Effect of Codon Message on Xylanase Thermal Activity†§

Received for publication, November 24, 2011, and in revised form, April 25, 2012 Published, JBC Papers in Press, June 15, 2012, DOI 10.1074/jbc.M111.327577

Liangwei Liu†, Linmin Wang§, Zhang Zhang§, Suya Wang†, and Hongge Chen§

From the †Life Science College, Henan Agricultural University, Zhengzhou 450002, the §Key Laboratory of Enzyme Engineering of Agricultural Microbiology, Ministry of Agriculture, Henan Agricultural University, Zhengzhou 450002, and the ‡Department of Food Science and Engineering, Nanjing University of Economics, Nanjing 210003, China

Background: Because it is known for degeneracy, the effect of codon on enzyme thermal property is elusive.

Results: Three purine-rich codons correlate positively with xylanase T_{opt}, and two pyridine-rich codons correlate negatively. Two positive codons have A-ends. One negative codon has a C-end.

Conclusion: Codons have effects on enzyme thermal property.

Significance: The effect of codon message is lost when thermal property is analyzed at the residual level.

Because the genetic codon is known for degeneracy, its effect on enzyme thermal property is seldom investigated. A dataset was constructed for GH10 xylanase coding sequences and optimal temperatures for activity (T_{opt}). Codon contents and relative synonymous codon usages were calculated and respectively correlated with the enzyme T_{opt} values, which were used to describe the xylanase thermostable tendencies without dividing them into two thermophilic and mesophilic groups. After analyses of codon content and relative synonymous codon usages were checked by the Bonferroni correction, we found five codons, with three (AUA, AGA, and AGG) correlating positively and two (CGU and AGC) correlating negatively with the T_{opt} value. The three positive codons are purine-rich codons, and the two negative codons have A-ends. The two negative codons are pyridine-rich codons, and one has a C-end. Comparable with the codon C- and A-ending features, C- and A-content within mRNA correlated negatively and positively with the T_{opt} value, respectively. Thereby, codons have effects on enzyme thermal property. When the issue is analyzed at the residual level, the effect of codon message is lost. The codons relating to enzyme thermal property are selected by thermostable force at nucleotide level.

Because thermostable enzymes are economic for biotechnology, it is important to reveal enzyme features adapting to higher temperatures at all levels, such as gene, amino acid residue, structure, etc. (1–7). Thermophilic enzymes have many features at the residual and structural levels, such as a larger number of charged residues, an increase of hydrogen bonds and intramolecular hydrophobic packing, etc. (4–7). The investigations provide profound insights into enzyme thermostable mechanism. However, when the issue is analyzed at the residual level, message decreases from 61 to 20 dimensions because the 20 residues are encoded by the 61 codons. Several lines of evidence indicate that codons affect the protein folding and structural formation. For example, different aminoacylated tRNAs of the same residue had different binding strengths to the ribosome (1). Thermophilic mRNA has more polyadenine tracts than mesophilic ones (2). The codon bias usage led to α-helix being translated faster than β-strand and irregular coil (8). A synonymous codon mutation increased the methyltransferase mRNA stability, and therefore, decreased its expression, and ultimately, resulted in illness (9). An organism optimal growth temperature (OGT)² was found relating to the codon bias usage and the GC%³ within genome or RNA (10–14). Moreover, thermophiles can be discriminated from mesophiles by protein residual compositions (15–17), which are also selected at the nucleotide level (18–20). Thereby, codon effect on enzyme thermal property needs to be investigated.

Thermophiles and enzyme data are important for analysis. In the investigation of enzyme thermostable features, data are often selected from different families and are divided into thermophilic and mesophilic groups according to whether or not the OGT values higher than 50 °C. However, the cutoff value is arbitrarily set to some extent (21). Additionally, there is a much weaker correlation between OGT value and enzyme thermostability (22). In particular, OGT values are unspecific when used to describe thermal properties of enzymes encoded by lateral transfer genes, which occur very often among organisms (23–27). Unlike OGT, an enzyme optimal reaction temperature for activity (T_{opt}) specifically describes its thermal activity, and the T_{opt} is specific to exclude interference of lateral gene transfer. Moreover, the T_{opt} value of an enzyme correlates well with its thermostability (28, 29). Because different structural proteins employ different strategies for thermostability (30, 31), enzymes selected from different families interfere with thermal analysis. The interference can be eliminated by selecting homologous enzymes from the same family (22).

