J Mol Model (2015) 21: 30
DOI 10.1007/s00894-014-2563-3

ORIGINAL PAPER

MaxMod: a hidden Markov model based novel interface to MODELLER for improved prediction of protein 3D models

Bikram K. Parida · Prasanna K. Panda · Namrata Misra · Barada K. Mishra

Received: 1 August 2014 / Accepted: 14 December 2014 / Published online: 31 January 2015

Abstract Modeling the three-dimensional (3D) structures of proteins assumes great significance because of its manifold applications in biomolecular research. Toward this goal, we present MaxMod, a graphical user interface (GUI) of the MODELLER program that combines profile hidden Markov model (profile HMM) method with Clustal Omega program to significantly improve the selection of homologous templates and target-template alignment for construction of accurate 3D protein models. MaxMod distinguishes itself from other existing GUIs of MODELLER software by implementing effortless modeling of proteins using templates that bear modified residues. Additionally, it provides various features such as loop optimization, express modeling (a feature where protein model can be generated directly from its sequence, without any further user intervention) and automatic update of PDB database, thus enhancing the user-friendly control of computational tasks. We find that HMM-based MaxMod performs better than other modeling packages in terms of execution time and model quality. MaxMod is freely available as a downloadable standalone tool for academic and non-commercial purpose at http://www.immt.res.in/maxmod/.

Keywords Clustal omega · Graphical user interface · Hidden markov model · Homology modeling · Modified residues

Introduction

Recent advancement in high throughput next-generation sequencing technologies has led to an exponential rise in genome sequence databases. However, the significance of the genomic data cannot be gained until functional inferences of these sequences are deciphered. Toward this end, elucidation of protein three-dimensional (3D) structure bears great importance in understanding the mechanism of protein function, its evolutionary features and catalytic activity, all of which can serve as important framework in designing further experimental studies. Keeping in view of the time consuming nature of experimental determination of protein structure, theoretical modeling based on homology is currently the most reliable, rapid, and cost-effective approach for deducing structural properties of sequences and to bridge the ever expanding gap between the number of known protein sequences and the number of structures solved [1]. Homology modeling method predicts the 3D structure of a given protein sequence (target) based primarily on its alignment to one or more proteins of known structure (template) [2]. Although the reliability of this method has been well established in recent years, selection of the most accurate template and correctness of the target-template alignment are still the challenging areas of research.For homologous protein sequences with sequence identity greater than 40 %, the alignment is generally considered to be almost accurate. However, as the overall sequence identity decreases, alignment becomes difficult and subsequently reduces the quality of the final model [1, 3]. Therefore, the choice of sequence alignment strategy plays a more

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2563-3) contains supplementary material, which is available to authorized users.

B. K. Parida · P. K. Panda (✉) · B. K. Mishra
Bioresources Engineering Department, CSIR-Institute of Minerals & Materials Technology, Bhubaneswar 751013, India
e-mail: pkpan@yahoo.com

P. K. Panda
e-mail: pkpanda@immt.res.in

P. K. Panda · N. Misra · B. K. Mishra
Academy of Scientific & Innovative Research, CSIR- CSIR-Institute of Minerals & Materials Technology, Bhubaneswar 751013, India

critical role in generating accurate protein models than the choice of the modeling program, with distinctly improved models obtained by employing the best available sequence alignment technique [4]. The widely used MODELLER program [5] for homology modeling uses standard pairwise comparison methods for template selection and target-template alignment [3]. The subsequently released graphical user interfaces (GUIs) of MODELLER program such as MINT (http://www.bioinf.org.uk/software/mint/), EasyModeller [6], SWIFT MODELLER [7] and PyMod [8] have also implemented pairwise comparison methods into their workflow for comparative protein structure modeling. A brief account of some essential features of these programs with their limitations is presented in Table 1.

Although pairwise comparison methods, which employ a dynamic programming algorithm guarantee an optimal alignment, the intensity and generality of the underlying substitution matrices (PAM and BLOSSUM) limit the reliability of such methods to cases of high sequence identity. On the other hand, alignment in the so called twilight zone (between 15–30 % sequence identity) requires additional information regarding the protein family to which the particular sequence belongs [9]. In the past several years probabilistic inference methods based on profile hidden Markov models (profile HMM) have emerged as an alternative to conventional pairwise alignment methods such as BLAST [10, 11] and FASTA [12] for creating sequence profiles in order to detect more distant remote homologous templates from database [13]. The key factor in HMM algorithm is in computing not just one best-scoring alignment but a sum of probabilities over the entire local alignment ensemble and therefore, contain more information about the sequence family than a single sequence [14, 15]. Furthermore, a number of recent studies have corroborated the principal advantage of profile-profile based alignment in template identification and overall model quality generation [16–18]. Despite these many advantages, implementation of HMM method in homology modeling software and tools is yet to be addressed adequately [13]. Here we describe the development and benchmarking of MaxMod, a unique Microsoft Windows based GUI of MODELLER that integrates HMMER3 program for template identification and Clustal Omega program for sequence alignment. HMMER3

