

Homology Modeling of Alpha-Glucosidase from *Candida albicans*: Sequence Analysis and Structural Validation Studies *in silico*

Haroon Ur Rashid,^{1b, a, b} Vanderlan S. Bolzani,^{1b, a} Khalid Khan,^{1b, c} Luiz Antonio Dutra,^{1b, a}
Nasir Ahmad^{1b, *, c} and Abdul Wadood^{1b, *, d}

^aInstituto de Química, Universidade Estadual Paulista Julio de Mesquita Filho,
14800-060 Araraquara-SP, Brazil

^bCentro de Ciências Químicas, Farmacêuticas e de Alimentos, Universidade Federal de Pelotas,
96010-900 Pelotas-RS, Brazil

^cDepartment of Chemistry, Islamia College University, Peshawar, 25120, Khyber Pakhtunkhwa,
Pakistan

^dDepartment of Biochemistry, Shankar Campus, Abdul Wali Khan University, Mardan, 23200,
Khyber Pakhtunkhwa, Pakistan

The alpha-glucosidase enzyme of *Candida albicans* plays a vital role in the pathogenesis of candidiasis, a serious fatal disease in immune-compromised patients. The unavailability of the three-dimensional crystallographic structure of this enzyme creates a hindrance in developing novel and potent inhibitors. Here, an attempt has been made to design a stable three-dimensional conformer of alpha-glucosidase through *in silico* analysis which may be helpful for the designing of effective drugs. For this purpose, the oligo-1-6-glucosidase enzyme is used as a template for homology modeling of the alpha-glucosidase structure by Molecular Operating Environment 2011-12 software. The generated model was validated through ERRAT and Ramachandran tools, whereas its stability was studied through the molecular dynamics simulation technique. The obtained results indicate that model of alpha-glucosidase has stable secondary and tertiary arrangements. This finding may spur new directions for the rational designing and development of new antifungal inhibitors. Nevertheless, additional experimental investigations and validation are needed to confirm the *in silico* results of this study.

Keywords: alpha-glucosidase, candidiasis, homology modeling, simulations

Introduction

The infections associated with the fungus have been enhanced at a large scale which leads to the enhancement of the mortality rate in humans. The species of *Candida* are mostly human pathogens responsible for infections of deep tissues and mucosa. The species of candida are related to the microbiota of a patient mucosal oral cavity vagina and gastrointestinal tract;¹ and cause different diseases ranging from an overgrowth of mucocutaneous to the infections of blood² which include candidemia a major disease that causes death. Candidiasis is included among the infections of *Candida* species.³ *Candida* infection may present as

angular cheilitis, acute pseudomembranous, acute atrophic, chronic atrophic, or chronic hyperplastic candidiasis. There are many reasons like diabetes, malignancies, and weak immunity which strengthen oropharyngeal candidiasis. Denture stomatitis and oropharyngeal candidiasis are caused by *Candida albicans*. The occult esophageal infection and oral fungal disease are linked with each other while the past study of odynophagia warned medical scientists about the development of systemic candidiasis.⁴ Normally, healthy humans also carry the above-mentioned fungus but they caused disease in human immunodeficiency virus (HIV) patients because they have weak immunity. The composition of the genus involved a diverse group that included 17 different *Candida* species which cause infection in humans about 90% of infections are caused by *Candida albicans*, *Candida glabrata*, *Candida parapsilosis*,

*e-mail: nasirdft@gmail.com, awadood@awkum.edu.pk
Editor handled this article: Paula Homem-de-Mello (Associate)



Candida tropicalis, and *Candida krusei*.⁵ A large number of proteins are secreted by *Candida albicans* which involved about 225 proteins and these are responsible for the formation of biofilms, invasion of tissues, damage of the immune system, and maintenance of cell walls as well as acquirement of nutrients like metal ions.⁶ The attachment of the *Candida albicans* to the host is made possible by the cell wall. The composition of the cell wall involved chitin which forms the inner layer and beta-1,3- and beta-1,6-glucans while mannoprotein forms the outer layer which is the 40% mass of cells of yeast.⁷ The mannoproteins have a significant role in adhesion and also contribute to antigenicity, and tempering of the host immune response; due to these mannoproteins the innate immune cells recognized the fungus.⁸ To understand the molecular basis of the relationship between *Candida albicans* and humans, it is essential to study the glycosylation of cell walls because the richer glycosylated mannoproteins play a significant role in the interaction of pathogen and host.^{9,10} *N*-Glycosylation is a posttranslational modification that is initiated in the endoplasmic reticulum (ER), where the Glc3Man9GlcNAc2 *N*-glycan (Glc-glucose, Man-mannose, GlcNA-*N*-acetylglucose amine) is processed by alpha-glucosidases I and II and alpha 1,2 mannosidase to generate Man8 (GlcNAc)2, Mannoctaosedi-(*N*-acetyl-D-glucosamine) D1, D3, oligomannose-8 glycan. The elaboration of this *N*-oligosaccharide takes place in the Golgi to structure *N*-glycans with highly branched outer chains rich in mannose. Therefore, *N*-oligosaccharide processing by ER-alpha-glycosidase to generate high-mannose *N*-glycans is vital for the host-fungus interaction and virulence.¹¹⁻¹⁵ It is evident from the above-mentioned studies that alpha-glucosidase plays an important role to initiate the *N*-glycosylation process and maintain the integrity of the cell wall which makes possible the normal host-guest interaction. This enzyme plays a significant role in causing various diseases in humans through *Candida albicans*, so it is important to find out novel drugs to inhibit it, but the unavailability of the 3D structure of this enzyme creates a hindrance in explaining the interaction of the enzyme with probable inhibitors.

