**Supplementary Material**

# Intriguing Cellular Processing of a Fluorinated Amino Acid During Protein Biosynthesis in *Escherichia coli*

Mark D. Vaughan[a,b], Zhengding Su[a,c], Elisabeth Daub[a] and John F. Honek[a,*]

[a]Department of Chemistry, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1 Canada

[b]Current address: MDV: QUEST University Canada, 3200 University Boulevard, Squamish, British Columbia V8B 0N8

[c]Current address: ZS: Key Laboratory of Industrial Fermentation of Ministry of Education and Provincial Collaborative Innovation Center for Industrial Fermentation, Hubei University of Technology, Wuhan, China

* To whom correspondence should be addressed: J. F. Honek Phone: 1-519-888-4567 x 35817; Fax: 1-519-746-0435; E-mail: jhonek@uwaterloo.ca

# Supplementary Methods

## SUPPLEMENTARY METHODS

L-Methionine and L-ethionine were obtained from Sigma (Mississauga, ON). 1-Bromo-2-fluoroethane was supplied from Acros Organics (New Jersey, USA). Acetylene (atomic absorption spectroscopic grade) was obtained from Linde (Hollriegelskreuth, Germany). Chlorodifluoromethane (Halocarbon-22) was from Matheson Canada and anhydrous ammonia was from Praxair Canada. The bacterial strains *E. coli* BL21(DE3) and methionine auxotroph B834(DE3) and plasmid pET-22b are products of Novagen (Madison, WI). Reagents for performing PCR and restriction endonucleases were purchased from New England Biolabs (Beverly, MA).

### SYNTHESES

#### *S*-(Trifluoromethyl)-L-homocysteine (L-TFM) and *S*-(Difluoromethyl)-L-homocysteine (L-DFM)

L-Trifluoromethionine (L-TFM) and L-difluoromethionine (L-DFM) were synthesized as previously described.[1-3]

#### *S*-(2-Fluoroethyl)-L-homocysteine (L-MFE)

L-MFE was synthesized as follows: L-Homocystine (0.20 g; 0.76 mmol) was dissolved in anhydrous liquid ammonia and sodium metal was added in small pieces until the maintenance of a dark blue colour for 20 min. After quenching excess sodium with $NH_4Cl$, 1-bromo-2-fluoroethane (160 □l; 0.27 g; 2.1 mmol) was added dropwise over 1 h. The ammonia was allowed to evaporate, and the residue was suspended in acetone, which

was then evaporated to remove remaining 1-bromo-2-fluoroethane. The residue was taken up in 5 ml Milli-Q water and applied to a Lobar $C_{18}$ column (31 cm × 2.5 cm, manufactured by Merck, Germany, purchased from VWR Canada). L-Monofluoroethionine was eluted with acetonitrile in 10% increments (L-monofluoroethionine eluted between 10 and 20% acetonitrile), yielding a white solid (182 mg, 66%) and consistent with previously reported characteristics for L-MFE.[4]

$R_f$ (*n*-BuOH/AcOH/$H_2O$ 4:1:1; $C_{18}$ tlc plates) 0.40; mp 225-226 ºC. $^1$H NMR (300 MHz, $D_2O$) $\delta$ 4.48 (dt, 2H, $J$ = 6.3 Hz, $J_{HF}$ = 45.7 Hz, $CH_2CH_2F$) 3.67 (t, 1H, $J$= 7.0 Hz, $C_\alpha H$) 2.71 (dt, 2H, J = 6.3 Hz, $J_{HF}$ = 22.5 Hz, $CH_2CH_2F$) 2.53 (t, 2H, $J$ = 7.7 Hz, $C_\gamma H_2$) 1.98 (m, 2H, $C_\beta H_2$); $^{13}$C NMR (75.5 MHz, $D_2O$) δ 170.9 ($CO_2^-$) 90.5 ($J_{CF}$ = 198.4 Hz, $CH_2CH_2F$) 51.8 ($C_\alpha H$) 31.1 ($J_{CF}$ = 22.3 Hz, $CH_2CH_2F$) 28.1 ($C_\beta H_2$) 22.2 ($C_\gamma H_2$); $^{19}$F (282.3 MHz) $\delta$ (relative to $CFCl_3$, 0.00 ppm) -206.2; HRMS (ESI) calculated for $C_6H_{12}NFO_2S$ 181.2331, observed 182.2341 ($M+H^+$).

