12 Katz, L. and Burge, C.B. (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.* 13, 2042–2051
13 Drummond, D.A. *et al*. (2006) A single determinant dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* 23, 327–337
14 Bloom, J.D. *et al*. (2005) Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U. S. A.* 102, 606–611
15 Subramanian, S. and Kumar, S. (2004) Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics* 168, 373–381
16 Duret, L. and Mouchiroud, D. (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* 17, 68–74
17 Hurst, L.D. and Smith, N.G. (1999) Do essential genes evolve slowly? *Curr. Biol.* 9, 747–750
18 Hirsh, A.E. and Fraser, H.B. (2001) Protein dispensability and rate of evolution. *Nature* 411, 1046–1049
19 Hirsh, A.E. (2003) Rate of evolution and gene dispensability – reply. *Nature* 421, 497–498
20 Yang, J. *et al*. (2003) Rate of protein evolution versus fitness effect of gene deletion. *Mol. Biol. Evol.* 20, 772–774
21 Krylov, D.M. *et al*. (2003) Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13, 2229–2235
22 Fraser, H.B. *et al*. (2002) Evolutionary rate in the protein interaction network. *Science* 296, 750–752
23 Teichmann, S.A. (2002) The constraints protein–protein interactions place on sequence divergence. *J. Mol. Biol.* 324, 399–407
24 Hahn, M.W. and Kern, A.D. (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* 22, 803–806
25 Bloom, J.D. and Adami, C. (2003) Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions. *BMC Evol. Biol.* 3, 21
26 Jordan, I.K. *et al*. (2003) No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol. Biol.* 3, 1
27 Fraser, H.B. (2005) Modularity and evolutionary constraint on proteins. *Nat. Genet.* 37, 351–352
28 Akashi, H. and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3695–3700
29 Lemos, B. *et al*. (2005) Evolution of proteins and gene expression levels are coupled in drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Mol. Biol. Evol.* 22, 1345–1354
30 Duret, L. and Mouchiroud, D. (1999) Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4482–4487
31 Salathé, M. *et al*. (2006) The effect of multifunctionality on the rate of evolution in yeast. *Mol. Biol. Evol.* 23, 721–722
32 Chen, Y. and Dokholyan, N.V. The coordinated evolution of yeast proteins constrained by functional modularity. *Trends Genet* (in press)
33 Parmley, J.L. *et al*. (2006) Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol. Biol. Evol.* 23, 301–309
34 Wilson, A.C. *et al*. (1977) Biochemical evolution. *Annu. Rev. Biochem.* 46, 573–639
35 Jordan, I.K. *et al*. (2002) Essential genes are more evolutionarily conserved than are nonessential genes in Bacteria. *Genome Res.* 12, 962–968
36 Pal, C. *et al*. (2003) Rate of evolution and gene dispensability. *Nature* 421, 496–497
37 Wall, D.P. *et al*. (2005) Functional genomic analysis of the rates of protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5483–5488
38 Fang, G. *et al*. (2005) How essential are nonessential genes? *Mol. Biol. Evol.* 22, 2147–2156
39 Dickerson, R.E. (1971) The structures of cytochrome c and the rates of molecular evolution. *J. Mol. Evol.* 1, 26–45
40 Castillo-Davis, C.I. *et al*. (2004) Cis-regulatory and protein evolution in orthologous and duplicate genes. *Genome Res.* 14, 1530–1536

0168-9525/$ - see front matter
doi:10.1016/j.tig.2006.06.004

Genome Analysis

The coordinated evolution of yeast proteins is constrained by functional modularity

Yiwen Chen and Nikolay V. Dokholyan

Department of Physics and Astronomy and Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Functional modularity is a key attribute of cellular systems and has important roles in evolution. However, the extent to which functional modularity affects protein evolution is largely unknown. Here, we analyzed the evolution of both sequence and expression level of proteins in the yeast *Saccharomyces cerevisiae* and found that proteins within the same functional modules evolve at more similar rates than those between different modules. We also found stronger co-evolution of expression levels between proteins within functional modules than between them. These results suggest that a coordinated evolution of both the sequence and expression level of proteins is constrained by functional modularity.

