J. Biochem. 2021;169(4):387–394 doi:10.1093/jb/mvaa140

JB THE JOURNAL OF BIOCHEMISTRY

JB Review

Escherichia coli amino acid auxotrophic expression host strains for investigating protein structure–function relationships

Received July 30, 2020; accepted November 16, 2020; published online December 8, 2020

Toshio Iwasaki [1,*], Yoshiharu Miyajima-Nakano[1], Risako Fukazawa[1], Myat T. Lin [2,†], Shin-ichi Matsushita[1], Emi Hagiuda[1], Alexander T. Taguchi [1,†], Sergei A. Dikanov [3], Yumiko Oishi [1] and Robert B. Gennis [2]

[1]Department of Biochemistry and Molecular Biology, Nippon Medical School, Sendagi, Tokyo 113-8602, Japan; [2]Department of Biochemistry; and [3]Department of Veterinary Clinical Medicine, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

*Toshio Iwasaki, Department of Biochemistry and Molecular Biology, Nippon Medical School, 1-1-5 Sendagi, Bunkyo-ku,Tokyo 113-8602, Japan. Tel.: +81-3-3822-2131 (ext. 5237), email: tiwasaki@nms.ac.jp

†Present addresses: Myat T. Lin, Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14850, USA; Alexander T. Taguchi, RubrYc Therapeutics, Inc., 733 Industrial Road, San Carlos, CA 94403, USA.

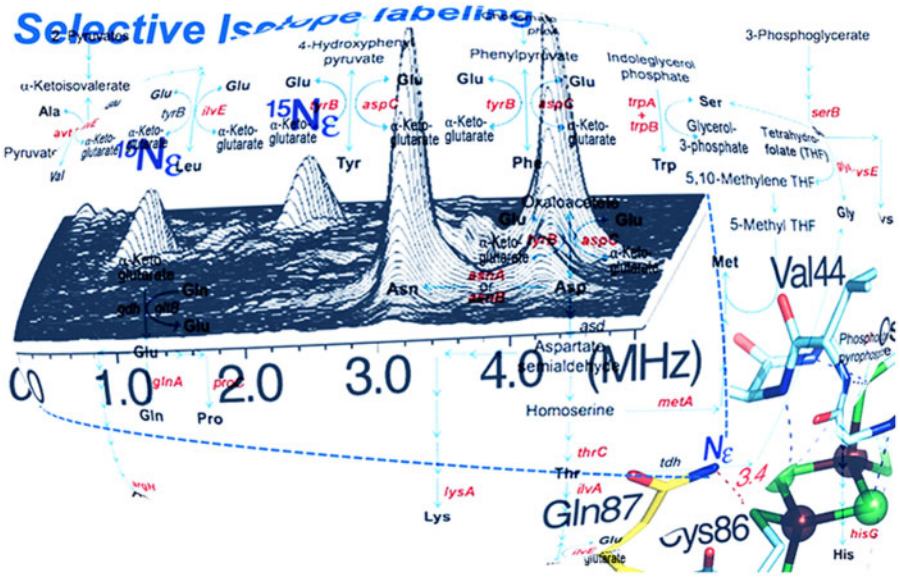
A set of C43(DE3) and BL21(DE3) *Escherichia coli* host strains that are auxotrophic for various amino acids is briefly reviewed. These strains require the addition of a defined set of one or more amino acids in the growth medium, and have been specifically designed for overproduction of membrane or water-soluble proteins selectively labelled with stable isotopes, such as ^{2}H, ^{13}C and ^{15}N. The strains described here are available for use and have been deposited into public strain banks. Although they cannot fully eliminate the possibility of isotope dilution and mixing, metabolic scrambling of the different amino acid types can be minimized through a careful consideration of the bacterial metabolic pathways. The use of a suitable auxotrophic expression host strain with an appropriately isotopically labelled growth medium ensures high levels of isotope labelling efficiency as well as selectivity for providing deeper insight into protein structure–function relationships.

Keywords: amino acid; auxotroph; *Escherichia coli*; isotope labelling; protein structure–function.

Abbreviations: EPR, electron paramagnetic resonance; FLP, flippase; IPTG, isopropyl-β-D-1-thiogalactopyranoside; LB, Luria–Bertani.

Amino acid-selective isotope labelling enriches a defined set of amino acid types in proteins with stable isotopes such as ^{2}H, ^{13}C and ^{15}N. This is a powerful approach for quantitatively resolving specific contributions of particular residues in reaction mechanisms and/or folding of a target protein by magnetic resonance [*e.g.* nuclear magnetic resonance and (pulsed) electron paramagnetic resonance (EPR)] and vibrational (*e.g.* resonance Raman and Fourier transform infrared) spectroscopies, often aided by the X-ray crystal structures (*1–13*).

Graphical Abstract



Isotopic labels can be incorporated into select amino acid residue types for protein samples by several ways. The *in vitro* labelling approach uses a cell-free protein synthesis system and has several advantages over *in vivo* biosynthesis (*2–4*). For example, it enables selective labelling of Asp or Glu residues within proteins (*4*), which is extremely difficult or impossible to do by other means. A technical limitation of this method is that it is usually only applicable to small and medium sized proteins with relatively simple subunit assembly and maturation processes. For more complicated proteins, such as multi-subunit membrane or soluble protein complexes with bound prosthetic groups, *in vivo* expression in a biological host is the only practical and cost-effective way for selective isotopic labelling (*5*,*8*).