Table 1 Comparison of various GUI's of MODELLER program available for protein homology modeling

Tool name	Features	Limitations	URL/reference
MINT	MINT is the first GUI of the MODELLER program that automates the process of protein homology modeling having basic features.	• It allows only the basic homology modeling functions of MODELLER to be used. It is built on an older version of MODELLER i.e., MODELLER8 with no further improvements in the software after the release of MINT v3.2.	http://www.bioinf.org.uk/software/mint/
EasyModeller	EasyModeller tool is a frontend graphical interface to MODELLER program. In contrast to MINT, EasyModeller performs visualization and optimization of protein structures in addition to the basic homology modeling protocol.	• For multi-template based modeling, users can load up to a maximum of ten templates. Customized features are unavailable for model optimization and refinement. No options are available for template search. Cannot include ligands from templates having modified residues. No option for copying selective ligands.	Kuntal et al., 2010; http://modellergui.blogspot.in/
SWIFT	MODELLER	SWIFT MODELLER was developed later to EasyModeller and follows a step-by-step approach where the flow of the software screen depicts the sequential steps of the homology modeling process. SWIFT MODELLER is capable of performing homology modeling, visualization, optimization, and generation of Ramachandran plots for the developed protein models.	• Requires plotting utilities as prerequisites for drawing evaluation plots. Customized features are unavailable for model optimization and refinement. Unable to model proteins, when templates are having modified residues. No options are available for copying ligands.
Mathur et al., 2011; http://www.bitmesra.ac.in/	swiftmodeller/swift.htm		
PyMod	PyMod developed as a plug-in to PYMOL, is used to perform homology modeling which is comparatively faster than the above mentioned graphical interfaces to MODELLER. It uses BLASTP for protein homology search.	• Unable to model proteins, when templates are having modified residues. No option for copying selective ligands. Lacks a sequential step-by-step approach making it more difficult for end-users.	Bramucci et al., 2012 http://schubert.bio.uniroma1.it/pymod/

makes profile HMM searches as fast as BLAST, while retaining the power of probabilistic inference technology [13]. In conjunction, implementation of Clustal Omega allows fast scalable generation of high quality multiple sequence alignment by using HHalign package of HMMER3 [19]. We believe that MaxMod will make the entire process of protein homology modeling much faster and user-friendly.

Methods

MaxMod has been developed using Visual Studio.NET platform with C# as the programming language for a high degree of flexibility in the development of user interface (UI) and creating an interactive modular system. The UI is built on a multiple document interface (MDI) for effective presentation of different user modules. The input and output (I/O) operations dominate the entire coding architecture for formatting Python scripts and input files of the backend MODELLER program.

The architecture of MaxMod (Fig. 1) consists of three distinct layers, (a) *Presentation layer*: All visual elements of MaxMod including user I/O, job directory management and PDB sequence database update are present in this layer. (b) *Business layer*: This layer contains standard programming features of the .NET framework base class library (BCL) such as collection classes, data type definitions, variables, security and IO operations along with some non-standard features *viz.*, drawing, classes for database interaction, and web support. Business layer takes input from the preceding presentation layer, processes data (formatting of python scripts and preparing inputs for other 3rd party programs) and sends it to the next level. (c) *Data access layer*: This is a virtual layer controlling various 3rd party programs such as HMMER3, Clustal Omega, Jmol, and PROCHECK [20], all of which have been integrated within MaxMod. The other programmes such as MODELLER and Python require pre-installation. The PDB database also resides in this layer for templates search. All processed data and instructions from BCL are received by the 3rd party programs of data access layer and are further executed to display the output in the presentation layer. Based on the above architecture, MaxMod follows a definite workflow as illustrated in Fig. 2.

Submission of protein sequence

The user is required to submit the target protein sequence in RAW format with a job title of a maximum of five characters. If no title is provided, the program assigns a default name (MODEL) to the submitted sequence along with date and time (format: YYYYMMDDHHMMSS) of submission. The job title also represents the working directory name, where the results are saved for accessing at a later time. At this stage the user can select one of the options *viz.*, "search templates", "upload templates" or "express modeling", depending on the requirement (Fig. 3a).