Homology modeling is the most authentic and promising tool of the computer-based technique for explaining the 3D structure of alpha-glucosidase of *Candida albicans* in the absence of the crystal structure of the protein. This article describes the homology modeling of alpha-glucosidase of *Candida albicans* and provides help in the designing of new drugs. Notably, this work is of speculative and theoretical nature. Therefore, experimental studies would be required in the future to confirm the results obtained in this *in silico* investigation.

Methodology

Protein sequence retrieval

The amino acid sequence of the alpha-glucosidase (accession No. Q02751) of *Candida albicans* was acquired in FASTA file format from the Universal Protein Resource (UniProt).¹⁶ As the protein data bank (PDB),¹⁷ did not contain the 3D structure of the protein, the task was to develop the 3D model of alpha-glucosidase.

Prediction of secondary structure

The sequence of the alpha-glucosidase was subjected to the operation to get the secondary structures in the following online available servers (DPM),¹⁸ Discrimination of protein secondary structure class (DSC),¹⁹ PHD,²⁰ PREDATOR,²¹ SIMPA96,²² SOPM,²³ self-optimized prediction method with alignment (SOPMA),²⁴ HNN²⁵ GOR V.²⁶

Prediction of intrinsic disorder

In alpha-glucosidase, the disordered regions and flexible regions were determined with the help of three online tools namely DisEMBLE,²⁷ Globplot,²⁸ and regional order neural network (RONN).²⁹ Prediction of the protein disordered is indispensable to finding out the function of protein and protein folding pathway.³⁰

Homology modeling of alpha-glucosidase and identification of active sites

Molecular Operating Environment (MOE) 2011-12 comparative modeling techniques were applied³¹ to get a deep understanding of the structural properties of the alpha-glucosidase protein of *Candida albicans*. With the lack of crystal structure of alpha-glucosidase, molecular modeling is the best tool that can provide an accurate structure model for broad-spectrum applications in the future and provide help in the field of drug design.

To generate the model of alpha-glucosidase the following four steps were taken into account: first, selection of the template; second, sequence alignment; the third involved modelgeneration with the identification of active sites present in the protein which were recognized from the 3D atomic coordinates of the receptor using Q-SITE FINDER. It is an energy-based method for the prediction of protein-ligand binding sites. Finally, refinement and evaluation of the model was done.

Template selection

The (PSI-BLAST) tool was used against PDB under the following default parameters like E-value threshold 10, size 3, and blossom 62 matrices to perform a similarity search. About three iterations of PSI-BLAST were considered as the blast search results. As a result, the high-resolution X-ray crystallography structure of the (Oligo-1,6-glucosidase IMA1 *Saccharomyces cerevisiae*) was selected as the template protein showing 49% identity with the target protein. It was used to construct a model on MOE 2011-12.

Target template alignment

The CLUSTAL W program of MUSCLE SERVER³² was used for the target template alignment and the sequence alignment was performed between the template and target. The result was obtained in the form of BLOSUM scoring matrix which was selected with a gap penalty of 10 and a gap extension of 0.05.

Model generation

The MOE 2011-12 model construction program was used for the generation of the model and the sequence of protein along with the template was used in it. After the adjustment of parameters, the job was run and a total of 10 models were formed while one of the stable models was selected. The MOE-Align application was used to align the sequence before the homology modeling. The MOE model was selected for the dynamics simulation as it was found more stable.

Molecular dynamics (MD) simulation

Out of ten generated models, the stable one was selected for MD simulation. The calculations were done using MOE software 28 employing MOE Force field AMBER99 by concerning calculation toward solvation energy with Bornimplicit solvation. The minimization of energy was done via RMS Gradient 0.1. The optimization of the total charge of protein was carried out with a partial charge option in MOE 2011-12. All parameters were kept at their default values which included the ensemble at NVT (N: number of atoms; V: volume, T: temperature) and NPA (Nosé-Poincaré-Anderson) for the algorithm to get the ensemble trajectory. The acceleration, velocity, and position were saved after every 0.5 picoseconds (ps). Before starting, the simulation of water molecules was added to the protein through the water soak option with soak mode BOX having

a layer width of 5 followed by energy minimization. The system was heated from 0 to 300 K in 20 ps (heat time) followed by a production time of 1200 ps. The system was cooled back to 0 K in 20 ps.

Evaluation and validation of the refined model

The structure obtained was validated to get the stereochemical quality. It was then evaluated using RAMPAGE SERVER³³ and ERRAT PROGRAM.³⁴ The above-mentioned programs were employed to evaluate the reliability of the homology model.

Results and Discussion

Recovery of primary sequence

The prime sequence of the alpha-glucosidase (accession No. Q02751) of *Candida albicans* was recovered in the FASTA set-up from the Universal Protein Resource (UniProt).¹⁶ The sequence of alpha-glucosidase was then characterized *in silico* using the Expasy-program tool.³⁵ The protein has 570 amino acids and the estimated molecular weight is 66210.3 Da. The isoelectric point (pI) is 5.18. The instability index (II) is worked out to be 29.70 which showed that the protein is stable.³⁶ The aliphatic index is 68.58.³⁷ The value of the Grand average of hydropathy (GRAVY) is -0.65 which shows that it is hydrophilic.³⁸

Secondary structure prediction

The properties of secondary structure were obtained on various tools that explain the position of given amino acid that lies in a helix, strand, and coil. After the comparison of results obtained through various secondary structures and prediction servers, it was observed that random coils percentages are high in all secondary structures properties above 50%, followed by alpha helix at about 30%, followed by extended strand at about 15%, and finally by beta sheets at 7% given by server DPM. The details of the secondary structure are given below in Table 1.

