J Mol Evol (2009) 68:248–255
DOI 10.1007/s00239-009-9199-4

Low Contents of Carbon and Nitrogen in Highly Abundant Proteins: Evidence of Selection for the Economy of Atomic Composition

Ning Li · Jie Lv · Deng-Ke Niu

Received: 29 July 2008 / Accepted: 12 January 2009 / Published online: 10 February 2009

Abstract Proteins that assimilate particular elements were found to avoid using amino acids containing the element, which indicates that the metabolic constraints of amino acids may influence the evolution of proteins. We suspected that low contents of carbon, nitrogen, and sulfur may also be selected for economy in highly abundant proteins that consume large amounts of the resources of cells. By analyzing recently available proteomic data in *Escherichia coli*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*, we found that at least the carbon and nitrogen contents in amino acid side chains are negatively correlated with protein abundance. An amino acid with a high number of carbon atoms in its side chain generally requires relatively more energy for its synthesis. Thus, it may be selected against in highly abundant proteins either because of economy in building blocks or because of economy in energy. Previous studies showed that highly abundant proteins preferentially use cheap (in terms of energy) amino acids. We found that the carbon content is still negatively correlated with protein abundance after controlling for the energetic cost of the amino acids. However, the negative correlation between protein abundance and energetic cost disappeared after controlling for carbon content. Building blocks seem to be more restricted than energy. It seems that the amino acid sequences of highly abundant proteins have to compromise between optimization for their biological functions and reducing the consumption of limiting resources. By contrast, the amino acid sequences of weakly expressed proteins are more likely to be optimized for their biological functions.

Keywords Atomic content · Energetic cost · Amino acid usage · Resource availability · Protein abundance · Protein turnover rate

Introduction

It has long been recognized that proteins with a low sulfur content are preferentially expressed in microorganisms under conditions of sulfur starvation (Cuhel et al. 1981; Mazel and Marliere 1989; Fauchon et al. 2002; Boer et al. 2003). In *Escherichia coli*, *Saccharomyces cerevisiae*, and *Salmonella typhimurium*, enzymes responsible for the assimilation of particular element (carbon, nitrogen, or sulfur) are found to have a lower content of the element they assimilate than the whole proteome (Pardee 1966; Baudouin-Cornu et al. 2001).

Is selection for economy in atomic usage limited to these specific groups of proteins? Baudouin-Cornu et al. (2004) found that proteome-wide atomic compositions vary considerably among species. Recently, we found an association between the depth of isolation of *Prochlorococcus* strains and the atomic composition of their proteomes, which further supports that the elemental composition of proteomes is shaped by the availability of resources in the environment (Lv et al. 2008). A highly

Electronic supplementary material The online version of this article (doi:10.1007/s00239-009-9199-4) contains supplementary material, which is available to authorized users.

N. Li · J. Lv · D.-K. Niu (✉)
MOE Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing 100875, China
e-mail: dkniu@bnu.edu.cn; dengkeniu@hotmail.com

abundant protein may have 10^5 to 10^6 copies per cell, while a scarce protein has only 50 to 100 copies per cell (Ghaemmaghami et al. 2003; Ishihama et al. 2008). Consequently, a substitution that reduces one carbon atom per protein (for example, from a leucine to valine) will reduce the number of carbon atoms per cell by 10^5 to 10^6 if it occurs in a highly abundant protein, but only by tens of carbons if it occurs in a scarce protein. The selective coefficient for the substitution in a highly abundant protein is thus expected to be >1000 times higher than that in a scarce protein, provided that the biological function of the protein is not affected. Thus, it could be conjectured that highly abundant proteins are subject to selection for economy because their expression consumes a large amount of resources. By contrast, selection for economy is expected to be very weak in scarce proteins. A pattern thus predicted is that highly abundant proteins have lower carbon, nitrogen, and sulfur contents in their amino acid side chains than scarce proteins. Elser et al. (2006) showed that the nitrogen content of plant proteins is lower than that of animal proteins. More importantly, they found that protein nitrogen usage is negatively correlated with mRNA abundance in plants, but not in animals. Their results indicate a selective force imposed by nitrogen limitation operative on plants, but not on animals. In addition, some other studies mostly using transcript abundance or codon bias as a measure of gene expression show that highly expressed proteins and weakly expressed proteins differ in their amino acid usages (Lobry and Gautier 1994; Akashi and Gojobori 2002; Greenbaum et al. 2002; Akashi 2003).

