Two distinct states of *Escherichia coli* cells that overexpress the recombinant heterogeneous

β-galactosidase *

Yun Zhao #1,2, Wei He #1,2, Wei-Feng Liu 3, Chun-Chun Liu 1,2, Li-Kui Feng 3, Lei Sun 4, Yong-Bin Yan *3, Hai-Ying Hang *1

*Running title: Two distinct states of *E. coli* cells expressing heterogeneous protein

1 Laboratory of Protein and Peptide Drugs, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. 2 Graduate School of the Chinese Academy of Sciences, Beijing 100039, China. 3 State Key Laboratory of Biomembrane and Membrane Biotechnology, School of Life Sciences, Tsinghua University, Beijing 100084, China. 4 Center for Biological Imaging, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.

# These authors contributed equally to this work.

* To whom all correspondence should be addressed: Dr. Yong-Bin Yan, 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; Dr. Hai-Ying Hang, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. Tel: +86-10-6488 8473; Fax: +86-10-6488 8473; Email: hh91@sun5.ibp.ac.cn

**Keywords:** β-D-galactosidase; flow cytometry; inclusion body formation; recombinant protein production; *Escherichia coli*; protein folding and aggregation

**Background:** Overexpressed heterogeneous protein often exists as soluble and insoluble forms.

**Results:** *Escherichia coli* 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.

**SUMMARY**

The mechanism by which inclusion bodies form is still not well understood, partly because the dynamic processes of the inclusion body formation and its solubilisation 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 *E. 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 IPTG-free medium after an initial protein expression induction in IPTG-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 due to 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 a 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) (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 make it often impossible to predict whether a recombinant protein will be soluble, partially or completely deposited into IBs (11).

Protein folding and misfolding is complex process 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 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 over-production. Single cell studies provide an approach to monitor the inner cell information and its kinetic behaviors. Flow cytometry is capable of making rapid, multi-parametric 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 status. 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 recombinant plasmids were constructed following the standard DNA cloning procedure (31) using E. coli strain Top10 (Invitrogen). The gene encoding Aspergillus 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:

ATACGCAGTCATGAAGCTTTCCTCCGCTTG and Gal-Sacl-R:

ATACGCAGCTCGGTACCGTATGCACCCTTCGCTTCTT. The PCR fragment was digested with NheI and Sacl 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':

ATACGCGGTACCGCCTCCTCCGAGGACGTCAT mRFP-NotI-3':

ATACGCGCGGCCGCTTAAGATCTGGCGCCGTGG. The PCR fragment was digested with KpnI and NotI, then ligated into pET28-galactosidase with a 2-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 His6-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 2ml 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 OD600 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 given hours.
at 37°C. The bacterial cells were harvested by centrifugation at 8000 g for 5 min, and 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, C12FDG, was added to the harvested cell cultures to a final concentration of 20 μM/ml. The reaction was carried out at 37°C in a water bath for 1 h. After the substrate loading, PETG, 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 5 min at 4°C, washed twice in PBS buffer, 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 FACSChantage Diva flow cytometer. 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 manufacture.

**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% OsO4 for 2 h. After dehydration by gradient concentrations of ethanol, 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 (TEM, FEI Tecnai 20) at 120 kV.

**Western blotting**—The harvested cells were disrupted by sonicaton in an ice-bath. Fractionation of the supernatant and precipitation proteins was performed by centrifugation of the sonicated cells for 10 min at 10000 rpm in a micro-centrifuge. 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 (Prod #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).

**RESULTS**

Recombinant β-galactosidase from *Aspergillus phoenicis* used as a 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 C12FDG, which can produce a green fluorescent product that is retained inside the cell (32). In this study, β-galactosidase from *Aspergillus 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, 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 designated times. The mRFP1 fluorescence increased very fast and almost reached the plateau at 4 hr, and the red fluorescence dynamic changes largely paralleled to the total expressed β-galactosidase-mRFP1 levels measured by Western blotting (Figure 1A and B). In contrast, β-galactosidase activity was still at a very low level at 4 hr and increased at later times (Figure 1A). These results suggest that the mRFP1 in the fusion protein folds fast largely independently of the β-galactosidase folding status and 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 (Figure 3) as well as in both inclusion bodies and the other cellular parts (Figure 5B, Figure 7A 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 of individual cells—After the induction for 2 hours, three dominant populations (R1, R2 and R3) appeared on a flow cytometric plot of side scattering light (SSC) versus forward scattering light (FSC) (Figure 2A). 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). SSC and FSC 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 hours after adding IPTG) with large cell sizes but little granularity (Figure 2A and B). The cells in the R2 region were small with little granularity (Figure 2A). 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 (Figure 2F). 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 while 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 (Figure 2C 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 (Figure 2D). A single peak with very high β-galactosidase activity was found in the R3 region cells (Figure 2C 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 hr or longer times according to their different green fluorescence intensities: a low-intensity peak and a high-intensity peak (Figure 2C). 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 (Figure 2B, bottom). 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. Since the cells in R2 region had almost no expression of the recombinant protein (Figure 2F) and the R3 region contained very few cells relative to those of the total population (Figure 2D), we will not 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 (Figure 3A). 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 Figure 3B clearly indicate that the high-β-galactosidase-activity protein was fully soluble, while 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. Since a fresh single colony of the bacteria was picked for each of the experiments in this study, the distinct statuses of the cells did not derived from a starting mixed populations.

