Journal of Biomolecular NMR, 8 (1996) 184–192
ESCOM

J-Bio NMR 367

Genetic tools for selective labeling of proteins with α-^{15}N-amino acids

David S. Waugh*

Department of Physical Chemistry, Roche Research Center, Hoffmann-La Roche Inc., 340 Kingsland Street, Nutley, NJ 07110, U.S.A.

Received 29 March 1996
Accepted 16 May 1996

Keywords: ^{15}N-amino acids; Amino acid metabolism; Aminotransferase; Genetic lesions; Residue-specific labeling; Selective labeling; Stable isotopes; Transaminase

Summary

A collection of genetic tools that can be used to manipulate amino acid metabolism in *Escherichia coli* is described. The set comprises 21 strains of bacteria, each containing a different genetic defect that is closely linked to a selectable transposon marker. These tools can be used to construct strains of *E. coli* with ideal genotypes for residue-specific, selective labeling of proteins with nearly any ^{15}N-amino acid. By using strains which have been modified to contain the appropriate genetic lesions to control amino acid biosynthesis, dilution of the isotope by endogenous amino acid biosynthesis and scrambling of the label to other types of residues can be avoided.

Introduction

Selective ^{15}N-labeling of the peptide backbone of a protein by residue type often enables sequential assignments to be achieved more rapidly and with a greater degree of confidence. This is particularly true for larger proteins, which typically have crowded spectra. Selective enrichment is accomplished biosynthetically, by cultivating cells that overproduce the protein of interest in defined medium supplemented with one or more specific ^{15}N-amino acids.

Two concerns arise when the label is to be restricted to specific types of residues. The first is dilution of the isotope by endogenous amino acid biosynthesis. The second and greater concern is that the label will be scrambled to other types of residues, either by the specific metabolic conversion of one amino acid into another, or as a result of aminotransferase (transaminase) activity. Even a small amount of cross-labeling can present a serious problem for experiments that lead to the assignment of chemical shifts. Because several factors contribute to signal strength in an NMR experiment, the strongest signals may not always correspond to residues that were intended to be labeled.

Amino acid biosynthesis is regulated by feedback inhibition, and so both of these concerns can be mitigated to some degree by supplementing the growth medium with a high concentration of all 20 amino acids (e.g., Torchia et al., 1989; Yamazaki et al., 1991; Roth et al., 1992; Ramesh et al., 1994). A more satisfactory approach, however, is to use hosts that have been modified to contain the appropriate genetic lesions to control amino acid biosynthesis (e.g., Muchmore et al., 1989; Henry and Sykes, 1992; Yamasaki et al., 1992; Byeon et al., 1993; Muto et al., 1993). Because it is currently the only organism in which the pathways of amino acid metabolism are both firmly established and amenable to genetic manipulation, *Escherichia coli* is the preferred host for residue-specific labeling. Moreover, *E. coli* is readily cultivated in defined medium and methods for overproducing recombinant proteins in this organism are straightforward and reliable. To further advance the technology for residue-specific isotopic enrichment of recombinant proteins, I have assembled a collection of genetic lesions that can be used to manipulate amino acid biosynthesis in *E. coli*. Whereas some of these lesions already have found widespread use in the NMR community (see above), others have been generally unavailable until now. With these

*Present address: ABL Basic Research Program, NCI-FCRDC, P.O. Box B, 7th Street, Ft. Detrick, Frederick, MD 21702, U.S.A.

Abbreviations: ^{15}N-amino acid, α-^{15}N-amino acid; Cam^R, chloramphenicol-resistant; DPA, diaminopimelic acid; Hfr, high-frequency recombinant; LB, Luria broth; Kan^R, kanamycin resistant; P1, bacteriophage P1; pfu, plaque-forming units; Str^R, streptomycin-resistant; Tet^R, tetracycline-resistant.

0925-2738/$ 6.00 + 1.00 © 1996 ESCOM Science Publishers B.V.