However, mRNA concentrations and codon bias are not direct measures of protein abundance (Nie et al. 2006). Transcriptional responses may be not translated into quantitatively identical changes at the protein level (Griffin et al. 2002; Kolkman et al. 2006). In recent years, global proteomic data sets have become available. Using three recently published high-coverage data sets on protein abundance (Brockmann et al. 2007; Schmidt et al. 2007; Ishihama et al. 2008), we analyzed the relationships between protein abundance and atomic (carbon, nitrogen, and sulfur) contents in the side chains of amino acids and, also, between protein abundance and amino acid usage. Then we compared our results with previous studies (Lobry and Gautier 1994; Akashi and Gojobori 2002; Greenbaum et al. 2002; Akashi 2003; Elser et al. 2006).

As the carbon content of amino acids is highly correlated with the amount of energy for their synthesis (Spearman's $\rho > 0.7$, $p < 0.001$; see Materials and Methods for data sources), selection on both the economy of energy and the economy of atomic composition would reduce the carbon content of highly abundant proteins. Other studies have shown that selection to minimize the energetic cost of amino acids plays a significant role in shaping the composition of proteins (Craig and Weber 1998; Akashi and Gojobori 2002; Heizer et al. 2006; Swire 2007). Thus, we attempted to disentangle the effects of selection on the economy of energy and the economy of atomic composition by analyzing recently available proteomic data.

Materials and Methods

Estimation of Atomic Contents in Amino Acid Side Chains

The protein sequences of *S. cerevisiae*, *Schizosaccharomyces pombe*, and *E. coli* k12 were downloaded from *Saccharomyces* Genome Database (SGD; http://www.yeastgenome.org/) (Hong et al. 2008), GeneDB (http://www.genedb.org/) (Hertz-Fowler et al. 2004), and the Universal Protein Resource (UniProt; http://beta.uniprot.org/) (The UniProt Consortium 2008), respectively. These three organisms were selected based on the availability of their corresponding protein abundance data. The carbon content of each protein (C_C), specifically the number of carbon atoms per amino acid side chain, was calculated as

$$C_C = \left(\sum C_i \times F_i\right) \Big/ L$$

where C_i is the number of carbon atoms in the amino acid side group i, F_i is the frequency of the ith amino acid, and L is the total number of amino acids in the protein. Nitrogen content (C_N) and sulfur content (C_S) were calculated similarly, except that the initial residues (namely, methionine) were removed from all of the protein sequences in the calculation of sulfur content.

Estimation of the Energetic Cost of Amino Acids

The total energy required for the synthesis of each of the 20 amino acids was estimated in previous studies by considering the energy required to synthesize precursor molecules and that required to convert the precursors into the amino acids (Craig and Weber 1998; Akashi and Gojobori 2002; Wagner 2005; Heizer et al. 2006). In this study, we used the results of Wagner (2005), who calculated the energetic costs of the amino acids of *S. cerevisiae* under respiratory conditions, and those of Heizer et al. (2006), who calculated the energetic costs of the amino acids of *E. coli*. The energetic costs of the amino acids of *S. cerevisiae* were used in the calculation of the energetic costs of the amino acids of *S. pombe*, because of the metabolic similarity between these two species. The energetic cost of each

protein (C_E), specifically the total energetic cost of the building blocks, but not including that of translation or transcription, was calculated as

$$C_E = \left(\sum E_i \times F_i\right) \Big/ L$$

where E_i is the number of ATP molecules used in the synthesis of the amino acid i, F_i is the frequency of the ith amino acid, and L is the total number of amino acids in the protein.

For readers' convenience, we list the atom number in side chain and energy costs of each amino acid in Supplementary Table S1.

Protein Abundance, Turnover Rate, Translation Rate, and Functional Category

Global proteomic data sets, especially those on model organisms, have increased in size in recent years. We chose data sets with the consideration of their coverage and the date of their release, with the assumption that most recent data may be more accurate because of the advancement of analyzing techniques. The abundance values of 1456 *S. pombe* proteins and 1103 *E. coli* proteins were obtained from the supplementary information file of Schmidt et al. (2007) and Ishihama et al. (2008), respectively. The global protein abundance of *S. cerevisiae* has been more extensively studied. Brockmann et al. (2007) recently integrated four published data sets of *S. cerevisiae* (Futcher et al. 1999; Gygi et al. 1999; Ghaemmaghami et al. 2003; Liu et al. 2004) and released a larger data set, including the concentrations of 4152 proteins under standard conditions. They first normalized the data by nonlinear regression and then took the median as the reference value for each protein. We regard their data set as the best proteomic data set for *S. cerevisiae* under normal conditions.