Search templates

The PDB sequence database and "phmmer" program of HMMER3 software suite are packaged together with MaxMod in order to search templates. On selecting the "search templates" option, HMMER3 program executes to find remote homologs from PDB for the target protein sequence and the output is presented in a tabular format outlining the PDB code with chain name of the crystal structure, E-value, bit-score, E-value of domain hits, bit-score of domain hits and percentage of sequence identity. The user can

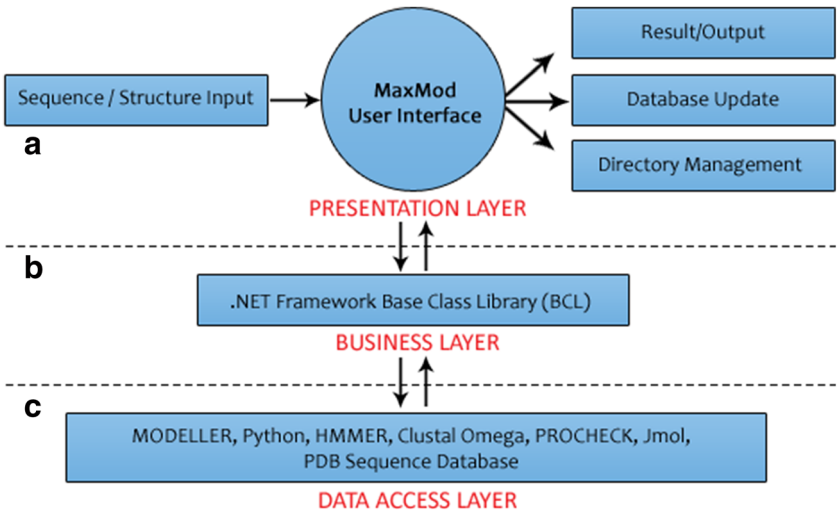

Fig. 1 Architecture of MaxMod demonstrating three different layers of data processing. (**a**) Presentation layer (**b**) Business layer (**c**) Data access layer

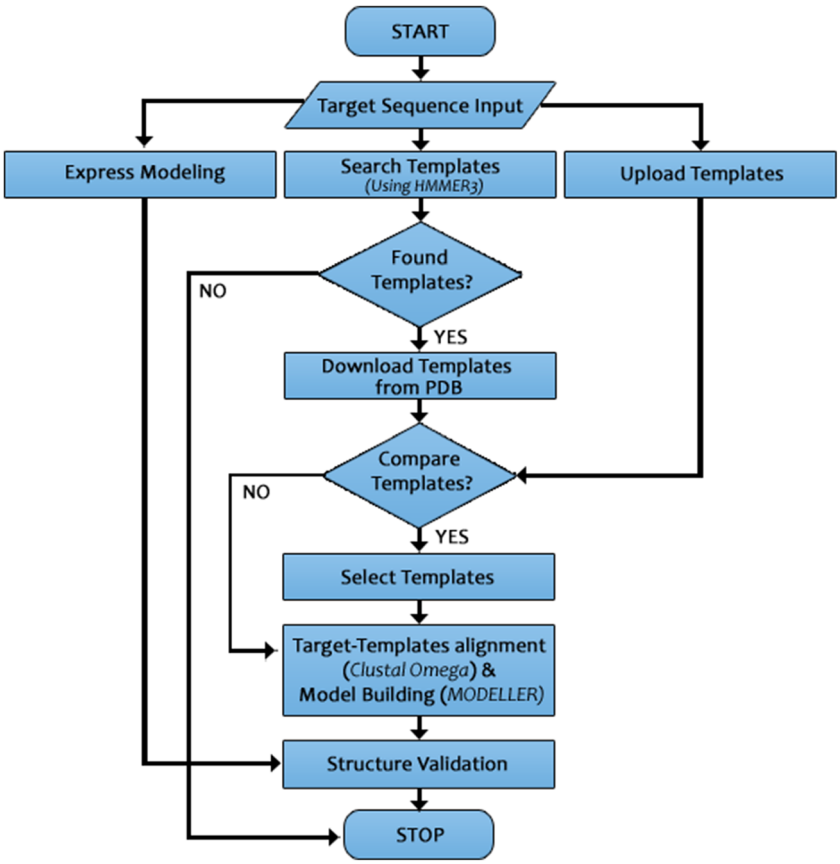

Fig. 2 Workflow of MaxMod for predicting protein 3D model from target sequence

select desired number of templates for viewing more detailed information of the crystal structure available in PDB and their alignment with target sequence. The window will then be directed to RCSB website (www.rcsb.org) for extracting the atomic coordinates of the selected structures (Fig. 3b).

