Japan Journal of Medicine
2018; 1(6): 285 - 290 . doi: 10.31488/jjm.1000129

Research article

Evolutionary Analysis of Post-translational Modification Sites in Translation Elongation Factor 1A

Yosuke Kondo[1], Satoru Miyazaki[*1]

1. Department of Medicinal and Life Science, Faculty of Pharmaceutical Sciences, Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba 278-8510, Japan

[*]Corresponding author: Department of Medicinal and Life Science, Faculty of Pharmaceutical Sciences, Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba 278-8510, Japan, Tel: +81-4-7121-3630; Fax: +81-4-7121-3630; E-mail: smiyazak@rs.noda.tus.ac.jp

Received: July 18, 2018; Accepted: August 17, 2018; Published: August 21, 2018

Abstract

Evolutionary conservation is one of the powerful analyses to predict protein functional sites. There have been many studies proposing a variety of degrees of conservation. In this paper, we propose and compare two mathematical measures that calculate the degree of conservation at each site in the molecular evolutionary process, named "the degree of conservation" and "the degree of specific conservation". The degree of conservation can identify the sites which show a conserved pattern in all proteins of a protein family. The degree of specific conservation is the degree that can identify the sites conserved in a subfamily but variable in the whole family. In this study, the conservation analyses were applied to a protein that has many functions in the cell. Many researches of eukaryotic elongation factor 1A (eEF1A) revealed that eEF1A is involved in not only protein biosynthesis but also moonlighting functions such as cytoskeletal modification, apoptosis and viral infection. And functional divergences of the EF-Tu/1A protein family have occurred in the evolutionary process. Our results showed that the degree of conservation can predict the regions that are important for binding molecules of the eEF-Tu/1A proteins. The degree of specific conservation can predict the residues important for the functional divergences and post-translational modifications (PTMs). These results suggest that PTM sites of eEF1A are not conserved in the whole family but only conserved in a subfamily. It would be necessary to verify whether the conserved sites are responsible for the moonlighting functions of eEF1A to develop new drugs that can be effective for cancers.

Keywords: evolutionary conservation; peptide elongation factor; post-translational modification

Abbreviations: eEF: eukaryotic elongation factor; MSA: multiple sequence alignment; PTM: post-translational modification

Introduction

Eukaryotic translation elongation factor 1A (eEF1A) is an abundant protein in cells and delivers aminoacyl-tRNA onto a ribosome [1]. It is said that the principal function of eEF1A is protein biosynthesis. However, eEF1A is revealed as one of the multifunctional proteins [2]. Novel eEF1A functions such as cytoskeletal modification, apoptosis, viral infection and so on [3] have been clarified from the discovery of the actin binding function [4]. eEF1A is also identified as a protein attaching cell membranes to induce anoikis by inactivating β1-integrin [5]. The antiadhesive function from the extracellular matrix is effective to develop new drugs for cancers such as leukemia [6].

In addition to the moonlighting functions, many functional divergences have been observed in the eEF1A family. There are two known isoforms of eEF1A proteins, eEF1A1 and eEF1A2, in which the sequences share 92% identity [7,8]. Expression of the gene coding eEF1A depends on cell types [9]; the majority of cells express eEF1A1, neuron or muscle cells often express eEF1A2 and some of other cell types express both eEF1A isoforms. In order to elongate a polypeptide chain by

eEF1As, GDP/GTP binding is necessary. However, eEF1As do not have identical binding affinities with GDP and GTP [10]; the GTP binding affinity of eEF1A1 is higher than eEF1A2 but that of eEF1A2 is higher than eEF1A1.

Furthermore, eEF1A1 promotes apoptosis [11] but overexpression of eEF1A2 causes cancers [12]. Such functional differences are important for elucidating various cell functions. There are 36 non-identical residues between human eEF1A1 and eEF1A2. It is known that some of these residues or corresponding residues of eEF-Tu are estimated to be important for actin binding or fibronectin binding functional divergences [13, 14]. On the other hand, identical residues were estimated to be also important for the functional divergences because GDP/GTP binding affinity is different because of non-identical residues that are close to the identical residues [15]. Recent studies showed that post-translational modification (PTM) sites of eEF1A determined by mass spectrometry analysis [16-18] are located on regions close to the non-identical residues [15]. PTM residues should be important for analyzing the functional divergences of eEF1As.

