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

Takashi Makino and Takashi Gojobori

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Yata, Mishima, Japan

Rates of protein evolution are thought to be influenced by features of protein-protein interaction (PPI). However, the most important features of interaction for determining the evolutionary rate are poorly understood. Here, we consider four categories for PPIs in *Saccharomyces cerevisiae*. Properties we consider are the extent to which proteins interact with proteins of the same function or different function (DF) and the extent to which these interactions involve connections in the dense part or sparse part (SP) of a PPI network. Our findings are that proteins with DF-SP interactions evolve at the slowest rate of all the proteins examined.

Introduction

The rate of amino acid substitution for a protein is determined mostly by the mutation rate of its encoding gene and its functional constraints on the gene product. In particular, the functional constraints are composed of several factors. For example, the so-called fitness effects as well as the gene expression level are typical factors of functional constraints because these factors are known to be negatively correlated with the rate of amino acid substitutions (Wilson, Carlson, and White 1977; Hirsh and Fraser 2001; Pal, Papp, and Hurst 2001; Jordan et al. 2002). Moreover, the functional constraints are probably influenced by the interactions among their gene products. The recent advancement of molecular technologies such as the high-throughput yeast two-hybrid system (Ito et al. 2000; Uetz et al. 2002) and mass spectrometry of communoprecipitated protein complexes (Gavin et al. 2002; Ho et al. 2002) has enabled us to understand the actual features of protein-protein interactions (PPIs). Of interest has been investigating the extent to which PPIs influence the rate of amino acid substitutions for proteins within the network. Here, we use interchangeably the phrase “rate of amino acid substitutions” and the phrase “evolutionary rate,” following the convention of molecular evolutionary studies.

When a two-dimensional presentation of PPI networks is made using a node and a line between neighboring nodes as a protein and an interaction between neighboring proteins, respectively, the PPI network is represented by a very complex structure of spider web-like networks. It has been reported, in this type of representation, that there are proteins tightly clustered in a particular part of the PPI network (Spirin and Mirny 2003). In particular, the proteins sharing a particular functional class tend to appear in the same part of a PPI network, making a cluster of the so-called “functional module” in the PPI network (Yook, Olsvai, and Barabasi 2004). Here, a functional class represents a category into which a group of particular proteins are classified according to the functional definitions. In other words, a functional module of the network is generally defined as a cluster of proteins sharing the same functional class that occupies a specific part of the network. In the PPI networks, the proteins building up a functional module have more interactions to other proteins within the functional module than to those outside the module. For example, VPS16 of *Saccharomyces cerevisiae* is clustered in a functional module that is required for sorting proteins in vacuole (fig. 1a).

On the other hand, there are proteins known to interact with those having different functional classes (Han et al. 2004). Calmodulin, which is a master regulator of calcium-mediated signaling (Davis et al. 1986), interacts with several proteins of different functional classes such as homeostasis of cations, proteins folding and stabilization, budding, cell polarity, and filament formation (Han et al. 2004). For these proteins, the gene expression patterns do not correlate with those of their PPI partner proteins (Han et al. 2004), 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 serine/threonine protein kinase of *S. cerevisiae* is in a framework module and interacts with proteins classified into different functional classes (fig. 1b). 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. We used the PPIs in *S. cerevisiae* that have well been documented based not only on hundreds of small-scale experiments but also on high-throughput methods. 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.

Materials and Methods

Protein Sequences

Protein sequences of *S. cerevisiae* and *Saccharomyces paradoxus* were downloaded from *Saccharomyces* Genome Database (ftp://genome-ftp.stanford.edu/pub/yeast/data_download/sequence/).
functional classes ‘protein fate,’ ‘cell cycle/DNA processing,’ ‘metabolism,’ ‘cellular transport,’ and ‘transcription,’ respectively.

Protein-Protein Interactions

Although the information on approximately 14,000 PPIs in S. cerevisiae is stored in the Database of Interacting Proteins (DIPs, http://dip.doe-mbi.ucla.edu/), we used only the CORE data set of DIP that contained 6,205 PPIs after we excluded proteins that interact with themselves (self-interactions) from the present analysis. We removed 231 PPIs from the CORE because those were related to proteins derived from pseudogenes and erroneously predicted genes (Kellis et al. 2003). Thus, we used 5,974 PPIs included in the CORE. The CORE has the following three characteristic features. First, the interactions were determined by hundreds of small-scale experiments. They are regarded as reliable PPIs because they are derived from individual research papers. Second, each interaction was identified by independent high-throughput experiments at least twice. PPIs identified by high-throughput methods in the CORE have high correlation with respect to function and cellular location. Thus, the data quality in this feature is as reliable as that produced by small-scale experiments (Deane et al. 2002). Therefore, the data quality in this feature is also high. Third, each interaction was confirmed by examining if paralogues interact with the same proteins or the paralogues themselves. If a pair of proteins encoded by a duplicated gene pair is known to share the same PPI partners, the PPIs between the proteins and the same partners are regarded as reliable relationships. In fact, paralogues often interact with the same protein (Deane et al. 2002).

