Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2012, **8**, 3325–3334

www.rsc.org/molecularbiosystems

PAPER

Measuring the dynamics of *E. coli* ribosome biogenesis using pulse-labeling and quantitative mass spectrometry†

Stephen S. Chen,[a] Edit Sperling,[a] Josh M. Silverman,[a] Joseph H. Davis[a] and James R. Williamson*[ab]

Received 4th August 2012, Accepted 8th October 2012
DOI: 10.1039/c2mb25310k

The ribosome is an essential organelle responsible for cellular protein synthesis. Until recently, the study of ribosome assembly has been largely limited to *in vitro* assays, with few attempts to reconcile these results with the more complex ribosome biogenesis process inside the living cell. Here, we characterize the ribosome synthesis and assembly pathway for each of the *E. coli* ribosomal protein (r-protein) *in vivo* using a stable isotope pulse-labeling timecourse. Isotope incorporation into assembled ribosomes was measured by quantitative mass spectrometry (qMS) and fit using steady-state flux models. Most r-proteins exhibit precursor pools ranging in size from 0% to 7% of completed ribosomes, and the sizes of these individual r-protein pools correlate well with the order of r-protein binding *in vitro*. Additionally, we observe anomalously large precursor pools for specific r-proteins with known extra-ribosomal functions, as well as three r-proteins that apparently turnover during steady-state growth. Taken together, this highly precise, time-dependent proteomic qMS approach should prove useful in future studies of ribosome biogenesis and could be easily extended to explore other complex biological processes in a cellular context.

Introduction

Understanding the dynamics of the proteome is currently at the forefront of systems biology, as changes in protein synthesis and degradation rates are required for cellular responses to perturbations in the environment and for progression through developmental programs.[1–3] Historically, pulse-chase experiments using radioactive labeling have been used to study proteome flux, but this approach is limited by low sample-throughput and poor quantitation precision. Recently, however, time-dependent proteomics using quantitative mass spectrometry has emerged as a more facile and accurate alternative to studying proteome flux in a variety of biological systems.[3]

One such system, the ribosome translation apparatus, is central to the regulation of cell growth and differentiation, and, as a major constituent of the proteome, deserves special attention for understanding proteome dynamics. Ribosomal components make up as much as 50% of the proteome by weight in *E. coli*,[4] and substantial resources are dedicated to their production. As the central hub of cell physiology, the ribosome is a key target for small molecule antibiotics.[5–7] Furthermore, aberrant ribosome production has been linked to human diseases such as carcinogenesis[8,9] and Diamond Blackfan Anemia.[10] Clearly, establishing the *in vivo* dynamics and regulation of ribosome biogenesis will greatly advance our understanding of the cellular economy.

The *E. coli* ribosome consists of the small 30S and large 50S subunits, which is composed of three large ribosomal RNA molecules (5S, 16S, and 23S rRNA) and 54 structural ribosomal proteins (r-proteins). The *in vitro* assembly process for the two subunits has been investigated extensively in a series of reconstitution experiments by the Nomura and Nierhaus laboratories,[11,12] resulting in two canonical ribosome assembly maps that depict the thermodynamic binding dependencies of the *E. coli* r-proteins. More recently, 30S subunit assembly kinetics were investigated using a series of novel biophysical techniques including hydroxyl radical footprinting,[13] stable isotope pulse-chase mass spectrometry,[14,15] and time-resolved electron microscopy.[16] Together, these studies suggest that the assembly of the 30S particle proceeds in a 5′-to-3′ direction along the 16S rRNA, and also follows the same top-to-bottom binding hierarchy found in the Nomura assembly map.[17]

Understanding how the well-characterized *in vitro* assembly pathway relates to the process of ribosome biogenesis inside living cells is a major goal in the ribosome field. In *in vitro* reconstitution experiments, ribosome assembly involves the synchronous binding of r-proteins to mature full-length rRNAs, whereas ribosome biogenesis inside cells involves the asynchronous, co-transcriptional binding of r-proteins to nascent rRNA precursors. In addition,

[a] *Department of Molecular Biology, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA*
[b] *Department of Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. E-mail: jrwill@scripps.edu; Fax: +1 858 784 2199; Tel: +1 858 784 8740*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb25310k

ribosome assembly *in vivo* requires the participation of many assembly cofactors and chaperones that modify the rRNA and facilitate RNA folding and r-protein binding.[18,19] Finally, the process of ribosome biogenesis is strongly coupled to cell growth, which involves transcriptional and translational regulation networks governed by global signaling molecules such as ppGpp.[20,21] Through these complex regulatory networks, cells maintain constant pools of free r-proteins and assembly intermediates to meet the demand for new ribosomes during exponential growth.

Radioisotope pulse-chase experiments were first applied to measure these precursor pools in order to understand the dynamics of *E. coli* ribosome assembly over forty years ago.[22–28] Measurements of total free r-proteins and assembly intermediates were estimated to be less than 5% of total ribosomes,[22,23] indicating that the *in vivo* ribosome assembly process is rapid and highly efficient. In conjunction with 2D-gel electrophoretic separation of the r-proteins, these radiolabeling experiments provided estimates of the amounts of individual r-proteins in ribosome assembly intermediates and assigned tentative orders of binding to the 30S and 50S r-proteins *in vivo*.[24,28] However, the data in these classic studies were limited in measurement precision and accuracy, and the resulting pools were difficult to reconcile with the emerging views of ribosome assembly from current *in vitro* studies. More recently, changes in total r-protein levels in the cell upon perturbation of the r-protein auto-regulatory network were characterized using a high-precision quantitative mass spectrometry (qMS) approach.[29] Although gross differences in total r-protein were readily observed upon perturbation, the high complexity of these whole-cell samples resulted in relatively large measurement errors of 7%, and any fine differences in the r-protein levels from the relatively small amounts of wild-type ribosome assembly intermediates were not detected.

Here, we present a next-generation stable isotope pulse-labeling study of the ribosome biogenesis system, as characterized by qMS and mathematical frameworks of pulse-labeling kinetics. Purification of completed ribosomal particles and an eight-point pulse-labeling timecourse allowed for precise measurements of the labeling kinetics for 53 of the 54 *E. coli* r-proteins under steady-state cell growth. Fitting this data to mathematical models of ribosome biogenesis produced precise values for individual r-protein precursor pool sizes encompassing both free proteins and ribosome assembly intermediates. These measurements cover the flux of a third of the bacterial proteome by mass, and the measured pool sizes provide a clear and consistent picture of the *in vivo* ribosome assembly landscape. In addition, anomalous pool sizes and significant turnover rates were found for a number of r-proteins with both known and possibly as-yet-uncharacterized extra-ribosomal functions. The qMS pulse-labeling method described herein is of general applicability and should find widespread use in the quantitative study of protein synthesis and turnover in a variety of biological systems.

Results

Stable isotope pulse-labeling kinetics in *E. coli*

An overview of the *in vivo* pulse-labeling qMS approach is illustrated in Fig. 1. Briefly, *E. coli* strain MRE600 was grown at 37 °C in M9 minimal medium containing ^{15}N as the sole nitrogen source. Early in exponential phase growth ($OD_{600} \approx 0.5$), an equal volume of ^{14}N medium was added to the culture, and samples were taken at eight time points spanning approximately one doubling period after the isotope pulse. A constant doubling time of 36 minutes was maintained throughout the pulse-labeling experiment. To improve downstream ribosomal peptide identification,[15] the pulse-labeled cells were mixed with reference cells grown in ^{14}N-only medium. The ribosomal particles were isolated from lysed cells using ultra-centrifugation on a non-dissociating sucrose gradient, and 70S ribosomes were purified, digested with trypsin and submitted to LC-MS analysis. The m/z trace for each peptide was extracted from raw MS data and the fully ^{14}N-labeled, 50% ^{15}N-labeled, and fully ^{15}N-labeled isotope distributions were quantified using a Least Squares Fourier Transform convolution algorithm.[30] Fits of the isotope distributions for each peptide were filtered for quality (see Experimental procedures).