The abbreviations used are: OGT, organism optimal growth temperature; CDS, coding sequence; T_{opt}, optimal temperature for activity; C_{i}, content of each codon within a CDS; N_{i}, content of each nucleotide within a CDS; RSCU, relative synonymous codon usage.

2 The abbreviations used are: OGT, organism optimal growth temperature; CDS, coding sequence; T_{opt}, optimal temperature for activity; C_{i}, content of each codon within a CDS; N_{i}, content of each nucleotide within a CDS; RSCU, relative synonymous codon usage.

3 Throughout this study, A%, G%, C%, AG%, UC%, and GC% indicate percentages of adenine, guanine, cytosine, adenine plus guanine, uracil plus cytosine, and guanine plus cytosine within mRNA, separately.
**Codon Affects Enzyme Thermal Activity**

GH10 xylanase exhibits \((4/8)\) structure, and the typical fold forms \(\sim 10\%\) enzymes, including \(\alpha\)-amylase, glucanase, \(\beta\)-mananase, triosephosphate isomerase, etc. (28). GH10 xylanase is used as a model to analyze codon effect on enzyme thermal property because \((4/8)\) can provide more general information. The dataset is constructed for GH10 xylanase coding sequences (CDSs) and \(T_{opt}\) values. Codon content and relative synonymous codon usage (RSCU) are calculated within the CDSs and are respectively linearly regressed with the \(T_{opt}\) values, which can be used to describe the xylanase thermophilic tendencies without dividing them into two clear-cut groups of thermophile and mesophile. This study can provide more insights into the enzyme thermophilic mechanism.

**EXPERIMENTAL PROCEDURES**

*Data Construction*—A dataset was constructed by retrieving GH10 xylanase CDSs from GenBank\textsuperscript{Tm} (Release 165.0). To eliminate sequence redundancy, the CDSs were categorized and listed according to the xylanase annotations in Swiss-Prot (Release 55.4). From published literature, the \(T_{opt}\) values were taken to describe xylanase thermophilic tendencies. \(T_{opt}\) values were found for 38 CDSs: O13436, O30427, O30700, O32374, P29417, P33559, P36917, P40942, P40943, P40944, P45703, P07986, P10478, P14768, P23360, P23556, P26223, P26514, P29417, P33559, P36917, P40942, P40943, P45703, P51584, P56588, P74912, Q00177, Q12603, Q2PGV8, Q5NUI5, Q60037, Q60042, Q60043, Q60044, Q76BV3, and Q8GJ37 (supplemental Table 1). The sequence identities range from 6.9% to 88.7%, and the sequence identities range from 6.9% to 88.7%.

*Method 1*—Counts and contents were calculated for the 61 codons and the 4 nucleotides within each CDS with our own designed programs (Compaq Visual FORTRAN 6). Codon content \((C_i)\) was defined as a ratio of each codon relative to the sum of codons within a CDS. Nucleotide content \((N_j)\) was defined as a ratio of each nucleotide relative to sum of nucleotides within a CDS.

\[
C_i = \frac{\text{Obsi}}{\sum_{i=1}^{61} \text{obsi}} \quad \text{(Eq. 1)}
\]

where \(C_i, \text{Obsi}, \) and \(\sum_{i=1}^{61} \text{obsi}\) indicate content, observation frequency of each codon, and sum of codons within a CDS.

\[
N_j = \frac{n_j}{\sum_{j=1}^{4} n_j} \quad \text{(Eq. 2)}
\]

where \(i\) indicates A, U, C, and G, respectively, and contents of A+G, U+C, A+C, U+G, A+U, and G+C were respectively calculated. \(N_j\) and \(\sum_{j=1}^{4} n_j\) indicate the content of each nucleotide and the sum of nucleotides within a CDS.

*Method 2*—RSCUi was defined as a ratio of each codon relative to expected frequency if all synonymous codons are used equally (32).

\[
\text{RSCUi} = \frac{\text{Obsi}}{\text{Expi}} \quad \text{(Eq. 3)}
\]

\[
\text{Expi} = \sum_{i=1}^{59} \frac{\text{aai}}{\sum_{i=1}^{59} \text{syni}} \quad \text{(Eq. 4)}
\]

There are 59 RSCUi values because Trp and Met have only one codon. Obsi is the observational frequency of each codon within a CDS, \(\Sigma_{aai}\) is the sum of residues within a xylanase, and \(\Sigma_{i=1}^{59} \text{syni}\) indicates the number of synonymous codon encoding for the same residue.

