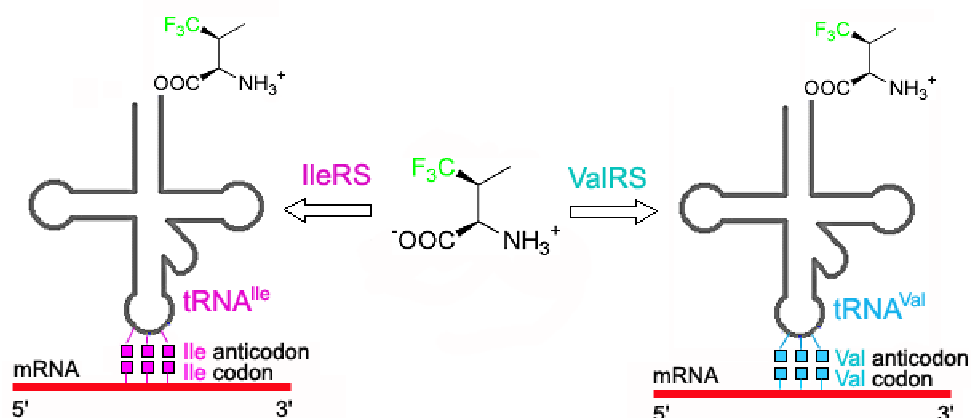
Chapter 4

It Depends on How You Read It: Alternative Translations of a Single RNA Message

Portions of this chapter are adapted from a submitted paper by Pin Wang, Alfio Fichera, Krishna Kumar and David A. Tirrell

Abstract

The genetic code links the sequence of a messenger RNA to the sequence of the corresponding protein. Codon assignments are established in the aminoacylation step, wherein the aminoacyl-tRNA synthetases catalyze the attachment of the amino acids to their cognate tRNAs. Here we show that engineering of bacterial expression hosts can allow a single RNA message to be read in different ways, depending on the relative rates of competing aminoacylation reactions. Specifically, we show that the *2S,3R*-form of 4,4,4-trifluorovaline can be assigned either to isoleucine or to valine codons, depending on whether the bacterial host overexpresses the isoleucyl- or the valyl-tRNA synthetase.



1. Introduction

The genetic code links the sequence of a messenger RNA to the sequence of the corresponding protein. Codon assignments are established in the aminoacylation step, wherein the aminoacyl-tRNA synthetases (aaRSs) catalyze the attachment of the amino acids to their cognate tRNAs (*1*). Here we show that engineering of bacterial expression hosts can allow a single RNA message to be read in two different ways, depending on the relative rates of competing aminoacylation reactions.

This work was motivated by previous demonstrations that fluorinated amino acids can be used to engineer the stability and dimerization specificity of coiled-coil peptides and proteins (*2-8*). Trifluoroleucine and hexafluoroleucine can be introduced into recombinant proteins by using a leucine auxotroph as the bacterial expression host (*7*). High-level incorporation of either of the fluorinated amino acids requires depletion of leucine in the culture medium prior to induction of gene expression. For hexafluoroleucine, depletion of the natural amino acid is not enough; efficient incorporation demands the use of an engineered bacterial host with enhanced leucyl-tRNA synthetase activity (*5*). One of our continuing objectives is to expand the availability of fluorinated amino acids for use in protein engineering.

2. Materials and Methods

2.1. Materials

Amino acids **1** and **2** were obtained from Sigma (St. Louis, MO). **5** were bought from Matrix Scientific (Columbia, SC). **3** and **4** were synthesized and purified by Xing et al (*9*). [^{32}P]-labeled sodium pyrophosphate was purchased from NEN Life Sciences.

Oligonucleotides were synthesized at the Caltech Biopolymer Synthesis Center. General cloning was performed in XL-1 blue cells from Stratagene (San Diego, CA). pQE vector system and Ni-NTA purification system for His tagged proteins were purchased from Qiagen.

2.2. Synthetase Cloning, Expression and Purification

The cloning of *E. coli* isoleucyl-tRNA synthetase (IleRS) was described in reference (*10*); the resulting plasmid is designated as pQE-ileS. A similar cloning scheme was used to clone *E. coli* valyl-tRNA synthetase (ValRS). *E. coli* genomic DNA was prepared by using the DNEasy kit from Qiagen (Chatsworth, CA). The ValRS gene was amplified from genomic DNA by PCR using the following primers: 5'-GCT CAA CCT GAA TAC GGA GCT CTG GAA AAA TGG-3' (valSSac1fw); 5'-GGG AGT TAT GCC TTC TTG GTA CCA TTT TCT GTA AGA G-3' (valSKpn1bw). *Pfx* polymerase was used for PCR to ensure high fidelity amplification. The resulting 2900 base pair DNA fragment was ligated into the expression plasmid pQE31 through the *Sac*I and *Kpn*I sites to yield pQE-valS. The integrity of the cloned gene was confirmed by DNA sequencing. The cloned synthetase has the N-terminal leader sequence MRGSHHHHHHTDPHASSGK.