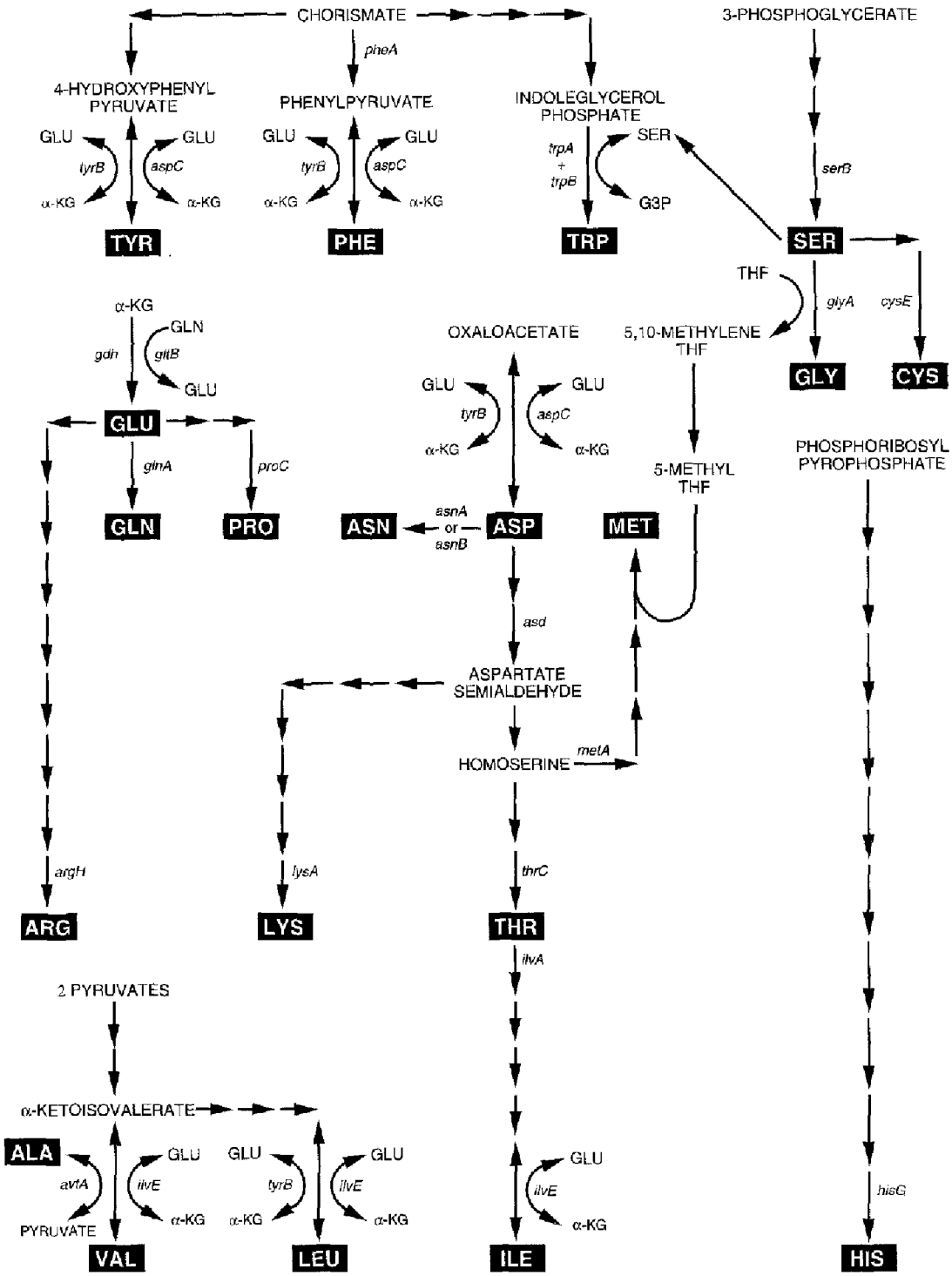

Fig. 1. Schematic view of amino acid biosynthesis in *Escherichia coli*. Pathways are summarized in schematic form, with each arrow representing a discrete biochemical transformation. Amino acids are abbreviated in the three-letter code. Genetic acronyms for defective genes that can be used to control amino acid biosynthesis (Table 1) are shown in italics adjacent to the chemical transformations catalyzed by the corresponding enzymes. See the text for further details. Other abbreviations used are: α-KG, α-ketoglutarate; G3P, glycerol-3-phosphate; THF, tetrahydrofolate.

tools, virtually any strain can be converted to an appropriate genotype for selective labeling.

Yet, why would one ever need to construct more than a single bacterial strain with a given genotype for selective labeling? Implicit in this question is the assumption that all laboratory strains of *E. coli* are essentially equivalent (isogenic), but this is clearly not the case (Bachmann, 1987). Consequently, the yield, solubility and chemical purity of a recombinant protein can vary widely in different host strains. At the very least, this necessitates that one or more pilot experiments be performed with each new bacterial strain before any expensive ^{15}N-amino acids

are utilized, and in some instances it may not be possible to identify suitable conditions for the production of a particular protein in the chosen host strain. Moreover, the choice of a bacterial strain may be dictated in part by other considerations, such as the requirement for an inducible T7 RNA polymerase gene in the chromosome (Studier et al., 1991) or the absence of an endogenous protease (e.g., Lon, Clp, etc.). Particularly when a series of selective labeling experiments are anticipated, it may ultimately be more efficient to begin by choosing a host strain that is well suited for overproduction of the target protein and convert it to the appropriate genotype(s) for selective labeling. The intent of this article is to describe, in a step-by-step fashion, how to construct new bacterial strains with ideal genotypes for almost any selective labeling experiment.

Materials and Methods

Bacterial strains

BL21/DE3 (Studier et al., 1991) was purchased from Novagen, Inc. (Madison, WI). EA1 (*asnA*, *asnB*, *aspC*, *tyrB*) and DL39avtA (Muchmore et al., 1989) were obtained from F. Dahlquist (University of Oregon). N5649 (Oppenheim et al., 1982) was a gift from M. Gottesmann (Columbia University). MM91 (Kumamoto and Beckwith, 1983) was kindly provided by J. Beckwith (Harvard University). JF436 (Felton et al., 1980) was obtained from A. Wright (Tufts School of Medicine). NK6860 [TN(*rrnD-rrnE*)$_1$, Str^R, Δ*lac-pro*, *leu*::Tn9, *hisG*::Tn10] was a gift from N. Kleckner (Harvard University). CBK007 (Shaw and Berg, 1979) was a gift from C. Berg (University of Connecticut). Other bacterial strains were either obtained from the *E. coli* Genetic Stock Center (Yale University), or constructed as described.

Preparation of P1 lysates

Bacteriophage P1$_{vir}$ was a gift from E. Raleigh (New England Biolabs). Phage stocks were prepared and titered as described (Miller, 1972). Titers ranged between 2×10^9 and 1×10^{11} plaque-forming units (pfu) per ml.

Generalized transduction of E. coli with P1 lysates

Approximately 10^8 pfu of a P1 lysate was added to 1 ml of cells grown in LB broth + 5 mM $CaCl_2$ ($A_{600}=0.5$; ca. 4×10^8 cells). The cells were incubated with the phage for 20 min in a 37 °C heat block, and then recovered by centrifugation for 4 min at 4000 rpm (ca. $1000\times g$) in a microcentrifuge. The pellets were resuspended in 1 ml of LB broth + 0.1 M sodium citrate, and then shaken vigorously for about 45 min at 37 °C. In most cases, the cells were then diluted 1:1 with 1 M sodium citrate, and either 0.1 or 0.2 ml of the dilution was spread on selective agar plates. However, if few or no transductants were obtained under these conditions, as was often the case when BL21/DE3 cells were infected with P1 lysates derived from strains of *E. coli* K-12, then instead the cells were concentrated by centrifugation, resuspended in 0.1 ml of 0.5 M sodium citrate, and spread on a single selective plate. Selective plates were composed of LB agar plus the appropriate antibiotic(s) at the following concentrations: chloramphenicol, 15 μg/ml; kanamycin, 25 μg/ml; tetracycline, 15 μg/ml.

