J Mol Model (2012) 18:2971–2979
DOI 10.1007/s00894-011-1320-0

ORIGINAL PAPER

Structure prediction and binding sites analysis of curcin protein of *Jatropha curcas* using computational approaches

Mugdha Srivastava · Shishir K. Gupta · P. C. Abhilash · Nandita Singh

Received: 9 June 2011 / Accepted: 22 November 2011 / Published online: 7 December 2011
© Springer-Verlag 2011

Abstract Ribosome inactivating proteins (RIPs) are defense proteins in a number of higher-plant species that are directly targeted toward herbivores. *Jatropha curcas* is one of the biodiesel plants having RIPs. The Jatropha seed meal, after extraction of oil, is rich in curcin, a highly toxic RIP similar to ricin, which makes it unsuitable for animal feed. Although the toxicity of curcin is well documented in the literature, the detailed toxic properties and the 3D structure of curcin has not been determined by X-ray crystallography, NMR spectroscopy or any *in silico* techniques to date. In this pursuit, the structure of curcin was modeled by a composite approach of 3D structure prediction using threading and *ab initio* modeling. Assessment of model quality was assessed by methods which include Ramachandran plot analysis and Qmean score estimation. Further, we applied the protein-ligand docking approach to identify the r-RNA binding residue of curcin. The present work provides the first structural insight into the binding mode of r-RNA adenine to the curcin protein and forms the basis for designing future inhibitors of curcin. Cloning of a future peptide inhibitor within *J. curcas* can produce non-toxic varieties of *J. curcas*, which would make the seed-cake suitable as animal feed without curcin detoxification.

Keywords Curcin · *Jatropha curcas* · Ribosome inactivating proteins · Structure prediction · Toxicity

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-1320-0) contains supplementary material, which is available to authorized users.

M. Srivastava · S. K. Gupta · P. C. Abhilash · N. Singh
Eco-Auditing Laboratory,
National Botanical Research Institute (CSIR),
Lucknow 226001, Uttar Pradesh, India

M. Srivastava
e-mail: mugdha.srivastava@gmail.com

S. K. Gupta
e-mail: shishir.bioinfo@gmail.com

N. Singh
e-mail: nanditasingh8@yahoo.co.in

P. C. Abhilash (✉)
Institute of Environment & Sustainable Development,
Banaras Hindu University,
Varanasi 221005, Uttar Pradesh, India
e-mail: pca.iesd@bhu.ac.in

Introduction

Jatropha curcas is a small tree belonging to the family Euphorbiaceae [1]. The seeds are highly exploited for biodiesel production. Seed-cake or press-cake is a by-product of oil extraction. Seed-cake of Jatropha contains curcin, a toxic protein mainly reserved in the endosperm of seeds [2], making it unsuitable for animal feed. Besides being a plant defense protein, curcin has an additional role as storage glycoprotein, because they constitute 20% of the total storage-seed proteins in Jatropha [3]. Therefore, the seed-cake of Jatropha may serve as highly nutritious protein supplement in animal feed if the toxins are removed [4].

Felke (1914) was the first to isolate a toxalbumin from seeds of *J. curcas* and he designated it as curcin [5]. Curcin belongs to type-I ribosome inactivating proteins [6] and have been reported for a lethal toxicity in sheep [7], goat [7, 8], rat [9–11], mice [10], chicks [12, 13], calves [14] and human beings, especially in children [10].

The cytotoxicity by curcin is the consequence of its ability to inhibit protein synthesis [15] of intact eukaryotic cells by catalytically damaging ribosomes [6] because of N-glycosidase action, which cleaves the N-glycosidic bond of adenine, making ribosome unable to bind elongation factors 1 or 2, consequently arresting protein synthesis. The breakage of N-glycosidic bond linking the universally conserved position adenine A4324 to the polyphosphate backbone of

the 28S rRNA of the rat liver ribosome has been illustrated as the RIP mediated cleavage action [16, 17].

A limited number of other plant species have been shown to contain type-II RIP's, which make them potentially toxic to animals and to human [18]. Type-II RIP's consist of two amino acid chains connected by disulfide linkage. Chain B is associated with galactose-binding part which binds to the target cell surface glycolipids and glycoproteins allowing chain A (toxin) to enter into the cell by endocytosis [15]. Inside the cell, the toxin is transported via endosomes to the golgi apparatus and the endoplasmic reticulum [19]. The translocated chain A acts as a glycosylase, which removes a highly conserved adenine from 28S ribosomal RNA, and thereby inhibits the protein synthesis. The ultimate result of these reactions is widespread cytotoxicity that can be severe and even fatal [20]. As the velocity of this reaction is much greater than the speed of re-synthesis/repair, in theory, one molecule of RIP would be enough to kill one cell [19]. The mechanism described above has been reported in ricin, a type-II RIP having >50% sequence level similarity of its chain A with curcin [21]. However, not much information about the toxic properties of curcin is currently available in literature. Owing to similarity with ricin it may be assumed that curcin have a similar mode of action to ricin. However, the definite molecular mechanism of curcin toxicity is under investigation and needs to be demonstrated at the structure level for more accurate information and development of future toxicity inhibition studies.

