ABSTRACT: An RNA G-quadruplex in the protein coding segment of mRNA is translatable (TG4) and may potentially impact protein translation. This can be consequent to staggered ribosomal synthesis and/or result in an increased frequency of missense translational events. A mathematical model of the peptides that encompass the substituted amino acids, ie, the TG4-mapped peptidome, has been previously studied. However, the significance and relevance to disease biology of this model remains to be established. ProTG4 computes a confidence-of-sequence-identity ($\gamma$)-score, which is the average weighted length of every matched TG4-mapped peptide in a generic protein sequence. The weighted length is the product of the length of the peptide and the probability of its non-random occurrence in a library of randomly generated sequences of equivalent lengths. This is then averaged over the entire length of the protein sequence. ProTG4 is simple to operate, has clear instructions, and is accompanied by a set of ready-to-use examples. The rationale of the study, algorithms deployed, and the computational pipeline deployed are also part of the web page. Analyses by ProTG4 of taxonomically diverse protein sequences suggest that there is significant homology to TG4-mapped peptides. These findings, especially in potentially infectious and infesting agents, offer plausible explanations into the aetiology and pathogenesis of certain proteopathies. ProTG4 can also provide a quantitative measure to identify and annotate the canonical form of a generic protein sequence from its known isoforms. The article presents several case studies and discusses the relevance of ProTG4-assisted peptide analysis in gaining insights into various mechanisms of disease biology (mistranslation, alternate splicing, amino acid substitutions).

KEYWORDS: Bioinformatics, mathematical and computational biology, peptidome and peptide analysis, protein sequences, translatable G-quadruplex

Background
Non-canonical or atypical translation is the synthesis of proteins by one or more mechanism(s) that depart significantly from the usual molecular biology. This includes the synthesis of short peptides (<100 aa) from upstream open reading frames (uORF), short open reading frames (sORF)-encoded polypeptides (SEPs), and translation from atypical start sites.1-4 A translatable G-quadruplex (TG4) is defined as the presence of Hoogsteen or reverse-Hoogsteen pairing between inter-spersed repeats of Guanosine residues in the protein coding segment (PCS) of a messenger RNA (mRNA).5-4 The presence of one or more TG4 s may stall the ribosomal machinery and/or facilitate missense translations. In fact, the stress-independent amino acid substitutions that result from the latter can account for errors rates $\approx 10^{-4}$ – $10^{-5}$ for proteins of approximately 300 amino acids.5,10,11 Although most of these substitutions favour misfolding and thence disorder, the same may also contribute to alternate splicing and a compensatory adaptation to missing tRNA molecules under perturbed conditions.11-15

The combinatorial superset of protein/peptide sequences that will result under the assumptions of the ‘wobble’- and ‘superwobble’-hypotheses is referred to as the ‘statistical’-proteome/peptidome.5,10,11 While the myriad of forms has precluded a comprehensive empirical validation, reverse-mapping, ie, populating the ‘empirical’-peptidome (mass spectrometry, microarrays) and thence characterizing the same might be an alternative strategy. In fact, protocols such as the data-dependent acquisition (DDA) and quantification of rare amino acid substitutions (QRAS) have made much progress.10 In addition, generic software which are platform-dependent and -independent may utilize advanced data analytics to compare and infer biological relevance.16-21 However, a web-based tool which can scan the sequence of a generic protein for the occurrences of this ’statistical’-peptidome and thereby establish biological relevance is not available.

A mathematical model (codon-association with third-base wobble) of a short TG4 was able to establish a map with a subset of the ‘statistical’-peptidome (TG4 $\rightarrow$ PTG4).5 Interestingly, the model was able to distinguish and quantitate the differential occurrence of TG4-mapped peptides in a 6-frame translation model of empirically validated protein sequences.5 This study also established a strong association (co-occurrence, correlation) between TG4-mapped peptides and disorder-favouring short linear motifs (SLiMS) within and across taxa.5,22 Despite these preliminary studies, a detailed assessment of the biomedical relevance of TG4-mapped peptides (PTG4) is awaited. ProTG4 will compare the full complement of TG4-mapped peptides with a user-defined/generic protein sequence, assign a weight to the selected matches and compute a confidence-of-sequence-identity ($\gamma$)-score.
The impact of ProTG4 on basic and disease biology will be gauged by comparing the distribution of TG4-mapped peptides across taxonomically diverse protein sequences. In particular, ProTG4 will examine protein sequences from the non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenase (2OyGd) and the carbohydrate active enzyme (CAZy) enzyme superfamilies.23–27 Members are present in all kingdoms of life, possess a conserved active site, are well characterized (in silico, laboratory), and are clinically relevant.27–29 Furthermore, the γ-score generated by ProTG4 will be parameterized and evaluated as indices to identify and annotate the canonical form of a generic protein sequence from its known isomers. This objective will be accomplished by evaluating known isomers of protein sequences from a diverse set of curated and reference proteomes.

