OPEN ACCESS

GAVIN PUBLISHERS

Advances in Proteomics and Bioinformatics

Research Article

Zafar NUA, et al. Adv Proteomics Bioinform: APBI-104.

DOI: 10.29011/APBI -104. 100004

Computational Prediction of Functional and Structural Aspects of the Protein FTDG_01454 From *Francisella tularensis*

Janneth González-Santos[1*], Raúl A. Poutou-Piñales[2], David Ramirez[3], George E. Barreto[1]

[1]Group of Computational and Structural Biochemistry, Department of Nutrition and Biochemistry, Faculty of Sciences, Pontifical Javeriana University, Bogotá D.C, Colombia

[2]Molecular Biotechnology Laboratory. Group of Environmental and Industrial Biotechnology (GBAI). Department of Microbiology. Science Faculty. Pontifical Javeriana University, Bogotá D.C, Colombia

[3]Center for Bioinformatics and Molecular Simulation, University of Talca. Talca, Chile

[*]**Corresponding author:** Janneth González-Santos, Group of Computational and Structural Biochemistry, Department of Nutrition and Biochemistry, Faculty of Sciences, Pontifical Javeriana University, Bogotá D.C, Colombia. Tel: +5713208320; Email: Janneth.gonzalez@javeriana.edu.co

Citation: González-Santos J, Poutou-Piñales RA, Ramirez D, BarretoGE (2018) Computational Prediction of Functional and Structural Aspects of the Protein FTDG_01454 From *Francisella tularensis*. Adv Proteomics Bioinform: APBI-105. DOI: 10.29011/APBI -105. 100005

Received Date: 02 March, 2018; **Accepted Date:** 10 April, 2018; **Published Date:** 17 April, 2018

Abstract

Francisella tularensis is a pathogenic bacterium that causes tularemia in humans and infects over 250 species of animals. The recent sequencing of the genome of *F. tularensis* allowed the identification of the protein FTDG_01454. Although, the FTDG_01454 protein could be involved in the bacterium pathogenesis its ability to produce disease is not completely understood. Our aim was to predict the 3D structure and function of FTDG_01454 protein by computational tools, and infer related functional domains important for its biological activity. In this study, the primary and secondary structure was analyzed computationally. The 3D structure was predicted by homology modeling and the final model was equilibrated by 40 ns of molecular dynamics simulations. Two domains were identified as hydrolase enzyme catalyst motifs. Our findings demonstrate that FTDG_014 may be related to hydrolytic activity that may provide resistance to antibiotics. These results allow a better understanding on the role of this protein in the physiology of this bacterium.

Keywords: Carboxyl Esterase; FTDG_01454; Hydrolase, *In silico*; Structural Modeling; Structural Bioinformatics; Tularemia

Introduction

Tularemia is a zoonotic disease caused by bacterium *Francisella tularensis*. The bacteria is capable of infecting over 250 species, including humans, other mammals, fish, birds, amphibians, arthropods and protozoa [1]. Humans can become infected through several routes, including: Skin contact with infected animals, Ingestion of contaminated wáter or Inhalación of contaminate aerosol or agricultural dusts [2]. The Genus *Francisella* comprises Gram negative bacilli and includes four *F. tularensis* species with three subspecies including *tularensis* (Type A), *holarctica (*Type B) and *mediasiatica*. The other members of the genus are *F. novicida*, *F. philomiragia* and *F. noatunensis* (formerly known as *F. philomiragia* subsp. *noatunensis*); this species includes two subspecies that are noatunensis and orientalis [3,4]. *F. tularensis* subsp. *tularensis*is the most virulent species and causes severe disease in humans; whereas the disease caused by *F. tularensis subsp. holarctica* is less severe. Infectivity and lethality of *F. tularensis* has made it considered a potential bioterrorism agent [4,5]. *Francisella tularensis*is a not mobile facultative intracellular microorganism, and with aerobic metabolism [5,6,7,] which infects and multiplies "*in vivo*" within macrophages; however, it can also infect and survive in a number of non-phagocytic mammalian cells like hepatocytes, endothelial, epithelial and fibroblast cells; additionally it can survive in the neutrophils [4]. The genus *Francisella* can survive at low temperatures and in the absence of direct sunlight, in a wide number of terrestrial and aquatic habitats[5,8] .

