18. Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
19. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
20. Hedges,S.B. and Kumar,S. (2002) Genomics. Vertebrate genomes compared. *Science*, **297**, 1283–1285.
21. Huttley,G.A., Wakefield,M.J. and Easteal,S. (2007) Rates of genome evolution and branching order from whole genome analysis. *Mol. Biol. Evol.*, **24**, 1722–1730.
22. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
23. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Buillard,V., Cerutti,L., Copley,R. *et al.* (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.
24. Endo,T., Ogishima,S. and Tanaka,H. (2002) ETools: Tools to Handle Biological Sequences and Alignments for Evolutionary Studies. *Genome Inform.*, **13**, 543–544.
25. Ota,S. and Li,W.H. (2001) NJML+: an extension of the NJML method to handle protein sequence data and computer software implementation. *Mol. Biol. Evol.*, **18**, 1983–1992.
26. Nei,M. and Gojobori,T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, **3**, 418–426.
27. Zhang,J., Rosenberg,H.F. and Nei,M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. USA*, **95**, 3708–3713.
28. Takeda,J., Suzuki,Y., Nakao,M., Barrero,R.A., Koyanagi,K.O., Jin,L., Motono,C., Hata,H., Isogai,T., Nagai,K. *et al.* (2006) Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res.*, **34**, 3917–3928.
29. Gu,Z., Cavalcanti,A., Chen,F.C., Bouman,P. and Li,W.H. (2002) Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol. Biol. Evol.*, **19**, 256–262.

Volff J-N (ed): Gene and Protein Evolution.
Genome Dyn. Basel, Karger, 2007, vol 3, pp 13–29

Evolution of Protein-Protein Interaction Network

T. Makino[a,b,c], *T. Gojobori*[a,d]

[a]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Yata, Mishima, [b]Immunotherapy Division, Shizuoka Cancer Center Research Institute, Shimonagakubo, Nagaizumi-cho, Shizuoka, Japan; [c]Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin, Ireland; [d]Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Aomi, Koto-ku, Tokyo, Japan

Abstract

Protein-protein interactions (PPIs) are one of the most important components of biological networks. It is important to understand the evolutionary process of PPIs in order to elucidate how the evolution of biological networks has contributed to diversification of the existent organisms. We focused on the evolutionary rates of proteins involved with PPIs, because it had been shown that for a given protein-coding gene the number of its PPIs in a biological network was one of the important factors in determining the evolutionary rate of the gene. We studied the evolutionary rates of duplicated gene products that were involved with PPIs, reviewing the current situation of this subject. In addition, we focused on how the evolutionary rates of proteins were influenced by the characteristic features of PPIs. We, then, concluded that the evolutionary rates of the proteins in the PPI networks were strongly influenced by their PPI partners. Finally, we emphasized that evolutionary considerations of the PPI proteins were very important for understanding the building up of the current PPI networks.

Protein-Protein Interaction Network as a Typical Example of Biological Networks

Interactions between proteins and various molecules including proteins themselves are absolutely necessary for sustaining life as a whole. For example, cells are controlled by interacting proteins in metabolic and signaling pathways, such as the molecular machines that replicate, translate and transcribe genes, and build up cell structures. We can classify the biological networks consisting

of such various interactions basically into five types according to the molecules interacting with proteins.

(i) Protein-chemical compound interaction: In the metabolic network, some proteins interact with low-molecular chemical compounds. For example, galactose is metabolized through a series of steps involving the enzymes that are encoded by *GAL1, GAL5, GAL7*, and *GAL10* [1]. These enzymes interact with the appropriate metabolic products.

(ii) Protein-DNA interaction: In the regulatory network, transcriptional factors interact with DNA segments such as the promoter region for transcriptional regulation. For example, genes involved in the galactose metabolism are regulated by the transcriptional factors encoded by *GAL3, GAL4*, and *GAL80*. They interact with the appropriate upstream regions of open reading frames in the DNA.

(iii) Protein-RNA interaction: For the interactions between proteins and nucleotides, proteins interact with not only DNA but also RNA. For example, proteins in the ribosomes in the translation machinery interact with messenger RNAs.

(iv) Protein-lipid interaction: There are proteins interacting with lipids such as phosphoinositides. The phosphoinositides serve as the second messengers that regulate diverse cellular processes [2, 3]. For example, steroid hormone receptors that are transcriptional factors interact with steroid hormones for the transcriptional regulation of target genes [4].

(v) Protein-protein interaction: Finally, protein-protein interactions (PPIs) are well-studied components of biological networks. PPIs are involved in a number of biological processes such as protein transportation and degradation, cell cycle progression, polarity, gene expression and DNA repair. For example, the transcriptional factor encoded by *GAL80* as already mentioned above interacts with the other transcriptional factors encoded by *GAL3* and *GAL4* for the regulation of galactose utilization.

Recently, global studies on PPIs have been investigated not only in prokaryotes, which are *Helicobacter pylori* [5] and *Escherichia coli* [6], but also in eukaryotes, which are *Plasmodium falciparum* [7], *Caenorhabditis elegans* [8], *Drosophila melanogaster* [9, 10] and human [11, 12]. In particular, *Saccharomyces cerevisiae* provides a great advantage for the study of PPIs, because a vast amount of information about PPIs has been produced not only by hundreds of small-scale experiments but also by the high-throughput yeast two-hybrid system (Y2H system; [13, 14]) and mass spectrometry of coimmunoprecipitated protein complexes (Co-IP; [15–17]). However, the high-throughput data on PPIs are known to contain a number of false-negative and false-positive interactions. In the case of the false-negative interactions, the PPIs sometimes could not be detected in the Y2H system for full-length ORFs. This is because

full-length proteins often show much weaker signals than appropriately trimmed protein regions containing interacting proteins. On the other hand, proteins having low expression levels will not be able to be identified by the Co-IP, because of the limitation of the sensitivity for the system. Therefore, the detection of the PPIs should be conducted by the both methods that are mutually complementary. Proteins such as transcriptional factors activate the expression of a reporter gene in the Y2H system and lead to false-positive interactions. Contaminant proteins with high expression levels tend to be recovered in coimmunoprecipitated protein complexes (Co-IP), even if they do not actually interact with one another. Consequently, the high-throughput data require further examination for their accuracy. Several methods for removing dubious PPIs from the original data were developed, and as a result, the credible PPIs have become enriched [18–20].

Evolutionary Studies of Protein-Protein Interaction Networks

Until now, molecular evolutionary analyses have mainly focused on individual genes regardless of how they are involved with the interactions among their gene products. However, it is interesting to carry out evolutionary analyses of a group of genes in which the encoding proteins interact with one another in the PPI network. In these analyses, it is important to examine how selective pressures affect gene products as the components of PPI networks. It is of particular interest to study how the organization of proteins as members of PPI network affects the evolutionary rates of their corresponding genes.