PTM sites of eEF1A is previously estimated as non-conservation sites [19]. Meanwhile, our previous studies showed that residues involving binding molecules or functional divergences of the EF-Tu/1A family can be predicted by evolutionary analysis [20, 21]. Therefore, evolutionary conservation of the EF-Tu/1A protein family is useful to investigate important residues of the proteins. In this paper, we propose degrees of conservation and discuss what degree of conservation can accurately predict the PTM sites of eEF1A. This analysis would clarify how the PTM sites of eEF1A proteins are evolved.

Materials and Methods

Construction of the multiple sequence alignment and phylogenetic tree

From UniProtKB/Swiss-Prot [22], we collected 984 entries, which (1) are annotated as 'Classic translation factor GTPase family. EF-Tu/EF-1A subfamily', (2) do not include 'X' in the sequence and (3) are not a fragment. The sequences were aligned by MAFFT 7 program [23]. Distances of sequence pairs were com-puted by maximum likelihood method [24] using the Jones-Taylor-Thornton model [25] as a substitution matrix and the Dayhoff method [26] for computing equilibrium frequencies. From all combinations of the distances, a phylogenetic tree was written by unweighted pair group method with arithmetic mean [27].

Preparation for defining degrees of conservation

Let M = (m_{ij}) denote a given multiple sequence alignment (MSA) and here mij denote an amino acid symbol in sequence *i* of site *j* on the MSA. Let $X_j = \{m_{1j}, m_{2j}, \ldots, m_{nj}\}$ be a multiset of amino acid symbols. Let F_j denote a field of sets of X_j. And, let F_{jt} be an element of F_j and a group by node t in a phylogenetic tree.

Mathematical formulation for a degree of variability
Let

$$\upsilon(F_{jt}) = \frac{1}{|F_{jt}|} \sum_{y,z \in F_{jt}, y \neq z} [s(y, z)\omega(y) + s(z, y)\omega(z)] \quad (1)$$

be a degree of variability, where $\upsilon(F_{jt}) = 0$ if $|F_{jt}| = 1$. Let

A = {A,C,D,E, F,G,H, I,K,L,M,N, P,Q,R, S, T, V,W, Y }

be a set of amino acid symbols and $G_j = \{\Upsilon_{j1}, \Upsilon_{j2}, \ldots, \Upsilon_{jg}\}$ be a set of gaps at site j, where g is the number of gaps at site j. Let

$$s(y, z) = \begin{cases} \dfrac{S(y-y) - S(y,z)}{S(y,y)} (y \in A \wedge z \in A) \\ \dfrac{S(y-y)S_{\min}}{S(y,y)} (y \in A \wedge z \in G_j) \\ \dfrac{S_{\max} - S_{\min}}{S_{\max}} (y \in G_j) \end{cases} \quad (2)$$

be a normalized substitution matrix, where S_{max}, S_{min}, $S(y, y)$ and $S(y, z)$ are the maximum, the minimum, a diagonal element and an off-diagonal element in an amino acid substitution matrix, respectively, and $w(y)$ is a weight of sequence *y*. The Gonnet matrix [28] was used as a substitution matrix. The weight was computed by Sibbald and Algos algorithm [29] and the iteration number was 100,000.

Mathematical formulation for degrees of conservation

Let

$$c_\alpha(F_j) = \frac{1}{N} \sum_{t=1}^{N} [1 - \upsilon(F_{jt})] \quad (3)$$

be a degree of conservation, where N is the number of nodes in the phylogenetic tree.

Let

$$c_\beta(F_j) = \frac{\upsilon(F_{jr}) + 1}{\upsilon(F_{ja}) + 1} - 1 \quad (4)$$

be a degree of specific conservation, where r is the root of the phylogenetic tree and a is the group interested in the study.

Data collection of functional sites in eEF-Tu/1A

Binding residues were collected from three-dimensional structures of eEF-Tu/1A described in our previous study [21]. Actin-binding residues were obtained from site-directed mutagenesis data [30]. PTM sites were obtained from the PhosphoSitePlus database [31].

Results

Predicted sites by degrees of conservation

The multiple sequence alignment constructed from the sequences in the EF-Tu/1A family consisted of 739 sites. The degrees of conservation and specific conservation were calculated at each site. In order to determine a threshold of each degree of conservation, we used ROC (receiver operating characteristic) curves.