**Rationale, Mathematical Derivation, and Algorithm Deployed by ProTG4**

A detailed account of the mathematics (formulation, derivation, enumeration) involved in constructing a model of a short TG4 in the protein coding segment (PCS) of a mRNA and its mapping has already been described. Briefly, SEPs with molecular weights (~0.8–2.3KD, ~7–20aa) were used to define the boundaries of the peptides that comprise PTG4.5 The corresponding TG4 was then modelled as an intra-strand subsequence of the mRNA (~20–60 Mer) of the gene of a hypothetical protein.5

\[
TG4 = \left( \left( G_{i,j} \right)_{1 \leq i,j \leq 3} \right)_{n=1} \text{ Def. (1)}
\]

\[
t = \text{Number of Guanines per } G - \text{rich cluster}
\]

\[
h = \text{Number of loop - forming generic intervening nucleotides}
\]

\[
k = \text{Cluster index}
\]

\[
m = \text{Number of strands}
\]

\[
G = \text{Guanine}
\]

\[
A = \text{Adenine}
\]

\[
T = \text{Thymine}
\]

\[
C = \text{Cytosine}
\]

\[
N = \text{Any nucleotide}
\]

A codon-association with third-base ‘wobble’ model was deployed to annotate a subset of vertebrate codons which was then used to identify amino acids for the Guanosine stretches \((y \in Y)\) and loops \((z \in Z)\),5

\[
pTG4_y = \left( \left( G_{i,j} \right)_{1 \leq i \leq 3} \right)_{y \in Y} \left( \left( G_{i,j} \right)_{1 \leq i \leq 3} \right)_{z \in Z} \text{ Def. (2)}
\]

\[
PTG4 = \text{Peptidome corresponding to } TG4
\]

\[
pTG4_y = j^{th} \text{ canonical amino acid form of } PTG4
\]

\[
i = \text{Number of amino acids that comprise the modelled } PTG4
\]

\[
J = \text{Maximum number of canonical } pTG4
\]

\[
y \in Y \text{ (set of specialized amino acids)}
\]

\[
z \in Z \text{ (set of all amino acids)}
\]

The TG4-mapped peptidome \((pTG4_y \in PTG4)\) is then,

\[
PTG4 = \bigcup_{i=20}^{i=20} \bigcup_{j=1}^{j=J} \left| pTG4_y \right|
\]

The association \(f(TG4, PTG4)\) is clearly a surjection, \(f : TG4 \rightarrow PTG4,^5\)

**Assessing the occurrence of TG4-mapped peptides in a generic protein sequence**

ProTG4 is a web server that queries a protein sequence for the full complement of TG4-mapped peptides. The occurrences of every matched peptide sequence \((pTG4_y \in pTG4 \subseteq PTG4)\) are then computed in a library of randomly generated sequences \((pTG4_y \in V)\) of equivalent lengths.

\[
\phi_{pTG4_y} = \sum \nu \in V pTG4_y
\]

The weight \((\omega)\) of each positive match in a generic protein sequence is the probability that this occurrence is not due to chance, ie,

\[
\omega = 1 - \frac{\phi_{pTG4_y}}{\#V} = 1 - \left( \sum pTG4_y \right) \text{ Def. (3)}
\]

Here,

\[
\#V = \left\{ \begin{array}{ll}
10^4, & L \in [7, 50] \text{aa} \\
10^5, & L \in (50, \infty) \text{aa}
\end{array} \right.
\]

\[
L = \text{Length of generic target protein sequence}
\]

\[
V = \text{Library of randomly generated sequences}
\]

\[
aa = \text{Amino acids}
\]

The weighted length of a positive match in a generic protein sequence is then,

\[
(\omega_\nu L)_{\nu \in V} = \omega_y L
\]

The confidence-of-sequence-identity \((\gamma)\) over the entire length of the protein is,

\[
\gamma = \left( \frac{1}{L} \right) \sum_{\nu \in V} \omega_y L
\]
Clearly, $\gamma$ can be computed for a single protein sequence or be utilized to derive the corresponding data from entire proteomes (Table 1).