**_S_-(2-Hydroxyethyl)-L-homocysteine (L-HEHC)**

L-HEHC was synthesized using a reported method[5] but utilizing L-homocystine as for the preparation of L-MFE. Isolated product was in complete agreement with previously reported data for this compound.[4-6]

**_S_-Vinyl-L-homocysteine (L-vinthionine)**

L-Homocystine (0.20 g, 0.76 mmol) was placed in a three-neck round bottom flask fitted with a dry ice/acetone condenser. Anhydrous ammonia (~25 ml) was condensed into the flask under a stream of argon. Sodium metal was added in small

pieces until the resulting blue colour was maintained for at least 20 min. Excess sodium was quenched by the addition of solid ammonium chloride until the blue colour disappeared. The ammonia was evaporated under a stream of argon and the resulting residue was taken up in 25 ml dimethyl sulfoxide (anhydrous, deoxygenated by bubbling argon into the liquid). The suspension was cooled to 0 ºC and acetylene was bubbled in for 4 h, during which time the reaction mixture was allowed to warm to ambient temperature. The solution was neutralized by adding 1 N HCl, and the solvent was removed *in vacuo*. The residue was taken up in 5 ml water and applied to a $C_{18}$ Lobar medium pressure chromatographic system and purified as described above for L-monofluoroethionine (L-vinthionine eluted between 10 and 20% acetonitrile) to yield a white solid (142 mg, 59%) and having properties consistent with previously reported L-vinthionine.[7]

$R_f$ (*n*-BuOH/$H_2O$/AcOH; $C_{18}$ tlc plates) 0.32; mp 220-222 ºC. $^1$H NMR (300 MHz, $D_2O$) $\delta$ (relative to TMS, 0.00 ppm) 6.25 (1H, dd, SC$H_X$=C$H_A$$H_B$, $J_{AX}$ = 10 Hz, $J_{BX}$ = 17 Hz) 5.16 (1H, d, SC$H_X$=C$H_A$$H_B$, $J_{AX}$ = 9 Hz) 5.03 (1H, d, SC$H_X$=C$H_A$$H_B$, $J_{BX}$ = 17 Hz) 3.72 (1H, t, C$_\alpha$*H*, *J* = 6 Hz) 2.73 (2H, t, C$_\gamma$$H_2$, *J* = 8 Hz) 2.01 (2H, m, C$_\beta$$H_2$); $^{13}$C NMR (75.5 MHz, $D_2O$) $\delta$ (relative to TMS, 0.00 ppm) 173.8 (*C*O) 130.5 (S*C*H=$CH_2$) 112.7 (S*C*H=$CH_2$) 53.7 ($C_\alpha$) 30.2 ($C_\gamma$) 26.3 ($C_\beta$); HRMS (ESI) calculated for $C_6H_{11}NO_2S$ 161.2268, observed 162.2259 ($M+H^+$).

**AMINO ACID INCORPORATION STUDIES**

Incorporation protocols using L-MFE and L-vinthionine were identical to those previously reported for the successful incorporation of the analogues L-DFM[2] and L-TFM[3] into bacteriophage lambda lysozyme. In brief, the previously reported *E. coli* expression system (*E. coli* BL21(DE3)) hosting a suitable plasmid for the overproduction of LaL (pLR102) was used to produce LaL or fluorinated amino acid analogue-labeled LaL.[2, 3, 8] Unless otherwise stated, cells were grown in $M_9$ minimal medium supplemented with 0.4% glucose and 40 mg/L ampicillin at 37 °C which was supplemented with 0.1 mM L-Met to an absorbance of 0.65 at 600 nm. Cells were centrifuged (x 9950*g*), washed with the same medium and then resuspended in one-half the original culture volume of the medium containing 0.75 mM IPTG and either 1.0 mM fluorinated amino acids (i.e. L-DFM, L-TFM or L-MFE) for incorporation (at 37 °C for 9.5 h). Cells were harvested by centrifugation, resuspended in 50 mM potassium phosphate (pH 7) and disrupted either by sonication or passage through a French press (10,000 psi). Cell debris was removed by centrifugation (x 37,000 *g*). Labeled lysozymes were purified by sequential chromatography over S-Sepharose Fast Flow, Mono-S, and Phenyl-Superose (Pharmacia-LKB Biotechnology AB, Uppsala, Sweden), then dialyzed against 5 mM phosphate buffer (pH 7), and lyophilized as described previously.[2, 3, 8]

