Medical Laboratory Journal
Online ISSN: 2538-4449
Golestan University of Medical Sciences

Original Article

Bioinformatical Analysis of Lipase-Subtilisin Protein Fusion

Neda Gharagozloo Hesari
(PhD Candidate) Department of Microbiology, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran

Davoud Esmaeili
(PhD) Department of Microbiology and Applied Microbiology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran and Applied Virology Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran

Taher Mohammadian
(PhD) Department of Microbiology, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran

Mohammad Hassan Shahhosseiny
(PhD) Department of Microbiology, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran

Atousa Ferdosi
(PhD) Department of Microbiology, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran

Corresponding author: Taher Mohammadian
Address: Department of Microbiology, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran
Tel: +989125267006
Email: tmohammadian@qodsiau.ac.ir

Received: 2020/03/24
Revised: 2020/06/2
Accepted: 2020/06/29

ABSTRACT

Background and objectives: Industrial wastewater is worldwide health concern. Microorganisms present in the environment have an important role in the biodegradation of lipids, fats and proteins from wastewater. In this regard, microbial lipases and proteases are interesting research targets because of high stability, broad substrate specificity, high yields and availability. In this study, we analyze sequences encoding lipase of *Pseudomonas putida* and subtilisin of *Bacillus subtilis* for generation of a new recombinant protein for degradation of environmental contaminations caused by lipids and proteins.

Methods: In this study, sequences of the genes encoding lipase and subtilisin were obtained from GenBank. To predict the 3D structure of the protein, modeling was carried out. The prediction of secondary structure, tertiary *structure* and solvent accessibility was carried using bioinformatics tools including I-TASSER, GoR4 and ExPasy.

Results: The lipase-subtilisin fusion protein was well-characterized by bioinformatical studies with appropriate spatial and secondary structures. The protein had appropriate hydrophilicity, biological half-life and thermal and acidic stability. The codon optimization was performed appropriately.

Conclusion: Overall, the bioinformatical analysis of the designed protein showed that the recombinant lipase-subtilisin protein has a stable structure both in vitro and in vivo, a negative normalized B-factor and lipolytic and proteolytic activities, which makes it suitable for treatment of lipid and protein contaminations.

Keywords: *Pseudomonas putida*, *Bacillussubtilis*, Lipase, subtilisin, Fusion protein, Bioinformatic analysis

DOI: 10.29252/mlj.14.6.23

INTRODUCTION

Bioremediation utilizes microorganisms to make pollutants and degradable biological materials less harmful. Enzymes, particularly microbial enzymes including laccases, oxygenases, proteases, lipases and cellulases have an important role in bioremediation (1). Lipases belong to industrial enzymes that are included in the superfamily of αß hydrolases. Bacterial lipases hydrolyze trioleoyglcerols (2,3). Microorganisms play an important role in the production of extracellular enzymes that are used in different fields, including the bioremediation of wastewater (4). While microbial pollutants are enhanced in wastewater treatment systems, their beneficial activities result in the efficient removal of organic matters such as lipids, proteins and toxic substances. Microorganisms produce a wide spectrum of lipases with variations in substrate specificity, reaction rate, thermal rate, optimum pH, etc. (5-7).

Several *Pseudomonas* and *Bacillus* species are able to produce extracellular lipolytic and proteolytic enzymes. Subtilisins produced by *Bacillus subtilis*, particularly alkaline protease subtilisin, are key extracellular enzymes with industrial applications (7-10). In this study, we analyze sequences encoding lipase of *Pseudomonas putida* and subtilisin of *Bacillus subtilis* for generation of a new recombinant protein for degradation of environmental contaminations caused by lipids and proteins.