**Effects of IPTG concentration on the ratio of the 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 (Figure 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 (Figure 4C), while 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 (Figure 4A 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 the ratio of the 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 lowly 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 hr, the cells were harvested by centrifugation and washed twice with fresh LB medium. Cells were re-cultured in pre-warmed conditioned LB medium which was collected from a bacterial culture without IPTG for 4 hours at 37 °C by centrifuging and removing E. coli. Cell samples were then collected at 2 hr interval. As shown in Figure 5B, 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 IPTG-free conditioned medium (Figure 5A and C). The preconditioned medium was used to keep everything else the same except IPTG for the experiments. We also replaced the old 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 hr, cells were either washed or unwashed, and then incubated for further 4 hr. The data showed that adding 10μg/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 (Figure 6). The “no effect” is not due to that tetracycline did not inhibit translation because adding tetracycline to cells at 0 hour completely blocked the translation of the recombinant protein (Figure 6).

**Effects of conditioned LB medium on the 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. Figure 3 demonstrates that the majority of the β-galactosidase-mRFP1 in the low activity cells was insoluble while 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 (TEM) demonstrated that there were mainly two types of the IPTG-induced unwashed cells regarding their internal protein distribution (Figure 7A). One of them contained small dense patches and the rest large area in these cells appeared dark. The dense patches are most likely to be IBs as reported previously (35,36) while 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 much less-dense patches in much larger area in 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 hr 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 (Figure 7 A). The randomly picked “feature-free areas” in the unwashed and washed cells were amplified 6 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 (Figure 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% to 91.2% of the total cells at 10 hr) after the removal of IPTG (Figure 7C, white bar). 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 hr (Figure 5A 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 hr) 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 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.

4. Discussion

In this study three dominant groups of an E. coli cell population overexpressing β-galactosidase-mRFP1 fusion protein were identified in SSC vs FSC plots (Figure 2A and B). Red fluorescence was shown to be able to measure the fusion protein levels in individual cells (Figure 1) and green fluorescence to distinguish the individual cells bearing the highly active and soluble β-galactosidase from those containing the lowly active and insoluble β-galactosidase (Figure 2 and 3). The average enzyme activities of the high-activity enzyme-containing cells were 13.1 to 17.5 folds of those of the low-activity enzyme-containing cells from 4 to 10 hr after IPTG induction (Supplemental table 1). This capability of simultaneously measuring multiple parameters in individual E. coli cells by flow cytometry rendered 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 lowly active β-galactosidase proteins. Further analyses using flow cytometric sorting and Western blotting indicates that the highly active enzyme-containing cells carried only soluble form of the recombinant protein while the lowly active enzyme-containing cells carried mainly insoluble form of this protein (Figure 3). TEM (Transmission Electronic Microscopy) analysis demonstrates that there were three types of cells that were induced by IPTG to overexpress the recombinant protein (Figure 7A, top row). In the first type of cells very dense patch(s) (IBs) was located in very small area(s) and darkness filled in the rest 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 the area inside the cells. We infer that the first and second types of cells were respectively the lowly and highly active enzyme-containing cells, and the third type of cells was those that 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 TEM 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 while the dark IB-bearing cells are the lowly active enzyme-containing cells. After IPTG withdraw and further culturing for another 6 hr, the second type of cells, including IB-free and IB-bearing cells (Figure 7C), 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 (Figure 7). This pair of data strongly supports the above supposition on the cell classification. We designate the dark and IB-bearing cells as DIB cells, and the loose white-material-dominated cells as LW cells (Figure 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 \( \beta \)-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 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 \( \beta \)-galactosidase-mRFP1 protein than R1 cells (Figure 2F) and the average enzyme activity in R3 cells was much higher than that of R1 cells (Figure 2E), suggesting that more protein in R3 folded into the final active form than that in R1 cells. Actually, almost all the cells in R3 group had the average enzyme activity comparable to 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 \( \beta \)-galactosidase-mRFP1) one subpopulation (LW) contained soluble and highly active form of \( \beta \)-galactosidase-mRFP1 while the other subpopulation (DIB) contained insoluble and lowly active form of \( \beta \)-galactosidase-mRFP1, although the concentrations of \( \beta \)-galactosidase-mRFP1 in individual cells of the two subpopulations were similar (Figure 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 (Figure 5 and 7). Third, the addition of tetracycline to cell culture at 4 hr (to inhibit IPTG-induced protein synthesis) did not change the percentage of the LW cells and the \( \beta \)-galactosidase activity of R1 cells at 8 hr whether IPTG was removed or not from the culture (Figure 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 the 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 \( \beta \)-galactosidase activity of R1 cells through changing the ratio of the high-activity (LW) and low-activity (DIB) cells (Figure 4). Second, removal of IPTG from the culture increased the ratio of soluble protein over IBs through the increase of LW cell percentage (Figure 5 and 7). We also found that OD 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 while 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 IBs-containing and IBs-free cells though they expressed β-galactosidase-mRFP1 protein at equally high levels (Figure 3), and that replacing old IPTG-containing medium with conditional IPTG free medium dramatically enhanced the percentage of the high enzyme activity population while the fusion protein levels were not changed (Figure 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 4 other heterozygous proteins before and after IPTG removal (Supplemental Fig. S5 and Supplemental table S2). The soluble portions of the two proteins tBgl1 and tBgl3 were increased compared to 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 pre-induced by IPTG can increase the ratio of soluble protein on some other proteins besides the β-galactosidase from Aspergillus phoenicis, and that there might be the two-state distribution E. coli expressing some other proteins in addition to β-galactosidase from Aspergillus 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 Aspergillus phoenicis used in this study.
REFERENCES