Classification of Interacting Proteins Based on Their Coefficients of the Same Functional Class

We defined a protein having PPI partners of the same functional class with a high frequency as a same function (SF) protein, on the other hand, a protein having PPI partners of different functional classes with a high frequency as a different function (DF) protein. According to this definition, the SF and DF proteins should build up a functional module and a framework module, respectively. In other words, the SF proteins interact with one another within a functional module, and the DF proteins mediate different functions by interacting their PPI partners with different functional classes. To classify the SF and DF proteins in a precise fashion, we devised a measure named as the coefficient of the same functional class, which is defined as the proportion of PPI partners belonging to the same functional class in all its PPI partners. For example, if a given protein has 10 PPI partners in which eight partners belong to the same functional class, the coefficient is computed as 0.8 (=8/10). For the functional class, we used the functional classifications that have been defined by the Munich Information Center for Protein Sequences (MIPS) database (Mewes et al. 2002).

In this database, each protein is assigned to one or more functional classes in the total of 10 classes that were based on the functional information reported in the literature (table 1). It followed that 1,891 proteins were assigned to at least one functional class out of 2,435 proteins in the CORE (see Materials and Methods). Then, we computed the coefficient of the same functional class for each protein. When a particular protein was assigned to more than one functional class, we computed the coefficients for all functional classes to which the protein belongs allowing the double count of the protein. We then took a functional class showing the largest value of the coefficients as the representative functional class for this protein.

For the classification of the SF and DF proteins, we computed the average over all the 1,891 clustering coefficients. We used the average value (0.63) obtained as the criterion for their quantitative identification. In other words, if the clustering coefficient for a protein is larger than 0.63, the protein is identified as a SF protein and otherwise identified as a DF protein. Using this criterion, we classified the 1,891 proteins into 1,079 SF and 812 DF proteins. We used the average value of the clustering coefficients as cutoffs to classify the proteins into DF or SF proteins. Although we used the cutoff over a wide range of the values, our results were unchanged.

Classification of Interacting Proteins Based on Their Clustering Coefficients

As mentioned earlier, the SF proteins tend to occupy a dense part of interactions, like a cluster, in the PPI network. In order to distinguish between the dense and sparse parts of interactions in the PPI networks, we defined those parts independently of the SF and DF proteins. For this reason, denoting proteins in dense and sparse parts of the PPI

![Figure 1](http://example.com/image.png)

**Figure 1.** (A) A protein in a functional module and (B) a protein in a framework module of the PPI network. The filled circles and lines represent proteins and PPIs, respectively. The black lines indicate interactions between VPS16 and its PPI partners and between SPS1 and its PPI partners. The gray lines indicate interactions among PPI partners. VPS16 interacts with proteins classified into different functional classes “protein fate,” “cell cycle/DNA processing,” “metabolism,” “cellular transport,” and “transcription,” respectively.

<table>
<thead>
<tr>
<th>Functional Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolism</td>
</tr>
<tr>
<td>Energy</td>
</tr>
<tr>
<td>Cell cycle and DNA processing</td>
</tr>
<tr>
<td>Transcription</td>
</tr>
<tr>
<td>Protein synthesis</td>
</tr>
<tr>
<td>Protein fate (folding, modification, destination)</td>
</tr>
<tr>
<td>Cellular transport, transport facilitation, and transport routes</td>
</tr>
<tr>
<td>Cellular communication/signal transduction mechanism</td>
</tr>
<tr>
<td>Cell rescue, defense, and virulence</td>
</tr>
<tr>
<td>Interaction with the cellular environment</td>
</tr>
</tbody>
</table>

**Table 1**

Functional Classifications in the Munich Information Center for Protein Sequences Database