**Implementation and usage of ProTG4**

The computations outlined, *vide supra*, are dependent on the length of protein sequences. A reasonable (~10–20) for the results may be obtained by restricting the number of user-defined sequences (~5–10). The server is simple to use and provides the user with a brief description of the rationale, algorithm and pipeline deployed. There are also several important instructions and precautions that the user must adhere to for relevant and timely feedback. Several ready-to-use examples (radio button) are provided to explore and comprehend the functioning of ProTG4. If ProTG4 finds suitable peptides in the user-defined sequence(s), it outputs these as tables to independent files which can be downloaded (summary, details). While the summarized data includes the sequence ID, length of the protein, number of positive matches, and confidence-of-sequence-identity ($\gamma$), details mention the amino acid sequence, start and end positions (Figure 1). ProTG4 gives consistent results when tested in three independent browsers (Chrome, Firefox, Microsoft Edge). The coding is done using in-house developed PERL scripts along standard HTML for the design and layout.

**Biomedical relevance of ProTG4**

The translatable G-quadruplex ($\text{TG}_4$) may be an important cause of abnormal amino acid substitutions in de novo
Table 1. (Continued)

| DATA SETS                  | SEQUENCES | $\gamma_{\text{prot}}$ | $\gamma_{\text{max}} \in [\gamma_{\text{prot}}, \gamma_{\text{inf}}]$ | REFERENCE(S)                      |
|---------------------------|-----------|-------------------------|-------------------------------------------------|-----------------------------------|
| 4. *C. trachomatis* (UP000000431) |           |                         |                                                 |                                   |
| Generic                   | 1992      | 0.92                    | 0.63-1.00                                       | Supplementary Texts 5a and 5b      |
| Combined                  | 1993      |                         |                                                 |                                   |
| Isoforms (canonical & curated) | 742     |                         |                                                 |                                   |
| Misc.                     | 1,251     |                         |                                                 |                                   |
| 5. *C. albicans* (UP000000559) |           |                         |                                                 |                                   |
| Generic                   | 1447      | 0.88                    | 0.40-1.00                                       | Supplementary Texts 6a and 6b      |
| Combined                  | 1478      |                         |                                                 |                                   |
| Isoforms (canonical & curated) | 264     |                         |                                                 |                                   |
| Misc.                     | 1,214     |                         |                                                 |                                   |
| 5. *D. rerio* (UP000000437) |           |                         |                                                 |                                   |
| Generic                   | 3216      | 0.92                    | 0.58-1.00                                       | Supplementary Texts 7a and 7b      |
| Combined                  | 3456      |                         |                                                 |                                   |
| Isoforms (canonical & curated) | 441     |                         |                                                 |                                   |
| Misc.                     | 3015      |                         |                                                 |                                   |
| *D. melanogaster* (UP000000803) |      |                         |                                                 |                                   |
| 6. *Genetic*              | 4886      | 0.92                    | 0.33-1.00                                       | Supplementary Texts 8a and 8b      |
| Combined                  | 6442      |                         |                                                 |                                   |
| Isoforms (canonical & curated) | 2335     |                         |                                                 |                                   |
| Misc.                     | 4107      |                         |                                                 |                                   |
| 7. *G. gallus* (UP000000539) |           |                         |                                                 |                                   |
| Generic                   | 2605      | 0.92                    | 0.41-1.00                                       | Supplementary Texts 9a and 9b      |
| Combined                  | 2867      |                         |                                                 |                                   |
| Isoforms (canonical & curated) | 414     |                         |                                                 |                                   |
| Misc.                     | 2453      |                         |                                                 |                                   |
| 8. *C. elegans* (UP0000001940) |           |                         |                                                 |                                   |
| Generic                   | 4747      | 0.92                    | 0.34-1.00                                       | Supplementary Texts 10a and 10b    |
| Combined                  | 6880      |                         |                                                 |                                   |
| Isoforms (canonical & curated) | 3259     |                         |                                                 |                                   |
| Misc.                     | 3621      |                         |                                                 |                                   |
| 9. *B. taurus* (UP0000009136) |           |                         |                                                 |                                   |
| Generic                   | 6904      | 0.92                    | 0.00-1.00                                       | Supplementary Texts 11a and 11b    |
| Combined                  | 7353      |                         |                                                 |                                   |
| Isoforms (canonical & curated) | 791     |                         |                                                 |                                   |
| Misc.                     | 6562      |                         |                                                 |                                   |
Figure 1. Implementation and usage of ProTG4: ProTG4, computes a confidence-of-sequence-identity (γ)-score, which is the average weighted length of every matched TG4-mapped peptide in a generic protein sequence. The weighted length is the product of the length of the peptide and the probability of its non-random occurrence in a library of randomly generated sequences of equivalent lengths. This is then averaged over the entire length of the protein sequence. ProTG4, is simple to operate, has clear instructions and is accompanied by a set of ready-to-use examples. The rationale of the study, algorithms deployed and the computational pipeline deployed are also part of the web page. The output of ProTG4 is in tabular format and written to independent files which can be downloaded (summary, details). The summarized data include the sequence ID, length of the protein, number of positive matches, and the γ-score. Details include the amino acid sequence of each matching pattern, start and end positions.

