Modeling and structural analysis of cellulases using Clostridium thermocellum as template

Nathan Vinod Kumar1, Mary Esther Rani1*, Rathinasamy Gunaseeli2, Narayanan Dhiraviam Kannan3 & Jayavel Sridhar4

1Research Centre, Department of Botany and Microbiology, Lady Doak College, Madurai - 625002; 2Center for Environmental Studies, Lady Doak College, Madurai - 625002; 3Department of Plant Biotechnology, School of Biotechnology, Madurai Kamaraj University, Madurai-625021; 4UGC-NRCBS, School of Biological Sciences, Madurai Kamaraj University, Madurai-625021; Mary Esther Rani – Email: ranimam@yahoo.com; Phone +91 9443064133; *Corresponding author

Abstract:
Cellulase is one of the most widely distributed enzymes with wide application. They are involved in conversion of biomass into simpler sugars. Cellulase of Trichoderma longibrachiatum, a known cellulolytic fungus was compared with Clostridium thermocellum [AAA23261.1] cellulase. Blast was performed with AAA23261.1 as query sequence to obtain nine similar sequences from NCBI protein data bank. The physicochemical properties of cellulase were analyzed using ExPaSy’s ProtParam tool namely ProtParam, SOPMA and GOR IV. Homology modeling was done using SWISS MODEL and checked quality by RMSD values using VMD1.9.1. Active sites of each model were predicted using automated active site prediction server of SCFBio. Study revealed instability of cellulase of two eukaryotic strains namely Trichoderma longibrachiatum [CAA43059.1] and Melanocarpus albozymes [CAD56665.1]. The negative GRAVY score value of cellulases ensured better interaction and activity in aqueous phase. It was found that molecular weight (M. Wt) ranges between 25-127.56 kDa. Iso-electric point (pI) of cellulases was found to be acidic in nature. GOR IV and SOPMA were used to predict secondary structure of cellulase, which showed that random coil, was dominated. Neighbor joining tree with C. thermocellum [AAA23261.1] cellulase as root showed that cellulases of Thermoaerobacter subterranean [ZP_07835928] and C. thermocellum [CAA43051.1] were more similar to eukaryotic cellulases supported by least boot strap values. Pseudalteromonas haloplanktis cellulase was found to be the ideal model supported by least RMSD score among the predicted structures. Trichoderma longibrachiatum cellulase was found to be the best compared to other cellulases, which possess high number of active sites with ASN and THR rich active sites. CYS residues were also present ensuring stable interaction and better bonding. Hydrophilic residues were found high in active sites of all analyzed models and template.

Background:
Cellulases are important enzymes in many proposed processes for producing fuels and chemicals from plant biomass [1]. They are multienzyme complexes, comprising three major components; endo- β-glucanase (EC 3.2.1.4), exo- β -glucanase (EC 3.2.1.9.1) and β- glucosidase (EC 3.2.1.21), which have been shown to act synergistically in the hydrolysis of cellulose [2]. These enzymes are being isolated from many microbial sources and characterized. Cellulase plays a major role in carbon cycling in the biosphere by contributing towards the major carbon source [3]. The role of cellulase in host- pathogen interaction is quite important when considering the cellulase producing ability of pathogenic strains [4]. These enzymes help in hydrolysis of substrates available and for the utilization of microbial growth and for their metabolism. There are no much reports available on computational characterization of the cellulase enzyme. Clostridium thermocellum is a thermophilic bacterial strain with high cellulolytic activity of about 5.32IU/ L [5]. However, when eukaryotic cellulases are concern, Trichoderma sp. are widely used as a cellulase source and reported to possess very high activity. It was found to be the best strain for conversion of substrate into glucose of about 0.92mg/ 0.5ml which shows its higher cellulolytic activity [6].