MATERIALS AND METHODS

Gene sequences of lipase of *Pseudomonas putida* and subtilisin of *Bacillus subtilis* were extracted from GenBank and then checked by Gene Runner. The construct was designed and was free of any stop codons. NCBI database was used in order to blast the genes. The sequences were cloned into XhoI and BamHI restriction sites of the pET22b vector. ExPASy was used for proteomics, genomics and systems biology analyses (11). The following parameters were predicted: secondary structure, solvent accessibility and normalized B-factor. I-TASSER modeling showed ten threading templates by ITASSER. I-TASSER modeling starts from the structure templates *identified by LOMETS* from the PDB library. LOMETS is a meta-server threading method consisting of multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER only uses templates of the highest significance in the threading alignments, the significance of which is measured by the Z-score. The templates are selected by the LOMETS threading programs (12). The secondary structure and solvent accessibility are available in I-TASSER. Normalized B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. For each target, I-TASSER simulations generate a large ensemble of structural conformations called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity and reports up to three models that correspond to the three structure clusters. The confidence of each model is quantitatively measured by the C-score, which is calculated based on the significance of the threading template alignments and the convergence parameters. C-score is typically in the range of -5 to 2, where a higher C-score signifies a higher confidence and vice-versa (13). TM-score and root-mean-square deviation are estimated based on the C-score and protein length following the correlation observed between these qualities. Since the top three models are ranked by the cluster size, the lower-rank models may have a higher C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower rank models have a better quality than the higher rank models as seen in our benchmark tests. If the I-TASSER simulations converge, it is possible to have less than three clusters generated; this is usually an indication that the models have a good quality because of the converged simulations (12,13). The secondary structure of the protein was predicted by GOR4 analysis. The PredictProtein server was used for sequence analysis and prediction of protein structure and functions, such as low-complexity regions, regions lacking regular structure, secondary structure, solvent accessibility, transmembrane helices, coiled-coil regions, disulfide bonds and subcellular localization as well as functional annotations (13, 14). Extinction coefficient, theoretical isoelectric point, molecular weight, total number of positive and negative residues, half-life, instability index, aliphatic index and grand average hydropath (GRAVY) as well as

physicochemical parameters were computed using the *ExPASy* Logo *ProtParam* tool (http://us.expasy.org/tools/ protparam.html) (15).

RESULTS

The Blast analysis showed that the sequence of lipase –subtilisin is appropriate for the synthesis of the construct. The results showed 100% query cover with an E-value of 0.29. As stated, the B-factor indicates the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template proteins obtained from the PDB library combined with the sequence profiles derived from the sequence databases. The reported B-factor profile corresponds to the normalized B-factor of the target protein, defined by B=(B'-u)/s, where B' is the raw B-factor value, and u and s are the mean and standard deviation of the raw B-factors, respectively (Figure 1). The diagram is below the threshold line and has a negative value, which indicates that the structure has good stability. As shown in (Figure 1). , the Z-score of most of the alignments is lower than one, indicating a suitable structure alignment for the recombinant protein.

Figure 2 shows the structure alignment between the I-TASSER model and the top three similar structure templates in PDB. The C- score of >-1.5 indicates a model of correct global topology. $Cscore^{EC}$ is the confidence score for the enzyme commission (EC) number prediction and usually ranges between 0 and 1, where a higher score indicates a more reliable EC number prediction (Figure 3).

Secondary structure prediction of the recombinant protein using GOR4 software showed abundance of random coils compared to extended strands and alpha-helices.

The results of the ExPasy analysis displayed that alanine is the major amino acid in the new structure. In addition, there are 70 negatively-charged and 56 positively-charged amino acids. The formula of the protein is $C_{3399}H_{6314}N_{924}O_{1044}S_{15}$ and the total number of atoms is 10,696. Extinction coefficients are in units of M^{-1} cm^{-1} at 280 nm and measured in water. The N-terminal of the sequence was met. The estimated half-life was 30 hours in mammals, less than 20 hours in yeasts and 10 hours in vivo. The instability index was computed as 37.50 and the protein was classified in a stable group. The aliphatic index was 87.52 and the GRAVY was predicted to be -0.028.