Determining the 3D structure of protein molecules is a cornerstone for many aspects of modern biological research [22]. Bioinformatics and computational biology have added an unavoidable contribution in protein structure prediction. Several methods such as X-ray crystallography, NMR spectroscopy, cryo-electron microscopy and others are available for the structural characterization of proteins however, due to the technical difficulties and labor intensiveness of these methods, the numbers of proteins are frequently modeled by computational techniques by modern researchers to annotate the biological function of a protein molecule whose structure is not available in PDB (Protein Data Bank) library. Historically, protein structure prediction methods have been divided into three broad categories: comparative modeling (CM), threading and *ab initio* modeling [22]. The first two methods are template based and in the absence of complete query coverage the full-length structure can not be determined. The *ab initio* modeling algorithm is free of templates and the structure is built from scratches however, the success of this method is limited to small proteins with<120 amino acids [23, 24]. Recent community-wide critical assessment of protein structure prediction (CASP) experiments [24–27] have demonstrated significant advantages of composite approaches in protein structure prediction, which combine various techniques of protein structure prediction. In the present study, we have implemented such a composite approach of structure prediction for modeling of curcin protein. The interactions of curcin protein with the adenine of r-RNA have been investigated using the docking approaches and possible binding sites are identified.

Materials and methods

Sequence retrieval and 3D structure prediction

The amino acid sequence of curcin from *J. curcas* (Genbank accession no: ACO53803.1) was retrieved from the Entrez protein database available at NCBI (http://www.ncbi.nlm.nih.gov) and is composed of 293 residues. Similarity search of retrieved curcin sequence were performed by BLASTp [28] at NCBI (http://www.ncbi.nlm.nih.gov) with PDB [29] as a reference database to identify the suitable templates for modeling of curcin. It was inspected that no single template with lower e-value and acceptable identity were able to satisfy 100% query coverage. Therefore, to enhance the query coverage we opted for combination of multiple templates. However, the manual inspection confirmed that even the combinations of multiple templates were unable to achieve 100% query coverage. Hence, we used the composite approach of full-length protein structure prediction, implemented in I-TASSER server (http://zhang.bioin-formatics.ku.edu/I-TASSER). The complete methodology of I-TASSER algorithm has been described in details elsewhere [30, 31].

I-TASSER uses restraints from templates identified by multiple threading programs to build a full length model using replica-exchange Monte-Carlo simulations. The unaligned region of target protein is further modeled by *ab initio* modeling. Consequently, the fragment assembly is performed using a modified replica-exchange Monte Carlo simulation technique [32]. Cluster centroids were then obtained by averaging the 3D coordinates of all the clustered structural decoys. Further, to remove steric clashes and to refine the global topology of the cluster centroids, structural analogs were identified and a second round of simulations were performed. The external constraints were pooled from the threading alignments and the PDB structures that were structurally closest to the cluster centroids. The template modeling score (TM-score) calculation [33] was implemented to assess the topological similarity of target and template protein structures

$$\mathrm{TM-score} = \mathrm{Max}\left[\frac{1}{\mathrm{L}}\sum_{i=1}^{L_{ali}}\frac{1}{1+d_i^2/d_0^2}\right], \quad (1)$$

where L is the length of the target protein, L_{ali} is the number of the equivalent residues in two proteins, *di* is the distance of the

i^{th} pair of the equivalent residues between the two structures, which depends on the superposition matrix; the 'max' means the procedure to identify the optimal superposition matrix that maximizes the sum in Eq. 1. The scale $d_o = \sqrt[3]{L-15} - 1.8\,\mathrm{d}$ is defined to normalize the TM-score in a way that the magnitude of the average TM-score for random protein pairs is independent on the size of the proteins.

The decoys generated during the second round of simulations were clustered again, and the lowest energy structures were selected as input to generate the final structural models by building all-atom models from Cα traces through the optimization of hydrogen bonding networks [34]. The quality of the best predicted structure of curcin by I-TASSER was estimated by a confidence score [35] named C-score

$$\mathrm{C-score} = \ln\left(\frac{\mathrm{M}}{\mathrm{M}_{tot}} \times \frac{1}{\langle \mathrm{RSMD} \rangle} \times \frac{1}{7}\sum_{i=1}^{7}\frac{Z(\mathrm{i})}{Z_0(\mathrm{i})}\right), \qquad (2)$$

where M is the number of structure decoys in the cluster, M_{tot} is the total number of decoys generated during the I-TASSER simulations, ⟨RMSD⟩ is the average root mean squared deviation (RMSD) of the decoys to the cluster centroid, Z(i) is the Z-score of the best template generated by i^{th} threading in the seven LOMETS programs and $Z_0(i)$ is a program-specified Z-score cutoff for distinguishing between good and bad templates.

Model correction and validation

Minimum energy arrangements of the atoms correspond to stable states of the system and energy minimization can repair distorted geometries by moving atoms to release internal constraints. Energy minimizations with Tripos force-field [36] were performed on the best model of curcin to make the structure closer to the native by refining the local side chain and protein-backbone packing using SYBYL software.

Estimation of model quality is critically important in protein structure prediction, since ultimately the accuracy of a model determines its suitability for specific biological and biochemical experimental design. The quality assessment of the refined energy minimized curcin model was performed by inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK [37] analysis and QMEAN Z-score [38] estimation using QMEAN server (http://swissmodel.expasy.org/qmean/).

Active site prediction

Protein-protein superposition, brings ligand into the target active site with an orientation similar to the one found in the original ligand-receptor crystal structure [39]. The adenine binding site of curcin was predicted by superposition of curcin with the ricin-adenine complex (PDB id: 2P8N). MatchMaker extension of UCSF Chimera v.1.4.1 [40] was used to analyze the structural conservancy between curcin model and ricin. The structure matching protocol uses the alpha-carbon pairs corresponding to every column in the generated alignment in least-squares fit [41]. The RMSD was calculated using the superimposition between matched pairs

$$\mathrm{RMSD} = \sqrt{\sum_{i}^{N}\frac{[di * di]}{\mathrm{N}}}, \qquad (3)$$

where di is the distance between matched pair i, and N is the number of matched pairs.