It should be noted that duplicated genes encoding proteins in PPI networks provide us with a unique opportunity of making fair comparisons of the genes under the same initial condition. The pair of proteins encoded by a duplicated gene pair often share PPI partners [20, 21], although some of the PPI partners may be lost later in the evolutionary process. In fact, there are a lot of duplicated pairs encoding proteins not having the shared PPI partners [21]. Therefore, we examined the relationship in the evolutionary rates between a duplicated protein in PPIs and its counterpart ('*Differential evolutionary rates of duplicated genes in protein interaction network*' in this chapter; [22]).

It has been shown that proteins sharing functions tend to interact in the PPI networks [23, 24]. There is a strong correlation between the structure of the PPI network and the functions of the proteins in the network [25]. In other words, many functions appear to be particular parts in the PPI networks. On the other hand, the recent study gave us an interesting insight [26]. The authors have shown that there are many proteins interacting with their PPI partners having different functions. For example, mitogen-activated protein kinase (MAPK)

interacts with proteins having different functions that are involved in ribosomal biogenesis, cytoskeleton and directional cell growth. In particular, it has been shown that such PPIs have biological importance according to an experiment of double gene deletion for genes encoding the protein and its PPI partner. PPIs are not in a uniform state as mentioned above. It is of great interest to study how the interacting proteins have been evolutionarily influenced by their PPI partners in the PPI network. Therefore, we examined the differences in evolutionary rate among the interacting proteins involved in different PPIs (*'The evolutionary rate of a protein is influenced by features of the interacting partners'* in this chapter; [27]).

Differential Evolutionary Rates of Duplicated Genes in Protein Interaction Network

The functional constraints of proteins involved in the PPI network are composed of several factors. The so-called fitness effects as well as the gene expression level are typical factors, because they are known to be negatively correlated with the rate of amino acid substitutions [28–31]. The number of PPIs for a given protein is also an important factor for determining its evolutionary rate. It has been reported that the number of PPI partners for proteins is negatively correlated with their evolutionary rates [32, 33]. Therefore, after gene duplication, the differentiation of PPIs through the PPI losses and/or PPI gains during evolution may affect the evolutionary rates of duplicated pairs. For a duplicated gene pair, it has been shown that one copy usually has more PPI partners than the other [34].

Gene duplication is one of the major evolutionary mechanisms for generating novel genes [35]. After gene duplication, one of the pair may be redundant, such that functional constraint is relaxed to allow one or both to differentiate as long as the original function is retained as a whole. Three pathways have been proposed for functional differentiation of duplicated genes [36]. First, one copy may be silenced by accumulation of deleterious mutations and eventually become indistinguishable from the nearby noncoding genomic regions in the absence of functional constraints, while the other copy retains the original function. Second, while one copy maintains the original function, the other acquires a novel function possibly by advantageous mutations. Third, both copies accumulate mutations that alter the original function, but compensate for the original function cooperatively. When a duplicated gene pair functionally differentiates, the evolutionary rate may be accelerated in one or both due to the relaxation of negative selection or the enhancement of positive selection [37]. In yeast, it has been proposed that the differentiation process is asymmetrical rather than symmetrical to minimize the risk of deleterious mutations [34]. It is

therefore expected that the acceleration of evolutionary rates occurs mainly in one of two copies after gene duplication. However, it is not yet known how the duplicated gene products affect their PPIs in evolution.

Duplicated products often interact with the same proteins [20]. One proposed model for the losses and/or gains of PPIs provides the reason why the products of a duplicated gene pair often share PPI partners [21]. In this model, although some duplicated pairs lose PPIs during the evolutionary process, many duplicated pairs retain some shared PPI partners. In a recent study, the magnitude of functional divergence for duplicated pairs was measured by using the number of shared PPI partners between all pairs in the PPI networks [38]. To examine the relationship between the evolutionary rate and the functional differentiation of duplicated gene products, we focused on the shared PPI partners that were considered to represent characteristics of the functional differentiation of the duplicated gene products, because the products sharing PPI partners would not have largely diverged.

The purpose of the study is to understand how gene duplication influences the evolution of PPI networks. To study the relationship between gene duplication and the evolutionary rates of the gene products with PPI partners, we used the PPIs in *Saccharomyces cerevisiae* that have well been documented based not only on hundreds of small-scale experiments but also on high throughout methods. We set up and examined the hypothesis that the ratios of evolutionary rates (faster rate/slower rate) for the pairs sharing any PPI partners are lower than those for the pairs sharing no PPI partners. We then discuss the mechanisms of the functional differentiation after gene duplication on the basis of the results obtained.

Losses of PPIs for Proteins Encoded by Duplicated Genes

Soon after gene duplication, the protein encoded by one copy should interact with the same set of proteins as the other, because both proteins are identical. It has been proposed that PPI partners of proteins encoded by duplicated genes change through PPI losses or PPI gains during evolution [21]. For a duplicated gene pair, it has been shown that one copy usually has more PPI partners than the other [34]. However, it was unclear which of the two mechanisms, namely PPI losses and PPI gains, is the major force of the evolution of PPIs. Proteins under strong functional constraints would be hard to change their PPI partners during evolution, because they are conservative. The PPI losses of the proteins may accelerate their evolutionary rates, because it has been reported that the evolutionary rate is negatively correlated with the number of PPIs [32, 33]. If the PPI losses occur more often than the PPI gains for a duplicated pair, the protein encoded by one copy evolving at a slower rate would have more PPI partners than the other.

To examine this possibility, we used duplicated pairs generated by genome duplication in *Saccharomyces cerevisiae*, which occurred about 100 million years ago [39, 40]. For each pair of gene products, we examined whether the protein with more PPI partners evolved more slowly than the other with less partners. We then found that a protein with more PPI partners evolved at a slower rate in 134 (62%) out of the 216 pairs examined, which was significantly greater than expected under the null hypothesis of random association between the number of PPI partners and the evolutionary rate (50%). We found that the protein encoded by one copy evolving at a slower rate had more PPI partners than the other copy. The results indicated that the PPI losses have occurred more often than the PPI gains for a copy evolving at a faster rate, on the assumption that PPIs of a copy evolving at a slower rate are conservative in the evolutionary process.

Functional Divergence through Changes in PPIs

After gene duplication, there are at least two possible pathways for PPI divergence of the proteins encoded by a duplicated gene pair. First, one encoded by a duplicated pair keeps the shared PPI partners, and the other loses all the shared PPI partners. The evolutionary rate of the former would be slower than that of the latter, because the former has to maintain the original function while the latter is free from it. In other words, they are likely to evolve at different rates. Second, both proteins share some of the PPI partners. In this case, both proteins will still have similar functions, and their sequences would not change by mutations as drastically as in the latter of the first case. The evolutionary rates of the gene products sharing PPI partners may not significantly differ from one another. If duplicated gene products lose the shared PPI partners, the ratio of evolutionary rates for the pair (faster rate/slower rate) may be higher than that for functionally similar pairs.

