Journal of Biochemistry, Biotechnology and Biomaterials, 1(1), 28-35, 2016

Journal Home page: http://molcare.org/jbcbb

In Silico Structural and Functional Characterization of a Hypothetical Protein of Vaccinia Virus

Malik Siddique Mahmood[1], Naeem Mahmood Ashraf[2], Muhammad Bilal3, Fouzia Ashraf[4], Aadil Hussain[2], Muhammad Zubair Mehboob[2], Iqra Butt[2], Komal Tariq[2], Attha Tallat[2], Kiran Tanveer[1], Ayesha Aslam[1], Umer Rashid[2], Muhammad Naveed[2].*

Abstract

The recent advancements in computational software has increased the chance for the identification of available gene number present in the genome of different organisms. In-silico approach also facilitates to characterize hypothetical proteins with respect to their structure as well as function. These approaches lead the annotation of vaccinia virus genome and to explore many hypothetical genes. This study is designed to estimate the effects of one of the hypothetical protein of vaccinia virus. It exhibit conserved domain of PHA02934 superfamily and its homologues reviled this hypothetical protein a putative apoptotic inhibitor. Secondary structure elements of modeled protein showed a maximum number of random coils (46.46%) with alpha helix (24.34%) and extended strands (29.20%) with high frequency of negatively charged amino acids. The 3D structure prediction revealed the presence of functional BH3 domain having same confirmation in its closest homologue.

Key words: Vaccinia virus. Hypothetical protein, PHA02934, Protein Modeling,

Full length article: Received: 19 Jan, 2016 Revised:10 Feb, 2016 Accepted:10 Feb, 2016 Available online:20 Feb, 2016
Affiliations of authors: *[1]Institute of Biochemistry and Biotechnology, University of the Punjab, Lahore, Pakistan [2]Department of Biochemistry & Molecular Biology, University of Gujrat, Gujrat, Pakistan, [3]Department of Pathology, NSMC, University of Gujrat, Gujrat, [4]Punjab AIDS Control Program, Molecular Diagnostics Section, Jinnah Hospital, Lahore, Pakistan*
***Corresponding Author**: dr.naveed@uog.edu.pk

1. Introduction

Vaccinia viruses belong to the class of poxvirus. These viruses have large linear, dsDNA genome ranges from 130 to 300 kbp which encodes >200 proteins (Broyles, 2003). Poxviruses are unique in that it replicate into the cytoplasm of the host cell (Tolonen *et al.,* 2001). Vaccinia virus during its life cycle forms four infectious forms. These are different in outer membranes, intracellular mature virion (IMV), the intracellular enveloped virion (IEV), the cell-associated enveloped virion (CEV) and the extracellular enveloped virion (EEV) (Smith *et al.,* 2002). It is considered that IMV is the most abundant infectious form that spread between hosts (Smith *et al.,* 2002).

Despite the mild pathogenicity of vaccinia viruses, they also have their role in vaccination as well as they make their usage in research tool (Smith *et al.,* 2002). It first time used to wipe out smallpox by vaccination Vaccinia viruses play major role to eradicate the small pox disease from the world through vaccination (Smith *et al.,* 2003) and still used as a vector for prophylactic vaccine and cell biology study. Numeral fascinating features of vaccinia viruses put forward them as competitive candidates for using them as viral vectors (Hruby, 1990; Smith, 1991).

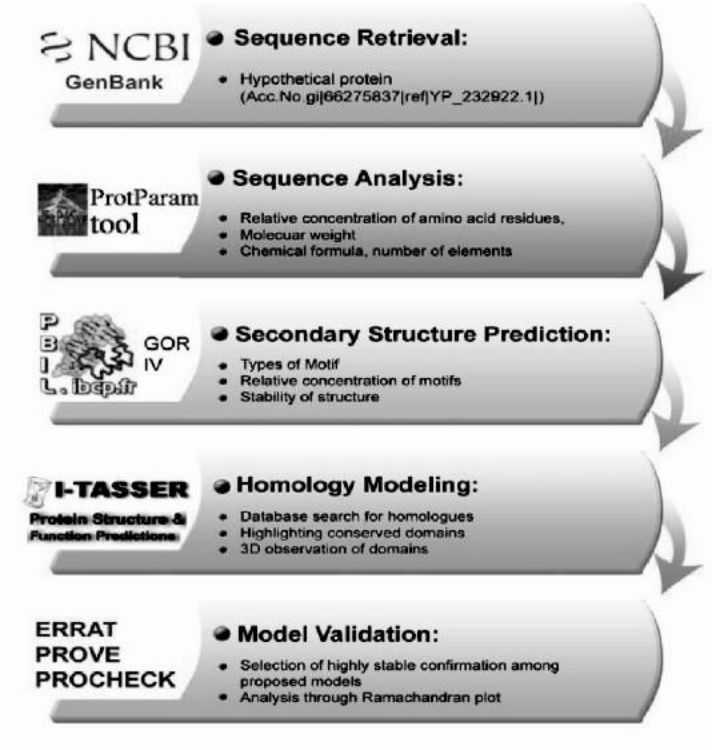
Hypothetical proteins are referred as those proteins which are encoded by an orphan gene whose sequence is known but structurally and functionally is not characterized (Bidkar, 2014). Recent innovation in DNA sequencing automation techniques reveals the bulk amount of information about the genomic sequence of various organisms these techniques provide the evidence for the existence of a great pool of Hypothetical proteins (Eisenstein *et al.,* 2000).

