The Pharma Research (T. Pharm. Res.), (2009), 2; 16-29.
Received: 14 Aug 2009



*Original Article*

# PROTEIN MODELING OF APICAL MEMBRANE ANTIGEN-1(AMA-1) OF *PLASMODIUM FALCIPARUM*

*Ankitash Tulyani

**Affiliated to:** Pharmacy Department, Birla Institute of Technology & Science, Pilani, India

## ABSTRACT

Apical membrane antigen1 (AMA-1) is a Plasmodium asexual blood-stage antigen. It is very essential for malarial parasite *Plasmodium falciparum* for Erythrocyte Invasion. It has been suggested that positive selection operates on the AMA-1 gene in regions coding for antigenic sites. It is thus a significant molecule to analyze for obtaining novel methods to block such an epitope in combating malaria. The results were analyzed through MolProbity, anolea, Kinemage, errat and co-relating them with the ramachandran plots.

**Keywords;** Plasmodium, Falciparum, Erythrocyte, Invasion, Apical, Membrane, Antigen1, AMA-1

*Corresponding Author:
Ankitash Tulyani,
3/427, Mukta Prasad Colony,
Bikaner. (Rajasthan)
**Email: ankitash.1990@gmail.com**

## 1.0 Introduction:

Proteins are the building blocks of an organism as they are involved in regulating the various activities. Protein is created by a process which starts from polypeptide and later this polypeptide takes the functional three-dimensional conformation. Each protein structure develops from unstructured polypeptide which is translated from a sequence of mRNA as a linear chain of amino acids. Each amino acid incorporates certain 'gross' chemical features. These amino acids then interact with each other and their cell surroundings to produce a well defined, 3-D shape called as native state, which is responsible for a protein's biological function.

Structural representations are the most useful parameters for understanding protein function. So, the Structure Prediction aims at deriving detailed structural information of high resolution for understanding biological function qualitatively. Understanding structure has potential applications in the various genome projects being undertaken, such as mapping the functions of proteins in metabolic pathways for whole genomes and deducing significant evolutionary relationships.

The Structure Prediction problem originates from the fact that electron densities around the amide nitrogen and oxygen atoms of amino acid residues are quite similar so that the locations of these atoms can be determined to high precision but not their identity. Hence, the problem seems to be specific for X-ray analysis of protein crystals.

We are involved in predicting the protein structure because the output of experimentally determined protein structures by X-ray crystallography or NMR spectroscopy is lagging far behind the number of DNA Sequences. There are many problems associated with both these methods. First of all starting from, X-Ray crystallography; there are many associated problems like an extremely pure protein sample is needed. The protein sample must form crystals, which is generally the biggest problem as many proteins aren't amenable to crystallization at all. NMR again is also limited to small, soluble proteins only and RMSD is also slightly higher, which makes it a later choice for predicting protein structures. Therefore, the experimental data obtained from X-ray and NMR experiments are generally refined by various refinement protocols before the final coordinates are deposited and we have to conclude that the refinement protocols used are also unable to correct the unfavorable configuration or perhaps that they even introduce such errors.

Apical membrane antigen1 (AMA-1) is a Plasmodium asexual blood-stage antigen. It is very essential for malarial parasite *Plasmodium falciparum* for Erythrocyte Invasion. It has been suggested that positive selection operates on the AMA-1 gene in regions coding for antigenic sites. It is thus a significant molecule to analyze for

obtaining novel methods to block such an epitope in combating malaria [3, 4].

This study focuses on building molecular model of AMA-1 protein of Plasmodium falciparum using molecular modeling techniques using Modeller 8 standalone tool. We used it for comparative modeling and by comparative analysis we got the structure with the least possible free energy [5].

## 2.0 MATERIALS AND METHOD

Homology modeling is carried out, if the native sequence similarity with the similar crystal/NMR structure of already known protein is greater than 35%.This process is conceptually very simple. First perform a multiple alignment through PDB-BLAST tool (PSI-BLAST) available on ncbi web server with the already modeled sequences for which the structure has been determined by experimental methods. The query sequence for AMA-1 gene is