Identification of transductants with the desired phenotype(s)

Transductants were purified by streaking twice on selective plates. In most cases, individual colonies were screened for the desired phenotype (auxotrophy) by streaking on M9 minimal agar medium (Miller, 1972) supplemented with the recommended concentrations (Davis et al., 1980) of all but one amino acid. This is how the cotransduction frequencies reported in Table 1 were measured. One exception is the *asd* lesion; strains that contained this mutation were identified by their inability to grow on LB agar plates (Miller, 1972) that do not contain diaminopimelic acid (10 μg/ml). Another exception is the *glyA*42::Tn5 lesion. The leaky phenotype of this allele necessitated that glycine auxotrophy be scored on minimal agar plates supplemented only with whichever amino acids were absolutely required for survival of the strain in minimal medium (e.g., tyrosine, phenylalanine, isoleucine, valine and leucine for DL39G). In some cases, additional nutrients were added to support the growth of strains with genetic defects that are unrelated to amino acid metabolism.

Results and Discussion

Wild-type *E. coli* can synthesize all 20 of the amino acids that are commonly found in proteins. A schematic view of the anabolic pathways in this organism, based on information compiled from several reviews and texts (Umbarger, 1978; Herrmann and Somerville, 1983; Bender, 1985; Ingraham et al., 1987), is presented in Fig. 1. It is chiefly the interplay between these pathways that poses a challenge for selective labeling. Some amino acids are the direct metabolic precursors of others, while some are substrates for one or more of the general aminotransferases. In either case, scrambling of the label is a potential threat. Dilution of the label by endogenous amino acid biosynthesis is a lesser concern because, in most cases, feedback regulatory mechanisms assure that the biosynthetic pathways will remain quiescent as long as there is a high concentration of amino acids in the medium.

Though it may not always be necessary to do so, the most reliable and efficient way to accomplish residue-specific labeling is to use bacterial hosts with the appropriate genetic defects to control amino acid metabolism. The collection of genetic lesions I have assembled for this

TABLE 1
MOBILE GENETIC DEFECTS IN AMINO ACID METABOLISM

Defective gene[a]	Defective enzyme[b]	Strain[c]	Donor/recipient[d]	Sex[e]	Selectable marker[f]	Cotransduction frequency[g]	Phenotype[h]	Counterselection[i]
argH	arginosuccinate lyase	MF28	JW383/W3678	F⁻	*zii*-510::Tn10	0.3	Arg⁻	
asd	aspartate semialdehyde dehydrogenase	MF7	TST3/CGSC#5081	Hfr	*malT*54::Tn10	0.5	DPA⁻	Mal⁺
asnA	asparagine synthetase A	KT2	SK2210/EA1	?	*zie*-296::Tn10	0.8		
asnB	asparagine synthetase B	JF436		F⁻	*asnB*50::Tn5	0.4		
aspC	aspartate aminotransferase	MF22	LCB273/DL39	F⁻	*aroA*273::Tn10	0.5	Trp⁻	
avtA	alanine-valine aminotransferase	DL39avtA		F⁻	*avtA*::Tn5	0.7		
cysE	serine acetyltransferase	MF32	NK6701/JM15	F⁻	*mtl*-16:Tn9	0.4	Cys⁻	Mtl⁺
		MM91		F⁻	*cysE*::Tn5	1.0	Cys⁻	
glnA	glutamine synthetase	MF29	RS3087/M5004	F⁻	*fad*-751:Tn10	0.3	Gln⁻	
glyA	serine hydroxymethyltransferase	DL39G		F⁻	*glyA*42::Tn5	1.0	Gly⁻	
		MF14	NK6042/AT2457	Hfr	*nadB*51::Tn10	0.1	Gly⁻	Nad⁺
hisG	ATP phosphoribosyl transferase	NK6860		F⁻	*hisG*::Tn10	1.0	His⁻	
ilvA	threonine deaminase	MF2	CBK007/Gif 41	Hfr	*ilvA*700::Tn5	1.0	Ile⁻	
ilvE	branched chain amino acid aminotransferase	MF15	RK4349/DL39	F⁻	*metE*163::Tn10	0.3	Ile⁻	Met⁺
lysA	diaminopimelate decarboxylase	MF36	X2904/KL334	Hfr	*thyA*748::Tn10	0.7	Lys⁻	Thy⁺
metA	homoserine succinyltransferase	MF21	TST1/DL41	F⁻	*malE*52::Tn10	0.1	Met⁻	Mal⁺
proC	1-pyrroline 5-carboxylate reductase	N5649		F⁻	*proC*::Tn10	0.8	Pro⁻	
serB	phosphoserine phosphatase	MF30	SK472/MF2	Hfr	*zjj*-202::Tn10	0.3	Ser⁻	Thr⁺
thrC	threonine synthase	MF30	SK472/MF2	Hfr	*zjj*-202::Tn10	0.1	Thr⁻	Ser⁺
trpB	tryptophan synthase	NK7402		F⁻	*trpB*83::Tn10	1.0	Trp⁻	
		LCB273		F'	*trp*::Tn5	0.9	Trp⁻	
tyrB	aromatic amino acid aminotransferase	MF16	TST1/DL39	F⁻	*malE*52::Tn10	0.5		Mal⁺

[a] The acronym for the defective gene in each strain (Bachmann, 1990).
[b] The name of the enzyme encoded by each defective gene.
[c] The name of the strain that contains both the defective gene and a cotransducible marker.
[d] When a strain was constructed by P1 transduction, the names of the donor (P1 lysate) and recipient strains are listed. Unless otherwise noted, all of the strains listed in this column can be obtained from the *E. coli* Genetic Stock Center at Yale University.
[e] The sex of each strain is indicated: F⁻ = female; F⁺, F' or Hfr = male.
[f] The defective gene in each strain is closely linked to a selectable transposon marker, with which it can be cotransduced at a reasonable frequency. In each case, the location of the transposon insertion is indicated either by the acronym for the gene that it disrupts (Bachmann, 1990) or by the three-letter nomenclature based on the genetic map (Chumley et al., 1979). In either case, the identity of the transposon marker also is indicated. For example, *asnB*50::Tn5 indicates that the selectable marker in this strain is a Tn5 insertion (kanamycin resistance) within or adjacent to the *asnB* locus. The selectable phenotypes associated with the different transposon markers are: Tn10, tetracycline resistance; Tn5, kanamycin resistance; Tn9, chloramphenicol resistance.
[g] Indicates the approximate frequency with which recombinants that have incorporated the selectable marker also inherit the defective gene.
[h] The phenotype (auxotrophy) associated with each defective gene in an otherwise wild-type genetic background is given, using the three-letter code for amino acids. For example, His⁻ indicates that a strain containing this defective gene is unable to grow on minimal medium that is not supplemented with histidine. Some of the defective genes have no discernible phenotype under these conditions, but can be detected in combination with other genetic defects (see text for discussion). DPA = diaminopimelic acid.
[i] In some cases, a counterselection can be employed to remove the selectable marker from a strain without always removing the associated genetic defect. The three-letter code for amino acids is employed. Other abbreviations used are: Mal⁺, selection for utilization of maltose as a carbon source; Mtl⁺, selection for utilization of mannitol as a carbon source; Nad⁺, selection for the ability to grow on minimal medium without nicotinic acid; Thy⁺, selection for the ability to grow on minimal medium without thymidine.

