In-silico prediction and modeling of the *Entamoeba histolytica* proteins: Serine-rich *Entamoeba histolytica* protein and peroxiredoxin

Kumar Manochitra [1] , **Subhash Chandra Parija** Corresp. [1]

[1] Department of Microbiology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, Puducherry, India

Corresponding Author: Subhash Chandra Parija
Email address: subhashparija@yahoo.co.in

Background: Amoebiasis is the third most common parasitic cause of morbidity and mortality particularly in countries with poor hygienic settings. There exists an ambiguity in the diagnosis of amoebiasis, and hence arises a necessity for a better diagnostic approach. Serine-rich *Entamoeba histolytica* protein (SREHP), peroxiredoxin and Gal/GalNAc lectin are pivotal in *E. histolytica* virulence and are extensively studied as diagnostic and vaccine targets. For elucidating the cellular function of these proteins, details regarding their respective quaternary structures are essential which are not available till date. Hence, this study was carried out to predict the structure of these target proteins and characterize them structurally as well as functionally using relevant *in-silico* methods.

Methods: The amino acid sequences of the proteins were retrieved from National Centre for Biotechnology Information database and aligned using ClustalW. Bioinformatic tools were employed in the secondary structure and tertiary structure prediction. The predicted structure was validated, and final refinement was carried out.

Results: The protein structures predicted by i-TASSER were found to be more accurate than Phyre2 based on the validation using SAVES server. The prediction suggests SREHP to be a extracellular protein, peroxiredoxin was a peripheral membrane protein, while Gal/GalAc was found to be a cell-wall protein. Signal peptides were found in the amino-acid sequences of SREHP and Gal/GalNAc, whereas they were not present in the peroxiredoxin sequence. Gal/GalNAc lectin showed better antigenicity than the other two proteins studied. All three proteins exhibited similarity in their structures and were mostly composed of loops.

Discussion: The structures of SREHP and peroxiredoxin were predicted successfully, while the structure of Gal/GalNAc lectin could not be predicted as it was a complex protein composed of three sub-units. Also, this protein showed less similarity with the available structural homologs. The quaternary structures predicted from this study would provide better structural and functional insights into these proteins and may aid in development of newer diagnostic assays or enhancement of the available treatment modalities.

In-silico prediction and modeling of the *Entamoeba histolytica* proteins: Serine-rich *Entamoeba histolytica* protein and peroxiredoxin

Kumar Manochitra and Subhash Chandra Parija*

Department of Microbiology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

Corresponding Author:
Dr. Subhash Chandra Parija,
Director,
Jawaharlal Institute of Postgraduate Medical Education and Research (An Institution of National Importance under the Ministry of Health & Family Welfare, Govt. of India), Dhanvantari Nagar, Puducherry – 605006, India
e-mail ID: subhashparija@yahoo.co.in

ABSTRACT

Background: Amoebiasis is the third most common parasitic cause of morbidity and mortality particularly in countries with poor hygienic settings. There exists an ambiguity in the diagnosis of amoebiasis, and hence arises a necessity for a better diagnostic approach. Serine-rich *Entamoeba histolytica* protein (SREHP), peroxiredoxin and Gal/GalNAc lectin are pivotal in *E. histolytica* virulence and are extensively studied as diagnostic and vaccine targets. For elucidating the cellular function of these proteins, details regarding their respective quaternary structures are essential which are not available till date. Hence, this study was carried out to predict the structure of these target proteins and characterize them structurally as well as functionally using relevant *in-silico* methods.

Methods: The amino acid sequences of the proteins were retrieved from National Centre for Biotechnology Information database and aligned using ClustalW. Bioinformatic tools were employed in the secondary structure and tertiary structure prediction. The predicted structure was validated, and final refinement was carried out.

Results: The protein structures predicted by i-TASSER were found to be more accurate than Phyre2 based on the validation using SAVES server. The prediction suggests SREHP to be a extracellular protein, peroxiredoxin a peripheral membrane protein while Gal/GalAc was found to be a cell-wall protein. Signal peptides were found in the amino-acid sequences of SREHP and Gal/GalNAc, whereas they were not present in the peroxiredoxin sequence. Gal/GalNAc lectin showed better antigenicity than the other two proteins studied. All the three proteins exhibited similarity in their structures and were mostly composed of loops.

