

Detection of Key Residues Involving Functional Divergence into the Translation Elongation Factor Tu/1A Family Using Quantitative Measurements for Specific Conservation of Protein Subfamilies

Yosuke Kondo, Yeondae Kwon and Satoru Miyazaki*

Department of Medical and Life Science, Faculty of Pharmaceutical Sciences, Japan

Abstract

Delivery of an aminoacyl-tRNA to the ribosomal A-site during protein biosynthesis is mediated by elongation factor Tu/1A (EF-Tu/1A). This function is inferred as a common function of the EF-Tu/1A family. Moonlighting functions and several functional divergences are speculated in the EF-Tu/1A molecules such as actin and fibronectin binding functions. Two variant eEF1A forms, referred to as eEF1A1 and eEF1A2, are surmised to have different actin binding affinities. *Mycoplasma pneumoniae* EF-Tu has higher fibronectin binding affinity than *M. genitalium* EF-Tu. Incidentally, quantitative description for specific conservation of protein subfamilies could be helpful for assessment of the functional differences. Our paper defines two types of variability measurements of a multiple sequence alignment site. One is based upon a substitution matrix and sequence weights. The other is based upon information entropy. These variabilities are converted into a specific conservation score by the comparison of different two groups in the evolutionary branches including a target protein. Our paper describes whether the conservation score can divide different residues between two sequences with functional differences into the actin or fibronectin binding residues and the others. The result shows that the functional divergence involving the actin and fibronectin binding functions of the EF-Tu/1A molecules highly correlates with the evolutionary branches supposedly dividing their sequences.

This implies that an inherent property of amino acids is an essential factor for the functional differences of the actin and fibronectin binding residues. Our paper describes one possible story for identification of key residues involving functional divergences of the human eEF1A1 and eEF1A2 molecules under the conjecture that the property of amino acids is critical. We expect that such quantitative approach is effective for further assessment of functional differences of the EF-Tu/1A subfamilies and helpful for detection of key residues involving functional divergence of protein families.

Introduction

The well-known function of the translation elongation factor Tu (EF-Tu) and eukaryotic translation elongation factor 1A (eEF1A) molecules is a carrier of aminoacyl-tRNAs to the A site of the ribosome in bacteria and eukaryotes, respectively [1]. Then, the EF-1A-GTP-tRNA complex elongates a new polypeptide chain upon the ribosome. Such an elongation event triggers the ribosome to induce GTP hydrolysis. The inactive GDP forms of the eEF1A molecule are recycled to the active GTP form by guanine nucleotide exchange factors including eukaryotic translation elongation factor 1B α (eEF1B α) [2]. These functions are inferred as common functions of the EF-Tu/1A family.

Some species have several eEF1A encoding genes. There are two variant human eEF1A forms, referred to as eEF1A1 and eEF1A2, shared with ~96% similarity between their protein sequences [3]. Both eEF1A proteins seem to exhibit similar translation activities although they have different binding affinities for the GTP/GDP and eEF1B α molecules [4,5]. The expression patterns of the two isoforms are exclusive in the human tissues [6,7]. eEF1A2 is highly expressed in skeletal muscle, heart muscle and brain whereas eEF1A1 is lowly expressed in these tissues but highly expressed in other tissues like lung, liver, and placenta. Loss of expression of eEF1A2 in mice has an effect on motor neuron degradation [8,9].

Moonlighting functions have been observed in the eEF1A proteins [10,11]. In around 1990, *Dictyostelium discoideum* eEF1A was identified as an actin-binding protein [12]. Gross et al. [13] have identified actin binding residues by site-directed mutagenesis studies of the yeast eEF1A molecule. *Saccharomyces cerevisiae* has two eEF1A encoding genes [14]. Overexpression of these genes resulted in defective budding

and enlarged cells [15]. eEF1A2 proteins induce filopodia production in rodent and human cell lines [16]. This indicates an important role for eEF1A2 in controlling actin remodeling and cell motility. eEF1A2 is likely to be an important human oncogene [17]. Overexpression of the eEF1A2 protein causes ovarian cancer probably triggered by inhibition of apoptosis [18]. The protection mechanism exerted by the eEF1A2 protein may correlate with regulation of caspase-3 activation whereas the increase in eEF1A1 protein levels may facilitate rapid death of cells [19]. Elucidation of the complete mechanism how the eEF1A proteins regulate these cellular processes would be a painstaking task because they have numerous moonlighting functions.

