A Bioinformatics Approach for Homology Modeling and Binding Site Identification of Triosephosphate Isomerase from *Plasmodium falciparum* 3D7

Ullah M, Hira J[1], Ghosh T, Ishaque N, Absar N

Department of Biochemistry and Biotechnology, University of Science and Technology Chittagong (USTC), Bangladesh, [1]Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Address for correspondence: Mr. M. Ullah, E-mail: babugene11@yahoo.com

ABSTRACT

Malaria is a major public health concern, and malarial parasites have developed resistance against the commonly available drugs. So now a days it is a major concern to find out a new target for drug therapy. *Plasmodium falciparum* 3D7, one of the strains of plasmodium species also lacks in a functional tricarboxylic acid cycle and solely dependent on glycolysis for its energy supply like other plasmodium species. Although enzymes of malarial parasite have been considered as potential antimalarial drug targets, a little is known about their structural biology. The tertiary structure of triose phosphate isomerase of *P. falciparum* 3D7 was determined by means of homology modeling through multiple alignment followed by intensive optimization and validation. The modeling was done by Swiss-Model Workspace. The obtained model was verified with the structure validation programs such as, PROCHECK, Verify3D, and QMEAN for reliability. The verify3D value of 0.69 indicates that the environment profile of the model is good. A self-optimized prediction method with alignment or SOPMA is employed for calculation of the secondary structural features of triose phosphate isomerase. The secondary structure indicates that the predicted 3D structure of triosephosphate isomerase of *P. falciparum* 3D7 contains 48.37% α-helix, 29.27% random coil, and 16.67% extended strand. Active site determination through CASTp suggests that this protein can be utilized as a potential drug target. However, these will further be tested by wet lab studies for a targeted vaccine design against *P. falciparum* 3D7.

Key words: Active site, homology modeling, *Plasmodium falciparum* 3D7

INTRODUCTION

Malaria afflicts 200 million people each year, making it a major cause of human morbidity and mortality worldwide.[1] The disease is caused by the parasites of the *Plasmodium* species (*Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale,* and *Plasmodium malariae*) and is transmitted by female mosquito of the Anopheles genus.[2] *P. falciparum* is the most virulent of the four species causing malaria and responsible for most malarial deaths. The particular virulence of *P. falciparum* is partly due to the ability of infected erythrocytes to adhere to a variety of host receptors and avoid splenic clearance.[3,4] Because of the increased prevalence of *P.*

falciparum strains to be resistant to current chemotherapy, treatment and control of this disease are becoming progressively more difficult.[1] The phenotypes *P. falciparum* 3D7 originated from the Netherlands represent distinct drug resistance and/or sensitivity to common antimalarial chemotherapeutics such as chloroquine and sulfadoxine.[5] The parasite also lacks a functional tricarboxylic acid cycle.[6] The asexual stage of the parasite, residing in the mature human erythrocyte, solely depends on glycolysis for its Adenosine-5'-triphosphate (ATP) requirements. Therefore, the glycolytic enzymes provide good drug targets.[7] The glucose consumption of the infected erythrocytes increases by 50- to 100-fold over that of uninfected erythrocytes. In infected erythrocytes, the flux of several glycolytic enzymes are elevated.[8] Glycolytic enzymes are believed to be associated with membrane components facilitating channeling of the substrate between consecutive glycolytic enzymes during triosephosphate metabolism.[9] Glycolysis in the parasite might be inhibited either by designing specific inhibitors for the enzymes of the pathway or by disrupting the microassembly on the cytoskeleton. Three-dimensional structures of glycolytic enzymes of *P. falciparum* might be of undoubted value in designing new strategies for therapeutic intervention. Triosephosphate isomerase is a dimeric glycolytic enzyme, which catalyzes the isomerization of d-glyceraldehyde-3-phosphate to dihydroxyacetone phosphate. Extensive mutagenesis experiments on the yeast protein suggest that triosephosphate isomerase is an "evolutionarily perfect enzyme."[10] Triosephosphate isomerase also plays an important role in gluconeogenesis, the hexose mono phosphate shunt, and fatty acid biosynthesis. It might be argued that the same triosephosphate isomerase is present in humans too and how a drug can be targeted in such a way that it does not affect the host. The triosephosphate isomerase of *P. falciparum* when compared to human triosephosphate isomerase shows many important differences.[11] There has been a recent report that during malarial infection, patients mount an antibody response to triosephosphate isomerase resulting in prolonged hemolytic anemia.[12] This raises the possibility that the protein may become localized on the erythrocyte membrane and getting partly exposed. Indeed, the triosephosphate isomerase from the parasite *Schistosoma mansoni* was also shown to be a surface antigen and has been considered for vaccine development.[13] In this paper, we have reported the 3D structure of *P. falciparum* strain 3D7 and also predicted its active site for a potential drug target by computer simulation.