To test this hypothesis, we examined whether F_1/S_1 were higher than F_2/S_2, where F and S denote faster rate and slower rate, respectively, and subscripts 1 and 2 refer to the cases of sharing no PPIs and sharing PPIs, respectively (fig. 1). Here we defined duplicated pairs sharing PPIs as the pairs sharing at least one PPI partner. There were 124 duplicated pairs sharing no PPI partners and 130 duplicated pairs interacting with one another or sharing PPI partners. F_1/S_1 was significantly higher than F_2/S_2 (fig. 2).

For a duplicated gene pair, if the protein encoded by one copy evolving at a faster rate has not been silenced during evolution, it would have lost its PPI partners and have a chance of finding a new PPI partner under the weak or no functional constraints. On the other hand, the PPIs for the protein encoded by one copy evolving at a slower rate would be conservative with relatively strong functional constraints. For duplicated pairs, the gene product evolving at a

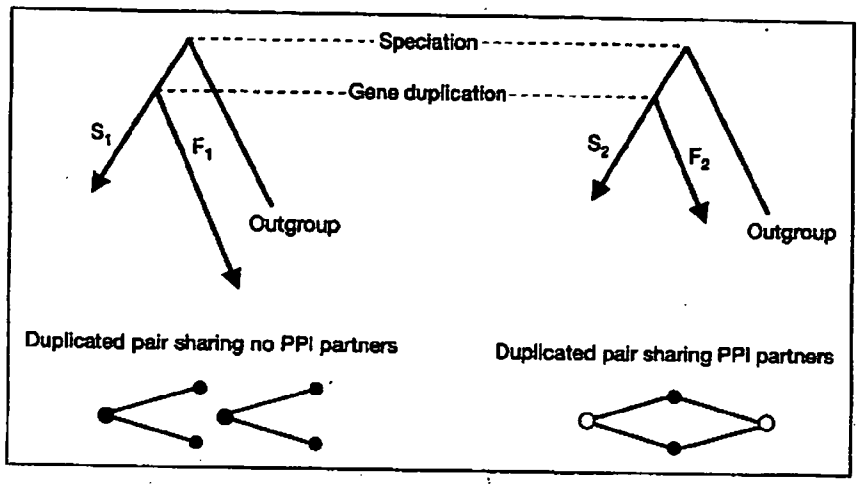

Fig. 1. Schematic representations of F_1, S_1, F_2, and S_2. Closed circles and open circles respectively mean proteins encoded by duplicated gene pair sharing no PPI and sharing PPIs. F (light gray arrow) and S (gray arrow) mean faster rate and slower rate, respectively, and subscripts 1 and 2 refer to the cases of sharing no PPI and sharing PPIs for duplicated pairs, respectively. The ratio of evolutionary rates for duplicated pairs after gene duplication was estimated by a faster evolutionary rate of one copy/a slower rate of another copy (F_1/S_1; F_2/S_2).

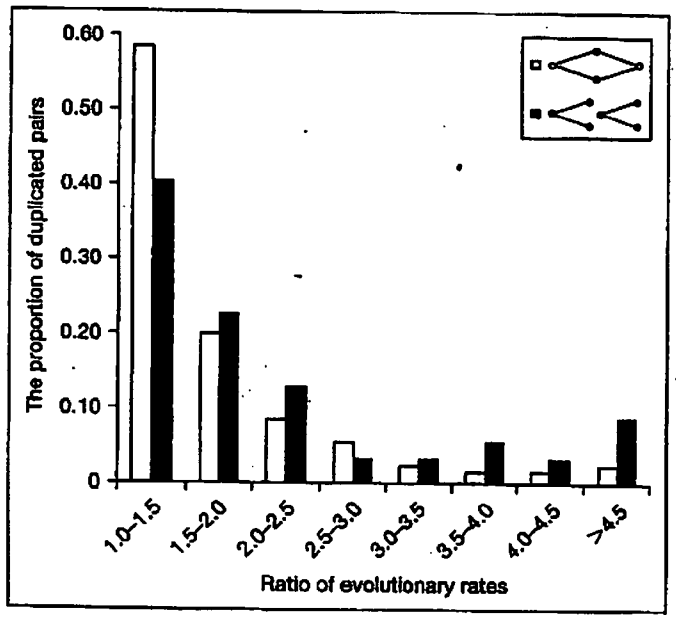

Fig. 2. Ratios of evolutionary rates for duplicated pairs sharing PPI partners and sharing no PPI partners. Open bars indicate duplicated pairs interacting with one another or sharing PPI partners, while closed bars indicate duplicated pairs sharing no PPI partners.

Table 1. Results of relative rate test for duplicated pairs having PPI partners and sharing no PPI partners in functional class 'transcription'

	Number of duplicated pairs	
	sharing PPI partners	not sharing PPI partners
Significant difference of rates	10	19
No significant difference of rates	13	7

faster rate will lose the shared PPI partners more frequently than the other. This implies that a pair of proteins encoded by a duplicated gene pair having few shared PPI partners evolves at different rates. In fact, the present study indicates that pairs sharing no PPI partners show a larger ratio of evolutionary rates than those sharing PPI partners, although it has been reported that a simple relationship between sequence divergence and their functional divergence revealed by the PPI network analysis could not be established [38]. When a duplicated gene pair shares no PPI partners, it is possible that the gene products interact with different PPI partners with different functions. This means that gene duplication will lead to the functional differentiation of the duplicated gene products through the PPI losses and/or PPI gains, which will then cause a change in their evolutionary rates.

Tendency of PPI Divergence for Duplicated Pair in Different Functional Classes

For investigating the functions of duplicated gene products, we used functional classification established by the MIPS database [41]. In the functional class of 'transcription', there were significantly many duplicated pairs sharing no PPI partners and having significant difference in evolutionary rates (table 1). There were also statistically significant differences in the rate between the two copies in the functional class of 'protein fate' (table 2). These results indicate that the PPIs of the proteins included in these functional classes tend not to be conservative in the evolutionary process, resulting in a change in their evolutionary rates. The other functional classes showed no significant difference in ratio of evolutionary rates between duplicated pairs sharing PPI partners and those sharing no PPI partners.