<table>
<thead>
<tr>
<th>Functional Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolism</td>
</tr>
<tr>
<td>Energy</td>
</tr>
<tr>
<td>Cell cycle and DNA processing</td>
</tr>
<tr>
<td>Transcription</td>
</tr>
<tr>
<td>Protein synthesis</td>
</tr>
<tr>
<td>Protein fate (folding, modification, destination)</td>
</tr>
<tr>
<td>Cellular transport, transport facilitation, and transport routes</td>
</tr>
<tr>
<td>Cellular communication/signal transduction mechanism</td>
</tr>
<tr>
<td>Cell rescue, defense, and virulence</td>
</tr>
<tr>
<td>Interaction with the cellular environment</td>
</tr>
</tbody>
</table>

Evolution of Protein-Protein Interactions 785
network as the dense part (DP) and sparse part (SP) proteins, respectively, we defined them using the clustering coefficient (Watts and Strogatz 1998). The definition of the clustering coefficient is as follows. Suppose that a protein \(v\) has \(k\) interacting proteins as the PPI partners. Then, the number of possible interactions between those PPI partners is given to be at most \(k(k,-1)/2\). This is because the maximum number of interactions is taken when every interacting protein of \(v\) is connected to every other interacting protein of \(v\). The clustering coefficient is defined as the fraction of the actual interactions in all the possible interactions (Watts and Strogatz 1998). For example, if a proteins has its five PPI partners and two of them have a single interaction between them, the clustering coefficient of the protein is computed as \(0.10 = 1/(5(5-1)/2)\) because the maximum number of possible interactions for this protein’s PPI partners is \(10 = 5(5-1)/2\). For 1,707 proteins with more than one PPI partners out of the 2,435 proteins in the CORE, we computed the clustering coefficient for each protein.

For the classification of the DP and SP proteins, we computed the average over all the 1,707 clustering coefficients. We used the average value (0.30) obtained as the criterion for their quantitative identification. In other words, if the clustering coefficient for a protein is larger than 0.30, the protein is identified as a DP protein and otherwise identified as a SP protein. Using this criterion, we classified the 1,707 proteins into 693 DP and 1,014 SP proteins. Although we used the cutoff over a wide range of the values, our results were unchanged.

### Classification of Interacting Proteins Based on Two Criteria

In the previous section, we invented two categorical classifications; one category for the SF and DF proteins and the other for the SP and DP proteins. When the features of the PPI network is described in a more detailed way, the proteins in the PPI network may be classified further into the four categories, setting up simultaneously the criteria described in the previous section. They were the SF-DP, SF-SP, DF-DP, and DF-SP proteins. Out of the 2,435 proteins in the CORE, we chose 1,446 proteins that have more than one PPI partners and are assigned to at least one functional class. By the way mentioned above, we then classified those proteins into four categories of the SF-DP, the SF-SP, the DF-DP, and the DF-SP proteins. A total of the 1,446 proteins were classified finally into 461 SF-DP, 377 SF-SP, 128 DF-DP, and 480 DF-SP proteins.

### Estimation of Evolutionary Distances

Using the number of amino acid substitutions per site as the relative evolutionary rate, we estimated the evolutionary distances of the orthologous genes encoding proteins in the PPI network between \(S.\ cerevisiae\) and \(S.\ paradoxus\). This is because \(S.\ paradoxus\) appeared to be the most closely related species to \(S.\ cerevisiae\) among all the organisms whose whole genome sequences were currently available. In their orthologous genes that have been identified based on the sequence similarities and their conserved gene order in the genome (Kellis et al. 2003), we found that there were the orthologous genes encoding a total of 2,152 proteins that were used in this study. There were 1,035 SF, 763 DF, 668 DP, 965 SP, 443 SF-DP, 353 SF-SP, 122 DF-DP, and 457 DF-SP proteins by the orthologous genes. A pairwise alignment was conducted using ClustalW (version 1.81; Thompson, Higgins, and Gibson 1994) for each set of the 2,152 orthologous pairs. In order to compute the rates of amino acid substitution between the orthologous pair of gene products, we estimated the number of amino acid substitutions by Kimura’s methods (1983), which is implemented in PHYLIP (version 3.573c). In this estimation, YJR159W was removed from this analysis because of a considerably low overlap (less than 20%) in the sequence alignment between the protein sequence and its orthologues.

