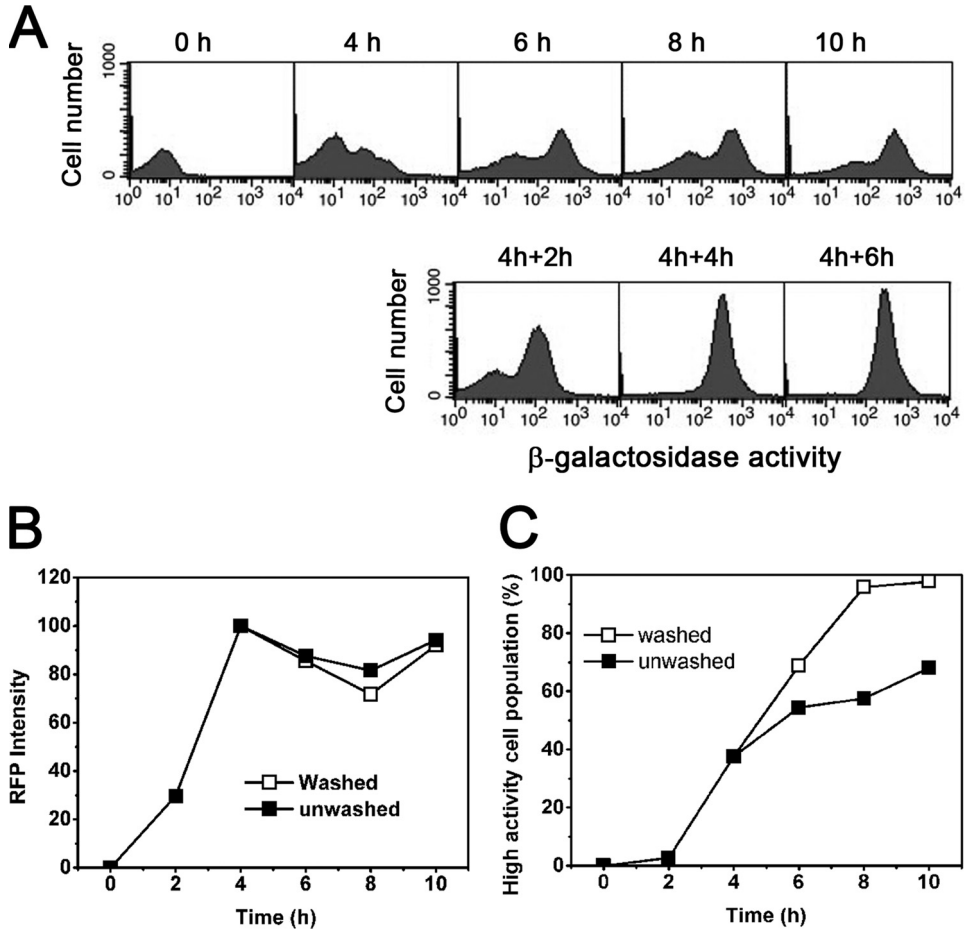

FIGURE 5. **Effects of IPTG removal after being induced for 4 h on the percentages of two β-galactosidase activity populations in the R1.** *A*, the β-galactosidase activity distribution without IPTG removal (*top row*) or when IPTG was removed (*bottom row*). *B*, RFP intensity. *C*, the percentages of high activity populations.

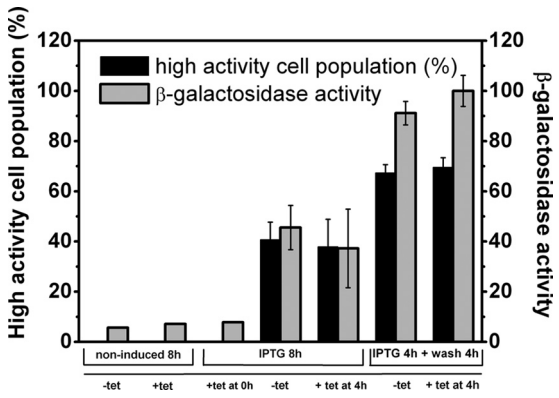

FIGURE 6. **Effects of tetracycline on the β-galactosidase activity profiles in the R1 region.** Tetracycline (10 μg/ml) was added to the medium to inhibit protein translation. After induction for 4 h, IPTG was either washed or reserved for a further 4 h of induction. The percentages of high activity populations and enzyme activities were determined by flow cytometry detailed under "Experimental Procedures."