Proteome-wide data on protein half-life in *S. cerevisiae* were obtained from the supplementary information file of Belle et al. (2006).

Protein translation rates were calculated by the following equation, obtained from Beyer et al. (2004) and Brockmann et al. (2007):

$$TR = \frac{\ln 2 \times PA}{HL}$$

where TR is the translation rate, PA is the protein abundance, and HL is the protein half-life.

Protein functional categories were integrated from the Gene Ontology (GO) (Ashburner et al. 2000) paths in SPIDer (Wu et al. 2006) and GO annotations of *S. cerevisiae*, *S. pombe*, and *E. coli*. In each proteome, high-expression proteins and low-expression proteins were defined as those above the median value of protein abundance and those lower than the median value, respectively. Functional categories that have at least five high-expression proteins and five low-expression proteins were retained. Totally, we obtained 27, 13, and 4 functional categories in *S. cerevisiae*, *S. pombe*, and *E. coli*, respectively.

Results and Discussion

Negative Correlation Between Protein Abundance and Atomic Contents

As shown in Fig. 1, the carbon contents of all three proteomes decreased significantly with an increase in protein abundance. The relationships between protein abundance and nitrogen and sulfur contents are much weaker, except for the nitrogen content in *S. cerevisiae*. Nonparametric correlation analysis revealed significant negative correlations between protein abundance and the contents of all three atoms in all three species analyzed (Table 1). Parametric correlation analysis gave similar results except that the correlation between nitrogen content and protein abundance is marginally significant in *S. pombe* and positive in *E. coli* (Supplementary Table S2). An interesting observation is that some of the most highly abundant proteins do not follow the global trend that nitrogen content decreases with an increase in protein abundance (Fig. 1). We then removed ribosomal proteins from the data sets and obtained considerably higher correlations between nitrogen content and protein abundance (Fig. 1, Table 1, and Supplementary Table S2). By contrast, after removal of the ribosomal proteins, there were no obvious changes in the correlations between protein abundance and carbon content; the correlations between protein abundance and sulfur content were still weak or even became nonsignificant (Table 1 and Supplementary Table S2). Consistently, the ribosomal proteins have significant higher contents of nitrogen and lower contents of carbon and sulfur than other proteins of each proteome (Supplementary Table S3). Furthermore, we found that the most preferred amino acids in ribosomal proteins (compared with other proteins in each proteome) are two nitrogen-containing basic amino acids: Arg and Lys. The frequencies of these two amino acids in ribosomal proteins are $\geq$1.48 times more than those in other proteins of each proteome (Supplementary Table S4). We suppose that nitrogen-containing basic amino acids are selected in ribosomal proteins because of the requirement to neutralize the acidic rRNA or the possible function of ribosomes to reserve resources for starved conditions (Kaplan and Apirion 1975; Kraft et al. 2008; Nakatogawa and Ohsumi 2008).

The side chains of eight amino acids (Arg, Asp, Glu, His, Lys, Trp, Cys, and Met) contain at least two of the