L-MFE/L-vinthionine labeled protein production levels were comparable to previously observed expression yields seen for the incorporation of L-difluoromethionine into LaL and were in the range of 15-20 mg of isolated protein per liter of bacterial growth. Due to the complexity of the multiple modifications and the dependence of analogue incorporation upon the ratio of [L-MFE] to [L-Met] in the bacterial growth

media, as previously observed for L-trifluoromethionine incorporation, labeling efficiency was conservatively estimated by analysis of the mass spectra obtained from various experiments to occur at a minimum of 70%. Higher levels were observed when bacterial cells were washed free from any L-methionine present in the initial growth medium before protein production in the presence of L-MFE (2 mM) as the sole methionine source in the medium during protein production. An analysis of the masses observed in the ESI spectrum of labeled protein (Figure 3) provides a crude estimate of the ratio of L-MFE:L-vinthionine under these specific conditions of approximately 1:6, assuming a roughly similar ESI sensitivity for each labeled protein. There exists the potential that some labeled proteins will also contain the hydroxylated L-HEHC analogue, and may add to the broadness of several masses that are observed (such as 17883) as deconvoluted by the MaxEnt algorithm in the MassLynx software (Supplementary Table S1). Mass peaks which are not observed in the representative mass spectrum shown in Figure 3 of the main text, such as for example, the protein containing only one L-MFE residue (Mr = 17857) have been observed in other MFE-LaL preparations produced by variation of the L-MFE to L-Met ratio in the bacterial growth medium.

In order to assign the $^{19}$F NMR resonances to each of the three positions in the L-MFE labeled phage lysozyme, protocols successfully utilized to assign the $^{19}$F NMR resonances for L-DFM and L-TFM incorporated into phage lysozyme were used.[9] This included the utilization of the paramagnetic line broadening agent Gd-EDTA as previously described to detect solvent exposed MFE residues at positions 1 and 107, with the MFE fluorine nucleus at position 14 being protected from the broadening effect as it is positioned in the hydrophobic core of the lysozyme.[9] As well, the gene coding for the

Met107Leu mutation was also utilized to prepare the L-MFE labeled variant, an approach that was also used successfully in a previous study.[9] Results are presented in the main text and in Figure 3a.

## METHIONYL-tRNA SYNTHETASE (MetRS)

### *Recombinant E. coli methionyl-tRNA synthetase*

The fully functional recombinant 551-amino acid tryptic fragment of the *E. coli* Met-tRNA synthetase (MetRS) was overproduced and purified as previously described.[10] The kinetic assays were undertaken based on the exchange of radiolabeled pyrophosphate into ATP in the absence of tRNA, but in the presence of L-methionine or an analogue.[10-12]

## MASS SPECTROMETRY

Electrospray mass spectrometry was performed on either a Fisons VG Quatro II triple-quadrupole mass spectrometer or a Micromass Q-tof Global Ultima instrument using delivery solvents of 1:1 $H_2O/CH_3CN$ (containing 0.1% TFA or acetic acid) and data analyzed with the Masslynx software. Multiply charged electrospray spectra were subjected to the MaxEnt algorithm to produce true molecular mass spectra with associated errors. MaxEnt data are quantitative and bar spectrum proportional to the area under individual peaks were created to estimate the relative quantities of the components in each spectrum. Cone voltages utilized for various mass spectrometry experiments were optimized for sensitivity of the analyte and were generally centered at 165 volts. No evidence for gas phase chemistry in the mass spectrometer was observed at a variety of

cone voltages for the protein samples nor for the amino acids themselves, also analyzed by ESI mass spectrometry.