purpose is described in Table 1. In Fig. 1, the catalytic steps associated with each of these lesions are labeled with the corresponding genetic acronyms (Bachmann, 1990). Considering this collection of genetic lesions together with the information in Fig. 1, an ideal genotype for selective labeling with each kind of amino acid can be derived. A list of these genotypes is provided in Table 2. It is quite similar to one that appears in an excellent review article (Muchmore et al., 1989), which also includes a discussion of the rationale behind the various genotypes. It is important to note that, because multistep pathways engender a choice of genetic defects, in many cases several different genotypes can be considered ideal for selective labeling with a particular ^{15}N-amino acid. The list in Table 2 is restricted to those genotypes which can be created by combining the genetic lesions in Table 1.

To construct bacterial hosts with appropriate genotypes for selective labeling, defective genes must be moved from the chromosome of one strain to that of another. Two general methods are widely used in *E. coli*: general-

ized transduction (Miller, 1972; Margolin, 1987) and conjugational transfer (Miller, 1972). Both methods rely on homologous recombination (Weinstock, 1987) to incorporate genetic material from the donor into the chromosome of the recipient strain, but they differ in the way that the DNA is delivered to the cell. Generalized transduction uses a bacteriophage called P1, which occasionally packages chromosomal DNA instead of phage DNA inside its proteinaceous coat, as a vehicle for delivering DNA to recipient cells. In conjugational transfer, the donor (male, Hfr) and recipient (female, F^-) cells associate with each other via the F-pillus produced by the male strain, and then some or all of the chromosome from the donor strain is transferred to the recipient, where homologous recombination can ensue. Generalized transduction is considered to be a superior method for strain construction because, due to the finite size of the phage particle, only a relatively small amount of DNA (ca. 2% of the *E. coli* chromosome) can be exchanged between strains. In conjugational transfer, large portions of the chromosome are delivered to the recipient cells during mating of the two strains, and so a great many more genetic markers can be exchanged in a single cross.

It is not possible to select directly for recombinants that have incorporated the defective genes. Rather, one must rely on cotransduction of these genetic lesions with nearby selectable markers. First, a selection is imposed for recombinants that have inherited the nearby selectable marker (usually antibiotic resistance), and then these are screened individually to identify recombinants that have also inherited the defective gene. Transposable genetic elements make convenient makers for this purpose, because they can insert themselves at a variety of locations in the bacterial chromosome, and they usually confer a selectable phenotype. The locations of many random transposon insertions in the *E. coli* chromosome have been mapped with respect to nearby markers (Berg and Berg, 1987). Some strains with transposon insertions adjacent to defective alleles of enzymes involved in amino acid metabolism can be obtained from the *E. coli* Genetic Stock Center, and many others can be constructed in one step from readily available strains.

The collection of mobile genetic defects that I have used to manipulate amino acid metabolism in *E. coli* (Table 1) includes lesions in all of the biosynthetic pathways depicted in Fig. 1. Each genetic defect is closely linked to a selectable transposon marker so that the defective alleles are cotransduced with the selectable markers between 10% and 100% of the time. This makes it easy to identify individual transductants that have inherited the genetic defect in amino acid metabolism. Some of these selectable markers have the added advantage that they can be removed by a genetic counterselection in a subsequent transduction, which facilitates the construction of strains with multiple genetic defects (see below). Most of the genetic defects in my collection are linked to Tn10 insertions, simply because the parental strains were easy to obtain from the *E. coli* Genetic Stock Center. Similar lesions which are cotransducible with alternative markers like Tn5 have been described in the literature (e.g., Shaw and Berg, 1979; Berg and Berg, 1987), but these would have to be obtained from the individual investigators. A few of the strains in Table 1 are high-frequency recombinants (Hfr), so in some cases this mode of transmission also can be used to move the defective alleles into new strains.

Arg, Cys, Gln, Gly, His, Ile, Lys, Met, Pro, Thr

In many cases, only one lesion is needed to create an ideal genotype for the incorporation of a particular ^{15}N-amino acid. Except for threonine and isoleucine, all of the residues in this category lie at the ends of biosynthetic pathways, and are not substrates for the general aminotransferases. The others are: arginine, cysteine, glutamine, glycine, histidine, lysine, methionine, proline, and probably tryptophan as well (but see below). However, proline is rarely used for ^{15}N-labeling because it has no backbone N-H. New bacterial strains that have an ideal genotype

TABLE 2
IDEAL *E. coli* GENOTYPES FOR SELECTIVE ISOTOPIC LABELING OF PROTEINS WITH ^{15}N-AMINO ACIDS

Amino acid	Ideal genotype[a]
Ala[b]	*avtA, ilvE, aspC, tyrB*
Arg	*argH*
Asn	*asnA, asnB*
Asp	*aspC, tyrB, asnA, asnB, asd*
Cys	*cysE*
Gln	*glnA*
Glu	*gdh, gltB, argH, glnA, proC, ilvE, aspC, tyrB*
Gly	*glyA*
His	*hisG*
Ile	*ilvE*
Leu	*ilvE, tyrB*
Lys	*lysA*
Met	*metA*
Phe	*ilvE, aspC, tyrB*
Pro	*proC*
Ser	*cysE, glyA, serB, trpB*
Thr	*thrC*
Trp[c]	*trpB, tyrB*
Tyr	*tyrB, aspC*
Val	*ilvE, avtA*

[a] Ideal genotype identifies only the defective genes in each strain that are required for selective labeling with a particular ^{15}N-amino acid; irrelevant markers have been omitted for clarity. In some cases, such as when mutations are available in more than one step of the same pathway, equivalent genetic lesions can be substituted. The ideal genotypes listed here are derived from the collection of lesions in Table 1.