The labeling kinetics of each r-protein is defined by the accumulation of 50% ^{15}N-labeled proteins in completed 70S ribosomes as a function of time. Specifically, for each peptide, the fraction-labeled value in 70S ribosomes at pulse time t is given by $f_r(t)$, the ratio of 50% ^{15}N-labeled peptide to total peptide (*i.e.* fully ^{15}N-labeled + 50% ^{15}N-labeled peptides) from the pulse-labeled sample (Fig. 1). The peptide $f_r(t)$ values were averaged for each r-protein except L36 in the timecourse experiment, for which no good isotope fits were found (Fig. 2). As expected, a monotonic increase in the fraction-labeled values was observed for each r-protein as a function of pulse time. Additionally, at each time-point significant variations were observed in the fraction-labeled values between individual r-proteins, and many r-proteins were consistently underlabeled as compared to a calculated maximum labeling value, $f_{max}(t)$ (see Experimental procedures). Interestingly, a few proteins were consistently over-labeled, implying a synthesis rate in excess of that necessary for ribosome synthesis. These specific differences in labeling kinetics reflect differences in the dynamics of the individual proteins in the ribosome biogenesis pathway.

Qualitatively, these results were interpreted using a biological model for ribosome biogenesis (Fig. 3). In this model, newly synthesized r-protein is released into the free protein pool and mixes with pre-existing free proteins. These r-proteins, some of which may take part in extra-ribosomal functions, are incorporated into ribosome assembly intermediates in an ordered assembly process. As such, earlier binding r-proteins would be incorporated into completed ribosomes after those that bind later. In this context, a lag in the observed labeling kinetics relative to $f_{max}(t)$ is attributed to the presence of a precursor pool for a particular r-protein, which is composed of both free proteins and downstream ribosome assembly intermediates containing that r-protein. Some r-proteins were labeled faster than $f_{max}(t)$, and these were interpreted by a more detailed model involving exchange and turnover, as described below. Quantitatively, the precursor pool sizes and turnover rates for each r-protein were obtained by fitting the observed labeling kinetics to mathematical models for ribosome biogenesis.

A quantitative precursor pool model to describe pulse-labeling kinetics

The *E. coli* ribosome biogenesis pathway was modeled as a series of three pools through which isotope labels flow.

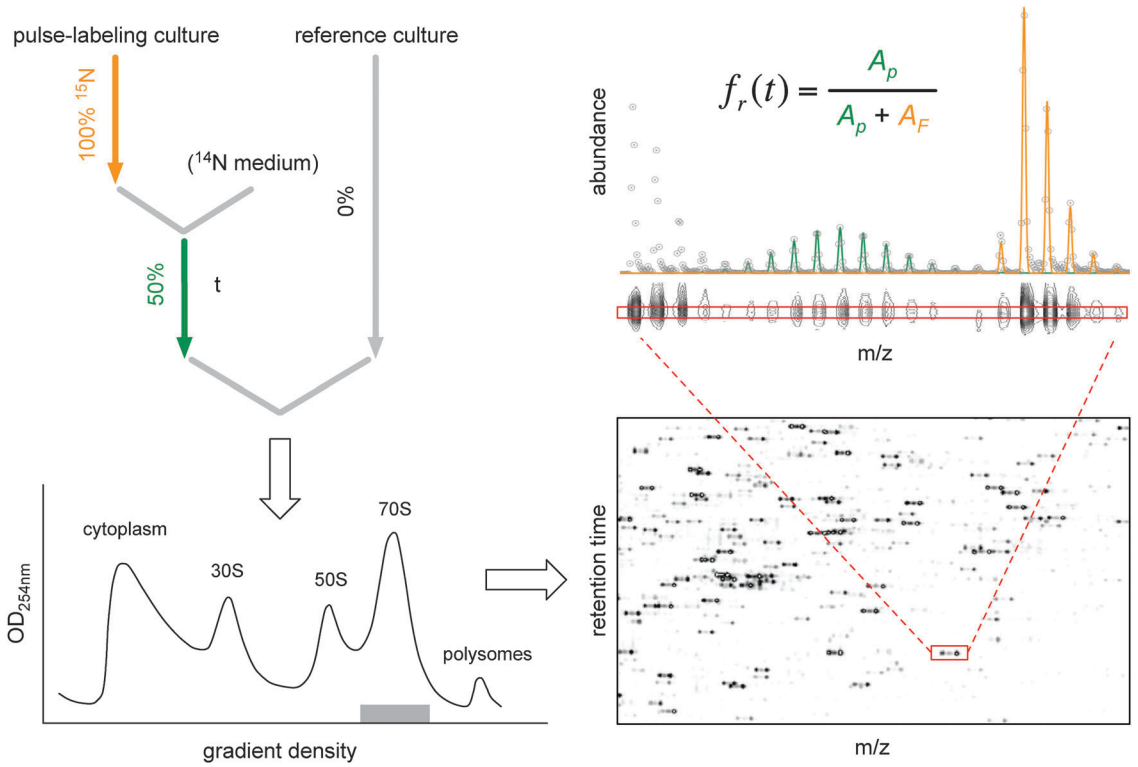

Fig. 1 Overview of the pulse-labeling and qMS approach. Cells growing exponentially in ^{15}N medium were pulsed with an equal volume of ^{14}N medium for time t. Pulsed cells were harvested and mixed with a reference culture grown entirely in ^{14}N medium. Ribosomes were purified *via* ultracentrifugation through a sucrose gradient, and fractions containing completed 70S ribosomes (gray bar) were combined and submitted to LC-MS analysis. The isotope distribution for each ribosomal peptide was extracted. Averaged abundance values from the raw data (dotted open circles) were fitted using a Least Squares Fourier Transform Convolution algorithm;[30] only fits to the partially labeled (green line) and fully labeled species (orange line) are shown. The fraction-labeled value $f_r(t)$ is calculated as the fraction of the amplitude of partially labeled peptides, A_P, relative to the sum of the amplitudes of the partially labeled and fully labeled peptides, A_P and A_F, respectively.