MATERIALS AND METHODS

Computational methods

The computational methods of 3D-model building involved template selection, alignment of template with the target, building of the model, and evolution of the structure. The sequence of triosephosphate isomerase of *P. falciparum* 3D7 was retrieved from NCBI database (Accession ID: XP_001348552.1).

Template selection and sequence alignment

The PDB (Brookhaven Protein Databank) database[14] was extensively screened using BLAST (Basic Local Alignment Tool)[15] server, developed and maintained in Adam Godzik's laboratory at the Burnham Institute, to find out the related homologs of the query sequence. PDB database was searched by PSI–BLAST using a profile generated from nr protein database for proteins exhibiting similarity to the unknown structure.

Model building

The program CPH models-3.0 server[16] was used to build the model as a PDB file of triosephosphate isomerase of *P. falciparum* 3D7 according to the homology modeling method. The model was submitted to the Swiss-Model Workspace[17,18] to obtain the 3D-structure.

Evolution and validation of model

The steriochemical quality of the models was verified with the program PROCHECK[19] in order to select the best model. The 3D-profiling of the residue was done by Verify3D Structure Evaluation Server,[20,21] and QMEAN.[22]

Secondary structure and active site analysis

The secondary structure evolution was done with the help of a self-optimized prediction method with alignment or SOPMA.[23] Active site analysis was performed with the help of CASTp[24] for the modeled protein to further work on its docking studies. Active site analysis gives us an idea to make a grid before docking.

RESULTS

Template selection and sequence alignment

The structure of the triosephosphate isomerase of *P. falciparum* 3D7 has been constructed based on the availability of the template sequence of chain A, triosephosphate

isomerase of *P. falciparum* (PDB code: 1YDV_A).[11]

3D-model

The 3D structure of triosephosphate isomerase of *P. falciparum* 3D7 was built by homology modeling based on the template, which is obtained from protein data bank [Figure 1].

Validation of model

Validation of the structure
The φ and ψ distribution value of the Ramachandran plot of non-glycine and non-proline residues are shown in Figure 2 and Table 1.

Verification of the 3D-model
The profile score above zero in the Verify3D graph correspond to the acceptable environment of the model [Figure 3]. The global quality of the model for the target sequence was estimated by the QMEAN6 score based on linear combination of six structural descriptors [Table 2], which reflects predicted model reliability ranges from 0 to 1.[22] The raw QMEAN6 score and average Z-score for triosephosphate isomerase of *P. falciparum* 3D7 model are shown in Figure 4.
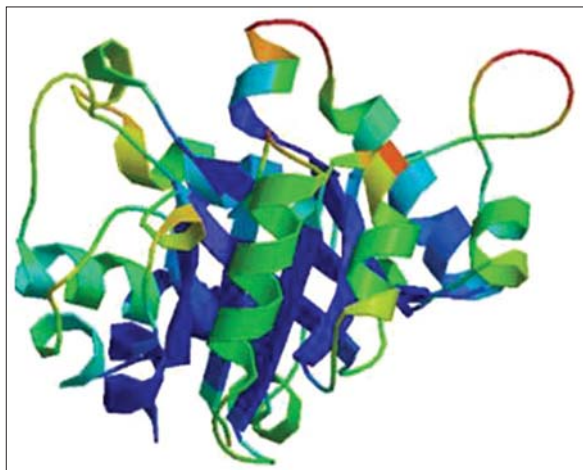
Figure 1: 3D model of triose phosphate isomerase of *P. falciparum* 3D7 by comparative modeling (Swiss Model Workspace)