Some bacteria seem to capitalize the membrane localized EF-Tu to infect their host cell. *Mycoplasma* species can be found in many different hosts but individual species have a strict host, organ and tissue specificity [20]. *Mycoplasma pneumoniae* is an established pathogen

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

Received December 20, 2013; Accepted January 15, 2013; Published January 18, 2014

Citation: Kondo Y, Kwon Y, Miyazaki S (2014) 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 7: 054-061. doi:10.4172/jcsb.1000138

Copyright: © 2014 Kondo Y, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

of the human respiratory tract [21] but has also been isolated from the urogenital tract [22]. *M. genitalium*, an emerging sexually transmitted disease pathogen [23], has also been associated with the human respiratory tract [24]. The tissue specific tropisms could be determined by genetic distinctions between them. Balasubramanian et al. [25] have shown that *M. pneumoniae* EF-Tu (MpEF-Tu) has higher fibronectin binding affinity than *M. genitalium* EF-Tu (MgEF-Tu). They have identified fibronectin binding residues by site-directed mutagenesis studies of the MpEF-Tu molecule.

Currently, many sequence data are accumulated into protein sequence databases. This makes more attractive to capture functional information from their sequences. There are many scoring methodologies for a multiple sequence alignment site of protein families [26]. Lichtarge et al. [27] have proposed an evolutionary trace method. This method can detect common functional surfaces of a protein family. Landgraf et al. [28] have introduced a specific conservation score based upon the variability consisting of a substitution matrix and sequence weights. The conservation score is effective not to detect common functional sites but to estimate specific sites of a protein subfamily. Mihalek et al. [29] have proposed the variability based upon information entropy. They have also proposed integrating methodology for quantitative measurements using all evolutionary branches of a protein family or only evolutionary branches including a target protein.

Our paper describes whether the specific conservation score can detect both actin and fibronectin binding residues which have been already identified by wet laboratory experiments of the EF-Tu/1A molecules. We define a novel scoring methodology, which is developed in the process of the verification, for multiple sequence alignment sites. Then, we discuss the inherent property of the actin and fibronectin binding residues and propose one possible story for assessment of key residues involving functional divergences between the human eEF1A1 and eEF1A2 molecules.

Materials and Methods

Data collection

1,024 entries were taken from UniProtKB/Swiss-Prot release 2013 11 (541,762 entries) [30] by searching for "GTP-binding elongation factor family. EF-Tu/EF-1A subfamily" in the annotation of protein similarities. 41 entries were excluded because 39 entries were annotated as a fragment and 2 entries included "X" in their sequences. Consequently, 983 entries were retained as a protein set of the EF-Tu/1A family.

Multiple sequence alignment and phylogenetic tree

Multiple sequence alignment of the protein set was performed by using the MAFFT 7 program [31]. Construction of a distance matrix was performed by the PHYLIP protdist program [32] with the Jones-Taylor-Thornton model [33] as an amino acid substitution model. A phylogenetic tree was written by the unweighted pair group method with arithmetic mean (UPGMA) using the PHYLIP neighbor program. The sequence alignment and phylogenetic tree were visualized by the Discovery Studio 3.1 software package [34] and the MEGA 5.2 software package [35], respectively. The 983 sequences were divided into groups at each node of the phylogenetic tree.

Variability based upon a substitution matrix and sequence weights

This paper employed the variability measure which was proposed

by Landgraf et al. [28]. The n th alignment site of a group g of variability $V_{g,n}^{sw}$ was defined as

$$V_{g,n}^{sw} = \frac{1}{N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{S(a_{g,i}, a_{g,i}) - S(a_{g,i}, a_{g,j})}{S(a_{g,i}, a_{g,i})} \times W_{g,i} \quad (1)$$

where $S(a_{g,i}, a_{g,j})$ is a substitution score between amino acids $a_{g,i}$ and $a_{g,j}$ and $W_{g,i}$ is a weight of the i th sequence of the group g . The sequence weight was estimated by the Sibbald and Argos methodology [36]. Then, random sequences were created by 100,000 iterations. The Gonnet substitution matrix was employed as a mutation model of amino acids [37].

Variability based upon information entropy

This paper employed another variability measure whose basis was proposed by Mihalek et al. [29]. The n th alignment site of a group g of variability $V_{g,n}^{ie}$ was defined as:

$$V_{g,n}^{ie} = - \sum_{a=1}^{21} \frac{f_{g,n,a}}{N_g} \log \frac{f_{g,n,a}}{N_g} \quad (2)$$

Where $f_{g,n,a}$ is frequency of an amino acid of the n th alignment site of the group g and N_g is the number of sequences in the group g . Here, a gap site is regarded as an extra amino acid.

Conservation score

The n th alignment site of a base group b_g and a target group t_g of a conservation score $C_n(b_g, t_g)$ was defined as

$$C_n(b_g, t_g) = \frac{V_{b_g,n} + 1}{V_{t_g,n} + 1} - 1 \quad (3)$$

This paper dealt with the case that the base group includes the target group and employed both variabilities described above.

ROC curve

Construction of a receiver operating characteristic (ROC) curve and calculation of an area under the curve (AUC) were conducted by the pROC R package [38]. A heat map of the AUC was created by the matplotlib Python package [39].