Hypothetical proteins structure and function identification make their way in novel drug discovery, vaccine production and in various metabolic pathways (Paul *et al.,* 2015). Moreover, Protein structural experimental techniques usage become restricted because of much time consuming, high-cost nature and lack of expertise and instrumentation (Feng *et al.,* 2011). Bioinformatics Tools provide a base for finding of Hypothetical protein's structure and function by Homology modeling or domain homology searches with the help of various improved algorithms and software's (Rambabu *et al.,* 2012). Therefore present study, used

bioinformatics tools based on different algorithms to annotate the structural and functional features of the hypothetical protein of vaccinia virus and help to understand the uncharacterized or hidden functions of hypothetical proteins.

2. Material and Methods

2.1. A couple of tools were performed in order to predict the structural and functional analysis of studied hypothetical protein sequence (Scheme 1).



Scheme 1: Methodology for characterization of Hypothetical protein

2.1. Sequence retrieval:

National Centre for Biotechnology Information (NCBI) (Pruitt *et al.,* 2009) based GenBank (Benson *et al.,* 1994) extension was accessed for retrieval of primary residue sequence of the hypothetical protein (Acc.No.gi|66275837|ref|YP_232922.1|).

2.2. Sequence analysis:

Homologous sequences were retrieved, against our protein sequence, present in online database NCBI by using Basic Local Alignment Search Tool (BLAST) (Altschul *et al.,* 1990; Altschul *et al.,* 1997). Furthermore, physiochemical properties' calculations were also made through ProtParam (Gasteiger *et al.,* 2005).

2.3. Secondary structure prediction:

On the basis of physicochemical properties of protein residues, secondary structures were predicted for a hypothetical protein sequence, aided by GOR IV server (Garnier *et al.,* 1996). NCBI embedded BLAST (Altschul *et al.,* 1990) tool was used for the retrieval of homologous sequences as well as to highlight the conserved domains (Marchler-Bauer *et al.,* 2014).

2.4. Homology Modeling:

The closest homologue sequence of hypothetical protein was retrieved through PSI-BLAST against Protein Data Bank (PDB) database (Berman *et al.,* 2000). Furthermore, the Multiple Sequence Alignment (MSA) by Clustal Omega (Sievers *et al.,* 2011) of the query with homologue was also performed (results not shown) to explore the conserved residues among homologue sequence.

The 3D model was predicted, by submitting primary residue sequence to a web-based 3D model prediction tool, I-TASSER (Yang *et al.,* 2015).

2.5. Model Validation:

The predicted 3D models were subjected to PROVE (Pontius *et al*., 1996) and ERRAT (Colovos and Yeates, 1993) in order to validate the structural confirmations as well as by Ramachandran Plot from PROCHECK (Laskowski *et al*., 1993). The 3 D model visualization was done by PyMOL (DeLano, 2002).

3. Results and Discussion

The sequence of the hypothetical protein of Vaccinia virus, retrieved from GenBank, revealed that protein is comprised of 226 amino acid residues. The BLAST results explored the homologue sequences of this protein belonging to two superfamilies' of proteins; Hypothetical protein provisional and Apoptosis Regulator M11L like proteins (Figure 1).