mRNA indicates messenger ribonucleic acid; PCS, protein coding segment; PTG4, peptidome associated with TG4; SLiMS, short linear motifs; TG4, translatable G-quadruplex.

protein synthesis. The persistent of these modifications reflect ‘escape’ mechanisms from the proof reading machinery and are implicated in the proteopathies or diseases associated with dysfunctional proteostasis. An interesting finding of the previous study was the redundancy of specific amino acids in proteins across taxa. These were
### Table 2. ProTG4-derived and confidence-of-sequence-identity based predictors of the canonical form of reference protein sequences with known isoforms.

| ORGANISM        | # iprot | γγ\text{min} | γγ\text{max} | γγ\text{min} ∨ γγ\text{max} | TP | FN | \(R = \frac{TP}{TP + FN}\) | REFERENCE(S)  |
|-----------------|---------|---------------|---------------|-------------------------------|----|----|-----------------------------|---------------|
| 1. *H. sapiens* | 10175   | −             | −             | −                             | 3230 | 6945 | = 32%                       | Supplementary Texts 4c, 4d and 4e |
|                 |         | −             | +             | −                             | 4104 | 6071 | = 40%                       |               |
|                 |         | −             | −             | +                             | 7334 | 2841 | = 72%                       |               |
| 2. *C. trachomatis* | 201    | +             | −             | −                             | 26  | 175  | = 13%                       | Supplementary Texts 5c, 5d and 5e |
|                 |         | −             | +             | −                             | 39  | 162  | = 19%                       |               |
|                 |         | −             | −             | +                             | 65  | 136  | = 32%                       |               |
| 3. *C. albicans* | 114     | +             | −             | −                             | 37  | 77   | = 33%                       | Supplementary Texts 6c, 6d and 6e |
|                 |         | −             | +             | −                             | 39  | 75   | = 34%                       |               |
|                 |         | −             | −             | +                             | 76  | 38   | = 67%                       |               |
| 4. *D. rerio*    | 201     | +             | −             | −                             | 84  | 117  | = 42%                       | Supplementary Texts 7c, 7d and 7e |
|                 |         | −             | +             | −                             | 106 | 95   | = 53%                       |               |
|                 |         | −             | −             | +                             | 190 | 11   | = 79%                       |               |
| 5. *D. melanogaster* | 759 | +             | −             | −                             | 303 | 456  | = 95%                       | Supplementary Texts 8c, 8d and 8e |
|                 |         | −             | +             | −                             | 293 | 466  | = 39%                       |               |
|                 |         | −             | −             | +                             | 596 | 163  | = 79%                       |               |
| 6. *G. gallus*   | 152     | +             | −             | −                             | 60  | 92   | = 40%                       | Supplementary Texts 9c, 9d and 9e |
|                 |         | −             | +             | −                             | 71  | 81   | = 47%                       |               |
|                 |         | −             | −             | +                             | 131 | 21   | = 86%                       |               |
| 7. *C. elegans*  | 1126    | +             | −             | −                             | 399 | 727  | = 35%                       | Supplementary Texts 10c, 10d and 10e |
|                 |         | −             | +             | −                             | 490 | 636  | = 44%                       |               |
|                 |         | −             | −             | +                             | 889 | 237  | = 79%                       |               |
| 8. *B. taurus*   | 342     | +             | −             | −                             | 162 | 180  | = 47%                       | Supplementary Texts 11c, 11d and 11e |
|                 |         | −             | +             | −                             | 159 | 183  | = 47%                       |               |
|                 |         | −             | −             | +                             | 321 | 21   | = 94%                       |               |
| 9. *R. norvegicus* | 966    | +             | −             | −                             | 349 | 617  | = 36%                       | Supplementary Texts 12c, 12d and 12e |
|                 |         | −             | +             | −                             | 475 | 491  | = 49%                       |               |
|                 |         | −             | −             | +                             | 824 | 142  | = 85%                       |               |
| 10. *A. thaliana* | 1871    | +             | −             | −                             | 732 | 1139 | = 39%                       | Supplementary Texts 13c, 13d and 13e |
|                 |         | −             | +             | −                             | 984 | 887  | = 52%                       |               |
|                 |         | −             | −             | +                             | 1716 | 155   | = 92%                       |               |