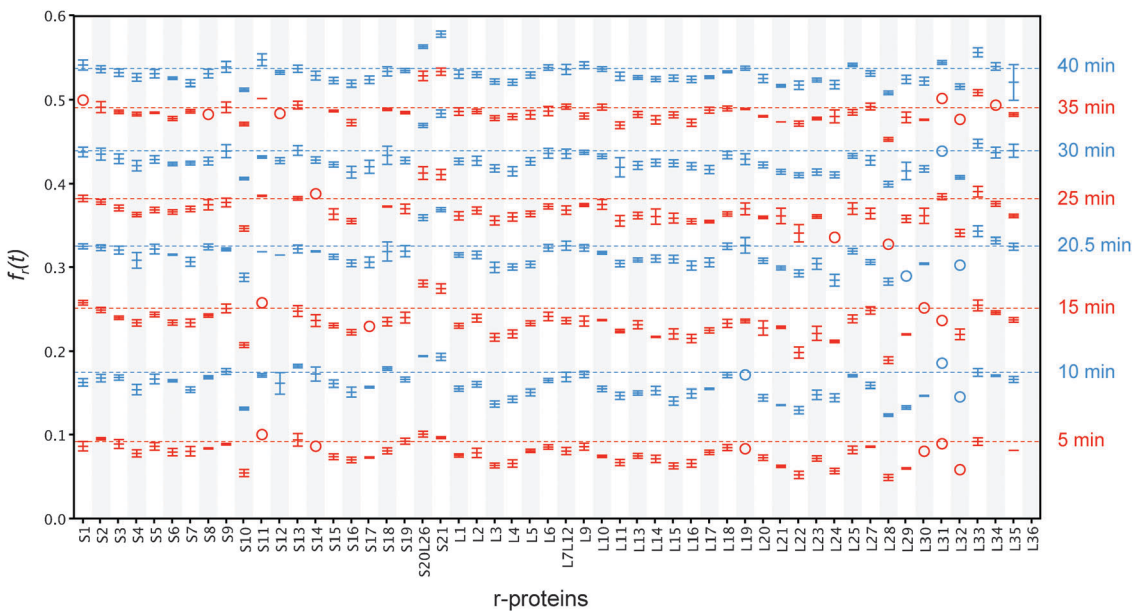

Fig. 2 Timecourse of 70S r-protein labeling kinetics. The average and standard deviation of the measured $f_r(t)$ values for all identified peptides for each r-protein is shown in alternating colors across the 8 pulse labeling time points. Proteins for which a single peptide was observed are represented by open circles, and the maximum labeling value at each pulse time, $f_{max}(t)$, is indicated by the dotted lines.

The first pool, n, consists of nutrients and unincorporated amino acids that flow with a constant flux ν_p into the second pool, p, which is composed of free r-proteins and r-proteins bound to ribosome assembly intermediates. These intermediates in turn

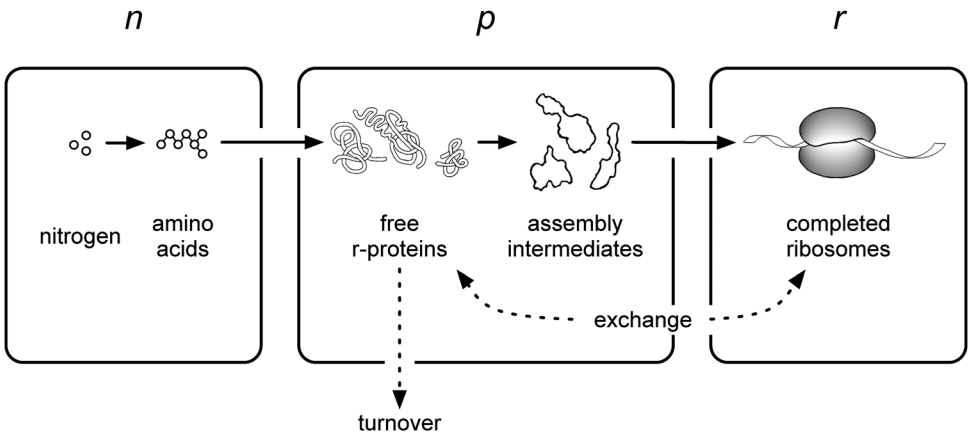

Fig. 3 Model of the ribosome biogenesis pathway in *E. coli*. Label is modeled to flow through three distinct compartments: the nutrients pool *n*, where nitrogen is taken up by the cell and incorporated into amino acids; the precursor pool *p*, where newly labeled r-protein enters a standing free pool involved in ribosome assembly, as well as extra-ribosomal functions; and the completed ribosomes *r*, where completed 30S and 50S subunits enter the translation cycle. Possible additional fluxes from pools *p* and *r* are indicated by dotted lines.

flow with flux ν_r into the third pool, *r*, which contains r-proteins incorporated in completed ribosomes:

$$n \xrightarrow{\nu_p} p \xrightarrow{\nu_r} r$$

The fraction-labeled value of each pool is given below, where x represents the total amount of material in a given pool and x^* represents the amount of material in that pool that is pulse-labeled:

$$f_x = \frac{x^*}{x} \tag{1}$$

It is assumed that the labeling dynamics of pool *n* are fast,[31] and that any r-proteins synthesized after the isotope pulse are immediately labeled, such that $f_n(t) = 1$ for all pulse times t, and that $f_p(t) = f_r(t) = 0$ at $t = 0$. The steady-state flux models of isotope flow outlined herein are derived based on the assumption of balanced growth, *i.e.* the net rate of synthesis for any pool x is $dx/dt = k \cdot x$, where k is the growth rate. The flux into a pool early in the pathway must support the growth of that pool as well as the growth of all subsequent pools. Assuming for now that no r-proteins leave completed ribosomes, the flux of material into the final pool r needs only to support the growth of the ribosome pool. The following equations relate the synthesis rates of pools *p* and *r* with the growth rate and sizes of these pools:

$$\nu_r = k \cdot r \tag{2}$$

$$\nu_p = k \cdot (p + r) \tag{3}$$

The set of equations describing dp^*/dt and dr^*/dt is then the difference between the synthesis and utilization rates, multiplied by the fractional labeling value (with $f_n = 1$) of their respective source pools:

$$\frac{dp^*}{dt} = \nu_p - \nu_r \cdot f_p \tag{4}$$

$$\frac{dr^*}{dt} = \nu_r \cdot f_p \tag{5}$$

These expressions can be used to obtain an integrable set of differential equations for the fraction-labeled values f_x using the quotient rule:

$$\frac{df_x}{dt} = \frac{d}{dt}\left(\frac{x^*}{x}\right) = \frac{1}{x^2} \cdot \left(x \cdot \frac{dx^*}{dt} - \frac{dx}{dt} \cdot x^*\right) = \frac{1}{x} \cdot \left(\frac{dx^*}{dt} - f_x \cdot \frac{dx}{dt}\right) \tag{6}$$

To solve for the rates of change in the fractional labeling values df_p/dt and df_r/dt, eqn (2)–(5) can be substituted into eqn (6) to arrive at:

$$\frac{df_p}{dt} = k \cdot (1 - f_p) \cdot \left(1 + \frac{1}{\mathbf{P}}\right) \tag{7}$$

$$\frac{df_r}{dt} = k \cdot (f_p - f_r) \tag{8}$$

where the relative precursor pool size $\mathbf{P} = p/r$ is given as the ratio of the precursor pool to the completed ribosome pool. Integration of this system of equations leads to a closed-form solution for the fractional labeling of ribosomes as a function of time:

$$f_r(t) = 1 + \mathbf{P} \cdot \exp\left[-k \cdot \left(1 + \frac{1}{\mathbf{P}}\right) \cdot t\right] - (1 + \mathbf{P}) \cdot \exp[-k \cdot t] \tag{9}$$

Thus, the relative precursor pool sizes can be obtained by fitting the experimental pulse-labeling timecourse in Fig. 2 to eqn (9), where **P** is the only free parameter.

The majority of r-protein labeling kinetics fit well to the precursor pool model derived above, and significant variations were observed in the **P** values of individual r-proteins (Table S1, ESI†). For example, some r-proteins such as S2, S11, and L34 exhibited virtually no lag in labeling (Fig. 2) and fit to a precursor pool size that is effectively zero. In contrast, primary binding r-proteins such as S4, S6 (Fig. 4a), S7, S15, and S17 exhibited a significant lag in their labeling kinetics, reflecting pool sizes between 2% to 3%, and remarkably large pool sizes of 6–7% were measured for the r-proteins S10 (Fig. 4b) and L28.