Table 1: Relative intensity of Amino Acid residues throughout the sequence

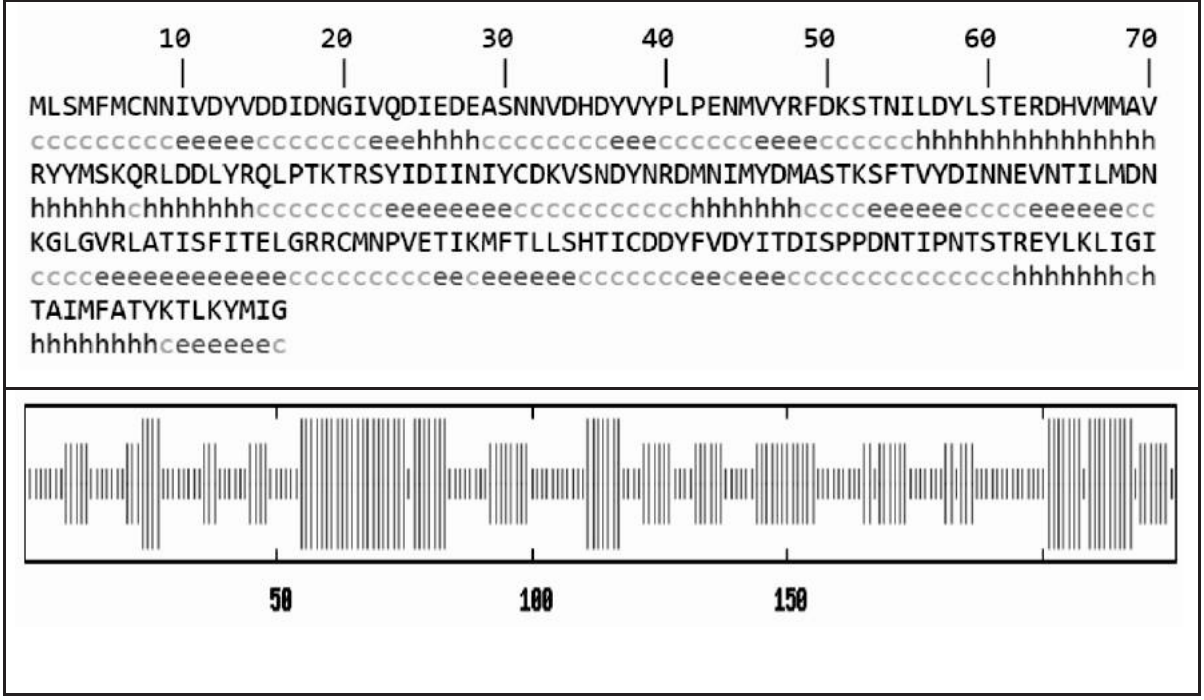
Amino acid	No. of residues	% of residues
Ala (A)	6	2.7%
Arg (R)	11	4.9%
Asn (N)	18	8.0%
Asp (D)	25	11.1%
Cys (C)	4	1.8%
Gln (Q)	3	1.3%
Glu (E)	8	3.5%
Gly (G)	6	2.7%
His (H)	3	1.3%
Ile (I)	23	10.2%
Leu (L)	16	7.1%
Lys (K)	10	4.4%
Met (M)	15	6.6%
Phe (F)	7	3.1%
Pro (P)	7	3.1%
Ser (S)	13	5.8%
Thr (T)	19	8.4%
Trp (W)	0	0.0%
Tyr (Y)	18	8.0%
Val (V)	14	6.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

The ProtParam categorized the residues on the basis of their physicochemical properties. The molecular weight of protein was 26428.1 with theoretical pI of 4.58. The most frequent amino acids were Aspartate, Isoleucine, Threonine, Asparagine and Tyrosine with 11.1%, 10.2%, 8.4% 8.0% and 8.0% respectively while the Glutamine, Histidine and Cysteine were present with minimum percentages of 1.3%, 1.3% and 1.8% respectively and Tryptophan was completely absent. The sequence contains 33 negatively charged (Asp+Glu) and 21 positively charged (Arg+Lys) amino acids. The atomic composition includes 3664 atoms having Carbon (1167), Hydrogen (1818), Nitrogen (296), Oxygen (364), Sulfur (19) atoms with chemical formula $C_{1167}H_{1818}N_{296}O_{364}S_{19}$. Aliphatic and instability Index was 87.92 and 41.13 respectively which declared this protein as unstable. The calculated Grand

Average of Hydropathicity (GRAVY) was about -0.247 (Table 1).