One of the most convenient procedures for selectively labelling proteins with isotopes is by using amino acid *auxotrophic* bacteria as the host strains for overproduction of target proteins. A wild-type strain of *Escherichia coli* can biosynthesize all 20 amino acids *in vivo*, whereas an *E.coli* amino acid *auxotroph*, lacking one or more essential genes involved in the biosynthesis of an amino acid, requires particular amino acids for growth (*5*,*8*). In this review, we provide an overview of a set of new C43(DE3) and BL21(DE3) auxotrophic expression host strains of *E.coli* (Table I and Supplementary Table S1), which have been designed by our research groups to facilitate in depth biochemical and biophysical characterization of either membrane or water-soluble proteins enriched with stable isotopes such as ^{2}H, ^{13}C and ^{15}N for select amino acids (*5–13*). The use of a suitable auxotrophic expression strain with the corresponding input isotopically labelled amino acid(s) in the growth medium ensures high levels of stable isotope labelling efficiency as well as selectivity.

Brief Overview of the Amino Acid Metabolic Network of *E.coli*

Wild-type *E.coli* has the ability to biosynthesize all 20 amino acids *in vivo*. Note that the bacterial biosynthetic and degradation pathways of many amino acids significantly overlap, and many amino acids serve as either precursors or breakdown products for other amino acids (*14*, *15*) (Fig. 1). The control over the network of metabolic pathways is strictly regulated and can be highly complex. Figure 2A depicts a schematic example of the partial metabolic network of the proposed Thr biosynthesis and its regulation in *E.coli* (*16*,*17*). Thr belongs to the aspartic family of amino acids, and its biosynthesis starting with the Asp building block (*14*) involving five successive reactions is strictly regulated to meet metabolic flow requirements. Some enzymes in the Thr terminal pathways are subject to partial or complete feedback inhibition by specific L-amino acids [bifunctional aspartate kinase I-homoserine dehydrogenase I (*thrA* gene product) partially by Thr; aspartate kinase III (*lysC* product) by Lys; homoserine kinase (*thrB* product) by Thr] (*16*, *17*) (Fig. 2A). In addition, the expression of some biosynthesis genes in these pathways is subject to repression (the *thrABC* operon by transcriptional attenuation mediated by Thr and Ile; *metL* by Met; *lysC* by Lys; *asd* by Lys, Thr and Met) (*16*,*17*). The depletion of Thr *in vivo* occurs mainly through the Lys and Met competing branches, and also by conversion to Ile and Gly (Fig. 2A). Such metabolic and regulatory information can be useful to select a suitable auxotrophic strain with the corresponding isotope-labelled amino acid supplied in the growth medium.

Graphical Abstract A major hurdle for selective labelling with certain amino acids is posed by the presence of at least four general transaminases (or aminotransferases) of *E.coli*, encoded, respectively, by the *ilvE*, *avtA*, *aspC* and *tyrB* genes (*14*). These transaminases catalyze the interconversion of amino acids and ketoacids by (reversible) transfer of amino groups in biosynthesis of Ile, Leu, Val, Tyr, Phe and Asp with overlapping specificities; except for the *avtA* gene product, the other three general transaminases can use multiple substrates (Fig. 2B). Ideally, the best option in such cases is to use a strain with defects in all four general transaminases (*ilvE*, *avtA*, *aspC* and *tyrB*) (*14*). As described below, such an 'ideal' strain has recently been made available with *E.coli* BL21(DE3) (Table I), although the *ilvE*/*tyrB* double deletion from the C43(DE3) chromosome has been unsuccessful so far, for reasons unclear to us.

Construction and Selection of Appropriate *E.coli* Auxotroph Strains

At the beginning of the *E.coli* amino acid auxotroph strain collection project, a major obstacle to selective isotopic enrichment of proteins with stable isotopes for biophysical studies was the shortage of suitable *auxotrophic host* strains that are compatible with commonly used expression vectors for overproduction of proteins. *Escherichia coli* BL21(DE3) and C43(DE3) strains incorporate an inducible T7 RNA polymerase gene and represent one of the most popular hosts for homologous and heterologous protein production (*18–20*).

For this purpose, we have generated auxotrophic C43(DE3) and BL21(DE3) strains, requiring the addition of a defined set of one or more select amino acids in the growth medium (Table I) (*5*,*7–9*). This is accomplished by genomic insertion/deletion mutagenesis techniques with the λ-Red recombination system (*5*,*21*), followed by the removal of the resistance cassette from the new knockout strain by the flippase (FLP) recombinase expressed from the pCP20 vector (*22*) where applicable (*5*,*7*,*8*) (Supplementary Fig. S1). Genes were also targeted which encode enzymes, such as general deaminases (*14*), which would otherwise result in interconverting amino acids and scrambling of the input isotope labels (Table I). In view of the crucial roles of Glu in *E.coli* central amino acid metabolism (Fig. 1) and nitrogen assimilation, as well as the requirement of many genetic defects to guard against scrambling of a Glu input label (*14*), engineering of an auxotroph of *E.coli* for isotopic enrichment of this

Table I. ***Escherichia coli*** **amino acid auxotrophic expression host strains and their properties**