Secondary structure and active site of triosephosphate isomerase

Analysis of secondary structure
The secondary structure analysis is used to know whether a given amino acid lies in a helix strand or coil. Secondary structural features as predicted using SOPMA are represented in Table 3. The secondary structure are predicted by using default parameters (window width: 17, similarity threshold: 8, and number of states: 4).

Active site determination
The active site of triosephosphate isomerase of *P. falciparum* 3D7 was predicted by using CASTp server [Figure 5a]. Further, in this study we have also reported the best active site area of the experimental enzyme as well as the number of amino acid involved in it [Figure 5b].

DISCUSSION

The homology modeling algorithm has proved to be one of the most robust modeling tools in bioinformatics. The protein model built in this study by using standard protocol indicates that the enzyme contains α-helix, β-sheets, and loops. The structure modeled by CPH models-3.0 server
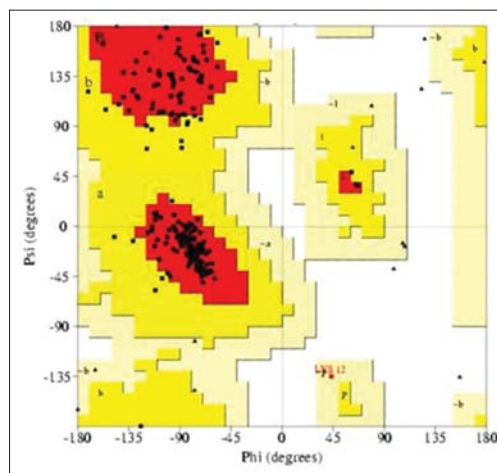
Figure 2: Ramachandran plot of triose phosphate isomerase of *P. falciparum* 3D7 obtained through the modeling tool
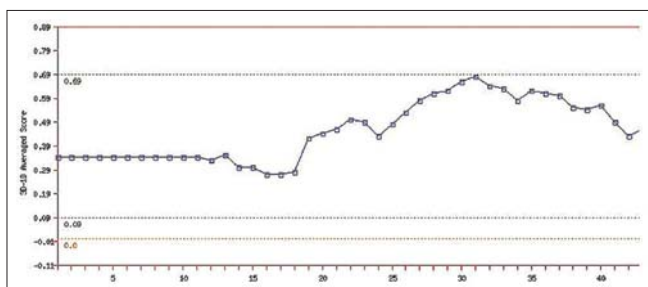
Table 1: Ramachandran plot statistics of triosephosphate isomerase of *P. falciparum* 3D7

Rama chandran plot statistics		(%)
Residues in the most favored regions [A, B, L]	204	90.3
Residues in the additional allowed regions [a, b, l, p]	21	9.3
Residues in the generously allowed regions [a, b, l, p]	1	0.4
Residues in the disallowed regions	0	0.0
Number of non-glycine and non-proline residues	226	100
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown in triangles)	14	
Number of proline residues	4	
Total number of residues	246	

Table 2: *Z* scores of individual component of QMEAN for triosephosphate isomerase (*P. falciparum* 3D7) models

Scoring function term	Z-score
C_β interaction energy	0.16
All-atom pairwise energy	−0.02
Solvation energy	1.72
Torsion angle energy	−0.38
Secondary structure agreement	0.12
Solvent accessibility agreement	0.46
QMEAN6 score	0.00

shows that the amino acid percentage in the favorable region is 90.3% and amino acid percentage in the disallowed region is 0.0%. Altogether, 100% of the residues are in favored and allowed regions. In the Ramachandran plot analysis, the residues were classified according to its regions in the quadrangle. The red area in the graph indicates the most allowed regions whereas the yellow area represent allowed regions. Glycine is represented by triangles, and other residues are represented by squares [Figure 2 and Table 1]. Analysis of PROCHECK reveals that all residues are within the limits of the Ramachandran plot. Therefore,

