

Predicting The Function Of A Putative Uncharacterized Hypothetical Protein From Leishmania Major

J Watkins

Citation

J Watkins. *Predicting The Function Of A Putative Uncharacterized Hypothetical Protein From Leishmania Major*. The Internet Journal of Genomics and Proteomics. 2012 Volume 6 Number 2.

Abstract

3M3I, a hypothetical protein from Leishmania Major was selected for functional studies using the Basic Local Alignment Search Tool (BLAST) to perform a similarity search through the National Center for Biotechnology Information (NCBI). Comparative analysis disclosed matches to the cupin family and purine nucleotide permease. To further elucidate the identity of 3M3I the STRING database was employed. STRING revealed protein interactions with glyoxalase II and adenosine deaminase. Based on the BLAST and STRING identifications 3M3I appears to be involved in nucleotide metabolism. In addition, analysis revealed possible participation in ATP-dependent protease activity and shuttling within metabolic pathways. This data has laid foundations for further studies to be accomplished linking 3M3I to the enzyme system responsible for the renewal of nucleic acids in Leishmania Major.

INTRODUCTION

Leishmania major is one of a few species of Leishmania, which is apart of the genus of Trypanosomatid protozoa. This intracellular pathogen is responsible for the disease Leishmaniasis, targeting dendritic cells and macrophages. Leishmaniasis has caused prominent morbidity and mortality to populations of more than 80 countries worldwide. Leishmania major infection is characterized by the presence of disfiguring cutaneous lesions. To better understand the pathogenic disease genome-wide and proteomic studies have been extensively underway. Studies have identified the genome of Leishmania as containing 8000 genes. The genes of the species have no introns and no alternative splicing thereby indicating that there should not be many more proteins than there are genes [1].

Proteomics can be defined as the study of global protein expression of biological samples [2]. This area of study employs qualitative and quantitative mass spectrometry approaches to examine global sample sets, which have provided a wealth of knowledge in the understanding of cellular function [3-6]. One of the key outcomes of global proteomic analysis is the identification of new proteins with unknown function. These hypothetical proteins are those having a predicted existence but no experimental data to define their function [7]. In order for a complete analysis and understanding of any proteome to take place we must

unravel all pieces of the puzzle.

In this paper we address the prediction of a hypothetical protein, 3M3I, from Leishmania Major. Elucidating the function and family classification of hypothetical proteins without experimental data involves the combined effort of a number of strategies [8,9]. These include, (1) identification of the sequence homology to proteins of known function, (2) phylogenetic analysis to show evolutionary relationships between biological species based upon similarities and differences in their DNA/protein sequence, and (3) comparative analysis for the identification of probable protein-protein interactions. One of the key points in identifying the function of hypothetical proteins will be to identify the domain(s) found and their organization in order to be able to classify these proteins into a family [10].

MATERIALS AND METHODS

A search of hypothetical proteins was performed at the Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>) and a hypothetical protein of unknown function, 3M3I (225 residues in length), of Leishmania major was selected.

PAIR-WISE ALIGNMENT

The sequence of 3M3I was extracted in FASTA format from PDB. Through the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov>), a similarity search using the Basic Local Alignment Search

Tool program, BLASTp, a protein-to-protein pair wise alignment was performed.

PREDICTED PROTEIN-PROTEIN INTERACTION

A search was carried out using the protein sequence of 3M3I in the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING 9.0). The STRING 9.0 database allowed for the identification of predicted and known protein interactions, encompassing physical and functional associations.

RESULTS AND DISCUSSION

The BLASTp tool found at NCBI identified 100 sequences that shared similarities to hypothetical protein 3M3I. A Max ID cut off of 70% was set identifying the top 6 proteins, (Table I), each representing a hypothetical protein with unknown function found amongst Leishmania. The search revealed 3M3I as possessing close similarity to the cupin super family of domains. To determine if there were any additional structural similarities in the hypothetical protein, a BLAST search was performed against the PDB database, revealing a total of four hits, (Table II), to the Cupin Super Family and the Yml079w Protein From Saccharomyces Cerevisiae. The Conserved Domain Architecture Retrieval Tool (CDART) identified that similarity existed with the Cupin Superfamily.

Figure 1

Table I. The following is a list of proteins identified in the BLASTp search with similarity to the 3M3I.