Cellulase being an enzyme of wide application needs to be characterized in all aspects to understand the structural and
functional relations. Present study is to identify more efficient 
cellulolytic enzyme producing microorganism for 
bioprospecting using the computational analysis. Protein 
sequences of cellulase retrieved from NCBI and were subjected 
to ProtParam to analyze physicochemical parameter, secondary 
structure prediction using GOR IV and SOPMA, homology 
modeling (Swiss model), phylogenetic analysis and active site 
prediction by SCFBio.

Methodology:

Sequence retrieval and alignment
Cellulase protein sequence of Clostridium thermocellum 
[AAA23226.1] was retrieved from the National Center for 
Biotechnology (NCBI) and made as the query sequence for the 
structure, properties prediction and modeling. Blastp 
was performed and obtained nine similar sequences of different 
strains. Clustal W multiple sequence alignment was done for 
those sequences using BioEdit5.0.

Secondary structure and physicochemical characterization 
cellulose
The sequences obtained were analyzed using various softwares 
available in the ExPaSy server [7]. The GOR IV analysis was 
performed to understand the presence of helices, beta turns and 
colls in the protein structure [8]. Self-optimized prediction 
method with alignment (SOPMA) analysis was done for 
analyzing the structural components [9]. Comparison was made 
between the GOR IV and SOPMA analysis results. ProtParam 
software analysis was done to understand about the amino acid 
composition, molecular weight, instability index, aliphatic 
index and grand average of hydropathicity (GRAVY) [7]. 
Hydropathy plot analysis for all cellulase sequences was 
performed and the nature of amino acid residues were studied 
using ProtScale [7] based on Kyte and Doolittle scale.

Homology modeling of cellulase
Homology models were predicted using SWISS- MODEL [10- 
12] and the quality was analysed using VMD 1.9.1 [13]. RMSD 
values were calculated using the RMSD calculator and the best 
homology model was selected. Ramachandran plot for the best 
predicted model was depicted by RAMPAGE software [14].

Phylogenetic analysis
Phylogenetic relation among the aligned cellulase sequences 
obtained from Blastp were analyzed based on neighbor joining 
method [15] using MEGA 4.0 [16]. The cellulase sequence of C. 
thermocellum [AAA23226.1] was considered as the root taxon for 
the analysis. Confidence level was analyzed using bootstrap of 
1000 replications.

Activity validation by active site comparison
Active sites of the predicted models and the template were 
analyzed using Automated Active Site prediction AADS server 
of SCFBio [17]. Amino acid compositions of all the cavities were 
analyzed and the frequency of amino acid occurrence in the 
cavities of each models were analyzed.

Discussion:

Blast analysis and sequence retrieval
The cellulase protein sequence of Clostridium thermocellum 
[AAA23226.1] was used as query sequence and nine sequences 
were obtained by performing Blastp. Multiple sequence 
alignment was done in BioEdit software and further used for 
phylogenetic analysis in MEGA.

Secondary structure and physicochemical analysis
SOPMA and GOR IV were used to predict the secondary 
structure, percentage of alpha, extended and random coils of 
the cellulase producing microorganism are presented Table 1 
(see supplementary material). SOPMA analysis for the structure 
prediction was also done and obtained the percentage of alpha, 
extended, beta and random coils (Table 1). The secondary 
structure indicates whether a given amino acid lies in a helix, 
strands or coil [18, 19]. SOPMA was used for structure prediction 
of cellulase protein [20]. Random coil dominates the other forms 
in the cellulase analyzed by SOPMA and GOR IV. It was 
identified that random coils of M. abomyces (58.72%) and T. 
longibrachiatum (57.88%) were dominant compared to other 
forms. However, followed by random coils, extended forms 
ranging from (10%-27%) was dominant over α and β helix. All 
the cellulases analyzed, α-helix was ranging from (12%-37%) 
dominates β-helix, which had less percentage of conformation 
(4%-10%).