The ROC curve of the degree of conservation was created from proximity of binding molecules or ions in three-dimensional structures of the EF-Tu/1A proteins as described in our previous paper [21]. Because Figure 1A shows that the threshold is 0.973, predicted sites were defined as the sites in which the degree of conservation is higher or equal to the threshold. The predicted sites shown in Figure 1B are mainly located on the left side of the three-dimensional structure.

The ROC curve of the degree of specific conservation was constructed from actin binding residues [13]. Because Figure 2A shows that the threshold is 0.190, predicted sites were defined as the sites in which the degree of specific conservation is higher or equal to the threshold. The predicted sites shown in Figure 2B are mainly located on the right side of the three-dimensional structure.

In order to clarify the difference of the predicted sites between each degree of conservation, we investigated overlapping predicted residues shown in Figure 3A. Hundred residues were overlapping residues that were predicted both degrees of conservation. Figure 3B also shows that the degrees of conservation and specific conservation mainly predicted the left and right side of the three-dimensional structure,respectively.

The degrees of conservation and PTM sites

PTM sites of eEF1A1 in human, mouse, rat and rabbit and eEF1A2 in human, mouse and rat were shown in Figure 4. In order to analyze which degree of conservation can accurately predict the PTM sites, the true positive rate and the false positive rate were calculated as shown in Table 1. The true positive rate of the degree of specific conservation is 0.827. This value is higher than that of the degree of conservation.

Meanwhile, the false positive rate of the degree of specific conservation is higher that of the degree of conservation. In addition, we created ROC curves of PTM sites and each degrees of conservation shown in Figure 5. The AUC (0.773) of the degree of specific conservation is higher than the AUC (0.649) of the degree of conservation.

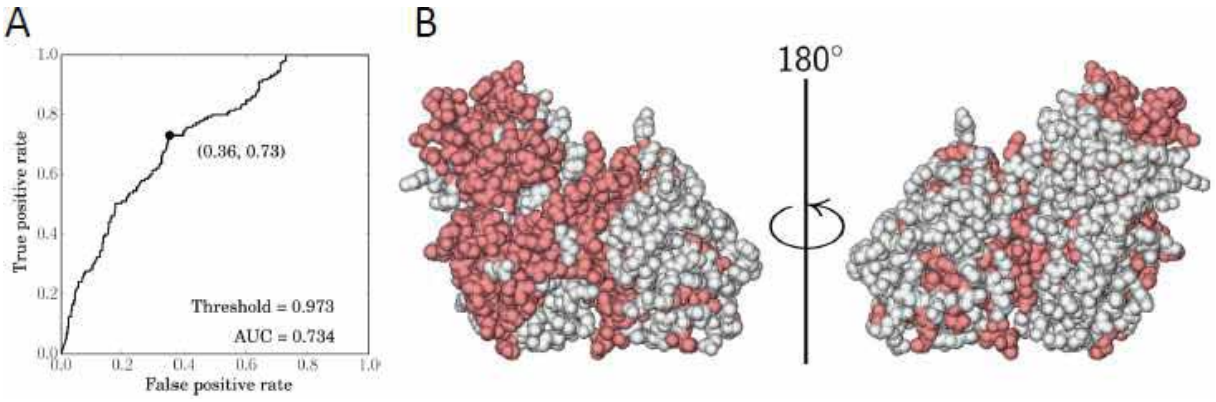

Figure 1. Prediction by degree of conservation. (A) An ROC curve created by degree of conservation and proximity that is derived from complex structures of EF-Tu/1As in the PDB [32]. The ROC curve was constructed by using residues binding with ions or molecules within 3 Å as functional sites. The threshold for the degree of conservation was determined by the black point, which makes the maximum of the sum of the true positive rate and the false positive rate. (B) The three-dimensional structure of rabbit eEF1A2 (PDBID: 4C0S) [33] visualized onto the degree of conservation. If the degree of conservation is higher than or equal to the threshold, the residue was colored by pink. If not, the residue was colored by white.

Table 1. Accuracy for predicting PTM sites by degree of conservation or specific conservation.