Figure 1. The B-factor diagram

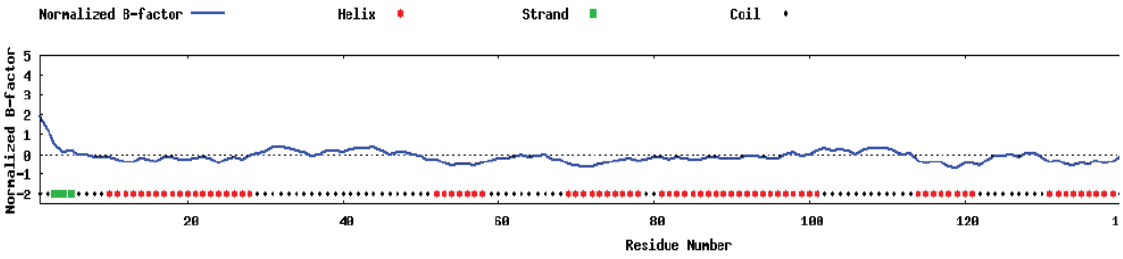


Figure 2. The structure alignment between the I-TASSER model and the top 10 most similar structure

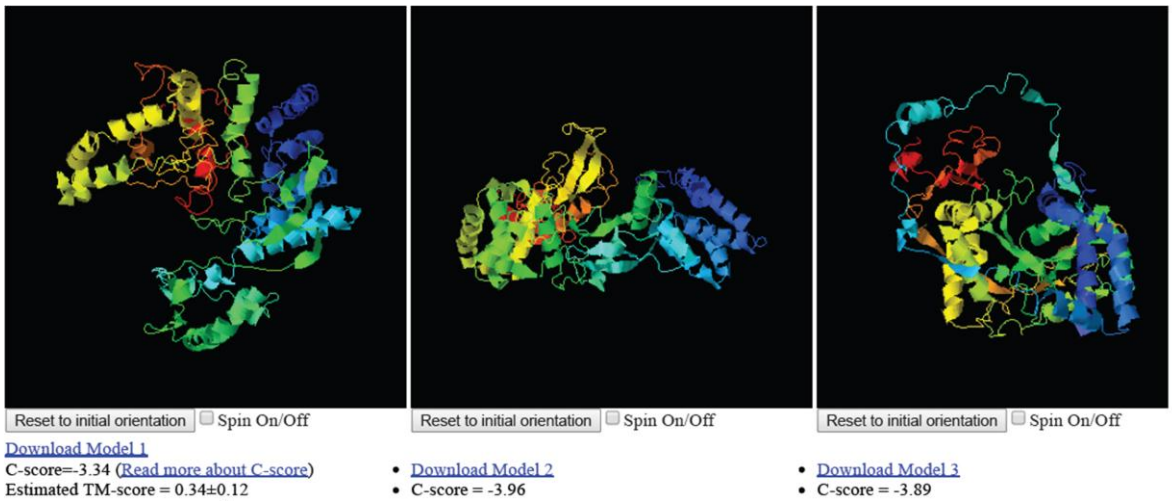
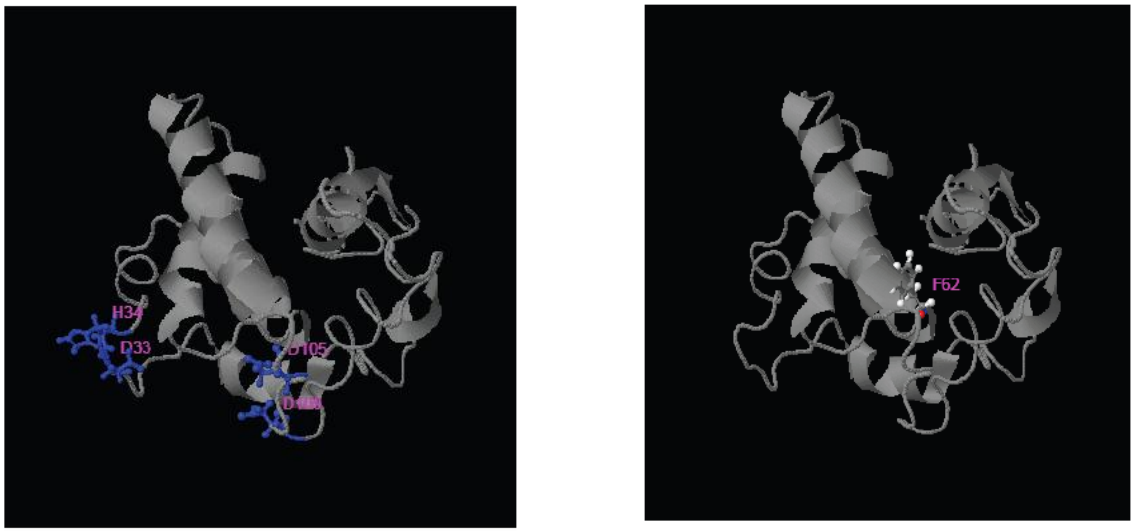


