

Insights from the analysis of a predicted model of gp63 in *Leishmania donovani*

Ali Razzazan^{1,*}, Mohammad Reza Saberi² and Mahmoud Reza Jaafari³

¹Department of Medical Physics and Engineering, School of Medicine, Shahid Beheshti University, M. C., Tehran, Iran; ²Department of Medicinal Chemistry, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran; ³Department of Pharmaceutics, School of Pharmacy, Mashhad University of Medical Sciences, Mashhad, Iran; Ali Razzazan* - Email: arazzazan@sbm.ac.ir; Phone: 98 21 22432020; Fax: 98 21 22402021; * Corresponding author

received August 28, 2008; accepted September 13, 2008; published November 02, 2008

Abstract:

Leishmaniasis is a protozoal disease of human that occurs in most parts of the world. By considering the progress of bioinformatics in molecular modeling, major surface glycoprotein of *Leishmania donovani* (gp63) structure was modeled using homology modeling with high accuracy based on the X-ray crystal structure of the *Leishmania major* gp63 as a template, and then analyzed 3D structure of gp63 which can reveal exact facts about its structure and interaction. The objective of this study was to find folding and three dimensional structure of the gp63 as potent antigen for human. In this project, we applied the theory of evolution method, including comparative modeling and threading. This study presented a simple protocol for rapid and precise finding 3D structure of gp63 and investigation of its structural properties. The translated amino acid sequence showed that *Leishmania donovani* gp63 contains 590 amino acids precursor protein consisting of an NH₂-terminal signal peptide of 39 amino acids for membrane targeting, a pro region of 48 amino acids, the mature protein of 478 amino acids containing glycosylation and putative catalytic sites, and a COOH-terminal signal peptide of 25 amino acids for GPI attachment. Based on our model, the protein consists of three domains: the N-terminal, central and C-terminal domains. Additionally, these results could guide future structure-function analyses of gp63 protein.

Keywords: Leishmaniasis; *Leishmania donovani*; gp63; homology modeling; comparative modeling; 3D structure analysis

Background:

Leishmaniasis is a parasitic disease that caused by infection with species of *Leishmania* and occurs in most parts of the world. Visceral leishmaniasis (VL) is caused by *Leishmania donovani* and motivates large-scale epidemics with a high fatality rate. [1] *Leishmania* species possess a membrane glycoprotein on their surface which is called leishmanolysin. This major surface glycoprotein of *Leishmania*, referred to as gp63, is a zinc metalloproteinase of 63,000 Daltons containing a glycosyl phosphatidyl inositol (GPI) membrane anchor, which is displayed abundantly on the surface of promastigotes (~5×10⁵ molecules/cell), however in a less density on the surface of amastigotes. [2] gp63 has been shown to function in the receptor mediated uptake of promastigotes by macrophages in the mammalian host. [3] It is one of the parasite receptors for host macrophages, and parasite mutants lacking the protein are non-virulent. [4] gp63 facilitates survival of the extracellular promastigote stage in the presence of host complement, likely by converting complement component C3b to C3bi. It also promotes attachment of promastigotes to macrophage receptors such as CR3, the receptor for C3bi. [5] There is evidence that gp63 promotes amastigote survival within macrophage phagolysosomes. gp63 protects parasite from lysosomal cytolysis and degradative activities of macrophage by its protease activity. [6] It can also protect liposome-encapsulated proteins from phagolysosomal degradation by macrophages. [7] Thus, gp63 plays a crucial role in macrophage binding to *Leishmania* and in intramacrophage survival and replication. [8] gp63 is a neutral site-specific

endopeptidase cleaving a variety of substrates, [9] such as clusters of differentiation (CD) 4 molecules on human T cells. [10] It is synthesized as an inactive precursor, which is activated via a cysteine switch mechanism [11] and has been reported to hydrolyze a variety of substrates including casein, azocasein, gelatin, albumin, hemoglobin and fibrinogen. [12] gp63 shares several characteristics with the members of the matrix metalloproteinase family including degradation of components of the extracellular matrix such as fibrinogen, location at the cell surface, requirement for Zn²⁺ and sequence similarity of the proposed active site, inhibition of the proteinase activity by chelating agents and α₂-macroglobulin, [13] secretion as a latent form of the enzyme and the activation by mercurial compounds. [14] gp63 is one of the promising candidates for subunit vaccine against leishmaniasis. As the three-dimensional structure of *Leishmania donovani* gp63 was not available, no correlations could have been made between the 3D structure and function of its protection as vaccine. By considering the progress of bioinformatics in molecular modeling, we can model gp63 structure using homology modeling with high accuracy, and then analyze 3D structure of gp63 which can reveal exact facts about its structure, interaction and function.