ProtParam analysis was performed and the number of amino 
acid residues, molecular weight, pl value, aliphatic index and 
GRAVY index was obtained for each sequence Table 2 (see 
 supplementary material). Comparison of the amino acid 
residue occurrence in cellulase sequences were done and the 
most dominant residues were highlighted Table 3 (see 
 supplementary material). It was found that molecular weight 
ranging from 25-127 kDa and it was higher in C. thermocellum 
(83 kDa) and lower in M. abomyces (25kDa). Comparing to the 
eukaryotic cellulase available, the higher aliphatic index of up 
to 97.51 was noted in T. subterraneus strains which indicate their 
stability over a wide range of temperatures. GRAVY value was 
negative in all species studied. It was notable that the bacterial 
strains had lower GRAVY values indicating the better 
possibilities of aqueous interaction. pl value showed that 
cellulase is acidic in nature. T. subterraneus had a slightly neutral 
pl value and the highest GRAVY value. Generally it was 
observed that towards acidic pl values the GRAVY tends to be 
low. In eukaryotic cellulases, the occurrence of α helices was 
found to be too low. In case of A. bisporus, α helices was similar 
to that of lower taxonomic groups. Moreover these cellulases 
possess higher percentages of random coils. A general pattern 
of inverse relationship between the percentage of occurrence of 
α helices and random coils were observed in both higher and 
lower taxonomic levels.

Cellulase of M. abomyces, T. longibrachiatum and R. flavfaciens 
FD-1 was classified as unstable (II > 40) with an instability 
index (II) of 53.54, 55.23 and 54.34 respectively. It is notable that 
the M. abomyces and T. longibrachiatum are eukaryotic isolates 
and possess the least percentage of alpha helices in their 
structure. P. haloplanktis and R. flavfaciens FD-1 with dominant 
amino acid residues Asn (10.1%) and Ser (11.6%) respectively 
which are hydrophilic residues, all the other sequences had 
ALA and GLY as dominant residues which are hydrophobic in 
nature. ALA was dominant in cellulases of A. bisporus, C. 
thermocellum, P. carotovarum, Saccharophagus sp. and T. 
subterraneus whereas, Gly was dominant for C. thermocellum, M. 
abomyces and T. longibrachiatum.
Compared to bacterial cellulases, fungal cellulases are widely used. Moreover, the cellulolytic activities are high for fungal cellulases. Highest cellulase activity for C. thermocellum was 12.05 IU/ml [5]. P. haloplankis being a psychrophilic bacterium the cellulase obtained is cold adaptable. Cellulase from the former has conserved five amino acid residues in their active sites [21]. C. thermocellum is a thermophilic bacteria and its cellulase has a better heat stability. It is known to be ethanogenic strain and cellulase from this source has high commercial applications [22]. Cysteine residues contribute to protein thermal stability [22]. Amongst fungi, species of Trichoderma and Aspergillus are well known for cellulolytic potential [23]. Apart from the above, other fungi used for cellulase production are Humicola and Aspergillus sp. [24].

Hydropathy plot for the cellulase sequence was constructed using ProtScale based on Kyte and Doolittle and the hydrophilicity and hydrophobicity nature was observed from the plot. It was observed that the majority of the residues were belonging to the hydrophilic regions confirming the interaction of the enzymes in aqueous medium. Aliphatic residues namely ALA, LEU, ILE and VAL were among the hydrophobic residues in the profile. Similarly, Phe which is an aromatic residue and sulfur containing residues may be favouring the enzyme activity in the profile. Hydrophobic residues were much similar to eukaryotic cellulase and it is suggested that cellulase of C. thermocellum is grouped among eukaryotic cellulase sequences.

Homology model validation
SWISS MODEL was used to predict the homology model of the cellulase sequences and the protein structure quality was analyzed. RMSD values for the models were calculated and the model with least value i.e. the best predicted model is shown in Figure 1. Ramachandran plot for the model was constructed using RAMPAGE software. Residue B 169 - LEU belonged to outlier region and the number of residues in the allowed and favoured region was very close to the expected values. It was observed that 94.8% of residues were in favored region and 5.5% in allowed region. It was found that 0.2% was found in outlier region.