infer that the first and second types of cell were, respectively, the slightly and highly active enzyme-containing cells, and the third type of cells did not overexpress the recombinant protein or those that contained IBs (IBs were small in size and did not appear in some demonstrated sections in transmission electron microscopy analysis). Therefore, the IB-containing third type of cells should also be included into those with low enzyme activity. The above data suggest that the loose white material-dominated cells are the highly active enzyme-containing cells, whereas the dark IB-bearing cells are the slightly active enzyme-containing cells. After IPTG withdraw and further culturing for another 6 h, the second type of cells, including IB-free and IB-bearing cells (Fig. 7*C*), was increased to 98% of the R1 cells, and concomitantly the ratio of the high activity cells also dramatically enhanced and accounted for 98.3% of the counted cells (Fig. 7). This pair of data strongly supports the above supposition on the cell classification. We designated the dark and IB-bearing cells as DIB cells, and the loose white-material-dominated cells as LW cells (Fig. 8). Here we emphasize the loose white and thin white materials in the whole intracellular space or in the most space as a defining point for a high enzyme activity cell. Some high enzyme activity cells, although containing IBs, fit into the above definition. The finding of the two states of cells overexpressing the heterozygous β-galactosidase can provide us with new insights into the mechanisms of IB formation and reversion as discussed below.

The first hypothesis we propose here is that the other factor(s) besides the high concentration of overexpressed protein can play a critical role in the equilibrium between soluble and IB-included proteins. Evidence accumulated by many research laboratories has suggested that IBs are formed when the concentration of partially folded polypeptides reaches a critical