Table 3: Computed secondary structure elements of triosephosphate isomerase *P. falciparum* 3D7 by SOPMA

Secondary structure	(%)
Alpha helix	48.37
3$_{10}$ helix	0.00
P$_i$ helix	0.00
Beta helix	0.00
Extended strand	16.67
Beta turn	5.69
Bend region	0.00
Random coil	29.27
Ambiguous state	0.00
Other states	0.00

it can be considered as a good model. The established model of 3Dstructure for the target sequence was verified by structure validation servers, and the best structure has been selected based on high scores. The high score of 0.69 in the Verify3D graph [Figure 3] indicates that the environment profile of the model is good. Premalatha *et al.,* showed that the 3D model of putative thioredoxine from *Helicobacter pylori* have a profile score of above zero in the Verify3D graph and a score of 0.69 indicates that the environment profile of the model is good.[25] The QMEAN scoring function estimates the global quality of the models on the basis of a linear combination of six structural descriptors, and four of them are the statistical
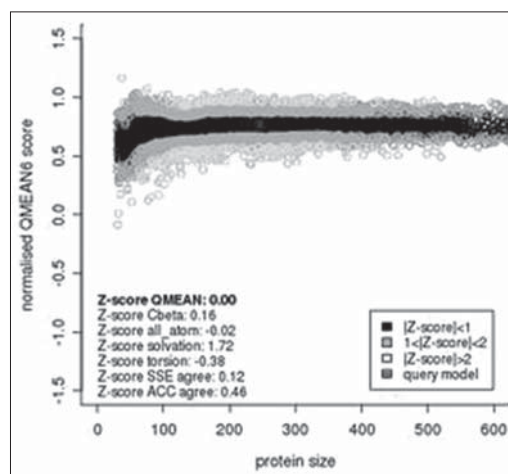
Figure 4: Graphical presentation of estimation of absolute quality of model triosephosphate isomerase (*P. falciparum* 3D7). The QMEAN *Z* scores were near "0" value for the model compared to the nonredundant set of protein structure. Here the dark zone indicates that the model has a of score <1 and the model that has have a score of >1 or <2 or >2 locates outside of this dark zone. Good models are generally located in the dark zone. The red marker indicates a query or a generated target model, which are considered to be a good model according to their position near or in the dark zone

Figure 3: Verify 3D graph of triose phosphate isomerase (*P. falciparum* 3D7)
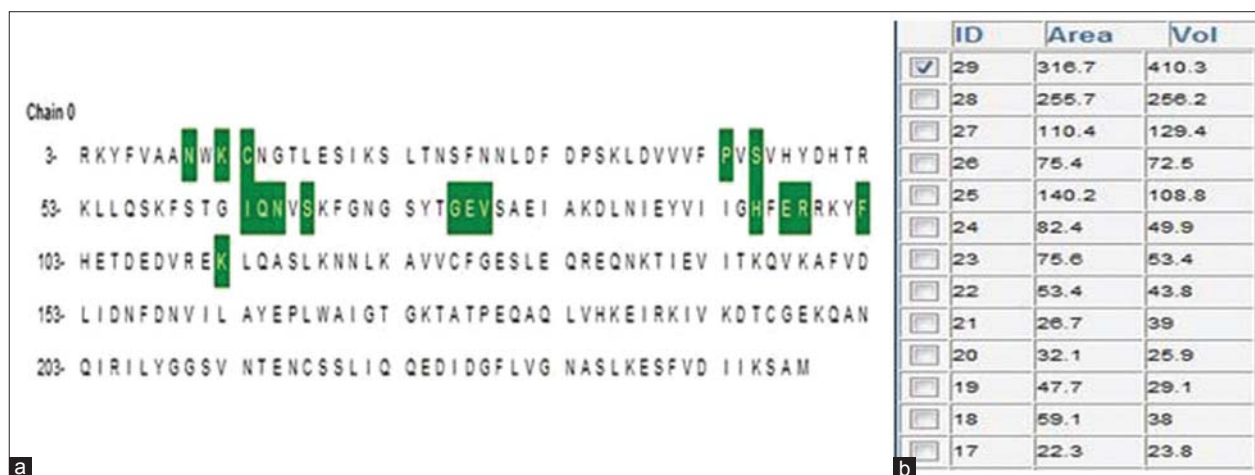
Figure 5: (a) Active site information by CASTp. Green color shows the active site position from 10 to 112 with the β-sheet in between them. (b) The table shows the area and the volume for different active sites of triosephosphate isomerase, and the best active site remains in an area of 316.7 and a volume of 410.3 amino acid