According to the CDC´s classification [2], *F. tularensis* has been classified as a microorganism that can cause massive infections given its high infectivity and virulence. The CDC has analyzed the consequences of a biological attack with *F. tularensis*

and reports that if a cloud of *F. tularensis* was dispersed over a 100.000 population, the number of cases of diseased people to be expected would be 82.500 with around 6.200 deaths; assumed that the majority of cases would be the pneumonic and typhoid forms (the most severe). That is, an infection rate of 82.5% with a mortality of 6.2% [2].

Even though *F. tularensis subsp. tularensis* causes the most serious systemic infection from tularemia, probably due to its intraorganic dissemination capacity [9], the factors that determine its virulence have not been sufficiently studied. Some studies have linked high resistance to lysis of virulent strains with the expression of lipopolysaccharide S type (Smooth),[10,11]. It has also been reported the relationship between virulence and expression ACPA acid phosphatase (E.C. 3.1.3.2), which acts as an inhibitor of the respiratory burst of phagocytes; allowing the bacteria to survive within these cells [12]. Another study has linked the virulence with the expression of stress inducible proteins, which give the bacteria the ability to survive in adverse environmental conditions [13].

The sequencing of the genome of *F. tularensis* [14] allowed the identification of 1741 genes, within which the FTDG_01454 protein coded, was identified as a protein that could be involved in the phatogenesis of the bacteria, making it a possible therapeutic target for immune protection against Tularemia [15].Obtaining 3D conformation of FTDG_01454 can be helpful for effective design of new experiments, such as site-directed mutagenesis, studies of disease-related mutations or structure-based drug design. For this purpose, in this study an "*in silico*" model of the three-dimensional structure of protein was predicted using templates with known 3D structure. In the next step, Molecular Dynamics (MD) simulation studies were performed to infer important functional regions associated with the biological activity. Our results describe some possible functions for FTDG_01454 related with the hydrolytic catalysis of carboxylesterase that would play a very important role in giving it a special antibiotic resistance as it allows the hydrolysis of CO, CN or CC and cellular detoxification of xenobiotics.

Materials and Methods

Computational Analysis of the Primary Structure of FTDG_01454

The protein sequence of FTDG_01454 (Acc.No. EDN38632.1) was obtained through GenBank [16]. Similarity searches with other reported sequences was performed with the BLAST algorithm [17], available in NCBI [16]. The physicochemical properties were represented using amino acid indices [18]. To estimate the amino acid preference and thereby determine the orientation of a protein, we counted the total number of arginines and lysines at a protein's N- and the C-terminal regions. Physicochemical properties, like accessibility profiles, stability index, aliphatic index, overall hydropathy, among others, were calculated with the Proteomic Tools Server [19]. To estimate the level of hydrophobicity of each region of a protein the Kyte-Doolittle scale was used. Individual residue's hydrophobicity scale was summed over the N- and the C-terminal regions. The isoelectric point (pI) and the molecular weight were obtained through ComputpI/Mw program [20]. To identify functional regions, we took the known catalysis of carboxylesterase sequences from PROSITE, built a multiple sequence alignment, and created 10 domain profiles using HMMER. For a query protein, HMMER produces 10 bit scores indicating the probability that the protein contains each of the 10 carboxylesterase domains. To enrich the preference of physicochemical properties in different parts of the protein sequence, the sequence was divided into three regions, the N-terminus, the middle, and the C-terminus.

FTDG_01454 Secondary and Tertiary Structure Prediction

To predict the FTDG_01454 secondary structure of *F. tularensis* we integrate 12 computational prediction algorithms to derive a consensus and having a prediction more accurately than using single predictors (SPOM, SOPMA, HNN, MLRC, DPM, DSC, GORI, GORRI, GORIV, PHD, PREDATOR and SIMPA96). [21]. The used methods are based on parameters determined by the relative occurrence frequencies of each amino acid and the relative frequencies obtained by Bayesian inference. The search for FTDG_01454 protein domains was performed by InterProscan [22].