Phylogenetic analysis
Phylogenetic tree was constructed using the ten sequences based on neighbour joining method with reference sequence C. thermocellum [AAA23226.1] as a root (Figure 2). It was observed that the cellulase of T. subterraneus [ZP_07835928.1] was found to be more related to the eukaryotic cellulases. T. longibrachiatum [CAA43035.1], T. subterraneus [ZP_07835928.1], M. albiomycetes [CAD56665.1], A. bisporus [CAA83971.1], C. thermocellum [AAA23226.1] were belonging to same group. It can be implied that cellulase sequence of T. subterraneus and C. thermocellum were much similar to eukaryotic cellulase and it is not much evolved from the C. thermocellum [AAA23226.1] cellulase sequence. But the higher boot strap values for the other sequences supports its divergence from the root sequence. However, all the taxa of the group belonged to prokaryotic origin. There was no much influence for evolutionary divergence of the sequence with respect to variations in secondary structure.
extreme environments and ASN rich cavities may be contributing towards better enzyme activity. Among the analysed models, 4 models and the template was found to possess ASN as the dominant residue in its active sites. Both C. thermocellum and R. flavefaciens FD-1 cellulases had LYS rich active sites. ARG was dominant in active sites of M. albobumyces [CAD56665.1] and T. subteraneus DSM 13965[ZP_07835928.1] cellulases. However P. haloplanktis, an extremophile had THR dominant active sites. In T. longibrachiatum ASN and THR was found to be dominant in active sites with a frequency of 10.58. It is clearly notable that the hydrophilic amino acid residues are high in the active sites of these enzyme structures ensuring their interaction with substrate in aqueous phase. However the least found residue was CYS which assures stable interaction and bonding. T. longibrachiatum -2063 found residue was CYS which assures stable interaction and bonding. T. subterraneus DSM 13965 is clea

**Conclusion:**
These studies provide an insight for better prospecting of cellulolytic isolates from the environment for various industrial applications. Among the microbial cellulase used in the present work, T. longibrachiatum cellulase was found to be best with high number of active sites.

**Acknowledgement:**
Authors are thankful to Department of Science and Technology, Government of India and TNSCST for providing the grant to facilitate the research. Moreover, we thank the Management, Lady Doak College, Madurai for providing us facilities and support for our work. JS thank the UGC Govt. of India for sponsoring the Networking Resource Centre in Biological Sciences at Madurai Kamaraj University, Madurai.

**References:**
[1] Wilson DB, Appl Microbiol Biotechnol. 2012 93: 497
[2] Ryu DDD & Mandels M, Enzym Microb Tech. 1980 2: 91
[3] Nowak J et al. Mater Sci Eng. 2005 C25: 119
[4] Gibson DM et al. Curr Opin Microbiol. 2011 14: 264
[5] Otajewu FD & Aluyi HAS, M eden A Appl Sci. 2011 5: 141
[6] Omojasola P et al. Nat Sci. 2008 6: 64
[7] http:// www.expasy.org/ tools/
[8] http:// npsa-pbil.ibcp.fr/cgi-
binc/npsa_automat.pl?page=npsa_gor4.html.
[9] Geourjon C & Delagre G, Prot Eng. 1994 7: 157
[10] Arnold K et al. Bioinformatics. 2006 22: 196
[11] Kiefer F et al. Nucleic Acids Res. 2009 37: 387 [PMID: 18931379]
[12] Peitsch MC, BioTechnol. 1995 13: 658
[13] Humphrey W et al. J Mol Graphics. 1996 14: 33 [PMID: 8744570]
[14] Love J et al. Proteins: Struct Funct Genet. 2002 50: 437
[15] Saitou N & Nei M, Mol Biol & Evol. 1987 4: 406 [PMID: 3447015]
[16] Tamura K et al. Mol Biol & Evol. 2007 24: 1596 [PMID: 1748638]
[17] Tany S et al. J Chem Inf Model. 2011 51: 2515 [PMID: 21877713]
[18] Jyotsna C et al. Int J Eng Sci Tech. 2010 2: 3468
[19] Ojeiru FE et al. Yonago Acta Med. 2010 53: 25
[20] Pradeep NV et al. Adv Life Sci Tech. 2012 4: 8
[21] Garou G et al. Biochem J. 2004 384: 247 [PMCID: PMC 2341252]
[22] Xu B et al. Eur J Biochem. 2001 268: 3718
[23] Lynd LR et al. Microbiol Molec Biol Rev. 2002 66: 506 [PMCID: PMC 120791]
[24] Ghori MI et al. Afr J Biotechnol. 2011 10: 5878