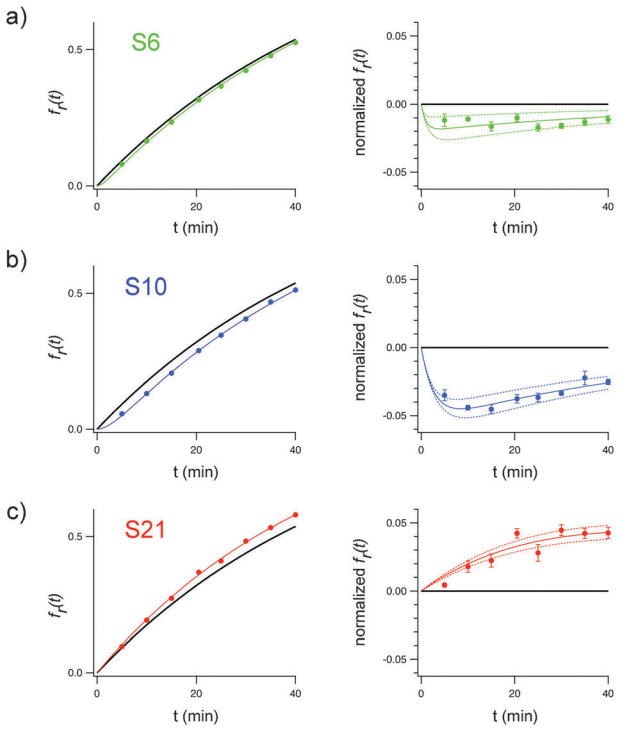

Fig. 4 Fits of select observed r-protein labeling kinetics using biological models of ribosome biogenesis. (a, b, c) Left, observed labeling kinetics $f_r(t)$ for r-protein S6, S10, and S21 (green, blue, and red circles, respectively) plotted against $f_{max}(t)$ (black line) and calculated labeling kinetics of the best-fit values of **P** = 2.0%, **P** = 5.6%, and turnover per generation of 8.4% (green, blue, and red lines, respectively). Right, observed labeling values and labeling kinetics normalized against $f_{max}(t)$ (same as left); additional calculated labeling kinetics for ±1% of the calculated fit are shown as dotted lines.

As a caveat, the precursor pool model may underestimate the actual precursor pool if there is some exchange of r-proteins between free proteins and complete ribosomes, *e.g.* due to the replacement of easily damaged r-proteins.[32] Mathematically, the r-protein labeling kinetics for smaller pools with no exchange (Fig. S1a, ESI†) is indistinguishable from that of a larger precursor pool with exchange (Fig. S1b, ESI†). Thus, given that a few r-proteins are known to exchange *in vivo*,[32–34] it is possible that in actuality these r-proteins have significantly larger precursor pools. Although our current experimental approach cannot distinguish between various rates of exchange, the derivation of the appropriate mathematical model for exchange between the precursor pool p and completed ribosome pool r is presented (see ESI†).

To confirm that the **P** values measured above accurately reflect the r-protein precursor pool, the free pool of the small subunit protein S9 was artificially increased by introducing a IPTG-inducible plasmid bearing a copy of this protein into the BL21 (DE3) Tuner strain. S9 levels were titrated by addition of 0 μM, 5 μM, or 15 μM IPTG. Upon induction, increases in total S9 total cellular protein of up to 66% were observed in the whole cell lysate relative to a 70S spike[29] (Fig. S2a, ESI†). Correspondingly, *in vivo* stable isotope pulse-labeling exhibited a specific delay in the labeling kinetics of S9 in completed ribosomes (Fig. S2b, ESI†), and there is a good correlation between the amount of extra S9 measured in whole cell lysates and the calculated precursor pool sizes (Fig. S2c, ESI†). Critically, this correlation between the increases in total S9 protein and its precursor pool size shows that the parameter **P** accurately reflects the *in vivo* precursor pools of *E. coli* r-proteins.

To eliminate the possibility of contamination by aggregates due to recombinant protein expression, 70S ribosomal particles were purified through a sucrose cushion and dissociated into 30S and 50S subunits in the S9 over-expression experiments. It is curious to note that here, in the dissociated 30S subunits, r-protein S20 displayed significantly slower labeling kinetics corresponding to an apparent precursor pool of ~15% (Fig. S2b, ESI†). The mechanism for this anomalous pulse-labeling result remains to be elucidated, but interference from possible extra-ribosomal copies of S20 associated with the 50S subunit[35,36] (data not shown) is briefly discussed (see ESI†).

An exchange-plus-turnover model to describe over-labeling

While the precursor pool model accurately described the labeling kinetics of most r-proteins, this model could not account for the observed labeling kinetics of r-proteins S20, S21 (Fig. 4c), and L33, which were each over-labeled relative to the theoretical maximum. To produce such over-labeling, the synthesis rate of an r-protein must exceed the rate required solely for pool replacement and ribosome synthesis. Under balanced growth, a counter-acting efflux must exist to compensate for this increase. These requirements can be satisfied by a mechanism in which specific r-proteins exchange from completed ribosome into the free protein pool[32–34] and are then degraded or sequestered from the ribosome biogenesis pathway. At biologically-relevant rates of exchange, the precursor pool p and completed ribosome pool r effectively merge into a single pool r'. Turnover of free proteins would then result in the over-labeling of the r-proteins in completed ribosomes (see ESI†). Specifically, this exchange-plus-turnover model is represented as:

$$n \xrightarrow{\nu_{r'}} r' \xrightarrow{d_{r'}}$$

where $r' = p + r$. The fluxes in and out of r' are given by:

$$\nu_{r'} = k \cdot r' + d' \cdot r' \quad (10)$$

$$d_{r'} = d' \cdot r' \quad (11)$$

where d' is the degradation rate of r'. The change in the amount of label in r' is then:

$$\frac{dr'^*}{dt} = \nu_{r'} - d_{r'} \cdot f_{r'} \quad (12)$$

By integrating eqn (12) and substituting in eqn (10) and (11), the closed form solution for $f_{r'}(t)$ can be written as:

$$f_{r'}(t) = 1 - \exp[-(k + d') \cdot t] \quad (13)$$

Thus, for the exchange-plus-turnover model the observed labeling kinetics $f_{r'}(t)$ is solely determined by the growth rate k and the turnover rate d'. As shown for S21 (Fig. 4c), the exchange-plus-turnover model describes the over-labeling kinetics

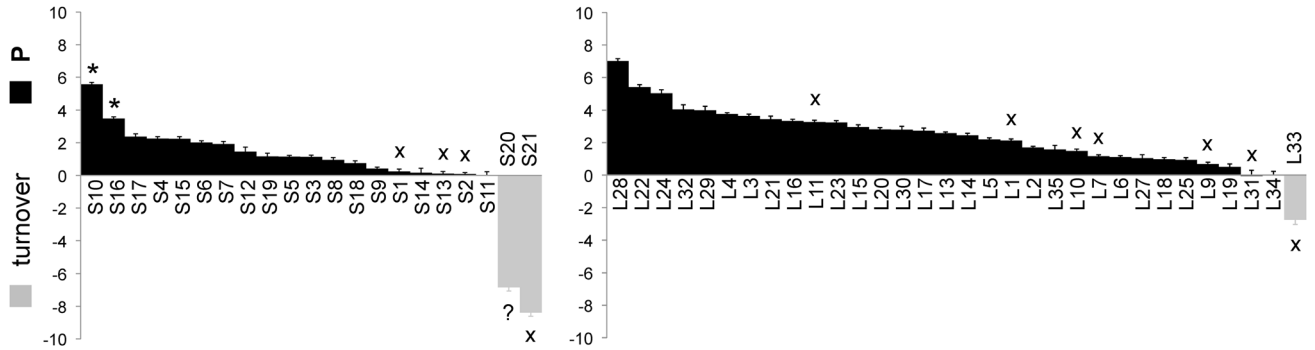

Fig. 5 Calculated relative precursor pool size and turnover rates for *E. coli* r-proteins. The relative pool size value **P** is defined as the % of the precursor pool *p* relative to the completed ribosome pool *r*. The turnover rate is expressed as % turnover per generation. Notable r-proteins are marked accordingly: those with extra-ribosomal functions (*),[41–43] those that exchange between completed ribosomes and a free proteins pool (x),[32–34] and S20 which has a questionable stoichiometry (?).[35,36]

observed for S20, S21 and L33 significantly better than the precursor pool model (Table S1, ESI†).