***NMR***

$^{1}$H-Decoupled $^{19}$F NMR spectra of LaL labeled with the fluorinated amino acid analogue were acquired at 470.3 MHz using a Bruker Avance 500 spectrometer at 25 °C. Standard parameters were 18,832 Hz sweep width, 0.887 s acquisition time, and a 1.0s relation delay. A 3 Hz line broadening was applied. Spectra were referenced to an external sample of $CFCl_3$ in $CD_3OD$ (set at 0.00 ppm). The protein concentration of NMR samples was 0.5 mM in 600 µl $D_2O$. The standard deviation in $^{19}$F chemical shift measurements was ±0.02 ppm based on three independent measurements of the same sample.

All $^{1}$H-NMR spectra of amino acid analogues and MFE-labeled LaL were collected using standard pulse sequences at 25 °C on a Bruker Avance 600 spectrometer equipped with a triple-resonance pulse-field gradient probe. All the NMR samples were prepared in the same buffer, i.e., 5 mM potassium phosphate at pH 7.0 except mentioned otherwise. TOCSY spectra ($\tau_{SL}$ = 45 ms) were recorded at 25 °C for the water and 100% $D_2O$ samples at 600 MHz. All spectra were referenced to 0 ppm for $^{1}$H by the $\delta_{DSS}$. The water signal was suppressed by using WATERGATE[13] for the spectra in $H_2O$ and by using presaturation in $D_2O$. The spin-lock used in the TOCSY spectra was an MLEV-17 sequence with a typical field strength of 8.3 kHz preceded and followed by a 2.0-ms trim pulse. The spectra were acquired with 2048 data points in the $t_2$ (acquisition) dimension, and 300-350 increments along $t_1$. Time proportional phase incrementation (TPPI) was

used for phase-sensitive acquisition in the $t_1$ dimension. The NMR data were processed by using the Bruker XWIN-NMR program. Squared cosine window functions were applied in both dimensions. The Fourier-transformed spectra had 2048 data points in the F2 dimension (resolution of 2.93 Hz/point) and 1024 data points in the F1 dimension (resolution of 5.87 Hz/point). A polynomial baseline correction was performed in the F2 dimension. Spectral processing, display and analysis were performed using the XwinNMR software package supplied with the spectrometer system.

NMR experiments for monitoring the conversion of L-MFE into *S*-(2-hydroxyethyl)-L-homocysteine at pH 6, 7 and 9 were performed at 25 °C on a Bruker Advance 500 MHz NMR spectrometer equipped with a 5 mm auto-tune gradient probe. The NMR samples were suspended in $D_2O$ or 10 mM potassium phosphate buffer (pH 7). The spectra were recorded every 6 hours. The 1D proton NMR experiments utilized a pre-saturation $^1H$ NMR pulse sequence (sweep width (sw): 14.0 ppm; transmitter offset (o1p): 6.0 ppm; relaxation delay (d1): 1 sec; number of scans (ns): 32 and time domain (TD): 16 k). All the proton NMR experiments, which included data acquisition, processing and analysis, were carried out using XwinNMR software

***Molecular Modeling***

Molecular graphics images were produced using the UCSF Chimera package (version 1.10.1) from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).[14]

# Supplementary Results

. **Table S1. Deconvolution of the ESI-MS Data for MFE-labeled LaL Proteins**

| Amino Acid Distribution[1] | Expected Molecular Mass (Da) | Observed Molecular Mass (Da) |
|---|---|---|
| Wild type | 17 825 | 17 824 |
| 1 Vin | 17 837 | 17 837 |
| 2 Vin | 17 849 | 17 850 |
| 3 Vin | 17 861 | 17 861 |
| 1 MFE<br>1 Vin | 17 869 | 17 870 |
| 2 Vin<br>1 MFE | 17 881 | 17 883 |
| 1 Vin<br>2 MFE | 17 901 | 17 905 |
| 3 MFE | 17 921 | 17 918 |

[1] Possible replacement at a subset of wild-type LaL methionine positions M1, M14, and M107.