Accession	Description	Query coverage	E value	Max ident
3M3I_A	Chain A, Hypothetical Protein From Leishmania Major	100%	7.00E-167	100%
XP_043212.1	hypothetical protein [Leishmania major strain Friedlin]	96%	4.00E-160	100%
XP_001468919.1	hypothetical protein, unknown function [Leishmania infantum JPCM5]	96%	6.00E-147	98%
CBZ38010.1	hypothetical protein, unknown function [Leishmania donovani]	96%	2.00E-146	97%
CBZ30654.1	hypothetical protein, unknown function [Leishmania mexicana MHCMB/QT/2001/U1103]	96%	2.00E-130	88%
XP_001568213.1	hypothetical protein, unknown function [Leishmania braziliensis MHCMB/BR/75/M2904]	92%	4.00E-115	78%

Figure 2

Table II. The following is a list of proteins identified from a BLAST search performed against the PDB database

Sequence	Accession No.	Protein Name	No. Entries	% Identity	% Positive	% Gaps
1	EFU11218.1	Cupin Family DUF985	5	39	55	18
2	EGS36471.1	Cupin Family	9	42	56	18
3	ZP_05746794.1	Purine Nucleoside Permease	1	42	56	18
4	1YUD_A	Protein SO0799	1	37	52	20

Based on the searches obtained in Table II, it was clear that many residues identical to 3M3I differed substantially. However NCBI Blast searches revealed matches to a cupin family domain containing protein and two a purine nucleoside permease. A PDB search of the 3M3I sequence

revealed a match to protein 1YUD_A, a chain A X-Ray crystal structure of protein So0799 from shewanella oneidensis (a bacterium capable of reducing poisonous heavy metal and has the ability to live in environments with or without oxygen.). This same match was identified in the NCBI Blastp search and showed 37 % identity and 52% similarity to protein 3M3I as seen in Table II. The protein 1YUD_A is known to contain a cupin family domain motif. The 3M3I protein also from the Blastp search showed 42% identity and 56% similarity to the protein purine nucleoside permease (ZP_05746794.1).

The Cupin Family of proteins were named based on the conservation of a beta-barrel fold, and the Cupins span archea, bacteria and eukaryotes [11]. Although this family of proteins is conserved across these groups there happens to be little sequence homology except for that seen for short partially conserved regions where length and residue sequence vary [11]. The cupins are functionally diverse and consists of enzymatic and non-enzymatic members, which may have one or two cupin domains. They are characterized at the biochemical level, however many are recognized only from gene cloning or sequencing projects [11].

Leishmania, Trypanosoma and Plasmodia are parasitic protozoa. All three of these protozoa are incapable of making purines de novo and are completely dependent on their surrogate hosts for all preformed purines [12]. This is a key biochemical distinction between the vertebrate host and the parasite, as all vertebrates have an established purine biosynthesis mechanism. In order for the parasite to overcome this evolutionary gap they have devised a salvage system encompassing transporters that will allow the intake of an array of purine nucleoside and nucleobases in order to convert them to essential purine metabolites [13]. The Leishmania group of nucleoside and nucleobase transporters are considered to be high affinity concentrative protein-coupled permeases that allow them to compete with the host for components of purine biosynthesis effectively [14, 15,16].

The NCBI search identified matches to purine nucleoside permease (PNP). This identifies the 3M3I protein as a possible cell membrane protein that functions as a mode for the transport of components involved in purine metabolism in or out of the cell. To assist in providing additional information that this protein may be a PNP, the proteomics software tool STRING 9.0 (<http://string-db.org/>) found at the bioinformatics resource portal ExPASy was utilized. STRING is a database of predicted and known protein

interactions, encompassing physical and functional associations. These predictions originate from high-throughput experiments, genomic context, (conserved) co-expression, and PubMed text mining. STRING searched for proteins similar to the 3M3I sequence and identified a total of 20 sequences (Table III).

Figure 3

Table III. The following is a list of proteins in the STRING database with similarity to the 3M3I sequence provided.