Intrinsic disorder identification in protein

The following online servers were used GlobPlot, DisEMBLE, and RONN to identify the intrinsic disorders in alpha-glucosidase and it was found that the regions (107-120 and 548-570) were common (Table 2). In the case of GlobPlot, Linding³⁰ designed a model, called GlobPlot, based on the amino-acid natural tendency. They proposed a new scale for tendencies, named Russell/Linding to be

Table 1. Details of secondary structure elements

Secondary structure	DPM / %	DSC / %	HNN / %	MLRC / %	PHD / %	Predator / %	SIMPA96 / %	GOR4 / %
Alpha helix	31	28.42	30.70	30.00	27.72	27.19	33.98	28.95
3 ₁₀ helix	0.00	0.00	0.00	0.00	00.00	0.00	0.00	0.00
Pi helix	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Beta bridge	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Extended strand	8.95	11.93	15.79	15.96	15.61	11.40	12.61	18.95
Beta turn	7.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bend region	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Random coil	52.2	59.65	53.51	54.04	56.67	61.40	53.24	52.11
Ambiguous states	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Other states	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00

DPM: Double Prediction Method; DSC: Discrimination of Protein Secondary Structure Class; HNN: Hierarchical Neural Network; MLRC: Multivariate Linear Regression Combination; PHD: Profile network from Heidelberg; Predator: Protein Secondary Structure prediction from a single sequence or a set of sequences; SIMPA 96: Self Optimized Prediction Method from Alignment; GOR4: Garnier-Osguthorpe-Robson.

either in regular secondary structures (alpha-helices or beta-strands) or outside of them ('random coil', loops, turns, etc.). In Figure 1a, the blue color indicates the disorder region and the green color is the ordered region while the range of amino acids is mentioned in Table 2.

The DisEMBL webserver (Figure 1b), shows predictions about alpha-glucosidase which involved the disordered regions expressed in Table 2. Plot (Figure 1b) consisted of the given curves; the green curve is the prediction for missing coordinates, red for the hot loop network, and

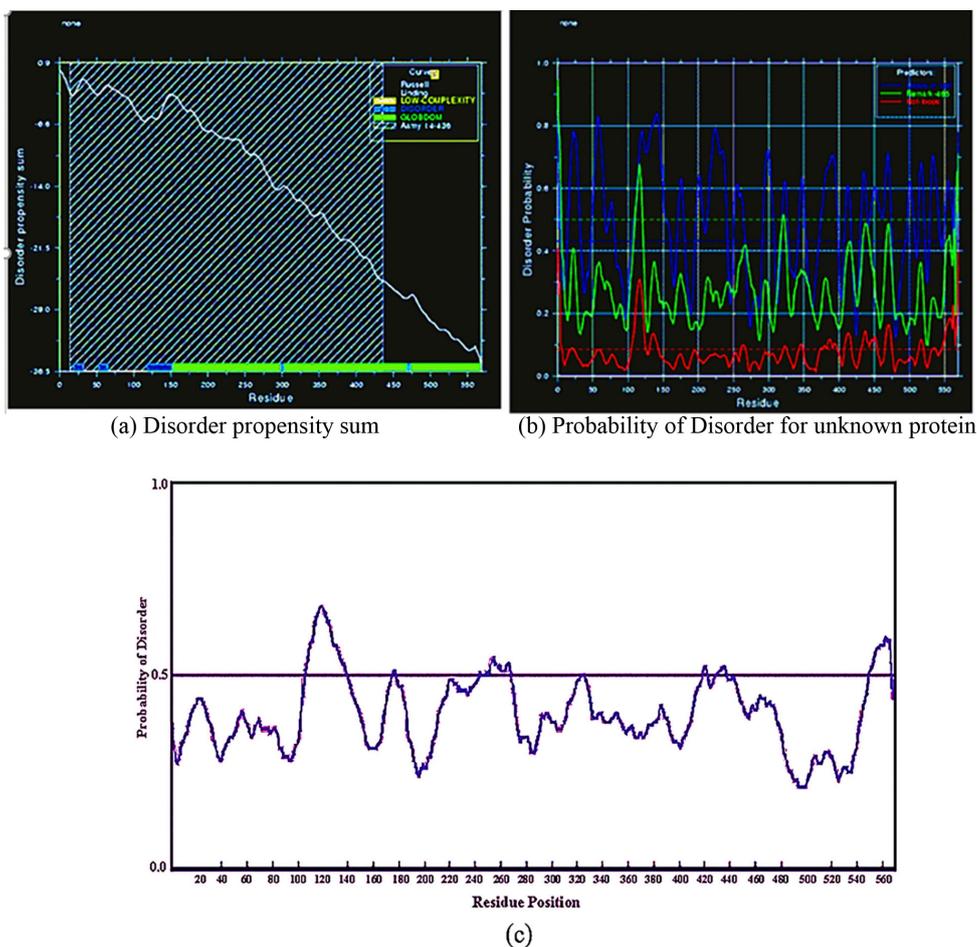
**Figure 1.** Protein intrinsic disorders of alpha-glucosidase (of *Candida albicans*) predicted by servers (a) DisEMBL (b) Globplot (c) RONN.