pQE-ileS and pQE-valS were individually transformed into XL-1 blue strain (Stratagene) to afford the expression strains XL[pQE-ileS] and XL[pQE-valS] respectively. Synthetase expressions were conducted in LB media and induced at OD_{600}=0.6 with 1 mM final concentration of IPTG. After three hours, the cells were harvested and lysed with sonication. The enzymes were purified using Ni-NTA agarose

resin under native conditions according to the manufacturer's protocol. Residual imidazole was removed on an ion-exchange column and proteins were eluted into Tris buffer (50 mM Tris-HCl, pH=7.4, 1 mM DTT)/50% wt glycerol. Aliquots were stored at –80 °C. The concentrations of the purified proteins were measured by absorbance at 280 nm under denaturing conditions.

2.3. Amino Acid Activation Assays

IleRS and ValRS activities were evaluated by the amino acid-dependent ATP-PP_i exchange reaction as described previously (*11, 12*). The assay measures the formation of $^{[32]}$P-labeled ATP by the enzyme-catalyzed exchange of $^{[32]}$P-pyrophosphate (PP_i). In a 200 μL reaction buffer (50 mM HEPES (pH 7.6), 20 mM $MgCl_2$, 1 mM DTT, 2 mM ATP and 2 mM [^{32}P]-PP_i (0.5 TBq/mol, in $NaPP_i$ form)), enzyme was added to a concentration of 100 nM and amino acid to a concentration of 3 μM to 10 mM (dependent on the rate of activation of the amino acid). At successive time point, aliquots (20 μL) were quenched into 500 μL quench solution (200 mM PP_i, 7% w/v $HClO_4$ and 3% w/v activated charcoal). The charcoal was washed twice with 10 mM PP_i and 0.5% $HClO_4$, and the absorbed ATP was counted via liquid scintillation counting. Kinetic parameters (k_{cat} and K_m) were determined by non-linear curve fitting of the data to a Michaelis-Menten model. The reported numbers in Table 4-1 are averages from duplicate experiments.

2.4. Plasmid Construction for *In Vivo* Assays of Codon Assignment

The IleRS and ValRS genes along with their endogenous promoters were amplified directly from *E. coli* genomic DNA. The primers used for IleRS were: 5'-GCC GGT ATT CGC CAG CTA GCT GGA AGT GCA TTT G-3' (ileSNhe1fw); 5'-GAG GAT CAG GTA TTT GCT AGC CCA GAT CGA TAA TCA G-3' (ileSNhe1bw). The primers used for ValRS were: 5'-GTC TGC GAA CAA GCT AGC AGA TTT TGC CAC-3' (valSNhe1fw); 5'-CCA GAT AAA GGC TTG CTA GCC AGT ATT TCA CGG G-3' (valSNhe1bw). The PCR conditions were 50 ng/100 μL of template DNA, 300 ng/100 μL of each primer, 95 ^{o}C denaturing temperature (0.5 min), 55 ^{o}C annealing temperature (1 min), 68 ^{o}C extension temperature (3 min). The resulting DNA fragments were digested by *Nhe*I and ligated into pQE15 (Qiagen) through the *Nhe*I site (pQE15 encodes murine dihydrofolate reductase (mDHFR) under control of a bacteriophage T5 promoter). The orientation of the inserts was checked by restriction enzyme digestion; plasmids containing the inserted genes with the same coding direction as the ORFs of mDHFR were selected. This orientation ensures that the transcription terminator sequence in the vector will be placed immediately downstream of IleRS or ValRS gene. The resulting vectors were denoted pQE15-ileS and pQE15-valS respectively. The isoleucine and valine double auxotrophic strain AIV-IQ (*10*) was individually transformed with pQE15-ileS and pQE15-valS to yield the expression strains AIV-IQ[pQE15-ileS] and AIV-IQ[pQE15-valS] respectively.

2.5. *In Vivo* Translational Assays of Codon Assignment

M9 medium supplemented with glucose (0.2 wt%), $MgSO_4$ (1 mM), $CaCl_2$ (0.1 mM), thiamine (5 mg/L), 20 natural amino acids and antibiotics (ampicillin and

chloramphenicol) was inoculated individually with overnight cultures of expression strains AIV-IQ[pQE15-ileS] and AIV-IQ[pQE15-valS]. When OD_{600} reached 0.8-1.0, the cultures were sedimented by centrifugation at 4 ^{o}C and the cell pellets were washed twice with 0.9% NaCl solution (sterilized). The cells were then resuspended in M9 minimal medium as described above, but without one natural amino acid (**1** or **2**) and supplemented with **3**, or **4** at a concentration of 250mg/L. After 10 min incubation, IPTG was added to a concentration of 1 mM to induce protein expression. Cells were harvested by centrifugation after 4 hour and lysed in buffer A (8 M urea, pH=8, 100 mM NaH_2PO_4, and 10 mM Tris). After a freeze/thaw cycle, the samples were subjected to SDS-PAGE analysis to evaluate protein expression. Samples of the test protein mDHFR were purified by nickel-affinity chromatography on Ni-NTA spin columns according to the manufacturer's instructions (Qiagen).