**Query Sequence**

1 gkcpvfgkgi iiensnttfl tpvatenqdl kdggfafppt eplmspmtld qmrhfykdnk

61 yvknldeltl csrhagnmip dndknsnyky pavyddkdkk chilyiaaqe nngprycnkd

121 qsirnsmfcf rpakdisfqn ytylsknvvd nwekvcprkn lenakfglwv dgncediphv

181 nefpaidlfe cnklvfelsa sdqpkqyeqh ltdyekikeg fknknasmik saflptgafk

241 adryksrgkg ynwgnynret qkceifnvkp tclinnssyi attalshpne vehnfpcsly

301 kdeikkeier eskriklndn ddegnkkiia prifisddid slkcpcdpei vsnstcnffv

361 ckcvekraev tsnnevvvke eykdeyadip ehkptydnmk iiiassaava vlatilmvyl

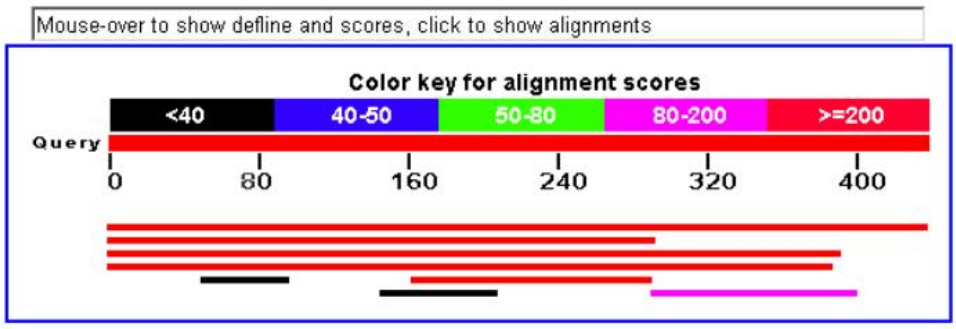
421 ykrkgnaeky dkmdqpq//

*(DBSOURCE* embl accession AJ408344.1)

**PDB BLAST Results:**

We got the following results for PDB Blast. We chose PSI-BLAST (Position-Specific Iterated BLAST) to use iteration method for best results.

Distribution of 8 Blast Hits on the Query Sequence



Distance tree of results  Related Structures

| Sequences producing significant alignments: | | Score (Bits) | E Value | |
|---|---|---|---|---|
| pdb\|2J5L\|A | Chain A, Structure Of A Plasmodium Falciparum Apic... | 783 | 0.0 | S |
| pdb\|1Z40\|A | Chain A, Ama1 From Plasmodium Falciparum >pdb\|1Z40... | 573 | 4e-164 | S |
| pdb\|1W81\|A | Chain A, Crystal Structure Of Apical Membrane Anti... | 487 | 2e-138 | S |
| pdb\|2J4W\|D | Chain D, Structure Of A Plasmodium Vivax Apical Me... | 483 | 4e-137 | S |
| pdb\|1YXE\|A | Chain A, Structure And Inter-Domain Interactions O... | 255 | 1e-68 | S |
| pdb\|1HN6\|A | Chain A, Solution Structure Of Plasmodium Falcipar... | 194 | 3e-50 | S |
| pdb\|2DN6\|A | Chain A, Solution Structure Of The Ph Domain Of Ki... | 29.6 | 1.5 | S |
| pdb\|1XYG\|A | Chain A, X-Ray Structure Of Gene Product From Arab... | 28.1 | 4.3 | S |

We selected 1Z40 sequence for the further work. Then using the FASTA sequence as the template for the query sequence, the alignment file (.ali file) was designed for using Modeler 8v2.

The model-default.py python file was run in the modeler making the necessary changes in it for the desired run for iteration. Now after processing this file using modeler 8v2, we iterated 30 times. It was done to predict the best out of these designed structures based on free energy parameter of a stable conformation. The structure with the least energy was then used for Ramachandran plot and was further asessessed through MolProbity, anolea, Kinemage, errat results.

## 3.0 RESULTS & DISCUSSION:

Out of the 8 results displayed in PDB BLAST results, the one chosen is 1Z40 protein. A simple reason is that it was having the minimum E-value i.e. 4e-164 as compared to other models. Though better than it one more model observed as template was 2J5L protein. But when this protein PDB file was seen and analyzed against its FASTA sequence and the Query Sequence, it was observed that even though it showed very high alignment aspect in the FASTA sequences, it showed a very minimum similarity in protein primary sequence aligned in PDB sequence for homology modeling. Thus, it made a requirement to search a PDB file with more number of query residues getting aligned in the PDB amino acid residue chain. Thus 1Z40 was chosen for the chain A modeling of AMA-I.

***Pairwise Alignment of Query with the Sorted 1Z40 FASTA format***

Pairwise alignment was performed now to see where exactly in the sorted PDB FASTA sequence, our most part of query sequence matches to model that part. It was observed that 287 residues were matching with very few gap introductions. Though it was revealed later that the identity shown for this 287 residues A chain, who was showing 96% similarity here with quite high score, is comparable to actual template PDB sequence with reasonable number of gaps. It showed that PDB sequence of A chain considered in template was having lot many atoms or amino acid residues missing, which may not have come

during X-Ray or NMR Crystal Analysis. So, to model them, gaps were incorporated.