The 3D structure prediction was performed by homology modeling with the MOE program [23] using as a template the crystal structure of the carboxylesterase (chain A) enzyme identified with phospholipase family (PDB ID: 4F21) [24] the constructed structure was visualized with pymol [25].

The model was subjected to a conjugate gradient energy minimization and 40 ns molecular dynamics simulation (MDs) in Desmond v3.0 using OPLS [26] force field [27]. The protein was solvated by an orthorhombic box of SPC water model, covering the whole surface of the system. Cl^- ions were used as counter ions in order to neutralize the system. Before starting the MDs, the system was minimized and pre-equilibrated using the default relaxation protocol implemented in Desmond. The temperature was maintained at 300°K, while pressure was kept at 1 atm, employing the Nose-Hoover thermostat method with a relaxation time of 1 ps using the MTK algorithm [28] Data were collected every 10 ps during the MDs for further analysis. The 40 ns - MDs were done applying a restraint spring constant of 0.5 kcal × mol-1 $\times Å^{-2}$ to the secondary structure of the protein. The structural evaluation was carried out by Ramachandran Plotvia PROCHECK [29,30]. Further validation of the model was done through flexible loop and side chain refinement.

Results and Discussion

Computational Analysis of the Primary Structure of FTDG_01454

To enrich the preference of physicochemical properties in different parts of the protein sequence, the sequence was divided into three regions, the N-terminus, the middle, and the C-terminus. The physicochemical properties obtained by ProtParam and ComputpI/Mw (Table 1) allowed determining the molecular weight, isoelectric point, instability index, lifetime, aliphatic index and hydropathy. On the other hand, the results obtained from ProtScale(Fig.1) revealed that FTDG_01454 is amphipathic, allowing the identification of its hydrophobic and hydrophilic regions, where the score function residue is observed. The highest peaks correspond to the hydrophobic regions, while the lowest close to -1.5 indicate the hydrophilic regions. It was determined that the protein has 8 hydrophobic regions and 7 hydrophilic regions. The secondary structure predicted showed that the protein has 6 alpha helices corresponding to 30.63 % of the protein, 8 unfolded strands corresponding to 18.47 % and 41.89 % of intercalated loops between helices and unfolded strains According to the molecular weight and the stability index of FTDG_01454, this protein can be classified as a powerful immunogen, given that the compounds with molecular weight between 10 to 100 kDa and stability indexes below 40 can be classified as such[31]. The aliphatic index, which determines the relative volume occupied by aliphatic side chains, showed a high value; this indicates that the protein FTDG_01454 tends to have a high thermostability. Regarding the overall hydropathy, the results show the capacity of FTDG_01454to stablish water interaction; this measure is based on the amount of energy (kal mol^{-1}) that is used to transfer a segment of defined sequence length from a hydrophobic medium to a hydrophilic medium. According to the hydrophilic and hydrophobic regions revealed by Protscale, one part of the structure might be in contact with aqueous medium and the other part might be in contact with the inside of the cell membrane.

Physicochemical properties of the hypothetical protein FTDG_01454 of *F. tularensis*	
Properties	value
Number of amino acids	222
Molecular weight (kDa)	24.64
Stability index	26.89
Isoelectric point	5.43
Time of life (Hours)	30
Aliphatic Index	102.75
Hydropathy overall average	0.081

Table 1: Physicochemical properties of the hypothetical protein FTDG_01454 of *F. tularensis* calculated by using ProtParam.