The variables \(C_i, \text{RSCUi}, \) and \(N_j\) were respectively linearly correlated with the \(T_{opt}\) values (Origin, Version 6.1). Cutoff of the \(p\) value was set for codons as \(p < 0.15\), and for nucleotides, it was set as \(p < 0.1\) because codon content was low. Effects of codons and nucleotides on enzyme thermal property were assessed by their correlation coefficients with \(T_{opt}\).

The analysis was based on multiple hypothesis testing that related to 61 codons. Thus, false positive codons found by linear analysis had to be excluded. Bonferroni correction is a standard statistics method in multiple analysis (33, 34). Thus, the related codons found above were corrected by the Bonferroni method. According to the method, the codons were ranked based on their \(p\) values. The significance was set as \(p < 0.15\) in multiple analysis, and these values were divided by the number of multiple factors to exclude false positive codons. After that, the false positive rate was analyzed.

**RESULTS**

*Codon Relating to Xylanase Thermal Activity*—23,005 codons were analyzed using codon content, and 15 codons were found relating to the \(T_{opt}\) value. However, after the Bonferroni correction analysis, five codons were found to reach a significant level, and the others were false positives (supplemental Table 2). The 23,005 codons were also analyzed using RSCU because it was usually used in codon analysis (32, 35–37), and 15 codons were found relating to the \(T_{opt}\) value. However, four codons were found to reach a significant level after Bonferroni correction analysis, and the others were false positives (supplemental Table 3). RSCU has no connection with codon content, and therefore, no straight connection with the \(T_{opt}\) value. However, the RSCU found four codons that found by codon content analysis, showing that codon content is fine for finding related codons. Thus, two (CGU and AGC) of the five codons correlated negatively, and three (AUA, AGA, and AGG) of the five codons correlated positively with the xylanase \(T_{opt}\) value (Table 1). Among the three codons of arginine, AGA, AGG, and CGU,
the first two correlate positively (with a ratio of 1.8:1), and the last one correlates negatively with the $T_{\text{opt}}$ value (Fig. 1). One codon of isoleucine (AUA) correlates positively with the $T_{\text{opt}}$ value (Table 1). After that, the five codons were respectively counted within the three genes having the lowest $T_{\text{opt}}$ values and within the three genes having highest $T_{\text{opt}}$ values. The ratios are 15:2, 10:1, and 10:3 for the contents of positive codons, AUA, AGA, and AGG, within thermophilic genes relative to mesophilic genes, respectively. In contrast, the ratios are 0:8 and 1:3 for the contents of negative codons, CGU and AGC, within thermophilic genes relative to mesophilic genes, respectively (Table 2).

Consistent with our result, prior genome analyses found that thermopiles had decreased GCN and increased AGR (36–39), although those studies had different purposes from this study. Two codons of arginine (AGA and AGG) and one of isoleucine (ATH) were shown to be the characteristic codon usage pattern of thermophiles. In addition, three codons of arginine, AGA, and AGG, within thermophilic genes relative to mesophilic genes, respectively. In contrast, the ratios are 0:8 and 1:3 for the contents of negative codons, CGU and AGC, within thermophilic genes relative to mesophilic genes, respectively. In general, all the other A- and C-ending codons correlate positively and negatively with the $T_{\text{opt}}$ value, although the $p$ values are significantly higher than the set value. The C- and A-ending feature of the codons shows that codon effect concentrates mainly on the 3rd nucleotide. Thus, in addition to keeping translation fidelity (40), which is well known to cause the degeneracy of the genetic codon, codons also contain effects of message on xylanase thermal activity.

The five related codon contents within the six selected genes

<table>
<thead>
<tr>
<th>CDS</th>
<th>$T_{\text{opt}}$</th>
<th>Positive codons</th>
<th>Negative codons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AUA (Ile)</td>
<td>AGA (Arg)</td>
</tr>
<tr>
<td>Q5NU15</td>
<td>40</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>P29417</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>O59859</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Q12603</td>
<td>85</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>Q60037</td>
<td>90</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Q60042</td>
<td>102</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Ratio</td>
<td>15:2</td>
<td>10:1</td>
<td>10:3</td>
</tr>
</tbody>
</table>

To get more general information, the codons were analyzed for nucleotide constituents. The three positive ones are purine-rich, and two of them have A-ends. The two negative codons are pyridine-rich, and one has a C-end. To examine whether or not the C- and A-ends are general features of the codons related to the $T_{\text{opt}}$ value, all other A- and C-ending codons were collected and analyzed (supplemental Tables 2 and 3). In general, all the other A- and C-ending codons correlate positively and negatively with the $T_{\text{opt}}$ value, although the $p$ values are significantly higher than the set value. The C- and A-ending feature of the codons shows that codon effect concentrates mainly on the 3rd nucleotide. Thus, in addition to keeping translation fidelity (40), which is well known to cause the degeneracy of the genetic codon, codons also contain effects of message on xylanase thermal activity.

