Available online at www.worldscientificnews.com

World Scientific News

An International Scientific Journal

WSN 137 (2019) 31-41 EISSN 2392-2192

The expression level of a recombinant lipase predicted *in silico* by different codon optimization algorithms

Adebayo J. Bello[1,*], Miriam Igwo-Ezikpe[2], Emuejevoke T. Toye[3] and **Joy Okpuzor[1]**

[1]Department of Cell Biology and Genetics, Faculty of Science, University of Lagos, Akoka, Lagos, Nigeria

[2]Department of Biochemistry, College of Medicine, University of Lagos, Idi-Araba, Lagos, Nigeria

[3]Molecular Biology and Bioinformatics Unit, Biologix Support Services Limited, Anthony, Lagos, Nigeria

*E-mail address: bayulinks@gmail.com

ABSTRACT

Optimizing codons of gene of interest has offered a value-added way to increase heterologous expression of proteins. Different algorithms have been designed to achieve this purpose. These algorithms make use of parameters such as codon Adaptation index (CAI), codon context (CC), percentage guanine-cytosine (GC) content and RNA instability for predicting optimal codons responsible for increased protein expression. Lipase 3646 is an enzyme with great potential for industrial applications. This enzymes has been described to possess thermostable properties with relative stability in high alkaline pH and at different concentrations of organic solvents and inhibitors. This research therefore used JCat, Codon optimization online (COOL), presyndocon and ExpOptimizer algorithms to predict expression level of lipase 3646 enzyme *in silico* by optimizing its gene coding sequence. The results showed that there were variations in the CAI generated by the algorithms for the same 3646 DNA sequence which suggests that each algorithm is specific for its own generated CAI. COOL algorithm prediction based on other parameters showed good results for potential expression of the lipase. Thus, we recommend COOL algorithm for codon optimization of the lipase 3646 gene for industrial applications.

(Received 05 October 2019; Accepted 22 October 2019; Date of Publication 23 October 2019)

***Keywords*:** Codon optimization, algorithm, JCat, COOL, presyndocon, ExpOptimizer, protein expression, lipase, *in silico, Escherichia coli, Cohnella*

1. INTRODUCTION

Enhancing codon usage for increased protein expression is one of the important modules in synthetic biology. Codon optimization is a method of interchanging synonymous codons based on the preferences of heterologous host. This approach makes use of codon engineering used for improving expressing recombinant protein, which is made possible because most amino acids are encoded by multiple codons termed degenerate and there is variation in the mRNAs coding for the same protein which can influence protein expression [1-3]. Weights or points are attached to each codon based on their abundance [4-5], which makes host organism to selectively choose best-fit synonymous codons with higher frequency based on codon bias during polypeptides manufacturing. These weights or points are determined by several parameters specific for each algorithm.

For example, codon Adaptation index (CAI) designed by [6] is the most widely used parameter for analyzing Codon usage bias. The CAI of a gene can be defined as the geometric mean of the weight attached to each codon over the length of the gene sequence which is measured in codons. The CAI measures the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes [7].

Also codon context (CC) associates some "rules" to organize adjacent codons as a result of potential tRNA-tRNA steric interaction within the ribosomes [8-9] but Chung and Lee [2] was first to report computational analysis to evaluate the performance of sequences generated by various CC optimization approaches. Codon context (CC) was previously shown to correlate with translation elongation rate such that the usage of rare codon pairs decreased protein translation rates [10].

Another important feature is the percentage guanine-cytosine (GC) content present in the sequences of genes to be optimized. Rich GC content has been demonstrated to contribute to codon bias as hosts prefer to use codon with rich GC for amino acid synthesis and increase in codon bias could increase protein expression [11-12]. Also, RNA instability has contributed to the increase in translation efficiency. Increase in the RNA instability RNA is less likely to fold on itself, allowing the ribosome easier access for initiation thus increase its efficiency.

Lipases (EC 3.1.1.3) are serine hydrolases and are one of the most significant biomolecules that catalyze important reactions in both aqueous and non-aqueous media. Due to enormous properties of lipases, they are of great importance in industrial applications such as food/dairy, detergent formulation, biodiesel production, cosmetic production, and in the paper and pharmaceutical industries.

They are enzymes that catalyze the hydrolysis of triacylglycerol to glycerol and fatty acids [13-14]. Lipase 3646 has been identified and characterized [15]. The enzyme has been proposed to have thermostable properties with relative stability in high alkaline pH and at different concentrations of organic solvents and inhibitors. Thus, for potential industrial applications, producing higher yield of this enzyme by genetic manipulations would be of tremendous benefit to the manufacturing industry. Different algorithms have been used to optimize coding sequences for different lipase enzymes [16-19], but no available literature has reported codon

optimization for the lipase 3646. Thus, we decided to optimize the codon of lipase 3646 gene using the online algorithms JCat [20] Codon optimization online (COOL) [21], presyndocon [22] and ExpOptimizer [23] and compared their efficiencies for predicting *in silico* protein translation based on the above described features. This study therefore, is expected to be potentially useful when chosen that contribute to increasing the production of this enzyme for industrial applications.