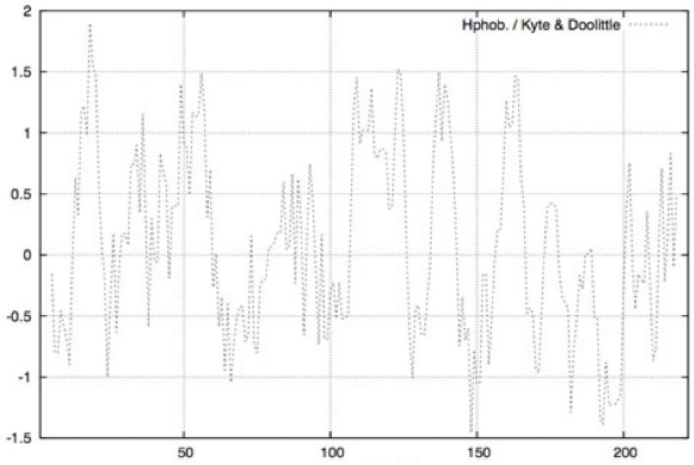

Figure 1: Analysis of the hydropathic of the hypothetical protein FTDG_01454 based on the values of Kyte and Doolittle, using a window of 9 residues.

According to the secondary structure of FTDG_01454, the stability of the protein is clearly evidenced by the prevalence of loops that make these molecules more flexible; this can be explained by the functional need for easy interaction with other proteins [32], which is corroborated by the stability index of 26.89. This finding is an indication that the protein might actually occur naturally [33].

Structural and Functional Aspects

The FTDG_01454 homology model was built using as a template the crystal structure of carboxylesterase/phospholipase family protein from *Francisella tularensis,* and was subjected to 40 ns - restrained MDs. The RMSD of the position for all backbone atoms of the model from their initial configuration as a function of simulation time is illustrated in the Figure 2A. The model was equilibrated around 2 ns of MDs. The RMSD value remains within 2 Å, demonstrating the conformational stability of the protein. Ramachandran plots were used to test how well the φ and ψ angles cluster before and after MDs (Figure 2B). The predicted 3D structure for FTDG_01454 protein is shown in Figure 3, In order to verify that the obtained model is similar to native proteins, a Ramachandran plot was calculated using the Procheck server [29]. As seen in Figure 2B, more than 90 % of the residues were located in the most favored regions of the plot. This shows

that the conformational characteristics of the modeled structure are similar to the native proteins.

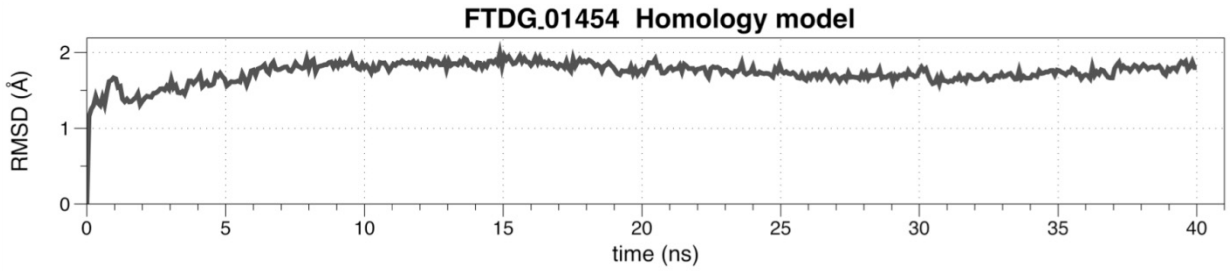

Figure 2A: Results of the 40 ns – MDs. **A.** RMSD for backbone atoms during restricted MDs.