The adenine binding residues of ricin were observed by LIGPLOT [42] furthermore the aligned residues at corresponding positions in curcin were identified to explore the adenine binding sites of curcin.

Docking experiments

Docking experiments were performed by using the Auto-Dock 4.2 software suit [43] to identify adenine interacting residues. The Kollman charges, solvation parameters and polar hydrogen were added to the water free curcin structure for the preparation of protein in docking simulation. The Gasteiger charge was assigned to ligand (adenine) and then non-polar hydrogen was merged. AutoDock requires pre-calculated three dimensional grid maps, one for each type of atom present in the ligand and its stores the interaction energy based on a macromolecular target using the AMBER force field. This grid must surround the region of interest in the macromolecule. AutoGrid 4.2 Program, supplied with AutoDock 4.2 was used to generate grid maps for the ligands. The grid box was fixed in such a way that the superimposed residue of adenine binding site of ricin may be involved in docking. The box size in x-, y- and z-axis was set at 60 Å×64 Å×50 Å with grid points separated by 0.375 Å. Lamarckian genetic search algorithm was selected to find suitable binding positions for a ligand on a protein receptor. Although the protein structure has to be fixed, the program allows torsional flexibility of the ligand.

Results and discussion

Structural information of the protein molecule is essential to annotate its biological function. As it was ascertained that the three-dimensional structure of curcin from *J. curcas* was not available in PDB, hence the present exercise of developing the 3D model of the curcin was undertaken. Similarity search revealed the absence of template with >40% sequence identity. Moreover, the combinations of multiple templates were not able to achieve

complete query coverage. No significant similarity was observed in any of the PDB templates against N-terminal sequence of curcin. The N-terminal leader sequence directs the RIPs to the endomembrane system [44, 45], from where the RIPs move to the subcellular compartments such as vacuoles, protein bodies or the periplasmic space [46]. N-terminal leader sequence of curcin it may also contribute in the reliable folding of curcin sequence in tertiary structure. In the N-terminal sequence we also found a secondary structure element and hydrophobic core as a building block for protein folding. It has been shown that removal of N-terminal building block from the structure may contribute in error during protein folding [47–49] as the protein may acquire a non-native stable conformation because of a mis-association of the adjunct building blocks. Further because of the cleavage of N-terminal sequence, the protein folding can be under kinetic control, since it is trapped in a thermodynamically less-stable state [50]. Owing to the significance of N-terminal leader sequence, development of full length model instead of truncated model was essential. Hence, I-TASSER server was used to model the complete sequence of curcin.

I-TASSER [35, 51] is a hierarchical protein structure modeling approach based on the multiple threading alignments and an iterative implementation of the Threading ASSEmbly Refinement (TASSER) program [33]. I-TASSER combines two protein structure prediction methods i.e., threading and *ab initio* prediction. To achieve the full-length model, the complete protein chain in I-TASSER is divided into threading aligned and unaligned regions, where the continuous fragments are excised from threading alignments; while the threading unaligned regions are built by *ab initio* modeling [52]. It has been observed that the average performance differs among different threading algorithms. There is not a single-threading program that can outperform other methods for every target [23] hence, the meta-servers for threading come in a sight that accesses the different servers via a single platform and gives more accurate structures. I-TASSER is a meta-server that implements different threading programs and uses the best multiple templates that result larger query coverage and better topology than the best templates identified by individual servers.

The quality of the threading alignment is usually judged based on the Z-score of the alignment, which is defined as the energy score in standard deviation units relative to the statistical mean of all alignments. Since, the templates are generated by different threading algorithms, a new parameter normalized Z-score is implemented in I-TASSER for quality estimation of threading alignment. The normalized Z-score accounts for the consensus of alignment confidence of multiple threading programs rather than one threading program. An alignment with a normalized Z-score>1 reflects a confident alignment [34]. The top ten threading templates used by I-TASSER along with normalized Z-score are listed in Table 1. The alignment of curcin with threading templates is depicted in supplementary Fig. 1.

Table 1 Threading templates along with normalized Z-score of the threading alignments

Rank	PDB hit	Normalized Z-score
1	3bwhA	2.73
2	2vlcA	3.29
3	1iftA	4.21
4	1hwmA	4.25
5	3bwhA	3.87
6	2vlcA	2.83
7	1br6A	6.87
8	1fmp	5.38
9	1hwmA	3.99
10	2vlcA	3.48

Protein structure similarity is often measured by root mean squared deviation, global distance test score and TM-score [53] among which RMSD calculation between target and template structure is commonly used to compare protein structures by calculating all the equivalent atom pairs after the optimal superposition of the two structures [54]. However, because all atoms in the structures are equally weighted in the RMSD calculation, one of the major drawbacks of RMSD is that it becomes more sensitive to the local structure deviation than to the global topology when the RMSD value is big [53, 55]. In comparison, the TM-score [32] counts all residue pairs using the Levitt–Gerstein weight [56] where the short distance is weighted stronger than the long distance. Hence, the TM-score is more sensitive to the global topology than local variations. TM-score stays in [(0–1)] where a higher value indicates a stronger similarity between structures. The top ten PDB structures of similar topology are listed in Table 2 along with their TM-score.