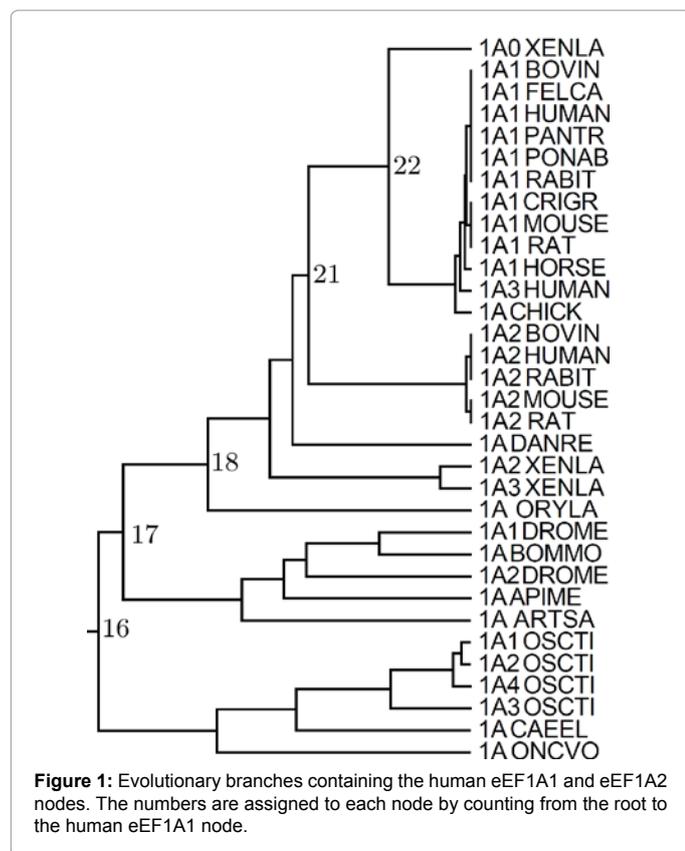
Results

Alignment and grouping of the EF-Tu/1A family

There were 831 sites into the multiple sequence alignment of the 983 EF-Tu/1A sequences. After the phylogenetic tree of the EF-Tu/1A family was written by using all alignment sites, division of each node created groups containing sequences of EF-Tu/1A subfamilies.

Analysis of the human eEF1A1 and eEF1A2 molecules

Figure 1 shows that the human eEF1A1 and eEF1A2 nodes are divided into the 21st node when numbers 1 to 28 is assigned to each node from the root to the human eEF1A1 node. Figure 2 shows that there are 36 different sites between the human eEF1A1 and eEF1A2 sequences. Figure 3A shows that the conservation score set between the 21st and 22nd nodes is the second highest score set. The highest score set is the score set between the 6th and 8th nodes. Figure 3B shows that the conservation score set between the 21st and 22nd nodes is the highest score set. The second highest score set is the score set between the 6th and 8th nodes. Table 1 show that each value between the 21st and 22nd nodes is higher than each value between the 6th and 8th nodes in both variabilities.



Analysis of the MpEF-Tu and MgEF-Tu molecules

Figure 4 shows that the MpEF-Tu and MgEF-Tu nodes are divided into the 23rd node when numbers 1 to 24 are assigned to each node from the root to the MpEF-Tu node. Figure 5 shows that there are 13 different sites between the MpEF-Tu and MgEF-Tu sequences. Figure 6A shows that the conservation score set between the 23rd and 24th nodes is the second highest score set. The highest score set is the score set between the 7th and 9th nodes. Figure 6B shows that the AUC value of the conservation score set between the 23rd and 24th nodes is 0.5. The highest score set is the score set between the 6th and 10th nodes.

Table 2 shows that each value between the 23rd and 24th nodes of both variabilities is higher than each value between the 7th and 9th nodes and the 6th and 10th nodes.

Discussion

Conservation scores based upon two types of variabilities

Landgraf et al. [28] have proposed the variability based upon a substitution matrix and sequence weights for quantitative description of a multiple sequence alignment site. The range of the variability is 0.0 to 1.0. The higher the value implies, the more the alignment site consists of hardly substituting amino acids.

Mihalek et al. [29] have proposed the variability based upon information entropy. The range of the variability is 0.0 to ln 20. The higher the value implies, the more the alignment site consists of various letters. In our paper, two changes have been added to this variability. One is that a gap site is regarded as an extra amino acid. The other is that the base of the logarithm is 21. Hereby, the range of this variability is 0.0 to 1.0.