Table 2. The detailed intrinsic disorder profile of alpha-glucosidase predicted by using different servers

Server	Disordered	Disordered by REM465	Disordered by Loop/Coil definition	Disordered by Hot-Loop definition	Disordered by Russell/Linding definition
GLOBPLOT					19-33,52-66, 117-152,298-303, 468-473
DISEMBL		111-121	16-34,49-82,107-161, 171-181,197-244, 251-267,293-208, 319-328,343, 353, 373-399, 409-418, 431-475,496-508, 513-523,530-544, 549-570	106-125,132-144, 409-418,432-455, 467-475,548-570	
RONN	107-138,176-177, 244-244,249-249, 251-267,420-422, 431-438,550-567				

RONN: regional order neural network

blue for the coil. The horizontal lines correspond to the random expectation level for each predictor; for coils and hot loops, the prior probabilities were used, while a neural network score of 0.5 is used for REMARK465. From this plot, it is seen that the predictors agree on residues present in Table 2 as being disordered. RONN in the graph (Figure 1c) composed of a black line that indicates the standard position. The residues which are above from that are in the disordered region while those below are in the ordered region (Table 2).

Target-template alignment

A precise sequence alignment is indispensable to the success of homology modeling. A multiple sequence alignment was performed using CLUSTAL W program of MUSCLE SERVER keeping all the parameters in the default range.³⁹ The sequence alignment results of the alpha-glucosidase with template oligo-1,6-glucosidase showed about 49% of the identity of both the target and template.

In Figure 2, the conservation of the strong groups is represented by (:) while weak groups are expressed by (.), deleted regions are denoted by (-), and fully conserved residues are shown by (*).

Homology modeling and identification of active sites

Homology modeling estimates the three-dimensional structure of a target protein sequence by using its alignment to one or more protein templates of known structure. Homology modeling is the most suitable method for structure-based protein molecule design and function investigation.

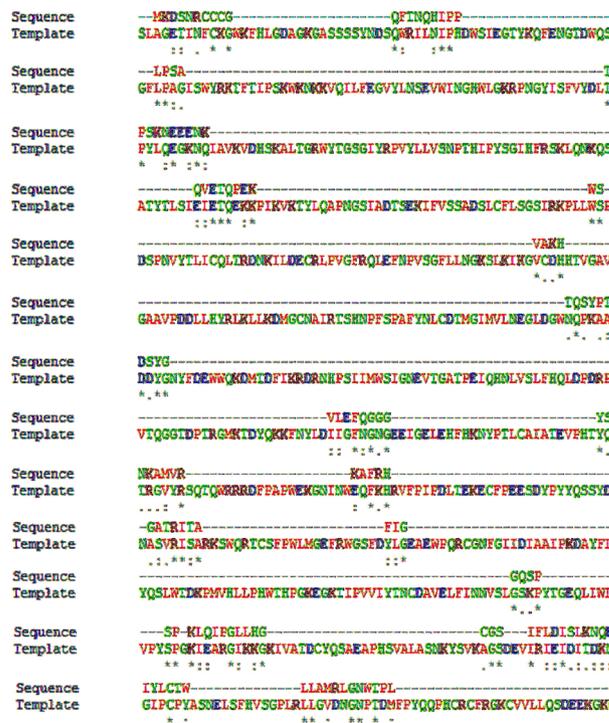


Figure 2. Sequence alignment between alpha-glucosidase (of *Candida albicans*) and template 3A47-A N-terminal performed by CLUSTAL W program of MUSCLE SERVER.

The modeling approach can be summarized into the following steps, (i) the derivation of a new model on MOE 2011-12 software based on relative objective functional values, (ii) the MD simulation step, and finally (iii) the validation step. MOE 2011-12 software was used to generate the model through the following requisite steps; first, the MOE was used to copy the initial partial geometry of the target sequence from the template structure whereas the residue’s identity was conserved. Then, insertion and

deletion phenomena were taken into account to treat those residues having no assigned backbone coordinates.⁴⁰ Furthermore, the loops were molded in random order. The contact energy function which is based on Boltzmann weighted averaging was chosen to analyze a list of possible candidates.^{41,42} An adjustment in the parameter of Model Scoring to Generalized Born/Volume Integral (GB/VI) methodology was made, and the model was subsequently generated.³¹ In the MOE software, force field Amber 99 is recommended for protein homology applications⁴³ so it was selected. Force field AMBER99 was used to minimize energy to 0.1 for the structure formed and AMBER99 is specifically used for the protein and nucleic acid and not recommended for the small molecules. At the end of the process, a new model was formed which was saved in the PDB format with a proper name for recognition (Figure 3a) and through MOE site finder application the active sites were identified. From the site finder tool, it was clear that following residues are active sites of the predicted model Val11, Tyr13, Val49, Leu51, Asp60, Gly63, Tyr64, Leu98, Leu100, Ile102, Asn103. Then, the superimposition of the native model on the template was done with root mean square deviation (RMSD) of 3.14 Å showing a close homology (Figure 4b) using MOE 2010-11. Consequently, the superimposed model was obtained.