Using these two models, we calculated the relevant biological parameters for each of the r-proteins in *E. coli* (Fig. 5). Measured **P** values and turnover rates per generation (see Experimental procedures for definition) ranged from 0–7% and 3–8%, respectively. As discussed below, this pulse-labeling strategy offers a powerful tool to investigate the dynamics of the free protein and assembly intermediates pool during ribosome biogenesis.

Discussions

This work describes, in detail, the process of ribosome biogenesis in exponentially growing *E. coli* cells using a novel, high-precision qMS stable isotope pulse-labeling approach. The average measured **P** value of 1–2% for all r-proteins is similar to previous measurements of total ribosomal precursors.[22,23,28,37–39] However, detailed inspection of the individual precursor pools reported here reveals some significant discrepancies with those found in previous studies.[27,40] Below, we compare the accuracy and precision in a previously published work to the current data. The values reported here represent a dramatic increase in precision over previous studies.[24,25,28]

A radioisotope pulse-labeling measurement of the individual r-protein precursor pools was previously performed in which r-protein spots were excised from 2D-gels and quantitated *via* scintillation counting.[25] One metric of measurement precision is the period of time for which significant labeling differences can be quantified. In their study, differences in the r-protein pulse-labeling values did not persist longer than one-sixth of a doubling time.[25] In comparison, in the present study significant differences are still observed after 1.1 doubling times, which translate into a 7-fold increase in precision over the previous radiolabeling approach.

There are several factors that contribute to this large improvement in precision. First, our measurements were performed on intact ribosomes obtained directly from sucrose gradients, thereby eliminating sample loss that often results from 2D-gel separation and excision. Second, the entire isotope distribution for the unlabeled and labeled species of each peptide was fit using a Least Squares Fourier Transform Convolution algorithm,[30] poor fits due to noise or peak overlaps were carefully excluded from further analysis, and multiple peptide quantitation values were averaged together to produce a much more precise estimate of r-protein labeling than from the quantitation of a single scintillation counting value. Third, pool size parameters were obtained by a least squares fit of $f_{obs}(t)$ across an eight-point pulse-labeling timecourse (Fig. 2), rather than from just a single timepoint. Together, these methodological improvements allowed us to quantitate the labeling differences of an *in vivo* pulse-labeling experiment at unprecedented levels of precision. The magnitudes of the pools between the two data sets are in general agreement, and differences between these two studies likely result from the improved precision of our qMS approach.

To demonstrate that the observed labeling kinetics and measured **P** values accurately reflect the precursor pools of r-proteins, S9 was over-expressed at various levels using an expression plasmid with an inducible promoter. Initial sucrose gradient purification of pulse-labeled 70S ribosomes showed the presence of contaminating S9 aggregates, and an alternative purification scheme was used to obtain pure 30S subunits for pulse-labeling studies (see Experimental procedures). A good correlation of the increase in total S9 protein and precursor pool size was observed under 0 μM (likely due to leaky expression), 5 μM, and 15 μM IPTG induction (Fig. S2c, ESI†).

For 50 of the 54 r-proteins, the observed fraction labeling value, $f_r(t)$, was at or below the calculated maximum labeling value, $f_{max}(t)$, which indicates finite precursor pools for these r-proteins (Fig. 2). Most of the r-proteins exhibited a significant lag in labeling and fit to pool sizes as large as 7%, implying that a significant fraction must either be sequestered upstream of the completed ribosomes as ribosomal assembly intermediates (*i.e.* 16S rRNA with an incomplete subset of the 30S r-proteins), as a free cytosolic pool, or, alternatively, take part in non-ribosomal functions orthogonal to the assembly process. A number of the r-proteins (*i.e.* S1, S2, S11, S13, S14, L31, and L34) exhibited no lag in their labeling kinetics, resulting in fitted pool sizes of effectively zero. This result implies that either these r-proteins are rapidly incorporated into completed 70S ribosomes upon protein synthesis, or as noted in previous studies, they may reversibly bind to

the ribosome and then exchange with free protein pools *in vivo*.[32–34]

Because the r-protein precursor pools include ribosome assembly intermediates, we expect that proteins that bind earlier during assembly would exhibit larger pools than those that bind later (Fig. 6a). To test this hypothesis, the **P** value for each r-protein was painted onto the Nomura and Nierhaus *in vitro* assembly maps (Fig. 6b and c, respectively). As expected, the observed **P** values decrease in a 5′-to-3′ direction along the rRNA and from top to bottom in the assembly hierarchies for both subunits. Aside from significant exceptions noted below, these trends agree with both the thermodynamic[11,12] and kinetic[14,15] binding dependencies observed *in vitro*. Moreover, this result is generally in agreement with prior work by Pichon *et al.*[25] that placed 30S and 50S r-proteins into early and late *in vivo* binding groups. Taken together, this is the clearest confirmation to date that the 30S and 50S assembly landscapes observed *in vitro* accurately reflect the complex process of ribosome biogenesis inside the cell.

A small number of r-proteins (*i.e.* S10, S16, L28, and L32) have unexpectedly large **P** values that are at odds with the assembly order expected from *in vitro* studies (Fig. 6b and c). For instance, the 30S tertiary binding r-protein S10 exhibited a **P** value of 5.6%, or roughly three times the average pool size for the primary binding r-proteins S4, S7, S15, and S17 (Fig. 5). Based strictly on the *in vitro* order of r-protein binding, one would expect the fraction of S10 present in assembly intermediates to be much less than that of the 30S primary binders. However, it is also known that S10 functions as a transcriptional elongation factor by binding directly to RNA polymerase.[41,42] Exchange between the extra-ribosomal S10 proteins and those in the free protein pool would then result in its large observed precursor pool. Similarly, the secondary binding r-protein S16 has a larger precursor pool than S4 (Fig. 6b), and is known to display endonuclease activity,[43] which suggests that S16 may also be involved in extra-ribosomal functions in *E. coli*.

The measured **P** value of S6 (2.0%) is significantly larger than that of S18 (0.8%) which, based on the order of *in vivo* assembly alone, suggests that S6 binds the 16S rRNA earlier than, and separately from, S18. However, it is well known that S6 and S18 assemble *in vitro* as a hetero-dimer,[44] which would imply an equimolar ratio in ribosome assembly intermediates. However, prior work suggests that there are multiple post-translationally modified forms of S6 in *E. coli*,[45] and that perhaps only the fully-matured form takes part in ribosome assembly. This would explain the larger precursor pool for S6 relative to S18. In light of these known extra-ribosomal processes, it is plausible that there are as yet unidentified extra-ribosomal functions for L28 and L32 as well.