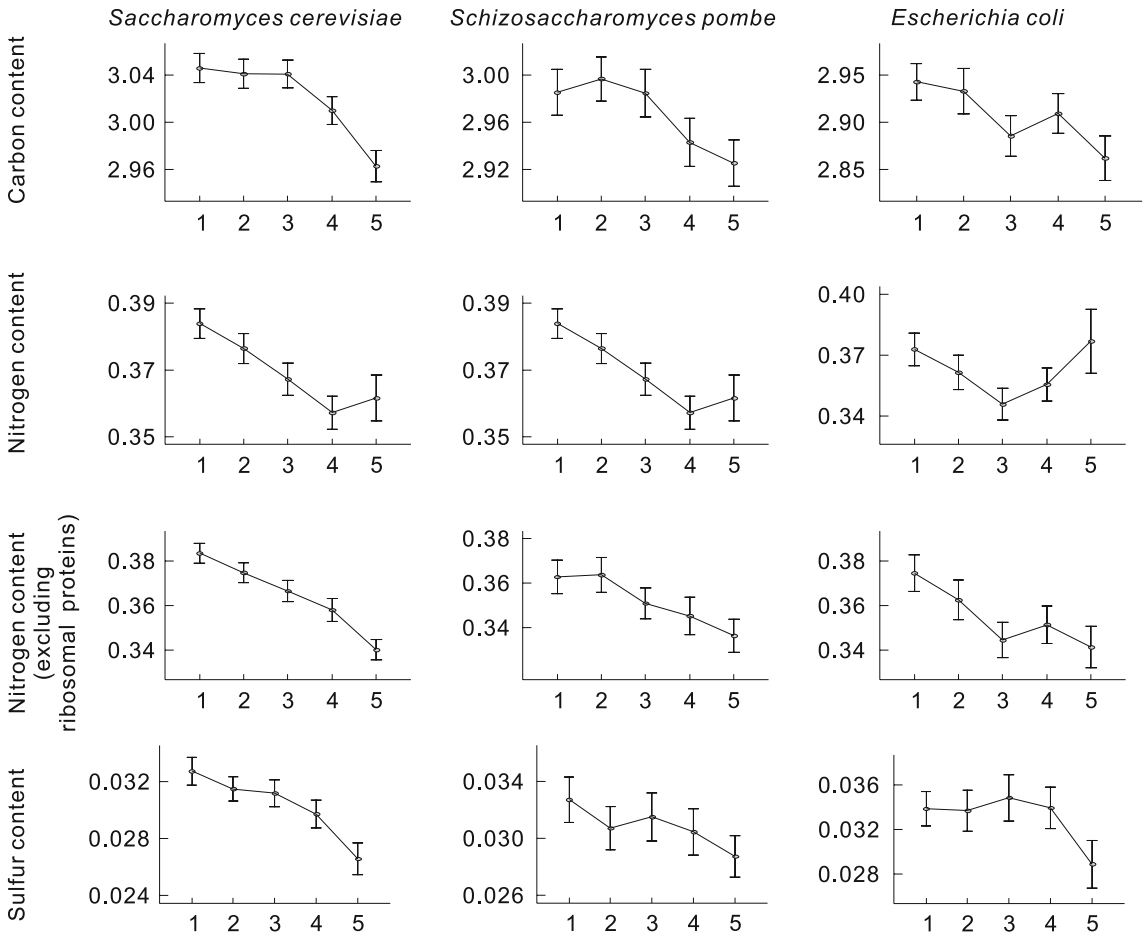

Fig. 1 Relationships between protein abundance and atomic contents per amino acid side chain. The *X* axis represents protein abundance. Proteins were binned into five categories according to their abundance, with abundance increasing from a value of 1 to a value of 5. The binning was just done for the figure, not for statistical analysis. We performed correlation analysis for the significance of global trends across all the data and the results are reported in Table 1 and Supplementary Table S2

three elements analyzed above. An abundant protein with a low content of arginine residues may be explained by selection for economy of carbon and nitrogen usage. We performed a partial correlation analysis after normalizing the data by a $\log_{10}$ transformation. As reported in Table 2, carbon contents are consistently negatively correlated with protein abundance after controlling for nitrogen and sulfur contents. However, only after excluding ribosomal proteins did the nitrogen content show consistently negative partial correlations with protein abundance. The sulfur contents are weakly correlated or uncorrelated with protein abundance after controlling for the contents of carbon and nitrogen in *S. pombe* and *E. coli*. Among the 20 common amino acids, there are 12 amino acids that do not contain nitrogen or sulfur in their side chains. Analyzing the relationship between protein abundance and the carbon contents of these 12 amino acids is analogous to partial correlation analysis controlling for nitrogen and sulfur contents. We found that protein abundance is negatively correlated with carbon content when considering these 12 amino acids only (Table 1 and Supplementary Table S2).

The results on sulfur content were inconsistent among species. We are inclined to think that selection for economy also minimizes the sulfur content of highly abundant proteins for the following reasons: (1) the appearance of sulfur-containing amino acids (Cys, and Met) is much lower in proteins than that of nitrogen- and carbon-containing amino acids, so the effect of random noise in sulfur content is expected to be higher than that in nitrogen or carbon content; (2) the sulfur content is negatively correlated with protein abundance at a level similar to the carbon and nitrogen contents in *S. cerevisiae*, the proteomic data on which have a much higher coverage than that on *S. pombe* and *E. coli* (Tables 1 and 2); and (3) low-sulfur-containing proteins have been observed to be preferentially expressed in microorganisms under conditions of sulfur starvation (Cuhel et al. 1981; Mazel and Marliere 1989; Fauchon et al. 2002; Boer et al. 2003).