**Template FASTA Sequence**

1 gnymgnpwte ymakydieev hgsgirvdlg edaevagtqy rlpsgkcpvf gkgiiiensn

61 ttfltpvatg nqylkdggfa fppteplmsp mtldemrhfy kdnkyvknld eltlcsrhag

121 nmipdndkns nykypavydd kdkkchilyi aaqenngpry cnkdeskrns mfcfrpakdi

181 sfqnytylsk nvvdnwekvc prknlqnakf glwvdgnced iphvnefpai dlfecnklvf

241 elsasdqpkq yeqhltdyek ikegfknkna smiksaflpt gafkadryks hgkgynwgny

301 ntetqkceif nvkptclinn ssyiattals hpieve//

*DBSOURCE* pdb: molecule 1Z40

gi: 75765674

**Template Structural Parameters**

| ***Parameters*** | Resolution[Å] | R-Value | R-Free | Space Group |
|---|---|---|---|---|
| | 1.90 | 0.195 (obs.) | 0.236 | P $3_1$ |

**Pairwise Alignment Result with the sorted Template**

```
 Score =  573 bits (1476),  Expect = 4e-164, Method: Composition-based stats.
 Identities = 283/292 (96%), Positives = 286/292 (97%), Gaps = 0/292 (0%)

Query  1    GKCPVFGKGIIIENSNTTFLTPVATENQDLKDGGFAFPPTEPLMSPMTLDQMRHFYKDNK  60
            GKCPVFGKGIIIENSNTTFLTPVAT NQ LKDGGFAFPPTEPLMSPMTLD+MRHFYKDNK
Sbjct  45   GKCPVFGKGIIIENSNTTFLTPVATGNQYLKDGGFAFPPTEPLMSPMTLDEMRHFYKDNK  104

Query  61   YVKNLDELTLCSRHAGNMIPDNDKNSNYKYPAVYDDKDKKCHILYIAAQENNGPRYCNKD  120
            YVKNLDELTLCSRHAGNMIPDNDKNSNYKYPAVYDDKDKKCHILYIAAQENNGPRYCNKD
Sbjct  105  YVKNLDELTLCSRHAGNMIPDNDKNSNYKYPAVYDDKDKKCHILYIAAQENNGPRYCNKD  164

Query  121  QSIRNSMFCFRPAKDISFQNYTYLSKNVVDNWEKVCPRKNLENAKFGLWVDGNCEDIPHV  180
            +S RNSMFCFRPAKDISFQNYTYLSKNVVDNWEKVCPRKNL+NAKFGLWVDGNCEDIPHV
Sbjct  165  ESKRNSMFCFRPAKDISFQNYTYLSKNVVDNWEKVCPRKNLQNAKFGLWVDGNCEDIPHV  224

Query  181  NEFPAIDLFECNKLVFELSASDQPKQYEQHLTDYEKIKEGFKNKNASMIKSAFLPTGAFK  240
            NEFPAIDLFECNKLVFELSASDQPKQYEQHLTDYEKIKEGFKNKNASMIKSAFLPTGAFK
Sbjct  225  NEFPAIDLFECNKLVFELSASDQPKQYEQHLTDYEKIKEGFKNKNASMIKSAFLPTGAFK  284

Query  241  ADRYKSRGKGYNWGNYNRETQKCEIFNVKPTCLINNSSYIATTALSHPNEVE  292
            ADRYKS GKGYNWGNYN ETQKCEIFNVKPTCLINNSSYIATTALSHP EVE
Sbjct  285  ADRYKSHGKGYNWGNYNTETQKCEIFNVKPTCLINNSSYIATTALSHPIEVE  336
```

Query is getting aligned from 1 to 292 residues with the subject sequence 45 to 336. This was done for the alignment file generation, which will be the next step.

***Alignment (.ali) File Generation***

Query was aligned with the FASTA sequence of the template selected above and based on it the gaps were incorporated for all the missing

residues and finally it is aligned and matched whether all amino acids considered are present in the ATOM file of PDB file selected. It was seen that here in this case, 287 residues chain was found aligned with some gaps incorporated in PDB at the sites where amino acids were not resolved properly in the original template PDB structure.

The Alignment file made had 2 fields one for the template on top with identifier as Structure X and the other one as Sequence (ama here) as the query sequence to be modeled. Here, it is a mandatory fact that the length which aligns in both Template and Query Sequence should be marked by the residues location by its locus number on start and ending of the alignment. In this case, it was seen that 108 to 438 of PDB structure were matching in the query sequence length of 287 residues total excluding number of the gaps incorporated to make a perfect alignment.