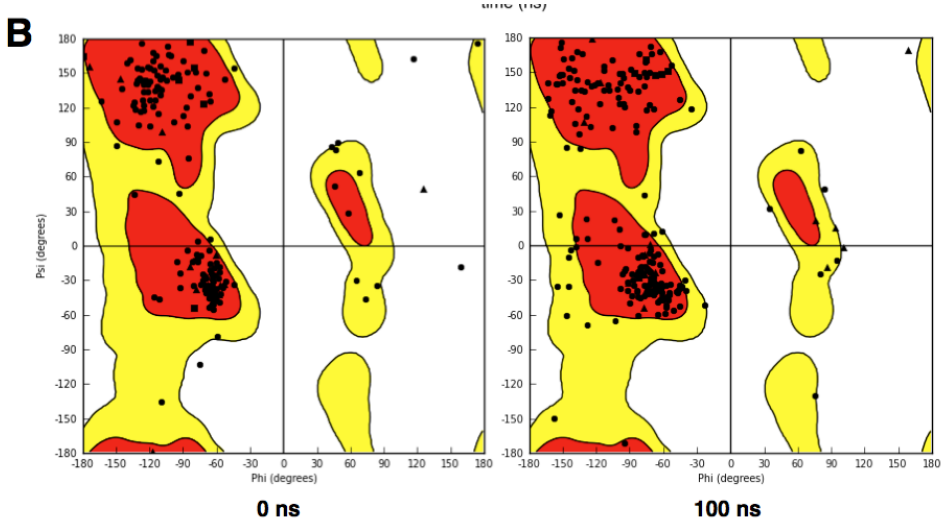

Figure 2B: Ramachandran plots at before (0 ns) and after (40 ns) MDs.

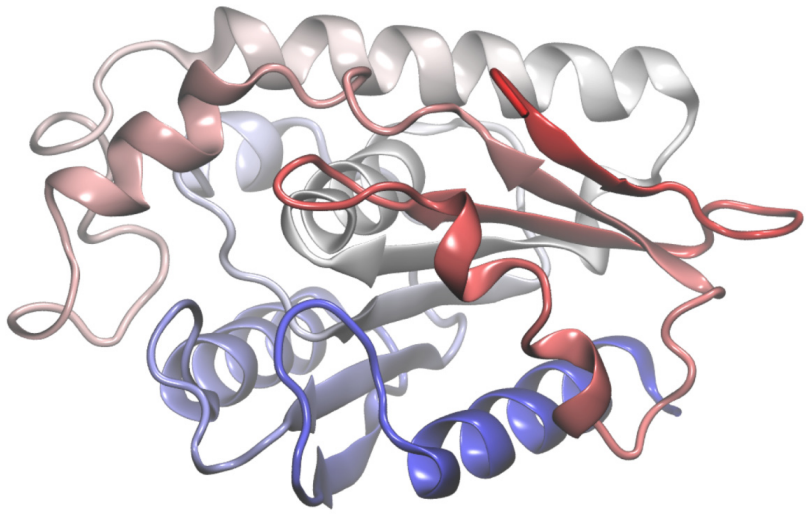

Figure 3: 3D structure of FTDG 01454, generated by homology modeling after 40 ns MDs.

Given that there is not available information for the hypothetical protein FTDG_01454 regarding to its possible involvement in the phatogenesis; the description of molecular functions and structure is a challenging task. The multiple sequence alignment created with 10 domain profiles of the protein sequence FTDG_01454 allowed the identification of Serine hydrolases domains. Serine hydrolases (EC 3.1) are members of the highly conserved α/β-hydrolase enzyme superfamily and are involved in a variety of key biological processes in both eukaryotes and prokaryotes [34]. Members of the serine hydrolase protein family are among current targets to treat diverse bacterial infections [35]. The catalytic triad typically involves three residues: (catalytic triad): a serine, a glutamate or aspartate and a histidine, and often the mechanism involves a nucleophilic attack on a carbonyl carbón atom. As for predicting the function of FTDG_01454 an alpha/ beta hydrolase fold was identified, which core is an alpha / beta sheet with 8 strands connected by helices with folds that form a catalytic triad [36] involved in the non-ribosomal synthesis of peptides.

From the predictable possible functions for FTDG_01454 protein, particularly hydrolytic catalysis of carboxylesterase, it would play a very important role in giving it a special antibiotic resistance as it allows the hydrolysis of CO, CN or CC [37] and cellular detoxification of xenobiotics [38]. The protein FTDG_01454 seems to also have a hydrolytic function of phospholipids for the production of aliphatic monocarboxylic acids. The future study of the carboxylesterase FTDG_01454 of *F. tularensis* and its possible interactions with different ligands and metabolites could generate valuable information that would allow the neutralization of the defense mechanisms of the bacteria and the prevention of the occurrence of biochemical cascades necessary for the development of the tularemia.