2.6. Amino Acid Identification at Valine and Isoleucine Codon Sites

Amino acid analysis, MALDI-MS and LC-MS/MS analysis of tryptic peptides were employed to identify the amino acids at the codon sites of interest. For amino acid analysis, the purified protein samples in 8 M urea were subjected to buffer exchange against water by ultrafiltration (Millipore, MWCO=5000) and submitted to the Molecular Structure Facility at the University of California at Davis for analysis. Quantitation was carried out based on standard chromatograms of the HCl (6 N) hydrolysis products.

For MALDI-MS analysis, 10 μL of protein sample in elution buffer (8 M urea, pH=8, 100 mM NaH_2PO_4, 10 mM Tris-HCl) was added to 90 μL of NH_4OAc (50 mM) solution. Modified trypsin (2 μL, Promega, 0.2 mg/mL) was added and the mixture was

digested overnight. Trifluoroacetic acid (TFA) was used to quench the reaction (pH<4.0) and the sample was subjected to chromatography on a $ZipTip_{C18}$ column (Millipore). The peptides were eluted into 3 μL 50% CH_3CN, of which 1 μL was added into 10 μL matrix solution (α-cyano-β-hydroxycinnamic acid, 10 mg/mL in 50% CH_3CN). The samples were analyzed on an Applied Biosystems Voyager DE Pro instrument.

LC-MS/MS analysis of tryptic peptides was carried out on a Finnigan LCQ ion trap mass spectrometer with HPLC pump and ESI probe. The purified tryptic peptide solution (3 μL) (eluted from a $ZipTip_{C18}$ column as above) was diluted with 25 μL of distilled water and injected into the HPLC. Peptides were separated on a C18 HPLC column and eluted at a flow rate of 30 μL/min using a linear gradient of 0-60% of solvent B (CH_3CN) in solvent A (H_2O with 0.1% TFA) over 45 min. The column eluent was directed to the electrospray source and tandem mass sequencing was carried out by fragmentation of the precursor ion with m/z corresponding to the desired tryptic fragments.

3. Results and Discussion

There are two stereoisomeric forms of L-3,3,3-trifluorovaline (L-Tfv): the *2S,3S*-form (**3**) and *2S,3R*-form (**4**). Previously we found that both the *2S,4S*- and the *2S,4R*-forms of 5,5,5-trifluoroleucine are readily incorporated into recombinant proteins; the leucyl-tRNA synthetase of *Escherichia coli* apparently activates both isomers (*7*). In evaluating L-Tfv as a substrate for protein synthesis, we considered two related questions. First, would L-Tfv function more efficiently as a surrogate for valine (**2**) or

for isoleucine (**1**)? Second, would either the valyl-tRNA synthetase or the isoleucyl-tRNA synthetase discriminate between the two diastereomeric forms of L-Tfv?

Initial efforts to incorporate **3**, **4** or **5** into recombinant proteins failed. Neither **3** nor **4** supported protein synthesis in *E. coli* cultures depleted of either isoleucine or valine. We then examined the kinetics of activation of **3** and **4** by the valyl- and isoleucyl-tRNA synthetases (*13*). The requisite genes were obtained by PCR amplification of chromosomal DNA of *E. coli*. A hexahistidine tag was placed at the amino terminus of each enzyme, and the His-tagged variants of the wild-type IleRS and ValRS were expressed in *E. coli* and purified by immobilized metal affinity chromatography. Amino acid activation kinetics were determined *in vitro* via the ATP-PP_i exchange assay (*11*). The results were striking; **4** is activated by both enzymes, **3** by neither (Table 4-1). When we compared the rates of activation by IleRS, we found the specificity constant k_{cat}/K_m for **4** to be reduced approximately 600-fold with respect to the native substrate **1**; for ValRS, k_{cat}/K_m was reduced approximately 2500-fold with respect to **2**.

In our previous work on incorporation of methionine and leucine analogues into recombinant proteins, we found that elevation of cellular aaRS activity could facilitate introduction of amino acids that are poor substrates for the wild-type synthetases (*5, 14*). The results shown in Table 4-1 encouraged us to test this approach with **4**. Expression plasmids harboring either of the synthetase genes *ileS* and *valS*, each outfitted with its endogenous promoter, were constructed from pQE15 and designated pQE15-ileS and pQE15-valS, respectively. Assays of whole-cell lysates from the corresponding transformants (designated AIV-IQ[pQE15-ileS] and AIV-IQ[pQE-valS], respectively

(*10, 15*)) showed enhanced reactivity toward the cognate amino acids (for AIV[pQE15-ileS], total synthetase activity elevated approximately 8-fold; for AIV[pQE15-valS], total synthetase activity elevated approximately 6-fold).