2. METHODOLOGY

2. 1. Retrieval of lipase 3646 nucleotide sequence from database

The gene encoding lipase 3646 originally isolated from thermophilic *Cohnella* sp. A01 (Golaki *et al.,* 2015) was obtained from nucleotide database of National Center for Biotechnology Information website (https://www.ncbi.nlm.nih.gov/nuccore/414303285). The 3646 gene coding sequence has 807 nucleotides. The FASTA format of the sequence was used for this study.

2. 2. *In silico* experiment

Different online codon optimization algorithms were used to optimize the codons of lipase 3646 gene. These include COOL, ExpOptimizer, JCat, and presyndocon. The DNA coding sequence for 3646 was copied from the NCBI database and pasted into each algorithm. The following features were chosen for each algorithm; for codon optimization online (COOL), CAI, CC, total GC content and 5' RNA Folding Instability were chosen. CAI and percentage GC context was automatically generated by ExpOptimizer and JCat. For GeneOptimizer, codon quality and GC content were generated automatically. To avoid potential restriction enzyme digestion, restrictions site for endonucleases Acc65I, BamHI, MluI, NdeI, SalI and SacI were removed from the gene coding sequence by COOL, ExpOptimizer and JCat. Acc65I, NdeI, SalI and SacI sites were removed by presyncodon. *Escherichia coli* [35] was selected in all functions as potential heterologous expression host.

3. RESULTS AND DISCUSSION

Different results were generated based on the imputed features into each algorithm. 50 different combinations were generated by COOL for the optimization.

For all the combinations, The range for each features generated were 58.24 – 71.75%, 0.66301 – 0.95522, -22.3 – -2.5, -0.13421 – -0.12157 for total GC content, CAI, 5’ RNA folding instability and CC fitness (data no shown). Apart from CAI with the best codon fitness score of 1 According to Chin *et al* [24] there is higher potential of increased expression efficiency when the values of 5’ RNA folding instability and CC are closer to zero. The plotted graph in Figure 1A represents the best fit point where the imputed features meet.

The summary report of the data point with arrow revealed the optimal value from each feature selectively, based on 5’ RNA folding instability followed by total GC content fitness, CAI and then CC fitness. The host versus (red) the optimized (blue) codon relative adaptiveness for the arrowed point is represented in Figure 1B. Chung and Lee (2) and Kelwick *et al.* [25] reported that codon context (CC) can significantly increase protein expression.