**Alignment file:**

>P1;1z40

structureX:1z40:108:A:438::epitope:*Plasmodium falciparum*: 1.90: 0.195

-----

NPWTEYMAKYDIEEVHGSGIRVDLGEDAEV

AGTQYRLPSGKCPVFGKGIIIENSN

TTFLTPVAT-----

KDGGFAFPPTEPLMSPMTLDEMRHFYKDNK

YVKNLDELTLCSRHAG

NMIPDNDKNSNYKYPAVYDDKDKKCHILYI

AAQENNGPRYC----------FCFRPAKDI

SFQNYTYLSKNVVDNWEKVCPRKNLQNAK

FGLWVDGNCEDIPHVNEFPAIDLFECNKLVF

ELSASDQPKQYEQHLTDYEKIKEGFKNKNA

SMIKSAFLPT-----DRYKSHGKGYNWGNY

NTETQKCEIFNVKPTCLINNSSYIATTALSHPI

EVE*

>P1;ama

sequence:ama:1 : :336: :epitope:*Plasmodium falciparum*::

----------

GKCPVFGKGIIIENSN

TTFLTPVATENQDLKDGGFAFPPTEPLMSPM

TLD-MRHFYKDNKYVKNLDELTLCSRHAG

NMIPDNDKNSNYKYPAVYDDKDKKCHILYI

AAQENNGPRYCNKDQSIRNSMFCFRPAKDI

SFQNYTYLSKNVVDNWEKVCPRKNL-

NAKFGLWVDGNCEDIPHVNEFPAIDLFECN

KLVF

ELSASDQPKQYEQHLTDYEKIKEGFKNKNA

SMIKSAFLPTGAFKADRYKS-GKGYNWGNY

N-ETQKCEIFNVKPTCLINNSSYIATTALSHP-

EVE*

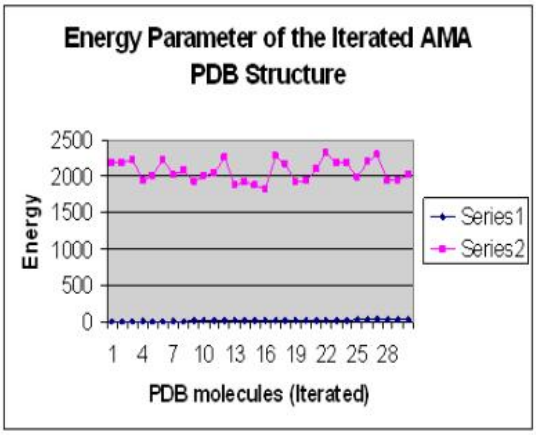
Now, the modeler 8v2 was run through the model-default.py python file by making the necessary changes in it for the desired run for iteration. The results shown were the following in brief.

| | |
|---|---|
| **Atoms** | : 3478 |
| **Energy** | : 2719.45728 |
| **Total CPU time [seconds]** | : 134.86 |

| Restraints | RMSD Values |
|---|---|
| Distance restraints 1 (CA-CA) | 0.222 |
| Distance restraints 2 (N-O) | 0.254 |
| Bond length potential | 0.005 |
| Bond angle potential | 1.817 |
| Stereochemical cosine torsion poten | 52.675 |
| Stereochemical improper torsion pot | 0.988 |

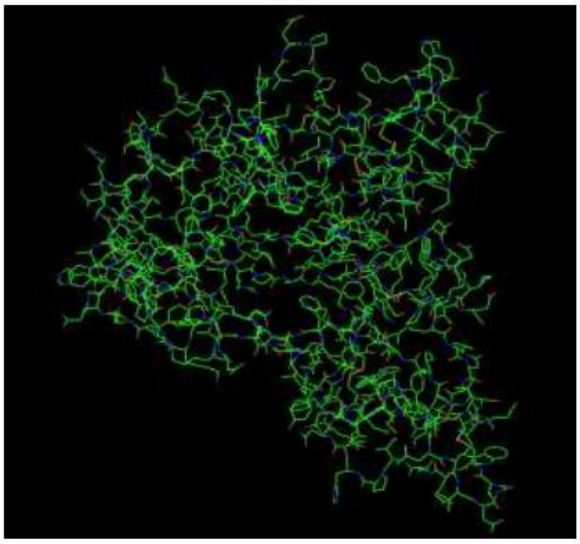
Thus overall first model generated was very closer to the original structure possible with very less value of RMS values with CPU solving time as 134.86 seconds.

Considering this model generated, 30 iterations were run and then the Energy Curve was plotted For all 30 models based on the log file generated for getting the Energy of each model. And finally minimum energy model was depicted.




Thus Structure 16 was chosen for the further analysis because it showed the minimum energy confirmation than other models. It was seen that model 16 was having the minimum energy of 1832.0832 J.