Landgraf et al. [28] have also proposed a specific conservation score using their variability. The score is obtained from two variabilities. One is the variability obtained from all sequences of a multiple sequence alignment site. The other is the variability obtained from a group, which includes a target protein, defined by a node of the phylogenetic tree. The range of the score is -0.5 to 1.0. The higher the score implies, the more the variability of the group relatively decreases than the variability of all sequences. They have discussed influences of grouping by setting to a threshold of the specific conservation score using random sequence clusters and differences of the residues with above the threshold using several groups. However, the conservation score obtained from their grouping method strongly depends upon how to construct a protein sequence set. Our study have employed comprehensive grouping of the evolutionary branches of a target protein. Then, we have analyzed whether the conservation score set can divide different residues between two sequences with functional divergence into the residues involving the functional difference and the others. Hereby, we could determine an evolutionary divergence highly correlating with the functional divergence.

Actin binding residues

Our study explores two EF-Tu/1A subfamilies, which make both specific conservation scores of 2 actin binding residues relative higher values than other 34 different residues between the human eEF1A1 and eEF1A2 molecules, in the evolutionary branches including the human eEF1A1 node. Conservation score sets of both variabilities with the first and second highest AUCs have included the score sets between the 21st and 22nd nodes of Figure 1. This implies that the functional divergence involving actin binding between the human eEF1A1 and eEF1A2 molecules highly correlates with the node dividing the human eEF1A1 and eEF1A2 sequences. Figure 2 shows that there are 29 sites consisting of unique letters in the sequences of the 22nd node of Figure 1. This implies that the value of the conservation score depends upon the substituting score of amino acids. This suggests that the property of amino acids is an essential factor for the functional difference of the actin binding function.

Conservation score sets of both variabilities with the first and second highest AUCs have also included the score sets between 6th and 8th nodes. Comparisons of the conservation score sets of 6th and 8th nodes and 21st and 22nd nodes show that the former values are considerable lower than the latter values. This shows that the conservation score sets between 6th and 8th nodes are interesting score sets although the high correlations might happen to obtain.

Key residues involving functional divergence

Table 1 shows that the first and second highest values of 36 conservation scores based upon information entropy are 0.25547 and 0.19407, respectively. The human eEF1A1 and eEF1A2 residues with these conservation scores include the 2 actin binding residues. The values of other 27 residues are also 0.19407. This shows that we cannot describe differences of the 27 residues by using the conservation scores based upon information entropy.

Conversely, there are various values in the conservation scores based upon a substitution matrix and sequence weights. We propose one possible story for assessment of functional differences of the human eEF1A1 and eEF1A2 molecules under the conjecture that the inherent property of amino acids is critical for their divergence. Here, the higher conservation score implies, the more the property of amino



Figure 2: Multiple sequence alignment of eEF1A1 and eEF1A2 sequences. The 18 sequences were extracted from the multiple sequence alignment of the EF-Tu/1A family. Alignment sites without amino acid residues of the 18 sequences were excluded. A white background site consists of unique amino acids and a black background site consists of two or more types of letters.

Residues	$C_n^{sw} (6, 8)$	$C_n^{sw} (21, 22)$	$C_n^{io} (6, 8)$	$C_n^{io} (21, 22)$	Actin binding
S83T	0.02403	0.08153	0.01511	0.19407	
V87I	-0.00468	0.06226	0.00442	0.19407	
P161A	0.02264	0.24512	0.00125	0.19407	
Q164E	0.03766	0.13165	0.01125	0.19407	
E168D	0.00917	0.13241	0.00984	0.19407	
T176A	0.00567	0.15339	0.00740	0.19407	
D186A	-0.00321	0.24408	0.00252	0.19407	
A189P	-0.00725	0.28690	-0.00221	0.19407	
N197H	-0.00107	0.25931	0.00426	0.19407	
A206P	0.00098	0.05119	0.01027	0.08149	
T217E	0.02192	0.23347	0.01228	0.19407	
D220E	0.01581	0.03949	-0.00070	0.11018	
T226V	-0.00024	0.22298	0.00951	0.19407	
T227S	-0.00233	0.07602	0.00579	0.19407	
C234T	0.00033	0.29774	0.00084	0.19407	
V271I	-0.00281	0.06226	-0.00071	0.19407	
K273R	-0.00389	0.11880	-0.00113	0.19407	
V285I	0.03537	0.06226	0.01110	0.19407	
V320I	0.00474	0.06226	-0.00023	0.19407	
A326C	0.00771	0.31076	0.00082	0.19407	
N331S	0.00691	0.20197	0.00510	0.25547	X
M335Q	0.03814	0.27765	0.01750	0.19407	X
G339Q	-0.02090	0.16000	-0.00380	0.15548	
A342S	-0.00375	0.10550	-0.00168	0.19407	
A358S	0.03212	0.10550	0.01011	0.19407	
L361I	0.02974	0.02915	0.01903	0.15548	
G390N	-0.00206	0.08175	-0.00219	0.11018	
F393S	0.01559	0.34059	0.02170	0.19407	
D403E	0.00208	0.11476	0.01418	0.19407	
D417Q	-0.00424	0.19312	-0.00304	0.19407	
A440N	-0.01732	0.24580	-0.00275	0.19407	
D442E	0.00573	0.03949	0.00231	0.11018	
A445S	-0.00473	0.10550	0.00362	0.19407	
A446G	-0.00585	0.14825	0.00157	0.11018	
K462G	-0.02254	0.33389	-0.00195	0.19407	
-463K	-0.00848	0.23160	-0.00183	0.19407	
AUC	0.83824	0.75000	0.77941	0.80147	
Threshold	0.00632	0.19754	0.00476	0.22477	
Sensitivity	0.67647	0.64709	0.61765	1.00000	
Specificity	1.00000	1.00000	1.00000	0.50000	