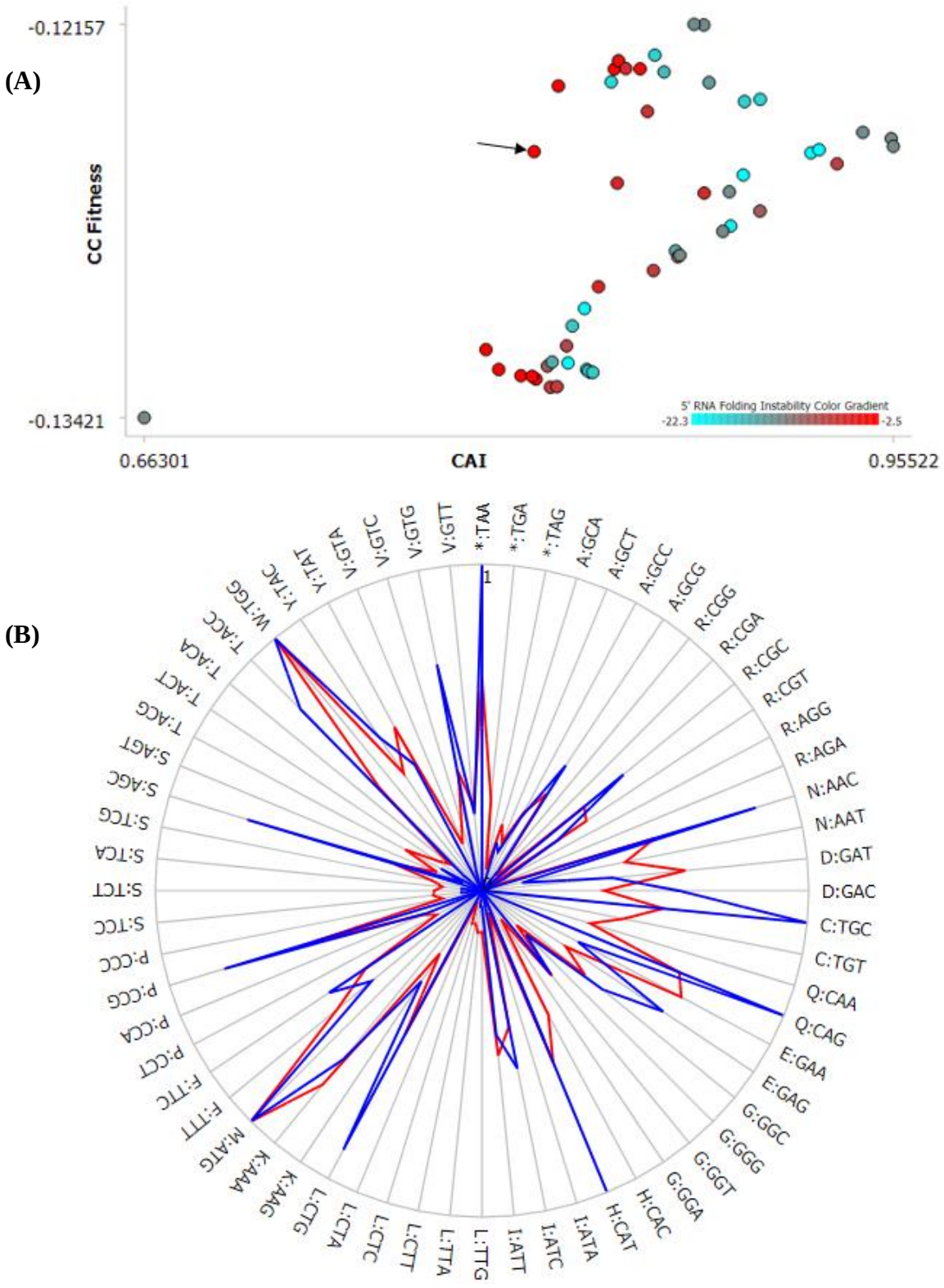


Figure 1. (A) CC fitness against CAI for codon optimized lipase 3646 generated by COOL. **(B)** Host Vs Optimized Codon Relative Frequency Radar Plot

The arrow indicates optimum coding sequence with the following Summary Report: Input Sequence GC content, 62.95%; User Specified Target Total GC content, 100%; Output Sequence Total GC content, 61.83%; Derived Total GC content fitness, -0.38157 Codon Context Fitness (Max), -0.1257; Codon Adaptation Index (Max), 0.8151; 5' RNA Folding Instability (Max), -2.5; 5' RNA Folding Instability Window Size 50 base pair.

The CAI generated by ExpOptimizer as shown in Figure 2 indicates the values before and after optimization were 0.46 and 0.82 respectively while the total GC content after optimization was 55.97%. On the other hand, JCat (Figure 3) revealed 0.19 and 1.00 CAI before and after optimization respectively while the total GC content after optimization was 58.33%. No actual value was reported for the CAI generated by presyncodon. The value was generated by imputing the optimized DNA sequence into COOL, ExpOptimizer, and JCat webpages. The CAI values recorded were 0.60, 0.81 and 0.23 for COOL, ExpOptimizer, and JCat respectively. The total GC content was 52.89% for presyncodon optimized sequence generated from the three algorithms.