Previous studies directly quantifying the amount of free r-protein in *E. coli* cell lysates have reported free protein pools in excess of 2% for r-proteins S1, S2, S6, L4, L7/L12, and L10.[27,40] In the present work, we measured the entire precursor pool, including both the free and intermediate bound proteins, and found an anomalously large **P** value for r-protein S6, which may indicate a significant free protein population as discussed above. In contrast, much smaller **P** values (0–1%) were observed for the late binding r-proteins S1, S2, L10, and L7/L12, which were previously found to have pool sizes as large as 20%. This discrepancy is likely a result of r-protein exchange between completed ribosomal particles and the free protein pool: in *E. coli* S1 is thought to be a loosely-associated translational factor,[46] and the 50S stalk proteins L10 and L7/L12 are known to exchange under active cell growth and protein translation.[47] Interestingly, large **P** values (6–7%) were observed for the late binding r-proteins S10 and L28, which have previously been reported to have undetectable pools. As noted above, this discrepancy may be due to their strong, but reversible, associations with target extra-ribosomal complexes,[42] which provide additional sources of r-proteins available for assembly that are not detectable as free proteins. Finally, the ~7% free pool previously reported[27] for r-protein L4 is in excess of the 3.8% total precursor pool reported here. There is no obvious explanation for this discrepancy, but the growth rate in the previous study is much faster (15 *vs.* 36 minutes doubling time), and the steady-state pool of L4 may be larger as a result.

Significant over-labeling, up to 8.4% per generation, was observed for r-proteins S21, S20, and L33 in 70S ribosomes. While some form of turnover is necessary to explain their over-labeling kinetics, there is no known mechanism for the targeted degradation of ribosome-bound r-proteins, nor is there any evidence for the global turnover of ribosomes (see ESI†). Instead, both r-proteins S21 and L33 are known to exchange between completed ribosomes and the free protein pool *in vivo*.[32–34] As described in the ESI,† turnover of these rapidly exchanging free r-proteins could account for the over-labeling observed for S21 and L33, while contaminating, extra copies of S20[35,36] could account for the over-labeling of this protein.

It should be noted that turnover does not necessarily imply r-protein degradation, as their irreversible inclusion into an orthogonal pathway to ribosome biogenesis would also account for the observed over-labeling patterns. To determine if protein degradation is the underlying mechanism of turnover, pulse-labeling experiments were carried out on *E. coli* W3110 cells as described above, and fraction-labeled values were measured for each r-protein in the whole cell lysate. Over-labeling of total cellular S21 and L33 was observed (data not shown), which suggests that these r-proteins are degraded inside the cell. In an attempt to identify the protease responsible, pulse-labeling experiments were carried out on *E. coli* strains bearing individual genetic knockouts[48] of each non-essential energy-dependent protease: ClpAP (*clpP*$^-$), ClpXP (*clpP*$^-$), Lon (*lon*$^-$), and HslUV (*hslV*$^-$). The final energy-dependent protease, FtsH, could not be tested as it is essential in *E. coli*.[49] No changes were observed in the over-labeling of S21 and L33 in completed ribosomes in any of the knockout strains as compared to the wild-type BW25113 parent strain (data not shown), indicating that none of these proteases are solely responsible for their degradation. Interestingly, S20 was not over-labeled in any of the BW25113-derived strains. This discrepancy may reflect physiological differences between the *E. coli* C-strain (MRE600) and the K-12 derived strain (BW25113).

Nomura and Nierhaus have previously demonstrated *in vitro* that a small subset of r-proteins in the 30S and 50S

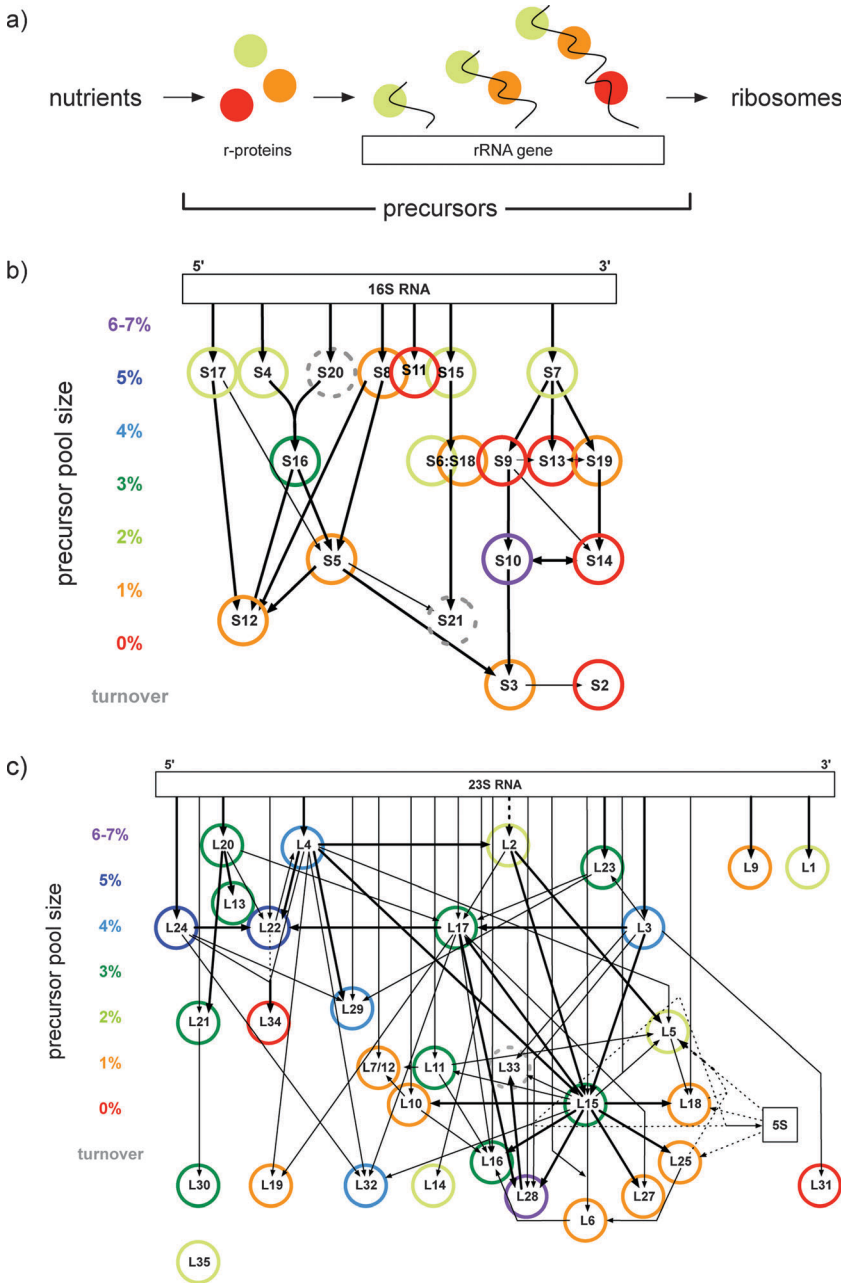

Fig. 6 Measured precursor pool/turnover mapped onto the 30S and 50S *in vitro* assembly maps. (a) During balanced growth, early binding r-proteins (green) in the co-transcriptional assembly process are enriched in the precursor pool compared to late binding r-proteins (red). (b, c) Calculated **P** values (solid circles) and turnover (dotted circles) painted onto the 30S and 50S *in vitro* assembly maps,[11,12] respectively.

subunits are "primary binders" and are required for the binding of subsequent r-proteins. In the 30S subunit, the primary binders S4, S7, S15, and S17 exhibited delays in labeling that correspond to precursor pools of ~2%, while in the 50S subunit the 5′ primary binders L24, L20, L22 and L4 were found to have precursor pools of ~4% (Fig. 5). As the primary binders are present in the earliest intermediates in the assembly pathway, their pulse-labeling kinetics provide upper bounds on the abundance of intermediates for each subunit. This result is in general agreement with a previous quantitative radiolabeling study which estimated that 30S and 50S precursors accounted for ~2% of total ribosomes.[28] The differences between the precursor pools of the 30S and 50S primary binders suggest that *E. coli* cells accumulate ~2-fold more 50S assembly intermediates than that of 30S during steady-state growth. Given that the scaffold 16S and 23S rRNAs are synthesized in stoichiometric quantities from a single precursor, a larger steady-state precursor pool implies that the 50S subunit assembles ~2-fold slower than the 30S subunit.[28] Under balanced growth, a pool of size **P** will be replaced by new material in approximately $t = \ln(1 + \mathbf{P})/k$ minutes (derivation not shown). Based on the intermediate pool sizes estimated above, ~60 seconds is required for 30S particle assembly, whereas assembly of the 50S particle requires ~120 seconds. Assuming a rate of rRNA transcription of 42 nucleotides per second,[50] the 16S and 23S transcription

phases are expected to account for only ~30 and ~60 seconds of 30S and 50S assembly, respectively. Therefore, it appears that significant portions of both 30S and 50S subunit assembly take place after their rRNAs has been completely transcribed.