[b] Since there are no true alanine auxotrophs, the ideal genotype for labeling with this amino acid is incompletely defined.

[c] The importance of the *tyrB* lesion for the incorporation of tryptophan is debatable. See text for further details.

for selective labeling with any of these amino acids can easily be constructed in a single step by generalized transduction, using P1 lysates prepared from the appropriate strains in Table 1. In fact, these strains themselves might be suitable hosts for selective labeling. Although threonine is the direct metabolic precursor of isoleucine (Fig. 1), transfer of ^{15}N from the former to the latter is not a concern because the α-amino group of isoleucine is derived from glutamate. The *ilvA* lesion still may be of some utility, however, in so far as it would be expected to reduce or eliminate loss of label by catabolism, and a *thrC ilvA* double mutant would be very useful for selective labeling with ^{13}C-threonine. Once again, it is worth noting that for all of the residues in this category except isoleucine, where scrambling of the label by transamination is a potential threat, a wild-type strain could probably be used to good effect as long as a high concentration of the labeled amino acid is maintained in the medium. In certain cases, however, this strategy could prove to be prohibitively expensive.

I have worked with two different alleles of *glyA*: *glyA*6, which is linked to a Tn10 insertion in *nadB*, and *glyA*42, which is linked to a Tn5 insertion. I have found the *glyA*42::Tn5 mutation to be somewhat leaky. If there is a high concentration of serine in the growth medium, as is typically the case in a selective labeling experiment, then strains with this lesion do not need glycine in order to grow. There is some concern that this enzyme also may catalyze the reverse reaction to a significant degree in vivo, which could cause the label to be scrambled from glycine to serine. On the other hand, strains with the *glyA*6 allele absolutely require glycine for survival. *glyA*6 strains grow rather slowly, especially in minimal medium where they have a tendency to revert to Gly^-, while *glyA*42 strains are relatively robust and stable. I do not know if dilution of the isotope or cross-labeling of serine occurs in strains that contain the *glyA*42 allele. This allele is widely available in *E. coli* DL39G (LeMaster and Richards, 1988), but so far neither problem has been reported in the literature.

Ala, Asn, Asp, Glu, Leu, Phe, Trp, Tyr, Ser and Val

For the remaining amino acids, a combination of genetic lesions is required to guard against dilution and scrambling of the isotope. With the exception of asparagine, all of the residues in this category are either the direct metabolic precursors of other residues, substrates for more than one of the general aminotransferase activities in *E. coli*, or both. The simpler cases, where transamination is not involved, will be considered first. The residues that fall into this category are asparagine and serine. Next, strategies for constructing strains with the ideal genotypes for selective labeling with amino acids for which scrambling of the label by transamination is a significant concern will be discussed. These are leucine, tyrosine, phenylalanine, valine, aspartate and possibly tryptophan. Finally, alanine and glutamate, the two cases for which genetic strategies are either impractical or only partially effective, will be addressed.

Since it is neither the direct precursor of another amino acid nor a substrate for the general transaminase activities in *E. coli*, asparagine would be included in the previous category were it not for the fact that the products of two unlinked genes, *asnA* and *asnB*, are capable of catalyzing its biosynthesis from aspartate. The complication which arises is that neither mutation has any phenotype by itself; both genes must be disabled to make an asparagine auxotroph (Felton et al., 1980). Fortunately, this problem can be overcome without much difficulty because each of the defective genes is cotransduced at a high frequency with its associated marker. First, an $asnA^+asnB^+$ host is transduced to tetracycline resistance (Tet^R) with a P1 lysate from KT2 (Table 1). At this point, it is not possible to determine which of the transductants have inherited the defective *asnA* gene because all of them are still $asnB^+$. However, the cotransduction frequency (ca. 0.8) is high enough so that if one simply picks a few individual transductants at random and transduces each of these to kanamycin resistance (Kan^R) with a P1 lysate from JF436 (Table 1), then many of the recombinants obtained in the second round of transduction will be asparagine auxotrophs ($asnA^-$ and $asnB^-$). This procedure can also be performed in reverse order, by first transducing the new host to Kan^R and then to Tet^R.

Four genetic defects are required to create a strain that has an ideal genotype for selective labeling with ^{15}N-serine. One lesion (*serB*) is needed to prevent the biosynthesis of serine. A second lesion (*glyA*) prevents the metabolic conversion of serine into glycine. A third genetic defect (*cysE*) is required to block the transformation of serine into cysteine. Finally, because the α-amino group of tryptophan originates from serine, another lesion (*trpB*) is required to prevent this metabolic conversion. A choice of selectable markers is available for several of these genetic lesions. However, the *cysE* defect should be used in conjunction with the Tn9 marker (Cam^R) for this application. Then one can either use the *glyA*6 lesion, which is linked to a Tn10 (Tet^R) insertion, or the *glyA*42 allele associated with a Tn5 (Kan^R) insertion. I prefer the Tn10-linked allele, because the *glyA*42 allele is somewhat leaky (see above). If one chooses to use the *glyA*6 allele, then it makes sense also to use the trp^- lesion that is linked to Tn5 (Kan^R). On the other hand, if the *glyA*42 allele is selected, then it should be used in conjunction with the *trpB*::Tn10 lesion. Because the *serB* defect is also linked to a Tn10 insertion (Tet^R), at some point one of the Tn10 markers will have to be removed before another one can be introduced. As indicated in Table 1, the *nadB*51::Tn10 marker associated with the *glyA*6 lesion and the *zjj*-202::Tn10 marker associated with the *serB*

lesion can be removed by transduction to Nad$^+$ or Thr$^+$, respectively. Hence, one could first transduce a naive strain to Tet^R with a P1 lysate from MF30, identify a recombinant that is both Ser$^-$ and Thr$^-$ (about 10% of the transductants should have this phenotype), and then transduce it back to Thr$^+$. Some of these Thr$^+$ transductants will have lost the Tn10 marker but retained the *serB* lesion (Thr$^-$, Ser$^-$, Tet^S). Alternatively, after transducing a strain to Tet^R with a P1 lysate from MF14, one can remove the Tn10 associated with the *glyA6* allele by transduction back to Nad$^+$; about 90% of the Nad$^+$ transductants will have the desired phenotype (Gly$^-$, Tet^S). The order in which the four genetic defects are introduced is not critical, but since the *glyA6* allele retards the growth of strains that contain it and has a tendency to revert to Gly$^+$, this lesion probably should be introduced last.