Table 1 Spearman's rank correlations between protein abundance and atomic contents per amino acid side chain

		Saccharomyces cerevisiae			*Schizosaccharomyces pombe*			*Escherichia coli*		
		N	ρ	*p*	*N*	ρ	*p*	*N*	ρ	*p*
All proteomic data	Carbon content	4183	−0.15	10^{-6}	1465	−0.15	10^{-6}	1033	−0.16	10^{-6}
	Nitrogen content	4183	−0.21	10^{-6}	1465	−0.09	4×10^{-4}	1033	−0.08	0.02
	Sulfur content	4183	−0.16	10^{-6}	1465	−0.08	0.001	1033	−0.11	4×10^{-4}
Excluding ribosomal proteins	Carbon content	4019	−0.15	10^{-6}	1342	−0.16	10^{-6}	980	−0.17	10^{-6}
	Nitrogen content	4019	−0.27	10^{-6}	1342	−0.19	10^{-6}	980	−0.20	10^{-6}
	Sulfur content	4019	−0.12	10^{-6}	1342	−0.04	0.17	980	−0.05	0.12
Excluding amino acids that contain nitrogen or sulfur in their side chains	Carbon content	4183	−0.17	10^{-6}	1465	−0.18	10^{-6}	1033	−0.15	10^{-6}

Note: We performed Spearman's rank analysis because the values of protein abundance are not normally distributed. The values of atomic contents and the logarithm values of the protein abundance are almost normally distributed, so we also performed Pearson correlation analysis and the results are reported in Supplementary Table S2

Table 2 Partial correlations between protein abundance and atomic contents per amino acid side chain

		Saccharomyces cerevisiae	*Schizosaccharomyces Pombe*	*E. coli*
All proteomic data	Carbon content	−0.11**	−0.13**	−0.19**
	Nitrogen content	−0.16**	−0.05*	0.13*
	Sulfur content	−0.18**	−0.07*	−0.13*
Excluding ribosomal proteins	Carbon content	−0.10**	−0.12**	−0.11*
	Nitrogen content	−0.22**	−0.16**	−0.14**
	Sulfur content	−0.13**	NS	−0.07*

Note: The partial correlations between the content of one element and protein abundance were performed by controlling for the other two elements, among carbon, nitrogen, and sulfur. The data were normalized by $\log_{10}$ transformation. The numbers of proteins analyzed are identical to those in Table 1

* $P < 0.05$

** $P < 10^{-5}$

NS not significant

Furthermore, we examined the biases in amino acid usages associated with protein abundance. Generally, the frequencies of most amino acids in proteins are significantly (positively or negatively) correlated with the protein abundance, and this pattern is consistent across functional categories (Table 3). However, in only a few amino acids (e.g., Ala, Gly, Val, Ser, and Leu) were strong correlations observed. These observations coincide with previous studies (Akashi and Gojobori 2002; Akashi 2003). Consistent with predictions based on the economy of atoms and energy, Gly and Ala, with a low atom and energetic cost, are overrepresented in high-expression proteins, and Leu, His, and Trp, with a high atom and energetic cost, are overrepresented in low-expression proteins (Table 3 and Supplementary Table S1). However, there are also exceptions that cannot be ignored (Table 3 and Supplementary Table S1). For example, Ser, which has a low atom and energetic cost, are also overrepresented in low-expression proteins. The biased usages of some amino acids in high-expression or low-expression proteins we observed are generally consistent with previous studies that used transcript abundance or codon bias as a measure of gene expression (Akashi and Gojobori 2002; Greenbaum et al. 2002; Akashi 2003). Together with other evidence, Akashi (2003) proposed that the amino acid composition of proteins may be optimized for the speed and accuracy of translation. We suppose that the translational selection and selection for economy may operate on the evolution of amino acid composition in concurrence.

We have confirmed the observations of Elser (2006) in unicellular organisms. At least in plants and unicellular organisms, the amino acid sequences of highly abundant proteins have to compromise between optimizing for their biological function and reducing the consumption of limiting resources. By contrast, the amino acid sequences of scarce proteins are more likely to be optimized for their