Corresponding author: Dokholyan, N.V. (dokh@med.unc.edu).
Available online 23 June 2006

Introduction

The accumulation of in-depth knowledge of cellular systems at the molecular level and progress in the development of functional genomic techniques [1–5] have both significantly advanced system-level studies of cellular phenomena [6]. From these system-level studies, we begin to grasp key features [7] and principles of cellular organization. One of the essential attributes of cellular

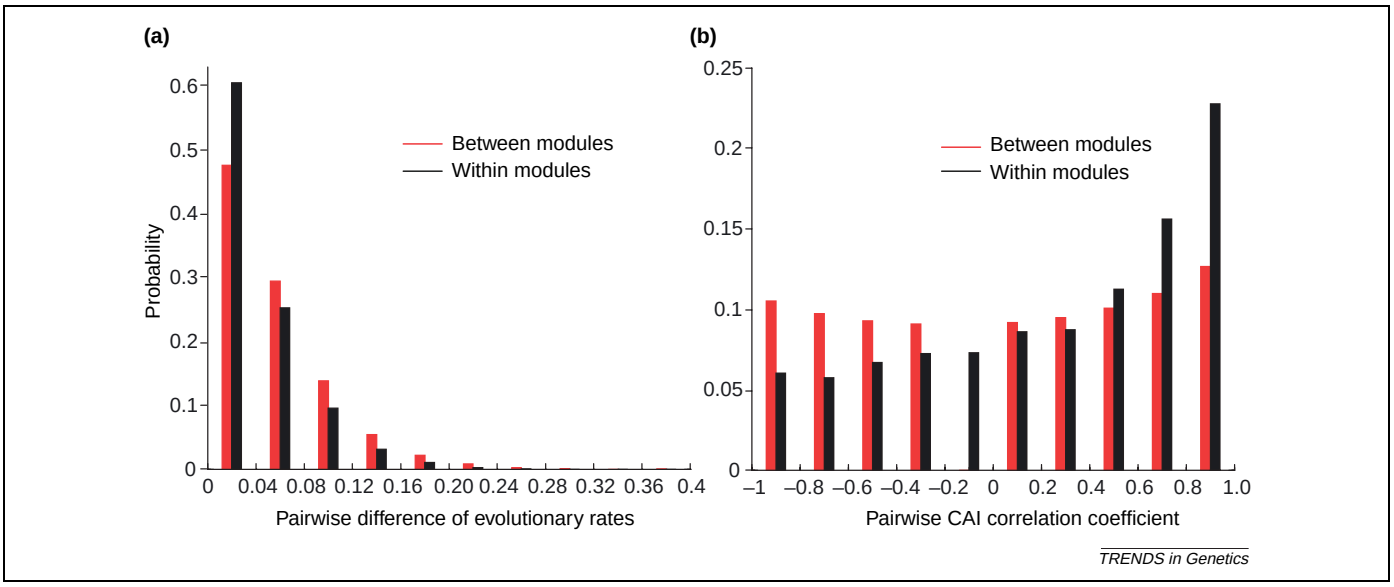

Figure 1. The histograms of evolutionary rate difference and co-evolution of expression levels. The histograms of **(a)** pairwise differences in evolutionary rates, **(b)** pairwise CAI correlation coefficients of proteins within (black) and between (red) modules. The histograms in (a) are plotted in the range between 0 and 0.4 with the bin size of 0.04; those in (b) are plotted in the range between −1 and 1 with the bin size of 0.2.

systems is modularity [7–13] – the modular organization of proteins.

A functional module consists of a spatially or chemically isolated set of functionally associated biological components that accomplishes a discrete collective function. The formation of functional modules is characterized by stronger functional associations within modules than between them and offers a natural solution to accommodating the evolvability of cellular systems [9,11–13]. Once modules are established, rewiring the connections between different modules can create new cellular functions. The looser functional inter-connections between modules reduce the pleiotropy, so that each module can be changed during the course of evolution without significantly affecting other modules. Despite the close relationship between modularity and evolution [8,9,11–13], it is largely unknown to what extent functional modularity affects protein evolution. To uncover the role of functional modularity in constraining protein evolution, we analyzed the evolution of both sequence and expression level of proteins in the yeast *Saccharomyces cerevisiae*, for which extensive functional genomics and genomic sequence data are available.