The general aminotransferases (or transaminases) of *E. coli* pose a major problem for selective labeling with certain amino acids. These enzymes catalyze the transfer of α-amino groups from amino acids to the α-keto precursors of other residues, and the reactions are freely reversible (Umbarger, 1978; Herrmann and Somerville, 1983; Bender, 1985). Amino acids that are substrates for one or more of these enzymes in *E. coli* are alanine, valine, isoleucine, leucine, tyrosine, phenylalanine, aspartate, glutamate, and possibly tryptophan. Thus far, genetic lesions affecting four general aminotransferases in *E. coli* have been described; they are the products of the *ilvE*, *aspC*, *tyrB* and *avtA* genes. There also is compelling evidence for an alanine-glutamate aminotransferase (Wang et al., 1987), but thus far no genetic defect in this activity has been reported. A further complication arises because of the overlapping specificities of some of these enzymes (Gelfand and Steinberg, 1977; Hayashi et al., 1993; Miyazawa et al., 1994). Frequently more than one aminotransferase gene must be disabled in order to prevent scrambling of the label from a particular amino acid.

Two of the general aminotransferases in *E. coli*, the products of the *ilvE* and *tyrB* genes, must be eliminated to prevent scrambling of the label from ^{15}N-leucine to other types of residues. By itself, a *tyrB* mutation does not confer auxotrophy upon the host, but an *ilvE* mutation results in a requirement for isoleucine, and an *ilvE tyrB* double mutant is a leucine auxotroph. Thus, the new strain is first transduced to Tet^R with a P1 lysate from MF15 (Table 1), and cotransduction of the *ilvE* lesion is confirmed by screening the individual transductants for isoleucine auxotrophy. Because the *tyrB* lesion is also associated with a Tn10 (Tet^R) marker, the Tn10 that is linked to the *ilvE* defect must be removed before the *tyrB* mutation can be combined with the *ilvE* lesion. This can be accomplished by transducing the *ilvE*$^-$ (Tet^R) strain to Met$^-$ with a P1 lysate from any *metE*$^+$ strain (but preferably the progenitor of the *ilvE*$^-$ transductant), and then screening the individual transductants for isoleucine auxotrophy to identify recombinants that have retained the *ilvE* allele but lost the Tn10 marker. About 70% of the transductants will have this phenotype. Now the *ilvE*$^-$, Tet^S strain can be transduced to Tet^R with a P1 lysate from MF16 (Table 1), and the individual transductants screened for leucine auxotrophy. This phenotype must be scored on plates that lack both leucine and tyrosine because the synthesis of the *tyrB* gene product is repressed by tyrosine (Gelfand and Steinberg, 1977). It appears that most leucine biosynthesis can be attributed to the action of IlvE, because *ilvE*$^-$ strains grow very slowly in the absence of leucine. Similarly, *tyrB* mutants grow more slowly than *tyrB*$^+$ strains on plates that lack tyrosine, and this phenotype is also useful for identifying recombinants. In practice, sometimes it can be difficult to distinguish between *tyrB*$^+$ and *tyrB*$^-$ genotypes in an *ilvE*$^-$ background, and so both methods for screening probably should be employed.

Because of the overlapping specificities of the general aminotransferases, multiple genetic lesions are required for selective labeling with the aromatic amino acids tyrosine (*tyrB*, *aspC*) and phenylalanine (*tyrB*, *aspC*, *ilvE*) . Once again, the problem is that in an otherwise wild-type genetic background, neither the *tyrB* nor the *aspC* mutation by itself confers amino acid auxotrophy on the host. However, starting with an *ilvE tyrB*$^-$ leucine auxotroph (above), the *aspC* lesion can be introduced in two steps. First the Tn10 associated with the mutant *tyrB* allele is removed by transducing the *ilvE*$^-$, *tyrB* (Tet^R) strain to Mal$^+$ (ability to utilize maltose as the sole carbon source), and then screening the recombinants to verify that they are still leucine auxotrophs (i.e. that they have not also lost the defective *tyrB* allele). Again, the best strain to use for this purpose would be the immediate progenitor of the *ilvE*$^-$*tyrB*$^-$ mutant. In practice, I have found this selection to be very leaky; essentially all the cells that are not killed by the phage are able to grow on minimal agar plates supplemented with maltose instead of glucose if amino acids are also included in the medium. Nevertheless, authentic Mal$^+$ transductants can be identified because they grow more rapidly and form larger colonies than Mal$^-$ cells do under these conditions. It is important to purify the putative Mal$^+$ transductants at least twice by streaking on selective plates, until homogeneous large colonies are obtained, before streak testing the isolates for the Tet^S and Leu$^-$ phenotypes. The *ilvE*$^-$, *tyrB*$^-$ (Tet^S) strain then can be transduced to Tet^R with a P1 lysate from MF22 (Table 1), and the transductants screened for aspartate auxotrophy. Cells become aspartate auxotrophs only when both the *tyrB* and *aspC* genes are disabled by mutation. Hence, the emergence of aspartate auxotrophy at this stage serves to confirm that the strain is indeed *tyrB*$^-$. If desired, the *aroA*::Tn10 insertion associated with the defective *aspC* allele can be removed by a subsequent transduction to Trp$^-$ (select transductants on minimal

agar plates that lack tryptophan), because the *aroA* gene product is required for tryptophan biosynthesis (Pittard and Wallace, 1966). DL39, an *ilvE*$^-$, *tyrB*$^-$, *aspC*$^-$ derivative of *E. coli* MG1655, was constructed in precisely this fashion (LeMaster and Richards, 1988).