We found many cases of pairs sharing no PPI partners in the functional classes such as 'transcription' and 'protein fate'. For example, YNR023W and YCR052W (a duplicated pair in 'transcription') do not share PPI partners, and

Table 2. Results of relative rate test for duplicated pairs having PPI partners and sharing no PPI partners in functional class 'protein fate'

	Number of duplicated pairs	
	sharing PPI partners	not sharing PPI partners
Significant difference of rates	2	11
No significant difference of rates	10	5

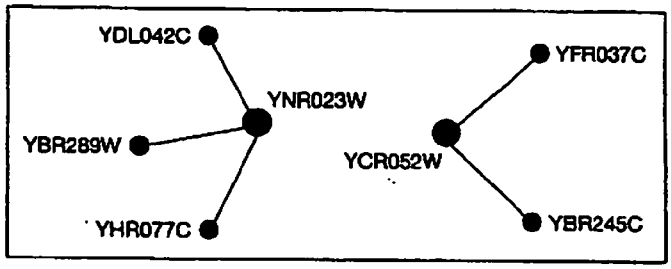

Fig. 3. An example for the pair of proteins encoded by duplicated gene pairs and their PPI partners. The circles and lines represent proteins and PPIs, respectively. The circles in gray are PPI partners. The closed circles represent proteins encoded by the duplicated gene pair (YNR023W and YCR052W), which are a subunit of SWI/SNF global transcription activator complex and a subunit of the RSC chromatin-remodeling complex, respectively.

have a significant difference in evolutionary rate between them. In addition, they are subunits in different protein complexes. YNR023W is a subunit of SWI/SNF global transcription activator complex, and YCR052W is a subunit of the RSC chromatin-remodeling complex (fig. 3; [42]). We consider the significant difference in evolutionary rate between the two copies is caused by drastic changes in the PPI partners during evolution. Although the proteins encoded by these duplicated gene pairs would have interacted with the same PPI partners immediately after the gene duplication, one of the copies would have subsequently changed its PPI partners and diverged its functions. It is thus suggested that YCR052W, which evolves at a faster rate than YNR023W, would have obtained novel functions by changing their PPI partners. Thus, the evolutionary comparison of the PPI partners of one copy in a duplicated pair with those of the other is important for understanding their functional differentiations through PPI network divergence.

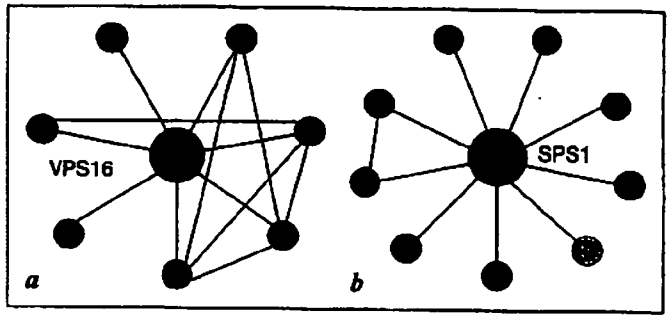

Fig. 4. *a* A protein in a functional module and (*b*) a protein in a framework module o the PPI network. The filled circles and lines represent proteins and PPIs, respectively. Th black lines indicate interactions between VPS16 and its PPI partners and between SPS1 an its PPI partners. The gray lines indicate interactions among PPI partners. VPS16 interact with proteins classified into the same functional class 'protein fate' on Munich Informatio Center for Protein Sequences database [41]. The different grey scales of the circles in *b* mea different functional classes. SPS1 interacts with proteins classified into different functiona classes 'protein fate', 'cell cycle/DNA processing', 'metabolism', 'cellular transport', an 'transcription', respectively.

The Evolutionary Rate of a Protein is Influenced by Features of the Interacting Partners

When a two-dimensional presentation of PPI networks is made using a node and a line between neighboring nodes as a protein and an interactio between neighboring proteins, respectively, the PPI network is represented by a very complex structure of spider web-like networks. It has been reported, in thi type of representation, that there are proteins tightly clustered in a particula part of the PPI network [43]. In particular, the proteins sharing a particula functional class tend to appear in the same part of a PPI network, making a clus ter of the so-called 'functional module' in the PPI network [25]. Here, a func tional class represents a category into which a group of particular proteins i classified according to the functional definitions. In other words, a functiona module of the network is generally defined as a cluster of proteins sharing th same functional class that occupies a specific part of the network. In the PP networks, the proteins building up a functional module have more interaction to other proteins within the functional module than to those outside the module For example, VPS16 of *Saccharomyces cerevisiae* is clustered in a functiona module that is required for sorting proteins in vacuolare (fig. 4a).

On the other hand, there are proteins known to interact with those havin different functional classes [26]. Calmodulin, which is a master regulator o

calcium-mediated signaling [44], interacts with several proteins of different functional classes such as homeostasis of cations, protein folding and stabilization, budding, cell polarity and filament formation [26]. For these proteins, the gene expression patterns do not correlate with those of their PPI partner proteins, suggesting that they interact with the PPI partners at different subcellular localizations or different time points. Let us call these the proteins in a framework module. In other words, the protein in a framework module is defined as a protein mediating different functions by interactions of proteins having different functional classes. For example, SPS1 encoding ser/thr protein kinase of *S. cerevisiae* is in a framework module, and interacts with proteins classified into different functional classes (fig. 4b). Therefore, the number of interactions among the PPI partners of these proteins in the framework module is expected to be smaller than that of the proteins in the functional module.

It is interesting to investigate the extent to which the evolutionary rate of proteins is influenced by the nature of PPIs. Therefore, we examined the differences in evolutionary rate among the proteins having different types of PPI partners. The difference in the evolutionary rate can be interpreted by the difference in functional constraints if the mutation rate does not vary much with the proteins. Thus, we would also discuss the differences in functional constraint among the proteins having different types of PPI partners in the PPI network.

SF vs. DF Proteins

Proteins in the PPI networks would have evolved under the influence of their PPI partners. It has been reported that the number of PPI partners is correlated significantly to their evolutionary rates [32, 33]. A recent study reported that proteins in the center of the PPI networks evolve more slowly, regardless of the number of PPI partners [45]. When the proteins lose or gain their PPI partners during evolution, an allowable degree of their amino acid substitutions may depend not only on the number of their PPI partners but also on the features of their PPI partners. It has been known that proteins sharing the same functional class tend to interact with each other [23, 24]. On the contrary, there are proteins that interact with those belonging to different functional classes [26]. Here, we defined a protein having PPI partners of the same functional class with a high frequency as an SF (the Same Function) protein, on the other hand, a protein having PPI partners of different functional classes with a high frequency as a DF (the Different Function) protein. It is of particular interest to know which of the SF or DF proteins is under stronger functional constraints in the evolutionary process. Therefore, we examined whether the evolutionary rates of the proteins in the PPI network have been strongly influenced by the PPI partners having the same or different functional classes. To answer the question, we compared the evolutionary rates of the SF proteins with those of

the DF proteins in yeast PPI networks. For this comparative study, we used the evolutionary distances for 1,035 SF and 763 DF proteins for the comparison. As a result, we found that the DF proteins evolved at a slower rate, with statistical significance, than the SF proteins. Thus, we concluded that the DF proteins are under much stronger functional constraints than the SF proteins.