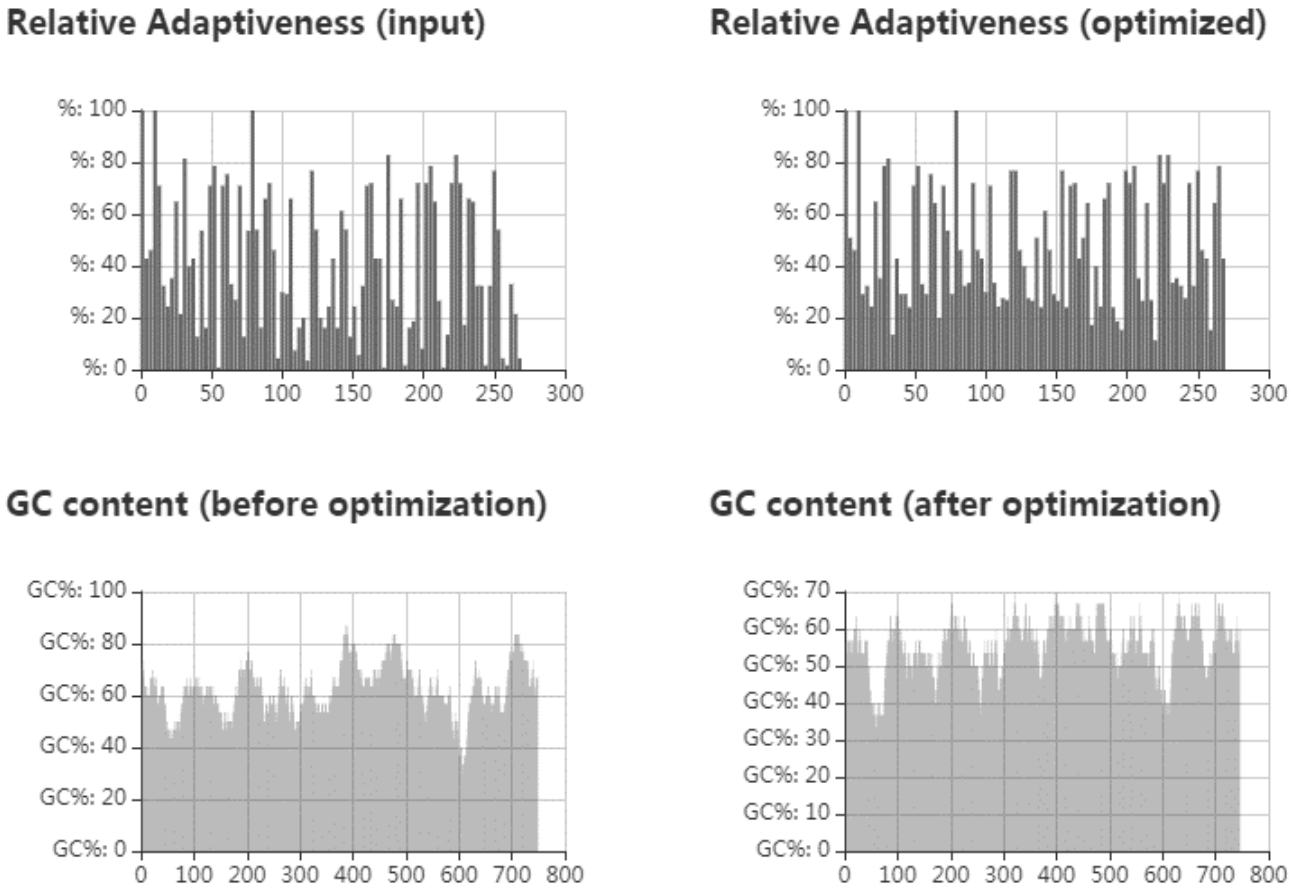


Figure 2. Relative adaptiveness and GC content before and after optimization of lipase 3646 codons generated by ExpOptimizer.

The Summary Report: CAI before optimization, 0.46; CAI after optimization, 0.82; GC content before optimization, 62.95%; GC content after optimization, 55.97%

The CAI and the total GC content of all the optimized sequences generated by the algorithms were compared (Figure 4A and 4B). There are variations in the CAI generated by the different algorithms for the same DNA sequence which is evident in the optimized DNA sequence originally generated by presyncodon and regenerated by other algorithms. This suggests that each algorithm is specific for its generated CAI. Several studies have reported that CAI maximization does not necessarily correlate with high protein yield [2, 26-28] thus CAI may not be only valuable parameter to consider for predicting protein expression level.

Reports on the effect of total GC content on protein expression level have shown that very high G + C content (>70%) may affect mRNA secondary structure formation and can slow down or inhibit translation while very low G + C content (<30%) can inhibit the speed of transcription elongation [26, 29-31]. Although, all the algorithms generated GC content between 53.73 and 61.83% for the imputed and optimized lipase 3646 gene, increase in GC content has been reported in literatures increase protein expression [29, 32]

COOL features 5' RNA instability as one of the determined parameters. It has been reported that translation rate can be positively influenced by RNA instability [33]. Our study predicts that this parameter couple with the codon content could significantly influence translation for lipase 3646 expression which is similar to the report of [34]. Thus, we suggest COOL to have advantage for increasing the expression of lipase 3646 compared to other algorithms used in this study. Future studies would be centered on verifying that COOL can better enhance optimum expression of lipase 3646 *in vivo.*