Strain	Precursor strain	Genotype	Selective amino acid labelling and/or requirement	Remarks
C43(DE3) derivatives				
ML2	CLY	*cyo*::*kan ilvE*	Ile, Leu[a]	Genotype *cyo*::*kan*
ML3	CLY	*cyo*::*kan hisG*	His	Genotype *cyo*::*kan*
ML6	ML2	*cyo*::*kan ilvE avtA*	Ile, Leu[a], Val	Genotype *cyo*::*kan*
ML8	CLY	*cyo*::*kan argH*	Arg	Genotype *cyo*::*kan*
ML12	ML6	*cyo*::*kan ilvE avtA aspC*	(Ala[b]), Ile, Leu[a], Tyr[b], Val	Genotype *cyo*::*kan*
ML14	C43(DE3)	*tyrA*	Tyr	
ML17	C43(DE3)	*glnA*	Gln	
ML21	ML14	*tyrA hisG*	Tyr, His	
ML24	ML23	*cyo ilvE avtA aspC hisG asnA*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His	
ML25	ML24	*cyo ilvE avtA aspC hisG asnA asnB*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His, Asn	
ML26	ML23	*cyo ilvE avtA aspC hisG argH*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His, Arg	
ML31	ML26	*cyo ilvE avtA aspC hisG argH metA*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His, Arg, Met	
ML36	ML23	*cyo ilvE avtA aspC hisG metA*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His, Met	
ML40[b]	ML31	*cyo ilvE avtA aspC hisG argH metA lysA*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His, Arg, Met, Lys	Almost normal growth rate
ML42	ML41	*cyo ilvE avtA aspC hisG argH metA lysA thrC asnB*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His, Arg, Met, Lys, Thr	
ML43	ML42	*cyo ilvE avtA aspC hisG argH metA lysA thrC asnA asnB*	(Ala[b]), Ile, Leu[a], Tyr[b], Val, His, Arg, Met, Lys, Thr, Asn	Very low recombinant protein production yield
YM138	C43(DE3)	*cysE*	Cys	Slow growth
MS1	YM138	*cysE hisG*	Cys, His	Very slow growth
RF11	C43(DE3)	*metA*	Met	
BL21(DE3) derivatives				
RF1	BL21 CodonPlus (DE3)-RIL	*glyA*	Gly	
RF2	BL21 CodonPlus (DE3)-RIL	*thrC*	Thr	Minor mixing of label
RF3	BL21 CodonPlus (DE3)-RIL	*aspC*	N.A. (not applicable)	No auxotrophy
RF4	RF3	*aspC tyrB*	Asp, (Tyr), (Phe)	Not tested for selective Asp labelling
RF5	RF4	*aspC tyrB hisG*	Asp, (Tyr), (Phe), His	Not tested for selective Asp labelling
RF6	BL21 CodonPlus (DE3)-RIL	*proC*	Pro	
RF8	BL21 CodonPlus (DE3)-RIL	*asnA asnB*	Asn	
RF10	BL21 CodonPlus (DE3)-RIL	*lysA*	Lys	
RF12	BL21 CodonPlus (DE3)-RIL	*trpA trpB*	Trp	*tyrB* repression by adding 0.4–1 mM Tyr recommended
RF13	RF4	*aspC tyrB trpA trpB*	Asp, Tyr, Trp, (Phe)	Not tested for selective Asp labelling
RF14	RF13	*aspC tyrB trpA trpB serB*	Asp, Tyr, Trp, (Phe), Ser	Not tested for selective Asp or Ser labelling
RF15	RF14	*aspC tyrB trpA trpB glyA serB*	Asp, Tyr, Trp, (Phe), Gly, Ser	Not tested for selective Asp or Ser labelling; growth inhibited by Cys (Fig. 4)
RF16	RF15	*aspC tyrB trpA trpB glyA serB cysE*	Asp, Tyr, Trp, (Phe), Gly, Ser, Cys, Ala	Ala auxotroph; not tested for selective Asp, Ser or Ala labelling; very slow growth
RF17	RF4	*aspC tyrB ilvE*	Asp, Tyr, Phe, Ile, Leu	Not tested for selective Asp labelling; slow growth
RF18	RF17	*aspC tyrB ilvE avtA*	Asp, Tyr, Phe, Ile, Leu, Val	Not tested for selective Asp labelling; slow growth
RF21	RF18	*aspC tyrB ilvE avtA yfbQ(alaA) yfdZ(alaC)*	Asp, Tyr, Phe, Ile, Leu, Val	Not tested for selective Asp labelling; slow growth
EH1	RF2	*thrC ilvA*	Thr, Ile	

[a]In the presence of 0.4–1 mM Tyr, *tyrB* is repressed and Leu is required for growth in minimal medium.
[b]In the presence of 0.4–1 mM Tyr, *tyrB* is repressed and Tyr is required for growth in minimal medium. Under these conditions, ML40 can also be used for selective L-^{15}N-Ala labelling (with minor mixing), albeit only for short-term cultivations.