Concluding Remarks

Despite intense efforts to characterize the disease caused by *Francisella tularensis* its ability to produce disease is not completely understood. Herein we present a structural and functional study of FTDG_01454 protein from *F. tularensis* to allow a better understanding on the role of this protein in the physiology of this bacterium. Although it has been demonstrated that *F. tularensis* can inhibit the development of the immune response [39,40], molecular basis for the infection by *Francisella* spp., have not been fully understood.

The structural resemblance of the conformation of FTDG_01454 to the Corresponding structural homologuerein forces the hypothesis that FTDG_01454 belong to a family of serine hydrolases and suggests catalytic activity and membrane association. Finally, FTDG_01454is a cytosol protein. Taken together with its involvement in carboxylesterase activities, suggests the need to predict the celular compartment as part of a comprehensive understanding of functionThe structural and functional study of FTDG_01454 could lead us to the understanding and provide an advance in the knowledge for targeting inhibitors of bacterial for the treatment of tularemia.

Conflictof Interest Statement

The authors declare that the research was conducted in the absence of any comercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

This work is part of the projects financed by the Pontificia Universidad Javeriana, Bogotá, D.C., Colombia (ID: 00004960 and 0048078).

References

1. Dai S, Mohapatra NP, Schlesinger LS, Gunn JS (2011) Regulation of *Francisella tularensis* virulence. Frontiers in Microbiology 1: 144.
2. Centers for Disease Control and Prevention C (2013)
3. Huber B, Escudero R, Busse H-J, Seibold E, Scholz HC, et al. (2010) Description of *Francisella hispaniensis* sp. nov., isolated from human blood, reclassification of *Francisella novicida* (Larson et al. 1955) Olsufiev et al. 1959 as *Francisella tularensis* subsp. *novicida* comb. nov. and emended description of the genus *Francisella*. Int J Sys Evol Microbiolv 60: 1887-1896.
4. Lederballe Meibom K, Charbit A (2010) *Francisella tularens* is metabolism and its relation to virulence. Front Microbiol 1: 140.
5. Dennis DT, Inglesby TV, Henderson DA, Bartlett JG, Ascher MS, et al. (2001) Tularemia as a biological weapon. JAMA 285: 2763-2773.
6. Sainz A, Kim CH, Tesouro MA, Hansen R, Amusategui I, et al. (2000) Serological evidence of exposure to *Ehrlichia* species in dogs in Spain. Ann NY Acad Sci 916: 635-642.
7. Morner T (1992) The ecology of tularaemia. Revscitech Off int Epiz 11: 1123-1130.
8. Anda P, Segura del Pozo J, DíazGarcía JM, Escudero R, García Peña FJ et al. (2001) Water borne outbreak of Tularemia associated with crayfish fishing. Emerg Infect Dis7: 575-582.
9. Champion MD (2011) Host-pathogen O-methyltransferase similarity and its specific presence in highly virulent strains of *Francisella tularensis* suggests molecular mimicry. PlosOne 6: e20295.
10. Pavlovich NV, Sorokin VM, Blagorodova NS (1996) The resistance of *Francisella tularensis* to the bactericidal action of normal serum as a criterion for evaluating the virulence of the bacterium. ZhMikrobiol Epidemiol Immunobiol 1: 7-10.
11. Maslova NN, Pavlovich NV (1999) The immunomodulating and antitoxic properties of preparations of lipopolysaccharides from representatives of the genus *Francisella*. Zh Mikrobiol Epidemiol Immunobiol 4: 50-53.
12. Baron GS, Reilly TJ, Nano FE (1999) The respiratory burst inhibiting acidphosphataseAcpA is not essential for the intramacrophage growth or virulence of *Francisella novicida*. FEMS Microbiol Lett 176: 85-90.

13. Basnak'ian IA, Bondarenko VM, Mel'nikova VA, Beliavskaia VA (2001) Stress-inducible bacterial proteins and virulence. Zh Mikrobiol Epidemiol Immunobiol 5: 101-108.

14. Svensson K, Sjödin A, Byström M, Granberg M, Brittnacher MJ et al (2012) Genome sequence of *Francisella tularensis* subspecies *holarctica* strain FSC200, isolated from a child with Tularemia. J Bacteriol 194: 6965-6966.