Table 3 Protein abundance and amino acid composition

	Saccharomyces cerevisiae		*Schizosaccharomyces pombe*		*E. coli*	
	r	OR	*r*	OR	*r*	OR
Ala	0.42***	1.27***	0.32***	1.22***	0.11**	1.04*
Gly	0.31***	1.19***	0.32***	1.17***	0.05*	1.06**
Val	0.31***	1.11***	0.29***	1.09**	0.23***	1.08**
Glu	0.07***	1.08***	−0.02	1.02	0.11**	1.03
Lys	0.07***	1.01	0.11**	1.07**	0.47***	1.26***
Asp	0.01	1.06***	−0.05**	1.04	−0.10**	1.07**
Arg	0.005	0.95***	0.04	0.96*	0.13**	0.91**
Thr	−0.03**	0.99	0.02	1.01	0.08**	1.02
Trp	−0.05**	0.89***	−0.11**	0.81**	−0.29***	0.73***
Pro	−0.06**	0.97**	−0.06**	1.03	−0.28***	0.85***
Tyr	−0.06**	0.96**	−0.03	0.90**	−0.14**	1.01
Gln	−0.07**	0.96**	−0.18***	0.88***	−0.19***	0.85***
Ile	−0.07***	0.97**	0.01	0.98	0.01	1.13***
Phe	−0.08***	0.96**	−0.08**	0.88***	−0.05	1.07**
Met	−0.09***	0.98	0.02	1.04	0.03	1.05
Cys	−0.10***	0.92***	−0.11**	0.87**	−0.15***	0.86**
His	−0.10***	0.91***	−0.11**	0.85***	−0.08**	0.87**
Leu	−0.14***	0.96***	−0.16***	0.94**	−0.35***	0.87***
Ser	−0.28***	0.87***	−0.30***	0.93***	−0.14**	0.93**
Asn	−0.30***	0.87***	−0.21***	0.91**	−0.01	1.03

Note: All 4183 proteins in *S. cerevisiae*, 1465 proteins in *S. pombe*, and 1033 proteins in *E. coli* were analyzed. We show the Pearson's correlations (r) between protein abundance ($\log_{10}$ transformed) and amino acid usage, and Cochran-Mantel-Haenszel chi-squared test (OR) for the biased usage of each amino acid in high-/low-expression proteins across functional categories. OR > 1 means that the amino acid is overrepresented in high-expression proteins and OR < 1 means that the amino acid is overrepresented in low-expression proteins. * $P < 0.1$; ** $P < 0.05$; *** $P < 10^{-5}$

biological functions. While this paper is under revision according to referees' comments, a paper was published online showing that single amino acid replacements that increase material costs can be subject to natural selection (Bragg and Wagner 2009).

Economy in Energy or Carbon?

Besides nutritional elements, energetic cost is another source of metabolic constraints. A conceivable idea is that highly expressed genes should avoid using expensive amino acids because of the selective force to minimize the energetic cost of gene expression. In several microorganisms, including *E. coli* and *S. cerevisiae*, highly abundant proteins were found to preferentially use cheap (in terms of energy) amino acids (Craig and Weber 1998; Akashi and Gojobori 2002; Heizer et al. 2006). Similarly, negative correlations between protein abundance and energy cost were observed by analyzing the currently available proteomic data sets we collected (Supplementary Table S5 and Fig. S1).

Most cellular energy is stored in carbon-containing organic compounds. An amino acid with a high number of carbon atoms in its side chain generally requires relatively more energy for its synthesis (Craig and Weber 1998; Akashi and Gojobori 2002; Wagner 2005; Heizer et al. 2006). The carbon content and mean energetic cost were shown to be highly correlated among yeast (*S. cerevisiae*) proteins (Bragg and Wagner 2007). The question can therefore be asked: Is there selection for the economy of carbon content, energetic cost, or both in highly abundant proteins?

To answer this, we discarded the eight amino acids (Arg, Asp, Glu, His, Lys, Trp, Cys, and Met) that contain nitrogen or sulfur in their side chains from the proteomes to exclude a potential influence of nitrogen and sulfur. Then we analyzed the relationship between the energy cost and the carbon content of the other 12 amino acids in proteomes. The negative correlations between protein abundance and energy cost were still significant (Supplementary Table S5 and Fig. S1). The carbon contents were negatively correlated with protein abundance after controlling for energetic cost (Table 4). However, there appeared to be significantly positive correlations between protein abundance and energetic cost after controlling for carbon content (Table 4). Partial correlation analysis including all

Table 4 Partial correlations of protein abundance with energetic cost or carbon content per amino acid side chain[a]

	N	Energetic cost (controlling for carbon content)		Carbon content (controlling for energetic cost)	
		r	*p*	*R*	*p*
Saccharomyces cerevisiae	4183	0.21	4×10^{-42}	−0.26	5×10^{-68}
Schizosaccharomyces pombe	1465	0.27	6×10^{-26}	−0.32	10^{-35}
E. coli	1033	0.19	6×10^{-10}	−0.25	3×10^{-16}