To get complete information about the codons, we further investigated the residues for their preferences for specific secondary structural elements (41, 42). The residues are compatible with the helices and strands, characteristics of the GH10 xylanase ($\alpha/\beta$) structure (28). The two negative codons encode for one Arg and one Ser. Arg is a non-helix- and strand-forming residue, and Ser prefers forming turns (Table 1). The three positive codons encode for two Arg residues and one Ile. These two residues are non-helix- and strand-forming residues (Table 1). In general, the negative and positive codons do not prefer specific secondary structural elements. Thereby, the codon effects depend on the residual propensities for relative secondary structures.

**mRNA Nucleotide Relation with the $T_{\text{opt}}$ Value—Nucleotide contents are calculated within mRNA and analyzed with the $T_{\text{opt}}$ value to investigate nucleotide effect on the codons. A% and C% correlate positively and negatively with the $T_{\text{opt}}$ value, respectively (Fig. 2). Thereby, AG% and UC% correlate positively and negatively with the $T_{\text{opt}}$ value, respectively. Consistent with our data, genome analysis found that A and G were typically involved in thermophilic mRNA (2). Thus, the codons relating to the $T_{\text{opt}}$ value were selected by thermophilic force at the nucleotide level.

The relationship of A% and C% within mRNA with the $T_{\text{opt}}$ value is consistent the A- and C-ending feature of positive and negative codons. Thereby, the nucleotides A and C are indicated as having positive and negative effects on the xylanase thermal activity.
Codon Affects Enzyme Thermal Activity

The nucleotides are A (adenine, $T_{\text{opt}} = 23.8 + 0.14 \times [A] \% (p = 0.03)$), C (cytosine, $T_{\text{opt}} = 28.2 - 0.13 \times [C] \% (p = 0.1)$), AG (adenine and guanine, $T_{\text{opt}} = 49.6 + 0.11 \times [AG] \% (p = 0.01)$), and CU (cytosine and uracil, $T_{\text{opt}} = 50.5 - 0.11 \times [CU] \% (p = 0.01)$). The figure was made by using Origin (version 6.1).

FIGURE 2. Correlation of nucleotide content within mRNA with xylanase $T_{\text{opt}}$. The nucleotides are A (adenine), C (cytosine), AG (adenine and guanine), and CU (cytosine and uracil).

Thermal activity. Among the three codons of Arg, AGA, AGG, and CGU, the first two have positive effects on xylanase $T_{\text{opt}}$ value, and the third one has negative effect on xylanase $T_{\text{opt}}$ value (Table 1). This result correlates with the low preference of Arg for specific secondary structures. The codons, AGA and AGG, have AG, whereas the CGU has CU at the 1st and 3rd nucleotides. The negative and positive effects of cytosine and adenine are attributed to the small and big size, respectively. Comparable with the three-dimensional structural analysis of the codon, the codon central C is reorganized by tRNA (G in the anti-codons) from the major groove side, whereas the central U is reorganized by tRNA (A in the anti-codons) from the minor groove side (43).

DISCUSSION

Influenced by the Anfinsen protein folding principle (44), enzyme thermal property is mainly investigated at the residual or structural levels. Presently, thermal property is analyzed at the codon level by using the GH10 xylanase as a model, and codons are found to have effects on the xylanase thermal property. Consistent with our data, Adzhubei (45) found the degeneracy of genetic codon to relate to the three-dimensional structure of the protein. Saunders and Deane (37) found that codons had more information than residues, and therefore, gave more insights into the translation role in protein folding. Thereby, synonymous codons might have slightly different functions to form the specific secondary structural elements.