Table 1: Four conservation score sets with the first and second highest AUC in 378 (= $_{28}C_2$) conservation score sets of both variabilities. The "X" letters denote the human eEF1A1 and eEF1A2 residues which are equivalent to the N329 and K333 residues involving actin binding of the yeast eEF1A molecule [13]. The threshold shows the arithmetic mean between two conservation scores whose sum of the sensitivity and specificity becomes maximum when the threshold is used as a cutoff value of the conservation score set.

acids of the site is changed. Soares et al. [40] have created structure models of the human eEF1A1 and eEF1A2 proteins by homology modeling using the yeast eEF1A as a template. They have compared interaction surfaces and different residues between the human eEF1A1 and eEF1A2 proteins based upon location of the three-dimensional structures. Table 3 shows hypothetical functional residues described by Soares et al. [40]. The N197H, A326C, and F393S residues have the highest values in the hypothetical residues involving guanosine binding, actin binding, and phosphorylation, respectively. This shows that quantitative measurement by the conservation score allows us to determine key residues involving functional divergences of protein subfamilies. However, since the story is only a possibility, further

assessment is necessary for identification of the key residues beyond the actin binding function.

Fibronectin binding residues

Our study explores two EF-Tu/1A subfamilies, which make specific conservation scores of 6 fibronectin binding residues relative higher values than other 7 different residues between the MpEF-Tu and MgEF-Tu molecules, in the evolutionary branches including the MpEF-Tu node. The conservation score set with the second highest AUC is the score set between the node dividing the MpEF-Tu and MgEF-Tu sequences and the MpEF-Tu node in the variability based upon a substitution matrix and sequence weights. This implies that the functional divergence involving fibronectin binding between the MpEF-Tu and MgEF-Tu molecules highly correlates with the node dividing the MpEF-Tu and MgEF-Tu sequences. The value of the conservation score depends upon the substituting scores of amino acids because the variability of a leaf node is 0. This suggests that the property of amino acids is an essential factor for the functional difference of the fibronectin binding function.

Balasubramanian et al. [25] have created the structure model of the MpEF-Tu protein by homology modeling using the *Thermus thermophilus* EF-Tu as a template. The fibronectin binding residues,

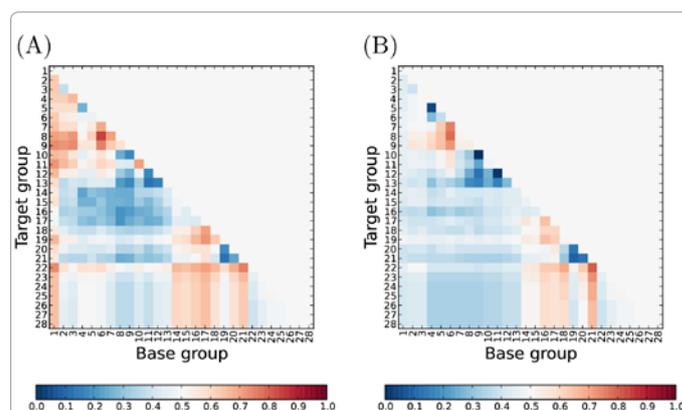


Figure 3: Two heat maps of areas under the ROC curve of 378 (= 28 C₂) conservation score sets using variabilities based upon (A) a substitution matrix and sequence weights and (B) information entropy. The AUC of the conservation score set of 36 different residues between the human eEF1A1 and eEF1A2 sequences was calculated by using the N331S and M335Q residues, which are equivalent to the N329 and K333 residues involving actin binding of the yeast eEF1A molecule [13], as learning data. The positive direction of the learning data defined as ascending order of the 13 conservation scores.