Figure 3. The predicted EC numbers and active sites



DISCUSSION

In recent years, treatment and removal of complex contaminants such as lipids and proteins from the water and wastewater attracted the attention of the scientific community. Imperfect water treatment will create many problems such as clogging of pipes, contamination of natural water resources, death of many organisms and *ecosystem disruption*. Nowadays, microbial enzymes are being used instead of chemical enzymes for various applications. There are numerous methods for the optimization and purification of these enzymes. In the present study, two enzymes of microbial origin were used for potential contaminant removal and wastewater treatment (1,6). In this study two enzymes of Lipase and Protease fused together in order to remove lipid and protein pollutions in water and wastewater.

Based on the results of the *Blast* search, the lipase-subtilisin protein had a high specificity. The E-value varied from 0.29 to 72 with 100% identity in case of subtilisin and 0.29 to 18 with 100% identity in case of lipase. Based on the I-TASSER and GOR4 predictions, the secondary structure of the recombinant protein is mainly composed of coils (47.86%) and helices (36.93%). Another highly relevant but often missed local features in the structure prediction pipelines is the inherent thermal mobility of residues in proteins. At absolute zero temperature, the atoms in a protein are assumed to stay at the equilibrium position of lowest energy, but as the temperature increases, the ambient thermal energy causes the atoms to oscillate around the equilibrium position, the extent of which often varies depending on the relative location on the 3D structure and the interaction with ligand and solvent atoms. The atomic motion can be experimentally measured by X-ray crystallography as a B-factor (or temperature factor), which was introduced as an amendment factor to the structure factor equation since the scattering effect of X-rays is reduced on the oscillating atoms compared to the atoms at rest. Because the distribution of the thermal motion factors in protein crystals can be affected by systematic errors, such as experimental resolution, crystal contacts and refinement procedures, the raw B-factor values are usually not comparable between different experimental structures. Therefore, we calculated a normalized B-factor with a Z-score-based transformation. The B-factor value indicates the extent of the inherent thermal mobility of residues/atoms in proteins.

We obtained negative B-factor values for most parts of the designed protein, indicating the high stability of the structure. I-TASSER predicted the top three models with the rate of C-score, which is acceptable as confidence protein.

I-TASSER utilizes the TM of top ten similar proteins to predict the TM of the target protein (12). In our case, the TM values ranged from 0.603 to 0.879 in similar proteins. This means that the designed protein will have the same TM range. In the prediction of ligands for the designed protein, magnesium was predicted by I-TASSER. Both I-TASSER and GOR4 predicted the same secondary structure. In addition, the most and least abundant amino acids were alanin and valin, respectively. The half-time of the protein was predicted to be 30 hours in vitro. The protein was classified as stable. The GOR4 analysis generated a functional tree that showed lipolytic and proteolytic activities. Based on the GOR4 predictions, the protein was extracellular in all three domains of life (Archaea, Bacteria and Eukaryota) that makes protein extraction easier.