Upload templates

If the "upload templates" option in the homepage is selected, the user will be redirected to a separate window where any number of PDB structures can be uploaded as templates and the appropriate chain can be further chosen from a drop down menu (Fig. 3c).

Compare templates

The user can select the most accurate template by clicking on the "compare templates" option, which performs comparison between the selected templates on the basis of better crystallographic resolution (R-factor) and higher overall sequence identity. MaxMod then displays a dendrogram from the generated log file with their respective R-factor (Fig. 3d).

Model construction and analysis

Successful submission of template structures by exercising any of the options *viz.*,"search templates", "upload templates" and "compare templates", the user will be redirected to the model construction window where template-wise arrangement of ligands are displayed in a tree-view topology. Required ligands may be selected to copy their atomic coordinates onto the modeled structure. Other advanced features are also available in MaxMod such as, "optimization and refinement" where each model is first optimized with the variable target function method with conjugate gradients, followed by its refinement using molecular dynamics with simulated annealing; "rapid optimization" enables the user to get an approximate model very quickly and, the "automatic loop refinement after model building" allows refinement of loop regions after constructing the 3D protein model (Fig. 3e). Selection of the "build model" option after indicating the number of models to be generated will automatically redirect to a new window where 'file name', 'molpdf (molecular probability density function)', and 'discrete optimized potential energy (DOPE) score' are shown in the left panel and options for 'PROCHECK', 'visualization', 'DOPE evaluation', and 'download' are available in the right panel (Fig. 3f). A low 'molpdf' or 'DOPE score' signifies a reliable model.