Note: The energetic cost was calculated after excluding the eight amino acids that contain nitrogen or sulfur in their side chains: Arg, Asp, Glu, His, Lys, Trp, Cys, and Met. The data on protein abundance were normalized by $\log_{10}$ transformation

the 20 common amino acids showed similar trends but much weaker correlations in *S. cerevisiae* and *S. pombe* (Supplementary Table S6). In *E. coli*, both of the significant partial correlations disappear when all the 20 common amino acids are included in partial correlation analysis (Supplementary Table S6). Considering the results overall, we are inclined to suggest that the partial correlation analysis including all 20 amino acids are not robust, perhaps because of the low coverage of the proteomic data. Although we do not know why highly abundant proteins prefer expensive (in terms of energy) amino acids, at least, the lower carbon content of highly abundant proteins should not be attributed to potential selection to minimize the energetic cost of gene expression. The observation that highly abundant proteins preferentially use cheap (in terms of energy) amino acids (Craig and Weber 1998; Akashi and Gojobori 2002; Heizer et al. 2006) seems to be superficial.

Effects of Protein Turnover Rates and Translation Rates on Atomic Composition

Besides abundance, protein turnover rate may also affect the atomic composition of proteins. The cost of using limiting elements in unstable proteins is expected to be lower than that in highly stable proteins, because the limiting atoms are recycled rapidly. However, in *S. cerevisiae*, we found that protein half-lives are only weakly correlated with nitrogen and sulfur contents, and not significantly correlated with carbon content (Spearman's rank correlations: $n = 3103$; nitrogen content, $\rho = -0.07$, $p = 2 \times 10^{-4}$; sulfur content, $\rho = -0.03$, $p = 0.05$). As proteins produced in large quantities are generally stable (Belle et al. 2006), we performed partial correlation analysis after normalizing the data by a $\log_{10}$ transformation. Neither carbon, nitrogen, nor sulfur content was significantly correlated with protein half-lives ($p > 0.05$). The half-lives of 95% of the proteins in *S. cerevisiae* fall between 2 and 318 min (Belle et al. 2006). Compared with the variation in protein abundance per cell (Ghaemmaghami et al. 2003; Ishihama et al. 2008), the magnitude of difference in half-lives between highly stable proteins and less stable proteins is much smaller. We conjectured that the effects of protein turnover rate on the evolution of atomic compositions are too weak to provide effective selection pressure.

Bragg and Wagner (2007) recently found that yeast proteins up-regulated under conditions of carbon starvation do not have a lower carbon content than the rest of the proteome. Similar to the effects of protein turnover rate, the magnitude of up-regulation may not be large enough for proteins with particular atomic compositions to be effectively selected.

By integrating the data on protein abundance and protein turnover rate, we estimated the protein translation rates of *S. cerevisiae* and found that they are significantly correlated with the atomic contents and the usage of most amino acids (Supplementary Tables S7 and S8). Overall, protein translation rate and protein abundance are similar in their correlations with atomic composition and amino acid composition. Further evidence is required to distinguish between selection for economy in atoms and selection for translation efficiency.

Acknowledgments We thank anonymous reviewers for very valuable comments on the manuscript and Xiaomei Wu, Jie Guo, and Yi-Fei Huang for their help. This study was supported by Program NCET-07-0094 and Beijing Normal University.

References

Akashi H (2003) Translational selection and yeast proteome evolution. Genetics 164:1291–1303

Akashi H, Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proc Natl Acad Sci USA 99:3695–3700

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genet 25:25–29

Baudouin-Cornu P, Surdin-Kerjan Y, Marliere P, Thomas D (2001) Molecular evolution of protein atomic composition. Science 293:297–300

Baudouin-Cornu P, Schuerer K, Marliere P, Thomas D (2004) Intimate evolution of proteins—proteome atomic content correlates with genome base composition. J Biol Chem 279:5421–5428

Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. Proc Natl Acad Sci USA 103:13004–13009

Beyer A, Hollunder J, Nasheuer HP, Wilhelm T (2004) Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. Mol Cell Proteomics 3:1083–1092

Boer VM, de Winde JH, Pronk JT, Piper MDW (2003) The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. J Biol Chem 278:3265–3274

Bragg JG, Wagner A (2007) Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. Proc R Soc B 274:1063–1070

Bragg JG, Wagner A (2009) Protein material costs: single atoms can make an evolutionary difference. Trends Genet 25:5–8. (doi: 10.1016/j.tig.2008.10.007)