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

**Citation:** Kumar et al. Bioinformation 8(22): 1105-1110 (2012)
### Supplementary material:

#### Table 1: Percentage of amino acids sequence forming secondary structure in GOR IV and SOPMA prediction

| Sl. No | Organisms                        | Accession Number | GOR IV Analysis | SOPMA Predictions |
|--------|----------------------------------|------------------|-----------------|-------------------|
|        |                                  |                  | α-helix (%)     | Extended strand (Ee) (%) | Random coils (%) | β-turns (Tt) (%) | Extended strand (Ee) (%) | Random coils (%) |
| 1      | Clostridium thermocellum         | AAA 23226.1      | 25.24           | 22.54             | 52.23            | 37.52            | 4.72              | 11.61             | 46.15            |
| 2      | Agaricus bisporus                | CAA 83971.1      | 25.57           | 17.12             | 57.31            | 30.82            | 7.08              | 10.5              | 51.6             |
| 3      | Melanocarpus albomycyes          | CAD 56665.1      | 9.36            | 24.26             | 66.38            | 13.62            | 5.96              | 21.7              | 58.72            |
| 4      | Clostridium thermocellum         | CAA 43035.1      | 23.14           | 25.71             | 51.15            | 35.45            | 2.84              | 13.8              | 47.9             |
| 5      | Pectobacterium carotovorum       | AAC 37033.1      | 20.27           | 26.8              | 50.89            | 29.95            | 6.53              | 22.52             | 40.99            |
| 6      | Trichoderma longibrachiatum      | CAA 43059.1      | 6.91            | 25.27             | 67.82            | 15.77            | 7.34              | 19.01             | 57.88            |
| 7      | Pseudoalteromonas haloplanktis   | CAA 76775.1      | 24.09           | 23.28             | 50.63            | 23.89            | 8.1               | 17.81             | 50.2             |
| 8      | Ruminococcus flavefaciens FD-1  | AAA 26468.1      | 26.56           | 19.38             | 54.06            | 31.56            | 6.25              | 23.44             | 38.75            |
| 9      | Sacharophagus sp. Myt-1          | BAL 42331.1      | 18.96           | 27.24             | 53.8             | 23.06            | 10.85             | 27.84             | 38.26            |
| 10     | Thermoanaerobacter subteraneus   | ZP_07835928.1    | 48.62           | 11.6              | 39.78            | 37.02            | 6.63              | 22.38             | 33.98            |