Table 2: Motif Predictions of linear residue sequence

Structural elements	Number of residues	Percentage of residues
Alpha helix (Hh)	55	24.34%
310 helix (Gg)	0	0.00%
Pi helix (Ii)	0	0.00%
Beta bridge (Bb)	0	0.00%
Extended strand (Ee)	66	29.20%
Beta turn (Tt)	0	0.00%
Bend region (Ss)	0	0.00%
Random coil (Cc)	105	46.46%
Ambiguous states (?)	0	0.00%
Other states	0	0.00%



The findings of secondary structure predictions, by GOR IV server, were as 105 residues with maximum abundance (46.46%) were supporting random coils to join 24.34% Alpha helices (55 residues) and 29.20% extended strands (66 residues) (Table 2). The structure lacked beta helices, beta bridges and beta turns. The high abundance of random coils is responsible for high stability and conservancy of protein (Neelamathi *et al.,* 2009)

Two conserved domains were found in studied protein sequence by Conserved Domain (CDD) BLAST search (Marchler-Bauer *et al.,* 2011). It declared M11L domain (pfam11099) a non-specific hit against our query (69-225 residues)

Table 3: Protein Homologues with closest hits by using BLASTP

Accession ID	Similar hits	Score	Query Cover	E-value	Identity
P68450.1	AAQ93136.1	452	100%	6e-159	99%
AIX99157.1	AIX99362.1	451	100%	2e-158	99%
AAW23437.1	AAW23719.1	450	100%	2e-158	99%
ABD98510.1	AEY72848.1	450	100%	2e-158	99%
Q9JFF2.1	AAF33891.1	449	100%	7e-158	99%

Table 4: Protein homologues found in PDB through BLAST

GI Number	PDB ID	Score	Query Cover	E-value	Identity
192988273	2VTY_A	336	74%	4e-117	98%
645985629	2D2L_A	334	74%	2e-116	97%
645985632	2D2M_A	330	74%	7e-115	96%
783283464	5AJK_A	311	72%	7e-108	93%
783283460	5AJJ_A	309	72%	6e-107	92%

with E-value 2.31e-52 (Table 4). This domain is responsible for an apoptotic regulatory function to modulate the apoptotic cascades and ultimately favors viral replication. M11L inhibits mitochondrial-dependent apoptosis by competing and mimicking the host proteins for blocking and binding with Bax and Bak, the key apoptotic factors (Douglas *et al.,* 2007). The specific hit explored by CDD-BLAST results was Hypothetical protein provisional (PHA02934 superfamily) with E-value: 1.45e-136 [Pssm ID: 165245, Accession No. cl19668]. This domain covers 1-225 residues of our protein sequence (Figure 1).

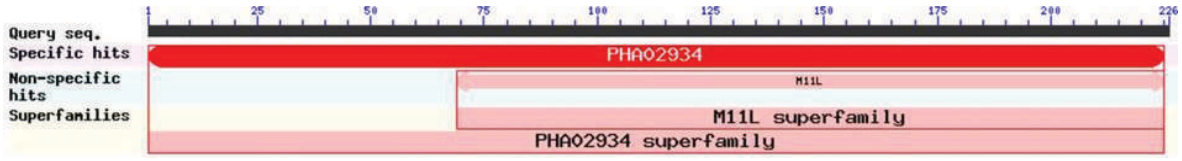


Fig. 1. Results found by Conserved Domain Search (CDD) using BLAST

The closest homologue obtained from BLASTP results was the Protein F1 of Vaccinia virus (Table 3). It also contains the same PHA02934 conserved domain. This protein interacts with pro-apoptotic BCL2L11 protein (a host protein) via BH3 domain and favors the viral replication by inhibiting apoptosis of infected cells (Kvansakul *et al.,* 2008). BLAST search against PDB also provided a 3D structure of Anti-Apoptotic F1 protein (GI: 192988273) of Vaccinia Virus (Table 3). Five proposed structural models from I-TASSER were further evaluated by different software. ERRAT evaluated one of the models with 95.872 overall quality factor and also approved by PROVE on the basis of Z-score value (0.44). Ramachandran plot proposed by PROCHECK (Figure 2) shows that majority of amino acid residues lies in the regions of highly, moderately and less stable confirmations, except following four residues, Met[4], Cys[7], Tyr[58] and Tyr[223], that lie in forbidden regions of plot.

The 3D visualization of hypothetical protein explored that the structure is only comprised of alpha helices, random coils and extended strands (Figure 3a). The structure was superimposed with closely related F1 protein (2VTY_A) to check for helices orientations (Figure 3b). The superimposed model revealed that functional domain (BH3 binding site) carries similar orientation in hypothetical protein as in F1 protein (Figure 3c), while the non-overlapping residues form extra helices and loops away from functional BH3 binding helix and do not hinder its function (Figure 3b).