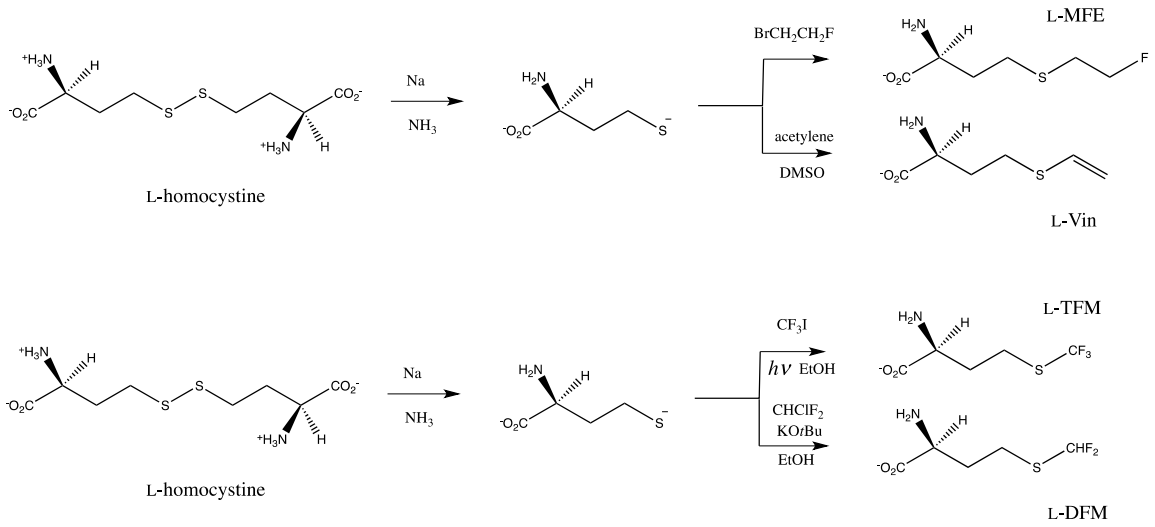
The following masses calculated for the presence of L-HEHC as a third component are as follows: 1HEHC (17855); **2HEHC (17885)**; **3HEHC (17915)**; 1HEHC/1VIN (17867); 1HEHC/1MFE (17887); 2HEHC/1VIN (17897); **2HEHC/1MFE (17917)**; 1HEHC/2VIN (17879); **1HEHC/2MFE (17919)**; 1HEHC/1VIN/1MFE (17899). Masses marked in bold for L-HEHC incorporation could also make contributions to the overall protein mass envelope although would appear, if present, to make only a minor contribution to the less abundant 17918 observed mass. A contribution to the observed broad mas centered at 17883 could also occur.