Murine dihydrofolate reductase (mDHFR), a marker protein encoded in pQE15, served as the test protein for experiments on *in vivo* incorporation of **3** and **4**. Assays were performed in *E. coli* cultures depleted in the canonical amino acid and supplemented with the analogue of interest. As shown in Figure 4-1, strains bearing either pQE15-ileS or pQE15-valS supported protein biosynthesis in media supplemented with **4**; cultures enriched in **3** showed only background levels of mDHFR (*16*). Amino acid analysis indicated that approximately 92% of the encoded isoleucine residues were replaced by **4** in mDHFR produced by the strain outfitted with additional copies of IleRS; approximately 86% of encoded valine residues were replaced by **4** in the strain outfitted with additional copies of ValRS. "Reassignment" of the isoleucine and valine codons was confirmed by MALDI-MS analysis of tryptic fragments of mDHFR (Figure 4-2). The fragment A corresponding to residues 124-147 contains a single residue encoded as isoleucine and four encoded as valine. When this fragment is derived from mDHFR produced in pQE15-ileS cultures supplemented with **4**, reassignment of the isoleucine codons to **4** causes the expected mass shift of +40 Da (Figure 4-2a). In contrast, reassignment of the valine codons causes a mass increase of 54 Da for fragment B derived from mDHFR expressed in the strain bearing pQE15-valS.

Liquid chromatography tandem mass spectrometry of the trypic peptides was employed to further confirm the codon reassignments. The precursor ion at m/z 485.91 Da, which corresponds to the doubly charged ion of the peptide QNLVI*MGR derived

from mDHFR expressed by the *E. coli* hosts overexpressing IleRS in media supplemented with **4**, is separated and fragmented with an ion trap mass spectrometer. As shown in Figure 4-3, the sequence information obtained from the spectrum clearly reveals the assignment of the isoleucine codon to **4**; we could find no indication of any other amino acid at the isoleucine site after carefully checking all possible peptides with the mass that could be associated with this peptide, suggesting that the fidelity of this assignment is better than 95%. Similarly, for the peptides derived from mDHFR expressed by the *E. coli* hosts overexpressing ValRS in media supplemented with **4**, we analyzed a precursor ion at m/z 492.91 Da, which likely corresponds to the doubly charged ion of peptide QNLV*IMGR (Figure 4-4). The sequence assignment indicates that the valine codon is unambiguously decoded by **4**.

Our finding of the convergence of the two structurally related aminoacyl-tRNA synthetases with one non-canonical amino acid may enlarge our understanding of catalytic mechanisms and substrate recognition properties of this class of enzymes. Both IleRS and ValRS require editing mechanisms to maintain their sufficient fidelity (*17*). According to the "double sieve" model of editing and structural studies, a synthetic active site for both synthetases excludes binding of amino acids larger than their cognate substrates, whereas a separate editing site ensures the mischarged amino acids to be corrected (*17-19*). Our *in vitro* assay showed that both enzymes are able to discriminate the isomers of Tfv at catalytic site and only *2S,3R*-form (**4**) is activated, which was not observed from LeuRS charging 5,5,5-trifluoroleucine. Failure of charging **3** and **5** by IleRS and ValRS could be due to their steric effect or the difference of electron property imposed by the fluorination. Our *in vivo* result of codon reassignments indicates that **4**

could escape from naturally evolved editing function of aaRSs. The editing site of IleRS allows binding of valine(*19*), but prevents from binding of **4**, an amino acid slightly larger than valine. Hydrophilic nature of the editing site of ValRS (*18*) excludes binding of the hydrophobic amino acid **4**.

4. Conclusion

In conclusion we explored the incorporation of fluorinated valines into recombinant proteins *in vivo* and found that only **4** can be introduced into proteins biosynthetically *in vivo*. In addition, we observed that **4** could be assigned to isoleucine codon or valine codon, depending on whether the bacterial hosts are engineered with the enhanced IleRS or ValRS activity. Thus we found an example of alternative translation of genetic code by importing a desired aminoacyl-tRNA synthetase and a non-canonical amino acid. This finding can open up new possibilities to diversify the composition of proteins *in vivo*.