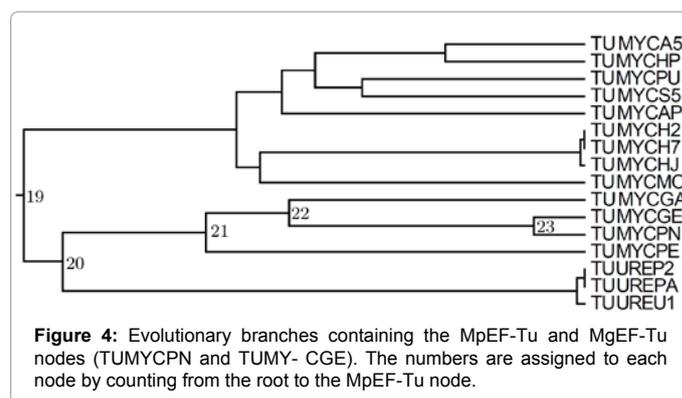


Figure 4: Evolutionary branches containing the MpEF-Tu and MgEF-Tu nodes (TUMYCPN and TUMYCGE). The numbers are assigned to each node by counting from the root to the MpEF-Tu node.

which are divided into binding regions 1 and 2, are surface accessible residues in the three-dimensional structure. Our result shows that these surface exposed residues have very high correlation with the conservation score sets because Table 2 shows the large AUC values.

The conservation score sets with the highest AUCs are the

score sets between 7th and 9th nodes and 6th and 10th nodes in the variabilities based upon a substitution matrix and sequence weights and information entropy, respectively. Comparisons of the conservation score sets of these nodes and 23rd and 24th nodes show that the former values are considerable lower than the latter values. This shows that the

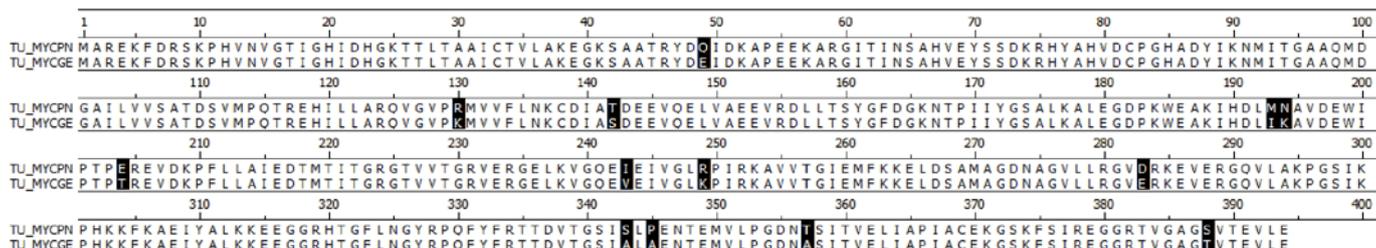


Figure 5: Pairwise sequence alignment of the MpEF-Tu and MgEF-Tu sequences. The 2 sequences were extracted from the multiple sequence alignment of the EF-Tu/1A family. Alignment sites without the MpEF-Tu and MgEF-Tu residues were excluded. A white background site consists of unique amino acids and a black background site consists of two types of letters.

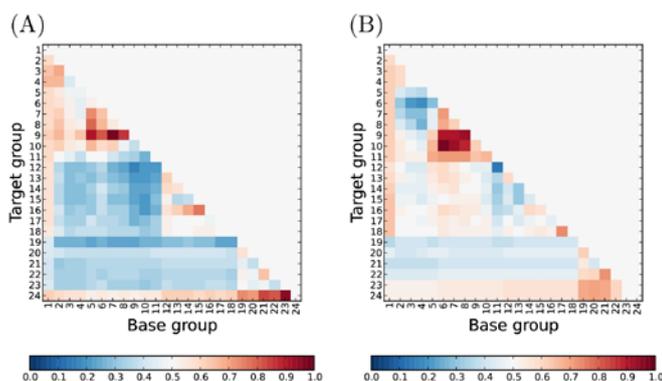


Figure 6: Two heat maps of areas under the ROC curve of 276 (= 24 C₂) conservation score sets using variabilities based upon (A) a substitution matrix and sequence weights and (B) information entropy. The AUC of the conservation score set of 13 different residues between the MpEF-Tu and MgEF-Tu sequences was calculated by using the M193, N194, E204, S343, P345, and T357 residues involving fibronectin binding of the MpEF-Tu molecule [25] as learning data. The positive direction of the learning data defined as ascending order of the 13 conservation scores.