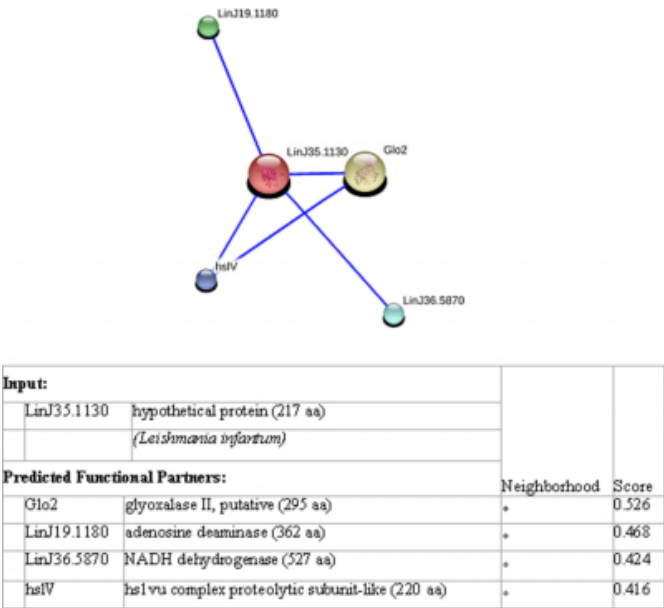
Protein	Species	Annotation	Bitscore	E-value
LinJ35.1130	Leishmania infantum	hypothetical protein	439	1e-122
Bcen2424_6946	Burkholderia cenocepacia 2424	hypothetical protein	147	5e-34
EF_3093	Enterococcus faecalis	hypothetical protein	147	4e-34
LAF_0106	Lactobacillus fermentum	hypothetical protein	144	4e-33
Beph_4665	Lysinibacillus sphaericus	hypothetical protein	144	5e-33
SH2224	Staphylococcus haemolyticus	hypothetical protein	143	7e-33
SDEG_0588	Streptococcus dysgalactiae	hypothetical protein	140	6e-32
SE_0438	Staphylococcus epidermidis 12228	hypothetical protein	139	1e-31
SERP0323	Staphylococcus epidermidis RP62A	hypothetical protein	139	1e-31
SP_2081	Streptococcus pneumoniae R6	hypothetical protein	132	2e-29
SP_2081	Streptococcus pneumoniae D39	hypothetical protein	132	2e-29
SP_2081	Streptococcus pneumoniae 700669	hypothetical protein	130	4e-29
SP_2081	Streptococcus pneumoniae G54	hypothetical protein	129	2e-28
SAS0635	Staphylococcus aureus MSSA476	hypothetical protein	128	3e-28
SP_2081	Streptococcus pneumoniae 70585	hypothetical protein	128	2e-28
SauJH1_0709	Staphylococcus aureus JH1	hypothetical protein	128	2e-28
SP_2081	Streptococcus pneumoniae OGSP14	hypothetical protein	128	3e-28
SAR0681	Staphylococcus aureus MRSA252	hypothetical protein	128	2e-28
SauJH9_0694	Staphylococcus aureus JH9	hypothetical protein	128	2e-28
MOCL_0410	Macrococcus caseolyticus	hypothetical protein	128	2e-28

The top match was to hypothetical protein LinJ35.1130 from Leishmania infantum with an e-value score of 1e-122. This match was then utilized to identify functional partners. In total there were four partners identified, (Figure 1). This data implicates 3M3I as a participant in nucleotide metabolism and metabolic shuttling. The 3M3I protein has a predicted interaction with an enzyme, glyoxalase II which is involved in initiating excision repair at various sites of damaged or improper bases in DNA. In addition a possible involvement in nucleotide metabolism is revealed through an interaction with adenosine deaminase. NADH dehydrogenase has a predicted interaction identifying that 3M3I may participate in shuttling within metabolic pathways. Finally 3M3I may interact with the hslvu complex proteolytic subunit-like revealing a possible connection in ATP-dependent protease

activity.

Figure 4

Figure 1. STRING Neighborhood Network Display – confidence of associations are represented by scores.