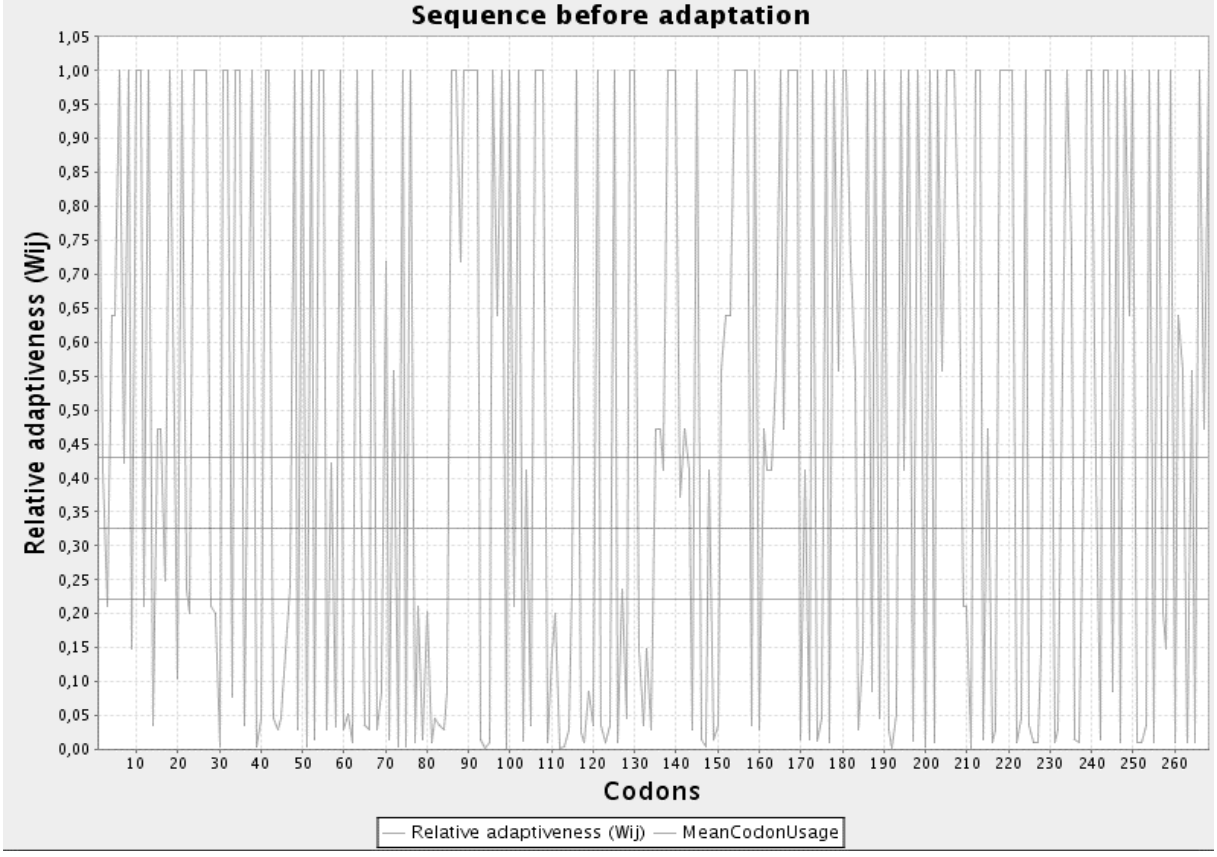


Figure 3. Relative adaptiveness before and after optimization of lipase 3646 codons generated by JCat.