Residues	C _n ^{sw} (7, 9)	C _n ^{sw} (23, 24)	C _n ^{ie} (6, 10)	C _n ^{ie} (23, 24)	Fibronectin binding
Q49E	-0.00082	0.14000	0.00043	0.22767	
R130K	-0.00006	0.11945	0.00081	0.22767	
T142S	0.00079	0.08863	0.00038	0.22767	
M193I	0.00278	0.15784	0.00323	0.22767	X
N194K	0.00166	0.23084	0.00064	0.22767	X
E204T	0.00253	0.25597	0.00271	0.22767	
I243V	-0.00197	0.06348	-0.00074	0.22767	
R249K	-0.00161	0.11945	-0.00013	0.22767	
D283E	-0.00022	0.13905	0.00058	0.22767	
S343A	0.00052	0.12061	0.00134	0.22767	X
P345A	0.00190	0.29930	0.00307	0.22767	X
T357A	0.00280	0.17341	0.00179	0.22767	X
S388T	-0.00322	0.08863	0.00059	0.22767	
AUC	0.97619	0.95238	0.97619	0.50000	
Threshold	0.00023	0.14892	0.00061		
Sensitivity	0.85714	1.00000	0.85714		
Specificity	1.00000	0.83333	1.00000		

Table 2: Two conservation score sets with the first and second highest AUC of the variability based upon a substitution matrix and sequence weights and two conservation score sets with the highest AUC and between the 23rd and 24th nodes of the variability based upon information entropy in 276 (= 24 C₂) conservation score sets, respectively. The "X" letter denotes the MpEF-Tu residue involving fibronectin binding [25]. The threshold shows the arithmetic mean between two conservation scores whose sum of the sensitivity and specificity becomes maximum when the threshold is used as a cutoff value of the conservation score set.

Residues	Hypothetical function
Q164E	Adjacent to the guanosine binding pocket
E168D	Adjacent to the guanosine binding pocket
N197H	Adjacent to the guanosine binding pocket
A206P	Adjacent to the guanosine binding pocket
T217E	Phosphorylatable
T227S	Phosphorylatable
V320I	Adjacent to aminoacyl-tRNA and eEF1B α binding sites
A326C	Adjacent to an actin binding residue
N331S	Actin binding
M335Q	Actin binding
A358S	Adjacent to an actin binding residue
F393S	Phosphorylatable
A445S	Phosphorylatable

Table 3: Hypothetical residues estimated from the human eEF1A1 and eEF1A2 structure models and potential phosphorylatable residues [40].

conservation score sets between former nodes are interesting score sets although the high correlations might occur by chance.

Conclusion

Our paper has described quantitative measurements for specific conservation of protein subfamilies. We have developed an exhaustive grouping approach, which is subject to all evolutionary branches involving the target nodes, to calculate the conservation scores. This approach has shown that the functional divergences involving actin and fibronectin binding of the EF-Tu/1A molecules highly correlate with the evolutionary branches supposedly dividing their sequences. Such quantitative descriptions, which are based upon a substitution matrix and sequence weights and information entropy, could be effective for objective evaluation of amino acid residues. We expect that such scoring methodology for multiple alignment sites prompts identification of key residues involving functional divergence of protein subfamilies or subtypes.