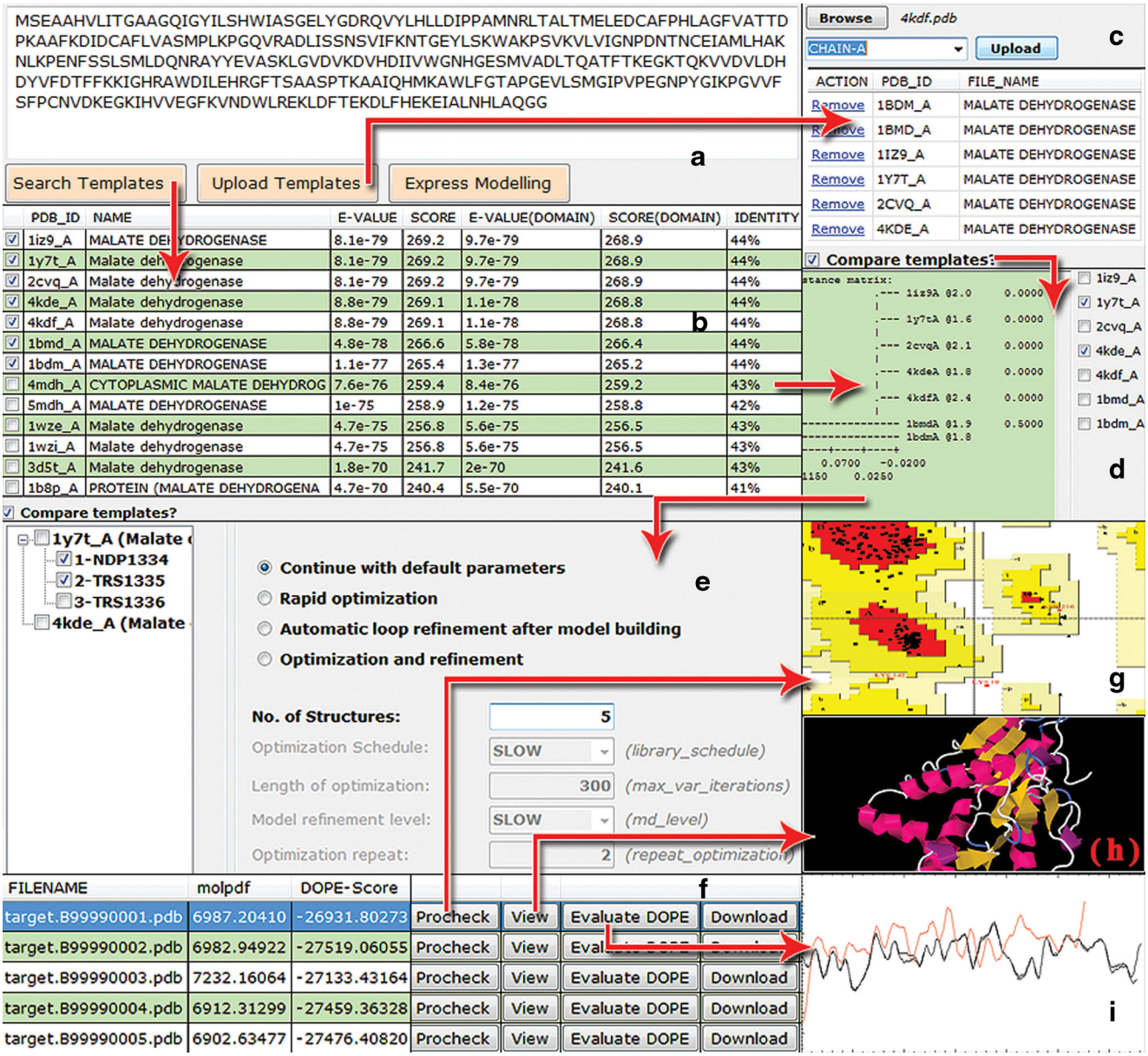

Fig. 3 Screenshots of various windows of MaxMod. (**a**) Sequence input window (**b**) Homologous proteins obtained from PDB using HMMER3 program (**c**) Template upload window (**d**) Window showing R-factor of the selected templates (**e**) Window showing ligands selection from templates and the default/advanced parameters of MODELLER (**f**) Output of the resulting protein models (**g**) Ramachandran plot generated using integrated PROCHECK program (**h**) Model visualization through Jmol (**i**) Residue-wise DOPE profile plot

PROCHECK and Jmol are programs used to generate the Ramachandran plot (Fig. 3g) and visualize 3D conformation of protein, respectively (Fig. 3h).

Express modeling

To make the homology modeling procedure simpler and user-friendly, especially for beginners and non-programmer biologists, another useful feature named "express modeling" option is provided in the home page of MaxMod, where submission of protein sequence in RAW format is the only requirement for building protein 3D model.

Loop optimization

Loops that connect elements of secondary structure for proper protein folding determine the functional specificity of the protein [21]. As a consequence, the accuracy of loop modeling is a crucial component in determining the usefulness of comparative models for studying protein-ligand interactions [22]. In this context we have included a "loop optimization" utility in

Table 2 Comparison of performance and features of MaxMod with other publicly available GUIs of MODELLER program

Activity	EasyModeller 4.0	SWIFT MODELLER 2.0	PyMod 1.0b	MaxMod 1.0
Time taken[a]				
Template search	Not available	49 s	12 s	5 s
Time taken for target-template alignment and model construction	78 s	75 s	57 s	13 s through rapid optimization and 35 s using default parameters
Model validation				
Ramachandran plot[b]	Allowed region: 99.7 % Disallowed region: 0.3 %	Allowed region: 99.7 % Disallowed region: 0.3 %	Allowed region: 99.9 % Disallowed region: 0.0 %	Allowed region: 99.9 % Disallowed region: 0.0 %
Verify-3D[c]	74.01 %	74.01 %	83.33 %	84.82 %
PROSA[d]	−9.09	−9.09	−9.27	−7.46
QMEAN Score[e]	0.676	0.676	0.695	0.689
Specific features				
Template search	Not available	profile.build() command of MODELLER	BLASTP & PSI-BLAST	HMMER3
Target-template alignment	ALIGN2D and SALIGN command of MODELLER for single template and multiple templates, respectively.	ALIGN2D and SALIGN command of MODELLER for single template and multiple templates, respectively.	SALIGN command of MODELLER for both single as well as multiple templates	SALIGN command of MODELLER or Clustal Omega, depending on user selection
Loop optimization	Present	Present	Absent	Present
Automatic loop refinement	Present	Absent	Absent	Present
Visualization	Rasmol	Jmol	PYMOL	Jmol
Model evaluation (DOPE profile plot)	Present	Requires NumPy and Matplotlib as prerequisites	Present	Present
Model validation (Ramac handran plot)	Present	Present	Absent	Present (using Procheck)
Model optimization and refinement feature during model building	Absent	Absent	Absent	Present
Rapid optimization to obtain fast and approximate models	Absent	Absent	Absent	Present
Copying selective ligands	Absent (copies all atomic co-ordinates	Absent	Absent(copies all atomic co-ordinates labeled as	Present

Table 2 (continued)

Activity	EasyModeller 4.0	SWIFT MODELLER 2.0	PyMod 1.0b	MaxMod 1.0
	labeled as "HETATM" which also includes ligands)		"HETATM" which also includes ligands)	
PDB sequence database update and job directory management	Absent	Absent	Absent	Present

[a] calculated for protein sequence lactate dehydrogenase (Uniprot Acc Id: 096445) using the crystal structure of malate dehydrogenase protein from *Thermus thermophilus* HB8 (PDB Id: 1IZ9 chain **a**) as template (E-value: 8.2e-79, Identity: 44 %) for all the programs, on a x86 based 2.80 GHz Pentium D dual-core processor with 1 GB primary memory (RAM) and 80 GB secondary memory (Hard-Disk) space.