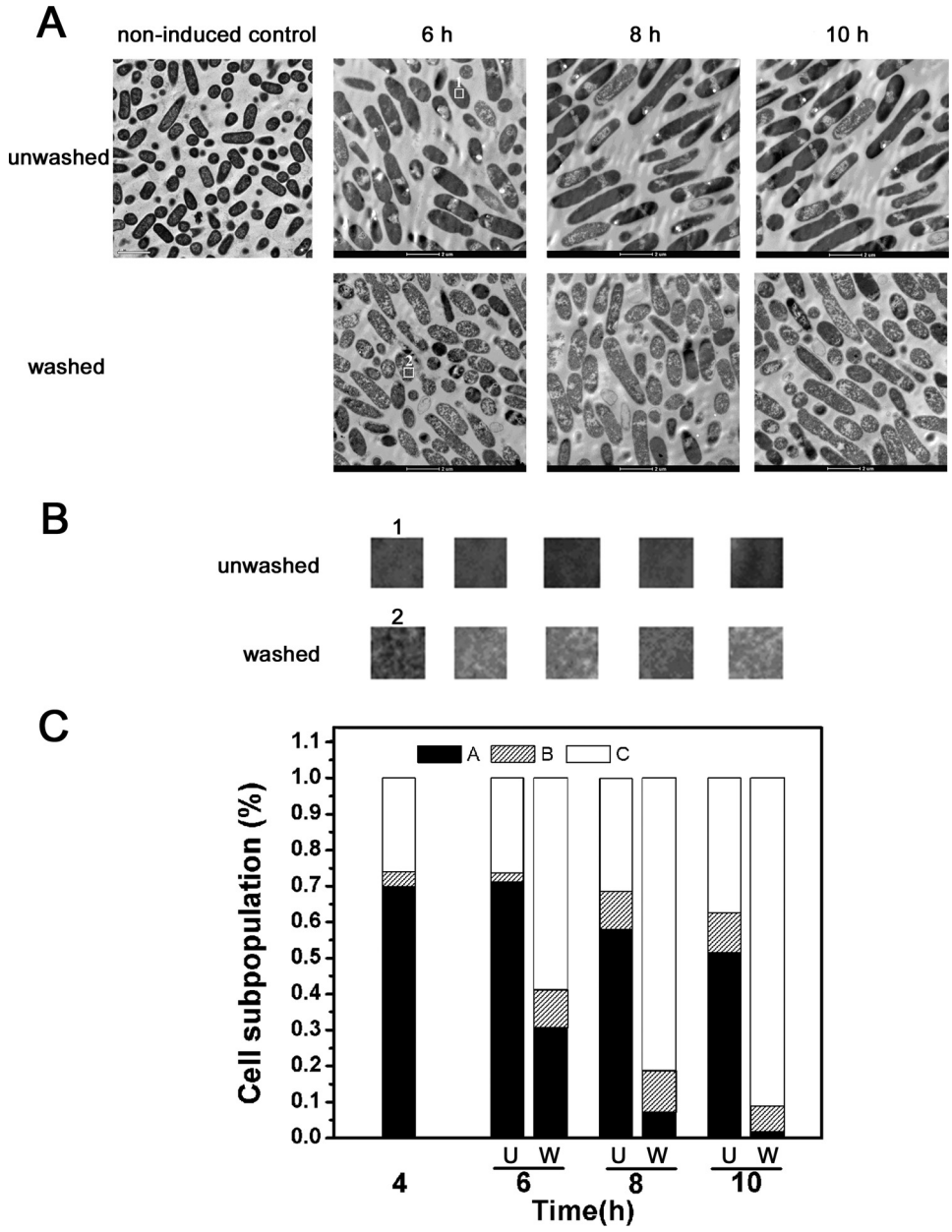

FIGURE 7. **Effects of IPTG removal on the percentages of non-inclusion body cells indentified by transmission electron microscopy.** *A*, the noninduced and IPTG-induced cells, as well as the cells first induced with IPTG for 4 h and then cultured in IPTG-free medium for the designated times, prepared and observed by transmission electron microscopy. *B*, the detailed morphological difference of the featureless areas of the cells continually cultured in IPTG-containing medium and the cells first induced with IPTG for 4 h and then cultured in IPTG-free medium. The featureless areas were purposely selected (for example *areas 1* and *2*) from *A* of this figure and amplified six times. White materials were clearly notable in the featureless areas of washed cells. *C*, the percentages of the cells with the following features: IB-containing accompanied with darkness in the rest of the large area (*black bars*, *A*), IB-bearing and loose white material-dominated (*hatched bars*, *B*), and IB-free and loose white material-dominated (*white bars*, *C*). The completely dark, small and hollow cells were not counted. At least 200 cells were counted for each bar from randomly selected microscopic photographs. *U*, unwashed; *W*, washed.

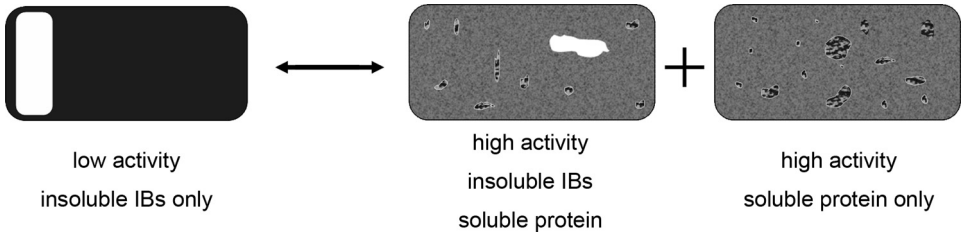

FIGURE 8. **A speculative model of two distinct states of cells overexpressing the heterogeneous β-galactosidase.** The *white blocks* stand for inclusion bodies. The *gray area* only appears in high enzyme activity cells (right in the equilibrium). The particulate-like materials spread in the cell represents loose form of heterogeneous β-galactosidase. We classify the right two types of cells as one state (LW, high activity and soluble protein-dominant) and the left type of cells as another state (DIB, low activity and only IB-containing).

level (18, 25, 26, 28–30). Indeed, the importance of the crowdedness of newly synthesized proteins in forming insoluble IBs is also reflected in this study. R3 cells contained significantly less β-galactosidase-mRFP1 protein than R1 cells (Fig. 2*F*), and the average enzyme activity in R3 cells was much higher than that of R1 cells (Fig. 2*E*), suggesting that more protein in R3 folded into the final active form than that in R1 cells. Actually, almost all of the cells in the R3 group had the average enzyme activity comparable with that of high enzyme activity-containing cells in R1 group. However, in this study, the other factor(s) beyond the protein concentration were shown to also play an important role in the equilibrium between soluble and IB-formed proteins by the three following pieces of evidence. First, within the R1 cell population (the major cell population that overexpressed β-galactosidase-mRFP1), one subpopulation (LW) contained the soluble and highly active form of β-galactosidase-mRFP1, whereas the other subpopulation (DIB) contained the insoluble and slightly active form of β-galactosidase-mRFP1, although the concentrations of β-galactosidase-mRFP1 in individual cells of the two subpopulations were similar (Fig. 3). Second, the percentage of LW cells increased significantly by shifting cells from IPTG-containing medium to IPTG-free medium, although the recombinant protein levels remained the same (Figs. 5 and 7). Third, the addition of tetracycline to cell culture at 4 h (to inhibit IPTG-induced protein synthesis) did not change the percentage of the LW cells and the β-galactosidase activity of R1 cells at 8 h whether IPTG was removed or not from the culture (Fig. 6).