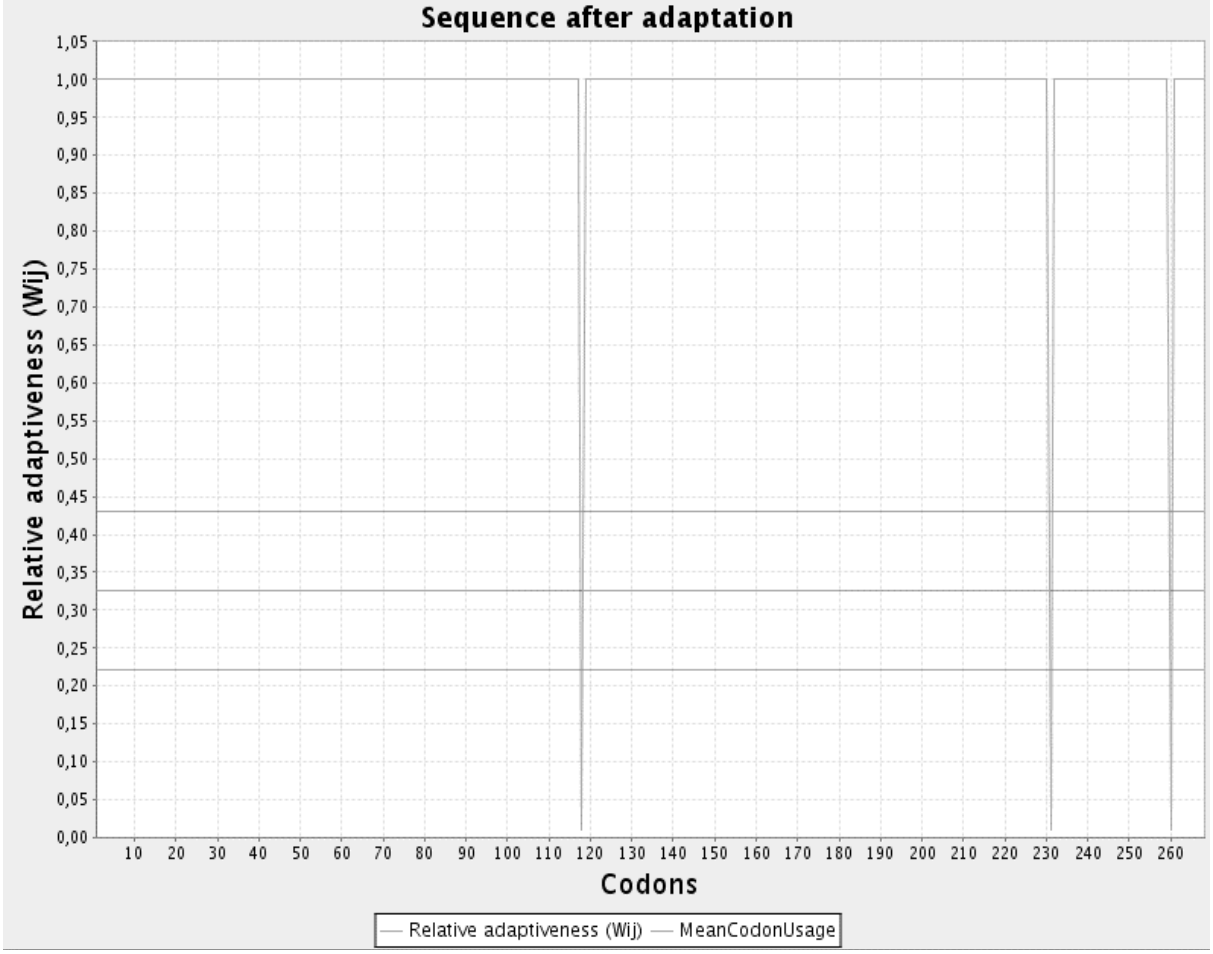
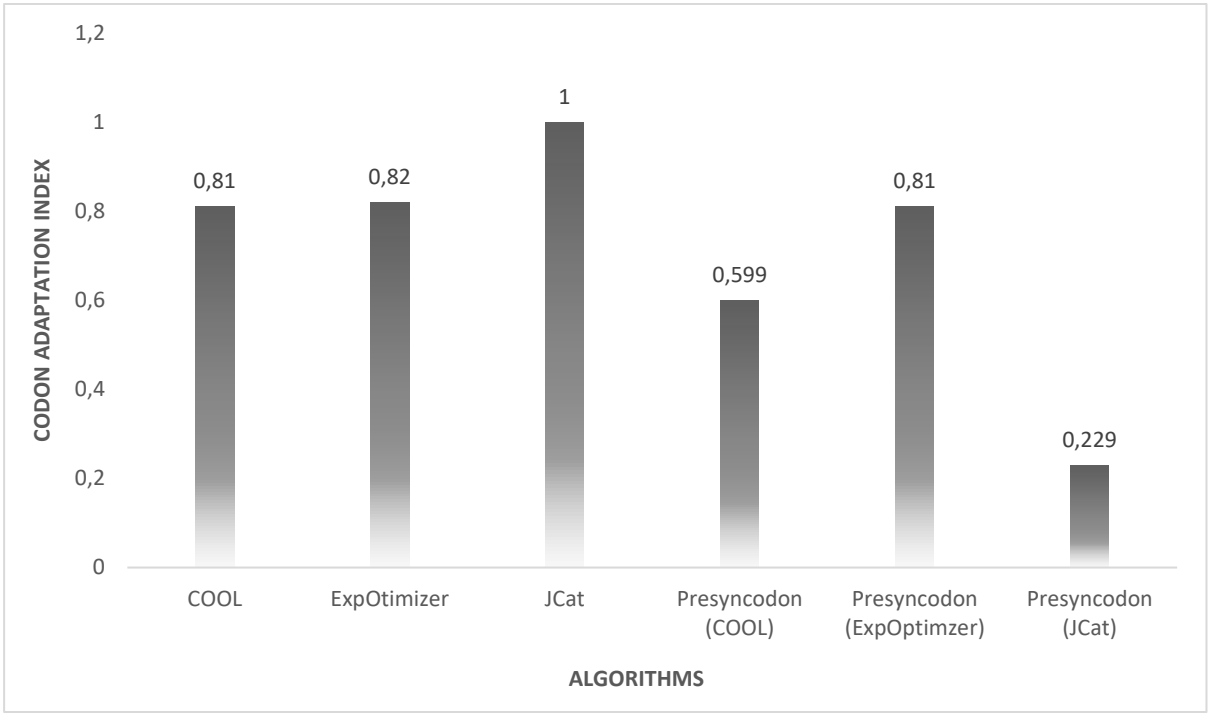


Figure 3(continue). Relative adaptiveness before and after optimization of lipase 3646 codons generated by JCat.

Summary report: CAI-Value of the pasted sequence, 0.19; CAI-Value of the improved sequence, 1.00; GC-Content of the pasted sequence, 62.95%; GC-Content of the improved sequence, 58.33%