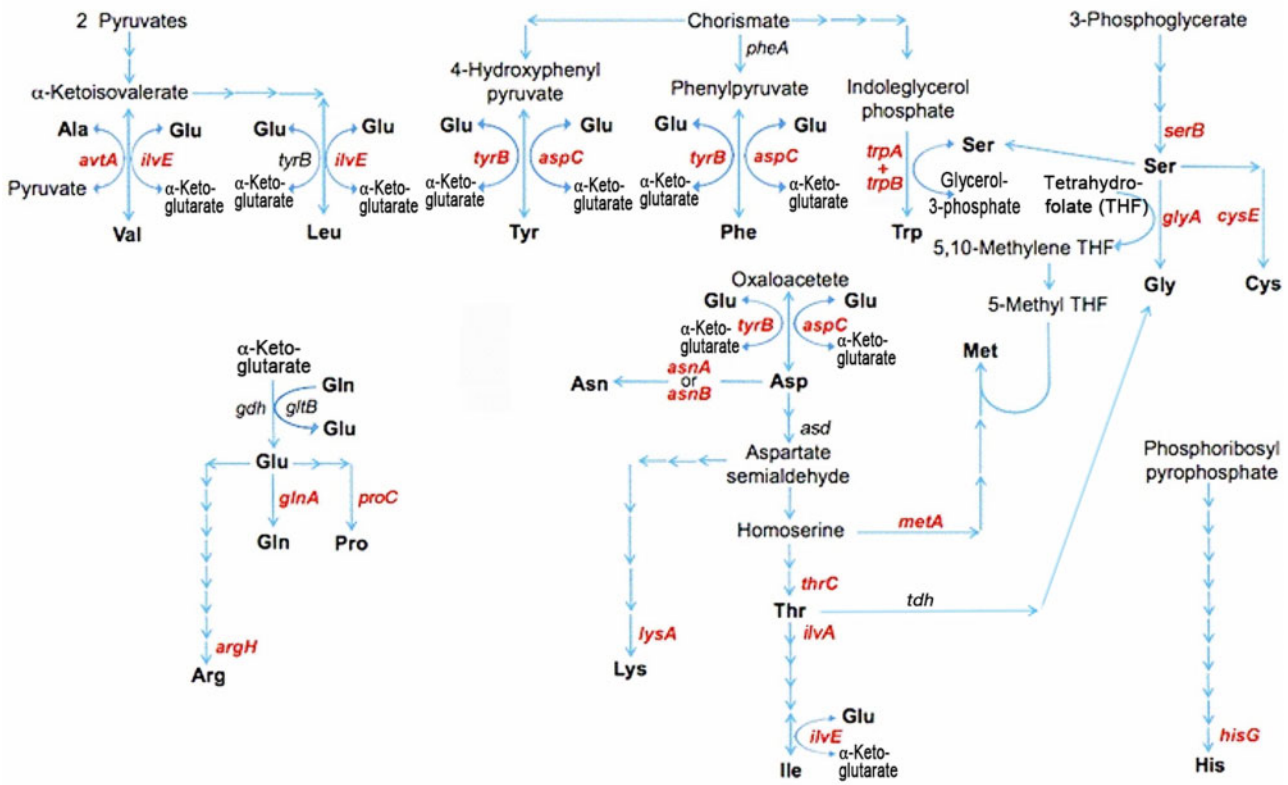


Fig. 1. Schematic view of L-amino acid biosynthesis/degradation network (main pathways) in *E.coli*. This scheme should be used with caution and only as a general guide. Coloured in red are genes subjected to genomic deletion in this study.

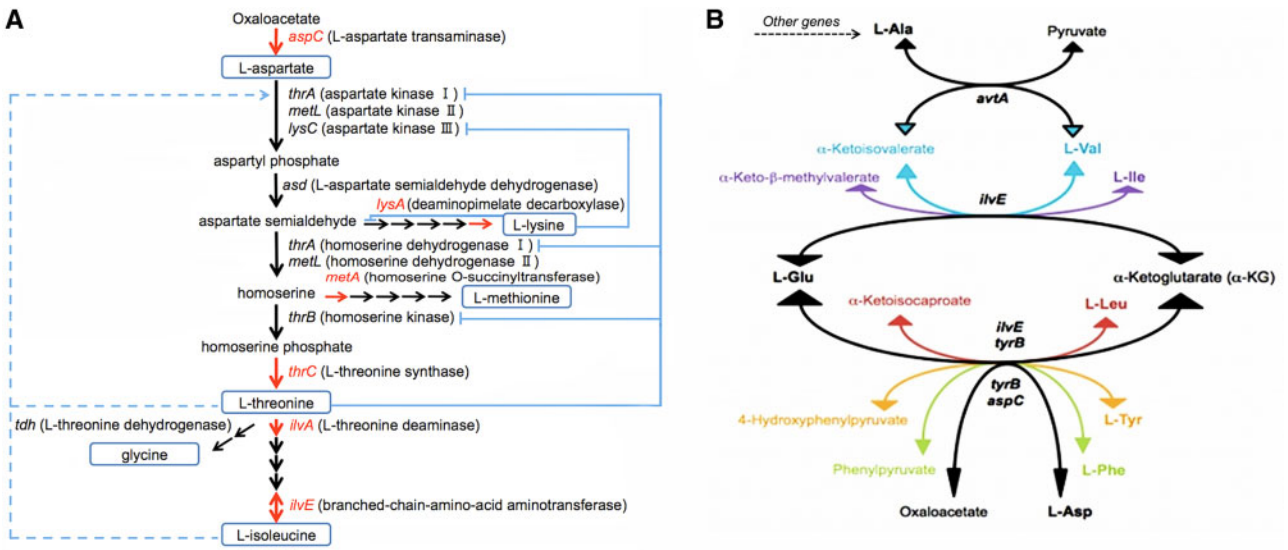


Fig. 2. (A) Proposed (partial) metabolic network of the Thr biosynthesis/degradation and regulation in *E.coli* (dashed lines, regulation by transcriptional attenuation; solid blue lines, regulation by feedback inhibition) and (B) schematic view of selected amino acid biosynthesis network in *E.coli*, catalyzed by four general transaminases (see Fig. 1).

amino acid is not considered in this study. For high-level expression of the foreign genes coding for metalloenzymes from extremophilic archaea and bacteria, which often requires extra copies of tRNA genes for the cognate rare codons, a pACYC-based plasmid harbouring tRNA genes (*argU*, *ileY* and *leuW*) for the *E.coli* rare codons was subsequently incorporated into the resulting cells (*7*,*9*) (Supplementary Fig. S1). These auxotrophic strains are designed for selective labelling of amino acids, and can be used for cost-effective, high-yield production of any recombinant water-soluble or membrane protein from a variety of species that can be expressed in *E.coli* (*5*–*13*).

For any given amino acid biosynthesis pathway in *E.coli*, there are only a limited number of junctions that could lead to scrambling of the input label with other amino acids (*14*) (Fig. 1). Consideration of these branching points is critical to the design of

auxotrophic strains with acceptably low levels of label scrambling for many amino acids, but in some cases only a single gene knockout is required: Arg, Cys, Gln, Gly, His, Ile, Lys, Met, Pro and Thr (*5*,*6*,*8*,*9*,*14*). Within this set, the C43(DE3) Cys auxotroph strains having deletions of *cysE* (**YM138** and **MS1** in Table I) grow very slowly but effectively eliminate most of the scrambling with media enriched in L-^{15}N-cysteine or L-3-^{13}C-cysteine (*8*,*9*). On the other hand, the Thr auxotrophic strains having the deletion of *thrC* [previously considered to be an ideal genotype for threonine labelling (*14*); strains **RF2** and **EH1** in Table I] eliminate most but not all of the scrambling for L-^{15}N-threonine enriched media (T. Iwasaki, R. Fukazawa, and S. A. Dikanov, unpublished results), and in this case possible metabolic scrambling by degradation to glycine may be reduced to lower levels by also deleting the threonine dehydrogenase (*tdh*) gene (*16*,*17*) (see Fig. 2A).