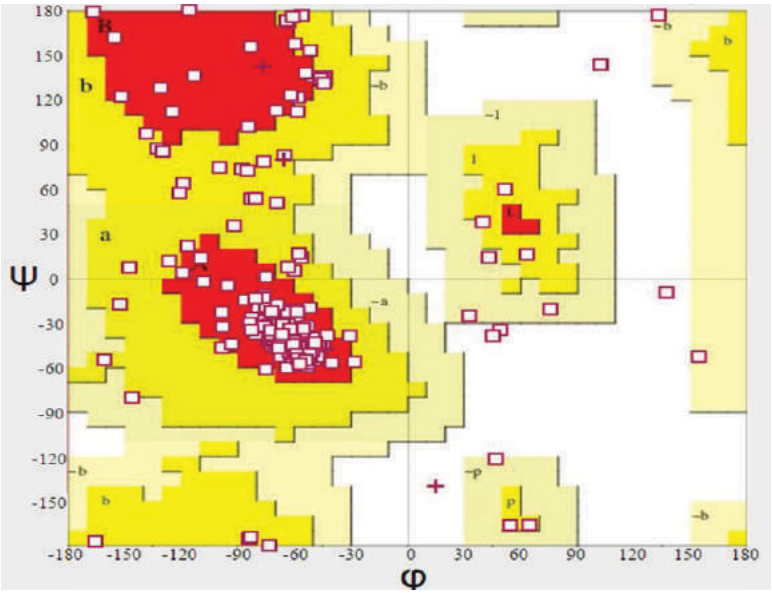


Fig. 2. Ramachandran Plot of hypothetical protein residues (Red region= highly stable confirmation, Yellow region= moderately stable confirmation, Skin region= less stable confirmation, White region= forbidden area)

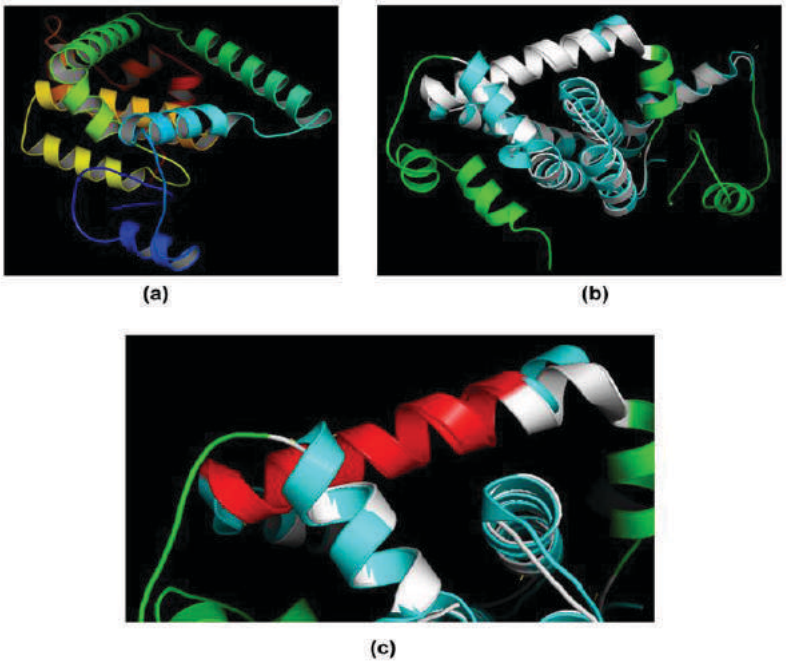


Fig. 3. (a). 3D model of a hypothetical protein of Vaccinia Virus. (b) Superimposed structural view of a hypothetical protein (white) with F1 protein (Cyan) of Vaccinia Virus with the non-overlapping region of the hypothetical protein. (c) a close view of overlapped conserved domain.

The structural based functional annotation of hypothetical proteins is very much helpful for recombinant technologist in vector designing strategies to eliminate the unnecessary sequences from vector construct and to make the room for exogenous DNA. These in-silico studies can let a researcher know the pros and cons of lab work prior to characterizing an unknown protein sequence and ultimately urges for a directed research to become a competent researcher.

6. References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990). "Basic local alignment search tool." Journal of molecular biology 215(3): 403-410.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipma 1997. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research 25(17): 3389-3402.