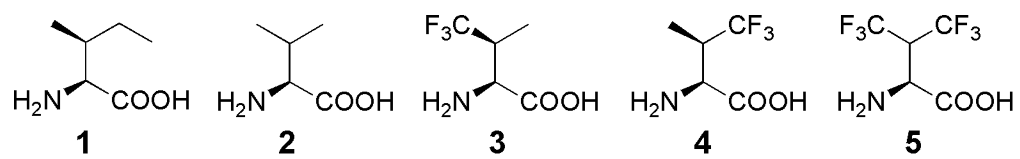

H2N
COOH
1
H2N
COOH
2
F3C
H2N
COOH
3
CF3
H2N
COOH
4
F3C
CF3
H2N
COOH
5

Table 4-1. Kinetic Parameters for Activation of Amino Acids (1-5) by *E. coli* IleRS and ValRS[a]

Substrate	Enzyme	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (rel)
1	IleRS	3.00	6.8	1
3	IleRS	n.a.[b]	n.a.	n.a.
4	IleRS	0.21	282	1/590[c]
5	IleRS	n.a.	n.a.	n.a.
2	ValRS	5.96	36.6	1
3	ValRS	n.a.	n.a.	n.a.
4	ValRS	0.36	5542	1/2490[d]
5	ValRS	n.a.	n.a.	n.a.

[a] **1** and **2** were used as the L-isomer; **3** as the (2S,3S)-form; **4** as the (2S,3R)-form; **5** was used as the D,L-mixture.
[b] not activated.
[c] relative to k_{cat}/K_m for **1** by IleRS.
[d] relative to k_{cat}/K_m for **2** by ValRS

Figure 4-1

SDS-PAGE analysis for *in vivo* incorporation of non-canonical amino acids into recombinant protein mDHFR. The analysis was conducted on whole-cell lysates and the expression conditions are noted on the top of each lane.

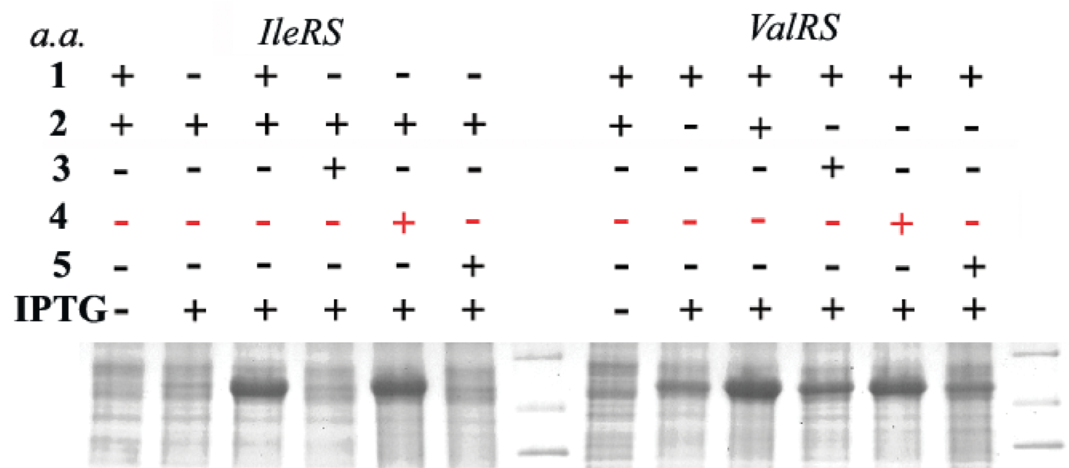
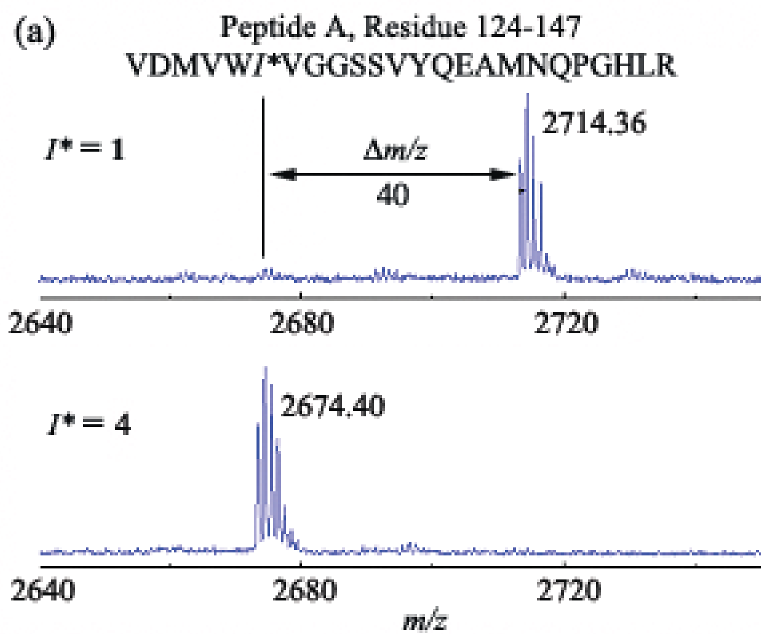

a.a.
IleRS
ValRS
1 + - + - - - + + + + + +
2 + + + + + + + - + - - -
3 - - - + - - - - - + - -
4 - - - - + - - - - - + -
5 - - - - - + - - - - - +
IPTG - + + + + + - + + + + +