Discussion: The structures of SREHP and peroxiredoxin were predicted successfully, while the structure of Gal/GalNAc lectin could not be predicted as it was a complex protein composed of three sub-units. Also, this protein showed less similarity with the available structural homologs. The quaternary structures predicted from this study would provide better structural and functional insights into these proteins and may aid in development of newer diagnostic assays or enhancement of the available treatment modalities.

KEYWORDS

Entamoeba histolytica, SREHP, 29KDa cysteine rich protease, Thioredoxin peroxidase, Peroxiredoxin, Gal/GalNAc lectin.

INTRODUCTION

Amoebiasis is one of the most common parasitic disease and is associated with high morbidity and mortality (Que & Reed, 2000), killing about 50 million people each year, predominantly in countries with poor hygienic settings (CDC, 2010). Amoebiasis remains a serious public health problem even today particularly in the developing and under-developed countries. Globally, the prevalence is 2%–60%, whereas in India it ranges between 3.6%–47.4% (Khairnar & Parija, 2007; Mukherjee 2010). Due to high level of uncertainty associated with the specificity of the available diagnostic assays, there is a need for a specific diagnostic target (Tanyuksel & Petri, 2003). Identifying new targets and exploring alternate strategies with high sensitivity and specificity for the early diagnosis of amoebiasis is important.

Proteins mediate most biological processes. Thus, identifying target proteins and ascertaining their role in pathogenesis will aid in selecting better diagnostic markers. The proteins involved in *E. histolytica* virulence and extensively studied as diagnostic and vaccine targets are Serine-rich *E. histolytica* protein (SREHP), peroxiredoxin or thioredoxin peroxidase or 29KDa cysteine-rich protease and Gal/GalNAc lectin (Stanley, 1991; Ravdin, 1989; Cheng, 2004). SREHP is highly immunogenic of all the *Entamoeba* proteins identified so far, possessing the largest number of conserved epitopes. It was found that more than 80% of the antibodies elicited among the patients with amoebic liver abscess are specific against SREHP. Peroxiredoxin also plays a significant role in regulating enzymatic activities, restoring oxidized proteins, cellular transcription and apoptosis (Arias, 2012). However, knowledge regarding quaternary structure, which is essential for elucidating the cellular and molecular ontology of these proteins, is currently lacking (Samuel & Stanley, 1997). Thus, detailed studies regarding the cellular function of these proteins are crucial to utilise them either as a diagnostic or a vaccine target. Moreover, accurate prediction of the protein structures and elucidation of their functions will aid in bridging the information gap necessary for identifying new diagnostic markers, vaccine candidates and drug targets precisely.

The aim of the current study is to predict the structure of these target proteins and to characterise them structurally as well as functionally using relevant *in-silico* methods.

MATERIALS AND METHODS

1. PROTEIN SEQUENCE ANALYSIS

Amino-acid sequences of the target proteins included in this study were retrieved from National Centre for Biotechnology Information database (NCBI) and aligned using ClustalW software to determine the appropriate sequence for protein structure prediction. Using sequence similarity model, the availability of the structural homologs for the retrieved sequences was verified from the available structures present in the protein data bank (PDB). The overall workflow of the present study has been summarized in Figure1.

2. PHYSIOCHEMICAL PROFILING

Using the target protein sequence as template, its molecular profile was determined using Protparam and Predict Protein and the structural properties of the protein were predicted using SOPMA, SAPS and FindMod. The solubility of these proteins was predicted using Predict Protein. The presence of signal peptides within the amino-acid sequence was verified using SignalP 4.1 server. Sub-cellular localization of the proteins within the cell was predicted using PSortB. The antigenicity of these proteins was predicted using Predicted Antigenic Peptides and the predicted results were further validated using EMBOSS.

3. COMPARATIVE STRUCTURE MODELING

The similarity with the available protein homologs in PDB was found to be less than 40%. Hence, the structure of the protein was predicted by fold recognition methodology using i-TASSER and Phyre2 prediction server.

4. STRUCTURE VALIDATION AND REFINEMENT

The protein structures generated using i-TASSER and Phyre2 servers were then validated by SAVes server. The energy levels were minimized, and the structures were reformed based on the generated Ramachandran plot.

RESULTS

The sequences AAA29117.1, P19476.2 and XP_656181.1, were found most suited for structure prediction of SREHP, peroxiredoxin and Gal/GalNAc lectin respectively as they had the entire stretch of amino acids comprising the N-terminal as well as C-terminal ends.