15. Larsson P, Oyston PCF, Chain P, Chu MC, Duffield M et al. (2005) The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. Nat Genet 37: 153-159.

16. National Center for Biotechnology Information (NCBI) (2013)

17. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. NuclAcid Res 25: 3389-3402.

18. Kawashima S, Ogata H, Kanehisa M (1999) AAindex: Amino acid index database. Nucleic Acids Research 27: 368-369.

19. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G et al. (2012) ExPASy: SIB bioinformatics resource portal. Nucl Acid Res 40: W597-W603.

20. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR et al. (2005) Protein identification and analysis tools on the ExPASy server. Humana Press.

21. Pollastri G, Martin AJ, MooneyC, Vullo A (2007) Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information. BMC Bioinf 8: 201.

22. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N et al (2005) InterProScan: protein domains identifier. Nucl Acid Res 33: 116-120.

23. Chemical Computing Group (2013) MOE, Molecular Operating Environment.

24. Filippova EV, Weston LA, Kuhn ML, Geissler B, Gehring AM et al (2013). Large scale structural rearrangement of a serine hydrolase from *Francisella tularens* is facilitates catalysis. Journal of Biological Chemistry 288: 10522-10535.

25. DeLano WL (2002) The PyMOL Molecular Graphics System. Schrödinger LLC wwwpymolorgVersion 1.

26. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. Journal of Physical Chemistry B 105: 6474-6487.

27. Schrödinger LLC The PyMOL (2007) Molecular Graphics System.

28. Martyna GJ, Klein ML, Tuckerman M (1992) Nosé-Hooverchains: The canonical ensemble via continuous dynamics. The Journal of Chemical Physics 97: 2635-2643.

29. Laskowski RA, MacArthur MA, Moss DS, Thornton JM (1993) PROCHECK - a program to check the stereochemical quality of protein structures. J App Cryst 26: 283-291.

30. De Bakker P, Lovell S (2013) RAMPAGE, Ramachandran Plot Assessment.

31. Arancibia S, Del Campo M, Nova E, Salazar F, Becker MI (2012) Enhanced structural stability of *Concholepas* hemocyanin increases its immunogenicity and maintains its non-specific immuno stimulatory effects. Eur J Immunol 42: 688-699.

32. Sáenz-Suárez H, Chávez-Zobbel A, Lareo LR, Oribio-Quinto C, Martínez-Mendoza J (2011) Predicción computacional de estructura terciaria de las proteín as humanas Hsp27, aB-cristalina y HspB8. Univ Sci 16: 15-28.

33. Guruprasad K, Reddy BVB, Pandit MW (1990) Corelation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering, Design and Selection 4: 155-161.

34. Bachovchin DA, Ji T, Li W, Simon GM, Blankman JL, et al (2010) Superfamily-wide portrait of serine hydrolase inhibition achieved by library-versus-library screening. Proceedings of the National Academy of Sciences 107: 20941-20946.

35. Ortega C, Anderson LN, Frando A, Sadler NC, Brown RW (2016) Systematic Survey of Serine Hydrolase Activity in *Mycobacterium tuberculosis* Defines Changes Associated with Persistence. Cell Chemical Biology 23: 290-298.

36. Carr PD, Ollis DL (2009) Alpha/beta hydrolase fold: an update. Protein Pept Lett 16: 1137-1148.

37. Ollis DL, Cheah E,Cygler M, Dijkstra B, Frolow F, et al (1992) The alpha/beta hydrolase fold. Prot Eng 5: 197-211.

38. Ross MK, Crow JA (2007) Human carboxylesterases and their role in xenobiotic and endobiotic metabolism. J Biochem Mol Toxicol 21: 87-196.

39. Woolard M, Navratil A, Studt A, Bryan J (2011) Phospholipases are required for *Francisella tularensis* induction of macrophage prostaglandin E2 synthesis. J Immunol 186: 110-115.

40. Finn RD, Tate J, Mistry J, Coggill PC, Sammut JS, et al. (2008) The Pfam protein families database. Nucl Acid Res 36: 281-288.