Dynamics simulation

The MD simulation method was used to refine the model. It was performed through the MOE 2011-12 software simulation program.⁴⁴ A few appropriate adjustments were made to initiate the process of MD simulation of the generated model. At first, the saturation of the generated structure with partial charges was carried out and after that, the energy of the system was

minimized to 0.1 RMS (root mean square) with the adjusted AMBER 99 force field of MOE 2011-12.⁴⁵ The NVT ensemble which involves N for the number of atoms, V for volume, and T for temperature, in MD simulation these all values will be kept constant.⁴⁶ For the generation of true trajectories of simulations the highly accurate and sensitive method NPT algorithm while the NPA method produces theoretically correct NVE, NVT, NPH, and NPT ensembles (N, V, T, H, E, P shows the number of atoms, volume, temperature, enthalpy, energy, and pressure, respectively). The parameters adjustment involved the nominal temperature in Kelvin fixed from 0 K to initiate simulation and 20 K for heating while 20 K for cooling of the system when the temperature was set then simulation was run for 1200 ps and finally equilibrium state was reached in a range of 1000-1200 which expressed the stability of protein at human body temperature. To attain stable bond energies the system was cooled down for 20 ps. The interpretation of results can be done through different methods, while in this work the potential energy graph for different conformations was a plot against time. The resulting conformation expressed the three-dimensional accomplished order of protein which can be altered without fluctuating covalent bonds.⁴⁷ The RMSD plot expressed stability of the model at 1100 to 1200 ps and the RMSD value is 0.2 Å (Figure 4) whereas similar stable conditions were observed in the potential energy plot of protein conformation *versus* time given in (Figure 5).

The RMSD value of the model was noted to be 0.2 Å after the simulation and was found in range with the template. Similarly, the superposition of initial and final models gave the RMSD value of 3.14 Å after simulation. These results expressed the stability of the model at 300 K temperature. It is given in Figure 6.

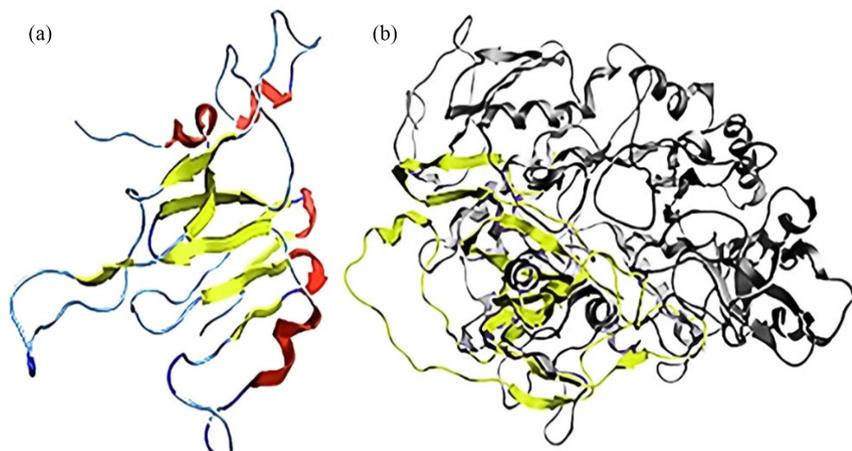


Figure 3. (a) Predicted structure of alpha-glucosidase of *Candida albicans*; yellow color represents beta sheets, red color represents alpha helix, the light blue color represents random coiled regions. (b) Structural superimposition of predicted structure (yellow color), and 3A47-A (gray color) showing high structural similarity.

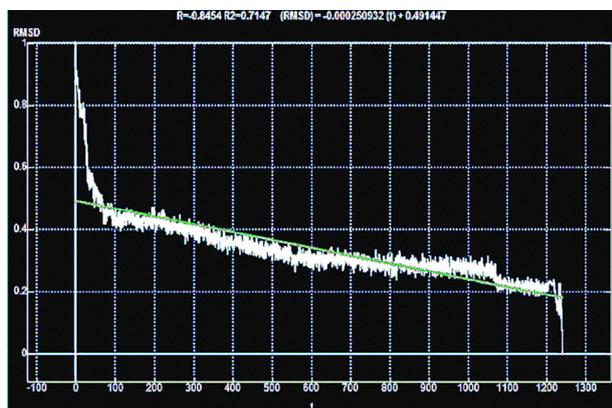


Figure 4. RMSD plot of alpha-glucosidase of *Candida albicans* obtained during MD simulation at 300 K.

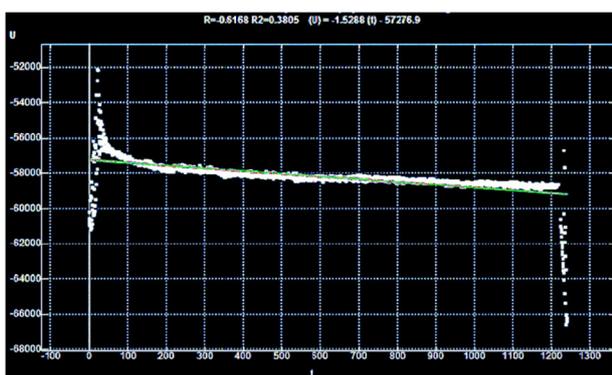


Figure 5. Potential energy plot of alpha-glucosidase of *Candida albicans* obtained during MD simulation at 300 K.

Evaluation and validation of the refined model

The appropriate steps were taken to assess the quality of the model and therefore the accuracy of the predicted model was subjected to a series of tests. Stereochemical quality was calculated using a Ramachandran plot obtained



Figure 6. Superposition of initial yellow color and final gray color during MD simulation of alpha-glucosidase.

from the RAMPAGE server. Stereochemical assessment of backbone Psi and Phi dihedral angles divulged that 76.4% were present inside the most favored regions, 16.4% of residues were found in additionally allowed regions, and 6.7% of residues were falling within disallowed regions of the Ramachandran plot (Figure 7). Besides that, Ramachandran plot analysis of MOE 2011-12 software showed that several residues GLN-49, GLN-34, and ILE-112 LYS-46 were placed out of energetically favored regions in the plot. The remaining residues are in the core regions of Ramachandran's plot, which indicated that the final structure is highly reliable for further studies. Totally, 98% of the residues are in the most favored and allowed region (Figures 7a and 7b).