Auxotrophs for the remaining 10 amino acids each require two or more genes to be deleted, but most of them share the requirement of deleting genes encoding the four general aminotransferase genes, *aspC*, *avtA*, *ilvE* and *tyrB* (*14*) (Figs 1 and 2B). We have generated *E.coli* strains that are appropriate for selective labelling of eight other amino acids, *i.e.* Ala, Asn, Leu, Phe, Ser, Trp, Tyr and Val (*5*,*8*) (Table I and Supplementary Table S1). In principle, the potential use of strains having knockouts of all four general transaminases [*aspC*, *tyrB*, *ilvE* and *avtA* (*14*)] of *E.coli* should help minimize possible dilution and scrambling of the input amino acid label whenever available (*13*)—in practice, some of these strains appear to exhibit complicated patterns of amino acid combination requirements for growth or they grow very slowly, so careful planning of experimental conditions is required (Fig. 3): (i) strains having knockouts in both *aspC* and *tyrB* genes (**RF4**, **RF5**, **RF13**, **RF14**, **RF15**, **RF16**, **RF17**, **RF18** and **RF21**) require Asp (but not Glu) for growth in M63 minimal media, and some of them (**RF15**, **RF16**, **RF17**, **RF18** and **RF21**) grow only slowly in M63 minimal media in the presence of the L-amino acids specified in Table I; (ii) strain **RF17**, having knockouts in the *aspC*, *tyrB* and *ilvE* genes, requires the presence of Asp, Tyr, Phe, Ile and Leu for (slow) growth in M63 minimal media; (iii) strains **RF18** and **RF21**, having knockouts in all four of the general transaminase genes [*aspC*, *tyrB*, *ilvE* and *avtA* (*14*)], appear to require the presence of Asp, Tyr, Phe, Ile, Leu and Val for slow growth in M63 minimal media. Of these, although strain **RF21** has further knockouts in genes *yfbQ* (*alaA*) and *yfdZ* (*alaC*) [*cf.* (*15*)], it requires the presence of Asp, Tyr, Phe, Ile, Leu and Val for slow growth in M63 minimal media like **RF18**, and is not an Ala auxotroph either (Fig. 3). Considering also the regulation of the Asp degradation pathways by feedback inhibition by Thr and Lys, and some form of repression by Thr, Ile, Lys and Met (*16*, *17*) (Fig. 2A), further deletions of *asnA* and *asnB* (*14*) in **RF18** would result in a BL21(DE3) derivative requiring the presence of Asp, Tyr, Phe, Ile, Leu and Asn for growth in M63 minimal media, which is expected to be applicable to selective labelling of Asp at least for short-term cultivations grown in medium supplied with sufficient amounts of Thr, Ile, Lys and Met.

Fig. 3. Amino acid requirements of RF3 (top left) and RF4 (top right) grown for ~20 h, and RF17 (first middle), RF18 (second middle) and RF21 (bottom) grown for ~40–43 h in M63 minimal medium supplied with the specified L-amino acid(s) or in LB medium. **RF3** (Δ*aspC*) grows in M63 minimal medium; **RF4** (Δ*aspC*, Δ*tyrB*) is an Asp auxotroph; **RF17** (Δ*aspC*, Δ*tyrB*, Δ*ilvE*) requires Asp, Ile, Leu, Tyr and Phe for growth; both **RF18** (Δ*aspC*, Δ*tyrB*, Δ*ilvE*, Δ*avtA*) and **RF21** [Δ*aspC*, Δ*tyrB*, Δ*ilvE*, Δ*avtA*, Δ*yfbQ* (Δ*alaA* (*15*)), Δ*yfdZ* (Δ*alaC* (*15*))] require Asp, Ile, Leu, Tyr, Phe and Val (but not Ala) for growth. Concentration(s) of amino acid(s) added to M63 minimal medium (where they exist) (*24*): 0.4 g/l L-aspartate, 0.17 g/l L-tyrosine, 0.13 g/l L-phenylalanine, 0.23 g/l L-isoleucine, 0.23 g/l L-leucine, 0.23 g/l L-valine, 0.5 g/l L-alanine, 2.1 g/l L-serine. This figure can only be used as a general guide for the amino acid requirements of the selected strains, where precaution must be taken to optimize expression conditions for each target protein to be tested, *e.g.* by using LB medium first.

Notably, some of these strains can also be used for selective labelling of select aromatic amino acid(s) such as Tyr (**RF4**, **RF5**, **RF13**, **RF14** and **RF15**) and/or Phe (**RF13**, **RF14** and **RF15**) (Figs 3 and 4) for studying protein folding events and/or protein–ligand interactions (*7*,*8*). Alternatively, Tyr is known to suppress expression of *tyrB* (*23*), and this can be useful in minimizing label scrambling using one of the C43(DE3) auxotrophic strains (*5*,*6*) [Table I; applicable only for short-term cultivations (*7*)]. For example, by adding L-tyrosine (0.4–1 mM) into the