**Table S2. Kinetics of Met-tRNA synthetase for various substrates**

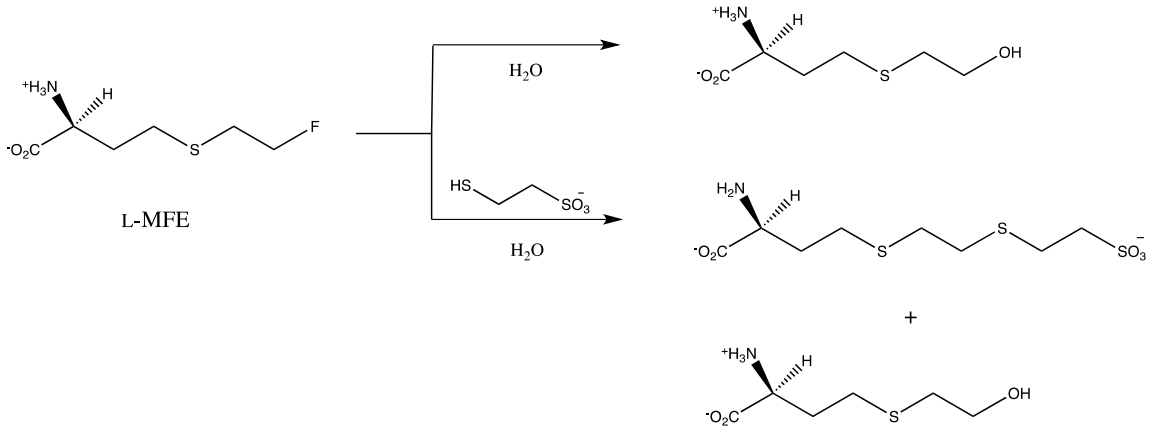
| *Substrate* | *Incorporation[a]* | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($mM^{-1}s^{-1}$) | $K_m$ (mM) |
|---|---|---|---|---|
| *L-Met* | *100* | *10.0 ± 0.4* | *1000 ± 10* | *0.0095 ± 0.002* |
| *L-DFM* | *>95* | *8.0 ± 1.0* | *4.0 ± 2.0* | *1.9 ± 0.9* |
| *L-TFM[b]* | *<70* | *-* | *-* | *> 5* |
| *L-Ethionine* | *>95* | *8.4 ± 0.6* | *8.0 ± 3.0* | *1.0 ± 0.3* |
| *L-MFE* | *60[c]* | *5.0 ± 0.2* | *1.0 ± 0.2* | *4.8 ± 0.9* |
| *L-Vinthionine* | *ND[d]* | *-* | *-* | *-* |
| *L- HEHC* | *ND[d]* | *-* | *-* | *-* |

*[a] Percent incorporation in LaL using previously reported methods. [b] Saturation kinetics not observed within the accessible concentration range. [c] Total incorporation of L-MFE and L-vinthionine. [d] incorporation not observed as determined by SDS-PAGE analysis*