Brockmann R, Beyer A, Heinisch JJ, Wilhelm T (2007) Posttranscriptional expression regulation: what determines translation rates? PLoS Comput Biol 3:531–539

Craig CL, Weber RS (1998) Selection costs of amino acid substitutions in ColE1 and ColIa gene clusters harbored by *Escherichia coli*. Mol Biol Evol 15:774–776

Cuhel RL, Taylor CD, Jannasch HW (1981) Assimilatory sulfur metabolism in marine microorganisms: Sulfur metabolism, growth, and protein synthesis of *Pseudomonas halodurans* and *Alteromonas luteo-violaceus* during sulfate limitation. Arch Microbiol 130:1–7

Elser JJ, Fagan WF, Subramanian S, Kumar S (2006) Signatures of ecological resource availability in the animal and plant proteomes. Mol Biol Evol 23:1946–1951

Fauchon M, Lagniel G, Aude JC, Lombardia L, Soularue P, Petat C, Marguerie G, Sentenac A, Werner M, Labarre J (2002) Sulfur sparing in the yeast proteome in response to sulfur demand. Mol Cell 9:713–723

Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JI (1999) A sampling of the yeast proteome. Mol Cell Biol 19:7357–7368

Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. Nature 425:737–741

Greenbaum D, Jansen R, Gerstein M (2002) Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. Bioinformatics 18:585–596

Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. Mol Cell Proteomics 1:323–333

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19:1720–1730

Heizer EM Jr, Raiford DW, Raymer ML, Doom TE, Miller RV, Krane DE (2006) Amino acid cost and codon-usage biases in 6 prokaryotic genomes: a whole-genome analysis. Mol Biol Evol 23:1670–1680

Hertz-Fowler C, Peacock CS, Wood V, Aslett M, Kerhornou A, Mooney P, Tivey A, Berriman M, Hall N, Rutherford K, Parkhill J, Ivens AC, Rajandream MA, Barrell B (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. Nucleic Acids Res. 32:D339–D343

Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Krieger CJ, Livstone MS, Miyasato SR, Nash RS, Oughtred R, Skrzypek MS, Weng S, Wong ED, Zhu KK, Dolinski K, Botstein D, Cherry JM (2008) Gene Ontology annotations at SGD: new data sources and annotation methods. Nucleic Acids Res 36:D577–D581

Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner M, Frishman D (2008) Protein abundance profiling of the *Escherichia coli* cytosol. BMC Genomics 9:102

Kaplan R, Apirion D (1975) The fate of ribosomes in *Escherichia coli* cells starved for a carbon source. J Biol Chem 250:1854–1863

Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MMA, Pronk JT, Slijper M, Heck AJR (2006) Proteome analysis of yeast response to various nutrient limitations. Mol Syst Biol 2:2006.0026

Kraft C, Deplazes A, Sohrmann M, Peter M (2008) Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. Nat Cell Biol 10:602–610

Liu HB, Sadygov RG, Yates JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76:4193–4201

Lobry JR, Gautier C (1994) Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. Nucleic Acids Res 22:3174–3180

Lv J, Li N, Niu D-K (2008) Association between the availability of environmental resources and the atomic composition of organismal proteomes: evidence from *Prochlorococcus* strains living at different depths. Biochem Biophys Res Commun 375:241–246

Mazel D, Marliere P (1989) Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. Nature 341:245–248

Nakatogawa H, Ohsumi Y (2008) Starved cells eat ribosomes. Nature Cell Biol 10:505–507

Nie L, Wu G, Zhang W (2006) Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations. Biochem Biophys Res Commun 339:603–610

Pardee AB (1966) Purification and properties of a sulfate-binding protein from *Salmonella typhimurium*. J Biol Chem 241:5886–5892

Schmidt MW, Houseman A, Ivanov AR, Wolf DA (2007) Comparative proteomic and transcriptomic profiling of the fission yeast *Schizosaccharomyces pombe*. Mol Syst Biol 3:79

Swire J (2007) Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. J Mol Evol 64:558–571

The UniProt Consortium (2008) The Universal Protein Resource (UniProt). Nucleic Acids Res 36:D190–D195

Wagner A (2005) Energy constraints on the evolution of gene expression. Mol Biol Evol 22:1365–1374

Wu XM, Zhu L, Guo J, Fu C, Zhou HJ, Dong D, Li ZB, Zhang DY, Lin K (2006) SPIDer: *Saccharomyces* protein-protein interaction database. BMC Bioinformatics 7:S16