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. 1994. "GenBank." Nucleic Acids Research 22(17): 3441-3444.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne. 2000. "The protein data bank." Nucleic Acids Research 28: 235-242.

Bidkar, A. K. K. Thakur, N. B. Bolshette, J. Dutta and R. Gogoi. 2014. "In-silico Structural and Functional analysis of Hypothetical proteins of *Leptospira Interrogans*." Biochem Pharmacol 3: 2167-0501.1000136.

Broyles, S. S. 2003. "Vaccinia virus transcription." Journal of General Virology 84: 2293-2303.

Colovos, C. and T. O. Yeates 1993. "Verification of protein structures: patterns of nonbonded atomic interactions." Protein science: a publication of the Protein Society 2: 1511.

DeLano, W. L. 2002. "The PyMOL molecular graphics system."

Douglas, A. E., K. D. Corbett, J. M. Berger, G. McFadden and T. M. Handel. 2007. "Structure of M11L: A myxoma virus structural homolog of the apoptosis inhibitor, Bcl-2." Protein science 16: 695-703.

Eisenstein, E., G. L. Gilliland, O. Herzberg, J. Moult, J. Orban, R. J. Poljak, L. Banerjei, D. Richardsonm and A. J. Howard. 2000. "Biological function made crystal clear—annotation of hypothetical proteins via structural genomics." Current opinion in biotechnology 11: 25-30.

Feng, W., G. L. Gilliland, O. Herzberg, J. Moult, J. Orban, R. J. Poljak, L. Banerjei, D. Richardsonm and A. J. Howard. 2011. "Combination of NMR spectroscopy and X-ray crystallography offers unique advantages for elucidation of the structural basis of protein complex assembly." Science China Life Sciences 54: 101-111.

Garnier, J., J. F. Gibrat and B. Robson. 1996. "GOR secondary structure prediction method version IV." Methods Enzymol 266: 540-553.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch. 2005. Protein identification and analysis tools on the ExPASy server, Springer.

Hruby, D. E. 1990. "Vaccinia virus vectors: new strategies for producing recombinant vaccines." Clinical microbiology reviews 3: 153-170.

Kvansakul, M., *et al.,* (2008). "Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands." Cell Death & Differentiation 15(10): 1564-1571.

Laskowski, R. A., H. Yang, W. D. Fairlie, P. E. Czabotar, S. F. Fischer, M. A. Perugini, D. C. Huang, P. M. Colman. 1993. "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of applied crystallography 26: 283-291.

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng, S. H. Bryant. 2014. "CDD: NCBI's conserved domain database." Nucleic Acids Research: gku1221.

Marchler-Bauer, A., S. Lu, B. J. Anderson, F. Chitsaz, M. K. Derbyshire, C. D. Weese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, I. D. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng and S. H. Bryant. 2011. "CDD: a Conserved Domain Database for the functional annotation of proteins." Nucleic Acids Research 39(suppl 1): D225-D229.

Neelamathi, E., E. Vasumathi, S. Bagyalakshmi, and R. Kannan. 2009. "*In silico* prediction of structure and functional aspects of a hypothetical protein of Neurospora crassa." J Cell Tissue Re 9: 1889-1894.

Pontius, J., J. Richellea, S. J. Wodak. 1996. "Deviations from standard atomic volumes as a quality measure for protein crystal structures." Journal of molecular biology 264: 121-136.

Pruitt, K. D., T. Tatusova, W. Klimke, D. R. Maglott. 2009. "NCBI Reference Sequences: current status, policy and new initiatives." Nucleic Acids Research 37(suppl 1): D32-D36.

Rambabu, R. 2012. "Computational analysis and function prediction of a hypothetical protein 1RW0." Int. J. Comp. Bioinformatics and In Silico: 58-62.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins. 2011. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Molecular systems biology 7: 539.

Smith, G. L. 1991. "Vaccinia virus vectors for gene expression." Current opinion in biotechnology 2(5): 713-717.

Smith, G. L., B. J. Murphy and M. Law. 2003. "Vaccinia virus motility." Annual Reviews in Microbiology 57: 323-342.

Smith, G. L., A. Vanderplasschen and M. Law. 2002. "The formation and function of extracellular enveloped vaccinia virus." Journal of General Virology 83: 2915-2931.

Tolonen, N., L. Doglio, S. Schleich, J. K. Locker. 2001. "Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei." Molecular biology of the cell 12: 2031-2046.

Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang. 2015. "The I-TASSER Suite: protein structure and function prediction." Nature methods 12: 7-8.