Table 2 PDB structures structurally closest to model along with TM-score of the structural alignment

Rank	PDB hit	TM-score
1	1br6A	0.8343
2	2vlcA	0.8137
3	1abrA	0.8069
4	3ku0A	0.801
5	1hwnA	0.8
6	2zr1A	0.7997
7	3bwhA	0.7897
8	1momA	0.7896
9	1bryY	0.7894
10	3ctkA	0.7883

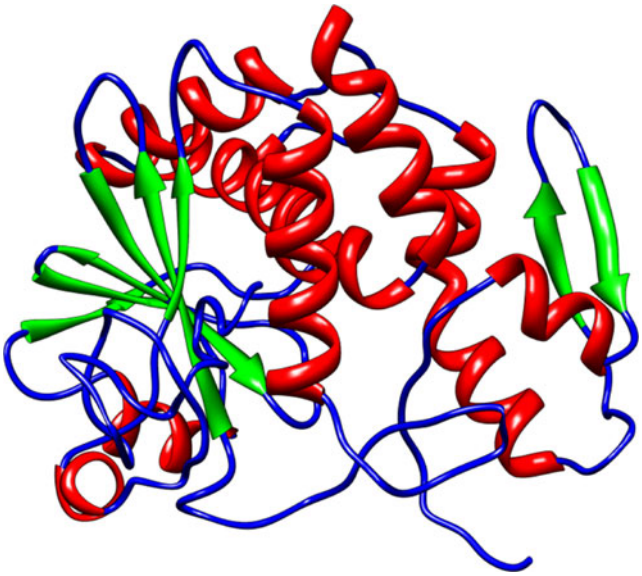

Fig. 1 Schematic representation showing the arrangements of α-helices, β-sheets, and loop regions in the predicted three-dimensional structure of curcin

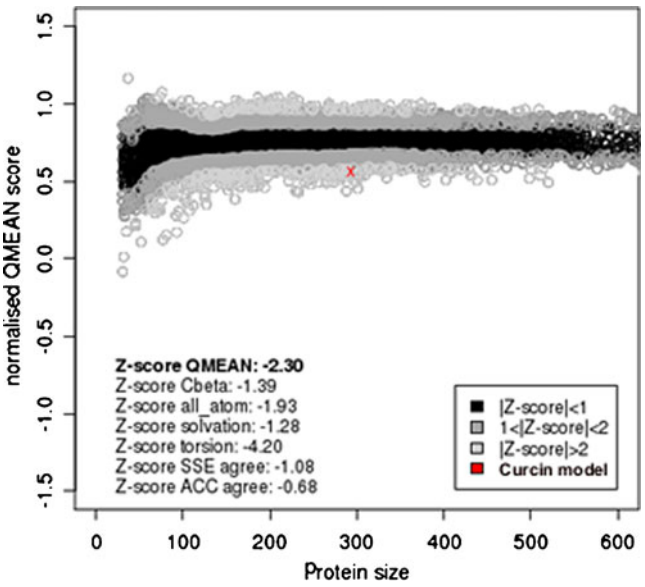

Fig. 3 Model quality estimation plot obtained by QMEAN server. The area built by the circles colored in different shades of gray in the plot represents the QMEAN scores of the reference structures from the PDB

C-score is a confidence score for estimating the quality of predicted models by I-TASSER. C-score is calculated based on the significance of threading template alignments and the convergence parameters of the I-TASSER's structure assembly refinement simulations. C-score of best model of curcin was −0.80 which was acceptable as it lie within the range of reliable models C-score, i.e., -0.5 to +2.0. Figure 1 depicts the 3D structure of curcin.

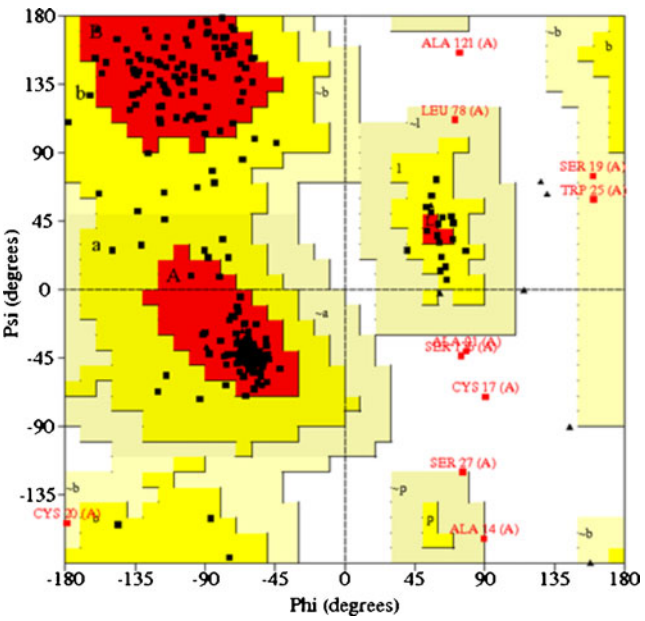

Fig. 2 Ramachandran plot of the curcin model. The most favored regions are colored red, additional allowed, generously allowed and disallowed regions are indicated as yellow, light yellow and white fields, respectively

The molecular refinement procedure implemented in I-TASSER exploits knowledge-based approaches for model corrections. Although it is suggested that a purely physics-based *ab initio* simulation has the advantage in revealing the pathway of protein folding hence the best current free modeling results come from those which combine both knowledge-based and physics-based approaches [23]. Therefore, the best model of curcin was energy minimized with physics-based force fields Tripos using SYBYL. The force-field contains terms associated with bond lengths, angles, torsion angles, van der Waals and electrostatics interactions. Slight RMSD deviation (0.207 Å) was