**The Minimum Energy Structure is as shown below:**



| Model No. | Energy | Model No. | Energy |
|---|---|---|---|
| 1 | 2173.187 | 16 | 1832.038 |
| 2 | 2181.284 | 17 | 2272.321 |
| 3 | 2220.731 | 18 | 2158.718 |
| 4 | 1945.198 | 19 | 1926.053 |
| 5 | 2002.511 | 20 | 1937.768 |
| 6 | 2220.157 | 21 | 2103.277 |
| 7 | 2023.343 | 22 | 2319.212 |
| 8 | 2078.338 | 23 | 2175.86 |
| 9 | 1923.541 | 24 | 2174.032 |
| 10 | 1996.077 | 25 | 1993.6 |
| 11 | 2036.177 | 26 | 2203.188 |
| 12 | 2256.98 | 27 | 2292.764 |
| 13 | 1868.111 | 28 | 1936.555 |
| 14 | 1916.066 | 29 | 1940.539 |
| 15 | 1878.384 | 30 | 2023.27 |

**MODEL ASSESSMENT**

Model 16 was submitted to the Swiss Model Workspace WebServer and different important results obtained are as follows.

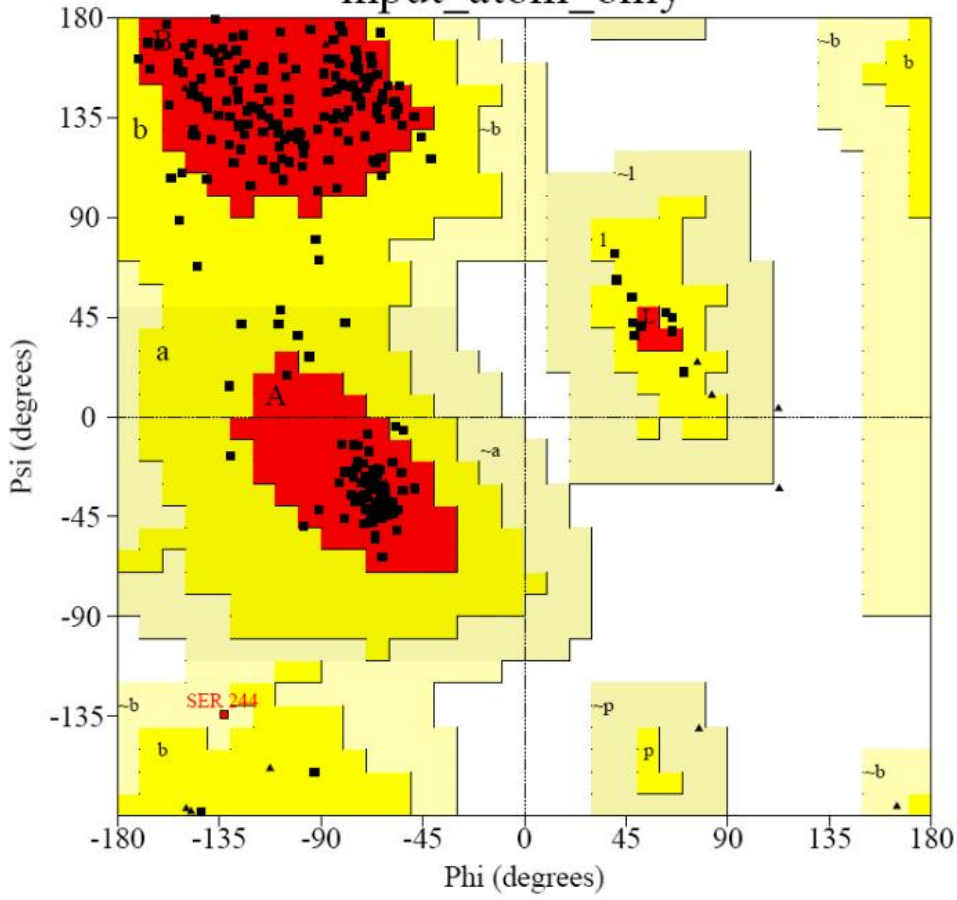
**1. Ramchandran Plot Analysis**

87.4% residues are in the most favored regions. Thus overall model is quite very well with stable confirmation of its most of the residues.

PROCHECK

# Ramachandran Plot

## input_atom_only




| | | |
|---|---|---|
| Residues in most favoured regions [A,B,L] | 222 | 87.1% |
| Residues in additional allowed regions [a,b,l,p] | 32 | 12.5% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 1 | 0.4% |
| Residues in disallowed regions | 0 | 0.0% |
| Number of non-glycine and non-proline residues | 255 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 1 | |
| Number of glycine residues (shown as triangles) | 14 | |
| Number of proline residues | 17 | |
| Total number of residues | 287 | |