DP vs. SP Proteins

It has been reported that there are proteins tightly clustered in a particular part of the PPI network [43]. Denoting proteins in dense and sparse parts of the PPI network as the DP (Dense Part) and SP (Sparse Part) proteins, respectively, we defined them using the clustering coefficient [46]. We examined the differences in evolutionary rates between DP proteins in a dense part of PPI networks and SP proteins in a sparse part of PPI networks. When we compared the evolutionary rates of the 668 DP proteins with those of the 965 SP proteins, we found that the SP proteins evolved at a slower rate, with statistical significance, than the DP proteins. Interestingly enough, this is also opposite to our expectation. Before conducting the present study, we speculated that the DP proteins would have slower rates, because it has been reported that proteins having cohesive patterns of PPIs are more evolutionarily conservative than other proteins in the PPI network [47]. In contrast, our observation suggests that the proteins in a sparse part of the PPI network could be more important than those in a dense part. It is possible that the PPI partners in a sparse part in the PPI network are indispensable because of possible scarceness of substitutable PPI partners. This is an interesting and meaningful finding.

Comparison of Evolutionary Rates among SF-DP, SF-SP, DF-DP and DF-SP Proteins

According to the results described above, we reasonably hypothesized that the DF-SP proteins would evolve at the slowest rate in the proteins examined. To test the hypothesis, we statistically compared the evolutionary rates among the 443 SF-DP, 353 SF-SP, 122 DF-DP and 457 DF-SP proteins. We found that out of all proteins examined the DF-SP proteins evolved certainly at the slowest relative evolutionary rate (fig. 5). The result suggests that the proteins having the PPI partners belonging to different functional classes and being in a sparse part of the PPI network are under the strongest functional constraints, implying that those proteins are possibly important for the maintenance and survival of the PPI network.

We have found that the DF proteins evolved at a slower rate than the SF proteins. The observation suggests that the proteins involved with multi-different biological processes in the PPI network are under strong functional constraints. We have also shown that the SP proteins evolved at a slower rate than the DP

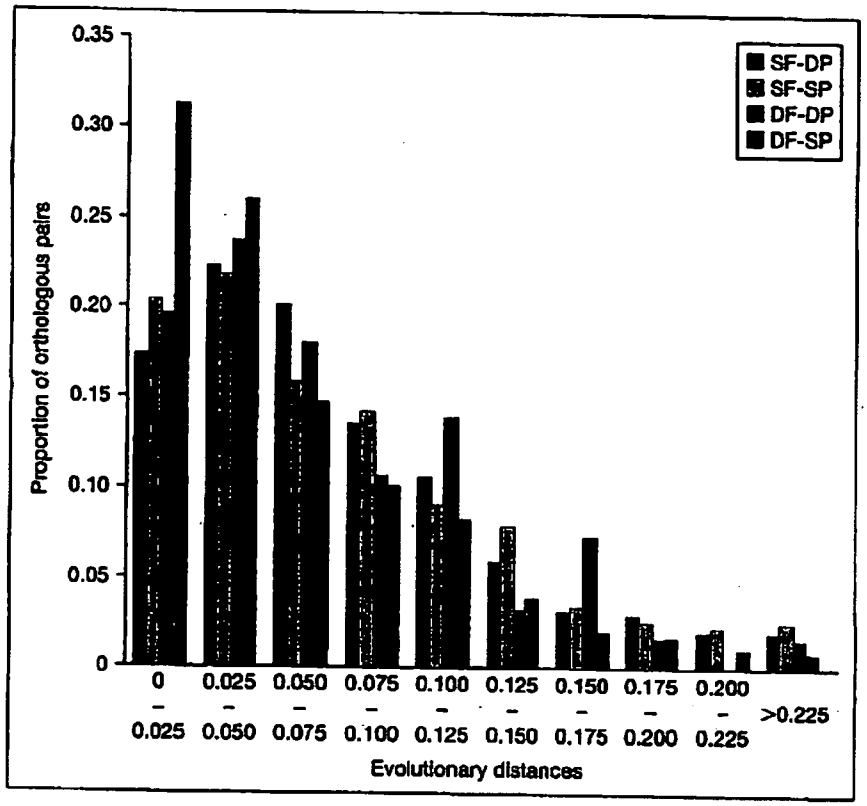

Fig. 5. Distribution of evolutionary distances for the SF-DP, SF-SP, DF-DP, and DF-SP proteins. The evolutionary distance is measured as the number of amino acid substitutions per site.

proteins. In fact, we have shown that the DF-SP proteins evolved at the slowest rate among all interacting proteins studied. This might be explained if loss of function in DF-SP proteins affected multiple biological processes more so than that of proteins with other interaction properties. These results strongly suggest that the evolutionary rates of proteins depend on the nature of interacting proteins in the PPI network.

For the evolutionary studies of proteins in the PPI networks, it has been shown that proteins involved in protein complexes are more evolutionarily conservative than other proteins in the PPI networks [48]. A protein complex can be considered as a typical example of SF proteins, because all the subunits are regarded as belonging to the same functional class due to a particular functional manifestation of the whole protein complex. To confirm this consideration, we compared a proportion of subunits in protein complexes for the SF proteins with that for DF proteins using the protein complex data set in the MIPS database [41]. As expected, we found that the SF proteins contained more subunits

of protein complexes than the DF proteins (data not shown). Although the SF proteins contained relatively many subunits of a protein complex, our results clearly showed that the SF proteins are evolutionarily much less conservative than the DF proteins. Moreover, it has been reported that proteins having cohesive patterns of PPIs are more evolutionarily conservative than other proteins in the PPI network, and tend to be subunits of protein complexes [47]. The proteins would be under strong structural constraints, because many of the proteins are in an extremely dense part of the PPI network. Although the authors particularly showed high evolutionary conservation of the proteins having cohesive patterns of PPIs, our finding is that the DF-SP proteins are under the strongest functional constraints among all interacting proteins studied. This conclusion highlights the importance of studying the evolution of the DF-SP proteins for understanding essential features of PPI network evolution.

Prospect of Studies in PPI Network Evolution

We focused on two themes to study the evolution of protein-protein interaction networks as a typical example of biological networks.

First, we focused on a relationship between the PPI divergences of duplicated gene products and their evolutionary rates, and examined whether the difference in evolutionary rate exists between a duplicated pair of genes encoding proteins involved in PPIs. Our results showed the evolutionary rate of a protein having more PPI partners is much slower than that of the other having fewer PPI partners. Moreover, we found that the ratios for duplicated pairs sharing PPI partners are significantly lower than the ratios for pairs sharing no PPI partners. When a duplicated pair shares no PPI partners, it is possible that the gene products interact with the PPI partners having different functions. These results clearly indicate that gene duplication leads to the functional differentiation of the duplicated gene pairs through PPI losses and/or PPI gains. The functional differentiation would cause eventually the change in their evolutionary rates. The evolutionary comparison of the PPI partners of one copy in a duplicated pair with those of the other copy gives an important clue for understanding their functional differentiations through PPI network divergence.