## Schemes




**Scheme S1.** Synthetic approaches to methionine analogues utilized in this study.




**Scheme S2.** Chemical reactions observed for L-MFE.

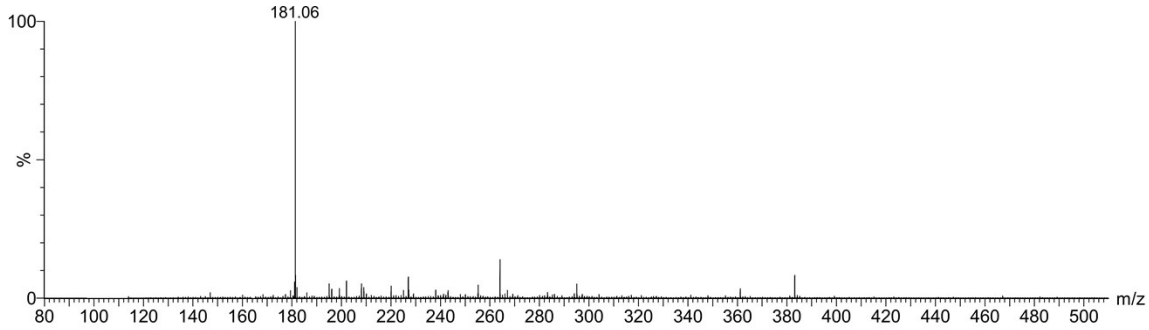



**Figure S1**. **Mass spectrum of L-MFE**.

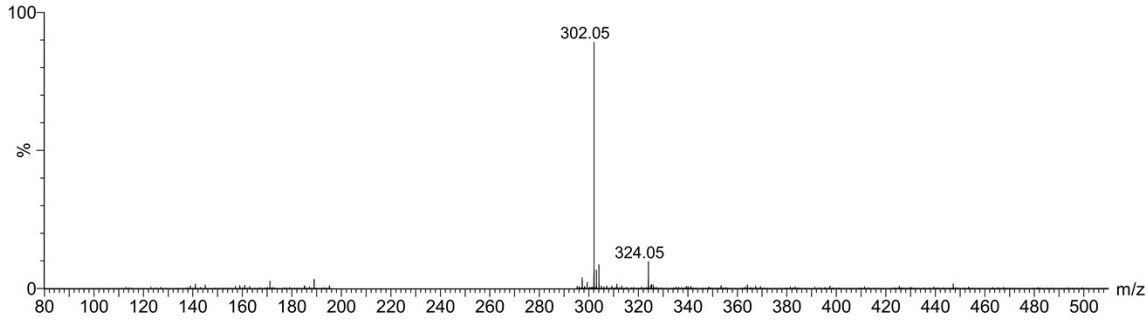



**Figure S2**. **Mass spectrum of L-MFE after a 4 day incubation in the presence of β-mercaptoethane sulfonate (MCES) at room temperature at pH 9. Molecular mass for adduct ($C_8H_{16}NO_5S_3$) was calculated to be 302.02. The mass at 324 is likely an adduct with sodium ion.**

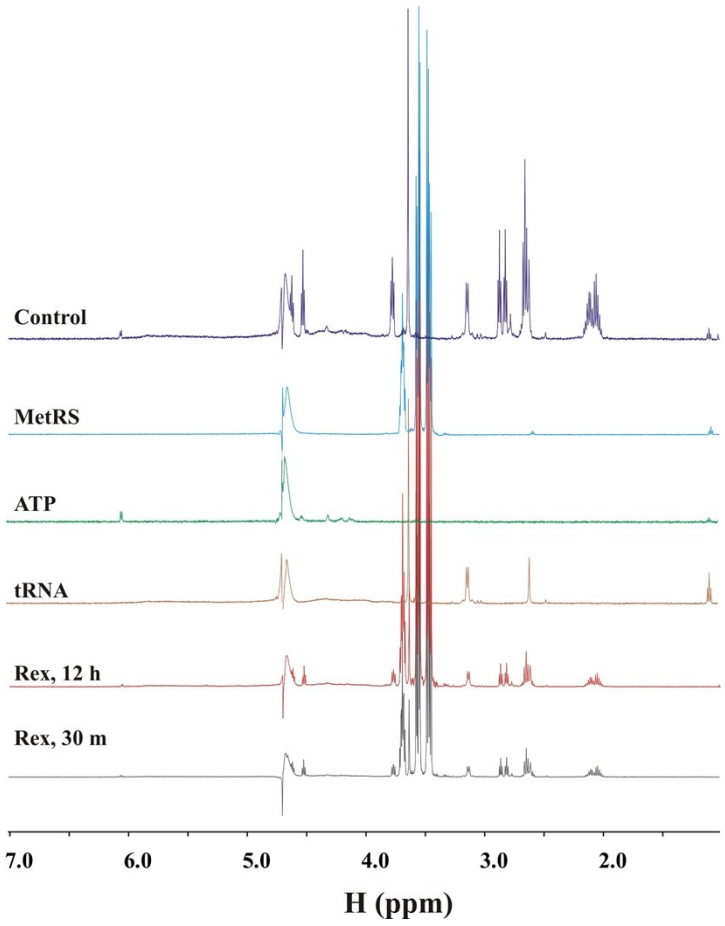



**Figure S3. Lack of L-vinthionine formation during incubation of L-MFE in the presence of ATP, $tRNA^{Met}$ and methionyl-tRNA synthetase as monitored by NMR spectroscopy**. Reaction mixture contained L-MFE (10 mM), ATP (10 mM), $tRNA^{Met}$ (10mM) and MetRS (10 ul) in 10 mM phosphate buffer (pH 7). The control experiment was done having the same reaction components except lacking MetRS in the reaction mixture.