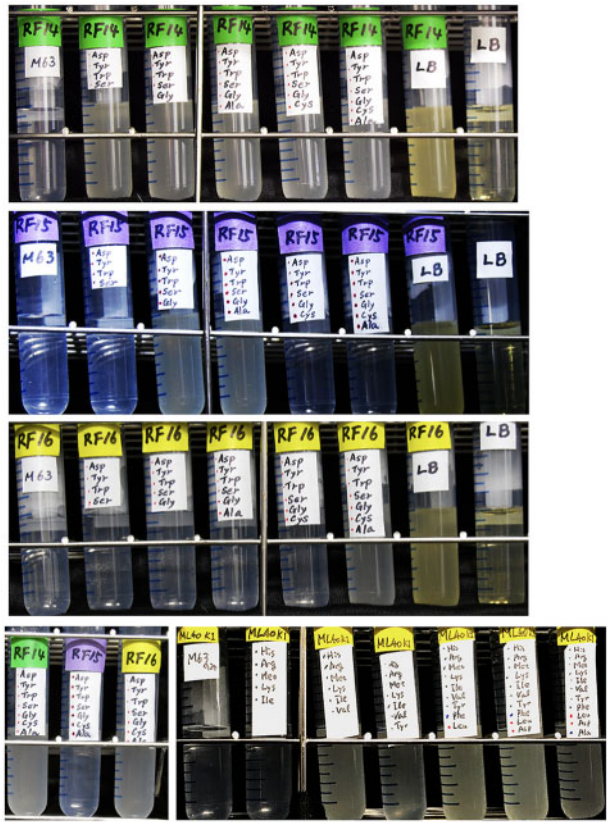


Fig. 4. Amino acid requirements of RF14 (top), RF15 (first middle) and RF16 (second middle) grown for ~40–43 h, and ML40 (bottom right) grown for ~20 h in M63 minimal medium supplied with the specified L-amino acid(s) or in LB medium. **RF14** (Δ*aspC*, Δ*tyrB*, Δ*trpA*, Δ*trpB*, Δ*serB*) requires Asp, Tyr, Trp and Ser for growth in M63 minimal medium; **RF15** (Δ*aspC*, Δ*tyrB*, Δ*trpA*, Δ*trpB*, Δ*glyA*, Δ*serB*) requires Asp, Tyr, Trp, Gly and Ser for growth in M63 minimal medium, but further addition of Cys inhibits growth (first middle and bottom left); **RF16** (Δ*aspC*, Δ*tyrB*, Δ*trpA*, Δ*trpB*, Δ*glyA*, Δ*serB*, Δ*cysE*) unexpectedly appears to be a new Ala auxotroph of BL21(DE3) that requires Asp, Tyr, Trp, Gly, Ser, Cys and Ala for very slow growth in M63 minimal medium (second middle) and was not inhibited by the presence of Cys (bottom left). **ML40** (Δ*aspC*, Δ*ilvE*, Δ*avtA*, Δ*hisG*, Δ*argH*, Δ*metA*, Δ*lysA*, Δ*cyo*) requires His, Arg, Met, Lys, Ile and Val for growth in M63 minimal medium; in the presence of 0.4–1 mM Tyr, **ML40** requires His, Arg, Met, Lys, Ile, Val, Tyr, Phe and Leu for slow growth due to repression of *tyrB* under the applied conditions (*5*,*23*) (bottom right). Concentration(s) of amino acid(s) added to M63 minimal medium (where they exist) (*24*): 0.4 g/l L-aspartate, 0.17 g/l L-tyrosine, 0.1 g/l L-tryptophan, 2.1 g/l L-serine, 0.55 g/l glycine, 0.05 g/l L-cysteine, 0.5 g/l L-alanine, 0.1 g/l L-histidine, 0.4 g/l L-arginine, 0.25 g/l L-methionine, 0.42 g/l L-lysine, 0.23 g/l L-isoleucine, 0.23 g/l L-valine, 0.13 g/l L-phenylalanine and 0.23 g/l L-leucine. This figure can only be used as a general guide for the amino acid requirements of the selected strains, where precaution must be taken to optimize expression conditions for each target protein to be tested, *e.g.* by using LB medium first.

medium, we were able to selectively label leucine using strain **ML6** (*5*). In another example, strains **ML14** and **ML21**, having *tyrA* deletions, are Tyr auxotrophs (Table I), although they are suitable only for selective labelling of the tyrosine side chain (*5*,*12*) and should not be used for selective labelling of proteins with L-^{15}N-Tyr due to the possible scrambling of the ^{15}N isotope by the transaminases encoded by *aspC* and *tyrB* (*14*) (Fig. 2B).

The complicated patterns of amino acid requirements for bacterial growth in these examples clearly reflect the significant overlap in cognate biosynthetic and degradation pathways (*14*) (Fig. 1). Our knowledge about the complicated regulation of the metabolic flow is still incomplete (Fig. 4): (i) strain **RF14**, having knockouts in *aspC*, *tyrB*, *trpA*, *trpB* and *serB* genes, requires the presence of Asp, Tyr, Trp and Ser for growth in M63 minimal medium, but it is not ideal to selectively label Ser (Table I); (ii) strain **RF15**, having knockouts in *aspC*, *tyrB*, *trpA*, *trpB*, *glyA* and *serB* genes, requires the presence of Asp, Tyr, Trp, Gly and Ser for growth in M63 minimal medium, but further addition of Cys inhibits growth; (iii) strain **RF16**, having knockouts in *aspC*, *tyrB*, *trpA*, *trpB*, *glyA*, *serB* and *cysE* genes and representing an ideal genotype to selectively label Ser (*14*), unexpectedly requires the presence of Asp, Tyr, Trp, Gly, Ser, Cys and Ala for very slow growth in M63 minimal medium, but does not grow in M63 minimal medium in the presence of Asp, Tyr, Trp, Gly, Ser and Cys like **RF15**. Thus, **RF16** unexpectedly represents a new BL21(DE3) derived Ala auxotroph (Fig. 4), whose underlying metabolic flow and regulation in *E.coli* remain elusive. These results suggest that appropriate growth conditions must be considered on a case-by-case basis, *e.g.* for 'reverse labelling' experiments see (*6*–*8*).