Compared to the template (PDB ID 3A47-A) three-dimensional crystallographic structure, the residues in the outlier region are different, the template Ramachandran outliers have zero value while the predicted model has

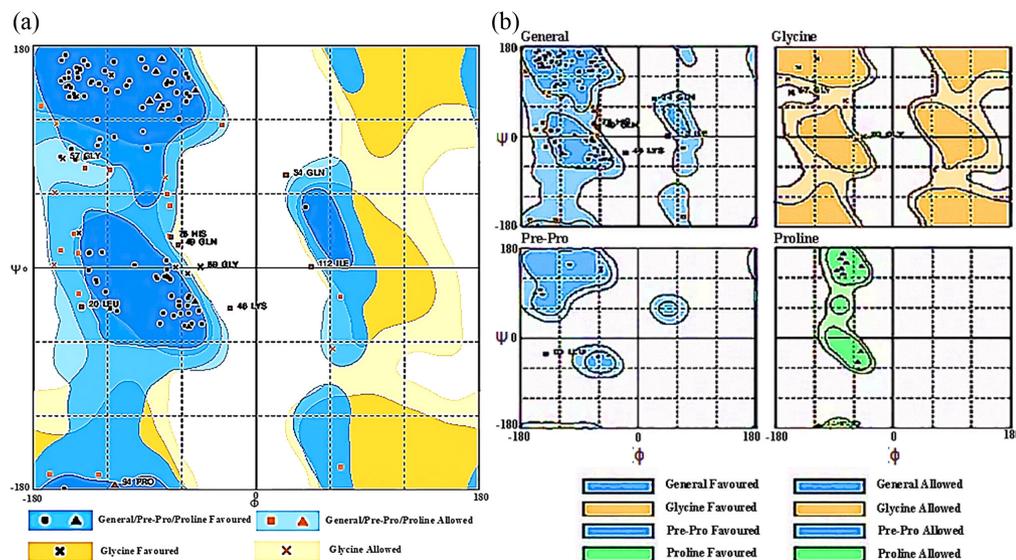


Figure 7. RAMPAGE defined results in groups of favored, allowed, and outlier categories. RAMPAGE showed 76.4% of the favored region, 16.4% allowed region and 6.7% outlier region of alpha-glucosidase. The GLN-49, GLN-34, and ILE-112 LYS-46 were out of the energetically favorable region.

6.7% residues in the disallowed region. For this model, the ERRAT score is 82.31 which covered the normal range for a high-quality model (Figure 8); on the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed the error value. The errate is used to explicate the statistics of non-bonded interaction between different atoms and a score of 50% is

normally acceptable. In Figure 8, the gray bars illustrate the region of error, the black bars show the misfolded area and the white bars demonstrate the region for the folding of the protein having less error rate. While comparatively difference exists between template ERRAT values which is 94.810% and predicted model value 82.31% but the standard and acceptable value is 50% (Figures 9a and 9b).

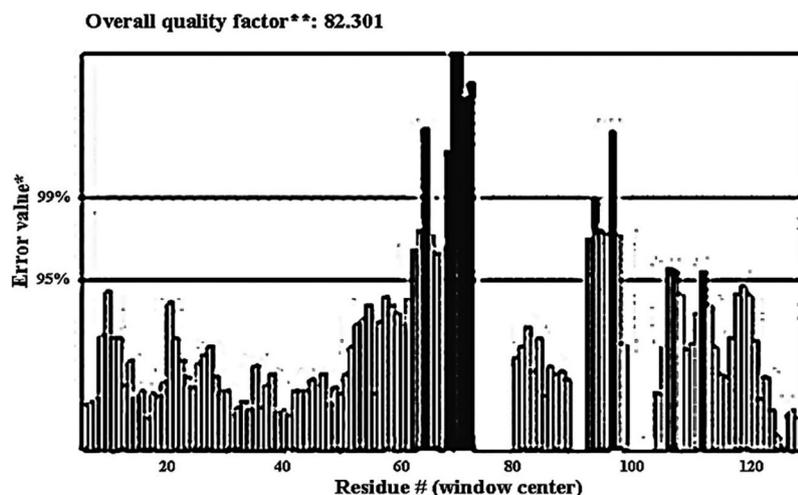


Figure 8. Overall quality factor (by ERRAT 2) of alpha-glucosidase of *Candida albicans*.