#### Table 2: ProtParam analysis of cellulosic sequences

| Sl. No | Organisms                        | Accession Number | No. of acids | M. Wt | pI | Aliphatic Index | GRAVY | Instability Index (II) | Stability |
|--------|----------------------------------|------------------|--------------|-------|----|-----------------|-------|------------------------|-----------|
| 1      | Clostridium thermocellum         | AAA 23226.1      | 741          | 83558.3 | 5.26 | 63.36           | -0.497 | 29.82                  | Stable    |
| 2      | Agaricus bisporus                | CAA 83971.1      | 438          | 46209.5 | 5.01 | 74.73           | -0.166 | 34.57                  | Stable    |
| 3      | Melanocarpus albomycyes          | CAD 56665.1      | 235          | 25000.8 | 5.31 | 55.32           | -0.251 | 53.54                  | Unstable  |
| 4      | Clostridium thermocellum         | CAA 43035.1      | 739          | 82088.9 | 5.19 | 74.22           | -0.324 | 29.93                  | Stable    |
| 5      | Pectobacterium carotovorum       | AAC 37033.1      | 444          | 48300.6 | 5.32 | 77.77           | -0.369 | 27.88                  | Stable    |
| 6      | Trichoderma longibrachiatum      | CAA 43059.1      | 463          | 48337.0 | 4.80 | 55.23           | -0.374 | 41.07                  | Unstable  |
| 7      | Pseudoalteromonas haloplanktis   | CAA 76775.1      | 494          | 52873.4 | 4.21 | 69.13           | -0.403 | 18.68                  | Stable    |
| 8      | Ruminococcus flavefaciens FD-1  | AAA 26468.1      | 320          | 35938.0 | 4.43 | 85.94           | -0.374 | 54.34                  | Unstable  |
| 9      | Sacharophagus sp. Myt-1          | BAL 42331.1      | 1171         | 127561.5 | 4.77 | 71.85           | -0.380 | 23.35                  | Stable    |
| 10     | Thermoanaerobacter subteraneus   | ZP_07835928.1    | 362          | 39191.0 | 6.28 | 97.51           | -0.067 | 38.89                  | Stable    |

#### Table 3: Occurrence percentage of amino acid residues (Highlighted are the highest amino acid frequencies)

| Amino acid Residues | AAA 23226.1 | CAA 83971.1 | CAD 56665.1 | CAA 43035.1 | CAA 76775.1 | AAA 23226.1 | ZP_07835928.1 |
|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|---------------|
| Percentage of Occurrence in % | 8.60 | 13.20 | 10.20 | 9.60 | 8.60 | 6.50 | 9.30 | 5.90 | 10.20 | 13.90 |
| Ala                 | 3.50 | 2.30 | 5.10 | 3.50 | 4.50 | 2.40 | 1.40 | 2.50 | 3.10 | 8.80 |
| Arg                 | 4.50 | 6.80 | 3.80 | 4.90 | 7.40 | 7.60 | 10.10 | 7.80 | 7.90 | 0.60 |
| Asn                 | 6.70 | 5.70 | 6.00 | 6.80 | 7.00 | 4.50 | 8.90 | 7.20 | 6.00 | 5.80 |
| Asp                 | 0.30 | 2.10 | 6.00 | 0.80 | 0.00 | 4.80 | 1.80 | 0.60 | 1.00 | 0.30 |
| Cys                 | 2.80 | 5.70 | 4.30 | 2.70 | 3.80 | 3.70 | 3.00 | 0.90 | 4.70 | 3.00 |
| Gin                 | 5.10 | 2.10 | 3.40 | 4.90 | 3.20 | 2.20 | 3.80 | 8.80 | 5.10 | 6.90 |
| Glu                 | 8.60 | 8.70 | 12.20 | 8.40 | 9.50 | 22.10 | 8.90 | 2.50 | 9.20 | 10.20 |
| Gly                 | 1.60 | 1.10 | 1.70 | 1.90 | 1.60 | 1.10 | 1.60 | 1.60 | 2.70 | 3.90 |
| His                 | 4.20 | 3.90 | 1.70 | 4.60 | 5.40 | 3.00 | 5.10 | 10.60 | 5.20 | 4.30 |
Table 4: Amino acid frequencies in active sites of predicted cellulase models and template