While most of the auxotroph strains with multiple gene deletions [*e.g.* **RF15**, **RF16** (very slow growth), **RF17**, **RF18** and **RF21**] often exhibit slow growth in Luria–Bertani (LB) media, the almost normal growth of **ML40** compared with the wild-type C43(DE3) strain is a remarkable exception (*5*,*7*,*8*). **ML40** has knockouts in *aspC*, *ilvE*, *avtA*, *hisG*, *argH*, *metA*, *lysA* and *cyo* genes, and requires the presence of His, Arg, Met, Lys, Ile and Val (but not Asp or Ala) for growth in M63 minimal medium (Fig. 4, bottom right). In the presence of 0.4–1 mM Tyr in the growth medium, *tyrB* gene is repressed (*5*,*23*), and the resulting **ML40** grows very slowly with the requirement of His, Arg, Met, Lys, Ile, Val, Tyr, Phe and Leu (but not Asp or Ala) in the M63 minimal medium (Fig. 4, bottom right); under the latter *tyrB*-repressed conditions in the presence of 0.4–1 mM Tyr, this strain can be used for selective labelling of L-^{15}N-Ala, albeit only for short-term cultivations [after isopropyl-*β*-D-1-thiogalactopyranoside (IPTG) induction] for heterologous expression of foreign genes (T. Iwasaki, R. Fukazawa, and S. A. Dikanov, unpublished results).

Lab Tips

Table I and Supplementary Table S1 summarize the current set of new C43(DE3) and BL21(DE3) auxotrophic expression strains of *E.coli* designed to facilitate isotopic labelling of specific amino acid types for either membrane proteins or water-soluble proteins. Almost all of these *E.coli* auxotrophic expression host strains are available through the public strain banks, Addgene, USA (https://www.addgene.org/Toshio_Iwasaki) and RIKEN Bioresource Research Center, Japan (https://dnaconda.riken.jp/search/depositor/

dep006963.html). At the time of preparing this review article, about 220 strains have been distributed from Addgene, USA to many laboratories in different countries. Additional information can be found on the research website https://fesworld.jp, and the list of PCR primers used for the construction and verification of these auxotrophs (before deposition into the public strain banks) are given in Supplementary Tables S2 and S3, respectively.

In our previous publications (*5*–*9*), detailed protocols for the usage of some of these strains have been described. In short, these strains must be made competent first (they are not competent cells). Then, the expression efficiency of each target protein should be optimized, *e.g.* using LB medium first and finding out the optimal expression conditions for further experiments. One important note, here, is that the protein expression using the **ML**, **YM138**, **MS1** and **RF11** strains based on C43(DE3) [a derivative of the BL21(DE3) strain available from Lucigen Inc., Middleton, WI, USA, and optimized for the successful overproduction of membrane proteins (*19*,*20*)] is strictly IPTG-controlled, whereas using BL21(DE3)-based strains (**RF** and **EH** strains) is much more leaky, at least under our experimental conditions tested [*e.g.* see (*7*–*9*)]. In other words, one might need longer cell growth times for effective expression of a target protein with C43(DE3) derivative strains. Therefore, one should decide carefully how to proceed with the growth conditions in each case [*e.g.* bacterial growth temperature, aeration, IPTG induction time for protein overproduction, and, when required for proper protein maturation processes, co-expression *e.g.* with *groELS*, *trx* or other cofactor biosynthesis gene(s)] (*5*,*8*).

In the case of running selective amino acid isotope labelling, we strongly recommend adding the amino acid isotope label together with other non-labelled amino acids to minimize mixing of the input amino acid isotope label (see Figs 1 and 2). One can make such a mixture, or if applicable, can use a commercially available algal medium containing extracts of amino acids and other compounds (*7*–*9*). Conversely, for the reverse labelling experiment, we recommend using a commercially available isotope-labelled (*e.g.* algae) medium and adding unlabelled amino acid(s) of interest prior to IPTG induction (*6*–*8*).

Conclusions

In this review, we provided a brief overview of a set of C43(DE3) and BL21(DE3) auxotrophic expression host strains of *E.coli*, all requiring the addition of a defined set of one or more amino acids in the growth medium (Table I and Supplementary Table S1). These strains have been generated by genomic insertion/deletion mutagenesis of some key genes involved in cognate amino acid biosynthesis and the degradation pathway network. They are suitably designed for overproduction of selectively labelled membrane or water-soluble proteins with stable isotopes such as ^{2}H, ^{13}C and ^{15}N, and have been deposited to public strain banks. The bacterial metabolic and regulatory pathways are often worth considering to reduce the potential dilution and mixing of the input isotope labels. The use of a suitable auxotrophic expression host strain with the corresponding isotopically labelled amino acid(s) in the appropriate growth medium will ensure high levels of efficiency as well as selectivity in stable isotope labelling, and is expected to solve many selective labelling problems.

Supplementary Data

Supplementary Data are available at *JB* Online.

Author Contributions

The manuscript was written by contributions of all authors. All authors approve of the final version of the manuscript.

Funding

This investigation was supported in part by the International Collaborations in Chemistry Grant from the Japan Society for the Promotion of Science (T.I.) and the National Science Foundation (CHE-1026541 to S.A.D.), the Japan Society for the Promotion of Science Grants-in-Aid [24659202, 26670215 (T.I.), 14F04415 (T.I., A.T.T.), and 20H03679, 20H04956 (Y.O.)], the Nagase Science and Technology Foundation Research Grant (T.I.), the Japan Agency for Medical Research and Development Grant [20gm5910021h9904, 20gm6210023h0001 (Y.O.)], the DE-FG02-87ER13716 (R.B.G.) and DE-FG02-08ER15960 (S.A.D., pulsed EPR work) Grants from the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Sciences, United States Department of Energy, and National Institute of General Medical Sciences Roadmap Initiative (R01GM075937) and National Institutes of Health Grant GM062954 (S.A.D.). A.T.T. gratefully acknowledges support as a Japan Society for the Promotion of Science Postdoctoral Fellow (P14415).

Conflict of Interest

None declared.