	True Positive Rate	False Positive Rate
C_α	0.395	0.291
C_β	0.827	0.376

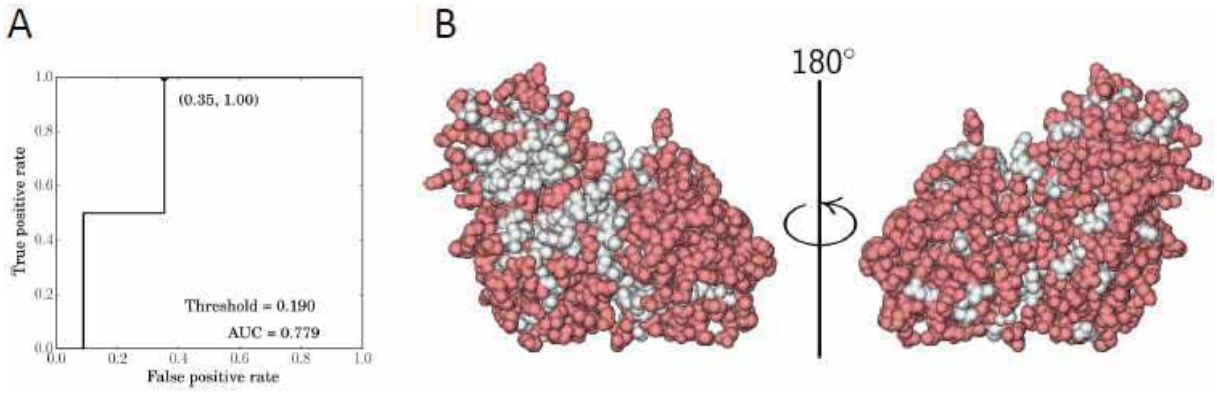

Figure 2. Prediction by degree of specific conservation. (A) An ROC curve created by degree of specific conservation and actin binding sites. The functional sites are actin binding residues determined by site-directed mutational analysis for yeast eEF1A. The threshold for the degree of specific conservation was determined by the black point, which makes the maximum of the sum of the true positive rate and the false positive rate. (B) The three-dimensional structure of rabbit eEF1A2 (PDBID: 4C0S) visualized onto the degree of specific conservation. If the degree of specific conservation is higher than or equal to the threshold, the residue was colored by pink. If not, the residue was colored by white.

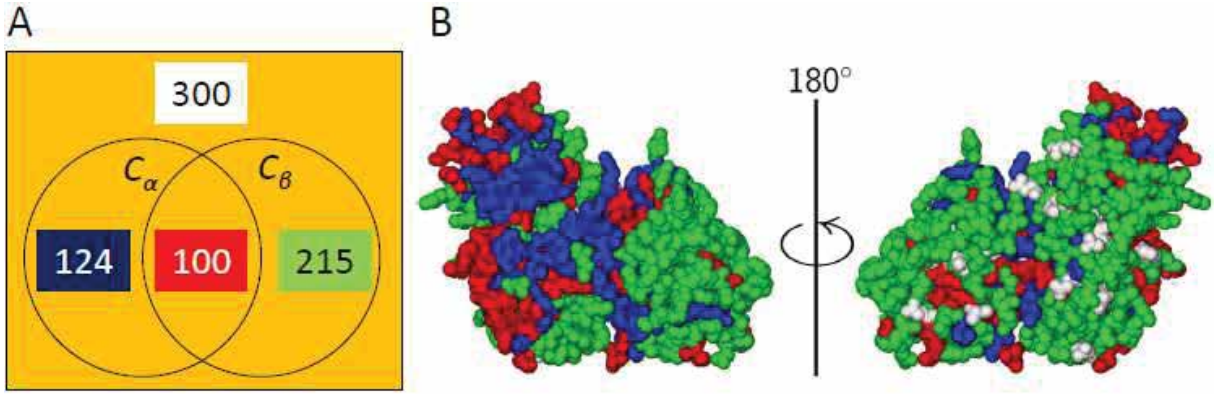

Figure 3. Prediction overlapping with degrees of conservation. (A) The numbers of overlapping predicted sites and other sites by degrees of conservation and specific conservation. (B) The overlapping and not overlapping residues visualized onto the three-dimensional structure of rabbit eEF1A2 (PDBID: 4C0S). Sites predicted by both methods were shown by red, sites only predicted by C_α were shown by blue, sites only predicted by C_β were shown by green and sites not predicted by both methods were shown by white.