FOOTNOTES

1. This investigation was supported by grant 2011CBA00906 (to H.-Y. Hang) from the National Key Basic Research Program of China (973) of Ministry of Science and Technology and 30970559 (to Y.-B. Yan) from the National Natural Science Foundation of China.

2. The abbreviations used are: DTT, dithiothreitol; FDG, fluorescein-di-β-D-galactopyranoside; IB, inclusion body; ITPG, Isopropyl-1-thio-β-D-galactopyranoside; PETG, phenylathylthio β-D-galactoside; RFP, red fluorescence protein.

FIGURE LEGENDS

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 induced for 2, 4, 6, 8 and 10 hours 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.

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 hr. The cell size (reflected by FSC) and granularity (reflected by SSC) 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 hr (See the B section in this figure). B. R1 region cells. The recombinant cells were induced by 0.5 mM IPTG at 37°C for designated times. Three dominant cell populations (R1, R2 and R3) in dot plots were characterized (top); RFP fluorescence intensities vs β-galactosidase activities in R1 region were shown in dot plots (middle); RFP fluorescence intensities in R1 regions were demonstrated in histogram plots (bottom). 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 of this figure. 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.

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 hr 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 number of both cell population were loaded for Western blotting analysis.

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. Cells were cultured for 5.5 hr.

FIGURE 5. Effects of IPTG removal after being induced for 4 hr on the percentages of two β-galactosidase activity populations in the R1. A. The β-galactosidase activity distribution without IPTG removal (top) or when IPTG was removed (bottom). 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 (10μg/ml) was added to the medium to inhibit protein translation. After induction for 4 hr, IPTG was either washed or reserved for further 4 hr induction. The percentages of high-activity populations and enzyme activities were determined by flow cytometry detailed in Materials and Methods.

FIGURE 7. Effects of IPTG removal on the percentages of non-inclusion-body cells identified by transmission electron microscopy. A. The non-induced and IPTG-induced cells as well as the cells first induced with IPTG for 4 hr and then cultured in IPTG-free medium for designated times were 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 hr and then cultured in IPTG-free medium. The featureless areas were purposely selected (for example area 1 and 2) from A of this figure and amplified 6 times. White materials were clearly notable in the featureless areas of washed cells. C. The percentages of the cells with the following features: (a) IB-containing accompanied with darkness in the rest large area, (b) IB-bearing and loose white material-dominated and (c) IB-free and loose white material-dominated. 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 IBs-containing).
Figure 1

A

![Graph showing β-galactosidase activity and RFP intensity over time.]

B

![A blot showing protein bands at different time points.]

0 2 4 6 8 10 h

0.00 0.54 0.76 0.93 0.81 1.00

±0.13 ±0.05 ±0.11 ±0.09
Figure 2
Figure 3

A

B

supematant
precipitation
supematant
precipitation

high-activity population
low-activity population
Figure 4

A

Cell number

β-galactosidase activity

0 mM 0.1 mM 0.3 mM 0.5 mM 0.8 mM 1.0 mM 2.0 mM IPTG

B

Cell subpopulation (%)

- low activity
- high-activity

[ IPTG ] (mM)

C

RFP intensity

[ IPTG ] (mM)
Figure 5

A

![Graph showing cell number over time with β-galactosidase activity](image)

B

![Graph showing RFP intensity over time with 'Washed' and 'unwashed' symbols](image)

C

![Graph showing high activity cell population over time with 'washed' and 'unwashed' symbols](image)
Figure 7

A

non-induced control  6 h  8 h  10 h

unwashed

washed

B

unwashed

washed

C

Cell subpopulation (%) vs. Time (h)

4  6  8  10

U  W  U  W  U  W
Figure 8

low activity
insoluble IBs only

high activity
insoluble IBs
soluble protein

high activity
soluble protein only
Two distinct states of Escherichia coli cells that overexpress the recombinant heterogeneous β-galactosidase

Yun Zhao, Wei He, Wei-Feng Liu, Chun-Chun Liu, Li-Kui Feng, Lei Sun, Yong-Bin Yan and Hai-Ying Hang

J. Biol. Chem. published online February 2, 2012

Access the most updated version of this article at doi: 10.1074/jbc.M111.327668

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2012/02/02/M111.327668.DC1

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2012/02/02/jbc.M111.327668.full.html#ref-list-1