Fig. 4 Stereo ribbon diagram of superposition of C^{α} atoms of ricin (yellow) and curcin (magenta). Adenine is shown in blue color

Fig. 5 The pairwise sequence alignment of the ricin and curcin. The sequence data of the ricin was extracted from the PDB atom coordinate file (PDB id. 2P8N|A). In the figure the gray boxes denote structurally conserved regions (SCRs) between curcin and ricin

observed by superimposition of minimized and unminimized curcin structure. The per residue RMSD deviation between the unminimized and energy minimized 3D structure of curcin is shown in supplementary Fig. 2.

PROCHECK is a highly reputed program to check the stereochemical quality of protein structure. Backbone conformation evaluation by the inspection of the Psi/Phi Ramachandran plot of the curcin model indicated that only five amino acids (Ala14, Cys17, Ser19, Ser27 and Ala121) had a disallowed geometry (Fig. 2).

Since, these residues were far away from the binding site, thereby their influence on the inferences derived here can be considered negligible. The statistical analysis of calculated

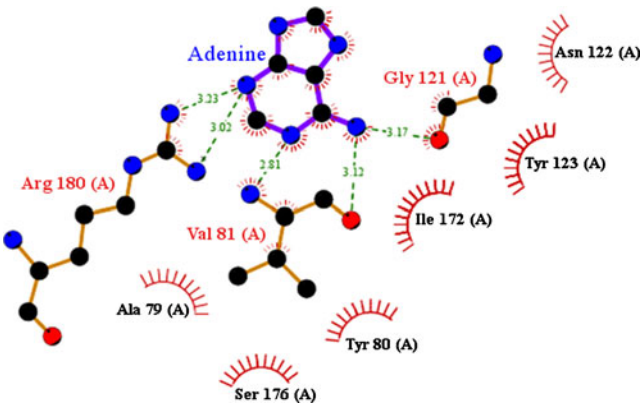

Fig. 6 A schematic diagram of the distributions of hydrophobicity and hydrogen bonds for ricin-adenine complex (PDB entry 1P8N), generated by the program LIGPLOT. The red arcs with radiating spokes represents the amino acids showing hydrophobic interaction with adenine and the green dotted lines represent hydrogen bond interactions. Carbon, nitrogen and oxygen atoms are shown in black, blue and red, respectively

Table 3 Adenine interacting amino acid residues of ricin and the curcin residues at the corresponding position on structure alignment

Ricin	Curcin
Ala79	Ala117
Tyr80	Tyr118
Val81	Leu119
Phe93	Phe131
Gly121	Gly157
Asn122	Ser158
Tyr123	Tyr159
Ser176	Pro208
Glu177	Glu209
Arg180	Arg212

Ramachandran plot suggested that 84.0%, 12.3%, 1.9%, and 1.9% of the residues in derived curcin model were in the most favored, additional allowed, generously allowed and disallowed regions, respectively. Thus, altogether 96.30% of the residues were placed into the favored and allowed categories; therefore PROCHECK validated the folding integrity of the curcin model and indicated that the model structure derived from I-TASSER was of higher quality in terms of protein folding.

Several scoring functions have been developed which are able to return a quality estimate for the protein structure predicted by computational techniques. In order to evaluate a predicted structure comprehensively, it is now common to construct a composite score which combines several features [57]. Since, the combination of broadly orthogonal information has been shown to improve model quality estimation [58], qualitative model energy analysis (QMEAN) model quality estimation was employed that provides an

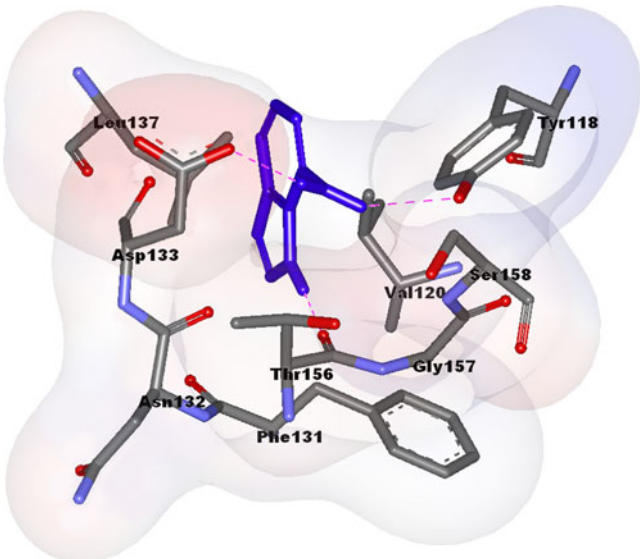

Fig. 7 Adenine binding site residues of curcin. Adenine is shown in blue and the hydrogen bond interactions are depicted by pink dotted lines

Table 4 Various energies and ligand efficiency details of the best conformation of adenine docked into the binding site of modeled curcin for the effective binding site residues prediction in kJ mol^{-1}

Binding energy	Ligand efficiency	Intermol	Total internal	Unbound	Vdw-hb-desolv	Electrostatic
−4.08	−0.41	−4.39	0.03	0.0	−4.23	0.16

estimate of the 'degree of nativeness' of the structural features observed in a model and describes the likelihood that a given model is of comparable quality to experimental structures. QMEAN is a composite scoring function consisting of a linear combination of four statistical potential terms covering the major aspects of protein stability and two additional terms describing the agreement of predicted and calculated secondary structure and solvent accessibility, respectively [59, 60]. QMEAN Z-score provides an estimate of the absolute quality of a model by relating it to reference structures of similar size (model size +/− 10%) deposited in PDB and solved by experimental techniques. Figure 3 depicts the plot for model quality estimation by comparison of curcin model with non-redundant set of PDB structures.