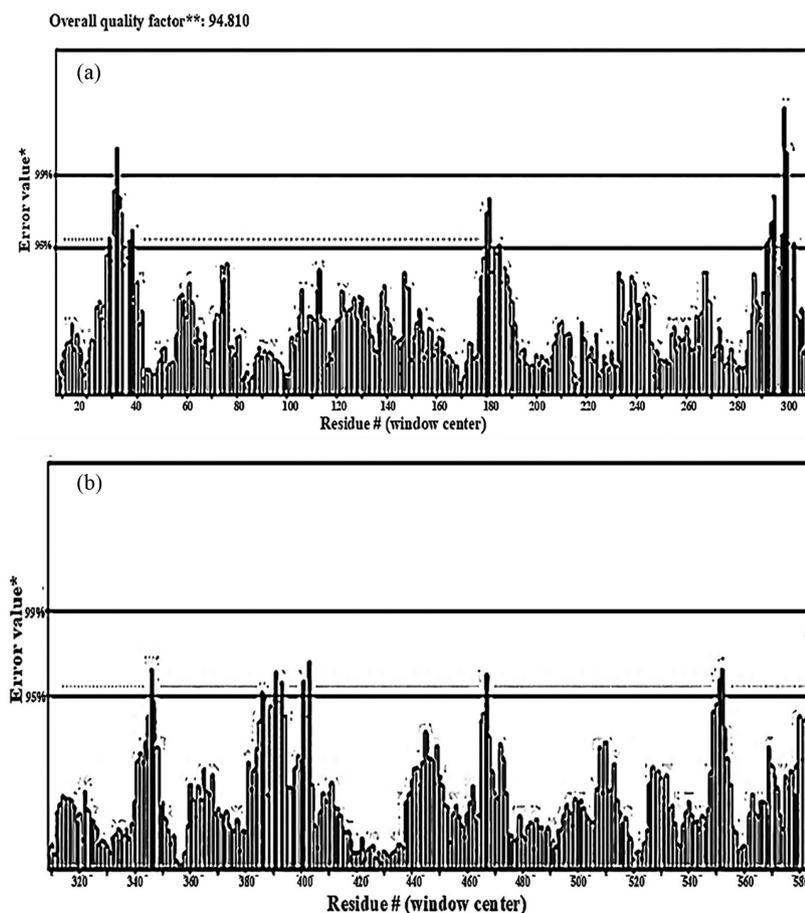


Figure 9. Overall quality factor of template PBD ID 3A47-A predicted by ERRAT 2.

Conclusions

Candida albicans causes lethal diseases to human beings including candidemia and candidiasis while the alpha-glucosidase enzyme plays a significant role in its pathogenesis. Comparative protein modeling is of great assistance during the rational design of drug molecules. In the dearth of experimental data, model building based on the known crystal structure of a homologous protein is the lone unswerving method to gain structural information. We presented the stable 3D model of alpha-glucosidase by homology modeling method. This was built on the template of the oligo-1-6-glucosidase enzyme. The intrinsically disordered regions were found through various servers which will help to explore those intrinsically unstructured proteins involved in several fungal diseases. The validation study of the predicted model was done through ERRAT server and Ramachandran plot in comparison to the template. The resulting values established the validity of the model whereas further stability of the predicted model was assessed through MD simulation. The RMSD plot confirmed its stability by a high-resolution RMSD value of 0.2 Å between 1100 to 1200 ps. The model obtained in this study may serve many significant purposes and it may be used as a valid novel drug target for designing antifungal drugs. In the absence of the experimental structure, this model will provide a foundation for elucidating the structure-function relationship and pave a way for rational drug design.

Acknowledgments

The authors thank the financial support from National Council for Scientific and Technological Development (Conselho Nacional de Tecnologia e Desenvolvimento) (CNPq)-project number-INCT-CNPq 465.637/2014-0 and the Dean of Research and Graduate Studies, São Paulo State University, Araraquara, SP, Brazil. The authors are also thankful to Universidade Federal de Pelotas (UFPeL), Pelotas, RS, Brazil for providing support to perform this project.

Author Contributions

Haroon ur Rashid was responsible for data curation, investigation, writing-review and editing, and software; Vanderlan da Silva Bolzani for formal analysis, funding acquisition, and visualization; Khalid Khan for project administration, conceptualization, and visualization; Luiz Antonio Dutra for reviewing, editing, and software; Nasir Ahmad for validation, software, writing an original draft, and writing-review and editing; Abdul wadood for visualization, investigation and software.

References

1. Ajenjo, H. M. C.; Aquevedo, S. A.; Guzmán, D. A. M.; Poggi, M. H.; Calvo, A. M.; Castillo, V. C.; León, C. E.; Andresen, H. M.; Labarca, L. J.; *Rev. Chil. Infectol.* **2011**, *28*, 118. [Crossref]
2. Seneviratne, C. J.; Wong, R. W. K.; Samaranyake, L. P.; *Mycoses* **2007**, *51*, 30. [Crossref]
3. Pappas, P. G.; Kauffman, C. A.; Andes, D. R.; Clancy, C. J.; Marr, K. A.; Ostrosky-Zeichner, L.; Reboli, A. C.; Schuster, M. G.; Vazquez, J. A.; Walsh, T. J.; Zaoutis, T. E.; Sobel, J. D.; *Clin. Infect. Dis.* **2016**, *62*, e1. [Crossref]
4. Summers, S. A.; Tilakaratne, W. M.; Fortune, F.; Ashman, N.; *Am. J. Med.* **2007**, *120*, 568. [Crossref]
5. Calderone, R. A.; Fonzi, W. A.; *Trends Microbiol.* **2001**, *9*, 327. [Crossref]
6. Sörgo, A. G.; Heilmann, C. J.; Brul, S.; de Koster, C. G.; Klis, F. M.; *FEMS Microbiol Lett.* **2013**, *338*, 10. [Crossref]
7. Almirante, B.; Rodríguez, D.; Park, B. J.; Cuenca-Estrella, M.; Planes, A. M.; Almela, M.; Mensa, J.; Sanchez, F.; Ayats, J.; Gimenez, M.; Saballs, P.; Fridkin, S. K.; Morgan, J.; Rodríguez-Tudela, J. L.; Warnock, D. W.; Pahissa, A.; *J. Clin. Microbiol.* **2005**, *43*, 1829. [Crossref]
8. Albrecht, A.; Felk, A.; Pichova, I.; Naglik, J. R.; Schaller, M.; de Groot, P.; Maccallum, D.; Odds, F. C.; Schäfer, W.; Klis, F.; Monod, M.; Hube, B.; *J. Biol. Chem.* **2006**, *281*, 688. [Crossref]
9. Buerth, C.; Heilmann, C. J.; Klis, F. M.; de Koster, C. G.; Ernst, J. F.; Tielker, D.; *Microbiology* **2011**, *157*, 2493. [Crossref]
10. Chaffin, W. L.; *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 495. [Crossref]
11. Chaffin, W. L.; Lopez-Ribot, J. L.; Casanova, M.; Gozalbo, D.; Martinez, J. P.; *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 130. [Crossref]
12. Abeijon, C.; Chen, L. Y.; *Mol. Biol. Cell.* **1998**, *10*, 2729. [Crossref]
13. Ballou, L.; Hitzeman, R. A.; Lewis, M. S.; Ballou, C. E.; *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 3209. [Crossref]
14. Bates, S.; MacCallum, D. M.; Bertram, G.; Munro, C. A.; Hughes, H. B.; Buurman, E. T.; Brown, A. J. P.; Odds, F. C.; Gow, N. A. R.; *J. Biol. Chem.* **2005**, *280*, 23408. [Crossref]
15. Bates, S.; Hughes, H. B.; Munro, C. A.; Thomas, W. P. H.; MacCallum, D. M.; Bertram, G.; Atrih, A.; Ferguson, M. A. J.; Brown, A. J. P.; Odds, F. C.; Gow, N. A. R.; *J. Biol. Chem.* **2006**, *281*, 90. [Crossref]
16. Find your protein, <http://www.uniprot.org>, accessed in April 2023.
17. RCSB Protein Data bank, <https://www.rcsb.org>, accessed in April 2023.
18. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *The Proteomics Protocols Handbook*; Walker, J. M., ed.; Springer Humana Press: New Jersey, 2005, p. 571. [Crossref]