Proteins within the same functional modules evolve at more similar rates than those between different modules

The functional modules in *S. cerevisiae* were systematically characterized in a recent study [14], by integrating the information obtained from different functional genomics experiments [1–5] (supplementary data online). First, we compared the pairwise differences in evolutionary rates between proteins within the same functional modules with those between different modules. The difference in evolutionary rates (supplementary data online) between two proteins i and j is defined as $d=|d_i-d_j|$, where d_i and d_j are the evolutionary rates of protein i and j (the ratio of non-synonymous substitution per site to synonymous substitution per site), respectively. If proteins evolve in a coordinated manner, we expect that they evolve at similar rates.

Indeed, we found that the average difference in evolutionary rates of the proteins between different modules is $\sim$30% more than the difference in evolutionary rates of the proteins within the same module (within the same module $d=0.042$; between different modules $d=0.054$). The histograms of pairwise differences in evolutionary rates within and between modules are significantly different (Figure 1a), as supported by the Kolmogorov–Smirnov statistical test (KS test, $P<10^{-109}$) [15]. This difference indicates that the proteins within the same modules evolve at more similar rates than those that are in separate modules. To assess the effect that physically interacting proteins have on evolutionary rates, we performed the same analysis while controlling for the protein pairs that have experimentally supported physical interactions in the same module. We found that the differences in evolutionary rates between the protein pairs with no experimentally supported physical interactions are almost indistinguishable from the original set that contains all pairs (KS test, $P>0.6$), indicating that the observed pattern for all pairs in the same module is not simply due to physically interacting proteins (supplementary data online).

Co-evolution of expression level between proteins within the same functional modules is more pronounced than that between different modules

The evolutionary history of a protein is imprinted not only in its sequence but also in its expression level. To uncover the influences of functional modularity on evolutionary change of the protein expression level, we analyzed the co-evolution of expression levels between

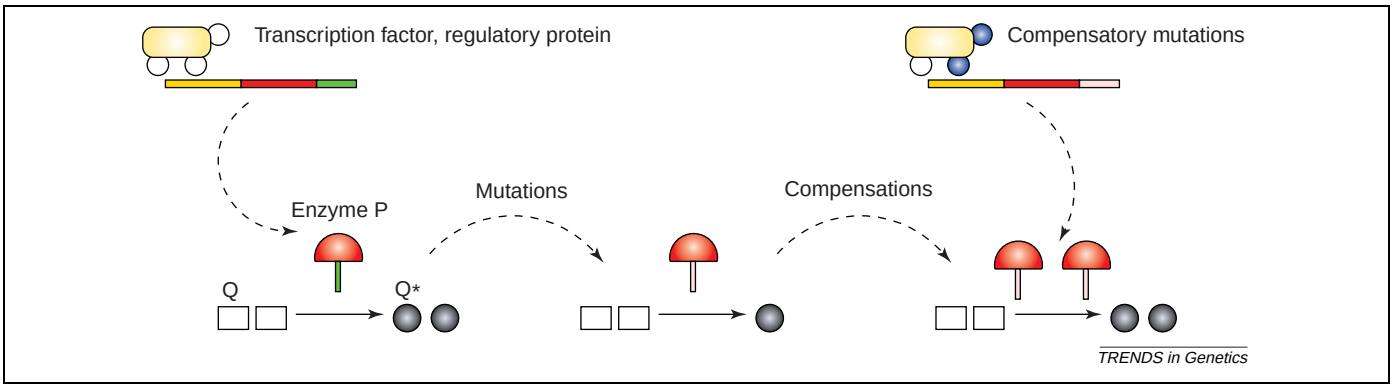

Figure 2. A hypothetical example of the coordinated evolution of sequences between proteins within the same module. The reaction converting Q to Q* is catalyzed by an enzyme P. When mutations in P cause a decrease of the conversion efficiency of the reaction, there can be compensation on the expression level of P to maintain the conversion efficiency of the reaction and thus the balance in metabolism. The increase in the expression level might require mutations in the transcription factors, repressors or activators to alter the binding affinity between these proteins and *cis*-regulatory elements or transcriptional machinery.