potentials of the mean force: The local geometry is analyzed by a torsion angle potential over three consecutive amino acids. Two distance-dependent interaction potentials based on Cβ atoms and all atoms, respectively, are used to assess long-range interactions.[26] A solvation potential describes the burial status of the residues. Cheng *et al.,* reported two terms reflecting the agreement between predicted and calculated secondary structure[27] and solvent accessibility.[28] It might be possible to consider as a better model as Z-scores of C_β interaction energy, salvation energy, torsion angle energy, secondary structure, and solvent accessibility are 0.016, 1.72, −0.38, 0.12, and 0.46, respectively, as shown in Table 2. The predicted global model reliability is ranging from 0 to 1.[26] According to this, the global score (0.76) of triosephosphate isomerase (*P. falciparum* 3D7) has the most possibility to be a better model. Moreover, quality of the model can be compared to reference structure of high resolution obtained from X-ray crystallography analysis through Z score and "0" is the average Z score for the good model.[22] According to Benkert *et al.*, QMEAN Z-score provides an estimation of the "degree of nativeness" of the structural features observed in a model and indicates that the model is of comparable quality to experimental structures.[22] In the present analysis, the average Z score for triosephosphate isomerase (*P. falciparum* 3D7) 3D-model is 0.00 [Table 2 and Figure 4]. The analysis of secondary structure reveals that α-helix (48.37%) dominated among secondary structure elements are followed by random coils (29.27%), extended strands (16.67%), and β-turns (5.69%) for the sequences [Table 3]. The dominance of the coiled regions indicates the high level of conservation and stability of the protein structure.[29] A binding site is a region on a protein, DNA or RNA to which other specific molecules and ions bind. The binding sites also exhibit chemical specificity, a measure of the types of ligand that will bind, and affinity, which is the measure of the strength of the chemical bond. Active site position of the building protein starting from amino acid 10 to some interaction of amino acid 112 shown in green color [Figure 5a]. Here, alphabetic letters indicate different amino acids. The major coverage of amino acid to be in building cavities lies in between 63–66, 76–78, and 95–98.

The server gives the best active site as an area of 316.7 and a volume of 410.3 amino acid is involved in the active site formation as shown in Figure 5b.

CONCLUSION

In this study, the sequence and structure analysis of triosephosphate isomerase from *P. falciparum* 3D7 was

done by various tools and software. On the basis of the findings, it could be concluded that further characterization of triosephosphate isomerase from *P. falciparum* 3D7 will be important for the regulation of the malaria. Comparative molecular modeling allows expanding the number of protein sequences for which we have structural information. Such models can be effectively used to design mutagenesis experiments and to support the drug design project. Knowledge gained from this study will be used in broad screening on inhibitors of the protein and can be further implemented in future drug designing.

REFERENCES

1. Gay F, Bustos MD, Diquet B, Rojas-Rivero L, Litaudon M, Pichet C, *et al*. Cross-resistance between mefloquine and halofantrine. Lancet 1990;336:1262.