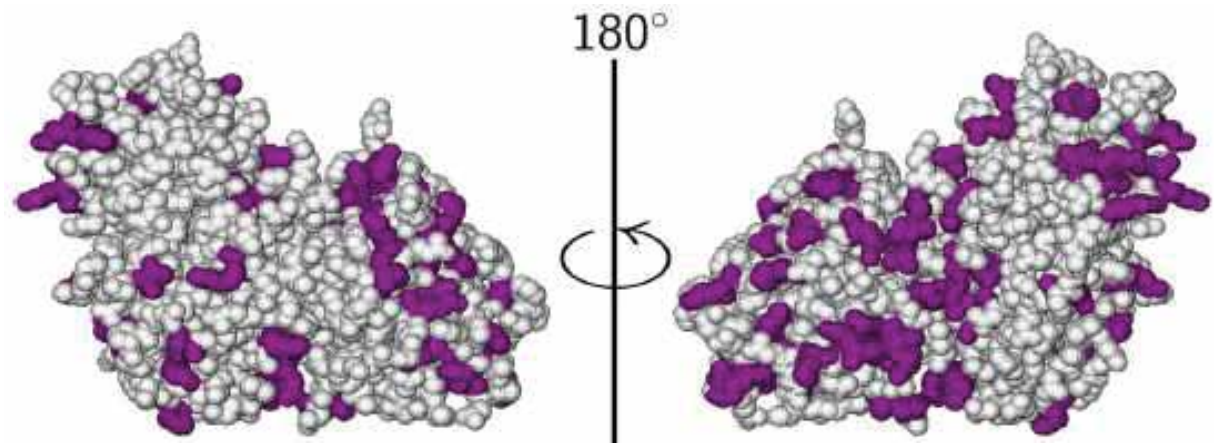

Figure 4. PTM sites of eEF1A. PTM sites visualized onto rabbit eEF1A2 (PDBID: 4C0S). If the residue has PTM, the residue is shown by purple

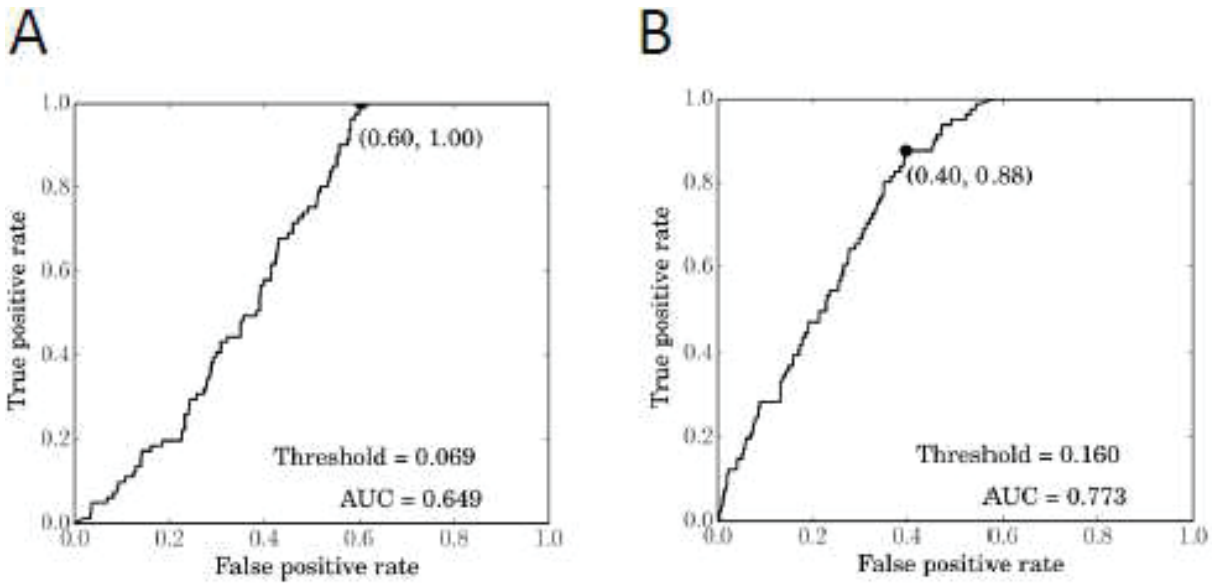

Figure 5. ROC curves of PTM sites. (A) An ROC curve created by degree of conservation and PTM sites. (B) An ROC curve created by degree of specific conservation and PTM sites

Discussion

In this paper, we investigated the conserved sites of eEF1A using two degrees of conservation. One is the degree that can identify the sites which show a conserved pattern in all proteins of a protein family. As shown in our previous study [21], the degree of conservation can predict close sites from binding molecules in the three-dimensional structures of eEF-Tu/1A. eEF1A has the conserved face and the variable face in the three-dimensional structure [15]. These faces were decided from variable residues between eEF1A1 and eEF1A2. The conserved face is similar to the regions that are determined by the degree of conservation shown in Figure 1B. This shows that the degree of conservation can predict the conserved face of eEF1A.