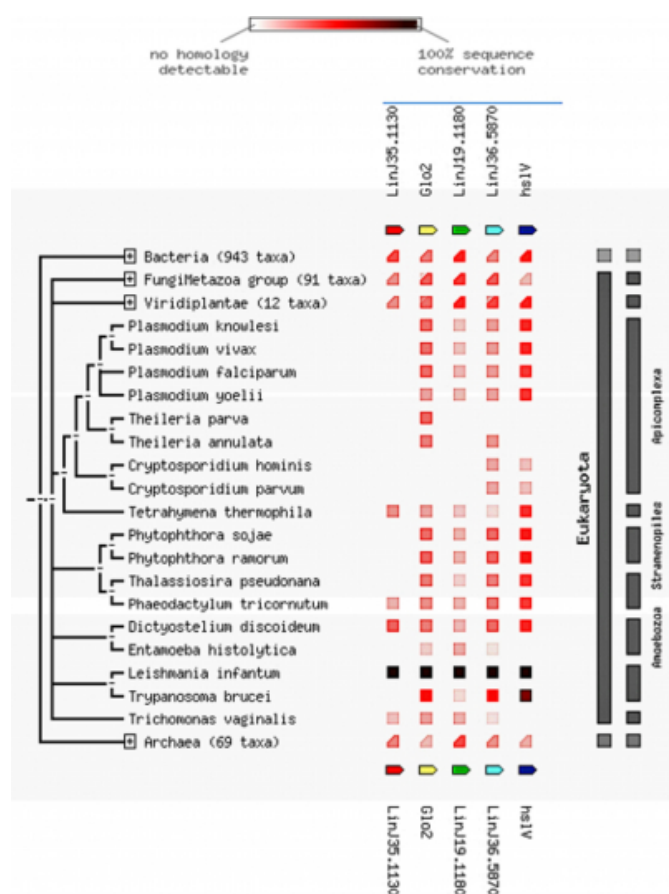


To determine how conservative these sequence hits were within and outside the species, an occurrence overview was carried out (Figure 2). It is important to note that all four partners are conserved within Leishmania, Trypanosoma and Plasmodia, however the LinJ35.1130 protein which had the closest match to our 3M3I protein in the STRING search is only found in Leishmania, identifying the need of a specialized transporter within the nucleotide metabolism pathway unique to this species. The second hit, adenosine deaminase (ADA), is involved in nucleotide metabolism, specifically purine metabolism. ADA irreversibly deaminates adenosine, converting it to the related nucleoside inosine by the substitution of the amino group for a hydroxyl group within the purine metabolism pathway.

This data provides evidence that 3M3I has possible involvement in nucleotide metabolism, however this data is superficial at best and will have to be confirmed. Further studies within Leishmania Major will be required to confirm if and how 3M3I is involved in the continuous cellular renewal of nucleic acids. Data will have to be produced on the enzyme systems participating and 3M3I inclusion in nucleotide renewal. The focus of these studies will elucidate the presence, concentration and interaction of primary and intermediate products in collaboration with 3M3I in nucleotide metabolism.

Figure 5

Figure 2. STRING Occurrence view -This figure identifies which organisms the LinJ35.1130 protein is conserved in.



ACKNOWLEDGMENTS

I would like to thank Mr. Jerry Watkins and Dr. Damon Love for critical review of the manuscript.

References

1. Drummelsmith J, Brochu V, Girard I, Messier N, and Ouellette M: Proteome Mapping of the Protozoan Parasite *Leishmania* and Application to the Study of Drug Targets and Resistance Mechanisms. *Molecular and Cellular*

- Proteomics. 2003 March 1; 2, 146-155
2. Anderson NL, Anderson NG: Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*. 1998 Aug;19(11):1853-61
3. Eberl HC, Mann M, Vermeulen M: Quantitative Proteomics for Epigenetics. *Chembiochem*. 2010 Nov 17.
4. Nanjo Y, Nouri MZ, Komatsu S: Quantitative proteomic analyses of crop seedlings subjected to stress conditions; a commentary. *Phytochemistry*. 2010 Nov 15.
5. Becker CH, Bern M: Recent developments in quantitative proteomics. *Mutat Res*. 2010 Jul 8.
6. Zarembinski TI, Hung LW, Mueller-Dieckmann HJ, Kim KK, Yokota H, Kim R, and Kim SH: Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. USA* 1998; 95:15189-15193.
7. Nan J, Brostromer E, Liu XY, Kristensen O, Su XD: Bioinformatics and structural characterization of a hypothetical protein from *Streptococcus mutans*: implication of antibiotic resistance. *PLoS One*. 2009 Oct 2;4(10):e7245.
8. George RA, and Heringa J: Protein Domain Identification and Improved Sequence Similarity Searching Using PSI-BLAST. *Proteins* 2002; 48:672-81.
9. Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; 215:403-410.
10. Detke S: Cloning of the *Candida albicans* nucleoside transporter by complementation of nucleoside transport-deficient *Saccharomyces*. *Yeast* 14 1257-65 1998.
11. Martha H. Stipanuk, Chad R. Simmons, P. Andrew Karplus and John E. Dominy: Thiol dioxygenases: unique families of cupin protein. *AMINO ACIDS* Volume 41, Number 1, 91-102. From the issue entitled "Sulfur- and selenium-containing amino acids"
12. Hammond DJ, Gutteridge WE. Purine and pyrimidine metabolism in the Trypanosomatidae. *Mol Biochem Parasitol*. 1984; 13:243-261.
13. Carter NS, Yates P, Arendt CS, Boitz JM, Ullman B. Purine and pyrimidine metabolism in *Leishmania*. *Adv Exp Med Biol*. 2008; 625:141-154.
14. Landfear SM, Ullman B, Carter N, Sanchez M. Nucleoside and nucleobase transporters in parasitic protozoa. *Eukaryotic Cell*. 2004; 3:245-254.
15. Ortiz D, Sanchez MA, Koch HP, Larsson HP, Landfear SM. An acid-activated nucleobase transporter from *Leishmania major*. *J Biol Chem*. 2009a; 284:16164-16169.
16. Stein A, Vasudevan G, Carter N, Ullman B, Landfear SM, Kavanaugh MP. Equilibrative nucleoside transporter family members from *Leishmania donovani* are electrogenic proton symporters. *J.Biol. Chem*. 2003; 278:35127-35134.

Author Information

Jermel Watkins, Ph.D.

DNALC Department, Cold Spring Harbor Laboratory