Figure 4-2

New assignment of isoleucine and valine codons can be detected by MALDI-MS analysis of tryptic fragment of mDHFR. (a) Spectrum for peptide A (Val124 to Arg147, digested from mDHFR produced by strain bearing pQE15-ileS with mass 2674.40 when the single isoleucine codon is assigned to **1** (bottom panel, generated in the cultures supplemented with **1**); with mass 2714.36 when this codon is assigned to **4** (top panel, generated in the cultures supplemented with **4**). (b) Spectrum for peptide B (Gln62 to Arg69, digested from mDHFR produced by strain bearing pQE15-valS) with mass 931.11 when the single valine codon is assigned to **2** (bottom panel, generated in the cultures supplemented with **2**); with mass 985.10 when this codon is assigned to **4** (top panel, generated in the cultures supplemented with **4**).


(a)
Peptide A, Residue 124-147
VDMVWI*VGGSSVYQEAMNQPGHLR
I* = 1
Δm/z
40
2714.36
2640
2680
2720
I* = 4
2674.40
2640
2680
2720
m/z

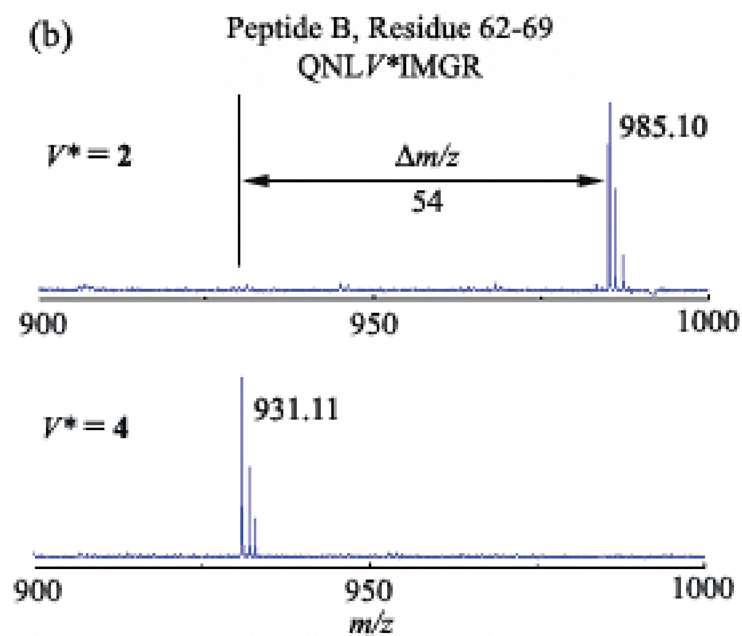

(b)
Peptide B, Residue 62-69
QNLV*IMGR
V* = 2
Δm/z
54
985.10
900
950
1000
V* = 4
931.11
900
950
1000
m/z

Figure 4-3

The tandem mass spectrum of the peptide (QNLVI*MGR) derived from mDHFR expressed in media supplemented with **4** (250 mg/L). mDHFR was expressed in an isoleucine and valine double auxotrophic strain (AIV) transformed with pQE-ileS. Partial sequence of NLVI* can be assigned from the annotated y ion series.