The molecular profile of the proteins as predicted by SOPMA, SAPS and FindMod servers has been described in Table 1.

The results of Predict Protein suggest that SREHP is an extracellular protein; peroxiredoxin is a peripheral membrane protein and Gal/GalNAc lectin is a cell-wall protein. Signal peptides were found within the amino-acid sequences of SREHP and Gal/GalNAc lectin. However, no signal peptides were found within the peroxiredoxin sequence, and this finding is consistent with that from a previous study (Clark, 2007).

SREHP contained three antigenic determinants with an average antigenic propensity of 0.9748; peroxiredoxin possessed 11 antigenic determinants with an average antigenic propensity of 1.0318. But, Gal/GalNAc lectin had 51 antigenic determinants with the maximum average antigenic propensity of 1.0410. Thus, it is known to be critical in eliciting anti-amoebic host immune response mechanism(s) (Rasti, 2006).

The predicted structures suggest that SREHP contained 51.5% loop, 30.9% helix and 17.6% strands; peroxiredoxin had 57.51% loop, 27.9% helix and 14.59% strands and Gal/GalNAc lectin comprised 67% loop, 25.5% helix and 7.4% strand. Thus, all the three proteins were found to be primarily composed of loops followed by helix and strands.

The tertiary structures of SREHP and peroxiredoxin were successfully predicted using i-TASSER & Phyre2 server via threading (Yang, 2015; Roy, Kucukural & Zhang, 2010; Zhang, 2008; Kelley & Sternberg, 2009).

The protein structures predicted by i-TASSER (Figs. 2 and 3) were found to be more accurate than Phyre2 based upon the analysis of SAVes server (Procheck, WHATCHECK, Verify-3D, Errat& Prove)[Laskowski, 1993; Hooft, 1996; Luthy, Bowie & Eisenberg, 1992; Pontius, Richelle & Wodak, 1996]. (Supplementary files)

DISCUSSION

The enteric protozoan parasite *E. histolytica* usually resides in the large bowel of the host causing amoebic colitis. However, it can occasionally penetrate the intestinal mucosa and spread to liver or other organs causing amoebic liver abscess (Mukherjee, 2010). The ability of the parasite to cope up with increasing oxygen pressures and high concentration of reactive oxygen species (ROS) and reactive nitrogen species (RNS), contributes to its virulence (Stanley, Koester & Li, 1995) and a previous study has demonstrated the involvement of peroxiredoxin in this regard (Arias, 2012). Gal/GalNAc lectin is accountable for the virulence of *E. histolytica* and is reported to be involved in almost all the steps of pathogenesis (Boettner, 2002). Hence, it serves as a potential target for diagnosis and vaccination.

The details regarding physiochemical properties of these proteins such as their quaternary structure, antigenicity, structural and functional properties will be informative and may assist in identifying their role in disease progression. As the crystal structures of these proteins are not available, we have predicted the structures using *in-silico* methods which would assist in further exploring these target proteins as diagnostic markers, drug targets and vaccine candidates.

The structures of SREHP and peroxiredoxin were predicted successfully, and on validation they were found to be more than 95% accurate which implies a good probability of the predicted structure being existent in nature. As, Gal/GalNAc lectin is a complex heteromeric protein composed of three sub-units, and its similarity with the available protein homologs was comparatively low. Therefore, the functional structure of this protein could not be predicted. However, the structures of the subunits have been predicted although they could not be assembled accurately. The structure of Gal/GalNAc lectin needs to be determined either by X-ray crystallography or NMR methodology.

Peroxiredoxin plays an important role in the parasite defence against the reactive species of the host. This protein is critical in the extra-intestinal phase of amoebic infection (Cheng, 2004). In-depth characterization of its activity and its functional properties are available (Arias, 2012), however, its structural properties are undetermied. In our study, we found peroxiredoxin to be the most stable of the three proteins with an instability index of 54.79, which is remarkable. Given its high stability and its pathophysiological role in extra-intestinal amoebic infection, this