2. Snow RW, Trape JF, Masdh K. The past, present and future of the child hood malaria mortility in Africa. Trends Parasirol 2001;17:593-7.

3. David PH, Hommel M, Miller LH, Udeinya IJ, Oligino LD. Parasite sequestration in *Plasmodium falciparum* malaria: Spleen and antibody modulation of cytoadherence of infected erythrocytes. Proc Natl Acad Sci U S A 1983;80:5075-9.

4. Howard RJ, Barnwell JW. Roles of surface antigens on malaria-infected red blood cells in evasion of immunity. Contemp Top Immunobiol 1984;12:127-200.

5. Rathod PK, McErlean T Lee PC. Variations in frequencies of drug resistance in Plasmodium falciparum. Proc Natl Acad Sci U S A 1997;94:9389-93.

6. Sherman IW. Biochemistry of *Plasmodium* (malarial parasites). Microbiol Rev 1979;43:453-95.

7. Lakhdar-Ghazal F, Blonski C, Willson M, Michels P, Perie J. Glycolysis and proteases as targets for the design of new anti-trypanosome drugs. Curr Top Med Chem 2002;2:439-56.

8. Roth E Jr, Joulin V, Miwa S, Yoshida A, Akatsuka J, Cohen-Solal M, *et al*. The use of enzymopathic humanred cells in study of malarial parasite glucose metabolism. Blood 1988;71:1408-13.

9. Orosz F, Ovadi J. Dynamic interactions of enzymes involved in triosephosphate metabolism. Eur J Biochem 1986;160: 615-9.

10. Knowles JR. Enzyme catalysis: Not different, just better. Nature 1991;350:121-4.

11. Velankar SS, Ray SS, Gokhale RS, Suma S, Balaram H, Balaram P *et al*. Triosephosphate isomerase from Plasmodium falciparum: The crystal structure provides insights into antimalarial drug design. Structure 1997;l5:751-61.

12. Ritter K, Kuhlencord A, Thomssen R, Bommer W. Prolonged hemolytic anemia in malaria and autoantibodies against triosephosphate isomerase. Lancet 1993;342:1333-4.

13. Harn DA, Gu W, Oligono DL, Mitsuyama M, Gebremichael A Ritcher D. A protective monoclonal antibody specifically recognizes and alters the catalytic activity of triosephosphate isomerase. J Immunol 1992;148: 562-71.

14. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. The Protein Data Bank. Nucleic Acids Res 2000;28:235-42.

15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.

16. Nielsen M, Lundegaard C, Lund O, Petersen TN. CPHmodels-3.0 - Remote homology modeling using structure guided sequence profiles. Nucleic Acids Res 2010;38(Web Server issue):W576-81.

17. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 2009;4:1-13.

18. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace:

A web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195-201.

19. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477-86.

20. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164-70.

21. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356:83-5.

22. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 2011;27:343-50.

23. Geourjon C, Deleage G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995;11:681-4.

24. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, and Liang J. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated resiudes. Nucleic Acids Res 2006;34(Web Server issue):W116-8.

25. Premalatha D, Ravindra P, Venkateshwer Rao L. Homology modeling of putative thioredoxin from *Helicobacerr pylori*. Indian J Biotechnol 2007;6:4859.

26. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res 2009;37(Web Server issue):10-4.

27. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: A protein structure and structural feature prediction server. Nucleic Acids Res 2005;33(Web Server issue):72-6.

28. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195-202.

29. Neelamathi E, Vasumathi E, Bagyalakshmi S, Kannan R. Insilico prediction of structure and functional aspects of a hypothetical protein of Neurospora crassa. J Cell Tissue Re 2009;9:1889-94.

Announcement

Android App

A free application to browse and search the journal's content is now available for Android based mobiles and devices. The application provides "Table of Contents" of the latest issues, which are stored on the device for future offline browsing. Internet connection is required to access the back issues and search facility. The application is compatible with all the versions of Android. The application can be downloaded from https://market.android.com/details?id=comm.app.medknow. For suggestions and comments do write back to us.