proteins within the same modules and those between different modules. Methods that measure expression levels include mRNA abundance (measured in the laboratory) [16] and the codon adaptation index (CAI), a quantitative measure of codon bias [17] (supplementary data online), and these have been used in recent studies [18–20]. A CAI value is more likely to reflect the expression level related to protein evolution because (i) the laboratory conditions, under which abundance of mRNA is measured, might not reflect evolutionarily relevant environmental conditions; and (ii) the CAI value estimates expression during the recent evolutionary history of a gene instead of during the much shorter time-scale in the laboratory (supplementary data online). For a given pair of orthologous gene sets, the Pearson correlation coefficient between the CAI values in the corresponding set is used as a measure of the co-evolution strength of expression levels between this pair. We calculated all pairwise CAI correlation coefficients within and between modules (supplementary data online). We found stronger co-evolution of expression level between proteins within the same modules, indicated by a distribution significantly skewed towards greater CAI correlation coefficients, (Figure 1b) compared with the expression level of those between different modules (KS test, $P<10^{-162}$). The exclusion of the protein pairs that have experimentally supported physical interactions in the same module shows no significant effects on the observed pattern (supplementary data online).

Functional modularity and co-evolution

It has been suggested that physically interacting proteins evolve in a coordinated manner in both sequence and expression level. Early evidence came from the finding that physically interacting proteins have similar phylogenetic trees or significant correlations between sequence distances [21–24]. Recently, Fraser *et al.* showed that physically interacting proteins evolve at similar rates [25] and co-evolve at the expression level [18]. By contrast, our observation suggests that coordinated evolution occurs not only between physically interacting proteins but also between proteins within the same modules (i.e. between proteins that have strong functional associations). Modules are cohesive functional blocks in cellular systems. To preserve a specific function of each module, a substitution in one protein can result in selection pressure for compensatory changes in other proteins in the same module. For example, in a metabolic reaction catalyzed by an enzyme, P, a metabolite Q is transformed to Q* (Figure 2). Mutations in protein P causing the decrease in converting efficiency from Q to Q* can affect the cellular function and fitness of the organism. To maintain the stable conversion efficiency from Q to Q* and the balance in metabolism, there can be a compensatory increase in the expression level of enzyme P. The increase in the expression level can require mutations in the transcription factors, repressors or activators to alter the binding affinity between these proteins and the *cis*-regulatory elements or transcriptional machinery. This hypothetical example is a simplification of processes underlying molecular evolution. The intricate and non-linear nature of functional associations within modules, such as feedback loops [7], could give rise to further scenarios that are more complex.

To maintain the functional integrity of each module, the coordinated change does not only occur in sequence, but also at the protein expression level. One of the major selection pressures causing the coordinated changes of expression level is to maintain proper stoichiometry of proteins within the same modules. For example, normal functioning of a protein complex (either transient or stable) requires a proper stoichiometry between the subunits that form the complex. The change of the expression level of one subunit in the complex without compensating changes of other subunits can cause malfunction owing to the imbalance of the stoichiometry. The correct stoichiometry is also required between the enzymes in the same module that catalyze closely coupled reactions in a metabolic network to maintain the balance of metabolisms.

Although coordinated evolution offers an appealing explanation for the current observations, alternative interpretations should be examined. For example, recent studies indicate that the expression level of a protein has important effects on its evolutionary rate [26–28]. Thus, the greater similarity of the evolutionary rates observed within the same modules could be simply due to the greater similarity in expression levels.

To address this issue, we calculated Kendall's partial tau-values [29], a nonparametric measure of partial correlation between the pairwise evolutionary rate difference and the following two variables (while controlling for the effect of each variable): (i) the category in which each protein pair belongs; and (ii) the difference in pairwise expression level. The category is a discrete variable: 'category 1' for proteins that belong to the same module; 'category 2' for those that belong to different modules. This quantitative measure enables us to examine the correlation between two variables while controlling for the third, potentially related variable. We found that the expression level contributes much less significantly than functional modularity to the observed patterns (supplementary data online).

The existence of many recently duplicated genes within the same module might provide another explanation for the greater similarity in evolutionary rates within rather than between modules. To test this hypothesis, we use the gene family data of *S. cerevisiae* that are catalogued as clusters of orthologous groups (COG) [30] to identify all the duplicate genes in the same module (supplementary data online). The duplicate gene pairs only represent a small fraction of all pairs in the same module. Hence, the exclusion of those pairs has almost no effect on the distribution of similarity in evolutionary rates (KS test, $P \sim 1$).