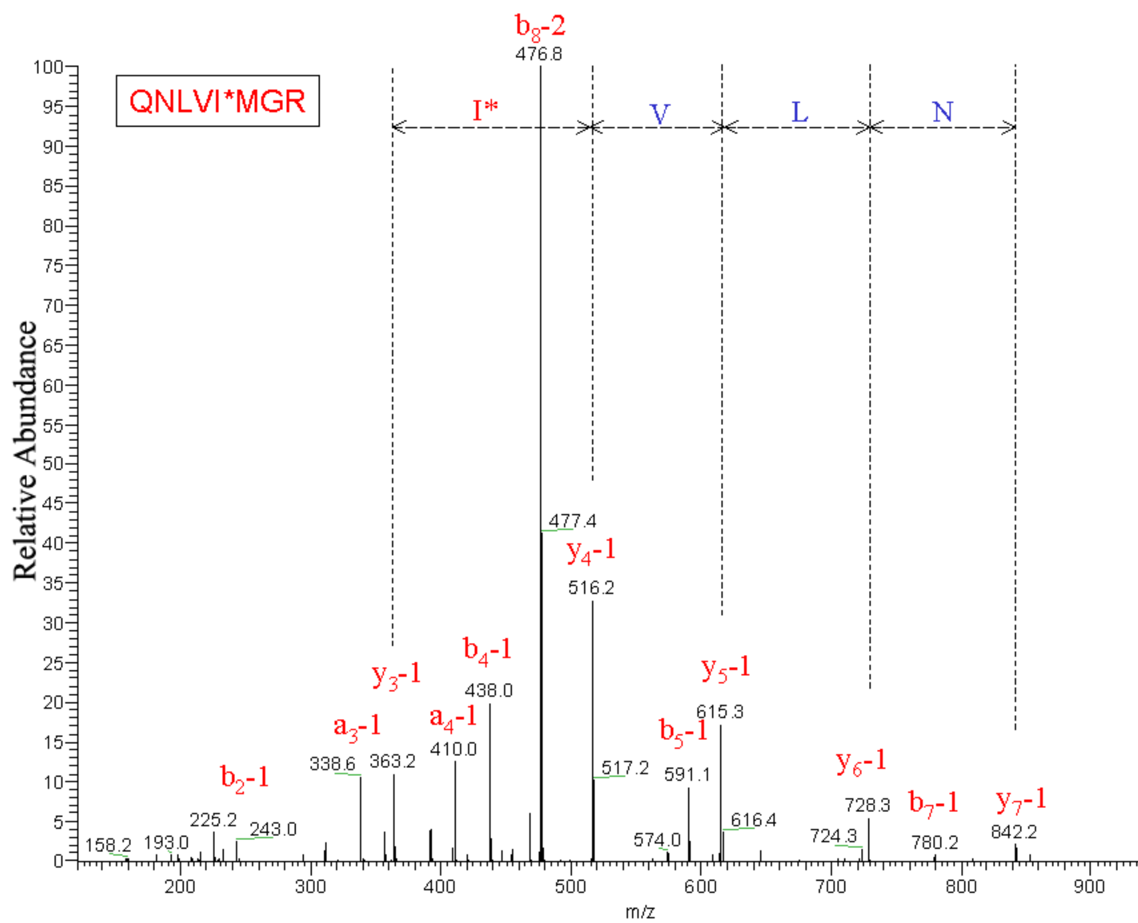
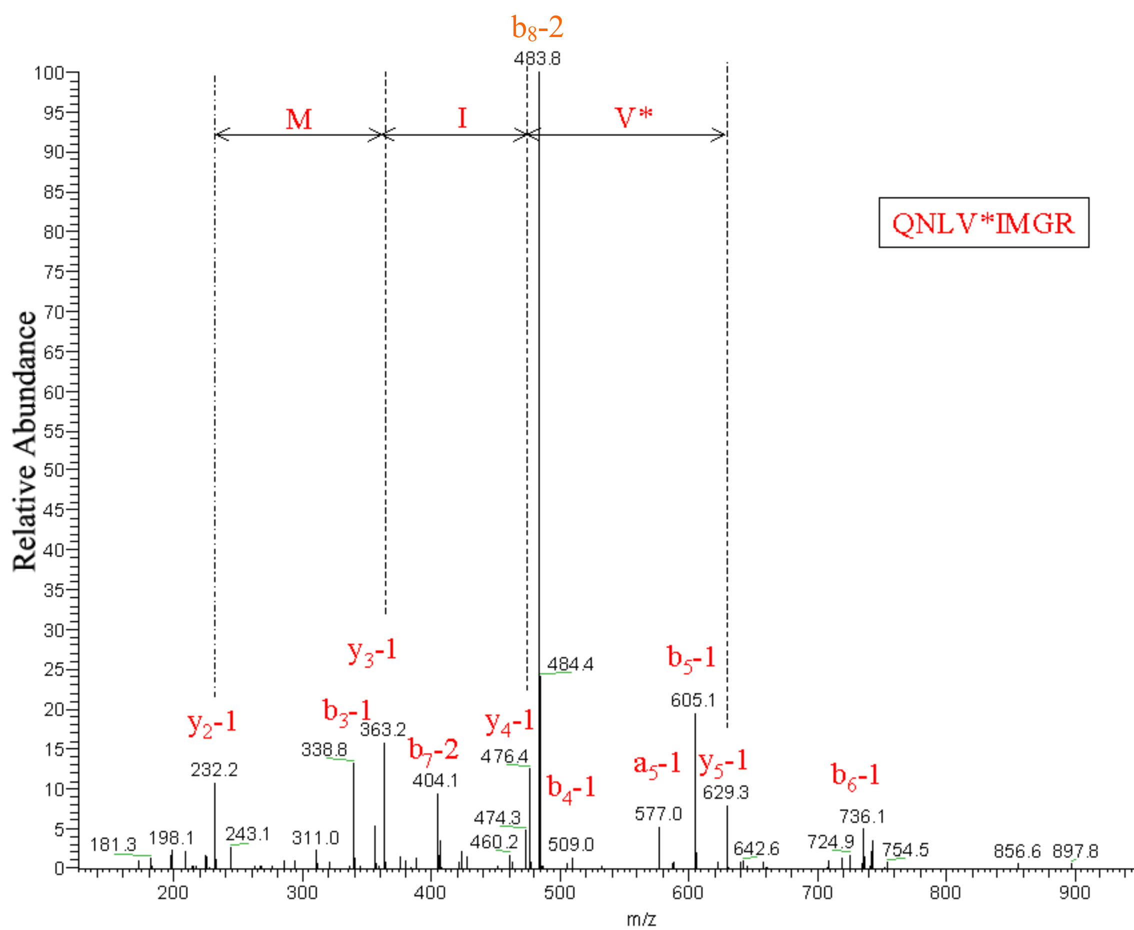