156 protein can be considered as a potential candidate for vaccine trials or enhanced treatment
157 strategies.
158 The SREHP molecule serves as a potent chemoattractant for amoebic trophozoites and is
159 unique when compared to other *E. histolytica* proteins because of its phosphorylation and
160 glycosylation patterns (Ravdin, 1989). In our study, we predicted that SREHP is an extracellular
161 protein, thus being easily accessible to the host immune system. The amino-acid residues within
162 the peptide sequence of SREHP were predicted to be highly conserved when compared with
163 other *E. histolytica* proteins. Findings from our study suggest that SREHP possesses multi-
164 hydrophilic conserved dodecapeptides, a detail that has also been reported previously from *in-*
165 *vitro* analysis of this protein (Stanley, Koester & Li, 1995). Thus, results generated from the
166 bioinformatic analysis employed in the present study are not mere pre-experimental findings but
167 can also serve as a reliable lead for future *in-vitro* experiments. SREHP being a highly conserved
168 protein can serve as vaccine candidate with other *E. histolytica* antigenic proteins such as
169 Gal/GalNAc lectin, and help in enhancing host immunity.
170 Gal/GalNAc lectin being a multimeric protein with a light subunit, heavy subunit and an
171 intermediate subunit, surmounted the other two proteins in all aspects of antigenicity with 51
172 potent antigenic determinants within its sequence. Apart from its antigenic propensity,
173 Gal/GalNAc lectin is structurally a highly conserved antigen (Boettner, 2002). Moreover,
174 Gal/GalNAc lectin is a cell-wall protein that is easily accessible and recognized by the host
175 immune system (Stanley, 1991; Boettner, 2002), thereby enhancing its antigenic profile. It
176 mediates attachment of trophozoites to colonic mucins, increases parasite phospholipase A
177 activity, maintains an acidic pH in amoebic intracellular vesicles and enhances cytolytic activity
178 (Ravdin, 1989). Thus, by hydrolyzing this protein, the host immune system can counteract
179 invasion by the parasite. Considering all these molecular features of Gal/GalNAc lectin, our
180 study suggests that, this protein could be a prime vaccine candidate and diagnostic target. Many
181 studies have been carried out regarding Gal/GalNAc lectin; however, they are inadequate whilst
182 considering its significance. A thorough investigation is essential as its impact would be far-
183 reaching.

184 **GRANT DISCLOSURES**

The first author (MK) is grateful to the Council of scientific and industrial research (CSIR), Govt. of India, for providing financial support (grant no. 09/05(0007)/2012-EMR-I). The study was also partly supported by JIPMER Institute Research Council Intramural Grant.

COMPETING INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENT

The authors would like to acknowledgement Mr. Yugendran Thiyagarajan for his assistance in formulating the study design.

AUTHOR CONTRIBUTIONS

Kumar Manochitra and Subhash Chandra Parija conceived and designed the study. Kumar Manochitra performed the experiments, analysed the data and prepared the manuscript. Subhash Chandra Parija reviewed the manuscript.

REFERENCES

Arias DG, Regner EL, Iglesias AA, Guerrero SA. 2012. *Entamoeba histolytica* thioredoxin reductase: Molecular and functional characterization of its atypical properties. *Biochimica et Biophysica Acta* 1820:1859–1866. DOI: 10.1016/j.bbagen.2012.08.020.

Boettner DR, Huston C, Petri WA. 2002. Galactose/N-acetylgalactosamine lectin: The coordinator of host cell killing. *Journal of Biosciences* **27**:553–557. DOI 10.1007/BF02704847.

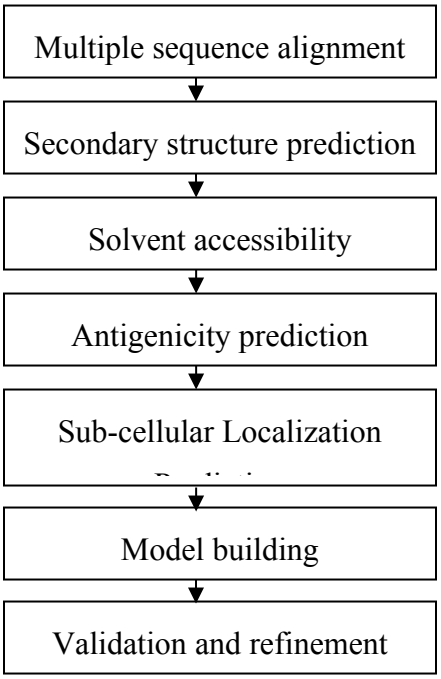
Centres for Disease control and Prevention. 2010. Parasites – Amebiasis. Retrieved from http://www.cdc.gov/parasites/amebiasis/ June 2015.