Experimental procedures

Cell growth and pulse-labeling

For wild-type pulse-labeling experiments, MRE600 cells were inoculated from a non-saturated overnight culture into 500 mL of M9 medium containing ^{15}N ammonium sulfate, trace minerals, and vitamins, as described previously.[30] The culture was grown at 37 °C to ~0.5 OD_{600nm} and pulsed with an equal volume of pre-warmed ^{14}N M9 medium. Cell growth was arrested by adding the culture directly into an equal volume of ice after 5, 10, 15, 20.5, 25, 30, 35, or 40 minutes of pulse-labeling, and cells were harvested *via* centrifugation at 4 °C. To assist in downstream ribosomal peptide identification by LC-MS analysis, a reference culture of unlabeled cells was mixed with each sample prior to cell lysis.[15] Both the addition of fresh ^{14}N medium to the ^{15}N culture and the cell harvest procedures took ~20 seconds. Because these delays have opposing effects on the observed labeling kinetics, neither was factored into the reported pulse-labeling times. For S9 over-expression, BL21(DE3) Tuner cells containing a pET24b-S9 expression vector were grown in the presence of 50 μg mL^{-1} kanamycin with 0 μM, 5 μM, or 15 μM IPTG and pulsed as describe above.

Balanced growth was maintained throughout the pulse-labeling experiments. Based on the measured growth rate k and pulse times t, maximum labeling values $f_{max}(t)$ for the simple precursor pool model were calculated as:

$$f_{max}(t) = 1 - \exp[-k \cdot t]$$

Ribosome purification and LC-MS analysis

Wild-type 70S ribosome samples were prepared as described previously[51] with minor modifications. Specifically, pulse-labeled cells were mixed with a reference culture grown entirely in ^{14}N medium. The sample was then lysed using a BioSpec Mini-Beadbeater, and the cell lysate was spun down at 5500×g for 10 minutes followed by an additional spin at 30 000×g for 40 minutes. The clarified supernatant was loaded onto a 13–51% (w/v) non-dissociating linear sucrose gradient containing 10 mM $MgCl_2$ and centrifuged in a Beckman SW32 rotor at 26 000 rpm for 18 hours at 4 °C. Approximately 34 fractions were collected from each sucrose gradient using a Brandel gradient fractionator. Based on the UV 254 nm trace, gradient fractions corresponding to the 70S ribosome peak were pooled together, TCA precipitated, and digested with trypsin.

In the S9 over-expression pulse-labeling experiments, completed ribosomes were first pelleted through a 48% w/v non-dissociating sucrose cushion to remove contaminating S9 aggregates. The resultant 70S pellets were then resuspended in dissociating Buffer A (20 mM Tris-HCl pH 7.5, 100 mM NH_4Cl, 1 mM $MgCl_2$, 0.5 mM EDTA, 6 mM b-mercaptoethanol) and resolved over a 13–51% (w/v) dissociating sucrose gradient. Finally, the gradient fractions corresponding to the purified 30S subunit peak were pooled together and processed as described above.

Digested r-protein samples were purified using Pierce C-18 columns (Thermo Scientific) and submitted to LC-MS analysis on an Agilent ESI-TOF mass spectrometer. Subsequent peptide identification and isotope quantitation were carried out as described previously[30] with the following differences: peptides were eluted off of a concave 5–50% ACN gradient over 100 minutes, and the MS detection range was set at 250–1300 m/z. The fraction-labeled value for each r-protein, $f_r(t)$, was calculated as the fitted amplitude of the partially labeled peptides isotope distribution, A_P, relative to that of both the partially labeled and fully labeled peptide isotope distributions, A_P and A_F, respectively (Fig. 1):

$$f_r = \frac{r^*}{r} = \frac{A_P}{A_P + A_F}$$

For each sample, a fixed ^{15}N fractional enrichment value (*i.e.* the fraction of ^{15}N atoms in the partially labeled peptides) was set to the median value of the calculated enrichment values from a preliminary round of free-floating fits. To improve measurement precision, each isotope distribution and its local chromatographic contour map was carefully examined, and fits with low signal-to-noise ratios or isobaric interferences[52] were excluded from further analysis. Each sample was submitted in duplicate or triplicate to LC-MS analysis, and the technical replicates were combined to generate the average fraction-labeled values for the r-proteins at each pulse time point.

Quantitation of r-protein labeling kinetics

Rate equations for the labeling of precursor pool p and completed ribosomes r were derived based on steady-state balanced-flux models (see Results). Turnover rates per generation are defined as $1 - \exp[-d' \cdot 36]$. The closed-form solution to each model was obtained by integrating the appropriate rate equations. Curve fitting was carried out using the corresponding feature in Igor Pro software (WaveMetrics, Inc.). Final **P** values, turnover rates, and the corresponding errors in fitted parameter were obtained from the best fits of r-protein labeling kinetics using the closed-form solution of the appropriate biological model (Table S1, ESI†).

Conclusions

In this study, we carried out an eight-point stable isotope pulse-labeling experiment to quantitate the process of ribosome biogenesis in wild-type *E. coli* cells. The labeling kinetics for each r-protein in completed ribosomes was measured using a previously developed high-precision qMS approach, and the results were interpreted through a set of flux-based mathematical models. Reported here are the relative precursor pool sizes and turnover rates for 53 of the 54 r-proteins in *E. coli*. These biological parameters shed new light on the relatively unknown *in vivo* assembly landscapes of the 30S and 50S subunits, and measure the times for subunit assembly to be 1 and 2 minutes, respectively. In addition, anomalous pool size values and labeling kinetics for r-proteins S6, S10, S16, S20, S21, L28, L32, and L33 highlight their potential

extra-ribosomal functions. We expect this approach to be directly applicable to the study of ribosome biogenesis in a variety of model organisms. Furthermore, the general methodology and mathematics can be easily adapted to the high-throughput, high-resolution LC-MS/MS studies of a wide range of other interesting biological systems, such as the assembly of other macromolecular machineries and the measurement of whole-proteome dynamics and turnover.

Conflict of interest

The authors declare that they have no conflict of interest.

Note added after first publication

This article replaces the version published on 22nd October 2012, which contained errors in equations 7, 9, 10 and 12, the Fig. 5 legend and the Acknowledgements section.

Acknowledgements

We acknowledge Drs. C. Wittenberg, E. Valentine, G. Ring, and Z. Shajani for critical comments on the manuscript, and J. Figueroa for providing figure art. A BL21(DE3) strain containing a pET24b-S9 plasmid vector used in preliminary experiments was provided courtesy of Dr. G. Culver at the University of Rochester. This work was supported by a grant from the National Institute of Health to J.R.W. (R37-GM053757). J.H.D. is supported by a fellowship from the Jane Coffin Childs Foundation, and J. M. S. is supported in part by a fellowship from the ARCS Foundation.