The second hypothesis is that the transition between DIB and LW cells is a key intermediate element through which factors modulate the ratio of soluble and IB-included protein in the whole cell population. In other words, the ratio change between soluble and IB-included protein is not through a gradual molecular transition in individual cells but due to the ratio change of DIB and LW cells. This hypothesis is supported by the following two pieces of evidence. First, IPTG concentration modification changed the average β-galactosidase activity of R1 cells through changing the ratio of the high activity (LW) and low activity (DIB) cells (Fig. 4). Second, removal of IPTG from the culture increased the ratio of soluble protein over IBs through the increase of LW cell percentage (Figs. 5 and 7). We also found that the *A* value (cell concentration) of the culture influenced the percentages of LW and DIB cells and led to the change of the average β-galactosidase activity of R1 cells (supplemental Fig. S4).

How does the quantum-like transition between the two cell states occur? According to previous studies from other research groups, the kinetics of IB formation was nucleation-dependent (17, 19, 20). The nucleation results in a limited copy number of the inclusion bodies in the bacteria (20). After induction, the newly produced polypeptides rapidly folded into a partially folded state, as proposed by *in vitro* studies for many proteins (21). With increasing incubation time, the accumulation of the partially folded proteins to a level above the critical concentration leads to the formation of an aggregation nucleus in some cells. Once the nucleus was formed, the fate of the newly produced polypeptides was triggered by the nucleus to the formation of IBs. This nucleation mechanism can explain a phenomenon that IBs are formed when heterogeneous protein is overexpressed, whereas no or less IB protein is formed when heterogeneous protein is expressed at a low level. However, the nucleation hypothesis cannot fully explain the phenomena that there were both IB-containing and IB-free cells, although they expressed β-galactosidase-mRFP1 protein at equally high levels (Fig. 3) and that replacing old IPTG-containing medium with conditional IPTG-free medium dramatically enhanced the percentage of the high enzyme activity population, whereas the fusion protein levels were not changed (Fig. 5). The above analyses indicate that the other currently unknown intracellular factors besides the concentration-dependent nucleation also play an important role in determining the folding status of overexpressed heterogeneous protein in a cell.

There are critical open questions on the distribution of two-state cells overexpressing heterozygous protein: 1) What are the other intracellular factors important in determining the folding status of overexpressed heterogeneous protein in a cell? 2) How do factors such as protein concentration and chaperones influence the transition between the two states of the cell? 3) Is this two-state distribution a common phenomenon for other heterozygous overexpressed proteins? The wider significance of the two-state cell distribution demonstrated in this study will not be known before the above questions have been answered; thus, the above questions warrant further investigation.

To further probe for the generality of the protein state transition after IPTG removal, we measured soluble and insoluble portions of four other heterozygous proteins before and after IPTG removal (supplemental Fig. S5 and Table S2). The soluble portions of the two proteins tBgl1 and tBgl3 were increased compared with the insoluble portions after IPTG removal. There were no solubility changes after IPTG removal on the two other proteins hCaf1 and hPop2. Note that the later two proteins were only in IB form when overexpressed. Probably the two-state distribution is a precondition for the transition. In summary, the above experimental results suggest that continued culturing without IPTG after preinduced by IPTG can increase the ratio of soluble protein on some other proteins besides the β-galactosidase from *A. phoenicis* and that there might be a two-state distribution *E. coli* expressing some other proteins in addition to β-galactosidase from *A. phoenicis*. To adequately prove this point, it needs protein systems in which protein state can be monitored in individual cells such as the β-galactosidase from *A. phoenicis* used in this study.

REFERENCES

1. Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* **22,** 1399–1408
2. Schein, C. H. (1989) Production of soluble recombinant proteins in bacteria. *Nat. Biotechnol.* **7,** 1141–1149
3. Sørensen, H. P., and Mortensen, K. K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* **4,** 1
4. Gatenby, A. A., Viitanen, P. V., and Lorimer, G. H. (1990) Chaperonin assisted polypeptide folding and assembly. Implications for the production of functional proteins in bacteria. *Trends Biotechnol.* **8,** 354–358
5. Teschke, C. M., and King, J. (1992) Folding and assembly of oligomeric proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **3,** 468–473
6. Carrio, M. M., and Villaverde, A. (2001) Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Lett.* **489,** 29–33