Cheng XJ, Yoshihara E, Takeuchi T, Tachibana H. 2004. Molecular characterization of peroxiredoxin from *Entamoeba moshkovskii* and a comparison with *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **138**:195–203. DOI 10.1016/j.molbiopara.2004.08.009.

211 Clark CG, Alsmark UCM, Tazreiter M, Saito-Nakano Y, Ali V, Marion S. 2007. Structure and
212 content of the *Entamoeba histolytica* genome. *Advances in Parasitology* **65**:51–189. DOI
213 10.1016/S0065-308X(07)65002-7.
214 Colovos C, Yeates TO. 1993. Verification of protein structures: Patterns of non-bonded atomic
215 interactions. *Protein Sciences* **2**:1511–1519. DOI:10.1002/pro.5560020916.
216 Hooft, RWW, Vriend G, Sander C, Abola EE. 1996. Errors in protein structures. *Nature*
217 **381**:272–272. DOI 10.1038/381272a0.
218 Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the web: A case study using
219 the Phyre server. *Nature Protocols* **4**:363–371. DOI 10.1038/nprot.2009.2.
220 Khairnar K, Parija SC. 2007. A novel nested multiplex PCR assay for differential detection of
221 Entamoeba histolytica, E. moshkovskii and E. dispar DNA in stool samples. *BMC*
222 *Microbiology* **7**:47. DOI 10.1186/1471-2180-7-47.
223 Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK – A program to
224 check the stereochemical quality of protein structures. *Journal of Applied Crystallography*
225 **26**:283–291. DOI 10.1107/S0021889892009944.
226 Luthy R, Bowie JU, Eisenberg D. 1992. Assessment of protein models with three-dimensional
227 profiles. *Nature* **356**:83–85. DOI:10.1038/356083a0.
228 Mukherjee AK, Das K, Bhattacharya MK, Nozaki T, Ganguly S. 2010. Trend of *Entamoeba*
229 *histolytica* infestation in Kolkata. *Gut Pathogens* **2**:12. DOI 10.1186/1757-4749-2-12.
230 Pontius J, Richelle J, Wodak SJ. 1996. Deviations from standard atomic volumes as a quality
231 measure for protein crystal structures. *Journal of Molecular Biology* 264:121–136. DOI
232 10.1006/jmbi.1996.0628.
233 Que X, Reed SL. 2000. Cysteine Proteinases and the pathogenesis of amoebiasis. *Clinical*
234 *Microbiology Reviews* **3**:196–206. DOI 10.1128/CMR.13.2.
235 Rasti S, Haghighi A, Kazemi B, Rezaian M. 2006. Cloning and characterization of serine-rich
236 *Entamoeba histolytica* protein gene from an Iranian *E. histolytica* Isolate. *Pakistan Journal*
237 *of Biological Sciences* **9**:654–658. DOI 10.3923/pjbs.2006.654.658.
238 Ravdin JI. 1989. *Entamoeba histolytica*: from adherence to enteropathy. *Journal of Infectious*
239 *Diseases* **159**:420–429.
240 Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: A unified platform for automated protein
241 structure and function prediction. *Nature Protocols* **5**:725–738. DOI 10.1038/nprot.2010.5.

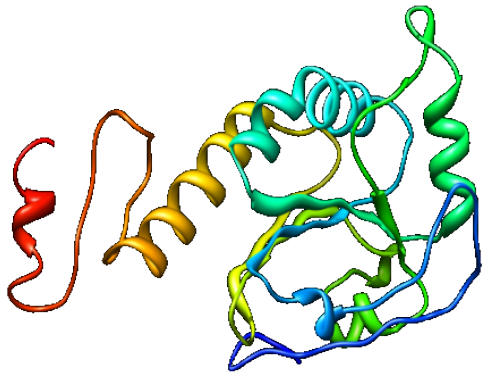
242 Samuel L Stanley JR. 1997. Progress towards development of a vaccine for amoebiasis.
243 *Clinical Microbiology Reviews* **10**:637–639.
244 Stanley SL, Jackson TFHG, Reed SL, Calderon J, Jenkins CK, Gathiram V, Li E. 1991.
245 Serodiagnosis of invasive amoebiasis using a recombinant *Entamoeba histolytica* protein.
246 *Journal of American Medical Association* **266**:984–986.
247 Stanley Jr SL, Koester T, Li E. 1995. The serine-rich *Entamoeba histolytica* protein is a
248 phosphorylated membrane protein containing *O*-linked terminal *N*-acetylglucosamine
249 Residues. *Journal of Biological Chemistry* **270**:4121–4126.
250 Tanyuksel M, Petri WA Jr. 2003. Laboratory diagnosis of amoebiasis. *Clinical Microbiology*
251 *Reviews* **4**:713–729. DOI 10.1128/CMR.16.
252 Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: Protein
253 structure and function prediction. *Nature Methods* **12**:7–8. DOI 10.1038/nmeth.3213.
254 Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*
255 **9**:40. DOI 10.1186/1471-2105-9-40.