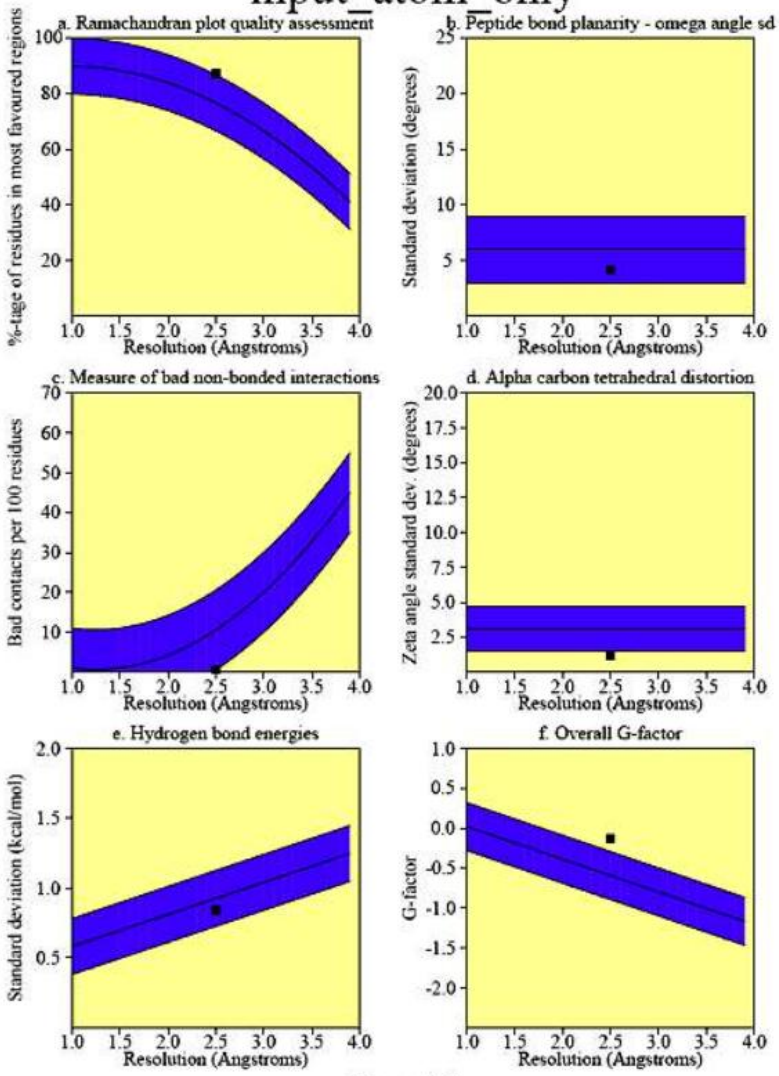
**2. Main Chain Parameters.**

Thus, model showed a significant Resolution in various parameters with overall very low Standard Deviation from the standard. Main Results are as diplayed.

PROCHECK

# Main-chain parameters

## input_atom_only




Plot statistics

| Stereochemical parameter | No. of data pts | Parameter value | Comparison values Typical value | Band width | No. of band widths from mean | |
|---|---|---|---|---|---|---|
| a. %-tage residues in A, B, L | 255 | 87.1 | 76.6 | 10.0 | 1.0 | BETTER |
| b. Omega angle st dev | 284 | 4.2 | 6.0 | 3.0 | -0.6 | Inside |
| c. Bad contacts / 100 residues | 2 | 0.7 | 10.5 | 10.0 | -1.0 | Inside |
| d. Zeta angle st dev | 273 | 1.1 | 3.1 | 1.6 | -1.3 | BETTER |
| e. H-bond energy st dev | 135 | 0.8 | 0.9 | 0.2 | -0.4 | Inside |
| f. Overall G-factor | 287 | -0.1 | -0.6 | 0.3 | 1.6 | BETTER |

### 3.0 ANOLEA Results

ANOLEA is used to figure out the level of energy for the different protein residues along with calculating the number of non-local interactions and non-local energy of the protein.

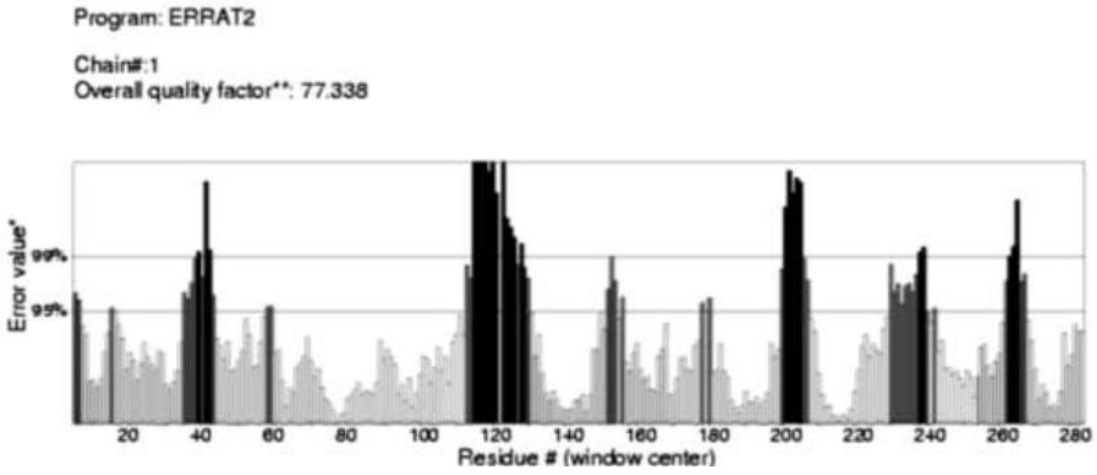
*Summary of the results is as shown:*

Total amino acids with high energy = 49
Percentage = 17.07
Total number of aminoacids = 287
Total number of atoms = 2311