Model with QMEAN Z-score above −3 most probably have the correct fold. The predicted QMEAN Z-score of −2.3 means that the quality/energy of the model deviates by approximately two standard deviations from values expected for structures solved experimentally, which is reasonable for a homology model and indicates that the model has a correct overall fold. The validated model of curcin was further used for the docking studies.

It has been previously suggested that ricin has a very high specificity to adenine as it depurinates adenosine from roughly 7000 nucleotides in eukaryotic ribosomes. This specificity has been attributed either to direct base recognition or to steric hindrance as all the other ricin binding sites are sterically blocked by the folded rRNA structure complexed with proteins [61]. The X-ray diffraction structure, 2P8N [62] is a ricin A-chain complex with adenine which is resolved at 1.94 Å. The overall RMSD deviation of 0.971 Å between ricin and curcin model indicated the curcin structural model was consistent with experimental coordinates of ricin and the packing of helices and sheets (Fig. 4). The pairwise sequence alignment of ricin A-chain and curcin on the basis of their structure alignment is depicted in Fig. 5.

To compare the secondary structures of the two proteins ricin and curcin, Stride program was used. Stride [63] is a program to recognize secondary structural elements in proteins from their atomic coordinates. The program utilizes hydrogen bond energy and main chain dihedral angles. Further, it relies on database-derived recognition parameters with the crystallographers' secondary structure definitions as a standard-of-truth. Whereas ricin includes seven α-helices and nine β-strands, curcin had eight and seven, respectively (Supplementary Table 1).

The amino acid residues in adenine binding site of ricin-A chain structure are depicted in Fig. 6. Docking analysis on curcin structure was performed by Autodock and after scoring of docking results it was observed that residues Tyr118, Val120, Phe131, Asn132, Asp133, Leu137, Thr156, Gly157, and Ser158 were in close contact with adenine. Among these residues the residues at corresponding positions of Tyr118, Phe131, Gly157 and Ser158 were involved in binding with adenine in ricin-adenine complex (Table 3). The residues in close contact with adenine and adenine curcin interactions in docked complex are visualized in Fig. 7 and the energies calculated by Autodock software are listed in Table 4. The binding site residues involved in hydrogen bonding are listed in Table 5.

The conservation of the sequence encoding the binding site within a target family leads putatively also to conservation of the shape and physicochemical properties of the ligand binding sites, resulting in a similarity of the structural requirements for ligands [64]. Among the ten residues in the adenine binding site seven were exactly the same in the structure of both the RIP family members. Further, we found the 0.971 Å RMSD between curcin and ricin structure that interprets the coordinates of both the structures as appreciably similar. Accordingly, structure-activity relationship homology concept of Frhe (1999), the reference set formed by the ligands of the previously investigated target may be assumed to be valid for a whole family of targets [65]. In the adenine binding site of ricin Tyr80 is reputed as a critical target for ricin inhibitors. Most of the good ricin inhibitors displace the Tyr80 ring and bind in the adenine pocket by making specific hydrogen bonds to active site residues for inhibition of ricin [66–68]. Besides, it was also visualized after careful assay that in the proposed curcin structure, tyrosine was at exactly the same corresponding position, i.e., Tyr118 (Table 3) providing

Table 5 Ligand-receptor residues in close contact and residues involved in hydrogen bond formation

Close contact	Ligand (adenine): Adenine	Receptor (curcin) residues: Tyr118, Val120, Phe131, Asn132, Asp133, Leu137, Thr156, Gly157 and Ser158
H-bond formation	Donors: Adenine: H1, Adenine: H51, Adenine: H52	Acceptors: THR156:O, ASP133:OD2, TYR118:OH

the similar H-bond interactions with adenine (Fig. 7). Therefore, considering all the facts, it is possible that the inhibitors of ricin may be used as probable inhibitors for prevention of curcin toxicity.

Structure-based inhibitor design has become increasingly important in the rational process of discovery of new analog and inhibitor compounds. The molecules similar to molecules of a reference set formed by ligands of a previously investigated target can be expected to have some activity also on other targets within the same family [64]. Because of significant structural similarity it is expected that ricin inhibitors may inhibit curcin and both the proteins can be thought of as SAR homologs. Further, using a reference set of ligands of a known target; ligands can be screened and used as probable candidates for any target that is similar to the already known receptor target [69]. Hence, creation of a knowledge based virtual compound database of ricin inhibitors [67], their derivatives [68] and analogs [70] can be examined by using preliminary *in silico* virtual screening approach before being subjected to wet lab experiments for their usage as future perspective of curcin toxicity inhibition.

Bagaria and coworkers (2006) proposed the less toxic nature of RIP agglutinin is because of the fewer adenine interactions with agglutinin than RIP abrin [71]. Moreover, in a recent study Cheng and co-workers (2010) explained the toxicity of agglutinin by its 3D structure determination and comparative study of 28S rRNA's adenine binding site analysis in agglutinin and abrin [72], which supports the methodology applied in this study.