Notes and references

1 M. Doherty, C. Whitehead, H. Mccormack, S. Gaskell and R. Beynon, *Proteomics*, 2005, **5**, 522–533.
2 I. V. Hinkson and J. E. Elias, *Trends Cell Biol.*, 2011, **21**, 293–303.
3 Q. Li, *Mass Spectrom. Rev.*, 2010, **29**, 717–736.
4 M. Scott and T. Hwa, *Curr. Opin. Biotechnol.*, 2011, **22**, 559–565.
5 A. Yonath, *Annu. Rev. Biochem.*, 2005, **74**, 649–679.
6 J. Poehlsgaard and S. Douthwaite, *Nat. Rev. Microbiol.*, 2005, **3**, 870–881.
7 B. A. Maguire, *Microbiol. Mol. Biol. Rev.*, 2009, **73**, 22–35.
8 L. Montanaro, D. Trere and M. Derenzini, *Am. J. Pathol.*, 2008, **173**, 301–310.
9 S. Belin, A. Beghin, E. Solano-Gonzalez, L. Bezin, S. Brunet-Manquat, J. Textoris, A. C. Prats, H. C. Mertani, C. Dumontet and J. J. Diaz, *PLoS ONE*, 2009, **4**, e7147.
10 V. Choesmel, D. Bacqueville, J. Rouquette, J. Noaillac-Depeyre, S. Fribourg, A. Cretien, T. Leblanc, G. Tchernia, L. Da Costa and P. E. Gleizes, *Blood*, 2007, **109**, 1275–1283.
11 W. A. Held, B. Ballou, S. Mizushima and M. Nomura, *J. Biol. Chem.*, 1974, **249**, 3103–3111.
12 M. Herold and K. H. Nierhaus, *J. Biol. Chem.*, 1987, **262**, 8826–8833.
13 T. Adilakshmi, D. L. Bellur and S. A. Woodson, *Nature*, 2008, **455**, 1268–1272.
14 M. W. Talkington, G. Siuzdak and J. R. Williamson, *Nature*, 2005, **438**, 628–632.
15 A. E. Bunner, S. A. Trauger, G. Siuzdak and J. R. Williamson, *Anal. Chem.*, 2008, **80**, 9379–9386.
16 A. M. Mulder, C. Yoshioka, A. H. Beck, A. E. Bunner, R. A. Milligan, C. S. Potter, B. Carragher and J. R. Williamson, *Science*, 2010, **330**, 673–677.
17 M. T. Sykes and J. R. Williamson, *Annu. Rev. Biophys.*, 2009, **38**, 197–215.
18 M. Kaczanowska and M. Ryden-Aulin, *Microbiol. Mol. Biol. Rev.*, 2007, **71**, 477–494.
19 Z. Shajani, M. T. Sykes and J. R. Williamson, *Annu. Rev. Biochem.*, 2011, **80**, 501–526.
20 D. N. Wilson and K. H. Nierhaus, *Crit. Rev. Biochem. Mol. Biol.*, 2007, **42**, 187–219.
21 M. Nomura, R. Gourse and G. Baughman, *Annu. Rev. Biochem.*, 1984, **53**, 75–117.
22 R. Schleif, *J. Mol. Biol.*, 1967, **27**, 41–55.
23 K. Gausing, *Mol. Gen. Genet.*, 1974, **129**, 61–75.
24 J. Marvaldi, J. Pichon, M. Delaage and G. Marchis-Mouren, *J. Mol. Biol.*, 1974, **84**, 83–96.
25 J. Pichon, J. Marvaldi and G. Marchis-Mouren, *J. Mol. Biol.*, 1975, **96**, 125–137.
26 J. Marvaldi, J. Pichon and G. Marchis-Mouren, *Biochim. Biophys. Acta*, 1972, **269**, 173–177.
27 B. Ulbrich and K. H. Nierhaus, *Eur. J. Biochem.*, 1975, **57**, 49–54.
28 L. Lindahl, *J. Mol. Biol.*, 1975, **92**, 15–37.
29 M. T. Sykes, E. Sperling, S. S. Chen and J. R. Williamson, *Anal. Chem.*, 2010, **82**, 5038–5045.
30 E. Sperling, A. E. Bunner, M. T. Sykes and J. R. Williamson, *Anal. Chem.*, 2008, **80**, 4906–4917.
31 L. Reitzer, *Annu. Rev. Microbiol.*, 2003, **57**, 155–176.
32 A. Pulk, A. Liiv, L. Peil, U. Maivali, K. Nierhaus and J. Remme, *Mol. Microbiol.*, 2010, **75**, 801–814.
33 A. R. Subramanian and J. van Duin, *Mol. Gen. Genet.*, 1977, **158**, 1–9.
34 W. R. Robertson, S. J. Dowsett and S. J. Hardy, *Mol. Gen. Genet.*, 1977, **157**, 205–214.
35 S. J. Hardy, *Mol. Gen. Genet.*, 1975, **140**, 253–274.
36 M. Tal, I. Weissman and A. Silberstein, *J. Biochem. Biophys. Methods*, 1990, **21**, 247–266.
37 B. H. Sells and F. C. Davis Jr., *J. Mol. Biol.*, 1970, **47**, 155–167.
38 P. Voynow and C. G. Kurland, *Biochemistry*, 1971, **10**, 517–524.
39 R. S. Gupta and U. N. Singh, *J. Mol. Biol.*, 1972, **69**, 279–301.
40 S. Ramagopal, *Eur. J. Biochem.*, 1976, **69**, 289–297.
41 D. I. Friedman, A. T. Schauer, M. R. Baumann, L. S. Baron and S. L. Adhya, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 1115–1118.
42 S. W. Mason and J. Greenblatt, *Genes Dev.*, 1991, **5**, 1504–1512.
43 J. Oberto, E. Bonnefoy, E. Mouray, O. Pellegrini, P. M. Wikstrom and J. Rouviere-Yaniv, *Mol. Microbiol.*, 1996, **19**, 1319–1330.
44 P. Svensson, L. M. Changchien, G. R. Craven and H. F. Noller, *J. Mol. Biol.*, 1988, **200**, 301–308.
45 S. Reeh and S. Pedersen, *Mol. Gen. Genet.*, 1979, **173**, 183–187.
46 M. A. Sorensen, J. Fricke and S. Pedersen, *J. Mol. Biol.*, 1998, **280**, 561–569.
47 S. Deroo, S. J. Hyung, J. Marcoux, Y. Gordiyenko, R. K. Koripella, S. Sanyal and C. V. Robinson, *ACS Chem Biol*, 2012, **7**, 1120–1127.
48 T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner and H. Mori, *Mol. Syst. Biol.*, 2006, **2**, 2006 0008.
49 C. Herman, T. Ogura, T. Tomoyasu, S. Hiraga, Y. Akiyama, K. Ito, R. Thomas, R. D'Ari and P. Bouloc, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 10861–10865.
50 R. L. Gourse, T. Gaal, M. S. Bartlett, J. A. Appleman and W. Ross, *Annu. Rev. Microbiol.*, 1996, **50**, 645–677.
51 M. T. Sykes, Z. Shajani, E. Sperling, A. H. Beck and J. R. Williamson, *J. Mol. Biol.*, 2010, **403**, 331–345.
52 M. Bantscheff, M. Schirle, G. Sweetman, J. Rick and B. Kuster, *Anal. Bioanal. Chem.*, 2007, **389**, 1017–1031.