7. Makrides, S. C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60,** 512–538
8. Thomas, J. G., Ayling, A., and Baneyx, F. (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold. *Appl. Biochem. Biotechnol.* **66,** 197–238
9. Sørensen, H. P., and Mortensen, K. K. (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115,** 113–128
10. Hunt, I. (2005) From gene to protein. A review of new and enabling technologies for multi-parallel protein expression. *Protein Expr. Purif.* **40,** 1–22
11. Ventura, S., and Villaverde, A. (2006) Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **24,** 179–185
12. Markossian, K. A., and Kurganov, B. I. (2004) Protein folding, misfolding, and aggregation. Formation of inclusion bodies and aggresomes. *Biochemistry* **69,** 971–984
13. Lei, H. X., and Duan, Y. (2011) Protein folding by computer: flow far have we gone? *Acta Biophys. Sin.* **27,** 187–194
14. Taylor, G., Hoare, M., Gray, D. R., and Marston, F. A. (1986) Size and density of protein inclusion bodies. *Nat. Biotechnol.* **4,** 553–557
15. Bowden, G. A., Paredes, A. M., and Georgiou, G. (1991) Structure and morphology of protein inclusion bodies in *Escherichia coli*. *Biotechnology* **9,** 725–730
16. Speed, M. A., Wang, D. I., and King, J. (1996) Specific aggregation of partially folded polypeptide chains. The molecular basis of inclusion body composition. *Nat. Biotechnol.* **14,** 1283–1287
17. Fink, A. L. (1998) Protein aggregation. Folding aggregates, inclusion bodies and amyloid. *Fold Des.* **3,** R9–R23
18. Ami, D., Natalello, A., Gatti-Lafranconi, P., Lotti, M., and Doglia, S. M. (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. *FEBS Lett.* **579,** 3433–3436
19. Morell, M., Bravo, R., Espargaró, A., Sisquella, X., Avilés, F. X., Fernàndez-Busquets, X., and Ventura, S. (2008) Inclusion bodies. Specificity in their aggregation process and amyloid-like structure. *Biochim. Biophys. Acta* **1783,** 1815–1825
20. Kopito, R. R. (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10,** 524–530
21. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426,** 884–890
22. Lei, H. (2010) Amyloid and Alzheimer's disease. *Protein Cell* **1,** 312–314
23. Strandberg, L., and Enfors, S. O. (1991) Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. *Appl. Environ. Microbiol.* **57,** 1669–1674
24. Wilkinson, D. L., and Harrison, R. G. (1991) Predicting the solubility of recombinant proteins in *Escherichia coli*. *Biotechnology* **9,** 443–448
25. García-Fruitós, E., Carrió, M. M., Arís, A., and Villaverde, A. (2005) Folding of a misfolding-prone β-galactosidase in absence of DnaK. *Biotechnol. Bioeng.* **90,** 869–875
26. García-Fruitós, E., Martínez-Alonso, M., Gonzàlez-Montalbán, N., Valli, M., Mattanovich, D., and Villaverde, A. (2007) Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*. *J. Mol. Biol.* **374,** 195–205
27. Lavergne-Mazeau, F., Maftah, A., Cenatiempo, Y., and Julien, R. (1996) Linear correlation between bacterial overexpression of recombinant peptides and cell light scatter. *Appl. Environ. Microbiol.* **62,** 3042–3046
28. Patkar, A., Vijayasankaran, N., Urry, D. W., and Srienc, F. (2002) Flow cytometry as a useful tool for process development. Rapid evaluation of expression systems. *J. Biotechnol.* **93,** 217–229
29. Choe, W. S., Clemmitt, R. H., Chase, H. A., and Middelberg, A. P. (2003) Coupling of chemical extraction and expanded-bed adsorption for simplified inclusion-body processing. Optimization using surface plasmon resonance. *Biotechnol. Bioeng.* **81,** 221–232
30. Hedhammar, M., Stenvall, M., Lönneborg, R., Nord, O., Sjölin, O., Brismar, H., Uhlén, M., Ottosson, J., and Hober, S. (2005) A novel flow cytometry-based method for analysis of expression levels in *Escherichia coli*, giving information about precipitated and soluble protein. *J. Biotechnol.* **119,** 133–146
31. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
32. Plovins, A., Alvarez, A. M., Ibañez, M., Molina, M., and Nombela, C. (1994) Use of fluorescein-di-β-D-galactopyranoside (FDG) and C12-FDG as substrates for β-galactosidase detection by flow cytometry in animal, bacterial, and yeast cells. *Appl. Environ. Microbiol.* **60,** 4638–4641
33. García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R. M., Arís, A., Ventura, S., and Villaverde, A. (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb. Cell Fact.* **4,** 27
34. Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (2002) A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 7877–7882
35. Carrió, M. M., Cubarsi, R., and Villaverde, A. (2000) Fine architecture of bacterial inclusion bodies. *FEBS Lett.* **471,** 7–11
36. Carrió, M. M., and Villaverde, A. (2005) Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. *J. Bacteriol.* **187,** 3599–3601