[b] The allowed region in the Ramachandran plot corresponds to confirmations in which shorter van der Waal radii used in the calculation. Disallowed regions involve steric hindrance between the side chain group and main chain atoms.

[c] Percentage of residues with VERIFY3D average score >0.2. VERIFY3D is used to analyze the compatibility of the 3D protein model with its own amino acid sequence.

[d] PROSA Z-score indicates overall model quality score and measures the deviation of the total energy of the modeled structure with respect to energy distribution derived from random conformations.

[e] Qualitative model energy analysis (QMEAN) is a composite scoring function describing four major geometrical aspects of protein structure *viz.*, c-beta interaction energy, all atom pairwise energy, solvation energy, and torsion angle. The QMEAN score should fall within the reliability zone of 0 and 1.

MaxMod where PDB structures can be uploaded or obtained directly from the job directory. The user is required to specify the loop region to be refined as well as the number of structures to be generated. The resulting optimized 3D protein models are displayed in a separate window to analyze and download.

Results and discussion

MaxMod is a rich user-friendly standalone tool for protein homology modeling that implements profile HMM method in the modeling framework, unlike other existing GUIs like EasyModeller, SWIFT MODELLER, and PyMod, which employ pairwise comparison methods such as ALIGN2D or SALIGN commands for target-template alignment. The advantage of using profile HMM over pairwise comparison method in MaxMod is that it turns a multiple sequence alignment into a position-specific scoring system which is more suitable for identifying distant homologous relationships. MaxMod can also effortlessly construct protein models using templates bearing modified residues, a feature not present in any other GUIs. Additionally other important features are available such as loop optimization, model validation, and visualization, automated update of PDB database, and express modeling to enable users, to build 3D model by simply submitting the protein sequence.

On comparing MaxMod with other MODELLER-based GUIs with respect to the total time taken to construct 3D model for the protein sequence lactate dehydrogenase (UniProt Acc Id: O96445), it was observed that MaxMod takes around 18 s which is approximately three times faster than PyMod and five times faster than EasyModeller and SWIFT MODELLER (Table 2). The rapid construction of protein model by MaxMod can be attributed to improved template search and target-template alignment using HMMER3 and Clustal Omega programs, respectively. Moreover on assessing the above four modeling programs in relation to their ability to build 3D models with template bearing modified residues, specifically using the crystal structure (PKR kinase domain-eIF2alpha- AMP-PNP complex; PDB Id-2A19) containing a modified residue named phosphothreonine, it was observed that unlike other programs which, either completely failed to construct any model or were unable to copy the atomic coordinates of ligands, MaxMod successfully completed protein modeling without any difficulty. Furthermore, the overall performance of these programs was compared by assessing the stereochemical quality of the various 3D structures generated from modeling a test set of 15 randomly selected proteins, ranging sequences identity from as low as 27 % to as high as 84 % (Table 3). PROCHECK results indicated that all 3D models determined using MaxMod were of better stereochemical quality with

Table 3 Comparative analysis of structure validation results obtained for the homology models determined through MaxMod, and other available GUIs of MODELLER program