The other degree of conservation is a specific conservation that can identify the sites that are only conserved in a subfamily of a protein family. In other words, the degree of specific conservation can identify the sites conserved in the subfamily but variable in the whole family. In this study, the subfamily is specified as the group that contains vertebrate eEF1A1 and eEF1A2. The whole family is specified as the group that contains whole eEF-Tu/1A family. Therefore, the degree of specific conservation predicts few non-identical sites between eEF1A1 and eEF1A2. However, the degree of specific conservation can mainly predict close sites from the non-identical residues shown in Figure 2. This suggests that the degree of specific conservation can predict the residues that are important for functional divergences between eEF1A1 and eEF1A2. In addition, the regions that are predicted from the degree of specific conservation are similar to the variable face described above. Because there are many PTM sites in the variable face, the specific conservation may be useful for predicting PTM sites.

As shown in Table 1 and Figure 5, our results show that the PTM sites were predicted well by the degree of specific conservation than degree of conservation. In addition, Table 1 shows that the true positive rate is high but the false positive rate is high in the degree of specific conservation. Therefore, the degree of specific conservation can accurately predict the known PTM sites that are registered in the database of PTM sites. This shows that the degree of specific conservation is superior to the degree of conservation to predict PTM sites of eEF1A proteins. These results suggest that one degree of conservation is inadequate to predict all important residues of eEF1A.

In conclusion, various degrees of conservation may be useful when we predict functional sites of multifunctional proteins like eEF1A proteins. In addition, our results suggest that PTM sites of eEF1A1and eEF1A2 are not conserved in the whole family but conserved in the subfamily.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Andersen G, Nissen P, Nyborg J. Elongation factors in protein biosynthesis. Trends Biochem Sci. 2003; 28: 434-441.
2. Ejiri S. Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization. Biosci Biotechnol Biochem. 2002; 66: 1-21.
3. Mateyak MK, Kinzy TG. eEF1A: Thinking Outside the Ribosome. J Biol Chem. 2010; 285: 21209-21213.
4. Murray J, Edmonds B, Liu G, et al. Bundling of actin filaments by elongation factor 1 alpha inhibits polymerization at filament ends. J Cell Biol. 1996; 135: 1309-1321.
5. Itagaki K, Naito T, Iwakiri R, et al. Eukaryotic translation elongation factor 1A induces anoikis by triggering cell detachment. J Biol Chem. 2012; 287: 16037-16046.
6. Matsunaga T, Fukai F, Miura S, et al., Combination therapy of an antican-

cer drug with the FNIII14 peptide of fibronectin effectively overcomes cell adhesion-mediated drug resistance of acute myelogenous leukemia. Leukemia. 2008; 22: 353-360.