CONCLUSION

Overall, the bioinformatical analysis of the designed protein showed that the recombinant lipase-subtilisin protein has a stable structure both in vitro and in vivo, a negative normalized B-factor and lipolytic and proteolytic activities, which makes it suitable for treatment of lipid and protein contaminations.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding publication of this article.

REFERENCES

1. Dzionek A, Wojcieszyńska D, Guzik U. *Natural carriers in bioremediation: A review.* Electronic Journal of Biotechnology. 2016; 23: 28-36.

2. Jaeger K-E, Ransac S, Dijkstra BW, Colson C, van Heuvel M, Misset O. *Bacterial lipases.* FEMS microbiology reviews. 1994; 15(1): 29-63.

3. Rai S, Vasava H, Upadhaya D. *Bacterial Lipase: A Review.* Studies in Indian Place Names. 2020; 40(71): 200-8.

4. Hasan F, Shah AA, Hameed A. *Industrial applications of microbial lipases.*Enzyme and Microbial technology. 2006; 39(2): 235-51.

5. Abdel-Raouf N, Al-Homaidan A, Ibraheem I. *Microalgae, and wastewater treatment.* Saudi journal of biological sciences. 2012; 19(3): 257-75.

6. Chipasa K, Mędrzycka K. *Behavior of lipids in biological wastewater treatment processes.* Journal of industrial microbiology and biotechnology. 2006; 33(8): 635-45.

7. Grady Jr CL, Daigger GT, Love NG, Filipe CD. *Biological wastewater treatment*. CRC press; 2011.

8. Ali CH, Qureshi AS, Mbadinga SM, Liu J-F, Yang S-Z, Mu B-Z. *Biodiesel production from waste cooking oil using onsite produced purified lipase from Pseudomonas aeruginosa FW_SH-1: Central composite design approach.* Renewable energy. 2017; 109: 93-100.

9. Nadeem M, Baig S, Qurat-ul-Ain S, Qazi J. *Microbial production of alkaline proteases by locally isolated Bacillus subtilis PCSIR-5.* Pakistan Journal of Zoology (Pakistan). 2006; 39(2): 109-114.

10. Shivasharana C, Naik GR. *Ecofriendly applications of thermostable alkaline protease produced from a Bacillus sp. JB-99 under solid-state fermentation.* International Journal of Environmental Sciences. 2012; 3(3): 956-64.

11. Esmaeili D, Daymad SF, Neshani A, Rashki S, Marzhoseyni Z, Khaledi A. *Alerting prevalence of MBLs producing Pseudomonas aeruginosa isolates.* Gene Reports. 2019; 16: 100460

12. Zhang Y. *I-TASSER server for protein 3D structure prediction.* BMC bioinformatics. 2008; 9(1): 40. doi: 10.1186/1471-2105-9-40.

13. Hodaei MH, Anduhjerdi RB, Mehrabadi JF, Esmaeili D. *Cloning and expression of the L1 immunogenic protein of human papillomavirus genotype 16 by using Lactobacillus expression system.* Gene Reports. 2019; 17: 100521.

14. Sen TZ, Jernigan RL, Garnier J, Kloczkowski A. *GOR V server for protein secondary structure prediction.* Bioinformatics. 2005; 21(11): 2787-8.

15. MohabatiMobarez A, Salmanian AH, Hosseini AZ, Esmaeili D. *Clearance of Helicobacter pylori with formulation rCagA and LPS in a mouse model.* Gene Reports. 2020; 19: 100588.

16. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, et al. *ExPASy: SIB bioinformatics resource portal.* Nucleic acids research. 2012; 40(W1): W597-W603.

How to Cite:

This paper should be cited as: Gharagozloo Hesari N, Esmaeili D, Mohammadian T, Shahhosseiny MH, Ferdosi A. [Bioinformatical Analysis of Lipase-Subtilisin Protein Fusion]. mljgoums. 2020; 14(6): 23-27. DOI: 10.29252/mlj.14.6.23.