**Abbreviations:** FN, false negative; R, recall or sensitivity analysis of predictor; TP, true positive; \# iprot, cardinal number of the set of unique protein sequences with one or more curated isoforms; γγ\text{min}, minimum value of the confidence-of-sequence-identity among isoforms of a generic protein sequence; γγ\text{max}, maximum value of the confidence-of-sequence-identity among isoforms of a generic protein sequence; γγ\text{min} ∨ γγ\text{max}, minimum or maximum value of the confidence-of-sequence-identity among isoforms of a generic protein sequence.
purported to be the reason for molecular mimicry, a mechanistic explanation for the secondary proteopathies. Here, an offline analysis by ProTG4 of protein sequences in taxonomically diverse protein sequences (i2OGdd, CAZY) suggests that there is significant homology of TG4-mapped peptides (Table 1; Supplementary Texts 1-3). These findings, especially in potentially infectious and infesting (bacteria, virus, fungi, helminths) agents, offer plausible explanations into the aetiology and pathogenesis of certain proteopathies. Isomorphs of proteins (iprot) share similar functionality and may be arise from gene duplications, re-insertional events (retrotransposition), polyprolidy (anepuolidy, polyplody) and atypcal recombination. The canonical form of a generic protein sequence is annotated upon the basis of several sequence-dependent and sequence-independent strategies. Here, too, the χ-score generated by ProTG4 can be parameterized (χmin, χmax, χmin, χmax) and may offer a quantitative measure to identify and thereby annotate the canonical form of a generic protein sequence from its known isoforms (Tables 1 and 2; Supplementary Texts 4-13).

Conclusions

ProTG4 is a web server that examines the distribution of short stretches of a specialized subset of amino acids that correspond to a translatable G-quadruplex, ie, the TG4-mapped peptide in a generic protein sequence. Here, the implementation, usage and scope of ProTG4 are presented. The article also discusses the relevance of ProTG4-assisted peptide analysis in gaining insights into probable mechanisms (mistranslation, alternate splicing, amino acid substitution) of disease causation and/or progression.

Author Contributions

S.K. designed the study, formulated and developed the algorithms, collated the data and conducted the analysis, and wrote all the code and the manuscript.

Availability and Implementation

The web server is available at the following URL (http://204.152.217.16/ProTG4.html), is free and does not require a login ID.