The presence of phorbol ester is another potential cause of Jatropha toxicity. Therefore, it is also essential to remove phorbol ester toxicity before serving seed-cake to animals. Fortunately a variety of Jatropha (*J. mahafalensis*) has been reported in Mexico and Central America, which does not contain toxic phorbol esters and can be a potential source of oil for human consumption, and the seed cake can be a good protein source for humans as well as for livestock after inhibition of curcin toxicity. Therefore, the inhibition of the curcin in *J. mahafalensis* like varieties would enable the Jatropha seed meal to be used as an animal feed without the necessity of detoxification, thus increasing the economic value of the crop.

Conclusions

In the present work the 3D model of curcin was constructed in order to accomplish docking studies to reveal the residues involved in adenine binding. The residues Val81, Gly121 and Arg180 form the critical H-bond with adenine of 28S rRNA in ricin, while the residues at corresponding positions of Val81 and Gly121 in curcin were not in a more compact complementary relationship and the corresponding residue to Arg180 was not even in close contact of adenine in curcin. Moreover, the docking study also revealed that adenine of 28S rRNA forms H-bonds with Tyr118, Thr156 and ASP133 residues of curcin, among which only residues at the corresponding position of Tyr118 were present in close contact with adenine in ricin-adenine complex. Glu177 and Arg180 are particularly important residues for ricin toxicity as Glu177 affects the speed of enzymatic reaction and Arg180 facilitates the breakage of N-glycosidic bond by donating a proton to N3 of adenine from substrate. However, in curcin structure the Glu209 and Arg212 were found at the corresponding positions on structure alignment that were not present in close contact with adenine. Therefore, the fewer interactions involved with the substrate adenine in the curcin-adenine complex than ricin-adenine complex may explain the lower toxicity of curcin than ricin, despite the similarity in structure of both RIPs. Our preliminary findings thus warrant molecular mechanisms of the curcin mediated potential toxicity of *J. curcas* seeds.

Furthermore, the proposed 3D structure of curcin was in the good agreement with experimental structure of high resolution proteins. Therefore, this model could be profitably used to filter off knowledge base compound libraries for potential inhibitors able to target this toxic protein and the findings may be the key step to address the toxicity issues during further studies of these mysterious RIPs.

Acknowledgments We wish to thank Dr. Pascal Benkert (Swiss Institute of Bioinformatics, University of Basel, Switzerland) for helpful discussions regarding Q-mean server. The authors are grateful to the Director of National Botanical Research Institute for his continuous encouragement. Authors are also thankful to Ministry of Environment & Forest, Govt. of India for financial support.

References

1. Jamil S, Abhilash PC, Singh N et al. (2009) J Hazard Mat 172:269–275
2. Mourgue M, Delphaut J, Baret R et al. (1961) Bull Soc Chim Biol (Paris) 43:517–531
3. Lin J, Zhou X, Wang J, Jiang P, Tang Kexuan (2010) Prep Biochem Biotechnol 40:107–118
4. Makkar HPS, Becker K (1997) Potential of Jatropha seedcake as protein supplement in livestock feed and constraints to its utilization. In: Proc Jatropha 97: International symposium on Biofuel and Industrial Products from *Jatropha curcas* and other Tropical Oil Seed Plants. Managua/Nicaragua, Mexico
5. Felke J (1914) Landw Versuchsw 82:427–430
6. Barbieri L, Battelli MG, Stirpe F (1993) Biochem Biophys Acta 1154:237–282
7. Ahmed MM, Adam EI (1979) Res Vet Sci 27:89–96
8. Adam EI, Magzoub M (1975) Toxicology 4:347–354
9. Adam EI (1974) Toxicology 2:67–76
10. Abdu-Aguye I, Sannusi A, Alafiya-Tayo RA, Bhusnurmath SR (1986) Hum Toxicol 5:269–274
11. Goonasekera MM, Gunawardana VK, Jayasena K, Mohammed SG, Balasubramaniam S (1995) J Ethnopharmacol 47:117–123