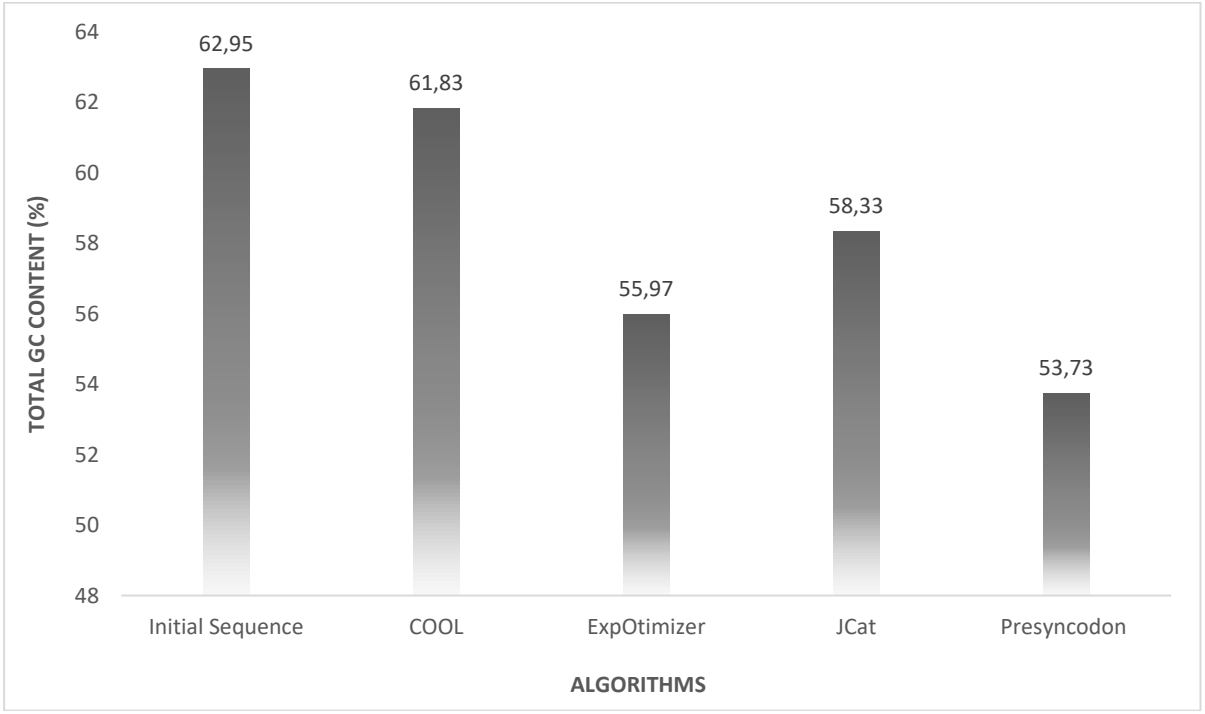


Figure 4. (A) Codon Adaptation Index generation for lipase 3646 gene from different codon optimization algorithms. The COOL ExpOptimzer and JCat were used to generate different CAI for optimized codons pre-generated by presyncodon. **(B)** Total GC content generated for lipase 3646 gene from different codon optimization algorithms.

4. CONCLUSION

Codon optimization is a powerful tool for the enhancement of recombinant protein synthesis and is based on the notion that protein synthesis is limited at the level of translation elongation. In this study, COOL algorithm prediction showed a greater potential for expression of the lipase gene *in silico* based on codon context fitness, total GC content, 5' RNA instability parameters. Thus, we recommend COOL compared to other algorithms used in this study for codon optimization of the lipase 3646 gene for improving protein expression *in vivo* in industrial applications. However, *in vivo* experimental approach is necessary to provide an enhanced data to support this claim.

Acknowledgement

This study was supported by the Central Research Committee Grant CRC/2017/01 of the University of Lagos, Lagos, Nigeria.

References

[1] Menzella, H.G., 2011. Comparison of two codon optimization strategies to enhance recombinant protein production in Escherichia coli. *Microbial Cell Factories*, 10(1), 15.

[2] Chung, B.K.S. and Lee, D.Y., 2012. Computational codon optimization of synthetic gene for protein expression. *BMC Systems Biology*, 6(1), 134.

[3] Mauro, V.P. and Chappell, S.A., 2018. Considerations in the Use of Codon Optimization for Recombinant Protein Expression. In *Recombinant Protein Expression in Mammalian Cells* (pp. 275-288). Humana Press, New York, NY.

[4] Nakamura, Y., Gojobori, T. and Ikemura, T., 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research*, 28(1), 292-292.

[5] Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L.V., Katneni, U., Simonyan, V. and Kimchi-Sarfaty, C., 2017. A new and updated resource for codon usage tables. *BMC Bioinformatics*, 18(1), 391.

[6] Sharp, P.M. and Li, W.H., 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3), 1281-1295.

[7] Jansen, R., Bussemaker, H.J. and Gerstein, M., 2003. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Research*, 31(8), 2242-2251.

[8] Smith, D. and Yarus, M., 1989. tRNA-tRNA interactions within cellular ribosomes. *Proceedings of the National Academy of Sciences*, 86(12), 4397-4401.

[9] Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E. and Mueller, S., 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science*, 320 (5884), 1784-1787.

[10] Singer, G.A. and Hickey, D.A., 2000. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Molecular Biology and Evolution*, 17(11), 1581-1588.

[11] Bohlin, J., Brynildsrud, O., Vesth, T., Skjerve, E. and Ussery, D.W., 2013. Amino acid usage is asymmetrically biased in AT-and GC-rich microbial genomes. *PLoS One*, 8(7), e69878.

[12] Arpigny, J.L. and Jaeger, K.E., 1999. Bacterial lipolytic enzymes: classification and properties. *Biochemical Journal*, 343(1), 177-183.

[13] Jaeger, K.E. and Eggert, T., 2002. Lipases for biotechnology. *Current Opinion in Biotechnology*, 13(4), 390-397.