QNLVI*MGR
I*
V
L
N
b8-2
476.8
477.4
y4-1
516.2
b4-1
438.0
y3-1
363.2
a3-1
338.6
a4-1
410.0
y5-1
615.3
b5-1
591.1
517.2
b2-1
225.2
243.0
158.2
193.0
574.0
616.4
y6-1
728.3
724.3
b7-1
780.2
y7-1
842.2
Relative Abundance
m/z

Figure 4-4

The tandem mass spectrum of the peptide (QNLV*IMGR) derived from mDHFR expressed in media supplemented with **4** (250 mg/L). mDHFR was expressed in an isoleucine and valine double auxotrophic strain (AIV) transformed with pQE-valS. Partial sequence of V*IM can be assigned from the annotated y ion series.


b8-2
483.8
M
I
V*
QNLV*IMGR
Relative Abundance
y3-1
484.4
b5-1
605.1
y2-1
b3-1
363.2
y4-1
338.8
b7-2
476.4
232.2
404.1
b4-1
a5-1
y5-1
629.3
b6-1
577.0
736.1
474.3
181.3
198.1
243.1
311.0
460.2
509.0
642.6
724.9
754.5
856.6
897.8
m/z

5. References and Notes

1. Schimmel, P. (1996) *Proc. Natl. Acad. Sci. USA 93*, 4521-4522.

2. Yoder, N. C., and Kumar, K. (2002) *Chem. Soc. Rev. 31*, 335-341.

3. Francklyn, C., Perona, J. J., Puetz, J., and Hou, Y. M. (2002) *RNA 8*, 1363-1372.

4. Bilgicer, B., Xing, X., and Kumar, K. (2001) *J. Am. Chem. Soc. 123*, 11815-11816.

5. Tang, Y., and Tirrell, D. A. (2001) *J. Am. Chem. Soc. 123*, 11089-11090.

6. Bilgicer, B., Fichera, A., and Kumar, K. (2001) *J. Am. Chem. Soc. 123*, 4393-4399.

7. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) *Angew. Chem., Int. Edit. 40*, 1494-1496.

8. Tang, Y., Ghirlanda, G., Vaidehi, N., Kua, J., Mainz, D. T., Goddard, W. A., DeGrado, W. F., and Tirrell, D. A. (2001) *Biochemistry 40*, 2790-2796.

9. Xing, X., Fichera, A., and Kumar, K. (2002) *J. Org. Chem. 67*, 1722-1725.

10. Wang, P., Tang, Y., and Tirrell, D. A. (2003) *J. Am. Chem. Soc. Accepted*.

11. Hendrickson, T. L., Nomanbhoy, T. K., and Schimmel, P. (2000) *Biochemistry 39*, 8180-8186.

12. Tang, Y., and Tirrell, D. A. (2002) *Biochemistry 41*, 10635-10645.

13. de Pouplana, L. R., and Schimmel, P. (2001) *Trends Biochem. Sci. 26*, 591-596.

14. Kiick, K. L., van Hest, J. C. M., and Tirrell, D. A. (2000) *Angew. Chem., Int. Edit. 39*, 2148-2152.

15. Expression strains were derived from the isoleucine and valine double auxotrophic strain AIV-IQ (*E. coli B F- ompT hsdS*(r_B^- m_B^-) *gal dcm l(DE3) ilvD691*), which has been described previously (Wang, P., Tang, Y., Tirrell, D. A., *J. Am. Chem. Soc*., in press).

16. Our medium shift protocol does not completely exhaust the intracellular pool of **2**, so significant background expression was observed in the strain bearing pQE15-valS in cultures supplemented with **3** or **5**. MALDI-MS analysis of tryptic peptides from these background proteins showed the valine codons in mDHFR were exclusively assigned to **2**.

17. Fersht, A. R. (1998) *Science 280*, 541-541.

18. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J. S., Vassylyev, D. G., and Yokoyama, S. (2000) *Cell 103*, 793-803.

19. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P., and Yokoyama, S. (1998) *Science 280*, 578-582.