References

1. Rotsaert, F.A.J., Pikus, J.D., Fox, B.G., Markley, J.L., and Sanders-Loehr, J. (2003) N-isotope effects on the Raman spectra of Fe_2S_2 ferredoxin and Rieske ferredoxin: evidence for structural rigidity of metal sites. *J. Biol. Inorg. Chem.* **8**, 318–326
2. Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A., and Güntert, P. (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* **440**, 52–57
3. Staunton, D., Schlinkert, R., Zanetti, G., Colebrook, S.A., and Campbell, I.D. (2006) Cell-free expression and selective isotope labelling in protein NMR. *Magn. Reson. Chem.* **44**, S2–S9
4. Yokoyama, J., Matsuda, T., Koshiba, S., Tochio, N., and Kigawa, T. (2011) A practical method for cell-free protein synthesis to avoid stable isotope scrambling and dilution. *Anal. Biochem.* **411**, 223–229
5. Lin, M.T., Sperling, L.J., Frericks Schmidt, H.L., Tang, M., Samoilova, R.I., Kumasaka, T., Iwasaki, T., Dikanov, S.A., Rienstra, C.M., and Gennis, R.B. (2011) A rapid and robust method for selective isotope labeling of proteins. *Methods* **55**, 370–378
6. Lin, M.T., Baldansuren, A., Hart, R., Samoilova, R., Narasimhulu, K.V., Yap, L.L., Choi, S.K., O'Malley, P.J., Gennis, R.B., and Dikanov, S.A. (2012) Interaction of intermediate semiquinone with

surrounding protein residues at the Q_H site of wild-type and D75H mutant cytochrome bo_3 from *Escherichia coli*. *Biochemistry* **51**, 3827–3838
7. Iwasaki, T., Fukazawa, R., Miyajima-Nakano, Y., Baldansuren, A., Matsushita, S., Lin, M.T., Gennis, R.B., Hasegawa, K., Kumasaka, T., and Dikanov, S.A. (2012) Dissection of hydrogen bond interaction network around an iron-sulfur cluster by site-specific isotope labeling of hyperthermophilic archaeal Rieske-type ferredoxin. *J. Am. Chem. Soc.* **134**, 19731–19738
8. Lin, M.T., Fukazawa, R., Miyajima-Nakano, Y., Matsushita, S., Choi, S.K., Iwasaki, T., and Gennis, R.B. (2015) *Escherichia coli* auxotroph host strains for amino acid-selective isotope labeling of recombinant proteins. *Methods Enzymol.* **565**, 45–66
9. Taguchi, A.T., Miyajima-Nakano, Y., Fukazawa, R., Lin, M.T., Baldansuren, A., Gennis, R.B., Hasegawa, K., Kumasaka, T., Dikanov, S.A., and Iwasaki, T. (2018) Unpaired electron spin density distribution across reduced [2Fe-2S] cluster ligands by $^{13}C_{\beta}$-cysteine labeling. *Inorg. Chem.* **57**, 741–746
10. Hartmann, J.-B., Zahn, M., Matečko Burmann, I., Bibow, S., and Hiller, S. (2018) Sequence-specific solution NMR assignments of the β-barrel insertase BamA to monitor its conformational ensemble at the atomic level. *J. Am. Chem. Soc.* **140**, 11252–11260
11. Iwata, T., Nagai, T., Ito, S., Osoegawa, S., Iseki, M., Watanabe, M., Unno, M., Kitagawa, S., and Kandori, H. (2018) Hydrogen bonding environments in the photocycle process around the flavin chromophore of the AppA-BLUF domain. *J. Am. Chem. Soc.* **140**, 11982–11991
12. Streit, B.R., Celis, A.I., Moraski, G.C., Shisler, K.A., Shepard, E.M., Rodgers, K.R., Lukat-Rodgers, G.S., and DuBois, J.L. (2018) Decarboxylation involving a ferryl, propionate, and a tyrosyl group in a radical relay yields heme *b*. *J. Biol. Chem.* **293**, 3989–3999
13. Kooijman, L., Ansorge, P., Schuster, M., Baumann, C., Löhr, F., Jurt, S., Güntert, P., and Zerbe, O. (2020) Backbone and methyl assignment of bacteriorhodopsin incorporated into nanodiscs. *J. Biomol. NMR* **74**, 45–60
14. Waugh, D.S. (1996) Genetic tools for selective labeling of proteins with α-^{15}N-amino acids. *J. Biomol. NMR* **8**, 184–192
15. Kim, S.H., Schneider, B.L., and Reitzer, L. (2010) Genetics and regulation of the major enzymes of alanine synthesis in *Escherichia coli*. *J. Bacteriol.* **192**, 5304–5311
16. Lee, K.H., Park, J.H., Kim, T.Y., Kim, H.U., and Lee, S.Y. (2007) Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol. Syst. Biol.* **3**, 149
17. Lee, S.Y. and Park, J.H. (2010) Integration of systems biology with bioprocess engineering: L-threonine production by systems metabolic engineering of *Escherichia coli*. *Adv. Biochem. Eng. Biotechnol.* **120**, 1–19
18. Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130
19. Miroux, B. and Walker, J.E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298
20. Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K.J., Slotboom, D.J., Persson, J.O., and de Gier, J.-W. (2008) Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. USA* **105**, 14371–14376
21. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645
22. Cherepanov, P.P. and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9–14
23. Yang, J., Camakaris, H., and Pittard, J. (2002) Molecular analysis of tyrosine- and phenylalanine-mediated repression of the *tyrB* promoter by TyrR protein of *Escherichia coli*. *Mol. Microbiol.* **45**, 1407–1419
24. Chae, Y.K., Abildgaard, F., Mooberry, E.S., and Markley, J.L. (1994) Multinuclear, multidimensional NMR studies of *Anabaena* 7120 heterocyst ferredoxin: sequence-specific resonance assignments and secondary structure of the oxidized form in solution. *Biochemistry* **33**, 3287–3295