7. Knudsen S, Frydenberg J, Clark B, et al. Tissue-dependent variation in the expression of elongation factor-1-alpha isoforms - isolation and characterization of a cdna-encoding a novel variant of human elongation-factor-1-alpha. European J Biochem. 1993; 215: 549-554.
8. Lund A, Knudsen S, Vissing H, et al. Assignment of human elongation factor 1 alpha genes: EEF1A maps to chromosome 6q14 and EEF1A2 to 20q13.3. Genomics. 1996; 36:359-361.
9. Lee S, Francoeur A, Liu S, et al. Tissue-specific expression in mammalian brain, heart, and muscle of S1, a member of the elongation factor 1 alpha gene family. J Biol Chem. 1992; 267:24064-24068.
10. Kahns S, Lund A, Kristensen P, et al. The elongation factor 1 A-2 isoform from rabbit: cloning of the cDNA and characterization of the protein. Nucleic Acids Res. 1998; 26:1884-1890.
11. Duttaroy A, Bourbeau D, Wang X, et al. Apoptosis rate can be accelerated or decelerated by overexpression or reduction of the level of elongation factor-1 alpha. Experimental Cell Res. 1998; 238: 168-176.
12. Anand N, Murthy S, Amann G, et al. Gene encoding protein elongation factor EEF1A2 is a putative oncogene in ovarian cancer. Nature Genet. 2002; 31: 301-305.
13. Soares DC, Barlow PN, Newbery HJ, et al. Structural models of human eEF1A1 and eEF1A2 reveal two distinct surface clusters of sequence variation and potential differences in phosphorylation. PLoS ONE. 2009; 4.
14. Balasubramanian S, Kannan TR, Hart PJ, et al. Amino acid changes in elongation factor Tu of Mycoplasma pneumoniae and Mycoplasma genitalium influence fibronectin binding. Infect Immun. 2009; 77: 3533-3541.
15. Soares DC, Abbott CM, Highly homologous eEF1A1 and eEF1A2 exhibit differential post-translational modification with significant enrichment around localised sites of sequence variation. Biology Direct. 2013; 8.
16. Jakobsson ME, Matecki J, Nilges BS, et al. Methylation of human eukaryotic elongation factor alpha (eEF1A) by a member of a novel protein lysine methyltransferase family modulates mRNA translation. Nucleic Acids Res. 2017; 45: 8239-8254.
17. Jank T, Belyi Y, Wirth C, et al. Protein glutaminylation is a yeast-specific posttranslational modification of elongation factor 1A. J Biol Chem. 2017; 292: 16014-16023.
18. Hamey JJ, Wilkins MR. Methylation of elongation factor 1a: Where, who, and why? Trend Biochem Sci. 2018; 43: 211 - 223.
19. Cavallius J, ZollW, Chakraburtty K, et al. Characterization of yeast ef-1: Non-conservation of post-translational modifications. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology. 1993; 1163: 75 - 80.
20. Kondo Y, Kwon Y, Miyazaki S. Detection of key residues involving functional divergence into the translation elongation factor Tu/1A family using quantitative measurements for specific conservation of protein subfamilies. J Comput Sci Syst Biol. 2014; 7: 054-061.
21. Kondo Y, Miyazaki S. Protein functional site prediction using a conservative grade and a proximate grade. J Data Mining Genomics Proteomics. 2015; 6: 175.
22. Consortium TU. Uniprot: a hub for protein information. Nucleic Acids Res 2015; 43: D204-D212.
23. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772-780.
24. Kishino H, Miyata T, et al. Maximum-likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol. 1990; 31: 151-160.
25. Jones D, Taylor W, Thornton J. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992; 8: 275-282.
26. Dayhoff MO, Schwartz RM. 1978 Chapter 22: A model of evolutionary change in proteins. In: in Atlas of Protein Sequence and Structure.
27. Sneath PHA, Sokal RR. 1973 Numerical taxonomy: the principles and practice of numerical classification. San Francisco: W. H. Freeman.
28. Benner S, Cohen M, Gonnet G. Amino-acid substitution during functionally constrained divergent evolution of protein sequences. Protein Eng. 1994; 7: 1323-1332.
29. Sibbald P, Argos P. Weighting aligned protein or nucleic acid sequences to correct for unequal representation. J Mol Biol. 1990; 216: 813-818.
30. Gross SR, Kinzy TG. Improper organization of the actin cytoskeleton affects protein synthesis at initiation. Mol Cell Biol. 2007; 27: 1974-1989.
31. Hornbeck PV, Zhang B, Murray B, et al. PhosphoSitePlus, 2014:mutations, PTMs and recalibrations. Nucleic Acids Res. 2015; 43: D512-D520.
32. Berman H, Westbrook J, Feng Z, et al. The protein data bank. Nucleic Acids Res. 2000; 28: 235-242.
33. Crepin T, Shalak VF, Yaremchuk AD, et al., Mammalian translation elongation factor eEF1A2: X-ray structure and new features of GDP/GTP exchange mechanism in higher eukaryotes. Nucleic Acids Res. 2014; 42: 12939-12948.

To cite this article: Kondo Y, Miyazaki S. Evolutionary Analysis of Post-translational Modification Sites in Translation Elongation Factor 1A. Japan Journal of Medicine. 2018: 1:6. doi: 10.31488/jjm.1000129