12. El Badwi SMA, Mousa HM, Adam SEI, Hapke HJ (1992) Vet Hum Toxicol 34:304–306
13. El Badwi SM, Adam SEI, Hapke HJ (1995) Dt Tierärztl Wochenschrift 102:75–77
14. Ahmed MM, Adam EI (1979) Vet Path 16:476–482
15. Devappa RK, Makkar HPS, Becker (2010) J Agric Food Chem 58:6543–6555
16. Luo MJ, Liu WX, Yang XY, Xu Y, Yan F, Huang P, Chen F (2007) Russian J Plant Physiol 54:202–206
17. Huang M-X, Hou P, Wei Q, Xu Y, Chen F (2008) Plant Growth Regul 54:115–123
18. Scientific Opinion of the Panel on Contaminants in the Food Chain on a request from the European Commission on ricin (from Ricinus communis) as undesirable substances in animal feed (2008) The EFSA J 726:1–38
19. Lurie Y, Fainmesser P, Yosef M et al. (2008) Remote Identification of Poisonous Plants by Cell-Phone Camera and Online Communication. IMAJ 10:802–803
20. Palmer M, Betz JM (2006) Plants. In: Flomenbaum NE, Goldfrank LR, Hoffman RS, Howland MA, Lewin NA, Nelson LS (eds) Goldfrank's Toxicological Emergencies, 8th edn. McGraw-hill, New York, pp 1577–1602
21. Juan L, Fang Y, Lin T, Fang C (2003) Acta Pharmacol Sin 24:241–246
22. Zhang Y (2009) Curr Opin Struct Biol 19:145–155
23. Zhang Y (2008) Curr Opin Struct Biol 18:342–348
24. Jauch R, Yeo HC, Kolatkar PR, Clarke ND (2007) Proteins 69:57–67
25. Battey JN, Kopp J, Bordoli L, Read RJ, Clarke ND, Schwede T (2007) Proteins 69:68–82
26. Moult J, Fidelis K, Kryshtafovych A, Rost B, Hubbard T, Tramontano A (2007) Proteins 69:3–9
27. Kopp J, Bordoli L, Battey JN et al. (2007) Proteins 69:38–56
28. Altschul SF, Thomas LM, Alejandro AS et al. (1997) Nucleic Acids Res 25:3389–3402
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
30. Wu S, Skolnick J, Zhang Y (2007) BMC Biol 5:17
31. Zhang Y (2009) Proteins 77:100–113
32. Zhang Y, Kihara D, Skolnick J (2002) Proteins 48:192–201
33. Zhang Y, Skolnick J (2004) Proc Natl Acad Sci USA 101:7594–7599
34. Roy A, Kucukural A, Zhang Y (2010) Nat Protoc 5:725–738
35. Zhang Y (2008) BMC Bioinforma 9:40
36. Clark M, Kramer RD, van Opdenbosch N (1989) J Comput Chem 10:982–1012
37. Laskowski RA, MacArthur MW, Moss DS et al. (1993) J Appl Cryst 26:283–291
38. Benkert P, Biasini M, Schwede T (2011) Bioinformatics 27:343–350
39. Srivastava M, Akhoon BA, Gupta SK et al. (2010) Fungal Genet Biol 47:800–808
40. Pettersen EF, Goddard TD, Huang CC et al. (2004) J Comput Chem 25:1605–1612
41. Meng EC, Pettersen EF, Couch GS et al. (2006) BMC Bioinform 7:339
42. Wallace AC, Laskowski RA, Thornton JM (1995) Protein Eng 8:127–134
43. Morris GM, Goodsell DS, Halliday RS et al. (1998) J Comput Chem 19:1639–1662
44. Lord JM (1985) Eur J Biochem 146:411–416
45. Hartley MR, Lord JM (1993) Structure, function and applications of ricin and related cytotoxic proteins. In: Griesson D (ed) Biosynthesis and Manipulation of Plant Products. Chapman & Hall, New York, pp 210–239
46. Nielsen K, Boston RS (2001) Annu Rev Plant Physiol Plant Mol Biol 52:785–816
47. Ma B, Tsai CJ, Nussinov R (2000) Protein Eng 13:617–627
48. Sham YY, Ma B, Tsai CJ, Nussinov R (2001) Protein Sci 10:135–148
49. Kumar S, Sham YY, Tsai C-J, Nussinov R (2001) Biophys J 80:2439–2454
50. Tsai CJ, Ma B, Nussinov R (2009) Phys Biol 6:013001
51. Zhang Y (2007) Proteins 8:108–117
52. Roy A, Wu S, Zhang Y (2010) Rangwala H, Karypis G (eds) Composite Approaches to Protein Tertiary Structure Prediction: A Case-Study by I-Tasser, in Introduction to Protein Structure Prediction: Methods and Algorithms. Wiley , Hoboken, NJ, USA. doi: 10.1002/9780470882207.ch11
53. Xu J, Zhang Y (2010) Bioinformatics 26:889–895
54. Kabsch W (1978) Acta Cryst A34:827–828
55. Zhang Y, Skolnick J (2005) Nucleic Acids Res 33:2302–2309
56. Levitt M, Gerstein M (1998) Proc Natl Acad Sci 95:5913–5920
57. Kihara D, Chen H, Yang YD (2009) Curr Protein Pept Sci 10:216–28
58. Benkert P, Schwede T, Tosatto SC (2009) BMC Struct Biol 9:35
59. Benkert P, Tosatto SC, Schomburg D (2008) Proteins 71:261–277
60. Benkert P, Künzli M, Schwede T (2009) Nucleic Acids Res 37: W510–W514
61. Corell CC, Munishkin A, Chan YL (1998) Proc Natl Acad Sci 95:13436–13441
62. Carra JH, McHugh CA, Mulligan S (2007) BMC Struct Biol 7:72
63. Heinig M, Frishman D (2004) Nucl Acids Res 32:W500–W502
64. Jacoby E (2011) Quant Struct Act Relat 20:115–123
65. Frye SV (1999) Chem Biol 6:R3–R7
66. Yan X, Hollis T, Svinth M, Day P, Monzingo AF, Milne GW, Robertus JD (1997) J Mol Biol 266:1043–1049
67. Miller DJ, Ravikumar K, Shen H, Suh JK, Kerwin SM, Robertus JD (2002) J Med Chem 45:90–98
68. Bai Y, Monzingo AF, Robertus JD (2009) Arch Biochem Biophys 483:23–28
69. Schuffenhauer A, Floersheim P, Acklin P, Jacoby E (2003) J Chem Inf Comput Sci 43:391–405
70. Robertus JD, Yan X, Ernst S, Monzingo A, Worley S, Day P, Hollis T, Svinth M (1996) Toxicon 34:1325–1334
71. Bagaria A, Surendranath K, Ramagopal UA, Ramakumar S, Karande AA (2006) J Biol Chem 281:34465–34474
72. Cheng J, Lu TH, Liu CL, Lin JY (2010) J Biomed Sci 17:34