Methodology:

Searching databases to find right sequences

Target protein was searched by name in the Swiss-Prot and TrEMBL databases through ExPASy server to find raw sequence. Visceral leishmaniasis (Kala azar), caused by

Leishmania donovani, gives rise to large-scale epidemics with a high fatality rate and its treatment is a serious problem and is also widely endemic in Iran. [1] Therefore, among 452 sequences for gp63 in Swiss-Prot and TrEMBL databases, the sequence of *Leishmania donovani* gp63 was selected from Swiss-Prot database.

Comparative modeling

The process of comparative protein structure modeling usually requires the use of many programs, to identify template structure, generate sequence-structure alignments, build and evaluate models. Our results can be divided into four categories: 1) Initial construction of comparative model was achieved by MODELLER and MOE. 2) Refinement of comparative models was facilitated by MODELLER, MOE and Swiss-PdbViewer. 3) Inspection and analysis of models was made by ERRAT, VERIFY3D, PROCHECK and WHAT IF. 4) The models were got better when template selection, target-template alignment and model building were improved.

Finding structures and sequences related to the target sequence

Before any modeling can begin, the sequences and segments with known 3D structures that are related to the sequence being modeled must be found. This was achieved by searching a database of structures that are representative structures of the whole Protein Data Bank (PDB), using BLAST programs through different servers. The similarity search was performed at the SIB by the BLAST network service using NCBI-BLAST 2 program under BLOSUM62 comparison matrix. The similarity search was also carried out through EBI and NCBI servers and the results were quite similar. Another method to find template is fold assignment or threading. It was achieved through FUGUE and 3D-PSSM web servers.

Template selection

Several factors need to be taken into account when selecting a template. The quality of a model increases with overall sequence similarity of template to the target and decreases with the number and length of gaps in the alignment. The simplest template selection rule is to select the structure with the highest sequence similarity to the target sequence. Another important parameter is the length of the template. Therefore, it is necessary to consider family, identity and the number of residues participating in the alignment in order to select the best template. CATH and SCOP databases were used to find the related super family of target protein. At the super family level, proteins probably share an evolutionary relationship, even if they share relatively low amino acid sequence identity in pairwise alignments. Thus, the selected templates must be assigned based on classification. The best result for alignment of gp63 sequence against PDB entries through ExpASy server was obtained.

Alignment of the target sequence and template

3D structure solution does not always end up with full length structure due to the complexity of some regions in X-ray diffraction map. To avoid mismatching between the number of the residues in the sequence and the PDB file, we derived the sequence directly from the related PDB file

through Swiss-PdbViewer not the databases such as SwissProt and TrEMBL. Although MODELLER could do sequence and structure alignment, the alignments were also carefully prepared by other resources such as ClustalX, T-COFFEE, CPHmodels, ESyPred3D, FUGUE and 3D-PSSM. These diverse alignments with different gap regions created various models that had different ERRAT and VERIFY3D scores.

Evaluating the models

After building the models, they checked for possible errors and determined whether or not were acceptable. Initial models were subjected to detailed evaluation, mainly by visual inspection of structural consistency in Swiss-PdbViewer and ViewerLite, secondary structure matching with the secondary structure prediction and using ERRAT program. Two types of evaluation were accomplished: stereo chemical quality and side chain environment. PROCHECK was utilized to assess the quality of the conformation of the polypeptide backbone and side chains using a Ramachandran plot. The steric overlap of atoms (bad clashes) was checked using bump check within WHAT IF, and selection of 'aa Making Clashes' command in Swiss-PdbViewer program.

VERIFY3D and ERRAT were used to evaluate the compatibility between the amino acid sequence and the environment of the amino acid side chains in the model. The result of ERRAT program was the main assignment factor for investigating of the progress in comparative modeling of target protein. Also, there were reasonable agreement between the position of helices and sheets in each generated model and those in the secondary structure prediction by Jpred server. This agreement confirmed the accuracy of modeling procedure for target protein.