Second, we focused on the differences in evolutionary rates among interacting proteins having different types of PPI partners, because it is of particular interest to know how the PPIs influence the evolutionary rate, namely the rate of amino acid substitutions. In fact, we showed that the DF proteins, which interact with PPI partners in different functional classes with a high frequency, evolve at a slower rate than the SF proteins do, which interact with PPI partners in the same functional class with a high frequency. It suggests that the

interacting proteins involved in multi-different biological processes would be under strong functional constraints. We also showed that SP proteins, which are in sparse parts of the PPI networks, evolve at a slower rate than the DP proteins, which are in dense parts of the networks. The result indicates that the weaker relationship among PPI partners of proteins is, the more slowly the interacting proteins evolve. These results strongly suggested that the evolutionary features of the interacting proteins have been influenced by the type of their PPIs such as functional and framework modules.

We clearly pointed out the advantage of utilizing a vast amount of information about PPIs in the molecular evolutionary studies of biological networks. In particular, we successfully showed that the evolution of proteins as the components of PPI networks can be understood, to a reasonably great extent, through the evolutionary rates. Finally, we would like to emphasize that this line of studies will give us an important insight into the understanding of evolutionary processes of the PPI networks.

Acknowledgements

This project is, in part, supported by the Genome Network Project of MEXT (Ministry of Education, Culture, Sports, Science and Technology) and BIRC (Biological Information Research Center) at AIST (National Institute of Advanced Industrial Science and Technology).

References

1 Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, et al: Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001;292:929–934.
2 Odorizzi G, Babst M, Emr SD: Phosphoinositide signaling and the regulation of membrane trafficking in yeast. Trends Biochem Sci 2000;25:229–235.
3 Wera S, Bergsma JC, Thevelein JM: Phosphoinositides in yeast: genetically tractable signalling. FEMS Yeast Res 2001;1:9–13.
4 Tsai MJ, O'Malley BW: Molecular mechanisms of action of steroid/thyroid receptor superfamily members. Annu Rev Biochem 1994;63:451–486.
5 Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, et al: The protein-protein interaction map of *Helicobacter pylori*. Nature 2001;409:211–215.
6 Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, et al: Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 2005;433:531–537.
7 LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, et al: A protein interaction network of the malaria parasite *Plasmodium falciparum*. Nature 2005;438:103–107.
8 Li S, Armstrong CM, Bertin N, Ge H, Milstein S, et al: A map of the interactome network of the metazoan *C. elegans*. Science 2004;303:540–543.
9 Formstecher E, Aresta S, Collura V, Hamburger A, Meil A, et al: Protein interaction mapping: a *Drosophila* case study. Genome Res 2005;15:376–384.
10 Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al: A protein interaction map of *Drosophila melanogaster*. Science 2003;302:1727–1736.

11 Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al: Towards a proteome-scale map of the human protein-protein interaction network. Nature 2005;437:1173–1178.
12 Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al: A human protein-protein interaction network: a resource for annotating the proteome. Cell 2005;122:957–968.
13 Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, et al: Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci USA 2000;97:1143–1147.
14 Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al: A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 2000;403:623–627.
15 Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415:141–147.
16 Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et al: Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 2002;415:180–183.
17 Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al: Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 2006;440:637–643
18 Bader GD, Hogue CW: Analyzing yeast protein-protein interaction data obtained from different sources. Nat Biotechnol 2002;20:991–997.
19 Bader JS, Chaudhuri A, Rothberg JM, Chant J: Gaining confidence in high-throughput protein interaction networks. Nat Biotechnol 2004;22:78–85.
20 Deane CM, Salwinski L, Xenarios I, Eisenberg D: Protein interactions: two methods for assessment of the reliability of high throughput observations. Mol Cell Proteomics 2002;1:349–356.
21 Wagner A: The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Mol Biol Evol 2001;18:1283–1292.
22 Makino T, Suzuki Y, Gojobori T: Differential evolutionary rates of duplicated genes in protein interaction network. Gene 2006;385:57–63.
23 Ge H, Liu Z, Church GM, Vidal M: Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat Genet 2001;29:482–486.
24 Schwikowski B, Uetz P, Fields S: A network of protein-protein interactions in yeast. Nat Biotechnol 2000;18:1257–1261.
25 Yook SH, Oltvai ZN, Barabasi AL: Functional and topological characterization of protein interaction networks. Proteomics 2004;4:928–942.
26 Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al: Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 2004;430:88–93.
27 Makino T, Gojobori T: The evolutionary rate of a protein is influenced by features of the interacting partners. Mol Biol Evol 2006;23:784–789.
28 Hirsh AE, Fraser HB: Protein dispensability and rate of evolution. Nature 2001;411:1046–1049.
29 Jordan IK, Rogozin IB, Wolf YI, Koonin EV: Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res 2002;12:962–968.
30 Pal C, Papp B, Hurst LD: Highly expressed genes in yeast evolve slowly. Genetics 2001;158:927–931.
31 Wilson AC, Carlson SS, White TJ: Biochemical evolution. Annu Rev Biochem 1977;46:573–639.
32 Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW: Evolutionary rate in the protein interaction network. Science 2002;296:750–752.
33 Fraser HB, Wall DP, Hirsh AE: A simple dependence between protein evolution rate and the number of protein-protein interactions. BMC Evol Biol 2003;3:11.
34 Wagner A: Asymmetric functional divergence of duplicate genes in yeast. Mol Biol Evol 2002;19:1760–1768.
35 Ohno S: Evolution by Gene Duplication. Springer, Berlin, 1970.
36 Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: Preservation of duplicate genes by complementary, degenerative mutations. Genetics 1999;151:1531–1545.
37 Li WH, Gojobori T: Rapid evolution of goat and sheep globin genes following gene duplication. Mol Biol Evol 1983;1:94–108.
38 Baudot A, Jacq B, Brun C: A scale of functional divergence for yeast duplicated genes revealed from analysis of the protein-protein interaction network. Genome Biol 2004;5:R76.

39 Kellis M, Birren BW, Lander ES: Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 2004;428:617–624.
40 Wolfe KH, Shields DC: Molecular evidence for an ancient duplication of the entire yeast genome. Nature 1997;387:708–713.
41 Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, et al: MIPS: a database for genomes and protein sequences. Nucleic Acids Res 2002;30:31–34.
42 Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, et al: RSC, an essential, abundant chromatin-remodeling complex. Cell 1996;87:1249–1260.
43 Spirin V, Mirny LA: Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci USA 2003;100:12123–12128.
44 Davis TN, Urdea MS, Masiarz FR, Thorner J: Isolation of the yeast calmodulin gene: calmodulin is an essential protein. Cell 1986;47:423–431.
45 Hahn MW, Kern AD: Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol 2005;22:803–806.
46 Watts DJ, Strogatz SH: Collective dynamics of 'small-world' networks. Nature 1998;393:440–442.
47 Wuchty S, Oltvai ZN, Barabasi AL: Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat Genet 2003;35:176–179.
48 Teichmann SA: The constraints protein-protein interactions place on sequence divergence. J Mol Biol 2002;324:399–407.