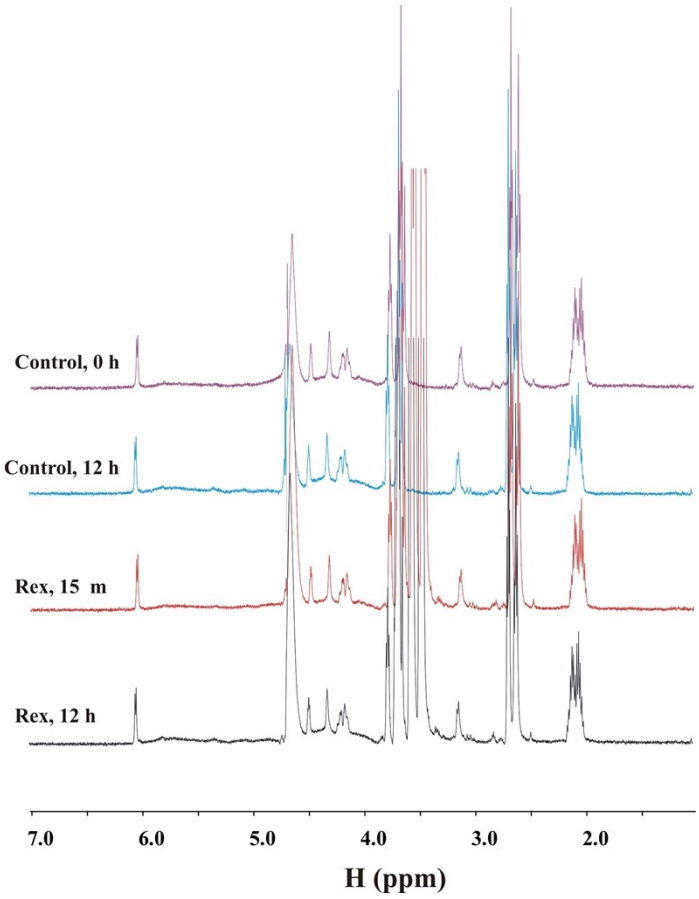



**Figure S4. Charging $tRNA^{Met}$ with L-HEHC by methionyl-tRNA synthetase monitored by NMR spectroscopy**. Reaction mixture contained L-MFE (10 mM), ATP (10 mM), $tRNA^{Met}$ (10mM) and MetRS (10 ul) in 10 mM phosphate buffer (pH 7). The control experiment was done having the same reaction components except lacking MetRS in the reaction mixture. No reduction of L-HEHC proton resonances were observed.

**References**


1. T. Crepin, E. Schmitt, Y. Mechulam, P. B. Sampson, M. D. Vaughan, J. F. Honek and S. Blanquet, *J Mol Biol*, 2003, **332**, 59-72.
2. M. D. Vaughan, P. Cleve, V. Robinson, H. S. Duewel and J. F. Honek, *J Am Chem Soc*, 1999, **121**, 8475-8478.
3. H. Duewel, E. Daub, V. Robinson and J. F. Honek, *Biochemistry*, 1997, **36**, 3404-3416.
4. T. Bourdier, C. J. R. Fookes, T. Q. Pham, I. Greguric and A. Katsifis, *J Labelled Compd Rad*, 2008, **51**, 369-373.
5. T. A. Connors, Peuthere.Ma, A. B. Mauger and W. C. J. Ross, *J Chem Soc*, 1962, 4601-4607.
6. T. Nishi, N. Kouda, K. Watanabe, Y. Tanaka, K. Araki, K. Furihata, M. Onodera and K. Toda, *J Biosci Bioeng*, 2004, **97**, 134-137.
7. W. R. Leopold, J. A. Miller and E. C. Miller, *Cancer Res*, 1982, **42**, 4364-4374.
8. H. S. Duewel, E. Daub and J. F. Honek, *Biochim Biophys Acta*, 1995, **1247**, 149-158.
9. H. S. Duewel, E. Daub, V. Robinson and J. F. Honek, *Biochemistry*, 2001, **40**, 13167-13176.
10. M. D. Vaughan, P. B. Sampson, E. Daub and J. F. Honek, *Med Chem*, 2005, **1**, 227-237.
11. S. Blanquet, G. Fayat, J. P. Waller and M. Iwatsubo, *Eur J Biochem*, 1972, **24**, 461-469.

12. F. Lawrence, S. Blanquet, M. Poiret, M. Robert-Gero and J. P. Waller, *Eur J Biochem*, 1973, **36**, 234-243.

13. M. Piotto, V. Saudek and V. Sklenar, *J Biomol NMR*, 1992, **2**, 661-665.

14. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605-1612.