ORCID iD

Siddhartha Kundu https://orcid.org/0000-0003-3962-776X

Data Availability

All data described in the article is available as supplementary material.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Werfmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A. sORFdb: a comprehensive literature database on eukaryotic sORF biology. Nucleic Acids Res. 2014;42:D60-D67.
2. Smirnov SA, Mitchell AJ, Schwaab AG, et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol. 2013;9:59-64.
3. Frith MC, Forrest AR, Nourbakhsh E, et al. The abundance of short proteins in the mammalian proteome. PLoS Genet. 2006;2:e52.
4. Kundu S. Mathematical model of a short translatable G-quadruplex and an assessment of its relevance to misfolding-induced proteinosis. Math Biosci Eng. 2017;14:2470-2493.
5. Agarwal P, Pandey S, Maiti S. The tale of RNA G-quadruplex. Organometallics. 2013;15:5570-5585.
6. Millevoi S, Moine H, Vagner S. G-quadruplexes in RNA biology. Wiley Interdiscip. Rev:RNA. 2014;4:495-531.
7. Hoogenstr K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylhyamine and 9-methylenedine. Acta Crystallographica. 1963;6:907-916.
8. Garant JM, Luce MJ, Smith BS, et al. sRNA G-quadruplex database (sQDbase). Database (Oxford). 2019;2019:bav059.
9. Garofalo R, Wohlgemuth I, Pearson M, et al. Broad range of missense error frequencies in cellular proteins. Nucleic Acids Res. 2019;47:2932-2945. doi:10.1093/nar/gky1319.
10. Ou X, Cao J, Cheng A, et al. Errors in translational decoding: tRNA wobbling or misincorporation? PLoS Genet. 2019;15:e1008017.
11. Hutchinson S, Furger A, Halliday D, et al. Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype? Hum Mol Genet. 2003;12:2269-2276.
12. Davey NE, Trave G, Gibson TJ. How viruses hijack cell regulation. Trends Biochem Sci. 2011;36:159-169.
13. Jucker M, Walker L.C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. Nature. 2013;501:45-51.
14. Piovesan D, Tabaro F, Micetic I, et al. DivoProT 7.0: a major update of the database of disordered proteins. Nucleic Acids Res. 2015;43:D2219-D2227.
15. Valiongk A1, Sonn T, Ele L.L. A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. Brief Bioinform. 2018;19:1344-1355.
16. Schmidt A, Forne I, Imhof A. Bioinformatic analysis of proteomics data. BMC Biol. 2014;12:Suppl 1.
17. Efthimiou G, Antonakis AN, Pavlopoulos GA, et al. ProteoSign: an end-user online differential proteomics statistical analysis platform. Nucleic Acids Res. 2017;45:W300-W306. doi:10.1093/nar/gkw444.
18. Tyranow S, Temu T, Sinicyn P, et al. The Perseus computational platform for comprehensive analysis of (pro)teomics data. Nat Methods. 2016;13:731-740.
19. Rainer J, Sanchez-Cabo F, Stocker G, et al. CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. Nucleic Acids Res. 2006;34:W498-W503.
20. Codaert L, Helsen K, Impens F, Vandenkercchove J, Gevaert K. A tool to visualize and validate quantitative proteomics data from different sources. Proteomics. 2010;10:1226-1229.
21. van der Lee R, Buljan M, Lang B, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014;114:6659-6631.
22. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 2009;37:D233-D238.
23. Martinez S, Hauinger RP. Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. Chem Rev. 2013;114:6589-6631.
24. Martinez S, Hauinger RP. Catalytic mechanisms of Fe(II) and 2-oxoglutarate-dependent oxygenases. Annu Rev Biochem. 2018;87:585-620.
25. Kundu S. Distribution and prediction of catalytic domains in 2-oxoglutarate-dependent dioxygenases. BMC Res Notes. 2012;5:410.
26. Kundu S. Insights into the mechanism(s) of digestion of crystalline cellulose by plant class C GH9 endoglucanases. J Biol Chem. 2015;290:20702-20711.
27. Islam MS, Leising TM, Chowdhury R, et al. 2-oxoglutarate-dependent oxygenases. Annu Rev Biochem. 2019;88:519-560.
28. Kundu S. Insights into the mechanism(s) of digestion of crystalline cellulose by plant class C GH9 endoglucanases. J Biol Chem. 2015;290:20702-20711.
29. Kundu S. Insights into the mechanism(s) of digestion of crystalline cellulose by plant class C GH9 endoglucanases. J Biol Chem. 2015;290:20702-20711.