Model regeneration and refinement

In the case of unsuccessful models (bad points and scores in evaluation), we stepped backward and realigned or regenerated models sometimes using totally different alignment or template till an acceptable model was created. To clarify the affect of alignment in refining a model an example is given here. For instance, only with the following change in the pairwise alignment of gp63, ERRAT percentage rise from 85.043 to 91.649 presenting huge improvement in model quality.

Discussion:

Overall structure

Gp63 is a compact molecule containing predominantly β sheet secondary structure. This is against of the other protozoan surface antigen structures determined to date. [15] The physico-chemical properties of gp63 calculated by ProtParam (an online tool on the ExpASy server) showed that the exact molecular weight was 62950.3 Dalton, isoelectric point (pI) = 6.41 and the total charge was -5. The three-dimensional structure of gp63 protein was modeled using comparative modeling of this protein based on the X-ray crystal structure of the *Leishmania major* gp63 (PDB code: 1LML) as a template. The alignment of target sequence and template showed 81% sequence identity. According to data in Ramachandran plot of the model, percentage of residues in most favored

regions was above 84%. Consequently, this result suggested the proper geometry for model. Because the model optimized and refined by MODELLER, it had no bump (bad clashes). Furthermore, the result of ERRAT program (overall quality factor = 91.649) showed the gp63 model had acceptable side chain environment. Also, the Accessible Surface Area (ASA) of gp63 model, calculated by RVP-NET program, indicated agreement between the model and the results (Figure 1a). The translated amino

acid sequence shows that *Leishmania donovani* gp63 contains 590 amino acids precursor protein consisting of an NH₂-terminal signal peptide of 39 amino acids for membrane targeting, a pro region of 48 amino acids, the mature protein of 478 amino acids containing glycosylation and putative catalytic sites, and a COOH-terminal signal peptide of 25 amino acids for GPI attachment [16] (Figure 1a). The protein consists of three domains: the N-terminal, central and C-terminal domains (Figure 1b).