Because gene transfer is a common phenomenon among organisms (23–27), OGT value is unspecific when used to describe enzyme thermophilic tendency, whereas thermophilic tendency is specifically described by the enzyme $T_{\text{opt}}$ value. There is a big difference between enzyme $T_{\text{opt}}$ and OGT value (supplemental Table 1). For example, the Cryptococcus xylanase $T_{\text{opt}}$ value is 47.5 °C (50); however, its OGT value is 4–12 °C, lower than that of Aspergillus aculeatus. However, the A. aculeatus xylanase $T_{\text{opt}}$ value is 40 °C (51). The Penicillium simplicissimum xylanase $T_{\text{opt}}$ value is 67.0 °C (52), whereas its OGT value is 30 °C, lower than that of A. aculeatus. However, the A. aculeatus xylanase $T_{\text{opt}}$ value is higher. The big differences probably indicate that the genes were caused by lateral gene transfer. Data selection is also important for enzyme thermophilic analysis because different structural enzymes employ different strategies for thermostability (30, 31). Therefore, we used the $T_{\text{opt}}$ values and CDSs of GH10 xylanases, the homologous enzymes of the same family, to exclude interference from heterologous enzymes and unspecific OGT values.

In the present study, codon effect was analyzed on enzyme thermal property by using linear correlation. Normal practice is to use a cutoff value of 0.05. The $p$ value was set as 0.15 to find as many codons as possible because the 61 codon contents are very low within each CDS. Moreover, the analysis was based on multiple hypothesis, and false positives had to be excluded by the Bonferroni correction method (33, 34). If the cutoff value was set too low, codons would be hard to find. The false positive rate is 67% in codon content analysis (supplemental Table 2). The false positive rate is 75% in RSCU analysis (supplemental Table 3). The false positive rates in both methods are approximately equal to the Bonferroni correction factor, 61. The RSCU has the drawback of having no relation with codon content. Moreover, the RSCU value is 1 for Met and Trp. However, the RSCU analysis found four shared codons by codon content analysis, showing that the analysis is accurate. Higher $p$ values and lower coefficient values indicate codons having only slight effects on enzyme thermal activity; however, the accumulation of the slight effects is of real importance.

We can infer that codons affect enzyme thermal property at different levels. Basically, the rate of amino acid translation differed by a 6-fold order magnitude between common and infrequent synonymous codons (53, 54). At the protein expression level, $\alpha$-helix is encoded faster than $\beta$-strand and irregular coil (8). Codon bias usage could be used to distinguish different secondary structural elements (55) because synonymous codons are not randomly used. At the structural level, $\alpha$-helix and irregular coil correlated positively and negatively with the xylanase $T_{\text{opt}}$ value (56), and we successfully increased the xylanase thermostability by deleting terminal residues that do not form helix and strand (57). Protein is translated through a successive translocation-and-pause cycle, with a median pause length of 2.8 s (58). Thereby, peptide bond is not formed at an even rate, and in this process, protein structure is organized (59). Thus, we postulate that codon usage affects protein translational rate, and therefore, affects its folding and hydrogen-bond formation, and ultimately, affects enzyme thermal activ-
ity. The resurrected ancestral elongation factor had higher thermostability than the other homologues (3), demonstrating directly that codons affected protein thermal property. Zhang et al. (60) also perturbed the protein folding efficiency by mutating some synonymous codons. Mendez et al. (61) found GC bias mutation affecting the protein folding stability, making it less hydrophobic and less stable to unfolding but also less susceptible to misfolding and aggregation.

In summary, the GH10 xylanase CDSs and $T_{opt}$ values were used to eliminate the interference of residual bias usage that arose from using different structures and unspecific OGT values as the thermostable index. The codon effect was analyzed on xylanase thermal property by respectively analyzing codon content and RSCU, and we found five codons, with two (CGU and AGC) showing negative and three (AUA, AGA, and AGG) showing positive correlations with the xylanase $T_{opt}$ value. Thus, in addition to the well known degeneracy, genetic codons also have effects of message on enzyme thermal property. However, the message is lost when the thermal property is analyzed at the residual level.

Acknowledgments—We thank the anonymous reviewer for the sincere comments on the paper, especially for the Bonferroni correction analysis. We also thank Youwei Wu of Jiangnan University for teaching us about the Bonferroni correction method. We are grateful to Dr. Peter Morgan of the Neurobiology laboratory of Rostock University, Youfa Cheng of University of California at San Diego, and Guoqing Xia of Tübingen University, for critical review of the paper.

REFERENCES

Codon Affects Enzyme Thermal Activity