Takashi Gojobori
Center for Information Biology and DNA Data Bank of Japan
National Institute of Genetics
1111 Yata, Mishima-shi, Shizuoka-ken 411-8540, Japan
Tel. +81-55-981-6847, Fax +81-55-981-6848, E-mail tgojobor@genes.nig.ac.jp

Genetic Stability of Human T Lymphotropic Virus Type I despite Antiviral Pressures by CTLs[1]

Ryuji Kubota,[2]* Kousuke Hanada,[§] Yoshitaka Furukawa,[†] Kimiyoshi Arimura,[‡] Mitsuhiro Osame,[‡] Takashi Gojobori,[§] and Shuji Izumo*

Human T lymphotropic virus type I (HTLV-I)-associated myelopathy/tropical spastic paraparesis (HAM/TSP) is an inflammatory neurological disease. Patients with HAM/TSP show high proviral load despite increased HTLV-I Tax-specific CTL. It is still unknown whether the CTL efficiently eliminate the virus in vivo and/or whether a naturally occurring variant virus becomes predominant by escaping from the CTL. To address these issues, we sequenced a large number of HTLV-I *tax* genes from HLA-A*02 HAM/TSP patients and estimated synonymous and nonsynonymous changes of the genes to detect positive selection pressure on the virus. We found the pressures in three of six CTL epitopes in HTLV-I Tax, where amino acid substitutions preferentially occurred. Although some of variant viruses were not recognized by the CTL, no variant viruses accumulated within 3–8 years, indicating genetic stability of HTLV-I *tax* gene. These results suggest that CTL eliminate the infected cells in vivo and naturally occurring variant viruses do not predominate. As Tax is a regulatory protein which controls viral replication, the amino acid substitutions in Tax may reduce viral fitness for replication. Viral fitness and host immune response may contribute to the viral evolution within the infected individuals. Furthermore, the genetic stability in the epitopes despite the antiviral pressures suggests that the three epitopes can be the candidate targets for HTLV-I vaccine development. ***The Journal of Immunology,*** **2007, 178: 5966–5972.**

Human T lymphotropic virus type I (HTLV-I)[3] is a retrovirus, which causes two different human diseases in some infected individuals: HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) and adult T cell leukemia (1, 2). Adult T cell leukemia is severe leukemia, for which an effective treatment has not yet been established. HAM/TSP is an inflammatory disease in the spinal cord, where $CD4^+$ and $CD8^+$ T cells infiltrate to the perivascular area (3). The patients show spastic gait and sphincter dysfunction with mild sensory dysfunction (4). They have increased proviral load as compared with HTLV-I carriers, which is a strong predictor for the development of HAM/TSP from the carrier state (5). Furthermore, an increase of proviral load is associated with disease progression (6). These suggest that reducing the proviral load prevents the development and progression of HAM/TSP. However, an effective treatment to reduce the virus has not yet been developed.

HAM/TSP patients have high frequency of circulating CTL specific for HTLV-I Tax and CTL efficiently kill Ag-expressed target cells in an in vitro assay (7–9). However, the fact that the proviral load is still high despite these vigorous CTL responses may raise the question of whether the CTL really eliminate the virus in vivo. Recently, it was proposed that the killing activity of Tax-specific CTL may be disturbed (10). It has been difficult to show that CTL kill virus-infected cells in vivo; however, calculation of synonymous (without amino acid substitution) and nonsynonymous (with amino acid substitution) changes of virus genes has been developed to show an immunological antiviral pressure in vivo (11). If the rate of nonsynonymous change is greater than that of synonymous change in a region of the virus, this will suggest that an in vivo positive selection pressure occurs on the region. In HTLV-I infection, it is shown that the ratio of nonsynonymous changes to synonymous changes in the *tax* gene is greater in HTLV-I carriers than in HAM/TSP patients (12). We had previously sequenced the HTLV-I *tax* gene that codes for an immunodominant and viral regulatory protein, Tax, in a large number of viruses from patients with HAM/TSP. In this study, using the sequence data, we estimated the numbers of nonsynonymous nucleotide substitutions per nonsynonymous sites (dn) and the numbers of synonymous nucleotide substitutions per synonymous sites (ds) in the genes coding the CTL epitopes as well as in the remaining regions and compared those.

Virus-specific $CD8^+$ CTL recognize viral peptide on the MHC and play a pivotal role in controlling viral infections (13). To develop a CTL vaccine in viral infections, it is fundamental to know whether virus-specific CTL exist, what the CTL epitopes are, and whether the virus escapes from the host immune system. In HIV infection, naturally occurring mutants escape from CTL and predominate in infected individuals (14–16). This is a strong obstacle to establish an effective CTL vaccine for HIV

*Center for Chronic Viral Diseases, [†]Division of Blood Transfusion Medicine, and [‡]Department of Neurology and Geriatrics, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan; and [§]Center for Information Biology, National Institutes of Genetics, Shizuoka, Japan

Received for publication September 26, 2006. Accepted for publication February 14, 2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by a Grant-in-Aid for Research on Brain Science of the Ministry of Health, Labor and Welfare of Japan, and a Grant-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

[2] Address correspondence and reprint requests to Dr. Ryuji Kubota, Center for Chronic Viral Diseases, Graduate School of Medical and Dental Sciences, Kagoshima University, 8-35-1 Sakuragaoka, Kagoshima 890-8544, Japan. E-mail address: kubotar@m2.kufm.kagoshima-u.ac.jp

[3] Abbreviations used in this paper: HTLV-I, human T lymphotropic virus type I; HAM/TSP, HTLV-I-associated myelopathy/tropical spastic paraparesis; Cn, count of nonsynonymous substitutions; Cs, count of synonymous substitutions; sN, nonsynonymous site; sS, synonymous site.

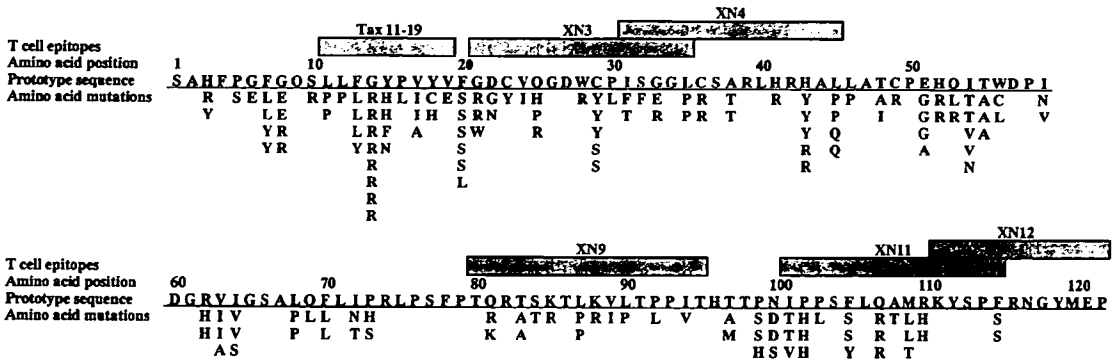

FIGURE 1. Amino acid replacements in HTLV-I Tax 1–122 in three HAM/TSP patients with HLA-A*02. The consensus *tax* gene sequence in each patient is the same as the ATK-1 sequence first reported. The prototype amino acid sequence from the ATK-1 is described above the line. The transverse bars indicate all the known CTL epitopes which can bind to HLA-A*02. The numbers above the prototype sequence indicate position number of the Tax protein. The capital letter under the line represents a single occurrence of amino acid replacement found at the position.

infection. In HTLV-I infection, it is unclear whether naturally occurring variant viruses escape from the host immune system and become predominant in the infected individuals. We analyzed longitudinal changes of variant virus proportion in association with variant virus-specific CTL.