Total number of non-local atomic interactions = 40632
Total non-local energy of the protein (E/kT units) = -2109
Non-local normalized energy Z-score = 0.57

### 4.0 Errat Results

It mentions overall quality factor for each aminoacid residue with the error value also.




Overall Quality factor shown was 77.338

On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value.

### 5. MOL PROBITY

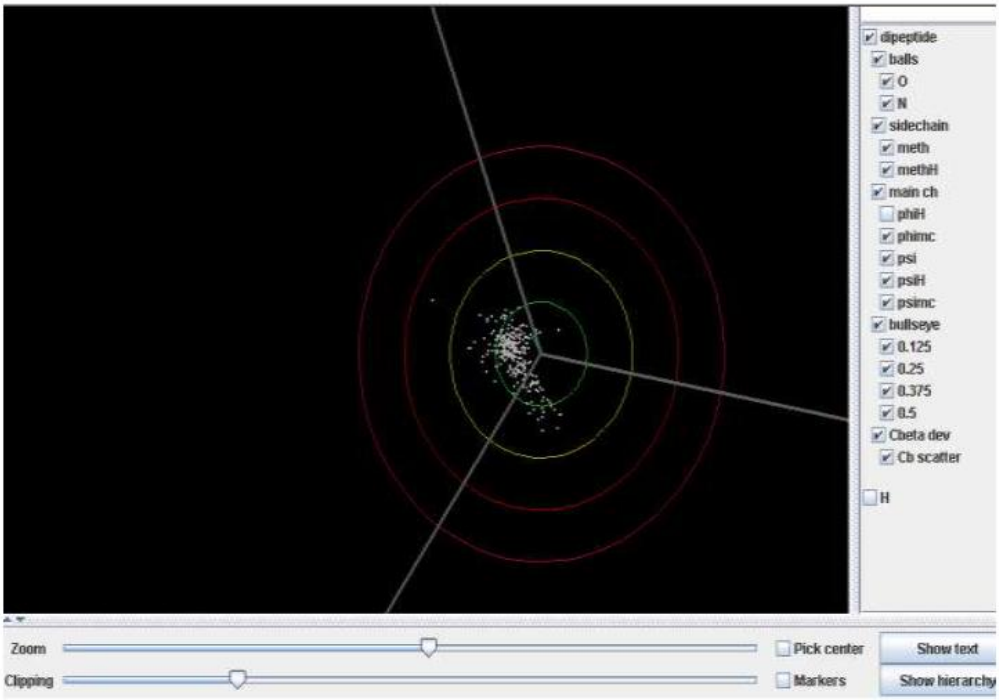
***Results Obtained in brief are as follows:***

| | | | |
|---|---|---|---|
| Protein Geometry | Rotamer outliers | 2.72% | Goal: <1% |
| | Ramachandran outliers | 0.00% | Goal: <0.2% |
| | Ramachandran favored | 96.14% | Goal: >98% |
| | Cβ deviations >0.25Å | 1 | Goal: 0 |

**Molprobity output scores:**

| | | | |
|---|---|---|---|
| Bad rotamers | : 2.7% | 7/257 | (target 0-1%) |
| Ramachandran outliers | : 0.0% | 0/285 | (target 0.2%) |
| Ramachandran favored | : 96.1% | 274/285 | (target 98.0%) |