257 **Figure 1:** Flowchart summarizing the methodology of the study



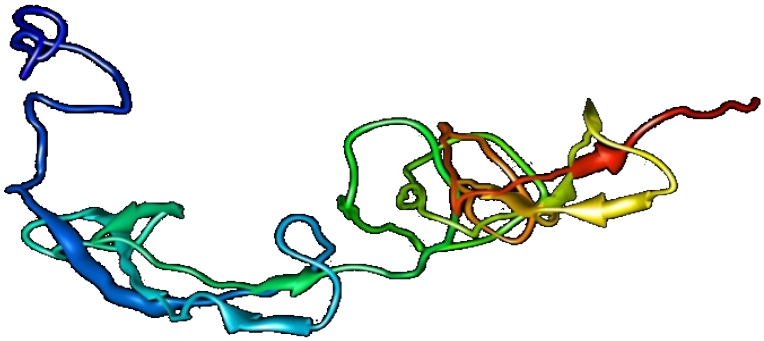
258

260 **Figure 2** Structure of Peroxiredoxin.



261

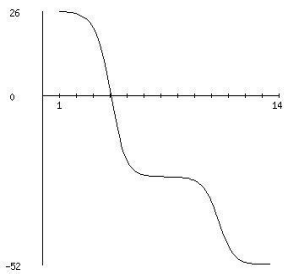
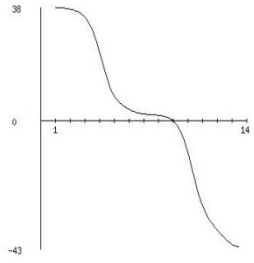
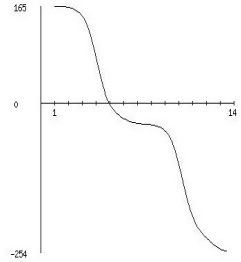
263 **Figure 3** Structure of SREHP.



264

266 **Table 1:** Molecular profile of the proteins SREHP, peroxiredoxin and Gal/GalNAc lectin

267

No.	Properties	SREHP	Peroxiredoxin	Gal/Gal/NAcLectin
1	**No of amino acids**	233	233	1286
2	**Molecular weight**	24.72 kDa	26.25 kDa	144.33 KDa
3	**Formula**	$C_{1032}H_{1623}N_{281}O_{418}S_2$	$C_{1162}H_{1837}N_{307}O_{342}S_{21}$	$C_{6205}H_{9714}N_{1668}O_{2054}S_{118}$
4	**Total no. of atoms**	3,356	3,669	19,759
5	**Net charge of the protein**	–25	+4	–26
6	**Theoretical pI**	4.26 	7.79 	5.16 
8	**Ext. coefficient**	1490	32400	159925
9	**Estimated half-life**	30 hr (mammalian reticulocytes, *in vitro*). >20 hr(yeast, *in vivo*) >10 hr (*E. coli*, *in vivo*)	30 hr (mammalian reticulocytes, *in vitro*). >20 hr (yeast, *in vivo*). >10 hr (*E. coli, in vivo*)	30 hr (mammalian reticulocytes, *in vitro*). >20 hr (yeast, *in vivo*). >10 hr (*E. coli, in vivo*)
10	**Aliphatic index**	41.63	76.57	63.20

11	**Grand average of hydropathicity (GRAVY)**	–1.218	–0.320	–0.546
12	**Localization Scores:**			
	Cytoplasmic	1.50	9.06	–
	Cellwall	3.50	0.02	–
	Extracellular	4.50	0.01	–
	Peripheral Membrane	-	9.96	–
	Final Prediction	Extracellular	Peripheral membrane protein	_
13	**Instability index**	54.79 (protein is stable)	30.44 (protein is stable)	36.34 (protein is stable)

268