Paper Evaluation: Metabolic efficiency and amino acid composition...

What is the hypothesis of the paper?
There is a difference in the amount of energy in terms of high-energy phosphate bonds needed to synthesize the various amino acids (12-74 phosphate bonds per molecule). If *E. coli* and *B. subtilis* synthesize a large portion of the amino acids that they use and the required energy for synthesis is diverted from that needed to survive or reproduce then the resulting differences in energy requirements and the availability of energy should exert an evolutionary pressure on the distribution of amino acids in their proteomes. In other words, natural selection favors metabolic efficiency which is reflected in the amino acids distribution.

What methodology is used?
The methodology is based on calculating the 1) biosynthetic cost of amino acids, 2) determining the MCU of each gene and 3) the amino acid complement of each protein.

Cost of Biosynthesis
This biosynthetic amino acid cost includes the calculation of the cost of precursor metabolites in different energy source environments, i.e., glucose, acetate and malate. Next the cost of amino acid biosynthesis from the precursors was determined.

Protein Coding Genes
ORF data was used to specify coding regions and gene functionality data was used to classify genes. Excluded from the list of genes to be analyzed were: 1) ORF’s that were potentially horizontally transferred, 2) genes that were suspected of being phage-related, and 3) paralogous genes (only one of the pairs was included).

Major Codons
Major codons, as identified by Kanaya *et al.* for *E. coli* and *B. subtilis* were use to determine major codon usage as (# major codons)/(# major codons + # minor codons) for each gene comprised of at least 100 codons. Note: major codons are defined as codons that contribute to the major trend in codon usage. Note: MCU is a measure of codon bias.

Amino Acid Composition Analysis
Correlation between MCU (codon bias) and amino acid composition were analyzed with functional gene classes and proteome-wide. In the case of functional gene class analysis, each class was divided into two subclasses, those above the MCU median and those below. The Mantel-Haenzel procedure was used to determine the probability amino acid distribution departs from equal usage. Excluded from analysis were 1) categories with fewer than ten genes, 2) proteins similar to those of other organisms, 3) proteins similar to those of this organism, 3) “miscellaneous”, 4) “no similarity” (for *B. subtilis*), 5) “hypothetical, unclassified, unknown” (for *E. coli*).
What is the relevant background?

Amino acid composition is determined principally by mutations and functional requirements of protein structure. Critical locations in proteins are more highly constrained in terms of amino acid composition. The amino acid composition of less constrained sections more strongly reflect mutational bias.

Is there some relevant background that the author(s) seem unaware of?

No, the background seems complete.

What are the results?

Metabolic Costs of Amino Acid Biosynthesis

Metabolic costs of amino acids are presented in table 1. The costs range from 11.7 to 74.3 and are calculated as:

\[ \text{# of phosphate bonds} + \text{# of hydrogen atoms} \]

lost to the metabolic pathway by the diversion of the precursor metabolite.

Figure 1 illustrates the pathways involved in diverting precursor metabolites from the metabolic pathway to produce amino acids. In general amino acids resulting from diverting metabolites early in the metabolic pathway entail higher cost. The cost calculations for different growth substrates were highly correlated > 0.9. Average costs were used for subsequent analysis of correlation between cost and codon usage bias.

Correlation Between Metabolic Cost and Synonymous Codon Usage Bias

For each of the 3397 genes in *E. coli* and 3055 genes in *B. subtilis*, average cost per amino acid and MCU were calculated. Next correlation between cost and codon bias were determined. The result is a statistically significant negative correlation:

\[ n = 3397, r_s = -0.240, Z = 14.43, P < 10^{-5} \text{ for } E. \text{ coli} \]

\[ n = 3055, r_s = -0.383, Z = 22.92, P < 10^{-5} \text{ for } B. \text{ subtilis} \]

The authors are using MCU as a surrogate for gene expression level. High MCU corresponds to high expression level. Figure 2 shows the correlation of cost and MCU for both proteomes. High MCU correlates with low metabolic cost. From this the authors infer high expression level correlates with low metabolic cost. The authors note that the fitness of an amino acid at a particular site in a protein may outweigh its metabolic cost but that by considering large numbers of codons this influence of this effect should be reduced. The authors note that A/T nucleotide bias in *B. subtilis* and observe that A/T-rich codons tend to encode more costly amino acids.
Separate Analysis of Metabolic Costs and Codon Bias for H and ~P

Separate costs evaluations for ~P and H were obtained to show that the interpretation of the relation between amino acid distribution and gene expression level is not confounded using a combination of these costs.

Metabolic Costs and Codon Usage Bias Within Functional Categories of Proteins

Separate analysis of functional categories and MCU were carried out (Table 2 and Table 3) to control for the possibility that different functional categories could differ in the relationship between amino acid composition and expression level. Table 2 and Table 3 show the same negative correlation between cost and MCU on a per category basis.

Metabolic Costs and Codon Usage Bias Within Physicochemical Classes of Amino Acids

Next the authors investigated cost versus codon bias with respect to three equivalence classes of amino acids. The three classes are:

1. “internal” amino acids which have hydrophobic groups: Phe, Leu, Ile, Met, and Val.
2. “external” amino acids which have hydrophilic groups: His, Arg, Lys, Gln, Glu, Asn, and Asp.
3. “ambivalent” amino acids which can function in either category: Trp, Tyr, Cys, Ala, Ser, Gly, Pro, and Thr.

The results of this analysis for the two proteomes are shown in Figure 3. These results also demonstrate a negative correlation between cost and codon bias.

Gene Expression and Amino Acid Composition

Table 4 lists the change in abundance of amino acids as a function of MCU for the whole proteome and within functional categories. Figure 4 shows the associations between individual amino acids and MCU. In this analysis, amino acid usage in each functional category is compared with high and low MCU classes. A Mantel-Haenszel test was used to test amino acid increase or decrease relative to the high and low MCU classes. The results indicate that amino acid abundance changes within a functional category as a function of the estimated translation rate.

Figure 4 shows that costly amino acids such as Trp, Phe, His, Cys, and Leu decrease in abundance in highly expressed genes. The authors also point out that AT-rich codons and GC-rich codons do not exhibit common trends related to gene expression.

To what extent do the results support the hypothesis?

The results demonstrate statistically significant negative correlation between estimated amino acid costs and MCU. The authors present a case for accepting MCU as being indicative of expression level.
What questions are still unresolved?

How can expression rate be determined?

Why are there still unresolved questions?

The authors have not bothered to perform the type of experiments such as microarray that would definitively characterize expression levels.

What new questions have been raised?

Figure 4 indicates that Lysine which is relatively metabolically expensive shows a statistically significant increase with MCU. This contradicts the overall trend. Also, the results do not show significant decreases for costly amino acids in *E. coli* relative to *B. subtilis*. Why is this?

What are the weaknesses of the paper?

Method?

The approach depends on a relationship between codon bias and translation rate. This a general weakness since it requires an argument that codon bias is highly positively correlated with translation rate. This approach is therefore not applicable to organisms which do not exhibit major codon preference.

Results?

There are no quantitative results demonstrating a high correlation between codon bias and translation rate.