***Cβ Scatter Plot Analysis Result***

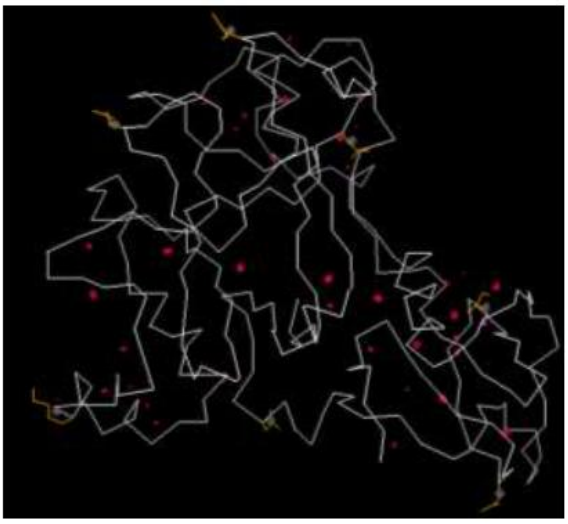


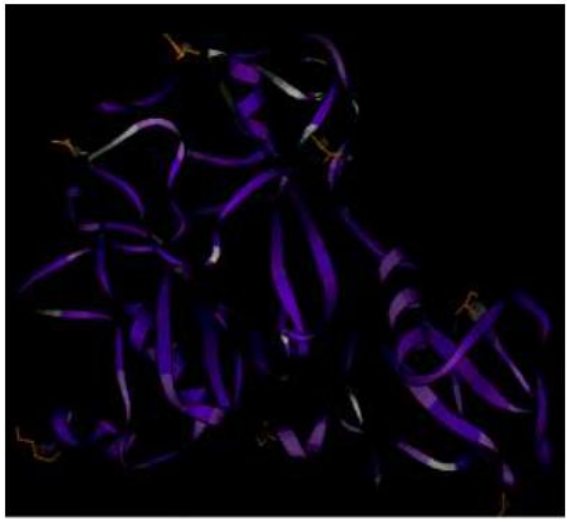

Here, it clearly shows that most of the residues are in the range from 0 to 0.125 and still some residues were just only seen from 0.125 to 0.25 scattered range from the central axis. The range limits 0.125, 0.25,0.375 and 0.5 are shown with four different circles to bound the Cβ Scatter values in the range for analysis.

Thus overall, most of the residues were just closer to 0 Cβ Scatter value as in the range before 0.125 circular axis range. Few residues were just seen outside this range. Thus, it is depicted that Cβ

Scatter was observed minimum for the proposed model and most of the structural parameters are in favor of stability and significance of this model.

***Multi-criterion Visualizations: Kinemage Result***





Here, all spots show the bad overlap and the points where the small lines at edges show the rotamer boundaries. The same rotamer with the ribbon structure is shown below:

## 5.0 Conclusion

Using protein modeling standalone tool modeler 8, building a comparative model for AMA-1 protein has been accomplished and visualized. In the absence of experimentally determined 3D structure of AMA-1 protein of *P. falciparum*, the generated protein model of AMA-1 may be used to understand the antigenic sites and regulatory function by visualizing its domains and structural conformation. This structure can be used for docking analysis for drug design against the antigenic epitopes to combat this malarial disease. The conserved nature of AMA-1 sequence and structure entails a preserved function for this molecule across various Plasmodium species. Apical membrane Antigen (AMA-1), which is an asexual blood–stage antigen of Plasmodium, is an important candidate for testing as a component of a malarial vaccine (1,2) . However, before further incorporation of this model for further work, it should be checked for its geometrical, stereo chemical and conformational accuracy before taking up for rational drug design, to avoid later complications.

**Acknowledgement:**

I wish to thank all the scientists who have evolved this beautiful field as a converging horizon of various fields.

**References:**

1. Crewther PE, Culvenor JG, Silva A, Cooper JA, Anders RF (1990). Plasmodium falciparum : two antigens of similar size are located in different compartments of the rhoptry. Exp. Parasitol. 70: 193-206.
2. Crewther PE, Matthe L, Flegg RH, Anders RF (1996). Protective immune responses to apical membrane antigen 1 of Plasmodium chabaudi involve recognition of strain-specific epitopes, Infect. Immun., 64(8): 3310-3317.
3. Dutta S, Malhotra P, Chauhan VS (1995). Sequence analysis of apical membrane antigen 1 (AMA-1) of Plasmodium *cynomolgi bastianelli*. Mol. Biochem. Parasitol., 73: 267-270.
4. Marshall VM, Peterson MG, Lew AM, Kemp DJ (1989). Structure of the apical membrane antigen I (AMA-1) of Plasmodium chabudi. Mol. Biochem. Parasitol., 37: 281-283.
5. Mary LS, Matthew M, Gavin ER, Robert LM, Richard JS, Robin FA (1996). Disulfide Bond Structure of Plasmodium Apical Membrane Antigen-1, TASSBMS 271(46): 446-452.
6. Yogesh K, Verma R, Gurudutta U, Gangenahalli VK, Singh PG, RameshChand HG, Raj R, Kumar S (2005). Homology Modeling of Anti-Apoptopic Protein Bcl-2. Bioinform. Ind. 3(3): 23-27.

Bioinformatics tools and server used

1. www.nbi.nlm.nih.gov/BLAST
2. http://swissmodel.expasy.org/workspace/index.php?func=tools_structureassessment1&userid=swissmodel@unibas.ch&token=TOKEN
3. http://protein.bio.puc.cl/anolea
4. http://www.doe-mbi.ucla.edu/People/Software/
5. http://www.doembi.ucla.edu/People/Yeates/Gallery/Errat.html
6. http://molprobity.biochem.duke.edu/index.php
7. www.rcsb.org/**pdb**/