### Results

#### Comparison of Evolutionary Rates Between Different Categorical Proteins

**SF Versus 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 (Fraser et al. 2002; Fraser, Wall, and Hirsh 2003). A recent study reported that proteins in the center of the PPI networks evolve more slowly, regardless of the number of PPI partners (Hahn and Kern 2005). 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 (Schwikowski, Uetz, and Fields 2000; Ge et al. 2001). On the contrary, there are proteins that interact with those belonging to different functional classes (Han et al. 2004). Thus, 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. For this comparative study, we used the evolutionary distances for 1,035 SF and 763 DF proteins for the comparison. As mentioned above, the distances were used as the relative evolutionary rates in this study. As a result, we found that the DF proteins evolved at a slower rate, with statistical significance, than the SF proteins (fig. 2; Mann-Whitney U test, \(P < 10^{-5}\)). Thus, we concluded that the DF proteins are under much stronger functional constraints than the SF proteins.

**DP Versus SP Proteins**

It has been reported that there are proteins tightly clustered in a particular part of the PPI network (Spirin and Mirny 2003). 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 (fig. 3; Mann-Whitney U test, \( P < 10^{-5} \)). 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 (Wuchty, Oltavi, and Barabasi 2003). 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. When we classified the proteins into four categories, we found that DF proteins were disproportionately located in sparse parts of the PPI network. On the other hand, we found that SF proteins tended to be not only located in dense parts of the network but also in the sparse parts of the network. We found that of all proteins examined, the DF-SP proteins tended to evolve at a slower rate than the SF-DP and SF-SP proteins regardless of the degrees of the connectivity.

Comparison of Evolutionary Rate for Different Degrees of Connectivity

One might argue that the difference in evolutionary rate between the DF-SP proteins and the other proteins was caused by different degrees of the connectivity that is the absolute number of PPI partners because it has been reported that the proteins having many PPI partners evolve slowly (Fraser et al. 2002; Fraser, Wall, and Hirsh 2003). Therefore, we made the comparison of the evolutionary rates among the proteins for different degrees of the connectivity. The result indicated that the DF-SP proteins tended to evolve at a slower rate than the SF-DP and SF-SP proteins regardless of the degrees of the connectivity.
(fig. 5). In particular, even when we removed the proteins having many PPI partners (more than five or 10) from this analysis, the DF-SP proteins still had slower evolutionary rates, with statistical significance, than the other proteins examined (Mann-Whitney $U$ test, $P < 0.05$). DF-SP proteins showed no significant difference in evolutionary rate from DF-DP proteins. The number of the DF-DP proteins was very small after removing the proteins having many PPI partners. This would be the reason why there was no statistical difference in their evolutionary rates. The result suggests that the DF-SP proteins are under strong functional constraint. This can not be explained as a consequence of the absolute number of PPI partners.

Structural Features of the Proteins Influenced by Their PPI Partners

Let us describe the structural features of these proteins. To examine the difference in molecular weights among the SF-DP, SF-SP, DF-DP, and DF-SP proteins, we counted the number of amino acid residues for each protein. As a result, we found that the SF-DP proteins had, on the average, the smallest number of amino acid residues (497.8 residues/protein). On the other hand, the SF-SP and DF-SP proteins had, on the average, the large number of amino acid residues (643.4 and 640.8 residues/protein, respectively). There were no statistically significant differences in the number of amino acid residues among the SF-SP, DF-DP, and DF-SP proteins. In particular, the number of amino acid residues for the SF-DP proteins was found to be smaller, with statistical significance, than that of the other three categorical proteins (fig. 6; Mann-Whitney $U$ test, $P < 0.005$). Thus, it suggests that the molecular weights of proteins are also influenced by the nature of PPIs during evolution.

Discussion

We have found that the DF proteins evolved at a slower rate than the SF proteins. The observation suggests that the proteins involved with multidifferent 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 proteins. In fact, we have shown that the DF-SP proteins evolved at the slowest rate among the four categories of interacting proteins (SF-DP, SF-SP, DF-SP, and DF-SP proteins). 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 are dependent on the nature of interacting proteins in the PPI network. We also found that the molecular weights of the SF-DP proteins are the smallest. It is possible that there is a great degree of differences in structural constraints among the proteins having different types of their PPI partners.

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 (Teichmann 2002). 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 (Mewes et al. 2002). As expected, we found that the SF proteins contained subunits of protein complexes more 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 (Wuchty, Oltvai, and Barabasi 2003). 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 in all interacting proteins mentioned above. This conclusion highlights the importance of studying the evolution of the DF-SP proteins for understanding the essential features of PPI network evolution.

Acknowledgments

We are grateful to Yoshio Tateno, Kazuho Ikeo, Yoshiyuki Suzuki, and all the members of the DNA Analysis Laboratory for valuable comments and discussion. This research was supported by Grants-in-Aid for Creative Basic Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to T.G.).

Literature Cited


Peter Lockhart, Associate Editor

Accepted December 21, 2005