Materials and Methods

Patients

Three patients with HAM/TSP (nos. 31, 38, 48) who had HLA-A*0201 allele were included (17). They were residing in Kagoshima, an endemic area of HTLV-I in Japan. The patients were diagnosed as HAM/TSP based on the neurological symptoms and seroreactivity to HTLV-I in accordance with the World Health Organization guidelines. These patients had not been treated with any antiretroviral drugs. PBMC were separated by Ficoll gradient centrifugation from heparinized blood repeatedly obtained from the patients and stored in liquid nitrogen until use. The Institutional Ethical Committee of Kagoshima University approved this study and informed consents were obtained from the patients.

Sequence analysis of HTLV-I tax *gene*

We used three samples from each patient as previously described (17). Cellular immune responses have predominantly been detected in a regulatory protein, HTLV-I Tax, in HTLV-I-infected individuals (7–9). The T cell epitopes restricted to HLA-A*02 accumulate in the N-terminal portion of the HTLV-I Tax protein (18). We therefore sequenced N-terminal of Tax (amino acid position 1–133). The method was previously described (17). Briefly, 100 ng of DNA extracted from the PBMC was amplified by 35 cycles of PCR. The first PCR products were further amplified by 20 cycles of nested PCR. The amplified products were purified using the QIA quick purification kit (Qiagen). The purified *tax* gene was subcloned into pCR-Blunt II-TOPO cloning vector (Invitrogen Life Technologies). After linearization by *Eco*RI digestion, the vector was purified by the QIA quick purification kit. The *tax* gene was sequenced using the Dye Terminator DNA Sequencing kit (Applied Biosystems) in an automatic sequencer (377 DNA Sequencer; Applied Biosystems). Approximately 50 clones were sequenced in each sample.

Comparison of selective pressures between CTL epitopes and the remaining regions

The CTL epitopes in the HTLV-I Tax were previously reported by epitope mapping (18–21). The reported CTL epitopes that restricted to HLA-A*02 in aa 1–133 are as follows: Tax 11–19 (aa 11–19; LLFGYPVYV), XN3 (aa 21–35; GDCVQGDWCPISGGL), XN4 (aa 31–45; ISGGLCSARL HRHAL), XN9 (aa 80–95; TQRTSKTLKVLTPPIT), XN11 (aa 101–115; IPPSFLQAMRKYSPF), and XN12 (aa 111–125; KYSPFRNGYMEP). Based on our sequence data of the *tax* genes, three phylogenetic trees were independently constructed by the maximum likelihood method for each patient (22). The ancestral sequence was inferred at each node in the phylogenetic tree using the maximum parsimony method (23). Then, the numbers of synonymous and nonsynonymous substitutions throughout each phylogenetic tree were estimated for each codon site. The total numbers (count) of synonymous (Cs) and nonsynonymous substitutions (Cn) independently occurring in three patients were summed in each codon site. The total numbers of Cs and Cn were counted in six regions identified as CTL epitopes and the remaining regions in *tax* genes. Also, we computed the total numbers of synonymous (sS) and nonsynonymous (sN) sites in the regions of compared sequences. To examine selective pressure in the regions, the test of significance between the rate of Cs to sS and the rate of Cn to sN was performed in the regions by the two-tailed χ^2 test (24). Values of $p < 0.05$ were considered significant.

Detection of positively selected regions of the tax *gene*

Positive selection pressure to the *tax* gene was examined by the modified method of Suzuki and Gojobori (25, 26) by three sequence data isolated from three patients. In this method, a phylogenetic tree was reconstructed and the ancestral sequence was inferred as described above. Then, the average number of synonymous (sS) and nonsynonymous (sN) sites and the total number of synonymous (Cs) and nonsynonymous (Cn) substitutions throughout the phylogenetic tree were estimated for each codon site by the Nei-Gojobori method (11). To examine positively selected regions in the *tax* gene, Cs, Cn, sS, and sN for a window size of five codon sites were calculated by sliding the window on the *tax* gene. The test of significance between the rate of Cs to sS and the rate of Cn to sN was performed in each window by the two-tailed Fisher's exact test (22).

Peptides

Substituted amino acids were predicted from the obtained sequence data of the *tax* gene. The variant epitope peptides of Tax 11–19 and influenza virus M1 peptide (GILGFVFTL) were synthesized using F-moc solid-phase methodology (Kurabo). All the variant epitopes were designated as G4R, when the glycine at position 4 of the Tax 11–19 was substituted to arginine. Influenza virus M1 peptide was used as a control peptide that binds to HLA-A*02 (27). Purity of the peptides was over 90% by HPLC analysis. The synthetic peptides were resolved in 50% DMSO in PBS at 1 mM.

Intracellular cytokine detection by flow cytometry

The assay was conducted by a modified protocol as previously described (17). Briefly, Hmy2.C1R cells transfected with HLA-A*0201 (Hmy-A2) were prepulsed with 1 μM Tax 11–19 or variant epitopes for 1 h and were washed. Cryopreserved PBMC were quickly thawed and washed. A total of 5×10^5 PBMC were cocultivated with the same number of peptide-prepulsed Hmy-A2 cells for 6 h. Brefeldin A (Sigma-Aldrich) was added to the cells at a final concentration of 10 μg/ml at the beginning of the culture to minimize the endogenous expression of HTLV-I protein on the infected cell surface. After culture, cells were harvested, washed, and stained with anti-human CD8 Ab conjugated with PC5 (Beckman Coulter) at 4°C for 20 min. Cells were washed and fixed with 4% paraformaldehyde for 5 min, then washed again. The cells, resuspended in 50 μl of permeabilization buffer containing 0.1% saponin (Sigma-Aldrich), were stained with anti-human IFN-γ Ab conjugated with FITC (BD Pharmingen) at 4°C for 20 min. Epics-XL flow cytometer and SYSTEM II software were used for fluorescent signal detection and data analysis (Beckman Coulter). Lymphocytes were readily distinguished from Hmy-A2 cells by size and were