In order to label proteins with ^{15}N-valine, both aminotransferases that can utilize this amino acid as a substrate must be inactivated. One of these, the product of the *ilvE* gene, is the same enzyme that is required for the biosynthesis of isoleucine. The other transaminase is the product of the *avtA* gene. This enzyme, the alanine-valine aminotransferase, catalyzes the interconversion of alanine and valine, and appears to be a major route for alanine biosynthesis. On its own, an *avtA* mutant has no discernible phenotype because there are alternative routes to both isoleucine (*ilvE*) and alanine (see below). Thus, the *ilvE* defect must be introduced first and the recombinants identified by isoleucine auxotrophy, as described above. Then the introduction of the *avtA* defect in a second round of transduction will give rise to valine auxotrophy in the *ilvE*$^-$ background.

Although only one genetic defect is needed to generate tryptophan auxotrophs (e.g. *trpB*$^-$ in Table 1), this amino acid evidently can serve as a substrate for the aromatic amino acid aminotransferase in vitro (Hayashi et al., 1993; Miyazawa et al., 1994). Hence, while dilution of the label is of no concern in a *trpB*$^-$ auxotroph, scrambling of ^{15}N from tryptophan to other residues is still a potential threat. The fact that *trpB* mutants require tryptophan for viability would seem to argue that little if any tryptophan is generated by TyrB in vivo. Nevertheless, it may be wise to use a *trpB*$^-$ derivative of a transaminase deficient strain (i.e. *trpB*$^-$, *tyrB*$^-$) for labeling with this amino acid.

A minimum of five genetic defects are required to construct a strain with an ideal genotype for labeling with ^{15}N-aspartate. A convenient starting point is the *ilvE*, *tyrB*, *aspC* triple-mutant from which the Tn10 in *aroA* has been removed by transduction to Trp$^+$ (see above). The *tyrB* and *aspC* defects together will prevent the biosynthesis of aspartate and keep the label from being scrambled to other residues by these aminotransferases. However, the pathways leading from aspartate to its direct metabolic derivatives asparagine, lysine, threonine and methionine also must be blocked. A convenient way to prevent the label from being incorporated into lysine, methionine and threonine is to take advantage of the *asd* lesion, which lies upstream of the branchpoints for all three pathways (Fig. 1). *Asd* mutants are unable to synthesize diaminopimelic acid (DPA), which is required for cell wall synthesis, so this compound must be added to the medium. In fact, the DPA requirement serves as a useful test for the presence of the *asd* lesion in any strain. Finally, the pathway leading from aspartate to asparagine also must be disabled by mutation. As discussed above, two genetic defects are necessary to accomplish this. Since one of these (*asnA*) is linked to a Tn10 marker, the *malA*::Tn10 associated with the *asd* lesion must be removed by transduction to Mal$^+$ before the *asnA* defect can be introduced. Having accomplished this, the *asnA* and *asnB* defects can be introduced in either order, as described above.

Selective labeling with alanine presents a problem because there are no true alanine auxotrophs. Evidently there are several pathways by which this residue can be synthesized in *E. coli*. The alanine-valine aminotransferase, the product of the *avtA* gene, appears to be one of the major routes for alanine biosynthesis, and some of the other general aminotransferases also may be capable of producing alanine by transamination of pyruvate (Herrmann and Somerville, 1983). Another genetic locus that appears to be involved in the biosynthesis of alanine has been identified (Wang et al., 1987), but not characterized further. This locus may encode the glutamate-alanine aminotransferase postulated to be responsible for the scrambling of label from ^{15}N-glutamate to alanine (Muchmore et al., 1989). Currently, the best option is to use a strain with defects in all four of the general aminotransferases (*ilvE*, *tyrB*, *aspC* and *avtA*). Enrichment of T4 lysozyme with ^{15}N-alanine has been performed in an *avtA*$^-$ strain without reported transfer of label (Muchmore et al., 1989; Muto et al., 1993). Moreover, other investigators have described conditions under which ^{15}N-alanine can be incorporated into various proteins without using a transaminase deficient strain, and without scrambling of the label (Yamazaki et al., 1991; Henry and Sykes, 1992; Ramesh et al., 1994).

Because of its central role in amino acid metabolism and nitrogen assimilation, labeling with ^{15}N-glutamate is not recommended. Since so many different genetic defects would be required to guard against scrambling of the label (Table 2), it is uncertain whether such a strain would be viable, even if it could be constructed. Some investigators have reported using prototrophic (Yamazaki et al., 1991; Ramesh et al., 1994) or transaminase-deficient strains (Muchmore et al., 1989) of *E. coli* for the incorporation of ^{15}N-glutamate, but a significant degree of cross-labeling was observed in these cases.

Practically speaking, a collection of only a few key strains can solve most selective labeling problems. As discussed above, as long as there is a high concentration of amino acids in the growth medium, probably any strain can be used to label recombinant proteins with the ^{15}N-amino acids arginine, cysteine, glutamine, glycine, histidine, lysine, methionine, proline, threonine and perhaps tryptophan, although the use of auxotrophs is likely to be more cost-effective. A strain with defective alleles of all four general aminotransferases (*ilvE*, *tyrB*, *aspC*, and *avtA*) is ideal for selective labeling with isoleucine, leucine, valine, tyrosine, phenylalanine and alanine. The *ilvE*, *tyrB*, *aspC*, *asd*, *asnA*, *asnB* genotype, which is ideal for

the incorporation of aspartate, also can be used to label proteins with isoleucine, leucine, tyrosine, phenylalanine, asparagine, lysine, methionine and threonine, thus making it particularly valuable. A *serB, glyA, cysE, trpB* strain can be used to label proteins with cysteine and glycine as well with serine, although in this case the *glyA6* allele could prove to be a liability for labeling with cysteine, since the strain will grow more slowly and the yield of recombinant protein may be lower than in a *cysE* mutant or a wild-type strain.