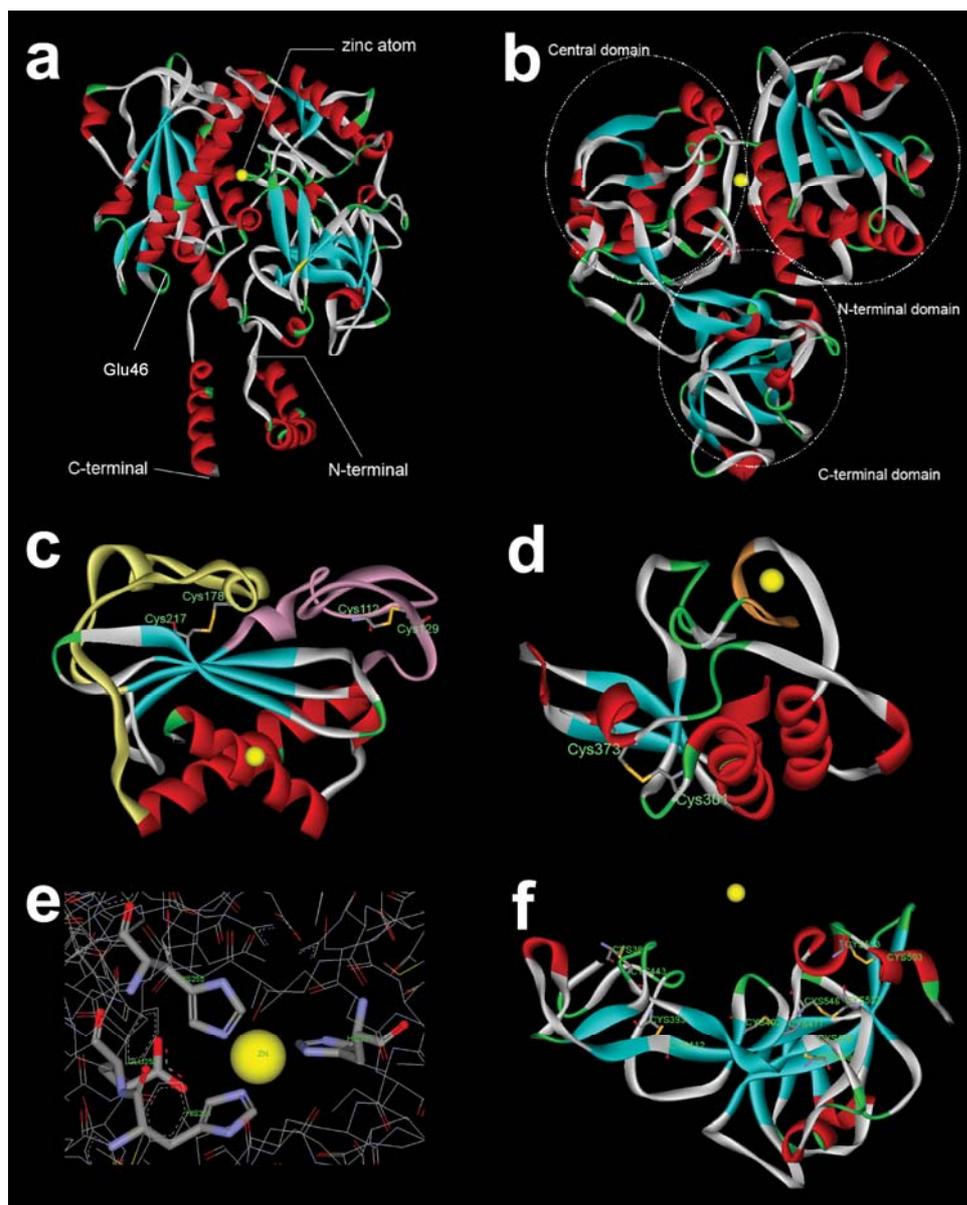


Figure 1: a) Three-dimensional structure of gp63 model based on 1LML, 1CM5, 1TZC and 1HGV. α -helices and β -sheets are in red and blue respectively. Green color represents loop while the grey areas are turns. Zinc atom is shown as a yellow sphere. ASA calculation for Glu46 was 48.4 which indicate to surfaced exposed amino acid. b) 3D structure of gp63 domains. c) The N-terminal domain of gp63 model. Pink ribbon represents the 35-residue flap while the yellow ribbon shows 38-residue flap. Dark yellow stick style indicates disulfide bond. d) The central domain of gp63 model. Orange ribbon represents the met-turn. e) The active site of gp63 model. His251, Glu252, His255 and His321 are within 4 Å radius of the zinc atom. The residues are shown as stick display style and zinc atom as yellow sphere. f) The C-terminal domain of gp63 model.

N-terminal domain

The N-terminal domain includes Val87-Ser260 of the model and has a fold similar to that of the catalytic modules of zinc proteinases. [17] This domain contains a catalytic zinc atom and an active site helix containing the two histidines of the zinc proteinase sequence motif HEXXH (His-Glu-Xaa-Xaa-His). There are deep surface indentations at domain interfaces that join the active site cleft, which is located between the N-terminal domain and the central domain of the molecule. [15] Three-dimensional structures of zinc proteinases containing the HEXXH sequence motif have been grouped into two families: the bacterial thermolysins and the metzincins. [18] Molecules of both classes have a major catalytic domain with a topology consisting of two α -helices (A and B) packed against one side of a five-stranded twisted β sheet. The HEXXH motif is located on helix B and the two histidine side chains are ligands of the zinc atom active site. The N-terminal domain of gp63 includes 35- and 38-residue inserted 'flaps' which cover the side of the β sheet opposite of the helices (Figure 1c). The 38-residue flap replaces the second strand of the central sheet in the conserved topology. [15] There are two disulfide bonds within this domain, one between residues Cys112-Cys129 and the other is between residues Cys178-Cys217 (Figure 1c).

Central domain

Residues Val261-Glu378 of gp63 folds into a compact domain with antiparallel helices. A single disulfide bond (residues Cys301-Cys373) connects the C terminus of the domain to the beginning of the long loop (Figure 1d). Metzincin class zinc proteinases have an extended zinc proteinase motif HExxHxxGxxH, where the glycine residue is part of a tight turn at the end of active site helix B, allowing the extra histidine of the motif to form the third ligand of the zinc atom active site. Metzincins share two additional structural features, an unusual tight 1,4 'met-turn' with a conserved methionine at the base of the active site, and a following helix (helix C), which is roughly parallel to the active site helix B. The central domain of gp63 contains a corresponding met-turn (Asp329-Ala335) (Figure 1d) and a helix C, but there is an unexpected 62 amino acids insertion between the glycine and third histidine residue (Gly258-His321) of the metzincin motif that represents half of the central domain. gp63 is clearly a metzincin class zinc proteinase, but does not satisfy the defining sequence motif HExxHxxGxxH because of the insertion between the glycine (Gly258) and the last histidine (His321).