Sequence (Uniprot ID)	Template (PDB ID)	E-Value	Sequence identity (%)	Tools	Ramachandran plot		Verify3d (%)[c]	Prosa (Z-score)[d]	Qmean score[e]
					Allowed (%)[a]	Disallowed (%)[b]			
A6SYY7	2XMN_A	7.9E-06	27 %	EasyModeller/SWIFT MODELLER	98.6	1.3	35.29	−5.01	0.414
				PyMod	99.5	0.5	41.88	−3.32	0.236
				MaxMod	99.7	0.3	56.99	−5.61	0.441
E8XAY8	3H90_A	1.4E-15	28 %	EasyModeller/SWIFT MODELLER	98.9	1.1	33.55	−2.43	0.261
				PyMod	99.3	0.7	36.74	−3.9	0.321
				MaxMod	99.6	0. 4	51.78	−3.66	0.409
D0K097	4KOE_A	5.1E-167	32 %	EasyModeller/SWIFT MODELLER	Unable to build model				
				PyMod	99.3	0.7	72.55	−11.43	0.539
				MaxMod	99.5	0.5	77.60	−11.5	0.539
I6W5K9	1WP1_A	7.7E-66	35 %	EasyModeller/SWIFT MODELLER	99.6	0.4	86.62	−7.13	0.672
				PyMod	100	0.0	85.50	−7.50	0.641
				MaxMod	100	0.0	87.08	−5.14	0.625
Q5LTM6	3J1Z_P	4.5E-17	35 %	EasyModeller/SWIFT MODELLER	98.6	1.4	22.43	−3.57	0.336
				PyMod	98.6	1.4	16.82	−3.5	0.417
				MaxMod	98.7	1.3	18.66	−2.69	0.284
K4QHM3	1B9M_A	2.3E-35	36 %	EasyModeller/SWIFT MODELLER	99.6	0.4	86.62	−7.13	0.672
				PyMod	100	0.0	85.50	−7.50	0.641
				MaxMod	100	0.0	87.08	−5.14	0.625
G2SX02	1SR8_A	2.6E-12	38 %	EasyModeller/SWIFT MODELLER	98.9	1.1	33.55	−2.43	0.261
				PyMod	99.1	0.9	43.80	−1.33	0.201
				MaxMod	99.7	0.3	47.47	−1.03	0.284
A0A1F2	1R0K_A	2.2e-100	45 %	EasyModeller/SWIFT MODELLER	99.5	0.5	84	−8.35	0.685
				PyMod	97.8	2.2	83.37	−6.9	0.656
				MaxMod	100	0.0	89.64	−9.58	0.722
A0A168	2P3X_A	3.8e-111	49 %	EasyModeller/SWIFT MODELLER	97.6	2.4	49.24	−1.64	0.409
				PyMod	98.1	1.9	52.44	−2.52	0.396
				MaxMod	98.8	1.2	53.64	−6.99	0.498
G0JM30	1ZK7_A	1.2e-183	61 %	EasyModeller/SWIFT MODELLER	98.2	1.9	78.66	−6.48	0.600
				PyMod	99.4	0.6	75.77	−10.15	0.756
				MaxMod	99.5	0.5	84.38	−10.77	0.761
F0J5U8	1I9D_A	4.2e-56	65 %	EasyModeller/SWIFT MODELLER	100.0	0.0	100.00	−7.37	0.821
				PyMod	100.0	0.0	100.00	−7.37	0.821
				MaxMod	100.0	0.0	100.00	−3.83	0.550
A0A061	2ZA0_A	2.6e-75	67 %	EasyModeller/SWIFT MODELLER	98.7	1.3	83.51	−4.77	0.609

Table 3 (continued)

Sequence (Uniprot ID)	Template (PDB ID)	E-Value	Sequence identity (%)	Tools	Ramachandran plot		Verify3d (%)[c]	Prosa (Z-score)[d]	Qmean score[e]
					Allowed (%)[a]	Disallowed (%)[b]			
				PyMod	98.7	1.3	83.51	−4.77	0.609
				MaxMod	100	0.0	90	−4.39	0.571
A0ZAS2	2Z51_A	6e-13	68.0 %	EasyModeller/SWIFT MODELLER	97.9	2.1	29.86	−2.46	0.275
				PyMod	99.6	0.4	43.17	−2.67	0.249
				MaxMod	100	0.0	56.88	−3.79	0.321
E3HKQ6	1JZW_A	7.7e-60	68 %	EasyModeller/SWIFT MODELLER	99.1	0.9	100.00	−7.48	0.750
				PyMod	99.1	0.9	100.00	−7.48	0.750
				MaxMod	99.9	0.0	100.00	−4.21	0.646
A0A2P5	1OCC_A	0.0	84 %	EasyModeller/SWIFT MODELLER	98.4	1.6	76.88	−5.55	0.542
				PyMod	99.9	0.1	76.88	−5.55	0.542
				MaxMod	100	0.0	82.35	−5.27	0.478

[a] The allowed region in the Ramachandran Plot corresponds to confirmations in which shorter van der Waal radii used in the calculation, that is the atoms that are allowed steric clashes

[b] Disallowed regions involve steric hindrance between the side chain group and main chain atoms and typically occur in turn regions of proteins

[c] Percentage of residues with VERIFY3D average score >0.2. VERIFY3D is used to analyze the compatibility of the 3D protein model with its own amino acid sequence

[d] PROSA Z-score indicates overall model quality score and measures the deviation of the total energy of the modeled structure with respect to energy distribution derived from random conformations . The Z-score of the modeled structure should fall within the range of scores typically found for native protein of similar size