References

- Andersen GR, Nissen P, Nyborg J (2003) Elongation factors in protein biosynthesis. *Trends Biochem Sci* 28: 434-441.
- Andersen GR, Pedersen L, Valente L, Chatterjee I, Kinzy TG, et al. (2000) Structural basis for nucleotide exchange and competition with tRNA in the yeast elongation factor complex eEF1A:eEF1B α . *Mol Cell* 6: 1261-1266.
- Lund A, Knudsen SM, Vissing H, Clark B, Tommerup N (1996) Assignment of human elongation factor 1 α genes: EEF1A maps to chromosome 6q14 and EEF1A2 to 20q13.3. *Genomics* 36: 359-361.
- Kahns S, Lund A, Kristensen P, Knudsen CR, Clark BF, et al. (1998) The elongation factor 1 A-2 isoform from rabbit: cloning of the cDNA and characterization of the protein. *Nucleic Acids Res* 26: 1884-1890.
- Mansilla F, Friis I, Jadidi M, Nielsen KM, Clark BF, et al. (2002) Mapping the human translation elongation factor eEF1H complex using the yeast two-hybrid system. *Biochem J* 365: 669-676.
- Lee S, Francoeur AM, Liu S, Wang E (1992) Tissue-specific expression in mammalian brain, heart, and muscle of S1, a member of the elongation factor-1 α gene family. *J Biol Chem* 267: 24064-24068.
- Knudsen SM, Frydenberg J, Clark BF, Leffers H (1993) Tissue-dependent variation in the expression of elongation factor-1 α isoforms: isolation and characterisation of a cDNA encoding a novel variant of human elongation-factor 1 α . *Eur J Biochem* 215: 549-554.
- Chambers DM, Peters J, Abbott CM (1998) The lethal mutation of the mouse wasted (wst) is a deletion that abolishes expression of a tissue-specific isoform of translation elongation factor 1 α , encoded by the Eef1a2 gene. *Proc Natl Acad Sci U S A* 95: 4463-4468.
- Newbery HJ, Gillingwater TH, Dharmasaroja P, Peters J, Wharton SB, et al. (2005) Progressive loss of motor neuron function in wasted mice: effects of a spontaneous null mutation in the gene for the eEF1 A2 translation factor. *J Neuropathol Exp Neurol* 64: 295-303.
- Ejiri S (2002) Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization. *Biosci Biotechnol Biochem* 66: 1-21.
- Mateyak MK, Kinzy TG (2010) eEF1A: thinking outside the ribosome. *J Biol Chem* 285: 21209-21213.
- Yang F, Demma M, Warren V, Dharmawardhane S, Condeelis J (1990) Identification of an actin-binding protein from Dictyostelium as elongation factor 1 α . *Nature* 347: 494-496.
- Gross SR, Kinzy TG (2007) Improper organization of the actin cytoskeleton affects protein synthesis at initiation. *Mol Cell Biol* 27: 1974-1989.
- Nagashima K, Kasai M, Nagata S, Kaziro Y (1986) Structure of the two genes coding for polypeptide chain elongation factor 1 α (EF-1 α) from *Saccharomyces cerevisiae*. *Gene* 45: 265-273.
- Munshi R, Kandl KA, Carr-Schmid A, Whitacre JL, Adams AE, et al. (2001) Overexpression of translation elongation factor 1A affects the organization and function of the actin cytoskeleton in yeast. *Genetics* 157: 1425-1436.
- Amiri A, Noei F, Jeganathan S, Kulkarni G, Pinke DE, et al. (2007) eEF1A2 activates Akt and stimulates Akt-dependent actin remodeling, invasion and migration. *Oncogene* 26: 3027-3040.
- Thornton S, Anand N, Purcell D, Lee J (2003) Not just for housekeeping: protein initiation and elongation factors in cell growth and tumorigenesis. *J Mol Med (Berl)* 81: 536-548.
- Lee JM (2003) The role of protein elongation factor eEF1A2 in ovarian cancer. *Reprod Biol Endocrinol* 1: 69.
- Ruest LB, Marcotte R, Wang E (2002) Peptide elongation factor eEF1A-2/S1 expression in cultured differentiated myotubes and its protective effect against caspase-3-mediated apoptosis. *J Biol Chem* 277: 5418-5425.
- Herrmann R, Reiner B (1998) *Mycoplasma pneumoniae* and *Mycoplasma genitalium*: a comparison of two closely related bacterial species. *Curr Opin Microbiol* 1: 572-579.
- Waites KB, Talkington DF (2004) *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin Microbiol Rev* 17: 697-728, table of contents.
- Goulet M, Dular R, Tully JG, Billowes G, Kasatiya S (1995) Isolation of *Mycoplasma pneumoniae* from the human urogenital tract. *J Clin Microbiol* 33: 2823-2825.
- Jensen JS (2004) *Mycoplasma genitalium*: the aetiological agent of urethritis and other sexually transmitted diseases. *J Eur Acad Dermatol Venereol* 18: 1-11.
- Baseman JB, Dallo SF, Tully JG, Rose DL (1988) Isolation and characterization of *Mycoplasma genitalium* strains from the human respiratory tract. *J Clin Microbiol* 26: 2266-2269.
- Balasubramanian S, Kannan TR, Hart PJ, Baseman JB (2009) Amino acid changes in elongation factor Tu of *Mycoplasma pneumoniae* and *Mycoplasma genitalium* influence fibronectin binding. *Infect Immun* 77: 3533-3541.
- Valdar WS (2002) Scoring residue conservation. *Proteins* 48: 227-241.
- Lichtarge O, Bourne HR, Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 257: 342-358.
- Landgraf R, Fischer D, Eisenberg D (1999) Analysis of heregulin symmetry by weighted evolutionary tracing. *Protein Eng* 12: 943-951.
- Mihalek I, Res I, Lichtarge O (2004) A family of evolution-entropy hybrid methods for ranking protein residues by importance. *J Mol Biol* 336: 1265-1282.
- Apweiler R, Martin MJ, O'Donovan C, Magrane M, Alam-Faruque Y, et al. (2013) Update on activities at the universal protein resource (UniProt) in 2013. *Nucleic Acids Res* 41: D43-D47.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30: 772-780.
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6.

- Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
33. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8: 275-282.
 34. Accelrys Inc (2011) Discovery Studio, Version 3.1. San Diego, CA.
 35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
 36. Sibbald P, Argos P (1990) Weighting aligned protein or nucleic acid sequences to correct for unequal representation. *J Mol Biol* 216: 813-818.
 37. Benner SA, Cohen MA, Gonnet GH (1994) Amino acid substitution during functionally constrained divergent evolution of protein sequences. *Protein Eng* 7: 1323-1332.
 38. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12: 77.
 39. Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing Sci Eng* 9: 90-95.
 40. Soares DC, Barlow PN, Newbery HJ, Porteous DJ, Abbott CM (2009) Structural models of human eEF1A1 and eEF1A2 reveal two distinct surface clusters of sequence variation and potential differences in phosphorylation. *PLoS One* 4: e6315.