| Amino acid residues | M o d e l 1 | M o d e l 2 | M o d e l 3 | M o d e l 4 | M o d e l 5 | M o d e l 6 | M o d e l 7 | M o d e l 8 | M o d e l 9 | M o d e l 10 | Template |
|---------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|----------|
| ALA                 | 3.91       | 5.34       | 5.89       | 3.85       | 6.22       | 5.38       | 7.2        | 4.12       | 6.82       | 8.63      | 6.7      |
| ARG                 | 5.12       | 1.90       | 9.55       | 7.74       | 6.94       | 3.73       | 3.64       | 9.28       | 5.08       | 13.4      | 3.81     |
| ASN                 | 4.48       | 10.50      | 4.59       | 4.05       | 8.41       | 10.58      | 9.34       | 5.38       | 11.26      | 0.22      | 9.63     |
| ASP                 | 6.80       | 5.57       | 7.98       | 7.62       | 6.52       | 5.63       | 7.15       | 4.22       | 6.18       | 4.71      | 6.96     |
| CYS                 | 0.13       | 0.01       | 2.50       | 1.1        | 0          | 3.11       | 0          | 0          | 0.14       | 0         | 0        |
| GLN                 | 3.46       | 1.71       | 5.38       | 4.03       | 7.14       | 5.22       | 4.68       | 7.47       | 6.64       | 3.77      | 5.34     |
| GLU                 | 7.30       | 7.41       | 6.12       | 6.97       | 6.74       | 3.07       | 7.86       | 6.69       | 6.64       | 8.36      | 7.75     |
| GLY                 | 5.41       | 8.78       | 6.49       | 4.67       | 6.06       | 6.71       | 6.87       | 3.83       | 3.6        | 5.22      | 6.28     |
| HIE                 | 1.75       | 4.05       | 2.50       | 2.37       | 1.92       | 0.64       | 2.86       | 2.89       | 4.43       | 5.89      | 3.22     |
| ILE                 | 3.20       | 5.63       | 2.55       | 3.44       | 3.84       | 3.09       | 5.28       | 3.24       | 6.7        | 3.9       | 4.9      |
| LEU                 | 2.97       | 6.71       | 4.03       | 5.44       | 7.46       | 3.03       | 2.77       | 4.22       | 5.84       | 8.75      | 3.32     |
| LYS                 | 10.57      | 3.62       | 3.10       | 10.54      | 6.39       | 4.48       | 6.93       | 9.34       | 6.8        | 4.06      | 7.68     |
| MET                 | 1.46       | 3.62       | 0.00       | 0.97       | 0.48       | 1.58       | 0.69       | 6.42       | 1.74       | 1.88      | 0.8      |
| PHE                 | 4.11       | 3.19       | 5.80       | 3.96       | 3.13       | 1.86       | 2.46       | 6.65       | 2.38       | 3.9       | 2.79     |
| PRO                 | 5.55       | 6.67       | 7.79       | 4.91       | 5.15       | 6.43       | 2.43       | 3.15       | 2.23       | 6.9       | 3.06     |
| SER                 | 5.26       | 3.59       | 6.40       | 4.13       | 6.68       | 10.61      | 4.49       | 3.46       | 4.46       | 1.97      | 4.59     |
| THR                 | 9.11       | 5.34       | 5.05       | 4.84       | 5.11       | 10.58      | 10.79      | 4.58       | 4.49       | 2.43      | 9.51     |
| TRP                 | 6.09       | 8.27       | 5.15       | 5.7        | 5.69       | 3.39       | 6.15       | 8.19       | 5.78       | 1.7       | 5.7      |
| TYR                 | 9.82       | 3.87       | 4.22       | 9.7        | 3.26       | 7.11       | 4.14       | 2.52       | 4.16       | 3.67      | 3.94     |
| VAL                 | 3.40       | 4.22       | 4.91       | 3.88       | 2.86       | 3.97       | 4.27       | 4.35       | 4.77       | 10.5      | 4.12     |
| **Total Cavities**  | **84**     | **34**     | **26**     | **58**     | **39**     | **192**    | **71**     | **29**     | **37**     | **49**     | **85**   |