[14] Saifuddin, N., Samiuddin, A. and Kumaran, P., 2015. A review on processing technology for biodiesel production. *Trends in Applied Sciences Research*, 10(1), 1.

[15] Poornima, K. and Preetha, R., 2017. Biosynthesis of food flavours and fragrances-a review. *Asian Journal of Chemistry*, 29, 2345-2352.

[16] Golaki, P.B., Aminzadeh, S., Karkhane, A.A., Yakhchali, B., Farrokh, P., Khaleghinejad, S.H., Tehrani, A.A. and Mehrpooyan, S., 2015. Cloning, expression, purification, and characterization of lipase 3646 from thermophilic indigenous Cohnella sp. A01. *Protein Expression and Purification*, 109, 120-126.

[17] Chang, S.W., Lee, G.C. and Shaw, J.F., 2006. Codon optimization of *Candida rugosa* lip 1 gene for improving expression in Pichia pastoris and biochemical characterization of the purified recombinant LIP1 lipase. *Journal of Agricultural and Food Chemistry*, 54(3), 815-822.

[18] Xu, L., Jiang, X., Yang, J., Liu, Y. and Yan, Y., 2010. Cloning of a novel lipase gene, lipJ08, from Candida rugosa and expression in Pichia pastoris by codon optimization. *Biotechnology Letters*, 32(2), 269-276.

[19] Zhou, W.J., Yang, J.K., Mao, L. and Miao, L.H., 2015. Codon optimization, promoter and expression system selection that achieved high-level production of Yarrowia lipolytica lipase in Pichia pastoris. *Enzyme and Microbial Technology*, 71, 66-72.

[20] Ghahremanifard, P., Rezaeinezhad, N., Rigi, G., Ramezani, F. and Ahmadian, G., 2018. Designing a novel signal sequence for efficient secretion of Candida antarctica lipase B in E. coli: The molecular dynamic simulation, codon optimization and statistical analysis approach. *International Journal of Biological Macromolecules*, 119, 291-305.

[21] Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C. and Jahn, D., 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, 33 (suppl_2), W526-W531.

[22] Chin, J.X., Chung, B.K.S. and Lee, D.Y., 2014. Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics*, 30(15), 2210-2212.

[23] Tian, J., Yan, Y., Yue, Q., Liu, X., Chu, X., Wu, N. and Fan, Y., 2017. Predicting synonymous codon usage and optimizing the heterologous gene for expression in E. coli. *Scientific Reports*, 7(1), 9926.

[24] Villalobos, A., Ness, J.E., Gustafsson, C., Minshull, J. and Govindarajan, S., 2006. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics*, 7(1), 285.

[25] Kelwick, R., MacDonald, J.T., Webb, A.J. and Freemont, P., 2014. Developments in the tools and methodologies of synthetic biology. *Frontiers in Bioengineering and Biotechnology*, 2, 60.

[26] Kudla, G., Murray, A.W., Tollervey, D. and Plotkin, J.B., 2009. Coding-sequence determinants of gene expression in Escherichia coli. *Science*, 324(5924), 255-258.

[27] Allert, M., Cox, J.C. and Hellinga, H.W., 2010. Multifactorial determinants of protein expression in prokaryotic open reading frames. *Journal of Molecular Biology*, 402(5), 905-918.

[28] Parret, A.H., Besir, H. and Meijers, R., 2016. Critical reflections on synthetic gene design for recombinant protein expression. *Current Opinion in Structural Biology*, 38, 155-162.

[29] Kudla, G., Lipinski, L., Caffin, F., Helwak, A. and Zylicz, M., 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biology*, 4(6), e180.

[30] Bauer, A.P., Leikam, D., Krinner, S., Notka, F., Ludwig, C., Längst, G. and Wagner, R., 2010. The impact of intragenic CpG content on gene expression. *Nucleic Acids Research*, 38(12), 3891-3908.

[31] Newman, Z.R., Young, J.M., Ingolia, N.T. and Barton, G.M., 2016. Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proceedings of the National Academy of Sciences*, 113(10), E1362-E1371.

[32] Li, J., Zhou, J., Wu, Y., Yang, S. and Tian, D., 2015. GC-content of synonymous codons profoundly influences amino acid usage. *G3: Genes, Genomes, Genetics*, 5(10), 2027-2036.

[33] Lorenz, R., Bernhart, S.H., Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L., 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1), 26.

[34] Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K.H., Su, M., Luff, J.D., Valecha, M., Everett, J.K., Acton, T.B. and Xiao, R., 2016. Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature*, 529 (7586), 58.

[35] Abimbola O. Adekanmbi, Adedoyin O. Adeyemi, Oreoluwa M. Olajide. Occurrence of multidrug resistant and extended spectrum β-lactamase (ESBL) – producing Escherichia coli in wastewater of two healthcare facilities in Ibadan, Nigeria. *World News of Natural Sciences* 26 (2019) 167-175