Conclusions

In principle, the collection of genetic tools that I have described can be used to convert virtually any strain of *E. coli* to the appropriate genotype for selective labeling with ^{15}N-amino acids, provided only that it can be infected by bacteriophage P1. Indeed, using the approaches outlined above, I have been able to construct nearly complete sets of otherwise isogenic *E. coli* K-12 (MG1655) and *E. coli* B (BL21) strains that, collectively, include all of the genotypes listed in Table 2, except that which is ideal for labeling with glutamate. Thus far, derivatives of the latter strain have been used to produce samples of the catalytic domain of human stromelysin selectively labeled with Gly, His, Leu, Phe and Thr without any significant scrambling of the label (details to be published elsewhere).

Acknowledgements

I thank Kwei-Lan Tsao, Tri Li, and Melanie Fleming for expert technical assistance, and Elena Rios-Fredrickson for assistance with the preparation of this manuscript. I am especially grateful to Barbara Bachmann and the staff of the *E. coli* Genetic Stock Center at Yale University; this work would not have been possible without their help.

References

Abeygunawardana, C., Weber, D.J., Frick, D.N., Bessman, M.J. and Mildvan, A.S. (1993) *Biochemistry*, **32**, 13071–13080.

Bachmann, B. (1987) In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Vol. 2 (Eds, Ingraham, J.L., Neidhardt, F.C., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E.), American Society for Microbiology, Washington, DC, U.S.A., pp. 1190–1219.

Bachmann, B. (1990) *Microbiol. Rev.*, **54**, 130–197.

Bender, D.A. (Ed.) (1985) *Amino Acid Metabolism*, Wiley, New York, NY, U.S.A.

Berg, C.M. and Berg, D.E. (1987) In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Vol. 2 (Eds, Ingraham, J.L., Neidhardt, F.C., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E.), American Society for Microbiology, Washington, DC, U.S.A., pp. 1071–1109.

Byeon, I.J., Yan, H., Edison, A.S., Mooberry, E.S., Abildgaard, F., Markley, J.L. and Tsai, M.D. (1993) *Biochemistry*, **32**, 12508–12521.

Chumley, F.G., Menzel, R. and Roth, J.R. (1979) *Genetics*, **91**, 639–655.

Davis, R.W., Botstein, D. and Roth, J.R. (1980) *Advanced Bacterial Genetics: A Manual for Genetic Engineering*, Cold Spring Harbor Press, Cold Spring Harbor, NY, U.S.A., p. 207.

Felton, J., Michaelis, S. and Wright, A. (1980) *J. Bacteriol.*, **142**, 221–228.

Gelfand, D.H. and Steinberg, R.A. (1977) *J. Bacteriol.*, **130**, 429–440.

Hayashi, H., Inoue, K., Nagata, T., Kuramitsu, S. and Kagamiyama, H. (1993) *Biochemistry*, **32**, 12229–12239.

Henry, G.D. and Sykes, B.D. (1992) *Biochemistry*, **31**, 5284–5297.

Herrmann, K.M. and Somerville, R.L. (Eds.) (1983) *Amino Acids: Biosynthesis and Genetic Regulation*, Addison-Wesley, London, U.K.

Ingraham, J.L., Neidhardt, F.C., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. (Eds.) (1987) *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Vol. 2, American Society for Microbiology, Washington, DC, U.S.A.

Kumamoto, C.A. and Beckwith, J. (1983) *J. Bacteriol.*, **154**, 253–260.

LeMaster, D.M. and Richards, F.M. (1988) *Biochemistry*, **27**, 142–150.

Margolin, P. (1987) In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Vol. 2 (Eds, Ingraham, J.L., Neidhardt, F.C., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E.), American Society for Microbiology, Washington, DC, U.S.A., pp. 1154–1168.

Miller, J.H. (Ed.) (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Press, Cold Spring Harbor, NY, U.S.A.

Miyazawa, K., Kawaguchi, S., Okamoto, A., Kato, R., Ogawa, T. and Kuramitsu, S. (1994) *J. Biochem.*, **115**, 568–577.

Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) *Methods Enzymol.*, **177**, 44–73.

Muto, Y., Yamasaki, K., Ito, Y., Yajima, S., Masaki, H., Uozumi, T., Walchli, M., Nishimura, S., Miyazawa, T. and Yokoyama, S. (1993) *J. Biomol. NMR*, **3**, 165–184.

Oppenheim, A.B., Gottesman, S. and Gottesman, M. (1982) *J. Mol. Biol.*, **158**, 327–346.

Pittard, J. and Wallace, B.J. (1966) *J. Bacteriol.*, **91**, 1494–1508.

Ramesh, B., Frederick, R.O., Syed, S.E.H., Gibson, C.F., Yang, J. and Roberts, G.C.K. (1994) *Eur. J. Biochem.*, **225**, 601–608.

Roth, S.M., Schneider, D.M., Strobel, L.A., Van Berkum, M.R.A., Means, A.R. and Wand, A.J. (1992) *Biochemistry*, **31**, 1443–1451.

Shaw, K.J. and Berg, C.M. (1979) *Genetics*, **92**, 741–747.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1991) *Methods Enzymol.*, **185**, 60–89.

Torchia, D.A., Sparks, S.W. and Bax, E. (1989) *Biochemistry*, **28**, 5509–5524.

Umbarger, H.E. (1978) *Annu. Rev. Biochem.*, **47**, 533–606.

Wang, M.D., Buckley, L. and Berg, C.M. (1987) *J. Bacteriol.*, **169**, 5610–5614.

Weinstock, G.M. (1987) In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Vol. 2 (Eds, Ingraham, J.L., Neidhardt, F.C., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E.), American Society for Microbiology, Washington, DC, U.S.A., pp. 1034–1043.

Yamasaki, K., Muto, Y., Ito, Y., Walchli, M., Miyazawa, T., Nishimura, S. and Yokoyama, S. (1992) *J. Biomol. NMR*, **2**, 71–82.

Yamazaki, T., Yoshida, M., Kanaya, S., Nakamura, H. and Nagayama, K. (1991) *Biochemistry*, **30**, 6036–6047.