19. King, R. D.; Sternberg, M. J.; *Protein Sci.* **1996**, *11*, 2298. [Crossref]
20. Rost, B.; Sander, C.; *J. Mol. Biol.* **1993**, *232*, 584. [Crossref]
21. Frishman, D.; Argos, P.; *Protein Eng.* **1996**, *9*, 133. [Crossref]
22. Levin, J. M.; Robson, B.; Garnier, J.; *FEBS Lett.* **1985**, *205*, 303. [Crossref]
23. Geourjon, C.; Deleage, G.; *Protein Eng. Des. Sel.* **1994**, *7*, 157. [Crossref]
24. Geourjon, C.; Deleage, G.; *Comput. Appl. Biosci.* **1995**, *11*, 681. [Crossref]
25. Guermeur, Y.; Geourjon, C.; Gallinari, P.; Deléage, G.; *Bioinformatics* **1999**, *15*, 413. [Crossref]
26. Sen, T. Z.; Jernigan, R. L.; Garnier, J.; Kloczkowski, A.; *Bioinformatics* **2005**, *21*, 2787. [Crossref]
27. Combet, C.; Blanchet, C.; Geourjon, C.; Deléage, G.; *Trends Biochem. Sci.* **2000**, *25*, 147. [Crossref]
28. Edgar, R. C.; *Nucleic Acids Res.* **2004**, *32*, 1792. [Crossref]
29. Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B.; *Structure* **2003**, *11*, 1453. [Crossref]
30. Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J.; *Nucleic Acids Res.* **2003**, *31*, 3701. [Crossref]
31. Labute, P.; *Molecular Operating Environment (MOE)*, 2011. 10; Chemical Computing Group Inc., Montreal, QC, Canada, 2012.
32. Multiple Sequence Alignment, <http://www.ebi.ac.uk/Tools/msa/muscle>, accessed in April 2023.
33. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V.; *RAMPAGE (CCP4: Supported Program), rampage-Ramachandran plots using the Richardsons' data*, Linux Foundation, San Francisco, California, United States, 1963; Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V.; *J. Mol. Biol.* **1963**, *7*, 95. [Crossref]
34. Colovos, C.; Yeates, T. O.; *ERRAT: An Empirical Atom-Based Method for Validating Protein Structures, Online server*, National Health Institute, University of California, USA, 1993-2015; Colovos, C.; Yeates, T. O.; *Protein Sci.* **1993**, *2*, 1511. [Crossref]
35. Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W.; *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5148. [Crossref]
36. Guruprasad, K.; Reddy, B. V. B.; Pandit, M. W.; *Protein Eng.* **1990**, *4*, 155. [Crossref]
37. Kyte, J.; Doolittle, R. F.; *J. Mol. Biol.* **1982**, *157*, 105. [Crossref]
38. Ikai, A. J.; *J. Biochem.* **1980**, *88*, 1895. [Crossref]
39. Needleman, S. B.; Wunsch, C. D.; *J. Mol. Biol.* **1970**, *48*, 443. [Crossref]
40. Sippl, M. J.; *J. Comp. Aid. Mol. Des.* **1993**, *7*, 473. [Crossref]
41. Labute, P.; *J. Comp. Chem.* **2008**, *29*, 1693. [Crossref]
42. Summa, C. S.; Levitt, M.; *PNAS* **2007**, *104*, 3177. [Crossref]
43. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A.; *J. Comp. Chem.* **1986**, *7*, 230. [Crossref]
44. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A.; *J. Am. Chem. Soc.* **1995**, *117*, 5179. [Crossref]
45. Sturgeon, J. B.; Laird, B. B.; *J. Chem. Phys.* **2000**, *112*, 3474. [Crossref]
46. Bond, S. D.; Benedict, J. L.; Laird, B. B.; *J Comp Phys.* **1999**, *151*, 114. [Crossref]
47. Bhagavan, N. V.; *Medical Biochemistry*, 4th ed.; Academic Press: San Diego, 2002.

Submitted: February 11, 2023

Published online: July 27, 2023