The active site

To find likely residues involved in the active site, residues within a 4 Angstrom radius of the zinc atom were selected using Swiss-PdbViewer program because the zinc atom has already been identified as a major part of the active site. [15] The zinc atom is coordinated to the nitrogen atoms of His251, His255 and His321. The gp63 zinc active site appears to be tetra coordinated, with Glu252 (Figure 1e).

C-terminal domain

The elongated C-terminal domain (residues Lys379-Asn565) of gp63 consists of mainly antiparallel β strand and random coil structure with only minor helical

contributions. The domain contains six disulfide bonds (Cys380-Cys443, Cys393-Cys412, Cys402-Cys477, Cys454-Cys498, Cys503-Cys553 and Cys523-Cys546) (Figure 1f). The N-terminal part of the domain also contains three two-stranded antiparallel sheets. These sheets are connected *via* two short helices to an antiparallel six-stranded β sheet folded into a sandwich structure and this is in turn attached *via* two short helices to the GPI anchor. [15]

Conclusion:

Although there is no evidence for similar antigenic variation in *Leishmania*, gp63 sequences derived from different species do differ. *Leishmania* sequences are at least 60% identical and the mature proteins share all 18 cysteines, the active site zinc ligand and methionine turn residues with the less conserved gp63 homologues from *Criethidia* and *Trypanosoma brucei*. The *Leishmania donovani* carbohydrate-linked residue (Asn287) and the GPI anchor attachment site at Asn565 are not conserved in *Leishmania* species. As expected, surface residues are less conserved than interior residues, and a surface representation of variability suggests sequence variability is correlated with structural flexibility. In summary, this work provides evidence that the homology modeling approach based on a reliable template could result in accurate and precise structural models that bear significant quality for finding crucial residues which affect protein structure, interaction and function.

Acknowledgment:

The authors wish to thank Dr. Najmeh Zarei (Pharm.D.) for his kind helps and comments in preparation of this manuscript.

References:

- [01] P. Desjeux, *Clin Dermatol.*, 14: 417 (1996)
- [02] P. B. Joshi *et al.*, *Mol Biochem Parasitol.*, 120: 33 (2002) [PMID: 11849703]
- [03] D. G. Russell and H. Wilhelm, *J Immunol.*, 136: 2613 (1986) [PMID: 3950420]
- [04] K. P. Chang *et al.*, *Annu Rev Microbiol.*, 44: 499 (1990)
- [05] A. Brittingham *et al.*, *J Immunol.*, 155: 3102 (1995) [PMID: 7673725]
- [06] J. Alexander *et al.*, *J Cell Sci.*, 112: 2993 (1999) [PMID: 10462516]
- [07] G. Chaudhuri *et al.*, *J Biol Chem.*, 264: 7483 (1989) [PMID: 2708373]
- [08] S. Corradin *et al.*, *Biochem J.*, 367: 761 (2002) [PMID: 12137567]
- [09] H. S. Ip *et al.*, *Mol Biochem Parasitol.*, 40: 163 (1990) [PMID: 2194121]
- [10] A. S. Hey *et al.*, *J Immunol.*, 152: 4542 (1994) [PMID: 7908919]
- [11] M. H. Macdonald *et al.*, *Biochim Biophys Acta.*, 1253: 199 (1995) [PMID: 8519803]
- [12] J. Bouvier *et al.*, *Biochemistry*, 29: 10113 (1990) [PMID: 2271643]
- [13] D. Heuman, *et al.*, *Mol Biochem Parasitol.*, 33: 67 (1989) [PMID: 2469010]
- [14] L. L. Button *et al.*, *Gene*, 134: 75 (1993) [PMID: 8244034]

- [15] E. Schlagenhauf *et al.*, *Structure*, 6: 1035 (1998) [PMID: 9739094]
- [16] J. R. Webb *et al.*, *Mol Biochem Parasitol.*, 48: 173 (1991) [PMID: 1762629]
- [17] W. Stocker and W. Bode, *Curr Opin Struct Biol.*, 5: 383 (1995) [PMID: 7583637]
- [18] W. Stocker *et al.*, *Protein Sci.*, 4: 823 (1995) [PMID: 7663339]

Edited by P. Kanguane

Citation: Razzazan *et al.*, *Bioinformatics* 3(3): 114-118 (2008)

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.