In the coordinated evolution scenario, it is expected that proteins that physically interact with each other co-evolve more strongly than those that do not, even if they are in the same module, owing to the intense selection pressures caused by physical interaction. Consistent with this expectation, we observed that the similarity in evolutionary rates and the co-evolution strength in expression level are greater between proteins that physically interact than between those that do not interact in the same module (supplementary data online). These observations cannot be explained using the alternative scenarios discussed earlier. Taken together, coordinated evolution is the main, although perhaps not the sole, factor contributing to our observations.

Concluding remarks

We presented several lines of evidence suggesting that the coordinated evolution of both sequence and expression level of proteins is constrained by functional modularity. The constraints applied by modularity on protein evolution reflect the complex interplay between evolutionary changes across molecular and cellular levels. We believe such complex interplay will be better elucidated in the context of systems biology and a comprehensive understanding of the cellular phenomena at the system level will offer a more mechanistic picture of evolution.

Acknowledgements

We thank Aaron. E. Hirsh for sharing his data sets of evolutionary rates and CAI values. We also thank Kyle Wilcox and Jason Lieb for their comments on this manuscript.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2006.06.008

References

1 Ho, Y. *et al*. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183
2 Tong, A.H. *et al*. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294, 2364–2368
3 Uetz, P. *et al*. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627
4 Gavin, A.C. *et al*. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147
5 Gollub, J. *et al*. (2003) The Stanford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res.* 31, 94–96
6 Kitano, H. (2002) Systems biology: a brief overview. *Science* 295, 1662–1664
7 Alon, U. (2003) Biological networks: the tinkerer as an engineer. *Science* 301, 1866–1867
8 Fraser, H.B. (2005) Modularity and evolutionary constraint on proteins. *Nat. Genet.* 37, 351–352
9 Hartwell, L.H. *et al*. (1999) From molecular to modular cell biology. *Nature* 402, 47–52
10 Rives, A.W. and Galitski, T. (2003) Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1128–1133
11 Schlosser, G. (2002) Modularity and the units of evolution. *Theory in Biosciences* 121, 1–80
12 Schlosser, G. and Wagner, G.P. (2004) *Modularity in Development and Evolution*, University of Chicago Press
13 Gerhart, J. and Kirschner, M. (1997) *Cells, Embryos and Evolution*, Blackwell Science
14 Lee, I. *et al*. (2004) A probabilistic functional network of yeast genes. *Science* 306, 1555–1558
15 Press, W.H. (1992) *Numerical recipes in C: The Art of Scientific Computing*, 2nd edn, Cambridge University Press
16 Akashi, H. (2003) Translational selection and yeast proteome evolution. *Genetics* 164, 1291–1303
17 Sharp, P.M. and Li, W.H. (1987) The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295
18 Fraser, H.B. *et al*. (2004) Coevolution of gene expression among interacting proteins. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9033–9038
19 Pal, C. *et al*. (2003) Rate of evolution and gene dispensability. *Nature* 421, 496–497
20 Wall, D.P. *et al*. (2005) Functional genomic analysis of the rates of protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5483–5488
21 Pazos, F. and Valencia, A. (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng.* 14, 609–614
22 Pazos, F. *et al*. (1997) Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.* 271, 511–523
23 Altschuh, D. *et al*. (1987) Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *J. Mol. Biol.* 193, 693–707
24 Goh, C.S. *et al*. (2000) Co-evolution of proteins with their interaction partners. *J. Mol. Biol.* 299, 283–293
25 Fraser, H.B. *et al*. (2002) Evolutionary rate in the protein interaction network. *Science* 296, 750–752
26 Drummond, D.A. *et al*. (2005) Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14338–14343
27 Pal, C. *et al*. (2001) Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927–931
28 Rocha, E.P.C. and Danchin, A. (2004) An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol. Biol. Evol.* 21, 108–116
29 Gibbons, J.D. (1993) *Nonparametric Measures of Association*, Sage Publications
30 Tatusov, R.L. *et al*. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41

doi:10.1016/j.tig.2006.06.008