[e] QMEAN(Qualitative Model Energy ANalysis)is a composite scoring function describing four major geometrical aspects of protein structure viz., c-beta interaction energy, all atom pairwise energy, solvation energy and torsion angle. The QMEAN score should fall within the reliability zone of 0 and 1

approximately more than 99 % of residues in the allowed region of Ramachandran plot (Table 3). Furthermore, to check the compatibility of inter-residues interactions, Verify3D [23, 24] tool was employed where the scores indicated that models generated through MaxMod have relatively greater percentage of residues with an average score >0.2, as compared to the models generated by other programs. Similarly, to detect potential errors in the proteins, their Z-score and total energy plots were calculated using ProSA-web program [25]. The Z-score indicates overall model quality and measures the deviation of the total energy of the modeled structure with respect to energy distribution derived from random conformations [26]. The score outside a range characteristic for native proteins indicates erroneous structures. The ProSA energy plot indicated that all the 3D models generated using MaxMod fall within the range of experimentally determined structures (Supplementary Fig. 1). Thus, the overall results (Table 3) conclusively demonstrate the reliability of MaxMod for significant improvement in model accuracy.

Conclusions

MaxMod is a rich user-friendly GUI to the MODELLER program for prediction of protein 3D structures. Its unique strengths are, (i) the use of profile HMM methods such as HMMER and Clustal Omega for template identification and target-template alignment, respectively; (ii) effortless modeling of protein using templates having modified residues (iii) other useful features such as (a) loop optimization, (b) express modeling, (c) model validation, and (d) PDB database update facility. Additionally, the processing time required for model building as well as the overall model quality is significantly improved due to substitution of progressive alignment with profile HMM method. The program runs on any version of Microsoft Windows and we plan to release regular updates, twice annually.

Acknowledgments NM is grateful to Council of Scientific and Industrial Research, Govt. of India for the award of Senior Research Fellowship. The authors would also like to thank the members of CNeM department, CSIR-IMMT for providing server space and hosting the website of MaxMod.

References

1. Cavasotto CN, Patak SS (2009) Homology modeling in drug discovery: current trends and applications. Drug Discov Today 14(13–14): 676–683
2. Barton GJ (1998) Protein Sequence alignment techniques. Acta Cryst D 54:1139–1146
3. Vyas VK, Ukawala RD, Ghate M, Chintha C (2012) Homology modeling a fast tool for drug discovery: current perspectives. Indian J Pharm Sci 74(1):1–17
4. Dalton JA, Jackson RM (2007) An evaluation of automated homology modeling methods at low target-template sequence similarity. Bioinformatics 23((5):1901–1908
5. Sali A, Blundell TL (1993) Comparative protein Ssructure modelling by satisfaction of spacial restrants. J Mol Biol 234(3):779–815
6. Kuntal BK, Aporoy P, Reddanna P (2010) EasyModeller: a graphical interface to MODELLER. BMC Res Notes 3:226–230
7. Mathur A, Vidyarthi AS (2011) SWIFT MODELLER: a JAVA based GUI for molecular modeling. J Mol Model 17(10):2601–2607
8. Bramucci E, Paiardini A, Bossa F, Pascarella S (2012) PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. BMC Bioinformatics 13: Suppl 4:S2
9. Saxena A, Sangwan RS, Mishra A (2013) Fundamentals of homology modelling steps and comparison among important bioinformatics tools: an overview. Sci Int 1(7):237–252
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410
11. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nucleic. Acids Res 25(17)): 3389–3402
12. Pearson WR (2000) (2000) Flexible sequence similarity searching with the FASTA3 program package. Methods Mol Biol 132:185–219
13. Eddy SR (2011) Accelerated Profile HMM Searches. PLoS Comput Biol 7(10):e1002195
14. Krogh A, Brown M, Mian IS, Sjolander K, Haussler D (1994) Hidden Markov models in computational biology: applications to protein modeling. J Mol Biol 235(5):1501–1531
15. Eddy SR (1998) Profile hidden markov models. Bioinformatics 4(9): 755–763
16. Yan R, Xu D, Yang J, Walker S, Zhang Y (2013) A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. Sci Rep 3:2619–2627
17. Sauder JM, Arthur JW, Dunbrack RL Jr (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. Proteins 40(1):6–22
18. Edgar RC, Sjolander KA (2004) Comparison of scoring functions for protein sequence profile alignment. Bioinformatics 20(8):1301–1308
19. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539
20. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
21. Fetrow JS (1995) Omega loops: nonregular secondary structures significant in protein function and stability. FASEB J 9(9)):708–717
22. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Sci 9(9):1753–1773
23. Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253(5016):164–170
24. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356(6364):83–85
25. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucl Acids Res 35:W407–410
26. Sippl MJ (1995) Knowledge-based potentials for proteins. Curr Opin Struct Biol 5(2):229–235